



LEANDRO DA SILVA PEREIRA

**ABORDAGEM GEOMÉTRICA À TEORIA
DOS MODELOS DE GAUSS-MARKOV**

LAVRAS - MG

2013

LEANDRO DA SILVA PEREIRA

**ABORDAGEM GEOMÉTRICA À TEORIA DOS MODELOS DE
GAUSS-MARKOV**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-graduação em Estatística e Experimentação Agropecuária, área de concentração em Estatística e Experimentação Agropecuária, para a obtenção do título de Mestre.

Orientador
Dr. Lucas Monteiro Chaves

Coorientador
Dr. Devanil Jaques de Souza

**LAVRAS - MG
2013**

**Ficha Catalográfica Elaborada pela Divisão de Processos Técnicos da
Biblioteca da UFLA**

Pereira, Leandro da Silva.

Abordagem geométrica à teoria dos modelos de Gauss-Markov /
Leandro da Silva Pereira. – Lavras : UFLA, 2013.

130 p. : il.

Dissertação (mestrado) – Universidade Federal de Lavras, 2013.

Orientador: Lucas Monteiro Chaves.

Bibliografia.

1. Identidade de Fisher. 2. Dados faltantes e adicionais. 3. Geometria de delineamentos. 4. Delineamento experimental. I. Universidade Federal de Lavras. II. Título.

CDD - 519.57

LEANDRO DA SILVA PEREIRA

**ABORDAGEM GEOMÉTRICA À TEORIA DOS MODELOS DE
GAUSS-MARKOV**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-graduação em Estatística e Experimentação Agropecuária, área de concentração em Estatística e Experimentação Agropecuária, para a obtenção do título de Mestre.

APROVADA em 18 de abril de 2013.

Dr. Marcelo Silva de Oliveira

UFLA

Dra. Maria do Carmo Pacheco de Toledo Costa

UFLA

Dr. Lucas Monteiro Chaves
Orientador

**LAVRAS - MG
2013**

“Por vezes sentimos que aquilo que fazemos não é senão uma gota de
água no mar. Mas o mar seria menor se lhe faltasse uma gota.”

Madre Teresa de Calcutá

AGRADECIMENTOS

A Deus, pelo sol que ilumina os meus dias, pelos corpos celestes que clareiam as minhas noites mais escuras, pelas chuvas que lavam a minha alma, pelas primaveras que colorem a minha vida com as suas flores, pelos pássaros que alegam o meu espírito com os seus cantos, pelos alimentos que fortificam o meu corpo, pela verdade que aos poucos me liberta, pelo trabalho que me faz sentir útil, pela vida que me proporciona novas experiências e oportunidades de crescimento, e por tudo que ainda não tenho consciência de que ELE faz por mim.

Aos meus pais Messias e Valdelice que, pelo exemplo de humildade e dignidade, me mostram o sentido da vida.

À minha eterna companheira Mayumi, sinônimo de vida e esperança, por ser meu forte nas horas difíceis e meu encanto nas horas de paz.

À minha irmã Celiani e sua família, pelo apoio direto e indireto.

A todos os meus familiares e amigos, pelo apoio e carinho.

Ao meu orientador Lucas Monteiro Chaves, que pelos conhecimentos e esclarecimentos intelectuais e morais confiados a mim, me ensinou o verdadeiro sentido de se ensinar.

Ao coorientador Devanil Jaques de Souza, que de forma síncrona com as idéias do professor Lucas, contribui para minha formação moral e intelectual.

À Universidade Federal de Lavras (UFLA) e ao Departamento de Ciências Exatas (DEX), pela oportunidade de cursar o mestrado.

Aos professores do programa de pós-graduação em estatística e experimentação agropecuária da UFLA, pelas contribuições na minha formação durante as suas disciplinas, em especial aos professores Renato Ribeiro de Lima e Daniel Furtado Ferreira.

À professora Maria do Carmo P. T. Costa, pelas mais relevantes contribuições à dissertação.

Às funcionárias do DEX: Edila, Josiane Cristina, Josiane Oliveira, Kelly, Maria, Miriam e Selma, pela amizade e carinho.

À CAPES, pela concessão da bolsa de estudos, tornando financeiramente possível a realização do mestrado. Aos colegas da Engenharia de Controle e Automação pelos momentos de lazer e pelos conselhos “utilizados”.

A todos os colegas de turma pela companhia nos estudos, nos momentos de lazer e nos momentos de tensão de véspera das provas.

A todos que de alguma forma contribuíram, mesmo inconsciente, para a realização deste trabalho.

RESUMO

Abordagens geométricas à duas identidades em soma de quadrados: a identidade de Fisher e uma identidade de David-Hartley-Pearson são apresentadas. Uma detalhada descrição dos modelos de Gauss-Markov, também baseadas em argumentos geométricos, é desenvolvida, obtendo-se uma nova interpretação geométrica do Teorema de Gauss-Markov. Nesta mesma linha, são apresentados dois artigos clássicos de Kruskal relativos a dados faltantes e dados adicionais, com exemplos em delineamentos experimentais.

Palavras-chave: Modelos de Gauss-Markov. Abordagens geométricas. Identidade de Fisher. Dados faltantes e adicionais.

ABSTRACT

Geometric approach for two identities in square sum: Fisher's identity and David-Hartley-Pearson's identity are shown. A detailed description of the Gauss-Markov models, based on geometric arguments as well, is developed acquiring a new geometric interpretation for the Gauss-Markov theorem. In this same train of thought, two classical Kruskal's articles, relating to missing and extra observations, are presented with applications in experimental designs.

Keywords: Gauss-Markov models. Geometric Approach. Fisher's identity. Missing and extra observations.

LISTA DE FIGURAS

Figura 1	Representação geométrica de um vetor de n observações	15
Figura 2	Ângulo entre os vetores \mathbf{x} e \mathbf{y}	17
Figura 3	Projeção do vetor \mathbf{v} na direção de \mathbf{w}	19
Figura 4	Identificação geométrica do triângulo de variâncias	23
Figura 5	Reta ajustada por quadrados mínimos	24
Figura 6	Matriz \mathbf{X} de transformação do \mathbb{R}^2 para \mathbb{R}^n	25
Figura 7	Matriz \mathbf{X} da transformação \mathbb{R}^p para \mathbb{R}^n , $n \geq p$	27
Figura 8	Interpretação geométrica das matrizes \mathbf{A} e \mathbf{A}'	33
Figura 9	Interpretação geométrica do Coeficiente de Determinação	37
Figura 10	Projeção de \mathbf{y} no subespaço de tratamentos	45
Figura 11	Decomposição de \mathbf{y} em componentes ortogonais	45
Figura 12	Vetores da decomposição da identidade de Pearson	47
Figura 13	Representação do vetor aleatório \mathbf{Y} e as correlações (projeções) de \mathbf{x} e \mathbf{z}	54
Figura 14	Representação das projeções do vetor \mathbf{Y} nas direções de \mathbf{x} e \mathbf{z} com relação ao novo produto interno	56
Figura 15	Relação entre os subespaços de parâmetros e do modelo	63
Figura 16	Esfera de variância	64
Figura 17	Comparação entre as projeções ortogonal e oblíqua	65
Figura 18	Transformação das nuvens de variâncias do espaço $Im(\mathbf{X})$ para o espaço de parâmetros	67
Figura 19	Relação dos vetores \mathbf{y} e $\hat{\mathbf{y}}$ com o subespaço ortogonal a $Im(\mathbf{X})$	70
Figura 20	Interpretação geométrica das somas de quadrados da tabela de análise de variância	82
Figura 21	Quadro de Análise de Variância dos dados do exemplo DIC	93
Figura 22	Interpretação geométrica de um delineamento em blocos casualizado (DIC)	105
Figura 23	Quadro de Análise de Variância dos dados do exemplo DBC	113
Figura 24	Projeções para dados faltantes	115
Figura 25	Projeções para dados extras	122

LISTA DE TABELAS

Tabela 1	Esquema da análise de variância para experimentos no delineamento inteiramente casualizado.	79
Tabela 2	Produtividade dos cultivares de mandioca, em t/ha.	86
Tabela 3	Esquema da análise de variância para experimentos no delineamento em blocos casualizado.	99
Tabela 4	Pesos médios dos frutos da macieira, em gramas.	106

SUMÁRIO

1	INTRODUÇÃO	12
2	ABORDAGEM GEOMÉTRICA DE ALGUMAS IDENTIDADES DE SOMA DE QUADRADOS UTILIZADAS EM ESTATÍSTICA	14
2.1	Interpretação geométrica do método dos quadrados mínimos	24
2.2	Uma interpretação geométrica do coeficiente de determinação	35
2.3	A identidade de Fisher	39
2.4	A identidade de David-Hartley-Pearson	46
3	ABORDAGEM GEOMÉTRICA À TEORIA DOS MODELOS DE GAUSS-MARKOV	50
3.1	O modelo	50
3.2	A geometria do teorema de Gauss-Markov	63
3.3	Condições para a igualdade entre os estimadores de Gauss-Markov e Mínimos Quadrados	67
3.4	Interpretação geométrica de um delineamento inteiramente casualizado	77
3.5	Interpretação geométrica de um delineamento em blocos casualizado	94
3.6	A teoria de Gauss-Markov na presença de dados faltantes . .	112
3.7	A teoria de Gauss-Markov na presença de dados adicionais .	121
4	CONCLUSÃO	125
	REFERÊNCIAS	126
	APÊNDICE	128

1 INTRODUÇÃO

Não é usual em livros textos e mesmo em artigos científicos relacionar Estatística e Geometria. Apesar de estes dois ramos da Matemática parecerem tão distantes entre si, tal fato não é verdade como expresso por

[...]historicamente, a análise de variância e os métodos de regressão foram desenvolvidos nas décadas iniciais deste século, usando geometria. Fisher usava a geometria como inspiração para desenvolvimento de suas idéias. Isso é evidente em suas anotações e é claramente relatado, por sua filha, em sua biografia.(SAVILLE et al., 1991, p.4 apud BOX,1978, tradução nossa).

A Geometria pode ser empregada com o intuito de deixar claro os conceitos estatísticos, permitindo uma maior compreensão da Estatística como um todo. “Uma figura forma uma ponte natural entre o problema e a solução.”(SAVILLE et al., 1991, p.3, tradução nossa), ou seja, é um caminho natural na resolução de problemas.

A aplicação da Geometria à teoria dos modelos lineares é intuitiva. No entanto, a abordagem dos livros textos se baseiam quase que exclusivamente na álgebra matricial, evitando os conceitos geométricos de subespaços vetoriais e projeções ortogonais. Mesmo para a teoria dos estimadores de quadrados mínimos, poucas referências adotam uma abordagem mais geométrica. Um fato ainda menos conhecido é que é possível uma abordagem geométrica a outras áreas estatísticas, notadamente à teoria dos delineamentos experimentais, bem como à análise de variância. Nesta linha, pode-se citar os livros “Statistical Methods: The Geometric Approach”(SAVILLE et al., 1991), “Linear Models in Statistics”(RENCHER et

al., 2008) e “Design of Comparative Experiments”(BAILEY, 2008).

Neste trabalho, um esforço foi feito no sentido do maior uso possível dos conceitos geométricos. Primeiramente, alguns fatos elementares como média e variância são estabelecidos em termos de vetores e projeções ortogonais, bem como a teoria da regressão linear. Estabelecidos estes conceitos, duas identidades em soma de quadrados, a identidade de Fisher e a identidade de David-Hartley-Pearson, são demonstradas do ponto de vista geométrico, ressaltando que a demonstração algébrica, usualmente utilizada, é complexa e não intuitiva.

O modelo de Gauss-Markov, em sua forma mais geral, é estudado do ponto de vista do uso da distância de Mahalanobis, o que permitiu uma abordagem bastante genérica. Em particular, fica evidenciado a distinção entre o estimador de quadrados mínimos e o estimador de Gauss-Markov. Uma interpretação geométrica e bastante didática do teorema de Gauss-Markov é apresentada. Acredita-se que tal interpretação seja original e didaticamente interessante.

Dois artigos clássicos de Kruskal são estudados em detalhes. No primeiro são demonstradas condições para a igualdade entre os estimadores de Gauss-Markov e de Quadrados Mínimos. No segundo, é apresentado um método geométrico para a imputação de dados. É também apresentado um método para a inclusão, no modelo linear, de dados adicionais. Uma aplicação à teoria geral dos delineamentos inteiramente casualizados e em blocos é apresentada.

2 ABORDAGEM GEOMÉTRICA DE ALGUMAS IDENTIDADES DE SOMA DE QUADRADOS UTILIZADAS EM ESTATÍSTICA

No presente trabalho, matrizes serão denotadas por letras maiúsculas em negrito ($\mathbf{A}, \mathbf{B}, \mathbf{\Gamma}$), vetores coluna por letras minúsculas em negrito ($\mathbf{a}, \mathbf{b}, \boldsymbol{\gamma}$), escalares por letras minúsculas em itálico (a, b, γ) e a transposta de uma matriz representada por um apóstrofo junto a esta, isto é, $\mathbf{X}^t = \mathbf{X}'$. Será utilizada também a notação $\mathbf{Y} = (Y_1, \dots, Y_n)'$ para representação de vetores aleatórios e $\mathbf{y} = (y_1, \dots, y_n)'$ para valores observados do vetor aleatório \mathbf{Y} . A imagem de uma transformação linear \mathbf{X} será representada por $Im(\mathbf{X})$, a dimensão do subespaço gerado por uma transformação linear \mathbf{X} é denotado por $Dim(\mathbf{X})$, o núcleo (ou *Kernel*) de uma transformação é denotado por $Ker(\mathbf{X})$ e o traço de uma matriz quadrada \mathbf{X} é representado por $tr(\mathbf{X})$.

Geometricamente, um vetor linha ou coluna com p elementos pode ser associado com um ponto num espaço p -dimensional. Os elementos no vetor são as coordenadas do ponto. Nesse sentido, um conjunto de n observações pode ser geometricamente interpretado como um vetor $\mathbf{y} = (y_1, \dots, y_n)'$, de dimensão $n \times 1$, em que cada coordenada deste corresponde à uma observação. A Figura 1 ilustra este tipo de representação.

Seguem agora algumas definições básicas, porém bastante relevantes aos resultados apresentados, conforme constam em Boldrini et al.(1986).

Definição 1. *O produto interno entre dois vetores $\mathbf{x} = (x_1, \dots, x_n)'$ e $\mathbf{y} = (y_1, \dots, y_n)'$ é definido por*

$$\mathbf{x} \cdot \mathbf{y} = \langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}'\mathbf{y} = (x_1, \dots, x_n) \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = x_1y_1 + \dots + x_ny_n.$$

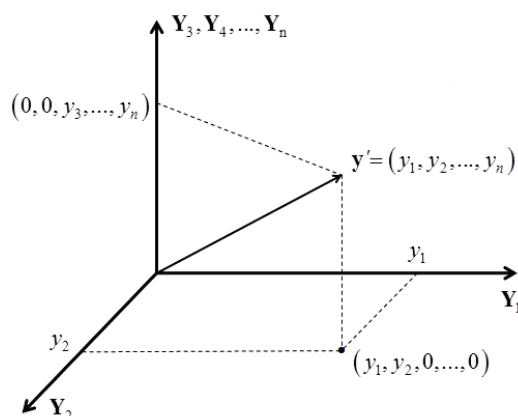


Figura 1 Representação geométrica de um vetor de n observações

Definição 2. A norma de um vetor \mathbf{x} , denotada por $\|\mathbf{x}\|$, é um número real associado a este tal que

$$\|\mathbf{x}\| = \sqrt{\mathbf{x} \cdot \mathbf{x}}.$$

Definição 3. Um vetor \mathbf{u} é dito unitário se $\|\mathbf{u}\| = \sqrt{\mathbf{u} \cdot \mathbf{u}} = 1$.

Proposição 1. O produto interno satisfaz

$$\mathbf{x} \cdot \mathbf{y} = \|\mathbf{x}\| \|\mathbf{y}\| \cos \theta,$$

em que θ é o ângulo formado pelos vetores \mathbf{x} e \mathbf{y} .

Demonstração. Aplicando-se a Lei dos Cossenos no triângulo da Figura 2 tem-se

$$\|\mathbf{x} - \mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 - 2 \|\mathbf{x}\| \|\mathbf{y}\| \cos \theta. \quad (2.1)$$

Desenvolvendo o primeiro membro da equação (2.1):

$$\|\mathbf{x} - \mathbf{y}\|^2 = \|(x_1, \dots, x_n)' - (y_1, \dots, y_n)'\|^2$$

$$\begin{aligned}
&= \left(\sqrt{(x_1 - y_1, \dots, x_n - y_n)' \cdot (x_1 - y_1, \dots, x_n - y_n)'} \right)^2 \\
&= (x_1 - y_1)^2 + \dots + (x_n - y_n)^2 \\
&= x_1^2 - 2x_1y_1 + y_1^2 + \dots + x_n^2 - 2x_ny_n + y_n^2 \\
&= x_1^2 + \dots + x_n^2 + y_1^2 + \dots + y_n^2 - 2x_1y_1 - \dots - 2x_ny_n \\
&= (x_1^2 + \dots + x_n^2 + y_1^2 + \dots + y_n^2) - 2(x_1y_1 + \dots + x_ny_n) \\
&= (x_1^2 + \dots + x_n^2 + y_1^2 + \dots + y_n^2) - 2(\mathbf{x} \cdot \mathbf{y}).
\end{aligned}$$

De forma análoga, desenvolvendo o segundo membro da equação (2.1) vem:

$$\begin{aligned}
\|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 - 2\|\mathbf{x}\|\|\mathbf{y}\|\cos\theta &= (\sqrt{\mathbf{x} \cdot \mathbf{x}})^2 + (\sqrt{\mathbf{y} \cdot \mathbf{y}})^2 - 2\|\mathbf{x}\|\|\mathbf{y}\|\cos\theta \\
&= (\mathbf{x} \cdot \mathbf{x}) + (\mathbf{y} \cdot \mathbf{y}) - 2\|\mathbf{x}\|\|\mathbf{y}\|\cos\theta \\
&= (x_1^2 + \dots + x_n^2) + (y_1^2 + \dots + y_n^2) \\
&\quad - 2\|\mathbf{x}\|\|\mathbf{y}\|\cos\theta \\
&= (x_1^2 + \dots + x_n^2 + y_1^2 + \dots + y_n^2) \\
&\quad - 2\|\mathbf{x}\|\|\mathbf{y}\|\cos\theta.
\end{aligned}$$

Assim, substituindo estes dois desenvolvimentos na equação (2.1), tem-se:

$$\begin{aligned}
(x_1^2 + \dots + x_n^2 + y_1^2 + \dots + y_n^2) - 2(\mathbf{x} \cdot \mathbf{y}) &= (x_1^2 + \dots + x_n^2 + y_1^2 + \\
&\quad + \dots + y_n^2) \\
&\quad - 2\|\mathbf{x}\|\|\mathbf{y}\|\cos\theta.
\end{aligned}$$

Assim, $-2(\mathbf{x} \cdot \mathbf{y}) = -2\|\mathbf{x}\|\|\mathbf{y}\|\cos\theta$ e portanto $\mathbf{x} \cdot \mathbf{y} = \|\mathbf{x}\|\|\mathbf{y}\|\cos\theta$.

□

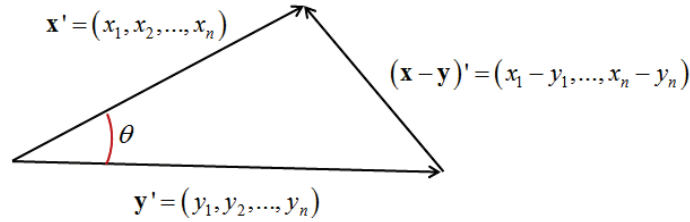


Figura 2 Ângulo entre os vetores \mathbf{x} e \mathbf{y}

O vetor unitário na direção de um vetor \mathbf{w} é dado por $\mathbf{u} = \frac{\mathbf{w}}{\|\mathbf{w}\|}$.

Proposição 2. A projeção ortogonal de um vetor \mathbf{v} qualquer na direção de um vetor \mathbf{w} é dada por

$$P_{\mathbf{w}}\mathbf{v} = \left(\mathbf{v} \cdot \frac{\mathbf{w}}{\|\mathbf{w}\|} \right) \frac{\mathbf{w}}{\|\mathbf{w}\|}.$$

Demonstração. Pela Figura 3, $P_{\mathbf{w}}\mathbf{v} = \alpha \frac{\mathbf{w}}{\|\mathbf{w}\|}$. Portanto, basta determinar o valor do escalar α de forma que $\left(\mathbf{v} - \alpha \frac{\mathbf{w}}{\|\mathbf{w}\|} \right) \cdot \left(\alpha \frac{\mathbf{w}}{\|\mathbf{w}\|} \right) = 0$ para se ter uma projeção ortogonal. Logo,

$$\begin{aligned} 0 &= \left(\mathbf{v} - \alpha \frac{\mathbf{w}}{\|\mathbf{w}\|} \right) \cdot \left(\alpha \frac{\mathbf{w}}{\|\mathbf{w}\|} \right) \\ &= \left((v_1, \dots, v_n)' - \frac{\alpha (w_1, \dots, w_n)'}{\|\mathbf{w}\|} \right) \cdot \left(\frac{\alpha (w_1, \dots, w_n)'}{\|\mathbf{w}\|} \right) \\ &= \left(v_1 - \frac{\alpha w_1}{\|\mathbf{w}\|}, \dots, v_n - \frac{\alpha w_n}{\|\mathbf{w}\|} \right)' \cdot \left(\alpha \frac{w_1}{\|\mathbf{w}\|}, \dots, \alpha \frac{w_n}{\|\mathbf{w}\|} \right)' \\ &= \left(v_1 - \frac{\alpha w_1}{\|\mathbf{w}\|} \right) \left(\alpha \frac{w_1}{\|\mathbf{w}\|} \right) + \dots + \left(v_n - \frac{\alpha w_n}{\|\mathbf{w}\|} \right) \left(\alpha \frac{w_n}{\|\mathbf{w}\|} \right) \\ &= \frac{\alpha w_1 v_1}{\|\mathbf{w}\|} - \frac{\alpha^2 w_1^2}{\|\mathbf{w}\|^2} + \dots + \frac{\alpha w_n v_n}{\|\mathbf{w}\|} - \frac{\alpha^2 w_n^2}{\|\mathbf{w}\|^2} \\ &= \frac{\alpha w_1 v_1 \|\mathbf{w}\| - \alpha^2 w_1^2 + \dots + \alpha w_n v_n \|\mathbf{w}\| - \alpha^2 w_n^2}{\|\mathbf{w}\|^2} \end{aligned}$$

$$\begin{aligned}
&= \alpha w_1 v_1 \|\mathbf{w}\| - \alpha^2 w_1^2 + \cdots + \alpha w_n v_n \|\mathbf{w}\| - \alpha^2 w_n^2 \\
&= \alpha (w_1 v_1 \|\mathbf{w}\| + \cdots + w_n v_n \|\mathbf{w}\|) - \alpha^2 (w_1^2 + \cdots + w_n^2) \\
&= (w_1 v_1 \|\mathbf{w}\| + \cdots + w_n v_n \|\mathbf{w}\|) - \alpha (w_1^2 + \cdots + w_n^2) \\
&= \|\mathbf{w}\| (w_1 v_1 + \cdots + w_n v_n) - \alpha (\mathbf{w} \cdot \mathbf{w}) \\
&= \|\mathbf{w}\| (w_1 v_1 + \cdots + w_n v_n) - \alpha \|\mathbf{w}\|^2 \\
&= (w_1 v_1 + \cdots + w_n v_n) - \alpha \|\mathbf{w}\| \\
&= (\mathbf{w} \cdot \mathbf{v}) - \alpha \|\mathbf{w}\|.
\end{aligned}$$

Portanto, $\alpha \|\mathbf{w}\| = (\mathbf{w} \cdot \mathbf{v})$, isto é,

$$\begin{aligned}
\alpha &= \frac{(\mathbf{w} \cdot \mathbf{v})}{\|\mathbf{w}\|} \\
&= \mathbf{v} \cdot \frac{\mathbf{w}}{\|\mathbf{w}\|}.
\end{aligned}$$

Desta forma, com o valor de α determinado, a projeção ortogonal é determinada por

$$P_{\mathbf{w}} \mathbf{v} = \alpha \frac{\mathbf{w}}{\|\mathbf{w}\|} = \left(\mathbf{v} \cdot \frac{\mathbf{w}}{\|\mathbf{w}\|} \right) \frac{\mathbf{w}}{\|\mathbf{w}\|}.$$

□

Pode-se também observar, na Figura 3, que

$$\|\mathbf{v}\|^2 = \|P_{\mathbf{w}} \mathbf{v}\|^2 + \|\mathbf{v} - P_{\mathbf{w}} \mathbf{v}\|^2. \quad (2.2)$$

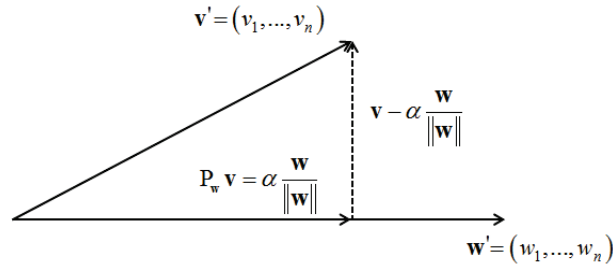


Figura 3 Projeção do vetor \mathbf{v} na direção de \mathbf{w}

Proposição 3. Dado um vetor $\mathbf{x} = (x_1, \dots, x_n)'$, sua projeção ortogonal na direção do vetor $\mathbf{u} = (a, \dots, a)'$, $a \in \mathbb{R}$, é o vetor $\bar{\mathbf{x}} = (\bar{x}, \dots, \bar{x})'$, em que $\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$ é a média aritmética das coordenadas do vetor \mathbf{x} .

Demonstração.

$$\begin{aligned}
 P_{\mathbf{u}}\mathbf{x} &= \left(\mathbf{x} \cdot \frac{\mathbf{u}}{\|\mathbf{u}\|} \right) \frac{\mathbf{u}}{\|\mathbf{u}\|} \\
 &= \left((x_1, \dots, x_n)' \cdot \frac{(a, \dots, a)'}{\sqrt{(a, \dots, a)' \cdot (a, \dots, a)'}} \right) \frac{(a, \dots, a)'}{\sqrt{(a, \dots, a)' \cdot (a, \dots, a)'}} \\
 &= \left((x_1, \dots, x_n)' \cdot \frac{(a, \dots, a)'}{\sqrt{\underbrace{a^2 + \dots + a^2}_{n \text{ termos}}}} \right) \frac{(a, \dots, a)'}{\sqrt{\underbrace{a^2 + \dots + a^2}_{n \text{ termos}}}} \\
 &= \left((x_1, \dots, x_n)' \cdot \frac{(a, \dots, a)'}{a\sqrt{n}} \right) \frac{(a, \dots, a)'}{a\sqrt{n}} \\
 &= \left(\frac{x_1}{\sqrt{n}} + \dots + \frac{x_n}{\sqrt{n}} \right) \left(\frac{1}{\sqrt{n}}, \dots, \frac{1}{\sqrt{n}} \right)' \\
 &= \left(\frac{x_1 + \dots + x_n}{\sqrt{n}} \right) \left(\frac{1}{\sqrt{n}}, \dots, \frac{1}{\sqrt{n}} \right)' \\
 &= \left(\frac{x_1 + \dots + x_n}{\sqrt{n}} \frac{1}{\sqrt{n}}, \dots, \frac{x_1 + \dots + x_n}{\sqrt{n}} \frac{1}{\sqrt{n}} \right)' \\
 &= \left(\frac{x_1 + \dots + x_n}{n}, \dots, \frac{x_1 + \dots + x_n}{n} \right)'
 \end{aligned}$$

$$\begin{aligned}
&= \left(\frac{\sum_{i=1}^n x_i}{n}, \dots, \frac{\sum_{i=1}^n x_i}{n} \right)' \\
&= (\bar{x}, \dots, \bar{x})'.
\end{aligned}$$

□

Desta forma, de acordo com (2.2), segue a identidade usual

$$\|\mathbf{v}\|^2 = \|\mathbf{P}_1 \mathbf{v}\|^2 + \|\mathbf{v} - \mathbf{P}_1 \mathbf{v}\|^2,$$

logo $v_1^2 + \dots + v_n^2 = n(\bar{v})^2 + \sum_{i=1}^n (v_i - \bar{v})^2$, pois $\mathbf{P}_1 \mathbf{v} = (\bar{v}, \dots, \bar{v})'$ e $\|\mathbf{P}_1 \mathbf{v}\|^2 = n(\bar{v})^2$.

Proposição 4. *Em uma amostra com média μ , se alguns termos são substituídos pelo valor médio destes, a nova amostra possui a mesma média e variância menor.*

Demonstração. A média se mantém já que, para uma amostra

$$\mathbf{x} = (x_1, x_2, \dots, x_r, x_{r+1}, \dots, x_n)',$$

substituindo os r primeiros termos pela média destes, então

$$\mathbf{x}_M = \left(\overbrace{\frac{x_1 + \dots + x_r}{r}, \dots, \frac{x_1 + \dots + x_r}{r}}^{r \text{ termos}}, x_{r+1}, x_{r+2}, \dots, x_n \right)'.$$

O cálculo da nova média fica:

$$\begin{aligned}
m &= \frac{\overbrace{\frac{x_1 + \dots + x_r}{r} + \dots + \frac{x_1 + \dots + x_r}{r}}^{r \text{ termos}} + \sum_{i=1}^{n-r} x_{r+i}}{n} \\
&= \frac{r \frac{(x_1 + \dots + x_r)}{r} + \sum_{i=1}^{n-r} x_{r+i}}{n} \\
&= \frac{\sum_{k=1}^r x_k + \sum_{i=1}^{n-r} x_{r+i}}{n} \\
&= \frac{\sum_{i=1}^n x_i}{n} \\
&= \mu.
\end{aligned}$$

$$\therefore \bar{\mathbf{x}}_{\mathbf{M}} = \bar{\mathbf{x}}.$$

Para verificar que a variância diminui, será utilizado um argumento geométrico. O triângulo pontilhado destacado na Figura 4 é retângulo em θ , ou seja, o produto interno é nulo, pois

$$\begin{aligned}
(\mathbf{x} - \mathbf{x}_{\mathbf{M}}) \cdot (\bar{\mathbf{x}} - \mathbf{x}_{\mathbf{M}}) &= (x_1 - \bar{x}_r, \dots, x_r - \bar{x}_r, x_{n-r} - x_{n-r}, \dots, x_n - x_n)' \\
&\quad \cdot (\mu - \bar{x}_r, \dots, \mu - \bar{x}_r, \mu - x_{n-r}, \dots, \mu - x_n)' \\
&= (x_1 - \bar{x}_r)(\mu - \bar{x}_r) + (x_2 - \bar{x}_r)(\mu - \bar{x}_r) + \\
&\quad + \dots + (x_r - \bar{x}_r)(\mu - \bar{x}_r) + 0 + 0 + \dots + 0 \\
&= (\mu - \bar{x}_r)(x_1 + x_2 + \dots + x_r - r\bar{x}_r)
\end{aligned}$$

$$\begin{aligned}
&= (\mu - \bar{x}_r) \left(\sum_{i=1}^r x_i - r \frac{\sum_{i=1}^r x_i}{r} \right) \\
&= (\mu - \bar{x}_r) \left(\sum_{i=1}^r x_i - \sum_{i=1}^r x_i \right) \\
&= 0.
\end{aligned}$$

Logo,

$$\begin{aligned}
&\|\mathbf{x} - \mathbf{x}_M\|^2 + \|\mathbf{x}_M - \bar{\mathbf{x}}\|^2 = \|\mathbf{x} - \bar{\mathbf{x}}\|^2 \\
&\frac{1}{n-1} \|\mathbf{x} - \mathbf{x}_M\|^2 + \frac{1}{n-1} \|\mathbf{x}_M - \bar{\mathbf{x}}\|^2 = \frac{1}{n-1} \|\mathbf{x} - \bar{\mathbf{x}}\|^2 \\
&\frac{1}{n-1} \|\mathbf{x}_M - \bar{\mathbf{x}}\|^2 \leq \frac{1}{n-1} \|\mathbf{x} - \bar{\mathbf{x}}\|^2 \Rightarrow \text{var}(\mathbf{x}_M) \leq \text{var}(\mathbf{x}).
\end{aligned}$$

□

2.1 Interpretação geométrica do método dos quadrados mínimos

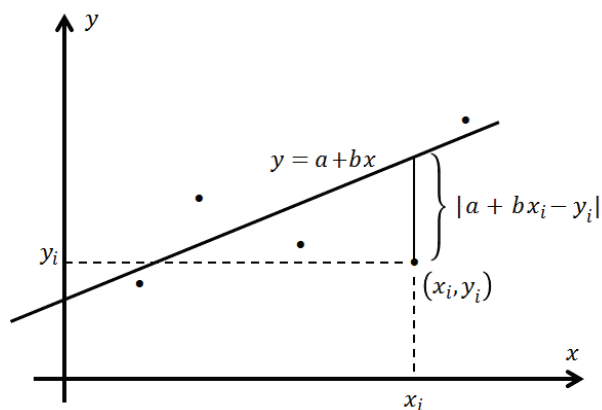


Figura 5 Retra ajustada por quadrados mínimos

Sejam $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ os valores observados. Deseja-se então ajustar uma reta $\hat{y} = bx + a$ de tal forma que a equação (2.3) seja minimizada.

$$\sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n (y_i - (a + bx_i))^2 \quad (2.3)$$

A Figura 5 ilustra tal contexto.

O método usual para se obter a e b é derivar (2.3) em relação a estes, igualar as equações derivadas a 0 e resolver o sistema. Uma abordagem geométrica é obtida da seguinte forma:

Se $\mathbf{Y}_{n \times 1} = (y_1, \dots, y_n)'$, $\boldsymbol{\beta}_{2 \times 1} = (a, b)'$, $\boldsymbol{\varepsilon}_{n \times 1} = (\varepsilon_1, \dots, \varepsilon_n)'$ e

$$\mathbf{X}_{n \times 2} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}, \text{ então}$$

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \Rightarrow \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

isto é, $y_i = a + bx_i + \varepsilon_i$, $i = 1, \dots, n$. Portanto, reescrevendo (2.3) tem-se

$$\sum_{i=1}^n (y_i - (a + bx_i))^2 = \|\mathbf{y}_{n \times 1} - \mathbf{X}_{n \times 2} \boldsymbol{\beta}_{2 \times 1}\|^2.$$

Em termos de transformações lineares em que a matriz $\mathbf{X}_{n \times 2}$ atua como uma transformação do espaço de parâmetros para o espaço dos dados, o conceito pode ser representado pela Figura 6.

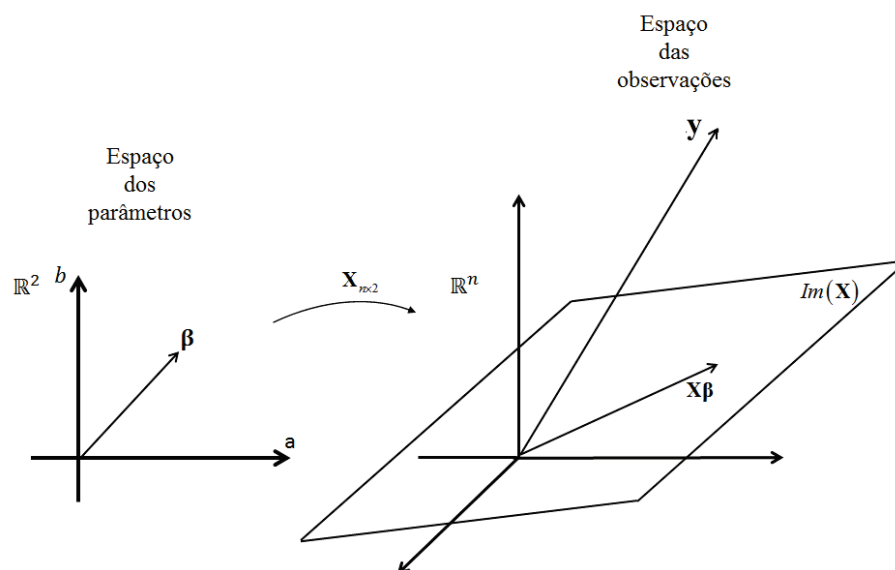


Figura 6 Matriz \mathbf{X} de transformação do \mathbb{R}^2 para \mathbb{R}^n

Assim,

$$\sum_{i=1}^n (y_i - (a + bx_i))^2 = \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2.$$

Logo, minimizar $\sum_{i=1}^n (y_i - (a + bx_i))^2$ equivale a minimizar $\|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2$ que, claramente, ocorre quando $\mathbf{X}\boldsymbol{\beta}$ é a projeção ortogonal do vetor de dados \mathbf{y} no subespaço $Im(\mathbf{X})$. Este vetor projeção será denominado $P_{Im(\mathbf{X})}\mathbf{y}$, também denotado por $\hat{\mathbf{y}}$, vetor de dados ajustados. Note que, como o posto (coluna) de \mathbf{X} é 2, a dimensão da imagem de \mathbf{X} é $Dim(Im(\mathbf{X})) = 2$.

O caso geral em que se tem várias covariáveis, isto é, (x_1, \dots, x_p) , que foram observadas n vezes, se quer ajustar o modelo

$$\boxed{y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \varepsilon_i} \quad i = 1, \dots, n$$

o que matricialmente equivale a

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}_{n \times 1} = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix}_{n \times p+1} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}_{p+1 \times 1} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}_{n \times 1} \quad (2.4)$$

e se quer minimizar $\sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^p x_{ij} \beta_j \right)^2 = \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2$. O valor mínimo é obtido pela projeção ortogonal do vetor de dados \mathbf{y} no subespaço $Im(\mathbf{X})$, que possui nesse caso dimensão p . A Figura 7 ilustra esta situação.

Geometricamente a representação é similar à Figura 6, diferindo apenas na dimensão do espaço dos parâmetros que agora passa a ser o \mathbb{R}^p (ver Figura 7). Considerando o posto de \mathbf{X} igual a p (posto completo), tem-se que

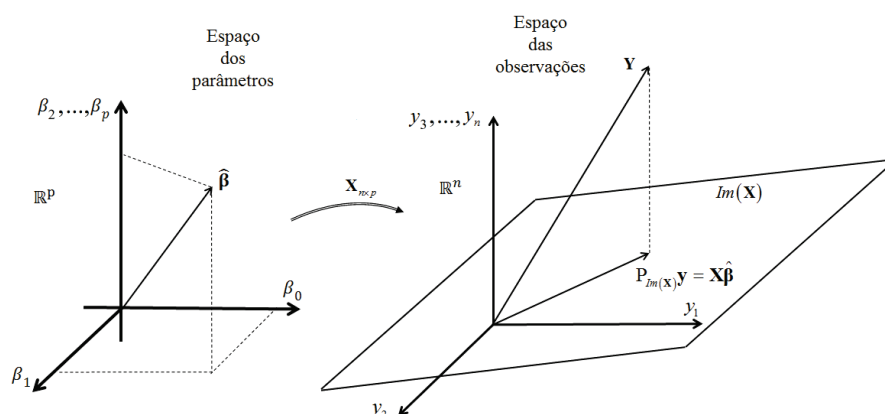


Figura 7 Matriz \mathbf{X} da transformação \mathbb{R}^p para \mathbb{R}^n , $n \geq p$

$\text{Dim}(Im(\mathbf{X})) = p$.

Para se obter uma expressão para $P_{Im(\mathbf{X})}\mathbf{y}$ é necessário a teoria das matrizes de projeção.

Definição 4. Uma matriz quadrada é dita matriz de projeção ou projetor se é idempotente, isto é, $\mathbf{A}^2 = \mathbf{A}$.

Pela definição 4, tem-se que um projetor \mathbf{A} restrito a $Im(\mathbf{A})$ é a identidade, ou seja,

$$\begin{aligned} \mathbf{A}(\mathbf{A}\mathbf{z}) &= \mathbf{A}^2\mathbf{z} \\ &= \mathbf{A}\mathbf{z}. \end{aligned}$$

De outra forma, $\mathbf{A}(\mathbf{A}\mathbf{z}) = \mathbf{I}(\mathbf{A}\mathbf{z})$, $\forall \mathbf{A}$ projetor e sendo $\mathbf{z} \in \mathbb{R}^n$ um vetor qualquer. Observe que, se \mathbf{I} é a matriz identidade, $\mathbf{I} - \mathbf{A}$ também é um projetor pois

$$\begin{aligned} (\mathbf{I} - \mathbf{A})^2 &= \mathbf{I} - 2\mathbf{A} + \mathbf{A}^2 \\ &= \mathbf{I} - 2\mathbf{A} + \mathbf{A} \end{aligned}$$

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

Como

$$\begin{aligned} \mathbf{A}((\mathbf{I} - \mathbf{A})\mathbf{z}) &= \mathbf{A}(\mathbf{z} - \mathbf{A}\mathbf{z}) \\ &= \mathbf{A}\mathbf{z} - \mathbf{A}^2\mathbf{z} \\ &= \mathbf{A}\mathbf{z} - \mathbf{A}\mathbf{z} \\ &= \mathbf{0}, \end{aligned}$$

segue que $Im(\mathbf{I} - \mathbf{A}) \subset Ker(\mathbf{A})$. Além do mais, se $\mathbf{z} \in Ker(\mathbf{A})$ então

$$\begin{aligned} (\mathbf{I} - \mathbf{A})\mathbf{z} &= \mathbf{z} - \mathbf{A}\mathbf{z} \\ &= \mathbf{z} - \mathbf{0}. \end{aligned}$$

Logo, $\mathbf{z} \in Im(\mathbf{I} - \mathbf{A})$ e portanto $Im(\mathbf{I} - \mathbf{A}) = Ker(\mathbf{A})$. Assim, segue então que $Im(\mathbf{A}) = Ker(\mathbf{I} - \mathbf{A})$.

Definição 5. Uma matriz de projeção \mathbf{A} é dita um projetor ortogonal se, para um dado vetor \mathbf{w} , $\mathbf{A}\mathbf{w} - \mathbf{w}$ é perpendicular ao subespaço $Im(\mathbf{A})$.

Proposição 5. Uma matriz de projeção é simétrica se, e somente se, é um projetor ortogonal.

Demonstração. Suponha \mathbf{A} simétrica e \mathbf{v} e \mathbf{w} vetores quaisquer. Utilizando as propriedades de produto interno e sendo $\mathbf{A}^2 = \mathbf{A}$, tem-se que

$$\begin{aligned} \langle \mathbf{v} - \mathbf{A}\mathbf{v}, \mathbf{A}\mathbf{w} \rangle &= \langle \mathbf{v}, \mathbf{A}\mathbf{w} \rangle - \langle \mathbf{A}\mathbf{v}, \mathbf{A}\mathbf{w} \rangle \\ &= \langle \mathbf{v}, \mathbf{A}\mathbf{w} \rangle - \langle \mathbf{v}, \mathbf{A}'\mathbf{A}\mathbf{w} \rangle \end{aligned}$$

$$\begin{aligned}
&= \langle \mathbf{v}, \mathbf{A}\mathbf{w} \rangle - \langle \mathbf{v}, \mathbf{A}\mathbf{A}\mathbf{w} \rangle \\
&= \langle \mathbf{v}, \mathbf{A}\mathbf{w} \rangle - \langle \mathbf{v}, \mathbf{A}^2\mathbf{w} \rangle \\
&= \langle \mathbf{v}, \mathbf{A}\mathbf{w} \rangle - \langle \mathbf{v}, \mathbf{A}\mathbf{w} \rangle \\
&= \mathbf{0}, \forall \mathbf{v}, \mathbf{w}.
\end{aligned}$$

Portanto, $\mathbf{v} - \mathbf{A}\mathbf{v}$ é perpendicular a $Im(\mathbf{A})$.

Seja agora $\mathbf{A}\mathbf{v} - \mathbf{v}$ perpendicular a $Im(\mathbf{A})$. Tem-se então que, para \mathbf{w} qualquer,

$$\begin{aligned}
0 &= \langle \mathbf{A}\mathbf{v} - \mathbf{v}, \mathbf{A}\mathbf{w} \rangle \\
&= \langle \mathbf{A}\mathbf{v}, \mathbf{A}\mathbf{w} \rangle - \langle \mathbf{v}, \mathbf{A}\mathbf{w} \rangle,
\end{aligned}$$

ou seja,

$$\langle \mathbf{A}\mathbf{v}, \mathbf{A}\mathbf{w} \rangle = \langle \mathbf{v}, \mathbf{A}\mathbf{w} \rangle. \quad (2.5)$$

Também, para $(\mathbf{A}\mathbf{w}) - \mathbf{w}$ perpendicular a $\mathbf{A}\mathbf{v}$,

$$\begin{aligned}
0 &= \langle \mathbf{A}\mathbf{w} - \mathbf{w}, \mathbf{A}\mathbf{v} \rangle \\
&= \langle \mathbf{A}\mathbf{w}, \mathbf{A}\mathbf{v} \rangle - \langle \mathbf{w}, \mathbf{A}\mathbf{v} \rangle,
\end{aligned}$$

ou seja, $\langle \mathbf{A}\mathbf{w}, \mathbf{A}\mathbf{v} \rangle = \langle \mathbf{w}, \mathbf{A}\mathbf{v} \rangle$ e portanto

$$\langle \mathbf{A}\mathbf{v}, \mathbf{A}\mathbf{w} \rangle = \langle \mathbf{A}\mathbf{v}, \mathbf{w} \rangle. \quad (2.6)$$

Consequentemente, das equações (2.5) e (2.6)

$$\langle \mathbf{A}\mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{v}, \mathbf{A}\mathbf{w} \rangle.$$

□

Como exemplo segue a proposição.

Proposição 6. *A matriz que projeta um vetor $\mathbf{y} = (y_1, \dots, y_n)'$ qualquer na direção do vetor $\mathbf{1} = (1, \dots, 1)'$, isto é, a matriz de projeção na média, é a matriz*

$$\begin{aligned} \mathbf{A} &= \begin{bmatrix} \frac{1}{n} & \frac{1}{n} & \dots & \frac{1}{n} \\ \frac{1}{n} & \frac{1}{n} & \dots & \frac{1}{n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{n} & \frac{1}{n} & \dots & \frac{1}{n} \end{bmatrix}_{n \times n} \\ &= \frac{1}{n} \begin{bmatrix} 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 1 \end{bmatrix}_{n \times n}. \end{aligned}$$

Demonstração. Basta mostrar que \mathbf{A} é uma matriz de projeção, \mathbf{A} é simétrica e que $\mathbf{A}(\mathbf{y}) = \bar{y}$. \mathbf{A} é matriz de projeção pois

$$\begin{aligned} \mathbf{A}^2 &= \mathbf{A}\mathbf{A} \\ &= \begin{bmatrix} \frac{1}{n} & \frac{1}{n} & \dots & \frac{1}{n} \\ \frac{1}{n} & \frac{1}{n} & \dots & \frac{1}{n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{n} & \frac{1}{n} & \dots & \frac{1}{n} \end{bmatrix}_{n \times n} \begin{bmatrix} \frac{1}{n} & \frac{1}{n} & \dots & \frac{1}{n} \\ \frac{1}{n} & \frac{1}{n} & \dots & \frac{1}{n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{n} & \frac{1}{n} & \dots & \frac{1}{n} \end{bmatrix}_{n \times n} \end{aligned}$$

$$\begin{aligned}
&= \begin{bmatrix} \overbrace{\frac{1}{n^2} + \dots + \frac{1}{n^2}}^{n \text{ termos}} & \overbrace{\frac{1}{n^2} + \dots + \frac{1}{n^2}}^{n \text{ termos}} & \dots & \overbrace{\frac{1}{n^2} + \dots + \frac{1}{n^2}}^{n \text{ termos}} \\ \overbrace{\frac{1}{n^2} + \dots + \frac{1}{n^2}}^{n \text{ termos}} & \overbrace{\frac{1}{n^2} + \dots + \frac{1}{n^2}}^{n \text{ termos}} & \dots & \overbrace{\frac{1}{n^2} + \dots + \frac{1}{n^2}}^{n \text{ termos}} \\ \vdots & \vdots & \ddots & \vdots \\ \overbrace{\frac{1}{n^2} + \dots + \frac{1}{n^2}}^{n \text{ termos}} & \overbrace{\frac{1}{n^2} + \dots + \frac{1}{n^2}}^{n \text{ termos}} & \dots & \overbrace{\frac{1}{n^2} + \dots + \frac{1}{n^2}}^{n \text{ termos}} \end{bmatrix}_{n \times n} \\
&= \begin{bmatrix} n \frac{1}{n^2} & n \frac{1}{n^2} & \dots & n \frac{1}{n^2} \\ n \frac{1}{n^2} & n \frac{1}{n^2} & \dots & n \frac{1}{n^2} \\ \vdots & \vdots & \ddots & \vdots \\ n \frac{1}{n^2} & n \frac{1}{n^2} & \dots & n \frac{1}{n^2} \end{bmatrix}_{n \times n} \\
&= \begin{bmatrix} \frac{1}{n} & \frac{1}{n} & \dots & \frac{1}{n} \\ \frac{1}{n} & \frac{1}{n} & \dots & \frac{1}{n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{n} & \frac{1}{n} & \dots & \frac{1}{n} \end{bmatrix}_{n \times n} \\
&= \mathbf{A},
\end{aligned}$$

e também

$$\begin{aligned}
\mathbf{A}' &= \begin{bmatrix} \frac{1}{n} & \frac{1}{n} & \dots & \frac{1}{n} \\ \frac{1}{n} & \frac{1}{n} & \dots & \frac{1}{n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{n} & \frac{1}{n} & \dots & \frac{1}{n} \end{bmatrix}_{n \times n} \\
&= \mathbf{A}.
\end{aligned}$$

Portanto, sendo $\mathbf{A}^2 = \mathbf{A}$ e $\mathbf{A}' = \mathbf{A}$ temos que \mathbf{A} é matriz de projeção ortogonal.

Como

$$\begin{aligned}
 \mathbf{A}(\mathbf{y}) &= \begin{bmatrix} \frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} \\ \frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} \end{bmatrix}_{n \times n} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}_{n \times 1} \\
 &= \begin{bmatrix} \frac{1}{n}y_1 + \frac{1}{n}y_2 + \cdots + \frac{1}{n}y_n \\ \frac{1}{n}y_1 + \frac{1}{n}y_2 + \cdots + \frac{1}{n}y_n \\ \vdots \\ \frac{1}{n}y_1 + \frac{1}{n}y_2 + \cdots + \frac{1}{n}y_n \end{bmatrix}_{n \times 1} \\
 &= \begin{bmatrix} \frac{1}{n}(y_1 + y_2 + \cdots + y_n) \\ \frac{1}{n}(y_1 + y_2 + \cdots + y_n) \\ \vdots \\ \frac{1}{n}(y_1 + y_2 + \cdots + y_n) \end{bmatrix}_{n \times 1} \\
 &= \begin{bmatrix} \frac{\sum_{i=1}^n y_i}{n} \\ \frac{\sum_{i=1}^n y_i}{n} \\ \vdots \\ \frac{\sum_{i=1}^n y_i}{n} \end{bmatrix}_{n \times 1} \\
 &= \begin{bmatrix} \bar{y} \\ \bar{y} \\ \vdots \\ \bar{y} \end{bmatrix} \\
 &= \bar{\mathbf{y}},
 \end{aligned}$$

temos então que $\mathbf{A}(\mathbf{y}) = \bar{\mathbf{y}}$.

Logo, \mathbf{A} é a matriz de projeção na média. □

Uma interpretação geométrica relativa à matriz transposta \mathbf{A}' (operador adjunto) pode ser visualizada na Figura 8.

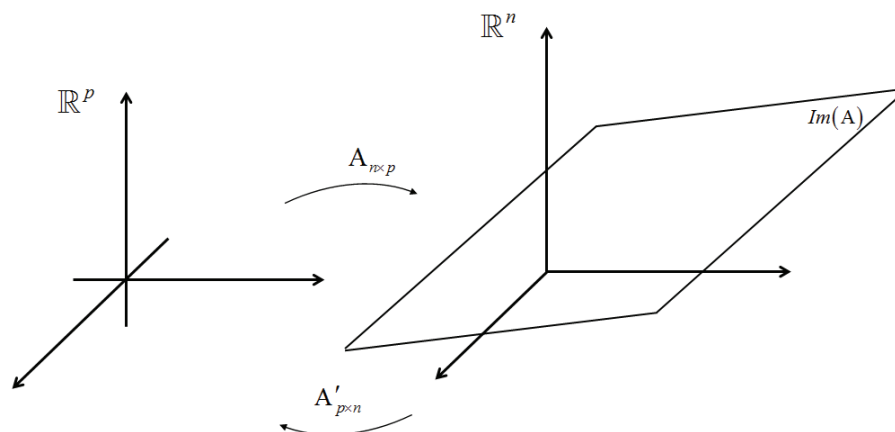


Figura 8 Interpretação geométrica das matrizes \mathbf{A} e \mathbf{A}'

Para se obter uma expressão para a projeção $P_{Im(\mathbf{X})}\mathbf{y} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$, será apresentada na Proposição 7 uma dedução baseada apenas em argumentos geométricos.

Proposição 7. *A projeção ortogonal do vetor de dados \mathbf{y} no subespaço $Im(\mathbf{X})$ é*

$$P_{Im(\mathbf{X})}\mathbf{y} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}.$$

Demonstração. Seja $\mathbf{Y} = (y_1, \dots, y_n)$. Em razão dos erros, tem-se que a probabilidade $P[\mathbf{Y} \in Im(\mathbf{X})] = 0$. Como queremos minimizar $L(\boldsymbol{\beta}) = \|\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}\|^2$ e o vetor $\mathbf{X}\boldsymbol{\beta}$ pertence à imagem de \mathbf{X} , segue que $L(\boldsymbol{\beta})$ é mínimo quando $\mathbf{X}\boldsymbol{\beta}$ é a projeção ortogonal de \mathbf{Y} na $Im(\mathbf{X})$, com notação $P_{Im(\mathbf{X})}(\mathbf{Y})$, e, consequentemente, existe um vetor $\hat{\boldsymbol{\beta}}$ de parâmetros tal que $\mathbf{X}\hat{\boldsymbol{\beta}} = P_{Im(\mathbf{X})}(\mathbf{Y})$ já

que a matriz \mathbf{X} é de posto completo. Considere ainda o vetor de parâmetros $\beta = \mathbf{X}'P_{Im(\mathbf{X})}(\mathbf{Y})$. Observe que, como $\mathbf{Y} - P_{Im(\mathbf{X})}(\mathbf{Y})$ é ortogonal a qualquer vetor na imagem de \mathbf{X} , $\mathbf{Y} - P_{Im(\mathbf{X})}(\mathbf{Y})$ está no espaço nulo de \mathbf{X} , isto é, $Ker(\mathbf{X})$. Segue então que $\mathbf{X}'(\mathbf{Y} - P_{Im(\mathbf{X})}(\mathbf{Y})) = \mathbf{0}$. Portanto, se pode escrever que

$$\begin{aligned}\beta &= \mathbf{X}'P_{Im(\mathbf{X})}(\mathbf{Y}) \\ &= \mathbf{X}'P_{Im(\mathbf{X})}(\mathbf{Y}) + \mathbf{X}'(\mathbf{Y} - P_{Im(\mathbf{X})}(\mathbf{Y})) \\ &= \mathbf{X}'\mathbf{Y}.\end{aligned}\tag{2.7}$$

Como $\mathbf{X}\hat{\beta} = P_{Im(\mathbf{X})}(\mathbf{Y})$, pode-se também escrever que

$$\beta = \mathbf{X}'\mathbf{X}\hat{\beta}.\tag{2.8}$$

Assim, de (2.7) e (2.8) segue que

$$\begin{aligned}\beta &= \mathbf{X}'\mathbf{X}\hat{\beta} \\ &= \mathbf{X}'\mathbf{Y} \\ \hat{\beta} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y},\end{aligned}$$

conforme representado na Figura 7. □

Segue da Proposição 7 que a estimativa do vetor de parâmetros β , que nos dá o ajuste de quadrados mínimos, é

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y},$$

pois

$$\begin{aligned}\mathbf{X}\hat{\boldsymbol{\beta}} &= \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \\ &= \mathbf{P}_{Im(\mathbf{X})}\mathbf{y}.\end{aligned}$$

O modelo ajustado fica então da forma

$$\hat{\mathbf{y}} = \hat{\beta}_0 + \hat{\beta}_1\mathbf{x}_1 + \hat{\beta}_2\mathbf{x}_2 + \cdots + \hat{\beta}_p\mathbf{x}_p.$$

2.2 Uma interpretação geométrica do coeficiente de determinação

Uma medida da qualidade do ajuste dado por $\hat{\mathbf{y}} = \hat{\beta}_0 + \hat{\beta}_1\mathbf{x}_1 + \cdots + \hat{\beta}_p\mathbf{x}_p$ é o coeficiente de determinação R^2 . Desta forma, seja SSR a Soma de Quadrados da Regressão (*Regression Sum of Squares*), definida por $SSR = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$ e seja também SST a Soma de Quadrados Total (*Total Sum of Squares*), definida por $SST = \sum_{i=1}^n (y_i - \bar{y})^2$. Nestas condições, segundo Rencher et al. (2008), o coeficiente de determinação é definido por:

$$\begin{aligned}R^2 &= \frac{SSR}{SST} \\ &= \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \\ &= \frac{\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} - n(\bar{y})^2}{\mathbf{y}'\mathbf{y} - n(\bar{y})^2}.\end{aligned}$$

Portanto, reescrevendo a igualdade anterior,

$$\begin{aligned}
R^2 &= \frac{\hat{\boldsymbol{\beta}}' \mathbf{X}' \mathbf{y} - n(\bar{y})^2}{\mathbf{y}' \mathbf{y} - n(\bar{y})^2} \\
&= \frac{(\mathbf{X} \hat{\boldsymbol{\beta}})' \mathbf{y} - \|\bar{\mathbf{y}}\|^2}{\|\mathbf{y}\|^2 - \|\bar{\mathbf{y}}\|^2} \\
&= \frac{\|\mathbf{X} \hat{\boldsymbol{\beta}}\| \|\mathbf{y}\| \cos \theta - \|\bar{\mathbf{y}}\|^2}{\|\mathbf{y}\|^2 - \|\bar{\mathbf{y}}\|^2} \\
&= \frac{\|\mathbf{X} \hat{\boldsymbol{\beta}}\|^2 - \|\bar{\mathbf{y}}\|^2}{\|\mathbf{y}\|^2 - \|\bar{\mathbf{y}}\|^2} \\
&= \frac{\|\hat{\mathbf{y}}\|^2 - \|\bar{\mathbf{y}}\|^2}{\|\mathbf{y}\|^2 - \|\bar{\mathbf{y}}\|^2} \\
&= \frac{\|\hat{\mathbf{y}} - \bar{\mathbf{y}}\|^2}{\|\mathbf{y} - \bar{\mathbf{y}}\|^2} \\
&= \left(\frac{\|\hat{\mathbf{y}} - \bar{\mathbf{y}}\|}{\|\mathbf{y} - \bar{\mathbf{y}}\|} \right)^2 \\
&= (\cos \alpha)^2 \\
&= \cos^2(\alpha).
\end{aligned}$$

R^2 admite uma interpretação geométrica conforme Figura 9.

Proposição 8. *O vetor de médias dos dados ajustados é igual ao vetor de média dos dados, isto é,*

$$\hat{\bar{\mathbf{y}}} = \bar{\mathbf{y}}$$

Demonstração. Seja $[\mathbf{1}]_{n \times n}$ a matriz em que se tem o elemento $a_{ij} = 1, \forall i, j \in \{1, 2, \dots, n\}$. Seja também $\mathbf{P}^{\perp}_{Im(\mathbf{X})} \mathbf{y}$ a componente do vetor \mathbf{y} que pertence ao espaço vetorial ortogonal a $Im(\mathbf{X})$. Assim,

$$\mathbf{y} = \mathbf{P}_{Im(\mathbf{X})} \mathbf{y} + \mathbf{P}^{\perp}_{Im(\mathbf{X})} \mathbf{y}.$$

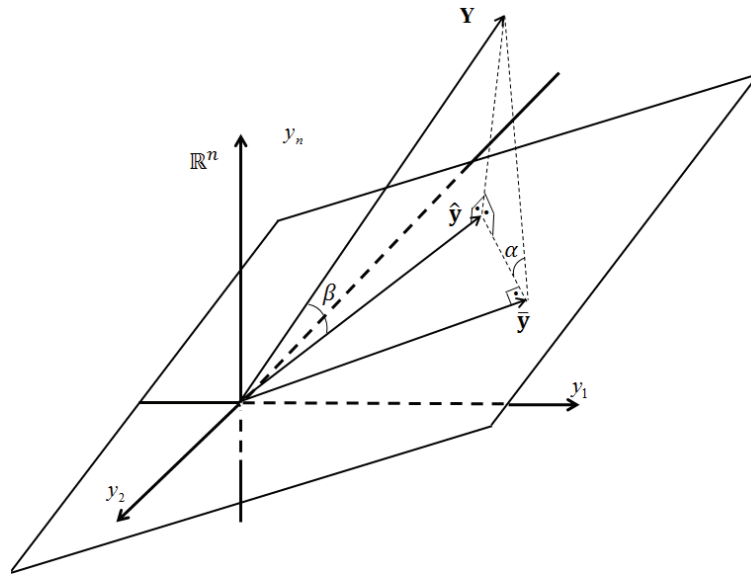


Figura 9 Interpretação geométrica do Coeficiente de Determinação

Queremos provar que $\left(\frac{1}{n}[1]_{n \times n}\right) \mathbf{y} = \left(\frac{1}{n}[1]_{n \times n}\right) \mathbf{X}\hat{\beta}$. Portanto,

$$\begin{aligned} \left(\frac{1}{n}[1]_{n \times n}\right) \mathbf{y} &= \left(\frac{1}{n}[1]_{n \times n}\right) \left[P_{Im(\mathbf{X})} \mathbf{y} + P^{\perp}_{Im(\mathbf{X})} \mathbf{y} \right] \\ &= \left(\frac{1}{n}[1]_{n \times n}\right) P_{Im(\mathbf{X})} \mathbf{y} + \left(\frac{1}{n}[1]_{n \times n}\right) P^{\perp}_{Im(\mathbf{X})} \mathbf{y}. \end{aligned}$$

Note que, sendo a matriz identidade $\mathbf{I}_{n \times n}$ um projetor ortogonal, então

$$P^{\perp}_{Im(\mathbf{X})} \mathbf{y} = \left(\mathbf{I} - \frac{1}{n}[1]_{n \times n} \right) \mathbf{y}. \quad (2.9)$$

Multiplicando a Equação (2.9) por $\frac{1}{n}[1]_{n \times n}$ tem-se

$$\frac{1}{n}[1]_{n \times n} \left(P^{\perp}_{Im(\mathbf{X})} \mathbf{y} \right) = \frac{1}{n}[1]_{n \times n} \left(\mathbf{I} - \frac{1}{n}[1]_{n \times n} \right) \mathbf{y}$$

$$\begin{aligned}
&= \left[\frac{1}{n} [1]_{n \times n} \mathbf{I} - \left(\frac{1}{n} [1]_{n \times n} \right) \left(\frac{1}{n} [1]_{n \times n} \right) \right] \mathbf{y} \\
&= \left[\frac{1}{n} [1]_{n \times n} - \left(\frac{1}{n} [1]_{n \times n} \right)^2 \right] \mathbf{y} \\
&= \left[\frac{1}{n} [1]_{n \times n} - \frac{1}{n} [1]_{n \times n} \right] \mathbf{y} \\
&= [0]_{n \times n} \mathbf{y} \\
&= [0]_{n \times n},
\end{aligned}$$

isto é, $\frac{1}{n} [1]_{n \times n} \mathbf{P}^\perp_{Im(\mathbf{X})} \mathbf{y} = [0]_{n \times n}$. Logo,

$$\begin{aligned}
\left(\frac{1}{n} [1]_{n \times n} \right) \mathbf{P}_{Im(\mathbf{X})} \mathbf{y} + \left(\frac{1}{n} [1]_{n \times n} \right) \mathbf{P}^\perp_{Im(\mathbf{X})} \mathbf{y} &= \left(\frac{1}{n} [1]_{n \times n} \right) \mathbf{P}_{Im(\mathbf{X})} \mathbf{y} \\
&\quad + [0]_{n \times n} \\
&= \left(\frac{1}{n} [1]_{n \times n} \right) \mathbf{X} \hat{\boldsymbol{\beta}}.
\end{aligned}$$

□

A subseção seguinte apresenta os resultados de SILVEIRA et al.(2012) que aborda geometricamente uma identidade proposta por Fisher.

2.3 A identidade de Fisher

Conforme dito inicialmente, os métodos fundamentais da experimentação (*análise de variância e métodos de regressão*) foram desenvolvidos com o auxílio da geometria. “Seu fundador, R. A. Fisher, era um astuto geômetra e obteve suas inspirações pensando em figuras”(SAVILLE et al., 1991, p.4, tradução nossa). Tal afirmação ressalta de forma clara a ligação entre geometria e estatística, presente nos trabalhos de Fisher. Neste sentido, segue uma interpretação geométrica para a Equação (2.10), identidade esta proposta por Fisher (SEARLE,1987).

$$\begin{aligned} (y_{ijk} - \bar{y}...) &= (\bar{y}_{i..} - \bar{y}...) + (\bar{y}_{.j.} - \bar{y}...) \\ &+ (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}...) + (y_{ijk} - \bar{y}_{ij.}) \end{aligned} \quad (2.10)$$

Tal identidade tem uma justificativa estatística razoável uma vez que os termos estão relacionados com a diferença de médias das várias hipóteses possíveis como média dos tratamentos, média geral e médias dos efeitos de cada fator individualmente. Mas tal fato por si só não justifica a identidade, pois certamente Fisher procurou uma identidade que tivesse a propriedade de uma decomposição em quadrados, como de fato ocorre, pois

$$\begin{aligned} \sum_{i,j,k} (y_{ijk} - \bar{y}...)^2 &= \sum_{i,j,k} \left[(\bar{y}_{i..} - \bar{y}...)^2 + (\bar{y}_{.j.} - \bar{y}...)^2 \right. \\ &+ (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}...)^2 \\ &\left. + (y_{ijk} - \bar{y}_{ij.})^2 \right] \end{aligned} \quad (2.11)$$

Assim, seja o espaço vetorial \mathbb{R}^n em que n é igual ao produto IJK , em que I representa o fator, J o nível e K a repetição de cada tratamento. As respostas y_{ijk} serão ordenadas de tal forma que cada uma destas corresponda a uma coordenada e, para tanto, será utilizada a ordem lexicográfica em relação a i, j, k . Portanto, as respostas podem ser consideradas como um vetor ordenado

$$\mathbf{y} = (y_{111}, y_{112}, \dots, y_{11K}, y_{121}, y_{122}, \dots, y_{12K}, \dots, y_{1J1}, \dots, y_{1JK}, y_{211}, y_{212}, \dots, y_{21K}, \dots, y_{IJ1}, \dots, y_{IJK})'$$

Utilizando-se a notação usual de modelos lineares tem-se as somas e médias em relação aos vários índices, como por exemplo

$$y_{...} = \sum_i \sum_j \sum_k y_{ijk} \qquad \bar{y}_{...} = \frac{1}{IJK} \sum_i \sum_j \sum_k y_{ijk}$$

$$y_{i..} = \sum_j \sum_k y_{ijk} \qquad \bar{y}_{i..} = \frac{1}{JK} \sum_j \sum_k y_{ijk}$$

De forma similar, se pode construir os vetores n -dimensionais relativos a estas somas e médias, conforme abaixo:

$$\mathbf{y}_{...} = (y_{...}, y_{...}, \dots, y_{...}, y_{...})'$$

$$\bar{\mathbf{y}}_{...} = (\bar{y}_{...}, \dots, \bar{y}_{...}, \dots, \bar{y}_{...}, \bar{y}_{...})'$$

$$\mathbf{y}_{.jk} = (y_{.11}, y_{.12}, y_{.13}, \dots, y_{.1K}, y_{.21}, y_{.22}, y_{.23}, \dots, y_{.2K}, \dots, y_{.J1}, y_{.J2}, \dots, y_{.JK})'$$

Seja o subespaço vetorial \mathbb{R}^n constituído pelos vetores que, para tratamentos iguais, possuem coordenadas iguais. Tais vetores, representando respostas, ocorreriam caso não houvesse aleatoriedade experimental, já que o vetor estaria

perfeitamente ajustado a um dado modelo. Este subespaço será denotado espaço dos tratamentos e será representado por

$$V_T = \{(\underbrace{a, a, \dots, a}_K, \underbrace{b, b, \dots, b}_K, \underbrace{c, c, \dots, c}_K, \dots, \underbrace{d, d, \dots, d}_K)'\}, a, b, c, \dots, d \in \mathbb{R},$$

onde se tem IJ seqüências com respostas iguais. Claramente, $Dim(V_T) = IJ$. De forma análoga, define-se o subespaço V_F formado pelos vetores que possuem coordenadas iguais para entradas com o mesmo índice i , isto é,

$$V_F = \{(\underbrace{a, a, \dots, a}_{JK}, \underbrace{b, b, \dots, b}_{JK}, \underbrace{c, c, \dots, c}_{JK}, \underbrace{d, d, \dots, d}_{JK})'\},$$

em que cada uma das seqüências de letras iguais é de tamanho JK e se tem I seqüências. Observa-se que $Dim(V_F) = I$. Da mesma forma, o subespaço V_G é formado pelos vetores que possuem coordenadas iguais para entradas com o mesmo índice j , ou seja,

$$V_G = \{(\underbrace{a, a, \dots, a}_K, \underbrace{b, b, \dots, b}_K, \underbrace{c, c, \dots, c}_K, \dots, \underbrace{d, d, \dots, d}_K, \dots, \underbrace{a, a, \dots, a}_K, \underbrace{b, b, \dots, b}_K, \underbrace{c, c, \dots, c}_K, \dots, \underbrace{d, d, \dots, d}_K, \dots, \underbrace{a, a, \dots, a}_K, \underbrace{b, b, \dots, b}_K, \underbrace{c, c, \dots, c}_K, \dots, \underbrace{d, d, \dots, d}_K)'\},$$

sendo que $Dim(V_G) = J$. Nota-se que

$$V_F \cap V_G = V_0$$

$$= \{(a, a, a, \dots, a)', a \in \mathbb{R}\}.$$

Denotando agora W_F e W_G como sendo subespaços de V_F e V_G respectivamente, definidos pelos vetores ortogonais ao subespaço V_0 , tem-se então que

$$W_F = \{(a, a, \dots, a, b, b, \dots, b, \dots, c, c, \dots, c, d, d, \dots, d)'\},$$

onde $JKa + JKb + \dots + JKc + JKd = 0$, e também

$$\begin{aligned} W_G = \{ & (a, a, \dots, a, b, b, \dots, b, \dots, c, c, \dots, c, d, d, \dots, d, a, a, \dots \\ & \dots, a, b, b, \dots, b, \dots, c, c, \dots, c, d, d, \dots, d, \dots, a, a, \dots, a, b, b, \dots \\ & \dots, b, \dots, c, c, \dots, c, d, d, \dots, d)'\}, \end{aligned}$$

sendo

$$\begin{aligned} & Ka + Kb + \dots + Kc + Kd + \\ & + Ka + Kb + \dots + Kc + Kd + \\ & + \dots + Ka + Kb + \dots + Kc + Kd = IKa + IKb + \dots + IKc + IKd \\ & = 0. \end{aligned}$$

Em termos vetoriais as equações (2.10) e (2.11) são expressas por

$$y - \bar{y}\dots = (\bar{y}_{i..} - \bar{y}\dots) + (\bar{y}_{.j.} - \bar{y}\dots) + [(\bar{y}_{ij.} - \bar{y}_{i..}) - (\bar{y}_{.j.} - \bar{y}\dots)] + (y_{ijk} - \bar{y}_{ij.}) \quad (2.12)$$

e

$$\begin{aligned} \|y - \bar{y}\dots\|^2 &= \|(\bar{y}_{i..} - \bar{y}\dots)\|^2 + \|(\bar{y}_{.j} - \bar{y}\dots)\|^2 + \\ &+ \|[(\bar{y}_{ij} - \bar{y}_{i..}) - (\bar{y}_{.j} - \bar{y}\dots)]\|^2 + \|(y_{ijk} - \bar{y}_{ij})\|^2 \end{aligned} \quad (2.13)$$

A idéia para demonstrar (2.13) é mostrar que cada um dos quatro fatores do segundo membro de (2.12) estão em subespaços mutuamente ortogonais. Será usada a notação $V \perp W$ para designar que V e W são subespaços mutuamente ortogonais e V^\perp para denotar o complementar ortogonal de V , que é o subespaço formado pelos vetores ortogonais a todos os vetores do subespaço V .

Por inspeção direta é verificado que

$$V_T^\perp = \{(x_{111}, x_{112}, \dots, x_{11K}, x_{121}, x_{122}, \dots, x_{12K}, \dots, x_{IJ1}, x_{IJ2}, \dots, x_{IJK})'\},$$

com $\sum_{j=1}^K x_{ij} = 0$.

Todos os quatro vetores da decomposição (2.12) têm média zero e portanto são perpendiculares a V_0 . Tem-se então

$$\bar{y}_{i..} - \bar{y}\dots \in W_F \quad \bar{y}_{.j} - \bar{y}\dots \in W_G \quad y_{ijk} - \bar{y}_{ij} \in V_T^\perp.$$

Proposição 9. *Os subespaços W_F e W_G são mutuamente ortogonais, isto é, $W_F \perp W_G$.*

Demonstração. Fazendo o produto interno entre dois vetores com $\mathbf{v} \in W_F$ e $\mathbf{u} \in W_G$ vem

$$\mathbf{v} = (\underbrace{a, a, \dots, a}_{JK}, \underbrace{b, b, \dots, b}_{JK}, \dots, \underbrace{c, c, \dots, c}_{JK}, \underbrace{d, d, \dots, d}_{JK})'$$

e

$$\mathbf{u} = \left(\underbrace{\alpha, \alpha, \dots, \alpha}_K, \underbrace{\beta, \beta, \dots, \beta}_K, \dots, \underbrace{\delta, \delta, \dots, \delta}_K, \underbrace{\lambda, \lambda, \dots, \lambda}_K, \right. \\ \left. \underbrace{\alpha, \alpha, \dots, \alpha}_K, \underbrace{\beta, \beta, \dots, \beta}_K, \dots, \underbrace{\delta, \delta, \dots, \delta}_K, \underbrace{\lambda, \lambda, \dots, \lambda}_K, \dots \right. \\ \left. \dots, \underbrace{\alpha, \alpha, \dots, \alpha}_K, \underbrace{\beta, \beta, \dots, \beta}_K, \dots, \underbrace{\delta, \delta, \dots, \delta}_K, \underbrace{\lambda, \lambda, \dots, \lambda}_K \right)',$$

com $a + b + \dots + c + d = 0$ e $\alpha + \beta + \dots + \delta + \lambda = 0$.

Portanto, efetuando o produto interno vem

$$\begin{aligned} \mathbf{v} \cdot \mathbf{u} &= Ka\alpha + Ka\beta + \dots + Kad + Ka\lambda + Kb\alpha + Kb\beta + \dots + Kbd + \\ &+ Kbl + \dots + Kc\alpha + Kc\beta + \dots + Kcd + Kc\lambda + Kd\alpha + Kd\beta + \\ &+ \dots + Kd\delta + Kd\lambda \\ &= Ka(\alpha + \beta + \dots + \delta + \lambda) + Kb(\alpha + \beta + \dots + \delta + \lambda) + \\ &+ \dots + Kc(\alpha + \beta + \dots + \delta + \lambda) + Kd(\alpha + \beta + \dots + \delta + \lambda) \\ &= 0. \end{aligned}$$

□

Resta portanto analisar em qual subespaço se encontra o vetor $(\bar{y}_{ij} - \bar{y}_{i..}) - (\bar{y}_{.j} - \bar{y}_{...})$. Observa-se que $(\bar{y}_{ij} - \bar{y}_{i..}) \in V_F^\perp \cap V_T$ e também $(\bar{y}_{.j} - \bar{y}_{...}) \in W_G$. Mas como $W_G \subset V_F^\perp \cap V_T$ tem-se $(\bar{y}_{ij} - \bar{y}_{i..}) \in V_F^\perp \cap V_T$.

Como se pode expressar $(\bar{y}_{ij} - \bar{y}_{i..}) - (\bar{y}_{.j} - \bar{y}_{...})$ como $(\bar{y}_{ij} - \bar{y}_{.j}) - (\bar{y}_{i..} - \bar{y}_{...})$ temos que $(\bar{y}_{ij} - \bar{y}_{.j}) \in V_G^\perp \cap V_T$ e $(\bar{y}_{i..} - \bar{y}_{...}) \in W_F$. Mas $W_F \subset V_G^\perp \cap V_T$. Logo, $(\bar{y}_{ij} - \bar{y}_{.j}) \in V_G^\perp \cap V_T$. Observa-se também que $V_F^\perp \cap V_T \cap V_G^\perp \cap V_T = (V_F + V_G)^\perp \cap V_T$.

Fica então demonstrado que a soma (2.12) está associada à soma direta de subespaços mutuamente ortogonais, ou seja,

$$W_F \oplus W_G \oplus (V_F \oplus V_G)^\perp \cap V_T \oplus V_T^\perp.$$

Abaixo, as Figuras 10 e 11 representam a decomposição do vetor de dados em termos do vetor de erros e do subespaço dos tratamentos.

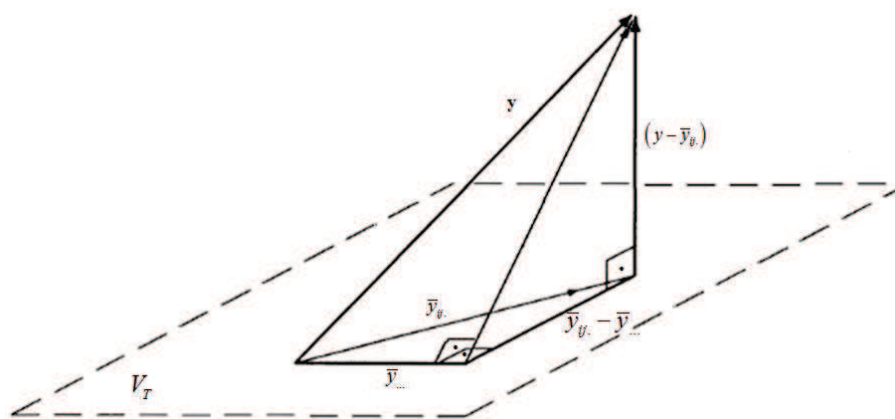


Figura 10 Projeção de y no subespaço de tratamentos

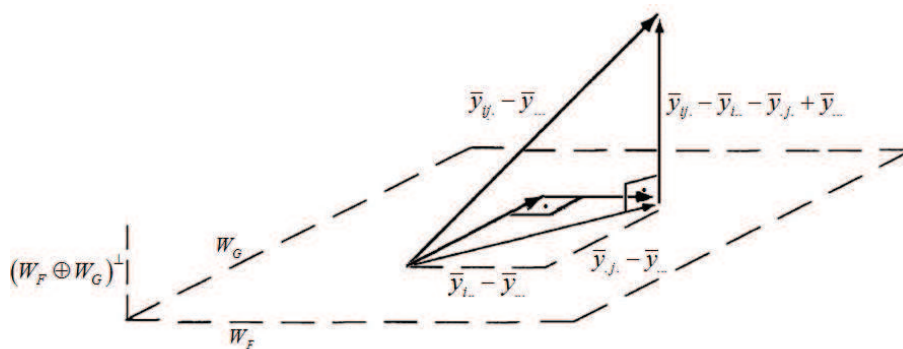


Figura 11 Decomposição de y em componentes ortogonais

2.4 A identidade de David-Hartley-Pearson

No artigo clássico *The distribution of the ratio, in a single normal sample, of range to standard deviation* (PEARSON et al., 1954), é apresentada a seguinte fórmula em soma de quadrados

$$(n-1)s^2 = \frac{1}{2}(y_j - y_k)^2 + \sum_{i \neq j, k}^n (y_i - \bar{y}')^2 + \frac{2(n-2)}{n}(\bar{y}' - \bar{y}'')^2, \quad (2.14)$$

em que $\bar{y}' = \frac{\sum_{i \neq j, k}^n y_i}{n-2}$ e $\bar{y}'' = \frac{y_j + y_k}{2}$.

Sem perda de generalidade, podemos supor $j = 1$ e $k = 2$. Desta forma, a equação (2.14) fica

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \frac{1}{2}(y_1 - y_2)^2 + \sum_{i \neq j, k}^n (y_i - \bar{y}')^2 + \frac{2(n-2)}{n}(\bar{y}' - \bar{y}'')^2. \quad (2.15)$$

Uma elaborada demonstração geométrica da equação (2.14) é apresentada em Ferreira et al.(2012). Observe que uma demonstração algébrica é bastante trabalhosa e também apresentada no apêndice do referido artigo. Será apresentada então uma demonstração essencialmente visual desta fórmula, ilustrada na Figura 12.

Sendo

$$\begin{aligned} (y_1 - \bar{y}'')^2 &= \left(y_1 - \frac{(y_1 + y_2)}{2} \right)^2 \\ &= \left(\frac{y_1 - y_2}{2} \right)^2 \end{aligned}$$

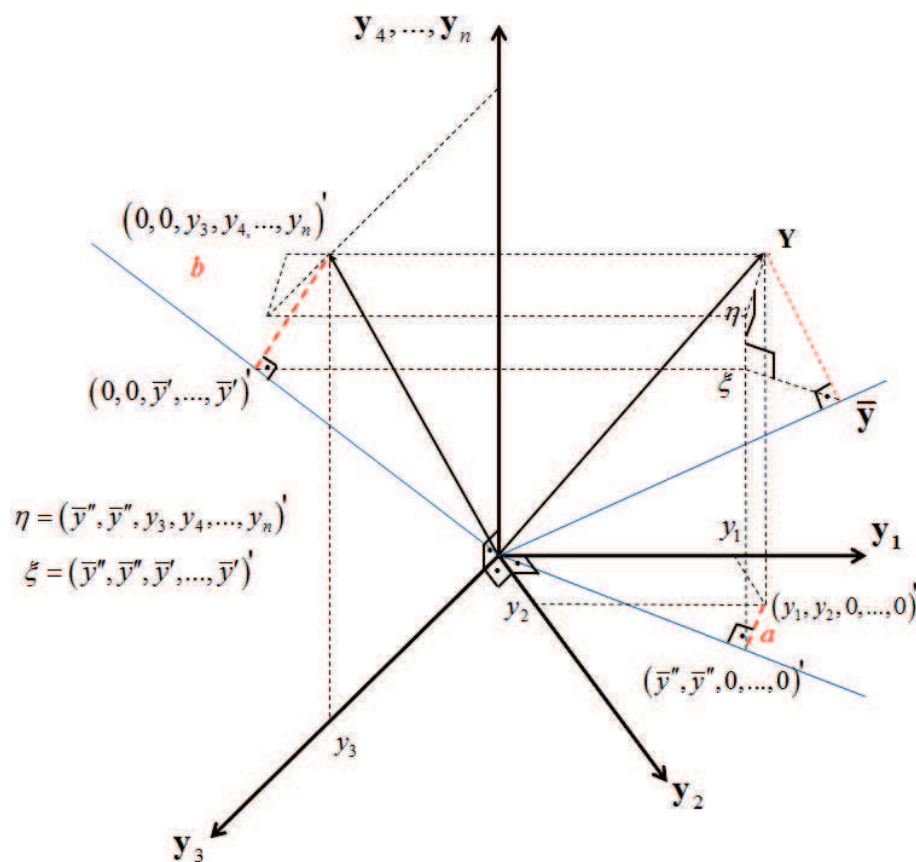


Figura 12 Vetores da decomposição da identidade de Pearson

e

$$\begin{aligned} (y_2 - \bar{y}'')^2 &= \left(y_2 - \frac{(y_1 + y_2)}{2} \right)^2 \\ &= \left(\frac{y_2 - y_1}{2} \right)^2, \end{aligned}$$

tem-se que

$$(y_1 - \bar{y}'')^2 + (y_2 - \bar{y}'')^2 = \left(\frac{y_1 - y_2}{2}\right)^2 + \left(\frac{y_2 - y_1}{2}\right)^2 = \frac{(y_1 - y_2)^2}{2}.$$

Substituindo a equação anterior em (2.15) tem-se

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^2 (y_i - \bar{y}'')^2 + \sum_{i \neq j, k}^n (y_i - \bar{y}')^2 + \frac{2(n-2)}{n} (\bar{y}' - \bar{y}'')^2.$$

Definindo-se então os vetores $\mathbf{y} = (y_1, y_2, \dots, y_n)$, $\bar{\mathbf{y}} = (\bar{y}, \bar{y}, \dots, \bar{y})$, $\boldsymbol{\eta} = (\bar{y}'', \bar{y}'', y_3, \dots, y_n)$ e $\boldsymbol{\xi} = (\bar{y}'', \bar{y}'', \bar{y}', \dots, \bar{y}')$, segue então a identidade

$$\mathbf{y} - \bar{\mathbf{y}} = (\mathbf{y} - \boldsymbol{\eta}) + (\boldsymbol{\eta} - \boldsymbol{\xi}) + (\boldsymbol{\xi} - \bar{\mathbf{y}}).$$

Provando-se que os três vetores do segundo membro da equação anterior são mutuamente ortogonais, tem-se que é válida a soma dos quadrados dos módulos, isto é,

$$\|\mathbf{y} - \bar{\mathbf{y}}\|^2 = \|\mathbf{y} - \boldsymbol{\eta}\|^2 + \|\boldsymbol{\eta} - \boldsymbol{\xi}\|^2 + \|\boldsymbol{\xi} - \bar{\mathbf{y}}\|^2. \quad (2.16)$$

Observe que

$$\begin{aligned} \|\boldsymbol{\xi} - \bar{\mathbf{y}}\|^2 &= (\bar{y}'' - \bar{y})^2 + (\bar{y}'' - \bar{y})^2 + \underbrace{(\bar{y}' - \bar{y})^2 + \dots + (\bar{y}' - \bar{y})^2}_{n-2} \\ &= 2(\bar{y}'' - \bar{y})^2 + (n-2)(\bar{y}' - \bar{y})^2. \end{aligned}$$

Como

$$\bar{y} = \frac{2\left(\frac{y_1 + y_2}{2}\right) + (n-2)\left(\frac{\sum_{i=3}^n y_i}{n-2}\right)}{n}$$

$$= \frac{2\bar{y}'' + (n-2)\bar{y}'}{n},$$

então

$$\begin{aligned} 2(\bar{y}'' - \bar{y})^2 + (n-2)(\bar{y}' - \bar{y})^2 &= 2\left[\bar{y}'' - \frac{2\bar{y}'' + (n-2)\bar{y}'}{n}\right]^2 \\ &\quad + (n-2)\left[\bar{y}' - \frac{2\bar{y}'' + (n-2)\bar{y}'}{n}\right]^2 \\ &= \frac{2}{n^2}(n\bar{y}'' - 2\bar{y}'' - (n-2)\bar{y}')^2 \\ &\quad + \frac{(n-2)}{n^2}(n\bar{y}' - 2\bar{y}'' - (n-2)\bar{y}')^2 \\ &= \frac{2}{n^2}(n\bar{y}'' - 2\bar{y}'' - n\bar{y}' + 2\bar{y}')^2 \\ &\quad + \frac{(n-2)}{n^2}(-2\bar{y}'' + 2\bar{y}')^2 \\ &= \frac{2}{n^2}((n-2)(\bar{y}'' - \bar{y}'))^2 \\ &\quad + \frac{n-2}{n^2}(2(\bar{y}' - \bar{y}''))^2 \\ &= \left[\frac{2}{n^2}(n-2)^2 + 4\frac{n-2}{n^2}\right](\bar{y}' - \bar{y}'')^2 \\ &= \frac{2(n-2)}{n}(\bar{y}' - \bar{y}'')^2. \end{aligned}$$

Segue então que a Equação (2.16) é o equivalente vetorial da identidade de David-Hartley-Pearson (Equação (2.15)).

3 ABORDAGEM GEOMÉTRICA À TEORIA DOS MODELOS DE GAUSS-MARKOV

A equação matricial obtida pelo método dos quadrados mínimos (Equação 2.4) é um caso particular de um modelo de Gauss-Markov, definido em geral por $\mathbf{Y}_{n \times 1} = \mathbf{X}_{n \times p} \boldsymbol{\beta}_{p \times 1} + \boldsymbol{\varepsilon}_{n \times 1}$ com $E(\mathbf{Y}_{n \times 1}) = \mathbf{X}_{n \times p} \boldsymbol{\beta}_{p \times 1}$ e matriz de variâncias e covariâncias dada por $D(\mathbf{Y}_{n \times 1}) = \boldsymbol{\Sigma}$.

3.1 O modelo

Para um vetor aleatório \mathbf{Z} , com matriz de variâncias e covariâncias $D(\mathbf{Z}) = \boldsymbol{\Sigma}$, a variância total de \mathbf{Z} é definida como o traço de $\boldsymbol{\Sigma}$, $tr(\boldsymbol{\Sigma})$. Em geral os livros textos (RAO,2002;RENCHER et al.,2008) apresentam o importante resultado que se segue, no caso particular de $D(\mathbf{Y}_{n \times 1}) = \sigma^2 \mathbf{I}$:

Teorema 1. (Gauss-Markov) *Se $E(\mathbf{Y}_{n \times 1}) = \mathbf{X}_{n \times p} \boldsymbol{\beta}_{p \times 1}$ e $D(\mathbf{Y}_{n \times 1}) = \sigma^2 \mathbf{I}$, então o estimador de quadrados mínimos $\hat{\boldsymbol{\beta}}$ é o estimador de menor variância dentre todos os estimadores lineares não viesados.*

Demonstração. Seja um estimador linear $\mathbf{A}(\mathbf{y})$ de $\boldsymbol{\beta}$. Vamos buscar a matriz \mathbf{A} para a qual $\mathbf{A}(\mathbf{y})$ seja o estimador não viesado de menor variância de $\boldsymbol{\beta}$. Para que $\mathbf{A}(\mathbf{y})$ seja não viesado, temos que $E[\mathbf{A}(\mathbf{y})] = \boldsymbol{\beta}$. Assumindo que $E[\mathbf{y}] = \mathbf{X}\boldsymbol{\beta}$, vem

$$\begin{aligned} E[\mathbf{A}(\mathbf{y})] &= \mathbf{A}E[\mathbf{y}] \\ &= \mathbf{A}\mathbf{X}\boldsymbol{\beta} \\ &= \boldsymbol{\beta}, \end{aligned}$$

o que implica na condição de não viesamento $\mathbf{AX} = \mathbf{I}$, desde que $\mathbf{AX}\beta = \beta$, $\forall \beta$. A matriz de variâncias e covariâncias para o estimador $\mathbf{A}(\mathbf{y})$ é dada por

$$\begin{aligned} \text{cov}(\mathbf{A}(\mathbf{y})) &= \mathbf{A}(\sigma^2\mathbf{I})\mathbf{A}' \\ &= \sigma^2\mathbf{AA}' \end{aligned}$$

Temos que as variâncias dos $\hat{\beta}_p$ estão na diagonal de $\sigma^2\mathbf{AA}'$ e se quer escolher \mathbf{A} (sujeito a $\mathbf{AX} = \mathbf{I}$) tal que os elementos da diagonal de \mathbf{AA}' sejam minimizados.

De forma a relacionar \mathbf{Ay} com $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$, deve-se adicionar e subtrair $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ para obter

$$\mathbf{AA}' = \left[\mathbf{A} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \right] \left[\mathbf{A} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \right]'$$

Sendo

$$\begin{aligned} \mathbf{AX} &= \mathbf{I} \\ (\mathbf{AX})' &= \mathbf{I}' \\ \mathbf{X}'\mathbf{A}' &= \mathbf{I} \end{aligned}$$

e $\left[(\mathbf{X}'\mathbf{X})^{-1} \right]' = (\mathbf{X}'\mathbf{X})^{-1}$ (pois $(\mathbf{X}'\mathbf{X})^{-1}$ é simétrica), tem-se

$$\begin{aligned} \mathbf{AA}' &= \left[\mathbf{A} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \right] \\ &\quad \left[\mathbf{A} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \right]' \\ &= \left[\left(\mathbf{A} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \right) + \left((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \right) \right] \end{aligned}$$

$$\begin{aligned}
& \left[\left(\mathbf{A} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \right)' + \left((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \right)' \right] \\
= & \left[\mathbf{A} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \right] \left[\mathbf{A} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \right]' \\
& + \left[\mathbf{A} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \right] \left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \right]' \\
& + \left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \right] \left[\mathbf{A} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \right]' \\
& + \left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \right] \left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \right]' \\
= & \left[\mathbf{A} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \right] \left[\mathbf{A} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \right]' \\
& + \mathbf{A} \left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \right]' - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \right]' \\
& + \left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \right] \left[\mathbf{A}' - \mathbf{X} \left[(\mathbf{X}'\mathbf{X})^{-1} \right]' \right] \\
& + \left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \right] \left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \right]' \\
= & \left[\mathbf{A} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \right] \left[\mathbf{A} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \right]' \\
& + \mathbf{A}\mathbf{X} \left[(\mathbf{X}'\mathbf{X})^{-1} \right]' - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X} \left[(\mathbf{X}'\mathbf{X})^{-1} \right]' \\
& + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{A}' - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X} \left[(\mathbf{X}'\mathbf{X})^{-1} \right]' \\
& + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X} \left[(\mathbf{X}'\mathbf{X})^{-1} \right]' \\
= & \left[\mathbf{A} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \right] \left[\mathbf{A} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \right]' + \mathbf{I}(\mathbf{X}'\mathbf{X})^{-1} \\
& - \mathbf{I}(\mathbf{X}'\mathbf{X})^{-1} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{I} \\
= & \left[\mathbf{A} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \right] \left[\mathbf{A} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \right]' + (\mathbf{X}'\mathbf{X})^{-1}.
\end{aligned}$$

Logo, $\mathbf{A}\mathbf{A}' = \left[\mathbf{A} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \right] \left[\mathbf{A} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \right]' + (\mathbf{X}'\mathbf{X})^{-1}$.

A matriz $\left[\mathbf{A} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \right] \left[\mathbf{A} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \right]'$ é positiva semidefinida (Teorema 2.6d (RENCHEER et al.,2008,p.27)), o que implica que os elementos de sua diagonal são maiores ou iguais a 0 (Teorema 2.6a (ii) (RENCHEER et al.,2008,p.26)). Pode-se fazer estes elementos serem 0 caso $\mathbf{A} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ (observe que este valor de \mathbf{A} também satisfaz a condição de não viesamento $\mathbf{A}\mathbf{X} = \mathbf{I}$).

O estimador de mínima variância de β resultante é

$$\mathbf{A}y = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'y,$$

o qual é igual ao estimador de mínimos quadrados $\hat{\beta}$. □

Para se estabelecer o teorema de Gauss-Markov numa situação mais geral em que $D(\mathbf{Y}) = \Sigma$, com Σ sendo uma matriz não singular, um contexto geométrico adequado é o de se utilizar um outro produto interno, distinto do produto usual, e portanto um novo conceito de ortogonalidade.

A importância da ortogonalidade na Estatística ocorre pois este conceito está intrinsecamente relacionado com o de correlação. Seja \mathbf{Y} um vetor aleatório n dimensional e \mathbf{x} e \mathbf{z} vetores fixos. As projeções ortogonais de \mathbf{Y} em \mathbf{x} e \mathbf{z} são dadas essencialmente pelos produtos internos $\langle \mathbf{x}, \mathbf{Y} \rangle = \mathbf{x}'\mathbf{Y}$ e $\langle \mathbf{z}, \mathbf{Y} \rangle = \mathbf{z}'\mathbf{Y}$. Uma questão natural é estudar a covariância entre estas duas variáveis aleatórias, representada na Figura 13.

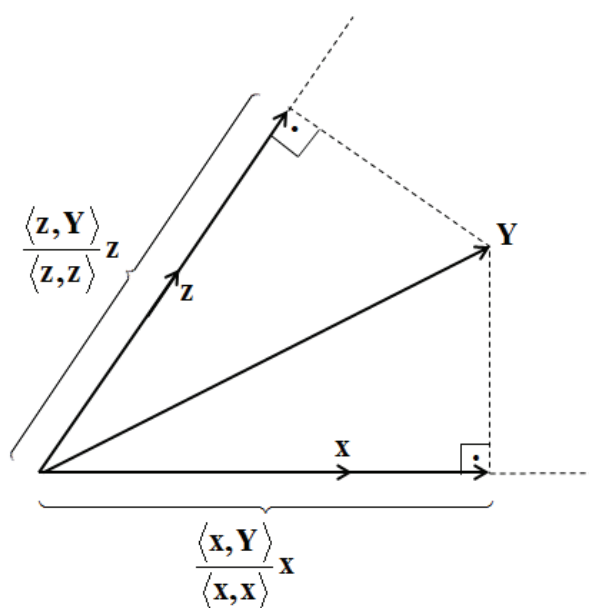


Figura 13 Representação do vetor aleatório Y e as correlações (projeções) de x e z

Proposição 10. A covariância entre os produtos internos $\langle x, Y \rangle$ e $\langle z, Y \rangle$ é dada por

$$\begin{aligned} \text{Cov} [\langle x, Y \rangle, \langle z, Y \rangle] &= \langle x, D(Y) z \rangle \\ &= \langle x, \Sigma z \rangle, \end{aligned}$$

em que Σ é a matriz de variâncias e covariâncias de Y .

Demonstração. Sejam os vetores x e z . Então,

$$\begin{aligned} \text{Cov} (L'Y, M'Y) &= E \{ [L'Y - E(L'Y)] [M'Y - E(M'Y)] \} \\ &= E \{ [L'Y - L'E(Y)] [M'Y - M'E(Y)] \} \end{aligned}$$

$$\begin{aligned}
&= E \{ \mathbf{L}' [\mathbf{Y} - E(\mathbf{Y})] \mathbf{M}' [\mathbf{Y} - E(\mathbf{Y})] \} \\
&= E \{ \mathbf{L}' [\mathbf{Y} - E(\mathbf{Y})] [\mathbf{Y} - E(\mathbf{Y})]' \mathbf{M} \} \\
&= \mathbf{L}' E \{ [\mathbf{Y} - E(\mathbf{Y})] [\mathbf{Y} - E(\mathbf{Y})]' \} \mathbf{M} \\
&= \mathbf{L}' \Sigma \mathbf{M}.
\end{aligned}$$

Logo,

$$Cov(\mathbf{L}'\mathbf{Y}, \mathbf{M}'\mathbf{Y}) = \mathbf{L}'\Sigma\mathbf{M}. \quad (3.1)$$

□

Como Σ é uma matriz positiva definida, a expressão (3.1) define um novo produto interno $\langle \mathbf{x}, \mathbf{z} \rangle = \langle \mathbf{x}, \Sigma \mathbf{z} \rangle = \mathbf{x}' \Sigma \mathbf{z}$. Desta forma, $Cov[\langle \mathbf{x}, \mathbf{Y} \rangle, \langle \mathbf{z}, \mathbf{Y} \rangle] = \langle \mathbf{x}, \mathbf{z} \rangle$. Entretanto esta fórmula não é totalmente adequada por utilizar simultaneamente dois produtos internos, uma vez que a projeção do vetor aleatório \mathbf{Y} na direção dos vetores \mathbf{x} e \mathbf{z} é dada pelo produto interno usual. Para se obter uma fórmula para a correlação entre duas projeções em termos de um único produto interno, este produto deve ser da forma $\langle \langle \mathbf{x}, \mathbf{z} \rangle \rangle = \langle \mathbf{x}, \Sigma^{-1} \mathbf{z} \rangle$. Este produto interno define a chamada distância de Mahalanobis (RENCHER et al., 2008). Assim,

$$\begin{aligned}
Cov[\langle \langle \mathbf{x}, \mathbf{Y} \rangle \rangle, \langle \langle \mathbf{z}, \mathbf{Y} \rangle \rangle] &= Cov[\langle \mathbf{x}, \Sigma^{-1} \mathbf{Y} \rangle, \langle \mathbf{z}, \Sigma^{-1} \mathbf{Y} \rangle] \\
&= \langle \mathbf{x}, D(\Sigma^{-1} \mathbf{Y}) \mathbf{z} \rangle \\
&= \langle \mathbf{x}, \Sigma^{-1} D(\mathbf{Y}) \Sigma^{-1} \mathbf{z} \rangle \\
&= \langle \mathbf{x}, \Sigma^{-1} \Sigma \Sigma^{-1} \mathbf{z} \rangle \\
&= \langle \mathbf{x}, \Sigma^{-1} \mathbf{z} \rangle \\
&= \langle \langle \mathbf{x}, \mathbf{z} \rangle \rangle.
\end{aligned}$$

Esse novo produto interno é interessante pois pode-se determinar direções para as quais as projeções ortogonais (ortogonal em relação a este produto interno) são não correlacionadas (Figura14).

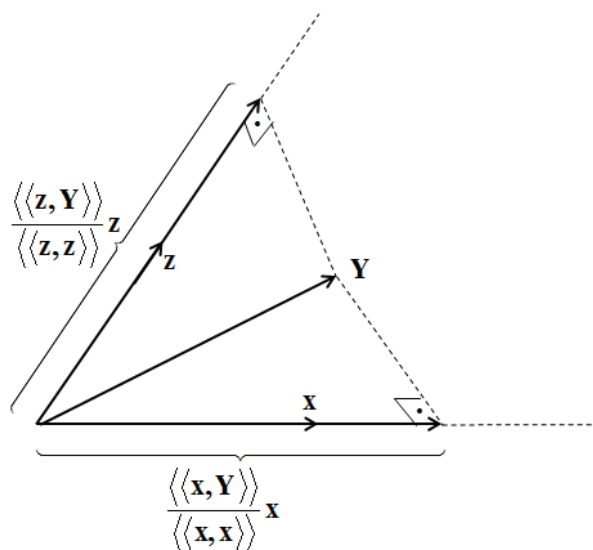


Figura 14 Representação das projeções do vetor \mathbf{Y} nas direções de \mathbf{x} e \mathbf{z} com relação ao novo produto interno

Deseja-se agora obter projetores ortogonais em relação a este novo produto interno, isto é, matrizes \mathbf{A} tais que $\mathbf{A}^2 = \mathbf{A}$ e $\langle\langle \mathbf{Ax}, \mathbf{z} \rangle\rangle = \langle\langle \mathbf{x}, \mathbf{Az} \rangle\rangle$. Considerando o modelo de Gauss-Markov $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ com $D(\mathbf{y}) = \boldsymbol{\Sigma}$, o estimador de quadrados mínimos em relação ao novo produto interno é dado pela projeção ortogonal de \mathbf{Y} na $Im(\mathbf{X})$.

Proposição 11.

$$P_{Im(\mathbf{X})}\mathbf{Y} = \mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{Y}.$$

Demonstração. Idempotência:

$$\mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1} = \mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{I}\mathbf{X}'\boldsymbol{\Sigma}^{-1}$$

Ortogonalidade:

$$\begin{aligned} \langle \langle \mathbf{y} - \mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{y}, \mathbf{X}\boldsymbol{\beta} \rangle \rangle &= \langle \mathbf{y} - \mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{y}, \\ &\quad \boldsymbol{\Sigma}^{-1}\mathbf{X}\boldsymbol{\beta} \rangle \\ &= \langle \mathbf{y}, \boldsymbol{\Sigma}^{-1}\mathbf{X}\boldsymbol{\beta} \rangle \\ &\quad - \langle \mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{y}, \\ &\quad \boldsymbol{\Sigma}^{-1}\mathbf{X}\boldsymbol{\beta} \rangle \\ &= \langle \mathbf{y}, \boldsymbol{\Sigma}^{-1}\mathbf{X}\boldsymbol{\beta} \rangle \\ &\quad - \langle \boldsymbol{\Sigma}^{-1}\mathbf{X}\boldsymbol{\beta}, \\ &\quad \mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{y} \rangle \\ &= \langle \mathbf{y}, \boldsymbol{\Sigma}^{-1}\mathbf{X}\boldsymbol{\beta} \rangle \\ &\quad - \langle \boldsymbol{\Sigma}^{-1}\mathbf{y}, \\ &\quad \mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\boldsymbol{\beta} \rangle \\ &= \langle \mathbf{y}, \boldsymbol{\Sigma}^{-1}\mathbf{X}\boldsymbol{\beta} \rangle - \langle \boldsymbol{\Sigma}^{-1}\mathbf{y}, \mathbf{X}\boldsymbol{\beta} \rangle \\ &= \langle \mathbf{y}, \boldsymbol{\Sigma}^{-1}\mathbf{X}\boldsymbol{\beta} \rangle - \langle \mathbf{y}, \boldsymbol{\Sigma}^{-1}\mathbf{X}\boldsymbol{\beta} \rangle \\ &= 0, \forall \boldsymbol{\beta}. \end{aligned}$$

□

Vale destacar que, um operador linear ser simétrico em relação ao produto interno $\langle \langle \mathbf{x}, \mathbf{z} \rangle \rangle = \langle \mathbf{x}, \boldsymbol{\Sigma}^{-1}\mathbf{z} \rangle$ implica em

$$\begin{aligned}
\langle\langle \mathbf{Ax}, \mathbf{z} \rangle\rangle &= \langle\langle \mathbf{x}, \mathbf{Az} \rangle\rangle \\
\langle \mathbf{Ax}, \boldsymbol{\Sigma}^{-1} \mathbf{z} \rangle &= \langle \mathbf{x}, \boldsymbol{\Sigma}^{-1} \mathbf{Az} \rangle \\
\langle \mathbf{x}, \mathbf{A}' \boldsymbol{\Sigma}^{-1} \mathbf{z} \rangle &= \langle \mathbf{x}, \boldsymbol{\Sigma}^{-1} \mathbf{Az} \rangle \\
\mathbf{A}' \boldsymbol{\Sigma}^{-1} &= \boldsymbol{\Sigma}^{-1} \mathbf{A} \\
\mathbf{A} &= \boldsymbol{\Sigma} \mathbf{A}' \boldsymbol{\Sigma}^{-1}.
\end{aligned}$$

Proposição 12. O projetor $(\mathbf{X}' \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}' \boldsymbol{\Sigma}^{-1}$ é simétrico em relação ao produto interno $\langle\langle \cdot, \cdot \rangle\rangle$.

Demonstração.

$$\begin{aligned}
\boldsymbol{\Sigma} \left((\mathbf{X}' \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}' \boldsymbol{\Sigma}^{-1} \right)' \boldsymbol{\Sigma}^{-1} &= \boldsymbol{\Sigma} \boldsymbol{\Sigma}^{-1} \mathbf{X} (\mathbf{X}' \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1} \boldsymbol{\Sigma}^{-1} \\
&= \mathbf{X} (\mathbf{X}' \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1} \boldsymbol{\Sigma}^{-1}
\end{aligned}$$

□

Antes de se enunciar o teorema de Gauss-Markov em um contexto de maior generalidade, seja a seguinte proposição.

Proposição 13. Se $\boldsymbol{\Sigma}$ é a matriz de variâncias e covariâncias de um vetor aleatório \mathbf{Y} , então $\boldsymbol{\Sigma}$ é positiva semidefinida e $\mathbf{A} \boldsymbol{\Sigma} \mathbf{A}'$ também é positiva semidefinida para qualquer matriz \mathbf{A} .

Demonstração. A matriz $[(\mathbf{Y} - \mathbf{E}[\mathbf{Y}])_{n \times 1} (\mathbf{Y} - \mathbf{E}[\mathbf{Y}])'_{1 \times n}]_{n \times n}$ é positiva semidefinida pois

$$\mathbf{a}'_{1 \times n} [(\mathbf{Y} - \mathbf{E}[\mathbf{Y}])_{n \times 1} (\mathbf{Y} - \mathbf{E}[\mathbf{Y}])'_{1 \times n}]_{n \times n} \mathbf{a}_{n \times 1} = ((\mathbf{Y} - \mathbf{E}[\mathbf{Y}])'_{1 \times n} \mathbf{a}_{n \times 1})'$$

$$\begin{aligned} & ((\mathbf{Y} - E[\mathbf{Y}])'_{1 \times n} \mathbf{a}_{n \times 1}) \\ & \geq 0. \end{aligned}$$

Aplicando a esperança em ambos os membros da inequação anterior tem-se

$$\begin{aligned} 0 & \leq E \left[\mathbf{a}'_{1 \times n} [(\mathbf{Y} - E[\mathbf{Y}])_{n \times 1} (\mathbf{Y} - E[\mathbf{Y}])'_{1 \times n}]_{n \times n} \mathbf{a}_{n \times 1} \right] \\ & = \mathbf{a}'_{1 \times n} E [(\mathbf{Y} - E[\mathbf{Y}])_{n \times 1} (\mathbf{Y} - E[\mathbf{Y}])'_{1 \times n}] \mathbf{a}_{n \times 1} \\ & = \mathbf{a}'_{1 \times n} \Sigma_{n \times n} \mathbf{a}_{n \times 1}. \end{aligned}$$

Logo, $\mathbf{a}'_{1 \times n} \Sigma_{n \times n} \mathbf{a}_{n \times 1} \geq 0$, o que implica que Σ é positiva semidefinida.

Tem-se então

$$\begin{aligned} \mathbf{a}'_{1 \times n} (\mathbf{A}_{n \times n} \Sigma_{n \times n} \mathbf{A}'_{n \times n}) \mathbf{a}_{n \times 1} & = (\mathbf{a}'_{1 \times n} \mathbf{A}_{n \times n}) \Sigma_{n \times n} (\mathbf{A}'_{n \times n} \mathbf{a}_{n \times 1}) \\ & = (\mathbf{a}'_{1 \times n} \mathbf{A}_{n \times n}) \Sigma_{n \times n} (\mathbf{a}'_{1 \times n} \mathbf{A}_{n \times n})' \\ & \geq 0, \end{aligned}$$

pois Σ é positiva semidefinida. Logo, $\mathbf{A} \Sigma \mathbf{A}'$ é positiva semidefinida.

□

O caso geral, denominado o teorema de Gauss-Markov para quadrados mínimos generalizados, fica da seguinte forma (RENCHEER et al., 2008):

Teorema 2. (Gauss-Markov).

Se $E(\mathbf{Y}_{n \times 1}) = \mathbf{X}_{n \times p} \boldsymbol{\beta}_{p \times 1}$ e $D(\mathbf{Y}_{n \times 1}) = \sigma^2 \Sigma$, então o estimador de Gauss-Markov dado por $\tilde{\boldsymbol{\beta}} = (\mathbf{X}' \Sigma^{-1} \mathbf{X})^{-1} \mathbf{X}' \Sigma^{-1} \mathbf{y}$ é o estimador de menor

variância total dentre todos os estimadores lineares não viesados.

Demonstração. De forma análoga ao Teorema 1

Seja $\mathbf{A} : \mathbb{R}^n \rightarrow \mathbb{R}^p$ tal que $\beta^* = \mathbf{A}\mathbf{y}$ seja um estimador não viesado, logo

$$\begin{aligned}\beta &= E[\beta^*] \\ &= E[\mathbf{A}\mathbf{y}] \\ &= \mathbf{A}E[\mathbf{y}] \\ &= \mathbf{A}\mathbf{X}\beta.\end{aligned}$$

Sendo β desconhecido, a equação anterior tem que ser válida para todo β , ou seja, $\mathbf{A}\mathbf{X} = \mathbf{I}$. A variância total de β^* é dada por

$$\begin{aligned}tr(D(\beta^*)) &= tr(D(\mathbf{A}\mathbf{y})) \\ &= tr(\mathbf{A}\Sigma\mathbf{A}').\end{aligned}$$

Como

$$\begin{aligned}D(\tilde{\beta}) &= (\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}'\Sigma^{-1}\Sigma\Sigma^{-1}\mathbf{X}(\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1} \\ &= (\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1},\end{aligned}$$

e sendo a variância total igual ao traço da matriz de variâncias e covariâncias, é necessário mostrar que

$$tr(\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1} \leq tr(\mathbf{A}\Sigma\mathbf{A}'), \forall \mathbf{A} \text{ tal que } \mathbf{A}\mathbf{X}=\mathbf{I}.$$

Assim, desenvolvendo o produto $\mathbf{A}\Sigma\mathbf{A}'$ tem-se

$$\begin{aligned}
& +(\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{I}\mathbf{A}' \\
& -(\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{I}(\Sigma^{-1})'\mathbf{X}\left[(\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\right]' \\
& +(\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{I}(\Sigma^{-1})'\mathbf{X}\left[(\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\right]' \\
= & \left[\mathbf{A} - (\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}'\Sigma^{-1}\right] \left[\Sigma\left(\mathbf{A} - (\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}'\Sigma^{-1}\right)'\right] \\
& +\mathbf{A}\Sigma\Sigma^{-1}\mathbf{X}(\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1} - \\
& -(\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{I}\Sigma^{-1}\mathbf{X}(\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1} \\
& +(\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{A}' \\
= & \left[\mathbf{A} - (\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}'\Sigma^{-1}\right] \left[\Sigma\left(\mathbf{A} - (\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}'\Sigma^{-1}\right)'\right] \\
& +\mathbf{A}\mathbf{I}\mathbf{X}(\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1} - (\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}'\Sigma^{-1}\mathbf{X}(\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1} \\
& +(\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{A}\mathbf{X} \\
= & \left[\mathbf{A} - (\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}'\Sigma^{-1}\right] \left[\Sigma\left(\mathbf{A} - (\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}'\Sigma^{-1}\right)'\right] \\
& +(\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1} - (\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1} + (\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1} \\
= & \left[\mathbf{A} - (\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}'\Sigma^{-1}\right] \left[\Sigma\left(\mathbf{A} - (\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}'\Sigma^{-1}\right)'\right] \\
& +(\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1} \\
= & \left[\mathbf{A} - (\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}'\Sigma^{-1}\right] \Sigma\left[\mathbf{A} - (\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}'\Sigma^{-1}\right]' \\
& +(\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}
\end{aligned}$$

Como Σ é positiva definida, temos que

$$\left[\mathbf{A} - (\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}'\Sigma^{-1}\right] \Sigma\left[\mathbf{A} - (\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}'\Sigma^{-1}\right]'$$

também é positiva semidefinida e portanto possui traço maior ou igual a 0. Segue então que

$$tr(\mathbf{A}\Sigma\mathbf{A}') \geq tr\left((\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\right).$$

As equações normais para o caso do produto interno $\langle\langle \mathbf{x}, \mathbf{y} \rangle\rangle$ são

$$\boldsymbol{\beta}^* = (\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{y},$$

isto é,

$$\boxed{\mathbf{X}\boldsymbol{\beta}^* = \mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{y}.}$$

3.2 A geometria do teorema de Gauss-Markov

A demonstração do teorema de Gauss-Markov é baseada totalmente em fatos algébricos, e portanto não possui nenhum apelo intuitivo. Segue então uma interpretação essencialmente geométrica que pretende justificar intuitivamente ao leitor este importante resultado. Seja a Figura 15, representando as relações entre os subespaços de parâmetros e subespaço do modelo.

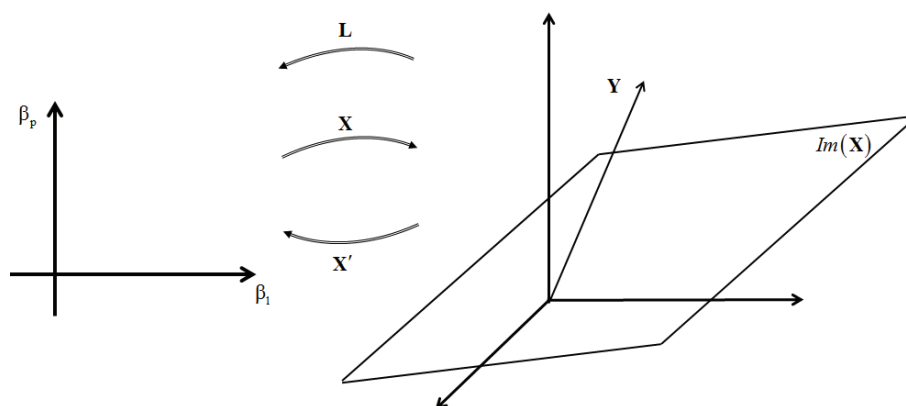


Figura 15 Relação entre os subespaços de parâmetros e do modelo

Como $P_{Im(\mathbf{X})} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$, aplicando-se a equação anterior no vetor

\mathbf{Y} , e a matriz \mathbf{X}' em ambos os membros da equação tem-se

$$\begin{aligned}\mathbf{X}'\mathbf{P}_{Im(\mathbf{X})}\mathbf{Y} &= \mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \\ &= \mathbf{X}'\mathbf{Y}.\end{aligned}\tag{3.2}$$

Conforme discutido anteriormente, $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$. Assim, relacionando-se este fato com a equação 3.2 tem-se que $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{P}_{Im(\mathbf{X})}(\mathbf{Y})$.

Seja agora $\hat{\boldsymbol{\beta}}^* = \mathbf{L}\mathbf{Y}$ um outro estimador linear não viesado. Como

$$\begin{aligned}\boldsymbol{\beta} &= E[\mathbf{L}\mathbf{Y}] \\ &= \mathbf{L}E[\mathbf{Y}] \\ &= \mathbf{L}\mathbf{X}\boldsymbol{\beta} \\ \mathbf{L}\mathbf{X} &= \mathbf{I}.\end{aligned}$$

Agora, como analisar a dispersão do estimador $\hat{\boldsymbol{\beta}}^*$? No caso em que $D(\mathbf{Y}) = \sigma^2\mathbf{I}$, a dispersão do vetor aleatório pode ser representada como uma nuvem esférica centrada em algum ponto $\mathbf{X}\boldsymbol{\beta} \in Im(\mathbf{X})$, conforme representado na Figura 16.

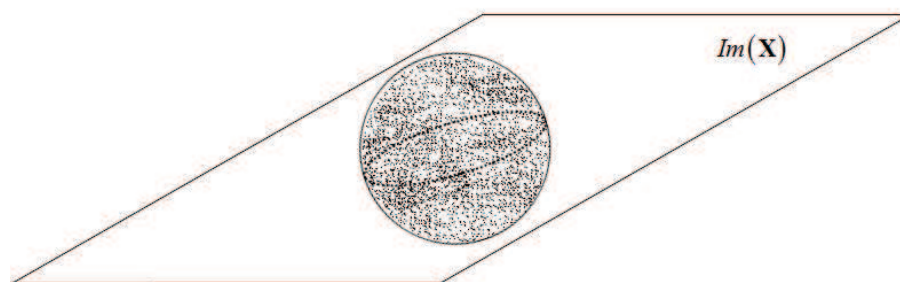


Figura 16 Esfera de variância

A projeção ortogonal desta nuvem fornece uma superfície esférica em

$Im(\mathbf{X})$. A imagem desta esfera por \mathbf{X}' dará origem, após a transformação $(\mathbf{X}'\mathbf{X})^{-1}$, à nuvem de dispersão do estimador $\hat{\beta}$.

Define-se então uma outra projeção em $Im(\mathbf{X})$ dada por $\tilde{\mathbf{P}}_{Im(\mathbf{X})}\mathbf{Y} = \mathbf{X}\mathbf{L}\mathbf{Y}$. A Figura 17 compara essa nova projeção com a projeção ortogonal.

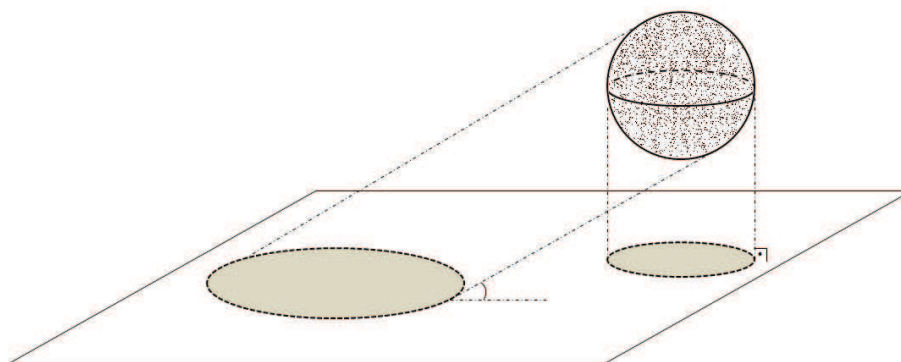


Figura 17 Comparação entre as projeções ortogonal e oblíqua

Claramente $\tilde{\mathbf{P}}$ é uma projeção pois

$$\begin{aligned}\tilde{\mathbf{P}}(\tilde{\mathbf{P}}\mathbf{Y}) &= \mathbf{X}\mathbf{L}(\mathbf{X}\mathbf{L}\mathbf{Y}) \\ &= \mathbf{X}(\mathbf{L}\mathbf{X})\mathbf{L}\mathbf{Y} \\ &= \mathbf{X}\mathbf{L}\mathbf{Y} \\ &= \tilde{\mathbf{P}}\mathbf{Y}.\end{aligned}$$

$\tilde{\mathbf{P}}$ é uma projeção oblíqua (o projetor ortogonal é $\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$). Note que

$$\mathbf{L}\tilde{\mathbf{P}}\mathbf{Y} = \mathbf{L}\mathbf{X}\mathbf{L}\mathbf{Y}$$

$$= \mathbf{LY}.$$

Para se estudar a dispersão de $\hat{\beta}^*$ a idéia é a mesma. Toma-se a esfera que representa a dispersão de \mathbf{Y} , a qual é projetada em $Im(\mathbf{X})$ pela projeção $\tilde{\mathbf{P}}$. Como $\tilde{\mathbf{P}}$ é oblíqua a imagem da esfera será uma região na forma elíptica, e portanto uma região maior que a obtida pela projeção ortogonal. Tal fato é razoável do ponto de vista intuitivo pois a projeção ortogonal minimiza distâncias.

A dispersão de $\hat{\beta}^*$ pode ser vista como a imagem por L desta região elíptica. Seja então a comparação da dispersão de $\hat{\beta}$ e de $\hat{\beta}^*$. Para tanto deve-se observar que

$$\begin{aligned} L(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}') &= L\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}', \end{aligned}$$

o que significa que a transformação L restrita a $Im(\mathbf{X})$ se comporta como a aplicação $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ que é a que define o estimador de quadrados mínimos $\hat{\beta}$. Portanto a nuvem de dispersão de $\hat{\beta}$ e de $\hat{\beta}^*$ são obtidas como imagem da mesma aplicação $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$.

Como a nuvem de dispersão de $\hat{\beta}$ é obtida como imagem de uma esfera que está “contida” na região elíptica, cuja imagem é a dispersão de $\hat{\beta}^*$, segue que a variância total de $\hat{\beta}^*$ é maior do que a de $\hat{\beta}$. A Figura 18 representa estes conceitos.

Uma demonstração geométrica do teorema para o caso de quadrados generalizados segue de forma análoga à anterior substituindo-se a esfera como nuvem de dispersão do vetor \mathbf{Y} por uma região elíptica.

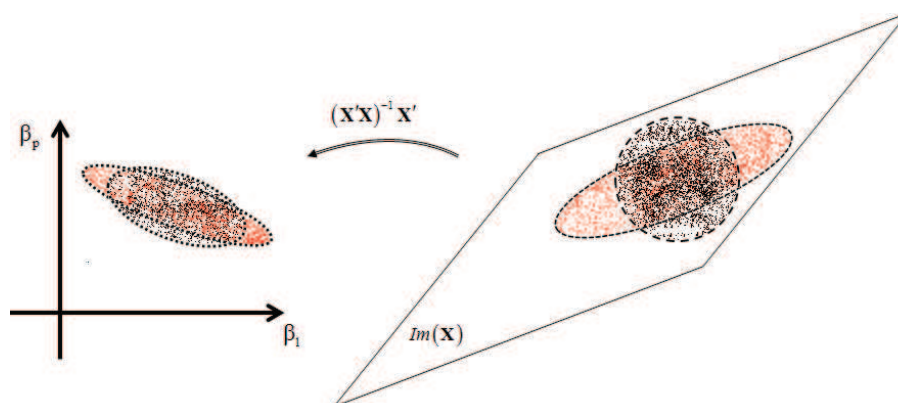


Figura 18 Transformação das nuvens de variâncias do espaço $Im(\mathbf{X})$ para o espaço de parâmetros

3.3 Condições para a igualdade entre os estimadores de Gauss-Markov e Mínimos Quadrados

Um estudo da interessante questão de quando o estimador de Gauss-Markov é igual ao estimador de quadrados mínimos é apresentado em Kruskal(1968). Tal fato é verdadeiro para o caso em que $D(\mathbf{Y}) = \sigma^2 \mathbf{I}$ que é o resultado dado pelo Teorema 1. No entanto, a igualdade entre estes dois estimadores é válida para uma situação mais geral, conforme teorema que se segue (KRUSKAL, op. cit.).

Teorema 3. Para o modelo de Gauss-Markov $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ com $D(\mathbf{Y}) = \sigma^2 \boldsymbol{\Sigma}$, o estimador de Gauss-Markov é igual ao estimador de quadrados mínimos se, e somente se, o subespaço imagem de \mathbf{X} é invariante por $\boldsymbol{\Sigma}$, isto é, $\boldsymbol{\Sigma}(Im(\mathbf{X})) = Im(\mathbf{X})$.

Demonstração. Supondo a igualdade

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \\ &= (\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{Y} \end{aligned}$$

$$= \widehat{\boldsymbol{\beta}}^*,$$

tem-se que

$$\begin{aligned} \mathbf{X}\widehat{\boldsymbol{\beta}} &= \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \\ &= \mathbf{P}^1_{Im(\mathbf{X})}\mathbf{Y}, \end{aligned}$$

e também

$$\begin{aligned} \mathbf{X}\widehat{\boldsymbol{\beta}}^* &= \mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{Y} \\ &= \mathbf{P}^2_{Im(\mathbf{X})}\mathbf{Y}, \end{aligned}$$

em que $\mathbf{P}^1_{Im(\mathbf{X})}$ é a projeção ortogonal de \mathbf{Y} no subespaço $Im(\mathbf{X})$ em relação ao produto interno $\langle \cdot, \cdot \rangle$ e $\mathbf{P}^2_{Im(\mathbf{X})}$ a projeção ortogonal de \mathbf{Y} em $Im(\mathbf{X})$ com relação ao produto $\langle \langle \cdot, \cdot \rangle \rangle = \langle \cdot, \boldsymbol{\Sigma}^{-1} \cdot \rangle$. Logo,

$$\begin{aligned} \widehat{\mathbf{Y}} &= \mathbf{P}^1_{Im(\mathbf{X})}\mathbf{Y} \\ &= \mathbf{P}^2_{Im(\mathbf{X})}\mathbf{Y}. \end{aligned}$$

Como $\langle \mathbf{y} - \widehat{\mathbf{y}}, \mathbf{z} \rangle = 0, \forall \mathbf{z} \in Im(\mathbf{X})$, e $\langle \langle \mathbf{y} - \widehat{\mathbf{y}}^*, \mathbf{z} \rangle \rangle = 0, \forall \mathbf{z} \in Im(\mathbf{X})$, então $\langle \mathbf{y} - \widehat{\mathbf{y}}, \boldsymbol{\Sigma}^{-1}\mathbf{z} \rangle = 0, \forall \mathbf{z} \in Im(\mathbf{X})$, de onde segue que $\mathbf{y} - \widehat{\mathbf{y}}$ é perpendicular a $Im(\mathbf{X})$ e a $\boldsymbol{\Sigma}(Im(\mathbf{X}))$. Como $\mathbf{y} - \widehat{\mathbf{y}}$ é um vetor genérico em $Im(\mathbf{X})^\perp$, segue que $\boldsymbol{\Sigma}(Im(\mathbf{X})) = Im(\mathbf{X})$.

Suponha agora que $\boldsymbol{\Sigma}(Im(\mathbf{X})) = Im(\mathbf{X})$. Tem-se que $\langle \mathbf{y} - \widehat{\mathbf{y}}, \mathbf{z} \rangle = 0, \forall \mathbf{z} \in Im(\mathbf{X})$ e $\langle \langle \mathbf{y} - \widehat{\mathbf{y}}^*, \mathbf{w} \rangle \rangle = \langle \mathbf{y} - \widehat{\mathbf{y}}^*, \boldsymbol{\Sigma}^{-1}\mathbf{w} \rangle = 0, \forall \mathbf{w} \in Im(\mathbf{X})$. Seja

\mathbf{z} um elemento genérico da $Im(\mathbf{X})$ e $\mathbf{w} = \Sigma\mathbf{z}$. Como

$$\begin{aligned}\langle \mathbf{y} - \widehat{\mathbf{y}}^*, \Sigma^{-1}\mathbf{w} \rangle &= \langle \mathbf{y} - \widehat{\mathbf{y}}^*, \Sigma^{-1}\Sigma\mathbf{z} \rangle \\ &= \langle \mathbf{y} - \widehat{\mathbf{y}}^*, \mathbf{z} \rangle \\ &= 0,\end{aligned}$$

segue que ambos $\mathbf{y} - \widehat{\mathbf{y}}^*$ e $\mathbf{y} - \widehat{\mathbf{y}}$ são elementos pertencentes a $Im(\mathbf{X})^\perp$. Portanto, $\widehat{\mathbf{y}} = \widehat{\mathbf{y}}^*$.

Como $(\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}$ e $(\mathbf{X}'\mathbf{X})^{-1}$ são injetivas, vem

$$\begin{aligned}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} &= 0 \\ \Leftrightarrow \mathbf{X}'\mathbf{Y} &= 0 \\ \Leftrightarrow \mathbf{Y} &\in Im(\mathbf{X})^\perp,\end{aligned}$$

onde $Im(\mathbf{X})^\perp$ é o subespaço perpendicular ao subespaço $Im(\mathbf{X})$. Logo, $\Sigma^{-1}\mathbf{Y} \in Ker(\mathbf{X}')$, o que implica que $\Sigma^{-1}\mathbf{Y} \in Im(\mathbf{X})^\perp$. Portanto, fica provado que

$$\Sigma^{-1}\mathbf{Y} \in Im(\mathbf{X})^\perp \text{ se } \mathbf{y} \in Im(\mathbf{X})^\perp$$

$$\Sigma^{-1}(Im(\mathbf{X})^\perp) \subset Im(\mathbf{X})^\perp.$$

Como Σ^{-1} é um isomorfismo, isto é, $\Sigma^{-1}(Im(\mathbf{X})^\perp) = Im(\mathbf{X})^\perp$, prova-se a seguir que $\Sigma^{-1}(Im(\mathbf{X})) = Im(\mathbf{X})$.

Se as duas projeções são iguais, sendo $\widehat{\mathbf{y}}$ o vetor projetado, então em relação aos dois produtos internos

$$\langle \mathbf{y} - \widehat{\mathbf{y}}, \widehat{\mathbf{y}} \rangle = 0$$

$$\langle \mathbf{y}, \hat{\mathbf{y}} \rangle = \langle \hat{\mathbf{y}}, \hat{\mathbf{y}} \rangle,$$

e também

$$\begin{aligned} \langle \langle \mathbf{y} - \hat{\mathbf{y}}, \hat{\mathbf{y}} \rangle \rangle &= 0 \\ \langle \mathbf{y} - \hat{\mathbf{y}}, \Sigma^{-1} \hat{\mathbf{y}} \rangle &= 0 \\ \langle \mathbf{y}, \Sigma^{-1} \hat{\mathbf{y}} \rangle &= \langle \hat{\mathbf{y}}, \Sigma^{-1} \hat{\mathbf{y}} \rangle. \end{aligned}$$

Logo, tem-se que

$$\langle \mathbf{y}, \hat{\mathbf{y}} - \Sigma^{-1} \hat{\mathbf{y}} \rangle = \langle \hat{\mathbf{y}}, \hat{\mathbf{y}} - \Sigma^{-1} \hat{\mathbf{y}} \rangle.$$

Portanto, $\langle \mathbf{y} - \hat{\mathbf{y}}, \hat{\mathbf{y}} - \Sigma^{-1} \hat{\mathbf{y}} \rangle = 0, \quad \forall \mathbf{y}$.

Esta relação é representada na Figura 19.

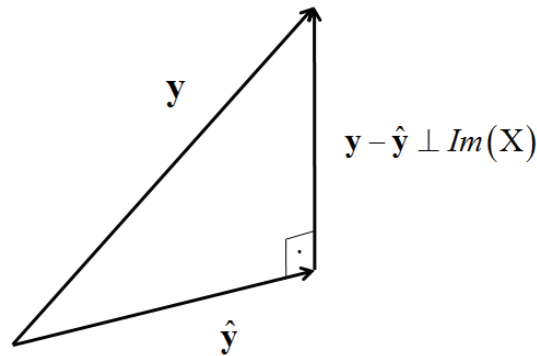


Figura 19 Relação dos vetores \mathbf{y} e $\hat{\mathbf{y}}$ com o subespaço ortogonal a $\text{Im}(\mathbf{X})$

Observe que \mathbf{y} pode ser qualquer vetor em $\hat{\mathbf{y}} + \text{Im}(\mathbf{X})^\perp$, isto é, qualquer vetor que se projete ortogonalmente em $\hat{\mathbf{y}}$. Logo, $\langle \mathbf{y} - \hat{\mathbf{y}}, \hat{\mathbf{y}} - \Sigma^{-1} \hat{\mathbf{y}} \rangle = 0$ implica que $\hat{\mathbf{y}} - \Sigma^{-1} \hat{\mathbf{y}} \in \text{Im}(\mathbf{X})$. Como $\hat{\mathbf{y}} \in \text{Im}(\mathbf{X}) \Rightarrow \Sigma^{-1} \hat{\mathbf{y}} \in \text{Im}(\mathbf{X})$.

□

Um exemplo clássico de matriz que caracteriza as condições de igualdade dos estimadores de Gauss-Markov e quadrados mínimos é a matriz

$$\Sigma = \begin{bmatrix} 1 & \rho & \dots & \rho \\ \rho & 1 & \dots & \rho \\ \dots & \dots & \dots & \dots \\ \rho & \rho & \dots & 1 \end{bmatrix}, \rho \in \mathbb{R}. \quad (3.3)$$

Proposição 14. A matriz Σ na Equação (3.3) atende às condições de igualdade do Teorema 3.

Demonstração. Basta provar que a matriz Σ , aplicada em qualquer vetor da base da matriz \mathbf{X} , pode ser escrita como combinação linear dos próprios vetores da base de \mathbf{X} . Assim, seja a matriz

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{21} & \dots & x_{m1} \\ x_{12} & x_{22} & \dots & x_{m2} \\ \dots & \dots & \dots & \dots \\ x_{1n} & x_{2n} & \dots & x_{mn} \end{bmatrix}_{n \times m}.$$

Aplicando-se a matriz Σ em uma coluna genérica p de \mathbf{X} se tem

$$\begin{bmatrix} 1 & \rho & \dots & \rho \\ \rho & 1 & \dots & \rho \\ \dots & \dots & \dots & \dots \\ \rho & \rho & \dots & 1 \end{bmatrix} \begin{bmatrix} x_{p1} \\ x_{p2} \\ \dots \\ x_{pn} \end{bmatrix} = \begin{bmatrix} x_{p1} + \rho x_{p2} + \dots + \rho x_{pn} \\ \rho x_{p1} + x_{p2} + \dots + \rho x_{pn} \\ \dots \\ \rho x_{p1} + \rho x_{p2} + \dots + x_{pn} \end{bmatrix}_{n \times 1}$$

$$\begin{aligned}
&= \begin{bmatrix} x_{p1} + \rho \sum_{i=1, i \neq 1}^n x_{pi} \\ x_{p2} + \rho \sum_{i=1, i \neq 2}^n x_{pi} \\ \dots \\ x_{pk} + \rho \sum_{i=1, i \neq k}^n x_{pi} \\ \dots \\ x_{pn} + \rho \sum_{i=1, i \neq n}^n x_{pi} \end{bmatrix}_{n \times 1} \\
&= \begin{bmatrix} x_{p1} + \rho \sum_{i=1}^n x_{pi} - \rho x_{p1} \\ x_{p2} + \rho \sum_{i=1}^n x_{pi} - \rho x_{p2} \\ \dots \\ x_{pk} + \rho \sum_{i=1}^n x_{pi} - \rho x_{pk} \\ \dots \\ x_{pn} + \rho \sum_{i=1}^n x_{pi} - \rho x_{pn} \end{bmatrix}_{n \times 1} \\
&= \begin{bmatrix} (1 - \rho) x_{p1} + \rho \sum_{i=1}^n x_{pi} \\ (1 - \rho) x_{p2} + \rho \sum_{i=1}^n x_{pi} \\ \dots \\ (1 - \rho) x_{pk} + \rho \sum_{i=1}^n x_{pi} \\ \dots \\ (1 - \rho) x_{pn} + \rho \sum_{i=1}^n x_{pi} \end{bmatrix}_{n \times 1}
\end{aligned}$$

$$= (1 - \rho) \begin{bmatrix} x_{p1} \\ x_{p2} \\ \dots \\ x_{pk} \\ \dots \\ x_{pn} \end{bmatrix}_{n \times 1} + \rho n \begin{bmatrix} \frac{\sum_{i=1}^n x_{pi}}{n} \\ \frac{\sum_{i=1}^n x_{pi}}{n} \\ \dots \\ \frac{\sum_{i=1}^n x_{pi}}{n} \\ \dots \\ \frac{\sum_{i=1}^n x_{pi}}{n} \end{bmatrix}_{n \times 1},$$

que é combinação linear das colunas de \mathbf{X} . \square

Nas mesmas condições do Teorema 3, isto é, para o modelo de Gauss-Markov $\mathbf{Y} = \mathbf{X}\beta + \varepsilon$ com $D(\mathbf{Y}) = \sigma^2\mathbf{I}$, o estimador não viesado usual para a variância é dado por

$$\begin{aligned} \hat{\sigma}^2 &= \frac{1}{\dim(\text{Im}(\mathbf{X})^\perp)} \|\mathbf{Y} - P_{\text{Im}(\mathbf{X})}(\mathbf{Y})\|^2 \\ &= \frac{1}{n-p} \langle \mathbf{y} - \hat{\mathbf{y}}, \mathbf{y} - \hat{\mathbf{y}} \rangle = \frac{1}{n-p} (\mathbf{y} - \hat{\mathbf{y}}) \cdot (\mathbf{y} - \hat{\mathbf{y}})'. \end{aligned}$$

Para o caso mais geral $\mathbf{Y} = \mathbf{X}\beta + \varepsilon$ com $D(\mathbf{Y}) = \sigma^2\mathbf{\Sigma}$ o estimador é o mesmo, mas com a norma definida pelo produto interno $\langle\langle \cdot, \cdot \rangle\rangle$, isto é,

$$\begin{aligned} \widehat{\sigma}^{*2} &= \frac{1}{n-p} \|\mathbf{Y} - \widehat{\mathbf{y}}^*\|^2 = \frac{1}{n-p} \langle\langle \mathbf{Y} - \widehat{\mathbf{y}}^*, \mathbf{Y} - \widehat{\mathbf{y}}^* \rangle\rangle \\ &= \frac{1}{n-p} (\mathbf{Y} - \widehat{\mathbf{y}}^*)' \mathbf{\Sigma}^{-1} (\mathbf{Y} - \widehat{\mathbf{y}}^*). \end{aligned} \quad (3.4)$$

Para se ver que este estimador é não viesado, tem-se o teorema seguinte

(RENCHEER et al.,2008).

Teorema 4. *Seja \mathbf{Z} um vetor aleatório com média $\boldsymbol{\mu}$ e matriz de covariância $D(\mathbf{Z}) = \boldsymbol{\Sigma}$. Seja também \mathbf{B} uma matriz simétrica $n \times n$. Nestas condições, a esperança da forma quadrática $\mathbf{Z}'\mathbf{B}\mathbf{Z}$ é*

$$E[\mathbf{Z}'\mathbf{B}\mathbf{Z}] = tr(\mathbf{B}\boldsymbol{\Sigma}) + \boldsymbol{\mu}'\mathbf{B}\boldsymbol{\mu}.$$

Demonstração. Como $\boldsymbol{\Sigma} = E(\mathbf{Z}\mathbf{Z}') - \boldsymbol{\mu}\boldsymbol{\mu}'$, então tem-se que $E(\mathbf{Z}\mathbf{Z}') = \boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}'$.

Sendo $\mathbf{Z}'\mathbf{B}\mathbf{Z}$ um escalar, então este equivale ao seu traço. Portanto,

$$\begin{aligned} E[\mathbf{Z}'\mathbf{B}\mathbf{Z}] &= E[tr(\mathbf{Z}'\mathbf{B}\mathbf{Z})] \\ &= E[tr(\mathbf{B}\mathbf{Z}\mathbf{Z}')] \\ &= tr[E(\mathbf{B}\mathbf{Z}\mathbf{Z}')] \\ &= tr[\mathbf{B}E(\mathbf{Z}\mathbf{Z}')] \\ &= tr[\mathbf{B}(\boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}')] \\ &= tr[\mathbf{B}\boldsymbol{\Sigma} + \mathbf{B}\boldsymbol{\mu}\boldsymbol{\mu}'] \\ &= tr(\mathbf{B}\boldsymbol{\Sigma}) + tr(\boldsymbol{\mu}'\mathbf{B}\boldsymbol{\mu}) \\ &= tr(\mathbf{B}\boldsymbol{\Sigma}) + \boldsymbol{\mu}'\mathbf{B}\boldsymbol{\mu}. \end{aligned}$$

□

Aplicando-se o Teorema 4 ao estimador da variância (3.4) tem-se:

Primeiramente, vamos obter a matriz do produto interno $\langle\langle \cdot, \cdot \rangle\rangle$.

$$\langle\langle \mathbf{y} - \widehat{\mathbf{y}}^*, \mathbf{y} - \widehat{\mathbf{y}}^* \rangle\rangle = (\mathbf{y} - \widehat{\mathbf{y}}^*)' \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \widehat{\mathbf{y}}^*)$$

$$= \mathbf{y}'(\mathbf{I} - \mathbf{A})'\Sigma^{-1}(\mathbf{I} - \mathbf{A})\mathbf{y}.$$

Logo, a matriz é $(\mathbf{I} - \mathbf{A})'\Sigma^{-1}(\mathbf{I} - \mathbf{A})$, em que \mathbf{A} é a projeção ortogonal em relação ao novo produto interno.

Como $\mathbf{A} = \Sigma\mathbf{A}'\Sigma^{-1}$ e $tr(\mathbf{BZ}) = tr(\mathbf{ZB})$, temos então que

$$\begin{aligned} E[\widehat{\sigma}^{*2}] &= \frac{1}{n-p} E\left[\left(\mathbf{Y} - \widehat{\mathbf{y}}\right)' \Sigma^{-1} \left(\mathbf{Y} - \widehat{\mathbf{y}}\right)\right] \\ &= \frac{1}{n-p} \left[tr\left(\sigma^2 \Sigma(\mathbf{I} - \mathbf{A})'\Sigma^{-1}(\mathbf{I} - \mathbf{A})\right) + \right. \\ &\quad \left. \left(E\left[\mathbf{Y} - \widehat{\mathbf{y}}\right]\right)' \Sigma^{-1} \left(E\left[\mathbf{Y} - \widehat{\mathbf{y}}\right]\right)\right] \\ &= \frac{1}{n-p} \sigma^2 tr\left(\Sigma(\mathbf{I} - \mathbf{A})'\Sigma^{-1}(\mathbf{I} - \mathbf{A})\right) + 0 \\ &= \frac{1}{n-p} \sigma^2 tr\left(\Sigma\Sigma^{-1} - \Sigma\mathbf{A}'\Sigma^{-1} - \Sigma\Sigma^{-1}\mathbf{A} + \Sigma\mathbf{A}'\Sigma^{-1}\mathbf{A}\right) \\ &= \frac{1}{n-p} \sigma^2 tr(\mathbf{I} - \mathbf{A} - \mathbf{A} + \mathbf{A}\mathbf{A}) \\ &= \frac{1}{n-p} \sigma^2 tr(\mathbf{I} - \mathbf{A} - \mathbf{A} + \mathbf{A}) \\ &= \frac{1}{n-p} \sigma^2 tr(\mathbf{I} - \mathbf{A}) \\ &= \frac{1}{n-p} \sigma^2 (n-p) = \sigma^2. \end{aligned}$$

Conforme visto anteriormente a condição necessária e suficiente para que os modelos $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, $D(\mathbf{Y}) = \sigma^2\mathbf{I}$ e $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, $D(\mathbf{Y}) = \sigma^2\Sigma$ definam estimadores de quadrados mínimos e de Gauss-Markov sejam iguais, é natural então perguntar sob quais condições, além de $\widehat{\mathbf{y}} = \widehat{\mathbf{y}}^*$, também se tenha $\widehat{\sigma}^2 = \widehat{\sigma}^{*2}$. Tal questão é respondida com o seguinte teorema:

Teorema 5. *As igualdades $\widehat{\mathbf{y}} = \widehat{\mathbf{y}}^*$ e $\widehat{\sigma}^2 = \widehat{\sigma}^{*2}$ ocorrem simultaneamente, se e somente se, $\Sigma(Im(\mathbf{X})) = Im(\mathbf{X})$ e $\Sigma(\mathbf{v}) = \mathbf{v}$, $\forall \mathbf{v} \in Im(\mathbf{X})^\perp$.*

Demonstração.

Sendo $\widehat{\mathbf{y}} = \widehat{\mathbf{y}}^*$ e $\widehat{\sigma}^2 = \widehat{\sigma}^{*2}$ então

$$\begin{aligned} \frac{1}{n-p} \langle \mathbf{y} - \widehat{\mathbf{y}}, \mathbf{y} - \widehat{\mathbf{y}} \rangle &= \frac{1}{n-p} \langle \langle \mathbf{Y} - \widehat{\mathbf{y}}^*, \mathbf{Y} - \widehat{\mathbf{y}}^* \rangle \rangle \\ &= \frac{1}{n-p} \langle \mathbf{Y} - \widehat{\mathbf{y}}, \boldsymbol{\Sigma}^{-1} (\mathbf{Y} - \widehat{\mathbf{y}}) \rangle. \end{aligned}$$

Como $\mathbf{Y} - \widehat{\mathbf{y}}$ é um vetor genérico de $Im(\mathbf{X})^\perp$ tem-se que

$$\begin{aligned} \langle \mathbf{v}, \mathbf{v} \rangle &= \langle \mathbf{v}, \boldsymbol{\Sigma}^{-1} \mathbf{v} \rangle \\ &= \langle \langle \mathbf{v}, \mathbf{v} \rangle \rangle, \quad \forall \mathbf{v} \in Im(\mathbf{X})^\perp. \end{aligned}$$

Utilizando a identidade de polarização para $\mathbf{v}, \mathbf{w} \in Im(\mathbf{X})^\perp$ tem-se

$$4 \langle \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{v} + \mathbf{w}, \mathbf{v} + \mathbf{w} \rangle - \langle \mathbf{v} - \mathbf{w}, \mathbf{v} - \mathbf{w} \rangle$$

e

$$4 \langle \langle \mathbf{v}, \mathbf{w} \rangle \rangle = \langle \langle \mathbf{v} + \mathbf{w}, \mathbf{v} + \mathbf{w} \rangle \rangle - \langle \langle \mathbf{v} - \mathbf{w}, \mathbf{v} - \mathbf{w} \rangle \rangle.$$

Logo,

$$\begin{aligned} \langle \mathbf{v}, \mathbf{w} \rangle &= \langle \langle \mathbf{v}, \mathbf{w} \rangle \rangle \\ &= \langle \mathbf{v}, \boldsymbol{\Sigma}^{-1} \mathbf{w} \rangle, \quad \forall \mathbf{v}, \mathbf{w} \in Im(\mathbf{X})^\perp, \end{aligned}$$

e portanto, $\mathbf{w} = \boldsymbol{\Sigma}^{-1} \mathbf{w}$.

Seja então $\mathbf{v} = \boldsymbol{\Sigma}^{-1} \mathbf{v}$, $\forall \mathbf{v} \in Im(\mathbf{X})^\perp$. Assim,

$$\begin{aligned}\langle \mathbf{y} - \hat{\mathbf{y}}, \mathbf{y} - \hat{\mathbf{y}} \rangle &= \langle \mathbf{y} - \hat{\mathbf{y}}, \Sigma^{-1} (\mathbf{y} - \hat{\mathbf{y}}) \rangle \\ &= \langle \langle \mathbf{y} - \hat{\mathbf{y}}, \mathbf{y} - \hat{\mathbf{y}} \rangle \rangle.\end{aligned}$$

Portanto,

$$\langle \langle \mathbf{y} - \hat{\mathbf{y}}^*, \mathbf{y} - \hat{\mathbf{y}}^* \rangle \rangle \Rightarrow \hat{\sigma}^2 = \widehat{\sigma}^{*2}$$

□

As duas próximas subseções (3.4 e 3.5) descrevem como os delineamentos inteiramente casualizados e blocos casualizados podem ser interpretados geometricamente. Os conceitos destas subseções são baseados nas idéias apresentadas em Adão(2011).

3.4 Interpretação geométrica de um delineamento inteiramente casualizado

Dentre os vários tipos de delineamentos experimentais, um dos mais simples é o Delineamento Inteiramente Casualizado (DIC), utilizados quando as condições experimentais são homogêneas, isto é, não há necessidade de se organizar as parcelas experimentais em blocos.

O modelo matemático referente a um delineamento inteiramente casualizado é definido por

$$y_{ij} = \mu + t_i + e_{ij}, \quad (3.5)$$

sendo:

- $i = i$ -ésimo tratamento, $i = 1, 2, \dots, I$;
- $j = j$ -ésima repetição, $j = 1, 2, \dots, J$;
- $y_{ij} =$ valor da parcela do tratamento i na repetição j ;
- $\mu =$ média da população;
- $t_i =$ efeito do tratamento i ;
- $e_{ij} =$ efeito na parcela dos fatores não controlados (erro);

Matricialmente tem-se

$$\begin{bmatrix} y_{11} \\ \dots \\ y_{1J} \\ y_{21} \\ \dots \\ y_{2J} \\ \dots \\ \dots \\ y_{I1} \\ \dots \\ y_{IJ} \end{bmatrix}_{IJ \times 1} = \begin{bmatrix} 1 & 1 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 1 & 1 & 0 & \dots & 0 \\ 1 & 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 1 & 0 & 1 & \dots & 0 \\ 1 & 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ 1 & 0 & 0 & \dots & 0 \\ 1 & 0 & 0 & \dots & 1 \\ \dots & \dots & \dots & \dots & \dots \\ 1 & 0 & 0 & \dots & 1 \end{bmatrix}_{IJ \times (I+1)} \begin{bmatrix} \mu \\ t_1 \\ t_2 \\ \dots \\ t_I \end{bmatrix}_{(I+1) \times 1} + \begin{bmatrix} e_{11} \\ \dots \\ e_{1J} \\ e_{21} \\ \dots \\ e_{2J} \\ \dots \\ \dots \\ e_{I1} \\ \dots \\ e_{IJ} \end{bmatrix}_{IJ \times (1)}.$$

Claramente, a matriz do delineamento, não é de posto completo, já que a primeira coluna desta pode ser escrita como a soma das demais. Assim, a menos da primeira coluna, que está diretamente relacionada com a média da população,

a matriz X seria de posto completo. Portanto, reescrevendo a Equação 3.5 tem-se $y_{ij} - \mu = t_i + e_{ij}$, que no formato do modelo de Gauss-Markov fica $y = X\beta + \varepsilon$ onde, nesta situação, diz-se que y foi corrigido para a média. A partir de então, a matriz do modelo X passa a ter posto completo, isto é, suas colunas são linearmente independentes, e portanto, a matriz $X'X$ passa a ser invertível, o que implica que o projetor ortogonal do vetor de dados no subespaço Ω do modelo é determinado por $X(X'X)^{-1}X'$.

Para se determinar se existe ou não diferença significativa entre os tratamentos, faz-se necessária a análise de variância do experimento, a qual pode ser representada pela Tabela 1, adaptada de Kronka et al.(2006), em que T_i é a soma de todas as parcelas de um tratamento i , isto é, $T_i = \sum_{j=1}^J y_{ij}$.

Tabela 1 Esquema da análise de variância para experimentos no delineamento inteiramente casualizado.

Fonte de Variação	GL	SQ	QM	F_c
Tratamentos	$I - 1$	$\frac{1}{J} \sum_{i=1}^I T_i^2 - C$	$\frac{S.Q. \text{ Tratamentos}}{I-1}$	$\frac{Q.M. \text{ Tratamentos}}{Q.M. \text{ Residuo}}$
Resíduo	$I(J - 1)$	Diferença	$\frac{S.Q. \text{ Residuo}}{I(J-1)}$	—
Total	$IJ - 1$	$\sum_{i=1}^I \sum_{j=1}^J y_{ij}^2 - C$	—	—

O valor C , denominado Correção, que aparece nas equações da soma de quadrados dos tratamentos (S.Q. Tratamentos) e na soma de quadrados total (S.Q. Total) da Tabela 1, é definido por

$$C = \frac{\left(\sum_{i=1}^I \sum_{j=1}^J y_{ij} \right)^2}{IJ}$$

$$= \frac{G^2}{IJ},$$

em que $G = \sum_{i=1}^I \sum_{j=1}^J y_{ij}$ é a soma de todas as parcelas do experimento.

Seja então o vetor \mathbf{y} , de dimensão $n \times 1$, em que cada uma de suas coordenadas representa o valor de uma parcela do experimento realizado. Note que $n = IJ$. Pela Proposição 3, temos que cada coordenada do vetor $\bar{\mathbf{y}}$, projeção de \mathbf{y} na direção do vetor $\mathbf{1}$, é a média de todas as observações, isto é,

$$\begin{aligned} \bar{\mathbf{y}} &= \left(\frac{\sum_{i=1}^I \sum_{j=1}^J y_{ij}}{n}, \dots, \frac{\sum_{i=1}^I \sum_{j=1}^J y_{ij}}{n} \right) \\ &= \left(\frac{G}{n}, \dots, \frac{G}{n} \right) \\ &= \left(\frac{G}{IJ}, \dots, \frac{G}{IJ} \right). \end{aligned}$$

Aplicando-se a Definição 2 na equação anterior tem-se que

$$\begin{aligned} \|\bar{\mathbf{y}}\|^2 &= \frac{G^2}{n^2} + \dots + \frac{G^2}{n^2} \\ &= n \left(\frac{G^2}{n^2} \right) \\ &= \frac{G^2}{n} \\ &= \frac{G^2}{IJ} \\ &= C. \end{aligned}$$

Conclui-se portanto que o valor C da correção, necessário para determinação do quadro de análise de variância, é o quadrado da norma do vetor de médias.

Como $\bar{\mathbf{y}}$ é projeção ortogonal do vetor \mathbf{y} , os vetores $\bar{\mathbf{y}}$, \mathbf{y} e $\mathbf{y} - \bar{\mathbf{y}}$ formam um triângulo retângulo, conforme se pode observar na Figura 20, e pelo teorema

de Pitágoras tem-se

$$\begin{aligned}\|\mathbf{y}\|^2 &= \|\bar{\mathbf{y}}\|^2 + \|\mathbf{y} - \bar{\mathbf{y}}\|^2 \\ \|\mathbf{y} - \bar{\mathbf{y}}\|^2 &= \|\mathbf{y}\|^2 - \|\bar{\mathbf{y}}\|^2.\end{aligned}\quad (3.6)$$

Desenvolvendo a equação 3.6 implica em

$$\begin{aligned}\|\mathbf{y} - \bar{\mathbf{y}}\|^2 &= \|\mathbf{y}\|^2 - \|\bar{\mathbf{y}}\|^2 \\ &= \|\mathbf{y}\|^2 - C \\ &= \mathbf{y} \cdot \mathbf{y} - C \\ &= \sum_{i=1}^I \sum_{j=1}^J y_{ij}^2 - C \\ &= \text{S.Q.Total.}\end{aligned}$$

O vetor de médias dos tratamentos, denominado $\hat{\mathbf{y}}$, é um vetor onde as coordenadas referentes a um mesmo tratamento têm o mesmo valor, sendo este valor a média de cada tratamento, isto é,

$$\hat{\mathbf{y}} = \left(\overbrace{\frac{T_1}{J}, \dots, \frac{T_1}{J}}^{\text{J elementos}}, \overbrace{\frac{T_2}{J}, \dots, \frac{T_2}{J}}^{\text{J elementos}}, \dots, \overbrace{\frac{T_I}{J}, \dots, \frac{T_I}{J}}^{\text{J elementos}} \right).$$

Como o vetor $\hat{\mathbf{y}}$ somente estaria na direção do vetor $\mathbf{1}$ caso todos os tratamentos tivessem o mesmo efeito médio, existe uma diferença (ou afastamento) entre os vetores $\hat{\mathbf{y}}$ e $\bar{\mathbf{y}}$. Novamente, como a projeção ortogonal do vetor $\hat{\mathbf{y}}$ na direção do vetor $\mathbf{1}$ resulta no vetor $\bar{\mathbf{y}}$ (conforme Proposição 8), tem-se um triângulo retângulo formado pelos vetores $\hat{\mathbf{y}}$, $\bar{\mathbf{y}}$ e $\hat{\mathbf{y}} - \bar{\mathbf{y}}$. Portanto, calculando-se o quadrado

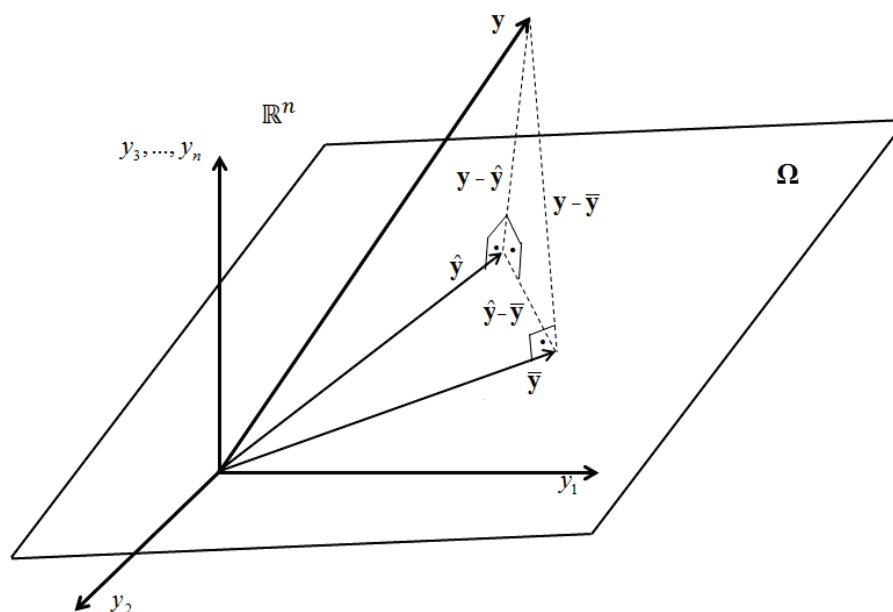


Figura 20 Interpretação geométrica das somas de quadrados da tabela de análise de variância

da norma do vetor $\hat{\mathbf{y}} - \bar{\mathbf{y}}$ pelo teorema de Pitágoras vem

$$\begin{aligned}
 \|\hat{\mathbf{y}}\|^2 &= \|\bar{\mathbf{y}}\|^2 + \|\hat{\mathbf{y}} - \bar{\mathbf{y}}\|^2 \\
 \|\hat{\mathbf{y}} - \bar{\mathbf{y}}\|^2 &= \|\hat{\mathbf{y}}\|^2 - \|\bar{\mathbf{y}}\|^2 \\
 &= \|\hat{\mathbf{y}}\|^2 - C \\
 &= \hat{\mathbf{y}} \cdot \hat{\mathbf{y}} - C \\
 &= \overbrace{\left(\frac{T_1}{J}\right)^2 + \dots + \left(\frac{T_1}{J}\right)^2}^{J \text{ elementos}} + \overbrace{\left(\frac{T_2}{J}\right)^2 + \dots + \left(\frac{T_2}{J}\right)^2}^{J \text{ elementos}} + \\
 &\quad + \dots + \overbrace{\left(\frac{T_1}{J}\right)^2 + \dots + \left(\frac{T_1}{J}\right)^2}^{J \text{ elementos}} - C \\
 &= J\left(\frac{T_1}{J}\right)^2 + J\left(\frac{T_2}{J}\right)^2 + \dots + J\left(\frac{T_1}{J}\right)^2 - C
 \end{aligned}$$

$$\begin{aligned}
&= J \frac{T_1^2}{J^2} + J \frac{T_2^2}{J^2} + \dots + J \frac{T_I^2}{J^2} - C \\
&= \frac{T_1^2}{J} + \frac{T_2^2}{J} + \dots + \frac{T_I^2}{J} - C \\
&= \frac{1}{J} \sum_{i=1}^I T_i^2 - C \\
&= \text{S.Q.Tratamentos.}
\end{aligned}$$

Caso não houvesse o erro experimental, as parcelas referentes a um mesmo tratamento teriam todas o mesmo valor, o qual seria a média do efeito de tal tratamento. Entretanto, devido aos fatores do acaso, o vetor resposta que deveria pertencer ao subespaço do modelo de delineamento (Ω) se afasta deste, formando um triângulo retângulo com hipotenusa igual a $\mathbf{y} - \bar{\mathbf{y}}$ e catetos $\hat{\mathbf{y}} - \bar{\mathbf{y}}$ e $\mathbf{y} - \hat{\mathbf{y}}$, conforme figura 20. Assim, pelo teorema de Pitágoras, é possível calcular $\|\mathbf{y} - \hat{\mathbf{y}}\|^2$ da seguinte maneira:

$$\begin{aligned}
\|\mathbf{y} - \bar{\mathbf{y}}\|^2 &= \|\hat{\mathbf{y}} - \bar{\mathbf{y}}\|^2 + \|\mathbf{y} - \hat{\mathbf{y}}\|^2 \\
\|\mathbf{y} - \hat{\mathbf{y}}\|^2 &= \|\mathbf{y} - \bar{\mathbf{y}}\|^2 - \|\hat{\mathbf{y}} - \bar{\mathbf{y}}\|^2 \\
&= \text{S.Q.Total} - \text{S.Q.Tratamentos,}
\end{aligned}$$

ou seja, S.Q.Resíduos é calculada pela diferença entre S.Q.Total e S.Q. Tratamentos, conforme indicado na Tabela 1.

A partir de agora, a matriz \mathbf{X} será sempre tratada no contexto corrigido para a média. Seja então a matriz \mathbf{X} de um Delineamento Inteiramente Casualizado, com I tratamentos, isto é, $\text{Dim}(Im(\mathbf{X})) = I$. Ao se projetar ortogonalmente um vetor de observações \mathbf{y} , de um espaço de dimensão IJ no subespaço $Im(\mathbf{X})$, obtendo assim o vetor $P_{Im(\mathbf{X})}\mathbf{y} = \hat{\mathbf{y}}$, tem-se que a dimensão do subes-

paço que contém o vetor $(\mathbf{y} - P_{I_m(\mathbf{x})}\mathbf{y})$ é a diferença $(IJ - I) = I(J - 1)$. Portanto, analisando a Tabela 1 e a Figura 20, fica claro que a dimensão do subespaço ao qual pertence o vetor $(\mathbf{y} - \hat{\mathbf{y}})$, cujo quadrado da norma é a S.Q.Resíduo, é o grau de liberdade de sua respectiva soma de quadrados (S.Q.). De forma análoga, tem-se que a dimensão do subespaço ao qual pertence o vetor $(\mathbf{y} - \bar{\mathbf{y}})$ é $IJ - 1$, que é o Grau de Liberdade (G.L) da S.Q.Total. Assim, procedendo com raciocínio análogo, a dimensão do subespaço que contém o vetor $(\hat{\mathbf{y}} - \bar{\mathbf{y}})$ é $(I - 1)$, isto é, o grau de liberdade dos tratamentos. Em suma, o grau de liberdade pode ser interpretado como a dimensão do subespaço ao qual o vetor, cujo quadrado da norma é uma soma de quadrados, pertence.

Com a definição dos graus de liberdade e das somas de quadrados, os quadrados médios (Q.M) podem então ser determinados dividindo-se cada soma de quadrados por seu respectivo grau de liberdade.

Antes de se analisar o teste F, seguem duas proposições (ADÃO, 2011).

Proposição 15. *Seja \mathbf{Y} um vetor normal com $E[\mathbf{Y}] = \boldsymbol{\mu}$ e $\text{Var}[\mathbf{Y}] = \boldsymbol{\Sigma}$. Seja também W um subespaço vetorial m -dimensional, consistindo de autovetores de $\boldsymbol{\Sigma}$ relativos a um autovalor ξ e ortogonal ao vetor de médias $\boldsymbol{\mu}$. Então, $\frac{\|P_W \mathbf{Y}\|^2}{\xi}$ tem distribuição qui-quadrado com m graus de liberdade.*

Demonstração. Seja $\{w_1, \dots, w_m\}$ uma base ortonormal para W . Tem-se que $P_W \mathbf{Y} = (w_1 \cdot \mathbf{Y}) w_1 + \dots + (w_m \cdot \mathbf{Y}) w_m$, $E[w_i \cdot \mathbf{Y}] = w_i \boldsymbol{\mu} = 0$ e $\text{Var}[w_i \cdot \mathbf{Y}] = w_i' \boldsymbol{\Sigma} w_i = \xi$. Portanto, $w_i \cdot \mathbf{Y}$ são variáveis normais independentes de média 0 e variância ξ . Segue então que $\frac{\|P_W \mathbf{Y}\|^2}{\xi} = \frac{(w_1 \cdot \mathbf{Y})^2 + \dots + (w_m \cdot \mathbf{Y})^2}{\xi}$ tem distribuição qui-quadrado com m graus de liberdade. \square

Proposição 16. *Seja \mathbf{Y} um vetor normal com $E[\mathbf{Y}] = \boldsymbol{\mu}$ e $\text{Var}[\mathbf{Y}] = \boldsymbol{\Sigma}$. Se W_1 e W_2 são subespaços de W de dimensões d_1 e d_2 consistindo de autovetores*

de um mesmo autovalor ξ e ortogonais entre si, então $\frac{\|P_{W_1}Y\|^2}{d_1} / \frac{\|P_{W_2}Y\|^2}{d_2}$ segue distribuição F com d_1 e d_2 graus de liberdade.

Demonstração. Sendo W_1 e W_2 subespaços ortogonais, então $P_{W_1}Y$ e $P_{W_2}Y$ são vetores aleatórios independentes. Também, conforme Proposição 15, os vetores $\frac{\|P_{W_1}Y\|^2}{\xi}$ e $\frac{\|P_{W_2}Y\|^2}{\xi}$ têm distribuição qui-quadrado com graus de liberdade, respectivamente, d_1 e d_2 , o que implica que possuem quociente com distribuição F . \square

Assim sendo, pela Proposição 15, temos que o Q.M. Tratamentos e o Q.M. Resíduo seguem ambos distribuição qui-quadrado com graus de liberdade $(I - 1)$ e $I(J - 1)$ respectivamente.

Portanto, pela Proposição 16, tem-se que $\frac{\text{Q.M. Tratamentos}}{\text{Q.M. Resíduo}}$ tem distribuição F , permitindo assim a realização do teste F para se detectar diferença entre as médias dos tratamentos.

Segue então um exemplo, com dados adaptados de Banzatto et al.(2006).

Na Tabela 2, constam valores referentes a um DIC em que as parcelas representam a produtividade de três variedades de mandioca. Portanto, a matriz corrigida do modelo é

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

O vetor y de dados é representado por

$$y' = \begin{bmatrix} 38,9 & 25,4 & 20,3 & 25,7 & 38,7 & 43,2 \\ 41,7 & 39,0 & 47,8 & 47,8 & 44,7 & 50,5 \end{bmatrix}.$$

Conforme Proposição 6, tem-se que a matriz que projeta o vetor y no subespaço gerado pelo vetor unitário $\mathbf{1}$, determinando assim o vetor \bar{y} , é a matriz

$$\begin{aligned} \mathbf{A}y &= \frac{1}{12} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} y \\ &= \left[\begin{array}{cccc} 38.64167 & 38.64167 & \dots & 38.64167 \end{array} \right]_{1 \times 12}' \\ &= \bar{y}. \end{aligned}$$

Tabela 2 Produtividade dos cultivares de mandioca, em t/ha.

TRATAMENTOS	REPETIÇÕES				TOTAIS
	1	2	3	4	
A-IAC 5	38,9	25,4	20,3	25,7	110,3
B-IRACEMA	38,7	43,2	41,7	39,0	162,6
C-MANTIQUEIRA	47,8	47,8	44,7	50,5	190,8
					463,7

Utilizando-se da Proposição 7, o projetor ortogonal do vetor \mathbf{y} no subespaço $Im(\mathbf{X})$ é dado por

$$P_{Im(\mathbf{X})} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$$

$$= \begin{bmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{bmatrix}.$$

Aplicando-se o projetor $\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ no vetor \mathbf{y} tem-se

$$\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \begin{bmatrix} 27.575 \\ 27.575 \\ 27.575 \\ 27.575 \\ 40.650 \\ 40.650 \\ 40.650 \\ 40.650 \\ 47.700 \\ 47.700 \\ 47.700 \\ 47.700 \end{bmatrix} = \hat{\mathbf{y}}.$$

Portanto, com os vetores \mathbf{y} , $\hat{\mathbf{y}}$ e $\bar{\mathbf{y}}$ determinados, deve-se calcular os quadrados das normas dos vetores diferenças destes, isto é, $\|\mathbf{y} - \bar{\mathbf{y}}\|^2$, $\|\hat{\mathbf{y}} - \bar{\mathbf{y}}\|^2$ e

$\|\mathbf{y} - \hat{\mathbf{y}}\|^2$. Logo,

$$\mathbf{y} - \bar{y} = \begin{bmatrix} 38,9 \\ 25,4 \\ 20,3 \\ 25,7 \\ 38,7 \\ 43,2 \\ 41,7 \\ 39,0 \\ 47,8 \\ 47,8 \\ 44,7 \\ 50,5 \end{bmatrix} - \begin{bmatrix} 38.64167 \\ 38.64167 \\ 38.64167 \\ 38.64167 \\ 38.64167 \\ 38.64167 \\ 38.64167 \\ 38.64167 \\ 38.64167 \\ 38.64167 \\ 38.64167 \\ 38.64167 \end{bmatrix}$$

$$= \begin{bmatrix} 0.258333 \\ -13.241666 \\ -18.341666 \\ -12.941666 \\ 0.058333 \\ 4.558333 \\ 3.058333 \\ 0.358333 \\ 9.158333 \\ 9.158333 \\ 6.058333 \\ 11.858333 \end{bmatrix}$$

$$\begin{aligned} &= \begin{bmatrix} -11.066667 \\ -11.066667 \\ -11.066667 \\ -11.066667 \\ 2.008333 \\ 2.008333 \\ 2.008333 \\ 2.008333 \\ 9.058333 \\ 9.058333 \\ 9.058333 \\ 9.058333 \end{bmatrix} \\ \|\hat{\mathbf{y}} - \bar{\mathbf{y}}\|^2 &= (-11.066667)^2 + (-11.066667)^2 + \dots + (9.058333)^2 \\ &= 834.23 \\ &= S.Q.Tratamentos. \end{aligned}$$

Finalmente, determinando a soma de quadrados do resíduo se tem

$$\begin{aligned}
 \mathbf{y} - \hat{\mathbf{y}} &= \begin{bmatrix} 38,9 \\ 25,4 \\ 20,3 \\ 25,7 \\ 38,7 \\ 43,2 \\ 41,7 \\ 39,0 \\ 47,8 \\ 47,8 \\ 44,7 \\ 50,5 \end{bmatrix} - \begin{bmatrix} 27.575 \\ 27.575 \\ 27.575 \\ 27.575 \\ 40.650 \\ 40.650 \\ 40.650 \\ 40.650 \\ 47.700 \\ 47.700 \\ 47.700 \\ 47.700 \end{bmatrix} \\
 &= \begin{bmatrix} 11.325 \\ -2.175 \\ -7.275 \\ -1.875 \\ -1.950 \\ 2.550 \\ 1.050 \\ -1.650 \\ 0.100 \\ 0.100 \\ -3.00 \\ 2.800 \end{bmatrix} \\
 \|\mathbf{y} - \hat{\mathbf{y}}\|^2 &= (11.325)^2 + (-2.175)^2 + \dots + (2.800)^2
 \end{aligned}$$

$$= 220.4175$$

$$= S.Q.Residuo.$$

Observa-se que a matriz \mathbf{X} tem posto igual a 4, isto é, $Dim(Im(\mathbf{X})) = 4$, enquanto que o espaço que contém o vetor y tem dimensão 12. Sendo o número de tratamentos igual a 3 e o número de repetições igual a 4, os graus de liberdade dos espaços são:

$$G.L.Total = IJ - 1 = 12 - 1 = 11.$$

$$G.L.Tratamentos = I - 1 = 3 - 1 = 2.$$

$$G.L.Residuo = IJ - I = 12 - 3 = 9.$$

A Figura 21 é o resultado da análise de variância realizada com o auxílio do software Sisvar. Conforme se observa, os valores calculados conferem com os valores apresentados no quadro de análise de variância.

```
Arquivo analisado:
E:\Mestrado\Dissertacao\Após Qualificação\DIC.DB
-----
Variável analisada: valor
Opção de transformação: Variável sem transformação ( Y )
-----
```

TABELA DE ANÁLISE DE VARIÂNCIA

FV	GL	SQ	QM	Fc	Pr>Fc
trat	2	834.231667	417.115833	17.032	0.0009
erro	9	220.417500	24.490833		
Total corrigido	11	1054.649167			
CV (%) =	12.81				
Média geral:	38.6416667	Número de observações:		12	

Figura 21 Quadro de Análise de Variância dos dados do exemplo DIC

3.5 Interpretação geométrica de um delineamento em blocos casualizado

Um dos mais utilizados dentre todos os tipos de delineamentos, tendo a característica de ser aplicado quando não há homogeneidade das condições experimentais, o delineamento em blocos casualizados se caracteriza por controlar localmente fontes de variações, criando assim ambientes homogêneos (blocos), sendo que em cada um destes blocos estão presentes todos os tratamentos (BANZATTO et al., 2006).

Define-se o modelo de delineamento em blocos casualizados por

$$y_{ij} = \mu + t_i + b_j + e_{ij}, \quad (3.7)$$

sendo:

i = i -ésimo tratamento, $i = 1, 2, \dots, I$;

j = j -ésimo bloco, $j = 1, 2, \dots, J$;

y_{ij} = valor da parcela do tratamento i no bloco j ;

μ = média da população;

t_i = efeito do tratamento i ;

b_j = efeito do bloco j ;

e_{ij} = efeito na parcela dos fatores não controlados (erro);

Representando o modelo em forma matricial, tem-se:

$$\begin{bmatrix}
 y_{11} \\
 y_{12} \\
 \dots \\
 y_{1J} \\
 y_{21} \\
 y_{22} \\
 \dots \\
 y_{2J} \\
 \dots \\
 \dots \\
 y_{I1} \\
 y_{I2} \\
 \dots \\
 y_{IJ}
 \end{bmatrix}_{IJ \times 1} = \begin{bmatrix}
 1 & 1 & 0 & \dots & \dots & 0 & 1 & 0 & \dots & \dots & 0 \\
 1 & 1 & 0 & \dots & \dots & 0 & 0 & 1 & \dots & \dots & 0 \\
 \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\
 1 & 1 & 0 & \dots & \dots & 0 & 0 & 0 & \dots & \dots & 1 \\
 1 & 0 & 1 & \dots & \dots & 0 & 1 & 0 & \dots & \dots & 0 \\
 1 & 0 & 1 & \dots & \dots & 0 & 0 & 1 & \dots & \dots & 0 \\
 \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\
 1 & 0 & 1 & \dots & \dots & 0 & 0 & 0 & \dots & \dots & 1 \\
 \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\
 \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\
 1 & 0 & 0 & \dots & \dots & 1 & 1 & 0 & \dots & \dots & 0 \\
 1 & 0 & 0 & \dots & \dots & 1 & 0 & 1 & \dots & \dots & 0 \\
 \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\
 1 & 0 & 0 & \dots & \dots & 1 & 0 & 0 & \dots & \dots & 1
 \end{bmatrix}_{IJ \times (I+J+1)}$$

$$\begin{bmatrix} \mu \\ t_1 \\ t_2 \\ \dots \\ t_I \\ b_1 \\ b_2 \\ \dots \\ b_J \end{bmatrix}_{(I+J+1) \times 1} + \begin{bmatrix} e_{11} \\ e_{12} \\ \dots \\ e_{1J} \\ e_{21} \\ e_{22} \\ \dots \\ e_{2J} \\ \dots \\ \dots \\ e_{I1} \\ e_{I2} \\ \dots \\ e_{IJ} \end{bmatrix}_{IJ \times 1} .$$

Logo, pode-se escrever que $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, isto é, o delineamento em blocos casualizados pode ser representado nos termos do modelo de Gauss-Markov.

De forma análoga ao DIC, observa-se que a matriz do delineamento não é de posto completo, já que se tem a primeira coluna (referente ao efeito da média μ) como sendo uma combinação linear das próximas I colunas. Portanto, reescrevendo a Equação 3.7 tem-se $y_{ij} - \mu = t_i + b_j + e_{ij}$. O problema é que, diferentemente do delineamento inteiramente casualizado, neste caso o fato de se corrigir os dados para a média não implica em uma matriz de delineamento com posto coluna completo, já que na matriz \mathbf{X} corrigida para a média, as I primeiras colunas são combinação linear das J últimas colunas. Entretanto, a matriz \mathbf{X} , a menos da primeira coluna, pode ser seccionada em duas matrizes de posto coluna completo, uma referente aos tratamentos e outra referente aos blocos. Seja então

\mathbf{X}_b a matriz de blocos, e \mathbf{X}_t a matriz de tratamentos, isto é,

$$\mathbf{X}_t = \begin{bmatrix} 1 & 0 & \dots & \dots & 0 \\ 1 & 0 & \dots & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 1 & 0 & \dots & \dots & 0 \\ 0 & 1 & \dots & \dots & 0 \\ 0 & 1 & \dots & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 1 & \dots & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \dots & 1 \\ 0 & 0 & \dots & \dots & 1 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \dots & 1 \end{bmatrix} \quad \text{e} \quad \mathbf{X}_b = \begin{bmatrix} 1 & 0 & \dots & \dots & 0 \\ 0 & 1 & \dots & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \dots & 1 \\ 1 & 0 & \dots & \dots & 0 \\ 0 & 1 & \dots & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \dots & 1 \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ 1 & 0 & \dots & \dots & 0 \\ 0 & 1 & \dots & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \dots & 1 \end{bmatrix} \quad .(3.8)$$

Pode-se agora observar que tanto \mathbf{X}_t quanto \mathbf{X}_b são matrizes de posto coluna completo.

O procedimento de se particionar a matriz \mathbf{X} , de posto coluna incompleto, em matrizes com posto coluna completo se faz necessário para que seja possível trabalhar matrizes invertíveis e, assim, se determinar projetores da forma

$$\mathbf{K}(\mathbf{K}'\mathbf{K})^{-1}\mathbf{K}'$$

onde \mathbf{K} é uma das matrizes seccionadas da matriz \mathbf{X} .

O posto da matriz \mathbf{X}_t é I , já que cada coluna desta matriz está associada a um tratamento, enquanto que o posto da matriz \mathbf{X}_b é J , sendo que cada

coluna desta está relacionada a um bloco. Assim, as dimensões dos subespaços gerados pelas matrizes \mathbf{X}_t e \mathbf{X}_b são, respectivamente, $Dim(Im(\mathbf{X}_t)) = I$ e $Dim(Im(\mathbf{X}_b)) = J$.

Como exemplificação, considere um DBC com 3 tratamentos e 4 blocos, isto é, $I = 3$ e $J = 4$. Note que o vetor de dados observados tem dimensão $12 = IJ$. Neste contexto, as matrizes \mathbf{X}_t e \mathbf{X}_b são, respectivamente, as seguintes:

$$\begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \quad e \quad \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Um vetor qualquer que pertença ao subespaço dos tratamentos tem a seguinte forma:

$$\mathbf{y}'_t = \left[a \ a \ a \ a \ b \ b \ b \ b \ c \ c \ c \ c \right], \ a, b, c, d \in \mathbb{R},$$

enquanto que um vetor do subespaço dos blocos tem a forma como se segue:

$$\mathbf{y}'_b = \left[a \ b \ c \ d \ a \ b \ c \ d \ a \ b \ c \ d \right], \ a, b, c, d \in \mathbb{R}.$$

Uma observação bastante relevante é que o vetor de médias \bar{y} pertence à intersecção dos subespaços de tratamento e de blocos.

Seja então a Tabela 3, adaptada de Kronka et al.(2006), a qual esquematiza a decomposição das variâncias de um Delineamento em Blocos Casualizados.

Tabela 3 Esquema da análise de variância para experimentos no delineamento em blocos casualizado.

Fonte de Variação	GL	SQ	QM	F_c
Tratamentos	$I - 1$	$\frac{1}{J} \sum_{i=1}^I T_i^2 - C$	$\frac{S.Q. \text{ Tratamentos}}{I-1}$	$\frac{Q.M. \text{ Tratamentos}}{Q.M. \text{ Residuo}}$
Blocos	$J - 1$	$\frac{1}{I} \sum_{j=1}^J B_j^2 - C$	$\frac{S.Q. \text{ Blocos}}{J-1}$	$\frac{Q.M. \text{ Blocos}}{Q.M. \text{ Residuo}}$
Resíduo	$(I - 1)(J - 1)$	Diferença	$\frac{S.Q. \text{ Residuo}}{(I-1)(J-1)}$	—
Total	$IJ - 1$	$\sum_{i=1}^I \sum_{j=1}^J y_{ij}^2 - C$	—	—

Primeiramente, comparando a Tabela 3 com as matrizes \mathbf{X}_b e \mathbf{X}_t em (3.8), observa-se que o grau de liberdade dos tratamentos é exatamente o posto da matriz de tratamentos menos um, bem como o grau de liberdade de blocos é determinado subtraindo-se uma unidade do posto da matriz de blocos. Já o G.L. total é obtido subtraindo-se uma unidade do valor da dimensão do espaço dos dados, isto é, $IJ - 1$. O grau de liberdade restante, ou seja, G.L. do resíduo, é obtido por diferença, isto é,

$$\begin{aligned}
 \text{G.L. Resíduo} &= \text{G.L. Total} - \text{G.L. Tratamentos} - \text{G.L. Blocos} \\
 &= (IJ - 1) - (I - 1) - (J - 1) \\
 &= IJ - 1 - I + 1 - J + 1 \\
 &= IJ - I - J + 1 \\
 &= (I - 1)(J - 1).
 \end{aligned}$$

Assim como demonstrado no delineamento inteiramente casualizado, o

valor da correção é dado por

$$\begin{aligned} C &= \frac{\left(\sum_{i=1}^I \sum_{j=1}^J y_{ij} \right)^2}{IJ} \\ &= \frac{G^2}{IJ} \\ &= \|\bar{\mathbf{y}}\|^2, \end{aligned}$$

em que $\bar{\mathbf{y}}$ é o vetor em que todas as coordenadas são iguais e têm o valor da média aritmética das coordenadas do vetor \mathbf{y} , isto é, a média de todas as parcelas do experimento.

Fazendo um comparativo com o modelo de Delineamento Inteiramente Casualizado, no DBC o espaço onde é projetado o vetor de observações não mais é denominado de espaço dos tratamentos, já que este novo espaço é composto por um subespaço referente aos tratamentos e a um outro subespaço referente aos blocos. Assim, adicionalmente à análise feita no DIC, agora se tem os vetores $\hat{\mathbf{y}}_b$ e $\hat{\mathbf{y}}_t$, respectivamente projeção do vetor \mathbf{y} nos subespaços de blocos e de tratamentos, onde tais vetores são representados por

$$\hat{\mathbf{y}}_t = \left(\underbrace{\left(\frac{T_1}{J}, \dots, \frac{T_1}{J} \right)}_{1^\circ \text{ conjunto}}, \underbrace{\left(\frac{T_2}{J}, \dots, \frac{T_2}{J} \right)}_{2^\circ \text{ conjunto}}, \dots, \underbrace{\left(\frac{T_I}{J}, \dots, \frac{T_I}{J} \right)}_{I\text{-ésimo conjunto}} \right)$$

e

$$\hat{\mathbf{y}}_b = \left(\underbrace{\left(\frac{B_1}{I}, \dots, \frac{B_J}{I} \right)}_{1^\circ \text{ conjunto}}, \underbrace{\left(\frac{B_1}{I}, \dots, \frac{B_J}{I} \right)}_{2^\circ \text{ conjunto}}, \dots, \underbrace{\left(\frac{B_1}{I}, \dots, \frac{B_J}{I} \right)}_{I\text{-ésimo conjunto}} \right).$$

Calculando a norma de cada um destes vetores, segue que

$$\begin{aligned}
\|\hat{\mathbf{y}}_b\|^2 &= \hat{\mathbf{y}}_b' \hat{\mathbf{y}}_b \\
&= \left(\frac{B_1}{I}\right)^2 + \cdots + \left(\frac{B_J}{I}\right)^2 + \left(\frac{B_1}{I}\right)^2 + \cdots + \left(\frac{B_J}{I}\right)^2 \\
&\quad + \cdots + \left(\frac{B_1}{I}\right)^2 + \cdots + \left(\frac{B_J}{I}\right)^2 \\
&= I\left(\frac{B_1}{I}\right)^2 + I\left(\frac{B_2}{I}\right)^2 + \cdots + I\left(\frac{B_J}{I}\right)^2 \\
&= I\left(\frac{B_1^2}{I^2} + \frac{B_2^2}{I^2} + \cdots + \frac{B_J^2}{I^2}\right) \\
&= I\left(\frac{1}{I^2} \sum_{j=1}^J B_j^2\right) \\
&= \frac{1}{I} \sum_{j=1}^J B_j^2,
\end{aligned}$$

e

$$\begin{aligned}
\|\hat{\mathbf{y}}_t\|^2 &= \left(\frac{T_1}{J}\right)^2 + \cdots + \left(\frac{T_1}{J}\right)^2 + \left(\frac{T_2}{J}\right)^2 + \cdots + \left(\frac{T_2}{J}\right)^2 \\
&\quad + \cdots + \left(\frac{T_I}{J}\right)^2 + \cdots + \left(\frac{T_I}{J}\right)^2 \\
&= J\left(\frac{T_1}{J}\right)^2 + J\left(\frac{T_2}{J}\right)^2 + \cdots + J\left(\frac{T_I}{J}\right)^2 \\
&= J\left(\frac{T_1^2}{J^2} + \frac{T_2^2}{J^2} + \cdots + \frac{T_I^2}{J^2}\right) \\
&= J\left(\frac{1}{J^2} \sum_{i=1}^I T_i^2\right) \\
&= \frac{1}{J} \sum_{i=1}^I T_i^2.
\end{aligned}$$

Sendo ortogonal a projeção do vetor de blocos no subespaço de médias,

tem-se um triângulo retângulo em que a hipotenusa é o próprio vetor de blocos. Logo,

$$\|\hat{\mathbf{y}}_b\|^2 = \|\bar{\mathbf{y}}\|^2 + \|\hat{\mathbf{y}}_b - \bar{\mathbf{y}}\|^2. \quad (3.9)$$

Reescrevendo a Equação 3.9 e aplicando os valores das normas (calculados anteriormente), tem-se que

$$\begin{aligned} \|\hat{\mathbf{y}}_b - \bar{\mathbf{y}}\|^2 &= \|\hat{\mathbf{y}}_b\|^2 - \|\bar{\mathbf{y}}\|^2 \\ &= \frac{1}{I} \sum_{j=1}^J B_j^2 - C \\ &= \text{S.Q.Blocos.} \end{aligned}$$

De forma semelhante, projetando-se ortogonalmente o vetor de tratamentos no subespaço das médias, um triângulo retângulo é formado, sendo hipotenusa o vetor de tratamentos. Portanto,

$$\|\hat{\mathbf{y}}_t\|^2 = \|\bar{\mathbf{y}}\|^2 + \|\hat{\mathbf{y}}_t - \bar{\mathbf{y}}\|^2, \quad (3.10)$$

podendo-se rearranjar a Equação 3.10 e aplicar, simultaneamente, as normas calculadas anteriormente para se concluir que

$$\begin{aligned} \|\hat{\mathbf{y}}_t - \bar{\mathbf{y}}\|^2 &= \|\hat{\mathbf{y}}_t\|^2 - \|\bar{\mathbf{y}}\|^2 \\ &= \frac{1}{J} \sum_{i=1}^I T_i^2 - C \\ &= \text{S.Q.Tratamentos.} \end{aligned}$$

A S.Q.Total é encontrada, em termos vetoriais, de forma estritamente análoga à apresentada no modelo de Delineamento Inteiramente Casualizado. Assim sendo, pode-se definir que

$$\begin{aligned}
 \|\mathbf{y} - \bar{\mathbf{y}}\|^2 &= \|\mathbf{y}\|^2 - \|\bar{\mathbf{y}}\|^2 \\
 &= \|\mathbf{y}\|^2 - C \\
 &= \mathbf{y} \cdot \mathbf{y} - C \\
 &= \sum_{i=1}^I \sum_{j=1}^J y_{ij}^2 - C \\
 &= \text{S.Q.Total.}
 \end{aligned}$$

Sobre a S.Q.Resíduo, esta também é obtida por diferença. Entretanto, surge um termo a mais quando comparado com o DIC, já que um novo subespaço (blocos) foi definido no processo. Assim, a soma de quadrados dos resíduos é calculada por

$$\begin{aligned}
 \text{S.Q.Resíduo} &= \text{S.Q.Total} - (\text{S.Q.Tratamentos} + \text{S.Q.Blocos}) \\
 &= \text{S.Q.Total} - \text{S.Q.Tratamentos} - \text{S.Q.Blocos.}
 \end{aligned}$$

Neste ponto, com os graus de liberdade definidos e as somas de quadrados determinadas, para se determinar os quadrados médios (Q.M) basta dividir cada Soma de Quadrado pelo seu respectivo grau de liberdade, assim como apresentado na Tabela 3.

Um fato, cuja demonstração será omitida, é que a projeção ortogonal do vetor de dados no subespaço das médias equivale à projeção do vetor de blocos (que por si só já é a projeção do vetor de dados no subespaço dos blocos) no su-

subespaço das médias. A importância desta afirmação é que tem-se assim formado um triângulo retângulo em que o teorema de Pitágoras pode ser aplicado da seguinte forma:

$$\|\mathbf{y} - \bar{\mathbf{y}}\|^2 = \|\mathbf{y} - \hat{\mathbf{y}}_b\|^2 + \|\hat{\mathbf{y}}_b - \bar{\mathbf{y}}\|^2.$$

Note que é possível manipular a equação anterior de forma a se isolar o termo $\|\hat{\mathbf{y}}_b - \bar{\mathbf{y}}\|^2$, que representa a S.Q.Blocos, sem para tanto utilizar a correção C (norma do vetor de médias). Fazendo análise similar para Tratamentos, tem-se que

$$\|\mathbf{y} - \bar{\mathbf{y}}\|^2 = \|\mathbf{y} - \hat{\mathbf{y}}_t\|^2 + \|\hat{\mathbf{y}}_t - \bar{\mathbf{y}}\|^2,$$

que também pode ser manipulada para se encontrar a S.Q.Tratamentos sem a utilização da correção C.

De forma análoga à análise feita no delineamento inteiramente casualizado, utiliza-se das Proposições 15 e 16 para demonstrar que $\frac{Q.M. \text{ Tratamentos}}{Q.M. \text{ Resíduo}}$ segue distribuição F, bem como $\frac{Q.M. \text{ Blocos}}{Q.M. \text{ Resíduo}}$ também o segue, o que permite a utilização de tal distribuição para testes de hipótese quanto à significância de igualdade das médias, tanto para tratamentos quanto para blocos.

Geometricamente, uma figura que representa as condições e situações até então descritas para o Delineamento em Blocos Casualizado é bastante similar àquela representada na Figura (20). Entretanto, uma diferença crucial é que o subespaço do modelo (Ω) no DBC é composto por dois outros subespaços ortogonais entre si, que são os subespaços de tratamento e de bloco. Presente na intersecção destes dois últimos está o subespaço gerado pelo vetor unitário, sendo que o vetor de médias μ está presente neste. Note que a S.Q.T. continua sendo calculada da mesma forma tanto no DBC quanto no DIC, bem como a ortogonalidade do

vetor de erros com o subespaço Ω . Ocorre porém que o vetor proveniente da diferença $\|\hat{y} - \bar{y}\|^2$ pode ser decomposto em duas componentes ortogonais: uma pertencente ao subespaço dos tratamentos e outra pertencente ao subespaço dos blocos, sendo que o vetor \bar{y} pertence à intersecção destes dois últimos. A Figura 22 representa as relações descritas anteriormente.

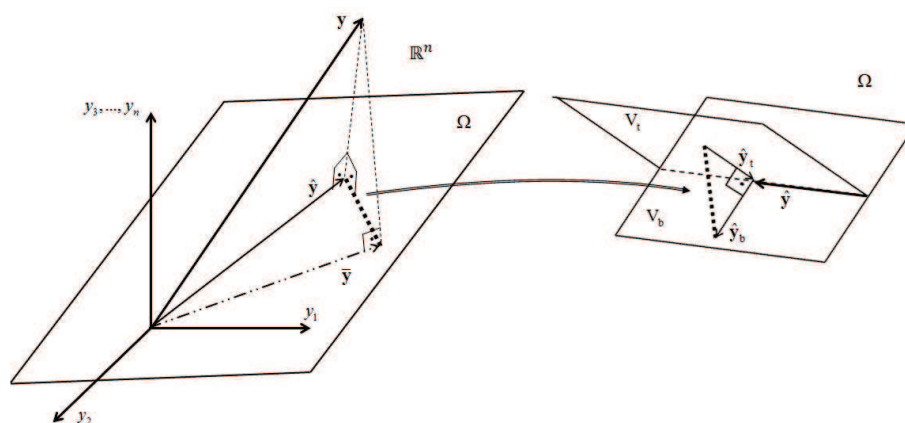


Figura 22 Interpretação geométrica de um delineamento em blocos casualizado (DIC)

Seja então um exemplo, com dados adaptados de Banzatto et al.(2006).

Na Tabela 4 estão relacionados os pesos, em gramas, de frutos de macieira num experimento realizado para se testar 3 tratamentos que são: A - 12,5 ppm do produto Promalin, B - 50,0 ppm do produto Promalin e C - Testemunha (0 ppm).

Tabela 4 Pesos médios dos frutos da macieira, em gramas.

TRATAMENTOS	BLOCOS				TOTAIS
	1	2	3	4	
A-12,5 ppm	142,4	144,8	145,2	138,9	571,3
B-50,0 ppm	140,7	134,1	136,1	144,1	555,0
C-TESTEMUNHA	153,5	165,0	151,8	150,2	620,5
TOTAIS	436,6	443,9	433,1	433,2	1746,8

Para este exemplo, em que se tem $I=3$ tratamentos e $J=4$ blocos, as matrizes dos subespaços de Tratamento e Blocos são, respectivamente,

$$\mathbf{X}_t = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{e} \quad \mathbf{X}_b = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

O vetor y de dados é representado por

$$y' = \begin{bmatrix} 142,4 & 144,8 & 145,2 & 138,9 & 140,7 & 134,1 \end{bmatrix}$$

$$136,1 \quad 144,1 \quad 153,5 \quad 165,0 \quad 151,8 \quad 150,2 \Big].$$

Pela Proposição 6, o vetor de médias \bar{y} pode ser encontrado por

$$\begin{aligned} \bar{y} &= \mathbf{A}y \\ &= \frac{1}{12} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} y \\ &= \left[145.566 \quad 145.566 \quad \cdots \quad 145.566 \right]_{1 \times 12}' \end{aligned}$$

Antes de se encontrar os vetores \hat{y}_t e \hat{y}_b , é necessário determinar as matrizes de projeção em cada um dos subespaços de Tratamentos e de Blocos. Para tanto, faz-se o emprego da Proposição 7, isto é, $P_{Im(\mathbf{X}_b)}\mathbf{y} = \mathbf{X}_b(\mathbf{X}'_b\mathbf{X}_b)^{-1}\mathbf{X}'_b\mathbf{y}$

e $P_{Im(\mathbf{X}_t)}\mathbf{y} = \mathbf{X}_t(\mathbf{X}'_t\mathbf{X}_t)^{-1}\mathbf{X}'_t\mathbf{y}$. Pertanto,

$$\mathbf{X}_t(\mathbf{X}'_t\mathbf{X}_t)^{-1}\mathbf{X}'_t = \begin{bmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{bmatrix}$$

e

$$\mathbf{X}_b(\mathbf{X}'_b\mathbf{X}_b)^{-1}\mathbf{X}'_b = \begin{bmatrix} \frac{1}{3} & 0 & 0 & 0 & \frac{1}{3} & 0 & 0 & 0 & \frac{1}{3} & 0 & 0 & 0 \\ 0 & \frac{1}{3} & 0 & 0 & 0 & \frac{1}{3} & 0 & 0 & 0 & \frac{1}{3} & 0 & 0 \\ 0 & 0 & \frac{1}{3} & 0 & 0 & 0 & \frac{1}{3} & 0 & 0 & 0 & \frac{1}{3} & 0 \\ 0 & 0 & 0 & \frac{1}{3} & 0 & 0 & 0 & \frac{1}{3} & 0 & 0 & 0 & \frac{1}{3} \\ \frac{1}{3} & 0 & 0 & 0 & \frac{1}{3} & 0 & 0 & 0 & \frac{1}{3} & 0 & 0 & 0 \\ 0 & \frac{1}{3} & 0 & 0 & 0 & \frac{1}{3} & 0 & 0 & 0 & \frac{1}{3} & 0 & 0 \\ 0 & 0 & \frac{1}{3} & 0 & 0 & 0 & \frac{1}{3} & 0 & 0 & 0 & \frac{1}{3} & 0 \\ 0 & 0 & 0 & \frac{1}{3} & 0 & 0 & 0 & \frac{1}{3} & 0 & 0 & 0 & \frac{1}{3} \\ \frac{1}{3} & 0 & 0 & 0 & \frac{1}{3} & 0 & 0 & 0 & \frac{1}{3} & 0 & 0 & 0 \\ 0 & \frac{1}{3} & 0 & 0 & 0 & \frac{1}{3} & 0 & 0 & 0 & \frac{1}{3} & 0 & 0 \\ 0 & 0 & \frac{1}{3} & 0 & 0 & 0 & \frac{1}{3} & 0 & 0 & 0 & \frac{1}{3} & 0 \\ 0 & 0 & 0 & \frac{1}{3} & 0 & 0 & 0 & \frac{1}{3} & 0 & 0 & 0 & \frac{1}{3} \end{bmatrix}.$$

Ao se aplicar estes dois projetores no vetor de dados \mathbf{y} , ficam determinados os

Portanto,

$$\hat{\mathbf{y}}_t = \begin{bmatrix} 142,825 \\ 142,825 \\ 142,825 \\ 142,825 \\ 138,750 \\ 138,750 \\ 138,750 \\ 138,750 \\ 138,750 \\ 155,125 \\ 155,125 \\ 155,125 \\ 155,125 \end{bmatrix} \quad \text{e} \quad \hat{\mathbf{y}}_b = \begin{bmatrix} 145,5333 \\ 147,9667 \\ 144,3667 \\ 144,400 \\ 145,5333 \\ 147,9667 \\ 144,3667 \\ 144,400 \\ 144,400 \\ 145,5333 \\ 147,9667 \\ 144,3667 \\ 144,400 \end{bmatrix} .$$

Assim, com os vetores \mathbf{y} , $\bar{\mathbf{y}}$, $\hat{\mathbf{y}}_t$ e $\hat{\mathbf{y}}_b$ encontrados, as somas de quadrados podem ser determinadas conforme descrito anteriormente, isto é,

$$\begin{aligned} \text{S.Q.Total} &= \|\mathbf{y} - \bar{\mathbf{y}}\|^2 \\ &= (\mathbf{y} - \bar{\mathbf{y}})'(\mathbf{y} - \bar{\mathbf{y}}) \\ &= (142,4 - 145,566)^2 + (144,8 - 145,566)^2 + \dots \\ &\quad + (150,2 - 145,566)^2 \\ &= 803,04667. \end{aligned}$$

$$\text{S.Q.Blocos} = (\hat{\mathbf{y}}_b - \bar{\mathbf{y}})'(\hat{\mathbf{y}}_b - \bar{\mathbf{y}})$$

$$\begin{aligned}
&= (145,5333 - 145,566)^2 + (147,9667 - 145,566)^2 + \dots \\
&\quad + (144,4 - 145,566)^2 \\
&= 25,686666.
\end{aligned}$$

$$\begin{aligned}
\text{S.Q.Tratamentos} &= (\hat{\mathbf{y}}_t - \bar{\mathbf{y}})' (\hat{\mathbf{y}}_t - \bar{\mathbf{y}}) \\
&= (142,825 - 145,566)^2 + (142,825 - 145,566)^2 + \dots \\
&\quad + (155,125 - 145,566)^2 \\
&= 581,38166.
\end{aligned}$$

$$\begin{aligned}
\text{S.Q.Residuo} &= \text{S.Q.Total} - \text{S.Q.Blocos} - \text{S.Q.Tratamento} \\
&= 803,04667 - 25,686666 - 581,38166 \\
&= 195,9783.
\end{aligned}$$

Conforme se pode observar na tabela da Figura 23, estes valores de soma de quadrados conferem com os valores executados no software Sisvar.

3.6 A teoria de Gauss-Markov na presença de dados faltantes

Suponha \mathbf{Y} um vetor aleatório em \mathbb{R}^n , e suponha sem perda de generalidade que as primeiras m coordenadas de \mathbf{Y} sejam observáveis, e as restantes $n - m$ sejam não observáveis ou faltantes. Usando a notação $\mathbf{Y} = \mathbf{Y}^1 + \mathbf{Y}^2$ em que \mathbf{Y}^1 é a parte observável e \mathbf{Y}^2 a não observável, considere o modelo $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ que considera todos os dados (observáveis ou não), $\mathbf{X} : \mathbb{R}^p \mapsto \mathbb{R}^n$.

```

Arquivo analisado:
E:\Mestrado\Dissertacao\Após Qualificação\DBC.DB
-----
Variável analisada: Resposta
Opção de transformação: Variável sem transformação ( Y )
-----

```

TABELA DE ANÁLISE DE VARIÂNCIA

FV	GL	SQ	QM	Fc	Pr>Fc
Trat	2	581.381667	290.690833	8.900	0.0160
Bloco	3	25.686667	8.562222	0.262	0.8504
erro	6	195.978333	32.663056		
Total corrigido	11	803.046667			
CV (%) =	3.93				
Média geral:	145.5666667	Número de observações:		12	

Figura 23 Quadro de Análise de Variância dos dados do exemplo DBC

Neste item $Im(\mathbf{X})$ será denotada por Ω . Desta forma tem-se a relação

$$\mathbb{R}^n = \mathbb{R}^m \oplus \mathbb{R}^{n-m}$$

Considere também $\Omega_1 = P_{\mathbb{R}^m}(\Omega)$ e $\Omega_2 = P_{\mathbb{R}^{n-m}}(\Omega)$. Observe que, em geral, $\Omega \neq \Omega_1 + \Omega_2$. O que se tem é que $\Omega \subset \Omega_1 + \Omega_2$. Vamos considerar a seguinte hipótese: $dim(\Omega) = dim(\Omega_1)$. Estatisticamente, tal fato significa que o número de parâmetros do modelo, mesmo com a ocorrência de dados faltantes, não se altera.

$P_{\mathbb{R}^m} : \Omega \mapsto \Omega_1$ é linear e sobrejetiva. Como $dim(\Omega) = dim(\Omega_1)$ então a projeção também é injetiva, isto é, $ker(P_{\mathbb{R}^m})$ restrito à Ω é o subespaço nulo. Logo, se $\mathbf{v}_1 + \mathbf{v}_2 \in \Omega$ e $\mathbf{v}'_1 + \mathbf{v}'_2 \in \Omega$ então

$$\begin{aligned}
 (\mathbf{v}_1 + \mathbf{v}_2) - (\mathbf{v}'_1 + \mathbf{v}'_2) &\in ker(P_{\mathbb{R}^m}) \\
 \mathbf{v}_2 - \mathbf{v}'_2 &= 0 \\
 \mathbf{v}_2 &= \mathbf{v}'_2,
 \end{aligned}$$

isto é, para cada vetor \mathbf{v}_1 em Ω_1 existe um único vetor \mathbf{v}_2 em Ω_2 tal que $\mathbf{v}_1 + \mathbf{v}_2 \in \Omega$, e portanto o *kernel* da projeção pertence à Ω . Pode-se então definir uma transformação linear $\mathbf{A} : \mathbb{R}^n \mapsto \mathbb{R}^n$ tal que $\mathbf{A}(\mathbf{v}_1) = \mathbf{v}_2$ e $\mathbf{A}(\mathbf{x}) = \mathbf{0}$ se $\mathbf{x} \in \Omega_1^\perp$. Seja a transformação linear $\mathbf{I} : \Omega_1 \mapsto \mathbb{R}^n$, inclusão, isto é, $\mathbf{I}(\mathbf{v}_1) = \mathbf{v}_1$, $\mathbf{v}_1 \in \Omega_1$. Desta forma, pode-se escrever, utilizando um certo abuso de notação, que $\Omega = (\mathbf{I} + \mathbf{A}) \Omega_1$.

Se $\boldsymbol{\mu}$ é o vetor de médias geral do experimento, o estimador de Gauss-Markov de $\boldsymbol{\mu}$ é $\hat{\boldsymbol{\mu}} = P_\Omega(\mathbf{Y})$. Como as m últimas componentes de \mathbf{Y} são dados faltantes, não é possível se obter $\hat{\boldsymbol{\mu}}$. A estratégia então é obter o estimador de Gauss-Markov para os dados observados utilizando o sub-modelo dado por Ω_1 , isto é, $\hat{\boldsymbol{\mu}}^1 = P_{\Omega_1}(\mathbf{Y}^1)$. Sendo $\boldsymbol{\mu}^1 = E[\mathbf{Y}^1]$, para $\hat{\boldsymbol{\mu}}^1$ existe um único $\hat{\boldsymbol{\mu}}^2$ dado por $\hat{\boldsymbol{\mu}}^2 = \mathbf{A}(\hat{\boldsymbol{\mu}}^1)$. Toma-se agora como estimador de $\hat{\boldsymbol{\mu}}$ o vetor obtido por

$$\hat{\boldsymbol{\mu}} = P_\Omega(\mathbf{Y}^1 + \hat{\boldsymbol{\mu}}^2) \quad (3.11)$$

A representação destes vetores e suas projeções nos subespaços discutidos consta na Figura 24. Justifica-se que o processo de imputação de dados é razoável quando m for bem menor do que n .

ção $\mathbf{Y}^1 + \hat{\boldsymbol{\mu}}^2$. Logo, $\left\| \underbrace{(\mathbf{Y}^1 + \hat{\boldsymbol{\mu}}^2)}_{\text{dados com imputação}} - \underbrace{(\hat{\boldsymbol{\mu}}^2 + \hat{\boldsymbol{\mu}}^1)}_{\in \Omega} \right\| = \|\mathbf{Y}^1 - \hat{\boldsymbol{\mu}}^1\|$. Como $P_{\Omega}(\mathbf{Y}^1 + \hat{\boldsymbol{\mu}}^2)$ é o ponto mais próximo de $(\mathbf{Y}^1 + \hat{\boldsymbol{\mu}}^2)$ em Ω , então qualquer ponto que pertença à Ω e que seja diferente de $\hat{\boldsymbol{\mu}}$ terá uma distância maior até o vetor $(\mathbf{Y}^1 + \hat{\boldsymbol{\mu}}^2)$. Assim sendo, $\hat{\boldsymbol{\mu}}^1 + \hat{\boldsymbol{\mu}}^2 \in \Omega$ e é diferente de $\hat{\boldsymbol{\mu}}$, o que implica em

$$\|(\mathbf{Y}^1 + \hat{\boldsymbol{\mu}}^2) - P_{\Omega}(\mathbf{Y}^1 + \hat{\boldsymbol{\mu}}^1)\| \leq \|\mathbf{Y}^1 - \hat{\boldsymbol{\mu}}^1\|.$$

□

Como exemplo para aplicação de dados faltantes, seja um delineamento inteiramente casualizado (DIC) com 3 tratamentos e 4 repetições $y_{ij} = \mu + \tau_i + \varepsilon_{ij}$, com $i = \{1, 2, 3\}$ e $j = \{1, 2, 3, 4\}$. y_{ij} é a observação referente ao i -ésimo tratamento e à j -ésima repetição. O vetor observado é representado de forma genérica como

$$\mathbf{y} = (y_{11}, y_{12}, \dots, y_{14}, y_{21}, \dots, y_{24}, y_{31}, \dots, y_{34})'.$$

Vamos supor então que a última observação tenha sido perdida, isto é, tem-se apenas o vetor

$$\mathbf{y}^1 = (y_{11}, y_{12}, \dots, y_{14}, y_{21}, \dots, y_{24}, y_{31}, \dots, y_{33}, 0)'.$$

Neste caso, temos:

$$\Omega = \{(a, a, a, a, b, b, b, b, c, c, c, c)', a, b, c \in \mathbb{R}\},$$

$$\Omega_1 = \{(a, a, a, a, b, b, b, b, c, c, c, 0)', a, b, c \in \mathbb{R}\},$$

e

$$\Omega_2 = \{(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, d)'\}, d \in \mathbb{R}\}.$$

Define-se então a função natural $\mathbf{A} : \Omega_1 \mapsto \Omega_2$, em que

$$\mathbf{A}((a, a, a, a, b, b, b, b, c, c, c, 0)') = (0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, c)'$$

Definindo $\bar{y}_1 = \frac{y_{11} + \dots + y_{14}}{4}$, $\bar{y}_2 = \frac{y_{21} + \dots + y_{24}}{4}$ e $\bar{y}'_3 = \frac{y_{31} + \dots + y_{33}}{3}$, se tem

$$P_{\Omega_1}(\mathbf{y}^1) = \left(\overbrace{\bar{y}_1, \dots, \bar{y}_1}^4, \overbrace{\bar{y}_2, \dots, \bar{y}_2}^4, \overbrace{\bar{y}'_3, \dots, \bar{y}'_3}^3, 0 \right)' = \hat{\boldsymbol{\mu}}^1.$$

$$\mathbf{A}(\hat{\boldsymbol{\mu}}^1) = (0, 0, 0, 0, 0, 0, 0, 0, 0, 0, \bar{y}'_3)' = \hat{\boldsymbol{\mu}}^2$$

Logo, o novo vetor de dados em que é levado em conta o dado faltante é

$$\begin{aligned} \mathbf{y}^* &= \hat{\boldsymbol{\mu}}^2 + \mathbf{y}^1 \\ &= (0, 0, \dots, 0, \bar{y}_3)' + (y_{11}, \dots, y_{14}, y_{21}, \dots, y_{24}, y_{31}, \dots, y_{33}, 0)' \\ &= (y_{11}, \dots, y_{14}, y_{21}, \dots, y_{24}, y_{31}, \dots, y_{33}, \bar{y}'_3)'. \end{aligned}$$

$$\text{Portanto, } P_{\Omega} \mathbf{y}^* = \left(\overbrace{\bar{y}_1, \dots, \bar{y}_1}^4, \overbrace{\bar{y}_2, \dots, \bar{y}_2}^4, \overbrace{\alpha, \dots, \alpha}^4 \right)' = \hat{\boldsymbol{\mu}}, \text{ em que}$$

$$\begin{aligned} \alpha &= \frac{y_{31} + y_{32} + y_{33} + \frac{y_{31} + y_{32} + y_{33}}{3}}{4} \\ &= \frac{\frac{3(y_{31} + y_{32} + y_{33}) + (y_{31} + y_{32} + y_{33})}{3}}{4} \end{aligned}$$

$$= \frac{y_{31} + y_{32} + y_{33}}{3}.$$

Em termos matriciais, o exemplo anterior é tratado da seguinte maneira:

Os subespaços são imagens das transformações lineares

$$\mathbf{X} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{e} \quad \mathbf{X}_1 = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

Note que a matriz \mathbf{X}_2 pode ser determinada por $\mathbf{X}_2 = \mathbf{X} - \mathbf{X}_1$.

O projetor que fornece $\hat{\mu}^1$ é $\mathbf{X}_1(\mathbf{X}'_1\mathbf{X}_1)^{-1}\mathbf{X}'_1$. Aplicando-se este proje-

tor no vetor de dados \mathbf{Y}^1 tem-se então

$$\begin{bmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{14} \\ y_{21} \\ y_{22} \\ y_{23} \\ y_{24} \\ y_{31} \\ y_{32} \\ y_{33} \\ 0 \end{bmatrix} = \begin{bmatrix} \bar{y}_{1\cdot} \\ \bar{y}_{1\cdot} \\ \bar{y}_{1\cdot} \\ \bar{y}_{1\cdot} \\ \bar{y}_{2\cdot} \\ \bar{y}_{2\cdot} \\ \bar{y}_{2\cdot} \\ \bar{y}_{2\cdot} \\ \bar{y}'_3 \\ \bar{y}'_3 \\ \bar{y}'_3 \\ 0 \end{bmatrix} = \hat{\boldsymbol{\mu}}^1.$$

A transformação linear \mathbf{A} que leva o vetor $\hat{\boldsymbol{\mu}}^1$ em Ω_2 , $\mathbf{A}(\hat{\boldsymbol{\mu}}^1) = \hat{\boldsymbol{\mu}}^2$, é

$$\begin{bmatrix}
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0
\end{bmatrix}
\begin{bmatrix}
\bar{y}_1. \\
\bar{y}_1. \\
\bar{y}_1. \\
\bar{y}_1. \\
\bar{y}_2. \\
\bar{y}_2. \\
\bar{y}_2. \\
\bar{y}_2. \\
\bar{y}_3' \\
\bar{y}_3' \\
\bar{y}_3' \\
0
\end{bmatrix}
=
\begin{bmatrix}
0 \\
0 \\
0 \\
0 \\
0 \\
0 \\
0 \\
0 \\
0 \\
0 \\
0 \\
0 \\
\bar{y}_3'
\end{bmatrix}
= \hat{\boldsymbol{\mu}}^2.$$

Portanto, os dados com a imputação são da forma

$$\hat{\boldsymbol{\mu}}^2 + \mathbf{Y}^1 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ \bar{y}'_3 \end{bmatrix} + \begin{bmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{14} \\ y_{21} \\ y_{22} \\ y_{23} \\ y_{24} \\ y_{31} \\ y_{32} \\ y_{33} \\ 0 \end{bmatrix} = \begin{bmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{14} \\ y_{21} \\ y_{22} \\ y_{23} \\ y_{24} \\ y_{31} \\ y_{32} \\ y_{33} \\ \bar{y}'_3 \end{bmatrix},$$

onde $\bar{y}'_3 = \frac{y_{31} + \dots + y_{33}}{3}$.

Conforme se pode visualizar na Figura 24, $\hat{\boldsymbol{\mu}}^2 + \mathbf{Y}^1 \notin \Omega$ e portanto

$$\begin{aligned} P_{\Omega}(\hat{\boldsymbol{\mu}}^2 + \mathbf{Y}^1) &= \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\hat{\boldsymbol{\mu}}^2 + \mathbf{Y}^1) \\ &= \hat{\boldsymbol{\mu}}. \end{aligned}$$

3.7 A teoria de Gauss-Markov na presença de dados adicionais

Suponha que se tenha o modelo de Gauss-Markov $\mathbf{Y}^1 = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, com $\mathbf{Y}^1 \in \mathbb{R}^n$. Suponha também que novas observações $\mathbf{Y}^2 \in \mathbb{R}^m$ sejam obtidas. A questão é como estes dados extras podem ser incorporados ao processo de estimação. Seja $\mathbf{Y} = \mathbf{Y}^1 + \mathbf{Y}^2$ o vetor de dados completo em \mathbb{R}^{n+m} . Quer se estender

o modelo $\Omega_1 = \text{Im}(\mathbf{X})$ em \mathbb{R}^n para Ω em \mathbb{R}^{n+m} . Novamente a hipótese central é que o número de parâmetros não é alterado, isto é, $\dim(\Omega_1) = \dim(\Omega)$. Desta forma, para cada vetor $\mathbf{v}_1 \in \Omega_1$ existe um único vetor \mathbf{v}_2 tal que $\mathbf{v}_1 + \mathbf{v}_2 \in \Omega$. Denominando Ω_2 como o conjunto destes vetores, fica definida a transformação linear $\mathbf{A} : \mathbb{R}^n \mapsto \mathbb{R}^{n+m}$ da forma $\mathbf{A}(\mathbf{v}_1) = \mathbf{v}_2$ se $\mathbf{v}_1 \in \Omega_1$ e $\mathbf{A}(\boldsymbol{\omega}) = \mathbf{0}$ se $\boldsymbol{\omega} \in \Omega_1^\perp$.

A idéia para se estimar $\boldsymbol{\mu}$ é projetar ortogonalmente o vetor de dados completos $\mathbf{Y} = \mathbf{Y}^1 + \mathbf{Y}^2$ no subespaço Ω e depois projetar novamente no subespaço Ω_1 do modelo original $\mathbf{X}\hat{\boldsymbol{\mu}} = P_{\Omega_1}P_{\Omega}(\mathbf{Y}^1 + \mathbf{Y}^2)$.

Segue na Figura 25 a representação geométrica das considerações descritas anteriormente.

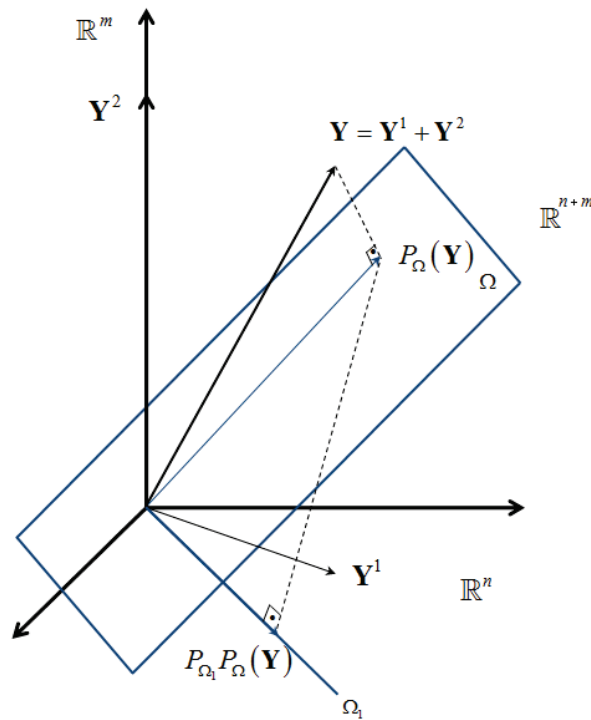


Figura 25 Projeções para dados extras

De forma similar ao caso de dados faltantes, seja um delineamento inteiramente casualizado (DIC) com três tratamentos e 4 repetições, sendo o modelo $y_{ij} = \mu + \tau_i + \varepsilon_{ij}$, com $i = \{1, 2, 3\}$ e $j = \{1, 2, 3, 4\}$. y_{ij} é a observação referente ao i -ésimo tratamento e à j -ésima repetição.

$$\mathbf{y}^1 = (y_{11}, \dots, y_{14}, y_{21}, \dots, y_{24}, y_{31}, \dots, y_{34})'.$$

Suponha então que se tenha uma observação a mais, referente ao terceiro tratamento. Considerando agora os vetores em $\mathbb{R}^{12} \oplus \mathbb{R}^1 = \mathbb{R}^{13}$ tem-se

$$\mathbf{y}^1 = (y_{11}, \dots, y_{14}, y_{21}, \dots, y_{24}, y_{31}, \dots, y_{34}, 0)'$$

$$\mathbf{y}^2 = (0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, y_{35})'$$

$$\mathbf{y} = \mathbf{y}^1 + \mathbf{y}^2 = (y_{11}, \dots, y_{14}, y_{21}, \dots, y_{24}, y_{31}, \dots, y_{34}, y_{35})'$$

e os subespaços

$$\Omega_1 = \{(a, a, a, a, b, b, b, b, c, c, c, c, 0)'\}$$

e

$$\Omega = \{(a, a, a, a, b, b, b, b, c, c, c, c, c)'\}.$$

Portanto, a projeção em Ω é dada por

$$P_{\Omega}\mathbf{y} = \{(\bar{\mathbf{y}}_{1\cdot}, \bar{\mathbf{y}}_{1\cdot}, \bar{\mathbf{y}}_{1\cdot}, \bar{\mathbf{y}}_{1\cdot}, \bar{\mathbf{y}}_{2\cdot}, \bar{\mathbf{y}}_{2\cdot}, \bar{\mathbf{y}}_{2\cdot}, \bar{\mathbf{y}}_{2\cdot}, \bar{\mathbf{y}}_{3\cdot}^*, \bar{\mathbf{y}}_{3\cdot}^*, \bar{\mathbf{y}}_{3\cdot}^*, \bar{\mathbf{y}}_{3\cdot}^*, \bar{\mathbf{y}}_{3\cdot}^*)'\},$$

$$\text{sendo } \bar{\mathbf{y}}_{1\cdot} = \frac{\sum_{i=1}^4 y_{1i}}{4}, \bar{\mathbf{y}}_{2\cdot} = \frac{\sum_{i=1}^4 y_{2i}}{4} \text{ e } \bar{\mathbf{y}}_{3\cdot}^* = \frac{\sum_{i=1}^5 y_{3i}}{5}.$$

Finalmente, o vetor $\boldsymbol{\mu}$ é determinado por

$$\boldsymbol{\mu} = P_{\Omega_1} (P_{\Omega} \mathbf{y}) = \{(\bar{y}_{1\cdot}, \bar{y}_{1\cdot}, \bar{y}_{1\cdot}, \bar{y}_{1\cdot}, \bar{y}_{2\cdot}, \bar{y}_{2\cdot}, \bar{y}_{2\cdot}, \bar{y}_{2\cdot}, \bar{y}_{3\cdot}^*, \bar{y}_{3\cdot}^*, \bar{y}_{3\cdot}^*, \bar{y}_{3\cdot}^*, 0)\}'.$$

4 CONCLUSÃO

A demonstração da identidade de Fisher e da identidade de David-Hartley-Pearson, utilizando conceitos geométricos, se apresenta de forma mais concisa e com aparente potencial didático.

O emprego de conceitos geométricos para a imputação de dados, no caso de dados faltantes é intuitiva, justificando os procedimentos usualmente adotados, como o uso de médias. De forma análoga, o uso da teoria para dados adicionais possui as mesmas qualidades.

A interpretação geométrica do Teorema de Gauss-Markov apresentada pode ser uma ferramenta útil na compreensão de seu resultado.

A utilização de construções geométricas como subespaços vetoriais, projeções e transformações lineares é uma forma natural de se abordar conceitos estatísticos.

REFERÊNCIAS

- ADÃO, A. S. Introdução à teoria geométrica dos delineamentos experimentais. **Dissertação (Mestrado em Estatística e Experimentação Agropecuária)** - Universidade Federal de Lavras, Lavras, MG.
- BAILEY, R. A. **Design of comparative experiments**. Cambridge: Cambridge University Press, 2008. 330 p.
- BANZATTO, D. A.; KRONKA, S. N. **Experimentação agrícola**. 4. ed. Jaboticabal: Funep, 2008. 237 p.
- BOLDRINI, J. L. et al. **Álgebra Linear**. 2. ed. São Paulo: Harbra, 1986. 411 p.
- DAVID, H. A; HARTLEY, H. O; PEARSON, E. S. The distribution of the ratio, in a single normal sample, of range to standard deviation. **Biometrika**, v. 41, p. 482-493, 1954.
- FERREIRA, D. F; CHAVES, L. M; SOUZA, D. J. The theory of the internally studentized range distribution revisited. *Preprint submitted to **Computational Statistics & Data Analysis***, 2012.
- GRAYBILL, F. A. **Theory and application of the linear model**. Belmont: Duxbury Press, 1976. 704 p.
- KRUSKAL, W. The coordinate-free approach to Gauss-Markov estimation and its application to missing and extra observations. **Proc. Fourth Berkeley Symp. Math. Statist. Probab.**, v. 1, p. 435-451, 1961.
- KRUSKAL, W. When are Gauss-Markov and least squares estimators identical? A coordinate-free approach. **The Annals of Mathematical Statistics**, v. 39, p. 70-75, 1968.
- MORAIS, A. R; CHAVES, L. M; COSTA, C. P. T. **Introdução à álgebra de matrizes**. Universidade Federal de Lavras: UFLA/FAEPE, 2001. 240 p.

RENCHEER, A. C; SCHAALJE, G.B. **Linear Models in statistics**. New Jersey: John Wiley & Sons, 2008. 672 p.

SAVILLE, D. J; WOOD, G. L. A Method for Teaching Statistics Using N-Dimensional Geometry. **The American Statistician**, v. 40, p. 205-214, 1986.

SAVILLE, D. J; WOOD, G. L. **Statistical Methods: The Geometric Approach**. New York: Springer-Verlag, 1991. 560 p.

SEARLE, S. R. **Linear Models for Unbalanced Data**. New York: John Wiley & Sons, 1987. 536 p.

SILVEIRA, F. G; COSTA, L. A.; PEREIRA, L. S, et al.; Uma demonstração geométrica para uma identidade de Fisher para o modelo de dois fatores. **Revista Brasileira de Biometria**, v. 30, n. 2, p. 199-222, 2012.

APÊNDICE

Determinação do estimador de quadrados mínimos, $\hat{\beta}$, utilizando cálculo diferencial

Algebricamente, seja $L(\beta) = \|\mathbf{Y} - \mathbf{X}\beta\|^2$. Então,

$$\begin{aligned} L(\beta) &= \|\mathbf{Y} - \mathbf{X}\beta\|^2 \\ &= (\mathbf{Y} - \mathbf{X}\beta)'(\mathbf{Y} - \mathbf{X}\beta) \\ &= (\mathbf{Y}' - \beta'\mathbf{X}')(\mathbf{Y} - \mathbf{X}\beta) \\ &= \mathbf{Y}'\mathbf{Y} - \mathbf{Y}'\mathbf{X}\beta - \beta'\mathbf{X}'\mathbf{Y} + \beta'\mathbf{X}'\mathbf{X}\beta. \end{aligned}$$

Claramente se pode observar que $\mathbf{Y}'\mathbf{X}\beta$ é um número e, portanto, $\mathbf{Y}'\mathbf{X}\beta = (\mathbf{Y}'\mathbf{X}\beta)'$. Assim, segue que

$$\begin{aligned} L(\beta) &= \mathbf{Y}'\mathbf{Y} - 2\beta'\mathbf{X}'\mathbf{Y} + \beta'\mathbf{X}'\mathbf{X}\beta \\ \frac{d}{d\beta}L(\beta) &= -2\mathbf{X}'\mathbf{Y} + 2\mathbf{X}'\mathbf{X}\beta. \end{aligned}$$

Igualando-se a derivada a 0 segue que

$$\begin{aligned} -2\mathbf{X}'\mathbf{Y} + 2\mathbf{X}'\mathbf{X}\hat{\beta} &= \mathbf{0} \\ \Rightarrow \mathbf{X}'\mathbf{X}\hat{\beta} &= \mathbf{X}'\mathbf{Y} \\ (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\hat{\beta} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \end{aligned}$$

$$\boxed{\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}}. \quad (4.1)$$

Determinação do estimador de Gauss-Markov, $\widehat{\beta}^*$, utilizando cálculo diferencial

Seja $\mathbf{Y} = (y_1, \dots, y_n)'$, $\beta = (\beta_1, \dots, \beta_p)'$, $D(\mathbf{Y}) = \Sigma_{n \times n}$ e $\mathbf{X}_{n \times p}$. Como $f(\mathbf{X}) = \mathbf{A} + \mathbf{X}'\mathbf{B} + \mathbf{X}'\mathbf{C}\mathbf{X} \Rightarrow \frac{\partial f}{\partial \mathbf{X}} = (\mathbf{B} + 2\mathbf{C}\mathbf{X})' = \mathbf{B}' + 2\mathbf{X}'\mathbf{C}'$ e $g(\mathbf{X}) = \mathbf{A}\mathbf{X} \Rightarrow \frac{\partial g}{\partial \mathbf{X}} = \mathbf{A}'$ tem-se que

$$\begin{aligned} \langle\langle (\mathbf{Y} - \mathbf{X}\beta), (\mathbf{Y} - \mathbf{X}\beta) \rangle\rangle &= \langle \mathbf{Y} - \mathbf{X}\beta, \Sigma^{-1}(\mathbf{Y} - \mathbf{X}\beta) \rangle \\ &= \langle \mathbf{Y} - \mathbf{X}\beta, \Sigma^{-1}\mathbf{Y} - \Sigma^{-1}\mathbf{X}\beta \rangle \\ &= \langle \mathbf{Y}, \Sigma^{-1}\mathbf{Y} \rangle - \langle \mathbf{Y}, \Sigma^{-1}\mathbf{X}\beta \rangle - \langle \mathbf{X}\beta, \Sigma^{-1}\mathbf{Y} \rangle \\ &\quad + \langle \mathbf{X}\beta, \Sigma^{-1}\mathbf{X}\beta \rangle \\ &= \mathbf{Y}'\Sigma^{-1}\mathbf{Y} - \mathbf{Y}'\Sigma^{-1}\mathbf{X}\beta - \beta'\mathbf{X}'\Sigma^{-1}\mathbf{Y} \\ &\quad + \beta'\mathbf{X}'\Sigma^{-1}\mathbf{X}\beta \\ &= \mathbf{Y}'\Sigma^{-1}\mathbf{Y} - \mathbf{Y}'\Sigma^{-1}\mathbf{X}\beta - (\beta'\mathbf{X}'\Sigma^{-1}\mathbf{Y})' \\ &\quad + \beta'\mathbf{X}'\Sigma^{-1}\mathbf{X}\beta \\ &= \mathbf{Y}'\Sigma^{-1}\mathbf{Y} - \mathbf{Y}'\Sigma^{-1}\mathbf{X}\beta - \mathbf{Y}'\Sigma^{-1}\mathbf{X}\beta \\ &\quad + \beta'\mathbf{X}'\Sigma^{-1}\mathbf{X}\beta \\ &= \mathbf{Y}'\Sigma^{-1}\mathbf{Y} - 2\mathbf{Y}'\Sigma^{-1}\mathbf{X}\beta + \beta'\mathbf{X}'\Sigma^{-1}\mathbf{X}\beta. \end{aligned}$$

Derivando em relação a β e igualando a 0 tem-se que

$$\mathbf{0} - 2\mathbf{X}'\Sigma^{-1}\mathbf{Y} + 2\beta'\mathbf{X}'\Sigma^{-1}\mathbf{X} = \mathbf{0}$$

$$\beta \mathbf{X}' \Sigma^{-1} \mathbf{X} = \mathbf{X}' \Sigma^{-1} \mathbf{Y}$$

$$\widehat{\beta}^* = (\mathbf{X}' \Sigma^{-1} \mathbf{X})^{-1} \mathbf{X}' \Sigma^{-1} \mathbf{Y}.$$