컴퓨터응용통계

6. 추정과 검정

최경미

표본분포

표본 $X_1, X_2, \cdots, X_n \sim iid\ N(\mu, \sigma^2)$

σ^2이 알려져 있다고 가정함.

$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$

$Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$

신뢰구간

$P(|Z| < z_{\alpha/2}) = 1 - \alpha$

$P\left(-z_{\alpha/2} < \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} < z_{\alpha/2}\right) = 1 - \alpha$

$P\left(-z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \bar{X} - \mu < z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$

$P\left(\mu \in \left(\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}},\ \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)\right) = 1 - \alpha$

μ의 $100(1-\alpha)\%$ 신뢰구간(CI)

$$= \left(\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \quad \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$$

가설검정

$H_0 : \mu = \mu_0$ vs. $H_1 : \mu \neq \mu_0$

기각역 R을 이용한 검정

검정통계량 $Z = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$

기각역 $R : |Z| \geq z_{\alpha/2}$

Z가 기각역 R에 속하면, H_0 기각.
Z가 기각역 R에 속하지 않으면, H_0 기각안함

의사결정	H_0 참	H_0 거짓
H_0 기각함 **H_1채택**	제1종의 오류	
H_0 기각안함 H_0채택		제2종의 오류

유의수준(significance level)

$\alpha = P($제1종 오류$)$

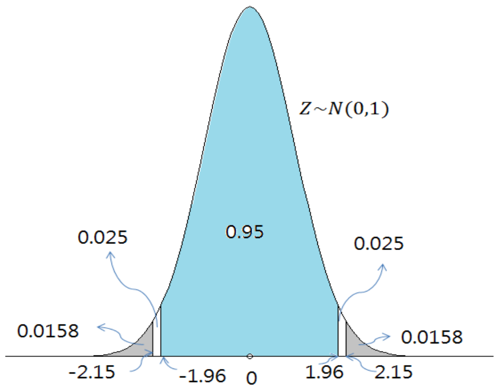
$= P(H_0$ 기각$|\ H_0$ 참$)$

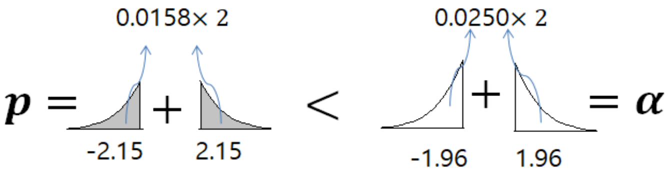
$= H_0$ 기각하는 결정이 틀릴 수

있도록 사회적으로 허용해주는 확률

양측검정에 대한 p − 값 $= P(|Z| \geq |z_0|)$

p − 값 $\leq \alpha$이면 H_0 을 기각한다.





의사결정

$\mu_o \notin CI, H_0$ 기각

$Z \in R$, H_0 기각

p − 값 $\leq \alpha$, H_0 기각

예제 로션 용량은 진짜 50ml일까?

25개 평균이 49.05ml, 표준편차가 2.21...!!! 점추정

진짜 50ml?

95% 신뢰구간 (48.18, 49.92)에 50이 포함되나? 구간추정

평균 용량이 50ml이라고 말할 수 있을까? 가설검정

점추정, 구간추정, 가설검정

- 용량이 50ml라고 표기되어 있는 화장품이 생각보다 빨리 떨어지면, 소비자 입장에서는 실제 용량이 50ml가 맞는지에 대한 합리적인 의심이 생길 수 있다.

- 점추정 (point estimator)

동일 제품을 25개의 평균 용량이 49.05ml이 모평균 μ의 점추정이 될 수 있다.

- 구간추정 (interval estimator)

모평균 μ이 포함될 확률이 0.95 (95%)인 구간 $(48.18, 49.92)$을 얻으면,

이를 μ의 구간 추정이라고 부른다.

- 가설검정 (hypothesis test)

50ml가 이 구간에 포함되지 않으므로, 소비자는 0.95의 확신을 가지고

화장품의 용량이 50ml가 아니라고 짐작한다.

이와 같이 화장품의 평균 용량이 50ml이라고 말할 수 있는지 없는지에 대하여

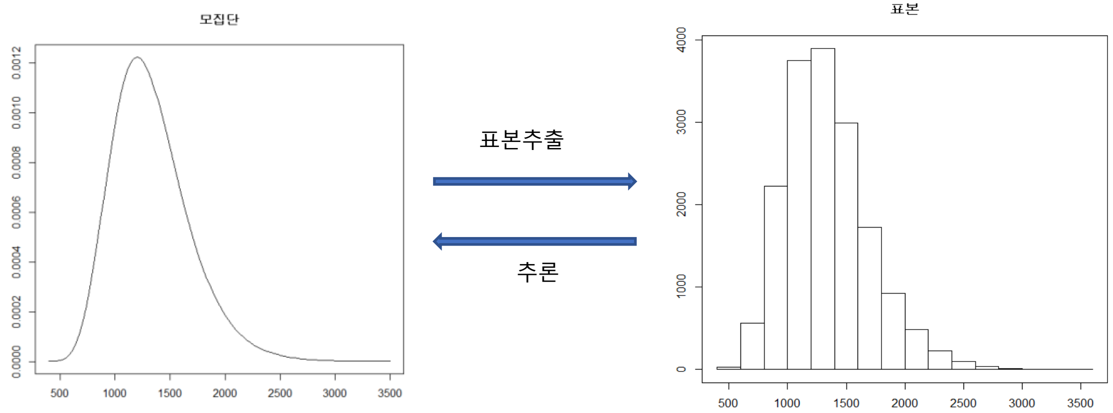
통계적으로 의사결정 내리는 방법을 가설검정이라고 부른다.

추론 = 추정과 검정 등

- 모수(parameter) : 모평균 μ, 모분산 σ^2, 성공 확률 p, 상관계수 ρ 등

- 추정(estimation)과 검정(hypothesis test) :
  모수를 한 값으로 점추정하거나,
  모수가 속할 가능성이 높은 신뢰구간을 추정하거나,
  가설을 세우고 검정할 수 있다.

- 추론 : 추정과 검정의 과정



점추정

표본 $X_1, X_2, \cdots, X_n$을 이용하여 모평균 μ를 한 값으로 추정

평균

$$\hat{\mu} = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$E[\bar{X}] = \mu$$

$$\mathrm{Var}(\bar{X}) = \frac{\sigma^2}{n}$$

$$SE(\bar{X}) = \sqrt{\mathrm{Var}(\bar{X})} = \frac{\sigma}{\sqrt{n}}$$ (표준오차, standard error)

표본 $X_1, X_2, \cdots, X_n$이 독립이고

동일분포 $N(\mu, \sigma^2)$을 따르며,

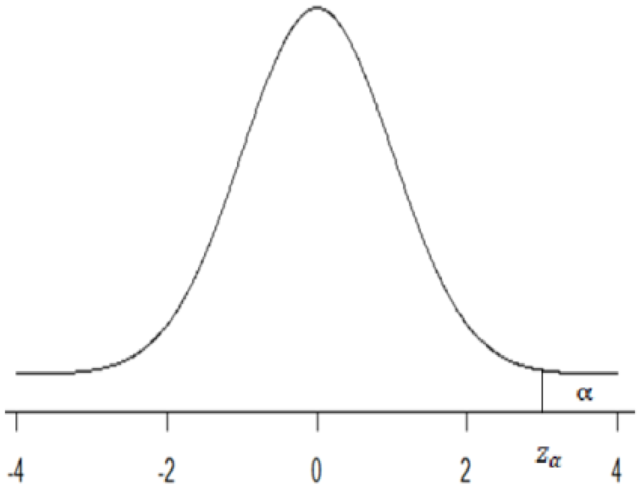
분산 σ^2이 알려져 있다고 가정하면,

다음이 성립한다.

$$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

100(1- α) 백분위수 또는 α 분위수 (quantile) z_α

- **정의 5.4** $Z \sim N(0,1)$ 일 때, $P(Z > z_\alpha) = \alpha$를 만족하는 z_α를
- 100(1- α) 백분위수 또는 α 분위수 (quantile)라고 정의한다.

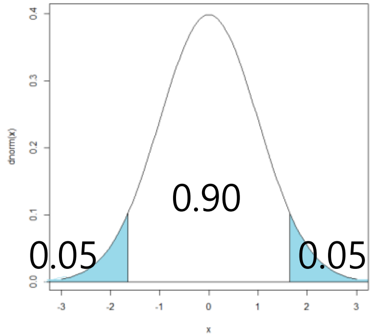


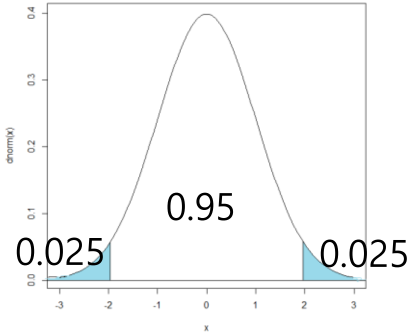
$$P(Z > z_{0.05}) = 0.05,\ z_{0.05} = 1.645$$

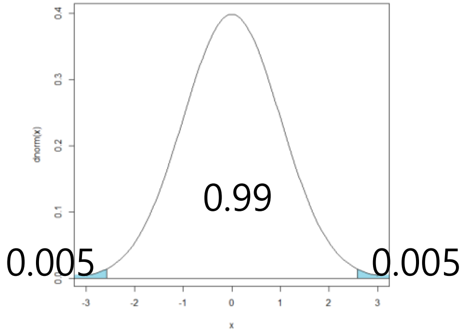
$$P(Z > z_{0.025}) = 0.025,\ z_{0.025} = 1.96$$

$$P(Z > z_{0.005}) = 0.005,\ z_{0.005} = 2.575$$

$$P(-z_{0.025} < Z < z_{0.025}) = 0.95$$







6.2 μ의 100$(1-\alpha)$% 신뢰구간(Confidence Interval)

표본 $X_1, X_2, \cdots, X_n$이 독립이고 동일분포 $N(\mu, \sigma^2)$을 따르며, 분산 σ^2이 알려져 있다고 가정하자. μ를 포함할 확률이 **100$(1-\alpha)$% 인 구간을 추정**하자.

$$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$$

$$Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

$$P\left(|Z| < z_{\alpha/2}\right) = 1-\alpha$$

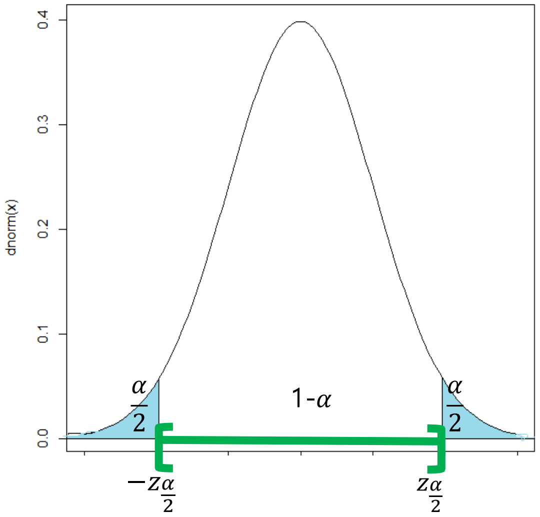
$$P\left(-z_{\alpha/2} < \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} < z_{\alpha/2}\right) = 1-\alpha$$

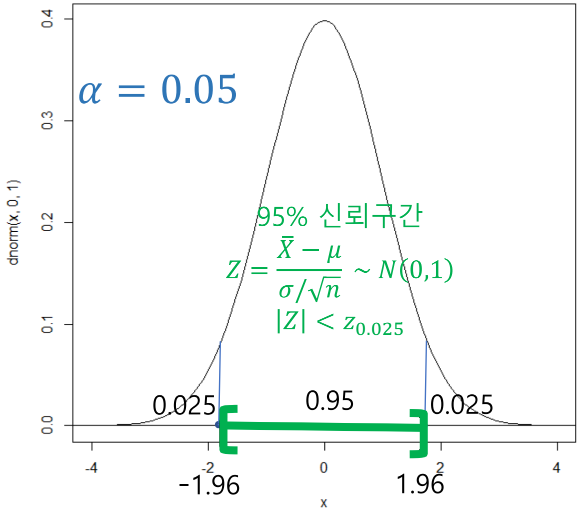
$$P\left(-z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \bar{X}-\mu < z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$$

$$P\left(\mu \in \left(\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \quad \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)\right) = 1-\alpha$$

μ의 100$(1-\alpha)$% 신뢰구간(CI)= $\left(\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}},\ \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$

μ의 95% 신뢰구간(CI)= $\left(\bar{X} - 1.96\frac{\sigma}{\sqrt{n}},\ \bar{X} + 1.96\frac{\sigma}{\sqrt{n}}\right)$





예제 6.2 화장품 평균용량 μ의 95% 신뢰구간

50ml ?

n=25

$\bar{x} = 49.05$

$\sigma = 2.21$ (알려짐)

$z_{0.025}$=1.96

μ의 95% 신뢰구간(CI)

$$= \bar{x} \pm z_{0.025} \frac{\sigma}{\sqrt{n}}$$

추정과 검정

$$= \left(49.05 - 1.96 \frac{2.21}{\sqrt{25}}, 49.05 + 1.96 \frac{2.21}{\sqrt{25}}\right)$$

$$= (48.18, 49.92)$$

μ의 95% 신뢰구간(CI)은 50을 포함하는가?

50 ∉ 95% CI

만약 50이 신뢰구간에 포함되면, 화장품용량이 50ml라고 볼 수 있다.

만약 50이 신뢰구간에 포함되지 않으면, 화장품용량이 50ml라고 볼 수 없다.

혼자 풀기

mtcars에서 32 대 차량의 연비(mpg)가 정규분포를 따르며, 표준편차가 6이라고 가정하고, 연비의 95%신뢰구간을 만들자. 평균연비가 20이라고 말할 수 있을까?

과제

- ppt 2쪽 제일 왼쪽 1번 쓰고 외우기
- 혼자풀기

6.3 가설과 검정: 가설

귀무가설 (null hypothesis) H_0

대립가설 (alternative hypothesis) H_1

세 가지 가설

① $H_0: \mu = \mu_0$ vs. $H_1: \mu \neq \mu_0$ (양측검정)

② $H_0: \mu \leq \mu_0$ vs. $H_1: \mu > \mu_0$ (단측검정)

③ $H_0: \mu \geq \mu_0$ vs. $H_1: \mu < \mu_0$ (단측검정)

화장품 용량 예제

회사 $H_0: \mu = 50$

소비자 $H_1: \mu \neq 50$

가설과 검정: 검정

- **가설** $H_0 : \mu = \mu_0$ vs. $H_1 : \mu \neq \mu_0$

✓ **검정통계량** $Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$

✓ **기각역** $R : |Z| \geq z_{\alpha/2}$ (H_0 에서 신뢰구간의 여집합 모양)

✓ 검정통계량 Z가 기각역 R에 속하면, 귀무가설 H_0을 기각한다.

✓ 검정통계량 Z가 기각역 R에 속하지 않으면, 귀무가설 H_0을 기각하지 않는다

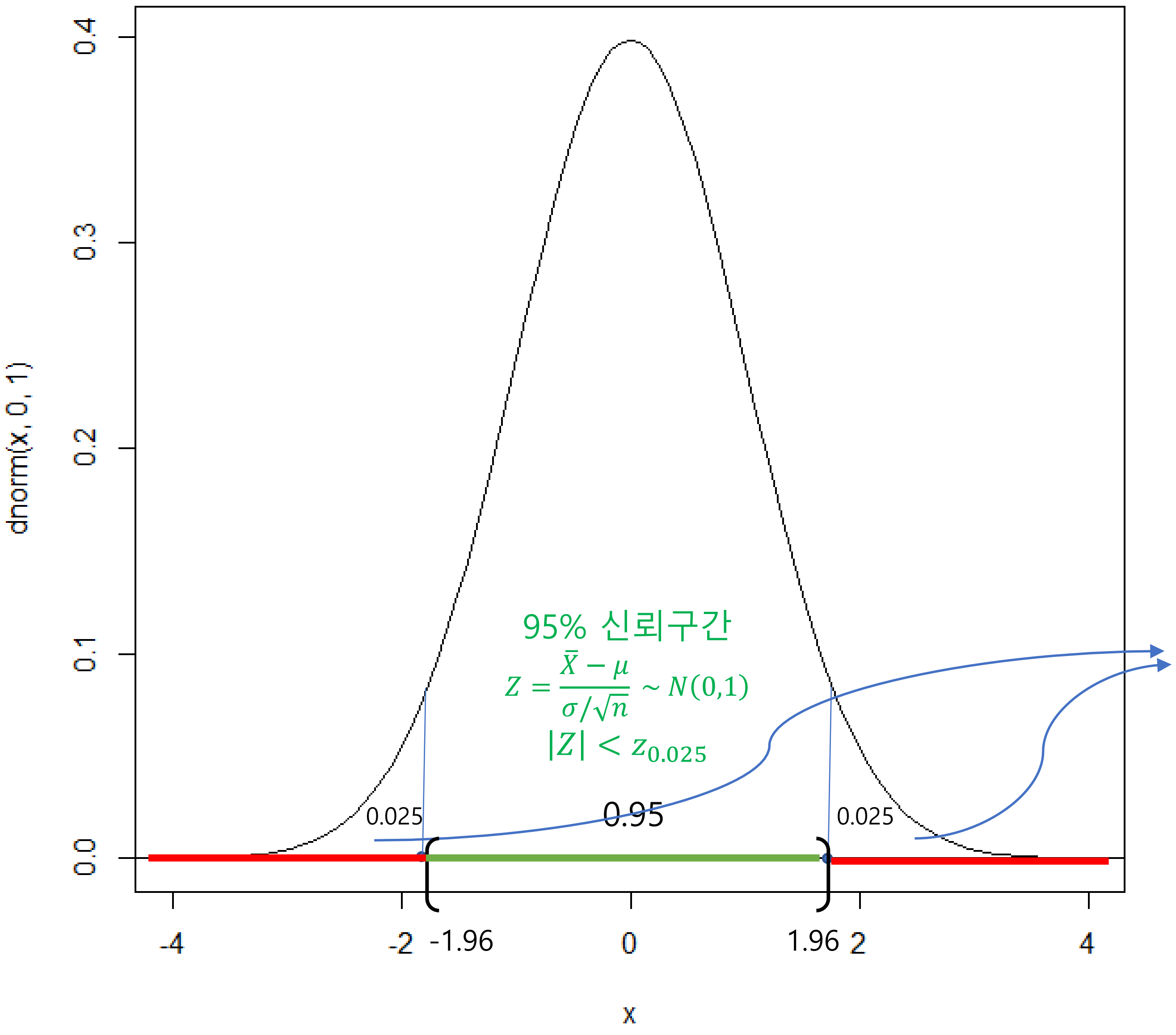
- **신뢰구간** CI을 이용한 검정

✓ $|Z| = \left|\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}\right| < z_{\alpha/2}$을 μ에 대해서 푼다.

✓ $\mathbf{CI} = \left(\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}},\ \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$

✓ 만약 $\mu_0 = 50$이 신뢰구간에 포함되면, 귀무가설 H_0이 타당하다고 볼 수 있다.

✓ 만약 $\mu_0 = 50$이 신뢰구간에 포함되지 않으면, 대립가설 H_1이 타당하다고 볼 수 있다.



가설 $H_0: \mu = \mu_0 \quad H_1: \mu \neq \mu_0$

검정통계량 $Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$

기각역 $|Z| > z_{0.025}$

Z 기각역에 속하면 H_0기각
Z 기각역에 안속하면 H_0채택

예제 화장품 평균용량 μ에 대한 가설검정에서 기각역

μ =50ml?

n=25, $\bar{x} = 49.05$, $\sigma = 2.21$ (알려짐)

유의수준 $\alpha = 0.05$ (양쪽 꼬리 확률, 면적*)*

가설검정

$H_0: \mu = 50$ vs. $H_1: \mu \neq 50$

검정통계량 $Z = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} = \frac{49.05-50}{2.21/\sqrt{25}} = -2.15$

$z_{0.05/2}$ = qnorm(0.975,0,1) = 1.96

기각역

R: $|Z| \geq 1.96$

$R = \{z; z \geq 1.96$ 또는 $z \leq -1.96\}$

$|Z| = |-2.15| = 2.15 > 1.96$이므로, 검정통계량이 기각역에 속한다. $Z \in R$

귀무가설 $H_0: \mu = 50.0$을 기각한다.

검정통계량 Z=-2.15가 기각역 R에 속하므로, 귀무가설 $H_0: \mu = 50.0$을 기각한다.

예제 화장품 평균용량 μ에 대한 가설검정에서 검정통계량

질문 μ =50ml?

자료 n=25, $\overline{x} = 49.05$, $\sigma = 2.21$ (알려짐)

가설검정

$H_0 : \mu = 50$ vs. $H_1 : \mu \neq 50$

검정통계량 Z 계산

신뢰구간 CI 만들기

기각역 R 만들기

의사결정

$Z \in R \Leftrightarrow \mu_0 \notin CI \Leftrightarrow H_0$ 기각

$Z \notin R \Leftrightarrow \mu_0 \in CI \Leftrightarrow H_0$ 기각하지 않음

유의수준 (significance level) α

의사결정	H_0 참	H_0 거짓
H_0을 기각함 $\boldsymbol{H_1}$**채택**	제1종의 오류	
H_0**을 기각하지 않음** $\boldsymbol{H_0}$**채택**		제2종의 오류

유의수준(significance level)

$$\alpha = P(\text{제1종 오류})$$

$$= P(H_0\text{을 기각함}|\, H_0\text{ 참})$$

$$= P(Z\ \in R|H_0\text{ 참})$$

6.4 유의확률 $p-$값

- **유의수준**

✓ 유의수준이 먼저 결정되어야 신뢰구간과 기각역이 정해진다.

✓ 유의수준은 사용자나 사용기관에 따라서 달라질 수 있다.

✓ 유의수준에 따라서 신뢰구간과 기각역이 달라진다. 검정통계량이 신뢰구간 또는 기각역에 속하는지 속하지 않는지를 근거로 의사결정을 내린다.

✓ 유의수준과 상관없이 주어진 자료의 검정통계량만으로 계산할 수 있도록 $p-$값을 정의하여 사용한다.

예제 화장품

- μ =50ml?
- n=25, $\bar{x} = 49.05$, $\sigma = 2.21$ (알려짐)
- 유의수준 $\alpha = 0.05$
- $H_0 : \mu = 50$ vs. $H_1 : \mu \neq 50$
- 검정통계량 $Z = \frac{49.05-50}{2.21/\sqrt{25}} = -2.15$
- 기각역 R: $|Z| \geq z_{0.05/2} = 1.96$, $Z \in R$
- 유의수준 0.05에서, $|Z| = |-2.15| = 2.15 > 1.96$이므로, 귀무가설 $H_0 : \mu = 50.0$을 기각한다.

- 그러나... 유의수준을 바꾸면 어떻게 될까?
- 유의수준 0.1에서, $|Z| = |-2.15| = 2.15 > 1.645$이므로, 귀무가설 $H_0 : \mu = 50.0$을 기각한다.
- 유의수준 0.01에서, $|Z| = |-2.15| = 2.15 < 2.575$이므로, 검정통계량 귀무가설 $H_0 : \mu = 50.0$을 기각하지 않는다.

유의확률 $p-$값

- 유의수준을 몰라도 자료가 주어지면, 계산할 수 있는 통계량.
- $\mathrm{p}-$값은 검정통계량의 위치에 해당하는 꼬리의 확률로 정의된다.
- 양측검정에 대한 $\mathrm{p}-$값 $= \mathrm{P}(|\mathrm{Z}| \geq |\mathrm{z}_0|)$
- $\mathrm{p}-$값 $\leq \alpha$이면 H_0 을 기각한다.

예제 화장품 평균용량 μ에 대한 가설검정에서 $p-$값

$H_0: \mu = 50$ vs. $H_1: \mu \neq 50$

$p-$값 $= \mathrm{P}(|Z| \geq |-2.15|) = (2)(0.0158) = 0.0316 < \alpha = 0.05$

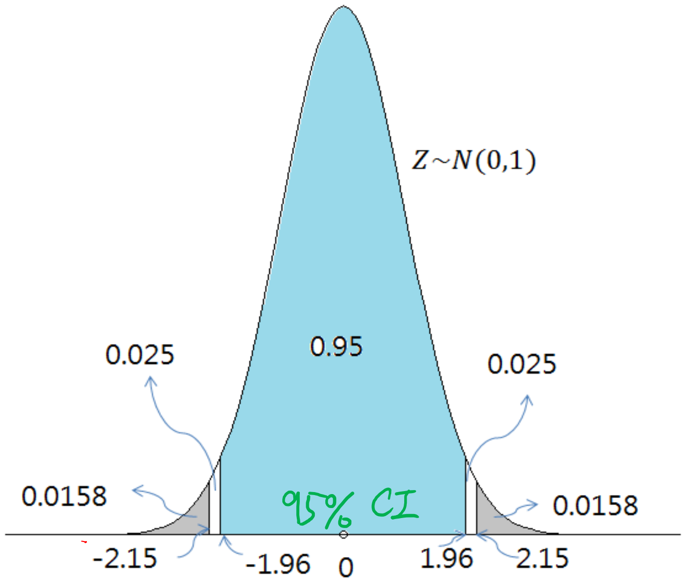
$p-$값 $= \mathrm{P}(|Z| \geq |-2.15|) = 2 * \left(pnorm(-2.15)\right)$

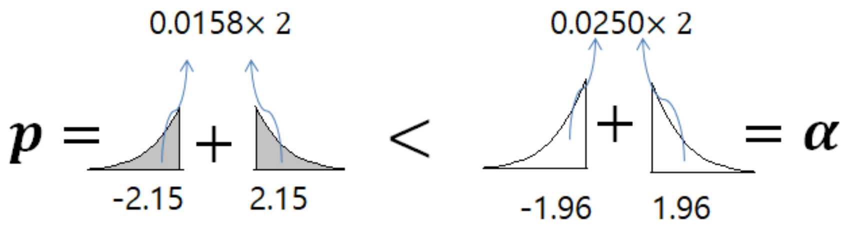
$= 2 * \left(1 - pnorm(abs(-2.15), 0, 1)\right)$

$= 0.0316 < \alpha = 0.05$

$Z \in R$

유의수준 0.05에서 귀무가설 $H_0: \mu = 50.0$을 기각한다





표본분포

표본 $X_1, X_2, \cdots, X_n \sim iid\ N(\mu, \sigma^2)$

σ^2이 알려져 있다고 가정함.

$$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$$

$$Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

신뢰구간

$$P(|Z| < z_{\alpha/2}) = 1-\alpha$$

$$P\left(-z_{\alpha/2} < \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} < z_{\alpha/2}\right) = 1-\alpha$$

$$P\left(-z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \bar{X}-\mu < z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$$

$$P\left(\mu \in \left(\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}},\ \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)\right) = 1-\alpha$$

μ의 100$(1-\alpha)$% 신뢰구간(CI)

$$= \left(\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}},\quad \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$$

가설검정

$H_0 : \mu = \mu_0$ vs. $H_1 : \mu \neq \mu_0$

기각역 R을 이용한 검정

검정통계량 $Z = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$

기각역 $R : |Z| \geq z_{\alpha/2}$

Z가 기각역 R에 속하면, H_0 기각.
Z가 기각역 R에 속하지 않으면, H_0 기각안함

의사결정	H_0 참	H_0 거짓
H_0 기각함 **H_1채택**	제1종의 오류	
H_0 기각안함 H_0채택		제2종의 오류

유의수준(significance level)

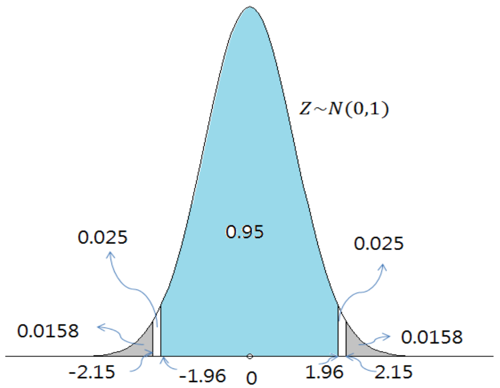
$\alpha = P$(제1종 오류)

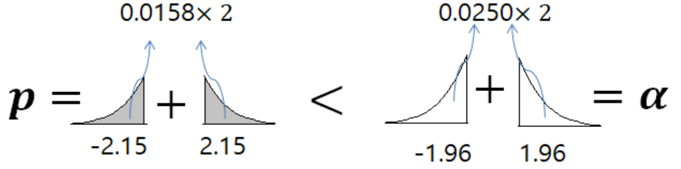
$= P(H_0$ 기각$|\ H_0$ 참$)$

$= H_0$ 기각하는 결정이 틀릴 수 있도록 사회적으로 허용해주는 확률

양측검정에 대한 p − 값 = P(|Z| ≥ |z_0|)

p − 값 ≤ α이면 H_0 을 기각한다.





의사결정

$\mu_o \notin CI, H_0$ 기각

$Z \in$ R, H_0 기각

p − 값 ≤ α, H_0 기각

예제

5. (1-5) 어느 수학 과목 수강생 9명의 학점자료를 이용하여 평균 $\bar{x}$ =3.311을 얻었다. 자료가 정규분포를 따르고, 표준편차가 0.6으로 알려져 있다고 가정하자. 수강생들의 평균평점이 3.8인지를 검정하고자 한다. 소수 3째자리까지만 계산하세요.

① 평점의 평균 μ 에 대한 95% 신뢰구간을 구하라.

(풀이) $(3.311 - 1.96 * \frac{0.6}{\sqrt{9}}, 3.311 + 1.96 * \frac{0.6}{\sqrt{9}}) = (2.919, 3.703)$

② 가설 $H_0 : \mu = 3.8$ vs. $H_1 : \mu \neq 3.8$를 검정하기 위한 검정통계량 Z를 구하라.

(풀이) $Z = \frac{3.311-3.8}{0.6/\sqrt{9}}$ =-2.445

③ 유의수준 $\alpha = 0.05$에서 기각역 R은 무엇이며, Z는 R에 속하는가?

(풀이) R: $|Z| \geq 1.96$ 검정통계량 Z는 기각역 R에 속한다.

④ 유의확률 $p-$값은 얼마인가? 유의확률p는 유의수준 $\alpha = 0.05$보다 작은가?

(풀이) 반올림하여,

$$p - 값 = P(|Z| \geq |-2.445|) = 2 * pnorm(-2.445) = 0.0145$$

$p - 값 = 0.0145 < \alpha = 0.05$. 유의확률이 유의수준보다 작다

⑤ (1-4)에 근거하여, 유의수준 0.05에서 귀무가설을 기각하는가?

(풀이) 유의수준 0.05에서 귀무가설을 기각한다.

⑥ 결론이 무엇인가?

(풀이) 유의수준 0.05에서, 평점 평균이 3.8이라고 말할 수 없다.

혼자 풀기

5. (1-5) 어느 수학 과목 수강생 16명의 학점자료를 이용하여 평균 $\bar{x}$ =3.4를 얻었다. 자료가 정규분포를 따르고, 표준편차가 0.7으로 알려져 있다고 가정하자. 수강생들의 평균평점이 3.6인지를 검정하고자 한다. 소수 3째자리까지만 계산하세요.

① 평점의 평균 μ 에 대한 95% 신뢰구간을 구하라.

② 가설 $H_0 : \mu = 3.6$ vs. $H_1 : \mu \neq 3.6$를 검정하기 위한 검정통계량 Z를 구하라.

③ 유의수준 $\alpha = 0.05$에서 기각역 R은 무엇이며, Z는 R에 속하는가?

④ 유의확률 p –값은 얼마인가? 유의확률p는 유의수준 $\alpha = 0.05$보다 작은가?

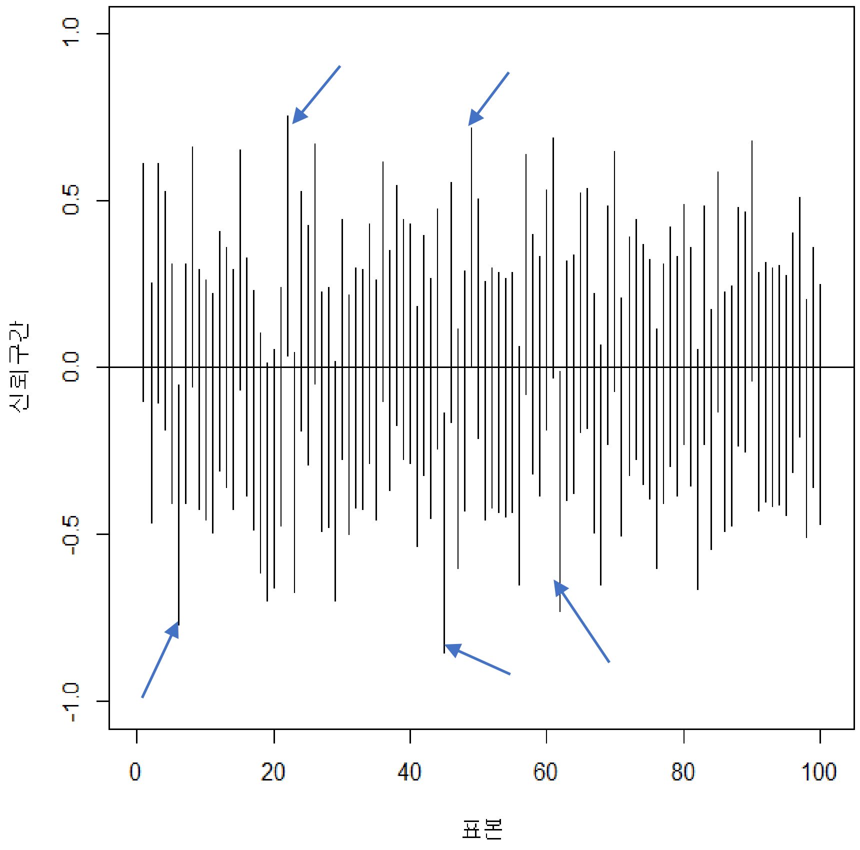
⑤ (1-4)에 근거하여, 유의수준 0.05에서 귀무가설을 기각하는가?

⑥ 결론이 무엇인가?

.

6.5 95% 신뢰구간의 의미

- $X_1, X_2, \cdots, X_n \sim iid\ N(0,1)$
- H_0: $\mu = 0$, H_1: $\mu \neq 0$
- $\alpha = 0.05$
- 표준정규분포 $N(0,1^2)$으로부터 표본크기가 30인 자료를 100개 만든 후, 신뢰구간을 100개 만들자.
- $\mu = 0$을 포함하지 못하는 신뢰구간은 몇 개인가?
- 6, 22, 45, 49, 62



표본 $X_1, X_2, \cdots, X_n \sim iid\ N(\mu, \sigma^2)$ CI $= \left(\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}},\ \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$

표본1 $X_1, X_2, \cdots, X_{30} \sim iid\ N(0,1)$ $\mathrm{CI}_1 = \left(\bar{X} - 1.96\frac{1}{\sqrt{30}},\ \bar{X} + 1.96\frac{1}{\sqrt{30}}\right)$

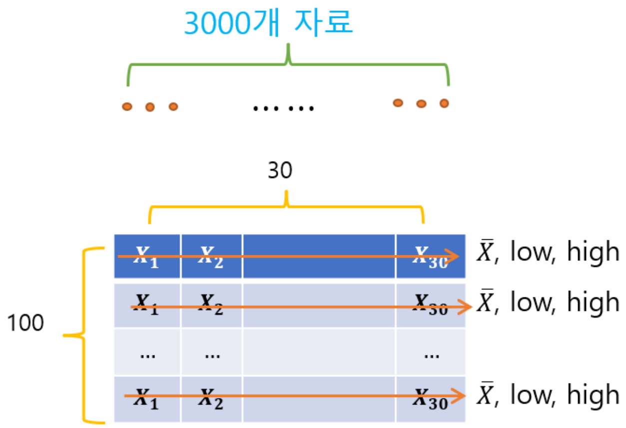
표본2 $X_1, X_2, \cdots, X_{30} \sim iid\ N(0,1)$ $\mathrm{CI}_2 = \left(\bar{X} - 1.96\frac{1}{\sqrt{30}},\ \bar{X} + 1.96\frac{1}{\sqrt{30}}\right)$

…

표본100 $X_1, X_2, \cdots, X_{30} \sim iid\ N(0,1)$ $\mathrm{CI}_{30} = \left(\bar{X} - 1.96\frac{1}{\sqrt{30}},\ \bar{X} + 1.96\frac{1}{\sqrt{30}}\right)$

95% 신뢰구간의 의미 시뮬레이션

```
# 그림 6.2

# 95% 신뢰구간 100개 만들기

x<-rnorm(30*100, 0, 1)          # 3000 개 난수 생성

x<-matrix(x, 100, 30)           # 100 x 30 행렬로 만들기

                                # 각 행이 표본크기 30인 표본에 해당함

x.bar<-apply(x, 1, mean)        # 각 표본(행)의 평균을 계산함

low <- x.bar - 1.96*1/sqrt(30)      #신뢰구간의 하한값 계산

high <- x.bar + 1.96*1/sqrt(30)     #신뢰구간의 상한값 계산

ci<- data.frame(low, high)          # (하한값, 상한값)을 data.frame으로 묶음.

ci
```




```
# 그래프 그리기

plot(NA, xlim=c(0,101), ylim=c(-1, 1), xlab="표본", ylab="신뢰구간")

ci$x0<-1:100 ; ci$y0<-ci$low

ci$x1<-1:100 ; ci$y1<-ci$high

segments(ci$x0,ci$y0, ci$x1,ci$y1)

abline(h=0)

# 0을 포함하지 못한 신뢰구간 찾기

which(high < 0)

which(low > 0)
```

유의수준α와 신뢰구간의 의미 $CI = \left(\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}},\ \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$

유의수준이란 신뢰구간이나 가설검정에서 귀무가설을 기각하는 의사결정을 내릴 때, 틀릴 수 있도록 '사회적으로 허용되는 오류의 확률'이다.

유의수준이 0.05이면, 참인 모평균 μ를 추정하기 위해서, 크기 n인 표본을 100번 추출하여 100개의 신뢰구간을 만들자. 그러면, 100개 중 95개의 신뢰구간은 모평균 μ를 포함하지만, 나머지 5개는 모평균 μ를 포함하지 못할 수도 있다.

실제로는 시간 또는 경제적인 이유로 표본을 한 번만 추출하여, 모평균 μ를 추론한다.

유의수준이 커지면, 신뢰구간이 짧아진다. 표본크기가 커지면, 신뢰구간의 길이가 짧아진다. 표준편차가 작아지면, 신뢰구간이 짧아진다.

조사 대상을 전수조사하지 않는 이상 절대로 알 수 없는 모평균 μ는 망망대해에 떠있는 물고기 한 마리에 비유될 수 있고, 신뢰구간은 그 물고기를 잡으려고 던지는 그물에 비유될 수 있다.

유의수준 0.05 또는 95% 신뢰구간은 그물을 100번 던지면, 95번 그 물고기를 잡고, 5번 그 물고기를 놓칠 수 있다는 의미와 유사하다.

유의수준 0.01 또는 99% 신뢰구간은 그물을 100번 던지면, 99번 그 물고기를 잡고, 1번 정도 물고기를 놓칠 수 있다는 의미와 유사하다.

표본분포

표본 $X_1, X_2, \cdots, X_n \sim iid\ N(\mu, \sigma^2)$

σ^2이 알려져 있다고 가정함.

$$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$$

$$Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

신뢰구간

$$P\left(|Z| < z_{\alpha/2}\right) = 1-\alpha$$

$$P\left(-z_{\alpha/2} < \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} < z_{\alpha/2}\right) = 1-\alpha$$

$$P\left(-z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \bar{X}-\mu < z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$$

$$P\left(\mu \in \left(\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}},\ \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)\right) = 1-\alpha$$

μ의 100$(1-\alpha)$% 신뢰구간(CI)

$$= \left(\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \quad \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$$

가설검정

$H_0 : \mu = \mu_0$ vs. $H_1 : \mu \neq \mu_0$

기각역 R을 이용한 검정

검정통계량 $\quad Z = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$

기각역 R : $|Z| \geq z_{\alpha/2}$

Z가 기각역 R에 속하면, H_0 기각.

Z가 기각역 R에 속하지 않으면, H_0 기각안함

의사결정	H_0 참	H_0 거짓
H_0 기각함 **H_1채택**	제1종의 오류	
H_0 기각안함 **H_0채택**		제2종의 오류

유의수준(significance level)

$\alpha = P$(제1종 오류)

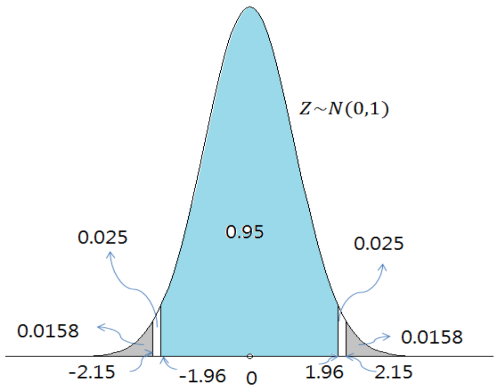
$= P(H_0$ 기각$|$ H_0 참$)$

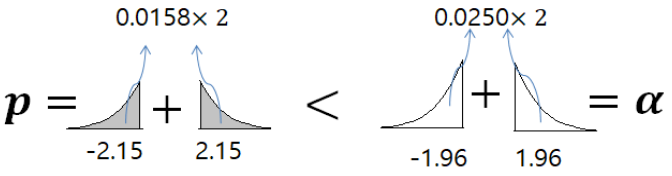
$= H_0$ 기각하는 결정이 틀릴 수

있도록 사회적으로 허용해주는 확률

양측검정에 대한 p − 값 $= P(|Z| \geq |z_0|)$

p − 값 $\leq \alpha$이면 H_0 을 기각한다.





의사결정

$\boldsymbol{\mu_o} \notin CI, H_0$ 기각

$Z \in$ R, H_0 기각

p − 값 $\leq \alpha$, H_0 기각

연습문제

3. 다음 중 옳은 설명은 무엇인가?

a. 주어진 유의수준에 대하여, 표본 수가 커지면 신뢰구간이 길어진다.

b. 주어진 표본의 수에 대하여, 유의수준이 작아지면 신뢰구간이 길어진다.

c. 95% 신뢰구간이라는 것은 동일한 크기의 표본을 100번 만들어서 100개의 신뢰구간을 만들 때, 이들 중 5개가 실제 모평균을 포함하지 못하는 것을 의미한다.

d. 유의확률 p –값이 유의수준보다 작으면 귀무가설을 채택한다.

e. 귀무가설이 참일 때, 신뢰구간과 기각역은 서로 여집합이다.

f. 제 1 종의 오류의 확률은 유의확률과 같다.

μ의 100$(1-\alpha)$% 신뢰구간(CI)= $\left(\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}},\ \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$

과제/퀴즈

- ppt 2쪽 1번 쓰기 (제출)
- 책 153쪽 연습문제 4번 풀기 (제출)
- ppt 24쪽 실습 따라하기 (제출)