

Aula 2 - Classes de dados e vetores no R

Eduardo Koerich Nery

Nesta aula iremos conhecer as classes de dados que existem no R e sua forma de organização básica. Assim como na língua portuguesa as palavras podem pertencer a diferentes classes gramaticais (e.g. substantivos, adjetivos, verbos, advérbios), na linguagem R os dados podem pertencer a diferentes classes: **numérico, inteiro, carácter e lógico**. Ainda, assim como na língua portuguesa nós organizamos as palavras em frases, na linguagem R nós organizamos os dados em **vetores**. Dado isto, veremos como essas classes de dados se comportam e como manipular os dados em vetores.

Numérico e Inteiro

Os dados numéricos e inteiros são aqueles que são passíveis das operações aritméticas: soma $+$, subtração $-$, multiplicação $*$ e divisão $/$. Além destas, também é possível aplicar a potenciação x^y , onde x é a base e y é o expoente; e transformações logarítmicas $\log(x, \text{base})$, onde x é o logaritmando e base é a base.

Para criar um vetor de numéricos ou de inteiros, inicie a expressão `c()`; digite dentro dos parênteses os números, separando os mesmos por vírgula `,`; atribua o vetor a um objeto. Por exemplo, para criar um vetor com as alturas de três pessoas, eu executei `alturas <- c(1.68, 1.84, 1.72)`. Para verificar se o vetor criado possui dados numéricos, você pode usar a função `is.numeric()` sobre o vetor.

Agora, você pode executar operações com o vetor, por exemplo `alturas + 10` ou `alturas/2`. Perceba que a mesma operação é executada em cada um dos elementos do vetor. Você também pode fazer operações com dois vetores de mesmo tamanho, por exemplo `alturas * c(2,3,4)`. Neste caso, cada elemento do primeiro vetor é operado com o elemento de posição correspondente no segundo vetor. Ainda, você pode fazer operações com vetores de tamanhos diferentes, por exemplo `alturas * c(2,3)`. O R avisa sobre essa diferença de tamanho e prossegue as operações repetindo os elementos do vetor menor. Esta forma de operar os dados é chamada de **comportamento orientado a vetor**.

Carácter

Os caracteres são aqueles dados que **não** são passíveis de operações aritméticas. Isto inclui letras, símbolos e acentos gráficos. Símbolos numéricos serão interpretados como caracteres quando inseridos em um vetor com outros caracteres. Neste último caso, símbolos numéricos poderão ser transformados em números ou inteiros pela função `as.numeric()`.

Para criar um vetor de caracteres, inicie a expressão `c()`; digite dentro dos parênteses os caracteres, cada dado entre aspas `"`; separe os dados por vírgula `,`; atribua o vetor a um objeto. Por exemplo, para criar um vetor com os nomes de pessoas, eu executei `nomes <- c("Ana", "Pedro", "Maria")`. Para verificar se o vetor criado possui dados do tipo carácter, você pode usar a função `is.character()` sobre o vetor.

Uma vez criado o vetor de caracteres, você pode extrair algumas informações do mesmo e manipular seus elementos. Você pode contar o número de caracteres de um vetor, por exemplo `nchar(nomes)`. Você também pode unir dois ou mais vetores de caracteres e uma única linha, por exemplo `paste(nomes, "tem", alturas, " metros de altura")`. Ainda, você pode separar uma linha em diferentes vetores de caracteres com base em algum critério, por exemplo `strsplit("Pedro tem 1.84 metros de altura", split=" ")`.

Lógico

Existem apenas dois valores lógicos, *TRUE* e *FALSE*, que sinalizam se uma dada condição foi satisfeita ou não, respectivamente. Nós vimos esses dados anteriormente, quando usamos a função `is.numeric()` sobre o objeto `alturas` e obtivemos a resposta *TRUE*. Os dados lógicos vão aparecer sempre que aplicarmos **operadores relacionais** sobre os dados. Os operadores relacionais são:

Operador	Leitura
<code>==</code>	igual
<code>!=</code>	diferente
<code>></code>	maior que
<code>>=</code>	maior ou igual a
<code><</code>	menor que
<code><=</code>	menor ou igual a

Por exemplo, para descobrir se existem alturas maiores que 1.80 metros, eu executei o comando `alturas > 1.80`. Isto gerou um vetor de dados lógicos, no qual o valor *TRUE* aparece onde a condição foi satisfeita.

Frequentemente nós precisaremos descobrir elementos que satisfazem mais de uma condição, para isso nós teremos que usar **operadores lógicos** sobre os valores lógicos. Os operadores lógicos são:

Operador	Leitura	Interpretação
<code>&</code>	E	<i>TRUE</i> , o elemento satisfaz todas as condições
<code> </code>	OU	<i>TRUE</i> , o elemento satisfaz pelo menos uma condição
<code>!</code>	NÃO	<i>TRUE</i> , o elemento não satisfaz as condições

Por exemplo, para descobrir se existem alturas maiores que 1.70 e menores que 1.80, eu executei o comando `alturas > 1.70 & alturas < 1.80`. Isto gerou um vetor de dados lógicos, no qual o valor *TRUE* aparece onde todas as condições foram satisfeitas.

Uma vez gerados os dados lógicos, nós podemos nos interessar pela posição (ordem) dos dados que satisfazem as condições. Para isso, nós podemos aplicar a função `which()` sobre o vetor de valores lógicos. Por exemplo, para descobrir *quais* alturas são maiores que 1.70 e menores que 1.80, eu executei o comando `which(alturas > 1.70 & alturas < 1.80)`. Isto gerou um vetor numérico indicando a posição dos dados que satisfazem as condições. Neste caso, apenas o segundo elemento.

Vetores

Um vetor é uma linha ou coluna de dados, ou seja, uma estrutura unidimensional para organizar dados. O tamanho de um vetor é definido pelo número de elementos que compõem o vetor. Para acessar o tamanho de um vetor, nós usamos a função `length()`. Por exemplo, para saber quantos nomes existem no vetor 'nomes', eu executei `length(nomes)`.

Dentro de um vetor, cada elemento está indexado a uma posição numérica. Para acessar uma posição dentro de um vetor, nós usamos os colchetes `[]` ao lado do vetor, e adicionamos o número da posição desejada dentro dos colchetes. Por exemplo, para verificar o terceiro nome dentro do vetor 'nomes', eu executei `nomes[3]`. Quando quisermos acessar mais de uma posição, nós adicionamos um vetor numérico dentro das colchetes. Por exemplo, para verificar o primeiro e o terceiro nome dentro do vetor 'nomes', eu executei `nomes[c(1,3)]`. Ainda, quando quisermos acessar todas as posições *exceto* algumas, nós usamos o sinal de menos - antes das posições que precisam ser excluídas. Por exemplo, para verificar todos os nomes exceto o primeiro, eu executei `nomes[-1]`. Por fim, quando quisermos substituir os dados dentro do vetor, nós acessamos a posição onde será feita a substituição e atribuímos um novo dado à posição com `<-`. Por exemplo, para substituir o segundo nome dentro do vetor 'nomes', eu executei `nomes[2] <- "Eduardo"`

Práticas

Uma empresa farmacêutica conduziu um estudo piloto para avaliar um novo medicamento no tratamento da ansiedade. Para isso foram selecionados alguns pacientes voluntários que reportaram pelo menos uma crise de ansiedade aguda (“ataque de pânico”) no último ano. Em um desenho duplo-cego, os pacientes foram aleatoriamente separados em dois grupos, controle e tratamento. O grupo controle recebeu placebo, enquanto que o grupo tratamento recebeu doses padrões do remédio. Após seis meses de acompanhamento, os pacientes foram entrevistados para diagnosticar seu estado psíquico e avaliar possíveis melhoras no quadro de ansiedade. Além disso, outras informações sobre os pacientes foram coletadas para auxiliar na interpretação dos resultados. As informações sobre cada paciente e seu respectivo estado psíquico são:

Idade dos pacientes:

```
idades<-c(25, 18, 45, 53, 61, 33, 29, 42, 37, 72, 59)
```

Massa dos pacientes em quilogramas:

```
massas<-c(89,67,88,74,72,61,60,71,83,63,99)
```

Altura dos pacientes em metros:

```
alturas<-c(1.79,1.62,1.89,1.77,1.84,1.71,1.79,1.84,1.87,1.55,1.96)
```

Grupos nos quais os pacientes estavam; onde *c* é controle, e *t* é tratamento:

```
grupos<-c("c","c","c","c","c","t","t","t","t","t","t")
```

Estado psíquico dos pacientes diagnosticado em entrevistas:

```
estados<-c("tranquilo","ansioso","preocupado","preocupado","ansioso",  
"tranquilo","ansioso","tranquilo","preocupado","tranquilo","preocupado")
```

Os vetores estão ordenados segundo os pacientes, ou seja, dados na mesma posição foram extraídos do mesmo paciente.

Com base nesses dados, faça as seguintes avaliações sobre esse estudo clínico:

- 1) A faixa etária geral dos pacientes pode influenciar o estudo, visto que pessoas muito jovens geralmente possuem menos experiência com situações de ansiedade. Para os fins deste estudo, são muito jovens os pacientes com 25 anos ou menos. Contabilize quantos pacientes são muito jovens.
- 2) A quantidade de massa dos pacientes pode influenciar na dosagem adequada do medicamento. Contudo, a massa absoluta é pouco informativa, pois não considera se a massa está concentrada ou não. Assim, como uma alternativa simplificada, podemos usar o índice de massa corpórea (IMC), que é calculado como a massa dividido pela altura ao quadrado. Calcule o IMC dos pacientes e guarde os resultados em um vetor.
- 3) A dosagem de medicamento que foi aplicada no estudo é excessiva para pacientes com IMC igual ou abaixo de 21. Descubra a posição/ordem dos pacientes que receberam dosagem excessiva de medicamento. [DICA: utilize o operador relacional `<=` e a função `which` para selecionar os pacientes segundo o IMC.]
- 4) Para podermos quantificar o efeito do medicamento, precisamos transformar os dados sobre o estado psíquico dos pacientes em pontuações (*scores*), onde “tranquilo” = 0, “preocupado” = 1 e “ansioso” = 3. Com base nessa escala, crie um vetor de pontuações a partir dos estados psíquicos dos pacientes. [DICA: utilize o operador relacional `==` e a função `which` para selecionar os pacientes segundo seu estado psíquico.]
- 5) Identifique o tipo de dado no vetor de pontuações criado anteriormente. Agora converta os dados do vetor de pontuações para que possamos fazer operações matemáticas [DICA: utilize a função `as.numeric()`]

- 6) O contraste das pontuações dos grupos pode revelar se o medicamento teve algum efeito ou não. Separe a pontuação de cada grupo de pacientes em vetores distintos. Isto é, um vetor com as pontuações do grupo controle, e um outro vetor com as pontuações do grupo tratamento. Calcule a média de pacientes ansiosos em cada grupo. Você percebe alguma diferença?
- 7) O efeito observado pode estar relacionado com a dosagem de medicamento que foi aplicada no estudo. Pacientes que receberam doses excessivas relativas ao IMC tendem a apresentar tranquilidade excessiva. Em que grupo do estudo estão os pacientes que receberam doses excessivas de medicamento?