Hypothesis Ⅱ

-Comparison of Proportions-

Contents

1. 모비율의 추정 – 일표본 & 이표본

모비율의 추정 – 일표본 & 이표본

모비율의 추정

- 질병의 발생률이나 유권자의 지지율 등 모집단의 비율에 대한 조사가 필요한 연구에 사용
- 모비율을 성공확률 p라 간주하면, n개의 표본에서 나온 성공 횟수 X의 분포는 이항분포

  $B(n, p)$를 따른다. X의 기댓값은 $E(X) = np$이므로, 모비율 p의 불편추정량 $\hat{p} = \frac{X}{n}$을 얻는다.
- X의 분산은 $V(X) = np(1-p)$이므로 중심극한정리를 적용하면

$$\frac{X - np}{\sqrt{np(1-p}} = \frac{\hat{p} - p}{\sqrt{p(1-p)\,/\,n}} \sim N(0,1)$$

모비율의 근사 신뢰구간

- 표본크기 n이 클 때, 모비율 p에 대한 $100(1-\alpha)\%$ 신뢰구간은

$$\left[\hat{p} \pm z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right]$$

※ 표준오차는 $\hat{p}$ 이 아닌 p가 들어가야 하나 모비율의 참값을 모르므로 점추정치인 $\hat{p}$을 사용함

모비율의 추정 – 일표본 & 이표본

모비율의 추정

- 기기 측정시험에서 200명의 대상자를 랜덤하게 선발하여 검사한 결과, 15명의 양성자가 발견되었다. 질환의 양성비율에 대한 95% 신뢰구간을 구하시오.

  - [구간추정] 양성비율의 점추정치는 $\hat{p} = \frac{15}{200} = 0.075$이고 $z_{0.975} \doteq 1.96$ 이므로 질환의 양성비율에 대한 95% 신뢰구간은

$$\left[0.075 \pm 1.96 \times \sqrt{\frac{0.075 \times 0.925}{200}}\right] \doteq [0.075 \pm 0.037] = [0.038, 0.112]$$

모비율의 추정 – 일표본 & 이표본

모비율의 추정 – 1) 일표본

모비율의 검정(*표본이 크지 않은 경우*)
● 귀무가설 $H_0: p = p_0$, 성공 횟수의 관측치: x 1) $H_1: p > p_0$ → 기각역: $P(X \geq x \mid H_0) \leq \alpha$ 2) $H_1: p < p_0$ → 기각역: $P(X \leq x \mid H_0) \leq \alpha$ 3) $H_1: p \neq p_0$ → 기각역: $\min[P(X \geq x \mid H_0), P(X \leq x \mid H_0)] \leq \alpha/2$

모비율의 추정 – 일표본 & 이표본

모비율의 추정 – 1) 일표본

모비율의 검정(*표본이 큰 경우*)

- 귀무가설 $H_0: p = p_0$, $Z_0 \equiv \frac{\bar{X} - np_0}{\sqrt{np_0(1-p_0)}} = \frac{\bar{X}/n - p_0}{\sqrt{p_0(1-p_0)/n}}$

1) $H_1: p > p_0$ → 기각역: $Z_0 > z_{1-\alpha}$

2) $H_1: p < p_0$ → 기각역: $Z_0 < z_\alpha = -z_{1-\alpha}$

3) $H_1: p \neq p_0$ → 기각역: $|Z_0| > z_{1-\alpha/2}$

모비율의 추정 – 일표본 & 이표본

모비율의 추정 – 1) 일표본

- 기기 측정시험에서 200명의 대상자를 랜덤하게 선발하여 검사한 결과, 15명의 양성자가 발견되었다. 질환의 양성비율이 0.1보다 작다고 할 수 있는지 유의수준 5%에서 검정하시오.

  - [검정] 귀무가설 $H_0: p = 0.1$, 대립가설 $H_1: p < 0.1$이며, 유의수준 $\alpha = 0.05$ 이므로 기각역은 $Z_0 < z_{0.05} \doteq -1.645$이다. 따라서 검정통계량은

$$Z_0 \doteq \frac{15 - 20}{\sqrt{200 \times 0.1 \times 0.9}} \doteq -1.179 > -1.645$$

  이므로 유의수준 5%에서 귀무가설을 기각할 수 없다. 즉, 유의수준 5% 하에서 질환의 양성비율이 0.1보다 작다는 증거가 없다.

  - *R 실습!*

모비율의 추정 – 일표본 & 이표본

모비율의 추정 – 2) 이표본

- 두 모집단의 모비율이 각각 p_1, p_2라 하고 각 모집단에서 충분히 큰 수의 표본 n_1, n_2개를 추출하여 각각 X개, Y개의 성공 횟수를 얻었다고 하자.
- 성공횟수 X와 Y는 각각 $X \sim B(n_1, p_1), Y \sim B(n_2, p_2)$와 같이 이항분포를 따르므로 $E(X) = n_1 p_1$, $E(Y) = n_2 p_2$이며, 따라서 $\hat{p}_1 = \frac{X}{n_1}$, $\hat{p}_2 = \frac{Y}{n_2}$은 각각 p_1, p_2에 대한 불편추정량이 된다.
- 모비율 차이에 대한 추정량 $\hat{p}_1 - \hat{p}_2$의 기댓값은 $\mathrm{E}(\hat{p}_1 - \hat{p}_2) = p_1 - p_2$, 분산은 $\mathrm{Var}(\hat{p}_1 - \hat{p}_2) = p_1(1 - p_1)/n_1 + p_2(1 - p_2) / n_2$이다.

모비율 차이의 근사 신뢰구간

- 표본크기 n_1, n_2가 클 때, 두 모집단의 모비율 차이 $p_1 - p_2$에 대한 $100(1-\alpha)\%$ 신뢰구간은

$$\left[(\hat{p}_1 - \hat{p}_2) \pm z_{1-\alpha/2}\sqrt{\frac{\hat{p}_1\,(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2\,(1-\hat{p}_2)}{n_2}}\right]$$

모비율의 추정 – 일표본 & 이표본

모비율의 추정 – 2) 이표본

모비율 차이의 검정(*표본이 큰 경우*)

- 귀무가설 H_0: $p_1 = p_2$, 검정통계량: $Z_0 \equiv \dfrac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_1}+\frac{1}{n_2})}}$

- 여기서 공통비율 p의 합동추정치는 $\hat{p} = \frac{X_1+X_2}{n_1+n_2}$를 사용한다.

1) H_1: $p_1 - p_2 > 0 \rightarrow$ 기각역: $Z_0 > z_{1-\alpha}$

2) H_1: $p_1 - p_2 < 0 \rightarrow$ 기각역: $Z_0 < z_\alpha = -z_{1-\alpha}$

3) H_1: $p_1 - p_2 \neq 0 \rightarrow$ 기각역: $|Z_0| > z_{1-\alpha/2}$

모비율의 추정 – 일표본 & 이표본

모비율의 추정 – 2) 이표본

- A기기와 B기기에서 측정한 질환의 양성비율을 비교하기 위하여 각각 160명과 200명의 표본을 추출해 검사한 결과 A기기에서는 12명, B기기에서는 13명의 양성자가 나왔다. 두 기기의 양성비율에 대한 95% 신뢰구간을 구하시오.

  - [구간추정] $\hat{p}_1 = \frac{12}{160} = 0.075$, $\hat{p}_2 = \frac{13}{200} = 0.065$이고 $z_{0.975} \doteq 1.96$ 이므로 95% 신뢰구간은

$$\left[(0.075 - 0.065) \pm 1.96 \times \sqrt{\frac{0.075 \times 0.925}{160} + \frac{0.065 \times 0.935}{200}}\right] \doteq [0.01 \pm 0.053] = [-0.043, 0.063]$$

모비율의 추정 – 일표본 & 이표본

모비율의 추정 – 2) 이표본

- A기기와 B기기에서 측정한 질환의 양성비율을 비교하기 위하여 각각 150명과 250명의 표본을 추출해 검사한 결과 A기기에서는 12명, B기기에서는 10명의 양성자가 나왔다. A기기의 양성비율이 B기기의 양성비율보다 크다고 할 수 있는지 유의수준 5%에서 검정하시오.

  - [검정] 귀무가설 $H_0: p_1 = p_2$, 대립가설 $H_1: p_1 > p_2$이며, 유의수준 $\alpha = 0.05$ 이므로 기각역은 $Z_0 > z_{0.95} \doteq 1.645$이다.
  - $\hat{p}_1 = \frac{12}{150} = 0.08,\ \ \hat{p}_2 = \frac{10}{250} = 0.04,\ \ \hat{p} = \frac{12+10}{150+250} = 0.055$ 이다.

  - 따라서 검정통계량은

$$Z_0 \doteq \frac{0.08 - 0.04}{\sqrt{0.055 \times 0.945 \times (\frac{1}{150} + \frac{1}{250})}} \doteq 1.699 > 1.645$$

이므로 유의수준 5%에서 귀무가설을 기각할 수 있다. 즉, 유의수준 5% 하에서 A기기의 양성비율이 B기기보다 양성비율이 크다고 할 만한 증거가 있다.

비율비교 vs. 평균비교

자료 상에서의 비교

ID	Group	Conti_var	Bin_var
1	1	12.2	0
2	1	13.5	0
3	1	12.0	1
4	2	13.3	1
5	2	14.2	0
6	2	13.9	1
⋮	⋮	⋮	⋮

2. 신뢰구간의 이해

신뢰구간의 이해

신뢰구간의 이해

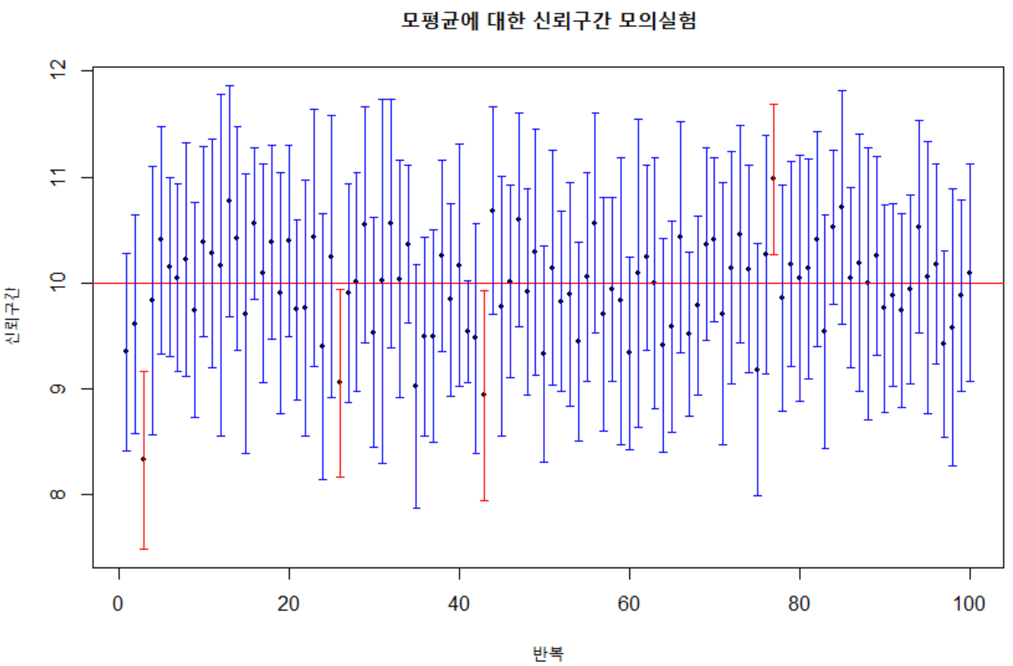
[신뢰구간에 대한 오해]

- 모수 θ가 신뢰구간 $[a, b]$에 포함될 확률은 $100(1-\alpha)\%$ 이다.

[신뢰구간에 대한 올바른 해석]

- 확률적인 신뢰구간 $[L, U]$가 모수 θ의 참값을 포함할 확률은 $100(1-\alpha)\%$인데, 그 신뢰구간 중 하나의 관측치가 $[a, b]$이다.

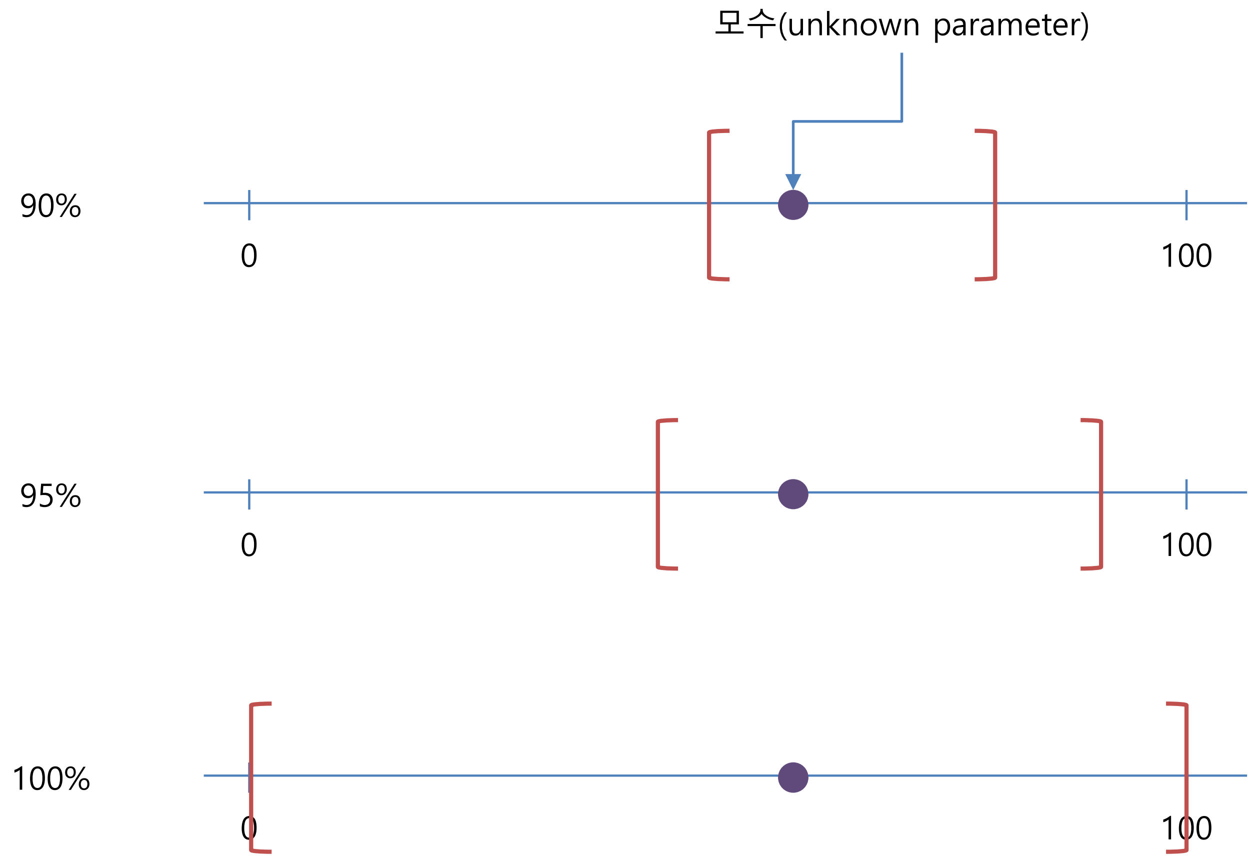
- *R 실습!*



신뢰구간의 이해

신뢰구간의 이해

[몇 % 신뢰구간을 사용해야 하는가?]



3. 검정력 함수

검정력 함수

검정력 함수(power of function)

- 가설검정에서 기각역을 설정할 때, 제1종 오류만을 기준으로 삼았지만, 모수 값이 귀무가설에서 벗어났을 때 가설검정 방법이 얼마나 민감하게 이를 발견하는지 알아볼 필요가 있다.
- 따라서 모수 값의 변화에 따라 귀무가설을 기각하는 확률의 함수로서 검정력 함수를 다음과 같이 정의한다.
- 모수 값의 변화에 따라 귀무가설 H_0을 기각할 확률의 함수로서,

$$\psi(\theta) = P\left(H_0\text{기각} \mid \theta\right)$$

- 재판과정에 비유하면, 범죄의 경중에 따라 의미 있는 증거를 확보하여 유죄 판결을 이끌어내는 검찰의 능력이라 할 수 있다.
- *R 실습!*

검정력 함수

검정력 함수(power of function)

- 모분산을 아는 경우 모평균에 대한 검정에서 $H_0: \mu = \mu_0$, $H_1: \mu > \mu_0$ 라고 한다면,

  기각역은 $Z_0 > z_{1-\alpha}$이므로 검정력 함수는 다음과 같다.

$$\psi(\mu) = P(Z_0 > z_{1-\alpha} |\, \mu) = P\left(\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} > z_{1-\alpha} \middle| \mu\right)$$

$$= P\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} > z_{1-\alpha} + \frac{\mu_0 - \mu}{\sigma/\sqrt{n}} \middle| \mu\right) = 1 - \Phi\left(z_{1-\alpha} + \frac{\mu_0 - \mu}{\sigma/\sqrt{n}}\right)$$

- 모분산을 아는 경우 모평균에 대한 검정에서 $H_0: \mu = \mu_0$, $H_1: \mu < \mu_0$ 라고 한다면, 기각역은

  $Z_0 < -z_{1-\alpha}$이므로 검정력 함수는 다음과 같다.

$$\psi(\mu) = P(Z_0 < -z_{1-\alpha} |\, \mu) = P\left(\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} < -z_{1-\alpha} \middle| \mu\right)$$

$$= P\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < -z_{1-\alpha} + \frac{\mu_0 - \mu}{\sigma/\sqrt{n}} \middle| \mu\right) = \Phi\left(-z_{1-\alpha} + \frac{\mu_0 - \mu}{\sigma/\sqrt{n}}\right)$$

검정력 함수

검정력 함수(power of function)

- 모분산을 아는 경우 모평균에 대한 검정에서 $H_0: \mu = \mu_0$, $H_1: \mu \neq \mu_0$ 라고 한다면,

  기각역은 $|Z_0| > z_{1-\alpha/2}$이므로 검정력 함수는 다음과 같다.

$$\psi(\mu) = P\left(|Z_0| > z_{1-\alpha/2} \middle| \mu\right) = P\left(\left|\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}\right| > z_{1-\alpha/2} \middle| \mu\right)$$

$$= \Phi\left(-z_{1-\alpha/2} + \frac{\mu_0 - \mu}{\sigma/\sqrt{n}}\right) + 1 - \Phi\left(z_{1-\alpha/2} + \frac{\mu_0 - \mu}{\sigma/\sqrt{n}}\right)$$

- 위의 검정력 함수를 활용하면 다음과 같은 표본 수 산출 수 있다.

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

  여기서 $\varepsilon = \mu_0 - \mu$ 이다.

검정력 함수

검정력 함수(power of function)

검정력 함수

- 귀무가설 $H_0: \mu = \mu_0$, 검정통계량: $Z_0 \equiv \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$

1) H_1: $\mu > \mu_0$ -> $\psi(\mu) = 1 - \Phi(z_{1-\alpha} + \frac{\mu_0-\mu}{\frac{\sigma}{\sqrt{n}}})$

2) H_1: $\mu < \mu_0$ -> $\psi(\mu) = \Phi(-z_{1-\alpha} + \frac{\mu_0-\mu}{\frac{\sigma}{\sqrt{n}}})$

3) H_1: $\mu \neq \mu_0$ -> $\psi(\mu) = \Phi\left(-z_{1-\alpha/2} + \frac{\mu_0-\mu}{\frac{\sigma}{\sqrt{n}}}\right) + 1 - \Phi(z_{1-\alpha/2} + \frac{\mu_0-\mu}{\frac{\sigma}{\sqrt{n}}})$

Thank you!