생생한 사례로 배우는 확률과 통계

[강의교안 이용 안내]

한빛아카데미
Hanbit Academy, Inc.

Chapter 10

생생한 사례로 배우는
확률과 통계
PROBABILITY & STATISTICS

대표본 가설검정

Large Sample Test of Hypotheses

목 차

10.1 통계적 가설검정

가설검정의 의미

통계적 가설statistical hypothesis
: 표본의 특성을 나타내는 모수에 대한 주장

예 **어느 사회단체에서 우리나라 근로자의 혈중 콜레스테롤 평균수치가 220mg/dl이라고 발표한다면, $\mu = 220$이라는 사회단체의 주장**

어느 철강회사 A에서 생산한 H빔의 평균 강도(μ_1)가 경쟁 회사 B에서 생산한 H빔의 평균 강도(μ_2)와 동일하다는 주장, 즉 $\mu_1 = \mu_2$이라는 주장

검정 이러한 주장이 참인지 거짓인지 표본을 이용하여 검정한다.

가설검정의 의미

가설검정hypothesis test : 표본통계량을 이용하여 모수에 대한 주장의 진위를 검정하는 과정

귀무가설null hypothesis : 거짓이 명확히 규명될 때까지 참인 것으로 인정되는 모수에 대한 주장, 즉 그 타당성을 입증해야 할 가설이고 H_0으로 나타낸다.

대립가설alternative hypothesis : 귀무가설을 부정하는 가설, 즉 귀무가설이 거짓이라면 참이 되는 가설이고 H_1로 나타낸다.

가설검정의 의미

예

어느 사회단체에서 우리나라 근로자의 혈중 콜레스테롤 평균수치가 220mg/dl이라고 발표한다면, $m = 220$이라는 사회단체의 주장

귀무가설은 $H_0 : \mu = 220$이고, 이에 대한 대립가설은 다음과 같다.

$$H_1 : \mu < 220, \quad H_1 : \mu \neq 220, \quad H_1 : \mu > 220$$

❖ 귀무가설은 항상 등호(=)를 사용하고 대립가설에는 등호를 사용하지 않는다.
즉, 모수 θ에 대한 귀무가설은 반드시 다음과 같이 ≥, =, ≤를 사용한다.

$$H_0 : \theta \leq \theta_0, \quad H_0 : \theta = \theta_0, \quad H_0 : \theta \geq \theta_0$$

이에 대한 대립가설은 각각 다음과 같다.

$$H_1 : \theta < \theta_0, \quad H_1 : \theta \neq \theta_0, \quad H_1 : \theta > \theta_0$$

가설검정의 의미

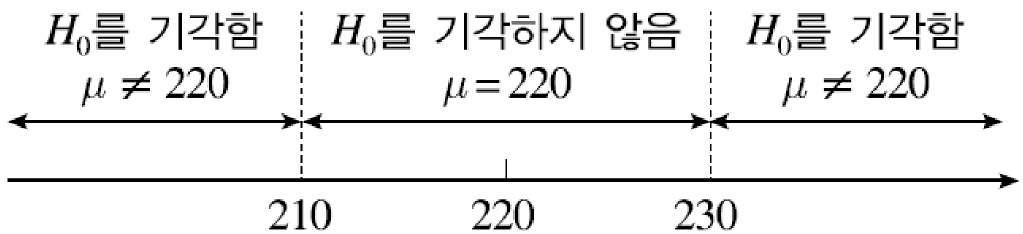
검정통계량test statistic : 귀무가설 H_0의 진위여부를 판정하기 위해 표본으로부터 얻은 통계량

채택accept 또는 **기각**rejec t : 검정 결과 H_0이 참인 결과를 얻으면, 귀무가설을 채택한다고 한다. 그리고 검정 결과 H_0이 거짓인 결과를 얻으면, 귀무가설을 기각한다고 한다.

채택역acceptance region : 귀무가설 H_0 을 채택하는 검정통계량의 영역

기각역critical region : 귀무가설 H_0 을 기각하는 검정통계량의영역

가설검정의 의미



❖ 검정 결과 :

실제 상황 / 검정 결과	H_0가 참	H_0가 거짓
H_0를 채택	옳은 결정	제2종 오류
H_0를 기각	제1종 오류	옳은 결정

❖ 제 1종 오류 : 실제 H_0이 참이지만 H_0을 기각함으로써 발생하는 오류

제 2종 오류 : 실제 H_0이 거짓이지만 H_0을 채택함으로써 발생하는 오류

가설검정의 의미

유의수준 significance level : 제 1종 오류를 범할 확률 (α로 표시)
보편적으로 유의수준은 0.01, 0.05, 0.1을 많이 사용한다.

NOTE **유의수준의 의미**

구간추정의 신뢰도와 비슷하게 유의수준이 $\alpha = 0.05$라는 것은 원칙적으로 기각할 것을 예상하여 설정한 가설을 기각한다고 하더라도 그것에 의한 오차는 최대 5%이하임을 나타낸다.

다시 말해서, 유의수준 $\alpha = 0.05$는 귀무가설 H_0이 참이지만 H_0을 기각함으로써 발생하는 오류를 범할 위험이 20회의 검정에서 최대 1회까지만 허용하는 것을 의미하며, 추정에서 사용하는 신뢰도 95%와 반대되는 개념으로 생각할 수 있다.

검정의 유형과 절차

❖ 모수 $\theta = \theta_0$에 대한 세 가지 유형의 귀무가설과 대립가설

$$\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta > \theta_0 \end{cases} \qquad \begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta \neq \theta_0 \end{cases} \qquad \begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta < \theta_0 \end{cases}$$

또는

$$\begin{cases} H_0 : \theta \leq \theta_0 \\ H_1 : \theta > \theta_0 \end{cases} \qquad \begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta \neq \theta_0 \end{cases} \qquad \begin{cases} H_0 : \theta \geq \theta_0 \\ H_1 : \theta < \theta_0 \end{cases}$$

❖ **양측검정** : 귀무가설 $H_0 : \theta = \theta_0$에 대한 대립가설 $H_1 : \theta \neq \theta_0$ 인 검정

하단측검정 : 귀무가설 $H_0 : \theta \geq \theta_0$에 대한 대립가설 $H_1 : \theta < \theta_0$ 인 검정

상단측검정 : 귀무가설 $H_0 : \theta \leq \theta_0$에 대한 대립가설 $H_1 : \theta > \theta_0$ 인 검정

검정 순서

❶ 대립가설 H_1을 설정한다. 이때 등호는 항상 귀무가설에서 사용한다.

❷ 유의수준 α를 정한다.

❸ 적당한 검정통계량을 선택한다.

❹ 유의수준 α에 대한 기각역을 구한다.

❺ 표본으로부터 검정통계량의 관찰값을 구한다.

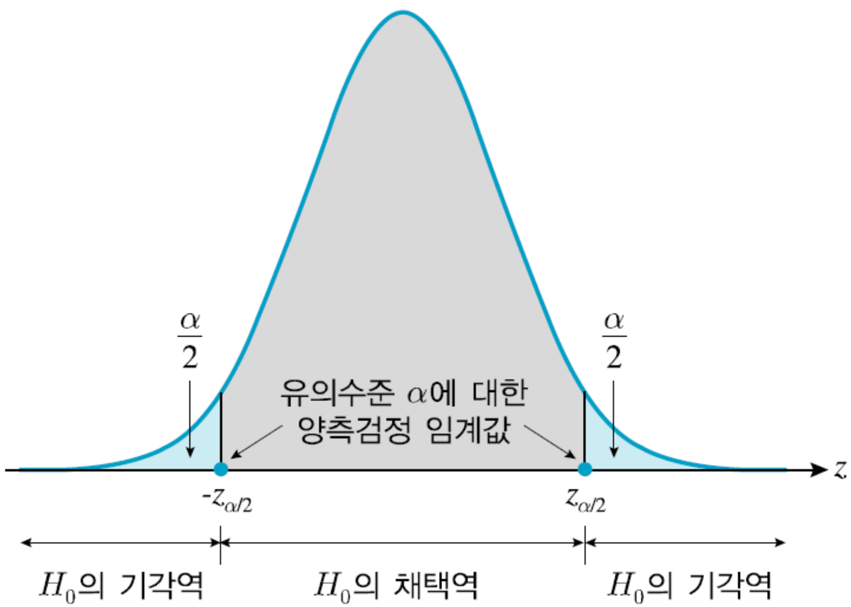
❻ 관찰값이 기각역 안에 들어 있으면 귀무가설 H_0을 기각시키고,
그렇지 않으면 H_0을 기각시키지 않는다.

양측검정

귀무가설 $H_0 : \theta = \theta_0$에 대한 대립가설 $H_1 : \theta \neq \theta_0$으로 구성된 검정

❖ 유의수준을 α라 하면, 양쪽 꼬리확률이 각각 $\alpha/2$가 되는 두 임계값 $\pm z_{\alpha/2}$에 의해 세 영역으로 분리된다.

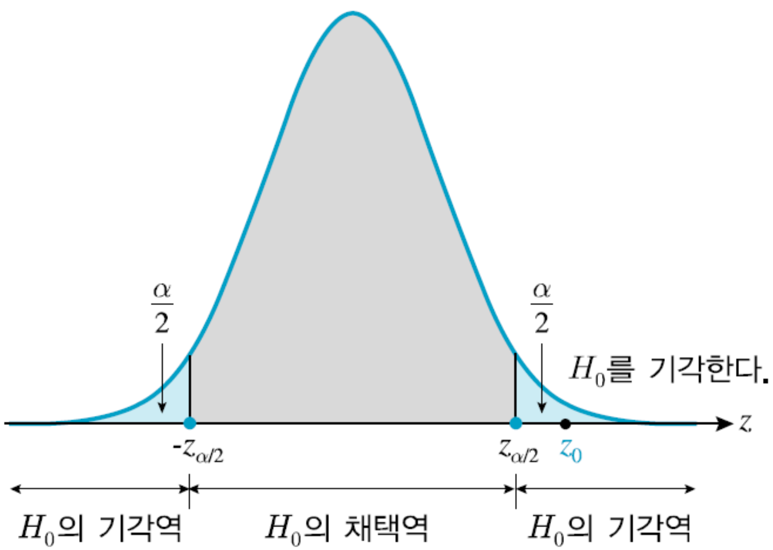
❖ 양쪽 꼬리부분은 귀무가설 H_0을 기각시키는 기각역이고 중심부분은 H_0을 기각시키지 못하는 채택역이다.



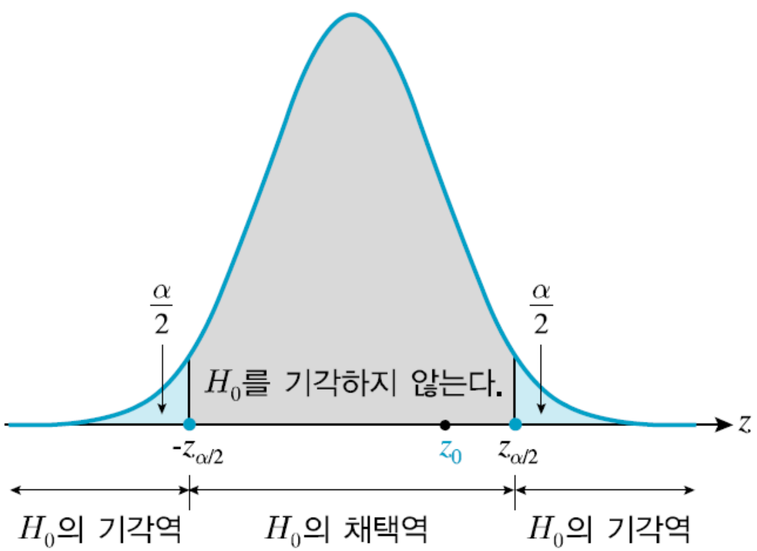
양측검정

❖ 검정통계량의 관찰값이 채택역 안에 놓이면 H_0을 채택하고, 기각역 안에 놓이면 H_0을 기각한다.

검정통계량의 관찰값 : z_0



귀무가설 H_0을 기각한다.



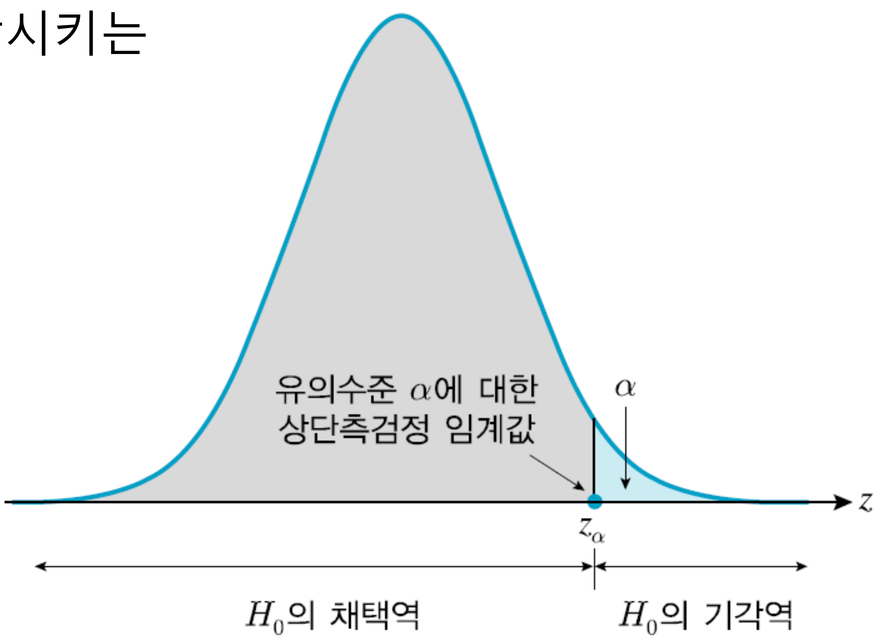
귀무가설 H_0을 기각하지 않는다.

상단측검정

귀무가설 $H_0 : \theta \le \theta_0$에 대하여 대립가설 $H_1 : \theta > \theta_0$으로 구성된 검정

❖ 유의수준을 α라 하면, 위쪽 꼬리확률이 α가 되는 임계값 z_α에 의해 두 영역으로 분리된다.

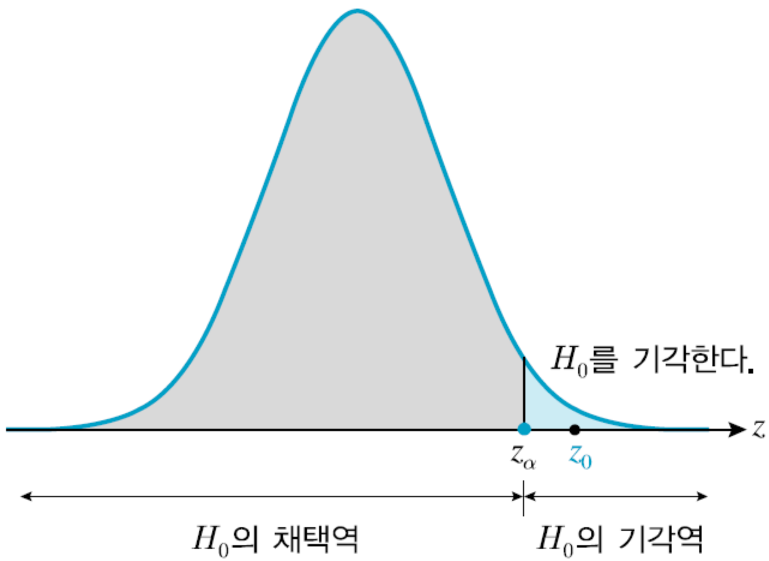
❖ 위쪽 꼬리부분은 귀무가설 H_0을 기각시키는 기각역이고 아래쪽부분은 H_0을 기각 시키지 못하는 채택역이다.



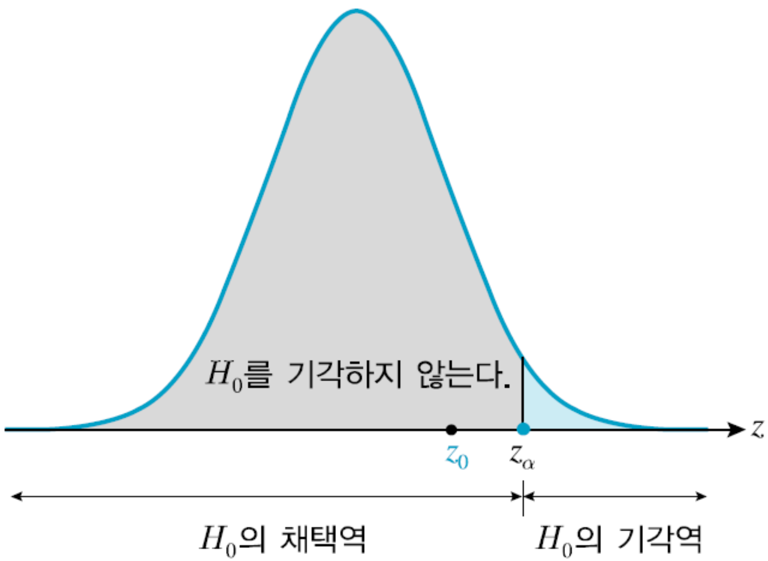
상단측검정

❖ 검정통계량의 관찰값이 채택역 안에 놓이면 H_0을 채택하고, 기각역 안에 놓이면 H_0을 기각한다.

검정통계량의 관찰값 : z_0



귀무가설 H_0을 기각한다.



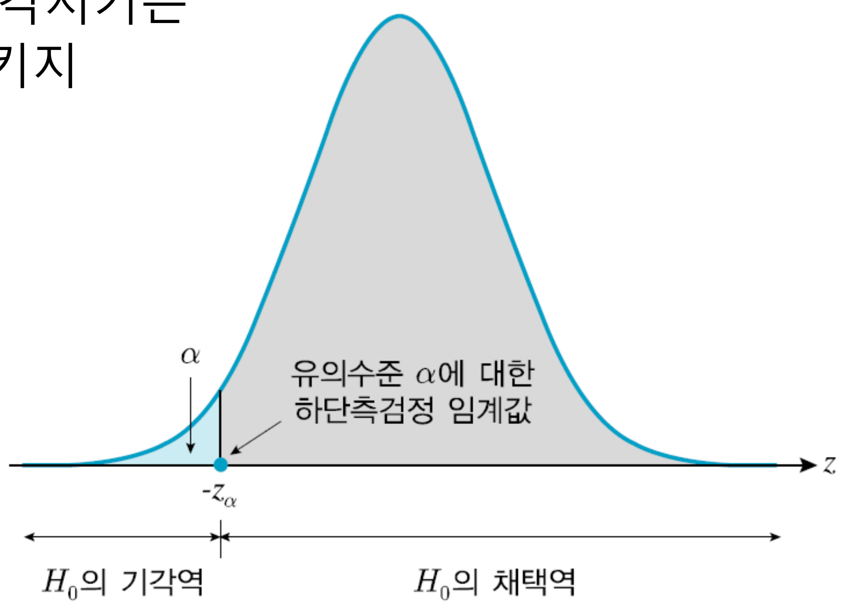
귀무가설 H_0을 기각하지 않는다.

하단측검정

귀무가설 $H_0 : \theta \geq \theta_0$에 대하여 대립가설 $H_1 : \theta < \theta_0$으로 구성된 검정

❖ 유의수준을 α라 하면, 아래쪽 꼬리확률이 α가 되는 임계값 $-z_\alpha$에 의해 두 영역으로 분리된다.

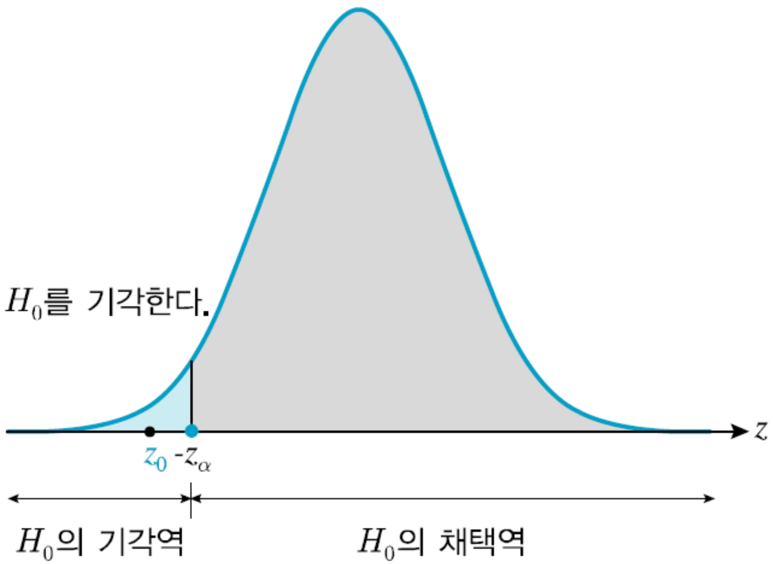
❖ 아래쪽 꼬리부분은 귀무가설 H_0을 기각시키는 기각역이고 위쪽부분은 H_0을 기각시키지 못하는 채택역이다.



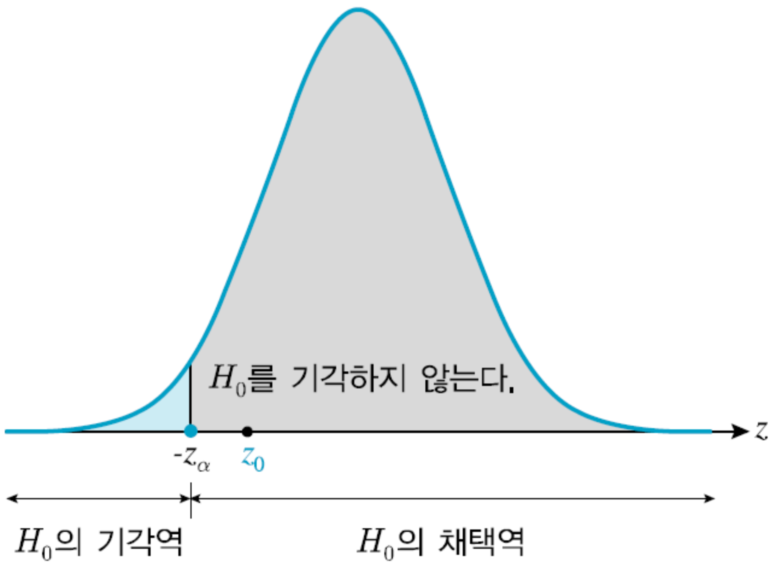
하단측검정

❖ 검정통계량의 관찰값이 채택역 안에 놓이면 H_0을 채택하고, 기각역 안에 놓이면 H_0을 기각한다.

검정통계량의 관찰값 : z_0



귀무가설 H_0을 기각한다.



귀무가설 H_0을 기각하지 않는다.

p - 값에 의한 검정 방법

p - 값p-value : 귀무가설 H_0이 참이라고 가정할 때, 관찰값에 의해 H_0을 기각시킬 가장 작은 유의수준

예 **유의수준 $\alpha = 0.05$와 0.01에서 귀무가설 $H_0 : \mu \leq 10$을 검정하기 위하여 임의로 선정한 표본의 표본평균 $\overline{x}$에 대하여 $z_0 = 2.24$라 하자.**

이때 H_0을 기각시킬 가장 작은 임계값은 2.24이며, 이에 대한 유의수준은 $P(Z \geq 2.24) = 0.0125$이고 따라서 p-값 $= 0.0125$이다.

p - 값에 의한 검정 방법

❖ 유의수준 $\alpha = 0.05$에서 상단측검정의 기각역은 $Z > z_{0.05} = 1.645$이고 관찰값 2.24는 기각역 안에 놓이므로 H_0을 기각시킨다.

❖ 유의수준 $\alpha = 0.01$에서 상단측검정의 기각역은 $Z > z_{0.01} = 2.58$이고 관찰값 2.24는 기각역 안에 놓이지 않으므로 H_0을 기각시키지 않는다.

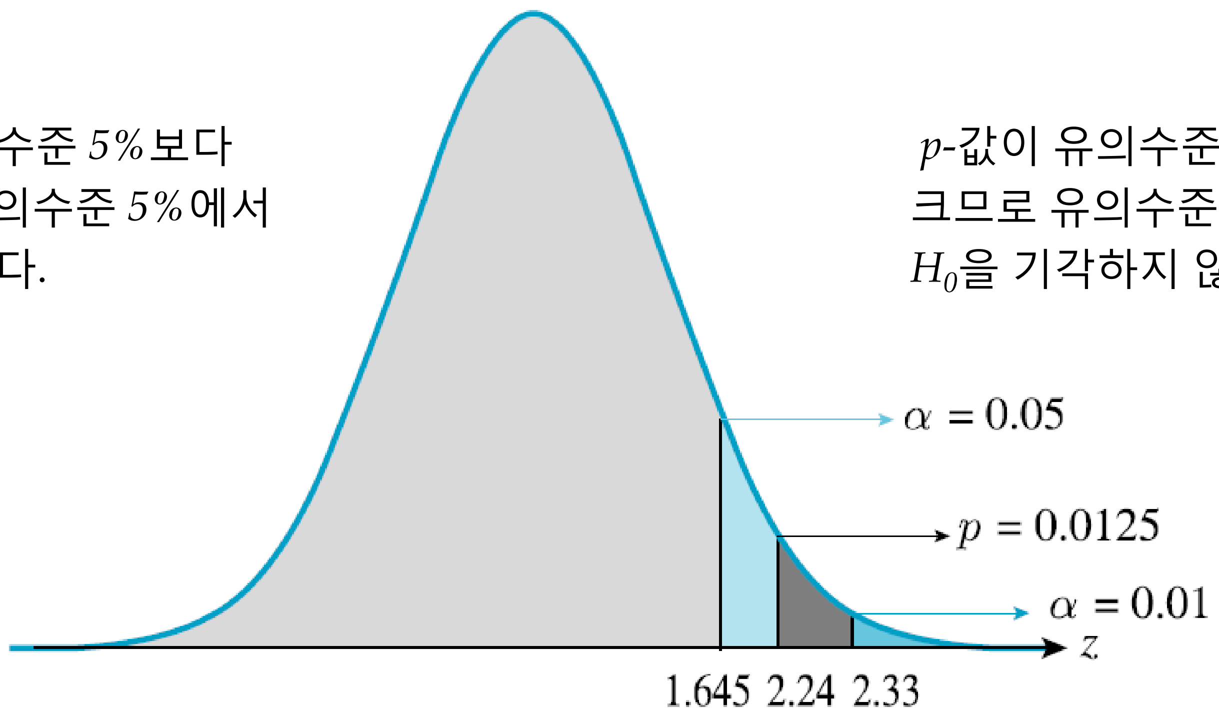
❖ 유의수준과 p-값의 비교 : $0.01 < p$-값 $= 0.0125 < 0.05$

p - 값에 의한 검정 방법

❖ p-값이 유의수준보다 작으면 귀무가설 H_0을 기각하고, p-값이 유의수준보다 크면 귀무가설 H_0을 기각하지 않는다.

p-값이 유의수준 5%보다 작으므로 유의수준 5%에서 H_0을 기각한다.

p-값이 유의수준 1%보다 크므로 유의수준 1%에서 H_0을 기각하지 않는다.



p - 값에 의한 검정 방법

p − 값	유의수준(α)		
	10%	5%	1%
p − 값 > 0.1	H_0를 채택	H_0를 채택	H_0를 채택
$0.05 < p$ − 값 ≤ 0.01	H_0를 기각	H_0를 채택	H_0를 채택
$0.01 < p$ − 값 ≤ 0.05	H_0를 기각	H_0를 기각	H_0를 채택
p − 값 ≤ 0.01	H_0를 기각	H_0를 기각	H_0를 기각

p - 값에 의한 검정 순서

❶ 대립가설 H_1을 설정한다. 이때 등호는 항상 귀무가설에서 사용한다.

❷ 유의수준 α를 정한다.

❸ 적당한 검정통계량을 선택한다.

❹ p-값을 구한다.

❺ p-값 $\leq \alpha$이면 귀무가설 H_0을 기각시키고, p-값 $> \alpha$이면 H_0을 기각시키지 않는다.

신뢰구간과 가설검정의 관계

❖ 유의수준 α 와 신뢰구간 $100(1 - \alpha)\%$는 서로 상반되는 개념이다.

모분산 σ^2이 알려진 정규모집단의 모평균 μ에 대한 $100(1 - \alpha)\%$ 신뢰구간:

$$\bar{x}_0 - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{x}_0 + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

모분산 σ^2이 알려진 정규모집단의 모평균 μ에 대한 $H_0 : \mu = \mu_0$에 대한 검정 :

기각역 : $|z_0| = \left|\frac{\bar{x}_0 - \mu_0}{\sigma / \sqrt{n}}\right| \geq z_{\alpha/2}$

채택역 : $|z_0| = \left|\frac{\bar{x}_0 - \mu_0}{\sigma / \sqrt{n}}\right| < z_{\alpha/2} \quad \Rightarrow \quad \bar{x}_0 - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu_0 < \bar{x}_0 + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$

❖ 귀무가설에서 주장하는 $\mu = \mu_0$이 모평균 μ에 대한 $100(1 - \alpha)\%$ 신뢰구간 안에 놓이면 귀무가설을 채택한다.

예제 10-1

$\sigma^2 = 4$ 인 정규모집단의 모평균을 추정하기 위해 크기 25 인 표본을 추출하여 $\overline{x} = 7$ 을 얻었다.

(a) 모평균에 대한 95% 신뢰구간을 구하라.

(b) 유의수준 5% 에서 $H_0 : \mu = 6.3$, $H_1 : \mu \neq 6.3$ 을 검정하라.

(c) 유의수준 5% 에서 $H_0 : \mu = 6.1$, $H_1 : \mu \neq 6.1$ 을 검정하라.

풀이

(a) $\bar{x} = 7,\ \sigma^2 = 4,\ n = 25$ 이므로 표본평균 $\bar{X}$ 의 표준오차는 $S.E(\bar{X}) = \dfrac{2}{\sqrt{25}} = 0.4$ 이다.

따라서 모평균 μ 에 대한 95% 신뢰구간의 오차한계 : $e = 1.96 \times 0.4 = 0.784$

μ 에 대한 95% 신뢰구간 : $(7 - 0.784,\ 7 + 0.784) = (6.216,\ 7.784)$

신뢰구간과 가설검정의 관계

(b) 유의수준 5%에서 $H_0 : \mu = 6.3$ 이 μ에 대한 95% 신뢰구간 안에 놓이므로 귀무가설 $\mu = 6.3$ 은 채택한다.

(c) 유의수준 5%에서 $H_0 : \mu = 6.1$ 이 μ에 대한 95% 신뢰구간 안에 놓이지 않으므로 귀무가설 $\mu = 6.1$ 은 기각한다.

10.2 모평균의 가설검정

모평균의 가설검정_(모분산 σ^2이 알려진 정규모집단)

❖ 모분산 σ^2이 알려진 정규모집단의 모평균을 μ이라 하면, 크기 n인 표본평균은 다음 분포에 따른다.

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

- 모분산 σ^2이 알려진 정규모집단의 모평균 μ에 대한 귀무가설과 대립가설

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{cases} \qquad \begin{cases} H_0 : \mu \le \mu_0 \\ H_1 : \mu > \mu_0 \end{cases} \qquad \begin{cases} H_0 : \mu \ge \mu_0 \\ H_1 : \mu < \mu_0 \end{cases}$$

- 모평균 μ에 대한 귀무가설을 검정하기 위해 표본평균 $\bar{X}$ 를 이용한다.

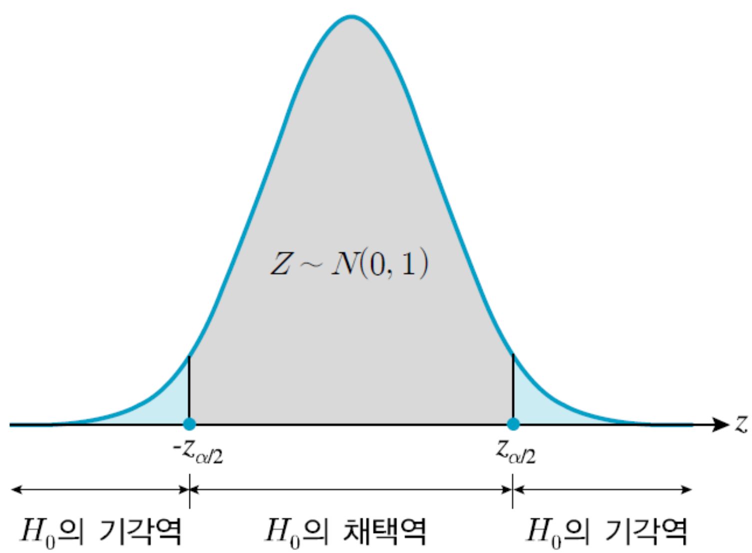
양측검정

귀무가설 $H_0 : \mu = \mu_0$ 에 대한 대립가설 $H_1 : \mu \neq \mu_0$ 으로 구성된 검정

귀무가설 $H_0 : \mu = \mu_0$이 정당한 것으로 가정하고, 이 주장에 대한 타당성을 검정

- 검정통계량과 확률분포 : $Z = \dfrac{\bar{X} - \mu_0}{\sigma / \sqrt{n}} \sim N(0,\ 1)$

- 미리 주어진 유의수준 α에 대한 $H_0 : \mu = \mu_0$의 기각역 : $Z \leq -z_{\alpha/2}$, $Z \geq z_{\alpha/2}$

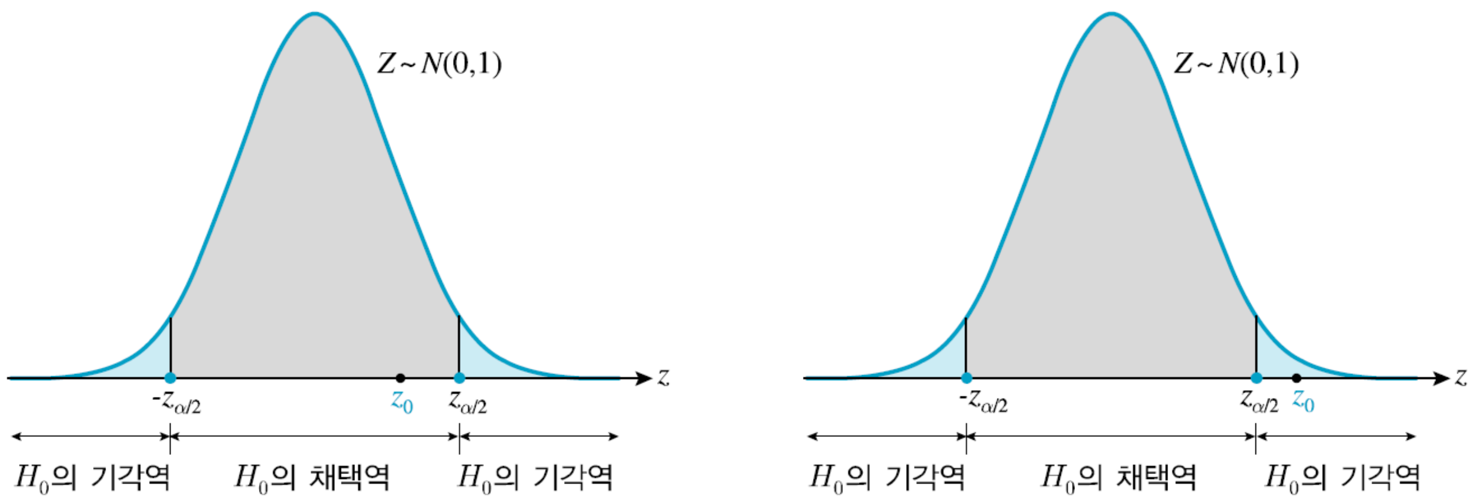


양측검정

- 표본으로부터 얻은 다음 검정통계량의 관찰값 $z_0 = \dfrac{\overline{x} - \mu_0}{\sigma / \sqrt{n}}$

  검정통계량의 관찰값 z_0이 채택역 안에 놓이는지 기각역 안에 놓이는지 판단한다.

- $H_0 : \mu = \mu_0$의 기각역 : $\boxed{z_0 = \dfrac{\overline{x} - \mu_0}{\sigma / \sqrt{n}} \le -z_{\alpha/2}, \quad z_0 = \dfrac{\overline{x} - \mu_0}{\sigma / \sqrt{n}} \ge z_{\alpha/2}}$



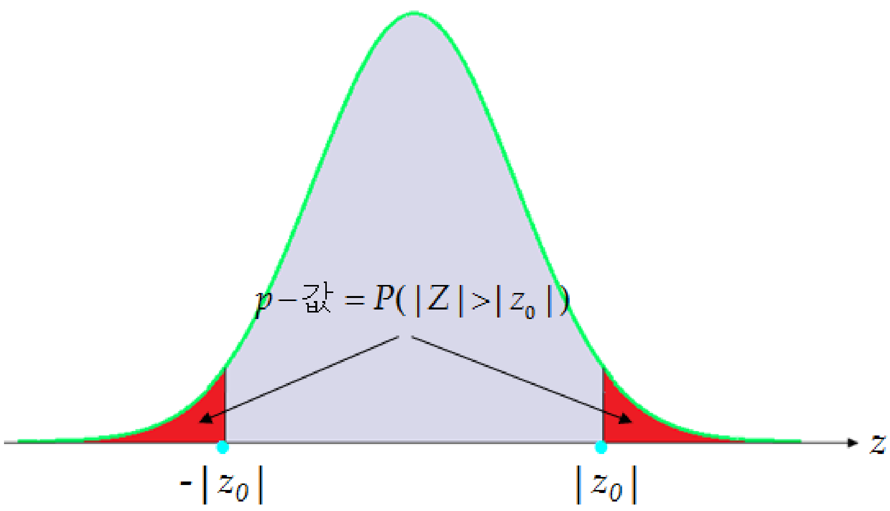
(a) H_0를 채택 (b) H_0를 기각

양측검정

- 검정통계량의 관찰값이 z_0인 양측검정에 대한 p – 값 :

$$p - \text{값} = P(|Z| > |z_0|) = 2[1 - P(Z < |z_0|)]$$

p – 값 $\leq \alpha$ 이면 H_0을 기각하고, p – 값 $> \alpha$ 이면 H_0을 채택한다.



모평균의 가설검정_(모분산 σ^2이 알려진 정규모집단)

예제 10-2

$\sigma^2 = 16$인 정규모집단에 대해 $\mu = 24$라는 주장을 검정하기 위해 크기 50인 표본을 추출하였다.

(a) 양측검정을 위한 귀무가설과 대립가설을 설정하라.

(b) 유의수준 5%에서 기각역을 구하라.

(c) $\overline{x} = 25$일 때, 유의수준 5%에서 귀무가설을 검정하라.

(d) $p-$값을 구하라.

(e) $p-$값을 이용하여 유의수준 5%에서 귀무가설을 검정하라.

(f) $p-$값을 이용하여 유의수준 10%에서 귀무가설을 검정하라.

풀이

(a) 귀무가설은 $H_0 : \mu = 24$이고 이에 대한 대립가설은 $H_1 : \mu \neq 24$이다.

(b) 유의수준 $\alpha = 0.05$에 대한 양측검정의 기각역은 $R : |Z| > z_{0.025} = 1.96$이다.

모평균의 가설검정_(모분산 σ^2이 알려진 정규모집단)

(c) 모표준편차가 $\sigma = 4$이고 $\bar{x} = 25$ 이므로 검정통계량과 관찰값은 각각 다음과 같다.

$$Z = \frac{\bar{X} - 24}{4/\sqrt{50}}, \quad z_0 = \frac{25-24}{4/\sqrt{50}} = 1.77$$

(d) 검정통계량의 관측값에 대하여 $|z_0| = 1.77$이므로 p – 값은 다음과 같다.

$$p - \text{값} = 2\left[1 - P(Z < 1.77)\right] = 2(1 - 0.9616) = 0.0768$$

(e) p – 값 $= 0.0768 > \alpha = 0.05$이므로 귀무가설 $H_0 : \mu = 24$를 유의수준 5%에서 기각할 수 없다.

(f) p – 값 $= 0.0768 < \alpha = 0.1$이므로 귀무가설 $H_0 : \mu = 24$를 유의수준 10%에서 기각한다.

기각역을 이용한 검정 방법

유의수준 5% 에서 기각역은 $R: |Z| > z_{0.025} = 1.96$이고 관찰값이 $z_0 = 1.77$이므로 관찰값이 기각역 안에 놓이지 않는다. 따라서 귀무가설을 기각할 수 없다.

유의수준 10% 에서 기각역은 $R: |Z| > z_{0.05} = 1.645$이고 관찰값이 $z_0 = 1.77$이므로 관찰값이 기각역 안에 놓인다. 따라서 귀무가설을 기각한다.

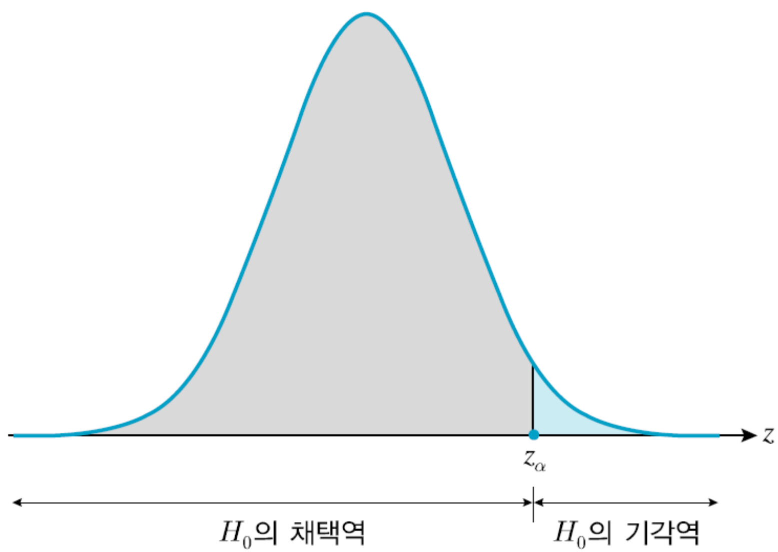
상단측검정

귀무가설 $H_0 : \mu \le \mu_0$에 대한 대립가설 $H_1 : \mu > \mu_0$으로 구성된 검정

귀무가설 $H_0 : \mu = \mu_0$이 정당한 것으로 가정하고, $\mu \le \mu_0$에 대한 타당성을 검정

- 검정통계량과 확률분포 : $Z = \dfrac{\bar{X} - \mu_0}{\sigma / \sqrt{n}} \sim N(0,\ 1)$

- 미리 주어진 유의수준 α에 대한 $H_0 : \mu \le \mu_0$의 기각역 : $Z \ge z_\alpha$

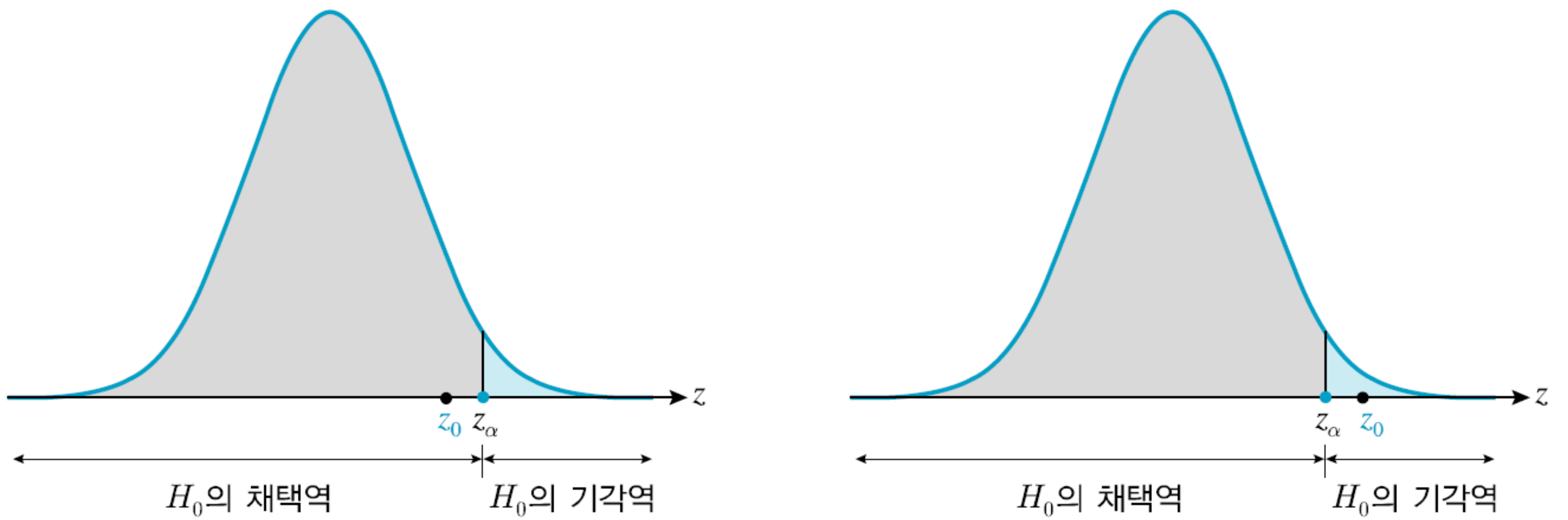


상단측검정

- 표본으로부터 얻은 다음 검정통계량의 관찰값 $z_0 = \dfrac{\overline{x} - \mu_0}{\sigma / \sqrt{n}}$

검정통계량의 관찰값 z_0이 채택역 안에 놓이는지 기각역 안에 놓이는지 판단한다.

- $H_0 : \mu \leq \mu_0$의 기각역 : $\boxed{z_0 = \dfrac{\overline{x} - \mu_0}{\sigma / \sqrt{n}} \geq z_\alpha}$

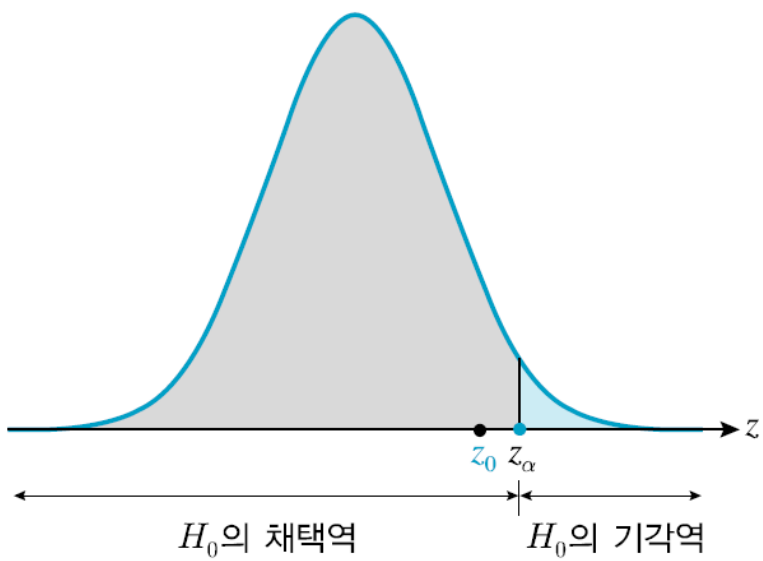


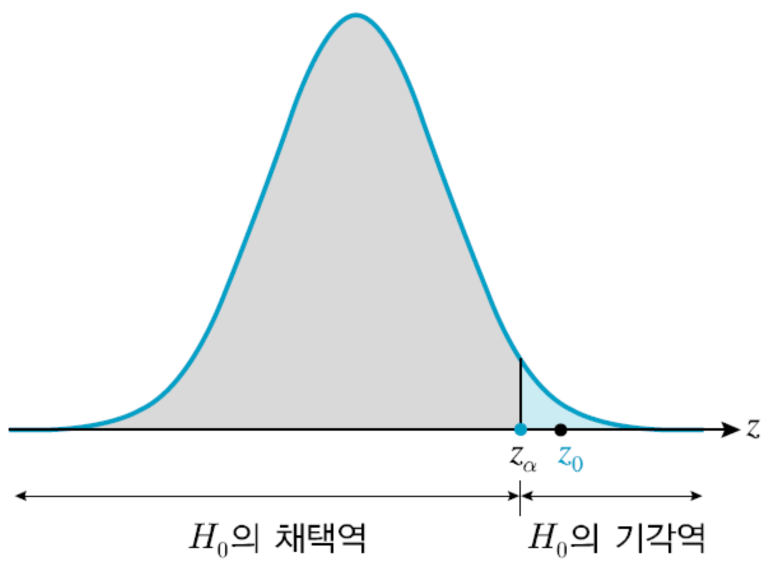
상단측검정

- 검정통계량의 관찰값이 z_0인 상단측검정에 대한 p – 값 :

$$p - \text{값} = P(Z > z_0)$$

p – 값 $\le \alpha$ 이면 H_0을 기각하고, p – 값 $> \alpha$ 이면 H_0을 채택한다.





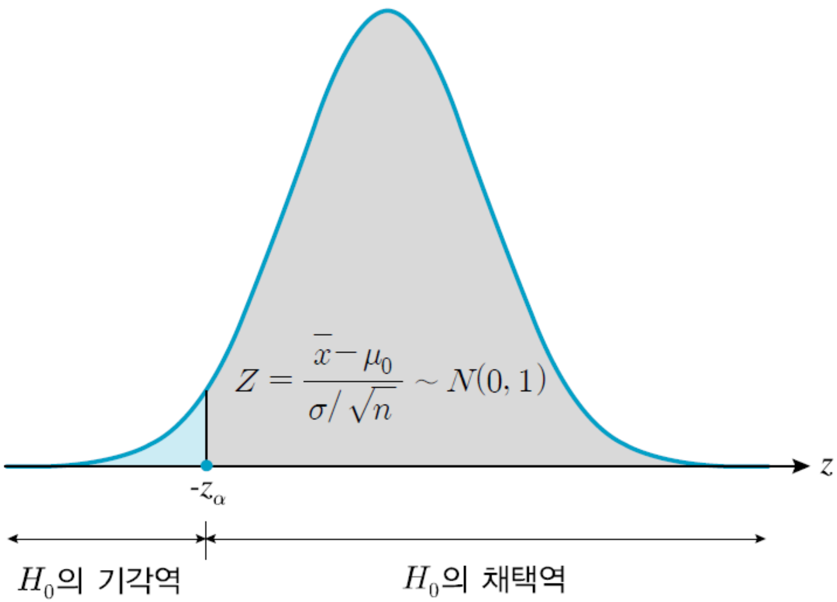
하단측검정

귀무가설 $H_0 : \mu \geq \mu_0$ 에 대한 대립가설 $H_1 : \mu < \mu_0$ 으로 구성된 검정

귀무가설 $H_0 : \mu = \mu_0$이 정당한 것으로 가정하고, $\mu \geq \mu_0$에 대한 타당성을 검정

- 검정통계량과 확률분포 : $Z = \dfrac{\bar{X} - \mu_0}{\sigma / \sqrt{n}} \sim N(0,\ 1)$

- 미리 주어진 유의수준 α에 대한
  $H_0 : \mu \geq \mu_0$의 기각역 : $Z \leq -z_\alpha$



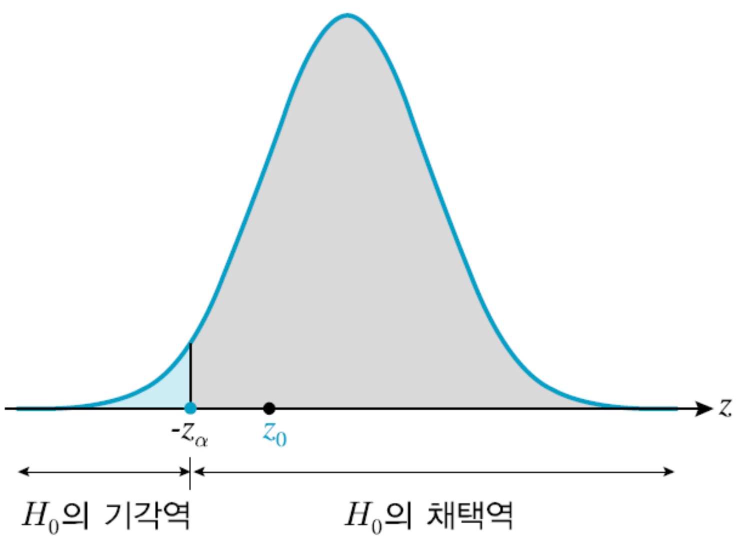
하단측검정

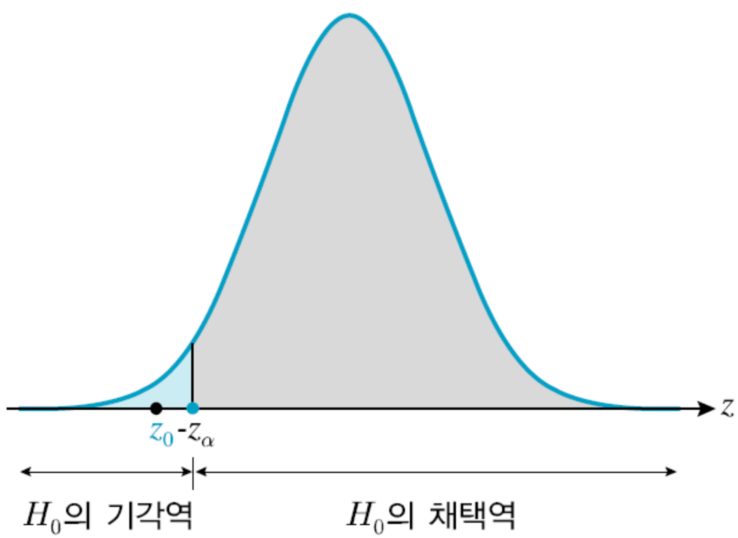
- 표본으로부터 얻은 다음 검정통계량의 관찰값 $z_0 = \dfrac{\overline{x} - \mu_0}{\sigma / \sqrt{n}}$

  검정통계량의 관찰값 z_0이 채택역 안에 놓이는지 기각역 안에 놓이는지 판단한다.

- $H_0 : \mu \geq \mu_0$의 기각역 :

$$\boxed{z_0 = \frac{\overline{x} - \mu_0}{\sigma / \sqrt{n}} \leq -z_\alpha}$$



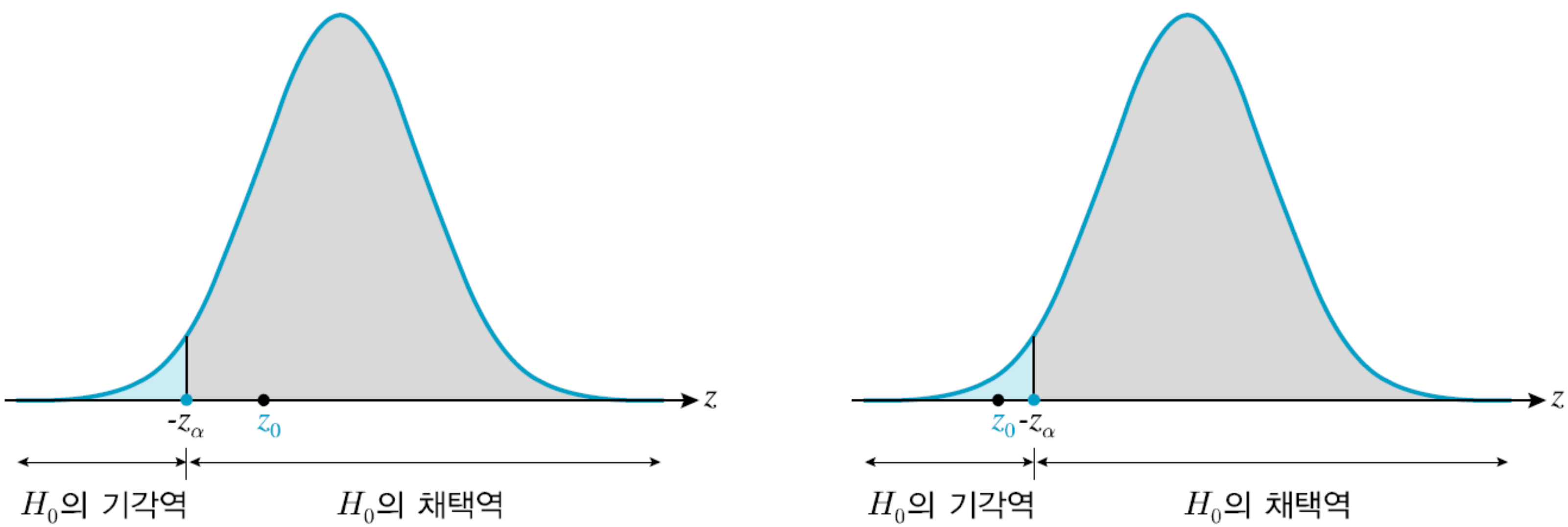


하단측검정

- 검정통계량의 관찰값이 z_0인 하단측검정에 대한 p – 값 :

$$p - \text{값} = P(Z < z_0)$$

p – 값 $\leq \alpha$ 이면 H_0을 기각하고, p – 값 $> \alpha$ 이면 H_0을 채택한다.



하단측검정

예제 10-3

$\sigma = 1.8$ 인 정규모집단에 대해 $H_0 : \mu = 13.5$, $H_1 : \mu > 13.5$ 를 검정하기 위해 크기 35 인 표본을 추출하여 $\bar{x} = 14$ 를 얻었다.

(a) 유의수준 5% 에서 귀무가설을 검정하라.
(b) p - 값을 이용하여 유의수준 5% 에서 귀무가설을 검정하라. 단, $P(Z > 1.643) = 0.0502$ 이다.
(c) p - 값을 이용하여 유의수준 10% 에서 귀무가설을 검정하라.

풀이

(a) 다음 순서에 따라 가설을 검정한다.

① 기각역을 구한다.
유의수준 $\alpha = 0.05$ 에 대한 상단측검정의 기각역은 $R : Z \geq z_{0.05} = 1.645$ 이다.

② 검정통계량을 선정한다.
모표준편차가 $\sigma = 1.8$ 이므로 검정통계량은 $Z = \dfrac{\bar{X} - 13.5}{1.8/\sqrt{35}}$ 이다.

하단측검정

③ 통계량의 관찰값을 구한다.

표본평균이 $\bar{x} = 14$ 이므로 관찰값은 $z_0 = \dfrac{14 - 13.5}{1.8/\sqrt{35}} = 1.643$ 이다.

④ 귀무가설 H_0의 기각을 결정한다.

검정통계량의 관찰값 $z_0 = 1.643$이 기각역 안에 놓이지 않으므로 $H_0 : \mu = 13.5$를 기각할 수 없다.

(b) 검정통계량의 관측값이 $z_0 = 1.643$이므로 p - 값은 다음과 같다.

$$p - \text{값} = P(Z > 1.643) = 0.0502$$

따라서 p - 값 $= 0.0502 > \alpha = 0.05$이므로 귀무가설 $H_0 : \mu = 13.5$를 유의수준 5% 에서 기각할 수 없다.

(c) p - 값 $= 0.0502 < \alpha = 0.1$이므로 귀무가설 $H_0 : \mu = 13.5$를 유의수준 10% 에서 기각한다.

모분산이 알려진 경우, 모평균에 대한 검정 유형과 기각역

기각역 및 p-값 / 검정 방법	귀무가설 H_0	대립가설 H_1	H_0 의 기각역	p-값
하단측검정	$\mu \geq \mu_0$ ($\mu = \mu_0$)	$\mu < \mu_0$	$Z \leq -z_\alpha$	$P(Z < z_0)$
상단측검정	$\mu \leq \mu_0$ ($\mu = \mu_0$)	$\mu > \mu_0$	$Z \geq z_\alpha$	$P(Z > z_0)$
양측검정	$\mu = \mu_0$	$\mu \neq \mu_0$	$\lvert Z \rvert \geq z_{\alpha/2}$	$2[1 - P(Z < \lvert z_0 \rvert)]$

모평균 차의 가설검정_두 모분산 σ_1^2 , σ_2^2 이 알려진 두 정규모집단

❖ 모분산 σ_1^2, σ_2^2이 알려진 두 정규모집단의 모평균을 μ_1, μ_2라 하면, 각각 크기 n, m인 표본평균 $\bar{X}$, $\bar{Y}$에 대해 다음 분포를 얻는다.

$$Z = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n} + \dfrac{\sigma_2^2}{m}}} \sim N(0,1)$$

- 모분산 σ_1^2, σ_2^2 이 알려진 두 정규모집단의 모평균 차 μ_1 - μ_2 = d_0에 대한 귀무가설과 대립가설

$$\begin{cases} H_0 : \mu_1 - \mu_2 \ge d_0 \\ H_1 : \mu_1 - \mu_2 < d_0 \end{cases} \qquad \begin{cases} H_0 : \mu_1 - \mu_2 = d_0 \\ H_1 : \mu_1 - \mu_2 \ne d_0 \end{cases} \qquad \begin{cases} H_0 : \mu_1 - \mu_2 \le d_0 \\ H_1 : \mu_1 - \mu_2 > d_0 \end{cases}$$

- 모평균 차 μ_1 - μ_2에 대한 귀무가설을 검정하기 위해 표본평균의 차 $\bar{X} - \bar{Y}$ 를 이용한다.

모평균 차에 대한 양측검정

귀무가설 $H_0 : \mu_1 - \mu_2 = d_0$에 대한 대립가설 $H_1 : \mu_1 - \mu_2 \neq d_0$으로 구성된 검정

검정통계량과 확률분포 : $Z = \dfrac{(\bar{X} - \bar{Y}) - d_0}{\sqrt{\dfrac{\sigma_1^2}{n} + \dfrac{\sigma_2^2}{m}}} \sim N(0, 1)$

검정통계량의 관찰값 : $z_0 = \dfrac{(\bar{x} - \bar{y}) - d_0}{\sqrt{\dfrac{\sigma_1^2}{n} + \dfrac{\sigma_2^2}{m}}}$

미리 주어진 유의수준 α에 대한 $H_0 : \mu_1 - \mu_2 = d_0$의 기각역 : $Z \leq -z_{\alpha/2}$, $Z \geq z_{\alpha/2}$

$$\boxed{z_0 = \frac{(\bar{x} - \bar{y}) - d_0}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}} \leq -z_{\alpha/2}, \quad z_0 = \frac{(\bar{x} - \bar{y}) - d_0}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}} \geq z_{\alpha/2}}$$

$p-$값 $= P(|Z| > |z_0|) = 2[1 - P(Z < |z_0|)]$, $p-$값 $\leq \alpha$이면 H_0을 기각한다.

모평균 차에 대한 상단측검정

귀무가설 $H_0 : \mu_1 - \mu_2 \le d_0$에 대한 대립가설 $H_1 : \mu_1 - \mu_2 > d_0$으로 구성된 검정

검정통계량과 확률분포 : $Z = \dfrac{(\bar{X} - \bar{Y}) - d_0}{\sqrt{\dfrac{\sigma_1^2}{n} + \dfrac{\sigma_2^2}{m}}} \sim N(0, 1)$

검정통계량의 관찰값 : $z_0 = \dfrac{(\bar{x} - \bar{y}) - d_0}{\sqrt{\dfrac{\sigma_1^2}{n} + \dfrac{\sigma_2^2}{m}}}$

미리 주어진 유의수준 α에 대한 $H_0 : \mu_1 - \mu_2 \le d_0$의 기각역 : $Z \ge z_\alpha$

$$\boxed{z_0 = \dfrac{(\bar{x} - \bar{y}) - d_0}{\sqrt{\dfrac{\sigma_1^2}{n} + \dfrac{\sigma_2^2}{m}}} \ge z_\alpha}$$

p – 값 = $P(Z > z_0)$, p – 값 $\le \alpha$이면 H_0을 기각한다.

모평균 차에 대한 상단측검정

귀무가설 $H_0 : \mu_1 - \mu_2 \geq d_0$에 대한 대립가설 $H_1 : \mu_1 - \mu_2 < d_0$으로 구성된 검정

검정통계량과 확률분포 : $Z = \dfrac{(\bar{X} - \bar{Y}) - d_0}{\sqrt{\dfrac{\sigma_1^2}{n} + \dfrac{\sigma_2^2}{m}}} \sim N(0, 1)$

검정통계량의 관찰값 : $z_0 = \dfrac{(\bar{x} - \bar{y}) - d_0}{\sqrt{\dfrac{\sigma_1^2}{n} + \dfrac{\sigma_2^2}{m}}}$

미리 주어진 유의수준 α에 대한 $H_0 : \mu_1 - \mu_2 \geq d_0$의 기각역 : $Z \leq - z_\alpha$

$$\boxed{z_0 = \dfrac{(\bar{x} - \bar{y}) - d_0}{\sqrt{\dfrac{\sigma_1^2}{n} + \dfrac{\sigma_2^2}{m}}} \leq -z_\alpha}$$

p – 값 = $P(Z < z_0)$, p – 값 $\leq \alpha$이면 H_0을 기각한다.

모분산이 알려진 경우, 모평균 차에 대한 검정 유형과 기각역

기각역 및 p-값 / 검정 방법	귀무가설 H_0	대립가설 H_1	H_0 의 기각역	p-값
하단측검정	$\mu_1 - \mu_2 \geq d_0$ $(\mu_1 - \mu_2 = d_0)$	$\mu_1 - \mu_2 < d_0$	$Z \leq -z_\alpha$	$P(Z < z_0)$
상단측검정	$\mu_1 - \mu_2 \leq d_0$ $(\mu_1 - \mu_2 = d_0)$	$\mu_1 - \mu_2 > d_0$	$Z \geq z_\alpha$	$P(Z > z_0)$
양측검정	$\mu_1 - \mu_2 = d_0$	$\mu_1 - \mu_2 \neq d_0$	$\lvert Z \rvert \geq z_{\alpha/2}$	$2[1 - P(Z < \lvert z_0 \rvert)]$

모분산이 알려진 경우, 모평균 차에 대한 검정 유형과 기각역

예제 10-4

순수한 동선에 비하여 합금을 사용한 전선의 저항이 평균 0.5만큼 더 많다고 한다. 이를 알아보기 위해 각각 64개씩 조사한 결과 $\overline{x} = 1.45$, $\overline{y} = 1.99$를 얻었다. 두 종류의 전선은 각각 $\sigma_1 = 0.128$, $\sigma_2 = 0.138$인 정규분포를 따르는 것으로 알려져 있으며, 단위는 Ω이다.

(a) 합금을 사용한 전선의 저항이 0.5만큼 더 많은지 유의수준 5%에서 검정하라.

(b) $p-$값을 구하고 유의수준 5%에서 검정하라.

풀이

(a) 순수한 동선과 합금을 사용한 전선의 평균 저항을 각각 μ_1, μ_2라 하고 다음 순서에 따라 가설을 검정한다.

① 귀무가설과 대립가설을 설정한다.
검정하고자 하는 가설 $\mu_2 - \mu_1 > 0.5$는 등호가 없으므로 대립가설로 설정한다.즉,
귀무가설은 $H_0 : \mu_2 - \mu_1 \leq 0.5$이고 대립가설은 $H_1 : \mu_2 - \mu_1 > 0.5$이다.

② 유의수준 $\alpha = 0.05$에 대한 상단측검정의 기각역은 $R : Z \geq z_{0.05} = 1.645$이다.

모분산이 알려진 경우, 모평균 차에 대한 검정 유형과 기각역

③ 검정통계량을 선정한다. $Z = \dfrac{(\bar{Y}-\bar{X})-0.5}{\sqrt{\dfrac{0.128^2}{64}+\dfrac{0.138^2}{64}}} = \dfrac{(\bar{Y}-\bar{X})-0.5}{0.0235}$

④ 검정통계량의 관찰값을 구한다.

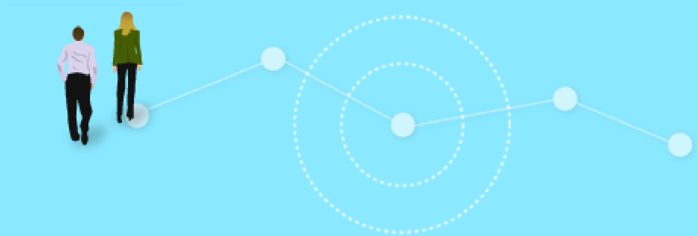
$\bar{x}=1.45,\ \bar{y}=1.99$ 이므로 검정통계량의 관찰값은 $z_0 = \dfrac{0.24-0.5}{0.0235} = 1.70$ 이다.

⑤ 귀무가설 H_0의 기각을 결정한다.

검정통계량의 관찰값 $z_0 = 1.70$이 기각역 안에 놓이므로 $H_0 : \mu_1 - \mu_2 \le 0.5$를 기각한다. 즉, 합금을 사용한 전선의 평균 저항이 순수한 동선의 저항보다 0.5 Ω보다 크다고 할 수 있다.

(2) 검정통계량의 관찰값이 $z_0 = 1.70$이므로 p – 값 = $P(Z > 1.70) = 0.0466 < 0.05$이고, 따라서 귀무가설 $H_0 : \mu_1 - \mu_2 \le 0.5$를 기각한다.

10.3 모비율의 가설검정



모비율의 가설검정

❖ 모비율 p인 모집단에서 크기 n인 표본을 선정할 때, n이 충분히 크면 표본비율은 다음 분포에 따른다.

$$Z = \frac{\hat{P} - p}{\sqrt{pq / n}} \approx N(0, 1)$$

- 모비율 p에 대한 귀무가설과 대립가설

$$\begin{cases} H_0 : p \ge p_0 \\ H_1 : p < p_0 \end{cases} \qquad \begin{cases} H_0 : p = p_0 \\ H_1 : p \ne p_0 \end{cases} \qquad \begin{cases} H_0 : p \le p_0 \\ H_1 : p > p_0 \end{cases}$$

- 이때 진위여부를 명확히 밝히기 전까지 귀무가설에 대한 주장을 정당한 것으로 간주하므로 모비율은 $p = p_0$으로 생각한다. 따라서 표본비율에 대해 다음 분포를 얻는다.

$$Z = \frac{\hat{P} - p_0}{\sqrt{p_0 q_0 / n}} \approx N(0, 1)$$

모비율에 대한 양측검정

귀무가설 $H_0 : p = p_0$에 대한 대립가설 $H_1 : p \neq p_0$으로 구성된 검정

검정통계량과 확률분포 : $Z = \dfrac{\hat{P} - p_0}{\sqrt{p_0 q_0 / n}} \sim N(0,1)$

검정통계량의 관찰값 : $z_0 = \dfrac{\hat{p} - p_0}{\sqrt{p_0 q_0 / n}}$

미리 주어진 유의수준 α에 대한 $H_0 : p = p_0$ 의 기각역 : $Z \leq -z_{\alpha/2}$, $Z \geq z_{\alpha/2}$

$$\boxed{z_0 = \frac{\hat{p} - p_0}{\sqrt{p_0 q_0 / n}} \leq -z_{\alpha/2}, \quad z_0 = \frac{\hat{p} - p_0}{\sqrt{p_0 q_0 / n}} \geq z_{\alpha/2}}$$

p – 값 = $P(|Z| > |z_0|) = 2[1 - P(Z < |z_0|)]$, p – 값 $\leq \alpha$이면 H_0을 기각한다.

모비율에 대한 상단측검정

귀무가설 $H_0: p \le p_0$에 대한 대립가설 $H_1: p > p_0$으로 구성된 검정

검정통계량과 확률분포 : $Z = \dfrac{\hat{P} - p_0}{\sqrt{p_0 q_0 / n}} \sim N(0,1)$

검정통계량의 관찰값 : $z_0 = \dfrac{\hat{p} - p_0}{\sqrt{p_0 q_0 / n}}$

미리 주어진 유의수준 α에 대한 $H_0: p \le p_0$의 기각역 : $Z \ge z_\alpha$

$$\boxed{z_0 = \dfrac{\hat{p} - p_0}{\sqrt{p_0 q_0 / n}} \ge z_\alpha}$$

p – 값 = $P(Z > z_0)$, p – 값 $\le \alpha$이면 H_0을 기각한다.

모비율에 대한 하단측검정

귀무가설 $H_0 : p \geq p_0$에 대한 대립가설 $H_1 : p < p_0$으로 구성된 검정

검정통계량과 확률분포 : $Z = \dfrac{\hat{P} - p_0}{\sqrt{p_0 q_0 / n}} \sim N(0, 1)$

검정통계량의 관찰값 : $z_0 = \dfrac{\hat{p} - p_0}{\sqrt{p_0 q_0 / n}}$

미리 주어진 유의수준 α에 대한 $H_0 : p \geq p_0$의 기각역 : $Z \leq - z_\alpha$

$$\boxed{z_0 = \frac{\hat{p} - p_0}{\sqrt{p_0 q_0 / n}} \leq -z_\alpha}$$

p – 값 = $P(Z < z_0)$, p – 값 $\leq \alpha$이면 H_0을 기각한다.

모비율에 대한 검정 유형과 기각역

기각역 및 $p-$값 / 검정 방법	귀무가설 H_0	대립가설 H_1	H_0 의 기각역	$p-$값
하단측검정	$p \geq p_0$ $(p = p_0)$	$p < p_0$	$Z \leq -z_\alpha$	$P(Z < z_0)$
상단측검정	$p \leq p_0$ $(p = p_0)$	$p > p_0$	$Z \geq z_\alpha$	$P(Z > z_0)$
양측검정	$p = p_0$	$p \neq p_0$	$\lvert Z \rvert \geq z_{\alpha/2}$	$2[1 - P(Z < z_0)]$

모비율에 대한 검정 유형과 기각역

예제 10-5

어느 제조회사에서 생산한 배터리의 불량률이 0.5% 이하라고 한다. 이를 알아보기 위해 생산한 배터리 1550개를 임의로 선정하여 조사한 결과 13개가 불량품이었다.

(a) 배터리의 불량률이 0.5% 이하라는 주장을 유의수준 5%에서 검정하라.

(b) $p-$값을 구하고 유의수준 5%에서 검정하라.

풀이

(a) 다음 순서에 따라 가설을 검정한다.

① 귀무가설과 대립가설을 설정한다.
귀무가설 $H_0 : p \le 0.005$에 대한 대립가설 $H_1 : p > 0.005$이다.

② 유의수준 $\alpha = 0.05$에 대한 상단측검정의 기각역은 $R : Z \ge z_{0.05} = 1.645$이다.

③ 검정통계량을 선정한다. $Z = \dfrac{\hat{P} - 0.005}{\sqrt{(0.005)(0.995)/1550}} = \dfrac{\hat{P} - 0.005}{0.00179}$

모비율에 대한 검정 유형과 기각역

④ 검정통계량의 관찰값을 구한다.

1550개의 배터리 중에서 13개가 불량품이므로 표본비율은 다음과 같다.

따라서 검정통계량의 관찰값은 다음과 같다.

$$z_0 = \frac{0.00839 - 0.005}{0.00179} = 1.89$$

⑤ 귀무가설 H_0의 기각을 결정한다.

검정통계량의 관찰값 $z_0 = 1.89$가 기각역 안에 놓이므로 $H_0 : p \leq 0.005$를 기각한다.

즉, 이 회사에서 생산한 배터리의 불량률이 0.5% 이하라는 주장은 타당성이 부족하다.

(2) 검정통계량의 관찰값이 $z_0 = 1.89$이므로 p – 값 $= P(Z > 1.89) = 1 - 0.9706 = 0.0294 < 0.05$이고, 따라서 귀무가설 $H_0 : p \leq 0.005$를 기각한다.

모비율 차의 가설검정

❖ 모비율 p_1, p_2인 두 모집단에서 각각 크기 n, m인 표본을 선정하여 표본비율을 각각 $\hat{P}_1, \hat{P}_2$ 라 하면, n과 m이 충분히 크면 표본비율은 다음 분포에 따른다.

$$Z = \frac{(\hat{P}_1 - \hat{P}_2) - (p_1 - p_2)}{\sqrt{\frac{p_1 q_1}{n} + \frac{p_2 q_2}{m}}} \sim N(0,1)$$

모비율의 차 $p_1 - p_2$에 대한 귀무가설과 대립가설

$$\begin{cases} H_0 : p_1 - p_2 \geq p_0 \\ H_1 : p_1 - p_2 < p_0 \end{cases} \qquad \begin{cases} H_0 : p_1 - p_2 = p_0 \\ H_1 : p_1 - p_2 \neq p_0 \end{cases} \qquad \begin{cases} H_0 : p_1 - p_2 \leq p_0 \\ H_1 : p_1 - p_2 > p_0 \end{cases}$$

이때 n과 m이 충분히 크면 $\hat{p}_1 \approx p_1, \hat{p}_2 \approx p_2$이므로 표본비율의 차에 대해 다음 분포를 얻는다.

$$Z = \frac{(\hat{P}_1 - \hat{P}_2) - p_0}{\sqrt{\frac{\hat{p}_1 \hat{q}_1}{n} + \frac{\hat{p}_2 \hat{q}_2}{m}}} \sim N(0,1)$$

모비율 차의 가설검정

❖ 모비율의 차 $p_1 - p_2 = 0$에 대한 귀무가설과 대립가설

$$\begin{cases} H_0 : p_1 - p_2 \geq 0 \\ H_1 : p_1 - p_2 < 0 \end{cases} \qquad \begin{cases} H_0 : p_1 - p_2 = 0 \\ H_1 : p_1 - p_2 \neq 0 \end{cases} \qquad \begin{cases} H_0 : p_1 - p_2 \leq 0 \\ H_1 : p_1 - p_2 > 0 \end{cases}$$

그러면 두 모비율이 동일하다, 즉 $p_1 = p_2 = p$에 대한 검정이므로 두 모집단으로부터 크기 $n + m$인 단일 표본을 선정하여 성공의 횟수가 각각 x와 y인 경우로 생각할 수 있다. 이때 표본의 성공률을 **합동표본비율**pooled sample proportion이라 하며, 다음과 같다.

$$\hat{p} = \frac{x + y}{n + m}$$

따라서 모비율의 차 $p_1 - p_2 = 0$에 대한 검정통계량은 다음과 같다.

$$Z = \frac{\hat{P}_1 - \hat{P}_2}{\sqrt{\hat{p}\hat{q}\left(\frac{1}{n} + \frac{1}{m}\right)}} \sim N(0, 1)$$

모비율 차에 대한 양측검정

귀무가설 $H_0 : p_1 - p_2 = 0$에 대한 대립가설 $H_1 : p_1 - p_2 \neq 0$으로 구성된 검정

검정통계량과 확률분포 : $Z = \dfrac{\hat{P}_1 - \hat{P}_2}{\sqrt{\hat{p}\hat{q}\left(\dfrac{1}{n} + \dfrac{1}{m}\right)}} \sim N(0, 1)$

검정통계량의 관찰값 : $z_0 = \dfrac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}\hat{q}\left(\dfrac{1}{n} + \dfrac{1}{m}\right)}}$

미리 주어진 유의수준 α에 대한 $H_0 : p = p_0$의 기각역 : $Z \leq -z_{\alpha/2}$, $Z \geq z_{\alpha/2}$

$$z_0 = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}\hat{q}\left(\frac{1}{n} + \frac{1}{m}\right)}} \leq -z_{\alpha/2}, \quad z_0 = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}\hat{q}\left(\frac{1}{n} + \frac{1}{m}\right)}} \geq z_{\alpha/2}$$

p – 값 $= P(|Z| > |z_0|) = 2[1 - P(Z < |z_0|)]$, p – 값 $\leq \alpha$이면 H_0을 기각한다.

모비율 차에 대한 상단측검정

귀무가설 $H_0 : p_1 - p_2 \le 0$에 대한 대립가설 $H_1 : p_1 - p_2 > 0$으로 구성된 검정

검정통계량과 확률분포 : $Z = \dfrac{\hat{P}_1 - \hat{P}_2}{\sqrt{\hat{p}\hat{q}\left(\dfrac{1}{n} + \dfrac{1}{m}\right)}} \sim N(0,1)$

검정통계량의 관찰값 : $z_0 = \dfrac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}\hat{q}\left(\dfrac{1}{n} + \dfrac{1}{m}\right)}}$

미리 주어진 유의수준 α에 대한 $H_0 : p \le p_0$의 기각역 : $Z \ge z_\alpha$

$$\boxed{z_0 = \dfrac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}\hat{q}\left(\dfrac{1}{n} + \dfrac{1}{m}\right)}} \ge z_\alpha}$$

p – 값 = $P(Z > z_0)$, p – 값 $\le \alpha$이면 H_0을 기각한다.

모비율 차에 대한 하단측검정

귀무가설 $H_0 : p_1 - p_2 \geq 0$에 대한 대립가설 $H_1 : p_1 - p_2 < 0$으로 구성된 검정

검정통계량과 확률분포 : $Z = \dfrac{\hat{P}_1 - \hat{P}_2}{\sqrt{\hat{p}\hat{q}\left(\dfrac{1}{n} + \dfrac{1}{m}\right)}} \sim N(0,1)$

검정통계량의 관찰값 : $z_0 = \dfrac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}\hat{q}\left(\dfrac{1}{n} + \dfrac{1}{m}\right)}}$

미리 주어진 유의수준 α에 대한 $H_0 : p \geq p_0$의 기각역 : $Z \leq - z_\alpha$

$$\boxed{z_0 = \dfrac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}\hat{q}\left(\dfrac{1}{n} + \dfrac{1}{m}\right)}} \leq -z_\alpha}$$

p – 값 = $P(Z < z_0)$, p – 값 $\leq \alpha$이면 H_0을 기각한다.

모비율 차에 대한 검정 유형과 기각역

기각역 및 p-값 / 검정 방법	귀무가설 H_0	대립가설 H_1	H_0 의 기각역	p-값
하단측검정	$p_1 - p_2 \geq 0$ $(p_1 - p_2 = 0)$	$p_1 - p_2 < 0$	$Z < -z_\alpha$	$P(Z < z_0)$
상단측검정	$p_1 - p_2 \leq 0$	$p_1 - p_2 > 0$	$Z > z_\alpha$	$P(Z > z_0)$
양측검정	$p_1 - p_2 = 0$	$p_1 - p_2 \neq 0$	$\lvert Z \rvert > z_{\alpha/2}$	$2[1 - P(Z < \lvert z_0 \rvert)]$

예제 10-6

어느 자동차 회사는 동종의 자동차를 두 생산라인 A와 B에서 생산한다. 이 회사에서 이미 판매된 자동차 중에서 결함이 발견된 자동차 수는 다음과 같다.

생산라인 \ 자동차 수	판매된 자동차 수	결함이 발견된 자동차 수
생산라인 A	185	4
생산라인 B	191	10

두 생산라인에서 생산된 자동차의 결함 비율이 동일하다는 주장을 유의수준 5%에서 다음 같은 방법으로 검정하라.

(a) 기각역을 이용하는 방법

(b) $p-$값을 이용하는 방법

풀이

모비율 차에 대한 검정 유형과 기각역

(a) 생산라인 A와 B의 결함 비율을 각각 p_1, p_2라 하고 다음 순서에 따라 가설을 검정한다.

① 귀무가설과 대립가설을 설정한다.

귀무가설 $H_0 : p_1 - p_2 = 0$에 대한 대립가설 $H_1 : p_1 - p_2 \neq 0$이다.

② 유의수준 $\alpha = 0.05$에 대한 상단측검정의 기각역은

$$R : Z \leq -z_{0.025} = -1.96,\ \ Z \geq z_{0.025} = 1.96$$

③ 합동표본비율은 다음과 같다.

$$\hat{p} = \frac{4+10}{185+191} = 0.03723$$

④ 검정통계량을 정한다.

$$z_0 = \frac{\hat{P}_1 - \hat{P}_2}{\sqrt{0.03723 \times 0.96277\left(\frac{1}{185} + \frac{1}{191}\right)}} = \frac{\hat{P}_1 - \hat{P}_2}{0.01953}$$

모비율 차에 대한 검정 유형과 기각역

⑤ 검정통계량의 관찰값을 구한다.

두 표본비율이 각각 $\hat{p}_1 = 0.02162,\ \hat{p}_2 = 0.05236$ 이므로 검정통계량의 관찰값은 다음과 같다.

$$z_0 = \frac{0.02162 - 0.05236}{0.01953} = -1.57$$

⑥ 귀무가설 H_0의 기각을 결정한다.

검정통계량의 관찰값 $z_0 = -1.57$이 기각역 안에 놓이지 않으므로 $H_0 : p_1 - p_2 = 0$을 기각하지 않는다.

(b) 검정통계량의 관찰값이 $z_0 = -1.57$이므로 p – 값 = $2[1-P(Z<1.57)] = 0.1164 > 0.05$이고, 따라서 귀무가설 $H_0 : p_1 - p_2 = 0$을 기각하지 않는다.

적합도 검정

❖ 단일 모비율에 대한 주장 또는 두 모비율의 등가성, 즉 $p_1 = p_2$에 대한 주장을 검정하기 위하여 정규분포를 사용하였다.

❖ 여러 개의 범주에 대한 등가성, 즉 $p_1 = p_2 = p_3 = \ldots = p_k = p$ 에 대한 주장을 검정하기 위하여 카이제곱분포를 사용한다.

기대도수 expected frequency : 이론적으로 각 범주에 대한 기대되는 도수

관측도수 observed frequency : 실험이나 관측에 의하여 실제로얻어진 각 범주의 도수

적합도 검정

적합도goodness of fit : 실험 또는 관찰로부터 얻은 관측도수와 기대도수가 어느 정도로 일치하는가를 나타내는 값

적합도 검정goodness-of-fit test : 관측값들이 어느 정도로 이론적인 분포에 따르고 있는가를 보이는 검정

예 주사위를 던져서 주사위 눈 i가 나온 비율 p_i가 동등한지 검정하기 위해서 공정한 주사위를 30번 던져서 다음 결과를 얻었다고 하자.

주사위 눈	1	2	3	4	5	6
기대도수	5	5	5	5	5	5
관측도수	4	6	5	7	3	5

적합도 검정

❖ 이 경우에 귀무가설과 대립가설은 다음과 같이 정의한다.

귀무가설 H_0 : $p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = 1/6$
대립가설 H_1 : H_0이 아니다.

- 적합도 검정에서 귀무가설을 채택하거나 기각하는 결정을 위하여 **카이제곱분포**를 사용하며,

i번째 범주의 도수 n_i에 대한 특정한 성질을 갖는 성분의 관측도수 o_i와 기대도수 e_i에 대하여 다음과 같은 χ^2-통계량을 이용한다.

$$\boxed{\chi^2 = \sum_{i=1}^{k} \frac{(o_i - e_i)^2}{e_i}}$$

- 이때 범주의 수 k에 대하여 자유도 $k-1$인 카이제곱분포를 사용하며, 적합도 검정은 항상 **상단측검검정**을 이용한다.

적합도 검정 방법

❶ 귀무가설 H_0과 대립가설 H_1을 설정한다.

❷ 유의수준 α에 대한 기각역을 구한다.

❸ 표본으로부터 검정통계량의 관찰값   을 구한다.

❹ 관찰값 χ_0^2이 기각역 안에 들어 있으면 귀무가설 H_0을 기각시키고, 그렇지 않으면 H_0을 기각시키지 않는다.

예 주사위 눈의 비율이 동일한지 유의수준 5%에서 검정해보자.

① 귀무가설과 대립가설을 설정한다.
귀무가설 H_0 : $p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = 1/6$
대립가설 H_1 : H_0이 아니다.

② 범주의 수가 6이므로 자유도 5인 카이제곱분포에서 유의수준 $\alpha = 0.05$에 대한 상단측검정의 기각역은 $R: \ \chi_0^2 \geq \chi_{0.05}^2(5) = 11.07$

③ 검정통계량의 관찰값을 구한다.

적합도 검정 방법

범주	관측도수(o_i)	비율(p)	기대도수 ($e_i = np_i$)	$o_i - e_i$	$(o_i - e_i)^2$	$\frac{(o_i - e_i)^2}{e_i}$
1	4	$\frac{1}{6}$	5	-1	1	0.2
2	6	$\frac{1}{6}$	5	1	1	0.2
3	5	$\frac{1}{6}$	5	0	0	0.0
4	7	$\frac{1}{6}$	5	2	4	0.8
5	3	$\frac{1}{6}$	5	-2	4	0.8
6	5	$\frac{1}{6}$	5	0	0	0.0
합계	–	–	–	–	–	$v_0 = 2.0$

④ 검정통계량의 관찰값 $\chi_0^2 = 2$는 기각역 안에 놓이지 않으므로 귀무가설을 기각하지 않는다. 즉, 주사위는 공정하게 만들어졌다고 할 수 있다.

적합도 검정 방법

- 귀무가설 $H_0:\ p_1 = p_2 = \ldots = p_k = p$의 동등한 비율 p의 값이 주어지지 않는 경우에 다음 합동표본비율을 사용한다.

$$\boxed{\hat{p} = \frac{x_1 + x_2 + \cdots + x_k}{n}}$$

- 이때 x_i는 i번째 범주의 관찰도수이고 기대도수는 $e_i = n_i\hat{p}$, $n = n_1 + n_2 + \ldots + n_k$ 이다.

적합도 검정 방법

예제 10-7

어느 자동차 회사는 동종의 자동차를 세 생산라인에서 생산한다. 이 회사에서 이미 판매된 자동차 중에서 결함이 발견된 자동차 수는 다음과 같다.

생산라인 \ 자동차 수	판매된 자동차 수	결함이 발견된 자동차 수
생산라인 1	185	4
생산라인 2	191	10
생산라인 3	201	6

세 생산라인에서 생산된 자동차의 결함 비율이 동일하다는 주장을 유의수준 5%에서 검정하라.

풀이

① 각 라인의 결함비율을 p_1, p_2, p_3이라 하고, 귀무가설과 대립가설을 설정한다.
귀무가설 $H_0 :\ p_1 = p_2 = p_3$
대립가설 $H_1 : H_0$이 아니다.

② 범주의 수가 3이므로 자유도 2인 카이제곱분포에서 유의수준 $\alpha = 0.05$에 대한 상단측검정의 기각역은 $R :\ \chi^2 \geq \chi^2_{0.05}(2) = 5.99$

적합도 검정 방법

③ 합동표본비율을 구하면 다음과 같다.

$$\hat{p} = \frac{4+10+6}{185+191+201} = 0.0347$$

따라서 각 생산라인의 결함이 있을 것으로 기대되는 자동차의 수는
각각 다음과 같다.

④ 검정통계량 $\chi_0^2 = \sum_{i=1}^{k} \frac{(o_i - e_i)^2}{e_i}$ 의 관찰값을 구한다.

범주	관측도수(o_i)	비율(p)	기대도수 ($e_i = np_i$)	$o_i - e_i$	$(o_i - e_i)^2$	$\frac{(o_i - e_i)^2}{e_i}$
1	4	0.0347	6.42	−2.42	5.8564	0.9122
2	10	0.0347	6.63	3.37	11.3569	1.7130
3	6	0.0347	6.97	−0.97	0.9409	0.1350
합계	–	–	–	–	–	2.7602

⑤ 검정통계량의 관찰값 $\chi_0^2 = 2.7602$ 는 기각역 안에 놓이지 않으므로 귀무가설을
기각하지 않는다. 즉, 세 생산라인의 결함 비율은 동일하다는 근거가 충분하다.

Q&A