**Tor Thogersen**

**DA 410 -- Multivariate Analysis -- Winter 2018**

**Project 6**

**Part 1**: Use R to solve Chapter 12 Page 433, #12.11. Make sure you include the commands and outputs, as well as the interpretations of the outputs.

12.11

Carry out a principle component Analysis separately for the two species in the beetle data of table 5.5. Compare the results for the two groups use S

**# Load Data**

> BEETLES <- read.table("~/School/DA 410/multivariate\_analysis - 3rd Ed/multivariate\_analysis - 3rd Ed/T5\_5\_FBEETLES.DAT", quote="\"", comment.char="", col.names = c("Exp", "Species", "y1", "y2", "y3", "y4"))

>

> View(BEETLES)

> attach(BEETLES)

The following objects are masked \_by\_ .GlobalEnv:

y1, y2

> head(BEETLES)

Exp Species y1 y2 y3 y4

1 1 1 189 245 137 163

2 2 1 192 260 132 217

3 3 1 217 276 141 192

4 4 1 221 299 142 213

5 5 1 171 239 128 158

6 6 1 192 262 147 173

**# covariance matrix S of the data**

> S.beetles <- cov(BEETLES[,2:6])

> head(S.beetles)

Species y1 y2 y3 y4

Species 0.2564103 -3.826586 6.089069 5.08502 5.975034

y1 -3.8265857 196.887989 56.937247 -34.47976 -19.071525

y2 6.0890688 56.937247 502.708502 239.42510 245.340081

y3 5.0850202 -34.479757 239.425101 216.04453 159.451417

y4 5.9750337 -19.071525 245.340081 159.45142 341.831309

**# eigenvalues of S**

> sum(diag(S))

[1] 292.8909

> s.eigen <- eigen(S)

> s.eigen

eigen() decomposition

$values

[1] 200.375372 36.090806 34.072122 14.967285 7.385324

$vectors

[,1] [,2] [,3] [,4] [,5]

[1,] 0.4727831 -0.57631826 0.41685181 0.2285789 0.4672469

[2,] 0.3918187 -0.10826396 -0.45239805 0.6558114 -0.4472185

[3,] 0.4875471 0.09600807 0.47945493 -0.3689607 -0.6221505

[4,] 0.4677199 -0.12056819 -0.61953744 -0.5803451 0.2146494

[5,] 0.4080320 0.79522446 0.08869553 0.2114962 0.3853964

**# eigenvectors of S**

> for (s in s.eigen$values) {

+ print(s / sum(s.eigen$values))

+ }

[1] 0.6841297

[1] 0.1232227

[1] 0.1163304

[1] 0.05110191

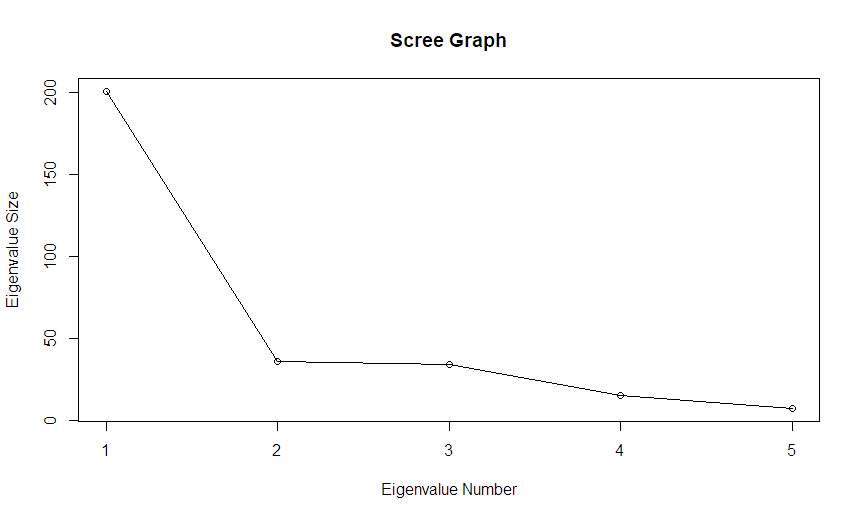
[1] 0.02521527

The first two principal components account for 80 % ( .68+ .12) of the total variance. A scree graph of the eigenvalues can be plotted to visualize the proportion of variance explained by each sub sequential eigenvalue.

**# cree graph of the eigenvalues**

> plot(s.eigen$values, xlab = 'Eigenvalue Number', ylab = 'Eigenvalue Size', main = 'Scree Graph')

> lines(s.eigen$values)



**# the S ‘coefficients’ of the principal components**

I would say the first two PC shows the largest portion of the sample and it finally levels out after 3rd PC, I would use 3 PC to develop my model.

> s.eigen$vectors

[,1] [,2] [,3] [,4] [,5]

[1,] 0.4727831 -0.57631826 0.41685181 0.2285789 0.4672469

[2,] 0.3918187 -0.10826396 -0.45239805 0.6558114 -0.4472185

[3,] 0.4875471 0.09600807 0.47945493 -0.3689607 -0.6221505

[4,] 0.4677199 -0.12056819 -0.61953744 -0.5803451 0.2146494

[5,] 0.4080320 0.79522446 0.08869553 0.2114962 0.3853964

> beetle.pc <- princomp(BEETLES[,-1],cor = T)

> beetle.pc

Call:

princomp(x = BEETLES[, -1], cor = T)

Standard deviations:

Comp.1 Comp.2 Comp.3 Comp.4 Comp.5

1.7124921 1.1269777 0.6709071 0.4684609 0.3573796

5 variables and 39 observations.

> summary(bird.pc,loadings=T)

Importance of components:

Comp.1 Comp.2 Comp.3 Comp.4 Comp.5

Standard deviation 1.9015726 0.7290433 0.62163056 0.5491498 0.4056199

Proportion of Variance 0.7231957 0.1063008 0.07728491 0.0603131 0.0329055

Cumulative Proportion 0.7231957 0.8294965 0.90678139 0.9670945 1.0000000

Loadings:

Comp.1 Comp.2 Comp.3 Comp.4 Comp.5

tot.length -0.452 0.690 0.420 -0.374

alar.length -0.462 -0.300 0.341 -0.548 0.530

beak.head.length -0.451 -0.325 -0.454 0.606 0.343

humerus.length -0.471 -0.185 -0.411 -0.388 -0.652

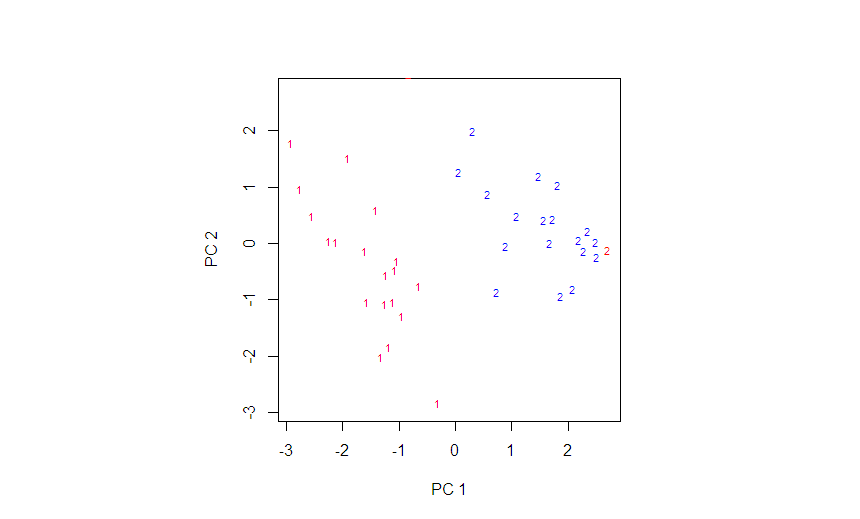
keel.stern.length -0.398 0.876 -0.178 0.192

> text(bird.pc$scores[,1], bird.pc$scores[,2], labels=Species, cex=0.7, lwd=2)

> par(pty="s")

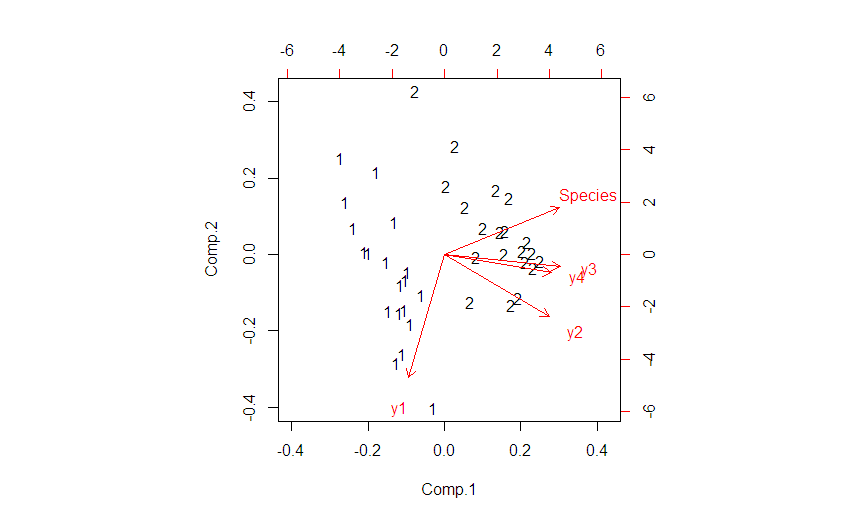
> plot(beetle.pc$scores[,1], beetle.pc$scores[,2], ylim=range(beetle.pc$scores[,1]), xlab="PC 1", ylab="PC 2", type ='n', lwd=2)

> text(beetle.pc$scores[,1], beetle.pc$scores[,2], labels=Species, cex=0.7, lwd=2, col=c(rep("red", times = 21), rep("blue", times=28)) )



**# The biplot**

biplot(beetle.pc, xlabs = Species)



Based on the arrows all the area vectors are point to the 2nd species of beetle.