

Lecture 5 Module 4: The Special Value NA

Exercises

Exercises

Exercise 4.1: Expressions with NA

What are the values of these expressions?

```
(7 <= NA) | (12 - 2 > 5)
```

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

```
(6 > NA) & (NA == 3) & (7 <= -5)
```

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

```
(TRUE & NA) | !(FALSE & TRUE)
```

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

Solution

Exercise 4.2: Working with NA

This is a very important technique.

Here's a vector that contains missing data:

```
exercise.4.2.data <-  
  c(3, 5, 7, NA, 4, 9, NA, 3)
```

I want to answer three questions about this vector:

- Does the vector contain any missing data?
- How many missing elements does the vector contain?
- What are the locations of the missing data?

Hint: use logical functions and the function `is.na()`.

Answer

Exercise 4.3: Summary statistics

Recall the vector from exercise 4.2:

```
exercise.4.2.data <-  
  c(3, 5, 7, NA, 4, 9, NA, 3)
```

Calculate the sum of the non-missing values of this vector.

Then calculate the sample mean of the non-missing values of this vector.

Solution

Solutions to the Exercises

Exercise 4.1: Expressions with NA

What are the values of these expressions?

```
(7 <= NA) | (12 - 2 > 5)
```

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

```
(6 > NA) & (NA == 3) & (7 <= -5)
```

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

```
(TRUE & NA) | !(FALSE & TRUE)
```

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

Solution

Exercise 4.2: Working with NA

This is a very important technique.

Here's a vector that contains missing data:

```
exercise.4.2.data <-  
  c(3, 5, 7, NA, 4, 9, NA, 3)
```

I want to answer three questions about this vector:

- Does the vector contain any missing data?
- How many missing elements does the vector contain?
- What are the locations of the missing data?

Hint: use logical functions and the function `is.na()`.

Answer

To determine if the vector contains missing data, we have:

```
any( is.na( exercise.4.2.data ) )
```

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

To count the number of missing elements, we have:

```
sum( is.na( exercise.4.2.data ) )
```

```
## [1] 2
```

The locations of the missing elements are:

```
which( is.na( exercise.4.2.data ) )
```

```
## [1] 4 7
```

Exercise 4.3: Summary statistics

Recall the vector from problem 4.2:

```
exercise.4.2.data <-  
  c(3, 5, 7, NA, 4, 9, NA, 3)
```

Calculate the sum of the non-missing values of this vector.

Then calculate the sample mean of the non-missing values of this vector.

Solution

The sum of the non-missing values is:

```
sum( exercise.4.2.data, na.rm = TRUE )
```

```
## [1] 31
```

The sample mean of the non-missing values is:

```
mean( exercise.4.2.data, na.rm = TRUE )
```

```
## [1] 5.166667
```