

범주형자료분석팀

2팀

김정훈
김상태
김민정
박정현
이윤희

INDEX

1. 자료의 형태
2. 분할표
3. 독립성 검정
4. 연구의 종류
5. 확률의 비교

자료의 형태



순서형 자료 : 증상 정도

증상 정도				
핵 심각	조금 심각	보통	괜찮음	리얼 쌔쌔함

“ 명목형 자료분석 방법을 사용할 수도 있음 ”

- 하지만 순서에 대한 정보 무시로 **검정력**에 심각한 **손실** 초래
- 따라서 순서형 자료들에 일정한 점수를 할당하여 **양적인 자료**의 형태로 다루는 경우도 있음

분할표

3차원 분할표

학과가 합쳐짐
(변수 Z 무시)

2차원 분할표

부분분할표

학과	성별	학회	
		합격	불합격
통계	남자	11	25
	여자	10	27
경영	남자	16	4
	여자	22	10
경제	남자	14	5
	여자	7	12

주변분할표

성별	학회	
	합격	불합격
남자	11 + 16 + 14	25 + 4 + 5
여자	10 + 22 + 7	27 + 10 + 12

X : 성별
Y : 합격 / 불합격
Z : 학과

기대도수

WHAT HOW WHERE HOWEVER

정의

- $\mu_{ij} = n\pi_{i+}\pi_{+j}$

$$\pi_{ij} = \pi_{i+} \times \pi_{+j}$$

성질

- X, Y가 독립이면 각 자리의 결합 확률은 주변확률의 곱과 같다.

X	Y		합계
	Y_1	Y_2	
X_1	π_{11}	π_{12}	π_{1+}
X_2	π_{21}	π_{22}	π_{2+}
	π_{+1}	π_{+2}	1

독립성 검정

WHAT HOW WHERE HOWEVER

$$H_0 : n_{ij}\pi_{ij} = \mu_{ij}$$

X와 Y는 독립이다

$$H_1 : n_{ij}\pi_{ij} \neq \mu_{ij}$$

독립이 아니다

값의 차가 크면 독립이 아닐 가능성 ↑

: 관계에 대한 연구 필요

피어슨 카이제곱 검정

WHAT HOW WHERE HOWEVER

통계량

$$\bullet \quad X^2 = \sum \frac{(n_{ij} - \mu_{ij})^2}{\mu_{ij}} \sim \chi^2_{(I-1)(J-1)}$$

(관측도수-기대도수) 차이에 집중!

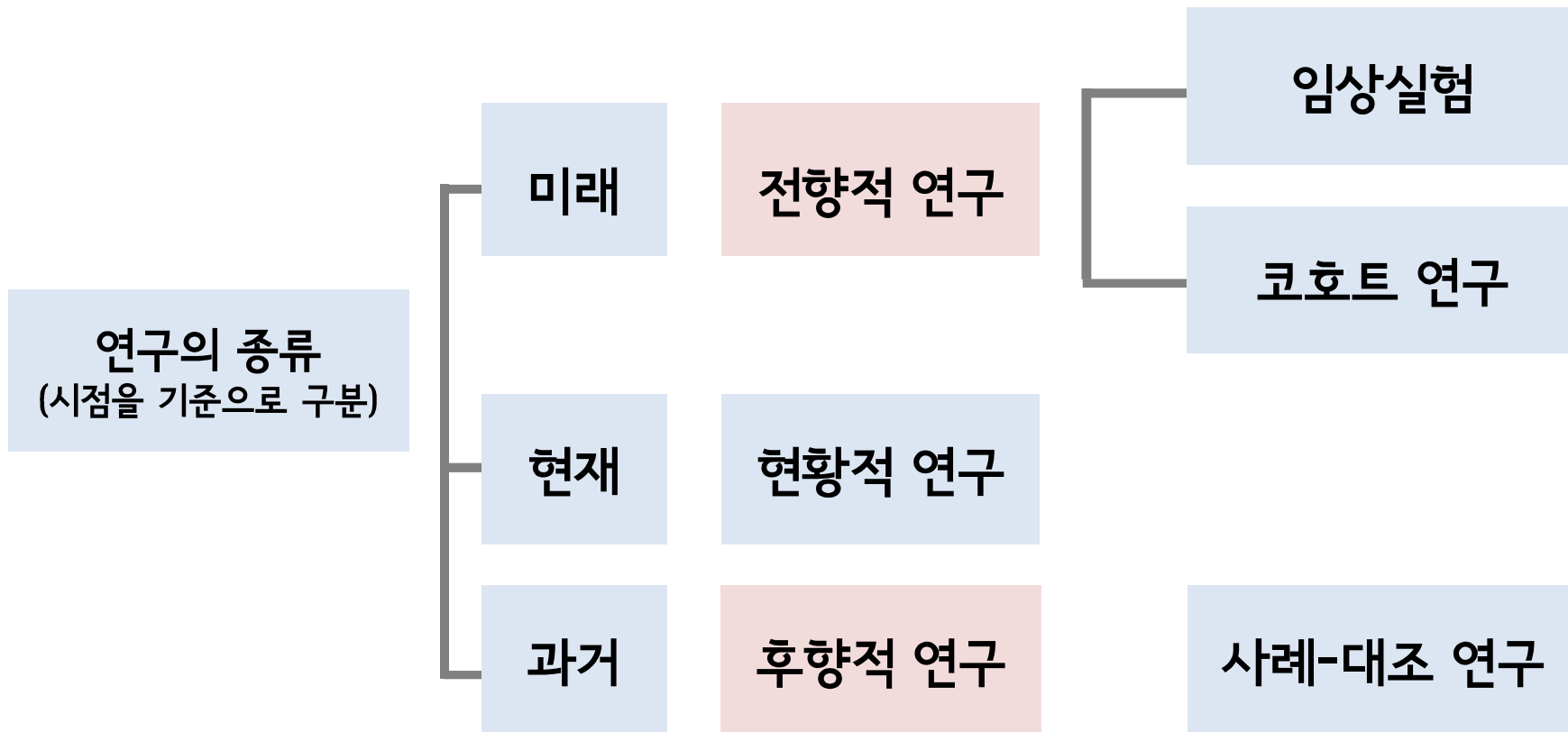
논리구조

표본크기 고정, $n_{ij} - \mu_{ij}$
클수록 통계량 커짐

통계량이 커지므로
p값 작아짐

독립이 아닐 확률
증가

연구의 종류



확률의 비교

확률의 비교 척도

이항 반응변수에 대하여
두 그룹을 비교하는 척도들을 제시

비율의 차이

상대 위험도

오즈비

오즈(odds)

$$\text{실패분의 성공} : \frac{\pi}{1-\pi}$$

오즈비(odds ratio)

$$\text{각 행의 오즈끼리의 비} : \text{odds ratio}(\theta) = \frac{\pi_1/(1-\pi_1)}{\pi_2/(1-\pi_2)} = \frac{\text{1행의 오즈}}{\text{2행의 오즈}}$$

$$\text{범위} : \theta \geq 0$$

$\theta > 1$: 첫 번째 행의 오즈가 두 번째 행의 오즈보다 크다

$1 > \theta \geq 0$: 두 번째 행의 오즈가 첫 번째 행의 오즈보다 크다

➡ 서로 역수 관계 : 같은 정도의 연관성

$$\text{독립} : \theta = 1$$