

# Prática 1 - Análise Multivariada

*Cibele Russo*

*08 de outubro de 2014*

## Prática 1

O conjunto de dados "Educação" disponível em na Coteia Wiki (Fonte: Ipeadata) mostra dados educacionais por estado de acordo com a descrição a seguir.

Sigla Sigla do estado

Estado Estado

- X1 Frequência escolar de pessoas com 7 a 14 anos
- X2 Defasagem escolar em mais de 1 ano de atraso de pessoas com 7 a 14 anos (2000)
- X3 Anos de estudo de pessoas com 25 anos ou mais (2000)
- X4 Analfabetos com 25 anos ou mais (2000)
- X5 Evasão escolar de pessoas com 7 a 14 anos (2000)
- X6 Evasão escolar de pessoas com 15 a 17 anos (2000)
- X7 Frequência escolar de pessoas com 7 a 22 anos

- (a) Descreva os dados em questão utilizando medidas resumo e análises gráficas.
- (b) Verifique se os dados podem ser considerados normalmente distribuídos.
- (c) Teste hipóteses como  $H_0 : \underline{\mu} = (93, 30, 5, 20, 7, 24, 82)'$  (versus  $\neq$ ).
- (d) Verifique se  $\underline{\mu} = (93, 30, 5, 20, 7, 24, 82)'$  está na região de confiança de  $\underline{\mu}$ .
- (e) Desenvolva e interprete uma análise de componentes principais.

Utilize alguns dos comandos sugeridos a seguir e acrescente o que achar necessário.

```
# ----- Leitura dos dados -----  
Educação <- read.table("http://wiki.icmc.usp.br/images/7/7f/Educacao.txt", header=TRUE)  
dados <- Educação[,3:9]  
rownames(dados) <- Educação[,1]  
n <- nrow(dados)  
p <- ncol(dados)  
# ----- Análise descritiva -----  
# Média amostral  
Xbarra <- colMeans(dados)  
# Matriz de variâncias e covariâncias amostrais  
# Existe diferença entre os comandos abaixo?  
S <- var(dados)  
S <- cov(dados)  
# Matriz de correlações amostrais  
cor(dados)
```

```
##          X1          X2          X3          X4          X5          X6          X7
## X1  1.0000 -0.6381  0.5052 -0.4778 -1.0000 -0.5599  0.7160
## X2 -0.6381  1.0000 -0.8161  0.8930  0.6399  0.2389 -0.4974
## X3  0.5052 -0.8161  1.0000 -0.9005 -0.5070 -0.4973  0.5884
## X4 -0.4778  0.8930 -0.9005  1.0000  0.4797  0.2340 -0.4174
## X5 -1.0000  0.6399 -0.5070  0.4797  1.0000  0.5608 -0.7171
## X6 -0.5599  0.2389 -0.4973  0.2340  0.5608  1.0000 -0.8105
## X7  0.7160 -0.4974  0.5884 -0.4174 -0.7171 -0.8105  1.0000
```

```
# ----- Verificando normalidade multivariada -----
# ----- Teste de Shapiro Wilk -----
library(mvShapiroTest)
mvShapiro.Test(as.matrix(dados))
```

```
##
## Generalized Shapiro-Wilk test for Multivariate Normality by
## Villasenor-Alva and Gonzalez-Estrada
##
## data: as.matrix(dados)
## MVW = 0.9575, p-value = 0.2643
```

```
# ----- Teste para a média -----
mu0= t(c(93, 30, 5, 20, 7, 24, 82))
T2 = n * as.matrix(Xbarra-mu0) %*% solve(S) %*% as.matrix(t(Xbarra-mu0))
#Comparar com (n-1) * p / (n-p) * qf(0.95, p, n-p)
T2 > (n-1) * p / (n-p) * qf(0.95, p, n-p)
```

```
##          [,1]
## [1,] TRUE
```

```
# Conclusão do teste ?
# Análise de Componentes principais
p.comp <- princomp(dados, cor = TRUE,scores = TRUE);
s <- summary(p.comp, loadings = TRUE );
# Interpretação dos resultados ?
```