

I. Excel 데이터 정렬

II. R 데이터 정렬

## 1. 데이터 정렬

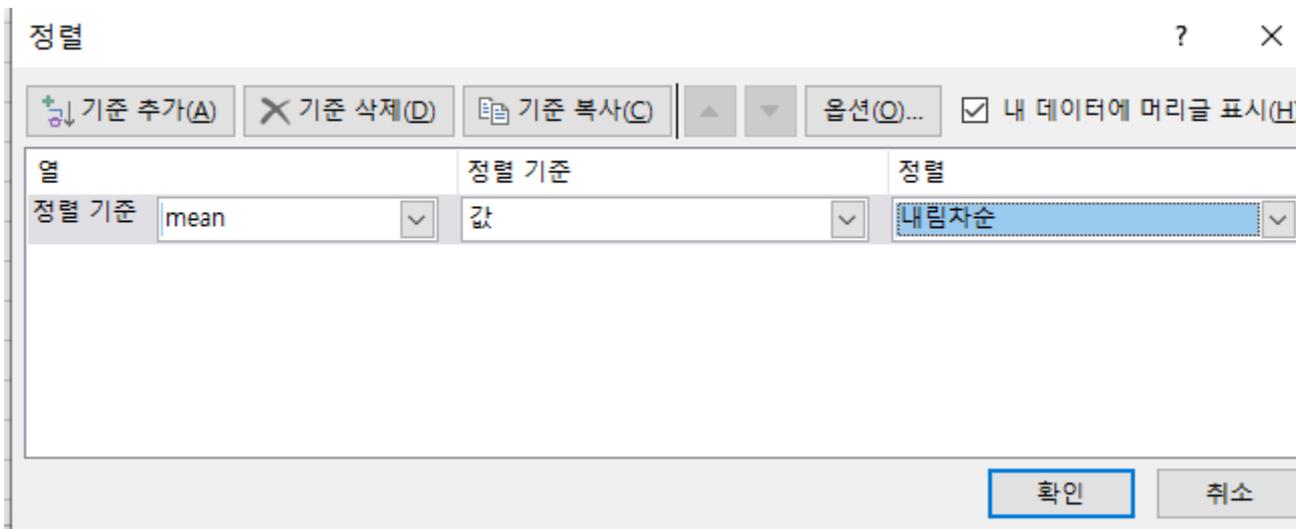
- 항목별로 일정한 순서에 따라 데이터를 나열해 주는 기능을 정렬이라고 함
- 데이터 정렬은 다음의 순서로 함
  - 정렬할 기준 셀로 셀 포인터를 이동
  - [홈]-[편집] 그룹에서 [정렬 및 필터]를 클릭하여 [숫자 내림차순 정렬]를 선택

## 2. 실습

- 강기춘 홈페이지에서 b3-ch1-3.xlsx 파일을 다운로드

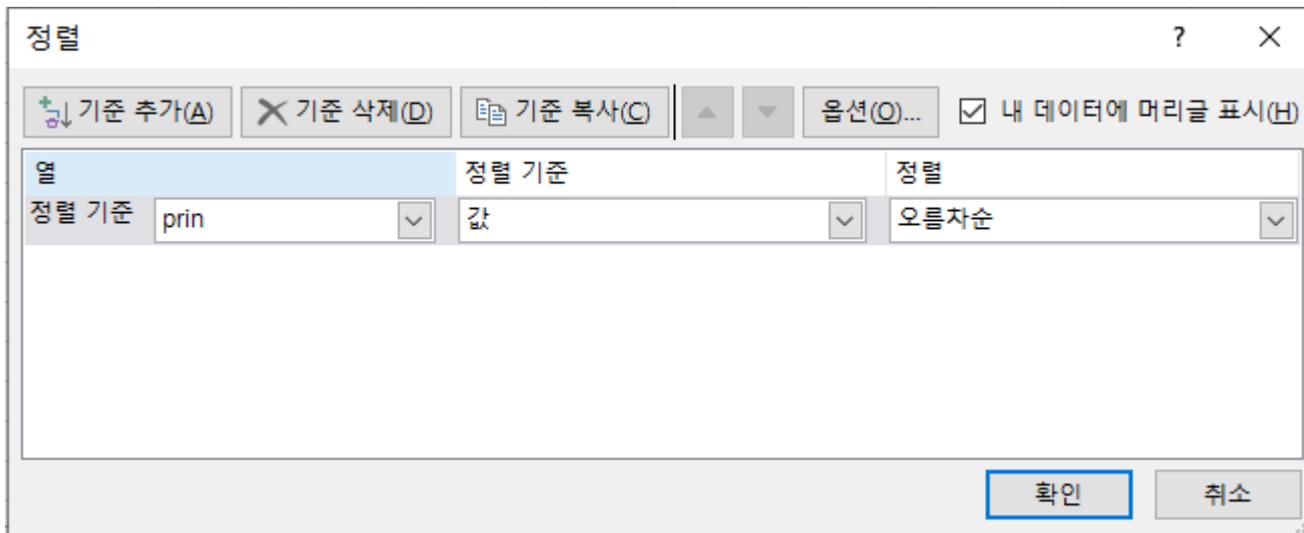
	A	B	C	D	E	F	G	H
1	name	prin	micro	macro	sum	mean	PF	grade
2	김기훈	98	82	95	275	91.66667	pass	A
3	박수동	100	92	80	272	90.66667	pass	A
4	원선희	50	45	75	170	56.66667	fail	F
5	위계영	50	100	100	250	83.33333	pass	B
6	최동팔	80	95	95	270	90	pass	A
7	최종열	90	60	60	210	70	pass	C
8	최종수	30	30	30	90	30	fail	F
9	김기팔	80	25	25	130	43.33333	fail	F
10	이상수	65	70	70	205	68.33333	pass	D
11	강창수	95	90	90	275	91.66667	pass	A

- 예 1 : 데이터를 평균을 중심으로 내림차순으로 정렬하려면 다음의 순서로 함
  - A2부터 H11까지 셀을 선택한 후 [데이터]-[정렬 및 필터] 그룹에서 [정렬]을 클릭하면 정렬 대화상자가 나타남
  - 정렬 대화상자에서 열-정렬 기준의 를 클릭하면 나타나는 선택 항목 중 mean(평균)을 선택
  - 정렬의 를 클릭하면 나타나는 선택 항목 중 내림차순을 선택하고 확인을 누름



	A	B	C	D	E	F	G	H
1	name	prin	micro	macro	sum	mean	PF	grade
2	김기훈	98	82	95	275	91.66667	pass	A
3	강창수	95	90	90	275	91.66667	pass	A
4	박수동	100	92	80	272	90.66667	pass	A
5	최동팔	80	95	95	270	90	pass	A
6	위계영	50	100	100	250	83.33333	pass	B
7	최종열	90	60	60	210	70	pass	C
8	이상수	65	70	70	205	68.33333	pass	D
9	원선희	50	45	75	170	56.66667	fail	F
10	김기팔	80	25	25	130	43.33333	fail	F
11	최종수	30	30	30	90	30	fail	F

- 예 2 : 데이터를 prin(경제원론)을 중심으로 오름차순으로 정렬하려면 다음의 순서로 함
  - A2부터 H11까지 셀을 선택한 후 [데이터]-[정렬 및 필터] 그룹에서 [정렬]을 클릭하면 정렬 대화상자가 나타남
  - 정렬 대화상자에서 열-정렬 기준의 를 클릭하면 나타나는 선택 항목 중 prin(경제원론)을 선택
  - 정렬의 를 클릭하면 나타나는 선택 항목 중 오름차순을 선택하고 확인을 누름



	A	B	C	D	E	F	G	H
1	name	prin	micro	macro	sum	mean	PF	grade
2	최종수	30	30	30	90	30	fail	F
3	원선희	50	45	75	170	56.66667	fail	F
4	위계영	50	100	100	250	83.33333	pass	B
5	이상수	65	70	70	205	68.33333	pass	D
6	최동팔	80	95	95	270	90	pass	A
7	김기팔	80	25	25	130	43.33333	fail	F
8	최종열	90	60	60	210	70	pass	C
9	강창수	95	90	90	275	91.66667	pass	A
10	김기훈	98	82	95	275	91.66667	pass	A
11	박수동	100	92	80	272	90.66667	pass	A

## 1. dplyr

- dplyr 패키지는 데이터 전처리(preprocessing)를 위한 패키지
- 데이터 전처리란 원하는 형태로 데이터를 가공하는 것을 말함
- dplyr 패키지를 install 하기 위해서 Console 창에서 `install.packages("dplyr")`을 입력
- dplyr에서 주로 사용하는 함수는 다음과 같음
  - `select()` : 특정 열(변수)을 추출
  - `arrange()` : 데이터를 오름차순이나 내림차순으로 정렬
  - `filter()` : 조건을 주어 행을 추출
  - `mutate()` : 기존의 변수(열)를 이용하여 새로운 변수(열) 생성
  - `summarise()(with group_by)` : 집단별로 통계량 계산
- `%>%`는 '무엇인가를 전송한다'는 의미의 파이프 연산자(pipe operator)로서 동일한 자료 집합에서 많은 작업을 할 때 유용
  - 왼쪽에서 계산된 결과 값을 파이프 오른쪽의 함수 첫 번째 입력 값으로 넘겨주는 역할을 함

### (1) select()

- 데이터의 특정 열(변수)을 선택
- 예 1 : `prin` 및 `micro`를 선택
- 예 2 : `prin`부터 `macro`까지 한꺼번에 선택
- 예 3 : `micro`만 제외하고 선택
- 예 4 : 열이름(변수명)이 `cro`로 끝나는 열만 선택
- 예 5 : 변수명을 변경



## (2) arrange()

- 특정변수를 기준으로 오름차순 또는 내림차순으로 정렬
- 예 1 : 평균을 기준으로 내림차순으로 정렬
- 예 2 : 경제원론을 기준으로 오름차순으로 정렬

```
b3-ch2-1.R

library(dplyr)

name<-c("김기훈","박수동","원선희","위계영","최동팔","최종열","최종수","김기팔","이상수","강창수")

prin<-c(98,100,50,50,80,90,30,80,65,95)

micro<-c(82,92,45,100,95,60,30,25,70,90)

macro<-c(95,80,75,100,95,60,30,25,70,90)

df<-data.frame(name,prin,micro,macro)

df$sum<-df$prin+df$micro+df$macro

df$mean<-df$sum/3

df$PF<-ifelse(df$mean >= 60, "pass","fail")

df$grade<-ifelse(df$mean >= 90, "A", ifelse(df$mean >= 80, "B", if
else(df$mean >= 70, "C", ifelse(df$mean >= 60, "D", "F"))))

df[order(df$mean,decreasing=T),]

df %>% arrange(desc(mean))

df[order(df$prin),]

df %>% arrange(prin)
```

```
> df %>% arrange(desc(mean))
  name prin micro macro sum mean PF grade
1 김기훈 98 82 95 275 91.66667 pass A
2 강창수 95 90 90 275 91.66667 pass A
3 박수동 100 92 80 272 90.66667 pass A
4 최동팔 80 95 95 270 90.00000 pass A
5 위계영 50 100 100 250 83.33333 pass B
6 최종열 90 60 60 210 70.00000 pass C
7 이상수 65 70 70 205 68.33333 pass D
8 원선희 50 45 75 170 56.66667 fail F
9 김기팔 80 25 25 130 43.33333 fail F
10 최종수 30 30 30 90 30.00000 fail F
```

```
> df %>% arrange(prin)
  name prin micro macro sum mean PF grade
1 최종수 30 30 30 90 30.00000 fail F
2 원선희 50 45 75 170 56.66667 fail F
3 위계영 50 100 100 250 83.33333 pass B
4 이상수 65 70 70 205 68.33333 pass D
5 최동팔 80 95 95 270 90.00000 pass A
6 김기팔 80 25 25 130 43.33333 fail F
7 최종열 90 60 60 210 70.00000 pass C
8 강창수 95 90 90 275 91.66667 pass A
9 김기훈 98 82 95 275 91.66667 pass A
10 박수동 100 92 80 272 90.66667 pass A
```