

사례-대조 연구에서의 비율 비교

서울시립대 통계학과 이용희

FALL 2018

1 사례대조 연구의 자료 구조

심장발작을 일으킨 환자와 그렇지 않은 사람들을 각각 214명씩 조사하여 과거에 약물남용을 한 경력이 있는지 조사한 사례-대조 연구의 자료이다 (교재 예제 3-2 자료)

		심장발작	
		유	무
약물남용	유	73	18
	무	141	196
합		214	214

이 연구의 목표는 약물남용과 심장발작의 연관성이 있는지를 알아보는 것이다. 이제 다음과 같은 사건들을 정의해 보자.

- $H+$: 심장발작이 발생했다.
- $H-$: 심장발작이 발생하지 않았다.
- $D+$: 약물남용을 했다.
- $D-$: 약물남용을 하지 않았다.

위에서 정의된 사건들을 고려할 때 사례-대조 연구의 자료에서 다음과 같은 조건부 확률에 대한 추정값을 구할 수 있다.

$$P(\text{약물남용을 했다}|\text{심장발작이 발생했다}) = P(D+|H+) = \frac{73}{214}$$

$$P(\text{약물남용을 하지 않았다}|\text{심장발작이 발생했다}) = P(D-|H+) = 1 - P(D+|H+) = \frac{141}{214}$$

$$P(\text{약물남용을 했다}|\text{심장발작이 발생하지 않았다}) = P(D+|H-) = \frac{18}{214}$$

$$P(\text{약물남용을 하지 않았다}|\text{심장발작이 발생하지 않았다}) = P(D-|H-) = 1 - P(D+|H-) = \frac{196}{214}$$

2 사례대조 연구의 목표와 가설

연구에서 비교하고 싶은 비율은 위에서 추정할 확률이 아니고 조건과 결과와 바뀐 다음과 같은 조건부 확률이다.

$$\begin{aligned} P(\text{심장발작이 발생했다}|\text{약물남용을 했다}) &= P(H+|D+) \\ P(\text{심장발작이 발생했다}|\text{약물남용을 하지 않았다}) &= P(H+|D-) \end{aligned}$$

즉 연구의 목표는 다음과 같은 가설을 검정하는 것이다.

$$H_0 : P(H+|D+) = P(H+|D-) \quad \text{vs} \quad H_1 : P(H+|D+) \neq P(H+|D-) \quad (1)$$

전체 모집단을 약물남용을 한 사람들과 하지 않은 사람들로 두 집단으로 나누었을 때 두 집단에 대한 심장발작의 확률이 같은지 다른지 비교하고 싶은 것이다.

위의 식에서 보듯이 추정하고 싶은 확률인 $P(H+|D+)$ 와 $P(H+|D-)$ 를 추정하려면 전체 모집단에 대한 심장발작 발병률 $P(H+)$ 와 약물남용의 비율 $P(D+)$ 를 알아야 한다. 즉

$$\begin{aligned} P(H+|D+) &= \frac{P(H+ \cap D+)}{P(D+)} \\ &= \frac{P(D+|H+)P(H+)}{P(D+)} \\ &\approx (73/214) \frac{P(H+)}{P(D+)} \end{aligned}$$

위의 식은 다음의 조건부 확률 공식을 각 단계마다 적용한 결과이다.

$$P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$$

사례-대조 연구의 자료만으로는 모집단에 대한 심장발작 발병률 $P(H+)$ 와 약물남용의 비율 $P(D+)$ 을 구할 수 없다. 또한 다른 외부의 자료가 있다 하더라도 약물남용의 비율을 정확하게 추정하는 것은 매우 어렵다.

3 오즈비의 비교

이러한 문제는 두 집단의 비율의 차이나 상대위험을 비교하지 않고 오즈비를 구하여 비교하면 심장발작 발병률과 약물남용의 비율을 추정하지 않고 사례-대조 연구의 자료만으로 추론이 가능하다.

다음의 가설은 두 비율의 비교를 오즈비로 표현한 것이다.

$$H_0 : \frac{P(H+|D+)/P(H-|D+)}{P(H+|D-)/P(H-|D-)} = 1 \quad \text{vs} \quad H_1 : \frac{P(H+|D+)/P(H-|D+)}{P(H+|D-)/P(H-|D-)} \neq 1 \quad (2)$$

위의 가설 (2)는 단순한 비율을 비교하는 가설 (1)과 동일한 가설이다.

가설 (2)에서 나타는 오즈비는 심장발작 발병률과 약물남용의 비율을 이용하지 않고 사례-대조 연구에서 추정할 수 있는 조건부 확률만으로 추정할 수 있다.

$$\begin{aligned}
\frac{P(H+|D+)/P(H-|D+)}{P(H+|D-)/P(H-|D-)} &= \frac{[P(H+|D+)P(D+)]/[P(H-|D+)P(D+)]}{[P(H+|D-)P(D-)]/[P(H-|D-)P(D-)]} \\
&= \frac{P(H \cap D+)/P(H- \cap D+)}{P(H \cap D-)/P(H- \cap D-)} \\
&= \frac{[P(D+|H+)P(H+)]/[P(D+|H-)P(H-)]}{[P(D-|H+)P(H+)]/[P(D-|H-)P(H-)]} \\
&= \frac{P(D+|H+)/P(D+|H-)}{P(D-|H+)/P(D-|H-)} \\
&= \frac{(73/214)/(142/214)}{(18/214)/(196/214)} \\
&= \frac{(73)(196)}{(141)(18)} \\
&= 5.64
\end{aligned}$$

결론적으로 사례-대조 연구에서는 연구의 목표에 대한 가설 검정을 비율의 차이나 상대위험으로 표현하여 수행할 수 없다. 하지만 오즈비를 검정하는 것으로 가설을 세우면 자료에서 쉽게 유도할 수 있는 오즈비로 가설 검정을 쉽게 수행할 수 있다.