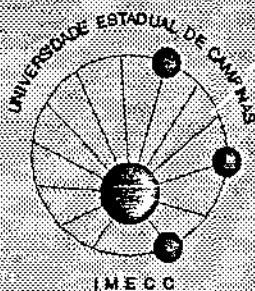

Um Curso em Modelos Lineares

ANTONIO RICARDO AMARANTE

Orientador: JOSÉ FERREIRA DE CARVALHO



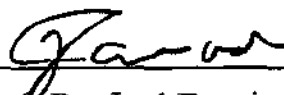
DISSERTAÇÃO APRESENTADA AO DEPARTAMENTO
DE ESTATÍSTICA DO INSTITUTO DE MATEMÁTICA,
ESTATÍSTICA E CIÊNCIA DA COMPUTAÇÃO DA
UNICAMP, PARA OBTENÇÃO DO TÍTULO DE
MESTRE EM ESTATÍSTICA

UNICAMP
BIBLIOTECA CENTRAL

UM CURSO EM MODELOS LINEARES

Este exemplar corresponde a redação final da tese devidamente corrigida e defendida pelo Sr. Antonio Ricardo Amarante e aprovada pela Comissão Julgadora.

Campinas, 1 de 6 de 1992.



Prof. Dr. José Ferreira de Carvalho
Orientador

Dissertação apresentada ao Instituto de Matemática, Estatística e Ciência da Computação, UNICAMP, como requisito parcial para obtenção do Título de Mestre em Estatística.

SUMÁRIO

<i>Introdução</i>	1
-------------------	---

1

<i>Alguns Conceitos Úteis</i>	7
1.1 Introdução	7
1.2 Distância	7
1.3 Operações Linha	8
1.4 Projeções	9
1.5 Matrizes de Projeção	11
1.6 Autovalores e Autovetores	13
1.7 Formas Quadráticas	14
1.8 Elipsóides	16

2

<i>Modelos Classificatórios</i>	19
2.1 Introdução	19
2.2 Fatores de Classificação	20
2.3 Representação Esquemática das Estruturas de Classificação	27
2.4 O Papel da Aleatorização	34
2.5 Os Modelos	37
2.6 A Matriz do Modelo	42

3

<i>O Ajuste de Mínimos Quadrados</i>	47
3.1 Introdução	47
3.2 Aproximações	48
3.3 O Método dos Mínimos Quadrados	52
3.4 Graus de Liberdade	64
3.5 As Equações Normais Reduzidas	65
3.6 O Modelo Linear a k-Partes	67
3.7 Redução na Soma de Quadrados: A Notação R(.)	70
3.8 Análise de Variância	75
3.9 Exemplo	80

4

<i>Estimabilidade</i>	87
4.1 Introdução	87
4.2 Funções Identificáveis	88
4.3 Funções Estimáveis	90
4.4 Exemplo	96

5

<i>Mínimos Quadrados em Modelo com Restrições nos Parâmetros</i>	99
5.1 Introdução	99
5.2 O Modelo Restrito	100
5.3 Restrições Envolvendo Funções Não Estimáveis	114
5.4 Restrições Envolvendo Funções Estimáveis	116
5.5 Exemplo	120

6

<i>Mínimos Quadrados Ponderados</i>	125
6.1 Introdução	125
6.2 Projeção Não Ortogonal (Oblíqua)	126
6.3 O Uso dos Mínimos Quadrados Ponderados	129

7

<i>Resolução Numérica das Equações Normais</i>	133
7.1 Introdução	133
7.2 O Processo de Eliminação de Gauss-Jordan	134
7.3 O Operador SWEEP(A,k)	140
7.4 O Caso de X com Posto Incompleto	150

8

<i>Forma Geral das Funções Estimáveis</i>	157
8.1 Introdução	157
8.2 Os Conjuntos Geradores das Funções Estimáveis	158
8.3 Exemplo	166

9

<i>Funções Estimáveis Tipos I, II, III e IV</i>	169
9.1 Introdução	169
9.2 Dados Desbalanceados	170
9.3 Hipóteses de Interesse	174

9.4 Efeitos Contidos em Outros	175
9.5 Funções Estimáveis Tipo I	178
9.6 Funções Estimáveis Tipo II	187
9.7 Funções Estimáveis Tipo III	191
9.8 Funções Estimáveis Tipo IV	195
9.9 Comparação das Somas de Quadrados	201
9.10 Exemplo	203

10

<i>Distribuição de Formas Quadráticas</i>	215
10.1 Introdução	215
10.2 Distribuição de uma Forma Quadrática	215
10.3 Independência de Formas Quadráticas	219
10.4 Aplicação aos Modelos Lineares	226
10.5 Análise de Variância no Modelo de Gauss-Markov Normal	232
10.6 O Caso Geral	234

11

<i>Testes Estatísticos</i>	237
11.1 Introdução	237
11.2 Testes de Significância e Nível de Significância	238
11.3 Hipóteses Testáveis	244
11.4 Testando Hipóteses Testáveis	248
11.5 Hipóteses Não Testáveis	249
11.6 Exemplo 1	250
11.7 Intervalos de Consonância	253
11.8 Intervalo de Consonância para l/β	258
11.9 Intervalo de Consonância para σ^2	260
11.10 Regiões de Consonância	262
11.11 Comparações Múltiplas	264
11.11.1 Método de Tukey	266
11.11.2 Método de Scheffé	272
11.12 Exemplo 2	277

12

<i>Análise de Covariância</i>	281
12.1 Introdução	281
12.2 Objetivos	283
12.3 O Modelo	284
12.4 Aplicações da Análise de Covariância	286
12.5 Exemplo	290

13

<i>Exame dos Resíduos</i>	295
13.1 Introdução	295
13.2 Métodos Gráficos para o Exame dos Resíduos	297
13.2.1. Diagrama de pontos	297
13.2.2. Sequência temporal	300
13.2.3. Resíduos \times valores preditos	301
13.2.4. Resíduos \times possíveis variáveis de interesse	304
13.2.5. Outros	304
13.3 Exemplo	305
 <i>Referências Bibliográficas</i>	 311
 <i>Índice Remissivo</i>	 315

Introdução

Os Modelos Lineares tratam do ajuste de funções lineares nos parâmetros pelo critério dos mínimos quadrados. Sua aplicação em Estatística é muito ampla como, por exemplo, num processo de amostragem, para estimar a média μ de uma população, assumimos que cada observação feita, y_i , é igual à média mais um erro ou $y_i = \mu + e_i$. E, assim, cada observação foi modelada através de um modelo linear. Outro exemplo é o emprego da técnica de análise de variância para comparação de tratamentos em experimentos.

Essencialmente, o ajuste de um modelo é uma questão de aproximação numérica. Nesse sentido, é desnecessário se colocar, a princípio, imposições ou restrições a componentes do modelo. Pretende-se, assim, desenvolver a teoria dos Modelos Lineares de modo gradativo, colocando imposições apenas onde necessário. Para tal, usaremos a idéia dos projetores, de modo que a visão geométrica predominará, acarretando importante simplificação algébrica, como se verá.

A presente Dissertação de Mestrado tem como objetivos fundamentais a) apresentar a teoria dos Modelos Lineares sob uma óptica basicamente geométrica e b) auxiliar no uso do sistema SAS*, que ocupa papel de destaque na Estatística aplicada,

construindo uma ponte entre a teoria e a prática. Por exemplo, as Somas de Quadrados tipos I, II, III e IV, popularizadas pelo sistema SAS*, constituem parte importante da Dissertação. Isso em razão de entendermos a sua utilidade prática e a falta de informações mais detalhadas sobre o tema, mesmo no material de referência que acompanha o sistema SAS*.

Além disso, para testar hipóteses e construir intervalos estatísticos, usaremos os testes de significância e intervalos de consonância no lugar dos clássicos testes de hipóteses e intervalos de confiança. Isso corresponde a uma mudança de filosofia em relação a outros textos sob Modelos Lineares no que concerne ao modo de encarar tais temas, já que os cálculos matemáticos envolvidos não se alteram.

Este trabalho não deve se encerrar em si mesmo: ele é o primeiro passo para a produção de um texto didático sobre Modelos Lineares.

Ao longo do texto, estaremos estudando e resolvendo um sistema de equações onde o número de incógnitas (parâmetros) é menor que o número de equações. Portanto, em geral, um sistema inconsistente. Esse fato não deve ser perdido de vista pelo leitor já que ele norteia todo o desenrolar do trabalho.

No Capítulo 1, apresentaremos resultados sobre vetores, matrizes e Álgebra Linear que serão usados no restante da Dissertação. O objetivo deste capítulo é recordar tópicos que estarão envolvidos no desenvolvimento da teoria nos capítulos seguintes. No Capítulo 2, estudaremos as estruturas de classificação. Definiremos, também, os Modelos Classificatórios, que serão o alvo principal do trabalho.

A partir do Capítulo 3, iniciaremos o estudo do sistema inconsistente referido

acima: suas características, suas soluções através do método dos Mínimos Quadrados e o surgimento das Equações Normais. No mesmo capítulo, veremos o significado dos graus de liberdade, a notação $R(\cdot)$ e a técnica de Análise de Variância.

Estudaremos o conceito de Estimabilidade e algumas características das funções estimáveis no Capítulo 4. Impor restrições, estimáveis ou não estimáveis, aos modelos, leva ao surgimento dos Modelos Restritos, que veremos no Capítulo 5.

Até aqui, as projeções associadas às Equações Normais foram ortogonais. Caso tais projeções sejam não ortogonais, necessitaremos desenvolver o método dos Mínimos Quadrados Ponderados para resolver o sistema. Isso será feito no Capítulo 6.

O Capítulo 7 trata da resolução numérica das Equações Normais através do operador $\text{SWEEP}(A,k)$. O $\text{SWEEP}(A,k)$, modificado por questões de implementação computacional, é o método usado pelo sistema SAS*. Ele fornece subprodutos de interesse no ajuste de um modelo linear.

Nos Capítulos 8 e 9, trataremos do estudo das Somas de Quadrados tipos I, II, III e IV que serão usadas nos Testes de Significância. Seu surgimento, construção e uso são vistos em detalhe. A importância do tema já foi enfatizada anteriormente.

Com exceção da necessidade da existência de Esperança e Variância de variáveis aleatórias para o conceito de Estimabilidade, nada se exigiu em termos de distribuição probabilística. Neste ponto, introduziremos as condições de Gauss-Markov aos modelos. No Capítulo 10 estudaremos as distribuições de estatísticas que serão empregadas nos Testes Estatísticos.

Os Testes de Significância e os Intervalos e Regiões de Consonância, cuja presença aqui já foi mencionada, serão vistos no Capítulo 11.

Finalmente, os dois últimos capítulos são uma complementação do estudo anterior: a Análise de Covariância, no Capítulo 12, e o Exame dos Resíduos, no Capítulo 13. A Análise de Covariância não mais vista como um tema à parte dentro dos Modelos Lineares, mas como uma aplicação de grande importância. Conforme enfatizado no capítulo, sua separação tornou-se irrelevante, do ponto de vista computacional, na era dos computadores.

Capítulo 1

Alguns Conceitos Úteis

1.1 INTRODUÇÃO

O objetivo do presente capítulo é resumir alguns conceitos que serão úteis no decorrer deste trabalho. Como a teoria dos Modelos Lineares será vista sob um ponto de vista essencialmente geométrico, veremos aqui alguns resultados importantes para esse desenvolvimento, como a noção dos projetores. Não nos deteremos em conceitos básicos de Álgebra Linear. Para tal, o leitor deverá consultar textos específicos sobre o assunto, citados na bibliografia.

1.2 DISTÂNCIA

Definição 1. Em um espaço vetorial X , para todo elemento x , y e z pertencentes a X , uma função real $d(.,.)$, tal que

1. $d(x,y) \geq 0$;

2. $d(x,y) = 0$ se, e somente se, $x = y$;

3. $d(x,y) = d(y,x)$ e

4. $d(x,y) \leq d(x,z) + d(z,y)$

é definida como a *distância* entre os elementos x e y .

Poderíamos tomar como a função $d(\cdot, \cdot)$, por exemplo,

1. $d(x,y) = \sqrt{(x-y)'(x-y)}$,

2. $d(x,y) = |x_1 - y_1| + \dots + |x_n - y_n|$ ou

3. $d(x,y) = \max_{1 \leq i \leq n} \{|x_i - y_i|\}$,

A função 1 é facilmente tratável do ponto de vista matemático, está envolvida no método dos mínimos quadrados - de largo uso prático - e será a medida de distância usada neste trabalho.

1.3 OPERAÇÕES LINHA

Algumas operações realizadas com os vetores linha de uma matriz A e que serão úteis são:

1. troca de uma linha por outra,

2. multiplicação de uma linha por uma constante e
3. adição de um múltiplo de uma linha à outra linha.

Cada uma dessas operações é equivalente à premultiplicação de A por uma matriz não singular e que não altera o posto de A . Por exemplo, para $A = \begin{bmatrix} 4 & 2 & 2 \\ 2 & 2 & 0 \\ 2 & 0 & 2 \end{bmatrix}$, com

$p(A) = 2$, sejam as matrizes $B = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, $C = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ e $D = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ de posto

completo. $F = BA = \begin{bmatrix} 2 & 2 & 0 \\ 4 & 2 & 2 \\ 2 & 0 & 2 \end{bmatrix}$, $G = CA = \begin{bmatrix} 4 & 2 & 2 \\ 6 & 6 & 0 \\ 2 & 0 & 2 \end{bmatrix}$ e $H = DA = \begin{bmatrix} 10 & 6 & 4 \\ 2 & 2 & 0 \\ 2 & 0 & 2 \end{bmatrix}$ são os

resultados da aplicação das operações linha 1,2 e 3 acima. É facilmente verificado que as três matrizes resultantes são de posto 2.

1.4 PROJEÇÕES

Suponhamos um ponto y em um espaço n -dimensional. Queremos encontrar sua distância a uma reta, digamos a reta na direção do vetor a , conforme mostra a Figura 1 para o caso de $n=3$. Estamos procurando o ponto q sobre essa reta que está mais próximo de y . A linha conectando y a q é perpendicular ao vetor a e podemos calcular a distância entre eles. A situação é a mesma quando, ao invés de uma reta na direção

de a , temos um plano ou, de modo mais geral, um subespaço S de \mathbb{R}^n . O problema é encontrar o ponto q naquele subespaço que está mais próximo de y . Esse ponto é a projeção de y no subespaço e será encontrado traçando uma perpendicular de y a S . q será o ponto de S onde a perpendicular a S encontra o subespaço.

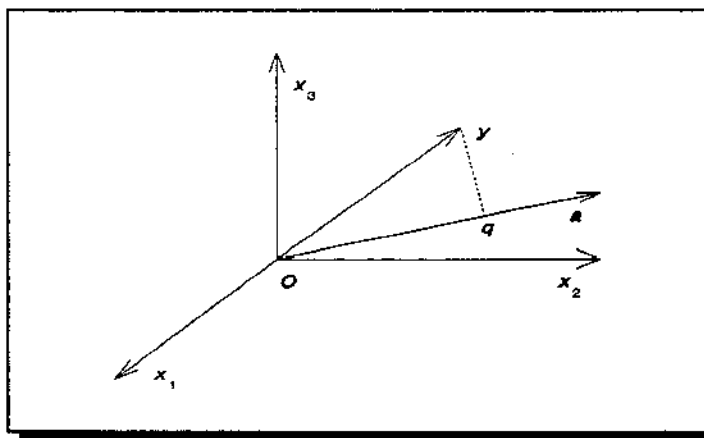


Figura 1. Projeção em um espaço tridimensional

Aqui, algumas questões surgem.

1. Este problema ocorre em aplicações práticas?
2. Existe uma expressão que defina o ponto q ?
3. Existe um modo computacionalmente estável de calcular q ?

As respostas às duas primeiras perguntas é sim. Nossa questão é o problema da *solução de mínimos quadrados de um sistema superdeterminado*. O vetor y representa os dados fornecidos por experimentos ou questionamentos e, conforme

veremos, surgem equações que não tem solução. O método dos mínimos quadrados seleciona o ponto q que está mais próximo do ponto y e, assim, podemos dizer que ele é a melhor escolha possível se queremos representar y por um ponto pertencente a S .

A segunda questão, encontrar q , é fácil quando o subespaço, ou é uma reta, ou está representado por sua base. No entanto, se S é descrito por um conjunto de vetores que o geram, mas que são linearmente dependentes, a questão leva a problemas adicionais. Nessa situação, o ponto q continua unicamente determinado por ser o mais próximo de y . Porém, ele pode ser expresso de muitas maneiras.

Com relação à terceira questão, a resposta não é simples. Se o subespaço S é especificado por um conjunto de vetores que estão muito próximos, é virtualmente impossível decidir se eles são, ou não, independentes. Como consequência, o ponto q será representado por combinações lineares cujos coeficientes são numericamente instáveis.

1.5 MATRIZES DE PROJEÇÃO

Consideremos o sistema $Az = c$, onde A é uma matriz $n \times p$, z , um vetor $p \times 1$ de incógnitas e c , um vetor de n observações. O número n de observações é maior que o número p de incógnitas e, assim, é de se esperar que o sistema não tenha solução. Provavelmente, não existirá um valor de z que se ajuste perfeitamente às observações c . Ou, de outra forma, provavelmente o vetor c não será uma combinação linear das colunas de A .

O problema, então, é escolher um valor \tilde{z} que torne o erro que estamos cometendo o mínimo possível. Esta minimização será feita no sentido dos mínimos quadrados. O erro, $r = \|A\tilde{z} - c\|$, é a distância de c ao ponto $A\tilde{z}$ no espaço coluna de A . Assim, a solução de mínimos quadrados corresponde a encontrar o ponto $q = A\tilde{z}$ que está mais próximo de c que qualquer outro ponto de $\mathcal{E}(A)$.

A projeção de c no espaço coluna de A , $\mathcal{E}(A)$, será $q = A\tilde{z} = P_A c$. A matriz P_A é a *matriz de projeção ortogonal* ou *projetor ortogonal* no espaço coluna de A . O índice A está sendo usado para enfatizar que a projeção é feita sobre esse subespaço. Quando não houver necessidade de se especificar, ou não houver dúvida sobre qual espaço a projeção está sendo feita, o índice será omitido. A projeção de c em $\mathcal{E}(A)$ fornece $P_A c$, nesse subespaço, e $c - P_A c = (I - P_A)c$ em $\mathcal{E}(A)^\perp$. Assim, $I - P_A$ também é um projetor: ele projeta c no ortocomplemento de A . O projetor ortogonal tem as seguintes propriedades:

P₁. $P^2 = P$, ou seja, ele é *idempotente*.

P₂. $P' = P$, ou seja, ele é *simétrico*.

Podemos notar que para qualquer vetor c , o vetor Pc pertence ao subespaço sobre o qual c está sendo projetado. Assim, se o projetarmos novamente sobre o mesmo subespaço, fornecendo $P(Pc) = P^2c$, nada é alterado e $Pc = P^2c$ para todo vetor c . Então, $P = P^2$.

Para verificar a simetria, consideremos quaisquer dois vetores c e d . Se c é projetado, como acima, fornecendo Pc e d é projetado no ortocomplemento de A , fornecendo $(I-P)d$, as duas projeções são perpendiculares entre si e

$$(Pc)'(I-P)d = c'P'(I-P)d = 0$$

Como isso vale para quaisquer c e d , concluímos que

$$P'(I-P) = \emptyset \rightarrow P' = P'P \rightarrow P = (P'P)' = P'P$$

de modo que $P' = P$.

Pelo mesmo raciocínio, é facilmente verificado que $I-P$ é simétrico e idempotente.

1.6 AUTOVALORES E AUTOVETORES

Definição 2. Um *autovalor* de uma matriz A , $p \times p$, é um escalar λ , tal que $Ax = \lambda x$, para algum vetor $x \neq \emptyset$. O vetor x é chamado um *autovetor* da matriz A e cada autovetor está associado a um autovalor.

Então, se λ é um autovalor de A , $Ax - \lambda x = \emptyset$ ou $(A - \lambda I)x = \emptyset$. Assim, λ

é um escalar, tal que esse sistema homogêneo tem outra solução que não $x = \emptyset$ e se verifica que o determinante $|A - \lambda I| = 0$. Portanto, os autovalores de uma matriz A podem ser definidos como os escalares λ , tais que $|A - \lambda I| = 0$. $|A - \lambda I|$ é um polinômio e suas raízes são os autovalores de A .

Alguns resultados sobre autovalores e autovetores serão usados nos próximos capítulos.

1. O número de autovalores diferentes de zero de uma matriz é igual a seu posto.
2. Os autovalores de uma matriz idempotente são todos iguais a 0 ou 1. Portanto, uma matriz de projeção tem autovalores iguais a 0 ou 1.
3. Os autovalores de uma matriz simétrica são reais.
4. O produto dos autovalores de uma matriz é igual a seu determinante.

1.7 FORMAS QUADRÁTICAS

Definição 3. Uma *forma quadrática* nas n variáveis y_1, \dots, y_n é uma função da forma

$$Q = \sum_{i=1}^n \sum_{j=1}^n a_{ij} y_i y_j$$

onde $\{a_{ij}\}$ são constantes.

Podemos escrever uma forma quadrática em notação matricial, definindo o vetor $y' = [y_1 \ \dots \ y_n]$ e a matriz $A = (a_{ij})$, de modo que

$$Q = y'Ay$$

A é chamada a *matriz da forma quadrática*. Pode-se assumir que a matriz de uma forma quadrática é simétrica, pois sempre é possível representar uma forma quadrática cuja matriz não é simétrica através de uma matriz simétrica.

Definição 4. A matriz simétrica A , $n \times n$, e a forma quadrática $y'Ay$, para $y \neq \emptyset$, são chamadas *positivas definidas* se $y'Ay > 0$ e, *positiva semidefinida*, se $y'Ay \geq 0$.

Se uma matriz é positiva definida, então seus autovalores são positivos e se ela é positiva semidefinida, eles são não negativos.

Pode-se provar que se uma matriz A é positiva semidefinida, existe T , tal que $A = TT'$. Se A for positiva definida, T será não singular.

1.8 ELIPSÓIDES

Definimos uma *esfera* sólida n -dimensional com o centro em a e raio r , como o conjunto de todos os pontos cuja distância, a partir de a , é menor ou igual a r . Em outras palavras, é o conjunto de pontos x , satisfazendo $\|x - a\| \leq r$ ou $\|x - a\|^2 \leq r^2$. Essa última expressão pode ser escrita de outra forma:

$$\|x - a\|^2 \leq r^2 \rightarrow (x - a)'(x - a) \leq r^2 \rightarrow (x - a)'I(x - a) \leq r^2$$

Se a matriz identidade for substituída por uma matriz M simétrica positiva definida qualquer, então a desigualdade $(x - a)'M(x - a) \leq r^2$ definirá um *elipsóide* em \mathbb{R}^n com centro em a . Um elipsóide é a extensão em \mathbb{R}^n do caso bidimensional em que $(x - a)'M(x - a) \leq r^2$ é uma elipse. Os comprimentos dos eixos, em cada uma das n dimensões do elipsóide, são inversamente proporcional aos autovalores de M . Assim, entende-se porque, quando a matriz da forma quadrática é a identidade (autovalores iguais a 1), tem-se uma esfera em \mathbb{R}^n . Portanto, esse é um caso especial da situação geral em que a matriz M é qualquer matriz positiva definida.

Capítulo 2

Modelos Classificatórios

2.1 INTRODUÇÃO

Uma das noções básicas da teoria dos conjuntos é a de *membro* de um conjunto. Se o elemento a é membro do conjunto A , representaremos tal fato por $a \in A$. Um conjunto será completamente determinado por seus membros e, assim, uma das formas mais comuns de se definir um conjunto é listando seus elementos. De outra forma, podemos dizer que o conjunto A é o conjunto de todos os elementos a que possuem uma certa propriedade. Em geral, pensamos em um conjunto como tendo alguns membros, mas, por vezes, é conveniente se considerar conjuntos que não possuem membros e que são chamados de *vazio*.

A propriedade que define os elementos de um conjunto e, portanto, o conjunto, estabelece a condição ou regra de pertinência a ele. Por exemplo, o conjunto dos municípios de um estado, o conjunto das construções de uma cidade.

Podemos ampliar a noção de conjunto e introduzir o conceito de *partição de um conjunto*. Uma regra que classifica os elementos de um conjunto A em

subconjuntos mutuamente exclusivos e exaustivos cria uma partição em A . Isto é, atribui a cada elemento de A , uma propriedade que o coloca como pertencente a um, e só um, desses subconjuntos. Assim, nenhum elemento de A deixa de pertencer a alguma partição; e o conjunto das partições recria A . Como exemplo, tomemos como conjunto A , os habitantes de uma cidade. A regra: idade menor que 18 anos, idade entre 18 e 60 anos e idade maior que 60 anos, cria uma partição em A . Ou seja, divide A em três subconjuntos: A_1 , idade menor que 18 anos; A_2 , idade entre 18 e 60 anos; e A_3 , idade maior que 60 anos. Então, todo habitante da cidade em questão pertence obrigatoriamente a A_1 , ou a A_2 , ou a A_3 . Pertencer à primeira partição implica em não pertencer à segunda e à terceira. Pertencer à segunda implica em não pertencer à primeira e à terceira. Igualmente, pertencer à terceira implica em não pertencer à primeira nem à segunda.

A regra que classifica, no exemplo acima, os habitantes da cidade em faixas etárias, é uma regra que determina um conjunto: o conjunto das pessoas que tem idade em certa faixa etária. Assim, a partição pode ser vista como uma extensão da noção de conjunto já que ela é criada por várias regras de pertinência aplicadas simultaneamente ao conjunto A .

2.2 FATORES DE CLASSIFICAÇÃO

Nas aplicações estatísticas, as noções que estamos discutindo, através do uso da teoria dos conjuntos, assumem nomes específicos com significados especiais que

veremos a seguir. O conjunto básico, que contém os elementos de interesse e que é submetido, eventualmente, a uma ou mais partições, é chamado *população*. A regra que define uma partição na população é um fator ou fator de classificação. Formalmente, podemos usar a seguinte definição.

Definição 1. Um *fator de classificação* de um conjunto de indivíduos é uma partição do conjunto desses indivíduos em subconjuntos disjuntos e exaustivos. Os subconjuntos formados pela partição são chamados *níveis do fator*.

Para um grupo de indivíduos, considerados seu sexo e país de nascimento, temos dois fatores de classificação. Isso induz duas partições no grupo: uma, segundo o sexo, com dois níveis e outra, segundo o país de nascimento, com tantos níveis quanto o número de países considerados. Cada subconjunto determinado pela partição criada por um fator estará, por definição, associado a um nível do fator. Vale enfatizar que se pode definir vários fatores de classificação simultaneamente na mesma população, sendo que cada um deles determina uma diferente partição na população. Esse procedimento fará surgir classificações derivadas que ocorrem em virtude das interseções dos diversos subconjuntos das partições.

Vejamos outro exemplo de partição induzida por dois fatores. Seja um estado da federação, o conjunto que chamamos população. O conjunto dos municípios do estado induz uma primeira partição, sendo, portanto, um fator de classificação. Consideremos, como outro fator, bacia hidrográfica. Admitamos que existam duas no estado. Cada ponto geográfico do estado estará em um município e em uma bacia. Assim, a cada elemento da população é possível associar um par de atributos: município e bacia. Ou ainda, podemos especificar o subconjunto ao qual pertence um elemento da população, referindo-nos a seu par de atributos.

Consideremos, agora, um terceiro fator no exemplo acima: distrito. Cada município pode conter diversos distritos que terão seus níveis denotados por nomes. Admita-se, ainda, que os nomes dos distritos sejam atribuídos por seus habitantes sem respeitar usos prévios do nome em outros municípios. Dessa forma, o nome, digamos Santana, em mais de um distrito, cria uma situação especial se desejamos nos referir a um elemento do estado através de seus atributos. Ao especificarmos 'Santana', é necessário especificar, também, a qual 'Santana' estamos nos referindo. O fator distrito, ao ser especificado, exige que se nomeie o município para sua correta localização. Dizemos que ele está embutido no fator município. Em contraste com essa situação, a especificação de bacia não necessita a menção do município. Dizemos que os fatores município e bacia são cruzados.

Definição 2. O fator de classificação B está *embutido* no fator de classificação A , se todos os indivíduos tendo o mesmo nível do fator B , necessariamente estão no mesmo nível do fator A .

A representação de fatores embutidos através de diagrama de Venn pode ser vista na Figura 1. Ali, A tem três níveis e B , dois.

Havendo três fatores A , B e C , onde B está embutido em A e C está embutido em B , é claro, pela definição, que C também estará embutido em A .

Para denotar que o fator B está embutido em A , usaremos $B(A)$.

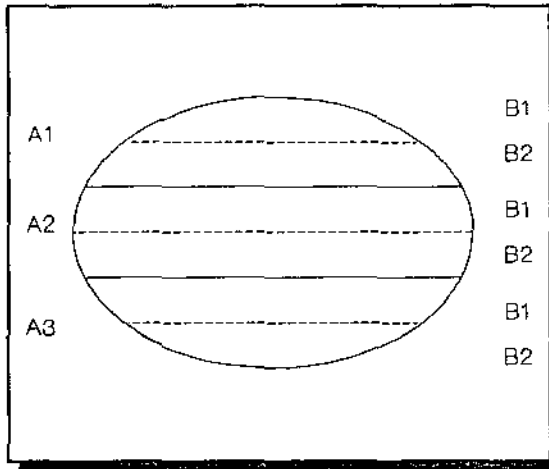


Figura 1. Fator B embutido no fator A

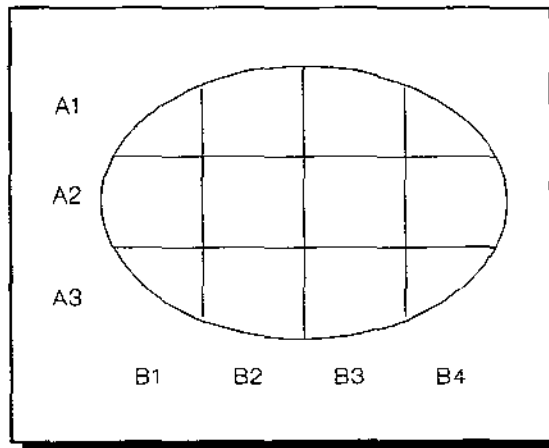


Figura 2. Fatores A e B cruzados

Definição 3. Dois fatores, tais que um não está embutido no outro, são ditos *cruzados*.

O diagrama de Venn, representando o fator *A* com três níveis e o fator *B* com quatro e, onde *A* e *B* são cruzados, pode ser visto na Figura 2.

Suponhamos que uma coleção de indivíduos seja classificada pelo fator sexo. Evidentemente, ainda é possível classificá-los por país de nascimento, origem religiosa, cor dos olhos, etc. Indivíduos, aqui, serão considerados como entidades que podem ser reconhecidas como algo simples e único como uma pessoa, um animal, tempos de observação de um sistema ou um pedaço de terra. Associado a cada indivíduo, existe um conjunto de atributos ou variáveis que identificam as p partições às quais os indivíduos pertencem. Aqui, elas serão representadas por x_1, x_2, \dots, x_p . Assim, x_j , $j = 1, \dots, p$, é um rótulo arbitrário, acessório à classificação, que identifica a partição.

Para trabalharmos sempre com variáveis matemáticas bem definidas, podemos associar a cada nível de um fator, uma variável indicadora. Por exemplo: seja um elemento genérico da população. Para um nível i do fator A , $\delta_i(x_j) = 1$, se o nível de x_j é i , e $\delta_i(x_j) = 0$, caso contrário. Se o fator A tem a níveis e, portanto, determina a subconjuntos, existirão a variáveis indicadoras associadas ao fator. Por exemplo, poderíamos ter $\delta_1(x_j) = 1$, para denotar que o indivíduo é um homem e $\delta_2(x_j) = 0$, para denotar que o indivíduo é mulher. No entanto, nada se alteraria se os valores associados aos níveis dos fatores fossem, por exemplo, 17 e 36, 3 e 10, -1 e 0 ou quaisquer outros dois valores. O que importa é estabelecer uma convenção que indique que um determinado indivíduo pertence ou não pertence a cada um dos níveis do fator em questão. Sem dúvida, a definição dada acima para $\delta_i(x_j)$ é bastante conveniente e será a utilizada daqui para frente.

As noções de classificação e medida podem se relacionar. Se fizer sentido, é possível associar números aos níveis do fator. Essa situação ocorrerá quando existir

Assim, a cada elemento do estado foi associado um conjunto de variáveis indicadoras que define exatamente aquele elemento. Por exemplo, 1 0 0 1 0 0 0 0 0 0 0 0 0 define o distrito a do município A , enquanto que 0 1 0 0 0 0 0 1 0 0 0 0 0 define o distrito a do município B . Note que a especificação do distrito a não define exatamente um elemento dentro da população: é preciso se fazer referência, também, ao município. Como já vimos, isso ocorre porque o fator *distrito* está embutido no fator *município*. Para simplificar a notação, usaremos daqui para frente, x_{ij} no lugar de $\delta_i(x_j)$.

Consideremos, agora, que a cada elemento da população esteja associado um valor numérico de interesse. Por exemplo, o número de habitantes de cada distrito dos municípios do estado. Assim, para um elemento específico da população podemos expressar o conjunto de variáveis indicadoras, x_1, \dots, x_p , que o define e, também, seu valor numérico de interesse. Digamos que nos municípios A , B e C , estamos interessados no peso dos adultos, segundo seu sexo. Portanto, temos dois fatores de classificação: município, com três níveis e sexo, com dois. A cada indivíduo adulto de cada sexo e de cada município existe associado um valor numérico que é seu peso. Seja i , um índice que especifique os municípios A , B e C ou municípios 1, 2 e 3. Seja j , um índice que especifique os sexos masculino e feminino ou sexos 1 e 2. Neste caso, os fatores são cruzados e induzem duas partições independentes na população. Cada subconjunto conterà n_{ij} indivíduos. Pictoricamente, podemos representar essa situação numa tabela 3×2 , onde as linhas representam os municípios e as colunas, os sexos. Cada combinação de níveis do primeiro fator com níveis do segundo fator será uma casela da tabela.

Definição 4. Se $n_{ij} = n_{i'j'}$, para quaisquer i, i', j, j' , ou seja, o número de observações por casela é igual para todas as combinações dos níveis dos fatores, a estrutura é dita *balanceada*. Caso contrário, será *desbalanceada*.

2.3 REPRESENTAÇÃO ESQUEMÁTICA DAS ESTRUTURAS DE CLASSIFICAÇÃO

É conveniente supor a existência de dois fatores triviais, ambos com um nível somente. Em uma ponta, um fator universal, a média geral μ , que embute todos os demais fatores. Na outra, um fator e , específico para cada observação e embutido em todos os outros fatores.

Uma forma útil de representar as estruturas de classificação é vista na Figura 3.

Cada 'o' corresponde a um fator de classificação e a regra básica para a construção desses esquemas é a seguinte: se um fator B está embutido em um fator A , ligá-los por uma linha como na Figura 3(2). O topo da figura é, de certa forma, desnecessário, mas é útil para indicar que o diagrama está completo superiormente. Associar o rótulo μ a ele, para indicar que todas as observações estão no mesmo corpo de dados. O diagrama da Figura 3(1) sugere que um indivíduo pode estar no nível i de A e no nível j de B independentemente. O diagrama da Figura 3(2) sugere que uma observação, estando em um nível de B , obrigatoriamente estará em um certo nível de A . Ao ponto inferior do diagrama, associar o rótulo e .

Cap 2

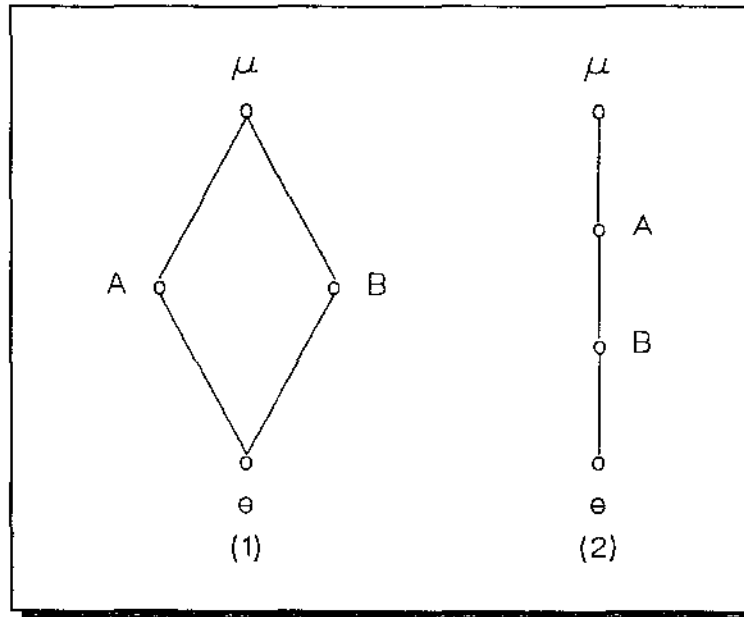


Figura 3. As duas possíveis estruturas com dois critérios de classificação

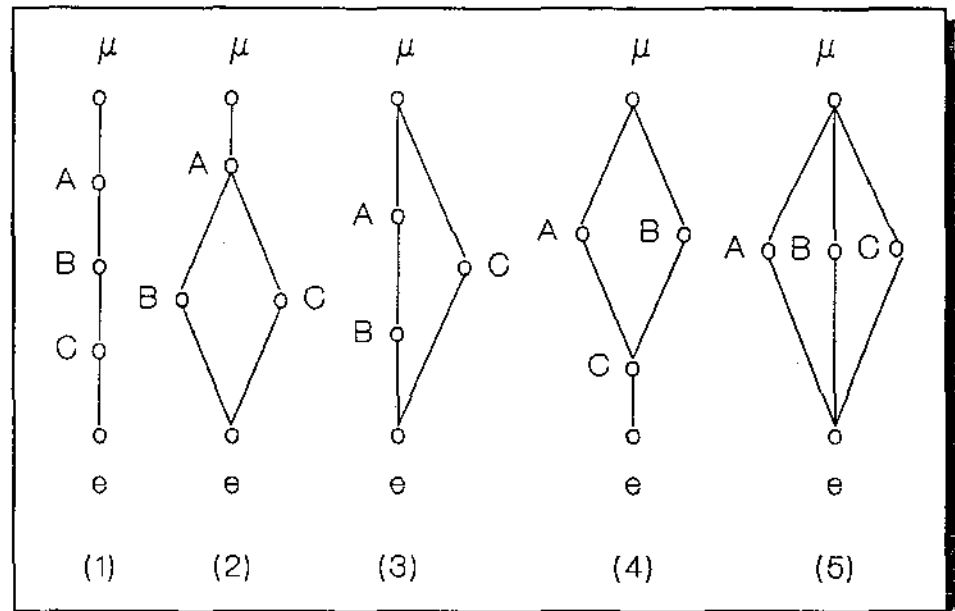


Figura 4. As cinco possíveis estruturas com três fatores de classificação

O diagrama da Figura 3(1) diz-nos que temos dois critérios de classificação: os fatores A e B . O diagrama da Figura 3(2) diz-nos que o fator B está embutido no fator A .

Ilustremos esses diagramas para casos mais complexos. Consideremos os diagramas da Figura 4.

1. Aqui, temos três fatores de classificação: A , $B(A)$ e $C(B)$. Um exemplo seria: casas C , em cidades B , nos estados A e no Brasil μ .
2. Aqui também, temos três fatores de classificação: A , B e C , com $B(A)$, $C(A)$ e B e C cruzados. Continuando com o exemplo acima, μ é o Brasil; A , a cidade; B , o tipo de casa (alvenaria ou não); C , o número de cômodos, enquanto que e se refere a uma casa individual.
3. Novamente temos os três fatores A , B e C . μ poderia ser o Brasil; A , o estado; B , a cidade e C , a localização (urbana ou rural). e se refere a uma casa individual.
4. Uma possível interpretação seria: μ , o Brasil; A , meio urbano ou rural; B , casa própria ou alugada; C , o número de cômodos dentro das classes A e B . Para e , a interpretação é equivalente às situações anteriores.
5. Neste caso poderíamos interpretar os símbolos da seguinte forma: μ , o Brasil; A , meio urbano ou rural; B , casa própria ou alugada; C , com ou sem telefone. Mais uma vez, e se refere a uma casa individual.

Já vimos que a cada um dos indivíduos da população pode estar associada uma característica numérica, ou resposta, de interesse. Representemos essa característica por $y_{ij\dots qr}$, onde cada um dos índices i, j, \dots, q representa um fator e r , as $n_{ij\dots q}$ replicações dentro daquela combinação de fatores.

Definição 5. Uma média das respostas sobre qualquer subconjunto dos índices que compõem a resposta é definida como *média parcial*.

Por exemplo, com dois critérios de classificação, uma forma útil de representar as médias parciais e que será usada a seguir é a seguinte:

$$y_{i.} = \sum_{r=1}^{n_{ij}} y_{ijr}, \quad \bar{y}_{i.} = \frac{1}{n_{ij}} y_{i.}$$

$$y_{i..} = \sum_{j=1}^b \sum_{r=1}^{n_{ij}} y_{ijr}, \quad \bar{y}_{i..} = \frac{1}{b} y_{i..}$$

$$y_{.j.} = \sum_{i=1}^a \sum_{r=1}^{n_{ij}} y_{ijr}, \quad \bar{y}_{.j.} = \frac{1}{a} y_{.j.}$$

onde n_{ij} é o número de replicações na combinação do nível i do fator A com o nível j do fator B ; a , o número de níveis do fator A e b , o número de níveis do fator B . Para a média geral, usaremos

$$y_{...} = \sum_{i=1}^a \sum_{j=1}^b \sum_{r=1}^{n_y} y_{ijr}, \quad \bar{y}_{...} = \frac{1}{n_{...}} y_{...}$$

Definição 6. Uma média parcial é *admissível* se, presente um índice correspondente a um fator, também estão presentes todos os índices dos fatores que o embutem. Caso contrário, a média é *inadmissível*.

Na estrutura da Figura 4(3), vemos que as médias dos subconjuntos $\{\mu\}$, $\{\mu, A\}$, $\{\mu, A, B\}$, $\{\mu, A, C\}$, $\{\mu, C\}$ e $\{\mu, A, B, C\}$ são admissíveis. Por outro lado, as médias dos subconjuntos $\{\mu, B\}$ e $\{\mu, B, C\}$ são inadmissíveis.

Consideremos, na mesma estrutura, que o fator A seja indexado por i ; o fator B , por j e C , por k . As médias parciais eventualmente possíveis de serem calculadas serão indexadas por $i...$, $.j..$, $..k.$, $ij..$, $i.k.$, $.jk.$ e $ijk.$. Pela Definição 6, as médias indexadas por $.j..$ e $.jk.$ são inadmissíveis. Note que podemos, também, considerar y_{ijkr} como uma média admissível.

As respostas individuais podem ser representadas através de identidades algébricas como veremos a seguir. Essas identidades são úteis para se compreender como os efeitos estão combinados para fornecer a resposta observada.

A identidade algébrica é formada, à esquerda, pela resposta indexada e, à direita, por uma soma de parcelas construídas, cada uma, da seguinte forma: o primeiro termo é uma média admissível; os termos seguintes são formados pela eliminação de um ou mais índices dessa média, dentre aqueles correspondentes a

fatores que não embutem outro fator naquela média admissível. Aos termos de onde se retira um número ímpar de índices, atribui-se o sinal negativo. Se o número de índices retirados é par, seu sinal será positivo. O significado de cada parcela será visto à frente.

Como exemplo, consideremos a estrutura da Figura 3(1). As médias admissíveis são: $\bar{y}_{...}$, $\bar{y}_{i..}$, $\bar{y}_{.j.}$, $\bar{y}_{ij.}$ e y_{ijr} . A identidade algébrica será:

$$y_{ijr} = \bar{y}_{...} + (\bar{y}_{i..} - \bar{y}_{...}) + (\bar{y}_{.j.} - \bar{y}_{...}) + (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...}) + (y_{ijr} - \bar{y}_{ij.})$$

Para a estrutura da Figura 3(2), a identidade algébrica será

$$y_{ijr} = \bar{y}_{...} + (\bar{y}_{i..} - \bar{y}_{...}) + (\bar{y}_{.j.} - \bar{y}_{i..}) + (y_{ijr} - \bar{y}_{ij.})$$

Uma idéia que será útil à frente é a da população conceitual de respostas. Tomemos Y_{ij} , $i=1,\dots,N$, $j=1,\dots,t$, como o valor da resposta do indivíduo i ao tratamento j . Se o indivíduo i sob o efeito do primeiro tratamento forneceu Y_{i1} , o valor Y_{i2} não será observado. No entanto, ele existe: o indivíduo i teria fornecido Y_{i2} se estivesse sob o efeito do segundo tratamento. Assim, $\{Y_{ij}\}$, é o conjunto de respostas de N indivíduos sob os efeitos de t tratamentos. Poderíamos definir $\{Y_{ij}\}$ como uma *população conceitual de respostas*.

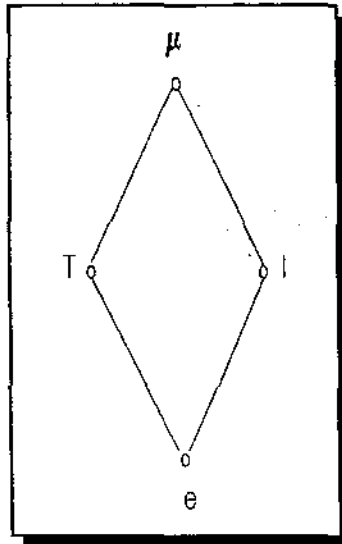


Figura 5. Estrutura de classificação para a população conceitual de respostas

É possível, para a população conceitual de respostas, considerando T o fator tratamento e I o fator indivíduo, representar a estrutura populacional como na Figura 5 e escrever a identidade algébrica

$$Y_{ij} = \bar{Y}_{..} + (\bar{Y}_i - \bar{Y}_{..}) + (\bar{Y}_j - \bar{Y}_{..}) + (Y_{ij} - \bar{Y}_i - \bar{Y}_j + \bar{Y}_{..}) \tag{1}$$

$\bar{Y}_{..}$ denota a média geral das respostas dos N indivíduos sob os efeitos dos t tratamentos. \bar{Y}_i denota a média das respostas do indivíduo i sob o efeito dos t tratamentos

e \bar{Y}_j , a média das respostas dos N indivíduos sob o efeito do tratamento j . Assim, $(\bar{Y}_j - \bar{Y}_{..})$ é a diferença entre a resposta média de todos os indivíduos sob o efeito do tratamento j e a média geral. Ou ainda, o efeito do tratamento j . $(\bar{Y}_i - \bar{Y}_{..})$ é a diferença entre a resposta média aos efeitos dos níveis dos tratamentos no indivíduo i e a média geral. $(Y_{ij} - \bar{Y}_i - \bar{Y}_j + \bar{Y}_{..})$ mede o desvio da resposta do indivíduo i da média dos demais indivíduos e da média dos efeitos dos tratamentos. Essa última parcela é o efeito do fator e presente nos diagramas das estruturas de classificação.

A população conceitual de respostas não pode ser obtida na prática. Nessa situação, o que ocorre é a atribuição de indivíduos a tratamentos dando origem a uma estrutura que pode ser representada como na Figura 6. Note que I está embutido em T . As médias admissíveis são $\bar{y}_{...}$, $\bar{y}_{i..}$, $\bar{y}_{ij.}$ e y_{ijr} . A identidade algébrica será

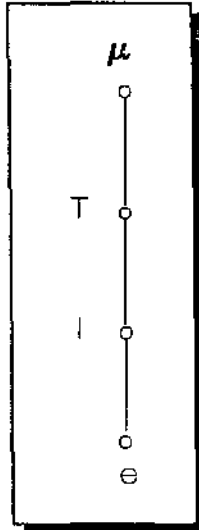


Figura 6. Estrutura de classificação amostral

$$y_{ij} = \bar{y}_{..} + (\bar{y}_j - \bar{y}_{..}) + (y_{ij} - \bar{y}_j)$$

Pode-se verificar que elevando-se ao quadrado ambos os membros das identidades algébricas acima e somando para todo i e j , os produtos cruzados desaparecem. Esse fato tem grande importância e será entendido ao estudarmos a geometria envolvida nas estruturas de classificação no Capítulo 3.

2.4 O PAPEL DA ALEATORIZAÇÃO

Seja a identidade (1). Para simplificar, consideremos o último termo à direita igual a zero. Não é trivial se assumir tal fato, mas mesmo que isso não seja verdade, a interpretação que se seguirá continuará válida. Dessa forma, a identidade (1) se torna

$$Y_{ij} = \bar{Y}_{..} + (\bar{Y}_i - \bar{Y}_{..}) + (\bar{Y}_j - \bar{Y}_{..}) \quad (2)$$

Por definição,

$$\sum_i (\bar{Y}_i - \bar{Y}_{..}) = \sum_j (\bar{Y}_j - \bar{Y}_{..}) = 0$$

Tomemos, agora, um experimento completamente aleatorizado onde será observado um subconjunto dos Y_{ij} . Escolhamos $n = rt$ unidades experimentais, divididas em t grupos de tamanho r . Denotemos os valores observados nas unidades experimentais por y_{jk} , $j = 1, \dots, t$, $k = 1, \dots, r$. Além disso, definamos a seguinte função indicadora: $\delta_{ij}^{(k)} = 1$, se a replicação k do tratamento j cai na unidade experimental i ; $\delta_{ij}^{(k)} = 0$, caso contrário. Note que o plano experimental determina as probabilidades

$$P\{\delta_{ij}^{(k)} = 1\} = \frac{1}{N}$$

$$P\{\delta_{ij}^{(k)} = 1 | \delta_{ij'}^{(k)} = 1\} = 0$$

Assim, a probabilidade de que uma determinada replicação de um certo tratamento caia no indivíduo i da população conceitual de respostas é $1/N$. Além disso, a probabilidade de que determinada replicação de um certo tratamento caia no indivíduo i da população, dado que a esse indivíduo foi atribuído outro tratamento, é zero. Ou seja, a cada indivíduo da população é atribuído um, e somente um, tratamento. Usando essa função, podemos representar os valores observados nas unidades experimentais através dos Y_{ij} da seguinte forma:

$$y_{jk} = \sum_{i=1}^N \delta_{ij}^{(k)} Y_{ij} \tag{3}$$

Substituindo em (3), o valor de Y_{ij} em (2) e simplificando, obteremos

$$y_{jk} = \bar{Y}_{..} + (\bar{Y}_{.j} - \bar{Y}_{..}) + \sum_i \delta_{ij}^{(k)} (Y_{ij} - \bar{Y}_{.j}) \quad (4)$$

Ou ainda, denotando os três termos à direita de (4), respectivamente, por μ , A_j e e_{jk} , temos

$$y_{jk} = \mu + A_j + e_{jk} \quad (5)$$

Em (5), μ denota a média geral, A_j o efeito do tratamento j e e_{jk} , um termo que pode ser descrito como uma falha em serem idênticas, repetidas observações feitas sob condições similares. Esse termo é descrito como o *erro*. Ele depende do esquema de aleatorização usado para definir as unidades experimentais que serão observadas. Pode-se mostrar que $E(e_{jk}) = 0$, $var(e_{jk}) = \sigma^2$ e $cov(e_{jk}) = -\rho\sigma^2$, onde ρ e σ^2 são parâmetros populacionais. Convém salientar que o cálculo desses momentos foi realizado apenas se levando em conta a aleatorização. Nenhuma outra hipótese foi necessária.

Com o raciocínio desenvolvido acima, vemos que cada unidade experimental observada pode ser representada através de certos parâmetros. Tais parâmetros dependem dos valores fornecidos pelos elementos componentes da população.

Mesmo quando a aleatorização não é realizada, frequentemente é razoável se supor que uma expressão com a estrutura semelhante a (5), seja adequada para a

representação do vetor de observações y .

2.5 OS MODELOS

Definição 7. Um *modelo* é uma equação matemática que envolve variáveis aleatórias, variáveis matemáticas e parâmetros.

Definição 8. Um *modelo linear* é uma equação que envolve variáveis aleatórias, variáveis matemáticas e parâmetros e que é linear nos parâmetros, assim como nas variáveis aleatórias.

Por exemplo, se τ_0 , τ_1 e τ_2 são parâmetros desconhecidos e X e Y são, respectivamente, variáveis matemática e aleatória, então, $\tau_0 + \tau_1 X + \tau_2 Y = 0$ é um modelo linear. Da mesma forma, $\tau_0 + \tau_1 e^X + \tau_2 Y \cos X = 0$, também é. Porém, modelos como $\tau_0^2 + X \sin \tau_1 + \tau_2 = 0$ não são lineares nos parâmetros.

Consideremos o caso em que temos observações indexadas por i, j, r , com $i = 1, 2, \dots, a$, $j = 1, 2, \dots, b$ e $r = 1, 2, \dots, n_{ij}$. Haverá ab classes diferentes de observações e

suas médias, $\mu_{ij} = \sum_{r=1}^{n_{ij}} y_{ijr}$, podem ser representadas como na Tabela I, onde

$$\mu_i = \sum_{j=1}^b \mu_{ij} / b \text{ e } \mu_j = \sum_{i=1}^a \mu_{ij} / a.$$

Tabela I. Médias das classes para dois critérios de classificação

$i \quad j$	1	2	...	b	Média
1	μ_{11}	μ_{12}	...	μ_{1a}	$\mu_{1.}$
2	μ_{21}	μ_{22}	...	μ_{2b}	$\mu_{2.}$
.
.
.
a	μ_{a1}	μ_{a2}	...	μ_{ab}	$\mu_{a.}$
Média	$\mu_{.1}$	$\mu_{.2}$...	$\mu_{.b}$	$\mu_{..}$

Se dois fatores A e B são cruzados, é possível representá-los pela estrutura da Figura 3(1). Observações e médias com o mesmo valor de i pertencem a uma classe do fator com a níveis. Observações com o mesmo valor de j pertencem a uma classe do fator com b níveis. Levando em conta as médias das classes, podemos escrever a identidade

$$\mu_{ij} = \mu_{..} + (\mu_{i.} - \mu_{..}) + (\mu_{.j} - \mu_{..}) + (\mu_{ij} - \mu_{i.} - \mu_{.j} + \mu_{..}) \quad (6)$$

Cada um dos termos de (6) tem um significado especial e a eles podem ser dados nomes. O primeiro, $\mu_{..}$, é a média geral. O segundo, $(\mu_{i.} - \mu_{..})$, a diferença entre a média do nível i do fator A e a média geral, é chamado efeito de A_i . De modo equivalente, $(\mu_{.j} - \mu_{..})$ é o efeito de B_j . O último termo é chamado *interação* do nível A_i com o nível $(\mu_{.j} - \mu_{..})$. Ele mede a quantidade pela qual μ_{ij} deixa de ser a soma

entre a média geral e os efeitos de A_i e B_j . Note que se pode medir o efeito de A em cada nível de B : $A(B_j) = \mu_{ij} - \mu_{.j}$. Se esses efeitos dependerem de j , ou seja, do nível do fator B , dizemos que há interação entre os dois fatores e escrevemos, $A*B_{ij} \neq 0$. De modo equivalente, $B(A_i) = \mu_{ij} - \mu_{i.}$ e o resultado de $A*B_{ij}$ será o mesmo. Assim, a interação é simétrica, quer olhada através de A , quer através de B . Consideremos, por exemplo, homens ou mulheres como fator A e droga I ou droga II como fator B . Na ausência de interação, dizemos que a diferença entre a droga II e a droga I é a mesma para ambos os sexos. Desse modo, as drogas I e II não interagem com sexo em relação a alguma medida, por exemplo, temperatura corporal. Aqui, os fatores A e B são cruzados, as médias μ_{ij} , $\mu_{i.}$, $\mu_{.j}$ e $\mu_{..}$ podem ser calculadas, tem sentido e são admissíveis.

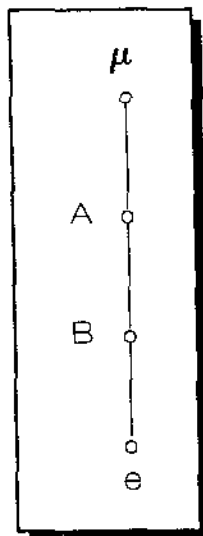


Figura 7. Estrutura de classificação para fatores embutidos

Por outro lado, suponhamos que o fator B esteja embutido no fator A . É ainda conveniente representar as populações e suas médias como na Tabela I. Entretanto, o índice j isolado não representa os níveis do fator B , mas os níveis desse fator, dentro de um nível do fator A . O conjunto de subpopulações com índice $j = 3$, por exemplo, não pertence à classe formada por uma partição. O fator embutido tem, de fato, ab níveis. Para este caso, a representação da estrutura de classificação é a da Figura 6 e uma identidade em termos de médias de classes é

$$\mu_{ij} = \mu_{..} + (\mu_{i.} - \mu_{..}) + (\mu_{ij} - \mu_{i.})$$

A média da casela (i,j) é dada pela média geral $\mu_{..}$, mais o efeito do nível i de A ,

mais o efeito do nível j de B , dentro do nível i de A . As médias admissíveis são μ_{ij} , μ_i e $\mu_{..}$. A média μ_j não é admissível.

Com base nas considerações feitas para as médias populacionais, é possível transferir a análise para um grupo de indivíduos observados.

Se estivermos observando uma coleção de indivíduos classificados por um único fator A , podemos representar os valores individuais observados por

$$y_{ir} = \mu_{..} + (\mu_i - \mu_{..}) + (y_{ir} - \mu_i) \quad (7)$$

Cada termo à direita de (7) pode ser substituído da seguinte forma:

$$\mu_{..} = \mu$$

$$\mu_i - \mu_{..} = A_i$$

$$y_{ir} - \mu_i = e_{ir}$$

μ é a média geral, A_i o efeito do nível i do fator A e e_{ir} , um desvio particular de cada indivíduo. Então, o modelo (7) pode ser expresso como

$$y_{ir} = \mu + A_i + e_{ir}$$

Havendo dois fatores cruzados A e B , podemos escrever a identidade

$$y_{ijr} = \mu_{...} + (\mu_{i..} - \mu_{...}) + (\mu_{.j.} - \mu_{...}) + (\mu_{ij.} - \mu_{i..} - \mu_{.j.} + \mu_{...}) + (y_{ijr} - \mu_{ij.}) \quad (8)$$

De modo equivalente, é possível representar cada termo à direita de (8) por

$$\mu_{...} = \mu$$

$$\mu_{i..} - \mu_{...} = A_i$$

$$\mu_{.j.} - \mu_{...} = B_j$$

$$\mu_{ij.} - \mu_{i..} - \mu_{.j.} + \mu_{...} = A * B_{ij}$$

$$y_{ijr} - \mu_{ij.} = e_{ijr}$$

B_j é o efeito do nível j do fator B . Então, o modelo (8) pode ser expresso como

$$y_{ijr} = \mu + A_i + B_j + A * B_{ij} + e_{ijr} \quad (9)$$

Outra situação é a dos fatores embutidos. Uma identidade será

$$y_{ijr} = \mu_{...} + (\mu_{i..} - \mu_{...}) + (\mu_{ij.} - \mu_{i..}) + (y_{ijr} - \mu_{ij.}) \quad (10)$$

Substituindo os termos de (10) da seguinte forma

$$\begin{aligned}\mu_{...} &= \mu \\ \mu_{i..} - \mu_{...} &= A_i \\ \mu_{ij.} - \mu_{i..} &= B(A)_{ij} \\ y_{ijr} - \mu_{ij.} &= e_{ijr}\end{aligned}$$

obteremos

$$y_{ijr} = \mu + A_i + B(A)_{ij} + e_{ijr} \quad (11)$$

2.6 A MATRIZ DO MODELO

O modelo (11) pode ser usado para representar a estrutura do exemplo da seção 2.2 sobre as cidades e distritos. Ali, A_i representa as cidades A , B e C ; $B(A)_{ij}$ representa os distritos e y_{ijr} poderia ser, por exemplo, o consumo de energia elétrica da cidade i , distrito j dentro da cidade i . Supondo $r = 1, 2$, o modelo (11) pode ser representado em forma matricial por

$$y = X\beta + e$$

onde y e e são vetores 20×1 , β , o vetor 13×1 de parâmetros, representando os

efeitos, e X , uma matriz 20×13 composta de 0's e 1's. Escrevendo explicitamente o modelo (11) para esse exemplo, teremos

$$\begin{array}{l}
 y_{111} \\
 y_{112} \\
 y_{121} \\
 y_{122} \\
 y_{131} \\
 y_{132} \\
 y_{141} \\
 y_{142} \\
 y_{211} \\
 y_{212} \\
 y_{221} \\
 y_{222} \\
 y_{231} \\
 y_{232} \\
 y_{311} \\
 y_{312} \\
 y_{321} \\
 y_{322} \\
 y_{331} \\
 y_{332}
 \end{array}
 =
 \begin{array}{cccccccccccc}
 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
 \end{array}
 \begin{array}{l}
 A_1 \\
 A_2 \\
 A_3 \\
 B(A)_{11} \\
 B(A)_{12} \\
 B(A)_{13} \\
 B(A)_{14} \\
 B(A)_{21} \\
 B(A)_{22} \\
 B(A)_{23} \\
 B(A)_{31} \\
 B(A)_{32} \\
 B(A)_{33}
 \end{array}
 +
 \begin{array}{l}
 e_{111} \\
 e_{112} \\
 e_{121} \\
 e_{122} \\
 e_{131} \\
 e_{132} \\
 e_{141} \\
 e_{142} \\
 e_{211} \\
 e_{212} \\
 e_{221} \\
 e_{222} \\
 e_{231} \\
 e_{232} \\
 e_{311} \\
 e_{312} \\
 e_{321} \\
 e_{322} \\
 e_{331} \\
 e_{332}
 \end{array}$$

Note que cada vetor linha x'_i da matriz é equivalente a uma linha da estrutura do exemplo, acrescida de um primeiro elemento igual a 1. Define-se a matriz X , como a *matriz do modelo*.

Consideremos outro exemplo. Seja o modelo (9), com A e B tendo dois níveis cada um. Suponhamos que existam duas replicações em cada combinação de tratamentos. Um indivíduo observado no primeiro nível dos dois fatores terá seu vetor x'_i representado por $[1 \ 1 \ 0 \ 1 \ 0 \ 1 \ 0 \ 0 \ 0]$. O primeiro 1 representa o efeito da

média. O segundo 1, o efeito do primeiro nível de A e, conseqüentemente, o 0 a seguir, representa a ausência do efeito do segundo nível de A . Os dois elementos seguintes dizem respeito à presença e à ausência, respectivamente, dos efeitos de B_1 e B_2 . Como os dois fatores são cruzados e cada um tem dois níveis, eles podem se combinar de quatro formas diferentes e dar origem às classes A_1B_1 , A_1B_2 , A_2B_1 e A_2B_2 . A representação simbólica dessas classes no vetor x'_i segue o mesmo padrão. Dos quatro últimos elementos do vetor acima, o 1 representa a influência conjunta de ambos os efeitos A_1 e B_1 . Os três 0's representam a ausência das outras combinações dos níveis de A e B . Note que o valor que representa A_1B_1 é igual à multiplicação dos elementos de x'_i correspondentes a A_1 e B_1 : 1 e 1. Para as demais classes temos a multiplicação de 1 por 0 e, assim, surgem os três últimos elementos 0 do vetor. Se, de dez indivíduos, tivéssemos os seis primeiros sob o efeito de A_1 e os quatro restantes sob A_2 ; os quatro primeiros, também sob o efeito de B_1 e os seis últimos, também sob B_2 , teríamos

$$X = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix} \quad (12)$$

Conforme veremos no Capítulo 7, a construção de X com as colunas representadas em uma ordem diferente daquela de (12) não altera qualquer análise que se faça. A construção de X para a ocorrência de fatores embutidos é equivalente.

Em razão de sua estrutura, a matriz X , assim construída, pode apresentar uma característica que chama atenção: por exemplo, em (12), as somas da segunda coluna com a terceira e da quarta com a quinta são iguais à primeira. E também, a soma das quatro últimas colunas é igual à primeira. Portanto, existem dependências lineares entre as linhas e/ou colunas de X e seu posto é menor que sua menor dimensão. Ou seja, essas matrizes tem posto incompleto. Esse fato é de grande importância, conforme veremos nos próximos capítulos.

Capítulo 3

O Ajuste de Mínimos Quadrados

3.1 INTRODUÇÃO

Seja um conjunto de medidas y_1, y_2, \dots, y_n tomadas em um grupo de n indivíduos. Desejamos explicar ou descrever tais medidas através das variáveis x_1, x_2, \dots, x_p , $p < n$. Podemos estar interessados em fornecer uma explicação exata. Por exemplo, da Física sabemos que um móvel em movimento retilíneo uniforme percorre uma distância d com a velocidade v no tempo t , segundo a relação $d = vt$. Dadas duas variáveis, temos exatamente o valor da terceira.

Em outras situações, lidaremos com explicações ou descrições aproximadas. Por exemplo, podemos descrever os dados dizendo que eles constituem um conjunto de valores que variam de 10 a 50. Ou ainda, com um conjunto de pares $\{y_i, x_i\}$, $i = 1, 2, \dots, n$, onde y_i é o peso de uma criança e x_i sua idade, temos, para $n = 3$, (12.5;2), (13.5;3) e (17.0;5). Aproximadamente, $y_i = 8 + 2x_i$, $i = 1, 2, 3$. Ou ainda,

$$y_i \doteq 8 + 2x_i \quad (1)$$

3.2 APROXIMAÇÕES

A expressão (1) pode ser escrita de outra forma:

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 12.5 \\ 13.5 \\ 17.0 \end{bmatrix} \doteq \begin{bmatrix} 8 \\ 8 \\ 8 \end{bmatrix} + 2 \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = 8 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + 2 \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

Assim, o vetor de dados, $\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$, foi representado aproximadamente por uma

combinação linear dos vetores $\mathbf{1} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ e $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$. Então,

$$\mathbf{y} \doteq 8\mathbf{1} + 2\mathbf{x}$$

De forma geral, podemos representar o vetor de medidas por

$$\mathbf{y} \doteq \beta_1 \mathbf{1} + \beta_2 \mathbf{x} \quad (2)$$

À medida que os valores β_1 e β_2 variam, temos diferentes aproximações.

A expressão (2) pode ser escrita em forma matricial como

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} \quad (3)$$

onde \mathbf{y} é um vetor $n \times 1$, \mathbf{X} , uma matriz $n \times p$ e $\boldsymbol{\beta}$, um vetor $p \times 1$.

A estrutura da matriz do modelo pode ser bastante diversa. No Capítulo 2, vimos matrizes \mathbf{X} cujas colunas eram formadas arbitrariamente por 1's e 0's e que indicavam condição de pertinência a classes. No entanto, chamou-se a atenção para a eventualidade de existirem colunas cujos elementos não eram símbolos arbitrários, mas valores numéricos associados às classes às quais os indivíduos pertenciam.

Por outro lado, podem existir matrizes \mathbf{X} onde todas as colunas são valores numéricos. Esse é um caso especial da situação geral onde \mathbf{X} é construída com variáveis indicadoras. Daqui para frente, trataremos, salvo menção em contrário, destes casos, de modo que a particularização se torna bastante simples.

A matriz \mathbf{X} pode ser vista como um conjunto de vetores coluna \mathbf{x}_i , $i = 1, 2, \dots, p$:

$$\mathbf{X} = [\mathbf{x}_1 \ \mathbf{x}_2 \ \dots \ \mathbf{x}_p]$$

e, assim,

$$\mathbf{y} \doteq \beta_1 \mathbf{x}_1 + \beta_2 \mathbf{x}_2 + \dots + \beta_p \mathbf{x}_p$$

ou seja, \mathbf{y} está sendo representado, aproximadamente, por uma combinação linear dos vetores $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p$.

Também no Capítulo 2, vimos que a matriz X pode apresentar dependências lineares entre suas colunas e, portanto, pode ter posto incompleto.

Devemos determinar o valor do vetor β para representar \mathbf{y} aproximadamente. Resolvendo o sistema $\mathbf{y} = X\hat{\beta}$, encontraremos o valor de $\beta \in \mathbb{R}^p$, $\hat{\beta}$, tal que \mathbf{y} é aproximadamente igual a $X\beta$.

Para descrever ou explicar um conjunto de medidas aproximadamente, torna-se necessário definir 'aproximadamente'. Poderíamos, por exemplo para a expressão (1), considerar o módulo da diferença entre seus dois membros, $|y_i - (8 + 2x_i)|$. Quanto menor esse valor, mais próximo y_i estará de sua representação $8 + 2x_i$. De outra forma, poderíamos analisar a norma da diferença entre os vetores \mathbf{y} e $X\beta$, $\|\mathbf{y} - X\beta\|$. Queremos, então, encontrar o número $d(\mathbf{y}, X\beta)$ que é a distância de \mathbf{y} a $X\beta$ e que desejamos seja mínima.

No presente trabalho, tomaremos como medida de distância, não a norma $\|\mathbf{y} - X\beta\|$, mas a norma ao quadrado, $\|\mathbf{y} - X\beta\|^2$. Esse é o método dos *mínimos quadrados*. Portanto, quando nos referirmos à mínima distância entre dois vetores - ou pontos em \mathbb{R}^n - estaremos nos referindo à mínima distância no sentido dos mínimos

quadrados.

Para um β escalar, a questão da aproximação de mínimos quadrados pode ser representada graficamente como na Figura 1, onde b é o valor que minimiza $d(y, x\beta)$.

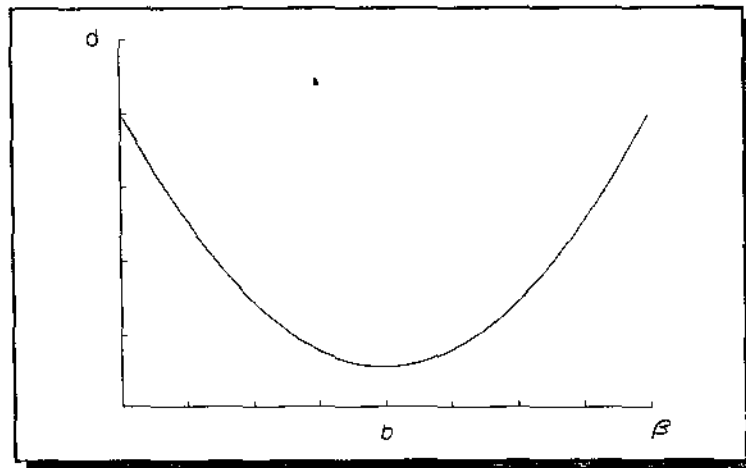


Figura 1. Valor de β que minimiza a distância $d(y, x\beta)$

Seja um vetor arbitrário β de dimensão $p \times 1$ em \mathbb{R}^p . O ajuste será dado pelo ponto em \mathbb{R}^n , $X\beta = [x'_1\beta_1 \ x'_2\beta_2 \ \dots \ x'_n\beta_n]$. Cada uma das coordenadas desse ponto corresponde a um componente do vetor de dados $y' = [y_1 \ y_2 \ \dots \ y_n]$ e $d(y, X\beta) = \sqrt{(y - X\beta)'(y - X\beta)}$.

3.3 O MÉTODO DOS MÍNIMOS QUADRADOS

Para o modelo $y = X\beta$, desejamos determinar um valor b para $\beta \in \mathbb{R}^p$ que minimize a distância $d(y, X\beta)$ ou $[d(y, X\beta)]^2$, o que é equivalente. Queremos, portanto, um valor b para β , tal que

$$\begin{aligned} \|y - Xb\|^2 &= \min_{\beta \in \mathbb{R}^p} \|y - X\beta\|^2 = \min_{\beta \in \mathbb{R}^p} [(y - X\beta)'(y - X\beta)] = \\ &= \min_{\beta \in \mathbb{R}^p} \left[\sum_{i=1}^n \left(y_i - \sum_{j=1}^p x_{ij}\beta_j \right)^2 \right] = \min_{\beta \in \mathbb{R}^p} Q(\beta) \end{aligned}$$

Será provado adiante que resolvendo o sistema de derivadas parciais $\frac{\partial Q(\beta)}{\partial \beta_j} = 0, j=1,2,\dots,p$, obtém-se um mínimo para $Q(\beta)$ e que esse mínimo é global.

Assim,

$$Q(\beta) = (y - X\beta)'(y - X\beta) = y'y - 2y'X\beta + \beta'X'X\beta$$

fornece o seguinte sistema:

$$\frac{\partial Q(\beta)}{\partial \beta} = -2X'y + 2X'X\beta = \emptyset \Rightarrow X'X\beta = X'y$$

ou

$$X'Xb = X'y \quad (3)$$

Definição 1. O sistema (3) é chamado *sistema das equações normais*.

Resolvido tal sistema, obteremos um b que fornece o mínimo procurado.

Podemos verificar

1. que o sistema (3) é consistente,
2. que o mínimo fornecido por b é global e
3. algumas propriedades e consequências do sistema.

Para essa análise, alguns resultados adicionais são convenientes.

► **Lema 1.** $a'a = 0 \leftrightarrow a = \emptyset$.

Prova. $a'a = \sum_i a_i^2$. Como cada termo é não negativo, $\sum_i a_i^2 = 0 \rightarrow a_i = 0$. A

implicação do lado contrário é trivial. ◀

Definição 2. O sistema $Az = b$ é *consistente*, se existe pelo menos um z_0 tal que $Az_0 = b$. Caso contrário, o sistema é dito ser *inconsistente*.

Por exemplo, o sistema

$$\begin{cases} z_1 + z_2 = 2 \\ 2z_1 + 2z_2 = 6 \end{cases}$$

é inconsistente.

► **Lema 2.** $Az = b$ é consistente se, e somente se, para todo a , tal que $a'A = 0$, tivermos $a'b = 0$.

Prova. Se o sistema é consistente, então existe z_0 , tal que $Az_0 = b$. Portanto, se $a'A = 0$, $a'Az_0 = 0 \rightarrow a'b = 0$.

Suponhamos, agora, que para todo a , tal que $a'A = 0$, $a'b = 0$. Assim, todo a ortogonal às colunas de A é ortogonal a b . Isso implica que $b \perp \mathcal{E}^\perp(A)$. Como $\mathcal{E}(A)$ é de dimensão finita, $b \in \mathcal{E}(A)$ e existe z_0 , tal que $Az_0 = b$ ◀

Podemos, então, dizer que $Az = b$ é consistente se, e somente se, $b \in \mathcal{E}(A)$.

► **Teorema 1.** As equações normais são consistentes.

Prova. Pelo Lema 2, o sistema (3) será consistente se $a'X'X = 0 \rightarrow a'X'y = 0$, para qualquer y . Verifiquemos:

$$a'X'X = 0 \rightarrow a'X'Xa = 0 \rightarrow a'X' = 0 \rightarrow a'X'y = 0$$

para qualquer y , usando o Lema 1. ◀

► **Teorema 2.** Para qualquer b , solução das equações normais, a forma quadrática $Q(\beta)$ toma seu mínimo, isto é, o método dos mínimos quadrados fornece um mínimo global à soma de quadrados dos desvios do ajuste.

Prova. Seja b uma solução das equações normais.

$$y - X\beta = (y - Xb) + X(b - \beta)$$

Assim,

$$(y - X\beta)'(y - X\beta) = (y - Xb)'(y - Xb) + 2(b - \beta)'X'(y - Xb) + (b - \beta)'X'X(b - \beta)$$

O segundo termo à direita da expressão acima é igual a zero já que $X'y = X'Xb$. O terceiro termo, que pode ser reescrito como $[X(b - \beta)]'[X(b - \beta)]$, é maior que zero. Então,

$$(y - X\beta)'(y - X\beta) > (y - Xb)'(y - Xb)$$

a menos que $X\beta = Xb$. ◀

► Teorema 3. Xb ; onde b é solução não única das equações normais, é único.

Prova. Sejam b_1 e b_2 duas soluções diferentes do sistema (3).

$$X'Xb_1 = X'y$$

$$X'Xb_2 = X'y$$

$$\therefore X'Xb_1 = X'Xb_2 \Rightarrow X'Xb_1 - X'Xb_2 = \emptyset \Rightarrow X'X(b_1 - b_2) = \emptyset \Rightarrow$$

$$\rightarrow (b_1 - b_2)'X'X(b_1 - b_2) = 0 \rightarrow X(b_1 - b_2) = \emptyset \rightarrow Xb_1 = Xb_2 \quad \blacktriangleleft$$

Seja uma matriz B , tal que

$$X'XB = X' \quad (4)$$

Esse sistema é consistente pois tem a forma das equações normais. Usando no lugar de y , as sucessivas colunas da matriz identidade, constata-se tal fato. Temos, então, o seguinte teorema:

- Teorema 4. 1. XB é simétrica,
 2. XB é idempotente,
 3. $KBX = X$ e
 4. XB é única.

- Prova.
1. $(XB)' = B'X' = B'X'XB = (X'XB)'B = XB$.
 2. $(XB)^2 = XBXB = B'X'XB = XB$.
 3. $XBX = B'X'X = (X'XB)' = X$.
 4. Sejam B_1 e B_2 distintas. Então,

$$X'XB_1 = X'$$

$$X'XB_2 = X'$$

$$\therefore X'XB_1 = X'XB_2 \Rightarrow X'XB_1 - X'XB_2 = \emptyset \Rightarrow X'(B_1 - B_2) = \emptyset \Rightarrow$$

$$\rightarrow (B_1 - B_2)'X'(B_1 - B_2) = \emptyset \rightarrow X(B_1 - B_2) = \emptyset \rightarrow XB_1 = XB_2 \quad \blacktriangleleft$$

A equação (4) será frequentemente usada a seguir. B , por ser de grande importância, será sempre referida como tal e não designaremos outra matriz por esse nome. Portanto, daqui para frente, sempre que se mencionar B , estaremos nos referindo à matriz de (4)¹. Veremos no Capítulo 7 que BB' é automaticamente obtida com a aplicação do operador SWEEP(A,k).

Pode-se verificar que o conjunto dos vetores b , solução das equações normais, pode ser escrito como $b_0 + (I - BX)z$, onde $b_0 = By$ e z , um vetor arbitrário de dimensão adequada:

¹Essa matriz satisfaz as propriedades de uma inversa generalizada de X . Além disso, $X'XBB'X'X = X'X$ e, portanto, BB' é uma inversa generalizada de $X'X$.

$$\begin{aligned} X'Xb &= X'XB y + X'X(I-BX)z = X'y + X'(X-GBX)z = \\ &= X'y + X'(X-X)z = X'y \end{aligned}$$

Portanto, $b = b_0 + (I-BX)z$ é solução não única das equações normais.

► Teorema 5. Seja $XB = P_X$. O vetor Xb pode ser representado por $P_X y$.

Prova. $X'Xb = X'y$. Então, para $Xb = P_X y$, temos

$$X'P_X y = (P_X'X)'y = (GBX)'y = X'y. \quad \blacktriangleleft$$

Analisemos o que está ocorrendo do ponto de vista geométrico. O subespaço gerado pelas colunas de X , $\mathcal{E}(X)$, é o mesmo subespaço gerado pelas colunas de P_X .

Para verificar, tomemos um elemento arbitrário $v \in \mathcal{E}(X)$. Assim,

$$v = Xu = GBXu = P_X Xu = P_X w \in \mathcal{E}(P_X) \rightarrow \mathcal{E}(X) \subset \mathcal{E}(P_X)$$

para algum u e $w = Xu$. Por outro lado,

$$v^* = P_X w^* = GBw^* = Xu^* \in \mathcal{E}(X) \rightarrow \mathcal{E}(P_X) \subset \mathcal{E}(X)$$

para algum w^* e $u^* = Bw^*$. Então, $\mathcal{E}(X) = \mathcal{E}(P_X)$.

► Teorema 6. O complemento ortogonal de $\mathcal{E}(X)$ em \mathbb{R}^n é $\mathcal{E}(I - P_X)$.

Prova. Seja $u \in \mathcal{E}(X) \rightarrow u = Xa = XBXa = P_X Xa = P_X z_1$ para algum a e $z_1 = Xa$. Seja $v \in \mathcal{E}(I - P_X) \rightarrow v = (I - P_X)z_2$ para algum z_2 . Então,

$$u'v = z_1' P_X (I - P_X) z_2 = z_1' (P_X - P_X) z_2 = 0 \therefore \mathcal{E}(X) \perp \mathcal{E}(I - P_X) \quad \blacktriangleleft$$

Podemos dizer que $y \in \mathbb{R}^n$ pode ser decomposto unicamente em duas partes: uma pertencente a $\mathcal{E}(X) = \mathcal{E}(P_X)$ e outra pertencente ao complemento ortogonal desse subespaço, $\mathcal{E}(I - P_X)$. Isso pode ser verificado abaixo.

$$y = P_X y - P_X y + y = P_X y + (I - P_X) y = u_1 + u_2 \quad (5)$$

com $u_1 \in \mathcal{E}(P_X)$ e $u_2 \in \mathcal{E}(I - P_X)$. Além disso, a decomposição é única pois, caso contrário, $y \in \mathbb{R}^n$ também poderia ser representado como em (5), através de $s_1 \in \mathcal{E}(P_X)$ e $s_2 \in \mathcal{E}(I - P_X)$ e

$$y = u_1 + u_2 = s_1 + s_2$$

Então,

$$u_1 - s_1 = s_2 - u_2 \quad (6)$$

Como o único vetor comum a $\mathcal{E}(P_X)$ e a $\mathcal{E}(I-P_X)$ é o vetor nulo, de (6) vem que

$$u_1 = s_1 \text{ e } u_2 = s_2.$$

Como $y \in \mathbb{R}^n$ está decomposto unicamente em dois vetores pertencentes a subespaços ortogonais, levando em conta (5), dizemos que $u_1 = P_X y$ é a projeção ortogonal de y em $\mathcal{E}(X)$ e $u_2 = (I-P_X)y$ é a projeção ortogonal em $\mathcal{E}(X)^\perp$. Note que $P_X = XB$, sendo simétrico e idempotente, é um projetor ortogonal² em $\mathcal{E}(X)$. Além disso, $(I-P_X)$ também é simétrico e idempotente e, portanto, é um projetor em $\mathcal{E}(X)^\perp = \mathcal{E}(I-P_X)$. Pode-se escrever $\mathbb{R}^n = \mathcal{E}(P_X) \oplus \mathcal{E}(I-P_X)$, onde o símbolo \oplus denota a soma direta dos dois subespaços.

Definição 3. O vetor $r = (I-P_X)y$ é chamado *vetor de resíduos* do ajuste. O vetor $\hat{y} = P_X y$ é chamado *vetor de valores preditos* pelo ajuste.

O teorema a seguir condensa a questão da aproximação ou ajuste de mínimos quadrados.

²Na literatura, é frequente o uso de inversa generalizada e, assim, o projetor no espaço coluna de X é representado por $X(X'X)^-X'$. Com BB' no lugar de $(X'X)^-$, o projetor será representado por $XBB'X' = XBXB = XB$. Frequentemente a seguir, usaremos $XBB'X'$ para representar o projetor ortogonal, para facilidade do leitor que estiver habituado ao uso de inversa generalizada. Além disso, nos modelos onde a matriz X tem posto completo, caso particular dos modelos lineares, basta entender a matriz BB' como $(X'X)^{-1}$, a inversa de $X'X$. Portanto, apesar da notação condensada XB para o projetor, daremos, muitas vezes, preferência à outra forma.

► **Teorema 7.** O ajuste de mínimos quadrados do modelo $\mathbf{y} = \mathbf{X}\boldsymbol{\beta}$ é a projeção ortogonal de \mathbf{y} em $\mathcal{E}(\mathbf{X})$. Essa projeção será o vetor $\hat{\mathbf{y}} = \mathbf{P}_X \mathbf{y}$ e $\mathbf{y}'\mathbf{y} = \hat{\mathbf{y}}'\hat{\mathbf{y}} + \mathbf{r}'\mathbf{r}$.

Prova. O ajuste e respectivo resíduo são fornecidos por

$$\mathbf{X}\mathbf{b} = \mathbf{P}_X \mathbf{y}$$

$$\mathbf{r} = \mathbf{y} - \mathbf{P}_X \mathbf{y} = (\mathbf{I} - \mathbf{P}_X) \mathbf{y}$$

Então,

$$\mathbf{y} = \mathbf{P}_X \mathbf{y} + (\mathbf{I} - \mathbf{P}_X) \mathbf{y} = \hat{\mathbf{y}} + \mathbf{r}$$

e

$$\begin{aligned} \mathbf{y}'\mathbf{y} &= (\hat{\mathbf{y}} + \mathbf{r})'(\hat{\mathbf{y}} + \mathbf{r}) = \hat{\mathbf{y}}'\hat{\mathbf{y}} + 2\hat{\mathbf{y}}'\mathbf{r} + \mathbf{r}'\mathbf{r} = \hat{\mathbf{y}}'\hat{\mathbf{y}} + 2\mathbf{y}'\mathbf{P}_X(\mathbf{I} - \mathbf{P}_X)\mathbf{y} + \mathbf{r}'\mathbf{r} = \\ &= \hat{\mathbf{y}}'\hat{\mathbf{y}} + \mathbf{r}'\mathbf{r} \end{aligned}$$

Assim, quando desejamos representar $\mathbf{y} \in \mathbb{R}^n$ aproximadamente por $\mathbf{X}\boldsymbol{\beta} \in \mathcal{E}(\mathbf{X})$, através do método dos mínimos quadrados, devemos projetar ortogonalmente \mathbf{y} em $\mathcal{E}(\mathbf{X})$, obtendo $\mathbf{X}\mathbf{b}$ e em $\mathcal{E}(\mathbf{X})^\perp$, obtendo \mathbf{r} . As projeções de \mathbf{y} são únicas apesar de \mathbf{b} não ser. O vetor diferença entre \mathbf{y} e sua representação, $\mathbf{X}\mathbf{b}$, é de mínimo comprimento, segundo os critérios que consideramos. Tais vetores podem ser vistos

na Figura 2.

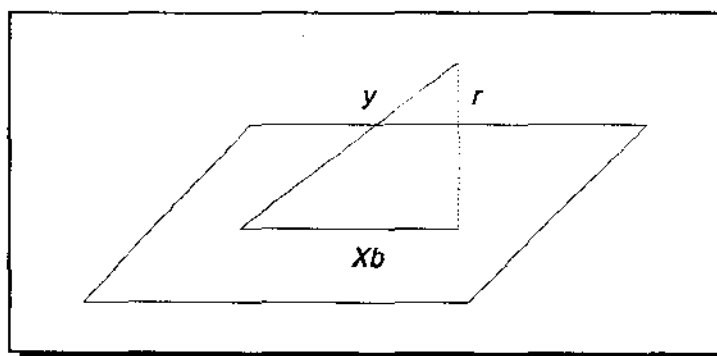


Figura 2. Projeção de y no espaço das colunas de X

A matriz X não é, em si, a determinadora da solução (e do problema). Qualquer que seja a matriz U , tal que $\mathcal{E}(U) = \mathcal{E}(X)$, a resolução de $U'U\hat{\gamma} = U'y$ e $X'X\hat{\beta} = X'y$ fornecerá a mesma aproximação

$$\hat{y} = U\hat{\gamma} = X\hat{\beta}$$

Assim, o modelo $y \doteq X\beta$ pode, de modo equivalente, ser expresso como $y \doteq U\gamma$, se $\mathcal{E}(U) = \mathcal{E}(X)$. Nesse caso, diz-se que os dois modelos são *reparametrizações* de um mesmo modelo. Note que se pode escolher U de posto completo.

Os comprimentos (normas) dos vetores podem ser usados para avaliar a

qualidade do ajuste. Quanto mais próximo y estiver do espaço coluna de X , menor será a norma do resíduo r e, de certa forma, melhor será o ajuste. Assim, é importante calcular as normas de y , Xb e r . Além disso, como Xb e r são ortogonais entre si, pode-se aplicar o teorema de Pitágoras a eles. Para isso, necessitaremos calcular suas normas ao quadrado, o que faremos a seguir.

A norma ao quadrado de y será simplesmente $y'y$. A norma ao quadrado de Xb será

$$(Xb)'Xb = y'XBB'X'XBB'X'y = y'XBB'X'y = y'XBy$$

e a norma ao quadrado de r será

$$r'r = y'(I - XBB'X')y = y'(I - XB)y$$

Assim,

$$y'y = (Xb)'Xb + r'r = y'XBy + y'(I - XB)y \quad (7)$$

Note que os termos à direita de (7) são formas quadráticas em y . Elas serão muito usadas e as denominaremos *somas de quadrados*.

Mais uma vez, é conveniente se chamar a atenção para o fato de que a matriz X nos modelos classificatórios, em geral, tem posto incompleto e a solução das equações normais não é única. No entanto, as somas de quadrados associadas ao

ajuste, $y'XB y$, e ao resíduo, $y'(I - XB)y$, são únicas já que XB é única.

3.4 GRAUS DE LIBERDADE

O vetor $y \in \mathbb{R}^n$ é projetado em um subespaço de \mathbb{R}^n , $\mathcal{E}(X) = \mathcal{E}(P_X)$, de dimensão $k = p(X)$, o posto de X , fornecendo $P_X y$. O vetor de resíduos será $(y - Xb) \in \mathcal{E}(I - P_X)$. Este último subespaço tem dimensão $n - k$, já que $\mathbb{R}^n = \mathcal{E}(P_X) \oplus \mathcal{E}(I - P_X)$.

Definição 4. $p(X)$, a dimensão de $\mathcal{E}(X)$, corresponde aos *graus de liberdade* associados ao ajuste do modelo $y \doteq X\beta$. Os graus de liberdade associados ao resíduo desse ajuste correspondem à dimensão de $\mathcal{E}(I - P_X)$, $n - k$.

Note que

$$X'Xb = X'y - X'P_X y = X'y - X'\hat{y} = X'y$$

Assim, dado um vetor y , o vetor do ajuste, \hat{y} , não varia livremente: ele fica restrito a um subespaço de dimensão k . O mesmo vale para o vetor resíduo: ele fica restrito a um subespaço de dimensão $n - k$. Essa é uma outra forma de se encarar os graus de liberdade.

3.5 AS EQUAÇÕES NORMAIS REDUZIDAS

Se, no modelo $y = X\beta$, particionarmos X em $[X_1 \ X_2]$ e, de modo equivalente, β , teremos

$$y = [X_1 \ X_2] \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = X_1\beta_1 + X_2\beta_2 \quad (8)$$

Partindo de $y = X_1\beta_1$, haverá melhora na aproximação de y se acrescentarmos a X_1 o subconjunto X_2 ? Para responder parte dessa questão na seção 3.7, necessitaremos de resultados que serão desenvolvidos nesta e na próxima seção.

As equações normais particionadas para o modelo (8) serão

$$X_1'X_1\beta_1 + X_1'X_2\beta_2 = X_1'y \quad (9a)$$

$$X_2'X_1\beta_1 + X_2'X_2\beta_2 = X_2'y \quad (9b)$$

A questão, aqui, é encontrar equações normais envolvendo somente β_1 , eliminando β_2 .

Sabemos que existe B_2 , tal que $X_2'X_2B_2 = X_2'$. Multiplicando-se (9b) por $X_1'B_2'$, temos

$$X_1'B_2'X_2'X_1\beta_1 + X_1'B_2'X_2'X_2\beta_2 = X_1'B_2'X_2'y$$

de onde, fazendo $X_2B_2 = P_2$, resulta

$$X_1'P_2X_1\beta_1 + X_1'X_2\beta_2 = X_1'P_2y \quad (10)$$

Subtraindo (10) de (9a), obtém-se

$$X_1'(I - P_2)X_1\beta_1 = X_1'(I - P_2)y \quad (11)$$

um sistema de equações em β_1 somente. Claramente, $I - P_2$ é uma matriz simétrica e idempotente e, assim, (11) tem a estrutura das equações normais e, portanto, é consistente para qualquer y .

Definição 5. O sistema (11) é chamado *sistema das equações normais reduzidas*.

Analogamente, as equações normais reduzidas para β_2 , eliminando β_1 , serão

$$X_2'(I - P_1)X_2\beta_2 = X_2'(I - P_1)y$$

onde $P_1 = X_1B_1$, com B_1 tal que $X_1'X_1B_1 = X_1'$.

Observando as equações normais reduzidas, por exemplo, para β_1 , nota-se que y está sendo projetado não em $\mathcal{E}(X)$, mas na projeção de $\mathcal{E}(X_1)$ em $\mathcal{E}^\perp(X_2)$.

3.6 O MODELO LINEAR A k -PARTES

Discutiremos aqui o caso $k = 2$. Para $k > 2$, a situação é análoga, sendo trivial, a menos de notação, a extensão.

Consideremos o modelo particionado em duas partes de (8). Ajustando, inicialmente, o submodelo $y = X_1\beta_1$, temos

$$\hat{y} = P_1y$$

A soma de quadrados de y será, conseqüentemente, particionada em

$$y'P_1y$$

Cap 3

devida a β_1 e

$$y'(I - P_1)y$$

devida ao resíduo.

Como P_1 é um projetor ortogonal (em $\mathcal{E}(X_1)$), é simétrica, idempotente e $P_1 X_1 = X_1$.

Considerando, agora, o modelo completo, temos

$$P_{12} = [X_1 \ X_2]B$$

P_{12} é, portanto, o projetor ortogonal em $\mathcal{E}(X)$ e, então, é uma matriz simétrica, idempotente e, além disso,

$$P_{12}X = X$$

$$P_{12}X_1 = X_1$$

já que $X_1 \subset X$.

Teremos, em consequência do ajuste do modelo completo, a seguinte partição da soma de quadrados:

$$y'P_{12}y$$

devida a $\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}$ e

$$y'(I - P_{12})y$$

devida ao resíduo.

Como P_1 e P_{12} são ambas simétricas, $P_{12} - P_1$ também será. Além disso,

$$(P_{12} - P_1)^2 = P_{12} - P_1P_{12} - P_{12}P_1 + P_1 \quad (12)$$

E, como $P_{12}P_1 = P_{12}X_1B_1 = X_1B_1 = P_1$ e $P_1 = P_1' = (P_{12}P_1)' = P_1P_{12}$, a expressão (12) se torna

$$(P_{12} - P_1)^2 = P_{12} - P_1$$

de modo que $P_{12} - P_1$ é um projetor ortogonal.

Consideremos, também, os graus de liberdade associados aos ajustes. Como o posto de uma matriz idempotente é igual a seu traço e $\text{tr}(U - V) = \text{tr}(U) - \text{tr}(V)$ para quaisquer matrizes U e V ,

$$\begin{aligned} p(P_{12} - P_1) &= \text{tr}(P_{12} - P_1) = \text{tr}(P_{12}) - \text{tr}(P_1) = p(P_{12}) - p(P_1) = \\ &= p(X) - p(X_1) \end{aligned}$$

Portanto, os graus de liberdade associados ao ajuste do submodelo contendo somente X_1 será igual ao posto dessa matriz; os graus de liberdade associados ao ajuste do modelo contendo o restante das colunas de X será $p(X) - p(X_1)$. Os graus de liberdade associados ao resíduo serão $n - p(X)$.

Para o caso onde $k = 2$, existem duas formas de se particionar a soma de quadrados total, $y'y$, conforme se acrescenta inicialmente β_1 ou β_2 . Em geral, existirão $k!$ possíveis formas de se particionar a soma de quadrados, dependendo da ordem em que se acrescentem os k subconjuntos das colunas de X . Uma situação especial ocorrerá quando as partições de X forem ortogonais: haverá somente uma forma de se quebrar a soma de quadrados total.

3.7 REDUÇÃO NA SOMA DE QUADRADOS: A NOTAÇÃO R(.)

Tomemos, como exemplo, o modelo $y_{ijr} = \mu + A_i + B_j + A * B_{ij}$, $i = 1, \dots, a$, $j = 1, \dots, b$ e $r = 1, \dots, n_{ij}$. a é o número de níveis de A ; b , o número de níveis de B e n_{ij} , o número de replicações na combinação do nível i de A com o nível j de B . Além

disso,

μ é a média geral,

A_i é o i ésimo nível do primeiro critério de classificação,

B_j é o j ésimo nível do segundo critério de classificação e

$A*B_{ij}$ é a ij ésima interação entre os dois critérios de classificação.

Consideremos o ajuste feito de forma sequencial, ou seja, adicionemos ao modelo um subconjunto das colunas de X de cada vez.

1. Ajustar $y_{ijr} \doteq \mu$ e calcular SQ_1 .
2. Ajustar $y_{ijr} \doteq \mu + A_i$ e calcular SQ_2 .
3. Ajustar $y_{ijr} \doteq \mu + A_i + B_j$ e calcular SQ_3 .
4. Ajustar $y_{ijr} \doteq \mu + A_i + B_j + A*B_{ij}$ e calcular SQ_4 .

SQ_t é o valor da soma de quadrados do modelo ajustado no t ésimo passo e que será diminuído da soma de quadrados total. A diferença $SQ_t - SQ_{t-1}$, $t = 2,3,4$ é a quantidade adicional que será retirada da soma de quadrados total, ao se ajustar o modelo do passo t , tendo sido anteriormente ajustado o modelo do passo $t-1$. Assim, por exemplo, após o passo 2, a soma de quadrados total, $y'y$, foi diminuída de uma quantidade igual a SQ_2 e a soma de quadrados do resíduo será igual a $y'y - SQ_2$. Ao ajustar o modelo do passo 3, o valor $SQ_3 - SQ_2$ provocará uma queda adicional na soma de quadrados do resíduo. Isso demonstrará a importância da introdução do efeito

B no modelo $y_{ijr} = \mu + A_i$.

Uma notação conveniente, para o último caso do exemplo acima, será $R(B | \mu, A)$. Ela representa a redução que ocorrerá na soma de quadrados do resíduo ao se acrescentar o efeito B no modelo que contém os efeitos μ e A . Para evitar mal entendidos, vale a pena uma observação sobre a notação $R(\cdot)$. Seja, por exemplo, o modelo $y = 1\beta_1 + X_2\beta_2$ onde 1 representa o efeito da média μ e X_2 , um efeito A . $R(A | \mu)$ será a redução que ocorrerá na soma de quadrados do resíduo com a introdução do efeito A , no modelo que já contém μ . Por vezes, em situações semelhantes, usaremos, por conveniência, $R(\beta_2 | \beta_1)$, cujo significado será o mesmo de $R(A | \mu)$, já que o modelo acima é equivalente a $y = \mu + A$. No exemplo, o uso de $R(\cdot)$ dará origem às seguintes reduções:

1. $R(\mu) = SQ_1 =$ redução na soma de quadrados total após ajustar μ ;
2. $R(A | \mu) = SQ_2 - SQ_1 = R(\mu, A) - R(\mu) =$ redução na soma de quadrados do resíduo após ajustar A no modelo que contém μ .
3. $R(B | \mu, A) = SQ_3 - SQ_2 = R(\mu, A, B) - R(\mu, A) =$ redução na soma de quadrados do resíduo após ajustar B no modelo que contém μ e A ;
4. $R(A * B | \mu, A, B) = SQ_4 - SQ_3 = R(\mu, A, B, A * B) - R(\mu, A, B) =$ redução na soma de quadrados do resíduo após ajustar $A * B$ no modelo que contém μ , A e B .

As somas de quadrados vistas acima são chamadas *somas de quadrados tipo I* para μ , A , B e $A * B$, respectivamente, e serão estudadas em detalhe posteriormente.

Vejamos geometricamente o que está ocorrendo. Sejam X_1 e X_2 , respectivamente, um subconjunto qualquer de X e seu complementar. Para poder representar graficamente os ajustes sucessivos, consideremos que esses subconjuntos sejam vetores. Portanto, X será um plano em \mathbb{R}^3 . Note que X_1 pode ser o subconjunto de X referente, por exemplo, a μ ou μ e A , ou ainda, μ , A e B .

Graficamente, o ajuste $y = x_1\beta_1$ pode ser visto na Figura 3a e o ajuste de $y = x_1\beta_1 + x_2\beta_2$, na Figura 3b. a e b são os vetores resultantes dessas projeções e suas normas ao quadrado serão, respectivamente, $R(\beta_1)$ e $R(\beta_1, \beta_2)$. Ao passarmos do ajuste $y = x_1\beta_1$ para $y = x_1\beta_1 + x_2\beta_2$, a soma de quadrados associada ao modelo será aumentada de $R(\beta_1)$ para $R(\beta_1, \beta_2)$. Esse aumento está representado na Figura 3c pelo vetor c , cuja norma ao quadrado é $R(\beta_1, \beta_2) - R(\beta_1) = R(\beta_2 | \beta_1)$.

Uma situação especial é termos um modelo que contenha todos os efeitos menos um e que gerará uma soma de quadrados SQ_1 . O modelo com todos os efeitos vai gerar uma soma de quadrados SQ_2 . Aqui, também, podemos usar a notação $R(\cdot)$. Assim, por exemplo, se B for o último efeito a entrar no modelo, teremos $R(B | \mu, A) = SQ_1 - SQ_2$. Essas somas de quadrados para cada efeito são chamadas *somas de quadrados tipo II* para aquele efeito. As somas de quadrados tipo II serão também estudadas em detalhe posteriormente.

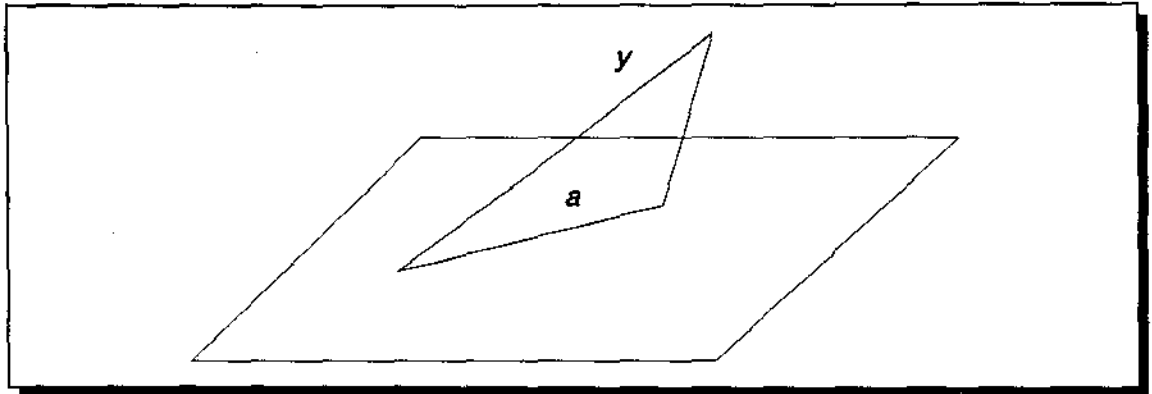


Figura 3a. Ajuste da primeira variável

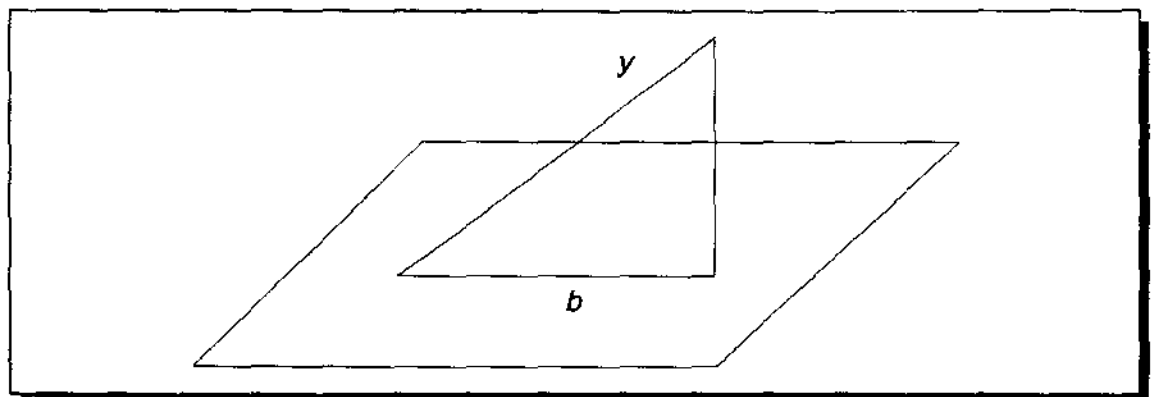


Figura 3b. Acréscimo da segunda variável

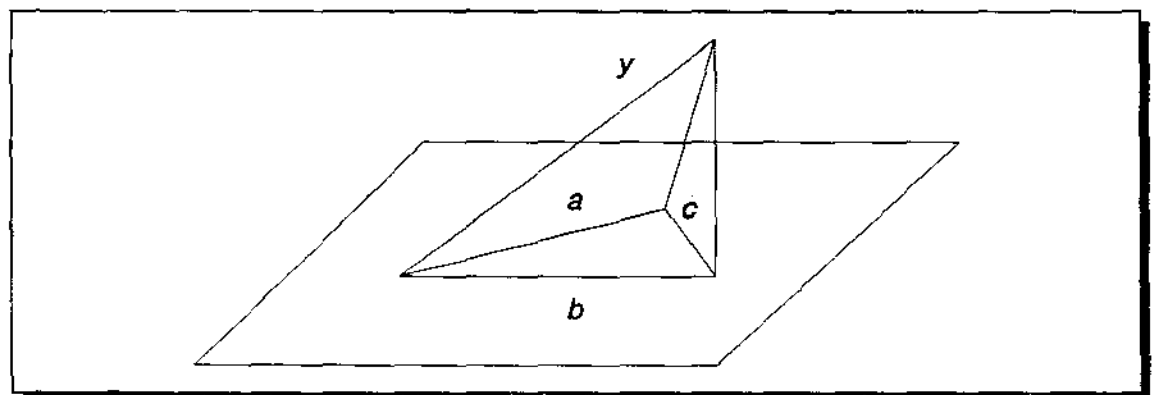


Figura 3c. Redução na soma de quadrados devido ao ajuste da segunda variável dada a primeira

3.8 ANÁLISE DE VARIÂNCIA

Vimos nas seções anteriores que o vetor de dados, y , $n \times 1$, pode ser representado (ou aproximado) por um vetor que é combinação linear dos vetores coluna de alguma matriz. Essa matriz pode ser constituída, por exemplo, apenas por um vetor $\mathbf{1}' = [1 \dots 1]$, o que implica em explicar o vetor de dados somente através do efeito da média. De forma equivalente, a matriz do modelo poderia conter colunas referentes à presença de um ou mais efeitos, associados ou não ao efeito da média. A Figura 4 mostra a aproximação de y por Xb e o vetor resíduo resultante r , para o caso onde temos a média e um efeito x_1 .

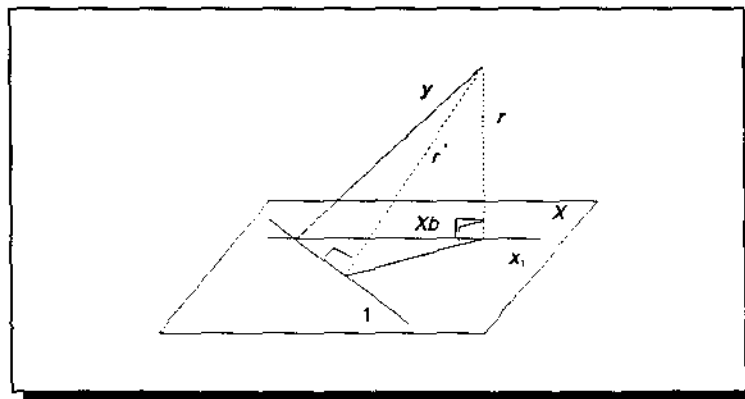


Figura 4. Aproximação de y por Xb

Em geral, inicialmente se projeta y sobre o vetor $\mathbf{1}$, que representa o efeito da média. Esse procedimento leva ao surgimento de um primeiro resíduo mostrado na

figura como r^* . Ele é o vetor *y corrigido pela média*. A seguir, r^* é projetado sobre \tilde{X} , fornecendo, finalmente, Xb e r . Por definição, $\tilde{X} = \left(I - \frac{1}{n}\mathbf{1}\mathbf{1}'\right)X$ é a matriz do modelo, excluída a coluna que representa a média. Quase sempre, é nisso que estamos interessados, já que o efeito da média sempre estará presente (Capítulo 2). O que desejamos, é estudar o comportamento de y apenas através dos efeitos das variáveis representadas nas colunas de X . Assim, a matriz do modelo foi substituída por

$$\tilde{X} = \left(I - \frac{1}{n}\mathbf{1}\mathbf{1}'\right)X$$

O modelo $y = X\beta$ é equivalente a

$$\left(I - \frac{1}{n}\mathbf{1}\mathbf{1}'\right)y = \left(I - \frac{1}{n}\mathbf{1}\mathbf{1}'\right)X\beta$$

ou

$$\tilde{y} = \tilde{X}\beta$$

onde o componente de β referente à média foi igualado a zero. Pode-se verificar que

$I - \frac{1}{n}\mathbf{1}\mathbf{1}'$ é simétrica, idempotente e de posto $n - 1$. Então, a soma de quadrados total,

corrigida pela média, será

$$\bar{y}'\bar{y} = y' \left(I - \frac{1}{n} \mathbf{1}\mathbf{1}' \right) y$$

e o número de graus de liberdade associados à soma de quadrados total corrigida pela média será $n - 1$.

Com base na geometria já estudada e denotando as somas de quadrados devidas ao modelo, ao resíduo, à média, aos tratamentos (excluindo a média) e total, respectivamente por SQM , SQR , SQM_m , SQM_t e SQT , podemos escrever

$$SQT = SQM + SQR$$

ou

$$SQT + SQM_m - SQM_m = SQM_m + SQM - SQM_m + SQR \quad (13)$$

E, assim,

$$\begin{aligned} y'y &= y' \frac{1}{n} \mathbf{1}\mathbf{1}' y + y' P y - y' \frac{1}{n} \mathbf{1}\mathbf{1}' y + y' (I - P) y \\ &= y' \frac{1}{n} \mathbf{1}\mathbf{1}' y + y' \left(P - \frac{1}{n} \mathbf{1}\mathbf{1}' \right) y + y' (I - P) y \end{aligned}$$

A expressão (13) é equivalente a

$$SQT - SQM_m = SQM - SQM_m + SQR$$

onde o membro esquerdo representa a soma de quadrados total corrigida pela média. Então,

$$y'y - y'\frac{1}{n}\mathbf{1}\mathbf{1}'y = y'Py - y'\frac{1}{n}\mathbf{1}\mathbf{1}'y + y'(I-P)y \Rightarrow$$

$$y'\left(I - \frac{1}{n}\mathbf{1}\mathbf{1}'\right)y = y'\left(P - \frac{1}{n}\mathbf{1}\mathbf{1}'\right)y + y'(I-P)y$$

Os graus de liberdade correspondentes a cada soma de quadrados, obviamente, são quebrados de modo equivalente.

A técnica de decomposição de y em vetores mutuamente ortogonais é conhecida como *análise de variância*. É conveniente resumir as informações referentes a essa análise em uma *tabela de análise de variância*:

Fonte de Variação	Graus de Liberdade	Soma de Quadrados	Quadrado Médio
Média	1	$y' \frac{1}{n} \mathbf{11}' y$	$y' \frac{1}{n} \mathbf{11}' y$
Tratamento	k	$y' \left(P - \frac{1}{n} \mathbf{11}' \right) y$	$y' \left(P - \frac{1}{n} \mathbf{11}' \right) y / k$
Resíduo	$n-k-1$	$y' (I-P) y$	$y' (I-P) y / (n-k-1)$
Total	n	$y' y$	

Comumente, a tabela de análise de variância é construída sem a linha referente ao efeito da média:

Fonte de Variação	Graus de Liberdade	Soma de Quadrados	Quadrado Médio
Modelo	k	$y' \left(P - \frac{1}{n} \mathbf{11}' \right) y$	$y' \left(P - \frac{1}{n} \mathbf{11}' \right) y / k$
Resíduo	$n-k-1$	$y' (I-P) y$	$y' (I-P) y / (n-k-1)$
Total	$n-1$	$y' \left(I - \frac{1}{n} \mathbf{11}' \right) y$	

Outras versões da tabela de análise de variância com a quebra da soma de quadrados do modelo em outras parcelas podem ser construídas. Isso será visto quando tratarmos das somas de quadrados tipos I, II, III e IV.

3.9 EXEMPLO

O exemplo a seguir ilustra conceitos introduzidos neste capítulo. Os dados apresentados abaixo são parte dos dados de Vannucchi e colaboradores (1991). O estudo analisa o comportamento dos níveis séricos de aminoácidos em pacientes clinicamente portadores de pelagra. Aqui, apresentamos os valores de valina, em micromoles/ml, de 15 pacientes pelagrosos e de 13 pacientes não pelagrosos. Os dados estão mostrados no Quadro I.

O seguinte programa³ SAS[®] pode ser usado para analisar os possíveis efeitos da doença e do nível de ingestão alcoólica sobre os valores séricos de valina:

```
PROC GLM;
  CLASS PACIENTE INGESTAO;
  MODEL VALINA = PACIENTE INGESTAO / SOLUTION XPX I;
RUN;
```

O resultado da análise encontra-se no Quadro II. O quadro mostra uma tabela de análise de variância. Outras informações presentes na tabela serão vistas nos próximos capítulos. (Os números abaixo se referem aos ítems numerados dos quadros.)

1. Variável dependente, cujo comportamento quer-se explicar.
2. Fontes de variação da variável dependente: o modelo, composto pelas variáveis

³A estrutura dos programas SAS[®] apresenta uma lógica bastante simples. Assim, PROC GLM (*procedure general linear models*) invoca o procedimento que realiza o ajuste; as variáveis classificatórias são declaradas dentro da declaração CLASS; através do comando MODEL, declaram-se as variáveis componentes do modelo: à esquerda a variável dependente e à direita, as independentes, já citadas através de CLASS; SOLUTION, XPX e I são opções do comando MODEL e solicitam, respectivamente, uma solução das equações normais, a matriz $X'X$ e a inversa generalizada $g_2(BB')$.

- PACIENTE e INGESTAO e o resíduo, aqui expresso como 'ERROR'.
3. Graus de liberdade associados a cada uma das fontes de variação. Note o total corrigido pela média, 27, indicando a retirada desse efeito.
 4. Somas de quadrados associadas ao modelo, resíduo e total.

Nos Quadros IIIa e IIIb temos dois conjuntos de valores encontrados para β , mostrando que eles não são únicos. Pode-se verificar que ambos os conjuntos satisfazem as equações normais.

5. Nome do parâmetro.
6. Valor encontrado para cada um dos parâmetros. Em geral, tais valores não são estimativas dos parâmetros e a referência a esse fato se encontra na nota na parte inferior do quadro.

Em termos do modelo visto anteriormente neste capítulo, INTERCEPT corresponde a β_1 , PELAGRA a β_2 e NPELAGRA a β_3 . Nos Quadros IVb e V, encontramos as matrizes $X'X$ e BB'^4 .

⁴Ou a inversa generalizada g_c de $X'X$.

Cap 3

Quadro I. Dados para ajuste

OBS	PACIENTE	INGESTAO	VALINA
1	PELAGRA	MEDIA	0.0616
2	PELAGRA	MEDIA	0.1176
3	PELAGRA	MEDIA	0.1254
4	PELAGRA	BAIXA	0.2173
5	PELAGRA	BAIXA	0.0958
6	PELAGRA	BAIXA	0.1379
7	PELAGRA	BAIXA	0.2072
8	PELAGRA	MEDIA	0.1064
9	PELAGRA	ALTA	0.2843
10	PELAGRA	ALTA	0.1764
11	PELAGRA	ALTA	0.2846
12	PELAGRA	ALTA	0.1751
13	PELAGRA	ALTA	0.2400
14	PELAGRA	ALTA	0.2290
15	PELAGRA	BAIXA	0.1988
16	NPELAGRA	BAIXA	0.2576
17	NPELAGRA	BAIXA	0.2822
18	NPELAGRA	BAIXA	0.2156
19	NPELAGRA	BAIXA	0.3248
20	NPELAGRA	BAIXA	0.2296
21	NPELAGRA	BAIXA	0.2333
22	NPELAGRA	BAIXA	0.2560
23	NPELAGRA	BAIXA	0.2182
24	NPELAGRA	BAIXA	0.2940
25	NPELAGRA	BAIXA	0.2300
26	NPELAGRA	BAIXA	0.2500
27	NPELAGRA	BAIXA	0.2583
28	NPELAGRA	BAIXA	0.3160

Quadro II. Tabela de análise de variância

General Linear Models Procedure					
Dependent Variable: VALINA ¹					
Source ²	DF ³	Sum of Squares ⁴	Mean Square	F Value	Pr > F
Model	3	0.08659641	0.02886547	17.03	0.0001
Error	24	0.04066813	0.00169451		
C Total	27	0.12726454			
R-Square			Root MSE		VALINA Mean
0.680444			0.04116437		0.21510714

Quadro IIIa. Solução das equações normais

General Linear Models Procedure				
Dependent Variable: VALINA				
Parameter ^a	Estimate ^b	T for H0: Parameter=0	Pr > T	Std Error of Estimate
INTERCEPT	0.3190589744 B	11.64	0.0001	0.02741651
PACIENTE PELAGRA	-.0874923077 B	-4.04	0.0005	0.02166213
NPELAGRA	0.0000000000 B			
INGESTAO MEDIA	-.1288166667 B	-4.85	0.0001	0.02657149
BAIXA	-.0601666667 B	-2.41	0.0238	0.02492627
ALTA	0.0000000000 B			

NOTE: The X'X matrix has been found to be singular and a generalized inverse was used to solve the normal equations. Estimates followed by the letter 'B' are biased, and are not unique estimators of the parameters.

Quadro IIIb. Solução das equações normais

General Linear Models Procedure				
Dependent Variable: VALINA				
Parameter	Estimate	T for H0: Parameter=0	Pr > T	Std Error of Estimate
INTERCEPT	0.1027500000 B	4.99	0.0001	0.02741651
PACIENTE PELAGRA	0.0874923077 B	4.04	0.0005	0.02166213
NPELAGRA	0.0000000000 B			
INGESTAO MEDIA	.1288166667 B	4.85	0.0001	0.02657149
BAIXA	.0686500000 B	2.49	0.0203	0.02492627
ALTA	0.0000000000 B			

NOTE: The X'X matrix has been found to be singular and a generalized inverse was used to solve the normal equations. Estimates followed by the letter 'B' are biased, and are not unique estimators of the parameters.

Cap 3

Quadro IVa. A Matriz $X'X$

General Linear Models Procedure		
Matrix Element Representation		
Dependent Variable: VALINA		
Effect		Representation
INTERCEPT		INTERCEPT
PACIENTE	NPELAGRA	DUMMY001
	PELAGRA	DUMMY002
INGESTAO	ALTA	DUMMY003
	BAIXA	DUMMY004
	MEDIA	DUMMY005

Quadro IVb. A Matriz $X'X$

General Linear Models Procedure							
The $X'X$ Matrix							
	INTERCEPT	DUMMY001	DUMMY002	DUMMY003	DUMMY004	DUMMY005	
INTERCEPT	28	13	15	6	18	4	
DUMMY001	13	13	0	0	13	0	
DUMMY002	15	0	15	6	5	4	
DUMMY003	6	0	6	6	0	0	
DUMMY004	18	13	5	0	18	0	
DUMMY005	4	0	4	0	0	4	

Quadro V. A Matriz BB'

General Linear Models Procedure							
$X'X$ Generalized Inverse (g2)							
	INTERCEPT	DUMMY001	DUMMY002	DUMMY003	DUMMY004	DUMMY005	
INTECEPT	0.25	4.839607E-17	0	-0.25	-0.25	0	
DUMMY001	4.839607E-17	0.2769230769	0	-3.39152E-17	-0.2	0	
DUMMY002	0	0	0	0	0	0	
DUMMY003	-0.25	-3.39152E-17	0	0.4166666667	0.25	0	
DUMMY004	-0.25	-0.2	0	0.25	0.45	0	
DUMMY005	0	0	0	0	0	0	

Capítulo 4

Estimabilidade

4.1 INTRODUÇÃO

No modelo $y = X\beta = \beta_1x_1 + \beta_2x_2 + \dots + \beta_px_p$, podemos estar interessados em estudar o valor de um dos β_i , $i=1, \dots, p$, isoladamente. Podemos ainda estar interessados em estudar uma função de um subconjunto deles. Por exemplo: $\beta_k = 2$? ou $\beta_k = 0$? ou ainda $\beta_k - \beta_m = 0$, para quaisquer k, m ? No entanto, é necessário, inicialmente, avaliar a possibilidade de se obter informação nos dados que permita tal estudo. Neste capítulo veremos que, apesar de β não poder ser unicamente determinado, algumas funções lineares de seus componentes podem: é possível encontrar combinações lineares dos elementos do vetor de dados que são estimativas dessas funções. Esse é um fato de grande importância e está diretamente ligado aos testes de significância.

4.2 FUNÇÕES IDENTIFICÁVEIS

Seja um sistema de equações com representação matricial $y = X\beta$

$$\begin{cases} \beta_1 + \beta_2 = 4 \\ \beta_1 + \beta_3 = 6 \end{cases}$$

O que se pode dizer sobre o valor exato de β_1 , β_2 ou de β_3 ? O sistema $y = X\beta$ admite infinitas soluções dadas por $\tilde{\beta} = By + (I - BX)z$, com B como na expressão (4) do Capítulo 3 e z um vetor arbitrário de dimensão adequada. Note que

$$X\tilde{\beta} = XBy + Xz - XBXz = XBX\beta + Xz - Xz = X\beta = y$$

Portanto, $\tilde{\beta}$ é solução do sistema.

Se escolhermos um valor para β_1 , digamos $\beta_1 = 2$, então, $\beta_2 = 4 - \beta_1 = 2$. E, como $\beta_3 = 6 - \beta_1$, $\beta_3 = 4$. O mesmo é válido se tomarmos β_2 ou β_3 no lugar de β_1 . Assim, nenhum dos elementos de β pode ser determinado unicamente: o valor de dois deles dependerá do valor que se atribuir ao terceiro.

E se estivermos interessados no valor de $\beta_3 - \beta_2$? Vejamos:

$$\beta_3 - \beta_2 = (6 - \beta_1) - (4 - \beta_1) = 2$$

Uma função de β , ou uma combinação linear de β , $l'\beta$, com $l' = [0 \ -1 \ 1]$ foi identificada a partir dos dados, apesar de cada elemento isolado de β não ter sido.

Assim, dado o modelo $y = X\beta$, os valores dos componentes do vetor β nem sempre podem ser identificados unicamente. No entanto, algumas funções desses elementos podem.

Temos, então, a seguinte definição:

Definição 1. Seja o modelo $y = X\beta$. Dizemos que uma função linear, $l'\beta$, é *identificável* se, e somente se, dada a equação consistente $X\beta = c$, o valor de $l'\beta$ é determinado unicamente.

Dados o modelo e uma função linear dos parâmetros, necessitamos, portanto, de meios para determinar se tal função é identificável. Isso pode ser determinado a partir do teorema seguinte.

► **Teorema 1.** No modelo $y = X\beta$, a função linear dos parâmetros, $l'\beta$, é identificável se, e somente se, l' pertence a $\mathfrak{R}(X)$, o espaço gerado pelas linhas de X .

Prova. Na equação consistente $X\beta = c$, os valores de β podem ser

representados por $\tilde{\beta} = Bc + (I - BX)z$. Então,

$$l'\tilde{\beta} = l'Bc + l'(I - BX)z$$

Para que $l'\tilde{\beta}$ seja determinado unicamente, já que z é arbitrário, devemos ter

$$l'(I - BX) = 0 \rightarrow l' = l'BX = a'X$$

com $a' = l'B$. Portanto, $l' \in \mathfrak{R}(X)$. Por outro lado, se $l' \in \mathfrak{R}(X)$, $l' = a'X$ para algum a e

$$l'\beta = a'X\beta = a'c$$

Assim, $l'\beta$ estará determinado. ◀

4.3 FUNÇÕES ESTIMÁVEIS

Consideremos, agora, o caso onde o vetor y é o valor observado de um vetor aleatório Y de distribuição conhecida. Tal fato pode ser usado para ampliar o conceito de identificabilidade estudado acima. Vejamos a seguinte definição:

Definição 2. Seja Y um vetor aleatório com distribuição conhecida. Uma função $g(\beta)$ é dita ser *estimável*, se existe $h(Y)$, tal que $E[h(Y)] = g(\beta)$, para qualquer $\beta \in \mathbb{R}^p$.

O que a definição diz, é que se existe a , tal que $E(a'Y) = l'\beta$, $l'\beta$ é estimável.

Apesar da distinção entre o vetor aleatório, Y , e um valor particular que ele possa assumir, y , usaremos y para representar qualquer um dos dois. O leitor, no entanto, deve estar atento para a diferença conceitual envolvida.

Usando a definição de estimabilidade, podemos ir um pouco mais adiante.

O modelo $y \doteq X\beta$ pode também ser representado por $y = X\beta + e$. Note a substituição do símbolo \doteq por $=$. A diferença e , entre y e sua projeção em $\mathcal{E}(X)$, $X\beta$, foi incorporada ao modelo. Além disso, frequentemente é razoável assumir que $E(e) = \emptyset$ e, como consequência, $E(y) = X\beta$. Então,

$$E(a'y) = a'E(y) = a'X\beta = l'\beta \rightarrow a'X = l' \rightarrow l' \in \mathcal{R}(X)$$

como já visto acima para a identificabilidade. Os dois conceitos tem os mesmos fundamentos, divergindo no fato de que a estimabilidade implica no uso de distribuição probabilística dos dados.

Assim, acabamos de discutir o seguinte teorema:

► **Teorema 2.** Seja o modelo $y = X\beta + e$, $E(e) = \emptyset$. Uma função paramétrica $l'\beta$ é linearmente estimável se, e somente se, $l' \in \mathfrak{R}(X)$.

Prova. Se $l'\beta$ é estimável, então

$$E(a'y) = l'\beta \rightarrow a'E(y) = l'\beta \rightarrow a'X\beta = l'\beta \rightarrow a'X = l' \in \mathfrak{R}(X)$$

E, se $l' \in \mathfrak{R}(X) \rightarrow l' = a'X$, para algum a , então,

$$l'\beta = a'X\beta = a'E(y) = E(a'y) \quad \blacktriangleleft$$

Como vimos, uma ou mais combinações lineares dos parâmetros podem ser estimáveis. Seja $\{l'_i\beta\}$, $i = 1, \dots, k$, o conjunto de todas as k funções estimáveis, linearmente independentes, dos parâmetros do modelo $y = X\beta$.

$$\{l'_i\beta\}_{1 \leq i \leq k} = \begin{bmatrix} l'_1\beta \\ \vdots \\ l'_k\beta \end{bmatrix} = \begin{bmatrix} l'_1 \\ \vdots \\ l'_k \end{bmatrix} \beta = L\beta$$

onde L é uma matriz $k \times p$.

No entanto, se o vetor $l'_i \notin \mathfrak{R}(X)$, dizemos que a função $l'_i\beta$ é não estimável.

Isso pode ser formalizado na seguinte definição:

Definição 3. Se $\mathfrak{R}(L) \cap \mathfrak{R}(X) = \emptyset$, as funções $L\beta$ são ditas *não estimáveis*.

Algumas características das funções estimáveis necessitam ser avaliadas.

1. $l'b$ é invariante com respeito a qualquer valor b de β , solução das equações normais. Pode-se notar que

$$l'b = a'Xb = a'XBB'X'y = a'XBXB'y = a'XB'y$$

Como XB é única, $l'b$ também será, para qualquer b .

2. Em qualquer conjunto de funções lineares estimáveis, o número máximo de funções estimáveis linearmente independentes é igual a k , o posto de X . Seja A , de dimensão $n \times n$ e posto completo ($n = p(A)$), tal que $L = AX$. Seja também, $L\beta$ um conjunto de n funções estimáveis. Como $p(L) = p(X)$, existem $p(X)$ linhas linearmente independentes em L e, portanto, $p(X)$ funções linearmente estimáveis.
3. ► **Teorema de Gauss-Markov.** No modelo $y = X\beta + e$, $E(e) = \emptyset$, $\text{var}(e) = \sigma^2 I$, o melhor estimador linear não viciado da função estimável $l'\beta$, $(l'\beta)_b = l'b$, onde b é qualquer solução das equações normais.

Prova. Sabemos que uma função linear estimável, $l'\beta$, é tal que $l' \in \mathfrak{R}(X)$ e, assim,

$$l' = a'X$$

para algum vetor a . Isso leva a

$$l'b = a'Xb = a'P_X y$$

Seja $k'y$ um estimador linear não viciado de $l'\beta$, para um certo vetor k . Como $k'y$ é não viciado,

$$E(k'y) = k'X\beta = l'\beta$$

para qualquer β . Então,

$$k'X = l' = a'X \Rightarrow k'XB = a'XB \Rightarrow k'P_X = a'P_X$$

Pode-se escrever

$$\begin{aligned}k' &= a'P_X + (k' - a'P_X) \\ k'y &= a'P_X y + (k' - a'P_X)y\end{aligned}$$

de modo que

$$\frac{1}{\sigma^2} \text{var}(k'y) = a'P_X a + (k' - a'P_X)(k - P_X a) + 2(k' - a'P_X)P_X a$$

O produto cruzado é igual a zero, pois

$$(k' - a'P_X)P_X = k'P_X - a'P_X$$

Portanto,

$$\frac{1}{\sigma^2} \text{var}(k'y) \geq a'P_X a$$

com igualdade se, e somente se, $k' = a'P_X$, isto é, se $(l'\beta)_b = l'b = a'P_X y$. ◀

Como $l'b$ tem a menor variância que qualquer outro estimador linear não viciado de $l'\beta$, nesse sentido, é o melhor. ◀

As duas propriedades associadas ao erro no Teorema de Gauss-Markov são chamadas *condições de Gauss-Markov*.

4.4 EXEMPLO

Usando os dados do exemplo da seção 3.9, podemos estimar diferenças entre as médias de valina para os três níveis de ingestão alcoólica. Essas diferenças de médias são claramente estimáveis. O seguinte comando¹ SAS*, associado ao programa apresentado anteriormente, fornece a estimativa desejada.

```
ESTIMATE 'BAIXA VS ALTA' INGESTAO I 0 -1;
```

A saída fornecida pelo comando é mostrada no Quadro I.

Quadro I. Estimativa da diferença entre as médias de ingestão alcoólica alta e baixa

General Linear Models				
Parameter	Estimate	T for H0: Parameter=0	Pr > T	Std Error of Estimate
BAIXA VS ALTA	0.06016667	-2.41	0.0238	0.02492627

¹O comando ESTIMATE pede a presença de um rótulo, no caso 'BAIXA VS ALTA', do fator de interesse e dos coeficientes correspondentes aos níveis desse fator. Esses coeficientes são, na realidade, a parte do vetor I referente àquele fator. Os demais componentes do vetor I , referentes aos outros fatores são zero e não necessitam ser explicitados.

Capítulo 5

Mínimos Quadrados em Modelo com Restrições nos Parâmetros

5.1 INTRODUÇÃO

A resolução das equações normais fornece um valor b para o vetor β que minimiza a distância Euclidiana entre y e o espaço gerado pelas colunas da matriz X . Nada mais se exige do vetor solução. No entanto, se sabemos que β está restrito a alguma região do espaço ao qual ele pertence, \mathbb{R}^p , esse fato deve ser levado em conta quando se resolve o sistema das equações normais. Por exemplo, ao estimar o valor dos ângulos internos A , B e C de um triângulo, é preciso considerar que $A + B + C = 180^\circ$.

Os valores b , de β no modelo $y = X\beta$ no qual existem restrições no vetor de parâmetros, também minimizam a distância entre y e um certo espaço. No entanto, como β está restrito a determinada região de seu espaço, essa distância pode ser maior, aqui, do que no caso onde não existe restrição.

5.2 O MODELO RESTRITO

Quer-se, portanto, $\min_{\beta} (y - X\beta)'(y - X\beta)$ com β satisfazendo o sistema consistente $C\beta = c$, onde C é uma matriz $k \times p$ e de posto k . Consideremos o Lagrangeano

$$Q_r(\beta, \lambda) = (y - X\beta)'(y - X\beta) + 2\lambda'(C\beta - c)$$

onde $2\lambda'$ é um vetor de multiplicadores de Lagrange. Desenvolvendo a última expressão, derivando e igualando a zero, teremos

$$Q_r(\beta, \lambda) = y'y - 2X'y\beta + \beta'X'X\beta + 2\lambda'(C\beta - c)$$

$$\frac{\partial Q_r(\beta, \lambda)}{\partial \beta} = -2X'y + 2X'X\beta + 2C'\lambda = \emptyset$$

$$\frac{\partial Q_r(\beta, \lambda)}{\partial \lambda} = 2(C\beta - c) = \emptyset$$

Essas expressões levam ao sistema

$$\begin{cases} X'X\beta + C'\lambda = X'y \\ C\beta = c \end{cases} \quad (1)$$

que, resolvido, fornece a solução procurada. O sistema (1) pode também ser escrito:

$$\begin{bmatrix} X'X & C' \\ C & \emptyset \end{bmatrix} \begin{bmatrix} \beta \\ \lambda \end{bmatrix} = \begin{bmatrix} X'y \\ c \end{bmatrix}$$

As questões que acabamos de ver podem ser formalizadas nas seguintes definições:

Definição 1. O modelo $y = X\beta$, com β sujeito ao sistema consistente $C\beta = c$, é chamado *modelo restrito*.

Definição 2. O sistema (1) é chamado *sistema das equações normais do modelo restrito*.

Devemos levar em conta que as restrições existentes nos parâmetros fazem parte da estrutura do modelo. C tem posto completo: as restrições paramétricas devem ser linearmente independentes, já que não há sentido prático em se colocar duas ou mais restrições equivalentes. O vetor de parâmetros, já sabemos de antemão, está dentro de uma subregião do espaço paramétrico total. Não existindo a restrição, o modelo restrito se transforma no modelo visto anteriormente.

Alguns teoremas serão úteis na compreensão do modelo restrito.

► **Teorema 1.** O sistema (1) é consistente.

Prova. No Lema 2 do Capítulo 3, vimos condições para que um sistema seja

consistente. Assim, um vetor $[v' \ w']$, com

$$[v' \ w'] \begin{bmatrix} X'X & C' \\ C & \emptyset \end{bmatrix} = \emptyset \quad (2)$$

leva a

$$\begin{cases} v'X'X + w'C = \emptyset \\ Cv = \emptyset' \end{cases}$$

Mas

$$\begin{aligned} v'X'X + w'C = \emptyset &\rightarrow v'X'Xv + w'Cv = 0 \rightarrow v'X'Xv = 0 \rightarrow \\ &\rightarrow v'X' = \emptyset \rightarrow v'X'y = 0, \forall y \end{aligned}$$

Como $v'X'X + w'C = \emptyset$ e $v'X' = \emptyset$, então

$$w'C = \emptyset \rightarrow w'C\beta = 0 \rightarrow w'c = 0$$

Dessa forma, $[v' \ w']$ satisfazendo (2), implica em

$$\begin{bmatrix} \mathbf{v}' & \mathbf{w}' \end{bmatrix} \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{c} \end{bmatrix} = \emptyset$$

Portanto, (1) é consistente. ◀

O valor obtido através da resolução de (1) fornece um mínimo global a $Q_r(\beta, \lambda)$ como pode ser verificado a partir do seguinte teorema.

► **Teorema 2.** Qualquer solução, (\mathbf{b}_r, λ) , do sistema (1) fornece um mínimo global a $(\mathbf{y} - \mathbf{X}\beta)'(\mathbf{y} - \mathbf{X}\beta)$ sujeito a $\mathbf{C}\beta = \mathbf{c}$ consistente.

Prova. Suponhamos $\beta = \mathbf{b}$. Então,

$$\mathbf{y} - \mathbf{X}\mathbf{b} = (\mathbf{y} - \mathbf{X}\mathbf{b}_r) + \mathbf{X}(\mathbf{b}_r - \mathbf{b})$$

Assim,

$$\begin{aligned} (\mathbf{y} - \mathbf{X}\mathbf{b})'(\mathbf{y} - \mathbf{X}\mathbf{b}) &= (\mathbf{y} - \mathbf{X}\mathbf{b}_r)'(\mathbf{y} - \mathbf{X}\mathbf{b}_r) + 2(\mathbf{b}_r - \mathbf{b})'\mathbf{X}'(\mathbf{y} - \mathbf{X}\mathbf{b}_r) + \\ &+ (\mathbf{b}_r - \mathbf{b})'\mathbf{X}'\mathbf{X}(\mathbf{b}_r - \mathbf{b}) \end{aligned}$$

Entretanto,

$$(\mathbf{b}_r - \mathbf{b})' \mathbf{X}' (\mathbf{y} - \mathbf{X}\mathbf{b}_r) = (\mathbf{b}_r - \mathbf{b})' \mathbf{C}' \mathbf{l} = \mathbf{l}' \mathbf{C} (\mathbf{b}_r - \mathbf{b}) = 0$$

Portanto,

$$(\mathbf{y} - \mathbf{X}\mathbf{b})' (\mathbf{y} - \mathbf{X}\mathbf{b}) > (\mathbf{y} - \mathbf{X}\mathbf{b}_r)' (\mathbf{y} - \mathbf{X}\mathbf{b}_r)$$

a menos que $\mathbf{X}\mathbf{b} = \mathbf{X}\mathbf{b}_r$, isto é, a menos que \mathbf{b} seja solução das equações normais. ◀

No modelo irrestrito, a projeção de \mathbf{y} em $\mathcal{E}(\mathbf{X})$ é única. Aqui, também, a projeção é única.

► **Teorema 3.** Para qualquer \mathbf{b}_r , solução das equações normais do modelo restrito, $\mathbf{X}\mathbf{b}_r$ é único.

Prova.

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{C}' \\ \mathbf{C} & \emptyset \end{bmatrix} \begin{bmatrix} \mathbf{b}_{r,1} \\ \lambda_1 \end{bmatrix} = \begin{cases} \mathbf{X}'\mathbf{X}\mathbf{b}_{r,1} + \mathbf{C}'\lambda_1 = \mathbf{X}'\mathbf{y} \\ \mathbf{C}\mathbf{b}_{r,1} = \mathbf{c} \end{cases}$$

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{C}' \\ \mathbf{C} & \emptyset \end{bmatrix} \begin{bmatrix} \mathbf{b}_{r,2} \\ \lambda_2 \end{bmatrix} = \begin{cases} \mathbf{X}'\mathbf{X}\mathbf{b}_{r,2} + \mathbf{C}'\lambda_2 = \mathbf{X}'\mathbf{y} \\ \mathbf{C}\mathbf{b}_{r,2} = \mathbf{c} \end{cases}$$

Então,

$$\begin{aligned} & \begin{cases} X'Xb_{r,1} + C'\lambda_1 = X'Xb_{r,2} + C'\lambda_2 \\ Cb_{r,1} = Cb_{r,2} \end{cases} \rightarrow \\ & \Rightarrow \begin{cases} X'Xb_{r,1} - X'Xb_{r,2} + C'\lambda_1 - C'\lambda_2 = \emptyset \\ Cb_{r,1} - Cb_{r,2} = \emptyset \end{cases} \rightarrow \\ & \rightarrow \begin{cases} X'X(b_{r,1} - b_{r,2}) + C'(\lambda_1 - \lambda_2) = \emptyset \\ C(b_{r,1} - b_{r,2}) = \emptyset \end{cases} \rightarrow \\ & \rightarrow \begin{cases} (b_{r,1} - b_{r,2})'X'X(b_{r,1} - b_{r,2}) + (b_{r,1} - b_{r,2})'C'(\lambda_1 - \lambda_2) = 0 \\ (b_{r,1} - b_{r,2})'C' = \emptyset \end{cases} \rightarrow \\ & \Rightarrow (b_{r,1} - b_{r,2})'X'X(b_{r,1} - b_{r,2}) = 0 \Leftrightarrow X(b_{r,1} - b_{r,2}) = \emptyset \Rightarrow Xb_{r,1} = Xb_{r,2} \end{aligned}$$

e o teorema está provado. ◀

Um fato, importante por si mesmo e que será utilizado posteriormente, é a dimensão dos espaços onde estão os vetores ajustado e resíduo do modelo restrito. Essas dimensões são determinadas através do próximo teorema.

► **Teorema 4.** O número de graus de liberdade associados ao ajuste do modelo

$y \doteq X\beta$, sujeito ao sistema consistente $C\beta = c$, é igual a $p\left(\begin{bmatrix} X \\ C \end{bmatrix}\right) - p(C)$. O

número de graus de liberdade associados ao resíduo será $n - p\left(\begin{bmatrix} X \\ C \end{bmatrix}\right) + p(C)$.

Prova. O conjunto das soluções de um sistema consistente do tipo $Qv = t$ será dado por $v = Q^-t + (I - Q^-Q)z$, onde z é um vetor arbitrário e Q^- uma matriz tal que $Q'QQ^- = Q'$ e, portanto, $QQ^-Q = Q$ é uma expressão válida. Note que Q^- é equivalente à matriz B de $X'XB = X'$, vista anteriormente.

Seja

$$A = \begin{bmatrix} X'X & C' \\ C & \emptyset \end{bmatrix}$$

As soluções do sistema (1) serão dadas por

$$\begin{bmatrix} b_r \\ \tilde{\lambda} \end{bmatrix} = A^{-1} \begin{bmatrix} X'y \\ c \end{bmatrix} + (I - A^{-1}A)z$$

Consideremos

$$A^- = \begin{bmatrix} D & E \\ E' & F \end{bmatrix}$$

Como A é simétrica, A^- também é. Suponhamos que não seja. De qualquer modo $A \left[\frac{1}{2}(A^-)' + \frac{1}{2}A^- \right] A = A$ e, portanto, $\frac{1}{2}[(A^-)' + A^-]$ é simétrica.

Teremos, então,

$$\begin{bmatrix} b_r \\ \tilde{\lambda} \end{bmatrix} = \begin{bmatrix} DX'y + Ec \\ E'X'y + Fc \end{bmatrix} + (I - A^-A)z$$

e o vetor ajustado será

$$Xb_r = XDX'y + XEc$$

O subespaço no qual ele está contido é $\mathcal{E}(XDX')$. Portanto, necessitamos determinar $p(XDX')$. Para isso, consideremos:

$$1. \mathfrak{R} \left(\begin{bmatrix} X'X \\ C \end{bmatrix} \right) = \mathfrak{R} \left(\begin{bmatrix} X \\ C \end{bmatrix} \right) \tag{3}$$

pois $\mathfrak{R}(X'X) = \mathfrak{R}(X)$.

$$2. \mathcal{E}\left(\begin{bmatrix} X'X \\ C \end{bmatrix}\right) \cap \mathcal{E}\left(\begin{bmatrix} C' \\ \emptyset \end{bmatrix}\right) = \emptyset \quad (4)$$

A validade de (4) é verificada a seguir. Determinemos a interseção dos dois subespaços.

$$\begin{aligned} \begin{bmatrix} X'X \\ C \end{bmatrix} v = \begin{bmatrix} C' \\ \emptyset \end{bmatrix} \delta &\rightarrow \begin{cases} X'Xv = C'\delta \\ Cv = \emptyset \end{cases} \rightarrow \begin{cases} v'X'Xv = v'C'\delta \\ v'C' = \emptyset' \end{cases} \\ &\rightarrow \begin{cases} v'X'Xv = 0 \\ Cv = \emptyset \end{cases} \rightarrow \begin{bmatrix} v'X'X \\ C \end{bmatrix} v = \begin{bmatrix} 0 \\ \emptyset \end{bmatrix} \end{aligned}$$

Portanto, a interseção é o vetor nulo.

3. Como decorrência de (4),

$$\mathcal{E}\left(\begin{bmatrix} X'X & C' \\ C & \emptyset \end{bmatrix}\right) = \mathcal{E}\left(\begin{bmatrix} X'X \\ C \end{bmatrix}\right) \cup \mathcal{E}\left(\begin{bmatrix} C' \\ \emptyset \end{bmatrix}\right) \quad (5)$$

4. Usando (3) e (5),

$$p(A) = p\left(\begin{bmatrix} X'X & C' \\ C & \emptyset \end{bmatrix}\right) = p\left(\begin{bmatrix} X'X \\ C \end{bmatrix}\right) + p(C') = p\left(\begin{bmatrix} X \\ C \end{bmatrix}\right) + p(C)$$

$$5. \text{CEC} = \text{C} \quad (6)$$

Para verificar (6), consideremos inicialmente o sistema

$$\begin{bmatrix} X'X & C' \\ C & \emptyset \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \tilde{\lambda} \end{bmatrix} = \begin{bmatrix} X'y \\ c = C\beta_0 \end{bmatrix}$$

que é consistente para qualquer $y \in \mathbb{R}^n$. Em particular, tomemos como y o vetor nulo e, como β_0 , as colunas da matriz identidade, sucessivamente. Tal processo leva à equação

$$\begin{bmatrix} X'X & C' \\ C & \emptyset \end{bmatrix} \begin{bmatrix} B \\ \tilde{\Lambda} \end{bmatrix} = \begin{bmatrix} \emptyset \\ C \end{bmatrix}$$

ou

$$\begin{cases} X'XB + C'\tilde{\Lambda} = \emptyset \Rightarrow B'X'X + \tilde{\Lambda}'C = \emptyset' \Rightarrow B'X'X = -\tilde{\Lambda}'C \\ CB = C \Rightarrow B'C' = C' \end{cases} \quad (7)$$

Premultiplicando $AA^{-1}A = A$ por $[B' \ \tilde{\Lambda}']$ temos

$$\begin{aligned}
& [B' \quad \tilde{A}'] \begin{bmatrix} X'X & C' \\ C & \emptyset \end{bmatrix} \begin{bmatrix} D & E \\ E' & F \end{bmatrix} \begin{bmatrix} X'X & C' \\ C & \emptyset \end{bmatrix} = [B' \quad \tilde{A}'] \begin{bmatrix} X'X & C' \\ C & \emptyset \end{bmatrix} \rightarrow \\
& \rightarrow [B'X'X + \tilde{A}'C \quad B'C'] \begin{bmatrix} D & E \\ E' & F \end{bmatrix} \begin{bmatrix} X'X & C' \\ C & \emptyset \end{bmatrix} = [B'X'X + \tilde{A}'C \quad B'C'] \rightarrow \\
& \rightarrow [\emptyset \quad C'] \begin{bmatrix} D & E \\ E' & F \end{bmatrix} \begin{bmatrix} X'X & C' \\ C & \emptyset \end{bmatrix} = [\emptyset \quad C'] \rightarrow \\
& \rightarrow [C'E' \quad C'F] \begin{bmatrix} X'X & C' \\ C & \emptyset \end{bmatrix} = [\emptyset \quad C'] \rightarrow \\
& \rightarrow \begin{cases} C'E'X'X + C'FC = \emptyset \\ C'E'C' = C' \rightarrow CEC = C \end{cases} \tag{8}
\end{aligned}$$

6. $p(C) = p(CEC) \leq p(CE) \leq p(C)$. Portanto, $p(CE) = p(C)$. Por (6), CE é um projetor e, portanto, uma matriz simétrica e idempotente. Então,

$$p(CE) = \text{tr}(CE) = \text{tr}(E'C') = \text{tr}(C'E') = p(C)$$

7. $XD'X'$ é simétrica e idempotente. Para verificar tal fato, tomemos $AA'A = A$. Pré e posmultiplicando essa igualdade, respectivamente por

$$[B' \quad \tilde{A}'] \text{ e } \begin{bmatrix} B \\ \tilde{A} \end{bmatrix}, \text{ temos}$$

$$\begin{bmatrix} \emptyset & C' \end{bmatrix} \begin{bmatrix} D & E \\ E' & F \end{bmatrix} \begin{bmatrix} \emptyset \\ C \end{bmatrix} = C'\bar{\Lambda} \rightarrow C'FC = C'\bar{\Lambda}$$

Como F é simétrica, transpondo a última igualdade obteremos

$$C'FC = \bar{\Lambda}'C = -BX'X$$

em decorrência de (7).

Também de $AA^{-1}A = A$, vem que

$$X'XDX'X + C'E'X'X + X'XEC + C'FC = X'X \quad (9)$$

De (8), $C'E'X'X = X'XEC = -C'FC$. Então, (9) se torna

$$X'XDX'X - C'FC = X'X$$

que, posmultiplicando por $DX'X$, fornece

$$X'XDX'XDX'X - C'FCDX'X = X'XDX'X$$

Ainda de $AA^{-1}A = A$, vem que $CDX'X + CEC = C$ ou, em razão de (6),

$$CDX'X + C = C \rightarrow CDX'X = \emptyset$$

e, portanto,

$$X'XDX'XDX'X = X'XDX'X \rightarrow (X'XDX'XD - X'XD)X'X = \emptyset \quad (10)$$

No entanto, para alguma matriz S , $SX'X = \emptyset \rightarrow SX' = \emptyset$, pois considerando $X'X\tilde{B} = X'$, então

$$SX'X = \emptyset \rightarrow S'X'X\tilde{B} = \emptyset \rightarrow SX' = \emptyset$$

A implicação no sentido contrário é óbvia. Por essa razão, (10) implica em

$$\begin{aligned} (X'XDX'XD - X'XD)X' &= \emptyset \rightarrow X'X(DX'XDX' - DX') = \emptyset \rightarrow \\ &\rightarrow X(DX'XDX' - DX') = \emptyset \rightarrow \\ &\rightarrow (XDX')(XDX') = XDX' \end{aligned}$$

A partir dos resultados acima, continuemos. Seja $T = AA^-$. T é uma matriz idempotente, pois $TT = AA^-AA^- = AA^- = T$. Então, $p(T) = \text{tr}(T)$. Mas

$$T = \begin{bmatrix} X'X & C' \\ C & \emptyset \end{bmatrix} \begin{bmatrix} D & E \\ E' & F \end{bmatrix} = \begin{bmatrix} X'XD + C'E' & X'XE + C'F \\ CD & CE \end{bmatrix}$$

Agora, $p(T) = p(A)$, pois

$$p(A) = p(AA^{-}A) \leq p(AA^{-}) = p(T) = p(AA^{-}) \leq p(A)$$

E

$$\text{tr}(T) = \text{tr}(X'XD + C'E') + \text{tr}(CE) = \text{tr}(X'XD) + \text{tr}(C'E') + \text{tr}(CE)$$

Usando os resultados acima, podemos dizer que

$$\begin{aligned} \text{tr}(T) &= p(T) = p(A) = p\left(\begin{bmatrix} X \\ C \end{bmatrix}\right) + p(C) = \\ &= \text{tr}(XDX') + \text{tr}(C'E') + \text{tr}(CE) = p(XDX') + p(C) + p(C) \Rightarrow \\ &\Rightarrow p(XDX') = p\left(\begin{bmatrix} X \\ C \end{bmatrix}\right) - p(C) \end{aligned}$$

que é o número de graus de liberdade associados ao ajuste do modelo restrito.

Assim, o número de graus de liberdade associados ao resíduo será

$$n - p\left(\begin{bmatrix} X \\ C \end{bmatrix}\right) + p(C).$$

5.3 RESTRIÇÕES ENVOLVENDO FUNÇÕES NÃO ESTIMÁVEIS

As consequências de impor, como restrições em um modelo, funções paramétricas não estimáveis, são decorrentes do seguinte teorema:

► **Teorema 5.** Se as condições $C\beta = c$ são conjuntamente não estimáveis no modelo $y = X\beta$, então qualquer solução das equações normais para o modelo restrito satisfaz as equações normais do modelo irrestrito.

Prova. Seja b_r uma solução qualquer das equações normais do modelo restrito. Então,

$$\begin{aligned} X'Xb_r + C'\lambda &= X'y \Rightarrow X'y - X'Xb_r = C'\lambda \Rightarrow \\ &\Rightarrow X'(y - Xb_r) = C'\lambda \Rightarrow \lambda'C = (y - Xb_r)'X \end{aligned}$$

O lado esquerdo da última igualdade acima é um vetor pertencente ao subespaço gerado pelas linhas de C , $\mathfrak{R}(C)$, enquanto que o vetor do membro direito pertence ao subespaço gerado pelas linhas de X , $\mathfrak{R}(X)$. Mas, como $C\beta$ é um conjunto de funções não estimáveis, então cada vetor linha c' de C não pertence a $\mathfrak{R}(X)$. Assim, $\mathfrak{R}(X) \cap \mathfrak{R}(C) = \emptyset$ e

$$\emptyset = \lambda'C = (y - Xb_r)'X \Rightarrow X'Xb_r = X'y$$

Portanto, o mesmo \hat{b} , que é solução das equações normais do modelo restrito, também é solução das equações normais do modelo irrestrito. ◀

Se as condições impostas aos parâmetros são conjuntamente estimáveis, então $C \in \mathfrak{R}(X)$ e as soluções para os dois modelos não são as mesmas.

A colocação de restrições não estimáveis no modelo $y = X\beta$ tem algumas consequências. Como as soluções das equações normais para os modelos com e sem restrição são as mesmas, as somas de quadrados associadas também serão as mesmas. As funções estimáveis no modelo irrestrito também serão estimáveis no modelo restrito. Isso porque, com a imposição de restrições, estaremos restringindo o espaço paramétrico. Consequentemente, estaremos deixando de lado algumas funções estimáveis que não satisfazem as restrições. Porém, as funções que permaneceram pertencem ao conjunto inicial das funções que eram estimáveis antes das restrições serem impostas.

Funções estimáveis no modelo restrito poderão deixar de ser estimáveis no modelo irrestrito. Por exemplo, seja a função $\mu + 1/2(A_1 + A_2)$ estimável no modelo irrestrito. A colocação da restrição $A_2 = 0$, torna a função igual a $\mu + 1/2A_1$ que, embora estimável no modelo restrito, não o é no modelo irrestrito. Por outro lado, funções não estimáveis no modelo irrestrito podem ser estimáveis no modelo restrito.

5.4 RESTRIÇÕES ENVOLVENDO FUNÇÕES ESTIMÁVEIS

Já foi visto que as soluções para β serão encontradas resolvendo o sistema (1).

Consideremos, então:

1. o vetor dos valores ajustados $\hat{y} = Xb$.
2. b , uma solução das equações normais do modelo irrestrito, com $b = BB'X'y$.

Note que b satisfaz as equações normais, pois

$$X'XBB'X'y = X'XBXBy = X'XBy = X'y$$

3. Como o conjunto de restrições $H_0: C\beta = c$ é estimável, $C \in \mathfrak{R}(X)$, de modo que $C = WX$ para alguma matriz W .
4. C tem posto completo, de forma que $(CBB'C')^{-1}$ existe. Isso pode ser verificado da seguinte forma: C tem dimensão $s \times p$, $p(C) = s$ e $C = WX = ZX'X$, para alguma matriz Z . Agora,

$$s = p(C) = p(ZX'X) \leq p(Z) \leq s \quad (11)$$

já que Z tem dimensão $s \times p$. Então, $p(Z) = s$. Colocando $p(ZX')$ no lugar de $p(Z)$ em (11), teremos $p(ZX') = s$. Como

$$CBB'C' = ZX'XBB'X'XZ' = ZX'XZ'$$

e

$$p(CBB'C') = p(ZX') = s$$

que é a ordem de $CBB'C'$, ela é não singular e, portanto, a inversa existe.

Seja a primeira equação de (1).

$$\begin{aligned} X'Xb_r + C'\lambda &= X'y \Rightarrow BB'X'Xb_r + BB'C'\lambda = BB'X'y \Rightarrow \\ &\Rightarrow BB'X'Xb_r = BB'X'y - BB'C'\lambda \Rightarrow \\ &\Rightarrow BB'X'\hat{y} = b - BB'C'\lambda \Rightarrow \\ &\Rightarrow b_r = b - BB'C'\lambda \end{aligned}$$

Substituindo o valor de b_r na segunda equação de (1), teremos

$$\begin{aligned} C(b - BB'C'\lambda) &= c \Rightarrow Cb - CBB'C'\lambda = c \Rightarrow CBB'C'\lambda = Cb - c \Rightarrow \\ &\Rightarrow \lambda = (CBB'C')^{-1}(Cb - c) \end{aligned}$$

Assim, a solução do sistema (1) será

$$\mathbf{b}_r = \mathbf{b} - \mathbf{B}\mathbf{B}'\mathbf{C}'(\mathbf{C}\mathbf{B}\mathbf{B}'\mathbf{C}')^{-1}(\mathbf{C}\mathbf{b} - \mathbf{c})$$

que corresponde ao vetor solução das equações do modelo irrestrito, menos uma certa quantidade.

O ajuste do modelo leva diretamente à necessidade do cálculo da soma de quadrados do resíduo, a fim de se avaliar a qualidade do ajuste. Obtido \mathbf{b}_r , calculemos o valor de SQR_r , a soma de quadrados do resíduo do modelo restrito. Sejam $\mathbf{m} = \mathbf{B}\mathbf{B}'\mathbf{C}'(\mathbf{C}\mathbf{B}\mathbf{B}'\mathbf{C}')^{-1}(\mathbf{C}\mathbf{b} - \mathbf{c})$ e SQR a soma de quadrados do resíduo do modelo irrestrito.

$$\begin{aligned} SQR_r &= (\mathbf{y} - \mathbf{X}\mathbf{b}_r)'(\mathbf{y} - \mathbf{X}\mathbf{b}_r) = \mathbf{y}'\mathbf{y} - 2\mathbf{y}'\mathbf{X}\mathbf{b}_r + \mathbf{b}_r'\mathbf{X}'\mathbf{X}\mathbf{b}_r = \\ &= \mathbf{y}'\mathbf{y} - 2\mathbf{y}'\mathbf{X}(\mathbf{b} - \mathbf{m}) + (\mathbf{b} - \mathbf{m})'\mathbf{X}'\mathbf{X}(\mathbf{b} - \mathbf{m}) = \\ &= \mathbf{y}'\mathbf{y} - 2\mathbf{y}'\mathbf{X}\mathbf{b} + \mathbf{b}'\mathbf{X}'\mathbf{X}\mathbf{b} + 2\mathbf{y}'\mathbf{X}\mathbf{m} - 2\mathbf{b}'\mathbf{X}'\mathbf{X}\mathbf{m} + \mathbf{m}'\mathbf{X}'\mathbf{X}\mathbf{m} = \\ &= SQR + 2\mathbf{y}'\mathbf{X}\mathbf{m} - 2\mathbf{b}'\mathbf{X}'\mathbf{X}\mathbf{m} + \mathbf{m}'\mathbf{X}'\mathbf{X}\mathbf{m} \end{aligned}$$

O desenvolvimento dos três últimos termos da última equação fornece

$$\begin{aligned} 2\mathbf{y}'\mathbf{X}\mathbf{m} &= 2\mathbf{b}'\mathbf{C}'(\mathbf{C}\mathbf{B}\mathbf{B}'\mathbf{C}')^{-1}(\mathbf{C}\mathbf{b} - \mathbf{c}) \\ -2\mathbf{b}'\mathbf{X}'\mathbf{X}\mathbf{m} &= -2\mathbf{b}'\mathbf{C}'(\mathbf{C}\mathbf{B}\mathbf{B}'\mathbf{C}')^{-1}(\mathbf{C}\mathbf{B} - \mathbf{c}) \\ \mathbf{m}'\mathbf{X}'\mathbf{X}\mathbf{m} &= (\mathbf{C}\mathbf{b} - \mathbf{c})'(\mathbf{C}\mathbf{B}\mathbf{B}'\mathbf{C}')^{-1}(\mathbf{C}\mathbf{b} - \mathbf{c}) \end{aligned}$$

Assim,

$$SQR_r = SQR + (Cb - c)'(CBB'C')^{-1}(Cb - c)$$

Portanto, a imposição de restrições estimáveis no modelo provoca um aumento na soma de quadrados do resíduo do modelo irrestrito (note que o segundo termo do segundo membro da equação acima é sempre positivo, já que é uma forma quadrática).

A questão de impor restrições nos parâmetros faz com que não mais projetemos y sobre $\mathcal{E}(X)$, mas sobre $\mathcal{E}\left(\begin{bmatrix} X \\ C \end{bmatrix}\right)$. A norma do resíduo da projeção de y sobre esse subespaço tende a ser maior. Somente não o é, em condições especiais: quando o vetor resultante da projeção for o mesmo nos dois casos, ou seja $Xb = Xb_r$.

A imposição das restrições estimáveis $H_0: C\beta = c$ ao modelo provoca alteração no vetor solução das equações normais e nas somas de quadrados associadas. No entanto, isso não afeta a estimabilidade de funções que são estimáveis no modelo irrestrito: funções estimáveis no modelo irrestrito, também serão estimáveis no modelo restrito. Porém, como tais funções são funções paramétricas, elas estarão sujeitas às restrições impostas aos parâmetros.

Resumindo, vimos que se $l'\beta$ é estimável, então $l' \in \mathfrak{R}(X)$. Mas isso é equivalente a $l \in \mathcal{E}(X')$ ou $l \in \mathcal{E}(X'X)$. E, $l = X'Xt$ para algum t . No modelo

restrito, a matriz do modelo será $\begin{bmatrix} \mathbf{X} \\ \mathbf{C} \end{bmatrix}$ e o vetor de parâmetros, $\begin{bmatrix} \beta \\ \lambda \end{bmatrix}$. Então, $\mathbf{l}'\beta$ deve

ser escrito como

$$\mathbf{l}'\beta = \left(\begin{bmatrix} \mathbf{X}' & \mathbf{C}' \end{bmatrix} \begin{bmatrix} \mathbf{X} \\ \mathbf{C} \end{bmatrix} \begin{bmatrix} t_1 \\ t_2 \end{bmatrix} \right)' \begin{bmatrix} \beta \\ \lambda \end{bmatrix} = t_1'(\mathbf{X}'\mathbf{X} + \mathbf{C}'\mathbf{C})\beta + t_2'(\mathbf{X}'\mathbf{X} + \mathbf{C}'\mathbf{C})\lambda$$

Para que $\mathbf{l}'\beta$ seja função somente de β , o segundo termo à direita da última igualdade deve ser zero. Por hipótese, sabemos que $t_1'(\mathbf{X}'\mathbf{X} + \mathbf{C}'\mathbf{C})' \in \mathfrak{R}\left(\begin{bmatrix} \mathbf{X} \\ \mathbf{C} \end{bmatrix}\right)$. Se as restrições

$H_0: \mathbf{C}\beta = \mathbf{c}$ são estimáveis, $\mathbf{C} \in \mathfrak{R}(\mathbf{X})$, $t_1'\mathbf{C}'\mathbf{C} \in \mathfrak{R}(\mathbf{X})$ e $t_1'\mathbf{C}'\mathbf{C}\beta$ é estimável.

Certamente, $t_1'\mathbf{X}'\mathbf{X}\beta$ também é estimável. Como a soma de duas funções estimáveis é estimável, a imposição de restrições estimáveis torna uma função estimável no modelo restrito, também estimável no modelo irrestrito.

5.5 EXEMPLO

Continuando com o exemplo da seção 3.9, podemos ajustar um modelo àqueles dados, considerando a restrição H_0 : o valor sérico de valina é o mesmo para os níveis

alto e médio de ingestão alcoólica. A diferença entre as médias de valina nos níveis baixo e alto de ingestão era estimável anteriormente e continua estimável no modelo restrito. Para ajustar esse modelo restrito, o conjunto de dados apresentados no Quadro I do Capítulo 3 foi alterado de modo que os valores de INGESTÃO correspondentes a 'MEDIA' foram substituídos por 'ALTA'. No Quadro I, temos a tabela de análise de variância referente ao ajuste desse modelo restrito, realizado pelo seguinte programa SAS*:

```
PROC GLM;  
  CLASS INGESTAO;  
  MODEL VALINA = INGESTAO;  
RUN;
```

No Quadro II, temos a tabela de análise de variância referente ao ajuste do modelo irrestrito pelo mesmo programa SAS*.

Observando os Quadros I e II, pode-se notar que a presença da restrição, além de provocar queda no número de graus de liberdade associados ao ajuste do modelo (1, contra 2 para o modelo irrestrito), causou alteração no valor da soma de quadrados do resíduo. A soma de quadrados do resíduo para o modelo irrestrito foi $SQR = 0.0683$. Com a colocação da restrição, a parcela $(Cb - c)'(CBB'C')^{-1}(Cb - c) = 0.0569$ foi acrescentada àquela soma de quadrados e obteve-se $SQR_r = 0.0683 + 0.0569 = 0.1252$.

Quadro I. Tabela de análise de variância para ajuste de modelo restrito

General Linear Models Procedure					
Dependent Variable: VALINA					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	0.00206881	0.00206881	0.43	0.5179
Error	26	0.12519573	0.00481522		
C. Total	27	0.12726454			
	R-Square		Root MSE		VALINA Mean
	0.016256		0.069392		0.21510714

Quadro II. Tabela de análise de variância para ajuste de modelo irrestrito

General Linear Models Procedure					
Dependent Variable: VALINA					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	0.05895370	0.02947685	10.49	0.0004
Error	25	0.06831084	0.00273243		
C. Total	27	0.12726454			
	R-Square		Root MSE		VALINA Mean
	0.463237		0.052273		0.21510714

Capítulo 6

Mínimos Quadrados Ponderados

6.1 INTRODUÇÃO

Ao estudar o ajuste do modelo $y = X\beta$ pelo método dos mínimos quadrados, consideramos como medida de distância, a grandeza $Q(\beta) = (y - X\beta)'(y - X\beta)$ que desejávamos ser mínima. $Q(\beta)$ é uma forma quadrática e sua matriz, a identidade. E, também, $Q(\beta) \geq 0$ como se desejaria de uma medida de distância.

Consideremos, agora, como medida de distância, uma forma quadrática semelhante a $Q(\beta)$, mas cuja matriz não é a identidade e sim, uma matriz simétrica qualquer $W \neq I$. Assim, teremos

$$Q^*(\beta) = (y - X\beta)'W(y - X\beta)$$

Obviamente, precisamos ter $Q^*(\beta) \geq 0$ para qualquer $(y - X\beta)$ e $Q^*(\beta) = 0$ se, e somente se, $y - X\beta = \emptyset$. Portanto, W deve ser positiva definida.

6.2 PROJEÇÃO NÃO ORTOGONAL (OBLÍQUA)

Para obter $\min_{\beta} Q^*(\beta)$, derivemos e igualemos a zero a expressão

$$Q^*(\beta) = (y - X\beta)'W(y - X\beta) = y'Wy - 2\beta'X'Wy + \beta'X'WX\beta$$

Teremos, então,

$$\frac{\partial Q^*(\beta)}{\partial \beta} = -2X'Wy + 2X'WX\beta = 0 \Rightarrow X'WX\beta = X'Wy \quad (1)$$

As equações do sistema (1) são chamadas *equações de Aitken*. Como W é positiva definida, existe T não singular, tal que $W = TT'$ e (1) pode ser escrita como

$$X'TT'X\beta = X'TT'y$$

Esse método de ajuste pode ser chamado ajuste de *norma- W mínima*. Ou, ainda, pode-se levar em conta que os elementos de T estão ponderando os elementos de X e, assim, temos o *ajuste de mínimos quadrados ponderados*.

Podemos contornar os detalhes envolvidos no caso de X com posto incompleto. Nesse caso, é possível usar um modelo reparametrizado, $y = Z\gamma$, em que a matriz do modelo tem posto completo e cujo ajuste será o mesmo de $y = X\beta$.

Considerando X de posto completo, $(X'WX)^{-1}$ existe e

$$Xb^* = X(X'WX)^{-1}X'Wy = Py$$

onde b^* é solução das equações de Aitken e $P = X(X'WX)^{-1}X'W$. Note que

$$P^2 = X(X'WX)^{-1}X'WX(X'WX)^{-1}X'W = X(X'WX)^{-1}XW = P$$

e

$$PX = X(X'WX)^{-1}X'WX = X$$

Como P é uma matriz de projeção, seus autovalores são 0 e 1. No entanto, $P' \neq P$, a menos que $W = cI$, onde c é alguma constante. Com o ajuste de mínimos quadrados ponderados, estamos ajustando y por uma projeção não ortogonal. (Uma projeção é ortogonal se, e somente se, sua matriz é simétrica.)

É possível visualizar o ajuste de norma- W mínima e a projeção não ortogonal associada, na Figura 1. Ali, $OV = y$ e $OU = Py$. Note que o ângulo OUV não é reto.

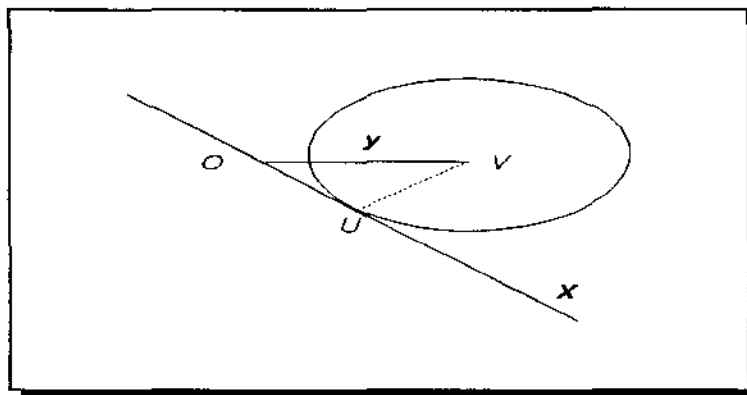


Figura 1. Projeção não ortogonal

Para entender a figura, sejam z e y pontos em \mathbb{R}^n , respectivamente arbitrário e fixo. Então,

$$(y - z)'W(y - z) = k^2$$

representa um elipsóide em \mathbb{R}^n (elipsóide, porque W é positiva definida). O ponto de melhor ajuste no espaço das colunas de X é o ponto desse subespaço mais perto de y . Tal ponto é o de tangência de $(y - z)'W(y - z) = k^2$ em $\mathcal{E}(X)$. Se $W = cI$, os contornos são esféricos e o ponto de tangência é dado por uma projeção perpendicular (ortogonal).

6.3 O USO DOS MÍNIMOS QUADRADOS PONDERADOS

O método dos mínimos quadrados ponderados pode ser usado para reduzir a influência de algumas observações. Observações com pequeno peso contribuem menos para a soma de quadrados e, assim, exercem menor influência no ajuste. Observações com grande peso, por sua vez, contribuem mais para a soma de quadrados e, portanto, exercem maior influência no ajuste. Assim, é razoável assinalar pequeno peso a observações com grande variância e peso maior às observações com variância pequena. Pode ser mostrado que melhores estimadores lineares não viciados são obtidos se os pesos são inversamente proporcionais às variâncias dos resíduos resultantes do ajuste. No entanto, a matriz de variâncias e covariâncias dos resíduos é desconhecida e necessitamos de meios que contornem essa dificuldade. Após o ajuste de mínimos quadrados, detectado o problema de variâncias desiguais através do estudo dos resíduos, devemos encontrar valores ligados às variâncias dos resíduos que possam ser usados como peso para as observações. Por exemplo, o inverso dos quadrados dos valores preditos ou o inverso dos quadrados dos resíduos podem ser pesos para outro ajuste. Os valores dos novos resíduos encontrados podem ser usados para cálculo de novos pesos e nova análise, se necessário. Novas iterações podem ser realizadas se os resultados indicarem nessa direção.

Outra questão a se considerar é a das funções paramétricas $L\beta$ que são estimáveis. Observando as equações de Aitken, vemos que a matriz X foi substituída por $H^{-1}X$. Assim, $L\beta$ é estimável se, e somente se, $L \in \mathfrak{R}(H^{-1}X)$. Como $\mathfrak{R}(H^{-1}X) = \mathfrak{R}(X)$, $L\beta$ é estimável se, e somente se, $L \in \mathfrak{R}(X)$.

Grizzle, Starmer e Koch (1969) propõem o método dos mínimos quadrados ponderados para o ajuste de modelos lineares a dados categóricos, em virtude da simplicidade que o método acarreta na manipulação de dados desse tipo.

Um exemplo de mínimos quadrados ponderados será visto no Capítulo 13.

Capítulo 7

Resolução Numérica das Equações Normais

7.1 INTRODUÇÃO

Pelo Teorema 1 do Capítulo 3, as equações normais são consistentes e, portanto, admitem uma ou mais soluções. Vimos também como obtê-las do ponto de vista algébrico. No entanto, sua solução numérica pode ser de difícil obtenção e, assim, existe a necessidade do desenvolvimento de algoritmos que permitam realizar essa tarefa. Tais cálculos, em geral, são muito trabalhosos, mesmo quando estão envolvidas matrizes de pequenas dimensões e o uso de computadores torna-se imperativo.

Como já foi visto, o vetor solução não é único no caso dos modelos com matriz de posto incompleto. No entanto, através dele se obtém resultados importantes como os valores preditos para as observações, estimativas de funções paramétricas e os resíduos.

O algoritmo que estudaremos fornecerá uma solução das equações normais e a soma de quadrados associada ao resíduo do ajuste. Não nos preocuparemos com a

implementação computacional do SWEEP(A,k). O leitor interessado poderá consultar o trabalho de Goodnight (1979).

A seguir, veremos algumas ferramentas que serão úteis para o estudo do operador SWEEP(A,k).

7.2 O PROCESSO DE ELIMINAÇÃO DE GAUSS-JORDAN

Um sistema de equações pode ser resolvido numericamente através do processo conhecido como *eliminação de Gauss-Jordan*.

Seja o sistema S ,

$$\begin{cases} 6\beta_1 + 2\beta_2 = 5 \\ 5\beta_1 + 13\beta_2 = 7 \end{cases}$$

O processo para a sua resolução é o seguinte:

A. Multiplicar a primeira equação por $1/6$:

$$\begin{cases} \beta_1 + \frac{1}{3}\beta_2 = \frac{5}{6} \\ 5\beta_1 + 13\beta_2 = 7 \end{cases}$$

B. Adicionar -5 vezes a primeira equação à segunda:

$$\begin{cases} \beta_1 + \frac{1}{3}\beta_2 = \frac{5}{6} \\ 0\beta_1 + \frac{34}{3}\beta_2 = \frac{17}{6} \end{cases}$$

C. Multiplicar a segunda equação por 3/34:

$$\begin{cases} \beta_1 + \frac{1}{3}\beta_2 = \frac{5}{6} \\ 0\beta_1 + \beta_2 = \frac{1}{4} \end{cases}$$

D. Adicionar -1/3 vezes a segunda equação à primeira:

$$\begin{cases} \beta_1 + 0\beta_2 = \frac{3}{4} \\ 0\beta_1 + \beta_2 = \frac{1}{4} \end{cases}$$

E obtivemos $\beta_1 = \frac{3}{4}$ e $\beta_2 = \frac{1}{4}$, os valores das incógnitas do sistema.

O sistema S é equivalente a

$$\begin{bmatrix} 6 & 2 \\ 5 & 13 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} 5 \\ 7 \end{bmatrix}$$

ou ainda, a

$$X\beta = y$$

e se pode aplicar à matriz aumentada A , formada a partir da matriz X e do vetor y , as etapas do processo de Gauss-Jordan da seguinte forma:

$$A = [X|y] = \left[\begin{array}{cc|c} 6 & 2 & 5 \\ 5 & 13 & 7 \end{array} \right]$$

1. O passo A é equivalente a premultiplicar A por $L_1 = \begin{bmatrix} \frac{1}{6} & 0 \\ 0 & 1 \end{bmatrix}$, resultando

$$A_1 = L_1 A = \begin{bmatrix} \frac{1}{6} & 0 \\ 0 & 1 \end{bmatrix} \left[\begin{array}{cc|c} 6 & 2 & 5 \\ 5 & 13 & 7 \end{array} \right] = \left[\begin{array}{cc|c} 1 & \frac{1}{3} & \frac{5}{6} \\ 5 & 13 & 7 \end{array} \right]$$

A operação realizada sobre A é uma operação linha. Note que A , ao ser

premultiplicada por L_1 , sofre uma alteração em sua primeira linha. L_1 é uma matriz derivada da matriz identidade onde o primeiro elemento de sua diagonal foi alterado de 1 para $1/6$. Esse valor é que multiplica a primeira equação de S no passo A.

2. O passo B é equivalente a premultiplicar A_1 por $L_2 = \begin{bmatrix} 1 & 0 \\ -5 & 1 \end{bmatrix}$, resultando

$$A_2 = L_2 A_1 = \begin{bmatrix} 1 & 0 \\ -5 & 1 \end{bmatrix} \left[\begin{array}{cc|c} 1 & \frac{1}{3} & \frac{5}{6} \\ 5 & 13 & 7 \end{array} \right] = \left[\begin{array}{cc|c} 1 & \frac{1}{3} & \frac{5}{6} \\ 0 & \frac{34}{3} & \frac{17}{6} \end{array} \right]$$

As operações realizadas sobre A_1 são também operações linha. A operação linha sobre A_1 promovida por L_2 , é do mesmo tipo da promovida por L_1 sobre A . Aqui também, L_2 é derivada da matriz identidade com o primeiro elemento da segunda linha alterado de 0 para -5, o valor que multiplica a primeira equação de S no passo B.

3. Este passo é semelhante ao passo 1, assim como o passo C é semelhante ao passo A. A matriz responsável pela operação linha correspondente é

$$L_3 = \begin{bmatrix} 1 & 0 \\ 0 & \frac{3}{34} \end{bmatrix}, \text{ de forma que}$$

$$A_3 = L_3 A_2 = \begin{bmatrix} 1 & 0 \\ 0 & \frac{3}{34} \end{bmatrix} \left[\begin{array}{c|c} 1 & \frac{1}{3} \left| \frac{5}{6} \right. \\ 0 & \frac{34}{3} \left| \frac{17}{6} \right. \end{array} \right] = \left[\begin{array}{c|c} 1 & \frac{1}{3} \left| \frac{5}{6} \right. \\ 0 & 1 \left| \frac{1}{4} \right. \end{array} \right]$$

4. Este passo é equivalente ao passo D acima. A operação linha realizada pela matriz L_4 sobre A_3 fornece

$$A_4 = L_4 A_3 = \begin{bmatrix} 1 & -\frac{1}{3} \\ 0 & 1 \end{bmatrix} \left[\begin{array}{c|c} 1 & \frac{1}{3} \left| \frac{5}{6} \right. \\ 0 & 1 \left| \frac{1}{4} \right. \end{array} \right] = \left[\begin{array}{c|c} 1 & 0 \left| \frac{3}{4} \right. \\ 0 & 1 \left| \frac{1}{4} \right. \end{array} \right]$$

E obtivemos uma matriz aumentada composta pela matriz identidade à esquerda e, à direita, pelo valor do vetor β que é solução do sistema S .

Portanto, partimos de um sistema do tipo $Az = c$, construímos uma matriz aumentada $[A|c]$, aplicamos sobre ela sucessivas operações linhas L_1, L_2, \dots, L_s e obtivemos o vetor solução do sistema. Esquemáticamente, podemos representar o processo da seguinte forma:

$$[A|c] \xrightarrow{L_1, L_2, \dots, L_s} [I|z]$$

As operações linha podem ser agrupadas em uma única matriz $L = L_s L_{s-1} \dots L_1$.

No caso do exemplo acima teremos

$$L = L_4 L_3 L_2 L_1 = \begin{bmatrix} \frac{3}{10} & -\frac{1}{10} \\ -\frac{1}{10} & \frac{1}{5} \end{bmatrix}$$

Mas $L = A^{-1}$, a inversa de A , de modo que o processo é equivalente a

$$A^{-1}[A|c] = [I|z]$$

Como o que foi visto acima vale para um sistema consistente qualquer, também valerá para as equações normais. Consideremos nesta, e na próxima seção, X de posto completo. Esquemáticamente, teremos para o sistema $X'X\beta = X'y$,

$$[X'X|X'y] \underline{L_1, L_2, \dots, L_s} [I|\beta]$$

ou, de modo equivalente,

$$[(X'X)^{-1}][X'X|X'y] = [I|\beta]$$

Após a aplicação do algoritmo, a matriz aumentada inicial teve a submatriz à esquerda transformada em uma matriz identidade. As operações linha são sempre escolhidas de forma a se obter tal resultado.

7.3 O OPERADOR SWEEP(A,k)

Uma das primeiras referências a esse método foi feita por Efron em Ralston e Welf, em 1960, para a análise de regressão múltipla. No entanto, segundo Goodnight (1979), o termo SWEEP foi criado por Beaton, em 1964.

Já dispomos de um meio para solucionar numericamente o sistema das equações normais. No entanto, é possível se aprimorar o processo para torná-lo mais geral e fornecer uma quantidade maior de informação.

Tomemos uma matriz aumentada Z de estrutura ligeiramente diferente da matriz aumentada da seção anterior:

$$Z = [X'X | X'y | I]$$

Aplicando operações linha sobre Z de modo a transformá-la em uma matriz aumentada cuja componente à esquerda seja a matriz identidade, teremos esquematicamente

$$[X'X | X'y | I] \xrightarrow{L_1, L_2, \dots, L_s} [I | \beta | (X'X)^{-1}]$$

Além de β , também $(X'X)^{-1}$ foi obtida.

É possível modificar as operações linha de modo a se conseguir, no final, as mesmas informações, partindo-se, porém, de uma estrutura matricial menor. Assim,

$$[X'X | X'y] \xrightarrow{L_1^*, L_2^*, \dots, L_s^*} [(X'X)^{-1} | \beta]$$

O método usado acima, chamado *inversão no lugar* de $X'X$, apresenta a vantagem de economia de memória computacional.

Consideremos o sistema de equações normais $X'X\beta = X'y$ para o modelo $y = X\beta$. Formando, a partir dele, a matriz

$$A = \left[\begin{array}{c|c|c} X'X & X'y & I \\ \hline y'X & y'y & \emptyset \end{array} \right]$$

e aplicando sobre ela determinadas operações linha, obteremos

$$A^* = \left[\begin{array}{c|c|c} I & b & (X'X)^{-1} \\ \hline \emptyset & S(X) & -b' \end{array} \right]$$

onde $S(X)$ é a soma de quadrados associada ao resíduo do ajuste do modelo e b , o valor de β que é solução do sistema. Neste capítulo, por conveniência de notação, usaremos $S(X)$ para denotar a soma de quadrados do resíduo. As operações linha podem ser representadas por uma matriz L , de modo que

$$A^* = LA = \left[\begin{array}{c|c} (X'X)^{-1} & \emptyset \\ \hline -y'X(X'X)^{-1} & I \end{array} \right] \left[\begin{array}{c|c|c} X'X & X'y & I \\ \hline y'X & y'y & \emptyset \end{array} \right] = \left[\begin{array}{c|c|c} I & b & (X'X)^{-1} \\ \hline \emptyset & S(X) & -b' \end{array} \right]$$

Se tomarmos, agora, A^* e sobre ela aplicarmos operações linha equivalentes a

$$L^{-1} = \left[\begin{array}{c|c} X'X & \emptyset \\ \hline y'X & I \end{array} \right] \text{ teremos}$$

$$L^{-1}A^* = \left[\begin{array}{c|c} X'X & \emptyset \\ \hline y'X & I \end{array} \right] \left[\begin{array}{c|c|c} I & b & (X'X)^{-1} \\ \hline \emptyset & S(X) & -b' \end{array} \right] = \left[\begin{array}{c|c|c} X'X & X'y & I \\ \hline y'X & y'y & \emptyset \end{array} \right] = A$$

Esse fato revela a reversibilidade do processo. As matrizes L e L^{-1} não necessitam ser calculadas.

Formando a partir das equações normais a matriz

$$A = \left[\begin{array}{c|c} X'X & X'y \\ \hline y'X & y'y \end{array} \right]$$

e aplicando sobre ela determinadas operações linha, obteremos

$$A^* = \left[\begin{array}{c|c} (X'X)^{-1} & b \\ \hline -b' & S(X) \end{array} \right]$$

onde $b = (X'X)^{-1}X'y$ é, novamente, o valor de β que satisfaz o sistema e $S(X) = yy - y'X(X'X)^{-1}X'y$, a soma de quadrados do resíduo devida ao ajuste do modelo. As operações linha realizadas sobre A , fornecendo A^* , representam o algoritmo que constitui o operador SWEEP(A,k).

Definição. Dada uma matriz A , originalmente simétrica positiva definida, o operador SWEEP(A,k) a modifica da seguinte forma:

1. fazer $D = a_{kk}$;
2. dividir a linha k por D ;
3. para todas as outras linhas $i \neq k$, fazer $B = a_{ik}$; subtrair B vezes a linha k da linha i ; fazer $a_{ik} = -B/D$;
4. fazer $a_{kk} = 1/D$.

A matriz A^* é obtida após aplicações sucessivas do operador SWEEP(A,k), para $k = 1, 2, \dots, p$.

Tomar a linha k da matriz A para a aplicação do operador SWEEP(A,k), é um processo chamado *pivoteamento*. Quando $k = 1$, o pivô é a linha 1. Quando $k = 2$, o

pivô é a linha 2 e, assim, por diante.

Quando o operador SWEEP(A,1) usa como pivô a linha 1 da matriz A , está sendo ajustado o modelo $y \doteq x_1\beta_1$ e que possui apenas x_1 como variável independente. Assim, teremos ao final do processo, b_1 e $S(x_1)$, a soma de quadrados do resíduo devida ao ajuste do modelo. Após a entrada de x_1 , usando como pivô a linha $k = 2$, obteremos informações referentes ao ajuste do modelo $y \doteq x_1\beta_1 + x_2\beta_2$. A operação fornecerá b_1 , b_2 e $S(x_2|x_1)$. Pode-se usar como pivô inicial, $k = 2$ ajustando o modelo $y \doteq x_2\beta_2$ e, posteriormente, acrescentar, digamos x_4 . Obteremos inicialmente b_2 , $S(x_2)$ e, a seguir, b_2 , b_4 e $S(x_4|x_2)$.

O operador SWEEP(A,k) também apresenta a característica da reversibilidade. A sua aplicação, tomando como pivô a linha $k = j$, cria a matriz A_j e promove a introdução no modelo da variável x_j . Sobre A_j , a aplicação do operador, tomando como pivô a linha $k = i$, cria a matriz A_i , que representa um modelo contendo as variáveis x_j e x_i . Se o operador for aplicado sobre essa última matriz, tomando como pivô, novamente, a linha $k = i$, a matriz resultante será A_j . Esquemáticamente, temos:

$$\text{SWEEP}(A_{h,j}) = A_j$$

$$\text{SWEEP}(A_j,i) = A_i$$

$$\text{SWEEP}(A_i,i) = A_j$$

Usando um vetor auxiliar, $V = [1 \ 1 \ \dots \ 1]$, pode-se controlar quais linhas já foram tomadas como pivô: quando a linha k for usada, v_k será feito $-v_k$. Após a aplicação do operador para qualquer linha k , obtém-se informações sobre submodelos entre as variáveis independentes conforme veremos no exemplo a seguir.

Consideremos os valores da Tabela I (Goodnight, 1977).

Tabela I. Valores para aplicação do SWEEP(A,k)

x_1	x_2	x_3	y
1	1	1	1
1	2	1	3
1	3	1	3
1	1	-1	2
1	2	-1	2
1	2	-1	2
1	3	-1	1

0. Formar a matriz A sobre a qual será aplicado o SWEEP(A,k).

$$A = \left[\begin{array}{c|c} X'X & X'y \\ \hline y'X & y'y \end{array} \right] = \begin{bmatrix} 6 & 12 & 0 & 12 \\ 12 & 28 & 0 & 25 \\ 0 & 0 & 6 & 2 \\ 12 & 25 & 2 & 28 \end{bmatrix}$$

1. Aplicar o operador sobre A , tomando como pivô a linha $k = 1$. Serão ajustados os modelos $y \doteq x_1\beta_1$, $x_2 \doteq x_1\alpha_1$ e $x_3 \doteq x_1\alpha_1^*$. Os valores dos coeficientes e das somas de quadrados dos resíduos são mostrados na Tabela II. Os símbolos '.' abaixo representam produtos cruzados entre as variáveis; $S_w(x_j)$ é a soma de quadrados do resíduo para o ajuste de w sobre x_j , onde w é, ou uma coluna da matriz X , ou y ; a_i e a_i^* são, respectivamente, as estimativas de α_i e α_i^* .

$$\text{SWEEP}(A,1) = A_1 = \begin{bmatrix} \frac{1}{6} & 2 & 0 & 2 \\ -2 & 4 & 0 & 1 \\ 0 & 0 & 6 & 2 \\ -2 & 1 & 2 & 4 \end{bmatrix} = \begin{bmatrix} (x_1'x_1)^{-1} & a_1 & a_1^* & b_1 \\ -a_1 & S_{x_2}(x_1) & . & . \\ -a_1^* & . & S_{x_3}(x_1) & . \\ -b_1 & . & . & S_y(x_1) \end{bmatrix}$$

2. Aplicar o operador sobre A_1 , tomando como pivô a linha $k = 2$. Isso implica em acrescentar x_2 ao modelo anterior e obteremos informações sobre o ajuste do modelo $y \doteq x_1\beta_1 + x_2\beta_2$. O modelo $x_3 \doteq x_1\alpha_1 + x_2\alpha_2$ também será ajustado. Abaixo, v_{ij} são os elementos da inversa da submatriz de $X'X$ que envolve as variáveis independentes presentes no ajuste. Por exemplo, neste passo temos

$$\begin{bmatrix} x_1'x_1 & x_1'x_2 \\ x_2'x_1 & x_2'x_2 \end{bmatrix}^{-1}$$

$$\text{SWEEP}(A_{1,2}) = A_2 = \begin{bmatrix} \frac{7}{6} & -\frac{1}{2} & 0 & \frac{3}{2} \\ -\frac{1}{2} & \frac{1}{4} & 0 & \frac{1}{4} \\ 0 & 0 & 6 & 2 \\ -\frac{3}{2} & -\frac{1}{4} & 2 & \frac{15}{4} \end{bmatrix} = \begin{bmatrix} v_{11} & v_{12} & a_1 & b_1 \\ v_{21} & v_{22} & a_2 & b_2 \\ -a_1 & -a_2 & S_{x_3}(x_1, x_2) & \cdot \\ -b_1 & -b_2 & \cdot & S_y(x_1, x_2) \end{bmatrix}$$

3. Aplicar o operador sobre A_2 , tomando como pivô a linha $k = 3$. Isso implica em acrescentar uma variável a mais ao modelo. Como $k = 3$, a variável será x_3 e o modelo ajustado, $y = x_1\beta_1 + x_2\beta_2 + x_3\beta_3$.

$$\text{SWEEP}(A_{2,3}) = A_3 = \begin{bmatrix} \frac{7}{6} & -\frac{1}{2} & 0 & \frac{3}{2} \\ -\frac{1}{2} & \frac{1}{4} & 0 & \frac{1}{4} \\ 0 & 0 & \frac{1}{6} & \frac{1}{3} \\ -\frac{3}{2} & -\frac{1}{4} & -\frac{1}{3} & \frac{37}{12} \end{bmatrix} = \begin{bmatrix} v_{11} & v_{12} & v_{13} & b_1 \\ v_{21} & v_{22} & v_{23} & b_2 \\ v_{31} & v_{32} & v_{33} & b_3 \\ -b_1 & -b_2 & -b_3 & S_y(x_1, x_2, x_3) \end{bmatrix}$$

4. Aplicar o operador sobre A_3 , tomando como pivô a linha $k = 2$. Tal procedimento resultará na retirada da variável x_2 do modelo, em virtude da reversibilidade do operador. Lembre que no passo 2, o pivô era a linha $k = 2$. O presente passo fornecerá informações sobre os modelos $y = x_1\beta_1 + x_3\beta_3$ e

$x_2 = x_1 \alpha_1 + x_3 \alpha_3$. v_{ij} , abaixo, são os elementos de $\begin{bmatrix} x'_1 x_1 & x'_1 x_3 \\ x'_3 x_1 & x'_3 x_3 \end{bmatrix}^{-1}$.

$$\text{SWEEP}(A_3, 2) = A_4 = \begin{bmatrix} \frac{1}{6} & 2 & 0 & 2 \\ -2 & 4 & 0 & 1 \\ 0 & 0 & \frac{1}{6} & \frac{1}{3} \\ -2 & 1 & -\frac{1}{3} & \frac{10}{3} \end{bmatrix} = \begin{bmatrix} v_{11} & a_1 & v_{12} & b_1 \\ -a_1 & S_{x_2}(x_1, x_3) & \cdot & \cdot \\ v_{21} & a_3 & v_{22} & b_3 \\ -b_1 & \cdot & -b_3 & S_y(x_1, x_3) \end{bmatrix}$$

5. Aplicar o operador sobre A_4 , tomando como pivô a linha $k = 3$. Como no passo 4, o resultado será a retirada de outra variável: x_3 . Portanto, ficaremos com os modelos do passo 1.

$$\text{SWEEP}(A_4, 3) = A_1 = \begin{bmatrix} \frac{1}{6} & 2 & 0 & 2 \\ -2 & 4 & 0 & 1 \\ 0 & 0 & 6 & 2 \\ -2 & 1 & 2 & 4 \end{bmatrix}$$

6. Finalmente, aplicar o operador sobre A_1 , tomando como pivô a linha $k = 1$. Isso provocará a retirada da variável x_1 do modelo ajustado no passo 5 e fornecerá, de volta, a matriz inicial A .

$$\text{SWEEP}(A_1,1) = A = \begin{bmatrix} 6 & 12 & 0 & 12 \\ 12 & 28 & 0 & 25 \\ 0 & 0 & 6 & 2 \\ 12 & 25 & 2 & 28 \end{bmatrix}$$

Tabela II. Resumo das aplicações do SWEEP(A,k) à matriz construída com os valores da Tabela I

Passo	Modelo	Coeficientes $S_w(x_j)$	
0	$y \doteq \emptyset$		28
1	$y \doteq x_1 \beta_1$	2	4
	$x_2 \doteq x_1 \alpha_1$	2	4
	$x_3 \doteq x_1 \alpha_1^*$	0	6
2	$y \doteq x_1 \beta_1 + x_2 \beta_2$	$\frac{3}{2}, \frac{1}{4}$	$\frac{15}{4}$
	$x_3 \doteq x_1 \alpha_1 + x_2 \alpha_2$	0,0	6
3	$y \doteq x_1 \beta_1 + x_2 \beta_2 + x_3 \beta_3$	$\frac{3}{2}, \frac{1}{4}, \frac{1}{3}$	$\frac{37}{12}$
4	$y \doteq x_1 \beta_1 + x_3 \beta_3$	$2, \frac{1}{3}$	$\frac{10}{3}$
	$x_2 \doteq x_1 \alpha_1 + x_3 \alpha_3$	2,0	4
5	$y \doteq x_1 \beta_1$	2	4
	$x_2 \doteq x_1 \alpha_1$	2	4
	$x_3 \doteq x_1 \alpha_1^*$	0	6
6	$y \doteq \emptyset$		28

7.4 O CASO DE X COM POSTO INCOMPLETO

Como discutido anteriormente, os elementos da diagonal de A correspondentes às linhas ainda não usadas como pivô representam soma de quadrados de resíduos para submodelos entre as variáveis independentes. Suponhamos que tenham sido acrescentadas ao modelo as variáveis x_1, \dots, x_{k-1} . O elemento a_{kk} de A será a soma de quadrados do resíduo para o modelo

$$x_k = x_1 \alpha_1 + \dots + x_{k-1} \alpha_{k-1} \quad (1)$$

O ajuste do modelo (1) é equivalente a projetar x_k no espaço formado pelos vetores x_1, \dots, x_{k-1} . Para o modelo (1), pode-se definir o *coeficiente de determinação*, denotado por R_k^2 , como

$$R_k^2 = \frac{v}{u}$$

onde v é o quadrado da norma da projeção de x_k corrigido pela média e u , o quadrado da norma de x_k corrigido pela média. Nas tabelas de análise de variância vistas como exemplo nos capítulos anteriores, aparece o coeficiente de determinação, referido como 'R-Square'. Quando considerarmos um modelo no qual o efeito da média não estiver presente (por exemplo, uma reta passando pela origem), o coeficiente de determinação deverá ser redefinido. Os vetores envolvidos na relação podem ser vistos

na Figura 1.

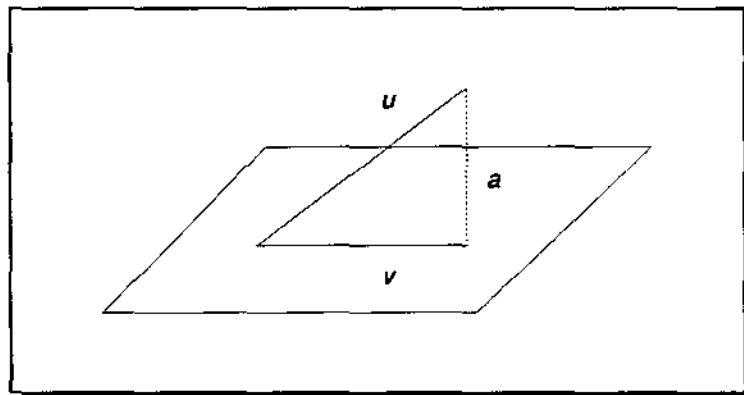


Figura 1. Vetores envolvidos no coeficiente de determinação e suas normas ao quadrado

Ali, v corresponde à diferença entre o quadrado da norma de x_k corrigido, que denotaremos por SQC_k , e a soma de quadrados do resíduo, a_{kk} . Este último vetor é representado na figura por a . Assim,

$$R_k^2 = \frac{SQC_k - a_{kk}}{SQC_k}$$

Caso x_k seja combinação linear das $k-1$ variáveis precedentes, a_{kk} terá o valor zero. Se x_k for aproximadamente uma combinação linear daquelas variáveis, a_{kk} estará próximo de 0 e resultará em um valor de R_k^2 próximo de 1, digamos 0,9999. Então,

x_k será, pelo menos estatisticamente, uma combinação linear das $k-1$ outras variáveis. Geometricamente, isso implica que existirá uma quase colinearidade em X com consequente instabilidade numérica.

R_k^2 e a *tolerância*, definida como $1 - R_k^2$, são ferramentas preciosas para testar a singularidade de X . No entanto, escolher um valor para R_k^2 , acima do qual declarar a presença estatística de dependência linear entre as colunas de X , não é tarefa fácil. O fato decorre da falta de teoria sobre a distribuição probabilística de R_k^2 .

Havendo dependência linear entre as colunas de X , é desnecessário que x_k dependa obrigatoriamente das $k-1$ colunas anteriores. Não sendo este o caso, as colunas de X podem ser permutadas de forma a se obter a situação representada em (1). Isso é conseguido através de uma matriz M , obtida por permutação das colunas da matriz identidade. M é tal que $MM' = I$. Assim,

$$y \doteq X\beta = XI\beta = XMM'\beta$$

Fazendo $Z = XM$ e $\gamma = M'\beta$, $y \doteq Z\gamma$, onde $Z = [Z_1 \ Z_2]$, com Z_1 contendo colunas linearmente independentes e $Z_2 = Z_1H$, para alguma matriz H . Pelo Teorema 4 do Capítulo 3, sabemos que $XB = B'X'$. Portanto, $XMM'B = B'MM'X'$. Fazendo $M'B = C$, temos $ZC = C'Z'$. E, como $y \doteq Z\gamma$,

$$g = CC'Z'y - M\hat{\gamma} = MCC'Z'y = MM'BB'MM'X'y = BB'X'y = b$$

onde g e b são os valores obtidos para γ e β . Portanto, solucionando o sistema para γ , obtém-se b .

O operador SWEEP(A,k) pode ser aplicado a matrizes cuja estrutura seja diferente da estrutura de Z , fornecendo os mesmos resultados. Então, para a aplicação do operador é desnecessária a prévia permutação das colunas de X .

Nos Modelos Lineares onde X tem posto incompleto, a matriz $X'X$ que aparece nas equações normais, também tem posto incompleto. Assim, $(X'X)^{-1}$ não existe e o algoritmo deve sofrer alteração para contornar o problema: detectada a dependência linear na k -ésima coluna, ambas, a linha e a coluna k da matriz que está sendo submetida ao operador, serão tornadas iguais a zero. Como consequência, o k -ésimo elemento do vetor de parâmetros, b , torna-se igual a zero. O último elemento da diagonal de A^* permanece sendo $S(X)$ e a submatriz $p \times p$ do canto superior esquerdo de A^* é a matriz BB' . Essa matriz obedece às seguintes propriedades:

1. $X'XBB'X'X = (X'XB)(X'XB)' = X'X$
 2. $BB'X'XBB' = BB'$
- (2)

Como BB' surge da aplicação de operações linha sobre $X'X$, $p(BB') = p(X'X)$. Pode-

se verificar (ver Rao e Mitra (1971)) que nesta condição a expressão (2) é válida¹.

Mais uma vez é bom se chamar atenção para o fato de que a solução encontrada não é única. O vetor b não é uma estimativa de β no modelo $y = X\beta$. O vetor encontrado é apenas uma das soluções das equações normais. BB' também não é única.

¹A expressão (2) indica que $X'X$ é uma inversa generalizada de BB' . Assim, BB' é uma inversa generalizada reflexiva.

Capítulo 8

Forma Geral das Funções Estimáveis

8.1 INTRODUÇÃO

Já vimos que um conjunto de funções, $\{l_i\beta\}_{1 \leq i \leq k}$ ou $L\beta$, será estimável se, e somente se, existe uma combinação linear de y cujo valor esperado seja $L\beta$, para qualquer β . Vimos, também, que $L \in \mathcal{R}(X)$ ou $L = AX$, para algum A . Então, L deverá ser uma combinação linear das linhas de X . Portanto, as linhas de X formam um conjunto gerador a partir do qual um L , tal que $L\beta$ é estimável, pode ser construído.

Porém,

$$L = AX = A(XBX) = AX(BB'X'X) = U(BB'(X'X)) = V(X'X)$$

Assim, $X'X$ e $BB'X'X$ também formarão conjuntos geradores para L , tais que $L\beta$ será estimável. Além disso, operações linha sobre X , $X'X$ e $BB'X'X$ que não

diminuam o posto dessas matrizes, resultarão em matrizes que também serão conjuntos geradores para L .

8.2 OS CONJUNTOS GERADORES DAS FUNÇÕES ESTIMÁVEIS

Tomemos como exemplo,

$$X = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}$$

e

$$\beta = \begin{bmatrix} \mu \\ A_1 \\ A_2 \\ A_3 \end{bmatrix}$$

Assim como X forma um conjunto gerador para L ,

$$X^* = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}$$

também formar. Já que todos os L devem ser funções lineares das linhas de X^* , um L , para uma estimativa com 1 grau de liberdade, pode ser representado simbolicamente como

$$a'X^* = [l_1 \quad l_2 \quad l_3] \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}$$

(Os elementos do vetor a' foram e serão representados por l_i por conveniência notacional.) Então, $L = [l_1 + l_2 + l_3 \quad l_1 \quad l_2 \quad l_3]$. Neste caso, $L\beta$ é estimável se, e somente se, o primeiro elemento de L é igual à soma dos demais.

Se operações linha forem aplicadas a X^* para formar uma matriz identidade à esquerda, teremos

$$X^{**} = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{bmatrix}$$

e

$$a'X^{**} = [l_1 \quad l_2 \quad l_3] \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{bmatrix}$$

Aqui, $L = [l_1 \quad l_2 \quad l_3 \quad l_1 - l_2 - l_3]$ e, novamente, o primeiro elemento de L é igual à soma dos demais.

Como infinitos conjuntos geradores estão disponíveis, o que se questiona é qual deles usar. Qual representará L mais convenientemente? É claro que aquele conjunto que contiver o menor número de linhas (portanto de posto linha completo) e a maior quantidade de elementos iguais a zero, será mais conveniente.

Ao ajustar um modelo aplicando o operador SWEEP(A,k) sobre uma matriz conveniente, a matriz BB' é automaticamente produzida. Tal matriz apresenta numerosos elementos iguais a zero e suas linhas não zero são linearmente independentes (ver Quadro V, seção 3.9). Além disso, $BB'X'X$ é um conjunto gerador

para L . Então, é razoável usar $BB'X'X$ para gerar L .

Para exemplificar, vejamos duas situações: modelo com um critério de classificação e modelo com três critérios de classificação.

Modelo com um critério de classificação

Para o modelo $y_{ir} = \mu + A_i + e_{ir}$, $i = 1, 2, 3$ e $r = 1, 2, \dots, n$, a forma geral das funções estimáveis e das restrições estimáveis será

$$L\beta = [l_1 \quad l_2 \quad l_3 \quad l_1 - l_2 - l_3] \begin{bmatrix} \mu \\ A_1 \\ A_2 \\ A_3 \end{bmatrix}$$

$$= l_1\mu + l_2A_1 + l_3A_2 + (l_1 - l_2 - l_3)A_3$$

Assim, funções estimáveis envolvendo os parâmetros A_1 , A_2 e A_3 devem ter L da forma $L = [0 \quad l_2 \quad l_3 \quad -l_2 - l_3]$. Como estão envolvidos somente dois símbolos, l_2 e l_3 , no máximo duas funções estimáveis podem ser obtidas: fazendo $l_2 = 1$ e $l_3 = 0$ e, então, $l_2 = 0$ e $l_3 = 1$, teremos

$$L = \begin{bmatrix} 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{bmatrix}$$

E L , como acima, poderá ser usado para construir as restrições estimáveis $H_0: A_1 = A_2 = A_3$, pois

$$\emptyset = L\beta = \begin{bmatrix} 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} \mu \\ A_1 \\ A_2 \\ A_3 \end{bmatrix} = \begin{bmatrix} A_1 - A_3 \\ A_2 - A_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Modelo com três critérios de classificação

Seja o modelo $y_{ijk} = \mu + A_i + B_j + C_k$, $i = 1, 2, 3$, $j = 1, 2$, $k = 1, 2, 3$ e $r = 1, \dots, 5$, com três critérios de classificação onde os valores componentes da matriz X são os da Tabela I.

Tabela I. Elementos da matriz X para um modelo com três critérios de classificação

μ	A_1	A_2	B_1	B_2	C_1	C_2	C_3
1	1	0	0	1	1	0	0
1	1	0	1	0	0	1	0
1	0	1	1	0	0	0	1
1	0	1	0	1	0	1	0
1	0	1	0	1	0	1	0

Tabela II. Forma geral das funções estimáveis para um modelo com três critérios de classificação

Parâmetro	Coefficiente
μ	l_1
A_1	l_2
A_2	$l_1 - l_2$
B_1	l_4
B_2	$l_1 - l_4$
C_1	l_6
C_2	$l_1 + l_2 - l_4 - 2l_6$
C_3	$-l_2 + l_4 + l_6$

A forma geral das funções estimáveis será definida por

$$L = a'BB'X'X = [l_1 \ l_2 \ l_3 \ l_4 \ l_5] \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & -2 & 1 \end{bmatrix}$$

Os coeficientes simbólicos para a média e para cada um dos níveis das três variáveis classificatórias podem ser vistos na Tabela II.

Analisemos a forma geral das funções estimáveis e as restrições estimáveis que poderão ser construídas:

1. Funções envolvendo os parâmetros de A deverão ter os coeficientes simbólicos referentes a μ , B_1 , B_2 , C_1 , C_2 e C_3 iguais a zero. Assim, $l_1 = l_2 = l_4 = l_6 = 0$ e os coeficientes simbólicos de A_1 e A_2 foram igualados a zero. Portanto, funções envolvendo A terão posto zero, $R(A|\mu, B, C) = 0$ e a restrição $A_1 - A_2$ não é estimável.
2. Funções envolvendo os parâmetros de B deverão ter os coeficientes simbólicos referentes a μ , A_1 , A_2 , C_1 , C_2 e C_3 iguais a zero. Isso também tornará igual a zero os coeficientes de B . Mais uma vez essas funções terão posto zero,

$R(B|\mu, A, C) = 0$ e a restrição $B_1 - B_2$ não é estimável.

3. Para funções envolvendo os parâmetros de C , $l_1 = l_2 = l_4 = 0$ e os coeficientes de cada nível de C serão respectivamente l_6 , $-2l_6$ e l_6 . Como apenas um coeficiente, l_6 , está envolvido, a restrição estimável terá 1 grau de liberdade:

$H_0: C_1 - 2C_2 + C_3 = k$, para qualquer k . Comumente se deseja $k = 0$. Aqui,

$$L = [0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 1 \quad -2 \quad 1].$$

Veremos no Capítulo 11, que para cada restrição estimável H_0 , calcula-se a soma de quadrados $Q = (Lb - k)'(LBB'L')^{-1}(Lb - k)$, que será usada para determinar o nível de significância de H_0 . Para cada restrição estimável, existe um L que dará origem a uma soma de quadrados adequada. Ou ainda, cada restrição estimável tem sua respectiva soma de quadrados.

Algumas restrições são particularmente úteis e são usadas com frequência. É o caso, por exemplo para o modelo $y_{hijr} = \mu + A_h + B_i + C_j + e_{hijr}$, das restrições que tem como somas de quadrados associadas $R(A|\mu)$, $R(B|\mu, A)$ e $R(C|\mu, A, B)$. Ou ainda, $R(A|\mu, B, C)$, $R(B|\mu, A, C)$ e $R(C|\mu, A, B)$. É natural, então, que tais restrições sejam de alguma forma padronizadas. Essa é razão do uso das somas de quadrados tipos I, II, III e IV, cada uma tendo suas utilidades e limitações. Isso não invalida ou limita o uso de outras restrições que não se enquadram na padronização acima: desde que a restrição seja estimável, determina-se o L correspondente e com ele o valor de Q .

8.3 EXEMPLO

Para o ajuste do modelo da seção 3.9, a forma geral das funções estimáveis pode ser vista no Quadro I abaixo. Ela é obtida pelo seguinte comando¹ SAS[®].

```
MODEL VALINA = PACIENTE INGESTAO / E;
```

Quadro I. Forma geral das funções estimáveis para o ajuste do modelo da seção 3.9

General Linear Models Procedure		
General Form of Estimable Functions		
Effect		Coefficients
INTERCEPT		L1
PACIENTE	NPELAGRA	L2
	PELAGRA	L1-L2
INGESTAO	ALTA	L1
	MEDIA	L4
	BAIXA	L1-L4-L5

¹A opção /E do comando MODEL fornece a forma geral das funções estimáveis.

Capítulo 9

Funções Estimáveis Tipos I, II, III e IV

9.1 INTRODUÇÃO

Antes da discussão das funções estimáveis tipos I, II, III e IV (FE I, FE II, FE III e FE IV) e das respectivas somas de quadrados geradas por elas, é importante se referir às situações que levaram à sua criação.

Yates, em 1934, já chamava atenção e propunha técnicas computacionais para a análise de dados desbalanceados: o método do ajuste de constantes e o método dos quadrados ponderados de médias. O primeiro seria útil para o caso onde as interações fossem assumidas ausentes. No segundo, as interações estariam presentes.

Em 1969, Overall e Spiegel, apesar de não fazerem referência ao operador $SWEEP(A,k)$, nem às considerações prévias de Yates, delineiam a construção das somas de quadrados tipos I, II e III e o uso da redução das somas de quadrados.

O surgimento das somas de quadrados tipos I, II, III e IV está intimamente ligado à existência de delineamentos experimentais com números diferentes de

replicações por casela e a ocorrência de caselas vazias. Em ambos os casos, há dificuldade na interpretação dos resultados dos testes de significância (ver Capítulo 11): nem sempre o uso das técnicas habituais produz somas de quadrados que testam o que se deseja. Além disso, os resultados obtidos podem levar a conclusões incorretas sem que o pesquisador tome conhecimento do fato se não estiver alerta.

No Capítulo 11, veremos que o ajuste de um modelo restrito com restrições estimáveis produz somas de quadrados que se convenientemente comparadas às somas de quadrados resultantes do ajuste do modelo irrestrito, fornece um meio de se avaliar a validade das restrições. Essas restrições são hipóteses sobre os parâmetros do modelo que serão testadas através de testes de significância. Neste capítulo, estudaremos a construção das somas de quadrados resultantes do ajuste de um modelo restrito com restrições estimáveis.

9.2 DADOS DESBALANCEADOS

Um exemplo servirá para chamar a atenção sobre a questão dos dados desbalanceados. Seja o modelo $y_{ijr} = \mu + A_i + B_j$ (onde, por simplicidade, a interação e o erro foram excluídos), $i, j = 1, 2$ e $r = 1, \dots, n_{ij}$, com n_{ij} , o número de replicações na linha i , coluna j . Consideremos os dados seguintes:

	B_1	B_2
A_1	7 9	5
A_2	8	4 6

onde $n_{11} = n_{22} = 2$, $n_{12} = n_{21} = 1$.

No nível B_1 , a média no primeiro nível de A é $\bar{y}_{11} = (7+9)/2 = 8$ e, no segundo, $\bar{y}_{21} = 8$. Portanto, não há evidência de diferença entre os níveis de A , dentro de B_1 . Da mesma forma, não há evidência de diferença entre os níveis de A , dentro de B_2 , pois $\bar{y}_{12} = 5$ e $\bar{y}_{22} = (4+6)/2 = 5$. Pode-se concluir, então, que não há evidência, nos dados, de diferença entre os níveis de A . Entretanto, as médias marginais de A são $\bar{y}_{1..} = (7+9+5)/3 = 7$ e $\bar{y}_{2..} = (8+4+6)/3 = 6$. A diferença, $7 - 6 = 1$, entre essas médias pode ser erroneamente interpretada como medindo um efeito geral do fator A . O que está ocorrendo é que a diferença observada entre as médias marginais dos dois níveis de A mede o efeito do fator B , somado ao efeito de A . Isso pode ser mostrado da seguinte forma: seja o modelo acima em que a diferença entre as médias marginais de A_1 e A_2 é dada por

$$\bar{y}_{1..} - \bar{y}_{2..} = \frac{\sum_{r=1}^{n_{11}} y_{11r} + \sum_{r=1}^{n_{12}} y_{12r}}{n_{11} + n_{12}} - \frac{\sum_{r=1}^{n_{21}} y_{21r} + \sum_{r=1}^{n_{22}} y_{22r}}{n_{21} + n_{22}}$$

Assim,

$$\begin{aligned} \bar{y}_{1..} - \bar{y}_{2..} &= \frac{n_{11}(A_1 + B_1) + n_{12}(A_1 + B_2)}{n_{11} + n_{12}} - \frac{n_{21}(A_2 + B_1) + n_{22}(A_2 + B_2)}{n_{21} + n_{22}} \\ &= (A_1 - A_2) + \frac{n_{22}n_{11} - n_{12}n_{21}}{(n_{11} + n_{12})(n_{21} + n_{22})} (B_1 - B_2) \end{aligned} \quad (1)$$

A expressão (1) é uma soma ponderada das diferenças entre os níveis de A e de B e não apenas a diferença entre os níveis de A como se desejava. Esta última está contaminada pelo efeito B . No entanto, se $n_{11} = n_{12} = n_{21} = n_{22}$, o peso de B será zero e $\bar{y}_{1..} - \bar{y}_{2..}$ estará refletindo a diferença entre as médias marginais de A_1 e A_2 .

Outra situação de interesse ocorre quando as frequências das caselas são proporcionais. Considere $n_{12} = qn_{11}$ e $n_{22} = qn_{21}$, o que leva a

$$\frac{n_{22}n_{11} - n_{12}n_{21}}{n_{11}n_{21}} = \frac{qn_{21}n_{11} - qn_{11}n_{21}}{n_{11}n_{21}} = 0$$

e, apesar de desbalanceados, os dados se comportam como balanceados.

A estatística a ser usada num teste de significância onde os dados são desbalanceados necessita ser escolhida com critério. Digamos que a estatística $\bar{y}_{1.} - \bar{y}_{2.}$ seja usada para testar $H_0: A_1 = A_2$. Acabamos de ver que quando os dados são desbalanceados, não estamos testando tal hipótese, mas sim, algo que pode carecer completamente de sentido para o pesquisador.

Veremos, no Capítulo 11, que um teste de significância envolve o cálculo da soma de quadrados $Q = (Lb - k)'(LBB'L')^{-1}(Lb - k)$. Tanto L como B são derivadas de $X'X$, que é composta de somas dos n_{ij} . Por exemplo, para o modelo $y_{ijr} = \mu + A_i + B_j + A*B_{ij}$, $i, j, r = 2$, teremos

$$X'X = \begin{bmatrix} n_{.} & n_{1.} & n_{2.} & n_{.1} & n_{.2} & n_{11} & n_{12} & n_{21} & n_{22} \\ n_{1.} & n_{1.} & 0 & n_{11} & n_{12} & n_{11} & n_{12} & 0 & 0 \\ n_{2.} & 0 & n_{2.} & n_{21} & n_{22} & 0 & 0 & n_{21} & n_{22} \\ n_{.1} & n_{11} & n_{21} & n_{.1} & 0 & n_{11} & 0 & n_{21} & 0 \\ n_{.2} & n_{12} & n_{22} & 0 & n_{.2} & 0 & n_{12} & 0 & n_{22} \\ n_{11} & n_{11} & 0 & n_{11} & 0 & n_{11} & 0 & 0 & 0 \\ n_{12} & n_{12} & 0 & 0 & n_{12} & 0 & n_{12} & 0 & 0 \\ n_{21} & 0 & n_{21} & n_{21} & 0 & 0 & 0 & n_{21} & 0 \\ n_{22} & 0 & n_{22} & 0 & n_{22} & 0 & 0 & 0 & n_{22} \end{bmatrix}$$

Então, é razoável que os n_{ij} estejam envolvidos no valor de Q .

9.3 HIPÓTESES DE INTERESSE

Consideremos o modelo

$$y_{ijr} = \mu + A_i + B_j + A * B_{ij} + e_{ijr} \quad (2)$$

onde $i = 1, \dots, a$, $j = 1, \dots, b$, $r = 1, \dots, n_{ij}$, $n_{ij} > 0$, para quaisquer i, j e $n_{ij} > 1$, para algum i, j . Para facilitar o entendimento, coloquemos os termos do membro direito de (2), excluindo o erro, iguais a μ_{ij} . O modelo pode ser reescrito como

$$y_{ijr} = \mu_{ij} + e_{ijr}$$

Frequentemente, as hipóteses de interesse serão:

1. $H_0: \mu_{1.} = \mu_{2.} = \dots = \mu_{a.}$
2. $H_0: \mu_{.1} = \mu_{.2} = \dots = \mu_{.b}$
3. $H_0: \frac{\sum_j n_{1j} \mu_{1j}}{n_{1.}} = \frac{\sum_j n_{2j} \mu_{2j}}{n_{2.}} = \dots = \frac{\sum_j n_{aj} \mu_{aj}}{n_{a.}}$
4. $H_0: \frac{\sum_i n_{i1} \mu_{i1}}{n_{.1}} = \frac{\sum_i n_{i2} \mu_{i2}}{n_{.2}} = \dots = \frac{\sum_i n_{ib} \mu_{ib}}{n_{.b}}$
5. $H_0: \mu_{ij} - \mu_{i'j} - \mu_{ij'} + \mu_{i'j'} = 0$, para quaisquer i, i', j, j' ($i \neq i', j \neq j'$)

6. H_0 : hipótese 1, sujeita a $\mu_{ij} - \mu_{i'j} - \mu_{ij'} + \mu_{i'j'} = 0$

7. H_0 : hipótese 2, sujeita a $\mu_{ij} - \mu_{i'j} - \mu_{ij'} + \mu_{i'j'} = 0$

A hipótese para testar a interação é a número 5. Entretanto, testes apropriados para efeitos principais (assumindo que a interação não é significativa) não estão claros. As hipóteses 3 e 4 não são adequadas para testar efeitos principais, a menos que os dados tenham sido obtidos através de amostra aleatória de tamanho suficientemente grande para que as proporções n_{ij}/n_i e n_{ij}/n_j sejam boas estimadoras das verdadeiras proporções populacionais. Restam, então, as hipóteses 1, 2, 6 e 7. Sem dúvida, as de números 1 e 2 são as mais adequadas e mais frequentemente desejadas, desde que as de números 6 e 7 implicam em testar efeitos sujeitos à ausência de interação. O caso especial onde as caselas apresentam frequências proporcionais pode ser encarado como se os dados fossem balanceados e as hipóteses 6 e 7 podem ser usadas para testes de efeitos principais, somente se a interação não é significativa.

9.4 EFEITOS CONTIDOS EM OUTROS

Definição. Dado um efeito E_1 e outro efeito E_2 , E_1 está contido em E_2 se:

1. ambos os efeitos envolvem as mesmas variáveis contínuas, se estas existirem;
2. E_2 tem mais variáveis classificatórias que E_1 . Se E_1 tem variáveis

classificatórias, todas elas aparecem em E_2 .

Nota: o efeito μ está contido em todas os efeitos classificatórios, mas não está contido em qualquer efeito envolvendo variáveis contínuas. Nenhum efeito está contido em μ .

Se E_2 contém¹ E_1 , todas as colunas de X associadas com E_1 podem ser representadas como combinações lineares das colunas associadas com E_2 . Por exemplo, no modelo $y_{ijr} = \mu + A_i + B_j + A*B_{ij}$,

μ está contido em A , B e em $A*B$;

A está contido em $A*B$;

B está contido em $A*B$ e

$A*B$ não está contido em qualquer outro efeito.

Além disso, se E_2 contém E_1 , $R(E_1|E_2) = 0$, pois

$$R(E_1|E_2) = R(E_1, E_2) - R(E_2) = R(E_2) - R(E_2) = 0$$

Analisemos geometricamente a questão dos efeitos contidos em outros. Sejam X_1 e X_2 as colunas de X relacionadas respectivamente a E_1 e E_2 . Se E_1 está contido em E_2 , o subespaço gerado pelas colunas de X_1 está contido no subespaço gerado pelas

¹Deve-se notar a diferença entre 'efeitos contidos em outros' e 'fatores embutidos' (Capítulo 2)

colunas de X_2 e

$$\mathcal{E}(X_1) \subset \mathcal{E}(X_2) \rightarrow P_2 P_1 = P_1 \quad (3)$$

onde $P_i = X_i B_i$ é o projetor ortogonal em $\mathcal{E}(X_i)$. A expressão (3) é verdadeira pois

$$P_2 P_1 = X_2 B_2 X_1 B_1 = X_1 B_1 = P_1$$

Por outro lado, se $P_2 P_1 = P_1$,

$$\mathcal{E}(X_1) = \mathcal{E}(P_1) = \mathcal{E}(P_2 P_1) \rightarrow \mathcal{E}(P_1) \subset \mathcal{E}(P_2) \rightarrow \mathcal{E}(X_1) \subset \mathcal{E}(X_2)$$

Como $\mathcal{E}(X_1) \subset \mathcal{E}(X_2)$, $X_1 = X_2 V$, para alguma matriz V . Assim, o modelo $y = X_1 \beta_1 + X_2 \beta_2$ será, de fato, $y = X_2 V \beta_1 + X_2 \beta_2$ ou $y = X_2 (V \beta_1 + \beta_2)$. Portanto, o modelo com E_1 e E_2 , mas no qual E_2 contém E_1 , é equivalente ao modelo com E_2 somente. Dessa forma, $R(E_1, E_2) = R(E_2)$.

Por exemplo, se $E_1 = A$ e $E_2 = A * B$, então,

$$R(A|A*B) = R(A, A*B) - R(A*B) = R(A*B) - R(A*B) = 0$$

9.5 FUNÇÕES ESTIMÁVEIS TIPO I

Esta forma de testar hipóteses é baseada no método do *ordenamento a priori* que Overall e Spiegel sugerem em 1969 e que é um caso especial do método do *ajuste de constantes*, descrito por Yates em 1934.

As FE I e as somas de quadrados que elas geram (SQ I) testam hipóteses sobre os efeitos do modelo, de forma sequencial. Por isso, as SQ I são ordem dependentes e variam de acordo com o modelo usado: cada efeito é ajustado após os efeitos precedentes já estarem no modelo. Por exemplo, para o modelo $y = X_1\beta_1 + X_2\beta_2 + X_3\beta_3$, as SQ I correspondem a

<i>Efeito</i>	<i>SQ I</i>
β_1	$R(\beta_1)$
β_2	$R(\beta_2 \beta_1)$
β_3	$R(\beta_3 \beta_1, \beta_2)$

Para calcular a SQ I para um efeito, necessitamos um vetor I' adequado. Além disso, L representa o conjunto das funções estimáveis linearmente independentes, cujo número total é igual ao posto de $X'X$. Portanto, L tem posto completo.

Nosso objetivo é determinar L , o que pode ser feito de várias formas. Entretanto, para calcular $R(\cdot)$ não é necessária a determinação dessa matriz (ou vetor), pois as SQ I são obtidas a partir de subprodutos do operador $\text{SWEEP}(A, k)$. Porém, a análise de L permite conhecer o que está sendo testado já que na presença de dados desbalanceados as somas de quadrados de determinados efeitos poderão conter funções de outros parâmetros.

A matriz $X'X$ forma um conjunto gerador para as funções estimáveis. Ela também serve como base para o operador $\text{SWEEP}(A, k)$: sua aplicação na matriz A , usando como pivô as colunas de $X'X$ relacionadas a X_q , ajusta um modelo contendo esse efeito, fornece $\hat{\beta}_q$ e a soma de quadrados do resíduo de tal ajuste. Assim, o subconjunto de elementos de $X'X$ relacionado a X_q fornecerá um conjunto gerador para funções estimáveis tipo I para β_q .

Consideremos o modelo $y = X_1\beta_1 + X_2\beta_2 + X_3\beta_3$. Aqui, a matriz $X'X$ pode ser representada por

$$X'X = \begin{bmatrix} X_1'X_1 & X_1'X_2 & X_1'X_3 \\ X_2'X_1 & X_2'X_2 & X_2'X_3 \\ X_3'X_1 & X_3'X_2 & X_3'X_3 \end{bmatrix}$$

Para o ajuste do submodelo $y = X_1\beta_1$ e cálculo de $R(\beta_1)$, portanto SQ I para β_1 ,

aplica-se o operador SWEEP(A,k), tomando como pivô as colunas da submatriz $X_1'X_1$.

Assim, $G_1 = [X_1'X_1 \quad X_1'X_2 \quad X_1'X_3]$ formará um conjunto gerador para as FE I para β_1 .

Após a introdução desse efeito no modelo, a matriz $X'X$ foi modificada pelo operador, formando

$$A_1 = \begin{bmatrix} B_1B_1' & B_1B_1'X_1'X_2 & B_1B_1'X_1'X_3 \\ -X_2'X_1B_1B_1' & X_2^*X_2^* & X_2'M_1X_3 \\ -X_3'X_1B_1B_1' & X_3'M_1'X_2 & X_3'M_1X_3 \end{bmatrix}$$

onde B_1 é tal que $X_1'X_1B_1 = X_1$ e $X_2^*X_2^* = X_2'M_1X_2$ com $M_1 = I - X_1B_1B_1'X_1'$, o projetor ortogonal no espaço de erro de X_1 . Usando o raciocínio anterior,

$G_2 = [\emptyset \quad X_2^*X_2^* \quad X_2'M_1X_3]$ formará um conjunto gerador das FE I para $\beta_2 | \beta_1$.

Aplicando o operador sobre as colunas correspondentes a X_2^* , obteremos, de forma

semelhante, $G_3 = [\emptyset \quad \emptyset \quad X_3^*X_3^*]$, onde $X_3^*X_3^* = X_3'M_2X_3$ com

$M_2 = M_1 - M_1X_2B_2^*B_2^*X_2'M_1$ e $B_2^*B_2^*$, uma matriz, tal que $X_2^*X_2^*B_2^* = X_2^*$. G_3 será um

conjunto gerador das FE I para $\beta_3 | \beta_1, \beta_2$. Além disso, conforme visto no estudo do

operador SWEEP(A,k), o último elemento à direita e abaixo da matriz sobre a qual estamos trabalhando (linha e coluna não representadas acima e que envolvem y),

corresponde ao valor da soma de quadrados do resíduo do ajuste. Ela é usada para a

obtenção de $R(\cdot|\cdot)$.

Como sabemos, existe um número infinito de conjuntos geradores para as funções estimáveis. Deseja-se, obviamente, um conjunto simples. Vimos que $BB'X'X$ constitui um tal conjunto. De forma equivalente, combinações lineares de G_1 , G_2 e G_3 , tais que $L = VG_i$, $i = 1,2,3$, é de posto linha completo e igual ao de G_i , também fornecerão conjuntos geradores para as FE I. Fazendo $V = B_1B_1'$ ou $V = B_i^*B_i^{*'}$, a matriz L obtida, de modo semelhante, é de estrutura simples. Portanto, um conjunto gerador conveniente das FE I para β_1 , β_2 e β_3 , onde B_3^* é tal que $X_3^{*'}X_3^*B_3^* = X_3^{*'}$, será

$$G_1^* = B_1B_1'G_1$$

$$G_2^* = B_2^*B_2^{*'}G_2$$

$$G_3^* = B_3^*B_3^{*'}G_3$$

Analisemos um exemplo. Sejam os dados abaixo e o modelo

$$y_{ijr} = \mu + A_i + B_j + A * B_{ij} \quad i,j,r = 1,2 \quad (4)$$

	B_1	B_2
A_1	23 24	28 29
A_2	8 10	5 7

$$l' = a'BB'X'X = [l_1 \ l_2 \ l_3 \ l_4 \ l_5 \ l_6 \ l_7 \ l_8 \ l_9] \times$$

$$\times \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & -1 & 0 & 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & -1 & -1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

A forma geral das funções estimáveis para este modelo encontra-se na Tabela I.

Aplicamos o SWEEP(A,1) na primeira coluna de $X'X$, o que corresponde a ajustar o modelo $y_{ijr} = \mu$. Tomemos a submatriz G_2 , calculemos G_2^* e o vetor l' , conforme indicado acima. Procedendo da mesma forma, obteremos G_3^* e G_4^* . As FE I para A , B e $A+B$ estão mostradas na Tabela II.

Vejamos agora o caso em que os dados são desbalanceados. Para isso, retiremos uma observação da casela A_1B_1 .

	B_1	B_2
A_1	23	28 29
A_2	8 10	5 7

Adotando o mesmo procedimento usado acima, as FE I para A , B e $A*B$ serão as da Tabela III.

Tabela I. Forma geral das funções estimáveis para o modelo (4)

<i>Parâmetro</i>	<i>Coefficiente</i>
μ	l_1
A_1	l_2
A_2	$l_1 - l_2$
B_1	l_4
B_2	$l_1 - l_4$
A_1B_1	l_6
A_1B_2	$l_2 - l_6$
A_2B_1	$l_4 - l_6$
A_2B_2	$l_1 - l_2 - l_4 + l_6$

Tabela II. Funções estimáveis tipo I com dados balanceados para o modelo (4)

<i>Coefficientes do efeito</i>			
<i>Efeito</i>	<i>A</i>	<i>B</i>	<i>A*B</i>
μ	0	0	0
A_1	l_2	0	0
A_2	$-l_2$	0	0
B_1	0	l_4	0
B_2	0	$-l_4$	0
A_1B_1	$0.5l_2$	$0.5l_4$	l_6
A_1B_2	$0.5l_2$	$-0.5l_4$	$-l_6$
A_2B_1	$-0.5l_2$	$0.5l_4$	$-l_6$
A_2B_2	$-0.5l_2$	$-0.5l_4$	l_6

Tabela III. Funções estimáveis tipo I com dados desbalanceados para o modelo (4)

<i>Coefficientes do efeito</i>			
<i>Efeito</i>	<i>A</i>	<i>B</i>	<i>A*B</i>
μ	0	0	0
A_1	l_2	0	0
A_2	$-l_2$	0	0
B_1	$-0.167l_2$	l_4	0
B_2	$0.167l_2$	$-l_4$	0
A_1B_1	$0.333l_2$	$0.4l_4$	l_6
A_1B_2	$0.667l_2$	$-0.4l_4$	$-l_6$
A_2B_1	$-0.5l_2$	$0.6l_4$	$-l_6$
A_2B_2	$-0.5l_2$	$-0.6l_4$	l_6

A forma geral das funções estimáveis é a mesma, sendo os dados balanceados ou não. No entanto, com as FE I o mesmo não ocorre. Note que para dados balanceados, a SQ I para A não envolve o efeito B . O mesmo não acontece quando os dados são desbalanceados. Além disso, a interação apresenta coeficientes simbólicos que darão origem a somas de quadrados que certamente serão de difícil interpretação, também no caso de B (apesar de não envolver A). Vejamos o que está sendo testado em cada uma das duas situações.

O teste tipo I para A com dados balanceados será

$$H_0: A_1 - A_2 + 0.5A_1B_1 + 0.5A_1B_2 - 0.5A_2B_1 - 0.5A_2B_2 = k$$

Para $k = 0$,

$$H_0: A_1 + \frac{A_1B_1 + A_1B_2}{2} = A_2 + \frac{A_2B_1 + A_2B_2}{2}$$

Para os dados desbalanceados, com $k = 0$,

$$H_0: A_1 + \frac{B_2}{6} + \frac{A_1B_1 + 2A_1B_2}{3} = A_2 + \frac{B_1}{6} + \frac{A_2B_1 + A_2B_2}{2}$$

Essa soma ponderada dos níveis dos efeitos do modelo é, sem dúvida, de difícil interpretação.

Para B , as hipóteses com $k = 0$ serão, no caso balanceado,

$$H_0: B_1 + \frac{A_1 B_1 + A_2 B_1}{2} = B_2 + \frac{A_1 B_2 + A_2 B_2}{2}$$

Havendo desbalanceamento,

$$H_0: B_1 + \frac{2A_1 B_1 + 3A_2 B_1}{5} = B_2 + \frac{2A_1 B_2 + 3A_2 B_2}{5}$$

Como para A , o teste acima é de difícil interpretação.

Para $A*B$, no entanto, tais problemas não ocorrem.

Assim, com dados desbalanceados, as FE I para efeitos contidos em outros são dependentes das frequências das caselas. Seu uso nessas condições leva, portanto, ao surgimento de testes que além de não representarem aquilo que se deseja, são difíceis de interpretar.

Pode-se apontar como desvantagens das SQ I:

1. não são invariantes com respeito ao ordenamento dos efeitos no modelo;
2. para delineamentos experimentais desbalanceados, cada hipótese geralmente envolve os parâmetros do efeito sendo testado mais todos os efeitos restantes

que o seguem no modelo e

3. como pode ser demonstrado variando o número de observações por casela, algumas hipóteses tipo I são dependentes das frequências das caselas.

Entretanto, existem modelos para os quais as hipóteses tipo I são consideradas apropriadas:

1. dados balanceados,
2. modelos com fatores embutidos somente,
3. modelos de regressão polinomial e
4. modelos com covariáveis, se estas são especificadas antes das variáveis classificatórias.

9.6 FUNÇÕES ESTIMÁVEIS TIPO II

Este é o método do *delineamento experimental* que Overall e Spiegel sugerem em 1969 ou o método do *ajuste de constantes* criado por Yates em 1934.

As somas de quadrados tipo II (SQ II) são utilizadas para testar hipóteses sobre uma variável, como se ela tivesse sido a última a entrar no modelo. Para um modelo de regressão como $y = x_1\beta_1 + x_2\beta_2 + x_3\beta_3$, as SQ II serão

<i>Efeito</i>	<i>SQ</i>
β_1	$R(\beta_1 \beta_2, \beta_3)$
β_2	$R(\beta_2 \beta_1, \beta_3)$
β_3	$R(\beta_3 \beta_1, \beta_2)$

E, para um modelo como $y_{ijr} = \mu + A_i + B_j + A*B_{ij}$, teremos

<i>Efeito</i>	<i>SQ</i>
A	$R(A \mu, B)$
B	$R(B \mu, A)$
$A*B$	$R(A*B \mu, A, B)$

Para modelos envolvendo interações e fatores embutidos, é impossível obter, para efeitos principais, testes tipo II que deixem de envolver parâmetros de efeitos que os contenham. A SQ II para um efeito E_1 (que pode ser um efeito principal ou interação) é calculada, dado um efeito E_2 se, e somente se, E_2 não contém E_1 . Para o modelo acima, as SQ II para $A|A*B$ e $B|A*B$ não podem ser calculadas, pois $A*B$ contém A e B . SQ II para $A|B$ é válida pois B não contém A . Do mesmo modo,

$B|A$ e $A*B|A,B$. Já foi visto que se E_2 contém E_1 , $R(E_1|E_2) = 0$.

Os conjuntos geradores para as FE II são os mesmos conjuntos geradores para as FE I, quando a variável em questão é a última a entrar no modelo. Assim, G_3^* é um conjunto gerador para as FE II.

A questão que se levanta, agora, é qual o comportamento das SQ II, para dados desbalanceados. Para um fatorial 2^2 , balanceado, com duas replicações por casela, as FE II são iguais às FE I e estão representadas na Tabela II. No entanto, se a casela A_1B_1 contiver somente um dado, as FE II serão as da Tabela IV.

Tabela IV. Funções estimáveis tipo II com dados desbalanceados para um fatorial 2^2

Efeito	Coeficientes do efeito		
	A	B	A*B
μ	0	0	0
A_1	l_2	0	0
A_2	$-l_2$	0	0
B_1	0	l_4	0
B_2	0	$-l_4$	0
A_1B_1	$0.4l_2$	$0.4l_4$	l_6
A_1B_2	$0.6l_2$	$-0.4l_4$	$-l_6$
A_2B_1	$-0.4l_2$	$0.6l_4$	$-l_6$
A_2B_2	$-0.6l_2$	$-0.6l_4$	l_6

No caso das hipóteses tipo II, pode-se notar que o teste para A não envolve parâmetros de B nem o teste para B envolve aqueles de A , mesmo quando o delineamento é desbalanceado. No entanto, nesse último caso as hipóteses testadas podem ter utilidade contestada. Para A ,

$$H_0: A_1 + \frac{2A_1B_1 + 3A_1B_2}{5} = A_2 + \frac{2A_2B_1 + 3A_2B_2}{5}$$

Para B ,

$$H_0: B_1 + \frac{2A_1B_1 + 3A_2B_1}{5} = B_2 + \frac{2A_1B_2 + 3A_2B_2}{5}$$

Como nas FE I, o teste para efeitos que contém outros, como por exemplo $A*B$, não envolve os parâmetros dos efeitos contidos. Aqui também, as hipóteses testadas para efeitos A e B , no caso desbalanceado, dependem das frequências das caselas. No entanto, quando os dados são proporcionais, o comportamento das FE II, como nas FE I, é o mesmo se há desbalanceamento ou não.

As SQ II são apropriadas para:

1. qualquer modelo balanceado,
2. qualquer modelo de regressão e
3. modelos com efeitos principais.

9.7 FUNÇÕES ESTIMÁVEIS TIPO III

As SQ III, bem como as SQ IV, que serão vistas a seguir, são equivalentes ao método dos *quadrados ponderados de médias*, descrito por Yates em 1934, ou ao método de *mínimos quadrados completos*, sugerido por Overall e Spiegel em 1969.

Quando um efeito está contido em outro, as hipóteses tipo II para aquele efeito são dependentes das frequências das caselas. A filosofia por trás das hipóteses tipos III e IV é que o teste realizado para qualquer efeito deve ser o mesmo para todos os delineamentos com a mesma forma geral das funções estimáveis. Assim, para a maioria dos casos de dados desbalanceados, é possível testar o mesmo conjunto de hipóteses que seria testado se os dados fossem balanceados. Para isso, é necessária a determinação de uma matriz L adequada.

Observando as hipóteses tipo II que estão sendo testadas em um fatorial 2^2 , verifica-se que as hipóteses para os efeitos A e $A*B$ são ortogonais. O mesmo ocorre com B e $A*B$. O princípio de ortogonalidade entre um efeito e qualquer efeito que o contenha é mantido para todos os delineamentos balanceados. Note que $\mathbf{l}_A \perp \mathbf{l}_{A*B}$ para os vetores da Tabela II.

$$\begin{bmatrix} 0 & l_2 & -l_2 & 0 & 0 & 0.5l_2 & 0.5l_2 & -0.5l_2 & -0.5l_2 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ l_6 \\ -l_6 \\ -l_6 \\ l_6 \end{bmatrix} = 0$$

O mesmo pode ser verificado para l_B e l_{A+B} .

No entanto, se os dados são desbalanceados, pela Tabela II nota-se que l_A não é ortogonal a l_{A+B} , pois

$$l'_A l_{A+B} = -\frac{1}{3} l_2 l_6$$

Portanto, a forma geral das funções estimáveis deve ser manipulada no sentido de se obter hipóteses que preencham o requisito da ortogonalidade.

Para construir uma hipótese tipo III para um efeito E_1 , partindo da forma geral das funções estimáveis, seguem-se os seguintes passos:

1. Igualar a zero os coeficientes simbólicos dos efeitos do modelo, com exceção de E_1 e daqueles que contém E_1 . Se E_1 não está contido em outro efeito, este passo define a hipótese tipo III para E_1 (como também para os tipos II e IV). Se E_1 está contido em outros efeitos, ir ao passo 2.
2. Igualar a zero os símbolos l_i , presentes em E_1 e nos efeitos que contém E_1 e que foram zerados no passo 1, a fim de se obter uma forma mais simples para esses coeficientes.
3. Igualar os coeficientes simbólicos dos efeitos que contém E_1 , a funções lineares dos símbolos de E_1 , de modo a tornar o vetor simbólico de E_1 ortogonal ao vetor simbólico dos efeitos que o contém.

O exemplo a seguir ilustra a construção de hipóteses tipo III para um fatorial 2^2 sem caselas vazias.

Para obter a hipótese para a interação, $A*B$, começar com a forma geral das funções estimáveis da Tabela I e igualar a zero os coeficientes de μ , A e B , como visto na Tabela V. A última coluna da tabela mostra a forma da hipótese tipo III para $A*B$. Para obter a hipótese tipo III para A , deve-se novamente começar com a forma geral das funções estimáveis e igualar a zero os coeficientes simbólicos de μ e B ($l_1 = l_2 = 0$). A seguir, fazer $l_6 = ql_2$ e encontrar o valor de q que torne a hipótese para A ortogonal à hipótese para $A*B$. Aqui, $q = 0.5$. Esses passos são mostrados na Tabela VI.

Para o modelo $y_{ijr} = \mu + A_i + B_j + A*B_{ij}$, $i, j, r = 2$, as hipóteses tipo III são as

mesmas mostradas na Tabela II e, usando a notação $R(\cdot)$, elas podem ser representadas por

<i>Efeito</i>	<i>SQ</i>
<i>A</i>	$R(A \mu, B, A * B)$
<i>B</i>	$R(B \mu, A, A * B)$
<i>A * B</i>	$R(A * B \mu, A, B)$

Tabela V. Hipótese tipo III para $A * B$ para um fatorial 2^2 sem caselas vazias

<i>Efeito</i>	<i>Forma Geral</i>	$l_1 = l_2 = l_4 = 0$
μ	l_1	0
A_1	l_2	0
A_2	$l_1 - l_2$	0
B_1	l_4	0
B_2	$l_1 - l_4$	0
$A_1 B_1$	l_6	l_6
$A_1 B_2$	$l_2 - l_6$	$-l_6$
$A_2 B_1$	$l_4 - l_6$	$-l_6$
$A_2 B_2$	$l_1 - l_2 - l_4 + l_6$	l_6

Tabela VI. Hipótese tipo III para A

Efeito	Forma Geral	$l_1=l_4=0$	$l_6=ql_2$	$q=0.5$
μ	l_1	0	0	0
A_1	l_2	l_2	l_2	l_2
A_2	l_1-l_2	$-l_2$	$-l_2$	$-l_2$
B_1	l_4	0	0	0
B_2	l_1-l_4	0	0	0
A_1B_1	l_6	l_6	ql_2	$0.5l_2$
A_1B_2	l_2-l_6	l_2-l_6	$(1-q)l_2$	$0.5l_2$
A_2B_1	l_4-l_6	$-l_6$	$-ql_2$	$-0.5l_2$
A_2B_2	$l_1-l_2-l_4+l_6$	$-l_2+l_6$	$(q-1)l_2$	$-0.5l_2$

Um fato de grande importância que merece ser reenfocado é que as hipóteses tipo III são independentes das frequências de cada casela do delineamento. Por essa razão, desde que não existam caselas vazias, as SQ III são as mais adequadas para dados desbalanceados onde existam efeitos contidos em outros.

9.8 FUNÇÕES ESTIMÁVEIS TIPO IV

Quando todas as combinações de tratamentos são observadas (todas as caselas contém dados), as hipóteses tipo III são as mesmas, tanto para os casos onde as frequências de cada casela são iguais, quanto para os casos onde isso não ocorre.

Quando algumas caselas estão vazias, tais hipóteses não podem ser testadas já que elas envolvem parâmetros sobre os quais não temos informações. Por exemplo, para um modelo com dois critérios de classificação e cada fator com três níveis, sejam os dados:

	B_1	B_2	B_3
A_1	2;6		8;6
A_2	3	14	12;9
A_3	6	9	

é impossível estimar funções de interesse como $A_1 - A_3$, já que não se dispõe de informação sobre A_1B_2 e A_3B_3 . Da mesma forma, não se pode estimar funções envolvendo B_2 e B_3 . Assim, é impossível testar $H_0: A_1 = A_2 = A_3$ ou $H_0: B_1 = B_2 = B_3$.

As hipóteses tipos I e II para dados desbalanceados dependem do número de observações em cada casela. Desde que exista pelo menos uma observação para determinada combinação de tratamentos, sua média pode ser estimada. Assim, funções estimáveis dos parâmetros dependem de quais combinações de tratamentos são observadas e não de quantas vezes elas são observadas.

Olhando novamente as hipóteses tipo II para um delineamento fatorial 2^2 balanceado na Tabela II, pode-se notar outra característica das hipóteses com dados desse tipo: os coeficientes dos efeitos principais são distribuídos equitativamente entre os coeficientes dos efeitos de ordem mais alta que contém o efeito em questão. Por

exemplo, para A , os coeficientes de A_1B_1 e A_1B_2 são iguais à metade do coeficiente de A_1 . Os coeficientes de A_2B_1 e de A_2B_2 são iguais à metade do coeficiente de A_2 . Assim, o conceito básico usado para construir as hipóteses tipo IV é que os coeficientes de qualquer efeito E_1 são distribuídos de modo equitativo entre os efeitos de nível mais alto que contém E_1 . No entanto, quando existem caselas vazias, esse princípio deve ser empregado com cuidado.

As hipóteses tipo IV para A referentes aos dados acima serão:

$$1. \frac{A_1B_1 + A_1B_3}{2} = \frac{A_2B_1 + A_2B_3}{2}$$

$$2. \frac{A_2B_1 + A_2B_2}{2} = \frac{A_3B_1 + A_3B_2}{2}$$

$$3. A_1B_1 = A_2B_1$$

$$4. A_1B_1 = A_3B_1$$

$$5. A_2B_1 = A_3B_1$$

$$6. A_2B_2 = A_3B_2$$

$$7. A_1B_3 = A_2B_3$$

Para B existirão hipóteses equivalentes.

Veremos a seguir como construir as FE IV. No entanto, com tal algoritmo estaremos testando as hipóteses 2 e 4 acima. Isso porque as SQ IV para um efeito E_1

testam o último nível de E_1 contra os seus outros níveis. Portanto, as hipóteses 1, 3, 5, 6 e 7 não serão testadas. Reordenar ou renumerar os tratamentos antes de fazer a análise, pode resultar em diferentes hipóteses tipo IV, conforme mostrado por Hudson e Searle (1982). Como será testado o último nível do efeito contra os demais, reordená-los de modo a alterar o nome do último nível, deverá produzir uma hipótese diferente.

Outro fato a se notar é que as hipóteses tipo IV para um delineamento sem dados perdidos são iguais às hipóteses tipo III.

Veremos agora um algoritmo que gera L para o cálculo das SQ IV.

Partindo da forma geral das funções estimáveis, uma hipótese tipo IV para um efeito E_1 pode ser construída da seguinte forma:

1. Igualar a zero os coeficientes dos efeitos principais diferentes de E_1 e dos que não contém E_1 . Se E_1 não está contido em nenhum outro efeito, as hipóteses tipo IV estarão definidas. Caso contrário ir ao passo 2.
2. Simplificar os coeficientes do bloco E_1 e dos efeitos que contém E_1 , igualando a zero os símbolos l_i , presentes nos efeitos principais, que foram igualados a zero no passo 1.
3. Para cada símbolo l_i na sua vez, proceder como abaixo:
 - a. Selecionar o símbolo l_i e colocar todos os outros símbolos $l_{i'}$, $i \neq i'$, iguais a zero.

- b. Se algum nível de E_1 tem zero como coeficiente, igualar a zero todos os coeficientes dos efeitos que contém aquele nível de E_1 .
 - c. Verificar se algum coeficiente de efeito que contém o nível de E_1 em questão é zero, quando o coeficiente desse nível é diferente de zero. Se isso ocorre, as FE IV para E_1 não são únicas.
 - d. Se existirem caselas vazias no último nível de E_1 , igualar a zero os coeficientes dos parâmetros correspondentes, nos níveis de E_1 anteriores ao último.
 - e. Se o nível v de E_1 definido pelo l_i em questão contiver caselas vazias, igualar a zero os coeficientes dos parâmetros correspondentes, nos níveis de E_1 superiores a v .
 - f. Para cada nível de E_1 , na sua vez, se seu coeficiente é diferente de zero, contar o número de vezes em que ele ocorre nos efeitos que contém E_1 e que são diferentes de zero. Igualar cada nível desses efeitos de ordem mais alta, ao coeficiente de E_1 dividido pela contagem.
4. Somar elemento a elemento cada vetor obtido em f.

As SQ IV assumem um papel importante nos delineamentos com dois critérios de classificação e ausência de observações em uma ou mais combinações de tratamentos. Havendo três ou mais critérios em um modelo com interações, dependendo da ordem destas, o comportamento das SQ IV irá variar: com interações de segunda ordem, as SQ IV serão iguais às SQ III. Se o modelo contiver interações de ordem mais alta, os testes gerados podem ter utilidade questionável, além do fato

de que interações de terceira ordem ou superiores quase sempre são de difícil interpretação, na prática.

Tabela VII. Funções estimáveis tipo IV para um fatorial 3×3 com dados desbalanceados

Efeito	Forma Geral	Coeficientes do Efeito		
		A	B	A*B
μ	l_1	0	0	0
A_1	l_2	l_2	0	0
A_2	l_3	l_3	0	0
A_3	$l_1 - l_2 - l_3$	$-l_2 - l_3$	0	0
B_1	l_5	0	l_5	0
B_2	l_6	0	l_6	0
B_3	$l_1 - l_5 - l_6$	0	$-l_5 - l_6$	0
A_1B_1	l_8	l_2	$0.5l_5$	l_8
A_1B_3	$l_2 - l_8$	0	$-0.5l_5$	$-l_8$
A_2B_1	l_{10}	$0.5l_3$	$0.5l_5$	l_{10}
A_2B_2	$-l_1 + l_2 + l_3 + l_5 + l_6 - l_8 - l_{10}$	$0.5l_3$	l_6	$-l_8 - l_{10}$
A_2B_3	$l_1 - l_2 - l_5 - l_6 + l_8$	0	$-0.5l_5 - l_6$	l_8
A_3B_1	$l_5 - l_8 - l_{10}$	$-l_2 - 0.5l_3$	0	$-l_8 - l_{10}$
A_3B_2	$l_1 - l_2 - l_3 - l_5 + l_8 + l_{10}$	$-0.5l_3$	0	$l_8 + l_{10}$

Para os dados do início desta seção, as FE IV serão as da Tabela VII. Ali, pode-se verificar que as hipóteses para A são as de número 2 e 4 citadas acima.

Como as hipóteses tipo IV não são únicas, ao serem usadas através de programas computacionais, é necessário examinar a forma das FE IV. Assim, será possível conhecer quais hipóteses estão sendo testadas.

9.9 COMPARAÇÃO DAS SOMAS DE QUADRADOS

O modelo $y_{ijr} = \mu + A_i + B_j + A * B_{ij}$ pode também ser representado por $y_{ijr} = \mu_{ij}$. Usando tal parametrização, Freund, Littell e Spector (1986), baseados em Speed, Hocking e Hackney (1978), mostram a interpretação das SQ I, II, III e IV para o caso onde os $n_{ij} > 0$, para quaisquer i, j , tornando as SQ III iguais às SQ IV. Isso é mostrado na Tabela VIII. Através de sua análise, é possível verificar alguns fatos citados anteriormente, como por exemplo, a igualdade dos quatro tipos de somas de quadrados quando os dados são balanceados.

Consideremos os quatro tipos de estrutura de dados, fazendo n_{ij} denotar o número de observações no nível i do fator A e nível j do fator B :

1. Caselas com frequências iguais: $n_{ij} = n$, para quaisquer i, j .
2. Caselas com frequências proporcionais: $n_{ij}/n_{ii} = n_{kj}/n_{kl}$, para quaisquer i, j, k, l .
3. Caselas com frequências não proporcionais: $n_{ij}/n_{ii} \neq n_{kj}/n_{kl}$, para algum i, j, k, l , mas $n_{ij} > 0$, para quaisquer i, j .
4. Casela(s) vazia(s): $n_{ij} = 0$ para pelo menos um i, j .

Tabela VIII. Interpretação das somas de quadrados

Efeito A

$$\text{Tipo I: } (\sum_j n_{1j} \mu_{1j}) / n_{1.} = \dots = (\sum_j n_{aj} \mu_{aj}) / n_{a.}$$

$$\text{Tipo II: } \sum_j n_{1j} \mu_{1j} = \sum_i \sum_j (n_{1j} n_{ij} \mu_{ij} / n_j), \dots,$$

$$\sum_j n_{aj} \mu_{aj} = \sum_i \sum_j (n_{aj} n_{ij} \mu_{ij} / n_j)$$

$$\text{Tipos III e IV: } \mu_{1.} = \dots = \mu_{a.}$$

Efeito B

$$\text{Tipos I e II: } \sum_i n_{i1} \mu_{i1} = \sum_i \sum_j (n_{i1} n_{ij} \mu_{ij} / n_i), \dots,$$

$$\sum_i n_{ib} \mu_{ib} = \sum_i \sum_j (n_{ib} n_{ij} \mu_{ij} / n_i)$$

$$\text{Tipos III e IV: } \mu_{.1} = \dots = \mu_{.b}$$

Efeito A*B

$$\text{Tipos I,II,III,IV: } \mu_{ij} - \mu_{im} - \mu_{lj} + \mu_{lm} = 0 \quad \forall i,j,l,m$$

A Tabela IX mostra a comparação dos tipos de somas de quadrados para as várias estruturas de dados onde, por exemplo, a notação III = IV indica que a SQ III é igual à SQ IV.

Tabela IX. Comparação das SQ I, II, III e IV

Estrutura dos Dados				
Efeito	1	2	3	4
A	I=II=III=IV	I=II,III=IV	III=IV	
B	I=II=III=IV	I=II,III=IV	I=II,III=IV	I=II
A*B	I=II=III=IV	I=II,III=IV	I=II=III=IV	I=II=III=IV

9.10 EXEMPLO

Para ilustrar os conceitos introduzidos neste capítulo, tomemos novamente o exemplo da seção 3.9. No Quadro I daquele capítulo, vemos que todas as observações com 'NPELAGRA' possuem o nível de ingestão 'BAIXA'. Assim, se considerarmos no ajuste do modelo a presença de interação, esta terá zero graus de liberdade. Tomemos as observações de números 19, 20, 21, 22 e substituamos 'BAIXA' por 'MEDIA'. Com esses dados, ajustemos um modelo com os efeitos principais 'PACIENTE' e 'INGESTAO' e a interação 'PACIENTE*INGESTAO'. O seguinte programa² SAS[®] ajustará esse modelo:

```

PROC GLM;
  CLASS PACIENTE INGESTAO;
  MODEL VALINA = PACIENTE INGESTAO PACIENTE*INGESTAO / E E1 E2 E3 E4;
RUN;
    
```

Com esse modelo, uma combinação de tratamentos ('NPELAGRA', 'ALTA') não

²As opções E1, E2, E3 e E4 do comando MODEL fornece a forma das FE I, II, III e IV, respectivamente.

contém dados. Assim, a Soma de Quadrados tipo IV será útil na análise. Os resultados encontram-se nos Quadros I a XV abaixo.

No Quadro I, temos a forma geral das funções estimáveis que é a mesma, quer os dados sejam balanceados, quer não.

No Quadro II, nota-se que, em razão da estrutura desbalanceada dos dados, as FE I para PACIENTE envolvem parâmetros que o seguem no modelo (ver comando MODEL acima). Além disso, não se estará testando $H_0: NPELAGRA = PELAGRA$. O mesmo ocorre com as FE I para INGESTAO, como se vê no Quadro III.

Para a interação PACIENTE*INGESTAO, a forma das FE I se encontra no Quadro IV. Como discutido no capítulo, apesar da estrutura desbalanceada, o teste para a interação não envolve outros parâmetros.

Nos Quadros V e VI, temos as FE II para PACIENTE*INGESTAO. Os testes para um deles não envolvem os parâmetros do outro, no entanto, ambos envolvem parâmetros da interação. Isso torna tais testes difíceis de serem interpretados.

Como para as FE I, as FE II para a interação não envolvem outros parâmetros (Quadro VII). Note que a forma das FE I e II para a interação é a mesma.

Comparando os Quadros VII e X, verifica-se que as FE III para PACIENTE são ortogonais às FE III para o efeito no modelo que o contém, a interação. Assim, os testes realizados através das SQ III com dados desbalanceados são equivalentes aos testes realizados com dados balanceados, mesmo se o efeito sendo testado, está contido em outro.

A comparação dos Quadros IX e X mostra, de modo semelhante, a ortogonalidade entre as FE III para INGESTAO e a interação PACIENTE*INGESTAO.

No Quadro XI, pode-se verificar alguns aspectos discutidos com relação às FE IV: os testes para PACIENTE não envolvem parâmetros de INGESTAO, são ortogonais aos testes para a interação (Quadro XIII) e os coeficientes simbólicos são equitativamente distribuídos entre os efeitos que contém cada nível de PACIENTE.

Para o Quadro XII, valem as mesmas considerações. Note que as hipóteses para INGESTAO tem 2 graus de liberdade e a distribuição equitativa dos coeficientes simbólicos ocorre para cada grau de liberdade considerado.

Note a existência de outras FE IV para PACIENTE já que o coeficiente simbólico de PELAGRA ALTA é 0, enquanto que o coeficiente de PELAGRA é -L2. O mesmo pode ser verificado para INGESTAO, considerando cada grau de liberdade separadamente.

A tabela de análise de variância referente ao ajuste do modelo é vista no Quadro XIV. No Quadro XV, os valores numéricos das somas de quadrados são mostrados.

Quadro I. Forma geral das funções estimáveis

General Linear Models Procedure General Form of Estimable Functions		
Effect		Coefficients
INTERCEPT		L1
PACIENTE	NPELAGRA	L2
	PELAGRA	L1-L2
INGESTAO	ALTA	L4
	BAIXA	L5
	MEDIA	L1-L4-L5
PACIENTE*INGESTAO	NPELAGRA BAIXA	L7
	NPELAGRA MEDIA	L2-L7
	PELAGRA ALTA	L4
	PELAGRA BAIXA	L5-L7
	PELAGRA MEDIA	L1-L2-L4-L5+L7

Quadro II. Forma das funções estimáveis tipo I para PACIENTE

General Linear Models Procedure Type I Estimable Functions for : PACIENTE		
Effect		Coefficients
INTERCEPT		0
PACIENTE	NPELAGRA	L2
	PELAGRA	-L2
INGESTAO	ALTA	-0.4*L2
	BAIXA	0.359*L2
	MEDIA	0.041*L2
PACIENTE*INGESTAO	NPELAGRA BAIXA	0.6923*L2
	NPELAGRA MEDIA	0.3077*L2
	PELAGRA ALTA	-0.4*L2
	PELAGRA BAIXA	-0.3333*L2
	PELAGRA MEDIA	-0.2667*L2

Quadro III. Forma das funções estimáveis tipo I para INGESTAO

General Linear Models Procedure		
Type I Estimable Functions for: INGESTAO		
Effect	Coefficients	
INTERCEPT		0
PACIENTE	NPELAGRA	0
	PELAGRA	0
INGESTAO	ALTA	L4
	BAIXA	L5
	MEDIA	-L4-L5
PACIENTE*INGESTAO	NPELAGRA BAIXA	0.3082*L4+0.5548*L5
	NPELAGRA MEDIA	-0.3082*L5-0.5548*L5
	PELAGRA ALTA	L4
	PELAGRA BAIXA	-0.3082*L4+0.4452*L5
	PELAGRA MEDIA	-0.6918*L4-0.4452*L5

Quadro IV. Forma das funções estimáveis tipo I para a interação PACIENTE*INGESTAO

General Linear Models Procedure		
Type I Estimable Functions for: PACIENTE*INGESTAO		
Effect	Coefficients	
INTERCEPT		0
PACIENTE	NPELAGRA	0
	PELAGRA	0
INGESTAO	ALTA	0
	BAIXA	0
	MEDIA	0
PACIENTE*INGESTAO	NPELAGRA BAIXA	L7
	NPELAGRA MEDIA	-L7
	PELAGRA ALTA	0
	PELAGRA BAIXA	-L7
	PELAGRA MEDIA	L7

Quadro V. Forma das funções estimáveis tipo II para PACIENTE

General Linear Models Procedure		
Type II Estimable Functions for: PACIENTE		
Effect	Coefficients	
INTERCEPT		0
PACIENTE	NPELAGRA	L2
	PELAGRA	-L2
INGESTAO	ALTA	0
	BAIXA	0
	MEDIA	0
PACIENTE*INGESTAO	NPELAGRA BAIXA	0.6164*L2
	NPELAGRA MEDIA	0.3836*L2
	PELAGRA ALTA	0
	PELAGRA BAIXA	-0.6164*L2
	PELAGRA MEDIA	-0.3836*L2

Quadro VI. Forma das funções estimáveis tipo II para INGESTAO

General Linear Models Procedure		
Type II Estimable Functions for: INGESTAO		
Effect	Coefficients	
INTERCEPT		0
PACIENTE	NPELAGRA	0
	PELAGRA	0
INGESTAO	ALTA	L4
	BAIXA	L5
	MEDIA	-L4-L5
PACIENTE*INGESTAO	NPELAGRA BAIXA	0.3082*L4+0.5548*L5
	NPELAGRA MEDIA	-0.3082*L4-0.5548*L5
	PELAGRA ALTA	L4
	PELAGRA BAIXA	-0.3082*L4+0.4452*L5
	PELAGRA MEDIA	-0.6918*L4-0.4452*L5

Quadro VII. Forma das funções estimáveis tipo II para a interação PACIENTE*INGESTAO

General Linear Models Procedure		
Type II Estimable Functions for: PACIENTE*INGESTAO		
Effect	Coefficients	
INTERCEPT		0
PACIENTE	NPELAGRA	0
	PELAGRA	0
INGESTAO	ALTA	0
	BAIXA	0
	MEDIA	0
PACIENTE*INGESTAO	NPELAGRA BAIXA	L7
	NPELAGRA MEDIA	-L7
	PELAGRA ALTA	0
	PELAGRA BAIXA	-L7
	PELAGRA MEDIA	L7

Quadro VIII. Forma das funções estimáveis tipo III para PACIENTE

General Linear Models Procedure		
Type III Estimable Functions for: PACIENTE		
Effect	Coefficients	
INTERCEPT		0
PACIENTE	NPELAGRA	L2
	PELAGRA	-L2
INGESTAO	ALTA	0
	BAIXA	0
	MEDIA	0
PACIENTE*INGESTAO	NPELAGRA BAIXA	0.5*L2
	NPELAGRA MEDIA	0.5*L2
	PELAGRA ALTA	0
	PELAGRA BAIXA	-0.5*L2
	PELAGRA MEDIA	-0.5*L2

Quadro IX. Forma das funções estimáveis tipo III para INGESTAO

General Linear Models Procedure		
Type III Estimable Functions for: INGESTAO		
Effect	Coefficients	
INTERCEPT		0
PACIENTE	NPELAGRA	0
	PELAGRA	0
INGESTAO	ALTA	L4
	BAIXA	L5
	MEDIA	-L4-L5
PACIENTE*INGESTAO	NPELAGRA BAIXA	$0.25*L4 + 0.5*L5$
	NPELAGRA MEDIA	$-0.25*L4 - 0.5*L5$
	PELAGRA ALTA	L4
	PELAGRA BAIXA	$-0.25*L4 + 0.5*L5$
	PELAGRA MEDIA	$-0.75*L4 - 0.5*L5$

Quadro X. Forma das funções estimáveis tipo III para a interação PACIENTE*INGESTAO

General Linear Models Procedure		
Type III Estimable Functions for: PACIENTE*INGESTAO		
Effect	Coefficients	
INTERCEPT		0
PACIENTE	NPELAGRA	0
	PELAGRA	0
INGESTAO	ALTA	0
	BAIXA	0
	MEDIA	0
PACIENTE*INGESTAO	NPELAGRA BAIXA	L7
	NPELAGRA MEDIA	-L7
	PELAGRA ALTA	0
	PELAGRA BAIXA	-L7
	PELAGRA MEDIA	L7

Quadro XI. Forma das funções estimáveis tipo IV para PACIENTE

General Linear Models Procedure Type IV Estimable Functions for: PACIENTE		
Effect	Coefficients	
INTERCEPT		0
PACIENTE	NPELAGRA	L2
	PELAGRA	-L2
INGESTAO	ALTA	0
	BAIXA	0
	MEDIA	0
PACIENTE*INGESTAO	NPELAGRA BAIXA	0.5*L2
	NPELAGRA MEDIA	0.5*L2
	PELAGRA ALTA	0
	PELAGRA BAIXA	-0.5*L2
	PELAGRA MEDIA	-0.5*L2

NOTE: Other type IV estimable functions exist.

Quadro XII. Forma das funções estimáveis tipo IV para INGESTAO

General Linear Models Procedure Type IV Estimable Functions for: INGESTAO		
Effect	Coefficients	
INTERCEPT		0
PACIENTE	NPELAGRA	0
	PELAGRA	0
INGESTAO	ALTA	L4
	BAIXA	L5
	MEDIA	-L4-L5
PACIENTE*INGESTAO	NPELAGRA BAIXA	0.5*L5
	NPELAGRA MEDIA	-0.5*L5
	PELAGRA ALTA	L4
	PELAGRA BAIXA	0.5*L5
	PELAGRA MEDIA	-L4-0.5*L5

NOTE: Other type IV estimable functions exist.

Quadro XIII. Forma das funções estimáveis tipo IV para a interação PACIENTE*INGESTAO

General Linear Models Procedure		
Type IV Estimable Functions for: PACIENTE*INGESTAO		
Effect	Coefficients	
INTERCEPT		0
PACIENTE	NPELAGRA	0
	PELAGRA	0
INGESTAO	ALTA	0
	BAIXA	0
	MEDIA	0
PACIENTE*INGESTAO	NPELAGRA BAIXA	L7
	NPELAGRA MEDIA	-L7
	PELAGRA ALTA	0
	PELAGRA BAIXA	-L7
	PELAGRA MEDIA	L7

Quadro XIV. Tabela de análise de variância

General Linear Models Procedure					
Dependent Variable: VALINA					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	0.08669628	0.02167407	12.29	0.0001
Error	23	0.04056826	0.00176384		
C. Total	27	0.12726454			
	R-Square		Root MSE		VALINA Mean
	0.681229		0.04199806		0.21510714

Quadro XV. Somas de quadrados tipos I, II, III e IV

General Linear Models Procedure					
Dependent Variable: VALINA					
Source	DF	Type I SS	Mean Square	F Value	Pr > F
PACIENTE	1	0.04652261	0.04652261	26.38	0.0001
INGESTAO	2	0.03330229	0.01665115	9.44	0.0010
PACIENTE*INGESTAO	1	0.00687138	0.00687138	3.90	0.0605
Source	DF	Type II SS	Mean Square	F Value	Pr > F
PACIENTE	1	0.06809749	0.06809749	38.61	0.0001
INGESTAO	2	0.03330229	0.01665115	9.44	0.0010
PACIENTE*INGESTAO	1	0.00687138	0.00687138	3.90	0.0605
Source	DF	Type III SS	Mean Square	F Value	Pr > F
PACIENTE	1	0.07457507	0.07457507	42.28	0.0001
INGESTAO	2	0.03561274	0.01780637	10.10	0.0007
PACIENTE*INGESTAO	1	0.00687138	0.00687138	3.90	0.0605
Source	DF	Type IV SS	Mean Square	F Value	Pr > F
PACIENTE	1*	0.07457507	0.07457507	42.28	0.0001
INGESTAO	2*	0.04014940	0.02007470	11.38	0.0004
PACIENTE*INGESTAO	1	0.00687138	0.00687138	3.90	0.0605

* NOTE: Other Type IV Testable Hypotheses exist which may yield different SS.

Capítulo 10

Distribuição de Formas Quadráticas

10.1 INTRODUÇÃO

Técnicas de análise de variância envolvem a quebra da soma de quadrados total de um modelo em componentes cujas razões (sob condições apropriadas) levam à estatística F , adequada para testar certas hipóteses. Esses componentes são formas quadráticas, $\mathbf{y}'\mathbf{A}\mathbf{y}$, onde \mathbf{y} é um vetor de valores observados e \mathbf{A} uma matriz simétrica, positiva semidefinida. Alguns teoremas relacionados a essas grandezas serão necessários. Passaremos agora a estudá-los.

10.2 DISTRIBUIÇÃO DE UMA FORMA QUADRÁTICA

No teorema a seguir surgirá a distribuição qui quadrado não central, que representaremos por $\chi^2(n, \lambda)$, onde n é o número de graus de liberdade e λ , o

parâmetro de não centralidade. A prova apresentada é a dada por Searle (1971).

► **Teorema 1.** Quando $y \sim N(\mu, V)$, então $y'Ay$ tem distribuição $\chi^2[p(A), \frac{1}{2}\mu'A\mu]$ se, e somente se, AV é idempotente.

Prova. AV idempotente $\Rightarrow y'Ay \sim \chi^2[p(A), \frac{1}{2}\mu'A\mu]$.

Seja $M(t)$ a função geratriz de momentos de $y'Ay$.

$$\begin{aligned} M(t) &= E(e^{ty'Ay}) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} e^{ty'Ay} f(y_1, \dots, y_n) dy_1 \dots dy_n = \\ &= (2\pi)^{-\frac{1}{2}n} |V|^{-\frac{1}{2}} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \exp\left\{ty'Ay - \frac{1}{2}(y - \mu)'V^{-1}(y - \mu)\right\} dy_1 \dots dy_n = \\ &= \frac{e^{-\frac{1}{2}\mu'V^{-1}\mu}}{(2\pi)^{\frac{n}{2}} |V|^{\frac{1}{2}}} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}y'(I - 2tAV)V^{-1}y + \mu'V^{-1}y\right\} dy_1 \dots dy_n = \\ &= \frac{e^{-\frac{1}{2}\mu'V^{-1}\mu} (2\pi)^{\frac{n}{2}} |V(I - 2AV)^{-1}|^{\frac{1}{2}} \exp\left\{\frac{1}{2}\mu'V^{-1}V(I - 2tAV)^{-1}V^{-1}\mu\right\}}{(2\pi)^{\frac{n}{2}} |V|^{\frac{1}{2}}} \end{aligned}$$

onde o numerador vem da integral da densidade multivariada $N(\emptyset, [(I - 2tAV)V^{-1}]^{-1})$. Isso fornece

$$\begin{aligned}
 M(t) &= |I - 2tAV|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\mu'[I - (I - 2tAV)^{-1}]V^{-1}\mu\right\} \\
 &= \prod_{i=1}^n (1 - 2t\lambda_i)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\mu'\left[-\sum_{h=1}^{\infty} (2t)^h (AV)^h\right]V^{-1}\mu\right\}
 \end{aligned} \tag{1}$$

onde λ_i , $i = 1, \dots, n$, são os autovalores de AV . Se AV é idempotente e de posto $p(AV) = k$, então k valores de λ_i são iguais a um, $n - k$ são zero e $(AV)^k = AV$. Assim,

$$\begin{aligned}
 M(t) &= \prod_{i=1}^k (1 - 2t)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\mu'\left[-\sum_{h=1}^{\infty} (2t)^h\right]AVV^{-1}\mu\right\} = \\
 &= (1 - 2t)^{-\frac{1}{2}k} \exp\left\{-\frac{1}{2}\mu'[1 - (1 - 2t)^{-1}]A\mu\right\} = \\
 &= (1 - 2t)^{-\frac{1}{2}k} \exp\left\{-\frac{1}{2}\mu'A\mu[1 - (1 - 2t)^{-1}]\right\}
 \end{aligned} \tag{2}$$

(2) é a função geratriz de momentos de uma variável aleatória com distribuição $\chi^2\left[k, \frac{1}{2}\mu'A\mu\right]$. Como V é não singular, $p(AV) = p(A)$. Portanto, $y'Ay - \chi^2\left[p(A), \frac{1}{2}\mu'A\mu\right]$.

$y' Ay \sim \chi^2 \left[k, \frac{1}{2} \mu' A \mu \right] \Rightarrow AV$ é idempotente.

A função geratriz de momentos de uma variável aleatória com distribuição qui quadrado não central é dada por (2) e é equivalente à expressão (1). Em particular, para $\mu = \emptyset$, igualando (1) e (2),

$$M(t) = (1 - 2t)^{-\frac{1}{2}k} = |I - 2tAV|^{-\frac{1}{2}}$$

Fazendo $u = 2t$, temos

$$\begin{aligned} (1 - u)^k &= |I - uAV| \\ &= \prod_{i=1}^n (1 - u\lambda_i) \end{aligned}$$

Como essa expressão é uma identidade em u , no membro direito não existe potência de u maior que k . Portanto, pelo menos um dos $\lambda_i = 0$. O uso repetido desse argumento mostra que $n - k$ dos λ_i são iguais a zero, de modo que

$$(1 - u)^k = \prod_{i=1}^k (1 - u\lambda_i) \Rightarrow k \ln(1 - u) = \sum_{i=1}^k \ln(1 - u\lambda_i) \quad (3)$$

(3) é um polinômio em u e, como dois polinômios são iguais quando seus coeficientes são iguais, igualando termo a termo a expressão à direita de (3),

$$\ln(1-u) = \ln(1-u\lambda_i) \rightarrow 1-u = 1-u\lambda_i \rightarrow u = u\lambda_i$$

Isso fornece $\lambda_i = 1$ para k dos autovalores e $\lambda_i = 0$ para os $n-k$ restantes.

Assim, AV é idempotente. ◀

► **Corolário.** Quando $y \sim N(\emptyset, I)$, $y'Ay$ tem distribuição qui quadrado central com k graus de liberdade, denotada por χ_k^2 , se, e somente se, A é idempotente e de posto k .

A prova do corolário é imediata com a substituição de V por I . ◀

10.3 INDEPENDÊNCIA DE FORMAS QUADRÁTICAS

Além do Teorema 1, alguns outros são necessários na construção de testes de significância. Tais teoremas dizem respeito à independência de formas quadráticas .

► **Teorema 2.** Sejam A_1, \dots, A_q matrizes simétricas, $n \times n$, positivas semidefinidas com $p(A_i) = k_i$. Quaisquer duas das condições seguintes implicam a terceira:

1. A_1, \dots, A_q são idempotentes.
2. $D = \sum_{i=1}^q A_i$, $p(D) = k$, é idempotente.
3. $A_i A_j = \emptyset$, para quaisquer i, j , $i \neq j$.

Prova. a) 1 e 2 \Rightarrow 3

Seja $A_0 = I - D$. A_0 é idempotente pois

$$(I - D)^2 = I - 2D + D^2$$

e como $D^2 = D$,

$$A_0^2 = (I - D)^2 = I - D = A_0$$

Além disso,

$$p(A_0) = p(I - D) = \text{tr}(I - D) = \text{tr}(I) - \text{tr}(D) = n - k$$

Por outro lado, $A_0 + A_1 + \dots + A_q = I$ e, portanto, $p(A_0 + A_1 + \dots + A_q) = n$. Então,

$$p(A_1) + \dots + p(A_q) = n - (n - k) = k = p(D)$$

ou seja, $\sum_{i=1}^q p(A_i) = p(D)$. Esse resultado é importante por si mesmo. Agora,

$$\begin{aligned} p(I - A_i - A_j) &= p\left(\sum_{l=0}^q A_l - A_i - A_j\right) = p\left(\sum_{l \neq i, j} A_l\right) \leq \sum_{l \neq i, j} p(A_l) = \\ &= n - k + \sum_{l=1}^q k_l - k_i - k_j = n - k + k - k_i - k_j = n - k_i - k_j = \\ &= p(I) - p(A_i) - p(A_j) = p(I) - [p(A_i) + p(A_j)] \leq \\ &= p(I) - p(A_i + A_j) \end{aligned}$$

Portanto,

$$\begin{aligned} p(I - A_i - A_j) \leq p(I) - p(A_i + A_j) &\Rightarrow p(I - A_i - A_j) + p(A_i + A_j) \leq \\ &\leq p(I) = n \end{aligned}$$

Porém,

$$\begin{aligned} A_i + A_j + I - A_i - A_j &= I \Rightarrow p(A_i + A_j + I - A_i - A_j) = p(I) = n \Rightarrow \\ &\Rightarrow p(A_i + A_j) + p(I - A_i - A_j) \geq n \end{aligned}$$

e assim,

$$p(A_i + A_j) + p(I - A_i - A_j) = n$$

A partir dessa expressão, $A_i + A_j$ é idempotente pois

$$\begin{aligned} p(A_i + A_j) &= n - p(I - A_i - A_j) = n - p[(I - A_i) - A_j] = \\ &= n - [p(I - A_i) - p(A_j)] = n - [p(I) - p(A_i) - p(A_j)] = \\ &= n - n + k_i + k_j = k_i + k_j \end{aligned}$$

já que $I - A_i$ é idempotente. Como

$$(A_i + A_j)^2 = A_i^2 + A_j^2 + A_i A_j + A_j A_i = A_i + A_j + A_i A_j + A_j A_i = A_i + A_j$$

então, $A_i A_j + A_j A_i = \emptyset$. Logo,

$$(A_i A_j + A_j A_i) A_j = A_i A_j + A_j A_i A_j = \emptyset \quad (4)$$

e

$$A_j (A_i A_j + A_j A_i A_j) = 2A_j A_i A_j = \emptyset \rightarrow A_j A_i A_j = \emptyset$$

Substituindo em (4), $A_i A_j = \emptyset$.

b) 1 e 3 \Rightarrow 2

$$D^2 = \left(\sum_{i=1}^q A_i \right)^2 = \sum_{i=1}^q A_i^2 + \sum_{i \neq j} A_i A_j = \sum_{i=1}^q A_i^2 = \sum_{i=1}^q A_i = D$$

c) 2 e 3 \Rightarrow 1

$$DA_i = (A_1 + \dots + A_q)A_i = A_1 A_i + \dots + A_i^2 + \dots + A_q A_i = A_i^2 \quad (5)$$

Como D é idempotente, de (5) temos

$$\begin{aligned} DA_i = A_i^2 &\Rightarrow D^2 A_i = A_i^2 \Rightarrow D(DA_i) = A_i^2 \Rightarrow DA_i^2 = A_i^2 \Rightarrow \\ &\Rightarrow (DA_i)A_i = A_i^2 \Rightarrow A_i^2 A_i = A_i^2 \Rightarrow A_i^3 = A_i^2 \end{aligned}$$

Então,

$$A_i^2(I - A_i) = A_i^2 - A_i^3 = \emptyset$$

Como

$$p(A_i^2) = p(A_i A_i) = p(A_i' A_i) = p(A_i)$$

$A_i = GA_i^2$, para alguma matriz G não singular. Porém,

$$A_i^2(I - A_i) = \emptyset \rightarrow GA_i^2(I - A_i) = \emptyset \rightarrow A_i(I - A_i) = \emptyset \rightarrow A_i = A_i^2 \quad \blacktriangleleft$$

► Teorema 3. Seja $y \sim N(\mu, V)$. Então, $y'Ay$ e Cy são independentes se, e somente se, $CVA = \emptyset$.

Prova. $CVA = \emptyset \Rightarrow$ independência.

Como A é simétrica, $A = LL'$ para algum L de posto coluna completo. Dessa forma,

$$CVA = \emptyset \rightarrow CVLL' = \emptyset \rightarrow CVLL'L(L'L)^{-1} = \emptyset \rightarrow CVL = \emptyset$$

pois $(L'L)^{-1}$ existe. Mas $cov(Cy, y'L) = CVL = \emptyset$. Como y é um vetor de variáveis aleatórias com distribuição normal, Cy e $y'L$ são independentes. Consequentemente, Cy e $y'Ay = y'LL'y$ são independentes.

$y'Ay$ e Cy independentes $\Rightarrow CVA = \emptyset$.

Independência implica que $cov(Cy, y'Ay) = \emptyset$. Mas

$$cov(Cy, y'Ay) = E[(Cy - C\mu)(y'Ay - E(y'Ay))]$$

Levando em conta que

$$\begin{aligned}
E(y'Ay) &= \text{tr} E(y'Ay) = E[\text{tr}(y'Ay)] = E[\text{tr}(Ayy')] = \\
&= \text{tr}[AE(yy')] = \text{tr}(AV + A\mu\mu') = \text{tr}(AV) + \text{tr}(A\mu\mu') = \\
&= \text{tr}(AV) + \mu'A\mu
\end{aligned}$$

então,

$$\begin{aligned}
\text{cov}(Cy, y'Ay) &= E[(Cy - C\mu)(y'Ay - \mu'A\mu - \text{tr}(AV))] = \\
&= E\{(Cy - C\mu)[(y - \mu)'A(y - \mu) + 2(y - \mu)'A\mu - \text{tr}(AV)]\} = \\
&= \emptyset + 2CVA\mu - \emptyset
\end{aligned}$$

Como $2CVA\mu = \emptyset$ deve valer para todo μ , então, $CVA = \emptyset$. ◀

► **Teorema 4.** Se $y \sim N(\mu, V)$ e A e B são matrizes simétricas positivas semidefinidas, então $y'Ay$ e $y'By$ são independentes se, e somente se, $AVB = \emptyset$.

Prova. $AVB = \emptyset \Rightarrow$ independência

Como A e B são matrizes simétricas, $A = LL'$ e $B = MM'$ para L e M de posto coluna completo. Dessa forma,

$$\begin{aligned}
AVB = \emptyset &= LL'VMM' = \emptyset \Rightarrow (LL')^{-1}L'LL'VMM'M(M'M)^{-1} = \emptyset \\
&\Rightarrow L'VM = \emptyset
\end{aligned}$$

pois $(L'L)^{-1}$ e $(M'M)^{-1}$ existem. Mas $cov(L'y, y'M) = L'VM = \emptyset$.

Como y é um vetor de variáveis aleatórias com distribuição normal, $L'y$ e $y'M$ são independentes. Consequentemente, $y'LL'y = y'Ay$ e $y'MM'y = y'By$ são independentes.

$y'Ay$ e $y'By$ independentes $\Rightarrow AVB = \emptyset$

Independência implica que $cov(y'Ay, y'By) = 0$. Mas

$$cov(y'Ay, y'By) = E[(y'Ay - E(y'Ay))(y'By - E(y'By))]$$

Como $E(y'Ay) = \text{tr}(AV) + \mu'A\mu$ e $E(y'By) = \text{tr}(BV) + \mu'B\mu$, então,

$$\begin{aligned} cov(y'Ay, y'By) &= E\{[y'Ay - \mu'A\mu - \text{tr}(AV)][y'By - \mu'B\mu - \text{tr}(BV)]\} = \\ &= E\{[(y - \mu)'A(y - \mu) + 2(y - \mu)'A\mu - \text{tr}(AV)] \times \\ &\quad \times [(y - \mu)'B(y - \mu) + 2(y - \mu)'B\mu - \text{tr}(BV)]\} = \\ &= 4\mu'AVB\mu = 0 \end{aligned}$$

Como $4\mu'AVB\mu = \emptyset$ deve valer para qualquer μ , então, $AVB = \emptyset$. ◀

► **Teorema de Cochran.** Sejam o vetor $n \times 1$ $y \sim N(\emptyset, \sigma^2 I)$, as matrizes A_i simétricas positivas semidefinidas, $i = 1, \dots, q$ e $p(A_i) = k_i$ os postos dessas matrizes.

a) As variáveis aleatórias $y'A_i y$ são independentes se, e somente se,

1. A_i é idempotente e

2. $A_i A_j = \emptyset$, para quaisquer i, j , $i \neq j$.

b) Se $y'y = \sum_{i=1}^q y'A_i y$, $y'A_i y \sim \chi_{k_i}^2$ se, e somente se, $\sum_{i=1}^q k_i = n$.

A aplicação dos Teoremas 1, 2 e 4 provam o Teorema de Cochran. ◀

10.4 APLICAÇÃO AOS MODELOS LINEARES

Consideremos o modelo

$$\begin{cases} y = X\beta + e, \\ e \sim N(\emptyset, \sigma^2 I) \end{cases} \quad (6)$$

que denominaremos *modelo de Gauss-Markov normal* (MGMN). Ele difere dos modelos vistos anteriormente, em razão da presença, aqui, da normalidade e das

condições de Gauss-Markov impostas ao erro.

Vejamos as consequências de se assumir o MGMN sobre as SQ I, II, III e IV geradas, respectivamente, pelas FE I, II, III e IV.

► **Teorema 5.** Seja $l'\beta$ estimável, tal que $a'X = l'$, para algum vetor a' . No MGMN $l'b \sim N(l'\beta, \sigma^2 l'BB'l)$, onde b é solução das equações normais e B , tal que $X'XB = X'$.

Prova. O melhor estimador linear não viciado de $l'\beta = (l'\beta)_b = l'b$. Como $e \sim N(\emptyset, \sigma^2 I)$, $y \sim N(X\beta, \sigma^2 I)$, pois $y = X\beta + e$. Portanto,

$$l'b = l'BB'X'y = a'XBB'X'y = a'XBXBy = a'XBy$$

e como

$$E(l'b) = E(a'XBy) = a'XBX\beta = a'X\beta = l'\beta$$

e

$$\text{var}(l'b) = \text{var}(a'XBy) = \sigma^2 a'XBB'X'a = \sigma^2 l'BB'l$$

$$l'b \sim N(l'\beta, \sigma^2 l'BB'l).$$

► Corolário. $l'b - m \sim N(l'\beta - m, \sigma^2 l'BB'l)$ ◀

► Teorema 6. Se $l'\beta$ é estimável, então, no MGMN, $l'b$ e a soma de quadrados do resíduo $y'(I - XBB'X')y$ são independentes.

Prova. $l'b = a'XBB'X'y = a'XBXB'y = (B'X'a)'y$. A independência existirá pelo Teorema 3 se, e somente se, $(B'X'a)'(I - XBB'X') = \emptyset$. Mas

$$\begin{aligned} (B'X'a)'(I - XBB'X') &= a'XB - a'XBXB'X' \\ &= a'XB - a'XBXB = a'XB - a'XB = \emptyset \end{aligned} \quad \blacktriangleleft$$

► Teorema 7. No MGMN, onde o posto de X , $p(X) = k$ e n é o número de observações, $\frac{(n-k)S^2}{\sigma^2} = \frac{y'(I - XBB'X')y}{\sigma^2} \sim \chi_{n-k}^2$.

Prova. $P_X = XBB'X'$ é uma matriz simétrica, idempotente e com $p(P_X) = k$.

A matriz $I - P_X$ é simétrica, idempotente e

$$p(I - P_X) = p(I) - p(P_X) = n - k$$

Usando o Teorema 1,

$$\frac{y'(I - XBB'X')y}{\sigma^2} \sim \chi_{n-k}^2$$

O Teorema 1 exige $E(y) = \emptyset$ e, no entanto, $E(y) = X\beta$. Porém, $(y - X\beta) \sim N(\emptyset, \sigma^2 I)$ e, portanto, $\frac{(y - X\beta)}{\sigma} \sim N(\emptyset, I)$. Mas

$$(y - X\beta)'(I - XBB'X')(y - X\beta) = y'(I - XBB'X')y$$

pois

$$\begin{aligned} \beta'X'(I - XBB'X')y &= \beta'(X'y - X'XBB'X'y) = \beta'(X'y - X'B'X'y) = \\ &= \beta'(X'y - X'y) = 0 \end{aligned}$$

e

$$\beta'X'(I - XBB'X')X\beta = \beta'(X'X - X'XBB'X'X)\beta = \beta'(X'X - X'X)\beta = 0$$

► **Corolário.** No MGMN, $S^2 = y'(I - XBB'X')y/(n - k)$ é um estimador não viciado de σ^2 e é independente do melhor estimador linear não viciado de qualquer função estimável.

Prova.

$$\begin{aligned} E \frac{y'(I - XBB'X')y}{n-k} &= \frac{\text{tr}[(I - XBB'X')\sigma^2] + \beta'X'(I - XBB'X')X\beta}{n-k} = \\ &= \frac{(n-k)\sigma^2 + (\beta'X'X\beta - \beta'X'XBB'X'X\beta)}{n-k} = \\ &= \frac{(n-k)\sigma^2 + (\beta'X'X\beta - \beta'X'X\beta)}{n-k} = \sigma^2 \end{aligned}$$

A prova da independência segue da prova do Teorema 6. ◀

10.5 ANÁLISE DE VARIÂNCIA NO MODELO DE GAUSS-MARKOV NORMAL

Vimos no Capítulo 3 que a aproximação de y por combinação linear de vetores do espaço coluna de X dava origem a vetores cujos comprimentos ao quadrado eram formas quadráticas. Sob o modelo

$$\begin{cases} y = X\beta + e \\ e \sim N(\mu, \sigma^2 I) \end{cases}$$

essas formas quadráticas terão distribuição qui quadrado não central e razões entre elas darão origem a estatísticas com distribuição conhecida. Seja a soma de quadrados

associada ao ajuste do modelo, denotada por SQM ; a soma de quadrados associada ao resíduo, por SQR ; k e $n-k$ os respectivos graus de liberdade, onde $k = p(X)$. Em razão dos Teoremas 1 e 7, SQM/σ^2 terá distribuição qui quadrado não central com k graus de liberdade. SQR/σ^2 terá distribuição qui quadrado central com $n-k$ graus de liberdade. Como essas duas razões são independentes, pela aplicação do Teorema 4, por definição

$$F = \frac{SQM/k}{SQR/(n-k)}$$

terá distribuição F não central, com k e $n-k$ graus de liberdade e parâmetro de não centralidade igual a $\beta'X'X\beta/(2\sigma^2)$. Denotaremos essa distribuição por $F'[k, n-k, \beta'X'X\beta/(2\sigma^2)]$.

Se o modelo (6) contiver a restrição estimável $l'\beta = m$, a razão

$$F_r = \frac{SRQ_r}{SQR/(n-k)}$$

onde SRQ_r , é a soma de quadrados do resíduo do modelo restrito, terá distribuição F central com 1 e $n-k$ graus de liberdade. Note que o parâmetro de não centralidade se torna igual a zero em razão do corolário do Teorema 5.

As estatísticas F fornecem um meio para se julgar a qualidade do ajuste e se pode acrescentá-las às tabelas de análise de variância.

Fonte	gl	SQ	QM	F	$P \geq F$
Modelo	k	$y'Py$	$y'Py/k$	$\frac{y'Py/k}{y'(I-P)y/(n-k)}$	p
Resíduo	$n-k$	$y'(I-P)y$	$y'(I-P)y/(n-k)$		
Total	n	$y'y$			

10.6 O CASO GERAL

No modelo de Gauss-Markov normal, o vetor $e \sim N(\emptyset, \sigma^2 I)$. Supondo V conhecida, se $\text{var}(e) = \sigma^2 V$, com V positiva definida, as considerações feitas para tal modelo, a princípio, não são válidas. No entanto, é possível se reparametrizar o modelo e utilizar a mesma teoria do caso em que $\text{var}(e) = \sigma^2 I$.

Como V é positiva definida, $V = HH'$ para algum H não singular. Assim, é possível a transformação

$$z = H^{-1}y \Rightarrow y = Hz \Rightarrow Hz = X\beta + e \Rightarrow z = H^{-1}X\beta + H^{-1}e = U\beta + f$$

onde $U = H^{-1}X$, $f = H^{-1}e$, $E(f) = H^{-1}E(e) = \emptyset$ e

$$\text{var}(f) = \sigma^2 H^{-1} V (H^{-1})' = \sigma^2 H^{-1} H H' (H')^{-1} = \sigma^2 I$$

Ou seja, $z = U\beta + f$ é um modelo que satisfaz as condições de Gauss-Markov. Para tal modelo, as equações normais serão

$$U'U\beta = U'z$$

ou, equivalentemente,

$$\begin{aligned} X'(H^{-1})'H^{-1}X\beta &= X'(H^{-1})'H^{-1}y \rightarrow X'(HH')^{-1}X\beta = X'(HH')^{-1}y \Rightarrow \\ &\Rightarrow X'V^{-1}X\beta = X'V^{-1}y \end{aligned}$$

Denotando V^{-1} por W , a última expressão acima será escrita como

$$X'WX\beta = X'Wy$$

e teremos as equações de Aitken. Assim, no ajuste de mínimos quadrados ponderados, a ponderação será realizada por elementos dependentes da matriz de variâncias e covariâncias do erro.

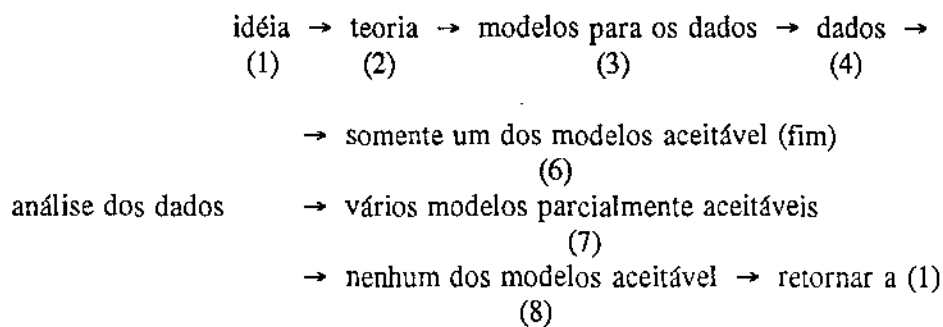
No modelo $y = X\beta + e$, com $e \sim N(\emptyset, \sigma^2 V)$, o melhor estimador linear não viciado de $X\beta$ será dado por $\tilde{y} = X\tilde{\beta}$, onde $\tilde{\beta}$ é solução das equações de Aitken.

Capítulo 11

Testes Estatísticos

11.1 INTRODUÇÃO

Procedimentos para avaliação de dados são úteis na Ciência. Para entendê-los, é importante se ter uma visão de como o conhecimento se acumula e como a validade de hipóteses sobre certos fatos é verificada. Consideremos o diagrama



Assumiremos que alcançamos o passo (7) onde vários modelos, ou um *continuum* de modelos, são parcialmente consonantes com os dados. Isso quer dizer que a probabilidade da ocorrência de nossos dados, sob os modelos parcialmente aceitáveis, é diferente de zero. Note que a idéia inicial pode ser muito vaga, como por

exemplo, que o aumento da temperatura ambiente pode causar aumento na velocidade de multiplicação de certo tipo de bactéria.

O problema, então, é como se ter uma idéia da validade dos diferentes modelos. Segundo Kempthorne e Folks, esse processo remonta, de certa forma, à antiguidade e um dos primeiros exemplos foi dado por Arbuthnot em 1710. Entretanto, seu real desenvolvimento começou com Karl Pearson em 1900 e continuou com Gosset (sob o pseudônimo de Student) em 1908. Ainda segundo Kempthorne e Folks, nos 20 anos seguintes, Fisher desenvolveu muitos testes de significância sem, no entanto, dar uma rigorosa definição do que ele considerava um teste de significância. Subsequentemente, Neyman e Pearson desenvolveram uma teoria geral sobre os testes de hipóteses que pareciam idênticos aos testes de significância. Controvérsias sobre as semelhanças e diferenças entre eles se seguiram. No presente trabalho, não discutiremos os testes de hipóteses: consideraremos, apenas, os testes de significância.

Sob o título 'Testes Estatísticos' englobaremos, também, os intervalos de consonância. Eles surgem da inversão dos testes de significância e foram incorporados, aqui, em virtude dessa ligação com os testes. Eles podem, também, ser usados para testar hipóteses, como veremos.

11.2 TESTES DE SIGNIFICÂNCIA E NÍVEL DE SIGNIFICÂNCIA

Suponhamos que somos levados a considerar um modelo $M(\theta)$ dependendo de um parâmetro θ (que pode ser um vetor) não especificado. Frequentemente, é possível

considerar, ou um valor específico, ou um conjunto de valores para θ que tenha significado científico. Por exemplo, consideremos n animais cujos pesos foram medidos em algum instante. Após submetê-los a certo tratamento (por exemplo, uma dieta durante um período de tempo), tomamos novamente seus pesos e calculamos as diferenças d_1, \dots, d_n entre eles. Sejam esses d_i os dados para estudo. Note que fizemos uma condensação dos dados: inicialmente tínhamos n pares de valores, $\{x_i; y_i\}$, onde x_i era o peso inicial e y_i o peso final após o tratamento. Tomemos, agora, a seguinte posição: o tratamento produziu uma alteração no peso final dos animais. Tal alteração, no entanto, pode ser indistinguível daquela que teria ocorrido sem o tratamento, apenas com flutuações decorrentes do mecanismo probabilístico do experimento. Assim, necessitamos de algum tipo de teste. Fisher, citado por Kempthorne e Folks, fez a seguinte afirmação em 1956: 'A necessidade de testar se algum efeito aparente poderia, razoavelmente, ser devido ao acaso, foi sentida desde os experimentos de Darwin sobre taxa de crescimento'.

Perguntamos, agora, se os dados podem fornecer uma medida da evidência de que tenha ocorrido uma alteração. A resposta é afirmativa e uma forma de produzir a quantificação dessa evidência é construir um teste de significância.

Digamos que sob um modelo especificado, M_0 , obtivéssemos todos os conjuntos de dados possíveis, D_i . Denotemos esses conjuntos por $\{D_i\}$. Conhecendo $\{D_i\}$, podemos conhecer a probabilidade associada a cada D_i . A comparação de nosso conjunto de dados, D_0 , com a população dos D_i poderá fornecer uma idéia da distância, em algum sentido, de D_0 em relação a $\{D_i\}$, se M_0 é realmente verdadeiro. Entretanto, não temos idéia do que constitui um grande ou pequeno desvio e devemos

buscar meios de medir a distância entre D_0 e $\{D_i\}$ especificado pelo modelo.

A idéia básica nos testes de significância é calcular uma medida do desvio dos dados em relação a valores esperados sob o modelo e computar a probabilidade de que uma amostra aleatória seja tão ou mais discrepante que os dados. Por exemplo, suponhamos o modelo de que uma observação y tenha vindo de uma distribuição binomial, $B(1000, 1/2)$. Se obtivemos $y=500$ sucessos em 1000 provas independentes, é óbvio que esta amostra é a menos discrepante de todas as amostras aleatórias possíveis, sob o modelo considerado.

Definição 1. Sejam todos os possíveis conjuntos de dados sob um modelo M_0 , $\{D_i\}$. Um teste de significância consiste em: 1) arranjar todos os conjuntos de dados, D_i , como um conjunto parcialmente ordenado: se D_i não ocorre na ordenação parcial depois de D_j , escrevemos $D_i > D_j$. 2) Seja $P(D_i, M_0)$, a probabilidade associada a D_i para o modelo especificado M_0 . Associar a cada conjunto de dados observado, D_0 , o número

$$NS(D_0, M_0) = \sum_{D > D_0} P(D, M_0)$$

que é chamado *nível de significância* do conjunto de dados D_0 , em relação ao modelo M_0 , para o ordenamento parcial escolhido.

Para ilustrar, suponhamos que existam sete diferentes conjuntos de dados

possíveis sob M_0 e que denotaremos por D_1, \dots, D_7 . Seja o ordenamento parcial $(D_1) \succ (D_2, D_7) \succ (D_3, D_6) \succ (D_4) \succ (D_5)$. Nesse arranjo, D_1 é menos discrepante que todos os outros D_i . D_2 e D_7 são, ambos, mais discrepantes que D_1 , são equivalentes e nenhum deles é mais discrepante que o outro. Com esse ordenamento parcial, escrevemos

$$\begin{aligned}
 &D_1 \succ D_2, D_1 \succ D_3, \dots, D_1 \succ D_7 \\
 &D_2 \succ D_3, D_2 \succ D_4, \dots, D_2 \succ D_7 \\
 &\dots\dots\dots \\
 &D_7 \succ D_2, D_7 \succ D_3, \dots, D_7 \succ D_6
 \end{aligned}$$

Se $D_i \succ D_j$, dizemos que o conjunto de dados D_i é tão ou mais discrepante que D_j em relação ao modelo M_0 .

É razoável construir a distância de um conjunto de dados em relação à população dos conjuntos de dados possíveis sobre uma escala probabilística. A distância está representada pelo nível de significância e, qualquer que seja o problema, obteremos uma distância entre 0 e 1. Um nível de significância igual a 1 indica que a distância é pequena, enquanto que um nível de significância igual a 0 indica que a distância é muito grande. Ou ainda, quanto menor o nível de significância, maior a evidência contra o modelo postulado.

Consideremos, agora, um importante teorema sobre os níveis de significância.

► Teorema 1. $P\{NS(D) \leq \alpha | M_0\} = \alpha, 0 \leq \alpha \leq 1.$

Prova. O nível de significância de um conjunto de dados D é menor ou igual a α se, e somente se, $D > D_1$, onde $NS(D_1) = \alpha$. Mas

$$\begin{aligned}P(D > D_1) &= \sum_{D > D_1} P(D) \\ &= NS(D_1) \\ &= \alpha\end{aligned}$$

Este teorema é a base dos testes de significância para o caso de um modelo completamente especificado, M_0 . Para ilustrar, consideremos o caso em que todos os valores entre 0 e 1 são níveis de significância possíveis de serem obtidos, como no caso contínuo. Então, a distribuição de um nível de significância, sob M_0 , pode ser representada pela distribuição uniforme $U(0,1)$.

A motivação para o uso de um teste de significância é que podemos substituir os dados por um número entre 0 e 1, cuja distribuição de probabilidade é conhecida, se amostrarmos sob M_0 . Nosso interesse em calcular o nível de significância é, portanto, formar uma idéia da validade do modelo sob M_0 .

Como vimos, o ponto inicial para os testes de significância é um conjunto de dados. Com base em conhecimento prévio e nos dados em si, suspeita-se que uma certa classe de modelos pode ser usada para representar os dados. O analista examinará os dados e obterá resultados que indicarão quanto o modelo que ele está inclinado a usar é consonante com os dados.

Um teste de significância consiste em:

1. Especificar uma classe de modelos, referida como *hipótese nula*, H_0 , e outra classe referida como *hipótese alternativa*, H_a . Por exemplo, para um modelo $M(\theta)$, poderíamos tomar como hipótese nula, $H_0: \theta \geq \theta_0$ para um θ_0 especificado. O conjunto de valores de θ , tal que $\theta < \theta_0$, constitui a hipótese alternativa, $H_a: \theta < \theta_0$.
2. Escolher uma função das observações, T . No exemplo da avaliação do peso final dos animais submetidos a uma certa dieta, a função T escolhida foi a diferença entre o peso final e inicial.
3. Avaliar o nível de significância, isto é, $NS = P(T \geq t)$, onde t é o valor observado de T e onde a probabilidade é calculada para a hipótese nula.

Mais uma vez, vale mencionar que se deseja formar uma idéia sobre quanto o modelo especificado é consonante com os dados e não, se o modelo é verdadeiro.

No caso dos modelos lineares, relações entre os parâmetros de um modelo constituem, na realidade, restrições dentro do espaço paramétrico total. Podemos, então, ajustar um modelo restrito, onde as restrições, conforme veremos à frente, deverão ser funções paramétricas estimáveis. A seguir, comparação entre os vetores resultantes desse ajuste e do ajuste do modelo irrestrito pode fornecer meios de se julgar a validade do modelo restrito. Já sabemos que o quadrado da norma de tais vetores, as somas de quadrados, sob o modelo de Gauss-Markov normal, tem distribuição de probabilidade conhecida, são independentes e, dessa forma, o quociente entre elas, também tem distribuição conhecida. Assim, dispomos de meios

probabilísticos para quantificar a ligação entre os modelos.

Pode-se encarar o ajuste do modelo restrito com restrições estimáveis como o ajuste sob determinada hipótese nula, H_0 . Isso é equivalente a testar a hipótese de que a restrição é válida e que pode ser realizado através de um teste de significância.

Consideremos o conjunto de restrições $L\beta = k$ no modelo de Gauss-Markov normal. Necessitamos de uma estatística T para o teste. Digamos que seja escolhido o melhor estimador linear não viciado de $L\beta$, Lb : isso somente será possível se $L\beta$ for estimável. Assim, o conjunto de restrições deve ser estimável.

Os intervalos de consonância que veremos na seção 11.7 podem ser usados para testar hipóteses, pois fornecem um conjunto de valores para o parâmetro, ou para uma função paramétrica estimável, consonante com os dados, a um dado nível de consonância. Se esses valores são compatíveis com aqueles da hipótese sendo testada, atribuímos à hipótese um nível de significância igual a 1 menos o nível de consonância considerado e, assim, os intervalos de consonância podem ser vistos como uma forma de testar hipóteses.

11.3 HIPÓTESES TESTÁVEIS

Consideremos o modelo de Gauss-Markov normal. Um objetivo na análise do modelo é estimar ou testar, onde possível, os elementos ou certas combinações lineares

dos elementos de β .

Para testar $H_0: L\beta = k$, é necessário inicialmente determinar se a hipótese é testável. Hipótese testável é aquela que pode ser expressa em termos de funções estimáveis. Os cálculos efetuados para testar uma hipótese testável poderão, também, ser realizados a partir de funções não estimáveis. Essa é uma situação especial que será analisada na próxima seção. Por ora, restringir-nos-emos às hipóteses construídas de funções estimáveis.

Se $H_0: L\beta = k$ vai ser testada e se $Lb - k$ for parte do teste estatístico, tal estatística necessitará ser invariante com respeito a valores de b , solução das equações normais. E será, somente se $L\beta$ for estimável.

Para realizar o teste, devemos ajustar um modelo levando em conta a hipótese que desejamos testar. Seja ela, $H_0: L\beta = k$. Assim, ajustar o modelo $y = X\beta + e$, sob $H_0: L\beta = k$, é equivalente a ajustar um modelo restrito, onde as equações normais são

$$\begin{cases} X'X\beta + L'\lambda = X'y \\ L\beta = k \end{cases}$$

conforme estudado na seção 5.2.

Note que situações particulares de interesse estão enquadradas nessa hipótese:

1. para $H_0: \beta = \emptyset$, cada um dos elementos de β é igual a zero, $L=I$ e $k = [0 \dots 0]$;
2. para $H_0: \beta = \beta_0$, β_p , $i = 1, \dots, p$, é igual a um valor β_{i0} , $L=I$ e $k' = [\beta_{10}, \dots, \beta_{p0}]$;
3. para $H_0: l'\beta = k$, uma combinação linear particular de β é igual a uma constante e $L=l'$. Por exemplo, se $H_0: \beta_2 - \beta_1 = 2$, $L=l' = [-1 \ 1 \ 0 \ \dots \ 0]$ e $k = 2$;
4. para $\beta_q = \emptyset$, alguns dos p β_p , $q < p$, são iguais a zero. Por exemplo, se

$$H_0: \beta_2 = \beta_4 = 0, \quad \text{c o m} \quad \beta' = [\beta_1 \ \beta_2 \ \beta_3 \ \beta_4], \quad L = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad \text{e}$$

$$k = [0 \ 0 \ 0 \ 0].$$

Após o ajuste do modelo restrito sob a hipótese nula, o quadrado da norma do vetor de resíduos deve ser comparado à norma ao quadrado do vetor de resíduos resultante do ajuste do modelo irrestrito. A primeira norma ao quadrado será $(y - Xb_r)'(y - Xb_r)$, onde $b_r = b - BB'L'(LBB'L')^{-1}(Lb - k)$ é a solução de mínimos quadrados para o modelo restrito.

$$(y - Xb_0)'(y - Xb_0) = [y - X(b - BB'L'(LBB'L')^{-1}(Lb - k))] \times \\ \times [y - X(b - BB'L'(LBB'L')^{-1}(Lb - k))]$$

Desenvolvendo a última expressão, teremos

$$(y - Xb_0)'(y - Xb_0) = (y - Xb)'(y - Xb) + (Lb - k)'(LBB'L')^{-1}(Lb - k) = SQR + Q$$

SQR é a soma de quadrados do resíduo no modelo irrestrito $y = X\beta + e$ e Q , uma soma de quadrados decorrente do ajuste do modelo sob H_0 . Ao avaliar o ajuste sob a hipótese nula, pode-se usar o valor de Q comparado de forma adequada à SQR . Detalhes foram estudados nos Capítulos 9 e 10.

No cálculo de SQR está envolvido Lb que, como já visto, é o melhor estimador linear não viciado de $L\beta$. Assim, é necessário, primeiramente, avaliar a estimabilidade de $L\beta$.

Para que $L\beta$ seja estimável,

$$L = AX \rightarrow LBB'X'X = AXBX = AX = L$$

Portanto, sempre que $L = LBB'X'X$, $L\beta$ será estimável.

11.4 TESTANDO HIPÓTESES TESTÁVEIS

No modelo de Gauss-Markov normal, pelo Teorema 1 e pelo corolário do Teorema 5 do Capítulo 10, a forma quadrática Q é tal que

$$Q/\sigma^2 = \chi^2[p(L), (L\beta - k)'(LBB'L')^{-1}(L\beta - k)/\sigma^2]$$

Como

$$k = LL'(LL')^{-1}k = AXL'(LL')^{-1}k = AXBB'X'XL'(LL')^{-1}k = LBB'X'XL'(LL')^{-1}k$$

e

$$b = BB'X'y$$

pode-se escrever que

$$Q = [y - XL'(LL')^{-1}k]'XBB'L'(LBB'L')^{-1}LBB'X'[y - XL'(LL')^{-1}k]$$

Além disso,

$$SQR = [y - XL'(LL')^{-1}k]'(I - XBB'X')[y - XL'(LL')^{-1}k]$$

pois $X'(I - XBB'X') = \emptyset$. Também por essa razão, o produto das matrizes dessas duas formas quadráticas é nulo e SQR e Q são independentes. Assim,

$$F = \frac{Q/p(L)}{SQR/(n-p(X))} \sim F'[p(L), n-p(X), (L\beta - k)'(LBB'L')^{-1}(L\beta - k)/2\sigma^2]$$

Sob a hipótese nula $H_0: L\beta = k$, o parâmetro de não centralidade torna-se igual a zero e F tem distribuição F central com $p(L)$ e $n - p(X)$ graus de liberdade. Então, a estatística F pode ser usada para um teste de significância de H_0 .

11.5 HIPÓTESES NÃO TESTÁVEIS

Desde que $LBB'L'$ seja não singular, Q poderá ser calculado, mesmo que $L\beta$ seja não testável. Nessa situação, que hipótese estará sendo testada com $H_0: L\beta = k$?

A resposta é

$$H_0: LBB'X'X\beta = k \quad (1)$$

como veremos a seguir. Note que (1) é sempre testável pois

$$LBB'X'X(BB'X'X) = LBB'X'X$$

Seja Q^* a soma de quadrados para o teste de tal hipótese.

$$Q^* = (LBB'X'Xb - k)'(LBB'X'XBB'X'XBB'L')^{-1}(LBB'X'Xb - k)$$

Porém,

$$LBB'X'Xb = LBB'X'XBB'X'y = LBB'X'y = Lb$$

e

$$LBB'X'XBB'X'XBB'L' = LBB'X'XBB'L' = LBB'L'$$

Assim,

$$Q^* = (Lb - k)'(LBB'L')^{-1}(Lb - k) = Q$$

11.6 EXEMPLO 1

Vejamos um exemplo do último caso discutido. Dos Quadros IIIa, IIb e V da seção 3.9, temos

$$X'X = \begin{bmatrix} 28 & 15 & 13 & 6 & 18 & 4 \\ 15 & 15 & 0 & 6 & 5 & 4 \\ 13 & 0 & 13 & 0 & 13 & 0 \\ 6 & 6 & 0 & 6 & 0 & 0 \\ 18 & 5 & 13 & 0 & 18 & 0 \\ 4 & 4 & 0 & 0 & 0 & 4 \end{bmatrix} \quad BB' = \begin{bmatrix} 0.25 & 0.00 & 0.00 & -0.25 & -0.25 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.28 & 0.00 & -0.20 & 0.00 \\ -0.25 & 0.00 & 0.00 & 0.42 & 0.25 & 0.00 \\ -0.25 & 0.00 & -0.20 & 0.25 & 0.45 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \end{bmatrix}$$

$$b' = [0.32 \quad -0.09 \quad 0.00 \quad -0.13 \quad -0.06 \quad 0.00]$$

Para a hipótese não testável H_0 :BAIXA = 0,

$$I' = [0 \quad 0 \quad 0 \quad 0 \quad 1 \quad 0]$$

$$I'BB'I' = 0.45$$

$$Q = (-0.06)(2.22)(-0.06) = 0.01$$

$$SQR = 0.04$$

$$F = \frac{0.01}{0.04/24} = 6.00$$

Para a hipótese testável H_0 :BAIXA - ALTA = 0,

$$l' = [0 \ 0 \ 0 \ 0 \ 1 \ -1]$$

$$l'BB'l = 0.45$$

$$Q = 0.01$$

$$F = 6.00$$

como para a hipótese não testável acima. De fato, note que

$$l'BB'X'X = [0 \ 0 \ 0 \ 0 \ 1 \ -1] = l'$$

e ao tentar testar $H_0: \text{BAIXA} = 0$, estávamos, realmente, testando $H_0: \text{BAIXA} - \text{ALTA} = 0$.

As várias somas de quadrados vistas no Capítulo 9 padronizam algumas hipóteses frequentemente úteis. No entanto, é possível, se assim se desejar, a partir de matrizes L adequadas, testar-se qualquer outra hipótese testável. Suponhamos que para os dados do Quadro I do Capítulo 3, desejamos testar que a média dos valores séricos de valina dos indivíduos com alta ingestão alcoólica seja a metade da média daqueles valores nos indivíduos com ingestão alcoólica baixa e média. A matriz adequada será $L = l' = [0 \ 0 \ 0 \ -1 \ 0.5 \ 0.5]$. O programa SAS® para o ajuste do modelo e teste dessa hipótese será:

```
PROC GLM;
  CLASS PACIENTE INGESTAO;
  MODEL VALINA=PACIENTE INGESTAO;
  ESTIMATE 'ALTA VS 0.5(B+M)' INGESTAO 0 0 0 -1 0.5 0.5;
  RUN;
```


As informações referentes ao ajuste do modelo já foram vistas. Com relação ao teste da hipótese referida acima, o resultado da análise encontra-se no Quadro I.

Quadro I. Teste de hipótese

General Linear Models Procedure				
Dependent Variable: VALINA				
Parameter	Estimate	T for H0: Parameter=0	Pr > T	Std Error of Estimate
ALTA VS 0.5 (B+M)	-0.09449167	-4.34	0.0002	0.02174970

Como o nível de significância (0.0002) é bastante baixo, conclui-se que a hipótese não é consonante com os dados.

11.7 INTERVALOS DE CONSONÂNCIA

A especificação de uma margem de erro é prática habitual na ciência, bem como na indústria ou engenharia. O investigador, após realizar um experimento, raramente estará satisfeito em saber que suas observações simplesmente falam contra uma hipótese. Se os achados mostram, por exemplo, que uma nova droga ou tratamento tem algum efeito significativo além daquele do placebo ou padrão prévio, o experimentador preferiria saber não somente uma estimativa desse efeito, mas também alguma variação de valores razoáveis do efeito.

Portanto, um intervalo ou um conjunto de valores para o parâmetro (ou para uma função paramétrica estimável) que torne o conjunto de dados sob o modelo em questão não muito improvável, é desejável.

Frequentemente, deseja-se formar uma opinião sobre o grau de concordância dos parâmetros de uma distribuição com os dados. Ou seja, deseja-se avaliar se os dados são consonantes com determinado modelo e em que grau; ou ainda, avaliar se um grupo de modelos está consonante com os dados. Por exemplo, consideremos o modelo de que no lançamento de uma moeda, a probabilidade de 'cara' é igual à probabilidade de 'coroa'. Entretanto, ao examinar a moeda, descobrimos que ela tem duas 'caras'. O modelo obviamente não é consonante com nossa observação. Porém, suponhamos que sob o mesmo modelo, após realizarmos 1000 lançamentos, observamos 408 'caras'. Suponhamos, também, que a análise da sequência dos 1000 resultados nos leve a admitir o modelo de independência das provas. Os dados, portanto, são consonantes com o modelo de provas independentes com p próximo de 0.41. O mesmo se daria para as hipóteses $p=0.40$ ou $p=0.39$. Assim, com base nos dados, temos meios de formar uma opinião sobre quais valores do parâmetro são plausíveis.

Consideremos outro exemplo. Seja o modelo de Gauss-Markov normal $y_i = \beta_0 + x_i\beta_1 + e_i$. Digamos que o seguinte conjunto de dados foi obtido

x_i	y_i
1	10
2	12
3	14
4	15
5	14
6	16

O melhor estimador linear não viciado para β_1 será $b_1 = 1.06$. O desvio padrão da estimativa será $S_{b_1} = 0.24$. Note que $(b_1 - \beta_1)/S_{b_1}$ tem distribuição t de Student com 5 graus de liberdade.

Fixemos, agora, b_1 calculado da amostra. Podemos variar β_1 até encontrarmos um valor β_{10} , tal que $P(b < \beta_{10}) = \alpha/2$, para α escolhido entre 0 e 1. Igualmente, podemos obter um valor $\beta_{10'}$, tal que $P(b > \beta_{10'}) = \alpha/2$.

O intervalo $[\beta_{10}; \beta_{10'}]$ é um conjunto de valores para β_1 , consonantes com a amostra observada, ao nível $1-\alpha$ de consonância. Para o exemplo,

$$\beta_{10} = 1.06 - (2.06)(0.24) = 0.57$$

$$\beta_{10'} = 1.06 + (2.06)(0.24) = 1.55$$

onde o valor $t = 2.06$ vem da distribuição t de Student com 5 graus de liberdade.

Assim, os valores de β_1 pertencentes a $[0.57; 1.55]$ são consonantes com os dados ao nível de consonância 0.90.

De um modo mais geral, podemos dizer que um conjunto de valores para um parâmetro, cujos níveis de significância excedem um certo valor α , constituem um intervalo de consonância, ao nível de consonância $1-\alpha$, para o parâmetro. Tais valores são aqueles compatíveis com os dados observados.

As considerações feitas até aqui podem ser generalizadas em dois sentidos: no primeiro, falaremos em um vetor de parâmetros, situação na qual teremos regiões de consonância. No segundo, funções paramétricas estimáveis. Evidentemente, poderemos ter $g(\beta) = \beta$ e/ou $\beta = \beta$.

Deve-se notar que o processo que acabamos de ver é o inverso de um teste de significância.

Tendo obtido um intervalo de consonância com um dado nível de consonância, poderíamos questionar qual a probabilidade de que o intervalo obtido contenha o verdadeiro parâmetro. Para resolver essa questão, deveríamos supor que repetidamente retirássemos amostras da mesma distribuição e determinássemos um intervalo de consonância para cada amostra. Tal processo geraria uma população de intervalos. O intervalo correspondente a uma amostra aleatória é, então, um intervalo aleatório com, pelo menos um de seus extremos, uma variável aleatória.

No exemplo acima, o intervalo de consonância para β_1 com nível de

consonância de 0.90, era $[b_1 - tS_{b_1} \leq \beta_1 \leq b_1 + tS_{b_1}]$. Desejamos determinar

$$P\{b_1 - tS_{b_1} \leq \beta_1 \leq b_1 + tS_{b_1}\} = P\left\{\left|\frac{b_1 - \beta_1}{S_{b_1}}\right| \leq t\right\} = P\{|T| \leq 2.06\}$$

Assim, num repetido processo de amostragem, a probabilidade de que obtivéssemos um intervalo de consonância que contivesse β_1 , ao nível de consonância de 0.90, é 0.90.

A expressão $T = (b_1 - \beta_1)/S_{b_1}$ representa uma função dos dados e do parâmetro de interesse, β_1 . T tem distribuição de probabilidade conhecida e que não depende de parâmetros desconhecidos. Além disso, a expressão pode ser invertida ou 'pivotada'. Uma expressão desse tipo permite a determinação de um intervalo de consonância e é conhecida como *quantidade pivotal*.

Resumindo, suponhamos que uma amostra aleatória de n observações tenha sido obtida de alguma população com função densidade de probabilidade $f(y, \beta)$. O intervalo de consonância para o nível de consonância $1 - \alpha$ de uma função paramétrica estimável $g(\beta)$ é o conjunto de valores

$$t_1 \leq g(\beta) \leq t_2$$

com pontos extremos, t_1 e t_2 , computados através da distribuição de probabilidade de

uma função conjunta de $g(\beta)$ e das observações. A função $h(y, g(\beta))$, é uma quantidade pivotal. A desigualdade $q_1 \leq h(y, g(\beta)) \leq q_2$ pode ser invertida ou 'pivotada', fornecendo $t_1(y) \leq g(\beta) \leq t_2(y)$. Desse modo,

$$P\{t_1 \leq g(\beta) \leq t_2\} = 1 - \alpha \quad (2)$$

Na verdade, existirão infinitos intervalos que satisfazem (2). Os pontos extremos do intervalo, em geral, são escolhidos de modo que seu comprimento seja o mínimo possível.

11.8 INTERVALO DE CONSONÂNCIA PARA $l'\beta$

Se $l'\beta$ é estimável, então, no modelo de Gauss-Markov normal, seu melhor estimador linear não viciado, $(l'\beta)_b$, será

$$(l'\beta)_b = l'\hat{\beta} = l'b = l'BB'X'y$$

Pelo Teorema 5 do Capítulo 10,

$$l'b \sim N(l'\beta, l'BB'l\sigma^2)$$

e pelo Teorema 7 do mesmo capítulo,

$$\frac{S^2}{\sigma^2}(n-k) \sim \chi_{n-k}^2$$

onde $S^2 = \frac{1}{n-k}(y - Xb)'(y - Xb)$ e $k = p(X)$.

Podemos, então, tomar

$$Z = \frac{l'b - l'\beta}{\sigma\sqrt{h}} \sim N(0,1)$$

onde $h = l'BB'l$. Como Z e S^2 são independentes,

$$T = \frac{Z}{\sqrt{\frac{S^2(n-k)}{\sigma^2(n-k)}}} = \frac{l'b - l'\beta}{S\sqrt{h}} \sim t_{n-k}$$

T é função dos dados e de $l'\beta$. Como, além disso, sua distribuição não depende de parâmetros desconhecidos, é uma quantidade pivotal. Então, podemos escrever

$$P\left\{-t_{n-k,1-\frac{\alpha}{2}} \leq \frac{l'b - l'\beta}{S\sqrt{h}} \leq t_{n-k,1-\frac{\alpha}{2}}\right\} = 1 - \alpha \quad (3)$$

onde $t_{n-k,1-\frac{\alpha}{2}}$ é o percentil $1 - \frac{\alpha}{2}$ da distribuição t de Student com $n-k$ graus de liberdade. Rearranjando a expressão (3), temos

$$P\left\{l'b - t_{n-k,1-\frac{\alpha}{2}}S\sqrt{h} \leq l'\beta \leq l'b + t_{n-k,1-\frac{\alpha}{2}}S\sqrt{h}\right\} = 1 - \alpha$$

e, assim,

$$\left[l'b - t_{n-k,1-\frac{\alpha}{2}}S\sqrt{h}; l'b + t_{n-k,1-\frac{\alpha}{2}}S\sqrt{h}\right]$$

é um intervalo de consonância para $l'\beta$ ao nível de consonância $1-\alpha$.

11.9 INTERVALO DE CONSONÂNCIA PARA σ^2

Pelo Teorema 7 do Capítulo 10, sabemos que $Q = S^2(n-k)/\sigma^2$, no modelo de Gauss-Markov normal, tem distribuição qui quadrado central. Q é uma quantidade

pivotal e, portanto, pode ser usada para calcular um intervalo de consonância para a variância do erro. Assim,

$$P\left\{q_1 \leq \frac{S^2}{\sigma^2}(n-k) \leq q_2\right\} = 1 - \alpha$$

onde q_1 e q_2 são pontos da distribuição qui quadrado central que deixam regiões sob os extremos da curva da distribuição com áreas cuja soma é igual a α .

A expressão acima pode ser invertida, de modo que

$$P\left\{\frac{S^2(n-k)}{q_2} \leq \sigma^2 \leq \frac{S^2(n-k)}{q_1}\right\} = 1 - \alpha$$

e o intervalo

$$\left[\frac{S^2(n-k)}{q_2}, \frac{S^2(n-k)}{q_1} \right]$$

constitue um conjunto de valores para σ^2 , consonante com os dados a um nível de consonância igual a $1-\alpha$.

11.10 REGIÕES DE CONSONÂNCIA

Consideremos um conjunto de m funções estimáveis linearmente independentes, $L'\beta$, com $p(L) = m$. Seu melhor estimador linear não viciado, no modelo de Gauss-Markov normal será $(L'\beta)_b = L'b$, onde b é qualquer solução das equações normais. Para que $L'\beta$ seja estimável, $L' \in \mathfrak{R}(X)$ ou, de modo equivalente, $L' \in \mathfrak{R}(X'X)$. Por isso, podemos escrever, para um conjunto de funções estimáveis linearmente independentes, que $L' = Q'X'X$ para algum Q e

$$(L'\beta)_b = L'b = L'BB'X'y = Q'X'XBB'X'y = Q'X'B'X'y = Q'X'y$$

Como $E(L'b) = L'\beta$ e

$$\text{var}(L'b) = L'BB'X'XBB'L\sigma^2 = Q'X'XBB'X'XBB'X'XQ\sigma^2 = Q'X'XQ\sigma^2 = L'Q\sigma^2$$

então, $L'b \sim N(L'\beta, L'Q\sigma^2)$.

Seja, agora, a matriz $G = L'Q$ e o vetor a , tal que

$$XQa = 0 \rightarrow X'XQa = 0 \rightarrow La = 0$$

Como $p(L) = m$, então, $a = 0$. Portanto,

$$XQa = \emptyset \rightarrow a = \emptyset$$

e $p(XQ) = m$. Assim,

$$G = L'Q = Q'X'XQ = (XQ)'XQ$$

é não singular, de modo que a forma quadrática

$$J = (L'b - L'\beta)'G^{-1}(L'b - L'\beta) \sim \sigma^2\chi_m^2$$

Pode-se verificar tal fato, fazendo $G^{-1} = A'A$ e a transformação $V = A(L'b - L'\beta)$.

Como J é composta de funções estimáveis, ela é independente de $\frac{n-k}{\sigma^2}S^2 \sim \chi_{n-k}^2$, onde

$k = p(X)$. Assim,

$$\frac{(L'b - L'\beta)'G^{-1}(L'b - L'\beta)}{mS^2} \sim F_{m, n-k; 1-\alpha}$$

e o conjunto de consonância é dado por

$$(L'b - L'\beta)'G^{-1}(L'b - L'\beta) \leq mS^2F_{m, n-k; 1-\alpha} \quad (4)$$

onde $F_{m, n-k; 1-\alpha}$ é o percentil $1-\alpha$ da distribuição F com m e $n-k$ graus de liberdade.

(4) representa um elipsóide em \mathbf{R}^n com centro em $L'b$.

11.11 COMPARAÇÕES MÚLTIPLAS

Em uma análise de variância, um teste F indicando a presença de diferença estatisticamente significativa para as médias de um fator, frequentemente não termina a análise. Em geral, estamos interessados em informações adicionais, como por exemplo, quais das médias são diferentes.

Os testes de *comparações múltiplas de médias* são úteis para responder tais questões. Numerosos procedimentos estão disponíveis para realizá-los. Em geral, é calculado um intervalo de consonância para a diferença observada entre quaisquer duas médias, ao nível de consonância desejado. Se o intervalo não contém o valor zero, as duas médias são declaradas estatisticamente diferentes. Como os comprimentos dos intervalos de consonância variam conforme o procedimento usado, os resultados obtidos para um mesmo conjunto de dados podem variar de acordo com o método que estiver sendo empregado.

A interpretação das comparações múltiplas deve ser feita com cuidado. A falha em detectar uma diferença estatisticamente significativa entre duas ou mais médias não deve levar à conclusão de que as médias populacionais são iguais. Na realidade, isso significa que a diferença, se existe, não deve ser grande o suficiente para ser detectada com o tamanho de amostra analisada. Além disso, a não significância não é transitiva:

dadas três médias, a maior e a menor podem ser estatisticamente diferentes entre si, sem que elas sejam estatisticamente diferentes daquela de valor intermediário. Esse é um fato de ocorrência frequente em comparações múltiplas.

Em comparações múltiplas, intervalos de consonância podem ser mais úteis que testes de significância. Intervalos de consonância mostram o grau de incerteza em cada comparação de uma forma facilmente interpretável. Eles tornam mais fácil o julgamento da significância prática de uma diferença.

Uma forma simples de realizar comparações múltiplas é construir um intervalo de consonância para cada par de médias, usando a distribuição t de Student:

$$|\mu_i - \mu_j| \leq (\bar{y}_i - \bar{y}_j) \pm t_{\frac{\alpha}{2}, \nu} S \sqrt{\frac{1}{n_i} + \frac{1}{n_j}}$$

onde $t_{\frac{\alpha}{2}, \nu}$ é o percentil $1 - \frac{\alpha}{2}$ da distribuição t de Student com ν graus de liberdade usados para estimar a variância; S é a raiz quadrada da estimativa da variância; \bar{y}_i e \bar{y}_j , as estimativas das médias de interesse; n_i e n_j , os números de observações correspondentes.

No entanto, existe um problema com repetidos intervalos de consonância desse tipo. Suponhamos que existam cinco médias e, para cada diferença entre quaisquer duas delas, seja construído um intervalo de consonância ao nível de consonância 0.95. Haverá dez comparações, cada qual com probabilidade 0.05 de erroneamente não

conter a média. Com dez comparações, a probabilidade de pelo menos um erro certamente é maior que 0.05, sendo difícil o cálculo da probabilidade exata. Grosseiramente, assumindo independência, a probabilidade de pelo menos um erro será

$$1 - (1 - 0.05)^{10} = 0.40$$

Possivelmente, não se estará satisfeito com um intervalo de consonância para uma função paramétrica com probabilidade 0.40 de ser consonante com os dados.

Muitos métodos tem sido propostos na literatura. Como ilustração veremos dois deles: o método de Tukey e o de Scheffé.

11.11.1 Método de Tukey

Sejam y_{ij} , $i = 1, \dots, p$, $j = 1, \dots, n$, p amostras independentes de n variáveis aleatórias independentes, normalmente distribuídas, com média μ_i e variância comum σ^2 . Esse é um delineamento experimental balanceado e com um critério de classificação. Desejamos testar

$$H_0: \mu_1 = \dots = \mu_p$$

ou, equivalentemente,

$$H_0: \mu_i - \mu_{i'} = 0 \quad i \neq i'$$

O problema pode ser resolvido construindo um intervalo de consonância para $\mu_i - \mu_j$ e que será dado por

$$(\bar{y}_i - \bar{y}_j) \pm q_{p,v,\alpha} \frac{S}{\sqrt{n}} \quad (5)$$

onde $q_{p,v,\alpha}$ é o nível de significância α da *variação estudentizada* para p médias e v graus de liberdade, usados na estimativa S^2 da variância σ^2 . A largura do intervalo de consonância, para qualquer dado nível de consonância, é maior quando q é usado no lugar de t . Os valores críticos de q encontram-se tabelados (por exemplo, Pearson e Hartley (1966), Tabela 29).

Para entender o surgimento da expressão (5), note que a desigualdade

$$\frac{|(y_i - y_j) - (\mu_i - \mu_j)|}{S} \leq c$$

será verdadeira se, e somente se,

$$\frac{\max_{ij} \{|(y_i - \mu_i) - (y_j - \mu_j)|\}}{S} \leq c$$

Como $R_p = \max_{ij} \{|(y_i - \mu_i) - (y_j - \mu_j)|\}$, no MGMN, são p variáveis aleatórias,

independentes e com distribuição $N \sim (0, \sigma^2)$, o resultado (5) segue imediatamente.

A principal aplicação do método de Tukey é a comparação emparelhada de p médias, para determinar se qualquer das médias populacionais, μ_1, \dots, μ_p , diferem. No entanto, aplicação a outros tipos de delineamentos é também possível. Consideremos, por exemplo, que existam dois critérios de classificação com r níveis do primeiro fator, c níveis do segundo e n observações para cada combinação de níveis do primeiro e segundo fatores. Para a comparação das médias dos níveis do primeiro fator, $A_i - A_{i'}$, o intervalo de consonância será dado por

$$\bar{y}_{i..} - \bar{y}_{i'..} \pm q_{p,v,\alpha} \frac{S}{\sqrt{cn}}$$

Uma desvantagem do método é o fato de que todas as médias amostrais devem ser baseadas no mesmo número de observações, de modo que elas tenham a mesma variância σ^2/n . Na prática, se o desvio da condição de balanceamento é pequeno, é razoável usar o método, colocando a média dos n_i no lugar de n . Para situações onde o desbalanceamento é mais evidente, não há outra alternativa senão usar outro método como, por exemplo, o método de Scheffé.

Uma generalização da comparação emparelhada de médias é o estudo dos contrastes. Definimos um *contraste* de parâmetros μ_i como qualquer função linear

$$\sum_{i=1}^p l_i \mu_i$$

cujos coeficientes l_i tem a propriedade

$$\sum_{i=1}^p l_i = 0$$

Assim, $\mu_1 - \mu_2$ e $2\mu_1 - (\mu_2 + \mu_3)$ são contrastes, mas $\mu_1 - (\mu_2 + \mu_3)$ não é.

Seja l um contraste e \mathcal{L} o espaço linear dos contrastes. Um intervalo de consonância para funções lineares $\sum_{i=1}^p l_i \mu_i$ será dado por

$$\sum_{i=1}^p l_i \bar{y}_i \pm q_{p, \nu, \alpha} \frac{S}{\sqrt{n}} \sum_{i=1}^p \frac{|l_i|}{2} \quad (6)$$

para qualquer $l \in \mathcal{L}$. Note que para a comparação emparelhada de médias,

$$\sum_{i=1}^p |l_i|/2 = 1.$$

A prova de (6) segue do Lema 1, para uma escolha adequada da constante c

referida ali.

► **Lema 1.** Se $|y_i - y_i'| \leq c$, para quaisquer $i, i' = 1, \dots, p$, então

$$\left| \sum_{i=1}^p l_i y_i \right| \leq c \sum_{i=1}^p \frac{|l_i|}{2} \quad (7)$$

para todo $l \in \mathcal{Q}$.

Prova. Se $l_i \equiv 0$, $i = 1, \dots, p$, (7) é trivialmente verdadeira. Suponhamos que

$l_i \neq 0$, $i = 1, \dots, p$. Sejam $P = \{i: l_i > 0\}$ e $N = \{i: l_i < 0\}$. Representando $\sum_{i=1}^p |l_i|/2$

por g ,

$$g = \frac{1}{2} \sum_{i \in P} l_i + \frac{1}{2} \sum_{i \in N} (-l_i)$$

Como

$$0 = \sum_{i=1}^p l_i = \sum_{i \in P} l_i + \sum_{i \in N} l_i$$

segue que

$$\sum_{i \in P} l_i = \sum_{i \in N} (-l_i) = g$$

Multiplicando e dividindo $\sum_{i=1}^P l_i y_i$ por g , temos

$$\begin{aligned} \sum_{i=1}^P l_i y_i &= \frac{\sum_{i' \in N} (-l_{i'}) \sum_{i \in P} l_i y_i + \sum_{i \in P} l_i \sum_{i' \in N} l_{i'} y_{i'}}{g} \\ &= \frac{\sum_{i \in P} \sum_{i' \in N} l_i (-l_{i'}) y_i + \sum_{i \in P} \sum_{i' \in N} l_{i'} y_{i'}}{g} \\ &= \frac{\sum_{i \in P} \sum_{i' \in N} l_i (-l_{i'}) (y_i - y_{i'})}{g} \end{aligned}$$

Mas, para $i \in P$ e $i' \in N$,

$$|l_i (-l_{i'}) (y_i - y_{i'})| = l_i (-l_{i'}) |y_i - y_{i'}| \leq l_i (-l_{i'}) c$$

e, assim,

$$\left| \sum_{i=1}^P l_i y_i \right| \leq \frac{\sum_{i \in P} \sum_{i' \in N} l_i (-l_{i'}) c}{g} = \frac{g^2 c}{g} = cg = c \sum_{i=1}^P \frac{|l_i|}{2}$$

11.11.2 Método de Scheffé

Seja o vetor $n \times 1$, $y \sim N(\mu, \sigma^2 I)$, onde $\mu = X\beta$. Para o caso particular de análise de variância, Scheffé mostrou que os intervalos de consonância simultâneos, com nível de consonância $1-\alpha$ para todos os contrastes de μ_i , tem a forma

$$l' \bar{y} \pm S \sqrt{(p-1) F_{p-1, n-p; \alpha} l' B B' l} \quad (8)$$

onde p é o número de tratamentos; S é a raiz quadrada da estimativa da variância; n o total de observações e $F_{p-1, n-p; \alpha}$, o percentil $1-\alpha$ da distribuição F , com $p-1$ e $n-p$ graus de liberdade.

Para verificar a validade da expressão (8), necessitaremos do lema e do teorema abaixo.

► **Lema 2.** Se $c > 0$, $\left| \sum_{i=1}^k a_i y_i \right| < c \left(\sum_{i=1}^k a_i^2 \right)^{\frac{1}{2}}$, para todo a_i , $i=1, \dots, k$ se, e

somente se, $\sum_{i=1}^k y_i^2 \leq c^2$.

Prova. Pela desigualdade de Cauchy-Schwartz (Mood, Graybill e Boes, 1974, pag. 162),

$$\left| \sum_{i=1}^k a_i y_i \right| \leq \left(\sum_{i=1}^k y_i^2 \right)^{\frac{1}{2}} \left(\sum_{i=1}^k a_i^2 \right)^{\frac{1}{2}}$$

Considerando $\sum_{i=1}^k y_i^2 \leq c^2$, segue que

$$\left| \sum_{i=1}^k a_i y_i \right| \leq c \left(\sum_{i=1}^k a_i^2 \right)^{\frac{1}{2}}$$

Por outro lado, escolhendo $a_i = y_i$, $i = 1, \dots, k$, a desigualdade

$$\left| \sum_{i=1}^k a_i y_i \right| \leq c \left(\sum_{i=1}^k a_i^2 \right)^{\frac{1}{2}}$$

torna-se

$$\left| \sum_{i=1}^k y_i^2 \right| \leq c \left(\sum_{i=1}^k y_i^2 \right)^{\frac{1}{2}}$$

► **Teorema 2.** Se o vetor aleatório $y \sim N(X\beta, \sigma^2 I)$, onde X é $n \times p$, \mathcal{L} é um subespaço de dimensão d do espaço Euclidiano p -dimensional, então

$$P\left\{|l'(\hat{\beta} - \beta)| \leq S\sqrt{dF_{d,n-p,\alpha} l'BB'l}\right\} = 1 - \alpha$$

para qualquer $l \in \mathcal{L}$.

Prova. Seja a matriz L , $d \times p$, de posto completo, cujas linhas constituem uma base para \mathcal{L} . Seja o conjunto de funções $G = L\beta$ e $\hat{G} = L\hat{\beta}$. Para $l \in \mathcal{L}$, a combinação linear $l'\beta$ (ou $l'\hat{\beta}$) é igual a $\lambda'G$ (ou $\lambda'\hat{G}$), para algum λ . Assim,

$$l'\beta = \lambda'G = \lambda'L\beta$$

Como a igualdade vale para qualquer β , então, $l' = \lambda'L$. Agora, no modelo de Gauss-Markov normal, $\hat{\beta} \sim N(\beta, \sigma^2 BB')$ e, então, $L\hat{\beta} \sim N(L\beta, \sigma^2 LBB'L')$ ou $\hat{G} \sim N(G, \sigma^2 LBB'L')$. Também, $l'\beta = \lambda'\hat{G} \sim N(l'\beta, \sigma^2 l'BB'l)$ ou $\lambda'\hat{G} \sim N(\lambda'G, \sigma^2 \lambda'LBB'L'/\lambda)$. Como $LBB'L'$ é positiva definida, existe uma matriz T , tal que $TLBB'L'T' = I$.

Sejam agora, $\delta = TG$ e $\hat{\delta} = T\hat{G}$. Assim, $\hat{\delta} \sim N(TG, \sigma^2 TLBB'L'T')$ ou $\hat{\delta} \sim N(\delta, \sigma^2 I)$ e os $\hat{\delta}_i$, $i=1, \dots, d$, são variáveis aleatórias normais, independentes, com média δ_i e variância comum σ^2 .

A razão $\sum_{i=1}^d (\delta_i - \delta_i)^2 / dS^2$ tem distribuição F , pois S^2 é independente de $\hat{\delta}$ (já que é independente de $\hat{\beta}$).

Pelo lema acima,

$$\sum_{i=1}^d (\delta_i - \delta_i)^2 \leq dF_{d,n-p,\alpha} S^2$$

se, e somente se,

$$\left| \sum_{i=1}^d a_i (\delta_i - \delta_i) \right| \leq S \sqrt{dF_{d,n-p,\alpha} \sum_{i=1}^d a_i^2}$$

para todo $\mathbf{a} = [a_1 \dots a_d]$ e, então,

$$P\left\{ |\mathbf{a}'(\hat{\delta} - \delta)| \leq S \sqrt{dF_{d,n-p,\alpha} \mathbf{a}'\mathbf{a}} \right\} = 1 - \alpha \tag{9}$$

Mas,

$$\begin{aligned} \mathbf{a}'\delta &= \mathbf{a}'TG = \mathbf{a}'TL\beta \\ \mathbf{a}'\hat{\delta} &= \mathbf{a}'T\hat{G} = \mathbf{a}'TL\hat{\beta} \\ \mathbf{a}'\mathbf{a} &= \mathbf{a}'I\mathbf{a} = \mathbf{a}'TLBB'L'T'\mathbf{a} \end{aligned}$$

Vimos que

$$l' \in \mathfrak{R}(L) \rightarrow l'\beta \in \mathfrak{R}(L\beta) \rightarrow l'\beta = \lambda'L\beta$$

para algum λ' . Mas $a'TL\beta \in \mathfrak{R}(L\beta)$ e, então, existe uma correspondência entre a e λ através de $\lambda' = a'T$. Portanto, a expressão (9) pode ser escrita como

$$P\left\{|\lambda'L(\hat{\beta} - \beta)| \leq S\sqrt{dF_{d,n-p,\alpha}(\lambda'LB B'L'\lambda)}\right\} = 1 - \alpha$$

e, como $l' = \lambda'L$, a última expressão se torna

$$P\left\{|l'(\hat{\beta} - \beta)| \leq S\sqrt{dF_{d,n-p,\alpha}l'BB'l'}\right\} = 1 - \alpha \quad \blacktriangleleft$$

Considerando $l'\hat{\beta} = l'\bar{y}$ e $d = p - 1$, o resultado (8) é imediato.

O método de Scheffé é uma técnica muito versátil e pode ser aplicada a quaisquer funções lineares estimáveis. No entanto, sua versatilidade constitui, ao mesmo tempo, sua maior deficiência. Por exemplo, para a comparação de pares de médias, o método de Tukey fornece intervalos mais estreitos que o método de Scheffé. Portanto, é mais sensível que este último nessas situações.

11.12 EXEMPLO 2

Como ilustração do uso dos intervalos de consonância, vejamos um exemplo. Para o conjunto de dados do Quadro I do Capítulo 3 foi ajustado um modelo. No entanto, por aquela análise, não se pode inferir quais níveis do fator, digamos nível de ingestão alcoólica, são estatisticamente significativos. Para isso, apliquemos os métodos de Tukey e Scheffé. O seguinte programa¹ SAS[®] pode ser usado para essa análise:

```
PROC GLM;
  CLASS PACIENTE INGESTAO;
  MODEL VALINA = PACIENTE INGESTAO;
  MEANS INGESTAO / TUKEY SCHEFFE;
RUN;
```

O resultado da análise (excluída a tabela de análise de variância já mostrada no Capítulo 3) encontra-se no Quadro I. Na parte superior do quadro, para o método de Tukey, nota-se o valor do nível de significância $\alpha=0.05$ e do nível de consonância, referido como 'Confidence'. Além disso, observa-se, também, o valor (3.532) da quantidade pivotal (variação estudentizada) que deixa sob os extremos da curva da distribuição desta, regiões cujas áreas somam α . As mesmas informações para o método de Scheffé encontram-se no centro do quadro. As comparações significantes, para ambos os métodos, estão indicados por '****'. Note que as larguras dos intervalos, para cada uma das comparações de médias, são maiores no método de Scheffé que no de Tukey. No entanto, apesar desse fato, as diferenças estatisticamente significativas encontradas foram as mesmas em ambos os métodos.

¹O comando MEANS calcula a média das observações para cada nível do fator indicado (INGESTAO). As opções TUKEY e SCHEFFE fornecem os intervalos de consonância para as várias combinações de níveis do fator.

Quadro I. Comparação múltipla de médias

General Linear Models			
Tukey's Studentized Range (HSD) Test for variable: VALINA			
NOTE: This test controls the type I experimentwise error rate.			
Alpha=0.05 Confidence=0.95 df=24 MSE=0.001695			
Critical Value of Studentized Range=3.532			
Comparisons significant at the 0.05 level are indicated by '***'.			
INGESTAO Comparison	Simultaneous Lower Confidence Limit	Difference Between Means	Simultaneous Upper Confidence Limit
BAIXA - ALTA	-0.0454	0.0030	0.0515
BAIXA - MEDIA	0.0750	0.1318	0.1887 ***
ALTA - BAIXA	-0.0515	-0.0030	0.0454
ALTA - MEDIA	0.0625	0.1288	0.1952 ***
MEDIA - BAIXA	-0.1887	-0.1318	-0.0750 ***
MEDIA - ALTA	-0.1952	-0.1288	-0.0625 ***
Scheffe's test for variable: VALINA			
NOTE: This test controls the type I experimentwise error rate but generally has a higher type II error rate than Tukey's for all pairwise comparisons.			
Alpha=0.05 Confidence=0.95 df=24 MSE=0.001695			
Critical Value of P=3.40283			
Comparisons significant at the 0.05 level are indicated by '***'.			
INGESTAO Comparison	Simultaneous Lower Confidence Limit	Difference Between Means	Simultaneous Upper Confidence Limit
BAIXA - ALTA	-0.0476	0.0030	0.0536
BAIXA - MEDIA	0.0725	0.1318	0.1912 ***
ALTA - BAIXA	-0.0536	-0.0030	0.0476
ALTA - MEDIA	0.0595	0.1288	0.1981 ***
MEDIA - BAIXA	-0.1912	-0.1318	-0.0725 ***
MEDIA - ALTA	-0.1981	-0.1288	-0.0595 ***

Capítulo 12

Análise de Covariância

12.1 INTRODUÇÃO

Consideremos a situação onde observamos um conjunto de variáveis aritméticas, Y , resultantes da aplicação de algum plano experimental sobre um sistema. Se, juntamente com Y , forem observadas outras variáveis, Z , ou seja, variáveis que variam junto com Y (ou covariáveis), e que não foram controladas pelo plano experimental, estamos diante do que é referido como *análise de covariância*. Pretende-se estudar o comportamento de Y com base nas variáveis de planejamento, bem como nos valores das covariáveis. Conceitualmente, não existe diferença entre a análise de covariância e os modelos lineares vistos até agora. Eles diferem apenas na origem dos elementos da matriz do modelo. Dentro desse contexto, os procedimentos para resolver as equações normais, testar hipóteses e calcular reduções nas somas de quadrados seguem o mesmo padrão já estudado. Nenhum conceito adicional está envolvido.

Podemos considerar a análise de covariância como o ajuste do modelo $y = X^* \beta^*$ onde $X^* = [X \ Z]$ e $\beta^* = [\beta \ \gamma]$. X é uma matriz de variáveis indicadoras

como anteriormente; Z , uma matriz de covariáveis e β e γ os parâmetros correspondentes. Além do significado estrutural do modelo de análise de covariância, a quebra da matriz do modelo (e a correspondente partição do vetor de parâmetros) em duas partes, foi conveniente na era anterior aos computadores. Ela representava uma estratégia útil para a resolução das equações normais. Hoje em dia, do ponto de vista computacional, a presença de variáveis não classificatórias em uma ou mais colunas na matriz do modelo, tornou-se irrelevante.

Como ilustração, consideremos o modelo particionado (8) da seção 3.5 e façamos $X_1 = X$, $X_2 = Z$, $\beta_1 = \beta$ e $\beta_2 = \gamma$. As equações normais reduzidas para Z , eliminando X serão

$$Z'(I - P_X)Z\hat{\gamma} = Z'(I - P_X)y$$

Retirando o valor de $Z'Z\hat{\gamma}$ dessa última equação, substituindo em (9b) da seção 3.5 e simplificando temos que

$$Xb = P_X(y - Z\hat{\gamma})$$

Pode-se verificar que a soma de quadrados associada ao modelo (8) da seção 3.5 é

$$SQM = b'X'y + \hat{\gamma}'Z'y$$

Mas isso é igual a

$$[P_X(y - Z\hat{\gamma})]'y + \hat{\gamma}'Z'y = y'P_Xy - \hat{\gamma}'Z'P_Xy + \hat{\gamma}'Z'y = y'P_Xy + \hat{\gamma}'Z'(I - P_X)y$$

A soma de quadrados do resíduo será

$$SQR = y'(I - P_X)y - \hat{\gamma}'Z'(I - P_X)y$$

Portanto, a soma de quadrados do resíduo para o modelo com covariáveis será igual à soma de quadrados do resíduo ignorando $Z\gamma$, menos uma certa quantidade que depende das covariáveis. Assim, a introdução da covariável no modelo, reduziu a soma de quadrados associada ao resíduo.

12.2 OBJETIVOS

O objetivo básico da análise de covariância é fazer inferência sobre médias de grupos ou de tratamentos de uma variável resposta Y , medida em cada unidade de estudo. A outra variável, a covariável Z , fornecerá também informação sobre Y , predizendo, em algum grau, sua resposta final. Isso é feito de duas formas:

1. a variação em Y que está associada com Z é removida da variância do erro, resultando em estimativas mais precisas e testes mais poderosos e
2. as médias da variável Y para cada grupo são ajustadas para as médias das covariáveis, dentro de cada grupo. Assim, as médias de Y tornam-se mais comparáveis.

Como exemplo, Snedecor e Cochran (1980, pag 365) citam a aplicação da análise de covariância sugerida por Fisher em experimentos sobre produção de chá. Uma importante fonte de erro, nesses casos, é a coincidência de tratamentos com plantas cuja produção é maior que outras. No entanto, os resultados apresentam certa estabilidade de ano para ano. O uso da covariável Z , onde esta é a produção anterior à aplicação dos tratamentos, serve como indicador da habilidade associada ao padrão genotípico de produção de cada planta. Ajustando o resultado final após os tratamentos para essa produção esperada, obteremos um erro experimental menor e, portanto, comparação mais precisa entre os tratamentos. Este uso da análise de covariância é possivelmente o mais comum.

12.3 O MODELO

O modelo mais simples de análise de covariância, composto por um fator e por uma covariável, será

$$y_{ij} = \mu + A_i + \gamma Z_{ij} \quad (1)$$

onde o último termo à direita corresponde a introdução da covariável no modelo linear já estudado anteriormente.

No caso do exemplo da seção anterior, y_{ij} é a produção da planta j após o tratamento A_i , Z_{ij} a produção da mesma planta, anteriormente ao tratamento A_i , e γ

um parâmetro desconhecido. A média observada para o tratamento A_i será $\bar{y}_i = \mu + A_i + \gamma \bar{Z}_i$. Assim, $A_i = \bar{y}_i - \mu - \gamma \bar{Z}_i$. O último termo à direita da última equação é o fator que ajusta a média da produção de chá, \bar{y}_i , após os tratamentos. Suponha que no ano anterior, a planta que está recebendo o tratamento I forneceu 10 Kg a mais de chá. Suponha, também, que o coeficiente de regressão de y sobre z , onde z é o vetor de valores observados da covariável Z , foi 0.2, significando que cada quilo de aumento em Z correspondeu a 0.2 Kg de aumento na produção final de chá. A média observada \bar{y}_i , deve ser diminuída de $(0.2)(10) = 2$ Kg, para tornar o tratamento I mais comparável aos demais tratamentos.

O modelo pode ser ampliado em dois sentidos: o uso de mais que um critério de classificação, interações e efeitos hierárquicos, por um lado e, por outro, a ocorrência de mais que uma covariável.

O modelo (1) representa, também, um feixe de linhas paralelas com coeficientes angulares comuns iguais a γ e interceptos iguais a $\mu + A_i$.

De uma forma mais geral, o modelo de análise de covariância pode ser escrito como

$$y = X\beta + Z\gamma \quad (2)$$

Fazendo $X^* = [X \ Z]$ e $\beta^{*'} = [\beta' \ \gamma']$ teremos o modelo $y = X^*\beta^*$. As equações normais serão

$$X'^*X^*\beta^* = X'^*y$$

Sob o aspecto computacional, vale lembrar que o ajuste do modelo de análise de covariância com as covariáveis entrando em primeiro lugar no ajuste, já foi visto quando tratamos das Somas de Quadrados tipo I. Portanto, será possível analisar através de testes de significância realizados com a estatística F , o efeito das variáveis classificatórias corrigidas pelas covariáveis.

12.4 APLICAÇÕES DA ANÁLISE DE COVARIÂNCIA

O teste de heterogeneidade de coeficientes angulares é uma extensão natural da análise de covariância. A análise de covariância testa a diferença dos interceptos de um feixe de retas, assumindo coeficientes angulares constantes. O teste para heterogeneidade de coeficientes angulares verifica a validade desse fato: testa se os coeficientes angulares são constantes ao longo dos grupos. Por exemplo, consideremos o modelo com uma covariável e um critério de classificação

$$y_{ij} = \beta_{0i} + \beta_{1i}Z_{ij} + e_{ij} \quad (3)$$

onde i denota os tratamentos. Note que o modelo (3) é equivalente ao modelo (1). A hipótese de interesse será

$$H_0: \beta_{1i} = \beta_{1i'} \quad (4)$$

para qualquer $i \neq i'$.

A eventual diferença entre os grupos está refletida na presença ou não de interação entre os grupos e as covariáveis. Denotando a interação por $Z*A$, as SQ I para o modelo $y_{ij} = \mu + A_i + Z_{ij} + Z*A_{ij} + e_{ij}$ podem ser usadas para o teste da hipótese (4).

Outra aplicação é o cálculo das *médias ajustadas* já referida na seção 12.2. Essa aplicação é, segundo Searle, Speed e Milliken (1980), uma extensão das *médias de mínimos quadrados*¹. Consideremos o modelo de Gauss-Markov normal com dois critérios de classificação, $y_{ijr} = \mu + A_i + B_j + A*B_{ij}$. Podemos, também, escrever

$$E(y_{ijr}) = \mu_{ij} = \mu + A_i + B_j + A*B_{ij}$$

Médias de mínimos quadrados podem ser calculadas para cada nível de cada fator (incluindo os níveis da interação). Define-se a média de mínimos quadrados, por exemplo para o nível i do fator A , $MMQ(A_i)$, como

$$MMQ(A_i) = \mu_i = \sum_{j=1}^b \mu_{ij} / b \quad (5)$$

onde b é o número de níveis do fator B . Equivalentemente, define-se $MMQ(B_j)$ e

¹Os autores sugerem o uso da denominação *médias marginais populacionais*. Manteremos aqui o termo média de mínimos quadrados que é o utilizado pelo sistema SAS*

$MMQ(A*B_{ij})$, respectivamente, como

$$MMQ(B_j) = \mu_j = \sum_{i=1}^a \mu_{ij} / a \quad (6)$$

$$MMQ(A*B_{ij}) = \mu_{ij} \quad (7)$$

onde a é o número de níveis do fator A . Note que estas expressões não correspondem aos valores esperados das médias observadas. Por exemplo, para A_i ,

$$E(\bar{y}_{i..}) = \sum_{j=1}^b n_{ij} \mu_{ij} / \sum_{j=1}^b n_{ij}$$

onde n_{ij} é o número de observações na combinação do nível i de A com o nível j de B .

O melhor estimador linear não viciado das médias de mínimos quadrados será a média dos melhores estimadores não viciados dos μ_{ij} e, assim, uma determinada média de mínimos quadrados existirá se, e somente se, todos os μ_{ij} implicados forem estimáveis. Se existirem $n_{ij} = 0$, Searle, Speed e Milliken sugerem o emprego de uma *média de mínimos quadrados modificada*.

Por (5), (6) e (7) vê-se que essas médias marginais são independentes dos n_{ij} .

Se o plano experimental é balanceado, essa é uma consequência óbvia. Havendo desbalanceamento, as médias de mínimos quadrados são equivalentes às médias das classes e subclasses em um plano experimental balanceado, com as covariáveis em seu valor médio².

Consideremos o nível A_1 e um vetor de covariáveis z no modelo $E(y) = X\beta + Z\gamma$. A $MMQ(A_1)$ correspondendo a $z = z_0$ é definida, supondo que existam dois critérios de classificação, por

$$MMQ(A_1|z=z_0) = \mu_1 + z_0'\gamma = MMQ(A_1) + z_0'\gamma$$

Para um plano completamente aleatorizado (um critério de classificação) com uma covariável,

$$MMQ(A_1|z=z_0) = MMQ(A_1) + \gamma z_0 = \mu + A_1 + \gamma z_0$$

Seu melhor estimador linear não viciado será

$$EMMQ(A_1|z=z_0) = \bar{y}_1 + \hat{\gamma}z_0$$

Usando $-\bar{z}_1$ no lugar de z_0 ,

$$EMMQ(A_1|z=-\bar{z}_1) = \bar{y}_1 - \hat{\gamma}\bar{z}_1$$

²O manual do PROC GLM do SAS[®] fornece um algoritmo para esse cálculo.

que corresponde à estimativa da média de A_1 corrigida pela covariável.

12.5 EXEMPLO

Como ilustração do emprego da análise de covariância, vejamos um exemplo (Freund, Littell e Spector, 1986). Os dados do Quadro I referem-se a um experimento sobre o crescimento de ostras. Seu objetivo é determinar se o crescimento é afetado pela exposição à água aquecida e pela posição (na superfície ou no fundo) dentro da água. Quatro sacos com dez ostras cada um foram aleatoriamente colocados em cada uma de quatro posições dentro de um canal. Um quinto saco foi colocado fora do canal. As classificações foram:

1. entrada-fundo
2. entrada-superfície
3. saída-fundo
4. saída-superfície
5. fora do canal

As posições na entrada do canal agem como controle já que a temperatura na saída do canal é maior. A posição fora do canal serve como controle geral para o caso da influência de algum outro fator que não a temperatura. Os pesos iniciais de cada ostra foram tomados. Eles foram novamente medidos um mês após. O seguinte programa SAS[®] foi usado para a análise que se encontra nos Quadros IIa, IIb e III.

```

PROC GLM;
  CLASS TRT;
  MODEL FINAL=INICIAL TRT;
  LSMEANS TRT;
  RUN;

```

O comando LSMEANS foi usado para o cálculo da média de cada tratamento corrigida pela covariável (o tamanho inicial). Elas são apresentadas no Quadro III. As médias finais por tratamento sem essa correção foram, respectivamente: 34.475; 31.650; 30.850; 32.225 e 25.025. É importante se verificar que a SQ I para INICIAL (Quadro IIb) foi 342.36. Essa soma de quadrados foi retirada da soma de quadrados do resíduo (Quadro I), resultando o valor de 4.22. Caso a covariável não tivesse sido usada, essa soma de quadrados seria 346.58 e o efeito dos tratamentos não teria sido detectado.

Quadro I. Dados para análise de covariância

OBS	TRT	REP	INICIAL	FINAL
1	1	1	27.2	32.6
2	1	2	32.0	36.6
3	1	3	33.0	27.7
4	1	4	26.8	31.0
5	2	1	28.6	33.8
6	2	2	26.8	31.7
7	2	3	26.5	30.7
8	2	4	26.8	30.4
9	3	1	28.6	35.2
10	3	2	22.4	29.1
11	3	3	23.2	28.9
12	3	4	24.4	30.2
13	4	1	29.3	35.0
14	4	2	21.8	27.0
15	4	3	30.3	36.4
16	4	4	24.3	30.5
17	5	1	20.4	24.6
18	5	2	19.6	23.4
19	5	3	25.1	30.3
20	5	4	18.1	21.8

Cap 12

Quadro IIa. Resultado da análise de covariância

General Linear Models Procedure					
Dependent Variable: FINAL					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	354.4471767	70.8894353	235.05	0.0001
Error	14	4.2223233	0.3015945		
C. Total	19	358.6695000			
		R-Square	Root MSE	FINAL Mean	
		0.988228	0.549176	30.8450000	

Quadro IIb. Resultado da análise de covariância

General Linear Models Procedure					
Dependent Variable: FINAL					
Source	DF	Type I SS	Mean Square	F Value	Pr > F
INICIAL	1	342.3578175	342.3578175	1135.16	0.0001
TRT	4	2.0893593	3.0223398	10.02	0.0005
Source	DF	Type III SS	Mean Square	F Value	Pr > F
INICIAL	1	156.0401767	156.0401767	517.38	0.0001
TRT	4	12.0893593	3.0223398	10.02	0.0005

Quadro III. Médias dos tratamentos corrigidas pela covariável

General Linear Models Least Squares Means	
TRT	FINAL LSMEAN
1	30.1531125
2	30.1173006
3	32.0523296
4	31.5046854
5	30.3975719

Capítulo 13

Exame dos Resíduos

13.1 INTRODUÇÃO

Os resíduos de um ajuste são definidos como as n diferenças $r_i = y_i - \hat{y}_i$, $i = 1, \dots, n$, onde y_i é uma observação e \hat{y}_i o correspondente valor obtido através do ajuste do modelo linear. O conjunto dos n resíduos constitui o vetor $r = y - \hat{y} = y - Xb$. Os r_i são as quantidades que restam após as contribuições associadas ao modelo assumido terem sido removidas. Eles são as diferenças entre o que é observado e o que é predito pelo ajuste, ou ainda, é o que o ajuste do modelo foi incapaz de explicar. Se eles contêm tendências sistemáticas não explicadas, devemos suspeitar do modelo. Assim, o estudo dos resíduos antes da análise estatística propriamente dita, deve sempre ser realizado.

Uma característica importante dos resíduos e que será útil na sua análise é sua sua variância. No modelo de Gauss-Markov normal,

$$\text{var}(r) = \text{var}((I - P_X)y) = (I - P_X)(I - P_X)' \sigma^2 = (I - P_X) \sigma^2$$

onde $(I - P_X)$ é o projetor no espaço de erro de X e, portanto, é simétrico e idempotente. Assim, nota-se que os resíduos não são independentes. Todavia, tal dependência, em geral, não causa problema na análise.

O exame dos resíduos pode fornecer informações sobre o ajuste do modelo de Gauss-Markov normal e é possível se verificar os seguintes fatos:

1. se alguma variável importante não está presente ou foi considerada em forma inadequada (por exemplo, linear quando deveria ser quadrática);
2. se os y_i foram corretamente observados em razão da presença eventual de valores aberrantes;
3. se as observações são homocedásticas e não correlacionadas; ou ainda,
4. se o erro tem distribuição normal.

Além disso, se o intercepto estiver no modelo,

$$\sum_{i=1}^n r_i = \mathbf{1}'\mathbf{r} = \mathbf{1}'(I - P_X)\mathbf{y} = (\mathbf{1}' - \mathbf{1}')\mathbf{y} = 0$$

Como consequência, $\bar{r} = 0$. Esse fato pode ser usado como uma forma de se verificar se os cálculos realizados para o ajuste do modelo estão corretos. Caso a soma dos resíduos não seja igual a zero (e o vetor de 1's estiver presente na matriz do modelo), deve-se suspeitar de problemas numéricos no ajuste.

13.2 MÉTODOS GRÁFICOS PARA O EXAME DOS RESÍDUOS

Os métodos gráficos para o exame dos resíduos que serão vistos a seguir são:

1. Diagrama de pontos.
2. Sequência temporal, se esta é conhecida.
3. Resíduos plotados contra os valores preditos.
4. Resíduos plotados contra possíveis variáveis de interesse.
5. Outros.

Tomemos um exemplo para o estudo dos métodos acima. Seja o modelo $y_{ij} = \mu + A_i + Z_{ij} + e_{ij}$, $i = 1, 2$, $j = 1, \dots, n_i$. Z é uma covariável. Os dados, os valores ajustados e respectivos resíduos encontram-se na Tabela I.

13.2.1. Diagrama de pontos

O diagrama de pontos dos dados do exemplo é mostrado na Figura 1. Se o modelo estiver correto, os resíduos devem ser compatíveis com dez observações de uma distribuição normal de média zero. Quando o número de observações é pequeno, considerável flutuação ocorrerá de modo que a falta de normalidade pode ser apenas aparente. Quando desvios grosseiros ocorrem, explicações devem ser buscadas. No caso do exemplo, não existe motivo para duvidar da presença de normalidade.

Tabela I. Dados, valores ajustados e respectivos resíduos após ajuste

y	μ	A_1	A_2	z	\hat{y}	r
53	1	1	0	5	50	3
49	1	1	0	5	50	-1
48	1	1	0	3	49	-1
48	1	1	0	8	50	-2
50	1	1	0	4	49	1
59	1	0	1	8	62	-3
61	1	0	1	2	61	0
62	1	0	1	8	62	0
63	1	0	1	3	61	0
65	1	0	1	10	62	3

Um procedimento alternativo é o gráfico probabilístico onde, pela característica da escala dos eixos, espera-se encontrar uma linha reta caso a normalidade esteja presente.

Quando o número dos resíduos é grande, em lugar do diagrama de pontos pode-se empregar um histograma. Além disso, o cálculo dos momentos amostrais pode ser útil no sentido de dar uma idéia da distribuição probabilística dos valores dos resíduos.

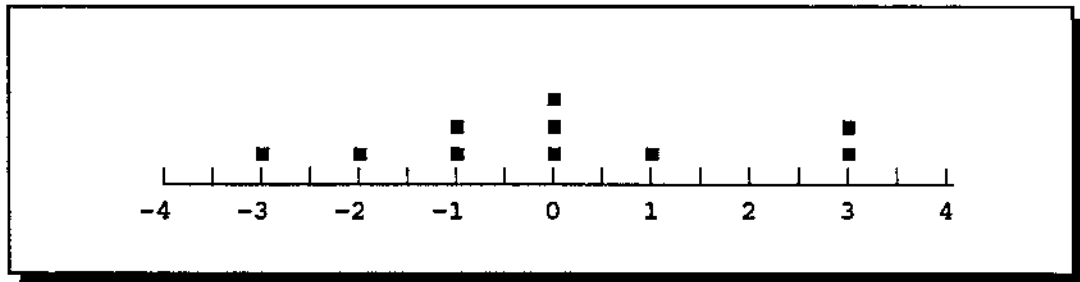


Figura 1. Diagrama de pontos dos resíduos

Observações aberrantes

A patologia mais comum detectada com o diagrama de pontos é a presença de um ou mais resíduos muito maiores em valor absoluto que os demais. O ponto aberrante indica uma observação que não é típica em relação às restantes. Detectado o ponto aberrante, ele deve ser submetido a cuidadoso exame para se buscar a razão de seu aparecimento.

Uma atitude frente a um ponto aberrante é seu descarte do conjunto de dados e reanálise dos mesmos sem essa observação. Rejeição automática não é um procedimento prudente, pois, muitas vezes, aquela observação está fornecendo informação não presente nos outros pontos: o ponto aberrante pode ter surgido de uma combinação não usual de circunstâncias que pode ser de interesse. Como regra geral, os pontos aberrantes devem ser descartados somente se são provenientes de erro no registro dos dados.

13.2.2. Sequência temporal

Consideremos que os resíduos do exemplo tenham ocorrido na seguinte ordem e em intervalos de tempo igualmente espaçados: -1, 3, 0, 1, -1, -2, 0, -3, 3, 0. O gráfico é o mostrado na Figura 2. Ele dá a impressão de uma banda horizontal como na Figura 3, indicando que com o passar do tempo não existe alteração do sistema. Por exemplo: no experimento pode estar envolvida uma substância química cuja atividade diminui aos poucos, ou ainda, um técnico que vai se tornando mais treinado. Tanto uma situação, quanto a outra, podem provocar distorções nos resultados, passíveis de serem detectadas com os resíduos plotados em sequência temporal. Os dados do exemplo não mostram evidência desse tipo de efeito.

Ao contrário da banda horizontal da Figura 3, outros padrões podem ser encontrados, como os das Figuras 4a, 4b e 4c.

Na Figura 4a, a variância não é constante, mas aumenta com o tempo. Assim, o método dos mínimos quadrados ponderados deve ser usado. Na Figura 4b, um termo linear no tempo deve ser incluído no modelo. Finalmente, na Figura 4c, a curvatura indica que um termo quadrático ou de ordem superior no tempo necessita ser adicionado ao modelo. Combinações e variações desses efeitos podem obviamente ocorrer.

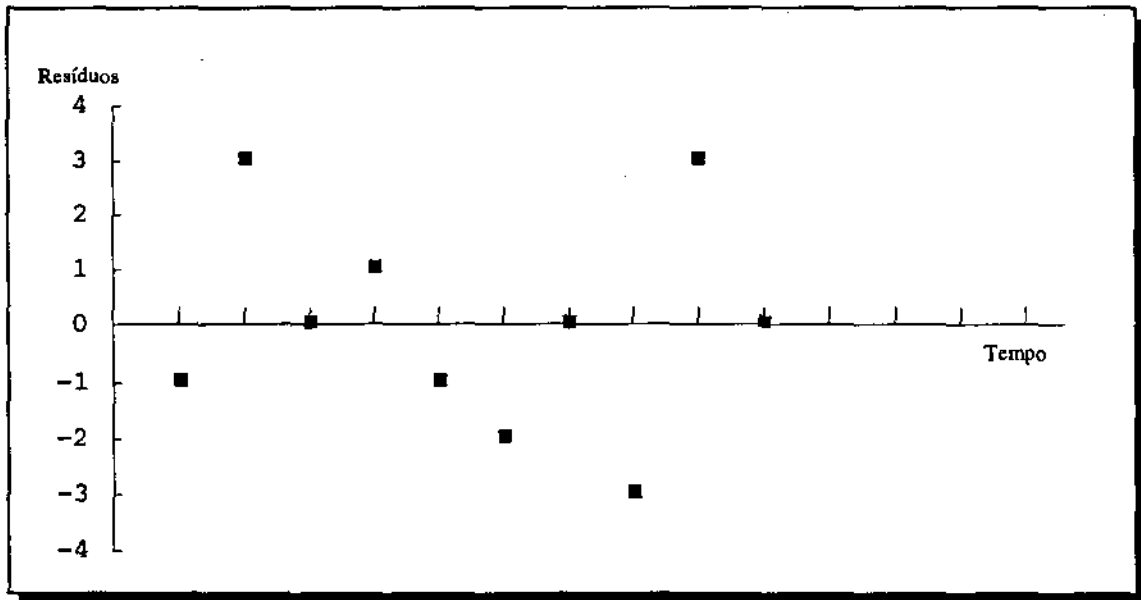


Figura 2. Resíduos plotados em ordem temporal

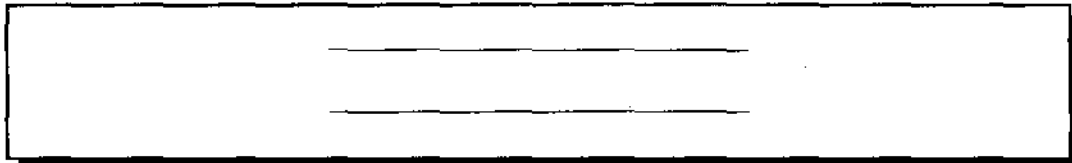


Figura 3. Aspecto satisfatório do gráfico dos resíduos

13.2.3. Resíduos \times valores preditos

O gráfico resíduos \times valores preditos para os dados da Tabela I é mostrado na Figura 5. A banda horizontal, como na Figura 3, não mostra evidência de anormalidade, de modo que o ajuste parece razoável.

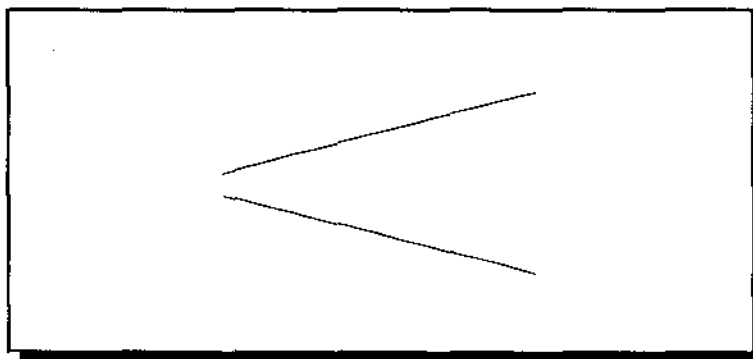


Figura 4a. Variância crescente dos dados

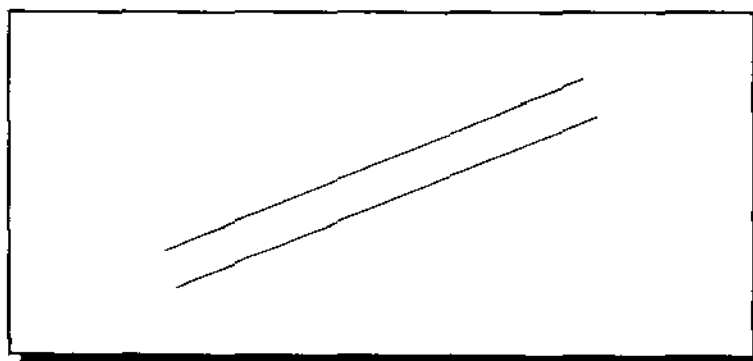


Figura 4b. Modelo inadequado

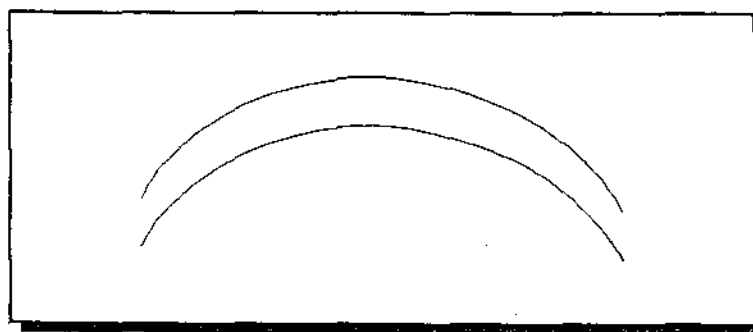


Figura 4c. Modelo inadequado

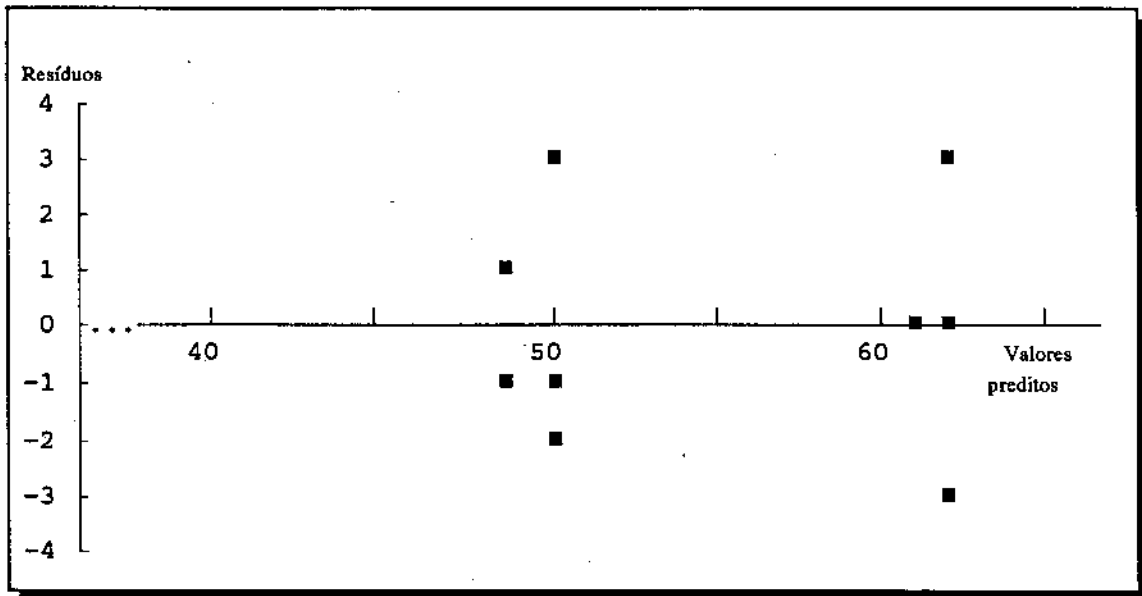


Figura 5. Resíduos plotados contra os valores preditos pelo ajuste

Problemas estariam indicados por gráficos como os da Figura 4a, 4b e 4c. Na Figura 4a, a variância não é constante como assumido e é preciso, ou usar o método dos mínimos quadrados ponderados, ou realizar uma transformação em y antes do ajuste. A Figura 4b indica erro na análise: resíduos negativos associados a valores baixos de \hat{y}_i e resíduos positivos, a valores altos de \hat{y}_i . Já a Figura 4c mostra que o modelo é inadequado: faltam, por exemplo, termos quadráticos ou produtos cruzados. O padrão também pode indicar a necessidade de uma transformação em y antes do ajuste.

Pode-se questionar o uso dos \hat{y}_i no lugar dos y_i . Isso é feito pois os resíduos e os valores observados, em geral, são correlacionados ao passo que os resíduos e os valores preditos não o são.

$$\text{cov}(\hat{y}, r) = \text{cov}(P_X y, (I - P_X)y) = \sigma^2 P_X (I - P_X) = \emptyset$$

13.2.4. Resíduos \times possíveis variáveis de interesse

Os resíduos devem ser plotados contra qualquer variável de eventual relevância. Suponhamos que os resíduos do exemplo fossem plotados contra a temperatura ambiental no momento da obtenção de cada y_i . Uma correlação entre essas variáveis, além de auxiliar o entendimento do processo, pode sugerir a inclusão da temperatura no modelo.

O gráfico pode ter qualquer uma das formas incluídas na Figura 4 e a interpretação é equivalente à interpretação do gráfico resíduos \times valores preditos.

13.2.5. Outros

O conhecimento do problema em estudo pode sugerir outros tipos de gráficos. Por exemplo, suponhamos que os onze resíduos do exemplo da Tabela I tenham sido obtidos de três máquinas diferentes:

A: 0, -1, -1

B: -2, 0, -3

C: 3, 1, 0, 3

Os diagramas de pontos para cada máquina separadamente são mostrados na Figura 6. Nota-se que o padrão da distribuição dos resíduos para a máquina C é diferente das duas outras. Isso sugere a introdução de outra variável classificatória no modelo correspondendo à máquina na qual y_i foi medido.

Draper e Smith (1981, Cap 3) discutem em detalhe a análise de resíduos.

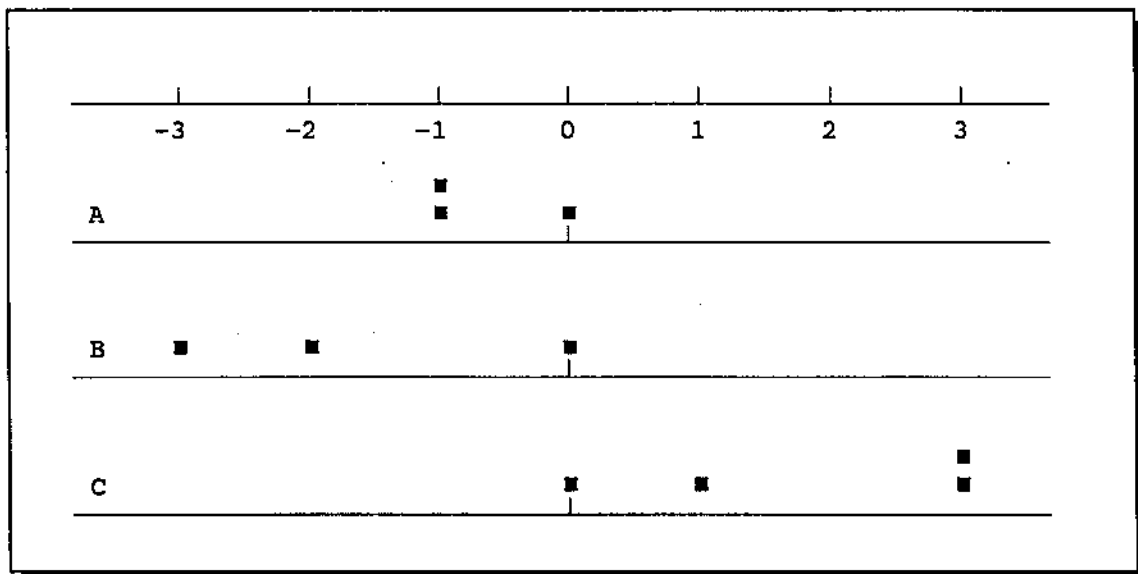


Figura 6. Resíduos indicando efeito não incorporado no modelo ajustado

13.3 EXEMPLO

Veremos agora um exemplo em que o estudo dos resíduos indica a direção em que deve caminhar a análise. Considere os dados do Quadro I. Eles foram gerados de

modo que as 'observações' para cada tratamento possuem variâncias desiguais.

Os dados foram analisados pelo programa SAS [®]:

```
PROC GLM;
  CLASS TRT;
  MODELO Y=TRT;
  OUTPUT OUT=SAIDA P=PY R=RY;

PROC PLOT;
  PLOT RY*PY / VREF=0;
RUN;
```

A observação do gráfico dos resíduos - $RY \times PY$ (Quadro II) - antes da leitura da tabela de análise de variância, mostra a patologia presente no conjunto de dados. Considerando o inverso do quadrado dos valores preditos ($1/(PY)^2$) como peso, novo ajuste foi feito:

```
PROC GLM;
  CLASS TRT;
  MODEL Y=TRT;
  WEIGHT PESO;
RUN;
```

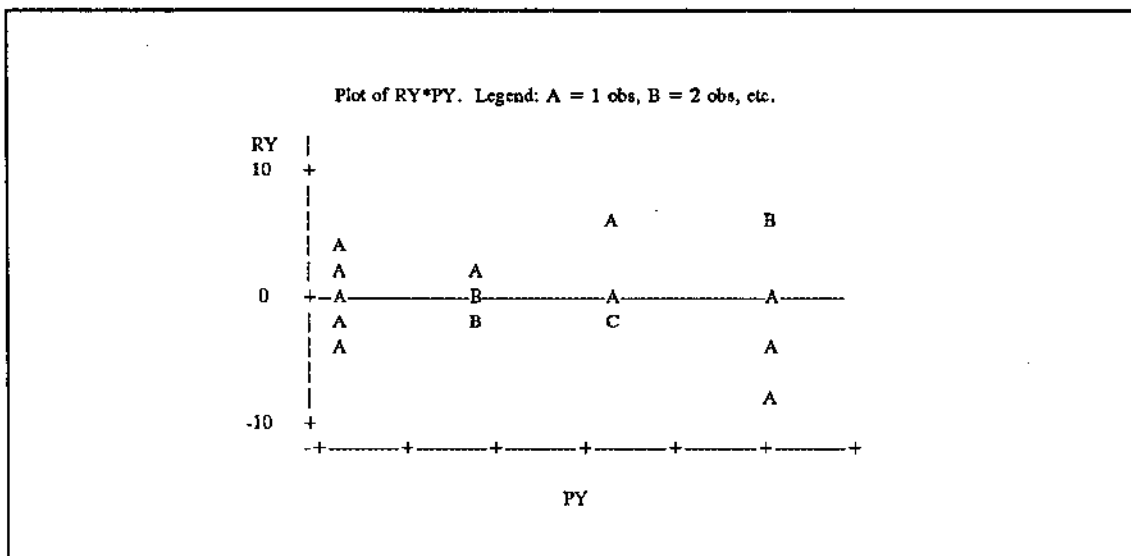
O gráfico dos resíduos (RES) contra os valores preditos (PRED) para o novo ajuste (Quadro III), mostra sensível melhora. Obviamente, é possível tomar-se estes valores preditos para ponderar novo ajuste.

A tabela de análise de variância referente a esse ajuste encontra-se no Quadro IV. Deve-se levar em conta, entretanto, que estas somas de quadrados não são comparáveis às somas de quadrados do ajuste sem a ponderação.

Quadro I. Dados para ajuste de mínimos quadrados ponderados

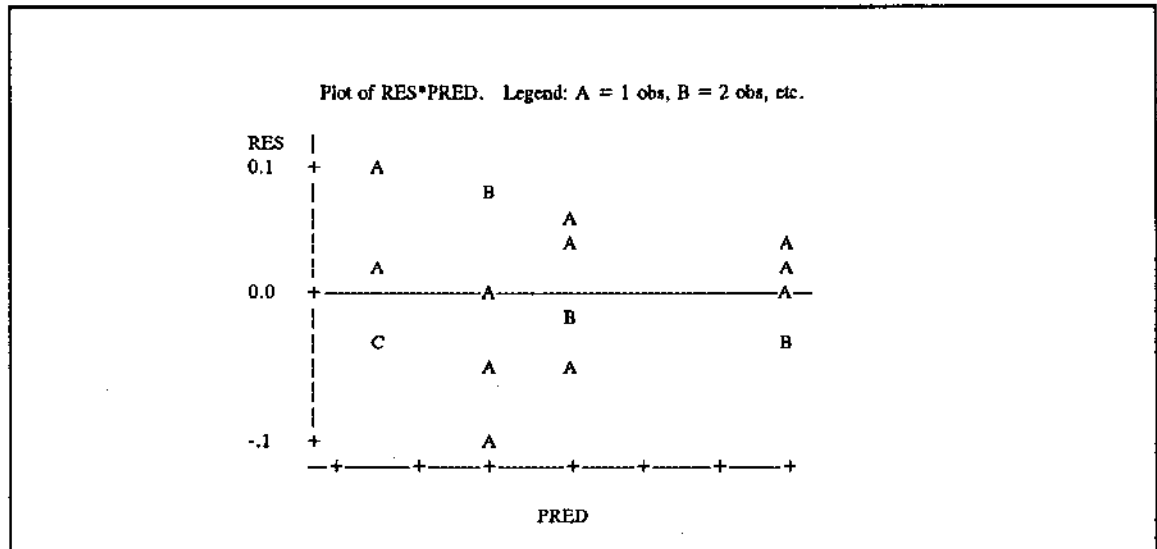
OBS	TRT	REP	Y
1	A	1	63.6096
2	A	2	59.8402
3	A	3	60.7932
4	A	4	57.8334
5	A	5	64.4766
6	B	1	67.5031
7	B	2	72.0546
8	B	3	69.6536
9	B	4	67.6233
10	B	5	70.1276
11	C	1	75.5732
12	C	2	78.4992
13	C	3	84.1100
14	C	4	75.1751
15	C	5	75.5343
16	D	1	83.6360
17	D	2	92.7257
18	D	3	87.5959
19	D	4	79.2012
20	D	5	93.4616

Quadro II. Gráfico resíduos \times valores preditos para o ajuste do Quadro II



Cap 13

Quadro III. Gráfico resíduos × valores preditos para o ajuste do modelo sem ponderação



Quadro IV. Tabela de análise de variância

General Linear Models Procedure					
Dependent Variable: Y					
Weight: PESO					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	0.33947141	0.11315714	45.71	0.0001
Error	16	0.03960675	0.00247542		
C. Total	19	0.37907816			
	R-Square		Root MSE		Y Mean
	0.8955183		0.0497533		73.9514

REFERÊNCIAS BIBLIOGRÁFICAS

1. Bolgiano, D. C. (1981): Making Sense of Types I, II, III and IV Sums of Squares in SAS[®] GLM, *Proceedings of the Sixth Annual SAS[®] Users Group International Conference*, Orlando, FL.
2. Box, G. E. P., Hunter, W. G., Hunter, J. S. (1978): *Statistics for Experimenters*, New York, NY: John Wiley & Sons, Inc.
3. Carvalho, J. F. (1978): Notas de aula, não publicadas.
4. Draper, N., Smith, H. (1981): *Applied Regression Analysis*, New York, NY: John Wiley & Sons, Inc.
5. Freund, R. J. (1980): The Case of the Missing Cell, *The American Statistician*, **34,2**, 94-98.
6. Freund, R. J., Littell, R. C., Spector, P. C. (1986): *SAS[®] System for Linear Models*, Cary, NC: SAS Institute, Inc.
7. Goodnight, J. H. (1978): *Tests of Hypothesis in Fixed Effects Linear Models*, SAS[®] Technical Report R-101, Cary, NC: SAS Institute Inc.
8. Goodnight J. H. (1979): A Tutorial on the SWEEP Operator, *The American Statistician*, **33,3**, 149-158.
9. Graybill, F. A. (1961): *An Introduction to Linear Statistical Models*, New York, NC: McGraw-Hill Book Company, Inc.
10. Graybill, F. A. (1983): *Matrices with Applications in Statistics*, Belmont, CA: Wadsworth International Group.
11. Grizzle, J. E., Starmer, C. F., Koch, G. G. (1969): Analysis of Categorical Data by Linear Models, *Biometrics*, **25**, 489-504.
12. Halmos, P. R. (1974): *Finite-Dimensional Vector Spaces*, New York, NY: Springer-Verlag, Inc.
13. Hudson, G. F. S., Searle, S. R. (1982): Hypothesis Testing with Type IV Sums of Squares of the Computer Routine SAS[®] GLM, *Proceedings of the Seventh Annual SAS[®] Users Group International Conference*, San Francisco, CA.
14. Kempthorne, O. (1980): Notas de aula, não publicadas.
15. Kempthorne, O., Folks, L. (1971): *Probability, Statistics and Data Analysis*, Ames, IO: The Iowa State University Press.
16. Kempthorne, O., Zyskind, G., Addelman, S., Throckmorton, T. N., White, R. F. (1961): *Analysis of Variance Procedures*, Contract AF 33(616)-5599, Project 7071, Task 70418, Aeronautical Research Laboratory, Office of Aerospace Research, United States Air Force, Wright-Patterson Air Force Base, Ohio.
17. Kutner, M. H. (1974): Hypothesis Testing in Linear Models (Eisenhart Model I), *The American*

- Statistician*, **28,3**, 98-100.
18. Miller, R. G. (1981): *Simultaneous Statistical Inference*, New York, NY: Springer-Verlag, Inc.
 19. Milliken, G. A., Johnson, D. E. (1974): *THE ANALYSIS OF MESSY DATA*, Van Nostrand Reinhold, New York, NY.
 20. Mood, A. M., Graybill, F. A., Boes, D. C. (1974): *Introduction to the Theory of Statistics*, New York, NY: McGraw-Hill, Inc.
 21. Overall, J. E., Spiegel, D. K. (1969): Concerning Least Squares Analysis of Experimental Data, *Psychological Bulletin*, **72,5**, 311-322.
 22. Pearson, E. S., Hartley, H. O. (1966): *Biometrika Tables for Statistician*, Vol 1, Cambridge University Press.
 23. Rao, C. R., Mitra, S. K. (1971): *Generalized Inverse of Matrices and its Applications*, New York, NY: John Wiley & Sons, Inc.
 24. SAS Institute, Inc. (1985): *SAS/STAT™ Guide for Personal Computers*, Version 6 Edition, Cary, NC: SAS Institute, Inc.
 25. Searle, S. R. (1971): *Linear Models*, New York, NY: John Wiley & Sons, Inc.
 26. Searle, S. R., Speed, F. M., Milliken, G. A. (1980): Population Marginal Means in the Linear Model: An Alternative to Least Squares Means, *The American Statistician*, **34,4**, 216-221.
 27. Searle, S. R. (1982): *Matrix Algebra Useful for Statistics*, New York, NY: John Wiley & Sons, Inc.
 28. Snedecor, G. W., Cochran, W. G. (1980): *Statistical Methods*, Ames, IO: The Iowa State University Press.
 29. Speed, F. M., Hocking, R. R. (1976): The Use of R()-Notation with Unbalanced Data, *The American Statistician*, **73,361**, 105-112.
 30. Speed, F. M., Hocking, R. R., Hackney, O. P. (1978): Methods of Analysis of Linear Models with Unbalanced Data, *The American Statistician*, **30,1**, 30-33.
 31. Strang, G. (1980): *Linear Algebra and its Applications*, Orlando, FL: Harcourt Brace Jovanovich, Inc.
 32. Vannucchi, H., Moreno, F. S., Amarante, A. R., Oliveira, J. E. D., Marchini, J. S. (1991): Plasma Amino Acid Patterns in Alcoholic Pellagra Patients, *Alcohol and Alcoholism*, **26,4**, 431-436.
 33. Wilk, M. B., Kempthorne, O. (1955): Fixed, Mixed, and Random Models, *Journal of The American Statistical Association*, **50**, 1144-1167.
 34. Yates, F (1934): The Analysis of Multiple Classification With Unequal Numbers in the Different Classes,

Journal of the American Statistical Association, 30, 51-66.

ÍNDICE REMISSIVO

A

Admissível *ver Média*

Ajuste 1, 3, 11, 47, 51, 55, 60, 61, 63,
64, 68, 70, 71, 73, 74,
80, 82, 105, 113, 118,
121, 122, 125-130, 133,
141, 143, 144, 146, 150,
166, 169, 170, 178-180,
187, 203, 205, 232-234,
243, 244, 246, 247, 251,
253, 281, 286, 295, 296,
298, 301, 303, 306, 307,
308

de constantes 169, 178, 187
sequencial 71

Aleatorização 34, 36

Análise

de covariância 4, 25, 281-286,
290, 292, 293
de variância 1, 3, 75, 78-80, 82,
121, 122, 150, 205, 212,
215, 231, 233, 264, 272,
277, 306, 308

Aproximação 1, 48, 49, 51, 60, 62, 65,
75, 231

Autovalor 13-16, 127, 217, 219

Autovetor 13, 14

simbólico 164, 185, 193, 205

Colinearidade 152

Combinação

de tratamentos 43, 195, 196,
199, 204

linear 11, 48, 50, 75, 87, 89, 92,
151, 152, 157, 176, 181,
231, 244, 246, 274

Comparação emparelhada 268, 269

Comparações múltiplas 264, 265

Complemento ortogonal (*v. tb*
Ortocomplemento) 59

Comprimento (*v. tb Norma*) 16, 61, 62,
231, 258, 264

Condições de Gauss-Markov 4, 96, 228,
234

Conjunto

de consonância 263

gerador 157, 158, 160, 179-181,
189

parcialmente ordenado 240

Consistente 53, 54, 56, 66, 89, 100-103,
106, 109, 133, 139

Contraste 22, 268, 269

Covariáveis 187, 281-287, 289-291, 293,
297

Crítérios de classificação 28-30, 38, 71,
161-163, 196, 199, 266,
268, 285, 286, 287, 289

C

Caselas vazias 170, 193-195, 197, 199

Coefficiente

angular 285, 286

de determinação 150, 151

D

Dados

balanceados 172, 175, 184-187,
189-191, 196, 201, 204,
266, 268, 289

desbalanceados 169, 170, 172,
 173, 179, 183-186,
 189-192, 195, 196, 200,
 204, 268, 289
 perdidos 198
 proporcionais 190
 Dados categóricos 130
 Decomposição 59, 78
 Delineamento experimental 169, 186,
 187, 266
 Densidade multivariada 216
 Dependência linear 152, 153
 Desigualdade de Cauchy-Schwartz 272
 Diagrama de pontos 297-299
 Dimensão 45, 51, 54, 57, 64, 88, 93,
 105, 116, 273
 Distância 7-9, 12, 16, 47, 50-52, 99, 125,
 239-241
 Euclidiana 99
 Distribuição
 de formas quadráticas 215
 normal 224, 226, 296, 297
 probabilística 4, 91, 152, 298
 qui quadrado 216, 217, 268
 t de Student 255, 260, 265

E

Efeito 31-33, 36, 38-41, 43, 44, 71-73,
 75, 76, 79-81, 150, 171,
 172, 175, 176, 178-180,
 185, 186, 188, 190-193,
 195-199, 203-205, 239,
 253, 285, 286, 291, 300,
 305
 contido em outro 175, 176, 186,
 195
 principal 175, 188, 190, 196,
 198, 203
 Efeitos classificatórios 176
 Eliminação de Gauss-Jordan 134
 Elipsóide 16, 128, 264
 Equações

de Aitken 126, 127, 129, 234,
 237
 normais 3, 53-58, 63, 65-67, 80,
 81, 83, 93, 99, 101, 104,
 114-116, 119, 133,
 139-142, 153, 154, 228,
 234, 245, 262, 281, 282,
 285
 normais do modelo restrito 101,
 104, 114, 115
 normais particionadas 65
 normais reduzidas 65-67, 282
 Erro experimental 284
 Error 81-83, 96, 122, 212, 253, 278,
 292, 308
 Esfera 16
 Espaço
 coluna 12, 60, 63, 231
 Euclidiano 273
 paramétrico 101, 115, 243
 vetorial 7
 base de 11, 40, 77, 179,
 242, 254, 274,
 281
 dimensão de 22, 23
 Estatística 1, 2, 1, 2, 152, 173, 215, 244,
 245, 249, 286, 295
 F 215, 249, 286
 Estimativa 96, 154, 159, 253, 255, 265,
 267, 272, 290

Estrutura
 de classificação 33, 39
 populacional 33
 desbalanceada 27, 204
 Exame dos resíduos 4, 295-297
 Experimento completamente aleatorizado
 35

F

Fator
 cruzado 40
 de classificação 20-22, 25-29
 embutido 22, 23, 26, 27, 29, 33,

39, 41, 45, 176,
187, 188
numérico 25, 26
Fatores cruzados 22, 23, 26, 29, 34,
38-40, 44, 95, 146, 303
Fatorial 189, 191, 193, 194, 196, 200
Feixe de linhas paralelas 285, 286
Formas quadráticas 14-16, 55, 63, 119,
125, 215, 219, 231, 248,
249, 263
Frequências não proporcionais 201
Frequências proporcionais 175, 201
Função
geratriz de momentos 216-218
indicadora 35
linear 89, 94, 268
Funções
estimáveis 3, 87, 90-93, 115,
116, 119-120, 158, 161,
163, 164, 166, 169, 178,
179, 181-185, 187, 189,
191-193, 195, 196, 198,
200, 203, 204, 206-212,
228, 230, 245, 247, 262,
263
forma geral das 161,
163, 164, 166,
182, 183, 185,
191-193, 198,
204, 206
linearmente independentes
93, 178, 262,
tipo I 169, 178-186,
189, 190, 203,
204, 206, 207,
228
tipo II 169, 187, 189,
190, 204, 208,
209
tipo III 169, 191, 204,
205, 209, 210
tipo IV 169, 195, 197,
199-201, 205,
211, 212
identificáveis 88, 90, 91
lineares 1, 87, 93, 159, 193,
269, 276

paramétricas 92, 114, 119, 129,
133, 243, 244,
254, 256, 257,
266

G

Gráfico probabilístico 298
Graus de liberdade 3, 64, 69, 70, 77-79,
81, 105, 106, 113, 121,
203, 205, 215, 219, 232,
231, 249, 255, 260, 263,
265, 267, 272

H

Heterogeneidade de coeficientes angulares
286
Hipótese
alternativa 243
nula 243, 244, 246, 247, 249
Hipóteses
de interesse 174
não testáveis 249, 251, 252
testáveis 244, 245, 251, 252
tipo I 187
tipo II 190, 191, 196
tipo III 192-195, 198
tipo IV 197, 198, 201
Histograma 298
Homocedasticidade 296

I

Identidade algébrica 31-33
Inadmissível *ver Média*
Inconsistente 2, 3, 53, 54
Independência 11, 26, 80, 92, 93, 101,
144-146, 150, 152, 160,

178, 195, 219,
224-227, 229-
232, 240, 243,
249, 254, 259,
262, 263 266,
268, 274, 275,
288, 296

Independência de formas quadráticas 219

Instabilidade numérica 152

Interação 38, 39, 71, 169, 170, 175, 185,
188, 193, 199, 203-205,
207, 209, 210, 212, 285,
287

Interações
de segunda ordem 199
de terceira ordem 200

Intervalo de consonância 2, 238, 244,
253, 256-258, 260, 261,
264-269, 272, 277

Inversa generalizada 57, 60, 80, 81, 154
g2 80, 81
reflexiva 154

Inversão no lugar 141

L

LSMEANS 291

M

Matriz

augmentada 136, 138-140
de posto incompleto 133
de projeção 11, 12, 14, 127
de variâncias e covariâncias 129,
234
do modelo 42, 43, 49, 75, 76,
120, 126, 281, 282, 296
idempotente 12-14, 56, 60, 66,
68, 69, 76, 110, 112,
216-220, 222, 223, 227,

229, 296

identidade 16, 56, 109, 137-140,
152, 159

positiva definida 15, 16, 125,
126, 128, 143, 233, 274

positiva semidefinida 15, 215

simétrica 14, 15, 66, 68, 110,
125, 215, 229

Média 1, 27, 30-33, 36-40, 44, 71, 72,
75-79, 81-84, 96, 121,
150, 164, 166, 169, 171,
172, 191, 196, 203,
206-212, 252, 264-269,
274, 276-278, 283, 285,
287-291, 293, 297

admissível 31-33, 39, 40

ajustada 287

de mínimos quadrados 287, 288

geral 27, 30, 33, 36, 38-40, 71

inadmissível, 31

marginal 171, 172, 287, 288

marginal populacional 287

parcial 30, 31

Melhor estimador linear não viciado 93,
129, 228, 230, 237, 244,
247, 255, 258, 262, 288,
289

Método
de Scheffé 266, 268, 272, 276,
277
de Tukey 266, 268, 276-278

Métodos gráficos para análise dos resíduos
297

Mínimo global 55, 103

Mínimos quadrados 1, 3, 8, 10-12, 47,
50-52, 55, 60, 61, 99,
125-127, 129, 130, 191,
234, 246, 287-289, 300,
303, 307

aproximação de 51

com restrições nos parâmetros 67

completos 191

ponderados 3, 125-127, 129,
130, 234, 300, 303, 307

Modelo
irrestrito 104, 114-116, 118-122,

170, 243, 246,
247

reparametrizado 62, 126, 233

Modelos

a k-partes 67

classificatórios 3, 19, 63

com covariáveis 187

com dois critérios de classificação
196

com três critérios de classificação
161-163

com um critério de classificação
161

completos 68

de Gauss-Markov 227-230, 231,
233, 243, 244, 248, 254,
258, 260, 267, 274, 287,
295, 296

de regressão 187

lineares 1, 2, 1-4, 7, 37, 60, 67,
130, 153, 227, 243, 281,
284, 295

restritos 3, 67, 100, 101, 104,
105, 113-115, 118-122,
170, 232, 243-246

Multiplicadores de Lagrange 100

N

Níveis do fator 21, 22, 24, 25, 30, 39,
277, 287, 288

Nível

de consonância 244, 256, 257, 260, 261,
264, 265, 267, 272, 277

de significância 165, 238,
240-244, 253, 267, 277

Norma 50, 63, 73, 119, 126, 127, 150,
151, 243, 246

ao quadrado 50, 63, 73, 151, 246

W mínima 126, 127

Normalidade (v. *tb Distribuição normal*)
227, 297, 298

O

Observação aberrante 299

Operação linha 8, 9, 136-143, 153, 157,
159

Ordenamento a priori 178

Ordenamento parcial 240, 241

Ortocomplemento (v. *tb Complemento
ortogonal*) 12, 13

Ortogonal (v. *tb Ortogonalidade*) 12, 54,
59-61, 68, 69, 126-128,
177, 180, 192, 193

Ortogonalidade (v. *tb Ortogonal*) 191,
192, 205

P

Parâmetro 1, 2, 36, 37, 42, 81, 89, 92,
99, 101, 115, 119, 120,
153, 161, 164, 165, 170,
179, 186, 188, 190, 196,
199, 204, 205, 216, 230,
232, 238, 243, 244, 249,
254, 256, 257, 259, 268,
282, 285

de não centralidade 216, 232,
249

Partição da soma de quadrados 70

Partições 20, 21, 24, 26, 70

Permutação

das colunas da matriz identidade
152

das colunas de X 153

Pivoteamento 143-148, 150, 179, 180

Plano completamente aleatorizado 289

Ponderação 234, 306, 308

Ponto aberrante 299

População 1, 21, 24, 26, 30, 32, 33, 35,
36, 239, 241, 256, 257

conceitual de respostas 32, 33,
35

Populações 39

Posto 9, 14, 45, 50, 60, 62-64, 69, 70,
76, 93, 100, 101, 116,
126, 127, 133, 139, 150,
153, 158, 160, 164, 178,
181, 217, 219, 224, 225,
229, 274
completo 9, 60, 62, 93, 101,
116, 126, 127,
139, 178, 274
incompleto 45, 50, 63, 126, 133,
150, 153
Probabilidade 35, 237, 239, 240, 242,
243, 254, 256, 257, 265,
266
PROC GLM 80, 121, 203, 252, 277, 289,
291, 306
Processo de Gauss-Jordan 134, 136
Produtos cruzados 34, 146, 303
Projeção 3, 9-14, 60-62, 67, 73, 91, 104,
119, 126-128, 150
não ortogonal 126-128
ortogonal 12, 60, 61
Projeções não ortogonais 3
Projektor 1, 7, 12, 60, 68, 69, 110, 177,
180, 296
ortogonal 12, 60, 68, 69, 177, 180

Q

Quadrado médio 79, 233
Quadrados ponderados de médias 169, 191
Qualidade do ajuste 63, 118, 233
Quantidade pivotal 257-260, 277
Quebra da soma de quadrados 79, 215
Qui quadrado
central 219, 232, 260, 261
não central 215, 218, 231, 232

R

R-Square 82, 122, 150, 212, 292, 308

Redução na soma de quadrados 70, 72, 74
Região
de consonância 4, 256, 261
Regressão 140, 187, 190, 285
Replicação 30, 35, 43, 70, 170, 189
Resíduos 4, 60, 61, 63, 64, 68-72, 75,
77, 79, 81, 105, 106,
113, 118, 119, 121, 129,
133, 141, 143, 144, 146,
150, 151, 179, 180, 229,
232, 233, 246, 247, 283,
291, 295-301, 303-308
em sequência temporal 297, 300
vetor de 60, 64, 246
vs. possíveis variáveis de interesse
297
vs. valores preditos 297, 301,
303, 304, 307, 308
Resolução numérica das equações normais
3, 133
Restrições
envolvendo funções estimáveis
116, 119, 120, 161, 162,
164, 170, 244
envolvendo funções não
estimáveis 114, 115
estimáveis 3, 165, 232
não estimáveis 3
Reversibilidade 142, 144, 147

S

SAS 2, 3, 80, 96, 121, 166, 203, 252,
277, 287, 289, 290, 306
Simetria 13
Simétrica 12-16, 39, 56, 60, 66, 68, 76,
107, 110, 111, 125, 127,
143, 215, 224, 229, 296
Singularidade 152
Sistema inconsistente 2, 3
Soma de quadrados 2, 3, 55, 63, 67, 68,
70-74, 76-79, 81, 115,
118, 119, 121, 129, 133,
141, 143, 144,

146, 150, 151,
165, 169, 170,
173, 178-180,
185, 187,
201-205, 213,
215, 228, 229,
231, 232, 243,
247, 250, 252,
281-283, 286,
291, 306

corrigida pela média 77, 78
do modelo 71, 73, 79, 231, 282
do resíduo 71, 72, 77, 118, 119,
121, 133, 141, 143, 144,
146, 150, 151, 179, 180,
229, 232, 247, 249, 283,
291

tipo I 72, 178, 179, 185, 186,
201, 203, 228, 286, 287,
291

tipo II 73, 187-190
tipo III 191, 195, 199, 201, 202,
204
tipo IV 191, 197-199, 201, 202,
204
total 70-72, 76-78, 215

Soma direta 60
Soma ponderada 172, 185
Subespaços ortogonais 60
Submodelos 67, 70, 145, 150, 179
SWEEP(A,k) 3, 57, 134, 140, 143-145,
149, 153, 160, 169, 179,
180

T

Tabela de análise de variância 78-80, 82,
121, 122, 205, 212, 277,
306, 308

Teorema
de Cochran 227
de Gauss-Markov 93, 96
de Pitágoras 63

Termo quadrático 300

Testes

de hipótese 2, 238
de significância 2-4, 87, 170,
173, 219, 238-240,
242-244, 249, 256, 265,
286
estatísticos 4, 237, 238

Tolerância 152

Tratamento 1, 32, 33, 35, 36, 43, 77, 79,
195, 196, 198, 199, 204,
239, 253, 272, 283-286,
291, 293, 306

U

Unidade experimental 35, 36

V

Valores

numéricos 25, 26, 49, 205
observados 35, 90, 215, 243,
285, 303
preditos 60, 129, 133, 297, 301,
303, 304, 306-308

Varição estudantizada 266, 277

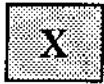
Variância 1, 3, 75, 78-80, 82, 95, 121,
122, 129, 150, 205, 212,
215, 231, 233, 261,
264-268, 272, 274, 277,
283, 295, 300, 302, 303,
306, 308

Variáveis

aritméticas 281
classificadoras 80, 164, 175,
187, 286
contínuas 175, 176
dependentes 80
independentes 144-146, 150
indicadoras 24-26, 49, 281
matemáticas 24, 37

Vetor

arbitrário 51, 57, 88, 106
auxiliar 145
de resíduos 60, 64, 246
de valores preditos 60
solução 99, 118, 119, 133, 138
solução das equações normais 55,
57, 63, 80, 83, 93, 104,
114-116, 119, 133, 228,
245, 262



X com posto incompleto 126, 150