

Lista 3- Subsetting

Lucas C. França

2022-03-15

Instale e carregue o pacote `quantreg` e, em seguida, execute o comando `data(Mammals)` para permitir a utilização do banco de dados `Mammals`.

```
data("Mammals")
```

Questão 1

Com base nos dados contidos em `Mammals`, responda os itens a seguir.

a) Descreva o conteúdo do banco de dados `Mammals`.

Abaixo uma prévia deste conjunto de dados e sua estrutura:

```
head(Mammals)
```

```
##   weight speed 1hoppers 1specials
## 1   6000    35     FALSE     FALSE
## 2   4000    26     FALSE     FALSE
## 3   3000    25     FALSE     FALSE
## 4   1400    45     FALSE     FALSE
## 5    400    70     FALSE     FALSE
## 6    350    70     FALSE     FALSE
```

```
str(Mammals)
```

```
## 'data.frame':   107 obs. of  4 variables:
## $ weight : num  6000 4000 3000 1400 400 350 300 260 250 3800 ...
## $ speed : num  35 26 25 45 70 70 64 70 40 25 ...
## $ 1hoppers : logi  FALSE FALSE FALSE FALSE FALSE FALSE ...
## $ 1specials: logi  FALSE FALSE FALSE FALSE FALSE FALSE ...
```

O conjunto de dados contém 107 observações, representando espécies de mamíferos, e 4 variáveis, representando a velocidade máxima de corrida, massa corporal e dois grupos de especial interesse, “hoppers” que representa animais saltitantes, como cangurus, e “specials” tal como a preguiça e o hipopótamo, cujos estilos de vida não têm a velocidade como fator importante.

Dicionário de dados:

Variável	Definição	Key
Weight	Peso em kg	
Speed	Velocidade máxima de corrida	
Hoppers	Se o animal é saltitante	TRUE = Sim, FALSE = Não
Specials	Animais com estilos de vida onde a velocidade não é um fator importante	TRUE = Sim, FALSE = Não

Abaixo atributos de cada variável:

```
summary(Mammals)
```

```
##      weight      speed      hoppers      specials
## Min.   : 0.016   Min.   : 1.60   Mode :logical   Mode :logical
## 1st Qu.: 1.700   1st Qu.: 22.50   FALSE:96       FALSE:97
## Median : 34.000   Median : 48.00   TRUE :11       TRUE :10
## Mean   : 278.688   Mean    : 46.21
## 3rd Qu.: 142.500   3rd Qu.: 65.00
## Max.    :6000.000   Max.     :110.00
```

b) Qual é a maior velocidade dos mamíferos apresentados?

Para encontrar a amplitude de um conjunto de dados, podemos usar a função `range()`, que retornará o valor mínimo e máximo, respectivamente, da seguinte forma:

```
range(Mammals$speed)
```

```
## [1] 1.6 110.0
```

Para encontrar somente o valor máximo, podemos usar a função `max()`:

```
max(Mammals$speed)
```

```
## [1] 110
```

Já para retornar a observação em que o valor máximo aparece, podemos usar a seguinte linha de código:

```
Mammals[Mammals$speed == max(Mammals$speed),]
```

```
##      weight speed hoppers specials
## 53      55   110   FALSE     FALSE
```

Como retornado pelas funções, a maior velocidade apresentada é de 110, pelo mamífero de nº 53.

c) Qual é o peso do mamífero mais veloz?

Como já encontrado na questão anterior, o mamífero mais rápido é o de nº 53, cujo o peso é de 55.

Uma forma alternativa de se encontrar o mesmo resultado é através da função `by()`, que retornará um `head()` (como segundo argumento), em que a condição lógica (como primeiro argumento) for verdadeira e outro em que ela for falsa.

```
by(Mammals, Mammals$speed == max(Mammals$speed), head)
```

```
## Mammals$speed == max(Mammals$speed): FALSE
##   weight speed  hoppers specials
## 1   6000    35    FALSE    FALSE
## 2   4000    26    FALSE    FALSE
## 3   3000    25    FALSE    FALSE
## 4   1400    45    FALSE    FALSE
## 5    400    70    FALSE    FALSE
## 6    350    70    FALSE    FALSE
## -----
## Mammals$speed == max(Mammals$speed): TRUE
##   weight speed  hoppers specials
## 53     55   110    FALSE    FALSE
```

Como retornado pela função, somente o animal de n. 53 tem como condição verdadeira a velocidade máxima.

d) Qual é a velocidade do mamífero mais pesado?

De forma análoga, podemos retornar o mamífero mais pesado através das duas seguintes alternativas:

```
Mammals[Mammals$weight == max(Mammals$weight),]
```

```
##   weight speed  hoppers specials
## 1   6000    35    FALSE    FALSE
```

```
by(Mammals, Mammals$weight == max(Mammals$weight), head)
```

```
## Mammals$weight == max(Mammals$weight): FALSE
##   weight speed  hoppers specials
## 2   4000    26    FALSE    FALSE
## 3   3000    25    FALSE    FALSE
## 4   1400    45    FALSE    FALSE
## 5    400    70    FALSE    FALSE
## 6    350    70    FALSE    FALSE
## 7    300    64    FALSE    FALSE
## -----
## Mammals$weight == max(Mammals$weight): TRUE
##   weight speed  hoppers specials
## 1   6000    35    FALSE    FALSE
```

Ambas retornam o mamífero de n° 1, cuja velocidade é de 35.

e) Tem alguma velocidade não disponível (NA)?

Podemos checar se uma variável é não disponível(NA), através da função `is.na()`, que retornará um valor verdadeiro ou falso. Ao adicionarmos como linha ao conjunto de dados, retornaremos o subconjunto em que a condição lógica for verdadeira:

```
Mammals[is.na(Mammals$speed) == TRUE, ]
```

```
## [1] weight    speed      hoppers    specials
## <0 linhas> (ou row.names de comprimento 0)
```

Em que nenhuma velocidade não disponível foi encontrada.

f) Qual é a velocidade mínima dos mamíferos “não especiais”?**

Podemos encontrar o valor mínimo de uma variável em um conjunto de dados através da função `min()`, que receberá como argumento o subconjunto da variável `speed` em que a variável `specials` for falsa, da seguinte forma:

```
min(Mammals$speed[which( Mammals$specials == F)])
```

```
## [1] 2.4
```

Em que o valor retornado foi de 2,4.

g) Quantos mamíferos especiais ou saltadores o banco de dados contém?

Podemos encontrar o número de ocorrências em que a condição lógica for verdadeira através da função `table()`, que receberá “ em que o animal é saltador ou especial” como argumento.

```
table(Mammals$hoppers == T | Mammals$specials == T)
```

```
##  
## FALSE  TRUE  
##    86    21
```

Em que valor retornado de ocorrências verdadeiras é de 21.

Uma solução alternativa é usar a função `count()` do pacote `dplyr`, da seguinte forma:

```
library("dplyr")  
Mammals %>%  
  dplyr::count(Mammals$hoppers == T | Mammals$specials == T)
```

```
##   Mammals$hoppers == T | Mammals$specials == T  n  
## 1                                     FALSE 86  
## 2                                     TRUE  21
```

Em que o resultado de 21 foi igualmente encontrado.

h) Qual é a velocidade média dos mamíferos do item g)?

Podemos encontrar a média de um conjunto de dados através da função `mean()`, que receberá como argumento o subconjunto `speed` em que `hoppers` ou `specials` for verdadeiro.

```
mean(Mammals$speed[which(Mammals$hoppers == T | Mammals$specials == T)])
```

```
## [1] 35.81905
```

Que encontrou como média o valor de 35.81905.

i) Tem algum mamífero com o mesmo peso do seu? Quais são?**

Podemos encontrar uma ocorrência em um conjunto de dados adicionando a condição lógica na função `which()`, da seguinte forma:

```
which(Mammals$weight == 60)
```

```
## [1] 52
```

Que encontrou o mamífero de número 52 como tendo o mesmo peso que o meu.

j) Altere o peso do mamífero encontrado em i) para NA. (se não encontrou nenhum, substitua por NA o peso do mamífero mais leve).

Podemos sobrescrever o valor antigo como NA da seguinte forma:

```
Mammals$weight[Mammals$weight == 60] <- NA
```

Para checarmos se o valor foi devidamente alterado, basta retornar a posição do conjunto de dados:

```
Mammals[52,]
```

```
##      weight speed 1hoppers 1specials  
## 52      NA     60     FALSE     FALSE
```

Que recebeu NA como novo valor para o peso.