

## 제2주 : 자료형과 함수

R 언어를 잘 활용하려면, 기본 자료형과 자료구조, 그리고 연산작업하는 방법에 관한 정확한 이해가 요구됩니다. 본 장에서는 자료(값이나 변수 등)가 저장되는 형식인 자료형(data type)의 종류를 설명합니다.

### 2.1. R의 기본 자료형

아래와 같이 5가지 기본 자료형이 있습니다. 기본 자료형이란, 더 이상 쪼갤 수 없는 원자수준의 자료형입니다.

자료형	예	설명
정수 (integer)	2L	L부호는 정수형으로 데이터를 저장하도록 R에게 알려준다
실수 (numeric; double)	2, 15.5	
복소수 (complex)	1+4i, 1i	a+bi형태로, 허수부 숫자에 i를 붙여서 복소수임을 나타낸다.
문자 (character)	"a", "+rw", "disk drive"	홑 따옴표('...')나 쌍따옴표("...")로 묶어준다.
논리형 (logical)	TRUE, FALSE	

아래 값들은 특수한 상태를 나타내는 상수들이다. NULL을 제외하고는 각기 해당하는 기본 자료형을 갖고있지만 크게 신경쓰지 않는다.

- NULL : 비어있는 값, 정의되지 않은 값
- NA : 결측값 (Not Available). 자료형이 경우에 따라 다름.
- -Inf, Inf : 음과 양의 무한대. 실수형을 갖는다.
- NaN : 수의 연산에서 불능인 경우를 표현. 0/0, Inf/Inf 등 (Not a Number)

## 각 기본 자료형의 사용 예

```
> x <- 3
> y <- 2
> x / y
[1] 1.5
> xi <- 1 + 2i
> yi <- 1 - 2i
> xi + yi
[1] 2+0i
> str <- "string"
> str
[1] "string"
> TRUE
[1] TRUE
> T
[1] TRUE
> FALSE
[1] FALSE
> F
[1] FALSE
> xinf <- Inf
> yinf <- -Inf
> xinf / yinf
[1] NaN
```

## 2.2. 자료형을 위한 R 내장 함수

값, 변수의 자료형과 관련하여 다음과 같은 R 내장 함수를 갖추고 있습니다. 자료의 자료형에 대한 판단과 변환 등을 통해 필요한 작업을 원활히 할 수 있도록 도와줍니다.

### ■ 자료형 정보를 보여주는 함수:

아래 함수들은 비슷한 결과를 가져오지만 각각 차이가 있습니다.

- `class()` : R 객체지향 관점에서의 자료형
- `typeof()` : 원시 자료형 (R 에서의)

- `mode()` : 원시 자료형 (S 에서의<sup>1)</sup>)
- 자료형 확인 함수 (`is.“RClass명”()`)
  - `is.integer(x)` : x 가 정수형이면 TRUE, 아니면 FALSE
  - `is.numeric(x)` : x가 실수형이면 TRUE, 아니면 FALSE
  - `is.complex(x)` : x 가 복소수형이면 TRUE, 아니면 FALSE
  - `is.character(x)` : x 가 문자열이면 TRUE, 아니면 FALSE
  - `is.logical(x)`
  - `is.na(x)` : x 가 NA 이면 TRUE, 아니면 FALSE
    - \* `is.na(NaN)`을 입력하여 결과를 확인해보시오.
    - \* `is.nan(NA)`를 입력하여 결과를 확인해보시오.
- `object.size(x)` : R 변수x를 저장하는 데 쓰이는 메모리 추정치를 확인

## 2.3. 함수 (Function, method)

변수는 자료를 저장하는 역할을 한다면 함수는 행위를 나타냅니다. 예를 들어 “1부터 10 까지의 합을 구하라”는 행위를 R에게 시키기 위해 합을 구하는 행위에 맞는 함수의 이름을 표시하고 합을 구하기 위해 필요한 자료를 전달해 주어야 합니다. 이를 위해 R은

**함수명 (필요로 하는 데이터 , ...)**

의 구조를 통해 원하는 행위의 결과를 얻습니다.

다음은 1부터 10 까지 합을 구하는 함수의 사용예제입니다.

```
> sum(c(1, 2, 3, 4, 5, 6, 7, 8, 9, 10))
[1] 55
```

함수 `sum` 은 주어진 값들의 합을 구하는 함수인데 위에서 보면 `sum` 함수에 전달하는 필요로 하는 데이터 부분에 또 다른 함수 `c()`가 사용

1) “S programming language” by Becker, Chambers & Wilks (1998)에서 분류한 자료형으로, S는 R이 나오기 이전 상업용 Package로서 R에게 모티브를 제공한 S-PLUS에서 사용하는 언어입니다.

했습니다. 이 함수의 역할은 R의 자료구조중 벡터를 만드는 역할을 하는 함수로 필요로 하는 데이터를 각각 콤마 (,)로 구별했음을 볼 수 있습니다. 함수에서 필요로 하는 데이터는 전달인자(argument)라고 부르며 전달인자가 여러개일 경우 콤마(,)로 구별합니다.

다음 함수는 앞서 본 ls 함수로 R이 현재 갖고 있는 자료들의 list(줄여서 줄여서 ls)들을 출력하는 함수입니다.

```
> ls()  
[1] "age"          "character"    "characters"    .....
```

어떤 함수들은 위와 같이 전달인자가 없을 수도 있습니다. 이 경우에는 그냥 소괄호를 열고 닫으면 됩니다.

그리고 어떤 전달인자들은 입력하지 않으면 해당 함수들의 기본값으로 처리할 경우가 있습니다. 다음의 예를 보겠습니다.

```
> seq(1, 10)  
[1] 1 2 3 4 5 6 7 8 9 10  
> seq(1, 10, by=2)
```

seq 함수는 순열 (sequence)를 만드는 함수로 처음의 예제처럼 by 전달인자가 생략될 경우 1씩 증가/감소하는 순열을 만들지만 사용자의 의해 by 전달인자가 지정되면 주어진 값만큼 증가/감소하면서 순열을 만들어 줍니다.

## 2.4. 도움말 함수 help()

R에는 사용할 수 있는 많은 함수와 자원이 있습니다(패키지, 데이터, ...). 또한 외부 Package를 통해 계속해서 증가하는 다양한 자원들을 모두 알기는 힘듭니다. 이럴 경우 사용자들에게 기본적인 도움말을 help() 함수를 통해 제공해 줍니다. 다음은 help()함수의 사용예입니다.

> help(c)

➔ 함수 c() 의 도움말을 기본 브라우저를 통해 보여줍니다.

> ?c

➔ help() 함수의 단축키는 ?로 물음표 (?) 뒤에 도움말을 얻을 함수 이름을 적습니다.

> help("<")

> help("for")

➔ R의 연산자나 문법에 사용되는 예약어(keyword) 등은 큰 따옴표로 묶어 질의합니다.

> help(package="datasets")

➔ 패키지에 대한 도움말은 위와같이 package="패키지명"으로 합니다.

> help.search("Normal Distribution")

➔ 찾고자 하는 대상을 명확히 모를 경우 검색어를 통한 검색을 실시할 수 있습니다.

> ?? "Normal Distribution"

➔ 검색어를 통한 검색의 검색의 단축키는 ?? ( 물음표 두개 )입니다.

> example(mean)

mean> x <- c(0:10, 50)

mean> xm <- mean(x)

mean> c(xm, mean(x, trim = 0.10))

[1] 8.75 5.50

➔ 함수의 사용예제를 제공합니다.