의학통계학 과제 1

이다연

2018. 09. 21

Problem 1

산후 우울증에 대한 연구에서, 연구자는 초산인 여성 116명을 뽑아 산후 우울증에 대한 위험요인으로 배우자의 무관심이나 애정결핍을 고려하여 조사하고, 산후 1개월때 산후 우울증에 대해 조사하였다. 결과는 다음의 표와 같다.

위험요인	산후우울증		총계
	유	무	
유	5	21	26
무	8	82	90
총계	13	103	116

1. 상대위험률과 오즈비를 구하라.
2. 상대위험률과 오즈비에 대한 95% 신뢰구간을 구하라.
3. 위의 1, 2로부터 어떤 해석을 할 수 있는가?

Solution 1.1

$$\begin{aligned} \text{상대위험률 (RR)} &= \frac{n_{11}/n_{1+}}{n_{21}/n_{2+}} \\ &= \frac{5/26}{8/90} = 2.1635 \end{aligned}$$

$$\begin{aligned} \text{오즈비 (OR)} &= \frac{n_{11}/n_{12}}{n_{21}/n_{22}} = \frac{n_{11} * n_{22}}{n_{12} * n_{21}} \\ &= \frac{5/8}{21/82} = \frac{5 * 82}{21 * 8} = 2.4405 \end{aligned}$$

Solution 1.2

상대위험률 (RR) 95% 신뢰구간

$$\begin{aligned} v_1 &= \frac{1 - n_{11}/n_{1+}}{n_{11}} + \frac{1 - n_{21}/n_{2+}}{n_{21}} \\ &= \frac{1 - 5/26}{5} + \frac{1 - 8/90}{8} = 0.2754 \end{aligned}$$

이때,

$$\begin{aligned} &(RR \times \exp\{-z_{\alpha/2}\sqrt{v_1}\}, RR \times \exp\{z_{\alpha/2}\sqrt{v_1}\}) \\ &= (2.1635 \times \exp\{-1.96\sqrt{0.2754}\}, 2.1635 \times \exp\{1.96\sqrt{0.2754}\}) \\ &= (0.7735, 6.0515) \end{aligned}$$

오즈비 (OR) 95% 신뢰구간

$$\begin{aligned} v_2 &= \frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}} \\ &= \frac{1}{5} + \frac{1}{8} + \frac{1}{21} + \frac{1}{82} = 0.3848 \end{aligned}$$

이때,

$$\begin{aligned} &(OR \times \exp\{-z_{\alpha/2}\sqrt{v_2}\}, OR \times \exp\{z_{\alpha/2}\sqrt{v_2}\}) \\ &= (2.4405 \times \exp\{-1.96\sqrt{0.3848}\}, 2.4405 \times \exp\{1.96\sqrt{0.3848}\}) \\ &= (0.7235, 8.2320) \end{aligned}$$

Solution 1.3

상대위험률이 1보다 크므로 위험요인이 있을 때, 그렇지 않은 것에 비해 산후우울증이 일어날 위험이 2.1635배정도 커진다. 또한, 오즈비가 2.4405로 산후우울증이 있는 집단에서 위험요소를 가질 오즈가 산후우울증이 없는 집단에서 위험요소를 가질 오즈에 2.4404배이다.

상대위험률과 오즈비의 신뢰구간이 모두 1을 포함한다. 이는 위험요인이 있을 때와 없을 때 산후우울증 발생도에 유의한 차이가 있다고 말하기 어렵다. 따라서 산후우울증의 위험요인이 배우자의 무관심이나 애정결핍이라고 말하기에는 근거다 부족하다고 할 수 있다.

Problem 2

구강피임약의 복용이 유방암과 관련이 있는지 알아보기 위한 연구를 실시하였다. 이전에 구강피임약을 사용한 2914명의 여성 가운데에서 273명이 유방암 환자였으며, 구강피임약을 복용한 경력이 없는 7976명의 여성 가운데에서 유방암이 발생한 사람은 716명이었다. 구강피임약을 사용하는 것이 유방암 발생을 높인다고 할 수 있는지 오즈비와 95% 신뢰구간을 구하고 결과를 해석해라.

Solution 2

	유방암		
구강피임약	O	X	계
O	273	2641	2914
X	716	7260	7976
계	989	9901	10890

$$\text{오즈비 (OR)} = \frac{273 * 7260}{716 * 2641} = 1.0481$$

오즈비 (OR) 95% 신뢰구간

$$v_2 = \frac{1}{273} + \frac{1}{716} + \frac{1}{2641} + \frac{1}{7260} = 0.0056$$

$$(1.0481 \times \exp\big\{-1.96\sqrt{0.0056}\big\}, 1.0481 \times \exp\big\{1.96\sqrt{0.0056}\big\}) = (0.9054, 1.2133)$$

오즈비는은 1.0481로 1개 매우 가깝고, 오즈비의 95% 신뢰구간은 (0.9054, 1.2133)로 1을 포함하고 있으므로 구강피임약을 복용했을때와 복용하지 않았을 때 유방암 발병률이 유의한 차이가 있다고 할 수 없다. 따라서 구강피임약 복용이 유방암과 관련이 있다고 보기 어렵다.

Problem 3

한 비확률화 연구에서 125명의 췌장염 환자들중 61명은 참여를 거부했으며, 33명은 금주를 권고받았고, 31명은 수술을 시행하였다. 이들의 췌장염 재발여부를 표로 나타낸 것이 아래와 같다. 세 그룹간 재발률이 같은지 검정하라.

그룹	재발		총계
	있음	없음	
참여거부	49	12	61
금주권고	24	9	33
수술	2	29	31
총계	75	50	125

Solution 3

세 그룹간 재발률이 같은지 검정하기 위해 카이제곱 동질성 검정을 실시한다.

$$H_0 : P_{1j} = P_{2j} = P_{3j} (j = 1, 2) \text{ vs } H_1 : not H_0$$

귀무가설 하에서 기대도수표 $E_{ij} = \frac{n_{i+}n_{j+}}{n}$ 를 나타내면 다음과 같다.

E_{ij}	1	2
1	36.6	24.4
2	19.8	13.2
3	18.6	12.4

한편, 카이제곱 통계량은

$$\begin{aligned}
\chi^2 &= \sum_{i=1}^{3}\sum_{j=1}^{2}\{\frac{(n_{ij}-E_{ij})^2}{E_{ij}}\} \\
&= \frac{(49-36.6)^2}{36.6} + \frac{(12-24.4)^2}{24.4} + \frac{(24-19.8)^2}{19.8} \\
&+ \frac{(9-13.2)^2}{13.2} + \frac{(2-18.6)^2}{18.6} + \frac{(29-12.4)^2}{12.4} \\
&= 49.7677 \sim \chi^2(2)
\end{aligned}$$

위 통계량은 자유도가 2인 카이제곱 분포를 따르며, 유의수준 5%일 때 카이제곱 임계치는 5.99로 귀무가설을 기각한다.

$$\chi^2 = 49.7677 > \chi^2_{0.05}(2) = 5.99$$

따라서, 세 그룹의 재발률은 같다고 할 수 없다.

Problem 4

한 연구자는 분자 하이브리드형성에 의해 탐지되는 인간 유두종바이러스(human papilloma virus; HPV)에 이미 전염된 HIV(human immuno-deficiency virus) 전염 여성은 둘 중의 한 바이러스에 전염되거나 둘 다 전염되지 않은 사람보다 더 쉽게 세포학적 자궁경부 이상이 나타나는지에 대해 알기 위해 96명의 여성을 조사하였다. 이 결과는 다음의 표와 같다. HPV감염 여부와 HIV감염 여부에 대해 관계가 있는지 유의수준 5%로 검정하시오.

HPV	HIV			총계
	혈청반응양성 증후성	혈청반응양성 무증후성	혈청반응음성	
양성	23	4	10	37
음성	10	14	35	59
총계	33	18	45	96

Solution 4

HPV감염 여부와 HIV감염 여부에 대해 관계를 보기위해 카이제곱 독립성 검정을 실시한다.

$$H_0 : \text{HPV감염여부와 HIV감염여부는 서로 독립이다. vs } H_1 : not H_0$$

귀무가설 하에서 기대도수표 $E_{ij} = \frac{n_{i+}n_{j+}}{n}$ 를 나타내면 다음과 같다.

E_{ij}	1	2	3
1	12.71	6.94	17.34
2	20.28	11.06	27.66

한편, 카이제곱 통계량은

$$\begin{aligned} \chi^2 &= \sum_{i=1}^{2}\sum_{j=1}^{3}\{\frac{(n_{ij}-E_{ij})^2}{E_{ij}}\} \\ &= \frac{(23-12.71)^2}{12.71} + \frac{(4-6.94)^2}{6.94} + \frac{(10-17.34)^2}{17.34} \\ &+ \frac{(10-20.28)^2}{20.28} + \frac{(14-11.06)^2}{11.06} + \frac{(35-27.66)^2}{27.66} \\ &= 20.6061 \sim \chi^2(2) \end{aligned}$$

위 통계량은 자유도가 2인 카이제곱 분포를 따르며, 유의수준 5%일 때 카이제곱 임계치는 5.99로 귀무가설을 기각한다.

$$\chi^2 = 20.6061 > \chi^2_{0.05}(2) = 5.99$$

따라서, HPV 감염여부와 HIV 감염여부는 독립이 아니라 볼 수 있으며 서로 연관이 있다고 판단할 수 있다.

Problem 7

다음 자료는 투표권이 있는 1600명의 임의 표본을 대상으로 국무총리의 임무 수행에 관한 의견을 조사한 후 6개월 뒤에 동일한 1600명에 대하여 다시 의견을 조사한 자료이다. 이 자료에 대해서 첫 번째 투표와 두 번째 투표에서 찬성확률이 동일한지를 검정하여라.

첫 번째 투표	두 번째 투표		총계
	찬성	반대	
찬성	794	150	944
반대	86	570	656
총계	880	720	1600

Solution 7

동일한 대상에 대하여 전, 후 처리를 비교한 문제이므로 맥니마 검정을 실시한다.
첫번째 투표에서 찬성률을 p_1, 두번째 투표에서 찬성률을 p_2라 하자.

$$H_0 : p_1 = p_2 \text{ vs } H_1 : p_1 \neq p_2$$

이때, 맥니마 검정통계량(Q_M)은

$$\begin{aligned} Q_M &= \frac{\left(n_{12} - n_{21}\right)^2}{\left(n_{12} + n_{21}\right)} \\ &= \frac{\left(150 - 86\right)^2}{\left(150 + 86\right)} \\ &= 17.3559 \sim \chi^2\left(1\right) \end{aligned}$$

위 통계량은 자유도가 1인 카이제곱 분포를 따르며, 유의수준 5%일 때 카이제곱 임계치는 3.84로 귀무가설을 기각한다.

$$Q_M = 17.3559 > \chi^2_{0.05}\left(1\right) = 3.84$$

따라서 첫번째 투표에서 찬성률과 두번째 투표에서 찬성률은 다르다고 할 수 있다.

Problem 8

신생아에게 치명적인 적아구증(erythroblastosis)에 관한 자료를 조사하였다. 적아구증은 혈액형이 Rh+형인 신생아의 혈액 안에 있는 anti-Rh 항체로 인해서 발병한다. 적아구증을 치료하기 위한 방법은 anti-Rh 항체에 영향을 안주는 혈액을 수혈 받는것이다. 혈액 기증자들의 성별과 환자들의 질병의 심각도에 따른 생존 상태 자료가 아래 표와 같다. 위의 자료를 바탕으로 기증자의 성별과 생존 상태의 연관성을 코크란-맨텔-핸젤방법을 이용하여 검정하라.

질병의 심각도	기증자 성별	생존 상태		총계
		사망	생존	
심각하지 않음	남 여	2 0	21 10	23 10
약간 심각	남 여	2 0	40 18	42 18
심각	남 여	6 0	33 10	39 10
매우 심각	남 여	17 0	16 4	33 4
총계		27	152	179

Solution 8

층 변수는 질병의 심각도로 4개의 층이 있고 기증자 성별이 남자일 때 사망률을 p_1, 여자일 때 사망률을 p_2라 하자. 코크란-맨텔-핸젤 검정의 가설은 다음과 같다.

$$H_0 : p_1 = p_2 \text{ vs } H_1 : p_1 \neq p_2$$

코크란-맨텔-핸젤 검정 통계량이 다음과 같을때,

$$Q_{CMH} = \Big\{\sum_{k=1}^{4} n_{k11} - \sum_{k=1}^{4} E\big(n_{k11}|H_0\big)\Big\}^2 \Big/ \sum_{k=1}^{4} Var\big(n_{k11}|H_0\big)$$

$$\sum_{k=1}^{4} n_{k11} = 2 + 2 + 6 + 17 = 27$$

$$\begin{aligned}
\sum_{k=1}^{4} E\big(n_{k11}|H_0\big) &= \sum_{k=1}^{4} \frac{n_{k1+}n_{k+1}}{n_k} \\
&= \frac{23*2}{33} + \frac{42*2}{60} + \frac{39*6}{49} + \frac{33*17}{37} \\
&= 22.7316
\end{aligned}$$

$$\begin{aligned}\sum_{k=1}^{4} Var\big(n_{k11}|H_0\big) &= \sum_{k=1}^{4} \frac{n_{k1+}n_{k2+}n_{k+1}n_{k+2}}{n_k^2\big(n_k-1\big)} \\ &= \frac{23*10*2*31}{33^2*\big(33-1\big)}+\frac{42*18*2*58}{60^2*\big(60-1\big)} \\ &+ \frac{39*10*6*43}{49^2*\big(49-1\big)}+\frac{33*4*17*20}{37^2*\big(37-1\big)} \\ &= 2.6058\end{aligned}$$

$$\begin{aligned}Q_{CMH} &= \big\{\sum_{k=1}^{4} n_{k11}-\sum_{k=1}^{4} E\big(n_{k11}|H_0\big)\big\}^2/\sum_{k=1}^{4} Var\big(n_{k11}|H_0\big) \\ &= \frac{\big(27-22.7316\big)^2}{2.6058} \\ &= 6.9918 \sim \chi^2\big(1\big)\end{aligned}$$

위 통계량은 자유도가 1인 카이제곱 분포를 따르며, 유의수준 5%일 때 카이제곱 임계치는 3.84로 귀무가설을 기각한다.

$$Q_{CMH} = 6.9918 > \chi^2_{0.05}\big(1\big) = 3.84$$

따라서 질병의 심각도를 층변수로 제어했을때, 남자일 때의 사망률과 여자일때의 사망률에는 차이가 있다. 즉, 기증자의 성별과 생존 상태는 서로 연관성이 있다고 볼 수 있다.