정보통계학과 **백재욱** 교수

제2강 (2장)

# 두 모집단의 비교

2.1 기본 용어

2.2 두 모집단의 비교

2.3 짝지어진 비교

2.4 두 모분산비에 대한 추론

# 2.1 기본 용어

- 모집단 (population)
- 랜덤추출 (random sampling)
- 랜덤표본 (random sample)
- 모수 (parameter)
- 통계량 (statistic)
- 추정 (estimation)
- 점추정 (point estimation)

## 2.1 기본 용어

- 구간추정 (interval estimation)
- 신뢰수준 (confidence level)
- 통계적 가설 (statistical hypothesis)
- 검정통계량 (test statistic)
- 귀무가설( $H_0$ : null hypothesis)과 대립가설( $H_1$ : alternative hypothesis)
- 귀무가설 $H_0$의 기각역 (rejection region)

# 2.1 기본 용어

## 제1종 오류(type I error)와 제2종 오류(type II error)

| 통계적 결정 | 귀무가설 $H_0$ | |
|---|---|---|
| | 참 | 거짓 |
| $\boldsymbol{H_0}$을 채택함 | 옳은 결정($1-\alpha$) | 제2종 오류($\beta$) |
| $\boldsymbol{H_1}$을 채택함 | 제1종 오류($\alpha$) | 옳은 결정($1-\beta$) |

- 유의수준(significance level) : 1종 오류를 범하는 최대허용확률
- 유의확률(significance probability) : 귀무가설이 맞다는 가정하에 주어진 데이터가 우연히 대립가설을 지지할 확률

# 2.2 두 모집단의 비교

| | 모집단 1 | 모집단 2 |
|---|---|---|
| 모평균 | $\mu_1$ | $\mu_2$ |
| 모분산 | ${\sigma_1}^2$ | ${\sigma_2}^2$ |
| 표본의 크기 | $n_1$ | $n_2$ |
| 랜덤표본 | $x_{11}, x_{12}, \ldots , x_{1n_1}$ | $x_{21}, x_{22}, \ldots , x_{2n_2}$ |
| 표준평균 | $\bar{x}_1 = \frac{\sum x_{1i}}{n_1}$ | $\bar{x}_2 = \frac{\sum x_{2i}}{n_2}$ |
| 표본분산 | $\mathrm{V}_1 = \frac{\sum (x_{1i} - \bar{x}_1)^2}{n_1 - 1}$ | $\mathrm{V}_2 = \frac{\sum (x_{2i} - \bar{x}_2)^2}{n_2 - 1}$ |

# 2.2 두 모집단의 비교

**가정** 모집단 1은 정규분포 $N(\mu_1, \sigma^2)$ 를 따르고,
모집단 2는 정규분포 $N(\mu_2, \sigma^2)$ 를 따른다.

- 공통분산 $\sigma^2$은 다음의 합동표본분산(pooled sample variance)으로 추정함

$$s_p{}^2 = \frac{(n_1-1)V_1+(n_2-1)V_2}{n_1+n_2-2}$$

- 검정통계량 : $t = \dfrac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s_p\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$ 는

  자유도 $(n_1 + n_2 - 2)$ 인 $t$분포를 따름

# 2.2 두 모집단의 비교

**예 1.5** 다음 표는 약의 생산 후와 1년 후의 약 효과를 측정한 결과이다.

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 표본1 | 10.2 | 10.5 | 10.3 | 10.8 | 9.8 | 10.6 | 10.7 | 10.2 | 10.0 | 10.1 |
| 표본2 | 9.8 | 9.6 | 10.1 | 10.2 | 10.1 | 9.7 | 9.5 | 9.6 | 9.8 | 9.9 |

- $\mu_1$, $\mu_2$를 각각 생산직 후와 1년 묵은 약의 평균약효라 하자. 1년이 지나도 약효가 그대로 보존되는지 가설 검정하라.

# 2.2 두 모집단의 비교

## 풀이

1) 가설의 설정 $H_0: \mu_1 - \mu_2 = 0, \quad H_1: \mu_1 - \mu_2 \neq 0$

2) 검정통계량의 값

$n_1 = 10, \quad \bar{x}_1 = 10.32, \quad \mathrm{V}_1 = 0.104$

$n_2 = 10, \quad \bar{x}_2 = 9.83, \quad \mathrm{V}_2 = 0.058$

$$s_p{}^2 = \frac{9 \times 0.104 + 9 \times 0.058}{10 + 10 - 2} = 0.081$$

$$s_p = \sqrt{0.081} = 0.285$$

$$t = \frac{10.32 - 9.83}{0.285\sqrt{\frac{1}{10} + \frac{1}{10}}} = 3.85$$

# 2.2 두 모집단의 비교

## 풀이

3) 의사결정 : 검정통계량의 $t$ 절대값 〉 $t(18\ ;\ 0.025)=2.101$

→ 귀무가설 기각

(유의 수준 0.05에서 약을 오래 보존하면 달라진다고 결론내림)

4) 신뢰구간

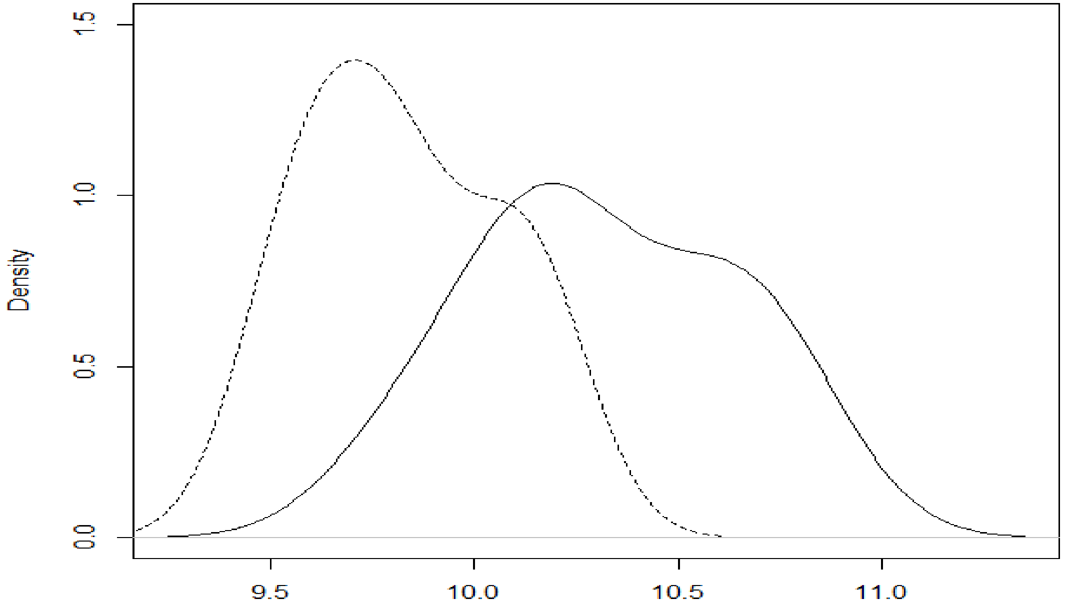
$$(10.32 - 9.83) \pm 2.101(0.285)\sqrt{\frac{1}{10}+\frac{1}{10}}$$

$$= 0.49 \pm 0.268$$

# 2.2 두 모집단의 비교

## R 실습

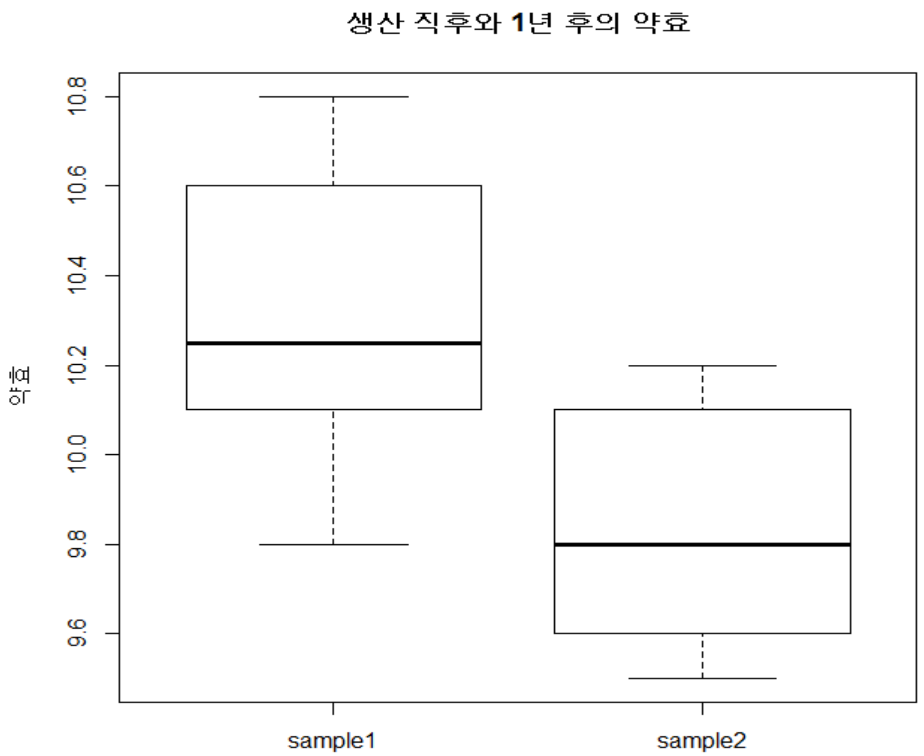
```
sample1 = c(10.2, 10.5, 10.3, 10.8, 9.8, 10.6, 10.7, 10.2, 10.0, 10.1)
sample2 = c(9.8, 9.6, 10.1, 10.2, 10.1, 9.7, 9.5, 9.6, 9.8, 9.9)
plot(density(sample1), lty=1, ylim=c(0, 1.5))
lines(density(sample2), lty=2, ylim=c(0, 1.5))
```



# 2.2 두 모집단의 비교

## R 실습(연속)

```
boxplot(sample1, sample2, ylab="약효",
names=c("sample1","sample2"), main="생산 직후와 1년 후의 약효")
```



# 2.2 두 모집단의 비교

## R 실습(연속)

t.test(sample1, sample2, var.equal=T) # default는 양측검정임

*Two Sample t-test*

*data: sample1 and sample2*

*t = 3.8511, df = 18, p-value = 0.00117*

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

*95 percent confidence interval:*

*0.222688 0.757312*

*sample estimates:*

*mean of x mean of y*

*10.32 9.83*

# 2.3 짝지어진 비교

〈표 2-1〉 짝지어진 비교에 대한 데이터 구조

| 쌍 | 표본 1 | 표본 2 | 차이 $d = x_1 - x_2$ |
|---|---|---|---|
| 1 | $x_{11}$ | $x_{21}$ | $d_1 = x_{11} - x_{21}$ |
| 2 | $x_{12}$ | $x_{22}$ | $d_2 = x_{12} - x_{22}$ |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |
| $n$ | $x_{1n}$ | $x_{2n}$ | $d_n = x_{1n} - x_{2n}$ |

# 2.3 짝지어진 비교

## 가정

$$d_1, d_2, \cdots d_n \sim N(\delta, {\sigma_\delta}^2), \qquad \delta = \mu_1 - \mu_2$$

- 귀무가설 : $H_0 : \delta = \delta_0$
- 검정통계량

$$\bar{d} = \frac{\sum_{i=1}^{n} d_i}{n}, \ \mathrm{V_d} = \frac{\sum_{i=1}^{n}(d_i - \bar{d})^2}{n-1} \quad \cdots\cdots\cdots\cdots\cdots\cdots\cdots \ (2.6)$$

이라고 하는 경우 $t = \frac{\bar{d}-\delta_0}{s_d/\sqrt{n}}$ 은 자유도 $(n-1)$인 $t$분포를 따름
(단, $s_d = \sqrt{V_d}$ )

## 2.3 짝지어진 비교

**예 2.2** 운동화의 밑창에 사용되는 두 가지 재질을 어떻게 비교할 수 있을까?

〈표 2-2〉 운동화 밑창의 마모도 비교 데이터

| 아이 | 재질 $A$ | 재질 $B$ | 차이 $d = B - A$ |
|---|---|---|---|
| 1 | $13.2(L)$ | $14.0(R)$ | 0.8 |
| 2 | $8.2(L)$ | $8.8(R)$ | 0.6 |
| 3 | $10.9(R)$ | $11.2(L)$ | 0.3 |
| 4 | $14.3(L)$ | $14.2(R)$ | $-0.1$ |
| 5 | $10.7(R)$ | $11.8(L)$ | 1.1 |

## 2.3 짝지어진 비교

**예 2.2** 운동화의 밑창에 사용되는 두 가지 재질을 어떻게 비교할 수 있을까?

〈표 2-2〉 운동화 밑창의 마모도 비교 데이터 (연속)

| 아이 | 재질 A | 재질 B | 차이 $d = B - A$ |
|---|---|---|---|
| 6 | $6.6(L)$ | $6.4(R)$ | $-0.2$ |
| 7 | $9.5(L)$ | $9.8(R)$ | $0.3$ |
| 8 | $10.8(R)$ | $11.3(L)$ | $0.5$ |
| 9 | $8.8(L)$ | $9.3(R)$ | $0.5$ |
| 10 | $13.3(R)$ | $13.6(L)$ | $0.3$ |
| | | | 평균차이 $0.41$ |

# 2.3 짝지어진 비교

## 풀이

1) 가설의 설정 $H_0: \ \delta = 0$ VS $H_1: \ \delta > 0$ $\quad (\delta = \mu_B - \mu_A)$

2) 검정통계량의 값 $\bar{d} = 0.41, \quad \mathrm{V_d}(= {s_d}^2) = 0.149, \quad s_d = 0.387$

$$\frac{s_d}{\sqrt{n}} = \frac{0.387}{\sqrt{10}} = 0.122$$

$$t = \frac{0.41}{0.122} = 3.4$$

3) 의사결정 $t\,(9\,;\,0.05) = 1.833 \ < \ 3.4$ ➡ 귀무가설 기각

(유의 수준 5%에서 재질 B로 만든 밑창이
재질 A로 만든 밑창보다 더 많이 닳는다고 결론내림)

# 2.4 두 모분산비에 대한 추론

**2.2절의 절차(두 평균차 검정)를 적용하기 전에**
**두 모분산이 같은지 먼저 검정을 해야 한다.**

- 모분산이 ${\sigma_1}^2$인 정규분포에서 크기 $n_1$의 랜덤표본이 추출되고
  또 모분산이 ${\sigma_2}^2$인 정규분포에서 크기 $n_2$의 두 번째 랜덤표본이 추출될 때
  표본분산 $V_1$과 $V_2$는 각각 ${\sigma_1}^2$ 과 ${\sigma_2}^2$의 추정량으로서
  다음 통계량은 F분포를 따른다.

$$\frac{V_1/{\sigma_1}^2}{V_2/{\sigma_2}^2} \sim \mathrm{F}\,(n_1-1, n_2-1)$$

# 2.4 두 모분산비에 대한 추론

**예 2.3** 예 2.1의 데이터의 경우 두 모집단의 공통분산 가정을 유의수준 α=0.05에서 가설검정하라.

1) 가설의 설정

$$H_0 : \ \sigma_1{}^2 = \sigma_2{}^2, \quad H_1 : \sigma_1{}^2 \neq \sigma_2{}^2$$

2) 검정통계량의 값

$$F = \frac{V_1}{V_2} = \frac{0.105}{0.058} = 1.81$$

3) 의사결정

$$1.81 < F(9, 9 : 0.025) = 4.03$$

➡ 공통분산 가정 타당

정보통계학과 **백재욱** 교수

## 다음시간 안내

제3강 (3장)

# 일원배치법