

[데이터 전처리-II]

4. formula ~ +

(1) **lm()** : 선형회귀식 도출 함수(독립변수, 종속변수, 연속적 변수에서 사용)

ex)

```
x1 <- c(1, 2, 3, 4, 5) # 공부량
```

```
x2 <- c(10, 20, 3, 4, 5) # 기출문제 풀 개수 y
```

```
y <- c(50,60,52,84,95) # 종속변수
```

```
fit <- lm(y ~ x1+x2) #x1과 x2 사이의 상관이 없다는 전제하에 설정
```

5. 데이터 분리

(1) **split**(데이터명, 분리조건)

(2) **subset**(데이터명, subset=(불러올 요소 조건), select =c(불러올 열))

ex) iris데이터에서 setosa종의 꽃받침 길이 출력

```
subset(iris, (iris$Species=='setosa'), select=c('Sepal.Length')) # ( )안에 'subset='은 생략 가능
```

6. 데이터 합치기

(1) **cbind**(합칠 열이름1, 합칠 열이름2); 열 합치기

(2) **rbind**(합칠 행이름1, 합칠 행이름2); 행 합치기 (합치려는 행의 열 구성이 같아야 가능)

(3) **merge**(병합할 데이터1, 병합할 데이터2); 병합

7. 데이터 정렬 **sort()** 정렬된 데이터 반환 / **order()** 정렬된 데이터의 index 반환

ex) iris데이터 꽃받침 길이 기준 내림차순 정렬(**decreasing = T**) 하기 ※ decreasing = T가 없으면 오름차순

```
sort(iris$Sepal.Length, decreasing = T) # 데이터 추출
```

```
order(iris$Sepal.Length, decreasing = T) # 데이터에 해당하는 index 번호 추출
```