

## 4.6. Overview of multivariate analysis

**Example:**

	Observations
<b>Population 1</b>	(0.37, 1.06, -0.36), (1.06, 1.65, 2.06), (0.81, 0.92, 0.80) (-0.98, 1.75, 1.08), (0.92, -0.85, 0.73)
<b>Population 2</b>	(1.13, 2.53, 1.42), (0.71, 3.19, 2.01), (2.28, 1.00, 2.16) (1.54, 1.99, 0.21), (2.11, 1.45, 1.46)
<b>Population 3</b>	(1.63, 2.62, 1.15), (2.66, 4.23, 0.23), (1.55, 1.36, 2.05) (-0.82, 1.22, 0.75), (2.15, 3.10, 1.00)

Please do the following:

- (a) For all data in the three populations, please produce pairwise scatter plots for the variables and find the correlation matrix for these variables.
- (b) Find the principal components by the 95% criterion and also give the screeplot.
- (c) For the data in all populations, please use Fisher's discrimination method to find  $\hat{a}_1$  and  $\hat{a}_2$ . Also, find the error rate for the 15 observations based on the two discriminants.
- (d) Please find the error rate for the above data as using K-means method with number of clusters equal to 3.
- (e) Please use classification tree method to allocate the observations (1, 1, 1), (2, 2, 2) and (3, 3, 3). Also, use the classification tree to allocate the observations in population 2 and compute the error rate for the 5 observations.
- (f) Also, please use  $X_2$  as response and  $X_1, X_3$  as covariates. Then, for an observation with  $(X_1, X_3) = (1, 2)$ , use the regression tree method to predict the value of  $X_2$ . Also, please use  $X_1$  as response and  $X_3$  as covariates. Then, for an observation with  $X_3 = 2.2$ , use the regression tree method to predict the value of  $X_1$ .

**Example (Splus):**

```
### (a)
p1=rbind(c(0.37,1.06,-0.36),c(1.06,1.65,2.06),c(0.81,0.92,0.80),c(-0.98,1.75,1.08),
c(0.92,-0.85,0.73))
p2=rbind(c(1.13,2.53,1.42),c(0.71,3.19,2.01),c(2.28,1.00,2.16),c(1.54,1.99,0.21),
c(2.11,1.45,1.46))
p3=rbind(c(1.63,2.62,1.15),c(2.66,4.23,0.23),c(1.55,1.36,2.05),c(-0.82,1.22,0.75),
c(2.15,3.10,1.00))
x=rbind(p1,p2,p3)
```

```

com=cor(x)
list("correlation matrix ="=com)
pairs(x)

#### (b)
varx=var(x)
eigenvarx=eigen(varx)
cumvar=cumsum(eigenvarx$values)/sum(eigenvarx$values)
xprin95=eigenvarx$vectors[,1:(sum(cumvar<0.95)+1)] # the principal
# components
plot(princomp(x),style="lines")      # the screeplot

#### (c)
xmean1=apply(x[1:5],2,mean)
xmean2=apply(x[6:10],2,mean)
xmean3=apply(x[11:15],2,mean)
xmean=apply(x,2,mean)
b1=5*(xmean1-xmean)%*%t(xmean1-xmean)
b2=5*(xmean2-xmean)%*%t(xmean2-xmean)
b3=5*(xmean3-xmean)%*%t(xmean3-xmean)
B=b1+b2+b3
sum1=4*var(x[1:5,])
sum2=4*var(x[6:10,])
sum3=4*var(x[11:15,])
W=sum1+sum2+sum3
invW=solve(W)
spool=W/(5+5+5-3)
eectorsB=eigen(invW%*%B)$vectors
a1hat=eectorsB[,1]/sqrt(t(eectorsB[,1])%*%spool%*%eectorsB[,1])
a2hat=eectorsB[,2]/sqrt(t(eectorsB[,2])%*%spool%*%eectorsB[,2])
a12hat=cbind(a1hat,a2hat)
y1bar=t(a12hat)%*%xmean1
y2bar=t(a12hat)%*%xmean2
y3bar=t(a12hat)%*%xmean3
cls=rep(0,15)
correct=c(rep(1,5),rep(2,5),rep(3,5))
for(i in 1:15)
{

```

```

x0=x[i,]
yhat=t(a12hat)%*%x0
cls[i]=order(c(sum((yhat-y1bar)^2),sum((yhat-y2bar)^2),sum((yhat-
y3bar)^2)))[1]
}
errorrate2=sum(cls!=correct)/length(correct)
list(a1hat=a1hat,a2hat=a2hat,"error rate for the 15 observations"=errorrate2)

### (d)
correct=c(rep(1,5),rep(2,5),rep(3,5))
errorrate3=sum(kmeans(x,3)$cluster!=correct)/length(correct)
errorrate3 # error rate

### (e)
spe=c(rep("1",5),rep("2",5),rep("3",5))
xdata=list(spe=spe,x=x)
popu=data.frame(xdata)
auto.tree1=tree(popu)
plot(auto.tree1,type="u")
text(auto.tree1)
e=rbind(c(1,1,1),c(2,2,2),c(3,3,3))
ecls1=rep(0,3)
for(i in 1:3)
{
  x0=e[i,]
  if(x0[1]<1.095)
    ecls1[i]=1
  else
    ecls1[i]=2
}
ecls2=rep(0,5)
ecorrect=rep(2,5)
for(i in 1:5)
{
  x0=p2[i,]
  if(x0[1]<1.095)
    ecls2[i]=1
  else

```

```

ecls2[i]=2
}
errorrate4=sum(ecls2!=ecorrect)/length(ecorrect)
list("allocate (1,1,1),(2,2,2),(3,3,3)"=ecls1,"error rate for population 2"=errorrate4)

### (f)
auto.tree2=tree(x2~x1+x3,popu)
plot(auto.tree2,type="u")
text(auto.tree2)
preX2=0
if(c(1,2)[1]<1.095)
  preX2=1.277
else
  preX2=2.285
auto.tree3=tree(x1~x3,popu)
plot(auto.tree3,type="u")
text(auto.tree3)
preX1=1.4960
list("prediction value of X2"=preX2,"prediction value of X1"=preX1)

```