



SME0263

Análise Multivariada

2º semestre de 2011

Prof. Cibele Russo

cibele@icmc.usp.br

<http://www.icmc.usp.br/cibele>

Sala 3-162, ramal 6618

Coteia Wiki: <http://wiki.icmc.usp.br/index.php/SME-263>

Principais objetivos do curso

- Estudar e desenvolver métodos estatísticos para **descrever e analisar dados multivariados**

Programa - 1ª parte

- Organização de dados e análise descritiva
- Vetores aleatórios e álgebra de matrizes
- Média e matriz de covariâncias amostrais
- Distribuições multivariadas discretas e contínuas.
- Inferências sobre um vetor de médias
- Análise de variância multivariada
- Regressão linear multivariada

Dados multivariados - Exemplo

Tabela 1.1: Taxas de delitos por 100.000 habitantes por divisão territorial das polícias do Estado de São Paulo (Deinter), em 2002

Deinter	Homicídio doloso	Furto	Roubo	Roubo e furto de veículos
SJRP	10,85	1.500,80	149,35	108,38
RP	14,13	1.496,07	187,99	116,66
Bauru	8,62	1.448,79	130,97	69,98
Campinas	23,04	1.277,33	424,87	435,75
Sorocaba	16,04	1.204,02	214,36	207,06
SP	43,74	1.190,94	1.139,52	909,21
SJC	25,39	1.292,91	358,39	268,24
Santos	42,86	1.590,66	721,90	275,89
Média	23,08	1.375,19	415,92	298,90
DP	13,69	152,05	351,62	273,35

fonte: Secretaria de Segurança Pública do Estado de São Paulo

<http://www.ssp.sp.gov.br/estatisticas/criminais/>, acessada em 11/02/2003.

SJRP: São José do Rio Preto

RP: Ribeirão Preto

SP: São Paulo (capital)

SJC: São José dos Campos

Fonte: Barroso, L. P. e Artes, R. Análise Multivariada. 2005.

Disponível em <http://www.ime.usp.br/lbarroso/Livro.pdf>.

Programa - 2ª parte

- Análise de componentes principais
- Análise fatorial
- Análise de correlação canônica
- Análise de agrupamentos
- Análise discriminante

Programa - 2ª parte

- Análise de componentes principais (Pearson, 1901)
 - 1 Explicar a estrutura de variância e covariância dos dados através de combinações lineares
 - 2 Redução da dimensão dos dados de p variáveis para k componentes principais não correlacionados, ordenados decrescentemente pela explicação da variabilidade total dos dados
 - 3 Interpretações

Programa - 2ª parte

- Análise fatorial (Spearman, 1904)
 - 1 Explicar a variabilidade original dos dados através de fatores que meçam aspectos comuns dos dados
 - 2 Com várias motivações na psicologia e psicometria, surgiu inicialmente com a necessidade de medir quantidades latentes (por exemplo: ansiedade, inteligência)
 - 3 Identificação do número de fatores
 - 4 Interpretações
 - 5 Pode ser considerada uma extensão da análise de componentes principais

Programa - 2ª parte

- Análise de correlação canônica
 - 1 Identificar e quantificar associações entre dois conjuntos de dados
 - 2 Resumir as informações de cada conjunto de dados de modo a maximizar a correlação existente entre ambos

Programa - 2ª parte

- Análise de agrupamentos

- 1 Análise de conglomerados ou análise de cluster

- 2 Dividir os elementos da amostra de forma a ter grupos de elementos similares entre si em relação às variáveis observadas

Programa - 2ª parte

Análise de agrupamentos - Exemplo

Deseja-se agrupar 11 idiomas de acordo com a sua similaridade para a primeira letra dos numerais de 1 a 10.

Table 12.2 Numerals in 11 Languages

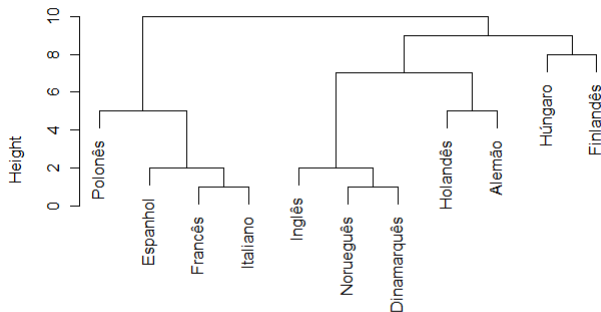
English (E)	Norwegian (N)	Danish (Da)	Dutch (Du)	German (G)	French (Fr)	Spanish (Sp)	Italian (I)	Polish (P)	Hungarian (H)	Finnish (Fi)
one	en	en	een	eins	un	uno	uno	jeden	egy	yksi
two	to	to	twee	zwei	deux	dos	due	dwa	ketto	kaksi
three	tre	tre	drie	drei	trois	tres	tre	trzy	harom	kolme
four	fire	fire	vier	vier	quatre	cuatro	quattro	cztery	negy	neljä
five	fem	fem	vijf	funf	cinq	cinco	cinque	piec	ot	viisi
six	seks	seks	zes	sechs	six	seis	sei	szesc	hat	kuusi
seven	sju	syv	zeven	sieben	sept	siete	sette	siedem	het	seitsemän
eight	atte	otte	acht	acht	huit	ocho	otto	osiem	nyolc	kahdeksän
nine	ni	ni	negen	neun	neuf	nueve	nove	dziewiec	kilenc	yhdeksän
ten	ti	ti	tien	zehn	dix	diez	dieci	dziesiec	tiz	kymmenen

Fonte: Johnson e Wichern (2007)

Programa - 2ª parte

Análise de agrupamentos - Exemplo

Cluster Dendrogram



```
dist(dados, "maximum")  
hclust(*, "complete")
```

Programa - 2ª parte

- Análise discriminante

- 1 Diferenciar populações

- 2 Classificar objetos em populações pré-definidas

Algumas áreas de aplicação de Análise Multivariada

- Psicologia
- Marketing
- Biologia
- Esportes
- Economia
- Controle de Qualidade

Bibliografia principal



Johnson, R. A. and Wichern, D. W. (2007)
Applied Multivariate Statistical Analysis.
5th edition. Prentice-Hall



Mingoti, S. A. (2005). Análise de Dados
Através de Métodos de Estatística Multiva-
riada: Uma Abordagem Aplicada. Belo Ho-
rizonte: UFMG

Bibliografia complementar

- Mardia, K. V., Kent, J. T. and Bibby, J. M. (1979). Multivariate Analysis. Academic Press.
- Hair, J. F., Tatham, R. L., Anderson, R. E. and Black, W. (1998). Multivariate Data Analysis, 5th edition, Prentice Hall.
- Morrison, D. F. (1967). Multivariate Statistical Methods. McGraw-Hill.
- Greenacre, M. J. (1984). Theory and Applications of Correspondence Analysis. London: Academic Press.

Critério de Avaliação

- $MF = (4 P1 + 4 P2 + 2 NL) / 10$

P1: Prova 1; P2: Prova 2; NL: Média de listas de exercícios

(até 2 pessoas)

(Sub: somente com justificativa oficial, por exemplo atestado médico reconhecido pela UBAS)

- Critério de Recuperação (Rec)

Se $3 \leq MF < 5$: O aluno poderá fazer a prova Rec

Nota da Rec: NR

Nova média final: NMF

$NMF = MF + (NR/2,5)$, se $NR \geq 7,5$; ou $\text{Max } MF, NR$, se

$NR \leq 5,0$; ou $5,0$, se $5,0 \leq NR < 7,5$

Datas importantes

- Datas importantes

1ª prova: 26/09/2011 das 21h às 23h.

2ª prova: 05/12/2011 das 21h às 23h.

Prova de recuperação: 20/01/2012 das 19h às 21h.

Horários

- **Aulas:** Segundas-feiras e sextas-feiras das 21h às 22h40 na sala 5-002
- **Atendimento:** Segundas-feiras das 16h às 17h ou mediante agendamento por e-mail

Apoio computacional



Algumas regras

- A USP exige no mínimo 70% de presença nas aulas
- A docente poderá fazer a chamada em qualquer instante do horário da aula, mesmo se os alunos já tiverem assinado a lista de presença, e eventuais ausências não serão abonadas
- Questionamentos a respeito da matéria poderão ser feitos a qualquer momento
- Conversas paralelas sobre outros assuntos não serão toleradas - os outros alunos têm direito a assistir a aula em silêncio

Algumas regras

- Os alunos devem desligar seus telefones celulares durante a aula
- Não é permitido o uso de notebook durante as aulas, exceto em possíveis aulas práticas
- Ao enviar e-mail para a docente, identificar o código da disciplina no assunto SME0263
- Eventuais dúvidas, consultar Normas da USP
<http://www.usp.br/leginf/>

Algumas regras - Sobre provas

- Será permitido usar 1 folha A4 com anotações próprias, que será entregue junto com a prova
- Será permitido o uso de calculadora própria
- Fraude em prova implicará a reprovação direta do(s) envolvido(s)
- Falsidade ideológica na realização da prova (ou assinatura de presença) implicará abertura de processo disciplinar

“Feliz aquele que transfere o que sabe e aprende o que ensina” (Cora Coralina)

Até a próxima aula!