Assignment 4 - Canonical Correlation Analysis

GROUP 03
Agustin Valencia - ***aguva779***
Bayu Brahmantio - ***baybr878***
Joris van Doorn - ***jorva845***
Marcos F Mourao - ***marfr825***

13 December 2019

Canonical correlation analysis by utilizing suit able software

Look at the data described in Exercise 10.16 of Johnson, Wichern. You may find it in the file P10-16.DAT. The data for 46 patients are summarized in a covariance matrix, which will be analyzed in R. Read through the description of the different R packages and functions so you may chose the must suitable one for the analysis. Supplement with own code where necessary.

The given matrix is the following:

V1	V2	V3	V4	V5
1106.000	396.700	108.400	0.787	26.230
396.700	2382.000	1143.000	-0.214	-23.960
108.400	1143.000	2136.000	2.189	-20.840
0.787	-0.214	2.189	0.016	0.216
26.230	-23.960	-20.840	0.216	70.560

Thus, separating the variance-covariance matrix it is obtained that

$$\Sigma_{11} =$$

V1	V2	V3
1106.0	396.7	108.4
396.7	2382.0	1143.0
108.4	1143.0	2136.0

$$\Sigma_{22} =$$

	V4	V5
4	0.016	0.216
5	0.216	70.560

$$\Sigma_{21} =$$

	V1	V2	V3
4	0.787	-0.214	2.189
5	26.230	-23.960	-20.840

$$\Sigma_{12} =$$

V4	V5
0.787	26.23
-0.214	-23.96
2.189	-20.84

It can be computed that

$$A = S_{11}^{-1/2} S_{12} S_{22}^{-1} S_{21} S_{11}^{-1/2} =$$

45000.99	25864.44	141869.5
25864.44	62898.07	182340.5
141869.53	182340.55	658795.9

Its eigen-decomposition is given by

Eigenvalues $\overrightarrow{\alpha}$:

x
739837.84
26857.14
0.00

Eigenvectors $\overrightarrow{\alpha}$:

-0.2024159	0.7395280	-0.6419705
-0.2619092	-0.6725418	-0.6921640
-0.9436267	0.0280330	0.3298224

Also,

$$B = S_{22}^{-1/2} S_{21} S_{11}^{-1} S_{12} S_{22}^{-1/2} =$$

0.0009011	0.2611067
0.2611067	81.1397279

Its eigen-decomposition is given by

Eigenvalues $\vec{\beta}$:

x
81.1405681
0.0000609

Eigenvectors $\vec{b}$:

0.0032180	-0.9999948
0.9999948	0.0032180

$$\rho_1 = \sqrt{\vec{\alpha}}$$

```
ro1 <- sqrt(aEigVals)
kable(ro1)
```

x
860.1382668
163.8814822
0.0000092

$$\rho_2 = \sqrt{\vec{\beta}}$$

```
ro2 <- sqrt(bEigVals)
kable(ro2)
```

x
9.0078060
0.0078016

a) Test at the 5% level if there is any association between the groups of variables.

```
alpha <- 0.05
#critical value
crit <- qchisq(p = (1-alpha), df = p*q)
#test statistic
```

b) How many pairs of canonical variates are significant?

c) Interpret the "significant" squared canonical correlations. Tip: Read section "Canonical Correlations as Generalizations of Other Correlation Coefficients".

d) Interpret the canonical variates by using the coefficients and suitable correlations.

e) Are the "significant" canonical variates good summary measures of the respective data sets? Tip: Read section "Proportions of Explained Sample Variance".

f) Give your opinion on the success of this canonical correlation analysis.

Appendix A - Code

```
RNGversion('3.5.1')
knitr::opts_chunk$set(echo = TRUE)
library(expm)
library(knitr)
data <- read.table("./Data/P10-16.DAT")
#number of observations (patients)
n <- 46
#number of primary variables
p <- 3
#number of secondary variables
q <- 2
kable(data)
#separating the variance-covariance matrix
sigma11 <- as.matrix(data[1:3,1:3])
sigma22 <- as.matrix(data[4:5, 4:5])
sigma21 <- as.matrix(data[4:5, 1:3])
sigma12 <- as.matrix(data[1:3, 4:5])

A <- sqrtm(sigma11) %*% sigma12 %*% solve(sigma22) %*% sigma21 %*% sqrtm(sigma11)
B <- sqrtm(sigma22) %*% sigma21 %*% solve(sigma11) %*% sigma12 %*% sqrtm(sigma22)

kable(sigma11)
kable(sigma22)
kable(sigma21)
kable(sigma12)
kable(A)
aEigenDecom <- eigen(A)
aEigVect <- aEigenDecom$vectors
aEigVals <- aEigenDecom$values
kable(aEigVals)
kable(aEigVect)
kable(B)
bEigenDecom <- eigen(B)
bEigVect <- bEigenDecom$vectors
bEigVals <- bEigenDecom$values
kable(bEigVals)
kable(bEigVect)
ro1 <- sqrt(aEigVals)
kable(ro1)
ro2 <- sqrt(bEigVals)
kable(ro2)
alpha <- 0.05
#critical value
crit <- qchisq(p = (1-alpha), df = p*q)
#test statistic
```