

9장. 유의성 검정2 – 이표본 검정

독립인 자료(Independent data)

- 비교하고자 하는 두 집단의 측정치들이 서로 다른 개체에서 얻어진 것으로 한 집단의 측정치가 다른 집단의 측정치에 영향을 주지 않는 경우에 얻어진 자료로 각 개체들은 서로 독립이어야 한다.
- 대표적인 통계기법
 - 모수적 기법 ; T-test
 - 비모수적 기법 ; Wilcoxon rank sum test(Mann-Whitney U Test), Median test

9.1. 이표본 T-Test (독립표본)

· 가설검정 절차(T-test)

① 가설 $H_0 : \mu_1 = \mu_2$ v.s. $H_1 : \mu_1 \neq \mu_2$

② 유의수준 α

③ 검정통계량

④ 대표본인 경우 : $Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \approx N(0, 1)$: **생략**

⑤ 정규모집단, **등분산인 경우** ($\sigma_1^2 = \sigma_2^2 = \sigma^2$) :

$$T = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t(n_1 + n_2 - 2), \text{ 여기서 } S_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

⑥ 정규모집단, **등분산이 아닌 경우(이분산)** :

$$T = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \sim t(df), \text{ 여기서 } df = \frac{[\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}]^2}{\frac{[s_1^2/n_1]^2}{n_1 - 1} + \frac{[s_2^2/n_2]^2}{n_2 - 1}}$$

⑦ H_0 를 기각 if $|T| \geq t_{\alpha/2, n_1 + n_2 - 2}$

또는 $p\text{-value} \leq \alpha$, where $p\text{-value} = \Pr(t_{n_1 + n_2 - 2} \geq |T|)$

0) 자료의 형태

A-type(1강의)

| 남자 키 | 여자 키 |
|------|------|
| 179 | 158 |
| 168 | 165 |
| 170 | 160 |
| 180 | 160 |
| 174 | 160 |
| 168 | 170 |
| | 153 |

B type(2강의)

| gender | height |
|--------|--------|
| 1 | 179 |
| 1 | 168 |
| 1 | 170 |
| 2 | 158 |
| 2 | 165 |
| 1 | 180 |
| 1 | 174 |
| 2 | 160 |
| 1 | 168 |
| 2 | 160 |
| 2 | 170 |
| 2 | 153 |

A. 자료의 형태 (A-type)

A-1. 정규모집단, 등분산인 경우 (R 제공)

(1) 예제자료 1 - 지혈제 A와 지혈제 B의 지혈시간

```
x1 = c ( 1.1, 2.3, 4.3, 2.2, 5.3 )      # 지혈제 A의 지혈시간
x2 = c ( 2.3, 4.3, 3.5 )                # 지혈제 B의 지혈시간
t.test(x1, x2, var.equal = T)           #  $H_1 : \mu_1 \neq \mu_2$  등분산 가정하에서
```

Two Sample t-test

```
data: x1 and x2
t = -0.2956, df = 6, p-value = 0.7775
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -3.030714  2.377381
sample estimates:
mean of x mean of y
 3.040000  3.366667
```

해석 :

- 추정 : 지혈제 A의 평균 지혈시간은 3.04시간, 지혈제 B의 평균 지혈시간은 3.36시간이다.
지혈제 A와 지혈제 B의 평균 지혈시간차에 대한 95% 신뢰구간은 (-3.03, 2.37)이다.

• 가설검정

- ① 가설 H_0 : 지혈제 A와 지혈제 B의 평균 지혈시간은 같다.
 H_1 : 지혈제 A와 지혈제 B의 평균 지혈시간은 다르다.
- ② 유의수준 $\alpha=0.05$
- ③ 검정통계량 $T값 = -0.2956$
- ④ $P값 = 0.7775 > \alpha \Rightarrow H_0$ 를 기각할 수 없다
- ⑤ 결론 : 유의수준 5%에서 지혈제 A와 지혈제 B의 평균 지혈시간은 통계적으로 유의한 차이가 없었다.

A-2. 정규모집단, 등분산이 아닌 경우(이분산) (R 제공)

```
t.test(x1, x2, var.equal = F)           #디폴트  $H_1 : \mu_1 \neq \mu_2$ 
```

Welch Two Sample t-test

```
data: x1 and x2
t = -0.33995, df = 5.9717, p-value = 0.7455
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -2.680674  2.027341
sample estimates:
mean of x mean of y
 3.040000  3.366667
```

A-3. 이표본 분산비 F-Test (R 제공)

[NOTE]

A-1 또는 A-2를 하기 전에 등분산인지, 이분산인지 검정하고, A-1, 또는 A-2를 선택
($p\text{-값} > 0.05 \Rightarrow$ 등분산 \Rightarrow A-1) , ($p\text{-값} < 0.05 \Rightarrow$ 이분산 \Rightarrow A-2)

[NOTE] F분포

```
x = seq(0,5, by=0.01)          # -5 ~ 5까지 0.01 간격으로..  
yf = df (x , 5, 6)              # =>  
plot(x, yf, type = "l", col = "blue" ) # => 그림
```

· 가설검정 절차(F-test)

① 가설 $H_0 : \sigma_1^2 = \sigma_2^2$ v.s. $H_1 : \sigma_1^2 \neq \sigma_2^2$

② 유의수준 α

③ 검정통계량 $F = \frac{S_1^2}{S_2^2} \sim F(n_1 - 1, n_2 - 1)$

④ H_0 를 기각 if $F \geq F_{\alpha/2, n_1-1, n_2-1}$ 또는 $F \leq F_{1-\alpha/2, n_1-1, n_2-1}$
if $p\text{-value} \leq \alpha$

```
var.test(x1, x2)                # 정규모집단에서 사용  $H_1 : \sigma_1^2 \neq \sigma_2^2$   
#  $H_1 : \sigma_1^2 \neq \sigma_2^2$  : 더 나은 통계량 Levene's test, Bartlett.test
```

```
F test to compare two variances  
  
data:  x1 and x2  
F = 2.8895, num df = 4, denom df = 2, p-value = 0.5465  
alternative hypothesis: true ratio of variances is not equal to 1  
95 percent confidence interval:  
 0.07362013 30.77032496  
sample estimates:  
ratio of variances  
2.889474
```

해석 :

· 가설검정

① 가설 H_0 : 지혈제 A와 지혈제 B의 분산은 같다.

H_1 : 지혈제 A와 지혈제 B의 분산은 다르다.

② 유의수준 $= 0.05$

③ 검정통계량 $F\text{-값} = 2.8895$

④ $P\text{-값} = 0.5465 > \alpha \Rightarrow H_0$ 를 기각할 수 없다

⑤ 결론 : 지혈제 A와 지혈제 B의 지혈시간은 분산이 통계적으로 유의한 차이가 없었다.

=> 이표본 T-Test에서 등분산 가정 적용(A-1)

A-4. 이표본 T-Test (독립표본) (R 프로그래밍)

```

T_test_2A = function(x , y) {
  n1 = length(x) ; n2 = length(y)      # 등분산 F-test
  s1 = var(x) ; s2 = var(y)
  F = s1 / s2
  pvalue = min( 2*pf(F, n1-1, n2-1) , 2*( 1 - pf(F, n1-1, n2-1) ) )
  cat( " ===== 이표본 분산비 검정 =====", "\n", "\n")
  cat( "          F = " , F, " , P - value = " , pvalue , "\n", "\n")

  xbar = mean(x) ; ybar = mean(y)      # 등분산인 경우 t-test
  sp = sqrt ( ( (n1 - 1) * s1 + (n2 - 1) * s2 ) / ( n1 + n2 - 2) )
  T = ( xbar - ybar ) / ( sp * sqrt ( 1/n1 + 1/n2) )
  pvalue = 2 *( 1 - pt( abs(T), n1 + n2 - 2) )
  cat( " ===== 이표본 평균차 검정 =====", "\n", "\n")
  cat( " 등분산인 경우 : T = " , T, " , P - value = " , pvalue , "\n")

  df = ( s1/n1 + s2/n2 )^2 / ( (s1/n1)^2 / (n1 - 1) + (s2/n2)^2 / (n2 - 1) )
  T = ( xbar - ybar ) / sqrt ( s1/n1 + s2/n2)      # 이분산인 경우 t-test
  pvalue = 2 *( 1 - pt( abs(T), df ) )
  cat( " 이분산인 경우 : T = " , T, " , P - value = " , pvalue , "\n")

}
T_test_2A(x1, x2)

```

===== 이표본 분산비 검정 =====

F = 2.889474 , P - value = 0.5465404

===== 이표본 평균차 검정 =====

등분산인 경우 : T = -0.295603 , P - value = 0.7774965
 이분산인 경우 : T = -0.3399501 , P - value = 0.7455324

[NOTE] SPSS 결과

독립표본 검정

| | | Levene의 등분산 검정 | | 평균의 동일성에 대한 T 검정 | | | | | | 차이의 95% 신뢰구간 | |
|------|--------------|----------------|------|------------------|-------|-----------|---------|---------|--|--------------|---------|
| | | F | 유의확률 | t | 자유도 | 유의확률 (양측) | 평균차이 | 표준오차 차이 | | 하한 | 상한 |
| 지혈시간 | 등분산을 가정함 | 2.366 | .175 | -.296 | 6 | .777 | -.32667 | 1.10509 | | -3.03071 | 2.37738 |
| | 등분산을 가정하지 않음 | | | -.340 | 5.972 | .746 | -.32667 | .96093 | | -2.68067 | 2.02734 |

[과제23] (여러분은 따라서 해보시고, 아래 실습문제를 과제로 제출하시길)

• 과제방법 :

- ① R에서 제공 결과 => 결과분석
- ② R 프로그래밍 결과 => ①의 결과와 같음을 확인

[실습1] 12가게를 조사하여 6가게는 새제품을 판매하였고, 또 다른 6가게는 기존 제품을 판매하였더니 다음과 같은 자료를 얻었습니다.

새제품과 기존제품의 판매량에 차이가 있는지 검정하시오.

| 새 제품 | 기존제품 |
|------|------|
| 50 | 45 |
| 48 | 43 |
| 47 | 45 |
| 45 | 44 |
| 46 | 43 |
| 43 | 40 |

[실습2] 두 사료의 우유생산량에 차이가 있는지 검정하시오.

| | | | | | | | | | |
|-----|----|----|----|----|----|----|----|----|----|
| 사료1 | 54 | 60 | 66 | 53 | 62 | 61 | 42 | 50 | |
| 사료2 | 53 | 60 | 62 | 67 | 59 | 45 | 60 | 52 | 52 |

첨부파일 : 학번이름23.hwp (예 : 20192260홍길동23.hwp)