

R-Object

Due date: 4, Apr, 2018

- ※ 수업시간 시작 시 프린트물로 제출
- ※ 숙제로 작성한 코드는 github에 upload 해서 관리해야 한다.

주어진 작업을 수행하기 위한 코드를 작성하여라

1. 다음 데이터를 벡터 x에 다음 숫자열이 입력되어 있다.

17 16 20 24 22 15 21 18

- 1) 벡터 x의 값중 20 이상인 것만 출력하라.
- 2) 벡터 x의 값중 20 이상인 숫자를 100으로 바꾸어 y로 저장하여라.

2. 아래는 입력하고자 하는 행렬값이다.

3	-1	-1	-1	-1
-1	4	-1	-1	-1
-1	-1	5	-1	-1
-1	-1	-1	6	-1
-1	-1	-1	-1	7

- 1) 다음 행렬을 x로 저장하여라.
- 2) 5열을 제거한 부분행렬을 y 로 저장하여라.
- 3) y 행렬의 행과 열의 수를 yinfo 벡터로 저장하여라.
- 4) y 행렬에서 -1 에 해당하는 숫자를 0으로 바꾼 후 y1 행렬로 저장하여라.

3. rowdata.txt 파일이 c:\wtmp 폴더에 저장되어 있다. 파일 데이터의 1-5줄은 다음과 같다.

v1	v2	v3	v4
4		3	2
5	2	2	3
6	3		
1	4	1	2

파일 데이터는 총 101줄이며, 첫 행은 데이터의 이름이다. 데이터에는 결측치가 포함되어 있고 결측치는 r 작업공간으로 데이터를 불러올 때 NA로 표시된다. 이 때 변수 v2 와 v3가 결측치가 아닌 데이터만 모아서 데이터 프레임으로 저장하고자 한다.

- 1) 다음 파일을 데이터 프레임 rdata 로 저장하는 코드를 작성하여라.
- 2) NA를 확인할 수 있는 함수를 적어라.
- 3) 변수 v2 와 v3가 동시에 NA가 아닌 행의 위치를 표시하는 코드를 적어라.
- 4) 변수 v2 와 v3가 결측치가 아닌 데이터만 모아서 데이터 프레임 rdata1으로 저장하여라.

4. 길이가 7인 리스트를 작성하고자 한다. 리스트의 첫 번째 위치에는 논리값을 가지는 벡터 (TRUE,FALSE)를 저장하고, 두 번째 위치에는 대각원소가 1인 2행 2열 대각행렬을, 세 번째 위치에는 길이가 0부터 1까지 길이가 100이며 숫자들의 간격이 동일한 벡터를 입력하고자 한다. 네 번째부터 7번째 까지는 1부터 4까지 숫자를 입력한다.

- 1) 이름이 temp인 리스트를 만드는 코드를 작성하여라.
- 2) temp 리스트의 두 번째 원소를 제거하라
- 3) 2)의 결과 temp 리스트의 3번째 위치하는 결과값은 무엇인가?
- 4) 2)의 결과 temp 리스트의 길이를 확인하는 코드를 작성하여라.

주어진 작업의 결과를 적어라.

5. 다음은 벡터의 연산이다. 벡터연산의 결과를 적고 설명하여라.

1)

```
a1 <- -1:2
a2 <- 1:2
a1 + a2
```

2)

```
a1 <- -(1:2)
a2 <- 1:2
a1 + a2
```

3)

```
a1 <- matrix(0,2,2)
a2 <- c(3,4)
a1 + a2
```

4)

```
a1 <- matrix(1:4,2,2)
a1[a1>2] = 0
```

5)

```
a1 <- 1:5
```

```
a1[-1] - a1[-length(a1)]
```

R-Programming

1. 수열 $a_{n+2} = 0.9a_{n+1} - 0.1a_n + 1$ ($a_1 = 1, a_2 = 3$) 일 때 a_{20} 의 값을 구하고자 한다. R코드를 작성하여라.
2. 1번 문제에서 처음으로 a_n 이 4보다 커지는 n 을 구하는 코드를 작성하라.
3. `A <- matrix(runif(100), 50, 5)`를 실행하면 50행 5열 매트릭스가 생성된다. 각 행의 합을 벡터 v 로 저장하는 반복문을 작성하여라.
4. 다음 R 코드의 실행결과(tmp의 값)를 적어라.

```
tmp = rep(0, 10)
a <- 10:1
idx = 1
for ( j in a)
{
  if (j<5)
  {
    tmp[idx] <- a[j]
    idx <- idx + 1
  }
}
```

5. x 는 행이 1000, 열이 5인 행렬이며, 각 행은 관측치 하나를 의미한다. 예를 들어 첫 번째 관측치는 5차원 데이터이며 그 값을 R에서 코드 `print(x[1,])`으로 확인할 수 있다. 우리는 각 관측치에 1부터 10까지의 숫자를 랜덤하게 할당하고자 한다. 할당 결과를 `sid`라 저장하고자 한다면 R코드를 어떻게 써야하는가?
6. '문제 5번'의 결과 각 관측치는 `sid` 벡터를 통해 1~10사이의 숫자를 할당받게 된다. 특정한 데이터가 어떤 번호를 할당 받았는지는 `sid`를 통해 알 수 있다. 예를 들어 `sid[1]`의 값을 확인해보면 첫 번째 데이터가 어떤 번호를 할당 받았는지 알 수 있다. `m.mat`은 10행 5열 행렬이다. `m.mat[j,]`는 1에서 10사이 숫자중에서 j 값을 할당받은 데이터들의 평균벡터를 저장하고자 한다. 예를 들어 `m.mat[1,]`은 `sid` 값이 1인 데이터들의 평균 벡터가 저장되어 있다. R 코드를 작성하여라.
6. '문제 6번' 결과는 `sid`를 통해 할당 받은 번호에 해당하는 데이터 셋들의 평균을 나타내 준다. 새로운 1000행 10열짜리 행렬 `idist`는 특정 관측치와 어떤 데이터 셋의 평균과의 거리를 저장하는 행렬이다. 예를 들어 `idist[1,1]`의 값은 1번 데이터 (`x[1,]`)와 1번 그룹의 평균

(m.mat[1,]) 의 거리가 입력된다. 여기서 5차원 상의 두 점을

$x_i = (x_{i1}, \dots, x_{i5}), y_i = (y_{i1}, \dots, y_{i5})$ 거리를 $\frac{\langle x_i, y_i \rangle}{\|x_i\| \times \|y_i\|}$ 라고 한다면, idist를 구하는 R코드를

작성하여라.

$$\text{단, } \langle x_i, y_i \rangle = \sum_{j=1}^5 x_{ij}y_{ij}, \quad \|x_i\| = \sqrt{\langle x_i, x_i \rangle}$$

7. '문제 6번'의 결과를 통해 각 데이터와 각 그룹의 중심간의 거리들을 알 수 있다. 각 데이터가 어떤 그룹과 거리가 가장 짧은지 표시한 벡터를 ivec이라고 하면 이를 계산하는 R코드를 작성하여라.

8. 아래 코드를 실행하면 리스트 형태의 데이터 a를 얻게 된다.

```
set.seed(1)
a = list()
for (i in 1:1000)
{
  x = rpois(1,4)+1
  x = min(x,10)
  a[[i]] = sample(1:10, x)
}
```

10명의 게이머가 1000번의 게임에 참가하여 등수를 매겼다. 각 게임의 결과는 a에 저장되어 있다. 예를 들어 a[[1]] 은 4 6 8 2 9 인 벡터이며, 첫 번째 게임에 2,4,6,8,9 선수 5명이 출전하였고, 4가 1등 6이 2등 8이 3등 2가 4등 9가 5등을 했다는 결과를 나타낸다. 즉, a[[k]] 는 k번째 게임에서 결과를 순위대로 선수의 번호로 나열한 데이터다. 한편, 각 선수들은 아래의 규칙대로 점수를 얻게 된다.

게임당 출전선수의 수	점수
2~4명 출전	1등: 1점, 그 외 0점
4~6명 출전	1등: 2점, 2등 1점, 그 외 0점
7~10명 출전	1등: 3점, 2등:2점, 3등:1점, 그 외 0점

1) 경기에 출전한 사람 수에 대한 빈도 테이블을 구하여라.

참여 인원	2명	3명	4명	5명	6명	7명	8명	9명	10명
경기수									

2) 위 경기에서 1등은 누구인가? (동점자가 있는 경우 모두 적어라)

답: ()

