Lab4

Naveen (navga709), Sridhar(sriad858), Juan(juado206), Samia(sambu064)

December 13, 2019

Contents

Assignment 1

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

```
data_1 <- read.table("P10-16.DAT")
data_1
```

```
        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
```

```
p = 3
q = 2
n = 46
alpha = 0.05

#Split the covarience matrix
S11 = as.matrix(data_1[1:p, 1:p])
S22 = as.matrix(data_1[(p+1):(p+q), (p+1):(p+q)])
S12 = as.matrix(data_1[1:p, (p+1):(p+q)])
S21 = as.matrix(data_1[(p+1):(p+q), 1:p])
S = as.matrix(data_1)

test_stat = (n - 1 - 0.5*(p + q + 1)) * log((det(S11) * det(S22))/det(S))
critical_value = qchisq(alpha, df=p*q)
cat("\nTest Stat : ", test_stat,
    ",\tCritical Value : ", critical_value,
    ",\tResult : ", test_stat > critical_value, "\n")
```

```
Test Stat :  13.74948 , Critical Value :  1.635383 ,    Result :  TRUE
```

Using the formulas from the book we perfoemed the test for association between the groups of variables at 5% level of significance. The result was that the two groups have a strong association and hence we are successful at rejecting the null hypothesis(zero association between the groups).

B - How many pairs of canonical variates are significant?

```
#calculte the eigen values and eigen vectors
e_mat = solve(sqrt(S11)) %*% S12 %*% solve(S22) %*% S21 %*% solve(sqrt(S11))
e = eigen(e_mat)

f_mat = solve(sqrt(S22)) %*% S21 %*% solve(S11) %*% S12 %*% solve(sqrt(S22))
f = eigen(f_mat)

sqCanonicalCorr = e$values

for(k in 1:q){
  temp = c("\nFirst", "\nSecond")
  test_stat = ( - log(prod(1 - sqCanonicalCorr[(k+1):3])) )*(n - 1 - 0.5 * (p+q+1))
  critical_value = qchisq(alpha, df=(p-k)*(q-k))
  cat(temp[k], "Canonical Correlation : ", test_stat,
      ",\tCritical Value : ", critical_value,
      ",\tResult : ", test_stat > critical_value, "\n")
}
```

```
First Canonical Correlation :  1.176238 ,   Critical Value :  0.1025866 ,   Result :  TRUE

Second Canonical Correlation :  4.662937e-15 ,  Critical Value :  0 ,   Result :  TRUE
```

There are two significant canonical variates. This is maximum possible number of variates for this data as the smaller group has just two members. Both the possible canonical variates are significant.

C - Interpret the significant squared canonical correlations.

```
sqCanonicalCorr[1:q]
```

```
[1] 1.55627405 0.02761715
```

The first two eigen values of the matrix

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

are the significant squared canonical correlations.

D - Interpret the canonical variates by using the coeffcients and suitable correlations.

```
#calculate the linear constants
a1 = t(t(as.matrix(e$vectors[1,])) %*% solve(sqrt(S11)))
a2 = t(t(as.matrix(e$vectors[2,])) %*% solve(sqrt(S11)))

b1 = t(t(as.matrix(f$vectors[1,])) %*% solve(sqrt(S22)))
b2 = t(t(as.matrix(f$vectors[2,])) %*% solve(sqrt(S22)))
```

$$a_1 = e_1' * S_{11}^{-1/2}$$
$$U_1 = a_1 * X^1$$

```
t(a1)
```

```
              1           2          3
[1,] 0.03190054 -0.06528672 0.05237954
```

$$a_2 = e_2' * S_{11}^{-1/2}$$
$$U_2 = a_2 * X^1$$

```
t(a2)
```

```
               1           2          3
[1,] -0.02687517 0.002375188 0.02001867
```

$$b_1 = e_1' * S_{22}^{-1/2}$$
$$V_1 = b_1 * X^2$$

```
t(b1)
```

```
            4         5
[1,] -9.72274 0.5197661
```

$$b_2 = e_2' * S_{22}^{-1/2}$$
$$V_2 = b_2 * X^2$$

```
t(b2)
```

```
            4          5
[1,] 2.057672 -0.2314993
```

E - Are the significant canonical variates good summary measures of the respective data sets?

Yes, we think that the canonical variates are a good summary measure of the respective dataset. As we can see from the dataset of covariance matrix, it is not calculated from a scaled dataset. This is the reason some of the variables have large variance and some have very small. The canonical covariates assigned to the variables with large variance are vary small and the variables with small variance is large. So, we think that the canonical covariates are a good summary measure for this dataset.

F - Give your opinion on the success of this canonical correlation analysis.

In canonical correlation analysis we measure the realtionship between the variables. CCA is usefull and gives you successful results in some conditions. If data doesnt satify the conditions then the results might not be satisfactory, not successfull analysis as well. For instance, canonical correlation will be greatly effected by the outliers and the variables should not be completly redundant. In our case, results were satisfactory and seems a successfull analysis of canonical correlation.