

### 3 - 데이터 탐색

#### 이 - R 기초

##### 1. R 기본 문법

###### (1) R data type

- 문자형
- 숫자형
- 논리형
- NaN / NA / NULL

###### (2) 기본 문법

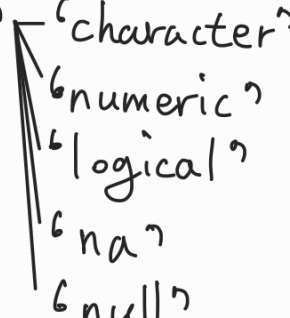
###### - 대입 연산자

: ' $<-$ ', ' $<<-$ ', ' $=$ ' (오른쪽 값을 왼쪽)

' $\rightarrow$ ', ' $\rightarrow>$ ' (왼쪽 값을 오른쪽)

###### - 비교 연산자

: ' $==$ ', ' $<$ '/' $>$ ', ' $!=$ ', ' $<=$ '/' $>=$ '

'is, ~' 

★ NA는 NA 변환

###### - 산술 연산자

: ' $\wedge$ '/' $**$ ' (power),  $\exp()$

### (3) Data Struc.

#### - Vector

: type 같은 여러 data를

하나의 row로 저장하는 1차원 data str.

'c(값1, 값2, ...)' 로 씀

#### - Matrix

: 2차원 구조 가진 vector.

mat.에 저장된 모든 data는 같은 type!

'matrix(name, nrow, ncol)' 으로 mat. 생성.

• dim(행렬)

: row의 개수, column의 개수 반환

#### - Array (배열)

: 3차원 이상의 구조를 가진 vector

'array(name, dim)' 로 생성

#### - List

: data의 type/struc. 상관없이

서로 다른 오브젝트 저장 가능.

#### - Data Frame

: data 분석을 위한

2차원 구조 갖는 관계형 data struc.

여러 개의 vector로 구성. 각 column은 서로 다른 type data

#### (4) 통계함수

- mean()

: 평균

- median()

: 중앙값

- var()

: 표본 분산

- sd()

: 표본 표준편차

- range()

: 최댓값, 최솟값

- skewness()

: 왜도

- kurtosis()

: 첨수

) "Basics" package 설치 필요

#### (5) 제어문

- 반복문

: for, while

- 조건문

: if, else if, else

- 사용자 정의 func. ex) func-name <- function(~){

- comment

: #으로.

⋮  
}

## (6) 자주 사용되는 R 함수

### - 숫자 연산

: `sqrt`, `abs`, `exp`, `pi`, `round`, ...

### - 문자 연산

: `tolower`, `toupper`, `nchar(string length)`,  
`substr(extract str)`, `strsplit`,  
`grep(str.에 주어진 char 있는지)`,  
`gsub(str. 일부분 대체)`

### - vector 연산

: `length`, `paste` (두 vector를 구분자 기준 결합),  
`cov` (두 수치 vector의 covar-),  
`cor` (, , '의 cor-),  
`table(data의 개수)`, `order(vector의 order)`

### - matrix 연산

: `t(transpose)`, `diag(diagonal)`, `%*%` (두 mat.의 곱)

### - data 탐색

: `head`, `tail`, `quantile` (4분위수)

### - data 처리

: `subset` (조건식 맞는 data 추출),  
`merge` (두 data를 특정 common row 기준으로 merge),  
`apply` (data에 row/column 별로  
주어진 func. 적용)

## - Normal Distribution

:  $\text{mean}=0$ ,  $\text{sd}=1$  인 건  
↳ Standard N ~ D

### • dnorm

: 함수 값

### • rnorm

: 주어진 개수만큼 표본 추출

### • pnorm

: 주어진 값보다 작은 확률

### • qnorm

: 주어진 넓이 값에 맞는 값

## - 표본 추출

### • runif

: Uniform Dist-에서

주어진 개수만큼 표본 추출

### • sample

: 주어진 data에서

주어진 개수만큼 "

## - 날짜

### • as.POSIXct

: timestamp을 날짜·시간으로.

## - 산점도

### • plot

• abline (추가 직선)

