

Análise de dados longitudinais

Versão parcial preliminar

junho 2018

Julio M. Singer

Juvêncio S. Nobre

Francisco Marcelo M. Rocha

Departamento de Estatística
Universidade de São Paulo
Rua do Matão, 1010
São Paulo, SP 05508-090
Brasil

Conteúdo

1	Introdução	1
1.1	Conceitos básicos	1
1.2	Exemplos	12
1.3	Análise descritiva e medidas resumo	24
1.3.1	Análise do comportamento de perfis	29
1.3.2	Análise de desfecho	30
1.3.3	Análise da área sob a curva	32
1.4	Exercícios	36
2	Modelos lineares para dados gaussianos	39
2.1	Introdução	39
2.2	Modelos para a estrutura de covariância	42
2.3	Inferência por máxima verossimilhança	46
2.4	Solução das equações de estimação	54
2.5	Diagnóstico	63
2.5.1	Análise de resíduos	64
2.5.2	Análise de influência global	67
2.5.3	Análise de influência local	73
2.6	Notas de capítulo	75
2.7	Exercícios	75
3	Modelos para dados não gaussianos	77
3.1	Modelos elipticamente simétricos e assimétricos	77
3.2	Modelos lineares generalizados mistos (GLMM)	82

3.3	Modelos baseados em equações de estimação generalizadas (GEE) . . .	89
3.4	Modelos para dados categorizados	93
3.5	Modelos não paramétricos para análise de perfis	93
3.5.1	O modelo	93
3.6	Exemplo	98
3.7	Exercícios	103
4	Tópicos especiais	105
4.1	Dados omissos	105
4.2	Planejamento e tamanho de amostras	105
4.3	Modelos não-lineares	105
5	Análise de dados	107
5.1	Estratégias de análise	107
5.2	Implementação computacional	110
5.3	Estudos pré-teste/pós-teste	113
5.4	Análise de perfis	126
5.5	Análise de medidas repetidas	141
5.6	Ajuste de curvas de crescimento	151
5.7	Exercícios	158
A	Matrizes e espaços vetoriais	161
A.1	Matrizes	161
A.1.1	Operações básicas	162
A.1.2	Tipos especiais de matrizes	164
A.1.3	Submatrizes e matrizes particionadas	165
A.1.4	Independência linear e espaço coluna	166
A.1.5	Determinante de uma matriz	167
A.1.6	Inversão de matrizes	168
A.1.7	Traço de uma matriz	170
A.1.8	Soma direta e produto de Kronecker	170
A.1.9	Operadores vec e vech	172

A.2	Tópicos de Álgebra Linear	173
A.3	Formas lineares, bilineares e quadráticas	179
A.4	Decomposição de matrizes	180
A.5	Derivadas de vetores e matrizes	181
A.6	Exercícios	190
B	O método Delta	195
B.1	O caso univariado	195
B.2	O caso multivariado	196
C	Análise de Regressão	199
C.1	Introdução	199
C.2	Método de mínimos quadrados	210
C.3	Método de máxima verossimilhança	215
C.4	Partição da soma de quadrados	216
C.5	Diagnóstico	218
C.5.1	Análise de resíduos	219
C.5.2	Análise da suposição de normalidade	222
C.5.3	Análise de sensibilidade	226
C.5.4	Análise da suposição de correlação nula	233
C.5.5	Multicolinearidade	243
C.6	Parametrização de modelos lineares	244
C.7	Regressão logística	249
C.8	Exemplos	260
C.9	Exercícios	265
	Bibliografia	287

Capítulo 1

Introdução

Neste capítulo caracterizamos estudos com estrutura longitudinal contrastando-os com outros tipos, identificamos os conceitos que permeiam essa modalidade de investigação, introduzimos uma notação apropriada para a descrição dos dados coletados sob esse molde, descrevemos diversos exemplos, identificando suas peculiaridades dentro desse contexto e, finalmente, consideramos alguns métodos básicos para sua análise. Mais especificamente, na Seção 1.1 ocupamo-nos da mencionada caracterização, salientando a dependência entre as observações que distingue esse tipo de estudo e damos os primeiros passos no estabelecimento da notação empregada no texto. Na Seção 1.2 descrevemos um conjunto de exemplos com complexidade crescente, apontando suas particularidades e relacionando-as com os conceitos descritos na seção anterior. Finalmente, na Seção 1.3, apresentamos algumas técnicas simples para a descrição e análise de dados longitudinais.

1.1 Conceitos básicos

Em muitas situações práticas, há interesse em modelar o comportamento de uma ou mais variáveis respostas medidas nas unidades de uma ou mais populações ao longo de alguma dimensão ordenada. Um exemplo simples envolve a modelagem da variação diária da pressão sanguínea de indivíduos normais ou hipertensos durante um certo período. Num campo diferente, um estudo cujo objetivo é estimar as concentrações de partículas em suspensão em diferentes distâncias de uma fonte poluidora constitui um segundo exemplo. Com objetivo simplificador, referir-nos-emos a essa escala ordenada ao longo da qual se fazem as medidas repetidas como **tempo**. Embora o caso geral possa envolver múltiplas variáveis respostas, concentraremos nossa atenção no caso univariado.

Neste contexto, podemos identificar duas grandes estratégias para coleta de da-

dos. A primeira envolve uma única observação (realizada num instante especificado) da variável resposta para cada elemento (pacientes, por exemplo) de uma amostra de cada população de interesse (de indivíduos normais ou hipertensos, por exemplo). A segunda estratégia envolve duas ou mais observações (realizadas em instantes diferentes) da variável resposta em cada unidade amostral sob investigação. No primeiro caso, dizemos que o estudo tem um planejamento **transversal** e no segundo, referimo-nos ao planejamento como **longitudinal**. Em Bioestatística, esta última forma de coleta de dados também é conhecida como **coorte**, ao passo que em outros campos do conhecimento, como Sociologia, Economia ou Administração, ela é cognominada **painel**.

Convém esclarecer que os problemas nos quais temos interesse diferem daqueles usualmente conhecidos sob a denominação de **séries de tempo** ou **séries cronológicas** na medida em que nestes, em geral, uma única unidade amostral é avaliada em muitos (200 ou mais, por exemplo) instantes enquanto que naqueles, várias (5 ou mais, por exemplo) unidades amostrais são observadas em poucas (2 a 20, por exemplo) ocasiões. Para contrastar os dois tipos de estudo, podemos considerar de um lado, a investigação sobre o regime diário de chuvas numa determinada região nos últimos 50 anos, e de outro, a pesquisa sobre os padrões mensais de crescimento de recém-nascidos no primeiro ano de vida. Leitores interessados em análise de séries cronológicas podem consultar Morettin & Tolo (2006), entre outros.

Estudos longitudinais constituem um caso especial daqueles conhecidos sob a denominação de **medidas repetidas**, que englobam os planejamentos com **lotes subdivididos** (*split-plot*) e **com intercâmbio** (*crossover*).

Planejamentos do tipo *split-plot* envolvem dois fatores; as unidades experimentais (*whole-plots*) são aleatoriamente alocadas aos diferentes níveis do primeiro fator e os níveis do segundo fator são aplicados a unidades observacionais (*split-plots*). Num estudo para avaliar o efeito da temperatura e do tipo de matéria prima na consistência de um tipo de bolo, as unidades experimentais (*whole-plots*) são bandejas em que os bolos são colocados para serem levados ao forno e as unidades observacionais (*split-plots*) são bolos com tamanho padronizado. As bandejas são aleatoriamente tratadas sob os diferentes níveis do fator Temperatura e os diferentes níveis do fator Matéria prima são aleatoriamente atribuídos aos bolos de cada bandeja.

O caso mais simples de planejamento com intercâmbio (dois tratamentos em dois períodos) envolve dois grupos de unidades experimentais; o primeiro é submetido a um tratamento *A* no primeiro período e a um tratamento *B* no segundo período; para as unidades experimentais do segundo grupo, a ordem dos tratamentos é invertida (tratamento *B* no primeiro período e tratamento *A* no segundo período).

Tanto no caso de experimentos do tipo *split-plot* quanto no caso de experimen-

tos com intercâmbio, os tratamentos intraunidades experimentais são alocados aleatoriamente às unidades observacionais. A característica que distingue os estudos longitudinais é a dimensão ordenada ao longo da qual as medidas intraunidades experimentais são repetidas. A análise de dados longitudinais tem uma longa história na literatura estatística e tem sido objeto de inúmeros textos, dentre os quais destacamos Goldstein (1979), Duncan & Kalton (1987), Crowder & Hand (1990), Jones (1993), Lindsey (1999), Diggle, Heagerty, Liang & Zeger (2002), Baltagi (1995), Vonesh & Chinchilli (1997), Brunner, Domhof & Langer (2002), Molenberghs & Verbeke (2000), Molenberghs & Verbeke (2005) e Demidenko (2013), entre outros. O leitor interessado em planejamentos do tipo *split-plot* ou em planejamentos com intercâmbio poderá consultar Cochran & Cox (1992), Cox (1992) ou Ratkowsky, Aldredge & Evans (1992), por exemplo.

Estudos longitudinais têm interesse especial quando o objetivo é avaliar tanto mudanças globais quanto individuais ao longo do tempo. Em primeiro lugar, eles permitem que a variável resposta seja observada em unidades amostrais sob níveis constantes de outras covariáveis que podem influenciá-la. No exemplo descrito acima, variações temporais na pressão sanguínea medida num esquema longitudinal deveriam estar (pelo menos parcialmente) livres da influência dos hábitos de vida dos pacientes selecionados; isso não aconteceria num estudo transversal, em que as flutuações cronológicas da pressão são avaliadas com base em diferentes indivíduos. Esta característica é de especial interesse nos casos em que a variabilidade interunidades amostrais é maior do que a variabilidade intraunidades amostrais. Em segundo lugar, planejamentos longitudinais proveem informações sobre variações individuais nos níveis da variável resposta. Finalmente, é importante notar que alguns parâmetros dos modelos estatísticos subjacentes podem ser estimados de forma mais eficiente sob planejamentos longitudinais do que sob planejamentos transversais com o mesmo número de observações.

Consideremos uma situação em que o interesse recai na comparação dos valores esperados de uma certa variável resposta sob duas condições diferentes, por exemplo, na comparação das frequências cardíacas médias de indivíduos em condições normais e sob estresse. Sejam X e Y , respectivamente, essas frequências cardíacas e suponhamos que ambas seguem distribuições normais com uma variância comum σ^2 .

Num contexto transversal, a comparação de interesse seria baseada nos dados de duas amostras independentes de indivíduos com n unidades amostrais cada, digamos (X_1, \dots, X_n) e (Y_1, \dots, Y_n) . A estatística adequada para testar a igualdade dos valores esperados seria $t = (\bar{X} - \bar{Y})/s\sqrt{2/n}$, em que \bar{X} e \bar{Y} denotam as médias amostrais de X e Y , respectivamente, e s^2 representa a variância amostral combinada (*pooled*).

Num contexto longitudinal, a comparação seria baseada na observação da resposta nas mesmas n unidades amostrais, antes (X) e depois (Y) de elas serem

submetidas ao estímulo estressante. Para essa finalidade, utilizar-se-ia a conhecida estatística t -pareada, $t_d = (\bar{X} - \bar{Y})/s_d\sqrt{1/n}$, em que s_d^2 denota a variância amostral de $D = X - Y$, uma estimativa não-enviesada da variância $2\sigma^2 - 2\sigma_{XY}$, com σ_{XY} representando a covariância entre X e Y . Quando σ_{XY} é positiva, esperamos que o denominador de t_d seja menor que o de t e, conseqüentemente, que o teste correspondente seja mais poderoso para detectar diferenças entre as respostas esperadas correspondentes às populações sob investigação.¹

A maior desvantagem dos estudos longitudinais está relacionada com seu custo, pois em muitas situações exige-se um grande esforço para garantir a observação das unidades amostrais nos instantes pré-determinados, e em outras, o período de observação pode ser muito longo. Em muitos ensaios clínicos, por exemplo, é necessário acompanhar os pacientes com extremo cuidado para que cumpram o protocolo experimental e não abandonem o estudo. Os aspectos técnicos também podem ser considerados como uma desvantagem, pois a análise estatística de dados obtidos sob esse tipo de planejamento é, em geral, mais difícil que a análise de dados obtidos sob esquemas transversais.

Essencialmente, os problemas com que nos deparamos no contexto de estudos longitudinais são similares àqueles com que nos defrontamos em estudos transversais. Para dados com distribuições normais, eles podem ser classificados como problemas de Análise de Variância (ANOVA) ou Análise de Regressão (linear ou não linear). A diferença básica entre eles reside numa possível dependência (estatística) entre as observações intraunidades amostrais, presente apenas nos dados provenientes de estudos longitudinais. A consequência prática desse tipo de dependência reflete-se às vezes num fenômeno conhecido como **trilhamento** (*tracking*), segundo o qual, unidades amostrais com níveis de resposta mais altos (ou mais baixos) no início da coleta de observações tendem a manter suas posições relativas ao longo de todo o estudo. O esforço adicional requerido na análise de dados longitudinais relativamente àquele exigido em estudos transversais concentra-se praticamente na modelagem dessa estrutura de dependência estatística.

Em estudos longitudinais, os dados associados a cada unidade amostral podem ser expressos na forma de um vetor cujos elementos são os valores da variável resposta (pressão sanguínea, por exemplo) em cada instante de observação e de uma matriz cujos elementos correspondem aos valores das variáveis explicativas (ou independentes) que podem ter uma natureza classificatória (tratamento, gênero, por exemplo) ou não (tempo, temperatura, etc.). Dentre essas, algumas podem variar entre unidades amostrais (tratamento, gênero, por exemplo) e outras podem ter variação intraunidades amostrais (hábitos tabagistas, exposição à poluentes etc.);

¹Na realidade o teste t pareado pode ser empregado em situações mais gerais em que X e Y têm variâncias σ_X^2 e σ_Y^2 , respectivamente, não necessariamente iguais. Nesse caso s_d^2 é um estimador não-enviesado da variância de $D = X - Y$, nomeadamente $\sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY}$.

estas são cognominadas **covariáveis dependentes do tempo** (*time dependent covariates*).

O vetor com as m_i respostas para a i -ésima unidade amostral ($i = 1, \dots, n$) é conhecido como **perfil individual de resposta** (*individual response profile*) e pode ser expresso como

$$\mathbf{y}_i = (y_{i1}, \dots, y_{im_i})^\top \quad (1.1.1)$$

em que y_{ik} denota o valor de variável resposta para a i -ésima unidade amostral no k -ésimo instante de observação, ($k = 1, \dots, m_i$) e \mathbf{a}^\top denota o vetor \mathbf{a} transposto. A natureza longitudinal desses perfis é caracterizada por sua matriz de covariâncias $\mathbf{\Omega}_i$, com dimensão $(m_i \times m_i)$, para a qual alguma estrutura diferente de $\sigma^2 \mathbf{I}_{m_i}$ é admitida. A matriz $m_i \times p$ com os valores de p variáveis explicativas pode ser expressa como

$$\mathbf{X}_i = (\mathbf{x}_{i1}, \dots, \mathbf{x}_{ip}), \quad (1.1.2)$$

em que $\mathbf{x}_{ij} = (x_{ij1}, \dots, x_{ijm_i})^\top$, com x_{ijk} representando o valor da j -ésima variável independente ($j = 1, \dots, p$) para a i -ésima unidade amostral observada no k -ésimo instante ($k = 1, \dots, m_i$).

Quando o esquema de coleta de dados determina que todas as unidades amostrais devem ser observadas nos mesmos instantes (igualmente espaçados ou não), dizemos que o planejamento é **balanceado com relação ao tempo**. Se, por outro lado, o planejamento determina que conjuntos diferentes de unidades amostrais sejam observados em conjuntos de instantes diferentes [como nos chamados planejamentos **encadeados** (*linked*) ou mais geralmente **transversais mistos** segundo a nomenclatura apresentada por Rao & Rao (1966)] ou permite que os dados sejam coletados irregularmente ao longo do tempo, ele é dito **desbalanceado com relação ao tempo**. Tanto neste caso quanto naqueles planejados de forma balanceada mas para os quais existem **observações omissas**, os dados são ditos desbalanceados com relação ao tempo. Um resumo dos principais conceitos empregados em pesquisas sobre dados longitudinais bem como uma tentativa de padronização da terminologia correspondente podem ser encontrados em Helms (1992).

Para efeito de análise, dados de estudos longitudinais devem ser dispostos de forma a conter uma indicação da unidade amostral a que os valores das variáveis respostas e variáveis explicativas estão associados. Na Tabela 1.1.1 apresentamos um paradigma para a disposição de um conjunto de dados em que as diferentes unidades amostrais são observadas em conjuntos de instantes de tempo possivelmente diferentes. Em cada unidade amostral, tanto a resposta (Y) quanto as variáveis explicativas (X , W , V e Z) são observadas em cada instante. As variáveis X , W e V representam características fixas da unidade amostral e por isso são covariáveis independentes do tempo ao passo que a variável Z é uma covariável dependente do tempo pois seus valores não são constantes ao longo do tempo. Num estudo clínico,

por exemplo, X e W poderiam representar o sexo e o tipo de droga administrado aos participantes, caracterizando fatores no sentido usual de ANOVA. A variável V poderia corresponder à idade dos pacientes, caracterizando uma covariável no espírito de Análise de Covariância (ANCOVA). A variável Z , por sua vez, poderia indicar a quantidade de gordura animal ingerida pelos participantes entre instantes consecutivos avaliação, justificando a classificação acima designada.

Tabela 1.1.1: Uma estrutura para disposição de dados longitudinais

Unidade amostral	Resposta	Tempo	Covariáveis			
			X	W	V	Z
1	y_{11}	t_{11}	x_1	w_1	v_1	z_{11}
1	y_{12}	t_{12}	x_1	w_1	v_1	z_{12}
.
1	y_{1m_1}	t_{1m_1}	x_1	w_1	v_1	z_{1m_1}
2	y_{21}	t_{21}	x_2	w_2	v_2	z_{21}
2	y_{22}	t_{22}	x_2	w_2	v_2	z_{22}
.
2	y_{2m_2}	t_{2m_2}	x_2	w_2	v_2	z_{2m_2}
.
n	y_{n1}	t_{n1}	x_n	w_n	v_n	z_{n1}
n	y_{n2}	t_{n2}	x_n	w_n	v_n	z_{n2}
.
n	y_{nm_n}	t_{nm_n}	x_n	w_n	v_n	z_{nm_n}

Neste caso, num estudo em que a variável resposta é a nota em um certo tipo de teste aplicado a adolescentes em diferentes ocasiões, o fator X representa um método de ensino (*e.g.*, Método A = 0 e Método B = 1), W representa o sexo (Feminino = 0 e Masculino = 1), a variável V indica a idade e a variável Z representa o número de horas dedicadas ao estudo num certo período, a matriz (1.1.2) associada a uma adolescente com 14 anos submetida ao método de ensino B seria especificada como

$$\mathbf{X}_i = \begin{pmatrix} t_{i1} & 1 & 0 & 14 & z_{i1} \\ t_{i2} & 1 & 0 & 14 & z_{i2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ t_{im_i} & 1 & 0 & 14 & z_{im_i} \end{pmatrix}.$$

Quando os fatores X e W têm mais do que dois níveis podemos representá-los por meio do acréscimo de colunas à matriz \mathbf{X}_i . Com objetivos computacionais, é comum concatenar os perfis de resposta e as matrizes de variáveis explicativas individuais,

respectivamente, na forma de um vetor com dimensão $(N \times 1)$, em que $N = \sum_{i=1}^n m_i$,

$$\mathbf{y} = (\mathbf{y}_1^\top, \dots, \mathbf{y}_n^\top)^\top$$

e de uma matriz com dimensão $(N \times p)$,

$$\mathbf{X} = (\mathbf{X}_1^\top, \dots, \mathbf{X}_n^\top)^\top.$$

Quando os dados são balanceados em relação ao tempo (com $m_i = m$) e não há covariáveis contínuas nem covariáveis dependentes do tempo, também é possível apresentá-los no formato da Tabela 1.1.2 para facilitar certo tipo de análises clássicas que discutiremos adiante. Para efeito de simplificação, admitimos que o fator X tem 3 níveis e que o fator W tem apenas 2 níveis. Na tabela, y_{hijk} corresponde ao valor da variável resposta para j -ésima unidade amostral submetida ao h -ésimo nível do fator X , i -ésimo nível do fator W no tempo k , $h = 1, \dots, 3$, $i = 1, 2$, $j = 1, \dots, n_{hi}$ e $k = 1, \dots, m$.

Neste contexto, é comum dispor os perfis de resposta

$$\mathbf{y}_{hij} = (y_{hij1}, y_{hij2}, \dots, y_{hijm})^\top$$

na forma de uma matriz com dimensão $(n \times m)$ em que $n = \sum_{h=1}^3 \sum_{i=1}^2 n_{hi}$, dada por

$$\mathbf{Y} = (\mathbf{y}_{111}, \mathbf{y}_{112}, \dots, \mathbf{y}_{11n_{11}}, \dots, \mathbf{y}_{321}, \mathbf{y}_{322}, \dots, \mathbf{y}_{32n_{32}})^\top.$$

As correspondentes matrizes individuais de variáveis explicativas, por outro lado, são geralmente agrupadas na forma de uma matriz com dimensão $(n \times 6)$ dada por

$$\mathbf{X} = \text{diag}(\mathbf{X}_{11}, \mathbf{X}_{12}, \mathbf{X}_{21}, \mathbf{X}_{22}, \mathbf{X}_{31}, \mathbf{X}_{32})$$

com $\mathbf{X}_{hi} = \mathbf{1}_{n_{hi}}$, ou compactamente, $\mathbf{X} = \oplus_{h=1}^3 \oplus_{i=1}^2 \mathbf{1}_{n_{hi}}$.

A forma mais simples de análise de dados longitudinais envolve o uso de técnicas de ANOVA paramétrica ou não paramétrica tendo como variável resposta algum tipo de **medida resumo** uni ou bivariada do perfil de respostas (que é essencialmente multivariado). Exemplos típicos dessas medidas resumo são a **área sob a curva** ou o **desfecho** (*endpoint*). Esse enfoque incorpora a estrutura de correlação intraunidades amostrais da mesma forma que as diferenças consideradas em problemas analisados por meio de testes t pareados o fazem. Ele é útil quando o objetivo é a comparação das populações sob investigação com respeito a alguma característica da distribuição da variável resposta e não inclui uma avaliação de seu comportamento longitudinal. O leitor poderá consultar Rowell & Walters (1976) ou Bryant & Gillings (1985) para uma revisão de estratégias de análise que empregam esse enfoque. Algumas delas serão revistas na Seção 1.3.

Alternativamente, modelos lineares ou não lineares podem ser empregados para avaliar a relação entre a variável resposta e as variáveis explicativas. Esses modelos podem ser classificados como **populacionais médios** (*population-averaged*) ou **individuais** (*subject-specific*). Modelos populacionais médios são aqueles em que a atenção está focada no valor esperado da resposta (entre todos os indivíduos da população) e modelos individuais são aqueles em que o interesse recai nas respostas de cada indivíduo. O leitor poderá consultar Zeger, Liang & Albert (1988) para detalhes sobre o assunto, embora as diferenças entre as duas classes possa ser esclarecida com os exemplos apresentados na próxima seção.

Tabela 1.1.2: Uma estrutura para disposição de dados longitudinais balanceados em relação ao tempo

Fatores		Unidade amostral	Instantes de avaliação			
X	W		1	2	...	m
1	1	1	y_{1111}	y_{1112}	...	y_{111m}
1	1	2	y_{1121}	y_{1122}	...	y_{112m}
.
1	1	n_{11}	$y_{11n_{11}1}$	$y_{11n_{11}2}$...	$y_{11n_{11}m}$
1	2	1	y_{1211}	y_{1212}	...	y_{121m}
1	2	2	y_{1221}	y_{1222}	...	y_{122m}
.
1	2	n_{12}	$y_{12n_{12}1}$	$y_{12n_{12}2}$...	$y_{12n_{12}m}$
.
2	1	1	y_{2111}	y_{2112}	...	y_{211m}
2	1	2	y_{2121}	y_{2122}	...	y_{212m}
.
2	1	n_{21}	$y_{21n_{21}1}$	$y_{21n_{21}2}$...	$y_{21n_{21}m}$
2	2	1	y_{2211}	y_{2212}	...	y_{221m}
2	2	2	y_{2221}	y_{2222}	...	y_{222m}
.
2	2	n_{22}	$y_{22n_{22}1}$	$y_{22n_{22}2}$...	$y_{22n_{22}m}$
3	1	1	y_{3111}	y_{3112}	...	y_{311m}
3	1	2	y_{3121}	y_{3122}	...	y_{312m}
.
3	1	n_{31}	$y_{31n_{31}1}$	$y_{31n_{31}2}$...	$y_{31n_{31}m}$
3	2	1	y_{3211}	y_{3212}	...	y_{321m}
3	2	2	y_{3221}	y_{3222}	...	y_{322m}
.
3	2	n_{32}	$y_{32n_{32}1}$	$y_{32n_{32}2}$...	$y_{32n_{32}m}$

Sob outra perspectiva, modelos para dados longitudinais podem ser classificados como **incondicionais** ou **condicionais**. Os primeiros são aqueles em que o valor esperado da variável resposta, $E(y_{ik})$, é expresso exclusivamente em termos das variáveis explicativas x_{i1k}, \dots, x_{ipk} . Se, entre elas, o tempo é tomado como uma variável discreta que indica a ordem em que a resposta é observada em cada unidade amostral sob um planejamento balanceado (possivelmente com dados omissos), os modelos correspondentes são conhecidos como **modelos de perfis** e são equivalentes àqueles costumeiramente considerados em ANOVA ou ANCOVA. Nos casos em que o tempo é encarado como uma variável contínua, *i.e.*, em que o interesse recai na sua relação funcional com a variável resposta, os modelos correspondentes são designados **modelos de crescimento** ou **curvas de crescimento**. Modelos condicionais, por outro lado, são aqueles em que a relação entre a variável resposta e as variáveis explicativas num certo instante é condicionada a valores prévios da resposta. Por exemplo,

$$y_{ij} = \mathbf{x}_{ij}^\top \boldsymbol{\beta} + \gamma(y_{i,j-1} - \mathbf{x}_{i,j-1}^\top \boldsymbol{\beta}) + e_{ij},$$

em que y_{ij} denota a resposta da unidade i no instante j , \mathbf{x}_{ij} denota o correspondente vetor com os valores das variáveis explicativas, $\boldsymbol{\beta}$ e γ indicam parâmetros e e_{ij} representa um erro aleatório associado.

Duas abordagens distintas podem ser consideradas para o processo de construção desses modelos. A primeira é essencialmente orientada pelos dados e considera funções polinomiais para descrever a relação entre a variável resposta e as variáveis explicativas. Talvez seja este o enfoque mais comumente empregado na prática, principalmente em razão da simplicidade das técnicas utilizadas na análise. Apesar de que em muitos casos esse tipo de modelos não incorpora o verdadeiro mecanismo gerador dos dados, eles podem ser satisfatórios dentro de certos limites. No que tange aos parâmetros de localização (ou de regressão), esse enfoque não difere muito daquele usualmente adotado na análise de dados com estrutura transversal; o componente distintivo é obtido por meio do relaxamento da suposição de independência para as observações intraunidades amostrais. Um exemplo típico envolve a relação linear entre uma medida anatômica (distância entre a fissura pterigomaxilar e o centro da glândula pituitária) e idade (com quatro níveis: 8, 10, 12 e 14 anos) de crianças de ambos os sexos num estudo ortodôntico descrito no artigo clássico de Potthoff & Roy (1964). Não fosse pelo fato de as medidas serem tomadas em cada criança aos 8, 10, 12 e 14 anos, o planejamento associado a esse estudo seria classificado como cruzado com dois fatores (sexo e idade) e poderia ser analisado por intermédio de uma ANOVA. A incorporação de covariâncias não nulas no modelo exige uma análise por meio de **ANOVA para medidas repetidas**.

A segunda é fundamentada em considerações teóricas sobre o fenômeno investigado e geralmente produz modelos baseados em equações diferenciais. Modelos gerados por esse tipo de enfoque são usualmente não lineares e os parâmetros cor-

respondentes (frequentemente da ordem de dois ou três) descrevem características importantes do objeto de estudo, como taxas de crescimento ou assíntotas. Eles são conhecidos como **modelos mecanísticos**. Um exemplo para o crescimento de animais (em termos de peso) é o modelo de Brody (1945), segundo o qual a taxa de crescimento do animal no instante t é proporcional à diferença entre o peso na maturidade e o peso nesse instante. O modelo é obtido da solução da equação diferencial

$$\frac{dy(t)}{dt} = k[A - y(t)], \quad (1.1.3)$$

em que $y(t)$ representa o peso no instante t , A é o peso na maturidade e k é a constante de proporcionalidade. A solução de (1.1.3) é

$$y(t) = A[1 - B \exp(-kt)],$$

com $B = 1 - y(0)/A$.

Em muitos casos, as equações diferenciais são baseadas no processo de **Ornstein-Ühlenbeck** que envolve soluções de equações do tipo

$$dY_i(t)/dt = -\alpha Y_i(t) + \sigma dB_i(t)/dt \quad (1.1.4)$$

em que $Y_i(t)$ representa a resposta para a i -ésima unidade amostral no tempo t , α é uma constante positiva e $B_i(t)$ representa um **movimento browniano** com variância infinitesimal σ^2 . Segundo este modelo, a taxa de variação da resposta é proporcional ao valor da resposta em cada instante. O leitor interessado poderá consultar Izeman & Williams (1989), Dwyer (1992), Davidian & Giltinan (1995) ou Vonesh & Chinchilli (1997), para maiores detalhes.

Por muitos anos, quase todos o desenvolvimento de técnicas estatísticas para a análise de dados longitudinais teve seu foco na órbita de modelos univariados para respostas contínuas, com especial concentração naqueles cuja distribuição subjacente é gaussiana. A facilidade do tratamento matemático aliada à propriedade de que a dependência entre as observações intraunidades amostrais pode ser facilmente modelada por meio dos momentos de segunda ordem (variâncias e covariâncias) talvez sejam as maiores razões para isso. Além disso, esses momentos de segunda ordem, em muitos casos, podem ser modelados independentemente dos momentos de primeira ordem, facilitando a análise. Mais recentemente, métodos estatísticos para **modelos lineares generalizados mistos** têm sido considerados para a análise de dados com distribuição na família exponencial, **modelos lineares mistos elípticos** vêm sendo empregados para dados com distribuições elípticamente simétricas e **modelos lineares mistos assimétricos** para dados com distribuições elípticas assimétricas. Detalhes podem ser obtidos em Arellano-Vale, Bolfarine & Lachos (2005), Lachos, Ghosh & Arellano-Valle (2010), Pinheiro, Liu & Wu (2001) e Savalli, Paula & Cysneiros (2006), entre outros. Casos em que apenas a relação entre a variância e a

média é especificada sem menção à forma da distribuição podem ser analisados por meio de **modelos baseados em equações de estimação generalizadas** (GEE) com matriz de correlação de trabalho.

1.2 Exemplos

A estrutura mais simples para estudos com medidas repetidas é aquela correspondente a dados usualmente analisados por intermédio do bem conhecido teste *t*-pareado. Essencialmente, esse tipo de estudo envolve medidas de alguma resposta de interesse nas mesmas unidades amostrais sob duas condições diferentes, *e.g.*, tratamento com placebo ou com droga ativa em ensaios clínicos. Quando essas duas condições de avaliação são ordenadas no tempo, *e.g.*, antes e depois de uma certa intervenção, pode-se classificar o estudo como longitudinal.

Exemplo 1.2.1: Os dados da Tabela 1.2.1 foram extraídos de um estudo realizado na Faculdade de Medicina da Universidade de São Paulo em que um dos objetivos era avaliar o efeito da ingestão de sódio na pressão arterial de pacientes hipertensos. Mais detalhes sobre o estudo podem ser encontrados em Singer & Magalhães (1998).

Tabela 1.2.1: Pressão arterial sistólica (mm Hg)

Paciente	Baixo teor de sódio	Alto teor de sódio	Paciente	Baixo teor de sódio	Alto teor de sódio
1	138	143	11	156	151
2	147	154	12	117	116
3	146	147	13	157	154
4	154	147	14	143	149
5	142	157	15	127	126
6	156	158	16	134	138
7	134	164	17	112	115
8	146	156	18	144	159
9	143	151	19	117	124
10	175	182	20	128	125

Este é um estudo com medidas repetidas, planejamento balanceado em relação ao “tempo” e dados completos. Há apenas um nível (hipertensos) para o único fator (tratamento) interunidades de investigação e dois níveis (baixo teor de sódio e alto teor de sódio) associados ao único fator (tipo de dieta) intraunidades de investigação. Embora sua estrutura seja similar a de um estudo longitudinal com as mesmas características, convém lembrar que a medida realizada com a dieta de alto teor de sódio poderia ter sido avaliada antes daquela com baixo teor de sódio. Essa permutabilidade das condições de avaliação obviamente não é possível em estudos longitudinais. Embora isso não tenha implicações nos casos em que existem apenas duas condições de avaliação, em casos mais gerais, essa permutabilidade pode tor-

nar certos modelos (uniforme, por exemplo) para a estrutura de covariância mais aceitáveis que outros (autorregressivo, por exemplo).

Exemplo 1.2.2: Os dados apresentados na Tabela 1.2.2 são oriundos de um estudo realizado na Escola de Educação Física e Esporte da Universidade de São Paulo cujo objetivo era avaliar o efeito de um programa de treinamento no desempenho de idosos relativamente a certas atividades físicas controlando o sexo (m=masculino, f=feminino). Em particular, focamos nossa atenção no menor tempo (s) gasto para calçar meia no pé de preferência em duas tentativas. Detalhes podem ser obtidos em Elia & Okazaki (1998).

Tabela 1.2.2: Tempo gasto para calçar meia (s)

Sexo	Treinamento		Gênero	Treinamento	
	Antes	Depois		Antes	Depois
m	6.02	4.99	f	6.77	4.97
m	5.30	5.37	f	4.80	3.89
m	8.56	5.85	f		4.44
m	5.15	5.64	f	6.87	8.59
m	4.96	6.29	f	9.86	7.99
m	11.03	10.39	f		3.93
m	9.20	6.77	f	7.54	3.77
m	10.00	5.53	f	6.34	5.04
m	9.99	8.74	f	4.72	3.88
m		5.63	f	6.04	5.11
m		8.47	f	8.12	6.64
m	6.96	6.66	f	10.38	7.26
f	4.54	8.33	f	8.01	7.79
f	5.86	5.41	f	6.39	4.95
f	5.28	4.40	f	8.87	8.08
f	6.47	5.74	f	6.48	4.08
f	5.89	7.45	f	6.98	7.51
f	4.93	4.96	f		8.41
f	5.51	5.76	f	15.84	9.68
f	4.81				

Este é um estudo longitudinal (pois as medidas são realizadas de forma não permutável) com planejamento similar àquele do Exemplo 1.2.1. As diferenças residem na existência de dois níveis (masculino e feminino) para o único fator interunidades de investigação (sexo) e de dados omissos, o que permite dizer que os dados são desbalanceados em relação ao tempo.

Exemplo 1.2.3: Na Tabela 1.2.3 apresentamos dados de um estudo realizado

no Laboratório Experimental de Poluição Atmosférica da Faculdade de Medicina da Universidade de São Paulo para avaliar os efeitos de agentes oxidantes no sistema respiratório e analisado em Singer & Andrade (2000). Espera-se que a exposição a maiores concentrações de agentes oxidantes possa causar danos crescentes às células ciliares e excretoras de muco, que constituem a principal defesa do sistema respiratório contra agentes externos. Cinquenta e seis palatos de sapos foram equitativamente e aleatoriamente alocados a um de seis grupos; cada grupo de 8 palatos foi imerso por 35 minutos numa solução de peróxido de hidrogênio numa concentração especificada, nomeadamente 0, 1, 8, 16, 32 ou 64 μM . A variável resposta de interesse é a velocidade de transporte mucociliar relativa (mm/s), definida como o quociente entre a velocidade de transporte mucociliar num determinado instante e aquela obtida antes da intervenção experimental. Essa variável foi observada a cada cinco minutos após a imersão.

Este estudo generaliza aquele apresentado no Exemplo 1.2.2 em duas direções: o número de níveis do único fator (concentração peróxido de hidrogênio) interunidades de investigação é 6 em vez de 2 e o número de níveis do único fator (tempo após imersão na solução oxidante) intraunidades de investigação é 7 em vez de 2. Ele é um estudo longitudinal com planeamento balanceado em relação ao tempo e com omissão de dados.

Tabela 1.2.3: Velocidade de transporte mucociliar relativa

Concentração de H_2O_2 (μM)	Tempo (min)						
	5	10	15	20	25	30	35
0	1.24	1.21	1.16	1.08	1.08	0.87	
0	1.47	0.91	1.21	1.35	1.26	1.18	
0	1.14	0.90	0.81	1.00	1.05	0.93	
0	0.84	1.02	0.86	0.98	0.98	1.07	
0	0.96	1.19	1.33	1.37	1.22	1.41	
0	0.84	0.89	0.81	0.84	1.05	0.95	
1	1.12	1.22	1.14	1.22	1.46	1.76	1.72
1	1.07	1.36	1.49	1.49	1.46	1.39	1.46
1	0.84	0.75	0.58	0.59	0.52	0.43	0.57
1	1.77	1.28	1.32	1.53	1.30	1.32	1.44
1	0.92	1.16	1.06	1.28	0.97	1.28	1.24
1	1.27	1.42	1.45	1.59	1.55	1.42	1.53
1	1.05	1.09	0.98	1.28	0.98	1.30	1.33
1	0.88	1.02	0.82	0.74	0.75	0.75	0.71
1	1.13	1.24	1.13	1.10	1.17	0.86	1.06
1	0.89	0.84	0.93	1.02	0.93	0.89	0.86
8	0.81	1.18	0.99	1.08	1.00	1.09	1.09
8	1.69	1.62	1.36	1.60	2.00	1.60	1.86
8	0.89	0.85	0.88	0.82	0.95	1.09	1.18
8	1.10	1.57	1.32	1.56	1.83	1.83	1.86
8	1.84	2.36	2.25	2.25	2.05	2.08	2.13
8	1.25	1.22	0.88	0.73	0.66	0.58	0.57
8	0.90	1.05	0.90	0.88	0.86	0.85	0.92
8	1.12	1.06	0.99	0.92	0.91	1.01	1.00
8	1.43	1.25	1.19	1.12	1.13	1.12	1.18
8	1.26	1.24	1.20	1.13	1.03	1.08	1.13
16	0.88	0.73	0.61	0.58	0.61	0.48	0.52
16	0.87	0.69	0.50	0.42	0.38	0.38	0.44
16	1.43	0.98	0.70	0.54	0.43	0.41	0.43
16	1.30	1.00	0.67	0.68	0.53	0.67	0.62
16	0.92	0.86	0.88	0.85	0.79	0.66	0.82
16	1.21	1.01	0.75	0.79	0.69	0.70	0.82
16	0.68	0.86	0.86	0.59	0.57	0.61	0.66
16	0.75	0.74	0.96	0.87	1.00	0.90	1.14
16	0.97	0.84	0.71	0.83	0.67	0.73	0.72
16	1.08	1.12	0.86	1.02	0.85	0.90	0.89

Tabela 1.2.3: Velocidade de transporte mucociliar relativa

Concentração de H_2O_2 (μM)	Tempo (min)						
	5	10	15	20	25	30	35
32	0.97	0.67	0.49	0.46	0.38	0.35	0.29
32	0.60	0.54	0.71	0.33	0.29	0.27	0.36
32	1.37	0.83	0.65	0.56	0.41	0.35	0.40
32	1.45	1.43	0.99	0.92	0.58	0.51	0.46
32	0.92	0.58	0.39	0.24	0.23	0.29	0.23
32	1.12	0.96	0.80	0.71	0.80	0.90	1.00
32	0.79	0.84	0.75	0.69	0.53	0.53	0.65
32	1.02	1.08	0.92	0.79	0.64	0.62	0.57
32	0.83	0.66	0.55	0.54	0.48	0.45	0.54
32	0.72	0.48	0.63	0.57	0.63	0.53	0.54
64	0.12	0.09	0.22	0.11	0.16	0.11	0.09
64	0.48	0.48	0.48	0.41	0.33	0.39	0.30
64	0.20	0.47	0.27	0.34	0.26	0.30	0.44
64	0.79	0.74	0.86	0.63	0.67	0.86	0.92
64	0.38	0.66	0.94	0.84	0.82	0.88	0.89
64	0.94	0.84	0.64	0.53	0.48	0.50	0.50
64	0.17	0.54	0.67	0.66	0.66	0.81	0.96
64	0.96	1.01	1.23	0.93	0.87	0.81	0.81
64	0.62	0.90	0.81	0.86	0.66	0.84	0.78
64	0.23	0.40	0.25	0.16	0.27	0.27	0.47

Exemplo 1.2.4: Os dados da Tabela 1.2.4 são provenientes de um estudo realizado na Faculdade de Odontologia da Universidade de São Paulo para avaliar o efeito do uso contínuo de uma solução para bochecho no pH da placa bacteriana dentária. Com essa finalidade, o pH da placa dentária retirada de 21 voluntários antes e depois do uso de uma solução para bochecho foi avaliado ao longo de 60 minutos após a adição de sacarose ao meio em que as unidades experimentais foram colocadas. Detalhes podem ser obtidos em Grande, Oliveira, Singer, Santos & Nicolau (1998).

Este também é um estudo balanceado com relação ao tempo e com dados completos. A diferença básica entre estes estudos e aqueles apresentados anteriormente deve-se à existência de dois fatores intraunidades de investigação: período de avaliação com dois níveis (antes e depois do uso da solução para bochecho) e tempo após a adição de sacarose com 7 níveis (0, 5, 10, 15, 30, 45 e 60 minutos).

Exemplo 1.2.5: Os dados da Tabela 1.2.5 provêm de um estudo realizado na Faculdade de Odontologia da Universidade de São Paulo cujo principal interesse era avaliar o efeito da irradiação de “laser” na microdureza média do esmalte dentário

Tabela 1.2.4: pH da placa bacteriana dentária

Voluntário	Período	Tempo (min)						Período	Tempo (min)							
		0	5	10	15	30	45		60	0	5	10	15	30	45	60
1	Antes	6.3	6.2	6.0	5.8	5.5	5.3	5.1	Depois	6.4	6.3	6.1	5.9	5.6	5.4	5.2
2	Antes	6.7	6.5	6.4	6.4	6.2	6.2	5.9	Depois	6.3	6.1	6.0	5.9	5.7	5.5	5.3
3	Antes	6.3	6.3	6.2	6.3	5.9	5.7	5.5	Depois	6.4	6.1	5.9	5.8	5.4	5.1	5.0
4	Antes	6.2	6.1	5.9	5.8	5.6	5.4	5.2	Depois	6.3	6.0	5.9	5.8	5.4	5.2	5.1
5	Antes	6.6	6.5	6.5	6.4	6.2	6.1	5.9	Depois	6.2	6.1	6.0	5.9	5.7	5.4	5.3
6	Antes	6.2	5.8	5.7	5.5	5.2	5.1	5.0	Depois	6.3	6.0	5.9	5.8	5.5	5.3	5.2
7	Antes	6.4	6.3	6.2	6.1	6.0	5.9	5.8	Depois	6.4	6.2	6.0	5.9	5.6	5.3	5.1
8	Antes	6.5	6.3	6.1	6.0	5.9	5.7	5.5	Depois	6.7	6.3	6.2	6.1	5.8	5.6	5.4
9	Antes	6.4	6.3	6.3	6.2	6.1	6.1	6.0	Depois	6.3	6.1	6.0	5.9	5.7	5.5	5.3
10	Antes	6.4	6.2	6.1	5.9	5.8	5.7	5.5	Depois	6.3	5.9	5.6	5.5	5.2	5.0	4.9
11	Antes	6.2	6.1	5.9	5.9	5.7	5.6	5.4	Depois	6.3	6.2	6.1	6.0	5.8	5.6	5.5
12	Antes	6.2	5.9	5.7	5.5	5.2	5.1	5.0	Depois	6.4	6.0	5.8	5.6	5.2	5.0	4.9
13	Antes	6.5	6.3	6.1	6.0	5.9	5.7	5.6	Depois	6.4	6.2	6.0	5.8	5.5	5.3	5.2
14	Antes	6.3	6.0	5.7	5.6	5.2	5.1	4.9	Depois	6.5	6.4	6.3	6.1	6.0	5.7	5.7
15	Antes	6.5	6.0	6.3	6.2	6.1	5.9	5.9	Depois	6.3	6.0	5.7	5.6	5.3	5.1	5.0
16	Antes	6.5	6.2	6.2	6.0	5.9	5.7	5.7	Depois	6.3	6.0	5.9	5.7	5.4	5.2	5.0
17	Antes	6.3	6.2	6.1	5.9	5.8	5.7	5.6	Depois	6.4	5.9	5.7	5.5	5.2	5.5	4.9
18	Antes	6.5	6.3	6.2	6.0	5.8	5.7	5.7	Depois	6.3	6.0	5.8	5.7	5.4	5.2	5.1
19	Antes	6.7	6.6	6.5	6.3	6.2	6.1	6.1	Depois	5.9	5.8	5.6	5.5	5.4	5.2	5.1
20	Antes	6.6	6.1	5.8	5.6	5.4	5.3	5.3	Depois	6.4	6.0	5.7	5.5	5.1	4.8	4.7
21	Antes	6.6	6.4	6.3	6.3	6.3	6.2	6.2	Depois	6.4	6.2	6.0	5.9	5.6	5.5	5.4

de dentes expostos a desafios cariogênicos. Três blocos de esmalte dentário foram confeccionados a partir de cada um de 32 terceiros molares humanos extraídos por indicação ortodôntica, totalizando 96 blocos. Esses blocos foram alocados aleatoriamente a um de três tratamentos de forma balanceada (32 blocos por tratamento), a saber: Controle extra bucal (controle), Controle *in situ* (esmalte) e Laser (laser). Os blocos submetidos aos tratamentos Laser e Controle *in situ* foram colocados em aparelhos intrabuciais devidamente fabricados para cada um de oito voluntários e usados durante 28 dias consecutivos para sofrerem a ação de agentes cariogênicos. Durante esse tempo, os blocos submetidos ao tratamento Controle extra bucal não sofreram ação de agentes cariogênicos. Após esse período, os blocos dos três tratamentos foram recolhidos e medidas da microdureza em diferentes profundidades (10, 20, 40, 60, 120 e 180 μm) em cada bloco foram observadas em triplicata. Para efeito do estudo, trabalhou-se com as médias das triplicatas. Mais detalhes podem ser encontrados em Korytnicki, Mayer, Daronch, Singer & Grande (2006).

Este também é um estudo longitudinal em que a dimensão ao longo da qual as observações são realizadas é a profundidade. No entanto, ele também tem outra dimensão que o caracteriza como um estudo com medidas repetidas, pois tanto os três níveis do fator Tratamentos (laser, controle e esmalte) quanto os blocos (dentes) são alocados intraunidades experimentais (voluntários).

Tabela 1.2.5: Microdureza do esmalte dentário (em KHN)

Voluntário	Dente	Tratamento	Profundidades (μm)						
			10	20	40	60	80	120	180
1	1	laser	239	304	282	300	308	330	318
1	1	controle	162	135	166	207	290	289	266
1	1	esmalte	271	297	301	302	304	313	314
1	2	laser	275	331	310	268	272	316	367
1	2	controle	193	168	178	196	275	317	330
1	2	esmalte	288	295	302	308	309	310	302
1	3	laser	196	301	296	286	256	318	318
1	3	controle	159	165	165	255	295	308	319
1	3	esmalte	289	296	319	318	302	316	324
1	4	laser	251	339	285	299	317	297	299
1	4	controle	189	175	162	218	293	273	287
1	4	esmalte	280	311	305	311	308	296	313
2	5	laser	265	335	292	308	284	294	311
2	5	controle	164	202	199	241	287	333	301
2	5	esmalte	289	301	304	311	309	317	322
2	6	laser	268	337	300	290	311	311	309
2	6	controle	175	191	151	234	285	317	307
2	6	esmalte	325	305	316	309	308	322	312
2	7	laser	246	307	323	270	278	319	310
2	7	controle	178	193	172	264	298	310	310
2	7	esmalte	282	273	297	305	305	309	318
2	8	laser	266	347	326	298	290	300	339
2	8	controle	170	192	159	216	317	301	311
2	8	esmalte	289	288	292	290	299	298	325
.
8	29	laser	242	381	295	306	312	339	313
8	29	controle	167	171	205	261	296	294	296
8	29	esmalte	283	313	296	309	308	322	322
8	30	laser	248	350	306	322	344	351	328
8	30	controle	159	159	197	246	277	302	303
8	30	esmalte	279	303	295	311	317	310	319
8	31	laser	237	322	313	281	303	326	355
8	31	controle	135	173	158	250	292	289	300
8	31	esmalte	286	302	277	300	302	313	314
8	32	laser	265	328	323	335	309	325	353
8	32	controle	170	153	148	243	302	285	286
8	32	esmalte	287	299	306	307	317	317	310

Exemplo 1.2.6: Na Tabela 1.2.6 encontram-se dados obtidos de um estudo realizado na Faculdade de Medicina da Universidade de São Paulo com o objetivo de estudar a evolução de um conjunto de características cardiocirculatórias em recém-nascidos pré-termo e comparar os valores correspondentes a 40 semanas da concepção com aqueles obtidos de recém-nascidos a termo (com 40 semanas de gestação). Um objetivo secundário era comparar a evolução desses parâmetros obtidos de recém-nascidos pré-termo com peso adequado para a idade gestacional (AIG) ou pequeno para a idade gestacional (PIG). Maiores detalhes sobre o estudo podem ser obtidos em Afiune (2000) ou Afiune, Singer & Leone (2005). Para efeito didático, concentramos a atenção no diâmetro sistólico da aorta por unidade de peso.

Tabela 1.2.6: Diâmetro sistólico da aorta por unidade de peso (mm/kg)

Grupo	Semanas pós-concepção														
	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
AIG	7.3	9.4	10.3		8.7	8.2	6.7	6.1		5.6		5.1			4.9
AIG					6.1	6.1	6.2		5.4	5.2	4.9				
AIG					5.8	6.3	5.8	5.1	4.9	4.6					
AIG				9.7	9.2		9.5		7.3		6.1	5.4	4.8		4.5
AIG							6.4	5.8	5.2	4.7					
AIG						5.4		4.9	4.6	4.3					
AIG			8.3	8.5	8.6		7.9		6.2		5.5		4.2		
AIG			7.7	8.6	7.9			6.6		5.7					
AIG									5.9	6.1		5.4			4.1
AIG								7.0	6.5						
AIG					5.2			4.8	4.2	4.1		3.7	.		
.
.
AIG					6.2	6.1		6.2			6.0		5.3		
PIG							7.2	6.8		5.5			4.7		
PIG								7.1	8.0	7.7			6.5	5.6	
PIG						7.4	8.3	9.4	10.0	9.2			8.0		
PIG							7.7		6.6		5.5		4.6		
PIG									6.5					4.4	
PIG					7.6	8.6	9.3	8.0		6.6		5.0		4.7	
PIG					6.6	8.4		8.2	7.6		6.6				
PIG							7.1	6.3			6.1	5.9	5.7		4.8
PIG						8.5	8.4					4.9			
PIG				8.3	7.4		6.2			4.6		3.8			
PIG				9.8			9.1		7.3		5.3				
.
.
PIG		8.5													
PIG				10.9	10.7	9.4		8.0			5.8		4.9		

Aqui temos um estudo com um fator interunidades amostrais com dois níveis (AIG e PIG). O planejamento é longitudinal com dados colhidos irregularmente ao

longo do tempo.

Exemplo 1.2.7: Os dados da Tabela 1.2.7 são oriundos de um estudo cujo objetivo era avaliar a eficácia de um tratamento para infecção urinária no que concerne ao desaparecimento de um de seus sintomas. Cinquenta pacientes com esse tipo de infecção foram examinadas em três instantes: no dia de início do tratamento e 14 e 21 dias após essa primeira avaliação. A característica observada foi o nível de corrimento vaginal, classificado como ausente (0), leve (1), moderado (2) ou intenso (3). Observações omissas, bastante comuns neste tipo de problema, estão representadas por pontos.

Mais especificamente, as questões de interesse são: *i)* avaliar se a distribuição da resposta se altera favoravelmente com o tratamento, *i.e.*, se apresenta menores frequências de pacientes com corrimento vaginal de maior intensidade 14 dias após o início do tratamento e *ii)* em caso afirmativo, saber se o tratamento pode ser interrompido após 14 dias, ou seja, se alguma característica relevante (*e.g.*, a proporção de pacientes com corrimento vaginal moderado ou intenso) das distribuições da resposta se mantém inalterada da segunda para a terceira avaliação.

Neste caso estamos diante de um estudo longitudinal com apenas um fator intraunidades amostrais (Avaliação) com três níveis (inicial, após 14 dias e após 21 dias). A natureza categorizada ordinal da resposta (Nível de corrimento vaginal) é a característica que distingue este exemplo daqueles descritos anteriormente.

Exemplo 1.2.8: Na Tabela 1.2.8 estão resumidos os dados de um estudo realizado na Faculdade de Odontologia da Universidade de Mogi das Cruzes, SP, para avaliar o efeito de dois adesivos dentinários (*Single Bond*) e (*Prime bond NT*) e de duas condições de aplicação (dentina seca ou úmida) na variação (pré e pós-operatória) da sensibilidade dentinária (presente ou ausente) de pacientes submetidos a um certo tipo de tratamento odontológico. Para maiores detalhes, o leitor deve consultar Singer & Polli (2004).

Aqui, deseja-se saber se há mudança na distribuição da sensibilidade dentinária após a restauração e se o resultado depende do tipo de adesivo e da condição da dentina durante a sua aplicação. Neste exemplo, além do aspecto categorizado da resposta (Sensibilidade dentinária) podemos salientar a presença de dois fatores interunidades amostrais (Material, com 2 níveis) e Dentina (seca ou úmida) e de um fator intraunidades amostrais (Presença de sensibilidade dentinária) com dois níveis (pré e pós-operatória).

Os exemplos descritos acima ilustram o tipo de problemas que pretendemos avaliar nesta obra. Nos capítulos subsequentes não só discutiremos as técnicas comumente empregadas para analisá-los como também concretizaremos suas análises. Para discutir aspectos mais específicos do tema abordado, outros exemplos serão apresentados e analisados ao longo do texto.

Tabela 1.2.7: Nível de corrimento vaginal em três avaliações

Paciente	Avaliação			Paciente	Avaliação		
	inicial	14 d	21 d		inicial	14 d	21 d
1	1	0	0	26	2	0	0
2	2	0	.	27	2	3	.
3	1	0	0	28	3	0	1
4	2	0	0	29	2	2	1
5	2	1	.	30	2	0	0
6	2	.	.	31	3	.	0
7	2	.	.	32	0	.	0
8	1	1	1	33	1	1	0
9	3	0	0	34	1	0	0
10	2	1	2	35	1	0	0
11	2	1	3	36	1	1	0
12	1	1	0	37	0	0	1
13	2	0	0	38	0	0	1
14	2	0	0	39	1	0	0
15	2	1	1	40	2	0	0
16	2	1	1	41	1	1	0
17	2	1	0	42	1	0	0
18	1	0	0	43	2	.	.
19	1	0	0	44	2	2	.
20	2	0	0	45	2	0	1
21	1	1	0	46	2	.	0
22	3	1	0	47	3	1	0
23	3	0	0	48	3	0	0
24	2	1	1	49	2	1	1
25	2	0	0	50	3	0	0

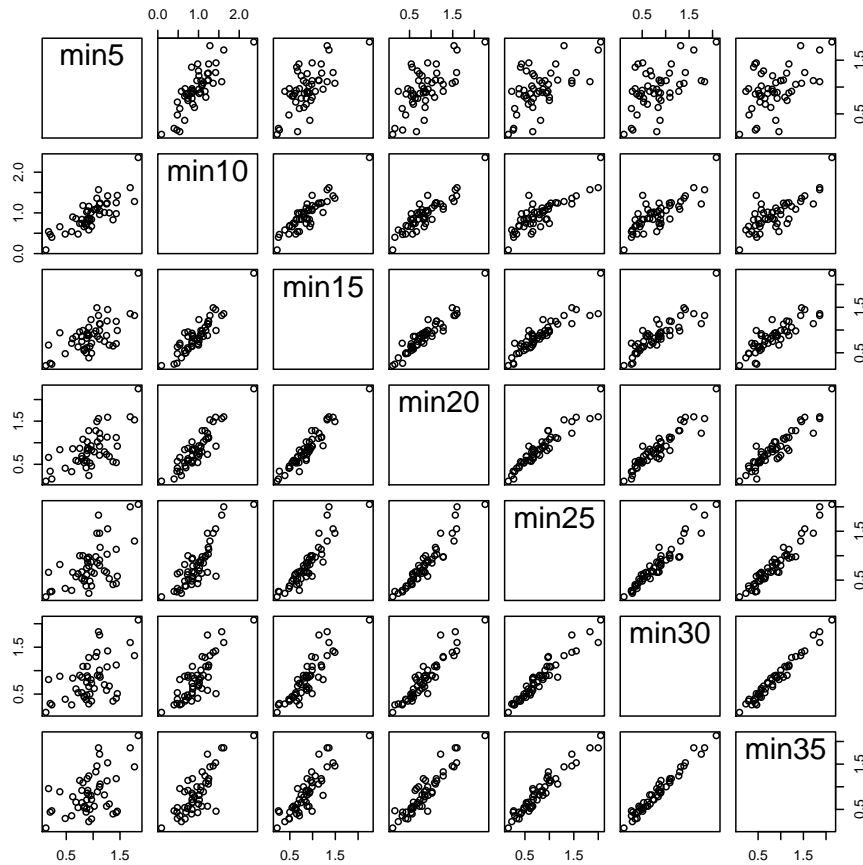
Tabela 1.2.8: Sensibilidade dentinária pré- e pós-operatória

Material	Dentina	Sensibilidade pré-operatória	Sensibilidade pós-operatória		Total
			Ausente	Presente	
Single Bond	Seca	Ausente	22	1	23
		Presente	3	6	9
		Subtotal	25	7	32
Single Bond	Úmida	Ausente	12	10	22
		Presente	7	4	11
		Subtotal	19	14	33
Prime Bond	Seca	Ausente	10	6	16
		Presente	12	3	15
		Subtotal	22	9	31
Prime Bond	Úmida	Ausente	5	13	18
		Presente	11	3	14
		Subtotal	16	16	32

1.3 Análise descritiva e medidas resumo

A principal característica que distingue dados longitudinais de dados transversais é a possível correlação entre as observações realizadas na mesma unidade amostral. Para avaliar essa característica, várias técnicas descritivas podem ser utilizadas. Para dados balanceados em relação ao tempo (especialmente quando não há ou há poucas observações omissas), a mais simples delas corresponde a gráficos de dispersão entre as respostas observadas em dois instantes diferentes. Em geral, esses gráficos são dispostos em forma de matriz (às vezes cognominada *draftman's plot* em inglês) como mostra a Figura 1.3.1 correspondente aos dados do Exemplo 1.2.3 com a exclusão do tratamento controle correspondente à concentração de $0 \mu\text{M}$ em que há dados omissos.

Figura 1.3.1: Gráficos de dispersão para os dados do Exemplo 1.2.3



Nesses gráficos pode-se notar claramente que há correlação entre as medidas

intraunidades amostrais com maior intensidade para as observações realizadas em instantes mais próximos entre si (gráficos situados em torno da diagonal principal da matriz) do que para aquelas realizadas em instantes mais espaçados entre si (gráficos mais afastados da diagonal principal da matriz).

Uma quantificação da intensidade dessas correlações pode ser obtida por intermédio da **matriz de covariâncias amostrais** ou da **matriz de correlações amostrais**. Para os dados do Exemplo 1.2.3, a matriz de covariâncias amostrais para as observações obtidas sob a i -ésima concentração $i = 1, \dots, 5$ pode ser calculada como

$$\hat{\Omega}_i = (n_i - 1)^{-1} \sum_{j=1}^{n_i} (\mathbf{y}_{ij} - \bar{\mathbf{y}}_i)(\mathbf{y}_{ij} - \bar{\mathbf{y}}_i)^\top,$$

em que \mathbf{y}_{ij} é o vetor com as respostas da j -ésima unidade amostral submetida à concentração i , $\bar{\mathbf{y}}_i = \sum_{j=1}^{n_i} \mathbf{y}_{ij}$, com n_i representando o número de unidades amostrais submetidas à concentração i . A correspondente matriz de correlações amostrais é dada por

$$\hat{\Psi}_i = \hat{\mathbf{D}}_i^{-1/2} \hat{\Omega}_i \hat{\mathbf{D}}_i^{-1/2}.$$

em que $\hat{\mathbf{D}}_i = \text{diag}(\hat{\Omega}_i)$. As matrizes de correlações amostrais (sem os elementos redundantes) correspondentes aos dados obtidos sob as concentrações de 0 μM e 32 μM são respectivamente

	min5	min10	min15	min20	min25	min30
min5	1.00	-0.05	0.47	0.55	0.61	-0.04
min10		1.00	0.65	0.38	0.14	0.27
min15			1.00	0.91	0.83	0.66
min20				1.00	0.90	0.79
min25					1.00	0.69
min30						1.00

e

	min5	min10	min15	min20	min25	min30	min35
min5	1.00	0.77	0.41	0.56	0.26	0.22	0.08
min10		1.00	0.82	0.86	0.50	0.46	0.32
min15			1.00	0.86	0.65	0.56	0.50
min20				1.00	0.80	0.68	0.58
min25					1.00	0.95	0.88
min30						1.00	0.94
min35							1.00

Em ambos os casos pode-se confirmar as conclusões tiradas por meio do exame da matriz de gráficos de dispersão apresentada na Figura 1.3.1.

Quando um exame das matrizes de covariâncias amostrais dos diferentes tratamentos (concentrações, no Exemplo 1.2.3) não sugere heterocedasticidade, convém combiná-las para garantir maior precisão na estimação da matriz de covariâncias (populacional) comum. Com essa finalidade, pode-se considerar

$$\hat{\Omega} = (n - p)^{-1} \sum_{i=1}^p (n_i - 1) \hat{\Omega}_i.$$

em que t é o número de tratamentos (concentrações) considerados. Para o exemplo sob investigação, a matriz de covariâncias amostrais combinada (*pooled*) obtida com a eliminação do grupo avaliado sob a concentração 0 μM (em que há observações omissas) é

	min5	min10	min15	min20	min25	min30	min35
min5	1.00	0.75	0.59	0.58	0.47	0.41	0.37
min10		1.00	0.86	0.86	0.73	0.70	0.66
min15			1.00	0.92	0.84	0.80	0.78
min20				1.00	0.91	0.89	0.87
min25					1.00	0.92	0.94
min30						1.00	0.97
min35							1.00

Gráficos de perfis (individuais) talvez sejam as ferramentas descritivas mais importantes para a análise de dados longitudinais. Eles são essencialmente gráficos de dispersão (com o tempo na abscissa e a resposta na ordenada) em que os pontos associados a uma mesma unidade amostral são unidos por segmentos de reta. Em geral, os **perfis médios** são sobrepostos a eles. Esse tipo de gráfico tem sido utilizado por inúmeros autores, como Rao & Rao (1966) ou Weiss & Lazaro (1992) não só para representar dados longitudinais como também para ajudar a identificação de modelos apropriados para inferência estatística. Nas Figuras 1.3.2, 1.3.3 e 1.3.4 apresentamos gráficos de perfis correspondentes aos Exemplos 1.2.3 e 1.2.6, respectivamente, com a sobreposição do perfil de médias (e/ou de curvas alisadas) correspondentes.

Figura 1.3.2: Gráfico de perfis para os dados do Exemplo 1.2.3

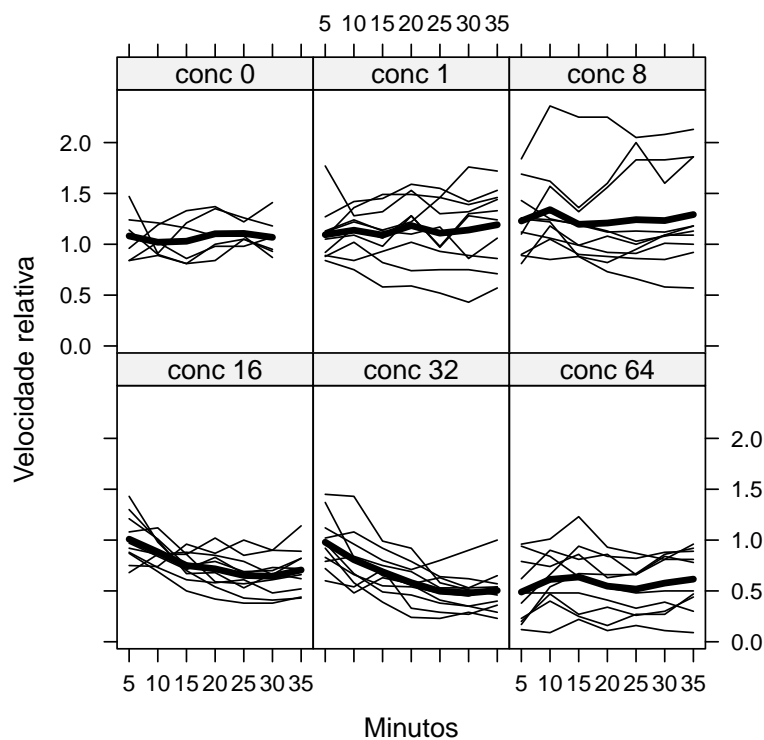
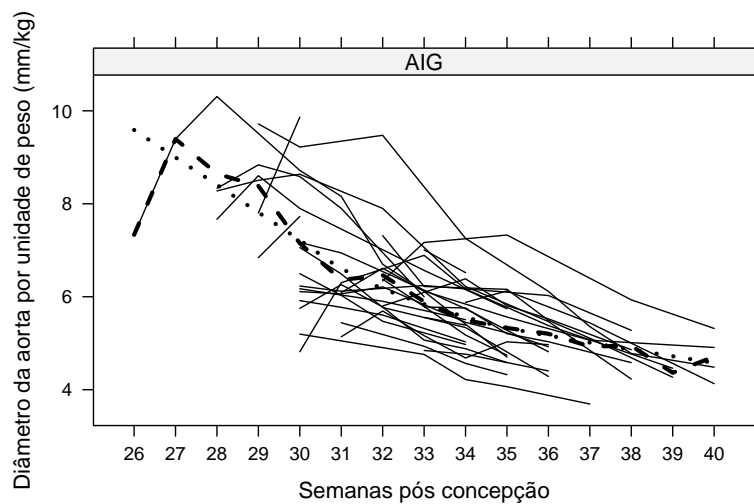
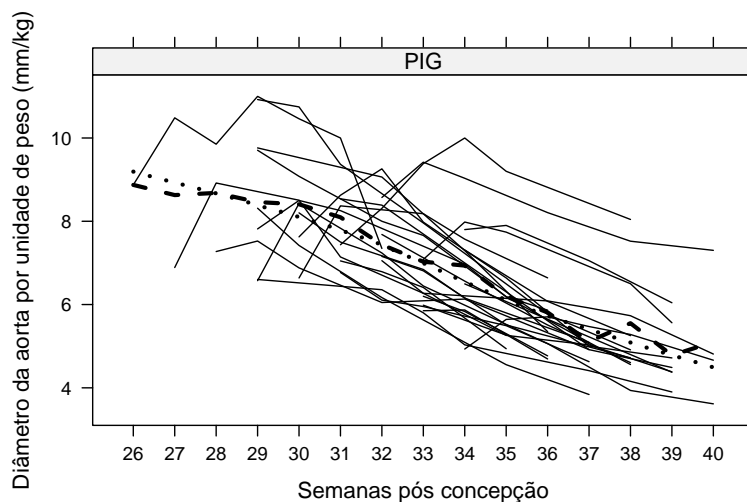
Figura 1.3.3: Gráfico de perfis para os dados do Exemplo 1.2.6 (grupo AIG)
(Linha pontilhada: curva alisada; linha interrompida: perfil médio)

Figura 1.3.4: Gráfico de perfis para os dados do Exemplo 1.2.6 (grupo FIG)
(Linha pontilhada: curva alisada; linha interrompida: perfil médio)



Esse tipo de gráfico permite

- i) identificar possíveis correlações intraunidades amostrais por meio da observação de perfis dispostos acima ou abaixo do perfil médio (trilhamento);
- ii) identificar uma possível heterocedasticidade por intermédio da observação de diferentes variabilidades da resposta ao longo do tempo;
- iii) sugerir possíveis formas para a curva a ser adotada para explicar a variação do perfil médio ao longo do tempo;
- iv) comparar as variabilidades inter- e intraunidades amostrais;
- v) identificar unidades amostrais e/ou observações discrepantes (*outliers*).

Na Figura 1.3.2 pode-se notar que a resposta média (velocidade de transporte mucociliar relativa) correspondente às concentrações 0-8 μM se mantém aproximadamente constante em torno de 1.0, é decrescente para as concentrações de 16 e 32 μM e é aproximadamente constante em torno de 0.5 para a concentração de 64 μM . Os gráficos de perfis individuais sugerem homocedasticidade da resposta para as observações realizadas sob uma mesma concentração, mas indicam heterocedasticidade relativamente às diferentes concentrações. Além disso, pelo menos um perfil individual observado sob a concentração 8 μM se destaca dos demais, sugerindo que pelo menos uma unidade amostral pode ser considerada discrepante.

Embora os dados do Exemplo 1.2.6 correspondam a um estudo com observações coletadas em instantes irregularmente distribuídos no tempo, o gráfico da Figura 1.3.3

sugere homocedasticidade e pode ajudar a escolha de uma estrutura de covariância para os modelos adotados para sua análise. Dois ou três perfis discrepantes também podem ser identificados.

Há situações em que o interesse recai não nos perfis de resposta mas sim em alguma característica deles. Por exemplo, num estudo longitudinal em que se avalia a variação da pressão arterial de um grupo de indivíduos submetidos a um tratamento anti-hipertensivo, o foco pode estar centrado na resposta esperada após seis meses do início da intervenção. Em alguns estudos de bioequivalência, por exemplo, a atenção é dirigida à máxima concentração sérica de alguma droga. Em estudos relacionados com hábitos tabagistas, pode haver interesse na quantidade de nicotina absorvida pelo organismo após um determinado período de tempo. Em todos esses casos, a resposta de interesse pode ser obtida como uma função dos perfis individuais. No primeiro caso, essa resposta é o **desfecho** (*endpoint*), no segundo, é o **pico** (*peak*) e no último, a **área sob a curva** (*area under the curve*). Todos esses exemplos podem ser classificados sob a denominação de **medidas resumo** que essencialmente corresponde à redução de cada perfil (multivariado) a um único valor. Nesse contexto, as possíveis covariâncias intraunidades amostrais são levadas em conta na estimação da variância das medidas resumo. O caso mais simples corresponde a estudos do tipo pré-teste/pós-teste em que a análise é baseada em testes t pareados.

1.3.1 Análise do comportamento de perfis

Uma estrutura experimental bastante comum em estudos longitudinais homocedásticos balanceados em relação ao tempo envolve a observação dos perfis de resposta $\mathbf{y}_{ij} = (y_{ij1}, \dots, y_{ijm})^\top$ de n_i unidades amostrais (indexadas por j) submetidas a um tratamento i , $i = 1, \dots, p$. Esses tratamentos podem corresponder às combinações dos níveis de vários fatores. Nesse contexto, muitas medidas resumo de interesse podem ser expressas na forma

$$f_{ij} = \mathbf{c}^\top \mathbf{y}_{ij} \quad (1.3.1)$$

em que \mathbf{c} é um vetor de constantes conhecidas com dimensão m . Tomando, por exemplo, $\mathbf{c} = m^{-1}\mathbf{1}_m$, f_{ij} corresponderá à média das observações associadas à j -ésima unidade amostral submetida ao i -ésimo tratamento; se $\mathbf{c} = (\mathbf{0}_{m-1}^\top, 1)^\top$, f_{ij} corresponderá à medida realizada no último instante de avaliação (desfecho); quando $\mathbf{c} = (1, \mathbf{0}_{m-2}^\top, -1)^\top$, f_{ij} corresponderá à diferença entre as medidas realizadas no primeiro e último instantes de avaliação. Denotando o vetor do valor esperado de \mathbf{y}_{ij} por $\boldsymbol{\mu}_i$ e a matriz de covariâncias intraunidades amostrais comum por \mathbf{R} , o valor esperado e a variância das medidas resumo são, respectivamente, $\phi_i = \mathbf{c}^\top \boldsymbol{\mu}_i$ e $\sigma^2 = \mathbf{c}^\top \mathbf{R} \mathbf{c}$. Esse procedimento consiste na substituição do perfil de repostas (\mathbf{y}_{ij}) por um escalar (f_{ij}), ou seja, na redução do problema multivariado a um problema

univariado, cuja análise pode ser realizada por intermédio de técnicas de ANOVA ou regressão, por exemplo.

Essa ideia pode ser estendida para casos em que há desbalanceamento em relação ao tempo ou observações omissas, *i.e.*, em que os perfis \mathbf{y}_{ij} têm dimensão m_{ij} , não necessariamente iguais; nesse caso, basta considerar $f_{ij} = \mathbf{c}_{ij}^\top \mathbf{y}_{ij}$ com \mathbf{c}_{ij} denotando um vetor de constantes conhecidas com dimensão m_{ij} e ter em mente que as medidas resumo podem ser heterocedásticas, dado que suas variâncias são $\sigma_{ij}^2 = \mathbf{c}_{ij}^\top \mathbf{R}_{ij} \mathbf{c}_{ij}$.

Casos em que a medida resumo f_{ij} não pode ser expressa na forma (1.3.1) também podem ser considerados. Um exemplo frequente ocorre quando o interesse recai na avaliação da variação da resposta expressa como porcentagem de algum valor basal, por exemplo, $f_{ij} = (y_{ijm} - y_{ij1})/y_{ij1}$ em que y_{ijm} corresponde à resposta obtida na última avaliação e y_{ij1} à resposta obtida no instante inicial (basal). Nesse caso, a variância da medida resumo deve ser obtida de forma aproximada. Com essa finalidade, pode-se empregar o método Delta descrito no Apêndice C.

Exemplo 1.3.1: Os dados do arquivo Singer&Nobre&Rocha2017exemp131.xls, disponível em www.ime.usp.br/~jmsinger/Dados são provenientes de um experimento realizado do Instituto de Ciências Biomédicas da Universidade de São Paulo com o objetivo de avaliar a ação antitumoral de um quimioterápico (BCNU) associado a uma emulsão lipídica (LDE), uma partícula sintética cuja atuação é semelhante ao LDL (*low density lipoprotein*). Após o tempo necessário para o desenvolvimento de um implante tumoral, um conjunto de ratos foi subdividido em 8 grupos (nem todos de mesmo tamanho) e cada um deles submetido a um tratamento (solução salina, LDE sem BCNU, BCNU sem LDE nas doses de 3.5, 7.0 e 15 ppm, LDE com BCNU nas doses de 3.5, 7.0 e 15 ppm). A massa tumoral (g) de cada animal foi medida antes e após 2, 4, 6, 8 e 10 dias da administração do tratamento. Um dos objetivos é comparar os tratamentos com relação à capacidade de diminuição da massa tumoral após 10 dias de tratamento.

1.3.2 Análise de desfecho

A análise de desfecho é importante em ensaios clínicos em que o objetivo é avaliar o valor esperado de algum sintoma ou resposta depois da administração de um ou mais tratamentos por um certo tempo. Embora não haja interesse direto nas avaliações intermediárias, elas geralmente são realizadas para detecção de efeitos colaterais ou outros eventos que possam estar relacionados com o acompanhamento do paciente.

Com dados balanceados em relação ao tempo, esse tipo de análise corresponde a um caso particular daquela baseada em (1.3.1) com $\mathbf{c} = (0, \dots, 0, 1)^\top$. A presença de **desistências** (*dropout*) é que justifica a atenção particularizada. O abandono ocorre quando algum paciente deixa de ser avaliado a partir de um certo instante, seja pelo

efeito do tratamento seja por outra causa, gerando **dados omissos** (*missing data*). Dentre os enfoques mais comuns para a abordagem desse problema, destacamos:

- i) eliminação das unidades com dados omissos;
- ii) utilização da última observação disponível;
- iii) Ponderação da última resposta disponível segundo algum critério que incorpore o motivo da desistência;
- iv) utilização de Análise de Covariância para modelar as causas da desistência;
- v) modelagem do processo gerador das desistências.

A estratégia i) pode gerar conclusões enviesadas, pois não leva em conta o fato de que desistências podem ter ocorrido pela ausência ou deficiência do efeito terapêutico do tratamento ou por intolerância a ele. A estratégia só deve ser utilizada sob a hipótese de que as causas da desistência não são relacionadas com o efeito do tratamento; além disso deve-se avaliar a perda de eficiência da análise gerada pela diminuição do tamanho amostral.

Uma suposição fundamental para o emprego da estratégia ii) é que a última observação disponível do paciente que desistiu do estudo corresponde àquela que se teria observado se ele tivesse completado o ciclo de observações previsto.

O enfoque iii), sugerido por Gould (1973) é aplicável quando as respostas podem ser ordenadas segundo alguma escala que incorpore o motivo pelo qual o paciente desistiu do estudo. Por exemplo, os pacientes que não completaram o estudo podem ser classificados como

- i) curado;
- ii) não tolerou o tratamento;
- iii) foi imune ao tratamento;
- iv) desistiu do estudo por razões não relacionadas com o tratamento.

Gould (1973) sugere que se atribuam escores arbitrários,

$$M_1 > \max(y_{im}) > \min(y_{im}) > M_2 > M_3$$

em que $\min(y_{im})$ e $\max(y_{im})$ correspondem, respectivamente, ao menor e maior valor das respostas observadas no instante m , conforme o paciente seja classificado em i), ii) ou iii). As observações de pacientes classificados em iv) são descartadas. Para os pacientes que terminaram o estudo, os escores correspondem aos valores observados no instante m . A análise pode ser conduzida por meio de técnicas não-paramétricas.

A estratégia iv) sugerida por Bryant & Gillings (1985) consiste em utilizar uma Análise de Covariância que inclua uma variável indicadora para identificar se a

última observação de cada paciente foi realizada no instante m ou antes. Por exemplo, considere um estudo em que se planeja observar pacientes submetidos a p tratamentos em m ocasiões com a finalidade de avaliar o efeito desses tratamentos numa determinada resposta no último instante. Seja y_{ij*} a última resposta do j -ésimo paciente submetido ao i -ésimo tratamento. Um modelo de ANCOVA para acomodar as desistências é

$$y_{ij*} = \mu_i + \beta_i x_{ij} + e_{ij},$$

$i = 1, \dots, p$, $j = 1, \dots, n_i$ em que $x_{ij} = 1$ se o j -ésimo paciente submetido ao i -ésimo tratamento desistiu do estudo antes do instante m ou $x_{ij} = 0$ em caso contrário, e_{ij} representa um erro aleatório com média zero e variância σ^2 independente dos demais, μ_i representa a resposta esperada para pacientes submetidos ao i -ésimo tratamento que não desistiram estudo e β_i representa a variação na resposta esperada do i -ésimo tratamento para pacientes que desistiram do estudo.

A estratégia v) será abordada no capítulo de tópicos especiais.

1.3.3 Análise da área sob a curva

Uma medida resumo de interesse principalmente em estudos farmacológicos é a **área sob a curva** da função que descreve a variação de uma determinada variável resposta ao longo do tempo. Essa medida corresponde a uma resposta cumulativa no intervalo de tempo considerado. Por exemplo, suponhamos que se pretende avaliar o efeito de programa educativo nos hábitos alimentares de voluntários. Com essa finalidade, em diferentes ocasiões durante 48 dias, cada participante dos grupos controle e experimental preencheu um inventário alimentar a partir do qual foi calculada a quantidade diária de gordura *trans* consumida. Suponhamos que esse consumo varie ao longo do período do estudo segundo uma função $f(x)$, desconhecida. De uma forma geral, o consumo efetivo para o participante i acumulado no período compreendido entre dois instantes x_1 e x_2 corresponde a

$$ASC_i = \int_{x_1}^{x_2} f(x) dx. \quad (1.3.2)$$

Os dados correspondentes a um dos participantes podem ser representados graficamente como indicado na Figura 1.3.5. A área representada pela região hachurada corresponde a uma estimativa da quantidade de gordura *trans* consumida nos 48 dias do estudo, ou seja, uma estimativa de (1.3.2) entre os instantes $x = 0$ e $x = 48$. Aqui estamos considerando uma interpolação linear entre dois instantes de observação. Outros tipos de interpolação podem ser adotados. Por exemplo, multiplicando a resposta média (ponderada ou não) pelo tempo de duração do estudo. Para detalhes, ver Bryant & Gillings (1985). Nesse contexto, a estimativa de

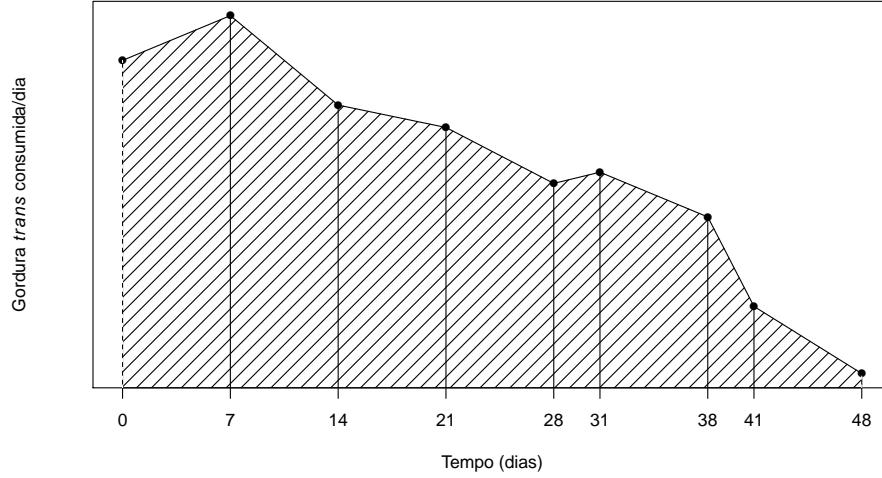


Figura 1.3.5: Exemplo da aproximação da área sob a curva

(1.3.2) é a soma das áreas dos trapézios (ou retângulos) formados pela região delimitada inferiormente pelo eixo das abscissas e superiormente pelo perfil individual de respostas.

Suponha que y_{ij} seja o consumo diário de gordura *trans* observado para o i -ésimo indivíduo no instante x_j , $i = 1, \dots, n$ e $j = 1, \dots, m_i$. A área ASC_i sob a curva correspondente ao i -ésimo indivíduo pode ser estimada por:

$$\widehat{ASC}_i = \sum_{j=1}^{m_i-1} \frac{(y_{ij} + y_{i,j+1})(x_{j+1} - x_j)}{2}. \quad (1.3.3)$$

A comparação entre os grupos experimental e controle poderá ser concretizada por técnicas univariadas paramétricas ou não-paramétricas, dado que reduzimos o perfil (multivariado) de cada participante a um único valor (a medida resumo). No exemplo, a comparação entre o valor esperado da área sob a curva para os indivíduos do grupo controle e aquele correspondente ao grupo experimental pode ser concretizada por um teste t ou pelo equivalente não-paramétrico.

Um procedimento alternativo para comparar tratamentos com relação a perfis esperados foi sugerido por Zerbe & Walker (1977) e envolve a distância entre dois perfis esperados $f(x)$ e $g(x)$, definida como

$$d(f, g) = \left\{ \int_{x_1}^{x_2} [f(x) - g(x)]^2 dx \right\}^{1/2}.$$

Esses autores sugerem estimadores das curvas de respostas individuais no caso em que podem ser representadas por polinômios e propõem um teste não-paramétrico

com o mesmo espírito daquele empregado em ANOVA, ou seja, comparando as distâncias entre as curvas individuais e a curva média geral com aquelas entre as curvas individuais e as curvas médias associadas a cada tratamento.

O conceito de área sob a curva pode ser generalizado para outros casos nos quais também se deseja uma medida resumo cumulativa como aquele descrito no Exemplo 1.2.4. Dado que o risco de desenvolvimento de cárie é maior quando o pH da boca está abaixo de 5.5, uma medida resumo de interesse deve avaliar o efeito cumulativo do tempo de exposição a pH menor que 5.5. Essa medida, que denominamos **potencial cariogênico**, corresponde à área compreendida entre a reta paralela ao eixo das abscissas passando pelo ponto pH=5.5 e a porção da curva resposta individual situada abaixo dessa reta. Na Figura 1.3.6 observam-se as medidas de pH para um dos voluntários da amostra, antes do bochecho. O correspondente poten-

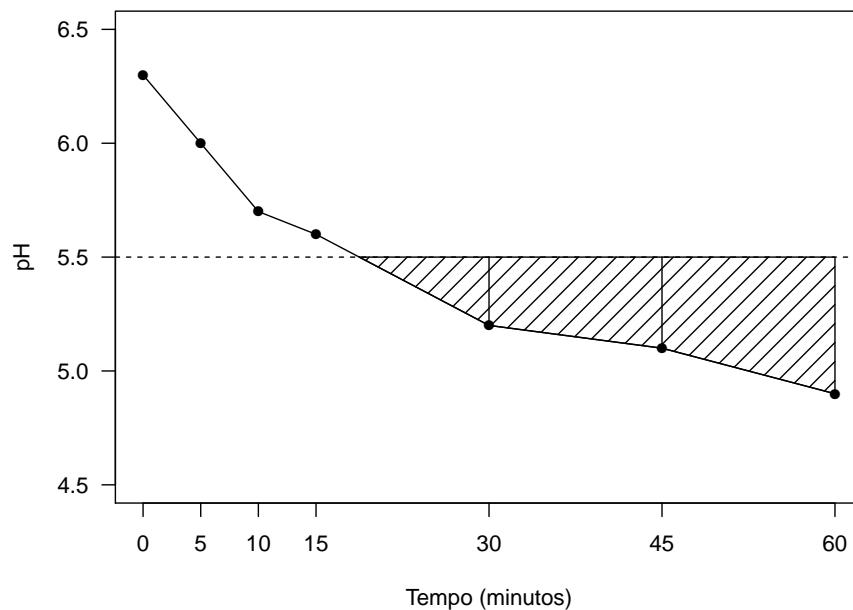


Figura 1.3.6: Área correspondente ao potencial cariogênico pré-bochecho para um dos voluntários (Tabela 1.2.4)

cial cariogênico pode ser estimado pela soma das áreas dos polígonos que compõem a área hachurada. Na Figura 1.3.6, o potencial cariogênico estimado corresponde à soma das áreas do triângulo e dos dois trapézios hachurados. Para o cálculo da área do triângulo, é necessário determinar o ponto de intersecção do segmento de reta do perfil de respostas que passa pelos pontos (15, 5.6) e (30, 5.2) com a reta correspondente a $y = 5.5$. A equação da reta que contém o segmento mencionado é

$$y - 5.6 = \alpha(x - 15)$$

em que $\alpha = (5.2 - 5.6)/(30 - 15) \approx -0,03$. Para $y = 5.5$, obtém-se $x = 18.75$ e pode-se então calcular a área do triângulo em questão. O cálculo das áreas dos dois trapézios pode ser feito conforme (1.3.3) e a soma das áreas dos três polígonos corresponde a uma estimativa do potencial cariogênico pré-bochecho para esse voluntário. O cálculo dos potenciais cariogênicos pré- e pós-bochecho para cada um dos participantes permite que o problema (de comparação do efeito esperado do bochecho sobre o potencial cariogênico) possa ser resolvido por meio de um teste t pareado ou de um teste de Wilcoxon.

Os valores observados dos potenciais cariogênicos pré- e pós-bochecho estão dispostos na Tabela 1.3.1 e algumas estatísticas descritivas, incluindo aquelas para as diferenças do potencial cariogênico (pré-bochecho - pós-bochecho) são apresentadas na Tabela 1.3.2.

Tabela 1.3.1: Potencial cariogênico (Exemplo 1.2.4)

Voluntário	Bochecho		Diferença (antes – depois)
	Antes	Depois	
1	6.00	3.37	2.63
2	0.00	1.50	-1.50
3	0.00	10.69	-10.69
4	3.37	8.44	-5.07
5	0.00	2.50	-2.50
6	14.25	5.25	9.00
7	0.00	5.50	-5.50
8	0.00	0.37	-0.37
9	0.00	1.50	-1.50
10	0.00	16.50	-16.50
11	0.37	0.00	0.37
12	14.25	15.94	-1.69
13	0.00	5.25	-5.25
14	14.44	0.00	14.44
15	0.00	12.25	-12.25
16	0.00	9.25	-9.25
17	0.00	9.00	-9.00
18	0.00	8.50	-8.50
19	0.00	9.00	-9.00
20	12.37	22.50	-10.13
21	0.00	0.75	-0.75

Embora a suposição de normalidade não seja plausível para as observações do

potencial cariogênico estimado pré-bochecho (a distribuição amostral é claramente assimétrica), ela é admissível para a diferença. Um teste t -pareado aplicado aos dados indica um aumento significativo ($p < 0.01$) do potencial cariogênico esperado após o bochecho (7.1) relativamente àquele correspondente ao pré-bochecho (3.10); o intervalo de confiança com coeficiente de confiança 0.95 para a diferença é $(-7.2; -0.7)$. Esse resultado é corroborado por um teste de postos sinalizados de Wilcoxon ($p = 0.006$).

Tabela 1.3.2: Estatísticas descritivas para o potencial cariogênico (Exemplo 1.2.4)

Bochecho	n	Média	DP	Mediana	Máximo	Mínimo
Antes	21	3.10	5.54	0.00	14.44	0.00
Depois	21	7.05	6.13	5.50	22.50	0.00
Diferença	21	-3.95	7.16	-5.07	14.44	-16.50

Embora a análise de dados com natureza longitudinal por intermédio de medidas resumo seja útil em muitas situações práticas, em geral há interesse em mais do que um aspecto do comportamento da resposta ao longo do período em que são realizadas as observações. A comparação de curvas de crescimento é um exemplo. Nesses casos, a análise requer modelos mais sofisticados e esse é o tópico abordado nos demais capítulos deste texto.

1.4 Exercícios

1.4.1. Construa os vetores de observações (\mathbf{y}_i) e matrizes com valores das covariáveis (\mathbf{X}_i) correspondentes aos dados dos Exemplos 1.2.1, 1.2.2, 1.2.3, 1.2.4, 1.2.5 e 1.2.6.

1.4.2. Considere os dados do Exemplo 1.2.4.

- Obtenha as matrizes de covariâncias e correlações intraunidades amostrais.
- Construa gráficos de perfis individuais para os dados do Exemplo 1.2.4. com perfis médios superpostos.
- Avalie o comportamento das variâncias e correlações correspondentes.

1.4.3. Analise os dados do Exemplo 1.3.1 levando em conta o objetivo indicado,

- com eliminação das unidades amostrais em que a medida no dia 10 está omissa;
- com a utilização da última observação disponível para cada unidade amostral.

1.4.4. Os dados do arquivo Singer&Nobre&Rocha2017exerc143.xls, disponível em www.ime.usp.br/~jmsinger/Dados são provenientes de um experimento realizado na Faculdade de Medicina da USP com o objetivo de comparar pacientes com hepatite C tratados com dois inibidores de protease (Telaprevir e Boceprevir) relativamente à ocorrência de disfunção renal. Uma das respostas de interesse é a gravidade da disfunção renal medida em termos da magnitude da diferença entre o nível sérico de creatinina e o valor crítico correspondente a 15% acima do nível basal e do tempo em que essa diferença persiste. Outras respostas têm definições similares mas são baseadas em valores 15% abaixo do nível basal da taxa de filtração glomerular calculada quer pelo método de Cockcroft-Gault quer pelo método MDRD.

- a) Construa gráficos de perfis individuais para os dados com perfis médios superpostos.
- b) Obtenha as respostas individuais definidas como a área entre os perfis individuais e a reta paralela ao eixo das abscissas correspondente a 15% acima do nível basal da creatinina.
- c) Obtenha as respostas individuais definidas como a área entre os perfis individuais e a reta paralela ao eixo das abscissas correspondente a 15% abaixo do nível basal da taxa de filtração glomerular calculada pelo método de Cockcroft-Gault.
- d) Obtenha as respostas individuais definidas como a área entre os perfis individuais e a reta paralela ao eixo das abscissas correspondente a 15% abaixo do nível basal da taxa de filtração glomerular calculada pelo método MDRD.
- c) Utilize modelos lineares para comparar as respostas médias dos dois grupos (Telaprevir e Boceprevir) controlando sexo, idade, hipertensão arterial (HAS), diabetes (DM), anemia na segunda semana de tratamento e cirrose.

Capítulo 2

Modelos lineares para dados gaussianos

A facilidade de tratamento matemático em conjunto com a indisponibilidade de meios computacionais adequados constituem as principais razões para que a análise de dados com medidas repetidas, ou, mais especificamente, de dados longitudinais tenha-se focado na distribuição normal desde as primeiras incursões realizadas por Wishart (1938). Os trabalhos clássicos de Box (1950), Geisser & Greenhouse (1958), Potthoff & Roy (1964), Rao (1959), Rao (1965), Rao (1966), Rao (1967) e Grizzle & Allen (1969), entre outros, são exemplos típicos dos primeiros esforços que se transformaram numa profícua área de pesquisa. Como consequência, uma grande variedade de modelos para dados gaussianos pode ser encontrada na literatura estatística. Como as propriedades de estimadores e estatísticas de teste obtidas sob esse modelo probabilístico são bem conhecidas, os métodos dele oriundos são excelentes ferramentas para aplicações práticas.

2.1 Introdução

Em geral, podemos analisar dados provenientes de estudos com medidas repetidas, por meio de modelos mistos da forma

$$\mathbf{y}_i = \mathbf{g}(\mathbf{X}_i, \mathbf{Z}_i, \boldsymbol{\beta}, \mathbf{b}_i) + \mathbf{e}_i \quad i = 1, \dots, n \quad (2.1.1)$$

em que, $\mathbf{y}_i = (y_{i1}, \dots, y_{im_i})^\top$ com dimensão $(m_i \times 1)$ é o perfil de respostas da i -ésima unidade experimental, $\boldsymbol{\beta}$ é um vetor com dimensão $(p \times 1)$ de parâmetros (efeitos fixos ou parâmetros de localização) desconhecidos, $\mathbf{X}_i = (\mathbf{x}_{i1}, \dots, \mathbf{x}_{ip})$ é uma matriz de especificação dos efeitos fixos com dimensão $(m_i \times p)$, conhecida e de posto completo, em que $\mathbf{x}_{ij} = (x_{ij1}, \dots, x_{ijm_i})^\top$ representa o vetor com os

m_i valores da j -ésima variável independente ($j = 1, \dots, p$) para a i -ésima unidade amostral, \mathbf{b}_i é um vetor com dimensão $(q \times 1)$ de variáveis latentes, comumente denominadas efeitos aleatórios, que refletem o comportamento individual da i -ésima unidade experimental, \mathbf{Z}_i é uma matriz de especificação dos efeitos aleatórios (com dimensão $(m_i \times q)$, conhecida e de posto completo), \mathbf{g} é uma função com derivadas contínuas até à segunda ordem e \mathbf{e}_i é um vetor de erros aleatórios com dimensão $(m_i \times 1)$.

Em muitos casos é razoável supor que $\mathbf{b}_i \sim N_q(\mathbf{0}, \mathbf{G})$ e $\mathbf{e}_i \sim N_{m_i}(\mathbf{0}, \mathbf{R}_i)$, em que \mathbf{G} , com dimensão $(q \times q)$ e \mathbf{R}_i , com dimensão $(m_i \times m_i)$ são matrizes simétricas definidas positivas e, além disso, que \mathbf{b}_i e \mathbf{e}_i são variáveis aleatórias independentes. Quando \mathbf{g} é uma função linear dos efeitos fixos $\boldsymbol{\beta}$ e dos efeitos aleatórios \mathbf{b}_i , (2.1.1) se reduz ao **modelo linear misto**

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \mathbf{e}_i, \quad i = 1, \dots, n. \quad (2.1.2)$$

Sob esse modelo, o vetor de respostas associado à i -ésima unidade amostral tem distribuição normal multivariada com vetor de médias e matriz de covariâncias dados, respectivamente, por

$$\mathbb{E}(\mathbf{y}_i) = \mathbf{X}_i\boldsymbol{\beta} \quad (2.1.3)$$

e

$$\mathbb{V}(\mathbf{y}_i) = \boldsymbol{\Omega}_i = \mathbf{Z}_i\mathbf{G}\mathbf{Z}_i^\top + \mathbf{R}_i. \quad (2.1.4)$$

A primeira componente da decomposição de $\mathbb{V}(\mathbf{y}_i)$ em (2.1.4) modela a dispersão dos perfis individuais de resposta e a segunda componente está relacionada com a dispersão da resposta em torno dos perfis individuais. As matrizes \mathbf{R}_i e \mathbf{G} são funções conhecidas de t_1 e t_2 parâmetros desconhecidos, respectivamente, ou seja, $\mathbf{G} = \mathbf{G}(\boldsymbol{\theta})$ e $\mathbf{R}_i = \mathbf{R}_i(\boldsymbol{\theta})$, de modo que a matriz de covariâncias para a i -ésima unidade amostral, $\boldsymbol{\Omega}_i = \boldsymbol{\Omega}_i(\boldsymbol{\theta})$, dependerá do vetor $\boldsymbol{\theta}$ com dimensão $(t \times 1)$, $t = t_1 + t_2$.

Se eliminarmos os componentes aleatórios \mathbf{b}_i [e consequentemente as matrizes \mathbf{Z}_i de (2.1.2)], teremos os chamados **modelos marginais**, nos quais a estrutura de covariância é modelada diretamente por meio das matrizes \mathbf{R}_i .

O modelo (2.1.2) também pode ser interpretado como um **modelo linear em dois estágios** como mostram Laird & Ware (1982). No primeiro estágio, consideramos fixos os efeitos aleatórios \mathbf{b}_i , de forma que

$$\mathbf{y}_i | \mathbf{b}_i \sim N(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i, \mathbf{R}_i). \quad (2.1.5)$$

No segundo estágio supomos que os vetores \mathbf{b}_i são independentes com distribuição $N_q(\mathbf{0}, \mathbf{G})$ e consequentemente, o modelo para a distribuição marginal de \mathbf{y}_i é

$$\mathbf{y}_i \sim N(\mathbf{X}_i\boldsymbol{\beta}, \mathbf{Z}_i\mathbf{G}\mathbf{Z}_i^\top + \mathbf{R}_i). \quad (2.1.6)$$

Quando $\mathbf{R}_i = \sigma^2 \mathbf{I}_{m_i}$, o modelo é chamado de **modelo de independência condicional homocedástico**, indicando que as m_i observações associadas à i -ésima unidade amostral são condicionalmente independentes dado \mathbf{b}_i . Uma análise ingênua de um conjunto de dados (Exemplo B.5.2) sob esse enfoque está apresentada no Apêndice B.

O modelo (2.1.2) pode ser escrito compactamente como

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \mathbf{e}, \quad (2.1.7)$$

em que $\mathbf{y} = [\mathbf{y}_1^\top, \dots, \mathbf{y}_n^\top]^\top$ é um vetor com dimensão $(N \times 1)$ em que $N = \sum_{i=1}^n m_i$ contendo as respostas das n unidades amostrais, $\mathbf{X} = [\mathbf{X}_1^\top, \dots, \mathbf{X}_n^\top]^\top$ com dimensão $(N \times p)$ é a matriz de especificação dos efeitos fixos, $\mathbf{Z} = \oplus_{i=1}^n \mathbf{Z}_i$ com dimensão $(N \times nq)$ é a matriz de especificação dos efeitos aleatórios, $\mathbf{b} = [\mathbf{b}_1^\top, \dots, \mathbf{b}_n^\top]^\top$ é um vetor com dimensão $(nq \times 1)$ que engloba os efeitos aleatórios e $\mathbf{e} = [\mathbf{e}_1^\top, \dots, \mathbf{e}_n^\top]^\top$ é um vetor com dimensão $(N \times 1)$ de erros aleatórios. Sob esta formulação, $\mathbf{b} \sim N_{nq}[\mathbf{0}, \boldsymbol{\Gamma}(\boldsymbol{\theta})]$ em que $\boldsymbol{\Gamma}(\boldsymbol{\theta}) = \mathbf{I}_n \otimes \mathbf{G}(\boldsymbol{\theta})$ e \mathbf{b} é independente de $\mathbf{e} \sim N_N[\mathbf{0}, \mathbf{R}(\boldsymbol{\theta})]$ com $\mathbf{R}[(\boldsymbol{\theta})] = \oplus_{i=1}^n \mathbf{R}_i(\boldsymbol{\theta})$. Consequentemente, $\mathbf{y} \sim N_N[\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Omega}(\boldsymbol{\theta})]$ em que

$$\boldsymbol{\Omega}(\boldsymbol{\theta}) = \mathbf{Z}\boldsymbol{\Gamma}(\boldsymbol{\theta})\mathbf{Z}^\top + \mathbf{R}(\boldsymbol{\theta}).$$

Para facilitar o desenvolvimento metodológico e aplicações práticas, apresentamos um resumo das dimensões dos componentes do modelo linear misto na Tabela 2.1.1.

Tabela 2.1.1: Resumo das dimensões dos componentes do modelo linear misto

Formulação individual		Formulação compacta	
Componente	Dimensão	Componente	Dimensão
\mathbf{y}_i	$m_i \times 1$	\mathbf{y}	$N \times 1$
\mathbf{X}_i	$m_i \times p$	\mathbf{X}	$N \times p$
$\boldsymbol{\beta}$	$p \times 1$	$\boldsymbol{\beta}$	$p \times 1$
\mathbf{Z}_i	$m_i \times q$	\mathbf{Z}	$N \times nq$
\mathbf{b}_i	$q \times 1$	\mathbf{b}	$nq \times 1$
\mathbf{e}_i	$m_i \times 1$	\mathbf{e}	$N \times 1$
\mathbf{G}	$q \times q$	$\boldsymbol{\Gamma}$	$nq \times nq$
\mathbf{R}_i	$m_i \times m_i$	\mathbf{R}	$N \times N$
$\boldsymbol{\Omega}_i$	$m_i \times m_i$	$\boldsymbol{\Omega}$	$N \times N$
$\boldsymbol{\theta}$	$t \times t$	$\boldsymbol{\theta}$	$t \times t$

Mais detalhes sobre modelos mistos para estudos com medidas repetidas ou longitudinais podem ser encontrados em Crowder & Hand (1990) Jones (1993), Crowder

& Hand (1996), Verbeke & Molenberghs (1997), Vonesh & Chinchilli (1997), Singer & Andrade (2000), Diggle et al. (2002), Davis (2002) ou Demidenko (2013), entre outros.

2.2 Modelos para a estrutura de covariância

Grande parte do esforço empregado na modelagem de dados com medidas repetidas se concentra na estrutura de covariância. Em geral, o modelo para a matriz de covariâncias $\mathbf{\Omega}_i$ deve depender da maneira pela qual as observações foram obtidas e do conhecimento sobre o mecanismo gerador das observações. Diggle (1988) e Diggle et al. (2002) comentam que a matriz de covariâncias deve ser suficientemente flexível para incluir no mínimo três fontes diferentes de variação aleatória, nomeadamente: *i)* a variação devida a efeitos aleatórios, quando as unidades de investigação formam uma amostra aleatória da população de interesse; *ii)* a variação que pode ser explicada por correlação serial, em que se esperam observações próximas mais fortemente correlacionadas que observações mais distantes e *iii)* a variação devida a erros de medida.

No contexto dos modelos mistos, a covariância entre as observações obtidas em uma mesma unidade amostral poderá ser modelada diretamente por meio da matriz \mathbf{R}_i , que representa a estrutura de covariância das observações intraunidades amostrais ou indiretamente, por meio de uma combinação da matriz \mathbf{G} associada aos efeitos aleatórios \mathbf{b}_i , que representam a variabilidade entre as unidades amostrais com a matriz \mathbf{R}_i . Diggle et al. (2002, Section 5.2) apresentam sugestões sobre as diferentes maneiras em que modelos para matrizes \mathbf{G} e \mathbf{R}_i podem incluir três possíveis fontes de variação na estrutura de covariância: aquelas correspondente aos efeitos aleatórios, à correlação serial e aos erros de medida.

No modelo linear misto homocedástico com independência condicional, $\mathbf{R}_i = \sigma^2 \mathbf{I}_{m_i}$ e \mathbf{G} é uma matriz de componentes de variância como aquelas sugeridas em Searle (1971) e Searle, Casela & McCulloch (1992). Modelos desse tipo são utilizados, por exemplo, em estudos com planejamentos em blocos aleatórios ou do tipo *split-plot*.

Das possíveis estruturas de covariância disponíveis na literatura estatística, algumas estão descritas abaixo, por intermédio de exemplos com $m_i = 4$. Essas estruturas podem ser adotadas tanto para a matriz $\mathbf{\Omega}_i$ como para suas componentes \mathbf{R}_i e \mathbf{G} .

1. Estrutura Uniforme [$\boldsymbol{\theta} = (\sigma^2, \tau)^\top$]

$$\mathbf{V}_i(\boldsymbol{\theta}) = \begin{bmatrix} \sigma^2 + \tau & \tau & \tau & \tau \\ \tau & \sigma^2 + \tau & \tau & \tau \\ \tau & \tau & \sigma^2 + \tau & \tau \\ \tau & \tau & \tau & \sigma^2 + \tau \end{bmatrix} \quad (2.2.1)$$

2. Estrutura AR(1) [$\boldsymbol{\theta} = (\sigma^2, \phi)^\top$]

$$\mathbf{V}_i(\boldsymbol{\theta}) = \sigma^2 \begin{bmatrix} 1 & \phi & \phi^2 & \phi^3 \\ \phi & 1 & \phi & \phi^2 \\ \phi^2 & \phi & 1 & \phi \\ \phi^3 & \phi^2 & \phi & 1 \end{bmatrix} \quad (2.2.2)$$

3. Estrutura ARMA(1,1) [$\boldsymbol{\theta} = (\sigma^2, \gamma, \phi)^\top$]

$$\mathbf{V}_i(\boldsymbol{\theta}) = \sigma^2 \begin{bmatrix} 1 & \gamma & \gamma\phi & \gamma\phi^2 \\ \gamma & 1 & \gamma & \gamma\phi \\ \gamma\phi & \gamma & 1 & \gamma \\ \gamma\phi^2 & \gamma\phi & \gamma & 1 \end{bmatrix} \quad (2.2.3)$$

4. Estrutura antedependência de ordem 1 [$\boldsymbol{\theta} = (\sigma_1^2, \sigma_2^2, \sigma_3^2, \sigma_4^2, \rho_1, \rho_2, \rho_3)^\top$]

$$\mathbf{V}_i(\boldsymbol{\theta}) = \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho_1 & \sigma_1\sigma_3\rho_1\rho_2 & \sigma_1\sigma_4\rho_1\rho_2\rho_3 \\ \sigma_1\sigma_2\rho_1 & \sigma_2^2 & \sigma_2\sigma_3\rho_2 & \sigma_2\sigma_4\rho_2\rho_3 \\ \sigma_1\sigma_3\rho_1\rho_2 & \sigma_2\sigma_3\rho_2 & \sigma_3^2 & \sigma_3\sigma_4\rho_3 \\ \sigma_1\sigma_4\rho_1\rho_2\rho_3 & \sigma_2\sigma_4\rho_2\rho_3 & \sigma_3\sigma_4\rho_3 & \sigma_4^2 \end{bmatrix} \quad (2.2.4)$$

5. Estrutura Toeplitz [$\boldsymbol{\theta} = (\sigma^2, \sigma_1, \sigma_2, \sigma_3)^\top$]

$$\mathbf{V}_i(\boldsymbol{\theta}) = \begin{bmatrix} \sigma^2 & \sigma_1 & \sigma_2 & \sigma_3 \\ \sigma_1 & \sigma^2 & \sigma_1 & \sigma_2 \\ \sigma_2 & \sigma_1 & \sigma^2 & \sigma_1 \\ \sigma_3 & \sigma_2 & \sigma_1 & \sigma^2 \end{bmatrix} \quad (2.2.5)$$

6. Estrutura espacial ou de Markov [$\boldsymbol{\theta} = (\sigma^2, \rho)^\top$]

$$\mathbf{V}_i(\boldsymbol{\theta}) = \sigma^2 \begin{bmatrix} 1 & \rho^{d_{12}} & \rho^{d_{13}} & \rho^{d_{14}} \\ \rho^{d_{21}} & 1 & \rho^{d_{23}} & \rho^{d_{24}} \\ \rho^{d_{31}} & \rho^{d_{32}} & 1 & \rho^{d_{34}} \\ \rho^{d_{41}} & \rho^{d_{42}} & \rho^{d_{43}} & 1 \end{bmatrix} \quad (2.2.6)$$

7. **Estrutura uniforme heterogênea** $[\boldsymbol{\theta} = (\sigma_1^2, \sigma_2^2, \sigma_3^2, \sigma_4^2, \rho)^\top]$

$$\mathbf{V}_i(\boldsymbol{\theta}) = \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho & \sigma_1\sigma_3\rho & \sigma_1\sigma_4\rho \\ \sigma_2\sigma_1\rho & \sigma_2^2 & \sigma_2\sigma_3\rho & \sigma_2\sigma_4\rho \\ \sigma_3\sigma_1\rho & \sigma_3\sigma_2\rho & \sigma_3^2 & \sigma_3\sigma_4\rho \\ \sigma_4\sigma_1\rho & \sigma_4\sigma_2\rho & \sigma_4\sigma_3\rho & \sigma_4^2 \end{bmatrix} \quad (2.2.7)$$

8. **Estrutura ARH(1)** $[\boldsymbol{\theta} = (\sigma_1^2, \sigma_2^2, \sigma_3^2, \sigma_4^2, \phi)^\top]$

$$\mathbf{V}_i(\boldsymbol{\theta}) = \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\phi & \sigma_1\sigma_3\phi^2 & \sigma_1\sigma_4\phi^3 \\ \sigma_2\sigma_1\phi & \sigma_2^2 & \sigma_2\sigma_3\phi & \sigma_2\sigma_4\phi^2 \\ \sigma_3\sigma_1\phi^2 & \sigma_3\sigma_2\phi & \sigma_3^2 & \sigma_3\sigma_4\phi \\ \sigma_4\sigma_1\phi^3 & \sigma_4\sigma_2\phi^2 & \sigma_4\sigma_3\phi & \sigma_4^2 \end{bmatrix} \quad (2.2.8)$$

9. **Estrutura Toeplitz heterogênea** $[\boldsymbol{\theta} = (\sigma_1^2, \sigma_2^2, \sigma_3^2, \sigma_4^2, \rho_1, \rho_2, \rho_3)^\top]$

$$\mathbf{V}_i(\boldsymbol{\theta}) = \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho_1 & \sigma_1\sigma_3\rho_2 & \sigma_1\sigma_4\rho_3 \\ \sigma_2\sigma_1\rho_1 & \sigma_2^2 & \sigma_2\sigma_3\rho_1 & \sigma_2\sigma_4\rho_2 \\ \sigma_3\sigma_1\rho_2 & \sigma_3\sigma_2\rho_1 & \sigma_3^2 & \sigma_3\sigma_4\rho_1 \\ \sigma_4\sigma_1\rho_3 & \sigma_4\sigma_2\rho_2 & \sigma_4\sigma_3\rho_1 & \sigma_4^2 \end{bmatrix} \quad (2.2.9)$$

10. **Estrutura baseada em efeitos aleatórios**

Para um modelo linear com coeficientes lineares e angulares aleatórios, (2.1.4) se reduz a

$$\boldsymbol{\Omega}_i = \mathbf{Z}_i \begin{bmatrix} \sigma_0^2 & \sigma_{01} \\ \sigma_{01} & \sigma_1^2 \end{bmatrix} \mathbf{Z}_i^\top + \mathbf{R}_i \quad (2.2.10)$$

com

$$\mathbf{Z}_i^\top = \begin{bmatrix} 1 & 1 & 1 & 1 \\ x_1 & x_2 & x_3 & x_4 \end{bmatrix}.$$

No caso em que $\mathbf{R}_i = \sigma^2 \mathbf{I}_4$, temos $\boldsymbol{\theta} = (\sigma_0^2, \sigma_1^2, \sigma_{01}, \sigma^2)^\top]$

11. **Não-estruturada (NE)** $[\boldsymbol{\theta} = (\sigma_1^2, \sigma_2^2, \sigma_3^2, \sigma_4^2, \sigma_{12}, \sigma_{13}, \sigma_{14}, \sigma_{23}, \sigma_{24}, \sigma_{34})^\top]$

$$\mathbf{V}_i(\boldsymbol{\theta}) = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ \sigma_{12} & \sigma_2^2 & \sigma_{23} & \sigma_{24} \\ \sigma_{13} & \sigma_{23} & \sigma_3^2 & \sigma_{34} \\ \sigma_{14} & \sigma_{24} & \sigma_{34} & \sigma_4^2 \end{bmatrix} \quad (2.2.11)$$

Muitas dessas estruturas podem ser expressas na forma linear

$$\mathbf{V}_i(\boldsymbol{\theta}) = \sum_{h=1}^d \phi_h \mathbf{F}_{hi} \quad (2.2.12)$$

em que \mathbf{F}_{hi} são matrizes convenientes conhecidas e ϕ_h são parâmetros desconhecidos. Esse modelo é adequado quando os parâmetros da matriz de covariâncias são aditivos, como é o caso das estruturas Uniforme e Toeplitz. Se os parâmetros forem multiplicativos como ocorre, por exemplo, na estrutura AR(1), que possui apenas dois parâmetros, precisaremos de quatro parâmetros para escrevê-la na forma linear. O modelo Toeplitz poderá ser escrito nessa forma com $\phi_1 = \sigma^2, \phi_2 = \sigma_1, \phi_3 = \sigma_2, \phi_4 = \sigma_3$ e

$$\begin{aligned} \mathbf{F}_{1i} &= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, & \mathbf{F}_{2i} &= \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \\ \mathbf{F}_{3i} &= \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}, & \mathbf{F}_{4i} &= \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}. \end{aligned} \quad (2.2.13)$$

Galecki (1994) propõe uma classe de estruturas de covariância baseada no produto de Kronecker, para estudos com medidas repetidas com mais de um fator intraunidades amostrais, como no Exemplo 1.2.5. Dentre as estruturas propostas, citamos

1. Não-estruturada \otimes autorregressiva

$$\mathbf{R}(\boldsymbol{\theta}) = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_2^2 & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_3^2 \end{bmatrix} \otimes \begin{bmatrix} 1 & \phi & \phi^2 & \phi^3 & \phi^4 & \phi^5 & \phi^6 \\ \phi & 1 & \phi & \phi^2 & \phi^3 & \phi^4 & \phi^5 \\ \phi^2 & \phi & 1 & \phi & \phi^2 & \phi^3 & \phi^4 \\ \phi^3 & \phi^2 & \phi & 1 & \phi & \phi^2 & \phi^3 \\ \phi^4 & \phi^3 & \phi^2 & \phi & 1 & \phi & \phi^2 \\ \phi^5 & \phi^4 & \phi^3 & \phi^2 & \phi & 1 & \phi \\ \phi^6 & \phi^5 & \phi^4 & \phi^3 & \phi^2 & \phi & 1 \end{bmatrix} \quad (2.2.14)$$

2. Não-estruturada \otimes uniforme

$$\mathbf{R}(\boldsymbol{\theta}) = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_2^2 & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_3^2 \end{bmatrix} \otimes \begin{bmatrix} 1 & \tau & \tau & \tau & \tau & \tau & \tau \\ \tau & 1 & \tau & \tau & \tau & \tau & \tau \\ \tau & \tau & 1 & \tau & \tau & \tau & \tau \\ \tau & \tau & \tau & 1 & \tau & \tau & \tau \\ \tau & \tau & \tau & \tau & 1 & \tau & \tau \\ \tau & \tau & \tau & \tau & \tau & 1 & \tau \\ \tau & \tau & \tau & \tau & \tau & \tau & 1 \end{bmatrix}. \quad (2.2.15)$$

Mais detalhes sobre a escolha de estruturas para as matrizes de covariâncias intraunidades amostrais podem ser encontrados em Rocha (2004), por exemplo.

2.3 Inferência por máxima verossimilhança

Vários métodos de estimação dos parâmetros do modelo (2.1.7) estão disponíveis na literatura estatística; dentre eles convém destacar o método bayesiano detalhado em Tountenburg (1982), Maritz & Lwin (1989) e Searle et al. (1992), os métodos de Máxima Verossimilhança (MV) e Máxima Verossimilhança Restrita (MVR), discutidos em Patterson & Thompson (1971), Harville (1977b), Robinson (1991), Searle et al. (1992) e Jiang (1996) e o método de Mínimos Quadrados (MQ), analisado em Searle et al. (1992), Draper & Smith (2002), por exemplo.

Supondo que $\mathbf{\Gamma}(\boldsymbol{\theta})$ e $\mathbf{R}(\boldsymbol{\theta})$ são conhecidas, todos esses métodos são equivalentes, desde que no método bayesiano seja atribuída uma distribuição a priori não informativa para \mathbf{b} como mencionado em Jiang (1997).

O método MV consiste em obter os estimadores dos parâmetros por meio da maximização do logaritmo da verossimilhança marginal dos dados, nomeadamente

$$l(\boldsymbol{\beta}, \boldsymbol{\theta}) = -\frac{N}{2} \log 2\pi - \frac{1}{2} \log |\boldsymbol{\Omega}(\boldsymbol{\theta})| - \frac{1}{2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top [\boldsymbol{\Omega}(\boldsymbol{\theta})]^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \quad (2.3.1)$$

ou alternativamente

$$\begin{aligned} l(\boldsymbol{\beta}, \boldsymbol{\theta}) = & -\frac{1}{2} \sum_{i=1}^n m_i \log 2\pi - \frac{1}{2} \sum_{i=1}^n \log |\boldsymbol{\Omega}_i(\boldsymbol{\theta})| \\ & - \frac{1}{2} \sum_{i=1}^n (\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})^\top [\boldsymbol{\Omega}_i(\boldsymbol{\theta})]^{-1} (\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}). \end{aligned} \quad (2.3.2)$$

A maximização de (2.3.2) pode ser realizada por meio dos seguintes passos:

i) Igualar $\partial l(\boldsymbol{\beta}, \boldsymbol{\theta}) / \partial \boldsymbol{\beta}$ a $\mathbf{0}$, obtendo

$$\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}) = \left[\sum_{i=1}^n \mathbf{X}_i^\top [\boldsymbol{\Omega}_i(\boldsymbol{\theta})]^{-1} \mathbf{X}_i \right]^{-1} \left[\sum_{i=1}^n \mathbf{X}_i^\top [\boldsymbol{\Omega}_i(\boldsymbol{\theta})]^{-1} \mathbf{y}_i \right]. \quad (2.3.3)$$

Como

$$\partial^2 l(\boldsymbol{\beta}, \boldsymbol{\theta}) / \partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top = - \sum_{i=1}^n \mathbf{X}_i^\top [\boldsymbol{\Omega}_i(\boldsymbol{\theta})]^{-1} \mathbf{X}_i$$

é uma matriz definida negativa, (2.3.3) corresponde ao ponto em que $l(\boldsymbol{\beta}, \boldsymbol{\theta})$ atinge o máximo. Também pode-se concluir que (2.3.3) corresponde ao estimador de **mínimos quadrados generalizados** de $\boldsymbol{\beta}$ sob a suposição de que $\boldsymbol{\Omega}_i(\boldsymbol{\theta})$ é conhecida, *i.e.*, $\hat{\boldsymbol{\beta}}(\boldsymbol{\theta})$ minimiza a forma quadrática

$$\sum_{i=1}^n [\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}(\boldsymbol{\theta})]^\top [\boldsymbol{\Omega}_i(\boldsymbol{\theta})]^{-1} [\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}(\boldsymbol{\theta})].$$

- ii) Substituir $\hat{\beta}(\theta)$ na expressão (2.3.2) obtendo a função **log-verossimilhança perfilada** $l[\hat{\beta}(\theta), \theta]$;
- iii) Igualar $\partial l[\hat{\beta}(\theta), \theta]/\partial \theta$ a $\mathbf{0}$, obtendo o sistema de equações de estimação

$$-\frac{1}{2} \sum_{i=1}^n \text{tr} \left[[\Omega_i(\hat{\theta})]^{-1} \dot{\Omega}_i(\hat{\theta}) \right] - \frac{1}{2} \sum_{i=1}^n [\partial Q_i(\theta)/\partial \theta_j]_{\theta=\hat{\theta}} = 0, \quad (2.3.4)$$

$j = 1, \dots, t$, em que

$$\dot{\Omega}_i(\hat{\theta}) = [\partial \Omega_i(\theta)/\partial \theta_j]_{\theta=\hat{\theta}}^\top$$

e

$$Q_i(\theta) = [\mathbf{y}_i - \mathbf{X}_i \hat{\beta}(\theta)]^\top [\Omega_i(\theta)]^{-1} [\mathbf{y}_i - \mathbf{X}_i \hat{\beta}(\theta)].$$

Como a matriz $\mathbf{H} = \mathbf{H}(\theta) = \partial^2 l[\beta(\theta), \theta]/\partial \theta \partial \theta^\top$ cujo elemento (r, s) , $r, s = 1, \dots, t$ é dado por

$$\begin{aligned} [\mathbf{H}]_{rs} &= -\frac{1}{2} \sum_{i=1}^n \text{tr} \{ \Omega_i^{-1} \dot{\Omega}_{ir} \Omega_i^{-1} (2\epsilon_i \epsilon_i^\top - \Omega_i) \Omega_i^{-1} \dot{\Omega}_{is} \} + \\ &\quad + \frac{1}{2} \text{tr} \{ \Omega_i^{-1} (\epsilon_i \epsilon_i^\top - \Omega_i) \Omega_i^{-1} \ddot{\Omega}_{i,rs} \} \end{aligned}$$

em que $\Omega_i = \Omega_i(\theta)$, $\epsilon_i = \mathbf{y}_i - \mathbf{X}_i \beta$, $\dot{\Omega}_{ir} = \partial \Omega_i(\theta)/\partial \theta_r$ e $\ddot{\Omega}_{i,rs} = \partial^2 \Omega_i(\theta)/\partial \theta_r \partial \theta_s$ é definida negativa, a solução $\hat{\theta}$ de (2.3.4) corresponde ao ponto de máximo de $l[\beta(\theta), \theta]$ e consequentemente é o estimador de MV de θ . Se substituirmos θ por $\hat{\theta}$ em (2.3.3) obtemos o estimador de MV de β .

Detalhes sobre o cálculo das derivadas do logaritmo da função de verossimilhança estão esboçados na Seção A.5 do Apêndice A.

A metodologia de máxima verossimilhança gera estimadores não-enviesados para os efeitos fixos mas produz estimadores viesados para os parâmetros da matriz de covariâncias intraunidades amostrais por não levar em consideração a perda de graus de liberdade na estimação dos primeiros. A fim de reduzir o viés desses estimadores, que pode ser grande quando o número de parâmetros de localização é grande em relação ao número de observações, muitos autores, como Laird & Ware (1982), recomendam o uso do método de **máxima verossimilhança restrita** (MVR). Este método foi proposto por Patterson & Thompson (1971) para estimar componentes de variância e consiste em maximizar a verossimilhança de uma transformação linear ortogonal do tipo $\mathbf{y}^\dagger = \mathbf{U}^\top \mathbf{y}$ em que \mathbf{U} tem dimensão $[N \times (N - p)]$, e é tal que $\mathbb{E}(\mathbf{y}^\dagger) = \mathbf{0}$, ou seja, $\mathbf{U}^\top \mathbf{X} = \mathbf{0}$. A verossimilhança obtida a partir desta transformação não depende dos efeitos fixos β e tampouco da particular matriz \mathbf{U} escolhida. Em geral obtém-se a verossimilhança restrita fazendo $\mathbf{U} = \mathbf{I} - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top$,

ou seja, utilizando a matriz de projeção que gera os resíduos do ajuste obtido por mínimos quadrados ordinários. Daí a denominação de **verossimilhança residual** empregada por alguns autores. Para os dados transformados, temos

$$\mathbb{E}(\mathbf{y}^\dagger) = \mathbf{0} \quad (2.3.5)$$

$$\mathbb{V}(\mathbf{y}^\dagger) = \mathbf{U}\boldsymbol{\Omega}(\boldsymbol{\theta})\mathbf{U}^\top, \quad (2.3.6)$$

e além disso, $\mathbf{y}^\dagger \sim N_{N-p}[\mathbf{0}, \mathbf{U}^\top \boldsymbol{\Omega}(\boldsymbol{\theta}) \mathbf{U}]$. O logaritmo da verossimilhança marginal restrita é

$$\begin{aligned} l_R(\boldsymbol{\theta}) = & -\frac{1}{2} \log |\boldsymbol{\Omega}(\boldsymbol{\theta})| - \frac{1}{2} \log |\mathbf{X}^\top [\boldsymbol{\Omega}(\boldsymbol{\theta})]^{-1} \mathbf{X}| \\ & - \frac{N-p}{2} \{\mathbf{y} - \mathbf{X}[\hat{\boldsymbol{\beta}}(\boldsymbol{\theta})]\}^\top [\boldsymbol{\Omega}(\boldsymbol{\theta})]^{-1} \{\mathbf{y} - \mathbf{X}[\hat{\boldsymbol{\beta}}(\boldsymbol{\theta})]\} - \frac{N-p}{2} \log(2\pi), \end{aligned} \quad (2.3.7)$$

e sua maximização gera os estimadores de MVR de $\boldsymbol{\theta}$.

Alternativamente, a função log-verossimilhança restrita (2.3.7) pode ser escrita como

$$\begin{aligned} l_R(\boldsymbol{\theta}) = & -\frac{1}{2} \sum_{i=1}^n m_i \log 2\pi - \frac{1}{2} \sum_{i=1}^n \log |\boldsymbol{\Omega}_i(\boldsymbol{\theta})| \\ & - \frac{1}{2} \sum_{i=1}^n [\mathbf{y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}}(\boldsymbol{\theta})]^\top [\boldsymbol{\Omega}_i(\boldsymbol{\theta})]^{-1} [\mathbf{y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}}(\boldsymbol{\theta})] \\ & - \frac{1}{2} \log \left| \sum_{i=1}^n \mathbf{X}_i^\top [\boldsymbol{\Omega}_i(\boldsymbol{\theta})]^{-1} \mathbf{X}_i \right|. \end{aligned} \quad (2.3.8)$$

Neste caso, no processo de maximização, o sistema de equações de estimação (2.3.4) é substituído por

$$\begin{aligned} & -\frac{1}{2} \sum_{i=1}^n \text{tr}\{[\boldsymbol{\Omega}_i(\hat{\boldsymbol{\theta}})]^{-1} \dot{\boldsymbol{\Omega}}_i(\hat{\boldsymbol{\theta}})\} - \frac{1}{2} \sum_{i=1}^n [\partial Q_i(\boldsymbol{\theta}) / \partial \boldsymbol{\theta}_j |_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}] \\ & - \frac{1}{2} \sum_{i=1}^n \text{tr}\{[\boldsymbol{\Omega}_i(\hat{\boldsymbol{\theta}})]^{-1} \mathbf{X}_i^\top [\boldsymbol{\Omega}_i(\hat{\boldsymbol{\theta}})]^{-1} \dot{\boldsymbol{\Omega}}_i(\hat{\boldsymbol{\theta}}) [\boldsymbol{\Omega}_i(\hat{\boldsymbol{\theta}})]^{-1} \mathbf{X}_i\} = 0, \end{aligned} \quad (2.3.9)$$

$j = 1, \dots, t$. A solução de (2.3.9), digamos, $\hat{\boldsymbol{\theta}}_R$, em conjunto com $\hat{\boldsymbol{\beta}}_R = \hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\theta}}_R)$ constituem os estimadores de MVR desejados. Detalhes podem ser obtidos em Demidenko (2013).

Preditores dos efeitos aleatórios podem ser obtidos juntamente com estimadores para os parâmetros de localização por meio da função de distribuição conjunta dos efeitos aleatórios \mathbf{b} e das observações \mathbf{y} , nomeadamente

$$f(\mathbf{y}, \mathbf{b}) = f(\mathbf{y}|\mathbf{b})f(\mathbf{b}), \quad (2.3.10)$$

em que $f(\mathbf{y}|\mathbf{b})$ e $f(\mathbf{b})$ são as funções densidade de $\mathbf{y}|\mathbf{b}$ e \mathbf{b} , respectivamente. Como $\mathbf{y}|\mathbf{b} \sim N_N(\mathbf{X}\boldsymbol{\beta} + \mathbf{Zb}, \mathbf{R})$ e $\mathbf{b} \sim N_{nq}(\mathbf{0}, \mathbf{\Gamma})$, a função (2.3.10) pode ser diretamente obtida, como indicam Henderson (1975), Searle (1971), Searle et al. (1992) ou McCulloch & Searle (2001), por exemplo.

O **melhor estimador linear não-enviesado** (*best linear unbiased estimator* - **BLUE**) para $\boldsymbol{\beta}$ e o **melhor preditor não-enviesado** (*best linear unbiased predictor* - **BLUP**) para o vetor de efeitos aleatórios \mathbf{b} podem ser obtidos por meio do teorema de Gauss-Markov para efeitos aleatórios apresentado em Harville (1977a). Diferentes formas de obtenção do BLUP e do BLUE, tanto sob o ponto de vista clássico como bayesiano e aplicações podem ser encontradas em Robinson (1991), Searle et al. (1992), Doganaksoy & Balakrishnan (1997), Jiang (1997) ou McCulloch & Searle (2001), por exemplo.

Hilden-Minton (1995) e Hodges (1998) comentam que existe uma série de vantagens em se utilizar **casos com restrição** (*constraint-cases*) para a obtenção do BLUE e do BLUP. A ideia básica é reexpressar o modelo linear misto (2.1.7) na forma de um modelo linear usual por meio da inclusão dos “casos artificiais” com variâncias desconhecidas apresentados em Hodges (1998). Dentre as vantagens citadas, destacam-se a obtenção “imediata” das equações de estimação utilizadas para determinar os BLUE e os BLUP e sua consequente utilização para diagnóstico. Nesse contexto, consideremos o modelo (2.1.7) com a inclusão do seguinte “caso artificial”,

$$\mathbf{0}_{nq \times 1} = \mathbf{0}_{nq \times 1} - \mathbf{I}_{nq} \mathbf{b} + \mathbf{d}, \quad (2.3.11)$$

em que \mathbf{d} é um vetor de dimensão $(nq \times 1)$ que faz o papel de erro, $\mathbb{V}(\mathbf{e}) = \sigma^2 \mathbf{R}(\boldsymbol{\theta})$, $\mathbb{V}(\mathbf{d}) = \sigma^2 \mathbf{\Gamma}(\boldsymbol{\theta})$ e $\text{Cov}(\mathbf{e}, \mathbf{d}) = \mathbf{0}$ e reescrevamos as equações (2.1.7) e (2.3.11) em forma matricial como

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} \mathbf{X} & \mathbf{Z} \\ \mathbf{0} & -\mathbf{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta} \\ \mathbf{b} \end{bmatrix} + \begin{bmatrix} \mathbf{e} \\ \mathbf{d} \end{bmatrix}. \quad (2.3.12)$$

Pré-multiplicando (2.3.12) por $[\mathbf{R}(\boldsymbol{\theta})]^{-1/2} \oplus [\mathbf{\Gamma}(\boldsymbol{\theta})]^{-1/2}$ obtemos

$$\mathbf{y}^* = \mathbf{X}^* \boldsymbol{\beta}^* + \mathbf{e}^*, \quad (2.3.13)$$

em que,

$$\mathbf{y}^* = \begin{bmatrix} [\mathbf{R}(\boldsymbol{\theta})]^{-1/2} \mathbf{y} \\ \mathbf{0} \end{bmatrix}, \quad \mathbf{X}^* = \begin{bmatrix} [\mathbf{R}(\boldsymbol{\theta})]^{-1/2} \mathbf{X} & [\mathbf{R}(\boldsymbol{\theta})]^{-1/2} \mathbf{Z} \\ \mathbf{0} & -[\mathbf{\Gamma}(\boldsymbol{\theta})]^{-1/2} \end{bmatrix}, \quad (2.3.14)$$

$\boldsymbol{\beta}^* = [\boldsymbol{\beta}^\top, \mathbf{b}^\top]^\top$ e $\mathbf{e}^* = [\mathbf{R}(\boldsymbol{\theta})]^{-1/2} \oplus [\mathbf{\Gamma}(\boldsymbol{\theta})]^{-1/2} (\mathbf{e}^\top, \mathbf{d}^\top)^\top$ é tal que $\mathbb{V}(\mathbf{e}^*) = \sigma^2 \mathbf{I}$.

Desta forma, (2.3.12) pode ser considerado como um modelo linear “homocedástico”. Portanto, o BLUE para $\boldsymbol{\beta}$ e o BLUP para \mathbf{b} podem ser obtidos por meio de (2.3.12) usando o método de mínimos quadrados, sob o qual se obtém

$$\mathbf{X}^{*\top} \mathbf{X}^* \hat{\boldsymbol{\beta}}^* = \mathbf{X}^{*\top} \mathbf{y}^*$$

ou seja

$$\begin{bmatrix} \mathbf{X}^\top [\mathbf{R}(\boldsymbol{\theta})]^{-1} \mathbf{X} & \mathbf{X}^\top [\mathbf{R}(\boldsymbol{\theta})]^{-1} \mathbf{Z} \\ \mathbf{Z}^\top [\mathbf{R}(\boldsymbol{\theta})]^{-1} \mathbf{X} & \mathbf{Z}^\top [\mathbf{R}(\boldsymbol{\theta})]^{-1} \mathbf{Z} + [\boldsymbol{\Gamma}(\boldsymbol{\theta})]^{-1} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{b}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^\top [\mathbf{R}(\boldsymbol{\theta})]^{-1/2} \mathbf{y} \\ \mathbf{Z}^\top [\mathbf{R}(\boldsymbol{\theta})]^{-1/2} \mathbf{y} \end{bmatrix}. \quad (2.3.15)$$

Essas equações são conhecidas como **equações de Henderson**. Note que, se $[\boldsymbol{\Gamma}(\boldsymbol{\theta})]^{-1} \equiv \mathbf{0}$ (o que implica que \mathbf{b} é um efeito fixo), então (2.3.15) coincide com as equações de estimação obtidas por meio do método de **mínimos quadrados generalizados** (MQG). O BLUP e o BLUE são obtidos como solução das equações (2.3.15), que independem das distribuições de \mathbf{b} e \mathbf{e} .

Pode-se mostrar que o BLUE de $\boldsymbol{\beta}$ e o BLUP de \mathbf{b} são dados, respectivamente, por

$$\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}) = [\mathbf{X}^\top [\boldsymbol{\Omega}(\boldsymbol{\theta})]^{-1} \mathbf{X}]^{-1} \mathbf{X}^\top [\boldsymbol{\Omega}(\boldsymbol{\theta})]^{-1} \mathbf{y} \quad (2.3.16)$$

e

$$\hat{\mathbf{b}}(\boldsymbol{\theta}) = \boldsymbol{\Gamma}(\boldsymbol{\theta}) \mathbf{Z}^\top [\boldsymbol{\Omega}(\boldsymbol{\theta})]^{-1} [\mathbf{y} - \mathbf{X} \hat{\boldsymbol{\beta}}(\boldsymbol{\theta})] = \boldsymbol{\Gamma}(\boldsymbol{\theta}) \mathbf{Z}^\top \mathbf{Q}(\boldsymbol{\theta}) \mathbf{y} \quad (2.3.17)$$

em que

$$\mathbf{Q}(\boldsymbol{\theta}) = [\boldsymbol{\Omega}(\boldsymbol{\theta})]^{-1} - [\boldsymbol{\Omega}(\boldsymbol{\theta})]^{-1} \mathbf{X} [\mathbf{X}^\top [\boldsymbol{\Omega}(\boldsymbol{\theta})]^{-1} \mathbf{X}]^{-1} \mathbf{X}^\top [\boldsymbol{\Omega}(\boldsymbol{\theta})]^{-1}. \quad (2.3.18)$$

Propriedades de $\hat{\boldsymbol{\beta}}(\boldsymbol{\theta})$ e $\hat{\mathbf{b}}(\boldsymbol{\theta})$ são apresentadas em Henderson (1975), Robinson (1991), Searle et al. (1992) ou McCulloch & Searle (2001), por exemplo. Em particular, Henderson (1975) mostra que

$$\mathbb{E}(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}, \quad \mathbb{V}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}^\top \boldsymbol{\Omega}^{-1} \mathbf{X})^{-1}, \quad (2.3.19)$$

$$\mathbb{E}(\hat{\mathbf{b}}) = \mathbf{0}, \quad \mathbb{V}(\hat{\mathbf{b}}) = \boldsymbol{\Gamma} \mathbf{Z}^\top \mathbf{Q} \mathbf{Z} \boldsymbol{\Gamma}. \quad (2.3.20)$$

e que

$$\mathbb{Cov} \begin{bmatrix} \hat{\boldsymbol{\beta}}(\boldsymbol{\theta}) - \boldsymbol{\beta} \\ \hat{\mathbf{b}}(\boldsymbol{\theta}) - \mathbf{b} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^\top [\mathbf{R}(\boldsymbol{\theta})]^{-1} \mathbf{X} & \mathbf{X}^\top [\mathbf{R}(\boldsymbol{\theta})]^{-1} \mathbf{Z} \\ \mathbf{Z}^\top [\mathbf{R}(\boldsymbol{\theta})]^{-1} \mathbf{X} & \mathbf{Z}^\top [\mathbf{R}(\boldsymbol{\theta})]^{-1} \mathbf{Z} + [\boldsymbol{\Gamma}(\boldsymbol{\theta})]^{-1} \end{bmatrix}^{-1}. \quad (2.3.21)$$

Além disso, o estimador do parâmetro de escala σ^2 é

$$\hat{\sigma}^2 = \frac{1}{N - p} [\mathbf{y} - \mathbf{X} \hat{\boldsymbol{\beta}}(\boldsymbol{\theta})]^\top \{ \mathbf{Z} \boldsymbol{\Gamma}(\boldsymbol{\theta}) \mathbf{Z}^\top + [\mathbf{R}(\boldsymbol{\theta})] \}^{-1} [\mathbf{y} - \mathbf{X} \hat{\boldsymbol{\beta}}(\boldsymbol{\theta})]. \quad (2.3.22)$$

Alternativamente, podemos expressar $\hat{\boldsymbol{\beta}}(\boldsymbol{\theta})$ e $\hat{\mathbf{b}}_i(\boldsymbol{\theta})$ como

$$\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}) = \sum_{i=1}^n \{ \mathbf{X}_i^\top [\boldsymbol{\Omega}_i(\boldsymbol{\theta})]^{-1} \mathbf{X}_i \}^{-1} \mathbf{X}_i [\boldsymbol{\Omega}_i(\boldsymbol{\theta})]^{-1} \mathbf{y}_i \quad (2.3.23)$$

e

$$\hat{\mathbf{b}}_i(\boldsymbol{\theta}) = \mathbf{G}(\boldsymbol{\theta}_1) \mathbf{Z}_i^\top [\boldsymbol{\Omega}_i(\boldsymbol{\theta})]^{-1} [\mathbf{y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}}(\boldsymbol{\theta})] \quad (2.3.24)$$

e então,

$$\mathbb{V}[\hat{\mathbf{b}}_i(\boldsymbol{\theta})] = \mathbf{G}(\boldsymbol{\theta}) \mathbf{Z}_i^\top \mathbf{Q}(\boldsymbol{\theta}) \mathbf{Z}_i \mathbf{G}(\boldsymbol{\theta}). \quad (2.3.25)$$

Como $\boldsymbol{\Gamma}(\boldsymbol{\theta})$ e $\mathbf{R}(\boldsymbol{\theta})$ dependem do vetor de parâmetros de covariância desconhecido $\boldsymbol{\theta}$, é razoável calcular o BLUE e BLUP com base num estimador $\hat{\boldsymbol{\theta}}$ de $\boldsymbol{\theta}$; esses “preditores” são denominados BLUE e BLUP empíricos (**EBLUE** e **EBLUP**). Se $\hat{\boldsymbol{\theta}}$ é o estimador de MV de $\boldsymbol{\theta}$, então o EBLUE e EBLUP, são respectivamente, o estimador de MV de $\boldsymbol{\beta}$ e o preditor empírico de Bayes dos efeitos aleatórios sob a hipótese de normalidade de \mathbf{b} e \mathbf{e} . Sob algumas condições, tanto o BLUP quanto o BLUE empíricos continuam não-enviesados, como mostram Kackar & Harville (1984) e Jiang (1999). Verbeke & Lesaffre (1996) mostram que o EBLUE e o estimador de MV de $\boldsymbol{\theta}$ são assintoticamente normais, mesmo quando a distribuição de \mathbf{b} é especificada incorretamente. Jiang (1998) obtém algumas propriedades assintóticas do EBLUP e EBLUE em modelos de componentes de variância, considerando o estimador de MVR para $\boldsymbol{\theta}$ sem supor normalidade de \mathbf{b} e \mathbf{e} .

Com base na suposição de normalidade adotada para os dados \mathbf{y} , e na expressão (2.3.3), podemos concluir que

$$\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}) \sim N_p\{\boldsymbol{\beta}, \mathbf{V}_{\hat{\boldsymbol{\beta}}}(\boldsymbol{\theta})\}.$$

em que $\mathbf{V}_{\hat{\boldsymbol{\beta}}}(\boldsymbol{\theta}) = \{\sum_{i=1}^n \mathbf{X}_i^\top [\boldsymbol{\Omega}_i(\boldsymbol{\theta})]^{-1} \mathbf{X}_i\}^{-1}$ e que

$$\hat{\boldsymbol{\theta}} \sim N_t\{\boldsymbol{\theta}, \mathbf{V}_{\hat{\boldsymbol{\theta}}}(\boldsymbol{\theta})\}$$

em que $\mathbf{V}_{\hat{\boldsymbol{\theta}}}(\boldsymbol{\theta})$ é uma matriz com dimensão $(t \times t)$ cujo elemento (r, s) , $r, s = 1, \dots, t$, é dado por

$$[\mathbf{V}_{\hat{\boldsymbol{\theta}}}(\boldsymbol{\theta})]_{rs} = \frac{1}{2} \sum_{i=1}^n \text{tr}\{[\boldsymbol{\Omega}_i(\boldsymbol{\theta})]^{-1} \dot{\boldsymbol{\Omega}}_{ir}(\boldsymbol{\theta}) [\boldsymbol{\Omega}_i(\boldsymbol{\theta})]^{-1} \dot{\boldsymbol{\Omega}}_{is}(\boldsymbol{\theta})\}.$$

Por outro lado, como tanto $\hat{\boldsymbol{\theta}}$ quanto $\hat{\boldsymbol{\theta}}_R$ são estimadores consistentes de $\boldsymbol{\theta}$, uma aplicação do Teorema de Sverdrup permite-nos mostrar que para amostras suficientemente grandes, *i.e.*, para n suficientemente grande, tanto a distribuição de $\hat{\boldsymbol{\beta}}$ quanto a de $\hat{\boldsymbol{\theta}}$ podem ser aproximadas, respectivamente, pelas mesmas distribuições utilizadas para $\hat{\boldsymbol{\beta}}(\boldsymbol{\theta})$ e $\hat{\boldsymbol{\theta}}(\boldsymbol{\theta})$ sob certas condições de regularidade para as matrizes de especificação \mathbf{X}_i . Resultados semelhantes valem para os estimadores de máxima verossimilhança restrita. Detalhes podem ser obtidos em Sen, Singer & Pedroso-de Lima (2009), Jiang (1996), Verbeke & Lesaffre (1996) e Demidenko (2013), por exemplo.

Esses resultados servem de base para a construção de testes de hipóteses sobre os parâmetros da matriz de covariâncias ($\boldsymbol{\theta}$) e sobre os parâmetros de localização $\boldsymbol{\beta}$. Testes de hipóteses do tipo

$$H : \mathbf{C}\boldsymbol{\theta} = \mathbf{0} \quad (2.3.26)$$

em que \mathbf{C} denota uma matriz com dimensão $(c \times t)$ de posto c são úteis para a seleção de modelos mais parcimoniosos para a estrutura de covariâncias intraunidades amostrais sob o mesmo modelo para os efeitos fixos.

Em princípio, para testar hipóteses dessa forma, podemos empregar a estatística de **Wald**

$$Q_W = (\mathbf{C}\hat{\boldsymbol{\theta}})^\top [\mathbf{C}\mathbf{V}(\hat{\boldsymbol{\theta}})\mathbf{C}^\top]^{-1}(\mathbf{C}\hat{\boldsymbol{\theta}}), \quad (2.3.27)$$

ou a estatística da razão de verossimilhanças (RV)

$$Q_{RV} = -2[l_1(\hat{\boldsymbol{\theta}}) - l_2(\hat{\boldsymbol{\theta}})], \quad (2.3.28)$$

em que $l_1(\hat{\boldsymbol{\theta}})$ e $l_2(\hat{\boldsymbol{\theta}})$ denotam, respectivamente, a função log-verossimilhança maximizada sob o modelo restrito *i.e.*, induzido por (2.3.26) e aquela maximizada sob o modelo irrestrito. Sob as condições de regularidade usuais e para amostras suficientemente grandes, as distribuições de (2.3.27) e de (2.3.28) podem ser aproximadas por distribuições χ_c^2 quando a hipótese nula é verdadeira. No entanto, em muitos casos, as hipóteses de interesse são do tipo

$$H_0 : \sigma_a^2 = 0 \text{ versus } H_1 : \sigma_a^2 \in (0, \infty) \quad (2.3.29)$$

em que σ_a^2 , um dos componentes de $\boldsymbol{\theta}$, é localizado na fronteira do espaço paramétrico. Nesse contexto, tanto os testes de Wald quanto os testes da razão de verossimilhanças apresentam problemas com relação às condições de regularidade conforme mencionam Cox & Hinkley (1974). O problema é causado pela violação da primeira condição de regularidade, que determina que o espaço paramétrico Ω deve ter dimensão finita, ser fechado e compacto e que o verdadeiro valor do parâmetro deverá estar em seu interior. Tanto num caso como no outro, a obtenção da distribuição aproximada das estatísticas de teste são baseadas em expansões de Taylor numa vizinhança do verdadeiro valor do parâmetro.

Self & Liang (1987) mostram que nesses casos, a distribuição da estatística RV pode ser aproximada por uma mistura de distribuições qui-quadrado. Stram & Lee (1994) discutem o comportamento da estatística RV para componentes de variância em estudos longitudinais com efeitos aleatórios, aplicando o método descrito por Self & Liang (1987) e particularizam os resultados para várias situações de interesse. De uma maneira geral, para testar hipótese sobre a inclusão de um efeito aleatório em um modelo com q efeitos aleatórios, a estatística do teste deverá ser comparada a uma mistura 50%:50% de distribuições qui-quadrado, χ_q^2 e χ_{q+1}^2 . O índice $q + 1$

representa o número de parâmetros adicionados ao modelo sob a hipótese alternativa, H_1 , e o correspondente valor-P é dado por $p = 0.5\mathbb{P}[\chi_q^2 > Q_{RV}(obs)] + 0.5\mathbb{P}[\chi_{q+1}^2 > Q_{RV}(obs)]$ em que $Q_{RV}(obs)$ é o valor observado da estatística (2.3.28).

Giampaoli & Singer (2009) observam que Stram & Lee (1994) não levaram em consideração o fato de que, em geral, em estudos longitudinais as variáveis não são identicamente distribuídas e estudam o comportamento de testes de hipóteses sobre os elementos da matriz de covariância num modelo de efeitos aleatórios utilizando a estatística da razão de verossimilhanças apresentada por Vu & Zhou (1997). Esses autores estendem os resultados obtidos por Self & Liang (1987) para o caso em que as variáveis aleatórias não são identicamente distribuídas. Giampaoli & Singer (2009) mostram que nos modelos com até dois efeitos aleatórios, mesmo após as devidas adaptações, os resultados de Stram & Lee (1994) continuam válidos.

Como uma ferramenta auxiliar para a seleção de estruturas de covariância intraunidades amostrais, consideramos os critérios de informação de Akaike (AIC_R) e de Schwarz (BIC_R) definidos por

$$AIC_R = -2l(\hat{\boldsymbol{\theta}}) + 2t \quad (2.3.30)$$

$$BIC_R = -2l(\hat{\boldsymbol{\theta}}) + t \log(N - p), \quad (2.3.31)$$

em que $l(\hat{\boldsymbol{\theta}})$ é o máximo da função log-verossimilhança restrita e t é o número de parâmetros da matriz de covariâncias, p é o número de efeitos fixos e N é o número efetivo de observações. A estrutura que apresenta os menores valores de AIC_R e BIC_R é considerada mais adequada. Este critério é conveniente por permitir a comparação de quaisquer estruturas de covariância, exigindo apenas que o modelo para os parâmetros de localização seja o mesmo. Mais detalhes sobre o uso deste critério na seleção de estruturas de covariância em estudos com medidas repetidas são encontrados em Keselman, James, Rhonda & Russell (1998).

Funções do tipo $\mathbf{L}\boldsymbol{\beta} + \mathbf{D}\mathbf{b}$, em que \mathbf{L} e \mathbf{D} são matrizes conhecidas com dimensões $(l \times p)$ e $(d \times nq)$ com postos l e d , respectivamente, são chamadas **funções previsíveis** por Henderson (1975) e constituem generalizações das **funções estimáveis** discutidas em Searle (1971), por exemplo. Elas podem ser utilizadas para definir hipóteses sobre efeitos fixos ou sobre combinações de efeitos fixos e aleatórios. Nesse contexto, basta tomar $\mathbf{K} = [\mathbf{L}\mathbf{D}]$ e considerar

$$H : \mathbf{K} \begin{bmatrix} \boldsymbol{\beta} \\ \mathbf{b} \end{bmatrix} = \mathbf{0}. \quad (2.3.32)$$

A estatística de Wald para testar (2.3.32) é

$$Q_W = \left(\mathbf{K} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{b}} \end{bmatrix} \right)^\top [\mathbf{K}\widehat{\mathbf{W}}\mathbf{K}^\top]^{-1} \mathbf{K} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{b}} \end{bmatrix} \quad (2.3.33)$$

em que $\widehat{\mathbf{W}}$ é uma estimativa da matriz de covariâncias de $(\widehat{\boldsymbol{\beta}}^\top, \widehat{\mathbf{b}}^\top)^\top$ (2.3.21). A distribuição aproximada da estatística Q_W é $\chi^2_{\text{posto}(\mathbf{K})}$. Se estivermos interessados apenas em contrastes dos efeitos fixos, basta considerar nulas as colunas associadas aos efeitos aleatórios na matriz \mathbf{K} .

De uma forma mais conservadora, podemos utilizar uma estatística F dividindo (2.3.33) pelo posto de \mathbf{K} . O número de graus de liberdade do numerador é dado pelo posto de \mathbf{K} e o do denominador, deve ser obtido por meio de um procedimento de Satterthwaite conforme a sugestão de Searle (1971).

Intervalos de confiança aproximados para funções previsíveis em que \mathbf{K} tem uma única linha podem ser construídos como

$$\mathbf{K} \begin{bmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\mathbf{b}} \end{bmatrix} \pm t_{v, \frac{\alpha}{2}} \sqrt{[\mathbf{K}\widehat{\mathbf{W}}\mathbf{K}^\top]}. \quad (2.3.34)$$

com $t_{v, \frac{\alpha}{2}}$ denotando o quantil de ordem $1 - \alpha/2$ de uma distribuição t -Student com v graus de liberdade.

Hipóteses do tipo (2.3.32) também podem ser testadas por meio de testes da razão de verossimilhanças.

2.4 Solução das equações de estimação

Embora em geral, o sistema de equações de estimação formado por (2.3.3) e (2.3.4) ou (2.3.9) precise ser resolvido por métodos iterativos, em certos casos especiais com dados balanceados, existem soluções explícitas. O exemplo clássico é o modelo de **análise de perfis** com matriz de covariâncias intraunidades amostrais não-estruturada. Esse modelo é apropriado para comparações de perfis de respostas médias de unidades amostrais submetidas a diferentes tratamentos e avaliadas nos mesmos instantes (não necessariamente igualmente espaçados). Essencialmente, o modelo corresponde a (2.1.2) com $m_i = m$, $\mathbf{Z}_i = \mathbf{0}$ e em que $\mathbb{V}(\mathbf{e}_i) = \mathbf{R}_i = \mathbf{R}$ é uma matriz de covariâncias não-estruturada. Por razões computacionais, convém reescrever o modelo na forma

$$\mathbf{Y} = \mathbf{X}\mathbf{M} + \mathbf{E}, \quad (2.4.1)$$

em que $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_n]^\top$ é uma matriz com dimensão $(n \times m)$, $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_p]$ é uma matriz com dimensão $(n \times p)$ cujos elementos \mathbf{x}_i são vetores de incidência que associam as n unidades amostrais aos tratamentos a que foram submetidas, \mathbf{M} é uma matriz com dimensão $(p \times m)$ cujo elemento na posição (i, j) , $i = 1, \dots, p$, $j = 1, \dots, m$ corresponde à resposta esperada para unidades submetidas ao tratamento

i no instante j e $\mathbf{E} = [\mathbf{e}_1, \dots, \mathbf{e}_n]^\top$ é uma matriz de erros aleatórios com dimensão $(n \times m)$ cujas linhas seguem distribuições normais m -variadas com vetor de médias nulo e matriz de covariâncias \mathbf{R} .

Utilizando métodos padrão de **Análise de Variância Multivariada** (MANOVA) podemos mostrar que os estimadores de MV restrita dos parâmetros de localização e de covariância são respectivamente

$$\widehat{\mathbf{M}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y}$$

e

$$\mathbf{S} = \frac{1}{n-p} \mathbf{Y}^\top [\mathbf{I} - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top] \mathbf{Y}.$$

A expressão acima com a substituição de $n-p$ por n corresponde ao estimador de MV. Neste caso, as hipóteses de interesse podem ser expressas na forma

$$H : \mathbf{C}\mathbf{M}\mathbf{U} = \mathbf{0} \quad (2.4.2)$$

em que \mathbf{C} e \mathbf{U} são matrizes conhecidas com dimensões $(c \times p)$ e $(m \times u)$ respectivamente. Testes de hipóteses dessa forma podem ser obtidos por meio de diferentes princípios como máxima verossimilhança, **união-intersecção** etc. e as estatísticas de teste correspondentes são funções das raízes características da matriz $\mathbf{H}\mathbf{E}^{-1}$ em que

$$\mathbf{H} = (\widehat{\mathbf{C}\mathbf{M}\mathbf{U}})^\top [\mathbf{C}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{C}^\top]^{-1} (\widehat{\mathbf{C}\mathbf{M}\mathbf{U}})$$

e

$$\mathbf{E} = (\mathbf{Y}\mathbf{U})^\top [\mathbf{I} - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top] (\mathbf{Y}\mathbf{U})$$

são, respectivamente, as matrizes de somas de quadrados e produtos cruzados devidas à hipótese e ao erro. Mais especificamente, definindo $\theta_i = \lambda_i(1 + \lambda_i)^{-1}$ com λ_i denotando a i -ésima raiz característica de $\mathbf{H}\mathbf{E}^{-1}$, $\lambda_1 \leq \dots \leq \lambda_s$ com $s = \min(u, c)$, as principais estatísticas de teste são

i) $\theta_s = \max(\theta_i)$ (**estatística de Roy**);

ii) $\Lambda = \prod_{i=1}^s (1 - \theta_i)$ (**estatística de Wilks**);

iii) $T = \sum_{i=1}^s \theta_i(1 - \theta_i)$ (**estatística de Lawley-Hotelling**);

iv) $P = \sum_{i=1}^s \theta_i$ (**estatística de Pillai**).

A distribuição (exata) dessas estatísticas depende de $k_1 = (|u - c| - 1)/2$, $k_2 = (n - p - u - 1)/2$ e $s = \min(u, c)$ e pode ser aproximada por distribuições F . Em particular, quando $s = 1$, a estatística $F = (k_2 + 1/k_1 + 1)[(1 - \Lambda)/\Lambda]$ tem distribuição (exata) F com $2(k_1 + 1)$ graus de liberdade no numerador e $2(k_2 + 1)$

graus de liberdade no denominador sob a hipótese nula $\mathbf{CMU} = \mathbf{0}$. Intervalos de confiança simultâneos com coeficiente de confiança $1 - \alpha$ para funções do tipo $\mathbf{a}^\top \mathbf{M} \mathbf{b}$ em que \mathbf{a} e \mathbf{b} são vetores de constantes podem ser construídos como

$$\mathbf{a}^\top \widehat{\mathbf{M}} \mathbf{b} \pm \theta_s(\alpha) \{[\mathbf{b}^\top \mathbf{E} \mathbf{b}][\mathbf{a}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{a}]\}^{1/2},$$

em que $\theta_s(\alpha)$ denota o percentil de ordem $100(1 - \alpha)\%$ da distribuição da estatística de Roy. O leitor poderá consultar Morrison (1988) para detalhes sobre esse problema ou para outros mais complexos, em que os tratamentos correspondem aos cruzamentos dos níveis de diferentes fatores ou em que existem covariáveis. Andrade & Singer (1998) discutem uma estratégia de análise em que planejamentos em blocos aleatorizados complexos podem ser acomodados.

Métodos MANOVA também podem ser empregados para analisar dados para os quais $\mathbf{M} = \mathbf{BQ}$ em que \mathbf{B} é uma matriz com dimensão $(p \times r)$ de coeficientes desconhecidos, \mathbf{Q} é uma matriz conhecida com dimensão $(r \times m)$ cujas linhas correspondem aos valores de polinômios ortogonais de graus 0 a $r - 1$ nos p instantes de observação. Esse é o modelo de **curvas de crescimento** de Potthoff & Roy (1964), nomeadamente

$$\mathbf{Y} = \mathbf{XBQ} + \mathbf{E}. \quad (2.4.3)$$

Khatri (1966) mostrou que os estimadores propostos por Potthoff & Roy (1964), a saber,

$$\widehat{\mathbf{B}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y} \widehat{\mathbf{R}}^{-1} \mathbf{Q}^\top (\mathbf{Q} \widehat{\mathbf{R}}^{-1} \mathbf{Q}^\top)^{-1}$$

e

$$\widehat{\mathbf{R}} = \frac{1}{n} [\mathbf{Y} - \mathbf{XBQ}]^\top [\mathbf{Y} - \mathbf{XBQ}]$$

são estimadores de MV. O estimador de MVR de \mathbf{R} é obtido com a substituição de n por $n - p$ na expressão acima. Testes de hipóteses sobre os parâmetros de localização \mathbf{B} podem ser obtidos similarmente àqueles considerados para análise de perfis.

Muitos autores contribuíram para o estudo deste modelo nas últimas décadas. Entre eles, mencionamos Grizzle & Allen (1969), que sugeriram um procedimento interessante para implementação da técnica de **ajuste por covariância** proposta por Rao (1966). Essencialmente, essa técnica consiste de *i*) pós-multiplicação de ambos os membros de (2.4.3) por uma matriz $\mathbf{H} = [\mathbf{H}_1 \ \mathbf{H}_2]$ com dimensão $(m \times m)$ tal que $\mathbf{QH}_1 = \mathbf{I}_r$ and $\mathbf{QH}_2 = \mathbf{0}$ de forma que o modelo de curvas de crescimento fique reduzido a (2.4.1) com \mathbf{Y} , \mathbf{M} e \mathbf{E} respectivamente substituídas por $\mathbf{Y}_1 = \mathbf{YH}_1$, \mathbf{B} e $\mathbf{E} = \mathbf{EH}_1$ e *ii*) utilização de todas ou algumas colunas de $\mathbf{Y}_2 = \mathbf{YH}_2$, cujos valores esperados são nulos, como covariáveis para aumentar a eficiência dos estimadores dos parâmetros de localização. Essa técnica foi aperfeiçoada por Baksalary, Corsten & Kala (1978), Kenward (1985), Calinski & Caussinus (1989) e revista no texto de

Kshirsagar & Smith (1995). Soler & Singer (2000) propuseram um procedimento ótimo para a escolha das covariáveis (colunas de \mathbf{Y}_2).

Soluções explícitas para as equações de estimação também podem ser obtidas sob o modelo (2.4.1) quando não há omissão e a matriz de covariâncias intraunidades amostrais tem uma estrutura uniforme do tipo

$$\mathbf{R} = \nu^2 \mathbf{1}_m \mathbf{1}_m^\top + \tau^2 \mathbf{I}_m$$

em que ν^2 e τ^2 são os parâmetros de covariância. Sob essas condições, o estimador de MVR (e de MV) do vetor de parâmetros de localização é o mesmo que aquele obtido quando essa matriz é não-estruturada. Pode-se mostrar que os estimadores de MVR de τ^2 e ν^2 são respectivamente

$$\hat{\tau}^2 = \frac{1}{n(m-1)} \text{tr}\{\mathbf{Y}[\mathbf{I}_m - m^{-1} \mathbf{1}_m \mathbf{1}_m^\top] \mathbf{Y}^\top\}$$

e

$$\hat{\nu}^2 = \frac{1}{m} \left\{ \frac{1}{n-p} \mathbf{1}_m^\top \mathbf{Y}^\top [\mathbf{I}_n - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top] \mathbf{Y} \mathbf{1}_m - \hat{\tau}^2 \right\}.$$

Os estimadores de MV correspondentes são obtidos com a substituição de $n-p$ por n na expressão de $\hat{\nu}^2$.

Como a matriz de covariâncias intraunidades amostrais com estrutura uniforme pode ser induzida por modelos mistos com um efeito aleatório como aqueles utilizados para análise de estudos com planejamento do tipo *split-plot*, podemos empregar técnicas usuais de ANOVA para a análise desse tipo de problemas.

Trabalhando independentemente, Rouanet & Lépine (1970) e Huynh & Feldt (1970) mostraram que os testes F para hipóteses do tipo $\mathbf{CMU} = \mathbf{0}$, em que \mathbf{U} é uma matriz de dimensão $(m \times u)$ resultantes da análise por meio desse enfoque permanecem válidos se a matriz de covariâncias intraunidades amostrais satisfizer a condição menos restritiva de **esfericidade** ou **circularidade** em relação à \mathbf{U} , *i.e.*, se

$$\mathbf{P}^\top \mathbf{R} \mathbf{P} = \lambda \mathbf{I}_u, \quad (2.4.4)$$

em que $\mathbf{P} = \mathbf{U}(\mathbf{U}^\top \mathbf{U})^{-1/2}$ e λ é uma constante. Sob essa condição, um certo nível de heterocedasticidade pode ser acomodado. Sob homocedasticidade, as condições de uniformidade e esfericidade são equivalentes.

Por exemplo, a matriz

$$\mathbf{R} = \begin{bmatrix} 2.0 & 1.2 & 2.0 & 1.0 \\ 1.2 & 2.5 & 1.5 & 1.0 \\ 2.0 & 1.5 & 3.5 & 1.8 \\ 1.0 & 1.0 & 1.8 & 3.0 \end{bmatrix}$$

claramente não tem estrutura uniforme, mas é esférica com relação à matriz

$$\mathbf{U}_1^\top = \begin{bmatrix} 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \end{bmatrix}.$$

Consequentemente, as estatísticas de teste para hipóteses do tipo $\mathbf{CMU} = \mathbf{0}$, em que a matriz $\mathbf{U} = \mathbf{U}_1$, têm distribuição F exata sob a hipótese nula. Isto não ocorre quando

$$\mathbf{U}^\top = \mathbf{U}_2^\top = \begin{bmatrix} 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \end{bmatrix},$$

pois a matriz \mathbf{R} não é esférica relativamente à matriz \mathbf{U}_2 . Nesse contexto, Singer & Andrade (1994) sugeriram uma estratégia para a obtenção de testes F para sub-hipóteses relativas a interações e interações parciais como aquelas consideradas em Boik (1979). Um teste para esfericidade foi proposto por Mauchly (1940).

Com base no trabalho de Box (1954a) e Box (1954b), Geisser & Greenhouse (1958) propuseram testes F aproximados para casos em que \mathbf{R} não satisfaz a condição de esfericidade. Sua sugestão envolve a multiplicação dos graus de liberdade das estatísticas F geradas pela ANOVA por um fator de correção dado por

$$\varepsilon = \frac{[tr(\mathbf{P}^\top \mathbf{R} \mathbf{P})]^2}{(m-1)tr(\mathbf{P}^\top \mathbf{R} \mathbf{P})^2}.$$

Quando \mathbf{R} é esférica, temos $\varepsilon = 1$ e o valor mínimo para ε é $(m-1)^{-1}$ como indicado em Box (1954a) e Box (1954b). O teste sem correção no número de graus de liberdade ($\varepsilon = 1$) tende a ser liberal (rejeitando mais hipóteses do que se deveria) e aquele baseado na correção com $\varepsilon = (m-1)^{-1}$ tende a ser conservador. Fundamentados nesses resultados, Greenhouse & Geisser (1959) propuseram o seguinte procedimento em três estágios para testar as hipóteses de interesse:

- i) Testar a hipótese com o teste F sem correção; se o resultado for não-significativo, não rejeitar a hipótese nula e em caso contrário, passar para o segundo estágio.
- ii) Considerar o teste F com graus de liberdade multiplicados por $\varepsilon = (m-1)^{-1}$; se o resultados for significativo, rejeitar a hipótese nula e em caso contrário, passar para o terceiro estágio.
- iii) Utilizar o teste F com os graus de liberdade multiplicados por uma estimativa de ε .

O estimador de máxima verossimilhança de ε é

$$\hat{\varepsilon} = \frac{[tr(\mathbf{P}^\top \mathbf{S} \mathbf{P})]^2}{(m-1)tr(\mathbf{P}^\top \mathbf{S} \mathbf{P})^2}$$

em que \mathbf{S} é a matriz de covariâncias amostral. Por intermédio de estudos de simulação, Collier, Baker, Mandeville & Hayes (1967) e Stoloff (1970) concluíram que $\widehat{\varepsilon}$ é negativamente enviesado para pequenos valores de ε , especialmente quando n é pequeno. Huynh & Feldt (1976) propuseram o estimador

$$\bar{\varepsilon} = [n(m-1)\widehat{\varepsilon} - 2] \{ / (m-1)[n-p-(m-1)\widehat{\varepsilon}] \}$$

e mostraram que ele é menos enviesado que $\widehat{\varepsilon}$ quando $\varepsilon \leq 0.75$. Nos casos em que este estimador tem valores maiores que 1, deve-se utilizar $\bar{\varepsilon} = 1$.

Soluções explícitas para as equações de estimação também podem ser obtidas em outros casos particulares. Tendo em vista a expressão (2.3.3) é possível obter um estimador de mínimos quadrados generalizados em dois estágios para o parâmetro de localização $\boldsymbol{\beta}$ se tivermos disponível estimadores consistentes de $\boldsymbol{\Omega}_i(\boldsymbol{\theta})$. Um caso especial de interesse é aquele para o qual podemos adotar um modelo misto homocedástico com independência condicional (2.1.2) em que $\mathbf{Z}_i = \mathbf{X}_i$ e $\mathbf{R}_i = \tau^2 \mathbf{I}_{m_i}$. Esse modelo, também cognominado **modelo de coeficientes aleatórios** é apropriado para casos em que os perfis individuais e populacionais têm a mesma forma. Escrevendo $\boldsymbol{\beta}_i = \boldsymbol{\beta} + \mathbf{b}_i$, e com base num argumento condicional, podemos ajustar curvas aos dados de cada unidade amostral, separadamente, por meio de mínimos quadrados ordinários, obtendo

$$\widehat{\boldsymbol{\beta}}_i = (\mathbf{X}_i^\top \mathbf{X}_i)^{-1} \mathbf{X}_i^\top \mathbf{y}_i$$

com $\mathbb{E}(\widehat{\boldsymbol{\beta}}_i) = \boldsymbol{\beta}$ e $\mathbb{V}(\widehat{\boldsymbol{\beta}}_i) = \mathbf{G} + \tau^2 (\mathbf{X}_i^\top \mathbf{X}_i)^{-1}$. Estimadores consistentes de τ^2 e $\mathbb{V}(\widehat{\boldsymbol{\beta}}_i)$ são dados por

$$\widehat{\tau}^2 = \left[\sum_{i=1}^n (m_i - p) \right]^{-1} \sum_{i=1}^n \mathbf{y}_i^\top [\mathbf{I}_{m_i} - \mathbf{X}_i (\mathbf{X}_i^\top \mathbf{X}_i)^{-1} \mathbf{X}_i^\top] \mathbf{y}_i$$

e

$$\widehat{\mathbb{V}}(\widehat{\boldsymbol{\beta}}_i) = (n-1)^{-1} \sum_{i=1}^n (\widehat{\boldsymbol{\beta}}_i - \widehat{\boldsymbol{\beta}})(\widehat{\boldsymbol{\beta}}_i - \widehat{\boldsymbol{\beta}})^\top$$

em que $\widehat{\boldsymbol{\beta}} = n^{-1} \sum_{i=1}^n \widehat{\boldsymbol{\beta}}_i$, de forma que os estimadores de \mathbf{G} e $\boldsymbol{\Omega}_i(\boldsymbol{\theta})$ podem ser obtidos a partir de

$$\widehat{\mathbf{G}} = \widehat{\mathbb{V}}(\widehat{\boldsymbol{\beta}}_i) - \widehat{\tau}^2 n^{-1} \sum_{i=1}^n (\mathbf{X}_i^\top \mathbf{X}_i)^{-1}$$

e

$$\boldsymbol{\Omega}_i(\widehat{\boldsymbol{\theta}}) = \mathbf{X}_i \widehat{\mathbf{G}} \mathbf{X}_i^\top + \widehat{\tau}^2 \mathbf{I}_{m_i}.$$

Consequentemente, inferências sobre o parâmetro $\boldsymbol{\beta}$ podem ser concretizadas por meio de $\widehat{\boldsymbol{\beta}} \approx N(\boldsymbol{\beta}, \mathbf{G} + \widehat{\tau}^2 n^{-1} \sum_{i=1}^n (\mathbf{X}_i^\top \mathbf{X}_i)^{-1})$.

Esse enfoque é adotado por Graybill (1976), Laird, Lange & Stram (1987) e Vonesh & Carter (1987), que também apresentam as propriedades assintóticas desses estimadores. Um caso particular está detalhado no Apêndice B.5. Uma generalização para casos em que as colunas de \mathbf{Z}_i correspondem a subconjuntos das colunas de \mathbf{X}_i está apresentada em Rocha & Singer (2017).

Métodos mais gerais do que aqueles descritos acima para resolução das equações de estimação são necessários para acomodar casos com observações omissas ou com dados coletados irregularmente ao longo do tempo. Essencialmente, três enfoques têm sido considerados com essa finalidade. O primeiro é baseado no **algoritmo de Newton-Raphson** ou em uma adaptação estatística denominada **algoritmo *scoring* de Fisher**; o segundo utiliza o **algoritmo EM** e o terceiro utiliza uma **representação do tipo espaço/estados** e o **filtro de Kalman**.

Dado um valor inicial $\boldsymbol{\theta}^{(0)}$ e denotando a **função escore** e a **matriz hessiana** associadas à função log-verossimilhança $l(\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}), \boldsymbol{\theta})$ respectivamente por $\mathbf{u}(\boldsymbol{\theta}) = \partial l[\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}), \boldsymbol{\theta}] / \partial \boldsymbol{\theta}$ e $\mathbf{H}(\boldsymbol{\theta}) = \partial^2 l[\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}), \boldsymbol{\theta}] / \partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top$, o algoritmo de Newton-Raphson para a solução das equações de estimação (2.3.4) corresponde a iterações de

$$\boldsymbol{\theta}^{(l)} = \boldsymbol{\theta}^{(l-1)} - \{\mathbf{H}[\boldsymbol{\theta}^{(l-1)}]\}^{-1} \mathbf{u}[\boldsymbol{\theta}^{(l-1)}] \quad (2.4.5)$$

para $l = 1, 2, \dots$ até que algum critério de convergência seja satisfeito, *e.g.*, $\|\boldsymbol{\theta}^{(l)} - \boldsymbol{\theta}^{(l-1)}\| < \varepsilon$ para algum valor pré-especificado $\varepsilon > 0$. Da última iteração de (2.4.5) obtemos a estimativa de MV de $\boldsymbol{\theta}$, nomeadamente $\hat{\boldsymbol{\theta}}$, que substituída em (2.3.3), gera a estimativa de MV de $\boldsymbol{\beta}$, a saber, $\hat{\boldsymbol{\beta}}$. Se utilizarmos (2.3.9) no lugar de (2.3.4) nas definições de $\mathbf{u}(\boldsymbol{\theta})$ e de $\mathbf{H}(\boldsymbol{\theta})$, os dois procedimentos geram estimativas de MVR.

Nos casos em que a matriz de informação empírica $-\mathbf{H}(\hat{\boldsymbol{\theta}})$ não é definida positiva em algum passo do processo iterativo, o algoritmo de Newton-Raphson pode ficar muito lento ou não convergir. Autores como Jennrich & Schluchter (1986) e Lindstrom & Bates (1988) apresentam diferentes propostas para acelerar o algoritmo. Outra solução consiste na utilização do algoritmo *scoring* de Fisher, que corresponde a (2.4.5) com $\mathbf{H}(\boldsymbol{\theta})$ substituída por sua esperança, que é sempre definida positiva. Esse algoritmo é quase tão rápido quanto o algoritmo de Newton-Raphson e é mais robusto com relação ao ponto inicial.

Uma estimativa consistente da matriz de covariâncias assintótica de $\hat{\boldsymbol{\theta}}$ pode ser obtida a partir da inversa da **matriz de informação empírica**, $-\mathbf{H}(\hat{\boldsymbol{\theta}})$ gerada no último passo dos algoritmos. Uma estimativa consistente da matriz de covariâncias assintótica de $\hat{\boldsymbol{\beta}}$ é $\{\sum_{i=1}^n \mathbf{X}_i^\top [\boldsymbol{\Omega}_i(\hat{\boldsymbol{\theta}})]^{-1} \mathbf{X}_i\}^{-1}$.

Alternativamente, o algoritmo EM proposto por Dempster, Laird & Rubin (1977) para estimação no contexto de dados incompletos (ou com observações omissas) pode ser utilizado para cálculo de estimativas de MV e MVR dos parâmetros de alguns dos modelos utilizados para análise de dados com medidas repetidas. A

base desse algoritmo é a função de verossimilhança dos dados “aumentados”, *i.e.*, $\mathbf{y}^* = (\mathbf{y}^\top, \boldsymbol{\nu}^\top)^\top$, em que \mathbf{y} denota o vetor de dados efetivamente observados e $\boldsymbol{\nu}$ representa quer as observações omissas num contexto de dados incompletos, quer os valores latentes (efeitos aleatórios) integrantes de modelos mistos. Denotando por \mathbf{s} o vetor de estatísticas suficientes para os parâmetros $\boldsymbol{\beta}$ e $\boldsymbol{\theta}$, o l -ésimo passo do algoritmo pode ser descrito como:

Passo E: Cálculo da esperança condicional da estatística suficiente dadas as observações e os valores atualizados dos parâmetros, *i.e.*,

$$\mathbf{s}^{(l)} = E[\mathbf{s}(\mathbf{y}^*) | \mathbf{y}, \boldsymbol{\beta}^{(l-1)}, \boldsymbol{\theta}^{(l-1)}]$$

Passo M: Solução da equação

$$\mathbf{s}^{(l)} = E[\mathbf{s}(\mathbf{y}^*) | \boldsymbol{\beta}, \boldsymbol{\theta}]$$

em termos de $\boldsymbol{\beta}$ e $\boldsymbol{\theta}$ e obtenção dos novos valores dos parâmetros, $\boldsymbol{\beta}^{(l)}$ e $\boldsymbol{\theta}^{(l)}$. A iteração desses dois passos até à convergência produz os estimadores de MV desejados. Detalhes sobre a extensão desse algoritmo para cálculo de estimativas de MVR podem ser obtidos em Jennrich & Schluchter (1986).

O algoritmo EM é conveniente para casos em que há soluções explícitas para as equações do Passo M. Em caso contrário, outros métodos iterativos precisam ser considerados para resolver essas equações, o que gera um procedimento duplamente iterativo a ser evitado.

No caso particular do modelo misto homocedástico com independência condicional, *i.e.*, (2.1.2) com $\mathbf{R}_i = \sigma^2 \mathbf{I}_{m_i}$ em (2.1.4), em que as estatísticas suficientes para o vetor de parâmetros $\boldsymbol{\theta}$ são $s_1 = \sum_{i=1}^n \mathbf{e}_i^\top \mathbf{e}_i$ e os $q(q+1)/2$ elementos não-redundantes de $\mathbf{S}_2 = \sum_{i=1}^n \mathbf{b}_i \mathbf{b}_i^\top$, Laird & Ware (1982) mostram que o Passo E se reduz a

$$\begin{aligned} s_1^{(l)} &= E\left\{\sum_{i=1}^n \mathbf{e}_i^\top \mathbf{e}_i | \mathbf{y}_i, \boldsymbol{\beta}(\boldsymbol{\theta}^{(l-1)}), \boldsymbol{\theta}^{(l-1)}\right\} \\ &= \sum_{i=1}^n [\mathbf{e}_i^\top(\boldsymbol{\theta}^{(l-1)}) \mathbf{e}_i(\boldsymbol{\theta}^{(l-1)}) + \text{tr} \mathbb{V}\{\mathbf{e}_i | \mathbf{y}_i, \boldsymbol{\beta}(\boldsymbol{\theta}^{(l-1)}), \boldsymbol{\theta}^{(l-1)}\}] \end{aligned} \quad (2.4.6)$$

e

$$\begin{aligned} \mathbf{S}_2^{(l)} &= \mathbb{E}\left\{\sum_{i=1}^n \mathbf{b}_i \mathbf{b}_i^\top | \mathbf{y}_i, \boldsymbol{\beta}(\boldsymbol{\theta}^{(l-1)}), \boldsymbol{\theta}^{(l-1)}\right\} \\ &= \sum_{i=1}^n [\mathbf{b}_i(\boldsymbol{\theta}^{(l-1)}) \mathbf{b}_i^\top(\boldsymbol{\theta}^{(l-1)}) + \mathbb{V}\{\mathbf{b}_i | \mathbf{y}_i, \boldsymbol{\beta}(\boldsymbol{\theta}^{(l-1)}), \boldsymbol{\theta}^{(l-1)}\}] \end{aligned} \quad (2.4.7)$$

em que

$$\mathbf{e}_i(\boldsymbol{\theta}^{(l-1)}) = \mathbb{E}[\mathbf{e}_i | \mathbf{y}_i, \boldsymbol{\beta}(\boldsymbol{\theta}^{(l-1)}), \boldsymbol{\theta}^{(l-1)}] = \mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}(\boldsymbol{\theta}^{(l-1)}) - \mathbf{Z}_i \mathbf{b}_i(\boldsymbol{\theta}^{(l-1)})$$

e

$$\mathbf{b}_i(\boldsymbol{\theta}^{(l-1)}) = \mathbf{G}(\boldsymbol{\theta}^{(l-1)}) \mathbf{Z}_i^\top [\mathbf{V}_i(\boldsymbol{\theta}^{(l-1)})]^{-1} [\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}(\boldsymbol{\theta}^{(l-1)})].$$

O Passo M é dado por

$$\sigma^{2(l)} = \sum_{i=1}^n \mathbf{e}_i^\top(\boldsymbol{\theta}^{(l-1)}) \mathbf{e}_i(\boldsymbol{\theta}^{(l-1)}) / \sum_{i=1}^n m_i = s_1^{(l)} / \sum_{i=1}^n m_i \quad (2.4.8)$$

e

$$\mathbf{G}^{(l)} = n^{-1} \sum_{i=1}^n \mathbf{b}_i(\boldsymbol{\theta}^{(l-1)}) \mathbf{b}_i^\top(\boldsymbol{\theta}^{(l-1)}) = n^{-1} \mathbf{S}_2^{(l)}. \quad (2.4.9)$$

Para obter a estimativa de MV de $\boldsymbol{\theta}$, iniciamos o algoritmo como um valor conveniente $\boldsymbol{\theta}^{(0)}$ e iteramos (2.4.6) – (2.4.9) até que o critério de convergência adotado seja satisfeito. Detalhes da extensão desse procedimento para a obtenção das estimativas de MVR podem ser obtidos em Laird & Ware (1982), por exemplo. Andrade & Helms (1984) mostram como o algoritmo EM pode ser empregado para obter estimativas de MV nos casos em que a matriz de covariâncias intraunidades amostrais satisfaz a estrutura linear (2.2.12).

A particularização dos algoritmos de Newton-Raphson, *scoring* de Fisher e EM para o cálculo de estimativas de parâmetros de modelos para análise de dados com medidas repetidas e mais especificamente para análise de dados longitudinais é considerada por muitos autores, dentre os quais destacamos Jennrich & Schluchter (1986) que delineiam um procedimento geral para acomodar diferentes estruturas de covariância e Laird et al. (1987) ou Lindstrom & Bates (1988), que avaliam a velocidade de convergência dos algoritmos. Uma desvantagem do algoritmo EM é que ele não produz estimativas dos erros padrões das estimativas de MV; essas estimativas precisam ser obtidas por meio de uma iteração do algoritmo de Newton-Raphson tendo como valor inicial a estimativa obtida por meio do algoritmo EM. Esse processo híbrido requer mais iterações que o algoritmo EM, mas sempre converge para um máximo local conforme mencionam Lindstrom & Bates (1988).

A terceira alternativa para obtenção das estimativas de MV baseia-se numa representação do tipo espaço/estados para o problema de dados longitudinais e no uso de um filtro de Kalman para a maximização da função de verossimilhança. Esse enfoque é útil para a análise de modelos lineares dinâmicos e modelos de efeitos aleatórios, especialmente quando a matriz de covariâncias intraunidades amostrais tem uma estrutura de correlação serial. O leitor poderá consultar Jones (1993) para detalhes.

Métodos bayesianos para a análise de dados longitudinais também têm sido considerados por muitos autores, dentre os quais destacamos Lee & Geisser (1972), Geisser (1970), Geisser (1980), Rao (1987), Lee (1988), Kass & Steffey (1989) e Lesaffre & Lawson (2012).

2.5 Diagnóstico

Técnicas de diagnóstico são importantes na especificação de modelos estatísticos, especialmente nos casos em que a base teórica para explicar a interrelação entre as variáveis envolvidas não está bem estabelecida. Elas são úteis para identificar características inadequadas dos modelos propostos e indicar possíveis modificações.

Análise de resíduos e análise de sensibilidade são ferramentas importantes para avaliar o ajuste de qualquer modelo estatístico a um determinado conjunto de dados, para verificar a validade das suposições adotadas e, conseqüentemente, para garantir a confiabilidade das inferências nele baseadas. Em modelos lineares padrão, resíduos são utilizados para avaliar as suposições de linearidade dos efeitos, de normalidade, de independência, de homocedasticidade dos erros e também para identificar a presença de observações discrepantes (*outliers*) ou de observações influentes. A análise de sensibilidade consiste de um conjunto de técnicas úteis para avaliar alterações no modelo ajustado quando perturbações são introduzidas nos dados ou em suas suposições. Essas técnicas incluem análise de pontos-alavanca (*leverage analysis*, que é baseada no efeito das variáveis explicativas medido por meio da **matriz chapéu** (*hat matrix*), análise de omissão de casos (*case deletion analysis*) ou análise de influência local (*local influence analysis*), em que se avalia o efeito de uma perturbação infinitesimal dos dados ou de algum componente do modelo.

Existem muitas alternativas para avaliar o ajuste de modelos lineares padrão como indicado em Hoaglin & Welsch (1978), Belsley, Kuh & Welsch (1980), Cook & Weisberg (1982), Cook (1986), Chatterjee & Hadi (1988), Wei, Hu & Fung (1998) ou Atkinson & Riani (2000) entre outros. Generalizações para modelos lineares mistos são mais recentes. Por exemplo, Beckman, Nachtsheim & Cook (1987), Christensen, Pearson & Johnson (1992), Banerjee & Frees (1997), Lesaffre & Verbeke (1998), Tan, Ouwen & Berger (2001), Fung, Zhu, Wei & He (2002), Demidenko (2013), Demidenko & Stukel (2005), Zewotir & Galpin (2005), Zewotir (2008), Nobre & Singer (2007), Nobre & Singer (2011), Gumedze, Welham, Gogel & Thompson (2010) ou Schützenmeister & Piepho (2012) apresentam uma série de métodos de diagnóstico que descrevemos e atualizamos nas seções seguintes.

2.5.1 Análise de resíduos

As ferramentas de diagnóstico para modelos lineares mistos são mais complexas do que no caso usual porque as fontes de variação associadas (\mathbf{e} and \mathbf{b}) geram os seguintes três tipos de resíduos:

- i) **Resíduos marginais**, $\hat{\boldsymbol{\xi}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}$, preditores dos erros marginais $\boldsymbol{\xi} = \mathbf{y} - \mathbb{E}[\mathbf{y}] = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}$.
- ii) **Resíduos condicionais**, $\hat{\mathbf{e}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{Z}\hat{\mathbf{b}}$, preditores dos erros condicionais $\mathbf{e} = \mathbf{y} - \mathbb{E}[\mathbf{y}|\mathbf{b}] = \mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{b}$.
- iii) **Resíduos de efeitos aleatórios**, $\mathbf{Z}\hat{\mathbf{b}}$, preditores dos efeitos aleatórios, $\mathbf{Z}\mathbf{b} = \mathbb{E}[\mathbf{y}|\mathbf{b}] - \mathbb{E}[\mathbf{y}]$.

Uma complicação adicional está relacionada com o conceito de confundimento. Como indicado em Hilden-Minton (1995), um resíduo é considerado confundido para um determinado tipo de erro se ele depende de erros diferentes daquele do qual ele é supostamente preditor. Por exemplo, dado que $\hat{\mathbf{e}} = \mathbf{RQe} + \mathbf{RQZb}$, resíduos condicionais e de efeitos aleatórios podem estar confundidos. Uma exceção ocorre quando as colunas de \mathbf{Z} pertencem ao espaço gerado pelas colunas de \mathbf{X} . Isto implica que, por exemplo, $\hat{\mathbf{e}}$ pode não ser adequado para avaliar a suposição de normalidade de \mathbf{e} porque quando \mathbf{b} não satisfaz a suposição de normalidade, $\hat{\mathbf{e}}$ pode não apresentar um comportamento gaussiano mesmo quando \mathbf{e} segue uma distribuição gaussiana.

Resíduos marginais

Lesaffre & Verbeke (1998) indicam que quando a estrutura de covariância intraunidades amostrais é adequada, $\mathcal{V}_i = \|\mathbf{I}_{m_i} - \boldsymbol{\mathcal{E}}_i \boldsymbol{\mathcal{E}}_i^\top\|^2$, em que $\boldsymbol{\mathcal{E}}_i = \hat{\boldsymbol{\Omega}}_i^{-1/2} \hat{\boldsymbol{\xi}}_i$ com $\boldsymbol{\Omega}_i = \boldsymbol{\Omega}_i(\boldsymbol{\theta})$ deveria ter valores próximos de zero. Unidades amostrais com valores “grandes” de \mathcal{V}_i são aquelas para as quais a estrutura de covariância proposta pode não ser adequada. Como a matriz de covariâncias de $\hat{\boldsymbol{\xi}}_i$ é $\mathbb{V}(\hat{\boldsymbol{\xi}}_i) = \boldsymbol{\Omega}_i - \mathbf{X}_i(\mathbf{X}_i^\top \boldsymbol{\Omega}_i^{-1} \mathbf{X}_i)^{-1} \mathbf{X}_i^\top$ e não $\boldsymbol{\Omega}_i$, consideramos a substituição de $\boldsymbol{\mathcal{E}}_i$ na expressão de \mathcal{V}_i pelos resíduos marginais padronizados $\hat{\boldsymbol{\xi}}_i^* = [\hat{\mathbb{V}}(\hat{\boldsymbol{\xi}}_i)]^{-1/2} \hat{\boldsymbol{\xi}}_i$ em que $\mathbb{V}(\hat{\boldsymbol{\xi}}_i)$ corresponde ao bloco diagonal de $\boldsymbol{\Omega} - \mathbf{X}(\mathbf{X}^\top \boldsymbol{\Omega}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top$ associado à i -ésima unidade amostral.

Além disso, para evitar a atribuição de um peso excessivo às unidades amostrais com muitas observações, consideramos a utilização de $\mathcal{V}_i^* = \sqrt{\mathcal{V}_i}/m_i$ como uma medida padronizada da adequação da estrutura de covariância intraunidades amostrais. Gráficos de \mathcal{V}_i^* versus os índices i , podem ser úteis na identificação de unidades para as quais a estrutura de covariância deveria ser modificada.

Para avaliar a linearidade dos efeitos fixos no modelo (2.1.2), consideramos a construção de um gráfico dos componentes dos resíduos marginais padronizados $\hat{\xi}_{ij}^* = \hat{\xi}_{ij} / [\text{diag}_j(\hat{\mathbb{V}}(\hat{\xi}_i))]^{1/2}$, em que $\text{diag}_j(\mathbb{V}(\hat{\xi}_i))$ denota o j -ésimo elemento da diagonal principal de $\mathbb{V}(\hat{\xi}_i)$, *versus* os valores de cada variável explicativa bem como *versus* os valores ajustados. Também recomendamos a construção de um gráfico de $\hat{\xi}_{ij}^*$ *versus* os índices das observações intraunidades amostrais como uma ferramenta para a detecção de observações discrepantes.

Resíduos condicionais

Dado que $\mathbb{V}(\hat{\mathbf{e}}) = \mathbf{R}[\mathbf{\Omega}^{-1} - \mathbf{\Omega}^{-1}\mathbf{X}(\mathbf{X}^\top\mathbf{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{\Omega}^{-1}]\mathbf{R} = \mathbf{RQR}$, Nobre & Singer (2007) observam que os resíduos condicionais podem ser heterocedásticos e sugerem a construção de gráficos dos resíduos condicionais padronizados, $\hat{e}_{ij}^* = \hat{e}_{ij} / \text{diag}_{ij}(\mathbf{RQR})^{1/2}$, em que $\text{diag}_{ij}(\mathbf{RQR})$ indica o elemento da diagonal principal de \mathbf{RQR} correspondente à j -ésima observação da i -ésima unidade *versus* valores ajustados para avaliar a homocedasticidade dos erros condicionais (nos casos de modelos com homocedasticidade condicional) ou *versus* índices de observações para detecção de observações discrepantes.

Como sugere Hilden-Minton (1995), a habilidade para avaliação da normalidade dos erros condicionais aumenta com a minimização da fração de confundimento do k -ésimo resíduo condicional, nomeadamente

$$0 \leq \frac{\mathbf{u}_k^\top \mathbf{RQZTZ}^\top \mathbf{QRu}_k}{\mathbf{u}_k^\top \mathbf{RQu}_k} = 1 - \frac{\mathbf{u}_k^\top \mathbf{RQRQRu}_k}{\mathbf{u}_k^\top \mathbf{RQu}_k} \leq 1$$

em que \mathbf{u}_k é a k -ésima coluna de \mathbf{I}_N . Com base nesse resultado, ele propõe a utilização de **resíduos condicionais minimamente confundidos** (*least confounded conditional residuals*), *i.e.*, de uma transformação linear dos resíduos condicionais que minimiza a fração de confundimento.

Os resíduos condicionais minimamente confundidos são dados por

$$\mathbf{c}_k^\top \hat{\mathbf{e}} = \lambda_k^{-1/2} \boldsymbol{\ell}_k^\top \mathbf{R}^{-1/2} \hat{\mathbf{e}} = \lambda_k^{1/2} \boldsymbol{\ell}_k^\top \mathbf{R}^{-1/2} \mathbf{y}, \quad k = 1, \dots, N-p$$

em que $1 \geq \lambda_1 \geq \dots \geq \lambda_{N-p} > 0$ são os valores ordenados de $\mathbf{\Lambda}$, obtidos da decomposição de valores singulares de

$$\mathbf{R}^{1/2} \mathbf{QR}^{1/2} = \mathbf{LAL}^\top, \quad \mathbf{L}^\top \mathbf{L} = \mathbf{I}_{N-p},$$

$\mathbf{c}_k = \lambda_k^{-1/2} \mathbf{R}^{-1/2} \boldsymbol{\ell}_k$ e $\boldsymbol{\ell}_k$ representa a k -ésima coluna de \mathbf{L} .

Resíduos condicionais minimamente confundidos padronizados, $\mathbf{c}_k^\top \hat{\mathbf{e}}^*$ podem ser obtidos por meio do quociente entre $\mathbf{c}_k^\top \hat{\mathbf{e}}$ e a raiz quadrada do elemento correspondente de $\mathbf{C} \hat{\mathbf{R}} \mathbf{Q} \hat{\mathbf{R}} \mathbf{C}^\top$ em que $\mathbf{C} = [\mathbf{c}_1, \dots, \mathbf{c}_{N-p}]^\top$. Gráficos QQ dos resíduos condicionais minimamente confundidos padronizados, $\mathbf{c}_k^\top \hat{\mathbf{e}}^*$, (ou dos próprios resíduos condicionais padronizados quando não há confundimento) podem ser utilizados para avaliação da suposição de normalidade.

Recentemente Schützenmeister & Piepho (2012) comentam que transformações lineares de resíduos podem não ser tão úteis como ferramentas para diagnóstico, pois primeiramente, os resíduos transformados não estão associados a observações específicas e em segundo lugar, eles podem amplificar o efeito conhecido como supranormalidade, *i.e.*, podem parecer mais normais do que os efeitos subjacentes realmente o são. Como alternativa, esses autores propõem analisar o modelo linear misto como se fosse um modelo linear padrão e considerar a utilização dos correspondentes resíduos padronizados como ferramentas para diagnóstico. Eles também propõem um método baseado em simulação para construir intervalos de tolerância para os gráficos QQ correspondentes. Essa proposta, no entanto, é computacionalmente intensiva e, em muitos casos, pode não ser razoável como ferramenta exploratória, especialmente quando muitos modelos precisam ser avaliados.

Resíduos de efeitos aleatórios

Quando não há confundimento e os efeitos aleatórios seguem uma distribuição gaussiana q -dimensional, $\mathcal{M}_i = \hat{\mathbf{b}}_i^\top \{\hat{\mathbf{V}}[\hat{\mathbf{b}}_i - \mathbf{b}_i]\}^{-1} \hat{\mathbf{b}}_i$ (a distância de Mahalanobis entre $\hat{\mathbf{b}}_i$ e $\mathbb{E}(\mathbf{b}_i) = \mathbf{0}$) deveria obedecer a uma distribuição qui-quadrado com q graus de liberdade. Consequentemente, um gráfico QQ baseado na distribuição χ_q^2 para \mathcal{M}_i pode ser utilizado para verificar se os efeitos aleatórios seguem uma distribuição gaussiana q -variada.

Gráficos de \mathcal{M}_i em função dos índices de unidades amostrais também podem ser empregados para identificar a presença daquelas que podem ser consideradas discrepantes.

As diferentes utilidades diagnósticas dos três tipos de resíduos obtidos do ajuste de modelos lineares mistos estão resumidas na Tabela (2.5.1) adaptada de Nobre & Singer (2007).

Tabela 2.5.1: Resíduos e sua utilização para diagnóstico

Diagnóstico para	Tipo de resíduo	Gráfico
Linearidade dos efeitos fixos ($\mathbb{E}[\mathbf{y}] = \mathbf{X}\boldsymbol{\beta}$)	Marginal	$\hat{\boldsymbol{\xi}}_{ij}^*$ vs valores ajustados ou variáveis explicativas
Presença de observações discrepantes (\mathbf{y}_{ij})	Marginal	$\hat{\boldsymbol{\xi}}_{ij}^*$ vs índices de observações
Matriz de covariância intraunidades ($\boldsymbol{\Omega}_i$)	Marginal	\mathcal{V}_i^* vs índices de unidades
Presença de observações discrepantes (\mathbf{y}_{ij})	Condicional	\hat{e}_{ij}^* vs índices de observações
Homocedasticidade dos erros condicionais (\mathbf{e}_{ij})	Condicional	\hat{e}_{ij}^* vs valores ajustados
Normalidade dos erros condicionais (\mathbf{e}_{ij})	Condicional	QQ gaussiano para $\mathbf{c}_k^\top \hat{\mathbf{e}}^*$
Presença de unidades discrepantes (\mathbf{b}_i)	Efeitos aleatórios	\mathcal{M}_i^* vs índices de unidades
Normalidade dos efeitos aleatórios (\mathbf{b}_i)	Efeitos aleatórios	QQ χ_q^2 para \mathcal{M}_i

2.5.2 Análise de influência global

Análise de pontos alavanca

No modelo linear clássico, $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$, o **poder de alavanca** (*leverage*) da i -ésima observação é definido como $h_{ii} = \partial \hat{y}_i / \partial y_i$ com \hat{y}_i representando o i -ésimo valor ajustado, (*i.e.*, o i -ésimo elemento de $\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$) e corresponde ao i -ésimo elemento da diagonal principal da matriz de projeção (chapéu) $\mathbf{H} = \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top$. Observações discrepantes no espaço vetorial gerado pelas colunas de \mathbf{X} têm **alto poder de alavanca** (*high leverage*) com respeito aos valores ajustados \hat{y}_i , como discutido em Cook & Weisberg (1982), Chatterjee & Hadi (1988) ou Wei et al. (1998), por exemplo.

Para o modelo linear misto (2.1.7), seguimos as ideias de Fung et al. (2002) e definimos a **matriz de alavanca marginal generalizada** (*generalized marginal leverage matrix*) como

$$\begin{aligned} \mathbf{L}_1 &= \frac{\partial \hat{\mathbf{y}}}{\partial \mathbf{y}^\top} = \frac{\partial \mathbf{X}(\mathbf{X}^\top \boldsymbol{\Omega}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \boldsymbol{\Omega}^{-1} \mathbf{y}}{\partial \mathbf{y}^\top} \\ &= \mathbf{X}(\mathbf{X}^\top \boldsymbol{\Omega}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \boldsymbol{\Omega}^{-1}. \end{aligned} \quad (2.5.1)$$

Denotando por $\mathbf{L}_{1i} = \mathbf{X}_i(\mathbf{X}^\top \boldsymbol{\Omega}^{-1} \mathbf{X})^{-1} \mathbf{X}_i^\top \boldsymbol{\Omega}_i^{-1}$ o i -ésimo bloco da diagonal principal de \mathbf{L}_1 e por $\mathbf{L}_{1i(jj)}$, o j -ésimo elemento da diagonal principal de \mathbf{L}_{1i} , temos

$$\text{tr}(\mathbf{L}_1) = \sum_{i=1}^n \sum_{j=1}^{m_i} \mathbf{L}_{1i(jj)} = p.$$

Então podemos dizer que a unidade amostral i tem alto poder de alavanca com respeito aos valores marginais ajustados, $\hat{\mathbf{y}}$, se $\text{tr}(\mathbf{L}_{1i})/m_i$ for maior que algum valor arbitrário, digamos $2p/n$ e, analogamente, que a observação j associada à unidade amostral i tem alto poder de alavanca com respeito aos valores marginais ajustados se $\mathbf{L}_{1i(jj)} \geq 2p/n$, embora seja mais comum utilizar critérios subjetivos (baseados em inspeção visual) para esses propósitos.

Dado que no modelo (2.1.7) uma unidade amostral ou uma observação (intraunidade amostral) pode afetar tanto os valores marginais quanto os condicionais ajustados, é razoável avaliar a influência conjunta de cada unidade (ou observação) em ambos. Consequentemente, podemos considerar $\hat{\mathbf{y}}^* = \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{Z}\hat{\mathbf{b}}$ em vez de $\hat{\mathbf{y}}$ para definir a **matriz de alavanca conjunta generalizada** (*generalized joint leverage matrix*) como

$$\mathbf{L} = \frac{\partial \hat{\mathbf{y}}^*}{\partial \mathbf{y}^\top} = \frac{\partial \hat{\mathbf{y}}}{\partial \mathbf{y}^\top} + \frac{\partial \mathbf{Z}\hat{\mathbf{b}}}{\partial \mathbf{y}^\top} = \mathbf{L}_1 + \mathbf{Z}\boldsymbol{\Gamma}\mathbf{Z}^\top \mathbf{Q} = \mathbf{L}_1 + \mathbf{L}_2 \mathbf{Q}, \quad (2.5.2)$$

em que \mathbf{L}_1 é a matriz de alavanca marginal generalizada (2.5.1) e

$$\mathbf{L}_2 = \mathbf{Z}\boldsymbol{\Gamma}\mathbf{Z}^\top. \quad (2.5.3)$$

A matriz \mathbf{L}_2 representa a porção da variabilidade intraunidades amostrais explicada pela presença dos efeitos aleatórios. Observações discrepantes no espaço vetorial gerado pelas colunas de \mathbf{Z} devem afetar o componente aleatório dos valores condicionais ajustados e, consequentemente, a estimativa da variância intraunidade explicada pelos efeitos aleatórios como preconizado por Tan et al. (2001). Então, estimativas dos componentes de variância do modelo devem ser afetados por observações como alto poder de alavanca com respeito ao componente aleatório dos valores condicionais ajustados. Demidenko & Stukel (2005) sugerem que

$$\mathbf{H}_2 = \mathbf{Z}\boldsymbol{\Gamma}\mathbf{Z}^\top \mathbf{Q} \quad (2.5.4)$$

poderia ser utilizada como uma matriz de alavanca generalizada para esse propósito. Nobre & Singer (2011), por outro lado, observam que

$$\mathbf{H}_2 = \mathbf{L}_2 \boldsymbol{\Omega}^{-1} [\mathbf{I}_n - \mathbf{L}_1]$$

também depende da matriz de alavanca marginal generalizada \mathbf{L}_1 e consideram que o poder de alavanca com respeito aos efeitos aleatórios associado aos valores

condicionais ajustados pode estar mascarado pelo poder de alavanca com respeito aos valores marginais ajustados. Então, à luz da expressão (2.5.2), sugerem que uma alternativa mais conveniente para avaliar o poder de alavanca das observações e unidades amostrais com respeito aos componentes aleatórios dos valores condicionais ajustados é a **matriz de alavanca de componentes aleatórios generalizada** (*generalized random component leverage matrix*) \mathbf{L}_2 em vez de \mathbf{H}_2 .

Paralelamente às considerações apresentadas no caso dos valores marginais ajustados, a matriz de alavanca de componentes aleatórios generalizada associada à i -ésima unidade é definida como $\mathbf{L}_{2i} = \mathbf{Z}_i \mathbf{G} \mathbf{Z}_i^\top$. A matriz de alavanca de componentes aleatórios generalizada associada à j -ésima observação correspondente à i -ésima unidade amostral é $\mathbf{L}_{2i(jj)}$, *i.e.*, o j -ésimo elemento da diagonal principal de \mathbf{L}_{2i} . Similarmente, a decisão sobre que unidades amostrais ou observações devem ser consideradas como tendo alto poder de alavanca é, usualmente, baseada em critérios subjetivos. Na prática, como as matrizes de covariâncias são desconhecidas, podemos substituí-las por estimativas.

Análise de omissão de casos

Uma ferramenta simples e facilmente interpretável para avaliar o impacto de uma observação ou de um conjunto de observações em alguma característica de interesse (a estimativa de um parâmetro ou o erro padrão correspondente, por exemplo) é baseada na **omissão de casos** (*case deletion*) inicialmente proposta por Cook (1977) e posteriormente estudada por diversos autores como Belsley et al. (1980), Cook & Weisberg (1982) ou Chatterjee & Hadi (1988). Em modelos lineares mistos como (2.1.7), é razoável considerar as medidas repetidas realizadas na mesma unidade amostral (caso) ou mesmo um conjunto de unidades amostrais como o conjunto a ser omitido. Nesse contexto, parte do esforço reside na obtenção de BLUEs ou de BLUPs sem a necessidade de reajustar o modelo quando cada unidade amostral ou conjunto delas é omitido.

Supondo que $\mathbf{\Gamma}$ and \mathbf{R} são conhecidas e que nosso interesse recai no exame do impacto do conjunto de unidades amostrais $I = \{i_1, i_2, \dots, i_k\}$ ($1 \leq i_1 \leq i_2 \leq \dots \leq i_k \leq n$), podemos considerar o modelo

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \mathbf{U}_I\boldsymbol{\phi}_{(I)} + \mathbf{e}, \quad (2.5.5)$$

em que $\boldsymbol{\phi}_{(I)}$ é um vetor k -dimensional de parâmetros (fixos) e $\mathbf{U}_I = [\mathbf{u}_{i_1}, \mathbf{u}_{i_2}, \dots, \mathbf{u}_{i_k}]$ com \mathbf{u}_i denotando a i -ésima coluna de \mathbf{I}_N . Hilden-Minton (1995) e Fung et al. (2002) mostram que o BLUE de $\boldsymbol{\beta}$ e o BLUP de \mathbf{b} no modelo (2.5.5) são idênticos àqueles obtidos sob o modelo (2.1.7) com as unidades contidas no conjunto I omitidas. Denotando esses estimadores por $\hat{\boldsymbol{\beta}}_{(I)}$ and $\hat{\mathbf{b}}_{(I)}$, respectivamente, eles mostram que

o BLUE de $\phi_{(I)}$ é

$$\hat{\phi}_{(I)} = (\mathbf{U}_I^\top \mathbf{Q} \mathbf{U}_I)^{-1} \mathbf{U}_I^\top \mathbf{Q} \mathbf{y} \quad (2.5.6)$$

e que

$$\hat{\beta}_{(I)} = \hat{\beta} - (\mathbf{X}^\top \mathbf{\Omega}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{\Omega}^{-1} \mathbf{U}_I \hat{\phi}_{(I)} \quad (2.5.7)$$

$$\mathbf{b}_{(I)} = \hat{\mathbf{b}} - \mathbf{\Gamma} \mathbf{Z}^\top \mathbf{Q} \mathbf{U}_I \hat{\phi}_{(I)}. \quad (2.5.8)$$

Além disso, temos

$$\mathbb{V}(\hat{\beta} - \hat{\beta}_{(I)}) = (\mathbf{X}^\top \mathbf{\Omega}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{\Omega}^{-1} \mathbf{U}_I (\mathbf{U}_I^\top \mathbf{Q} \mathbf{U}_I)^{-1} \mathbf{U}_I^\top \mathbf{\Omega}^{-1} \mathbf{X} (\mathbf{X}^\top \mathbf{\Omega}^{-1} \mathbf{X})^{-1} \quad (2.5.9)$$

$$\mathbb{V}(\hat{\mathbf{b}} - \hat{\mathbf{b}}_{(I)}) = \mathbf{\Gamma} \mathbf{Z}^\top \mathbf{Q} \mathbf{U}_I (\mathbf{U}_I^\top \mathbf{Q} \mathbf{U}_I)^{-1} \mathbf{U}_I \mathbf{Q} \mathbf{Z} \mathbf{\Gamma} \quad (2.5.10)$$

e

$$\mathbb{V}(\hat{\beta}_{(I)}) = \mathbb{V}(\hat{\beta}) - \mathbb{V}(\hat{\beta} - \hat{\beta}_{(I)}). \quad (2.5.11)$$

Com esses ingredientes, podemos considerar a seguinte generalização da distância de Cook

$$D_I = \frac{(\hat{\beta} - \hat{\beta}_{(I)})^\top (\mathbf{X}^\top \mathbf{\Omega}^{-1} \mathbf{X}) (\hat{\beta} - \hat{\beta}_{(I)})}{p} = \frac{(\hat{\mathbf{y}} - \hat{\mathbf{y}}_{(I)})^\top \mathbf{\Omega}^{-1} (\hat{\mathbf{y}} - \hat{\mathbf{y}}_{(I)})}{p}, \quad (2.5.12)$$

como uma medida da influência das unidades do conjunto I na estimativa dos parâmetros fixos β .

Tan et al. (2001), entre outros autores, discutem as limitações da distância de Cook (2.5.12) como uma medida de influência no contexto de modelos mistos. Especificamente, esses autores salientam que medidas construídas para a avaliação do efeito de unidades amostrais podem não ser convenientes para detecção daquelas discrepantes em função da posição relativa das observações influentes intra- e interunidades amostrais e propõem um enfoque condicional baseado em observações influentes.

Embora as propostas existentes na literatura adotem a suposição de que $\mathbf{R}_i = \sigma^2 \mathbf{I}_{m_i}$, é possível adaptá-las para o caso mais geral em que \mathbf{R}_i tem outras estruturas. Com essa finalidade, consideremos o modelo condicional

$$\mathbf{y} = \mathbf{X}^* \beta^* + \mathbf{e}$$

em que $\mathbf{X}^* = [\mathbf{X} \ \mathbf{Z}]$ e $\beta^* = (\beta^\top, \mathbf{b}^\top)^\top$. A **distância de Cook condicional** é definida como

$$D_{i(j)}^{cond} = \sum_{i=1}^n \frac{(\hat{\mathbf{y}}_i^* - \hat{\mathbf{y}}_{i(j)}^*)^\top \mathbb{V}[\mathbf{y}_i | \mathbf{b}_i]^{-1} (\hat{\mathbf{y}}_i^* - \hat{\mathbf{y}}_{i(j)}^*)}{(n-1)q + p} = \sum_{i=1}^n \frac{(\hat{\mathbf{y}}_i^* - \hat{\mathbf{y}}_{i(j)}^*)^\top \mathbf{R}_i^{-1} (\hat{\mathbf{y}}_i^* - \hat{\mathbf{y}}_{i(j)}^*)}{(n-1)q + p}. \quad (2.5.13)$$

em que $\hat{\mathbf{y}}_i^* = \mathbf{X}_i \hat{\boldsymbol{\beta}} + \mathbf{Z}_i \hat{\mathbf{b}}$, $\hat{\mathbf{y}}_{i(j)}^* = \mathbf{X}_i \hat{\boldsymbol{\beta}}_{(i(j))} + \mathbf{Z}_i \hat{\mathbf{b}}_{(i(j))}$ e $\hat{\boldsymbol{\beta}}_{(i(j))}$ e $\hat{\mathbf{b}}_{(i(j))}$ denotam, respectivamente, os BLUEs de $\boldsymbol{\beta}$ e \mathbf{b} obtidos após a eliminação da j -ésima observação da i -ésima unidade amostral.

Como sugerem Tan et al. (2001), (2.5.13) pode ser decomposta como $D_{i(j)}^{cond} = D_{1i(j)}^{cond} + D_{2i(j)}^{cond} + D_{3i(j)}^{cond}$ em que

$$D_{1i(j)}^{cond} = [(n-1)q + p]^{-1} (\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(i(j))})^\top \mathbf{X}^\top \mathbf{R}^{-1} \mathbf{X} (\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(i(j))})$$

é uma medida útil para a avaliação da influência da j -ésima observação da i -ésima unidade amostral na estimação de $\boldsymbol{\beta}$,

$$D_{2i(j)}^{cond} = [(n-1)q + p]^{-1} (\hat{\mathbf{b}} - \hat{\mathbf{b}}_{(i(j))})^\top \mathbf{Z}^\top \mathbf{R}^{-1} \mathbf{Z} (\hat{\mathbf{b}} - \hat{\mathbf{b}}_{(i(j))})$$

é apropriada para avaliar a influência da j -ésima observação da i -ésima unidade amostral na estimação de \mathbf{b} e

$$D_{3i(j)}^{cond} = [(n-1)q + p]^{-1} (\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(i(j))})^\top \mathbf{X}^\top \mathbf{R}^{-1} \mathbf{Z} (\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(i(j))})$$

está relacionada com mudanças na covariância entre as estimativas de $\boldsymbol{\beta}$ e \mathbf{b} quando j -ésima observação da i -ésima unidade amostral é eliminada da análise. Esses autores também sugerem que quando o interesse recai na avaliação da influência de uma unidade amostral nas estimativas dos parâmetros, é suficiente somar os termos (2.5.13) correspondentes àquela unidade.

Um enfoque diferente daquele que envolve medidas condicionais de omissão de casos está proposto em Fei & Pan (2003). Zewotir (2008), por outro lado, considera aspectos computacionais associados a essas ferramentas de diagnóstico.

O quociente dos elipsoides de variância,

$$\begin{aligned} \rho_{(I)} &= \frac{|\hat{\mathbf{W}}(\hat{\boldsymbol{\beta}}_{(I)})|}{|\hat{\mathbf{W}}(\hat{\boldsymbol{\beta}})|} \\ &= \frac{|(\mathbf{X}^\top \boldsymbol{\Omega}^{-1} \mathbf{X})^{-1} (\mathbf{I}_p + \mathbf{X}^\top \boldsymbol{\Omega}^{-1} \mathbf{U}_I (\mathbf{U}_I^\top \mathbf{Q} \mathbf{U}_I)^{-1} \mathbf{U}_I^\top \boldsymbol{\Omega}^{-1} \mathbf{X} (\mathbf{X}^\top \boldsymbol{\Omega}^{-1} \mathbf{X})^{-1})|}{|(\mathbf{X}^\top \boldsymbol{\Omega}^{-1} \mathbf{X})^{-1}|} \\ &= |\mathbf{I}_p + \mathbf{X}^\top \boldsymbol{\Omega}^{-1} \mathbf{U}_I (\mathbf{U}_I^\top \mathbf{Q} \mathbf{U}_I)^{-1} \mathbf{U}_I^\top \boldsymbol{\Omega}^{-1} \mathbf{X} (\mathbf{X}^\top \boldsymbol{\Omega}^{-1} \mathbf{X})^{-1}| \end{aligned} \quad (2.5.14)$$

é outra medida que pode ser utilizada para avaliar a influência de um conjunto I de unidades amostrais na matriz de covariâncias de $\hat{\boldsymbol{\beta}}$, como sugerido por Hilden-Minton (1995).

Na prática, técnicas de diagnóstico global podem ser implementadas por meio de gráficos das medidas apropriadas em função dos índices de unidades amostrais ou de observações. Um resumo está apresentado na Tabela 2.5.2.

Tabela 2.5.2: Gráficos de influência global para unidades ou observações

Diagnóstico para efeito sobre	Medida de influência	Gráfico de
Componente fixo de valor ajustado ($\mathbf{X}\hat{\boldsymbol{\beta}}$)	Matriz alavanca generalizada \mathbf{L}_1 (2.5.1)	$\mathbf{L}_{1i(jj)}$ $[tr(\mathbf{L}_{1i})/m_i]$ <i>vs</i> observações ou unidades
Componente aleatório de valor ajustado ($\mathbf{Z}\hat{\mathbf{b}}$)	Matriz alavanca para componentes aleatórios generalizada \mathbf{L}_2 (2.5.3)	$\mathbf{L}_{2i(jj)}$ $[tr(\mathbf{L}_{2i})/m_i]$ <i>vs</i> observações ou unidades
Coefficientes de regressão ($\hat{\boldsymbol{\beta}}$)	Distância de Cook D_I (2.5.12)	$D_{i(j)}$ $[D_i]$ <i>vs</i> observações ou unidades
Matriz de covariância de coeficientes de regressão $\hat{\mathbb{V}}(\hat{\boldsymbol{\beta}})$	Quociente de elipsoides de variância $\rho_{(I)}$ (2.5.10)	$\rho_{i(j)}$ $[\rho_i]$ <i>vs</i> observações ou unidades

Também podemos considerar a generalização decomposição da distância de Cook condicional (2.5.13) proposta originalmente para modelos de independência condicional por Tan et al. (2001) como resumido na Tabela 2.5.3. Recentemente, Mun

Tabela 2.5.3: Gráficos de influência global baseados na decomposição da distância de Cook condicional

Diagnóstico para efeito sobre	Gráfico de índices
Coefficientes de regressão ($\hat{\boldsymbol{\beta}}$)	$D_{1i(j)}^{cond}$ <i>vs</i> observações
Efeitos aleatórios ($\hat{\mathbf{b}}$)	$D_{2i(j)}^{cond}$ <i>vs</i> observações
Mudanças na covariância entre $\hat{\boldsymbol{\beta}}$ e $\hat{\mathbf{b}}$	$D_{3i(j)}^{cond}$ <i>vs</i> observações

& Lindstrom (2013) propõem um método baseado em soma de quadrados residuais studentizadas (TRSS) que não exige a eliminação de casos e é mais eficiente do que os métodos baseados na distância de Cook para a identificação de unidades amostrais ou observações discrepantes.

2.5.3 Análise de influência local

O conceito de influência local foi introduzido por Cook (1986) com o objetivo de avaliar as mudanças na análise que resultam de “pequenas” perturbações nos dados ou em algum elemento do modelo. Essencialmente, a ideia é investigar o comportamento do **deslocamento de verossimilhança** (*likelihood displacement*), definido como

$$LD(\omega) = 2 \left\{ L(\hat{\psi}) - L(\hat{\psi}_\omega | \omega) \right\}, \quad (2.5.15)$$

em que L denota a verossimilhança associada ao modelo proposto, ψ é um vetor de parâmetros p -dimensional, $\omega \in \Omega \subset \mathbb{R}^q$ é um vetor q -dimensional de “perturbações” e $\hat{\psi}$ e $\hat{\psi}_\omega$ são, respectivamente, os estimadores de máxima verossimilhança de ψ baseados em $L(\psi)$ e $L(\psi | \omega)$.

Supomos que existe $\omega_0 \in \Omega$ tal que $\psi_{\omega_0} = \psi$ e que $L(\psi | \omega)$ tem primeira e segunda derivadas contínuas numa vizinhança de ω_0 . Nesse contexto, podemos considerar a **curvatura normal** (*normal curvature*) do gráfico $G(\omega) = [\omega^\top, LD(\omega)]$ na direção de um vetor q -dimensional com norma unitária, \mathbf{d} no ponto ω_0 , *i.e.*,

$$C_{\mathbf{d}} = 2 \left| \mathbf{d}^\top \mathbf{H}^\top \ddot{\mathbf{F}}^{-1} \mathbf{H} \mathbf{d} \right|, \quad (2.5.16)$$

em que $\ddot{\mathbf{F}} = [\partial^2 L(\psi) / \partial \psi \partial \psi^\top]_{\psi=\hat{\psi}}$ e $\mathbf{H} = [\partial^2 L(\psi | \omega) / \partial \psi \partial \omega^\top]_{\omega=\omega_0; \psi=\hat{\psi}}$. Denotando o menor e o maior autovetor de $-\mathbf{H}^\top \ddot{\mathbf{F}}^{-1} \mathbf{H}$ respectivamente por C_{min} e C_{max} , é possível mostrar que $C_{min} \leq C_{\mathbf{d}} \leq C_{max}$. O autovetor normalizado \mathbf{d}_{max} correspondente a C_{max} pode ser utilizado para identificar a combinação linear dos elementos de ω mais influente na curvatura de $LD(\omega)$. Usualmente essa ferramenta é empregada para avaliar o efeito de perturbações na variável resposta, *i.e.*, $\mathbf{y}_i(\omega_i) = \mathbf{y}_i + \omega_i$, nas variáveis explicativas, *i.e.*, $\mathbf{X}_i(\mathbf{W}_i) = \mathbf{X}_i + \mathbf{W}_i$, em que $\mathbf{W}_i = [\omega_{i1}, \dots, \omega_{ip}]$ com $\omega_{ij} = (\omega_{ij1}, \dots, \omega_{ijm_i})^\top$, na matriz de covariâncias dos efeitos aleatórios, *i.e.*, $\mathbf{G}(\omega_i) = \omega_i \mathbf{G}$ ou na matriz de covariâncias dos erros, $\mathbf{R}_i(\omega_i) = \omega_i \mathbf{R}_i$. Em cada caso, gráficos dos valores absolutos dos elementos de \mathbf{d}_{max} (obtidos para determinados \mathbf{H} e $\ddot{\mathbf{F}}$) em função dos índices de unidades ou de observações podem ser utilizados para identificar aquelas que têm maior impacto no deslocamento de verossimilhança.

A escolha do esquema de perturbação apropriado não é simples, principalmente por que “perturbar arbitrariamente um modelo pode gerar inferências inadequadas sobre as causas (*e.g.*, observações influentes) de um efeito considerável” e “os componentes de um vetor de perturbações podem não ser ortogonais entre si” o que pode provocar dificuldades na interpretação, como indicado em Zhu, Ibrahim, Lee & Zhang (2007).

Um enfoque diferente e mais prático para avaliação de influência local em modelos lineares mistos é apresentada por Beckman et al. (1987) e Lesaffre & Verbeke

(1998). Esses autores consideram a verossimilhança marginal de $\boldsymbol{\psi} = (\boldsymbol{\beta}^\top, \boldsymbol{\theta}^\top)^\top$, nomeadamente

$$L(\boldsymbol{\psi}) = \sum_{i=1}^n L_i(\boldsymbol{\psi}) = \sum_{i=1}^n -\frac{1}{2} \left\{ \log |\boldsymbol{\Omega}_i| + (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta})^\top \boldsymbol{\Omega}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}) \right\}, \quad (2.5.17)$$

e perturbam seus componentes tomando

$$L_i(\boldsymbol{\psi}_\omega) = \sum_{i=1}^n \omega_i L_i(\boldsymbol{\psi}), \quad (2.5.18)$$

em que L_i denota a verossimilhança associada à i -ésima unidade amostral e $\boldsymbol{\omega} = (\omega_1, \dots, \omega_n)^\top$. Aqui, $\boldsymbol{\omega}_0 = \mathbf{1}_n$ em que $\mathbf{1}_n$ denota um vetor n -dimensional com todos os elementos iguais a 1. Definindo \mathbf{u}_i como um vetor n -dimensional com 1 na posição i e zero nas demais, a curvatura normal (2.5.16) calculada na direção de \mathbf{u}_i é

$$C_{\mathbf{u}_i} = 2 \left| \mathbf{u}_i^\top \mathbf{H}^\top \ddot{\mathbf{F}}^{-1} \mathbf{H} \mathbf{u}_i \right| = 2 \left| \mathbf{h}_i^\top \ddot{\mathbf{F}}^{-1} \mathbf{h}_i \right|, \quad (2.5.19)$$

em que \mathbf{h}_i denota a i -ésima coluna de \mathbf{H} . Valores grandes de (2.5.19) indicam um grande impacto da i -ésima unidade amostral no deslocamento da verossimilhança num sentido local, *i.e.*, uma pequena perturbação no correspondente vetor de pesos ω_i induz um grande deslocamento na log-verossimilhança correspondente. Lesaffre & Verbeke (1998) também mostram que $C_i = C_{\mathbf{u}_i} = -2(\hat{\boldsymbol{\psi}} - \hat{\boldsymbol{\psi}}_{(i)}^\top)^\top \ddot{\mathbf{F}}(\hat{\boldsymbol{\psi}} - \hat{\boldsymbol{\psi}}_{(i)})$ em que $\hat{\boldsymbol{\psi}}_{(i)}^\top$ corresponde à aproximação em um passo para $\hat{\boldsymbol{\psi}}_{(i)}$ e que para n grande, é equivalente à proposta de Pregibon (1981) para medir a **influência global local** (**global local influence**) da i -ésima unidade amostral. Esses autores mostram que o cálculo da curvatura normal (2.5.16) na direção de \mathbf{d}_{max} correspondente ao maior autovetor de $-\mathbf{H}^\top \ddot{\mathbf{F}}^{-1} \mathbf{H}$ indica a perturbação do modelo que induz a maior mudança no deslocamento da verossimilhança.

Como C_i está relacionado com o impacto da i -ésima unidade amostral no vetor de parâmetros $\boldsymbol{\psi}$, Lesaffre & Verbeke (1998) propõem uma decomposição dessa medida que pode ser utilizada para indicar que componente do modelo é mais afetada pela unidade amostral considerada. Em particular, eles recomendam gráficos de índices de $\|\mathbf{I}_{n_i} - \hat{\boldsymbol{\epsilon}}_i \hat{\boldsymbol{\epsilon}}_i^\top\|^2$, $\|\hat{\boldsymbol{\epsilon}}_i \hat{\boldsymbol{\epsilon}}_i^\top\|^2$, $\|\hat{\mathbf{X}}_i \hat{\mathbf{X}}_i^\top\|^2$, $\|\hat{\mathbf{Z}}_i \hat{\mathbf{Z}}_i^\top\|^2$ e de $\|\hat{\boldsymbol{\Omega}}_i^\top\|^2$ em que $\hat{\boldsymbol{\epsilon}}_i = \hat{\boldsymbol{\Omega}}_i^{-1/2} \hat{\boldsymbol{\xi}}_i$, $\hat{\mathbf{X}}_i = \hat{\boldsymbol{\Omega}}_i^{-1/2} \mathbf{X}_i$ e $\hat{\mathbf{Z}}_i = \hat{\boldsymbol{\Omega}}_i^{-1/2} \mathbf{Z}_i$ como uma ferramenta diagnóstica quando unidades amostrais influentes são detectadas por meio de C_i . A análise baseada em $\mathcal{V}_i = \|\mathbf{I}_{n_i} - \hat{\boldsymbol{\epsilon}}_i \hat{\boldsymbol{\epsilon}}_i^\top\|^2$ está intimamente ligada à análise de resíduos, como já indicado na Tabela 2.5.1.

Finalmente, esses autores mencionam que seus resultados são obtidos com os estimadores de máxima verossimilhança dos parâmetros da matriz de covariâncias

e que podem não ser válidos quando o método de máxima verossimilhança restrita é considerado. Todavia, como essas propostas devem ser encaradas sob o ponto de vista exploratório, sua utilização em quaisquer circunstâncias pode trazer mais benefícios do que perdas.

2.6 Notas de capítulo

2.7 Exercícios

2.7.1. Indique como matrizes de covariância com estrutura uniforme ou AR(1) podem ser expressa no formato (2.2.12).

2.7.2. Proponha modelos no formato (2.1.4) para a matriz de covariâncias intraunidades amostrais associadas aos Exemplos 1.2.1, 1.2.4, 1.2.5 e 1.2.6 que levem em conta todas as fontes de variabilidade correspondentes.

2.7.3. Mostre que $\mathbb{V}(\mathbf{e}^*) = \sigma^2 \mathbf{I}$ em (2.3.13).

2.7.4. Mostre que (2.3.16) e (2.3.17) são soluções de (2.3.15).

2.7.5. Mostre que sob o modelo (2.1.7) o vetor de resíduos condicionais pode ser expresso como $\hat{\mathbf{e}} = \mathbf{RQe} + \mathbf{RQZb}$.

2.7.6. Considere o modelo $\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{1}_m a_i + \mathbf{e}_i$, $i = 1, \dots, n$, em que $a_i \sim N(0, \sigma_a^2)$ e $\mathbf{e}_i \sim N(\mathbf{0}, \mathbf{I}_m \sigma_e^2)$. Obtenha os estimadores de máxima verossimilhança de $\boldsymbol{\beta}$, σ_a^2 e σ_e^2 .

2.7.7. Considere o modelo de coeficientes aleatórios homocedástico com independência condicional balanceado, *i.e.*, (2.1.2) com $\mathbf{Z}_i = \mathbf{X}_i = \mathbf{X}$ e $\mathbf{R}_i = \mathbf{I}_m \sigma^2$. Mostre que o estimador de máxima verossimilhança de $\boldsymbol{\beta}$ é $\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \bar{\mathbf{y}}$ em que $\bar{\mathbf{y}} = n^{-1} \sum_{i=1}^n \mathbf{y}_i$.

Sugestão: Considere a parametrização $\boldsymbol{\Omega}_i = \sigma^2(\mathbf{XG}\mathbf{X}^\top + \mathbf{I}_m)$, use a decomposição $\mathbf{y}_i - \mathbf{X}\boldsymbol{\beta} = (\mathbf{y}_i - \bar{\mathbf{y}}) + (\bar{\mathbf{y}} - \mathbf{X}\boldsymbol{\beta})$ na forma quadrática a ser minimizada e o resultado vi) da Seção A.1.6 do Apêndice A para obter a inversa da matriz de covariâncias intraunidades amostrais.

Capítulo 3

Modelos para dados não gaussianos

3.1 Modelos elipticamente simétricos e assimétricos

A falta de robustez de modelos lineares mistos com distribuições gaussianas para os efeitos aleatórios e erros tem motivado uma intensa atividade de pesquisa baseada em outras distribuições, principalmente aquelas classificadas como **elipticamente simétricas** (*elliptically symmetric*) ou **assimétricas** (*skew-elliptical*). Dentre os diversos autores que adotam esse enfoque, destacamos Pinheiro et al. (2001), Savalli et al. (2006), Osorio, Paula & Galea (2007), Arellano-Vale et al. (2005), Arellano-Vale, Bolfarine & Lachos (2007), Bolfarine, Montenegro & Lachos (2007) e Lachos et al. (2010).

A classe de distribuições elipticamente simétricas engloba a distribuição normal, a distribuição t multivariada, a **distribuição potência exponencial** (*power exponential*) bem como outras com densidades da forma

$$f(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \alpha) = |\boldsymbol{\Sigma}|^{-1/2} g[(\mathbf{y} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \boldsymbol{\mu})] \quad (3.1.1)$$

em que \mathbf{y} é um vetor aleatório m -dimensional, $\mathbb{E}(\mathbf{y}) = \boldsymbol{\mu}$, $\mathbb{V}(\mathbf{y}) = \boldsymbol{\Omega} = \alpha \boldsymbol{\Sigma}$ com α representando uma constante $[= \nu/(\nu - 2), \nu > 2]$ para a distribuição- t multivariada com ν graus de liberdade, por exemplo] e g é uma função com valores não-negativos definida em \mathbb{R} e tal que $\int_0^\infty u^{m/2-1} g(u) du < \infty$.

Modelos lineares mistos com efeitos aleatórios e erros com distribuição (3.1.1)

podem ser definidos de forma hierárquica como nos modelos gaussianos, *i.e.*,

$$\begin{aligned} \mathbf{y}_i | \mathbf{b}_i &\sim ES_{m_i}[\mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{b}_i, \mathbf{R}_i(\boldsymbol{\theta}), \alpha_i, \gamma_i], \\ \mathbf{b}_i &\sim ES_q[\mathbf{0}, \mathbf{G}(\boldsymbol{\theta}), \alpha_i], \\ \mathbf{e}_i &\sim ES_{m_i}[\mathbf{0}, \mathbf{R}_i(\boldsymbol{\theta}), \gamma_i], \end{aligned}$$

$i = 1, \dots, n$ com \mathbf{b}_i e \mathbf{e}_i não correlacionados. Apesar de que em alguns casos especiais (*e.g.*, a distribuição t multivariada) a distribuição conjunta de $(\mathbf{y}_i^\top, \mathbf{b}_i^\top)^\top$ pertence à mesma classe, isto não vale em geral. Por razões computacionais e de generalizabilidade, muitos autores consideram a seguinte especificação alternativa para o modelo

$$\begin{bmatrix} \mathbf{y}_i \\ \mathbf{b}_i \end{bmatrix} \sim ES_{m_i+q} \left\{ \begin{bmatrix} \mathbf{X}_i \boldsymbol{\beta} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{Z}_i \mathbf{G}(\boldsymbol{\theta}) \mathbf{Z}_i^\top + \mathbf{R}_i(\boldsymbol{\theta}) & \mathbf{Z}_i \mathbf{G}(\boldsymbol{\theta}) \\ \mathbf{G}(\boldsymbol{\theta}) \mathbf{Z}_i^\top & \mathbf{G}(\boldsymbol{\theta}) \end{bmatrix}, \alpha_i, \gamma_i \right\}$$

e trabalham com o modelo marginal

$$\mathbf{y}_i \sim ES_{m_i}[\mathbf{X}_i \boldsymbol{\beta}, \mathbf{D}_i(\boldsymbol{\theta}), \alpha_i],$$

com $\mathbf{D}_i(\boldsymbol{\theta}) = \mathbf{Z}_i \mathbf{G}(\boldsymbol{\theta}) \mathbf{Z}_i^\top + \mathbf{R}_i(\boldsymbol{\theta})$ geralmente assumindo que $\mathbf{R}_i(\boldsymbol{\theta}) = \sigma^2 \mathbf{I}_{m_i}$. Em teoria, os parâmetros α_i e γ_i podem variar entre as unidades amostrais, mas na prática, é comum adotar $\alpha_i = \alpha$ e $\gamma_i = \gamma$ ou considerar apenas alguns valores para esses parâmetros, *e.g.*, um para cada conjunto de unidades amostrais agrupadas segundo algum critério, conforme sugerem Pinheiro et al. (2001).

A estimação dos parâmetros por máxima verossimilhança pode ser obtida de forma paralela àquela utilizada no caso gaussiano e consiste na iteração das seguintes expressões

$$\boldsymbol{\beta}^{(h+1)} = \left\{ \sum_{i=1}^n v(u_i^{(h)}) \mathbf{X}_i^\top [\mathbf{D}_i(\boldsymbol{\theta})^{(h)}]^{-1} \mathbf{X}_i \right\}^{-1} \left\{ \sum_{i=1}^n v(u_i^{(h)}) \mathbf{X}_i^\top [\mathbf{D}_i(\boldsymbol{\theta})^{(h)}]^{-1} \mathbf{y}_i \right\}$$

e

$$\boldsymbol{\theta}^{(h+1)} = \operatorname{argmax}_{\boldsymbol{\theta}} \{l(\boldsymbol{\beta}^{(h)}, \boldsymbol{\theta})\}$$

em que $l(\boldsymbol{\beta}, \boldsymbol{\theta})$ denota a log-verossimilhança,

$$u_i^{(h)} = (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}^{(h)})^\top [\mathbf{D}_i(\boldsymbol{\theta})^{(h)}]^{-1} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}^{(h)})$$

e $v(u) = -2g'(u)/g(u)$, for $h = 0, 1, 2, \dots$. Como $v(u_i)$ pode ser interpretado como um peso que, em geral, atribui menos importância às observações mais extremas na ótica da distância de Mahalanobis, u_i , modelos lineares mistos baseados em distribuições elípticamente simétricas podem acomodar unidades amostrais discrepantes detectadas sob modelos gaussianos correspondentes.

Para efeito inferencial, precisamos nos apoiar em resultados assintóticos mesmo em casos especiais e é comum [ver Pinheiro et al. (2001), por exemplo] considerar aproximações por distribuições gaussianas para os estimadores de máxima verossimilhança $\hat{\beta}$ e $\hat{\theta}$. Savalli et al. (2006) sugerem que para n suficientemente grande

$$\hat{\beta} \approx N_p(\beta, \mathbf{K}_\beta^{-1})$$

em que $\mathbf{K}_\beta = \sum_{i=1}^n 4d_{gi}/m_i \mathbf{X}_i^\top [\mathbf{D}_i(\theta)]^{-1} \mathbf{X}_i$ com $d_{gi} = \mathbb{E}\{[g'(u_i)/g(u_i)]^2 u_i\}$ é a matriz de informação de Fisher correspondente. Um resultado similar pode ser considerado para $\hat{\theta}$.

Assumindo θ conhecido, preditores para os efeitos aleatórios podem ser obtidos por meio de um argumento bayesiano e podem ser expressos como aqueles obtidos no caso gaussiano, a saber

$$\hat{\mathbf{b}}_i = \mathbb{E}(\mathbf{b}_i | \mathbf{y}_i, \hat{\beta}, \theta) = \mathbf{G}(\theta) \mathbf{Z}_i^\top [\mathbf{D}_i(\theta)]^{-1} (\mathbf{y}_i - \mathbf{X}_i \hat{\beta}).$$

A matriz de covariância de $\hat{\mathbf{b}}_i$ não é tão facilmente obtida como no caso bayesiano por causa dos pesos $v(u_i)$. Assumindo pesos fixos e usando a notação compacta, temos

$$\mathbb{V}(\hat{\mathbf{b}}) = \mathbf{\Gamma}(\theta) \mathbf{Z}^\top [\mathbf{D}(\theta)]^{-1} \mathbb{V}(\mathbf{y} - \mathbf{X}\hat{\beta}) [\mathbf{D}(\theta)]^{-1} \mathbf{Z} \mathbf{\Gamma}(\theta)$$

com

$$\mathbb{V}(\mathbf{y} - \mathbf{X}\hat{\beta}) = \mathbb{V}(\hat{\xi}) = \mathbf{D}(\theta)^* \mathbf{Q}(\theta)^* \mathbf{\Omega}(\theta) \mathbf{Q}(\theta)^* \mathbf{D}(\theta)^*,$$

$$\mathbf{D}(\theta)^* = \otimes_{i=1}^n [v(u_i) \mathbf{D}_i(\theta)],$$

$$\mathbf{Q}(\theta)^* = [\mathbf{D}(\theta)^*]^{-1} - [\mathbf{D}(\theta)^*]^{-1} \mathbf{X} (\mathbf{X}^\top [\mathbf{D}(\theta)^*]^{-1} \mathbf{X})^{-1} \mathbf{X}^\top [\mathbf{D}(\theta)^*]^{-1}.$$

Em aplicações práticas, deve-se substituir os termos desconhecidos por estimativas convenientes.

Análises de resíduos e de influência global para modelos lineares mistos baseados em distribuições elipticamente simétricas ainda não estão bem estabelecidas. Contudo, a similaridade entre tipo de modelos e os modelos gaussianos sugere algumas ferramentas exploratórias. Por exemplo, gráfico de resíduos marginais ponderados, $\hat{\xi}_i = \sqrt{v(\hat{u}_i)} [\mathbf{y}_i - \mathbf{X}_i \hat{\beta}]$ em função dos índices das unidades amostrais podem ser utilizados para identificar aquelas possivelmente discrepantes, conforme sugerem Savalli et al. (2006). Notemos que, mesmo para as distribuições elipticamente simétricas mais comuns, como a distribuição t multivariada ou a distribuição de potência exponencial, os pesos $v(\hat{u}_i) = (\mathbf{y}_i - \mathbf{X}_i \hat{\beta})^\top [\mathbf{D}_i(\hat{\theta})]^{-1} (\mathbf{y}_i - \mathbf{X}_i \hat{\beta})$ tendem a reduzir o impacto de unidades amostrais associadas aos maiores valores da distância de Mahalanobis e, conseqüentemente, acomodar possíveis unidades amostrais discrepantes detectadas por meio do ajuste de modelos gaussianos.

Uma análise semelhante pode ser realizada com os resíduos condicionais $\hat{\mathbf{e}}_i = \mathbf{y}_i - \mathbf{X}_i\hat{\boldsymbol{\beta}} - \mathbf{Z}_i\hat{\mathbf{b}}_i$. Acreditamos que as sugestões de Schützenmeister & Piepho (2012) poderiam ser úteis nesse contexto, mas esse é um tópico sobre o qual ainda não há resultados estabelecidos.

Análises de influência local, por outro lado, é considerada em detalhe por Osorio et al. (2007). Esses autores obtêm expressões gerais para \mathbf{H} e $\ddot{\mathbf{F}}$ em (2.5.16) correspondentes a perturbações nos pesos (w_i), na matriz de escala (\mathbf{D}_i), ou nas variáveis resposta (\mathbf{y}_i) e explicativas (\mathbf{X}_i). Na prática, devemos escolher um membro específico da classe de distribuições elípticamente simétricas, especificando g , α e $\mathbf{D}(\boldsymbol{\theta})$ para obter a medida de influência local correspondente, (2.5.19).

Como sugerem Mudholkar & Hutson (2000), os efeitos de assimetria nos resultados de análises baseadas em modelos gaussianos são mais sérios do que aqueles causados por distribuições com caudas pesadas e por essa razão, um esforço considerável tem sido dirigido a modelos lineares mistos com distribuições assimétricas para efeitos aleatórios e erros. Jara, Quintana & San Martín (2008) adotam um enfoque bayesiano para tratar distribuições elípticamente assimétricas nesse contexto. Especificamente, eles consideram um modelo em dois estágios definido como

$$\mathbf{y}_i | (\boldsymbol{\beta}, \mathbf{b}_i, \sigma^2, \mathbf{R}_i, \boldsymbol{\Lambda}_{e_i}, g_{\theta_e}^{m_i}) \sim SE_{m_i}(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i, \sigma^2\mathbf{R}_i, \boldsymbol{\Lambda}_{e_i}, g_{\theta_e}^{m_i}) \quad (3.1.2)$$

$$\mathbf{b}_i | (\mathbf{G}, \boldsymbol{\Lambda}_b, g_{\theta_b}^q) \sim SE_q(\mathbf{0}, \mathbf{G}, \boldsymbol{\Lambda}_b, g_{\theta_b}^q) \quad (3.1.3)$$

em que $SE_r(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Lambda}, g_\theta)$ denota uma distribuição elípticamente assimétrica r -dimensional com vetor de localização r -dimensional $\boldsymbol{\mu}$, matriz de escala $\boldsymbol{\Sigma}$, matriz de assimetria r -dimensional $\boldsymbol{\Lambda}$ e função geradora de densidade g_θ . Embora os modelos propostos por Jara et al. (2008) sejam bem gerais, ajustá-los a problemas práticos pode ser bastante complicado e nesse caso, devemos nos restringir a modelos mais específicos, como aqueles pertencentes à classe dos **modelos normais assimétricos** (*skew-normal*).

Arellano-Vale et al. (2007) também adotam um enfoque bayesiano para analisar modelos lineares mistos com distribuições gaussianas para os efeitos aleatórios e erros, supondo que

$$\mathbf{y}_i | [\boldsymbol{\beta}, \mathbf{b}_i, \mathbf{R}_i(\boldsymbol{\theta}), \boldsymbol{\Lambda}_{e_i}] \sim SN_{m_i}[\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i, \mathbf{R}_i(\boldsymbol{\theta}), \boldsymbol{\Lambda}_{e_i}] \quad (3.1.4)$$

$$\mathbf{b}_i | [\mathbf{G}(\boldsymbol{\theta}), \boldsymbol{\Lambda}_b] \sim SN_q[\mathbf{0}, \mathbf{G}(\boldsymbol{\theta}), \boldsymbol{\Lambda}_b] \quad (3.1.5)$$

em que $SN_r(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Lambda})$ denota uma distribuição normal assimétrica com vetor de localização r -dimensional $\boldsymbol{\mu}$, matriz de escala r -dimensional $\boldsymbol{\Sigma}$ e $\boldsymbol{\Lambda} = \bigoplus_{i=1}^r \lambda_i$ com $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_r)^\top$ representando os parâmetros de assimetria. A função densidade correspondente é

$$f(\mathbf{y} | \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Lambda}) = 2^r \phi_r(\mathbf{y} | \boldsymbol{\mu}, \boldsymbol{\Sigma} + \boldsymbol{\Lambda}^2) \Phi_r[\boldsymbol{\Lambda}((\boldsymbol{\Sigma} + \boldsymbol{\Lambda}^2)^{-1}(\mathbf{y} - \boldsymbol{\mu}) | \mathbf{0}, \mathbf{I}_r + \boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Lambda})]$$

em que $\phi_r(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ e $\Phi_r(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotam, respectivamente, as funções densidade e distribuição de uma distribuição gaussiana com vetor de médias $\boldsymbol{\mu}$ e matriz de covariâncias $\boldsymbol{\Sigma}$ calculada no ponto \mathbf{y} . Isto implica que

$$\mathbb{E}(\mathbf{y}) = \boldsymbol{\mu} + \sqrt{2/\pi} \boldsymbol{\lambda}, \text{ e } \mathbb{V}(\mathbf{y}) = \boldsymbol{\Sigma} + (1 - 2/\pi) \boldsymbol{\Lambda}^2.$$

Para o caso especial em que $\mathbf{R}_i = \sigma^2 \mathbf{I}_{m_i}$ e $\boldsymbol{\Lambda}_{e_i} = \lambda \mathbf{I}_{m_i}$ Arellano-Vale et al. (2007) mostram como ajustar o modelo baseado na distribuição marginal de \mathbf{y} por meio de metodologia bayesiana, alertando que “Mesmo com densidades de difícil trato analítico, mostramos que esses modelos podem ser facilmente ajustados por intermédio de métodos MCMC”.

Bolfarine et al. (2007) consideram modelos lineares mistos do tipo (2.1.2) com uma formulação ligeiramente modificada para acomodar distribuições normais assimétricas. Especificamente, esses autores consideram distribuições com função densidade da forma

$$f(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}) = 2\phi_r(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma} + \boldsymbol{\lambda}\boldsymbol{\lambda}^\top) \Phi_1 \left[\frac{\boldsymbol{\lambda}^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y} - \boldsymbol{\mu})}{(1 + \boldsymbol{\lambda}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\lambda})^{1/2}} \right]$$

para a qual

$$\mathbb{E}(\mathbf{y}) = \boldsymbol{\mu} + \sqrt{2/\pi} \boldsymbol{\lambda}, \text{ e } \mathbb{V}(\mathbf{y}) = \boldsymbol{\Sigma} + (1 - 2/\pi) \boldsymbol{\lambda}\boldsymbol{\lambda}^\top.$$

Eles restringem a atenção ao caso especial em que $\mathbf{R}_i = \sigma^2 \mathbf{I}_{m_i}$ e expressam o modelo numa formulação em dois estágios que facilita a obtenção da função densidade marginal de \mathbf{y}_i , nomeadamente,

$$f(\mathbf{y}_i|\boldsymbol{\mu}, \mathbf{G}(\boldsymbol{\theta}), \boldsymbol{\lambda}) = 2\phi_{m_i}\{\mathbf{y}_i|\mathbf{X}_i\boldsymbol{\beta}, \mathbf{Z}_i([\mathbf{G}(\boldsymbol{\theta})]^{-1} + \boldsymbol{\lambda}\boldsymbol{\lambda}^\top)\mathbf{Z}_i^\top\} \Phi_1\{\bar{\boldsymbol{\lambda}}_i(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})\}$$

em que

$$\bar{\boldsymbol{\lambda}}_i = \frac{\boldsymbol{\lambda}^\top [\mathbf{G}(\boldsymbol{\theta})]^{-1} [\mathbf{D}_i(\boldsymbol{\theta})]^{-1} \mathbf{Z}_i^\top [\mathbf{R}_i(\boldsymbol{\theta})]}{\{1 + \boldsymbol{\lambda}^\top [\mathbf{G}(\boldsymbol{\theta})]^{-1} \boldsymbol{\lambda} + \boldsymbol{\lambda}^\top [\mathbf{G}(\boldsymbol{\theta})]^{-1} [\mathbf{D}_i(\boldsymbol{\theta})]^{-1} [\mathbf{G}(\boldsymbol{\theta})]^{-1} \boldsymbol{\lambda}\}^{1/2}}$$

com

$$\mathbf{D}_i(\boldsymbol{\theta}) = \{[\mathbf{G}(\boldsymbol{\theta}) + \boldsymbol{\lambda}\boldsymbol{\lambda}^\top]^{-1} + \mathbf{Z}_i[\mathbf{R}_i(\boldsymbol{\theta})]^{-1} \mathbf{Z}_i^\top\}.$$

Nesse contexto, eles obtêm um algoritmo EM para cálculo dos estimadores de máxima verossimilhança dos parâmetros do modelo e propõem ferramentas de diagnóstico de influência local baseadas na perturbação da matriz de dispersão associada aos efeitos aleatórios $[\mathbf{G}(\boldsymbol{\theta})]$, nos pesos (w_i), nas variáveis explicativas (\mathbf{X}_i) e na variável resposta (\mathbf{y}_i). Eles também consideram ferramentas de diagnóstico baseadas na omissão de casos mas mencionam que a avaliação da qualidade do ajuste ainda é um problema que requer investigação.

3.2 Modelos lineares generalizados mistos (GLMM)

Modelos lineares generalizados mistos (GLMM) constituem uma extensão natural de modelos lineares generalizados (GLM) nos quais os efeitos aleatórios são introduzidos para acomodar alguma estrutura de covariância intraunidades amostrais. Sua natureza permite incorporar não só distribuições não gaussianas como também possíveis relações não lineares entre as variáveis resposta e explicativas.

O modelo também pode ser especificado em dois estágios. No primeiro, assumimos que a distribuição condicional da resposta y_{ij} dados os efeitos aleatórios \mathbf{b}_i , pertence à família exponencial, *i.e.*, tem densidade

$$f(y_{ij}|\mathbf{b}_i) = \exp\{\phi[y_{ij}\vartheta_{ij} - a(\vartheta_{ij})] + c(y_{ij}, \phi)\} \quad (3.2.1)$$

em que a e c são funções conhecidas e ϕ é um parâmetro de escala. Isso implica que $\mathbb{E}(y_{ij}|\mathbf{b}_i) = \mu_{ij} = da(\vartheta_{ij})/d\vartheta_{ij}$ e $\mathbb{V}(y_{ij}|\mathbf{b}_i) = \phi^{-1}d^2(\vartheta_{ij})/d\vartheta_{ij}^2$. Então, tomamos

$$g(\mu_{ij}) = \eta_{ij} = \mathbf{x}_{ij}^\top \boldsymbol{\beta} + \mathbf{z}_{ij}^\top \mathbf{b}_i \quad (3.2.2)$$

em que g é uma função de ligação conveniente (a função identidade quando quisermos ajustar um modelo linear misto) e \mathbf{x}_{ij}^\top e \mathbf{z}_{ij}^\top denotam, respectivamente, a j -ésima linha das matrizes \mathbf{X}_i e \mathbf{Z}_i de especificação dos efeitos fixos e aleatórios, respectivamente. No segundo estágio, usualmente supomos que $\mathbf{b}_i \sim N[\mathbf{0}, \mathbf{G}(\boldsymbol{\theta})]$ embora em teoria, outras distribuições possam ser consideradas.

Modelos lineares generalizados mistos têm uma estrutura similar àquela considerada para os modelos lineares mistos, que na realidade constituem um caso particular daqueles. No entanto a relação entre os parâmetros de localização e de covariância por eles induzida traz mais dificuldades analíticas e de interpretação.

Considere, por exemplo, os dados mencionados no Exercício 1.4.4, provenientes de um experimento realizado na Faculdade de Medicina da USP com o objetivo de comparar pacientes com hepatite C tratados com dois inibidores de protease (Telaprevir e Boceprevir) relativamente à ocorrência de disfunção renal durante o período de observação de 48 semanas. A resposta (ocorrência de disfunção renal) assume o valor 1 quando o nível de creatinina numa determinada semana tem valor 15% maior que o correspondente valor basal e um dos objetivos era comparar os dois tratamentos com relação às chances de ocorrência de disfunção renal ao longo do período de observação controlando outras covariáveis (sexo, idade, hipertensão arterial etc.). Para efeito didático, suponhamos que a ocorrência de disfunção renal dependa apenas do tempo de tratamento e que cada indivíduo tenha uma susceptibilidade própria.

Um GLMM em que supomos que a distribuição condicional da ocorrência de disfunção renal para o indivíduo i no instante t_{ij} é a distribuição Bernoulli e que a

função de ligação é a função logito pode ser representado como

$$\log \left\{ \frac{P(Y_{ij} = 1|t_{ij}, b_i)}{1 - [P(Y_{ij} = 1|t_{ij}, b_i)]} \right\} = \alpha + b_i + \beta t_{ij} \quad (3.2.3)$$

ou, equivalentemente, como

$$P(Y_{ij} = 1|t_{ij}, b_i) = \frac{\exp(\alpha + b_i + \beta t_{ij})}{1 + \exp(\alpha + b_i + \beta t_{ij})} \quad (3.2.4)$$

em que $Y_{ij} = 1$ corresponde à ocorrência de disfunção renal para o indivíduo i na semana t_{ij} , b_i é um efeito aleatório com distribuição $N(0, \sigma^2)$, $\alpha + b_i$ representa o logaritmo da chance (*log-odds*) de ocorrência de disfunção renal para o indivíduo i no início do estudo ($t_{ij} = 0$) e β indica o correspondente logaritmo da razão de chances para duas semanas consecutivas. Note que o parâmetro α denota a chance de ocorrência de disfunção renal para um indivíduo com $b_i = 0$ no início do estudo.

O fato de o parâmetro β estar associado à mudança no logaritmo da chance de ocorrência de disfunção renal para **um indivíduo específico** sugere que modelos lineares generalizados mistos não são apropriados para situações em que o interesse recai no parâmetro populacional correspondente. Esse parâmetro marginal é o valor esperado (em relação à distribuição dos efeitos aleatórios) dos parâmetros individuais (β).

Para ilustrar a diferença entre os parâmetros individuais e o correspondente parâmetro marginal, consideremos o modelo

$$\log \left\{ \frac{P(Y = 1|t, b_i)}{1 - [P(Y = 1|t, b_i)]} \right\} = \alpha + b_i + \beta t$$

com $\alpha = -5.5$, $\beta = 0.90$, $t \in (0, 12)$ e $b_i \sim N(0, 4)$. Um gráfico da probabilidade de sucesso (ocorrência de disfunção renal, no exemplo) em função do tempo para uma amostra hipotética de 15 indivíduos está apresentado na Figura 3.2.1 juntamente com a curva marginal (linha sólida). A resposta marginal não tem o mesmo padrão que as respostas individuais, pois corresponde a

$$\begin{aligned} \mathbb{E}_{N(0, \sigma_b^2)}[\mathbb{E}_{f(\alpha, \beta)}(Y|b_i)] &= \int_{-\infty}^{\infty} \frac{\exp(\alpha + b_i + \beta t)}{1 + \exp(\alpha + b_i + \beta t)} \frac{1}{\sqrt{2\pi\sigma_b^2}} \exp[-b_i^2/(2\sigma_b^2)] db_i \\ &\neq \frac{\exp(\alpha + \beta t)}{1 + \exp(\alpha + \beta t)}. \end{aligned}$$

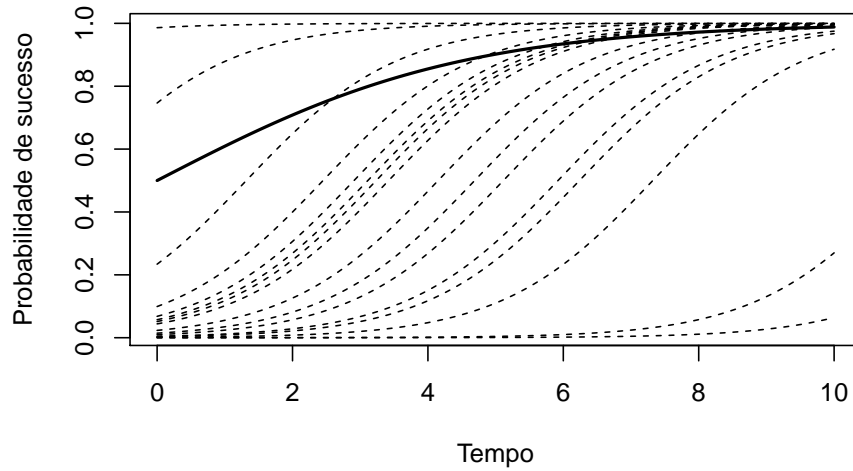
Esta integral não tem forma fechada e nem tem a estrutura (3.2.4). Neste caso especial, Zeger et al. (1988) mostram que

$$\log \left[\frac{\mathbb{E}_{f(\alpha, \beta)}(Y)}{1 - \mathbb{E}_{f(\alpha, \beta)}(Y)} \right] = \log \left[\frac{P(Y = 1|t)}{1 - [P(Y = 1|t)]} \right] \approx (c^2 \sigma_b^2 + 1)^{-1/2} \beta t$$

com $c^2 \approx 0.346$ e consequentemente o coeficiente β^* correspondente ao modelo marginal pode ser aproximado como $\beta^* = (0.346\sigma_b^2 + 1)^{-1/2}\beta$.

Em geral, a integral envolvida na eliminação dos efeitos aleatórios para obtenção da curva de resposta populacional não tem solução explícita e precisa ser obtida por métodos numéricos. Embora se possa construir essa curva, estimativas dos parâmetros populacionais não podem ser obtidos a não ser em casos simples como aquele que acabamos de descrever. Detalhes podem ser obtidos em Molenberghs & Verbeke (2005).

Figura 3.2.1: Probabilidade de sucesso para um exemplo hipotético



A função de verossimilhança marginal para o GLMM definido por (3.2.1)-(3.2.2) é

$$\begin{aligned} L(\boldsymbol{\beta}, \boldsymbol{\theta}, \phi | \mathbf{y}) &= \kappa \prod_{i=1}^n \int f(\mathbf{y}_i | \boldsymbol{\beta}, \mathbf{b}_i) \Phi[\mathbf{b}_i | \mathbf{0}, \mathbf{G}(\boldsymbol{\theta})] d\mathbf{b}_i \\ &= \kappa \prod_{i=1}^n \int f(\mathbf{y}_i | \boldsymbol{\beta}, \mathbf{b}_i) |\mathbf{G}(\boldsymbol{\theta})|^{-1/2} \exp\left(-\frac{1}{2} \mathbf{b}_i^\top [\mathbf{G}(\boldsymbol{\theta})]^{-1} \mathbf{b}_i\right) d\mathbf{b}_i \end{aligned} \quad (3.2.5)$$

em que κ é uma constante. No caso em que a distribuição da resposta é Bernoulli, temos

$$f(\mathbf{y}_i | \boldsymbol{\beta}, \mathbf{b}_i) = \prod_{j=1}^{m_i} \left[\frac{\exp(\mathbf{x}_{ij}^\top \boldsymbol{\beta} + \mathbf{z}_{ij}^\top \mathbf{b}_i)}{1 + \exp(\mathbf{x}_{ij}^\top \boldsymbol{\beta} + \mathbf{z}_{ij}^\top \mathbf{b}_i)} \right]^{y_{ij}} \left[\frac{1}{1 + \exp(\mathbf{x}_{ij}^\top \boldsymbol{\beta} + \mathbf{z}_{ij}^\top \mathbf{b}_i)} \right]^{1-y_{ij}}.$$

A obtenção dos estimadores de máxima verossimilhança de β e θ envolve a maximização de (3.2.5) que exige o cálculo de n integrais (com dimensão q) que, em geral, são algebricamente intratáveis e precisam ser aproximadas. Vários métodos de aproximação podem ser encontrados na literatura estatística e uma excelente revisão pode ser encontrada em Tuerlinckx, Rijmen & Verbeke (2006).

Em geral, esses métodos são baseados quer na aproximação direta dessas integrais quer na aproximação do integrando de forma que a integral do integrando aproximado tenha uma forma tratável. Em ambos os casos, a maximização da função resultante dessas aproximações pode ser concretizada por métodos padrão. A seguir, apresentamos as ideias que regem esses procedimentos.

Aproximação da integral

Um método comumente utilizado para a aproximação de integrais de funções é aquele baseado em quadratura de Gauss-Hermite. Nesse contexto, integrais de funções do tipo $g(x) \exp(-x^2)$ podem ser aproximadas como

$$\int g(x) e^{-x^2} dx \approx \sum_{k=1}^K g(x_k) w_k$$

em que x_k e w_k , $k = 1, \dots, K$ são, respectivamente os m nós e pesos da quadratura de Gauss-Hermite. Em particular, a função de verossimilhança marginal correspondente à i -ésima unidade amostral em (3.2.5) é aproximada como

$$L_i(\beta, \theta, \phi | \mathbf{y}) \approx \sum_{k_1=1}^K \dots \sum_{k_q=1}^K f(\mathbf{y}_i | \beta, [\mathbf{G}(\theta)]^{1/2} \mathbf{x}, \phi) w_{k_1} \dots w_{k_q} \quad (3.2.6)$$

em que $\mathbf{x} = (x_{k_1}, \dots, x_{k_q})^\top$ e w_{k_1}, \dots, w_{k_q} são, respectivamente, nós e pesos apropriados.

Alternativamente (3.2.5) pode ser aproximada pelo método de Monte Carlo. Nesse contexto, a i -ésima componente dessa expressão, nomeadamente

$$L_i(\beta, \theta, \phi | \mathbf{y}_i) = K \prod_{i=1}^n \int f(\mathbf{y}_i | \beta, \mathbf{b}_i) \Phi[\mathbf{b}_i | \mathbf{0}, \mathbf{G}(\theta)] d\mathbf{b}_i = \mathbb{E}_{\Phi[\mathbf{0}, \mathbf{G}(\theta)]} [f(\mathbf{y}_i | \beta, \mathbf{b}_i)]$$

é aproximada por

$$L_i(\beta, \theta, \phi | \mathbf{y}_i) \approx \frac{1}{K} \sum_{k=1}^K f(\mathbf{y}_i | \beta, \mathbf{b}_i^{(k)}) \quad (3.2.7)$$

em que $\mathbf{b}_i^{(1)}, \dots, \mathbf{b}_i^{(K)}$ é uma amostra aleatória gerada de uma distribuição $\Phi[\mathbf{0}, \mathbf{G}(\hat{\theta})]$.

Aproximação do integrando

O objetivo da aproximação do integrando é obter uma integral para a qual existam soluções mais tratáveis que facilitam a maximização da função de verossimilhança. Nesse contexto, uma alternativa bastante conveniente é o método de Laplace, apropriado para obter aproximações de integrais da forma

$$I = \int \exp[Q(\mathbf{b})] d\mathbf{b}$$

em que $Q(\mathbf{b})$ é uma função unimodal e limitada de uma variável q -dimensional, \mathbf{b} . O método é baseado na seguinte expansão de Taylor de segunda ordem de $Q(\mathbf{b})$ em torno do ponto $\hat{\mathbf{b}}$ que maximiza essa função

$$\begin{aligned} Q(\mathbf{b}) &\approx Q(\hat{\mathbf{b}}) + Q'(\hat{\mathbf{b}})^\top (\mathbf{b} - \hat{\mathbf{b}}) + \frac{1}{2} (\mathbf{b} - \hat{\mathbf{b}})^\top Q''(\hat{\mathbf{b}}) (\mathbf{b} - \hat{\mathbf{b}}) \\ &\approx Q(\hat{\mathbf{b}}) + \frac{1}{2} (\mathbf{b} - \hat{\mathbf{b}})^\top Q''(\hat{\mathbf{b}}) (\mathbf{b} - \hat{\mathbf{b}}) \end{aligned}$$

em que $Q'(\hat{\mathbf{b}}) = \partial Q(\mathbf{b}) / \partial \mathbf{b}|_{\mathbf{b}=\hat{\mathbf{b}}}$ e $Q''(\hat{\mathbf{b}}) = \partial^2 Q(\mathbf{b}) / \partial \mathbf{b} \partial \mathbf{b}^\top|_{\mathbf{b}=\hat{\mathbf{b}}}$.

Consequentemente, podemos escrever

$$\begin{aligned} I &= \int \exp[Q(\mathbf{b})] d\mathbf{b} \approx \exp[Q(\hat{\mathbf{b}})] \int -\frac{1}{2} \exp[(\mathbf{b} - \hat{\mathbf{b}})^\top [-Q''(\hat{\mathbf{b}})] (\mathbf{b} - \hat{\mathbf{b}})] d\mathbf{b} \\ &= \exp[Q(\hat{\mathbf{b}})] (2\pi)^{q/2} |Q''(\hat{\mathbf{b}})|^{-1/2} \end{aligned} \quad (3.2.8)$$

tendo em vista que o integrando corresponde ao núcleo de uma distribuição normal q -variada com vetor de médias nulo e matriz de covariâncias $[-Q''(\hat{\mathbf{b}})]^{-1}$.

No caso da função de verossimilhança (3.2.5), a integral correspondente à i -ésima unidade amostral pode ser escrita como

$$\int f(\mathbf{y}_i | \boldsymbol{\beta}, \mathbf{b}_i) \Phi[\mathbf{b}_i | \mathbf{0}, \mathbf{G}(\boldsymbol{\theta})] d\mathbf{b}_i = \int h(\mathbf{b}_i | \mathbf{y}_i, \boldsymbol{\beta}, \boldsymbol{\theta}) d\mathbf{b}_i = \int \exp\{\log[h(\mathbf{b}_i | \mathbf{y}_i, \boldsymbol{\beta}, \boldsymbol{\theta})]\} d\mathbf{b}_i.$$

Podemos substituir $\log[h(\mathbf{b}_i | \mathbf{y}_i, \boldsymbol{\beta}, \boldsymbol{\theta})]$ por sua expansão de Taylor de segunda ordem em torno de sua moda $\hat{\mathbf{b}}_i$, que pode ser calculada como solução da equação $\partial \log[h(\mathbf{b}_i | \mathbf{y}_i, \boldsymbol{\beta}, \boldsymbol{\theta})] / \partial \mathbf{b}_i = \mathbf{0}$, obtendo

$$f(\mathbf{y}_i | \boldsymbol{\beta}, \mathbf{b}_i) \Phi[\mathbf{b}_i | \mathbf{0}, \mathbf{G}(\boldsymbol{\theta})] = f(\mathbf{y}_i | \boldsymbol{\beta}, \hat{\mathbf{b}}_i) \Phi[\hat{\mathbf{b}}_i | \mathbf{0}, \mathbf{G}(\boldsymbol{\theta})] + \frac{1}{2} (\mathbf{b}_i - \hat{\mathbf{b}}_i)^\top [\mathbf{H}(\hat{\mathbf{b}}_i)] (\mathbf{b}_i - \hat{\mathbf{b}}_i)$$

em que $\mathbf{H}(\hat{\mathbf{b}}_i) = \partial^2 \log[h(\mathbf{b}_i | \mathbf{y}_i, \boldsymbol{\beta}, \boldsymbol{\theta})] / \partial \mathbf{b}_i \partial \mathbf{b}_i^\top$.

Então, por meio de (3.2.8) temos

$$L(\boldsymbol{\beta}, \boldsymbol{\theta}, \phi | \mathbf{y}) \approx \kappa \prod_{i=1}^n (2\pi)^q |[-\mathbf{H}(\hat{\mathbf{b}}_i)]|^{-1} f(\mathbf{y}_i | \boldsymbol{\beta}, \hat{\mathbf{b}}_i) \Phi[\hat{\mathbf{b}}_i | \mathbf{0}, \mathbf{G}(\boldsymbol{\theta})]$$

Convém lembrar que o cálculo da moda $\hat{\mathbf{b}}_i$ de $\log[h(\mathbf{b}_i|\mathbf{y}_i, \boldsymbol{\beta}, \boldsymbol{\theta})]$ depende dos valores de $\boldsymbol{\beta}$ e $\boldsymbol{\theta}$, que são desconhecidos. Uma solução para esse problema é utilizar os valores atualizados $\boldsymbol{\beta}^{(r)}$ e $\boldsymbol{\theta}^{(r)}$ no passo r do algoritmo de maximização.

Outra alternativa para a obtenção dos estimadores de máxima verossimilhança de $\boldsymbol{\beta}$ e $\boldsymbol{\theta}$ envolve os chamados métodos de **quase-verossimilhança**, a saber, **quase-verossimilhança penalizada** (PQL) e **quase-verossimilhança marginal** (MQL). Um ponto de partida para esses métodos é a decomposição dos dados como

$$y_{ij} = \mu_{ij} + e_{ij} = h(\mathbf{x}_{ij}^\top \boldsymbol{\beta} + \mathbf{z}_{ij}^\top \mathbf{b}_i) + e_{ij} \quad (3.2.9)$$

em que h é a inversa da função de ligação g e e_{ij} é um erro aleatório. No caso do modelo logístico (3.2.3), temos

$$h(\mathbf{x}_{ij}^\top \boldsymbol{\beta} + \mathbf{z}_{ij}^\top \mathbf{b}_i) = P(Y_{ij} = 1 | t_{ij}, b_i) = \pi_{ij} = \frac{\exp(\alpha + b_i + \beta t_{ij})}{1 + \exp(\alpha + b_i + \beta t_{ij})}$$

e o erro e_{ij} é igual a π_{ij} ou $1 - \pi_{ij}$ consoante y_{ij} seja igual a 1 ou 0.

A fundamentação do método PQL é baseada na expansão de Taylor de primeira ordem de (3.2.9) em torno da estimativa atual $\hat{\boldsymbol{\beta}}$ de $\boldsymbol{\beta}$ e do preditor atual $\hat{\mathbf{b}}_i$ do efeito aleatório \mathbf{b}_i , nomeadamente

$$\begin{aligned} y_{ij} &\approx h(\mathbf{x}_{ij}^\top \hat{\boldsymbol{\beta}} + \mathbf{z}_{ij}^\top \hat{\mathbf{b}}_i) + h'(\mathbf{x}_{ij}^\top \hat{\boldsymbol{\beta}} + \mathbf{z}_{ij}^\top \hat{\mathbf{b}}_i) \mathbf{x}_{ij}^\top (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) \\ &\quad + h'(\mathbf{x}_{ij}^\top \hat{\boldsymbol{\beta}} + \mathbf{z}_{ij}^\top \hat{\mathbf{b}}_i) \mathbf{z}_{ij}^\top (\mathbf{b}_i - \hat{\mathbf{b}}_i) + e_{ij} \\ &= \mu_{ij}^* + v(\mu_{ij}^*) \mathbf{x}_{ij}^\top (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) + v(\mu_{ij}^*) \mathbf{z}_{ij}^\top (\mathbf{b}_i - \hat{\mathbf{b}}_i) + e_{ij} \end{aligned}$$

em que $h'(w) = \partial h(w) / \partial w$, $\mu_{ij}^* = h(\mathbf{x}_{ij}^\top \hat{\boldsymbol{\beta}} + \mathbf{z}_{ij}^\top \hat{\mathbf{b}}_i)$ corresponde à média de y_{ij} calculada em termos de $\hat{\boldsymbol{\beta}}$ e $\hat{\mathbf{b}}_i$ e $v(\mu_{ij}^*) = h'(\mathbf{x}_{ij}^\top \hat{\boldsymbol{\beta}} + \mathbf{z}_{ij}^\top \hat{\mathbf{b}}_i)$ corresponde à variância aproximada de e_{ij} .

Primeiramente, organizemos as observações (y_{ij}) e os erros (e_{ij}) lexicograficamente em vetores \mathbf{y} e \mathbf{e} , respectivamente, e os vetores \mathbf{x}_{ij} e \mathbf{z}_{ij} em matrizes \mathbf{X} e \mathbf{Z} como em (2.1.7). Em seguida, consideremos o vetor $\boldsymbol{\mu}^*$ cujos elementos são os termos μ_{ij}^* , também organizados lexicograficamente e a matriz diagonal \mathbf{V}^* , cujos elementos não nulos são os termos $v(\mu_{ij}^*)$ com organização similar. Então podemos escrever

$$\mathbf{y} \approx \boldsymbol{\mu}^* + \mathbf{V}^* \mathbf{X} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) + \mathbf{V}^* \mathbf{Z} (\mathbf{b} - \hat{\mathbf{b}}) + \mathbf{e}.$$

Subtraindo $\boldsymbol{\mu}^*$ de ambos os membros e multiplicando-os à esquerda por $[\mathbf{V}^*]^{-1}$, obtemos

$$[\mathbf{V}^*]^{-1} (\mathbf{y} - \boldsymbol{\mu}^*) \approx \mathbf{X} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) + \mathbf{Z} (\mathbf{b} - \hat{\mathbf{b}}) + [\mathbf{V}^*]^{-1} \mathbf{e}$$

e, consequentemente,

$$\mathbf{y}^* = [\mathbf{V}^*]^{-1} (\mathbf{y} - \boldsymbol{\mu}^*) + \mathbf{X} \hat{\boldsymbol{\beta}} + \mathbf{Z} \hat{\mathbf{b}} + [\mathbf{V}^*]^{-1} \mathbf{e} \approx \mathbf{X} \boldsymbol{\beta} + \mathbf{Z} \mathbf{b} + \mathbf{e}^* \quad (3.2.10)$$

com $\mathbf{e}^* = [\mathbf{V}^*]^{-1}\mathbf{e}$.

A expressão (3.2.10) tem a forma de (2.1.7) e pode ser encarada como um modelo linear misto tendo as “pseudo-variáveis” \mathbf{y}^* como respostas. Notemos que $\mathbb{V}(\mathbf{e}) = \mathbf{V}^*$, o que implica $\mathbb{V}(\mathbf{e}^*) = [\mathbf{V}^*]^{-1}\mathbf{V}^*[\mathbf{V}^*]^{-1} = [\mathbf{V}^*]^{-1}$ e $\mathbb{V}(\mathbf{y}^*) = \mathbf{Z}\mathbf{G}(\boldsymbol{\theta})\mathbf{Z}^\top + [\mathbf{V}^*]^{-1}$. Esse contexto sugere o seguinte algoritmo para ajustar o modelo original:

Passo 1 : Inicie o processo com valores para $\boldsymbol{\beta}$ e $\boldsymbol{\theta}$.

Passo 2 : Obtenha preditores de \mathbf{b}_i com base nos valores correntes de $\boldsymbol{\beta}$ e $\boldsymbol{\theta}$ e calcule os valores das pseudo-variáveis \mathbf{y}^* .

Passo 3 : Reajuste o modelo linear misto (3.2.10) obtendo novos valores para $\boldsymbol{\beta}$ e $\boldsymbol{\theta}$.

Passo 4 : Repita os passos 2 e 3 até que algum critério de convergência seja atingido.

Os estimadores resultantes desse processo são conhecidos como **estimadores de quase-verossimilhança penalizados** pois dependem apenas dos dois primeiros momentos condicionais penalizados relativamente aos efeitos aleatórios.

O princípio que norteia o método de máxima quase-verossimilhança marginal (MQL) é essencialmente o mesmo que fundamenta o método PQL; neste caso, considera-se uma expansão de Taylor em torno do valor corrente de $\hat{\boldsymbol{\beta}}$ e do valor esperado nulo dos efeitos aleatórios, ou seja

$$\begin{aligned} y_{ij} &\approx h(\mathbf{x}_{ij}^\top \hat{\boldsymbol{\beta}}) + h'(\mathbf{x}_{ij}^\top \hat{\boldsymbol{\beta}}) \mathbf{x}_{ij}^\top (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) + h'(\mathbf{x}_{ij}^\top \hat{\boldsymbol{\beta}}) \mathbf{z}_{ij}^\top \mathbf{b}_i + e_{ij} \\ &= \mu_{ij}^\bullet + v(\mu_{ij}^\bullet) \mathbf{x}_{ij}^\top (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) + v(\mu_{ij}^\bullet) \mathbf{z}_{ij}^\top \mathbf{b}_i + e_{ij} \end{aligned}$$

em que $\mu_{ij}^\bullet = h(\mathbf{x}_{ij}^\top \hat{\boldsymbol{\beta}})$ e $v(\mu_{ij}^\bullet) = h'(\mathbf{x}_{ij}^\top \hat{\boldsymbol{\beta}})$. Organizando os componentes desse modelo de maneira similar àquela que consideramos no caso PQL, obtemos

$$\mathbf{y}^\bullet = [\mathbf{V}^\bullet]^{-1}(\mathbf{y} - \boldsymbol{\mu}^\bullet) + \mathbf{X}\hat{\boldsymbol{\beta}} \approx \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \mathbf{e}^\bullet \quad (3.2.11)$$

em que \mathbf{y}^\bullet representa o vetor de pseudo-observações correspondente e $\mathbf{e}^\bullet = [\mathbf{V}^\bullet]^{-1}\mathbf{e}$.

Um algoritmo similar àquele utilizado para obtenção dos estimadores PQL pode ser construído no caso MQL. Os estimadores resultantes são erroneamente conhecidos como estimadores de quase-verossimilhança marginal por que são calculados supondo-se que os efeitos aleatórios são nulos. Embora em muitos casos os métodos PQL e MQL produzam resultados semelhantes, estimadores MQL são, em geral, mais enviesados do que os estimadores PQL correspondentes.

Outros métodos, como aqueles baseados em expansões de Taylor de maior ordem (PQL2, MQL2) podem ser considerados para a maximização da verossimilhança (3.2.5). Detalhes podem ser encontrados em Tuerlinckx et al. (2006).

Com o objetivo de desenvolver ferramentas de diagnóstico, Xu, Lee & Poon (2006) consideram um enfoque baseado no conceito de dados omissos em que \mathbf{y} corresponde aos dados observados e \mathbf{b}_i aos dados faltantes, de forma que a função log-verossimilhança para os dados “completos”, $\mathbf{y}_C = (\mathbf{y}^\top, \mathbf{b}^\top)^\top$, pode ser expressa como

$$l(\boldsymbol{\beta}, \boldsymbol{\theta}, \phi | \mathbf{y}_C) = \sum_{i=1}^n \left(\sum_{j=1}^{m_i} \{ \phi[y_{ij}\vartheta_{ij} - a(\vartheta_{ij})] + c(y_{ij}, \phi) \} - \frac{1}{2} \mathbf{b}_i^\top [\mathbf{G}(\boldsymbol{\theta})]^{-1} \mathbf{b}_i - \frac{1}{2} |\mathbf{G}(\boldsymbol{\theta})|^{-1/2} \right).$$

Esses autores sugerem um algoritmo que requer a geração de uma amostra por meio de técnicas MCMC a partir da distribuição de $\mathbf{b} | (\mathbf{y}, \boldsymbol{\beta}, \boldsymbol{\theta}, \phi)$ *i.e.*, da distribuição condicional de \mathbf{b} dados os valores correntes dos parâmetros $(\boldsymbol{\beta}, \boldsymbol{\theta}, \phi)$ e dos dados observados \mathbf{y} , de forma similar àquela do passo E do algoritmo EM. No passo final, os ingredientes necessários para a construção de ferramentas diagnósticas baseadas na omissão de casos são obtidos, tanto para unidades amostrais, quanto para observações intraunidades amostrais.

3.3 Modelos baseados em equações de estimação generalizadas (GEE)

Modelos baseados em GEE focam diretamente a distribuição marginal dos dados \mathbf{y} , sem referência aos efeitos aleatórios, com a especificação apenas do valor esperado e da variância, a saber

$$\mathbb{E}(y_{ij}) = \mu_{ij} \text{ e } \mathbb{V}(y_{ij}) = \phi^{-1} \nu(\mu_{ij}) \quad (3.3.1)$$

em que $\nu(\mu_{ij})$ é uma função do valor esperado e ϕ é um parâmetro de escala. A relação entre a resposta esperada e as variáveis exploratórias é especificada como

$$g(\mu_{ij}) = \eta_{ij} = \mathbf{x}_{ij}^\top \boldsymbol{\beta} \quad (3.3.2)$$

em que g é uma função de ligação, possivelmente não linear. Para dados não correlacionados, estimadores dos parâmetros de localização $\boldsymbol{\beta}$ são obtidos por meio das equações de estimação generalizadas

$$\sum_{i=1}^n \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}^\top} \mathbf{A}_i^{-1} [\mathbf{y}_i - \boldsymbol{\mu}_i(\mathbf{X}_i^\top \boldsymbol{\beta})] = \sum_{i=1}^n \mathbf{X}_i^\top \boldsymbol{\Delta}_i \mathbf{A}_i^{-1} [\mathbf{y}_i - \boldsymbol{\mu}_i(\mathbf{X}_i^\top \boldsymbol{\beta})] = \mathbf{0}$$

em que $\boldsymbol{\mu}_i = (\mu_{i1}, \dots, \mu_{im_i})^\top$, $\mathbf{A}_i = \oplus_{j=1}^{m_i} \nu(\mu_{ij})$ e $\boldsymbol{\Delta}_i = \oplus_{j=1}^{m_i} [\partial \mu_{ij} / \partial \eta_{ij}]$.

Com a finalidade de incorporar a estrutura de covariância intraunidades amostrais, consideramos uma **matriz de covariâncias de trabalho** (*working covariance matrix*) definida como

$$\Omega_{W_i}(\boldsymbol{\theta}) = \phi \mathbf{A}_i^{1/2} \mathbf{R}_W(\boldsymbol{\theta}) \mathbf{A}_i^{1/2} \quad (3.3.3)$$

em que $\mathbf{R}_W(\boldsymbol{\theta})$ é uma matriz definida positiva dependente de parâmetros $\boldsymbol{\theta}$. Se $\mathbf{R}_W(\boldsymbol{\theta})$ for a verdadeira matriz de correlações intraunidades amostrais de \mathbf{y}_i , então $\Omega_{W_i}(\boldsymbol{\theta}) = \mathbb{V}(\mathbf{y}_i)$.

Estimadores dos parâmetros de regressão são obtidos como solução para as equações de estimação generalizadas

$$\sum_{i=1}^n \mathbf{X}_i^\top \Delta_i [\Omega_{W_i}(\boldsymbol{\theta})]^{-1} [\mathbf{y}_i - \boldsymbol{\mu}_i(\mathbf{X}_i^\top \boldsymbol{\beta})] = \mathbf{0}. \quad (3.3.4)$$

Para resolvê-las, usualmente utiliza-se o método de momentos para $\boldsymbol{\theta}$ e ϕ atualizados após cada iteração de

$$\begin{aligned} \boldsymbol{\beta}^{(h+1)} = \boldsymbol{\beta}^{(h)} + & \left[\sum_{i=1}^n \mathbf{X}_i^\top \Delta_i^{(h)} [\Omega_{W_i}(\boldsymbol{\theta}^{(h)})]^{-1} \Delta_i^{(h)} \mathbf{X}_i \right]^{-1} \times \\ & \left[\sum_{i=1}^n \mathbf{X}_i^\top \Delta_i^{(h)} [\Omega_{W_i}(\boldsymbol{\theta}^{(h)})]^{-1} [\mathbf{y}_i - \boldsymbol{\mu}_i(\mathbf{X}_i^\top \boldsymbol{\beta}^{(h)})] \right] \end{aligned} \quad (3.3.5)$$

$h = 0, 1, \dots$, com $\boldsymbol{\beta}^{(0)}$ denotando uma estimativa inicial conveniente de $\boldsymbol{\beta}$.

Sob certas condições de regularidade, a solução $\hat{\boldsymbol{\beta}}$ de (3.3.4) é tal que

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \rightarrow N_p(\mathbf{0}, \mathbf{A}^{-1})$$

em que

$$\mathbf{A}^{-1} = \lim_{n \rightarrow \infty} n \mathbf{S}^{-1} \left[\sum_{i=1}^n \mathbf{X}_i^\top \Delta_i [\Omega_{W_i}(\boldsymbol{\theta})]^{-1} \mathbb{V}(\mathbf{y}_i) [\Omega_{W_i}(\boldsymbol{\theta})]^{-1} \Delta_i \mathbf{X}_i \right] \mathbf{S}^{-1}. \quad (3.3.6)$$

com $\mathbf{S} = \left[\sum_{i=1}^n \mathbf{X}_i^\top \Delta_i [\Omega_{W_i}(\boldsymbol{\theta})]^{-1} \Delta_i \mathbf{X}_i \right]$. Em aplicações práticas, os termos desconhecidos são substituídos por estimativas convenientes. Como a matriz de covariâncias de trabalho é escolhida arbitrariamente, convém substituir $\mathbb{V}(\mathbf{y}_i)$ em (3.3.6) por $[\mathbf{y}_i - \boldsymbol{\mu}_i(\mathbf{X}_i^\top \hat{\boldsymbol{\beta}})][\mathbf{y}_i - \boldsymbol{\mu}_i(\mathbf{X}_i^\top \hat{\boldsymbol{\beta}})]^\top$, gerando um estimador mais robusto, *i.e.*, o **estimador sanduíche** (*sandwich estimator*). Quando $\mathbf{R}_W(\boldsymbol{\theta})$ estiver corretamente especificada, (3.3.6) fica reduzida a \mathbf{S}^{-1} de forma que $\|\mathbf{A}^{-1} - \mathbf{S}^{-1}\|$ pode ser utilizada como uma medida de adequação da matriz de correlações de trabalho.

Venezuela, Botter & Sandoval (2007) desenvolvem uma série de ferramentas diagnósticas nesse contexto. Com essa finalidade, eles observam inicialmente que utilizando o **algoritmo de mínimos quadrados iterativamente ponderados** (*iteratively weighted least squares algorithm*) com pseudo-observações

$$\tilde{\mathbf{y}}_i = \mathbf{g}[\boldsymbol{\mu}_i(\mathbf{X}_i^\top \hat{\boldsymbol{\beta}})] + \hat{\boldsymbol{\Delta}}_i^{-1}[\mathbf{y}_i - \boldsymbol{\mu}_i(\mathbf{X}_i^\top \hat{\boldsymbol{\beta}})]$$

e uma matriz de ponderação $\mathbf{W}_i(\hat{\boldsymbol{\theta}}) = \hat{\boldsymbol{\Delta}}_i \boldsymbol{\Omega}_{W_i}(\hat{\boldsymbol{\theta}}) \hat{\boldsymbol{\Delta}}_i$, (3.3.5) fica reduzida a

$$\boldsymbol{\beta}^{(h+1)} = \left[\sum_{i=1}^n \mathbf{X}_i^\top \mathbf{W}_i(\hat{\boldsymbol{\theta}}^{(h)}) \mathbf{X}_i \right]^{-1} \left[\sum_{i=1}^n \mathbf{X}_i^\top \mathbf{W}_i(\hat{\boldsymbol{\theta}}^{(h)}) \tilde{\mathbf{y}}_i^{(h)} \right]. \quad (3.3.7)$$

Quando a convergência é atingida, temos

$$\hat{\boldsymbol{\beta}} = [\mathbf{X}^\top \mathbf{W}(\hat{\boldsymbol{\theta}}) \mathbf{X}]^{-1} \mathbf{X}^\top \mathbf{W}(\hat{\boldsymbol{\theta}}) \tilde{\mathbf{y}}$$

em que $\mathbf{W}(\hat{\boldsymbol{\theta}}) = \oplus_{i=1}^n \mathbf{W}_i(\hat{\boldsymbol{\theta}})$, $\mathbf{X} = (\mathbf{X}_1^\top, \dots, \mathbf{X}_n^\top)^\top$ e $\tilde{\mathbf{y}} = (\tilde{\mathbf{y}}_1^\top, \dots, \tilde{\mathbf{y}}_n^\top)^\top$.

Neste caso, os resíduos são definidos como

$$\tilde{\mathbf{r}} = \mathbf{W}(\hat{\boldsymbol{\theta}})^{1/2} \hat{\boldsymbol{\Delta}}^{-1}[\mathbf{y} - \boldsymbol{\mu}(\mathbf{X}^\top \hat{\boldsymbol{\beta}})]$$

em que $\hat{\boldsymbol{\Delta}} = \oplus_{i=1}^n \hat{\boldsymbol{\Delta}}_i$ e $\boldsymbol{\mu}(\mathbf{X}^\top \hat{\boldsymbol{\beta}}) = [\boldsymbol{\mu}_1(\mathbf{X}_1^\top \hat{\boldsymbol{\beta}})^\top, \dots, \boldsymbol{\mu}_n(\mathbf{X}_n^\top \hat{\boldsymbol{\beta}})^\top]^\top$. Essas autoras mostram que $\mathbb{V}(\tilde{\mathbf{r}}) \approx (\mathbf{I}_N - \mathbf{H})$ em que $\mathbf{H} = \oplus_{i=1}^n \mathbf{H}_i$ com

$$\mathbf{H}_i = \mathbf{W}_i(\hat{\boldsymbol{\theta}})^{1/2} \mathbf{X}_i [\mathbf{X}_i^\top \mathbf{W}_i(\hat{\boldsymbol{\theta}}) \mathbf{X}_i]^{-1} \mathbf{X}_i^\top \mathbf{W}_i(\hat{\boldsymbol{\theta}})^{1/2}$$

faz o papel da matriz chapéu correspondente ao modelo linear clássico. Desta forma, valores grandes dos elementos da diagonal de H_{kk} ($k = 1, \dots, N$) podem ser empregados para identificação de observações com alto grau de alavanca, segundo critérios similares àqueles utilizados no caso dos modelos lineares mistos.

Os resíduos padronizados

$$\tilde{\mathbf{r}}_{Sij} = (1 - H_{i(jj)})^{-1/2} \mathbf{d}_{ij}^\top \mathbf{W}(\hat{\boldsymbol{\theta}})^{1/2} \hat{\boldsymbol{\Delta}}^{-1}[\mathbf{y} - \boldsymbol{\mu}(\mathbf{X}^\top \hat{\boldsymbol{\beta}})]$$

em que \mathbf{d}_{ij} denota um vetor com elementos nulos à exceção daquele na posição (ij) (sob uma ordem lexicográfica) que é igual a 1 e $H_{i(jj)}$ representa o j -ésimo elemento da diagonal principal de \mathbf{H}_i , podem ser utilizados para identificar possíveis observações discrepantes.

Além disso, Venezuela et al. (2007) sugerem que uma estimativa de $\boldsymbol{\beta}$ obtida quando a observação (ij) é omitida pode ser aproximada por

$$\hat{\boldsymbol{\beta}}_{(ij)} = \hat{\boldsymbol{\beta}} - \frac{1}{1 - H_{i(jj)}} [\mathbf{X}^\top \mathbf{W}(\hat{\boldsymbol{\theta}}) \mathbf{X}]^{-1} [\mathbf{X}^\top \mathbf{W}(\hat{\boldsymbol{\theta}})^{1/2} \mathbf{d}_{ij}] \{ \mathbf{d}_{ij}^\top \mathbf{W}(\hat{\boldsymbol{\theta}})^{1/2} \hat{\boldsymbol{\Delta}}^{-1} [\mathbf{y} - \boldsymbol{\mu}(\mathbf{X}^\top \hat{\boldsymbol{\beta}})] \}$$

e consideram a utilização da distância de Cook

$$D_{(ij)} = \frac{1}{p}(\hat{\beta} - \hat{\beta}_{(ij)})^\top [\mathbf{X}^\top \mathbf{W}(\hat{\theta}) \mathbf{X}] (\hat{\beta} - \hat{\beta}_{(ij)}) = (\tilde{\mathbf{r}}_{sij})^2 \frac{H_{i(jj)}}{p(1 - H_{i(jj)})}$$

para avaliar a influência de cada observação intraunidades amostrais nos coeficientes de regressão. Se o interesse recai na influência de uma unidade amostral, elas sugerem a utilização de $D_{(i)} = \sum_{j=1}^{m_i} D_{(ij)}$.

Finalmente, Venezuela et al. (2007) propõem um algoritmo para gerar gráficos QQ meio-normais com envelopes simulados, úteis para avaliar a adequação do ajuste do modelo.

Recentemente, Venezuela, Sandoval & Botter (2011) consideram uma generalização dos resultados de Cadigan & Farrell (2002) e propõem ferramentas para diagnóstico de influência local para uma classe mais ampla de modelos de regressão que inclui aqueles para os quais a **função de ajuste** (*fit function*), $\mathcal{F}(\beta)$, *e.g.*, as funções de verossimilhança ou quase-verossimilhança, não é especificada. Particularmente, elas consideram modelos da forma (3.3.1)-(3.3.2) para os quais as funções de estimação generalizadas são

$$\Psi(\beta, \theta) = \frac{\partial \mathcal{F}(\beta)}{\partial \beta} = \mathbf{X}^\top \Delta^\top \Omega_W(\theta)^{-1} \mathbf{b}(\mathbf{y}, \beta) = \mathbf{X}^\top \mathbf{W}(\hat{\theta}) \Delta^{-1} \mathbf{b}(\mathbf{y}, \beta) = \mathbf{0} \quad (3.3.8)$$

em que $\mathbf{b}(\mathbf{y}, \beta) = [\mathbf{b}_1(\mathbf{y}_1, \beta)^\top, \dots, \mathbf{b}_n(\mathbf{y}_n, \beta)^\top]^\top$ com

$$\mathbf{b}_i(\mathbf{y}_i, \beta) = [b_{i1}(y_{i1}, \beta), \dots, b_{im_i}(y_{im_i}, \beta)]^\top$$

denotando um vetor de variáveis aleatórias com valor esperado nulo, $\Delta = \oplus_{i=1}^n \Delta_i$, $\Delta_i = \mathbb{E}[\partial \mathbf{b}_i(\mathbf{y}_i, \beta) / \partial \boldsymbol{\eta}_i^\top]$, $\Omega_W(\theta) = \oplus_{i=1}^n \Omega_i(\theta)$, $\Omega_i(\theta) = \mathbf{A}_i^{1/2} \mathbf{R}_W(\theta) \mathbf{A}_i^{1/2}$, $\mathbf{A}_i = \oplus_{j=1}^{m_i} \mathbb{V}[b_{ij}(y_{ij}, \beta)]$, $\mathbf{W} = \Delta \Omega_W(\theta)^{-1} \Delta$ e $\mathbf{R}_W(\theta)$ é uma matriz de correlações de trabalho. Notemos que (3.3.4) é um caso especial de (3.3.8) com $\mathbf{b}(\mathbf{y}, \beta) = [\mathbf{y} - \boldsymbol{\mu}(\mathbf{X}\beta)]$.

Nesse contexto, Venezuela et al. (2011) consideram o comportamento do **deslocamento da função de ajuste** (*fit function displacement*), $2[\mathcal{F}(\hat{\beta}) - \mathcal{F}(\hat{\beta}_\omega)]$ em que ω é um vetor de perturbações com o mesmo espírito de (2.5.15), que, de fato, constitui um caso especial em que a função de ajuste é a função de verossimilhança. Elas consideram esquemas de perturbação de casos, da variável resposta, do parâmetro de precisão e da matriz de correlações de trabalho.

3.4 Modelos para dados categorizados

3.5 Modelos não paramétricos para análise de perfis

Estudos com um ou mais fatores interunidades amostrais são usualmente analisados por meio de MANOVA ou ANOVA para medidas repetidas (ver Seção 2.4) quando a distribuição subjacente é normal. Nos casos em que essa suposição não é sustentável ou é questionável, técnicas não paramétricas baseadas no conceito de efeito relativo de tratamentos proposto por Brunner, Munzel & Puri (1999) podem ser utilizadas. Essas técnicas produzem resultados análogos àqueles obtidos por meio de ANOVA para medidas repetidas e podem ser utilizadas mesmo quando a forma da distribuição da resposta não é conhecida. Além disso, não necessitam que a estrutura de covariância intraunidades amostrais seja especificada e permitem a análise de variáveis respostas contínuas, discretas ou qualitativas ordinais.

Neste trabalho, consideramos o modelo proposto por Brunner et al. (2002) explicitando a análise para estudos com medidas repetidas e descrevemos uma sub-rotina programada em R que pode ser utilizada para analisar estudos com mais que dois fatores entre unidades amostrais cruzados com celas estruturalmente vazias ou não. Ilustramos o procedimento por meio de exemplos práticos.

3.5.1 O modelo

Consideramos inicialmente um estudo com medidas repetidas com um fator (fixo) interunidades amostrais e denotemos o vetor das observações associadas à j -ésima unidade amostral submetida ao i -ésimo nível desse fator por $\mathbf{X}_{ij} = (X_{ij1}, \dots, X_{ijt})^\top$, em que $i = 1, \dots, g$, $j = 1, \dots, n_i$ e $k = 1, \dots, t$. Supomos que a função distribuição marginal de cada variável X_{ijk} , nomeadamente, F_{ik} é a mesma para todas as observações obtidas sob as combinações dos níveis dos dois fatores (inter e intraunidades amostrais), *i.e.*, para todo j , com i e k fixos. Nos casos em que a função distribuição de X_{ijk} não é contínua, consideramos a versão normalizada

$$F_{ik}(x) = \frac{1}{2}[F_{ik}^+(x) + F_{ik}^-(x)] \quad (3.5.1)$$

em que $F_{ik}^+(x) = P(X_{ijk} \leq x)$ e $F_{ik}^- = P(X_{ijk} < x)$.

O efeito relativo do tratamento correspondente à combinação do nível i do fator interunidades amostrais e nível k do fator intraunidades amostrais, p_{ik} , com $i =$

$1, \dots, g$ e $k = 1, \dots, t$ é definido como

$$p_{ik} = \int H dF_{ik}$$

em que $H(x) = N^{-1} \sum_{i=1}^g \sum_{k=1}^t n_i F_{ik}(x)$ com $N = \sum_{i=1}^g n_i$. Os valores de $p_{ik} < \frac{1}{2}$ ($p_{ik} > \frac{1}{2}$) indicam que as respostas das unidades amostrais correspondentes ao nível i do fator interunidades amostrais e nível k do fator intraunidades amostrais tendem a assumir valores menores (maiores) que a resposta média correspondente a todos os tratamentos. Um estimador consistente de p_{ik} [ver Brunner, Domhof e Langer (2002)] é

$$\hat{p}_{ik} = \frac{1}{N} (\bar{R}_{i.k} - \frac{1}{2})$$

em que $\bar{R}_{i.k} = n_i^{-1} \sum_{j=1}^{n_i} R_{ijk}$ é a média aritmética dos postos no i -ésimo nível do fator interunidades amostrais e k -ésimo do fator intraunidades amostrais.

Similarmente, podemos definir o efeito relativo marginal de tratamento como

$$p_i = \int H dF_i,$$

em que $H(x) = N^{-1} \sum_{i=1}^g n_i F_i(x)$, $N = \sum_{i=1}^g n_i$ e os valores de $p_i < \frac{1}{2}$ ($p_i > \frac{1}{2}$) indicam que as respostas das unidades amostrais correspondentes ao nível i do fator interunidades amostrais tendem a assumir valores menores (maiores) que a resposta média correspondente a todos os tratamentos. Os efeitos relativos p_i podem ser estimados por

$$\hat{p}_i = \int \hat{H} d\hat{F}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \hat{H}(X_{ij}) = \frac{1}{N} (\bar{R}_i - \frac{1}{2})$$

com $\bar{R}_i = n_i^{-1} \sum_{j=1}^{n_i} R_{ij}$ denotando a média aritmética dos postos no i -ésimo nível do fator interunidades amostrais e R_{ij} é o posto de X_{ij} entre todas as observações. Similarmente, o efeito relativo marginal p_{ik} correspondente ao i -ésimo nível do fator interunidades amostrais no k -ésimo nível do fator intraunidades amostrais, com $i = 1, \dots, g$ e $k = 1, \dots, t$ é definido como $p_{ik} = \int H dF_{ik}$, em que $H(x) = N^{-1} \sum_{i=1}^g \sum_{k=1}^t n_i F_{ik}(x)$. Os valores de $p_{ik} < \frac{1}{2}$ ($p_{ik} > \frac{1}{2}$) indicam que as respostas das unidades amostrais correspondentes ao nível i do fator interunidades amostrais e nível k do fator intraunidades amostrais tendem a assumir valores menores (maiores) que a resposta média correspondente a todos os tratamentos no instante k .

As hipóteses de interesse, nomeadamente, de inexistência de interação e de efeitos principais dos fatores, além de outras, podem ser expressas na forma

$$\mathbf{CF} = \mathbf{0} \tag{3.5.2}$$

ou equivalentemente $\mathbf{C}\mathbf{p} = \mathbf{0}$, em que \mathbf{C} é uma matriz de constantes. Com a finalidade de testar essas hipóteses, Brunner e Langer (2000) propuseram a seguinte estatística de Wald

$$\mathbf{Q}_W(\mathbf{C}) = n\hat{\mathbf{p}}^\top \mathbf{C}^\top [C\hat{\mathbf{V}}_n \mathbf{C}^\top]^{-1} \mathbf{C}\hat{\mathbf{p}} \quad (3.5.3)$$

em que $\hat{\mathbf{V}}_n$ é um estimador consistente da matriz de covariância de \mathbf{e} e $n = \sum_{i=1}^g n_i$. Sob a hipótese nula, \mathbf{Q}_W tem distribuição assintótica χ_f^2 central com $f = \text{posto}(\mathbf{C})$ graus de liberdade. Estudos de simulação mostraram que a distribuição da estatística \mathbf{Q}_W converge vagarosamente para a distribuição Qui-quadrado e, conseqüentemente que os níveis descritivos para os testes são bastante liberais. Alternativamente, Brunner e Langer (2000) também consideram a seguinte estatística tipo ANOVA

$$\mathbf{Q}_A(\mathbf{C}) = \frac{n}{\text{tr}(\mathbf{T}\hat{\mathbf{V}}_n)} \hat{\mathbf{p}}^\top \mathbf{T}\hat{\mathbf{p}} \quad (3.5.4)$$

em que $\mathbf{T} = \mathbf{C}^\top [\mathbf{C}\mathbf{C}^\top]^{-1} \mathbf{C}$ e $\text{tr}()$. Sob a hipótese nula, a estatística \mathbf{Q}_A tem distribuição aproximada $\chi_{\hat{f}}^2$, com

$$\hat{f} = \frac{[\text{tr}(\mathbf{T}\hat{\mathbf{V}}_n)]^2}{\text{tr}(\mathbf{T}\hat{\mathbf{V}}_n)} \mathbf{T}\hat{\mathbf{V}}_n \quad (3.5.5)$$

para $\text{tr}(\mathbf{T}\hat{\mathbf{V}}_n) > 0$. A convergência dessa estatística para sua distribuição assintótica é mais rápida do que a da estatística \mathbf{Q}_W . Esse resultado é útil quando o tamanho amostral é pequeno. Para mais detalhes sobre o uso dessas estatísticas ver Brunner, Domhof e Langer (2002).

Para analisar dados com medidas repetidas com um fator entre e outro intraunidades amostrais e testar as hipóteses do tipo (2) construímos uma sub-rotina (*ANOVA.NPar*) escrita em linguagem *R* (*R Development Core Team, 2012*) baseada em macros escritas para o pacote *SAS* disponibilizadas em <http://www.ams.med.uni-goettingen.de/amsneu/sasmakr-de.shtml>, em sub-rotinas *Excel* disponibilizadas em <http://www.ime.usp.br/~jmsinger/Medidas%20repetidas%20NP.zip> e na função *nparLD* do *software R*.

A sub-rotina *ANOVA.NPar* produz gráficos de perfis individuais para cada nível do fator interunidades amostrais, gráfico de perfis médios e gráfico de intervalos de confiança, com 95% de confiança, para os efeitos relativos de tratamento, uma tabela com estimativas dos efeitos relativos de tratamento e limites inferiores e superiores de intervalos de confiança, com coeficiente de confiança de 95% além de tabelas com os resultados dos testes usuais (efeitos de interação, de grupo e de tempo). Quando outras hipóteses são especificadas pelo usuário a sub-rotina gera resultados similares àqueles produzidos no caso usual. Os gráficos são dispostos em janelas distintas.

Apresentamos abaixo a estrutura necessária para a disposição dos dados a serem analisados.

Tabela 3.5.1: Disposição de dados para análise por meio da sub-rotina *ANOVA.NPar*

Grupo	Unidade Amostral	Tempo	Resposta
1	1	1	X_{111}
1	1	2	X_{112}
\vdots	\vdots	\vdots	\vdots
1	1	t	X_{11t}
1	2	1	X_{121}
1	2	2	X_{122}
\vdots	\vdots	\vdots	\vdots
1	2	t	X_{12t}
\vdots	\vdots	\vdots	\vdots
1	n_1	1	X_{1n_11}
1	n_1	2	X_{1n_12}
\vdots	\vdots	\vdots	\vdots
1	n_1	t	X_{1n_1t}
\vdots	\vdots	\vdots	\vdots
g	1	1	X_{g11}
g	1	2	X_{g12}
\vdots	\vdots	\vdots	\vdots
g	1	t	X_{g1t}
g	2	1	X_{g21}
g	2	2	X_{g22}
\vdots	\vdots	\vdots	\vdots
g	2	t	X_{g2t}
\vdots	\vdots	\vdots	\vdots
g	n_g	1	X_{gn_g1}
g	n_g	2	X_{gn_g2}
\vdots	\vdots	\vdots	\vdots
g	n_g	t	X_{gn_gt}

Para utilizar a sub-rotina, os dados devem estar disponíveis usando o comando `attach()`.

Argumentos obrigatórios:

Os argumentos serão apresentados usando o exemplo de uma estrutura similar a Tabela 1, após o uso do comando `attach()`.

var: vetor contendo os valores numéricos da variável resposta. Os valores omisso (*missing*) devem ser indicados por `NA`, exceto para os dados faltantes devido à cela vazia, que devem ser excluídos. No exemplo, `var = Resposta`.

time: vetor com os valores dos níveis do fator intraunidades amostrais. No exemplo,

time = Tempo.

subject: vetor com os valores da variável que identifica as unidades amostrais. No exemplo, **subject** = Unidade Amostral.

group: no caso de o estudo conter apenas o fator intraunidades amostrais, não é preciso declarar o vetor de valores correspondentes ao fator entre unidades amostrais. No caso de estudos ter fatores entre e intraunidades amostrais (com ou sem cela vazia) o parâmetro **group** recebe o vetor com os valores do fator entre unidades amostrais. No exemplo, **group** = Grupo. Se o estudo tiver dois ou mais fatores entre unidades amostrais deve-se criar uma nova variável cujos níveis sejam a combinação dos níveis dos fatores. O *default* desse parâmetro é NULL.

time.order: vetor com os níveis do fator intraunidades amostrais ordenados. O formato (numérico ou não) dos níveis declarados deve ser igual ao formato em **time**. No exemplo, se no vetor **time** os níveis forem declarados como nomes, então **time.order** = `c("1","2","3", ...,"t")`, caso contrário **time.order** = `c(1,2,3,...,t)`. O *default* desse parâmetro é NULL.

emptyCell: parâmetro usado apenas se o estudo possuir cela(s) vazia(s), sendo necessário declará-la(s). Como exemplo, suponha que em um estudo com 3 grupos e 5 tempos ocorram celas vazias no grupo 2, tempo 4 e no grupo 3, tempo 2, então as celas vazias seriam declaradas **emptyCell** = `rbind(c("G2","T4"),c("G3","T2"))`. O *default* desse parâmetro é NULL.

Argumentos opcionais:

Para os parâmetros opcionais, rótulo da variável resposta (**var.name**), rótulo dos níveis do fator intraunidades amostrais (**time.name**) e rótulo dos níveis do fator entre unidades amostrais (**group.name**) o usuário pode declarar o rótulo dessas variáveis que aparecerão nos resultados da sub-rotina, o *default* é NULL.

group.order: vetor com os níveis do fator entre unidades amostrais ordenados. O formato (numéricos ou não) dos níveis deve ser igual ao formato declarados em **group**. No exemplo, se no vetor **group** os níveis do fator entre unidades amostrais forem declarados como nomes, então **group.order** = `c("1","2","3",...,"g")`, caso contrário **group.order** = `c(1,2,3,...,g)`. O *default* desse parâmetro é NULL.

contrast: parâmetro obrigatório em estudos com celas vazias; nele o usuário declara a(s) matriz(es) de contraste(s) da(s) hipótese(s) de interesse. Por exemplo, suponha que o usuário queira testar h hipóteses. Inicialmente são montadas as h matrizes de contrastes, nomeando-as, por exemplo, C_1, C_2, \dots, C_h e declarando-as em uma lista usando o comando `list()` da forma **contrast** = `list(C1,C2,...,Ch)`. O *default* desse parâmetro NULL.

rounds: parâmetro que define quantas casas decimais serão apresentadas nos resultados das tabelas, o *default* é com 4 casas decimais.

alpha: parâmetro que define o nível de significância adotado, o *default* é de 0.05.

Saídas geradas pela sub-rotina *ANOVA.NPar*:

summary: tabela resumo contendo os níveis do fator entre unidades amostrais (*Group*), os níveis fator intraunidades amostrais (*Time*), quantidade de unidades amostrais (*Nobs*), médias dos postos (*RankMeans*), estimativa do efeito relativo de tratamento (*RTE*), estimativa da variância (*Variance*), e os limites inferior e superior do IC(95%) dos efeitos relativos de tratamentos (*Lower* e *Upper*).

out: resultados das estatísticas de Wald e tipo ANOVA para os testes usuais (efeitos de interação, de grupo e tempo). Essa saída não existe quando o estudo é com cela(s) vazia(s).

Stat: apresenta os resultados (estatística de Wald e tipo ANOVA) para os testes definidos pelo usuário com suas respectivas matrizes de contrastes.

Stat\$Tests[[h]]: o usuário seleciona apenas um dos resultados dos testes definidos em **contrast**, pela sua posição ocupada na lista. Na saída de **Stat\$Tests[[h]]** o usuário encontrará a matriz de contrastes e seus respectivos testes de Wald e tipo-ANOVA. Por exemplo, para selecionar os resultados dos testes cuja matriz de contraste ocupa a posição 3 na lista, o usuário deve escrever **Stat\$Tests[[3]]**.

Covariance: o usuário obtém a matriz de variâncias e covariâncias estimada.

3.6 Exemplo

A sub-rotina pode ser utilizada para análise de dados com respostas completas (*i.e.*, quando todas as unidades amostrais possuem resposta em todos os tempos) ou com respostas omissas (há uma ou mais unidades amostrais sem observação de resposta em algum instante); além disso ela possibilita a análise de estudos com dois ou mais fatores entre unidades amostrais cruzados, com celas vazias ou não. Para esse tipo de análise, basta construir uma nova variável cujos níveis correspondem à combinação dos níveis dos fatores com celas não vazias e especificar as hipóteses de interesse por meio de uma escolha adequada da matriz **C** em (2).

Como exemplo consideremos um estudo com 3 grupos e 5 tempos. Na Figura 1a apresentamos o perfil (hipotético) de efeitos relativos de tratamento em um caso em que não há celas vazias. Na Figura 1b representamos os perfis com duas celas vazias (uma no grupo 2, tempo 4 e outra no grupo 3, tempo 2). Para um estudo com dois fatores sem celas vazias (Figura 1a) o vetor de parâmetros

Figura 3.6.1: Gráficos de perfis (hipotético) de efeitos relativos de tratamento: (a) sem celas vazias e (b) com duas celas vazias (uma no grupo 2, tempo 4 e outra no grupo 3, tempo 2).

é $\mathbf{p} = (p_{11}, p_{12}, \dots, p_{15}, p_{21}, \dots, p_{25}, p_{31}, p_{32}, \dots, p_{35})^\top$; a matriz especificadora da hipótese de inexistência de interação para o estudo apresentado na Figura 1a é

$$\mathbf{C}_{1a} = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & -1 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & -1 \end{pmatrix}$$

Para o estudo da Figura 1b, o vetor de parâmetros é $\mathbf{p} = (p_{11}, \dots, p_{15}, p_{21}, p_{22}, p_{23}, p_{25}, p_{31}, p_{33}, p_{34}, p_{35})^\top$ e a matriz especificadora da hipótese de inexistência de interação nesse caso é definida por

$$\mathbf{C}_{1b} = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & -1 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & -1 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & -1 & 1 \end{pmatrix}$$

A seguir apresentamos, com finalidade puramente ilustrativa, dois exemplos do uso da sub-rotina, ambos com um fator interunidades e um fator intraunidades amostrais (o primeiro exemplo, sem celas vazias e o segundo, com uma cela vazia). Estes exemplos foram extraídos de um estudo realizado no Laboratório Experimental de Poluição Atmosférica da Faculdade de Medicina da Universidade de São Paulo para avaliar os efeitos de agentes oxidantes no sistema respiratório. Para mais detalhes, ver Singer e Andrade (2000). Em ambos os casos, a variável resposta é a velocidade de transporte mucociliar relativa definida como o quociente entre a velocidade de transporte mucociliar num determinado instante e aquela obtida antes da intervenção experimental.

Exemplo 1: Neste exemplo, sem cela vazia ou dados omissos, selecionamos as unidades amostrais submetidas aos níveis 1, 8 e 32 μM de concentração de H_2O_2 (fator interunidades amostrais) avaliadas aos 5, 10, 15, 20, 25, 30 e 35 minutos (fator intraunidades amostrais), testamos as hipóteses usuais e especificamos algumas hipóteses para mostrar a construção de contrastes de interesse.

Para este exemplo, X_{ijk} representa o valor observado da velocidade de transporte mucociliar relativa observada na j -ésima unidade amostral ($j = 1, \dots, n_i$) submetida à i -ésima concentração de H_2O_2 ($i = 1, 2, 3$) no k -ésimo instante ($k = 1, 2, \dots, 7$) e $n_i = 10$. Os parâmetros de interesse são $\mathbf{p} = (p_{11}, \dots, p_{17}, p_{21}, \dots, p_{27}, p_{31}, \dots, p_{37})^\top$.

Os resultados da ANOVA não paramétrica usual estão apresentados na Tabela 2. Usando um nível de significância de 5%, notamos que há evidências de interação (valor- $p < 0,001$); em sequência testamos as seguintes hipóteses.

- 1) Efeito de concentração aos 5 minutos: $H_0 : p_{11} = p_{21} = p_{31}$,
- 2) Efeito das concentrações 1 e 8 μM : $H_0 : p_{1.} = p_{2.}$ ($k = 1, \dots, 7$),
- 3) $H_0 : p_{35} = p_{36} = p_{37}$.

em que, $p_{i.} = \frac{1}{7} \sum_{k=1}^7 p_{ik}$, $i = 1, 2, 3$ e $k = 1, \dots, 7$ e matrizes especificadoras das hipóteses dadas, respectivamente, por.

$$\mathbf{C}_1 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$\mathbf{C}_2 = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$\mathbf{C}_3 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 \end{pmatrix}$$

Figura 3.6.2: Estudo sem cela vazia: (a) Gráficos de perfis individuais por nível de concentração. (b) Intervalos de confiança, com 95% de confiança, para os efeitos relativos de tratamento.

Os resultados dos testes das hipóteses indicadas em 1, 2 e 3 estão dispostos na Tabela 3, e há evidências (i) de que aos 5 minutos, os efeitos relativos de tratamentos nas três concentrações são iguais (Hipótese 1), (ii) que não há diferença entre as concentrações 0 e 8 μM (Hipótese 2) e (iii) que aos 25, 30 e 35 minutos os efeitos relativos da concentração 32 μM são iguais (Hipótese 3).

Exemplo 2: Neste exemplo, com cela vazia, selecionamos as unidades amostrais submetidas aos níveis 0, 16 e 64 μM de concentração de H_2O_2 (fator interunidades

Tabela 3.6.2: Estatísticas de teste (tipo Wald e tipo ANOVA), graus de liberdade e níveis descritivos para as hipóteses usuais.

Estudo dos efeitos	Tipo	Estatística	g.l.	Valor-p
Concentração	Wald	39,402	2,00	< 0,001
	ANOVA	13,838	1,84	< 0,001
Tempo	Wald	36,737	6,00	< 0,001
	ANOVA	6,309	2,23	0,001
Concentração x Tempo	Wald	64,574	12,00	< 0,001
	ANOVA	5,474	4.19	< 0,001

Tabela 3.6.3: Estatísticas de teste (tipo Wald e tipo ANOVA), graus de liberdade e níveis descritivos para as hipóteses de interesse.

Estudo dos efeitos	Tipo	Estatística	g.l.	Valor-p
Hipótese 1	Wald	2,993	2,00	0,224
	ANOVA	1,672	1,97	0.188
Hipótese 2	Wald	0,123	1,00	0,726
	ANOVA	0,123	2,23	0,726

amostrais) avaliadas aos 5, 10, 15, 20, 25, 30 e 35 minutos (fator interunidades amostrais). Não há dados disponíveis para a cela correspondente á concentração 0 μM aos 35 minutos e também não há dados omissos.

Para este exemplo, X_{ijk} representa o valor observado da velocidade de transporte mucociliar relativa observada na j -ésima unidade amostral, $j = 1, \dots, n_i$, submetida à i -ésima concentração de H_2O_2 , $i = 1, 2, 3$, no k -ésimo instante, $k = 1, 2, \dots, 7$. Além disso, temos $n_1 = 6$ (com $k = 1, 2, \dots, 6$) e $n_2 = n_3 = 10$, (com $k = 1, \dots, 7$). Os parâmetros de interesse são $\mathbf{p} = (p_{11}, \dots, p_{16}, p_{21}, \dots, p_{27}, p_{31}, \dots, p_{37})^\top$.

Em estudos com celas vazias, não é possível obter resultados da ANOVA não paramétrica usual e as hipóteses da ANOVA em termos dos efeitos relativos de tratamentos podem ser reexpressas como

- 1) Interação entre concentração e tempo: $H_0 : p_{ik} = p_{i.} + p_{.k} - p_{..}$ ($i = 1, 2, 3$) e ($k = 1, \dots, 6$),
- 2) Efeito principal de concentração: $H_0 : p_{1.} = p_{2.} = p_{3.} = p_{..}$ ($k = 1, \dots, 7$),
- 3) Efeito principal de tempo: $H_0 : p_{.1} = p_{.2} = \dots = p_{.7} = p_{..}$ ($i = 1, 2, 3$).

em que, $p_{1.} = \frac{1}{6} \sum_{k=1}^6 p_{1k}$, $p_{i.} = \frac{1}{7} \sum_{k=1}^7 p_{ik}$, $i = 2, 3$, $p_{.k} = \frac{1}{3} \sum_{i=1}^3 p_{ik}$, $k = 1, \dots, 6$, $p_{.7} =$

$\frac{1}{2} \sum_{i=2}^3 p_{ik}$, $p_{..} = \frac{1}{20} \left(\sum_{k=1}^6 p_{1k} + \sum_{i=2}^3 \sum_{k=1}^7 p_{ik} \right)$ e matrizes especificadoras das hipóteses de interação e de efeitos principais de concentração e tempo dadas, respectivamente, por

$$\mathbf{C}_{C*T} = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 & 0 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 & 0 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 \end{pmatrix}$$

$$\mathbf{C}_C = \begin{pmatrix} \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & -\frac{1}{7} & -\frac{1}{7} & -\frac{1}{7} & -\frac{1}{7} & -\frac{1}{7} & -\frac{1}{7} & -\frac{1}{7} & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\frac{1}{7} & -\frac{1}{7} & -\frac{1}{7} & -\frac{1}{7} & -\frac{1}{7} & -\frac{1}{7} \end{pmatrix}$$

$$\mathbf{C}_T = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{3} & 0 & 0 & 0 & 0 & 0 & \frac{1}{3} & -\frac{1}{2} & 0 & 0 & 0 & 0 & \frac{1}{3} & -\frac{1}{2} \end{pmatrix}$$

Figura 3.6.3: Estudo com cela vazia: (a) Gráficos de perfis individuais por nível de concentração. (b) Intervalos de confiança, com 95% de confiança, para os efeitos relativos de tratamento.

(*)como discutido em Brunner, Domhof e Langer (2002), opta-se pelo resultado inferencial da estatística do tipo ANOVA, que é mais apropriada para amostras pequenas.

A análise não paramétrica para medidas repetidas evita os problemas que envolvem a escolha da estrutura de correlação e requer suposições pouco restritivas. Essa metodologia pode ser aplicada com respostas contínuas, discretas, ordinais, e mesmo dicotômicas. Além disso, ela é bastante confiável para amostras pequenas, sendo que neste caso a estatística mais indicada é a do tipo ANOVA. Para mais detalhes da comparação entre análise não paramétrica para medidas repetidas e as

Tabela 3.6.4: Estatísticas de teste (tipo Wald e tipo ANOVA), graus de liberdade e níveis descritivos para as hipóteses de interesse.

Estudo parcial dos efeitos	Tipo	Estatística	g.l.	Valor-p
Concentração ²	Wald	40,616	2,00	< 0,001
	ANOVA	17,575	1,76	< 0,001
Tempo ³	Wald	12,129	6,00	0,0592
	ANOVA*	4,452	2,60	0,0061
Concentração x Tempo ¹	Wald	36,624	10,00	< 0,001
	ANOVA	4,339	4,54	< 0,001

análises paramétricas o leitor pode consultar Singer, Poletto e Rosa (2004). O desenvolvimento de uma sub-rotina que permite a visualização do comportamento dos efeitos relativos de tratamento por intermédio de um gráfico com intervalo de confiança, mesmo quando o planejamento fatorial não é totalmente cruzado aumenta a flexibilidade e o potencial de utilização desse método.

3.7 Exercícios

3.7.1. Os dados da planilha disponível em

<http://www.ime.usp.br/~jmsinger/Dados/Singer&Nobre&Rocha2017exerc361.xls>

foram obtidos de um estudo sobre fertilização animal. Um dos objetivos é avaliar o efeito de uma suplementação alimentar (relativamente a uma dieta controle) e do tempo de sua administração no número de ovócitos aspirados de diferentes vacas doadoras.

- Faça uma análise descritiva dos dados, incluindo gráficos de perfis individuais e médios.
- Analise os dados por meio de modelos generalizados mistos (GLMM) admitindo que a distribuição da resposta é Poisson e utilizando uma função de ligação logarítmica.
- Reanalise os dados por meio de modelos baseados em equações de estimação generalizadas (GEE) com duas propostas para a matriz de covariâncias de trabalho.
- Repita a análise adotando modelos lineares generalizados mistos aproximando a distribuição das respostas por uma distribuição normal e utilizando uma função de ligação logarítmica.

- e) Repita a análise do item d) por meio de modelos lineares mistos considerando o logaritmo do número de ovócitos como variável resposta. Apresente a análise de resíduos correspondente ao ajuste desse modelo.
- f) Resuma os resultados de suas análises em uma tabela e discuta-os comparativamente.

Capítulo 4

Tópicos especiais

4.1 Dados omissos

4.2 Planejamento e tamanho de amostras

4.3 Modelos não-lineares

Capítulo 5

Análise de dados

Neste capítulo apresentamos exemplos de análise de dados, aproveitando-os para discutir alguns tópicos metodológicos não abordados nos demais.

5.1 Estratégias de análise

Convém notar que as técnicas de diagnóstico resumidas na Seção 2.5 dependem da correta especificação da estrutura de covariância adotada conforme sugerem Zewotir & Galpin (2007). Embora Christensen et al. (1992) indiquem que boas estimativas dos parâmetros de covariância possam servir como base para diagnóstico, algum esforço no refinamento do modelo proposto deve ser empregado para prevenir, ou ao menos diminuir, a probabilidade de detecção inadequada de unidades amostrais ou observações.

Com esse espírito, Rocha & Singer (2017) apresentam algumas propostas no caso de modelos polinomiais de coeficientes aleatórios, *i.e.*, aqueles em que as colunas de \mathbf{X}_i e \mathbf{Z}_i em (2.1.2) correspondem a potências da variável tempo. Esses autores propõem (i) testes t baseados nas estimativas dos coeficientes de regressões lineares padrão ajustadas aos dados de cada unidade amostral como ferramenta para seleção dos efeitos fixos e (ii) intervalos de confiança corrigidos pela técnica de Bonferroni para a seleção dos efeitos aleatórios (e da correspondente matriz de covariâncias). Eles também observam que quando $\mathbf{R}_i = \sigma^2 \mathbf{I}_{m_i}$, a estrutura das colunas da matriz de covariâncias intraunidades amostrais é similar àquela dos perfis individuais. Explicitamente, denotando por \mathbf{z}_{is}^\top a s -ésima linha da matriz \mathbf{Z}_i , a contribuição dos efeitos aleatórios para a s -ésima coluna da matriz de covariâncias (2.1.4) é dada por $\mathbf{Z}_i \mathbf{G} \mathbf{z}_{is}$, que tem a mesma forma que a componente do perfil individual $\mathbf{Z}_i \mathbf{b}_i$. Nos casos em que as observações de todas as unidades amostrais são obtidas nos mesmos

instantes, temos $\mathbf{Z}_i = \mathbf{Z}^*$ de forma que $\mathbf{V} = \mathbf{Z}^* \mathbf{G} \mathbf{Z}^{*\top}$ pode ser estimada por

$$\hat{\mathbf{V}} = n^{-1} \sum_{i=1}^n (\mathbf{y}_i - \bar{\mathbf{y}})(\mathbf{y}_i - \bar{\mathbf{y}})^\top - \hat{\sigma}^2 \mathbf{I}_m$$

em que m é o número de observações por unidade amostral, $\bar{\mathbf{y}} = n^{-1} \sum_{i=1}^n \mathbf{y}_i$ e $\hat{\sigma}^2$ é uma estimativa de σ^2 . Rocha & Singer (2017) sugerem o ajuste de modelos de regressão linear padrão às linhas de $\hat{\mathbf{V}}$ e a utilização de intervalos de referência corrigidos pela técnica de Bonferroni para os coeficientes como ingredientes para a decisão sobre que efeitos aleatórios devem ser incluídos no modelo. Com base em um estudo de simulação, eles sugerem que mesmo para tamanhos amostrais moderados, *e.g.*, 25 unidades com 5 observações cada o procedimento é eficiente, mesmo com efeitos aleatórios ou erros não gaussianos.

Grady & Helms (1995) sugerem que gráficos das covariâncias e correlações *versus* o tempo entre instantes de observação podem ser utilizados como ferramenta para a identificação de padrões autorregressivos e propõem uma estratégia para comparar diferentes modelos. Singer & Cúri (2006) consideram um exemplo em que as relações entre estimativas de variâncias e covariâncias obtidas de um modelo saturado para os parâmetros fixos e uma variável explicativa são exploradas para sugerir a forma das matrizes \mathbf{Z}_i e \mathbf{G} em (2.1.2).

A escolha dos modelos para a estrutura de covariância intraunidades amostrais depende de informações sobre o tipo de estudo com medidas repetidas (longitudinal ou não), sobre seu planejamento (balanceado ou não) e sobre a existência de dados omissos. Estudos com medidas repetidas não-longitudinais, em geral, têm caráter experimental e a correspondente estrutura de covariância intraunidades amostrais pode ser modelada com o auxílio de informações sobre seu planejamento. Para estudos longitudinais, é mais razoável utilizar informações sobre o comportamento da resposta ao longo das condições de avaliação na modelagem da estrutura de covariância intraunidades amostrais.

Para estudos não-longitudinais, sugerimos a seguinte estratégia de seleção da estrutura de covariância intraunidades amostrais:

1. Análise do planejamento do estudo para identificação de um modelo saturado para os efeitos fixos e de possíveis efeitos aleatórios.
2. Análise das matrizes de covariâncias e correlações amostrais, quando for possível calculá-las, para identificar seus padrões de variação.
3. Ajuste de modelos saturados para os parâmetros de localização identificados no item 2 com matrizes de covariâncias identificadas no item 1.

4. Utilização dos critérios de informação AIC_R e BIC_R como ferramenta auxiliar na escolha da estrutura de covariância identificada nos itens 1 e 2.
5. Utilização de técnicas de diagnóstico para avaliar a qualidade do ajuste.
6. Ações corretivas para modificação do modelo.

Para estudos longitudinais, uma estratégia para a identificação da estrutura de covariância intraunidades amostrais é a seguinte:

1. Análise do gráfico de perfis médios para propor um modelo para os efeitos fixos.
2. Análise dos gráficos de perfis individuais e de perfis individuais centralizados para identificar padrões que indiquem possíveis efeitos aleatórios e heterocedasticidade ao longo das condições de avaliação.
3. Análise das matrizes de covariâncias e correlações intraunidades amostrais, quando for possível calculá-las, para identificar seus padrões de variação.
4. Análise da matriz de gráficos de dispersão das observações padronizadas (*draftman's plot*) quando for possível construí-los.
5. Análise de gráficos de linhas das matrizes de covariâncias e correlações amostrais *versus* defasagens entre os instantes de observação.
6. Análise do variograma amostral como sugerido por Diggle et al. (2002).
7. No caso de estudos longitudinais sem dados omissos, análise do gráfico das linhas da matriz de covariâncias intraunidades amostrais por meio de técnicas de regressão polinomial para a identificação de efeitos aleatórios.
8. No caso de estudos longitudinais com dados omissos, análise dos perfis individuais por meio de técnicas de regressão polinomial para a identificação de efeitos aleatórios.
9. Utilização dos critérios de informação como ferramenta auxiliar para a seleção das estruturas de covariância intraunidades amostrais identificadas.
10. Utilização de técnicas de diagnóstico para avaliar a qualidade do ajuste.
11. Ações corretivas para modificação do modelo.

5.2 Implementação computacional

Embora atrativos, os numerosos enfoques para a análise de medidas repetidas ou de dados longitudinais disponíveis na literatura estatística podem se constituir numa fonte de dificuldades quando a questão é ajustar modelos a conjuntos de dados provenientes de problemas práticos. Um exemplo típico em que vários modelos alternativos podem ser utilizados é apresentado em Alencar, Singer & Rocha (2012). Em primeiro lugar, é difícil decidir que classe de modelos e que estratégia de análise de ser adotada. Em segundo lugar, “software” adequado para implementação do modelo adotado nem sempre está disponível ou nem sempre acomodam as particularidades usualmente encontradas em problemas práticos. Obviamente, é sempre possível desenvolver ou modificar o “software” disponível para que satisfaça os requisitos do problema sob análise; embora isso possa ser uma fonte de deleite para investigadores interessados no desenvolvimento de métodos computacionais, em geral, analistas não dispõem do tempo necessário para essa atividade.

A estratégia de modelagem deve incluir um processo iterativo segundo o qual após o ajuste de cada modelo, ferramentas de diagnóstico sejam empregadas para avaliar se a qualidade do ajuste do modelo atual é melhor do que a do anterior. Com grande probabilidade, um ajuste ótimo não será atingido, mas a crescente familiaridade com os dados deverá melhorar a análise.

Para efeito computacional, em geral, os dados devem ser dispostos na forma indicada em (2.1.7). Convém lembrar que uma variável que indique as unidades de (ou sub-unidades) amostrais em que são realizadas as medidas repetidas deve ser sempre incluída no arquivo de dados. Por exemplo, os dados correspondentes à segunda e à sexta unidades amostrais (nascidos pré-termo) do Exemplo 1.2.6 devem ser colocados no formato indicado na Tabela 5.2.1.

Os pacotes computacionais mais populares para o ajuste de modelos lineares mistos são `proc MIXED` no pacote `SAS` e as programotecas `lme4` e `nlme` do pacote `R`. Restringimo-nos ao pacote `R` dado que é “software” livre e que permite tanto o desenvolvimento de funções apropriadas para a análise desse tipo de modelos quanto modificação de funções já existentes.

As funções disponíveis estão resumidas na Tabela 5.2.2.

“Software” para diagnóstico só está disponível em poucos pacotes de análise estatística ou deve ser obtido diretamente dos autores e geralmente é dirigido para casos particulares. Uma revisão pode ser encontrada em West & Galecki (2011). Um conjunto de funções para construção dos gráficos apresentados na Seção 2.5 pode ser obtido em www.ime.usp.br/~jmsinger/lmmdiagnosics.zip. Em particular, a função `diaglme4` gera esses gráficos a partir de ajustes obtidos via `lmer` e a função `diagnlme` os gera a partir de ajustes obtidos via `lme`.

Tabela 5.2.1: Formatação dos dados do Exemplo 1.2.6 para efeito computacional

Peso ao ao nascer	Unidade amostral	Diâmetro sistólico da aorta/peso	Semana
AIG	2	6.1	30
AIG	2	6.1	31
AIG	2	6.2	32
AIG	2	5.4	34
AIG	2	5.2	35
AIG	2	4.9	36
AIG	6	5.4	31
AIG	6	4.9	33
AIG	6	4.6	34
AIG	6	4.3	35

Tabela 5.2.2: “Software” R para ajuste de modelos mistos

Programoteca	Função	Ajustes	Distribuição dos efeitos aleatórios	Matrizes \mathbf{G} ou \mathbf{R}_w	Distribuição dos erros	Matriz \mathbf{R}_i
lme4	lmer	LMM	gaussiana	\mathbf{G} não estruturada	gaussian	$\sigma^2 \mathbf{I}_{m_i}$
	nlmer	NLMM	gaussiana	\mathbf{G} não estruturada	gaussiana	estruturada
	glmer	GLMM	gaussiana	\mathbf{G} não estruturada	Família	NA
nlme	lme	LMM	gaussiana	\mathbf{G} estruturada	exponencial gaussiana	estruturada
	nlme	NLMM	gaussiana	\mathbf{G} estruturada	gaussiana	estruturada
	gls	LM	NA	NA	gaussiana	estruturada
gee	gee	Modelos GEE	NE	\mathbf{R}_w estruturada	Família	NA
geepack	geeglm	Modelos GEE	NE	\mathbf{R}_w estruturada	exponencial ou NE Família	NA
heavy	heavyLme	ES-LMM	elípticamente simétrica	\mathbf{G} não estruturada	exponencial ou NE elípticamente simétrica	NA
robustlmm	rlmer	LMM	simétrica	\mathbf{G} diagonal ou	simétrica	NA
		robusto	não estruturada			$\sigma^2 \mathbf{I}_{m_i}$

NA: não aplicável
NE: não especificada

5.3 Estudos pré-teste/pós-teste

Estudos do tipo pré-teste/pós-teste são frequentemente usados em áreas como Biologia, Medicina, Odontologia e Psicologia, entre outras. O objetivo, em geral, é avaliar o efeito de algum tipo de intervenção (um tratamento, por exemplo) na distribuição de alguma resposta. O procedimento utilizado para sua concretização consiste na observação da variável resposta antes e depois da intervenção. Consideramos estudos do tipo pré-teste/pós-teste sob dois cenários que descrevemos por meio de exemplos.

Exemplo 5.3.1: Num estudo realizado no Instituto de Ciências Biomédicas da Universidade de São Paulo para avaliar o efeito de $MgSO_4$ na pressão arterial média (PAM, em mmHg) de cães, a resposta foi medida antes e depois da aplicação desse sal em dois grupos de animais, o primeiro ($n = 12$) previamente tratado com indometacina e o outro ($n = 12$) previamente tratado com nifedipina. Para detalhes, ver Singer, Seoanes & Ogando (1988). Os dados estão apresentados na Tabela 5.3.1, um gráfico de perfis individuais, na Figura 5.3.1 e um gráfico de perfis médios, na Figura 5.3.2.

Tabela 5.3.1: Pressão arterial média (mmHg) de cães antes e depois de uma aplicação de $MgSO_4$

Cão	Grupo	Antes	Depois	Cão	Grupo	Antes	Depois
1	indometacina	148	132	13	nifedipina	82	68
2	indometacina	100	76	14	nifedipina	90	72
3	indometacina	120	108	15	nifedipina	102	84
4	indometacina	116	96	16	nifedipina	110	100
5	indometacina	140	128	17	nifedipina	86	58
6	indometacina	92	88	18	nifedipina	90	76
7	indometacina	112	100	19	nifedipina	90	78
8	indometacina	120	108	20	nifedipina	84	72
9	indometacina	128	100	21	nifedipina	98	80
10	indometacina	124	100	22	nifedipina	116	80
11	indometacina	108	92	23	nifedipina	90	80
12	indometacina	100	80	24	nifedipina	94	76

Gráficos do tipo *boxplot* correspondentes estão dispostos na Figura 5.3.3. Neste exemplo, temos uma intervenção (aplicação de $MgSO_4$) e uma população com dois estratos, que correspondem aos dois tratamentos (indometacina e nifedipina). O interesse é:

- comparar o efeito da aplicação de $MgSO_4$ (PAM esperada pós-teste - PAM esperada pré-teste) nos cães tratados com indometacina com o efeito da aplicação

Figura 5.3.1: Perfis individuais da PAM de cães, antes (pré-teste) e depois (pós-teste) de uma aplicação de $MgSO_4$

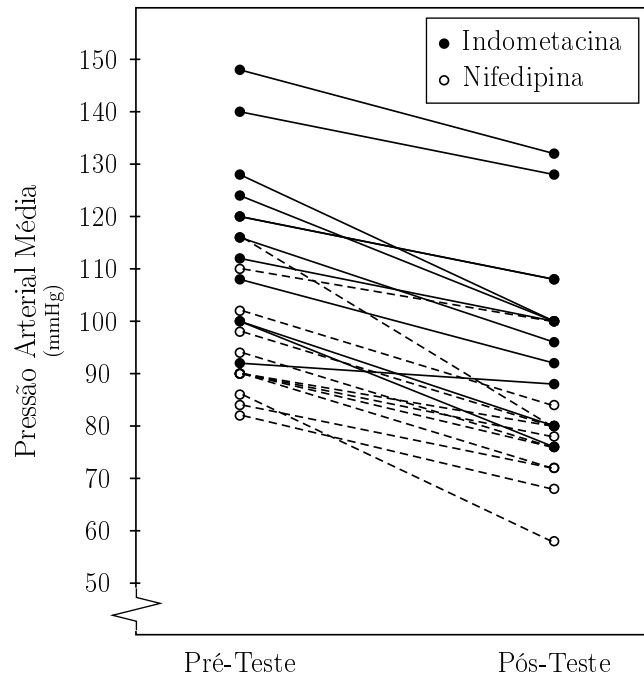
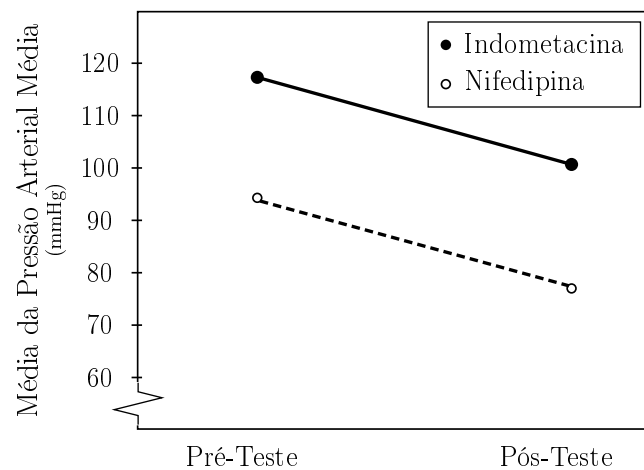
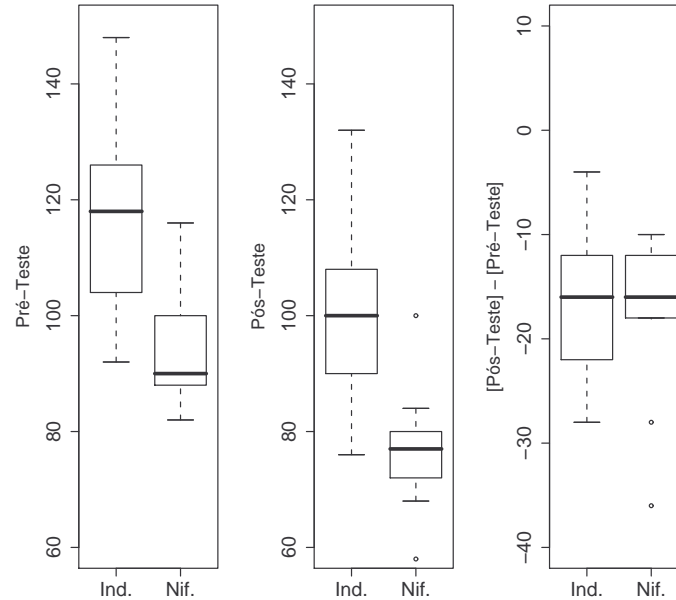


Figura 5.3.2: Perfis médios e médias amostrais (círculos) para a PAM, antes (pré-teste) e depois (pós-teste) de uma aplicação de $MgSO_4$.



de $MgSO_4$ nos cães tratados com nifedipina (interação entre tratamento e tempo),

Figura 5.3.3: Boxplots para a PAM de cães, antes (pré-teste) e depois (pós-teste) de uma aplicação de $MgSO_4$ e para o efeito da aplicação



- b) avaliar o efeito principal da aplicação de $MgSO_4$ (se não houver interação ou se ela for não-essencial),
- c) avaliar o efeito da aplicação de $MgSO_4$ para cada tratamento (se houver interação essencial),
- d) avaliar o efeito principal de tratamento sobre a PAM esperada (se não houver interação ou se ela for não-essencial)
- e) avaliar o efeito de tratamento sobre a PAM esperada pré-teste (se houver interação essencial).

Para situações como aquela descrita no Exemplo 5.3.1, com n_i unidades amostrais no estrato i , $i = 1, 2$, o modelo linear básico pode ser escrito como

$$y_{ijk} = \mu_{ik} + e_{ijk}, \quad i = 1, 2, \quad j = 1, \dots, n_i, \quad k = 1, 2 \quad (5.3.1)$$

em que y_{ijk} corresponde à resposta observada para a j -ésima unidade amostral do i -ésimo estrato no k -ésimo tempo, μ_{ik} denota a resposta esperada para as unidades amostrais do i -ésimo estrato no k -ésimo tempo e e_{ijk} é o erro aleatório correspondente. Sob a perspectiva de Análise de Variância Multivariada (MANOVA), esse modelo pode ser colocado na forma (2.4.1), *i.e.*, $\mathbf{Y} = \mathbf{X}\mathbf{M} + \mathbf{E}$, com

$\mathbf{Y} = (\mathbf{y}_{11}, \dots, \mathbf{y}_{1n_1}, \mathbf{y}_{21}, \dots, \mathbf{y}_{2n_2})^\top$, $\mathbf{y}_{ij} = (y_{ij1}, y_{ij2})^\top$, $\mathbf{X} = \bigoplus_{i=1}^2 \mathbf{1}_{n_i}$, $\mathbf{M} = (\boldsymbol{\mu}_1, \boldsymbol{\mu}_2)^\top$ com $\boldsymbol{\mu}_i = (\mu_{i1}, \mu_{i2})^\top$ e $\mathbf{E} = (\mathbf{e}_{11}, \dots, \mathbf{e}_{1n_1}, \mathbf{e}_{21}, \dots, \mathbf{e}_{2n_2})^\top$, $\mathbf{e}_{ij} = (e_{ij1}, e_{ij2})^\top$. Nesse contexto, assumimos que $\mathbf{e}_{ij} \sim N_2(\mathbf{0}, \boldsymbol{\Omega})$ com

$$\boldsymbol{\Omega} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}$$

são erros independentes, $i = 1, 2$, $j = 1, \dots, n_i$. As hipóteses mencionadas acima podem ser expressas na forma (2.4.2), *i.e.*, $\mathbf{CMU} = \mathbf{0}$, com

- i) $\mathbf{C} = (1, -1)$ e $\mathbf{U} = (1, -1)^\top$ (interação entre tratamento e tempo),
- ii) $\mathbf{C} = (1, 1)$ e $\mathbf{U} = (1, -1)^\top$ (efeito principal de tempo na ausência de interação entre tratamento e tempo),
- iii) $\mathbf{C} = (1, 0)$ (ou $\mathbf{C} = (0, 1)$) e $\mathbf{U} = (1, -1)^\top$ [efeito de tempo para o primeiro (segundo) tratamento],
- iv) $\mathbf{C} = (1, -1)$ e $\mathbf{U} = (1, 1)^\top$ (efeito principal de tratamento na ausência de interação entre tratamento e tempo),
- v) $\mathbf{C} = (1, -1)$ e $\mathbf{U} = (1, 0)^\top$ (efeito de tratamento no pré-teste).

Explicitamente, sob o modelo (2.4.1), as hipóteses i) - v) correspondem respectivamente a $\mu_{11} - \mu_{12} = \mu_{21} - \mu_{22}$, $\mu_{11} + \mu_{21} = \mu_{12} + \mu_{22}$, $\mu_{11} = \mu_{12}$, $\mu_{21} = \mu_{22}$, $\mu_{11} + \mu_{12} = \mu_{21} + \mu_{22}$, $\mu_{11} = \mu_{21}$. As estatísticas de teste para essas hipóteses são baseadas nas raízes características das matrizes de soma de quadrados e produtos cruzados indicadas na Seção 2.4.

Muitos textos e pacotes computacionais (REF) utilizam o modelo (5.3.1) sob a **reparametrização de desvios médios**

$$\mu_{ik} = \mu + \alpha_i + \beta_k + \alpha\beta_{ik}, \quad i = 1, 2, \quad k = 1, 2 \quad (5.3.2)$$

com as **restrições de identificabilidade**

$$\sum_{i=1}^2 \alpha_i = \sum_{k=1}^2 \beta_k = \sum_{i=1}^2 \alpha\beta_{ik} = \sum_{k=1}^2 \alpha\beta_{ik} = 0.$$

A expressão das hipóteses i) - v) em termos dos parâmetros do modelo (5.3.2) ou de outros com diferentes parametrizações está explicitada no Apêndice B.

Um outro enfoque clássico para o problema envolve um modelo misto obtido de (5.3.1) por intermédio da inclusão da seguinte estrutura para o erro aleatório

$$e_{ijk} = b_{ij} + d_{ijk} \quad (5.3.3)$$

em que $b_{ij} \sim N(0, \sigma_b^2)$ e $d_{ijk} \sim N(0, \sigma^2)$ com todas as variáveis envolvidas supostas independentes. Esse modelo, conhecido na literatura por **modelo de parâmetros irrestritos** (*unconstrained parameter model - UP*)¹ implica

$$\mathbb{V}(\mathbf{e}_{ij}) = \mathbf{\Omega} = \begin{bmatrix} \sigma_b^2 + \sigma^2 & \sigma_b^2 \\ \sigma_b^2 & \sigma_b^2 + \sigma^2 \end{bmatrix}$$

e consequentemente, é mais restritivo que o modelo MANOVA, na medida em que impõe homocedasticidade para as observações intraunidades amostrais e não permite covariâncias negativas entre elas. Essencialmente, este modelo, tem a forma (2.1.7) com

$$\begin{aligned} \mathbf{y} &= (\mathbf{y}_{11}^\top, \dots, \mathbf{y}_{1n_1}^\top, \mathbf{y}_{21}^\top, \dots, \mathbf{y}_{2n_2}^\top)^\top, \\ \mathbf{X} &= \bigoplus_{i=1}^2 (\mathbf{1}_{n_i} \otimes \mathbf{I}_2), \mathbf{\beta} = (\mu_{11}, \mu_{12}, \mu_{21}, \mu_{22})^\top, \mathbf{Z} = \bigotimes_{l=1}^{n_1+n_2} \mathbf{1}_2, \\ \mathbf{b} &= (b_{11}, \dots, b_{1n_1}, b_{21}, \dots, b_{2n_2})^\top \end{aligned}$$

e

$$\mathbf{e} = (\mathbf{e}_{11}^\top, \dots, \mathbf{e}_{1n_1}^\top, \mathbf{e}_{21}^\top, \dots, \mathbf{e}_{2n_2}^\top)^\top.$$

Além disso, ele pode ser considerado como um caso particular dos modelos do tipo *split-plot* e a correspondente análise de dados pode ser implementada por meio de ANOVA para medidas repetidas. Nesse contexto, as hipóteses de interesse podem ser testadas por meio de estatísticas como aquelas indicadas na Tabela 5.3.2, em que as somas de quadrados podem ser expressas em termos das somas e diferenças entre as respostas intraunidades amostrais, nomeadamente, $s_{ij} = y_{ij1} + y_{ij2}$ e $d_{ij} = y_{ij1} - y_{ij2}$, ou seja

$$SQTr = \sum_{i=1}^2 n_i (\bar{s}_i - \bar{s})^2,$$

com $\bar{s}_i = n_i^{-1} \sum_{j=1}^{n_i} s_{ij}$ e $\bar{s} = (n_1 + n_2)^{-1} \sum_{i=1}^2 \sum_{j=1}^{n_i} s_{ij}$,

$$SQE1 = \sum_{i=1}^2 \sum_{j=1}^{n_i} (s_{ij} - \bar{s}_i)^2,$$

$$SQTrTp = \sum_{i=1}^2 n_i (\bar{d}_i - \bar{d})^2,$$

¹Embora o modelo *UP* seja o mais comumente utilizado na prática, um modelo alternativo, conhecido por **modelo de parâmetros restritos** (*constrained parameter model - CP*) também pode ser adotado para os mesmos propósitos. Esses dois modelos constituem o cerne da questão conhecida como *controvérsia dos modelos mistos*, discutida em Voss (1999), Lencina, Singer & Stanek III (2005), Lencina & Singer (2006) e Lencina, Singer & Stanek III (2008).

com $\bar{d}_i = n_i^{-1} \sum_{j=1}^{n_i} d_{ij}$ e $\bar{d} = (n_1 + n_2)^{-1} \sum_{i=1}^2 \sum_{j=1}^{n_i} d_{ij}$,

$$SQTp = [(n_1 n_2)/(n_1 + n_2)](\bar{d}_1 + \bar{d}_2)^2,$$

e

$$SQE2 = \sum_{i=1}^2 \sum_{j=1}^{n_i} (d_{ij} - \bar{d}_i)^2.$$

Sob as correspondentes hipóteses nulas, essas estatísticas tem distribuição F com 1 grau de liberdade no numerador e $n - 2$ (com $n = n_1 + n_2$) graus de liberdade no denominador.

Tabela 5.3.2: Tabela ANOVA para estudos com planejamento *split-plot*

Fonte de variação	Graus de liberdade	Soma de quadrados	Quadrado médio	Estatística F
Tratamentos	1	$SQTr$	$SQTr/1$	F_{Tr}
Unidades amostrais	$n - 2$	$SQE1$	$SQE1/(n - 2)$	
Tempo	1	$SQTp$	$SQTp/1$	F_{Tp}
Tempo \times Tratamento	1	$SQTpTr$	$SQTpTr/1$	F_{TpTr}
Observações	$n - 2$	$SQE2$	$SQE2/(n - 2)$	

A estatística $F_{TpTr} = (n - 2)SQTpTr/SQE2$ é apropriada para o teste da inexistência de interação entre tratamento e tempo (i) e as estatísticas $F_{Tr} = (n - 2)SQTr/SQE1$ e $F_{Tp} = (n - 2)SQTp/SQE2$, para testar as hipóteses de inexistência de efeitos principais de tratamento (iv) e tempo (ii), respectivamente, caso não exista interação ou ela seja não-essencial. Para testar as demais hipóteses de interesse, convém observar que as estatísticas F apresentadas acima podem ser substituídas por estatísticas t . Em particular, tendo em vista que a variância das diferenças d_{ij} e das somas s_{ij} das respostas intraunidades amostrais são respectivamente $s_d^2 = SQE2/(n - 2)$ e $s_s^2 = SQE1/(n - 2)$, podemos escrever

$$F_{TpTr} = \{(\bar{d}_1 - \bar{d}_2)/[s_d \sqrt{(n_1 + n_2)/(n_1 n_2)}]\}^2,$$

$$F_{Tp} = \{(\bar{d}_1 + \bar{d}_2)/[s_d \sqrt{(n_1 + n_2)/(n_1 n_2)}]\}^2,$$

e

$$F_{Tr} = \{(\bar{s}_1 - \bar{s}_2)/[s_s \sqrt{(n_1 + n_2)/(n_1 n_2)}]\}^2.$$

Então, pode-se deduzir que para testar a hipótese (iii), basta utilizar $F_{TpTr1} = \{\bar{d}_1/[s_d/\sqrt{n_1}]\}^2$ ou $F_{TpTr2} = \{\bar{d}_2/[s_d/\sqrt{n_2}]\}^2$, consoante o interesse recaia no primeiro ou no segundo tratamento.

Finalmente, para testar a hipótese definida em (v), pode-se utilizar a estatística

$$F_{TrTp1} = \{\bar{y}_{11} - \bar{y}_{11}/[s\sqrt{(n_1 + n_2)/(n_1 n_2)}]\}^2$$

em que $s^2 = (n_1 + n_2 - 4)^{-1} \sum_{i=1}^2 \sum_{j=1}^{n_i} \sum_{k=1}^2 (y_{ijk} - \bar{y}_{ik})^2$ com $\bar{y}_{ik} = n_i^{-1} \sum_{j=1}^{n_i} y_{ijk}$. Sob a hipótese nula de que as respostas esperadas das unidades amostrais dos dois tratamentos são iguais no pré-teste, a estatística F_{TrTp1} segue uma distribuição $F_{1, n_1 + n_2 - 4}$.

Convém notar que a suposição de homocedasticidade das medidas intraunidades amostrais (*i.e.*, igualdade das variâncias no pré- e pós-teste) não é necessária para que as estatísticas de teste consideradas acima sigam distribuições F . Basta que as matrizes de covariâncias intraunidades amostrais sejam iguais. Isso implica que os resultados são válidos sob modelos mais gerais do que aquele definido por (5.3.1)-(5.3.3). Os resultados da análise dos dados do Exemplo 5.3.2 estão dispostos na Tabela 5.3.3.

Tabela 5.3.3: Tabela ANOVA para os dados do Exemplo 5.3.1

Fonte de variação	Graus de liberdade	Soma de quadrados	Quadrado médio	Estatística F	Valor-P
Tratamento	1	6533.3	6533.3	18.36	0.0003
Cães	22	7829.3	355.9		
Tempo	1	3468.0	3468.0	131.85	0.0000
Tratamento \times Tempo	1	1.3	1.3	0.05	0.8239
Observações	22	578.7	26.3		

Estimativas das matrizes de covariâncias intraunidades amostrais para cada tratamento são, respectivamente

$$\hat{\Omega}_1 = \begin{bmatrix} 271.52 & 255.03 \\ & 284.61 \end{bmatrix} \text{ e } \hat{\Omega}_2 = \begin{bmatrix} 108.24 & 74.75 \\ & 100.00 \end{bmatrix}$$

e a estimativa da matriz de covariâncias intraunidades amostrais comum, obtida sob a hipótese de homocedasticidade entre os dois tratamentos é

$$\hat{\Omega} = \begin{bmatrix} 189.88 & 164.79 \\ & 192.30 \end{bmatrix}.$$

Apesar de as matrizes de covariâncias amostrais sugerirem uma possível heterocedasticidade, *i.e.*, que $\Omega_1 \neq \Omega_2$, as correspondentes estimativas das variâncias das diferenças entre as respostas no pré-teste e pós-teste são respectivamente $\hat{\sigma}_{d_1}^2 = 46.05$, $\hat{\sigma}_{d_2}^2 = 59.16$, e apontam para a homocedasticidade das correspondentes variâncias

populacionais, $\sigma_{d_1}^2$ e $\sigma_{d_2}^2$. Consequentemente, podemos considerar uma estimativa para a variância das diferenças entre as respostas no pré-teste e pós-teste combinada (*pooled*), a saber, $\hat{\sigma}_d^2 = 52.60$, e concluir que a análise baseada no modelo (5.3.1)-(5.3.3) é adequada para as hipóteses de inexistência de interação Tratamento \times Tempo e de inexistência efeito principal de Tempo.

Por outro lado, as estimativas das variâncias das somas das respostas no pré-teste e pós-teste são respectivamente $\hat{\sigma}_{s_1}^2 = 1066.19$, $\hat{\sigma}_{s_2}^2 = 357.84$, e nesse caso, não há evidências para concluirmos a homocedasticidade das correspondentes variâncias populacionais, $\sigma_{s_1}^2$ e $\sigma_{s_2}^2$. Por essa razão, a análise da hipótese de inexistência de efeito principal de Tratamento deveria ser adaptada para acomodar heterocedasticidade.

Finalizamos, observando que, como esperado,

$$t_{TrTp}^2 = \{(-16.63 + 17.33)/\sqrt{52.60(1/12 + 1/12)}\}^2 = 0.05$$

é exatamente o valor da estatística F_{TrTp} obtido na Tabela 5.3.3.

Há situações experimentais em que não existe interesse nas comparações envolvendo as medidas pré-teste. O exemplo descrito a seguir é um caso típico.

Exemplo 5.3.2: Consideramos um estudo realizado na Faculdade de Medicina da Universidade de São Paulo para avaliar a resistência respiratória de ratos com respeito a gases de diferentes densidades. Dentre as várias respostas avaliadas no estudo, selecionamos a resistência homogênea do sistema respiratório (RHSR) para nossos propósitos. Primeiramente, a RHSR de cada um de 29 ratos foi medida logo após serem submetidos à ventilação com ar sintético. Posteriormente, 15 dos ratos escolhidos aleatoriamente foram submetidos novamente à ventilação com ar sintético (*C*) e os outros 14 ratos foram submetidos à ventilação com uma mistura de hélio e oxigênio (*T*). A mistura de hélio e oxigênio representa o “tratamento” de interesse e o ar sintético corresponde ao “controle”. Ao término da intervenção, uma segunda medição da RHSR foi realizada. Os dados estão apresentados na Tabela 5.3.4, um gráfico de perfis individuais, na Figura 5.3.4 e um gráfico de perfis médios, na Figura 5.3.5.

Neste exemplo temos uma população não estratificada medida num ponto de referência (pré-teste), que corresponde à medição realizada logo após a ventilação com ar sintético. A cada indivíduo dessa população foi aplicado um de dois tipos de intervenção de forma aleatória. Essas intervenções correspondem à ventilação com *C* ou à ventilação com *T*. Neste cenário, o interesse é:

- a) comparar os efeitos (RHSR esperada pós-teste - RHSR esperada pré-teste) da ventilação com *T* e com *C* (interação entre tratamento e tempo),
- b) avaliar os efeitos da ventilação com *C* e com *T* individualmente no caso de existência de interação,

Tabela 5.3.4: Resistência homogênea do sistema respiratório

Rato	Gás	Pré-teste	Pós-teste	Rato	Gás	Pré-teste	Pós-teste
1	<i>T</i>	0.0828	0.0347	15	<i>C</i>	0.0477	0.0461
2	<i>T</i>	0.0274	0.0172	16	<i>C</i>	0.0313	0.0114
3	<i>T</i>	0.0115	0.0022	17	<i>C</i>	-0.0437	0.0417
4	<i>T</i>	0.0112	0.0157	18	<i>C</i>	0.0358	0.0021
5	<i>T</i>	0.0065	0.0149	19	<i>C</i>	0.0477	0.0488
6	<i>T</i>	0.0025	0.0000	20	<i>C</i>	0.0579	0.0591
7	<i>T</i>	0.0357	0.0147	21	<i>C</i>	0.0664	0.0986
8	<i>T</i>	0.0232	0.0291	22	<i>C</i>	0.0871	0.0529
9	<i>T</i>	0.0356	0.0092	23	<i>C</i>	0.0517	0.0577
10	<i>T</i>	0.0458	0.0490	24	<i>C</i>	0.1102	0.1018
11	<i>T</i>	0.0836	0.0208	25	<i>C</i>	0.0776	0.0329
12	<i>T</i>	0.0638	0.0221	26	<i>C</i>	0.0970	0.0110
13	<i>T</i>	0.0909	0.0972	27	<i>C</i>	0.0157	0.0258
14	<i>T</i>	0.0281	0.0204	28	<i>C</i>	0.1203	0.1171
				29	<i>C</i>	0.0279	0.0176

Ventilação com *T*: ar sintético/hélio-oxigênioVentilação com *C*: ar sintético/ar sintético

Figura 5.3.4: Perfis individuais para RHSR de ratos

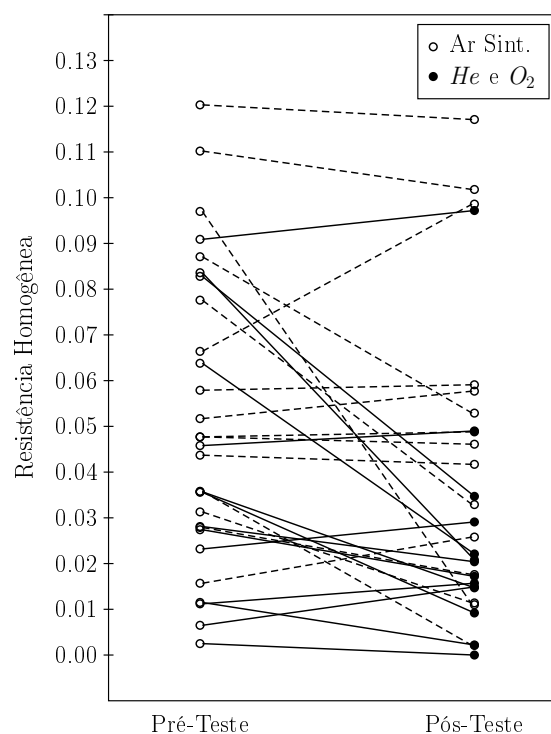
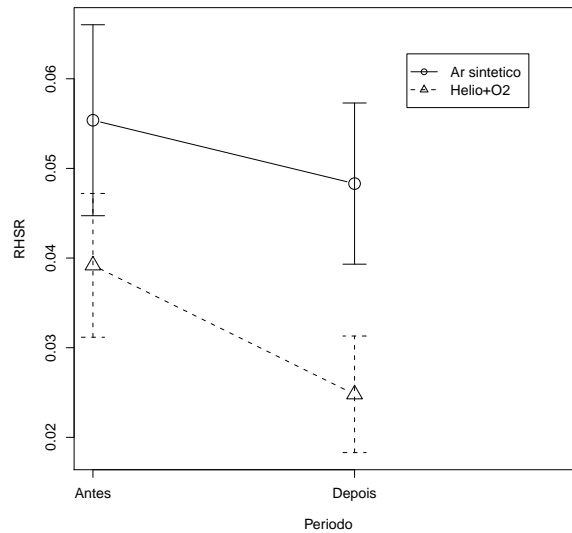
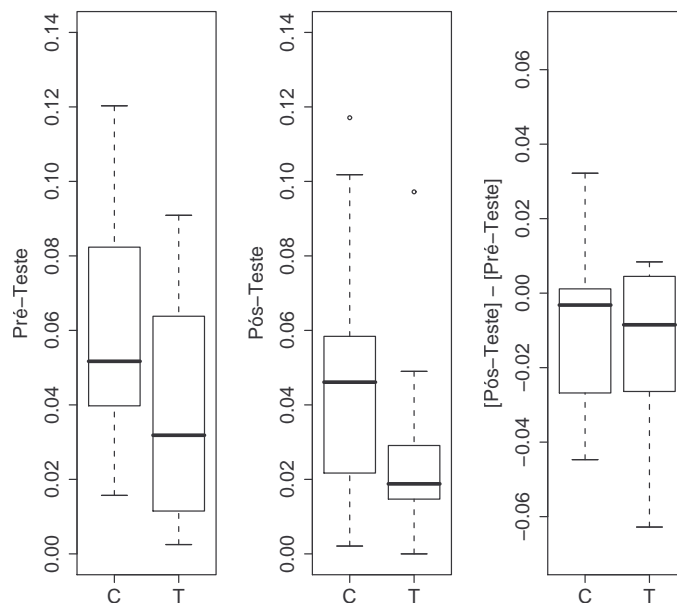


Figura 5.3.5: Perfis médios (desvios padrões) da RHSR.



Gráficos do tipo *boxplot* correspondentes estão dispostos na Figura 5.3.6.

Figura 5.3.6: Boxplots para a RHSR antes (pré-teste), depois (pós-teste) e efeito da ventilação com ar sintético, ou com a mistura hélio/oxigênio



Embora os modelos considerados acima possam ser empregados nesse caso, um

modelo de análise de covariância (ANCOVA) também pode ser utilizado para atacar as questões pertinentes, nomeadamente, (a) e (b). Para essa abordagem, a análise é condicional aos valores das respostas no pré-teste, y_{ij1} , $i = 1, 2$ $j = 1, \dots, n_i$, que admitimos fixas. Um modelo de ANCOVA é

$$y_{ij2} = \mu_2 + \alpha_i + \beta(y_{ij1} - \bar{y}_1) + e_{ij}, \quad i = 1, 2, \quad j = 1, \dots, n_i, \quad (5.3.4)$$

em que $\bar{y}_1 = n^{-1} \sum_{i=1}^2 \sum_{j=1}^{n_i} y_{ij1}$, $\alpha_1 + \alpha_2 = 0$, e_{ij} é um erro (de medida) aleatório tal que $\mathbb{E}(e_{ij}) = 0$, $\mathbb{V}(e_{ij}) = \sigma^2$ e as variáveis e_{ij} são não correlacionadas duas a duas, $i = 1, 2$ $j = 1, \dots, n_i$, e μ_2 , α_i e β correspondem, respectivamente, ao valor esperado da resposta no pós-teste quando não se considera tratamento (média geral), ao efeito do tratamento i no valor esperado da resposta e ao coeficiente de regressão linear comum (de y_{ij2} em função de y_{ij1}).

Uma das vantagens desse tipo de modelo é que aborda diretamente a pergunta: os efeitos dos tratamentos são equivalentes sob a suposição de que as unidades amostrais têm os mesmos valores da resposta pré-teste?

Um segundo modelo de ANCOVA é dado por

$$d_{ij} = \nu + \gamma_i + \delta(y_{ij1} - \bar{y}_1) + e_{ij}, \quad i = 1, 2, \quad j = 1, \dots, n_i, \quad (5.3.5)$$

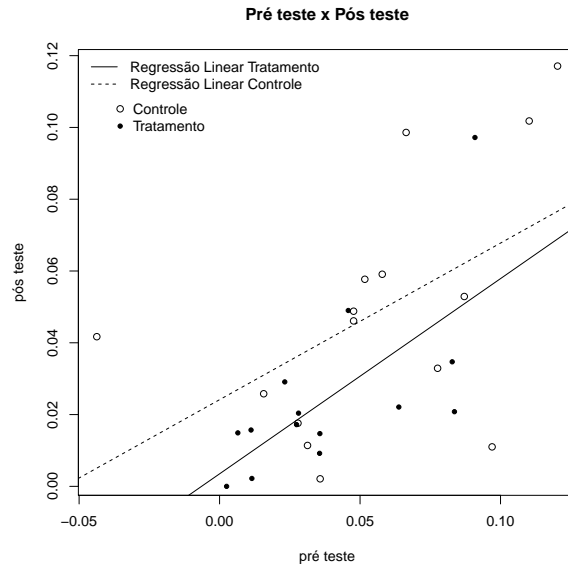
em que $d_{ij} = y_{ij1} - y_{ij2}$, $\bar{y}_1 = n^{-1} \sum_{i=1}^2 \sum_{j=1}^{n_i} y_{ij1}$, $\gamma_1 + \gamma_2 = 0$, e_{ij} é um erro (de medida) aleatório tal que $\mathbb{E}(e_{ij}) = 0$, $\mathbb{V}(e_{ij}) = \sigma^2$ e as variáveis e_{ij} são não correlacionadas duas a duas, $i = 1, 2$ $j = 1, \dots, n_i$, e ν , γ_i e δ correspondem, respectivamente, ao valor esperado da diferença entre respostas pós-teste e pré-teste quando tratamento não é levado em consideração, ao efeito do tratamento i no valor esperado da diferença entre respostas pós-teste e pré-teste e ao coeficiente de regressão linear comum (de y_{ij2} em função de y_{ij1}).

Ambos os modelos (5.3.4) e (5.3.5) produzem resultados equivalentes e podem ser ajustados com a utilização de sub-rotinas existentes em quase todos os pacotes computacionais (comerciais ou não). Em geral, esses pacotes geram testes para as hipóteses $\mu_2 = 0$ ou $\nu = 0$. No primeiro caso, a hipótese correspondente é de que o valor esperado ajustado da resposta no pós-teste é nulo e por isso, não tem interesse prático. No segundo caso, a hipótese correspondente é de que não há diferença entre os valores esperados ajustados da resposta no pós-teste e pré-teste.

Tanto o modelo (5.3.4) quanto o modelo (5.3.5) são aplicáveis quando os coeficientes angulares correspondentes às regressões das observações pós-teste em função das observações pré-teste são iguais. Em caso contrário as médias ajustadas dependerão do valor da medida pré-teste (\bar{y}_1) tomada como referência.

O diagrama de dispersão entre as observações pós-teste e pré-teste para os dados do Exemplo 5.3.2 está apresentado na Figura 5.3.7.

Figura 5.3.7: Diagrama de dispersão da RHSR, antes (pré-teste) *versus* depois (pós-teste) de submeter os ratos à ventilação com ar sintético, ou com hélio/oxigênio e retas ajustadas segundo o modelo (5.3.6)



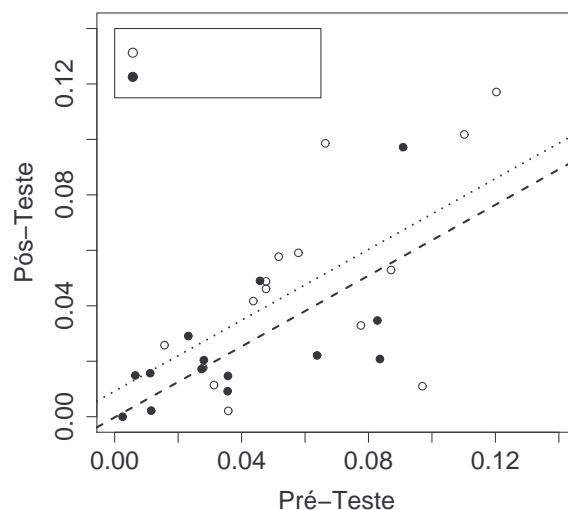
Com a finalidade de avaliar se ANCOVA pode ser empregada, consideramos o modelo

$$y_{ij2} = \alpha_i + \beta_i y_{ij1} + e_{ij} \quad (5.3.6)$$

com $e_{ij} \sim N(0, \sigma^2)$ independentes, cujo ajuste gerou um coeficiente de determinação $R^2 = 0.35$ e estimativas (\pm erro padrão) $\hat{\beta}_1 = 0.54 \pm 0.24$ e $\hat{\beta}_2 = 0.44 \pm 0.17$. Como o valor-P associado ao teste da razão de verossimilhanças para a hipótese $\beta_1 = \beta_2$ foi $P = 0.71$, concluímos que os modelos de ANCOVA descritos acima podem ser considerados como alternativa para a comparação de interesse.

O diagrama de dispersão para a RHSR dos ratos e as retas $\widehat{E}(y_{ij2}) = \hat{\mu}_2 + \hat{\alpha}_i + \hat{\beta}(y_{ij1} - \bar{y}_1)$, $i = 1, 2$, obtidas via mínimos quadrados estão apresentados na Figura 5.3.8.

Figura 5.3.8: Diagrama de dispersão da RHSR, antes (pré-teste) *versus* depois (pós-teste) de submeter os ratos à ventilação com ar sintético, ou com hélio/oxigênio, e retas ajustadas segundo o modelo (5.3.4)



Esse gráfico serve para ilustrar que a comparação entre as respostas esperadas pós-teste ajustadas independe do valor da resposta pré-teste tomada como referência.

Ao teste da hipótese $\alpha_i = 0$ no modelo (5.3.4) corresponde $P = 0.11$, o que sugere uma igualdade das respostas pós-teste esperadas ajustadas. A resposta esperada pós-teste comum ajustada pode ser obtida mediante o reajuste do modelo sem o termo α_i e um intervalo de confiança com coeficiente de confiança de 95% para esse valor é $XX \pm YY$. Como esperado, sob o modelo (5.3.5), ao teste da hipótese $\gamma_i = 0$ também corresponde $p = 0.11$, ilustrando a equivalência de ambos os modelos descritos com respeito à capacidade de detectar diferenças nas respostas pós-teste esperadas ajustadas. Neste caso, para o teste da hipótese $\nu = 0$, obtemos $P = 0.29$, o que sugere que não há diferença entre as respostas esperadas pós-teste e pré-teste ajustadas.

Detalhes sobre as relações identificadas acima além de outras considerações sobre a análise de estudos do tipo pré-teste/pós-teste podem ser encontrados em Brogan & Kutner (1980), Grieve (1981), Levin (1981), Schafer (1981), Kutner & Brogan (1981), Laird (1983) e Stanek III (1988).

5.4 Análise de perfis

O termo análise de perfis se refere ao conjunto de técnicas utilizadas para comparar p diferentes tratamentos (possivelmente definidos por meio do cruzamento dos níveis de dois ou mais fatores) com respeito à variação da resposta esperada ao longo de m diferentes condições de avaliação (tempo) em que as unidades amostrais são observadas. Em princípio, todas as unidades amostrais são observadas nos mesmos instantes, não necessariamente igualmente espaçados. Possíveis desbalanceamentos são devidos a observações omissas. Essencialmente, a análise tem como objetivo responder as seguintes questões:

- i) comparação dos diferentes tratamentos quanto à variação das respectivas distribuições de respostas ao longo do tempo, isto é, avaliação da existência de interação entre os fatores que definem os tratamentos e o fator tempo;
- ii) comparação dos diferentes tratamentos quanto às respectivas distribuições marginais (em relação ao tempo) de respostas, isto é, avaliação da existência de efeitos principais dos fatores que definem os tratamentos;
- iii) comparação dos diferentes instantes de avaliação quanto às respectivas distribuições marginais (em relação aos diferentes tratamentos) de respostas, isto é avaliação da existência de efeito principal do fator tempo;
- iv) identificação de modelos polinomiais que expliquem a variação das respostas médias (ou de outras características das distribuições das respostas) como função dos níveis do fator tempo.

Os objetivos (i)-(iii) podem ser operacionalizados por meio das hipóteses:

- H_I : os perfis médios de resposta correspondentes aos diferentes tratamentos são paralelos;
- H_T : os perfis médios de resposta correspondentes aos diferentes tratamentos são coincidentes;
- H_P : os perfis médios de resposta correspondentes aos diferentes tratamentos são paralelos ao eixo das abscissas.

Com essa finalidade, procuramos representar os parâmetros originais das distribuições de probabilidade geradoras dos dados por meio de modelos estruturais simples (isto é, com um número menor de parâmetros) que reflitam os aspectos fundamentais da variação da resposta. Nesse contexto, o perfil individual de resposta com as

m observações referentes à i -ésima unidade amostral ($i = 1, \dots, n_j$) submetida ao j -ésimo tratamento ($j = 1, \dots, p$) é dado por

$$\mathbf{y}_{ij} = (y_{ij1}, \dots, y_{ijm})^\top$$

em que y_{ijk} denota o valor da variável resposta para a i -ésima unidade amostral submetida ao j -ésimo tratamento no k -ésimo instante ($k = 1, \dots, m$). Os perfis médios de respostas nos m níveis da condição de avaliação são dados por $\mathbb{E}(\mathbf{y}_{ij}^\top) = \boldsymbol{\mu}_j = (\mu_{j1}, \dots, \mu_{jm})^\top$, $j = 1, \dots, p$.

Quando os dados são balanceados em relação ao tempo, *i.e.*, quando não há observações omissas, a forma mais adequada para representar o problema é (2.4.1) em que

$$\mathbf{M} = \begin{bmatrix} \mu_{11} & \mu_{12} & \dots & \mu_{1m} \\ \mu_{21} & \mu_{22} & \dots & \mu_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ \mu_{p1} & \mu_{p2} & \dots & \mu_{pm} \end{bmatrix} \quad (5.4.1)$$

e $\mathbf{X} = \oplus_{j=1}^p \mathbf{1}_{n_j}$.

Em termos dos parâmetros μ_{jk} , a hipótese i) pode ser expressa como

$$H_I : \begin{bmatrix} \mu_{11} - \mu_{12} \\ \mu_{12} - \mu_{13} \\ \vdots \\ \mu_{1(m-1)} - \mu_{1m} \end{bmatrix} = \begin{bmatrix} \mu_{21} - \mu_{22} \\ \mu_{22} - \mu_{23} \\ \vdots \\ \mu_{2(m-1)} - \mu_{2m} \end{bmatrix} = \dots = \begin{bmatrix} \mu_{p1} - \mu_{p2} \\ \mu_{p2} - \mu_{p3} \\ \vdots \\ \mu_{p(m-1)} - \mu_{pm} \end{bmatrix}.$$

Na ausência de interação, ou seja, quando não rejeitamos H_I , as hipóteses ii) e iii) podem ser definidas como

$$H_T : \sum_{k=1}^m \mu_{1k} = \sum_{k=1}^m \mu_{2k} = \dots = \sum_{k=1}^m \mu_{pk}$$

e

$$H_P : \sum_{j=1}^p \mu_{j1} = \sum_{j=1}^p \mu_{j2} = \dots = \sum_{j=1}^p \mu_{jm}.$$

Outra forma para representar as hipóteses ii) e iii) independentemente da presença ou não da interação é

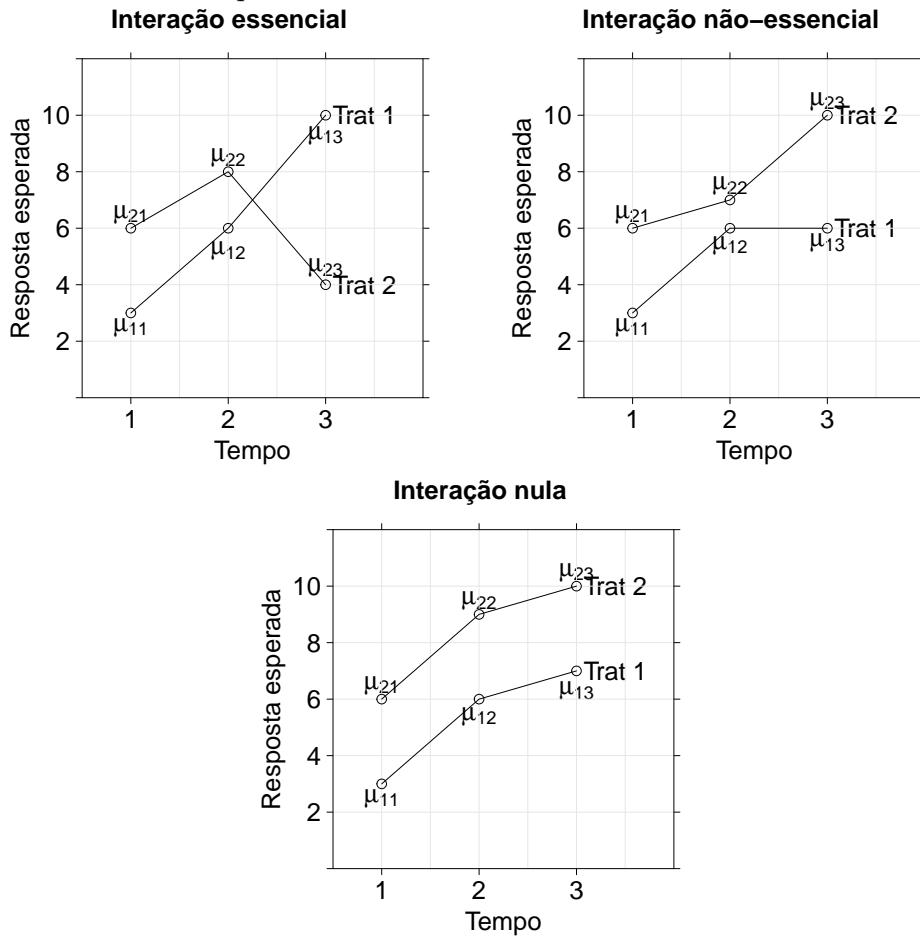
$$H_{TI} : \begin{bmatrix} \mu_{11} \\ \mu_{12} \\ \vdots \\ \mu_{1m} \end{bmatrix} = \begin{bmatrix} \mu_{21} \\ \mu_{22} \\ \vdots \\ \mu_{2m} \end{bmatrix} = \dots = \begin{bmatrix} \mu_{p1} \\ \mu_{p2} \\ \vdots \\ \mu_{pm} \end{bmatrix}$$

e

$$H_{PI}^* : \begin{bmatrix} \mu_{11} \\ \mu_{21} \\ \vdots \\ \mu_{p1} \end{bmatrix} = \begin{bmatrix} \mu_{12} \\ \mu_{22} \\ \vdots \\ \mu_{p2} \end{bmatrix} = \dots = \begin{bmatrix} \mu_{1m} \\ \mu_{2m} \\ \vdots \\ \mu_{pm} \end{bmatrix}.$$

As situações hipotéticas apresentadas na Figura 5.4.1, ilustram os seguintes casos:

Figura 5.4.1: Perfis médios de resposta correspondentes a situações hipotéticas com 2 tratamentos e 3 tempos



i) inexistência de interação e existência dos efeitos principais de tratamento e de tempo; *ii*) existência de interação (não-essencial), de efeitos principais de tratamento e de tempo e *iii*) existência de interação (essencial); neste caso não se pode falar em efeitos principais de tratamento e de tempo e sim em efeitos de tratamento (ou tempo) em cada nível de tempo (ou tratamento).

As hipóteses descritas acima podem ser expressas na forma (2.4.2). A matriz \mathbf{C} é utilizada para definir as comparações das respostas esperadas entre tratamentos e a matriz \mathbf{U} é utilizada para definir as comparações das respostas esperadas entre tempos. Lembrando que a escolha dessas matrizes para a definição de cada uma das hipóteses em questão não é única, podemos definir H_I por meio de (2.4.2) com

$$\mathbf{C} = \mathbf{C}_1 = \begin{bmatrix} 1 & -1 & 0 & \dots & 0 \\ 1 & 0 & -1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \dots & -1 \end{bmatrix},$$

uma matriz com dimensão $(p-1) \times p$ e

$$\mathbf{U} = \mathbf{U}_1 = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ -1 & 1 & 0 & \dots & 0 \\ 0 & -1 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ 0 & 0 & 0 & \dots & -1 \end{bmatrix},$$

uma matriz com dimensão $m \times (m-1)$. Para a hipótese H_{TI} , podemos utilizar

$$\mathbf{C} = \mathbf{C}_1 \text{ e } \mathbf{U} = \mathbf{1}_m$$

e para a hipótese H_{PI} ,

$$\mathbf{C} = \mathbf{1}_p^\top \text{ e } \mathbf{U} = \mathbf{U}_1.$$

Também podemos expressar modelos para análise de perfis na forma (2.1.7) com $\mathbf{X} = \oplus_{j=1}^p \mathbf{1}_{n_j} \otimes \mathbf{I}_m$, $\boldsymbol{\beta} = (\mu_{11}, \mu_{12}, \dots, \mu_{1m}, \dots, \mu_{p1}, \mu_{p2}, \dots, \mu_{pm})^\top$ e $\mathbf{Z} = \mathbf{0}$. As hipóteses H_I , H_{TI} e H_{PI} podem ser expressas na forma (2.3.32), que neste caso se reduz a

$$\mathbf{L}\boldsymbol{\beta} = \mathbf{0} \quad (5.4.2)$$

com $\mathbf{L} = \mathbf{C} \otimes \mathbf{U}^\top$.

Como exemplo, consideremos um estudo com dois tratamentos ($p = 2$) e três condições de avaliação ($m = 3$). Expressando o modelo na forma (2.4.1), a matriz de parâmetros é

$$\mathbf{M} = \begin{bmatrix} \mu_{11} & \mu_{12} & \mu_{13} \\ \mu_{21} & \mu_{22} & \mu_{23} \end{bmatrix};$$

as matrizes \mathbf{C} e \mathbf{U} associadas à hipótese H_I são

$$\mathbf{C} = \begin{bmatrix} 1 & -1 \end{bmatrix} \text{ e } \mathbf{U} = \begin{bmatrix} 1 & 0 \\ -1 & 1 \\ 0 & -1 \end{bmatrix}.$$

Sob a formulação (2.1.7) temos

$$\boldsymbol{\beta} = \text{vec}(\mathbf{M}) = [\mu_{11}, \mu_{12}, \mu_{13}, \mu_{21}, \mu_{22}, \mu_{23}]^\top.$$

e a matriz \mathbf{L} associada à hipótese H_I é

$$\mathbf{L} = \begin{bmatrix} 1 & -1 & 0 & -1 & 1 & 0 \\ 0 & 1 & -1 & 0 & -1 & 1 \end{bmatrix}.$$

A formulação (2.1.7) permite a especificação de hipóteses mais gerais que a formulação (2.4.1), pois no primeiro caso, \mathbf{L} pode ser qualquer matriz de posto completo.

Apresentamos a seguir análises de perfis de conjuntos de dados reais. Nossos objetivos são essencialmente didáticos e não pretendemos fazer uma análise exaustiva do problema considerado.

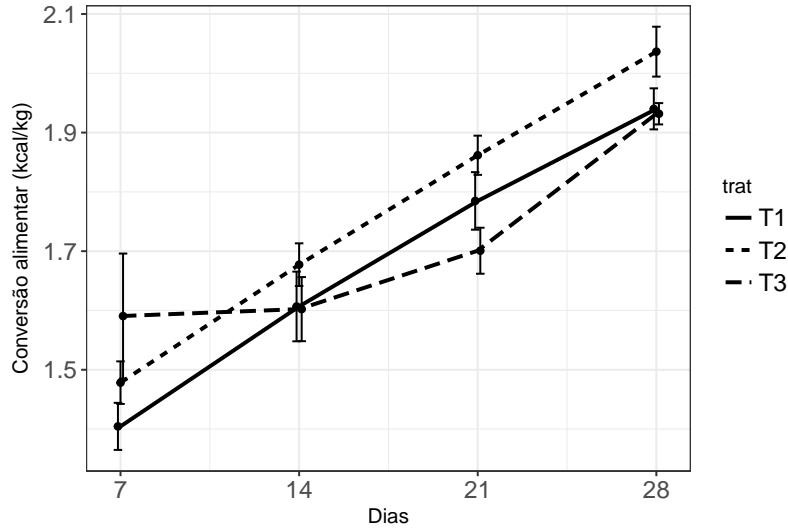
Exemplo 5.4.1: Os dados apresentados na Tabela 5.4.1 são oriundos de experimento na área de nutrição animal realizado no Centro Nacional de Pesquisa de Suínos e Aves da EMBRAPA. O principal objetivo é comparar os efeitos de duas dietas experimentais (T1 e T2) e de uma dieta controle (T3) no ganho de peso e consumo alimentar de frangos de corte. A resposta é a conversão alimentar média, expressa em kg de consumo de ração por kg de ganho de peso após 7, 14, 21 e 28 dias do início do experimento.

Tabela 5.4.1: Conversão alimentar (kg ração/kg ganho de peso)

Dieta	dia 7	dia 14	dia 21	dia 28
T1	1.399	1.526	1.720	1.899
T1	1.369	1.616	1.783	1.927
T1	1.389	1.618	1.800	1.955
T1	1.461	1.667	1.836	1.979
T2	1.427	1.627	1.826	1.981
T2	1.480	1.681	1.843	2.034
T2	1.501	1.689	1.880	2.049
T2	1.505	1.712	1.898	2.082
T3	1.484	1.549	1.658	1.910
T3	1.529	1.563	1.678	1.926
T3	1.629	1.641	1.730	1.939
T3	1.720	1.656	1.737	1.952

Um gráfico de perfis médios com barras de desvios padrões está apresentado na Figura 5.4.2.

Figura 5.4.2: Gráfico de perfis médios (com barras de desvios padrões) para os dados da Tabela 5.4.1



Para a análise sob o enfoque multivariado, consideramos o modelo $\mathbf{Y} = \mathbf{XM} + \mathbf{E}$ com

$$\mathbf{M} = \begin{bmatrix} \mu_{11} & \mu_{12} & \mu_{13} & \mu_{14} \\ \mu_{21} & \mu_{22} & \mu_{23} & \mu_{24} \\ \mu_{31} & \mu_{32} & \mu_{33} & \mu_{34} \end{bmatrix},$$

em que μ_{ij} representa a conversão alimentar esperada de animais após $7 \times j$ dias de submissão à dieta i ($i = 1$ correspondendo à dieta T1, $i = 2$, à dieta T2, $i = 3$, à dieta T3) e $\mathbf{X} = \bigoplus_{j=1}^3 \mathbf{1}_4$.

Estimativas de máxima verossimilhança dos parâmetros μ_{ij} são

$$\widehat{\mathbf{M}} = \begin{bmatrix} 1.405 & 1.607 & 1.785 & 1.940 \\ 1.478 & 1.677 & 1.862 & 2.037 \\ 1.591 & 1.602 & 1.701 & 1.932 \end{bmatrix}.$$

A hipótese de inexistência de interação entre os fatores dieta e tempo pode ser expressa na forma $\mathbf{CMU} = \mathbf{0}$ com

$$\mathbf{C} = \begin{bmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \end{bmatrix} \quad \text{e} \quad \mathbf{U} = \begin{bmatrix} 1 & 1 & 1 \\ -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}. \quad (5.4.3)$$

Os valores das estatísticas de teste para essa hipótese com os respectivos valores-P estão apresentados na Tabela 5.4.2 e indicam evidências de interação entre os

Tabela 5.4.2: Estatísticas de teste para a hipótese de inexistência de interação entre tratamento e tempo para os dados do Exemplo 5.4.1

Estatística	GL	Valor	Aproximação	GL	GL	Valor P
			F	Num	Denom	
Pillai	2	1.133	3.491	6	16	0.021
Wilks	2	0.021	10.690	6	14	< 0.001
Hotelling-Lawley	2	24.984	24.984	6	12	< 0.001
Roy	2	24.776	24.776	3	8	< 0.001

GL: graus de liberdade

Num: numerador

Denom: denominador

dois fatores. Uma estratégia para o prosseguimento da análise consiste em avaliar a natureza da interação. Um exame do gráfico de perfis médios sugere que as respostas sob os tratamentos T1 e T2 têm um comportamento longitudinal semelhante e a hipótese de inexistência de interação entre esses tratamentos e o tempo pode ser expressa com a matriz $\mathbf{C} = \begin{bmatrix} 1 & -1 & 0 \end{bmatrix}$ e a mesma matriz \mathbf{U} utilizada em (5.4.3). Nesse caso, as distribuições das quatro estatísticas multivariadas de teste podem ser aproximadas por uma distribuição F com 3 graus de liberdade no numerador e 7 graus de liberdade no denominador. O valor observado é $F=0.588$ com valor P igual a 0.642, sugerindo que não há evidências para rejeição da hipótese.

Nesse contexto, a hipótese de inexistência de efeito de tratamento (comparando T1 e T2) faz sentido e pode ser expressa com $\mathbf{C} = \begin{bmatrix} 1 & -1 & 0 \end{bmatrix}$ e $\mathbf{U} = \mathbf{1}_4$. Aqui, a estatística de teste tem distribuição (exata) F com 1 grau de liberdade no numerador e 9 graus de liberdade no denominador. O valor observado é $F=6.452$ com valor P igual a 0.032, sugerindo uma diferença entre as respostas esperadas sob os dois tratamentos. Outra hipótese de interesse é a de inexistência de efeito de tempo para o conjunto de tratamentos (T1, T2) que pode ser expressa com $\mathbf{C} = \begin{bmatrix} 1 & 1 & 0 \end{bmatrix}$ e \mathbf{U} como em (5.4.3). Assim como no caso da hipótese de inexistência de interação entre o conjunto de tratamentos (T1, T2) e o tempo, as distribuições das quatro estatísticas multivariadas de teste podem ser aproximadas por uma distribuição F com 1 grau de liberdade no numerador e 3 graus de liberdade no denominador. O valor observado é $F=373.4$ com valor $P < 0.001$, sugerindo um efeito significativo de tempo para esses dois tratamentos.

A forma do comportamento longitudinal da resposta esperada pode ser avaliada por meio de testes das hipóteses de efeito linear de tempo para os tratamentos T1 e T2 e de efeito quadrático para o tratamento T3. Uma maneira conveniente para expressar essas hipóteses envolve a utilização de polinômios ortogonais (ou ortonormais). Para a primeira delas, podemos utilizar $\mathbf{C} = \begin{bmatrix} 1 & 1 & 0 \end{bmatrix}$ e a matriz

de contrastes ortogonais

$$\mathbf{U} = \begin{bmatrix} 1 & -1 \\ -1 & 3 \\ -1 & -3 \\ 1 & 1 \end{bmatrix}$$

e, para a segunda, $\mathbf{C} = [0 \ 0 \ 1]$ e $\mathbf{U} = [-1 \ 3 \ -3 \ 1]^\top$. Para a hipótese de efeito linear de tempo para os tratamentos T1 e T2, todas as estatísticas multivariadas têm distribuição aproximada F com 2 graus de liberdade no numerador e 8 graus de liberdade no denominador. O valor observado é $F=3.675$ com quadrático de tempo para o tratamento T3, todas as estatísticas multivariadas têm distribuição aproximada F com 1 grau de liberdade no numerador e 9 graus de liberdade no denominador. O valor observado é $F=3.145$ com valor $P=0.110$. Nos dois casos, os testes não sugerem evidências contrárias às hipóteses nulas.

Como complementação da análise podemos testar a hipótese de que a conversão alimentar esperada no período de 28 dias é igual para os três tratamentos. Com essa finalidade, podemos utilizar \mathbf{C} como em (5.4.3) e $\mathbf{U} = [1/4 \ 1/4 \ 1/4 \ 1/4]^\top$. O valor observado da estatística F correspondente, a saber, 3.433 indica um valor $P=0.078$ obtido da distribuição F com 2 graus de liberdade no numerador e 9 graus de liberdade no denominador, sugerindo a homogeneidade da conversão alimentar esperada nesse período. Uma estimativa para essa conversão alimentar esperada é 1.718 kg ração/kg de peso, com limites inferior e superior para um intervalo de confiança aproximado (95%) iguais a 1.687 e 1.749, respectivamente.

Exemplo 5.4.2: Os dados considerados neste exemplo foram obtidos de um estudo cuja finalidade era avaliar o efeito do *clordiazepóxido* no mecanismo de bloqueio de ansiedade em ratos. Essa droga funciona como agente depressor do sistema nervoso central e é empregada clinicamente com tal fim. O estudo foi baseado numa amostra de 90 ratos de mesma linhagem com pesos e idades comparáveis e consistiu das três seguintes etapas:

Modelagem: 90 ratos foram condicionados a responder a estímulos elétricos de 1.3 mA com duração de 0.5 segundo a cada intervalo de 3.0 segundos. A resposta, conhecida na literatura como “resposta de esquia”, consiste no acionamento de uma barra apropriada, pelo animal, com a finalidade de interromper o estímulo elétrico e proporcionar um período inerte de 30 segundos em média. O objetivo desta etapa é familiarizar os animais com o procedimento e avaliar as diferenças entre eles relativamente ao número de estímulos elétricos gerados nesta etapa. O processo de modelagem é interrompido quando o animal aciona o mecanismo 3 vezes entre dois choques consecutivos.

Manutenção: nesta fase os animais foram treinados com o objetivo de atingir uma estabilização das taxas de resposta de esquia, ou seja da frequência de respostas de

esquiva por minuto. Nesse sentido os animais foram submetidos a estímulos elétricos nos padrões descritos acima e as taxas de resposta de esquiva foram observadas em 12 ocasiões correspondentes aos pontos finais de 12 intervalos consecutivos de 15 minutos cada.

Extinção: os 90 ratos foram aleatoriamente subdivididos em 9 conjuntos de 10 animais e cada um desses conjuntos foi submetido a um dos tratamentos descritos a seguir:

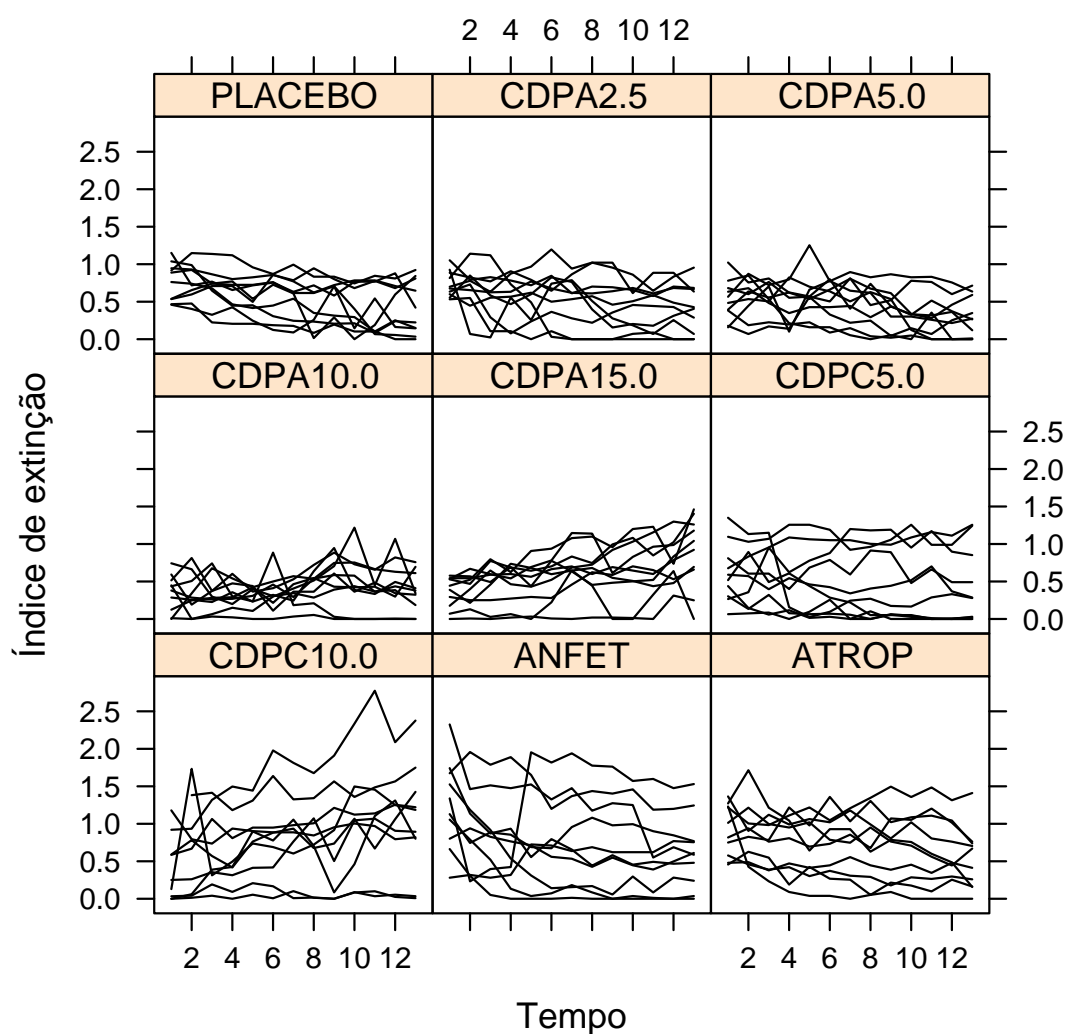
- **PLACEBO:** injeção de 1.0 mg/kg de solução salina;
- **CDPA2.5:** injeção de 2.5 mg/kg de clordiazepóxido;
- **CDPA5.0:** injeção de 5.0 mg/kg de clordiazepóxido;
- **CDPA10.0:** injeção de 10.0 mg/kg de clordiazepóxido;
- **CDPA15.0:** injeção de 15.0 mg/kg de clordiazepóxido;
- **CDPC5.0:** injeção de 5.0 mg/kg de clordiazepóxido após um pré-tratamento de 4 injeções com a mesma dosagem da droga durante 4 dias;
- **CDPC10.0:** injeção de 10.0 mg/kg de clordiazepóxido após um pré-tratamento de 4 injeções com a mesma dosagem da droga durante 4 dias;
- **ANFET:** injeção de 1.0 mg/kg de anfetamina;
- **ATROP:** injeção de 1.0 mg/kg de atropina.

As drogas utilizadas nos dois últimos tratamentos são estimulantes do sistema nervoso central e foram incluídas no estudo para efeito de comparação. Durante esta etapa foram observadas as taxas de resposta de esquiva em 13 ocasiões correspondentes aos pontos finais de 13 intervalos consecutivos de 30 minutos cada. A variável de interesse, neste caso, é o chamado índice de extinção, que corresponde ao quociente entre a taxa de resposta de esquiva obtida na ocasião avaliada e uma taxa padrão de resposta de esquiva obtida na etapa de manutenção. Essa taxa padrão corresponde à média das taxas de resposta de esquiva nas últimas 5 ocasiões daquela etapa.

Espera-se que o índice médio de extinção decresça com o tempo a não ser para animais que estejam sob o efeito de estimulantes, como aqueles submetidos aos tratamentos com anfetamina ou atropina. Tendo essa expectativa em vista, o objetivo do estudo está essencialmente associado à comparação dos tratamentos com relação às taxas de decréscimo desses índices.

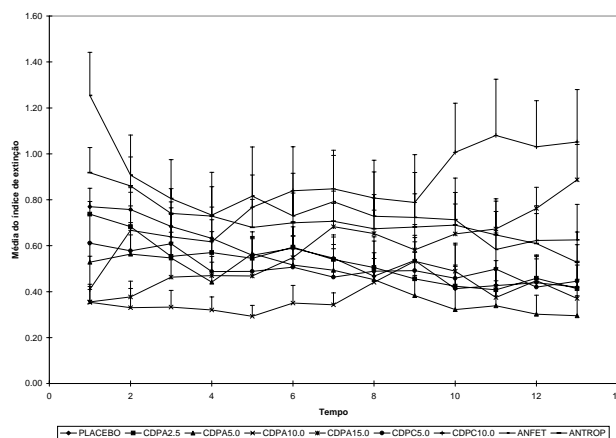
Gráficos de perfis individuais para os nove tratamentos estão apresentados na Figura 5.4.3 e sugerem a existência de efeitos aleatórios de animais dado o “paralelismo” dos perfis individuais. Grosso modo, pode-se observar que animais com índices de extinção acima (ou abaixo) da média tendem a manter esse padrão ao longo do tempo.

Figura 5.4.3: Perfis individuais para os nove tratamentos



O gráfico dos perfis médios apresentado na Figura 5.4.4 sugere uma possível interação entre tratamento e tempo.

Figura 5.4.4: Perfis médios para os nove tratamentos



A matriz de correlações intraunidades amostrais é

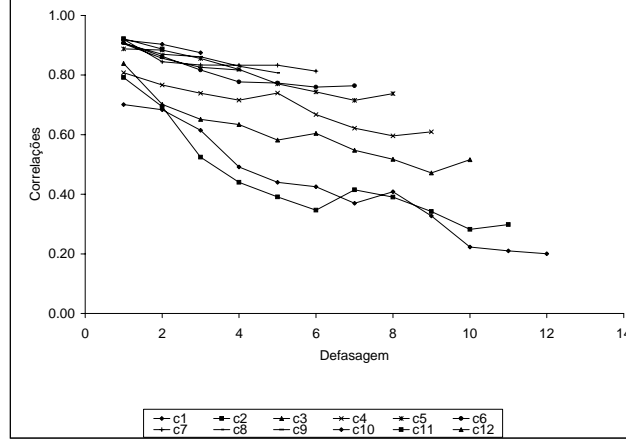
$$\begin{bmatrix}
 1 \\
 \mathbf{0.70} & 1 \\
 0.68 & \mathbf{0.79} & 1 \\
 0.61 & 0.69 & \mathbf{0.84} & 1 \\
 0.49 & 0.53 & 0.70 & \mathbf{0.81} & 1 \\
 0.44 & 0.44 & 0.65 & 0.77 & \mathbf{0.89} & 1 \\
 0.43 & 0.39 & 0.63 & 0.74 & 0.88 & \mathbf{0.91} & 1 \\
 0.37 & 0.35 & 0.58 & 0.72 & 0.86 & 0.86 & \mathbf{0.92} & 1 \\
 0.41 & 0.42 & 0.61 & 0.74 & 0.82 & 0.82 & 0.84 & \mathbf{0.91} & 1 \\
 0.33 & 0.39 & 0.55 & 0.67 & 0.77 & 0.78 & 0.83 & 0.87 & \mathbf{0.91} & 1 \\
 0.22 & 0.34 & 0.52 & 0.62 & 0.74 & 0.77 & 0.83 & 0.86 & 0.86 & \mathbf{0.92} & 1 \\
 0.21 & 0.28 & 0.47 & 0.60 & 0.72 & 0.76 & 0.83 & 0.83 & 0.83 & 0.90 & \mathbf{0.92} & 1 \\
 0.20 & 0.29 & 0.52 & 0.61 & 0.74 & 0.76 & 0.81 & 0.81 & 0.82 & 0.88 & 0.89 & \mathbf{0.91} & 1
 \end{bmatrix}
 \quad (5.4.4)$$

Seguindo a sugestão de Grady & Helms (1995), construímos gráficos de perfis para cada coluna da matriz de correlações, tendo como ordenadas os coeficientes de correlação e como abscissas as defasagens entre os instantes em que foram observadas as respostas. Esses gráficos estão resumidos na Figura 5.4.5.

Essas ferramentas permitem-nos observar que os valores dos coeficientes de correlação em cada coluna da matriz decrescem com a defasagem, sugerindo a presença de correlação serial. Além disso, notamos um aumento do valor dos coeficientes de correlação em cada subdiagonal. Esses padrões observados sugerem que a estrutura de antedependência para a matriz de correlações intraunidades amostrais.

Ainda com o objetivo de identificar uma estrutura de covariância adequada para

Figura 5.4.5: Correlações em função da defasagem



os dados, ajustamos o modelo

$$y_{ijk} = \mu_{ij} + e_{ijk}$$

em que, $i = 1, \dots, 10$ indexa as unidades amostrais (dentro de tratamento), $j = 1, \dots, 9$ indexa tratamentos (1=PLACEBO, 2=CDPA2.5, 3=CDPA5.0, 4=CDPA10.0, 5=CDPA15.0, 6=CDPC5.0, 7=CDPC10.0, 8=ANFET e 9=ATROP), $k = 1, \dots, 13$ indexa os tempos e e_{ijk} são erros aleatórios. Alternativamente, o modelo pode ser escrito como

$$\mathbf{y}_{ij} = \boldsymbol{\mu}_j + \mathbf{e}_{ij}, \quad (5.4.5)$$

em que \mathbf{y}_{ij} denota o vetor com os índices de extinção da i -ésima unidade amostral submetida ao j -ésimo tratamento ao longo dos k instantes de observação, $\boldsymbol{\mu}_j$ representa o vetor com as médias dos índices de extinção para o j -ésimo tratamento ao longo dos k instantes de observação e \mathbf{e}_{ij} representa o vetor com dos erros de medida associados, para o qual assumimos uma distribuição $N(\mathbf{0}; \mathbf{R})$.

Ainda com objetivo exploratório, adotamos as estruturas de covariância AR1, ARH1 e Antedependência de ordem 1 para a matriz \mathbf{R} . Consideramos também o ajuste de seguinte modelo com coeficientes lineares aleatórios

$$\mathbf{y}_{ij} = \boldsymbol{\mu}_j + \mathbf{Z}_i \mathbf{b}_i + \mathbf{e}_{ij}, \quad (5.4.6)$$

em que $\mathbf{Z}_i = \mathbf{1}_{13}$ e $\mathbf{R} = \mathbb{V}(\mathbf{e}_{ij})$ tem uma estrutura AR1 para incorporar a correlação serial observada; a estrutura de covariâncias marginal obtida com esse modelo é rotulada AR1CL. Estatísticas de ajuste dos modelos exploratórios adotados estão apresentados na Tabela 5.3.1 e confirmam as evidências favoráveis ao modelo de antedependência de ordem 1 para a matriz de covariâncias intraunidades amostrais.

Tabela 5.4.3: Estatísticas de ajuste de modelos com diferentes estruturas de covariâncias intraunidades amostrais

Estruturas de covariância	Número de parâmetros	AIC	BIC
AR1	2	70.8	75.8
ARH1	14	43.0	78.0
ANTE1	25	11.2	73.7
AR1CL	3	50.2	57.7
NE	91	-5.2	222.3

Adotando a matriz de covariâncias intraunidades amostrais identificada na Tabela 5.4.3, inicialmente testamos a hipótese de inexistência de interação entre Tratamento e Tempo (H_{0I}). Utilizando a aproximação F para a estatística (2.3.33) obtivemos $p = 0.0036$, o que sugere a rejeição da hipótese nula. Em situações semelhantes, muitos autores, como Lindsey (1999), sugerem o prosseguimento da análise por meio da realização da comparação dos tratamentos em cada instante de avaliação. Uma alternativa mais interessante corresponde a um estudo da **natureza da interação** por meio de um exame da estrutura dos tratamentos. No caso em consideração, uma justificativa para a interação entre Tratamento e Tempo é uma possível diferença no comportamento longitudinal dos perfis médios de resposta para os seguintes três grupos de tratamentos:

GRUPO 1: PLACEBO, CDPA2.5, CDPA5.0, ANFET, ATROP;

GRUPO 2: CDPA10.0, CDPC5.0;

GRUPO 3: CDPA15.0, CDPC10.0.

A composição desses grupos tem uma explicação biológica que não discutiremos aqui. A inclusão dos tratamentos ANFET e ATROP no mesmo grupo do PLACEBO pode ser justificada em função da pequena dosagem dos estimulantes administrada aos animais; caso uma dosagem maior tivesse sido considerada, esses tratamentos teriam sido incluídos no GRUPO 3. Uma representação gráfica dos índices de extinção médios correspondentes aos tratamentos de cada um desses três grupos é indicada na Figura 5.4.6. O tratamento PLACEBO foi incluído em todos os gráficos para comparação.

Tendo em vista esse agrupamento dos tratamentos, uma hipótese de interesse é a de paralelismo dos perfis médios de resposta dentro de cada um desses três grupos, isto é, a hipótese de inexistência de interação entre Tratamento e Tempo dentro de

Tabela 5.4.4: Testes para as hipóteses de existência de efeito de Tratamento e de Tempo dentro de grupos

Hipótese	C	U	F	gl	P
Tratamento	(5.4.7)	$\mathbf{1}_{13}$	1.75	(6, 963)	0.106
Tempo Grupo 1	(1 1 1 0 0 0 0 1 1)	(5.4.8)	2.85	(12, 963)	0.001
Tempo Grupo 2	(0 0 0 1 0 1 0 0 0)	(5.4.8)	0.76	(12, 963)	0.697
Tempo Grupo 3	(0 0 0 0 1 0 1 0 0)	(5.4.8)	3.64	(12, 963)	< 0.001

cada grupo. Essa hipótese pode ser expressa na forma (5.4.2), com $\mathbf{L} = \mathbf{C} \otimes \mathbf{U}^\top$ em que

$$\mathbf{C} = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 & 0 \end{bmatrix} \quad (5.4.7)$$

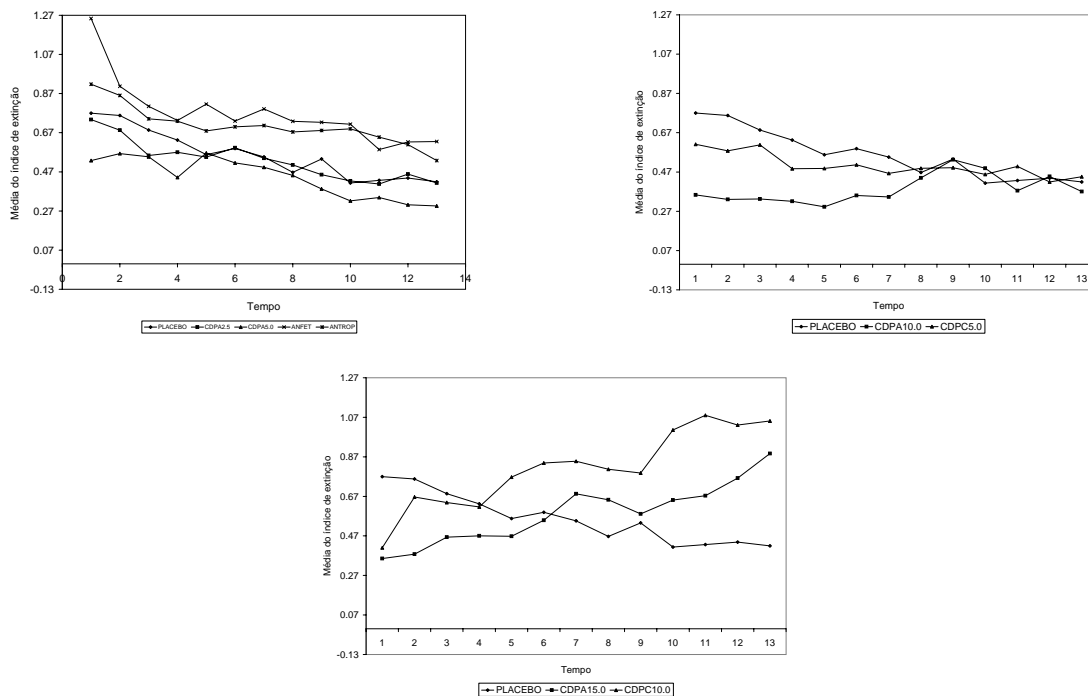
e

$$\mathbf{U} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 \end{bmatrix} \quad (5.4.8)$$

O valor-P correspondente ao teste dessa hipótese é $p = 0.7223$, e sugere que não há razões para a sua rejeição. Como consequência, podemos considerar paralelos os perfis médios de resposta dos tratamentos dentro de cada um dos grupos descritos acima e então faz sentido testar hipóteses sobre a inexistência de efeito de tratamentos nos moldes de H_{TI} e de inexistência de efeito de tempo nos moldes de H_{PI} dentro de cada um dos três grupos. A especificação dessas hipóteses na forma (5.4.2) com $\mathbf{L} = \mathbf{C} \otimes \mathbf{U}^\top$, assim como os valores das estatísticas F aproximadas, os respectivos graus de liberdade e os valores-P dos testes correspondentes estão indicados na Tabela 5.3.2.

Esses resultados sugerem que:

Figura 5.4.6: Perfis médios para os três grupos de tratamentos



- i) os 9 tratamentos podem ser classificados em 3 grupos de tratamentos equivalentes;
- ii) tratamentos pertencentes a grupos diferentes têm efeitos diferentes;
- iii) somente os tratamentos do segundo grupo produzem respostas médias essencialmente constantes ao longo das condições de avaliação.

Na Tabela 5.3.4 apresentamos as médias e respectivos erros padrões estimados sob o modelo que incorpora o resultado da análise.

Tabela 5.4.5: Médias estimadas (erros padrões) para os três grupos de tratamentos

Grupo	Tempo	Erro		Grupo	Tempo	Erro	
		Estimativa	padrão			Estimativa	padrão
1	1	0.84	0.05	3	1	0.36	0.09
1	2	0.75	0.05	3	2	0.52	0.09
1	3	0.67	0.05	3	3	0.55	0.08
1	4	0.62	0.05	3	4	0.54	0.08
1	5	0.63	0.06	3	5	0.62	0.09
1	6	0.63	0.06	3	6	0.69	0.09
1	7	0.62	0.06	3	7	0.77	0.09
1	8	0.57	0.06	3	8	0.73	0.09
1	9	0.56	0.06	3	9	0.68	0.10
1	10	0.51	0.06	3	10	0.83	0.10
1	11	0.48	0.06	3	11	0.88	0.10
1	12	0.49	0.06	3	12	0.90	0.09
1	13	0.46	0.06	3	13	0.97	0.10
2	1-13	0.44	0.07				

5.5 Análise de medidas repetidas

Nesta seção apresentamos um exemplo em que as medidas repetidas não são longitudinais. O objetivo é detalhar características do modelo por meio de um exemplo simples em que tanto os estimadores dos efeitos fixos quanto os preditores dos efeitos aleatórios têm expressões explícitas, além de mostrar como utilizar técnicas de diagnóstico para modificar o modelo.

Exemplo 5.5.1: Os dados da Tabela 5.5.1 são oriundos de um estudo realizado no Laboratório de Poluição Atmosférica Experimental da faculdade de Medicina da Universidade de São Paulo com o objetivo de avaliar a concentração de ozônio (como indicador de poluição atmosférica) em diferentes períodos. Em função do alto custo dos equipamentos para medida direta da concentração atmosférica de ozônio, a reflectância medida em filtros passivos impregnados com uma solução de índigo-carmim, cuja coloração desbota quando são expostos ao poluente, foi utilizada como uma forma indireta de medida. A intensidade da descoloração medida por intermédio de um reflectômetro foi utilizada como variável resposta.

Cada um de 9 conjuntos de 3 filtros foi colocado ao lado de uma estação em que o equipamento padrão para medir a concentração de ozônio estava disponível em nove períodos de 7 dias com espaçamento de 10 dias entre eles. Depois de cada período de 7 dias, as respostas de cada um dos 3 filtros foram registradas.

A análise dos dados deve fornecer um valor latente para a reflectância esperada em cada período. Esses valores latentes serão utilizados para a construção de uma

curva de calibração. Dadas as características de construção dos filtros, as medidas podem ser influenciadas por fatores como chuva, vento etc. e por isso, deve-se levar em conta a presença de dados discrepantes.

Tabela 5.5.1: Reflectância de filtros passivos

Período	Filtro	Reflectância	Período	Filtro	Reflectância
1	1	27.0	6	1	47.9
1	2	34.0	6	2	60.4
1	3	17.4	6	3	47.3
2	1	24.8	7	1	50.4
2	2	29.9	7	2	50.7
2	3	32.1	7	3	55.9
3	1	35.4	8	1	54.9
3	2	63.2	8	2	43.2
3	3	27.4	8	3	52.1
4	1	51.2	9	1	38.8
4	2	54.5	9	2	59.9
4	3	52.2	9	3	61.1
5	1	77.7			
5	2	53.9			
5	3	48.2			

Um modelo adequado para os propósitos do estudo é

$$y_{ij} = \mu + a_i + e_{ij}, \quad i = 1, \dots, 9, \quad j = 1, 2, 3, \quad (5.5.1)$$

em que y_{ij} denota a reflectância observada no filtro j no período i , μ é a reflectância esperada (populacional), $a_i \sim N(0, \sigma_a^2)$ e $e_{ij} \sim N(0, \sigma^2)$, com independência entre todos os termos aleatórios. O coeficiente $\rho = \sigma_a^2 / (\sigma_a^2 + \sigma^2)$ é conhecido como **coeficiente de correlação intraclasse**.

Esse modelo pode ser expresso no formato (2.1.2) com

$$\mathbf{y}_i = \begin{pmatrix} y_{i1} \\ y_{i2} \\ y_{i3} \end{pmatrix}, \quad \mathbf{X}_i = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \quad \boldsymbol{\beta} = \mu, \quad \mathbf{b}_i = a_i, \quad \mathbf{e}_i = \begin{pmatrix} e_{i1} \\ e_{i2} \\ e_{i3} \end{pmatrix}$$

$$\mathbf{G} = \sigma_a^2, \quad \mathbf{R}_i = \mathbf{I}_3 \sigma^2, \quad \boldsymbol{\Omega}_i = \begin{pmatrix} \sigma_a^2 + \sigma^2 & \sigma_a^2 & \sigma_a^2 \\ \sigma_a^2 & \sigma_a^2 + \sigma^2 & \sigma_a^2 \\ \sigma_a^2 & \sigma_a^2 & \sigma_a^2 + \sigma^2 \end{pmatrix}$$

Dada a simplicidade do problema sob investigação é possível expressar estimadores e preditores em forma explícita; em particular,

$$\hat{\beta} = \hat{\mu} = \bar{y} = (9 \times 3)^{-1} \sum_{i=1}^9 \sum_{j=1}^3 y_{ij}, \quad \text{e} \quad \hat{\mathbf{b}}_i = \hat{a}_i = \hat{k}(\bar{y}_i - \bar{y})$$

com

$$\bar{y}_i = 3^{-1} \sum_{j=1}^3 y_{ij} \quad \text{e} \quad \hat{k} = \hat{\sigma}_a^2 / (\hat{\sigma}_a^2 + \hat{\sigma}^2 / 3).$$

A constante \hat{k} é conhecida como **constante de encolhimento**. O valor latente predito para o período i é

$$\hat{y}_i = \bar{y} + \hat{k}(\bar{y}_i - \bar{y}) = \hat{k}\bar{y}_i + (1 - \hat{k})\bar{y}.$$

Note que o preditor \hat{y}_i é uma combinação convexa entre a média amostral para o período i , \bar{y}_i e a média geral, \bar{y} . Quando a constante de encolhimento tem valor próximo de 1, *i.e.*, quando a variância interunidades amostrais ($\hat{\sigma}_a^2$) é muito maior que a variância intraunidades amostrais ($\hat{\sigma}^2$), o preditor do valor latente para o período i se aproxima de sua média amostral (\bar{y}_i); em caso contrário, quando a constante de encolhimento tem valor próximo de 0, o preditor do valor latente para o período i se aproxima da média geral (\bar{y}).

Com o ajuste do modelo (5.5.1) aos dados da Tabela 5.5.1 por meio de máxima verossimilhança restrita, obtemos $\hat{\mu} = 46.4$, $\hat{\sigma}_a^2 = 100.4$, $\hat{\sigma}^2 = 104.8$, $\hat{k} = 0.75$. As médias amostrais para os 9 períodos assim como s valores latentes preditos sob o modelo (5.5.1) encontram-se nas colunas 2 e 3 da Tabela 5.5.2.

O gráfico do índice de Lesaffre-Verbeke modificado apresentado na Figura (5.5.1) sugere que a estrutura de covariância proposta pode não ser adequada para os períodos 3 e 5.

Dado o planejamento experimental, não há razões para considerar uma modificação da estrutura de covariância interunidades amostrais definida por σ_a^2 . Tendo em vista que variáveis não controladas como direção do vento e umidade podem afetar a variabilidade intraunidades amostrais, propomos modificar o modelo (5.5.1) com a inclusão de variâncias específicas para os períodos detectados por meio do gráfico do índice de Lesaffre-Verbeke modificado. Mais especificamente, consideramos o modelo

$$y_{ij} = \mu + a_i + e_{ij} \text{ with } e_{ij} \sim N(0, \sigma_i^2), \quad (5.5.2)$$

com $\sigma_i^2 = \tau^2$, $i = 3, 5$ ou $\sigma_i^2 = \sigma^2$, em caso contrário. Nesse contexto, a constante de encolhimento é

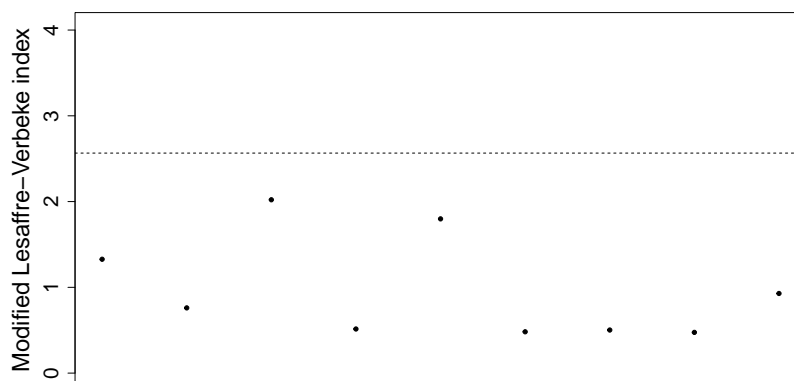
$$k_i = \sigma_a^2 / (\sigma_a^2 + \sigma_i^2 / 3)$$

Tabela 5.5.2: Preditores para os valores latentes dos 9 períodos (Exemplo 5.4.1)

Período	Média amostral	MLM (5.5.1) homocedástico	MLM (5.5.2) (3 variâncias)	Modelo t (gl=21.8)	Modelo Robusto
1	26.1	31.4	31.9	31.8	29.2
2	28.9	33.4	30.2	33.8	31.6
3	42.0	43.1	44.0	43.3	<i>39.4</i>
4	52.6	51.0	52.2	50.9	51.5
5	59.9	56.4	53.5	56.3	54.4
6	51.9	50.4	51.5	50.4	50.9
7	52.3	50.8	51.9	50.8	51.3
8	50.1	49.1	49.8	49.1	49.4
9	53.3	51.5	51.3	51.4	52.7
Média	46.4	46.4	46.2	46.6	45.7

Figura 5.5.1: Gráfico do índice de Lesaffre-Verbeke modificado *versus* índices de períodos para o modelo (5.5.1)

Linha pontilhada: 3o quartil + 1.5 distância interquartis

e o valor latente predito para o período i é

$$\hat{y}_i = \hat{\mu} + \hat{k}_i(\bar{y}_i - \hat{\mu})$$

com $\hat{\mu} = \sum_{i=1}^9 (w_i / \sum_{i=1}^9 w_i) \bar{y}_i$, e $w_i = (\sigma_a^2 + \sigma_i^2)^{-1}$. O resultado do ajuste do modelo (5.5.2) aos dados da Tabela 5.5.1 são $\hat{\mu} = 45.9$, $\hat{\sigma}_a^2 = 114.3$, $\hat{\sigma}^2 = 49.6$, $\hat{\tau}^2 = 274.0$, $\hat{k}_{i \neq 3,5} = 0.87$, $\hat{k}_{i=3,5} = 0.56$ e o gráfico do índice de Lesaffre-Verbeke modificado correspondente está disposto na Figura 5.5.2, sugerindo agora que a estrutura de covariância pode não ser adequada para os dados dos períodos 1 e 9.

Figura 5.5.2: Gráfico do índice de Lesaffre-Verbeke modificado *versus* índices de períodos para o modelo (5.5.2)

Linha pontilhada: 3o quartil + 1.5 distância interquartis



O modelo (5.5.2) agora com variâncias diferentes para esses períodos, *i.e.*, com $\sigma_i^2 = \tau^2, i = 3, 5$, $\sigma_i^2 = \nu^2, i = 1, 9$ e $\sigma_i^2 = \sigma_a^2, i = 2, 4, 6, 7, 8$, em caso contrário foi ajustado aos dados gerando os seguintes resultados: $\hat{\mu} = 46.2$, $\hat{\sigma}_a^2 = 103.8$, $\hat{\sigma}^2 = 123.6$, $\hat{\tau}^2 = 270.0$, $\hat{\nu}^2 = 23.3$. As constantes de encolhimento estimadas são $\hat{k}_{i=1,9} = 0.72$, $\hat{k}_{i=3,5} = 0.54$, $\hat{k}_{i=2,4,6,7,8} = 0.93$.

Gráficos de resíduos correspondentes a esse modelo estão apresentados nas Figuras 5.5.3 - 5.5.11 e não sugerem violações grosseiras das suposições subjacentes.

Preditores dos valores latentes correspondentes aos 9 períodos obtidos sob esse modelo estão dispostos na quarta coluna da Tabela 5.5.2, juntamente com preditores obtidos por meio do ajuste de um modelo baseado na distribuição t e de outro baseado em métodos robustos. Pode-se notar o efeito das constantes de encolhimento nos valores indicados em **negrito**. Eles estão mais próximos da média geral do que as correspondentes médias amostrais.

Figura 5.5.3: Gráfico do índice de Lesaffre-Verbeke modificado *versus* índices de períodos para o modelo (5.5.2) com componente de variância adicional para os períodos 1 e 9

Linha pontilhada: 3o quartil + 1.5 distância interquartis

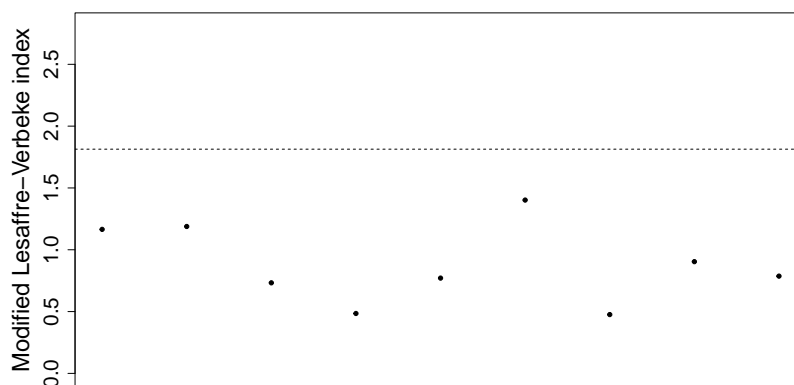


Figura 5.5.4: Gráfico de resíduos marginais padronizados *versus* índices de observações e histograma para o modelo (5.5.2) com componente de variância adicional para os períodos 1 e 9

Linha pontilhada: 3o quartil + 1.5 distância interquartis

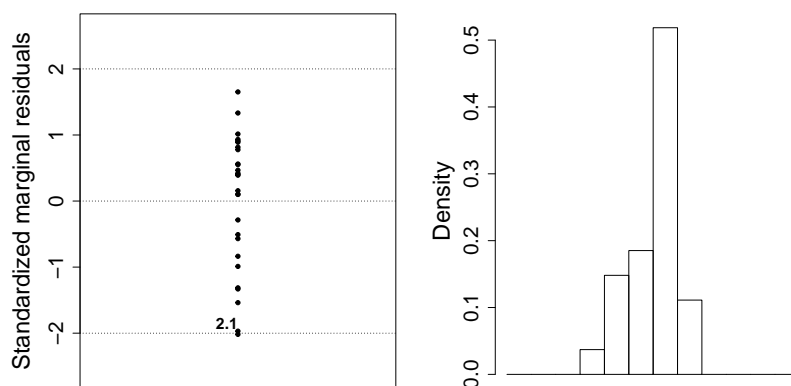


Figura 5.5.5: Gráfico da distância de Mahalanobis *versus* índices de períodos correspondentes ao modelo (5.5.2) com componente de variância adicional para os períodos 1 e 9

Linha pontilhada: 3o quartil + 1.5 distância interquartis

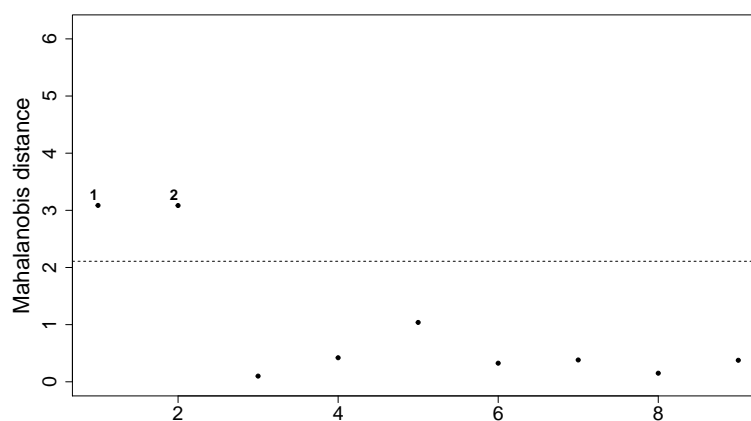


Figura 5.5.6: Gráfico QQ para a distância de Mahalanobis correspondente ao modelo (5.5.2) com componente de variância adicional para os períodos 1 e 9

Linha pontilhada: 3o quartil + 1.5 distância interquartis

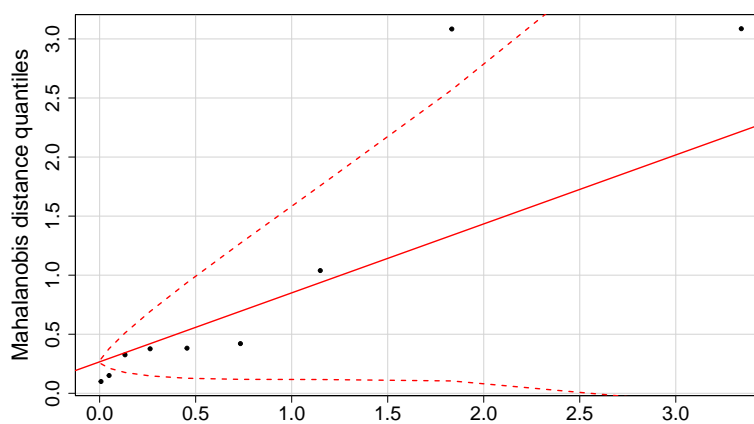


Figura 5.5.7: Gráfico QQ e histograma para resíduos condicionais minimamente confundidos para correspondente ao modelo (5.5.2) com componente de variância adicional para os períodos 1 e 9

Linha pontilhada: 3o quartil + 1.5 distância interquartis

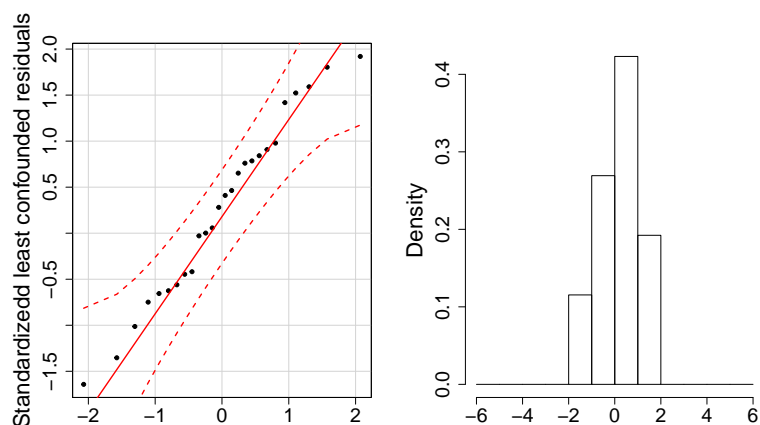


Figura 5.5.8: Gráfico da distância de Cook para correspondente ao modelo (5.5.2) com componente de variância adicional para os períodos 1 e 9

Linha pontilhada: 3o quartil + 1.5 distância interquartis

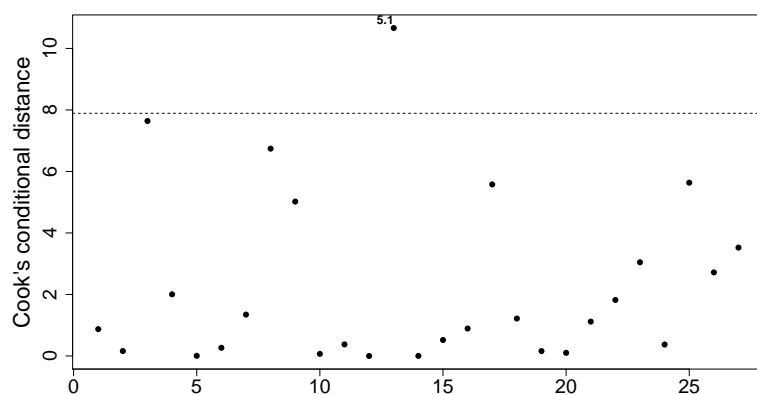


Figura 5.5.9: Gráfico da distância de Cook condicional para estimação do efeito fixo correspondente ao modelo (5.5.2) com componente de variância adicional para os períodos 1 e 9

Linha pontilhada: 3o quartil + 1.5 distância interquartis

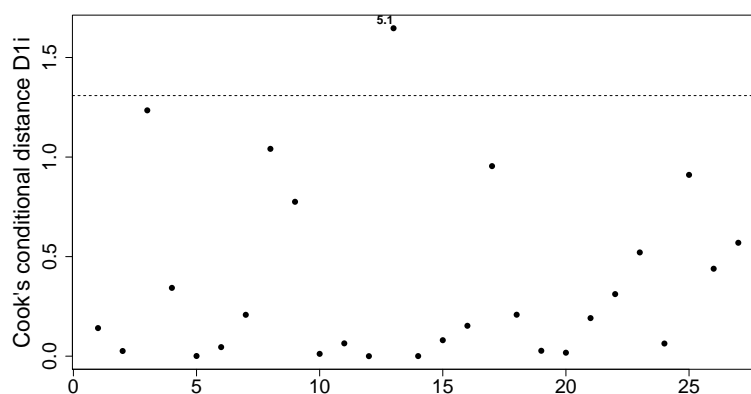


Figura 5.5.10: Gráfico da distância de Cook condicional para predição dos efeitos aleatórios correspondente ao modelo (5.5.2) com componente de variância adicional para os períodos 1 e 9

Linha pontilhada: 3o quartil + 1.5 distância interquartis

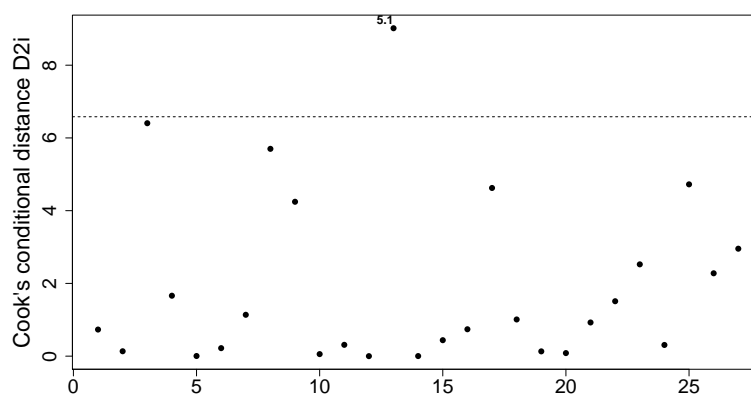
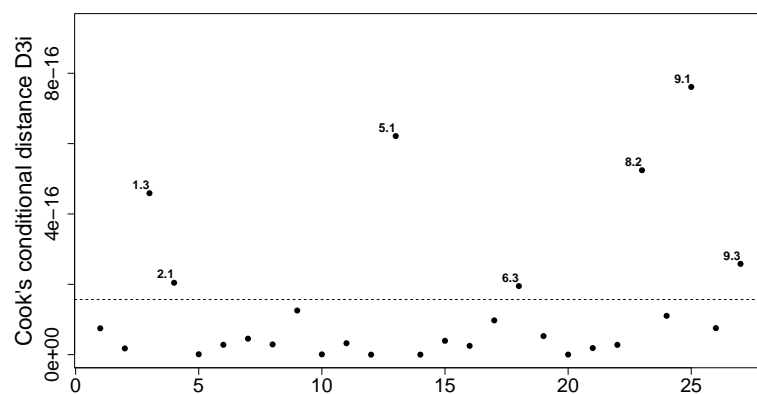


Figura 5.5.11: Gráfico da distância de Cook condicional para avaliação da correlação entre o estimador do efeito fixo e preditores dos efeitos aleatórios correspondente ao modelo (5.5.2) com componente de variância adicional para os períodos 1 e 9
Linha pontilhada: 3o quartil + 1.5 distância interquartis



5.6 Ajuste de curvas de crescimento

O objetivo desta seção é detalhar a análise de um conjunto de dados longitudinais cujo objetivo é ajustar curvas polinomiais de crescimento. A identificação tanto do grau do polinômio adotado quanto dos efeitos aleatórios a serem incluídos na análise é realizada por meio das sugestões apresentadas em Rocha & Singer (2017).

Exemplo 5.6.1: Consideramos aqui o problema descrito no Exemplo 1.2.6. Os perfis individuais e médios bem como as curvas alisadas estão dispostos nas Figuras 1.3.3 e 1.3.4. A escolha do grau dos polinômios para representar o crescimento do diâmetro sistólico da aorta é complicada pelo padrão irregular dos perfis, embora seja razoável supor que curvas do segundo grau com efeitos aleatórios dos interceptos, coeficientes angulares e quadráticos possam ser considerados como candidatos.

Com a finalidade de fundamentar a escolha, iniciamos com o ajuste de polinômios quadráticos a cada perfil individual com a exclusão das unidades amostrais com três observações ou menos (6 no grupo AIG e 9 no grupo PIG), a saber, aquelas com identificação ind=10, 12, 13, 15, 23, 25, 34, 38, 43, 44, 45, 46, 49, 51 e 60 no arquivo Singer&Nobre&Rocha2017exemp126.xls, disponível em www.ime.usp.br/~jmsinger/Dados.

Para reduzir possíveis problemas relacionados com multicolinearidade, transformamos a variável tempo fazendo $t_{ij}^* = (t_{ij} - 33)/4.29$ de forma que sua origem fica localizada na semana 33 e a unidade correspondente é mês em substituição a semana (note que $4.29 = 30 \text{ dias} / 7 \text{ dias por semana}$). Adotamos um nível de significância de 5% para selecionar os coeficientes fixos; para a construção dos intervalos de referência corrigidos pela técnica de Bonferroni utilizados para a seleção dos efeitos aleatórios, adotamos níveis de significância de 5% e também de 1% para sermos mais conservadores. Estimativas dos parâmetros e valores das estatísticas t correspondentes para os polinômios ajustados às unidades amostrais dos grupos AIG e PIG estão dispostos nas Tabelas 5.6.1 e 5.6.2.

Como apenas as médias dos interceptos e coeficientes angulares das curvas individuais para o grupo AIG são significativos [$(\alpha = 1.67\% (= 0.05/3))$], esses dois termos são os candidatos para efeitos fixos no modelo.

Com a adoção de um nível de significância global $\alpha = 5\%$ e a correção de Bonferroni, o nível de significância individual para as $69 = (23 \times 3)$ comparações é $\alpha^* = 0.0003623 (= 0.05/69)$, de forma que o intervalo de referência correspondente, obtido da distribuição t com $\nu = 54 (= 123 - 69)$ graus de liberdade é $(-3.6, 3.6)$.

Os resultados da Tabela 5.6.1 indicam que 52% ($=12/23$) dos interceptos estimados, 13% ($=3/23$) dos coeficientes angulares estimados não pertencem ao intervalo de referência; por outro lado, todos os coeficientes quadráticos pertencem a esse intervalo. Esses resultados sugerem que apenas os interceptos e os coeficientes an-

gulares são candidatos a efeitos aleatórios. Um intervalo mais rigoroso, $(-4.1, 4.1)$, obtido com nível de significância individual $\alpha = 1\%$ conduz à mesma decisão.

Com propósito confirmatório, ajustamos modelos com interceptos, coeficientes angulares e quadráticos fixos e aleatórios aos dados e testamos hipóteses para avaliar se os termos quadráticos poderiam ser eliminados. O teste padrão correspondente à hipótese de nulidade do efeito quadrático fixo indica $p = 0.30$. Para o efeito quadrático aleatório, utilizamos o teste proposto por Stram & Lee (1994), obtendo $p = 0.94$. Esses dois resultados sugerem que polinômios do primeiro grau tanto para os termos fixos quanto para os aleatórios são adequados para a análise.

Uma análise similar foi conduzida para os dados referentes ao grupo PIG. Estimativas dos parâmetros e valores t para os coeficientes de um polinômio quadrático ajustado aos dados desse grupo estão apresentados na Tabela 5.6.3 e sugerem que interceptos, coeficientes angulares e coeficientes do segundo grau, tanto fixos quanto aleatórios, devem ser incluídos no modelo.

Os resultados do ajuste de um modelo que incorpora as sugestões obtidas da análise descritiva estão apresentados na Tabela 5.6.3 sob o rótulo *Modelo 1* e sugerem que o padrão de crescimento do diâmetro sistólico esperado da aorta difere entre os neonatos pré-termo nascidos com peso adequado ou pequenos para a idade gestacional.

Também ajustamos outros 21 modelos obtidos a partir de um modelo homocedástico com independência condicional global por meio de eliminação ou combinação de termos. A especificação dos termos de cada um deles está apresentada na Tabela 5.6.4.

O modelo global é

$$y_{ijk} = \alpha_i + \beta_i t_{ijk}^* + \gamma_i t_{ijk}^{*2} + a_{ij} + b_{ij} t_{ijk}^* + c_{ij} t_{ijk}^{*2} + e_{ijk}, \quad i = 1, 2, \quad j = 1, \dots, n_i, \quad k = 1, \dots, m_{ij} \quad (5.6.1)$$

em que y_{ijk} denota a k -ésima observação do diâmetro da aorta do j -ésimo recém-nascido no i -ésimo grupo ($i = 1$: AIG, $i = 2$: PIG), t_{ijk}^* denota o tempo (centrado na semana 33, em meses) no qual a observação foi realizada, α_i , β_i e γ_i denotam, respectivamente, os coeficientes lineares, angulares e quadráticos, a_{ij} , b_{ij} and c_{ij} , representam, respectivamente os correspondentes coeficientes aleatórios associados ao j -ésimo recém-nascido no i -ésimo grupo e e_{ijk} denota um erro aleatório. Adotamos uma matriz de covariâncias não estruturada para os efeitos aleatórios com variâncias representadas por σ_{ai}^2 , σ_{bi}^2 e σ_{ci}^2 e covariâncias, por σ_{abi} , σ_{aci} e σ_{bci} . A variância do erro é representada por σ^2 .

Resultados para 3 modelos selecionados entre os 21 mencionados também estão dispostos na Tabela 5.6.3. O modelo proposto por meio da análise descritiva tem um ajuste melhor que os demais segundo o critério AIC mas não segundo o critério BIC.

Uma análise de resíduos como aquela sugerida no Capítulo 2 é recomendada para obtenção de mais detalhes que possibilitem a escolha do modelo mais adequado.

Tabela 5.6.1: Estimativas dos interceptos e dos coeficientes dos termos de primeiro e segundo graus além de valores t para indivíduos do grupo AIG

Indivíduo	Intercepto		Linear		Quadrático	
	Estim	Valor t	Estim	Valor t	Estim	Valor t
1	6.90	11.82	-1.50	-6.56	-0.06	0.32
2	5.84	-0.86	-0.92	0.14	-0.69	0.09
3	5.37	0.19	-1.53	-0.04	-0.94	0.07
4	7.89	11.34	-2.29	-5.68	-0.05	0.51
5	5.76	-1.79	-2.53	0.62	0.70	0.81
6	4.90	-5.16	-1.22	0.70	-0.12	0.62
7	7.21	5.36	-1.94	-5.51	-0.66	-1.00
8	6.88	0.94	-2.25	-0.64	-1.08	-0.45
9	5.91	-9.98	0.55	2.50	-1.01	-0.12
11	4.58	-2.90	-0.96	0.22	-0.04	0.61
14	6.68	5.49	-1.63	0.58	-0.05	0.59
16	5.14	-0.53	-0.44	1.61	0.14	0.75
17	6.26	-2.04	-1.59	0.23	-3.86	-0.74
18	5.35	-3.32	-1.21	1.43	-0.57	0.45
19	6.32	5.09	0.45	3.29	-1.39	-2.28
20	6.04	1.60	0.37	2.61	-1.24	-0.50
21	6.19	-4.51	-1.71	-1.30	-0.38	0.39
22	5.96	-5.03	-6.07	-0.85	-3.39	-1.44
24	5.19	-3.71	-1.20	1.44	1.01	1.03
26	5.54	-2.55	-1.88	-1.70	0.27	0.88
27	5.86	-4.37	-1.23	1.67	0.10	0.72
28	6.91	5.61	1.05	3.29	-1.32	-1.19
29	6.29	-0.73	-0.16	1.94	-0.56	-0.12
Estimativa	6.04		-1.30		-0.66	
Erro padrão	0.09		0.23		0.39	
Valor P	< 0.001	ID	< 0.001		0.094	

Tabela 5.6.2: Estimativas dos interceptos e dos coeficientes dos termos de primeiro e segundo graus além de valores t para indivíduos do grupo PIG

Indivíduo	Interceptos		Linear		Quadrático	
	Estim	Valor t	Estim	Valor t	Estim	Valor t
30	6.63	1.09	-2.53	1.34	0.75	1.68
31	7.22	-4.49	2.73	4.78	-2.83	-1.11
32	9.20	16.34	2.38	7.04	-2.95	-3.02
33	7.10	0.59	-2.44	1.25	0.25	1.28
35	7.89	8.96	-1.40	-0.07	-0.89	-0.91
36	8.20	3.24	-0.39	2.48	-2.94	-1.72
37	6.65	1.89	-0.59	2.63	-0.29	0.55
39	5.54	-6.65	-2.38	-4.20	0.59	1.95
40	8.28	1.27	-3.14	-3.43	-1.63	-0.93
41	5.70	0.03	-1.20	-0.36	-0.10	0.77
42	7.86	-12.27	0.34	3.48	-1.19	0.37
47	6.64	3.88	-1.44	-4.34	-0.30	-0.20
48	7.51	1.05	-1.73	-2.36	-1.13	-0.74
50	6.39	-4.73	-2.25	0.23	-1.84	0.33
52	8.92	12.50	-0.52	2.77	-0.36	0.46
53	4.61	-14.58	2.45	4.65	-1.50	-0.47
54	6.12	-0.96	-1.16	-1.98	-0.02	1.00
55	5.98	-2.11	-1.39	1.36	0.00	1.06
56	6.70	0.79	-1.94	-1.30	0.13	1.27
57	6.55	2.55	-1.42	-1.53	-0.38	-0.20
58	5.61	-13.24	-10.49	-2.20	-5.22	-4.22
59	7.49	2.83	-2.40	-2.83	-0.07	0.96
61	7.93	2.03	-3.13	-7.40	0.39	1.84
Estimativa	6.99		-1.48		-0.94	
Erro padrão	0.09		0.23		0.21	
Valor P	< 0.001		< 0.001		< 0.001	

Tabela 5.6.3: Estimativas dos parâmetros de modelos alternativos para os dados do Exemplo 1.2.6 (índice 1: AIG e índice 2: PIG)

Parâmetro	Modelo 1		Modelo 4		Modelo 8		Modelo 22	
	Estim	EP	Estim	EP	Estim	EP	Estim	EP
intercepto (α_1)	6.04	0.15	6.04	0.17	6.09	0.18	6.04	0.15
intercepto (α_2)	7.04	0.18	6.95	0.16	6.91	0.16	7.04	0.18
angular (β_1)	-1.23	0.12	-1.30	0.15	-1.20	0.18	-1.22	0.12
angular (β_2)	-1.36	0.2	-1.57	0.13	-1.63	0.13	-1.36	0.20
quadrático (γ_1)	-	-	-	-	-0.17	0.12	-	-
quadrático (γ_2)	-0.33	0.12	-	-	-	-	-0.33	0.12
$V(a) = \sigma_a^2$	-	-	0.712	-	0.779	-	-	-
$V(b) = \sigma_b^2$	-	-	0.287	-	0.563	-	-	-
$V(c) = \sigma_c^2$	-	-	-	-	0.085	-	-	-
$\mathbb{C}\mathbb{O}\mathbb{V}(a, b) = \sigma_{ab}$	-	-	-0.074	-	0.062	-	-	-
$\mathbb{C}\mathbb{O}\mathbb{V}(a, c) = \sigma_{ac}$	-	-	-	-	-0.115	-	-	-
$\mathbb{C}\mathbb{O}\mathbb{V}(b, c) = \sigma_{bc}$	-	-	-	-	-0.203	-	-	-
$V(a_1) = \sigma_{a_1}^2$	0.565	-	-	-	-	-	0.599	-
$V(a_2) = \sigma_{a_2}^2$	0.893	-	-	-	-	-	0.894	-
$V(b_1) = \sigma_{b_1}^2$	0.191	-	-	-	-	-	0.212	-
$V(b_2) = \sigma_{b_2}^2$	0.926	-	-	-	-	-	0.931	-
$V(c_1) = \sigma_{c_1}^2$	-	-	-	-	-	-	0.008	-
$V(c_2) = \sigma_{c_2}^2$	0.128	-	-	-	-	-	0.129	-
$\mathbb{C}\mathbb{O}\mathbb{V}(a_1, b_1) = \sigma_{ab_{11}}$	-0.232	-	-	-	-	-	-0.208	-
$\mathbb{C}\mathbb{O}\mathbb{V}(a_1, c_1) = \sigma_{ac_{11}}$	-	-	-	-	-	-	-0.045	-
$\mathbb{C}\mathbb{O}\mathbb{V}(b_1, c_1) = \sigma_{bc_{11}}$	-	-	-	-	-	-	-0.01	-
$\mathbb{C}\mathbb{O}\mathbb{V}(a_2, b_2) = \sigma_{ab_{22}}$	0.217	-	-	-	-	-	0.218	-
$\mathbb{C}\mathbb{O}\mathbb{V}(a_2, c_2) = \sigma_{ac_{22}}$	-0.260	-	-	-	-	-	-0.261	-
$\mathbb{C}\mathbb{O}\mathbb{V}(b_2, c_2) = \sigma_{bc_{22}}$	-0.277	-	-	-	-	-	-0.278	-
$V(e_{ijk}) = \sigma^2$	0.304	-	0.346	-	0.307	-	0.300	-
AIG	707.5	-	714.2	-	714.3	-	713.0	-
PIG	762.1	-	743.3	-	758.0	-	778.5	-

Estim: estimativa; EP: erro padrão

Tabela 5.6.4: Termos selecionados para modelos alternativos para os dados do Exemplo 1.2.6

Modelos	AIG						PIG						Coeficientes aleatórios comuns					
	Coeficientes fixos			Coeficientes aleatórios			Coeficientes fixos			Coeficientes aleatórios			Coeficientes aleatórios comuns			Int	Ang	Quad
	Int	Ang	Quad	Int	Ang	Quad	Int	Ang	Quad	Int	Ang	Quad	Int	Ang	Quad			
Modelo 1	x	x		x	x		x	x		x	x							
Modelo 2	x	x		x	x		x	x					x					
Modelo 3	x	x		x	x			x										
Modelo 4	x	x		x	x		x	x					x	x				
Modelo 5	x	x		x	x		x	x		x	x							
Modelo 6	x	x		x	x		x	x					x					
Modelo 7	x	x		x	x		x	x					x	x				
Modelo 8	x	x		x	x		x	x					x	x				
Modelo 9	x	x		x	x		x	x		x	x					x		
Modelo 10	x	x		x	x		x	x		x	x							
Modelo 11	x	x		x	x		x	x		x	x							
Modelo 12	x	x		x	x		x	x		x	x							
Modelo 13	x	x		x	x		x	x		x	x							
Modelo 14	x	x		x	x		x	x										
Modelo 15	x	x		x	x		x	x										
Modelo 16	x	x		x	x		x	x										
Modelo 17	x	x		x	x		x	x										
Modelo 18	x	x		x	x		x	x										
Modelo 19	x	x		x	x		x	x										
Modelo 20	x	x		x	x		x	x										
Modelo 21	x	x		x	x		x	x										
Modelo 22	x	x		x	x		x	x										

Int: intercepto; Ang: coeficiente angular; Quad: coeficiente quadrático

5.7 Exercícios

5.7.1. Os dados da planilha disponível em

<http://www.ime.usp.br/~jmsinger/Dados/Singer&Nobre&Rocha2017exerc571.xls>

foram obtidos de um estudo realizado pelo Dr Paulo Francisco Ramos Margarido na Clínica Ginecológica do Departamento de Obstetrícia e Ginecologia do Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo entre agosto de 1993 e dezembro de 1997. O volume do útero (em cm³) foi avaliado antes e depois de uma cirurgia para a retirada de um tumor de 51 mulheres tratadas com uma substância análoga ao hormônio liberador de gonadotrofina (GnRH) e de 39 mulheres não tratadas. Um dos objetivos é avaliar se o tratamento está relacionado com a diminuição do volume uterino. Utilize as técnicas descritas na Seção 5.3 para analisar os dados.

5.7.2. Os dados disponíveis em

<http://www.ime.usp.br/~jmsinger/Dados/Singer&Nobre&Rocha2017exerc572.xls>

foram obtidos de um estudo cujo objetivo é verificar se a diferença entre os fluxos salivares direito e esquerdo varia com o lado preferencial de mastigação de mulheres adultas e se depende de estímulos. O fluxo salivar de 12 mulheres que mastigam preferencialmente do lado direito e de 8 mulheres que mastigam preferencialmente do lado esquerdo foi medido em cinco ocasiões: a primeira sem estímulo, a segunda, terceira e quarta com estímulo e a quinta novamente sem estímulo.

- a) Defina a variável resposta como a diferença entre o fluxo salivar direito menos fluxo salivar esquerdo para mulheres com mastigação preferencial do lado direito e fluxo salivar esquerdo menos fluxo salivar direito para mulheres com mastigação preferencial do lado esquerdo.
- b) Impute os dados omissos.
- c) Construa um gráfico de perfis médios com as respectivas barras de desvios padrões.
- d) Analise os dados por intermédio de uma análise de perfis multivariada com parametrização de médias de celas, especificando todos os componentes do modelo assim como as matrizes definidoras das hipóteses testadas.
- e) Teste a hipótese de diferença entre as respostas esperadas com e sem estímulo, levando em conta os resultados do item d). Especifique as matrizes definidoras de cada hipótese testada.

- f) Repita as análises dos itens d) e e) sob uma parametrização de desvios de médias (soma nula).
- g) Reanalise os dados por meio de uma análise de perfis univariada.
- h) Apresente um relatório quantificando suas conclusões.

5.7.3. Os dados disponíveis em

<http://www.ime.usp.br/~jmsinger/Dados/Singer&Nobre&Rocha2017exerc573.xls>

são oriundos de um estudo desenvolvido na Faculdade de Medicina Veterinária e Zootecnia da USP para avaliar o efeito do cloranfenicol (um antibiótico de amplo espectro) em processos inflamatórios. Analise-os segundo a estratégia descrita no Exercício 5.7.2.

5.7.4. Considere o seguinte modelo para os dados do Exemplo 1.2.4:

$$y_{ijk} = \mu_{jk} + e_{ijk}^*$$

em que y_{ijk} representa o pH da placa bacteriana do i -ésimo voluntário ($i = 1, \dots, 21$) observado no instante t_k ($k = 1, \dots, 7$) do j -ésimo período ($j = 1$: antes do bochecho, $j = 2$: depois do bochecho), μ_{jk} indica a resposta esperada para o instante t_k no período j e e_{ijk}^* são erros aleatórios. Suponha que $a_i \sim N(0, \sigma_a^2)$, $b_{ij} \sim N(0, \sigma_b^2)$ e $e_{ijk} \sim N(0, \sigma^2)$ sejam variáveis aleatórias independentes e que e_{ijk}^* possa ser decomposto numa das seguintes formas

A) $e_{ijk}^* = a_i + e_{ijk}$

B) $e_{ijk}^* = b_{ij} + e_{ijk}$

C) $e_{ijk}^* = a_i + b_{ij} + e_{ijk}$

- i) Calcule as variâncias e covariâncias intraunidades amostrais induzidas por cada um dos três modelos e expresse os resultados na forma de matrizes de covariâncias intraunidades amostrais.
- ii) Esboce gráficos de perfis individuais para representar os elementos de cada modelo e interprete as diferenças entre eles.
- iii) Suponha que no modelo C, o erro e_{ijk} possa ser modelado por meio de $e_t = \rho e_{t-1} + u_t$, $t = 1, 2, \dots$ com $0 < \rho < 1$, $u_t \sim N(0, \sigma^2)$ e $e_0 = 0$. Obtenha as variâncias e covariâncias amostrais correspondentes e expresse os resultados na forma da matriz de covariâncias intraunidades amostrais.

- iv) Utilize as técnicas descritas em Rocha & Singer (2017) e consideradas no Exemplo 5.6.1 para identificar modelos polinomiais de coeficientes aleatórios apropriados para cada período separadamente.
- v) Utilize as ferramentas de diagnóstico (análise de resíduos, de pontos alavanca e de influência) descritas na Seção 2.5.1 para completar a identificação dos modelos considerados no item iv).
- vi) Ajuste um modelo que incorpore as conclusões obtidas aos dados dos dois períodos simultaneamente.

5.7.5. Expresse as hipóteses i)-v) indicadas no Exemplo 5.3.1 com a parametrização (5.3.2.

Apêndice A

Matrizes e espaços vetoriais

A.1 Matrizes

Neste apêndice apresentamos a notação matricial utilizada no texto e listamos alguns resultados relacionados com álgebra e derivação de matrizes além de conceitos de espaços vetoriais. Para detalhes, o leitor deve consultar Searle (1982), Magnus & Neudecker (1988) e Harville (1997), por exemplo.

Uma **matriz** \mathbf{A} de dimensão $m \times n$, é um arranjo retangular de elementos¹ com m linhas e n colunas, no qual o elemento a_{ij} situa-se no cruzamento da i -ésima linha com a j -ésima coluna:

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} = ((a_{ij}))_{1 \leq i \leq m, 1 \leq j \leq n}.$$

Um exemplo de uma matriz (2×4) é

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 4 & -8 \\ 3 & 5 & 2 & 0 \end{pmatrix}.$$

Um **vetor** de dimensão $(m \times 1)$ é uma matriz com m linhas e uma única coluna:

$$\mathbf{u} = \begin{pmatrix} u_{11} \\ u_{21} \\ \vdots \\ u_{m1} \end{pmatrix}.$$

¹Neste texto consideramos apenas matrizes com elementos reais.

Matrizes serão representadas por letras maiúsculas em negrito (por exemplo, \mathbf{A} , \mathbf{X} , \mathbf{G}) e vetores, por letras minúsculas em negrito (\mathbf{a} , \mathbf{x} , \mathbf{y} , por exemplo). Quando necessário, a dimensão será especificada entre parênteses; por exemplo, \mathbf{A} ($m \times n$). Uma matriz \mathbf{A} ($m \times n$), pode ser expressa como $\mathbf{A} = (\mathbf{a}_1 \dots \mathbf{a}_n)$, com \mathbf{a}_j denotando sua j -ésima coluna, ou seja,

$$\mathbf{a}_j = \begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{pmatrix}.$$

A.1.1 Operações básicas

Multiplicação por escalar: *Sejam k um número real e \mathbf{A} uma matriz ($m \times n$). O produto de \mathbf{A} por k , denotado $\mathbf{B} = k\mathbf{A}$ é uma matriz ($m \times n$) no qual o elemento $b_{ij} = ka_{ij}$, ou seja,*

$$\mathbf{B} = k\mathbf{A} = \begin{pmatrix} ka_{11} & ka_{12} & \dots & ka_{1n} \\ ka_{21} & ka_{22} & \dots & ka_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ ka_{m1} & ka_{m2} & \dots & ka_{mn} \end{pmatrix}.$$

Soma e subtração de matrizes: *Sejam \mathbf{A} e \mathbf{B} duas matrizes de mesma dimensão ($m \times n$). Sua soma, representada por $\mathbf{A} + \mathbf{B}$, é a matriz ($m \times n$) cujos elementos são dados por $c_{ij} = a_{ij} + b_{ij}$, ou seja,*

$$\mathbf{A} + \mathbf{B} = \begin{pmatrix} a_{11} + b_{11} & a_{12} + b_{12} & \dots & a_{1n} + b_{1n} \\ a_{21} + b_{21} & a_{22} + b_{22} & \dots & a_{2n} + b_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} + b_{m1} & a_{m2} + b_{m2} & \dots & a_{mn} + b_{mn} \end{pmatrix}.$$

Sejam \mathbf{A} , \mathbf{B} matrizes de dimensão ($m \times n$) e k um número real. Então valem as seguintes propriedades:

- i) $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$;
- ii) $\mathbf{A} + (\mathbf{B} + \mathbf{C}) = (\mathbf{A} + \mathbf{B}) + \mathbf{C}$;
- iii) $k(\mathbf{A} + \mathbf{B}) = k\mathbf{A} + k\mathbf{B}$.

A subtração de duas matrizes quaisquer \mathbf{A} e \mathbf{B} , de mesma dimensão $(m \times n)$, denotada $\mathbf{A} - \mathbf{B}$ é uma matriz de dimensão $(m \times n)$ cujos elementos são dados por $d_{ij} = a_{ij} - b_{ij}$.

Produto de matrizes: *O produto de uma matriz \mathbf{A} com dimensão $(m \times n)$ por uma matriz \mathbf{B} com dimensão $(n \times q)$ é uma matriz $\mathbf{C} = \mathbf{AB}$ com dimensão $(m \times q)$ e elementos*

$$c_{ij} = \sum_{k=1}^n a_{ik}b_{kj} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{in}b_{nj},$$

para $i = 1, \dots, m$ e $j = 1, \dots, q$.

Sejam \mathbf{A} , \mathbf{B} e \mathbf{C} matrizes com produtos \mathbf{AB} , \mathbf{AC} e \mathbf{BC} bem definidos. Então:

- i) $\mathbf{A}(\mathbf{BC}) = (\mathbf{AB})\mathbf{C}$;
- ii) $\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}$.

Em geral o produto de matrizes não é comutativo, ou seja, não necessariamente $\mathbf{AB} = \mathbf{BA}$. Por exemplo, dadas

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 2 & 1 \\ 0 & 0 & 5 \end{pmatrix} \text{ e } \mathbf{B} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix},$$

temos

$$\mathbf{AB} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 2 & 0 \\ 2 & 1 & 1 \end{pmatrix} \neq \mathbf{BA} = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 1 \\ 2 & 0 & 1 \end{pmatrix}.$$

Por outro lado, dadas

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 5 \end{pmatrix} \text{ e } \mathbf{B} = \begin{pmatrix} -1 & 3 \\ 1 & 2 \\ -2 & -2 \end{pmatrix},$$

temos

$$\mathbf{AB} = \begin{pmatrix} -1 & 3 \\ 2 & 4 \\ -10 & -10 \end{pmatrix},$$

mas o produto \mathbf{BA} não está definido.

Matriz transposta: A matriz transposta (às vezes chamada apenas de transposta) de uma matriz \mathbf{A} ($m \times n$), denotada por \mathbf{A}^\top , é a matriz com dimensão $(n \times m)$ cujos elementos a'_{ij} são dados por $a'_{ij} = a_{ji}$. Por exemplo, se

$$\mathbf{A} = \begin{pmatrix} -1 & 3 \\ 1 & 2 \\ -2 & -2 \end{pmatrix} \quad \text{então} \quad \mathbf{A}^\top = \begin{pmatrix} -1 & 1 & -2 \\ 3 & 2 & -2 \end{pmatrix}.$$

Para quaisquer matrizes \mathbf{A} e \mathbf{B} (para as quais as operações matriciais abaixo estejam definidas) valem as seguintes propriedades:

- i) $(\mathbf{A}^\top)^\top = \mathbf{A}$;
- ii) $(\mathbf{A} + \mathbf{B})^\top = \mathbf{A}^\top + \mathbf{B}^\top$;
- iii) $(\mathbf{AB})^\top = \mathbf{B}^\top \mathbf{A}^\top$.

A.1.2 Tipos especiais de matrizes

Matriz quadrada: Uma matriz \mathbf{A} com dimensão $(n \times n)$ é chamada de matriz quadrada de ordem n . Os elementos a_{11}, \dots, a_{nn} de uma matriz quadrada constituem sua **diagonal principal**.

Matriz simétrica: Uma matriz quadrada \mathbf{A} é simétrica se $\mathbf{A} = \mathbf{A}^\top$.

Matriz diagonal: Uma matriz quadrada \mathbf{A} é diagonal se todos os elementos não pertencentes à diagonal principal forem nulos, ou seja, se for da forma

$$\mathbf{A} = \begin{pmatrix} a_{11} & 0 & \dots & 0 \\ 0 & a_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a_{nn} \end{pmatrix}.$$

Seja \mathbf{A} uma matriz quadrada de ordem n e \mathbf{a} um vetor $(n \times 1)$ formado pelos elementos de sua diagonal principal. Então, o operador diagonal é definido como

$$\text{diag}(\mathbf{A}) = \mathbf{a}$$

e

$$\text{diag}(\mathbf{a}) = \begin{pmatrix} a_{11} & 0 & \dots & 0 \\ 0 & a_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a_{nn} \end{pmatrix}.$$

Matriz identidade: *Uma matriz diagonal de ordem n em que todos os elementos da diagonal principal são iguais a 1 é chamada matriz identidade de ordem n e é denotada por \mathbf{I}_n* ².

A matriz \mathbf{I} é o elemento neutro na multiplicação de matrizes, isto é, para qualquer matriz quadrada \mathbf{A} de ordem n

$$\mathbf{IA} = \mathbf{AI}.$$

Matriz triangular superior: *Uma matriz quadrada \mathbf{A} é triangular superior se todos os elementos abaixo da diagonal principal forem iguais a zero, ou seja, se for da forma:*

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ 0 & a_{22} & a_{23} & \dots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & a_{nn} \end{pmatrix},$$

Matriz triangular inferior: *Analogamente, se todos os elementos acima da diagonal principal de \mathbf{A} forem nulos, a matriz \mathbf{A} é triangular inferior.*

Matriz triangular: *Uma matriz é triangular se ela é triangular superior ou inferior.*

Se a matriz \mathbf{A} é triangular inferior (superior), então \mathbf{A}^\top é triangular superior (inferior); além disso, a matriz resultante da soma ou produto de matrizes triangulares superiores (inferiores) é triangular superior (inferior).

Matriz idempotente: *Uma matriz quadrada \mathbf{A} de ordem n é idempotente se*

$$\mathbf{AA} = \mathbf{A}^2 = \mathbf{A}.$$

A.1.3 Submatrizes e matrizes particionadas

Submatriz: *Uma submatriz de uma matriz \mathbf{A} é qualquer matriz obtida através da eliminação de linhas e/ou colunas.*

Por exemplo, se considerarmos

$$\mathbf{A} = \begin{pmatrix} 2 & 5 & 1 & 4 \\ -7 & 4 & 0 & 10 \\ 3 & 7 & 20 & 8 \end{pmatrix},$$

²Quando a ordem da matriz identidade for evidente, ela será denotada \mathbf{I} .

duas submatrizes de \mathbf{A} são

$$\begin{pmatrix} 2 & 5 & 1 \\ -7 & 4 & 0 \\ 3 & 7 & 20 \end{pmatrix} \text{ e } \begin{pmatrix} 2 & 5 & 1 & 4 \\ -7 & 4 & 0 & 10 \end{pmatrix},$$

quando eliminamos, respectivamente, a quarta coluna ou terceira linha de \mathbf{A} .

Matriz particionada: *Uma matriz particionada de dimensão $(m \times n)$ é uma matriz expressa na forma*

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} & \mathbf{A}_{13} & \dots & \mathbf{A}_{1s} \\ \mathbf{A}_{21} & \mathbf{A}_{22} & \mathbf{A}_{23} & \dots & \mathbf{A}_{2s} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{A}_{p1} & \mathbf{A}_{p2} & \mathbf{A}_{p3} & \dots & \mathbf{A}_{ps} \end{pmatrix},$$

com \mathbf{A}_{ij} representando uma submatriz de dimensão $(m_i \times n_j)$, $i = 1, \dots, p$; $j = 1, \dots, s$ com $m_1, \dots, m_p, n_1, \dots, n_s$ representando números inteiros positivos, tais que $\sum_{i=1}^p m_i = m$ e $\sum_{j=1}^s n_j = n$. Por exemplo, a matriz

$$\mathbf{A} = \begin{pmatrix} 2 & 5 & 1 & 4 \\ -7 & 4 & 0 & 10 \\ 3 & 7 & 20 & 8 \end{pmatrix},$$

pode ser representada como

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix},$$

em que

$$\mathbf{A}_{11} = \begin{pmatrix} 2 & 5 \\ -7 & 4 \end{pmatrix}, \quad \mathbf{A}_{12} = \begin{pmatrix} 1 & 4 \\ 0 & 10 \end{pmatrix}, \quad \mathbf{A}_{21}^\top = \begin{pmatrix} 3 \\ 7 \end{pmatrix} \text{ e } \mathbf{A}_{22}^\top = \begin{pmatrix} 20 \\ 8 \end{pmatrix},$$

com $m_1 = n_1 = n_2 = 2$ e $m_2 = 1$. Obviamente, uma matriz pode ser particionada de várias maneiras.

A.1.4 Independência linear e espaço coluna

Combinação linear: *Sejam $\mathbf{x}_1, \dots, \mathbf{x}_p$ vetores de dimensão $(n \times 1)$. O vetor \mathbf{u} , de dimensão $(n \times 1)$, é uma combinação linear dos vetores $\mathbf{x}_1, \dots, \mathbf{x}_p$, se existem c_1, \dots, c_p , números reais tais que $\mathbf{u} = c_1\mathbf{x}_1 + \dots + c_p\mathbf{x}_p$.*

Independência linear: *Os vetores $\mathbf{x}_1, \dots, \mathbf{x}_p$ de dimensão $(n \times 1)$ são linearmente independentes (L.I.) se, e somente se, a combinação linear $c_1\mathbf{x}_1 + \dots + c_p\mathbf{x}_p = \mathbf{0}$ implicar $c_1 = \dots = c_p = 0$.*

Espaço-coluna: Seja \mathbf{X} uma matriz de dimensão $(n \times p)$. O espaço-coluna da matriz \mathbf{X} , denotado $\mathcal{C}(\mathbf{X})$, é o conjunto de todas as combinações lineares dos vetores coluna da matriz \mathbf{X} , ou seja, $\mathcal{C}(\mathbf{X}) = \{\mathbf{y} \in \mathbb{R}^n : \mathbf{y} = \mathbf{X}\mathbf{a}, \mathbf{a} \in \mathbb{R}^p\}$.

Posto (rank) de uma matriz: Seja $\mathbf{A} = (\mathbf{a}_1, \dots, \mathbf{a}_n)$ uma matriz de dimensão $(m \times n)$. O posto de \mathbf{A} , denotado $r(\mathbf{A})$, é o número máximo de colunas (linhas) linearmente independentes de \mathbf{A} , ou seja, é a dimensão do espaço-coluna (linha) de \mathbf{A} .

Sejam \mathbf{A} , \mathbf{B} e \mathbf{C} matrizes de dimensões $(m \times n)$, $(n \times q)$ e $(q \times t)$, respectivamente, então valem as seguintes propriedades:

- i) $r(\mathbf{A}) = r(\mathbf{A}^\top)$;
- ii) $r(\mathbf{AB}) \leq \min[r(\mathbf{A}), r(\mathbf{B})]$;
- iii) Se $r(\mathbf{A}) = n$ e $r(\mathbf{B}) = q < n$, então $r(\mathbf{AB}) = q$;
- iv) $r(\mathbf{AB}) + r(\mathbf{BC}) \leq r(\mathbf{B}) + r(\mathbf{ABC})$.

Matriz de posto completo: Uma matriz \mathbf{A} de dimensão $(m \times n)$ tem posto completo quando $r(\mathbf{A}) = \min(m, n)$.

Matriz não singular: Uma matriz quadrada \mathbf{A} de ordem n é não singular se $r(\mathbf{A}) = n$.

A.1.5 Determinante de uma matriz

Determinante: O determinante de uma matriz quadrada \mathbf{A} de ordem n , denotado $|\mathbf{A}|$, é

$$|\mathbf{A}| = \sum_{k=1}^n a_{ik}(-1)^{i+k}|\mathbf{A}_{ik}|,$$

em que \mathbf{A}_{ik} é obtida a partir da matriz \mathbf{A} excluindo-se sua i -ésima linha e k -ésima coluna. O determinante $|\mathbf{A}_{ik}|$ é chamado **menor** de \mathbf{A} . Quando $i = k$, o determinante é chamado de **menor principal** de \mathbf{A} . O termo $(-1)^{i+k}|\mathbf{A}_{ik}|$ é denominado **cofator**. Se

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix},$$

então,

$$\begin{aligned} |\mathbf{A}| &= a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{32}a_{21} \\ &- a_{31}a_{22}a_{13} - a_{21}a_{12}a_{33} - a_{11}a_{23}a_{32}. \end{aligned}$$

O determinante satisfaz as seguintes propriedades:

- i) $|\mathbf{A}| \neq 0$ se e somente se a matriz \mathbf{A} tem posto completo. Quando $|\mathbf{A}| \neq 0$, a matriz \mathbf{A} é **não singular**;
- ii) $|\mathbf{A}| = |\mathbf{A}^\top|$;
- iii) $|c\mathbf{A}| = c^n |\mathbf{A}|$, $c \in \mathbb{R}$;
- iv) Se \mathbf{A} é uma matriz triangular, então $|\mathbf{A}| = \prod_{i=1}^n a_{ii}$;
- v) Sejam \mathbf{A} e \mathbf{B} duas matrizes quadradas de mesma ordem; então $|\mathbf{AB}| = |\mathbf{BA}| = |\mathbf{A}||\mathbf{B}|$;
- vi) Sejam \mathbf{A} uma matriz $(m \times n)$ e \mathbf{B} uma matriz $(n \times m)$; então

$$|\mathbf{I}_m + \mathbf{AB}| = |\mathbf{I}_n + \mathbf{BA}|.$$

A.1.6 Inversão de matrizes

Matriz inversa: *A matriz inversa (quando existe) de uma matriz matriz quadrada \mathbf{A} de ordem n é uma matriz \mathbf{A}^{-1} quadrada de ordem n tal que $\mathbf{AA}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}_n$.*

O teorema abaixo, relaciona a existência da inversa de \mathbf{A} com o fato de seu determinante ser diferente de zero.

Teorema A.1.1. *Uma matriz quadrada \mathbf{A} é inversível e sua inversa é única, se e somente se ela for não singular.*

Supondo que todas as matrizes inversas existam, valem as seguintes propriedades:

- i) $|\mathbf{A}^{-1}| = |\mathbf{A}|^{-1}$;
- ii) Se $|\mathbf{A}| \neq 0$, então \mathbf{A}^\top e \mathbf{A}^{-1} são matrizes não singulares e além disso $(\mathbf{A}^\top)^{-1} = (\mathbf{A}^{-1})^\top$;
- iii) $(c\mathbf{A})^{-1} = c^{-1}\mathbf{A}^{-1}$;

iv) $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$;

v) Sejam \mathbf{A} , \mathbf{B} e \mathbf{C} matrizes com dimensões $(k \times k)$, $(k \times n)$ e $(n \times k)$, respectivamente, com \mathbf{A} não singular. Então

$$|\mathbf{A} + \mathbf{BC}| = |\mathbf{A}||\mathbf{I}_k + \mathbf{A}^{-1}\mathbf{BC}|;$$

vi) Sejam \mathbf{A} , \mathbf{B} , \mathbf{C} e \mathbf{D} matrizes com dimensões $(m \times m)$, $(m \times n)$, $(n \times n)$ e $(n \times m)$, respectivamente. Então

$$(\mathbf{A} + \mathbf{BCD})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{C}^{-1} + \mathbf{DA}^{-1}\mathbf{B})^{-1}\mathbf{DA}^{-1}.$$

vii) Seja \mathbf{J}_n uma matriz quadrada de ordem n com todos elementos iguais a 1 e a, b números reais positivos. Então

$$[a\mathbf{I}_n + b\mathbf{J}_n]^{-1} = \frac{1}{a} \left[\mathbf{I}_n - \frac{b}{nb + a} \mathbf{J}_n \right];$$

viii) Sejam \mathbf{a} e \mathbf{b} vetores de dimensão $(n \times 1)$ e \mathbf{A} uma matriz quadrada não singular de ordem n . Se $1 \pm \mathbf{b}^\top \mathbf{A}^{-1} \mathbf{a} \neq 0$, então

$$(\mathbf{A} \pm \mathbf{ab}^\top)^{-1} = \mathbf{A}^{-1} \mp \frac{(\mathbf{A}^{-1}\mathbf{a})(\mathbf{b}^\top \mathbf{A}^{-1})}{1 \pm \mathbf{b}^\top \mathbf{A}^{-1} \mathbf{a}};$$

ix) Sejam \mathbf{A} e \mathbf{D} matrizes quadradas; então

$$\begin{vmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{vmatrix} = \begin{cases} |\mathbf{A}||\mathbf{D} - \mathbf{CA}^{-1}\mathbf{B}|, & \text{se } \mathbf{A} \text{ for não singular} \\ |\mathbf{D}||\mathbf{A} - \mathbf{BD}^{-1}\mathbf{C}|, & \text{se } \mathbf{D} \text{ for não singular} \end{cases}$$

Se ambas \mathbf{A} e \mathbf{D} forem não singulares, então

$$|\mathbf{D} - \mathbf{CA}^{-1}\mathbf{B}| = \frac{|\mathbf{D}|}{|\mathbf{A}|} |\mathbf{A} - \mathbf{BD}^{-1}\mathbf{C}|.$$

x) Sejam \mathbf{A} e \mathbf{D} matrizes quadradas; então

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^\top & \mathbf{D} \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{E} & \mathbf{F} \\ \mathbf{F}^\top & \mathbf{G} \end{pmatrix},$$

com

$$\begin{aligned} \mathbf{E} &= \mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{B}(\mathbf{D} - \mathbf{B}^\top \mathbf{A}^{-1} \mathbf{B})^{-1} \mathbf{B}^\top \mathbf{A}^{-1}, \\ \mathbf{F} &= -\mathbf{A}^{-1}\mathbf{B}(\mathbf{D} - \mathbf{B}^\top \mathbf{A}^{-1} \mathbf{B})^{-1}, \\ \mathbf{G} &= (\mathbf{D} - \mathbf{B}^\top \mathbf{A}^{-1} \mathbf{B})^{-1}. \end{aligned}$$

Inversa generalizada: Uma matriz inversa generalizada da matriz \mathbf{A} , $(m \times n)$, é qualquer matriz \mathbf{G} de dimensão $(n \times m)$ que satisfaz a relação

$$\mathbf{AGA} = \mathbf{A}.$$

Para detalhes sobre essa classe de matrizes veja Harville (1997), por exemplo.

A.1.7 Traço de uma matriz

Traço de uma matriz: O traço de uma matriz quadrada de ordem n é $\text{tr}(\mathbf{A}) = \sum_{i=1}^n a_{ii}$.

Por exemplo, se

$$\mathbf{A} = \begin{pmatrix} 6 & 7 & 4 \\ 5 & 9 & 1 \\ 3 & 8 & -2 \end{pmatrix},$$

então $\text{tr}(\mathbf{A}) = 6 + 9 - 2 = 13$.

Considere \mathbf{A} , \mathbf{B} e \mathbf{C} matrizes quadradas de ordem n , \mathbf{a} um vetor $(m \times 1)$ e a e b números reais. A função traço apresenta as seguintes propriedades:

- i) $\text{tr}(\mathbf{I}_n) = n$;
- ii) $\text{tr}(a\mathbf{A} \pm b\mathbf{B}) = a\text{tr}(\mathbf{A}) \pm b\text{tr}(\mathbf{B})$;
- iii) $\text{tr}(\mathbf{AB}) = \text{tr}(\mathbf{BA})$;
- iv) $\text{tr}(\mathbf{ABC}) = \text{tr}(\mathbf{CAB}) = \text{tr}(\mathbf{BCA})$;
- v) Se \mathbf{A} for idempotente, então $\text{tr}(\mathbf{A}) = r(\mathbf{A})$;
- vii) $\text{tr}(\mathbf{A}^\top) = \text{tr}(\mathbf{A})$;
- viii) $\text{tr}(\mathbf{AA}^\top) = \text{tr}(\mathbf{A}^\top \mathbf{A}) = \sum_{i=1}^n \sum_{j=1}^n a_{ij}^2$;
- ix) $\text{tr}(\mathbf{aa}^\top) = \mathbf{a}^\top \mathbf{a} = \sum_{i=1}^n a_i^2$.

A.1.8 Soma direta e produto de Kronecker

Soma direta: Sejam \mathbf{A} uma matriz de dimensão $(m \times n)$ e \mathbf{B} uma matriz de dimensão $(p \times q)$. A soma direta das matrizes \mathbf{A} e \mathbf{B} é a matriz diagonal em blocos de dimensão $[(m+p) \times (n+q)]$, definida por

$$\mathbf{A} \oplus \mathbf{B} = \begin{pmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{B} \end{pmatrix}.$$

De uma forma mais geral, a soma direta de matrizes \mathbf{A}_i , com dimensão $(n_i \times m_i)$, $i = 1, \dots, n$ é a matriz com dimensão $(\sum_{i=1}^n n_i \times \sum_{i=1}^n m_i)$ dada por

$$\bigoplus_{i=1}^n \mathbf{A}_i = \mathbf{A}_1 \oplus \mathbf{A}_2 \oplus \dots \oplus \mathbf{A}_n = \begin{pmatrix} \mathbf{A}_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_2 & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{A}_n \end{pmatrix}.$$

Produto de Kronecker: Sejam \mathbf{A} e \mathbf{B} matrizes de dimensões $(m \times n)$ e $(p \times q)$, respectivamente. O produto de Kronecker (**produto direto** ou **produto tensorial**) das matrizes \mathbf{A} e \mathbf{B} é a matriz de dimensão $(mp \times nq)$, definida por

$$\mathbf{A} \otimes \mathbf{B} = \begin{pmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \dots & a_{1n}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & \dots & a_{2n}\mathbf{B} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}\mathbf{B} & a_{m2}\mathbf{B} & \dots & a_{mn}\mathbf{B} \end{pmatrix}.$$

Em geral $\mathbf{A} \otimes \mathbf{B} \neq \mathbf{B} \otimes \mathbf{A}$.

Sejam \mathbf{A} , \mathbf{B} , \mathbf{C} , \mathbf{D} matrizes de dimensões $(m \times n)$, $(p \times q)$, $(n \times u)$, $(q \times v)$, respectivamente, \mathbf{a} , \mathbf{b} e \mathbf{d} vetores de dimensões $(m \times 1)$, $(n \times 1)$ e $(p \times 1)$, respectivamente, e x e y números reais. Então:

- i) $x \otimes \mathbf{A} = \mathbf{A} \otimes x = x\mathbf{A}$;
- ii) $\mathbf{a} \otimes \mathbf{b}^\top = \mathbf{b}^\top \otimes \mathbf{a} = \mathbf{a}\mathbf{b}^\top$;
- iii) $\mathbf{0}_{p \times q} \otimes \mathbf{A} = \mathbf{A} \otimes \mathbf{0}_{p \times q} = \mathbf{0}_{mp \times nq}$;
- iv) $\mathbf{I}_m \otimes \mathbf{I}_p = \mathbf{I}_{mp}$;
- v) Se $\mathbf{F} = \text{diag}(f_{11}, \dots, f_{kk})$, então $\mathbf{F} \otimes \mathbf{A} = \bigoplus_{i=1}^k f_{ii} \mathbf{A}$;
- vi) $\mathbf{I}_k \otimes \mathbf{A} = \bigoplus_{i=1}^k \mathbf{A}$;
- vii) $x\mathbf{A} \otimes y\mathbf{B} = xy(\mathbf{A} \otimes \mathbf{B})$;
- viii) $(\mathbf{A} \otimes \mathbf{B}) \otimes \mathbf{C} = \mathbf{A} \otimes (\mathbf{B} \otimes \mathbf{C})$;
- ix) $(\mathbf{A} \otimes \mathbf{B}) \otimes (\mathbf{C} \otimes \mathbf{D}) = (\mathbf{A}\mathbf{C}) \otimes (\mathbf{B}\mathbf{D})$;
- x) $\mathbf{A} \otimes \mathbf{B} = (\mathbf{A} \otimes \mathbf{I}_p)(\mathbf{I}_n \otimes \mathbf{B}) = (\mathbf{I}_m \otimes \mathbf{B})(\mathbf{A} \otimes \mathbf{I}_q)$;
- xi) $(\mathbf{A} \otimes \mathbf{d}^\top)(\mathbf{b} \otimes \mathbf{B}) = (\mathbf{d}^\top \otimes \mathbf{A})(\mathbf{B} \otimes \mathbf{b}) = \mathbf{A}\mathbf{b}\mathbf{d}^\top \mathbf{B}$;
- xii) $\mathbf{D} \otimes (\mathbf{A} + \mathbf{B}) = (\mathbf{D} \otimes \mathbf{A}) + (\mathbf{D} \otimes \mathbf{B})$;
- xiii) $(\mathbf{A} \otimes \mathbf{B})^\top = (\mathbf{A}^\top \otimes \mathbf{B}^\top)$;
- xiv) $r(\mathbf{A} \otimes \mathbf{B}) = r(\mathbf{A})r(\mathbf{B})$;
- xv) $\text{tr}(\mathbf{A} \otimes \mathbf{B}) = \text{tr}(\mathbf{A})\text{tr}(\mathbf{B})$;

Além disso,

- xvi) Se \mathbf{A} e \mathbf{B} são matrizes simétricas, então $(\mathbf{A} \otimes \mathbf{B})^\top = \mathbf{A} \otimes \mathbf{B}$;
- xvii) Se \mathbf{A} é uma matriz quadrada de ordem n e \mathbf{a} é um vetor $(m \times 1)$, então $(\mathbf{I}_n \otimes \mathbf{a})\mathbf{A}(\mathbf{I}_n \otimes \mathbf{a}^\top) = \mathbf{A} \otimes \mathbf{a}\mathbf{a}^\top$;
- xviii) Se \mathbf{A} e \mathbf{B} são matrizes não singulares, temos $(\mathbf{A} \otimes \mathbf{B})^{-1} = \mathbf{A}^{-1} \otimes \mathbf{B}^{-1}$;
- xix) Se \mathbf{A} e \mathbf{B} são matrizes quadradas de ordem m e n , respectivamente, então $|\mathbf{A} \otimes \mathbf{B}| = |\mathbf{A}|^n |\mathbf{B}|^m$;
- xx) Seja $\mathbf{A} = [\mathbf{A}_1 \ \mathbf{A}_2]$; então $[\mathbf{A}_1 \ \mathbf{A}_2] \otimes \mathbf{B} = [\mathbf{A}_1 \otimes \mathbf{B} \ \mathbf{A}_2 \otimes \mathbf{B}]$, mas $\mathbf{W} \otimes [\mathbf{B}_1 \ \mathbf{B}_2] \neq [\mathbf{W} \otimes \mathbf{B}_1 \ \mathbf{W} \otimes \mathbf{B}_2]$.

A.1.9 Operadores vec e vech

Operador vec : A operação de vetorização de uma matriz $\mathbf{A} = (\mathbf{a}_1 \dots \mathbf{a}_n)$ consiste em “empilhar” seus elementos na forma

$$\text{vec}(\mathbf{A}) = \begin{pmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{a}_n \end{pmatrix}.$$

Sejam \mathbf{A} , \mathbf{B} , \mathbf{C} matrizes reais de dimensões $(m \times n)$, $(n \times p)$, $(p \times q)$, respectivamente e vetores \mathbf{a} e \mathbf{b} de dimensões $(m \times 1)$ e $(n \times 1)$, respectivamente. Então:

- i) $\text{vec}(\mathbf{a}^\top) = \text{vec}(\mathbf{a})$;
- ii) $\text{vec}(\mathbf{a}\mathbf{b}^\top) = \mathbf{b} \otimes \mathbf{a}$;
- iii) $\text{vec}(\mathbf{A}\mathbf{B}) = (\mathbf{I}_p \otimes \mathbf{A})\text{vec}(\mathbf{B}) = (\mathbf{B}^\top \otimes \mathbf{I}_m)\text{vec}(\mathbf{A})$;
- iv) $\text{vec}(\mathbf{A}\mathbf{B}\mathbf{C}) = (\mathbf{C}^\top \otimes \mathbf{A})\text{vec}(\mathbf{B})$;
- v) $\text{vec}(\mathbf{A}\mathbf{B}\mathbf{C}) = (\mathbf{I}_q \otimes \mathbf{A}\mathbf{B})\text{vec}(\mathbf{C}) = (\mathbf{C}^\top \mathbf{B}^\top \otimes \mathbf{I}_n)\text{vec}(\mathbf{A})$.

Além disso,

- vi) Se \mathbf{A} e \mathbf{B} são matrizes de mesma dimensão, temos

$$\text{tr}(\mathbf{A}^\top \mathbf{B}) = \text{vec}(\mathbf{A})^\top \text{vec}(\mathbf{B})$$

e

$$\text{vec}(\mathbf{A}^\top)^\top \text{vec}(\mathbf{B}) = \text{vec}(\mathbf{B}^\top)^\top \text{vec}(\mathbf{A}) = \text{tr}(\mathbf{A}\mathbf{B});$$

vii) Se \mathbf{B} for uma matriz de dimensão $(n \times m)$, então

$$\text{tr}(\mathbf{AB}) = \text{vec}(\mathbf{A}^\top)^\top \text{vec}(\mathbf{B});$$

viii) Se \mathbf{A} e \mathbf{B} são matrizes simétricas de ordem n , então

$$\text{vec}(\mathbf{A})^\top (\mathbf{B} \otimes \mathbf{B}) \text{vec}(\mathbf{A}) = [\text{tr}(\mathbf{BA})]^2;$$

ix) Se \mathbf{C} é uma matriz de dimensão $(p \times m)$, temos

$$\begin{aligned} \text{tr}(\mathbf{ABC}) &= \text{vec}(\mathbf{A}^\top)^\top (\mathbf{C}^\top \otimes \mathbf{I}_n) \text{vec}(\mathbf{B}) \\ &= \text{vec}(\mathbf{A}^\top)^\top (\mathbf{I}_m \otimes \mathbf{B}) \text{vec}(\mathbf{C}) \\ &= \text{vec}(\mathbf{B}^\top)^\top (\mathbf{A} \otimes \mathbf{I}_p) \text{vec}(\mathbf{C}) \\ &= \text{vec}(\mathbf{B}^\top)^\top (\mathbf{I}_n \otimes \mathbf{C}) \text{vec}(\mathbf{A}) \\ &= \text{vec}(\mathbf{C}^\top)^\top (\mathbf{B} \otimes \mathbf{I}_m) \text{vec}(\mathbf{A}) \\ &= \text{vec}(\mathbf{C}^\top)^\top (\mathbf{I}_p \otimes \mathbf{A}) \text{vec}(\mathbf{B}) \end{aligned}$$

O operador $\text{vech}(\cdot)$ aplicado a uma matriz simétrica \mathbf{A} gera um vetor com os elementos distintos dessa matriz. Por exemplo, se

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \quad \text{então} \quad \text{vech}(\mathbf{A}) = \begin{pmatrix} a_{11} \\ a_{21} \\ a_{31} \\ a_{22} \\ a_{32} \\ a_{33} \end{pmatrix}.$$

A.2 Tópicos de Álgebra Linear

Espaço vetorial: Um espaço vetorial sobre \mathbb{R} é um conjunto \mathcal{V} não vazio de elementos chamados vetores no qual estão definidas:

i) Uma operação de adição que associa a cada par de vetores \mathbf{a} e \mathbf{b} de \mathcal{V} um vetor $\mathbf{a} + \mathbf{b} \in \mathcal{V}$ e para a qual são válidas as seguintes propriedades:

1. $\mathbf{a} + \mathbf{b} = \mathbf{b} + \mathbf{a}$ (comutatividade);
2. $\mathbf{a} + (\mathbf{b} + \mathbf{c}) = (\mathbf{a} + \mathbf{b}) + \mathbf{c}$ (associatividade);
3. Existe um vetor nulo $\mathbf{0} \in \mathcal{V}$, tal que $\mathbf{a} + \mathbf{0} = \mathbf{a}$ para todo $\mathbf{a} \in \mathcal{V}$;

4. Para cada vetor $\mathbf{a} \in \mathcal{V}$, existe um vetor $-\mathbf{a} \in \mathcal{V}$ tal que $\mathbf{a} + (-\mathbf{a}) = \mathbf{0}$;
- ii) Uma operação de multiplicação por números reais que associa a cada $\alpha \in \mathbb{R}$ e a cada vetor $\mathbf{a} \in \mathcal{V}$ um vetor $\alpha\mathbf{a} \in \mathcal{V}$ para a qual são válidas as seguintes propriedades:
1. $1\mathbf{a} = \mathbf{a} \quad \forall \mathbf{a} \in \mathcal{V}$;
 2. $(\alpha\beta)\mathbf{a} = \alpha(\beta\mathbf{a}), \quad \forall \alpha, \beta \in \mathbb{R}$
 3. $\alpha(\mathbf{a} + \mathbf{b}) = \alpha\mathbf{a} + \alpha\mathbf{b}$;
 4. $(\alpha + \beta)\mathbf{a} = \alpha\mathbf{a} + \beta\mathbf{a}$.

O espaço vetorial mais utilizado nas aplicações em Estatística é o espaço das matrizes reais de dimensão $(n \times p)$, denotado $\mathbb{R}^{n \times p}$.

Subespaço vetorial: Sejam \mathcal{V} um espaço vetorial sobre \mathbb{R} e \mathcal{W} um subconjunto não vazio de \mathcal{V} . Dizemos que \mathcal{W} é um subespaço vetorial de \mathcal{V} se valem as seguintes propriedades:

- i) se \mathbf{a} e $\mathbf{b} \in \mathcal{W}$, então $\mathbf{a} + \mathbf{b} \in \mathcal{W}$;
- ii) se $\alpha \in \mathbb{R}$ e $\mathbf{a} \in \mathcal{W}$, então $\alpha\mathbf{a} \in \mathcal{W}$.

Base de um espaço vetorial: Seja \mathcal{V} um espaço vetorial. Se qualquer vetor $\mathbf{x} \in \mathcal{V}$ puder ser escrito como uma combinação linear de um conjunto de vetores linearmente independentes $\{\mathbf{a}_1, \dots, \mathbf{a}_n\} \subset \mathcal{V}$, então dizemos que $\{\mathbf{a}_1, \dots, \mathbf{a}_n\}$ é uma base do espaço vetorial \mathcal{V} .

Dimensão de um espaço vetorial: A dimensão do espaço vetorial \mathcal{V} , denotada $\dim \mathcal{V}$, é igual ao número de vetores que formam uma base de \mathcal{V} .

Transformação linear: Sejam \mathcal{U} e \mathcal{V} espaços vetoriais. Uma transformação linear de \mathcal{U} em \mathcal{V} , denotada por $\mathbf{T} : \mathcal{U} \rightarrow \mathcal{V}$, é uma função que associa a cada vetor $\mathbf{v} \in \mathcal{U}$ um vetor $\mathbf{T}(\mathbf{v}) \in \mathcal{V}$, de modo que, para quaisquer vetores $\mathbf{a}, \mathbf{b} \in \mathcal{V}$ e $k \in \mathbb{R}$, valem as seguintes propriedades:

1. $\mathbf{T}(\mathbf{a} + \mathbf{b}) = \mathbf{T}(\mathbf{a}) + \mathbf{T}(\mathbf{b})$;
2. $\mathbf{T}(k\mathbf{a}) = k\mathbf{T}(\mathbf{a}), \quad \forall k \in \mathbb{R}$.

As transformações lineares frequentemente utilizadas em Estatística são aquelas em que \mathbf{T} é uma função vetorial do tipo $\mathbf{T} : \mathbb{R}^p \rightarrow \mathbb{R}^n$ definida por uma matriz \mathbf{A}

de dimensão $(p \times n)$, tal que $\forall \mathbf{x} \in \mathbb{R}^p$,

$$\mathbf{T}(\mathbf{x}) = \mathbf{A}\mathbf{x} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1p} \\ a_{21} & a_{22} & \dots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{np} \end{pmatrix} \begin{pmatrix} x_{11} \\ x_{21} \\ \vdots \\ x_{p1} \end{pmatrix}.$$

Teorema A.2.1. *Sejam $\mathbf{x}_1, \dots, \mathbf{x}_p$ vetores de dimensão $(n \times 1)$ pertencentes ao espaço vetorial \mathbb{R}^n e \mathcal{W} o conjunto definido por*

$$\mathcal{W} = \{\mathbf{b} \in \mathbb{R}^n \mid \mathbf{b} = \sum_{i=1}^p \alpha_i \mathbf{x}_i = \mathbf{X}\boldsymbol{\alpha}, \mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_p) \in \mathbb{R}^{n \times p}, \boldsymbol{\alpha} \in \mathbb{R}^p\}$$

com \mathbf{X} denotando a matriz da transformação linear de \mathbb{R}^p em \mathbb{R}^n ($p < n$); então \mathcal{W} é um subespaço de \mathbb{R}^n .

Espaço coluna de uma matriz: *Seja \mathbf{X} uma matriz de dimensão $(n \times p)$; o espaço coluna de \mathbf{X} , denotado por $\mathcal{C}(\mathbf{X})$ é o espaço vetorial gerado por suas colunas.*

Espaço nulo de uma matriz: *O espaço nulo de uma matriz \mathbf{X} com dimensão $(n \times p)$, denotado $\mathcal{N}(\mathbf{X})$, é o conjunto de vetores $\mathbf{a} \in \mathbb{R}^p$ tais que $\mathbf{X}\mathbf{a} = \mathbf{0}$, ou seja, $\mathcal{N}(\mathbf{X}) = \{\mathbf{a} \in \mathbb{R}^p; \mathbf{X}\mathbf{a} = \mathbf{0}\}$.*

Teorema A.2.2. *Sejam \mathbf{X} uma matriz de dimensão $(n \times p)$ com $r(\mathbf{X}) = r$ e $\mathcal{C}(\mathbf{X})$ seu espaço coluna. Então $r(\mathbf{X}) = \dim[\mathcal{C}(\mathbf{X})] = r(\mathbf{X}^\top) = \dim[\mathcal{C}(\mathbf{X}^\top)] = r$ e além disso, $\dim[\mathcal{N}(\mathbf{X})] = p - r$ e $\dim[\mathcal{N}(\mathbf{X}^\top)] = n - r$.*

Teorema A.2.3. *Sejam \mathbf{A} e \mathbf{B} matrizes com dimensões $(n \times m)$ e $(m \times p)$, respectivamente. Então $\mathcal{C}(\mathbf{AB})$ é um subespaço de $\mathcal{C}(\mathbf{A})$.*

Teorema A.2.4. *Seja \mathbf{X} uma matriz de dimensão $(n \times p)$ com $r(\mathbf{X}) = r$. Então $\mathcal{C}(\mathbf{X}^\top) = \mathcal{C}(\mathbf{X}^\top \mathbf{X})$ e $r(\mathbf{X}^\top \mathbf{X}) = r$.*

Produto interno: *No espaço \mathbb{R}^n , o produto interno canônico dos vetores \mathbf{x} e \mathbf{y} é um número real dado por*

$$\mathbf{x} \bullet \mathbf{y} = \mathbf{x}^\top \mathbf{y} = \begin{pmatrix} x_1 & x_2 & \dots & x_n \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \sum_{i=1}^n x_i y_i.$$

Espaço euclidiano: *O espaço euclidiano \mathbb{R}^n é o espaço vetorial \mathbb{R}^n com soma e produto por escalar definidos da forma usual, munido do produto interno canônico.*

Norma de um vetor: No espaço euclidiano \mathbb{R}^n , a norma (ou comprimento) do vetor \mathbf{x} é o número $\|\mathbf{x}\| = (\mathbf{x}^\top \mathbf{x})^{\frac{1}{2}}$. Quando a norma de um vetor \mathbf{x} é igual a 1, diz-se que \mathbf{x} é um vetor unitário.

Distância euclidiana: A distância euclidiana entre os vetores \mathbf{x} e \mathbf{y} de \mathbb{R}^n é o número $\|\mathbf{x} - \mathbf{y}\|$.

Para \mathbf{a} , \mathbf{b} e $\mathbf{c} \in \mathbb{R}^n$ e $k \in \mathbb{R}$, o produto interno canônico tem as seguintes propriedades:

- i) $\mathbf{a}^\top \mathbf{b} = \mathbf{b}^\top \mathbf{a}$;
- ii) $\mathbf{a}^\top (\mathbf{b} + \mathbf{c}) = \mathbf{a}^\top \mathbf{b} + \mathbf{a}^\top \mathbf{c}$;
- iii) $k(\mathbf{a}^\top \mathbf{b}) = (k\mathbf{a})^\top \mathbf{b} = \mathbf{a}^\top (k\mathbf{b})$;
- iv) $\mathbf{a}^\top \mathbf{a} = \|\mathbf{a}\|^2 > 0$ se $\mathbf{a} \neq \mathbf{0}$;
- v) $\|\mathbf{a} \pm \mathbf{b}\|^2 = \|\mathbf{a}\|^2 + \|\mathbf{b}\|^2 \pm 2\mathbf{a}^\top \mathbf{b}$;
- vi) $|\mathbf{a}^\top \mathbf{b}| \leq \|\mathbf{a}\| \|\mathbf{b}\|$; **Desigualdade de Cauchy-Schwarz**
- vii) $\|\mathbf{a} + \mathbf{b}\| \leq \|\mathbf{a}\| + \|\mathbf{b}\|$; **Desigualdade triangular**

Ângulo entre vetores: O ângulo $\theta \in [0, \pi]$ entre dois vetores \mathbf{a} e $\mathbf{b} \in \mathbb{R}^n$ é

$$\arccos(\theta) = \frac{\mathbf{a}^\top \mathbf{b}}{\|\mathbf{a}\| \|\mathbf{b}\|}.$$

Produto interno de matrizes: No espaço $\mathbb{R}^{m \times n}$, o produto interno canônico das matrizes \mathbf{A} e \mathbf{B} é o número real $\text{tr}(\mathbf{A}^\top \mathbf{B}) = \text{tr}(\mathbf{A} \mathbf{B}^\top)$.

Norma de uma matriz: A norma da matriz $\mathbf{A} \in \mathbb{R}^{m \times n}$ (comumente denominada norma de Frobenius) é o número real

$$\|\mathbf{A}\| = [\text{tr}(\mathbf{A}^\top \mathbf{A})]^{\frac{1}{2}} = \left(\sum_{i=1}^m \sum_{j=1}^n a_{ij}^2 \right)^{\frac{1}{2}} = \|\text{vec}(\mathbf{A})\|.$$

Vetores ortogonais: Se \mathcal{V} é um espaço vetorial com produto interno canônico, dizemos que $\mathbf{x}, \mathbf{y} \in \mathcal{V}$ são ortogonais ($\mathbf{x} \perp \mathbf{y}$), se, e somente se $\mathbf{x}^\top \mathbf{y} = 0$.

Complemento ortogonal: Se \mathcal{V} é um espaço vetorial com produto interno e \mathcal{W} é um subespaço de \mathcal{V} , o conjunto $\mathcal{W}^\perp = \{\mathbf{a} \in \mathcal{V}; \mathbf{a} \bullet \mathbf{v} = 0 \forall \mathbf{v} \in \mathcal{W}\}$ é um subespaço vetorial de \mathcal{V} , denominado complemento ortogonal de \mathcal{W} .

Teorema A.2.5. *Seja uma matriz \mathbf{X} de dimensão $(n \times p)$. O espaço nulo de \mathbf{X}^\top e o complemento ortogonal do espaço coluna de \mathbf{X} são iguais, ou seja, $\mathcal{N}(\mathbf{X}^\top) = \mathcal{C}(\mathbf{X})^\perp$ e $\mathcal{N}(\mathbf{X}) = \mathcal{C}(\mathbf{X}^\top)^\perp$.*

Subespaço ortogonal: *Sejam \mathcal{V} um espaço vetorial com produto interno, \mathcal{U} e \mathcal{W} , subespaços de \mathcal{V} . O subespaço \mathcal{U} é ortogonal ao subespaço \mathcal{W} ($\mathcal{U} \perp \mathcal{W}$), se cada vetor de \mathcal{U} for ortogonal a cada vetor de \mathcal{W} . Além disso, dizemos que $\mathbf{v} \perp \mathcal{U}$ se $\mathbf{v} \bullet \mathbf{u} = 0 \quad \forall \mathbf{u} \in \mathcal{U}$.*

Sejam \mathbf{y} um vetor de dimensão $(m \times 1)$ e \mathbf{X} e \mathbf{Z} matrizes com dimensões $(m \times n)$ e $(m \times p)$, respectivamente. Então \mathbf{y} é ortogonal ao espaço coluna da matriz \mathbf{X} (com relação ao produto interno canônico de \mathbb{R}^m), nomeadamente, $\mathcal{C}(\mathbf{X})$, se e somente se $\mathbf{X}^\top \mathbf{y} = \mathbf{0}$. De modo similar, o espaço coluna de \mathbf{X} , $\mathcal{C}(\mathbf{X})$ é ortogonal ao espaço coluna de \mathbf{Z} , $\mathcal{C}(\mathbf{Z})$, se e somente se $\mathbf{X}^\top \mathbf{Z} = \mathbf{0}$.

Vetores ortonormais: *Seja \mathcal{V} um espaço vetorial com produto interno e \mathbf{x} e $\mathbf{y} \in \mathcal{V}$. Os vetores \mathbf{x} e \mathbf{y} são ortonormais se $\|\mathbf{x}\| = \|\mathbf{y}\| = 1$ e $\mathbf{x} \perp \mathbf{y}$.*

Base ortonormal: *Seja \mathcal{V} um espaço vetorial de dimensão finita n com produto interno. Uma base $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ de \mathcal{V} é dita ortonormal se seus elementos forem vetores de norma igual a 1 ($\|\mathbf{x}_i\| = 1$, para $i = 1, \dots, n$) e forem ortogonais dois a dois.*

Teorema A.2.6. *Sejam \mathbf{Y} uma matriz no espaço vetorial $\mathcal{V} \subset \mathbb{R}^{nm}$ das matrizes $(m \times n)$ e \mathcal{U} um subespaço de \mathcal{V} . Então existe uma única matriz $\mathbf{Z} \in \mathcal{U}$, tal que $(\mathbf{Y} - \mathbf{Z}) \in \mathcal{U}^\perp$. A matriz \mathbf{Z} é a projeção ortogonal de \mathbf{Y} em \mathcal{U} .*

Matriz base: *Uma matriz \mathbf{X} de dimensão $(m \times n)$ é uma matriz base do subespaço $\mathcal{U} \subset \mathbb{R}^m$, se os vetores coluna de \mathbf{X} formam uma base de \mathcal{U} ; se os vetores coluna de \mathbf{X} forem ortonormais, ela é uma base ortonormal de \mathcal{U} .*

Lema A.2.1. *Se \mathbf{X} de dimensão $(n \times p)$ é uma matriz base do subespaço $\mathcal{U} \subset \mathbb{R}^n$, então*

- i) \mathbf{X} é uma matriz de posto p e $\mathbf{X}^\top \mathbf{X}$ é inversível;*
- ii) $\mathbf{v} \in \mathcal{U}$ se e somente se, $\mathbf{v} = \mathbf{X}\mathbf{b}$ para algum $\mathbf{b} \in \mathbb{R}^p$.*

Projeção ortogonal de um vetor: *Sejam $\mathbf{y} \in \mathbb{R}^n$ e \mathcal{U} um subespaço do \mathbb{R}^n . A projeção ortogonal de \mathbf{y} em \mathcal{U} é um vetor $\mathbf{x} \in \mathcal{U}$ tal que $\mathbf{y} - \mathbf{x} \in \mathcal{U}^\perp$.*

Teorema A.2.7. *Sejam \mathcal{U} um subespaço do \mathbb{R}^n e um vetor $\mathbf{y} \in \mathbb{R}^n$. Então,*

- i) A projeção ortogonal de \mathbf{y} em \mathcal{U} é única;*

ii) Se \mathbf{X} é a matriz base do subespaço \mathcal{U} , a projeção ortogonal de \mathbf{y} em \mathcal{U} é o vetor $\mathbf{z} = \mathbf{X}\mathbf{b}^*$ de dimensão $(n \times 1)$ em que \mathbf{b}^* é a solução do sistema $\mathbf{X}^\top \mathbf{X} \mathbf{b} = \mathbf{X}^\top \mathbf{y}$, ou seja, $\mathbf{z} = \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}$. A matriz $\mathbf{P}_\mathbf{X} = \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top$ é denominada matriz de projeção.

iii) $\mathbf{P}_\mathbf{X} \mathbf{X} = \mathbf{X}$;

iv) $\mathbf{P}_\mathbf{X}$ e $\mathbf{I} - \mathbf{P}_\mathbf{X}$ são matrizes simétricas e idempotentes;

v) $\mathcal{C}(\mathbf{P}_\mathbf{X}) = \mathcal{C}(\mathbf{X})$;

vi) $r(\mathbf{P}_\mathbf{X}) = r(\mathbf{X})$ e $r(\mathbf{I} - \mathbf{P}_\mathbf{X}) = n - r(\mathbf{X})$

Autovalor: Seja \mathbf{A} uma matriz quadrada de ordem n . As raízes do polinômio característico $|\mathbf{A} - \lambda \mathbf{I}|$, denotadas $\lambda_1, \dots, \lambda_n$, são denominadas autovalores (ou raízes características) da matriz \mathbf{A} . A equação $|\mathbf{A} - \lambda \mathbf{I}| = 0$ é denominada equação característica da matriz \mathbf{A} .

Por exemplo, se

$$\mathbf{A} = \begin{pmatrix} 1 & 4 \\ 9 & 1 \end{pmatrix},$$

seus autovalores correspondem às soluções da equação característica

$$\begin{aligned} |\mathbf{A} - \lambda \mathbf{I}| &= \left| \begin{pmatrix} 1 & 4 \\ 9 & 1 \end{pmatrix} - \lambda \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right| = \\ &= (1 - \lambda)^2 - 36 = 0, \end{aligned}$$

ou sejam, $\lambda_1 = -5$ e $\lambda_2 = 7$.

Autovetor (Vetor característico): Seja \mathbf{A} uma matriz quadrada de ordem n e λ um autovalor de \mathbf{A} . Se \mathbf{v} é um vetor (não nulo) tal que $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$, então \mathbf{v} é denominado autovetor (ou vetor característico) da matriz \mathbf{A} .

Para o exemplo acima, o autovetor associado ao autovalor $\lambda_1 = -5$ é obtido do sistema

$$\begin{pmatrix} 1 & 4 \\ 9 & 1 \end{pmatrix} \begin{pmatrix} v_{11} \\ v_{12} \end{pmatrix} = -5 \begin{pmatrix} v_{11} \\ v_{12} \end{pmatrix}$$

que tem infinitas soluções; um possível autovetor associado ao autovalor $\lambda_1 = -5$ é $\mathbf{v}_1 = (2 \ -3)^\top$. De modo similar, obtemos um autovetor associado ao autovalor $\lambda_2 = 7$, nomeadamente, $\mathbf{v}_2 = (2 \ 3)^\top$.

Teorema A.2.8. Seja \mathbf{A} uma matriz quadrada de ordem n e $\lambda_1, \dots, \lambda_n$ seus autovalores; então

i) $|\mathbf{A}| = \prod_{i=1}^n \lambda_i$;

ii) $\text{tr}(\mathbf{A}) = \sum_{i=1}^n \lambda_i$.

A.3 Formas lineares, bilineares e quadráticas

Forma linear: Uma forma linear é uma função $f : \mathbb{R}^n \rightarrow \mathbb{R}$ que associa a cada vetor $\mathbf{x} \in \mathbb{R}^n$ o número real

$$f(\mathbf{x}) = \mathbf{a}^\top \mathbf{x} = \sum_{i=1}^n a_i x_i$$

em que $\mathbf{a} \in \mathbb{R}^n$ é denominado vetor de coeficientes.

Forma bilinear: Uma forma bilinear é uma função $f : \mathbb{R}^m \times \mathbb{R}^n \rightarrow \mathbb{R}$ que associa a cada par de vetores $\mathbf{x} \in \mathbb{R}^m$ e $\mathbf{y} \in \mathbb{R}^n$ o número real

$$f(\mathbf{x}, \mathbf{y}) = \mathbf{x}^\top \mathbf{A} \mathbf{y} = \sum_{i=1}^m \sum_{j=1}^n a_{ij} x_i y_j$$

em que \mathbf{A} é uma matriz de coeficientes de dimensão $(m \times n)$.

Forma quadrática: Uma forma quadrática é uma função $f : \mathbb{R}^n \rightarrow \mathbb{R}$ que associa ao vetor $\mathbf{x} \in \mathbb{R}^n$ o número real

$$f(\mathbf{x}) = \mathbf{x}^\top \mathbf{A} \mathbf{x} = \sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i x_j$$

em que \mathbf{A} é uma matriz de coeficientes quadrada de ordem n .

Matriz definida não negativa: Uma matriz \mathbf{A} , quadrada de ordem n é denominada matriz definida não negativa se $\mathbf{x}^\top \mathbf{A} \mathbf{x} \geq 0$ para todo $\mathbf{x} \in \mathbb{R}^n$.

Matriz definida positiva: Uma matriz \mathbf{A} , quadrada de ordem n é denominada matriz definida positiva se $\mathbf{x}^\top \mathbf{A} \mathbf{x} > 0$ para todo vetor não nulo $\mathbf{x} \in \mathbb{R}^n$ e $\mathbf{x}^\top \mathbf{A} \mathbf{x} = 0$ somente quando $\mathbf{x} = \mathbf{0}$.

Matriz semidefinida positiva: Uma matriz \mathbf{A} , quadrada de ordem n é denominada matriz semidefinida positiva se $\mathbf{x}^\top \mathbf{A} \mathbf{x} \geq 0$ para $\mathbf{x} \in \mathbb{R}^n$ e $\mathbf{x}^\top \mathbf{A} \mathbf{x} = 0$ para algum vetor \mathbf{x} não nulo.

Teorema A.3.1. Seja \mathbf{A} uma matriz de dimensão $(n \times n)$ e \mathbf{M} uma matriz de dimensão $(n \times m)$. Então

- i) Se \mathbf{A} for definida não negativa, então $\mathbf{M}^\top \mathbf{A} \mathbf{M}$ é definida não negativa;
- ii) Se \mathbf{A} for definida não negativa e $r(\mathbf{M}) < m$ então $\mathbf{M}^\top \mathbf{A} \mathbf{M}$ é semidefinida positiva;
- iii) Se \mathbf{A} for definida positiva e $r(\mathbf{M}) = m$ então $\mathbf{M}^\top \mathbf{A} \mathbf{M}$ é definida positiva;

Formas quadráticas envolvendo vetores com distribuição Normal são extremamente importantes para aplicações estatísticas. Nesse contexto, apresentaremos alguns resultados bastante úteis para inferência em modelos lineares em geral. O leitor poderá consultar Searle (1971) para detalhes e demonstrações.

Teorema A.3.2. *Se $\mathbf{y} \sim N_p(\boldsymbol{\mu}, \mathbf{V})$ e \mathbf{A} é uma matriz simétrica, então*

$$i) \mathbb{E}(\mathbf{y}^\top \mathbf{A} \mathbf{y}) = \text{tr}(\mathbf{A} \mathbf{V}) + \boldsymbol{\mu}^\top \mathbf{A} \boldsymbol{\mu};$$

ii) *o cumulante de ordem r de $\mathbf{y}^\top \mathbf{A} \mathbf{y}$ é*

$$K_r(\mathbf{y}^\top \mathbf{A} \mathbf{y}) = 2^{r-1}(r-1)![\text{tr}(\mathbf{A} \mathbf{V})^r + r \boldsymbol{\mu}^\top \mathbf{A} (\mathbf{V} \mathbf{A})^{r-1} \boldsymbol{\mu}];$$

$$iii) \mathbb{C}_{\text{ov}}(\mathbf{y}, \mathbf{y}^\top \mathbf{A} \mathbf{y}) = 2 \mathbf{V} \mathbf{A} \boldsymbol{\mu};$$

O item i) prescinde da suposição de normalidade. Tomando $r = 2$, uma aplicação direta desse resultado permite-nos calcular a variância de formas quadráticas envolvendo vetores com distribuição Normal, nomeadamente

$$\mathbb{V}(\mathbf{y}^\top \mathbf{A} \mathbf{y}) = 2 \text{tr}(\mathbf{A} \mathbf{V})^2 + 4 \boldsymbol{\mu}^\top \mathbf{A} (\mathbf{V} \mathbf{A}) \boldsymbol{\mu};$$

se além disso, $\boldsymbol{\mu} = \mathbf{0}$ então $\mathbb{V}(\mathbf{y}^\top \mathbf{A} \mathbf{y}) = 2 \text{tr}(\mathbf{A} \mathbf{V})^2$.

Teorema A.3.3. *Se $\mathbf{y} \sim N_p(\boldsymbol{\mu}, \mathbf{V})$ e \mathbf{A} é uma matriz simétrica com posto r , então $\mathbf{y}^\top \mathbf{A} \mathbf{y} \sim \chi_r^2(\delta)$, em que $\delta = \frac{1}{2} \boldsymbol{\mu}^\top \mathbf{A} \boldsymbol{\mu}$ é o parâmetro de não centralidade, se e somente se $\mathbf{A} \mathbf{V}$ for idempotente.*

Teorema A.3.4. *Se $\mathbf{y} \sim N_p(\boldsymbol{\mu}, \mathbf{V})$, \mathbf{A} é uma matriz simétrica com posto r e \mathbf{B} é uma matriz com dimensão $(b \times p)$ então $\mathbf{y}^\top \mathbf{A} \mathbf{y}$ e $\mathbf{B} \mathbf{y}$ têm distribuições independentes se e somente se $\mathbf{B} \mathbf{V} \mathbf{A} = \mathbf{0}$.*

Note que o teorema não envolve o produto $\mathbf{A} \mathbf{V} \mathbf{B}$, que pode não existir.

Teorema A.3.5. *Se $\mathbf{y} \sim N_p(\boldsymbol{\mu}, \mathbf{V})$, \mathbf{A} e \mathbf{B} são matrizes simétricas então $\mathbf{y}^\top \mathbf{A} \mathbf{y}$ e $\mathbf{y}^\top \mathbf{B} \mathbf{y}$ têm distribuições independentes se e somente se $\mathbf{A} \mathbf{V} \mathbf{B} = \mathbf{0}$ ou equivalentemente se $\mathbf{B} \mathbf{V} \mathbf{A} = \mathbf{0}$.*

A.4 Decomposição de matrizes

Teorema A.4.1. *Para toda matriz simétrica \mathbf{A} de dimensão $(n \times n)$ existe uma matriz não singular \mathbf{Q} tal que $\mathbf{Q}^\top \mathbf{A} \mathbf{Q}$ é uma matriz diagonal.*

Teorema A.4.2. *Seja \mathbf{A} uma matriz de dimensão $(n \times n)$. Então existem uma matriz não singular \mathbf{Q} e uma matriz diagonal \mathbf{D} tais que $\mathbf{A} = \mathbf{Q}^\top \mathbf{D} \mathbf{Q}$.*

Teorema A.4.3. *Uma matriz \mathbf{A} (não nula) de dimensão $(n \times n)$ é simétrica definida não negativa com $r(\mathbf{A}) = r$ se e somente se existe uma matriz \mathbf{Q} de dimensão $(r \times n)$ com $r(\mathbf{A}) = r$ tal que $\mathbf{A} = \mathbf{Q}^\top \mathbf{Q}$.*

Teorema A.4.4. *Uma matriz \mathbf{A} (não nula) de dimensão $(n \times n)$ é simétrica definida positiva se e somente se existe uma matriz não singular \mathbf{Q} tal que $\mathbf{A} = \mathbf{Q}^\top \mathbf{Q}$.*

Teorema A.4.5. *Uma matriz simétrica definida não negativa \mathbf{A} de dimensão $(n \times n)$ é definida positiva se e somente se ela for não singular (ou equivalentemente, ela é semidefinida positiva se e somente se ela for singular).*

Teorema A.4.6. *Uma matriz definida positiva \mathbf{A} de dimensão $(n \times n)$ tem uma única decomposição do tipo $\mathbf{A} = \mathbf{L}^\top \mathbf{D} \mathbf{U}$ em que \mathbf{L} é uma matriz triangular inferior, \mathbf{U} é uma matriz triangular superior e \mathbf{D} é uma matriz diagonal com todos os elementos da diagonal principal positivos.*

Teorema A.4.7. *Uma matriz simétrica definida positiva \mathbf{A} de dimensão $(n \times n)$ tem uma única decomposição do tipo $\mathbf{A} = \mathbf{U}^\top \mathbf{D} \mathbf{U}$ em que \mathbf{U} é uma matriz triangular superior e \mathbf{D} é uma matriz diagonal com todos os elementos da diagonal principal positivos.*

Teorema A.4.8. *Para qualquer matriz simétrica definida positiva \mathbf{A} de dimensão $(n \times n)$ existe uma única matriz triangular superior $\mathbf{A}^{1/2}$ com todos os elementos da diagonal principal positivos tal que $\mathbf{A} = [\mathbf{A}^{1/2}]^\top \mathbf{A}^{1/2}$. Este resultado é conhecido como **Decomposição de Cholesky**.*

A.5 Derivadas de vetores e matrizes

Neste texto consideramos funções de várias variáveis expressas na forma de

i) escalares do tipo

$$f(\mathbf{x}) = \mathbf{a}^\top \mathbf{x} = \sum_{i=1}^n a_i x_i;$$

ii) vetores do tipo

$$\mathbf{f}(\mathbf{x}) = \begin{pmatrix} f_1(\mathbf{x}) \\ f_2(\mathbf{x}) \\ f_3(\mathbf{x}) \end{pmatrix}$$

em que, por exemplo, $f_1(\mathbf{x}) = x_1 + x_2$, $f_2(\mathbf{x}) = e^{x_1 x_2}$ e $f_3(\mathbf{x}) = x_1 x_2$.

iii) matrizes do tipo

$$\begin{aligned}\mathbf{F}(\mathbf{x}) &= [\mathbf{f}_1(\mathbf{x}) \quad \mathbf{f}_2(\mathbf{x}) \quad \dots \quad \mathbf{f}_n(\mathbf{x})] \\ &= \begin{pmatrix} f_{11}(\mathbf{x}) & f_{12}(\mathbf{x}) & \dots & f_{1n}(\mathbf{x}) \\ f_{21}(\mathbf{x}) & f_{22}(\mathbf{x}) & \dots & f_{2n}(\mathbf{x}) \\ \vdots & \vdots & \vdots & \vdots \\ f_{m1}(\mathbf{x}) & f_{m2}(\mathbf{x}) & \dots & f_{mn}(\mathbf{x}) \end{pmatrix};\end{aligned}$$

por exemplo, se $\mathbf{f}_1(\mathbf{x}) = (x_1 + x_2, x_1x_2, x_1 - x_2)^\top$ e $\mathbf{f}_2(\mathbf{x}) = (x_1, x_1 + x_2, x_1x_2)^\top$, então

$$\mathbf{F}(\mathbf{x}) = [\mathbf{f}_1(\mathbf{x}) \quad \mathbf{f}_2(\mathbf{x})] = \begin{pmatrix} f_{11}(\mathbf{x}) & f_{12}(\mathbf{x}) \\ f_{21}(\mathbf{x}) & f_{22}(\mathbf{x}) \\ f_{31}(\mathbf{x}) & f_{32}(\mathbf{x}) \end{pmatrix} = \begin{pmatrix} x_1 + x_2 & x_1 \\ x_1x_2 & x_1 + x_2 \\ x_1 - x_2 & x_1x_2 \end{pmatrix}. \quad (\text{A.5.1})$$

Em muitas aplicações, é possível ainda encontrar funções do tipo $\mathbf{F}(\mathbf{X})$ com \mathbf{X} denotando uma matriz de coeficientes com dimensão $(m \times n)$.

No restante desta subseção admitimos a existência de todas as derivadas mencionadas.

Vetor gradiente: Seja $f(\mathbf{x})$ uma função do vetor \mathbf{x} de dimensão $(p \times 1)$. A derivada de primeira ordem ou vetor gradiente de $f(\mathbf{x})$ é o vetor de dimensão $(p \times 1)$ dado por

$$\nabla f(\mathbf{x}) = \frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} = \left(\frac{\partial f(\mathbf{x})}{\partial x_1}, \frac{\partial f(\mathbf{x})}{\partial x_2}, \dots, \frac{\partial f(\mathbf{x})}{\partial x_p} \right)^\top.$$

Também podemos definir $\partial f(\mathbf{x})/\partial \mathbf{x}^\top = (\partial f(\mathbf{x})/\partial x_1, \dots, \partial f(\mathbf{x})/\partial x_p) = (\partial f(\mathbf{x})/\partial \mathbf{x})^\top$.

Por exemplo, seja $\mathbf{x} = (x_1, x_2, x_3)^\top$ e $f(\mathbf{x}) = 2x_1^2 + 4x_2^2 + 5x_3^2$. O gradiente de f é dado por

$$\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} = \left(\frac{\partial f(\mathbf{x})}{\partial x_1}, \frac{\partial f(\mathbf{x})}{\partial x_2}, \frac{\partial f(\mathbf{x})}{\partial x_3} \right)^\top = (4x_1, 8x_2, 10x_3)^\top.$$

Matriz hessiana: Seja $f(\mathbf{x})$ uma função do vetor \mathbf{x} de dimensão $(p \times 1)$. A matriz de derivadas segundas ou matriz hessiana de $f(\mathbf{x})$ é a matriz quadrada de ordem p dada por

$$\frac{\partial f(\mathbf{x})}{\partial \mathbf{x} \partial \mathbf{x}^\top} = \begin{pmatrix} \partial^2 f(\mathbf{x})/\partial x_1^2 & \partial^2 f(\mathbf{x})/\partial x_1 \partial x_2 & \dots & \partial^2 f(\mathbf{x})/\partial x_1 \partial x_p \\ \partial^2 f(\mathbf{x})/\partial x_2 \partial x_1 & \partial^2 f(\mathbf{x})/\partial x_2^2 & \dots & \partial^2 f(\mathbf{x})/\partial x_2 \partial x_p \\ \vdots & \vdots & \vdots & \vdots \\ \partial^2 f(\mathbf{x})/\partial x_p \partial x_1 & \partial^2 f(\mathbf{x})/\partial x_p \partial x_2 & \dots & \partial^2 f(\mathbf{x})/\partial x_p^2 \end{pmatrix}.$$

A matriz hessiana de $f(\mathbf{x})$ também é comumente denotada por $\nabla^2 f(\mathbf{x})$.

Por exemplo, se $\mathbf{x} = (x_1, x_2, x_3)^\top$ e $f(\mathbf{x}) = 2x_1^2 + 4x_2^2 + 5x_3^2$, a matriz hessiana de f é dada por

$$\begin{aligned} \frac{\partial f(\mathbf{x})}{\partial \mathbf{x} \partial \mathbf{x}^\top} &= \begin{pmatrix} \partial^2 f(\mathbf{x})/\partial x_1^2 & \partial^2 f(\mathbf{x})/\partial x_1 \partial x_2 & \partial^2 f(\mathbf{x})/\partial x_1 \partial x_3 \\ \partial^2 f(\mathbf{x})/\partial x_2 \partial x_1 & \partial^2 f(\mathbf{x})/\partial x_2^2 & \partial^2 f(\mathbf{x})/\partial x_2 \partial x_3 \\ \partial^2 f(\mathbf{x})/\partial x_3 \partial x_1 & \partial^2 f(\mathbf{x})/\partial x_3 \partial x_2 & \partial^2 f(\mathbf{x})/\partial x_3^2 \end{pmatrix} \\ &= \begin{pmatrix} 4 & 0 & 0 \\ 0 & 8 & 0 \\ 0 & 0 & 10 \end{pmatrix} \end{aligned}$$

As definições acima podem ser estendidas para funções vetoriais ou matriciais.

Matriz jacobiana: Seja $\mathbf{f}(\mathbf{x}) = [f_1(\mathbf{x}), \dots, f_m(\mathbf{x})]^\top$ um vetor de dimensão $(m \times 1)$ de funções com argumento vetorial $\mathbf{x} = (x_1, \dots, x_p)$. A matriz jacobiana de $\mathbf{f}(\mathbf{x})$ é a matriz de dimensão $(m \times p)$ dada por

$$\nabla \mathbf{f}(\mathbf{x}) = \frac{\partial \mathbf{f}(\mathbf{x})}{\partial \mathbf{x}^\top} = \begin{pmatrix} \partial \mathbf{f}_1(\mathbf{x})/\partial x_1 & \partial \mathbf{f}_1(\mathbf{x})/\partial x_2 & \dots & \partial \mathbf{f}_1(\mathbf{x})/\partial x_p \\ \partial \mathbf{f}_2(\mathbf{x})/\partial x_1 & \partial \mathbf{f}_2(\mathbf{x})/\partial x_2 & \dots & \partial \mathbf{f}_2(\mathbf{x})/\partial x_p \\ \vdots & \vdots & \vdots & \vdots \\ \partial \mathbf{f}_m(\mathbf{x})/\partial x_1 & \partial \mathbf{f}_m(\mathbf{x})/\partial x_2 & \dots & \partial \mathbf{f}_m(\mathbf{x})/\partial x_p \end{pmatrix}.$$

Para o exemplo do início da seção, a matriz jacobiana de $\mathbf{f}(\mathbf{x})$ é

$$\frac{\partial \mathbf{f}(\mathbf{x})}{\partial \mathbf{x}^\top} = \begin{pmatrix} \partial \mathbf{f}_1(\mathbf{x})/\partial x_1 & \partial \mathbf{f}_1(\mathbf{x})/\partial x_2 \\ \partial \mathbf{f}_2(\mathbf{x})/\partial x_1 & \partial \mathbf{f}_2(\mathbf{x})/\partial x_2 \\ \partial \mathbf{f}_3(\mathbf{x})/\partial x_1 & \partial \mathbf{f}_3(\mathbf{x})/\partial x_2 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ x_2 e^{x_1 x_2} & x_1 e^{x_1 x_2} \\ x_2 & x_1 \end{pmatrix};$$

Similarmente, se $\mathbf{F}(\mathbf{x}) = [\mathbf{f}_1(\mathbf{x}) \ \mathbf{f}_2(\mathbf{x}) \ \dots \ \mathbf{f}_n(\mathbf{x})]$ for uma matriz de dimensão $(m \times n)$ de funções $\mathbf{f}(\mathbf{x})$, sua derivada de primeira ordem é a matriz de dimensão $(m \times np)$ dada por

$$\nabla \mathbf{F}(\mathbf{x}) = \frac{\partial \mathbf{F}(\mathbf{x})}{\partial \mathbf{x}^\top} = [\partial \mathbf{f}_1(\mathbf{x})/\partial \mathbf{x}^\top \ \partial \mathbf{f}_2(\mathbf{x})/\partial \mathbf{x}^\top \ \dots \ \partial \mathbf{f}_n(\mathbf{x})/\partial \mathbf{x}^\top].$$

Para o exemplo no início da seção, a derivada da função $\mathbf{F}(\mathbf{x})$ é dada por

$$\begin{aligned} \frac{\partial \mathbf{F}(\mathbf{x})}{\partial \mathbf{x}^\top} &= [\partial \mathbf{f}_1(\mathbf{x})/\partial x_1 \ \partial \mathbf{f}_1(\mathbf{x})/\partial x_2 \ \partial \mathbf{f}_2(\mathbf{x})/\partial x_1 \ \partial \mathbf{f}_2(\mathbf{x})/\partial x_2] \\ &= \begin{pmatrix} 1 & 1 & 1 & 0 \\ x_2 & x_1 & 1 & 1 \\ 1 & -1 & x_2 & x_1 \end{pmatrix}. \end{aligned}$$

Seja f uma função real da matriz \mathbf{X} de dimensão $(m \times n)$ definida por

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ x_{m1} & x_{m2} & \dots & x_{mn} \end{pmatrix};$$

a derivada de f com relação à matriz \mathbf{X} é uma matriz de dimensão $(m \times n)$ dada por

$$\frac{\partial f(\mathbf{X})}{\partial \mathbf{X}} = (\partial f / \partial x_{ij}) = \begin{pmatrix} \partial f / \partial x_{11} & \partial f / \partial x_{12} & \dots & \partial f / \partial x_{1n} \\ \partial f / \partial x_{21} & \partial f / \partial x_{22} & \dots & \partial f / \partial x_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ \partial f / \partial x_{m1} & \partial f / \partial x_{m2} & \dots & \partial f / \partial x_{mn} \end{pmatrix}.$$

As regras do produto e da cadeia podem ser empregadas para derivação vetorial e matricial. Se $\mathbf{f}_1(\mathbf{x})$ e $\mathbf{f}_2(\mathbf{x})$ são duas funções vetoriais diferenciáveis de dimensão $(m \times 1)$ com argumento \mathbf{x} , então

$$\partial[\mathbf{f}_1(\mathbf{x})^\top \mathbf{f}_2(\mathbf{x})] / \partial \mathbf{x}^\top = \mathbf{f}_1(\mathbf{x})^\top (\partial \mathbf{f}_2(\mathbf{x}) / \partial \mathbf{x}^\top) + \mathbf{f}_2(\mathbf{x})^\top (\partial \mathbf{f}_1(\mathbf{x}) / \partial \mathbf{x}^\top)$$

Se o vetor $\mathbf{g}(\mathbf{z})$ com dimensão $(p \times 1)$ é função de um vetor de variáveis \mathbf{z} de dimensão $(q \times 1)$ e $\mathbf{f}(\mathbf{x})$ é uma função com argumento $\mathbf{g}(\mathbf{z})$, então

$$\partial \mathbf{f}[\mathbf{g}(\mathbf{z})] / \partial \mathbf{z}^\top = (\partial \mathbf{f}(\mathbf{x}) / \partial \mathbf{x}^\top)|_{\mathbf{x}=\mathbf{g}(\mathbf{z})} (\partial \mathbf{g}(\mathbf{z}) / \partial \mathbf{z}^\top).$$

Algumas das derivadas vetoriais e matriciais mais usadas em aplicações estatísticas são:

- i) $\partial \mathbf{a}^\top \mathbf{x} / \partial \mathbf{x} = \mathbf{a}$;
- ii) $\partial \mathbf{x}^\top \mathbf{x} / \partial \mathbf{x} = 2\mathbf{x}$;
- iii) $\partial \mathbf{x}^\top \mathbf{A} \mathbf{x} / \partial \mathbf{x} = (\mathbf{A} + \mathbf{A}^\top) \mathbf{x}$ ($= 2\mathbf{A} \mathbf{x}$, se \mathbf{A} for simétrica);
- iv) Para \mathbf{A} e \mathbf{B} simétricas,

$$\partial(\mathbf{x}^\top \mathbf{A} \mathbf{x} / \mathbf{x}^\top \mathbf{B} \mathbf{x}) / \partial \mathbf{x} = 2\mathbf{A} \mathbf{x} / \mathbf{x}^\top \mathbf{B} \mathbf{x} - 2[\mathbf{x}^\top \mathbf{A} \mathbf{x} / (\mathbf{x}^\top \mathbf{B} \mathbf{x})^2] \mathbf{B} \mathbf{x};$$

- v) Para \mathbf{A} simétrica,

$$\partial[\mathbf{y} - \mathbf{g}(\mathbf{x})]^\top \mathbf{A} [\mathbf{y} - \mathbf{g}(\mathbf{x})] / \partial \mathbf{x} = -2\mathbf{D}(\mathbf{x})^\top \mathbf{A} [\mathbf{y} - \mathbf{g}(\mathbf{x})]$$

em que $\mathbf{D}(\mathbf{x}) = \partial \mathbf{g}(\mathbf{x}) / \partial \mathbf{x}^\top$;

vi) Para matrizes \mathbf{A} com dimensão $(m \times n)$, e \mathbf{B} com dimensão $(n \times q)$,

$$\partial \text{tr}(\mathbf{A}(\mathbf{x})\mathbf{B}(\mathbf{x}))/\partial \mathbf{x} = \partial \text{tr}(\mathbf{A}(\mathbf{x})\mathbf{B}(\mathbf{z}))/\partial \mathbf{x}|_{\mathbf{z}=\mathbf{x}} + \partial \text{tr}(\mathbf{A}(\mathbf{z})\mathbf{B}(\mathbf{x}))/\partial \mathbf{x}|_{\mathbf{z}=\mathbf{x}}.$$

vii) $\partial |\mathbf{X}|/\partial x_{ij} = |\mathbf{X}_{ij}|$, em que $|\mathbf{X}_{ij}|$ é o cofator de x_{ij} ;

viii) $\partial \ln |\mathbf{X}|/\partial x_{ij} = \text{tr}(\mathbf{X}^{-1}(\partial \mathbf{X}/\partial x_{ij}))$;

ix) $\partial \mathbf{X}^{-1}/\partial x_{ij} = -\mathbf{X}^{-1}(\partial \mathbf{X}/\partial x_{ij})\mathbf{X}^{-1}$;

x) $\partial [\text{tr} \mathbf{X}]/\partial x_{ij} = \text{tr}[\partial \mathbf{X}/\partial x_{ij}]$;

xi) $\partial \text{tr}(\mathbf{A}\mathbf{X}^{-1}\mathbf{B})/\partial \mathbf{X} = -(\mathbf{X}^{-1}\mathbf{B}\mathbf{A}\mathbf{X}^{-1})^\top$;

xii) $\partial |\mathbf{X}|/\partial \mathbf{X} = |\mathbf{X}|(\mathbf{X}^{-1})^\top$;

xiii) $\partial \ln(|\mathbf{X}|)/\partial \mathbf{X} = \{1/|\mathbf{X}|\}\{\partial |\mathbf{X}|/\partial \mathbf{X}\} = (\mathbf{X}^{-1})^\top$;

xiv) $\partial \text{tr}(\mathbf{A}\mathbf{X})/\partial \mathbf{X} = \mathbf{A}^\top$;

xv) $\partial |\mathbf{A}\mathbf{X}|/\partial \mathbf{X} = |\mathbf{A}\mathbf{X}|((\mathbf{A}\mathbf{X})^{-1}\mathbf{A})^\top$.

Se \mathbf{X} é uma matriz com dimensão $(p \times q)$ e $\mathbf{U}(\mathbf{X})$ é uma matriz quadrada de ordem p ,

xvi) $\partial \text{tr}(\mathbf{U}(\mathbf{X})^{-1}\mathbf{A})/\partial \mathbf{X} = -(\partial/\partial \mathbf{X})\text{tr}(\mathbf{U}(\mathbf{Z})^{-1}\mathbf{A}\mathbf{U}(\mathbf{Z})^{-1}\mathbf{U}(\mathbf{X}))|_{\mathbf{Z}=\mathbf{X}}$;

xvii) $\partial |\mathbf{U}(\mathbf{X})|/\partial \mathbf{X} = |\mathbf{U}(\mathbf{X})|(\partial/\partial \mathbf{X})\text{tr}(\mathbf{U}(\mathbf{Z})^{-1}\mathbf{U}(\mathbf{X}))|_{\mathbf{Z}=\mathbf{X}}$.

Algumas derivadas envolvendo o operador vec são:

xviii) $\partial \text{vec}(\mathbf{A}\mathbf{X}\mathbf{B})/\partial \text{vec}(\mathbf{X})^\top = \mathbf{B}^\top \otimes \mathbf{A}$;

xix) Sejam $\mathbf{U}(\mathbf{X})$ uma matriz com dimensão $(m \times n)$ e $\mathbf{V}(\mathbf{X})$ uma matriz com dimensão $(n \times r)$, então

$$\frac{\partial \text{vec}[\mathbf{U}(\mathbf{X})\mathbf{V}(\mathbf{X})]}{\partial \text{vec}(\mathbf{X})^\top} = (\mathbf{V} \otimes \mathbf{I}_m)^\top \frac{\partial \text{vec}[\mathbf{U}(\mathbf{X})]}{\partial \text{vec}(\mathbf{X})^\top} + (\mathbf{I}_r \otimes \mathbf{U})^\top \frac{\partial \text{vec}[\mathbf{V}(\mathbf{X})]}{\partial \text{vec}(\mathbf{X})^\top};$$

xx) Seja \mathbf{X} uma matriz quadrada; então

$$\frac{\partial \text{vec}(\mathbf{A}\mathbf{X}^{-1}\mathbf{B})}{\partial \text{vec}(\mathbf{X})^\top} = -(\mathbf{X}\mathbf{B})^\top \otimes (\mathbf{A}\mathbf{X}^{-1});$$

xxi) Para uma matriz não singular $\mathbf{U}(\mathbf{X})$,

$$\frac{\partial \text{vec}[\mathbf{U}(\mathbf{X})^{-1}]}{\partial \text{vec}(\mathbf{X})^\top} = -[(\mathbf{U}(\mathbf{X})^{-1})^\top \otimes \mathbf{U}(\mathbf{X})^{-1}] \frac{\partial \text{vec}[\mathbf{U}(\mathbf{X})]}{\partial \text{vec}(\mathbf{X})^\top};$$

$$\text{xxii)} \quad \frac{\partial \text{vec}\{\mathbf{F}[\mathbf{G}(\mathbf{X})]\}}{\partial \text{vec}(\mathbf{Z})^\top} = \frac{\partial \text{vec}[\mathbf{F}(\mathbf{X})]}{\partial \text{vec}(\mathbf{X})^\top} \bigg|_{\mathbf{X}=\mathbf{G}(\mathbf{Z})} \frac{\partial \text{vec}[\mathbf{G}(\mathbf{Z})]}{\partial \text{vec}(\mathbf{Z})^\top}.$$

Para ilustrar a aplicação de derivadas de vetores e matrizes, inicialmente, consideremos a função

$$\begin{aligned} g(\boldsymbol{\beta}, \boldsymbol{\theta}) &= [\mathbf{y} - \mathbf{f}(\boldsymbol{\beta})]^\top [\mathbf{V}(\boldsymbol{\theta})]^{-1} [\mathbf{y} - \mathbf{f}(\boldsymbol{\beta})] + \ln |\mathbf{V}(\boldsymbol{\theta})| \\ &= \text{tr}\{[\mathbf{V}(\boldsymbol{\theta})]^{-1} [\mathbf{y} - \mathbf{f}(\boldsymbol{\beta})][\mathbf{y} - \mathbf{f}(\boldsymbol{\beta})]^\top\} + \ln |\mathbf{V}(\boldsymbol{\theta})| \end{aligned}$$

com argumentos $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top$ e $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)^\top$, com $\boldsymbol{\theta}$ funcionalmente independente de $\boldsymbol{\beta}$, em que $\mathbf{V}(\boldsymbol{\theta})$ é uma matriz simétrica definida positiva. Além disso, considere que $\mathbf{f}(\boldsymbol{\beta})$ e $\mathbf{V}(\boldsymbol{\theta})$ sejam funções diferenciáveis. O gradiente de $g(\boldsymbol{\beta}, \boldsymbol{\theta})$ é dado por:

$$\nabla g(\boldsymbol{\beta}, \boldsymbol{\theta}) = \begin{pmatrix} (\partial/\partial \boldsymbol{\beta})g(\boldsymbol{\beta}, \boldsymbol{\theta}) \\ (\partial/\partial \boldsymbol{\theta})g(\boldsymbol{\beta}, \boldsymbol{\theta}) \end{pmatrix}$$

em que

i) a derivada em relação a $\boldsymbol{\beta}$ é

$$\begin{aligned} \frac{\partial g(\boldsymbol{\beta}, \boldsymbol{\theta})}{\partial \boldsymbol{\beta}} &= \frac{\partial [\mathbf{y} - \mathbf{f}(\boldsymbol{\beta})]^\top [\mathbf{V}(\boldsymbol{\theta})]^{-1} [\mathbf{y} - \mathbf{f}(\boldsymbol{\beta})]}{\partial \boldsymbol{\beta}} \\ &= -2 \left(\frac{\partial [\mathbf{f}(\boldsymbol{\beta})]^\top}{\partial \boldsymbol{\beta}} \right) [\mathbf{V}(\boldsymbol{\theta})]^{-1} [\mathbf{y} - \mathbf{f}(\boldsymbol{\beta})]. \end{aligned}$$

ii) a derivada em relação a θ_j para $j = 1, \dots, k$ é

$$\begin{aligned}
 \frac{\partial g(\boldsymbol{\beta}, \boldsymbol{\theta})}{\partial \theta_j} &= \frac{\partial [\mathbf{y} - \mathbf{f}(\boldsymbol{\beta})]^\top [\mathbf{V}(\boldsymbol{\theta})]^{-1} [\mathbf{y} - \mathbf{f}(\boldsymbol{\beta})]}{\partial \theta_j} + \frac{\partial \ln |\mathbf{V}(\boldsymbol{\theta})|}{\partial \theta_j} \\
 &= [\mathbf{y} - \mathbf{f}(\boldsymbol{\beta})]^\top \frac{\partial [\mathbf{V}(\boldsymbol{\theta})]^{-1}}{\partial \theta_j} [\mathbf{y} - \mathbf{f}(\boldsymbol{\beta})] + \frac{\partial \ln |\mathbf{V}(\boldsymbol{\theta})|}{\partial \theta_j} \\
 &= -[\mathbf{y} - \mathbf{f}(\boldsymbol{\beta})]^\top [\mathbf{V}(\boldsymbol{\theta})]^{-1} \frac{\partial \mathbf{V}(\boldsymbol{\theta})}{\partial \theta_j} [\mathbf{V}(\boldsymbol{\theta})]^{-1} [\mathbf{y} - \mathbf{f}(\boldsymbol{\beta})] \\
 &\quad + \text{tr} \left\{ [\mathbf{V}(\boldsymbol{\theta})]^{-1} \frac{\partial \mathbf{V}(\boldsymbol{\theta})}{\partial \theta_j} \right\} \\
 &= -\text{tr} \left\{ [\mathbf{V}(\boldsymbol{\theta})]^{-1} \frac{\partial [\mathbf{V}(\boldsymbol{\theta})]}{\partial \theta_j} [\mathbf{V}(\boldsymbol{\theta})]^{-1} [\mathbf{y} - \mathbf{f}(\boldsymbol{\beta})] [(\mathbf{y} - \mathbf{f}(\boldsymbol{\beta}))^\top] \right\} \\
 &\quad + \text{tr} \left\{ [\mathbf{V}(\boldsymbol{\theta})]^{-1} \frac{\partial \mathbf{V}(\boldsymbol{\theta})}{\partial \theta_j} \right\} \\
 &= \text{tr} \left\{ [\mathbf{V}(\boldsymbol{\theta})]^{-1} (-[\mathbf{y} - \mathbf{f}(\boldsymbol{\beta})][\mathbf{y} - \mathbf{f}(\boldsymbol{\beta})]^\top + \mathbf{V}(\boldsymbol{\theta})) [\mathbf{V}(\boldsymbol{\theta})]^{-1} \frac{\partial \mathbf{V}(\boldsymbol{\theta})}{\partial \theta_j} \right\}.
 \end{aligned}$$

A matriz de segundas derivadas de $g(\boldsymbol{\beta}, \boldsymbol{\theta})$ em relação a $\boldsymbol{\beta}$ é:

$$\begin{aligned}
 \frac{\partial^2 g(\boldsymbol{\beta}, \boldsymbol{\theta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} &= \frac{\partial \left\{ -2 \left(\frac{\partial [\mathbf{f}(\boldsymbol{\beta})]^\top}{\partial \boldsymbol{\beta}} \right) [\mathbf{V}(\boldsymbol{\theta})]^{-1} [\mathbf{y} - \mathbf{f}(\boldsymbol{\beta})] \right\}}{\partial \boldsymbol{\beta}^\top} \\
 &= -2 \frac{\partial^2 [\mathbf{f}(\boldsymbol{\beta})]^\top}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} [\mathbf{V}(\boldsymbol{\theta})]^{-1} [\mathbf{y} - \mathbf{f}(\boldsymbol{\beta})] \\
 &\quad + 2 \left(\frac{\partial [\mathbf{f}(\boldsymbol{\beta})]^\top}{\partial \boldsymbol{\beta}} \right) [\mathbf{V}(\boldsymbol{\theta})]^{-1} \left(\frac{\partial \mathbf{f}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^\top} \right).
 \end{aligned}$$

Para obter $\partial^2 g(\boldsymbol{\beta}, \boldsymbol{\theta}) / \partial \beta_i \partial \theta_j$, notemos que para $i = 1, \dots, p$

$$\begin{aligned}
 \frac{\partial g(\boldsymbol{\beta}, \boldsymbol{\theta})}{\partial \beta_i} &= \frac{\partial [\mathbf{y} - \mathbf{f}(\boldsymbol{\beta})]^\top [\mathbf{V}(\boldsymbol{\theta})]^{-1} [\mathbf{y} - \mathbf{f}(\boldsymbol{\beta})]}{\partial \beta_i} \\
 &= -2 \left(\frac{\partial [\mathbf{f}(\boldsymbol{\beta})]^\top}{\partial \beta_i} \right) [\mathbf{V}(\boldsymbol{\theta})]^{-1} [\mathbf{y} - \mathbf{f}(\boldsymbol{\beta})].
 \end{aligned}$$

e que $\partial^2 g(\boldsymbol{\beta}, \boldsymbol{\theta}) / \partial \beta_i \partial \theta_j = \partial^2 g(\boldsymbol{\beta}, \boldsymbol{\theta}) / \partial \theta_j \partial \beta_i$ para $i = 1, \dots, p$, $j = 1, \dots, k$; então

$$\begin{aligned} \frac{\partial^2 g(\boldsymbol{\beta}, \boldsymbol{\theta})}{\partial \beta_i \partial \theta_j} &= \frac{\partial \left\{ -2 \left(\partial [\mathbf{f}(\boldsymbol{\beta})]^\top / \partial \beta_i \right) [\mathbf{V}(\boldsymbol{\theta})]^{-1} [\mathbf{y} - \mathbf{f}(\boldsymbol{\beta})] \right\}}{\partial \theta_j} \\ &= -2 \left(\frac{\partial [\mathbf{f}(\boldsymbol{\beta})]^\top}{\partial \beta_i} \right) \frac{\partial [\mathbf{V}(\boldsymbol{\theta})]^{-1}}{\partial \theta_j} [\mathbf{y} - \mathbf{f}(\boldsymbol{\beta})] \\ &= -2 \left(\frac{\partial [\mathbf{f}(\boldsymbol{\beta})]^\top}{\partial \beta_i} \right) [\mathbf{V}(\boldsymbol{\theta})]^{-1} \frac{\partial \mathbf{V}(\boldsymbol{\theta})}{\partial \theta_j} [\mathbf{V}(\boldsymbol{\theta})]^{-1} [\mathbf{y} - \mathbf{f}(\boldsymbol{\beta})]. \end{aligned}$$

Fazendo uso da propriedade (x), as derivadas $\partial^2 g(\boldsymbol{\beta}, \boldsymbol{\theta}) / \partial \theta_s \partial \theta_j$ para $j, s = 1, \dots, k$ são

$$\begin{aligned} \frac{\partial^2 g(\boldsymbol{\beta}, \boldsymbol{\theta})}{\partial \theta_s \partial \theta_j} &= \frac{\partial}{\partial \theta_s} \left\{ \text{tr} \left([\mathbf{V}(\boldsymbol{\theta})]^{-1} \left\{ -[\mathbf{y} - \mathbf{f}(\boldsymbol{\beta})][\mathbf{y} - \mathbf{f}(\boldsymbol{\beta})]^\top + \mathbf{V}(\boldsymbol{\theta}) \right\} [\mathbf{V}(\boldsymbol{\theta})]^{-1} \frac{\partial \mathbf{V}(\boldsymbol{\theta})}{\partial \theta_j} \right) \right\} \\ &= \frac{\partial \left\{ -[\mathbf{y} - \mathbf{f}(\boldsymbol{\beta})]^\top [\mathbf{V}(\boldsymbol{\theta})]^{-1} [\partial \mathbf{V}(\boldsymbol{\theta}) / \partial \theta_j] [\mathbf{V}(\boldsymbol{\theta})]^{-1} [\mathbf{y} - \mathbf{f}(\boldsymbol{\beta})] \right\}}{\partial \theta_s} \\ &\quad + \frac{\partial \left\{ \text{tr} ([\mathbf{V}(\boldsymbol{\theta})]^{-1} [\partial \mathbf{V}(\boldsymbol{\theta}) / \partial \theta_j]) \right\}}{\partial \theta_s} \\ &= -[\mathbf{y} - \mathbf{f}(\boldsymbol{\beta})]^\top \frac{\partial ([\mathbf{V}(\boldsymbol{\theta})]^{-1} [\partial \mathbf{V}(\boldsymbol{\theta}) / \partial \theta_j] [\mathbf{V}(\boldsymbol{\theta})]^{-1})}{\partial \theta_s} [\mathbf{y} - \mathbf{f}(\boldsymbol{\beta})] \\ &\quad + \text{tr} \left\{ \frac{\partial ([\mathbf{V}(\boldsymbol{\theta})]^{-1} [\partial \mathbf{V}(\boldsymbol{\theta}) / \partial \theta_j])}{\partial \theta_s} \right\}. \end{aligned}$$

Utilizando a regra do produto para derivadas de matrizes, temos que

$$\begin{aligned} \frac{\partial \{ [\mathbf{V}(\boldsymbol{\theta})]^{-1} [\partial \mathbf{V}(\boldsymbol{\theta}) / \partial \theta_j] \}}{\partial \theta_s} &= \frac{\partial [\mathbf{V}(\boldsymbol{\theta})]^{-1}}{\partial \theta_s} \frac{\partial \mathbf{V}(\boldsymbol{\theta})}{\partial \theta_j} + [\mathbf{V}(\boldsymbol{\theta})]^{-1} \frac{\partial}{\partial \theta_s} \left[\frac{\partial \mathbf{V}(\boldsymbol{\theta})}{\partial \theta_j} \right] \\ &= -[\mathbf{V}(\boldsymbol{\theta})]^{-1} \frac{\partial \mathbf{V}(\boldsymbol{\theta})}{\partial \theta_s} [\mathbf{V}(\boldsymbol{\theta})]^{-1} \frac{\partial \mathbf{V}(\boldsymbol{\theta})}{\partial \theta_j} + [\mathbf{V}(\boldsymbol{\theta})]^{-1} \frac{\partial^2 \mathbf{V}(\boldsymbol{\theta})}{\partial \theta_s \partial \theta_j} \end{aligned}$$

e

$$\begin{aligned}
\frac{\partial}{\partial \theta_s} \left\{ [\mathbf{V}(\boldsymbol{\theta})]^{-1} \frac{\partial \mathbf{V}(\boldsymbol{\theta})}{\partial \theta_j} [\mathbf{V}(\boldsymbol{\theta})]^{-1} \right\} &= \frac{\partial \{ [\mathbf{V}(\boldsymbol{\theta})]^{-1} [\partial \mathbf{V}(\boldsymbol{\theta}) / \partial \theta_j] \}}{\partial \theta_s} \\
&\quad + [\mathbf{V}(\boldsymbol{\theta})]^{-1} + \mathbf{V}(\boldsymbol{\theta}) \frac{\partial \mathbf{V}(\boldsymbol{\theta})}{\partial \theta_j} \frac{\partial [\mathbf{V}(\boldsymbol{\theta})]^{-1}}{\partial \theta_s} \\
&= -[\mathbf{V}(\boldsymbol{\theta})]^{-1} \frac{\partial \mathbf{V}(\boldsymbol{\theta})}{\partial \theta_s} [\mathbf{V}(\boldsymbol{\theta})]^{-1} \frac{\partial \mathbf{V}(\boldsymbol{\theta})}{\partial \theta_j} [\mathbf{V}(\boldsymbol{\theta})]^{-1} \\
&\quad + [\mathbf{V}(\boldsymbol{\theta})]^{-1} \frac{\partial^2 \mathbf{V}(\boldsymbol{\theta})}{\partial \theta_s \partial \theta_j} [\mathbf{V}(\boldsymbol{\theta})]^{-1} \\
&\quad - [\mathbf{V}(\boldsymbol{\theta})]^{-1} \frac{\partial \mathbf{V}(\boldsymbol{\theta})}{\partial \theta_j} [\mathbf{V}(\boldsymbol{\theta})]^{-1} \frac{\partial \mathbf{V}(\boldsymbol{\theta})}{\partial \theta_s} [\mathbf{V}(\boldsymbol{\theta})]^{-1}.
\end{aligned}$$

Então

$$\begin{aligned}
\frac{\partial^2 g(\boldsymbol{\beta}, \boldsymbol{\theta})}{\partial \theta_s \partial \theta_j} &= [\mathbf{y} - \mathbf{f}(\boldsymbol{\beta})]^\top [\mathbf{V}(\boldsymbol{\theta})]^{-1} \frac{\partial \mathbf{V}(\boldsymbol{\theta})}{\partial \theta_s} [\mathbf{V}(\boldsymbol{\theta})]^{-1} \frac{\partial \mathbf{V}(\boldsymbol{\theta})}{\partial \theta_j} [\mathbf{V}(\boldsymbol{\theta})]^{-1} [\mathbf{y} - \mathbf{f}(\boldsymbol{\beta})] \\
&\quad - [\mathbf{y} - \mathbf{f}(\boldsymbol{\beta})]^\top [\mathbf{V}(\boldsymbol{\theta})]^{-1} \frac{\partial^2 \mathbf{V}(\boldsymbol{\theta})}{\partial \theta_s \partial \theta_j} [\mathbf{V}(\boldsymbol{\theta})]^{-1} [\mathbf{y} - \mathbf{f}(\boldsymbol{\beta})] \\
&\quad + [\mathbf{y} - \mathbf{f}(\boldsymbol{\beta})]^\top [\mathbf{V}(\boldsymbol{\theta})]^{-1} \frac{\partial \mathbf{V}(\boldsymbol{\theta})}{\partial \theta_j} [\mathbf{V}(\boldsymbol{\theta})]^{-1} \frac{\partial \mathbf{V}(\boldsymbol{\theta})}{\partial \theta_s} [\mathbf{V}(\boldsymbol{\theta})]^{-1} [\mathbf{y} - \mathbf{f}(\boldsymbol{\beta})] \\
&\quad + \operatorname{tr} \left\{ -[\mathbf{V}(\boldsymbol{\theta})]^{-1} \frac{\partial \mathbf{V}(\boldsymbol{\theta})}{\partial \theta_s} [\mathbf{V}(\boldsymbol{\theta})]^{-1} \frac{\partial \mathbf{V}(\boldsymbol{\theta})}{\partial \theta_j} + [\mathbf{V}(\boldsymbol{\theta})]^{-1} \frac{\partial^2 \mathbf{V}(\boldsymbol{\theta})}{\partial \theta_s \partial \theta_j} \right\}.
\end{aligned}$$

Além disso, como $[\mathbf{V}(\boldsymbol{\theta})]^{-1}$, $\partial \mathbf{V}(\boldsymbol{\theta}) / \partial \theta_j$ e $\partial \mathbf{V}(\boldsymbol{\theta}) / \partial \theta_s$ são matrizes simétricas e

$$[\mathbf{y} - \mathbf{f}(\boldsymbol{\beta})]^\top [\mathbf{V}(\boldsymbol{\theta})]^{-1} \frac{\partial \mathbf{V}(\boldsymbol{\theta})}{\partial \theta_s} [\mathbf{V}(\boldsymbol{\theta})]^{-1} \frac{\partial \mathbf{V}(\boldsymbol{\theta})}{\partial \theta_j} [\mathbf{V}(\boldsymbol{\theta})]^{-1} [\mathbf{y} - \mathbf{f}(\boldsymbol{\beta})]$$

é um escalar, obtemos a igualdade

$$\begin{aligned}
[\mathbf{y} - \mathbf{f}(\boldsymbol{\beta})]^\top [\mathbf{V}(\boldsymbol{\theta})]^{-1} \frac{\partial \mathbf{V}(\boldsymbol{\theta})}{\partial \theta_s} [\mathbf{V}(\boldsymbol{\theta})]^{-1} \frac{\partial \mathbf{V}(\boldsymbol{\theta})}{\partial \theta_j} [\mathbf{V}(\boldsymbol{\theta})]^{-1} [\mathbf{y} - \mathbf{f}(\boldsymbol{\beta})] &= \\
[\mathbf{y} - \mathbf{f}(\boldsymbol{\beta})]^\top [\mathbf{V}(\boldsymbol{\theta})]^{-1} \frac{\partial \mathbf{V}(\boldsymbol{\theta})}{\partial \theta_j} [\mathbf{V}(\boldsymbol{\theta})]^{-1} \frac{\partial \mathbf{V}(\boldsymbol{\theta})}{\partial \theta_s} [\mathbf{V}(\boldsymbol{\theta})]^{-1} [\mathbf{y} - \mathbf{f}(\boldsymbol{\beta})].
\end{aligned}$$

Portanto, utilizando as propriedades do traço de matrizes e os resultados acima, as derivadas de $g(\boldsymbol{\beta}, \boldsymbol{\theta})$ em relação a θ_j e θ_s para $j, s = 1, \dots, k$ são dadas por

$$\begin{aligned}
\frac{\partial^2 g(\boldsymbol{\beta}, \boldsymbol{\theta})}{\partial \theta_s \partial \theta_j} &= \operatorname{tr} \{ \mathbf{A} (2[\mathbf{y} - \mathbf{f}(\boldsymbol{\beta})][\mathbf{y} - \mathbf{f}(\boldsymbol{\beta})]^\top - \mathbf{V}(\boldsymbol{\theta})) \} \\
&\quad + \operatorname{tr} \{ \mathbf{B} (-[\mathbf{y} - \mathbf{f}(\boldsymbol{\beta})][\mathbf{y} - \mathbf{f}(\boldsymbol{\beta})]^\top + \mathbf{V}(\boldsymbol{\theta})) \},
\end{aligned}$$

em que

$$\begin{aligned}\mathbf{A} &= [\mathbf{V}(\boldsymbol{\theta})]^{-1} \frac{\partial \mathbf{V}(\boldsymbol{\theta})}{\partial \theta_s} [\mathbf{V}(\boldsymbol{\theta})]^{-1} \frac{\partial \mathbf{V}(\boldsymbol{\theta})}{\partial \theta_j} [\mathbf{V}(\boldsymbol{\theta})]^{-1}, \\ \mathbf{B} &= [\mathbf{V}(\boldsymbol{\theta})]^{-1} \frac{\partial^2 \mathbf{V}(\boldsymbol{\theta})}{\partial \theta_s \partial \theta_j} [\mathbf{V}(\boldsymbol{\theta})]^{-1}.\end{aligned}$$

A.6 Exercícios

A.6.1. Sejam \mathbf{x} e \mathbf{y} vetores com dimensão $(n \times 1)$. Prove a desigualdade de Cauchy-Schwarz:

$$(\mathbf{x}^\top \mathbf{y})^2 \leq (\mathbf{x}^\top \mathbf{x})(\mathbf{y}^\top \mathbf{y}).$$

Em que condições a igualdade é válida?

Sugestão: Use a desigualdade $\|\mathbf{x} + \lambda \mathbf{y}\|^2 \geq 0$, $\forall \mathbf{x}, \mathbf{y}$ e $\lambda \in \mathbb{R}$.

A.6.2. Seja $\mathbf{x} = (x_1, \dots, x_n)^\top$ um vetor com dimensão $(n \times 1)$ e defina $\|\mathbf{x}\| = \sqrt{\mathbf{x}^\top \mathbf{x}} = (\sum_{i=1}^n x_i^2)^{1/2}$. Mostre que

- i) $\mathbf{x}^\top \mathbf{x} \geq 0$, $\forall \mathbf{x}$;
- ii) $\|\mathbf{x}\| = 0 \Leftrightarrow \mathbf{x} = (0, \dots, 0)^\top = \mathbf{0}$;
- iii) $\|c\mathbf{x}\| = |c|\|\mathbf{x}\|$, $\forall c \in \mathbb{R}$;
- iv) $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$, $\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n$.

Sugestão: Use os resultados do Exercício A.6.1.

A.6.3. Sejam \mathbf{A} e \mathbf{B} duas matrizes quadradas de ordem n , \mathbf{C} e \mathbf{D} duas matrizes com dimensões $(m \times n)$ e $(n \times m)$, respectivamente e $\mathbf{x} = (x_1, \dots, x_n)^\top$, um vetor com dimensão $(n \times 1)$. Prove que

- i) $\text{tr}(\mathbf{A}^\top) = \text{tr}(\mathbf{A})$;
- ii) $\text{tr}(\mathbf{A} + \mathbf{B}) = \text{tr}(\mathbf{A}) + \text{tr}(\mathbf{B})$;
- iii) $\text{tr}(\mathbf{CD}) = \text{tr}(\mathbf{DC})$;
- iv) $\text{tr}(\mathbf{AA}^\top) = \text{tr}(\mathbf{A}^\top \mathbf{A}) = \sum_{i=1}^n \sum_{j=1}^n a_{ij}^2$;
- v) $\|\mathbf{x}\|^2 = \sum_{i=1}^n x_i^2 = \mathbf{x}^\top \mathbf{x} = \text{tr}(\mathbf{xx}^\top)$.

vi) Construa exemplos numéricos para ilustrar as propriedades acima.

A.6.4. Seja \mathbf{A} uma matriz quadrada *idempotente*, i.e., tal que $\mathbf{A}^2 = \mathbf{A}\mathbf{A} = \mathbf{A}$. Além disso, seja $\mathbf{0}_n$ a matriz nula de dimensão $(n \times n)$. Prove que

- i) $|\mathbf{A}| = 0$ ou $|\mathbf{A}| = 1$;
- ii) Os autovalores de \mathbf{A} são iguais a 0 ou 1;
- iii) $\mathbf{I}_n - \mathbf{A}$ é idempotente e $\mathbf{A}(\mathbf{I}_n - \mathbf{A}) = (\mathbf{I}_n - \mathbf{A})\mathbf{A} = \mathbf{0}_n$;
- iv) Suponha adicionalmente que \mathbf{A} é simétrica e prove que

- a) $r(\mathbf{A}) = \text{tr}(\mathbf{A})$;
- b) $r(\mathbf{A}) = n \Rightarrow \mathbf{A} = \mathbf{I}_n$.

Sugestão: Utilize a decomposição espectral de \mathbf{A} .

Observação: Os resultados são válidos mesmo quando a matriz \mathbf{A} não é simétrica.

v) Construa exemplos numéricos (com matrizes diferentes da matriz identidade) para ilustrar as propriedades acima.

A.6.5. Mostre que se \mathbf{X} é uma matriz tal que $\mathbf{X}^\top \mathbf{X} = \mathbf{X}$ então ela é simétrica e idempotente. Construa um exemplo numérico (com \mathbf{X} diferente da matriz identidade) para ilustrar a propriedade acima.

A.6.6. Seja \mathbf{A} uma matriz quadrada cujos autovalores são representados por $\lambda_1, \dots, \lambda_n$. Prove que

$$|\mathbf{A}| = \prod_{i=1}^n \lambda_i.$$

Construa um exemplo numérico (com \mathbf{X} diferente da matriz identidade) para ilustrar a propriedade acima.

A.6.7. Uma matriz \mathbf{A} quadrada de ordem n é positiva definida se $\mathbf{x}^\top \mathbf{A} \mathbf{x} > 0 \quad \forall \mathbf{x} \neq \mathbf{0}$. Seja \mathbf{X} uma matriz $(n \times p)$, $n > p$, de posto completo. Mostre que

- i) $\mathbf{X}^\top \mathbf{X}$ é simétrica;
- ii) $\mathbf{X}^\top \mathbf{X}$ é p.d.;
- iii) como uma matriz é positiva definida se e somente se todos os seus autovalores são positivos, prove que $\mathbf{X}^\top \mathbf{X}$ é inversível;

- iv) $\mathbf{H} = \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top$ é uma matriz simétrica e idempotente;
- v) $\mathbf{I}_n - \mathbf{H}$ é uma matriz simétrica e idempotente;
- vi) $r(\mathbf{I}_n - \mathbf{H}) = n - p$.
- vii) Construa exemplos numéricos (com matrizes diferentes da matriz identidade) para ilustrar as propriedades acima.

A.6.8. Seja a matriz

$$\mathbf{A} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}.$$

Para que valores de ρ a matriz \mathbf{A} é positiva definida?

A.6.9. Seja $f(\mathbf{X})$ uma função real de uma matriz \mathbf{X} com dimensão $(m \times n)$ e elementos x_{ij} , $i = 1, \dots, m$, $j = 1, \dots, n$. A derivada de f com respeito a \mathbf{X} é definida como sendo a matriz $(m \times n)$ de derivadas $\partial f(\mathbf{X})/\partial x_{ij}$:

$$\frac{\partial f(\mathbf{X})}{\partial \mathbf{X}} := \begin{pmatrix} \frac{\partial f(\mathbf{X})}{\partial x_{11}} & \cdots & \frac{\partial f(\mathbf{X})}{\partial x_{1n}} \\ \vdots & \ddots & \vdots \\ \frac{\partial f(\mathbf{X})}{\partial x_{m1}} & \cdots & \frac{\partial f(\mathbf{X})}{\partial x_{mn}} \end{pmatrix}.$$

Sejam $\mathbf{x} = (x_1, \dots, x_n)^\top$ e $\mathbf{a} = (a_1, \dots, a_n)^\top$, vetores com dimensão $(n \times 1)$, \mathbf{A} , uma matriz quadrada de ordem n e \mathbf{B} , uma matriz com dimensão $m \times n$. Mostre que

- i) $\partial \mathbf{a}^\top \mathbf{x} / \partial \mathbf{x} = \partial \mathbf{x}^\top \mathbf{a} / \partial \mathbf{x} = \mathbf{a}$;
- ii) $\partial \mathbf{Bx} / \partial \mathbf{x} = \mathbf{B}^\top$;
- iii) $\partial \mathbf{x}^\top \mathbf{Ax} / \partial \mathbf{x} = (\mathbf{A} + \mathbf{A}^\top) \mathbf{x}$;
- iv) $\partial \mathbf{x}^\top \mathbf{Ax} / \partial \mathbf{x} = 2\mathbf{Ax}$ se \mathbf{A} for simétrica.

A.6.10. Seja $\mathbf{x} = (x_1, \dots, x_n)^\top$ um vetor de n observações de uma certa variável X . Mostre que a média e a variância amostral das observações podem ser escritas na seguinte forma matricial:

$$\begin{aligned} \bar{x} &= \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{n} \mathbf{1}_n^\top \mathbf{x} \\ s_x^2 &= \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n-1} (\mathbf{x} - \mathbf{1}_n \bar{x})^\top (\mathbf{x} - \mathbf{1}_n \bar{x}) \\ &= \frac{1}{n-1} \mathbf{x}^\top (\mathbf{I}_n - \frac{1}{n} \mathbf{J}_n) \mathbf{x}, \end{aligned}$$

em que $\mathbf{1}_n$ representa um vetor de dimensão $(n \times 1)$ com todos elementos iguais a 1 e $\mathbf{J}_n = \mathbf{1}_n \mathbf{1}_n^\top$.

A.6.11. Mostre que a matriz $\mathbf{I}_n - n^{-1}\mathbf{J}_n$ é simétrica, idempotente e não negativa definida (n.n.d.), *i.e.*, $\mathbf{x}^\top (\mathbf{I}_n - \mathbf{J}_n) \mathbf{x} \geq 0$, $\forall \mathbf{x} \neq \mathbf{0}$

A.6.12. Considere as seguintes funções das variáveis aleatórias Y_1, Y_2, Y_3 e Y_4 :

$$\begin{aligned} W_1 &= Y_1 - Y_2 \\ W_2 &= Y_1 + Y_3 \\ W_3 &= Y_1 - Y_4 \end{aligned}$$

- i) Expresse as relações acima em notação matricial.
- ii) Obtenha a esperança do vetor $\mathbf{W} = (W_1, W_2, W_3)^\top$ em termos das esperanças de Y_1, Y_2, Y_3 e Y_4 .
- iii) Obtenha a matriz de covariâncias do vetor \mathbf{W} em termos das variâncias e covariâncias de Y_1, Y_2, Y_3 e Y_4 .

Apêndice B

O método Delta

Em muitas situações, o Teorema Limite Central pode ser empregado para obtenção de distribuições assintóticas de funções de variáveis aleatórias. No entanto, há casos, como aquele que é objeto da Seção C.7, em que as condições de regularidade não permitem sua aplicação. Uma alternativa conveniente é o chamado método Delta. Embora casos multivariados sejam os mais comuns, apresentaremos inicialmente o caso univariado por razões didáticas. Detalhes podem ser obtidos em Sen et al. (2009).

B.1 O caso univariado

Consideremos uma função contínua $f(x) : \mathbb{R} \rightarrow \mathbb{R}$, com derivadas contínuas até a ordem $k > 0$ em torno de um ponto $x_0 \in \mathbb{R}$. Sua expansão de Taylor de ordem k em torno do ponto x_0 é

$$f(x) = f(x_0) + \sum_{j=1}^k \frac{(x - x_0)^j}{j!} f^{(j)}(x_0) + R_k(x, x_0),$$

em que $f^{(j)}(x)$ denota a derivada de ordem j de f calculada no ponto x e

$$R_k(x, x_0) = \frac{(x - x_0)^k}{k!} \{f^{(k)}[hx_0 + (1 - h)x] - f^{(k)}(x_0)\}$$

para algum $0 < h < 1$ é o resto. O caso $k = +\infty$ é conhecido como série de Taylor.

Tomando $k = 1$, quando $x \rightarrow x_0$, podemos desprezar o resto e escrever

$$f(x) \approx f(x_0) + f^{(1)}(x - x_0)$$

em que $f'(x) = f^{(1)}(x)$.

Teorema B.1.1. *Sejam $\sqrt{n}(T_n - \theta)/\sigma \xrightarrow{\mathcal{D}} N(0, 1)$ e g uma função contínua tal que $g'(\theta)$ existe e $g'(\theta) \neq 0$. Então*

$$\sqrt{n}[g(T_n) - g(\theta)]/\sigma g'(\theta) \xrightarrow{\mathcal{D}} N(0, 1).$$

Notemos que por (B.1),

$$\mathbb{E}[g(T_n)] \approx \mathbb{E}[g(\theta)] + \mathbb{E}[g'(\theta)(T_n - \theta)] = g(\theta)$$

e que

$$\mathbb{V}[g(T_n)] \approx \mathbb{V}[g(\theta)] + \mathbb{V}[g'(\theta)(T_n - \theta)] = [g'(\theta)]^2 \mathbb{V}(T_n - \theta) = [g'(\theta)]^2 \sigma^2. \quad (\text{B.1.1})$$

B.2 O caso multivariado

Consideremos agora uma função contínua $f(\mathbf{x}) : \mathbb{R}^p \rightarrow \mathbb{R}$ com derivadas contínuas até a ordem $k \geq 1$. Então para cada $\mathbf{x} \in \mathbb{R}^p$, $\mathbf{x}_0 \in \mathbb{R}^p$, temos a seguinte expansão de Taylor multivariada:

$$f(\mathbf{x}) = f(\mathbf{x}_0) + \sum_{j=1}^k \frac{1}{j!} \sum_{i_1=1}^p \cdots \sum_{i_j=1}^p \frac{\partial^j}{\partial x_{i_1} \cdots \partial x_{i_j}} f(\mathbf{x}) \Big|_{\mathbf{x}_0} \prod_{l=1}^j (x_{i_l} - x_{0_{i_l}}) + R_k(\mathbf{x}, \mathbf{x}_0),$$

em que $R_k(\mathbf{x}, \mathbf{x}_0)$ é

$$\frac{1}{(k+1)!} \sum_{i_1=1}^p \cdots \sum_{i_{k+1}=1}^p \frac{\partial^{k+1}}{\partial x_{i_1} \cdots \partial x_{i_{k+1}}} f[h\mathbf{x}_0 + (1-h)\mathbf{x}] \prod_{l=1}^{k+1} (x_{i_l} - x_{0_{i_l}})$$

para algum $0 < h < 1$. No caso $k = 1$, a expansão de Taylor de primeira ordem pode ser escrita em notação matricial como

$$f(\mathbf{x}) = f(\mathbf{x}_0) + (\mathbf{x} - \mathbf{x}_0)^\top \mathbf{u}(\mathbf{x}_0) + o(\|\mathbf{x} - \mathbf{x}_0\|).$$

em que $\mathbf{u}(\mathbf{x}_0) = \partial f(\mathbf{x})/\partial \mathbf{x}|_{\mathbf{x}=\mathbf{x}_0}$ e $o(\|\mathbf{x} - \mathbf{x}_0\|)$ denota um termo que converge para 0 quando $\mathbf{x} \rightarrow \mathbf{x}_0$. Desprezando esse termo, na vizinhança de \mathbf{x}_0 , a função $f(\mathbf{x})$ pode ser aproximada por

$$f(\mathbf{x}) \approx f(\mathbf{x}_0) + (\mathbf{x} - \mathbf{x}_0)^\top \mathbf{u}(\mathbf{x}_0). \quad (\text{B.2.1})$$

Esta é a base para o seguinte teorema.

Teorema B.2.1. *Seja $\{\mathbf{T}_n\}$ uma sequência de variáveis aleatórias com dimensão p tal que $\sqrt{n}(\mathbf{T}_n - \boldsymbol{\theta}) \xrightarrow{\mathcal{D}} N(\mathbf{0}, \boldsymbol{\Sigma})$ e consideremos uma função $g(\mathbf{T}_n) : \mathbb{R}^p \rightarrow \mathbb{R}$ tal que $\mathbf{u}(\boldsymbol{\theta}) = \partial g(\mathbf{x})/\partial \mathbf{x}$ é não nula e contínua numa vizinhança de $\boldsymbol{\theta}$. Então*

$$\sqrt{n}[g(\mathbf{T}_n) - g(\boldsymbol{\theta})] \xrightarrow{\mathcal{D}} N(0, \gamma^2) \quad \text{com} \quad \gamma^2 = \mathbf{u}(\boldsymbol{\theta})^\top \boldsymbol{\Sigma} \mathbf{u}(\boldsymbol{\theta}).$$

Notemos que por (B.2.1),

$$\mathbb{E}[g(\mathbf{T}_n)] \approx \mathbb{E}[g(\boldsymbol{\theta})] + \mathbb{E}[(\mathbf{T}_n - \boldsymbol{\theta})^\top \mathbf{u}(\boldsymbol{\theta})] = g(\boldsymbol{\theta})$$

e que

$$\mathbb{V}[g(\mathbf{T}_n)] \approx \mathbb{V}[g(\boldsymbol{\theta})] + \mathbb{V}[(\mathbf{T}_n - \boldsymbol{\theta})^\top \mathbf{u}(\boldsymbol{\theta})] = \mathbf{u}(\boldsymbol{\theta})^\top [\mathbb{V}(\mathbf{T}_n - \boldsymbol{\theta})] \mathbf{u}(\boldsymbol{\theta}) = \mathbf{u}(\boldsymbol{\theta})^\top \boldsymbol{\Sigma} \mathbf{u}(\boldsymbol{\theta}). \quad (\text{B.2.2})$$

Apêndice C

Análise de Regressão

C.1 Introdução

O objetivo da análise de regressão é estudar a associação entre uma variável, denominada resposta (ou dependente ou explicada ou ainda endógena, como é geralmente conhecida em textos de econometria), e uma ou mais variáveis explicativas (independentes, preditoras, de controle ou exógenas). Especificamente, esse tipo de análise visa estimar o valor médio (ou predizer o valor) da variável resposta condicionalmente aos valores das variáveis explicativas.

Sejam y_i e $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^\top$ respectivamente os valores de uma variável resposta e de p variáveis explicativas observadas na i -ésima unidade de um conjunto de n unidades amostrais. Esses valores podem ser dispostos na forma de um vetor $\mathbf{y} = (y_1, \dots, y_n)^\top$ e de uma matriz $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)^\top$.¹

Suponhamos que a resposta média da i -ésima unidade amostral condicionalmente às variáveis explicativas \mathbf{x}_i pode ser expressa na forma $\mathbb{E}(y_i|\mathbf{x}_i) = \mu(\mathbf{x}_i, \boldsymbol{\beta})$ em que $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)^\top$ denota um vetor de parâmetros desconhecidos e fixos, denominados **coeficientes de regressão**. A função $\mu(\mathbf{X}, \boldsymbol{\beta})$ é denominada função de regressão ou simplesmente, regressão. A forma de $\mu(\mathbf{X}, \boldsymbol{\beta})$ pode ser proveniente de alguma teoria (como no caso do espaço y_i percorrido por um corpo em queda livre após um tempo x_i , em que essa função seria quadrática) ou sugerida por meio de uma análise descritiva. A função de regressão pode ser:

- i) linear nos parâmetros e nas variáveis explicativas, *e.g.*,

$$\mu(\mathbf{x}_i, \boldsymbol{\beta}) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip};$$

¹Embora seja possível desenvolver toda a metodologia em termos de matrizes \mathbf{X} com posto incompleto, isso não é necessário para efeitos práticos, e neste texto assumiremos que \mathbf{X} tem posto completo, ou seja, que $r(\mathbf{X}) = p$.

ii) linear somente nos parâmetros, *e.g.*,

$$\mu(\mathbf{x}_i, \boldsymbol{\beta}) = \beta_0 + \beta_1 x_{i1}^2 + \beta_2 \log(x_{i2}) + \dots + \beta_p \sqrt{x_{ip}};$$

iii) linearizável, *e.g.*,

$$\mu(\mathbf{x}_i, \boldsymbol{\beta}) = \exp(\beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip})$$

ou

$$\mu(\mathbf{x}_i, \boldsymbol{\beta}) = \beta_0 x_{i1}^{\beta_1} \dots x_{ip}^{\beta_p};$$

iv) não linear, *e.g.*,

$$\mu(\mathbf{x}_i, \boldsymbol{\beta}) = \beta_0 + \exp(\beta_1 x_{i1}^2) + \beta_2 \log(x_{i2}) + \dots + \beta_p \sqrt{x_{ip}}.$$

Nós trataremos somente do caso de funções de regressão lineares (ou linearizáveis) nos parâmetros. Detalhes sobre os modelos de regressão não lineares podem ser encontrados em Seber & Wild (1989) ou Souza (1998), por exemplo.

Condicionalmente aos valores das variáveis explicativas, \mathbf{x}_i , a resposta da i -ésima unidade amostral se situa em torno de $\mu(\mathbf{x}_i, \boldsymbol{\beta})$; expressando o desvio de y_i em relação a $\mu(\mathbf{x}_i, \boldsymbol{\beta})$ como $e_i = y_i - \mu(\mathbf{x}_i, \boldsymbol{\beta})$, o modelo pode ser expresso na forma

$$y_i = \mu(\mathbf{x}_i, \boldsymbol{\beta}) + e_i,$$

$i = 1, \dots, n$, ou na forma matricial, como

$$\mathbf{y} = \boldsymbol{\mu}(\mathbf{X}, \boldsymbol{\beta}) + \mathbf{e}$$

com $\boldsymbol{\mu}(\mathbf{X}, \boldsymbol{\beta}) = [\mu_1(\mathbf{X}, \boldsymbol{\beta}), \dots, \mu_n(\mathbf{X}, \boldsymbol{\beta})]^\top$ e $\mathbf{e} = [e_1, \dots, e_n]^\top$. Quando $\boldsymbol{\mu}(\mathbf{X}, \boldsymbol{\beta}) = \mathbf{X}\boldsymbol{\beta}$, temos

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e} \tag{C.1.1}$$

que é o **modelo de regressão linear múltipla**. Explicitamente esse modelo corresponde a

$$y_i = \mu(\mathbf{x}_i, \boldsymbol{\beta}) + e_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip} + e_i,$$

$i = 1, \dots, n$. Quando $p = 1$, ele é denominado **modelo de regressão linear simples**.

Modelos da forma

$$y_i = \mu(\mathbf{x}_i, \boldsymbol{\beta}) + e_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \dots + \beta_p x_i^p + e_i,$$

$i = 1, \dots, n$, $p \geq 2$ são conhecidos como **modelos de regressão polinomial**.

Usualmente assumimos que

$$\mathbb{E}(\mathbf{e}) = \mathbf{0}, \quad \mathbb{V}(\mathbf{e}) = \sigma^2 \mathbf{I}_n \quad (\text{C.1.2})$$

ou seja, que os erros têm média nula, são homocedásticos e não correlacionados. Nesse contexto, convém lembrar que as variáveis explicativas são consideradas fixas e observadas sem erro. Quando isso não acontece, a análise de regressão ainda pode ser realizada por meio de modelos condicionais. Por exemplo, num estudo em que o objetivo é avaliar a relação entre peso (Y) e altura (X) de adolescentes, podemos obter os valores (possivelmente com erro) de ambas as variáveis numa amostra de n indivíduos, nomeadamente $\{(x_1, y_1), \dots, (x_n, y_n)\}$ e analisar os dados por meio do modelo condicional $y_i|x_i = \alpha + \beta x_i + e_i$, $i = 1, \dots, n$ em que desconsideramos possíveis erros de medida da altura X . A ideia aqui é avaliar a distribuição do peso Y para dados valores da altura X . Alternativamente, é possível considerar os chamados **modelos com erros nas variáveis** ou **modelos com erros de medida** que também acomodam variáveis explicativas observadas com erro. A análise sob esse tipo de modelo foge aos objetivos deste texto e o leitor interessado pode consultar Fuller (1987) ou Chesher (1991) para detalhes.

Exemplo C.1.1: Os dados da Tabela C.1.1 correspondem à idade e a uma medida de pressão arterial sistólica para um conjunto de 20 pacientes de uma determinada clínica.

Tabela C.1.1: Idade (anos) e pressão arterial sistólica (mmHg)

Pressão			Pressão		
Paciente	sistólica	Idade	Paciente	sistólica	Idade
1	114	17	11	156	47
2	134	18	12	159	47
3	116	20	13	142	50
4	139	23	14	156	52
5	110	34	15	164	57
6	150	38	16	185	60
7	152	41	17	162	64
8	138	42	18	176	66
9	142	46	19	175	69
10	145	47	20	180	70

Para estudar a associação entre idade (X) e pressão arterial sistólica (Y), consideramos o modelo de regressão linear simples

$$y_i = \alpha + \beta x_i + e_i,$$

$i = 1, \dots, 20$ que pode ser expresso na forma matricial (C.1.1) com

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{20} \end{bmatrix}, \mathbf{X} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_{20} \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}, \mathbf{e} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_{20} \end{bmatrix}.$$

O parâmetro α corresponde ao valor esperado da pressão arterial sistólica de recém-nascidos e o β representa a variação na pressão arterial sistólica esperada por ano para pacientes com as mesmas características daqueles investigados. Como não há dados para pacientes com idades menores do que 17 anos, a extrapolação do modelo para alguém dessa idade não pode ser baseada em argumentos estatísticos (o modelo poderia, e deve ser, não linear quando consideradas todas as idades). Consequentemente, para efeito de interpretação, convém reescrever o modelo na forma

$$y_i = \alpha^* + \beta(x_i - x_0) + e_i,$$

$i = 1, \dots, 20$ em que x_0 é uma constante conhecida, por exemplo $x_0 = \bar{x}$ com \bar{x} denotando a idade média dos pacientes. Nesse caso, o parâmetro α^* corresponde ao valor esperado da pressão arterial sistólica para pacientes cuja idade é \bar{x} . Se escolhermos $x_0 = 17$, o parâmetro seria interpretado como valor esperado da pressão arterial sistólica para pacientes cuja idade é 17 anos. Nesses casos, a matriz de especificação do modelo seria dada por

$$\mathbf{X} = \begin{bmatrix} 1 & x_1 - x_0 \\ 1 & x_2 - x_0 \\ \vdots & \vdots \\ 1 & x_{20} - x_0 \end{bmatrix}.$$

Tendo em conta que esse modelo é equivalente a

$$y_i = \alpha + \beta x_i + e_i,$$

$i = 1, \dots, 20$ com $\alpha = \alpha^* - \beta x_0$, fica evidente que a alteração corresponde apenas a uma translação do eixo das ordenadas e que nem o valor do parâmetro β nem sua interpretação são afetados.

Exemplo C.1.2: A Tabela C.1.2 contém dados de capacidade instalada (ton), potência instalada (1000 kW) e área construída (100 m^2) de 10 empresas de uma certa indústria.

Tabela C.1.2: Capacidade instalada (ton), potência instalada (1000 kW) e área construída (100 m²) de empresas de uma certa indústria

Produção	4.5	5.9	4.2	5.2	8.1	9.7	10.7	11.9	12.0	12.3
Potência	0.9	2.5	1.3	1.8	3.1	4.6	6.1	6.0	5.9	6.1
Área	7.1	10.4	7.2	8.2	8.5	11.9	12.1	12.5	12.0	11.3

Para representar a distribuição da capacidade instalada (Y) a partir das informações sobre potência instalada (X_1) e área construída (X_2), consideramos o seguinte modelo:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + e_i,$$

$i = 1, \dots, 10$, que genericamente, *i.e.*, para n empresas, pode ser expresso na forma matricial (C.1.1) com

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{21} \\ 1 & x_{12} & x_{22} \\ \vdots & \vdots & \vdots \\ 1 & x_{1n} & x_{2n} \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}, \mathbf{e} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}.$$

Aqui, o parâmetro β_1 representa a variação na capacidade instalada esperada por unidade de potência instalada para empresas com a mesma área construída. O parâmetro β_2 tem interpretação semelhante com a substituição de potência instalada por área construída e vice-versa. Como não temos dados para empresas com potência instalada menor que $0.9 \times 1000kW$ e área construída menor que $7.1 \times 100m^2$, o parâmetro β_0 deve ser interpretado como um fator de ajuste do plano que aproxima a verdadeira função que relaciona o valor esperado da variável resposta com as variáveis explicativas na região em que há dados disponíveis. Se essa função for linear em todo o campo de variação das variáveis explicativas, um modelo linear mais adequado não deveria conter o termo β_0 , pois potência instalada ou área construída iguais a zero implicariam capacidade de produção nula.

Exemplo C.1.3: A Tabela C.1.3 contém dados de um estudo cujo objetivo é avaliar o efeito de diferentes níveis de um determinado fertilizante no número de frutos de boa qualidade por macieira de uma certa variedade.

Tabela C.1.3: Concentração do fertilizante e número de frutos de boa qualidade

Concentração do fertilizante	Número de frutos de boa qualidade	Concentração do fertilizante	Número de frutos de boa qualidade
10	16	40	24
10	18	40	25
10	19	40	28
10	20	50	26
20	18	50	28
20	20	50	30
20	21	50	32
20	22	50	35
20	23	50	25
30	22		
30	25		
30	26		

Se encararmos a variável Concentração de fertilizante como um fator (no espírito dos modelos usuais de ANOVA), o estudo pode ser modelado (de forma genérica) por intermédio de

$$y_{ij} = \mu_i + e_{ij}, \quad (\text{C.1.3})$$

$i = 1, \dots, I$, $j = 1, \dots, n_i$, em que y_{ij} representa o número de frutos de boa qualidade na j -ésima macieira tratada com a i -ésima concentração do fertilizante e μ_i o valor esperado correspondente.² Esse modelo pode ser escrito na forma matricial (C.1.1) com

$$\mathbf{y} = \begin{bmatrix} y_{11} \\ \vdots \\ y_{1n_1} \\ y_{21} \\ \vdots \\ y_{2n_2} \\ \vdots \\ y_{I1} \\ \vdots \\ y_{In_I} \end{bmatrix}, \mathbf{X} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ \vdots & \vdots & \dots & 0 \\ 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & 1 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_I \end{bmatrix}$$

² Aqui expressamos o modelo sob a parametrização de médias de celas. Outras parametrizações podem ser consideradas; o leitor encontrará mais detalhes na Seção C.6.

e $\mathbf{e} = [e_{11}, \dots, e_{1n_1}, e_{21}, \dots, e_{2n_2}, \dots, e_{k1}, \dots, e_{kn_k}]^\top$. Embora esse seja um modelo linear (nos parâmetros), ele não impõe uma relação linear entre o valor esperado do número de frutos de boa qualidade (μ_i) e a concentração do fertilizante (x_{ij}), ou seja, permite acomodar um efeito não linear desse fator (no número esperado de frutos de boa qualidade). Se considerarmos a natureza quantitativa dos níveis do fator Concentração de fertilizante e quisermos adotar um modelo linear (expresso na forma genérica) que inclua um efeito linear desse fator, podemos utilizar

$$y_{ij} = \beta_0 + \beta_1 x_i + e_{ij}, \quad (\text{C.1.4})$$

$i = 1, \dots, I$, $j = 1, \dots, n_i$ que pode ser expresso na forma matricial (C.1.1) com

$$\mathbf{y} = \begin{bmatrix} y_{11} \\ \vdots \\ y_{1n_1} \\ y_{21} \\ \vdots \\ y_{2n_2} \\ \vdots \\ y_{I1} \\ \vdots \\ y_{In_I} \end{bmatrix}, \mathbf{X} = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_I \\ \vdots & \vdots \\ 1 & x_I \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \text{ e } \mathbf{e} = \begin{bmatrix} e_{11} \\ \vdots \\ e_{1n_1} \\ e_{21} \\ \vdots \\ e_{2n_2} \\ \vdots \\ e_{I1} \\ \vdots \\ e_{In_I} \end{bmatrix}.$$

Exemplo C.1.4: As ondas epidêmicas estacionais de gripe constituem uma carga considerável para os serviços de saúde e caracterizam-se por sua grande variabilidade de ano a ano. O uso dos dados históricos pode servir para estabelecer um modelo preditor do número de médicos indispensáveis para atender a totalidade das consultas de forma eficiente. Os dados da Tabela C.1.4 correspondem aos números de consultas (acumuladas semanalmente) da temporada invernal realizadas nos anos de 2000 a 2004 numa certa localidade.

Tabela C.1.4: Consultas acumuladas na temporada invernal de 2000 a 2004

Semana	2000	2001	2002	2003	2004
1	23470	46041	18284	20868	18726
2	26101	86018	23800	24574	21674
3	29178	105393	30764	28399	25771
4	32460	124280	40738	31496	29270
5	34949	138060	59248	34459	32563
6	37698	151779	93457	37703	40797
7	41216	162918	132469	42823	56698
8	44661	169642	162121	46930	72239
9	48030	174107	179232	50527	87008
10	55254	175520	185623	64897	84654

Supondo que para cada ano, a relação entre o tempo (em semanas) e a distribuição do número de consultas possa ser representado por intermédio de regressões lineares simples, poderíamos considerar um modelo do tipo

$$y_{ij} = \beta_{0i} + \beta_{1i}x_{ij} + e_{ij}, \quad (\text{C.1.5})$$

$i = 1, \dots, I$, $j = 1, \dots, n_i$, em que β_{0i} representa o número esperado de consultas no início do inverno do ano i ($i = 1$ correspondendo ao ano 2000) e β_{1i} denota a variação no número esperado de consultas por semana para o ano i . Neste caso particular, $I = 5$ e $n_i = 10$. Na forma matricial (C.1.1), esse modelo pode ser expresso como

$$\mathbf{y} = \begin{bmatrix} y_{11} \\ \vdots \\ y_{1n_1} \\ y_{21} \\ \vdots \\ y_{2n_2} \\ \vdots \\ y_{I1} \\ \vdots \\ y_{In_I} \end{bmatrix}, \mathbf{X} = \begin{bmatrix} 1 & 0 & \dots & 0 & x_{11} & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & \dots & 0 & x_{1n_1} & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 & 0 & x_{21} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 1 & \dots & 0 & 0 & x_{2n_2} & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 & 0 & \dots & x_{I1} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 & 0 & 0 & \dots & x_{In_I} \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \beta_{01} \\ \beta_{02} \\ \vdots \\ \beta_{0I} \\ \beta_{11} \\ \beta_{12} \\ \vdots \\ \beta_{1I} \end{bmatrix}$$

e

$$\mathbf{e} = [e_{11}, \dots, e_{1n_1}, e_{21}, \dots, e_{2n_2}, \dots, e_{I1}, \dots, e_{In_I}]^\top.$$

Se considerarmos que existe um número fixo de consultas de pacientes gripados no início do inverno (semana 1), e que a variação desse número de gripados depende

das condições climáticas de cada ano, o modelo poderia ser escrito como

$$y_{ij} = \beta_0 + \beta_{1i}x_{ij} + e_{ij}, \quad (\text{C.1.6})$$

$i = 1, \dots, I$, $j = 1, \dots, n_i$ ou na forma matricial (C.1.1), com

$$\mathbf{y} = \begin{bmatrix} y_{11} \\ \vdots \\ y_{1n_1} \\ y_{21} \\ \vdots \\ y_{2n_2} \\ \vdots \\ y_{I1} \\ \vdots \\ y_{In_I} \end{bmatrix}, \mathbf{X} = \begin{bmatrix} 1 & x_{11} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n_1} & 0 & \dots & 0 \\ 1 & 0 & x_{21} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & x_{2n_2} & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & \dots & x_{I1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \dots & x_{In_I} \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_{11} \\ \beta_{12} \\ \vdots \\ \beta_{1I} \end{bmatrix}$$

e

$$\mathbf{e} = [e_{11}, \dots, e_{1n_1}, e_{21}, \dots, e_{2n_2}, \dots, e_{I1}, \dots, e_{In_I}]^\top.$$

Em ambas as situações descritas acima, podemos dizer que existe **interação** entre o tempo decorrido desde o início do inverno e o ano (possivelmente por causa de diferenças climáticas), pois as taxas de variação no número esperado de consultas (β_{1i}) dependem do ano em que foram colhidos os dados (as retas que representam a variação no número esperado de consultas ao longo das 10 semanas de observação não são paralelas). Num modelo sem interação, essas taxas são iguais para todos os anos, ou seja, $\beta_{1i} = \beta_1$. Nesse caso, admitindo efeito de ano, o modelo poderia ser escrito como

$$y_{ij} = \beta_{0i} + \beta_1 x_{ij} + e_{ij},$$

$i = 1, \dots, I$, $j = 1, \dots, n_i$ e na forma matricial (C.1.1), com

$$\mathbf{y} = \begin{bmatrix} y_{11} \\ \vdots \\ y_{1n_1} \\ y_{21} \\ \vdots \\ y_{2n_2} \\ \vdots \\ y_{I1} \\ \vdots \\ y_{In_I} \end{bmatrix}, \mathbf{X} = \begin{bmatrix} 1 & 0 & \dots & 0 & x_{11} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & 0 & \dots & 0 & x_{1n_1} \\ 0 & 1 & \dots & 0 & x_{21} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 1 & \dots & 0 & x_{2n_2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & x_{I1} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & x_{In_I} \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \beta_{01} \\ \beta_{02} \\ \vdots \\ \beta_{0I} \\ \beta_1 \end{bmatrix}$$

e

$$\mathbf{e} = [e_{11}, \dots, e_{1n_1}, e_{21}, \dots, e_{2n_2}, \dots, e_{I1}, \dots, e_{In_I}]^\top.$$

Exemplo C.1.5: Em muitas situações, o modelo de **regressão segmentada** proporciona maior flexibilidade para a caracterização do comportamento da resposta. Esse modelo se caracteriza pela existência de um ponto x_0 (conhecido) em que a taxa de variação da resposta média se altera. Para justificar o modelo, suponhamos que

$$y_i = \begin{cases} \beta_0 + \beta_1 x_i + e_i, & x_i \leq x_0, \quad i = 1, \dots, j \\ \beta_3 + \beta_2 x_i + e_i, & x_i \geq x_0, \quad i = j+1, \dots, n. \end{cases}$$

Admitindo que as duas retas têm em comum o ponto (x_0, y_0) , temos $\beta_0 + \beta_1 x_0 = \beta_3 + \beta_2 x_0$, ou seja $\beta_3 = \beta_0 + \beta_1 x_0 - \beta_2 x_0$. Substituindo essa expressão de β_3 no modelo original, obtém-se o seguinte modelo com três parâmetros $(\beta_0, \beta_1, \beta_2)$:

$$y_i = \begin{cases} \beta_0 + \beta_1 x_i + e_i, & x_i \leq x_0, \quad i = 1, \dots, j \\ \beta_0 + \beta_1 x_0 + \beta_2 (x_i - x_0) + e_i, & x_i \geq x_0, \quad i = j+1, \dots, n. \end{cases}$$

que pode ser expresso na forma matricial (C.1.1) com

$$\mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_j \\ y_{j+1} \\ \vdots \\ y_n \end{bmatrix}, \mathbf{X} = \begin{bmatrix} 1 & x_1 & 0 \\ \vdots & \vdots & \vdots \\ 1 & x_j & 0 \\ 1 & x_0 & (x_{j+1} - x_0) \\ \vdots & \vdots & \vdots \\ 1 & x_0 & (x_n - x_0) \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} \text{ e } \mathbf{e} = \begin{bmatrix} e_1 \\ \vdots \\ e_j \\ e_{j+1} \\ \vdots \\ e_n \end{bmatrix}.$$

Aqui, o coeficiente β_0 corresponde ao valor esperado da resposta Y quando o valor da variável explicativa X é zero, o coeficiente β_1 pode ser interpretado como a variação esperada da resposta Y por unidade de variação da variável explicativa X para valores $X \leq x_0$ e o coeficiente β_2 corresponde à variação esperada da resposta Y por unidade de variação da variável explicativa X para valores $X \geq x_0$.

Sob a formulação matricial (C.1.1), hipóteses lineares de interesse podem ser expressas na forma

$$H : \mathbf{C}\boldsymbol{\beta} = \mathbf{m} \quad (\text{C.1.7})$$

em que \mathbf{C} é uma matriz $(c \times p)$ de constantes conhecidas com $r(\mathbf{C}) = c$ e \mathbf{m} é um vetor $(c \times 1)$ de constantes conhecidas. Em geral, $\mathbf{m} = \mathbf{0}$.

No Exemplo C.1.1, a hipótese de que o intercepto é nulo é expressa como (C.1.7) com $\mathbf{C} = [1 \ 0]$ e $\mathbf{m} = 0$ e a hipótese de que o coeficiente angular é nulo, com $\mathbf{C} = [0 \ 1]$ e $\mathbf{m} = 0$. Além disso, supondo que, sob a parametrização alternativa lá

mencionada, $x_0 = 17$, a hipótese de que o valor esperado para a pressão sistólica de pacientes com idade igual a 30 anos é de 120 mmHg seria expressa na forma (C.1.7) com $\mathbf{C} = [1 \ 13]$ e $\mathbf{m} = 120$. Em todos esses casos, a matriz \mathbf{C} tem apenas uma linha, e consequentemente, $c = 1$.

No Exemplo C.1.2, uma hipótese de interesse é a de que nenhuma das duas variáveis é significativamente linearmente associada ao valor esperado da variável resposta. Esta hipótese pode ser expressa na forma (C.1.7) com

$$\mathbf{C} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

No Exemplo C.1.3, sob o modelo (C.1.3), a hipótese de que os valores esperados de frutos de boa qualidade por macieira sejam iguais, independentemente da concentração de fertilizante utilizada pode ser expressa na forma (C.1.7), em que $\mathbf{m} = \mathbf{0}$ é um vetor $[(I - 1) \times 1]$ e \mathbf{C} é uma matriz $[(I - 1) \times I]$ construída de forma a gerar $I - 1$ contrastes linearmente independentes dos elementos de $\boldsymbol{\beta} = [\mu_1 \ \mu_2 \ \dots \ \mu_I]^\top$. Como exemplo desse tipo de matriz, temos

$$\mathbf{C} = \begin{bmatrix} 1 & 0 & \dots & 0 & -1 \\ 0 & 1 & \dots & 0 & -1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & -1 \end{bmatrix}.$$

Sob o modelo (C.1.4), uma hipótese equivalente seria expressa com $\mathbf{C} = [0 \ 1]$ e $\mathbf{m} = 0$.

No Exemplo C.1.4, a hipótese a ser testada para avaliar se o modelo (C.1.5) pode ser reduzido ao modelo (C.1.6) também pode ser expressa na forma (C.1.7) em que \mathbf{C} é uma matriz $[(I - 1) \times 2I]$ com $r(\mathbf{C}) = I - 1$ dada por

$$\mathbf{C} = \begin{bmatrix} 1 & 0 & \dots & -1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & -1 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & 1 & -1 & 0 & \dots & 0 & 0 \end{bmatrix},$$

por exemplo, e $\mathbf{m} = \mathbf{0}$, um vetor de dimensão $[(I - 1) \times 1]$.

No Exemplo C.1.5, a hipótese de que os coeficientes angulares correspondentes a valores da variável explicativa menores ou maiores do que x_0 são iguais pode ser expressa na forma (C.1.7) com $\mathbf{C} = [0 \ 1 \ -1]$.

C.2 Método de mínimos quadrados

O método de mínimos quadrados (ordinários) pode ser utilizado para estimar o vetor de parâmetros β no modelo (C.1.1)-(C.1.2). A base do método é a minimização da forma quadrática

$$Q(\beta) = \mathbf{e}^\top \mathbf{e} = (\mathbf{y} - \mathbf{X}\beta)^\top (\mathbf{y} - \mathbf{X}\beta) = \sum_{i=1}^n e_i^2. \quad (\text{C.2.1})$$

Com essa finalidade, consideremos

$$Q(\beta) = \mathbf{y}^\top \mathbf{y} + \beta^\top \mathbf{X}^\top \mathbf{X} \beta - 2\beta^\top \mathbf{X}^\top \mathbf{y}$$

e utilizemos as expressões para derivadas matriciais apresentadas no Apêndice A para obter

$$\begin{aligned} \frac{\partial Q(\beta)}{\partial \beta} &= 2\mathbf{X}^\top \mathbf{X} \beta - 2\mathbf{X}^\top \mathbf{y}, \\ \frac{\partial^2 Q(\beta)}{\partial \beta \partial \beta^\top} &= 2\mathbf{X}^\top \mathbf{X}. \end{aligned}$$

Igualando a primeira derivada parcial a zero, obtemos o **sistema de equações de estimação** (também conhecido por **sistema de equações normais**) é ³

$$\mathbf{X}^\top \mathbf{X} \hat{\beta} = \mathbf{X}^\top \mathbf{y}.$$

Como assumimos que \mathbf{X} tem posto completo, $\mathbf{X}^\top \mathbf{X}$ é uma matriz não singular e a solução do sistema de equações normais

$$\hat{\beta} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}. \quad (\text{C.2.2})$$

Como a matriz $\mathbf{X}^\top \mathbf{X}$ é definida positiva, o estimador (C.2.2) corresponde ao ponto de mínimo de (C.2.1).

Sob a suposição de que $\mathbb{E}(\mathbf{e}) = \mathbf{0}$ e $\mathbb{V}(\mathbf{e}) = \sigma^2 \mathbf{I}_n$, temos

- i) $\mathbb{E}(\hat{\beta}) = \beta,$
- ii) $\mathbb{V}(\hat{\beta}) = \sigma^2 (\mathbf{X}^\top \mathbf{X})^{-1}.$

³ Aqui, o termo “normal” não se refere à distribuição normal e sim ao conceito de ortogonalidade. A razão para isso é que a teoria de mínimos quadrados pode ser desenvolvida por meio de projeções ortogonais.

Quando a matriz \mathbf{X} não tem posto completo, o sistema de equações de estimação tem infinitas soluções dadas por

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^- \mathbf{X}^\top \mathbf{y} + [\mathbf{I} - (\mathbf{X}^\top \mathbf{X})^- \mathbf{X}^\top \mathbf{X}] \mathbf{g},$$

em que $(\mathbf{X}^\top \mathbf{X})^-$ é uma inversa generalizada de $\mathbf{X}^\top \mathbf{X}$ e \mathbf{g} é um vetor arbitrário. Pode-se mostrar que $\mathbf{X}\hat{\boldsymbol{\beta}}$ é invariante com relação ao vetor \mathbf{g} e $Q(\boldsymbol{\beta})$ obtém seu ponto mínimo para qualquer solução de $\mathbf{X}^\top \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}^\top \mathbf{y}$. Nesse caso, $\hat{\boldsymbol{\beta}}$ deve ser encarado como uma solução das equações normais e não como um estimador de $\boldsymbol{\beta}$, que é dito **não estimável**.⁴

Há casos em que a matriz \mathbf{X} tem posto completo, mas suas colunas estão muito próximas de serem linearmente dependentes; isso ocorre, por exemplo, em alguns modelos de regressão polinomial. Como consequência, a matriz $\mathbf{X}^\top \mathbf{X}$ é “quase” não singular, com autovalores próximos de zero de forma a gerar estimadores $\hat{\boldsymbol{\beta}}$ muito imprecisos. A presença de um certo nível de dependência entre as colunas de \mathbf{X} é conhecida na literatura como **multicolinearidade**. Procedimentos para lidar com esse problema serão discutidos na Seção C.5.

O estimador de mínimos quadrados de $\boldsymbol{\beta}$ é o estimador linear não enviesado de variância mínima, (*best linear unbiased estimator* - BLUE) como se pode deduzir por meio do seguinte teorema:

Teorema C.2.1. Teorema de Gauss-Markov: *Considere o modelo linear $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$, com as hipóteses $\mathbb{E}(\mathbf{e}) = \mathbf{0}$ e $\mathbb{V}(\mathbf{e}) = \sigma^2 \mathbf{I}_n$. Seja $\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}$ o estimador de mínimos quadrados de $\boldsymbol{\beta}$. Se $\tilde{\boldsymbol{\beta}}$ é um outro estimador de $\boldsymbol{\beta}$ tal que $\tilde{\boldsymbol{\beta}} = \mathbf{C}\mathbf{y}$, em que \mathbf{C} é uma matriz $(p \times n)$, e $\mathbb{E}(\tilde{\boldsymbol{\beta}}) = \boldsymbol{\beta}$, então*

$$\mathbf{r}^\top \mathbb{V}(\tilde{\boldsymbol{\beta}}) \mathbf{r} \geq \mathbf{r}^\top \mathbb{V}(\hat{\boldsymbol{\beta}}) \mathbf{r}, \quad \forall \mathbf{r} \in \mathbb{R}^p.$$

Sob a suposição adicional de que os erros e_i têm distribuição Normal, a linearidade do estimador (C.2.2) permite-nos concluir que

$$\hat{\boldsymbol{\beta}} \sim N_p[\boldsymbol{\beta}, \sigma^2 (\mathbf{X}^\top \mathbf{X})^{-1}]. \quad (\text{C.2.3})$$

Em geral, a variância σ^2 é desconhecida e inferências sobre $\boldsymbol{\beta}$ dependem de um estimador desse parâmetro; um candidato é o estimador não enviesado

$$\begin{aligned} s^2 &= (n-p)^{-1} (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^\top (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \\ &= (n-p)^{-1} \mathbf{y}^\top [\mathbf{I}_n - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top] \mathbf{y}. \end{aligned}$$

⁴Uma função $\mathbf{a}^\top \boldsymbol{\beta}$ é estimável se tiver um estimador linear não enviesado, digamos $\mathbf{b}^\top \mathbf{y}$. Em muitos casos, a matriz \mathbf{X} define modelos não identificáveis cujos parâmetros são não estimáveis embora haja funções desses parâmetros que são estimáveis. Para detalhes, veja a Seção C.6 e Freedman (2005).

Como a matriz $\mathbf{I}_n - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top$ é idempotente pode-se recorrer ao Teorema A.3.3 para mostrar que $(n - p)s^2/\sigma^2 \sim \chi_{n-p}^2$.

Dados uma matriz \mathbf{C} de constantes conhecidas com dimensão $(c \times p)$ e $r(\mathbf{C}) = c$ e um vetor \mathbf{m} de constantes conhecidas com dimensão $(c \times 1)$ consideremos a forma quadrática

$$\begin{aligned} Q_{\mathbf{C}} &= (\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{m})^\top [\mathbf{C}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{C}^\top]^{-1} (\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{m}) \\ &= [\mathbf{C}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y} - \mathbf{m}]^\top [\mathbf{C}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{C}^\top]^{-1} [\mathbf{C}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y} - \mathbf{m}]. \end{aligned}$$

Tendo (C.2.3) em conta, podemos novamente recorrer ao Teorema A.3.3 para mostrar que $\sigma^{-2} Q_{\mathbf{C}} \sim \chi_c^2(\delta)$ com parâmetro de não centralidade

$$\delta = \frac{1}{2} (\mathbf{C}\boldsymbol{\beta} - \mathbf{m})^\top [\mathbf{C}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{C}^\top]^{-1} (\mathbf{C}\boldsymbol{\beta} - \mathbf{m}).$$

Como $\mathbf{C}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y} - \mathbf{m} = \mathbf{C}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top [\mathbf{y} - \mathbf{X}\mathbf{C}(\mathbf{C}^\top \mathbf{C})^{-1} \mathbf{m}]$, é possível expressar $Q_{\mathbf{C}}$ como uma forma quadrática na variável $\mathbf{y} - \mathbf{X}\mathbf{C}(\mathbf{C}^\top \mathbf{C})^{-1} \mathbf{m}$, ou seja,

$$Q_{\mathbf{C}} = [\mathbf{y} - \mathbf{X}\mathbf{C}(\mathbf{C}^\top \mathbf{C})^{-1} \mathbf{m}]^\top \mathbf{A} [\mathbf{y} - \mathbf{X}\mathbf{C}(\mathbf{C}^\top \mathbf{C})^{-1} \mathbf{m}]$$

com

$$\mathbf{A} = \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{C}^\top [\mathbf{C}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{C}^\top]^{-1} \mathbf{C}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top.$$

Também é possível escrever

$$\begin{aligned} (n - p)s^2 &= \mathbf{y}^\top [\mathbf{I}_n - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top] \mathbf{y} \\ &= [\mathbf{y} - \mathbf{X}\mathbf{C}(\mathbf{C}^\top \mathbf{C})^{-1} \mathbf{m}]^\top \mathbf{B} [\mathbf{y} - \mathbf{X}\mathbf{C}(\mathbf{C}^\top \mathbf{C})^{-1} \mathbf{m}] \end{aligned}$$

com $\mathbf{B} = \mathbf{I}_n - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top$. Observando que $\mathbf{A}\mathbf{B} = \mathbf{0}$, podemos utilizar o Teorema A.3.5 para mostrar que

$$F = c^{-1} s^{-2} (\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{m})^\top [\mathbf{C}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{C}^\top]^{-1} (\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{m}) \sim F_{(c, n-p)}(\delta), \quad (\text{C.2.4})$$

com parâmetro de não centralidade

$$\delta = \frac{1}{2} (\mathbf{C}\boldsymbol{\beta} - \mathbf{m})^\top [\mathbf{C}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{C}^\top]^{-1} (\mathbf{C}\boldsymbol{\beta} - \mathbf{m}).$$

Esse resultado nos dá todos os ingredientes necessários tanto para construir intervalos (ou regiões) de confiança para combinações lineares dos elementos de $\boldsymbol{\beta}$ quanto para testar hipóteses lineares de interesse. Por exemplo, para testar hipóteses na forma (C.1.7), basta notar que sob a hipótese nula $\delta = 0$ e a estatística (C.2.4) tem distribuição $F_{(c, n-p)}$.

Sob o modelo de regressão linear simples, $y_i = \alpha + \beta x_i + e_i$, $i = 1, \dots, n$ com as suposições mencionadas acima (incluindo a de normalidade de e_i), um teste da hipótese de que o coeficiente angular é igual a uma constante β_0 pode ser realizado por meio de (C.2.4) com $\mathbf{C} = [0 \ 1]$ e $\mathbf{m} = \beta_0$. Nesse caso, a estatística (C.2.4) pode ser expressa como

$$(\hat{\beta} - \beta_0)^2 / \{s^2 [(\mathbf{X}^\top \mathbf{X})^{-1}]_{22}\}$$

em que $[(\mathbf{X}^\top \mathbf{X})^{-1}]_{22}$ denota o elemento (2, 2) da matriz $(\mathbf{X}^\top \mathbf{X})^{-1}$ e segue uma distribuição $F_{(1, n-p)}$ quando a hipótese $H : \beta = \beta_0$ é verdadeira. A raiz quadrada dessa estatística munida de sinal $(\hat{\beta} - \beta_0)^5$, nomeadamente

$$[\text{sinal}(\hat{\beta} - \beta_0)](\hat{\beta} - \beta_0) / \{s \sqrt{[(\mathbf{X}^\top \mathbf{X})^{-1}]_{22}}\}$$

segue uma distribuição t_{n-p} quando a hipótese $H : \beta = \beta_0$ é verdadeira. O resultado para o caso particular em que $\beta_0 = 0$ é obtido de forma automática quando se utilizam os pacotes computacionais mais comuns.

Nos casos de dúvidas quanto à validade da suposição de normalidade, podemos recorrer ao Teorema Limite Central para efeito de inferência sobre os parâmetros do modelo. Para salientar que o resultado depende do tamanho da amostra, adicionamos o índice n aos componentes do modelo. Consideremos agora, uma sequência de elementos $\{\mathbf{y}_n, \mathbf{X}_n, \mathbf{e}_n\}$ com $\mathbf{X}_n = [\mathbf{x}_{n1}, \mathbf{x}_{n2}, \dots, \mathbf{x}_{nn}]^\top$ satisfazendo o modelo $\mathbf{y}_n = \mathbf{X}_n \boldsymbol{\beta} + \mathbf{e}_n$ sob as suposições do Teorema de Gauss-Markov e admitamos que

- a) $\max_{1 \leq i \leq n} \mathbf{x}_{ni}^\top (\mathbf{X}_n^\top \mathbf{X}_n)^{-1} \mathbf{x}_{ni} \rightarrow 0$ com $n \rightarrow \infty$,⁶
- b) $\lim_{n \rightarrow \infty} \mathbf{X}_n^\top \mathbf{X}_n = \mathbf{V}$, com \mathbf{V} finita e definida positiva.

Então, utilizando o Teorema Limite Central de Hájek-Šidak (ver Sen et al. (2009), por exemplo), podemos mostrar que

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) \xrightarrow{\mathcal{D}} N_p(\mathbf{0}, \sigma^2 \mathbf{V}^{-1})$$

ou equivalentemente, que

$$(\mathbf{X}_n^\top \mathbf{X}_n)^{1/2}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) \xrightarrow{\mathcal{D}} N_p(\mathbf{0}, \sigma^2 \mathbf{I}_p).$$

⁵ $\text{sinal}(a) = 1$ se $a > 0$, $\text{sinal}(a) = -1$ se $a < 0$ e $\text{sinal}(a) = 0$ se $a = 0$.

⁶ Esta suposição, conhecida como **condição de Noether** é válida na maioria dos casos de interesse prático. Em particular, no caso de problemas envolvendo a comparação de I médias, ela corresponde à exigência de que os tamanhos n_i das amostras colhidas de cada subpopulação investigada sejam tais que $n_i / \sum_{i=1}^I n_i$ convirja para uma constante λ_i com $n = \sum_{i=1}^I n_i \rightarrow \infty$.

Em termos práticos, esse resultado significa que para amostras com tamanhos suficientemente “grandes”,

$$\hat{\boldsymbol{\beta}} \approx N_p[\boldsymbol{\beta}, \sigma^2(\mathbf{X}^\top \mathbf{X})^{-1}]$$

ou seja, que a distribuição do estimador $\hat{\boldsymbol{\beta}}$ pode ser aproximada por uma distribuição normal com média $\boldsymbol{\beta}$ e matriz de covariâncias $\sigma^2(\mathbf{X}^\top \mathbf{X})^{-1}$.

Neste caso, para testar hipóteses da forma (C.1.7) podemos empregar a **estatística de Wald**

$$Q_W = s^{-2}(\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{m})^\top [\mathbf{C}(\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{C}^\top]^{-1}(\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{m}), \quad (\text{C.2.5})$$

cuja distribuição sob a hipótese nula pode ser aproximada por uma distribuição χ_c^2 .

Existem situações em que a suposição de **homocedasticidade** *i.e.*, variâncias constantes, não é válida. Como ilustração, tomemos o Exemplo C.1.1, em que um aumento da variância da pressão arterial sistólica com a idade não seria um fato inesperado. Por exemplo, poderíamos supor que $\mathbb{V}(e_i) = x_i \sigma^2$, ou seja, que $\mathbb{V}(\mathbf{e}) = \text{diag}\{x_1, \dots, x_{20}\} \sigma^2$. Nesse caso, o modelo é dito **heterocedástico**. De uma forma mais geral, podemos ter modelos em que a matriz de covariâncias de \mathbf{e} é uma matriz simétrica definida positiva, \mathbf{V} . Nesses casos, o ajuste do modelo pode ser concretizado por meio do método dos **mínimos quadrados generalizados** (MQG) que consiste em minimizar

$$Q(\boldsymbol{\beta}) = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top \mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}).$$

Quando a matriz \mathbf{V} é uma matriz diagonal com elementos $w_i \sigma^2$ ao longo da diagonal principal, o método é chamado de **mínimos quadrados ponderados** e a função a ser minimizada pode ser expressa como

$$Q(\boldsymbol{\beta}) = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top \mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = \frac{1}{\sigma^2} \sum_{i=1}^n w_i^{-1} (y_i - \mathbf{x}_i^\top \boldsymbol{\beta})^2.$$

Admitindo que \mathbf{V} é conhecida e considerando um procedimento similar àquele usado no caso de mínimos quadrados ordinários, podemos mostrar que o estimador de mínimos quadrados generalizados de $\boldsymbol{\beta}$ é

$$\tilde{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{V}^{-1} \mathbf{y}.$$

Se $\mathbf{e} \sim N(\mathbf{0}, \mathbf{V})$ com \mathbf{V} conhecida, então podemos demonstrar que

$$\tilde{\boldsymbol{\beta}} \sim N_p[\boldsymbol{\beta}, (\mathbf{X}^\top \mathbf{V}^{-1} \mathbf{X})^{-1}].$$

Quando há dúvida com relação à suposição de normalidade, mas $\mathbb{V}(\mathbf{e}) = \mathbf{V}$, com \mathbf{V} finita e conhecida, o Teorema Limite Central permite-nos concluir que para n suficientemente grande,

$$\tilde{\boldsymbol{\beta}} \approx N_p[\boldsymbol{\beta}, (\mathbf{X}^\top \mathbf{V}^{-1} \mathbf{X})^{-1}].$$

Como, em geral, \mathbf{V} não é conhecida, podemos substituí-la por um estimador consistente $\hat{\mathbf{V}}$ e considerar o estimador

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \hat{\mathbf{V}}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \hat{\mathbf{V}}^{-1} \mathbf{y}. \quad (\text{C.2.6})$$

O estimador consistente de \mathbf{V} pode ser obtido de fontes externas, *e.g.*, outros estudos com características similares; em muitos casos, é possível utilizar a matriz de covariâncias dos resíduos obtidos do ajuste de um modelo homocedástico com os mesmos parâmetros $\boldsymbol{\beta}$. Apelando para o Teorema Limite Central e para o Teorema de Sverdrup (ver Sen et al. (2009)), por exemplo, podemos mostrar que para n suficientemente grande

$$\hat{\boldsymbol{\beta}} \approx N_p[\boldsymbol{\beta}, (\mathbf{X}^\top \mathbf{V}^{-1} \mathbf{X})^{-1}]. \quad (\text{C.2.7})$$

Intervalos de confiança aproximados para combinações lineares da forma $\mathbf{C}\boldsymbol{\beta}$ ou testes para hipóteses do tipo (C.1.7) podem ser concretizados utilizando (C.2.7) ou a estatística de Wald

$$Q_W = (\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{m})^\top [\mathbf{C}(\mathbf{X}^\top \hat{\mathbf{V}}^{-1} \mathbf{X})^{-1} \mathbf{C}^\top]^{-1} (\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{m}), \quad (\text{C.2.8})$$

cujas distribuição sob a hipótese nula pode ser aproximada por uma distribuição χ_c^2 .

C.3 Método de máxima verossimilhança

Sob a suposição de normalidade, podemos considerar o método de máxima verossimilhança para a estimação dos parâmetros do modelo (C.1.1)-(C.1.2). Neste caso, o método consiste em maximizar a função de verossimilhança

$$L(\boldsymbol{\beta}, \sigma^2 | \mathbf{y}, \mathbf{X}) = (2\pi\sigma^2)^{-n/2} \exp\left[-\frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right]. \quad (\text{C.3.1})$$

com relação aos parâmetros $\boldsymbol{\beta}$ e σ^2 . Lembrando que a determinação dos pontos de máximo de (C.3.1) é equivalente à obtenção dos pontos de máximo de seu logaritmo e utilizando as regras de derivação matricial apresentadas no Apêndice A, podemos mostrar que os estimadores de máxima verossimilhança desejados são

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

e

$$\hat{\sigma}^2 = n^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^\top (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = n^{-1}\mathbf{y}^\top [\mathbf{I}_n - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top] \mathbf{y}.$$

Dado que a maximização de (C.3.1) relativamente a $\boldsymbol{\beta}$ corresponde à minimização da forma quadrática de seu expoente, não é de se estranhar que os estimadores de máxima verossimilhança e de mínimos quadrados sejam idênticos. Por outro lado, o estimador de máxima verossimilhança de σ^2 corresponde ao estimador proposto anteriormente (s^2) multiplicado por um fator $(n-p)/n$ e é, conseqüentemente, enviesado. A razão para isso é que o estimador $\hat{\sigma}^2$ depende de $\boldsymbol{\beta}$ e uma alternativa para se conseguir um estimador não enviesado é utilizar o método de máxima verossimilhança restrita que consiste na minimização de uma transformação linear dos dados que não dependa desse parâmetro. Uma transformação com essas características é dada por $\mathbf{y}^* = [\mathbf{I} - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top] \mathbf{y}$. A maximização da verossimilhança correspondente gera o estimador não enviesado s^2 . Mais detalhes sobre esse método podem ser encontrados em Diggle et al. (2002), por exemplo.

A metodologia de máxima verossimilhança também pode ser aplicada a modelos do tipo (C.1.1) com $\mathbf{e} \sim N(\mathbf{0}, \mathbf{V})$ em que os elementos não redundantes de \mathbf{V} são expressos na forma de um vetor de parâmetros (de covariâncias), $\boldsymbol{\theta}$, *i.e.*, $\mathbf{V} = \mathbf{V}(\boldsymbol{\theta})$. A função de verossimilhança correspondente é

$$L(\boldsymbol{\beta}, \boldsymbol{\theta} | \mathbf{y}, \mathbf{X}) = (2\pi)^{-n/2} |\mathbf{V}(\boldsymbol{\theta})|^{-1/2} \exp[-(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top \mathbf{V}^{-1}(\boldsymbol{\theta}) (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})]. \quad (\text{C.3.2})$$

Nesse caso, o estimador de $\boldsymbol{\beta}$ tem a mesma forma que o estimador de mínimos quadrados generalizados (C.2.6) mas com $\hat{\mathbf{V}} = \mathbf{V}(\hat{\boldsymbol{\theta}})$ representando o estimador de máxima verossimilhança de \mathbf{V} . É possível impor estruturas particulares à matriz de covariâncias, ou seja, é possível considerar situações em que \mathbf{V} é função de um vetor de parâmetros de covariâncias com dimensão reduzida, *e.g.*, 2 ou 3 em vez de $n(n-1)/2$, que é a dimensão de $\boldsymbol{\theta}$ quando não se impõem restrições. Modelos com essa natureza são considerados para a análise de medidas repetidas ou de dados longitudinais.

C.4 Partição da soma de quadrados

A utilidade do modelo de regressão $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$ para explicação da variação da resposta esperada como função das variáveis explicativas pode ser avaliada a partir da identidade

$$y_i - \bar{y} = (\hat{y}_i - \bar{y}) + (y_i - \hat{y}_i)$$

em que $\hat{y}_i = \mathbf{x}_i^\top \hat{\boldsymbol{\beta}}$ com \mathbf{x}_i denotando a i -ésima linha da matriz \mathbf{X} e $\bar{y} = n^{-1} \sum_{i=1}^n y_i$. Pode-se demonstrar que

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^n (y_i - \hat{y}_i)^2.$$

Os termos

$$SQT = \sum_{i=1}^n (y_i - \bar{y})^2 = \mathbf{y}^\top [\mathbf{I}_n - n^{-1} \mathbf{1} \mathbf{1}^\top] \mathbf{y},$$

$$SQR = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 = \mathbf{y}^\top [\mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top - n^{-1} \mathbf{1} \mathbf{1}^\top] \mathbf{y}$$

e

$$SQE = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \mathbf{y}^\top [\mathbf{I}_n - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top] \mathbf{y}$$

correspondem, respectivamente, à **soma de quadrados total**, **soma de quadrados devida à regressão** e **soma de quadrados residual** e representam a variabilidade (expressa em termos de quadrados de diferenças) da resposta em torno de sua média, a parcela dessa variabilidade explicada pelo modelo $\mathbb{E}(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$ e a variabilidade da resposta em torno deste modelo, ou seja a parcela da variabilidade total não explicada.

Quando o ajuste do modelo $\mathbb{E}(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$ é perfeito, *i.e.*, quando $\mathbf{y} = \hat{\mathbf{y}}$ temos $SQE = 0$ e quando o modelo $\mathbb{E}(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$ não melhora a explicação da variabilidade relativamente à resposta média, $SQR = 0$. Uma medida da parcela da variabilidade explicada pelo modelo $\mathbb{E}(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$ é o **coeficiente de determinação**

$$R^2 = \frac{SQR}{SQT} = \frac{SQT - SQE}{SQT} = 1 - \frac{SQE}{SQT}.$$

Como $0 \leq SQE \leq SQT$, temos $0 \leq R^2 \leq 1$ e quanto maior o valor de R^2 , maior é a redução da variabilidade da resposta explicada pela introdução das variáveis explicativas.

Um dos problemas associados à utilização do coeficiente de determinação como medida da qualidade do modelo é que ele não leva em consideração o número de parâmetros envolvidos. Quanto mais variáveis explicativas forem introduzidas no modelo, mais o coeficiente R^2 se aproximará de 1. Para contornar esse problema, podemos considerar o **coeficiente de determinação ajustado**:

$$R_a^2 = 1 - \left(\frac{n-1}{n-p} \right) \frac{SQE}{SQT}.$$

O coeficiente R_a^2 pode diminuir quando adicionamos variáveis explicativas ao modelo pois o decréscimo que isso acarretará em SQE pode ser compensado pela perda de graus de liberdade no denominador $n - p$.

C.5 Diagnóstico

Modelos estatísticos são utilizados como aproximações de processos complexos e são construídos sobre um conjunto de suposições. Para efeitos práticos, é importante avaliar se tais aproximações são aceitáveis. Isto pode ser concretizado por meio de **técnicas de diagnóstico**, que englobam a **avaliação do ajuste** e a **análise de sensibilidade**. No primeiro caso, o objetivo é avaliar se as suposições adotadas são compatíveis com os dados; no segundo, o objetivo é estudar a variação dos resultados da análise quando a formulação inicial do modelo é ligeiramente modificada. Se esta variação for “substancial” no sentido de mudar as conclusões, diz-se que o modelo não é **robusto**. Nesse caso, ou as conclusões devem ser tomadas (se tomadas) de forma cautelosa, ou então deve-se optar por outro modelo.

Para ilustrar a importância do uso das técnicas de diagnóstico, apresentamos na Tabela C.5.1, quatro conjuntos de dados (A, B, C e D) extraídos de Anscombe (1973). Cada um desses conjuntos contém os valores de uma variável explicativa (x) e de uma variável resposta (y).

Tabela C.5.1: Dados obtidos de Anscombe (1973)

A		B		C		D	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

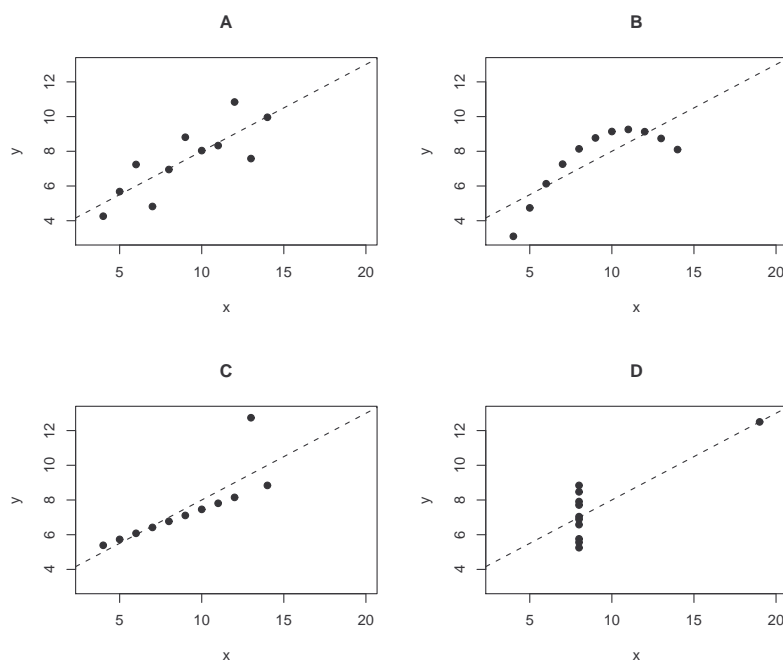
Consideremos o ajuste do seguinte modelo de regressão linear simples

$$y_i = \alpha + \beta x_i + e_i,$$

$i = 1, \dots, 11$ com $\mathbb{E}(e_i) = 0$, $\mathbb{V}(e_i) = \sigma^2$ e $\mathbb{Cov}(e_i, e_j) = 0$, $i \neq j$ a cada um dos quatro subconjuntos de dados. Na Figura C.5.1 apresentamos diagramas de dispersão correspondentes a cada subconjunto; as linhas tracejadas correspondem às retas

ajustadas pelo método de mínimos quadrados. No conjunto A, o modelo parece adequado; no conjunto B, um modelo quadrático seria mais conveniente e no conjunto C, o modelo parece ajustar-se bem aos dados, com exceção do ponto (13.0, 12.74) que se caracteriza como uma observação **discrepante** (*outlier*)⁷. No conjunto D, o coeficiente de regressão β é significativo apenas em função do ponto (19.0, 12.50). No entanto, para os quatro conjuntos de dados, temos $\hat{\alpha} = 3.00$, $\hat{\beta} = 0.50$, $\hat{\sigma}^2 = 1.53$ e $R^2 = 0.667$ o que evidencia que o coeficiente de determinação (R^2) nem sempre é uma boa medida para a avaliação da qualidade do ajuste. Essa avaliação precisa ser complementada por outros métodos, alguns dos quais descrevemos a seguir.

Figura C.5.1: Diagramas de dispersão (dados da Tabela C.5.1)



C.5.1 Análise de resíduos

Resíduos são utilizados para avaliar a validade de determinadas suposições de modelos estatísticos. No caso de modelos de regressão linear clássicos, podemos utilizá-los para verificar homocedasticidade, existência de pontos discrepantes, normalidade e independência dos erros. Nós adotamos a proposta de Cox & Snell (1968), que apre-

⁷Grosso modo, uma observação é chamada discrepante, se apresenta um comportamento distinto das demais.

sentam uma forma geral para definir resíduos para modelos que contêm uma única fonte de variação.

Consideremos o modelo de regressão linear

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e} \quad (\text{C.5.1})$$

com $\mathbf{e} \sim N_n(\mathbf{0}, \sigma^2 \mathbf{I}_n)$. Seja $\hat{\boldsymbol{\beta}}$ o estimador de mínimos quadrados (ordinários) de $\boldsymbol{\beta}$ e $\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y} = \mathbf{H}\mathbf{y}$, o vetor dos valores ajustados. A matriz

$$\mathbf{H} = \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top$$

é simétrica e idempotente e representa a matriz de projeção de \mathbf{y} no subespaço $\mathcal{C}(\mathbf{X})$ (para maiores detalhes, veja o Teorema A.2.7). O vetor de **resíduos ordinários** é definido como

$$\hat{\mathbf{e}} = [\hat{e}_1, \dots, \hat{e}_n]^\top = \mathbf{y} - \hat{\mathbf{y}} = (\mathbf{I} - \mathbf{H})\mathbf{y} \quad (\text{C.5.2})$$

$$= (\mathbf{I} - \mathbf{H})(\mathbf{X}\boldsymbol{\beta} + \mathbf{e}) = (\mathbf{I} - \mathbf{H})\mathbf{e}, \quad (\text{C.5.3})$$

de onde podemos concluir que $\hat{\mathbf{e}} \sim N_n(\mathbf{0}, \sigma^2[\mathbf{I} - \mathbf{H}])$. A relação entre o vetor de erros e o vetor de resíduos ordinários depende somente da **matriz chapéu** (*hat matrix*) \mathbf{H} .⁸

Apresentamos a seguir algumas transformações dos resíduos ordinários apropriadas para fins de diagnóstico. O leitor interessado pode consultar Cook & Weisberg (1982), por exemplo, para detalhes.

A distribuição do vetor de resíduos ordinários depende tanto de σ^2 quanto da matriz \mathbf{H} ; conseqüentemente, os resíduos podem ter variâncias distintas. Para efeito de comparação, convém construir **resíduos padronizados** que não dependam dessas quantidades. Se σ^2 for conhecido, uma padronização natural consistiria na divisão de \hat{e}_i pelo seu desvio padrão, a saber, $\sigma\sqrt{1 - h_{ii}}$, em que h_{ii} denota o i -ésimo elemento da diagonal principal de \mathbf{H} . Pode-se mostrar que $h_{ii} = \mathbf{x}_i^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{x}_i$ em que \mathbf{x}_i^\top é a i -ésima linha de \mathbf{X} . Com essa padronização, a distribuição conjunta dos resíduos transformados não depende da variância. Como, em geral, σ^2 é desconhecida, uma alternativa é considerar os **resíduos studentizados**, definidos por

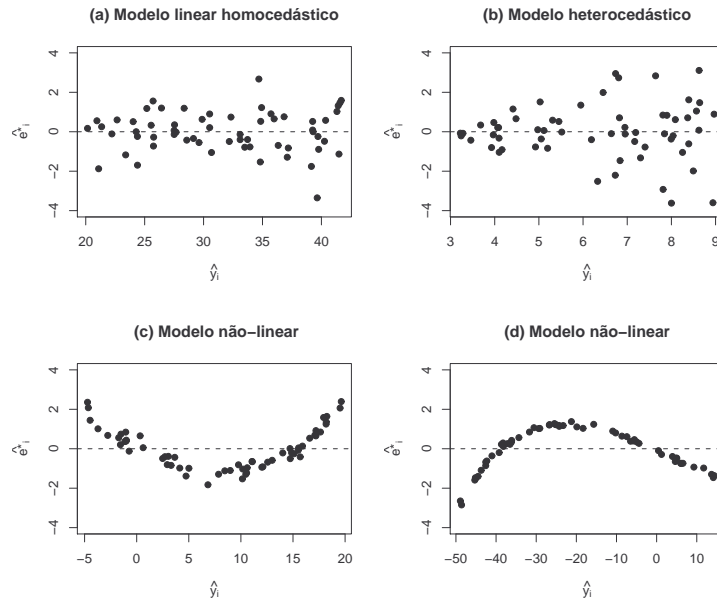
$$\hat{e}_i^* = \frac{\hat{e}_i}{s\sqrt{1 - h_{ii}}}, \quad i = 1, \dots, n.$$

Um gráfico de \hat{e}_i^* versus \hat{y}_i (o valor predito da i -ésima observação) é útil para verificar a plausibilidade da suposição de homocedasticidade; quando esta suposição é verdadeira, espera-se que o comportamento de \hat{e}_i^* em torno do valor zero seja aleatório

⁸Essa denominação, dada por J.W. Tukey, se deve ao fato de \mathbf{H} ser base da transformação linear do vetor dos valores observados no vetor de valores ajustados, $\hat{\mathbf{y}} = \mathbf{H}\mathbf{y}$, ou seja, \mathbf{H} “coloca” o chapéu ($\hat{}$) no vetor \mathbf{y} . Ela também é chamada de **matriz de predição**.

e que sua variabilidade tenha magnitude independente do valor de \hat{y}_i . Esse tipo de gráfico também pode ser utilizado para avaliar a suposição de linearidade assumida pelo modelo, além de revelar para quais observações essa suposição parece inadequada. Se a suposição for verdadeira, esperamos que nenhuma tendência na disposição dos resíduos seja detectada. Como ilustração, considere a Figura C.5.2 que corresponde aos resíduos associados a quatro ajustes fictícios. No gráfico C.5.2(a) não detectamos nenhum tipo de violação das suposições de homocedasticidade e de linearidade; no gráfico C.5.2(b), podemos notar que a variabilidade dos resíduos aumenta com a magnitude do valor ajustado, indicando que o modelo em questão deve ser heterocedástico e nos gráficos C.5.2(c) e C.5.2(d) podemos notar uma tendência não linear dos resíduos, o que sugere a má especificação da forma funcional. Nesses casos, uma função não linear talvez fosse mais adequada para representar a relação entre o valor esperado da variável resposta e os valores da variável explicativa.

Figura C.5.2: Resíduos studentizados *versus* valores ajustados



Os resíduos studentizados que apresentam um valor absoluto “muito grande” (maior que 2 ou 3, por exemplo) identificam observações discrepantes. Este critério tem cunho puramente descritivo, dado que na realidade $(\hat{e}_i^*)^2/(n-p) \sim \text{Beta}[1/2, (n-p-1)/2]$ ⁹, indicando que $|\hat{e}_i^*| \leq \sqrt{n-p}$. Além disso, pode-se mostrar, que $\mathbb{E}[\hat{e}_i^*] = 0$, $\mathbb{V}[\hat{e}_i^*] = 1$ e que $\text{Cov}[\hat{e}_i^*, \hat{e}_j^*] = -h_{ij}/[(1-h_{ii})(1-h_{jj})]^{1/2}$. Para detalhes e demonstração destas propriedades, veja Cook & Weisberg (1982), por exemplo.

⁹ $\text{Beta}(a, b)$ denota a distribuição beta com parâmetros $a, b > 0$.

Na definição dos resíduos studentizados, utilizamos o quadrado médio do resíduo (s^2) como estimador de σ^2 . Dada a dependência entre \hat{e}_i e s^2 , a distribuição de \hat{e}_i^* não possui uma densidade com tratamento matemático fácil. Isto motiva outra forma de padronização dos resíduos ordinários em que um estimador de σ^2 independente de e_i é utilizado. Com essa finalidade, podemos considerar o quadrado médio do resíduo obtido com a omissão da i -ésima observação, que pode ser calculado como

$$s_{(i)}^2 = s^2 \left(\frac{n - p - (\hat{e}_i^*)^2}{n - p - 1} \right).$$

Pode-se provar que \hat{e}_i^* e $s_{(i)}^2$ são independentes, de forma que definimos os **resíduos studentizados externamente** como

$$t_i = \frac{\hat{e}_i}{s_{(i)} \sqrt{1 - h_{ii}}}, \quad i = 1, \dots, n.$$

Cook & Weisberg (1982) mostram que $t_i \sim t(n - p - 1)$ e que

$$t_i = \hat{e}_i^* \left(\frac{n - p - 1}{n - p - (\hat{e}_i^*)^2} \right)^{1/2},$$

indicando que t_i é uma transformação monótona crescente de \hat{e}_i^* e proporcionando uma forma de cálculo que prescinde do ajuste do modelo sem a i -ésima observação.

Além disso, é possível mostrar que valores “grandes” de $|t_i|$, e consequentemente de $|\hat{e}_i^*|$, podem ser utilizados como identificadores de observações significativamente discrepantes; detalhes que incluem sugestões para pontos de corte podem ser obtidos em Cook & Weisberg (1982).

Outros tipos de resíduos, tais como **resíduos preditos** (*predicted residuals*) e **resíduos recursivos** (*recursive residuals*), também são discutidos em Cook & Weisberg (1982).

C.5.2 Análise da suposição de normalidade

Na teoria clássica de regressão, tanto intervalos de confiança quanto testes de hipóteses sobre os parâmetros de modelos lineares são baseados na suposição de normalidade dos erros. A verificação da plausibilidade dessa suposição é fundamental para a validade dos procedimentos inferenciais (exatos). Como os resíduos são essencialmente preditores dos erros do modelo, nada mais natural do que utilizá-los com essa finalidade. Nesse sentido, gráficos do tipo **QQ** (**quantis-quantis**), em que dispomos os resíduos studentizados ordenados (quantis observados) no eixo das ordenadas e os quantis obtidos da distribuição normal padrão (quantis teóricos) no

eixo das abscissas, são ferramentas utilíssimas. Quando a distribuição dos erros é gaussiana, espera-se que esses resíduos estejam dispostos numa vizinhança da reta com inclinação de 45 graus. Esse tipo de gráfico também pode ser útil para detectar a presença de observações discrepantes, para avaliar se a distribuição dos erros possui caudas mais pesadas que a distribuição normal, para avaliar se os erros são heterocedásticos etc.

Esses gráficos podem ser obtidos por meio do seguinte procedimento:

- i) Ajustar o modelo (C.5.1), obter o vetor de resíduos studentizados $\hat{\mathbf{e}}^* = (\hat{e}_1^*, \dots, \hat{e}_n^*)^\top$ e o vetor de resíduos studentizados ordenados, $(\hat{e}_{(1)}^* \leq \hat{e}_{(2)}^* \leq \dots \leq \hat{e}_{(n)}^*)^\top$.
- ii) Calcular $p_i = (i - 0.375)/n$, $i = 1, \dots, n$ e definir os quantis amostrais de ordem p_i , $i = 1, \dots, n$ como $Q_{p_i} = \hat{e}_{(i)}^*$.
- iii) Obter os quantis normais $Z_{p_i} = P[Z \leq Z(p_i)] = p_i$, $i = 1, \dots, n$ em que Z é uma variável com distribuição $N(0, 1)$.
- iv) Construir o gráfico de dispersão $Q_{p_i} \times Z_{p_i}$.

Como em gráficos do tipo QQ, é difícil avaliar afastamentos da normalidade visualmente, Atkinson (1985), sugere a construção de bandas de confiança, denominadas **envelopes simulados** (*simulated envelopes*). Essas bandas de confiança são obtidas por meio de simulação de dados com distribuição normal com vetor de médias iguais a zero e matriz de covariâncias $(\mathbf{I} - \mathbf{H})$.

Um algoritmo para sua construção é o seguinte

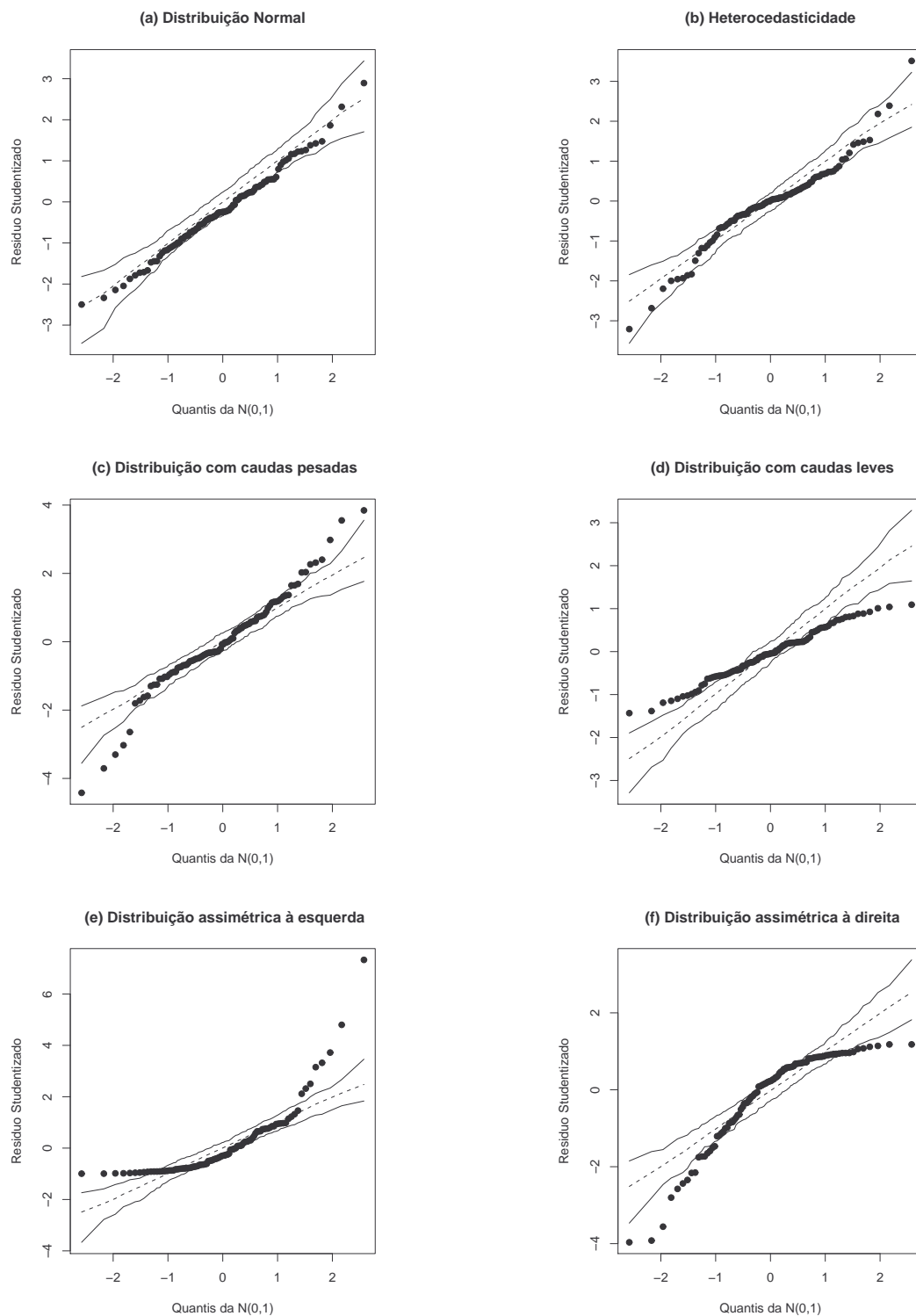
- i) Ajustar o modelo (C.5.1), obtendo $\hat{\beta}$, s^2 e o vetor de resíduos studentizados $\hat{\mathbf{e}}^*$.
- ii) Construir o gráfico QQ correspondente.
- iii) Gerar um vetor \mathbf{z} com n elementos correspondentes a valores independentes de uma distribuição $N(0, 1)$.
- iv) Obter um vetor de observações simuladas $\mathbf{y}_s = \mathbf{X}\hat{\beta} + \mathbf{sz}$.
- v) Ajustar o modelo $\mathbf{y}_s = \mathbf{X}\beta + \mathbf{e}_s$, obter o vetor de resíduos studentizados (simulados) $\hat{\mathbf{e}}_s^*$ e ordenar seus componentes.
- vi) Repetir os itens (iii)-(v) m vezes, gerando para cada $p_i = (i - 0.375)/n$, um conjunto de m resíduos simulados.
- vii) Selecionar para cada p_i , o menor e o maior resíduo simulado e incluí-los no gráfico QQ construído no item (ii).
- viii) Construir o limite inferior (superior) do envelope simulado ligando os pontos do conjunto de menores (maiores) resíduos simulados.

Se o modelo estiver correto, espera-se que todos os pontos observados no gráfico QQ fiquem contidos no envelope simulado em $[m/(m+1)] \times 100\%$ das vezes. Quando $m = 19$, isso corresponde a 95% das vezes.

Para detalhes a respeito de sua construção, veja Atkinson (1985), Atkinson & Riani (2000) ou Paula (2004), por exemplo. O mesmo tipo de procedimento pode ser empregado em situações em que outras distribuições para os erros são adotadas no modelo. Paula (2004) apresenta vários algoritmos (em linguagem **S-Plus** e **R**) para construção de gráficos envelope simulados, baseados em várias distribuições (Gama, Binomial, Poisson etc.).

Na Figura C.5.3, apresentamos exemplos de gráficos QQ (e envelopes simulados) com diferentes padrões de afastamento das suposições de normalidade, homocedasticidade ou simetria. Por exemplo, no gráfico C.5.3(a), não há indícios contrários às hipóteses de normalidade e homocedasticidade; no gráfico C.5.3(b), os resíduos estão próximos dos limites das bandas de confiança, evidenciando que a distribuição dos erros deve ser heterocedástica. O gráfico C.5.3(c) [C.5.3(d)] apresenta características típicas de casos em que a distribuição dos erros padronizados tem caudas mais pesadas (mais leves) que a distribuição normal padrão. Quando a distribuição dos erros padronizados possui caudas mais pesadas (leves) que a distribuição normal padrão, o gráfico QQ assume a forma de **S**, com os quantis extremos do resíduo studentizado maiores (menores) que os quantis teóricos da distribuição normal padrão. Por fim, o gráfico C.5.3(e) [C.5.3(f)] mostra uma situação em que a distribuição dos erros é assimétrica à direita (à esquerda). Quando se tem uma distribuição assimétrica à direita, o gráfico apresenta a forma de um **J**; se a distribuição for assimétrica à esquerda, o gráfico apresenta a forma de um **J** invertido.

Figura C.5.3: Envelopes simulados com coeficiente de confiança 95%



C.5.3 Análise de sensibilidade

A análise de sensibilidade visa avaliar o comportamento do ajuste de um modelo sujeito a algum tipo de perturbação, ou seja, sob alguma mudança nas hipóteses ou nos dados. Como cada observação não tem a mesma influência em todas as características do ajuste do modelo, é natural que se defina aquela na qual se quer focar a análise. Se o objetivo for fazer previsões, então é razoável medir a influência das observações nos valores preditos e não nos parâmetros de localização, como mencionam Chatterjee & Hadi (1986) e Chatterjee & Hadi (1988).

Existem medidas de influência baseadas nos resíduos, na curva de influência¹⁰, na verossimilhança, no volume dos elipsoides de confiança, em um subconjunto do vetor de parâmetros de localização (influência parcial) e nos pontos remotos do espaço vetorial gerado pelas colunas da matriz de especificação \mathbf{X} .

Dentre as abordagens mais utilizadas na prática para medir influência em modelos lineares, destacam-se aquelas baseadas em **influência local** considerada em Cook (1986) e aquelas obtidas por intermédio da eliminação de observações (**influência global**).

O **poder de alavanca** (*leverage*) da i -ésima observação (y_i) é a derivada parcial $\partial \hat{y}_i / \partial y_i$ e indica a taxa de variação do i -ésimo valor predito quando a i -ésima observação é acrescida de um infinitésimo. No modelo linear clássico, o vetor de valores preditos (ajustados) é dado por $\hat{\mathbf{y}} = \mathbf{H}\mathbf{y}$ e conseqüentemente,

$$\partial \hat{\mathbf{y}} / \partial \mathbf{y} = \mathbf{H},$$

indicando que o poder de alavanca da i -ésima observação é dado por h_{ii} , ou seja, pelo i -ésimo elemento da diagonal principal da matriz de projeção \mathbf{H} . Como essa matriz é simétrica e idempotente, é possível mostrar que¹¹

$$\begin{aligned} 0 &\leq h_{ii} \leq 1, \quad i = 1, \dots, n, \\ h_{ij} &\leq h_{ii}(1 - h_{ii}), \quad 1 \leq i < j \leq n. \end{aligned}$$

Desta forma, se $h_{ii} = 1$, então $\hat{y}_i = y_i$, implicando que a i -ésima observação tem influência total no seu valor predito. Além disso, como $\text{tr}(\mathbf{H}) = \sum_{i=1}^n h_{ii} = p$, então o valor médio do poder de alavanca é p/n . Quando todos os n elementos da diagonal principal de \mathbf{H} são próximos de p/n , nenhuma observação influencia o seu valor predito de forma desproporcional; então podemos dizer que a i -ésima

¹⁰ A curva de influência de uma estatística $T(x_1, \dots, x_n)$ corresponde a um gráfico com os valores de uma nova observação amostral x_{n+1} no eixo das abscissas e com os valores da estatística $T(x_1, \dots, x_n, x_{n+1})$ no eixo das ordenadas; ela dá uma ideia de como a estatística é influenciada pelo acréscimo de uma única observação amostral.

¹¹ Também é possível mostrar que $n^{-1} \leq h_{ii} \leq 1$ se o modelo incluir intercepto.

observação tem alto poder de alavanca (*high leverage*) se $h_{ii} \geq 2p/n$ ou $h_{ii} \geq 3p/n$, por exemplo. Esse critério é arbitrário e deve ser encarado com espírito puramente descritivo. Em função de (C.5.2), podemos concluir que, em geral, observações com alto poder de alavanca apresentam resíduos pequenos relativamente aos demais.

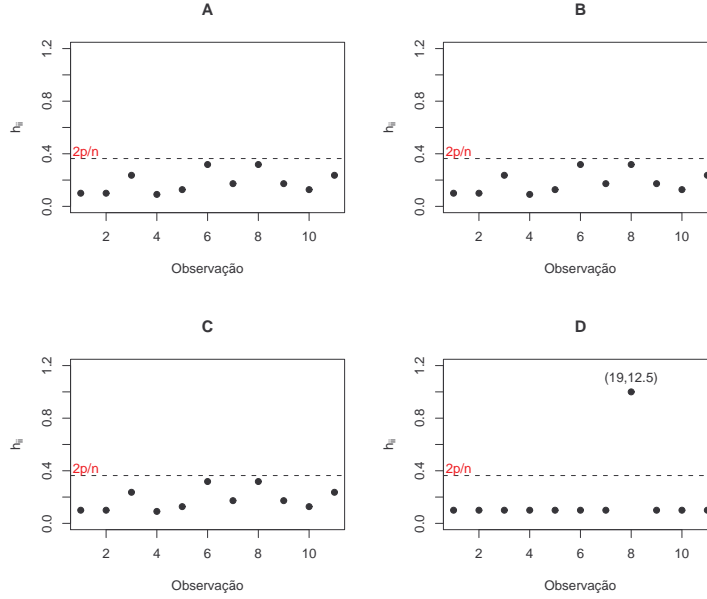
Para exemplificar, consideremos o modelo de regressão $y_i = \alpha + \beta x_i + e_i$, $i = 1, \dots, n$, com $\mathbb{E}[e_i] = 0$, $\mathbb{V}[e_i] = \sigma^2$ e $\mathbb{Cov}(e_i, e_j) = 0$, $i \neq j$. Nesse caso, o poder de alavanca da i -ésima observação é dado por

$$h_{ii} = \frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2},$$

e conseqüentemente, a i -ésima observação tem alto poder de alavanca se x_i for um valor discrepante no conjunto $\{x_1, \dots, x_n\}$.

De uma forma geral, é possível mostrar que observações com alto poder de alavanca são aquelas associadas a covariáveis discrepantes em $\mathcal{C}(\mathbf{X})$ como observado por Chatterjee & Hadi (1988) e Wei et al. (1998). Mais detalhes, podem ser obtidos em Cook & Weisberg (1982), Chatterjee & Hadi (1988), Wei et al. (1998), Atkinson & Riani (2000) e Paula (2004), por exemplo.

A título de ilustração, na Figura C.5.4, mostramos os gráficos de h_{ii} versus índice das observações para os quatro subconjuntos de dados apresentados na Tabela C.5.1. Nos subconjuntos A, B e C não existe indício de observações com alto poder de alavanca. Por outro lado, no subconjunto D, a observação 8 apresenta $h_{88} = 1$ indicando que o coeficiente angular da regressão é significativo apenas pela sua presença. Como os valores da variável explicativa são idênticos nos conjuntos A, B e C, as matrizes \mathbf{H} correspondentes também são as mesmas e conseqüentemente, os poderes de alavanca associados são iguais.

Figura C.5.4: Poder de alavanca *versus* índice das observações

Consideremos agora um modelo de regressão linear $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$ com $\mathbb{E}(\mathbf{e}) = \mathbf{0}$ e $\mathbb{V}(\mathbf{e}) = \sigma^2 \mathbf{I}$ e suponhamos que há interesse em incluir uma nova variável explicativa, digamos w , de forma que o modelo passa a ser

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \gamma \mathbf{w} + \mathbf{e}^* = \mathbf{X}^* \boldsymbol{\beta}^* + \mathbf{e}^* \quad (\text{C.5.4})$$

com γ representando o coeficiente da nova variável, $\mathbf{w} = (w_1, \dots, w_n)^\top$, $\mathbf{X}^* = (\mathbf{X}, \mathbf{w})$, $\boldsymbol{\beta}^* = (\boldsymbol{\beta}^\top, \gamma)^\top$ e \mathbf{e}^* representando um conjunto de erros aleatórios homocedásticos e não correlacionados. Sob normalidade dos erros, podemos verificar a plausibilidade da inclusão da nova covariável, por exemplo, por meio de um teste t para $\gamma = 0$. Uma forma alternativa para avaliar a importância da inclusão da covariável é descrita a seguir.

Inicialmente, observemos que as equações de estimação (equações normais) para a obtenção do estimador de mínimos quadrados ordinários de $\boldsymbol{\beta}^*$ sob o modelo (C.5.4) são

$$\begin{aligned} \mathbf{X}^\top \mathbf{X} \hat{\boldsymbol{\beta}} + \mathbf{X}^\top \mathbf{w} \hat{\gamma} &= \mathbf{X}^\top \mathbf{y} \\ \mathbf{w}^\top \mathbf{X} \hat{\boldsymbol{\beta}} + \mathbf{w}^\top \mathbf{w} \hat{\gamma} &= \mathbf{w}^\top \mathbf{y}. \end{aligned}$$

Resolvendo esse sistema, obtemos o estimador de mínimos quadrados

$$\hat{\gamma} = \frac{\mathbf{w}^\top (\mathbf{I} - \mathbf{H}) \mathbf{y}}{\mathbf{w}^\top (\mathbf{I} - \mathbf{H}) \mathbf{w}}.$$

Lembrando que $(\mathbf{I} - \mathbf{H})$ é uma matriz idempotente, podemos reexpressar esse estimador como

$$\hat{\gamma} = \frac{\mathbf{w}^\top (\mathbf{I} - \mathbf{H})(\mathbf{I} - \mathbf{H})\mathbf{y}}{\mathbf{w}^\top (\mathbf{I} - \mathbf{H})(\mathbf{I} - \mathbf{H})\mathbf{w}} = \frac{\hat{\mathbf{e}}_w^\top \hat{\mathbf{e}}}{\hat{\mathbf{e}}_w^\top \hat{\mathbf{e}}_w},$$

com $\hat{\mathbf{e}} = (\mathbf{I} - \mathbf{H})\mathbf{y}$ e $\hat{\mathbf{e}}_w = (\mathbf{I} - \mathbf{H})\mathbf{w}$ de forma que ele pode ser interpretado como o estimador de mínimos quadrados (ordinários) do coeficiente angular de uma regressão (sem intercepto) tendo como variável resposta os resíduos ordinários ($\hat{\mathbf{e}}$) obtidos do modelo inicial (sem a covariável w) e como variável explicativa os resíduos ($\hat{\mathbf{e}}_w$) obtidos do modelo de regressão $\mathbf{w} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}_w$.

O gráfico de dispersão entre os resíduos $\hat{\mathbf{e}}$ e $\hat{\mathbf{e}}_w$, conhecido como **gráfico da variável adicionada** ou **gráfico de regressão parcial** (*partial regression plot*) fornece informação sobre os ganhos com a inclusão da covariável w no modelo. Ele também pode ser útil para identificar pontos que se desviam da relação linear entre os resíduos, e que podem ser encarados como observações influentes na estimação de γ . Mais detalhes e extensões, que fogem ao escopo deste texto, podem ser encontrados em Cook & Weisberg (1989) e Cook (1996), por exemplo.

Em modelos de regressão linear múltipla, uma estratégia para identificar observações influentes na estimação dos coeficientes do vetor de parâmetros, consiste em construir um gráfico desse tipo para cada covariável do modelo.

Cook (1977), por outro lado, sugere que a influência de uma particular observação, ou de um conjunto de observações, seja avaliada por intermédio dos efeitos provocados por sua eliminação do conjunto de dados.

Consideremos o modelo de regressão linear (C.5.1) e denotemos por $\hat{\boldsymbol{\beta}}$ e $\hat{\boldsymbol{\beta}}_{(I)}$, respectivamente, os estimadores de mínimos quadrados de $\boldsymbol{\beta}$ obtidos com todos os dados da amostra e com a eliminação do conjunto de observações I . Nesse contexto, uma das medidas mais utilizadas é a **distância de Cook** definida por

$$D_I = \frac{[\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(I)}]^\top (\mathbf{X}^\top \mathbf{X}) [\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(I)}]}{ps^2} = \frac{[\hat{\mathbf{y}} - \hat{\mathbf{y}}_{(I)}]^\top [\hat{\mathbf{y}} - \hat{\mathbf{y}}_{(I)}]}{ps^2}, \quad (\text{C.5.5})$$

com $\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$ e $\hat{\mathbf{y}}_{(I)} = \mathbf{X}\hat{\boldsymbol{\beta}}_{(I)}$ representando respectivamente, o vetor de valores preditos pelo ajuste do modelo com todos os dados da amostra e aquele obtido com a eliminação do conjunto de observações I . A estatística D_I mede a influência das observações do conjunto I na estimativa de $\boldsymbol{\beta}$ segundo a métrica definida por $(ps^2)^{-1}\mathbf{X}^\top \mathbf{X}$ ou equivalentemente, a influência dessas observações no vetor de valores preditos. Valores grandes de D_I indicam que as observações do conjunto I são influentes na estimação de $\boldsymbol{\beta}$ ou nos valores de $\hat{\mathbf{y}}$. Sob essa abordagem, é essencial obter expressões que relacionem o estimador do parâmetro de interesse calculado com base em toda amostra com o respectivo estimador calculado após a eliminação de um

conjunto de observações sem a necessidade de reajustar o modelo. Em particular, quando apenas a i -ésima observação é eliminada, é possível mostrar que

$$\hat{\beta}_{(i)} - \hat{\beta} = -\frac{\hat{e}_i^*}{1 - h_{ii}}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{x}_i,$$

de tal forma que a distância de Cook correspondente é

$$D_i = \frac{[\hat{\beta} - \hat{\beta}_{(i)}]^\top (\mathbf{X}^\top \mathbf{X}) [\hat{\beta} - \hat{\beta}_{(i)}]}{ps^2} = \frac{\hat{e}_i^{*2}}{p} \frac{h_{ii}}{(1 - h_{ii})^2}. \quad (\text{C.5.6})$$

Dada a expressão (C.5.6), a influência da i -ésima observação depende tanto do respectivo resíduo studentizado quanto do grau de alavanca da observação em questão; se $h_{ii} \approx 0$ (indicando um baixo poder de alavanca) então D_i assume um valor “pequeno”, mesmo quando a i -ésima observação for altamente discrepante, indicando que a distância (C.5.6) pode não ser adequada nessas situações.

Outras propostas são sugeridas na literatura para contornar este fato. Por exemplo, Belsley et al. (1980) sugerem a utilização de

$$\text{DFITS}_i = \frac{|\hat{e}_i^*|}{s_{(i)}(1 - h_{ii})^{1/2}} \left(\frac{h_{ii}}{1 - h_{ii}} \right)^{1/2} = |t_i| \left(\frac{h_{ii}}{1 - h_{ii}} \right)^{1/2}, \quad (\text{C.5.7})$$

enquanto Atkinson (1981) sugere uma versão modificada da distância de Cook, obtida com a substituição de s^2 por $s_{(i)}^2$ e ajustada pelo tamanho da amostra, a saber,

$$C_i = \left(\frac{n - p}{p} \frac{h_{ii}}{1 - h_{ii}} \right)^{1/2} |t_i|. \quad (\text{C.5.8})$$

Quando todos os h_{ii} são iguais, $C_i = |t_i|$. Chatterjee & Hadi (1988) sugerem outros usos para esta versão da distância de Cook. Em particular, esses autores comentam que $C_i^* = \text{sign}(y_i - \hat{y}_i)C_i$ ($i = 1, \dots, n$) podem ser utilizados como resíduos.

Em tese, essas três medidas de influência competem entre si. Todavia, como destacam Cook, Peña & Weisberg (1988) e Paula (2004) elas servem para avaliar diferentes aspectos da influência das observações nos resultados do ajuste. Por exemplo, (C.5.6) é mais adequada para medir a influência das observações nos parâmetros de localização (β), enquanto que (C.5.7) tem o objetivo de medir a influência das observações nos parâmetros de localização e escala simultaneamente embora possa falhar em algumas situações, como indicam Cook et al. (1988). Segundo Paula (2004), observações discrepantes com baixo poder de alavanca dificilmente influem na estimação de β e não comprometem o potencial uso de (C.5.5).

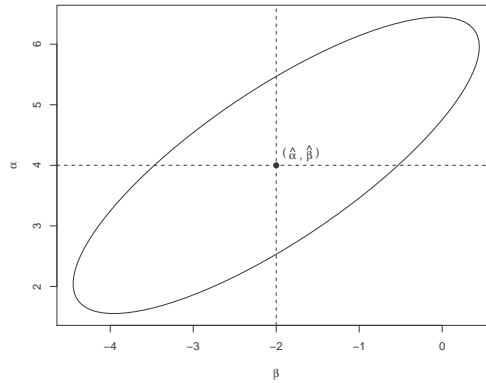
Gráficos das medidas de influência *versus* índices das observações são ferramentas úteis para a identificação observações influentes, *i.e.*, aquelas com valores da medida de influência “grandes” em relação aos demais.

Medidas de influência como a distância de Cook são baseadas na mudança do centro (definido em termos dos parâmetros de regressão) dos **elipsóides de confiança** para o vetor de parâmetros β , nomeadamente

$$\left\{ \beta \in \mathbb{R}^p; (\hat{\beta} - \beta)^\top (\mathbf{X}^\top \mathbf{X}) (\hat{\beta} - \beta) \leq ps^2 F_{p, (n-p)}(\alpha) \right\} \quad (\text{C.5.9})$$

em que $F_{p, (n-p)}(\alpha)$ denota o quantil de ordem $1 - \alpha$ da distribuição F com p graus de liberdade no numerador e $n - p$ no denominador. A avaliação da influência de um conjunto de observações também pode ser concretizada a partir do volume desses elipsóides. Na Figura C.5.5, ilustramos um elipsóide de confiança com coeficiente de confiança de 95% para os parâmetros de uma regressão linear simples. O volume do elipsóide está diretamente relacionado com a estimativa da matriz de covariâncias de $\hat{\beta}$ e por conseguinte, a avaliação da influência das observações no respectivo volume reveste-se de particular importância, uma vez que a variância dos estimadores dos coeficientes podem ser extremamente afetadas por poucas observações. Aqui apresentamos duas propostas para tal fim; para detalhes, sugerimos uma consulta a Andrews & Pregibon (1978), Belsley et al. (1980), Cook & Weisberg (1982) ou Chatterjee & Hadi (1988), por exemplo.

Figura C.5.5: Elipsóide de confiança com coeficiente de confiança 95%



O volume do elipsóide (C.5.9) é inversamente proporcional à raiz quadrada de $|\mathbf{X}^\top \mathbf{X}|$. Por isso, Andrews & Pregibon (1978) sugeriram avaliar a influência da i -ésima observação por meio de

$$\text{AP}_i = \frac{s_{(i)}^2 |\mathbf{X}_{(i)}^\top \mathbf{X}_{(i)}|}{s^2 |\mathbf{X}^\top \mathbf{X}|} = h_{ii} + (1 - h_{ii}) \frac{\widehat{e}_i^{*2}}{n - p}. \quad (\text{C.5.10})$$

Valores de AP_i muito distantes de 1 indicam que a observação em questão é potencialmente influente com relação à matriz de covariâncias de $\hat{\beta}$. Propriedades desta

medida podem ser encontrados em Cook & Weisberg (1982) ou Chatterjee & Hadi (1988), por exemplo.

Belsley et al. (1980) sugerem que se avalie a influência da i -ésima observação na matriz de covariâncias de $\hat{\beta}$ por meio de

$$\begin{aligned} \text{COVRATIO}_i &= \frac{|\hat{V}[\hat{\beta}_{(i)}]|}{|\hat{V}[\hat{\beta}]|} = \frac{|s_{(i)}^2 (\mathbf{X}_{(i)}^\top \mathbf{X}_{(i)})^{-1}|}{|s^2 (\mathbf{X}^\top \mathbf{X})^{-1}|} \\ &= \left(\frac{n - p - \hat{e}_i^*}{n - p - 1} \right)^p \frac{1}{1 - h_{ii}}. \end{aligned} \quad (\text{C.5.11})$$

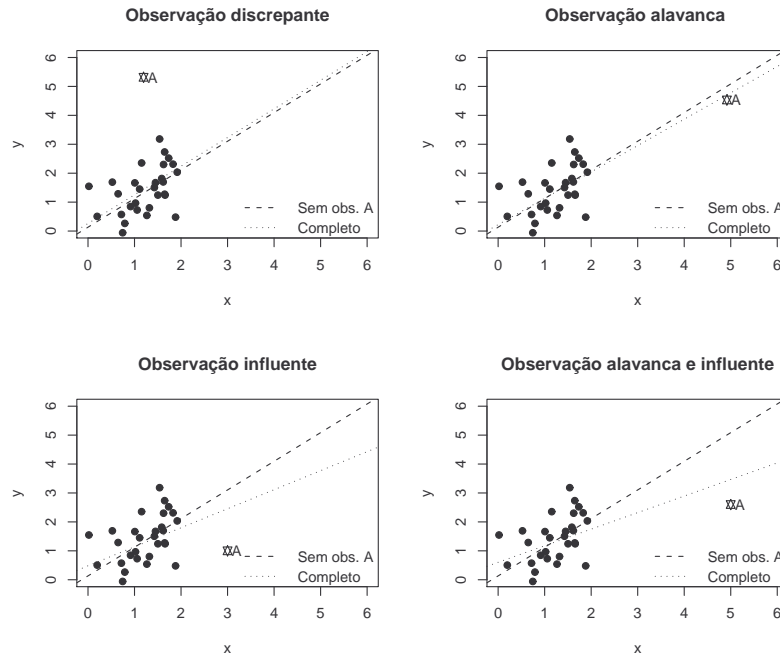
Se todas as observações tiverem o mesmo grau de influência, espera-se que $\text{COVRATIO}_i \approx 1$; afastamentos da unidade indicam que a observação correspondente é potencialmente influente com relação à matriz de covariâncias de $\hat{\beta}$. Belsley et al. (1980) sugerem utilizar o seguinte limiar para identificar observações influentes

$$|\text{COVRATIO}_i - 1| \geq 3p/n.$$

Esse limiar, assim como aqueles mencionados anteriormente, tem cunho totalmente descritivo.

Na Figura C.5.6 mostramos gráficos que ilustram a diferença entre observações discrepantes, alavanca e influentes.

Figura C.5.6: Diferença entre observações alavanca, discrepantes e influentes



Todos os procedimentos aqui discutidos são baseados na eliminação de conjuntos de observações e são conhecidos na literatura como métodos de **influência global**. Todavia existem técnicas de diagnóstico baseados na discrepância da função de verossimilhança quando perturbamos as observações de alguma forma. Tais técnicas, conhecidas sob a denominação geral de **influência local**, foram propostas por Cook (1986). Para detalhes, veja, por exemplo, Cook (1986) ou Paula (2004).

C.5.4 Análise da suposição de correlação nula

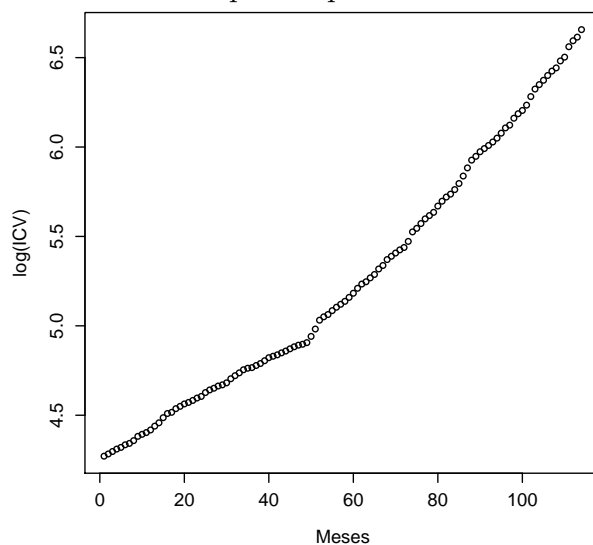
Em geral, a suposição de que os erros do modelo linear são não correlacionados deve ser questionada com base no procedimento de coleta de dados. Como ilustração, consideramos dois exemplos nos quais essa característica justifica a dúvida. O primeiro exemplo é um caso simples dos problemas abordados pelas técnicas de análise de séries cronológicas; o segundo exemplo é o caso típico daqueles que constituem o objeto do núcleo deste texto. Ambos são apresentados aqui com a finalidade de mostrar como as técnicas de análise de regressão podem ser empregadas para analisar modelos mais gerais do que aqueles governados pelo paradigma de Gauss-Markov.

Exemplo C.5.1: Na Tabela C.5.2 apresentamos valores do índice de custo de vida (ICV) na cidade de São Paulo colhidos pela Fundação Getúlio Vargas entre janeiro de 1970 e junho de 1979 com o objetivo de avaliar seu crescimento nesse período. O gráfico de dispersão correspondente está disposto na Figura C.5.7 Com base em argumentos de teoria econômica, é razoável supor que o ICV num determinado mês seja correlacionado com aqueles obtidos em meses anteriores.

Tabela C.5.2: Índice de custo de vida para São Paulo (jan/70 a jul/79)

Obs	ICV	Obs	ICV	Obs	ICV	Obs	ICV	Obs	ICV
1	71.6	24	100	47	133	70	223	93	415
2	72.5	25	102	48	134	71	227	94	424
3	73.5	26	104	49	135	72	230	95	436
4	74.5	27	105	50	140	73	238	96	449
5	75.2	28	106	51	146	74	251	97	456
6	76.3	29	107	52	153	75	256	98	474
7	76.9	30	108	53	156	76	263	99	486
8	78.1	31	110	54	158	77	270	100	495
9	80	32	112	55	162	78	275	101	510
10	80.9	33	114	56	165	79	280	102	535
11	81.7	34	116	57	167	80	290	103	558
12	82.9	35	117	58	170	81	298	104	572
13	84.7	36	118	59	174	82	305	105	586
14	86.3	37	119	60	178	83	310	106	602
15	88.8	38	120	61	183	84	318	107	617
16	90.9	39	122	62	188	85	329	108	628
17	91.5	40	124	63	190	86	343	109	653
18	93.4	41	125	64	194	87	359	110	667
19	94.6	42	126	65	198	88	375	111	707
20	95.9	43	128	66	204	89	383	112	731
21	96.7	44	129	67	208	90	393	113	746
22	97.8	45	131	68	215	91	400	114	778
23	99.1	46	132	69	219	92	407		

Figura C.5.7: Gráfico de dispersão para os dados do Exemplo C.5.1



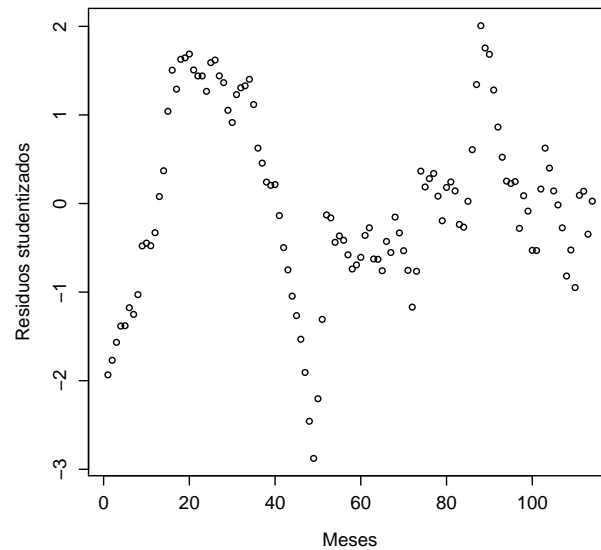
Tendo em vista o gráfico de dispersão apresentado na Figura C.5.7, uma primeira abordagem para a análise dos dados do Exemplo C.5.1 poderia envolver um modelo da forma

$$\log(y_t) = \alpha + \beta t + \gamma t^2 + e_t, \quad (\text{C.5.12})$$

$t = 1, \dots, n$ em que y_t representa o ICV no instante t , α denota o valor esperado do ICV no tempo $t = 0$, β e γ representam os componentes linear e quadrático da curva que rege a variação temporal do logaritmo do ICV e e_t denota um erro aleatório. Utilizamos t como índice para salientar que as observações são colhidas sequencialmente.

O coeficiente de determinação $R^2 = 0.9986$ indica que o ajuste (por mínimos quadrados) do modelo com $\hat{\alpha} = 4.310$ (EP = 0.007), $\hat{\beta} = 0.008$ (EP < 0.001) e $\hat{\gamma} = 0.0001$ (EP < 0.00001) é excelente (sob essa ótica, obviamente). Por outro lado, o gráfico de resíduos apresentado na Figura C.5.8 mostra sequências de resíduos positivos seguidas de sequências de resíduos negativos, sugerindo uma possível correlação positiva entre eles (autocorrelação).

Figura C.5.8: Resíduos studentizados obtidos do ajuste do modelo (C.5.12)



Uma maneira de contornar esse problema, é modificar os componentes aleatórios do modelo para incorporar essa possível autocorrelação nos erros. Nesse contexto, podemos considerar o modelo (C.5.12) com

$$e_t = \rho e_{t-1} + u_t, \quad t = 1, \dots, n \quad (\text{C.5.13})$$

em que $u_t \sim N(0, \sigma^2)$, $t = 1, \dots, n$, independentes e e_0 é uma constante (geralmente igual a zero). Essas suposições implicam que $\text{Var}(e_t) = \sigma^2/(1 - \rho^2)$ e que $\text{Cov}(e_t, e_{t-s}) = \rho^s[\sigma^2/(1 - \rho^2)]$.

Para testar a hipótese de que os erros são não correlacionados pode-se utilizar a **estatística de Durbin-Watson**:

$$D = \sum_{t=2}^n (\hat{e}_t - \hat{e}_{t-1})^2 / \sum_{t=1}^n \hat{e}_t^2, \quad (\text{C.5.14})$$

em que \hat{e}_t , $t = 1, \dots, n$ são os resíduos obtidos do ajuste do modelo (C.5.12) por mínimos quadrados. Expandindo (C.5.14) obtemos

$$\begin{aligned} D &= \frac{\sum_{t=2}^n \hat{e}_t^2}{\sum_{t=1}^n \hat{e}_t^2} + \frac{\sum_{t=2}^n \hat{e}_{t-1}^2}{\sum_{t=1}^n \hat{e}_t^2} - 2 \frac{\sum_{t=2}^n \hat{e}_t \hat{e}_{t-1}}{\sum_{t=1}^n \hat{e}_t^2} \\ &\approx 2 - 2 \frac{\sum_{t=2}^n \hat{e}_t \hat{e}_{t-1}}{\sum_{t=1}^n \hat{e}_t^2}, \end{aligned} \quad (\text{C.5.15})$$

Se os resíduos não forem correlacionados, então $\sum_{t=2}^n \hat{e}_t \hat{e}_{t-1} \approx 0$ e conseqüentemente, $D \approx 2$; se, por outro lado, os resíduos forem altamente correlacionados, esperamos que $\sum_{t=2}^n \hat{e}_t \hat{e}_{t-1} \approx \sum_{t=2}^n \hat{e}_t^2$ e então $D \approx 0$; finalmente, se os resíduos tiverem uma grande correlação negativa, esperamos que $\sum_{t=2}^n \hat{e}_t \hat{e}_{t-1} \approx -\sum_{t=2}^n \hat{e}_t^2$ e nesse caso, $D \approx 4$. Durbin & Watson (1950), Durbin & Watson (1951) e Durbin & Watson (1971) produziram tabelas da distribuição da estatística D que podem ser utilizados para avaliar a suposição de que os erros são não correlacionados.

Se a análise indicar que os erros são correlacionados, o modelo (C.5.12) - (C.5.13) poderá ser ajustado pelo método de mínimos quadrados generalizados. Com esse intuito, notemos primeiramente que (C.5.13) sugere o seguinte modelo de regressão linear simples sem intercepto

$$\hat{e}_t = \rho \hat{e}_{t-1} + u_t$$

de onde podemos obter estimadores de σ^2 e ρ ; mais especificamente,

$$\hat{\rho} = \sum_{t=2}^n \hat{e}_t \hat{e}_{t-1} / \sum_{t=1}^n \hat{e}_t^2$$

e

$$\hat{\sigma}^2 = (n-1)^{-1} \sum_{t=1}^n (\hat{e}_t - \hat{\rho} \hat{e}_{t-1})^2.$$

Expressando o modelo (C.5.12) - (C.5.13) na forma matricial, o vetor de parâmetros

é $\boldsymbol{\beta} = (\alpha, \beta, \gamma)^\top$ e a matriz de covariâncias é

$$\mathbf{V} = \frac{\sigma^2}{1 - \rho^2} \begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^{n-1} \\ \rho & 1 & \rho & \dots & \rho^{n-2} \\ \rho^2 & \rho & 1 & \dots & \rho^{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \dots & 1 \end{bmatrix}.$$

Substituindo os elementos de \mathbf{V} por seus estimadores, podemos utilizar (C.2.6) para estimar $\boldsymbol{\beta}$.

O valor da estatística de Durbin-Watson para os dados do Exemplo C.5.1 sob o modelo (C.5.12) é $D = 0.1259$ ($p < 0.0001$), sugerindo um alto grau de autocorrelação dos resíduos.

Exemplo C.5.2: Na Tabela C.5.3 apresentamos dados provenientes de um estudo em que o objetivo é avaliar a variação do peso de bezerros submetidos a diferentes dietas (tipos de pasto) entre 12 e 26 semanas após o nascimento. Como animais mais pesados (ou mais leves) no início do estudo tendem a permanecer mais pesados (mais leves) ao longo do tempo (pelo menos, ao longo das primeiras observações) é razoável supor que o peso de cada animal numa determinada semana seja correlacionado com seu peso na semana anterior.

Tabela C.5.3: Peso de bezerros (kg)

Semanas após nascimento							
12	14	16	18	20	22	24	26
54.1	65.4	75.1	87.9	98.0	108.7	124.2	131.3
91.7	104.0	119.2	133.1	145.4	156.5	167.2	176.8
64.2	81.0	91.5	106.9	117.1	127.7	144.2	154.9
70.3	80.0	90.0	102.6	101.2	120.4	130.9	137.1
68.3	77.2	84.2	96.2	104.1	114.0	123.0	132.0
43.9	48.1	58.3	68.6	78.5	86.8	99.9	106.2
87.4	95.4	110.5	122.5	127.0	136.3	144.8	151.5
74.5	86.8	94.4	103.6	110.7	120.0	126.7	132.2
50.5	55.0	59.1	68.9	78.2	75.1	79.0	77.0
91.0	95.5	109.8	124.9	135.9	148.0	154.5	167.6
83.3	89.7	99.7	110.0	120.8	135.1	141.5	157.0
76.3	80.8	94.2	102.6	111.0	115.6	121.4	134.5
55.9	61.1	67.7	80.9	93.0	100.1	103.2	108.0
76.1	81.1	84.6	89.8	97.4	111.0	120.2	134.2
56.6	63.7	70.1	74.4	85.1	90.2	96.1	103.6

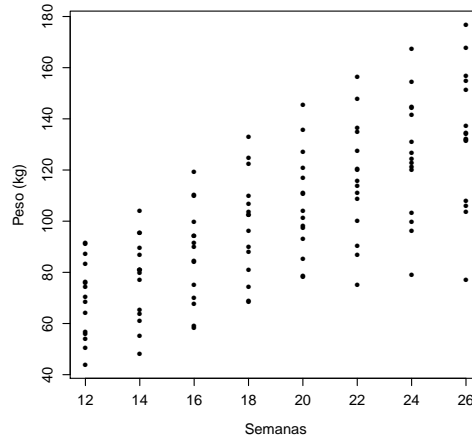
O **gráfico do desenhista** (*draftman's plot*) para os dados do Exemplo C.5.2 disposto na Figura C.5.9 sugere uma variação linear do peso médio dos animais ao longo do período estudado, para a qual podem-se usar modelos do tipo

$$y_{ij} = \alpha + \beta x_{ij} + e_{ij}, \quad i = 1, \dots, n, \quad j = 1, \dots, m_i, \quad (\text{C.5.16})$$

em que y_{ij} denota a j -ésima medida do peso do i -ésimo animal, x_{ij} indica o número de semanas pós-nascimento em que foi realizada essa medida, α representa o peso esperado dos animais ao nascer (admitindo que o modelo linear possa ser extrapolado para o período entre o nascimento e a décima segunda semana), β representa a variação esperada do peso dos animais por semana e e_{ij} corresponde a um erro aleatório com média nula e variância constante σ_e^2 . Esse modelo, particularizado para o Exemplo C.5.2, em todos os animais foram observados nos mesmos instantes, é tal que $n = 15$, $m_i = m = 8$ e $x_{ij} = x_j$.

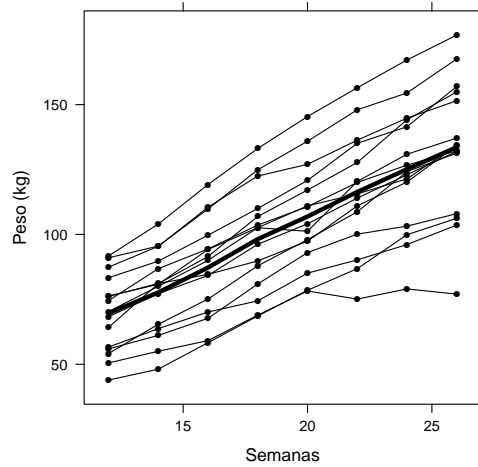
Sob a suposição de que os erros e_{ij} são não correlacionados, estimativas (e erros padrões) dos parâmetros de localização do modelo (C.5.16) obtidas por mínimos quadrados são $\hat{\alpha} = 13.5$ (EP= 7.8), $\hat{\beta} = 4.7$ (EP= 0.4); a estimativa do desvio padrão é $\hat{\sigma}_e = 20.1$ e o coeficiente de determinação ajustado é $R_a^2 = 0.53$. Uma análise de resíduos não indica violações das suposições de heterocedasticidade e normalidade dos erros.

Figura C.5.9: Gráfico de dispersão para os dados do Exemplo C.5.2



Tendo em vista que várias observações são realizadas em cada animal, podemos construir um gráfico de perfis como aquele apresentado na Figura C.5.10. Esse tipo de gráfico serve como ferramenta adicional para avaliação das suposições adotadas.

Figura C.5.10: Gráfico de perfis para os dados do Exemplo C.5.2



Em primeiro lugar, observa-se que não há razões para duvidar da hipótese de homocedasticidade, pois a variabilidade das observações é similar nos oito instantes de observação. Também não há evidências nem de unidades amostrais (animais) com perfis discrepantes nem de observações discrepantes. Além disso, a proposta de uma reta para explicar a variação do peso médio ao longo do período de observação não é contrariada, dada a forma do perfil médio. No entanto, observa-se um certo “paralelismo” dos perfis individuais, sugerindo uma correlação positiva entre as observações intraunidades amostrais; os animais mais pesados (leves) tendem a manter essa característica ao longo das quatorze semanas em que se desenvolveu o estudo.

Para incorporar essa informação no modelo a ser adotado, uma alternativa é substituir o termo aleatório do modelo (C.5.16), fazendo

$$e_{ij} = a_i + d_{ij} \quad (\text{C.5.17})$$

com a_i e d_{ij} denotando variáveis aleatórias não correlacionadas com médias nulas e variâncias σ_a^2 e σ^2 , respectivamente. As suposições adotadas implicam

- a) $\mathbb{V}(y_{ij}) = \sigma_a^2 + \sigma^2$
- b) $\mathbb{Cov}(y_{ij}, y_{kl}) = \sigma_a^2$ se $i = k$ e $j \neq l$
- c) $\mathbb{Cov}(y_{ij}, y_{kl}) = 0$ se $i \neq k$

de forma que observações realizadas na mesma unidade amostral são correlacionadas e aquelas realizadas em unidades amostrais diferentes, não. Embora a suposição de

homocedasticidade possa ser questionável as covariâncias (correlações) podem ser consideradas constantes, como sugerem tanto com a matriz de covariâncias (correlações) amostral disposta na Tabela C.5.4 quanto com a matriz de gráficos de dispersão (*draftman's plot*) apresentada na Figura C.5.11. Isso é uma indicação de que o modelo proposto é uma alternativa razoável para representar os dados.

Figura C.5.11: Gráfico do desenhista para o Exemplo C.5.2

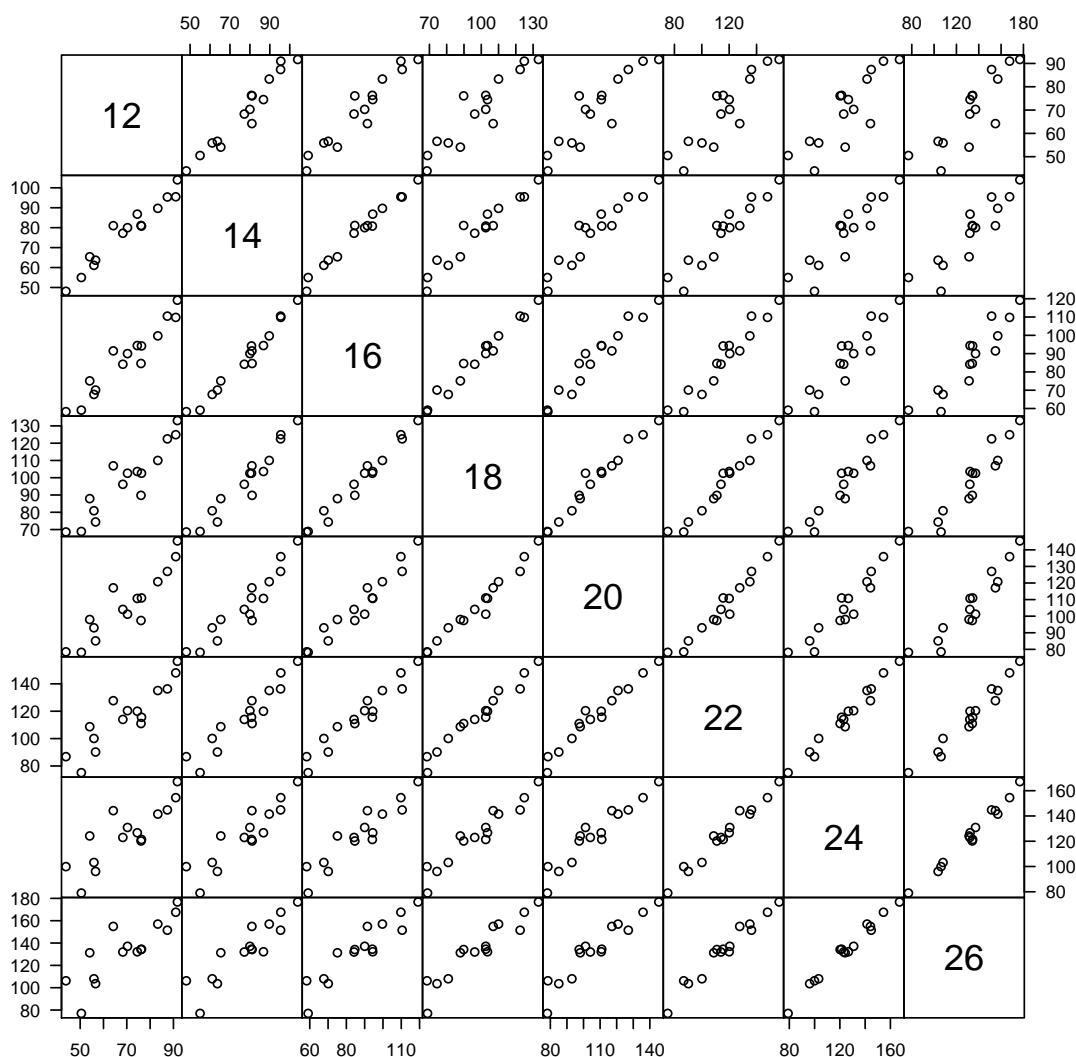


Tabela C.5.4: Matriz de covariâncias (correlações) para os dados do Exemplo C.5.2

Semanas	12	14	16	18	20	22	24	26
12	229.7	0.97	0.96	0.93	0.91	0.90	0.84	0.85
14	236.3	257.2	0.99	0.96	0.94	0.95	0.91	0.90
16	270.1	293.7	344.3	0.99	0.97	0.97	0.93	0.93
18	280.2	309.1	366.4	398.9	0.98	0.98	0.96	0.94
20	274.2	301.8	359.0	392.2	397.5	0.98	0.95	0.94
22	309.1	342.8	405.1	442.9	441.0	509.8	0.99	0.98
24	300.6	343.8	409.7	452.2	449.0	526.3	558.5	0.99
26	341.1	384.2	456.0	498.0	497.3	587.1	622.1	705.5

Essencialmente, o modelo (C.5.16)-(C.5.17) é um modelo misto com um efeito aleatório (a_i) e sugere que o peso esperado do i -ésimo animal varia linearmente com o tempo segundo o modelo condicional

$$y_{ij}|a_i = (\alpha + a_i) + \beta x_{ij} + d_{ij} = \alpha_i + \beta x_{ij} + d_{ij} \quad (\text{C.5.18})$$

i.e., com a mesma taxa de crescimento β , porém com pesos esperados ao nascer, $\alpha_i = \alpha + a_i$, diferentes.

Embora o ajuste de modelos mistos do tipo (C.5.16)-(C.5.17) possa ser facilmente realizado utilizando a metodologia descrita no Capítulo 2, no caso particular em que as unidades amostrais são observadas nos mesmos instantes, *i.e.*, em que $x_{ij} = x_j$ e $m_i = m$, é possível utilizar um enfoque ingênuo baseado na técnica de mínimos quadrados generalizados para concretizá-lo. Nesse contexto, o núcleo do problema é a estimação dos componentes de variância σ_a^2 e σ^2 . Com essa finalidade, convém escrever o modelo (C.5.18) na forma matricial

$$\mathbf{y} = \mathbf{W}\boldsymbol{\gamma} + \mathbf{e},$$

em que $\mathbf{W} = [\mathbf{I}_n \otimes \mathbf{1}_m, \mathbf{1}_n \otimes \mathbf{x}]$, com dimensão $nm \times (n+1)$, é a matriz de variáveis explicativas do modelo condicional, $\mathbf{x} = [x_1, \dots, x_m]^\top$, $\mathbf{1}_a$ é um vetor de ordem a com todos os elementos iguais a 1, $\mathbf{y} = [\mathbf{y}_1^\top, \dots, \mathbf{y}_n^\top]^\top$ com $\mathbf{y}_i = [y_{i1}, \dots, y_{im}]^\top$ e $\boldsymbol{\gamma} = [\alpha_1, \dots, \alpha_n, \beta]^\top$.

Um estimador consistente de σ^2 é

$$\hat{\sigma}^2 = [nm - (n+1)]^{-1} \mathbf{y}^\top [\mathbf{I}_{nm} - \mathbf{W}(\mathbf{W}^\top \mathbf{W})^{-1} \mathbf{W}^\top] \mathbf{y}. \quad (\text{C.5.19})$$

Consideremos agora o modelo

$$\bar{y}_i = \alpha + a_i + \beta \bar{x} + \bar{d}_i \quad (\text{C.5.20})$$

em que, \bar{y}_i e \bar{x} representam respectivamente a média dos m valores da variável resposta e da covariável para a i -ésima unidade amostral e $\bar{d}_i = m^{-1} \sum_{j=1}^m d_{ij}$, $i = 1, \dots, n$. Esse modelo pode ser expresso como

$$\bar{y}_i = \mu + d_i^*$$

em que $\mu = \alpha + \beta\bar{x}$ e $d_i^* = a_i + \bar{d}_i$, $i = 1, \dots, n$ são erros aleatórios independentes com média nula e variância $\tau^2 = \sigma_a^2 + \sigma^2/m$. Sob esse modelo, um estimador consistente de τ^2 é

$$\hat{\tau}^2 = (n-1)^{-1} \sum_{i=1}^n (\bar{y}_i - \bar{y})^2 \quad (\text{C.5.21})$$

com $\bar{y} = n^{-1} \sum_{i=1}^n \bar{y}_i$. Consequentemente, um estimador de σ_a^2 pode ser obtido de (C.5.19) e (C.5.21) por meio de

$$\hat{\sigma}_a^2 = \hat{\tau}^2 - \hat{\sigma}^2/m.$$

Na forma matricial, o modelo (C.5.16)-(C.5.17) pode ser escrito como

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$$

em que $\mathbf{X} = [\mathbf{1}_n \otimes \mathbf{1}_m, \mathbf{1}_n \otimes \mathbf{x}]$, $\boldsymbol{\beta} = (\alpha, \beta)^\top$ e \mathbf{e} é um vetor aleatório com média nula e matriz de covariâncias $\mathbb{V}(\mathbf{e}) = \mathbf{I}_n \otimes \mathbf{R}$ com $\mathbf{R} = \sigma^2 \mathbf{I}_m + \sigma_a^2 \mathbf{1}_m \mathbf{1}_m^\top$. Um estimador consistente, $\hat{\mathbf{R}}$, da matriz de covariâncias intraunidades amostrais \mathbf{R} pode ser obtido com a substituição dos parâmetros σ_a^2 e σ^2 pelos estimadores $\hat{\sigma}_a^2$ e $\hat{\sigma}^2$. O vetor de parâmetros de localização $\boldsymbol{\beta}$ pode ser facilmente estimado por meio de (C.2.6) tendo em vista que, segundo a propriedade *vii*) da Seção A.1.6,

$$\mathbf{R}^{-1} = \frac{1}{\sigma^2} \left[\mathbf{I}_m - \frac{\sigma_a^2}{m\sigma_a^2 + \sigma^2} \mathbf{1}_m \mathbf{1}_m^\top \right].$$

Mais especificamente, pode-se mostrar que, neste caso,

$$\hat{\boldsymbol{\beta}} = \left[\sum_{i=1}^n \mathbf{X}_i \hat{\mathbf{R}}^{-1} \mathbf{X}_i^\top \right]^{-1} \sum_{i=1}^n \mathbf{X}_i \hat{\mathbf{R}}^{-1} \mathbf{y}_i$$

em que $\mathbf{X}_i = [\mathbf{1}_m \ \mathbf{x}]$. Sob as condições de regularidade detalhadas na Seção C.2, concluímos que a distribuição aproximada de $\hat{\boldsymbol{\beta}}$ é dada por (C.2.7).

Para os dados do Exemplo C.5.2, obtemos $\hat{\sigma}^2 = 34.6$, $\hat{\tau}^2 = 393.5$ e $\hat{\sigma}_a^2 = 389.2$, $\hat{\alpha} = 13.5$ (EP= 5.6) e $\hat{\beta} = 4.7$ (EP= 0.1), resultados que coincidem com aqueles gerados pela metodologia de máxima verossimilhança (sob normalidade) descrita no Capítulo 2. Observamos que, em conformidade com os perfis apresentados na Figura C.5.10, a variância interunidades amostrais é cerca de 10 vezes a variância intraunidades amostrais. Além disso, notamos que, embora as estimativas dos coeficientes de regressão α e β sejam iguais àquelas obtidas por mínimos quadrados ordinários, os erros padrões correspondentes são menores, evidenciando o ganho de precisão proporcionado pelo ajuste de um modelo mais adequado.

C.5.5 Multicolinearidade

A maneira mais fácil para avaliar a existência de colinearidade entre duas variáveis explicativas é a construção de gráficos de dispersão e o exame dos coeficientes de correlação linear entre elas. Multicolinearidade (entre várias variáveis explicativas) é mais difícil de ser detectada, mas pode ser avaliada por meio da observação dos seguintes fenômenos:

- i) a inclusão de alguma variável altera consideravelmente as estimativas dos coeficientes de regressão;
- ii) os erros padrões das estimativas dos coeficientes de regressão são grandes;
- iii) as estimativas dos coeficientes de regressão não são significativas mesmo quando se espera uma associação linear entre a variável resposta e as variáveis preditoras;
- iv) as estimativas dos coeficientes de regressão têm sinais diferentes daqueles esperados.

Consideremos o modelo de regressão linear múltipla

$$y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip} + e_i,$$

$i = 1, \dots, n$ com erros e_i não correlacionados e $\mathbb{E}(e_i) = 0$ e $\mathbb{V}(e_i) = \sigma^2$. Se as variáveis X_i forem não correlacionadas, teremos

$$\mathbb{V}(\hat{\beta}_i) = \sigma^2 / n S_{X_i}^2$$

em que $S_{X_i}^2$ é a variância da variável X_i (lembramos que nesse caso $\mathbf{X}^\top \mathbf{X}$ é uma matriz diagonal). No caso geral, *i.e.*, em que as variáveis explicativas podem ser correlacionadas, teremos

$$\mathbb{V}(\hat{\beta}_i) = \sigma^2 [\mathbf{X}^\top \mathbf{X}]_{i+1, i+1}^{-1}.$$

O quociente entre as duas expressões da variância, nomeadamente,

$$FIV_i = n S_{X_i}^2 [\mathbf{X}^\top \mathbf{X}]_{i+1, i+1}^{-1}$$

é o conhecido como **fator de inflação da variância** (*variance inflation factor*) para o i -ésimo coeficiente. Valores grandes de FIV_i sugerem alto grau de multicolinearidade. Pode-se mostrar que

$$FIV_i = 1/(1 - R_i^2) \geq 1$$

em que R_i^2 é o coeficiente de determinação correspondente à regressão que tem X_i como variável resposta e as demais variáveis como explicativas. FIV_i aumenta com o aumento da correlação entre X_i e alguma combinação linear das demais variáveis explicativas. Como medida resumo, é comum calcular o fator de inflação da variância médio, $\overline{FIV} = \sum_{i=1}^p FIV_i/p$; valores de $\overline{FIV} \gg 1$ sugerem problemas sérios de multicolinearidade.

Para contornar esses problemas, algumas sugestões envolvem

- i) utilizar variáveis centradas em modelos de regressão polinomial;
- ii) eliminar algumas variáveis explicativas que sejam correlacionadas com as demais;
- iii) quando possível, adicionar observações que permitam romper a estrutura de correlação entre as variáveis explicativas.
- iv) substituir as variáveis originais por **componentes principais**.

Detalhes podem ser obtidos em Kutner, Nachtsheim, Neter & Li (2005).

C.6 Parametrização de modelos lineares

Consideremos um estudo em que se deseja comparar o “efeito” de I tratamentos (drogas, por exemplo) na distribuição de uma variável resposta Y (pressão diastólica, por exemplo). Há muitas situações práticas em que esse “efeito” consiste na modificação do valor esperado da resposta sem alteração da forma de sua distribuição ou de sua variância. Nesse contexto, supondo que m unidades amostrais escolhidas aleatoriamente sejam submetidas a cada um dos I tratamentos, um modelo bastante comum é

$$y_{ij} = \mu_i + e_{ij}, \quad (\text{C.6.1})$$

$i = 1, \dots, I$, $j = 1, \dots, m$, em que $\mathbb{E}(e_{ij}) = 0$, $\mathbb{V}(e_{ij}) = \sigma^2$ e $\mathbb{E}(e_{ij}e_{kl}) = 0$, $i \neq k$, $j \neq l$ ou seja, os e_{ij} são não correlacionados. Esta parametrização do modelo é conhecida como **parametrização de médias de celas** pois o parâmetro de localização μ_i pode ser interpretado como o valor esperado (médio) da resposta de unidades amostrais submetidas ao tratamento i . O “efeito” do i -ésimo tratamento é definido como uma função desses valores esperados; por exemplo, podemos definir o efeito do tratamento i em relação ao tratamento I como a diferença $\mu_i - \mu_I$.

Para permitir que a definição de “efeito” seja incorporada diretamente nos parâmetros do modelo, é comum utilizarem-se outras parametrizações. Muitos autores

sugerem que se escreva $\mu_i = \mu + \alpha_i$, o que implica o modelo

$$y_{ij} = \mu + \alpha_i + e_{ij}, \quad (\text{C.6.2})$$

$i = 1, \dots, I$, $j = 1, \dots, m$, e interpretam o parâmetro μ como “média geral” e α_i como o “efeito” do tratamento i . O problema é que esse modelo é **inidentificável**¹² e tanto μ quanto α_i nem podem ser interpretados dessa forma nem são **estimáveis**¹³. Nesses casos, é possível trabalhar com funções estimáveis (e em geral, aquelas nas quais se tem interesse, o são), mas a ideia de acomodar parâmetros interpretáveis no modelo não é concretizada. Uma possível solução para esse problema consiste na adoção de **restrições de identificabilidade** que não só implicam a identificabilidade do modelo como a estimabilidade de seus parâmetros. Dentre elas, as mais utilizadas na prática são

$$\sum_{i=1}^I \alpha_i = 0, \quad (\text{C.6.3})$$

que induz a chamada **parametrização de desvios de médias** e

$$\alpha_1 = 0, \quad (\text{C.6.4})$$

que induz a chamada **parametrização de cela de referência**. Definindo $\bar{y} = (Im)^{-1} \sum_{i=1}^I \sum_{j=1}^m y_{ij}$ e $\bar{y}_i = m^{-1} \sum_{j=1}^m y_{ij}$, e utilizando (C.6.3), obtemos

$$\mathbb{E}(\bar{y}) = (Im)^{-1} \sum_{i=1}^I \sum_{j=1}^m \mathbb{E}(y_{ij}) = \mu + I^{-1} \sum_{i=1}^I \alpha_i = \mu,$$

e consequentemente o termo μ pode ser interpretado como média geral (que essencialmente é uma média dos valores esperados da resposta associados aos I tratamentos). Além disso,

$$\mathbb{E}(\bar{y}_i) = m^{-1} \sum_{j=1}^m \mathbb{E}(y_{ij}) = \mu + \alpha_i$$

¹²Um modelo $F(\theta)$, dependendo do parâmetro $\theta \in \Theta$, é identificável se para todo $\theta_1, \theta_2 \in \Theta$, $\theta_1 \neq \theta_2$ temos $F(\theta_1) \neq F(\theta_2)$. Em caso contrário, o modelo é dito inidentificável. Por exemplo, consideremos o modelo $y_i \sim N(\mu + \alpha_i, \sigma^2)$, $i = 1, 2$ em que y_1 e y_2 são independentes. Tomando $\theta = (\mu, \alpha_1, \alpha_2)^\top$ como parâmetro, o modelo é inidentificável, pois tanto para $\theta_1 = (5, 1, 0)^\top$ quanto para $\theta_2 = (4, 2, 1)^\top \neq \theta_1$, a distribuição conjunta de (y_1, y_2) é $N_2[(6, 6)^\top, \sigma^2 \mathbf{I}_2]$. O leitor poderá consultar Bickel & Doksum (2001), entre outros, para detalhes.

¹³Uma função linear de um vetor de parâmetros θ , nomeadamente, $\mathbf{c}^\top \theta$, é estimável se ela for identicamente igual a uma combinação linear do valor esperado do vetor de observações, \mathbf{y} , *i.e.*, se existir um vetor \mathbf{t} tal que $\mathbf{c}^\top \theta = \mathbf{t}^\top \mathbb{E}(\mathbf{y})$. No modelo (C.6.2), nem μ nem α_i , $i = 1, \dots, I$ são funções estimáveis, embora tanto $\mu + \alpha_i$, $i = 1, \dots, I$ quanto $\alpha_i - \alpha_k$, $i \neq k$ o sejam. O leitor poderá consultar Searle (1971), entre outros, para detalhes.

o que implica que $\alpha_i = \mathbb{E}(\bar{y}_i) - \mu$, *i.e.*, a diferença entre o valor esperado das observações submetidas ao tratamento i e a média geral μ . Essa diferença pode ser interpretada como o efeito do tratamento i .

Se utilizarmos a restrição de identificabilidade (C.6.4), obtemos

$$\mathbb{E}(\bar{y}_1) = m^{-1} \sum_{j=1}^m E(y_{1j}) = \mu + \alpha_1 = \mu$$

e consequentemente, o termo μ pode ser interpretado como o valor esperado das observações submetidas ao tratamento 1. Além disso, para $k \neq 1$,

$$\mathbb{E}(\bar{y}_k) = m^{-1} \sum_{j=1}^m E(y_{kj}) = \mu + \alpha_k$$

o que sugere que α_k , $k \neq 1$, neste caso correspondendo à diferença entre o valor esperado das observações submetidas ao tratamento k e o valor esperado das observações submetidas ao tratamento 1, tomado como referência, pode ser interpretado como o efeito do tratamento k . Obviamente, qualquer dos tratamentos pode servir de referência, bastando para isso, modificar a restrição de identificabilidade convenientemente. Em geral, quando existe um tratamento controle, é ele que serve de referência.

Voltemos agora nossa atenção para estudos em que se deseja avaliar o efeito de dois fatores A (droga, por exemplo) e B (faixa etária, por exemplo), o primeiro com a e o segundo com b níveis, na distribuição de uma resposta (pressão diastólica, por exemplo). Admitamos que m unidades amostrais tenham sido observadas para cada tratamento, *i.e.*, para cada combinação dos níveis dos fatores A e B . Com base nos mesmos argumentos utilizados no caso anterior, suponhamos que o “efeito” de cada um dos fatores e sua interação possa ser definido apenas em termos dos valores esperados das distribuições da resposta sob os diferentes tratamentos. Um modelo comumente considerado para análise de dados com essa estrutura é

$$y_{ijk} = \mu_{ij} + e_{ijk}, \quad (\text{C.6.5})$$

$i = 1, \dots, a$, $j = 1, \dots, b$, $k = 1, \dots, m$, em que $\mathbb{E}(e_{ijk}) = 0$, $\mathbb{V}(e_{ijk}) = \sigma^2$ e $\mathbb{E}(e_{ijk}e_{i'j'k'}) = 0$, $i \neq i'$ ou $j \neq j'$ ou $k \neq k'$, ou seja, os e_{ijk} são não correlacionados. Esta é a parametrização de médias de celas pois o parâmetro de localização μ_{ij} pode ser interpretado como o valor esperado (médio) da resposta de unidades amostrais submetidas ao tratamento correspondente ao nível i do fator A e nível j do fator B .

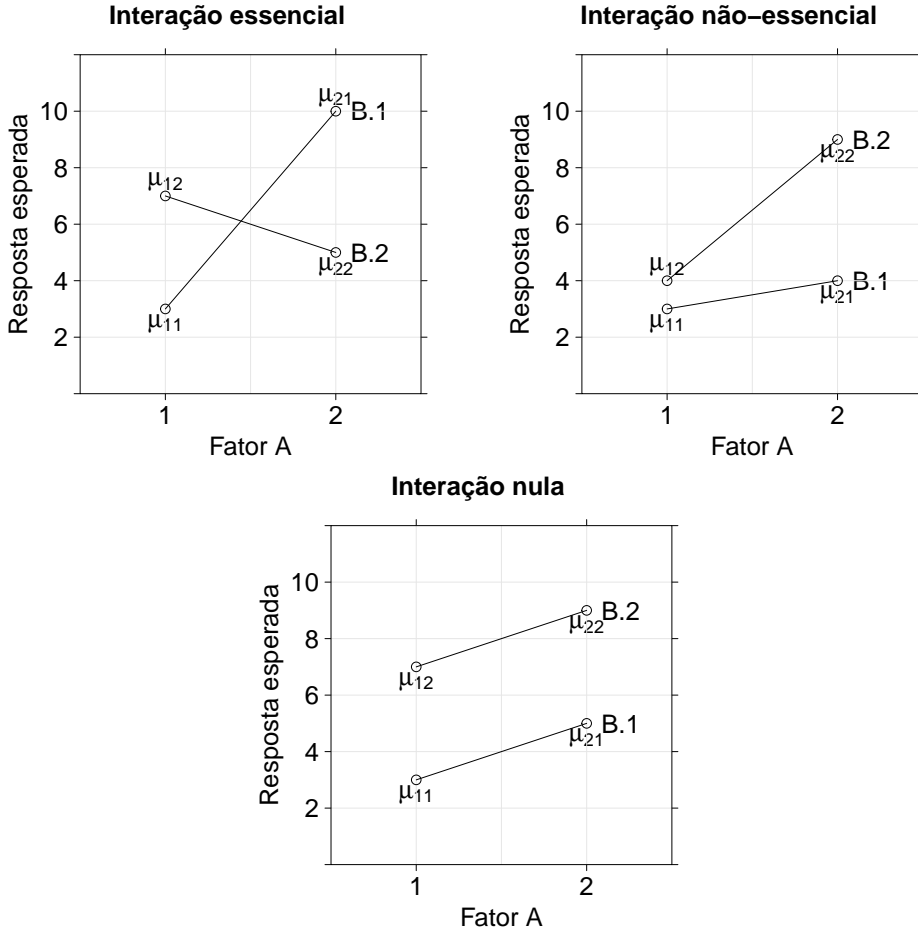
Tomando $a = b = 2$ para facilidade de exposição, o “efeito” do fator A para unidades amostrais no nível j do fator B pode ser definido como a diferença $\mu_{1j} - \mu_{2j}$, que, por exemplo, corresponde à diferença entre o valor esperado da pressão

diastólica de unidades amostrais na faixa etária j submetidas à droga 1 e o valor esperado da pressão diastólica de unidades amostrais na faixa etária j submetidas à droga 2. Analogamente, o “efeito” do fator B para unidades amostrais no nível i do fator A pode ser definido como a diferença $\mu_{i1} - \mu_{i2}$.

A interação entre os fatores A e B pode ser definida como a diferença entre o efeito do fator A para unidades amostrais no nível 1 do fator B e o efeito do fator A para unidades amostrais no nível 2 do fator B , nomeadamente, $(\mu_{11} - \mu_{21}) - (\mu_{12} - \mu_{22})$. Outras definições equivalentes, como $(\mu_{22} - \mu_{12}) - (\mu_{21} - \mu_{11})$ podem ser utilizadas. A escolha entre as alternativas deve ser feita em função dos detalhes do problema; por exemplo, se a droga 1 for uma droga padrão e a faixa etária 1 corresponder a indivíduos mais jovens, esta última proposta pode ser mais conveniente.

Quando a interação é nula, o efeito do fator A é o mesmo para unidades amostrais submetidas a qualquer dos níveis do fator B e pode-se definir o **efeito principal** do fator A como $(\mu_{11} + \mu_{12})/2 - (\mu_{21} + \mu_{22})/2$, que corresponde à diferença entre o valor esperado da resposta para unidades amostrais submetidas ao nível 1 do fator A e o valor esperado da resposta para unidades amostrais submetidas ao nível 2 do fator A (independentemente do nível do fator B). Similarmente, o efeito principal do fator B pode ser definido como $(\mu_{11} + \mu_{21})/2 - (\mu_{12} + \mu_{22})/2$. Em muitos casos, essas definições de efeitos principais podem ser consideradas mesmo na presença de interação, desde que ela seja não essencial. A interação entre os fatores A e B é não essencial quando as diferenças $\mu_{11} - \mu_{21}$ e $\mu_{12} - \mu_{22}$ têm o mesmo sinal, mas magnitudes diferentes. Por exemplo, se $\mu_{11} - \mu_{21} = K_1 > 0$ e $\mu_{12} - \mu_{22} = K_2 > 0$ com $K_1 \neq K_2$, a resposta esperada sob o nível 1 do fator A é maior que a resposta esperada sob o nível 2 do fator A tanto no nível 1 quanto no nível 2 do fator B , embora as magnitudes das diferenças não sejam iguais. Se essas magnitudes tiverem sinais diferentes, a interação é essencial. Por outro lado, se $K_1 = K_2$, não há interação. O leitor pode consultar Lencina et al. (2005) e Lencina et al. (2008) para uma discussão sobre a consideração de efeitos principais em situações com interação não essencial. Na Figura C.6.1 apresentamos gráficos de perfis médios com interações essencial e não essencial.

Figura C.6.1: Gráfico de perfis médios com diferentes tipos de interação



Com a finalidade de explicitar efeitos principais e interação no modelo, é comum considerar-se a reparametrização $\mu_{ij} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij}$, o que implica o modelo

$$y_{ij} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + e_{ijk}, \quad (\text{C.6.6})$$

$i = 1, \dots, a$, $j = 1, \dots, b$, $k = 1, \dots, m$. Muitos autores, como Nelder (1998), interpretam erroneamente os parâmetros μ , α_i , β_j , $\alpha\beta_{ij}$ como “média geral”, “efeito principal do nível i do fator A ”, “efeito principal do nível j do fator B ” e “interação entre os níveis i do fator A e j do fator B ”. Como no caso discutido acima, esse modelo também é inidentificável e seus parâmetros são não estimáveis e as restrições de identificabilidade mais frequentemente utilizadas e correspondentes às parametrizações de desvios de médias e ceta de referência são, respectivamente,

$$\sum_{i=1}^a \alpha_i = \sum_{j=1}^b \beta_j = \sum_{i=1}^a \alpha\beta_{ij} = \sum_{j=1}^b \alpha\beta_{ij} = 0 \quad (\text{C.6.7})$$

e

$$\alpha_1 = \beta_1 = \alpha\beta_{11} = \dots = \alpha\beta_{1b} = \alpha\beta_{21} = \dots = \alpha\beta_{a1} = 0 \quad (\text{C.6.8})$$

Sob as restrições (C.6.7), pode-se mostrar que

$$\mu = (ab)^{-1} \sum_{i=1}^a \sum_{j=1}^b \mu_{ij}, \quad \alpha_i = b^{-1} \sum_{j=1}^b \mu_{ij} - \mu, \quad \beta_j = a^{-1} \sum_{i=1}^a \mu_{ij} - \mu$$

e que

$$\alpha\beta_{ij} = \mu_{ij} - b^{-1} \sum_{j=1}^b \mu_{ij} - a^{-1} \sum_{i=1}^a \mu_{ij}$$

de forma que esses parâmetros podem ser interpretados como se desejava inicialmente. Sob as restrições (C.6.8), temos

$$\mu = \mu_{11}, \quad \alpha_i = \mu_{ij} - \mu_{1j}, \quad i = 2, \dots, a, \quad \beta_j = \mu_{ij} - \mu_{i1}, \quad j = 2, \dots, b,$$

e que

$$\alpha\beta_{ij} = \mu_{ij} - (\mu_{11} + \alpha_i + \beta_j), \quad i = 2, \dots, a, \quad j = 2, \dots, b,$$

de forma que os parâmetros α_i , $i = 2, \dots, a$ podem ser interpretados como efeitos diferenciais entre as respostas esperadas das unidades amostrais submetidas ao nível i do fator A relativamente às obtidas por unidades amostrais submetidas ao tratamento associado ao nível 1 do fator A , mantido fixo o nível correspondente ao fator B . Analogamente, os parâmetros β_j , $j = 2, \dots, b$ podem ser interpretados como efeitos diferenciais entre as respostas esperadas das unidades amostrais submetidas ao nível j do fator B relativamente às obtidas por unidades amostrais submetidas ao tratamento associado ao nível 1 do fator B , mantido fixo o nível correspondente do fator A . Os parâmetros $\alpha\beta_{ij}$, $i = 2, \dots, a$, $j = 2, \dots, b$ podem ser interpretados como diferenças entre as respostas esperadas das unidades amostrais submetidas ao tratamento correspondente à cela (i, j) e aquela esperada sob um modelo sem interação.

C.7 Regressão logística

Os dados da Tabela C.7.1 são extraídos de um estudo realizado no Hospital Universitário da Universidade de São Paulo com o objetivo de avaliar se algumas medidas obtidas ultrassonograficamente poderiam ser utilizadas como substitutas de medidas obtidas por métodos de ressonância magnética, considerada como padrão áureo, para avaliação do deslocamento do disco da articulação temporomandibular (doravante referido simplesmente como disco). Distâncias cápsula-côndilo (em mm) com boca

aberta ou fechada (referidas, respectivamente, como distância aberta ou fechada no restante do texto) foram obtidas ultrassonograficamente de 104 articulações e o disco correspondente foi classificado como deslocado (1) ou não (0) segundo a avaliação por ressonância magnética. A variável resposta é o *status* do disco (1 = deslocado ou 0 = não). Mais detalhes podem ser obtidos em Elias, Birman, Matsuda, Oliveira & Jorge (2006).

A diferença fundamental entre este problema e os demais abordados neste apêndice é a natureza da variável resposta, que é discreta em vez de contínua. Denotando a resposta da i -ésima unidade amostral por y_i , podemos definir $y_i = 1$ se houve deslocamento do disco e $y_i = 0$, em caso contrário. Admitindo que para a i -ésima unidade amostral, o disco pode estar deslocado com probabilidade θ_i , temos $P(y_i = 1) = \theta_i$ e $P(y_i = 0) = 1 - \theta_i$, de maneira que $\mathbb{E}(y_i) = \theta_i$ e $\mathbb{V}(y_i) = \theta_i(1 - \theta_i)$.

O objetivo da análise é modelar $\mathbb{E}(y_i)$ como função das variáveis explicativas. O modelo (linear) correspondente é

$$y_i = \mathbf{x}_i^\top \boldsymbol{\beta} + \varepsilon_i, \quad (\text{C.7.1})$$

$i = 1, \dots, n$ em que \mathbf{x}_i^\top é a i -ésima linha da matriz \mathbf{X} , que contém os valores p variáveis explicativas para as n unidades amostrais. Em notação matricial, o modelo pode ser escrito como

$$\mathbb{E}(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta} \quad \text{com} \quad \mathbb{V}(\mathbf{y}) = n^{-1} \text{diag}[\theta_1(1 - \theta_1), \dots, \theta_n(1 - \theta_n)]$$

em que $\mathbf{y} = (y_1, \dots, y_n)^\top$, \mathbf{X} e $\boldsymbol{\beta}$ têm as interpretações usuais. Tanto o problema é claramente heterocedástico quanto a distribuição da variável resposta é claramente não gaussiana. No entanto, ele tem a mesma estrutura daqueles que podem ser analisados por intermédio de métodos de mínimos quadrados ponderados, desde que estimativas consistentes dos parâmetros θ_i estejam disponíveis, o que em geral, não é verdade. Uma situação na qual essa opção pode ser aplicada é aquela em que todas as variáveis explicativas são discretas. Detalhes sobre análises específicas para esse caso podem ser encontrados em Paulino & Singer (2006).

O método de máxima verossimilhança é uma alternativa para o ajuste desses modelos no caso mais geral (em que as variáveis explicativas podem ter natureza discreta ou contínua). A função de verossimilhança a ser maximizada é

$$L(\boldsymbol{\beta}|\mathbf{y}, \mathbf{X}) = \prod_{i=1}^n (\mathbf{x}_i^\top \boldsymbol{\beta})^{y_i} (1 - \mathbf{x}_i^\top \boldsymbol{\beta})^{1-y_i}.$$

Com essa finalidade, podemos considerar seu logaritmo,

$$l(\boldsymbol{\beta}|\mathbf{y}, \mathbf{X}) = \sum_{i=1}^n [y_i \log(\mathbf{x}_i^\top \boldsymbol{\beta}) + (1 - y_i) \log(1 - \mathbf{x}_i^\top \boldsymbol{\beta})],$$

Tabela C.7.1: Dados de um estudo odontológico

Dist aberta	Dist fechada	Desloc disco	Dist aberta	Dist fechada	Desloc disco	Dist aberta	Dist fechada	Desloc disco
2.2	1.4	0	0.9	0.8	0	1.0	0.6	0
2.4	1.2	0	1.1	0.9	0	1.6	1.3	0
2.6	2.0	0	1.4	1.1	0	4.3	2.3	1
3.5	1.8	1	1.6	0.8	0	2.1	1.0	0
1.3	1.0	0	2.1	1.3	0	1.6	0.9	0
2.8	1.1	1	1.8	0.9	0	2.3	1.2	0
1.5	1.2	0	2.4	0.9	0	2.4	1.3	0
2.6	1.1	0	2.0	2.3	0	2.0	1.1	0
1.2	0.6	0	2.0	2.3	0	1.8	1.2	0
1.7	1.5	0	2.4	2.9	0	1.4	1.9	0
1.3	1.2	0	2.7	2.4	1	1.5	1.3	0
1.2	1.0	0	1.9	2.7	1	2.2	1.2	0
4.0	2.5	1	2.4	1.3	1	1.6	2.0	0
1.2	1.0	0	2.1	0.8	1	1.5	1.1	0
3.1	1.7	1	0.8	1.3	0	1.2	0.7	0
2.6	0.6	1	0.8	2.0	1	1.5	0.8	0
1.8	0.8	0	0.5	0.6	0	1.8	1.1	0
1.2	1.0	0	1.5	0.7	0	2.3	1.6	1
1.9	1.0	0	2.9	1.6	1	1.2	0.4	0
1.2	0.9	0	1.4	1.2	0	1.0	1.1	0
1.7	0.9	1	3.2	0.5	1	2.9	2.4	1
1.2	0.8	0	1.2	1.2	0	2.5	3.3	1
3.9	3.2	1	2.1	1.6	1	1.4	1.1	0
1.7	1.1	0	1.4	1.5	1	1.5	1.3	0
1.4	1.0	0	1.5	1.4	0	0.8	2.0	0
1.6	1.3	0	1.6	1.5	0	2.0	2.1	0
1.3	0.5	0	4.9	1.2	1	3.1	2.2	1
1.7	0.7	0	1.1	1.1	0	3.1	2.1	1
2.6	1.8	1	2.0	1.3	1	1.7	1.2	0
1.5	1.5	0	1.5	2.2	0	1.6	0.5	0
1.8	1.4	0	1.7	1.0	0	1.4	1.1	0
1.2	0.9	0	1.9	1.4	0	1.6	1.0	0
1.9	1.0	0	2.5	3.1	1	2.3	1.6	1
2.3	1.0	0	1.4	1.5	0	2.2	1.8	1
1.6	1.0	0	2.5	1.8	1			

Dist aberta: distância cápsula-côndilo com boca aberta (mm)

Dist fechada: distância cápsula-côndilo com boca fechada (mm)

Desloc disco: deslocamento do disco da articulação temporomandibular (1=sim, 0=não)

cujas primeira e segunda derivadas são, respectivamente,

$$\frac{\partial}{\partial \boldsymbol{\beta}} l(\boldsymbol{\beta} | \mathbf{y}, \mathbf{X}) = \mathbf{U}(\boldsymbol{\beta}) = \sum_{i=1}^n \frac{y_i - \mathbf{x}_i^\top \boldsymbol{\beta}}{\mathbf{x}_i^\top \boldsymbol{\beta} (1 - \mathbf{x}_i^\top \boldsymbol{\beta})} \mathbf{x}_i,$$

e

$$\frac{\partial^2}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} l(\boldsymbol{\beta} | \mathbf{y}, \mathbf{X}) = \mathbf{H}(\boldsymbol{\beta}) = - \sum_{i=1}^n \frac{(\sqrt{y_i} - \mathbf{x}_i^\top \boldsymbol{\beta})^2}{[\mathbf{x}_i^\top \boldsymbol{\beta} (1 - \mathbf{x}_i^\top \boldsymbol{\beta})]^2} \mathbf{x}_i \mathbf{x}_i^\top.$$

As equações de estimação correspondentes são

$$\mathbf{U}(\boldsymbol{\beta}) = \sum_{i=1}^n \frac{y_i - \mathbf{x}_i^\top \boldsymbol{\beta}}{\mathbf{x}_i^\top \boldsymbol{\beta} (1 - \mathbf{x}_i^\top \boldsymbol{\beta})} \mathbf{x}_i = \mathbf{0}$$

e o estimador de máxima verossimilhança de $\boldsymbol{\beta}$ pode ser obtido por meio do algoritmo de Newton-Raphson, ou seja, iterando

$$\boldsymbol{\beta}^{(l)} = \boldsymbol{\beta}^{(l-1)} - [\mathbf{H}(\boldsymbol{\beta}^{(l-1)})]^{-1} \mathbf{U}(\boldsymbol{\beta}^{(l-1)}), \quad l = 1, 2, \dots \quad (\text{C.7.2})$$

até que $\|\boldsymbol{\beta}^{(l)} - \boldsymbol{\beta}^{(l-1)}\| < \varepsilon$ com $\varepsilon > 0$ e $\boldsymbol{\beta}^{(0)}$ é um valor inicial, arbitrário. O valor $\hat{\boldsymbol{\beta}}$ obtido na convergência é o estimador de máxima verossimilhança de $\boldsymbol{\beta}$.

O problema com o modelo (C.7.1) é que ele não garante que as estimativas $\mathbf{x}_i^\top \hat{\boldsymbol{\beta}}$ das probabilidades θ_i fiquem restritas ao intervalo $(0, 1)$. Uma maneira de evitar esse problema é apelar para um modelo não linear do tipo $\theta_i = \theta(\mathbf{x}_i; \boldsymbol{\beta}) = F^{-1}(\mathbf{x}_i^\top \boldsymbol{\beta})$ em que F é uma função distribuição, *i.e.*, com imagem no intervalo desejado. Em particular, a função distribuição logística,

$$F(x) = [1 + \exp(-x)]^{-1}, \quad x \in \mathbb{R}$$

é uma candidata com excelentes propriedades. Nesse contexto, o modelo conhecido como **regressão logística** é

$$\theta(\mathbf{x}_i; \boldsymbol{\beta}) = \frac{\exp(\mathbf{x}_i^\top \boldsymbol{\beta})}{1 + \exp(\mathbf{x}_i^\top \boldsymbol{\beta})} \quad (\text{C.7.3})$$

ou, equivalentemente,

$$\log[\theta(\mathbf{x}_i; \boldsymbol{\beta})] / [1 - \theta(\mathbf{x}_i; \boldsymbol{\beta})] = \mathbf{x}_i^\top \boldsymbol{\beta}. \quad (\text{C.7.4})$$

Os termos $\log[\theta/(1 - \theta)]$ são conhecidos como **logitos** (*logits*).

Para efeito de interpretação, consideremos o seguinte modelo de regressão logística com apenas uma variável explicativa,

$$\log[\theta(x; \alpha, \beta)] / [1 - \theta(x; \alpha, \beta)] = \alpha + \beta x.$$

Então, $\alpha = \log[\theta(0; \alpha, \beta)]/[1 - \theta(0; \alpha, \beta)]$ e $\exp(\alpha)$ pode ser interpretado como a **chance** (*odds*) de resposta $Y = 1$ *versus* $Y = 0$ para unidades amostrais com valor da variável explicativa $x = 0$. Por outro lado, $\beta = \log[\theta(x+1; \alpha, \beta)]/[1 - \theta(x+1; \alpha, \beta)] - \log[\theta(x; \alpha, \beta)]/[1 - \theta(x; \alpha, \beta)]$ e então $\exp(\beta)$ pode ser interpretado como a **razão de chances** (*odds ratio*) correspondente a unidades amostrais com valor da variável explicativa $x+1$ relativamente a unidades amostrais com valor da variável explicativa x .

A função de verossimilhança correspondente é

$$L(\boldsymbol{\beta}|\mathbf{y}, \mathbf{X}) = \prod_{i=1}^n \left[\frac{\exp(\mathbf{x}_i^\top \boldsymbol{\beta})}{1 + \exp(\mathbf{x}_i^\top \boldsymbol{\beta})} \right]^{y_i} \left[\frac{1}{1 + \exp(\mathbf{x}_i^\top \boldsymbol{\beta})} \right]^{1-y_i}$$

e seu logaritmo é

$$l(\boldsymbol{\beta}|\mathbf{y}, \mathbf{X}) = \sum_{i=1}^n [y_i \mathbf{x}_i^\top \boldsymbol{\beta} - \log(1 + \mathbf{x}_i^\top \boldsymbol{\beta})],$$

com primeira e segunda derivadas dadas, respectivamente, por

$$\frac{\partial}{\partial \boldsymbol{\beta}} l(\boldsymbol{\beta}|\mathbf{y}, \mathbf{X}) = \mathbf{U}(\boldsymbol{\beta}) = \sum_{i=1}^n [y_i - \frac{\exp(\mathbf{x}_i^\top \boldsymbol{\beta})}{1 + \exp(\mathbf{x}_i^\top \boldsymbol{\beta})}] \mathbf{x}_i = \sum_{i=1}^n [y_i - \theta(\mathbf{x}_i; \boldsymbol{\beta})] \mathbf{x}_i$$

e

$$\begin{aligned} \frac{\partial^2}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} l(\boldsymbol{\beta}|\mathbf{y}, \mathbf{X}) &= \mathbf{H}(\boldsymbol{\beta}) = - \sum_{i=1}^n \frac{\exp(\mathbf{x}_i^\top \boldsymbol{\beta})}{[1 + \exp(\mathbf{x}_i^\top \boldsymbol{\beta})]^2} \mathbf{x}_i \mathbf{x}_i^\top \\ &= - \sum_{i=1}^n \theta(\mathbf{x}_i; \boldsymbol{\beta}) [1 - \theta(\mathbf{x}_i; \boldsymbol{\beta})] \mathbf{x}_i \mathbf{x}_i^\top. \end{aligned}$$

O estimador de máxima verossimilhança de $\boldsymbol{\beta}$ é a solução $\hat{\boldsymbol{\beta}}$ das equações de verossimilhança

$$\mathbf{U}(\boldsymbol{\beta}) = \mathbf{X}^\top [\mathbf{y} - \boldsymbol{\theta}(\boldsymbol{\beta})] = \mathbf{0} \quad (\text{C.7.5})$$

em que $\boldsymbol{\theta}(\boldsymbol{\beta}) = [\theta(\mathbf{x}_1; \boldsymbol{\beta}), \dots, \theta(\mathbf{x}_n; \boldsymbol{\beta})]^\top$. Observando que $\mathbf{H}(\boldsymbol{\beta}) = -\mathbf{X}^\top \mathbf{W}[\boldsymbol{\theta}(\boldsymbol{\beta})] \mathbf{X}$ em que

$$\mathbf{W}[\boldsymbol{\theta}(\boldsymbol{\beta})] = \text{diag}\{\theta(\mathbf{x}_1; \boldsymbol{\beta})[1 - \theta(\mathbf{x}_1; \boldsymbol{\beta})], \dots, \theta(\mathbf{x}_n; \boldsymbol{\beta})[1 - \theta(\mathbf{x}_n; \boldsymbol{\beta})]\},$$

pode-se demonstrar que

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{\mathcal{D}} N_p\{\mathbf{0}, [\mathbf{I}(\boldsymbol{\beta})]^{-1}\}$$

com $\mathbf{I}(\boldsymbol{\beta}) = -\mathbb{E}[\mathbf{H}(\boldsymbol{\beta})] = -\mathbb{E}[\frac{\partial^2}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} l(\boldsymbol{\beta}|\mathbf{y}, \mathbf{X})] = \mathbf{X}^\top \mathbf{W}(\boldsymbol{\theta}) \mathbf{X}$ denotando a matriz de informação de Fisher. Em termos práticos, isto significa que para n suficientemente grande

$$\hat{\boldsymbol{\beta}} \approx N_p\{\boldsymbol{\beta}, (\mathbf{X}^\top \mathbf{W}[\boldsymbol{\theta}(\hat{\boldsymbol{\beta})}] \mathbf{X})^{-1}\}.$$

Em notação matricial, as equações de estimação, obtidas por meio de uma aproximação de Taylor de primeira ordem de (C.7.5), podem ser escritas como

$$\hat{\beta} \approx \beta - \{\mathbf{X}^\top \mathbf{W}[\boldsymbol{\theta}(\beta)] \mathbf{X}\}^{-1} \mathbf{U}[\boldsymbol{\theta}(\beta)]$$

sugerindo o seguinte algoritmo de Newton-Raphson, que neste caso coincide com o algoritmo “Scoring” / de Fisher,

$$\beta^{(l)} = \beta^{(l-1)} - \{\mathbf{X}^\top \mathbf{W}[\boldsymbol{\theta}(\beta^{(l-1)})] \mathbf{X}\}^{-1} \mathbf{X}^\top \mathbf{W}[\boldsymbol{\theta}(\beta^{(l-1)})][\mathbf{y} - \boldsymbol{\theta}(\beta^{(l-1)})], \quad l = 1, 2, \dots$$

Fazendo $\mathbf{z} = \mathbf{X}^\top \beta^{(l-1)} + \{\mathbf{W}[\boldsymbol{\theta}(\beta^{(l-1)})]\}^{-1}[\mathbf{y} - \boldsymbol{\theta}(\beta^{(l-1)})]$, o algoritmo pode ser escrito como

$$\beta^{(l)} = \{\mathbf{X}^\top \mathbf{W}[\boldsymbol{\theta}(\beta^{(l-1)})] \mathbf{X}\}^{-1} \mathbf{X}^\top \mathbf{W}[\boldsymbol{\theta}(\beta^{(l-1)})] \mathbf{z}, \quad l = 1, 2, \dots$$

Em cada passo, esse algoritmo tem uma estrutura equivalente à solução de mínimos quadrados ponderados com pesos explicitados em $\mathbf{W}[\boldsymbol{\theta}(\beta^{(l-1)})]$ e pseudo-variáveis \mathbf{z} , ambos recalculados em cada iteração. Por isso é conhecido como algoritmo de **mínimos quadrados iterativamente ponderados** (*iteratively reweighted least squares*). Mais detalhes podem ser obtidos em Sen et al. (2009), lembrando que o modelo de regressão logística é um caso particular dos chamados **modelos lineares generalizados**.

Uma das características do modelo de regressão logística é que ele permite a estimação das probabilidades de sucesso θ_i , bastando para isto, substituir β por $\hat{\beta}$ em (C.7.3), ou seja

$$\hat{\theta}_i = \frac{\exp(\mathbf{x}_i^\top \hat{\beta})}{1 + \exp(\mathbf{x}_i^\top \hat{\beta})}. \quad (\text{C.7.6})$$

A distribuição aproximada de $\hat{\theta}_i$ pode ser obtido por meio do **Método Delta** (ver Apêndice B). Para isto, notemos que

$$\begin{aligned} \frac{\partial}{\partial \beta} \theta(\mathbf{x}_i; \beta) &= \frac{\exp(\mathbf{x}_i^\top \beta) [1 + \exp(\mathbf{x}_i^\top \beta)] - [\exp(\mathbf{x}_i^\top \beta)]^2}{[1 + \exp(\mathbf{x}_i^\top \beta)]^2} \mathbf{x}_i \\ &= \frac{\exp(\mathbf{x}_i^\top \beta)}{[1 + \exp(\mathbf{x}_i^\top \beta)]^2} \mathbf{x}_i = \theta(\mathbf{x}_i; \beta) [1 - \theta(\mathbf{x}_i; \beta)] \mathbf{x}_i. \end{aligned}$$

Então, utilizando (B.2.2), obtemos

$$\begin{aligned} \mathbb{V}(\hat{\theta}_i) &= \{\theta(\mathbf{x}_i; \beta) [1 - \theta(\mathbf{x}_i; \beta)]\}^2 \mathbf{x}_i^\top \mathbb{V}(\hat{\beta}) \mathbf{x}_i \\ &= \{\theta(\mathbf{x}_i; \beta) [1 - \theta(\mathbf{x}_i; \beta)]\}^2 \mathbf{x}_i^\top (\mathbf{X}^\top \mathbf{W}[\boldsymbol{\theta}(\beta)] \mathbf{X})^{-1} \mathbf{x}_i. \end{aligned}$$

Consequentemente, a matriz de covariâncias de $\hat{\boldsymbol{\theta}}$ pode ser escrita compactamente como

$$\mathbb{V}(\hat{\boldsymbol{\theta}}) = \mathbf{W}[\boldsymbol{\theta}(\beta)] \mathbf{X} (\mathbf{X}^\top \mathbf{W}[\boldsymbol{\theta}(\beta)] \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{W}[\boldsymbol{\theta}(\beta)] \quad (\text{C.7.7})$$

e pelo Método Delta, obtemos,

$$\hat{\theta} \approx N_n\{\theta, \mathbb{V}(\hat{\theta})\},$$

em que $\mathbb{V}(\hat{\theta})$ é dada por (C.7.7). Para efeito de aplicações, uma estimativa de (C.7.7) pode ser obtida por meio da substituição de β por $\hat{\beta}$. Para detalhes, o leitor poderá consultar Hosmer & Lemeshow (2000) entre outros.

Com intuito didático, voltemos aos dados da Tabela C.7.1 e consideremos um modelo logístico para a chance de deslocamento do disco, tendo apenas a distância aberta como variável explicativa. Nesse contexto, o modelo (C.7.4) corresponde a

$$\log[\theta(x_i; \alpha, \beta)]/[1 - \theta(x_i; \alpha, \beta)] = \alpha + x_i\beta \quad (\text{C.7.8})$$

$i = 1, \dots, 104$ em que $\theta(x_i; \alpha, \beta)$ representa a probabilidade de deslocamento do disco quando o valor da distância aberta é x_i , α denota o logaritmo da chance de deslocamento do disco quando a distância aberta tem valor $x_i = 0$ e β é interpretado como a variação no logaritmo da chance de deslocamento do disco por unidade de variação da distância aberta. Consequentemente, a razão de chances do deslocamento do disco correspondente a uma diferença de d unidades da distância aberta será $\exp(d \times \beta)$. Como não temos dados correspondentes a distâncias abertas menores que 0.50, convém substituir os valores x_i por valores “centrados”, ou seja por $x_i^* = x_i - x_0$. Uma possível escolha para x_0 é o mínimo de x_i , que é 0.50. Essa transformação na variável explicativa altera somente a interpretação do parâmetro α que passa a ser o logaritmo da chance de deslocamento do disco quando a distância aberta tem valor $x_i = 0.50$.

Estimativas (com erros padrões entre parênteses) dos parâmetros desse modelo ajustado por máxima verossimilhança aos dados da Tabela C.7.1, são, $\hat{\alpha} = -5.86$ (1.10) e $\hat{\beta} = 3.16$ (0.66) e então, segundo o modelo, uma estimativa da chance de deslocamento do disco para articulações com distância aberta $x = 0.50$ (que corresponde à distância aberta transformada $x^* = 0.00$) é $\exp(-5.86) = 0.003$; uma estimativa da razão entre a chance de deslocamento do disco para articulações com distância aberta $x + 1$ e um intervalo de confiança (95%) para essa chance pode ser obtido exponenciando os limites (LI e LS) do intervalo para o parâmetro α , nomeadamente

$$\begin{aligned} LI &= \exp[\hat{\alpha} - 1.96EP(\hat{\alpha})] = \exp(-5.16 - 1.96 \times 1.10) = 0.000 \\ LS &= \exp[\hat{\alpha} + 1.96EP(\hat{\alpha})] = \exp(-5.16 + 1.96 \times 1.10) = 0.024. \end{aligned}$$

Os limites de um intervalo de confiança para a razão de chances correspondentes a uma variação de uma unidade no valor da distância aberta podem ser obtidos de maneira similar e são 6.55 e 85.56.

Com base em (C.7.6) podemos estimar a probabilidade de sucesso (deslocamento do disco, no exemplo sob investigação); por exemplo, para uma articulação cuja distância aberta seja 2.10 (correspondente à distância aberta transformada igual a 1.60), a estimativa dessa probabilidade é

$$\hat{\theta} = \exp(-5.86 + 3.16 \times 1.60) / [1 + \exp(-5.86 + 3.16 \times 1.60)] = 0.31.$$

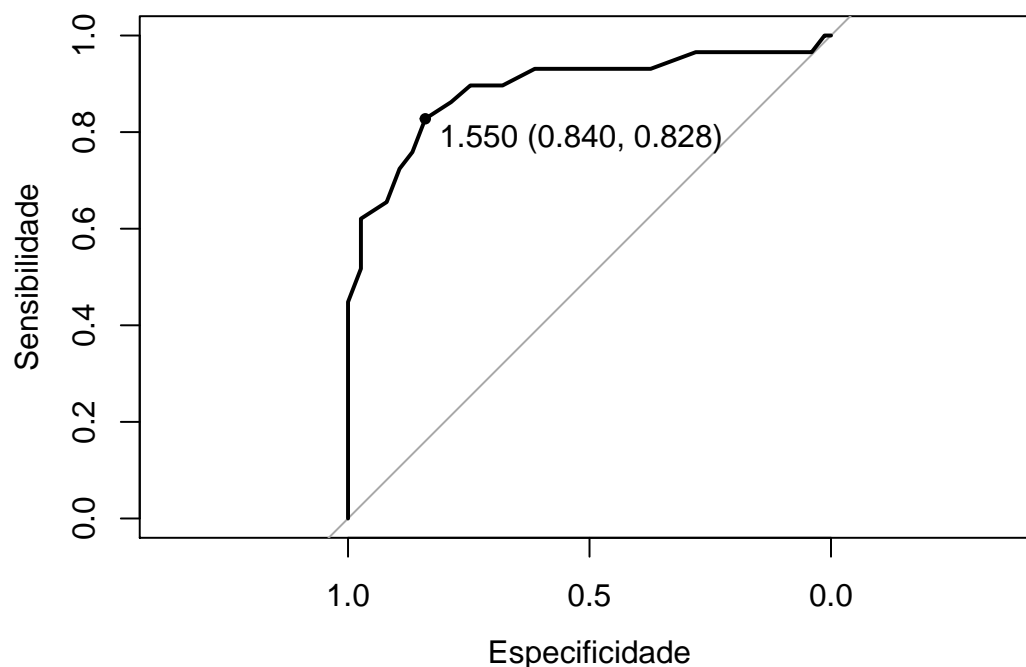
Lembrando que o objetivo do estudo é substituir o processo de identificação de deslocamento do disco realizado via ressonância magnética por aquele baseado na medida da distância aberta por meio de ultrassonografia, podemos estimar as probabilidades de sucesso para todas as articulações e identificar um **ponto de corte** d_0 segundo o qual, distâncias abertas com valores acima dele sugerem decidirmos pelo deslocamento do disco e distâncias abertas com valores abaixo dele sugerem a decisão oposta. Obviamente, não esperamos que todas as decisões tomadas dessa forma sejam corretas e conseqüentemente, a escolha do ponto de corte deve ser feita com o objetivo de minimizar os erros (decidir pelo deslocamento quando ele não existe ou *vice versa*). Nesse contexto, um contraste entre as decisões tomadas com base em um determinado ponto de corte d_0 e o padrão áureo definido pela ressonância magnética para todas as 104 articulações pode ser resumido por meio da Tabela C.7.2 em que as frequências da diagonal principal correspondem a decisões corretas e aquelas da diagonal secundária às decisões erradas. O quociente $n_{11}/(n_{11} + n_{21})$ é

Tabela C.7.2: Frequência de decisões para um ponto de corte d_0

		Deslocamento real do disco	
		sim	não
Decisão baseada na distância aberta d_0	sim	n_{11}	n_{12}
	não	n_{21}	n_{22}

conhecido como **sensibilidade** do processo de decisão e é uma estimativa da probabilidade de decisões corretas quando o disco está realmente deslocado. O quociente $n_{22}/(n_{12} + n_{22})$ é conhecido como **especificidade** do processo de decisão e é uma estimativa da probabilidade de decisões corretas quando o disco está realmente não está deslocado. A situação ideal é aquela em que tanto a sensibilidade quanto a especificidade do processo de decisão são iguais a 100%. O problema a resolver é determinar o ponto de corte d_{max} que gere o melhor equilíbrio entre sensibilidade e especificidade. Com essa finalidade, podemos construir tabelas com o mesmo formato da Tabela C.7.2) para diferentes pontos de corte e um gráfico cartesiano entre a sensibilidade e especificidade obtida de cada uma delas. Esse gráfico, conhecido como **curva ROC** (do termo inglês *Receiver Operating Characteristic*) gerado para os dados da Tabela C.7.1 está apresentado na Figura C.7.1.

Figura C.7.1: Curva ROC para os dados da Tabela C.7.1 baseada no modelo (C.7.8 com distância aberta como variável explicativa



O ponto de corte ótimo é aquele mais próximo do vértice superior esquerdo (em que tanto a sensibilidade quanto a especificidade seriam iguais a 100%). Para o exemplo, esse ponto está salientado na Figura C.7.1 e corresponde à distância aberta com valor $d_{max} = 2.05 (= 1.55 + 0.50)$. A sensibilidade e a especificidade associadas à decisão baseada nesse ponto de corte, são, respectivamente, 83% e 84% e as frequências de decisões corretas estão indicadas na Tabela C.7.3. Com

Tabela C.7.3: Frequência de decisões para um ponto de corte para distância aberta $d_{max} = 2.05$

		Deslocamento real do disco	
		sim	não
Decisão baseada na distância aberta $d_{max} = 2.05$	sim	24	12
	não	5	63

esse procedimento de decisão a porcentagem de acertos (**acurácia**) é 84% [= (24 +

63)/104]. A porcentagem de **falsos positivos** é 17% [= $5/(5 + 29)$] e a porcentagem de **falsos negativos** é 16% [= $12/(12 + 63)$].

Uma análise similar, baseada na distância fechada (transformada por meio da subtração de seu valor mínimo (0,4) gera a curva ROC apresentada na Figura C.7.2 e frequências de decisões apresentada na Tabela C.7.4.

Figura C.7.2: Curva ROC para os dados da Tabela C.7.1 baseada no modelo (C.7.8 com distância fechada como variável explicativa

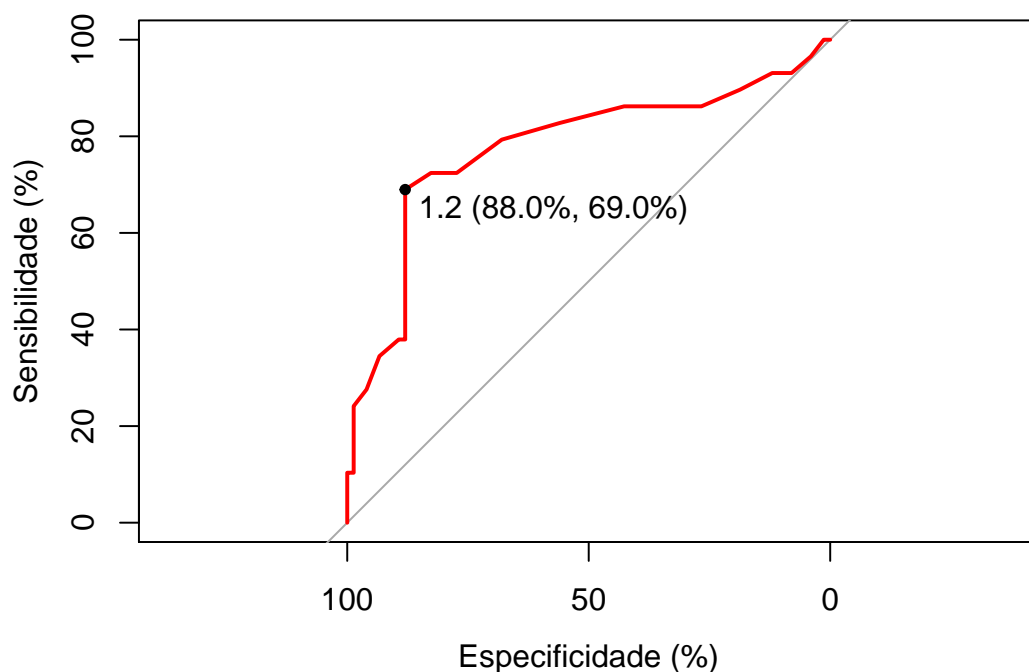


Tabela C.7.4: Frequência de decisões para um ponto de corte para distância fechada $d_{max} = 1,60$

		Deslocamento real do disco	
		sim	não
Decisão baseada na distância fechada $d_{max} = 1.60$	sim	20	9
	não	9	66

A acurácia associada a processo de decisão baseado apenas na distância fechada, 83% [= $(20 + 66)/104$] é praticamente igual àquela obtida com base apenas na

distância aberta; no entanto aquele processo apresenta um melhor equilíbrio entre sensibilidade e especificidade (83% e 84%, respectivamente, *versus* 88% e 69%).

Se quisermos avaliar o processo de decisão com base nas observações das distâncias aberta e fechada simultaneamente, podemos considerar o modelo

$$\log[\theta(x_i; \alpha, \beta, \gamma)]/[1 - \theta(x_i; \alpha, \beta, \gamma)] = \alpha + x_i\beta + w_i\gamma \quad (\text{C.7.9})$$

$i = 1, \dots, 104$ em que w_i corresponde à distância fechada observada na i -ésima articulação. Neste caso, γ corresponde à razão entre a chance de deslocamento do disco para articulações com distância fechada $w + 1$ e a chance de deslocamento do disco para articulações com distância fechada w para aquelas com mesmo valor da distância aberta; uma interpretação similar vale para o parâmetro β . Estimativas dos parâmetros (com erros padrões entre parênteses) do modelo (C.7.9) obtidas após a transformação das variáveis explicativas segundo o mesmo figurino adotado nas análises univariadas são $\hat{\alpha} = -6.38$ (1.19), $\hat{\beta} = 2.83$ (0.67) e $\hat{\gamma} = 0.98$ (0.54). A estimativa do parâmetro γ é apenas marginalmente significativa, ou seja a inclusão da variável explicativa distância fechada não acrescenta muito poder de discriminação além daquele correspondente à distância aberta. Uma das razões para isso é que as duas variáveis são correlacionadas (com coeficiente de correlação de Pearson igual a 0.46. A determinação de pontos de corte para modelos com duas ou mais variáveis explicativas é bem mais complexa do que no caso univariado e não será abordada neste texto. Para efeito de comparação com as análises anteriores, as frequências de decisões obtidas com os pontos de corte utilizados naquelas estão dispostas na Tabela C.7.5, e correspondem a uma sensibilidade de 62%, especificidade de 97% e acurácia de 88%.

Tabela C.7.5: Frequência de decisões correspondentes a pontos de corte $d_{max} = 2.05$ para distância aberta e $d_{max} = 1.60$ para distância fechada

		Deslocamento real do disco	
		sim	não
Decisão baseada em	sim	18	2
ambas as distâncias	não	11	73

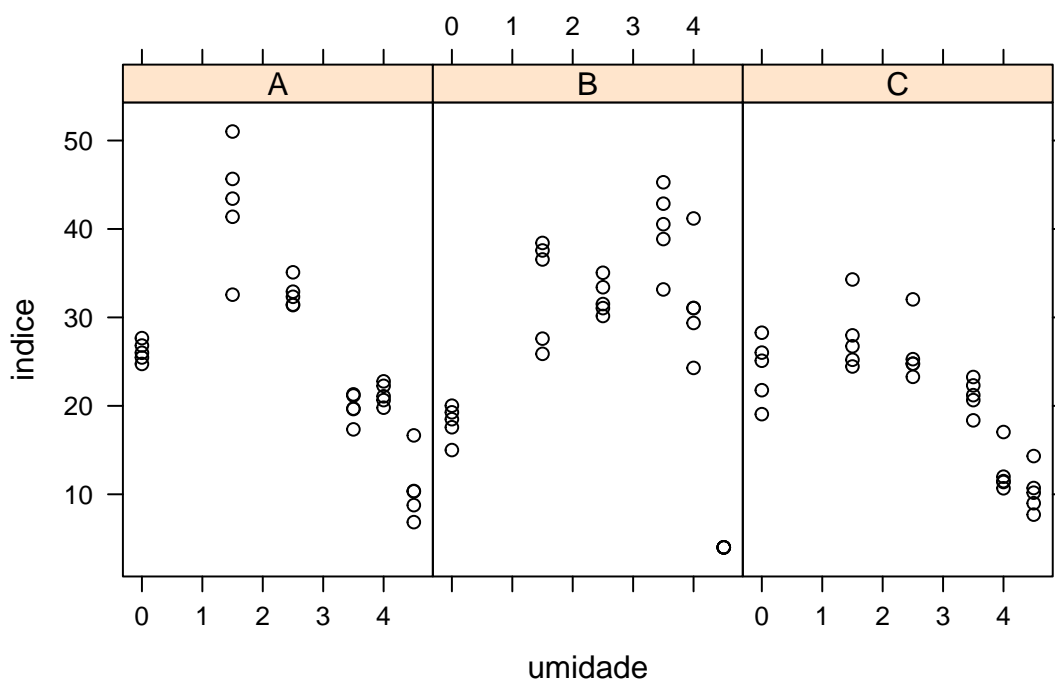
C.8 Exemplos

Exemplo C.8.1: Os dados representados na Tabela C.8.1 e disponíveis em

www.ime.usp.br/~jmsinger/Dados/Singer&Nobre&Rocha2018exempc81.xls

são provenientes de um estudo conduzido na Faculdade de Odontologia da Universidade de São Paulo com o objetivo de avaliar o efeito do nível de umidade na resistência de união (medida por meio de um índice) de três adesivos dentários. Cada um de três conjuntos de trinta molares extraídos foi tratado com um dos três adesivos (A, B ou C), com 5 dentes submetidos a cada nível de umidade (0.00, 1.50, 2.50, 3.50, 4.00 e 4.50). Depois de um certo tempo, a resistência de união foi avaliada em cada um dos 90 dentes. Detalhes podem ser obtidos em Reis, Loguercio, Azevedo, Carvalho, Singer & Grande (2003). Médias e desvios padrões correspondentes aos índices de resistência de união estão dispostos na Tabela C.8.2 e gráficos de dispersão correspondentes estão apresentados na Figura C.8.1.

Figura C.8.1: Gráficos de dispersão para os dados da Tabela C.8.1



Ambos sugerem um erro de observação ou transcrição para os dados associados às observações dos dentes submetidos ao adesivo B com nível de umidade 4.50. Esses valores serão eliminados na análise subsequente. O gráfico de dispersão sugere um

Tabela C.8.1: Índice de resistência de adesivos dentários

adesivo	umid	índice	adesivo	umid	índice	adesivo	umid	índice
A	0.0	24.74	B	0.0	15.00	C	0.0	28.27
A	0.0	27.66	B	0.0	20.02	C	0.0	19.06
A	0.0	26.00	B	0.0	19.27	C	0.0	26.02
A	0.0	25.47	B	0.0	18.49	C	0.0	25.10
A	0.0	26.82	B	0.0	17.58	C	0.0	21.77
A	1.5	45.66	B	1.5	25.88	C	1.5	24.45
A	1.5	51.02	B	1.5	36.55	C	1.5	34.29
A	1.5	32.57	B	1.5	38.42	C	1.5	25.21
A	1.5	41.38	B	1.5	37.57	C	1.5	26.74
A	1.5	43.43	B	1.5	27.60	C	1.5	27.97
A	2.5	32.91	B	2.5	31.05	C	2.5	24.75
A	2.5	35.10	B	2.5	35.05	C	2.5	24.77
A	2.5	31.39	B	2.5	30.16	C	2.5	32.04
A	2.5	32.33	B	2.5	33.41	C	2.5	25.28
A	2.5	31.45	B	2.5	31.51	C	2.5	23.28
A	3.5	19.69	B	3.5	45.28	C	3.5	18.36
A	3.5	21.17	B	3.5	38.86	C	3.5	23.26
A	3.5	19.64	B	3.5	40.54	C	3.5	22.31
A	3.5	21.30	B	3.5	42.86	C	3.5	20.65
A	3.5	17.34	B	3.5	33.16	C	3.5	21.21
A	4.0	21.08	B	4.0	24.29	C	4.0	11.39
A	4.0	22.77	B	4.0	41.19	C	4.0	10.68
A	4.0	19.79	B	4.0	29.37	C	4.0	17.04
A	4.0	22.26	B	4.0	31.07	C	4.0	11.49
A	4.0	20.65	B	4.0	31.05	C	4.0	11.99
A	4.5	6.85	B	4.5	4.00	C	4.5	8.99
A	4.5	8.77	B	4.5	4.00	C	4.5	10.71
A	4.5	16.66	B	4.5	4.00	C	4.5	10.18
A	4.5	10.33	B	4.5	4.00	C	4.5	14.32
A	4.5	10.37	B	4.5	4.00	C	4.5	7.70

Tabela C.8.2: Médias (desvios padrões) do índice de resistência de união

Adesivo	Umidade					
	0.00	1.50	2.50	3.50	4.00	4.50
A	26.1 (1.1)	42.8 (6.8)	32.6 (1.5)	19.8 (1.6)	21.3 (1.2)	10.6 (3.7)
B	18.1 (1.9)	33.2 (6.0)	32.2 (2.0)	40.1 (4.6)	31.4 (6.1)	4.0 (0.0)
C	24.0 (3.6)	27.7 (3.9)	26.0 (3.4)	21.2 (1.9)	12.5 (2.6)	10.4 (2.5)

efeito quadrático dos níveis de umidade. Uma análise inicial que incorpora essa sugestão pode ser concretizada por meio do ajuste do seguinte modelo quadrático

$$y_{ijk} = \alpha_i + \beta_i x_k + \gamma_i x_k^2 + e_{ijk}, \quad (\text{C.8.1})$$

$i = 1, \dots, 3$, $j = 1, \dots, 5$, $k = 1, \dots, 6$ em que y_{ijk} representa o índice de resistência para o j -ésimo dente submetido ao i -ésimo tratamento ($i = 1$ correspondendo ao adesivo A, $i = 2$, ao adesivo B e $i = 3$, ao adesivo C) sob o k -ésimo nível de umidade, x_k (lembrando que para o adesivo B, $k = 1, \dots, 5$ dada a eliminação dos valores correspondentes ao nível de umidade 4.50). Supomos que $e_{ijk} \sim N(0, \sigma^2)$ são erros aleatórios independentes.

Os resultados do ajuste desse modelo aos dados estão dispostos na Tabela C.8.3. O desvio padrão residual é $S = 4.55$ e o coeficiente de determinação ajustado,

Tabela C.8.3: Estimativas (erros padrões) dos coeficientes do modelo (C.8.1) aos dados da Tabela C.8.1

Adesivo	Coeficiente		
	linear	angular	quadrático
A	27.94 (1.97)	11.35 (1.96)	-3.43 (0.42)
B	18.18 (2.00)	12.25 (2.27)	-2.09 (0.54)
C	23.98 (1.97)	5.71 (1.96)	-1.99 (0.42)

$R_{aj}^2 = 0.972$. O excelente ajuste obtido sob essa ótica é confirmado por intermédio do gráfico de resíduos padronizados e do correspondente histograma, apresentados na Figura C.8.2. A sugestão de uma leve assimetria não deve ser importante no que tange à estimação dos parâmetros.

Figura C.8.2: Gráfico de resíduos padronizados e o correspondente histograma referentes ao ajuste do modelo (C.8.1)

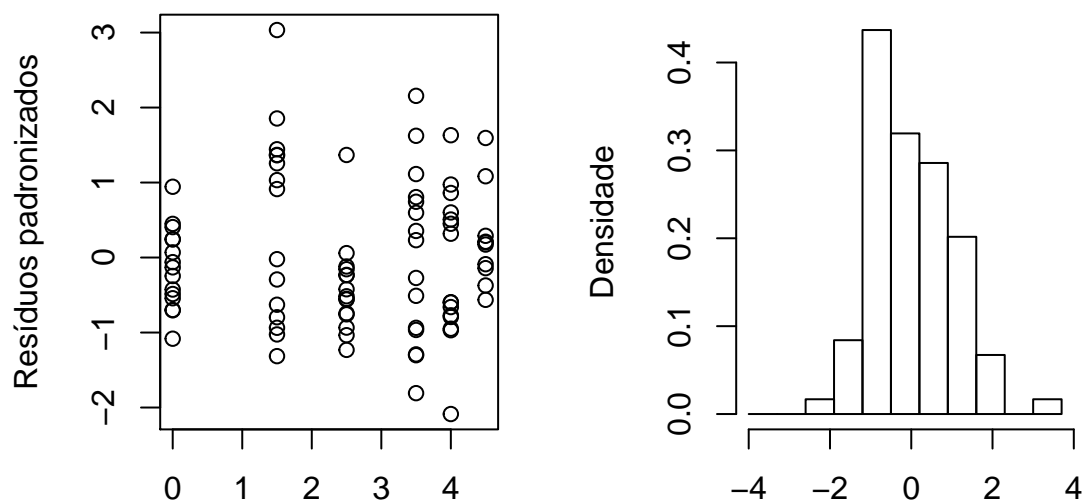
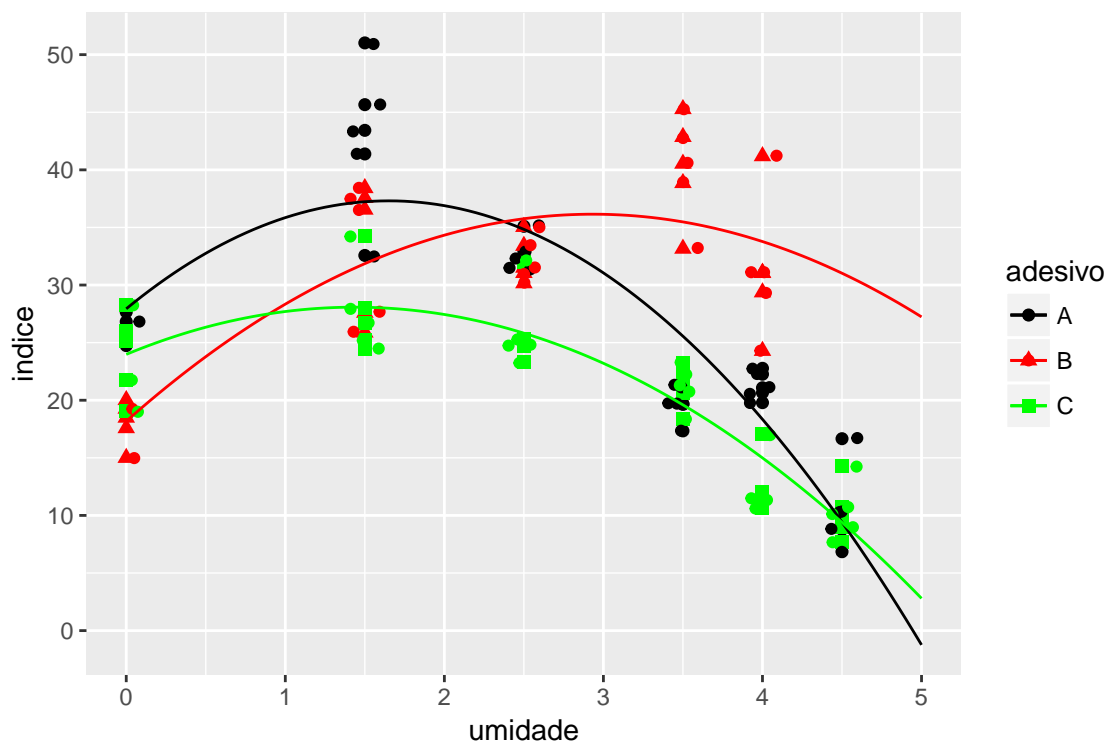


Figura C.8.3: Gráfico com curvas ajustadas pelo modelo (C.8.1)



Na Figura C.8.3 apresentamos as curvas ajustadas para os três adesivos.

A comparação entre as curvas que representam o efeito da umidade no índice de resistência correspondentes aos três adesivos pode ser realizada por meio de testes de hipóteses sobre seus parâmetros. A hipótese de que as três curvas são coincidentes pode ser expressa na forma $H : \mathbf{C}\boldsymbol{\beta} = \mathbf{0}$ com

$$\mathbf{C} = \begin{pmatrix} 1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 \end{pmatrix}$$

e $\boldsymbol{\beta} = (\alpha_1, \beta_1, \gamma_1, \alpha_2, \beta_2, \gamma_2, \alpha_3, \beta_3, \gamma_3)^\top$. O resultado do teste, baseado numa estatística F com 6 graus de liberdade no numerador e 76 graus de liberdade no denominador mostra forte evidência de diferença entre as curvas ($p < 0.001$). Comparações múltiplas podem ser concretizadas de forma semelhante. Por exemplo, para comparar as curvas associadas aos adesivos A e B, basta testar a hipótese utilizando apenas a primeira, terceira e quinta linhas da matriz \mathbf{C} . Nesse caso, a estatística F tem 3 graus de liberdade no numerador. Todas as comparações entre as curvas duas a duas sugerem diferenças altamente significativas ($p < 0.001$).

Os níveis ótimos de umidade (correspondentes ao máximo índice de resistência de união) são dados por $g(\beta_i, \gamma_i) = -\beta_i/(2\gamma_i)$, $i = 1, \dots, 3$ e suas estimativas obtidas por meio dos valores apresentados na Tabela C.8.3. Estimativas de suas variâncias podem ser obtidas via Método Delta, observando que

$$\mathbf{u}_i = \left[\frac{\partial g(\beta_i, \gamma_i)}{\partial \beta_i}, \frac{\partial g(\beta_i, \gamma_i)}{\partial \gamma_i} \right]^\top = \left[-\frac{1}{2\gamma_i}, \frac{\beta_i}{2\gamma_i^2} \right]^\top$$

de forma que $\mathbb{V}[g(\hat{\beta}_i, \hat{\gamma}_i)] = \mathbf{u}_i(\hat{\boldsymbol{\beta}})^\top \mathbb{V}_i(\hat{\boldsymbol{\beta}}) \mathbf{u}_i(\hat{\boldsymbol{\beta}})$ com $\mathbb{V}_i(\hat{\boldsymbol{\beta}})$ representando a submatriz (com dimensão 2×2) de $\mathbb{V}(\hat{\boldsymbol{\beta}})$ correspondente aos parâmetros β_i, γ_i . Estimativas dos pontos de resistência máxima e intervalos de confiança (aproximados) com coeficientes de confiança de 95% estão dispostos na Tabela C.8.4.

Tabela C.8.4: Estimativas e intervalos de confiança (95%) para os pontos de resistência máxima

Adesivo	Ponto de máximo	Limites do IC(95%)	
		inferior	superior
A	1.65	1.44	1.86
B	2.93	1.26	4.60
C	1.44	0.52	2.35

C.9 Exercícios

C.9.1. Seja x_1, \dots, x_n uma amostra aleatória de uma variável X cuja distribuição é Normal.

- Mostre que a média e a variância amostrais são independentes.
- Enuncie um resultado semelhante no contexto de regressão linear simples.
- Qual a utilidade desse resultado?

C.9.2. Mostre que se $\mathbf{y} \sim N_p(\boldsymbol{\mu}, \mathbf{V})$ então $(\mathbf{y} - \boldsymbol{\mu})^\top \mathbf{V}^{-1}(\mathbf{y} - \boldsymbol{\mu}) \sim \chi_p^2$.

C.9.3. Considere um vetor aleatório (X, Y) com distribuição normal bivariada. Mostre que a esperança condicional de Y dado $X = x$ é da forma $E(Y|X = x) = \alpha + \beta x$ explicitando os parâmetros α e β em termos dos parâmetros da distribuição normal bivariada adotada.

C.9.4. Obtenha os estimadores de máxima verossimilhança dos parâmetros do modelo de regressão

$$y_i = x_i^\beta e_i,$$

$i = 1, \dots, n$ em que os e_i representam erros aleatórios independentes log-normais com média $\exp(\sigma^2/2)$ e variância $\exp(\sigma^2)[\exp(\sigma^2) - 1]$.

- Obtenha a distribuição do estimador de β e construa um intervalo de confiança.
- Obtenha um intervalo de confiança aproximado para o valor esperado de y dado x_0 .

Sugestão: linearize o modelo e lembre-se que se $\log X \sim N(\mu, \sigma^2)$ então X tem distribuição log-normal com média $\exp(\mu + \sigma^2/2)$ e variância $[\exp(\sigma^2) - 1] \exp(2\mu + \sigma^2)$.

C.9.5. Suponha que as variáveis x_i e y_i estejam relacionadas de acordo com o modelo de regressão linear simples

$$y_i = \alpha + \beta x_i + e_i,$$

$i = 1, \dots, n$ em que os e_i representam erros aleatórios independentes de média zero e variância σ^2 . Mostre que o estimador

$$S^2 = \frac{1}{n-2} \sum_{i=1}^n (y_i - \hat{\alpha} - \hat{\beta}x_i)^2$$

em que $\hat{\alpha}$ e $\hat{\beta}$ denotam os estimadores de mínimos quadrados de α e β , respectivamente, é não enviesado para a variância σ^2 .

C.9.6. Expresse o modelo de regressão linear simples $y_i = \alpha + \beta x_i + e_i$, $i = 1, \dots, n$ em que os e_i representam erros aleatórios independentes de média zero e variância σ^2 na forma $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$ e calcule $\hat{\boldsymbol{\beta}}$ e $\mathbb{V}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}^\top \mathbf{X})^{-1} \sigma^2$.

C.9.7. Para avaliar a associação entre a pressão arterial sistólica e idade, colheram-se os dados dispostos na Tabela C.1.1. Utilize esses dados para avaliar se essa associação pode ser representada por um modelo de regressão linear simples. Com essa finalidade,

- Especifique o modelo, interpretando os parâmetros;
- Construa um diagrama de dispersão (rotulando os eixos convenientemente);
- Estime os parâmetros e os correspondentes erros padrões;
- Construa intervalos de confiança (com coeficiente de confiança de 95%) para os parâmetros;
- Obtenha o valor-p correspondente ao teste da hipótese de que o coeficiente angular é nulo.

C.9.8. Considere os modelos

$$y_i = \alpha + \beta x_i + e_i \quad \text{e} \quad y_i = \beta x_i + e_i$$

$i = 1, \dots, n$ em que os e_i representam erros aleatórios independentes de média zero e variância σ^2 . Para ambos os modelos, expresse o coeficiente de determinação R^2 em termos de x_i e y_i e discuta a diferença entre eles.

C.9.9. Para investigar a associação entre tipo de escola (particular ou pública), cursada por calouros de uma universidade e a média no curso de Cálculo I, obtiveram-se os seguintes dados:

Escola	Média no curso de Cálculo I								
Particular	8.6	8.6	7.8	6.5	7.2	6.6	5.6	5.5	8.2
Pública	5.8	7.6	8.0	6.2	7.6	6.5	5.6	5.7	5.8

Seja y_i a nota obtida pelo i -ésimo aluno, $x_i = 1$ se o aluno cursou escola particular e $x_i = -1$ se o aluno cursou escola pública, $i = 1, \dots, 18$. Considere o modelo $y_i = \alpha + \beta x_i + e_i$, $i = 1, \dots, 18$ em que os e_i são erros aleatórios não correlacionados com $E(e_i) = 0$ e $Var(e_i) = \sigma^2$.

- i) Interprete os parâmetros α e β .
- ii) Estime α e β pelo método de mínimos quadrados. Obtenha também uma estimativa de σ^2 .
- iii) Construa intervalos de confiança para α e β .
- iv) Com base nas estimativas obtidas no item ii), construa intervalos de confiança para os valores esperados das notas dos alunos das escolas particulares e públicas.
- v) Ainda utilizando o modelo proposto, especifique e teste a hipótese de que ambos os valores esperados são iguais.
- vi) Repita os itens i)-v) definindo $x_i = 1$ se o aluno cursou escola particular e $x_i = 0$ se o aluno cursou escola pública, $i = 1, \dots, 18$.

C.9.10. Num estudo realizado na Faculdade de Medicina da Universidade de São Paulo foram colhidos dados de 16 pacientes submetidos a transplante intervivos e em cada um deles obtiveram-se medidas tanto do peso (g) real do lobo direito do fígado quanto de seu volume (cm^3) previsto pré operatorialmente por métodos ultrassonográficos. O objetivo é estimar o peso real por meio do volume previsto. Os dados estão dispostos na tabela abaixo.

- i) Proponha um modelo de regressão linear simples para analisar os dados e interprete seus parâmetros.
- ii) Construa um gráfico de dispersão apropriado.
- iii) Ajuste o modelo e construa intervalos de confiança para seus parâmetros.
- iv) Avalie o ajuste do modelo por meio de medidas descritivas, de testes de hipóteses convenientes e de uma análise de resíduos.
- v) Construa uma tabela com intervalos de confiança para o peso esperado do lobo direito do fígado correspondentes a volumes (estimados ultrassonograficamente) de 600, 700, 800, 900 e 1000 cm^3 .
- vi) Repita os itens anteriores considerando um modelo linear simples sem intercepto.

Volume USG (cm^3)	Peso real (g)	Volume USG (cm^3)	Peso real (g)
656	630	737	705
692	745	921	955
588	690	923	990
799	890	945	725
766	825	816	840
800	960	584	640
693	835	642	740
602	570	970	945

C.9.11. Os dados abaixo são provenientes de uma pesquisa cujo objetivo é propor um modelo para a relação entre a área construída de um determinado tipo de imóvel e o seu preço.

Imóvel	Área (m^2)	Preço (R\$)
1	128	10 000
2	125	9 000
3	200	17 000
4	4.000	200 000
5	258	25 000
6	360	40 000
7	896	70 000
8	400	25 000
9	352	35 000
10	250	27 000
11	135	11 000
12	6.492	120 000
13	1.040	35 000
14	3.000	300 000

- i) Construa um gráfico de dispersão apropriado para o problema.
- ii) Ajuste um modelo de regressão linear simples e avalie a qualidade do ajuste (obtenha estimativas dos parâmetros e de seus erros padrões, calcule o coeficiente de determinação e construa gráficos de resíduos e um gráfico do tipo QQ).
- iii) Ajuste o modelo linearizável

$$y = \beta x^\gamma e$$

em que y representa o preço e x representa a área e avalie a qualidade do ajuste comparativamente ao modelo linear ajustado no item ii); construa um gráfico de dispersão com os dados transformados.

- iv) Utilizando o modelo com o melhor ajuste, construa intervalos de confiança com coeficiente de confiança (aproximado) de 95% para os preços esperados de imóveis com $200m^2$, $500m^2$ e $1000m^2$.

C.9.12. O arquivo Bosco (1998), disponível na forma de uma planilha Excel no sítio www.ime.usp.br/~jmsinger contém dados provenientes de um estudo observacional baseado numa amostra de 66 pacientes matriculadas na Clínica Ginecológica do Departamento de Obstetrícia e Ginecologia do Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo e no Setor de Mamas do Centro de Referência da Saúde da Mulher e de Nutrição, Alimentação e Desenvolvimento Infantil entre novembro de 1995 e outubro de 1997. Um dos objetivos é estudar a relação entre o tamanho clínico de tumores de mama e seu tamanho obtido ultrassonograficamente. O tamanho clínico do tumor é definido como a média dos valores encontrados nas colunas N e O da planilha mencionada; as três medidas ultrassonográficas (altura, comprimento e largura) estão disponíveis nas colunas U, V e W.

- i) Construa gráficos de dispersão apropriados para o problema.
- ii) Ajuste um modelo de regressão linear múltipla tendo como variável resposta o tamanho clínico do tumor e como variáveis explicativas as três medidas ultrassonográficas.
- iii) Avalie a qualidade do ajuste por meio do coeficiente de determinação, testes de hipóteses convenientes, gráficos de resíduos e um gráfico do tipo QQ.
- iv) Com base nas conclusões obtidas dos itens acima, verifique se é possível reduzir o modelo (eliminando uma ou duas variáveis explicativas).
- v) Repita a análise utilizando a raiz cúbica do volume ultrassonográfico do tumor, definido como

$$Volume = \frac{\pi}{6}(altura \times comprimento \times largura)$$

como única variável explicativa.

- vi) Compare o modelo que você julgou mais adequado por meio da análise realizada nos itens i)-iv) com aquele obtido no item v).

C.9.13. Os dados abaixo são provenientes de um estudo cujo objetivo era avaliar a eficácia de um tipo de escova de dentes na remoção de placa bacteriana. Para isso foram observados índices de placa bacteriana (maiores valores do índice correspondendo a maiores quantidades de placa) antes (X) e após (Y) a escovação numa amostra de 26 crianças.

- a) Assuma que o par (X, Y) segue uma distribuição normal bivariada. Proponha um modelo que permita a comparação das médias de X e Y , expresse-o em

- notação matricial e interprete todos os seus parâmetros. Teste a hipótese de que as médias de X e Y são iguais e construa um intervalo de confiança para a sua diferença. Indique precisamente como devem ser realizados os cálculos.
- b) Construa um gráfico de dispersão tendo o índice de placa bacteriana pré-escovação (X) no eixo das abscissas e o índice de placa bacteriana pós-escovação (Y) no eixo das ordenadas.
- c) Proponha um modelo de regressão para explicar a variação do índice pós-escovação como função do índice pré-escovação, levando em conta o fato de que índices pré-escovação nulos implicam índices pós-escovação nulos (em média). Explícite as suposições e interprete os parâmetros.
- d) Ajuste o modelo proposto e apresente os resultados de forma não técnica.
- e) Utilize técnicas de diagnóstico para avaliar o ajuste do modelo.
- f) Qual dos modelos você usaria para analisar os dados? Por que?

Índice de placa bacteriana			
pré-escovação	pós-escovação	pré-escovação	pós-escovação
2.18	0.43	1.40	0.24
2.05	0.08	0.90	0.15
1.05	0.18	0.58	0.10
1.95	0.78	2.50	0.33
0.28	0.03	2.25	0.33
2.63	0.23	1.53	0.53
1.50	0.20	1.43	0.43
0.45	0.00	3.48	0.65
0.70	0.05	1.80	0.20
1.30	0.30	1.50	0.25
1.25	0.33	2.55	0.15
0.18	0.00	1.30	0.05
3.30	0.90	2.65	0.25

C.9.14. Considere o modelo $y_i = \alpha + \beta x_i + e_i$, $i = 1, \dots, 20$ em que $e_i = \rho e_{i-1} + u_i$ com $u_i \sim N(0, 1)$ e assuma que $\alpha = 2$, $\beta = 0.5$, $e_0 = 3$ e $\rho = 0.9$.

- a) Utilizando um gerador de números aleatórios obtenha 20 valores de u_i e construa os valores correspondentes de e_i , $i = 1, \dots, 20$.
- b) Construa um gráfico de e_i em função de i .
- c) Obtenha os valores de y_i para $x_i = i$, $i = 1, \dots, 20$ e construa o gráfico de dispersão correspondente incluindo nele a reta $E(y_i|x_i) = \alpha + \beta x_i$.
- d) Obtenha os estimadores de mínimos quadrados de α e β a partir dos dados gerados no item c) e inclua a reta estimada no gráfico do item c).

- e) Calcule a estatística de Durbin-Watson e discuta os resultados, interpretando o efeito da autocorrelação dos erros.
- f) Repita os itens c) - e) com $\rho = 0.5$ e $\rho = 0.1$
- g) Compare os resultados obtidos com os diferentes valores de ρ e comente as diferenças encontradas.

C.9.15. Os dados do arquivo intitulado Singer&Andrade (1997), disponíveis na forma de uma planilha Excel no sítio www.ime.usp.br/~jmsinger são provenientes de um estudo cujo objetivo era avaliar a eficácia de dois tipos de escova de dentes (Hugger e Convencional) na remoção de placa bacteriana. Para isso foram observados índices de placa bacteriana (maiores valores do índice correspondendo a maiores quantidades de placa) antes (X) e após (Y) a escovação com cada tipo de escova numa amostra de $n_1 = 14$ crianças do gênero feminino (F) e $n_2 = 12$ do gênero masculino (M).

- a) Construa um gráfico de dispersão para cada tipo de escova, tendo o índice de placa bacteriana pré-escovação (X) no eixo das abscissas e o índice de placa bacteriana pós-escovação (Y) no eixo das ordenadas. Use símbolos diferentes para identificar crianças de cada gênero.
- b) Utilize métodos de regressão linear simples para ajustar modelos do tipo

$$Y_{ij} = \beta_i X_{ij}^{\gamma_i} e_{ij},$$

- $i = 1, 2, j = 1, \dots, n_i$ para cada tipo de escova separadamente, indicando as suposições adotadas.
- c) Teste a hipótese de que $\gamma_1 = \gamma_2$ e no caso de não rejeição, reajuste o modelo com $\gamma_1 = \gamma_2 = \gamma$.
- d) No modelo reduzido, teste a hipótese de que $\gamma = 1$ e em caso de não rejeição, ajuste um novo modelo reduzido que incorpore esse resultado.
- e) No terceiro modelo reduzido, teste a hipótese de que $\beta_1 = \beta_2 = \beta$ e, se for o caso, ajuste um novo modelo que incorpore o resultado.
- f) Avalie a qualidade do ajuste do último modelo ajustado por intermédio de técnicas de diagnóstico.
- g) Compare os resultados das análises dos dados das duas escovas.
- h) Escreva sua conclusão, interpretando as hipóteses testadas e construindo um tabela com valores esperados para os índices de placa pós-escovação com diferentes níveis de índices de placa pré-escovação.

C.9.16. Considere o modelo

$$y_{ij} = \alpha + \beta x_{ij} + a_i + e_{ij},$$

$i = 1, \dots, n, j = 1, \dots, m$ em que y_{ij} , α , β e x_{ij} têm as interpretações usuais e as variáveis aleatórias $a_i \sim N(0, \sigma_a^2)$, $e_{ij} \sim N(0, \sigma^2)$ são independentes.

- i) Mostre que $Var(y_{ij}) = \sigma_a^2 + \sigma^2$, $Cov(y_{ij}, y_{ij'}) = \sigma_a^2$, $j \neq j'$ e que $Cov(y_{ij}, y_{i'j'}) = 0$, $i \neq i'$.
- ii) Escreva o modelo descrito acima na forma matricial, especificando a matriz de covariâncias do vetor de erros, *i.e.*, do vetor cujos componentes são $a_i + e_{ij}$, $i = 1, \dots, n$, $j = 1, \dots, m$.
- iii) Assuma que tanto σ_a^2 quanto σ^2 são conhecidas e obtenha os estimadores de mínimos quadrados generalizados e máxima verossimilhança dos demais parâmetros do modelo e suas respectivas distribuições.

C.9.17. A Tabela C.1.2 contém dados de capacidade instalada (*ton*), potência instalada (*1000 kW*) e área construída (*100 m²*) de 10 empresas de uma certa indústria. Com o objetivo de estimar a capacidade instalada (*Y*) a partir das informações sobre potência instalada (*X₁*) e área construída (*X₂*),

- i) Construa gráficos de dispersão apropriados.
- ii) Especifique um modelo de regressão linear e interprete seus parâmetros.
- iii) Obtenha estimativas dos parâmetros e de seus erros padrões e calcule o coeficiente de determinação múltiplo.
- iv) Avalie o ajuste do modelo por meio de gráficos de resíduos e gráficos QQ.
- v) Avalie a perda de precisão dos estimadores decorrente do uso de cada uma das variáveis explicativas em modelos de regressão linear simples construídos com o mesmo propósito do modelo descrito no item ii).
- vi) Com base no modelo mais adequado dentre aqueles que você analisou, construa uma tabela com intervalos de confiança (coeficiente de confiança = 95%) para as capacidades de produção esperadas de empresas com todas as combinações de potências instaladas de 1.0, 2.5 e 5.0 ($\times 1000 \text{ kW}$) e áreas construídas 8.0, 10.0 e 12.0 ($\times 100 \text{ m}^2$).

C.9.18. Os dados do arquivo intitulado Braga (1998), disponível na forma de uma planilha Excel no sítio www.ime.usp.br/~jmsinger são oriundos de um estudo realizado na Faculdade de Medicina da Universidade de São Paulo para avaliar pacientes com insuficiência cardíaca. Foram estudados 87 pacientes com algum nível de insuficiência cardíaca, além de 40 pacientes controle (coluna K). Para cada paciente foram registradas algumas características físicas (altura, peso, superfície corporal, idade, sexo). Eles foram submetidos a um teste de esforço cardiopulmonar em cicloergômetro em que foram medidos a frequência cardíaca, o consumo de oxigênio, o equivalente ventilatório de oxigênio, o equivalente ventilatório de dióxido de carbono, o pulso de oxigênio e a pressão parcial de dióxido de carbono ao final da expiração, em três momentos diferentes: no limiar anaeróbio, no ponto de compensação respiratória e no pico do exercício.

Ajuste um modelo linear que permita comparar a relação entre o consumo de oxigênio no limiar anaeróbico do exercício (coluna X) e a carga na esteira ergométrica (coluna U) para pacientes com diferentes níveis de insuficiência cardíaca (medida segundo a classificação NYHA - coluna K). Com essa finalidade, você deve:

- Construir gráficos de dispersão convenientes.
- Interpretar os diferentes parâmetros do modelo.
- Estimar os parâmetros do modelo e apresentar os respectivos erros padrões.
- Avaliar a qualidade de ajuste do modelo por meio de gráficos de resíduos e gráficos QQ.
- Identificar possíveis valores discrepantes (*outliers*) e reajustar o modelo sem esses pontos.
- Comparar os ajustes dos modelos obtidos com e sem os *outliers*.
- Definir e testar hipóteses adequadas para avaliar se a relação entre consumo de oxigênio no limiar anaeróbico do exercício e carga na esteira ergométrica depende da classificação NYHA.
- Reajustar o modelo com base nas conclusões do item (g) e avaliar o seu ajuste.
- Apresentar conclusões que evitem jargão técnico.

C.9.19. As tabelas abaixo foram obtidas da análise de um conjunto de dados.

Coefficients	Estimate	Std. Error	<i>t</i> value	Pr(> <i>t</i>)
(Intercept)	4.1003	2.8355	1.446	0.1862
w	-0.8334	0.2640	-3.157	0.0135 *
x	2.5496	0.3151	8.093	4.02e-05 ***

Residual standard error: 2.153 on 8 degrees of freedom Multiple R-Squared: 0.905,
Adjusted R-squared: 0.8813 F-statistic: 38.11 on 2 and 8 DF, p-value: 8.141e-05

	(Intercept)	w	x
(Intercept)	1.46702786	-0.08622291	-0.11625387
w	-0.08622291	0.03978328	-0.01996904
x	-0.11625387	-0.01996904	0.03142415

- Quantas variáveis explicativas e quantas observações (*n*) foram utilizadas na análise?
- Especifique o modelo adotado.
- Há alguma evidência de que o modelo se ajusta bem aos dados? Justifique.

- d) Fixe um valor (entre 1 e 10) para cada variável explicativa e construa um intervalo de confiança com coeficiente de confiança de 95% para o valor esperado da resposta correspondente aos valores fixados para as variáveis explicativas.

C.9.20. Para estudar a associação entre gênero (1=Masc, 0=Fem) e idade (anos) e a preferência (1=sim, 0=não) pelo refrigerante Kcola, o seguinte modelo de regressão logística foi ajustado aos dados de 50 crianças escolhidas ao acaso:

$$\log \left\{ \frac{\pi_i(x_i, w_i)}{1 - \pi_i(x_i, w_i)} \right\} = \alpha + \beta x_i + \gamma(w_i - 5),$$

em que x_i (w_i) representa o gênero (idade) da i -ésima criança e $\pi_i(x_i, w_i)$ a probabilidade de uma criança do gênero x_i e idade w_i preferir Kcola. As seguintes estimativas para os parâmetros foram obtidas:

Parâmetro	Estimativa	Erro-padrão	Valor p
α	0.69	0.12	< 0.01
β	0.33	0.10	< 0.01
γ	-0.03	0.005	< 0.01

- Interprete os parâmetros do modelo por intermédio de chances e razões de chances.
- Com as informações acima, estime a razão de chances de preferência por Kcola correspondente à comparação de crianças do mesmo gênero com 10 e 15 anos.
- Construa intervalos de confiança (com coeficiente de confiança aproximado de 95%) para $\exp(\beta)$ e $\exp(\gamma)$ e traduza o resultado em linguagem não técnica.
- Estime a probabilidade de meninos com 15 anos preferirem Kcola.

C.9.21. No arquivo Singer&Ikeda (1996) disponível em www.ime.usp.br/~jmsinger você encontra dados provenientes de um estudo cuja finalidade é identificar fatores de risco para a doença aterosclerótica coronariana (definida como obstrução de mais de 50% de pelo menos uma coronária).

- Utilize modelos de regressão logística para verificar se a presença de angina estável (ANGEST), antecedentes hereditários (AH), infarto do miocárdio prévio (IMP), nível de triglicérides para pacientes sem medicamento (TRIGS), nível de colesterol para pacientes sem medicamento (COLS), idade (IDADE1) e sexo (SEXO) podem ser consideradas como fatores de risco para a doença aterosclerótica coronariana (LO3). Considere somente os participantes com dados completos para as variáveis indicadas.

- b) Com base no modelo selecionado, construa uma tabela com limites inferiores e superiores de intervalos de confiança (com coeficiente de confiança de 95%) para razões de chances definidas a partir de um conjunto de valores pré-especificados das variáveis explicativas.

C.9.22. Para avaliar a relação entre o tempo de uso (X) e o número de defeitos (Y) de um determinado componente eletrônico, obteve-se uma amostra de n itens desse tipo e em cada um observaram-se as duas variáveis. O estatístico encarregado da análise concluiu que o número de defeitos segue uma distribuição de Poisson, *i.e.*, que $P(Y = y) = \exp(-\mu)\mu^y/y!$, $y = 0, 1, \dots$ e propôs um modelo linear sem intercepto para a relação entre o número esperado de defeitos e o tempo de uso, *i.e.*, $\mu = \beta x$.

- Obtenha o estimador de máxima verossimilhança do parâmetro do modelo.
- Mostre que esse estimador é não enviesado.
- Obtenha a variância desse estimador.
- Construa um intervalo de confiança com coeficiente de confiança aproximado de 95% para o valor esperado do número de defeitos de componentes com tempo de uso igual a x_0 .

C.9.23. Os dados abaixo correspondem ao faturamento de empresas similares de um mesmo setor industrial nos últimos 15 meses.

mês	jan	fev	mar	abr	mai	jun	jul	ago	set	out	nov	dez	jan	fev	mar
vendas	1.0	1.6	1.8	2.0	1.8	2.2	3.6	3.4	3.3	3.7	4.0	6.4	5.7	6.0	6.8

Utilize técnicas de análise de regressão para quantificar o crescimento do faturamento de empresas desse setor ao longo do período observado. Com essa finalidade:

- Proponha um modelo adequado, interpretando todos os parâmetros e especificando as suposições.
- Estime os parâmetros do modelo e apresente os resultados numa linguagem não técnica.
- Utilize técnicas de diagnóstico para avaliar o ajuste do modelo.

C.9.24. Obtenha os estimadores de máxima verossimilhança dos parâmetros do modelo de regressão

$$y_i = \beta x_i + e_i,$$

$i = 1, \dots, n$ em que os e_i representam erros aleatórios independentes normais com média zero e variância σ^2 .

- i) Obtenha a distribuição do estimador de β e construa um intervalo de confiança.
- ii) Repita o item i) com a suposição adicional de que $Var(e_i) = x_i^2 \sigma^2$.

C.9.25. Utilizando a notação usual, considere o modelo linear

$$\begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \end{pmatrix} = \begin{pmatrix} E(y_{11}) \\ E(y_{12}) \\ E(y_{21}) \\ E(y_{22}) \end{pmatrix}$$

- a) Expresse os parâmetros do modelo em termos dos valores esperados $E(y_{ij})$.
- b) Repita o procedimento do item anterior sob as restrições $\alpha_1 = 0$ e $\alpha_1 + \alpha_2 = 0$.
- c) Repita o procedimento agora sob a reparametrização $\beta_1 = \mu + \alpha_1$ e $\beta_2 = \mu + \alpha_2$.
- d) Repita o procedimento agora sob a reparametrização $\beta_1 = \mu$ e $\beta_2 = \mu + \alpha_1$.

Interprete os parâmetros em cada caso.

C.9.26. A tabela abaixo contém dados obtidos de diferentes institutos de pesquisa coletados entre fevereiro de 2008 e março de 2010 e correspondem às porcentagens de eleitores favoráveis a cada um dos dois principais candidatos à presidência do Brasil.

- a) Construa um diagrama de dispersão apropriado, evidenciando os pontos correspondentes a cada um dos candidatos.
- b) Especifique um modelo polinomial de segundo grau, homocedástico, que represente a variação da preferência eleitoral de cada candidato ao longo do tempo.
- c) Ajuste o modelo especificado no item anterior.
- d) Avalie o ajuste do modelo e verifique, por meio de testes de hipóteses adequadas, se ele pode ser simplificado; em caso afirmativo, ajuste o modelo mais simples.
- e) Com base no modelo escolhido, estime a porcentagem esperada de eleitores favoráveis a cada um dos candidatos em 3 de outubro de 2010 e construa um intervalo de confiança para a diferença entre essas porcentagens esperadas.
- f) Faça uma crítica da análise e indique o que poderia ser feito para melhorá-la (mesmo não que não saiba implementar suas sugestões).

Porcentagem de eleitores favoráveis

Fonte	Data	Dilma	Serra	Fonte	Data	Dilma	Serra
sensus	16/02/2008	4.5	38.2	sensus	13/08/2009	19	39.5
dataf	27/03/2008	3	38	ibope	04/09/2009	14	34
sensus	25/04/2008	6.2	36.4	sensus	14/09/2009	21.7	31.8
sensus	19/09/2008	8.4	38.1	ibope	20/11/2009	17	38
dataf	28/11/2008	8	41	vox	30/11/2009	17	39
sensus	30/11/2008	10.4	46.5	vox	07/12/2009	18	39
ibope	12/12/2008	5	42	dataf	14/12/2009	23	37
sensus	14/12/2008	13.3	42.8	vox	18/12/2009	27	34
dataf	30/01/2009	11	41	sensus	17/01/2010	27	33.2
sensus	19/03/2009	16.3	45.7	ibope	29/01/2010	25	36
dataf	27/03/2009	16	38	dataf	06/02/2010	28	32
sensus	28/05/2009	23.5	40.4	ibope	25/02/2010	30	35
ibope	29/05/2009	18	38	dataf	27/03/2010	27	36
dataf	01/06/2009	17	36	vox	31/03/2010	31	34

C.9.27. Os dados da tabela abaixo foram obtidos de um estudo cujo objetivo era avaliar a relação entre a quantidade de um certo aditivo (X) e o tempo de vida (Y) de um determinado alimento. Os valores substituídos por ? ficaram ilegíveis depois que o responsável pelo estudo derramou café sobre eles.

X (g/kg)	5	10	15	20	30
Y (dias)	3.2	?	?	?	?

Um modelo de regressão linear simples (com a suposição de normalidade e independência dos erros) foi ajustado aos dados gerando os seguintes resultados:

Tabela de ANOVA

<i>Fonte de variação</i>	<i>gl</i>	<i>SQ</i>	<i>QM</i>	<i>F</i>	<i>Valor p</i>
Regressão	1	42,30	42,30	156,53	0,001
Resíduo	3	0,81	0,27		
Total	4	43,11			

Intervalos de confiança (95%)

Parâmetro	<i>Limite inferior</i>	<i>Limite superior</i>
Intercepto	-0,19	2,93
X	0,25	0,42

Resíduos

<i>Observação</i>	<i>Resíduos</i>
1	0,14
2	-0,55
3	0,66
4	-0,23
5	-0,01

$$(\mathbf{X}^t \mathbf{X})^{-1} = \begin{pmatrix} 0.892 & -0.043 \\ -0.043 & 0.003 \end{pmatrix}$$

- Escreva o modelo na forma matricial e interprete seus parâmetros.
- Construa um intervalo de confiança para o valor esperado do tempo de vida do produto quando a quantidade de aditivo utilizada é de 25 g/kg.
- Construa um intervalo de previsão para o valor do tempo de vida do produto quando a quantidade de aditivo utilizada é de 25 g/kg.
- Reconstrua a tabela dos dados, *i.e.*, calcule os valores de Y substituídos por ?.

Observação: O quantil de ordem 97,5% da distribuição t com 3 graus de liberdade é 3.182.

C.9.28. Os dados abaixo são provenientes de uma pesquisa cujo objetivo é avaliar o efeito da dosagem de uma certa droga (X) na redução de pressão arterial (Y) de pacientes hipertensos.

Homens		Mulheres	
	Redução		Redução
Dose	de pressão	Dose	de pressão
1	3	2	4
3	5	3	7
4	9	5	11
6	15	6	14
		6	13

O pesquisador sugeriu o seguinte modelo para a análise dos dados

$$y_{ij} = \alpha_i + \beta(x_{ij} - \bar{x}) + e_{ij}$$

$i = 1, 2$, $j = 1, \dots, n_i$ em que os erros e_{ij} seguem distribuições $N(0, \sigma^2)$ independentes e \bar{x} denota a dose média empregada no estudo.

- a) Interprete os parâmetros do modelo.
- b) Escreva o modelo na forma matricial.

C.9.29. Uma determinada empresa de transportes deseja saber se o número de passageiros está relacionado com o preço da gasolina. Com essa finalidade, adotaram o modelo

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + e$$

em que y corresponde ao número médio de passageiros no mês, x_1 representa a variação no preço médio mensal da gasolina relativamente à média do ano anterior, x_2 é uma variável indicadora do tipo de percurso do ônibus (0 = expresso e 1 = usual) e e é um erro aleatório de média zero.

- a) Interprete os parâmetros do modelo.
- b) Que hipótese você testaria para avaliar se a relação entre o número médio mensal de passageiros (y) e a variação no preço da gasolina (x_1) é igual para ônibus com os dois tipos de percurso.
- c) Suponha que as estimativas de mínimos quadrados dos parâmetros sejam $\hat{\beta}_0 = 500$, $\hat{\beta}_1 = 50$, $\hat{\beta}_2 = 5$ e $\hat{\beta}_3 = -10$. Esboce um gráfico que represente a relação entre y e x_1 para ônibus com cada tipo de percurso.

C.9.30. Num estudo cujo objetivo era avaliar o efeito do sexo (X) e da idade (Y) de indivíduos no envolvimento em acidentes automobilísticos foram coletadas informações de sobre essas três variáveis para uma amostra de clientes de uma empresa de seguros.

- a) Proponha um modelo de regressão logística para representar a associação entre as três variáveis e interprete os seus parâmetros.
- b) Calcule a chance de um homem com 40 anos se envolver num acidente.
- c) Por que valor fica multiplicada essa chance (de envolvimento em acidentes) para uma mulher com 50 anos?

C.9.31. Uma fábrica de cadeiras dispõe dos seguintes dados sobre sua produção mensal:

Número de cadeiras produzidas	105	130	141	159	160	172
Custos fixos e variáveis (R\$)	1700	1850	1872	1922	1951	1970

- a) Proponha um modelo de regressão linear simples para a relação entre o custo e o número de cadeiras produzidas e interprete os parâmetros;
- b) Utilize um intervalo de confiança com coeficiente de confiança de 95% para estimar o custo esperado de produção para 200 cadeiras;

- c) Admitindo que o preço de venda é de R\$ 20.00 por unidade, qual a menor quantidade de cadeiras que deve ser produzida para que o lucro seja positivo.

C.9.32. Considere o seguinte modelo de regressão

$$y_i = \begin{cases} \beta_0 + \beta_1 x_i + e_i, & x_i \leq x_0, \quad i = 1, \dots, j \\ \beta_0 + \beta_1 x_0 + \beta_2(x_i - x_0) + e_i, & x_i \geq x_0, \quad i = j + 1, \dots, n. \end{cases}$$

em que os termos y_i , x_i , e_i , têm a interpretação usual e x_0 é uma constante positiva. Escreva o modelo com notação matricial e interprete os parâmetros β_0 , β_1 , e β_2 .

C.9.33. Numa pesquisa realizada na Faculdade de Medicina da Universidade de São Paulo, foi observada uma amostra de 32 recém-nascidos pré-termo (RNPT), vulgarmente chamados de prematuros. Em cada um deles foram medidos o diâmetro da aorta (em mm) e a idade (em semanas) desde a concepção (instante em que houve a fecundação). O objetivo era verificar se 39 semanas após a concepção, o diâmetro médio da aorta dos RNPT era equivalente ao diâmetro médio da aorta de recém-nascidos a termo (RNT), vulgarmente conhecidos como normais. Para isso, foi ajustado um modelo de regressão linear simples, cujos resultados estão indicados abaixo.

Coeficiente	Estimativa	Erro padrão
Intercepto	8.42	0.20
Angular	-0.12	0.02

$$(\mathbf{X}^t \mathbf{X})^{-1} = \begin{pmatrix} 0.25 & -0.0125 \\ -0.0125 & 0.0025 \end{pmatrix}$$

em que \mathbf{X} denota a matriz com os valores das variáveis explicativas.

- Obtenha uma estimativa da variância dos erros (σ^2).
- Estime o valor esperado do diâmetro da aorta para 39 semanas pós-concepção e estime o seu erro padrão.
- Sabendo que o diâmetro médio (populacional) da aorta de RNT (39 semanas após a concepção) é de 3.23 mm, responda a pergunta que originou o estudo, justificando sua resposta.
- Repita a análise do item anterior no caso de o valor do diâmetro médio da aorta de RNT (3.23 mm) ser proveniente de uma amostra de tamanho 30 e que o erro padrão correspondente é de 0.50 mm.

Observações:

- Utilize um nível de significância de 5% em todas as suas análises.

- ii) Utilize uma aproximação normal para as distribuições envolvidas nos métodos inferenciais que você utilizar.
- iii) Indique claramente como os cálculos foram realizados.

C.9.34. Os dados da tabela abaixo foram obtidos de empresas de duas indústrias (Alimentos e Produtos de Limpeza) com o objetivo de avaliar a relação entre o valor investido em propaganda (X) num certo trimestre e o aumento de faturamento no trimestre subsequente (Y).

Indústria	Valor investido em propaganda	Aumento de faturamento
Alimentos	5	20
Alimentos	7	25
Alimentos	9	28
Alimentos	10	27
Alimentos	8	26
Alimentos	6	22
Limpeza	4	16
Limpeza	11	19
Limpeza	7	17
Limpeza	9	16
Limpeza	6	15

- a) Especifique um modelo em que a relação entre X e Y é quadrática, com coeficientes possivelmente diferentes para cada indústria. Interprete os parâmetros do modelo.
- b) Escreva o modelo na forma matricial.
- c) Utilizando notação matricial, especifique a hipótese de que as curvas correspondentes à relação entre X e Y são iguais para as duas indústrias.

C.9.35. Com a finalidade de comparar homens (H) e mulheres (M) quanto à relação entre o tempo gasto entre a chegada e a saída em um centro de compras, (X) e o valor dispendido (Y), considerou-se o modelo:

$$y_{ij} = \mu + \alpha_i + \gamma(x_{ij} - x_0) + e_{ij},$$

$\sum_{i=1}^2 \alpha_i = 0$, $i = 1 (M), 2 (H)$, $j = 1, \dots, n_i$ em que as variáveis aleatórias e_{ij} têm distribuições $N(0, \sigma^2)$ independentes.

- a) Descreva todos os símbolos utilizados, interpretando-os.
- b) Especifique o modelo na forma matricial.

- c) Utilizando notação matricial, especifique a hipótese de que a relação entre Y e X é igual para homens e mulheres.
- d) Esboce um gráfico que represente o modelo.

C.9.36. Num estudo cujo objetivo era avaliar a associação entre atividade (técnica ou administrativa), tempo de serviço (anos) e salário (R\$ 1000.00) e a participação em um programa de demissão voluntária (sim ou não), foram coletadas informações sobre essas quatro variáveis para uma amostra de empregados de uma grande empresa.

- a) Proponha um modelo de regressão logística para representar a associação entre as variáveis e interprete os seus parâmetros.
- b) Com base no modelo proposto, obtenha uma expressão para cálculo da chance de um empregado do setor administrativo com 5 anos de serviço e ganhando um salário de R\$ 3000.00 participar do programa de demissão voluntária.
- c) Segundo o modelo proposto, por que valor fica multiplicada essa chance para um funcionário da área técnica com salário de R\$ 2000.00 e de mesma idade?

C.9.37. A tabela abaixo contém dados de uma investigação cujo objetivo era estudar a relação entre a duração de diabetes e a ocorrência de retinopatia (uma moléstia dos olhos).

- a) Considere um modelo linear para avaliar a intensidade dessa relação e utilize o método de mínimos quadrados generalizados para ajustá-lo. **Sugestão:** Considere o ponto médio de cada intervalo como valor da variável explicativa e use as frequências relativas observadas para estimação das variâncias.
- b) Utilize o método de máxima verossimilhança para ajustar um modelo linear e um modelo logístico aos dados.
- c) Compare os resultados do ajuste dos modelos lineares obtidos pelos dois métodos por meio de uma tabela com as estimativas e erros padrões dos parâmetros.

Duração da Diabetes (anos)	Retinopatia	
	Sim	Não
0 - 2	17	215
3 - 5	26	218
6 - 8	39	137
9 - 11	27	62
12 - 14	35	36
15 - 17	37	16
18 - 20	26	13
21 - 23	23	15

C.9.38. Num estudo em que se desejava comparar um tipo de pneu experimental ($i = 2$) com um tipo de pneu convencional ($i = 1$) com relação ao desgaste Y (profundidade do sulco em mm) em função da distância percorrida X (em 1000 km), foram observados $n_1 = 4$ pneus convencionais e $n_2 = 5$ pneus convencionais. Os dados estão dispostos abaixo.

Tipo de pneu	Distância percorrida	Profundidade do sulco
Convencional	10	99
Convencional	20	95
Convencional	30	72
Convencional	40	56
Experimental	10	93
Experimental	20	86
Experimental	25	83
Experimental	30	77
Experimental	40	68

O estatístico responsável pela análise propôs o seguinte modelo

$$y_{ij} = \mu + \alpha_i + \beta_i x_{ij} + e_{ij},$$

$i = 1, 2, j = 1, \dots, m_i$ com $\alpha_1 = 0$ e $e_{ij} \sim N(0, \sigma^2)$, independentes.

- Escreva o modelo na forma matricial e interprete seus parâmetros.
- Represente o modelo graficamente indicando claramente o significado dos parâmetros.
- Expresse em termos matriciais, a hipótese de que os dois tipos de pneus têm desgaste esperado equivalente. Especifique a distribuição da estatística que você utilizaria para testar essa hipótese.

C.9.39. Considere o modelo

$$y_i = \alpha x_i + e_i,$$

$i = 1, \dots, n$ em que $e_i \sim N(0, \sigma^2)$ são variáveis aleatórias independentes.

- Obtenha o estimador de máxima verossimilhança de α e proponha um estimador não enviesado para σ^2 .
- Especifique a distribuição do estimador de α .
- Especifique um intervalo de confiança para o parâmetro α com coeficiente de confiança γ , $0 < \gamma < 1$.

Observação: a função densidade da distribuição $N(\mu, \sigma^2)$ é

$$f(x) = (\sqrt{2\pi}\sigma)^{-1} \exp(-2^{-1}[(x - \mu)/\sigma]^2).$$

C.9.40. Considere o modelo $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$ em que \mathbf{y} é um vetor de respostas com dimensão $(n_1 + n_2) \times 1$, $\mathbf{e} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_{n_1+n_2})$, $\boldsymbol{\beta}$ é um vetor de parâmetros,

$$\mathbf{X} = \begin{bmatrix} \mathbf{1}_{n_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{1}_{n_2} \end{bmatrix},$$

com \mathbf{I}_m representando uma matriz identidade dimensão m , $\mathbf{1}_m$, um vetor de dimensão m com todos os elementos iguais a 1 e $\mathbf{0}$ um vetor com todos os elementos iguais a zero.

- Interprete os parâmetros do modelo.
- Expresse o estimador de mínimos quadrados dos elementos de $\boldsymbol{\beta}$ em termos de somatórios.
- Expresse um estimador não enviesado de σ^2 em termos de somatórios.
- Dê um exemplo de situação prática em que esse modelo poderia ser aplicado.

C.9.41. As tabelas abaixo foram obtidas da análise de um conjunto de dados.

Ajuste do Modelo

Coefficients	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	13.32	6.93	1.92	0.084
x_1	4.37	0.81	5.42	0.000
x_2	-22.60	5.46	-4.14	0.002
x_3	-7.36	5.46	-1.35	0.208

Multiple R-Squared: 81.02%, Adjusted R-squared: 75.06%

ANOVA

Source	DF	SS	MS	F	p
Regression	3	2807.90	935.97	14.42	0.001
Error	10	649.09	64.91		
Total	13	3456.99			

- Quantas variáveis explicativas e quantas observações (n) foram utilizadas na análise?
- Especifique o modelo adotado e interprete seus parâmetros.

- c) Apresente uma estimativa para a variância dos erros baseada num estimador não enviesado.
- d) Indique como se calcula o coeficiente R^2 a partir das tabelas acima.
- e) Especifique (em termos dos parâmetros do modelo) a hipótese testada por meio da estatística F da tabela de ANOVA e indique sua distribuição.
- f) Especifique (em termos dos parâmetros do modelo) as hipóteses testadas por meio das estatísticas t da tabela de ajuste do modelo e indique suas distribuições.
- g) Esclareça as diferenças entre as hipóteses consideradas nos itens e) e f).
- h) Com base nos resultados da análise, proponha um modelo mais simples para os dados, justificando sua resposta.

C.9.42. Reproduza todas as análises do Exemplo C.8.1 com e sem os pontos associados ao adesivo B e nível de umidade 4.50 adotando as parametrizações correspondentes aos seguintes comandos da função `lm()` do pacote R:

- a) `lm(indice ~ adesivo + umidade + umidade2 + adesivo:umidade + adesivo:umidade2)`
- b) `lm(indice ~ adesivo + umidade + umidade2 + adesivo:umidade + adesivo:umidade2 -1)`

com `umidade2 = umidade2`. Em cada caso,

- i) especifique e interprete os parâmetros do modelo;
- ii) mostre como você obteve os resultados apresentados no Exemplo C.8.1;
- iii) avalie o impacto da eliminação dos dados mencionados nas estimativas dos parâmetros;
- iv) compare os coeficientes de determinação dos dois modelos e explique possíveis diferenças.

C.9.43. Os dados disponíveis em

www.ime.usp.br/~jmsinger/Dados/Singer&Nobre&Rocha2018exerc943.xls

são provenientes de um estudo conduzido no Instituto de Biociências da Universidade de São Paulo com o objetivo de avaliar o efeito do número de malárias contraídas durante a gestação em algumas características de recém nascidos. Avalie o efeito de quantidade de malárias (`qntmal`, coluna C), idade da mãe [`idade` (anos), coluna E], peso da mãe na triagem [`pesotriag` (kg), coluna H], peso da mãe no parto [`pesoparto` (kg), coluna I]), altura da mãe [`altura` (cm), coluna J] e idade gestacional [`ig` (sem), coluna K] no peso do recém nascido [`peso` (g), coluna M] por meio de modelos de regressão múltipla. Com essa finalidade,

- a) construa gráficos de dispersão e *boxplots* para descrever o comportamento da variável peso;
- b) explicita a estratégia de análise empregada;
- c) ajuste os modelos adotados para concretizar a estratégia adotada;
- d) utilize técnicas de diagnóstico para avaliar a qualidade dos ajustes;
- e) apresente os resultados na forma de um relatório descrevendo cada passo da análise e a sua conclusão, quantificando-a.

C.9.44. Considere o modelo

$$y_{ij} = \alpha + \beta t_j + a_i + b_i t_j + e_{ij}, \quad i = 1, \dots, n, \quad j = 1, \dots, m_i,$$

em que y_{ij} é a resposta da i -ésima unidade amostral no j -ésimo tempo, α e β são efeitos fixos, $\mathbf{b}_i = (a_i, b_i)^\top$ são efeitos aleatórios independentes com distribuição $N(\mathbf{0}, \mathbf{G})$ e e_{ij} são erros aleatórios independentes com distribuição $N(0, \sigma^2)$. Suponha que \mathbf{b}_i e e_{ij} são independentes. Obtenha expressões para a variância de y_{ij} e para a covariância entre y_{ij} e y_{il} , $j \neq l$ nos seguintes casos:

- a) $\mathbf{G} = \text{diag}[\sigma_a^2, \sigma_b^2]$
- b) $\mathbf{G} = \begin{pmatrix} \sigma_a^2 & \sigma_{ab} \\ \sigma_{ab} & \sigma_b^2 \end{pmatrix}$

Bibliografia

- Afiune, J. Y. (2000). *Avaliação ecocardiográfica evolutiva de recém-nascidos pré-termo, do nascimento até o termo*, PhD thesis, São Paulo: Instituto da Criança da Faculdade de Medicina, Universidade de São Paulo.
- Afiune, J. Y., Singer, J. M. & Leone, C. R. (2005). Evolução ecocardiográfica de recém-nascidos com persistência do canal arterial, *Jornal de Pediatria* **81**: 454–460.
- Alencar, A. P., Singer, J. M. & Rocha, F. M. M. (2012). Competing regression models for longitudinal data, *Biometrical Journal* **54**: 214–229.
- Andrade, D. F. & Helms, R. W. (1984). Maximum likelihood estimates in the multivariate normal distribution with patterned mean and covariance via the EM algorithm, *Communications in Statistics, Theory and Methods* **13**: 2239–2251.
- Andrade, D. F. & Singer, J. M. (1998). Profile analysis for randomized complete block experiments, **25**: 237–244.
- Andrews, D. F. & Pregibon, D. (1978). Finding outliers that matter, *Journal of the Royal Statistical Society B* **40**: 85–93.
- Anscombe, F. J. (1973). Graphs in statistical analysis, *The American Statistician* **27**: 17–21.
- Arellano-Vale, R., Bolfarine, H. & Lachos, V. H. (2005). Skew-normal linear mixed models, *Journal of Data Science* **3**: 415–438.
- Arellano-Vale, R., Bolfarine, H. & Lachos, V. H. (2007). Bayesian inference for skew-normal linear mixed models, *Journal of Applied Statistics* **34**: 663–682.
- Atkinson, A. C. (1981). Two graphical display for outlying and influential observations in regression, *Biometrika* **68**: 13–20.
-

- Atkinson, A. C. (1985). *Plots, transformations, and regression: an introduction to graphical methods of diagnostic regression analysis*, Oxford: Oxford University Press.
- Atkinson, A. C. & Riani, M. (2000). *Robust diagnostic regression analysis*, New York: Springer.
- Baksalary, J. K., Corsten, L. C. A. & Kala, R. (1978). Reconciliation of two different views on estimation of growth curve parameters, *Biometrika* **65**: 662–665.
- Baltagi, B. H. (1995). *Econometric analysis of panel data*, Chichester: Wiley.
- Banerjee, M. & Frees, E. W. (1997). Influence diagnostics for linear longitudinal modelss, *Journal of the American Statistical Association* **92**: 999–1005.
- Beckman, R. J., Nachtsheim, C. J. & Cook, R. D. (1987). Diagnostics for mixed-model analysis of variance, *Technometrics* **29**: 413–426.
- Belsley, D. A., Kuh, E. & Welsch, R. E. (1980). *Regression diagnostics: identifying influential data and sources of collinearity*, New York: Wiley.
- Bickel, P. J. & Doksum, K. A. (2001). *Mathematical Statistics, Volume 1*, 2 edn, Upper Saddle River, NJ: Prentice-Hall.
- Boik, R. J. (1979). Interactions, partial interactions and interaction contrasts in the analysis of variance, *Psychological Bulletin* **86**: 1084–1089.
- Bolfarine, H., Montenegro, L. C. & Lachos, V. H. (2007). Influence diagnostics for skew-normal linear mixed models, *Sankhya* **69**: 648–670.
- Box, G. E. P. (1950). Problems in the analysis of growth and wear curves, *Biometrics* **6**: 362–389.
- Box, G. E. P. (1954a). Some theorems on quadratic forms applied in the study of analysis of variance problems, I: effects of inequality of variance in the one-way classification, *Annals of Mathematical Statistics* **25**: 290–302.
- Box, G. E. P. (1954b). Some theorems on quadratic forms applied in the study of analysis of variance problems, II: effects of inequality of variance and of correlation between errors in the two-way classification, *Annals of Mathematical Statistics* **25**: 484–498.
- Brody, S. (1945). *Bioenergetics and Growth*, New York: Reinhold Publishing Corporation.

- Brogan, D. R. & Kutner, M. H. (1980). Comparative analysis of pretest-posttest research designs, *The American Statistician* **34**: 229–232.
- Brunner, E., Domhof, S. & Langer, F. (2002). *Nonparametric analysis of longitudinal data in factorial experiments*, New York: Wiley.
- Brunner, E., Munzel, U. & Puri, M. L. (1999). Rank-score tests in factorial designs with repeated measures, *Journal of Multivariate Analysis* **70**: 286–317.
- Bryant, E. & Gillings, D. (1985). Statistical analysis of longitudinal repeated measures designs, *Biostatistics, Statistics in Biomedical, Public Health and Environmental Sciences*, P.K. Sen (ed). Amsterdam: North Holland, 251–282, pp. 251–282.
- Cadigan, N. G. & Farrell, P. J. (2002). Generalized local influence with applications to fish stock cohort analysis, *Journal of Applied Statistics* **51**: 469–483.
- Calinski, T. & Caussinus, H. (1989). A note on the analysis of covariance: efficiency of concomitant variables, *Journal of Statistical Planning and Inference* **21**: 315–326.
- Chatterjee, S. & Hadi, A. S. (1986). Influential observations, high leverage points, and outliers in linear regression (with discussion), *Statistical Science* **1**: 379–393.
- Chatterjee, S. & Hadi, A. S. (1988). *Sensitivity analysis in linear regression*, New York: Wiley.
- Chesher, A. (1991). The effect of measurement error, *Biometrika* **78**: 451–462.
- Christensen, R., Pearson, L. M. & Johnson, W. (1992). Case-deletion diagnostics for mixed models, *Technometrics* **34**: 38–45.
- Cochran, W. G. & Cox, M. G. (1992). *Experimental designs*, 2 edn, New York: Wiley.
- Collier, R. O., Baker, F. B., Mandeville, G. R. & Hayes, T. F. (1967). Estimates of test size for several test procedures based on convention variance ratios in the repeated measures design, *Psychometrika* **32**: 339–353.
- Cook, R. D. (1977). Detection of influential observation in linear regression, *Technometrics* **19**: 15–18.
- Cook, R. D. (1986). Assessment of local influence (with discussion), *Journal of the Royal Statistical Society B* **48**: 133–169.

- Cook, R. D. (1996). Added variables plots and curvature in linear regression, *Technometrics* **38**: 275–278.
- Cook, R. D., Peña, D. & Weisberg, S. (1988). The likelihood: a unifying principle for influence measures, *Communications in Statistics, Theory and Methods* **17**: 623–640.
- Cook, R. D. & Weisberg, S. (1982). *Residuals and influence regression*, New York: Chapman and Hall.
- Cook, R. D. & Weisberg, S. (1989). Regression diagnostics with dynamic graphics, *Technometrics* **31**: 277–311.
- Cox, D. G. (1992). *Planning of experiments*, 2 edn, New York: Wiley.
- Cox, D. G. & Hinkley, D. V. (1974). *Theoretical Statistics*, London: Chapman & Hall.
- Cox, D. R. & Snell, E. J. (1968). A general definition of residuals (with discussion), *Journal of the Royal Statistical Society B* **30**: 248–275.
- Crowder, M. J. & Hand, J. (1990). *Analysis of repeated measures*, London: Chapman and Hall.
- Crowder, M. J. & Hand, J. (1996). *Practical longitudinal data analysis*, London: Chapman and Hall.
- Davidian, M. & Giltinian, D. M. (1995). *Nonlinear models for repeated measurement data*, London: Chapman and Hall.
- Davis, C. S. (2002). *Statistical methods for the analysis of repeated measures*, New York: Springer.
- Demidenko, E. (2013). *Mixed models: theory and applications with R*, 2 edn, New York: Wiley.
- Demidenko, E. & Stukel, T. A. (2005). Influence analysis for linear mixed-effects models, *Statistics in Medicine* **24**: 893–909.
- Dempster, A. P., Laird, N. M. & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm, *Journal of the Royal Statistical Society, B* **39**: 1–38.
- Diggle, P. J. (1988). An approach to the analysis of repeated measurements, *Biometrics* **44**: 959–971.

- Diggle, P. J., Heagerty, P., Liang, K. Y. & Zeger, S. L. (2002). *Analysis of longitudinal data*, 2 edn, Oxford: Oxford University Press.
- Doganaksoy, N. & Balakrishnan, N. (1997). A useful property of best linear unbiased predictors with applications to life-testing, *The American Statistician* **51**: 22–28.
- Draper, N. R. & Smith, H. (2002). *Applied regression analysis*, 3 edn, New York: Wiley.
- Duncan, G. J. & Kalton, G. (1987). Issues of design and analysis of surveys across time, *International Statistical Review* **55**: 97–117.
- Durbin, J. & Watson, G. S. (1950). Testing for serial correlation in least squares regression, I, *Biometrika* **37**: 409–428.
- Durbin, J. & Watson, G. S. (1951). Testing for serial correlation in least squares regression, II, *Biometrika* **38**: 159–178.
- Durbin, J. & Watson, G. S. (1971). Testing for serial correlation in least squares regression, III, *Biometrika* **58**: 1–19.
- Dwyer, J. (1992). Differential equation models for longitudinal data, *In Statistical Models for Longitudinal Studies of Health*, J.H. Dwyer, M. Feinleib, P. Lippert and H. Hoffmeister (eds). New York: Oxford University Press, pp. 71–98.
- Elian, S. N. & Okaze, S. M. (1998). Mudanças de desempenho em atividades motoras da vida diária em idosos participantes de um programa de educação física, *Technical Report RAE-CEA-98P07*, Centro de Estatística Aplicada, IME, Universidade de São Paulo.
- Elias, F. M., Birman, E. G., Matsuda, C. K., Oliveira, I. R. S. & Jorge, W. A. (2006). Ultrasonographic findings in normal temporomandibular joints, *Brazilian Oral Research* **20**: 25–32.
- Fei, Y. & Pan, J. (2003). Influence assessments for longitudinal data in linear mixed models, *18 th International Workshop on Statistical Modelling*, G. Verbeke, G. Molenberghs, M. Aerts and S. Fieuws (eds). Leuven: Belgium, pp. 143–148.
- Freedman, D. (2005). *Statistical Models: Theory and Practice*, Cambridge: Cambridge University Press.
- Fuller, W. A. (1987). *Measurement error models*, New York: Wiley.

- Fung, W. K., Zhu, Z. Y., Wei, B. C. & He, X. (2002). Influence diagnostics and outliers tests for semiparametric mixed models, *Journal of the Royal Statistical Society, B* **64**: 565–579.
- Galecki, T. G. (1994). General class of covariance structures for two or more repeated factors in longitudinal data analysis, *Communications in Statistics, Theory and Methods* **23**: 3105–3119.
- Geisser, S. (1970). A bayesian analysis of growth curves, *Sankhya A* **32**: 53–64.
- Geisser, S. (1980). Growth curve analysis, *Handbook of Statistics, Vol. 1, P. R. Krishnaiah (ed). Amsterdam: North Holland. New York: Oxford University Press*, pp. 89–115.
- Geisser, S. & Greenhouse, S. W. (1958). An extension of box's results on the use of the f-distribution in multivariate analysis, *Annals of Mathematical Statistics* **29**: 885–891.
- Giampaoli, V. & Singer, J. M. (2009). Generalized likelihood ratio tests for variance components in linear mixed models, *Journal of Statistical Planning and Inference* **139**: 1435–1448.
- Goldstein, H. (1979). *The design and analysis of longitudinal studies*, New York: Academic Press.
- Gould, A. L. (1973). A new approach to the analysis of clinical drug trials with withdrawals, *Biometrics* **36**: 721–727.
- Grady, J. J. & Helms, R. W. (1995). Model selection techniques for the covariance matrix for incomplete longitudinal data, *Statistics in Medicine* **14**: 1397–1416.
- Grande, R. H. M., Oliveira, E., Singer, J. M., Santos, J. F. F. & Nicolau, J. (1998). Efeitos do uso contínuo de solução para bochecho sobre o pH e o conteúdo mineral da placa bacteriana, *Revista da Pós-Graduação da Faculdade de Odontologia da Universidade de São Paulo* **5**: 143–147.
- Graybill, F. A. (1976). *Theory and applications of the linear model*, North Scituate, Mass.: Duxbury Press.
- Greenhouse, S. W. & Geisser, S. (1959). On methods in the analysis of profile data, *Psychometrika* **24**: 95–112.
- Grieve, A. P. (1981). Pretest-posttest designs: Letter to the editor, *The American Statistician* **35**: 177–178.

- Grizzle, J. E. & Allen, D. M. (1969). Analysis of dose and response curves, *Biometrics* **25**: 357–382.
- Gumedze, F. N., Welham, S. J., Gogel, B. J. & Thompson, R. (2010). A variance shift model for detection of outliers in the linear mixed model, *Computational Statistics and Data Analysis* **54**: 2128–2144.
- Harville, D. A. (1977a). Extension of the Gauss-Markov theorem to include the estimation of random effects, *The Annals of Statistics* **4**: 384–395.
- Harville, D. A. (1977b). Maximum likelihood approaches to variance component estimation and to related problems, *Journal of the American Statistical Association* **72**: 320–340.
- Harville, D. A. (1997). *Matrix algebra from a statistician's perspective*, New York: Springer.
- Helms, R. W. (1992). Intentionally incomplete longitudinal designs: I. methodology and comparison of some full span designs, *Statistics in Medicine* **11**: 1889–1913.
- Henderson, C. R. (1975). Best linear unbiased estimation and prediction under a selection model, *Biometrics* **31**: 423–447.
- Hilden-Minton, J. A. (1995). *Multilevel diagnostics for mixed and hierarchical linear models*, PhD thesis, Los Angeles: University of California.
- Hoaglin, D. & Welsch, R. (1978). The hat matrix in regression and anova, *The American Statistician* **32**: 17–22.
- Hodges, J. S. (1998). Some algebra and geometry for hierarchical models, applied to diagnostics (with discussion), *Journal of the Royal Statistical Society, Series B* **60**: 197–236.
- Hosmer, D. W. & Lemeshow, S. (2000). *Applied Logistic Regression*, 2 edn, New York: John Wiley.
- Huynh, H. & Feldt, L. S. (1970). Conditions under which mean square ratios in repeated measurements designs have exact F-distributions, *Journal of the American Statistical Association* **65**: 1582–1589.
- Huynh, H. & Feldt, L. S. (1976). Estimation for the Box correction for degrees of freedom from sample data in the randomized block and split-plot designs, *Journal of Educational Statistics* **1**: 69–82.
- Izeman, A. J. & Williams, J. S. (1989). A class of linear spectral models and analyses for the study of longitudinal data, *Biometrics* **45**: 831–849.

- Jara, A., Quintana, F. & San Martín, E. (2008). Linear mixed models with skew-elliptical distributions: a bayesian approach, *Computational Statistics and Data Analysis* **52**: 5033–5045.
- Jennrich, R. I. & Schluchter, M. D. (1986). Unbalanced repeated measures models with structured covariance matrices, *Biometrics* **42**: 805–820.
- Jiang, J. (1996). Reml estimation: asymptotic behavior and related topics, *The Annals of Statistics* **24**: 255–286.
- Jiang, J. (1997). A derivation of BLUP - Best Linear Unbiased Predictor, *Statistics & Probability Letters* **32**: 321–324.
- Jiang, J. (1998). Asymptotic properties of the empirical BLUP and BLUE in mixed linear models, *Statistica Sinica* **8**: 861–885.
- Jiang, J. (1999). On unbiasedness of the empirical BLUE and BLUP, *Statistics & Probability Letters* **41**: 19–24.
- Jones, R. H. (1993). *Longitudinal data with serial correlation: a state-space approach*, London: Chapman and Hall.
- Kackar, R. N. & Harville, D. A. (1984). Approximations for standard errors of estimators of fixed and random effects in mixed linear models, *Journal of the American Statistical Association* **79**: 853–862.
- Kass, R. E. & Steffey, D. (1989). Approximate Bayesian inference in conditionally independent hierarchical models (parametric empirical Bayes models), *Journal of the American Statistical Association* **84**: 717–726.
- Kenward, M. G. (1985). The use of fitted higher-order polynomial coefficients as covariates in the analysis of growth curves, *Biometrics* **41**: 19–28.
- Keselman, H. J., James, A., Rhonda, K. K. & Russell, D. W. (1998). A comparison of two approaches for selecting covariance structures in the analysis of repeated measurements, *Communications in Statistics - Simulation and Computation* **27**: 591–604.
- Khatri, C. G. (1966). A note on MANOVA model applied to problems in growth curves, *Annals of the Institute of Statistical Mathematics* **18**: 75–86.
- Korytnicki, D., Mayer, M. P. A., Daronch, M., Singer, J. M. & Grande, R. H. M. (2006). Effects of nd:yag laser on enamel microhardness and dental plaque composition: an in situ study, *Photomedicine and laser surgery* **24**: 59–63.

- Kshirsagar, A. M. & Smith, W. B. (1995). *Growth curves*, New York: Marcel Dekker.
- Kutner, M. & Brogan, D. (1981). Comparative analysis of pretest-posttest designs: Letter to the editor, *The American Statistician* **35**: 329–330.
- Kutner, M., Nachtsheim, C. J., Neter, J. & Li, W. (2005). *Applied Linear Statistical Models*, 5 edn, Boston: McGraw-Hill.
- Lachos, V. H., Ghosh, P. & Arellano-Valle, R. B. (2010). Likelihood based inference for skew-normal independent linear mixed models, *Statistica Sinica* **20**: 303–322.
- Laird, N. (1983). Further comparative analyses of pretest-posttest research designs, *The American Statistician* **37**: 179.
- Laird, N. M., 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. M. & Ware, J. H. (1982). Random-effects models for longitudinal data, *Biometrics* **38**: 963–974.
- Lee, J. C. (1988). Prediction and estimation of growth curves with special covariance structures, *Journal of the American Statistical Association* **83**: 432–440.
- Lee, J. C. & Geisser, S. (1972). Growth curve prediction, *Sankhya A* **34**: 393–412.
- Lencina, V. B. & Singer, J. M. (2006). Measure for measure: exact F-tests and the mixed models controversy, *International Statistical Review* **74**: 391–402.
- Lencina, V. B., Singer, J. M. & Stanek III, E. J. (2005). Much ado about nothing: the mixed models controversy revisited, *International Statistical Review* **73**: 9–20.
- Lencina, V. B., Singer, J. M. & Stanek III, E. J. (2008). Response to J.A. Nelder: What is the mixed models controversy?, *International Statistical Review* **76**: 134–139.
- Lesaffre, E. & Lawson, A. B. (2012). *Bayesian Biostatistics*, West Sussex: Wiley.
- Lesaffre, E. & Verbeke, G. (1998). Local influence in linear mixed models, *Biometrics* **54**: 570–582.
- Levin, J. R. (1981). Pretest-posttest designs: Letter to the editor, *The American Statistician* **35**: 178–179.

- Lindsey, J. K. (1999). *Models for repeated measurements*, 2 edn, Oxford: Oxford University Press.
- Lindstrom, M. J. & Bates, D. M. (1988). Newton-raphson and EM algorithms for linear mixed-effects models for repeated-measures data, *Journal of the American Statistical Association* **83**: 1014–1022.
- Magnus, J. R. & Neudecker, H. (1988). *Matrix differential calculus with applications in Statistics and Econometrics*, New York: Wiley.
- Maritz, J. S. & Lwin, T. (1989). *Empirical Bayes methods*, 2 edn, London: Chapman & Hall.
- Mauchly, J. W. (1940). Significance test for sphericity of a normal n-variate distribution, *The Annals of Mathematical Statistics* **11**: 204–209.
- McCulloch, C. E. & Searle, S. R. (2001). *Generalized, linear, and mixed models*, New York: Wiley.
- Molenberghs, G. & Verbeke, G. (2000). *Linear mixed models for longitudinal data*, New York: Springer.
- Molenberghs, G. & Verbeke, G. (2005). *Models for discrete longitudinal data*, New York: Springer.
- Morettin, P. A. & Toloi, C. M. C. (2006). *Análise de séries de tempo*, 2 edn, São Paulo: Edgard Blücher.
- Morrison, D. F. (1988). *Multivariate statistical methods*, 2 edn, New York: McGraw-Hill.
- Mudholkar, G. S. & Hutson, A. D. (2000). The epsilon-skew-normal distribution for analyzing near-normal data, *Journal of Statistical Planning and Inference* **83**: 291–309.
- Mun, J. & Lindstrom, M. J. (2013). Diagnostics for repeated measurements in linear mixed models, *Statistics in Medicine* **32**: 1361–1375.
- Nelder, J. A. (1998). The great mixed-model muddle is alive and flourishing, alas!, *Food quality and preference* **9**: 157–159.
- Nobre, J. S. & Singer, J. M. (2007). Residual analysis for linear mixed models, *Biometrical Journal* **49**: 863–875.
- Nobre, J. S. & Singer, J. M. (2011). Leverage analysis for linear mixed models, *Journal of Applied Statistics* **38**: 1063–1072.

- Osorio, F., Paula, G. A. & Galea, M. (2007). Assessment of local influence in elliptical linear models with longitudinal structure, *Computational Statistics and Data Analysis* **51**: 4354–4368.
- Patterson, H. D. & Thompson, R. (1971). Recovery of interblock information when block sizes are unequal, *Biometrika* **58**: 545–554.
- Paula, G. A. (2004). *Modelos de regressão com apoio computacional*, São Paulo: IME/USP.
- Paulino, C. D. & Singer, J. M. (2006). *Análise de dados categorizados*, São Paulo: Blucher.
- Pinheiro, J. C., Liu, C. & Wu, Y. N. (2001). Efficient algorithms for robust estimation in linear mixed-effects models using the multivariate t -distribution, *Journal of Computational and Graphical Statistics* **10**: 249–276.
- Potthoff, R. F. & Roy, S. N. (1964). A generalized multivariate analysis of variance model useful especially for growth curve problems, *Biometrika* **51**: 313–326.
- Pregibon, D. (1981). Logistic regression diagnostics, *Annals of Statistics* **9**: 705–724.
- Rao, C. R. (1959). Some problems involving linear hypotheses in multivariate analysis, *Biometrika* **46**: 49–58.
- Rao, C. R. (1965). The theory of least squares when the parameters are stochastic and its application to the analysis of growth curves, *Biometrika* **52**: 447–458.
- Rao, C. R. (1966). Covariance adjustment and related problems in multivariate analysis, *Multivariate Analysis II. P.R. Krishnaiah (ed). New York: Academic Press*, pp. 87–103.
- Rao, C. R. (1967). Least squares theory using an estimated dispersion matrix and its application to measurement of signals, *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability, Vol. I*, pp. 355–372.
- Rao, C. R. (1987). Prediction of future observations in growth curve models, *Statistical Science* **4**: 434–471.
- Rao, M. N. & Rao, C. R. (1966). Linked cross-sectional study for determining norms and growth rates - a pilot survey on indian school-going boys, *Sankhya, B* **28**: 237–258.
- Ratkowsky, D. A., Aldredge, R. & Evans, M. A. (1992). *Cross-over experiments*, New York: Marcel Dekker.

- Reis, A., Loguercio, A. D., Azevedo, C. L., Carvalho, R. M., Singer, J. M. & Grande, R. H. (2003). Moisture spectrum of demineralized dentin for adhesive systems with different solvent bases, *Journal of Adhesive Dentistry* **5**: 183–192.
- Robinson, G. K. (1991). The estimation of random effects (with discussion), *Statistical Science* **6**: 15–51.
- Rocha, F. M. M. (2004). *Seleção de estruturas de covariância para dados com medidas repetidas*, Master's thesis, Departamento de Estatística, Universidade de São Paulo, Brasil.
- Rocha, F. M. M. & Singer, J. M. (2017). Selection of terms in random coefficient regression models, *Journal of Applied Statistics* pp. ??–??
- Rouanet, H. & Lépine, D. (1970). Comparison between treatments in a repeated measurement design: ANOVA and multivariate methods, *The British Journal of Mathematical and Statistical Psychology* **23**: 147–163.
- Rowell, J. G. & Walters, D. E. (1976). Analysing data with repeated observations on each experimental unit, *Journal of Agricultural Science* **87**: 423–432.
- Savalli, C., Paula, G. A. & Cysneiros, F. J. A. (2006). Assessment of variance components in elliptical linear mixed models, *Statistical Modelling* **6**: 59–76.
- Schafer, W. D. (1981). Pretest-posttest designs: Letter to the editor, *The American Statistician* **35**: 179.
- Schützenmeister, A. & Piepho, H.-P. (2012). Residual analysis of linear mixed models using a simulation approach, *Computational Statistics and Data Analysis* **56**: 1405–1416.
- Searle, S. R. (1971). *Linear models*, New York: Wiley.
- Searle, S. R. (1982). *Matrix algebra useful for Statistics*, New York: Wiley.
- Searle, S. R., Casela, G. & McCulloch, C. E. (1992). *Variance components*, New York: Wiley.
- Seber, G. A. F. & Wild, C. J. (1989). *Nonlinear regression*, New York: Wiley.
- Self, S. G. & Liang, K. Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions, *Journal of the American Statistical Association* **82**: 605–610.
- Sen, P. K., Singer, J. M. & Pedroso-de Lima, A. C. (2009). *From finite sample to asymptotic methods in Statistics*, Cambridge: Cambridge University Press.

- Singer, J. M. & Andrade, D. F. (1994). On the choice of appropriate error terms in profile analysis, *The Statistician* **43**: 259–266.
- Singer, J. M. & Andrade, D. F. (2000). Analysis of longitudinal data, *Handbook of Statistics, Volume 18: Bio-environmental and Public Health Statistics*. Eds. P.K. Sen and C.R. Rao, pp. 115–160.
- Singer, J. M. & Cúri, M. (2006). Modelling regression and dispersion parameters in a complex repeated measures experiment, *Environmental and Ecological Statistics* **13**: 53–68.
- Singer, J. M. & Magalhães, L. M. (1998). Efeito do só dio na pressão arterial em hipertensos essenciais, *Technical Report RAE-CEA-97P37*, Centro de Estatística Aplicada, IME, Universidade de São Paulo.
- Singer, J. M. & Polli, D. A. (2004). Avaliação clínica de dois adesivos dentinários 18 meses após aplicação em dentina seca ou úmida, *Technical Report CEA-RAE-04P12*, Centro de Estatística Aplicada, IME, Universidade de São Paulo.
- Singer, J. M., Seoanes, M. & Ogando, M. A. (1988). Estudo do efeito da infusão aguda de $mgso_4$ nas funções hemodinâmicas e renais em cães, *Technical Report SEA-RAE*, Centro de Estatística Aplicada, IME, Universidade de São Paulo.
- Soler, J. M. P. & Singer, J. M. (2000). Optimal covariance adjustment in growth curve models, *Computational Statistics and Data Analysis* **33**: 101–110.
- Souza, G. S. (1998). *Introdução aos modelos de regressão linear e não-linear*, Brasília: EMBRAPA.
- Stanek III, E. J. (1988). Choosing a pretest-posttest analysis, *The American Statistician* **42**: 178–183.
- Stoloff, P. H. (1970). Correcting for heterogeneity of covariance for repeated measures designs of the analysis of variance, *Educational and Psychometrical Measurements* **30**: 909–924.
- Stram, D. O. & Lee, J. W. (1994). Variance components testing in the longitudinal mixed effects model, *Biometrics* **50**: 1171–1177.
- Tan, F. E. S., Ouwers, M. J. N. & Berger, M. P. F. (2001). Detection of influential observations in longitudinal mixed effects regression models, *The Statistician* **50**: 271–284.
- Tountenburg, H. (1982). *Prior information in linear models*, New York: Wiley.

- Tuerlinckx, F., Rijmen, F. & Verbeke, G. and De Boeck, P. (2006). Statistical inference in generalized linear mixed models: A review., *British Journal of Mathematical and Statistical Psychology* **59**: 225–255.
- Venezuela, M. K., Botter, D. A. & Sandoval, M. (2007). Diagnostic techniques in generalized estimating equations, *Journal of Statistical Computation and Simulation* **77**: 879–888.
- Venezuela, M. K., Sandoval, M. & Botter, D. A. (2011). Local influence in estimating equations, *Journal of Statistical Computation and Simulation* **55**: 1867–1883.
- Verbeke, G. & Lesaffre, E. (1996). Large sample properties of the maximum likelihood estimators in linear mixed models with misspecified random effects distributions, *Technical report*, Biostatistical Centre for Clinical Trials, Catholic University of Leuven, Belgium.
- Verbeke, G. & Molenberghs, G. (1997). *Linear mixed models in practice: a SAS-oriented approach*, New York: Springer.
- Vonesh, E. F. & Carter, R. L. (1987). Efficient inference for random-coefficient growth curve models with unbalanced data, *Biometrics* **42**: 601–610.
- Vonesh, E. F. & Chinchilli, V. M. (1997). *Linear and nonlinear models for the analysis of repeated measurements*, New York: Marcel Dekker.
- Voss, D. T. (1999). Resolving the mixed model controversy, *The American Statistician* **53**: 352–356.
- Vu, H. T. V. & Zhou, S. (1997). Generalization of likelihood ratio tests under nonstandard conditions, *The Annals of Statistics* **25**: 897–916.
- Wei, B. C., Hu, Y. Q. & Fung, W. K. (1998). Generalized leverage and its applications, *Scandinavian Journal of Statistics* **25**: 25–37.
- Weiss, R. E. & Lazaro, C. G. (1992). Residual plots for repeated measures, *Statistics in Medicine* **11**: 115–124.
- West, B. T. & Galecki, A. T. (2011). An overview of current software procedures for fitting linear mixed models, *The American Statistician* **65**: 274–282.
- Wishart, J. (1938). Growth rate determination in nutrition studies with the bacon pig and their analysis, *Biometrika* **30**: 16–24.
- Xu, L., Lee, S.-Y. & Poon, W.-Y. (2006). Deletion measures for generalized linear mixed effects models, *Computational Statistics and Data Analysis* **51**: 1131–1146.

- Zeger, S. L., Liang, K. Y. & Albert, P. S. (1988). Models for longitudinal data: a generalized estimating equation approach, *Biometrics* **44**: 1049–1060.
- Zerbe, G. O. & Walker, J. H. (1977). A randomization test for comparison of groups of growth curves with different polynomial design matrices, *Biometrics* **33**: 653–657.
- Zewotir, T. (2008). Multiple cases deletion diagnostics for linear mixed models, *Communications in Statistics - Theory and Methods* **37**: 1071–1084.
- Zewotir, T. & Galpin, J. S. (2005). Influence diagnostics for linear mixed models, *Journal of Data Science* **3**: 153–177.
- Zewotir, T. & Galpin, J. S. (2007). A unified approach on residuals, leverages and outliers in linear mixed models, *163* pp. 58–75.
- Zhu, H. T., Ibrahim, J. G., Lee, S. & Zhang, H. (2007). Perturbation selection and influence measures in local influence analysis, *The Annals of Statistics* **35**: 2565–2588.

Índice

- Area under the curve*, 28
 - Case deletion*, 61
 - Crossover design*, 2
 - Draftman's plot*, 23, 236
 - Dropout*, 29
 - Endpoint*, 7, 28
 - Fisher scoring*, 58
 - Fit function*
 - displacement*, 88
 - Fit function*, 88
 - Hat matrix*, 216
 - Least confounded residuals*, 63
 - Leverage analysis*, 61
 - Linked design*, 5
 - Local influence*, 61
 - Missing data*, 30
 - Peak*, 28
 - Population-averaged*, 7
 - Response profile*, 4
 - Skew-normal*, 76
 - Split-plot model*, 113
 - Subject-specific*, 7
 - Time dependent covariate*, 4
 - Tracking*, 4
 - Working covariance matrix*, 85
 - Acurácia, 253
 - Algoritmo
 - de mínimos quadrados iterativamente ponderados, 86, 250
 - EM, 58
 - Newton-Raphson, 57, 248
 - Análise
 - de covariância, 30
 - de dados longitudinais, 212
 - de desfecho, 29
 - de medidas repetidas, 212
 - de regressão, 4
 - de séries cronológicas, 229
 - de séries de tempo, 229
 - de variância, 200
 - ANCOVA, 5, 119
 - ANOVA, 4, 5, 200
 - para medidas repetidas, 9, 88
 - Area sob a curva, 7, 28, 31
 - Autocorrelação, 231
 - Chance, 249
 - Coefficiente
 - de correlação intraclasse, 138
 - de determinação, 213
 - de determinação ajustado, 213
 - de regressão, 195
 - Combinação linear, 162
 - Componentes principais, 240
 - Condição de Noether, 209
 - Confundimento, 61
 - Constante
 - de encolhimento, 139
 - Controvérsia dos modelos mistos, 113
 - Covariável
 - dependente do tempo, 4, 5
 - Critério de informação
 - Akaike, 51
 - Schwarz, 51
 - Curva
 - de crescimento, 9
 - de influência, 222
-

- ROC, 252
- Curvatura normal, 70
- Dados
- omissos, 30
- Desfecho, 7, 28
- Desigualdade
- de Cauchy-Schwarz, 172
 - triangular, 172
- Desistência, 29
- Deslocamento
- da função de ajuste, 88
 - de verossimilhança, 70
- DFFITS, 226
- Diagnóstico, 214
- análise de sensibilidade, 214
 - avaliação do ajuste, 214
- Distância
- de Cook, 87, 225
 - de Cook condicional, 68
 - de Mahalanobis, 74
- Distribuição
- elipticamente assimétrica, 73
 - elipticamente simétrica, 73
 - potência exponencial, 73
- Efeito, 240
- de tratamento, 241, 242
 - principal, 243
- Elipsóide de confiança, 227
- Elipsoide de confiança, 222
- Envelope simulado, 87
- Equação
- de estimação, 206
 - de estimação generalizada, 10, 85
 - de Henderson, 48
 - normal, 206
- Espaço
- coluna, 163
- Espaço vetorial, 170
- complemento ortogonal, 172
 - dimensão, 170
 - espaço coluna, 171
 - espaço euclidiano, 171
 - espaço nulo, 171
 - subespaço vetorial, 170
- Especificidade, 252
- Estatística
- de Wald, 50
 - da razão de verossimilhanças, 50
 - de Durbin-Watson, 232
 - de Lawley-Hotelling, 53
 - de Pillai, 53
 - de Roy, 53
 - de Wald, 210, 211
 - de Wilks, 53
- Estimador
- BLUE, 47, 207
 - de máxima verossimilhança restrita, 46
 - de mínimos quadrados generalizados, 44
 - de quase-verossimilhança penalizado, 84
 - EBLUE, 49
 - linear não enviesado de variância mínima, 207
 - sanduíche, 86
- Estudo
- de coorte, 2
 - longitudinal, 2
 - tipo painel, 2
 - transversal, 2
- Expansão
- de Taylor, 191
 - de Taylor multivariada, 192
- Falsos positivos, 254
- Fator
- de inflação da variância, 239
- Filtro de Kalman, 58
- Forma
- bilinear, 175

-
- linear, 175
 - quadrática, 175, 206
 - Fração de confundimento, 62
 - Função
 - de ajuste, 88
 - de ligação, 78, 85
 - escore, 58
 - estimável, 51
 - geradora de densidade, 76
 - gradiente, 178
 - previsível, 51
 - GEE, 10
 - Gráfico
 - da variável adicionada, 225
 - de perfis, 25, 109, 132
 - de perfis médios, 25
 - de regressão parcial, 225
 - do desenhista, 234
 - QQ, 218
 - Heterocedasticidade, 25, 27
 - Homocedasticidade, 215
 - Independência linear, 162
 - Influência
 - global, 222
 - global local, 71
 - local, 222
 - Interação, 203, 243
 - essencial, 111, 124, 243
 - não essencial, 243
 - não-essencial, 124
 - Logito, 78, 248
 - Média geral, 241
 - Método
 - da união-intersecção, 53
 - de Laplace, 81
 - de quase-verossimilhança, 82
 - de quase-verossimilhança marginal, 82
 - de quase-verossimilhança penalizada, 82
 - bayesiano, 44
 - da máxima verossimilhança restrita, 44
 - de ajuste por covariância, 54
 - de máxima verossimilhança, 44, 211
 - de máxima verossimilhança restrita, 212
 - de mínimos quadrados, 206, 215, 263
 - de mínimos quadrados generalizados, 210, 232, 237
 - de mínimos quadrados ponderados, 210
 - de Monte Carlo, 81
 - Delta, 29, 250, 260
 - MANOVA, 52, 88, 111
 - Matriz, 157
 - autovalor, 174
 - autovetor, 174
 - base, 173
 - base ortonormal, 173
 - chapéu, 61, 87, 216
 - circular, 55
 - cofator, 163
 - de alavanca conjunta generalizada, 65
 - de alavanca de componentes aleatórios generalizada, 66
 - de alavanca marginal generalizada, 65
 - de correlações amostrais, 24
 - de correlações de trabalho, 10, 88
 - de covariâncias amostrais, 24
 - de covariâncias de trabalho, 85
 - de informação de Fisher, 75, 249
 - de informação empírica, 58
 - de posto completo, 163
 - de predição, 216
 - de projeção, 216
 - definida não negativa, 175
 - definida positiva, 175
 - definida semipositiva, 175
-

- determinante, 163
diagonal, 160
diagonal em blocos, 166
esférica, 55
hessiana, 58, 179
idempotente, 161
identidade, 161
inversa, 164
inversa generalizada, 165, 207
jacobiana, 179
menor, 163
menor principal, 163
multiplicação por escalar, 158
não singular, 163, 164
norma, 172
operador vec, 168
operador vech, 169
particionada, 162
posto, 163
produto de, 159
produto de Kronecker, 167
produto direto, 167
produto tensorial, 167
quadrada, 160
raiz característica, 174
simétrica, 160
soma de, 158
soma direta, 167
submatriz, 161
traço, 166
transposta, 160
triangular, 161
triangular inferior, 161
triangular superior, 161
vetor característico, 174
vetorização, 168
- Medida resumo, 7, 28
- Modelo
- com erros de medida, 197
 - condicional, 8
 - de análise de perfis, 52
 - de coeficientes aleatórios, 57
 - de crescimento, 9
 - de curvas de crescimento, 54
 - de independência condicional, 38
 - de parâmetros irrestritos, 113
 - de parâmetros restritos, 113
 - de perfis, 9
 - de regressão linear múltipla, 196
 - de regressão linear simples, 196
 - de regressão polinomial, 196
 - de regressão segmentada, 204
 - em dois estágios, 38
 - heterocedástico, 210
 - homocedástico, 210
 - incondicional, 8
 - individual, 7
 - inidentificável, 241
 - linear generalizado, 250
 - linear generalizado misto, 10
 - linear misto, 38
 - linear misto assimétrico, 10
 - linear misto elíptico, 10
 - marginal, 38
 - mecanístico, 9
 - misto, 112, 237
 - normal assimétrico, 76
 - populacional médios, 7
- Movimento browniano, 10
- Multicolinearidade, 147, 207
- Norma
- de Frobenius, 172
- Observação
- discrepante, 215, 217
 - omissa, 5
- Omissão de casos, 66
- Operador diagonal, 160
- Parâmetro
- de assimetria, 76
 - de localização, 37

- de não centralidade, 176
- estimável, 241
- não estimável, 207
- Parametrização
 - de cela de referência, 241
 - de desvios de médias, 241
 - de desvios médios, 112
 - de médias de celas, 240
- Perfil
 - de resposta, 4
- Pico, 28
- Planejamento
 - balanceado no tempo, 5, 11, 13, 15
 - com intercâmbio, 2
 - com lotes subdivididos, 2
 - encadeado, 5
 - split-plot, 2
 - transversal misto, 5
- Poder de alavanca, 64, 222
- Polinômio
 - ortogonais, 54
- Ponto alavanca, 61
- Ponto de corte, 252
- Preditor
 - BLUP, 47
 - EBLUP, 49
- Processo de Ornstein-Ühlenbeck, 10
- Quadratura
 - de Gauss-Hermite, 81
- Razão de chances, 249
- Regressão
 - com erros nas variáveis, 197
 - logística, 248
- Resíduo
 - condicional, 61
 - envelope simulado, 219
 - marginal, 61
 - minimamente confundido, 63
 - ordinário, 216
 - padronizado, 216
 - predito, 218
 - recursivo, 218
 - studentizado, 216
 - studentizado externamente, 218
 - transformado, 63
- Restrição
 - de identificabilidade, 112, 241
- Série
 - de Taylor, 191
 - de tempo, 2
- Sensibilidade, 252
- Soma de quadrados
 - devida à regressão, 213
 - residual, 213
 - total, 213
- Supranormalidade, 63
- Teorema
 - de Hájek-Šidak, 209
 - de Sverdrup, 211
 - limite central, 209
- Transformação
 - linear, 170
- Trilhamento, 4, 27
- Unidade amostral
 - discrepante, 64, 74
- Variável
 - latente, 37
- Verossimilhança
 - residual, 46
 - restrita, 45
- Vetor, 158
 - distância euclidiana, 172
 - norma, 172
 - ortogonais, 172
 - ortonormal, 173
 - produto interno, 171
 - projeção ortogonal, 173
 - unitário, 172