Confidence intervals

Population $\longrightarrow$ Statistic estimate

$\hookrightarrow$ SE 발생!

역으로 SE를 이용하여 추정

Statistic estimate $\longrightarrow$ Population : 표본값으로부터 신뢰구간을 잡았을 때, 해당 구간에 실제 Parameter 값이 포함되어있을 확률 | ~%

$\hookrightarrow$ 추정

$\hat{p}$: 샘플값 (보통 평균, 비율), z, SE

$\boxed{\text{Confidence intervals : } [\hat{p} - z \cdot SE,\ \hat{p} + z \cdot SE]}$

$z = 1 \rightarrow 68\%$
$z = 1.645 \rightarrow 90\%$
$z = 1.96 \rightarrow 95\%$
$z = 2.576 \rightarrow 99\%$

z-score $\propto$ Confidence level

Margin of error : $|z \cdot SE|$

$\longrightarrow$ 샘플을 이용하여 모집단 매개변수에 대해 그럴듯한 값의 범위를 제공한다.

Confident intervals via the Central Limit Theorem

- CLT : Sample size가 클 때, 모집단이 Normal curve에 근사하는지 여부와 관계없이 Sample은 Normal Curve에 근사할 수 있다.

$\boxed{\text{Confidence intervals : } [\hat{p} - z \cdot SE,\ \hat{p} + z \cdot SE]}$

해당 식은 정규분포 기반이다 : CLT로 인하여 Population이 정규분포를 따르지 않아도 Sample로 신뢰구간을 구할 수 있다.

Bootstrap principle

- Population을 모르는 상태에서 Confident intervals를 구하기 위해 $SE = \frac{\sigma}{\sqrt{n}}$는 알아야 할 때, Sample의 s를 통해 σ를 추정하는 방법

Ex) 95% 신뢰구간 추정 (Population 정보 X)

$SE = \frac{\sigma}{\sqrt{n}}$, p : 모집단의 비율

$\sigma = \sqrt{p(1-p)}$ 이다. $\sigma = \sqrt{p(1-p)} \leq \frac{1}{2}$ (if $p = \frac{1}{2}$)

95%의 신뢰구간의 경우, $z \approx 2$이므로 신뢰구간 $\approx \hat{p} \pm 2 \cdot \frac{1}{2\sqrt{n}} = \hat{p} \pm \frac{1}{\sqrt{n}}$

$\hookrightarrow$ 보수적으로 사용하는 95% 신뢰구간 추정
다른 신뢰구간에서는 다르다.