

# R을 이용한 통계 기초(2일차)

인하대학교 대학원 통계학과 국성희

# Contents

1. 반복문
2. 조건문
3. 결측값
4. 데이터셋 읽기 및 합치기
5. 간단한 그래프
6. 연습문제



# 1. 반복문

- 1.1 for()를 이용한 루핑

1) for()를 이용하여 1, 2, 3, 4 값 출력하기

```
for(i in 1:4) print(i)
```

2) 100부터 200까지의 합을 구해보자

```
start=100; end=200
```

```
isum=0
```

```
for(i in start:end) ism=isum+i
```

```
print(isum)
```

3) 문자 변수의 루핑 예

```
transport = c( " bus " , " subway " , " car " , " bike " )
```

```
for(vehicle in transport) {print(vehicle)}
```

# 1. 반복문

- 1.2 while()

for()와 달리 while()은 전체 반복회수는 모르지만 어떤 조건을 만족할 때 까지 반복해야 하는 경우 사용된다.

자연수 1, 2, ..., n을 더했을 때 100이 넘으며 100과 가장 가까운 수가 되는 n을 구하는 경우

```
n=0
sumsofar=0
while(sumsofar<=100) {
n=n+1
sumsofar=sumsofar+n
}
print(n) ; print(sumsofar)
```

# 1. 반복문

- 1.3 repeat()

멈추어야 하는 조건을 만날 때 까지 해당 작업을 반복하는 경우 사용된다.

자연수 1, 2, ..., n을 더했을 때 100이 넘으며 100과 가장 가까운 수가 되는 n을 구하는 경우

```
n=0
sum.sofar=0
repeat{
n=n+1
sum.sofar=sum.sofar+n
if(sum.sofar>100) break
}
print(n) ; print(sum.sofar)
```

## 2. 조건문

- 1.1 if else 명령어

if else를 이용하여 x벡터에서 성분 값이 양수이면 새로운 변수 pos값이 1이 되고 그렇지 않으면 pos값이 0이 되도록 변수 pos를 만드는 경우

```
x=c(1,-2,-3,4,5)
pos=rep(0,5)
for( i in 1:5){
    if(x[i]>0) pos[i]=1
    else pos[i]=0
}
print(pos)
```

## 2. 조건문

- 1.2 ifelse 명령어

score가 2 이상이면 이기고(1) 미만이면 지는(0)경우

```
score=c(2,1,0,14,19,20,31)
```

```
ifelse(score>=2,1,0)
```

### 3. 결측값

- 데이터에 결측값이 포함된 경우 R에서는 NA값으로 표시가 되며 여러가지 함수를 통해 결측값을 다룰 수 있다.

데이터 벡터 x에 대해 결측값의 여부를 논리 연산자로 알아보는 경우

```
x=c(1,6,2,NA)
```

```
is.na(x)
```

평균을 구하는 경우, 결측값을 포함한 것과 그렇지 않은 경우

```
mean(x) # NA로 표시
```

```
mean(x,na.rm=TRUE) #NA를 제외하고 계산
```



### 3. 결측값

- 데이터에 결측값이 포함된 경우 R에서는 NA값으로 표시가 되며 여러가지 함수를 통해 결측값을 다룰 수 있다.

결측값을 제외한 데이터 벡터 생성하기

```
xx=na.omit(x)
```

```
na.action(na.omit(x)) #결측값의 위치 찾기
```

결측값을 가장 큰 값으로 주기

```
r=rank(x)
```

```
r=rank(x,na.last=TRUE)
```

## 4. 데이터셋 읽기 및 합치기

- 4.1 내장된 데이터셋 읽기

`data()` # 내장된 데이터셋 목록이 나타난다.

`data(Orange)`

`head(Orange)`

## 4. 데이터셋 읽기 및 합치기

- 4.2 txt 파일 데이터 읽기

다음의 자료를 메모장에 옮긴 후 read.table을 통해 txt파일을 읽는다.

x	y
1	2
2	4
3	5
4	7

```
txtdata=read.table("file위치",header=TRUE)
```

```
csvdata=read.csv("file위치",header=TRUE,sep=",") #csv file
```

## 4. 데이터셋 읽기 및 합치기

- 4.3 데이터 합치기

```
authors=data.frame(surname=c("Tukey","Venables","Tierney","Ripley","McNeil"),  
nationality=c("US","Australia","US","UK","Australia"))
```

```
books=data.frame(name=c("Tukey","Venables","Tierney","Ripley","McNeil"),  
year=c(1976,1995,1996,1998,2000))
```

```
d=merge(authors,books)
```

```
d=merge(authors,books,by.x="surname",by.y="name")
```

## 4. 데이터셋 읽기 및 합치기

- 4.3 데이터 합치기

```
books2=data.frame(name=c("Tukey","Venables","Tierney","Ripley","McNeil","Kim"),  
year=c(1976,1995,1996,1998,2000,2008))
```

```
d2=merge(authors,books2,by.x="surname",by.y="name")
```

```
d3=merge(authors,books2,by.x="surname",by.y="name",all=TRUE)
```

## 5. 간단한 그래프 그리기

- 5.1 간단한 그래프 그리기

`x=1:10`

`y=(x-5)^2`

`plot(x,y)`

`plot(y~x)`

## 6. 연습문제

- 다음은 1반과 2반 학생들의 국어 성적 데이터이다.
  - (a) 각 반의 평균을 구하시오
  - (b) 두 반의 국어 성적 데이터를 합해 Korean 데이터 벡터를 만들고 전체 데이터에 대한 평균을 구하시오
  - (c) 두 반의 데이터를 합치고 이름을 알파벳순으로 정렬하시오

1반	국어		2반	국어
kim	93		kang	90
lee	84		yun	95
park	87		park	88
og	95		cho	75
ang	98		yang	79
min	77		min	87
jung	82		jung	90
moon	92		choi	90

Q&A

