

[과목명] 01강. 주차 명 입력해 주세요.

01강. 서론, 이차원 분할표(1)

■ 주요용어

용어	해설
범주형 데이터 (Categorical data)	종속변수가 범주형인 데이터를 말함
최 대 가 능 도 추 정 (Maximum likelihood estimation)	미지인 모수의 함수로 표현된 관측자료의 확률을 최대로 하는 값을 추정값으로 결정하는 추정방법
두 변수의 독립성	X 의 각 수준에서 Y 의 조건부분포가 동일할 때 두 변수 X, Y 는 통계적으로 독립이라고 함. 두 변수 X, Y 가 모두 범주형 변수일 때 독립성은 모든 칸에서 결합확률과 이에 대응하는 주변확률의 곱이 같아지는 것임
상대 위험 (Relative risk)	2X2 분할표에서 두 집단의 “성공” 확률의 비(ratio) $\frac{\pi_1}{\pi_2}$ 를 의미함.
오즈비(Odds ratio)	2X2 분할표에서 행과 열의 연관성에 관한 척도로 각 행에 대하여 오즈값의 비(ratio)를 말함. 행과 열이 서로 독립이면 오즈비는 1이 되며, 1에서 멀어질수록 행과 열의 연관성이 높다고 해석

정리하기

1. 최대가능도추정(maximum likelihood estimator)

- 가능도함수가 최대값을 갖게 하는 모수(parameter) 값을 추정값으로 정하는 추정법
- 표본크기가 클 때 ML추정량은 최적(optimal)의 추정량이고, 근사적으로 정규분포를 따름

2. 두 변수의 독립성

- X와 Y는 통계적 독립 \Leftrightarrow Y의 조건부 확률이 X의 각각의 수준에서 동일
 $\Leftrightarrow \pi_{ij} = \pi_{i+} \cdot \pi_{+j}$, 모든 i, j 에 대하여

3. 상대위험도(Relative risk)

- 상대위험도 $= \frac{\pi_1}{\pi_2}$
- 두 비율들이 모두 0에 가까울 때 비율의 차이만으로 두 집단을 비교하는 것은 잘못된 결론을 가져올 수 있음

4. 오즈비(odds ratio)

- $\theta = \frac{\pi_1/(1-\pi_1)}{\pi_2/(1-\pi_2)}$
- 두 변수 X와 Y가 서로 독립이면 $\pi_1 = \pi_2$, $odds_1 = odds_2$ 을 만족하여 결과적으로 $\theta = 1$ 임
- 오즈비 θ 는 1로부터 멀리 떨어질수록 더 강한 연관성을 나타냄
- 행의 순서가 바뀌거나 열의 순서가 바뀌면 오즈비는 역수가 됨
- 분할표에서 행이나 열을 서로 바꾸더라도 오즈비는 변하지 않음

5. 오즈비와 상대위험의 관계

- 오즈비 $\theta = \frac{\pi_1/(1-\pi_1)}{\pi_2/(1-\pi_2)} = \text{상대위험도} \times \frac{1-\pi_2}{1-\pi_1}$
- π_1 과 π_2 가 모두 0에 가까우면 $\theta = \frac{\pi_1/(1-\pi_1)}{\pi_2/(1-\pi_2)} \approx \frac{\pi_1}{\pi_2}$ (상대위험도)

6. 연구방법의 구분

- 관측연구(observational studies), 실험연구(experimental studies)
- 전향적 연구(prospective study), 횡단면적 연구(cross-sectional study), 후향적 연구(restrospective study)

[과목명] 01강. 주차 명 입력해 주세요.

	과제하기
--	-------------

구분	내용											
과제 주제	1. 2015년에 미국 플로리다주 고속도로 안전국과 자동차국에 의해 수집된 사고 기록이다. 표본오즈비, 상대위험도와 비율의 차이를 계산하고 해석하시오. 그리고 오즈비와 상대위험도가 근사적으로 같은 이유를 설명하시오. R이나 다른 통계패키지를 이용하여 분석하시오.											
	<table><tr><th rowspan="2">안전장치</th><th colspan="2">치명적 사고</th></tr><tr><th>예</th><th>아니오</th></tr><tr><td>사용 안함</td><td>433</td><td>8,049</td></tr><tr><td>안전벨트</td><td>570</td><td>554,883</td></tr></table>	안전장치	치명적 사고		예	아니오	사용 안함	433	8,049	안전벨트	570	554,883
	안전장치		치명적 사고									
		예	아니오									
	사용 안함	433	8,049									
안전벨트	570	554,883										
2. 참고문헌에서 제시한 “오즈비와 상대위험도 관련 자료” 중 하나의 논문을 읽고 그 내용과 자기 생각을 A4 용지 1매로 작성하시오.												
3. 두 처리 (A, B)와 반응변수 (사망, 생존)간의 오즈비가 2.0이라고 하자. (1) “처리 A의 사망 확률은 처리 B에 비하여 2배이다”라는 해석이 왜 틀렸는지 설명하고 옳은 해석을 하시오. (2) (a)에서의 해석이 근사적으로 맞는 경우는 어느 경우인가? (3) 처리 A의 사망 오즈가 0.5일 때 처리 A와 처리 B의 사망 확률을 구하시오.												
목적	1주차 강의 내용을 복습하고, 오즈비와 상대위험도에 대한 심층적인 이해를 목적으로 함.											
제출 기간	1주차 강의 후 1주 후 토요일 밤 10시까지											
참고 자료	범주형 자료분석에서 R 사용방법 관련 자료는 참고문헌에 제시한 자료를 참고하기 바람											
기타 유의사항												