

9.3. 이표본 모비율차 검정

A그룹에서의 비율과 B그룹에서의 비율이 같은지 검정

• 가설검정 절차(T-test)

① 가설 $H_0 : p_1 = p_2$ v.s. $H_1 : p_1 \neq p_2$

② 유의수준 α

③ 검정통계량 $Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_1} + \frac{1}{n_2})}} \sim N(0,1)$, 여기서 $\hat{p} = \frac{x_1 + x_2}{n_1 + n_2}$

④ H_0 를 기각 if $p\text{-value} \leq \alpha$

(1) 예제자료

• R-제공

```
x = c ( 100, 170 )          # 성공횟수 입력
n = c ( 300, 400 )          # 시행횟수 입력
prop.test(x, n) # H1: p1 ≠ p2

2-sample test for equality of proportions with continuity correction

data: x out of n
X-squared = 5.6988, df = 1, p-value = 0.01698
alternative hypothesis: two.sided
95 percent confidence interval:
 -0.16664176 -0.01669158
sample estimates:
 prop 1    prop 2 
0.3333333 0.4250000
```

해석 :

• 추정 : D후보의 A도시에서의 지지율은 33.3%, B도시에서의 지지율은 42.5%였다

• 가설검정

① 가설 H_0 : D후보의 A도시에서의 지지율과 B도시에서의 지지율은 같다.

H_1 : D후보의 A도시에서의 지지율과 B도시에서의 지지율은 다르다.

② 유의수준 $\alpha=0.05$

③ 검정통계량 χ^2 값 = 5.6988

④ P값 = 0.01698 < $\alpha \Rightarrow H_0$ 를 기각

⑤ 결론 : 유의수준 5%에서 A도시에서의 지지율과 B도시에서의 지지율은 다르다고 할 수 있다.

B도시에서의 지지율(42.5%)이 A도시에서의 지지율(33.3%)보다 높다고 할 수 있다.

```
Z_prop_2 = function(x , n) {
  p1 = x[1] / n[1]
  p2 = x[2] / n[2]
  p = ( x[1] + x[2] ) / ( n[1] + n[2] )
  se = sqrt ( p * (1-p) * ( 1/n[1] + 1/ n[2] ) )
  Z = ( p1 - p2 ) / se
  pvalue = 2 * ( 1 - pnorm( abs(Z) ) )
  cat( " ===== 비 율 차 검정 =====", "\n", "\n")
  cat( " 비율차 = ", p1-p2, " Z = " , Z, " , P - value = " , pvalue , "\n")
}

Z_prop_2( x , n )

[NOTE] prop.test(x,n,correct=F)          # 연속성을 보정하지 않는 경우

===== 비 율 차 검정 =====

비율차   =   -0.09166667   Z =   -2.46567   , P - value =   0.01367572
```

(2) 우리자료

```
data
bmi = weight/(height^2) * 10000 ; bmi
n= length(bmi)
bmi1 = rep(0,n)
for ( i in 1 : n)
  { if ( bmi[i] <= 25) bmi1[i] = 1
    else bmi1[i] = 2  }

bmi1
bmi1 <- factor(bmi1, labels = c("정상", "비만") ); bmi1  # 1 => 정상, 2=> 비만

b=table(gender, bmi1)
addmargins(b)
x = c ( 12, 1 )
n = c ( 33, 37)

prop.test(x, n)
Z_prop_2( x , n )
```

[과제26] (여러분은 한번 따라서 해보시고, 아래 실습문제를 과제로 제출하시길)

• 과제방법 :

- ① R에서 제공 결과 => 결과분석
- ② R 프로그래밍 결과 => ①의 결과와 p값은 조금 다르지만 결과는 같음을 확인

첨부파일 : 학번이름26.hwp (예 : 20192260홍길동26.hwp)

[실습1]

부모의 안전벨트 착용유무에 따라 아이의 안전벨트 착용율에 차이가 있는지 검정하고, 해석하시오.

안전벨트 착용 유무	아이 착용	아이 미착용
부모착용	54	7
부모미착용	3	12

[실습2]

여론조사에서 A후보는 도시지역 유권자 중에서 5,000명을 임의 추출하여 조사하였더니 지지자가 2,400명 이었고, 농촌지역의 유권자중에서 2,000명을 임의로 추출하여 조사하였더니 지지자가 1,200명이었다.

도시와 농촌지역에서의 지지율이 차이가 있는지 검정하고, 해석하시오.