

# Exercícios Análise de Componentes Principais

21 de agosto de 2012

## 1 Exemplo de dimensões do corpo de Pardocas

Um grupo de biólogos, pretende estudar as componentes principais das dimensões do corpo de algumas aves (Pardocas), afim de verificar se há relação entre a morte das aves com tais dimensões.

Os dados analisados seguem a seguinte ordem (por coluna):

1. Comprimento total
2. Extensão alar
3. Comprimento do bico e cabeça
4. Comprimento do úmero
5. Comprimento da quilha do esterno

Os pássaros de 1 a 21 sobreviveram, os pássaros de 22 a 49 morreram.

Resolução:

1. Leitura do arquivo com os dados:

```
> library (RODBC)
> setwd("C://Users/User/Documents/Estatistica/ACP")
> dados <- odbcConnectExcel("corpo_pardocas.xls")
> X <- sqlFetch(dados, "p1")
> odbcClose(dados)
> #Verificando a ordem:
>
> dim(X)

[1] 48 5

> str(X)

'data.frame':      48 obs. of  5 variables:
 $ F1: num  154 153 153 155 163 157 155 164 158 158 ...
 $ F2: num  240 240 236 243 247 238 239 248 238 240 ...
 $ F3: num  30.4 31 30.9 31.5 32 30.9 32.8 32.7 31 31.3 ...
 $ F4: num  17.9 18.4 17.7 18.6 19 18.4 18.6 19.1 18.8 18.6 ...
 $ F5: num  19.6 20.6 20.2 20.3 20.9 20.2 21.2 21.1 22 22 ...
```

2. Calculando a matriz de covariância e de correlação:

```
> (S <- round(cov(X),2))
```

```

      F1   F2   F3   F4   F5
F1 13.55 14.06 1.97 1.36 2.22
F2 14.06 25.94 2.76 2.24 2.74
F3  1.97  2.76 0.64 0.35 0.42
F4  1.36  2.24 0.35 0.33 0.35
F5  2.22  2.74 0.42 0.35 1.00

```

```
> (R <- round(cor(X),2))
```

```

      F1   F2   F3   F4   F5
F1 1.00 0.75 0.67 0.65 0.60
F2 0.75 1.00 0.68 0.77 0.54
F3 0.67 0.68 1.00 0.76 0.53
F4 0.65 0.77 0.76 1.00 0.61
F5 0.60 0.54 0.53 0.61 1.00

```

3. Inserindo a matriz de correlação R e o número de observações, e o teste de Bartlett:

```
> library(psych)
> cortest.bartlett(R, n=nrow(X))
```

```
$chisq
[1] 148.8431
```

```
$p.value
[1] 6.447261e-27
```

```
$df
[1] 10
```

```
> z <- scale(X)
```

```
> #AUTOVALORES E AUTOVETORES
```

```
>
```

```
> autovalor.autovetor <- as.data.frame(eigen(R))
```

```
> #Recebendo porcentual das variâncias e porcentagem acumulativa
```

```
>
```

```
> var.porc = autovalor.autovetor$values / sum(autovalor.autovetor$values)*100
```

```
> var.acum = cumsum(var.porc)
```

```
> (porc.explic <- round(data.frame(autovalores = autovalor.autovetor$values, var.porc = var.porc
```

```

 autovalores var.porc var.acum
1      3.634  72.686  72.686
2      0.523  10.463  83.149
3      0.383   7.655  90.804
4      0.296   5.923  96.727
5      0.164   3.273 100.000

```

4. O gráfico 'Screeplot', pode ser visto na figura:

5. A matriz dos coeficientes das combinações lineares referentes as componentes principais:

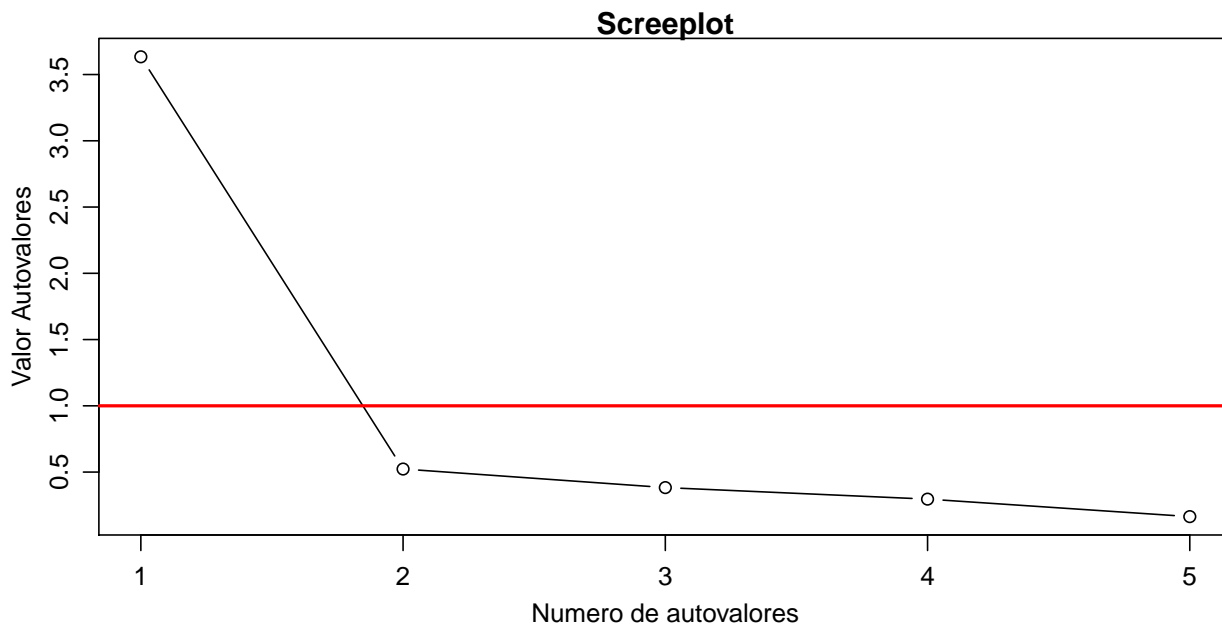
```
> (e <- round(as.matrix(autovalor.autovetor[,-1]),3))
```

```

 vectors.1 vectors.2 vectors.3 vectors.4 vectors.5
[1,]    -0.453    -0.010     0.690     0.398    -0.401
[2,]    -0.464    -0.286     0.325    -0.536     0.557
[3,]    -0.450    -0.318    -0.465     0.612     0.323

```

```
> plot(porc.explic$autovalores, main = "Screeplot", type = "b", ylab = "Valor Autovalores", xlab = "Numero de autovalores")
> abline(h=1, col=2, lwd=2)
```



```
[4,] -0.469 -0.157 -0.430 -0.422 -0.626
[5,] -0.397 0.890 -0.130 -0.023 0.180
```

```
> (V <- round(diag(autovalor.autovetor[,1]),3))
```

```
      [,1] [,2] [,3] [,4] [,5]
[1,] 3.634 0.000 0.000 0.000 0.000
[2,] 0.000 0.523 0.000 0.000 0.000
[3,] 0.000 0.000 0.383 0.000 0.000
[4,] 0.000 0.000 0.000 0.296 0.000
[5,] 0.000 0.000 0.000 0.000 0.164
```

6. E os escores:

```
> #ESCORES DAS COMPONENTES Y1 E Y2
>
> escores <- round(t(e) %*% t(z),2)
```

7. Matriz de correlação entre  $Y_k$  e as variáveis  $x_1$

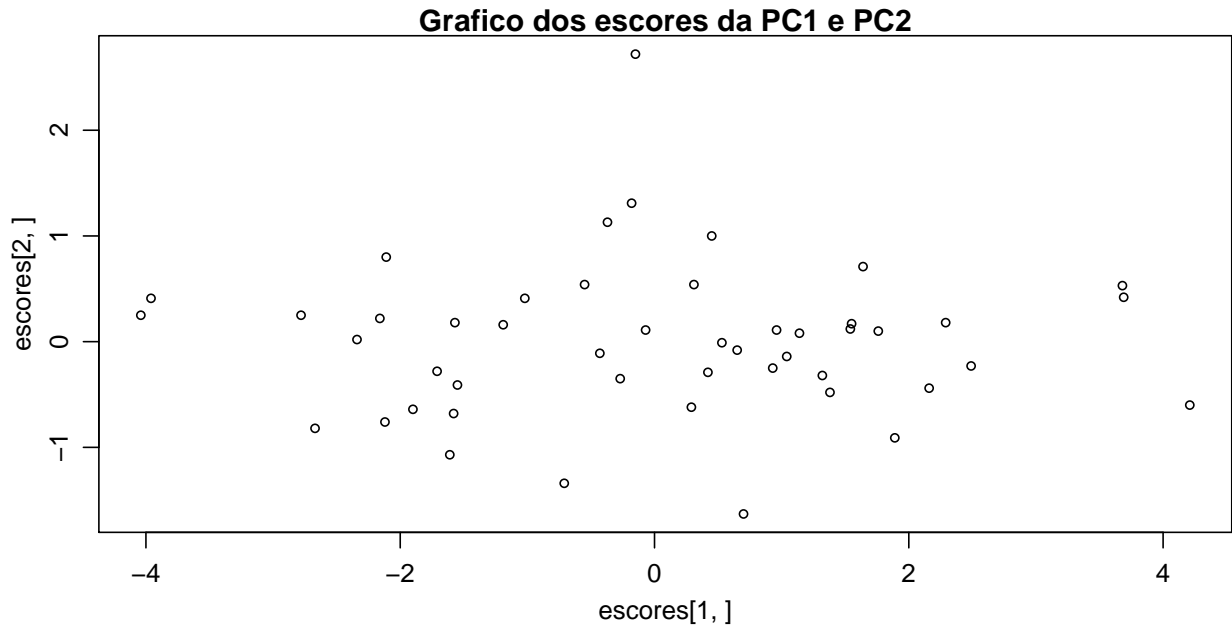
```
> (CORR.Y.COM.X <- round(e %*%sqrt(V),2))
```

```
      [,1] [,2] [,3] [,4] [,5]
[1,] -0.86 -0.01 0.43 0.22 -0.16
[2,] -0.88 -0.21 0.20 -0.29 0.23
[3,] -0.86 -0.23 -0.29 0.33 0.13
[4,] -0.89 -0.11 -0.27 -0.23 -0.25
[5,] -0.76 0.64 -0.08 -0.01 0.07
```

8. Algumas das possíveis visualizações de escores:

9. De outra forma:

```
> plot(escores[2,] ~ escores[1,], cex = 0.7, main = "Grafico dos escores da PC1 e PC2")
```



```
> library(stats)
> args(prcomp)
```

```
function (x, ...)
NULL
```

```
> PCA <- prcomp(X, scale = TRUE)
> print(PCA)
```

Standard deviations:

```
[1] 1.9055447 0.7250034 0.6225445 0.5452366 0.3980260
```

Rotation:

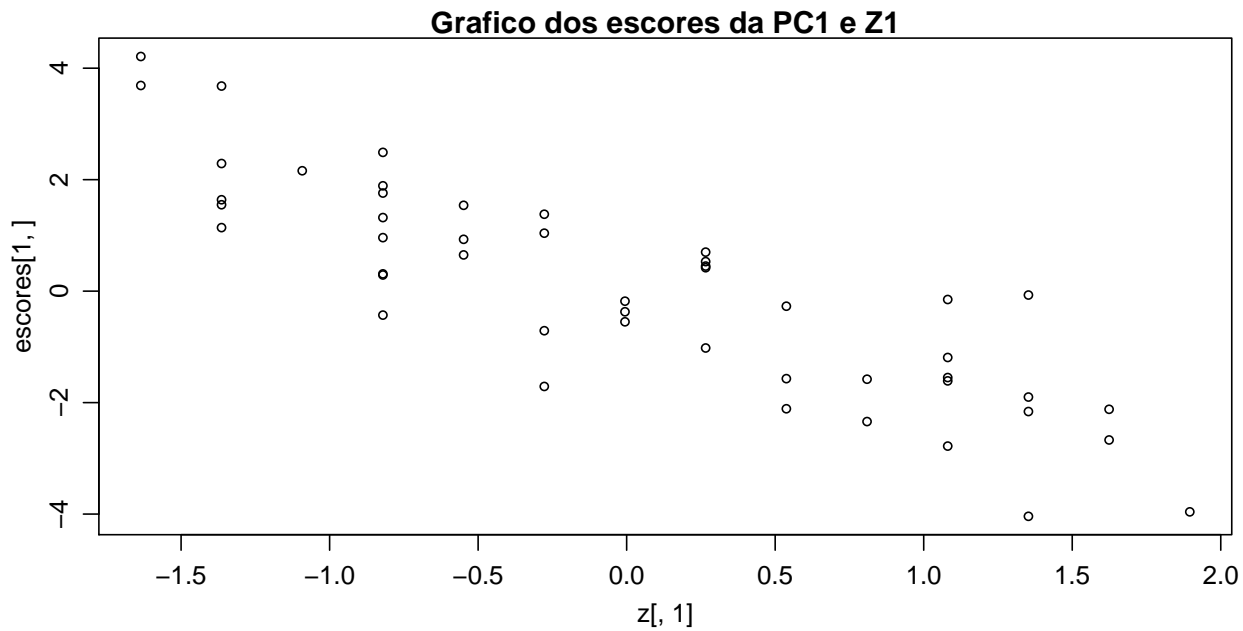
	PC1	PC2	PC3	PC4	PC5
F1	0.4526859	-0.02040548	0.6769268	0.42251233	0.3973822
F2	0.4633174	0.28272915	0.3513659	-0.52640838	-0.5521208
F3	0.4494818	0.32501600	-0.4722171	0.60061898	-0.3295430
F4	0.4695952	0.17798051	-0.4120750	-0.42635262	0.6294608
F5	0.3973146	-0.88449725	-0.1597417	-0.04310271	-0.1800847

```
> summary(PCA)
```

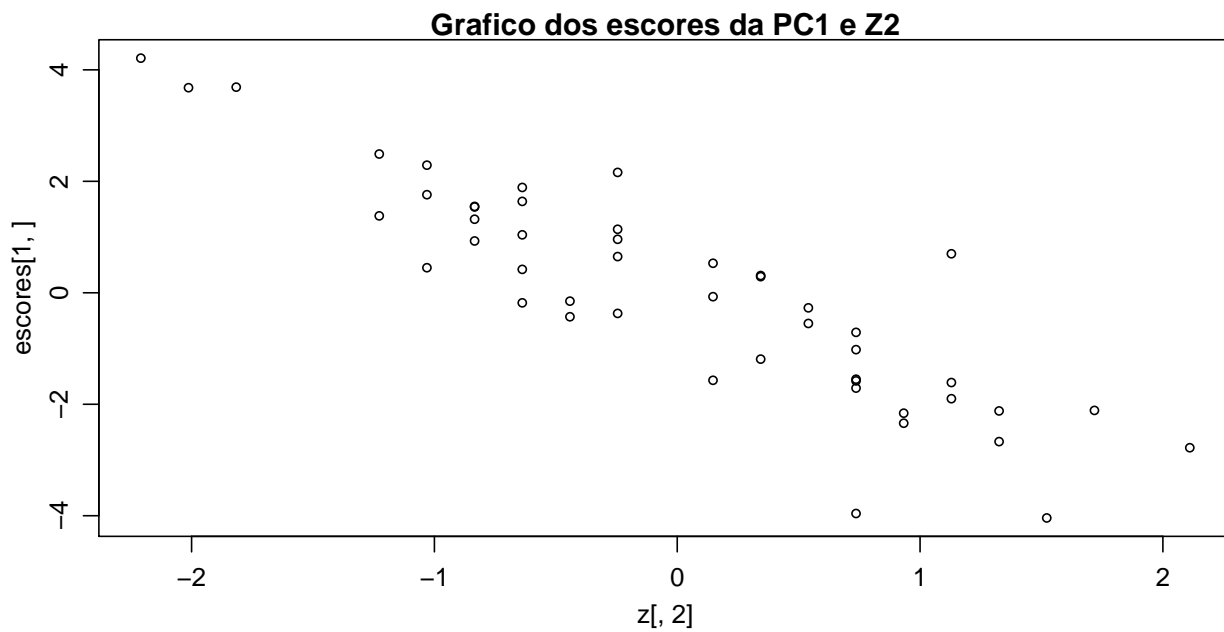
Importance of components:

	PC1	PC2	PC3	PC4	PC5
Standard deviation	1.9055	0.7250	0.62254	0.54524	0.39803
Proportion of Variance	0.7262	0.1051	0.07751	0.05946	0.03168
Cumulative Proportion	0.7262	0.8314	0.90886	0.96832	1.00000

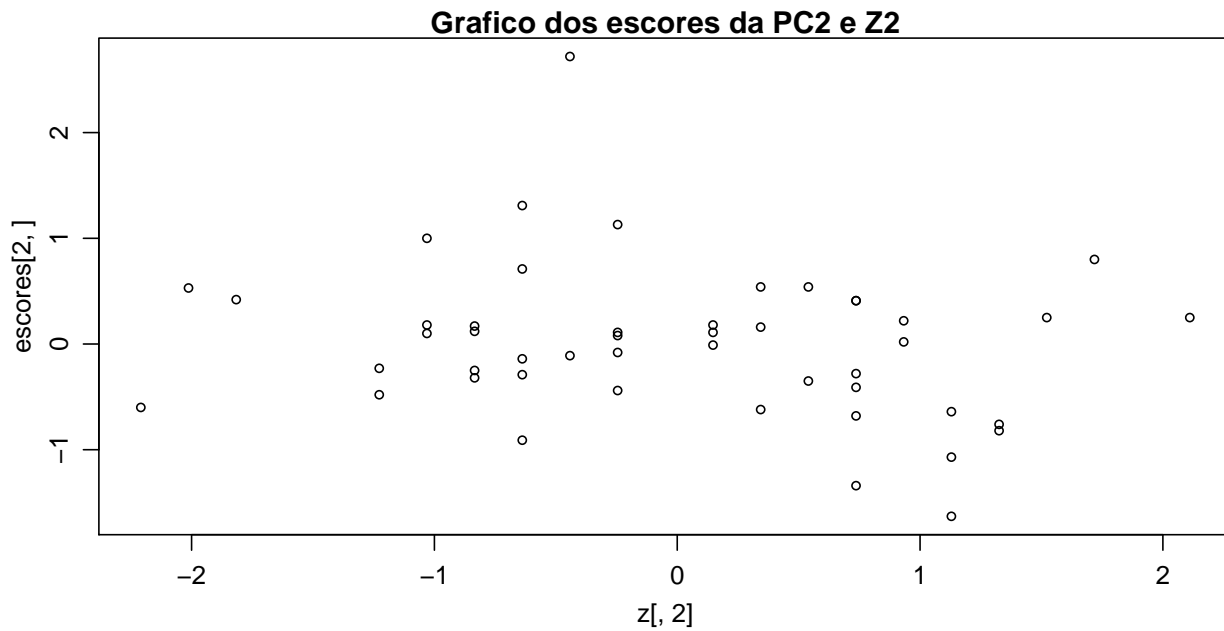
```
> plot(escores[1,] ~ z[,1], cex = 0.7, main = "Grafico dos escores da PC1 e Z1")
```



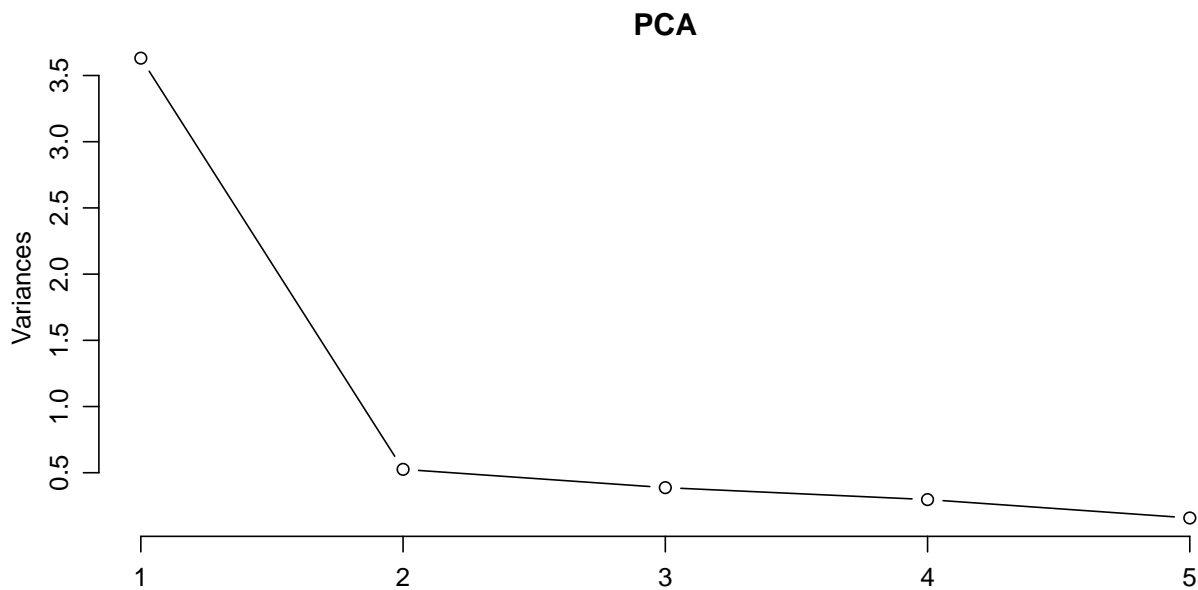
```
> plot(escores[1,] ~ z[,2], cex = 0.7, main = "Grafico dos escores da PC1 e Z2")
```



```
> plot(escores[2,] ~ z[,2], cex = 0.7, main = "Grafico dos escores da PC2 e Z2")
```



```
> screepLOT(PCA, type = c("lines"), min = deparse(substitute(PCA)))
```



```
> biplot(prcomp(X, cor = T), pc.biplot = T, cex = 0.7, expand = 0.8)
```

