

# 데이터과학을 위한 **R**프로그래밍

6주차. R을 이용한 통계분석



**이혜선** 교수

포항공과대학교 산업경영공학과



# 목차

## 6주차. R을 이용한 통계분석

---

1차시

두 그룹간 평균비교분석

2차시

짝을 이룬 그룹간 평균비교

3차시

분산분석(ANOVA)

An isometric illustration of a business meeting. In the center, a large white trapezoidal table is surrounded by several people. To the left, a large screen displays various charts and graphs. To the right, another screen shows a grid of data and gears. In the background, there's a red 3D bar chart on a pedestal. The overall scene is set on a light blue floor with a white trapezoidal area in the center.

6주차

2차시

# 짝을 이룬 그룹간 평균비교

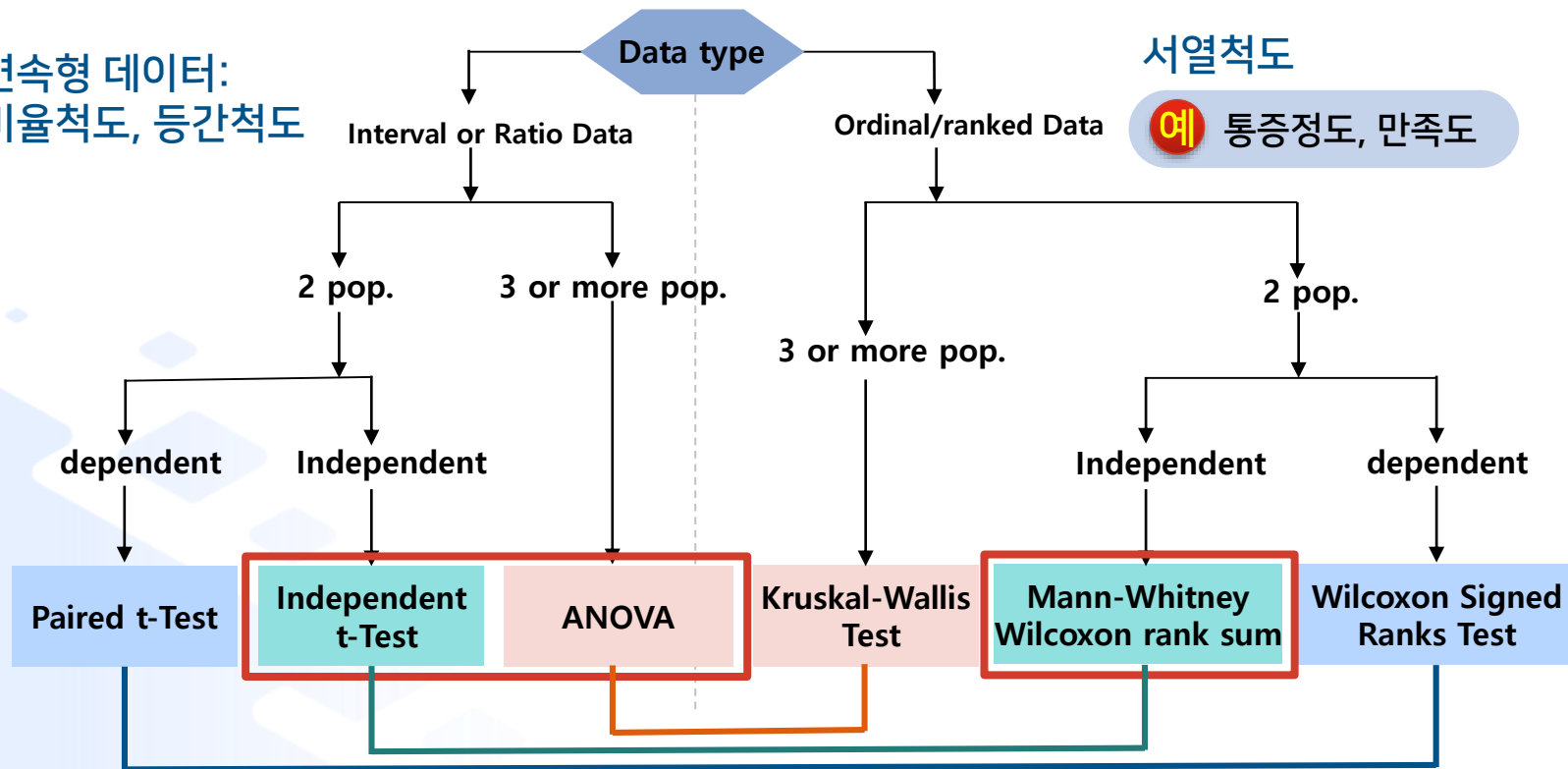
## 모집단간 차이에 대한 검정 (모수/비모수 검정)

연속형 데이터:  
비율척도, 등간척도

서열척도



통증정도, 만족도



## ● 짝을 이룬 그룹간 비교 (paired t-test)

✓ paired t-test : `t.test(before, after, mu=0, paired=T)`

- ▶ 특정 처리(treatment)의 효과를 비교분석 할 때 사용
- ▶ 동일한 표본 : before & after 측정

id	before	After	차이
1	130	125	5
2	140	120	20
3	145	130	15
4	160	125	35
5	125	120	5
...	...	...	...
...	...	...	...
...	...	...	...

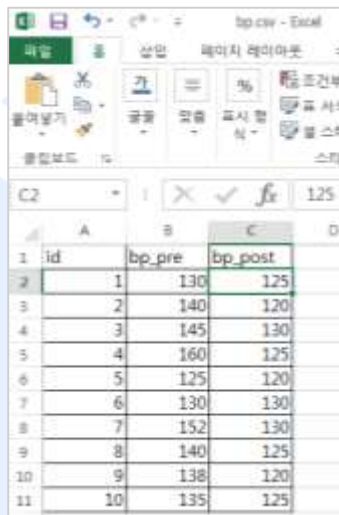
- 예 혈압강하제 투약효과
- 예 방과후 프로그램 성과 (학업흥미도)
- 예 다이어트 프로그램 효과
- 예 직무교육 생산성 효과

평균, 편차 계산 ⇒ 검정통계량

## ● 짝을 이룬 그룹간 비교 (paired t-test)

✓ paired t-test : `t.test(before, after, mu=0, paired=T)`

- ▶ 예제 1 : 고혈압 환자 10명에게 혈압강하제를 12주 동안 투여한 후 복용 전의 혈압과 복용 후의 혈압을 비교하였다. 새로운 혈압강하제가 효과가 있다고 할 수 있는가?



	A	B	C
	id	bp_pre	bp_post
1			
2	1	130	125
3	2	140	120
4	3	145	130
5	4	160	125
6	5	125	120
7	6	130	130
8	7	152	130
9	8	140	125
10	9	138	120
11	10	135	125

짝을 이룬 그룹간 비교 (paired t-test)

## ● 짝을 이룬 그룹간 비교 (paired t-test)

☑ paired t-test : `t.test(before, after, mu=0, paired=T)`

- ▶ 예제 1 : 고혈압 환자 10명에게 혈압강하제를 12주 동안 투여한 후 복용 전의 혈압과 복용 후의 혈압을 비교하였다. 새로운 혈압강하제가 효과가 있다고 할 수 있는가?

```
# paired t-test for two sample means

# set working directory
setwd("D:/tempstore/moocr")

## example 1: blood pressure data
bp<-read.csv("bp.csv")
attach(bp)

# paired t-test (two-sided)
t.test(bp_pre, bp_post, mu=0, paired=T)
```

양측검정 :  $H_0 : \mu_{(dif)} = 0$ ,  $H_1 : \mu_{(dif)} \neq 0$   
유의한 차이가 있는지 없는지에 대한 검정

p-value=0.0015 (매우 작음) :  
유의수준 0.05 ( $\alpha = 0.05$ )보다 작으므로  $H_0$ 를 기각  
⇒ 따라서 투약 전과 투약 후의 혈압에 유의한 차이가 있다고 볼 수 있음

t검정통계량, 자유도, p-value

```
> t.test(bp_pre, bp_post, mu=0, paired=T)

Paired t-test

data: bp_pre and bp_post
t = 4.5095, df = 9, p-value = 0.001469
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 7.226228 21.773772
sample estimates:
mean of the differences
      14.5
```

## ● 짝을 이룬 그룹간 비교 (paired t-test)

### ☑ paired t-test의 검정통계량

```
> t.test(bp_pre, bp_post, mu=0, paired=T)
```

Paired t-test

data: bp\_pre and bp\_post

t = 4.5095, df = 9, p-value = 0.001469

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

7.226228 21.773772

sample estimates:

mean of the differences

14.5

t검정통계량, 자유도, p-value

$$t = 4.5095 = \frac{(\text{평균}(Dif)) - 0}{\text{편차}(Dif)/\sqrt{n}} = \frac{14.5 - 0}{10.168/\sqrt{10}}$$



## ● 짜을 이론 그룹간 비교 (paired t-test)

☑ paired t-test : `t.test(before, after, mu=0, alternative="greater", paired=T)`

```
# paired t-test (one-sided)
t.test(bp_pre, bp_post, mu=0, alternative="greater", paired=T)
```

단측검정 :  $H_0 : \mu_{(dif)} = 0$  ,  $H_1 : \mu_{(dif)} > 0$  혈압(투약 전-투약 후)의 차이가 0보다 큰 가?

```
> # paired t-test (one-sided)
> t.test(bp_pre, bp_post, mu=0, alternative="greater", paired=T)

Paired t-test

data: bp_pre and bp_post
t = 4.5095, df = 9, p-value = 0.0007344
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 8.605783      Inf
sample estimates:
mean of the differences
      14.5
```

p-value=0.0007 (매우 작음), 유의수준 ( $\alpha=0.05$ )보다 작으므로  $H_0$  를 기각  
⇒ 따라서 투약효과가 매우 유의하다고 볼 수 있다 (즉, 투약 전보다 투약 후의 혈압이 유의하게 낮아진다는 것이 검정됨)

## ● 짝을 이룬 그룹간 비교 (paired t-test)

- ☑ 예제 2\* : 비만 대상자들(성인)에게 12주 동안 극저 칼로리 식이요법 (very low-calorie diet: VLCD)을 실시한 후 그 효과를 비교. 이 프로그램이 체중감소에 효과가 있다고 할 수 있는가?

weight.csv

	A	B	C
	id	wt_pre	wt_post
	1	117.3	83.3
	2	111.4	85.9
	3	98.6	75.8
	4	104.3	82.9
	5	105.4	82.3
	6	100.4	77.7
	7	81.7	62.7
	8	89.5	69
	9	78.2	63.9

very low-calorie\*\*

<=800 calories/day

low-calorie

1,000-1,200 calories/day for a woman  
1,200-1,600 calories/day for a man

standard-calorie

2000 calories/day

## ● 짜을 이론 그룹간 비교 (paired t-test)

### ☑ 예제 2 : 극저 칼로리 식이요법(very low-calorie diet: VLCD)의 효과

```
## example 2: Very Low-calroie diet
diet<-read.csv("weight.csv")
attach(diet)

# paired t-test (two-sided)
t.test(wt_pre, wt_post, mu=0, paired=T)
```

양측검정 :  $H_0 : \mu_{(dif)} = 0$ ,  $H_1 : \mu_{(dif)} \neq 0$   
극저칼로리 식이요법이 체중감량에 유의한 효과가 있는지 없는지에 대한 검정

```
> t.test(wt_pre, wt_post, mu=0, paired=T)
```

Paired t-test

data: wt\_pre and wt post

t = 12.74, df = 8, p-value = 1.357e-06

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

18.50003 26.67775

sample estimates:

mean of the differences

22.58889

p-value=0.000001357

참고 : 0.001을 1e-3으로 표기