



DEPARTAMENTO DE ESTATÍSTICA

03 maio 2023

Entrega 3 - Lista 4

Prof. Dr. George von Borries

Análise Multivariada 1

Aluno: Bruno Gondim Toledo | Matrícula: 15/0167636

6. Exercício 32 da Lista 4 Suponha que um pesquisador padronizou os dados de um estudo através da transformação de Mahalanobis ($\mathbf{Z} = \mathbf{X}\mathbf{D}^{-1/2}$), em que \mathbf{S} é a matriz de variância-covariâncias amostrais. Seria razoável aplicar componentes principais nos dados transformados? Justifique sua resposta.
7. Exercício 37 da Lista 4 - Johnson e Wichern - Exercício 8.12. Dados no arquivo Air Pollution (T1-5.DAT). Os dados correspondem a 42 medidas de poluição do ar observadas na área de Los Angeles em um mesmo horário.
- (a) Resumir os dados em em menos de 7 dimensões (se possível) através de análise de componentes principais utilizando a matrix de covariâncias \mathbf{S} e apresentar suas conclusões.

```
S <- cov(df)

avav <- eigen(S)

avav[1]

## $values
## [1] 304.2578640 28.2761046 11.4644830 2.5243296 1.2795247 0.5287288
## [7] 0.2096157
```

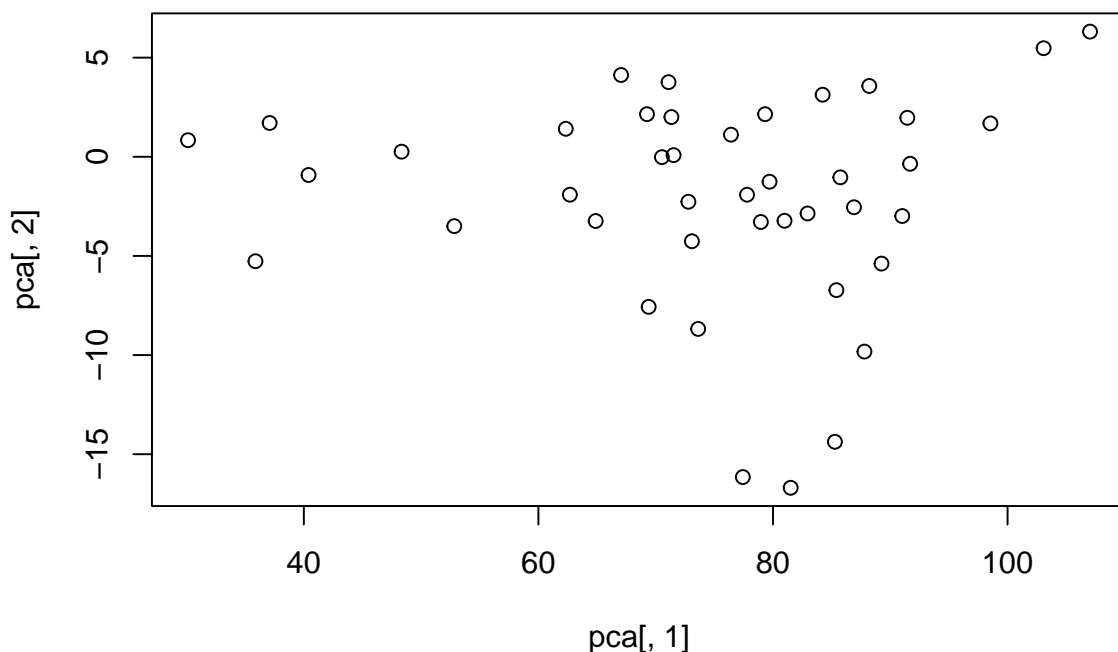
Aqui, notamos que a maior parte da informação está concentrada na primeira dimensão. As dimensões 2 e 3 também são relevantes. Da quarta em diante, apenas uma fração da informação está contida.

Irei então trabalhar com 3 dimensões.

```
transformada <- avav$vectors[, 1:3]

pca <- as.matrix(df) %*% transformada

plot(pca[,1], pca[,2])
```



- (b) Resumir os dados em em menos de 7 dimensões (se possível) através de análise de componentes principais utilizando a matrix de correlações \mathbf{R} e apresentar suas conclusões.

- (c) A escolha da matriz para análise faz alguma diferença? Explique.
- (d) Os dados podem ser resumidos em 3 ou menos dimensões?