Chapter03

통계분석 I

구간추정

FASTCAMPUS
ONLINE
금융공학/퀀트 I

강사. 장순용

I 키포인트

- 구간추정의 원리.
- 신뢰구간.

I 모평균의 구간추정

- 통계량을 바탕으로 신뢰구간 (confidence interval)을 계산한다.
- 신뢰구간: 표본평균의 확률분포에 모평균이 신뢰수준 확률로 포함되는 구간.
- 중심극한정리에 의하면 표본평균 $\bar{X}$는 근사적으로 정규분포를 따르고 표준화된 Z는 표준정규분포를 따른다: $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$.
- 그러면 다음을 정의한다:

  → 신뢰수준 확률: $(1-\alpha)$.

  → 오차율: α.

I 모평균의 구간추정

- 95% 신뢰구간을 만들어 보자. 표준정규분포의 대칭성에 의하면:

$$P(-1.96 < Z < 1.96) = 0.95$$

$$\Downarrow$$

$$P\left(-1.96 < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < 1.96\right) = 0.95$$

$$\Downarrow$$

$$P\left(-1.96\,{}^{\sigma}/_{\sqrt{n}} < \bar{X} - \mu < 1.96\,{}^{\sigma}/_{\sqrt{n}}\right) = 0.95$$

I 모평균의 구간추정

$$P\left(-1.96\,{}^{\sigma}\!/_{\sqrt{n}} < \bar{X} - \mu < 1.96\,{}^{\sigma}\!/_{\sqrt{n}}\right) = 0.95$$

$$\Downarrow$$

$$P\left(-\bar{X} - 1.96\,{}^{\sigma}\!/_{\sqrt{n}} < -\mu < -\bar{X} + 1.96\,{}^{\sigma}\!/_{\sqrt{n}}\right) = 0.95$$

$$\Downarrow$$

$$P\left(\bar{X} - 1.96\,{}^{\sigma}\!/_{\sqrt{n}} \le \mu \le \bar{X} + 1.96\,{}^{\sigma}\!/_{\sqrt{n}}\right) = 0.95$$

I 모평균의 구간추정: 모표준편차를 아는 경우

- 그러면 95% 신뢰구간은 다음 상한과 하한으로 구성되어 있다.

하한: $\bar{X} - 1.96\,{}^{\sigma}/_{\sqrt{n}}$ 상한: $\bar{X} + 1.96\,{}^{\sigma}/_{\sqrt{n}}$

$[$ ← <신뢰구간> → $]$

- 1.96이라는 수치는 어디에서 나온 것인가?

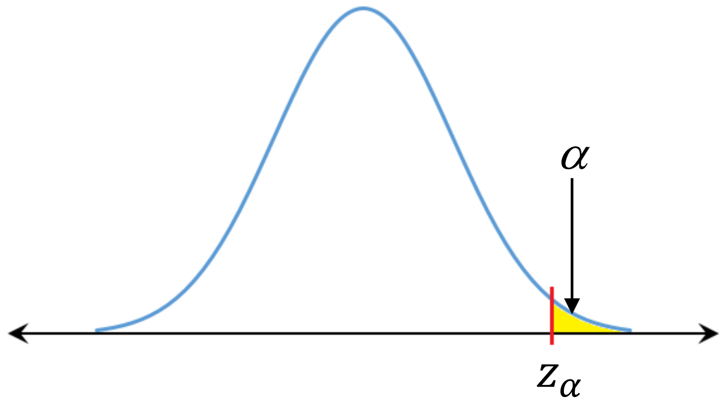
→ $z_{0.025}$에 해당하는 수치이다.

→ $z_{0.025}$는 표준정규확률분포에서 누적확률(CDF)가 0.975에 해당하는 위치이다.

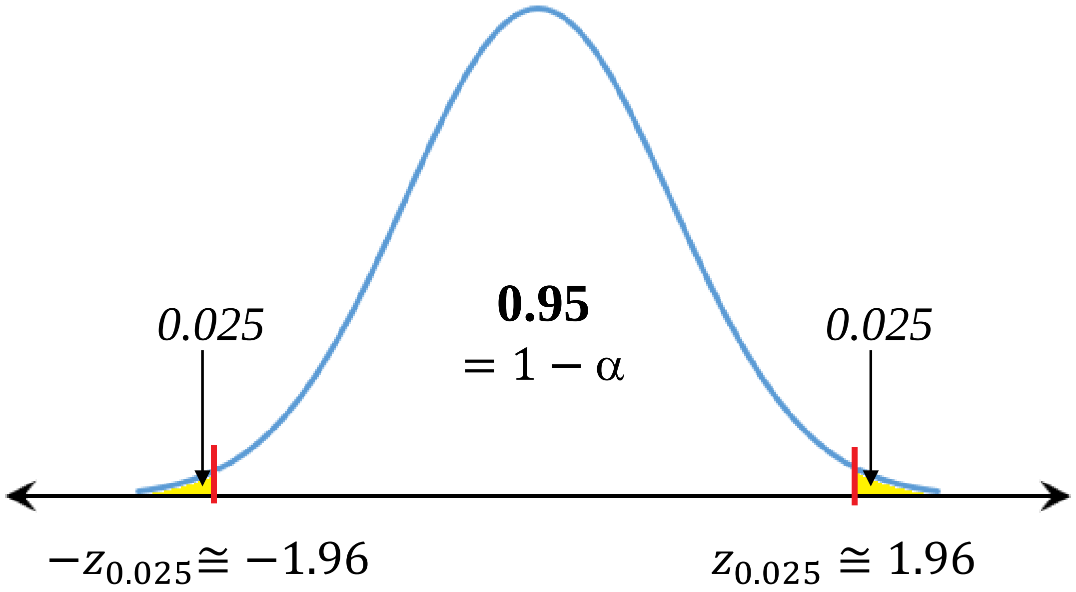
I 표준정규분포의 분위수 (Quantile of Standard Normal)

- z_α는 일종의 분위수.
- z_α는 우측 꼬리의 면적이 α와 같은 위치를 의미함.

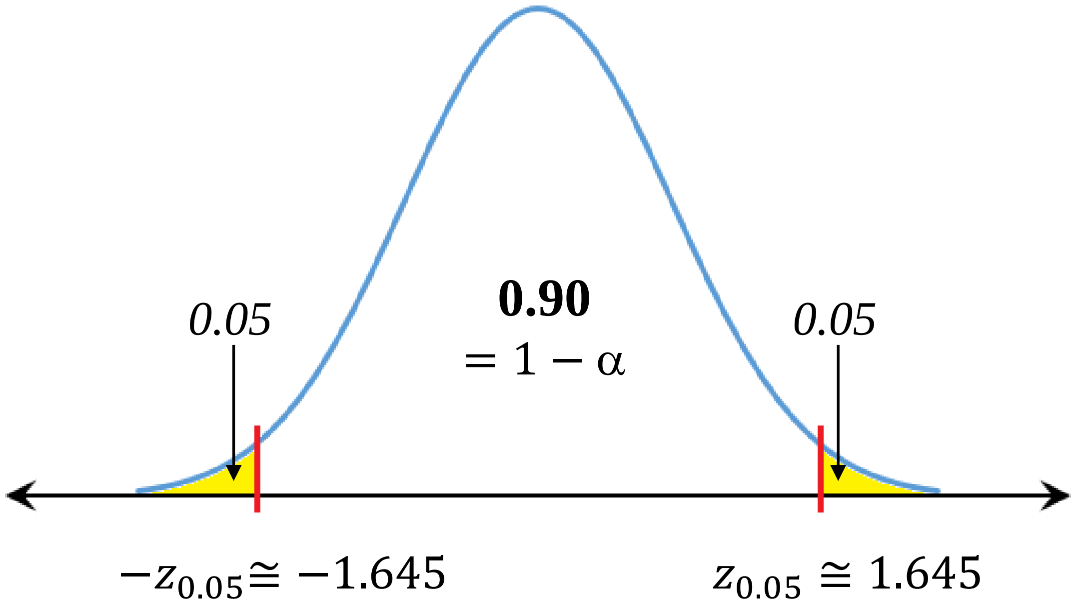
$$P(z_\alpha < Z) = \alpha$$



I 표준정규분포의 분위수 (Quantile of Standard Normal)



I 표준정규분포의 분위수 (Quantile of Standard Normal)



I 모평균의 구간추정: 모표준편차를 아는 경우 일반화

- 다음과 같이 임의의 신뢰수준 확률 $1-\alpha$에 해당하는 신뢰구간을 만들 수 있다.

하한: $\bar{X} - z_{\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{n}}$ 상

한: $\bar{X} + z_{\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{n}}$

[← <신뢰구간>

→]

I 모평균의 구간추정: 원리

Question: 신뢰수준 확률이 높아야 좋은 것 아닌가?

I 모평균의 구간추정: 원리

(　　　　←————————————→　　　　)

신뢰수준 99% 신뢰구간

(　　　←——————————→　　　)

신뢰수준 95% 신뢰구간

(　　←————————→　　)

신뢰수준 90% 신뢰구간

I 모평균의 구간추정: 원리

100% 신뢰구간 : 모국의 성인 남성의 신장은 0m ~ 3m 사이이다.

95% 신뢰구간 : 모국의 성인 남성의 신장은 1.60m ~ 1.90m 사이이다.

90% 신뢰구간 : 모국의 성인 남성의 신장은 1.70m ~ 1.80m 사이이다.

I 모평균의 구간추정: 원리

- 신뢰구간의 상한과 하한은 다음과 같이 계산하였다: $\bar{X} \pm z_{\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{n}}$.
- 오차율 α가 클수록 신뢰구간은 좁다 (컨트롤 가능하지만 그대로 둠).
- 표준편차 σ가 작을수록 신뢰구간은 좁다 (컨트롤 불가능).
- 표본크기 n이 클수록 신뢰구간은 좁다 (컨트롤 가능).

→ 표본크기를 키우면 오차율을 키우지 않고 (신뢰수준 유지) 신뢰구간을 좁힐 수 있다!

I 모평균의 구간추정: 원리

- W가 목표하는 신뢰구간의 폭이라고 한다면 $\bar{X} \pm W$, 다음 관계에 의해서 표본크기를 정한다.

$$\bar{X} \pm z_{\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{n}} = \bar{X} \pm W \quad \Rightarrow \quad z_{\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{n}} = W \quad \Rightarrow \quad \boldsymbol{n} = \left[\frac{\boldsymbol{z_{\alpha/2} \times \sigma}}{\boldsymbol{W}}\right]^{\boldsymbol{2}}$$

I 모평균의 구간추정: 모표준편차를 모르는 경우

- 다음과 같이 임의의 신뢰수준 확률 $1-\alpha$에 해당하는 신뢰구간을 만들 수 있다.

하한: $\bar{X} - t_{\frac{\alpha}{2}} \times \frac{s}{\sqrt{n}}$ 상한: $\bar{X} + t_{\frac{\alpha}{2}} \times \frac{s}{\sqrt{n}}$

[← <신뢰구간> →]

- 모분산을 아는 경우와 비교해서 바뀐 것은:

→ $z_{\alpha/2}$ 대신에 $t_{\alpha/2}$를 사용한다.

(자유도 $n-1$인 스튜던트 t 분포의 분위수)

→ σ 대신에 s를 사용한다.

I 끝.

감사합니다.

FAST CAMPUS
ONLINE
장순용 강사.