

R 통계

2025-09-16

오늘 목표

- R 기초 문법 이어서
- 기초 통계 with R
 - 이론
 - 실습

R

- List
- Data frame
- Control flow
- Loop
- Apply
- Save / Load

List

- vector는 항상 동일한 타입만 저장
- 달라야 하는 경우?
- `list()`로 생성
- 다른 프로그램과의 작업
 - 여러개의 엑셀 파일, DB, JSON 등

List

```
1 x <- list(  
2   a = 'Character',  
3   b = 123.456,  
4   c = NULL,  
5   e = TRUE  
6 )
```

List indexing

```
1 x[["a"]] # 따옴표  
2 x$a  
3 x[1]  
4 x[[1]] # [[ 2개 ]]
```

Data frame

- 1줄이 아닌, 행과 열을 가지는 표 데이터
 - 이전에는 `array`, `matrix`
- `df`로 표기하는 경우가 많은데, 비권장 (`df()`)
- 각 열은 벡터 -> 열 단위 벡터 작업

Data frame

- A: 20살, 85.5점
- B: 21살, 90.0점
- C: 22살, 78.5점
- D: 21살, 95.0점

을 데이터로 어떻게??

Data frame

```
1 students <- data.frame(  
2   name = c("A", "B", "C", "D"),  
3   age = c(20, 21, 22, 21),  
4   score = c(85.5, 90.0, 78.5, 95.0)  
5 )
```

name	age	score
A	20	85.5
B	21	90.0
C	22	78.5
D	21	95.0

Data frame (Tidy)

```
1 students_tidy <- data.frame(  
2   name = c('A','B'), # C, D...  
3   variable = c('age','score', 'age', 'score'),  
4   value = c(20, 85.5, 21, 90.0)  
5 )
```

name	variable	value
A	age	20
A	score	85.5
B	age	21
B	score	90.0

Data frame indexing

```
1 # named vector
2 students$name
3 students['name']
4 students[1]
5
6 # unnamed vector
7 students[['name']]
8 students[[1]]
9
10 # value (비권장)
11 students[1, 2]
12 students[2, 1]
```

Control Flow

- 프로그램을 조건(TRUE / FALSE, &&, ||)에 따라 실행
 - &, | -> 벡터 단위 비교
- if & else
- %in%

```
1 if (condition){  
2     CODE  
3 } else { # 줄 넘김 주의  
4     CODE2  
5 }
```

Control Flow

```
1 if(BMI >= 25){  
2     print('과체중')  
3 } else if(BMI >= 30){  
4     print('비만')  
5 }
```

```
1 if(BMI >= 25){  
2     print('과체중')  
3 }  
4 if(BMI >= 30){  
5     print('비만')  
6 }
```

- BMI ≥ 25 kg/m² : 과체중
- BMI ≥ 30 kg/m² : 비만

Control Flow

```
1 x <- 1
2 if (3 == 3) {
3     x <- 2
4 } else {
5     x <- 3
6 }
7 print(x)
```

- ifelse

Control Flow (ifelse)

```
1 # ifelse(Cond, True, False)
2 x <- ifelse(3 == 3, 2, 3)
3 print(x)
```

Loop

- 조건에 따라 코드를 반복 실행
- `while`, `continue`은 자주 안 쓰임

```
1 for(i in 1:10){  
2     CODE  
3     if(BREAK){break}  
4 }
```


Loop

- `in` 은 뒤의 내용을 앞의 변수가 반복하게 함

```
1 for(i in students){  
2     print(i)  
3 }
```

Apply

- 벡터화 된 `loop`
- 대부분 `for` 보다 빠름
- 간결한 사용법
- 결과를 벡터로 저장
- `sapply`, `lapply`, `apply`

```
1 v <- sapply(1:5, function(i){i^2})  
2  
3 v2 <- c()  
4 for(i in 1:5){  
5     v2[i] <- i^2  
6 }
```

Save / Load

- .Rdata, .RDS

```
1 getwd() # 현재 작업 위치 = 파일 저장 위치
2
3 save(students, file = 'students.RData')
4 rm(students) # REMOVE
5 load('student.Rdata')
6
7 saveRDS(students, 'students.RDS')
8 ss <- readRDS('students.RDS') # 오브젝트 이름 변경
```

Q&A?

Assignments

Score	Grade	Score	Grade
95	A+	90	A
85	B+	80	B
75	C+	70	C
65	D+	60	D
60 미만	F		

```

1 # grades.R
2 set.seed(916)
3 test = data.frame(
4   name = letters,
5   score = round(runif(26, 0, 100)),
6   grade = NA
7 )
8 # test$grade

```