Project 6

Code

JONATHAN FENG

Carry out a principal component analysis separately for the two species in the beetle data of Table 5.5. Compare the results for the two groups. Use S. [BEETLE DATA]

DATA IMPORTS

Hide

```
beet <- read.csv("C:\\Users\\Taterthot\\Desktop\\da 410\\a7\\beetles.csv", fileEncoding = 'UTF-8
-BOM')
```

```
Warning message:
R graphics engine version 14 is not supported by this version of RStudio. The Plots tab will be
disabled until a newer version of RStudio is installed.
```

Hide

```
beet
```

Experiment.Number <int>	y1 <int>	y2 <int>	y3 <int>	y4 <int>	Species <chr>
1	189	245	137	163	Haltica oleracea
2	192	260	132	217	Haltica oleracea
3	217	276	141	192	Haltica oleracea
4	221	299	142	213	Haltica oleracea
5	171	239	128	158	Haltica oleracea
6	192	262	147	173	Haltica oleracea
7	213	278	136	201	Haltica oleracea
8	192	255	128	185	Haltica oleracea
9	170	244	128	192	Haltica oleracea
10	201	276	146	186	Haltica oleracea

1-10 of 39 rows Previous **1** 2 3 4 Next

BEETLES CLEAN DATA

Hide

```
beet.o <- subset(beet, Species == "Haltica oleracea")
beet.c <- subset(beet, Species == "Haltica carduorum")
head(beet.o)
```

	Experiment.Number	y1	y2	y3	y4	Species
	<int>	<int>	<int>	<int>	<int>	<chr>
1	1	189	245	137	163	Haltica oleracea
2	2	192	260	132	217	Haltica oleracea
3	3	217	276	141	192	Haltica oleracea
4	4	221	299	142	213	Haltica oleracea
5	5	171	239	128	158	Haltica oleracea
6	6	192	262	147	173	Haltica oleracea

6 rows

Hide

```
head(beet.c)
```

	Experiment.Number	y1	y2	y3	y4	Species
	<int>	<int>	<int>	<int>	<int>	<chr>
20	1	181	305	184	209	Haltica carduorum
21	2	158	237	133	188	Haltica carduorum
22	3	184	300	166	231	Haltica carduorum
23	4	171	273	162	213	Haltica carduorum
24	5	181	297	163	224	Haltica carduorum
25	6	181	308	160	223	Haltica carduorum

6 rows

S MATRICES

Hide

```
s1 <- var(beet.o[2:5])
s2 <- var(beet.c[2:5])
s1
```

```
          y1        y2       y3        y4
y1 187.59649 176.86257 48.37135 113.58187
y2 176.86257 345.38596 75.97953 118.78070
y3  48.37135  75.97953 66.35673  16.24269
y4 113.58187 118.78070 16.24269 239.94152
```

Hide

```
s2
```

```
          y1        y2        y3        y4
y1 101.83947 128.06316  36.98947  32.59211
y2 128.06316 389.01053 165.35789  94.36842
y3  36.98947 165.35789 167.53684  66.52632
y4  32.59211  94.36842  66.52632 177.88158
```

EIGEN VECTORS FOR BOTH GROUPS

Hide

```
e1 <- eigen(s1)
e2 <- eigen(s2)
e1
```

```
eigen() decomposition
$values
[1] 561.30574 168.98584  65.27709  43.71203

$vectors
           [,1]         [,2]       [,3]       [,4]
[1,] -0.4997445  0.009204574  0.8230272  0.2698089
[2,] -0.7187015 -0.484408702 -0.4778690  0.1430301
[3,] -0.1739702 -0.220296505  0.2042647 -0.9378058
[4,] -0.4510631  0.846600812 -0.2292234 -0.1651236
```

Hide

```
e2
```

```
eigen() decomposition
$values
[1] 555.69314 145.44632  93.46372  41.66524

$vectors
           [,1]       [,2]       [,3]        [,4]
[1,] -0.2836552 -0.2007357  0.5315166 -0.77248627
[2,] -0.8068689 -0.3389760  0.1218433  0.46820095
[3,] -0.4222422  0.1359900 -0.7897513 -0.42368751
[4,] -0.3003563  0.9090144  0.2809577  0.06739234
```

VARIANCE EXPLAINED BY COMPONENT

E1

Hide

```
var <- rep(NA, 4)
for (i in 1:4) {
  var[i] <- e1$values[i] / (e1$values[1] + e1$values[2] + e1$values[3] + e1$values[4])
}
var
```

```
[1] 0.66879382 0.20134603 0.07777743 0.05208273
```

Variance of the Haltica Oleracea beetles is show as the first component explaining around 66-67% of the variation with our second component explaining around 20%.The rest of the components and variation percentages are shown above.

E2

Hide

```
var2 <- rep(NA, 4)
for (i in 1:4) {
  var2[i] <-
    e2$values[i] / (e2$values[1] + e2$values[2] + e2$values[3] + e2$values[4])
}
var2
```

```
[1] 0.6644914 0.1739230 0.1117628 0.0498228
```

Variance of the Haltica Carduorum beetles is show as the first component explaining around 66% of the variation with our second component explaining around 17%. The rest of the components and variation percentages are shown above.

PCA ANALYSIS

oleracea

Hide

```
summary(princomp(x = beet.o[2:5]))
```

```
Importance of components:
                          Comp.1     Comp.2     Comp.3     Comp.4
Standard deviation     23.0599943 12.6527406 7.86393386 6.43516871
Proportion of Variance  0.6687938  0.2013460 0.07777743 0.05208273
Cumulative Proportion   0.6687938  0.8701398 0.94791727 1.00000000
```

carduorum

Hide

```
summary(princomp(x = beet.c[2:5]))
```

```
Importance of components:
                          Comp.1     Comp.2    Comp.3    Comp.4
Standard deviation     22.9762592 11.7547439 9.4228730 6.2914206
Proportion of Variance  0.6644914  0.1739230 0.1117628 0.0498228
Cumulative Proportion   0.6644914  0.8384144 0.9501772 1.0000000
```

Of the two PCAs we ran on the two different species of beetles, the oleracea variant can account for more of the variation within the first two of the components than the other species. With the combination of our first two components we can say that for oleracea that it explains around 87% of the data versus our other one having around 3% less.

DECIDING WHICH PRINCIPLE COMPONENTS TO KEEP

Hide

```
e1$values > mean(e1$values)
```

```
[1]  TRUE FALSE FALSE FALSE
```

Hide

```
e2$values > mean(e2$values)
```

```
[1]  TRUE FALSE FALSE FALSE
```

Based on the mean values we can compare these values presented to tell us which one to keep based on this basic boolean logic of keeping it above the mean for the variance.

Due to the results and putting it into the boolean, we can only keep the first principle component of each species.