통계적 추론

-표본의 크기가 클 때-

모비율 p에 대한 가설검정

모비율 p에 대한 가설검정

모비율 p에 대한 가설검정

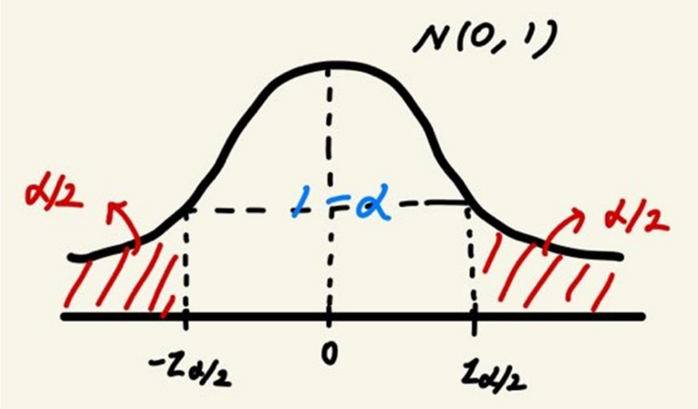
모비율의 추정(표본의 크기가 클 때) - 점추정

- 모수: 모집단의 비율 p (ex. 실업률)
- 자료: 모집단에서 임의로 n명을 추출하고, 그 중 해당하는 사람(실직자의 수)를 X라 하자.
- (모비율 p의) 점추정량: 표본비율 $\hat{p} = \frac{X}{n}$

$$E(\hat{p}) = p,\ Var(\hat{p}) = p(1-p)/n$$

- 표준오차: $S.E.(\hat{p}) = \sqrt{p(1-p)/n}$
- 추정된 표준오차: $\sqrt{\hat{p}(1-\hat{p})/n}$

모비율의 추정(표본의 크기가 클 때) - 구간추정



- $X \sim Bin(n, p)$
- $X = \sum_{i=1}^{n} X_i$ where $X_i \sim i.i.d.\ Bernoulli(p)$
- $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ by Central Limit Theorem
- $\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{X}{n} = \hat{p} \sim N(p, \frac{p(1-p)}{n})$
- $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} = \frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} \sim N(0,1)$
- $P\left(\left|\frac{\hat{p}-p}{\sqrt{p(1-p)/n}}\right| < z_{\alpha/2}\right) = P\left(|\hat{p}-p| < z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}\right) = 1-\alpha$ (근사치)

모비율의 추정(표본의 크기가 클 때) - 구간추정

- $P\left(\hat{p} - z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} < p < \hat{p} + z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}\right) = 1 - \alpha$: 근사치

- $100(1-\alpha)\%$ 신뢰구간

  $\left(\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$ $where\ 100(1-\alpha)\%$: 신뢰도(신뢰수준)

- 오차범위(error margin): $z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

모비율 p에 대한 가설검정

- 표본의 크기가 클 때 모비율 p에 대한 가설 $H_0: p = p_0$를 검정하기 위한 검정통계량은 다음과 같다.

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1 - p_0)}{n}}}$$

검정통계량의 분포는 H_0가 맞을 때 $N(0,1)$을 따른다. 각 대립가설에 대하여 유의수준 α를 갖는 기각역은 다음과 같다.

$H_1: p < p_0$ 일 때 $R: Z \le -z_\alpha$ (단측검정)

$H_1: p > p_0$ 일 때 $R: Z \ge z_\alpha$ (단측검정)

$H_1: p \ne p_0$ 일 때 $R: |Z| \ge z_{\alpha/2}$ (양측검정)

모비율 p에 대한 가설검정

- P-값(혹은 유의확률): 주어진 검정통계량의 관측치로부터 H_0를 기각하게 하는 최소의 유의수준을 말한다.

- $H_1: p < p_0$ 일 때 $R: Z \le -z_\alpha$ (단측검정) -> P-값$=P(Z \le z)$
  $H_1: p > p_0$ 일 때 $R: Z \ge z_\alpha$ (단측검정) -> P-값$=P(Z \ge z)$
  $H_1: p \ne p_0$ 일 때 $R: |Z| \ge z_{\alpha/2}$ (양측검정) -> P-값$=P(|Z| \le |z|)$

수고하셨습니다.