

오늘 실습 시간에는 R 에서 벡터 형식의 데이터를 만들고, 이를 수정하거나 조건에 맞는 값만을 추출하는 방법에 대해 배워볼 것입니다. 아래 주어진 문제를 풀고, 그 답안을 R 스크립트 파일로 저장하여 KLAS 를 통해 제출해 주세요. 답안은 하나의 R 파일로 작성되어야 하며, 제출한 파일은 R 에서 `source()` 함수를 통해 오류 없이 실행이 가능해야 합니다.

**첫 번째 문제는 조교님의 설명과 함께 진행합니다.**

1. [10 점] 다음은 이번 학기 빅데이터프로그래밍연습 과목을 수강하는 학생들 중 무작위로 선택된 학생 10 명의 통학 시간을 분(分) 단위로 기록한 것이다.

```
{30, 60, 30, 5, 25, 15, 100, 40, 20, 45}
```

- (1) [1 점] 위에서 주어진 데이터를 `commuting` 이라고 하는 벡터 객체에 입력하기 위한 R 코드를 적으시오.  
(2) [2 점] `commuting` 의 첫 번째와 마지막 데이터 값을 각각 20 과 NA 로 변경하는 R 코드를 적으시오.

벡터 객체에서 일부 내용만을 출력하기 위해서는 객체의 이름 뒤에 `[]` 를 붙여주면 됩니다. 예를 들어, 위에서 저장한 `commuting` 객체에서 두 번째 값을 출력하고자 한다면 아래와 같이 입력할 수 있습니다.

```
> commuting[2]  
[1] 60
```

만약 두 번째와 다섯 번째 값을 출력하고 싶다면, `c()` 함수를 사용해 원소의 위치를 지정하면 됩니다.

```
> commuting[c(2, 5)]  
[1] 60 25
```

- (3) [1 점] `commuting` 에 저장된 데이터에서 네 번째부터 여섯 번째까지의 값을 출력하기 위한 R 코드를 적으시오.  
(4) [2 점] 위와 다른 방법으로 `commuting` 에 저장된 데이터에서 네 번째부터 여섯 번째까지의 값을 출력하기 위한 R 코드를 적으시오.

R 함수 중 `is.na()` 는 주어진 객체 안에 저장된 각 값들이 NA 인지 여부를 판단하고, 그 결과에 따라 논리값을 반환합니다. 함수에 대한 자세한 내용은 도움말을 통해 확인할 수 있습니다.

- (5) [2 점] `commuting` 에 저장된 각각의 값들이 NA 인지 여부를 판단하고, 그 결과를 `findNA` 라는 객체에 저장하는 R 코드를 적으시오.

지금까지 만들고 사용한 `commuting` 객체는 숫자형 벡터입니다. 이는 아래와 같이 `mode()` 함수를 통해 확인이 가능합니다. 벡터의 유형을 변환하는 방법에는 여러 가지가 있는데, 대표적인 것이 `as()` 계열의 함수를 사용하는 것입니다. 그러나 특별한 함수를 사용하지 않더라도 형을 강제로 변환할 수도 있습니다.

```
> mode(commuting)  
[1] "numeric"
```

(6) [2 점] `as()` 계열의 함수 사용 없이 `commuting` 에 저장된 데이터를 문자열(character) 형식으로 변환하는 R 코드를 적으시오.

2. [10 점] 다음 표는 2019 년 8 월 13 일 기준, 한국프로야구(KBO)에서 방어율이 가장 좋은 상위 10 명 투수의 기록을 보여준다.

번호	이름	방어율	승리	패배	이닝	탈삼진
1	린드블럼	1.95	18	1	148	142
2	산체스	2.24	15	3	128 2/3	115
3	김광현	2.44	14	3	144	138
4	루친스키	2.62	6	7	137 2/3	93
5	켈리	2.66	10	11	142 1/3	100
6	양현종	2.68	13	8	144 1/3	124
7	윌슨	2.72	10	5	139	99
8	브리검	3.06	8	4	120 2/3	90
9	유희관	3.08	7	7	125 2/3	50
10	박종훈	3.18	7	7	116	81

- (1) [2 점] 위의 표에서 방어율과 승리, 패배를 각각 `era`, `win`, `lose` 라고 하는 벡터 객체에 입력하기 위한 R 코드를 적으시오.
- (2) [2 점] `lose` 에 저장된 데이터에서 마지막 값을 제외한 값들을 모두 출력하는 R 코드를 적으시오.
- (3) [2 점] `seq()` 함수를 사용하여 `win` 에 저장된 데이터에서 첫 번째, 세 번째, 다섯 번째, 일곱 번째 값을 출력하기 위한 R 코드를 적으시오.
- (4) [2 점] 비교 연산자 `<` 또는 `>`를 사용하여 `era` 가 2.70 보다 낮은 경우 `TRUE`, 높은 경우 `FALSE` 값을 갖는 벡터 객체를 만들 수 있도록 R 코드를 적으시오. 비교 연산자의 사용법에 대해 도움이 필요한 경우 `help(">")`를 콘솔 창에 입력하면 됩니다.
- (5) [2 점] 승률이 아래의 식과 같이 정의되어 있다고 가정할 때, 표에 주어진 투수들의 승률을 계산하여 출력하기 위한 R 코드를 적으시오.

$$\text{승률} = \frac{\text{승리}}{\text{승리} + \text{패배}}$$

### 기타 평가항목

R 파일의 실행 가능 여부 및 작성 스타일(Google's R Style Guide 준수 여부): 5 점

실습 결과물은 9 월 13 일까지 KLAS 를 통해 제출하면 됩니다. 파일 제출 전 `source(file.choose())` 함수를 사용해 다시 한 번 파일이 오류 없이 실행이 되는지 여부를 확인할 것을 권합니다.

2019 년 8 월 13 일  
홍성연([syhong@khu.ac.kr](mailto:syhong@khu.ac.kr))