



SME0822 Análise Multivariada e Aprendizado Não-Supervisionado 2º semestre de 2020

Prof. Cibele Russo

cibele@icmc.usp.br

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

Monitora: Milene

milene.santos@usp.br

Principal objetivo do curso

- Estudo de métodos multivariados de análise de dados e de aprendizado não supervisionado.

Programa - 1ª parte

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

Programa - 2ª parte

- 1 Análise de componentes principais
- 2 Análise fatorial
- 3 Análise de correlação canônica
- 4 Análise de agrupamentos
- 5 Análise discriminante
- 6 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
 - 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

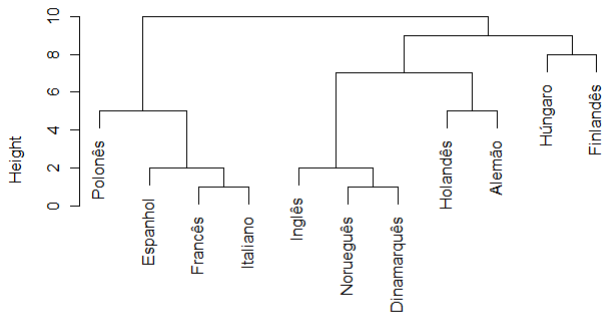
English (E)	Norwegian (N)	Danish (Da)	Dutch (Du)	German (G)	French (Fr)	Spanish (Sp)	Italian (I)	Polish (P)	Hungarian (H)	Finnish (Fi)
one	en	en	een	eins	un	uno	uno	jeden	egy	yksi
two	to	to	twee	zwei	deux	dos	due	dwa	ketto	kaksi
three	tre	tre	drie	drei	trois	tres	tre	trzy	harom	kolme
four	fire	fire	vier	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

Bibliografia principal

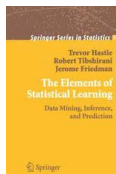


Johnson, R. A. and Wichern, D. W. (2007) Applied Multivariate Statistical Analysis. 5th edition. Prentice-Hall

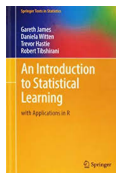


Lattin, J., Carroll, J. D., Green, P. E. (2011) Análise de dados multivariados - São Paulo: Cengage Learning.

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.

Critério de Avaliação

A avaliação será composta por **questionários no e-disciplinas** com frequência aproximadamente quinzenal, após a aula ao vivo semanal, que será encurtada nesses dias. Veja calendário no e-disciplinas. Ao final da disciplina **haverá uma avaliação final**, também feita no e-disciplinas, na data 17/12/2020 às 19h.

O critério de avaliação será:

$$MF = 0,6MQ + 0,4NP,$$

em que

- *MQ*: Média dos questionários
- *NP*: Nota da avaliação final

Critério de Avaliação

Importante: Os alunos devem aceitar o critério de avaliação em formulário que será enviado por e-mail.

Avaliações substitutivas: Caso o aluno perca alguma avaliação, deve justificar imediatamente à docente por e-mail, preferencialmente com documentação comprobatória e as avaliações substitutivas serão feitas ao final do semestre.

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

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

Horários

- **Aulas:** A docente fará aulas gravadas e o link será disponibilizado no e-disciplinas.

Uma vez por semana, às quintas-feiras às 19h haverá uma aula ao vivo no Google Meet para revisar o conteúdo das aulas gravadas da semana. A presença será contabilizada pela aula ao vivo e pelo acesso aos conteúdos da disciplina.

- **Atendimento docente:** Mediante agendamento por e-mail.
- **Atendimento monitora Milene:** Segundas-feiras às 18h via Google Meet.

Atenção: Em toda correspondência por e-mail sobre o curso, mencionar o código da disciplina no assunto: SME0822.

Apoio computacional



Recomendações para a prova

- É permitido consultar os materiais do e-disciplinas e usar calculadora, lápis e papel. Só será permitida uma tentativa para responder o questionário. O aluno é responsável por enviar o questionário ao final.
- O aluno assume a responsabilidade ética de não se comunicar com outros alunos durante a atividade, que é individual e individualizada.
- A senha dos sistemas USP é pessoal e intransferível. Qualquer tentativa de fraude será punida.

Importante: Comunicação

- E-disciplinas e e-mail são canais oficial de comunicação entre docente e alunos.
- Verifique se você recebe as notificações do e-disciplinas em seu e-mail.
- Se o aluno tiver problemas para acessar a internet ou acompanhar as atividades por outro motivo, deve notificar a docente o mais cedo possível.
- O Fórum do e-disciplinas pode ser usado para postar dúvidas.

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 às avaliações.
- Estudar primeiramente o conteúdo dado em aula, preferencialmente fazendo anotações, buscar apoio em referências bibliográficas ou buscar conteúdos similares na internet, resolver exemplos e exercícios.
- Repositório de livros eletrônicos da USP:
<https://www.aguia.usp.br/ebooks/>

Recadinho :

Ligue a câmera!

Ao menos quando

for interagir 😊