표 9-1 귀무가설에 대한 판정과 오류

검정결과	실제	
	H_0가 참	H_0가 거짓
채택	옳은 결정 확률 $= 1-\alpha$	제2종 오류 확률 $= \beta$(β 위험)
기각	제1종 오류 확률 $= \alpha$(유의수준)	옳은 결정 확률 $= 1-\beta$(검정력)

$$\text{검정통계량} = \frac{\text{표본 통계량} - \text{귀무가설에서 설정된 모수값}}{\text{표본 통계량의 표준오차}}$$

가설검정의 목적은 귀무가설의 타당성 여부 판단에 있는데, 이 결정의 기준이 되는 표본통계량-> 검정통계량(test statistic)

예) 모평균에 대한 가설검정-> 검정통계량은 모평균에 대한 통계량인 표본평균으로부터 구할 수 있음

결정규칙

표본정보에 근거하여 귀무가설을 채택하는 경우와 대립가설이 보다 타당하다고 판단되어 귀무가설을 기각하는 경우 중 하나를 결정하는 기준

제1종 오류

귀무가설이 참일 때 귀무가설을 기각하는 오류

제2종 오류

귀무가설이 거짓일 때 귀무가설을 채택하는 오류

α 위험 혹은 유의수준

제1종 오류의 발생확률

$\alpha = P(\text{제1종 오류}) = P(H_0\text{를 기각} \mid H_0\text{가 참})$

β 위험

제2종 오류의 발생확률

$\beta = P$(제2종 오류)$= P(H_0$를 채택 | H_0가거짓)

검정력

거짓 귀무가설을 기각할 확률$(1-\beta)$

검정통계량

귀무가설의 타당성 여부를 결정하는 기준이 되는 표본 통계량

임계값 CV(critical value)

주어진 유의수준에서 귀무가설을 채택하거나 기각하는 의사결정을 할 때 검정 통계량과의 비교기준이 되는 값

2. 정규모집단 평균에 대한 가설검정

1) 모분산을 아는 경우

모평균에 대한 가설검정을 위해 모집단이 정규분포하며 모분산을 안다고 가정하자. 물론 표본의 크기가 충분히 크다면 이런 가정은 필요 없게 된다. 모평균이 특정 값과 동일하다는 단순귀무가설

$$H_0 : \mu = \mu_0$$

을 검정하는 방법에 대해 살펴보자. 이 귀무가설이 잘못되었을 때 받아들이게 되는 대립가설로 모평균이 특정값보다 크다고 설정하자.

$$H_1 : \mu > \mu_0$$

이 가설검정의 경우 참인 귀무가설을 기각하는 확률인 유의수준 α를 정하고 그 수준에서 β를 최소화하는 결정규칙을 선택한다. 즉 확률변수

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

가 표준정규분포한다는 사실을 이용하여 주어진 유의수준에서 검정통계량이 어떤 영역에 속할 경우 귀무가설을 기각하게 되는데, 이 영역을 **기각영역** 혹은 **기각역**(rejection region)이라 한다.

$\bar{x}$가 $\bar{x}_U$보다 크면 H_0를 기각

$\alpha = P(H_0$를 기각 $|$ H_0가 참$)$

$\alpha = P(\bar{x} > \bar{x}_U \mid H_0$가 참$)$

을 만족해야 한다. 여기서 $\bar{x}$를 통계량$(\bar{x}-\mu_0)/(\sigma/\sqrt{n})$으로 변형하면 표준정규확률분포를 이용한 기각역을 정할 수 있다.

$$\alpha = P\left(\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} > \frac{\bar{x}_U-\mu_0}{\sigma/\sqrt{n}} \mid H_0\text{가 참}\right)$$
$$= P(Z > z_\alpha \mid H_0\text{가 참})$$

결정규칙

$$\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} > z_\alpha \text{이면 } H_0\text{를 기각}$$

혹은

$$\bar{x} > \mu_0 + z_\alpha \sigma/\sqrt{n} \text{ 이면 } H_0\text{를 기각}$$

대립가설 $H_1 : \mu > \mu_0$ 대신$H_1 : \mu < \mu_0$를 설정하면

$\bar{x}$가 $\bar{x}_L$보다 작으면 H_0를 기각

결정규칙

$$\frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} < -z_\alpha \text{이면 } H_0\text{를 기각}$$

$$\bar{x} < \mu_0 - z_\alpha \sigma / \sqrt{n} \text{ 이면 } H_0\text{를 기각}$$

단측대립가설 대신 양측대립가설 $H_1 : \mu \neq \mu_0$를 설정하면 $\bar{x}$가 μ_0보다 매우 크거나 매우 작을 때 귀무가설을 기각하게 될 것이다.

$\bar{x}$가 $\bar{x}_U$보다 크거나 $\bar{x}_L$보다 작으면 H_0를 기각

유의수준 α가 정해지면

$$\alpha = P(\bar{x} > \bar{x}_U \text{ 혹은 } \bar{x} < \bar{x}_L \mid H_0\text{가 참})$$
$$= P(\bar{x} > \bar{x}_U \mid H_0\text{가 참}) + P(\bar{x} < \bar{x}_L \mid H_0\text{가 참})$$

그림 9-2
$\bar{X}$ 의 표본분포 및
z 분포와 기각역 :
단측검정

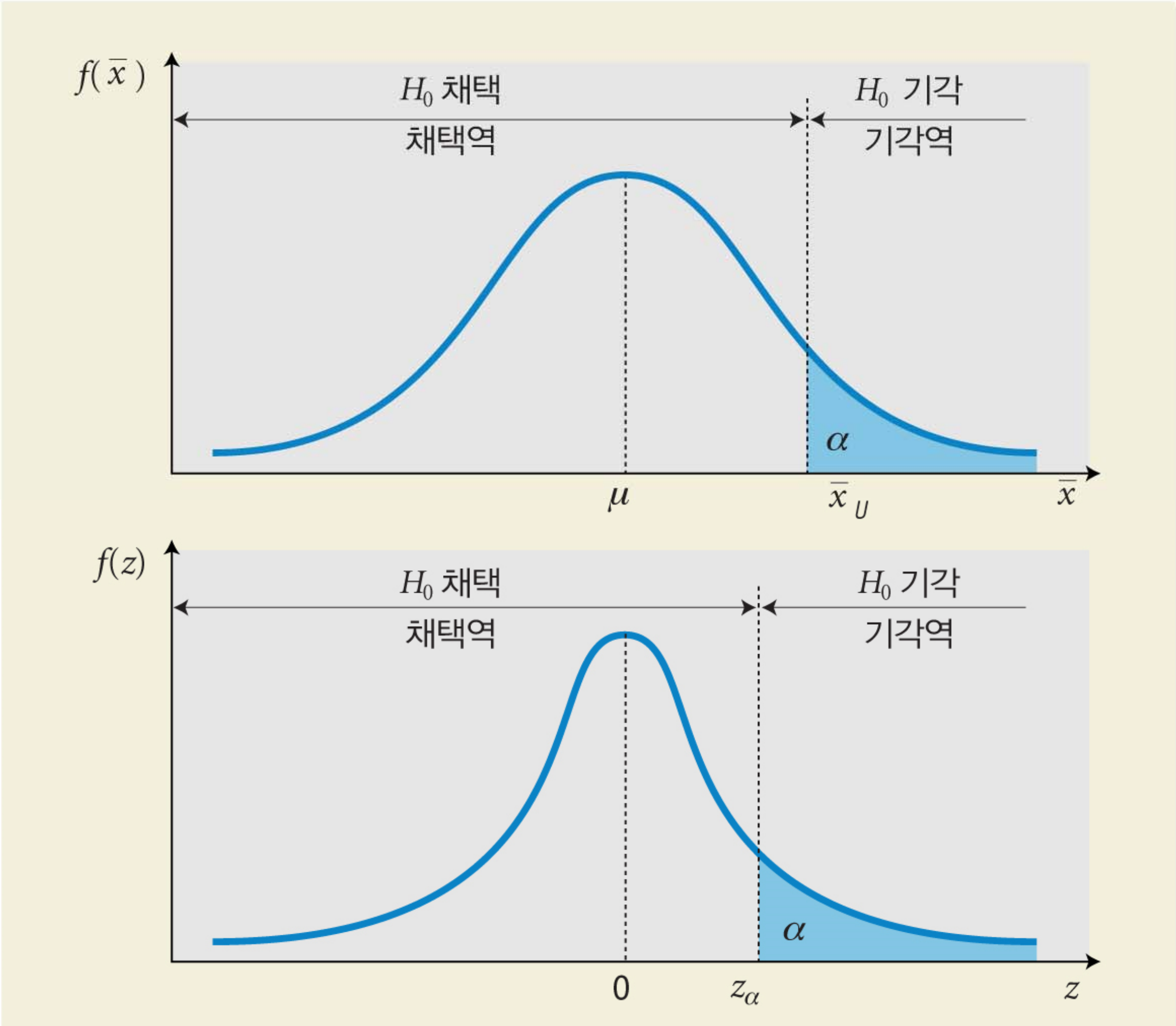
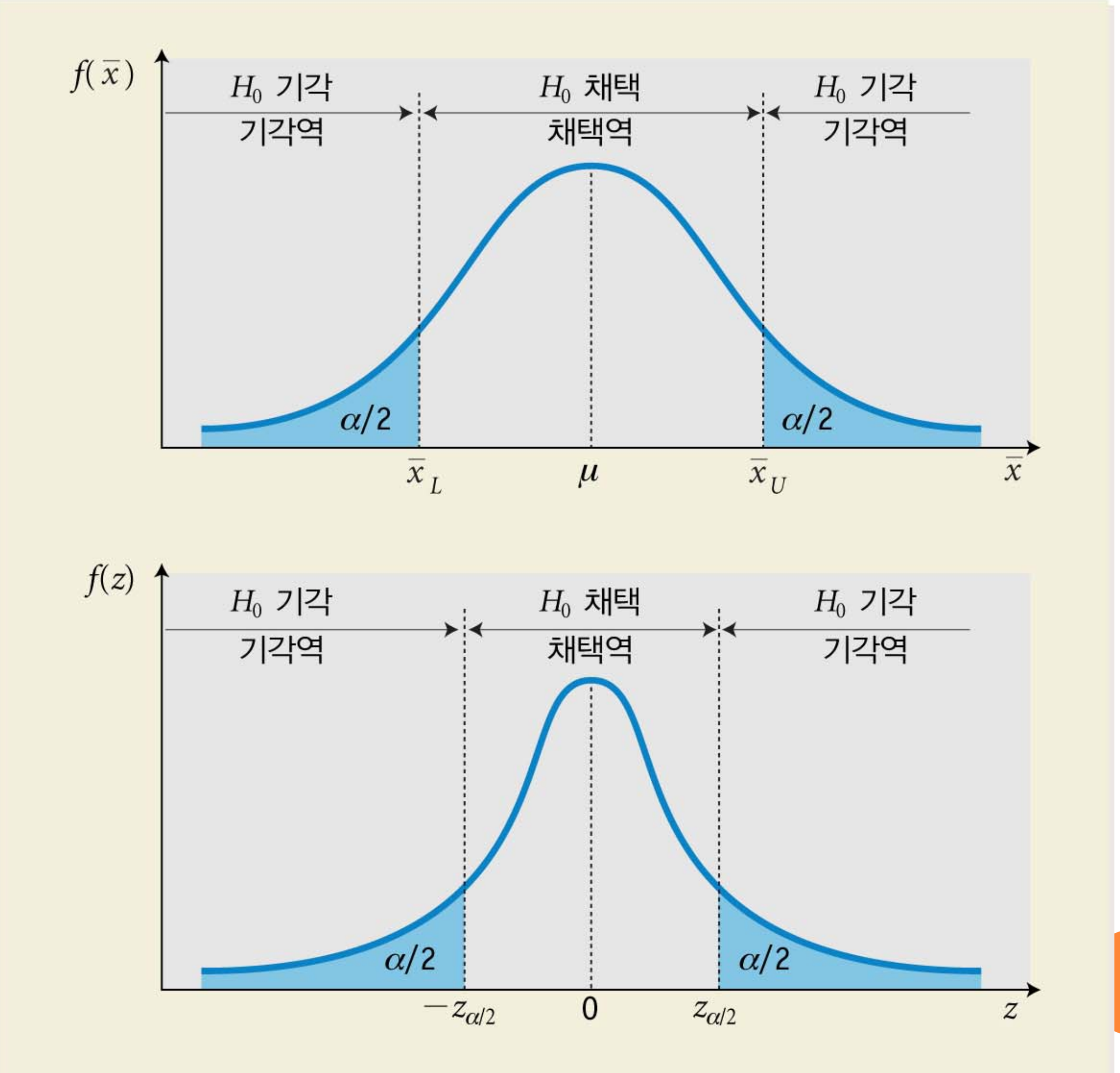


그림 9-3
$\bar{X}$의 표본분포 및 z분포와 기각역 : 양측검정



$\bar{x}$ 를 통계량$(\bar{x}-\mu_0)/(\sigma/\sqrt{n})$으로 전환하면

$$\alpha = P\left(\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} > \frac{\bar{x}_U-\mu_0}{\sigma/\sqrt{n}} \mid H_0\text{가 참}\right) + P\left(\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} < \frac{\bar{x}_L-\mu_0}{\sigma/\sqrt{n}} \mid H_0\text{가 참}\right)$$

$$= P(Z > z_{\alpha/2} \mid H_0\text{가 참}) + P(Z < -z_{\alpha/2} \mid H_0\text{가 참})$$

따라서 양측검정의 경우 결정규칙은 다음과 같다.

$$\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} > z_{\alpha/2}\text{이거나 } \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} < -z_{\alpha/2}\text{이면 } H_0\text{를 기각}$$

혹은

$$\bar{x} > \mu_0 + z_{\alpha/2}\sigma/\sqrt{n}\text{ 이거나 } \bar{x} < \mu_0 - z_{\alpha/2}\sigma/\sqrt{n}\text{ 이면 } H_0\text{를 기각}$$

정규분포하는 모집단 평균의 가설검정 : 모분산을 아는 경우

표본평균의 값은 $\bar{x}$ 이며 가설검정의 유의수준은 α이다.

1. 가설 설정 $H_0 : \mu=\mu_0$, $H_1 : \mu>\mu_0$

   검정통계량 $Z=\dfrac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$

   결정규칙 – 검정통계량이 임계값 z_α보다 크면 H_0를 기각

2. 가설 설정 $H_0 : \mu=\mu_0$, $H_1 : \mu<\mu_0$

   검정통계량 $Z=\dfrac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$

   결정규칙 – 검정통계량이 임계값 $-z_\alpha$보다 작으면 H_0를 기각

3. 가설 설정 $H_0 : \mu=\mu_0$, $H_1 : \mu\neq\mu_0$

   검정통계량 $Z=\dfrac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$

   결정규칙 – 검정통계량이 임계값 $z_{\alpha/2}$보다 크거나 $-z_{\alpha/2}$보다 작으면 H_0를 기각

자주 사용되는 임계값 : $z_{0.10}=1.283$, $z_{0.05}=1.645$, $z_{0.025}=1.960$, $z_{0.01}=2.326$

예 9-1

앞의 생활폐기물의 예에서 도시 거주민들의 1인당 생활폐기물량은 정규분포하며 표준편차는 0.5kg이라고 가정하자. 도시 거주민들의 평균 생활폐기물량이 1kg이라는 가설을 검정하기 위해 도시 거주민 중 20명을 임의로 선정하여 1일 1인당 생활폐기물량을 조사해 보니 표본평균이 1.3kg이었다.

- "1일 1인당 생활폐기물량은 1kg이 아니다."라는 대립가설을 세우고 5%의 유의수준에서 가설검정을 하라.
- "1일 1인당 생활폐기물량은 1kg을 넘는다."라는 대립가설을 세우고 5%의 유의수준에서 가설검정을 하라.

답

- 이 문제에서 가설검정의 대상은 모평균이며 1일 1인당 생활폐기물량이 1kg이라는 가설을 기각하거나 채택할 수 있는 충분한 통계적 증거를 얻어야 할 것이다. 가설은 다음과 같이 설정된다.

$$H_0 : \mu=1, \quad H_1 : \mu\neq1$$

주어진 정보 $\bar{x}=1.3$, $\mu_0=1$, $\sigma=0.5$, $n=20$을 이용하여 검정통계량을 계산하면

$$\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} = \frac{1.3-1}{0.5/\sqrt{20}} = 2.6833$$

이 된다.

이 된다. 5%의 유의수준에서 양측검정의 경우 $\alpha/2=0.025$이므로 임계값은 $z_{0.025}=1.96$이다. $(\bar{x}-\mu_0)/(\sigma/\sqrt{n})=2.6833>z_{0.025}=1.96$이므로 검정통계량이 $z_{\alpha/2}$보다 크거나 $-z_{\alpha/2}$보다 작으면 H_0를 기각하는 결정규칙에 의해 H_0를 기각한다. 즉 통계적 유의성을 가지고 도시민들의 1일 1인당 평균 생활폐기물량이 1kg은 아니라고 할 수 있다.

- $H_0 : \mu=1, \quad H_1 : \mu>1$

검정통계량은 2.6833으로 같지만 단측검정의 경우 $\alpha=0.05$로 임계값이 $z_{0.05}=1.645$가 된다. $2.6833>1.645$이므로 검정통계량이 z_α보다 크면 결정규칙에 의해 H_0를 기각한다. 따라서 5%의 유의수준에서 도시민들의 1일 1인당 평균 생활폐기물량은 1kg을 넘는다고 주장할 수 있다.

2) 모분산을 모르는 경우

(1) 대표본의 경우

검정통계량은

$$Z = \frac{\overline{X} - \mu_0}{S/\sqrt{n}}$$

예 9-2

금융전문가들은 우리나라 금융기관들의 평균 배당률이 10%라고 예상한다. 이 예상이 타당한지 알아보기 위해 금융기관 40개를 표본추출하여 배당률을 계산해 보니 평균은 9.3%, 표준편차는 3%였다. 우리나라 금융기관들의 평균 배당률이 10%라는 전문가들의 주장이 타당한지 알아보기 위해 0.10의 유의수준에서 평균 배당률이 10% 미만이라는 대립가설을 가지는 단측검정과 평균 배당률이 10%가 아니라는 양측검정을 하라.

답

- 단측검정

$$H_0 : \mu = 10, \ \ H_1 : \mu < 10$$

문제로부터 $\bar{x} = 9.3$, $\mu_0 = 10$, $S = 3$, $n = 40$임을 알았고, 이 정보를 이용하여 검정통계량을 계산하면

$$\frac{\bar{X} - \mu_0}{S/\sqrt{n}} = \frac{9.3 - 10}{3/\sqrt{40}} = -1.4757$$

0.10의 유의수준에서 $\alpha = 0.10$, 임계값은 $-z_{0.10} = -1.28$이다.
$(\bar{x} - \mu_0)/(S\sqrt{n}) = -1.4757 < -1.28$이므로 검정통계량이 $-z_\alpha$보다 작으면 H_0를 기각하는 결정규칙에 의해 귀무가설은 기각된다.

- 양측검정

$$H_0 : \mu=10,\ \ H_1 : \mu\neq10$$

0.10의 유의수준에서 양측검정을 할 경우 $\alpha/2=0.05$이며 임계값은 $-z_{0.05}=-1.645$이다. $(\bar{x}-\mu_0)/(S\sqrt{n})=-1.4757>-1.645$이므로 결정규칙에 의해 귀무가설이 채택된다. 이 예에서 알 수 있는 것처럼 단측검정과 양측검정의 결과가 다르게 나타날 수도 있다. 0.10의 유의수준에서 우리나라 금융기관들의 평균 배당률이 10% 미만이라는 통계적 증거는 제시되지만 우리나라 금융기관들의 평균 배당률이 10%가 아니라는 통계적 증거는 제시되지 못한다는 것이다.

(2) 소표본의 경우

정규분포하는 모집단 평균의 가설검정 : 모분산을 모르며 소표본인 경우

표본평균값은 $\bar{x}$ 이며 가설검정의 유의수준은 α이다.

1. 가설 설정 $H_0 : \mu = \mu_0$ 혹은 $\mu \leq \mu_0$

$H_1 : \mu > \mu_0$

검정통계량 $t_{n-1} = \dfrac{\bar{x} - \mu_0}{S/\sqrt{n}}$

결정규칙 − 검정통계량이 $t_{(n-1,\, \alpha)}$보다 크면 H_0를 기각

2. 가설 설정 H_0 : $\mu=\mu_0$ 혹은 $\mu \geq \mu_0$

H_1 : $\mu<\mu_0$

검정통계량 $t_{n-1}=\dfrac{\bar{x}-\mu_0}{S/\sqrt{n}}$

결정규칙−검정통계량이 임계값 $-t_{(n-1,\ \alpha)}$보다 작으면 H_0를 기각

3. 가설 설정 H_0 : $\mu=\mu_0$

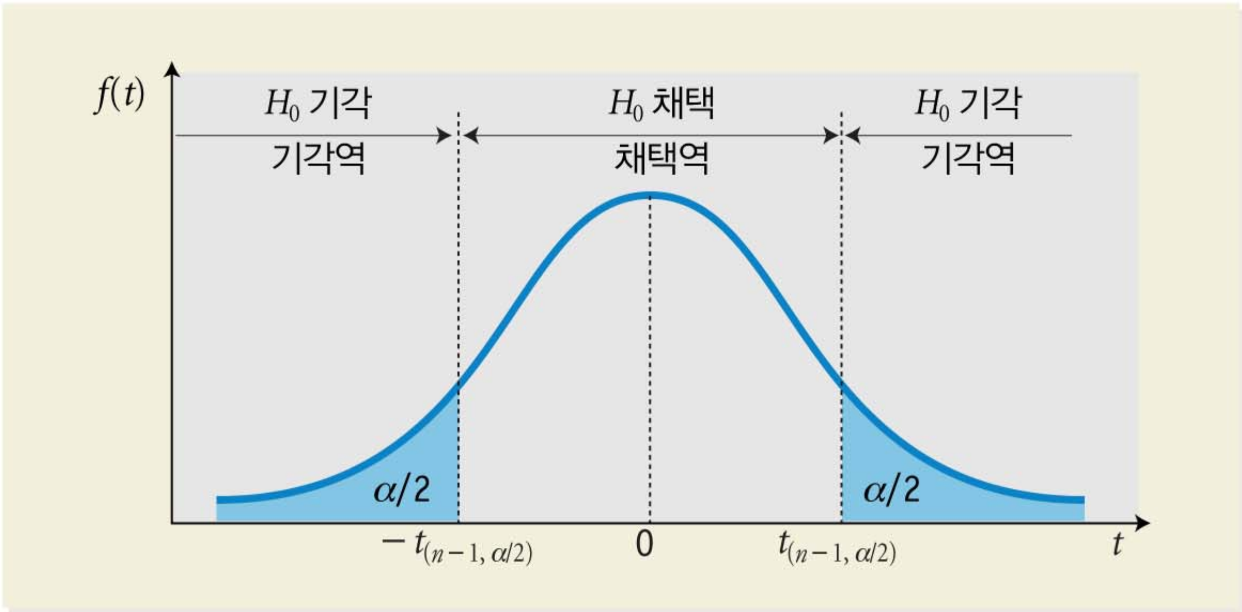
H_1 : $\mu\neq\mu_0$

검정통계량 $t_{n-1}=\dfrac{\bar{x}-\mu_0}{S/\sqrt{n}}$

결정규칙−검정통계량이 임계값 $t_{(n-1,\ \alpha/2)}$보다 크거나 $-t_{(n-1,\ \alpha/2)}$보다 작으면 H_0를 기각

그림 9-4

t-분포와 기각역 : 양측검정



예 9-3

어떤 아이스크림 회사의 영업부 사원은 체인점의 여름 판매량보다 겨울 판매량이 평균 34.5% 감소한다고 한다. 전국 체인점 중 15개를 표본추출하여 판매 감소량을 조사해 보니 다음과 같이 나타났다.

표 9-2 판매 감소량(단위 : %)

33.46	33.38	32.73	32.15	33.99	34.10	33.97	34.34
33.95	33.85	34.23	34.05	34.13	34.45	34.19	

모집단이 정규분포한다고 가정하고 판매량의 감소가 평균 34.5%라는 귀무가설을 10% 유의수준에서 양측검정하라.

답

검정하려는 귀무가설과 대립가설은

$$H_0 : \mu = 34.5, \quad H_1 : \mu \neq 34.5$$

이며 주어진 정보와 표본자료에 대한 계산결과에 의하면

$$\bar{x} = 33.798, \quad S = 0.630, \quad \mu_0 = 34.5, \quad n = 15$$

이다. 이 정보를 이용한 검정통계량의 계산값은

$$\frac{\bar{x} - \mu_0}{S/\sqrt{n}} = \frac{33.798 - 34.5}{0.630\sqrt{15}} = -4.314$$

이다. 10% 유의수준에서 $\alpha = 0.10$이며 양측검정의 경우 $\alpha/2 = 0.05$이고 임계값은 $-t_{(14,\ 0.05)} = -1.761$이다. $(\bar{x} - \mu_0)/(S/\sqrt{n}) = -4.314 < -1.761$이므로 결정규칙에 의해 H_0를 기각하게 된다. 따라서 체인점들의 평균 판매량 감소가 34.5%라는 주장은 부당하다고 결론지을 수 있다.