

Análise de dados longitudinais

Versão parcial preliminar

março 2011

Julio M. Singer

Juvêncio S. Nobre

Francisco Marcelo M. Rocha

Departamento de Estatística

Universidade de São Paulo

Caixa Postal 66281

São Paulo, SP 05314-970

Brasil

Conteúdo

1	Introdução	1
1.1	Conceitos básicos	1
1.2	Exemplos	12
1.3	Análise descritiva e medidas resumo	24
2	Modelos lineares para dados gaussianos	31
2.1	Introdução	31
2.2	Modelos para a estrutura de covariância	33
2.3	Inferência por máxima verossimilhança	37
2.4	Solução das equações de estimação	47
2.5	Estratégias de análise	55
2.6	Diagnóstico	57
2.7	Notas de capítulo	57
2.8	Exercícios	57
3	Análise de dados	59
3.1	Estudos pré-teste/pós-teste	59
3.2	Análise de perfis	71
3.2.1	Uma estratégia para análise de perfis	76
A	Matrizes e espaços vetoriais	85
A.1	Matrizes	85
A.1.1	Operações básicas	86

<i>CONTEÚDO</i>	<i>CONTEÚDO</i>
A.1.2 Tipos especiais de matrizes	88
A.1.3 Submatrizes e matrizes particionadas	90
A.1.4 Independência linear e espaço-coluna	91
A.1.5 Determinante de uma matriz	92
A.1.6 Inversão de matrizes	93
A.1.7 Traço de uma matriz	95
A.1.8 Soma direta e produto de Kronecker	96
A.1.9 Operadores vec e vech	98
A.2 Tópicos de Álgebra Linear	99
A.3 Formas lineares, bilineares e quadráticas	105
A.4 Decomposição de matrizes	107
A.5 Derivadas de vetores e matrizes	108
A.6 Exercícios	117
B Regressão	121
B.1 Introdução	121
B.2 Método de mínimos quadrados	133
B.3 Método de máxima verossimilhança	139
B.4 Partição da soma de quadrados	140
B.5 Diagnóstico	141
B.5.1 Análise de Resíduos	143
B.5.2 Análise da suposição de normalidade	147
B.5.3 Análise de sensibilidade	151
B.5.4 Análise da suposição de correlação nula	159
B.6 Parametrização de modelos lineares	170
B.7 Regressão logística	175
B.8 Exercícios	175
Bibliografia	187

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 a 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 dados. 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 do tipo *split-plot* e **com intercâmbio** (*crossover*). Planejamento 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 à 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 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 experimentos 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), Bruner, Domhof & Langer (2002), Molenberghs & Verbeke (2000) e Molenberghs & Verbeke (2005), 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 mudanças globais ou individuais ao longo do tempo. pode ser interpretado como o efeito do tratamento k , 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 inter-indivíduos é maior do que a variabilidade intra-indivíduos. 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, por exemplo, uma situação em que o interesse recai na comparação das médias 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 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 médias das 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.

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 $\sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY}$.

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 tem variação intraunidades amostrais (hábitos tabagistas, exposição à poluentes etc.); estas são cognominadas **covariáveis dependentes do tempo** (*time dependent covariates*).

O vetor com as p_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

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

em que y_{ik} denota o valor de variável resposta para a i -ésima unidade amostral no k -ésimo instante de observação, ($k = 1, \dots, p_i$) e \mathbf{a}^\top denota o vetor \mathbf{a} com os elementos transpostos. A matriz $p_i \times m$ com os valores das variáveis explicativas pode ser expressa como

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

em que $\mathbf{x}_{ij} = (x_{ij1}, \dots, x_{ijp_i})^\top$ com x_{ijk} representando o valor da j -ésima

variável independente ($j = 1, \dots, m$) para a i -ésima unidade amostral observada no k -ésimo instante ($k = 1, \dots, p_i$).

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 pode ser encontrado 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_{1p_1}	t_{1p_1}	x_1	w_1	v_1	z_{1p_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_{2p_2}	t_{2p_2}	x_2	w_2	v_2	z_{2p_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_{np_n}	t_{np_n}	x_n	w_n	v_n	z_{np_n}

Neste caso, admitindo que os fatores X e W tenham dois níveis cada e que tanto V quanto Z sejam variáveis contínuas, as matrizes (1.1.2) seriam dadas por

$$\mathbf{X}_i = \begin{pmatrix} t_{i1} & x_i & w_i & v_i & z_{i1} \\ t_{i2} & x_i & w_i & v_i & z_{i2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ t_{ip_i} & x_i & w_i & v_i & z_{ip_i} \end{pmatrix}.$$

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 $(\sum_{i=1}^n p_i \times 1)$

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

e de uma matriz com dimensão $(n \sum_{i=1}^n p_i \times 5)$

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

Quando os dados são balanceados em relação ao tempo 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, assumimos que o fator X tem m 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 , $h = 1, \dots, m$, $i = 1, 2$, $j = 1, \dots, n_{hi}$ e $k = 1, \dots, p$.

Neste caso, é comum dispor os perfis de resposta

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

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

$$\mathbf{Y} = (\mathbf{y}_{111}, \mathbf{y}_{112}, \dots, \mathbf{y}_{11n_{11}}, \dots, \mathbf{y}_{m21}, \mathbf{y}_{m22}, \dots, \mathbf{y}_{m2n_{m2}})^\top.$$

As matrizes individuais de variáveis explicativas são geralmente agrupadas na forma de uma matriz com dimensão $n \times 2m$ dada, por exemplo, por

$$\mathbf{X} = \text{diag}(\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_m)$$

com

$$\mathbf{X}_h = \begin{pmatrix} \mathbf{1}_{n_{h1}} & \mathbf{0}_{n_{h1}} \\ \mathbf{1}_{n_{h2}} & \mathbf{0}_{n_{h2}} \end{pmatrix}.$$

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** (*end-point*). 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.

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	...	p
1	1	1	y_{1111}	y_{1112}	...	y_{111p}
1	1	2	y_{1121}	y_{1122}	...	y_{112p}
.
1	1	n_{11}	$y_{11n_{11}1}$	$y_{11n_{11}2}$...	$y_{11n_{11}p}$
1	2	1	y_{1211}	y_{1212}	...	y_{121p}
1	2	2	y_{1221}	y_{1222}	...	y_{122p}
.
1	2	n_{12}	$y_{12n_{12}1}$	$y_{12n_{12}2}$...	$y_{12n_{12}p}$
.
2	1	1	y_{2111}	y_{2112}	...	y_{211p}
2	1	2	y_{2121}	y_{2122}	...	y_{212p}
.
2	1	n_{21}	$y_{21n_{21}1}$	$y_{21n_{21}2}$...	$y_{21n_{21}p}$
2	2	1	y_{2211}	y_{2212}	...	y_{221p}
2	2	2	y_{2221}	y_{2222}	...	y_{222p}
.
2	2	n_{22}	$y_{22n_{22}1}$	$y_{22n_{22}2}$...	$y_{22n_{22}p}$
.
m	1	1	y_{m111}	y_{m112}	...	y_{m11p}
m	1	2	y_{m121}	y_{m122}	...	y_{m12p}
.
m	1	n_{m1}	$y_{m1n_{m1}1}$	$y_{m1n_{m1}2}$...	$y_{m1n_{m1}p}$
m	2	1	y_{m211}	y_{m212}	...	y_{m21p}
m	2	2	y_{m221}	y_{m222}	...	y_{m22p}
.
m	2	n_{m2}	$y_{m2n_{m2}1}$	$y_{m2n_{m2}2}$...	$y_{m2n_{m2}p}$

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.

Sob outra perspectiva, modelos para dados longitudinais podem ser classificados como **condicionais** ou **incondicionais**. Os primeiros são aqueles em que o valor esperado da variável resposta, $\mathbb{E}(y_{ik})$, é expresso exclusivamente em termos das variáveis explicativas x_{i1k}, \dots, x_{imk} . 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.

Duas abordagens distintas podem ser consideradas para o processo de geração 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 segunda é fundamentada em considerações teóricas sobre o fenômeno investigado e geralmente produz modelos baseados em equações diferenciais estocásticas. Modelos gerados por esse tipo de enfoque são usualmente não lineares e os parâmetros correspondentes (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 típico é aquele baseado no processo de **Ornstein-Ühlenbeck** que envolve soluções de equações do tipo

$$dY_i(t)/dt = -\alpha Y_i(t) + \sigma dB_i(t)/dt \quad (1.1.1)$$

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 & Giltinian (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, esse momentos de segunda ordem, em muitos casos, podem ser modelados independentemente dos momentos de primeira ordem facilitando a análise. Mais recentemente, métodos estatísticos baseados nos **modelos lineares generalizados** têm sido considerados para a análise de dados com distribuição na família exponencial, **modelos lineares elípticos** vêm sendo empregados para dados com distribuição elíptica e **modelos lineares assimétricos** para dados com distribuições assimétricas .

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 estudos 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 tornar 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 gênero (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 Elian & Okaze (1998).

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 (gênero) e de dados omissores, 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 velocidade de transporte mucociliar (mm/s), foi observada a cada cinco minutos após a imersão. A variável resposta de interesse 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.

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 planejamento balanceado em relação ao tempo e com omissão de dados.

Tabela 1.2.2: Tempo gasto para calçar meia (s)

Gênero	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				

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. O pH da placa dentária retirada de 21 voluntários antes e depois de um período de 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).

Tabela 1.2.4: pH da placa bacteriana dentária

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

Tabela 1.2.4: pH da placa bacteriana dentária

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

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 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 *textitin situ* (esmalte) e Laser (laser). Os blocos submetidos aos tratamentos Laser e Controle *in situ* foram colocados em aparelhos intra-bucais devidamente fabricados para cada um de oito voluntários e usados durante 28 dias consecutivos para sofrerem a ação de agentes cariogênicos. Du-

rante 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
2									

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 cardio-circulatórias 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)

Peso	Semanas pós-concepção													
	27	28	29	30	31	32	33	34	35	36	37	38	39	40
AIG	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.

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

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).

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

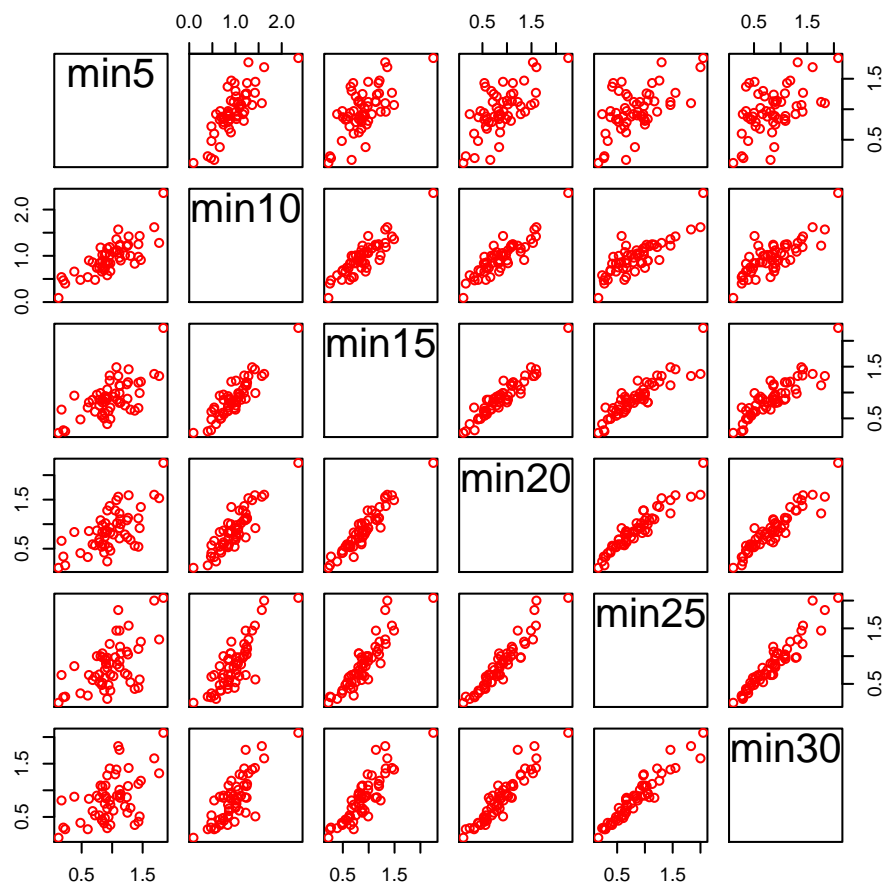
Os exemplos descritos acima ilustram o tipo de problemas que pretendemos avaliar nesta obra. Nos capítulos subsequentes não só discutiremos as técnicas comumente empregadas para analisá-los como também concretizaremos suas análises. Para discutir aspectos mais específicos do tema abordado, outros exemplos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

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 cognominados *draftman's plot* em inglês) como mostra a Figura 1.3.1 correspondente aos dados do Exemplo 1.2.3.

Figura 1.3.1: Gráficos de dispersão 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 (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

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

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

$$\hat{\mathbf{R}}_i = \hat{\mathbf{D}}_i^{-1/2} \hat{\mathbf{V}}_i \hat{\mathbf{D}}_i^{-1/2}.$$

em que $\hat{\mathbf{D}}_i = \text{diag}(\hat{\mathbf{V}}_i)$. 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

$$\hat{\mathbf{V}} = (n - t)^{-1} \sum_{i=1}^t (n_i - 1) \hat{\mathbf{V}}_i.$$

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

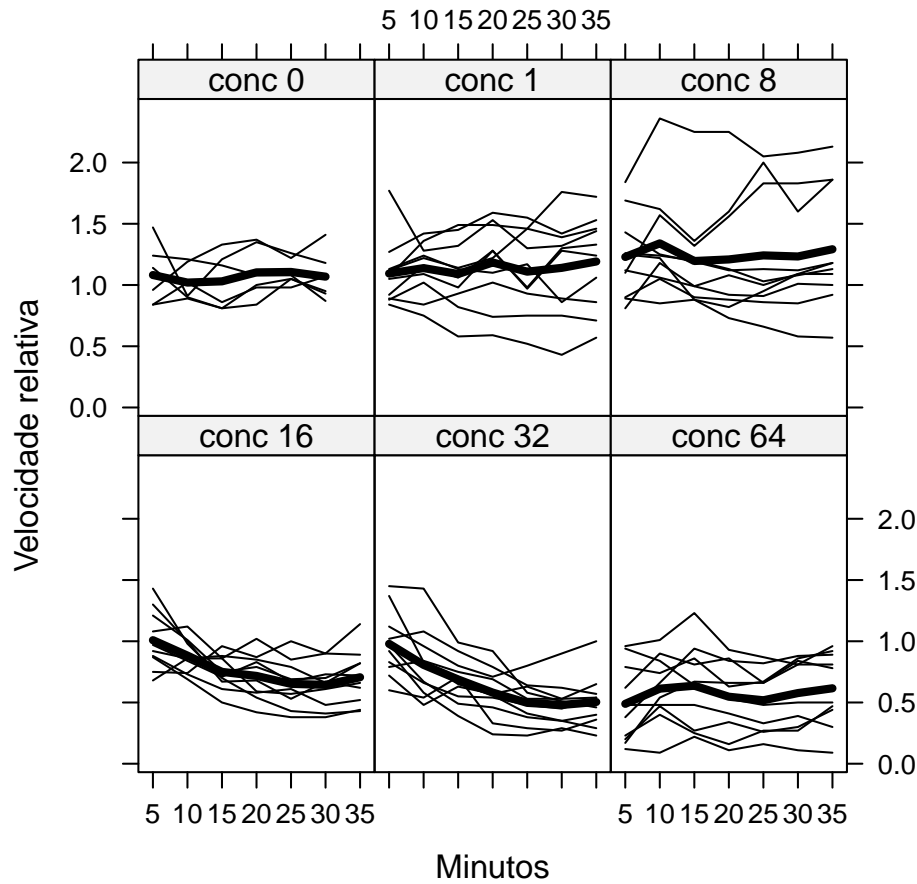
	min5	min10	min15	min20	min25	min30	min35
min5	1.00	0.75	0.59	0.58	0.47	0.41	0.37
min10		1.00	0.86	0.86	0.73	0.70	0.66
min15			1.00	0.92	0.84	0.80	0.78
min20				1.00	0.91	0.89	0.87
min25					1.00	0.92	0.94
min30						1.00	0.97
min35							1.00

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 de médias** são sobrepostos a eles. Esse tipo de gráficos 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 e 1.3.3 apresentamos gráficos de perfis correspondentes aos Exemplos 1.2.3 e 1.2.6, respectivamente, com a sobreposição do perfil de médias correspondentes.

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);

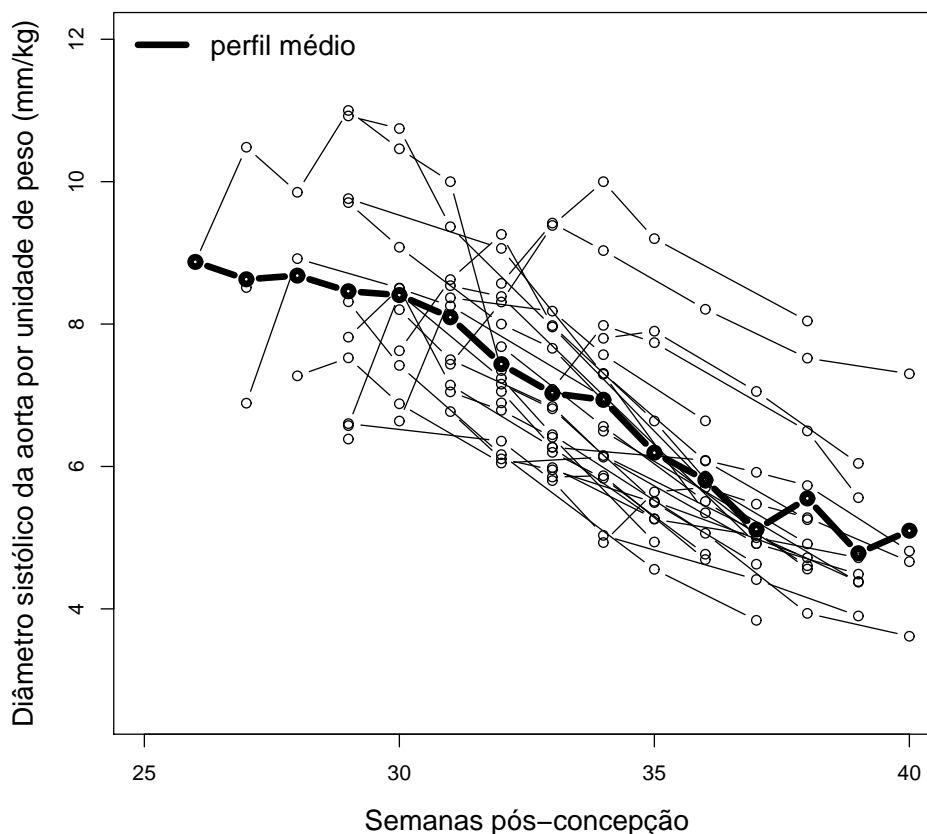
Figura 1.3.2: Gráfico de perfis para os dados do Exemplo 1.2.3



- 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

Figura 1.3.3: Gráfico de perfis para os dados do Exemplo 1.2.6



transporte mucociliar relativa) correspondente às concentrações 0-8 μM se mantêm aproximadamente constantes 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, porém com níveis dependentes das diferentes concentrações. Além disso, um dos perfis individuais observados sob a concentração 8 μM se destaca dos demais, sugerindo que a unidade amostral correspondente pode ser considerada como ponto discrepante.

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, no segundo, é o **pico** (*peak*) e no último, a área sob a curva. Todos esses exemplos podem ser classificados sob a denominação de **medidas resumo**, cuja análise remete-nos essencialmente à análises de regressão e/ou ANOVA usuais.

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, 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) and 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, $\mathbf{y}_i = (y_{i1}, \dots, y_{ip_i})^\top$ com dimensão $(p_i \times 1)$ 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, $\mathbf{X}_i = (\mathbf{x}_{i1}, \dots, \mathbf{x}_{im})$ é uma matriz de especificação dos efeitos fixos com dimensão $(p_i \times t)$, conhecida e de posto completo, em que $\mathbf{x}_{ij} = (x_{ij1}, \dots, x_{ijp_i})^\top$ representa o vetor com os p_i valores da j -ésima variável independente ($j = 1, \dots, m$) 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 $(p_i \times q)$, conhecida e de posto completo) e \mathbf{e}_i é um vetor de erros aleatórios com dimensão $(p_i \times 1)$.

Em muitos casos é razoável assumir que $\mathbf{b}_i \sim N_q(\mathbf{0}, \mathbf{G})$ e $\mathbf{e}_i \sim N_{p_i}(\mathbf{0}, \mathbf{R}_i)$, em que \mathbf{G} , com dimensão $(q \times q)$ e \mathbf{R}_i , com dimensão $(p_i \times p_i)$ são matrizes simétricas definidas positivas, são variáveis aleatórias independentes. Quando \mathbf{g} é uma função linear dos efeitos fixos $\boldsymbol{\beta}$ e dos efeitos aleatórios \mathbf{b}_i , podemos reescrever o modelo como

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

Portanto, 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

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

As matrizes \mathbf{R}_i e \mathbf{G} são funções conhecidas de m_1 e m_2 parâmetros desconhecidos, respectivamente ou seja, $\mathbf{G} = \mathbf{G}(\boldsymbol{\theta}_1)$ e $\mathbf{R}_i = \mathbf{R}_i(\boldsymbol{\theta}_2)$, de modo que a matriz de covariâncias para a i -ésima unidade amostral, $\mathbf{V}_i = \mathbf{V}_i(\boldsymbol{\theta})$, dependerá do vetor

$$\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top)^\top \quad (2.1.5)$$

com dimensão $(m \times 1)$ em que $m = m_1 + m_2$. Este modelo 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

$$\mathbf{y}_i | \mathbf{b}_i \sim N(\mathbf{R}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{b}_i, \mathbf{R}_i). \quad (2.1.6)$$

No segundo estágio assumimos que os vetores \mathbf{b}_i são independentes com distribuição $N_q(\mathbf{0}, \mathbf{G})$ e conseqüentemente, o modelo marginal é dado por

$$\mathbf{y}_i \sim N(\mathbf{X}_i\boldsymbol{\beta}, \mathbf{Z}_i\mathbf{G}\mathbf{Z}_i^\top + \mathbf{R}_i). \quad (2.1.7)$$

Quando $\mathbf{R}_i = \sigma^2\mathbf{I}_{p_i}$, o modelo é chamado de **modelo de independência condicional homocedástico**, indicando que as p_i observações associadas à i -ésima unidade amostral são condicionalmente independentes de \mathbf{b}_i .

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 (2004) 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 covariâncias. Em geral, o modelo para matriz de covariâncias \mathbf{V}_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 indiretamente por meio dos efeitos aleatórios, \mathbf{b}_i , que representa a variabilidade entre as unidades amostrais, diretamente por meio da matriz \mathbf{R}_i , que representa a covariância das observações intraunidades amostrais como nos modelos lineares usuais, ou ainda, como uma combinação, de \mathbf{b}_i e \mathbf{R}_i , como sugerem Vonesh & Chinchilli (1997), permitindo deste modo incluir na estrutura de covariância as três fontes de variação: os efeitos aleatórios

e a variação devida à erros de medida, na matriz \mathbf{G} e a correlação serial na matriz \mathbf{R}_i .

No modelo linear misto homocedástico com independência condicional, $\mathbf{R}_i = \sigma^2 \mathbf{I}_{p_i}$ 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 $p_i = 4$.

1. Estrutura Uniforme

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

2. Estrutura autorregressiva de ordem 1 AR(1)

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

3. Estrutura ARMA(1,1)

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

4. Estrutura antedependência de ordem 1

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

5. Estrutura Toeplitz

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

6. Estrutura espacial ou de Markov

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

7. Estrutura uniforme heterogênea

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

8. Estrutura ARH(1)

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

9. Estrutura Toeplitz heterogênea

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

10. Estrutura baseada em efeitos aleatórios

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

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

com

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

11. Não-estruturada (NE)

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

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 Kronecker autorregressiva

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

2. Não-estruturada Kronecker 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 os processos estocásticos que geram essas estruturas podem ser encontrados em Rocha (2004), por exemplo.

2.3 Inferência por máxima verossimilhança

Sob a suposição de normalidade e linearidade, o modelo (2.1.2) pode ser escrito compactamente como

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \mathbf{e}, \quad (2.3.1)$$

em que $\mathbf{y} = [\mathbf{y}_1^\top, \dots, \mathbf{y}_n^\top]^\top$ é um vetor com dimensão $(N \times 1)$ em que $N = \sum_{i=1}^n p_i$ contendo as respostas das n unidades amostrais, $\mathbf{X} = [\mathbf{X}_1^\top, \dots, \mathbf{X}_n^\top]$ com dimensão $(N \times t)$ é a matriz de especificação dos efeitos fixos, $\mathbf{Z} = \oplus_{i=1}^n [\mathbf{Z}_1, \dots, \mathbf{Z}_n]$ 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, $\mathbf{b} \sim N_{nq}[\mathbf{0}, \boldsymbol{\Gamma}(\boldsymbol{\theta}_1)]$ em que $\boldsymbol{\Gamma}(\boldsymbol{\theta}_1) = \mathbf{I}_n \otimes \mathbf{G}(\boldsymbol{\theta}_1)$ e \mathbf{b} é independente de $\mathbf{e} \sim N_N[\mathbf{0}, \mathbf{R}]$ com $\mathbf{R} = \oplus_{i=1}^n \mathbf{R}_i(\boldsymbol{\theta}_2)$. Consequentemente, $\mathbf{y} \sim N_N[\mathbf{X}\boldsymbol{\beta}, \mathbf{V}(\boldsymbol{\theta})]$ em que $\mathbf{V}(\boldsymbol{\theta}) = \mathbf{Z}\boldsymbol{\Gamma}(\boldsymbol{\theta}_1)\mathbf{Z}^\top + \mathbf{R}(\boldsymbol{\theta}_2)$.

Vários métodos de estimação dos parâmetros do modelo (2.3.1) 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 $\mathbf{G}(\boldsymbol{\theta}_1)$ e $\mathbf{R}(\boldsymbol{\theta}_2)$ 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

$$l(\boldsymbol{\beta}, \boldsymbol{\theta}) = -\frac{N}{2} \log 2\pi - \frac{1}{2} \log |\mathbf{V}(\boldsymbol{\theta})| - \frac{1}{2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top \mathbf{V}^{-1}(\boldsymbol{\theta}) (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \quad (2.3.2)$$

ou alternativamente

$$\begin{aligned} l(\boldsymbol{\beta}, \boldsymbol{\theta}) &= -\frac{1}{2} \log 2\pi \sum_{i=1}^n p_i - \frac{1}{2} \sum_{i=1}^n \log |\mathbf{V}_i(\boldsymbol{\theta})| \\ &\quad - \frac{1}{2} \sum_{i=1}^n (\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})^\top \mathbf{V}_i^{-1}(\boldsymbol{\theta}) (\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}). \end{aligned} \quad (2.3.3)$$

A maximização de (2.3.3) pode ser realizada por meio dos seguintes passos:

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

$$\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}) = \left[\sum_{i=1}^n \mathbf{X}_i^\top \mathbf{V}_i^{-1}(\boldsymbol{\theta}) \mathbf{X}_i \right]^{-1} \left[\sum_{i=1}^n \mathbf{X}_i^\top \mathbf{V}_i^{-1}(\boldsymbol{\theta}) \mathbf{y}_i \right]. \quad (2.3.4)$$

Este é o estimador de **mínimos quadrados generalizados** de $\boldsymbol{\beta}$ sob a suposição de que $\mathbf{V}_i(\boldsymbol{\theta})$ é conhecida, *i.e.*, $\hat{\boldsymbol{\beta}}(\boldsymbol{\theta})$ minimiza a forma quadrática

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

- ii) Substituir $\hat{\beta}(\theta)$ na expressão (2.3.3) 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, para $j = 1, \dots, m$,

$$-\frac{1}{2} \sum_{i=1}^n \text{tr}[\mathbf{V}_i^{-1}(\hat{\theta}) \dot{\mathbf{V}}_i(\hat{\theta})] - \frac{1}{2} \sum_{i=1}^n [\partial Q_i(\theta)/\partial \theta_j]_{\theta=\hat{\theta}} = 0 \quad (2.3.5)$$

em que

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

e

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

A solução $\hat{\theta}$ de (2.3.5) e $\hat{\beta} = \hat{\beta}(\hat{\theta})$, ou seja (2.3.4) com a substituição de θ por $\hat{\theta}$ constituem os estimadores de MV de θ e β .

A metodologia de máxima verossimilhança gera estimadores não-enviesados para os efeitos fixos mas produz estimadores enviesados 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 $\mathbf{y}^* = \mathbf{U}^{\top} \mathbf{y}$ em que \mathbf{U} tem dimensão $[N \times (N - t)]$, e é tal que $\mathbb{E}(\mathbf{y}^*) = \mathbf{0}$, 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 projecção 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

$$\mathbb{E}(\mathbf{y}^*) = \mathbf{0} \quad (2.3.6)$$

$$\mathbb{V}(\mathbf{y}^*) = \mathbf{U} \mathbf{V}(\theta) \mathbf{U}^{\top}, \quad (2.3.7)$$

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

$$l_R(\boldsymbol{\theta}) = -\frac{1}{2} \log |\mathbf{V}(\boldsymbol{\theta})| - \frac{1}{2} \log |\mathbf{X}^\top \mathbf{V}^{-1}(\boldsymbol{\theta}) \mathbf{X}| \\ - \frac{N-t}{2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top \mathbf{V}^{-1}(\boldsymbol{\theta}) (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) - \frac{N-t}{2} \log(2\pi), \quad (2.3.8)$$

e sua maximização gera os estimadores de $\hat{\boldsymbol{\theta}}^\top = [\hat{\boldsymbol{\theta}}_1^\top, \hat{\boldsymbol{\theta}}_2^\top]^\top$.

Alternativamente, a função log-verossimilhança restrita (2.3.8) pode ser escrita como

$$l_R(\boldsymbol{\theta}) = -\frac{1}{2} \log 2\pi \sum_{i=1}^n p_i - \frac{1}{2} \sum_{i=1}^n \log |\mathbf{V}_i(\boldsymbol{\theta})| \\ - \frac{1}{2} \sum_{i=1}^n [\mathbf{y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}}(\boldsymbol{\theta})]^\top \mathbf{V}_i^{-1}(\boldsymbol{\theta}) [\mathbf{y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}}(\boldsymbol{\theta})] \\ - \frac{1}{2} \log \left| \sum_{i=1}^n \mathbf{X}_i^\top \mathbf{V}_i^{-1}(\boldsymbol{\theta}) \mathbf{X}_i \right|. \quad (2.3.9)$$

Neste caso, no processo de maximização, a equação (2.3.5) é substituída por

$$-\frac{1}{2} \sum_{i=1}^n \text{tr}\{\mathbf{V}_i^{-1}(\hat{\boldsymbol{\theta}}) \dot{\mathbf{V}}_i(\hat{\boldsymbol{\theta}})\} - \frac{1}{2} \sum_{i=1}^n [\partial Q_i(\boldsymbol{\theta}) / \partial \boldsymbol{\theta}_j]_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \\ - \frac{1}{2} \sum_{i=1}^n \text{tr}\{\mathbf{V}_i^{-1}(\hat{\boldsymbol{\theta}}) \mathbf{X}_i^\top \mathbf{V}_i^{-1}(\hat{\boldsymbol{\theta}}) \dot{\mathbf{V}}_i(\hat{\boldsymbol{\theta}}) \mathbf{V}_i^{-1}(\hat{\boldsymbol{\theta}}) \mathbf{X}_i\} = 0. \quad (2.3.10)$$

A solução de (2.3.10) para $j = 1, \dots, m$, 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.

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

$$f(\mathbf{y}, \mathbf{b}) = f(\mathbf{y}|\mathbf{b})f(\mathbf{b}), \quad (2.3.11)$$

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 $\mathbf{y}|\mathbf{b} \sim N_N(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b}, \mathbf{R})$ e $\mathbf{b} \sim N_{nq}(\mathbf{0}, \boldsymbol{\Gamma})$ a função

(2.3.11) pode ser diretamente obtida como indicam Henderson (1975), Searle (1971), Searle et al. (1992) ou McCulloch & Searle (2001), por exemplo.

Por meio do teorema de Gauss-Markov para efeitos aleatórios apresentado em Harville (1977a) podem ser obtidos o **melhor estimador linear não-enviesado** (*best linear unbiased estimator* - **BLUE**) para β e o **melhor preditor não-enviesado** (*best linear unbiased predictor* - **BLUP**) para o vetor de efeitos aleatórios \mathbf{b} . Diferentes formas de obtenção do BLUP e 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 BLUP. A ideia básica é reexpressar o modelo linear misto (2.3.1) na forma de um modelo linear geral através da inclusão de “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 a consequente analogia que pode ser feita com as técnicas de diagnóstico existentes para modelos lineares. Nesse contexto, consideremos o modelo (2.3.1) com a inclusão do seguinte “caso artificial”,

$$\mathbf{0}_{nq \times 1} = \mathbf{0}_{nq \times 1} - \mathbf{I}_{nq} \mathbf{b} + \mathbf{d}, \quad (2.3.12)$$

em que \mathbf{d} é um vetor de dimensão $(nq \times 1)$ que faz o papel de erro, $\mathbb{V}(\mathbf{e}) = \sigma^2 \mathbf{R}(\theta_2)$, $\mathbb{V}(\mathbf{d}) = \sigma^2 \mathbf{\Gamma}(\theta_1)$ e $\mathbb{Cov}(\mathbf{e}, \mathbf{d}) = \mathbf{0}$ e reescrevamos as equações (2.3.1) e (2.3.12) em forma matricial como

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} \mathbf{X} & \mathbf{Z} \\ \mathbf{0} & -\mathbf{I} \end{bmatrix} \begin{bmatrix} \beta \\ \mathbf{b} \end{bmatrix} + \begin{bmatrix} \mathbf{e} \\ \mathbf{d} \end{bmatrix}. \quad (2.3.13)$$

Pré-multiplicando (2.3.13) por $\mathbf{R}^{-1/2}(\theta_2) \oplus \mathbf{\Gamma}^{-1/2}(\theta_1)$ tem-se

$$\mathbf{y}^{**} = \mathbf{X}^* \beta^* + \mathbf{H}, \quad (2.3.14)$$

em que,

$$\mathbf{y}^{**} = \begin{bmatrix} \mathbf{R}^{-1/2}(\theta_2) \mathbf{y} \\ \mathbf{0} \end{bmatrix}, \quad \mathbf{X}^* = \begin{bmatrix} \mathbf{R}^{-1/2}(\theta_2) \mathbf{X} & \mathbf{R}^{-1/2}(\theta_2) \mathbf{Z} \\ \mathbf{0} & -\mathbf{\Gamma}^{-1/2}(\theta_1) \end{bmatrix}, \quad (2.3.15)$$

e $\boldsymbol{\beta}^* = [\boldsymbol{\beta}^\top, \mathbf{b}^\top]^\top$, tal que $\mathbb{V}(\mathbf{H}) = \sigma^2 \mathbf{I}$.

Desta forma, (2.3.13) 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 da equação (2.3.13), usando o método de MQ, sob o qual se obtém

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

ou seja

$$\begin{bmatrix} \mathbf{X}^\top \mathbf{R}^{-1/2}(\boldsymbol{\theta}_2) \mathbf{y} \\ \mathbf{Z}^\top \mathbf{R}^{-1/2}(\boldsymbol{\theta}_2) \mathbf{y} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^\top \mathbf{R}^{-1}(\boldsymbol{\theta}_2) \mathbf{X} & \mathbf{X}^\top \mathbf{R}^{-1}(\boldsymbol{\theta}_2) \mathbf{Z} \\ \mathbf{Z}^\top \mathbf{R}^{-1}(\boldsymbol{\theta}_2) \mathbf{X} & \mathbf{Z}^\top \mathbf{R}^{-1}(\boldsymbol{\theta}_2) \mathbf{Z} + \boldsymbol{\Gamma}^{-1}(\boldsymbol{\theta}_1) \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{b}} \end{bmatrix} \quad (2.3.16)$$

Essas equações são conhecidas na literatura como **equações de Henderson**. Note que, se $\boldsymbol{\Gamma}^{-1}(\boldsymbol{\theta}_1) \equiv \mathbf{0}$ (o que implica que \mathbf{b} é um efeito fixo), então (2.3.16) coincide com a equação de estimação obtida por meio do método de **mínimos quadrados generalizados** (MQG). O BLUP e o BLUE são obtidos resolvendo-se as equações (2.3.16), 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

$$\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}) = [\mathbf{X}^\top \mathbf{V}^{-1}(\boldsymbol{\theta}) \mathbf{X}]^{-1} \mathbf{X}^\top \mathbf{V}^{-1}(\boldsymbol{\theta}) \mathbf{y} \quad (2.3.17)$$

e

$$\hat{\mathbf{b}}(\boldsymbol{\theta}) = \boldsymbol{\Gamma}(\boldsymbol{\theta}_1) \mathbf{Z}^\top \mathbf{V}^{-1}(\boldsymbol{\theta}) [\mathbf{y} - \mathbf{X} \hat{\boldsymbol{\beta}}(\boldsymbol{\theta})] \quad (2.3.18)$$

ou alternativamente por

$$l\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}) = \sum_{i=1}^n [\mathbf{X}_i^\top \mathbf{V}_i^{-1}(\boldsymbol{\theta}) \mathbf{X}_i]^{-1} \mathbf{X}_i \mathbf{V}_i^{-1}(\boldsymbol{\theta}) \mathbf{y}_i \quad (2.3.19)$$

e

$$\hat{\mathbf{b}}_i(\boldsymbol{\theta}) = \mathbf{G}(\boldsymbol{\theta}_1) \mathbf{Z}_i^\top \mathbf{V}_i^{-1}(\boldsymbol{\theta}) [\mathbf{y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}}(\boldsymbol{\theta})] \quad (2.3.20)$$

Além disso, o estimador do parâmetro de escala comum a $\boldsymbol{\theta}_1^\top$, $\boldsymbol{\theta}_2^\top$ é dado por

$$\hat{\sigma}^2 = \frac{1}{N-t} [\mathbf{y} - \mathbf{X} \hat{\boldsymbol{\beta}}(\boldsymbol{\theta})]^\top [\mathbf{Z} \boldsymbol{\Gamma}(\boldsymbol{\theta}_1) \mathbf{Z}^\top + \mathbf{R}(\boldsymbol{\theta}_2)]^{-1} [\mathbf{y} - \mathbf{X} \hat{\boldsymbol{\beta}}(\boldsymbol{\theta})]. \quad (2.3.21)$$

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

$$\mathbb{C}_{\Phi\mathbf{V}} \begin{bmatrix} \widehat{\boldsymbol{\beta}}(\boldsymbol{\theta}) - \boldsymbol{\beta} \\ \widehat{\mathbf{b}}(\boldsymbol{\theta}) - \mathbf{b} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^\top \mathbf{R}^{-1}(\boldsymbol{\theta}_2) \mathbf{X} & \mathbf{X}^\top \mathbf{R}^{-1}(\boldsymbol{\theta}_2) \mathbf{Z} \\ \mathbf{Z}^\top \mathbf{R}^{-1}(\boldsymbol{\theta}_2) \mathbf{X} & \mathbf{Z}^\top \mathbf{R}^{-1}(\boldsymbol{\theta}_2) \mathbf{Z} + \boldsymbol{\Gamma}^{-1}(\boldsymbol{\theta}_1) \end{bmatrix}^{-1}. \quad (2.3.22)$$

Em termos dos componentes individuais (2.3.20), temos

$$\mathbb{V}[\widehat{\mathbf{b}}_i(\boldsymbol{\theta})] = \mathbf{G}(\boldsymbol{\theta}_1) \mathbf{Z}_i^\top \boldsymbol{\Delta}(\boldsymbol{\theta}) \mathbf{Z}_i \mathbf{G}(\boldsymbol{\theta}_1). \quad (2.3.23)$$

com

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

Como $\boldsymbol{\Gamma}(\boldsymbol{\theta}_1)$ e $\mathbf{R}(\boldsymbol{\theta}_2)$ dependem do vetor de parâmetros de covariância desconhecido $\boldsymbol{\theta}$, é razoável calcular o BLUE e BLUP com base num estimador $\widehat{\boldsymbol{\theta}}$ de $\boldsymbol{\theta}$; esses “preditores” são denominados BLUE e BLUP empíricos (**EBLUE** e **EBLUP**). Se $\widehat{\boldsymbol{\theta}}$ 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 (2.3.4), podemos concluir que

$$\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta}) \sim N\{\boldsymbol{\beta}, [\sum_{i=1}^n \mathbf{X}_i^\top \mathbf{V}_i^{-1}(\boldsymbol{\theta}) \mathbf{X}_i]^{-1}\}.$$

Por outro lado, como tanto $\widehat{\boldsymbol{\theta}}$ quanto $\widehat{\boldsymbol{\theta}}_R$ 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 $\hat{\beta}$ quanto a de $\hat{\beta}_R$ podem ser aproximadas pela mesma distribuição utilizada para $\hat{\beta}(\theta)$ sob certas condições de regularidade para as matrizes de especificação \mathbf{X}_i . Detalhes podem ser obtidos em Sen & Singer (1993), Jiang (1996) e Verbeke & Lesaffre (1996), por exemplo.

Esses resultados servem de base para a construção de testes de hipóteses sobre os parâmetros da matriz de covariâncias θ e sobre os parâmetros de localização β . Testes de hipóteses do tipo

$$H : \mathbf{C}\theta = \mathbf{0} \quad (2.3.24)$$

em que \mathbf{C} denota uma matriz com dimensão $(c \times m)$ 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**

$$Q_W = (\mathbf{C}\hat{\theta})^\top [\mathbf{C}\mathbb{V}(\hat{\theta})\mathbf{C}^\top]^{-1}(\mathbf{C}\hat{\theta}), \quad (2.3.25)$$

ou a estatística da razão de verossimilhanças (RV)

$$Q_{RV} = -2[l_1(\hat{\theta}) - l_2(\hat{\theta})], \quad (2.3.26)$$

em que $l_1(\hat{\theta})$ e $l_2(\hat{\theta})$ denotam, respectivamente, a função log-verossimilhança maximizada sob o modelo restrito *i.e.*, induzido por (2.3.24) e aquela maximizada sob o modelo irrestrito. Sob as condições de regularidade usuais e para amostras suficientemente grandes, as distribuições de (2.3.25) e de (2.3.26) 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

$$H_0 : \sigma_a^2 = 0 \text{ versus } H_1 : \sigma_a^2 \in (0, \infty) \quad (2.3.27)$$

em que σ_a^2 , um dos componentes de θ 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 $p = 0.5P[\chi_q^2 > Q_{RV}(obs)] + 0.5P[\chi_{q+1}^2 > Q_{RV}(obs)]$ em que $Q_{RV}(obs)$ é o valor observado da estatística (2.3.26).

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 covariâncias intraunidades amostrais, consideramos os critérios de informação de Akaike (AIC_R) e de Schwarz (BIC_R) definidos por

$$AIC_R = -2l(\hat{\theta}) + 2q \quad (2.3.28)$$

$$BIC_R = -2l(\hat{\theta}) + q \log(N - t), \quad (2.3.29)$$

em que $l(\hat{\theta})$ é o máximo da função log-verossimilhança restrita e q é o número de parâmetros da matriz de covariância, t 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 e estudos com medidas repetidas são encontrados em Keselman, James, Rhonda & Russell (1998).

Funções do tipo $\mathbf{L}^\top \boldsymbol{\beta} + \mathbf{D}^\top \mathbf{b}$ em que \mathbf{L} e \mathbf{D} são matrizes conhecidas com dimensões $(l \times t)$ 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 $\mathbf{K} = [\mathbf{L}^\top \mathbf{D}^\top]$ e considerar

$$H : \mathbf{K} \begin{bmatrix} \boldsymbol{\beta} \\ \mathbf{b} \end{bmatrix} = \mathbf{0}. \quad (2.3.30)$$

A estatística de Wald para testar (2.3.30) é

$$Q_W = \left(\mathbf{K} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{b}} \end{bmatrix} \right)^\top [\mathbf{K} \widehat{\mathbf{W}} \mathbf{K}^\top]^{-1} \mathbf{K} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{b}} \end{bmatrix} \quad (2.3.31)$$

em que $\widehat{\mathbf{W}}$ é uma estimativa da matriz (2.3.22). Se estivermos interessados apenas em contrastes dos efeitos fixos, todas as linhas associadas aos efeitos aleatórios na matriz \mathbf{K} são nulas. A sua distribuição aproximada é $\chi^2_{\text{posto}(\mathbf{K})}$.

De uma forma mais conservadora, podemos utilizar uma estatística F dividindo (2.3.31) 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

$$\mathbf{K} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{b}} \end{bmatrix} \pm t_{v, \frac{\alpha}{2}} \sqrt{[\mathbf{K} \widehat{\mathbf{W}} \mathbf{K}^\top]}. \quad (2.3.32)$$

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 (2.3.30) 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 (2.3.4) e (2.3.5) ou (2.3.10) 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 $p_i = p$, $\mathbf{Z}_i = \mathbf{0}$ e em que $\mathbb{V}(\mathbf{e}_i) = \mathbf{R}_i = \mathbf{\Sigma}$ 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 $(n \times p)$, $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_t]$ é uma matriz com dimensão $(n \times t)$ 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 $(t \times p)$ cujo elemento na posição (i, j) , $i = 1, \dots, t$, $j = 1, \dots, p$ corresponde à resposta média do tratamento i no tempo j e $\mathbf{E} = [\mathbf{e}_1, \dots, \mathbf{e}_n]^\top$ é uma matriz de erros aleatórios com dimensão $(n \times p)$ cujas linhas seguem distribuições normais p -variadas com vetor de médias nulo e matriz de covariâncias $\mathbf{\Sigma}$.

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

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

A expressão acima com a substituição de $n-t$ 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 $(c \times t)$ e $(p \times u)$ 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{HE}^{-1} em que

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

e

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

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{HE}^{-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 dependem de $k_1 = (|u-c|-1)/2$, $k_2 = (n-t-u-1)/2$ 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 $\mathbf{a}^\top \mathbf{Mb}$ em que \mathbf{a} e \mathbf{b} são vetores de constantes podem ser construídos como

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

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 $(t \times r)$ de coeficientes desconhecidos \mathbf{Q} é uma matriz conhecida com dimensão $(r \times p)$ 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,

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

e

$$\hat{\Sigma} = \frac{1}{n} [\mathbf{Y} - \mathbf{X} \hat{\mathbf{B}} \mathbf{Q}]^\top [\mathbf{Y} - \mathbf{X} \hat{\mathbf{B}} \mathbf{Q}]$$

são estimadores de MV. O estimador de MVR de Σ é obtido com a substituição de $n - t$ for n na expressão acima. Testes de hipótese 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 três 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 lados de (2.4.3) por uma matriz $\mathbf{H} = [\mathbf{H}_1 \ \mathbf{H}_2]$ com dimensão $(p \times p)$ 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} and \mathbf{E} respectivamente substituídas por $\mathbf{Y}_1 = \mathbf{YH}_1$, \mathbf{B} and $\mathbf{E} = \mathbf{EH}_1$ 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 uma estrutura uniforme do tipo

$$\Sigma = \nu^2 \mathbf{1}_p \mathbf{1}_p^\top + \tau^2 \mathbf{I}_p$$

em que o ν^2 e τ^2 são os parâmetros de covariância é adotada para a matriz de covariâncias intraunidades amostrais. 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

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

e

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

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

Como a matriz de covariâncias intraunidades amostrais com estrutura uniforme é 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 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**, *i.e.*, se

$$\mathbf{P}^\top \Sigma \mathbf{P} = \lambda \mathbf{I}_{p-1} \quad (2.4.4)$$

em que \mathbf{P} é uma matriz de contrastes ortonormais com dimensão $(p \times (p-1))$ 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.

Com base no trabalho de Box (1954a) e Box (1954b), Geisser & Greenhouse (1958) propuseram testes F aproximados para casos em que Σ 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

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

Quando Σ é esférica, temos $\varepsilon = 1$. Estimadores do fator de correção foram propostos por Huynh & Feldt (1976), entre outros. 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).

Soluções explícitas para as equações de estimação também podem ser obtidas para problemas sem dados omissos sob modelos mistos homocedásticos com independência condicional. Os resultados, cujos detalhes podem ser obtidos em Graybill (1976) e Laird, Lange & Stram (1987), são menos interessantes, pois não tiram proveito das técnicas padrão de ANOVA ou MANOVA.

Métodos mais gerais do que aqueles descritos acima para resolução das equações de estimação são necessárias 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 $\theta^{(0)}$ e denotando a **função escore** por $\mathbf{u}(\theta) = \partial l[\hat{\beta}(\theta), \theta] / \partial \theta$ e a **matriz hessiana** por $\mathbf{H}(\theta) = \partial^2 l[\hat{\beta}(\theta), \theta] / \partial \theta \partial \theta^t$ associadas à função log-verossimilhança $l(\hat{\beta}(\theta), \theta)$ o algoritmo de Newton-Raphson para a solução das equações de estimação (2.3.5) corresponde a iterações de

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

para $l = 1, 2, \dots$ até que algum critério de convergência seja satisfeito, *e.g.*, $\|\theta^{(l)} - \theta^{(l-1)}\| < \varepsilon$ para algum valor pré-especificado $\varepsilon > 0$. Da última iteração de (2.4.5) obtemos a estimativa de MV de θ , nomeadamente $\hat{\theta}$, que substituída em (2.3.4), gera a estimativa de MV de β , a saber, $\hat{\beta}$. Uma estimativa consistente da matriz de covariâncias assintótica de $\hat{\beta}$ é dada por $\{\sum_{i=1}^n \mathbf{X}_i^T \mathbf{V}_i^{-1}(\hat{\theta}) \mathbf{X}_i\}^{-1}$. Uma estimativa consistente da matriz de covariâncias assintótica de $\hat{\theta}$ pode ser obtida a partir da inversa da **matriz de informação empírica**, $-\mathbf{H}(\hat{\theta})$.

O algoritmo *scoring* de Fisher corresponde a (2.4.5) com $\mathbf{H}(\theta)$ substituída por sua esperança. Se utilizarmos (2.3.10) no lugar de (2.3.5)) nas definições de $\mathbf{u}(\theta)$ e de $\mathbf{H}(\theta)$, os dois procedimentos geram estimativas de MVR.

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 "completos", *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 $\mathbf{R}_i = \sigma^2 \mathbf{I}$ em (2.1.4), expressando o vetor de estatísticas suficientes como $\mathbf{s} = (s_1, \mathbf{s}_2^\top)^\top$, Laird & Ware (1982) mostram que o Passo E se reduz a

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

e

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

em que

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

e

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

O Passo M é dado por

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

e

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

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. Além disso, ele 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.

Tendo em vista a expressão (2.3.4) é possível obter um estimador de mínimos quadrados generalizados em dois estágios para o parâmetro de localização β se tivermos disponível estimadores consistentes de $\mathbf{V}_i(\theta)$. Um caso especial de interesse é aquele para o qual podemos adotar um modelo misto homocedástico com independência condicional (2.1.2) em que $\mathbf{Z}_i = \mathbf{X}_i$ and $\mathbf{R}_i = \tau^2 \mathbf{I}_{p_i}$. Esse modelo, também cognominado **modelo de coeficientes aleatórios** é apropriado para casos em que os perfis individuais e populacionais têm a mesma forma. Escrevendo $\beta_i = \beta + \mathbf{b}_i$ e com base num argumento condicional, podemos ajustar curvas separadamente aos dados de cada unidade amostral por meio de mínimos quadrados ordinários, obtendo

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

com $\mathbb{E}(\hat{\beta}_i) = \beta$ e $\mathbb{V}(\hat{\beta}_i) = \mathbf{G} + \tau^2 (\mathbf{X}_i^\top \mathbf{X}_i)^{-1}$. Como consequência, estimadores consistentes de τ^2 e $\mathbb{V}(\hat{\beta}_i)$ são dados por

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

e

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

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

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

e

$$\mathbf{V}_i(\hat{\boldsymbol{\theta}}) = \mathbf{X}_i \hat{\mathbf{G}} \mathbf{X}_i^\top + \hat{\tau}^2 \mathbf{I}_{p_i}.$$

Esse enfoque é adotado por Vonesh & Carter (1987), que também apresentam as propriedades assintóticas desses estimadores.

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

2.5 Estratégias de análise

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 o 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 para a 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 regressão 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. análise de resíduos.

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 dos gráficos de dispersão múltiplos das observações padronizadas quando for possível construí-los;
5. análise dos gráficos de linhas das matrizes de covariâncias e correlações amostrais versus defasagens;
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. análise de resíduos.

Detalhes sobre os diferentes procedimentos mencionados acima serão discutidos por meio de exemplos no Capítulo 5.

2.6 Diagnóstico

2.7 Notas de capítulo

2.8 Exercícios

Capítulo 3

Análise de dados

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

3.1 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 5.1.1: 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) medida em mmHg de cães, a resposta foi medida antes e depois da aplicação 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 3.1.1, um gráfico de perfis individuais, na Figura 3.1.1 e um gráfico de perfis médios, na Figura 3.1.2.

Gráficos do tipo *boxplot* correspondentes estão dispostos na Figura 3.1.3. Neste exemplo, temos uma intervenção (aplicação de $MgSO_4$) e uma

Tabela 3.1.1: 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

população com dois estratos, que correspondem aos dois tratamentos (indometacina e nifedipina). O interesse é:

- comparar o efeito da aplicação de $MgSO_4$ (PAM média pós-teste - PAM média pré-teste) nos cães tratados com indometacina com o efeito da aplicação de $MgSO_4$ nos cães tratados com nifedipina (interação entre tratamento e tempo),
- avaliar o efeito principal da aplicação de $MgSO_4$ (se não houver interação ou se ela for não-essencial),
- avaliar o efeito da aplicação de $MgSO_4$ para cada tratamento (se houver interação essencial),
- avaliar o efeito principal de tratamento sobre a PAM média (se não houver interação ou se ela for não-essencial)
- avaliar o efeito de tratamento sobre a PAM média pré-teste (se houver interação essencial).

Para situações como aquela descrita no Exemplo 5.1.1, com n_i unidades amostrais no estrato i , $i = 1, 2$, o modelo linear básico pode ser escrito como

$$y_{ijk} = \mu_{ik} + e_{ijk}, \quad i = 1, 2, \quad j = 1, \dots, n_i, \quad k = 1, 2 \quad (3.1.1)$$

Figura 3.1.1: Perfis individuais da PAM de cães, antes (pré-teste) e depois (pós-teste) de uma aplicação de $MgSO_4$

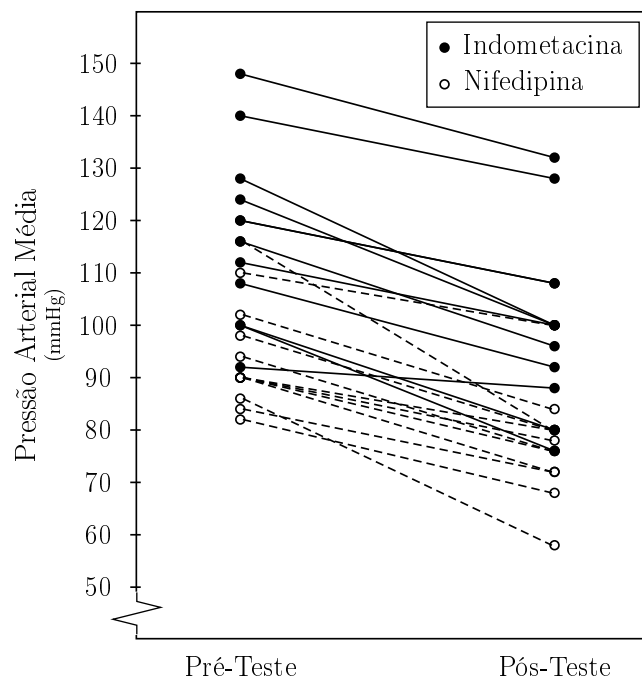


Figura 3.1.2: Perfis médios e médias amostrais (círculos) para a PAM, antes (pré-teste) e depois (pós-teste) de uma aplicação de $MgSO_4$.

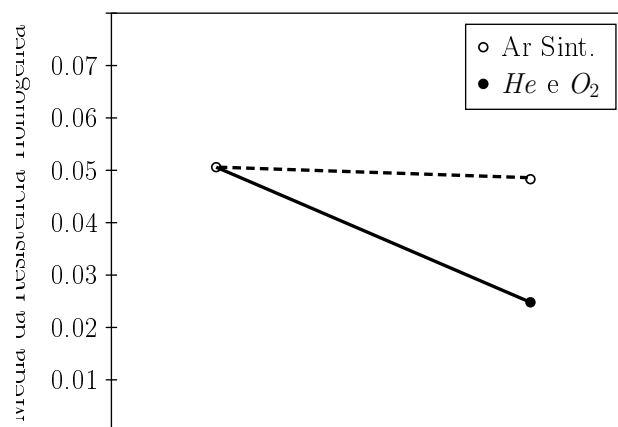
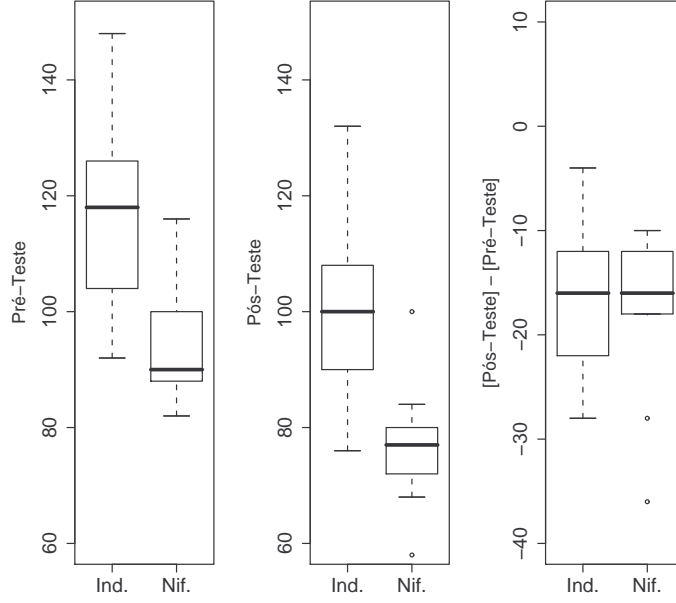


Figura 3.1.3: Boxplots para a PAM de cães, antes (pré-teste) e depois (pós-teste) de uma aplicação de $MgSO_4$ e para o efeito da aplicação



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 média para as unidades amostrais do i -ésimo estrato no k -ésimo tempo 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.*, $\mathbf{Y} = \mathbf{XM} + \mathbf{E}$, com $\mathbf{Y} = (\mathbf{y}_{11}, \dots, \mathbf{y}_{1n_1}, \mathbf{y}_{21}, \dots, \mathbf{y}_{2n_2})^\top$, $\mathbf{y}_{ij} = (y_{ij1}, y_{ij2})^\top$, $\mathbf{X} = \oplus_{i=1}^2 \mathbf{1}_{n_i}$, $\mathbf{M} = (\boldsymbol{\mu}_1, \boldsymbol{\mu}_2)^\top$ 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 $\mathbf{e}_{ij} \sim N_2(\mathbf{0}, \boldsymbol{\Sigma})$ com

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}$$

e 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),

- 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),
- 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 (3.1.1) sob a **reparametrização de desvios médios**

$$\mu_{ik} = \mu + \alpha_i + \beta_k + \alpha\beta_{ik} \quad i = 1, 2, \quad k = 1, 2 \quad (3.1.2)$$

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 (3.1.2) ou de outros com diferentes parametrizações está explicitada no Apêndice C.

Um outro enfoque clássico para o problema envolve um modelo misto obtido de (3.1.1) por intermédio da inclusão da seguinte estrutura para o erro aleatório

$$e_{ijk} = b_{ij} + d_{ijk} \quad (3.1.3)$$

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

$$\mathbb{V}(\mathbf{e}_{ij}) = \Sigma = \begin{bmatrix} \sigma_b^2 + \sigma^2 & \sigma_b^2 \\ \sigma_b^2 & \sigma_b^2 + \sigma^2 \end{bmatrix}$$

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).

e conseqüentemente, 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 (2.3.1) com

$$\mathbf{y} = (\mathbf{y}_{11}^\top, \dots, \mathbf{y}_{1n_1}^\top, \mathbf{y}_{21}^\top, \dots, \mathbf{y}_{2n_2}^\top)^\top,$$

$$\mathbf{X} = \bigoplus_{i=1}^2 (\mathbf{1}_{n_i} \otimes \mathbf{I}_2), \quad \boldsymbol{\beta} = (\mu_{11}, \mu_{12}, \mu_{21}, \mu_{22})^\top, \quad \mathbf{Z} = \bigotimes_{l=1}^{n_1+n_2} \mathbf{1}_2,$$

$$\mathbf{b} = (b_{11}, \dots, b_{1n_1}, b_{21}, \dots, b_{2n_2})^\top$$

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 com duas fontes de variação. Nesse contexto, as hipóteses de interesse podem ser testadas por meio de estatísticas como aquelas indicadas na Tabela 3.1.2 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

$$SQETr = \sum_{i=1}^2 n_i (\bar{s}_i - \bar{s})^2,$$

$$\text{com } \bar{s}_i = n_i^{-1} \sum_{j=1}^{n_i} s_{ij} \text{ e } \bar{s} = (n_1 + n_2)^{-1} \sum_{i=1}^2 \sum_{j=1}^{n_i} s_{ij},$$

$$SQE1 = \sum_{i=1}^2 \sum_{j=1}^{n_i} (s_{ij} - \bar{s}_i)^2,$$

$$SQTrTp = \sum_{i=1}^2 n_i (\bar{d}_i - \bar{d})^2,$$

$$\text{com } \bar{d}_i = n_i^{-1} \sum_{j=1}^{n_i} d_{ij} \text{ e } \bar{d} = (n_1 + n_2)^{-1} \sum_{i=1}^2 \sum_{j=1}^{n_i} d_{ij},$$

$$SQTp = [(n_1 n_2) / (n_1 + n_2)] (\bar{d}_1 + \bar{d}_2)^2,$$

e

$$SQE2 = \sum_{i=1}^2 \sum_{j=1}^{n_i} (d_{ij} - \bar{d}_i)^2.$$

Tabela 3.1.2: 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	$n - 2$	$SQE2$	$SQE2/(n - 2)$	

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.

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 que $s_s^2 = SQE1/(n - 2)$, podemos escrever

$$F_{TpTr} = \{(\bar{d}_1 - \bar{d}_2)/s_d\sqrt{(n_1n_2)/(n_1 + n_2)}\}^2,$$

$$F_{Tp} = \{(\bar{d}_1 + \bar{d}_2)/s_d\sqrt{(n_1n_2)/(n_1 + n_2)}\}^2,$$

e

$$F_{Tr} = \{(\bar{s}_1 - \bar{s}_2)/s_s\sqrt{(n_1n_2)/(n_1 + n_2)}\}^2.$$

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 (v), pode-se utilizar a estatística

$$F_{TrTp1} = \left\{ \sum_{j=1}^{n_i} (y_{11k} - y_{21k}) / s_1 \sqrt{(n_1n_2)/(n_1 + n_2)} \right\}^2$$

em que $s_1^2 = (n-2)^{-1} \sum_{i=1}^2 \sum_{j=1}^{n_i} (y_{ij1} - \bar{y}_{i1})^2$ com $\bar{y}_{i1} = n_i^{-1} \sum_{j=1}^{n_i} y_{ij1}$. Sob a hipótese nula de que as respostas médias das unidades amostrais dos dois tratamentos são iguais no pré-teste, a estatística F_{TrTp1} segue uma distribuição F_{1,n_1+n_2-2} .

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 (3.1.1)-(3.1.3).

Os resultados da análise dos dados do Exemplo 5.1.2 estão dispostos na Tabela 3.1.3.

Tabela 3.1.3: Tabela ANOVA para os dados do Exemplo 5.1.2

Fonte de variação	Graus de liberdade	Soma de quadrados	Quadrado médio	Estatística F	Valor-P
Grupo	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
Grupo \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 grupo são, respectivamente

$$\hat{\Sigma}_1 = \begin{bmatrix} 271.52 & 255.03 \\ & 284.61 \end{bmatrix} \text{ e } \hat{\Sigma}_2 = \begin{bmatrix} 108.24 & 74.75 \\ & 100.00 \end{bmatrix}$$

e a estimativa da matriz de covariâncias intraunidades amostrais comum, obtida sob a hipótese de homocedasticidade entre os dois grupos é

$$\hat{\Sigma} = \begin{bmatrix} 189.88 & 164.79 \\ & 192.30 \end{bmatrix}.$$

Além disso, as correspondentes estimativas das variâncias das diferenças das respostas no pré-teste e pós-teste são respectivamente $\hat{\sigma}_{d_1}^2 = 46.05$, $\hat{\sigma}_{d_2}^2 = 59.16$ e $\hat{\sigma}_d^2 = 52.60$ o que não traz evidências contrárias à suposição de homocedasticidade entre os dois grupos e que, consequentemente, não invalida a análise baseada no modelo (3.1.1)-(3.1.3). Em

particular, convém observar que

$$t_{TrTp}^2 = \{(-16.67 + 17.33)/\sqrt{52.60(1/12 + 1/12)}\}^2 = 0.05$$

é exatamente o valor da estatística F_{TrTp} obtido na Tabela 3.1.3.

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 5.1.2: 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 3.1.4, um gráfico de perfis individuais, na Figura 3.1.4 e um gráfico de perfis médios, na Figura 3.1.5.

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 média pós-teste - RHSR média 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,

Gráficos do tipo *boxplot* correspondentes estão dispostos na Figura 3.1.6.

Tabela 3.1.4: Resistência homogênea do sistema respiratório de ratos

Rato	Gás	Pré-teste	Pós-teste	Rato	Gás	Pré-teste	Pós-teste
1	T	0.0828	0.0347	15	C	0.0477	0.0461
2	T	0.0274	0.0172	16	C	0.0313	0.0114
3	T	0.0115	0.0022	17	C	-0.0437	0.0417
4	T	0.0112	0.0157	18	C	0.0358	0.0021
5	T	0.0065	0.0149	19	C	0.0477	0.0488
6	T	0.0025	0.0000	20	C	0.0579	0.0591
7	T	0.0357	0.0147	21	C	0.0664	0.0986
8	T	0.0232	0.0291	22	C	0.0871	0.0529
9	T	0.0356	0.0092	23	C	0.0517	0.0577
10	T	0.0458	0.0490	24	C	0.1102	0.1018
11	T	0.0836	0.0208	25	C	0.0776	0.0329
12	T	0.0638	0.0221	26	C	0.0970	0.0110
13	T	0.0909	0.0972	27	C	0.0157	0.0258
14	T	0.0281	0.0204	28	C	0.1203	0.1171
				29	C	0.0279	0.0176

Ventilação com T: ar sintético/hélio-oxigênio

Ventilação com C: ar sintético/ar sintético

-

Figura 3.1.4: Perfis individuais para RHSR de ratos

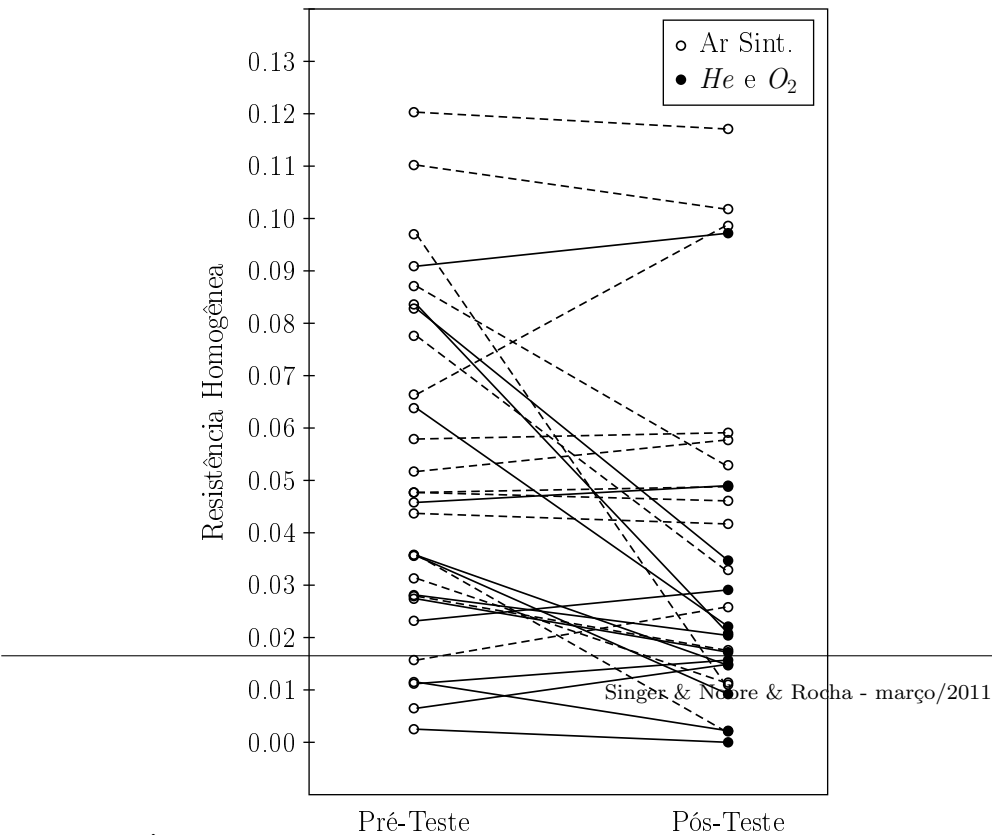


Figura 3.1.5: Valores das médias amostrais (círculos) da RHSR de ratos.

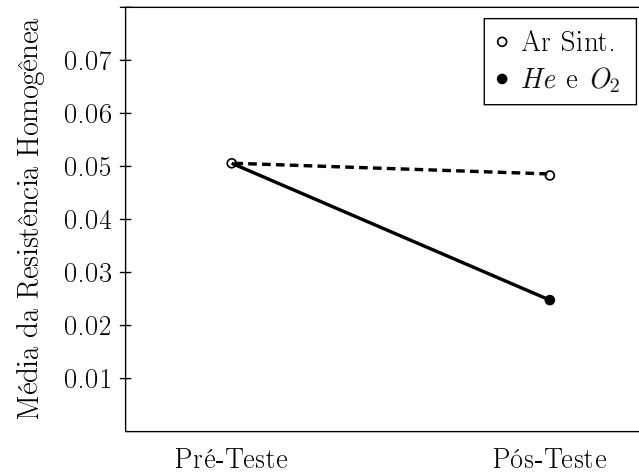
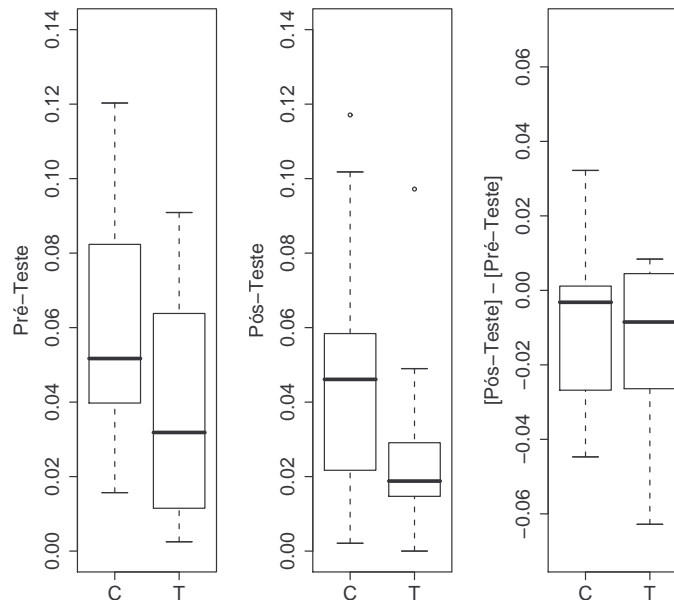


Figura 3.1.6: Boxplots para a RHSR de ratos antes (pré-teste), depois (pós-teste) e efeito da ventilação com ar sintético, ou com a mistura hélio/oxigênio



Embora os modelos considerados acima também possam ser emprega-

dos 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

$$y_{ij2} = \mu_2 + \alpha_i + \beta(y_{ij1} - \bar{y}_1) + e_{ij}, \quad i = 1, 2, \quad j = 1, \dots, n_i, \quad (3.1.4)$$

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 $\mathbb{E}(e_{ij}) = 0$, $\mathbb{V}(e_{ij}) = \sigma_e^2$ 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, à média geral da resposta no pós-teste, ao efeito do tratamento i resposta média e ao coeficiente de regressão linear comum (de y_{ij2} em função de y_{ij1}).

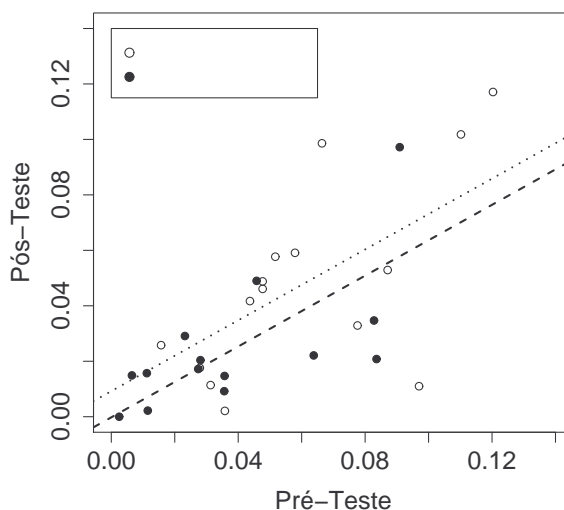
Um segundo modelo de ANCOVA é dado por

$$d_{ij} = \nu_2 + \gamma_i + \delta(y_{ij1} - \bar{y}_1) + e_{ij}, \quad i = 1, 2, \quad j = 1, \dots, n_i, \quad (3.1.5)$$

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 $\mathbb{E}(e_{ij}) = 0$, $\mathbb{V}(e_{ij}) = \sigma_e^2$ 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, à média geral da resposta no pós-teste, ao efeito do tratamento i resposta média e ao coeficiente de regressão linear comum (de y_{ij2} em função de y_{ij1}).

Para o Exemplo 5.1.2, o diagrama de dispersão para a RHSR dos ratos e as retas $\widehat{E}(y_{ij2}) = \widehat{\mu}_2 + \widehat{\alpha}_i + \widehat{\beta}(y_{ij1} - \bar{y}_1)$, $i = 1, 2$, obtidas via mínimos quadrados estão apresentados na Figura 3.1.7.

Figura 3.1.7: 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 esperadas segundo o modelo (3.1.4)



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).

3.2 Análise de perfis

O termo análise de perfis se refere ao conjunto de técnicas utilizadas para comparar t diferentes tratamentos (possivelmente definidos por meio do cruzamento dos níveis de dois ou mais fatores) com respeito à variação da resposta média ao longo de p 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 ao padrão de 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 médias (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 médias (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 p observações referentes à i -ésima unidade amostral ($i = 1, \dots, n_j$) submetida ao j -ésimo tratamento ($j = 1, \dots, t$) é dado por

$$\mathbf{y}_{ij} = (y_{ij1}, \dots, y_{ijp})^\top$$

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 ($k =$

$1, \dots, p$). Os perfis médios de respostas nos p níveis da condição de avaliação são dados por $\mathbb{E}(\mathbf{y}_{ij}^\top) = \boldsymbol{\mu}_j = (\mu_{j1}, \dots, \mu_{jp})^\top$, $j = 1, \dots, t$.

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

$$\mathbf{M} = \begin{bmatrix} \mu_{11} & \mu_{12} & \dots & \mu_{1p} \\ \mu_{21} & \mu_{22} & \dots & \mu_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ \mu_{t1} & \mu_{t2} & \dots & \mu_{tp} \end{bmatrix} \quad (3.2.1)$$

e $\mathbf{X} = \oplus_{j=1}^t \mathbf{1}_{n_j}$.

Em termos dos parâmetros μ_{jk} , a hipótese *i*) pode ser expressa como

$$H_I : \begin{bmatrix} \mu_{11} - \mu_{12} \\ \mu_{12} - \mu_{13} \\ \vdots \\ \mu_{1(p-1)} - \mu_{1p} \end{bmatrix} = \begin{bmatrix} \mu_{21} - \mu_{22} \\ \mu_{22} - \mu_{23} \\ \vdots \\ \mu_{2(p-1)} - \mu_{2p} \end{bmatrix} = \dots = \begin{bmatrix} \mu_{t1} - \mu_{t2} \\ \mu_{t2} - \mu_{t3} \\ \vdots \\ \mu_{t(p-1)} - \mu_{tp} \end{bmatrix}.$$

Na ausência de interação, ou seja, quando não rejeitamos H_I , as hipóteses *ii*) e *iii*) podem ser definidas como

$$H_T : \sum_{k=1}^p \mu_{1k} = \sum_{k=1}^p \mu_{2k} = \dots = \sum_{k=1}^p \mu_{tk}$$

e

$$H_P : \sum_{j=1}^t \mu_{j1} = \sum_{j=1}^t \mu_{j2} = \dots = \sum_{j=1}^t \mu_{jp}.$$

Outra forma para representar as hipóteses *ii*) e *iii*) independentemente da presença ou não da interação é

$$H_{TI} : \begin{bmatrix} \mu_{11} \\ \mu_{12} \\ \vdots \\ \mu_{1p} \end{bmatrix} = \begin{bmatrix} \mu_{21} \\ \mu_{22} \\ \vdots \\ \mu_{2p} \end{bmatrix} = \dots = \begin{bmatrix} \mu_{t1} \\ \mu_{t2} \\ \vdots \\ \mu_{tp} \end{bmatrix}$$

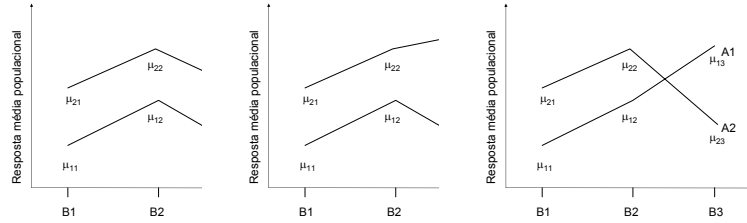
e

$$H_{PI}^* : \begin{bmatrix} \mu_{11} \\ \mu_{21} \\ \vdots \\ \mu_{t1} \end{bmatrix} = \begin{bmatrix} \mu_{12} \\ \mu_{22} \\ \vdots \\ \mu_{t2} \end{bmatrix} = \dots = \begin{bmatrix} \mu_{1p} \\ \mu_{2p} \\ \vdots \\ \mu_{tp} \end{bmatrix}.$$

Nas situações hipotéticas ilustradas na Figura 3.2, ilustramos os seguintes

o.

Figura 3.2.1: Perfis médios de resposta correspondentes a uma situações hipotéticas com 2 tratamentos e 3 tempos



casos: *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 médias entre tratamentos e a matriz \mathbf{U} é utilizada para definir as comparações das respostas médias 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 $(t-1) \times t$ 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 $p \times (p - 1)$. Para a hipótese H_{TI} , podemos utilizar

$$\mathbf{C} = \mathbf{C}_1 \text{ e } \mathbf{U} = \mathbf{1}_p$$

e para a hipótese H_{PI} ,

$$\mathbf{C} = \mathbf{1}_t^\top \text{ e } \mathbf{U} = \mathbf{U}_1.$$

Também podemos expressar modelos para análise de perfis na forma (2.3.1) com $\mathbf{X} = \oplus_{j=1}^t \mathbf{1}_{n_j} \otimes \mathbf{I}_p$, $\boldsymbol{\beta} = (\mu_{11}, \mu_{12}, \dots, \mu_{1p}, \dots, \mu_{t1}, \mu_{t2}, \dots, \mu_{tp})^\top$ e $\mathbf{Z} = \mathbf{0}$. As hipóteses H_I , H_{TI} e H_{PI} podem ser expressas na forma (2.3.30), que neste caso se reduz a

$$\mathbf{L}\boldsymbol{\beta} = \mathbf{0} \quad (3.2.2)$$

com $\mathbf{L} = \mathbf{C} \otimes \mathbf{U}^\top$.

Como exemplo, consideremos um estudo com dois tratamentos ($t = 2$) e três condições de avaliação ($p = 3$). 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 (2.3.1) 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 (2.3.1) 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.

3.2.1 Uma estratégia para análise de perfis

Exemplo 5.2.1: 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 esquiva”, 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 (MODELCHO) 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 esquiva, 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 3.2.2 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.

O gráfico dos perfis médios apresentado na Figura 3.2.3 sugere uma possível interação entre tratamento e tempo.

Figura 3.2.2: Perfis individuais para os nove tratamentos

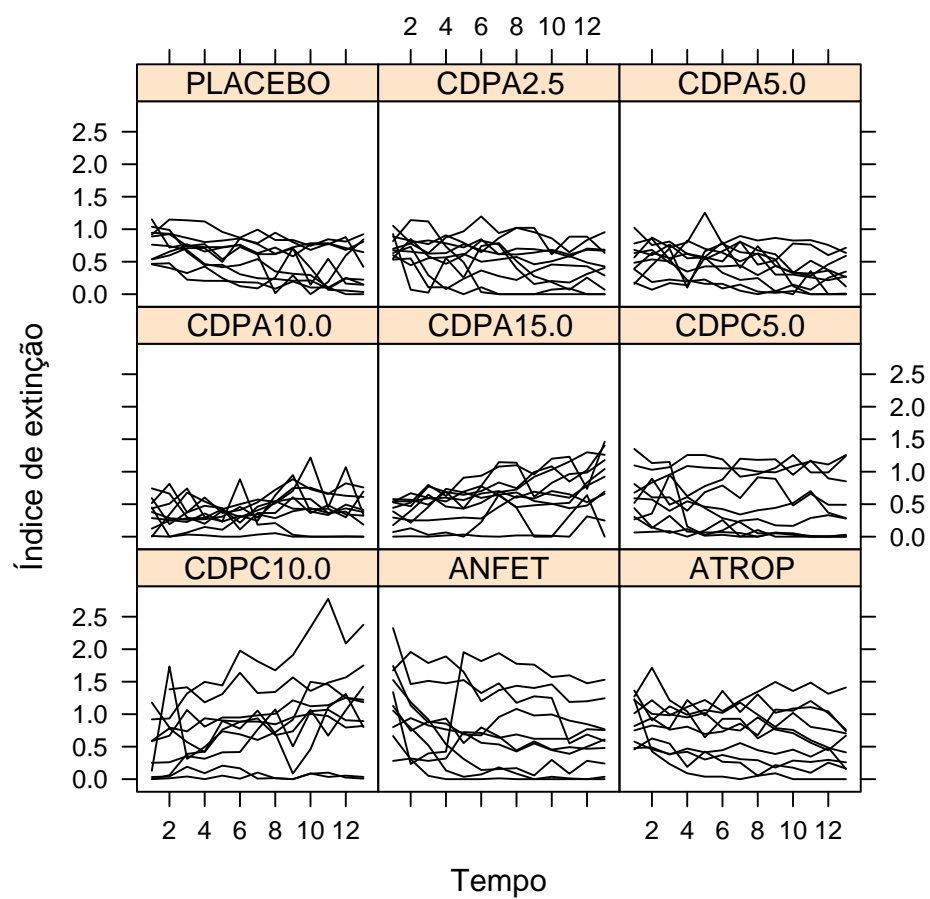
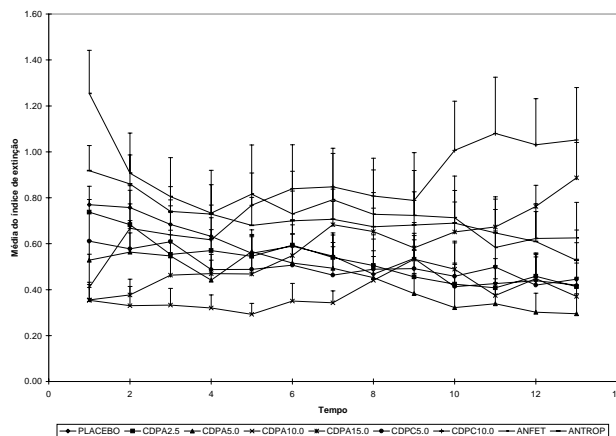


Figura 3.2.3: Perfis médios para os nove tratamentos



A matriz de correlações intraunidades amostrais é

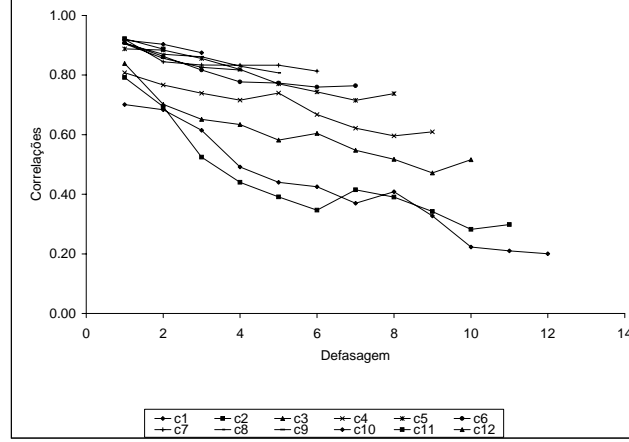
1												
0.70	1											
0.68	0.79	1										
0.61	0.69	0.84	1									
0.49	0.53	0.70	0.81	1								
0.44	0.44	0.65	0.77	0.89	1							
0.43	0.39	0.63	0.74	0.88	0.91	1						
0.37	0.35	0.58	0.72	0.86	0.86	0.92	1					
0.41	0.42	0.61	0.74	0.82	0.82	0.84	0.91	1				
0.33	0.39	0.55	0.67	0.77	0.78	0.83	0.87	0.91	1			
0.22	0.34	0.52	0.62	0.74	0.77	0.83	0.86	0.86	0.92	1		
0.21	0.28	0.47	0.60	0.72	0.76	0.83	0.83	0.83	0.90	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	0.91	1

(3.2.3)

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 3.2.4.

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

Figura 3.2.4: Correlações em função da defasagem



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 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

$$\mathbf{y}_{ij} = \boldsymbol{\mu}_j + \mathbf{e}_{ij}, \quad (3.2.4)$$

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

Tabela 3.2.1: 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

aleatórios

$$\mathbf{y}_{ij} = \boldsymbol{\mu}_j + \mathbf{Z}_i \mathbf{b}_i + \mathbf{e}_{ij}, \quad (3.2.5)$$

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 3.2.1 e confirmam as evidências favoráveis ao modelo de antedependência de ordem 1 para a matriz de covariâncias intra-unidades amostrais.

Adotando a matriz de covariâncias intraunidades amostrais identificada na Tabela 3.2.1, inicialmente testamos a hipótese de inexistência de interação entre Tratamento e Tempo (H_{0I}). Utilizando a aproximação F para a estatística (2.3.31) 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:

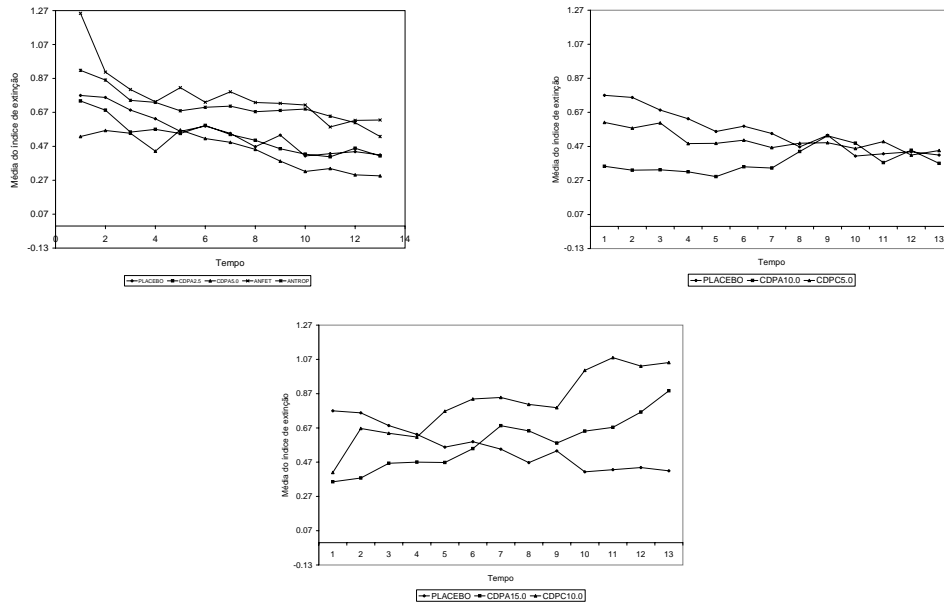
GRUPO 1: PLACEBO, CDPA2.5, CDPA5.0, ANFET, ATROP;

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 3.2.1. O tratamento PLACEBO foi incluído em todos os gráficos para comparação.

Figura 3.2.5: Perfis médios para os três grupos de tratamentos



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 cada grupo. Essa hipótese pode ser expressa na forma (3.2.2), com $\mathbf{L} = \mathbf{C} \otimes \mathbf{U}^T$ e

$$\mathbf{C} = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 \end{bmatrix} \quad (3.2.6)$$

Tabela 3.2.2: Testes para as hipóteses de existência de efeito de Tratamento e de Tempo dentro de grupos

Hipótese	C	U	F	gl	P
Tratamento	(3.2.6)	$\mathbf{1}_{13}$	1.75	(6, 963)	0.106
Tempo Grupo 1	(1 1 1 0 0 0 0 1 1)	(3.2.7)	2.85	(12, 963)	0.001
Tempo Grupo 2	(0 0 0 1 0 1 0 0 0)	(3.2.7)	0.76	(12, 963)	0.697
Tempo Grupo 3	(0 0 0 0 1 0 1 0 0)	(3.2.7)	3.64	(12, 963)	< 0.001

e

$$\mathbf{U} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \end{bmatrix}. \quad (3.2.7)$$

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 (3.2.2) com $\mathbf{L} = \mathbf{C} \otimes \mathbf{U}^\top$, assim como os valores das estatísticas F aproximadas, os respectivos graus de liberdade (g.l) e os valores-P dos testes correspondentes estão indicados na Tabela 3.2.1.

Esses resultados sugerem que:

- i) os 9 tratamentos podem ser classificados em 3 grupos de tratamentos equivalentes;

Tabela 3.2.3: Médias estimadas (erros padrões) para os três grupos de tratamentos

Grupo	Tempo	Estimativa	Erro padrão
1	1	0.84	0.05
1	2	0.75	0.05
1	3	0.67	0.05
1	4	0.62	0.05
1	5	0.63	0.06
1	6	0.63	0.06
1	7	0.62	0.06
1	8	0.57	0.06
1	9	0.56	0.06
1	10	0.51	0.06
1	11	0.48	0.06
1	12	0.49	0.06
1	13	0.46	0.06
2	1 - 13	0.44	0.07
3	1	0.36	0.09
3	2	0.52	0.09
3	3	0.55	0.08
3	4	0.54	0.08
3	5	0.62	0.09
3	6	0.69	0.09
3	7	0.77	0.09
3	8	0.73	0.09
3	9	0.68	0.10
3	10	0.83	0.10
3	11	0.88	0.10
3	12	0.90	0.09
3	13	0.97	0.10

- 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 3.2.1 apresentamos as médias e respectivos erros padrões estimados sob o modelo que incorpora o resultado da análise.

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 maiores 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:

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} = ((a_{ij}))_{1 \leq i \leq n, 1 \leq j \leq m}.$$

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

¹Neste texto consideramos apenas matrizes com elementos reais.

única coluna:

$$\mathbf{u} = \begin{pmatrix} u_{11} \\ u_{21} \\ \vdots \\ u_{m1} \end{pmatrix}.$$

Nesta obra, 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 \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 $(m \times q)$ com 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 tais que os produtos \mathbf{AB} , \mathbf{AC} e \mathbf{BC} estejam bem definidos. Então valem as seguintes propriedades:

- 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*

$$\mathbf{A} = \begin{pmatrix} -1 & 3 \\ 1 & 2 \\ -2 & -2 \end{pmatrix},$$

então

$$\mathbf{A}^\top = \begin{pmatrix} -1 & 1 & -2 \\ 3 & 2 & -2 \end{pmatrix}.$$

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

$$\mathbf{I}\mathbf{A} = \mathbf{A}\mathbf{I}.$$

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:*

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ 0 & a_{22} & a_{23} & \dots & a_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & a_{nn} \end{pmatrix},$$

²Quando a ordem da matriz identidade for evidente, ela será denotada \mathbf{I} .

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*

$$\mathbf{A}\mathbf{A} = \mathbf{A}^2 = \mathbf{A}.$$

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

$$\mathbf{A} = \begin{pmatrix} 2 & 5 & 1 & 4 \\ -7 & 4 & 0 & 10 \\ 0,3 & 7 & 20 & 8 \end{pmatrix},$$

podemos obter, por exemplo, as seguintes submatrizes de \mathbf{A}

$$\begin{pmatrix} 2 & 5 & 1 \\ -7 & 4 & 0 \\ 0,3 & 7 & 20 \end{pmatrix}$$

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 $m_2 = 1$. 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 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} é de 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 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.*

Assumindo que todas as 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

$$(\mathbf{A} \pm \mathbf{a}\mathbf{b}^\top)^{-1} = \mathbf{A}^{-1} \mp \frac{(\mathbf{A}^{-1}\mathbf{a})(\mathbf{b}^\top \mathbf{A}^{-1})}{1 \pm \mathbf{b}^\top \mathbf{A}^{-1} \mathbf{a}};$$

- 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 maiores 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

$$\mathbf{A} = \begin{pmatrix} 6 & 7 & 4 \\ 5 & 9 & 1 \\ 3 & 8 & -2 \end{pmatrix}, \quad (\text{A.1.1})$$

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 $(\sum_{i=1}^n n_i \times \sum_{i=1}^n m_i)$

$$\bigoplus_{i=1}^n \mathbf{A}_i = \mathbf{A}_1 \oplus \mathbf{A}_2 \oplus \dots \oplus \mathbf{A}_n = \begin{pmatrix} \mathbf{A}_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_2 & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{A}_n \end{pmatrix}.$$

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 reais 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. Pode-se mostrar que valem as seguintes propriedades:

- 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) $(\mathbf{A} \otimes \mathbf{B}) \otimes (\mathbf{C} \otimes \mathbf{D}) = (\mathbf{AC}) \otimes (\mathbf{BD})$;
- 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 $(n \times 1)$, 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} 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 valem as seguintes propriedades:

- i) $\text{vec}(\mathbf{a}^\top) = \text{vec}(\mathbf{a})$;
- ii) $\text{vec}(\mathbf{ab}^\top) = \mathbf{b} \otimes \mathbf{a}$;
- iii) $\text{vec}(\mathbf{AB}) = (\mathbf{I}_p \otimes \mathbf{A})\text{vec}(\mathbf{B}) = (\mathbf{B}^\top \otimes \mathbf{I}_m)\text{vec}(\mathbf{A})$;
- iv) $\text{vec}(\mathbf{ABC}) = (\mathbf{C}^\top \otimes \mathbf{A})\text{vec}(\mathbf{B})$;
- v) $\text{vec}(\mathbf{ABC}) = (\mathbf{I}_q \otimes \mathbf{AB})\text{vec}(\mathbf{C}) = (\mathbf{C}^\top \mathbf{B}^\top \otimes \mathbf{I}_n)\text{vec}(\mathbf{A})$. 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

$$\text{vec}(\mathbf{A}^\top)^\top \text{vec}(\mathbf{B}) = \text{vec}(\mathbf{B}^\top)^\top \text{vec}(\mathbf{A}) = \text{tr}(\mathbf{AB});$$

- 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

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix},$$

então

$$\text{vech}(\mathbf{A}) = \begin{pmatrix} a_{11} \\ a_{21} \\ a_{31} \\ a_{22} \\ a_{32} \\ a_{33} \end{pmatrix}.$$

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 $\mathbf{a} + \mathbf{b} \in \mathcal{V}$ para as quais 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 $\mathbf{a}, \mathbf{b} \in \mathcal{U}$ 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 $\forall \mathbf{x} \in \mathbb{R}^p$,

$$\mathbf{T}(\mathbf{x}) = \mathbf{Ax} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1p} \\ a_{21} & a_{22} & \dots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{np} \end{pmatrix} \begin{pmatrix} x_{11} \\ x_{21} \\ \vdots \\ x_{p1} \end{pmatrix}.$$

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 nulo de uma matriz: *Seja \mathbf{X} uma matriz de dimensão $(n \times p)$ e $\mathcal{C}(\mathbf{X})$ o espaço coluna de \mathbf{X} . O espaço nulo de \mathbf{X} , denotado $\mathcal{N}(\mathbf{X})$, é o conjunto de vetores $\mathbf{a} \in \mathbb{R}^p$ tais que $\mathbf{Xa} = \mathbf{0}$, ou seja, $\mathcal{N}(\mathbf{X}) = \{\mathbf{a} \in \mathbb{R}^p; \mathbf{Xa} = \mathbf{0}\}$.*

Teorema A.2.2. *Seja \mathbf{X} uma matriz de dimensão $(n \times p)$ com $r(\mathbf{X}) = r$. 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 possui 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 $\mathbf{v} \bullet \mathbf{u} = 0 \ \forall \mathbf{u} \in \mathcal{U}$.

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 $\mathbf{v} = (2 \ -3)^\top$. 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*

$$i) \quad |\mathbf{A}| = \prod_{i=1}^n \lambda_i;$$

$$ii) \quad \text{tr}(\mathbf{A}) = \sum_{i=1}^n \lambda_i.$$

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 quadrada de ordem n de coeficientes.

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 têm 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 $\mathbf{y} \sim \mathcal{N}_p(\boldsymbol{\mu}, \mathbf{V})$ 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 the ordem r the $\mathbf{y}^\top \mathbf{A} \mathbf{y}$ é

$$K_r(\mathbf{y}^\top \mathbf{A} \mathbf{y}) = 2^{r-1}(r-1)![\text{tr}(\mathbf{A} \mathbf{V})^r + r \boldsymbol{\mu}^\top \mathbf{A} (\mathbf{V} \mathbf{A})^{r-1} \boldsymbol{\mu}];$$

$$iii) \mathbb{Cov}(\mathbf{y}, \mathbf{y}^\top \mathbf{A} \mathbf{y}) = 2 \mathbf{V} \mathbf{A} \boldsymbol{\mu};$$

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 $\mathbf{y} \sim \mathcal{N}_p(\boldsymbol{\mu}, \mathbf{V})$ 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 $\mathbf{y} \sim \mathcal{N}_p(\boldsymbol{\mu}, \mathbf{V})$, \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 $\mathbf{y} \sim \mathcal{N}_p(\boldsymbol{\mu}, \mathbf{V})$, \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

$$\mathbf{f}(\mathbf{x}) = \begin{pmatrix} f_1(\mathbf{x}) \\ f_2(\mathbf{x}) \\ f_3(\mathbf{x}) \end{pmatrix}$$

com $f_1(\mathbf{x}) = x_1 + x_2$, $f_2(\mathbf{x}) = e^{x_1 x_2}$ e $f_3(\mathbf{x}) = x_1 x_2$, por exemplo.

iii) matrizes do tipo

$$\begin{aligned} \mathbf{F}(\mathbf{x}) &= (\mathbf{f}_1(\mathbf{x}) \quad \mathbf{f}_2(\mathbf{x}) \quad \dots \quad \mathbf{f}_n(\mathbf{x})) \\ &= \begin{pmatrix} f_{11}(\mathbf{x}) & f_{12}(\mathbf{x}) & \dots & f_{1n}(\mathbf{x}) \\ f_{21}(\mathbf{x}) & f_{22}(\mathbf{x}) & \dots & f_{2n}(\mathbf{x}) \\ \vdots & \vdots & \ddots & \vdots \\ f_{m1}(\mathbf{x}) & f_{m2}(\mathbf{x}) & \dots & f_{mn}(\mathbf{x}) \end{pmatrix}; \end{aligned}$$

por exemplo, se $\mathbf{f}_1(\mathbf{x}) = (x_1 + x_2, x_1 x_2, x_1 - x_2)^\top$ e $\mathbf{f}_2(\mathbf{x}) = (x_1, x_1 + x_2, x_1 x_2)^\top$, então

$$\begin{aligned} \mathbf{F}(\mathbf{x}) &= (\mathbf{f}_1(\mathbf{x}) \quad \mathbf{f}_2(\mathbf{x})) \\ &= \begin{pmatrix} f_{11}(\mathbf{x}) & f_{12}(\mathbf{x}) \\ f_{21}(\mathbf{x}) & f_{22}(\mathbf{x}) \\ f_{31}(\mathbf{x}) & f_{32}(\mathbf{x}) \end{pmatrix} \\ &= \begin{pmatrix} x_1 + x_2 & x_1 \\ x_1 x_2 & x_1 + x_2 \\ x_1 - x_2 & x_1 x_2 \end{pmatrix}. \end{aligned}$$

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

$$\begin{aligned} \frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} &= \left(\frac{\partial f(\mathbf{x})}{\partial x_1}, \frac{\partial f(\mathbf{x})}{\partial x_2}, \frac{\partial f(\mathbf{x})}{\partial x_3} \right)^\top \\ &= (4x_1, 8x_2, 10x_3)^\top. \end{aligned}$$

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

$$\frac{\partial f(\mathbf{x})}{\partial \mathbf{x} \partial \mathbf{x}^\top} = \begin{pmatrix} \partial^2 f(\mathbf{x})/\partial x_1^2 & \partial^2 f(\mathbf{x})/\partial x_1 \partial x_2 & \dots & \partial^2 f(\mathbf{x})/\partial x_1 \partial x_p \\ \partial^2 f(\mathbf{x})/\partial x_2 \partial x_1 & \partial^2 f(\mathbf{x})/\partial x_2^2 & \dots & \partial^2 f(\mathbf{x})/\partial x_2 \partial x_p \\ \vdots & \vdots & \ddots & \vdots \\ \partial^2 f(\mathbf{x})/\partial x_p \partial x_1 & \partial^2 f(\mathbf{x})/\partial x_p \partial x_2 & \dots & \partial^2 f(\mathbf{x})/\partial x_p^2 \end{pmatrix}.$$

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

$$\nabla \mathbf{f}(\mathbf{x}) = \frac{\partial \mathbf{f}(\mathbf{x})}{\partial \mathbf{x}^\top} = \begin{pmatrix} \partial \mathbf{f}_1(\mathbf{x})/\partial x_1 & \partial \mathbf{f}_1(\mathbf{x})/\partial x_2 & \dots & \partial \mathbf{f}_1(\mathbf{x})/\partial x_p \\ \partial \mathbf{f}_2(\mathbf{x})/\partial x_1 & \partial \mathbf{f}_2(\mathbf{x})/\partial x_2 & \dots & \partial \mathbf{f}_2(\mathbf{x})/\partial x_p \\ \vdots & \vdots & \ddots & \vdots \\ \partial \mathbf{f}_m(\mathbf{x})/\partial x_1 & \partial \mathbf{f}_m(\mathbf{x})/\partial x_2 & \dots & \partial \mathbf{f}_m(\mathbf{x})/\partial x_p \end{pmatrix}.$$

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

$$\nabla \mathbf{F}(\mathbf{x}) = \frac{\partial \mathbf{F}(\mathbf{x})}{\partial \mathbf{x}^\top} = \begin{pmatrix} \partial \mathbf{f}_1(\mathbf{x})/\partial \mathbf{x}^\top & \partial \mathbf{f}_2(\mathbf{x})/\partial \mathbf{x}^\top & \dots & \partial \mathbf{f}_n(\mathbf{x})/\partial \mathbf{x}^\top \end{pmatrix}.$$

Para o exemplo no início da seção, a derivada da função $\mathbf{F}(\mathbf{x})$ é dada por

$$\begin{aligned} \frac{\partial \mathbf{F}(\mathbf{x})}{\partial \mathbf{x}^\top} &= \begin{pmatrix} \partial \mathbf{f}_1(\mathbf{x})/\partial x_1 & \partial \mathbf{f}_1(\mathbf{x})/\partial x_2 & \partial \mathbf{f}_2(\mathbf{x})/\partial x_1 & \partial \mathbf{f}_2(\mathbf{x})/\partial x_2 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 1 & 1 & 0 \\ x_2 & x_1 & 1 & 1 \\ 1 & -1 & x_2 & x_1 \end{pmatrix}. \end{aligned}$$

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

$$\begin{aligned} \frac{\partial f}{\partial \mathbf{X}} &= (\partial f / \partial x_{ij}) \\ &= \begin{pmatrix} \partial f / \partial x_{11} & \partial f / \partial x_{12} & \dots & \partial f / \partial x_{1n} \\ \partial f / \partial x_{21} & \partial f / \partial x_{22} & \dots & \partial f / \partial x_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ \partial f / \partial x_{m1} & \partial f / \partial x_{m2} & \dots & \partial f / \partial x_{mn} \end{pmatrix}. \end{aligned}$$

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) $\partial \mathbf{x}^\top \mathbf{A} \mathbf{x} / \partial \mathbf{x} = (\mathbf{A} + \mathbf{A}^\top) \mathbf{x} (= 2\mathbf{A} \mathbf{x}, \text{ se } \mathbf{A} \text{ for simétrica});$

iv) Para \mathbf{A} e \mathbf{B} simétricas,

$$\partial (\mathbf{x}^\top \mathbf{A} \mathbf{x} / \mathbf{x}^\top \mathbf{B} \mathbf{x}) / \partial \mathbf{x} = 2\mathbf{A} \mathbf{x} / \mathbf{x}^\top \mathbf{B} \mathbf{x} - [\mathbf{x}^\top \mathbf{A} \mathbf{x} / (\mathbf{x}^\top \mathbf{B} \mathbf{x})^2] 2\mathbf{B} \mathbf{x};$$

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) $\partial|\mathbf{AX}|/\partial\mathbf{X} = |\mathbf{AX}|((\mathbf{AX})^{-1}\mathbf{A})^\top.$

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) $\partial\text{vec}(\mathbf{AXB})/\partial\text{vec}(\mathbf{X})^\top = \mathbf{B}^\top \otimes \mathbf{A};$

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

$$\begin{aligned} \frac{\partial\text{vec}[\mathbf{U}(\mathbf{X})\mathbf{V}(\mathbf{X})]}{\partial\text{vec}(\mathbf{X})^\top} &= (\mathbf{V} \otimes \mathbf{I}_m)^\top \frac{\partial\text{vec}[\mathbf{U}(\mathbf{X})]}{\partial\text{vec}(\mathbf{X})^\top} \\ &\quad + (\mathbf{I}_r \otimes \mathbf{U})^\top \frac{\partial\text{vec}[\mathbf{V}(\mathbf{X})]}{\partial\text{vec}(\mathbf{X})^\top}; \end{aligned}$$

xx) Seja \mathbf{X} uma matriz quadrada; então

$$\frac{\partial\text{vec}(\mathbf{AX}^{-1}\mathbf{B})}{\partial\text{vec}(\mathbf{X})^\top} = -(\mathbf{XB})^\top \otimes (\mathbf{AX}^{-1});$$

xxi) Para uma matriz não singular $\mathbf{U}(\mathbf{X})$,

$$\frac{\partial\text{vec}[\mathbf{U}(\mathbf{X})^{-1}]}{\partial\text{vec}(\mathbf{X})^\top} = -[(\mathbf{U}(\mathbf{X})^{-1})^\top \otimes \mathbf{U}(\mathbf{X})^{-1}] \frac{\partial\text{vec}[\mathbf{U}(\mathbf{X})]}{\partial\text{vec}(\mathbf{X})^\top};$$

xxii)

$$\frac{\partial\text{vec}\{\mathbf{F}[\mathbf{G}(\mathbf{X})]\}}{\partial\text{vec}(\mathbf{Z})^\top} = \frac{\partial\text{vec}[\mathbf{F}(\mathbf{X})]}{\partial\text{vec}(\mathbf{X})^\top} \bigg|_{\mathbf{X}=\mathbf{G}(\mathbf{Z})} \frac{\partial\text{vec}[\mathbf{G}(\mathbf{Z})]}{\partial\text{vec}(\mathbf{Z})^\top}.$$

Para ilustrar a aplicação de derivadas de vetores e matrizes, inicialmente, consideremos a função

$$\begin{aligned} g(\boldsymbol{\beta}, \boldsymbol{\theta}) &= (\mathbf{y} - \mathbf{f}(\boldsymbol{\beta}))^\top \mathbf{V}^{-1}(\boldsymbol{\theta})(\mathbf{y} - \mathbf{f}(\boldsymbol{\beta})) + \ln |\mathbf{V}(\boldsymbol{\theta})| \\ &= \text{tr}[\mathbf{V}^{-1}(\boldsymbol{\theta})(\mathbf{y} - \mathbf{f}(\boldsymbol{\beta}))(\mathbf{y} - \mathbf{f}(\boldsymbol{\beta}))^\top] + \ln |\mathbf{V}(\boldsymbol{\theta})| \end{aligned}$$

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:

$$\nabla g(\boldsymbol{\beta}, \boldsymbol{\theta}) = \begin{pmatrix} (\partial/\partial\boldsymbol{\beta})g(\boldsymbol{\beta}, \boldsymbol{\theta}) \\ (\partial/\partial\boldsymbol{\theta})g(\boldsymbol{\beta}, \boldsymbol{\theta}) \end{pmatrix}$$

com

$$\begin{aligned} \frac{\partial g(\boldsymbol{\beta}, \boldsymbol{\theta})}{\partial\boldsymbol{\beta}} &= \frac{\partial(\mathbf{y} - \mathbf{f}(\boldsymbol{\beta}))^\top \mathbf{V}^{-1}(\boldsymbol{\theta})(\mathbf{y} - \mathbf{f}(\boldsymbol{\beta}))}{\partial\boldsymbol{\beta}} \\ &= -2 \left(\frac{\partial \mathbf{f}^\top(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}} \right) \mathbf{V}^{-1}(\boldsymbol{\theta})(\mathbf{y} - \mathbf{f}(\boldsymbol{\beta})). \end{aligned}$$

Além disso, segue que para $j = 1, \dots, k$

$$\begin{aligned} \frac{\partial g(\boldsymbol{\beta}, \boldsymbol{\theta})}{\partial\theta_j} &= \frac{\partial[(\mathbf{y} - \mathbf{f}(\boldsymbol{\beta}))^\top \mathbf{V}(\boldsymbol{\theta})^{-1}(\mathbf{y} - \mathbf{f}(\boldsymbol{\beta}))]}{\partial\theta_j} + \frac{\partial \ln |\mathbf{V}(\boldsymbol{\theta})|}{\partial\theta_j} \\ &= (\mathbf{y} - \mathbf{f}(\boldsymbol{\beta}))^\top \frac{\partial \mathbf{V}(\boldsymbol{\theta})^{-1}}{\partial\theta_j} (\mathbf{y} - \mathbf{f}(\boldsymbol{\beta})) + \frac{\partial \ln |\mathbf{V}(\boldsymbol{\theta})|}{\partial\theta_j} \\ &= -(\mathbf{y} - \mathbf{f}(\boldsymbol{\beta}))^\top \mathbf{V}(\boldsymbol{\theta})^{-1} \frac{\partial \mathbf{V}(\boldsymbol{\theta})^{-1}}{\partial\theta_j} \mathbf{V}(\boldsymbol{\theta})^{-1} (\mathbf{y} - \mathbf{f}(\boldsymbol{\beta})) \\ &\quad + \text{tr} \left[\mathbf{V}(\boldsymbol{\theta})^{-1} \frac{\partial \mathbf{V}(\boldsymbol{\theta})}{\partial\theta_j} \right] \\ &= -\text{tr} \left[\mathbf{V}(\boldsymbol{\theta})^{-1} \frac{\partial \mathbf{V}(\boldsymbol{\theta})^{-1}}{\partial\theta_j} \mathbf{V}(\boldsymbol{\theta})^{-1} (\mathbf{y} - \mathbf{f}(\boldsymbol{\beta})) (\mathbf{y} - \mathbf{f}(\boldsymbol{\beta}))^\top \right] \\ &\quad + \text{tr} \left[\mathbf{V}(\boldsymbol{\theta})^{-1} \frac{\partial \mathbf{V}(\boldsymbol{\theta})}{\partial\theta_j} \right] \\ &= \text{tr} \left[\mathbf{V}(\boldsymbol{\theta})^{-1} [-(\mathbf{y} - \mathbf{f}(\boldsymbol{\beta})) (\mathbf{y} - \mathbf{f}(\boldsymbol{\beta}))^\top + \mathbf{V}(\boldsymbol{\theta})] \mathbf{V}(\boldsymbol{\theta})^{-1} \frac{\partial \mathbf{V}(\boldsymbol{\theta})}{\partial\theta_j} \right]. \end{aligned}$$

A matriz de segundas derivadas em relação ao vetor $\boldsymbol{\beta}$ é:

$$\begin{aligned} \frac{\partial^2 g(\boldsymbol{\beta}, \boldsymbol{\theta})}{\partial\boldsymbol{\beta}^\top \partial\boldsymbol{\beta}} &= \frac{\partial \left(-2 \left(\frac{\partial \mathbf{f}^\top(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}} \right) \mathbf{V}^{-1}(\boldsymbol{\theta})(\mathbf{y} - \mathbf{f}(\boldsymbol{\beta})) \right)}{\partial\boldsymbol{\beta}^\top} \\ &= -2 \frac{\partial^2 \mathbf{f}^\top(\boldsymbol{\beta})}{\partial\boldsymbol{\beta} \partial\boldsymbol{\beta}^\top} \mathbf{V}^{-1}(\boldsymbol{\theta})(\mathbf{y} - \mathbf{f}(\boldsymbol{\beta})) \\ &\quad + 2 \left(\frac{\partial \mathbf{f}^\top(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}} \right) \mathbf{V}^{-1}(\boldsymbol{\theta}) \left(\frac{\partial \mathbf{f}(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}^\top} \right). \end{aligned}$$

Para obter $\partial^2 g(\boldsymbol{\beta}, \boldsymbol{\theta}) / \partial \beta_i \partial \theta_j$, notemos que para $i = 1, \dots, p$

$$\begin{aligned} \frac{\partial g(\boldsymbol{\beta}, \boldsymbol{\theta})}{\partial \beta_i} &= \frac{\partial (\mathbf{y} - \mathbf{f}(\boldsymbol{\beta}))^\top \mathbf{V}^{-1}(\boldsymbol{\theta})(\mathbf{y} - \mathbf{f}(\boldsymbol{\beta}))}{\partial \beta_i} \\ &= -2 \left(\frac{\partial \mathbf{f}^\top(\boldsymbol{\beta})}{\partial \beta_i} \right) \mathbf{V}^{-1}(\boldsymbol{\theta})(\mathbf{y} - \mathbf{f}(\boldsymbol{\beta})). \end{aligned}$$

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 $i = 1, \dots, p$ e $j = 1, \dots, k$ de forma que

$$\begin{aligned} \frac{\partial^2 g(\boldsymbol{\beta}, \boldsymbol{\theta})}{\partial \beta_i \partial \theta_j} &= \frac{\partial \left(-2 \left(\frac{\partial \mathbf{f}^\top(\boldsymbol{\beta})}{\partial \beta_i} \right) \mathbf{V}^{-1}(\boldsymbol{\theta})(\mathbf{y} - \mathbf{f}(\boldsymbol{\beta})) \right)}{\partial \theta_j} \\ &= -2 \left(\frac{\partial \mathbf{f}^\top(\boldsymbol{\beta})}{\partial \beta_i} \right) \frac{\partial \mathbf{V}^{-1}(\boldsymbol{\theta})}{\partial \theta_j} (\mathbf{y} - \mathbf{f}(\boldsymbol{\beta})) \\ &= -2 \left(\frac{\partial \mathbf{f}^\top(\boldsymbol{\beta})}{\partial \beta_i} \right) \mathbf{V}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{V}^{-1}(\boldsymbol{\theta})}{\partial \theta_j} \mathbf{V}^{-1}(\boldsymbol{\theta})(\mathbf{y} - \mathbf{f}(\boldsymbol{\beta})). \end{aligned}$$

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

$$\begin{aligned} \frac{\partial^2 g(\boldsymbol{\beta}, \boldsymbol{\theta})}{\partial \theta_s \partial \theta_j} &= \\ &= \frac{\partial}{\partial \theta_s} \left\{ \text{tr} \left[\mathbf{V}(\boldsymbol{\theta})^{-1} [-(\mathbf{y} - \mathbf{f}(\boldsymbol{\beta}))(\mathbf{y} - \mathbf{f}(\boldsymbol{\beta}))^\top + \mathbf{V}(\boldsymbol{\theta})] \mathbf{V}(\boldsymbol{\theta})^{-1} \frac{\partial \mathbf{V}(\boldsymbol{\theta})}{\partial \theta_j} \right] \right\} \\ &= \frac{\partial \left\{ -(\mathbf{y} - \mathbf{f}(\boldsymbol{\beta}))^\top \mathbf{V}(\boldsymbol{\theta})^{-1} \frac{\partial \mathbf{V}(\boldsymbol{\theta})^{-1}}{\partial \theta_j} \mathbf{V}(\boldsymbol{\theta})^{-1} (\mathbf{y} - \mathbf{f}(\boldsymbol{\beta})) \right\}}{\partial \theta_s} \\ &\quad + \frac{\partial \left\{ \text{tr} \left[\mathbf{V}(\boldsymbol{\theta})^{-1} \frac{\partial \mathbf{V}(\boldsymbol{\theta})}{\partial \theta_j} \right] \right\}}{\partial \theta_s} \\ &= -(\mathbf{y} - \mathbf{f}(\boldsymbol{\beta}))^\top \frac{\partial \left[\mathbf{V}(\boldsymbol{\theta})^{-1} \frac{\partial \mathbf{V}(\boldsymbol{\theta})^{-1}}{\partial \theta_j} \mathbf{V}(\boldsymbol{\theta})^{-1} \right]}{\partial \theta_s} (\mathbf{y} - \mathbf{f}(\boldsymbol{\beta})) \\ &\quad + \text{tr} \left\{ \frac{\partial \left[\mathbf{V}(\boldsymbol{\theta})^{-1} \frac{\partial \mathbf{V}(\boldsymbol{\theta})}{\partial \theta_j} \right]}{\partial \theta_s} \right\}. \end{aligned}$$

Utilizando a regra do produto para derivadas de matrizes, temos que

$$\begin{aligned} \frac{\partial \left[\mathbf{V}(\boldsymbol{\theta})^{-1} \frac{\partial \mathbf{V}(\boldsymbol{\theta})}{\partial \theta_j} \right]}{\partial \theta_s} &= \frac{\partial \mathbf{V}^{-1}(\boldsymbol{\theta})}{\partial \theta_s} \frac{\partial \mathbf{V}(\boldsymbol{\theta})}{\partial \theta_j} + \mathbf{V}^{-1}(\boldsymbol{\theta}) \frac{\partial}{\partial \theta_s} \left[\frac{\partial \mathbf{V}(\boldsymbol{\theta})}{\partial \theta_j} \right] \\ &= -\mathbf{V}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{V}(\boldsymbol{\theta})}{\partial \theta_s} \mathbf{V}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{V}(\boldsymbol{\theta})}{\partial \theta_j} + \mathbf{V}^{-1}(\boldsymbol{\theta}) \frac{\partial^2 \mathbf{V}(\boldsymbol{\theta})}{\partial \theta_s \partial \theta_j} \end{aligned}$$

e

$$\begin{aligned} \frac{\partial}{\partial \theta_s} \left[\mathbf{V}(\boldsymbol{\theta})^{-1} \frac{\partial \mathbf{V}(\boldsymbol{\theta})^{-1}}{\partial \theta_j} \mathbf{V}(\boldsymbol{\theta})^{-1} \right] &= \\ &= \frac{\partial \left[\mathbf{V}(\boldsymbol{\theta})^{-1} \frac{\partial \mathbf{V}(\boldsymbol{\theta})}{\partial \theta_j} \right]}{\partial \theta_s} \mathbf{V}^{-1}(\boldsymbol{\theta}) + \mathbf{V}(\boldsymbol{\theta}) \frac{\partial \mathbf{V}(\boldsymbol{\theta})}{\partial \theta_j} \frac{\partial \mathbf{V}^{-1}(\boldsymbol{\theta})}{\partial \theta_s} \\ &= -\mathbf{V}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{V}(\boldsymbol{\theta})}{\partial \theta_s} \mathbf{V}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{V}(\boldsymbol{\theta})}{\partial \theta_j} \mathbf{V}^{-1}(\boldsymbol{\theta}) \\ &\quad + \mathbf{V}^{-1}(\boldsymbol{\theta}) \frac{\partial^2 \mathbf{V}(\boldsymbol{\theta})}{\partial \theta_s \partial \theta_j} \mathbf{V}^{-1}(\boldsymbol{\theta}) \\ &\quad - \mathbf{V}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{V}(\boldsymbol{\theta})}{\partial \theta_j} \mathbf{V}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{V}(\boldsymbol{\theta})}{\partial \theta_s} \mathbf{V}^{-1}(\boldsymbol{\theta}). \end{aligned}$$

Assim

$$\begin{aligned} \frac{\partial^2 g(\boldsymbol{\beta}, \boldsymbol{\theta})}{\partial \theta_s \partial \theta_j} &= (\mathbf{y} - \mathbf{f}(\boldsymbol{\beta}))^\top \mathbf{V}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{V}(\boldsymbol{\theta})}{\partial \theta_s} \mathbf{V}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{V}(\boldsymbol{\theta})}{\partial \theta_j} \mathbf{V}^{-1}(\boldsymbol{\theta}) (\mathbf{y} - \mathbf{f}(\boldsymbol{\beta})) \\ &\quad - (\mathbf{y} - \mathbf{f}(\boldsymbol{\beta}))^\top \mathbf{V}^{-1}(\boldsymbol{\theta}) \frac{\partial^2 \mathbf{V}(\boldsymbol{\theta})}{\partial \theta_s \partial \theta_j} \mathbf{V}^{-1}(\boldsymbol{\theta}) (\mathbf{y} - \mathbf{f}(\boldsymbol{\beta})) \\ &\quad + (\mathbf{y} - \mathbf{f}(\boldsymbol{\beta}))^\top \mathbf{V}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{V}(\boldsymbol{\theta})}{\partial \theta_j} \mathbf{V}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{V}(\boldsymbol{\theta})}{\partial \theta_s} \mathbf{V}^{-1}(\boldsymbol{\theta}) (\mathbf{y} - \mathbf{f}(\boldsymbol{\beta})) \\ &\quad + \text{tr} \left[-\mathbf{V}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{V}(\boldsymbol{\theta})}{\partial \theta_s} \mathbf{V}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{V}(\boldsymbol{\theta})}{\partial \theta_j} + \mathbf{V}^{-1}(\boldsymbol{\theta}) \frac{\partial^2 \mathbf{V}(\boldsymbol{\theta})}{\partial \theta_s \partial \theta_j} \right]. \end{aligned}$$

Além disso, como $\mathbf{V}^{-1}(\boldsymbol{\theta})$, $\partial \mathbf{V}(\boldsymbol{\theta})/\partial \theta_j$ e $\partial \mathbf{V}(\boldsymbol{\theta})/\partial \theta_s$ são matrizes simétricas

e

$$(\mathbf{y} - \mathbf{f}(\boldsymbol{\beta}))^\top \mathbf{V}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{V}(\boldsymbol{\theta})}{\partial \theta_s} \mathbf{V}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{V}(\boldsymbol{\theta})}{\partial \theta_j} \mathbf{V}^{-1}(\boldsymbol{\theta}) (\mathbf{y} - \mathbf{f}(\boldsymbol{\beta}))$$

é um escalar, obtemos a igualdade

$$\begin{aligned} (\mathbf{y} - \mathbf{f}(\boldsymbol{\beta}))^\top \mathbf{V}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{V}(\boldsymbol{\theta})}{\partial \theta_s} \mathbf{V}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{V}(\boldsymbol{\theta})}{\partial \theta_j} \mathbf{V}^{-1}(\boldsymbol{\theta}) (\mathbf{y} - \mathbf{f}(\boldsymbol{\beta})) &= \\ (\mathbf{y} - \mathbf{f}(\boldsymbol{\beta}))^\top \mathbf{V}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{V}(\boldsymbol{\theta})}{\partial \theta_j} \mathbf{V}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{V}(\boldsymbol{\theta})}{\partial \theta_s} \mathbf{V}^{-1}(\boldsymbol{\theta}) (\mathbf{y} - \mathbf{f}(\boldsymbol{\beta})). \end{aligned}$$

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

$$\begin{aligned} \frac{\partial^2 g(\boldsymbol{\beta}, \boldsymbol{\theta})}{\partial \theta_s \partial \theta_j} &= \text{tr} \left\{ \mathbf{A} \left[2(\mathbf{y} - \mathbf{f}(\boldsymbol{\beta}))(\mathbf{y} - \mathbf{f}(\boldsymbol{\beta}))^\top - \mathbf{V}(\boldsymbol{\theta}) \right] \right\} \\ &+ \text{tr} \left\{ \mathbf{B} \left[-(\mathbf{y} - \mathbf{f}(\boldsymbol{\beta}))(\mathbf{y} - \mathbf{f}(\boldsymbol{\beta}))^\top + \mathbf{V}(\boldsymbol{\theta}) \right] \right\}, \end{aligned}$$

em que

$$\begin{aligned} \mathbf{A} &= \mathbf{V}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{V}(\boldsymbol{\theta})}{\partial \theta_s} \mathbf{V}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{V}(\boldsymbol{\theta})}{\partial \theta_j} \mathbf{V}^{-1}(\boldsymbol{\theta}), \\ \mathbf{B} &= \mathbf{V}^{-1}(\boldsymbol{\theta}) \frac{\partial^2 \mathbf{V}(\boldsymbol{\theta})}{\partial \theta_s \partial \theta_j} \mathbf{V}^{-1}(\boldsymbol{\theta}). \end{aligned}$$

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 $\mathbf{A}^2 = \mathbf{AA} = \mathbf{A}$. 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 (p.d.) se $\mathbf{x}^\top \mathbf{A} \mathbf{x} > 0$, $\forall \mathbf{x} \neq \mathbf{0}$. 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 p.d. 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) $\text{r}(\mathbf{I}_n - \mathbf{H}) = n - p$.
- vii) Construa exemplos numéricos (com matrizes diferentes da matriz identidade) para ilustrar as propriedades acima.

A.6.8. Seja a matriz

$$\mathbf{A} = \begin{pmatrix} \rho & 1 \\ 1 & \rho \end{pmatrix}.$$

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 $\partial f / \partial x_{ij}$:

$$\frac{\partial f(\mathbf{X})}{\partial \mathbf{X}} := \begin{pmatrix} \frac{\partial f(\mathbf{X})}{\partial x_{11}} & \cdots & \frac{\partial f(\mathbf{X})}{\partial x_{1n}} \\ \vdots & \ddots & \vdots \\ \frac{\partial f(\mathbf{X})}{\partial x_{m1}} & \cdots & \frac{\partial f(\mathbf{X})}{\partial x_{mn}} \end{pmatrix}.$$

Sejam $\mathbf{x} = (x_1, \dots, x_n)^\top$ e $\mathbf{a} = (a_1, \dots, a_n)^\top$, vetores com dimensão $(n \times 1)$ e \mathbf{A} , uma matriz quadrada de ordem n . Mostre que

$$\begin{aligned} \frac{\partial \mathbf{a}^\top \mathbf{x}}{\partial \mathbf{x}} &= \frac{\partial \mathbf{x}^\top \mathbf{a}}{\partial \mathbf{x}} = \mathbf{a} \\ \frac{\partial \mathbf{x}^\top \mathbf{A} \mathbf{x}}{\partial \mathbf{x}} &= (\mathbf{A} + \mathbf{A}^\top) \mathbf{x}. \end{aligned}$$

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 .

A.6.13. Sejam \mathbf{A} uma matriz com dimensão $m \times n$, \mathbf{B} uma matriz quadrada de ordem n e \mathbf{x} um vetor com dimensão $(n \times 1)$. Mostre que

- i) $\partial \mathbf{Ax} / \partial \mathbf{x} = \mathbf{A}^\top$;
- ii) $\partial \mathbf{x}^\top \mathbf{Bx} / \partial \mathbf{x} = (\mathbf{B} + \mathbf{B}^\top) \mathbf{x}$;
- iii) $\partial \mathbf{x}^\top \mathbf{Bx} / \partial \mathbf{x} = 2\mathbf{Bx}$ se \mathbf{B} for simétrica.

Apêndice B

Regressão

B.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 denominada em textos de econometria), e uma ou mais variáveis explicativas (independentes, preditoras, de controle ou exógenas) ou, especificamente estimar o valor médio (ou prever 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

Assumamos que a resposta média da i -ésima unidade amostral condicional 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 percorrido y_i por um corpo em queda livre após um tempo

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 $r(\mathbf{X}) = p$.

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, *i.e.*, do tipo

$$\mu(\mathbf{x}_i, \boldsymbol{\beta}) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip};$$

ii) linear somente nos parâmetros, *i.e.*, do tipo

$$\mu(\mathbf{x}_i, \boldsymbol{\beta}) = \beta_0 + \beta_1 x_{i1}^2 + \beta_2 \ln(x_{i2}^2) + \dots + \beta_p \sqrt{x_{ip}};$$

iii) linearizável, *i.e.*, do tipo

$$\mu(\mathbf{x}_i, \boldsymbol{\beta}) = \exp(\beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip})$$

ou

$$\mu(\mathbf{x}_i, \boldsymbol{\beta}) = \beta_0 x_{i1}^{\beta_1} \dots x_{ip}^{\beta_p};$$

iv) não linear,

$$\mu(\mathbf{x}_i, \boldsymbol{\beta}) = \beta_0 + \exp(\beta_1 x_{i1}^2) + \beta_2 \ln(x_{i2}^2) + \dots + \beta_p \sqrt{x_{ip}}.$$

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).

Condicionalmente aos valores das variáveis explicativas, \mathbf{x}_i , a resposta da i -ésima unidade amostral se situa em torno de $\mu(\mathbf{x}_i, \boldsymbol{\beta})$; Inclusão da seção de parametrizaçãop 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

$$y_i = \mu(\mathbf{x}_i, \boldsymbol{\beta}) + e_i, \quad i = 1, \dots, n$$

ou na forma matricial, como

$$\mathbf{y} = \mu(\mathbf{X}, \boldsymbol{\beta}) + \mathbf{e}$$

com $\mathbf{e} = [e_1, \dots, e_n]^\top$. Quando $\mu(\mathbf{X}, \boldsymbol{\beta}) = \mathbf{X}\boldsymbol{\beta}$, temos

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e} \tag{B.1.1}$$

que é o **modelo de regressão linear múltipla**. Explicitamente esse modelo corresponde a

$$y_i = \mu(\mathbf{x}_i, \boldsymbol{\beta}) + e_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip} + e_i, \quad i = 1, \dots, n.$$

Quando $p = 1$, ele é denominado **modelo de regressão linear simples**.

Usualmente assumimos que

$$\mathbb{E}(\mathbf{e}) = \mathbf{0} \text{ e que } \mathbb{V}(\mathbf{e}) = \sigma^2 \mathbf{I} \quad (\text{B.1.2})$$

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 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 de 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 B.1.1: Os dados da Tabela B.1.1 correspondem à idade e a uma medida de pressão arterial sistólica para um conjunto de 20 pacientes de uma determinada clínica.

Tabela B.1.1: Idade (anos) e pressão arterial sistólica (mmHg)

Paciente	Pressão sistólica	Idade
1	114	17
2	134	18
3	116	20
4	139	23
5	110	34
6	150	38
7	152	41
8	138	42
9	142	46
10	145	47
11	156	47
12	159	47
13	142	50
14	156	52
15	164	57
16	185	60
17	162	64
18	176	66
19	175	69
20	180	70

Para estudar a associação entre idade (X) e pressão arterial sistólica (Y), consideramos o modelo de regressão linear simples

$$y_i = \alpha + \beta x_i + e_i, \quad i = 1, \dots, 20$$

que pode ser expresso na forma matricial (B.1.1) com

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{20} \end{bmatrix}, \mathbf{X} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_{20} \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \alpha \\ \beta \end{bmatrix} \text{ e } \mathbf{e} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_{20} \end{bmatrix}.$$

Para efeito de interpretação, convém reescrever o modelo na forma

$$y_i = \alpha^* + \beta(x_i - x_0) + e_i, \quad i = 1, \dots, 20$$

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 de 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

$$y_i = \alpha + \beta x_i + e_i, \quad i = 1, \dots, 20$$

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. Mais especificamente, β representa a variação na pressão arterial sistólica esperada por ano para pacientes com as mesmas características daqueles investigados.

Exemplo B.1.2: A Tabela B.1.2 contém dados de capacidade instalada (ton), potência instalada (1000 kW) e área construída (100 m²) de 10 empresas de uma certa indústria.

Tabela B.1.2: Capacidade instalada (ton), potência instalada (1000 kW) e área construída (100 m²) de empresas de uma certa indústria

Capacidade de produção	4.5	5.9	4.2	5.2	8.1	9.7	10.7	11.9	12.0	12.3
Potência instalada	0.9	2.5	1.3	1.8	3.1	4.6	6.1	6.0	5.9	6.1
Área construída	7.1	10.4	7.2	8.2	8.5	11.9	12.1	12.5	12.0	11.3

Para estimar a capacidade instalada (Y) a partir das informações sobre potência instalada (X_1) e área construída (X_2), consideramos o seguinte modelo:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + e_i, \quad i = 1, \dots, 10,$$

que genericamente, *i.e.*, para n empresas, pode ser expresso na forma matricial (B.1.1) com

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{21} \\ 1 & x_{12} & x_{22} \\ \vdots & \vdots & \\ 1 & x_{1n} & x_{2n} \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} \text{ e } \mathbf{e} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}.$$

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 a variável resposta e 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 e área construída iguais a zero implicariam capacidade de produção nula.

Exemplo B.1.3: A Tabela B.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 B.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

$$y_{ij} = \mu_i + e_{ij}, \quad i = 1, \dots, k, \quad j = 1, \dots, n_i \quad (\text{B.1.3})$$

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 (B.1.1) com

² Aqui expressamos o modelo sob a parametrização de médias de celas. Outras parametrizações podem ser consideradas; o leitor encontrará mais detalhes na Seção B.6.

$$\mathbf{y} = \begin{bmatrix} y_{11} \\ \vdots \\ y_{1n_1} \\ y_{21} \\ \vdots \\ y_{2n_2} \\ \vdots \\ y_{k1} \\ \vdots \\ y_{kn_k} \end{bmatrix}, \mathbf{X} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ \vdots & \vdots & \dots & 0 \\ 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & 1 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_k \end{bmatrix}$$

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, permita acomodar um efeito não-linear desse fator. 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

$$y_{ij} = \beta_0 + \beta_1 x_i + e_{ij}, \quad i = 1, \dots, k, \quad j = 1, \dots, n_i \quad (\text{B.1.4})$$

que pode ser expresso na forma matricial (B.1.1) com

$$\mathbf{y} = \begin{bmatrix} y_{11} \\ \vdots \\ y_{1n_1} \\ y_{21} \\ \vdots \\ y_{2n_2} \\ \vdots \\ y_{k1} \\ \vdots \\ y_{kn_k} \end{bmatrix}, \mathbf{X} = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_k \\ \vdots & \vdots \\ 1 & x_k \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \text{ e } \mathbf{e} = \begin{bmatrix} e_{11} \\ \vdots \\ e_{1n_1} \\ e_{21} \\ \vdots \\ e_{2n_2} \\ \vdots \\ e_{k1} \\ \vdots \\ e_{kn_k} \end{bmatrix}.$$

Exemplo B.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 numa forma eficiente. Os dados da Tabela B.1.4 correspondem aos números de consultas (acumuladas semanalmente) da temporada invernal realizadas nos anos de 2000 a 2004 numa certa localidade.

Tabela B.1.4: Consultas acumuladas na temporada invernal de 2000 a 2004

Semana N°	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 o número esperado de consultas possa ser representado por intermédio de regressões lineares simples, poderíamos considerar um modelo do tipo

$$y_{ij} = \beta_{0i} + \beta_{1i}x_{ij} + e_{ij}, \quad i = 1, \dots, k, \quad j = 1, \dots, n_i, \quad (\text{B.1.5})$$

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, $k = 5$ e $n_i = 10$. Na forma matricial (B.1.1), esse modelo

pode ser expresso como

$$\mathbf{y} = \begin{bmatrix} y_{11} \\ \vdots \\ y_{1n_1} \\ y_{21} \\ \vdots \\ y_{2n_2} \\ \vdots \\ y_{k1} \\ \vdots \\ y_{kn_k} \end{bmatrix}, \mathbf{X} = \begin{bmatrix} 1 & 0 & \dots & 0 & x_{11} & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & \dots & 0 & x_{1n_1} & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 & 0 & x_{21} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 1 & \dots & 0 & 0 & x_{2n_2} & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 & 0 & \dots & x_{k1} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 & 0 & 0 & \dots & x_{kn_k} \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \beta_{01} \\ \beta_{02} \\ \vdots \\ \beta_{0k} \\ \beta_{11} \\ \beta_{12} \\ \vdots \\ \beta_{1k} \end{bmatrix}$$

e

$$\mathbf{e} = [e_{11}, \dots, e_{1n_1}, e_{21}, \dots, e_{2n_2}, \dots, e_{k1}, \dots, e_{kn_k}]^\top$$

em que $n = n_1 + n_2 + \dots + n_k$.

Se considerarmos que existe um número fixo de pacientes com gripe 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

$$y_{ij} = \beta_0 + \beta_{1i}x_{ij} + e_{ij}, \quad i = 1, \dots, k, \quad j = 1, \dots, n_i \quad (\text{B.1.6})$$

ou na forma matricial (B.1.1), com

$$\mathbf{y} = \begin{bmatrix} y_{11} \\ \vdots \\ y_{1n_1} \\ y_{21} \\ \vdots \\ y_{2n_2} \\ \vdots \\ y_{k1} \\ \vdots \\ y_{kn_k} \end{bmatrix}, \mathbf{X} = \begin{bmatrix} 1 & x_{11} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n_1} & 0 & \dots & 0 \\ 1 & 0 & x_{21} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & x_{2n_2} & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & \dots & x_{k1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \dots & x_{kn_k} \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_{11} \\ \beta_{12} \\ \vdots \\ \beta_{1k} \end{bmatrix}$$

e \mathbf{e} como acima.

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

$$y_{ij} = \beta_{0i} + \beta_1 x_{ij} + e_{ij}, \quad i = 1, \dots, k, \quad j = 1, \dots, n_i$$

e na forma matricial (B.1.1), com

$$\mathbf{y} = \begin{bmatrix} y_{11} \\ \vdots \\ y_{1n_1} \\ y_{21} \\ \vdots \\ y_{2n_2} \\ \vdots \\ y_{k1} \\ \vdots \\ y_{kn_k} \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} 1 & 0 & \dots & 0 & x_{11} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & 0 & \dots & 0 & x_{1n_1} \\ 0 & 1 & \dots & 0 & x_{21} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 1 & \dots & 0 & x_{2n_2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & x_{k1} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & x_{kn_k} \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_{01} \\ \beta_{02} \\ \vdots \\ \beta_{0k} \\ \beta_1 \end{bmatrix}$$

e \mathbf{e} como acima.

Exemplo B.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 (B.1.1) com

$$\mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_j \\ y_{j+1} \\ \vdots \\ y_n \end{bmatrix}, \mathbf{X} = \begin{bmatrix} 1 & x_1 & 0 \\ \vdots & \vdots & \vdots \\ 1 & x_j & 0 \\ 1 & x_0 & (x_{j+1} - x_0) \\ \vdots & \vdots & \vdots \\ 1 & x_0 & (x_n - x_0) \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} \text{ e } \mathbf{e} = \begin{bmatrix} e_1 \\ \vdots \\ e_j \\ e_{j+1} \\ \vdots \\ e_n \end{bmatrix}.$$

Sob a formulação matricial (B.1.1), hipóteses lineares de interesse podem ser expressas na forma

$$H : \mathbf{C}\boldsymbol{\beta} = \mathbf{m} \quad (\text{B.1.7})$$

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 B.1.1, a hipótese de que o intercepto é nulo é expressa como (B.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 (B.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 B.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 (B.1.7) com

$$\mathbf{C} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

No Exemplo B.1.3, sob o modelo (B.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 (B.1.7) em que $\mathbf{m} = \mathbf{0}$ é um vetor $[(k-1) \times 1]$ e \mathbf{C} é uma matriz $[(k-1) \times k]$ construída de forma a gerar $k-1$ contrastes linearmente independentes dos elementos de $\boldsymbol{\beta} = [\mu_1 \ \mu_2 \ \dots \ \mu_k]^\top$. Como exemplo desse

tipo de matriz podemos ter

$$\mathbf{C} = \begin{bmatrix} 1 & 0 & \dots & 0 & -1 \\ 0 & 1 & \dots & 0 & -1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & -1 \end{bmatrix}.$$

Sob o modelo (B.1.4), uma hipótese equivalente seria expressa com $\mathbf{C} = [0 \ 1]$ e $\mathbf{m} = 0$.

No Exemplo B.1.4, a hipótese a ser testada para avaliar se o modelo (B.1.5) pode ser reduzido ao modelo (B.1.6) também pode ser expressa na forma (B.1.7) em que \mathbf{C} é uma matriz $[(k-1) \times 2k]$ com $r(\mathbf{C}) = k-1$ 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 $[(k-1) \times 1]$.

No Exemplo B.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 (B.1.7) com $\mathbf{C} = [0 \ 1 \ -1]$.

B.2 Método de mínimos quadrados

O método de mínimos quadrados (ordinários) pode ser utilizada para estimar o vetor de parâmetros $\boldsymbol{\beta}$ no modelo (B.1.1). A base do método é a minimização da forma quadrática

$$Q(\boldsymbol{\beta}) = \mathbf{e}^\top \mathbf{e} = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = \sum_{i=1}^n e_i^2. \quad (\text{B.2.1})$$

Com essa finalidade, consideremos

$$Q(\boldsymbol{\beta}) = \mathbf{y}^\top \mathbf{y} + \boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{X} \boldsymbol{\beta} - 2\boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{y}$$

e utilizemos as expressões para derivadas matriciais apresentadas no Apêndice A para obter

$$\begin{aligned}\frac{\partial Q(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} &= 2\mathbf{X}^\top \mathbf{X} \boldsymbol{\beta} - 2\mathbf{X}^\top \mathbf{y}, \\ \frac{\partial^2 Q(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^2} &= 2\mathbf{X}^\top \mathbf{X}.\end{aligned}$$

Igualando a primeira derivada parcial a zero, temos $\mathbf{X}^\top \mathbf{X} \hat{\boldsymbol{\beta}} = \mathbf{X}^\top \mathbf{y}$. Como assumimos que \mathbf{X} tem posto completo³, então $\mathbf{X}^\top \mathbf{X}$ é uma matriz não singular e a solução do **sistema de equações de estimação** (também conhecido por **sistema de equações normais**) é⁴

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}. \quad (\text{B.2.2})$$

Como a matriz $\mathbf{X}^\top \mathbf{X}$ é definida positiva, o estimador (B.2.2) corresponde ao ponto de mínimo de (B.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) $\mathbb{E}(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$,
- ii) $\mathbb{V}(\hat{\boldsymbol{\beta}}) = \sigma^2 (\mathbf{X}^\top \mathbf{X})^{-1}$.

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 B.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*

³Quando 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 $\mathbf{X}^\top \mathbf{X} \hat{\boldsymbol{\beta}} = \mathbf{X}^\top \mathbf{y}$.

⁴Aqui, 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.

$\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}$ 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 (B.2.2) permite-nos concluir que

$$\hat{\boldsymbol{\beta}} \sim \mathcal{N}_p[\boldsymbol{\beta}, \sigma^2 (\mathbf{X}^\top \mathbf{X})^{-1}]. \quad (\text{B.2.3})$$

Como, em geral a variância σ^2 é desconhecida, inferências sobre $\boldsymbol{\beta}$ dependem de um estimador desse parâmetro e um candidato é o estimador não enviesado

$$\begin{aligned} s^2 &= (n-p)^{-1} (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^\top (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \\ &= (n-p)^{-1} \mathbf{y}^\top [\mathbf{I}_n - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top] \mathbf{y}. \end{aligned}$$

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 $(n-p)\sigma^{-2}s^2 \sim \chi_{n-p}^2$.

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 (B.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 $\delta = (\mathbf{C}\boldsymbol{\beta} - \mathbf{m})^\top [\mathbf{C}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{C}^\top]^{-1} (\mathbf{C}\boldsymbol{\beta} - \mathbf{m})$. 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

$$\mathbf{A} = \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{C} [\mathbf{C}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{C}^\top]^{-1} \mathbf{C}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top.$$

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 $\mathbf{AB} = \mathbf{0}$, podemos utilizar o Teorema A.3.5 para mostrar que

$$F = c^{-1} s^{-2} (\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{m})^\top [\mathbf{C}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{C}^\top]^{-1} (\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{m}) \sim F_{(c, n-p)}(\delta), \quad (\text{B.2.4})$$

com $\delta = (\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{m})^\top [\mathbf{C}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{C}^\top]^{-1} (\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{m})$.

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 (B.1.7), basta notar que sob a hipótese nula $\delta = 0$ e a estatística (B.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 (B.2.4) com $\mathbf{C} = [0 \ 1]$ e $\mathbf{m} = \beta_0$. Nesse caso, a estatística (B.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 do sinal⁵ de $\hat{\beta} - \beta_0$, 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, consideremos 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

$$\text{a) } \max_{1 \leq i \leq n} \mathbf{x}_{ni}^\top (\mathbf{X}_n^\top \mathbf{X}_n)^{-1} \mathbf{x}_{ni} \rightarrow 0 \text{ com } n \rightarrow \infty, \quad ^6$$

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

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 & Singer (1993), por exemplo), podemos mostrar que

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) \xrightarrow{\mathcal{D}} \mathcal{N}_p(\mathbf{0}, \sigma^2 \mathbf{V}^{-1})$$

ou equivalentemente, que

$$(\mathbf{X}_n^\top \mathbf{X}_n)^{1/2}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) \xrightarrow{\mathcal{D}} \mathcal{N}_p(\mathbf{0}, \sigma^2 \mathbf{I}_p).$$

Em termos práticos, esse resultado significa que para amostras com tamanhos suficientemente “grandes”,

$$\hat{\boldsymbol{\beta}} \approx \mathcal{N}_p[\boldsymbol{\beta}, \sigma^2(\mathbf{X}^\top \mathbf{X})^{-1}]$$

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 (B.1.7) podemos empregar a estatística de Wald

$$Q_W = s^{-2}(\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{m})^\top [\mathbf{C}(\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{C}^\top]^{-1}(\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{m}), \quad (\text{B.2.5})$$

cujas 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 B.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

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

casos de interesse prático. Em particular, no caso de problemas envolvendo a comparação de k médias, ela corresponde à exigência de que os tamanhos n_i das amostras colhidas de cada subpopulação investigada sejam tais que $n_i / \sum_{i=1}^k n_i$ convirja para uma constante λ_i com $n = \sum_{i=1}^k n_i \rightarrow \infty$.

Quando a matriz \mathbf{V} é uma matriz diagonal com elementos $w_i\sigma^2$ na diagonal principal, o método é chamado de **mínimos quadrados ponderados** e a função a ser minimizada pode ser expressa como

$$Q(\boldsymbol{\beta}) = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top \mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = \sum_{i=1}^n w_i^{-1} (y_i - \mathbf{x}_i^\top \boldsymbol{\beta})^2.$$

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 $\mathbf{e} \sim \mathcal{N}(\mathbf{0}, \mathbf{V})$ com \mathbf{V} conhecida, então podemos demonstrar que

$$\tilde{\boldsymbol{\beta}} \sim \mathcal{N}_p[\boldsymbol{\beta}, (\mathbf{X}^\top \mathbf{V}^{-1} \mathbf{X})^{-1}]$$

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,

$$\tilde{\boldsymbol{\beta}} \approx \mathcal{N}_p[\boldsymbol{\beta}, (\mathbf{X}^\top \mathbf{V}^{-1} \mathbf{X})^{-1}]$$

Como, em geral, \mathbf{V} não é conhecida, podemos substituí-la por um estimador consistente $\hat{\mathbf{V}}$ e considerar o estimador

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \hat{\mathbf{V}}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \hat{\mathbf{V}}^{-1} \mathbf{y}. \quad (\text{B.2.6})$$

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 & Singer (1993)), por exemplo, podemos mostrar que para n suficientemente grande

$$\hat{\boldsymbol{\beta}} \approx \mathcal{N}_p[\boldsymbol{\beta}, (\mathbf{X}^\top \mathbf{V}^{-1} \mathbf{X})^{-1}]. \quad (\text{B.2.7})$$

Intervalos de confiança aproximados para combinações lineares da forma $\mathbf{C}\boldsymbol{\beta}$ ou testes para hipóteses do tipo (B.1.7) podem ser concretizados utilizando (B.2.7) ou a estatística de Wald

$$Q_W = (\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{m})^\top [\mathbf{C}(\mathbf{X}^\top \hat{\mathbf{V}}^{-1} \mathbf{X})^{-1} \mathbf{C}^\top]^{-1} (\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{m}), \quad (\text{B.2.8})$$

cuja distribuição sob a hipótese nula pode ser aproximada por uma distribuição χ_c^2 .

B.3 Método de máxima verossimilhança

Sob a suposição de normalidade, podemos considerar o método de máxima verossimilhança para a estimação dos parâmetros do modelo (B.1.1), que consiste em maximizar a função de verossimilhança

$$L(\boldsymbol{\beta}, \sigma^2 | \mathbf{y}, \mathbf{X}) = (2\pi\sigma^2)^{-n/2} \exp\left[-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right]. \quad (\text{B.3.1})$$

com relação aos parâmetros $\boldsymbol{\beta}$ e σ^2 . Lembrando que a determinação dos pontos de máximo de (B.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 (B.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 (B.1.1) com $\mathbf{e} \sim \mathcal{N}(\mathbf{0}, \mathbf{V})$ 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

$$L(\boldsymbol{\beta}, \boldsymbol{\theta} | \mathbf{y}, \mathbf{X}) = (2\pi)^{-n/2} |\mathbf{V}(\boldsymbol{\theta})|^{-1/2} \exp[-(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top \mathbf{V}^{-1}(\boldsymbol{\theta})(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})]. \quad (\text{B.3.2})$$

Nesse caso, o estimador de β tem a mesma forma que o estimador de mínimos quadrados generalizados (B.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.

B.4 Partição da soma de quadrados

A utilidade do modelo de regressão para explicação da variação da resposta esperada como função das variáveis explicativas pode ser avaliada a partir da identidade

$$y_i - \bar{y} = (\hat{y}_i - \bar{y}) + (y_i - \hat{y}_i)$$

que implica

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^n (y_i - \hat{y}_i)^2.$$

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 $\mathbb{E}(\mathbf{y}) = \mathbf{X}\beta$ e a variabilidade da resposta em torno deste modelo, ou seja a parcela da variabilidade total não explicada.

Quando o ajuste do modelo $\mathbb{E}(\mathbf{y}) = \mathbf{X}\beta$ é perfeito, *i.e.*, quando $\mathbf{y} = \hat{\mathbf{y}}$ temos $SQE = 0$ e quando o modelo $\mathbb{E}(\mathbf{y}) = \mathbf{X}\beta$ 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 devida à 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 $n-p$.

B.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 B.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 B.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 de um modelo de regressão linear simples do tipo

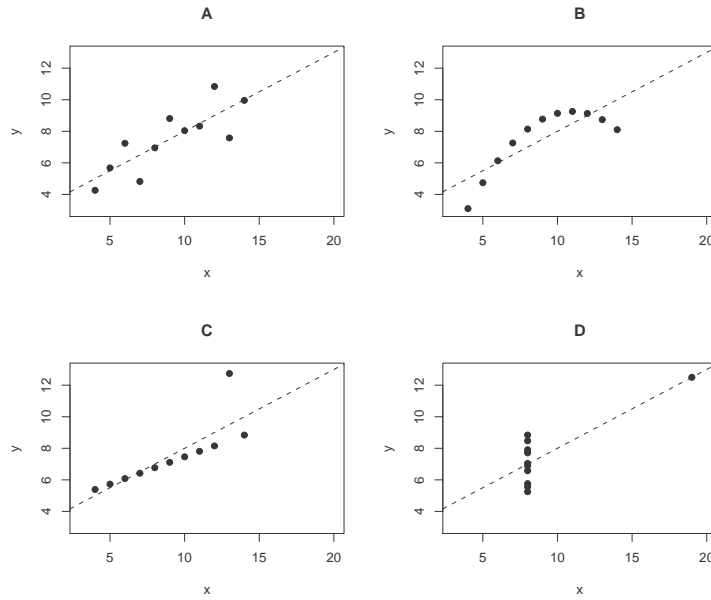
$$y_i = \alpha + \beta x_i + e_i, \quad i = 1, \dots, 11.$$

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 B.5.1 apresentamos diagramas de dispersão correspondentes a cada subconjunto; as linhas tracejadas correspondem às retas ajustadas pelo método de mínimos quadrados ordinários. 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

⁷Grosso modo, uma observação é chamada discrepante, se apresenta um comportamento distinto das demais.

avaliação da qualidade do ajuste. Essa avaliação precisa ser complementada por outros métodos que descrevemos a seguir.

Figura B.5.1: Diagramas de dispersão (dados da Tabela B.5.1)



B.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, como veremos a seguir. Uma referência clássica é Cox & Snell (1968); esses autores apresentam uma forma geral para definir resíduos para modelos que contêm uma única fonte de variação.

Consideremos o modelo de regressão linear

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e} \quad (\text{B.5.1})$$

com $\mathbf{e} \sim \mathcal{N}_n(\mathbf{0}, \sigma^2 \mathbf{I}_n)$. 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 ??). O vetor de **resíduos ordinários** é definido como

$$\hat{\mathbf{e}} = [\hat{e}_1, \dots, \hat{e}_n]^\top = \mathbf{y} - \hat{\mathbf{y}} = (\mathbf{I} - \mathbf{H})\mathbf{y}. \quad (\text{B.5.2})$$

Sob o modelo (B.5.1),

$$\hat{\mathbf{e}} = (\mathbf{I} - \mathbf{H})\mathbf{y} = (\mathbf{I} - \mathbf{H})(\mathbf{X}\boldsymbol{\beta} + \mathbf{e}) = (\mathbf{I} - \mathbf{H})\mathbf{e}, \quad (\text{B.5.3})$$

de onde podemos concluir que $\hat{\mathbf{e}} \sim \mathcal{N}_n(\mathbf{0}, \sigma^2[\mathbf{I} - \mathbf{H}])$. A relação entre o vetor de erros e o vetor de resíduos ordinários depende somente da **matriz chapéu** (*hat matrix*) \mathbf{H} .⁸

Algumas transformações dos resíduos ordinários têm sido sugeridas para fins de diagnóstico. Apresentaremos apenas algumas delas, mas o leitor interessado pode consultar Cook & Weisberg (1982), por exemplo, para maiores 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} .⁹ 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

$$\hat{e}_i^* = \frac{\hat{e}_i}{s\sqrt{1 - h_{ii}}}, \quad i = 1, \dots, n.$$

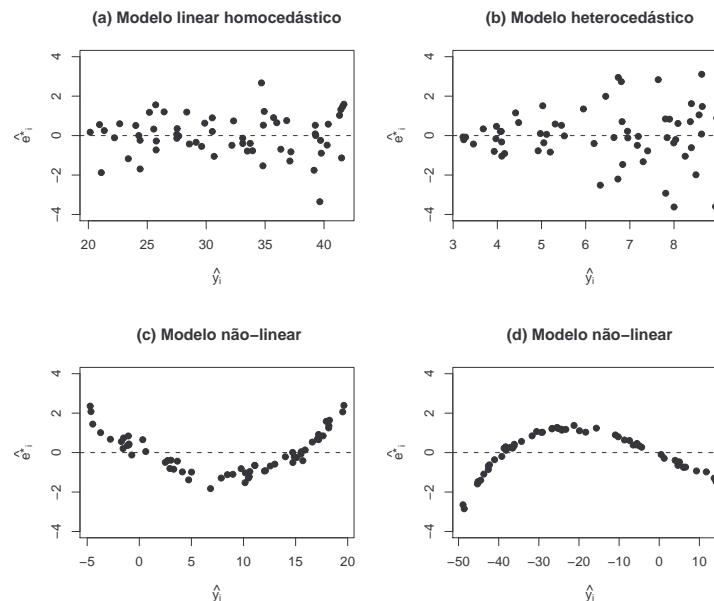
Um gráfico de \hat{e}_i^* *versus* \hat{y}_i (o valor predito da i -ésima observação) é útil para verificar a plausibilidade da suposição de **homocedasticidade**; quando esta suposição é verdadeira, espera-se que o comportamento de \hat{e}_i^* em torno do valor zero seja aleatório e que sua variabilidade tenha

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**.

⁹Pode-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} .

a mesma magnitude, independentemente 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 B.5.2 que corresponde aos resíduos associados a quatro ajustes fictícios. No gráfico B.5.2(a) não detectamos nenhum tipo de violação das suposições de homocedasticidade e de linearidade; no gráfico B.5.2(b), podemos notar que a variabilidade dos resíduos aumenta com a magnitude do valor ajustado, indicando que o modelo em questão é heterocedástico e nos gráficos B.5.2(c) e B.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 a variável explicativa.

Figura B.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 $(\widehat{e}_i^*)^2/(n-p) \sim \text{Beta}[1/2, (n-p-1)/2]$ ¹⁰, indicando que $|\widehat{e}_i^*| \leq \sqrt{n-p}$. Além disso, pode-se mostrar, que $\mathbb{E}[\widehat{e}_i^*] = 0$, $\mathbb{V}[\widehat{e}_i^*] = 1$ e que $\mathbb{Cov}[\widehat{e}_i^*, \widehat{e}_j^*] = -h_{ij}/[(1-h_{ii})(1-h_{jj})]^{1/2}$. Para detalhes e demonstração destas propriedades, veja Cook & Weisberg (1982), por exemplo.

Na definição dos resíduos studentizados, utilizamos o quadrado médio do resíduo (s^2) como estimador de σ^2 . Dada a dependência entre \widehat{e}_i e s^2 , a distribuição de \widehat{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

$$s_{(i)}^2 = s^2 \left(\frac{n-p-(\widehat{e}_i^*)^2}{n-p-1} \right).$$

Pode-se provar que \widehat{e}_i^* e $s_{(i)}^2$ são independentes, de forma que definimos os **resíduos studentizados externamente** por

$$t_i = \frac{\widehat{e}_i}{s_{(i)}\sqrt{1-h_{ii}}}, \quad i = 1, \dots, n.$$

Cook & Weisberg (1982) mostram que $t_i \sim t(n-p-1)$ e que

$$t_i = \widehat{e}_i^* \left(\frac{n-p-1}{n-p-(\widehat{e}_i^*)^2} \right)^{1/2},$$

indicando que t_i é uma transformação monótona crescente de \widehat{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 $|\widehat{e}_i^*|$, 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).

¹⁰ $\text{Beta}(a, b)$ denota a distribuição beta com parâmetros $a, b > 0$.

B.5.2 Análise da suposição de normalidade

Na teoria clássica, em geral, 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 **Q-Q** (**quantis-quantis**), em que dispomos os resíduos studentizados ordenados (quantis observados) no eixo das abscissas e os quantis obtidos da distribuição normal padrão (quantis teóricos) no eixo das ordenadas, 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, se o erro é heterocedástico, etc.

Esses gráficos podem ser obtidos por meio do seguinte procedimento

- i) Ajustar o modelo (B.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 obter os quantis amostrais $Q_{p_i} = \hat{e}_{(i)}^*$
- 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 $\mathcal{N}(0, 1)$
- iv) Construir o gráfico de dispersão $Q_{p_i} \times Z_{p_i}$.

Como em gráficos do tipo Q-Q, 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 (B.5.1), obtendo $\hat{\beta}$, s^2 e os resíduos studentizados $\hat{\mathbf{e}}^*$.

- ii) Construir o gráfico Q-Q correspondente.
 - iii)] Gerar um vetor \mathbf{z} com n elementos correspondentes a valores independentes de uma distribuição $\mathcal{N}(0, 1)$.
- iv) Obter um vetor de observações simuladas $\mathbf{y}_s = \mathbf{X}\hat{\boldsymbol{\beta}} + s\mathbf{z}$.
- v) Ajustar o modelo $\mathbf{y}_s = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}_s$ e obter os resíduos studentizados (simulados) $\hat{\mathbf{e}}_s^*$ e ordená-los.
- v) Repetir os itens (iii)-(v) m vezes gerando para cada p_i , um conjunto de m resíduos simulados.
- vi) Selecionar para cada p_i , o menor e o maior resíduo simulado e incluí-los no gráfico Q-Q construído no item (ii).
- vii) Construir o limite inferior (superior) do envelope simulado ligando os ponto do conjunto de menores (maiores) resíduos simulados.

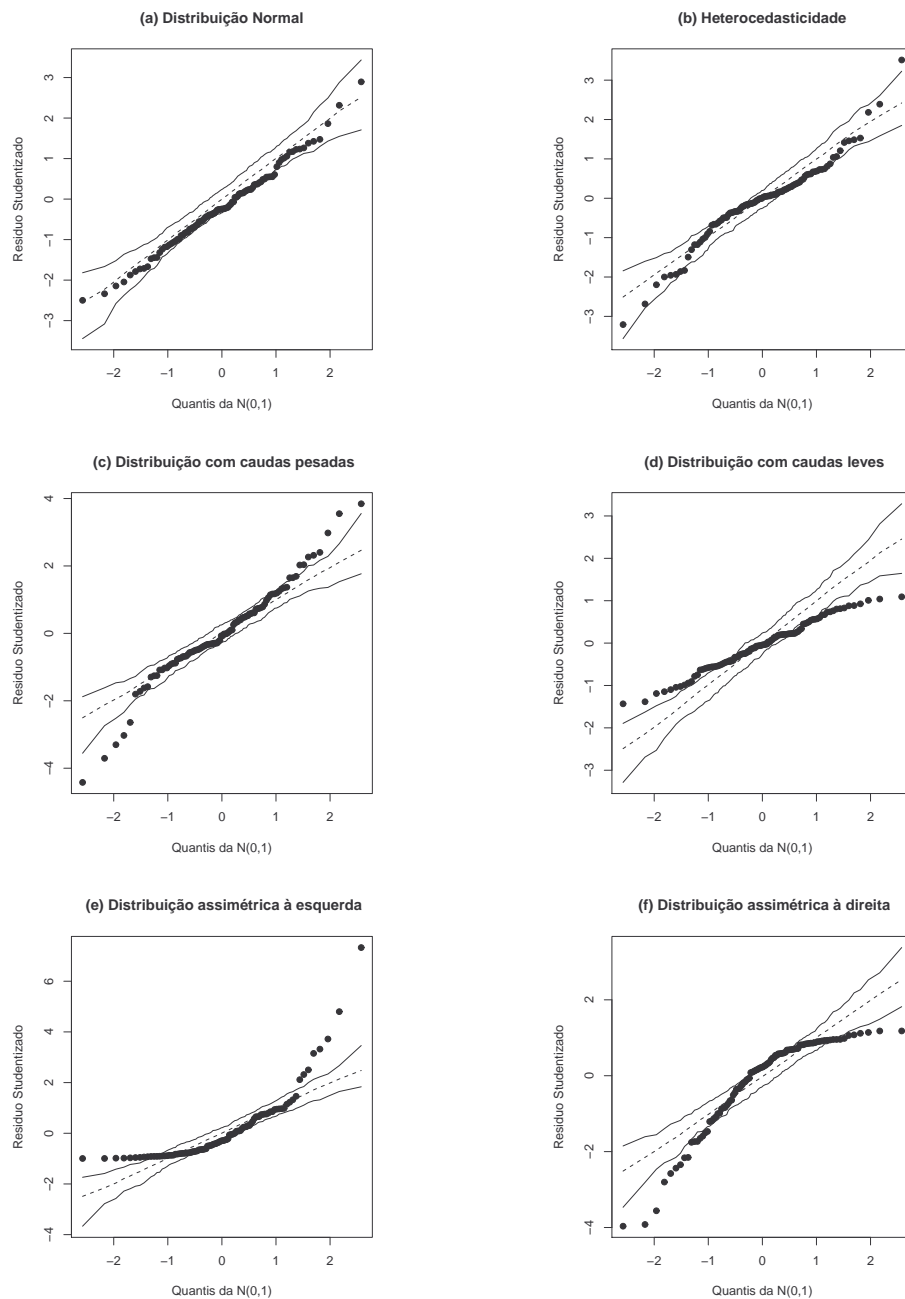
Se o modelo estiver correto, espera-se que todos os pontos observados no gráfico Q-Q fiquem contidos no envelope simulado em $[m/(m+1)] \times 100\%$ das vezes. Quando $m = 19$, isso corresponde a 95% das vezes.

Para maiores 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 B.5.3, apresentamos exemplos de gráficos Q-Q com diferentes padrões de afastamento das suposições de normalidade, homocedasticidade ou simetria. Por exemplo, no gráfico B.5.3(a), não há indícios contrários às hipóteses de normalidade e homocedasticidade; no gráfico B.5.3(b), os resíduos estão próximos dos limites das bandas de confiança, evidenciando que a distribuição dos erros é heterocedástica. O gráfico B.5.3(c) (B.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 Q-Q 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 B.5.3(e) (B.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 B.5.3: Envelopes simulados com coeficiente de confiança 95%



B.5.3 Análise de sensibilidade

A análise de sensibilidade visa avaliar o comportamento do ajuste de um modelo quando ele está 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 influência¹¹, na verossimilhança, no volume dos elipsóides 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 **grau 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 grau 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 que¹²

$$\begin{aligned} 0 &\leq h_{ii} \leq 1, \quad i = 1, \dots, n, \\ h_{ij} &\leq h_{ii}(1 - h_{ii}), \quad 1 \leq i < j \leq n. \end{aligned}$$

¹¹ 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 idéia de como a estatística é influenciada pelo acréscimo de uma única observação amostral.

¹² Também é possível mostrar que $n^{-1} \leq h_{ii} \leq 1$ se o modelo incluir intercepto.

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 grau de alavanca é p/n . Quando os 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 observação tem alto grau 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 (B.5.2), podemos concluir que, em geral, observações com alto grau de alavanca apresentam resíduos pequenos relativamente aos demais.

Para exemplificar, consideremos o seguinte modelo de regressão

$$y_i = \alpha + \beta x_i + e_i, \quad i = 1, \dots, n,$$

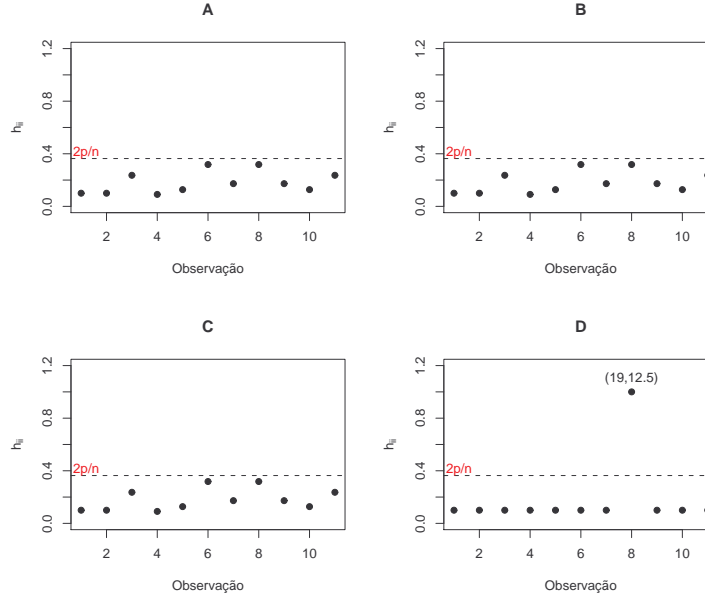
com $\mathbb{E}[e_i] = 0$, $\mathbb{V}[e_i] = \sigma^2$ e $\text{Cov}(e_i, e_j) = 0$, $i \neq j$. Nesse caso, o grau de alavanca da i -ésima observação é dado por

$$h_{ii} = \frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2},$$

e consequentemente, a i -ésima observação tem alto grau 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 grau de alavanca são aquelas associadas a covariáveis discrepantes em $\mathcal{C}(\mathbf{X})$ como observado por Chatterjee & Hadi (1988) e Wei, Hu & Fung (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 B.5.4, mostramos os gráficos de h_{ii} *versus* índice das observações para os quatro subconjuntos de dados apresentados na Tabela B.5.1. Nos subconjuntos A, B e C não existe indício de observações com alto grau 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 consequentemente, os graus de alavanca associados são iguais.

Figura B.5.4: Grau 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 a obter o seguinte modelo

$$\begin{aligned} \mathbf{y} &= \mathbf{X}\boldsymbol{\beta} + \gamma \mathbf{w} + \mathbf{e}^* \\ &= \mathbf{X}^* \boldsymbol{\beta}^* + \mathbf{e}^* \end{aligned} \quad (\text{B.5.4})$$

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 uma sequência 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 (B.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 . Maiores 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 (B.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

$$\begin{aligned} D_I &= \frac{[\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(I)}]^\top (\mathbf{X}^\top \mathbf{X}) [\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(I)}]}{ps^2} \\ &= \frac{[\hat{\mathbf{y}} - \hat{\mathbf{y}}_{(I)}]^\top [\hat{\mathbf{y}} - \hat{\mathbf{y}}_{(I)}]}{ps^2}, \end{aligned} \quad (\text{B.5.5})$$

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

$$\hat{\boldsymbol{\beta}}_{(i)} - \hat{\boldsymbol{\beta}} = -\frac{\hat{e}_i^*}{1 - h_{ii}}(\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{x}_i,$$

de tal forma que a distância de Cook correspondente é

$$\begin{aligned} D_i &= \frac{[\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(i)}]^\top (\mathbf{X}^\top \mathbf{X}) [\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(i)}]}{ps^2} \\ &= \frac{\hat{e}_i^{*2}}{p} \frac{h_{ii}}{(1 - h_{ii})^2}. \end{aligned} \quad (\text{B.5.6})$$

Dada a expressão (B.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 grau de alavanca) então D_i assume um valor “pequeno”, mesmo quando a i -ésima observação for altamente discrepante, indicando que a distância (B.5.6) pode não ser adequada nestas situações.

Outras propostas são sugeridas na literatura para contornar este fato. Por exemplo, Belsley, Kuh & Welsch (1980) sugerem a utilização de

$$\begin{aligned} \text{DFFITs}_i &= \frac{|\hat{e}_i^*|}{s_{(i)}(1 - h_{ii})^{1/2}} \left(\frac{h_{ii}}{1 - h_{ii}} \right)^{1/2} \\ &= |t_i| \left(\frac{h_{ii}}{1 - h_{ii}} \right)^{1/2}, \end{aligned} \quad (\text{B.5.7})$$

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,

$$C_i = \left(\frac{n - p}{p} \frac{h_{ii}}{1 - h_{ii}} \right)^{1/2} |t_i|. \quad (\text{B.5.8})$$

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, (B.5.6) é mais adequada para medir a influência das observações nos parâmetros de localização (β), enquanto que (B.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 grau de alavanca dificilmente influem na estimação de β e não comprometem o potencial uso de (B.5.5).

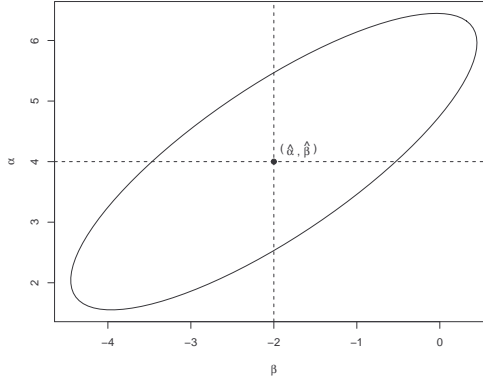
Gráficos das medidas de influência *versus* índices das observações são ferramentas úteis para a identificação observações influentes, *i.e.*, aquelas com valores da medida de influência “grandes” em relação aos demais.

Medidas de influência como a distância de Cook são baseadas na mudança do centro (parâmetros de localização) dos **elipsóides de confiança** para o vetor de parâmetros β , nomeadamente

$$\left\{ \beta \in \mathbb{R}^p; (\hat{\beta} - \beta)^\top (\mathbf{X}^\top \mathbf{X}) (\hat{\beta} - \beta) \leq ps^2 F_{p, (n-p)}(\alpha) \right\} \quad (\text{B.5.9})$$

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 B.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 maiores detalhes, sugerimos uma consulta a Andrews & Pregibon (1978), Belsley et al. (1980), Cook & Weisberg (1982) ou Chatterjee & Hadi (1988), por exemplo.

Figura B.5.5: Elipsóide de confiança com coeficiente de confiança 95%



O volume do elipsóide (B.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

$$AP_i = \frac{s_{(i)}^2 |\mathbf{X}_{(i)}^\top \mathbf{X}_{(i)}|}{s^2 |\mathbf{X}^\top \mathbf{X}|} = h_{ii} + (1 - h_{ii}) \frac{\hat{e}_i^{*2}}{n - p}. \quad (\text{B.5.10})$$

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

$$\begin{aligned} \text{COVRATIO}_i &= \frac{|\hat{V}[\hat{\beta}_{(i)}]|}{|\hat{V}[\hat{\beta}]|} = \frac{|s_{(i)}^2(\mathbf{X}_{(i)}^\top \mathbf{X}_{(i)})^{-1}|}{|s^2(\mathbf{X}^\top \mathbf{X})^{-1}|} \\ &= \left(\frac{n-p-\hat{e}_i^*}{n-p-1} \right)^p \frac{1}{1-h_{ii}}. \end{aligned} \quad (\text{B.5.11})$$

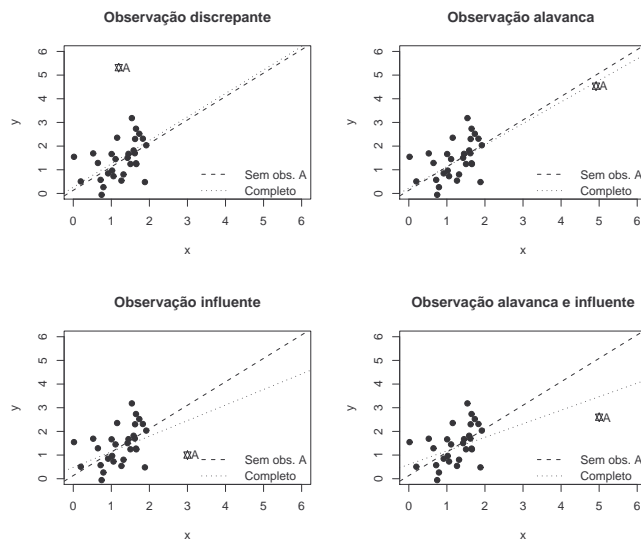
Se todas as observações tiverem o mesmo grau de influência, espera-se que $\text{COVRATIO}_i \approx 1$; afastamentos da unidade indicam que a observação correspondente é potencialmente influente com relação à matriz de covariâncias de $\hat{\beta}$. Belsley et al. (1980) sugerem utilizar o seguinte limiar para identificar observações influentes

$$|\text{COVRATIO}_i - 1| \geq 3p/n.$$

Esse limiar, assim como aqueles mencionados acima, tem cunho totalmente descritivo.

Na Figura B.5.6 mostramos um gráfico que ilustra a diferença entre observações discrepantes, alavanca e influentes.

Figura B.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).

B.5.4 Análise da suposição de correlação nula

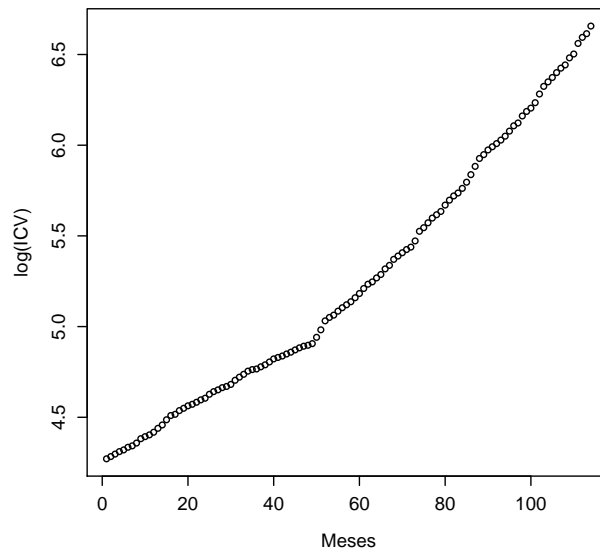
Em geral, a suposição de que os erros do modelo linear são não-correlacionados deve ser questionada com base no procedimento de coleta de dados. Como ilustração, consideramos dois exemplos nos quais essa característica justifica a dúvida. O primeiro exemplo é um caso simples dos problemas abordados pelas 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 B.5.1: Na Tabela B.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 B.5.7 Com base em argumentos de Economia, é razoável supor que o ICV num determinado mês seja correlacionado com aqueles obtidos em meses anteriores.

Tabela B.5.2: Índice de custo de vida para São Paulo (jan/70 a jul/79)

Observação	ICV	Observação	ICV	Observação	ICV
1	71.6	39	122	77	270
2	72.5	40	124	78	275
3	73.5	41	125	79	280
4	74.5	42	126	80	290
5	75.2	43	128	81	298
6	76.3	44	129	82	305
7	76.9	45	131	83	310
8	78.1	46	132	84	318
9	80	47	133	85	329
10	80.9	48	134	86	343
11	81.7	49	135	87	359
12	82.9	50	140	88	375
13	84.7	51	146	89	383
14	86.3	52	153	90	393
15	88.8	53	156	91	400
16	90.9	54	158	92	407
17	91.5	55	162	93	415
18	93.4	56	165	94	424
19	94.6	57	167	95	436
20	95.9	58	170	96	449
21	96.7	59	174	97	456
22	97.8	60	178	98	474
23	99.1	61	183	99	486
24	100	62	188	100	495
25	102	63	190	101	510
26	104	64	194	102	535
27	105	65	198	103	558
28	106	66	204	104	572
29	107	67	208	105	586
30	108	68	215	106	602
31	110	69	219	107	617
32	112	70	223	108	628
33	114	71	227	109	653
34	116	72	230	110	667
35	117	73	238	111	707
36	118	74	251	112	731
37	119	75	256	113	746
38	120	76	263	114	778

Figura B.5.7: Gráfico de dispersão para os dados do Exemplo B.5.1



Exemplo B.5.2: Na Tabela B.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 B.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

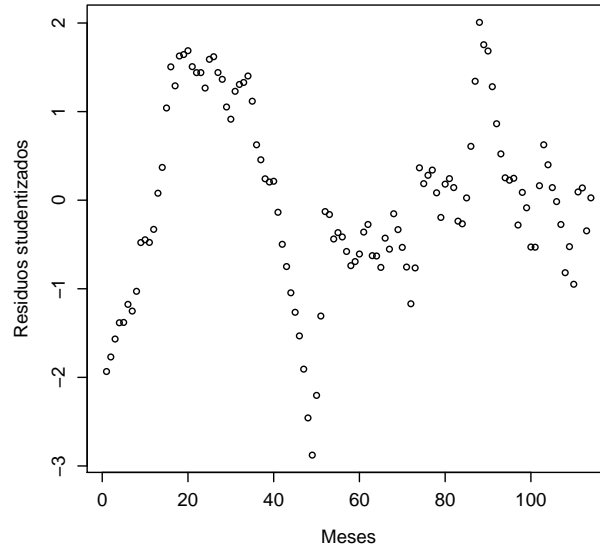
Tendo em vista o gráfico de dispersão apresentado na Figura B.5.7, uma primeira abordagem para a análise dos dados do Exemplo B.5.1 poderia envolver um modelo da forma

$$\ln(y_t) = \alpha + \beta t + \gamma t^2 + e_t, \quad t = 1, \dots, n \quad (\text{B.5.12})$$

em que y_t representa o incremento de vendas no instante t , α denota o valor esperado do incremento de vendas no tempo $t = 0$, β e γ representam os componentes linear e quadrático da curva que rege a variação temporal do logaritmo das vendas 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 B.5.8 mostra sequências de resíduos positivos seguidas de sequências de resíduos negativos, sugerindo uma possível correlação positiva entre eles (autocorrelação).

Figura B.5.8: Resíduos studentizados obtidos do ajuste do modelo (B.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 (B.5.12) com

$$e_t = \rho e_{t-1} + u_t, \quad t = 1, \dots, n \quad (\text{B.5.13})$$

em que $u_t \sim \mathcal{N}(0, \sigma^2)$, $t = 1, \dots, n$, 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:

$$D = \sum_{t=2}^n (\hat{e}_t - \hat{e}_{t-1})^2 / \sum_{t=1}^n \hat{e}_t^2, \quad (\text{B.5.14})$$

em que \hat{e}_t , $t = 1, \dots, n$ são os resíduos obtidos do ajuste do modelo

(B.5.12) por mínimos quadrados. Expandindo (B.5.14) obtemos

$$\begin{aligned} D &= \frac{\sum_{t=2}^n \hat{e}_t^2}{\sum_{t=1}^n \hat{e}_t^2} + \frac{\sum_{t=2}^n \hat{e}_{t-1}^2}{\sum_{t=1}^n \hat{e}_t^2} - 2 \frac{\sum_{t=2}^n \hat{e}_t \hat{e}_{t-1}}{\sum_{t=1}^n \hat{e}_t^2} \\ &\approx 2 - 2 \frac{\sum_{t=2}^n \hat{e}_t \hat{e}_{t-1}}{\sum_{t=1}^n \hat{e}_t^2}, \end{aligned} \quad (\text{B.5.15})$$

Se os resíduos não forem correlacionados, então $\sum_{t=2}^n \hat{e}_t \hat{e}_{t-1} \approx 0$ e consequentemente, $D \approx 2$; 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 (B.5.12) - (B.5.13) poderá ser ajustado pelo método MQG. Com esse intuito, notemos primeiramente que (B.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 (B.5.12) - (B.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 (B.2.6) para estimar $\boldsymbol{\beta}$.

O valor da estatística de Durbin-Watson para os dados do Exemplo B.5.1 sob o modelo (B.5.12) é $D = 0.1259$ ($p < 0.0001$) sugerindo um alto grau de autocorrelação dos resíduos.

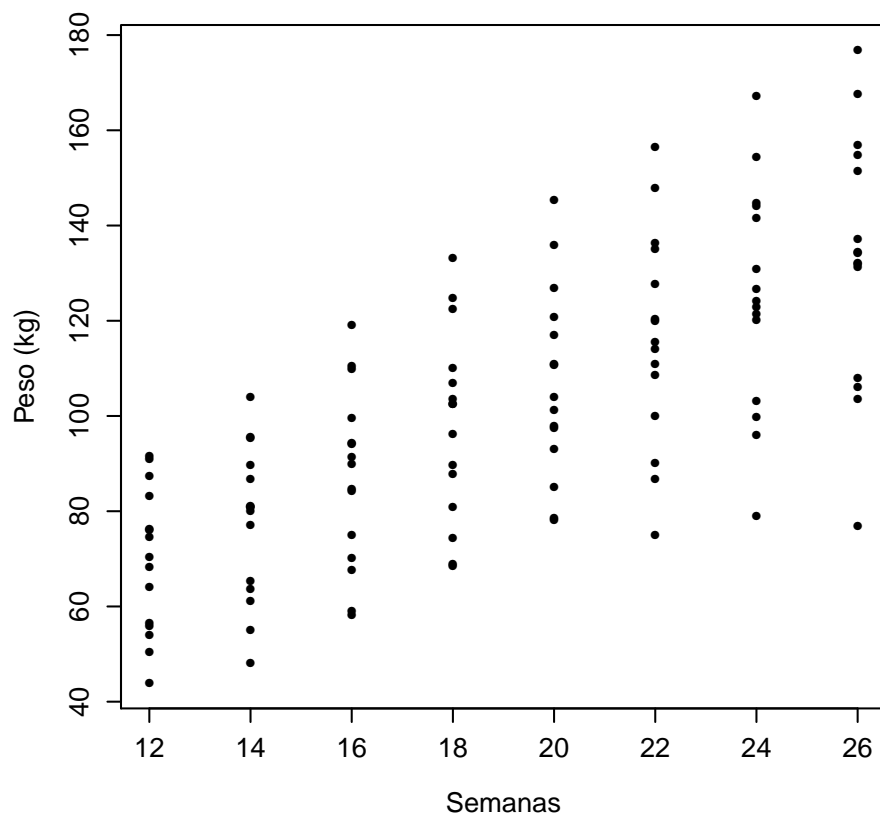
O gráfico de dispersão para os dados do Exemplo B.5.2 disposto na Figura B.5.9 sugere uma variação linear do peso médio dos animais ao longo do período estudado, para a qual podem-se usar modelos do tipo

$$y_{ij} = \alpha + \beta x_{ij} + e_{ij}, \quad i = 1, \dots, n, \quad j = 1, \dots, p_i, \quad (\text{B.5.16})$$

em que y_{ij} denota a j -ésima medida do peso do i -ésimo animal, x_{ij} indica o número de semanas 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 B.5.2, em todos os animais foram observados nos mesmos instantes, é tal que $n = 15$, $p_i = p = 8$ e $x_{ij} = x_j$.

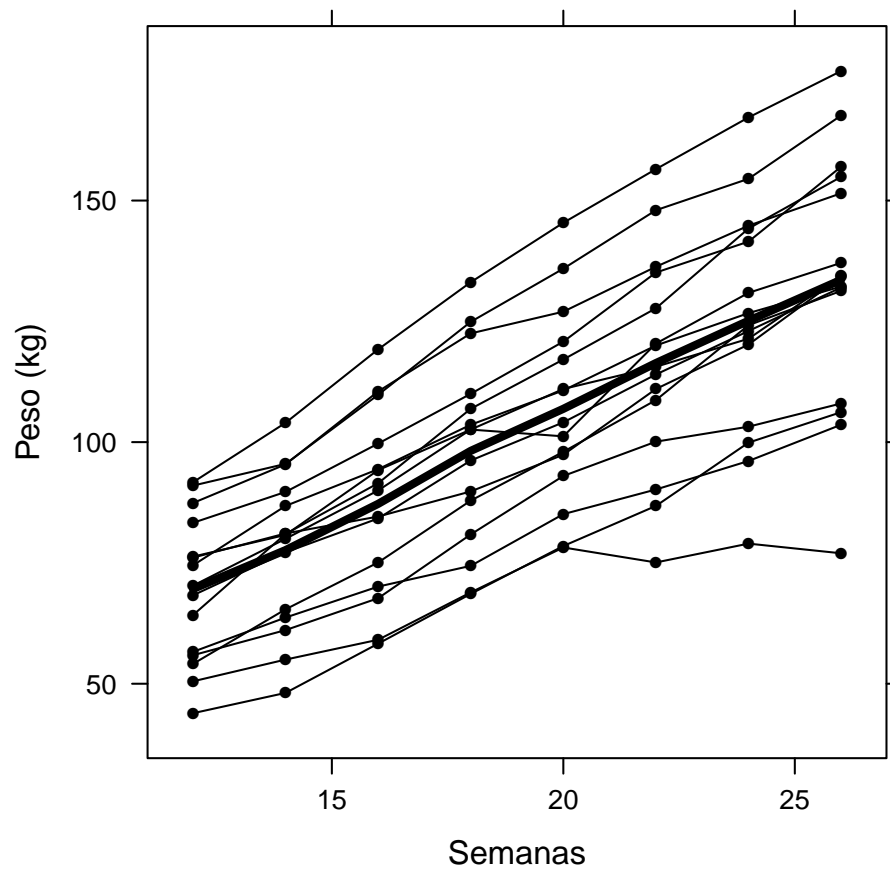
Sob a suposição de que os erros e_{ij} são não-correlacionados, estimativas (e erros padrões) dos parâmetros de localização do modelo (B.5.16) obtidas por mínimos quadrados são $\hat{\alpha} = 13.5$ (EP= 7.8), $\hat{\beta} = 4.7$ (EP= 0.4); a estimativa do desvio padrão é $\hat{\sigma}_e = 20.1$ e o 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 B.5.9: Gráfico de dispersão para os dados do Exemplo B.5.2



Tendo em vista que várias observações são realizadas em cada animal, podemos construir um gráfico de perfis como aquele apresentado na Figura B.5.10. Esse tipo de gráfico serve como ferramenta adicional para avaliação das suposições adotadas.

Figura B.5.10: Gráfico de perfis para os dados do Exemplo B.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 oito 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 (B.5.16), fazendo

$$e_{ij} = a_i + d_{ij} \quad (\text{B.5.17})$$

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. Essencialmente, o modelo (B.5.16)-(B.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

$$y_{ij}|a_i = (\alpha + a_i) + \beta x_{ij} + d_{ij} = \alpha_i + \beta x_{ij} + d_{ij} \quad (\text{B.5.18})$$

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 (B.5.16)-(B.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 $p_i = p$, 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 (B.5.18) na forma matricial

$$\mathbf{y} = \mathbf{W}\boldsymbol{\gamma} + \mathbf{e},$$

em que $\mathbf{W} = \mathbf{I}_n \otimes (\mathbf{1}_p \ \mathbf{x})$, com dimensão $np \times (n+1)$, é a matrix de variáveis explicativas do modelo condicional, $\mathbf{x} = [x_1, \dots, x_p]^\top$, $\mathbf{1}_a$ 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 $\mathbf{y}_i = [y_{i1}, \dots, y_{ip}]^\top$ e $\boldsymbol{\gamma} = [\alpha_1, \dots, \alpha_n, \beta]^\top$.

Um estimador consistente de σ^2 é

$$\hat{\sigma}^2 = [np - (n + 1)]^{-1} \mathbf{y}^\top [\mathbf{I}_{np} - \mathbf{W}(\mathbf{W}^\top \mathbf{W})^{-1} \mathbf{W}^\top] \mathbf{y}. \quad (\text{B.5.19})$$

Consideremos agora o modelo

$$\bar{y}_i = \alpha + a_i + \beta \bar{x} + \bar{d}_i \quad (\text{B.5.20})$$

em que, \bar{y}_i e \bar{x} representam respectivamente a média dos p valores da variável resposta e da covariável para a i -ésima unidade amostral e $\bar{d}_i = p^{-1} \sum_{j=1}^p d_{ij}$, $i = 1, \dots, n$. Esse modelo pode ser expresso como

$$\bar{y}_i = \mu + f_i$$

em que $\mu = \alpha + \beta \bar{x}$ e $f_i = a_i + \bar{d}_i$, $i = 1, \dots, n$ são erros aleatórios independentes com média nula e variância $\tau^2 = \sigma_a^2 + \sigma^2/p$. Sob esse modelo, um estimador consistente de τ^2 é

$$\hat{\tau}^2 = (n - 1)^{-1} \sum_{i=1}^n (\bar{y}_i - \bar{y})^2 \quad (\text{B.5.21})$$

com $\bar{y} = n^{-1} \sum_{i=1}^n \bar{y}_i$. Consequentemente, um estimador de σ_a^2 pode ser obtido de (B.5.19) e (B.5.21) por meio de

$$\hat{\sigma}_a^2 = \hat{\tau}^2 - \hat{\sigma}^2/p.$$

Na forma matricial, o modelo (B.5.16)-(B.5.17) pode ser escrito como

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$$

em que $\mathbf{X} = \mathbf{1}_n \otimes (\mathbf{1}_p \ \mathbf{x})$, $\boldsymbol{\beta} = (\alpha, \beta)^\top$ e \mathbf{e} é um vetor aleatório com média nula e matriz de covariâncias $\mathbb{V}(\mathbf{e}) = \mathbf{I}_n \otimes \boldsymbol{\Sigma}$ com $\boldsymbol{\Sigma} = \sigma^2 \mathbf{I}_p + \sigma_a^2 \mathbf{1}_p \mathbf{1}_p^\top$. Um estimador consistente, $\hat{\boldsymbol{\Sigma}}$ da matriz de covariâncias intraunidades amostrais $\boldsymbol{\Sigma}$ 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 (B.2.6) tendo em vista que, segundo a propriedade *vii*) da Seção A.1.6,

$$\boldsymbol{\Sigma}^{-1} = \frac{1}{\sigma^2} [\mathbf{I}_p - \frac{\sigma_a^2}{p\sigma_a^2 + \sigma^2} \mathbf{1}_p \mathbf{1}_p^\top].$$

Mais especificamente, pode-se mostrar que, neste caso,

$$\hat{\beta} = \left[\sum_{i=1}^n \mathbf{X}_i \hat{\Sigma}^{-1} \mathbf{X}_i^{\top} \right]^{-1} \sum_{i=1}^n \mathbf{X}_i \hat{\Sigma}^{-1} \mathbf{y}_i$$

em que $\mathbf{X}_i = [\mathbf{1}_p \ \mathbf{x}]$. Sob as condições de regularidade detalhadas na Seção B.2, concluímos que a distribuição aproximada de $\hat{\beta}$ é dada por (B.2.7).

Para os dados do Exemplo B.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 B.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.

B.6 Parametrização de modelos lineares

Consideremos um estudo em que se deseja comparar o “efeito” de a 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 p unidades amostrais escolhidas aleatoriamente sejam submetidas a cada um dos a tratamentos, um modelo bastante comum é

$$y_{ij} = \mu_i + e_{ij}, \quad i = 1, \dots, a, \quad j = 1, \dots, p, \quad (\text{B.6.1})$$

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 k como a diferença $\mu_i - \mu_k$.

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

$$y_{ij} = \mu + \alpha_i + e_{ij}, \quad i = 1, \dots, a, \quad j = 1, \dots, p, \quad (\text{B.6.2})$$

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**¹³ e tanto μ quanto α_i nem podem ser interpretados dessa forma nem são **estimáveis**¹⁴. 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

$$\sum_{i=1}^a \alpha_i = 0, \quad (\text{B.6.3})$$

que induz a chamada **parametrização de desvios de médias** e

$$\alpha_1 = 0, \quad (\text{B.6.4})$$

que induz a chamada **parametrização de cela de referência**. Definindo $\bar{y} = (ap)^{-1} \sum_{i=1}^a \sum_{j=1}^p y_{ij}$ e $\bar{y}_i = p^{-1} \sum_{j=1}^p y_{ij}$, e utilizando (B.6.3), obte-

¹³Um 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 $y_i \sim \mathcal{N}(\mu + \alpha_i, \sigma^2)$, $i = 1, 2$ 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 $\mathcal{N}_2[(6, 6)^\top, \sigma^2 \mathbf{I}_2]$. O leitor poderá consultar Bickel & Doksum (2001), entre outros, para detalhes.

¹⁴Uma 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 (B.6.2), nem μ nem α_i , $i = 1, \dots, a$ são funções estimáveis, embora tanto $\mu + \alpha_i$, $i = 1, \dots, a$ quanto $\alpha_i - \alpha_k$, $i \neq k$ o sejam. O leitor poderá consultar Searle (1971), entre outros, para detalhes.

mos

$$\mathbb{E}(\bar{y}) = (ap)^{-1} \sum_{i=1}^a \sum_{j=1}^p E(y_{ij}) = \mu + a^{-1} \sum_{i=1}^a \alpha_i = 0,$$

e conseqüentemente o termo μ pode ser interpretado como média geral (que essencialmente é uma média dos valores esperados da resposta associados aos a tratamentos). Além disso,

$$\mathbb{E}(\bar{y}_i) = p^{-1} \sum_{j=1}^p E(y_{ij}) = \mu + \alpha_i$$

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 (B.6.4), obtemos

$$\mathbb{E}(\bar{y}_1) = p^{-1} \sum_{j=1}^p E(y_{1j}) = \mu + \alpha_1 = \mu$$

e conseqüentemente, o termo μ pode ser interpretado como o valor esperado das observações submetidas ao tratamento 1. Além disso, para $k \neq 1$,

$$\mathbb{E}(\bar{y}_k) = p^{-1} \sum_{j=1}^p E(y_{kj}) = \mu + \alpha_k$$

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 p 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

$$y_{ijk} = \mu_{ijk} + e_{ijk}, \quad i = 1, \dots, a, \quad j = 1, \dots, b, \quad k = 1, \dots, p, \quad (\text{B.6.5})$$

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 $(\mu_{11} + \mu_{21})/2 - (\mu_{12} + \mu_{22})/2$. Essas definições de efeitos principais podem ser consideradas mesmo na presença de interação, desde que ela seja não-essencial. O leitor pode con-

sultar Lencina et al. (2005) e Lencina et al. (2008) para uma discussão sobre a consideração de efeitos principais em situações com interação não-essencial.

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

$$y_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + e_{ijk}, \quad i = 1, \dots, a, \quad j = 1, \dots, b, \quad k = 1, \dots, p. \quad (\text{B.6.6})$$

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 cela de referência são, respectivamente,

$$\sum_{i=1}^a \alpha_i = \sum_{j=1}^b \beta_j = \sum_{i=1}^a \alpha\beta_{ij} = \sum_{j=1}^b \alpha\beta_{ij} = 0 \quad (\text{B.6.7})$$

e

$$\alpha_1 = \beta_1 = \alpha\beta_{11} = \dots = \alpha\beta_{1b} = \alpha\beta_{21} = \dots = \alpha\beta_{a1} = 0 \quad (\text{B.6.8})$$

Sob as restrições (B.6.7), pode-se mostrar que

$$\mu = (ab)^{-1} \sum_{i=1}^a \sum_{j=1}^b \mu_{ij},$$

$$\alpha_i = b^{-1} \sum_{j=1}^b \mu_{ij} - \mu$$

$$\beta_j = a^{-1} \sum_{i=1}^a \mu_{ij} - \mu$$

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 (B.6.8), temos

$$\mu = \mu_{11},$$

$$\alpha_i = \mu_{ij} - \mu_{1j}, \quad i = 2, \dots, a,$$

$$\beta_j = \mu_{ij} - \mu_{i1}, \quad j = 2, \dots, b,$$

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 àquelas 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 àquelas 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 ‘a cela (i, j) e aquela esperada sob um modelo sem interação.

B.7 Regressão logística

B.8 Exercícios

B.8.1. Seja x_1, \dots, x_n uma amostra aleatória de uma variável X cuja distribuição é Normal.

- a) Mostre que a média e a variância amostrais são independentes.
- b) Enuncie um resultado semelhante no contexto de regressão linear simples.
- c) Qual a utilidade desse resultado?

B.8.2. Mostre que se $\mathbf{y} \sim N_p(\boldsymbol{\mu}, \mathbf{V})$ então $(\mathbf{y} - \boldsymbol{\mu})^\top \mathbf{V}^{-1}(\mathbf{y} - \boldsymbol{\mu}) \sim \chi_p^2$.

B.8.3. 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.

B.8.4. Obtenha os estimadores de máxima verossimilhança dos parâmetros do modelo de regressão

$$y_i = x_i^\beta e_i, \quad i = 1, \dots, n$$

em que os e_i representam erros aleatórios independentes log-normais com média $\exp(\sigma^2/2)$ e variância $\exp(\sigma^2)[\exp(\sigma^2) - 1]$.

- i) Obtenha a distribuição do estimador de β e construa um intervalo de confiança.
- ii) 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 $[\exp(\sigma^2) - 1] \exp(2\mu + \sigma^2)$.

B.8.5. Suponha que as variáveis x_i e y_i estejam relacionadas de acordo com o modelo de regressão linear simples

$$y_i = \alpha + \beta x_i + e_i, \quad i = 1, \dots, n$$

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 .

B.8.6. Para avaliar a associação entre a pressão arterial sistólica e idade, colheram-se os dados dispostos na tabela abaixo. Utilize esses dados para avaliar se essa associação pode ser representada por um modelo de regressão linear simples. Com essa finalidade,

- a) Especifique o modelo, interpretando os parâmetros;
- b) Construa um diagrama de dispersão (rotulando os eixos convenientemente);
- c) Estime os parâmetros e os correspondentes erros padrões;
- d) Construa intervalos de confiança (com coeficiente de confiança de 95%) para os parâmetros;
- e) Obtenha o valor-p correspondente ao teste da hipótese de que o coeficiente angular é nulo.

Paciente	Tensão sistólica (mm Hg)	Idade (anos)
1	114	16
2	134	20
3	116	18
4	139	34
5	110	23
6	150	38
7	152	41
8	138	39
9	142	46
10	145	45
11	156	47
12	159	47
13	142	50
14	156	52
15	164	45
16	185	60
17	162	64
18	176	60
19	175	69
20	180	68

B.8.7. 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	
particular	pública
8.6	5.8
8.6	7.6
7.8	8.0
6.5	6.2
7.2	7.6
6.6	6.5
5.6	5.6
5.5	5.7
8.2	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, $i = 1, \dots, 18$.

B.8.8. 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-operatoriamente 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 ultrassonográfico (cm^3)	Peso real (g)
656	630
692	745
588	690
799	890
766	825
800	960
693	835
602	570
737	705
921	955
923	990
945	725
816	840
584	640
642	740
970	945

B.8.9. 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 $1000m^2$.

B.8.10. O arquivo Bosco (1998), disponível na forma de uma planilha Excel em

www.ime.usp.br/simjmsinger

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 (C.R.S.M.N.A.D.I.) 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).

B.8.11. 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) na abscissa e o índice de placa bacteriana pós-escovação (Y) na ordenada.
- 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

B.8.12. 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 assumo 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.

B.8.13. Os dados do arquivo Singer&Andrade (1997), disponível em

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 sexo masculino (M).

- a) Construa um gráfico de dispersão para cada tipo de escova, tendo o índice de placa bacteriana pré-escovação (X) na abscissa e o índice de placa bacteriana pós-escovação (Y) na ordenada. 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 construído um tabela com valores esperados para os índices de placa pós-escovação com diferentes níveis de índices de placa pré-escovação.

B.8.14. Considere o modelo

$$y_{ij} = \alpha + \beta x_{ij} + a_i + e_{ij}, \quad i = 1, \dots, n, j = 1, \dots, p$$

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. Mostre que $Var(y_{ij}) = \sigma_a^2 + \sigma^2$, $Cov(y_{ij}, y_{ij'}) = \sigma_a^2$, $j \neq j'$ e que $Cov(y_{ij}, y_{i'j'}) = 0$, $i \neq i'$.

B.8.15. A tabela abaixo contém dados de capacidade instalada (*ton*), potência instalada (*1000 kW*) e área construída (*100 m²*) de 10 empresas de uma certa indústria.

Capacidade instalada (*ton*), potência instalada (*1000 kW*) e área construída (*100 m²*) de empresas de uma certa indústria

Capacidade de produção	4.5	5.9	4.2	5.2	8.1	9.7	10.7	11.9	12.0	12.3
Potência instalada	0.9	2.5	1.3	1.8	3.1	4.6	6.1	6.0	5.9	6.1
Área construída	7.1	10.4	7.2	8.2	8.5	11.9	12.1	12.5	12.0	11.3

Com o objetivo de estimar a capacidade instalada (Y) a partir das informações sobre potência instalada (X_1) e área construída (X_2),

- 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 ($\times 100 \text{ m}^2$).

B.8.16. Os dados do arquivo Braga (1998), disponível em

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 anaeróbio 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:

- a) Construir gráficos de dispersão convenientes.
- b) Interpretar os diferentes parâmetros do modelo.
- c) Estimar os parâmetros do modelo e apresentar os respectivos erros padrões.
- d) Avaliar a qualidade de ajuste do modelo por meio de gráficos de resíduos e gráficos QQ.
- e) Identificar possíveis valores discrepantes (*outliers*) e reajustar o modelo sem esses pontos.
- f) Comparar os ajustes dos modelos obtidos com e sem os *outliers*.
- g) 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.
- h) Reajustar o modelo com base nas conclusões do item (g) e avaliar o seu ajuste.
- i) Apresentar conclusões que evitem jargão técnico.

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, *Journal de Pediatria* **81**: 454–460.
- 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.
- 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.
- 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.
- 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.
- Brogan, D. R. & Kutner, M. H. (1980). Comparative analysis of pretest-posttest research designs, *The American Statistician* **34**: 229–232.
- Bruner, E., Domhof, S. & Langer, F. (2002). *Nonparametric analysis of longitudinal data in factorial experiments*, New York: Wiley.
- 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.

- 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.
- Cochran, W. G. & Cox, M. G. (1992). *Experimental designs*, 2 edn, New York: Wiley.
- 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. (2004). *Mixed models: theory and applications*, New York: Wiley.
- 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.
- Fuller, W. A. (1987). *Measurement error models*, New York: Wiley.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- Nelder, J. A. (1998). The great mixed-model muddle is alive and flourishing, alas!, *Food quality and preference* **9**: 157–159.
- 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.

- Potthoff, R. F. & Roy, S. N. (1964). A generalized multivariate analysis of variance model useful especially for growth curve problems, *Biometrika* **51**: 313–326.
- 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.
- Robinson, G. K. (1991). The estimation of random effects (with discussion), *Statistical Science* **6**: 15–51.
- 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.

- Schafer, W. D. (1981). Pretest-posttest designs: Letter to the editor, *The American Statistician* **35**: 179.
- 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. (1993). *Large sample methods in Statistics: an introduction with applications*, New York: Chapman and Hall.
- 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. & 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.
- Stram, D. O. & Lee, J. W. (1994). Variance components testing in the longitudinal mixed effects model, *Biometrics* **50**: 1171–1177.
- Tountenburg, H. (1982). *Prior information in linear models*, New York: Wiley.
- 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.

-
- Wishart, J. (1938). Growth rate determination in nutrition studies with the bacon pig and their analysis, *Biometrika* **30**: 16–24.
- Zeger, S. L., Liang, K. Y. & Albert, P. S. (1988). Models for longitudinal data: a generalized estimating equation approach, *Biometrics* **44**: 1049–1060.

Índice

- Área sob a curva, 8, 27
 - Endpoint*, 8
 - population-averaged*, 9
 - subject-specific*, 9
 - Algoritmo
 - scoring*, 49
 - EM, 49
 - Newton-Raphson, 49
 - Análise
 - de regressão, 5
 - de séries cronológicas, 158
 - de séries de tempo, 158
 - ANCOVA, 6, 68
 - ANOVA, 5, 6, 127
 - autocorrelação, 161
 - Coefficiente
 - de determinação, 141
 - de determinação ajustado, 141
 - Combinação linear, 89
 - Condição de Noether, 136
 - controvérsia dos modelos mistos, 61
 - Covariável
 - time dependent covariate*, 5
 - dependente do tempo, 5, 6
 - Critério de informação
 - Akaike, 43
 - Schwarz, 43
 - Curva
 - de crescimento, 10
 - de influência, 150
 - Desfecho, 8, 27
 - Desigualdade
 - de Cauchy-Schwarz, 100
 - triangular, 100
 - DFFITS, 155
 - Diagnóstico, 141
 - análise de sensibilidade, 141
 - avaliação do ajuste, 141
 - Distância de Cook, 154
 - Efeito
 - principal, 172
 - efeito de tratamento, 170
 - Elipsóide
 - de confiança, 150, 155
 - Equação
 - de estimação, 134
 - de Henderson, 40
 - normal, 134
 - Espaço vetorial, 98
 - complemento ortogonal, 101
 - dimensão, 98
 - espaço euclidiano, 100
 - espaço nulo, 99
 - subespaço vetorial, 98
 - Espaço-coluna, 89
 - Estatística
 - de Wald, 42
 - da razão de verossimilhanças, 42
 - de Durbin-Watson, 162
 - de Wald, 137, 138
 - Estimador
-

-
- BLUE, 39, 134
 - de máxima verossimilhança restrita, 38
 - de mínimos quadrados generalizados, 36
 - EBLUE, 41
 - linear não-enviesado de variância mínima, 134
 - Estudo
 - de coorte, 2
 - longitudinal, 2
 - tipo painel, 2
 - transversal, 2
 - Forma
 - bilinear, 104
 - linear, 103
 - quadrática, 104
 - Forma quadrática, 133
 - Função
 - escore, 49
 - gradiente, 107
 - Gráfico
 - de perfis de médias, 25
 - da variável adicionada, 153
 - de perfis, 57, 77
 - de perfis individuais, 24
 - de regressão parcial, 153
 - Q-Q, 147
 - Grau de alavanca, 150
 - heterocedasticidade, 26
 - Independência linear, 89
 - Influência
 - global, 150
 - local, 150
 - Interação, 131, 172
 - essencial, 58, 72
 - não-essencial, 172
 - Máxima verossimilhança restrita, 139
 - Média geral, 170
 - Método
 - da união-intersecção, 46
 - de mínimos quadrados generalizados, 137
 - de mínimos quadrados ordinários, 133
 - de mínimos quadrados ponderados, 138
 - bayesiano, 36
 - da máxima verossimilhança, 36
 - da máxima verossimilhança restrita, 36
 - de ajuste por covariância, 47
 - de mínimos quadrados generalizados, 167
 - MANOVA, 45, 60
 - Matriz, 83
 - autovalor, 102
 - autovetor, 103
 - base, 102
 - base ortonormal, 102
 - chapéu (*hat matrix*), 144
 - circular, 48
 - cofator, 90
 - de covariâncias amostrais, 27
 - de informação empírica, 49
 - de posto completo, 90
 - de predição, 144
 - de projeção, 144
 - definida não negativa, 104
 - definida positiva, 104
 - definida semi-positiva, 104
 - determinante, 90
 - diagonal, 87
-

-
- diagonal em blocos, 94
 - esférica, 48
 - hessiana, 49, 108
 - idempotente, 88
 - identidade, 87
 - inversa, 91
 - inversa generalizada, 93, 134
 - jacobiana, 109
 - menor, 90
 - menor principal, 90
 - multiplicação por escalar, 84
 - não-singular, 90, 91
 - norma, 101
 - operador vec, 96
 - operador vech, 97
 - particionada, 89
 - posto, 90
 - produto de, 85
 - produto de Kronecker, 94
 - produto direto, 94
 - produto tensorial, 94
 - quadrada, 86
 - raiz característica, 102
 - simétrica, 87
 - soma de, 85
 - soma direta, 94
 - submatriz, 88
 - traço, 93
 - transposta, 86
 - triangular, 88
 - triangular inferior, 88
 - triangular superior, 87
 - vetor caraterístico, 103
 - vetorização, 96
 - Medida resumo, 8, 27
 - Modelo
 - split-plot*, 62
 - de independência condicional, 31
 - condicional, 10
 - de análise de perfis, 45
 - de crescimento, 10
 - de curvas de crescimento, 47
 - de erros de medida, 123
 - de parâmetros irrestritos, 61
 - de parâmetros restritos, 61
 - de perfis, 10
 - de regressão linear múltipla, 123
 - de regressão linear simples, 123
 - de regressão segmentada, 131
 - em dois estágios, 30
 - heterocedástico, 137
 - homocedástico, 137
 - incondicional, 10
 - individual, 9
 - inidentificável, 170
 - linear assimétrico, 11
 - linear elíptico, 11
 - linear generalizado, 11
 - mecanístico, 11
 - misto, 61, 167
 - populacional médios, 9
 - Movimento browniano, 11
 - Norma
 - de Frobenius, 100
 - Observação
 - discrepante, 142, 145
 - omissa, 6
 - Operador diagonal, 87
 - Parâmetro
 - de localização, 30
 - de não-centralidade, 105
 - estimável, 170
 - Parametrização
 - de cela de referência, 170
 - de desvios de médias, 170
 - de médias de celas, 169
-

-
- Perfil
 response profile, 5
 de resposta, 5
 Pico, 27
 peak, 27
 Planejamento
 linked, 6
 balanceado com relação ao tempo, 6
 6
 com intercâmbio, 2
 crossover, 2
 encadeado, 6
 split-plot, 2
 transversal misto, 6
 Preditor
 BLUP, 39
 EBLUP, 41
 Processo de Ornstein-Ühlenbeck, 11
 Regressão
 com erros nas variáveis, 123
 Reparametrização
 de desvios médios, 61
 Resíduo
 envelope simulado, 147
 ordinário, 144
 predito, 146
 recursivo, 146
 studentizado, 144
 studentizado externamente, 146
 Restrição
 de identificabilidade, 61, 170
 Séries de tempo, 2
 Soma de quadrados
 devida à regressão, 140
 residual, 140
 total, 140
 Teorema
- de Hájek-Šidak, 137
 de Sverdrup, 138
 limite central, 136
 Transformação
 linear, 99
 Trilhamento, 5
 tracking, 5
 trilhamento, 26
 Variável
 latente, 30
 Verossimilhança
 residual, 37
 restrita, 37
 Vetor, 84
 distância euclidiana, 100
 norma, 100
 ortogonais, 101
 ortonormal, 101
 produto interno, 100
 projeção ortogonal, 102
 unitário, 100