01

추정의 이해

01. 추정의 이해

■ 추정의 개념

- **추정(estimation):** 표본의 정보를 근거로 모집단의 통계량을 근사적으로 추측하는 것.

01. 추정의 이해

■ 점추정량

- **점추정량(point estimator):** 모집단의 통계량을 추정하기 위하여 표본의 통계량 하나의 값을 사용하기도 하는데, 이때 표본 통계량 하나의 값.
- 이처럼 하나의 값을 사용하여 모집단의 통계량을 추정하는 것을 '점추정량을 통한 추정'이라 부른다.
- 점추정량을 통한 추정의 한계

  ① 표본의 크기가 작은 경우라면 추정이 부정확할 확률이 높다.

  ② 추정치와 모수의 차이가 어느 정도인지 가늠할 수가 없다.

  ③ 연속확률분포에는 점추정량을 사용할 수 없다.

01. 추정의 이해

■ 구간추정량

- **구간추정량(interval estimator):** 모집단의 통계량을 추정하기 위하여 일정 구간을 명시하는 방법으로 일정 구간 안에 모집단의 통계량이 있을 확률을 보여준다.
- **신뢰구간(confidence interval):** 주어진 확률로 모집단의 통계량이 존재하는 구간을 의미하며, 여기서 주어진 확률을 신뢰수준(confidence level)이라 한다.
- **유의수준(significance level):** 해당 구간에 모집단의 통계량이 존재하지 않을 확률을 말하며, 이것을 α 라고 표시하고 '알파'라 읽는다. 유의수준은 오차의 가능성을 의미한다.
- 신뢰수준은 '1-유의수준'과 동일하다.

01. 추정의 이해

■ 추정량의 바람직한 특징

불편추정량

- **편의(bias):** 표본을 통한 추정량의 기댓값과 실제 모집단의 통계량과의 차이.
- 평균을 예로 들어 보자. 표본평균의 기댓값, 즉 표본평균의 평균과 모집단의 평균의 차이가 편의인 것이다. 모집단의 확률변수를 X, 표본평균의 확률변수를 $\overline{X}$라 할 때 편의는 다음과 같다.

$$E\left(\overline{X}\right)-E(X)=\mu_{\overline{X}}-\mu_X$$

01. 추정의 이해

■ 추정량의 바람직한 특징

최소분산추정량

- **최소분산추정량(minimum variance estimator):** 추정량 표본분포의 분산이 제일 작은 추정량을 의미.

■ 추정량의 바람직한 특징

일치추정량

- **일치추정량(consistent estimator):** 표본의 크기가 커짐에 따라 모집단의 통계량과 일치함을 의미.

02

모평균의 추정

02. 모평균의 추정

하나 더 알기 모표준편차를 알고 있다는 가정의 비현실성

- 여기서 일부 학생은 모표준편차를 알고 있다는 가정이 비현실적이지 않냐는 질문을 하곤 한다. 이런 질문을 하는 학생들은 지금까지 통계학 학습을 제대로 한 학생일 가능성이 매우 높다. 실제 추정의 경우 모표준편차를 모르는 경우가 대부분이다. 그러나 모표준편차를 모른다고 가정하였을 경우만 고려하여 배우면 처음부터 다소 복잡한 내용으로 모평균의 추정이 시작된다. 따라서 다소 비현실적이긴 하지만 모표준편차를 알고 있다는 가정하에 모평균을 추정하는 방식을 우선 학습하고, 이후 보다 현실에 부합하는 모집단의 표준편차를 모른다는 가정하에서 모평균의 추정 방식을 배우도록 하자.

02. 모평균의 추정

■ 모표준편차를 알고 있을 때의 모평균의 추정

- 모집단의 확률변수를 X, 표본평균의 확률변수를 X라 하자. 모집단의 평균이 μ , 표준편차가 σ , 표본의 크기가 n이라고 가정한다. 여기서 모집단의 표준편차 σ 를 알고 있을 때 모평균 μ 를 추정하고자 하는 것이다.
- X는 평균이 모평균과 같은 μ , 표준편차는 $\frac{\sigma}{\sqrt{n}}$인 정규분포를 따르게 되어 다음과 같이 표현 할 수 있다.

$$\overline{X} \sim N\left(\mu, \left(\frac{\sigma}{\sqrt{n}}\right)^2\right)$$

02. 모평균의 추정

■ 모표준편차를 알고 있을 때의 모평균의 추정

- 확률변수에 평균을 빼준 값을 표준편차로 나누면 표준화가 되며 표준화를 거친 새로운 확률변수 Z는 평균이 0, 표준편차가 1인 표준정규분포를 따른다.

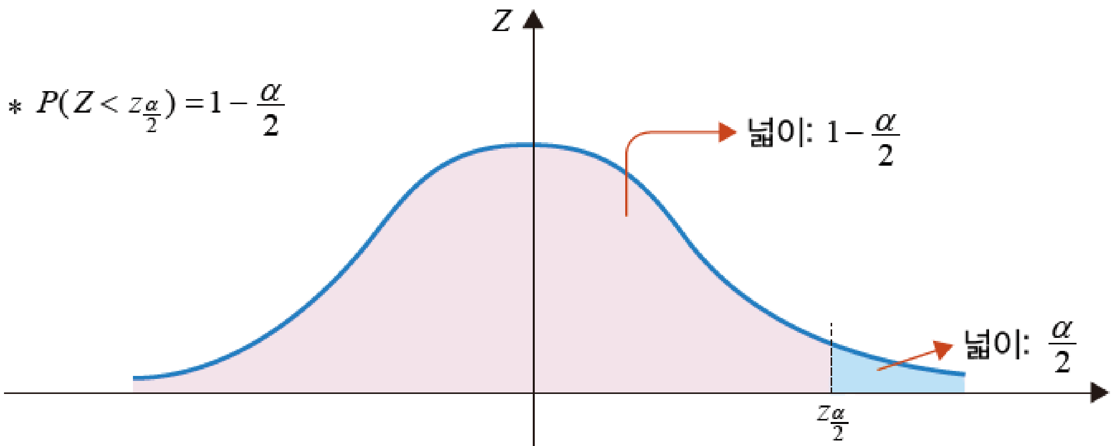
$$Z = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

- 신뢰수준이 1-α 라 할 때의 신뢰 구간을 정규분포의 특징을 이용하여 다음과 같이 표현할 수 있다.

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

- $z_{\frac{\alpha}{2}}$ 는 $P\left(Z > z_{\frac{\alpha}{2}}\right) = \frac{\alpha}{2}$ 인 표준정규확률변수의 값을 의미한다. 반대로, $P\left(Z < z_{\frac{\alpha}{2}}\right) = 1 - \frac{\alpha}{2}$ 이다.

■ 모표준편차를 알고 있을 때의 모평균의 추정

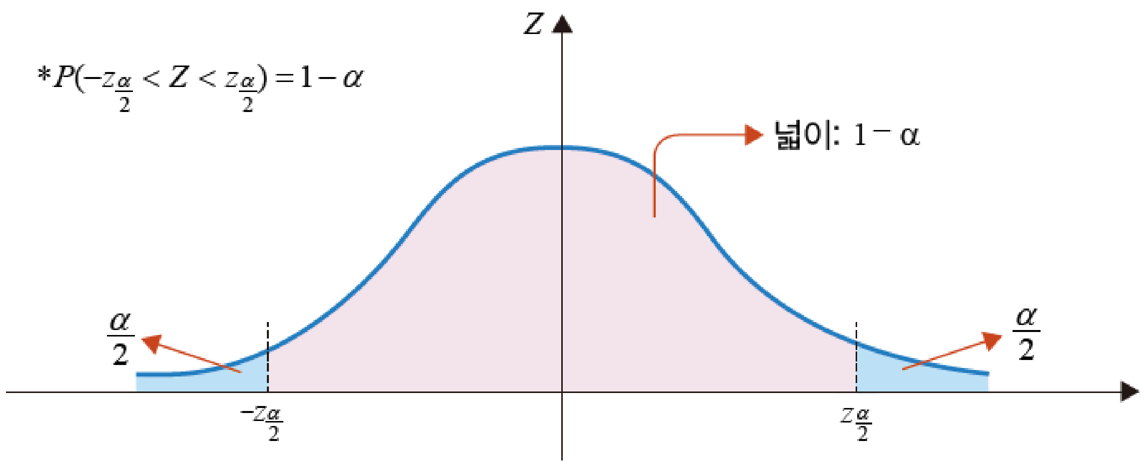


$z_{\frac{\alpha}{2}}$는 표준정규분포에서 $P\left(Z > z_{\frac{\alpha}{2}}\right) = \frac{\alpha}{2}$에 해당하는 표준정규확률변수의 값

02. 모평균의 추정

■ 모표준편차를 알고 있을 때의 모평균의 추정

- 표준정규분포에서 $P\left(Z < z_{\frac{\alpha}{2}}\right) = 1 - \frac{\alpha}{2}$ 라면 확률 $P\left(-z_{\frac{\alpha}{2}} < Z < z_{\frac{\alpha}{2}}\right) = 1 - \alpha$ 이다.



표준정규분포에서 $P\left(-z_{\frac{\alpha}{2}} < Z < z_{\frac{\alpha}{2}}\right)$에 해당하는 부분의 넓이

■ 모표준편차를 알고 있을 때의 모평균의 추정

- 확률 $P\left(\mu - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} < \overline{X} < \mu + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$ 를 표준화하면 다음과 같다.

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

- 여기서 우리가 관심을 갖는 변수는 모평균 μ 이므로 μ 를 기준으로 수식을 변환하면 다음과 같다.

$$\begin{aligned} P\left(-z_{\frac{\alpha}{2}} < \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} < z_{\frac{\alpha}{2}}\right) &= P\left(-z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} < \overline{X}-\mu < z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right) \\ &= P\left(-\overline{X} - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} < -\mu < -\overline{X} + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right) \\ &= P\left(\overline{X} - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right) \end{aligned}$$

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

■ 모표준편차를 알고 있을 때의 모평균의 추정

- 신뢰수준 90%, 95%, 99%를 예로 들어 $z_{\frac{\alpha}{2}}$를 계산하면 [표 9-1]과 같다.

[표 9-1] 신뢰수준에 따른 $z_{\frac{\alpha}{2}}$의 값

신뢰수준 $1-\alpha$	유의수준 α	$\frac{\alpha}{2}$	$z_{\frac{\alpha}{2}}$
90%	10%	0.05	$z_{0.05}=1.645$
95%	5%	0.025	$z_{0.025}=1.96$
99%	1%	0.005	$z_{0.005}=2.575$

예제 9-1 모표준편차를 알고 있을 때의 모평균의 추정 I

중고차 판매원들은 매년 본인이 판매할 차량을 매입하여 재고를 확보한다. 작년 한 해 동안 중고차 판매원 한 명당 평균적으로 몇 대의 자동차를 판매하였는지 추정하고자 한다. 다음은 25명의 표본으로 구성된 각 판매원의 작년 한 해 동안의 자동차 판매 대수이다.

37	34	42	39	33	37	38	38	36	35	34	45
31	37	35	43	39	37	34	43	38	34	36	31
39											

중고차 시장의 오랜 역사적 자료를 통하여 모집단의 표준편차가 4라는 것을 이미 알고 있다. 그렇다면 (1) 95%, (2) 99%, (3) 99.9% 신뢰수준하에서 신뢰구간에 대하여 각각 추정하시오. (단, $z_{0.025}=1.960$, $z_{0.005}=2.576$, $z_{0.0005}=3.291$ 로 계산한다. 여기서 z_t는 표준정규분포를 따르는 Z 확률변수에서 P(Z>z_t) = t 인 값을 의미한다).

예제 9-1 모표준편차를 알고 있을 때의 모평균의 추정 I

풀이

$$\overline{X} = \frac{1}{25}(37+34+42+39+33+\cdots\ \cdots\cdots+36+31+39) = 37.00$$

모표준편차σ = 4, 표본평균 $\overline{X} = 37$ 과 $P\left(\overline{X} - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$ 임을 이용하여 물음에 답하자.

(1) 95% 신뢰수준에서의 신뢰구간 추정

95% 신뢰수준이므로 1-α=95%이고 유의수준 α 는 5%=0.05, $\frac{\alpha}{2}$ 는 2.5%=0.025 이다. 주어진 값에 의하여 $z_{\frac{\alpha}{2}} = z_{0.025} = 1.960$ 이다.

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

$$P\left(37 - 1.96\frac{4}{\sqrt{25}} < \mu < 37 + 1.96\frac{4}{\sqrt{25}}\right) = 1-0.05 = 0.95$$

$$P(35.432 < \mu < 38.568) = 0.95$$

∴ 모집단의 평균이 35.432와 38.568 사이에 존재할 확률이 95%이다.

예제 9-1 모표준편차를 알고 있을 때의 모평균의 추정 I

풀이

(2) 99% 신뢰수준에서의 신뢰구간 추정

99% 신뢰수준이므로 1-α=99%이고 유의수준 α 는 1% 0.01, $\frac{\alpha}{2}$는 0.5%=0.005 이다. 주어진 값에 의하여 $z_{\frac{\alpha}{2}} = z_{0.005} = 2.576$ 이다.

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

$$P\left(37 - 2.576\frac{4}{\sqrt{25}} < \mu < 37 + 2.576\frac{4}{\sqrt{25}}\right) = 1 - 0.01 = 0.99$$

$$P(34.939 < \mu < 39.061) = 0.99$$

∴ 모집단의 평균이 34.939와 39.061 사이에 존재할 확률이 99%이다.

예제 9-1 모표준편차를 알고 있을 때의 모평균의 추정 I

풀이

(3) 99.9% 신뢰수준에서의 신뢰구간 추정

99.9% 신뢰수준이므로 1-α=99.9 %이고 유의수준 α 는 0.1% 0.001, $\frac{\alpha}{2}$는 0.05%=0.0005 이다. 주어진 값에 의하여 $z_{\frac{\alpha}{2}} = z_{0.0005} = 3.291$이다.

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

$$P\left(37 - 3.291\frac{4}{\sqrt{25}} < \mu < 37 + 3.291\frac{4}{\sqrt{25}}\right) = 1 - 0.001 = 0.999$$

$$P(34.367 < \mu < 39.633) = 0.999$$

∴ 모집단의 평균이 34.367과 39.633 사이에 존재할 확률이 99.9%이다.

02. 모평균의 추정

■ 모표준편차를 알고 있을 때의 모평균의 추정

예제 9-2 **모표준편차를 알고 있을 때의 모평균의 추정 II**

작년 한 해 동안 중고차 판매원 한 명당 평균적으로 몇 대의 자동차를 판매하였는지 추정하고자 한다. 표본추출의 결과 표본의 크기가 196명이고 평균이 37.8대였다. 중고차 시장의 오랜 역사적 자료를 통하여 모집단의 표준편차가 4라는 것을 이미 알고 있다.

그렇다면 (1) 95%, (2) 99%, (3) 99.9% 신뢰수준하에서 신뢰구간에 대하여 각각 추정하시오 (단, $z_{0.025}=1.960$, $z_{0.005}=2.576$, $z_{0.0005}=3.291$ 로 계산한다. 여기서 z_t는 표준정규분포를 따르는 Z 확률변수에서 P(Z > z_t)=t인 값을 의미한다).

예제 9-2 모표준편차를 알고 있을 때의 모평균의 추정 II

풀이

$$\overline{X} = 37.80$$

모표준편차 σ = 4와 표본평균 $\overline{X} = 37.80$, $P\left(\overline{X} - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$ 임을 이용하여 물음에 답하도록 하자.

(1) 95% 신뢰수준에서의 신뢰구간 추정

95% 신뢰수준이므로 1-α=95%이고 유의수준 α 는 5% 0.05, $\frac{\alpha}{2}$= 는 2.5% = 0.025이다. 문제에서 주어진 값에 의하여 $z_{\frac{\alpha}{2}} = z_{0.025} = 1.960$ 이다.

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

$$P\left(37.80 - 1.96\frac{4}{\sqrt{196}} < \mu < 37.80 + 1.96\frac{4}{\sqrt{196}}\right) = 1 - 0.05 = 0.95$$

$$P(37.240 < \mu < 38.360) = 0.95$$

∴ 모집단의 평균이 37.240과 38.360 사이에 존재할 확률이 95%이다.

예제 9-2 **모표준편차를 알고 있을 때의 모평균의 추정 II**

풀이

(2) 99% 신뢰수준에서의 신뢰구간 추정

99% 신뢰수준이므로 1-α=99%이고 유의수준 α는 1% 0.01, $\frac{\alpha}{2}$는 0.5%= 0.005 이다. 문제에서 주어진 값에 의하여 $z_{\frac{\alpha}{2}} = z_{0.005} = 2.576$ 이다.

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

$$P\left(37.80 - 2.576\frac{4}{\sqrt{196}} < \mu < 37.80 + 2.576\frac{4}{\sqrt{196}}\right) = 1 - 0.01 = 0.99$$

$$P(37.064 < \mu < 38.536) = 0.99$$

∴ 모집단의 평균이 37.064와 38.536 사이에 존재할 확률이 99%이다.

예제 9-2 모표준편차를 알고 있을 때의 모평균의 추정 II

풀이

(3) 99.9% 신뢰수준에서의 신뢰구간 추정

99.9% 신뢰수준이므로 1-α=99.9%이고 유의수준 α는 0.1% 0.001, $\frac{\alpha}{2}$ 는 0.05%= 0.0005이다. 문제에서 주어진 값에 의하여 $z_{\frac{\alpha}{2}} = z_{0.0005} = 3.291$ 이다.

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

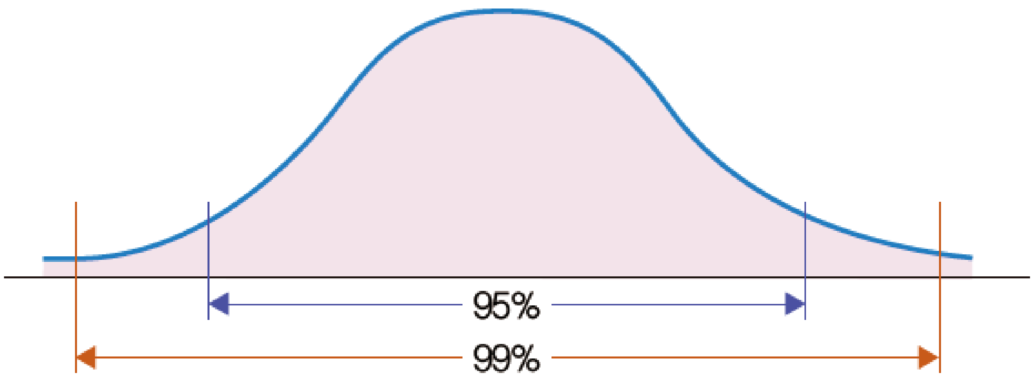
$$P\left(37.80 - 3.291\frac{4}{\sqrt{196}} < \mu < 37.80 + 3.291\frac{4}{\sqrt{196}}\right) = 1 - 0.001 = 0.999$$

$$P(36.860 < \mu < 38.740) = 0.999$$

∴ 모집단의 평균이 36.860과 38.740 사이에 존재할 확률이 99.9%이다.

■ 모표준편차를 알고 있을 때의 모평균의 추정

- 신뢰수준이 커지면 신뢰구간은 더 넓어진다.



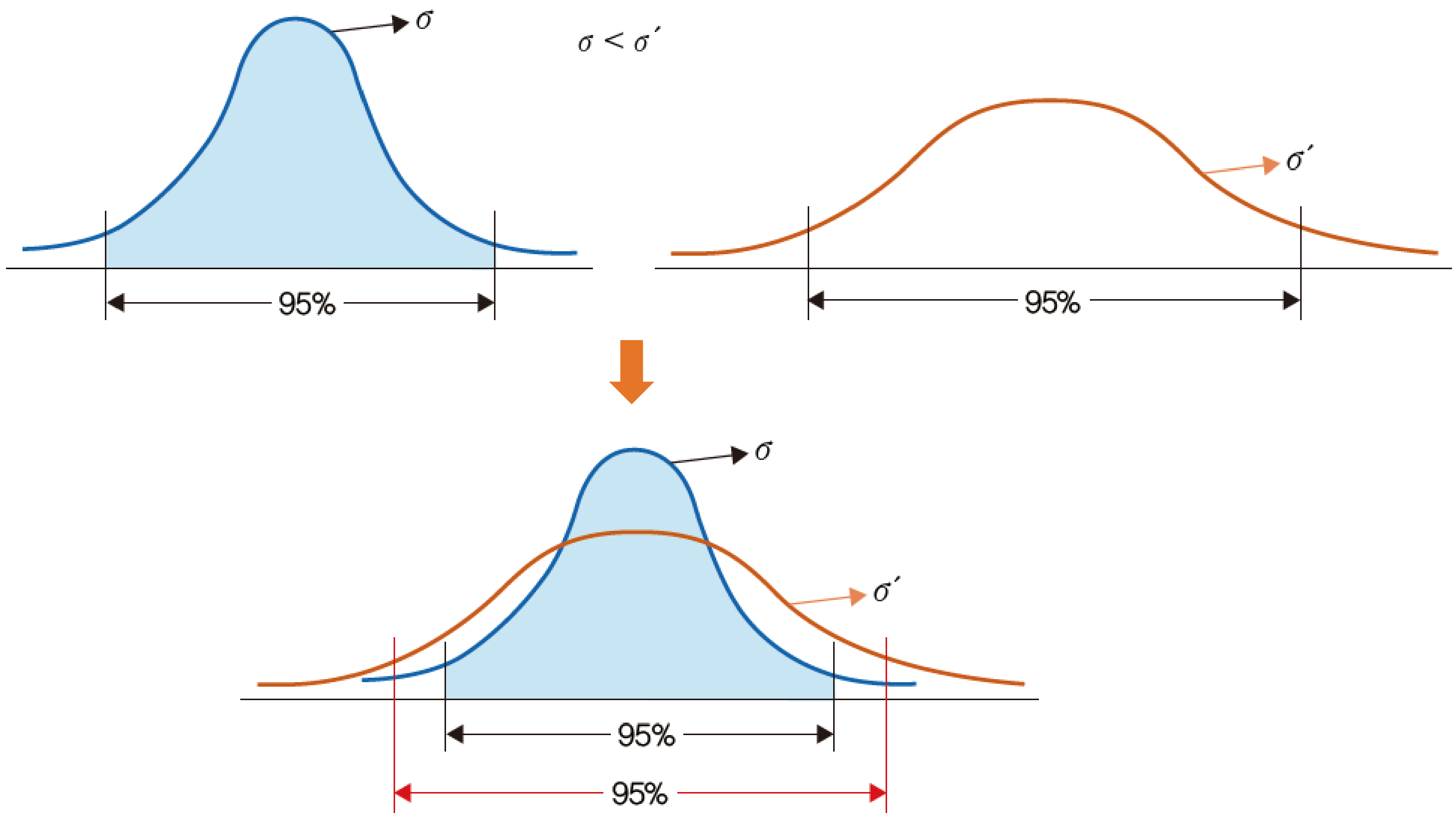
[동일분포에서 신뢰수준에 따른 신뢰구간의 차이]

- 모표준편차 σ 의 값이 커질수록 신뢰구간이 넓어지고 표본의 크기 n이 커질수록 신뢰구간은 더 좁아진다.

$$\overline{X} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

02. 모평균의 추정

■ 모표준편차를 알고 있을 때의 모평균의 추정



[동일 신뢰수준하에서 모표준편차가 다를 때 신뢰구간의 차이]

01

가설검정의 이해

01. 가설검정의 이해

가설검정의 개념

- **가설검정(hypothesis test):** 통계적 추론에 있어서 가장 중요한 영역으로, 표본자료에서 얻은 통계량을 이용하여 모집단의 특성인 모수에 대한 정보를 분석하는 과정.

01. 가설검정의 이해

■ 귀무가설과 대립가설

- **귀무가설(null hypothesis):** 검정하기 위한 모집단의 특성에 대한 가설을 말하고, H_0 로 표현.
- **대립가설(alternative hypothesis):** 표본 데이터 분석에 의하여 귀무가설 H_0 가 거짓이라는 통계적 증거가 충분할 때 대안적으로 채택하기 위한 가설을 말하고, H_1 으로 표현.

■ 귀무가설과 대립가설

[표 10-1] 귀무가설(H_0)검정에 따른 제1종 오류와 제2종 오류

가설검정	H_0 기각하지 않음	H_0 기각
H_0가 참	올바른 결정 $P(H_0$ 기각하지 않음 | H_0 참$) = 1-\alpha$ 신뢰수준(confidence level)	$P(H_0$ 기각 | H_0 참$) = \alpha$ 제1종 오류(Type I error) α위험(α risk) 유의수준(significance level)
H_0가 거짓	$P(H_0$ 기각하지 않음 | H_0 거짓$) = \beta$ 제2종 오류(Type II error) β위험(β risk)	올바른 결정 $P(H_0$ 기각 | H_0 거짓$) = 1-\beta$ 검정력(power of test)

01. 가설검정의 이해

■ 가설검정의 적용

- 어떤 피고인이 범죄 행위로 인하여 검사에게 기소를 당하였다고 가정하자. 실제로 피고인이 유죄인지 무죄인지는 알 수 없다. 다만 주어진 증거(표본)들을 최대한 확보하여 가설을 검정해야 한다. 법에서는 기본적으로 피고인이 무죄라는 가정에서 출발한다는 관점에서 귀무가설 H_0를 다음과 같이 설정하고, 반대로 대립가설 H_1은 다음과 같이 설정한다.

$$H_0\text{: 피고인은 무죄이다.}$$

$$H_1\text{: 피고인은 유죄이다.}$$

가설검정의 적용

[표 10-2] 재판에서 귀무가설(H_0)검정에 따른 제1종 오류와 제2종 오류

가설검정	H_0 기각하지 않음 (피고인에게 무죄 판결)	H_0 기각 (피고인에게 유죄 판결)
H_0가 참 (피고인이 무죄)	올바른 결정 P(무죄판결 | 무죄) = $1-\alpha$	P(유죄판결 | 무죄) = α 제1종 오류(Type I error) α위험(α risk) '무죄인 사람을 유죄로 판결'
H_0가 거짓 (피고인이 유죄)	P(무죄판결 | 유죄) = β 제2종 오류(Type II error) β위험(β risk) '유죄인 범법자에게 무죄 판결	올바른 결정 P(유죄판결 | 유죄) = $1-\beta$

02

모평균의 가설검정

02. 모평균의 가설검정

모평균의 가설검정 개요

- 일반적으로 모평균의 검정을 위한 귀무가설은 다음과 같다.

$$\text{귀무가설}(H_0)\ :\ \text{모집단의 평균}(\mu)\text{은 } \mu_0\text{이다.}$$

- 모평균에 대한 가설검정의 방법은 설정된 귀무가설과 대립가설에 따라 양측검정(two tail test)과 단측검정(one tail test)으로 구분된다. 만약 대립가설(H_1)이 '모집단의 평균(μ)이 μ_0 가 아니다'라면 양측검정일 것이고, '모집단의 평균(μ)이 μ_0 보다 크다' 또는 '모집단의 평균(μ)이 μ_0 보다 작다' 와 같은 진술인 경우 단측검정이 된다.

- 표본의 분포에서 올바른 H_0를 기각하게 될 영역을 기각역이라 한다. 기각역의 넓이는 유의수준이며, 제1종 오류의 확률 α 와 같다.

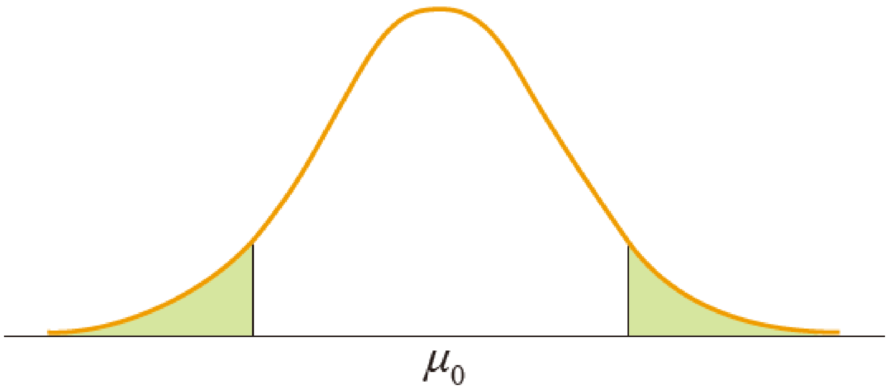
02. 모평균의 가설검정

■ 모평균의 가설검정 개요

- 다음과 같은 귀무가설과 대립가설의 쌍이 있다.

$$H_0\text{: 모평균}(\mu)\text{은 } \mu_0\text{이다. } (\mu = \mu_0)$$

$$H_1\text{: 모평균}(\mu)\text{은 } \mu_0\text{가 아니다. } (\mu \neq \mu_0)$$



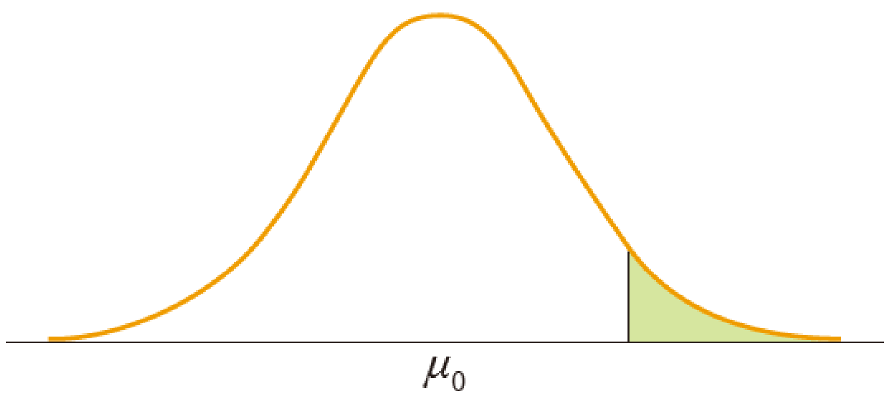
[양측검정의 기각역]

모평균의 가설검정 개요

- 만약 대립가설이 한쪽 영역만을 지정하고 있다면 다음과 같은 귀무가설과 대립가설의 쌍으로 구성된다.

H_0: 모평균(μ)은 μ_0보다 작거나 같다. $(\mu \le \mu_0)$

H_1: 모평균(μ)은 μ_0보다 크다. $(\mu > \mu_0)$



[단측검정(오른쪽 꼬리)의 기각역(right-tail)]

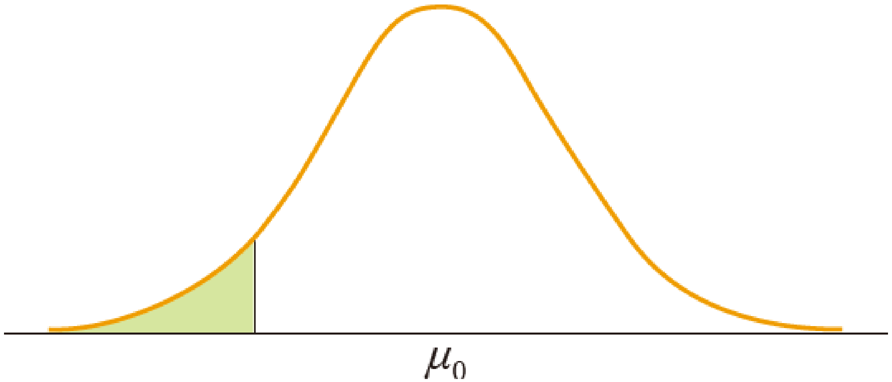
02. 모평균의 가설검정

■ 모평균의 가설검정 개요

- 반대로 대립가설이 분포의 좌측영역에 해당되도록 다음과 같이 가설이 설정되어 있다면 기각역은 반대로 나타날 것이다.

$$H_0\text{: 모평균}(\mu)\text{은 } \mu_0\text{보다 크거나 같다. } (\mu \geq \mu_0)$$

$$H_1\text{: 모평균}(\mu)\text{은 } \mu_0\text{보다 작다. } (\mu < \mu_0)$$



[단측검정(왼쪽 꼬리)의 기각역(left-tail)]

02. 모평균의 가설검정

■ 모표준편차를 알고 있을 때 모평균에 대한 가설검정

- 모표준편차(σ)를 알고 있을 때 모평균(μ)에 대한 검증을 진행해보자. 모표준편차를 알고 있다고 가정하면 표본의 크기 n이 충분히 클 때, 중심극한정리에 의하여 표본평균의 분포 $\overline{X}$는 정규분포를 따른다.

02. 모평균의 가설검정

■ 모표준편차를 알고 있을 때 모평균에 대한 가설검정

$$H_0: \text{모평균}(\mu)\text{은 } \mu_0\text{이다. } (\mu = \mu_0)$$

$$H_1: \text{모평균}(\mu)\text{은 } \mu_0\text{가 아니다. } (\mu \neq \mu_0)$$

- 표본평균 $\overline{X}$

$$\overline{X} \sim N\left(\mu, \left(\frac{\sigma}{\sqrt{n}}\right)^2\right)$$

- 확률변수 Z

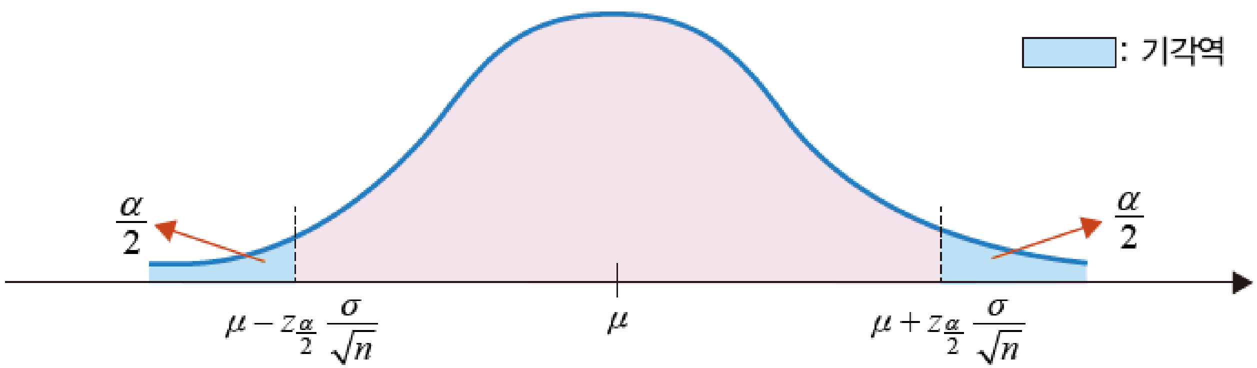
$$Z = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

- 기각할 확률

$$P\left(\overline{X} < \mu - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right) + P\left(\overline{X} > \mu + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right) = \alpha$$

02. 모평균의 가설검정

■ 모표준편차를 알고 있을 때 모평균에 대한 가설검정



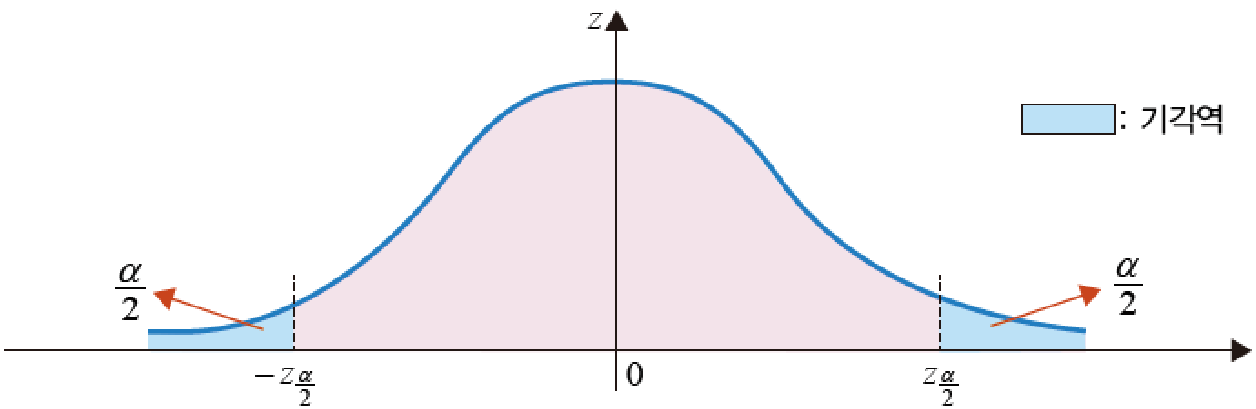
정규분포에서 $P\left(\overline{X} < \mu - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right)$와 $P\left(\overline{X} > \mu + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right)$에 해당하는 기각역의 넓이

- 확률 $P\left(\overline{X} < \mu - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right) + P\left(\overline{X} > \mu + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right) = \alpha$ 를 표준화하면 다음과 같다.

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

02. 모평균의 가설검정

■ 모표준편차를 알고 있을 때 모평균에 대한 가설검정



표준정규분포에서 검정통계량 $z = \dfrac{\overline{X} - \mu}{\dfrac{\sigma}{\sqrt{n}}}$에 따라 귀무가설 기각여부 판단

예제 10-1 **양측검정**

중고차 판매원들은 매년 본인이 판매할 차량을 매입하여 재고를 확보한다. 작년 한 해 동안 중고차 판매원 한 명당 평균적으로 35대를 판매한다는 귀무가설을 설정하였다. 중고차 시장의 오랜 역사적 자료를 통하여 모집단의 표준편차가 4라는 것을 이미 알고 있다. 다음은 25명의 표본으로 구성된 각 판매원의 작년 한 해 동안의 자동차 판매 대수이다.

37	34	42	39	33	37	38	38	36	35	34	45	31
37	35	43	39	37	34	43	38	34	36	31	39	

(1) 5% 유의수준하에서 귀무가설의 기각여부를 확인하시오.

(2) 1% 유의수준하에서 귀무가설의 기각여부를 확인하시오 (단, $z_{0.025} = 1.960$, . $z_{0.005} = 2.576$으로 계산한다. 여기서 z_t는 표준정규분포를 따르는 Z 확률변수에서 P(Z > z_t) = t인 값을 의미한다).

예제 10-1 양측검정

풀이

먼저 귀무가설과 대립가설을 다음과 같이 설정한다.

$$H_0\text{: 모평균}(\mu)\text{은 35이다. } (\mu = 35)$$

$$H_1\text{: 모평균}(\mu)\text{은 35가 아니다. } (\mu \neq 35)$$

또한 표본의 평균 $\overline{X}$를 우선 계산한다.

$$\overline{X} = \frac{1}{25}(37+34+42+39+33+\cdots\cdots\cdots+36+31+39) = 37.00$$

이미 알고 있는 모표준편차 σ = 4와 지금 계산한 표본평균 $\overline{X}$ = 37임을 이용하여 (1) ~ (2)번 물음에 답하도록 하자.

예제 10-1 양측검정

풀이

(1) 5% 유의수준하에서 귀무가설의 기각여부 확인

5% 유의수준이므로 α = 0.05이고 $\frac{\alpha}{2}$ = 0.025이다.

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

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

$$P\left(Z < -z_{\frac{\alpha}{2}}\right) + P\left(Z > z_{\frac{\alpha}{2}}\right) = \alpha$$

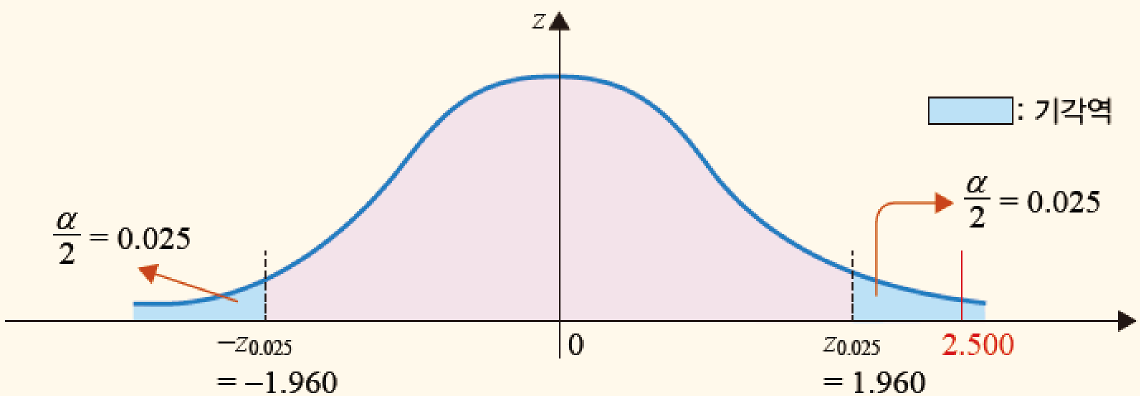
검정통계량은 문제에서 주어진 값에 의하여 다음과 같이 계산한다.

$$z = \frac{\bar{X}-\mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{37-35}{\frac{4}{\sqrt{25}}} = 2.500$$

예제 10-1 양측검정

풀이

2.5는 -1.960과 1.960 사이에 있지 않으므로 [그림 10-6]과 같이 기각역에 존재한다. 따라서 귀무가설을 기각한다.



5% 유의수준에서 검정통계량 $z = \dfrac{\overline{X} - \mu}{\dfrac{\sigma}{\sqrt{n}}}$에 따른 귀무가설 기각여부

∴ 결론적으로 중고차 판매원 한 명당 평균적으로 35대를 판매한다고 볼 수 없다

예제 10-1 양측검정

풀이

(2) 1% 유의수준하에서 귀무가설의 기각여부 확인

1% 유의수준이므로 $\alpha = 0.01$이고 $\frac{\alpha}{2} = 0.005$이다.

$$P\left(\overline{X} < \mu - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right) + P\left(\overline{X} > \mu + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right) = \alpha$$

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

$$P\left(Z < -z_{\frac{\alpha}{2}}\right) + P\left(Z > z_{\frac{\alpha}{2}}\right) = \alpha$$

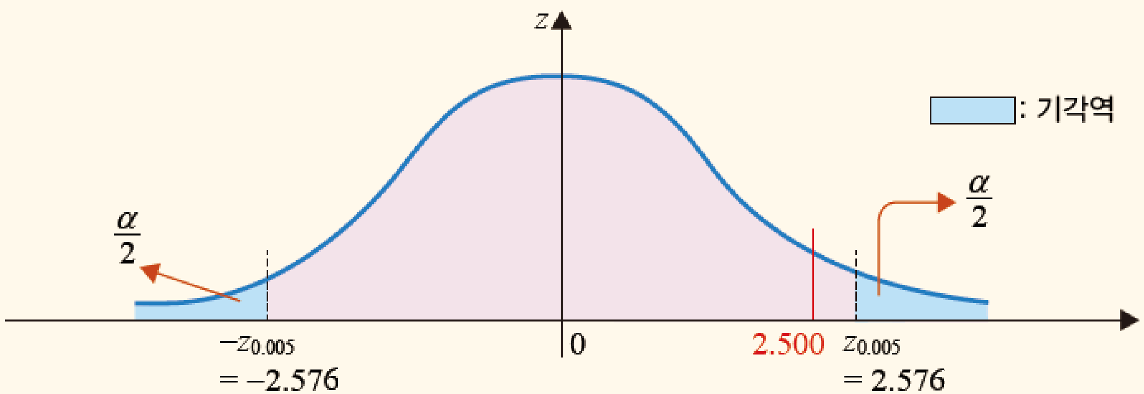
정통계량은 문제에서 주어진 값에 의하여 다음과 같이 계산한다.

$$z = \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{37-35}{\frac{4}{\sqrt{25}}} = 2.500$$

예제 10-1 양측검정

풀이

2.5는 -2.576과 2.576 사이에 있으므로 귀무가설을 기각할 수 없다. 중고차 판매원 한 명당 평균적으로 35대를 판매한다고 볼 수 있다.



1% 유의수준에서 검정통계량 $z = \dfrac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$에 따른 귀무가설 기각여부

∴ 결론적으로 '중고차 판매원 한 명당 평균적으로 35대를 판매한다는 귀무가설을 기각할 수 없다'라고 표현하는 것이 통계학적 진술이다.

02. 모평균의 가설검정

■ 모표준편차를 알고 있을 때 모평균에 대한 가설검정

H_0: 모평균(μ)은 μ_0보다 작거나 같다. ($\mu \le \mu_0$)

H_1: 모평균(μ)은 μ_0보다 크다. ($\mu > \mu_0$)

- 표본평균 $\overline{X}$

$$\overline{X} \sim N\left(\mu, \left(\frac{\sigma}{\sqrt{n}}\right)^2\right)$$

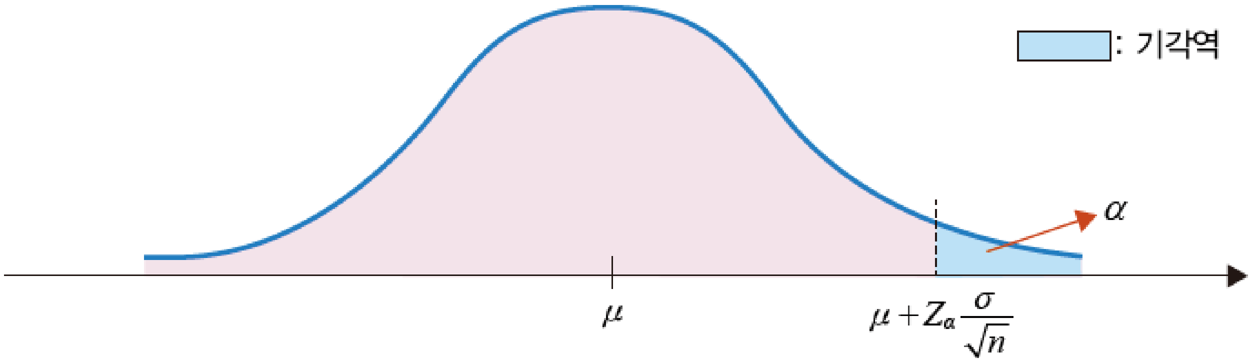
- 확률변수 Z

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

- 기각할 확률

$$P\left(\overline{X} > \mu + z_\alpha \frac{\sigma}{\sqrt{n}}\right) = \alpha$$

■ 모표준편차를 알고 있을 때 모평균에 대한 가설검정



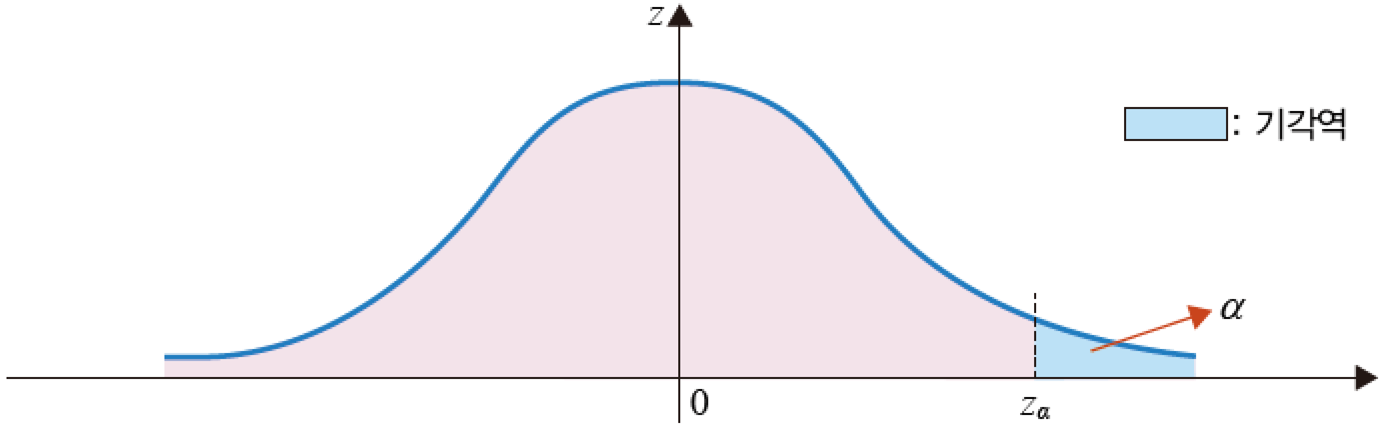
정규분포에서 $P\left(\overline{X} > \mu + z_\alpha \dfrac{\sigma}{\sqrt{n}}\right)$에 해당하는 기각역의 넓이

- 확률 $P\left(\overline{X} > \mu + z_\alpha \dfrac{\sigma}{\sqrt{n}}\right) = \alpha$ 를 표준화하면 다음과 같다.

$$P\left(\frac{\overline{X} - \mu}{\dfrac{\sigma}{\sqrt{n}}} > z_\alpha\right) = \alpha$$

02. 모평균의 가설검정

■ 모표준편차를 알고 있을 때 모평균에 대한 가설검정



표준정규분포에서 검정통계량 $z = \dfrac{\overline{X} - \mu}{\dfrac{\sigma}{\sqrt{n}}}$에 따라 귀무가설 기각여부 판단

예제 10-2 단측검정

중고차 판매원들은 매년 본인이 판매할 차량을 매입하여 재고를 확보한다. 작년 한 해 동안 중고차 판매원 한 명당 평균적으로 35대 이하를 판매한다는 귀무가설을 설정하였다. 중고차 시장의 오랜 역사적 자료를 통하여 모집단의 표준편차가 4라는 것을 이미 알고 있다. 다음은 25명의 표본으로 구성된 각 판매원의 작년 한 해 동안의 자동차 판매 대수이다.

37	34	42	39	33	37	38	38	36	35	34	45	31
37	35	43	39	37	34	43	38	34	36	31	39	

(1) 5% 유의수준하에서 귀무가설의 기각여부를 확인하시오.

(2) 1% 유의수준하에서 귀무가설의 기각여부를 확인하시오 (단, $z_{0.05}=1.645$, $z_{0.01}=2.326$으로 계산한다. 여기서 z_t는 표준정규분포를 따르는 Z 확률변수에서 P(Z > z_t) = t인 값을 의미한다).

예제 10-2 단측검정

풀이

먼저 귀무가설과 대립가설을 다음과 같이 설정한다.

$$H_0: \text{모평균}(\mu)\text{은 } 35 \text{ 이하이다. } (\mu \le 35)$$
$$H_1: \text{모평균}(\mu)\text{은 } 35 \text{ 초과이다. } (\mu > 35)$$

또한 표본의 평균 $\overline{X}$를 우선 계산한다.

$$\overline{X} = \frac{1}{25}(37+34+42+39+33+\cdots\cdots\cdots+36+31+39) = 37.00$$

이미 알고 있는 모표준편차 σ = 4와 지금 계산한 표본평균 $\overline{X}$ = 37임을 이용하여 (1) ~ (2)번 물음에 답하도록 하자.

예제 10-2 단측검정

풀이

(1) 5% 유의수준하에서 귀무가설의 기각여부 확인

5% 유의수준이므로 α = 0.05이다.

$$P\left(\overline{X} > \mu + z_\alpha \frac{\sigma}{\sqrt{n}}\right) = \alpha,\ P\left(\frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} > z_\alpha\right) = \alpha,\ P(Z > z_\alpha) = \alpha$$

검정통계량은 문제에서 주어진 값에 의하여 다음과 같이 계산한다.

$$z = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{37 - 35}{\frac{4}{\sqrt{25}}} = 2.500$$

2.5는 1.645보다 큰 기각역에 존재한다. 따라서 귀무가설을 기각한다.

예제 10-2 **단측검정**

풀이

표본평균을 기준으로 기각역을 확인한다면 $\dfrac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}>1.645$에서 표본평균 $\overline{X}$가 얼마인지 확인함으로써 비교해볼 수도 있다.

$$\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}>1.645,\ \frac{\overline{X}-35}{\frac{4}{\sqrt{25}}}>1.645$$

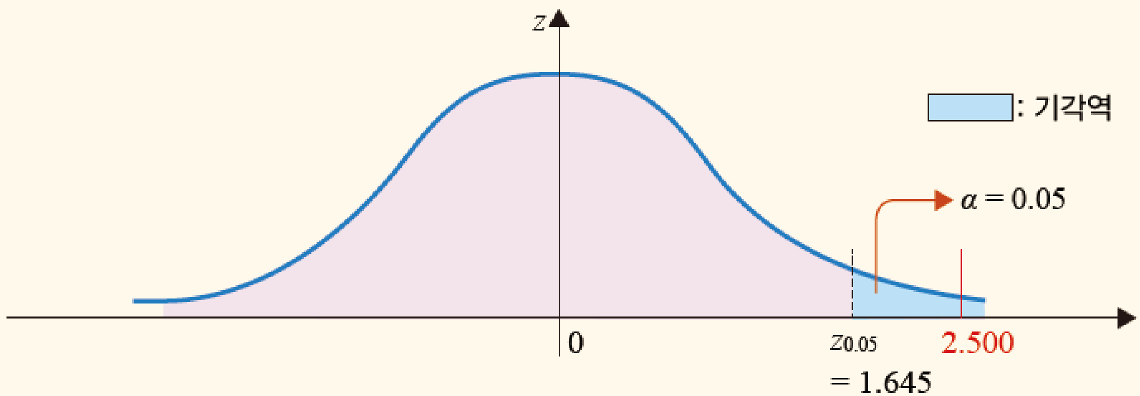
$$\overline{X}-35>1.645\frac{4}{\sqrt{25}},\ \overline{X}>35+1.645\frac{4}{\sqrt{25}}$$

$$\overline{X}>36.316$$

즉, 표본평균이 36.316보다 크면 귀무가설을 기각한다고 볼 수 있다. 따라서 이 문제에서 표본평균이 37이었으므로 귀무가설을 기각한다.

예제 10-2 **단측검정**

풀이



5% 유의수준에서 검정통계량 $z = \dfrac{\overline{X} - \mu}{\dfrac{\sigma}{\sqrt{n}}}$에 따른 귀무가설 기각여부

∴ 결론적으로 5% 유의수준하에서 중고차 판매원 한 명당 평균적으로 35대 이하를 판매한다고 볼 수 없다.

예제 10-2 단측검정

풀이

(2) 1% 유의수준하에서 귀무가설의 기각여부 확인

1% 유의수준이므로 α = 0.01이다.

$$P\left(\overline{X} > \mu + z_\alpha \frac{\sigma}{\sqrt{n}}\right) = \alpha$$

$$P\left(\frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} > z_\alpha\right) = \alpha$$

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

검정통계량은 문제에서 주어진 값에 의하여 다음과 같이 계산한다.

$$z = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{37 - 35}{\frac{4}{\sqrt{25}}} = 2.500$$

2.5는 2.326보다 큰 기각역에 존재한다. 따라서 귀무가설을 기각한다.

예제 10-2 단측검정

풀이

표본평균을 기준으로 기각역을 확인한다면 $z = \frac{\bar{X}-\mu_0}{\frac{\sigma}{\sqrt{n}}} > 2.326$에서 표본평균 $\overline{X}$가 얼마인지 확인함으로써 비교해볼 수도 있다.

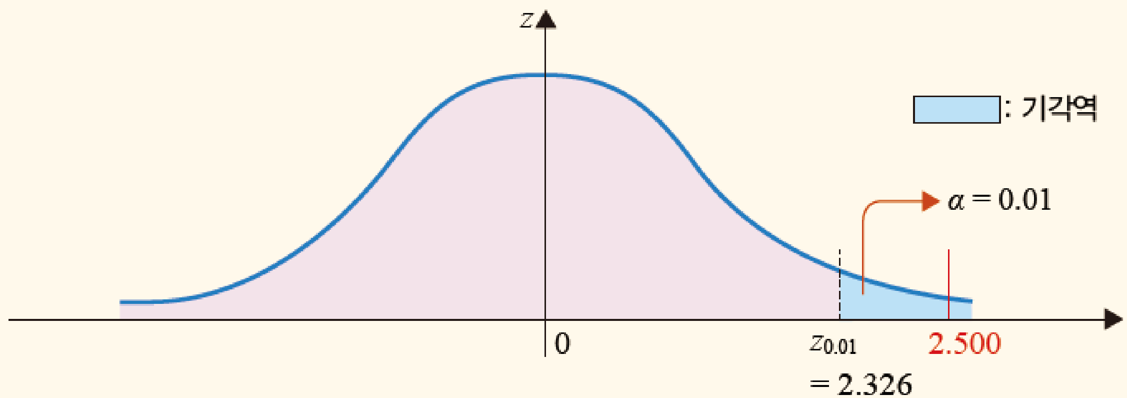
$$z = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} > 2.326, \qquad \frac{\bar{X} - 35_0}{\frac{4}{\sqrt{25}}} > 2.326$$

$$\overline{X} - 35 > 2.326\frac{4}{\sqrt{25}}, \quad \overline{X} > 35 + 2.326\frac{4}{\sqrt{25}}$$
$$\overline{X} > 36.861$$

즉, 표본평균이 36.861보다 크면 귀무가설을 기각한다고 볼 수 있다. 따라서 이 문제에서 표본평균이 37이었으므로 귀무가설을 기각한다.

예제 10-2 단측검정

풀이



1% 유의수준에서 검정통계량 $z = \dfrac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$에 따른 귀무가설 기각여부

∴ 결론적으로 1% 유의수준하에서 중고차 판매원 한 명당 평균적으로 35대 이하를 판매한다고 볼 수 없다.

02. 모평균의 가설검정

제2종 오류의 계산

- **제2종 오류:** H_0 가 거짓임에도 H_0 를 채택하는(기각하지 못하는) 오류를 범하는 확률이며 이것을 β 위험(β risk)이라 한다.
- 제1종 오류(Type I error)와 제2종 오류(Type II error)는 서로 상충 관계를 보인다.
- 제1종 오류의 확률 α 는 올바른 귀무가설을 기각할 확률로 유의수준이라고 한다. 반대로 1-α 는 올바른 귀무가설을 채택할 확률로 신뢰수준이라 한다. 제2종 오류의 확률 β 에서 1-β 는 틀린 귀무가설을 기각할 확률로 검정력(power of test)이라 한다.

예제 10-3 제2종 오류의 계산

중고차 판매원들은 매년 본인이 판매할 차량을 매입하여 재고를 확보한다. 작년 한 해 동안 중고차 판매원 한 명당 평균적으로 35대 이하를 판매한다는 귀무가설을 설정하였다. 중고차 시장의 오랜 역사적 자료를 통하여 모집단의 표준편차가 4라는 것을 이미 알고 있다. 다음은 25명의 표본으로 구성된 각 판매원의 작년 한 해 동안의 자동차 판매 대수이다.

37	34	42	39	33	37	38	38	36	35	34	45	31
37	35	43	39	37	34	43	38	34	36	31	39	

실제 모집단의 평균(μ)이 38일 때,

(1) 5% 유의수준하에서 제1종 오류(α)와 제2종 오류(β)를 계산하고 이를 그래프에서 표현하시오.

(2) 1% 유의수준하에서 제1종 오류(α)와 제2종 오류(β)를 계산하고 이를 그래프에서 표현하시오.

예제 10-3 **제2종 오류의 계산**

(3) 제1종 오류와 제2종 오류의 관계에 대하여 확인하고, 두 가지의 오류를 동시에 줄일 수 있는 방법을 제시하시오(단, $z_{0.05}=1.645$, $z_{0.01}=2.326$ 으로 계산한다. 여기서 z_t 는 표준정규분포는 따르는 Z 확률변수에서 P(Z > z_t) = t인 값을 의미한다).

풀이

먼저 귀무가설과 대립가설을 다음과 같이 설정한다.

$$H_0\text{: 모평균}(\mu)\text{은 35 이하이다. } (\mu \le 35)$$

$$H_1\text{: 모평균}(\mu)\text{은 35 초과이다. } (\mu > 35)$$

또한 표본의 평균 $\overline{X}$를 우선 계산한다.

$$\overline{X} = \frac{1}{25}(37+34+42+39+33+\cdots\cdots\cdots+36+31+39) = 37.00$$

(1) ~ (3)번 물음에 답하도록 하자.

예제 10-2 단측검정

풀이

(1) 5% 유의수준하에서 제1종 오류(α)와 제2종 오류(β)를 계산하고 이를 그래프에서 표현하시오.

5% 유의수준이므로 제1종 오류에 해당하는 α = 0.05이다. 해당 기각역은 $\overline{X} > 36.316$ 이었다.

$$\alpha = 0.05 = P\left(\overline{X} > 36.316 \mid \mu \le 35\right)$$

본 문제에서는 모평균 μ 가 38임이 주어졌다. 따라서 제2종 오류의 확률은 다음과 같다.

$$\beta = P\left(\overline{X} \le 36.316 \mid \mu = 38\right)$$

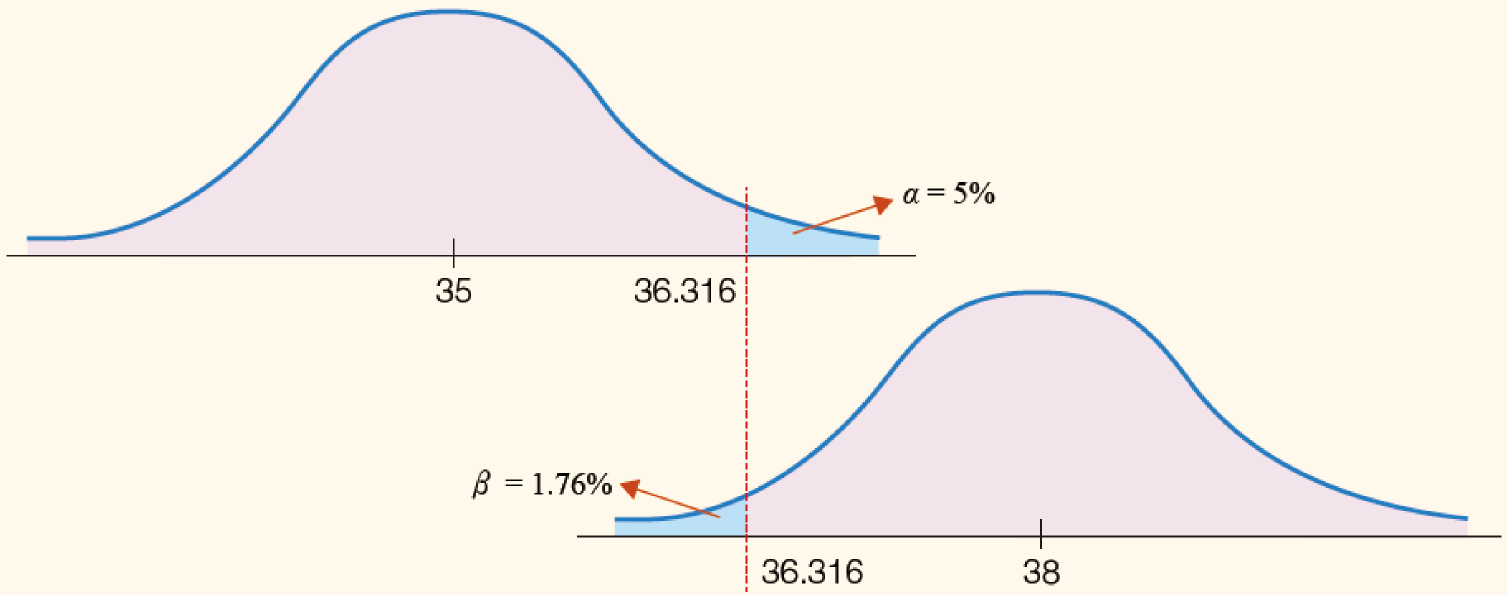
$$\beta = P\left(\overline{X} \le 36.316 \mid \mu = 38\right) = P\left(\frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \le \frac{36.316 - \mu}{\frac{\sigma}{\sqrt{n}}}\right)$$

$$= P\left(\frac{\overline{X} - 38}{\frac{4}{\sqrt{25}}} \le \frac{36.316 - 38}{\frac{4}{\sqrt{25}}}\right) = P(Z \le -2.105) = 0.0176$$

예제 10-2 단측검정

풀이

∴ 즉, 제2종 오류의 확률은 1.76%인 것이다. 검정력은 1-β 로 98.24%이다.



[5% 유의수준하에서 제1종 오류와 제2종 오류와의 관계]

예제 10-2 단측검정

풀이

(2) 1% 유의수준하에서 제1종 오류(α)와 제2종 오류(β)를 계산하고 이를 그래프에서 표현하시오.

1% 유의수준이므로 제1종 오류에 해당하는 α = 0.01이다. 해당 기각역은 .$\overline{X} > 36.861$ 이었다.

$$\alpha = 0.01 = P\left(\overline{X} > 36.861 \mid \mu \le 35\right)$$

본 문제에서는 모평균 μ 가 38임이 주어졌다. 따라서 제2종 오류의 확률은 다음과 같다.

$$\beta = P\left(\overline{X} \le 36.861 \mid \mu = 38\right)$$

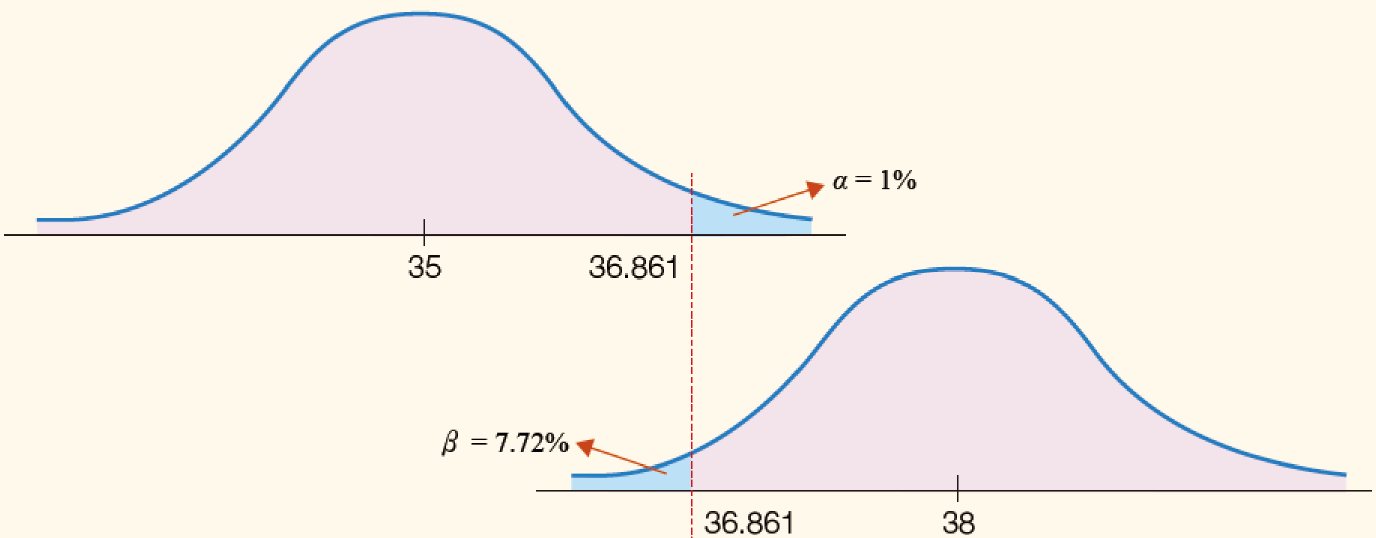
$$\beta = P\left(\overline{X} \le 36.861 \mid \mu = 38\right) = P\left(\frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \le \frac{36.861 - \mu}{\frac{\sigma}{\sqrt{n}}}\right)$$

$$= P\left(\frac{\overline{X} - 38}{\frac{4}{\sqrt{25}}} \le \frac{36.861 - 38}{\frac{4}{\sqrt{25}}}\right) = P(Z \le -1.424) = 0.0772$$

예제 10-2 단측검정

풀이

∴ 즉, 제2종 오류의 확률은 7.72%인 것이다. 동시에 검정력은 1-β 로 92.28%이다. 역시 제1종 오류의 확률을 5%에서 1%로 낮춤과 동시에 제2종 오류는 1.76%에서 7.72%로 올라가는 것을 확인하였다.



[1% 유의수준하에서 제1종 오류와 제2종 오류와의 관계]

예제 10-2 단측검정

풀이

(3) 제1종 오류와 제2종 오류의 관계에 대하여 확인하고, 두 가지의 오류를 동시에 줄일 수 있는 방법을 제시하시오.

가장 간편하고 강력한 방법은 표본의 수 n을 늘리는 방법이다. (2)번 문제에서 표본의 수가 49개라고 가정하고 문제를 풀어보면 명확히 확인할 수 있다.

1% 유의수준이므로 제1종 오류에 해당하는 α = 0.01이다. 해당 기각역을 다음과 같이 다시 계산할 수 있다.

$$\frac{\overline{X}-35}{\frac{4}{\sqrt{49}}} > 2.326$$

$$\overline{X}-35 > 2.326\frac{4}{\sqrt{49}}$$

$$\overline{X} > 35 + 2.326\frac{4}{\sqrt{49}}$$

$$\overline{X} > 36.329$$

예제 10-2 단측검정

풀이

표본의 수가 증가하면서 기각역은 $\overline{X}$ > 36.329로 바뀐다. 제1종 오류 1%는 귀무가설 H_0가 옳다는 가정하에 H_0를 기각할 확률을 말한다.

$$\alpha = 0.01 = P\left(\overline{X} > 36.329 \mid \mu \le 35\right)$$

제2종 오류 β 는 거짓인 귀무가설 H_0가 기각되지 않을 확률을 말하며 다음과 같이 계산한다.

$$\beta = P\left(\overline{X} \le 36.329 \mid \mu = 38\right)$$

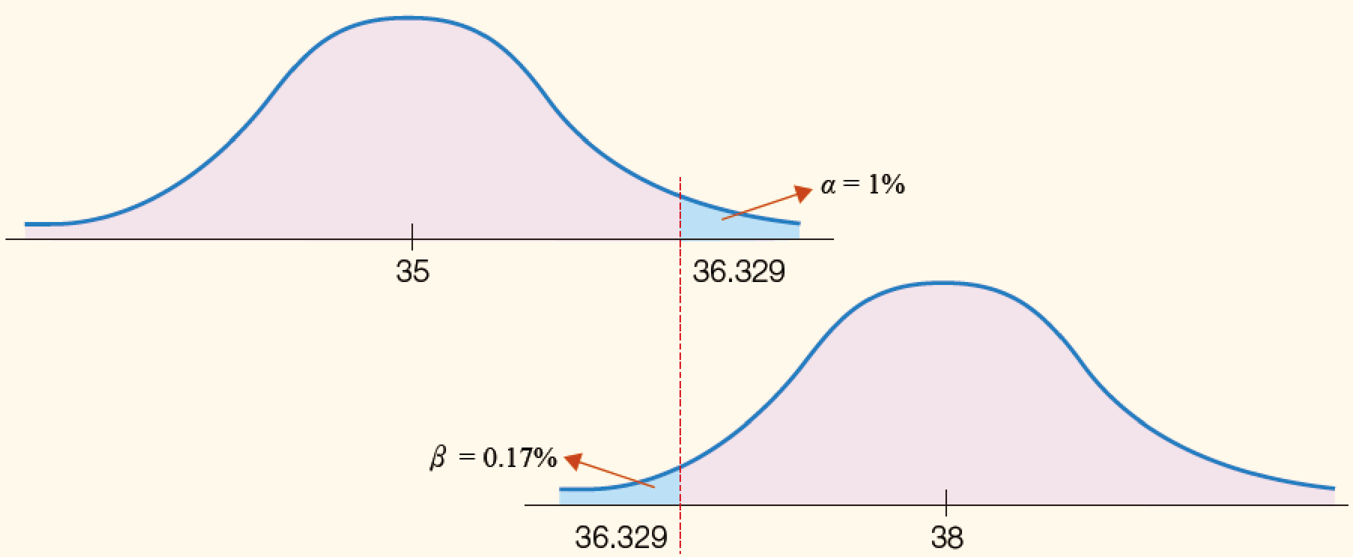
$$\beta = P\left(\overline{X} \le 36.329 \mid \mu = 38\right) = P\left(\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \le \frac{36.329-\mu}{\frac{\sigma}{\sqrt{n}}}\right) = P\left(\frac{\overline{X}-38}{\frac{4}{\sqrt{49}}} \le \frac{36.329-38}{\frac{4}{\sqrt{49}}}\right)$$

$$= P(Z \le -2.924) = 0.0017$$

예제 10-2 단측검정

풀이

∴ 즉 제2종 오류의 확률은 0.17%인 것이다. 표본의 수를 증가시킴으로 제1종 오류는 그대로인 상황에서 제2종 오류를 급격히 줄일 수 있다. 제2종 오류가 줄면서 검정력(1-β)이99.83%로 증가하게 된다.



[1% 유의수준과 증가된 표본하에서 제1종 오류와 제2종 오류와의 관계]