

Técnicas Multivariadas em Saúde

Lupércio França Bessegato
Dep. Estatística/UFJF

Introdução

Roteiro

1. Introdução
2. Dados Multivariados
3. Normal Multivariada
4. Análise de Componentes Principais
5. Análise Fatorial
6. Análise de Correlação Canônica
7. Análise de Conglomerados
8. Análise Discriminante
9. Análise de Correspondência
10. Referências

Técnicas Multivariadas em Saúde - 2015

Análise Multivariada

- Considera várias variáveis relacionadas simultaneamente
- Variáveis de interesse não-independentes uma das outras
- Associação entre conjuntos de medidas

Técnicas Multivariadas em Saúde - 2015

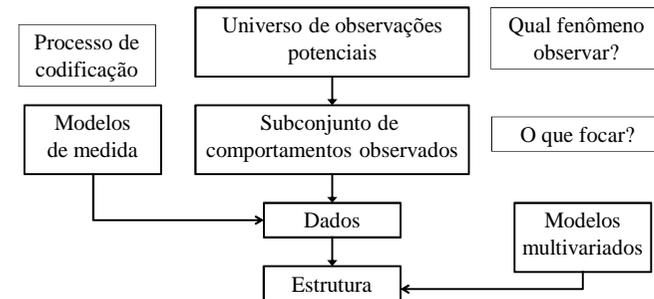
Objetos

- Entidades das quais são tomadas medidas
 - √ Itens, pessoas, organizações, etc.
 - √ São portadores de medidas
 - √ São medidos somente com respeito a certas variáveis de interesse

Técnicas Multivariadas em Saúde - 2015

Observação e Dados

- Dados:
 - √ Observações documentadas ou resultados de medição



Técnicas Multivariadas em Saúde - 2015

Variáveis

- Características ou propriedades
 - √ São os aspectos dos objetos que são medidos

Técnicas Multivariadas em Saúde - 2015

- Teoria subjacente sobre a área de interesse:
 - √ Necessária em cada etapa do processo de ir das observações aos dados
 - √ A interação do pesquisador (orientado pela teoria) com o ambiente de observações potenciais é que leva aos dados
- Última etapa:
 - √ Buscar estrutura associativa nos dados adequando os modelos multivariados

Técnicas Multivariadas em Saúde - 2015

Medida

- Processo pelo qual atribuem-se aos números (ou, algumas vezes, outros símbolos) características ou propriedades de objetos, de acordo com um procedimento predeterminado
- Escala de medida:
 - √ Refere-se à quantidade de informação que está contida na medida e o que ela nos informa sobre a relação entre dois objetos

Técnicas Multivariadas em Saúde - 2015

- Exemplo:

√ Cor de olhos

Variável original	Variáveis indicadoras	
Cor	Cor_1 = Preto	Cor_2 = Castanho
Preto	1	0
Castanho	0	1
Outros	0	0

Técnicas Multivariadas em Saúde - 2015

Escala Nominal

- Dados são descritos categoricamente
 - √ Objeto é classificado em um dos conjuntos de grupos exaustivos e mutuamente exclusivos
 - √ Números não transmitem nenhuma informação
 - São rótulos para as categorias
 - √ Em geral, precisam ser codificados apropriadamente

Técnicas Multivariadas em Saúde - 2015

Escala Ordinal

- Informa apenas se um objeto (i) tem mais ou menos a mesma quantidade de características medidas de outro objeto (j)
- Exemplo:
 - √ Classes sócio-econômicas
- Operações aritméticas não tem significado
 - √ Exemplo: magnitude das diferenças

Técnicas Multivariadas em Saúde - 2015

Escala Intervalar

- Ordena os objetos quanto à característica medida
- Possui uma unidade de medida constante
- A origem (ponto zero) dessa escala é arbitrária.
- É válido comparar diferenças entre itens nessa escala
- Exemplo
 - √ 0°C não corresponde à ausência de temperatura
 - √ Diferença entre 20°C e 40°C é maior que a diferença entre 30°C e 40°C

Técnicas Multivariadas em Saúde - 2015

Escala de Razão

- Ordena os objetos quanto à característica medida
- Possui uma unidade de medida constante
- Possui origem (ponto zero) única.
- Exemplo:
 - √ Idade, preço, comprimento, temperatura ($^{\circ}\text{K}$)
- Aumenta a quantidade de informações fornecidas
 - √ Permite o uso de todas as operações aritméticas

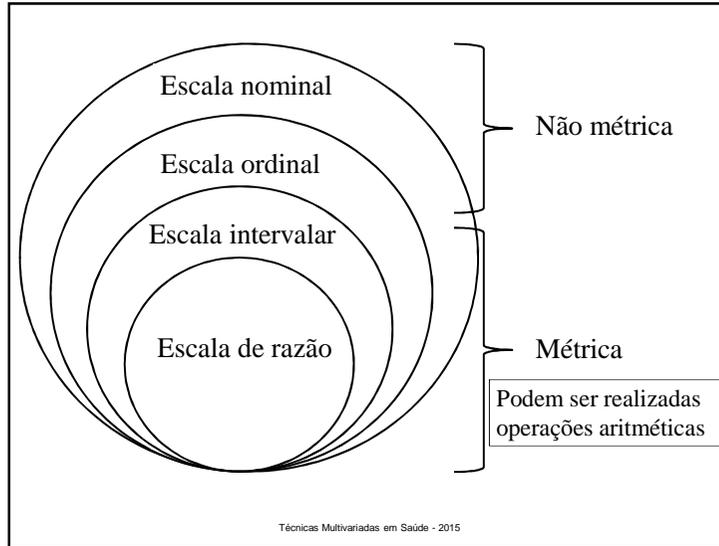
Técnicas Multivariadas em Saúde - 2015

- √ Um objeto com 40°C não é duas vezes mais quente que um objeto a 20°C
- √ Mesmos objetos na escala de temperatura Fahrenheit
 - $40^{\circ}\text{C} = (32 + 40 \times 9/5) = 104^{\circ}\text{F}$
 - $20^{\circ}\text{C} = (32 + 20 \times 9/5) = 68^{\circ}\text{F}$
- √ Razão:
 - 2 : 1, em $^{\circ}\text{C}$
 - 1,52:1 (26:17), em $^{\circ}\text{F}$
- √ Diferença:
 - $40^{\circ}\text{C} - 0^{\circ}\text{C}$ é o dobro de $20^{\circ}\text{C} - 0^{\circ}\text{C}$
 - $104^{\circ}\text{F} - 32^{\circ}\text{F}$ é o dobro de $68^{\circ}\text{F} - 32^{\circ}\text{F}$
- √ Invariante a transformações lineares positivas
 - Transformações da forma $X^* = a + b X$, $a > 0$

Técnicas Multivariadas em Saúde - 2015

- √ Um objeto com 4 m é duas vezes mais extenso que um objeto com 2m
- √ Mesmos objetos medidos em cm
 - 400 cm tem o dobro da extensão de 200 cm (mesma razão)
- √ Invariante sob transformações de proporção positivas
 - Transformações da forma $X^* = c X$, $c > 0$

Técnicas Multivariadas em Saúde - 2015



Conjunto de Dados #1

- Bumpus (1898)
- Pardais sobreviventes de tempestade
 - √ Dados de 1/Fev/1898
 - √ Medidas morfológicas e peso de 49 pássaros fêmeas
 - √ 28 morreram, 21 não morreram
- Dados: *birds.csv* (*birds.txt*)

Técnicas Multivariadas em Saúde - 2015

Organização de Dados Multivariados

$$\mathbf{X}_{n \times p} = \begin{matrix} & \begin{matrix} 1 & 2 & \dots & p \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ \vdots \\ n \end{matrix} & \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix} \end{matrix}$$

p: número de variáveis
n: número de objetos

- √ Matriz:
 - Organização em arranjo estrutural de dados multivariados (métricos ou não métricos)
- √ Colunas: variáveis (características medidas de objetos)
- √ Linhas: objetos (lista de características medidas de um objeto)

Técnicas Multivariadas em Saúde - 2015

- √ Variáveis:
 - sv: 1= vivo, 2= morto.
 - ag: 1= adulto, 2 = jovem.
 - tl: comprimento total (bico à ponta da cauda), em mm.
 - ae: extensão alar, (ponta a ponta de asas), em mm.
 - wt: peso, em gramas.
 - bh: comprimento bico e cabeça, em mm.
 - hl: comprimento do úmero (osso braço), em polegadas.
 - fl: comprimento do fêmur (osso coxa), em polegadas.
 - tt: comprimento da tíbia-tarso (osso perna), em polegadas.
 - sk: amplitude do crânio, em polegadas.
 - kl: comprimento da quilha do esterno, em polegadas.

Técnicas Multivariadas em Saúde - 2015

• Questões interessantes:

1. Como as variáveis estão relacionadas?
(Um valor grande de uma variável tende a ocorrer com valores grandes para as outras variáveis?)
2. Os sobreviventes e os não sobreviventes têm diferenças estatisticamente significantes para seus valores médios das variáveis?
3. Os sobreviventes e os não sobreviventes mostram quantidades similares de variação para a variável?
4. Se os sobreviventes e não sobreviventes diferem em termos das distribuições das variáveis, então é possível construir alguma função dessas variáveis que separe os dois grupos?
(Índice de ajuste Darwiano dos pardais: valores grandes da função tendem a ocorrer com os sobreviventes)

• Conclusão de Bumpus (1898):

- √ Os sobreviventes são mais curtos e pesam menos, tem ossos das asas mais longos, pernas mais longas, esternos mais longos e maior capacidade cerebral
- O processo de eliminação seletiva é mais severo com indivíduos extremamente variáveis (independente da direção)
 - √ É tão perigoso estar acima de um certo padrão de excelência orgânica como estar visivelmente abaixo desse padrão

Técnicas Multivariadas em Saúde - 2015

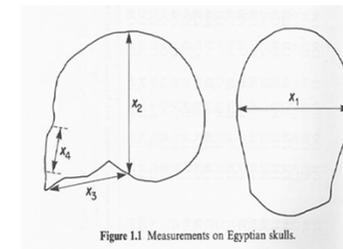
Conjunto de Dados #2

- Thomson e Randall-Maciver (1905)
- Medidas em crânios masculinos da área de Tebas
 - √ 5 amostras de 30 crânios cada uma:
 - Período pré-dinástico primitivo (~ 4.000 aC)
 - Período pré-dinástico antigo (~3.300 aC)
 - 12ª e 13ª dinastias (~ 1.850 aC)
 - Período Ptolemaico (~ 200 aC)
 - Período romano (~150 dC)
 - √ Dados: *skulls{ade4}* ou *skulls2.csv* (*skulls.txt*)

Técnicas Multivariadas em Saúde - 2015

• Variáveis:

- √ MB (X_1): Amplitude máxima do crânio
- √ BH (X_2): Altura basilobregmática do crânio
- √ BL (X_3): Comprimento basiloalveolar do crânio
- √ NH (X_4): Altura nasal do Crânio
- √ Ano: Ano aproximado de formação do crânio



Técnicas Multivariadas em Saúde - 2015

- Questões interessantes:

1. Como estão relacionadas as quatro medidas?
2. Existem diferenças estatisticamente significantes nas médias amostrais das variáveis?
 - Elas refletem mudanças graduais ao longo do tempo na forma e no tamanho dos crânios?
3. Existem diferenças significantes nos desvios-padrão para as variáveis?
 - Elas refletem mudanças ao longo do tempo na quantidade de variação?
4. É possível construir uma função das quatro variáveis que, em algum sentido, descreva as mudanças ao longo do tempo?

Técnicas Multivariadas em Saúde - 2015

Conjunto de Dados #3

- McKechnie et al. (1975)
- Distribuição de uma borboleta:
 - √ 16 colônias de borboletas *Euphydryas editha* na Califórnia e Oregon
- Dados: *butterfly*{ade4}

Técnicas Multivariadas em Saúde - 2015

- Conclusão de Thomson e Randall-Maciver (1905):

- √ Há diferenças entre as cinco amostras que podem ser explicadas parcialmente como tendências no tempo.
- √ As razões para as aparentes mudanças são desconhecidas
- √ A migração de outras raças dentro da região pode ter sido o fator mais importante.

Técnicas Multivariadas em Saúde - 2015

- Variáveis ambientais (*butterfly*\$*envir*):

- √ Altitude, em pés
- √ Precipitation: precipitação anual, em polegadas
- √ Temp_max: temperatura máxima, em °F
- √ Temp_min: temperatura mínima, em °F

- Variáveis genéticas: (*butterfly*\$*genet*):

- √ Frequências de mobilidade gênica *Fósforo Glucose-Isomerase (Pgi)* para colônias de borboletas
 - Níveis: 0,4; 0,6; 0,8; 1; 1,16; 1,3(representa diferentes tipos genéticos de Pgi de modo que as frequências para uma colônia (somando a 100%) mostram as frequências do diferentes tipos para a *E. editha* naquele local)

Técnicas Multivariadas em Saúde - 2015

• Dados agregados:

√ Variáveis ambientais e genéticas

```
> borboleta <- cbind(butterfly$envir, butterfly$genet)
> borboleta
  Altitude Precipitation Temp_Max Temp_Min 0.4 0.6 0.8 1 1.16 1.3
SS      500           43      98      17  0  3  22  57  17  1
SB      800           20      92      32  0 16  20  38  13 13
WSB     570           28      98      26  0  6  28  46  17  3
JRC     550           28      98      26  0  4  19  47  27  3
JRH     550           28      98      26  0  1  8  50  35  6
SJ      380           15      99      28  0  2  19  44  32  3
CR      930           21      99      28  0  0  15  50  27  8
UO      650           10     101      27 10 21  40  25  4  0
LO      600           10     101      27 14 26  32  28  0  0
DP     1500           19      99      23  0  1  6  80  12  1
PZ     1750           22     101      27  1  4  34  33  22  6
MC     2000           58     100      18  0  7  14  66  13  0
IF     2500           34     102      16  0  9  15  47  21  8
AF     2000           21     105      20  3  7  17  32  27 14
GH     7850           42      84       5  0  5  7  84  4  0
GL    10500           50      81     -12  0  3  1  92  4  0
```

Técnicas Multivariadas em Saúde - 2015

• Questões interessantes:

- √ As frequências Pgi são similares para as colônias espacialmente próximas?
- √ O quanto, se algum, as frequências Pgi estão relacionadas às variáveis ambientais?

Técnicas Multivariadas em Saúde - 2015

• Rótulos das linhas: Local das colônias

√ SS (Oregon), SB, WSB, JRC, JRH, SJ, CR, UO, LO, DP, PZ, MC, IF, AF, GH, GL.

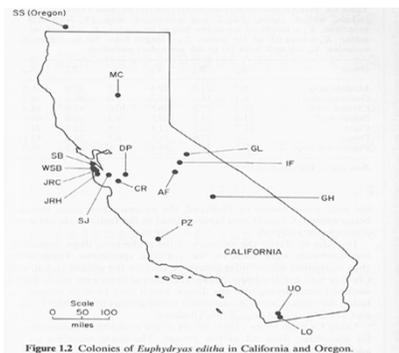


Figure 1.2. Colonies of *Euphydryas editha* in California and Oregon.

Fonte: Manly, 2005

Técnicas Multivariadas em Saúde - 2015

• Análise das respostas:

- √ Se a composição genética das colônias foi fortemente determinada pelas migrações passadas e presentes
 - Frequências gênicas tenderão a ser similares para colônias que estão localizadas nas proximidades
- √ Se o meio ambiente é mais importante
 - Deve aparecer relacionamento entre as frequências gênicas e as variáveis ambientais
 - Colônias próximas têm frequências gênicas similares somente se têm ambientes similares)
 - Pode ser difícil chegar a essa conclusão

Técnicas Multivariadas em Saúde - 2015

Conjunto de Dados #4

- Coleção de ossos caninos do período em torno de 3.500 aC até o presente
 - √ Local: Tailândia
- A origem dos cães pré-históricos é incerta
 - √ Podem descender dos chacais dourados (*canis aureus*) ou do lobo (não é nativo da Tailândia)
- Fontes de origem mais próximas:
 - √ Parte ocidental da China: *canis lupus chanco*
 - √ Subcontinente indiano: *canis lupus pallides*
- Dados: *canine.txt*

Técnicas Multivariadas em Saúde - 2015

• Variáveis:

- √ X_1 : comprimento da mandíbula, em mm.
- √ X_2 : largura da mandíbula, em mm.
- √ X_3 : largura do côndilo da mandíbula, em mm.
- √ X_4 : altura da mandíbula abaixo do 1º molar, em mm.
- √ X_5 : comprimento do 1º molar, em mm.
- √ X_6 : largura do 1º molar, em mm.
- √ X_7 : comprimento do 1º ao 3º molar, em mm.
- √ X_8 : comprimento do 1º ao 4º molar, em mm.
- √ X_9 : largura do maxilar inferior, em mm.

Técnicas Multivariadas em Saúde - 2015

- Higham et al. (1980)
- Objetivo do estudo:
 - √ Comparar medidas da mandíbula dos espécimes pré-históricos disponíveis com as mesmas medidas no chacal dourado, lobo chinês, lobo indiano, dingo (originário da Índia), cuon: *cuon alpinus* (originário do Sudeste da Ásia) e cães modernos da Tailândia.

Técnicas Multivariadas em Saúde - 2015

• Questões interessantes:

1. O que as medidas sugerem sobre o relacionamento entre os grupos?
2. Como os cães pré-históricos parecem se relacionar com os outros grupos?

Técnicas Multivariadas em Saúde - 2015

Conjunto de Dados #5

- Porcentagem da força de trabalho em nove diferentes tipos de indústrias para 30 países europeus
 - √ De 1980 a 1995
- Dados:

Técnicas Multivariadas em Saúde - 2015

- Variáveis:

- √ Group: Eastern= Leste europeu; EFTA: área europeia de livre comércio; EU: União Europeia; Other: outros países
- √ AGR: agricultura, floresta e pesca
- √ MIN: mineração e exploração de pedreiras
- √ MAN: fabricação
- √ PS: fornecimento de energia elétrica
- √ CON: construção
- √ SER: serviços
- √ FIN: finanças
- √ SPS: serviços sociais e pessoais
- √ TC: transportes e comunicações

Técnicas Multivariadas em Saúde - 2015

- Objetivo do estudo:
 - √ Isolar grupos de países com padrões similares
 - √ Auxiliar o entendimento dos relacionamentos entre países
- Cada linha do conjunto de dados soma 100%

Técnicas Multivariadas em Saúde - 2015

- Questões:

1. É possível isolar grupos de países com padrões similares de emprego?
2. Qual o relacionamento entre os países com relação aos níveis de emprego?
 - Podem ser de interesse diferenças entre países que estejam relacionadas a grupos políticos (União Europeia; área europeia de livre comércio; países do leste europeu e outros países)

Técnicas Multivariadas em Saúde - 2015

Conjunto de Dados #6

- Estudo de poluição do ar em 41 cidades dos EUA
 - √ Ano: 1970
- Dados: *Usairpollution*{MVA}

Técnicas Multivariadas em Saúde - 2015

• Questões:

- √ Como o nível de poluição, medido pela concentração de dióxido de enxofre está relacionado com as outras seis variáveis?
- √ Como reduzir a quantidade de variáveis que descrevem as cidades, simplificando os dados?
- √ É possível encontrar variáveis latentes que descrevam as cidades?

Técnicas Multivariadas em Saúde - 2015

• Variáveis:

- √ SO₂: conteúdo de dióxido de enxofre no ar, em $\mu\text{g}/\text{m}^3$.
- √ temp: temperatura média anual (°F)
- √ manu: quantidade de empresas manufatureiras empregando pelo menos 20 empregados.
- √ popul: população (censo 1970), em milhares.
- √ wind: velocidade média anual de vento, em milhas/h
- √ precip: precipitação média anual, em polegadas
- √ predays: número médio anual de dias com precipitação

Técnicas Multivariadas em Saúde - 2015

Técnicas em Estatística Multivariada

• Técnicas Exploratórias:

- √ Sintetização da estrutura de variabilidade dos dados
 - Análise de componentes principais, análise fatorial, análise de correlações canônicas, análise de agrupamentos, análise discriminante

√ Técnicas de Inferência Estatística:

- Métodos de estimação de parâmetros, testes de hipóteses, análise de variância, análise de covariância, análise de regressão multivariada

Técnicas Multivariadas em Saúde - 2015

Visão dos Métodos Multivariados

- **Análise de componentes principais:**
 - √ Redução do número de variáveis a um número menor de índices (componentes principais)
 - Componentes principais: combinações lineares das variáveis originais
 - √ Pode acontecer de que duas ou mais componentes principais forneçam um bom resumo de todas as variáveis originais

Técnicas Multivariadas em Saúde - 2015

- **Análise fatorial:**

- √ Estudar a variação em uma quantidade de variáveis originais, usando um número menor de fatores (variáveis índices)
- √ Assume-se que cada variável original pode ser expressa como uma combinação linear dos fatores mais um termo residual (reflete quanto a variável é independente das outras variáveis)

Técnicas Multivariadas em Saúde - 2015

- **Exemplo – Pardais sobreviventes**

- √ Muito da variação nas medidas do corpo dos pardais (X_1 a X_5) está relacionada com:
 - Tamanho geral dos pássaros (I_1)

$$I_1 = k_1X_1 + k_2X_2 + k_3X_3 + k_4X_4 + k_5X_5$$
 - Contraste entre as três primeiras medidas e as duas últimas

$$I_2 = k^*_1X_1 + k^*_2X_2 + k^*_3X_3 - k^*_4X_4 - k^*_5X_5$$
- √ **Análise de componentes principais:**
 - Maneira de simplificar os dados, reduzindo o número de variáveis

Técnicas Multivariadas em Saúde - 2015

- **Modelo:**

- $$X_1 = a_{11}F_1 + a_{12}F_2 + \epsilon_1,$$
- $$X_2 = a_{21}F_1 + a_{22}F_2 + \epsilon_2,$$
- $$X_3 = a_{31}F_1 + a_{32}F_3 + \epsilon_3,$$
- $$X_4 = a_{41}F_1 + a_{42}F_4 + \epsilon_4,$$
- $$X_5 = a_{51}F_1 + a_{52}F_5 + \epsilon_5.$$
- a_{ij} : constantes
 - $F_1; F_2$: fatores
 - ϵ_i : variação em X_i que é independente da variação nas outras variáveis
 - √ F_1 : fator tamanho
 - Alguns pássaros tendem a ser grandes, alguns pássaros tendem a ser pequenos, em todas as medidas do corpo
 - √ F_2 : mede algum aspecto da forma dos pássaros
 - Alguns coeficientes positivos e alguns negativos

Técnicas Multivariadas em Saúde - 2015

- **Análise discriminante:**

- √ Separação em diferentes grupos com base nas medidas disponíveis

- √ Encontrar combinações lineares convenientes das variáveis originais para atingir o objetivo desejado

- √ Exemplo:

- Quão bem os pardais sobreviventes e não sobreviventes podem ser separados usando suas medidas do corpo?

Técnicas Multivariadas em Saúde - 2015

- **Escalonamento multidimensional:**

- √ Mapeamento, com distância, entre objetos, mostrando como eles estão relacionados

- √ Visualização possível em até três dimensões

- √ Alternativa para a análise de agrupamentos

- √ Exemplos:

1. Encontrar similaridades entre cães pré-históricos tailandeses e outros animais

2. Agrupar os países europeus em termos de suas similaridade no padrão de empregos.

Técnicas Multivariadas em Saúde - 2015

- **Análise de agrupamento:**

- √ Identificação de grupos de objetos similares

- √ Exemplos:

1. Encontrar similaridades entre cães pré-históricos tailandeses e outros animais

2. Agrupar os países europeus em termos de suas similaridade no padrão de empregos.

Técnicas Multivariadas em Saúde - 2015

- **Métodos de ordenação:**

- √ Produção de eixos nos quais um conjunto de objetos de interesse pode ser representado

- √ Algumas técnicas de ordenação:

- Análise de componentes principais

- Escalonamento multidimensional

Técnicas Multivariadas em Saúde - 2015

- **Análise de correlação canônica:**
 - √ As variáveis (não os objetos) são divididos em dois grupos e o interesse está centrado no relacionamento entre elas
 - √ Exemplo – Colônias de borboleta:
 - Encontrar relacionamento entre as variáveis ambientais e as variáveis genéticas.

Técnicas Multivariadas em Saúde - 2015

Usos das Técnicas Multivariadas

- **Construção de índices:**
 - √ Sintetizar em uma única variável a informação de todas as variáveis que foram medidas sobre o fenômeno
 - Análise de componentes principais
 - Análise fatorial
 - Análise de correlação canônica

Técnicas Multivariadas em Saúde - 2015

- **Análise de correspondência:**
 - √ Dados sobre a abundância de cada uma das várias características para cada elemento de um conjunto de objetos
 - √ Exemplo – Ecologia:
 - Objetos de interesse: diferentes locais
 - Características: diferentes espécies
 - Dados: abundância de espécies em locais
 - Objetivo: tornar claro os relacionamentos entre locais (expressos por distribuições das espécies) e os relacionamentos entre as espécies (expressos por distribuições dos locais).

Técnicas Multivariadas em Saúde - 2015

- **Classificação e discriminação:**
 - √ Busca-se a divisão de conjunto de dados em grupos, de modo que os grupos tenham coesão interna e sejam heterogêneos entre si
 - Análise de agrupamentos (segmentação de mercado)
 - Análise discriminante (classificação de crédito)

Técnicas Multivariadas em Saúde - 2015

- Inferência estatística:
 - √ Comparação de grupos em relação às médias de variáveis medidas conjuntamente
 - √ Regressão multivariada:
 - Análise do efeito de fatores externos não controlados nas variáveis que são monitoras regularmente

Técnicas Multivariadas em Saúde - 2015

Conceitos Básicos

- Matriz ($\mathbf{A}_{m \times n}$):
 - √ Arranjo de números com m linhas e n colunas

$$\mathbf{A}_{m \times n} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}.$$

- √ Matriz quadrada: quantidade de linhas é igual à quantidade de colunas (m = n)

- Matriz transposta ($\mathbf{A}'_{n \times m}$):

$$\mathbf{A}'_{n \times m} = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{mn} \end{bmatrix}.$$

Técnicas Multivariadas em Saúde - 2015

Álgebra Matricial

- Matriz simétrica:
 - √ A matriz e sua transposta são iguais

$$\mathbf{A}_{n \times n} = \mathbf{A}'_{n \times n}$$

- √ Matriz simétrica será sempre quadrada

Técnicas Multivariadas em Saúde - 2015

- Vetores:
 - √ Vetor-coluna ($\mathbf{c}_{m \times 1}$): $\mathbf{c}_{1 \times m} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_m \end{bmatrix}$
 - √ Vetor-linha: ($\mathbf{r}_{1 \times n}$): $\mathbf{r}_{n \times 1} = [r_1, r_2, \dots, r_n]$
 - √ Notação: vetores serão sempre colunas
 - Vetor-linha: $\mathbf{c}'_{1 \times n}$

Técnicas Multivariadas em Saúde - 2015

- Igualdade de matrizes:
 - √ Duas matrizes são iguais somente se:
 - Tiverem o mesmo tamanho
 - Seus elementos correspondentes forem iguais
 - $\mathbf{A}_{m \times n} = \mathbf{B}_{m \times n} \Leftrightarrow a_{ij} = b_{ij}, \forall i = 1, 2, \dots, m \text{ e } j = 1, 2, \dots, n$
- Traço de uma matriz:
 - √ Soma dos elementos da diagonal principal
 - √ Válido apenas para matrizes quadradas
$$\text{tr}(\mathbf{A}_n) = a_{11} + a_{22} + \dots + a_{nn}.$$

Técnicas Multivariadas em Saúde - 2015

- Matriz nula ($\mathbf{0}_{m \times n}$):
 - √ Todos os elementos da matriz são zeros
- Matriz diagonal (\mathbf{D}_n):
 - √ Apenas os elementos da diagonal não são zeros
$$\mathbf{D}_n = \begin{bmatrix} d_1 & 0 & \dots & 0 \\ 0 & d_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & d_n \end{bmatrix}.$$
- Matriz identidade:
 - √ Matriz diagonal com os termos na diagonal iguais a 1
$$\mathbf{I}_n = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}.$$

Técnicas Multivariadas em Saúde - 2015

Operações com Matrizes

- Adição e Subtração:
 - √ Soma (subtração), elemento por elemento, de duas matrizes de mesmo tamanho

$$\mathbf{A}_{m \times n} + \mathbf{B}_{m \times n} = \begin{bmatrix} a_{11} + b_{11} & a_{12} + b_{12} & \dots & a_{1n} + b_{1n} \\ a_{21} + b_{21} & a_{22} + b_{22} & \dots & a_{2n} + b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} + b_{m1} & a_{m2} + b_{m2} & \dots & a_{mn} + b_{mn} \end{bmatrix}.$$

$$\mathbf{A}_{m \times n} - \mathbf{B}_{m \times n} = \begin{bmatrix} a_{11} - b_{11} & a_{12} - b_{12} & \dots & a_{1n} - b_{1n} \\ a_{21} - b_{21} & a_{22} - b_{22} & \dots & a_{2n} - b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} - b_{m1} & a_{m2} - b_{m2} & \dots & a_{mn} - b_{mn} \end{bmatrix}.$$

Técnicas Multivariadas em Saúde - 2015

• Multiplicação de matriz por escalar:

√ Escalar: número

$$k \mathbf{A}_{m \times n} = \begin{bmatrix} k a_{11} & k a_{12} & \dots & k a_{1n} \\ k a_{21} & k a_{22} & \dots & k a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ k a_{m1} & k a_{m2} & \dots & k a_{mn} \end{bmatrix}$$

Técnicas Multivariadas em Saúde - 2015

• Quando **A** e **B** são matrizes quadradas, os produtos **BxA** e **BxA** estão definidos.

• Multiplicação entre matrizes não é comutativa

√ Contra-exemplo:

$$\begin{bmatrix} 2 & -1 \\ 1 & 0 \end{bmatrix} \times \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 \times 1 - 1 \times 0 & 2 \times 1 - 1 \times 1 \\ 1 \times 1 + 1 \times 0 & 1 \times 1 + 1 \times 1 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \times \begin{bmatrix} 2 & -1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 \times 2 + 1 \times 1 & -1 \times 1 + 1 \times 1 \\ 0 \times 2 + 1 \times 1 & -1 \times 0 + 1 \times 1 \end{bmatrix} = \begin{bmatrix} 3 & 0 \\ 1 & 1 \end{bmatrix}$$

Técnicas Multivariadas em Saúde - 2015

• Multiplicação entre duas matrizes:

√ Definida apenas se o número de colunas da matriz à esquerda for igual ao número de linhas da matriz à esquerda

$$\mathbf{A}_{m \times n} \times \mathbf{B}_{n \times p} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \times \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1p} \\ b_{21} & b_{22} & \dots & b_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \dots & b_{np} \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{i=1}^n a_{1i}b_{i1} & \sum_{i=1}^n a_{1i}b_{i2} & \dots & \sum_{i=1}^n a_{1i}b_{ip} \\ \sum_{i=1}^n a_{2i}b_{i1} & \sum_{i=1}^n a_{2i}b_{i2} & \dots & \sum_{i=1}^n a_{2i}b_{ip} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^n a_{mi}b_{i1} & \sum_{i=1}^n a_{mi}b_{i2} & \dots & \sum_{i=1}^n a_{mi}b_{ip} \end{bmatrix}_{m \times p}$$

- No exemplo, a multiplicação **BxA** não está definida

Técnicas Multivariadas em Saúde - 2015

• Matriz Inversa (**A**⁻¹):

√ A inversa da matriz quadrada **A** é definida como:

$$\mathbf{A}_n \times \mathbf{A}_n^{-1} = \mathbf{A}_n^{-1} \times \mathbf{A}_n = \mathbf{I}_n$$

√ Se existir, a matriz inversa é a única matriz que satisfaz a expressão acima

√ Nem todas as matrizes quadradas têm inversas:

- A matriz não tem inversa se seu determinante for nulo (Matriz singular)

• Matriz ortogonal:

$$\mathbf{A}_n^{-1} = \mathbf{A}_n'$$

Técnicas Multivariadas em Saúde - 2015

Autovalores e Autovetores

- Considere o conjunto de equações lineares:

$$a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n = \lambda x_1$$

$$a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n = \lambda x_2$$

$$\vdots$$

$$a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nn}x_n = \lambda x_n$$

√ λ é escalar

√ Notação matricial do sistema de equações:

$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} \times \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \lambda \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

$$\mathbf{A}_{n \times n} \mathbf{x}_{n \times 1} = \lambda \mathbf{x}_{n \times 1}$$

Técnicas Multivariadas em Saúde - 2015

- Solução:

$$\mathbf{A} \mathbf{x} = \lambda \mathbf{x}$$

$$\mathbf{A} \mathbf{x} - \lambda \mathbf{x} = \mathbf{0}$$

$$(\mathbf{A} - \lambda \mathbf{I}) \mathbf{x} = \mathbf{0}$$

√ Essas equações valem apenas para certos valores de \mathbf{A}

(Podem existir apenas n desses valores particulares de λ)

√ Valores de λ :

- Autovalores ou raízes latentes

- Autovetores:

√ Para o i-ésimo autovalor λ_i , as equações podem ser resolvidas para um vetor resultante de valores de \mathbf{x} , $\mathbf{x}' = (x_1, x_2, \dots, x_n)$ ou qualquer múltiplo desse vetor.

Técnicas Multivariadas em Saúde - 2015