TÓPICOS DE GEOMETRIA COM APLICAÇÕES EM ESTATÍSTICA

IDEMAURO ANTONIO RODRIGUES DE LARA
Licenciado em Matemática

Orientador: Prof. Dr. **ANTONIO FRANCISCO IEMMA**

Dissertação apresentada à Escola Superior de
Agricultura “Luiz de Queiroz”, Universidade de
São Paulo, para obtenção do título de Mestre em
Agronomia. Área de Concentração: Estatística e
Experimentação Agronômica.

PIRACICABA
Estado de São Paulo - Brasil
Novembro - 1998

Dados Internacionais de Catalogação na Publicação (CIP)
DIVISÃO DE BIBLIOTECA E DOCUMENTAÇÃO - ESALQ/USP

Lara, Idemauro Antonio Rodrigues de
Tópicos de geometria com aplicações em estatística / Idemauro Antonio Rodrigues de
Lara. -- ed. rev. -- Piracicaba, 1998.
115 p. : il.

Dissertação (Mestrado) -- Escola Superior de Agricultura Luiz de Queiroz, 1998.
Bibliografia.

1. Análise de regressão 2. Análise de variância 3. Estatística aplicada 4. Geometria
5. Hipótese estatística 6. Inferência estatística 7. Probabilidade I. Título

CDD 519.5

"Permitida a cópia total ou parcial deste documento, desde que citada a fonte - O autor"

*“Nenhuma história humana é escrita sem a presença, de uma ou duas
mãos amigas, que se estendem em nossa direção”.*

Agradecimentos:

A Deus, por tudo.

Ao professor Dr. Antonio Francisco Iemma pela orientação, pelo estímulo e principalmente pela confiança transmitida durante todo o curso.

A professora Dr^a. Ângela Maria C. J. Corrêa pelos primeiros passos em Estatística, pelas sugestões dadas neste trabalho e pela participação em dois momentos especiais de minha carreira: graduação e pós-graduação.

Aos professores do Departamento de Matemática e Estatística da ESALQ-USP, pelos ensinamentos e pela amizade.

Aos funcionários do Departamento de Matemática e Estatística da ESALQ-USP, especialmente às secretárias Luciane, Rosa e Solange pelo carinho e dedicação.

À CAPES, pela concessão da bolsa de estudos.

À minha família por compreender os momentos de ausência e impaciência.

Aos colegas da pós-graduação Adílson, Afrânio, Denise, Francisco, Lêda, Paulo, Sandra, Sílvia e Wilson por todos os bons momentos de convívio e pelas trocas de experiências.

Ao amigo de mestrado Cristián pela constante ajuda na parte gráfica.

À grande amiga Maria Helena pela contribuição através de sugestões e sobretudo pelo companheirismo e encorajamento nos momentos difíceis.

À Jeanete, infelizmente faltam-me palavras que possam exprimir de forma ímpar sua contribuição para este trabalho, seu estímulo, seu companheirismo ... sua capacidade mágica de trazer felicidade.

SUMÁRIO

	Página
LISTA DE FIGURAS	vi
LISTA DE TABELAS.....	viii
RESUMO	ix
SUMMARY	x
1 INTRODUÇÃO.....	1
2 REVISÃO DE LITERATURA	4
3 METODOLOGIA.....	19
3.1 Vetores e operações com vetores	19
3.2 Dependência e independência linear	23
3.3 Espaços vetoriais e operações com espaços vetoriais	26
3.4 Projeções ortogonais.....	30
3.5 Autovalores, autovetores e a decomposição por valores singulares	35
3.6 Inversas generalizadas	37
3.7 Algumas idéias sobre modelos lineares.....	39
3.7.1 Equações normais e aproximação por mínimos quadrados	40
3.7.2 Projetores ortogonais e o modelo linear	42
4 RESULTADOS E DISCUSSÃO	44
4.1 Algumas estatísticas descritivas	44
4.1.1 A média e a variância amostrais	44
4.1.1.1 Exemplo	47
4.1.2 O coeficiente de correlação linear simples e a covariância amostral	49
4.1.2.1 Exemplo	50
4.2 A probabilidade na óptica da geometria.....	52
4.3 Tópicos de inferência estatística.....	63
4.3.1 Estimação de mínimos quadrados	63

4.3.2	Estimador de máxima verossimilhança para μ	64
4.3.3	A lógica geométrica dos testes de hipóteses.....	67
4.3.3.1	Teste de significância para a média.....	68
4.3.3.2	Teste de significância para a existência de correlação linear.....	73
4.3.4	Intervalo de confiança para a média.....	74
4.4	Análise da variância.....	75
4.4.1	Exemplo.....	76
4.4.2	Projetores ortogonais associados a hipóteses de interesse.....	83
4.4.2.1	Exemplo.....	84
4.5	Regressão linear.....	87
4.5.1	Uma visão geométrica da regressão.....	87
4.5.2	Análise da regressão.....	89
4.5.3	Teste da hipótese de exclusão de variável.....	90
4.5.4	Exemplo.....	92
4.5.5	Intervalo de confiança para o coeficiente de regressão linear.....	97
4.5.6	O modelo de regressão através da D.V.S.....	98
4.5.7	Multicolinearidade.....	101
4.5.8	Seleção de variáveis.....	106
5	CONCLUSÕES.....	111
	REFERÊNCIAS BIBLIOGRÁFICAS.....	112

LISTA DE FIGURAS

	Página
1 Decomposição ortogonal do vetor y	15
2 Visão geométrica de um vetor em três dimensões	20
3 Adição de vetores, subtração de vetores e multiplicação por escalar.....	21
4 O método de Gram-Schmidt no $\mathbb{R}^3$	25
5 Projeção ortogonal em 2 dimensões: x projetado em y	30
6 Projeção ortogonal em 3 dimensões	34
7 A média amostral como uma projeção ortogonal.....	45
8 Visão geométrica da moda e da mediana em caso de assimetria	46
9 Interpretação geométrica da média amostral	48
10 Coeficiente de correlação linear simples	49
11 Uma visão geométrica das variáveis aleatórias no espaço tridimensional	53
12 A esperança matemática como uma projeção ortogonal	58
13 Variáveis aleatórias centradas e o coeficiente de correlação.....	59
14 Variáveis aleatórias centradas e a Lei dos Cossenos	60
15 Esperança condicional	61
16 Contornos da função densidade de probabilidade, centrados no verdadeiro vetor do modelo, $\mu = [\mu, \mu, \dots, \mu]'$	65
17 O estimador de máxima verossimilhança para μ	66
18 Idéia básica sobre teste de hipóteses.....	67
19 Partição ortogonal do vetor $y = [2.1, 1.8, 2.1]'$	69
20 Teorema de Pitágoras aplicado a $y = \langle y, u_1 \rangle u_1 + \langle y, u_2 \rangle u_2 + \langle y, u_3 \rangle u_3$	70
21 Distribuições das projeções $Y.u_1$	70
22 Distribuição da projeção $Y.u_1$ para as hipóteses de nulidade e alternativa	71
23 Análise da variância: decomposição ortogonal do vetor y	76

24	Projeção ortogonal do vetor y no espaço coluna de X : visão no 1º octante.....	78
25	Projeção ortogonal do vetor y sobre o espaço estimação $C(X)$	88
26	Visão geométrica de uma superfície de regressão.....	88
27	Teste da hipótese de exclusão de variável.....	92
28	Decomposição ortogonal do vetor de observações para o caso de regressão linear simples	94
29	Forma simplificada da decomposição ortogonal na regressão linear simples	95
30	Um exemplo de multicolinearidade perfeita	102
31	Regressão com duas variáveis não correlacionadas: uma visão da partição do coeficiente de determinação: $R^2 = r_1^2 + r_2^2$	107
32	Coefficientes padronizados de regressão com duas variáveis não ortogonais	108
33	Uma idéia geométrica do procedimento <i>stepwise</i> na seleção de variáveis.....	109

LISTA DE TABELAS

	Página
1 Valores fictícios de duas características	50
2 Distribuição de probabilidade conjunta de X e Y	56
3 Dados de um experimento inteiramente casualizado	77
4 Quadro da análise da variância.....	80
5 Esperanças das somas de quadrados e hipóteses naturais	82
6 Dados, em Kg, de um experimento fatorial 2×3	84
7 Alturas, em cm, de pés de feijão.....	92
8 Quadro da análise da variância para a regressão linear simples.....	96
9 Valores fictícios de três variáveis.....	103

TÓPICOS DE GEOMETRIA COM APLICAÇÕES EM ESTATÍSTICA

Autor: Idemauro Antonio Rodrigues de Lara
Orientador: Prof. Dr. Antonio Francisco Iemma

RESUMO

Esta dissertação procura mostrar como a Geometria pode ser usada para auxiliar na compreensão de muitos conceitos de Estatística e Probabilidade. A metodologia está centrada nos conceitos básicos da geometria vetorial elementar. Como exemplos de aplicações do procedimento geométrico, são considerados os seguintes assuntos: estatística descritiva, teoria da probabilidade, tópicos de inferência, análise da variância e análise da regressão. Os dois primeiros temas são desenvolvidos com base na analogia que existe entre muitos conceitos de Geometria, Estatística e Probabilidade. A análise da variância e da regressão baseiam-se na teoria das projeções ortogonais e no Teorema de Pitágoras. Sobre a regressão, são também apresentados alguns aspectos geométricos sobre testes de hipóteses, multicolinearidade, descrição do modelo pela decomposição por valores singulares e seleção de variáveis. Verifica-se que, além de servir como um importante recurso didático, a Geometria possui uma linguagem simples e indutiva, permitindo concretizar linhas rigorosas de raciocínio.

TOPICS OF GEOMETRY WITH APPLICATIONS IN STATISTICS

Author: Idemauro Antonio Rodrigues de Lara
Adviser: Prof. Dr. Antonio Francisco Iemma

SUMMARY

This work is intended for showing how the Geometry may be used in helping the comprehension of many concepts of Statistics and Probability. The methodology is centered in basic concepts of the elementary vectorial geometry. As examples of applications of the geometric procedure, the following issues are considered: descriptive statistics, probability theory, inference topics, analysis of variance and regression analysis. The first two themes are developed based on the existent analogies among many concepts of Geometry, Statistics and Probability. The analysis of variance and regression analysis are based on the theory of orthogonal projections and on the Pitagoras Theorem. With regard to regression, some geometrics aspects are also presented related to hypotheses testing, multicollinearity, model description using singular value decomposition and variables selection. Beyond the important didactical resource, it is verified that Geometry has a simple and inductive language allowing to make real rigorous lines of reasoning.

1 INTRODUÇÃO

A partir da década de 80 surge no meio estatístico uma crescente preocupação na valorização da abordagem geométrica como um importante recurso didático, apesar de tal preocupação ser observada em textos clássicos mais antigos. Isto se deve, sem dúvida, ao fato de que a grande maioria dos métodos estatísticos pode ser desenvolvida através de conceitos geométricos. No entanto esse fato não é, em geral, apresentado em cursos de Estatística, basicamente por dois motivos: o tradicionalismo do método algébrico e a dificuldade de se fazer associações com outras áreas do saber, em particular com a Geometria. Além disso, nesses cursos geralmente existem duas tendências distintas: ensinar os conceitos de forma superficial ou apresentá-los de forma substancial e complexa. Então, como uma alternativa para preencher a lacuna existente entre os dois extremos, a literatura tem sugerido o procedimento geométrico como uma forma de transmitir o conhecimento de princípios matemáticos fundamentais em um nível elementar e que seja convincente na explicação de conceitos estatísticos.

Há muitas outras áreas da matemática aplicada, como por exemplo o estudo da óptica na Física, nas quais o procedimento geométrico simplifica sobremaneira tanto a apresentação como a compreensão de um determinado conceito ou desenvolvimento. *“Isto está se tornando cada vez mais evidente, tanto que muitos matemáticos do século XX sentem que talvez a melhor forma de descrever a geometria, hoje, seja não como um corpo de conhecimentos, algo em separado e determinado, mas sim como um ponto de vista: uma maneira particular de observar o assunto”*. (Eves, 1992, p.28)

“Além da linguagem da geometria ser freqüentemente mais simples que as linguagens da álgebra e da análise, muitas vezes torna-se possível concretizar linhas de raciocínio rigorosas em termos geométricos, sem traduzi-las para a álgebra e a análise. Disso resulta uma economia considerável, tanto de reflexões como de comunicações de reflexões. Outro aspecto e talvez um dos mais importantes, é que as imagens geométricas sugeridas freqüentemente levam a resultados e estudos adicionais, transformando-se num instrumento poderoso de raciocínio indutivo e criativo”. (ibid)

“Grande parte da análise moderna tornou-se significativamente compacta e unificada através do emprego da linguagem e das imagens geométricas. Parece não haver dúvidas que tal procedimento terá reflexos nos cursos elementares de análise e, com isso, os atuais textos de Cálculo e de Estatística, supervolumosos, tenderão a tornar-se mais exíguos e mais compreensíveis para os alunos, graças ao uso do ponto de vista geométrico”. (ibid, p.29)

Apesar da relevância do tema, tanto do ponto de vista didático, quanto da força incontestável do relacionamento que associa a Estatística e a Geometria, a revisão de literatura revelou que sua divulgação ao longo do tempo tem sido feita de modo pouco consistente, através de artigos de excelente qualidade, mas de limitada penetração entre os usuários potenciais. Nesse contexto, tem-se como objetivos desse estudo:

- redigir um texto que possa consolidar o papel da Geometria no processo ensino-aprendizagem da Estatística, de modo a contribuir para que o movimento atual de valorização da geometria como recurso didático não seja apenas uma moda passageira;
- fornecer subsídios para interpretação geométrica de conceitos de Estatística Descritiva, Probabilidade, Análise da Variância e da Regressão e Inferência;

- utilizar a geometria para elucidar conceitos considerados complexos em cursos iniciais de Estatística como, por exemplo, a colinearidade nos modelos de regressão, os testes de hipóteses, entre outros;
- incentivar e motivar professores e alunos com vistas a estudos correlatados.

2 REVISÃO DE LITERATURA

O uso de conceitos geométricos em Estatística não é tão recente. Há alguns artigos, como o de Fisher (1915), que datam do começo deste século. No entanto, esses primeiros artigos limitam-se a pequenas interpretações e, na maior parte, usam uma geometria sofisticada, o que torna difícil a sua compreensão. Nesse contexto histórico, alguns textos tornaram-se clássicos como os de Scheffé (1959) e Margolis (1979), pela preocupação em valorizar a abordagem geométrica. Pode-se dizer que esses textos foram os pioneiros de uma série de muitos outros, que principalmente a partir da década de 80 valorizaram a geometria como uma técnica explanatória útil para apresentação e interpretação de conceitos em Estatística.

Fisher (1915) estuda a distribuição do coeficiente de correlação de uma amostra de n pares de uma distribuição normal bivariada. Sua técnica tem sido classificada como geometria pura, na qual há muito pouco de uma natureza analítica entre a descrição verbal da geometria e a conclusão estabelecida como uma fórmula. Um exemplo típico e razoável, apresentado pelo autor, consiste em transformar as densidades normais $2n$ -variadas da amostra de pares (x_i, y_i) numa densidade em termos das quantidades $\bar{x}$, $\mu_1 = \left[n^{-1} \sum_{i=1}^n (x_i - \bar{x})^2 \right]$, $\bar{y}$, $\mu_2 = \left[n^{-1} \sum_{i=1}^n (y_i - \bar{y})^2 \right]$ e do coeficiente de correlação amostral r .

“Um elemento do volume neste espaço n -dimensional pode então, sem dificuldades, ser especificado em termos de $\bar{x}$ e μ_1 , para $\bar{x}$ e μ_1 dados; P , o ponto $(x_1, \dots, x_n)$, deve ser uma esfera em $n - 1$ dimensões, situado em ângulos direitos à linha que passa pela origem na direção do vetor $(1, 1, \dots, 1)$, e o elemento do volume é: $C\mu_1^{n-2}d\mu_1d\bar{x}$, onde C é uma constante que não precisa ser determinada” (Fisher, 1915, p.509).

Esse exemplo sugere a robustez da geometria utilizada pelo autor. Herr (1980) comenta que a geometria usada por Fisher (1915) nesse artigo é muito elegante, mas é difícil de ser visualizada e, portanto, pode ser compreensível para poucos talentosos mas não o é para a maioria.

Bartlett¹ (1933/1934), citado por Herr (1980), mostra as implicações e vantagens em se pensar numa amostra de tamanho n como um vetor n -dimensional. Sua técnica pode ser classificada como geometria analítica, na qual cada idéia geométrica é representada pelas fórmulas analíticas. Esse modo iterativo de exposição permite ao leitor uma profunda compreensão tanto da visão geométrica como da analítica. Um exemplo é apresentado e discutido pelo autor e relata a análise do vetor linha, S , das observações a partir de um delineamento em Quadrado Latino:

Tem-se uma classificação nas linhas, colunas e tratamentos, descrita por:

$$S = R + C + T + E + M, \quad (2.1)$$

onde $R = (\bar{x}_r - \bar{x})$ é o vetor representando as diferenças entre médias de linhas e média geral, $C = (\bar{x}_c - \bar{x})$ similarmente para as colunas, $T = (\bar{x}_t - \bar{x})$ para tratamentos, $M = (\bar{x})$ é a média geral e $E = (\bar{x} - \bar{x}_r - \bar{x}_c - \bar{x}_t + 2\bar{x})$ é o termo residual. A partir da relação algébrica: $RC' = RT' = \dots = EM' = 0$, tem-se, analogamente para (2.1):

$$S^2 = R^2 + C^2 + T^2 + E^2 + M^2.$$

¹ BARTLETT, M. S. The Vector Representation of a Sample. **Proceedings of the Cambridge Philosophical Society**, vol. 30: p. 327-340, 1933/1934.

O autor explica que as relações algébricas significam que os vetores correspondentes são ortogonais e que S^2 denota o quadrado do comprimento de S .

Durbin & Kendall (1951) em seus estudos sobre a geometria de estimação tentam reverter a aproximação geométrica pura usada por Fisher (1915), embora sem sucesso. Para se ter uma idéia, um dos exemplos apresentados pelos autores e citado em Herr (1980), consiste na obtenção da variância mínima e do estimador não viesado da média comum para uma amostra $x_1, \dots, x_n$, de variáveis aleatórias independentes e identicamente distribuídas com variância comum σ^2 . O estimador é da forma $\sum \lambda_i x_i$, com a restrição não viesada $\sum \lambda_i = 1$.

Os autores consideram um espaço euclidiano com coordenadas $\lambda_1, \dots, \lambda_n$, chamado de espaço do estimador. O hiperplano $\sum \lambda_i = 1$ corresponde ao limite dos valores de λ fornecendo estimadores não viesados. Ademais, qualquer ponto P do hiperplano determina apenas um estimador. Nesse contexto, a variância do estimador é $\sigma^2 \sum \lambda_j^2 = \sigma^2 OP^2$, onde O é a origem. Mostram, ainda, que isso é um mínimo quando P é a base da perpendicular a partir de O sobre o hiperplano. Só a simetria é suficiente para mostrar que os valores dos λ 's são, então, todos iguais.

Como se pode notar, a geometria utilizada pelos autores é tão sofisticada, quanto aquela apresentada por Fisher (1915) e, novamente, muito bonita na teoria, mas de difícil compreensão e, portanto, de pouca aplicabilidade.

Scheffé (1959) aborda em seu livro alguns aspectos geométricos da análise da variância, em particular, a melhor aproximação para o vetor y , obtida pelo método dos mínimos quadrados, que é dada por sua projeção ortogonal, bem como as somas de quadrados necessárias para a obtenção da estatística F . O referido texto traz em seu apêndice conceitos básicos da geometria vetorial n -dimensional, constituindo-se dessa maneira em uma ferramenta essencial para a compreensão do elo: Estatística e Geometria.

Kruskall² (1961), citado por Herr (1980), apresenta um artigo que tem como propósito básico descrever a aproximação da coordenada livre para a estimação de Gauss-Markov (Mínimos quadrados lineares). Essa aproximação é, como a de Bartlett (1933/1934), uma técnica da geometria analítica. O autor enfatiza nesse artigo que a aproximação da coordenada livre para a estimação de Gauss-Markov, não tem sido freqüentemente discutida na literatura sobre mínimos quadrados e análise da variância. Além disso, ele indica que há duas motivações maiores para enfatizar a aproximação da coordenada livre:

- permite um tratamento mais simples, mais elegante e mais direto da Teoria Geral de estimação linear do que suas competidoras notacionais: a matriz e as aproximações escalares;
- é útil como uma introdução aos espaços de dimensões infinitas, que são importantes, por exemplo, no estudo de processos estocásticos.

Zyskind (1967) estabelece condições sob as quais estimadores de mínimos quadrados amostrais são também melhores estimadores não viesados lineares e refere-se a r autovetores ortogonais da matriz de variâncias e covariâncias, formando uma base para o espaço coluna da matriz do delineamento. Na verdade, esse trabalho não apresenta grande ênfase em geometria e pode ser visto como um exemplo da aproximação geométrica analítica.

Watson (1967), faz uso extensivo e efetivo da decomposição espectral da matriz de variâncias e covariâncias para estudar os vetores de erro em regressão de mínimos quadrados. O que o autor pensa geometricamente está ilustrado pela solução da regressão de mínimos quadrados, onde uma perpendicular é extraída de um ponto no

² KRUSKAL, W. H. The Coordinate-free Approach to Gauss-Markov estimation and its application to missing and extra observations. 4^o Symposium on Mathematical, Statistics and Probability, vol 1: p. 435-451, 1961.

espaço n -dimensional com posição y sobre espaço da regressão, a base dessa perpendicular é Xb , onde em b estão as estimativas de mínimos quadrados de β .

Margolis (1979) publica um artigo onde cita vários exemplos do uso da geometria vetorial para derivação de resultados estatísticos elementares e defende a geometria como recurso didático. Seu texto acrescenta idéias ilustrativas não somente na análise de regressão, mas também em muitos conceitos básicos tais como médias, variâncias e testes de hipóteses, com interpretações de projeções ortogonais e propriedades de distância mínima. Para isso, o autor utiliza dois modelos lineares:

a) Com o modelo linear mais simples possível $y = \alpha J + e$, onde y representa um vetor de uma amostra de n observações da variável Y , α representa um parâmetro desconhecido, J é um vetor de “uns” e e um vetor não observável de erros aleatórios, o autor mostra que $\hat{\alpha} = \bar{y}$ é a projeção ortogonal do vetor y sobre o vetor J do espaço n -dimensional. A variância amostral é descrita como a distância quadrática média entre os vetores y e $\hat{y} = \bar{y}$;

b) Através do modelo de regressão linear simples $y = \alpha J + \beta X + e$, onde y , J , α e e são definidos como no modelo anterior, X é o vetor de variáveis regressoras e β o vetor dos parâmetros correspondentes, o autor descreve a estimação de mínimos quadrados geometricamente, ilustrando graficamente que as estimativas $\hat{\alpha}$ e $\hat{\beta}$ são, na realidade, os coeficientes da combinação linear do vetor dependente $\hat{y} = \hat{\alpha}J + \hat{\beta}X$. Margolis (1979) aborda ainda, embora sucintamente, o problema da colinearidade por meio do ângulo de proximidade entre os vetores J e X . O autor finaliza seu artigo apresentando um teste de hipóteses a respeito do coeficiente de inclinação do modelo, descrevendo-o geometricamente por meio de duas projeções ortogonais. Ademais, obtém a estatística F escrita em termos dessas projeções, tal qual Scheffé (1959).

Herr (1980) estabelece um paralelo entre os conceitos algébrico e geométrico da estimação de mínimos quadrados. O autor discute, com base no modelo linear $y = X\beta + e$, o problema de encontrar o mínimo local da função:

$$Q = \sum_{i=1}^n (y_i - \sum_{j=1}^k x_{ij}\beta_j)^2,$$

e salienta o fato de que a colocação da derivada $\frac{\delta Q}{\delta \beta}$ igual a zero não garante por ela mesma a existência da solução, ou que a solução, caso exista, produza o extremo desejado. O autor mostra que a função Q reescrita na forma:

$$Q = \|y - X\beta\|^2,$$

representa a distância quadrática de y ao subespaço de $\mathbb{R}^n$, gerado pelas colunas da matriz X . Minimizar Q corresponde, então, a encontrar o ponto desse subespaço mais próximo a y . A resposta é visualizada facilmente como a projeção ortogonal de y sobre esse subespaço. Sendo assim, se $X\hat{\beta}$ denota essa projeção, então $X\hat{\beta}$ é único e satisfaz:

$$\|y - X\hat{\beta}\|^2 \leq \|y - X\beta\|^2 \quad \forall \beta \in \mathbb{R}^k, \text{ onde } k \leq n.$$

O autor apresenta, ainda, dois exemplos adicionais salientando o fato de que a aproximação geométrica é mais simples e mais completa que a algébrica. No primeiro deles, mostra como encontrar a distribuição de $s^2 = (n-1)^{-1} \sum (X_i - \bar{X})^2$ de n variáveis aleatórias independentes e identicamente distribuídas (i.i.d.), tal que $X_i \sim N(\mu, \sigma^2)$. Para tanto, adota a forma $Y = \mu J + e$, onde Y é o vetor das variáveis aleatórias, J é um vetor de “uns” e e é um vetor de variáveis aleatórias i.i.d. $N(0, \sigma^2)$. Como $PY = \bar{X}J$ é a projeção ortogonal sobre o vetor J , decorre que $s^2 = (n-1)^{-1} \|(I - P)y\|^2$ e existe uma base ortonormal $\{g_1, \dots, g_{n-1}\}$ do complemento ortogonal de J de tal forma que:

$$\|(I - P)y\|^2 = (g_1'Y)^2 + \dots + (g_{n-1}'Y)^2,$$

onde $g'_i Y$ é uma combinação linear de variáveis aleatórias normais e independentes e em decorrência, $g'_i Y \cap N(\mu g'_i J, \sigma^2)$. Mas como $g'_i J = 0$ então $g'_i Y \cap N(0, \sigma^2)$. Dado que g_i é ortogonal a $g_j \quad \forall i \neq j$, os $g'_i Y$ são “não correlacionados” e desse modo independentes. Segue imediatamente que $(n-1)s^2 / \sigma^2$ é uma soma de $(n-1)$ variáveis aleatórias independentes com distribuições de qui-quadrados, cada uma com um grau de liberdade.

No outro exemplo, o autor cita a análise de um experimento com dois fatores desbalanceados, no qual é desejado, para o teste de hipóteses, que não exista diferença entre médias simples das caselas para o fator das colunas. Comenta que, se a hipótese da interação não é aceita, há duas somas de quadrados de linhas que podem ser utilizadas: uma supondo interação e a outra supondo que não há interação, sendo que esta última deve fornecer um teste mais poderoso que o anterior. Para provar, o autor explica que o parâmetro de não centralidade, presumida a não interação, é o comprimento da hipotenusa de um triângulo, enquanto que o outro parâmetro é o comprimento de um dos catetos do mesmo triângulo.

Em suma, os exemplos apresentados pelo autor têm o objetivo básico de evidenciar as vantagens da descrição geométrica. Nesse artigo, ele faz ainda, uma consideração sobre a história do uso da geometria no estudo dos modelos lineares, revisando artigos com inclinação geométrica como os de Fisher (1915) e Kruskal (1961), concluindo que *“a relativa impopularidade da geometria não é devida a uma falta de assimilação das pessoas, mas a uma combinação de inércia, pouca exposição e resistência à abstração”* (Herr, 1980, p.46).

Mandel (1982) analisa situações de colinearidade e quase colinearidade no modelo de regressão $y_1 = X_p \beta_1 + e_1$, a partir da Decomposição por Valores Singulares (D.V.S) da matriz X , procedimento correlato à regressão por componentes principais. O problema da multicolinearidade nos modelos de regressão é fundamentalmente o mau condicionamento da matriz X do modelo linear. Ele ocorre

quando as colunas, ou vetores, da matriz X são altamente correlacionados. O autor explica que esse mau condicionamento é refletido nos valores singulares. Assim, um valor singular nulo implica que existe uma relação linear perfeita entre pelo menos dois dos p vetores da matriz X . Analogamente, um valor singular próximo de zero implica numa forte relação linear entre pelos menos dois desses p vetores.

O autor também avalia a vantagem de se tomar um estimador viciado para β , desde que a variância da estimativa seja suficientemente menor que a dos estimadores não viesados obtidos pelo método de mínimos quadrados ordinários. O ponto interessante e de grande importância desse artigo é que ele fornece aspectos geométricos sobre a regressão por mínimos quadrados, a colinearidade e a própria técnica da D.V.S., procurando ilustrá-los principalmente com representação gráfica. Por outras palavras, o autor localiza a causa e apresenta uma possível solução através de boa base de interpretação geométrica.

Bryant (1984) enfatiza que não há fórmulas diferentes em Geometria, Estatística e Probabilidade, mas sim variações sobre um tema comum. Com esse propósito o autor demonstra claramente a simplicidade dos conceitos básicos da Geometria, expressos por vetores, norma, produto interno, ângulo e projeções e suas equivalências nas áreas de Estatística e Probabilidade.

Seguindo a mesma linha de pensamento de Margolis (1979), o autor mostra que as relações feitas entre os critérios de medidas de projeção e distância de um vetor a um outro vetor J de “uns” na direção da origem, como análogas ao cálculo de medidas de tendência central e dispersão. O ângulo θ , formado entre dois vetores, define uma medida de associação entre as variáveis e o cosseno do ângulo formado pelos seus componentes ortogonais ao vetor J é numericamente igual ao coeficiente de correlação amostral.

O autor comenta, ainda, que a interpretação de muitos conceitos, como por exemplo a regressão linear por mínimos quadrados, podem ser melhor explorados se as interpretações são dadas no campo da geometria: a projeção fornece o melhor

ajuste e o ângulo mede a adequação desse ajuste. A inovação desse artigo fica caracterizada pelo tratamento geométrico que o autor apresenta para a Teoria da Probabilidade. Segundo ele, isso é possível quando as variáveis aleatórias são interpretadas como vetores. Dessa forma, algumas das principais propriedades da Esperança Matemática podem ser reproduzidas por definições geométricas. É o caso, por exemplo, da esperança conjunta entre duas variáveis aleatórias, que pode ser expressa em função do produto interno. Centrando as variáveis aleatórias nas suas respectivas esperanças, alguns teoremas são reproduzidos pela simples e imediata aplicação da Lei dos Cossenos e do Teorema de Pitágoras.

Sem dúvida, esse artigo mostra de maneira muito elegante e simples a equivalência das noções de fundamentos básicos que são expressos em diferentes linguagens. O autor comenta que a partir do momento em que se visualizam tais variações, adquire-se habilidade para aplicar e compreender os princípios fundamentais das idéias estatísticas.

Schey (1985) relata uma descrição geométrica dos contrastes na análise da variância de modelos de classificação simples, justificando a condição de soma nula para os coeficientes dos contrastes nas combinações lineares das médias de tratamentos. O autor usa o fato de que a soma de quadrados devida aos parâmetros na análise da variância, traduzida geometricamente por $\|\hat{y}\|^2$, pode ser decomposta em duas parcelas: uma devida aos tratamentos e outra devida à correção.

A idéia desse artigo é decompor a soma de quadrados devida aos tratamentos em somas de quadrados relativas a contrastes ortogonais. O autor propõe que se use um vetor z de dimensão r , onde r é o número de tratamentos e cuja norma ao quadrado é numericamente igual a $\|\hat{y}\|^2$. Para tanto, define:

$$z' = [z_1, z_2, \dots, z_r] \quad \text{onde: } z_i = n_i^{1/2} \bar{y}_i,$$

de tal forma que:

$$\|z\|^2 = \sum_i n_i (\bar{y}_i - \bar{y})^2 + n\bar{y}^2 = \text{SQ tratamentos} + \text{Correção},$$

onde $\bar{y}$ é a média geral; $\bar{y}_i$ é a média inerente ao tratamento i e n_i é o número correspondente de observações.

Esse vetor é, então, decomposto em dois vetores: um relativo à constante e outro aos efeitos de tratamentos. Para determinar um vetor cuja norma ao quadrado corresponda a $n\bar{y}^2$, o autor define o vetor $\phi_0 = [(n_1/n)^{1/2}, \dots, (n_r/n)^{1/2}]$, pertencente a um espaço unidimensional. A projeção de z sobre o complemento ortogonal desse espaço ($\phi_0^\perp$) é feita através do projetor ortogonal $W = I_r - \phi_0\phi_0'$. Como W é simétrico e idempotente com $\text{Tr}(W) = r - 1 = \sum_i \lambda_i$, então W tem um autovalor igual a zero e $r - 1$ autovalores iguais a 1. Decorre que o autovetor associado a $\lambda = 0$ é o vetor ϕ_0 . Os demais autovetores $\phi_1, \phi_2, \dots, \phi_{r-1}$ estão associados aos autovalores iguais a 1 e são ortogonais a ϕ_0 . Desse modo, os autovetores de W formam uma base ortogonal em r dimensões e, conseqüentemente, z pode ser escrito como combinação linear desses vetores. Por outro lado, Wz surge como combinação linear dos autovetores associados aos autovalores iguais a 1, ou seja: $Wz = \sum_{j=1}^{r-1} L_j \phi_j$, onde os L_j são os coeficientes dessa combinação linear. Desse modo, é fácil justificar que a soma dos coeficientes de um contraste é zero, $\sum c_i = 0$, dada a ortogonalidade entre ϕ_j , ($j \neq 0$), e ϕ_0 , ou seja, pelo produto interno: $\phi_j' \phi_0 = 0$. Analogamente, pode-se mostrar a ortogonalidade entre dois contrastes. Assim, pode-se dizer que quando dois contrastes estimados L_j e L_k são ortogonais, dois vetores $L_j \phi_j$ e $L_k \phi_k$ são ortogonais.

Outra idéia interessante implícita nesse trabalho é o procedimento de desdobramento dos graus de liberdade de tratamentos, onde cada contraste ortogonal está associado a 1 grau de liberdade.

Saville & Wood (1986) utilizam elementos básicos da geometria n-dimensional vetorial para fazer análise da variância e análise da regressão de forma rigorosa, mas elementar. Os autores consideram basicamente dois objetos e dois processos, os quais são suficientes para as análises. Os objetos são:

a) O vetor de dados y do espaço n-dimensional: $y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$;

b) Um subespaço M do espaço n-dimensional, gerado por um conjunto de vetores ortonormais $\{u_1, u_2, \dots, u_k\}$, que por sua vez são determinados conforme o modelo assumido.

Uma vez determinado o espaço gerado pelas colunas da matriz do modelo, os dados são então usados para ajustar o modelo e testar hipóteses sobre seus parâmetros. Para isso, os autores utilizam dois processos:

a) O ajuste do modelo através da projeção de y sobre M , obtendo $\hat{y} = \langle y, u_1 \rangle u_1 + \dots + \langle y, u_k \rangle u_k$ e $\hat{e} = y - \hat{y}$;

b) Teste de hipóteses sobre os parâmetros através da aplicação do Teorema de Pitágoras para expressar $\|y\|^2$ como $\langle y, u_1 \rangle^2 + \dots + \langle y, u_n \rangle^2$, onde $\{u_1, u_2, \dots, u_n\}$ é um sistema de coordenadas ortogonais do espaço n-dimensional que inclui o conjunto de vetores $\{u_1, u_2, \dots, u_k\}$. Os testes estatísticos são produzidos comparando-se médias de conjuntos apropriados dessas normas ao quadrado. Para ilustrar o procedimento os autores utilizam delineamentos de experimentos inteiramente ao acaso, em blocos ao acaso e de regressão linear simples, embora como eles próprios salientam, o processo possa ser aplicado a ensaios mais complexos.

Saville & Wood (1991) baseiam-se no artigo publicado em 1986, e adotam a mesma técnica para fazer análise da variância de outros tipos de delineamentos, como, por exemplo, os quadrados latinos. Trata-se de um livro cujo

conteúdo utiliza a aproximação geométrica para os métodos estatísticos, constituindo-se assim em um material de grande relevância para os pesquisadores da referida área.

Lemma & Palm (1992) partem do modelo linear $y = X\theta + e$, onde:

y é o vetor de observações;

X é a matriz conhecida do delineamento;

θ é o vetor de parâmetros;

e é um vetor de variáveis aleatórias não observáveis, tal que $E[e] = \phi$ e $Var[e] = I\sigma^2$;

e descrevem a análise da variância como uma aplicação do Teorema de Pitágoras. Os autores usam o projetor ortogonal, $P = X(X'X)^G X' = XX^\ell = XX^+$, obtido a partir da matriz do delineamento, onde $(X'X)^G$, X^ℓ e X^+ denotam respectivamente uma inversa generalizada qualquer de $(X'X)$, uma inversa generalizada de mínimos quadrados de X e a inversa generalizada de Moore-Penrose de X . Os autores ilustram a decomposição do vetor y de observações:

$$y = \hat{y} + \hat{e}$$

onde $\hat{y} = Py$ e $\hat{e} = (I - P)y$, observando que $\hat{y} \in C(X)$ e $\hat{e} \in C^\perp(X)$, onde $C(X)$ é o espaço gerado pelas colunas de X e $C^\perp(X)$ seu complemento ortogonal.

Figura 1. Decomposição ortogonal do vetor y .

Dada a ortogonalidade, utilizam o Teorema de Pitágoras e obtêm a decomposição clássica da análise da variância:

$$\|y\|^2 = \|\hat{y}\|^2 + \|\hat{e}\|^2$$

onde:

$$\|y\|^2 = \langle y, y \rangle = y'y = \sum_{i=1}^n y_i^2 = \text{Soma de Quadrados Total não corrigida (SQ}_T\text{)}$$

$$\|\hat{y}\|^2 = \langle \hat{y}, \hat{y} \rangle = \hat{y}'\hat{y} = y'Py = \text{Soma de Quadrados devida aos Parâmetros (SQ}_0\text{)}$$

$$\|\hat{e}\|^2 = \langle \hat{e}, \hat{e} \rangle = \hat{e}'\hat{e} = y'(I - P)y = \text{Soma de Quadrados Residual (SQ}_R\text{)}$$

Os graus de liberdade associados a cada uma dessas somas de quadrados são, então, iguais às dimensões dos espaços vetoriais correspondentes e são dados pelo posto das matrizes de projeção. Os autores utilizam a simetria e a idempotência dos projetores ortogonais para obter as esperanças das somas de quadrados através do teorema clássico das formas quadráticas: $E[y'My] = \text{Tr}[M]\sigma^2 + \theta'X'MX\theta$ onde, dada a idempotência, $\text{Tr}[M] = r[M]$.

Johnson & Wichern (1992) reproduzem num livro sobre análise multivariada, alguns resultados da geometria amostral a partir da definição de uma matriz de dados $X_{(p \times n)}$, onde p representa o número de variáveis e n o número de observações. O texto procura elucidar as interpretações geométricas de conceitos básicos, tais como média, variância, covariância e coeficiente de correlação, como já mostrados em artigos anteriores como os de Margolis (1979) e Bryant (1984), entre outros.

Uma interpretação geométrica interessante é fornecida a partir da matriz de variâncias e covariâncias, S . Os autores assumem que o determinante $|S|$ é um número único denominado variância amostral generalizada, e mostram que:

$$|S| = (n-1)^{-p} (\text{volume})^2$$

isto é, a variância amostral generalizada para um conjunto fixo de dados, é proporcional ao quadrado do volume gerado pelos p vetores de desvios e_i , $i = 1, \dots, p$. Na verdade,

os autores buscam salientar algumas das muitas analogias existentes entre a Estatística e a Geometria. Além do mais, eles utilizam a geometria como uma importante ferramenta para a compreensão de outros temas como a estimação de mínimos quadrados para o modelo de regressão múltipla e o método da aproximação por componentes principais.

Lemma et al. (1993) apresentam um procedimento simples e imediato para se obter o projetor ortogonal associado a uma hipótese H_0 para qualquer nível de desbalanceamento dos dados.

Os autores baseiam-se na definição de Rao (1945) para uma função estimável no modelo linear já apresentado $y = X\theta + e$, “uma função linear paramétrica, $\lambda'\theta$, é dita estimável nesse modelo, se e somente se, existe ao menos uma combinação linear das observações, $a'y$, tal que $E[a'y] = \lambda'\theta$.”

Com isso, obtêm a soma de quadrados da hipótese através de:

$$SQH_0 = y'a(a'a)^{-1}a'y = y'P_{H_0}y$$

onde:

y é o vetor de observações;

a é o vetor de coeficientes da combinação linear dos componentes do vetor de observações y ;

$P_{H_0} = a(a'a)^{-1}a'$ é o projetor ortogonal associado à hipótese H_0 .

Além de simplificar consideravelmente a obtenção do projetor ortogonal associado a uma hipótese, os autores salientam a sua importância através do quociente

$\frac{SQH_0}{\sigma^2}$ que, sob normalidade, tem distribuição de χ^2 não central com graus de

liberdade dados pelo posto de a , e parâmetro de não centralidade

$\delta = \frac{1}{2\sigma^2}\theta'X'PX\theta = \frac{1}{2\sigma^2}\theta'\lambda(a'a)^{-1}\lambda'\theta$, explicando de modo objetivo a hipótese

testada.

Bring (1996) mostra a importância da geometria no estudo da análise de regressão. O artigo fornece uma introdução sumária do método geométrico e tem por objetivo principal “fornecer uma luz” sobre o problema de como comparar variáveis no modelo de regressão. Para isso, o autor propõe que sejam utilizadas variáveis padronizadas, na forma: $x_j = (x_j^* - \bar{x}) / s_{x_j^*} \sqrt{(n-1)}$ e $y_j = (y_j^* - \bar{y}) / s_{y_j^*} \sqrt{(n-1)}$, onde x^* e y^* são as variáveis originais. Com isso, os vetores ficam normalizados e consequentemente o coeficiente de determinação, R^2 , pode ser expresso como: $R^2 = \|\hat{y}\|^2 = \cos^2(\theta)$, onde θ é o ângulo agudo formado entre os vetores y e $\hat{y}$.

Desse modo, o autor procura elucidar que o tão usado método da partição do R^2 : quanto maior a magnitude da padronização de β_j mais x_j contribui para o valor predito y , pode ser válido se as variáveis são não correlacionadas. Nesse caso, os vetores x são ortogonais, e pelo Teorema de Pitágoras obtém a seguinte partição:

$$\|\hat{y}\|^2 = \beta_1^2 + \beta_2^2 + \dots + \beta_k^2$$

onde os coeficientes padronizados coincidem com os coeficientes de correlação, r_j , entre y e x_j . No entanto, o autor mostra que se as variáveis são correlacionadas, essa partição não está assegurada e, portanto, não existe justificativa para a contribuição dos β_j relativa ao plano de regressão. O autor ilustra geometricamente três modelos de partição do R^2 : a análise de regressão por etapas (*stepwise*), o *stepwise médio* e a medida do produto $\beta_j r_j$, procurando exaltar que não há uma definição única da importância do conceito e nenhuma resposta clara pode ser dada sobre que medida deve ser usada. O procedimento geométrico aumenta a compreensão do que elas estão medindo e, especialmente, indica situações onde as medidas não são apropriadas. Nesse sentido, o referido artigo mostra claramente que o papel da Geometria em Estatística, não se restringe apenas a interpretações de determinados conceitos, mas que serve também como uma importante ferramenta na compreensão e solução de problemas.

3 METODOLOGIA

O método proposto está centrado nos conceitos e resultados da Geometria elementar e na técnica da decomposição por valores singulares.

Neste contexto, apresentam-se a seguir, alguns conceitos básicos da Geometria, tais como linhas, planos, comprimentos, distâncias e ângulos, os quais podem ser expressos em termos de vetores e produto interno.

São apresentadas, também, algumas idéias sobre inversas generalizadas e modelos lineares, importantes no desenvolvimento deste trabalho.

3.1 Vetores e operações com vetores

Definição 3.1.1. Seja $\mathfrak{R}$ o conjunto dos números reais e seja m um número inteiro e positivo. Chama-se de espaço cartesiano m -dimensional, o conjunto de todas as possíveis m -uplas ordenadas $(x_1, x_2, \dots, x_m)$, tal que $x_i \in \mathfrak{R}$ para todo $i = 1, 2, \dots, m$, decorre, então, que $\mathfrak{R} \times \mathfrak{R} \times \dots \times \mathfrak{R} = \mathfrak{R}^m$. Nesse sentido, $x' = (x_1, x_2, \dots, x_m)$ é dito vetor m -dimensional, ou simplesmente vetor $m \times 1$.

Geometricamente, o vetor é um ponto do espaço cartesiano m -dimensional ou um segmento de reta que une esse ponto à origem do sistema cartesiano.

Figura 2. Visão geométrica de um vetor em três dimensões.

Definição 3.1.2. Dados dois vetores quaisquer x e y de um mesmo espaço cartesiano m -dimensional:

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} \quad \text{e} \quad y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix},$$

a adição e subtração entre ambos é definida por: $x \pm y = \begin{bmatrix} x_1 \pm y_1 \\ x_2 \pm y_2 \\ \vdots \\ x_m \pm y_m \end{bmatrix}$.

A adição entre vetores satisfaz as seguintes propriedades:

- a) Comutativa: $x + y = y + x$;
- b) Associativa: $(x + y) + z = x + (y + z)$;
- c) Elemento neutro: $x + \phi = \phi + x = x$;
- d) Inverso aditivo: $x + (-x) = (-x) + x = \phi$.

Definição 3.1.3. Se c é um escalar, define-se cx como o vetor obtido pela multiplicação de cada componente de x por c .

$$cx = \begin{bmatrix} cx_1 \\ cx_2 \\ \vdots \\ cx_m \end{bmatrix}.$$

Figura 3. Adição de vetores, subtração de vetores e multiplicação por escalar.

Geometricamente, como observa-se pela Figura 3, a adição e a subtração de vetores podem ser descritas através das diagonais de um paralelogramo.

Definição 3.1.4. Produto interno ou produto escalar entre dois vetores x e y de $\mathfrak{R}^m$, é definido como uma função que associa a cada par de vetores x e y um número real, denotado por $\langle x, y \rangle$, satisfazendo em relação aos vetores x, y e z e ao escalar c as propriedades:

- a) $\langle x, y \rangle = \langle y, x \rangle$ (comutativa ou simetria)
- b) $\langle x, y + z \rangle = \langle x, y \rangle + \langle x, z \rangle$ (distributiva ou linearidade)
- c) $\langle cx, y \rangle = c \langle x, y \rangle$ (homogeneidade)
- d) $\langle x, x \rangle \geq 0$; $\langle x, x \rangle = 0 \Leftrightarrow x = \phi$ (positividade)

A Definição 3.1.4 mostra que pode estar definido mais do que um produto interno num mesmo espaço cartesiano. Dentre os possíveis produtos internos que satisfazem as propriedades da Definição 3.1.4, destaca-se o produto interno usual, ou euclidiano:

$$\langle x, y \rangle = \sum_{i=1}^m x_i y_i ;$$

por ser o mais conhecido dentre todos e, também, porque é muito utilizado nesse trabalho.

Definição 3.1.5. Chama-se de norma euclidiana de um vetor x e denota-se por $\|x\|$, ao número real não negativo, definido em relação ao produto interno $\langle x, x \rangle$:

$$\|x\| = [\langle x, x \rangle]^{1/2} ;$$

satisfazendo os axiomas:

- a) $\|x\| > 0$ se $x \neq \phi$ (positividade)
- b) $\|x\| = 0$ se $x = \phi$ (nulidade)
- c) $\|cx\| = |c| \|x\|$ (homogeneidade)
- d) $\|x + y\| \leq \|x\| + \|y\|$ (desigualdade triangular)

Uma interpretação geométrica simples da norma euclidiana é que ela mede o comprimento de um vetor.

Definição 3.1.6. Sejam x e y dois vetores não nulos do $\Re^m$, o ângulo entre eles é definido por:

$$\cos \theta = \frac{\langle x, y \rangle}{\|x\| \|y\|} ; \quad -1 \leq \cos \theta \leq 1 \quad \Rightarrow \quad \theta = \arccos \left[\frac{\langle x, y \rangle}{\|x\| \|y\|} \right] ; \quad 0^\circ \leq \theta \leq 180^\circ .$$

Da Definição 3.1.6 decorre que:

$$\langle x, y \rangle = \|x\| \|y\| \cos \theta ;$$

que define uma forma alternativa de expressar o produto interno entre dois vetores.

Teorema 3.1.1. Uma condição necessária e suficiente para que dois vetores sejam ortogonais é que seu produto interno seja nulo.

A prova decorre imediatamente da Definição 3.1.4.

Observa-se também que a condição de nulidade de um dos vetores é apenas suficiente para que verifique o Teorema 3.1.1.

Definição 3.1.7. Diz-se que um vetor é normalizado e denota-se por $\tilde{x}$, se ele tem norma 1. Logicamente todo vetor não nulo pode ser normalizado, como se segue:

$$\tilde{x} = \frac{x}{\|x\|}.$$

Definição 3.1.8. Um conjunto de vetores do $\mathfrak{R}^m$ é dito ortonormal, se e somente se seus vetores são ortogonais dois a dois e todos têm norma 1.

3.2 Dependência e independência linear

Definição 3.2.1. Dado um conjunto de vetores $A = \{v_1, v_2, \dots, v_n\}$, onde $v_i \in \mathfrak{R}^m$, diz-se que A é linearmente dependente (L.D.) se, dado um conjunto de escalares $\alpha_1, \alpha_2, \dots, \alpha_n$, existe ao menos um não nulo tal que $\sum_{i=1}^n \alpha_i v_i = \phi$. Caso contrário o conjunto é dito linearmente independente (L.I.).

A Definição 3.2.1, estabelece a seguinte relação: $\sum \alpha_i v_i = \phi \Rightarrow {}_m A_n \alpha_1 = \phi$, onde ${}_m A_n \alpha_1 = \phi$ é um sistema homogêneo de equações lineares, na forma matricial. O sistema homogêneo é sempre consistente e nessas condições:

- a) $r[A] = n$, o sistema é determinado e o conjunto é L.I.;
 b) $r[A] < n$, o sistema é indeterminado e o conjunto é L.D.

Um algoritmo prático: Seja ${}_m A_n$ a matriz de posto k , que descreve esses n vetores de m componentes reais e, seja ${}_m B_n$ a matriz de sua forma escalonada canônica:

- a) o posto da matriz A é igual ao número de colunas L.I. em B ;
 b) as colunas L.I. de B são correspondentes às colunas L.I. de A ;
 c) as combinações lineares de A estão descritas nas colunas de B .

Teorema 3.2.1. Se o vetor nulo está incluso num conjunto de vetores então esse conjunto é linearmente dependente.

$$\text{Prova: } \sum_{i=1}^n \alpha_i v_i = \phi \Rightarrow \alpha_1 v_1 + \alpha_2 v_2 + \dots + \alpha_p \phi + \dots + \alpha_n v_n = \phi.$$

Decorre que $\alpha_p \phi = \phi$, $\forall \alpha_p$. Portanto, sem dúvida, existe $\alpha_p \neq 0$,

tal que $\sum_{i=1}^n \alpha_i v_i = \phi$, assim, o conjunto é L.D.

Teorema 3.2.2. Dado um conjunto $A = \{v_1, v_2, \dots, v_n\}$, de n vetores de m componentes reais, dos quais $k \leq n$ são linearmente independentes, então é possível descrevê-los através de:

- k vetores ortogonais $\{y_1, y_2, \dots, y_k\}$;
- k vetores ortonormais $\{z_1, z_2, \dots, z_k\}$, onde $z_i = \frac{y_i}{\|y_i\|}$.

Registra-se adicionalmente que esse conjunto de k vetores ortogonais pode ser obtido através do *algoritmo de Gram-Schmidt*:

Se v e y são dois vetores de um espaço cartesiano, com $y \neq \phi$, o vetor $\frac{\langle y, v \rangle}{\langle y, y \rangle} y$ diz-se a projeção de v sobre y . No método de *Gram-Schmidt*, determina-se o elemento y_{r+1} subtraindo-se de v_{r+1} a projeção de v_{r+1} sobre cada um dos elementos anteriores $y_1, \dots, y_r$. Sendo assim:

$$\begin{aligned} y_1 &= v_1 \\ y_2 &= v_2 - \frac{\langle y_1, v_2 \rangle}{\langle y_1, y_1 \rangle} y_1 \\ &\vdots \\ y_k &= v_k - \frac{\langle y_1, v_k \rangle}{\langle y_1, y_1 \rangle} y_1 - \frac{\langle y_2, v_k \rangle}{\langle y_2, y_2 \rangle} y_2 - \dots - \frac{\langle y_{k-1}, v_k \rangle}{\langle y_{k-1}, y_{k-1} \rangle} y_{k-1}. \end{aligned}$$

A Figura 4 mostra essa construção geométrica no espaço cartesiano tridimensional. Um conjunto ortogonal $\{y_1, y_2, y_3\}$ foi construído a partir de um conjunto L.I. dado $\{v_1, v_2, v_3\}$.

Figura 4. O método de Gram-Schmidt no $\mathbb{R}^3$.

3.3 Espaços vetoriais e operações com espaços vetoriais

Definição 3.3.1. Um conjunto V^m de vetores de m componentes reais, fechado em relação às operações de adição de vetores e a multiplicação de vetor por escalar, é dito espaço vetorial real.

Considerando os objetivos deste trabalho, tem-se interesse nos espaços vetoriais munidos de produto interno e portanto de norma euclidiana, também conhecidos como espaços métricos euclidianos.

Teorema 3.3.1. O conjunto $\mathfrak{R}^m$ de todos os possíveis vetores de m componentes reais é um espaço vetorial real.

Prova: Sabe-se que o conjunto dos números reais é fechado em relação às operações de adição e multiplicação por escalar. É imediato verificar, então, que a soma de quaisquer dois vetores do $\mathfrak{R}^m$, reproduz um vetor do $\mathfrak{R}^m$ e, o produto de um escalar real por um vetor do $\mathfrak{R}^m$, reproduz um vetor do $\mathfrak{R}^m$. De fato:

$$\mathfrak{R}^m = \{v / v = (x_1, x_2, \dots, x_m); x_i \in \mathfrak{R} \ \forall i\},$$

- a) $v_i + v_j = v_k \in \mathfrak{R}^m, \ \forall \ v_i \text{ e } v_j \in \mathfrak{R}^m;$
- b) $\alpha v_i = v_k \in \mathfrak{R}^m, \ \forall \ v_i \in \mathfrak{R}^m \text{ e } \alpha \in \mathfrak{R}.$

Definição 3.3.2. Se S^m é um subconjunto do $\mathfrak{R}^m$, fechado em relação às operações de adição de vetores e multiplicação por escalar, então S^m é um espaço vetorial real e por estar contido em $\mathfrak{R}^m$ é referido como subespaço vetorial e denota-se $S^m \subset \mathfrak{R}^m$.

Definição 3.3.3. Sejam V^m um espaço vetorial real e $G = \{v_1, v_2, \dots, v_p\}$ um subconjunto de V^m ($p \geq m$). Se qualquer vetor de V^m pode ser escrito como combinação linear desses p vetores, diz-se que eles formam um conjunto

de geradores de V^m . Além disso, se B é um subconjunto de G e tem $k \leq p$ vetores linearmente independentes e, ainda gera V^m , então B é uma *base* de V^m .

Definição 3.3.4. Considere V^m um espaço vetorial real. Define-se dimensão de V^m e denota-se $\text{Dim}(V^m)$, ao número de vetores de qualquer uma de suas bases. Se $V^m = \phi$, então $\text{Dim}(V^m) = 0$. Um espaço vetorial é dito de dimensão finita, se ele possui uma base finita.

Neste trabalho, os espaços vetoriais considerados têm dimensão finita.

Teorema 3.3.2. Seja V^m um espaço vetorial real e seja $B = \{u_1, u_2, \dots, u_k\}$ uma de suas bases. Decorre que qualquer vetor $v \in V^m$ pode ser obtido como combinação linear de B : $v = \sum_{i=1}^k \alpha_i u_i$, onde $\alpha \in \mathbb{R}$.

Definição 3.3.5. Seja ${}_m A_n = [u_1 \mid u_2 \mid \dots \mid u_n]$ uma matriz cujas colunas são vetores $m \times 1$. Se as colunas de A formam um conjunto de geradores de V^m , diz-se que V^m é o espaço gerado pelas colunas de A e denota-se $V^m = C(A)$. Analogamente se ${}_m A_n$ é uma matriz, cujas linhas são vetores transpostos $1 \times n$, e as linhas de A formam um conjunto de geradores de V^n , então diz-se que V^n é o espaço gerado pelas linhas de A e denota-se $V^n = L(A)$.

Da Definição 3.3.5, é imediato verificar que $C(A) \in \mathbb{R}^m$, $L(A) \in \mathbb{R}^n$ e $L(A) = C(A')$.

Definição 3.3.6. A intersecção de S_1 com S_2 , onde S_1 e S_2 são dois subespaços vetoriais de $\mathbb{R}^m$, é o subconjunto formado por todos os vetores que pertencem simultaneamente a S_1 e S_2 :

$$S_1 \cap S_2 = S \quad \text{onde} \quad S = \{y: y \in S_1 \wedge S_2\} \text{ é também um espaço vetorial.}$$

Definição 3.3.7. Sejam S_1 e S_2 dois subespaços de $\mathfrak{R}^m$. Define-se:

a) Soma direta entre eles e denota-se $S_1 \oplus S_2$, ao subespaço vetorial $W \subseteq \mathfrak{R}^m : S_1 \oplus S_2 = W$ tal que:

$$W = \{w: w = s_1 + s_2, s_1 \in S_1 \text{ e } s_2 \in S_2 \text{ com } S_1 \cap S_2 = \emptyset\};$$

b) Soma: $S_1 + S_2 = W \subseteq \mathfrak{R}^m$, tal que:

$$W = \{w: w = s_1 + s_2, s_1 \in S_1 \text{ e } s_2 \in S_2\}.$$

Definição 3.3.8. Dois subespaços quaisquer S_1 e S_2 de $\mathfrak{R}^m$ são ditos ortogonais ($S_1 \perp S_2$), se e somente se para todos os vetores $s_1 \in S_1$ e $s_2 \in S_2$ se verifique $\langle s_1, s_2 \rangle = 0$.

Definição 3.3.9. Considere S_1 e S_2 dois subespaços vetoriais de $S \subseteq \mathfrak{R}^m$, tais que: $S_1 \perp S_2$ e $S_1 \oplus S_2 = S$, diz-se, então, que S_2 é o complemento ortogonal de S_1 em relação a S e vice-versa. Sendo assim: $S_2 = S_1^\perp$ e equivalentemente $S_1 = S_2^\perp$.

No contexto da Definição 3.3.9, diz-se também que a soma direta entre S_1 e S_2 é uma soma direta ortogonal: $S_1 \oplus^\perp S_2$.

Teorema 3.3.3. Dados S_1 e S_2 dois subespaços vetoriais de $\mathfrak{R}^m$, sejam $V = S_1 \cap S_2$, $U = S_1 + S_2$, $W = S_1 \oplus S_2$ e $T = S_1 \oplus^\perp S_2$. Então:

- a) $\text{Dim}(S_1) + \text{Dim}(S_2) = \text{Dim}(V) + \text{Dim}(U)$;
- b) $\text{Dim}(S_1) + \text{Dim}(S_2) = \text{Dim}(W)$, pois $\text{Dim}(V) = \text{Dim}(\emptyset) = 0$;
- c) $\text{Dim}(S_1^\perp) + \text{Dim}(S_1) = \text{Dim}(T)$.

Teorema 3.3.4. Se S_1 é um subespaço vetorial de $S \subseteq \mathfrak{R}^m$, então seu complemento ortogonal existe e é único. Ademais, se $S_1 = S$, então $S_1^\perp = \{\emptyset\}$.

Teorema 3.3.5. Todo subespaço $S \subseteq \mathfrak{R}^m$ pode ser escrito como uma soma direta ortogonal entre um de seus subespaços S_1 e seu respectivo complemento ortogonal.

A prova deste Teorema é imediata, usando-se o Teorema 3.3.4 e a Definição 3.3.9:

$$S_1 \oplus S_1^\perp = S_1 \oplus S_1^\perp = S_1 + S_1^\perp = S \subseteq \mathfrak{R}^m.$$

Teorema 3.3.6. Seja S_1 um subespaço de $S \subseteq \mathfrak{R}^m$. Qualquer vetor $x \in S$ pode ser escrito de modo único como uma adição entre dois vetores $y \in S_1$ e $z \in S_1^\perp$.

Prova:

a) se $S_1 = S$, então $S_1^\perp = \{\phi\}$ e logo $z = \phi$, nesse caso $x = y$, onde z é ortogonal a qualquer vetor de dimensão compatível com a sua;

b) se $S_1 \neq S$ então $S_1^\perp \neq \phi$ existe e é único pelo Teorema 3.3.4. Sejam B_1 e B_2 bases para S_1 e $S_1^\perp$ respectivamente:

$$B_1 = \{u_1, u_2, \dots, u_p\} \quad \text{e} \quad B_2 = \{v_1, v_2, \dots, v_k\};$$

Neste caso: $\text{Dim}(S_1) = p$ e $\text{Dim}(S_1^\perp) = k$ e $B_1 \cup B_2$ é uma base para S :

$$B_1 \cup B_2 = \{u_1, u_2, \dots, u_p, v_1, v_2, \dots, v_k\}.$$

Logo, existem escalares $\alpha_i \in \mathfrak{R}$ e $\beta_j \in \mathfrak{R}$ tais que:

$$x = \alpha_1 u_1 + \alpha_2 u_2 + \dots + \alpha_p u_p + \beta_1 v_1 + \beta_2 v_2 + \dots + \beta_k v_k \quad \forall x \in S,$$

ou seja:
$$x = \sum_{i=1}^p \alpha_i u_i + \sum_{j=1}^k \beta_j v_j,$$

logo: $x = y + z$, pois y e z aparecem como combinações lineares de B_1 e B_2 respectivamente.

3.4 Projeções ortogonais

Genericamente a projeção de um vetor sobre um subespaço é um caso particular de uma transformação linear. Seja V um espaço vetorial e L um subespaço de V com dimensão finita. Uma função $T: V \rightarrow L$, que preserva a adição de vetores e a multiplicação por escalar é dita transformação linear de V em L , estabelecendo que se $x \in V$ então $T(x)$ é a projeção de x sobre L .

Considere então o seguinte tipo de problema: dado um elemento x de V , determinar um elemento de L cuja distância a x seja tão pequena quanto possível. Se $x \in L$, então é evidente que a solução é $s = x$. Se x não pertence a L , então o ponto s mais próximo define-se pelo pé da perpendicular tirada de x para o subespaço L .

Figura 5. Projeção ortogonal em 2 dimensões: x projetado em y .

Este exemplo simples sugere uma maneira de abordar o problema geral de aproximação e dá origem a discussão que se segue. Inicialmente são formalizadas algumas idéias essenciais sobre projetores ortogonais. A seguir definem-se as projeções ortogonais dentro do contexto de uma transformação linear, com enfoque matricial.

Definição 3.4.1. Seja S um espaço vetorial munido de produto interno, tal que $S = S_1 \oplus S_2 = S_1 \oplus S_1^\perp$. Considere um vetor arbitrário $x \in S$ que pode ser expresso de modo único por $x = x_1 + x_2$, tal que $x_1 \in S_1$ e $x_2 \in S_2 = S_1^\perp$. A função

$P: x \rightarrow x_1$ é uma transformação linear de S , denominada projetor ortogonal de S em S_1 , na direção de S_2 .

Teorema 3.4.1. Um operador linear P é um projetor ortogonal se e somente se P é idempotente, ou seja, $P^2 = P$.

Prova: Considere um espaço vetorial S como na Definição 3.4.1. Então, para todo $x \in S$, a decomposição $x = x_1 + x_2$ é única. Além disso, tem-se que $P(x) = x_1$, e $P(x_1) = x_1$ e para todo $x_2 \in S_2$ tem-se que $P(x_2) = \phi$. Nesse contexto:

$$P^2(x) = P[P(x_1 + x_2)] = P[P(x_1) + P(x_2)] = P(x_1) = x_1 = P(x).$$

Agora seja, $x_2 = x - x_1$, logo $P(x_2) = P(x - x_1) = P(x) - P(x_1) = \phi$.

Portanto, P é um projetor de S em S_1 , na direção de S_2 .

Corolário 3.4.1. Se P é o projetor de S em S_1 na direção de S_2 , então $(I - P)$ é o projetor de S em S_2 na direção de S_1 .

Prova: Desenvolvendo o binômio: $(I - P)^2 = I - 2P + P^2$, onde $P^2 = P$, vem $(I - P)^2 = (I - P)$ e, portanto $(I - P)$ é idempotente e é projetor. Agora se P é o projetor de S em S_1 na direção de S_2 , tem-se que $P(x) = x_1$ e portanto:

$$(I - P)x = x - P(x) = x - x_1 = x_2,$$

de modo que $(I - P)$ é o projetor ortogonal de S em S_2 na direção de S_1 .

Teorema 3.4.2. Se P é o projetor ortogonal de S em S_1 , então P além de idempotente é simétrico, isto é, $P' = P$, onde P' é o transposto de P .

Prova: A demonstração de que P é idempotente é dada pelo Teorema 3.4.1. Objetiva-se mostrar agora, que ele é simétrico. Para isso, considere dois vetores x e $y \in S$, onde $P(x) \in S_1$ e $(I - P)y \in S_1^\perp$, decorre que $\langle P(x), (I - P)y \rangle = 0$, ou seja, $\langle P'(I - P)y, x \rangle = 0 \quad \forall x, y \in S$. Então:

$$P'(I - P)y = 0, \forall y \in S \Rightarrow P'(I - P) = 0 \Rightarrow P' = P'P$$

e como $P = (P')' = (P'P)' = P'P = P'$, portanto $P' = P$.

Teorema 3.4.3. Seja P um projetor ortogonal de S em S_1 . Então, $\forall x \in S$, verifica-se que $\|x\| \geq \|P(x)\|$.

Prova:

a) Por definição, tem-se que $\|P(x)\|^2 \geq 0, \forall x \in S$. Então:

$$\|P(x)\|^2 = \langle Px, Px \rangle = x'P'Px = x'P^2x = x'Px = \langle Px, x \rangle \geq 0;$$

b) $\|(I - P)x\|^2 \geq 0, \forall x \in S$. Então:

$$\begin{aligned} \|(I - P)x\|^2 &= \langle (I - P)x, (I - P)x \rangle = x'(I - P)'(I - P)x = x'(I - P)^2x = \\ &= x'(I - P)x = \langle (I - P)x, x \rangle \geq 0. \end{aligned}$$

Agora seja:

$$\|x\|^2 - \|Px\|^2 = \langle x, x \rangle - \langle Px, Px \rangle \stackrel{(a)}{=} \langle x, x \rangle - \langle Px, x \rangle \geq 0^{(b)}$$

$$\therefore \|Px\| \leq \|x\|$$

Definição 3.4.2. Seja ${}_m A_n$ uma matriz real e sejam os vetores $x \in \mathbb{R}^n$ e $y \in \mathbb{R}^m$, tal que $y = Ax$. Diz-se que x foi transformado em y por meio de uma matriz de transformação ${}_m A_n$.

Definição 3.4.3. Considere ${}_m A_n$ uma matriz que transforme vetores $x_i \in \mathbb{R}^n$ em vetores $y_j \in \mathbb{R}^m$. Se para quaisquer x_1 e x_2 do $\mathbb{R}^n$ e quaisquer dois escalares c_1 e c_2 , sempre se verifique: $A(c_1x_1 + c_2x_2) = c_1Ax_1 + c_2Ax_2$, então a transformação é linear. As transformações $Ax = y$ além de lineares são homogêneas, visto que $A\phi = \phi$.

Definição 3.4.4. Se A_n é uma matriz não singular então a transformação $y = Ax$ é não singular. Se B_n é ortogonal então a transformação $y = Bx$ é ortogonal. Em ambos os casos as transformações são bijetoras.

Teorema 3.4.4. As transformações ortogonais preservam a distância euclidiana entre os vetores.

Prova: Seja $y = Ax$ uma transformação ortogonal, isto é, $A^{-1} = A'$.

Uma vez que $d^2(y_1, y_2) = \|y_2 - y_1\|^2 = (y_2 - y_1)'(y_2 - y_1)$, então:

$$\begin{aligned} d^2(y_1, y_2) &= (y_2 - y_1)'(y_2 - y_1) = (Ax_2 - Ax_1)'(Ax_2 - Ax_1) = \\ &= (x_2' A' - x_1' A')(Ax_2 - Ax_1) = (x_2' - x_1') A' A (x_2 - x_1). \end{aligned}$$

Mas, por hipótese $A' A = I$. Logo:

$$d^2(y_1, y_2) = (x_2' - x_1')(x_2 - x_1) = (x_2 - x_1)'(x_2 - x_1) = d^2(x_1, x_2) = \|x_2 - x_1\|^2,$$

$$\therefore \|y_2 - y_1\| = \|x_2 - x_1\|.$$

Isso posto, apresentam-se a seguir, alguns conceitos sobre um tipo especial de transformação linear - as projeções ortogonais - de importância particular neste trabalho.

Teorema 3.4.5. Seja $V \subseteq \mathbb{R}^m$ um espaço vetorial e seja $\{v_1, v_2, \dots, v_k\}$ um subconjunto de k vetores ortonormais de V . Se S_1 é um subespaço de V gerado por $G(S_1) = \{v_1, v_2, \dots, v_k\}$, então qualquer vetor $z \in V$ pode ser decomposto como $z = s_1 + s_2$, onde $s_1 \in S_1$ e $s_2 \in S_2$, com $S_1 \perp S_2$. Sendo assim:

$$s_1 = \langle z, v_1 \rangle v_1 + \langle z, v_2 \rangle v_2 + \dots + \langle z, v_k \rangle v_k$$

$$s_2 = z - s_1$$

onde s_1 é a projeção ortogonal de z no subespaço S_1 e s_2 é a direção dessa projeção.

Figura 6. Projeção ortogonal em 3 dimensões.

Definição 3.4.5. Considere $A = [v_1, v_2, \dots, v_k]$ uma matriz cujas colunas sejam vetores ortonormais, seja também $G(S_1) = C(A)$. De acordo com o Teorema 3.4.5:

$$s_1 = \langle z, v_1 \rangle v_1 + \langle z, v_2 \rangle v_2 + \dots + \langle z, v_k \rangle v_k$$

$$s_1 = (v_1' z) v_1 + (v_2' z) v_2 + \dots + (v_k' z) v_k$$

$$s_1 = (v_1' v_1) z + (v_2' v_2) z + \dots + (v_k' v_k) z$$

$$s_1 = [v_1' v_1 + v_2' v_2 + \dots + v_k' v_k] z$$

$$s_1 = Pz$$

Iemma (1989), mostra que $s_1 = Pz$, onde $P = A'A$ é uma matriz simétrica e idempotente, denominada projetor ortogonal de z no espaço gerado pelas colunas da matriz A . Além disso, $(I - P)z = s_2$, projeta ortogonalmente os vetores de V sobre $C^\perp(A) = S_2$, o complemento ortogonal do espaço coluna de A .

3.5 Autovalores, autovetores e a decomposição por valores singulares

Definição 3.5.1. Dada uma matriz quadrada A_n , a equação polinomial $\delta(A - \lambda I) = 0$ é definida como equação característica de A_n . As raízes da equação característica de A são ditas raízes características de A , autovalores de A , valores próprios de A ou simplesmente raízes de A .

Definição 3.5.2. Os vetores x que satisfazem o sistema homogêneo $(A - \lambda I)x = 0$ são ditos vetores característicos, autovetores ou vetores próprios de A .

Definição 3.5.3. De acordo com Eckart & Young (1936), a decomposição por valores singulares de uma matriz ${}_nX_p$ de posto $r \leq \min\{n, p\}$, consiste em fatorá-la como:

$${}_nX_p = {}_nU {}_nS {}_pV'$$

onde U e V são matrizes ortogonais, ou seja, $U'U = I$ e $V'V = I$ e S é uma matriz de dimensão $n \times p$ tal que:

$$S = \begin{bmatrix} s_1 & & & & & \\ & s_2 & & & & \\ & & \ddots & & & \\ & & & s_r & & \\ & & & & 0 & \\ & & & & & 0 \\ & & \phi & & 0 & \\ & & & & & \ddots \\ & & & & & & 0 \end{bmatrix}$$

sendo que os valores $s_1 \geq s_2 \geq \dots \geq s_r > 0$ são chamados de valores singulares da matriz X .

A decomposição por valores singulares é estritamente relacionada com os conceitos de autovalor e autovetor, uma vez que s_i^2 , $i = 1, 2, \dots, r$ é raiz característica não nula de XX' (e de $X'X$) e as n colunas de U formam um conjunto de autovetores ortogonais normalizados associados aos autovalores de XX' . Por outro lado, as linhas de V' são vetores normalizados correspondentes aos autovalores de $X'X$.

Deste modo, é possível particionar as matrizes, U e V , como se segue:

$$U = [U_1 \mid U_2] \quad \text{e} \quad V = [V_1 \mid V_2],$$

onde $U_1 = [u_1, u_2, \dots, u_r]$ são os autovetores normalizados correspondentes aos autovalores não nulos de XX' e $U_2 = [u_{r+1}, \dots, u_n]$ são os vetores próprios normalizados associados as raízes características nulas de XX' . Analogamente $V_1 = [v_1, \dots, v_r]$ são os autovetores normalizados associados aos autovalores não nulos de $X'X$, enquanto que $V_2 = [v_{r+1}, \dots, v_p]$ são vetores próprios normalizados correspondentes às raízes características nulas de $X'X$.

Sendo assim, U_1 constitui uma base ortonormal para o espaço coluna da matriz X , $C(X)$, enquanto que V_1 é uma base ortonormal para $C(X')$, o espaço linha da matriz X . Ademais, se U_1 é uma base ortonormal, então $XX'u_i = s_i^2 u_i$, $i = 1, 2, \dots, r$ e desse modo v_i , $i = 1, 2, \dots, r$ pode ser definido como $v_i = \frac{1}{s_i} X'u_i$.

Então, por construção tem-se que:

$$i) U = [U_1 \mid U_2] \quad \text{e} \quad V = [V_1 \mid V_2];$$

$$ii) XV_2 = \phi \quad \text{e} \quad X'U_2 = \phi.$$

E uma decomposição, muito usada para a matriz X é dada por:

$$X = U_1 S_1 V_1'$$

onde S_1 é uma matriz diagonal que exhibe os valores singulares de X . Esta decomposição é mais usualmente referida como a decomposição por valores singulares (**DVS**) de X .

3.6 Inversas generalizadas

As inversas generalizadas constituem-se num dos tópicos de grande aplicabilidade no desenvolvimento da teoria sobre as projeções ortogonais. Ademais, são de grande importância no estudo dos modelos lineares. Entre outros são apresentados aqui três tipos de interesse:

Definição 3.6.1. Inversa generalizada condicional

Dada a matriz ${}_m A_n$ de posto k , existe uma matriz ${}_n A_m^-$, dita inversa condicional de ${}_m A_n$, satisfazendo:

$$AA^-A = A.$$

Um algoritmo prático para se obter A^- foi proposto por Searle (1971):

- Tomar, em A , qualquer matriz não singular, M , de ordem k ;
- Calcular $[(M)^{-1}]'$;
- Substituir em A os elementos correspondentes pelos de $[(M)^{-1}]'$;
- Zerar os demais elementos;
- Transpor a matriz;
- A resultante é uma inversa condicional de A , que em geral, não é única.

Definição 3.6.2. Inversa generalizada de mínimos quadrados

Dada a matriz ${}_m A_n$ de posto k , toda matriz ${}_n A_m^\ell$, que satisfaz as duas condições seguintes, é dita inversa generalizada de mínimos quadrados de A :

- $AA^\ell A = A$
- $(AA^\ell) = (AA^\ell)'$ (simétrica)

Teorema 3.6.1. Toda matriz do tipo ${}_n A_m^\ell = (A'A)^- A'$ é inversa generalizada de mínimos quadrados de ${}_m A_n$.

Definição 3.6.3. Inversa generalizada de Moore-Penrose

Dada a matriz ${}_m A_n$ de posto k , a matriz ${}_n A_m^+$, de posto k , que satisfaz as quatro condições seguintes, é definida como inversa generalizada de Moore-Penrose de A :

a) $AA^+A = A$	c) $(AA^+) = (AA^+)'$	(simétrica)
b) $A^+AA^+ = A^+$	d) $(A^+A) = (A^+A)'$	(simétrica)

Uma regra para a obtenção de A^+ é dada por:
 $A^+ = C'(CC')^{-1}(B'B)^{-1}B'$, onde B é a matriz de posto coluna completo e C é a matriz de posto linha completo, obtidas pela fatoração de posto completo de A , tal que $A = BC$. Em geral as matrizes B e C não são únicas e, dentre outros modos, podem ser obtidas pelo algoritmo de Dwivedi (1975):

Dada a matriz ${}_m A_n = (a_{ij})$, de posto k :

- a) Escolher algum elemento $a_{pq} \neq 0$;

b) Obter o produto $u_1 v_1'$, onde $u_1 = \frac{1}{a_{pq}} \begin{bmatrix} a_{1q} \\ a_{2q} \\ \vdots \\ a_{mq} \end{bmatrix}$; $v_1' = [a_{p1} \quad a_{p2} \quad \cdots \quad a_{pn}]$;

- c) Fazer $A_1 = A - u_1 v_1'$;

d) Se $A_1 = \phi$, o processo está encerrado e $B = u_1$; $C = v_1'$ e $r[A] = 1$;

e) Se $A_1 \neq \phi$, repetir o processo para A_1 e assim sucessivamente, até obter $A_k = \phi$.

Nesse caso: $B = [u_1 \mid u_2 \mid \dots \mid u_k]$; $C = \begin{bmatrix} v_1' \\ v_2' \\ \vdots \\ v_k' \end{bmatrix}$ e $r[A] = k$.

Teorema 3.6.2. Existência e unicidade

Para cada matriz ${}_m A_n$ de posto k , existe uma e somente uma matriz ${}_n A_m^+$.

Pode-se verificar que A^+ assim definida sempre existe e é única. Além disso, as formas (A^+A) e (AA^+) são idempotentes.

3.7 Algumas idéias sobre modelos lineares

Face aos objetivos deste trabalho, serão adotados modelos lineares do tipo:

$${}_n y_1 = {}_n X_p \theta_1 + {}_n e_1$$

onde:

${}_n y_1$ é um vetor de observações;

${}_n X_p$ é a matriz conhecida do delineamento de posto $k \leq \min\{n, p\}$;

${}_p \theta_1$ é um vetor de parâmetros desconhecidos;

${}_n e_1$ é um vetor de variáveis aleatórias não observáveis, tal que $e \cap N(\phi, I\sigma^2)$.

Um modelo linear com essas caracterizações é definido como modelo linear de Gauss-Markov (G.M.). Naturalmente, conforme citado por Lemma (1987), esse modelo pode assumir outras formas e caracterizações.

3.7.1 Equações normais e aproximação por mínimos quadrados

Sendo e um vetor de variáveis aleatórias não observáveis, sua estimação depende da caracterização assumida e da solução do sistema de equações lineares $y = X\theta$, que geralmente é inconsistente. Uma solução interessante para esse impasse pode ser obtida através do sistema de equações normais. O problema consiste em encontrar θ que torne mínima a norma euclidiana: $Q(\theta) = \|y - X\theta\|^2$. Pelo princípio dos mínimos quadrados, tem-se que:

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

Derivando $Q(\theta)$ em relação a θ , obtém-se:

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

e, igualando-se a derivada a zero, tem-se o chamado sistema de equações normais:

$$X'X\theta = X'y$$

O sistema de equações normais é sempre consistente e qualquer solução exata, θ^0 , do sistema de equações normais é solução aproximada de mínimos quadrados para $y = X\theta$ inconsistente. Consequentemente, qualquer que seja a solução θ^0 do sistema de equações normais, o vetor $\hat{y} = X\theta^0$ é invariante e é a aproximação de mínimos quadrados para o vetor y das observações.

A justificativa principal da utilização do método dos mínimos quadrados, é dada pelo Teorema de Gauss-Markov: “se $\lambda'\theta$ é estimável no modelo $y = X\theta + e$,

então seu *blue* (*best linear unbiased estimator*) é dado de modo único por $\lambda'\theta^0$, para qualquer θ^0 solução das equações normais.” Esse Teorema garante que na classe dos estimadores lineares não tendenciosos de $\lambda'\theta$, o estimador $\lambda'\theta^0$ de mínimos quadrados, é o que tem variância mínima, isto é, o melhor estimador linear não tendencioso.

Segundo Lemma (1987), se a matriz X tem posto coluna completo, como é o caso dos modelos de regressão e dos modelos de planejamento com restrições, $X'X$ é positiva definida e portanto não singular. Nesse caso, o sistema de equações normais terá solução única: $\theta^0 = (X'X)^{-1} X'y$. E θ^0 , solução das equações normais, coincide com o estimador de mínimos quadrados $\hat{\theta}$, de θ . De fato:

$$E[\hat{\theta}] = E[(X'X)^{-1} X'y] = (X'X)^{-1} X'E[y] = (X'X)^{-1} X'X\theta = \theta,$$

$\hat{\theta}$ é um estimador não viesado de θ . Nesse caso, sendo $(X'X)$ simétrica, então a matriz de variâncias e covariâncias é dada por:

$$V[\hat{\theta}] = V[(X'X)^{-1} X'y] = (X'X)^{-1} X'V[y]X(X'X)^{-1} = (X'X)^{-1} \sigma^2.$$

Nesse contexto, cada elemento de θ é individualmente estimável. Além disso, visto que $X\theta^0 = X(X'X)^{-1} X'y = Py$, P é o operador de transformação linear que projeta ortogonalmente y no espaço euclidiano gerado pelas colunas da matriz X , $C(X)$ e, $(I - P)$ representa o projetor ortogonal do vetor y no complemento ortogonal do espaço coluna de X , $C^\perp(X)$.

No entanto, se a matriz X não tem posto coluna completo, $X'X$ é singular e o sistema de equações normais, não proporcionará solução única. Em outras palavras, qualquer θ^0 será mera solução aproximada de mínimos quadrados. Decorre que os elementos de θ não são individualmente estimáveis.

3.7.2 Projetores ortogonais e o modelo linear

Na seção 3.4 foram formalizadas algumas características sobre projetores ortogonais e na seção 3.6 foram postas idéias sobre inversas generalizadas. Considere-se agora, o teorema:

Teorema 3.7.2.1. Seja $V \subseteq \mathfrak{R}^n$ tal que $V = C(X) \oplus C^\perp(X)$. Então a matriz que projeta ortogonalmente qualquer vetor $y \in V$ no $C(X)$ na direção do $C^\perp(X)$ é a matriz $P = X(X'X)^G X' = XX^\ell = XX^+$. A matriz P é simétrica, idempotente e invariante para qualquer inversa generalizada de $X'X$ ou qualquer inversa de mínimos quadrados de X . Analogamente $(I - P)$ é a matriz que projeta ortogonalmente o vetor y no $C^\perp(X)$ na direção de $C(X)$.

Dessa forma, a aplicação de projetores ortogonais aos modelos lineares é imediata:

- a) $\hat{y} = Py = XX^+y = X\theta^0$, invariante para qualquer θ^0 solução das equações normais $X'X\theta = X'y$, é a aproximação de mínimos quadrados para $y \in \mathfrak{R}^n$, no sistema inconsistente $y = X\theta$;
- b) $\hat{e} = (I - P)y = y - Py = y - \hat{y}$ é o erro de ajuste devido à aproximação de mínimos quadrados para y no sistema inconsistente $y = X\theta$;
- c) $y = \hat{y} + \hat{e}$, onde $y \in \mathfrak{R}^n$, $\hat{y} \in C(X)$ e $\hat{e} \in C^\perp(X)$.

Teorema 3.7.2.2. Uma condição necessária e suficiente para que θ^0 seja solução aproximada de mínimos quadrados para $y = X\theta$ inconsistente é que $\hat{y} = X\theta^0 = Py$ seja a projeção ortogonal de y sobre $C(X)$.

Teorema 3.7.2.3. Dado o sistema inconsistente $y = X\theta$ e θ^0 uma solução aproximada qualquer. Uma condição necessária e suficiente para que a soma de quadrados dos erros devido à aproximação seja mínima é que o vetor de erros seja ortogonal ao espaço coluna de X .

Prova: $\|\hat{e}\|^2$ é mínimo se e somente se:

$$X'X\theta^0 = X'y \Leftrightarrow X'(y - X\theta^0) = \phi \Leftrightarrow X'(y - \hat{y}) = \phi \Leftrightarrow X'\hat{e} = \phi,$$

portanto $\hat{e} \perp C(X)$.

Teorema 3.7.2.4. Uma condição necessária e suficiente para que a soma de quadrados dos erros devido à aproximação θ^0 seja mínima é que $\hat{y} = Py$.

Prova: $\|\hat{e}\|^2$ é mínimo $\Leftrightarrow X'X\theta^0 = X'y \Leftrightarrow X\hat{y} = X'y \Leftrightarrow \hat{y} = XX^+y \Leftrightarrow \hat{y} = Py$.

4 RESULTADOS E DISCUSSÃO

4.1 Algumas estatísticas descritivas

4.1.1 A média e a variância amostrais

Suponha que se tenha uma amostra de n observações da variável Y e que se deseja representar esse conjunto de dados $(y_1, y_2, \dots, y_n)$ através de um único valor.

Com o objetivo de encontrar qual é o melhor valor para representar os dados, suponha também o modelo linear mais simples possível, apenas com intercepto e nenhuma variável independente: $y = \mu J + e$. Neste modelo, μ representa um parâmetro escalar desconhecido, J é um vetor de “uns” e e é um vetor não observável de erros aleatórios.

Pelo princípio de mínimos quadrados, o melhor valor para sumarizar os dados de y é o valor de μ que minimiza: $\sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - \mu)^2$. Usando o conceito de norma euclidiana, esta equação pode ser escrita como: $\|y - \mu J\|^2$.

Geometricamente, como mostra a Figura 7, a aproximação de mínimos quadrados para os dados de y significa que devem ser considerados todos os possíveis modelos μJ que sejam múltiplos de J e escolher aquele que seja mais próximo de y . Logo, o procedimento de mínimos quadrados leva à menor distância de y à μJ , quando se usa o produto interno euclidiano.

Na seção 3.4, verificou-se que Py fornece o ponto mais próximo a y na linha definida por J . Usando as propriedades do produto interno e das projeções tem-se que:

$$\begin{aligned} \langle y - Py, J \rangle &= 0 \Rightarrow \langle y, J \rangle - \hat{\mu} \langle J, J \rangle = 0 \Rightarrow \langle y, J \rangle = \hat{\mu} \langle J, J \rangle \Rightarrow \\ \Rightarrow \langle y, J \rangle &= \hat{\mu} n \Rightarrow \sum_{i=1}^n y_i = \hat{\mu} n \Rightarrow \hat{\mu} = \bar{y} \end{aligned}$$

Figura 7. A média amostral como uma projeção ortogonal.

Alternativamente, utilizando as equações normais na forma matricial $J'J\mu = J'y$ tem-se que, como J é de posto coluna completo, então a solução aproximada de mínimos quadrados para $y = \mu J$ inconsistente, coincide com o estimador de mínimos quadrados para μ :

$$\hat{\mu} = (J'J)^{-1} J'y = \frac{1}{n} J'y = \frac{\sum_{i=1}^n y_i}{n} = \bar{y}$$

Ademais, a aproximação de mínimos quadrados para o vetor y das observações fica $\hat{y} = Py = J(J'J)^{-1} J'y = \hat{\mu}J = \bar{y}J$. Logo, a média amostral de y é o estimador de mínimos quadrados para o parâmetro populacional μ e $\bar{y}$ é a projeção ortogonal de y sobre a linha definida por J .

Outras medidas de tendência central comumente usadas na Estatística Descritiva são a *moda* e a *mediana*. Como se sabe, se a distribuição dos dados é simétrica, média, moda e mediana amostrais coincidem e, portanto, geometricamente

representam a projeção ortogonal do vetor y de observações sobre o vetor J de “uns”. Em caso de distribuição assimétrica, mediana e moda, representam, geometricamente, *outras projeções* do vetor y sobre J . Essas projeções, correspondem a pontos localizados à direita ou esquerda de $\bar{y}J$, conforme a natureza da assimetria. A Figura 8 mostra essas projeções para os casos de distribuições assimétrica positiva e assimétrica negativa.

Figura 8. Visão geométrica da moda e da mediana em caso de assimetria.

Devido a ortogonalidade, observada na decomposição do vetor y (Figura 7), pode-se aplicar o Teorema de Pitágoras:

$$\|y\|^2 = \|\bar{y}J\|^2 + \|y - \bar{y}J\|^2 \Rightarrow \sum_{i=1}^n y_i^2 = n\bar{y}^2 + \sum_{i=1}^n (y_i - \bar{y})^2$$

obtendo-se a decomposição usual da soma de quadrados.

Se o conjunto de dados é composto de valores idênticos, então, a média será o próprio valor e $\|\hat{e}\|^2 = 0$. Isso significa, que y pode ser explicado satisfatoriamente por uma constante.

Pode-se medir “quão boa” é a representação de y por $\bar{y}J$, usando o ângulo θ entre y e sua aproximação de mínimos quadrados $\hat{y} = \bar{y}J$ ou $\|\hat{e}\|^2$. Sendo que a medida $\|\hat{e}\|^2$ corresponde a soma de quadrados dos erros (SQerro) e é diretamente proporcional ao valor de n , isto é, quanto maior o número de observações, mais variação ocorrerá. Para contabilizar essa medida, costuma-se dividir a SQerro pela

dimensão do subespaço ortogonal a J , $(n - 1)$. O resultado obtido é a média dos quadrados dos erros, ou seja, a variância amostral.

$$\frac{\|\hat{e}\|^2}{n-1} = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1}$$

Analogamente, aplicando a projeção ortogonal complementar em y , produz-se o vetor estimado de erros no modelo linear assumido:

$$(I - P)y = \hat{e} = y - \hat{y} = y - \bar{y}J$$

Também a menor soma de quadrados dos erros produz o numerador da variância:

$$\hat{e}'\hat{e} = y'(I - P)'(I - P)y = (y - \bar{y}J)'(y - \bar{y}J) = \sum_{i=1}^n (y_i - \bar{y})^2$$

Pode-se interpretar a variância amostral de y , em termos geométricos, como o comprimento quadrático médio de $\hat{e}$. Correspondentemente, o desvio padrão amostral de y representa comprimento médio de $\hat{e}$, sendo uma medida da distância média dos valores individuais y_i a $\bar{y}$.

4.1.1.1 Exemplo

Seja uma amostra aleatória de tamanho 3, $n = 3$, com os seguintes valores: 2,1; 1,8 e 2,1. A variável Y em estudo representa, por exemplo, a produção de milho, em toneladas/hectare, em fazendas de médio porte. Suponha que $Y \cap N(\mu, \sigma^2)$, onde μ e σ^2 são parâmetros desconhecidos. O objetivo, nesse caso, é obter as estimativas desses parâmetros, usando a abordagem geométrica.

De acordo com o procedimento descrito em 4.1.1, assumindo o modelo linear $y_i = \mu + e_i$, pode-se expressar o vetor y como:

4.1.2 O coeficiente de correlação linear simples e a covariância amostral

Imagine que foram feitas n medições de duas características X e Y . O objetivo, neste caso, é verificar que tipo de relação existe entre essas duas características. A intensidade da relação entre duas variáveis, geometricamente, pode ser quantificada pelo grau de proximidade entre os vetores x e y , que representam as n medidas de X e Y respectivamente (Figura 10). Para quantificar tal intensidade pode-se usar a medida $\|x - y\|^2$ ou alguma função do ângulo entre os vetores x e y .

Figura 10. Coeficiente de correlação linear simples.

Para avaliar tal relação tomam-se, então, as variações de x e y , representadas na Figura 10 por seus componentes ortogonais a J : $x - \bar{x}J = x - Px = x - \hat{x}$ e $y - \bar{y}J = y - Py = y - \hat{y}$. O produto interno entre os vetores $x - \bar{x}J$ e $y - \bar{y}J$, quando dividido pela dimensão do subespaço ortogonal a J , $(n - 1)$, é chamado covariância amostral de X e Y .

Uma medida usual que relaciona as variações de x e de y é o $\cos\theta$, chamado de correlação linear simples entre as variáveis X e Y , uma vez que:

$$\langle x - \bar{x}J, y - \bar{y}J \rangle = \sum_{i=1}^n [(x_i - \bar{x})(y_i - \bar{y})] = (n - 1)\text{Cov}(X, Y)$$

$$\|x - \bar{x}J\| = \sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} = \sqrt{n-1} s_x \quad \text{e} \quad \|y - \bar{y}J\| = \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2} = \sqrt{n-1} s_y$$

$$\text{logo: } \cos\theta = \frac{(n-1)\text{Cov}(X,Y)}{\sqrt{n-1} s_x \sqrt{n-1} s_y} = \frac{\text{Cov}(X,Y)}{s_x s_y} = r_{xy}; \quad -1 \leq r_{xy} \leq 1.$$

Quando $\cos\theta = 1$, então $\theta = 0^\circ$ ou quando $\cos\theta = -1$, então $\theta = 180^\circ$, as variações em x são exatamente proporcionais às variações em y , isto é, $x - \bar{x}J$ e $y - \bar{y}J$ são colineares. Se, no entanto, $\cos\theta = 0$, então $\theta = 90^\circ$, as variações de x não estão relacionadas com as variações de y : $x - \bar{x}J$ e $y - \bar{y}J$ são ortogonais.

4.1.2.1 Exemplo

Considere os dados da Tabela 1, através dos quais se quer verificar se existe algum tipo de relação linear entre as variáveis X e Y .

Tabela 1. Valores fictícios de duas características.

X	Y
2	18
4	12
5	10
6	8
8	7
11	5

Tomando-se os valores de X e Y como componentes dos vetores x e y respectivamente, pode-se representá-los conjuntamente em uma matriz como se segue:

$$\begin{bmatrix} 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \\ 1 & x_3 & y_3 \\ 1 & x_4 & y_4 \\ 1 & x_5 & y_5 \\ 1 & x_6 & y_6 \end{bmatrix} \Rightarrow \begin{bmatrix} 1 & 2 & 18 \\ 1 & 4 & 12 \\ 1 & 5 & 10 \\ 1 & 6 & 8 \\ 1 & 8 & 7 \\ 1 & 11 & 5 \end{bmatrix} \Rightarrow [J \mid x \mid y]$$

Sendo $P_J = J(J'J)^{-1}J' = \frac{1}{6}[J \mid J \mid J \mid J \mid J \mid J]$, então $P_J x = \hat{x} = \bar{x}J = 6J$ e $P_J y = \hat{y} = \bar{y}J$, onde as médias amostrais 6 e 10 são respectivamente as projeções ortogonais dos vetores x e y , respectivamente, sobre o vetor J de “uns”. A covariância amostral é dada pela razão entre o produto interno dos vetores $[x - Px]' = [-4 \ -2 \ -1 \ 0 \ 2 \ 5]$ e $[y - Py]' = [8 \ 2 \ 0 \ -2 \ -3 \ -5]$ e a dimensão do subespaço ortogonal a J , $(n - 1)$.

$$Cov(X, Y) = \frac{\langle x - \bar{x}J, y - \bar{y}J \rangle}{n - 1} = \frac{-67}{5} = -13,4.$$

O coeficiente de correlação linear simples entre X e Y , será dado pelo cosseno do ângulo formado entre os vetores $[x - Px]$ e $[y - Py]$:

$$\cos \theta = \frac{\langle x - \bar{x}J, y - \bar{y}J \rangle}{\|x - \bar{x}J\| \|y - \bar{y}J\|} = \frac{-67}{\sqrt{50}\sqrt{106}} = -0,9203 = r_{xy},$$

concluindo-se, portanto, que existe uma forte correlação linear negativa.

4.2 A probabilidade na óptica da geometria

Uma situação comum na análise de experimentos é a seguinte: verifica-se que o fenômeno estudado tem n possibilidades distintas de se manifestar, descritas no seu espaço amostral Ω e, a cada possível resultado (numérico ou não numérico) de Ω associa-se uma dada probabilidade. Contudo, em muitas experiências tem-se interesse na mensuração de uma característica particular e no seu registro como um número. Sempre que se associa um número real a cada resultado de Ω , considera-se uma função cujo domínio é o próprio espaço amostral e o contradomínio é conjunto dos números reais em questão. Tal função $X:\Omega \rightarrow \mathbb{R}$ é denominada variável aleatória. Nesse sentido, informalmente, a variável aleatória pode ser vista como a caracterização numérica do resultado do experimento. Interpretando-se as variáveis aleatórias como vetores, torna-se possível pensar na Teoria da Probabilidade geometricamente. A título de ilustração considere a seguinte situação:

Duas pessoas, A e B, fazem a seguinte aposta: lançam duas vezes uma única moeda. Se der duas caras, A ganhará 16 unidades monetárias (u.m.) de B, se der duas coroas, A ganhará 10 u.m. de B e, se der uma cara e uma coroa, B ganhará 14 u.m. de A. Então tem-se $\Omega = \{(cara, cara); (coroa, coroa); (cara, coroa); (coroa, cara)\}$; a variável aleatória do ponto de vista de A será $X = [16, 10, -14]$; do ponto de vista de B será $Y = [-16, -10, 14]$, onde o sinal negativo indica a perda na aposta e, $p = [1/4, 1/4, 1/2]$ é o vetor de probabilidades associado às variáveis aleatórias. Geometricamente, como ilustra a Figura 11, as variáveis aleatórias X e Y representam dois vetores do $\mathbb{R}^3$.

Figura 11. Uma visão geométrica das variáveis aleatórias no espaço tridimensional.

Por outro lado, o valor esperado de uma variável aleatória X com função de distribuição $F(x)$ é definido rigorosamente por alguns tratados da Estatística Matemática pela integral genérica de *Lebesgue-Stieltjes*, aqui denotada integral L.S.:

$$E(X) = \int_{-\infty}^{+\infty} x dF(x), \quad (4.2.1)$$

que existe se, e somente se, essa integral for convergente.

Corrêa (1981) e James (1981) comentam a vantagem da utilização dessa integral na definição da esperança matemática, bem como em outros conceitos. Assim, a integral L.S. evita a necessidade de se exporem todos os teoremas e definições duplamente e, desse modo, se $F(x)$ é a função de distribuição de uma variável aleatória discreta X , a integral L.S. reduz-se a uma série: se $P(X = x_i) = p(x_i) \geq 0$ e $\sum_i p(x_i) = 1$, isto é, se p é a função de probabilidade de X , então $p(x_i)$ é um salto de $F(x)$ em x_i e:

$$\int_{-\infty}^{+\infty} x dF(x) = \sum_i x_i p(x_i).$$

Sendo $F(x)$ uma função de distribuição de uma variável aleatória contínua que tem como função densidade de probabilidade $f(x)$, então $f(x)$ é a derivada de $F(x)$, isto é $dF(x) = f(x)dx$ e, a integral L.S. torna-se uma integral comum do tipo Riemann:

$$\int_{-\infty}^{+\infty} x dF(x) = \int_{-\infty}^{+\infty} x f(x) dx = \lim_{\substack{a \rightarrow -\infty \\ b \rightarrow +\infty}} \int_a^b x f(x) dx,$$

também como um caso particular.

A ligação deste conceito (4.2.1) com a geometria fica estabelecida pelo produto interno:

$$\langle (x_1, x_2, \dots, x_n), (y_1, y_2, \dots, y_n) \rangle = p_1 x_1 y_1 + p_2 x_2 y_2 + \dots + p_n x_n y_n$$

e, desse modo, é imediato verificar que em relação ao produto interno:

$$E(X) = \langle X, J \rangle \quad (4.2.2)$$

Logo, para o exemplo das apostas, pode-se calcular o valor esperado das variáveis aleatórias X e Y :

$$E(X) = \langle X, J \rangle = \frac{1}{4} \cdot 16 \cdot 1 + \frac{1}{4} \cdot 10 \cdot 1 + \frac{1}{2} \cdot (-14) \cdot 1 = -\frac{1}{2}$$

$$E(Y) = \langle Y, J \rangle = \frac{1}{4} \cdot (-16) \cdot 1 + \frac{1}{4} \cdot (-10) \cdot 1 + \frac{1}{2} \cdot 14 \cdot 1 = \frac{1}{2}$$

Pelos resultados obtidos para os valores esperados das variáveis aleatórias (v.a.'s) X e Y , verifica-se que o jogador B tem vantagem na aposta, pois o valor esperado por B é positivo, enquanto que o valor esperado por A é negativo. Esta aposta só seria justa se não houvesse vantagem para nenhum dos apostadores. O produto interno, nesse caso, apresenta uma aplicação bastante interessante na verificação de que o jogo é equitativo ou não. Um jogo equitativo é aquele em que o ganho esperado é nulo, ou seja, a longo prazo, ou em média, não se espera ganhar nem perder. Isto só ocorre quando o produto interno da variável aleatória pelo vetor J for zero, o que indica a ortogonalidade entre o vetor de probabilidades e o vetor de apostas descrito pela respectiva variável aleatória.

De modo análogo ao caso unidimensional pode-se definir a esperança conjunta de v.a.'s de dimensão mais elevada. Em particular, para o caso bidimensional

se (X,Y) é um vetor aleatório e $Z = g(X,Y)$ uma função real de (X,Y) então Z também é uma variável aleatória (unidimensional) e seu valor esperado pode ser dado pela integral L.S. (4.2.1), já apresentada. Então, sem perda de generalidade seja:

$$E(X.Y) = \int_{-\infty}^{+\infty} z dF(z) = \int_{-\infty}^{+\infty} xy dF(x,y) \quad (4.2.3)$$

Este é um resultado extremamente útil porque ele mostra que para calcular o valor esperado de (X,Y) não há necessidade de se conhecer a distribuição de probabilidade da variável aleatória Z . Sendo assim, se (X,Y) é um vetor de v.a.'s discretas e se $p(X = x_i, Y = y_j) = p(x_i, y_j) \geq 0$, isto é, $p(x_i, y_j)$ é a função de probabilidade conjunta de (X,Y) então:

$$E(X.Y) = \sum_i \sum_j x_i y_j p(x_i, y_j);$$

e, se (X,Y) é um vetor de v.a.'s contínuas com função densidade de probabilidade conjunta $f(x,y)$, então:

$$E(X.Y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xy f(x,y) dx dy;$$

verificando-se, em ambos os casos, as mesmas particularidades da integral L.S.

Assim como a esperança unidimensional, a esperança conjunta de duas variáveis aleatórias quaisquer (4.2.3) pode ser também interpretada como um produto interno, embora neste caso, o produto interno não seja usual. Sobre este aspecto, Bryant (1984) mostra que:

$$E(X.Y) = \langle X, Y \rangle = \|X\| \|Y\| \cos \theta \quad (4.2.4)$$

verificando-se que a equação (4.2.4) é válida em relação ao seguinte produto interno em $\mathfrak{R}^n$:

$$\langle (x_1, x_2, \dots, x_n), (y_1, y_2, \dots, y_n) \rangle = p_{11}x_1y_1 + p_{12}x_1y_2 + \dots + p_{nn}x_ny_n;$$

e, obviamente, a equação (4.2.4) satisfaz as propriedades básicas do produto interno. Considere o exemplo ilustrativo desta seção, cuja distribuição conjunta das variáveis aleatórias X e Y é apresentada na Tabela 2:

Tabela 2. Distribuição de probabilidade conjunta de X e Y .

$Y \backslash X$	-14	10	16	Marginal (Y)
-16	0	0	1/4	1/4
-10	0	1/4	0	1/4
14	1/2	0	0	1/2
Marginal (X)	1/2	1/4	1/4	1

Com base nestes dados tem-se:

$$E(X.Y) = \langle (16, 10, -14), (-16, -10, 14) \rangle = -187$$

Uma das vantagens da descrição da esperança através do produto interno é que, por meio de suas propriedades torna-se fácil a demonstração de alguns teoremas envolvendo a esperança matemática, a saber:

a) $E(k) = k$, onde k é uma constante real;

A demonstração é imediata pela definição (4.2.2), bastando efetuar o produto interno: $E(k) = \langle kJ, J \rangle = E(kJ) = k$.

b) $E(kX) = kE(X)$;

A demonstração decorre da propriedade homogeneidade do produto interno: $E(kX) = \langle kX, J \rangle = k \langle X, J \rangle = kE(X)$.

c) $E(kX+q) = kE(X) + q$, onde q é uma constante real;

A demonstração decorre das propriedades a e b .

d) $E(X+Y) = E(X) + E(Y)$;

A demonstração decorre da propriedade distributiva do produto interno:

$$E(X+Y, J) = \langle X+Y, J \rangle = \langle X, J \rangle + \langle Y, J \rangle = E(X) + E(Y).$$

A abordagem geométrica da esperança através do produto interno pode também ser utilizada para mostrar que muitas outras idéias da Teoria da Probabilidade correspondem a idéias comuns em Geometria. Desse modo, para qualquer variável aleatória X :

$$\langle X, X \rangle = \|X\|^2 = E(X \cdot X) = E(X^2),$$

ou seja, $E(X^2)$ corresponde geometricamente ao comprimento quadrático do vetor representativo da v.a. X . Tomando-se uma v.a. centrada, por analogia, pode-se definir a variância associada à v.a. X . Considere que $E(X) = \mu_X$, então:

$$\langle X - \mu_X, X - \mu_X \rangle = \|X - \mu_X\|^2 = E[(X - \mu_X)^2] = \sigma_X^2;$$

consequentemente, o desvio padrão da v.a. X é dado pela norma euclidiana:

$$\sigma_X = \|X - \mu_X\|$$

Aplicando essas idéias ao exemplo do jogo da moeda, tem-se:

$$\begin{aligned} \sigma_X^2 &= \langle (16, 10, -14) - (-1/2)(1, 1, 1), (16, 10, -14) - (-1/2)(1, 1, 1) \rangle = \|(33/2, 21/2, -27/2)\|^2 = 186,75 \\ \therefore \sigma_X &= \|(33/2, 21/2, -27/2)\| = 13,66 \end{aligned}$$

e de modo análogo para a v.a. Y , tem-se $\sigma_Y^2 = 186,75$ e $\sigma_Y = 13,66$.

Vistos sobre o prisma de uma projeção ortogonal, os conceitos de valor esperado e variância apresentam uma interpretação geométrica simples. Considere no espaço vetorial $\mathfrak{R}^n$, os vetores que denotam a v. a. X e o vetor J . Seja $P(X)$ a projeção ortogonal do vetor X sobre o vetor J , tal que $P(X) = k$, onde k é uma constante real. Nessas condições, a direção da projeção ortogonal é dada pelo vetor $X - P(X)$, que é ortogonal ao subespaço vetorial determinado pelo vetor J e, consequentemente:

$$\langle X - P(X), J \rangle = 0 \Rightarrow \langle X, J \rangle = \langle P(X), J \rangle \Rightarrow E(X) = E(P(X)) \Rightarrow \mu_X = P(X)$$

Figura 12. A esperança matemática como uma projeção ortogonal.

A Figura 12 mostra que a esperança matemática de uma variável aleatória X nada mais é do que um caso particular de projeção ortogonal. Do ponto de vista geométrico, a direção dessa projeção é a menor distância entre X e J , e cujo comprimento quadrático é probabilisticamente a variância da v. a. X . De modo semelhante ao da esperança, é possível verificar com facilidade alguns teoremas envolvendo a variância de uma variável aleatória pela aplicação imediata da definição

$$\sigma_X^2 = \|X - \mu_X\|^2:$$

a) $Var(k) = 0$, onde k é uma constante real;

$$Prova: Var(k) = \|X - \mu_X\|^2 = \|k - k\|^2 = 0.$$

b) $Var(kX) = k^2 Var(X)$;

$$Prova: \|kX - kE(X)\|^2 = \|k(X - E(X))\|^2 = |k|^2 \|X - E(X)\|^2 = k^2 Var(X).$$

c) $Var(kX + q) = k^2 Var(X)$, onde q é uma constante real;

$$Prova: \|kX + q - (kE(X) + q)\|^2 = \|kX - kE(X)\|^2 = k^2 Var(X).$$

Considerando a idéia de v.a.'s centradas, com o enfoque geométrico, o ângulo θ formado entre os vetores representativos das v.a.'s X e Y por exemplo, fornece uma importante regra que define o coeficiente de correlação na população:

$$\rho_{XY} = \cos\theta = \frac{\langle X - \mu_X, Y - \mu_Y \rangle}{\|X - \mu_X\| \|Y - \mu_Y\|}; \quad -1 \leq \rho_{XY} \leq 1,$$

conforme ilustra a Figura 13.

Figura 13. Variáveis aleatórias centradas e o coeficiente de correlação.

Retomando o exemplo do jogo, considere $x = X - \mu_X$ e $y = Y - \mu_Y$ os vetores que representam as variáveis aleatórias centradas:

$$x = [33/2, 21/2, -27/2] \quad \text{e} \quad y = [-33/2, -21/2, 27/2]$$

e usando $\sigma_X = \sigma_Y = 13,66$, obtém-se o coeficiente de correlação:

$$\rho_{xy} = \cos\theta = \frac{\langle x, y \rangle}{\|x\| \|y\|} = \frac{\langle (33/2, 21/2, -27/2), (-33/2, -21/2, 27/2) \rangle}{(13,66)(13,66)} = \frac{-186,75}{(13,66)^2} = -1$$

e, com base neste resultado, afirma-se estatisticamente que as variáveis tem correlação linear negativa perfeita. Geometricamente, isto significa que os vetores x e y são colineares, como pode ser observado na Figura 11 e através do ângulo formado entre eles:

$$\theta = \arccos(-1) = 180^\circ$$

Não perdendo de vista a idéia de v.a.'s centradas, considere x e y os vetores que as representam geometricamente, como no exemplo. Na Teoria da Probabilidade, é bastante conhecido o Teorema da Variância de uma soma de v.a.'s:

$$\text{Var}(x + y) = \text{Var}(x) + \text{Var}(y) + 2\text{Cov}(x, y) \quad (4.2.5)$$

Em (4.2.5) a covariância é igual a esperança conjunta de x e y , pois as v.a.'s estão centradas em suas esperanças e, assim, o Teorema clássico (4.2.5) pode ser interpretado geometricamente como uma decorrência direta da aplicação da Lei dos Cossenos, válida para triângulos quaisquer:

Figura 14. Variáveis aleatórias centradas e a Lei dos Cossenos.

Aplicando a Lei dos Cossenos ao triângulo ABC da Figura 14:

$$\|x + y\|^2 = \|x\|^2 + \|y\|^2 - 2\|x\| \cdot \|y\| \cdot \cos(180^\circ - \theta) ;$$

mas, como $\cos(180^\circ - \theta) = -\cos\theta$. Assim:

$$\|x + y\|^2 = \|x\|^2 + \|y\|^2 + 2\|x\| \cdot \|y\| \cdot \cos\theta \quad (4.2.6)$$

A equação (4.2.6) é, portanto, a expressão geométrica do Teorema (4.2.5) e, sendo assim, se duas variáveis são não correlacionadas, em termos geométricos

equivale a dizer que os dois vetores que as descrevem são ortogonais. Nesse caso, a equação (4.2.6) reduz-se ao Teorema de Pitágoras:

$$\|x + y\|^2 = \|x\|^2 + \|y\|^2.$$

Note que no exemplo que ilustra esta seção os vetores são colineares e não é possível formar um triângulo, mas sim uma linha. Nesse caso, a variância de $(x + y)$ é nula como se pode comprovar através equação (4.2.6):

$$Var(x + y) = \|x + y\|^2 = 186,75 + 186,75 + 2.(-186,75) = 0.$$

Considerou-se anteriormente um caso particular de projeção, no qual o valor esperado de uma v. a. corresponde a sua projeção ortogonal sobre o vetor J. Na Teoria da Probabilidade, os casos genéricos de projeções são as esperanças condicionais. Suponha que X, Y, e Z sejam variáveis aleatórias distribuídas conjuntamente com função de distribuição $F(x,y,z)$. Considere também que x , y e z são vetores que representam essas variáveis aleatórias centradas em seus respectivos valores esperados.

Nessas condições, Bryant (1984) mostra que o valor esperado de z dado x e y , denotado por $E(z/x,y)$, na visão geométrica é equivalente à projeção de z sobre o plano determinado por x , y . O plano x , y , por conseguinte, é fechado em relação às operações lineares da adição de vetores e multiplicação de vetor por escalar e dessa forma pode ser pensado como um subespaço vetorial.

Figura 15. Esperança condicional.

Considere três importantes propriedades da Esperança condicional:

- a) $E(z/x,y)$ é uma função de x,y ;
- b) $E[E(z/x,y)] = E(z)$;
- c) $E[g(z)/x,y] = gE(z/x,y)$.

Ademais, se g é uma função de x e y tem-se que:

$$\langle g, z - E(z/x,y) \rangle = \langle g, z \rangle - \langle g, E(z/x,y) \rangle = E[g(z)] - E[gE(z/x,y)] = 0$$

Do ponto de vista da Geometria, a primeira propriedade afirma que $E(z/x,y)$ está no plano L , enquanto que $\langle g, z - E(z/x,y) \rangle = 0$ diz que $z - E(z/x,y)$ é ortogonal a qualquer vetor pertencente ao subespaço L . Então, existe um projetor ortogonal P tal que $P(z) = E(z/x,y)$, podendo ser usado para prever um futuro valor de z em função dos valores já conhecidos x e y .

Outro conhecido Teorema da Teoria da Probabilidade estabelece que:

$$Var(z) = Var[E(z/x,y)] + E[Var(z/x,y)] \quad (4.2.7)$$

visto que $E[z - E(z/x,y)] = 0$, tem-se que $Var[z - E(z/x,y)] = E[z - E(z/x,y)]^2$, ou seja, a variância condicional é a variância relativa à distribuição condicional e a igualdade $E(z) = E[E(z/x,y)]$ implica que $Var[z - E(z/x,y)]$ é também igual a $E[Var(z/x,y)]$. Assim pode-se escrever a variância (4.2.7) como:

$$Var(z) = Var[E(z/x,y)] + Var[z - E(z/x,y)] \quad (4.2.8)$$

e, a expressão (4.2.8) nada mais é do que uma aplicação direta do Teorema de Pitágoras ao triângulo mostrado na Figura 15.

4.3 Tópicos de inferência estatística

4.3.1 Estimação de mínimos quadrados

Nas situações clássicas de análise de dados, como por exemplo a análise da variância e da regressão, um modelo é assumido e os dados são coletados. Isso fornece um espaço ao qual pertence o vetor de observações $y = [y_1, y_2, \dots, y_n]'$ e um subespaço deste espaço determinado pelo modelo assumido, que corresponde a alguma suposição sobre os dados e a relação dos dados observados com esse subespaço. Além disso, supõe-se que há um vetor de parâmetros $\theta = (\theta_1, \theta_2, \dots, \theta_k)$ pertencente a $\mathfrak{R}^k$, onde $k \leq n$, e uma matriz conhecida ${}_n X_k$. Assim, tem-se o modelo linear $y = X\theta + e$, onde e é um vetor de variáveis aleatórias não observáveis, tal que $e \sim N(\phi, I\sigma^2)$, conforme descrito na seção 3.7. Um dos objetivos da inferência é estimar θ ou funções paramétricas de θ .

O método dos mínimos quadrados consiste em encontrar o mínimo local da função:

$$Q = \sum_{i=1}^n (y_i - \sum_{j=1}^k x_{ij} \theta_j)^2,$$

ou seja, de:

$$Q = \|y - X\theta\|^2.$$

O ajuste pelo método dos mínimos quadrados tem uma interpretação geométrica simples e bem conhecida na literatura estatística. O vetor y pode ser observado como um ponto do espaço cartesiano n -dimensional. Para um dado valor de θ o vetor $X\theta$ de valores ajustados é um ponto no mesmo espaço. Como θ varia livremente sobre todos os valores que ele pode assumir, $X\theta$ determina um subespaço linear ou um hiperplano denominado *loco de solução*. Sendo assim, teoricamente é possível pensar que se y está no *loco de solução*, os valores observados podem ser

reproduzidos exatamente pelo modelo. Entretanto, isso raramente ocorre e nenhum valor de θ reproduz exatamente y , pois $y = X\theta$ é inconsistente e, em geral, utilizam-se as soluções aproximadas de mínimos quadrados. Se $X\theta$ representa um ponto no loco de solução então $\|y - X\theta\|^2$ é o quadrado da distância euclidiana entre y e $X\theta$. O método dos mínimos quadrados, consiste em escolher o ponto $X\theta^0$ que torne mínima a distância euclidiana entre y e o subespaço de projeção, nesse momento dado por $C(X)$.

Minimizar Q corresponde, então, a encontrar o ponto desse subespaço mais próximo a y . E a resposta é obtida pela projeção ortogonal de y sobre esse subespaço. Sendo assim, se $X\theta^0$ denota essa projeção, então $X\theta^0$ é único e satisfaz:

$$\|y - X\theta^0\|^2 \leq \|y - X\theta\|^2 \quad \forall \theta \in \mathbb{R}^k,$$

onde, conforme também visto em 3.7, $\hat{y} = X\theta^0$ é a aproximação de mínimos quadrados para y .

Um exemplo simples e de fácil compreensão é mostrado na seção 4.1.1, no qual a média amostral, $\bar{y}$, é o estimador de mínimos quadrados para o parâmetro populacional μ , tal que $\hat{y} = \bar{y}J$ é dado pela projeção ortogonal do vetor y dos dados, sobre o vetor J de “uns”.

4.3.2 Estimador de máxima verossimilhança para μ

Considere um experimento que forneceu a amostra aleatória $y_1, y_2, \dots, y_n$. Essa amostra aleatória foi uma realização da variável aleatória Y , onde $Y_i = \mu + e_i$. Tipicamente supõe-se que os erros são independentes e igualmente distribuídos. Do ponto de vista da Estatística Matemática esse modelo é chamado de modelo de locação a uma amostra, onde, em geral, $e_i \sim N(0, \sigma^2)$. Sendo assim, $Y_i \sim N(\mu, \sigma^2)$. Nesse contexto, a variável aleatória Y é caracterizada por uma distribuição de probabilidade que pertence a uma família $\mathfrak{F}$ de distribuições normais e, que possui um número finito de

parâmetros $\theta = (\mu, \sigma^2)$, desconhecidos. Como se sabe, no caso em que as amostras são independentes, a função de verossimilhança pode ser definida como:

$$L(\theta) = \prod_{i=1}^n f(y_i, \theta),$$

onde $f(y, \theta)$ é a função densidade ou função de probabilidade dado θ . A estimativa de máxima verossimilhança (EMV), é o valor de θ que maximiza $L(\theta)$. É possível mostrar que sob normalidade o estimador de mínimos quadrados $\hat{\mu} = \bar{y}$ é também um estimador de máxima verossimilhança para μ .

Suponha que se conheça a verdadeira média populacional μ , então, a função de verossimilhança é dada por:

$$L(\theta) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\frac{1}{2} \frac{\|y - \mu\|^2}{\sigma^2}}.$$

Esta equação depende somente da distância $\|y - \mu\|$, assim é esfericamente simétrica sobre o verdadeiro vetor de parâmetros do modelo $\mu = [\mu, \mu, \dots, \mu]'$, como ilustra a Figura 16.

Figura 16. Contornos da função densidade de probabilidade, centrados no verdadeiro

vetor do modelo, $\mu = [\mu, \mu, \dots, \mu]'$.

Saville & Wood (1991) salientam que é possível pensar na Figura 16 como uma distribuição normal familiar, mas expandida em todas as direções: vetores de observações próximos do vetor μ são mais verossímeis ou apropriados que vetores mais afastados. Visualizando muitas amostras, pode-se ver o vetor de observações $y = [y_1, y_2, \dots, y_n]'$ variando numa “nuvem de probabilidades” que é mais densa perto do vetor μ , e torna-se menos densa à medida que se afasta do vetor $\mu = [\mu, \mu, \dots, \mu]'$.

Isso posto, pode-se pensar na seguinte questão: dado um conjunto de observações $y = [y_1, y_2, \dots, y_n]'$ da variável Y , onde não se conhece μ , como escolher μ de tal forma que se maximize a função de verossimilhança?

A resposta, na visão geométrica, é encontrar μ para minimizar a distância $\|y - \mu\|$. Isto é obtido pela projeção ortogonal do vetor y das observações sobre o vetor $J = [1, 1, \dots, 1]'$ como ilustra a Figura 17. A estimativa $\hat{\mu} = \bar{y}$ é a mais apropriada para μ , desde que o comprimento de $y - \bar{y}$ seja menor que os comprimentos de $y - \mu'$ ou $y - \mu''$, onde μ' e μ'' representam outras projeções sobre o vetor $J = [1, 1, \dots, 1]'$.

Figura 17. O estimador de máxima verossimilhança para μ .

4.3.3 A lógica geométrica dos testes de hipóteses

Segundo Saville & Wood (1991), o procedimento geométrico padrão para se testar uma hipótese de interesse envolve a comparação do comprimento quadrático da projeção do vetor de observações $y = [y_1, y_2, \dots, y_n]'$ na direção da hipótese testada, com o comprimento quadrático médio da projeção do vetor y sobre espaço do erro ou, em outras palavras, comprimento quadrático médio da direção dessa projeção. A hipótese é testada pela divisão dessas duas medidas.

Figura 18. Idéia básica sobre teste de hipóteses.

Se a hipótese é falsa, o comprimento quadrático da projeção na direção da hipótese será, em geral, consideravelmente maior que o comprimento quadrático médio da projeção sobre o espaço do erro. Se a hipótese é verdadeira então a medida anterior será, em geral, similar a posterior. Dessa forma, tem-se um método para distinguir o verdadeiro do falso, se a razão entre as medidas é “grande” decide-se que a hipótese é falsa, e se a razão é “pequena” decide-se que ela é verdadeira.

Convém salientar, também, que esse método de testar hipóteses envolvendo razões de comprimentos quadráticos médios, em todos os casos, segue a distribuição da variável aleatória:

$$F_{p,q} = \frac{\left[\frac{W_1^2 + \dots + W_p^2}{p} \right]}{\left[\frac{W_{p+1}^2 + \dots + W_{p+q}^2}{q} \right]}$$

onde $W_1, \dots, W_{p+q}$ são variáveis aleatórias independentes $N(0, \sigma^2)$. Se H_0 é verdadeira tem-se uma v.a. F central com p e q graus de liberdade.

Um caso especial da distribuição $F_{p,q}$, particularmente usado neste trabalho ocorre quando $p = 1$, e leva à variável aleatória, ou estatística t , com q graus de liberdade:

$$t_q = \frac{W_1}{\sqrt{\frac{W_2^2 + \dots + W_{q+1}^2}{q}}}$$

onde $W_1, \dots, W_{q+1}$ são variáveis aleatórias independentes $N(0, \sigma^2)$.

4.3.3.1 Teste de significância para a média

Para ilustrar o procedimento geométrico de um teste de hipóteses, considere a situação descrita na seção 4.1.1, caracterizada pelo modelo $y = \mu J + e$, e o exemplo 4.1.1.1, com uma amostra de 3 observações referente a variável Y = produção de milho, onde $Y \cap N(\mu, \sigma^2)$. Suponha que se deseje testar $H_0: \mu = 0$ contra $H_a: \mu \neq 0$. Observou-se, naquele exemplo, a partição ortogonal do vetor y . Essa partição é reproduzida na Figura 19.

Figura 19. Partição ortogonal do vetor $y = [2,1 \ 1,8 \ 2,1]'$.

Observou-se também, em 4.1.1, que no exemplo em questão, espaço do erro é bidimensional. Seja, então, $\{u_1, u_2, u_3\}$ uma base ortonormal para o $\mathfrak{R}^3$, tal que u_1 é o vetor unitário na direção do vetor J e u_2 e u_3 formam uma base ortonormal para o vetor do erro:

$$\{u_1, u_2, u_3\} = \left\{ \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \frac{1}{\sqrt{6}} \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix} \right\}$$

Aplicando o Teorema de Pitágoras (Figura 19), tem-se:

$$\|y\|^2 = \|\bar{y}\|^2 + \|\hat{e}\|^2$$

como os vetores $\bar{y}$ e $\hat{e}$ podem ser decompostos em seus componentes ortogonais, pode-se reescrever o Teorema de Pitágoras como:

$$\begin{aligned} \|y\|^2 &= \langle y, u_1 \rangle^2 + \langle y, u_2 \rangle^2 + \langle y, u_3 \rangle^2 \\ 12,06 &= 12,00 + 0,045 + 0,015 \end{aligned}$$

Figura 20. Teorema de Pitágoras aplicado a $y = \langle y, u_1 \rangle u_1 + \langle y, u_2 \rangle u_2 + \langle y, u_3 \rangle u_3$.

Conforme já discutido neste texto, o teste estatístico envolve a razão de comprimentos quadráticos de projeções, representadas no exemplo por $\langle y, u_i \rangle^2$ e, assim, torna-se necessário considerar a distribuição das variáveis aleatórias associadas: $Y \cdot u_i$. Desde que $Y \sim N(\mu, \sigma^2)$, então:

$$E(Y \cdot u_1) = \sqrt{3}\mu \quad \text{e} \quad E(Y \cdot u_2) = E(Y \cdot u_3) = 0$$

Figura 21. Distribuições das projeções $Y \cdot u_i$.

Neste exemplo há somente uma direção para o subespaço gerado pela “matriz” J do modelo: a reta equiangular na direção do vetor u_1 (vetor J). Esta é, portanto, a direção associada com a hipótese testada, conforme ilustra a Figura 21. A Figura 22 mostra a distribuição da projeção do coeficiente $Y \cdot u_1$ sob a hipótese de nulidade $H_0: \mu = 0$ e sob a hipótese alternativa $H_a: \mu \neq 0$. Se $\mu = 0$, o valor

realizado $y \cdot u_1$ será relativamente “pequeno”, enquanto que se $\mu \neq 0$ o valor realizado $y \cdot u_1$ será relativamente “grande”.

Figura 22. Distribuição da projeção Yu_1 para as hipóteses de nulidade e alternativa.

O teste de $H_0: \mu = 0$ envolve simplesmente a comparação do comprimento quadrático da projeção $\langle y, u_1 \rangle^2$, com o comprimento quadrático médio no espaço do erro. O teste estatístico é portanto:

Sob H_0 :

$$\frac{\langle y, u_1 \rangle^2}{[\langle y, u_2 \rangle^2 + \langle y, u_3 \rangle^2] / 2} \cap F_{(1,2)}$$

No presente caso tem-se:

$$F = \frac{12,00}{(0,045 + 0,015) / 2} = 400$$

e assim rejeita-se $H_0: \mu = 0$ ao nível $\alpha = 0,05$ de probabilidade, uma vez que $F_{0.05(1,2)} = 18,51$.

Dado que $\langle y, u_1 \rangle^2 = \|\bar{y}\|^2$ e $\langle y, u_2 \rangle^2 + \langle y, u_3 \rangle^2 = \|y - \bar{y}\|^2$, então o teste estatístico pode ser genericamente expresso por:

$$F = \frac{\|\bar{y}\|^2}{\|y - \bar{y}\|^2 / (n-1)} = \frac{n\bar{y}^2}{s^2} \quad (4.3.3.1)$$

onde s^2 é a variância amostral. Essa forma simplificada do teste mostra que não há necessidade de se “escolher” uma base ortonormal para o espaço do erro, mas sim de se conhecer sua dimensão. Verifica-se ainda que o teste F pode ser reescrito em termos de projetores ortogonais associados ao numerador e ao denominador. No exemplo em questão, verificou-se na seção 4.1.1 que $Py = \bar{y}$ e $\hat{e} = (I - P)y$, onde $P = J(J'J)^{-1}J'$, logo:

$$F = \frac{\|Py\|^2}{\|(I - P)y\|^2 / n - 1} = \frac{y'Py}{y'(I - P)y / n - 1} \quad (4.3.3.2)$$

A forma expressa em (4.3.3.2) será de grande utilidade nos próximos tópicos, como a análise da variância e da regressão.

Na literatura estatística, é sugerida a estatística t para testar um valor suposto da população, que é normalmente distribuída, e tem σ desconhecido. Como $t^2 = F$ com 1 grau de liberdade, pode-se reformular o teste estatístico para o exemplo em discussão:

$$F = \frac{\|\bar{y}\|^2}{\|y - \bar{y}\|^2 / (n - 1)} = \frac{n\bar{y}^2}{s^2} = \frac{\bar{y}^2}{s^2 / n} = \left(\frac{\bar{y}}{s / \sqrt{n}} \right)^2 = t^2,$$

Logo:

$$t = \frac{\|\bar{y}\|}{\|y - \bar{y}\| / \sqrt{n - 1}} = \sqrt{n - 1} \frac{\|\bar{y}\|}{\|y - \bar{y}\|}$$

e pela Figura 19:

$$t = \sqrt{n - 1} \cdot \cot g\theta$$

e, sendo $t^2 = F$:

$$F = (n - 1) \cdot \cot g^2 \theta$$

Portanto, se a cotangente é grande, θ é pequeno, e dessa forma pode-se verificar que o teste t , rotineiramente calculado, é uma medida do ângulo entre os vetores y e $\hat{y}$, aqui $\bar{y}$. Para o exemplo:

$$\cot g\theta = \frac{t}{\sqrt{n - 1}} = \frac{20}{\sqrt{2}} = 14,14 \Rightarrow \theta = \arccot g(14,14) \cong 4^\circ$$

4.3.3.2 Teste de significância para a existência de correlação linear

De um modo geral, a hipótese nula de interesse é que o coeficiente de correlação da população é igual a zero, isto é $\rho = 0$. Assim, se esta hipótese for rejeitada ao nível de significância α estipulado, pode-se concluir que efetivamente existe uma relação linear entre as variáveis. Na seção 4.1.2, mostrou-se geometricamente que o coeficiente de correlação amostral dado por $r_{xy} = \cos \theta$, é um estimador de ρ . Na literatura, a seguinte estatística amostral envolvendo r é distribuída segundo uma distribuição t com $(n - 2)$ graus de liberdade, quando $\rho = 0$:

$$t = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}}$$

Geometricamente, testar a hipótese nula de que $\rho = 0$, equivale a estudar a função cotangente do ângulo θ (Figura 10), uma vez que $r_{xy} = \cos \theta$:

$$t = \frac{\cos \theta}{\sqrt{\frac{1 - \cos^2 \theta}{n-2}}} = \sqrt{n-2} \frac{\cos \theta}{\sqrt{\sin^2 \theta}} = \sqrt{n-2} \frac{\cos \theta}{\sin \theta} = \sqrt{n-2} \cot g \theta$$

Como exemplo, considere os dados descritos no exemplo 4.1.2.1, onde se tem uma amostra de 6 observações das variáveis X e Y e obteve-se $r = -0,9203$. Com base nessa estimativa, suponha que se queira testar a existência de correlação linear negativa entre X e Y , ou seja, deseja-se testar $H_0: \rho = 0$ contra $H_a: \rho < 0$.

$$\theta = \arccos (-0,9203) \cong 157^\circ$$

e a estatística para o teste é:

$$t = \sqrt{6-2} \cot g(157^\circ) = 2.(-2,36) = -4,71$$

e assim, rejeita-se H_0 , ao nível de 0,05 de probabilidade, uma vez que

$$t_{(0,05; 4)} = -2,132.$$

4.3.4 Intervalo de confiança para a média

Embora a média amostral seja útil como um estimador não tendencioso da média populacional, a maneira de melhor expressar o grau de acurácia desse estimador por ponto dá-se através de um intervalo de confiança. Quando uma variável aleatória segue a distribuição normal, a distribuição por amostragem da média dessa variável também segue a distribuição normal. Contudo, no processo de inferência cada valor da média é convertido para um valor normal padronizado, que quando σ é desconhecido, segue a distribuição t de *Student*.

Sendo assim, considerando as distribuições das projeções $Y.u_i$ apresentadas na seção 4.3.3.1 (ver Figura 21), pode-se especificar o intervalo de confiança para μ , do exemplo em questão. Dado que $Y.u_1 = \sqrt{3}\bar{Y}$ está distribuída como $N(\sqrt{3}\mu, \sigma^2)$ então a variável aleatória $W_1 = \sqrt{3}(\bar{Y} - \mu)$ tem distribuição $N(0, \sigma^2)$, como também ocorre com as variáveis aleatórias $W_2 = Yu_2$ e $W_3 = Yu_3$. Consequentemente a variável aleatória:

$$T = \frac{\sqrt{3}(\bar{Y} - \mu)}{\sqrt{[(Y.u_2)^2 + (Y.u_3)^2] / 2}} = \frac{Y.u_1 - E(Y.u_1)}{\sqrt{\text{estimador de } \sigma^2}},$$

tem distribuição t com 2 graus de liberdade e onde u_1 é a direção associada com a hipótese $H_0: \mu = 0$. Desse modo, com base nos valores realizados do exemplo 4.1.1.1 obtém-se:

$$T = \frac{\sqrt{3}(\bar{y} - \mu)}{\sqrt{[< y, u_2 >^2 + < y, u_3 >^2] / 2}} = \frac{\sqrt{3}(\bar{y} - \mu)}{\sqrt{\|y - \bar{y}\|^2 / 2}} = \frac{\bar{y} - \mu}{\sqrt{s^2 / 3}} = \frac{2 - \mu}{\sqrt{0,03 / 3}}$$

e o intervalo com nível de confiança $1 - \alpha = 0,95$ para μ é:

$$\bar{y} - \frac{s}{\sqrt{3}}t_2 \leq \mu \leq \bar{y} + \frac{s}{\sqrt{3}}t_2 \Rightarrow 1,5697 \leq \mu \leq 2,4303, \text{ onde } t_{(0,05; 2)} = 4,303.$$

Generalizando: $\bar{y} - \frac{s}{\sqrt{n}}t_{n-1} \leq \mu \leq \bar{y} + \frac{s}{\sqrt{n}}t_{n-1}.$

4.4 Análise da Variância

Considere o modelo linear $y = X\theta + e$, onde y é o vetor de observações, X é a matriz conhecida do delineamento, θ é o vetor de parâmetros e e é um vetor de variáveis aleatórias não observáveis tal que $E[e] = \phi$ e $Var[e] = I\sigma^2$.

Como já é sabido, o sistema de equações normais $X'X\theta = X'y$ é sempre consistente e suas soluções exatas são soluções aproximadas de mínimos quadrados para o sistema $y = X\theta$, que normalmente é inconsistente. Tais soluções podem ser obtidas por: $\theta^0 = (X'X)^G X'y$, onde $(X'X)^G$ é uma inversa generalizada qualquer de $X'X$.

Em grande parte dos casos a matriz X não tem posto coluna completo e o sistema de equações normais, embora consistente, é indeterminado. Consequentemente θ^0 não tem valor por si só, mas sim associado a X através de $\hat{y} = X\theta^0 = X(X'X)^G X'y = Py$.

Como visto na seção 3.7.2, a matriz $P = X(X'X)^G X' = XX^l = XX^+$, é uma matriz simétrica e idempotente, conhecida como o projetor ortogonal de y sobre o espaço coluna da matriz X . Como a matriz X do delineamento é conhecida, a obtenção de P é sempre facilitada. Além disso, a unicidade da inversa de Moore-Penrose garante a unicidade de P e a invariância de $\hat{y}$ para qualquer inversa generalizada de $X'X$ ou qualquer inversa de mínimos quadrados de X .

Por outro lado, o erro devido à aproximação de mínimos quadrados é dado por $\hat{e} = y - \hat{y} = (I - P)y$, sendo que $\hat{e}$ pertence ao complemento ortogonal do espaço coluna da matriz X . Esse resultado é particularmente importante dentro da Estatística, pois garante a verdadeira partição do vetor y de observações, na soma de dois vetores: $\hat{y} \in C(X)$ e $\hat{e} \in C^\perp(X)$, sendo $C(X) \oplus C^\perp(X) = \mathbb{R}^n$, exemplificando de modo objetivo as idéias metodológicas sobre projeções discutidas na seção 3.4.

Figura 23. Análise da variância: decomposição ortogonal do vetor y .

Dada a ortogonalidade, pode-se aplicar o Teorema de Pitágoras e com isso obter a decomposição ortogonal clássica da análise da variância:

$$\|y\|^2 = \|\hat{y}\|^2 + \|\hat{e}\|^2$$

onde:

$$\|y\|^2 = \langle y, y \rangle = y'y = \sum_{i=1}^n y_i^2 = \text{SQ}_T \text{ não corrigida}$$

$$\|\hat{y}\|^2 = \langle \hat{y}, \hat{y} \rangle = \hat{y}'\hat{y} = y'Py = \text{SQ}_0$$

$$\|\hat{e}\|^2 = \langle \hat{e}, \hat{e} \rangle = \hat{e}'\hat{e} = y'(I - P)y = \text{SQ}_R$$

4.4.1 Exemplo

A fim de ilustrar a abordagem geométrica através da noção de projeção ortogonal na análise da variância, considere um conjunto de apenas três dados mostrados na Tabela 3.

Tabela 3. Dados de um experimento inteiramente casualizado.

	i=1	i=2
j = 1	4	1
j = 2	3	-

Fonte: Iemma & Palm (1992), p.17.

Sob a forma matricial, o modelo $y = X\theta + e$ (G.M.) caracterizado por $y_{ij} = \mu + t_i + e_{ij}$:

$$\begin{bmatrix} 4 \\ 3 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ t_1 \\ t_2 \end{bmatrix} + \begin{bmatrix} e_{11} \\ e_{12} \\ e_{21} \end{bmatrix}$$

A matriz de projeção P é:

$$P = X(X'X)^G X',$$

sendo:

$$X'X = \begin{bmatrix} 3 & 2 & 1 \\ 2 & 2 & 0 \\ 1 & 0 & 1 \end{bmatrix} \quad (X'X)^- = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1/2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Logo,

$$P = \begin{bmatrix} 1/2 & 1/2 & 0 \\ 1/2 & 1/2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

consequentemente a aproximação de mínimos quadrados fica:

$$\hat{y} = Py = \begin{bmatrix} 1/2 & 1/2 & 0 \\ 1/2 & 1/2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 4 \\ 3 \\ 1 \end{bmatrix} = \begin{bmatrix} 7/2 \\ 7/2 \\ 1 \end{bmatrix}$$

e o vetor dos erros:

$$\hat{e} = y - \hat{y} = (I - P)y = \begin{bmatrix} 4 \\ 3 \\ 1 \end{bmatrix} - \begin{bmatrix} 7/2 \\ 7/2 \\ 1 \end{bmatrix} = \begin{bmatrix} 1/2 \\ -1/2 \\ 0 \end{bmatrix}$$

A Figura 24 mostra a decomposição ortogonal do vetor y do espaço tridimensional. Observe que o espaço gerado pelas colunas de X é um plano, pois X é de posto coluna incompleto igual a 2, constituindo-se no subespaço sobre o qual está a projeção ortogonal do vetor y , o vetor $\hat{y}$. Para representar esse plano basta escolher dois vetores L.I. dentre as colunas de X , como por exemplo:

$$X_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \quad X_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

Figura 24. Projeção ortogonal do vetor y no espaço coluna de X : visão no 1º octante.

Fonte: Iemma & Palm (1992), p.18.

É fácil ver que o subespaço dos parâmetros $C(X)$ e o subespaço dos erros $C^\perp(X)$ são ortogonais, uma vez que o produto escalar:

$$\langle \hat{y}, \hat{e} \rangle = \hat{y}'\hat{e} = \begin{bmatrix} 7/2 & 7/2 & 1 \end{bmatrix} \begin{bmatrix} 1/2 \\ -1/2 \\ 0 \end{bmatrix} = 0$$

A decomposição do vetor y em componentes ortogonais $\hat{y}$ e $\hat{e}$, permite a partição da soma de quadrados total em uma soma de quadrados devida aos parâmetros do modelo e outra devida ao resíduo:

$$y'y = y'Py + y'(I-P)y$$

$$SQ_T = SQ_\theta + SQ_R$$

$$26 = 25,5 + 0,5$$

A soma de quadrados dos parâmetros, por sua vez, pode ser partida em dois componentes: a soma de quadrados devida ao parâmetro μ , (SQ_μ), e a soma de quadrados devida aos tratamentos, quando μ já é considerado no modelo ($SQ_{t/\mu}$). É a chamada parametrização sucessiva do modelo, que em termos de projetores ortogonais é obtida pela partição da matriz do delineamento, como se segue:

$$X = [X_1 \mid X_2]$$

onde $X_1 = J$ é a 1ª coluna de X , cujos elementos são todos iguais a unidade e X_2 corresponde às demais colunas. Assim:

$$P_1 = X_1(X_1'X_1)^{-1}X_1' = \frac{1}{n}(X_1X_1') = \frac{1}{n}JJ'$$

e

$$P_2 = P - P_1$$

Então,

$$SQ_\mu = y'P_1y$$

$$SQ_{t/\mu} = y'P_2y$$

Para o exemplo em questão:

$$P_1 = \frac{1}{3} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \quad P_2 = \frac{1}{6} \begin{bmatrix} 1 & 1 & -2 \\ 1 & 1 & -2 \\ -2 & -2 & 4 \end{bmatrix}$$

Dessa forma:

$$SQ_{\mu} = y' P_1 y = 21,34$$

$$SQ_{v/\mu} = y' P_2 y = 4,16$$

Os graus de liberdade, associados a essas somas de quadrados, são iguais às dimensões dos espaços vetoriais correspondentes e são dados pelos postos das matrizes de projeção. Desse modo, o número de graus de liberdade da SQ_{θ} é igual ao posto da matriz P que, por sua vez é igual ao posto da matriz X , devido a uma propriedade da inversa generalizada de Moore-Penrose: $[r(X) = r(X^+) = r(XX^+)]$.

Com isso, é possível formalizar o Quadro da Análise da Variância em função de conceitos geométricos como projetores ortogonais e dimensões de espaços vetoriais.

Tabela 4. Quadro da Análise da Variância.

C.V.	G.L.	SQ	E[SQ]	E[QM]	H ₀
Parâmetros	$r[P] = r[X]$	$y'Py$	$r[P]\sigma^2 + \theta'X'X\theta$	$\sigma^2 + \{1/r[P]\}\theta'X'X\theta$	$H_0 : \theta'X'X\theta = 0$
Resíduo	$r[I-P] = n-r[X]$	$y'(I-P)y$	$\{n - r[P]\}\sigma^2$	σ^2	
Total	n	$y'y$	$n\sigma^2 + \theta'X'X\theta$	$\sigma^2 + \{1/n\}\theta'X'X\theta$	

A Tabela 4, refere-se a um modelo geral sem considerar parametrizações. As esperanças matemáticas decorrem da hipótese de que y é um vetor

de observações provenientes de uma distribuição multinormal, de média $\mu = X\theta$ e matriz de variâncias e covariâncias $I\sigma^2$. Nessas condições, a esperança matemática de uma soma de quadrados, descrita como uma forma quadrática $Q(y) = y' M y$ é, conforme Searle (1971):

$$E[y' M y] = \text{Tr}[M] \sigma^2 + \theta' X' M X \theta$$

mas sendo M simétrica e idempotente, então $\text{Tr}[M] = r[M]$. Deste modo:

$$E[SQ_T] = E[y' I y] = \text{Tr}[I] \sigma^2 + \theta' X' I X \theta = n \sigma^2 + \theta' X' X \theta$$

$$E[SQ_0] = E[y' P y] = \text{Tr}[P] \sigma^2 + \theta' X' P X \theta = r[P] \sigma^2 + \theta' X' X \theta$$

$$E[SQ_R] = E[y' (I - P) y] = \text{Tr}[I - P] \sigma^2 + \theta' X' (I - P) X \theta = \{n - r[P]\} \sigma^2$$

É evidente que efetuando-se partições sobre a matriz X e o vetor de parâmetros θ podem ser obtidas as esperanças referentes aos subconjuntos de parâmetros do modelo. Um outro procedimento também de grande utilidade consiste na obtenção da soma de quadrados para uma hipótese de interesse. Assim, para a hipótese $H_0: \lambda' \theta = \phi$, onde $\lambda' \theta$ é estimável e λ' tem posto linha completo, tem-se:

$$SQH_0: (\lambda' \theta^0)' [\lambda' (X' X)^G \lambda]^{-1} (\lambda' \theta^0)$$

onde θ^0 é qualquer solução das equações normais e verifica-se que:

$$\frac{SQH_0}{\sigma^2} \cap \chi^2_{[r[\lambda], \delta]}$$

onde $\delta = \frac{1}{2\sigma^2} \theta' X' P^* X \theta$ é o parâmetro de não centralidade e P^* é o projetor ortogonal tal $y' P^* y = SQH_0$. Além disso, sob normalidade, covariância nula é condição necessária e suficiente para independência, então, as formas quadráticas associadas a $P^*(I - P^*) = \phi$, são independentes e o quociente dos respectivos quadrados médios tem distribuição F não central.

$$\frac{QMH_0}{QMRes} \cap F_{\{r[\lambda]; r[I-P]; \delta\}}$$

Desse modo, com base no parâmetro de não centralidade δ ou com base na $E[SQH_0]$ pode-se verificar quais as verdadeiras hipóteses que estão sendo testadas no exemplo em questão, conforme descrito na Tabela 5.

Tabela 5. Esperanças das somas de quadrados e hipóteses naturais.

C.V.	G.L.	E(SQ)	δ	H_0
Parâmetros	$r [P] = (2)$	$2\sigma^2 + 2(\mu + t_1)^2 + (\mu + t_2)^2$	$\frac{2(\mu + t_1)^2 + (\mu + t_2)^2}{2\sigma^2}$	$H_0^1: \mu + t_1 = 0$
Constante	$r [P_1] = 1$	$\sigma^2 + 3\mu^2$	$\frac{3}{2\sigma^2} \mu^2$	$H_0^2: \mu = 0$
Tratamentos	$r [P_2] = 1$	$\sigma^2 + \frac{2}{3} [t_1 - t_2]^2$	$\frac{1}{3\sigma^2} (t_1 - t_2)^2$	$H_0^3: t_1 = t_2$
Resíduo	$r [I - P] = 1$	σ^2	0	
Total	3	$3\sigma^2 + 2(\mu + t_1)^2 + (\mu + t_2)^2$		

O exemplo considerado tem apenas o apelo didático, no sentido de evidenciar as noções de espaço vetorial e projeções ortogonais, relacionando-as na análise da variância. O procedimento é análogo para qualquer tipo de delineamento experimental, seja ele balanceado ou não.

4.4.2 Projetores ortogonais associados a hipóteses de interesse

A obtenção de projetores ortogonais a partir de hipóteses mais específicas não é trivial e pode gerar dificuldades, principalmente se os dados forem não balanceados e/ou apresentarem caselas vazias.

Sobre esse aspecto, Iemma et al. (1993) apresentam um procedimento simples e imediato para se obter o projetor ortogonal associado a uma hipótese H_0 , para qualquer nível de desbalanceamento dos dados.

A proposta dos autores baseia-se na seguinte definição de Rao (1945): “uma função linear paramétrica, $\lambda'\theta$, é dita estimável no modelo $y = X\theta + e$, onde $E[y] = X\theta$ e $V[y] = I\sigma^2$; se e somente se, existe ao menos uma combinação linear das observações, $a'y$, tal que $E[a'y] = \lambda'\theta$ ”.

Como qualquer combinação linear dos valores observados, $a'y$, tem $E[a'y] = a'X\theta$, pode-se também escrever que: $a'X\theta = \lambda'\theta$. Sendo assim, $\lambda'\theta$ é estimável se e somente se:

$$X'a = \lambda, \quad (4.4.2.1)$$

ou seja, se $\lambda \in C(X')$ ou $\lambda' \in L(X)$. Neste caso, $a = X^+ \lambda$ é uma solução para o sistema (4.4.2.1), pois $\lambda' \in L(X)$ e por conseguinte $X^+X\lambda = \lambda$.

Por outro lado, considere a hipótese $H_0 = \lambda'\theta = \phi$, λ' de posto linha completo e sua respectiva soma de quadrados de hipóteses:

$$SQH_0 = (\lambda'\theta^0)' [\lambda' (X'X)^G \lambda]^{-1} (\lambda'\theta^0),$$

onde θ^0 é qualquer solução do sistema de equações normais. Então, pelo Teorema de Gauss-Markov $\lambda'\theta^0$ é o *blue* de $\lambda'\theta$. Ademais, $\lambda'\theta^0$ é invariante para qualquer escolha de $(X'X)^G$. Sendo assim:

$$SQH_0 = y'X [(X'X)^G]' \lambda [\lambda' (X'X)^G \lambda]^{-1} \lambda' (X'X)^G X'y$$

como $X'a = \lambda$, segue que:

$$SQH_0 = y'X [(X'X)^G]' X'a [a'X(X'X)^G X'a]^{-1} a'X(X'X)^G X'y$$

$$\therefore SQH_0 = y' XX^+ a [a' XX^+ a]^{-1} XX^+ y$$

como $a = X^+ \lambda$, tem-se que:

$$SQH_0 = y' XX^+ X^+ \lambda [\lambda' X^+ XX^+ X^+ \lambda]^{-1} \lambda' X^+ XX^+ y$$

$$SQH_0 = y' X^+ \lambda [\lambda' X^+ X^+ \lambda]^{-1} \lambda' X^+ y$$

$$SQH_0 = y' a [a' a]^{-1} a' y$$

$$SQH_0 = y' P_{H_0} y$$

Portanto $P_{H_0} = a(a'a)^{-1}a'$ é uma forma simples e prática, apresentada por Iemma et al. (1993), para se obter um projetor ortogonal associado à hipótese H_0 .

4.4.2.1 Exemplo

Para ilustrar a idéia de projetor ortogonal associado a uma hipótese, considere o conjunto de dados descritos na Tabela 6, relativo a um experimento fictício envolvendo duas variedades de peras e três espaçamentos entre plantas.

Tabela 6. Dados, em Kg, de um experimento fatorial 2×3 .

	j = 1	j = 2	j = 3
i = 1	62	70	45
	---	78	58
i = 2	90	---	63
	87	---	---

Considerando esse conjunto de dados, serão construídos projetores ortogonais, testando duas hipóteses, dentre outras possíveis. Uma delas refere-se ao efeito de linhas e a outra sobre a interação, como sugere o procedimento *General Linear Models* (GLM) do sistema estatístico SAS. A fim de simplificar os cálculos, adota-se aqui o modelo de médias de caselas, o qual conforme Searle (1987), apresenta-

se na forma: $y = W\mu + e$, sendo $E[e] = \phi$ e $V[e] = I\sigma^2$, cuja classificação dupla cruzada e interação se caracteriza por: $y_{ijk} = \mu_{ij} + e_{ijk}$. Sob a forma matricial:

$$\begin{bmatrix} y_{111} \\ y_{121} \\ y_{122} \\ y_{131} \\ y_{132} \\ y_{211} \\ y_{212} \\ y_{231} \end{bmatrix} = \begin{bmatrix} 62 \\ 70 \\ 78 \\ 45 \\ 58 \\ 90 \\ 87 \\ 63 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_{11} \\ \mu_{12} \\ \mu_{13} \\ \mu_{21} \\ \mu_{23} \end{bmatrix} + \begin{bmatrix} e_{111} \\ e_{121} \\ e_{122} \\ e_{131} \\ e_{132} \\ e_{211} \\ e_{212} \\ e_{231} \end{bmatrix}$$

$$W^+ = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 0 & 0 \\ 0 & 0 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 0 & 1/2 \end{bmatrix}$$

Como W é de posto coluna completo, a obtenção de W^+ é imediata.

Hipóteses:

i) Hipótese sobre médias ponderadas de linhas, não ajustadas

$$H_0: \frac{\mu_{11} + 2\mu_{12} + 2\mu_{13}}{5} = \frac{2\mu_{21} + \mu_{23}}{3}$$

$$H_0: \lambda\mu = \phi \Rightarrow \lambda = [1/5 \quad 2/5 \quad 2/5 \quad -2/3 \quad -1/3]$$

$$a = W^+ \lambda \Rightarrow a' = [1/5 \quad 1/5 \quad 1/5 \quad 1/5 \quad 1/5 \quad -1/3 \quad -1/3 \quad -1/3]$$

$$P_{H_0} = a(a'a)^{-1}a'$$

$$P_{H_0} = \begin{bmatrix} & & & & & & & & \\ & & & & & & & & \\ & & & & & & & & \\ & & & & & & & & \\ & & & & & & & & \\ & & & & & & & & \\ & & & & & & & & \\ & & & & & & & & \\ & & & & & & & & \end{bmatrix}$$

cuja soma de quadrados correspondente é: $SQH_0 = y' P_{H_0} y = 567,675$.

ii) Hipótese sobre interação

$$H_0 = \mu_{11} - \mu_{13} - \mu_{21} + \mu_{23} = 0$$

$$H_0: \lambda'\mu = \phi \Rightarrow \lambda' = [1 \quad 0 \quad -1 \quad -1 \quad 1]$$

$$a = W^{++}\lambda \Rightarrow a' = [1 \quad 0 \quad 0 \quad -1/2 \quad -1/2 \quad -1/2 \quad -1/2 \quad 1]$$

$$P_{H_0} = a(a'a)^{-1}a'$$

$$P_{H_0} = \frac{1}{12} \begin{bmatrix} 4 & 0 & 0 & -2 & -2 & -2 & -2 & 4 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -2 & 0 & 0 & 1 & 1 & 1 & 1 & -2 \\ -2 & 0 & 0 & 1 & 1 & 1 & 1 & -2 \\ -2 & 0 & 0 & 1 & 1 & 1 & 1 & -2 \\ -2 & 0 & 0 & 1 & 1 & 1 & 1 & -2 \\ 4 & 0 & 0 & -2 & -2 & -2 & -2 & 4 \end{bmatrix}$$

cuja soma de quadrados correspondente é: $SQH_0 = y' P_{H_0} y = 75,00$.

4.5 Regressão linear

Seja o modelo de regressão linear múltipla $y = X\beta + e$, onde y é um vetor de observações de dimensões $(n \times 1)$; X é uma matriz conhecida dos valores das variáveis regressoras de dimensão $(n \times p)$; β é um vetor $(p \times 1)$ de parâmetros desconhecidos do modelo e e é um vetor $(n \times 1)$ de erros aleatórios.

Sem perda de generalidade, admitindo-se que a estrutura estocástica do vetor de erros tem valor esperado, $E[e] = \phi$ e matriz de variâncias e covariâncias, $V[e] = I\sigma^2$, então o modelo acima é definido tal qual na seção 3.7, como um modelo linear de Gauss-Markov.

O uso e a interpretação de um modelo de regressão linear múltipla frequentemente depende das estimativas dos coeficientes de regressão individuais. Além de estimar por ponto, o ajuste de um modelo de regressão tem, em geral, por objetivos básicos, realizar inferências que incluem testes de hipóteses e intervalos de confiança, predição e seleção de um conjunto de variáveis apropriadas para o modelo.

4.5.1 Uma visão geométrica da regressão

Como exemplo, considere o modelo de regressão linear com apenas duas variáveis regressoras, isto é:

$$y = \beta_1 x_1 + \beta_2 x_2 + e,$$

onde y , x_1 , x_2 são vetores colunas representando desvios em relação às médias das variáveis e e é um vetor de erros aleatórios. A Figura 25, mostra a regressão do vetor dependente y sobre o subespaço gerado pelos vetores independentes x_1 e x_2 . O vetor $\hat{y}$ aparece como combinação linear de x_1 e x_2 e de acordo com o princípio de mínimos quadrados está localizado a uma distância mínima de y .

Figura 25. Projeção ortogonal do vetor y sobre o espaço estimação $C(X)$.

Figura 26. Visão geométrica de uma superfície de regressão.

Uma outra abordagem, cuja representação gráfica está reproduzida na Figura 26, foi mostrada por Mandel (1982). Assumindo o mesmo modelo de regressão, imagina-se que os pontos (x_1, x_2) podem ser plotados em um plano $L = C(X)$. Em cada ponto, traça-se um segmento perpendicular ao plano L de altura y , onde *em* y estão os valores da variável resposta. Note que os extremos dos segmentos não estão exatamente no mesmo plano visto que y é acompanhado de erro experimental. Neste caso, devido aos erros em y , diz-se que R , o verdadeiro plano de resposta, não pode ser determinado exatamente, mas sim ajustado num plano chamado R_F .

4.5.2 Análise da regressão

Geometricamente, a análise da variância da regressão é a expressão do Teorema de Pitágoras, aplicado ao triângulo de lados y , $\hat{y}$ e $\hat{e}$ (Figura 25), representado assim, por analogia, as mesmas regras discutidas na seção 4.4. Desse modo, tomando um modelo no qual as variáveis são centradas, como o exemplo anterior, então:

$$\|y\|^2 = \langle y, y \rangle = y'y = \sum_{i=1}^n y_i^2 = SQ_{total}$$

$$\|\hat{y}\|^2 = \langle \hat{y}, \hat{y} \rangle = \hat{y}'\hat{y} = y'Py = SQ_{Regressão}$$

$$\|\hat{e}\|^2 = \langle \hat{e}, \hat{e} \rangle = \hat{e}'\hat{e} = y'(I - P)y = SQ_{Residual}$$

Outra medida importante na análise de regressão é o coeficiente de determinação, que geometricamente é expresso por:

$$R^2 = \frac{SQ_{Reg}}{SQ_{Tot}} = \frac{\|\hat{y}\|^2}{\|y\|^2} = \cos^2(\theta); \quad 0 \leq R^2 \leq 1,$$

onde θ é o ângulo agudo formado entre os vetores y e $\hat{y}$.

4.5.3 Teste da hipótese de exclusão de variável

Seja o modelo de regressão linear simples $y = \alpha J + \beta x + e$, onde J é um vetor de “uns” e x é um vetor de variáveis regressoras composto de variáveis centradas. Com as condições de Gauss-Markov em e , tem-se que $E[y] = \alpha J + \beta x$ e $V[y] = I\sigma^2$.

A média de y , $E[y]$, pertence ao subespaço linear bidimensional L , do espaço amostral. A hipótese típica H_0 , poderá ser a de que $E[y]$ pertence a um subespaço unidimensional M de L .

Esta hipótese descreve uma exclusão de variáveis. Por exemplo, suponha que se deseje testar a hipótese de que o coeficiente de inclinação do modelo é zero, isto é:

$$H_0 : \beta = 0 \quad \text{contra} \quad H_a : \beta \neq 0$$

O teste dessa hipótese envolve duas projeções ortogonais de y . De acordo com o modelo de regressão no espaço bidimensional, o vetor $\hat{y}$ deve pertencer ao subespaço L , formado por $k = 2$ vetores independentes, J e x . Porém, de acordo com a hipótese de nulidade, ele deve pertencer ao subespaço M formado por somente um único vetor J representando as $k - 1 = 1$ variáveis independentes não nulas.

A Figura 27 descreve os vários vetores, subespaços e projeções. O vetor $\hat{y}$ é a projeção ortogonal de y sobre L , e $\hat{y}_1$ é a projeção de y sobre M . O teste de hipóteses é baseado nas distâncias médias quadráticas de $\hat{y}$ a $\hat{y}_1$ e deles a y . Quanto mais próximo $\hat{y}$ é de $\hat{y}_1$ mais provável que se erre na rejeição da hipótese nula, dado que essa distância é fixa para y . Apresentam-se segundo Margolis (1979), três passos para o teste de hipóteses:

i) Obtenção das equações normais sem restrição:

$$[Jx]' \begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \end{bmatrix} = [Jx]' y$$

De onde resulta:

$$y'P_{[Jx]}y = \begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \end{bmatrix}' [Jx]' y$$

A soma de quadrados residual baseada na hipótese de que ambos os coeficientes são diferentes de zero é:

$$y' (I - P_{[Jx]})y = SQR_{NR}$$

onde SQR_{NR} significa “soma de quadrados residual sem restrição”. A SQR_{NR} tem $n - 2$ graus de liberdade e é representada pela distância quadrática do vetor y ao vetor $\hat{y}$ na Figura 27.

ii) No modelo restrito pela hipótese de nulidade, isto é:

$$y = \alpha J + e$$

obtem-se o sistema restrito de equações normais como: $J'J\hat{\alpha} = J'y$. De onde resulta:

$$y'P_Jy = \hat{\alpha}'J'y.$$

A soma de quadrados residual, sob H_0 , é: $y'(I - P_J)y = SQR_R$.

A SQR_R , restrita, tem $n-1$ graus de liberdade e é representada na Figura 27 pela distância quadrática do vetor y ao vetor $\hat{y}_1$.

iii) Cálculo da estatística F ou razão de variâncias:

Escrita em termos de projeções, a estatística F é:

$$F_{1, n-2(\alpha)} = \frac{y'(P_{[Jx]} - P_J)y/1}{y'(I - P_{[Jx]})y/n-2}$$

onde $(P_{[Jx]} - P_J)y = \hat{y} - \hat{y}_1$ e $(I - P_{[Jx]})y = y - \hat{y}$.

Figura 27. Teste da hipótese de exclusão de variável.

4.5.4 Exemplo

Um experimento foi conduzido durante 7 semanas para investigar a relação entre a idade (X) e a altura de feijão (Y). Os dados referentes às amostras aleatórias e independentes, encontram-se na Tabela 7.

Tabela 7. Alturas, em cm, de pés de feijão.

Idade em semanas (X)	1	2	3	4	5	6	7
Alturas em cm (Y)	5	13	16	23	33	38	40

Fonte: Snedecor & Cochran (1967), p.139.

Uma relação linear entre X e Y é assumida, de modo que se pode escrever o modelo na forma: $y = \alpha + \beta(x - \bar{x}) + e$, onde α e β são parâmetros desconhecidos, $\bar{x}$ é a média da variável X e e o vetor de erros aleatórios, tal que $e \cap N(\phi, I\sigma^2)$. O objetivo é estimar α e β e testar $H_0: \beta = 0$ contra $H_a: \beta \neq 0$.

O vetor de observações e sua decomposição são dados por:

$$\begin{bmatrix} 5 \\ 13 \\ 16 \\ 23 \\ 33 \\ 38 \\ 40 \end{bmatrix} = \alpha \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} + \beta \begin{bmatrix} -3 \\ -2 \\ -1 \\ 0 \\ 1 \\ 2 \\ 3 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \\ e_6 \\ e_7 \end{bmatrix}$$

A matriz dos vetores J e $(x - \bar{x})$, gera um espaço bidimensional para o modelo, uma vez que ela tem posto coluna completo. Desse modo as estimativas de mínimos quadrados são:

$$\begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \end{bmatrix} = \begin{bmatrix} 0,1429 & 0 \\ 0 & 0,0357 \end{bmatrix} \begin{bmatrix} 168 \\ 172 \end{bmatrix} = \begin{bmatrix} 24,000 \\ 6,1429 \end{bmatrix}$$

Sendo:

$$P_J = J(J'J)^{-1}J' = \frac{1}{7}JJ' = \frac{1}{7} \begin{bmatrix} J & J & J & J & J & J & J \end{bmatrix}$$

$$P_{Jx} = \begin{bmatrix} 0,4643 & 0,3571 & 0,2500 & 0,1429 & 0,0357 & -0,0714 & -0,1786 \\ 0,3571 & 0,2857 & 0,2143 & 0,1429 & 0,0071 & 0,0000 & -0,0714 \\ 0,2500 & 0,2143 & 0,1786 & 0,1429 & 0,1071 & 0,0714 & 0,0357 \\ 0,1429 & 0,1429 & 0,1429 & 0,1429 & 0,1429 & 0,1429 & 0,1429 \\ 0,0357 & 0,0714 & 0,1071 & 0,1429 & 0,1786 & 0,2143 & 0,2500 \\ -0,0714 & 0,0000 & 0,0714 & 0,1429 & 0,2143 & 0,2857 & 0,3571 \\ -0,1786 & -0,0714 & 0,0357 & 0,1429 & 0,2500 & 0,3571 & 0,4643 \end{bmatrix}$$

$$\hat{y} = P_{Jx}y = \begin{bmatrix} 5,57 \\ 11,71 \\ 17,85 \\ 24,00 \\ 30,14 \\ 36,28 \\ 42,43 \end{bmatrix} \quad \hat{y}_1 = P_J y = \begin{bmatrix} 24 \\ 24 \\ 24 \\ 24 \\ 24 \\ 24 \\ 24 \end{bmatrix} = \bar{y}J$$

Dessa forma o modelo ajustado, mostrado na Figura 28, pode ser expresso por:

$$y = \bar{y} + 6.1429(x - \bar{x}) + \hat{e}$$

Figura 28. Decomposição ortogonal do vetor de observações para o caso de regressão linear simples.

Tradicionalmente essa decomposição ortogonal é escrita com o vetor de observações centrado na média, $\bar{y}$:

$$y - \bar{y} = \hat{\beta}(x - \bar{x}) + \hat{e}$$

$$\begin{bmatrix} -19 \\ -11 \\ -8 \\ -1 \\ 9 \\ 14 \\ 16 \end{bmatrix} = 6,1429 \begin{bmatrix} -3 \\ -2 \\ -1 \\ 0 \\ 1 \\ 2 \\ 3 \end{bmatrix} + \begin{bmatrix} -0,57 \\ 1,29 \\ -1,85 \\ -1,00 \\ 2,86 \\ 1,72 \\ -2,43 \end{bmatrix}$$

Essa partição facilita a representação geométrica, como mostra a Figura 29.

Figura 29. Forma simplificada da decomposição ortogonal na regressão linear simples.

Aplicando o Teorema de Pitágoras ao triângulo retângulo formado na Figura 29 e observando às respectivas dimensões dos espaços associados aos vetores tem-se as bases para se escrever o tradicional quadro da análise da variância.

Tabela 8. Quadro da análise da variância para a regressão linear simples.

<i>Causas de Variação</i>	<i>G.L.</i>	<i>S.Q</i>	<i>Q.M.</i>	<i>F</i>
Regressão	1	$\hat{\beta}^2 \ x - \bar{x}\ ^2 = 1056,57$	1056,57	225,4885
Erro	5	$\ y - \hat{y}\ ^2 = 23,4286$	4,6857	
Total	6	$\ y - \bar{y}\ ^2 = 1080,00$		

Com base no valor de F, pode-se, então, testar a hipótese H_0 . Observa-se também, que tal estatística pode ser encontrada com base em P_{Jx} e P_J e as respectivas somas de quadrados SQR_{NR} e SQR_R , dadas por:

$SQR_{NR} = y'(I - P_{Jx})y = 23,4286$, a qual tem $n - 2 = 5$ g. l. e geometricamente representa a distância quadrática do vetor y ao vetor $\hat{y}$ (Figura 28).

$SQR_R = y'(I - P_J)y = 1.080,00$, com $n - 1 = 6$ g. l. e geometricamente representa a distância quadrática do vetor y ao vetor $\hat{y}_1$ (Figura 28).

A estatística F para o teste:

$$F_{1,n-2(\alpha)} = \frac{y'(P_{Jx} - P_J)y/1}{y'(I - P_{Jx})y/n-2} = \frac{1056,5714/1}{23,4286/5} = 225,4885$$

onde o numerador representa geometricamente a distância quadrática do vetor $\hat{y}_1$ ao vetor $\hat{y}$.

Como $F_{calc} = 225,48 > F_{1,5} = 6,61$, rejeita-se $H_0: \beta = 0$, ao nível 0,05 de probabilidade.

O coeficiente de determinação para o modelo pode ser expresso por:

$$R^2 = \cos^2 \theta = \frac{\|\hat{y}\|^2}{\|y - \bar{y}\|^2} = \frac{y'(P_{[Jx]} - P_J)y}{y'(I - P_J)y} = \frac{1056,5714}{1080,00} = 0,9783$$

isto é, 97,83% da soma de quadrados total é “explicada” pela regressão linear.

Sendo $R^2 = \cos^2 \theta$, então θ , o ângulo agudo formado entre os vetores $y - \bar{y}$ (hipotenusa) e $\hat{y} = \hat{\beta}(x - \bar{x})$ (cateto adjacente) do triângulo da Figura 29, é:

$$\theta = \arccos(\sqrt{0,9783}) \cong 8^\circ 30'$$

e, obviamente é um ângulo muito pequeno, expressando a boa qualidade do ajuste obtido.

4.5.5 Intervalo de confiança para o coeficiente de regressão linear

Desde que $\|x - \bar{x}\| \cdot \hat{\beta}$ apresenta distribuição $N(\|x - \bar{x}\| \cdot \beta, \sigma^2)$, então $\|x - \bar{x}\|(\hat{\beta} - \beta)$ apresenta distribuição $N(0, \sigma^2)$ e, em decorrência, $\|x - \bar{x}\|(\hat{\beta} - \beta) / \sqrt{s^2}$ tem distribuição t com 2 graus de liberdade. Sendo assim, uma estimativa por intervalo para β ao nível de confiança 0,95, pode ser obtida da seguinte forma:

$$t_{(0,025)n-2} \leq \frac{\|x - \bar{x}\|(\hat{\beta} - \beta)}{s} \leq t_{(0,975)n-2}$$

e, isolando β tem-se que:

$$\hat{\beta} - t \frac{s}{\|x - \bar{x}\|} \leq \beta \leq \hat{\beta} + t \frac{s}{\|x - \bar{x}\|}$$

onde $s_{\hat{\beta}} = \frac{s}{\|x - \bar{x}\|}$ é o erro padrão da declividade.

Para o exemplo 4.5.4, foram obtidos os seguintes resultados:

$$\hat{\beta} = 6,1429, \quad \|x - \bar{x}\|^2 = 28 \quad \text{e} \quad s^2 = \frac{\|\hat{e}\|^2}{n-2} = \frac{23,4286}{5} = 4,6857$$

conseqüentemente a estimativa por intervalo para β , ao nível de confiança 0,95, sendo $t_{(0,05; 5)} = 2,571$ é:

$$5,0911 \leq \beta \leq 7,1946.$$

4.5.6 O modelo de regressão através da D.V.S.

Considere o modelo de regressão $y = X\beta + e$, com as condições de Gauss-Markov em e e seja $X = USV' = U_1 S_1 V_1'$ a decomposição por valores singulares da matriz X , conforme visto na seção 3.5. Então o problema de encontrar β que minimize a função $Q(\beta) = \|y - X\beta\|^2$ pode também ser posto na forma:

$$\|y - USV'\beta\|^2 = \|U'y - SV'\beta\|^2 = \left\| \begin{pmatrix} U_1'y \\ U_2'y \end{pmatrix} - \begin{pmatrix} S_1 V_1'\beta \\ S_2 V_2'\beta \end{pmatrix} \right\|^2 = \|U_1'y - S_1 V_1'\beta\|^2 + \|U_2'y\|^2$$

o que é perfeitamente possível pois as transformações ortogonais preservam a distância euclidiana e portanto a norma.

Sendo assim, o sistema de equações normais, em função da D.V.S. da matriz X será: $S_1 V_1' \hat{\beta} = U_1' y$, com solução dada por: $\hat{\beta} = V_1 S_1^{-1} U_1' y = X^+ y$. É possível, também, determinar os projetores e as projeções em função da D.V.S.:

- a) $\hat{y} = X\hat{\beta} = U_1 S_1 V_1' V_1 S_1^{-1} U_1' y = U_1 U_1' y$
- b) $\hat{e} = y - \hat{y} = (I - U_1 U_1' y) = U_2 U_2' y$
- c) Projetor ortogonal de y em $C(X)$, $P = U_1 U_1'$
- d) Projetor ortogonal de y em $C^\perp(X) = (I - U_1 U_1') = U_2 U_2'$

Como se vê, a troca de X pela sua D.V.S. em nada afeta as idéias já formalizadas neste trabalho. Pelo contrário, fornece um critério alternativo para se obter tais resultados. Maiores detalhes sobre a D.V.S e modelos lineares podem ser encontrados em Carvalho Filho (1987), Lemma (1989), Costa (1995), entre outros. Esse procedimento fornece as bases para o estudo da *regressão por componentes principais* e sua utilização é, sem dúvida, particularmente importante no caso da existência da colinearidade ou da quase colinearidade entre as colunas de X .

Seja, nesse contexto, o modelo de regressão $y = X\beta + e$ e sua respectiva regressão por componentes principais. De acordo com Mandel (1982):

$$y = USV' \beta + e \quad (4.5.6.1)$$

A equação (4.5.6.1) pode também ser escrita como: $y = U(SV' \beta) + e$, onde $SV' \beta$ é um vetor de dimensão $r \times 1$. Denotando este vetor por δ , o modelo se reduz a:

$$y = U\delta + e \quad (4.5.6.2)$$

No modelo definido pela equação 4.5.6.2, o vetor y e a matriz U são conhecidos. A solução de mínimos quadrados para os r elementos desconhecidos de δ é obtida pelo sistema de equações normais:

$$\hat{\delta} = (U'U)^{-1} U' y = U' y \quad (4.5.6.3)$$

A equação (4.5.6.3) é de fácil resolução, uma vez que $\hat{\delta}$ é simplesmente o produto interno do vetor y pelo j -ésimo vetor u_j . Das equações 4.5.6.2 e 4.5.6.3 segue que:

$$E[\hat{\delta}] = E[U' y] = U' U \delta = \delta \quad (4.5.6.4)$$

de modo que $\hat{\delta}$ é não viesado. Além disso, a matriz de variâncias e covariâncias de $\hat{\delta}$ é dada por:

$$V[U' y] = U' V[y] U = I \sigma^2$$

mostrando que os $\hat{\delta}_j$ são não correlacionados.

Observe que o número de parâmetros em β é p , enquanto que o número de elementos em δ é r , que pode ser menor que p . Mas $\delta = S V' \beta$ e essa relação permanece válida para os estimadores de mínimos quadrados de δ e β , ou seja:

$$\hat{\delta} = S V' \hat{\beta} \quad (4.5.6.5)$$

Na relação (4.5.6.5), $S V'$ é uma matriz de dimensão $r \times p$. Então, a relação matricial representa r equações com p estimativas do parâmetro β , $\hat{\beta}$. Como é bem sabido, se $r = p$ (posto coluna completo), a solução é possível e única. Neste caso V' é uma matriz ortogonal de ordem p . Então a solução para o vetor β será dada por:

$$\hat{\beta} = V'^{-1} S^{-1} \hat{\delta} = V S^{-1} \hat{\delta} \quad (4.5.6.6)$$

Pode-se notar que $S^{-1} \hat{\delta}$ é um vetor $p \times 1$, obtido pela divisão de cada $\hat{\delta}_j$ pelo seu correspondente s_j .

A equação (4.5.6.6) além de fornecer as estimativas dos componentes de β como funções de $\hat{\delta}$, também pode ser usada para calcular as variâncias dos $\hat{\beta}_j$. Em

notação geral, ela pode ser escrita como: $\hat{\beta}_j = \sum_{k=1}^p v_{jk} \frac{\hat{\delta}_k}{s_k}$. Desde que os $\hat{\delta}_j$ são

mutuamente ortogonais e tem a mesma variância σ^2 , então: $V(\hat{\beta}_j) = \left[\sum_{k=1}^p \frac{v_{jk}^2}{s_k^2} \right] \sigma^2$. O

numerador de cada termo é o quadrado dos elementos da j -ésima linha da matriz V e os denominadores são os quadrados dos valores singulares.

Uma consequência desse fato é quando um s_j é considerado pequeno em relação aos demais, ele contribui de forma importante para o aumento do valor da variância das estimativas dos parâmetros. Como se nota, a técnica da DVS permite localizar a causa ou as causas de grandes variâncias encontradas para alguns estimadores (Mandel, 1982, p.18).

4.5.7 Multicolinearidade

Como se sabe, os valores singulares, s_j , da matriz X são as raízes quadradas dos autovalores de $X'X$. Então se um s_j é igual a zero, reciprocamente tem-se que um autovalor é nulo. Isso significa que existe uma relação linear perfeita entre pelo menos dois dentre os p vetores da matriz X . Um valor s_j próximo de zero implica uma forte relação linear entre esses vetores (Mandel, 1982, p.19). Caso contrário, isto é, se não existe relação linear entre as variáveis regressoras, então elas são ditas ortogonais e as inferências sobre o modelo de regressão podem ser feitas com relativa facilidade. Quando todas as colunas da matriz X são ortogonais entre si, a matriz $X'X$ é diagonal e as estimativas dos parâmetros da regressão múltipla coincidem com as estimativas dos coeficientes das regressões lineares simples de y_j contra cada uma das variáveis independentes. Portanto, a ortogonalidade entre os vetores de X facilita bastante a análise.

Infelizmente, na maioria das aplicações, as variáveis regressoras não são ortogonais. Algumas vezes a falta da ortogonalidade não é séria. No entanto, em algumas situações as variáveis são quase perfeitamente lineares, e nestes casos as inferências baseadas sobre o modelo de regressão podem ser enganosas. Quando existe uma quase dependência linear entre as variáveis regressoras, o problema da multicolinearidade existe.

Geometricamente, duas variáveis são perfeitamente colineares se os vetores regressores estão numa mesma reta, ou, de modo mais geral, estão no mesmo subespaço de dimensão um.

Nesse contexto, suponha que no exemplo que ilustra a seção 4.5.1, o vetor regressor x_1 em $y = \beta_1 x_1 + \beta_2 x_2 + e$, seja exatamente uma função linear de x_2 , dada por $x_1 = 2x_2$. Dado que x_1 e x_2 formam algum ângulo com y , o subespaço de estimação $L = C(X)$, mostrado na Figura 30, é uma linha ao invés de um plano como na Figura 25. Conseqüentemente nenhuma projeção de y sobre L , produzirá estimadores únicos para β_1 e β_2 .

Figura 30. Um exemplo de multicolinearidade perfeita.

Note que este é um caso específico com duas variáveis. Para o caso geral envolvendo p variáveis regressoras o raciocínio é análogo, e o termo colinear é substituído por coplanar, como uma extensão do conceito de colinearidade. Vale a pena ressaltar que embora o termo coplanar seja o mais indicado para o caso geral, ele não tem sido usualmente utilizado na literatura. Assim, p variáveis serão colineares se os vetores que as representam pertencem a um subespaço de dimensão menor que p .

Na prática, a colinearidade perfeita raramente ocorre. Para efeitos de estimação, duas variáveis regressoras serão ditas colineares se elas estiverem “quase” no mesmo subespaço, ou seja, se o ângulo entre elas for “pequeno”. Relacionando com as idéias da seção 4.1.2, equivale a dizer que um alto coeficiente de correlação entre elas é um forte indício de colinearidade. Para se ter uma idéia do problema da quase colinearidade, considere os dados da Tabela 9.

Tabela 9. Valores fictícios de três variáveis.

X_1	X_2	Y
6	11	37
15	18	44
9	24	52,5

Um modelo de regressão linear é assumido para os dados apresentados, tal que:

$$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + e$$

ou, sob a forma matricial:

$$\begin{bmatrix} 37 \\ 44 \\ 52,5 \end{bmatrix} = \begin{bmatrix} 1 & 6 & 11 \\ 1 & 15 & 18 \\ 1 & 9 & 24 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ e_3 \end{bmatrix}$$

Num primeiro momento, pode-se pensar em proceder a análise da forma tradicional através do já discutido método dos mínimos quadrados, visto que a matriz que representa as variáveis regressoras tem posto coluna completo, $r[X] = 3$. Embora esses vetores estejam no espaço tridimensional, o problema da colinearidade existe. Essa quase dependência linear se deve a proximidade entre os vetores J , x_1 e x_2 , como pode ser observado através dos ângulos formados entre eles:

$$\cos \theta_1 = \frac{\langle J, x_1 \rangle}{\|J\| \|x_1\|} = \frac{30}{32,03} = 0,9365 \Rightarrow \theta_1 = \arccos(0,9365) \cong 20^\circ$$

$$\cos \theta_2 = \frac{\langle J, x_2 \rangle}{\|J\| \|x_2\|} = \frac{53}{55,34} = 0,9576 \Rightarrow \theta_2 = \arccos(0,9576) \cong 17^\circ$$

$$\cos \theta_3 = \frac{\langle x_1, x_2 \rangle}{\|x_1\| \|x_2\|} = \frac{552}{590,75} = 0,9343 \Rightarrow \theta_3 = \arccos(0,9343) = 20^\circ$$

Em decorrência disso, espaço tridimensional gerado pelos vetores fica “mal definido” e as inferências baseadas no modelo de regressão podem ser enganosas. Esse fato é comumente referido como *mau condicionamento* da matriz X e é de certa forma, refletido no tamanho dos valores singulares. No entanto, o grau de mau condicionamento dependerá de quão pequeno é o menor valor singular em relação ao maior. Apresenta-se a seguir, os valores singulares da matriz X, descritos na diagonal da matriz S:

$$S = \begin{bmatrix} 36,5012 & \phi & \\ & 5,7842 & \\ & \phi & 0,4547 \end{bmatrix},$$

onde s_j^2 é raiz característica não nula de $X'X$. Alguns autores, como por exemplo Montgomery & Peck (1981), Iemma (1989) e Ferrari (1989), sugerem examinar o número de condição de $X'X$ definido por:

$$k = \frac{\lambda_{\max}}{\lambda_{\min}}$$

sendo que, em geral, se o número de condição é menor que 100, não há sérios problemas com multicolinearidade. Número de condição entre 100 e 1000 implica moderada a forte multicolinearidade e, se k excede 1000 é indicação de multicolinearidade severa. Para o exemplo em discussão:

$$k = \frac{(36,512)^2}{(0,4547)^2} = 6.444,13$$

não deixando dúvidas a cerca do grau de multicolinearidade.

Em casos como este, a literatura sugere uma atenção especial, através de métodos alternativos de estimação dos parâmetros de regressão linear múltipla, uma vez que o estimador usual de mínimos quadrados apesar de possuir a propriedade tão desejável de estimador não viesado, na presença de multicolinearidade, apresenta deficiência na obtenção de estimativas confiáveis para os parâmetros do modelo, pois nestes casos podem apresentar variâncias extremamente grandes, resultando numa forma de mau condicionamento da matriz $X'X$. Dentre os efeitos que a multicolinearidade pode acarretar sobre as estimativas de mínimos quadrados dos coeficientes de regressão $\hat{\beta}$, Hoerl & Kennard (1970) citados por Ferrari (1989) observam que quando se considera o quadrado da distância de $\hat{\beta}$ ao parâmetro β , isto é:

$$d^2 = \|\hat{\beta} - \beta\|^2 = (\hat{\beta} - \beta)'(\hat{\beta} - \beta)$$

tem-se que o valor esperado de d^2 :

$$E(d^2) = E[(\hat{\beta} - \beta)'(\hat{\beta} - \beta)] = \sum_{j=1}^p E(\hat{\beta}_j - \beta_j)^2 = \sum_{j=1}^p V(\hat{\beta}_j) = \sigma^2 \text{Tr}(X'X)^{-1}$$

implicando que:

$$E(d^2) = \sigma^2 \sum_{j=1}^p \lambda_j^{-1},$$

onde $\lambda_j > 0, j = 1, 2, \dots, p$, são os autovalores da matriz $X'X$, considerada na forma de correlações. Então se $X'X$ é mal condicionada em razão da multicolinearidade, pelo menos um dos λ_j será muito pequeno, implicando que a distância da estimativa de mínimos quadrados $\hat{\beta}$ ao parâmetro β poderá ser muito grande. Ainda, segundo esses autores tem-se, equivalentemente:

$$E(d^2) = E(\hat{\beta}'\beta - 2\hat{\beta}'\beta + \beta'\beta)$$

ou

$$E(\hat{\beta}'\hat{\beta}) = \beta'\beta + \sigma^2 \text{Tr}(X'X)^{-1}$$

ou seja, o vetor $\hat{\beta}$ geralmente possui maior norma do que o vetor β , mostrando que o método de mínimos quadrados produz coeficientes de regressão estimados que podem apresentar grandes valores absolutos.

Por estes aspectos, diz-se que a multicolinearidade não é um problema estatístico do modelo linear $y = X\beta + e$, mas sim um problema induzido pelas colunas da matriz X que descrevem as variáveis independentes e, mais precisamente, um problema geométrico.

4.5.8 Seleção de variáveis

Um dos objetivos do estudo da regressão é comparar a importância de diferentes variáveis explanatórias. Entre as muitas medidas sugeridas pela literatura, pode-se citar, por exemplo, os coeficientes de regressão padronizados e o procedimento *stepwise*. Entretanto, na falta de ortogonalidade entre as variáveis regressoras, tais procedimentos podem conduzir a conclusões equívocas. Para se ter uma idéia desse problema, considere um modelo de regressão linear com apenas duas variáveis, como na seção 4.5.1:

$$y = \beta_1 x_1 + \beta_2 x_2 + e,$$

suponha agora, que nesse modelo as variáveis independentes e a variável dependente sejam padronizadas segundo o critério proposto por Bring (1996), ou seja:

$$x_i = (x_i^* - \bar{x}) / s_{x_i^*} \sqrt{(n-1)} \quad \text{e} \quad y_i = (y_i^* - \bar{y}) / s_{y_i^*} \sqrt{(n-1)}, \quad (4.5.8.1)$$

onde x^* e y^* são as variáveis originais. A vantagem desta padronização é que os vetores tornam-se normalizados e, em consequência $R^2 = \|\hat{y}\|^2 = \cos^2(\theta)$, onde θ é o ângulo agudo formado entre os vetores y e $\hat{y}$.

Nas ciências médica e social, é muito comum usar o método de coeficientes de regressão padronizados, $\hat{\beta}$, obtidos a partir de um modelo com variáveis padronizadas com média zero e desvio padrão 1, para selecionar variáveis independentes. A seleção baseia-se na magnitude de $\hat{\beta}_j$, pois quanto maior for o seu valor mais x_j contribui para o valor predito de y . Estes coeficientes, geometricamente, como já visto também em 4.5.1, podem ser ilustrados como os coeficientes da combinação linear do vetor $\hat{y}$ com relação a x_1 e x_2 :

$$\hat{y} = \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2,$$

conforme ilustra a Figura 31. Note que $\hat{\beta}_j x_j$ para $j = 1, 2$, corresponde a projeção ortogonal do vetor $\hat{y}$ sobre o vetor x_j e cada coeficiente padronizado $\hat{\beta}_j$ é a distância considerando o sinal do deslocamento paralelo com x_j .

Figura 31. Regressão com duas variáveis não correlacionadas: uma visão da partição do coeficiente de determinação: $R^2 = r_1^2 + r_2^2$.

Quando as variáveis explanatórias são não correlacionadas, como na Figura 31, os $\hat{\beta}_j$'s podem ser utilizados como indicadores da contribuição da predição

de y . Isso se justifica pela ortogonalidade entre os vetores regressores pois, neste caso, aplicando o Teorema de Pitágoras obtém-se a seguinte partição:

$$\|\hat{y}\|^2 = \hat{\beta}_1^2 \|x_1\|^2 + \hat{\beta}_2^2 \|x_2\|^2$$

e, sendo as variáveis padronizadas como em 4.5.8.1 segue que:

$$\|\hat{y}\|^2 = \hat{\beta}_1^2 + \hat{\beta}_2^2, \quad (4.5.8.2)$$

onde os coeficientes $\hat{\beta}_j$'s coincidem com os coeficientes de correlações simples (r_j 's) entre y e x_j . Nessas condições, é possível usar tal procedimento para “medir” a importância relativa de cada uma das variáveis, observando quanto cada uma contribui para o comprimento quadrado do vetor $\hat{y}$. Ademais, isso é bastante plausível pois muitas medidas da importância de variáveis baseiam-se na partição do coeficiente de determinação do modelo, que neste caso pode ser particularmente expresso por:

$$R^2 = r_1^2 + r_2^2$$

No entanto, quando as variáveis são correlacionadas a partição sugerida na padronização de Bring (1996) em 4.5.8.1 não é assegurada, como pode ser observado através da Figura 32.

Figura 32. Coeficientes padronizados de regressão com duas variáveis não ortogonais.

Uma forma alternativa de selecionar variáveis apropriadas para o modelo consiste na regressão *stepwise*. Em tal procedimento, adiciona-se em cada etapa mais uma variável independente, recalculando-se a cada passo o termo constante, os

coeficientes de regressão parcial e o erro padrão de estimação. Quando este procedimento é usado, a maioria dos programas computacionais informa o incremento no R^2 para cada variável incluída sucessivamente. “*Infelizmente, estes incrementos são às vezes usados como indicadores da importância relativa das variáveis independentes*” (Bring, 1996, p.60).

Através da geometria é possível mostrar que o procedimento *stepwise* deve ser sempre analisado cuidadosamente, pois quando as variáveis são correlacionadas o método pode apresentar falhas. Suponha um modelo com apenas duas variáveis, como o que exemplifica esta seção, sobre o qual se deva escolher entre x_1 e x_2 . De um modo geral, a primeira variável independente incluída é aquela com maior grau de correlação com y . Geometricamente isto significa que o procedimento compara os comprimentos dos vetores a e c (Figura 33). Se a é maior que c , x_1 é a primeira a ser selecionada, porém se c é maior que a então x_2 será a primeira a ser selecionada.

Figura 33. Uma idéia geométrica do procedimento *stepwise* na seleção de variáveis.

A Figura 33 mostra que se x_1 é a primeira a ser selecionada, R^2 cresce de 0 para $\|a\|^2$. Quando x_2 é incluída como uma segunda variável R^2 cresce com $\|b\|^2$. Se x_2 fosse selecionada primeiro, sua contribuição a R^2 seria $\|c\|^2$ e o acréscimo a partir de x_1 seria $\|d\|^2$ ao invés de $\|a\|^2$. Somente se x_1 e x_2 forem

ortogonais é que $\|a\|^2 = \|d\|^2$. Assim quanto mais relacionadas as variáveis se apresentarem, mais a avaliação da importância depende sobre a ordem de seleção. *“Não existe, com raras exceções, nenhuma justificativa para selecionar variáveis na ordem como são selecionadas pelo procedimento *stepwise*”* (Bring, 1996, p.60).

Um método alternativo para contornar esse problema consiste em considerar todas as possíveis ordenações ao invés de usar somente a ordem selecionada pelo procedimento *stepwise*. Assim se modelo existem k variáveis regressoras, existem $k!$ possíveis ordenações. Considerando todas as possíveis ordenações, uma contribuição de uma variável ao R^2 pode ser calculada como o incremento médio em R^2 . Esse método recebe o nome de *stepwise médio* e, sendo assim, para o exemplo ilustrado na Figura 33 a importância de x_1 seria igual a média dos comprimentos $\|a\|^2$ e $\|d\|^2$ e a importância de x_2 seria igual à média entre $\|b\|^2$ e $\|c\|^2$. Uma outra vantagem desse método é que o problema da multicolinearidade é reduzido devido ao fato de que todas as possíveis ordenações das variáveis são consideradas.

5 CONCLUSÕES

Este trabalho ilustra algumas das muitas possíveis aplicações da abordagem geométrica vetorial na interpretação de conceitos estatísticos. Nos exemplos apresentados verifica-se que o método geométrico permanece inalterado, ou seja, é sempre o mesmo para qualquer que seja o assunto em pauta. Por exemplo, tanto nas estatísticas básicas como na análise da variância e da regressão, a projeção ortogonal fornece o melhor ajuste para o modelo e o ângulo pode ser usado para medir a qualidade desse ajuste. Essa possibilidade da unidade de apresentação, em todos os níveis de aprendizagem, assegura uma vantagem significativa da abordagem geométrica. Nesse contexto, a geometria pode ser vista, como um meio natural para ressaltar a união de idéias fundamentais comuns entre as diferentes áreas.

Outro aspecto positivo é que a geometria constitui-se numa ferramenta poderosa e de fácil manuseio, para a apresentação e demonstração de conceitos de Estatística e Probabilidade. Há situações onde a interpretação geométrica de muitos conceitos, como mínimos quadrados, graus de liberdade e testes de hipóteses, entre outros, podem ser mais facilmente explorados se suas apresentações são feitas no campo da geometria. Existem ainda alguns problemas com os quais comumente os estatísticos se deparam, como por exemplo a colinearidade e a seleção de variáveis, onde o papel da geometria torna-se indispensável, ultrapassando a barreira paradigmática de que o procedimento geométrico é meramente um recurso didático.

REFERÊNCIAS BIBLIOGRÁFICAS

BRING, J. A Geometric Approach to compare variables in a Regression Model. **The Amer. Statist.**, vol. 50 (1): p. 57-62, Feb. 1996.

BRYANT, P. Geometry, statistics, probability: variations on a common theme. **The Amer. Statist.**, vol. 38 (1): p. 38-48, Feb. 1984.

CARVALHO FILHO, A. A. A decomposição da matriz de planejamento por valores singulares em modelos de posto incompleto. Piracicaba, 1987. 116p. Tese (Doutorado) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo.

CORRÊA, A. M. C. J. Funções Geradoras de Momentos. Piracicaba, 1981. 190p. Dissertação (Mestrado) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo.

COSTA, P. A. B. Estudo da utilização de estimadores com problemas de estimabilidade e aplicações. Piracicaba, 1995. 111p. Dissertação (Mestrado) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo.

DURBIN, J.; KENDALL, M. G. The Geometry of Estimation. **Biometrika**, vol. 38: p. 150-158, 1951.

- DWIVEDI, T. D. A Method to Compute the Rank Factors of a Matrix. **Sankhyā**, vol. 37(B4): p. 463-64, 1975.
- ECKART, C.; YOUNG, G. The approximation of a matriz by another lower rank. **Psychometrika**, vol. 1: p. 211-18, 1936.
- EVES, H. **Tópicos de História da Matemática para uso em sala de aula - Geometria**. Tradução de Hygino H. Domingues. São Paulo, SP. Atual Editora. 1992. 77 p.
- FERRARI, F. Estimadores viesados para modelos de regressão em presença de multicolinearidade. Piracicaba, 1989. 127p. Tese (Doutorado) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo.
- FISHER, R. A. Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. **Biometrika**, vol. 10: p. 507-521, 1915.
- HERR, D. G. On the History of the Use of Geometry in the General Linear Model. **The Amer. Statist.**, vol. 34 (1): p. 43-47, Feb. 1980.
- HOERL, A. E.; KENARD, R. W. Ridge regression: Biased estimation for nonorthogonal problems. **Technometrics**, vol. 12(1): p. 55-68, 1970.
- IEMMA, A. F. **Modelos Lineares: uma introdução para profissionais da pesquisa agropecuária**. Londrina: Imprensa oficial do Estado do Paraná, 1987. 236p.

IEMMA, A. F. **Matrizes para Estatística**. Piracicaba: Departamento de Matemática e Estatística - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo. 1989. 418p.

IEMMA, A. F.; PALM, R. Les matrices inverses generalisées et leur utilisation dans le modele linéaire. **Notes Stat. Inform.**, vol. 43(2): p. 1 - 26. 1992.

IEMMA, A. F.; PALM, R.; CLAUSTRIAUX, J.J. Sobre a construção de projetores ortogonais. **Revista Matemática Estatística II**, vol.11: p. 133-142. 1993.

JAMES, B. R. **Probabilidade: um curso em nível intermediário**. Rio de Janeiro: Instituto de Matemática Pura e Aplicada, 1981. 304p.

JOHNSON, R. A.; WICHERN D. W. **Applied Multivariate Statistical Analysis**. New Jersey: Prentice-Hall, 1992. 639p.

MANDEL, J. Use of the Singular Value Decomposition in Regression Analysis. **The Amer. Statist.**, vol. 36(1): p.15-24, Feb. 1982.

MARGOLIS, M. S. Perpendicular projections and elementary statistics. **The Amer. Statist.**, vol. 33(3): p. 131-135, Aug. 1979.

MONTGOMERY, D. C.; PECK, E. A. **Introduction to linear regression analysis**. New York: Wiley & Sons, 1981. 504p.

RAO, C. R. On the linear combination of observations and the general theory of least squares. **San khyā**, vol. 7(3): p. 237-56, 1945.

SAVILLE, D.J.; WOOD, G.R. A method for teaching statistics using n-dimensional geometry. **The Amer. Statist.**, vol. 40(3): p. 205-214, Aug. 1986.

SAVILLE, D.J.; WOOD, G.R. **Statistical Methods: The Geometric Approach**. New York: Springer-Verlag, 1991. 559p.

SCHEFFÉ, H. **The analysis of Variance**. New York: Wiley & Sons, 1959. 477p.

SCHEY, H.M. A geometric description of ortogonal contrasts in on way analises of variance. **The Amer. Statist.**, vol. 39(2): p.104-106, May 1985.

SEARLE, S. R. **Linear Models**. New York: Wiley, 1971. 532p.

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

SNEDECOR, G. W.; COCHRAN, W. G. **Statistical Method**. The Iowa State College Press, 1967. 534p.

WATSON, G. S. Linear least Squares Regression. **Annals of Mathematical Statistics**, vol 38: p. 1679-1699, 1967.

ZYSKIND, G. On Canonical Forms, Non-negative Covariance Matrices and Best and Simple Least Squares Linear Estimators in Linear Models. **Annals of Mathematical Statistics**, vol 38: p. 1092-1109, 1967.