의학통계학 과제 4

이다연

2018. 11. 21

[문제 1번]

확률변수 X와 Y는 각각 독립적으로 다음과 같은 이항분포를 따른다.

$$X \sim Bin(n, p_1) \quad Y \sim Bin(n, p_2)$$

다음과 같은 가설을 검정하려고 한다.

$$H_0 : p_1 = p_2 \quad vs \quad H_0 : p_1 \neq p_2$$

유의수준은 5%, 검정력은 80%로 정한 경우 $p_1 = 0.4$, $p_2 = 0.2$일 때 필요한 연구대상자 수 n을 결정하시오.

1.

$$\alpha = 0.05, \quad 1 - \beta = 0.8$$
$$p_1 = 0.4, \quad p_2 = 0.2$$
$$Z_{\alpha/2} = 1.96, \quad Z_\beta = 0.84$$

이제 두 비율을 비교할 때 필요한 표본의 수를 구하는 공식은 다음과 같이 (1) 두 비율의 비교에서 검정 통계량을 정확하게 이용하는 교과서의 방법과 (2) 평균의 비교에서 사용되는 공식을 근사적으로 이용하는 방법이 있다.

1. 교과서 방법 - 두 비율의 검정 통계량을 정확하게 이용하는 경우

두 비율을 비교하는 경우는 사용하는 검정통계량은 다음과 같다.

$$Z^* = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\bar{p}(1-\bar{p}_1)(1/n+1/n)}} = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{v_1(1/n+1/n)}}$$

여기서 $\bar{p} = (p_1 + p_2)/2$ 이다.

위의 식과 다음의 계산에서 편의를 위하여 다음과 같이 v_1과 v_2를 정의하자.

$$v_1 = \bar{p}(1-\bar{p}) \quad v_2 = p_1(1-p_1) + p_2(1-p_2)$$

따라서 $p_1 - p_2 = \delta > 0$ 가정하에서의 검정력을 $1-\beta$로 하는 방정식은

$$\begin{aligned}
1-\beta &= P(Z^* > Z_{\alpha/2} | p_1 - p_2 = \delta) \\
&= P\left[\frac{\hat{p}_1 - \hat{p}_2}{\sqrt{v_1(1/n+1/n)}} > Z_{\alpha/2} | p_1 - p_2 = \delta\right] \\
&= P\left[\hat{p}_1 - \hat{p}_2 > Z_{\alpha/2}\sqrt{2v_1/n} | p_1 - p_2 = \delta\right] \\
&= P\left[(\hat{p}_1 - \hat{p}_2) - \delta > Z_{\alpha/2}\sqrt{2v_1/n} - \delta | p_1 - p_2 = \delta\right] \\
&= P\left[\frac{\hat{p}_1 - \hat{p}_2 - \delta}{\sqrt{p_1(1-p_1)/n + p_2(1-p_2)/n)}} > \frac{Z_{\alpha/2}\sqrt{2v_1/n} - \delta}{\sqrt{p_1(1-p_1)/n + p_2(1-p_2)/n}} | p_1 - p_2 = \delta\right] \\
&= P\left[Z > \frac{Z_{\alpha/2}\sqrt{2v_1/n} - \delta}{\sqrt{v_2/n}}\right] \\
&= 1 - P\left[Z < \frac{Z_{\alpha/2}\sqrt{2v_1/n} - \delta}{\sqrt{v_2/n}}\right]
\end{aligned}$$

위의 방정식에서 표본의 수 공식을 다음과 같이 유도할 수 있다.

$$\frac{Z_{\alpha/2}\sqrt{2v_1/n} - \delta}{\sqrt{v_2/n}} = -Z_\beta$$

따라서 표본의 수는 다음과 같이 주어진다.

$$n = \frac{\left[Z_{\alpha/2}\sqrt{2v_1} + Z_\beta\sqrt{v_2}\right]^2}{\delta^2} = \frac{\left[Z_{\alpha/2}\sqrt{2\bar{p}(1-\bar{p})} + Z_\beta\sqrt{p_1(1-p_1) + p_2(1-p_2)}\right]^2}{(p_1 - p_2)^2}$$

따라서,

$$
\begin{aligned}
n &= \frac{\left[Z_{\alpha/2}\sqrt{2\bar{p}(1-\bar{p})} + Z_{\beta}\sqrt{p_1(1-p_1)+p_2(1-p_2)}\right]^2}{\delta^2} \\
&= \frac{\left[(1.96)\sqrt{2(0.3)(0.7)} + (0.84)\sqrt{(0.4)(0.6)+(0.2)(0.8)}\right]^2}{(0.2)^2} \\
&= 81.13396
\end{aligned}
$$

따라서, $n = 82$이므로 필요한 환자 수는 $82 \times 2 = 164$명이다.

2. 평균을 비교하는 공식을 근사적으로 이용하는 방법

두 비율을 비교하는 경우 두 평균을 비교하는 식을 응용하면 표본의 수를 계산하는 공식을 유도할 수 있다.두 평균을 비교할 때 표본의 수 n에 대한 방정식은 다음과 같다.

$$
1 - P\left[Z < Z_{\alpha/2} - \frac{\delta}{\sigma\sqrt{2/n}}\right] = 1 - \beta
$$

위의 식에서 표준정규분포를 따르는 확률변수 Z는 두 평균을 비교할 때 다음과 같다.

$$
Z = \frac{\bar{X} - \bar{Y} - \delta}{\sqrt{Var(\bar{X} - \bar{Y})}} = \frac{\bar{X} - \bar{Y} - \delta}{\sqrt{\sigma^2/n + \sigma^2/n}} = \frac{\bar{X} - \bar{Y} - \delta}{\sigma\sqrt{2}/\sqrt{n}}
$$

즉 비율을 비교하는 경우에는 평균에 대한 표본의 수 공식에서 $\sigma\sqrt{2}$를 $\sqrt{p_1(1-p_1)+p_2(1-p_2)}$로 바꾸면 된다.

따라서,

$$
\begin{aligned}
n &= \frac{2(Z_{\alpha/2} + Z_{\beta})^2\sigma^2}{\delta^2} \\
&= \frac{(Z_{\alpha/2} + Z_{\beta})^2(p_1(1-p_1)+p_2(1-p_2))}{(p_1-p_2)^2} \\
&= \frac{(1.96 + 0.84)^2(0.4)}{0.2^2} \\
&= 78.49
\end{aligned}
$$

따라서, $n = 79$이므로 필요한 환자 수는 $79 \times 2 = 158$명이다. 이 숫자는 위의 교과서에서 나오는 공식으로 구한 표본의 수와 거의 유사하다.

[문제 2번] 교재 p.290 연습문제 1, 2, 3, 5

연습문제 1

표본의 크기(연구대상환자의 수)는 무엇에 영향을 받는가?

1.

$$n = \frac{2(Z_{\alpha/2} + Z_\beta)^2\sigma^2}{\delta^2}$$

이므로 표본의 크기는 유의수준 α, 검정력 $1-\beta$, 평균의 차이 δ, 표준편차 σ에 영향을 받는다.

연습문제 2

실험군에서의 반응률이 $0.4(=\ p_1)$이고, 대조군에서의 반응률이 $0.2(=\ p_2)$일 때 실험군과 대조군의 반응률이 동일한지를 검정하려 한다. 유의수준 α를 5%, 검정력 $1-\beta$를 90%로 하고 단측검정을 실시한다고 하자. 두 그룹에 같은 수의 환자를 할당한다고 할 때 필요한 환자수를 계산하여라.

다음과 같은 단측검정에서는

$$H_0 : p_1 = p_2 \quad \text{vs.} \quad H_1 : p_1 > p_2$$

기각역이 단측이다. 즉 기각역은

$$\text{rejection region } = \{|z| > z_\alpha\}$$

따라서 양측검정의 표본의 수 공식에서 $z_{\alpha/2}$를 z_α로 바꾸어 주면 단측검정에 대한 표본의 수 공식이 된다.

2.

$$\alpha = 0.05, \quad 1-\beta = 0.9$$
$$p_1 = 0.4, \quad p_2 = 0.2$$
$$Z_\alpha = 1.64, \quad Z_\beta = 1.28$$

$$n = \frac{2(Z_\alpha + Z_\beta)^2\sigma^2}{\delta^2} = \frac{(Z_\alpha + Z_\beta)^2(p_1(1-p_1) + p_2(1-p_2))}{(p_1 - p_2)^2} = \frac{(1.64 + 1.28)^2(0.4)}{0.2^2} = 85.64$$

따라서, $n = 86$이므로 필요한 환자 수는 $86 \times 2 = 172$명이다.

연습문제 3

모 제약회사에서는 혈당량 수치를 낮추는 두 개의 약을 비교하고자 한다. 새로운 약으로 처리 받은 집단의 혈당량 수치가 기존의 약을 처리 받은 집단에 비해 $50mg/dl$만큼 차이가 있음을 보이고자 한다. 이때 각 그룹의 분산은 $(15mg/dl)^2$로 추정되었다. 유의수준 $\alpha = 0.05$으로 하고 검정력은 80%로 하여 양측검정을 할 때 필요한 환자수를 계산하여라.

3.

$$\alpha = 0.05, \quad 1 - \beta = 0.8$$
$$\delta = 50, \quad \sigma^2 = 15^2$$
$$Z_{\alpha/2} = 1.96, \quad Z_\beta = 0.84$$

$$n = \frac{2(Z_{\alpha/2} + Z_\beta)^2\sigma^2}{\delta^2} = \frac{2(Z_{\alpha/2} + Z_\beta)^2}{(\delta/\sigma)^2} = \frac{2(1.96 + 0.84)^2}{(50/15)^2} = 1.41$$

따라서, $n = 2$이므로 필요한 환자 수는 $2 \times 2 = 4$명이다.

연습문제 5

한 연구에서는 유도분만으로 인한 고혈압(hypertension)과 독혈증(toxaemia)을 억제하는 치료제의 영향을 알아보고자 이러한 증상을 갖고 있는 임산부들을 대상으로 고혈압의 위험률을 측정하였다.

(1) 고혈압의 위험률이 30%에서 20%로 감소하는 것을 80%의 검정력으로 유의하다고(p-값<0.05) 말할 수 있으려면 필요한 표본수는 얼마인가?

(2) 실제 표본수는 65이었다. 그렇다면 (1)에서의 위험률의 차이를 검정하기 위한 검정력은 얼마나 되겠는가?

5.1

$$\alpha = 0.05, \quad 1-\beta = 0.8$$
$$p_1 = 0.3, \quad p_2 = 0.2$$
$$Z_{\alpha/2} = 1.96, \quad Z_\beta = 0.84$$

$$n = \frac{2(Z_{\alpha/2}+Z_\beta)^2\sigma^2}{\delta^2} = \frac{(Z_{\alpha/2}+Z_\beta)^2(p_1(1-p_1)+p_2(1-p_2))}{(p_1-p_2)^2} = \frac{(1.96+0.84)^2(0.37)}{0.1^2} = 290.08$$

따라서, $n = 291$이므로 필요한 환자 수는 $291 \times 2 = 582$명이다.

5.2

$$\begin{aligned}
\text{검정력 (power)} &= 1-\beta \\
&= 1 - P\big(Z < Z_{\alpha/2} - \frac{\delta}{\sigma\sqrt{2/n}}\big) \\
&= 1 - P\big(Z < 1.96 - \frac{0.1}{\sqrt{0.37}\sqrt{1/65}}\big) \\
&= 1 - 0.74 \\
&= 0.26
\end{aligned}$$

따라서, 검정력은 26%이다.