

제3강

매트릭스와 데이터프레임

Section 01

매트릭스

1. 매트릭스

1. 매트릭스의 개념

- 1차원 데이터 : '몸무게' 데이터와 같은 단일 주제의 데이터 → 벡터
- 2차원 데이터 : '키', '몸무게', '나이' 와 같은 여러 주제의 데이터
→ 매트릭스, 데이터 프레임

몸무게	키	몸무게	나이
62.4	168.4	62.4	29
65.3	169.5	65.3	27
59.8	172.1	59.8	26
46.5	185.2	46.5	25
49.8	173.7	49.8	26
58.7	175.2	58.7	28

(a) 1차원 데이터

(b) 2차원 데이터

그림 3-1 1차원 데이터와 2차원 데이터

- **매트릭스(matrix):** 데이터 테이블의 모든 셀의 값들이 동일한 자료형
- **데이터 프레임(data frame):** 자료형이 다른 컬럼들로 구성

1. 매트릭스

열(column), 컬럼,
변수(variable)

name	age	job	office	M_F

셀(cell)

행(row)
관측값(observation)

그림 3-2 데이터 테이블에 사용하는 용어

1. 매트릭스

2 매트릭스 만들기

2.1 기본적인 매트릭스 만들기

- 2차원 테이블 형태의 자료구조로, 매트릭스의 모든 셀에 저장되는 값은 동일한 자료형이어야 함

코드 3-1

```
z <- matrix(1:20, nrow=4, ncol=5)
```

z

매트릭스 z의 내용을 출력

```
> z <- matrix(1:20, nrow=4, ncol=5)
```

```
> z
```

매트릭스 z의 내용을 출력

```
      [,1] [,2] [,3] [,4] [,5]
```

```
[1,]    1     5     9    13    17
```

```
[2,]    2     6    10    14    18
```

```
[3,]    3     7    11    15    19
```

```
[4,]    4     8    12    16    20
```

matrix(1:20, nrow=4, ncol=5)

매트릭스에
저장될 값

행의 수

열의 수

1. 매트릭스

2.2 매트릭스에 저장될 값들을 행 방향으로 채우기

코드 3-2

```
z2 <- matrix(1:20, nrow=4, ncol=5, byrow=T)
z2                                     # 매트릭스 z2의 내용을 출력
```

```
> z2 <- matrix(1:20, nrow=4, ncol=5, byrow=T)
> z2                                     # 매트릭스 z2의 내용을 출력
```

```
      [,1] [,2] [,3] [,4] [,5]
[1,]    1    2    3    4    5
[2,]    6    7    8    9   10
[3,]   11   12   13   14   15
[4,]   16   17   18   19   20
```

여기서, `byrow=TRUE`는 데이터를 가로 즉 행으로 저장
을 하겠다라는 것이다. 지정하지 않으면 컬럼순으로 저
장되어진다.

1. 매트릭스

2.3 기존 매트릭스에 벡터를 추가하여 새로운 매트릭스 만들기

코드 3-3

```
x <- 1:4                                # 벡터 x 생성
y <- 5:8                                # 벡터 y 생성
z <- matrix(1:20, nrow=4, ncol=5) # 매트릭스 z 생성

m1 <- cbind(x,y)                        # x와 y를 열 방향으로 결합하여 매트릭스 생성
m1                                     # 매트릭스 m1의 내용을 출력
m2 <- rbind(x,y)                        # x와 y를 행 방향으로 결합하여 매트릭스 생성
m2                                     # 매트릭스 m2의 내용을 출력
m3 <- rbind(m2,x) # 매트릭스 m2와 벡터 x를 행 방향으로 결합
m3                                     # 매트릭스 m3의 내용을 출력
m4 <- cbind(z,x)                        # 매트릭스 z와 벡터 x를 열 방향으로 결합
m4                                     # 매트릭스 m4의 내용을 출력
```

rbind()에서 r은 row의 약자임.
cbind()에서 c는 column의 약자임

1. 매트릭스

```
> x <- 1:4           # 벡터 x 생성
> y <- 5:8           # 벡터 y 생성
> z <- matrix(1:20, nrow=4, ncol=5) # 매트릭스 z 생성
> m1 <- cbind(x,y)   # x와 y를 열 방향으로 결합하여 매트릭스 생성
> m1                # 매트릭스 m1의 내용을 출력

      x y
[1,] 1 5
[2,] 2 6
[3,] 3 7
[4,] 4 8

> m2 <- rbind(x,y)   # x와 y를 행 방향으로 결합하여 매트릭스 생성
> m2                # 매트릭스 m2의 내용을 출력

 [,1] [,2] [,3] [,4]
x    1    2    3    4
y    5    6    7    8
```


1. 매트릭스

```
> m3 <- rbind(m2,x)
```

```
> m3
```

```
  [,1] [,2] [,3] [,4]
```

```
x    1    2    3    4
```

```
y    5    6    7    8
```

```
x    1    2    3    4
```

```
> m4 <- cbind(z,x)
```

```
> m4
```

```
      x
```

```
[1,] 1 5  9 13 17 1
```

```
[2,] 2 6 10 14 18 2
```

```
[3,] 3 7 11 15 19 3
```

```
[4,] 4 8 12 16 20 4
```

```
# 매트릭스 m2와 벡터 x를 행 방향으로 결합
```

```
# 매트릭스 m3의 내용을 출력
```

```
# 매트릭스 z와 벡터 x를 열 방향으로 결합
```

```
# 매트릭스 m4의 내용을 출력
```

1. 매트릭스

3. 매트릭스에서의 값 추출

3.1 인덱스 값을 이용하여 매트릭스에서의 값 추출하기

- 매트릭스에서 특정 위치에 있는 값을 추출하는 방법은 벡터와 유사
- 값들의 위치를 나타내는 인덱스를 사용하는데, 2차원상에서 위치를 지정하려면 2개 필요

코드 3-4

```
z <- matrix(1:20, nrow=4, ncol=5) # 매트릭스 z 생성
z                                   # 매트릭스 z의 내용 출력
```

```
z[2,3] # 2행 3열에 있는 값
z[1,4] # 1행 4열에 있는 값
z[2,]  # 2행에 있는 모든 값
z[,4]  # 4열에 있는 모든 값
```

```
> z <- matrix(1:20, nrow=4, ncol=5) # 매트릭스 z 생성
> z                                   # 매트릭스 z의 내용 출력
```

```
      [,1] [,2] [,3] [,4] [,5]
[1,]    1    5    9   13   17
[2,]    2    6   10   14   18
[3,]    3    7   11   15   19
[4,]    4    8   12   16   20
```

1. 매트릭스

```
> z[2,3]           # 2행 3열에 있는 값  
[1] 10  
  
> z[1,4]           # 1행 4열에 있는 값  
[1] 13  
  
> z[2,]            # 2행에 있는 모든 값  
[1]  2  6 10 14 18  
  
> z[,4]            # 4열에 있는 모든 값  
[1] 13 14 15 16
```

	[,1]	[,2]	[,3]	[,4]	[,5]
[1,]	1	5	9	13	17
[2,]	2	6	10	14	18
[3,]	3	7	11	15	19
[4,]	4	8	12	16	20

그림 3-3 매트릭스에서 값의 위치 지정

1. 매트릭스

3.2 매트릭스에서 여러 개의 값을 동시에 추출하기

코드 3-5

```
z <- matrix(1:20, nrow=4, ncol=5) # 매트릭스 z 생성
z                                  # 매트릭스 z의 내용 출력

z[2,1:3]                          # 2행의 값 중 1~3열에 있는 값
z[1,c(1,2,4)]                     # 1행의 값 중 1, 2, 4열에 있는 값
z[1:2,]                           # 1, 2행에 있는 모든 값
z[,c(1,4)]                        # 1, 4열에 있는 모든 값
```

1. 매트릭스

```
> z <- matrix(1:20, nrow=4, ncol=5) # 매트릭스 z 생성
> z # 매트릭스 z의 내용 출력
      [,1] [,2] [,3] [,4] [,5]
[1,]    1    5    9   13   17
[2,]    2    6   10   14   18
[3,]    3    7   11   15   19
[4,]    4    8   12   16   20
>
> z[2,1:3] # 2행의 값 중 1~3열에 있는 값
[1] 2 6 10
> z[1,c(1,2,4)] # 1행의 값 중 1, 2, 4열에 있는 값
[1] 1 5 13
> z[1:2,] # 1, 2행에 있는 모든 값
      [,1] [,2] [,3] [,4] [,5]
[1,]    1    5    9   13   17
[2,]    2    6   10   14   18
> z[,c(1,4)] # 1, 4열에 있는 모든 값
      [,1] [,2]
[1,]    1   13
[2,]    2   14
[3,]    3   15
[4,]    4   16
```

1. 매트릭스

4. 매트릭스의 행과 열에 이름 지정

4.1 매트릭스의 행과 열에 이름을 지정하는 방법

코드 3-6

```
score <- matrix(c(90,85,69,78,  
                 85,96,49,95,  
                 90,80,70,60),  
               nrow=4, ncol=3)  
  
score  
rownames(score) <- c('John','Tom','Mark','Jane')  
colnames(score) <- c('English','Math','Science')  
score
```

```
> score <- matrix(c(90,85,69,78,  
+                 85,96,49,95,  
+                 90,80,70,60),  
+                 nrow=4, ncol=3)
```

```
> score  
      [,1] [,2] [,3]  
[1,]  90   85  90  
[2,]  85   96  80  
[3,]  69   49  70  
[4,]  78   95  60
```

1. 매트릭스

```
> rownames(score) <- c('John','Tom','Mark','Jane')  
> colnames(score) <- c('English','Math','Science')  
> score
```

	English	Math	Science
John	90	85	90
Tom	85	96	80
Mark	69	49	70
Jane	78	95	60

1. 매트릭스

4.2 행과 열에 지정한 이름을 이용하여 매트릭스값 추출하기

코드 3-7

```
score['John','Math']          # John의 수학 성적
score['Tom',c('Math','Science')] # Tom의 수학, 과학 성적
score['Mark',]                 # Mark의 모든 과목 성적
score[, 'English']             # 모든 학생의 영어 성적
rownames(score)                # score의 행의 이름
colnames(score)                # score의 열의 이름
colnames(score)[2]             # score의 열의 이름 중 두 번째 값
```


1. 매트릭스

```
> score['John','Math']           # John의 수학 성적
[1] 85

> score['Tom',c('Math','Science')] # Tom의 수학, 과학 성적
      Math Science
      96      80

> score['Mark',]                 # Mark의 모든 과목 성적
English  Math Science
      69      49      70

> score[, 'English']             # 모든 학생의 영어 성적
John Tom Mark Jane
      90      85      69      78

> rownames(score)                # score의 행의 이름
[1] "John" "Tom" "Mark" "Jane"

> colnames(score)                # score의 열의 이름
[1] "English" "Math" "Science"

> colnames(score)[2]             # score의 열의 이름 중 두 번째 값
[1] "Math"
```

Section 02

데이터 프레임

2. 데이터 프레임

1. 데이터프레임의 개념

- 숫자형 벡터, 문자형 벡터 등 서로 다른 형태의 데이터를 2차원 데이터 테이블 형태로 묶을 수 있는 자료구조
- 외관상으로는 매트릭스와 차이가 없지만 매트릭스에 저장되는 모든 값들이 동일한 자료형인 것과는 달리 데이터 프레임에는 서로 다른 자료형의 값들이 함께 저장

키	몸무게	키	몸무게	성별
168.4	62.4	168.4	62.4	M
169.5	65.3	169.5	65.3	F
172.1	59.8	172.1	59.8	F
185.2	46.5	185.2	46.5	M
173.7	49.8	173.7	49.8	M
175.2	58.7	175.2	58.7	F

(a) 매트릭스의 예

(b) 데이터프레임의 예

그림 3-4 매트릭스와 데이터프레임의 예

2. 데이터 프레임

2. 데이터프레임 만들기

코드 3-8

```
city <- c("Seoul","Tokyo","Washington") # 문자로 이루어진 벡터
rank <- c(1,3,2) # 숫자로 이루어진 벡터
city.info <- data.frame(city, rank) # 데이터프레임 생성
city.info # city.info의 내용 출력
```

```
> city <- c("Seoul","Tokyo","Washington") # 문자로 이루어진 벡터
> rank <- c(1,3,2) # 숫자로 이루어진 벡터
> city.info <- data.frame(city, rank) # 데이터프레임 생성
> city.info # city.info의 내용 출력
```

	city	rank
1	Seoul	1
2	Tokyo	3
3	Washington	2

2. 데이터 프레임

3. iris 데이터셋

- R에서 제공하는 실습용 데이터 셋 중의 하나로 데이터 프레임으로 되어 있음
- 150 종류의 붓꽃에 대해 4개 분야의 측정 데이터와 품종 정보를 결합하여 만든 데이터셋

```
> iris
```

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
1	5.1	3.5	1.4	0.2	setosa
2	4.9	3.0	1.4	0.2	setosa
3	4.7	3.2	1.3	0.2	setosa
4	4.6	3.1	1.5	0.2	setosa
5	5.0	3.6	1.4	0.2	setosa
6	5.4	3.9	1.7	0.4	setosa
... (이하생략)					

2. 데이터 프레임

열 이름	의미	자료형
Sepal.Length	꽃받침의 길이	숫자형
Sepal.Width	꽃받침의 폭	숫자형
Petal.Length	꽃잎의 길이	숫자형
Petal.Width	꽃잎의 폭	숫자형
Species	붓꽃의 품종	문자형(팩터)

표  iris 데

이터셋

코드 3-9

```
iris[,c(1:2)]           # 1, 2열의 모든 데이터
iris[,c(1,3,5)]         # 1, 3, 5열의 모든 데이터
iris[,c("Sepal.Length", "Species")] # 1, 5열의 모든 데이터
iris[1:5,]              # 1~5행의 모든 데이터
iris[1:5,c(1,3)]        # 1~5행의 데이터 중 1, 3열의 데이터
```

2. 데이터 프레임

```
> iris[,c(1:2)]
```

1, 2열의 모든 데이터

	Sepal.Length	Sepal.Width
1	5.1	3.5
2	4.9	3.0
3	4.7	3.2
...	(생략)	

```
> iris[,c(1,3,5)]
```

1, 3, 5열의 모든 데이터

	Sepal.Length	Petal.Length	Species
1	5.1	1.4	setosa
2	4.9	1.4	setosa
3	4.7	1.3	setosa
...	(생략)		

```
> iris[,c("Sepal.Length", "Species")]
```

1, 5열의 모든 데이터

	Sepal.Length	Species
1	5.1	setosa
2	4.9	setosa
3	4.7	setosa
...	(생략)	

2. 데이터 프레임

```
> iris[1:5,] # 1~5행의 모든 데이터
```

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
1	5.1	3.5	1.4	0.2	setosa
2	4.9	3.0	1.4	0.2	setosa
3	4.7	3.2	1.3	0.2	setosa
4	4.6	3.1	1.5	0.2	setosa
5	5.0	3.6	1.4	0.2	setosa

```
> iris[1:5,c(1,3)] # 1~5행의 데이터 중 1, 3열의 데이터
```

	Sepal.Length	Petal.Length
1	5.1	1.4
2	4.9	1.4
3	4.7	1.3
4	4.6	1.5
5	5.0	1.4

감사합니다.