

INDEX

1. 범주형 자료분석
2. 분할표
3. 독립성 검정
4. 연관성 측도

자료의 형태

자료(DATA)

양적 자료
(Quantitative, 수치형)

질적 자료
(Qualitative, 범주형)

이산형 자료
(Discrete)

연속형 자료
(Continuous)

명목형 자료
(Nominal)

순서형 자료
(Ordinal)

여러 차원의 분할표

3차원 분할표($I \times J \times K$)

X(설명변수), Y(종속변수), Z(제어변수)

고정된 Z의 한 수준에 대해서
XY의 관계를 보여줌
Z를 통제했을 때
Y에 대한 X의 효과를 알 수 있음

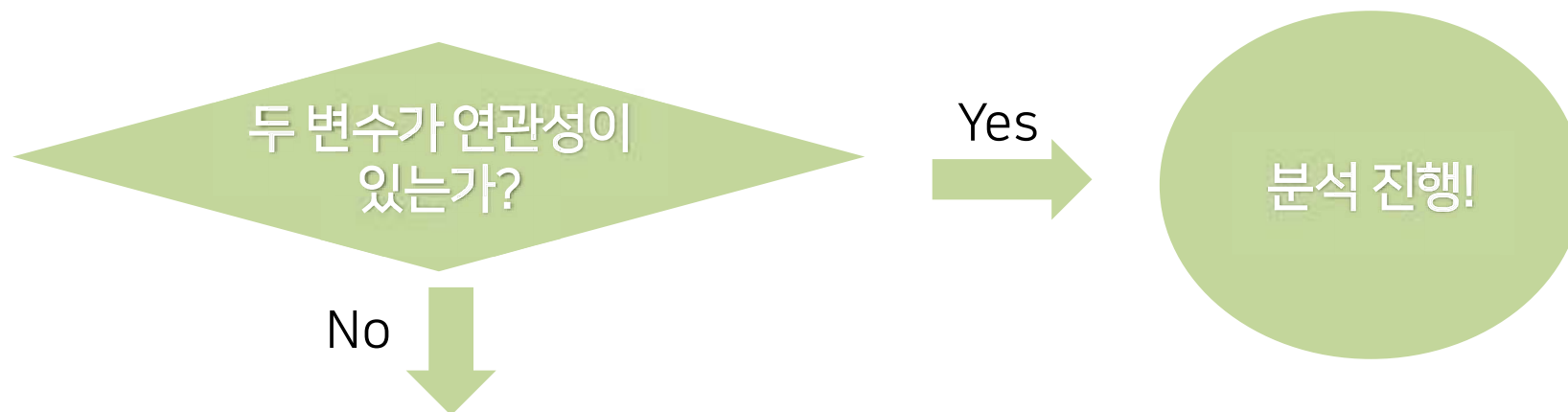
		Y		합계
Z	X	n_{111}	n_{121}	n_{1+1}
		n_{211}	n_{221}	n_{2+1}
	합계	n_{+11}	n_{+21}	n_{++1}
	X	n_{112}	n_{122}	n_{1+2}
		n_{212}	n_{222}	n_{2+2}
	합계	n_{+12}	n_{+22}	n_{++2}

「부분분할표」

: Z의 각 수준에서 X와 Y를 분류한 표

독립성 검정

독립성 검정 두 범주형 변수가 독립인지 검정하는 것!



범주형 변수는 연속형 변수의 상관계수처럼
변수 간의 상관성을 나타내기 어려움

"이런 데이터는 아예 쓸 데가 없어요"

기대도수와 관측도수

기대 도수와 관측 도수를 이용해 앞의 가설을 이렇게 바꿔 쓸 수 있다!

귀무가설 H_0 : 두 범주형 변수는 독립이다 ($\mu_{ij} = n \cdot \pi_{ij}$)

대립가설 H_1 : 두 범주형 변수는 독립이 아니다 ($\mu_{ij} \neq n \cdot \pi_{ij}$)



즉, 기대 도수와 관측 도수의 차이 ($\mu_{ij} - n \cdot \pi_{ij}$) 가
유의미하게 크다면, 귀무 가설을 기각할 가능성이 커지게 됨!

순서형 자료 검정

순서형 자료에 명목형 독립성 검정을 할 수는 있지만, 순서 정보의 손실이 일어나기 때문에 비추!

MH 검정 (Mantel-Haenszel test)

MH 검정통계량

$$M^2 = (n - 1)r^2$$

- 조건 : 두 범주형 변수 모두 순서형일 때 사용
- 원리 : 범주의 각 level에 점수를 할당하여 변수 간의 선형 추세 측정!
- 추세 연관성을 파악하기 위해 **피어슨 교차적률 상관계수** 사용! 없는데가 없는 갓-피어슨...
- 기각역 : $M^2 \geq \chi_{\alpha,1}^2$
- n 과 r 이 **커지면** 검정통계량 M^2 도 **커진다**. -> 귀무가설 **기각** -> 변수 간 **연관성**!

오즈비 (Odds Ratio, OR)

비율의 차이와 상대위험도는 직관적이지만,
한 변수의 수를 고정시킨 조사에서는 사용이 불가함

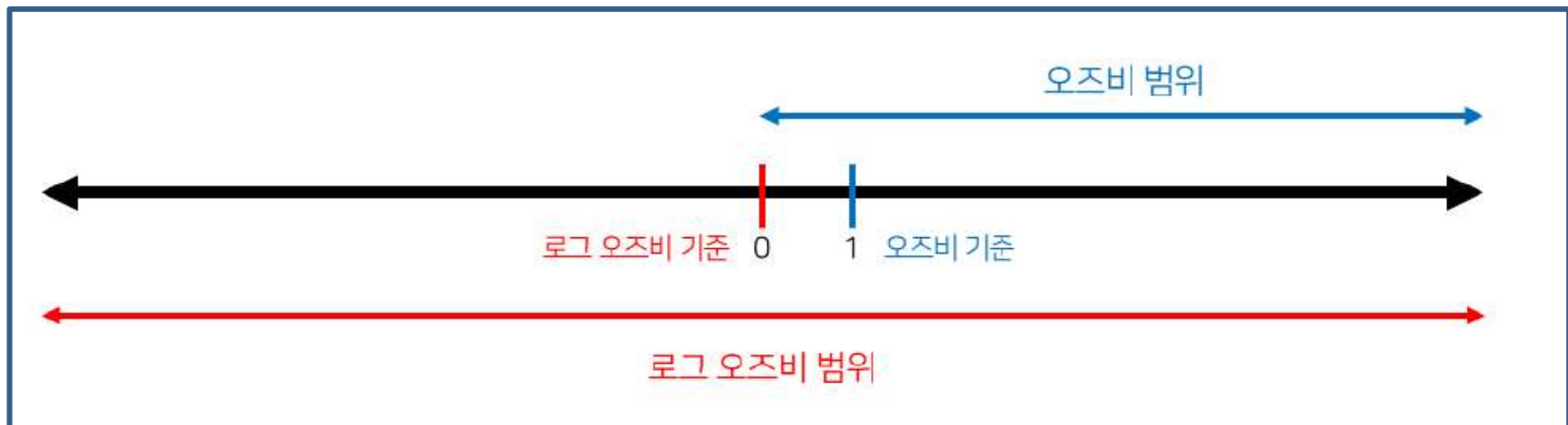
	심장질환 있음 ($Y = 1$)	심장질환 없음 ($Y = 0$)	합
알코올 중독 0 ($X = 1$)	4	2	6
알코올 중독 X ($X = 0$)	46	98	144
합	50	100	150

관측치를 랜덤하게 선택하지 않고,
전체 표본 중 심장질환자의 비율을 1/3로 고정해서 추출
➡ 비율의 차이와 상대위험도 대신에 **오즈비**를 사용

로그 오즈비 (Log Odds Ratio)

로그 오즈비란?

기존의 비대칭한 오즈의 범위를 교정한 측도



➡ 기존 오즈비의 기준인 1이 0이 되면서 범위가 **대칭**으로 교정

$$-\infty < \log \theta < \infty$$

오즈비의 장점

오즈비 = 교차적비(cross-product ratio)

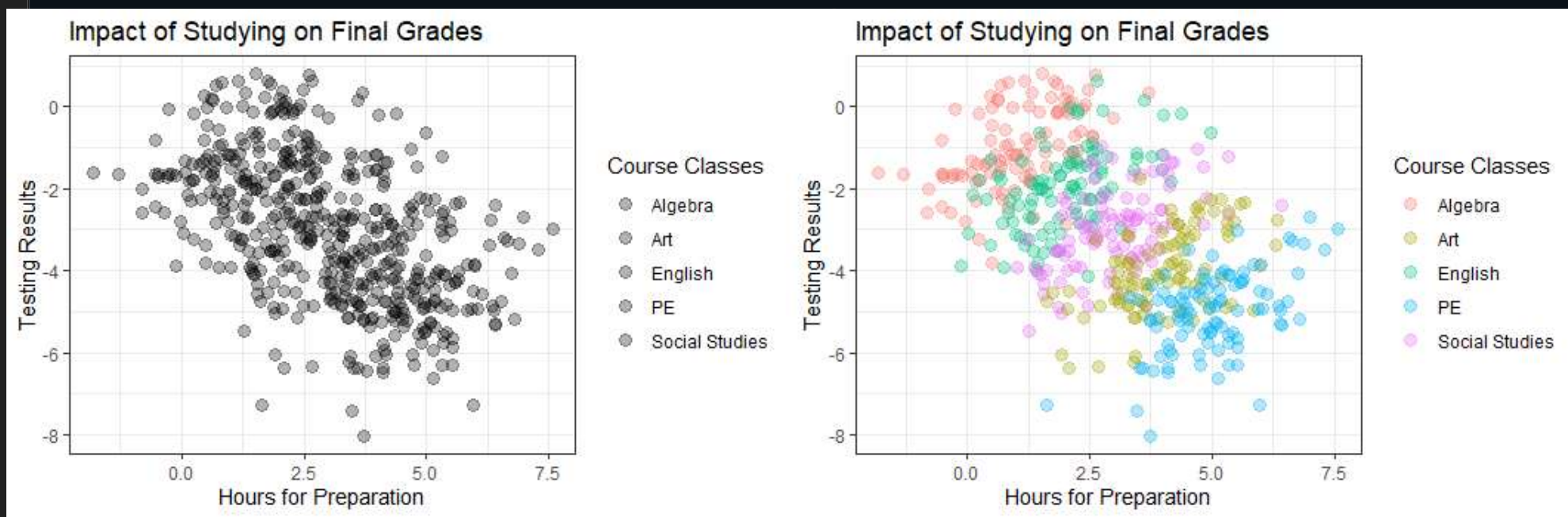
대각선에 있는 칸 도수 혹은 확률의 곱

$$\theta = \frac{\pi_{11}/\pi_{12}}{\pi_{21}/\pi_{22}} = \frac{\pi_{11}\pi_{22}}{\pi_{12}\pi_{21}} = \frac{n_{11}n_{22}}{n_{12}n_{21}}$$

정말 맛있는 오즈비 ...



조건부 독립성과 주변독립성이 성립한다고 해서
주변독립성이 성립되는 것은 아님!
제어변수 Z를 합친 주변분할표



주변 오즈비가 1일 때 $[\theta_{xy+} = 1]$ 주변독립성을 가짐

이 그래프를 해석해보자!