

Análise de Dados - Aula 1

Ítalo e José Antonio

2024-09-25

Análise de Dados - Aula 1

Questão 1

Descrição:

Nesta questão, carregamos o conjunto de dados `iris` e visualizamos sua estrutura utilizando a função `View()`.

```
# Carregar os dados do arquivo para um dataset - computer+hardware
data <- read.table("C:/Users/2019101100910126/Documents/Github/EstudoDirigindo-MinecaoDeDados/Computer I

# Visualizar o dataset
View(data)
```

Questão 2

Descrição:

Aqui, utilizamos as funções `dim()`, `nrow()` e `ncol()` para verificar as dimensões do conjunto de dados, o número de linhas e o número de colunas.

```
# Verificar dimensões do dataset
dim(iris)
```

```
## [1] 150 5
```

```
# Número de linhas
nrow(iris)
```

```
## [1] 150
```

```
# Número de colunas
ncol(iris)
```

```
## [1] 5
```

Questão 3

Descrição:

Utilizamos a função `str()` para exibir a estrutura do conjunto de dados `iris`, o que nos permite observar os tipos de variáveis e os primeiros valores de cada uma.

```
# Exibir a estrutura do dataset
str(iris)

## 'data.frame': 150 obs. of 5 variables:
## $ Sepal.Length: num 5.1 4.9 4.7 4.6 5 5.4 4.6 5 4.4 4.9 ...
## $ Sepal.Width : num 3.5 3 3.2 3.1 3.6 3.9 3.4 3.4 2.9 3.1 ...
## $ Petal.Length: num 1.4 1.4 1.3 1.5 1.4 1.7 1.4 1.5 1.4 1.5 ...
## $ Petal.Width : num 0.2 0.2 0.2 0.2 0.2 0.4 0.3 0.2 0.2 0.1 ...
## $ Species : Factor w/ 3 levels "setosa","versicolor",...: 1 1 1 1 1 1 1 1 1 1 ...
```

Questão 4

Descrição:

Nesta etapa, adicionamos dados ausentes ao conjunto de dados e em seguida removemos a linha adicional. Após isso, usamos a função `summary()` para gerar um resumo estatístico.

```
# Adicionar uma linha com dados ausentes
# iris <- rbind(iris, c(5.86, NA, NA, 1.89, as.factor("virginica")))

# Remover a última linha (caso tenhamos adicionado)
# iris <- iris[-151,]

# Resumo do dataset
summary(iris)
```

```
## Sepal.Length Sepal.Width Petal.Length Petal.Width
## Min. :4.300 Min. :2.000 Min. :1.000 Min. :0.100
## 1st Qu.:5.100 1st Qu.:2.800 1st Qu.:1.600 1st Qu.:0.300
## Median :5.800 Median :3.000 Median :4.350 Median :1.300
## Mean :5.843 Mean :3.057 Mean :3.758 Mean :1.199
## 3rd Qu.:6.400 3rd Qu.:3.300 3rd Qu.:5.100 3rd Qu.:1.800
## Max. :7.900 Max. :4.400 Max. :6.900 Max. :2.500
## Species
## setosa :50
## versicolor:50
## virginica :50
##
##
##
```

Questão 5

Descrição:

Aqui, usamos a função `summary()` para obter um resumo estatístico e calcular a porcentagem de valores ausentes.

```
# Resumo estatístico dos dados
summary(iris)
```

```
## Sepal.Length Sepal.Width Petal.Length Petal.Width
## Min. :4.300 Min. :2.000 Min. :1.000 Min. :0.100
## 1st Qu.:5.100 1st Qu.:2.800 1st Qu.:1.600 1st Qu.:0.300
## Median :5.800 Median :3.000 Median :4.350 Median :1.300
## Mean :5.843 Mean :3.057 Mean :3.758 Mean :1.199
## 3rd Qu.:6.400 3rd Qu.:3.300 3rd Qu.:5.100 3rd Qu.:1.800
## Max. :7.900 Max. :4.400 Max. :6.900 Max. :2.500
## Species
## setosa :50
## versicolor:50
## virginica :50
##
##
##
```

```
# Cálculo da porcentagem de valores ausentes (NA) em cada coluna
na_percent <- sapply(iris, function(x) sum(is.na(x)) / length(x) * 100)
na_percent
```

```
## Sepal.Length Sepal.Width Petal.Length Petal.Width Species
## 0 0 0 0 0
```

Questão 6

Descrição:

(Inclua a descrição e o código para a Questão 6)

Questão 7

Descrição:

(Inclua a descrição e o código para a Questão 7)

Questão 8

Descrição:

Calculamos as médias, os valores mínimos e máximos das variáveis `Sepal.Length` e `Petal.Length` para as espécies `setosa`, `versicolor` e `virginica`.

```
# Setosa
mean(iris$Sepal.Length[1:50])
```

```
## [1] 5.006
```

```
mean(iris$Petal.Length[1:50])
```

```
## [1] 1.462
```

```
min(iris$Sepal.Length[1:50])
```

```
## [1] 4.3
```

```
min(iris$Petal.Length[1:50])
```

```
## [1] 1
```

```
max(iris$Sepal.Length[1:50])
```

```
## [1] 5.8
```

```
max(iris$Petal.Length[1:50])
```

```
## [1] 1.9
```

```
# Versicolor
```

```
mean(iris$Sepal.Length[51:100])
```

```
## [1] 5.936
```

```
mean(iris$Petal.Length[51:100])
```

```
## [1] 4.26
```

```
# Virginica
```

```
mean(iris$Sepal.Length[101:150])
```

```
## [1] 6.588
```

```
mean(iris$Petal.Length[101:150])
```

```
## [1] 5.552
```

Questão 9

Descrição:

(Inclua a descrição e o código para a Questão 9)

Questão 10

Descrição:

(Inclua a descrição e o código para a Questão 10)

Questão 11**Descrição:**

(Inclua a descrição e o código para a Questão 11)

Questão 12**Descrição:**

(Inclua a descrição e o código para a Questão 12)

Questão 13**Descrição:**

(Inclua a descrição e o código para a Questão 13)

Questão 14**Descrição:**

(Inclua a descrição e o código para a Questão 14)

Questão 15**Descrição:**

(Inclua a descrição e o código para a Questão 15)