DA410_Project5_MattGraham

Finding canonical correlationss

```
library(nnspat)  # used for dist2full()
library("dplyr")  # used to select numeric datatypes
library("ggplot2")
library(reshape)  # used for melting matricies
library(klaR)
library(ggvis)
library(class)
library(gmodels)
library(MASS)
library(readxl)
library(psych)
```

11.10 Get data

```
glucose <- read.table("C:/mattgraham93.github.io/school/22_3_DA410/data/T3_9_GLUCOSE.DAT", he
ader=FALSE)[1:34,]

colnames(glucose) <- c('b_time1', 'b_time2', 'b_time3', 'a_time1', 'a_time2', 'a_time3')
# b = before, a = after

fasting <- glucose[, 1:3]
post.cons <- glucose[, 4:6]

glucose
```

	b_time1 <int>	b_time2 <int>	b_time3 <int>	a_time1 <int>	a_time2 <int>	a_time3 <int>
1	60	69	62	97	69	98
2	56	53	84	103	78	107
3	80	69	76	66	99	130
4	55	80	90	80	85	114
5	62	75	68	116	130	91
6	74	64	70	109	101	103
7	64	71	66	77	102	130
8	73	70	64	115	110	109
9	68	67	75	76	85	119
10	69	82	74	72	133	127

1-10 of 34 rows Previous **1** 2 3 4 Next

Find means

```
x.bar <- colMeans(fasting)
y.bar <- colMeans(post.cons)

cbind.data.frame(x.bar, y.bar)
```

	x.bar <dbl>	y.bar <dbl>
b_time1	71.08824	108.7353
b_time2	72.00000	103.0588
b_time3	74.20588	110.0882

3 rows

a - Find canonical correlations

```
cancor(y.bar, x.bar)
```

```
## $cor
## [1] 0.4055683
##
## $xcoef
##           [,1]
## [1,] 0.1895865
##
## $ycoef
##           [,1]
## [1,] 0.4411246
##
## $xcenter
## [1] 107.2941
##
## $ycenter
## [1] 72.43137
```

Interpretation

Overall, we can see that the centers of before and after consumption are pretty significant. There is slight positive correlation between before and after, however, we can see there is a higher correlation among our post-consumption results.

```
cancor2<-function(x,y,dec=4){
#Canonical Correlation Analysis to mimic SAS PROC CANCOR output.
#Basic formulas can be found in Chapter 10 of Mardia, Kent, and Bibby (1979).
# The approximate F statistic is exercise 3.7.6b.
    x<-as.matrix(x)
    y<-as.matrix(y)

    n<-dim(x)[1]
    q1<-dim(x)[2]
    q2<-dim(y)[2]
    q<-min(q1,q2)

    S11<-cov(x)
    S12<-cov(x,y)
    S21<-t(S12)
    S22<-cov(y)

    E1<-eigen(solve(S11)%*%S12%*%solve(S22)%*%S21)
    E2<-eigen(solve(S22)%*%S21%*%solve(S11)%*%S12)

    rsquared<-as.double(E1$values[1:q])

    LR<-NULL;pp<-NULL;qq<-NULL;tt<-NULL

    for (i in 1:q){
        LR<-c(LR,prod(1-rsquared[i:q]))
        pp<-c(pp,q1-i+1)
        qq<-c(qq,q2-i+1)
        tt<-c(tt,n-1-i+1)}

    m<-tt-0.5*(pp+qq+1);lambda<-(1/4)*(pp*qq-2);s<-sqrt((pp^2*qq^2-4)/(pp^2+qq^2-5))
    F<-((m*s-2*lambda)/(pp*qq))*((1-LR^(1/s))/LR^(1/s))
    df1<-pp*qq;df2<-(m*s-2*lambda)
    pval<-1-pf(F,df1,df2)
    outmat<-round(cbind(sqrt(rsquared),rsquared,LR,F,df1,df2,pval),dec)

    colnames(outmat) <- list("R","RSquared","LR","ApproxF","NumDF","DenDF","pvalue")
    rownames(outmat) <- as.character(1:q)
    xrels<-round(cor(x,x%*%E1$vectors)[,1:q],dec)
    colnames(xrels)<-apply(cbind(rep("U",q),as.character(1:q)),1,paste,collapse="")
    yrels<-round(cor(y,y%*%E2$vectors)[,1:q],dec)
    colnames(yrels)<- apply(cbind(rep("V",q),as.character(1:q)),1,paste,collapse="")
    list(Summary=outmat,
         a.Coefficients=E1$vectors,
         b.Coefficients=E2$vectors,
         XUCorrelations=xrels,YVCorrelations=yrels
     )
   }
## END FUNCTION
##################################################
```

b - Find standard coefficients

For canonical variables

Fasting coefficients

```
before.coefficients <- cancor2(fasting, post.cons)$a.Coefficients
after.coefficients <- cancor2(fasting, post.cons)$b.Coefficients

diag(before.coefficients)
```

```
## [1]  0.2916227 -0.5031850  0.9541986
```

Post-consumption coefficients

```
diag(after.coefficients)
```

```
## [1] -0.3903434  0.8212557 -0.7425049
```

c - Test significance for reach canonical correlation

```
cancor2(fasting, post.cons)
```

```
## $Summary
##        R RSquared     LR ApproxF NumDF   DenDF pvalue
## 1 0.5406   0.2923 0.6357  1.5527     9 68.2952 0.1475
## 2 0.3038   0.0923 0.8982  0.7996     4 58.0000 0.5304
## 3 0.1023   0.0105 0.9895  0.3172     1 30.0000 0.5775
##
## $a.Coefficients
##            [,1]        [,2]       [,3]
## [1,]  0.2916227 -0.86244010 -0.2440668
## [2,] -0.9116674 -0.50318496  0.1730216
## [3,]  0.2895147  0.05479022  0.9541986
##
## $b.Coefficients
##            [,1]      [,2]       [,3]
## [1,] -0.3903434 0.3161124  0.6221615
## [2,]  0.7868380 0.8212557 -0.2481965
## [3,] -0.4780356 0.4749863 -0.7425049
##
## $XUCorrelations
##              U1      U2      U3
## b_time1  0.3475 -0.9342 -0.0807
## b_time2 -0.8078 -0.4981  0.3151
## b_time3  0.2329 -0.2107  0.9494
##
## $YVCorrelations
##              V1     V2      V3
## a_time1 -0.3430 0.7441  0.5733
## a_time2  0.5108 0.8590  0.0354
## a_time3 -0.5682 0.5748 -0.5888
```

```
# It produces two other pieces of information:  An F-test for the
# significance of each canonical correlation, and the correlations between
# the original variables and the corresponding canonical variates.
```

Interpretation

Given all our p-values > 0.05, there is not enough evidence to conclude that there is at least one non-zero canonical correlation.