

# Multivariate Statistical Methods - Lab 04

Maximilian Pfundstein (maxpf364), Hector Plata (hecpl268), Aashana Nijhawan(aasni448),  
Lakshidaa Saigiridharan (laksa656)

2019-12-10

## Contents

<b>1</b>	<b>Canonical correlation analysis by utilizing suitable software</b>	<b>1</b>
1.1	Association Between Groups . . . . .	1
1.2	Number of Canonical Significant Variables . . . . .	2
1.3	Interpretation of the Significant Squared Canonical Correlations . . . . .	2
1.4	Interpretation of Canonical Variates . . . . .	2
1.5	Suitability of the Canonical Variates as a Summary Measure . . . . .	3
1.6	Opinion on the Success of the canonical analysis. . . . .	3

## 1 Canonical correlation analysis by utilizing suitable 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 choose the most suitable one for the analysis. Supplement with own code where necessary.

```
data = read.table("P10-16.DAT")
head(data)
```

```
##          V1          V2          V3          V4          V5
## 1 1106.000  396.700  108.400  0.787  26.230
## 2  396.700 2382.000 1143.000 -0.214 -23.960
## 3  108.400 1143.000 2136.000  2.189 -20.840
## 4   0.787  -0.214   2.189  0.016  0.216
## 5  26.230 -23.960 -20.840  0.216  70.560
```

### 1.1 Association Between Groups

**Task:** Test at the 5 percent level if there is any association between the groups of variables.

```
myFun = function(Sigma, p=3, q=2, n=46, alpha=0.05) {
  S11 = as.matrix(Sigma[1:p,1:p])
  S22 = as.matrix(Sigma[(p+1):(p+q),(p+1):(p+q)])
  S = as.matrix(Sigma)

  test_statistics = n * log((det(S11) * det(S22) / det(S)))
  critical_value = qchisq(1 - alpha, df=p*q)

  return(test_statistics > critical_value)
}
```

```
myFun(data)
```

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

So we reject  $H_0$  which means that we reject:  $H_0 : \Sigma_{12} = 0$ .

## 1.2 Number of Cononical Significant Variables

**Task:** How many pairs of canonical variates are significant?

```
give_me_Rho_sq_Plx = function(M, p=2, q=2) {
  R11 = as.matrix(M[1:p,1:p])
  R12 = as.matrix(M[1:p,(p+1):(p+q)])
  R21 = as.matrix(M[(p+1):(p+q), 1:p])
  R22 = as.matrix(M[(p+1):(p+q),(p+1):(p+q)])
  res = eigen(solve(sqrtm(R11)) %*% R12 %*% solve(R22) %*% R21 %*% solve(sqrtm(R11)))
  return(res$values)
}

significant_k = function(Sigma, alpha=0.05, n=46, p=3, q=2) {

  k_max = p
  Rho_sq = give_me_Rho_sq_Plx(Sigma, p=p, q=q)

  for (k in 1:k_max) {

    test_statistics = - (n - 1 - 0.5 * (p + q + 1)) * log(prod(1 - Rho_sq[(k+1):p]))
    critical_value = qchisq(1 - alpha, df=(p-k)*(q-k))

    if (test_statistics >= critical_value)
      return(k)
  }
  return(k_max)
}
```

The amount of significant canonical variates is:

```
significant_k(data)
```

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

## 1.3 Interpretation of the Significant Squared Canonical Correlations

**Task:** Interpret the “significant” squared canonical correlations.

**Tip:** Read section “Canonical Correlations as Generalizations of Other Correlation Coefficients”.

**Answer:** Because of its multiple correlation coefficient interpretation, the  $k$ th squared canonical correlation  $\rho_k^{*2}$  is the proportion of the variance of canonical variate  $U_k$  “explained” by the set  $\mathbf{X}^{(2)}$ . It is also the proportion of the variance of canonical variate  $V_k$  “explained” by the set  $\mathbf{X}^{(1)}$ . Therefore,  $\rho_k^{*2}$  is often called the *shared variance* between the two sets  $\mathbf{X}^{(2)}$  and  $\mathbf{X}^{(1)}$ . The largest value,  $\rho_k^{*2}$ , is sometimes regarded as a measure of set “overlap”.

## 1.4 Interpretation of Canonial Variates

**Task:** Interpret the canonical variates by using the coefficients and suitable correlations.

## 1.5 Suitability of the Canonical Variates as a Summary Measure

**Task:** Are the “significant” canonical variates good summary measures of the respective data sets?

**Tip:** Read section “Proportions of Explained Sample Variance”.

## 1.6 Opinion on the Success of the canonical analysis.

**Task:** Give your opinion on the success of this canonical correlation analysis.