

오즈와 오즈비, 왜 사용하지?

7월 06, 2015

로지스틱 회귀분석 등 범주형 데이터를 이용한 분석에서 오즈(odds)와 오즈비(odds ratio)라는 용어를 자주 접하게 된다. 어려운 개념이 아니지만 조금 혼란스러워하는 부분이 있어 간단하게 정리를 해본다.

먼저, 오즈는 $p/(1-p)$ 로써 실패할 확률 대비 성공 확률로 (또는 실패 횟수 대비 성공 횟수) 계산이 된다. 즉 실패 대비 성공의 비율을 말한다. *p: 성공 확률*

오즈비는 두 방법, 즉 방법 1과 방법 2의 각 오즈의 비율로 계산된다. $\{p1/(1-p1)\} / \{p2/(1-p2)\}$

이렇게 계산된 오즈비로 성공은 실패 대비 방법 1이 몇배 더 효과적인가를 알 수 있다. *방법 2 대비*

그런데 왜 이런 조금은 복잡해(?) 보이는 오즈비를 사용하는가?

그 이유는 예를 들어 방법 1과 방법 2를 적용할 때 일반적으로 모집단 크기를 알 수 없고, 단지 표본의 크기를 선택하여 성공과 실패 횟수를 얻을 수 있기 때문에 통계처리시 많이 이용하게 된다.

Odds와 Odds Ratios 이해하기

Making Sense of Odds and Odds Ratios

David A. Grimes, MD, and Kenneth F. Schulz, PhD, MBA

많은 의사들에게 있어 승산(Odds)과 승산비(Odds Ratios, 이하 OR)는 이해하기 어렵다. Odds는 "어떤 사건이 일어날 확률을 그 사건이 일어나지 않을 확률로 나눈 것"이다. Odds Ratio는 "한 그룹에서의 odds를 다른 그룹에서의 odds로 나눠준 것이다" (예를 들어 약을 복용한 그룹의 odds/약을 복용하지 않은 그룹의 Odds)

로지스틱 회귀분석

• 오즈비(Odds ratio)

• 왜 이런 복잡한 걸 쓸까?

- 오즈(odds)와 단순 확률(p)는 거의 같음
- 오즈를 오즈로 나누면 비교가 가능하다

• 비도심지역 대비 도심지역 발병 오즈비(odds ratio) = $\frac{1}{\frac{1}{7999}} \approx 4.0015$

- 비도심지역에 1명의 희귀질병 환자가 발병할 경우, 도심지역에는 4명 발생

• 여기에 log를 붙이면 더 활용범위가 높아진다

- $0 < p < 1$ (여기서 p는 단순확률) & $0 < 1-p < 1$
- p가 "0"에 가까울 수록 오즈비(odds ratio)는 "0"을 향해 접근
- p가 "1"에 가까울 수록 오즈비(odds ratio)는 "무한대"를 향해 접근
- $0 < \text{오즈비(odds ratio)} < \text{무한대} \rightarrow -\text{무한대} < \log(\text{오즈비}) = \text{로짓} < +\text{무한대}$

오즈비 해석