Missing values

Today we're training how to handle missing values in a data set. Before starting the exercises, please first read section 2.5 of An Introduction to R.

Solutions are available here.

Exercise 1
If X <- c (22,3,7,NA,NA,67) what will be the output for the R statement length(X)

Exercise 2
If X = c(NA,3,14,NA,33,17,NA,41) write some R code that will remove all occurrences of NA in X.
a. X[!is.na(X)]
b. X[is.na(X)]
c. X[X==NA]= 0

Exercise 3
If Y = c(1,3,12,NA,33,7,NA,21) what R statement will replace all occurrences of NA with 11?
a. Y[Y==NA]= 11
b. Y[is.na(Y)]= 11
c. Y[Y==11] = NA

Exercise 4
If X = c(34,33,65,37,89,NA,43,NA,11,NA,23,NA) then what will count the number of occurrences of NA in X?
a. sum(X==NA)
b. sum(X == NA, is.na(X))
c. sum(is.na(X))

Exercise 5
Consider the following vector W <- c (11, 3, 5, NA, 6)
Write some R code that will return TRUE for value of W missing in the vector.

Exercise 6
Load ‘Orange’ dataset from R using the command data(Orange) .
Replace all values of age=118 to NA.

Exercise 7
Consider the following vector A <- c (33, 21, 12, NA, 7, 8) .
Write some R code that will calculate the mean of A without the missing value.

Exercise 8
Let:
c1 <- c(1,2,3,NA) ;
c2 <- c(2,4,6,89) ;
c3 <- c(45,NA,66,101) .
If X <- rbind (c1,c2,c3, deparse.level=1) , write a code that will display all rows with missing values.

Exercise 9
Consider the following data obtained from df <- data.frame (Name = c(NA, “Joseph”, “Martin”, NA, “Andrea”), Sales = c(15, 18, 21, 56, 60), Price = c(34, 52, 21, 44, 20), stringsAsFactors = FALSE)
Write some R code that will return a data frame which removes all rows with NA values in Name column

Exercise 10
Consider the following data obtained from df <- data.frame(Name = c(NA, “Joseph”, “Martin”, NA, “Andrea”), Sales = c(15, 18, 21, NA, 60), Price = c(34, 52, 33, 44, NA), stringsAsFactors = FALSE)
Write some R code that will remove all rows with NA values and give the following output

Name Sales Price
2 Joseph 18 52
3 Martin 21 33

Array exercises

Exercise 1
Create an array (3 dimensional) of 24 elements using the `dim()` function.

Exercise 2
Create an array (3 dimensional) of 24 elements using the `array()` function.

Exercise 3
Assign some dimnames of your choice to the array using the `dimnames()` function.

Exercise 4
Assign some dimnames of your choice to the array using the arguments of the `array()` function.

Exercise 5
Instead of column-major array, make a row-major array (transpose).

Exercise 6
For this exercise, and all that follow, download this file, and read it into R using the `read.csv()` function, e.g.:
temp
Copy the column named `N` into a new variable `arr`.

Exercise 7
Set dimensions of this variable and convert it into a 3 * 2 * 4 array. Add dimnames.

Exercise 8
Print the whole array on the screen.

Exercise 9
Print only elements of height 2, assuming the first dimension represents rows, the second columns and the third heigth.

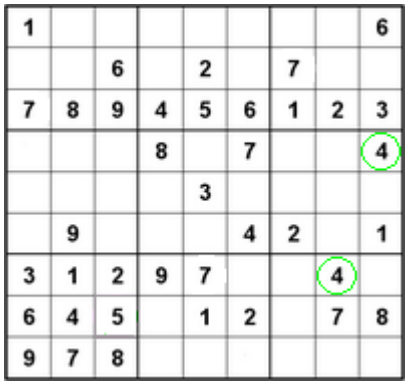
Exercise 10
Print elements of height 1 and columns 3 and 1.

Exercise 11
Print element of height 2, column 4 and row 2.

Exercise 12
Repeat the exercises 9-11, but instead of using numbers to reference row, column and height, use dimnames.

Image: Cubo completato" by Masakazu "Matto" Matsumoto from Nagoya, Japan – http://flickr.com/photos/vitroids/1527092739/. Licensed under CC BY 2.0 via Wikimedia Commons.

Matrix exercises



Please note, solutions are available here.
Exercise 1
Create three vectors x,y,z with integers and each vector has 3 elements. Combine the three vectors to become a 3×3 matrix A where each column represents a vector. Change the row names to a,b,c.

Think: How about each row represents a vector, can you modify your code to implement it?

Exercise 2
Please check your result from Exercise 1, using `is.matrix(A)`. It should return `TRUE`, if your answer is correct. Otherwise, please correct your answer. Hint: Note that `is.matrix()` will return `FALSE` on a non-matrix type of input. Eg: a vector and so on.

Exercise 3
Create a vector with 12 integers. Convert the vector to a 4*3 matrix `B` using `matrix()`. Please change the column names to `x, y, z` and row names to `a, b, c, d`.
The argument `byrow` in `matrix()` is set to be `FALSE` by default. Please change it to `TRUE` and print `B` to see the differences.

Exercise 4
Please obtain the transpose matrix of `B` named `tB` .

Exercise 5
Now `tB` is a 3×4 matrix. By the rule of matrix multiplication in algebra, can we perform `tB*tB` in R language? (Is a 3×4 matrix multiplied by a 3×4 allowed?) What result would we get?

Exercise 6
As we can see from Exercise 5, we were expecting that `tB*tB` would not be allowed because it disobeys the algebra rules. But it actually went through the computation in R. However, as we check the output result , we notice the multiplication with a single `*` operator is performing the componentwise multiplication. It is not the conventional matrix multiplication. How to perform the conventional matrix multiplication in R? Can you compute matrix `A` multiplies `tB` ?

Exercise 7
If we convert `A` to a `data.frame` type instead of a

matrix , can we still compute a conventional matrix multiplication for matrix A multiplies matrix A ? Is there any way we could still perform the matrix multiplication for two data.frame type variables? (Assuming proper dimension)

Exercise 8
Extract a sub-matrix from B named subB . It should be a 3×3 matrix which includes the last three rows of matrix B and their corresponding columns.

Exercise 9
Compute 3*A , A+subB , A-subB . Can we compute A+B? Why?

Exercise 10
Generate a n * n matrix (square matrix) A1 with proper number of random numbers, then generate another n * m matrix A2.
If we have A1*M=A2 (Here * represents the conventional multiplication), please solve for M.
Hint: use the runif() and solve() functions. E.g., runif(9) should give you 9 random numbers.

Want to practice matrices a bit more? We have more exercise sets on this topic here.

Image: "200px-Sudoku06u" by DrBorka from nl. Licensed under CC BY-SA 3.0 via Wikimedia Commons.

Logical vectors and operators

Before you start, enter the following code: data <- mtcars

Solutions are available here.

Exercise 1
Use logical operators to output only those rows of data where column mpg is between 15 and 20 (excluding 15 and 20).

Exercise 2
Use logical operators to output only those rows of data where column cyl is equal to 6 and column am is not 0.

Exercise 3
Use logical operators to output only those rows of data where column gear or carb has the value 4.

Exercise 4
Use logical operators to output only the even rows of data.

Exercise 5
Use logical operators and change every fourth element in column mpg to 0.

Exercise 6
Output only those rows of data where columns vs and am have the same value 1, solve this without using == operator.

Exercise 7
(TRUE + TRUE) * FALSE , what does this expression evaluate to and why?

Exercise 8
Output only those rows of `data` where at least `vs` or `am` have the value 1, solve this without using `==` or `!=`.

Exercise 9
Explain the difference between `|`, `||`, `&` and `&&`.

Exercise 10
Change all values that are 0 in the column `am` in `data` to 2.

Exercise 11
Add 2 to every element in the column `vs` without using numbers.

Exercise 12
Output only those rows of `data` where `vs` and `am` have different values, solve this without using `==` or `!=`.

Regular sequences

Before proceeding, please review Chapter 2 of An Introduction to R and the internal R Documentation for the seq() function (type ?seq in the R console).

Solutions are available here.

Excercise 1

Using the seq() function, generate the sequence 2, 5, 8, 11.

Exercise 2

Use the seq() function to generate the sequence 9, 18, 27, 36, 45.

Exercise 3

Generate the sequence 9, 18, 27, 36, 45, 54, 63, 72, 81, 90 using the length.out parameter.

Exercise 4

For this exercise, first write down your answer, without using R. Then, check your answer using R.

What is the output for the code:

```
seq(from = -10, to = 10, length.out = 5)
```

Exercise 5

Assign value 5 to variable x.

Write code 1:x-1 you should get 0, 1, 2, 3, 4.

Write code 1 : (x-1) you will get 1, 2, 3, 4.

Explain the discrepancy in the output.

Exercise 6

For this exercise, first write down your answer, without using R. Then, check your answer using R.

Create a vector a with values 1, 2, 3, 4

For the code seq(along.with = a), what will be the output?

Exercise 7

For this exercise, first write down your answer, without using R. Then, check your answer using R.

Generate a sequence using the below code.

seq(from=1, to=4, by=1)

What other ways can you generate the same sequence?

a. X <- 1:4

b. X <- seq(4)

c. X <- c(1,2,3,4)

d. All of the above

Exercise 8

Generate a backward sequence from 5, 4, 3, 2, 1

Exercise 9

Assign x <- c(1, 2, 3, 4)

Using the function rep() , create the below sequence

1, 2, 3, 4, 1, 2, 3, 4, 1, 2, 3, 4

Exercise 10

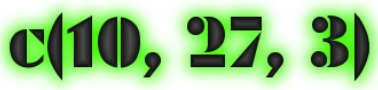
Assign x <- c(1, 2, 3, 4)

Using the rep() function generate the sequence:

1, 1, 1, 2, 2, 2, 3, 3, 3, 4, 4, 4

Want to practice regular sequences a bit more? We have more exercise sets on this topic here.

Vector exercises



A vector is a simple data structure in R. You will use it frequently, often as a building block of more complex data structures and operations on those structures. Before proceeding, please follow our short tutorial and review Chapter 2 of An Introduction to R. First, write down your answer, without using R and without looking at the answer options. Then, match the answer you wrote down with one of the choices given. Finally, check your answer using R.

Solutions are available here.

Exercise 1

Consider a vector:

```
x <- c(4,6,5,7,10,9,4,15)
```

What is the value of:

c(4,6,5,7,10,9,4,15) < 7

a. TRUE, FALSE, TRUE, FALSE, FALSE, FALSE, TRUE, FALSE

b. TRUE, TRUE, TRUE, FALSE, FALSE, FALSE, TRUE, FALSE

c. FALSE, TRUE, TRUE, FALSE, FALSE, FALSE, TRUE, FALSE

d. TRUE, TRUE, TRUE, TRUE, TRUE, FALSE, TRUE, FALSE

e. TRUE, TRUE, TRUE, FALSE, FALSE, FALSE, TRUE, FALSE

Exercise 2

Consider two vectors:

p <- c (3, 5, 6, 8)

and

q <- c (3, 3, 3)

What is the value of:

p+q

a. 6, 8, 6, 8

b. 6, 8, 0, 0

c. 6, 8, NA, NA

d. 3, 5, 6, 8 Warning message: In p+q : longer object length is not a multiple of shorter object length

e. 6, 8, 9, 11

Exercise 3

If:

```
Age <- c(22, 25, 18, 20)
Name <- c("James", "Mathew", "Olivia", "Stella")
Gender <- c("M", "M", "F", "F")
```

then what is the R-code for getting the following output;

```
##   Age   Name Gender
## 1  22  James      M
## 2  25 Mathew      M
```

a.

```
DataFrame = data.frame(c(Age), c(Name), c(Gender))
subset(DataFrame, Gender == "M")
```

b.

```
DataFrame = data.frame(c(Age),c(Name),c(Gender))
subset(Gender=="M"), eval=FALSE
```

c.

```
DataFrame = data.frame(Age,Name,Gender)
subset(DataFrame,Gender=="M")
```

d.

```
DataFrame = data.frame(c(Age,Name,Gender))
subset(DataFrame,Gender=="M")
```

Exercise 4

If

```
z <- 0:9
```

then what is the output from the following R-statements:

```
digits <- as.character(z)
as.integer(digits)
```

a. Error in subset. object 'z' not found

b. 0, 1, 2, 3, 4, 5, 6, 7, 8, 9

c. "NA", "NA", "NA", "NA", "NA", "NA", "NA", "NA", "NA"

d. "0", "1", "2", "3", "4", "5", "6", "7", "8", "9"

e. 0, 0, 0, 0, 0, 0, 0, 0, 0

Learn more about vectors in the online course R Programming A-Z™: R For Data Science With Real Exercises! This course had more than 68,000 students enrolled already and does not require prior knowledge of R.

Exercise 5

Consider the vector:

```
x <- c(1,2,3,4)
```

What is the value of k for:

```
(x+2)[(!is.na(x)) & x > 0] -> k
```

a. 1, 2, 3, 4

b. 1, 4, 9, 16

c. Error: object 'k' not found

d. 3, 4, 5, 6

e. numeric(0)

Exercise 6

Consider the AirPassenger data set

```
data(AirPassengers)
```

Which statement will produce the following output?

```
##  [1] 112 118 132 129 121 135 148 148 136 119 104 118
```

a. `AirPassengers[time(AirPassengers) >= 1949 & time(AirPassengers) < 1950, 12]`

b. `AirPassengers[AirPassengers >= 1949 & AirPassengers < 1950]`

c. `AirPassengers[time(AirPassengers) >= 1949 & time(AirPassengers) < 1950]`

d. `AirPassengers[AirPassengers >= 1949 & AirPassengers < 1950, 12]`

e. `c[[1]]`

Exercise 7

If

```
x <- c(2, 4, 6, 8)
```

and

```
y <- c(TRUE, TRUE, FALSE, TRUE)
```

What is the value of:

```
sum(x[y])
```

a. 20

b. 8

c. 14

d. NA

Exercise 8

Consider the vector:

```
x <- c(34, 56, 55, 87, NA, 4, 77, NA, 21, NA, 39)
```

Which R-statement will count the number of NA values in x?

a. count(is.na(X))

b. length(is.na(x))

c. sum(is.na(x))

d. count(!is.na(x))

e. sum(!is.na(x))

Want to practice vectors a bit more? We have more exercise sets on this topic here.