

04 행렬, 리스트, 팩터

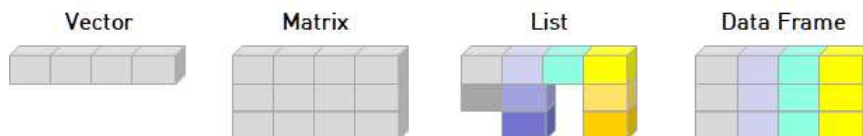
학습목표

1. 행렬의 특징을 이해하며, 행렬을 생성하고 원소에 접근할 수 있다.
2. 리스트의 특징을 이해하며, 리스트를 생성하고 원소에 접근할 수 있다.
3. 팩터의 특징을 이해하며, 팩터 자료형을 만들 수 있다.

강의내용

- R 자료구조 - 단일값들로 구성된 자료의 모음
- 벡터(vector), 행렬(matrix), 리스트(list), 데이터프레임(dataframe) 등

| | |
|-------------------|--|
| 벡터(vector) | <ul style="list-style-type: none"> ◦ 하나 이상의 원소로 이루어진 1차원 구조, R에서 가장 기본이 되는 자료구조 ◦ 동일한 자료형으로 구성되어야 함 |
| 행렬(matrix) | <ul style="list-style-type: none"> ◦ 행과 열로 구성된 2차원 벡터 |
| 리스트(list) | <ul style="list-style-type: none"> ◦ 다양한 자료형을 가질 수 있는 자료구조, 벡터의 확장형 |
| 데이터프레임(dataframe) | <ul style="list-style-type: none"> ◦ 데이터 분석에서 가장 많이 사용하는 테이블 형태의 2차원 자료구조 ◦ 각 열마다 다른 자료형을 가질 수 있으나 하나의 열은 동일한 자료형으로 구성 |



행렬(matrix) 생성 및 행과 열에 이름 붙이기

| | |
|----------------|---|
| matrix() 함수 이용 | <ul style="list-style-type: none"> - 원하는 값을 입력하여 행렬 생성 - matrix(vec, nrow, ncol, byrow=FALSE) |
| 기존 벡터 이용 | <ul style="list-style-type: none"> - cbind(), rbind() 함수를 사용하여 기존 벡터를 행렬로 변환 |
| 행과 열에 이름 부여 | <ul style="list-style-type: none"> - colnames(), rownames() 함수 이용 |

행렬 내 원소에 접근

- 행과 열의 인덱스로 접근
- 행과 열의 이름으로 접근

```
> score
      [,1] [,2] [,3] [,4]
[1,]  70  75  80  85
[2,]  71  76  81  86
[3,]  72  77  82  87
[4,]  73  78  83  88
[5,]  74  79  84  89
```

```
# matrix() 함수 이용하여 5행 행렬 생성
mv1 <- 1:20                                # 벡터 mv1 생성(자료 개수 20개)
mtx1 <- matrix(mv1, nrow = 5)              # 데이터 열 방향 우선 배치
mtx2 <- matrix(mv1, nrow = 5, byrow=T)    # 데이터 행 방향 우선 배

# rbind(), cbind() 함수로 기존 벡터를 결합
v1 <- 80:84; v2 <- 95:99                    # 벡터 v1, v2 생성(자료 개수 5개)
mtx3 <- cbind(v1, v2)                      # 열 방향으로 결합
mtx4 <- rbind(v1, v2)                     # 행 방향으로 결합

# 행과 열에 이름 붙이기
score <- matrix(70:89, nrow=5)
colnames(score) <- c('kor', 'eng', 'com', 'gym')
rownames(score) <- c('학생1', '학생2', '학생3', '학생4', '학생5')

# 행렬 내의 원소에 접근
score[1:3, c(1,3)]                        # 1~3행의 1열과 3열 자료
score[1:2,]                                # 1행과 2행의 자료
score[, -3]                                # 3열을 제외한 모든 자료
score[, "gym"]                             # "gym"열의 모든 자료
```

04 행렬, 리스트, 팩터

```
# 자료의 정보 확인
class(mtx)      # 자료형
length(mtx)     # 자료의 갯수
nrow(mtx)       # 행의 수
ncol(mtx)       # 열의 수
dim(mtx)        # 자료의 차원 (행, 열)
str(mtx)        # 자료의 구조 정보
```

리스트(list) 생성 및 이름 붙이기

- list() 함수 이용하여 다양한 자료형을 갖는 리스트 생성
- 벡터, 데이터프레임, 리스트 등 어떤 자료형도 원소로 포함될 수 있다.
- 벡터와 동일한 방법으로 리스트 내 원소에 이름 부여

리스트 내 원소에 접근

- 벡터와 유사하게 대괄호([]) 안에 리스트 내부 원소의 인덱스나 이름을 입력하여 접근
- 리스트의 원소는 크기가 일정하지 않기 때문에, 리스트 원소의 자료형은 기본적으로 리스트가 된다.
- 리스트의 개별 원소에 접근하려면 이중괄호([[]])를 사용하거나 \$ 기호를 사용

```
# list() 함수 이용하여 다양한 자료형을 갖는 리스트 생성
lt <- list(title = "리스트", num = 1:10, lst=list(1,2)) # 원소가 3개인 리스트 lt 생성

# 리스트 내의 원소에 접근, class()함수로 자료형 확인
lt[c(1,2)]      # 대괄호의 결과는 항상 리스트
lt["title"]     # 이름이 "title"인 원소
lt[1]           # 리스트 lt의 첫 번째 원소

# 리스트 내의 개별 원소에 접근
lt[[2]]         # 리스트 lt의 두 번째 개별 원소
lt$num          # 이름이 있는 경우 개별 원소에 접근 : $ 사용

# 함수사용
sum(lt[2])      # 결과는 ?
sum(lt$num)
```

팩터형(Factor)

- 범주형(Categorical) 데이터를 저장하는 자료형, 혈액형, 성별, 학점등급 등을 표현할 때 사용
- 명목형(크기 비교할 수 없는 데이터) : 혈액형(A,B,O, AB), 성별(여성, 남성)
- 순서형(순서 부여할 수 있는 데이터) : 성적등급(A,B,C,D,F), 크기(대,중,소)
- factor()함수로 생성, levels의 범위 이외의 값은 입력이 불가, order=T 옵션을 사용하여 순서 부여

```
# Factor형 자료 생성
gender <- factor(c('남','여','남','남','여'), levels=c('남','여'))
levels(gender)      # Factor의 고유한 값 목록 확인

# Factor에 서열 부여 order = T
grade <- factor(c("A", "B", "A", "C", "D", "F", "B", "C"),
               levels = c("A", "B", "C", "D", "F"), order=T)
levels(grade)       # Levels: A < B < C < D < F
```