

## 2.8 연습문제

① 다음의 데이터를 벡터 x에 입력해 보자.

17	16	20	24	22	15	21	18
----	----	----	----	----	----	----	----

- 1) 벡터 x에 입력된 데이터의 개수를 확인해 보자.
- 2) 벡터 x의 마지막 데이터의 값을 출력해 보자. 단, x[8]과 같이 자료의 위치를 숫자로 지정하지 않는다.

② 다음의 데이터를 벡터 y에 입력해 보자.

10.4	5.6	3.1	6.4	9.6	7.8	12.1
------	-----	-----	-----	-----	-----	------

- 1) 벡터 y에 입력된 데이터 개수를 확인해 보자.
- 2) 벡터 y에 입력된 데이터 중 마지막에서 두 번째 데이터의 값을 출력해 보자. 단, 자료의 위치를 숫자로 지정하지 않는다.

③ 다음의 문자형 데이터를 벡터 grade에 입력해 보자.

1st	1st	2nd	3rd	2nd	3rd	1st
-----	-----	-----	-----	-----	-----	-----

- 1) 문자형 벡터 grade를 명목형 요인으로 변환해 보자.
- 2) 문자형 벡터 grade를 순서형 요인으로 변환해 보자. 지정되는 수준의 순서는 3rd < 2nd < 1st가 된다.

④ 다음의 행렬 m1을 생성해 보자.

```
> m1
  var1 var2 var3
Case #1  12  21  32
Case #2  17  22  34
Case #3  19  25  35
```

- 1) 행렬의 두 번째 열을 다음과 같이 출력해 보자.

Case #1	Case #2	Case #3
21	22	25

- ⑤ 다음의 데이터 프레임 d1을 생성해 보자.

```
> d1
  var1 var2 var3
1   12   21   32
2   17   22   34
3   19   25   35
```

- 1) 데이터 프레임의 두 번째 열을 다음과 같이 출력해 보자.

```
var2
1   21
2   22
3   25
```

- ⑥ Base 패키지 중 하나인 datasets에 데이터 프레임 iris가 있다.

- 1) iris를 구성하고 있는 변수의 이름과 관찰값의 개수를 확인해 보자.
- 2) 함수 head()를 사용하여 데이터 프레임 iris의 처음 세 줄을 출력해 보자.
- 3) 함수 tail()을 사용하여 데이터 프레임 iris의 마지막 세 줄을 출력해 보자.

- ⑦ 다음의 tibble df1을 생성해 보자. 변수 w는 영문 소문자 중 임의로 9개를 sample(letters, 9)을 실행하여 선택하는 것이므로 다른 문자가 선택될 수 있다.

```
> df1
# A tibble: 9 x 4
   x     y     z     w
<dbl> <int> <int> <chr>
1     1     1     1     q
2     1     2     1     x
3     1     3     1     i
4     1     4     2     h
5     1     5     2     v
6     1     6     2     o
```



7	1	7	3	g
8	1	8	3	u
9	1	9	3	b

- 1) df1의 두 번째 열을 선택하여 벡터로 출력해 보자. 기호 `[[ ]]`와 `$`을 사용하는 것으로 세 가지 방법이 있다.
- 2) df1의 두 번째 열의 처음 다섯 개 자료를 행렬에서 적용되는 인덱싱 기법인 대괄호 하나를 사용하여 선택해 보자. 또한 df1을 전통적인 데이터 프레임으로 전환하여 동일한 작업을 실시해 보고, 차이점을 살펴보자.

**⑧** 다음의 벡터를 만들어 보자.

- 1) -5.0 -4.8 -4.6 ... 4.6 4.8 5.0
- 2) 1 3 5 7 9 2 4 6 8 10
- 3) 1 2 2 3 3 3 4 4 4 4
- 4) 1 2 3 1 2 3 1 2 3
- 5) "a" "a" "b" "b" "b"
- 6) "a1" "a2" "a3" "a4" "a5"
- 7) "aa" "ba" "ca" "ab" "bb" "cb" "ac" "bc" "cc"

**⑨** 벡터 `letters`는 영문 소문자로 이루어진 문자형 벡터이다.

- 1) 벡터 `letters`를 이용해서 다음의 벡터 `a1`을 만들어 보자.

```
> a1
[1] "a1" "b2" "c3" "d4" "e5" "f6" "g7" "h8"
[9] "i9" "j10" "k11" "l12" "m13" "n14" "o15" "p16"
[17] "q17" "r18" "s19" "t20" "u21" "v22" "w23" "x24"
[25] "y25" "z26"
```

- 2) 벡터 `a1`을 이용하여 다음의 벡터 `a2`를 만들어 보자.

```
> a2
[1] "a1-b2-c3-d4-e5-f6-g7-h8-i9-j10-k11-l12-m13-n14-o15-p16-q17-
r18-s19-t20-u21-v22-w23-x24-y25-z26"
```

- 3) 벡터 `a2`를 이용하여 다음의 벡터 `a3`를 만들어 보자.

```
> a3
[1] "a1b2c3d4e5f6g7h8i9j10k11l12m13n14o15p16q17r18s19t20u21v22w23x24y25z26"
```

⑩ 다음의 데이터를 벡터 x에 할당하라.

10.4    5.6    3.1    6.4    21.7

- 1) 벡터 x의 평균값을  $\sum_{i=1}^n x_i/n$ 의 공식으로 구하고 그 값을 mean(x)의 결과와 비교하라.
- 2) 벡터 x의 표준편차를  $\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 / (n-1)}$ 의 공식으로 구하고, 그 값을 sd(x)의 결과와 비교하라.
- 3) 벡터 x의 차분을  $x_{i+1} - x_i$ 의 공식으로 구하고, 그 값을 diff(x)의 결과와 비교하라.

⑪ 다음의 데이터를 벡터 y에 할당하라.

17    16    20    24    22    15    21    18

- 1) 최댓값인 24는 잘못 입력된 데이터라고 한다. 이것을 벡터의 인덱싱 기법을 활용하여 23으로 수정하라. 단, 'y[4] <- 23'와 같이 최댓값의 위치를 직접 입력하는 것은 좋은 방법이 아님.
- 2) 벡터 y의 자료 중 20보다 큰 값의 개수를 함수 sum()을 이용하여 구하라.
- 3) 벡터 y의 자료 중 18보다 작은 값의 비율을 함수 mean()을 이용하여 구하라.

⑫ 다음의 데이터를 벡터 score에 할당하라.

85    91    75    69    52    95    88    100

숫자형 벡터 score를 아래의 규칙에 의하여 요인 grade로 변환시키고자 한다.

90 ≤ score	grade=A	60 ≤ score < 70	grade=D
80 ≤ score < 90	grade=B	score < 60	grade=F
70 ≤ score < 80	grade=C		



- 1) 논리형 벡터를 이용하여 요인 grade를 생성해 보자.
- 2) 함수 cut()을 이용하여 요인 grade를 생성해 보자.
- 3) 다음과 같은 형식으로 score와 grade를 함께 출력해 보자.

```

score grade
1    85    B
2    91    A
3    75    C
4    69    D
5    52    F
6    95    A
7    88    B
8   100    A

```

- ⑬ 다음의 두 변수 X와 Y의 자료를 기반으로 단순 회귀모형  $Y = \beta_0 + \beta_1 X + \varepsilon$ 을 적합해 보자.

X	1	3	2	7	12	6	1	3	6	6	7
Y	6.1	9.7	10.3	18.8	28.3	16.1	5.7	12.6	16.1	15.8	18.8

모수  $\beta_0$ 와  $\beta_1$ 의 최소제곱추정량 값을 행렬 계산식인  $\hat{\beta} = (X^T X)^{-1} X^T Y$ 에 의하여 계산하고, 그 결과를 다음과 같이 출력해 보자.

```

> beta.hat
      Estimate
beta_0 4.790317
beta_1 1.955676

```

- ⑭ 다음과 같이 행렬과 데이터 프레임으로 구성되어 있는 리스트 A를 생성해 보자.

```

> A
$mat
      [,1] [,2] [,3]
[1,]  1.2  1.5  2.1
[2,]  2.5  2.7  2.1
[3,]  3.1  3.2  2.8

$df

```

```

      x1 x2
1 Park 14
2 Lee 16
3 Kim 21

```

- 1) 리스트 A의 첫 번째 요소인 행렬의 행 이름을 sub1, sub2, sub3로 부여하고 열 이름을 Trt1, Trt2, Trt3로 부여하라. 또한 두 번째 요소인 데이터 프레임의 변수 이름을 x1은 name으로 x2는 sales로 변경하라. 변경 후 리스트 A의 출력결과를 다음과 같다.

```

> A
$mat
      Trt1 Trt2 Trt3
Sub1  1.2  1.5  2.1
Sub2  2.5  2.7  2.1
Sub3  3.1  3.2  2.8

$df
  name sales
1 Park    14
2 Lee    16
3 Kim    21

```

- 2) 리스트 A의 첫 번째 요소인 행렬의 각 행의 평균을 계산하여 다음과 같이 출력하라.

```

      sub1      sub2      sub3
1.600000 2.433333 3.033333

```

- 3) 리스트 A의 두 번째 요소인 데이터 프레임의 변수 sales의 평균을 계산하라.