

2019년 3월 정화민 교수



## R 기초문법 (수치연산)

연산자와 함수	의미
<code>+, -, *, /</code>	사칙 연산
<code>n %% m</code>	n을 m으로 나눈 나머지
<code>n %/% m</code>	n을 m으로 나눈 몫
<code>n^m</code>	n의 m승
<code>exp(n)</code>	e의 n승
<code>log(x, base=exp(1))</code>	logbase(x). 만약 base가 지정되지 않으면 loge(x)를 계산
<code>log2(x), log10(x)</code>	각각 log2(x), log10(x)를 계산
<code>sin(x), cos(x), tan(x)</code>	삼각 함수

Source: R을 이용한 데이터처리분석 실무

# R 수치연산, 벡터연산

```
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
ited3 x PCA(주성분분석).R x Descriptive_Statistics.R x correlation.R x paired_Ttest.R x decisionTree.R x 수치연산.R x 벡터연산.R x
Source on Save Run Source
1 # 수치연산 예: 2에서 4까지 각각 3을 곱하고 1을 더한다
2 2:4 * 3+1
3
4 # 벡터 연산 예: 벡터 또는 리스트를 한번에 연산함
5 x <-c(2, 3, 4)
6 x + 2
7 |

7:1 (Top Level) R Script
Console D:/Rdata/
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

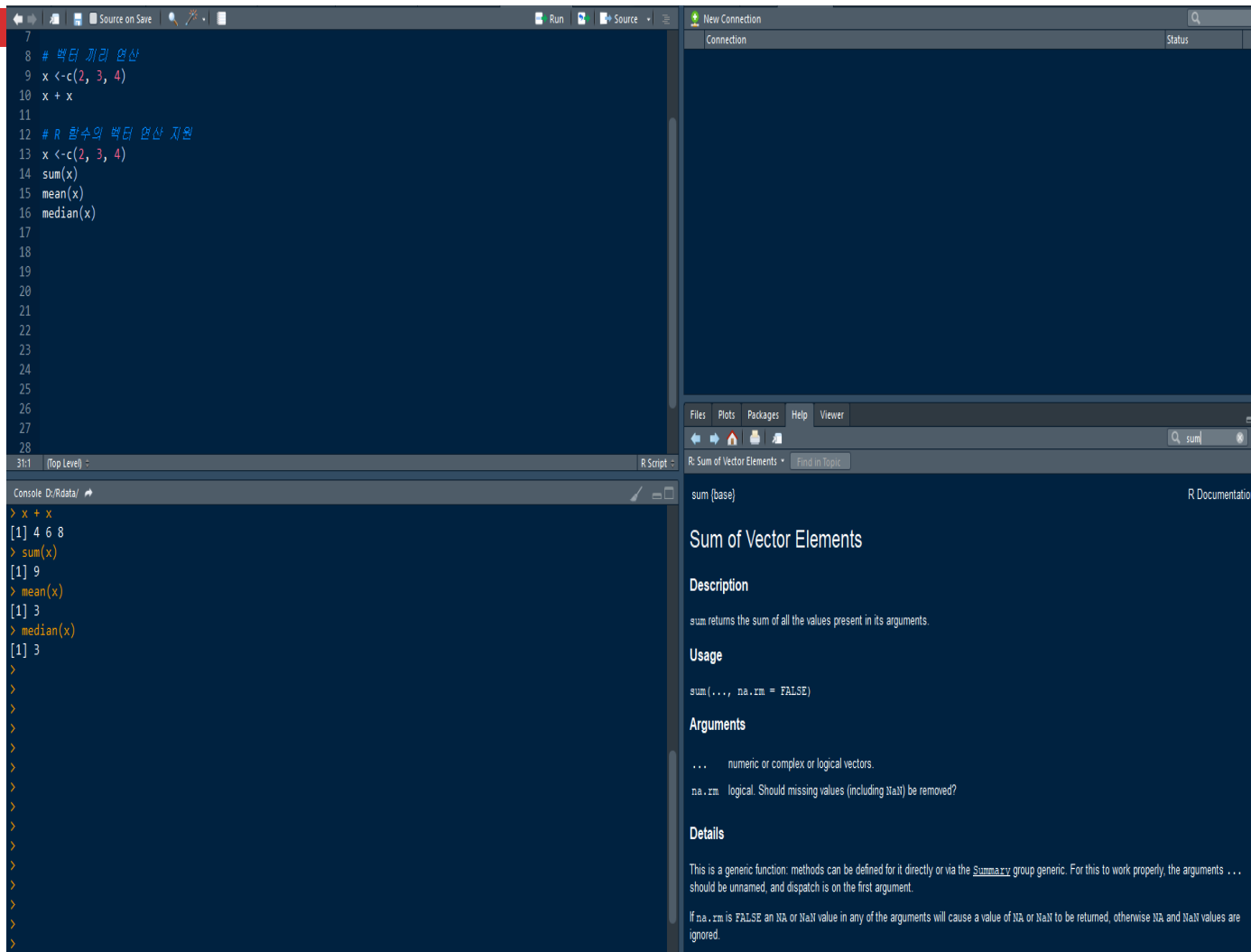
[Workspace loaded from D:/Rdata/.RData]

> 2:4 * 3+1
[1] 7 10 13
> x <-c(2, 3, 4)
> x + 2
Error: object 'X' not found
> # 벡터 연산 예: 벡터 또는 리스트를 한번에 연산함
> x <-c(2, 3, 4)
> x + 2
Error: object 'X' not found
> x + 2
[1] 4 5 6
> |
```

# 수치연산 예:  
2에서 4까지 각각  
3을 곱하고 1을  
더한다  
 $2:4 * 3+1$

# 벡터 연산 예:  
벡터 또는 리스트를  
한번에 연산함  
 $x <-c(2, 3, 4)$   
 $x + 2$   
(대소문자 구분  
안하면 에러)

# R 벡터끼리 연산, R 함수의 벡터연산 지원



The screenshot shows the RStudio interface. The top-left pane contains R code with comments in Korean. The bottom-left pane shows the console output. The right pane shows the documentation for the `sum` function.

```
# 벡터 끼리 연산
x <- c(2, 3, 4)
x + x

# R 함수의 벡터 연산 지원
x <- c(2, 3, 4)
sum(x)
mean(x)
median(x)
```

```
> x + x
[1] 4 6 8
> sum(x)
[1] 9
> mean(x)
[1] 3
> median(x)
[1] 3
```

**Sum of Vector Elements**

**Description**

`sum` returns the sum of all the values present in its arguments.

**Usage**

```
sum(..., na.rm = FALSE)
```

**Arguments**

- `...` numeric or complex or logical vectors.
- `na.rm` logical. Should missing values (including `NA`) be removed?

**Details**

This is a generic function: methods can be defined for it directly or via the `Summary` group generic. For this to work properly, the arguments ... should be unnamed, and dispatch is on the first argument.

If `na.rm` is `FALSE` an `NA` or `NaN` value in any of the arguments will cause a value of `NA` or `NaN` to be returned, otherwise `NA` and `NaN` values are ignored.

# 벡터 끼리 연산

x <-c(2, 3, 4)

x + x

# R 함수의 벡터  
연산 지원

x <-c(2, 3, 4)

sum(x)

mean(x)

median(x)

지식채널

money

키 크기는  
그들의 소득에 비례한다

The heights are proportional  
to the income.

## R 함수(R 데이터 mtcars)

# 함수는 특정한 작업을 수행하기 위해 일련의 구문들을 체계적으로 묶은 것, R은 수많은 내장 함수를 가지고 있음.

# 사용 Data : R 내장 데이터 mtcars:

1974년 미국 Motor trend US magazine에 나오는 data이다.  
32개의 차량에 대해서 각 자료들을 기재되어 있음

mpg - Miles/gallon (연비, 1갤런당 몇 마일을 가는가)

cyl - Number of cylinders (차량 엔진의 실린더의 개수, 펌프같이 움직이는 것)

disp - Displacement (배기량)

hp - Gross horsepower (마력)

drat - Rear axle ratio (후방 축 비율)

wt - Weight (1000lbs) 파운드 기준 차량무게

qseq - 1/4 mile time 1/4 마일 간 시간?

vs - V/S ??? versus??

am - Transmission( 0 = automatic, 1 = manual) 변속기가 자동이냐 아니냐

gear - Number of forward gears 전진기어의 수? (1,2,3 )

carb - Number of carburetors 카뷰레이터 수 (기화기수)

## mtcars 중앙값, 최빈값(prettyR 패키지 활용)

# mtcars 를 활용한 평균, 분산, 표준편차 함수  
# R에 기본으로 내장된 데이터 셋은 data() 함수를 통해 쉽게 불러올 수 있다.

```
data("mtcars")
```

```
head(mtcars)
```

# attach() 함수는 R 검색경로에 데이터프레임을 추가

```
attach(mtcars)
```

# 단순평균함수

```
mean(mpg)
```

# 중앙값

```
median(mpg)
```

#최빈값(R은 빌트인 함수로는 mode를 제공하지 않는다. prettyR패키지 설치하여 Mode함수 사용하기)

```
install.packages("prettyR")
```

```
library(prettyR)
```

```
← → Source on Save 🔍 🛠️ 📄 Run 📄 Source 📄  
1 # mtcars 를 활용한 평균, 분산, 표준편차 함수  
2 # R에 기본으로 내장된 데이터 셋은 data() 함수를 통해 쉽게 불러올 수 있다.  
3 data("mtcars")  
4 head(mtcars)  
5 # attach() 함수는 R 검색경로에 데이터프레임을 추가  
6 attach(mtcars)  
7 # 단순평균 함수  
8 mean(mpg)  
9 # 중앙값  
10 median(mpg)  
11 # 최빈값(R은 빌트인 함수로는 mode를 제공하지 않는다. prettyR 패키지 설치하여 Mode 함수 사용하기  
12 install.packages("prettyR")  
13 library(prettyR)  
14 # 실린더 갯수 전체 데이터셋에서 4, 6, 8개만 존재한다. table 함수로 실린더별 데이터 확인  
15 table(cyl)  
16 # 실린더 최빈값  
17 Mode(mtcars$cyl)  
18 # 분산 var() 함수 사용  
19 var(mpg)  
20 # 표준편차 sd()  
21 sd(mpg)  
22 # 최대값과 최소값을 구한다. range() 함수  
23 range(mpg)  
24 # range 차이, max()  
25 max(mpg)-min(mpg)
```



## mtcars 최빈값, 분산, 표준편차, Range, max, min, IRQ

# 실린더 갯수전체 데이터셋에서 4, 6, 8개만 존재한다. table 함수로 실린더별 데이터 확인

```
table(cyl)
```

#실린더 최빈값

```
Mode(mtcars$cyl)
```

#분산 var() 함수 사용

```
var(mpg)
```

# 표준편차 sd()

```
sd(mpg)
```

# 최대값과 최소값을 구한다. range()함수

```
range(mpg)
```

# range 차이, max()

```
max(mpg)-min(mpg)
```

#Interquartile Range (IQR): IQR은 4분위값(quantile 4개) = 3번째 분위값 - 1번째 분위값을 사용하는 것으로中间的 50%값을 보기위해 사용한다.

```
IQR(mtcars$mpg)
```

```

26 #Interquartile Range (IQR): IQR은 4분위값(quantile 4개) = 3번째 분위값 - 1번째 분위값을 사용하는 것으로
    중간의 50%값을 보기위해 사용한다.
27 IQR(mtcars$mpg)
28 # 공분산(Covariance): 공분산(covariance)은 2개의 확률변수의 상관정도를 나타내는 값이다.(1개의 변수의
    이산정도를 나타내는 분산과는 별개임) 만약 2개의 변수중 하나의 값이 상승하는 경향을 보일 때, 다른 값도
    상승하는 경향의 상관관계에 있다면, 공분산의 값은 양수가 될 것이다. 반대로 2개의 변수중 하나의 값이
    상승하는 경향을 보일 때, 다른 값이 하강하는 경향을 보인다면 공분산의 값은 음수가 된다. IRQ 함수.
29 cov(wt, mpg)
30 # 상관분석(Correlation Analysis):상관분석(Correlation Analysis)은 확률론과 통계학에서 두 변수간에 어떤
    선형적 관계를 갖고 있는 지를 분석하는 방법이다. 두변수는 서로 독립적인 관계로부터 서로 상관된 관계일 수
    있으며 이때 두 변수간의 관계의 강도를 상관관계(Correlation)라 한다. cov() 함수
31 cor(wt, mpg)
32 # summary 함수: summary() 함수를 통해서 아래의 지표들을 구할 수 있다. minimum, maximum, quartiles, the
    mean for numerical variables, frequencies
33 variables <- c("mpg", "hp", "wt")
34 summary(mtcars[variables])
35 #describe 함수: psych 패키지의 describe() 함수를 사용하면 아래의 지표들을 한번에 구할 수 있다. 표본수
    ,mean, standard deviation, median, trimmed mean, median absolute deviation, minimum, maximum, range,
    skew, kurtosis, standard error of the mean
36 install.packages("psych")
37 library(psych)
38 describe(mtcars[variables])

```

## mtcars 공분산, 상관분석, summary, describe

# 공분산(Covariance): 공분산(covariance)은 2개의 확률변수의 상관정도를 나타내는 값이다.(1개의 변수의 이산정도를 나타내는 분산과는 별개임) 만약 2개의 변수중 하나의 값이 상승하는 경향을 보일 때, 다른 값도 상승하는 경향의 상관관계에 있다면, 공분산의 값은 양수가 될 것이다. 반대로 2개의 변수중 하나의 값이 상승하는 경향을 보일 때, 다른 값이 하강하는 경향을 보인다면 공분산의 값은 음수가 된다. IRQ 함수.

```
cov(wt, mpg)
```

# 상관분석(Correlation Analysis):상관분석(Correlation Analysis)은 확률론과 통계학에서 두 변수간에 어떤 선형적 관계를 갖고 있는 지를 분석하는 방법이다. 두변수는 서로 독립적인 관계로부터 서로 상관된 관계일 수 있으며 이때 두 변수간의 관계의 강도를 상관관계(Correlation)라 한다. cov()함수

```
cor(wt, mpg)수
```

# summary 함수: summary() 함수를 통해서 아래의 지표들을 구할 수 있다. minimum, maximum, quartiles, the mean for numerical variables, frequencies

```
variables <- c("mpg", "hp", "wt")
```

```
summary(mtcars[variables])
```

#describe 함수: psych 패키지의 describe()함수를 사용하면 아래의 지표들을 한번에 구할 수 있다. 표본수,mean, standard deviation, median, trimmed mean, median absolute deviation, minimum, maximum, range, skew, kurtosis, standard error of the mean

```
install.packages("psych")
```

```
library(psych)
```

```
describe(mtcars[variables])
```

공식을 활용한 분산, 표준편차 구하기

---

강의시간에 실습