

1 Sparrow Data

The output attached gives a principal components analysis for five anatomical variates of 49 female sparrows. The body measurements, in mm, are total length (x_1), alar extent (x_2), length of beak and head (x_3), length of humerus (x_4) and length of keel of the sternum (x_5). Birds numbered 1-21 survived the period of observation while birds 22-49 did not.

1. Comment on why the principal components analysis is carried out using the correlation matrix.
2. Calculate the eigenvalues of the correlation matrix. Check these values sum to 5.
3. Give the proportion of variability explained by the first two principal components.
4. What variables are highly correlated with the first two principal components?
5. Use the correlations to find the missing values in the loadings output.
6. What can you say about the survivors given the plot of the scores for the first two principal components?

R output for Exercise 1, Question 2

```
> summary(sparrow)
```

	X1	X2	X3	X4
Min.	:152.0	Min. :230.0	Min. :30.10	Min. :17.20
1st Qu.:	155.0	1st Qu.:238.0	1st Qu.:30.90	1st Qu.:18.10
Median :	158.0	Median :242.0	Median :31.50	Median :18.50
Mean :	158.0	Mean :241.3	Mean :31.46	Mean :18.47
3rd Qu.:	161.0	3rd Qu.:245.0	3rd Qu.:32.00	3rd Qu.:18.80
Max. :	165.0	Max. :252.0	Max. :33.40	Max. :19.80
	X5			
Min.	:18.60			
1st Qu.:	20.20			
Median :	20.70			
Mean :	20.83			
3rd Qu.:	21.50			
Max. :	23.10			

```
> var(sparrow)
```

```

      X1      X2      X3      X4      X5
X1 13.353741 13.610969 1.9220663 1.3306122 2.1922194
X2 13.610969 25.682823 2.7136054 2.1977041 2.6578231
X3  1.922066  2.713605 0.6316327 0.3422662 0.4146471
X4  1.330612  2.197704 0.3422662 0.3184184 0.3393707
X5  2.192219  2.657823 0.4146471 0.3393707 0.9828231

> eigen(cor(sparrow))$values

[1] 3.6159783 0.5315041 0.3864245 0.3015655 0.1645275

> spar.pc = princomp(sparrow, cor = TRUE)
> names(spar.pc)

[1] "sdev"      "loadings"  "center"    "scale"     "n.obs"     "scores"    "call"

> spar.pc$loadings

Loadings:
  Comp.1 Comp.2 Comp.3 Comp.4 Comp.5
X1 -0.452      0.690  0.420 -0.374
X2 -0.462 -0.300  0.341 -0.548  0.530
X3 -0.451 -0.325 -0.454  0.606  0.343
X4 -0.471 -0.185 -0.411 -0.388 -0.652
X5 -0.398  0.876 -0.178      0.192

      Comp.1 Comp.2 Comp.3 Comp.4 Comp.5
SS loadings      1.0  1.0  1.0  1.0  1.0
Proportion Var   0.2  0.2  0.2  0.2  0.2
Cumulative Var   0.2  0.4  0.6  0.8  1.0

> summary(spar.pc)

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

> round(apply(spar.pc$scores, 2, cor, sparrow), 5)

      Comp.1  Comp.2  Comp.3  Comp.4  Comp.5
[1,] -0.85913  0.03698  0.42922  0.23087 -0.15166
[2,] -0.87792 -0.21839  0.21170 -0.30086  0.21501
[3,] -0.85674 -0.23663 -0.28253  0.33295  0.13904
[4,] -0.89514 -0.13464 -0.25545 -0.21322 -0.26433
[5,] -0.75621  0.63900 -0.11093 -0.03782  0.07806

> pcs = spar.pc$scores
> dim(pcs)

```

```
[1] 49 5
```

```
> survivaLabels = c(rep(1, 21), rep(0, 28))
```

```
> survivaLabels
```

```
[1] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0
```

```
[39] 0 0 0 0 0 0 0 0 0 0 0 0
```

```
> plot(pcs[, 1], pcs[, 2], main = "P.C. scores identifying survivors",
```

```
+ type = "n")
```

```
> text(pcs[, 1], pcs[, 2], survivaLabels)
```

