

Estatística Multivariada

UNIVERSIDADE FEDERAL DE LAVRAS

REITOR: Antônio Nazareno Guimarães Mendes

VICE-REITOR: Elias Tadeu Fialho

Diretoria Executiva

Renato Paiva (Diretor)

Elias Tadeu Fialho

Conselho Editorial

Renato Paiva (Presidente)

Amauri Alves de Alvarenga

Carlos Alberto Silva

Elias Tadeu Fialho

Luiz Carlos de Oliveira Lima

Daniel Furtado Ferreira

Estatística Multivariada



Lavras - MG

© 2008 by Daniel Furtado Ferreira, 1ª edição: 2008

Nenhuma parte desta publicação pode ser reproduzida, por qualquer meio ou forma, sem a autorização escrita e prévia dos detentores do copyright.

Direitos de publicação reservados à Editora UFLA.

Impresso no Brasil - ISBN: 978-85-87692-52-8

Editora UFLA

Campus Histórico - Caixa Postal 3037

37200-000 - Lavras - MG.

Tel: (35) 3829-1115 - Fax: (35) 3829-1551

E-mail: vendas_editora@ufla.br - editora@ufla.br

Homepage: www.editora.ufla.br

Projeto Gráfico: Daniel Furtado Ferreira

Secretaria: Glenda Fernanda Morton

Revisão de Texto: Jane Cherem

Revisão de Referências Bibliográficas: Márcio Barbosa de Assis

Editoração Eletrônica: Daniel Furtado Ferreira, Christyane Aparecida Caetano,

Luciana Carvalho Costa

Marketing e Comercialização: Bruna de Carvalho Naves

Capa: Daniel Furtado Ferreira, Helder Tobias

Ficha Catalográfica Preparada pela Divisão de Processos Técnicos da
Biblioteca Central da UFLA

Ferreira, Daniel Furtado.

Estatística multivariada / Daniel Furtado Ferreira. – 1. ed. – Lavras :
Ed. UFLA, 2008.

662 p. : il.

Bibliografia.

ISBN 978-85-87692-52-8

1. Estatística. 2. Normal multivariada. 3. Distância de Mahalanobis.
4. Máxima verossimilhança. 5. Análise de variância. 6. Covariância.

I. Título.

CDD - 519.535

Prefácio

A multivariada é uma das áreas da estatística de maior importância. Na investigação científica os fenômenos são estudados e analisados buscando-se soluções para problemas relevantes para a sociedade. As respostas desses fenômenos são mensuradas em mais de uma variável e, em geral, as análises são realizadas separadamente para cada uma delas. As técnicas e métodos científicos da estatística multivariada procuram contemplar todas essas variáveis de uma forma integrada, onde as inter-relações são exploradas em sua máxima profundidade e as soluções para os problemas são mais consistentes e úteis. A possibilidade de aplicar técnicas e métodos multivariados aos problemas específicos de cada área do conhecimento é constantemente aventada e muitas vezes abandonada. As razões para isso são a falta de informação sobre os detalhes das técnicas e, principalmente, a dificuldade de interpretação dos resultados alcançados. Esta obra foi idealizada para suprir uma lacuna e, principalmente, para minorar a carência de publicações brasileiras na área de estatística multivariada. Os aspectos didáticos não foram esquecidos. Este Livro poderá ser recomendado para os cursos regulares de estatística e para os cursos de pós-graduação em estatística e em áreas correlatas.

Esse Livro difere dos seus pares pela forma com que as demonstrações dos teoremas e as derivações das estatísticas dos testes são apresentados. A experiência que tenho no ensino e na pesquisa, na área de multivariada, foi a grande motivação para a compilação da obra. Muitos de meus alunos colaboraram para o enriquecimento e melhoria da qualidade do manuscrito. A necessidade de um livro-texto para o ensino de estatística multivariada foi um fator decisivo no desenvolvimento desse projeto. Também merece destaque o compromisso que tenho com o ensino público, de qualidade e gratuito. Adentrar no mundo da multivariada é navegar pelo hiperespaço da sabedoria e desvendar mistérios que parecem estar além da nossa capacidade de compreensão. Todos os pesquisadores, alunos, professores e cientistas estão convidados a realizar essa viagem.

Ao CNPq quero externar meus agradecimentos pelo suporte financeiro e pela confiança. Os recursos aportados à minha pesquisa permitiram uma adequada infraestrutura para a construção desse Livro. Outras instituições de fomento também

merecem meu agradecimento, como a CAPES e a FAPEMIG. À Universidade Federal de Lavras e ao Departamento de Ciências Exatas meu mais profundo agradecimento por propiciar um ambiente de trabalho que me deu condições amplas, em todos os sentidos dessa palavra, para a confecção desse Livro. Aos colegas de Departamento de Ciências Exatas da UFLA, que sempre me incentivaram e apoiaram, meu muito obrigado.

Gostaria de aproveitar o ensejo e agradecer à consultoria científica. Aos consultores, registro meus sinceros agradecimentos pelas considerações, elogios e críticas. A qualidade final dessa obra, com toda certeza, é co-responsabilidade desses consultores anônimos. Espero que os ensinamentos desse Livro consigam atingir o propósito de dirimir dúvidas e não de causá-las.

A maior parte dos exemplos apresentados nesse Livro foram compilados utilizando dados oriundos do projeto BiosBrasil (CSM-BGBD-www.bisobrasil.ufla.br), que tem intenção de mapear a biodiversidade dos solos da Amazônia. Assim, gostaria de agradecer a professora Fátima Maria de Souza Moreira, do Departamento de Ciências dos Solos da Universidade Federal de Lavras, pela cessão dos dados.

Aos meus Alunos, co-responsáveis por essa obra, meu muito obrigado. Não há sentido em uma obra como essa sem que haja a figura dos alunos. Aos meus orientados registro aqui um especial agradecimento pelo convívio, incentivo e apoio.

Gostaria de finalizar, direcionando meus agradecimentos àquelas pessoas que me acompanham todos os dias e que se sacrificaram para que meus ideais fossem alcançados. Aos meus familiares, mãe (Nadime), pai (João) (em memória), esposa (Lidiane) e filha (Aline) meu eterno amor, respeito e admiração.

Daniel Furtado Ferreira
Lavras, Minas Gerais

Sumário

Lista de Tabelas	14
Lista de Figuras	15
1 Introdução	19
1.1 Variáveis Aleatórias Multidimensionais	20
1.2 Amostras Aleatórias	27
1.3 Estatísticas Descritivas	28
1.4 Distâncias	34
1.5 Variâncias Generalizadas e Aspectos Geométricos	40
1.6 Exercícios	47
2 Álgebra Vetorial e Matricial	49
2.1 Álgebra Vetorial	49
2.1.1 Produto Interno	51
2.1.2 Comprimento, Ângulo e Distância	52
2.1.3 Ortogonalidade e Bases Ortonormais	55
2.1.4 Ortonormalização de Gram-Schmidt e Aplicações	56
2.2 Álgebra Matricial	59
2.2.1 Operações Matriciais Elementares	59
2.2.2 Matrizes Especiais	64
2.2.3 Operações Matriciais	68
2.2.4 Matrizes Particionadas	77
2.2.5 Formas Quadráticas e Fator de Cholesky	80
2.2.6 Maximização de Razão de Formas Quadráticas	89
2.2.7 Métodos Numéricos Para Determinação de Autovalores e Autovetores	91

2.2.8	Teorema do Valor Singular	96
2.2.9	Potências Fracionárias de Matrizes	97
2.2.10	Derivadas Vetoriais e Matriciais	100
2.3	Exercícios	104
3	Distribuições Multivariadas	107
3.1	Introdução	107
3.2	Distribuições Multivariadas	107
3.3	Jacobiano da Transformação	110
3.4	Distribuição Normal Multivariada	112
3.4.1	Propriedades	116
3.4.2	Distribuição Normal Bivariada	125
3.4.3	Elipsóides de Concentração	127
3.5	Verificação da Normalidade Multivariada	132
3.5.1	Testes Univariados de Normalidade	133
3.5.2	Testes Para Normalidade Multivariada	154
3.6	Distribuições Elípticas e Esféricas Simétricas	162
3.6.1	Definições e Exemplos	163
3.6.2	Propriedades das Distribuições Elípticas e Esféricas	165
3.7	Simulação de Dados de Distribuições Multivariadas	165
3.8	Exercícios	168
4	Distribuições Amostrais Multivariadas	173
4.1	Introdução	173
4.2	Distribuições de Formas Quadráticas	173
4.3	Distribuição Wishart e Suas Propriedades	175
4.4	Distribuição T^2 de Hotelling	179
4.5	Exercícios	180
5	Inferências Sobre Vetores de Médias	183
5.1	Introdução	183
5.2	Estimadores de Máxima Verossimilhança da Média e Covariância	183
5.3	Testes de Hipóteses Sobre a Média de Uma Normal Multivariada	187
5.4	Região de Confiança do Vetor de Médias	201
5.5	Intervalos de Confiança Simultâneos	205
5.6	Inferências Sobre Proporções Multinomiais	211
5.7	Exercícios	224

6	Testes de Hipóteses Sobre Matrizes de Covariâncias	227
6.1	Introdução	227
6.2	Testes Sobre Matrizes de Covariâncias de Uma População	227
6.2.1	Teste Para Uma Covariância Específica	228
6.2.2	Testes de Independência e Esfericidade	234
6.2.3	Teste Para Simetria Composta	242
6.3	Testes Para Matrizes de Covariâncias de Várias Populações	248
6.3.1	Teste Para Homogeneidade de Matrizes de Covariâncias	248
6.3.2	Teste de Independência de Grupos de Variáveis	257
6.3.3	Generalização de Alguns Testes	264
6.4	Exercícios	264
7	Inferências Sobre Vetores de Médias de Duas Populações	267
7.1	Introdução	267
7.2	Comparações Emparelhadas	268
7.2.1	Testes de Hipóteses	269
7.2.2	Regiões de Confiança	275
7.2.3	Intervalos de Confiança Simultâneos	277
7.3	Comparações Independentes	279
7.3.1	Pressuposições Básicas	280
7.3.2	Inferências Quando as Covariâncias São Homogêneas	281
7.3.3	Problema de Behrens-Fisher Multivariado e Suas Soluções Aproximadas	293
7.4	Exercícios	315
8	Análise de Variância Multivariada	317
8.1	Introdução	317
8.2	Modelos Lineares e Análise de Variância Multivariados	318
8.3	Testes de Hipóteses Multivariados e Aproximações F	331
8.4	Comparações Múltiplas	338
8.5	Exercícios	340
9	Análise de Agrupamento	341
9.1	Introdução	341
9.2	Medidas de Similaridade e Dissimilaridade	343
9.2.1	Medidas de Dissimilaridades Para Dados Binários	348
9.2.2	Medidas de Similaridades	350

9.2.3	Medidas de Similaridades Para Dados Binários	354
9.2.4	Agrupando Variáveis	355
9.3	Agrupamentos Hierárquicos	357
9.3.1	Agrupamento Hierárquico do Vizinho Mais Próximo	361
9.3.2	Agrupamento Hierárquico do Vizinho Mais Distante	364
9.3.3	Agrupamento Hierárquico da Ligação Média	367
9.3.4	Agrupamento Hierárquico do Centróide	370
9.3.5	Agrupamento Hierárquico da Mediana	375
9.3.6	Agrupamento Hierárquico de Ward	376
9.3.7	Agrupamento Hierárquico Baseado no Método Flexível Beta	382
9.3.8	Propriedades dos Métodos de Agrupamento Hierárquicos	384
9.4	Agrupamentos Não-Hierárquicos	385
9.5	Determinação do Número de Grupos	389
9.6	Validação do Agrupamento	391
9.7	Exercícios	393
10	Componentes Principais	395
10.1	Introdução	395
10.2	Componentes Principais Populacionais	396
10.3	Componentes Principais da Matriz de Correlação Populacional	408
10.4	Matrizes de Covariâncias ou de Correlações Especiais	415
10.5	Componentes Principais Parciais	419
10.6	Componentes Principais Amostrais	425
10.7	Número de Componentes	438
10.7.1	Porcentagem Cumulativa da Explicação da Variância Total	438
10.7.2	Média dos Autovalores	440
10.7.3	Gráfico dos Autovalores	442
10.7.4	Teste de Hipótese de Igualdade dos Últimos Autovalores	443
10.8	Seleção de Variáveis	444
10.9	Procedimentos Gráficos	447
10.10	Inferências	451
10.10.1	Propriedades Assintóticas dos Componentes Principais	452
10.10.2	Inferências Sobre Autovalores	454
10.10.3	Inferências Sobre Autovetores	463
10.10.4	Inferências Sobre Componentes Principais da Matriz de Correlações	467

10.10.5	Detecção de <i>Outliers</i>	468
10.11	Exercícios	469
11	Análise Fatorial	473
11.1	Introdução	473
11.2	Modelo Fatorial Ortogonal	474
11.3	Modelo Fatorial Oblíquo	480
11.4	Estimação dos Parâmetros do Modelo Fatorial Ortogonal	481
11.4.1	Método dos Componentes Principais	482
11.4.2	Método Fatores Principais	493
11.4.3	Método Iterativo dos Fatores Principais	501
11.4.4	Método da Máxima Verossimilhança	505
11.4.5	Método dos Fatores Canônicos	516
11.5	Rotação Fatorial	519
11.6	Teste da Falta de Ajuste do Modelo Fatorial	523
11.7	Escores Fatoriais	525
11.7.1	Preditores de Mínimos Quadrados Ponderado	526
11.7.2	Preditor de Regressão	528
11.7.3	Comparação dos Preditores dos Escores Fatoriais	529
11.8	Exercícios	530
12	Análise de Correlação Canônica	531
12.1	Introdução	531
12.2	Variáveis Canônicas e Correlação Canônica Populacional	532
12.3	Correlação Canônica Populacional de Variáveis Padronizadas	541
12.4	Variáveis Canônicas e Correlação Canônica Amostrais	546
12.5	Qualidade da Análise de Variáveis Canônicas	556
12.6	Inferências	563
12.7	Exercícios	569
13	Análise Discriminante	571
13.1	Introdução	571
13.2	Regras de Classificação	572
13.3	Classificação em Uma de Duas Populações Normais	579
13.3.1	Classificação em Duas Populações Normais Com Covariâncias Heterogêneas	584

13.3.2	Avaliação das Regras de Classificação Para Duas Populações Normais	585
13.3.3	Função Discriminante de Fisher Para Duas Populações	602
13.4	Classificação em Mais de Duas Populações Normais	608
13.4.1	Função Discriminante de Fisher Considerando Duas ou Mais Populações	613
13.4.2	Avaliação das Regras de Classificação Para Mais de Duas Populações Normais	621
13.5	Análise Discriminante e de Agrupamento	622
13.6	Exercícios	623
	Referências Bibliográficas	625
	Apêndices: Tabelas Estatísticas	635
	Índice Remissivo	651