



1. Organização e análise descritiva de dados multivariados

Prof. Cibele Russo

2º semestre de 2011

Organização de dados - Exemplo: dados de flores

Dados originalmente apresentados em

Fisher, R. A. (1936) The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7, Part II, 179–188.

Table I

<i>Iris setosa</i>				<i>Iris versicolor</i>				<i>Iris virginica</i>			
Sepal length	Sepal width	Petal length	Petal width	Sepal length	Sepal width	Petal length	Petal width	Sepal length	Sepal width	Petal length	Petal width
5.1	3.5	1.4	0.2	7.0	3.2	4.7	1.4	6.3	3.3	6.0	2.5
4.9	3.0	1.4	0.2	6.4	3.2	4.5	1.5	5.8	2.7	5.1	1.9
4.7	3.2	1.3	0.2	6.9	3.1	4.9	1.5	7.1	3.0	5.9	2.1
4.6	3.1	1.5	0.2	5.5	2.3	4.0	1.3	6.3	2.9	5.6	1.8
5.0	3.6	1.4	0.2	6.5	2.8	4.6	1.5	6.5	3.0	5.8	2.2
5.4	3.9	1.7	0.4	5.7	2.8	4.5	1.3	7.6	3.0	6.6	2.1
4.6	3.4	1.4	0.3	6.3	3.3	4.7	1.6	4.9	2.5	4.5	1.7
5.0	3.4	1.5	0.2	4.9	2.4	3.3	1.0	7.3	2.9	6.3	1.8
4.4	2.9	1.4	0.2	6.6	2.9	4.6	1.3	6.7	2.5	5.8	1.8
4.9	3.1	1.5	0.1	5.2	2.7	3.9	1.4	7.2	3.6	6.1	2.5
5.4	3.7	1.5	0.2	5.0	2.0	3.5	1.0	6.5	3.2	5.1	2.0
4.8	3.4	1.6	0.2	5.9	3.0	4.2	1.5	6.4	2.7	5.3	1.9
4.8	3.0	1.4	0.1	6.0	2.2	4.0	1.0	6.8	3.0	5.5	2.1
4.3	3.0	1.1	0.1	6.1	2.9	4.7	1.4	5.7	2.5	5.0	2.0
5.8	4.0	1.2	0.2	5.6	2.9	3.6	1.3	5.8	2.8	5.1	2.4
5.7	4.4	1.5	0.4	6.7	3.1	4.4	1.4	6.4	3.2	5.3	2.3
5.4	3.9	1.3	0.4	5.6	3.0	4.5	1.5	6.5	3.0	5.5	1.8
5.1	3.5	1.4	0.3	5.8	2.7	4.1	1.0	7.7	3.8	6.7	2.2
5.7	3.8	1.7	0.3	6.2	2.2	4.5	1.5	7.7	2.6	6.9	2.3
5.1	3.8	1.5	0.3	5.6	2.5	3.9	1.1	6.0	2.2	5.0	1.5
5.4	3.4	1.7	0.2	5.9	3.2	4.8	1.8	6.9	3.2	5.7	2.3
5.1	3.7	1.5	0.4	6.1	2.2	4.0	1.3	5.6	2.8	4.9	2.0

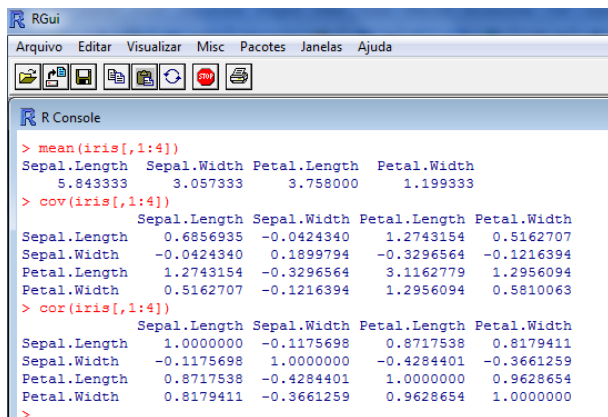
Exemplo de tabela de dados no R

> iris

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
1	5.1	3.5	1.4	0.2	setosa
2	4.9	3.0	1.4	0.2	setosa
3	4.7	3.2	1.3	0.2	setosa
4	4.6	3.1	1.5	0.2	setosa
5	5.0	3.6	1.4	0.2	setosa
6	5.4	3.9	1.7	0.4	setosa
7	4.6	3.4	1.4	0.3	setosa
8	5.0	3.4	1.5	0.2	setosa
9	4.4	2.9	1.4	0.2	setosa
10	4.9	3.1	1.5	0.1	setosa
11	5.4	3.7	1.5	0.2	setosa
12	4.8	3.4	1.6	0.2	setosa
13	4.8	3.0	1.4	0.1	setosa
14	4.3	3.0	1.1	0.1	setosa
15	5.8	4.0	1.2	0.2	setosa
16	5.7	4.4	1.5	0.4	setosa
17	5.4	3.9	1.3	0.4	setosa
18	5.1	3.5	1.4	0.3	setosa
19	5.7	3.8	1.7	0.3	setosa
20	5.1	3.8	1.5	0.3	setosa
21	5.4	3.4	1.7	0.2	setosa

Estatísticas descritivas - Variáveis quantitativas

- > `mean(iris[,1:4])`
- > `cov(iris[,1:4])`
- > `cor(iris[,1:4])`



```
RGui
Arquivo  Editar  Visualizar  Misc  Pacotes  Janelas  Ajuda
[Icons]
R Console
> mean(iris[,1:4])
Sepal.Length Sepal.Width Petal.Length Petal.Width
5.843333      3.057333      3.758000      1.199333
> cov(iris[,1:4])
      Sepal.Length Sepal.Width Petal.Length Petal.Width
Sepal.Length  0.6856935 -0.0424340  1.2743154  0.5162707
Sepal.Width  -0.0424340  0.1899794 -0.3296564 -0.1216394
Petal.Length  1.2743154 -0.3296564  3.1162779  1.2956094
Petal.Width   0.5162707 -0.1216394  1.2956094  0.5810063
> cor(iris[,1:4])
      Sepal.Length Sepal.Width Petal.Length Petal.Width
Sepal.Length  1.0000000 -0.1175698  0.8717538  0.8179411
Sepal.Width  -0.1175698  1.0000000 -0.4284401 -0.3661259
Petal.Length  0.8717538 -0.4284401  1.0000000  0.9628654
Petal.Width   0.8179411 -0.3661259  0.9628654  1.0000000
>
```

Estatísticas descritivas - Variáveis quantitativas

Exercício:

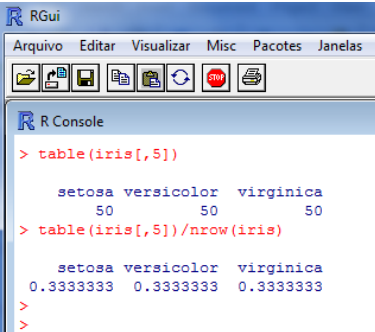
Obter estatísticas descritivas separadas por espécie de flores

Estatísticas descritivas - Variáveis qualitativas

Tabelas de frequências absolutas e relativas

> **table(iris[,5])**

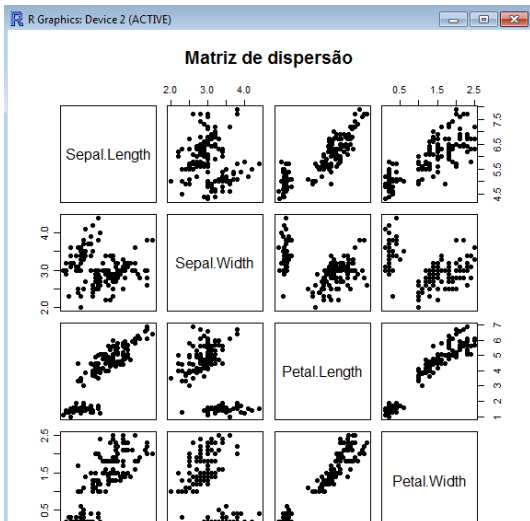
> **table(iris[,5])/nrow(iris)**



```
RGui
Arquivo  Editar  Visualizar  Misc  Pacotes  Janelas
[Icons]
R Console
> table(iris[,5])
      setosa versicolor virginica
      50         50         50
> table(iris[,5])/nrow(iris)
      setosa versicolor virginica
0.3333333  0.3333333  0.3333333
>
>
```

Gráficos de dispersão multivariados

> `pairs(iris[1:4], pch=16, main="Matriz de dispersão")`



Gráficos de dispersão multivariados

Exercício:

Obter gráficos de dispersão multivariados separadas por espécie de flores

Gráficos de dispersão multivariados

Exercício:

Obter gráficos de dispersão multivariados separadas por espécie de flores

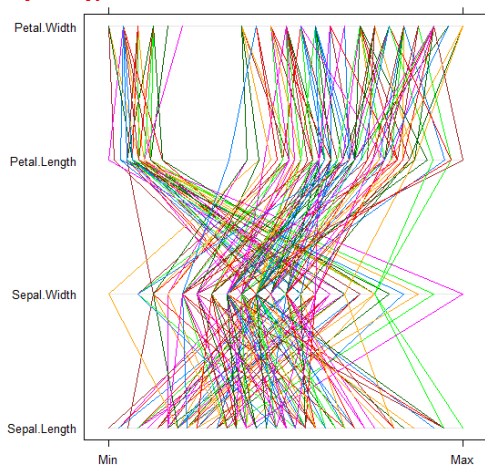
Exercício:

Incluir correlações no triângulo inferior e histogramas na diagonal

(Ver referência “Gráficos estadísticos con R”)

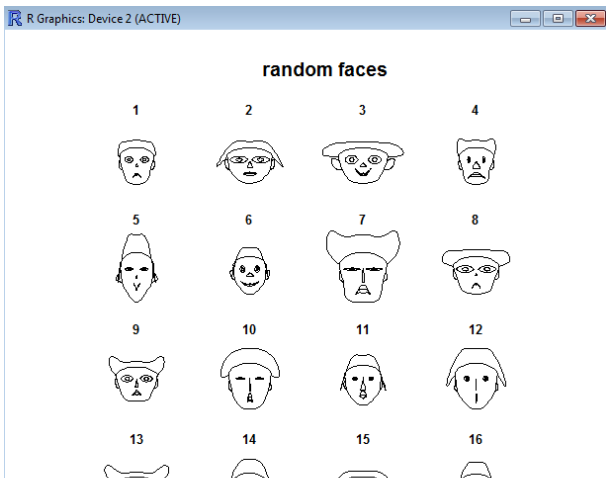
Gráfico de coordenadas paralelas

- > **library(lattice)**
- > **parallel(iris[,1:4])**



Faces de Chernoff

- > `require(TeachingDemos)`
- > `faces(matrix(sample(1:1000, 128,), 16, 8), main="random faces")`



Faces de Chernoff

Ver referência:

[http://addictedtor.free.fr/graphiques/
RGraphGallery.php?graph=87](http://addictedtor.free.fr/graphiques/RGraphGallery.php?graph=87)

Gráficos de estrelas

```
> publicfirms <- read.table("F:\\\\ datos \\\ publicfirms.txt")
> colnames(publicfirms) <-
c("X1", "X2", "X3", "X4", "X5", "X6", "X7", "X8", "Empresa")
> rownames(publicfirms) <- publicfirms[,9]
> stars(publicfirms[,1:8], key.loc=c(11,2), main='Gráfico de
estrelas para dados de empresas públicas')
```

Gráficos de estrellas

Gráfico de estrellas para datos de empresas públicas

