

Multivariate Statistical Inference HW# 5

Steven Francis

April 15, 2018

Question 1 – Part a

```
corr1.hw5 <- array(data = c(1.0000, 0.6328, 0.2412, 0.0586, 0.6328, 1.000,
                             -0.0553, 0.0655, 0.2412, -0.0553, 1.000, 0.4248,
                             0.0586, 0.0655, 0.4248, 1.000), dim = c(4,4),
dimnames = list(c("x1","x2", "y1", "y2"),c("x1", "x2", "y1", "y2")))
```

```
corr1.hw5
```

```
##          x1          x2          y1          y2
## x1 1.0000  0.6328  0.2412  0.0586
## x2 0.6328  1.0000 -0.0553  0.0655
## y1 0.2412 -0.0553  1.0000  0.4248
## y2 0.0586  0.0655  0.4248  1.0000
```

```
R1.xx <- corr1.hw5[1:2,1:2]
R1.xy <- corr1.hw5[1:2,3:4]
R1.yx <- corr1.hw5[3:4,1:2]
R1.yy <- corr1.hw5[3:4,3:4]
```

```
sp <- eigen(R1.xx)
```

```
R1.xx
```

```
##          x1          x2
## x1 1.0000  0.6328
## x2 0.6328  1.0000
```

```
sp$eigenvectors %*% diag(sp$values) %*% t(sp$eigenvectors)
```

```
##          [,1]      [,2]
## [1,] 1.0000  0.6328
## [2,] 0.6328  1.0000
```

```
R1xx.invsqrt <- sp$eigenvectors %*% diag(1/sqrt(sp$values)) %*% t(sp$eigenvectors)
```

```
sp <- eigen(R1.yy)
```

```
R1.yy
```

```
##          y1          y2
## y1 1.0000  0.4248
## y2 0.4248  1.0000
```

```
sp$eigenvectors %*% diag(sp$values) %*% t(sp$eigenvectors)
```

```
##          [,1]      [,2]
## [1,] 1.0000  0.4248
## [2,] 0.4248  1.0000
```

```
R1yy.invsqrt <- sp$variables %*% diag(1/sqrt(sp$values)) %*% t(sp$variables)
```

```
K1.hat <- R1xx.invsqrt %*% R1.xy %*% R1yy.invsqrt
```

```
#Calculating the A.hat and B.hat variables by performing SVD on K.hat
```

```
foo1 <- svd(K1.hat)
```

```
G <- foo1$u
```

```
L <- diag(foo1$d)
```

```
D <- foo1$v
```

```
#Sample canonical correlation coefficients
```

```
diag(L)
```

```
## [1] 0.39450592 0.06884787
```

```
#Sample canonical variables
```

```
A1.hat <- t(G) %*% R1xx.invsqrt
```

```
B1.hat <- t(D) %*% R1yy.invsqrt
```

```
A1.hat
```

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

```
## [1,] -1.2568447 1.0253167
```

```
## [2,] 0.2970177 0.7852413
```

```
B1.hat
```

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

```
## [1,] -1.10447218 0.4527216
```

```
## [2,] -0.01818009 1.0075875
```

Part b

```
n1 <- 140
```

```
q1 <- 2
```

```
p1 <- 2
```

```
r1 <- svd(K1.hat)$d
```

```
#Calculating Lambda0
```

```
Lambda1.0 <- prod(1 - r1^2)
```

```
Lambda1.0
```

```
## [1] 0.8403628
```

```
#Calculating the test statistic
```

```
-(n1 - (p1+q1+3)/2)*log(Lambda1.0)
```

```
## [1] 23.7403
```

```
#Calculating p-value for a test of H0: Sigma.xy = 0
```

```
1 - pchisq(23.7403, df=p1*q1)
```

```
## [1] 9.004155e-05
```

I would reject the notion that the Null Hypothesis of $\Sigma_{xy} = 0$ since the p-value is very close to zero.

Part c

```
#Since I would reject the Null Hypothesis from the previous part
k <- min(p1,q1)
k

## [1] 2

#Calculating Lambda1
Lambda1.1 <- prod(1 - r1[2:k]^2)
Lambda1.1

## [1] 0.99526

#Calculating the test statistic
-(n1 - (p1+q1+3)/2)*log(Lambda1.1)

## [1] 0.6485522

s1 <- 1

#Calculating p-value for a test of H0: rho2 = 0
1 - pchisq(0.6485522, df=(p1-s1)*(q1-s1))

## [1] 0.4206308
```

Part d

It seems as if there is some correlation between reading ability and arithmetic ability as indicated by the first canonical correlation's p-value. However, with a p-value of 0.421, it seems that the second canonical correlation is not significant at all. Therefore we would fail to reject new the null hypothesis.

Problem 2 – Part a

```
corr2.hw5 <- array(data = c(1.00, 0.37, 0.21, 0.26, 0.33, 0.37, 1.00,
                           0.35, 0.67, 0.59, 0.21, 0.35, 1.00, 0.34,
                           0.34, 0.26, 0.67, 0.34, 1.00, 0.80, 0.33,
                           0.59, 0.34, 0.80, 1.00), dim = c(5,5),
dimnames = list(c("x1","x2","x3","y1","y2"),
                 c("x1","x2","x3","y1","y2")))

corr2.hw5

##      x1  x2  x3  y1  y2
## x1 1.00 0.37 0.21 0.26 0.33
## x2 0.37 1.00 0.35 0.67 0.59
## x3 0.21 0.35 1.00 0.34 0.34
## y1 0.26 0.67 0.34 1.00 0.80
## y2 0.33 0.59 0.34 0.80 1.00

R2.xx <- corr2.hw5[1:3,1:3]
R2.xy <- corr2.hw5[1:3,4:5]
R2.yx <- corr2.hw5[4:5,1:3]
```

```

R2.yy <- corr2.hw5[4:5,4:5]

sp <- eigen(R2.xx)

R2.xx

##      x1  x2  x3
## x1 1.00 0.37 0.21
## x2 0.37 1.00 0.35
## x3 0.21 0.35 1.00

sp$eigenvectors %*% diag(sp$values) %*% t(sp$eigenvectors)

##      [,1] [,2] [,3]
## [1,] 1.00 0.37 0.21
## [2,] 0.37 1.00 0.35
## [3,] 0.21 0.35 1.00

R2xx.invsqrt <- sp$eigenvectors %*% diag(1/sqrt(sp$values)) %*% t(sp$eigenvectors)

sp <- eigen(R2.yy)

R2.yy

##      y1  y2
## y1 1.0 0.8
## y2 0.8 1.0

sp$eigenvectors %*% diag(sp$values) %*% t(sp$eigenvectors)

##      [,1] [,2]
## [1,] 1.0 0.8
## [2,] 0.8 1.0

R2yy.invsqrt <- sp$eigenvectors %*% diag(1/sqrt(sp$values)) %*% t(sp$eigenvectors)

K2.hat <- R2xx.invsqrt %*% R2.xy %*% R2yy.invsqrt

svd(K2.hat)

## $d
## [1] 0.6879479 0.1868654
##
## $u
##      [,1]      [,2]
## [1,] -0.2287748 0.9000953
## [2,] -0.9105151 -0.3326102
## [3,] -0.3444189 0.2814231
##
## $v
##      [,1]      [,2]
## [1,] -0.8094243 -0.5872242
## [2,] -0.5872242 0.8094243

foo2 <- svd(K2.hat)

G2 <- foo2$u
L2 <- diag(foo2$d)

```

```

D2 <- foo2$v

#Sample canonical correlation coefficients
diag(L2)

## [1] 0.6879479 0.1868654

n2 <- 70
q2 <- 3
p2 <- 2

r2 <- svd(K2.hat)$d

#Calculating Lambda0
Lambda2.0 <- prod(1 - r2^2)
Lambda2.0

## [1] 0.508335

#Calculating the test statistic
-(n2 - (p2+q2+3)/2)*log(Lambda2.0)

## [1] 44.65657

#Calculating p-value for a test of H0: Sigma.xy = 0
1 - pchisq(44.65657, df=p2*q2)

## [1] 5.476234e-08

```

It appears that since the p-value is very close to 0, there is definitely some correlation between the demographic backgrounds and consumption practices of this data set.

Part b

```

#Since I would reject the Null Hypothesis from the previous part
k <- min(p2,q2)
k

## [1] 2

#Calculating Lambda1
Lambda2.1 <- prod(1 - r2[2:k]^2)
Lambda2.1

## [1] 0.9650813

#Calculating the test statistic
-(n2 - (p2+q2+3)/2)*log(Lambda2.1)

## [1] 2.345833

s2 <- 1

#Calculating p-value for a test of H0: rho2 = 0
1 - pchisq(2.345833, df=(p2-s2)*(q2-s2))

## [1] 0.3094631

```

With a p-value of 0.309, it seems that the second canonical correlation is not significant on its own. Therefore we would fail to reject new the null hypothesis.

Part c

```
#Sample canonical variables
A2.hat <- t(G2) %*% R2xx.invsqrt
B2.hat <- t(D2) %*% R2yy.invsqrt

A2.hat

##           [,1]      [,2]      [,3]
## [1,] -0.04912716 -0.8975114 -0.1900411
## [2,]  1.00029422 -0.5836950  0.2955723

B2.hat

##           [,1]      [,2]
## [1,] -0.7689274 -0.2720729
## [2,] -1.4786914  1.6443096
```

Part d

By observing the first row of B2.hat, we can see that (since both numbers have the same sign) that this is an index of the preference of how often the people in this data set want to go out to eat. The second row of B2.hat shows a clear preference to one out of the house activity compared to the other. As we look at the x parameters represented by A2.hat's first row, it is apparent that annual family income holds the most weight in the information which seems to be the driving force behind why the people of this dataset prefer going to dinner rather than going to the movies. This makes sense because people need to be of a certain social status to afford to go out regularly.

By observing the second row of A2.hat, it seems to indicate that the age of the head of the household has a profound impact on how often people go out to the movies. This makes sense because young people are the ones that are usually willing to take a trip to the movie theater.

Part e

Yes, it appears that demographic values have something to say about the consumption variables and the consumption variables have something to say about the demographic variables as they give insight into the other grouping of variables as proven by parts c and d.