

ME731 - Métodos em Análise Multivariada

– Análise de Correlação Canônica II –

Prof. Carlos Trucíos
ctrucios@unicamp.br
ctruciosm.github.io

Instituto de Matemática, Estatística e Computação Científica,
Universidade Estadual de Campinas

Aula 20



UNICAMP

Agenda I

- 1 Introdução
- 2 Implementação
- 3 Correlação canônica para variáveis qualitativas
- 4 Correlação canônica para variáveis mixtas
- 5 Exemplos

Introdução

Introdução

- Na aula anterior vimos que a implementação do pacote stats e implementação feita *na mão* não concidiam.

Introdução

- Na aula anterior vimos que a implementação do pacote stats e implementação feita *na mão* não concidiam.
- Na aula de hoje iremos mais a fundo para entender o que está acontecendo.

Introdução

- Na aula anterior vimos que a implementação do pacote stats e implementação feita *na mão* não concidiam.
- Na aula de hoje iremos mais a fundo para entender o que está acontecendo.
- Além disso, aprenderemos como utilizar análise de correlação canônica quanto as variáveis são qualitativas.

Implementação

Implementação

```
library(expm)
cc_me731 = function(x, y) {
  S_xx <- cov(x)
  S_yy <- cov(y)
  S_xy <- cov(x, y)
  M <- sqrtm(solve(S_xx)) %*% S_xy %*% sqrtm(solve(S_yy))
  decomposicao_svd <- svd(M)
  a <- sqrtm(solve(S_xx)) %*% decomposicao_svd$u
  b <- sqrtm(solve(S_yy)) %*% decomposicao_svd$v
  lambda <- decomposicao_svd$d
  return(list(a, b, lambda))
}
```


Implementação

```
library(dplyr)
dados <- read.table("https://raw.githubusercontent.com/ctruciosm/ctruciosm/master/data/ctruciosm.csv",
  colnames(dados) <- c("name", "economy",
    "service", "value",
    "price", "design",
    "sporty", "safety",
    "handling")
X <- dados %>% dplyr::select(price, value)
Y <- dados %>% dplyr::select(-price, -value, -name)
cor_canonica_na_mao <- cc_me731(X, Y)
cc_results <- cancor(X,Y)
```

Implementação

```
# Variáveis canônicas (pacote stats)  
a_1 <- matrix(cc_results$xcoef[,1], ncol = 1)  
a_2 <- matrix(cc_results$xcoef[,2], ncol = 1)  
b_1 <- matrix(cc_results$ycoef[,1], ncol = 1)  
b_2 <- matrix(cc_results$ycoef[,2], ncol = 1)  
canonica_x_1 <- as.matrix(X) %*% a_1  
canonica_x_2 <- as.matrix(X) %*% a_2  
canonica_y_1 <- as.matrix(Y) %*% b_1  
canonica_y_2 <- as.matrix(Y) %*% b_2
```

Implementação

```
# Variáveis canônicas (nossa implementação)
our_a_1 <- matrix(cor_canonica_na_mao[[1]][,1], ncol = 1)
our_a_2 <- matrix(cor_canonica_na_mao[[1]][,2], ncol = 1)
our_b_1 <- matrix(cor_canonica_na_mao[[2]][,1], ncol = 1)
our_b_2 <- matrix(cor_canonica_na_mao[[2]][,2], ncol = 1)
our_canonica_x_1 <- as.matrix(X) %*% our_a_1
our_canonica_x_2 <- as.matrix(X) %*% our_a_2
our_canonica_y_1 <- as.matrix(Y) %*% our_b_1
our_canonica_y_2 <- as.matrix(Y) %*% our_b_2
```

Implementação

Lembre que $\text{Cov}(\eta_i, \phi_j) = 0$.

Implementação

Lembre que $\text{Cov}(\eta_i, \phi_j) = 0$.

```
c(cov(canonica_x_1, canonica_x_2),  
  cov(canonica_y_1, canonica_y_2),  
  cov(canonica_x_1, canonica_y_2))
```

```
## [1] -1.820505e-17 -3.770313e-17 -3.913477e-17
```

```
c(cov(our_canonica_x_1, our_canonica_x_2),  
  cov(our_canonica_y_1, our_canonica_y_2),  
  cov(our_canonica_x_1, our_canonica_y_2))
```

```
## [1] -2.322361e-16 -4.181443e-15 2.379331e-16
```

Implementação

Mas também $\mathbb{V}(\eta) = \mathbf{I}$ e $\mathbb{V}(\phi) = \mathbf{I}$

```
c(var(canonica_x_1), var(canonica_x_2),  
  var(canonica_y_1), var(canonica_y_2))
```

```
## [1] 0.04545455 0.04545455 0.04545455 0.04545455
```

```
c(var(our_canonica_x_1), var(our_canonica_x_2),  
  var(our_canonica_y_1), var(our_canonica_y_2))
```

```
## [1] 1 1 1 1
```

Implementação

```
cor(canonica_x_1, canonica_y_1)
```

```
##           [,1]
```

```
## [1,] 0.9791972
```

```
cor(our_canonica_x_1, our_canonica_y_1)
```

```
##           [,1]
```

```
## [1,] 0.9791972
```

Implementação

```
head(cbind(canonica_x_1, our_canonica_x_1))
```

```
##           [,1]      [,2]
## [1,]  0.02295986 -0.1076913
## [2,]  0.11736684 -0.5504993
## [3,] -0.28396106  1.3318954
## [4,]  0.14375179 -0.6742557
## [5,] -0.31610894  1.4826824
## [6,] -0.24098151  1.1303035
```


Implementação

```
head(cbind(canonica_y_1, our_canonica_y_1))
```

```
##           [,1]      [,2]
## [1,] -0.3990281  1.8716078
## [2,] -0.1694885  0.7949714
## [3,] -0.7107743  3.3338271
## [4,] -0.2402225  1.1267432
## [5,] -0.7852577  3.6831849
## [6,] -0.6355132  2.9808213
```

Implementação

Existe outra implementação disponível no pacote CCA que coincide com a nossa.

```
library(CCA)
cca_results <- cc(X, Y)
cca_results$cor
```

```
## [1] 0.9791972 0.8851224
```

```
cor_canonica_na_mao[[3]]
```

```
## [1] 0.9791972 0.8851224
```

Implementação

```
# Autovetores
```

```
round(cca_results$xcoef, 3)
```

```
##           [,1]  [,2]
```

```
## price -0.333 -1.602
```

```
## value  0.587 -1.686
```

```
round(cor_canonica_na_mao[[1]], 3)
```

```
##           [,1]  [,2]
```

```
## [1,] -0.333  1.602
```

```
## [2,]  0.587  1.686
```

Implementação

```
# Autovetores  
round(cca_results$ycoef, 3)
```

```
##           [,1]  [,2]  
## economy -0.433 -0.568  
## service  0.191 -0.544  
## design   0.005  0.012  
## sporty   0.458  0.096  
## safety   0.223  0.014  
## handling 0.376 -0.915
```

Implementação

```
# Autovetores
```

```
round(cor_canonica_na_mao[[2]], 3)
```

```
##           [,1]    [,2]  
## [1,] -0.433    0.568  
## [2,]  0.191    0.544  
## [3,]  0.005   -0.012  
## [4,]  0.458   -0.096  
## [5,]  0.223   -0.014  
## [6,]  0.376    0.915
```

Correlação canônica para variáveis qualitativas

Correlação canônica para variáveis qualitativas

	Biscoe	Dream	Torgersen
Adelie	44	56	52
Chinstrap	0	68	0
Gentoo	124	0	0

A Tabela de contingência acima apresenta informação dos dados mas não é uma matriz de dados (pois cada linha não representa observações e cada coluna não representa variáveis).

Correlação canônica para variáveis qualitativas

	Biscoe	Dream	Torgersen
Adelie	44	56	52
Chinstrap	0	68	0
Gentoo	124	0	0

A Tabela de contigência acima apresenta informação dos dados mas não é uma matriz de dados (pois cada linha não representa observações e cada coluna não representa variáveis).

Contudo, podemos construir matrizes de dados cujas linhas serão zeros e uns dependendo se a observação apresenta ou não alguma das categorias das variáveis.

Correlação canônica para variáveis qualitativas

Assim, para $k = 1, \dots, n$, $i = 1, \dots, p$ e $j = 1, \dots, q$

$$x_{ki} = \begin{cases} 1 & \text{se a } k\text{-ésima observação pertence à categoria } i \text{ das linhas.} \\ 0 & \text{c.c} \end{cases}$$

$$y_{kj} = \begin{cases} 1 & \text{se a } k\text{-ésima observação pertence à categoria } j \text{ das colunas.} \\ 0 & \text{c.c} \end{cases}$$

Correlação canônica para variáveis qualitativas

Assim, para $k = 1, \dots, n$, $i = 1, \dots, p$ e $j = 1, \dots, q$

$$x_{ki} = \begin{cases} 1 & \text{se a } k\text{-ésima observação pertence à categoria } i \text{ das linhas.} \\ 0 & \text{c.c} \end{cases}$$

$$y_{kj} = \begin{cases} 1 & \text{se a } k\text{-ésima observação pertence à categoria } j \text{ das colunas.} \\ 0 & \text{c.c} \end{cases}$$

Aplicaremos análise de correlação canônica nesse novo conjunto de dados.

Correlação canônica para variáveis mixtas

Correlação canônica para variáveis mixtas

- As ideias apresentadas anteriormente, podem também ser aplicadas quando trabalhamos com variáveis mixtas (qualitativas e quantitativas).

Correlação canônica para variáveis mixtas

- As ideias apresentadas anteriormente, podem também ser aplicadas quando trabalhamos com variáveis mixtas (qualitativas e quantitativas).
- Cada variável quantitativa será representada por ela mesma e cada variável qualitativa será representadas por, digamos, p variáveis com zeros e uns (representando as p categorias da variável).

Correlação canônica para variáveis mixtas

- As ideias apresentadas anteriormente, podem também ser aplicadas quando trabalhamos com variáveis mixtas (qualitativas e quantitativas).
- Cada variável quantitativa será representada por ela mesma e cada variável qualitativa será representadas por, digamos, p variáveis com zeros e uns (representando as p categorias da variável).
- Por último, aplicamos análise de correlação canônica nesse novo conjunto de dados.

Exemplos

Exemplos

Veremos alguns exemplos, bem como detalhes referentes à interpretação direto no notebook.

Referências

Referências

- Härdle, W. K., & Simar, L. (2019). Applied Multivariate Statistical Analysis. Fifth Edition. Springer Nature. Capítulo 16.
- Mardia, K. V., Kent, J. T., & Bibby, J. M. (1979). Multivariate Analysis. Academic Press. Capítulo 10.
- Peña, D. (2002). Análisis de Datos Multivariantes. Mc Graw Hill. Capítulo 16.