

연구대상자수의 결정

이용희

November 2, 2018

Contents

1	통계적 가설검정과 오류	2
2	평균의 차이와 검정력의 정의	3
3	가설검정의 기각역	4
4	검정력의 계산	5
5	연구대상자수 n 의 결정	8

임상시험을 수행하기 위해서는 표본의 수, 즉 연구대상자를 실험 전에 먼저 정하고 실험을 해야한다. 교과서에서도 언급되었듯이 많은 임상실험은 실험전에 필요한 연구대상자의 수를 미리 정하고 실험을 진행한다. 엄격한 증거가 요구되는 임상실험에서는 원하고자 하는 결과가 나올 때 까지 실험중에 대상자수를 늘리는 것이 일반적으로 받아들여 지지 않는다. 또한 충분하지 않은 소수의 대상자만으로 얻은 결론 또한 받아들여 지기 힘들다.

그러면 실험 전에 어떻게 필요한 연구대상자의 수를 결정할 수 있을까?

1 통계적 가설검정과 오류

연구대상자의 수를 실험전에 결정하기 위해서는 모집단에 대한 두 가지 정보가 필요하다. 이제 통계학에서 나오는 가장 기본적이고 가설검정인 두 정규분포의 평균의 차이를 검정하는 문제에서 연구대상자의 수를 결정하는 절차를 알아보자.

확률변수 X 는 정규분포 $N(\mu_1, \sigma^2)$ 를 따르고 Y 는 $N(\mu_2, \sigma^2)$ 를 따른다고 가정하자. 그리고 이제 다음과 같은 가설검정을 하려고 한다.

$$H_0 : \mu_1 = \mu_2 \quad \text{vs.} \quad H_1 : \mu_1 \neq \mu_2 \quad (1)$$

가설 (1)은 임상실험에서 가장 자주 사용되는 가설이다. 예를 들어 새로운 약을 개발하면 제일 먼저 해야할 일은 새로운 약이 병을 치료하는 유의한 효과가 있다는사실을 통계적으로 보여야 한다. 즉, 위약을 사용한 집단과 새로운 약을 사용한 집단의 평균이 다르다는 가설을 검정해야 한다.

참고로 위의 가설은 새로운 약의 효과가 위약의 효과와 다르다는 가설을 검정하는 것으로 대립가설 H_1 은 신약이 좋은 경우(예를 들어 $\mu_1 > \mu_2$)와 위약이 더 좋은 경우(예를 들어 $\mu_1 < \mu_2$)인 경우를 모두 포함하고 있다. 이러한 대립가설은 실험의 목적인 신약의 효과(effective)를 보이는 목적과는 약간 차이가 난다. 즉, 논리적으로 실험의 목적에 적합한 대립가설은 단측가설(one-sided hypothesis)인 $H_1 : \mu_1 > \mu_2$ 가 되어야 한다. 하지만 일반적인 임상실험에서는 통상적으로 대립가설을 양측가설인 가설 (1)로 놓고 가설검정을 하여 신약의 효과가 좋은 쪽으로 결과가 나타나면 성공이라고 판단한다(신약이 치료효과가 있다). 귀무가설을 기각하지 못하거나 위약이 더 효과적이라고 결론이 나오면 임상실험의 결과는 실패이다

이제 가설검정에서 발생할 수 있는 오류에 대해 알아보자. 다음의 표는 가설의 상태와 검정의 결과에 대하여 발생할 수 있는 오류들을 나타낸 표이다.

Table 1: Two types of error

가설의 상태	검정의 결론(action)	
	H_0 기각	H_0 채택
H_0 참	제 1 종 오류	ggod!
H_0 거짓	good!	제 2 종 오류

위의 표에서 알 수 있듯이 H_0 가 참인 경우 이를 기각하는 오류를 **제 1 종의 오류 (Type I error)**라고 한다. 임상실험의 목적으로 보면 신약이 효과가 없는데 효과가 있다고 잘못된 결론을 내리는 경우이다 (false positive). 또한 H_0 가 거짓인 경우 이를 기각하지 못하는 오류를 **제 2 종의 오류 (Type II error)**라고 하며 이는 신약이 효과가 있는데 효과가 없다고 잘못된 결론을 내리는 경우이다 (false negative).

이렇게 가설검정에서는 두 가지 오류를 범할 수 있는데 통계적 가설검정법은 일반적으로 제 1종 오류를 범할 확률을 유의수준 $\alpha = 0.05$ 보다 작게 하면서 제 2 종 오류를 범할 확률을 최소화 하도록 만들어 졌다. 이론 상 두 개의 오류를 범할 확률을 동시에 최소화 하지 못하므로 제 1종 오류를 범할 확률을 주어진 기준(유의수준)보다 작게 하고 제 2 종 오류를 범할 확률을 최소화 하게 만든 것이다.

이는 제 1 종 오류가 더 중대한 오류이기 때문이며 이유는 제 1 종 오류가 발생하는 것은 효과가 없는 약을 효과가 있다고 결론내리는 것이기 때문이다. 제 2 종 오류가 발생하면 단지 효과적인 약이 허가를 받지 못하여 개발한 기관만 손해를 보지만 제 1 종 오류가 발생하면 효과없는 약이 시판되어 훨씬 더 큰 피해가 나기 때문이다.

2 평균의 차이 와 검정력의 정의

이제 가설검정이 약이 효과가 있는 경우, 즉 H_0 가 거짓인 이를 기각할 확률을 생각해보자. 이는 실제로 약이 효과가 있는 경우 검정 이를 뒷받침하는 결론을 내릴 확률이며 이를 크게 하는 것이 임상실험을 수행하는 기관이 바라는 것이다.

이러한 확률을 **검정력(power)**라고 하며 다음과 같이 정의된다.

$$\begin{aligned}
 power &= P(\text{reject } H_0 | H_0 \text{ is false}) \\
 &= 1 - P(\text{accept } H_0 | H_0 \text{ is false}) \\
 &= 1 - P(\text{Type II error}) \\
 &= 1 - \beta
 \end{aligned}$$

위의 식에서 보듯이 제 1 종 오류를 범할 확률을 β 라고 하면 검정력은 $1 - \beta$ 이다.

이제 이러한 확률을 계산하려면 H_0 가 참이 아닌 경우를 가정해야 하는데 대립가설 (1)에서는 두 평균이 같지 않다고 가정하였기 때문에 두 평균의 차이 $\delta = \mu_1 - \mu_2$ 가 가질 수 있는 가능한 값은 무수히 많다. 검정력을 구하려면 이러한 모든 경우를 고려할 수 없으며 특정한 값이 주어져야 한다.

임상실험은 전체적으로 제 1상에서부터 제 3 상까지 일련의 많은 실험들로 구성되고(교과서 246-247 참조) 약의 효과를 보이는 비교임상실험을 수행하는 마지막 3 상 단계에서는 신약이 얼마나 효과적인지에 대한 정보를 이전의 실험으로부터 유추할 수 있다. 신약의 독성이나 부작용을 알아보는 제 1상의 실험들과 약이 최소한의 효과가 있는지에 대한 제 2 상 실험들에서 어느 정도 약의 효과에 대한 정보를 얻을 수 있다. 즉, 두 모집단의 평균의 차이 $\delta = \mu_1 - \mu_2$ 에 대한 정보가 어느 정도 생기는 것이다. 이렇게 얻는 정보를 이용하여 비교실험 전에 신약의 효과, 즉 **평균의 차이**를 정하는 것이며 이 유효크기가 바로 두 집단의 평균의 차이 $\delta = \mu_1 - \mu_2$ 이다.

이제 평균의 차이 $\delta > 0$ 의 값이 정해지면 검정력을 실제로 계산할 수 있게 된다.

$$power = 1 - \beta = P(\text{reject } H_0 | \mu_1 - \mu_2 = \delta > 0) \quad (2)$$

참고로 유념해야 할 점은 임상실험의 전 과정에서 평균의 차이 뿐만 아니라 모집단의 분산 σ^2 에 대한 정보도 얻을 수 있다. 표본의 수를 결정할 때 분산의 값도 가정해야 함을 유의하자. 실제로 분산의 값을 미리 예상하는 것은 매우 어려운 작업이며 흔히 선행연구의 결과를 참고하여 결정하기도 한다.

3 가설검정의 기각역

이제 일단 각 두 집단의 표본의 수를 n 이라고 하고 가정하며 또한 모집단의 분산 σ^2 를 알고 있다고 하자. 가설검정은 두 집단의 표본 평균 \bar{X} 와 \bar{Y} 를 이용한 z-통계량을 이용한다고 가정하자.

$$z = \frac{\bar{X} - \bar{Y}}{\sigma \sqrt{1/n + 1/n}} \quad (3)$$

양측가설 (1)의 귀무가설을 기각하려면 z-통계량의 값의 절대값이 커야 한다. 즉 어떤 양수 c 에 대하여 $|z| > c$ 이면 귀무가설을 기각한다고 하자. 가설검정의 기각역(rejection region)은 제 1 종의 오류가 유의수준 α 과 같아지도록 정한다.

H_0 가 옳은 경우 z-통계량은 표준정규분포를 따르므로

$$z = \frac{\bar{X} - \bar{Y}}{\sigma\sqrt{1/n + 1/n}} \sim N(0,1) \text{ under } H_0$$

이제 이러한 사실을 이용하여 제 1 종의 오류의 확률이 유의수준 α 와 같다고 하면

$$P(|z| > c | \mu_1 = \mu_2) = \alpha$$

기각역은 다음과 같이 주어진다. 여기서 z_α 는 표준정규분포의 α -상위 백분위 수이다 [$P(Z > z_\alpha) = \alpha$].

$$\text{rejection region} = \{|z| > z_{\alpha/2}\} \quad (4)$$

4 검정력의 계산

이제 약이 효과가 있는 경우, 즉 가설 (1)에서 H_1 이 참인 경우를 고려하자. 평균의 차이 $\delta > 0$ 의 값이 주어지면 다음과 같은 가설을 고려한다.

$$H_0 : \mu_1 - \mu_2 = 0 \quad \text{vs.} \quad H_1 : \mu_1 - \mu_2 = \delta \quad (5)$$

이제 $\mu_1 - \mu_2 = \delta$ 인 경우, 즉 가설 (5) 에서 H_0 를 기각하고 H_1 을 채택할 확률, 즉 검정력을 계산해 보자. 여기서 유의할 점은 가설 (5)에서는 검정력을 계산할 때 기각역 (4)의 한 쪽 부분만 고려해야 한다. 이는 신약의 효과가 보여지려면 z-통계량이 $z_{\alpha/2}$ 보다 커야한다. 만약 z-통계량이 $-z_{\alpha/2}$ 보다 작으면 가설 (5)에서 H_0 는 기각될 수 없다.

$$\begin{aligned}
power &= P(z > z_{\alpha/2} | \mu_1 - \mu_2 = \delta) \\
&= P\left[\frac{\bar{X} - \bar{Y}}{\sigma\sqrt{2/n}} > z_{\alpha/2} | \mu_1 - \mu_2 = \delta\right] \\
&= P\left[\frac{\bar{X} - \bar{Y} - \delta}{\sigma\sqrt{2/n}} > z_{\alpha/2} - \frac{\delta}{\sigma\sqrt{2/n}}\right] \\
&= P\left[Z > z_{\alpha/2} - \frac{\delta}{\sigma\sqrt{2/n}}\right] \\
&= 1 - P\left[Z < z_{\alpha/2} - \frac{\delta}{\sigma\sqrt{2/n}}\right]
\end{aligned}$$

위의 식은 $H_1 : \mu_1 - \mu_2 = \delta$ 이 참인 경우 다음의 z-통계량이 표준정규분포를 따른다는 사실을 이용한 것이다.

$$z = \frac{\bar{X} - \bar{Y} - \delta}{\sigma\sqrt{2/n}} \sim N(0,1) \text{ under } H_1$$

유의수준 $\alpha = 0.05$, 평균의 차이 $\delta = 2.0$, 표준편차가 $\sigma = 4.0$ 인 경우 검정력이 표본의 수 n 에 따라 변한다. 이를 그림으로 그려보자

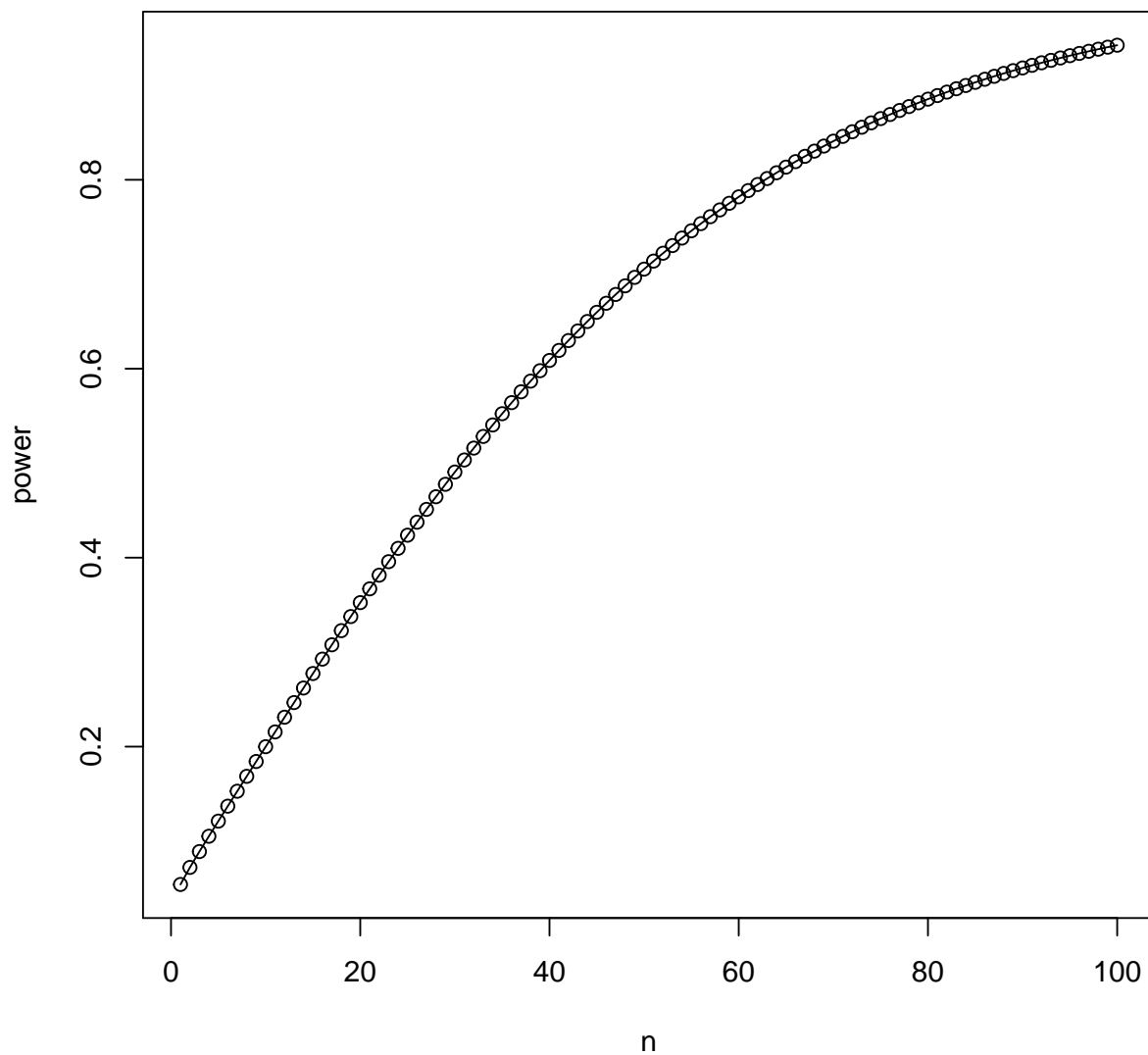
```

alpha <- 0.05
delta <- 2.0
sigma <- 4.0
n <- 1:100
zalpha2 <- -qnorm(alpha/2)
zalpha2

## [1] 1.959964

power <- 1-pnorm(zalpha2 - delta*sqrt(n)/(sigma*sqrt(2)))
plot(n,power)
lines(n,power)

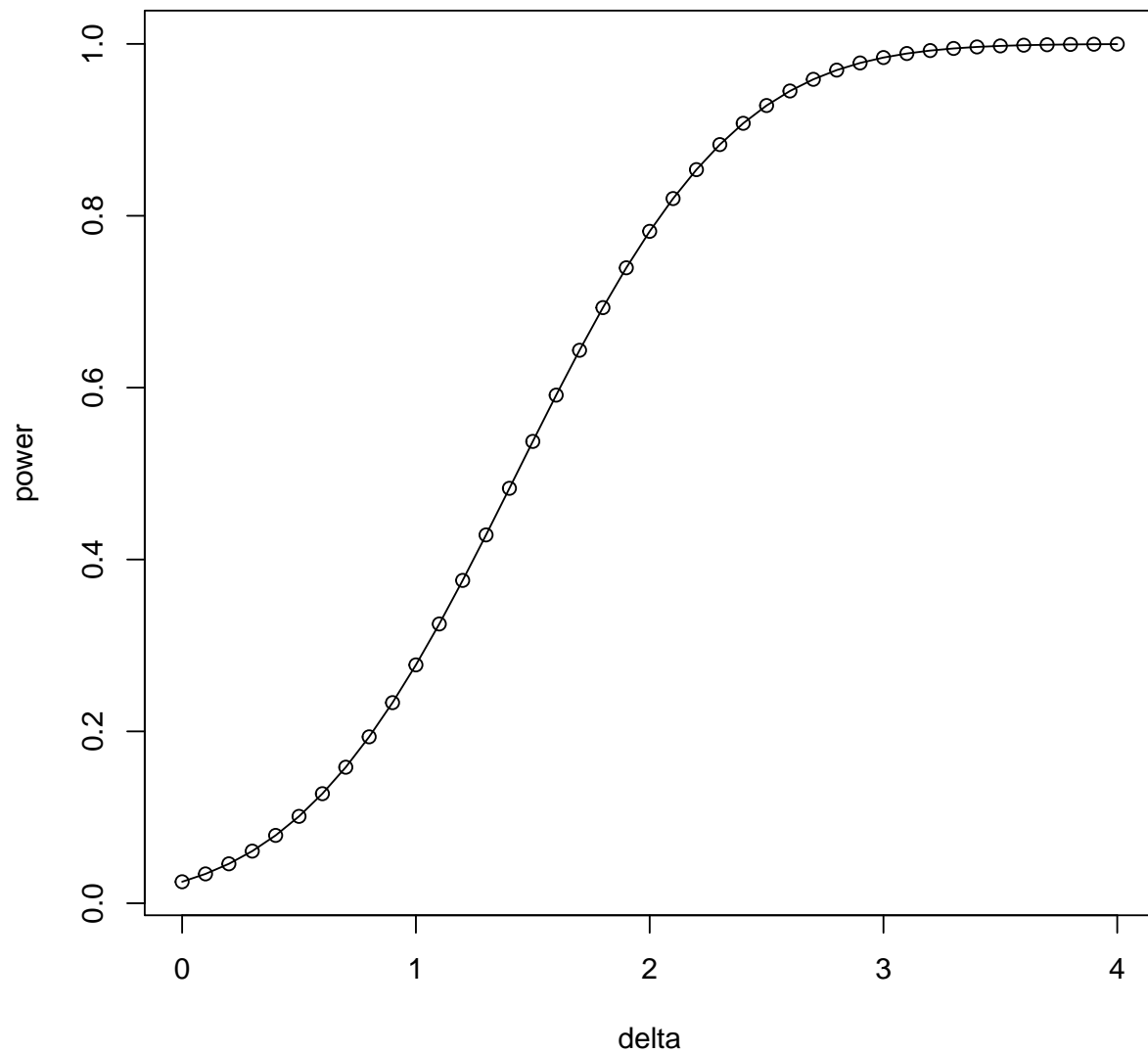
```



유의수준 $\alpha = 0.05$, 표준편차가 $\sigma = 4.0$, 표본의 수 $n = 60$ 인 경우 검정력이 평균의 차이 δ 에 따라 변화하는 그림을 그려보자

```
alpha <- 0.05
delta <- seq(0.0, 4.0, 0.1)
sigma <- 4.0
n <- 60
zalpha2 <- -qnorm(alpha/2)
power <- 1-pnorm(zalpha2 - delta*sqrt(n)/(sigma*sqrt(2)))
```

```
plot(delta,power)
lines(delta,power)
```



5 연구대상자수 n 의 결정

이제 평균의 차이 $\delta > 0$ 의 값이 주어진 경우 검정력, 즉 신약이 효과가 있다고 가설 (5) 에서 대립가설을 채택할 확률은 다음과 같다.

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

일반적으로 임상시험을 시작할 때 목표하는 검정력의 크기를 정한다. 위에서 검정력은 $1 - \beta$ 와 같으므로 제 2 종의 오류를 범할 확률 β 을 정하면 검정력의 크기도 정해지게 된다.

이제 제 2 종의 오류를 범할 확률 β 가 정해 졌다면 다음과 같은 방정식이 성립한다.

$$1 - P \left[Z < z_{\alpha/2} - \frac{\delta}{\sigma\sqrt{2/n}} \right] = 1 - \beta \quad (6)$$

방정식 (6)에서 평균의 차이 δ , 유의수준 α , 제 2 종의 오류를 범할 확률 β 그리고 표준편차 σ 의 값이 주어졌다면 정해지지 않은 값은 표본의 개수 n 이므로 방정식 (6)을 n 에 대하여 풀 수 있다.

$$\begin{aligned} 1 - P \left[Z < z_{\alpha/2} - \frac{\delta}{\sigma\sqrt{2/n}} \right] &= 1 - \beta \\ \Leftrightarrow P \left[Z < z_{\alpha/2} - \frac{\delta}{\sigma\sqrt{2/n}} \right] &= \beta \\ \Leftrightarrow z_{\alpha/2} - \frac{\delta}{\sigma\sqrt{2/n}} &= -z_{\beta} \\ \Leftrightarrow \frac{\delta}{\sigma\sqrt{2}}\sqrt{n} &= z_{\alpha/2} + z_{\beta} \\ \Leftrightarrow n &= \frac{2(z_{\alpha/2} + z_{\beta})^2\sigma^2}{\delta^2} \end{aligned}$$

이제 유효크기 δ , 유의수준 α , 제 2 종의 오류를 범할 확률 β 그리고 표본편차 σ 의 값이 주어졌다면 각 치료 그룹에 필요한 연구대상자의 수는 아래와 같이 주어지며 처리 그룹이 2개이기 때문에 전체적으로 필요한 연구대상자의 수는 $2n$ 명이다. 이때의 검정력은 $1 - \beta$ 임을 알 수 있다.

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

위의 식에서 평균의 차이 $\delta = \mu_1 - \mu_2$ 를 표준편차 σ 로 나눈 양을 **유효크기(effective size)**라고 부른다. 유효크기는 단위에 상관없는 양으로서 표준화된 평균의 차이로 볼 수 있다. 위의 공식에서 표본의 크기는 유효크기의 제곱에 반비례함을 알 수 있다.

$$\text{effective size} = \frac{\delta}{\sigma} = \frac{\mu_1 - \mu_2}{\sigma}$$

유의수준 $\alpha = 0.05$, 평균의 차이 $\delta = 2.0$, 표준편차가 $\sigma = 4.0$, 검정력이 80%인 경우($\beta = 0.2$) 필요한 표본의 수 n 를 구해보자.

```
alpha <- 0.05
delta <- 2.0
sigma <- 4.0
beta <- 0.2
zalpha2 <- -qnorm(alpha/2)
zbeta <- -qnorm(beta)
zbeta

## [1] 0.8416212

n <- 2*(zalpha2 + zbeta)^2 * sigma^2/ delta^2
n

## [1] 62.79104
```

계산된 n 이 정수가 아닌 경우 올림하여 구한다 ($n = 63$)