



SME0822 Análise Multivariada **2º semestre de 2019**

Prof. Cibele Russo

cibele@icmc.usp.br

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

Sala 3-250A

Estagiário PAE: Fabiano Rodrigues Coelho

fabiano_icmc@usp.br

Principais objetivos do curso

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

Programa - 1ª parte

- Organização de dados e análise descritiva

Programa - 1ª parte

- Organização de dados e análise descritiva
- Vetores aleatórios e álgebra de matrizes

Programa - 1ª parte

- Organização de dados e análise descritiva
- Vetores aleatórios e álgebra de matrizes
- Média e matriz de covariâncias amostrais

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.

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

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

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 - 2^a parte

- Análise de componentes principais

Programa - 2ª parte

- Análise de componentes principais
- Análise fatorial

Programa - 2ª parte

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

Programa - 2ª parte

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

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
- Análise fatorial
- Análise de correlação canônica
- Análise de agrupamentos
- Análise discriminante
- Análise de correspondência

Programa - 2ª parte

- Análise de componentes principais (Pearson, 1901)
 - ① Explicar a estrutura de variância e covariância dos dados através de combinações lineares

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

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

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)

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

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

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

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

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

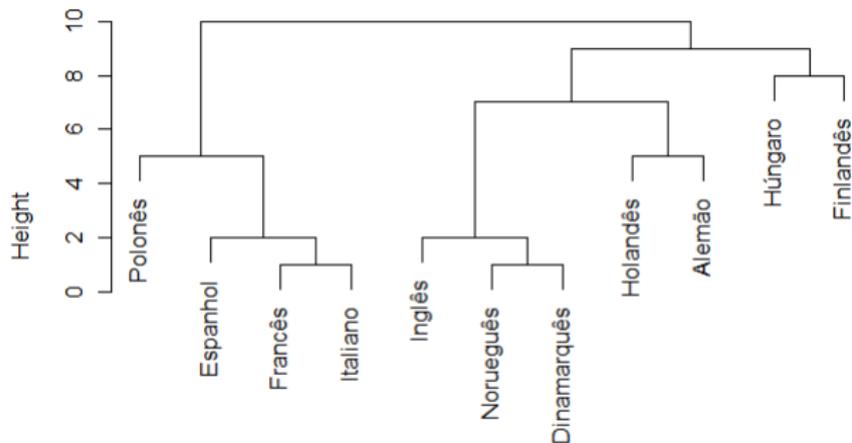
English (E)	Norwegian (N)	Danish (Da)	Dutch (Du)	German (G)	French (Fr)	Spanish (Sp)	Italian (I)	Polish (P)	Hungarian (H)	Finnish (Fi)
one	en	en	een	eins	un	uno	uno	jeden	egy	yksi
two	to	to	twee	zwei	deux	dos	due	dwa	ketto	kaksi
three	tre	tre	drie	drei	trois	tres	tre	trzy	harom	kolme
four	fire	fire	vier	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	seitseman
eight	atte	otte	acht	acht	huit	ocho	otto	osiem	nyolc	kahdeksan
nine	ni	ni	negen	neun	neuf	nueve	nove	dziewiec	kilenc	yhdeksan
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

Programa - 2ª parte

- Análise discriminante
 - 1 Diferenciar populações
 - 2 Classificar objetos em populações pré-definidas

Algumas áreas de aplicação de Análise Multivariada

Algumas áreas de aplicação de Análise Multivariada

- Psicologia

Algumas áreas de aplicação de Análise Multivariada

- Psicologia
- Marketing

Algumas áreas de aplicação de Análise Multivariada

- Psicologia
- Marketing
- Biologia

Algumas áreas de aplicação de Análise Multivariada

- Psicologia
- Marketing
- Biologia
- Esportes

Algumas áreas de aplicação de Análise Multivariada

- Psicologia
- Marketing
- Biologia
- Esportes
- Economia

Algumas áreas de aplicação de Análise Multivariada

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

Bibliografia principal

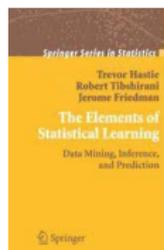


Johnson, R. A. and Wichern, D. W. (2007) Applied Multivariate Statistical Analysis. 5th edition. Prentice-Hall

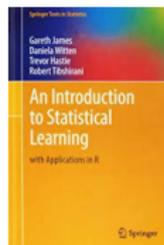


Lattin, J., Carroll, J. D., Green, P. E. (2011) Análise de dados multivariados - São Paulo: Cengage Learning.

Bibliografia principal



Friedman, J., Hastie, T., and; Tibshirani, R. (2001). The elements of statistical learning. New York, NY, USA:: Springer Series in Statistics.



James, G., Witten, D., Hastie, T., and; Tibshirani, R. (2013). An introduction to statistical learning (with applications in R). New York: Springer.

Bibliografia complementar

- Mingoti, S. A. (2005). Análise de Dados Através de Métodos de Estatística Multivariada: Uma Abordagem Aplicada. Belo Horizonte: UFMG
- 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

Sejam

P1: Prova 1;

P2: Prova 2.

NE: Nota de exercícios (que serão propostos ao longo do semestre)

Critério de Avaliação

Sejam

P1: Prova 1;

P2: Prova 2.

NE: Nota de exercícios (que serão propostos ao longo do semestre)

A média do semestre MS será calculada como

$$MS = \frac{4 P1 + 5 P2 + 1 NE}{10}$$

Critério de Avaliação

Sejam

P1: Prova 1;

P2: Prova 2.

NE: Nota de exercícios (que serão propostos ao longo do semestre)

A média do semestre MS será calculada como

$$MS = \frac{4 P1 + 5 P2 + 1 NE}{10}$$

Prova Substitutiva: somente com justificativa oficial, por exemplo atestado médico reconhecido pela UBAS.

Critério de Recuperação (REC)

O aluno poderá fazer a prova REC se, e somente se,

$$3 \leq MS < 5$$

Nota da REC: NR

Nova média após a REC: MF

$$MF = \begin{cases} 5, & \text{se } 5 \leq NR \leq (10 - MS); \text{ ou} \\ (MS + NR)/2 & \text{se } NR > (10 - MS); \text{ ou} \\ MS & \text{se } NR < 5 \end{cases}$$

Datas importantes

- Datas importantes
 - 23/09 - Prova 1
 - 25/11 - Prova 2
- Prova Substitutiva: 02/12/19
(somente para quem perder uma das provas, com justificativa oficial).
- Prova de Recuperação: a combinar.

Horários

- **Aulas:** Segundas-feiras das 19h às 20h40 e quartas-feiras das 21h às 22h40.
- **Atendimento PAE:** Quintas-feiras às 18h mediante agendamento por e-mail fabianoicmc@usp.br.
- **Atendimento docente:** Terças-feiras às 14h mediante agendamento por e-mail cibele@icmc.usp.br.

Atenção: Para toda correspondência sobre o curso, mencionar o código da disciplina no assunto SME0822.

Apoio computacional



Algumas regras

- A USP exige no mínimo 70% de presença nas aulas.

Algumas regras

- A USP exige no mínimo 70% de presença nas aulas.
- Questionamentos a respeito da matéria poderão ser feitos a qualquer momento.
- Evitar conversas de assuntos paralelos, a menos que seja imprescindível.

Algumas regras

- Os alunos devem desligar seus telefones celulares durante a aula.

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.

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 SME0822.

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 SME0822.
- Eventuais dúvidas, consultar Normas da USP
<http://www.usp.br/leginf/>

Algumas regras - Sobre provas

- Será permitido usar 1 folha A4, identificada com o nome do aluno, com anotações próprias (somente teoria).

Algumas regras - Sobre provas

- Será permitido usar 1 folha A4, identificada com o nome do aluno, com anotações próprias (somente teoria).
- Será permitido o uso de calculadora própria.

Algumas regras - Sobre provas

- Será permitido usar 1 folha A4, identificada com o nome do aluno, com anotações próprias (somente teoria).
- Será permitido o uso de calculadora própria.
- Fraude em prova (cola, whatsapp, etc) implicará a reprovação direta do(s) envolvido(s).

Algumas regras - Sobre provas

- Será permitido usar 1 folha A4, identificada com o nome do aluno, com anotações próprias (somente teoria).
- Será permitido o uso de calculadora própria.
- Fraude em prova (cola, whatsapp, etc) implicará a reprovação direta do(s) envolvido(s).
- Falsidade ideológica na realização da prova (ou assinatura de presença) pode implicar abertura de processo disciplinar.

Importante: E-mail

- E-mail é um canal oficial de comunicação da docente com os alunos.

Importante: E-mail

- E-mail é um canal oficial de comunicação da docente com os alunos.
- Checar e-mail cadastrado no JupiterWeb regularmente (pelo menos uma vez ao dia).

Importante: E-mail

- E-mail é um canal oficial de comunicação da docente com os alunos.
- Checar e-mail cadastrado no JupiterWeb regularmente (pelo menos uma vez ao dia).
- Ao enviar e-mail para a docente, identificar o código da disciplina no assunto SME0822.

Importante: E-mail

- E-mail é um canal oficial de comunicação da docente com os alunos.
- Checar e-mail cadastrado no JupiterWeb regularmente (pelo menos uma vez ao dia).
- Ao enviar e-mail para a docente, identificar o código da disciplina no assunto SME0822.
- É importante como forma de documentar ocorrências. Por exemplo: envio de listas, ausências por problemas de saúde, avisos sobre listas de exercícios, aulas.

Orientações de estudo

- Manter frequência de estudo do início ao final do semestre. O ideal é estudar o conteúdo de cada aula logo após ter sido ministrada, isso facilita o entendimento das aulas seguintes e diminui a quantidade de conteúdos para os dias anteriores à prova.

Orientações de estudo

- Manter frequência de estudo do início ao final do semestre. O ideal é estudar o conteúdo de cada aula logo após ter sido ministrada, isso facilita o entendimento das aulas seguintes e diminui a quantidade de conteúdos para os dias anteriores à prova.
- Estudar primeiramente o conteúdo dado em aula, buscando apoio em referências bibliográficas, em seguida resolver exemplos e exercícios. Não deixar para a véspera da prova.

Orientações de estudo

- Manter frequência de estudo do início ao final do semestre. O ideal é estudar o conteúdo de cada aula logo após ter sido ministrada, isso facilita o entendimento das aulas seguintes e diminui a quantidade de conteúdos para os dias anteriores à prova.
- Estudar primeiramente o conteúdo dado em aula, buscando apoio em referências bibliográficas, em seguida resolver exemplos e exercícios. Não deixar para a véspera da prova.
- Buscar referências bibliográficas logo no início do semestre, dando preferência às referências principais e complementares adotadas.

“We are drowning in information and starving for knowledge”
(Rutherford D. Rogers)