Canonical Population Analysis

Miquel Salicrú, Sergi Civit, Ferran Reverter

Packages

```r
library(car)
library(heplots)
```

Data

Reaven and Miller (1979) examined the relationship among blood chemistry measures of glucose tolerance and insulin in 145 nonobese adults.

```r
data(Diabetes)
str(Diabetes)
```

```
## 'data.frame':    145 obs. of  6 variables:
##  $ relwt : num  0.81 0.95 0.94 1.04 1 0.76 0.91 1.1 0.99 0.78 ...
##  $ glufast: int  80 97 105 90 90 86 100 85 97 97 ...
##  $ glutest: int  356 289 319 356 323 381 350 301 379 296 ...
##  $ instest: int  124 117 143 199 240 157 221 186 142 131 ...
##  $ sspg   : int  55 76 105 108 143 165 119 105 98 94 ...
##  $ group  : Factor w/ 3 levels "Normal","Chemical_Diabetic",..: 1 1 1 1 1 1 1 1 1 1 ...
```

Canonical Population Analysis

within variability

```r
sel1<-which(Diabetes$group=="Normal")
sel2<-which(Diabetes$group=="Chemical_Diabetic")
sel3<-which(Diabetes$group=="Overt_Diabetic")

S1<-cov(Diabetes[sel1,-6])
S2<-cov(Diabetes[sel2,-6])
S3<-cov(Diabetes[sel3,-6])


n1<-length(sel1)
n2<-length(sel2)
n3<-length(sel3)
n<-n1+n2+n3
g<-3

S1<-(n1-1)*S1/n1
S2<-(n2-1)*S2/n2
S3<-(n3-1)*S3/n3


W<-(n1*S1+n2*S2+n3*S3) # within
Sp<-W/(n-g) # pooled
```

between variability

```
m<-apply(Diabetes[,-6],2,mean)
m
```

```
##       relwt     glufast     glutest     instest        sspg
##   0.9773103 121.9862069 543.6137931 186.1172414 184.2068966
```

```
m1<-apply(Diabetes[sel1,-6],2,mean)
m2<-apply(Diabetes[sel2,-6],2,mean)
m3<-apply(Diabetes[sel3,-6],2,mean)

m1
```

```
##       relwt     glufast     glutest     instest        sspg
##   0.9372368   91.1842105 349.9736842 172.6447368 114.0000000
```

```
m2
```

```
##      relwt    glufast    glutest   instest       sspg
##   1.055833  99.305556 493.944444 288.000000 208.972222
```

```
m3
```

```
##       relwt     glufast      glutest     instest        sspg
##   0.9839394 217.6666667 1043.7575758 106.0000000 318.8787879
```

```
v1<-as.matrix(m1-m)
v2<-as.matrix(m2-m)
v3<-as.matrix(m3-m)

B<-n1*v1%*%t(v1)+n2*v2%*%t(v2)+n3*v3%*%t(v3)
B
```

```
##                  relwt       glufast       glutest        instest          sspg
## relwt        0.3454684      50.62677      558.7517       311.5103       313.2891
## glufast     50.6267698  392731.57914   2073038.9147   -304615.7392   569350.4416
## glutest    558.7517238 2073038.91472 11193292.4756  -1306220.7240  3211652.6115
## instest    311.5103294 -304615.73920  -1306220.7240    599297.5990  -193335.5172
## sspg       313.2890987  569350.44157   3211652.6115   -193335.5172   995189.3057
```

solving

```
A<-solve(Sp)%*%B
vecs<-eigen(A)$vectors

Y<-as.matrix(Diabetes[,-6])%*%vecs


ym1<-t(as.matrix(m1))%*%vecs
ym2<-t(as.matrix(m2))%*%vecs
ym3<-t(as.matrix(m3))%*%vecs

ym1
```

```
##          [,1]       [,2]       [,3]       [,4]      [,5]
## [1,] -2.25705 -0.9578447 -1.390952 -0.9486748 0.9076623
```

```
ym2
```

```
##           [,1]      [,2]      [,3]       [,4]      [,5]
## [1,] -3.790624 -1.427346 -1.390952 -0.9486748 0.9076623
```

```
ym3
```

```
##          [,1]       [,2]      [,3]       [,4]      [,5]
## [1,] -6.22582 -0.9102302 -1.390952 -0.9486748 0.9076623
```

```
ym<-rbind(ym1,ym2,ym3)
```

Outputs

```
Diabetes$g<-0
Diabetes[sel1,"g"]<-1
Diabetes[sel2,"g"]<-2
Diabetes[sel3,"g"]<-3

plot(Y[,1],Y[,2],type="n",xlab="1st canonical dim",ylab="2nd canonical dim")
points(Y[,1],Y[,2],col=Diabetes$g)
points(ym[,1],ym[,2],col="blue",pch=19)
```