8.3절~8.5절

11주1강 가설과 검정2

복습

가설검정

- **가설검정**이란 모집단에 대한 어떤 가설을 설정한 뒤에 표본관찰을 통하여 그 가설의 채택 여부를 결정하는 분석방법이다.
- 일반적으로 통계분석에서는 모집단의 모수에 대하여 관심이 있으므로 **가설은 모수에 대하여 설정**한다.

가설설정

- 가설검정에서 가설은 귀무가설(H_0)과 대립가설(H_1)로 설정한다.

① 귀무가설 : "모수가 특정한 값이다" , "두 모수의 값이 같다" 등과 같이 간단하고 구체적인 경우를 귀무가설로 설정한다.
② 대립가설 : "모수가 특정한 값이 아니다" , "한 모수의 값이 다른 모수의 값보다 크다" , "두 모수의 값이 다르다" 등과 같이 모수에 대한 관심의 영역 중에서 귀무가설로 지정되지 않은 모든 경우를 포괄적으로 대립가설로 설정한다.

- **가설검정**이란 두 가설 H_0와 H_1중에서 하나를 선택하는 과정이므로 H_0를 채택(accept)하면 H_1을 기각(reject)하게 되고 H_0를 기각하면 H_1을 채택하게 된다.
- 따라서 **가설검정이란 '두 가설 중에서 귀무가설 H_0 를 채택하든지 또는 기각하는 과정'**이라고 이해할 수 있다.

검정통계량

- **검정통계량**이란 가설검정에서 관찰된 표본으로부터 구하는 통계량으로 분포가 가설에서 주어지는 모수에 의존한다.
- 귀무가설이 옳다는 전제하에서 구한 검정통계량의 값이 나타날 가능성이 크면 귀무가설을 채택하고 나타날 가능성이 작으면 귀무가설을 기각한다.

유의수준

- **유의수준** α란 귀무가설이 옳은데도 불구하고 이를 기각하는 확률의 크기를 말하며 검정통계량을 구하는 것과는 무관하게 검정을 실시하는 사람의 판단에 따라 결정한다.

기각역

- **기각역**이란 가설검정에서 유의수준 α가 정해졌을 때, 검정통계량의 분포에서 이 유의수준의 크기에 해당하는 영역을 말하는데, 검정통계량의 분포에서 이 영역의 위치는 대립가설의 형태에 따라 다르다.

- **기각역 C 와 유의수준 α의 관계**

유의수준 α ; 귀무가설 하에서 검정통계량이 기각역 C에 속할 확률이다.

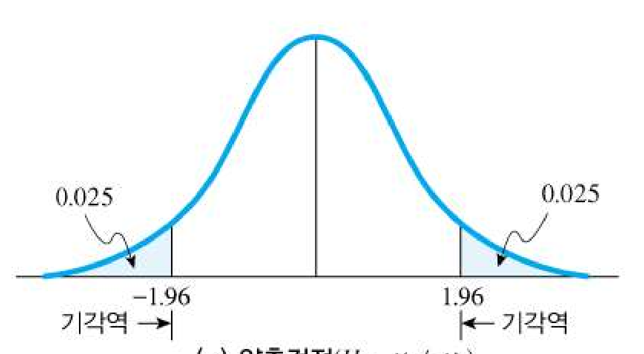
$$P_r(T(X) \in C \mid H_0) = \alpha$$

대립가설과 기각역, $\alpha = 0.05$

- 검정통계량의 분포에서 유의수준 α에 의해 기각역의 크기가 결정되며, 기각역의 위치는 대립가설 H_1의 형태에 의해 결정된다.
- 대립가설의 형태는 가설검정의 목적에 의하여 결정되는데 가설검정은 대립가설의 형태에 따라 양측검정과 단측검정으로 나누어지고, 단측검정은 다시 왼쪽 단측검정과 오른쪽 단측검정으로 분류된다.

① 양측검정
귀무가설이 "모수가 특정값이다"라고 할 때, 대립가설이 "모수가 특정값이 아니다"라고 주어지는 경우로 귀무가설과 대립가설은 다음과 같이 표현할 수 있다.

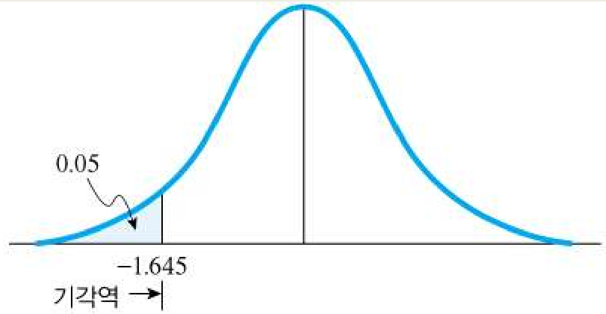
$H_0 : \mu = \mu_0$; μ_0은 고정된 상수
$H_1 : \mu \neq \mu_0$
기각역 $C = \{T(X) \leq -C_1$또는 $T(X) \geq C_1\}$



(a) 양측검정(H_1 ; $\mu \neq \mu_0$)

② 왼쪽 단측검정
귀무가설이 "모수가 특정값이다"라고 할 때, 대립가설 H_1이 "모수가 μ_0보다 작다"로 주어지는 경우로 귀무가설과 대립가설을 다음과 같이 표현할 수 있다.

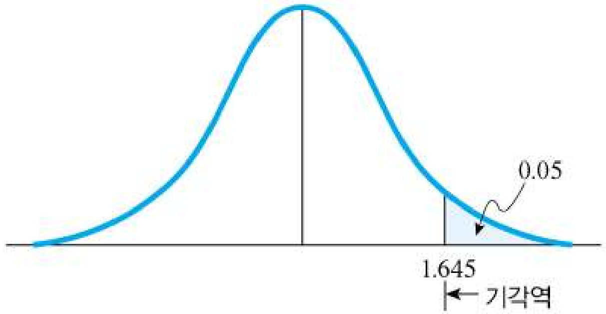
$H_0 : \mu = \mu_0$
$H_1 : \mu < \mu_0$
기각역 $C = \{T(X) \leq C_2\}$



(b) 왼쪽 단측검정(H_1 ; $\mu < \mu_0$)

③ 오른쪽 단측검정
귀무가설이 "모수가 특정값이다"라고 할 때, 대립가설 H_1이 "모수가 μ_0보다 크다"로 주어지는 경우로 귀무가설과 대립가설을 다음과 같이 표현할 수 있다.

$H_0 : \mu = \mu_0$
$H_1 : \mu > \mu_0$
기각역 $C = \{T(X) \geq C_3\}$



(c) 오른쪽 단측검정(H_1 ; $\mu > \mu_0$)

단일모평균 μ의 검정(t-검정)

- 단일집단의 모평균 μ에 대한 검정은 모집단의 분포가 평균 μ와 분산 σ^2을 갖는 정규분포 $N(\mu,\sigma^2)$을 따른다는 것을 전제한다고 할 수 있다. 검정을 하기 위하여 추출한 표본을 $(X_1,X_2,\ldots,X_n)$이라고 할 때, $(X_1,X_2,\ldots,X_n)$은 $N(\mu,\sigma^2)$으로부터의 확률표본이다.

① 가설의 설정

(a) 양측검정 : $H_0 : \mu = \mu_0,\ \ H_1 : \mu \neq \mu_0$

(b) 단측검정 : $H_0 : \mu = \mu_0,\ \ H_1 : \mu > \mu_0$ (또는 $\mu < \mu_0$)

→ 여기에서 μ_0은 구체적으로 주어지는 특정한 값이며 H_1에 주어지는 세 가지 서로 다른 가설은 가설검정에서 알고자 하는 목적에 따라 결정된다.

② 귀무가설 하에서의 검정통계량과 분포

(a) σ^2을 아는 경우　　　　(b) σ^2을 모르는 경우

② 귀무가설 하에서의 검정통계량과 분포

(a) σ^2을 아는 경우

$$T(X) = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0,\ 1)$$

(b) σ^2을 모르는 경우

(i) $n > 30$일 때

$$T(X) = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim N(0,\ 1)$$

(ii) $n \leq 30$일 때

$$T(X) = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim t_{n-1}$$

- [(a) & (b)-(i)] 표준정규분포를 이용하여 검정하는 경우
  모분산 σ^2이 알려져 있는 경우와 σ^2이 알려져 있지 않으나 표본의 수 n이 이상이면 검정통계량의 분포가 근사적으로 표준정규분포를 따르므로 표준정규분포를 이용하여 검정을 실시한다.
- [(b)-(ii)] t-분포를 이용하여 검정하는 경우
  모분산 σ^2이 알려져 있지 않고 표본의 수 n이 30 이하이면 자유도가 $n-1$인 t-분포를 이용하여 검정을 실시한다.
  이 검정에서 검정통계량 $T(X)$가 자유도 $n-1$인 t-분포를 따르므로 ***t*-검정**이라고 한다.

8.3절

8.3 두 집단 모평균의 동일성에 대한 검정

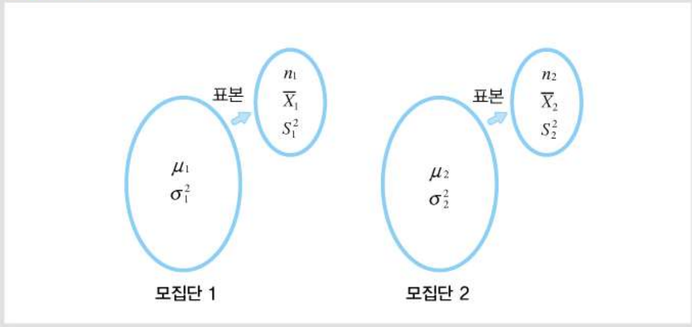
두 집단의 모평균의 동질성에 대한 검정

- 두 집단의 모평균의 동질성에 대한 검정에 있어서의 전제조건은 '**두 집단이 서로 독립이며 두 집단 모두 정규분포를 따른다**'는 것이다.
- 일반적으로 두 집단이 서로 독립이며 각 집단의 평균과 분산이 각각 $(\mu_1, {\sigma_1}^2)$과 $(\mu_2, {\sigma_2}^2)$인 정규분포를 따를 때 두 모평균의 동일성 $(\mu_1 = \mu_2)$에 대한 검정은 각 집단에서 '랜덤하게' n_1과 n_2개의 표본을 추출하여 실시한다.

표 10-2 두 독립집단의 모수와 통계량

모수와 통계량 / 집단	모수		표본의 크기	통계량	
	모평균	모분산		표본평균	표본분산
집단 1	μ_1	σ_1^2	n_1	$\overline{X}_1$	S_1^2
집단 2	μ_2	σ_2^2	n_2	$\overline{X}_2$	S_2^2

그림 10-2 두 모집단과 표본



- $\overline{X_1} \sim N\left(\mu_1, \frac{\sigma_1^2}{n_1}\right)$
- $\overline{X_2} \sim N\left(\mu_2, \frac{\sigma_2^2}{n_2}\right)$

$$\Longrightarrow \overline{X_1} - \overline{X_2} \sim N\left(\mu_1 - \mu_2,\ \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

$$\Longrightarrow T(X) = \frac{\overline{X_1} - \overline{X_2} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

① 가설의 설정

(a) 양측검정 : H_0 : $\mu_1 = \mu_2, H_1$: $\mu_1 \neq \mu_2$

(b) 단측검정 : H_0 : $\mu_1 = \mu_2, H_1$: $\mu_1 > \mu_2$ (또는 $\mu_1 < \mu_2$)

② 귀무가설하에서의 검정통계량의 값과 분포

(a) σ_1^2, σ_2^2이 알려져 있는 경우

$$T(X) = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,\ 1)$$

(a) σ_1^2, σ_2^2이 알려져 있지 않는 경우

(i) $\sigma_1^2 = \sigma_2^2$인 경우

$\sigma_1^2 = \sigma_2^2 = \sigma^2$이라 할 때, S_1^2, S_2^2 모두 σ^2의 추정량이다.

σ^2의 추정치로 S_1^2와 S_2^2의 가중평균 S_p^2 사용.

$$T(X) = \frac{\overline{X_1} - \overline{X_2}}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}, \quad S_p = \sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}}$$

$$T(X) \sim N(0,1), \qquad n_1 + n_2 > 30\text{인 경우}$$

$$T(X) \sim t_{n_1+n_2-2}, \qquad n_1 + n_2 \le 30\text{인 경우}$$

(i) $\sigma_1^2 \neq \sigma_2^2$인 경우

$T(X) = \dfrac{\overline{X_1} - \overline{X_2}}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$ 이며 $T(X)$의 분포는 $n_1 + n_2 > 30$일 때, $T(X) \sim N(0,1)$이다.

③ 검정

- 검정통계량 $T(X)$의 분포에서 **가설의 종류(단측검정 또는 양측검정)와 유의수준 α에 의하여 기각역을 설정**한다.
- 귀무가설하에서의 검정통계량의 값 **$T(X)$가 기각역에 속하면 귀무가설을 기각**하고 **기각역에 속하지 않으면 귀무가설을 채택**한다.

예8-4.

서울시내 남녀 고등학생들의 학력수준에 차이가 있는가를 알아보기 위하여 남녀 고등학생 각각 100명씩을 임의로 선발한 뒤에 학력고사를 실시하여 다음과 같은 결과를 얻었다.

남학생 : $\overline{X_1} = 82$점, $S_1 = 20$점

여학생 : $\overline{X_2} = 78$점, $S_2 = 18$점

두 집단의 모분산이 동일하다고 할 때 남녀학생들의 학력고사 성적이 차이가 있는지를 검정하라.

(sol)

(1) 가설의 설정 : 남학생 집단의 평균성적을 μ_1, 여학생 집단의 평균성적을 μ_2라고 할 때, 두 집단의 평균성적이 동일한가에 대한 검정을 하고자 하므로 양측검정이라고 할 수 있다.

$$H_0 : \mu_1 = \mu_2 \quad v.s \quad H_1 : \mu_1 \neq \mu_2$$

(2) 검정통계량과 분포 : $\sigma_1{}^2 = \sigma_2{}^2$이므로 합동표본분산은 다음과 같이 계산된다.

$$S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2} = \frac{99\times400 + 99\times324}{100+100-2} = 362$$

$$S_p = \sqrt{S_p^2} = 19.026$$

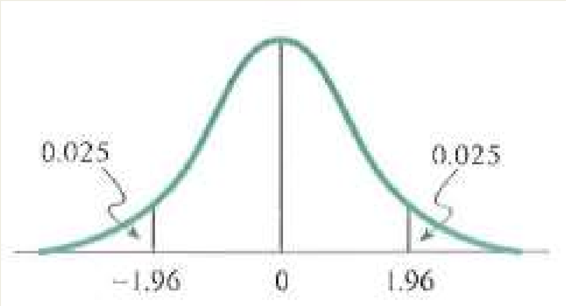
따라서 귀무가설하에서 검정통계량의 값은 다음과 같다.

$$T(X) = \frac{\overline{X}_1 - \overline{X}_2}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{82-78}{19.026\sqrt{\frac{1}{100}+\frac{1}{100}}}$$

$$= \frac{4}{2.6907} = 1.487$$

표본의 수는 $n_1 + n_2 = 200 > 30$이므로 $T(X) \sim N(0,1)$이라고 할 수 있다.

⑶ 검정 : 양측검정에서 유의수준 $\alpha = 0.05$에 대한 표준정규분포의 기각역은 아래 그림의 빗금 친 부분이고, $T(X) = 1.487$은 이 기각역에 속하지 않으므로 귀무가설은 기각되지 않는다.



즉, 위의 표본관측결과에 의할 때 남녀 고등학생들의 평균성적은 유의수준 5%하에서 동일하다고 할 수 있다.

[예 8. 4]에서 같이 표본의 크기가 동일한 경우($n_1 = n_2 = n$)에는 ${S_P}^2$과 $T(X)$를 다음과 같이 간단하게 구할 수 있다.

$$S_P^2 = \frac{1}{2}(S_1^2 + S_2^2)$$

$$T(X) = \frac{\overline{X}_1 - \overline{X}_2}{S_P\sqrt{\frac{2}{n}}}$$

예8-5.

두 가지 서로 다른 체중조절약의 체중조절효과를 분석하기 위하여 각각 40명씩 1년간 임상실험한 결과 다음과 같은 자료를 얻었다. 체중조절약 I의 체중조절효과가 체중조절약 II의 효과보다 더 크다고 볼 수 있는지를 5% 유의수준으로 검정하라.

체중조절약	표본크기	표본평균(체중평균감소량)	표본분산
I	40	10kg	4.3kg
II	40	8kg	5.7kg

(sol) 체중조절약 I 과 II의 체중조절효과의 평균과 분산을 각각 다음과 같다고 할 때, 검정을 실시한다.

체중조절약 I : $\mu_1, {\sigma_1}^2$, 체중조절약 II : $\mu_2, {\sigma_2}^2$

(1) 가설의 설정 : 체중조절약 I의 효과가 더 큰가(즉, μ_1이 μ_2보다 더 큰가)를 알고자 하므로 단측검정이며 H_0와 H_1은 각각 다음과 같이 설정된다.

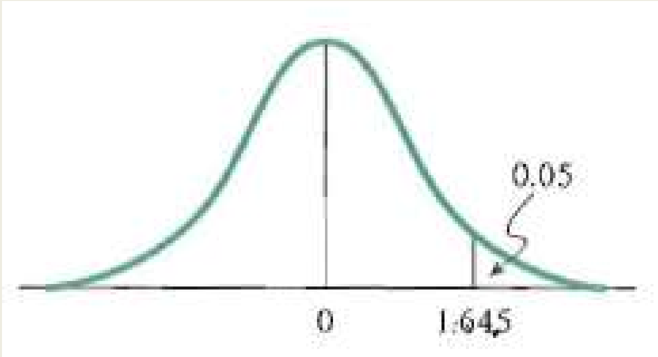
$$H_0 : \mu_1 = \mu_2 \;\; v.s \;\; H_1 : \mu_1 > \mu_2$$

(2) 검정통계량과 분포 : $\sigma_1{}^2 = \sigma_2{}^2$ 이라는 조건이 주어지지 않았으므로 귀무가설 하에서의 검정통계량은 다음과 같다.

$$T(X) = \frac{X_1 - X_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} = \frac{10 - 8}{\sqrt{\frac{4.3}{40} + \frac{5.7}{40}}} = \frac{2}{0.5} = 4.0$$

표본크기 $n_1 + n_2 = 80(> 30)$이 충분히 크다고 볼 수 있으므로 $T(X) \sim N(0, 1)$을 따른다고 할 수 있다.

(3) 검정 : 양측검정에서 유의수준 $\alpha = 0.05$에 대한 표준정규분포의 기각역은 아래 그림의 빗금 친 부분이고, $T(X) = 4$은 이 기각역에 속하지 않으므로 귀무가설은 기각된다.



즉, 체중조절약 I의 효과가 더 크다고 볼 수 있다.

예8-6.

두 개의 서로 다른 자동차 제동장치의 성능을 비교하기 위하여 각 제동장치를 10대의 동일모형 자동차에 장착한 후에 $80km/hr$로 달리다가 제동을 하는 데 필요한 거리(m)를 관측한 결과가 다음과 같다. 두 제동장치의 제동거리의 분포는 동일한 분산 σ^2을 가지며 정규분포를 따른다고 할 때, 두 제동장치의 성능이 동일한가에 대한 검정을 유의수준 5%하에서 실시하라.

제동장치 1	10.2	10.5	10.3	10.8	9.8	10.6	10.7	10.2	10.0	10.6
제동장치 2	9.8	9.6	10.1	10.2	9.7	9.5	9.6	10.1	9.8	9.9

(sol)

(1) 가설의 설정 : 두 제동장치의 성능이 동일한가를 알고자 하므로 양측검정이며
H_0와 H_1은 각각 다음과 같이 설정한다.

$$H_0 : \mu_1 = \mu_2 \ v.s \ H_1 : \mu_1 \neq \mu_2$$

(2) 검정통계량과 분포 : 두 집단으로부터의 표본에 의하여 다음을 구할 수 있다.

$$n_1 = 10,\ \overline{X}_1 = 10.37,\ S_1^2 = 0.105$$
$$n_2 = 10,\ \overline{X}_2 = 9.83,\ S_2^2 = 0.058$$

표본크기가 $n_1 = n_2 = 10$이고, 두 집단의 모분산이 동일하므로 합동표본분산과 검정통계량은 다음과 같이 구할 수 있다.

$$S_P^2 = \frac{S_1^2 + S_2^2}{2} = \frac{0.105 + 0.058}{2} = 0.0815$$
$$S_P = \sqrt{S_P^2} = 0.285$$

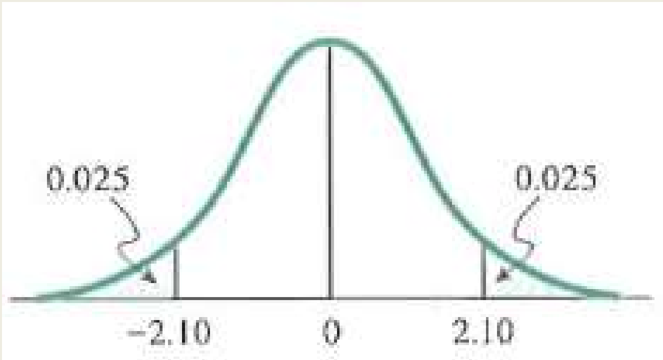
$$T(X) = \frac{\overline{X}_1 - \overline{X}_2}{S_P\sqrt{\frac{2}{n}}} = \frac{10.37 - 9.83}{0.285\sqrt{\frac{2}{10}}} = 4.24$$

$T(X)$는 자유도가 $n_1 + n_2 - 2 = 18$인 t-분포를 따른다.

⑶ 검정 : 유의수준 5%하에서 양측검정에 대한 $t_{(18)}$ 분포의 기각역은 부록V의 [표 4]에 의하여 다음 그림의 빗금친 부분과 같다.

$T(X) = 4.24$가 기각역에 속하므로 귀무가설은 기각된다.

즉, 두 제동장치의 성능이 동일하다고 볼 수 없다.



8.4절

8.4 짝진표본의 모평균에 대한 검정

- 성장환경이 어린이의 지능발달에 영향을 미치는가를 조사하고자 할 때,

- 우선 생각할 수 있는 검정방법은 성장환경이 좋은 지역의 어린이 n_1명과 성장환경이 좋지 않은 지역의 어린이 n_2명을 임의로 선발해 동일한 지능검사를 실시한 후 지능검사 결과가 차이가 있는가를 분석해 검정 실시한다.
- 그러나, 이 경우에 지능검사에 미치는 요인이 환경이 아니라 유전적 요인일 수 있음을 배제하기 어렵다.
- 이 문제에 대한 바람직한 접근 방법은 일란성 쌍둥이 n쌍을 선택하여 일정기한 동안 각 쌍둥이를 하나는 좋은 환경에 다른 하나는 좋지 않은 환경에 두는 것이다.

짝진표본의 모평균에 대한 검정

- n개의 쌍 $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$으로 관측된 표본에서, 관찰값의 차이를 $d_i = x_i - y_i$, $i = 1, 2, \dots, n$이라고 할 때, 두 집단 평균 μ_X와 μ_Y의 동일성에 대한 검정은 $\mu_D = \mu_X - \mu_Y$이므로 다음과 같이 실시한다.

표 10-3 짝진표본의 관찰값

쌍번호	1	2	3	4	⋯	n
X	x_1	x_2	x_3	x_4	⋯	x_n
Y	y_1	y_2	y_3	y_4	⋯	y_n
D	d_1	d_2	d_3	d_4	⋯	d_n

① 가설 설정

(a) 양측검정 : $H_0 : \mu_D = 0, \quad H_1 : \mu_D \neq 0$

(b) 단측검정 : $H_0 : \mu_D = \mu_2, \quad H_1 : \mu_D > 0$(또는 $\mu_D < 0$)

② 검정통계량과 분포

두 집단 차이의 평균과 분산을 다음과 같다고 할 때, 귀무가설 검정통계량은 다음과 같다.

$$\bar{d} = \frac{1}{n}\sum_{i=1}^{n} d_i \ , \ S_d^{\ 2} = \frac{1}{n-1}\sum_{i=1}^{n}(d_i - \bar{d})^2 \qquad T(X) = \frac{\bar{d}}{S_d\sqrt{n}}$$

검정통계량 $T(X)$의 분포는 n에 따라 다음과 같이 나타난다.

$T(X) \sim N(0,\ 1),\ \ n > 30$인 경우,

$T(X) \sim t_{n-1}, \qquad n \le 30$인 경우이다.

③ 검정

검정통계량 $T(X)$의 분포에서 **가설의 종류(단측검정 또는 양측검정)와 유의수준 α에 의하여 기각역을 설정**한다.

귀무가설하에서의 검정통계량의 값 $\boldsymbol{T(X)}$가 **기각역에 속하면 귀무가설을 기각**하고 **기각역에 속하지 않으면 귀무가설을 채택**한다.

예8-7.

- **두 종류의 타이어(A, B)의 성능을 비교하기 위하여 5대의 자동차를 임의로 선발하여 각 자동차의 뒷바퀴 중에서 한 쪽에는 A타이어를 그리고 다른 한쪽에는 B타이어를 끼우고 $500km$를 주행한 뒤에 타이어의 마모상태를 조사한 결과가 주어져 있다. 두 타이어의 마모율에 차이가 있는가를 유의수준 5%하에서 검정하라.**

표 10-4 두 타이어의 마모율

자동차	타이어 A	타이어 B	$D=A-B$
1	10.6	10.2	0.4
2	9.4	9.8	−0.4
3	12.3	11.8	0.5
4	9.7	9.1	0.6
5	8.3	8.8	−0.5

(sol)

(1) 가설의 설정 : 두 타이어의 마모율이 동일한가에 대한 검정이므로 A, B, D의 모평균을 각각 μ_A, μ_B, μ_D라고 할 때 귀무가설과 대립가설은 각각 다음과 같이 설정할 수 있다.

$$H_0 : \mu_D = 0 \ (즉, \mu_A = \mu_B)$$

$$H_1 : \mu_D \neq 0 \ (즉, \mu_A \neq \mu_B)$$

(2) 검정통계량과 분포 : 자료에 의하여 D의 표본평균과 표본표준편차는 각각 다음과 같다.

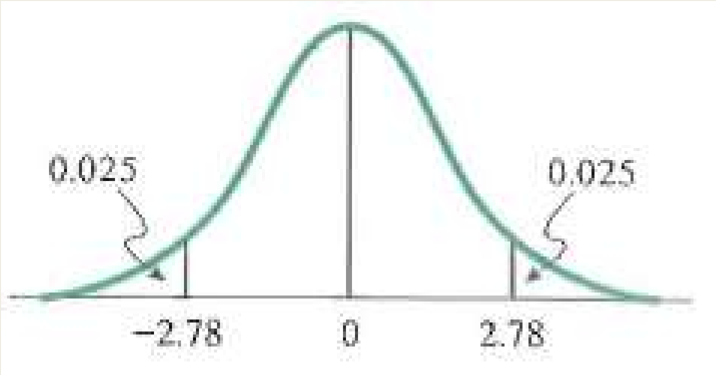
$$\bar{d} = 0.12, \qquad S_d = 0.5263$$

귀무가설 하에서의 검정통계량의 값은 다음과 같다.

$$T(X) = \frac{\bar{d}}{S_d/\sqrt{n}} = \frac{0.12}{0.5263/\sqrt{5}} = 0.510$$

쌍의 크기 $n = 5 \leq 30$이므로 $T(X)$는 자유도가 $n - 1 = 4$인 t-분포를 따른다.

(3) 검정 : 유의수준 5%하에서 자유도가 4인 t-분포의 양측검정에 대한 기각역은 부록V의 [표 4]에 의하여 아래 그림의 빗금친 부분과 같다. 귀무가설 하에서의 검정통계량의 값 $T(X) = 0.510$ 은 기각역에 속하지 않으므로 귀무가설은 기각되지 않는다. 즉, 5% 유의수준 하에서 두 타이어의 성능은 동일하다고 할 수 있다.



예8-8.

- 8명의 사람이 한 다이어트 프로그램에 한 달 동안 참석하였는데, 처음 시작할 때와 한 달 후 다이어트 프로그램이 끝났을 때 측정한 체중이 다음 표와 같다. 다이어트 프로그램이 체중을 줄이는 효과가 있는지를 5% 유의수준 하에서 검정하라.

표 10-5 다이어트 프로그램 참가자의 체중값

사람	처음 체중(A)	한 달 후 체중(B)	$D=A-B$
1	110	95	15
2	95	91	4
3	87	85	2
4	105	98	7
5	70	68	2
6	123	110	13
7	77	74	3
8	99	95	4

(sol)

(1) 가설의 설정 : 체중이 다이어트 프로그램이 끝난 후에 줄어들었는지의 여부에 관심이 있으므로 귀무가설과 대립가설은 각각 다음과 같이 설정할 수 있다.

$$H_0 : \mu_D = 0 \ (즉, \mu_A = \mu_B)$$

$$H_1 : \mu_D > 0 \ (즉, \mu_A > \mu_B)$$

(2) 검정통계량과 분포 : 주어진 자료에 의하여 D의 표본평균과 표본 표준편차는 각각 다음과 같이 계산된다.

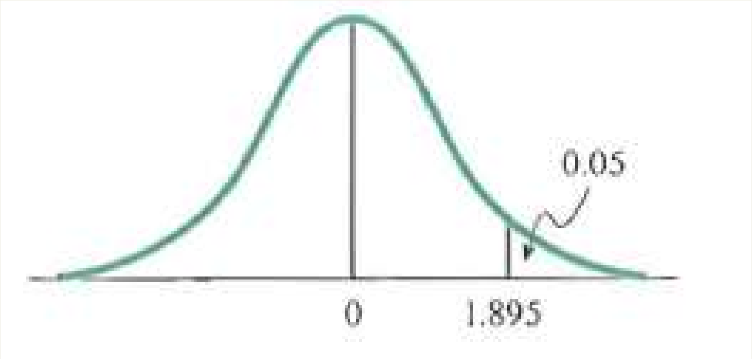
$$\bar{d} = 6.25, \qquad S_d = 5.06$$

따라서 귀무가설 하에서의 검정통계량의 값은 다음과 같이 계산된다.

$$T(X) = \frac{\bar{d}}{S_d/\sqrt{n}} = \frac{6.25}{5.06/\sqrt{8}} = 3.491$$

표본크기 $n = 8 < 30$이므로 $T(X)$는 자유도가 $n - 1 = 7$인 t-분포를 따른다.

⑶ 검정 : 5% 유의수준 하에서 자유도가 7 인 t-분포의 오른쪽 단측검정에 대한 기각역은 부록 V의 [표 4]에 의하여 아래 그림의 빗금친 부분과 같다.



귀무가설 하에서의 검정통계량 값 $T(X) = 3.901$이 기각역에 속하므로 귀무가설은 기각된다. 즉, 다이어트 프로그램이 체중을 줄이는 효과가 있다고 할 수 있다.

8.5절

8.5 모비율 P에 대한 검정

단일 모비율 P의 검정

- 단일 모집단의 특성의 비율 P에 대한 검정에서 n개의 표본을 관찰한 결과 모집단의 특성을 만족하는 경우가 X개라고 할 때 모비율 P가 특성값 P_0과 같은가에 대한 검정이다.

① 가설의 설정

(a) 양측검정 : $H_0 : P = P_0,\ H_1 : P \neq P_0$

(b) 단측검정 : $H_0 : P = P_0,\ H_1 : P > P_0$(또는 $P < P_0$)

② 귀무가설하에서의 검정통계량의 값과 분포

$\hat{P} = \frac{X}{n}$라 할 때 $\hat{P} \sim N\left(P, \frac{P(1-P)}{n}\right)$ 귀무가설 하에서 검정통계량의 값과 분포는 다음과 같다.

$$T(X) = \frac{\hat{P} - P_0}{\sqrt{\frac{P_0(1-P_0)}{n}}} \sim N(0,1)$$

③ 검정

검정통계량 $T(X)$의 분포에서 **가설의 종류(단측검정 또는 양측검정)와 유의수준 α에 의하여 기각역을 설정**한다.

귀무가설 하에서의 검정통계량의 값 $T(X)$가 **기각역에 속하면 귀무가설을 기각**하고 **기각역에 속하지 않으면 귀무가설을 채택**한다.

예8-9.

한 전구회사에서는 자기 회사 제품의 불량률이 5% 이하라고 주장하였다. 그 주장이 사실인가를 알아보기 위하여 그 회사 제품의 전구 300개를 임의로 선택하여 조사한 결과 10개가 불량품임을 발견하였다. 그 회사의 주장이 타당하다고 볼 수 있는가를 유의수준 5%하에서 검정하라.

(sol)

⑴ 가설의 설정 : 모비율 P가 0.05 이하인 경우에 관심이 있으므로 귀무가설과 대립가설은 각각 다음과 같이 설정한다.

$$H_0 : P = 0.05$$

$$H_1 : P < 0.05$$

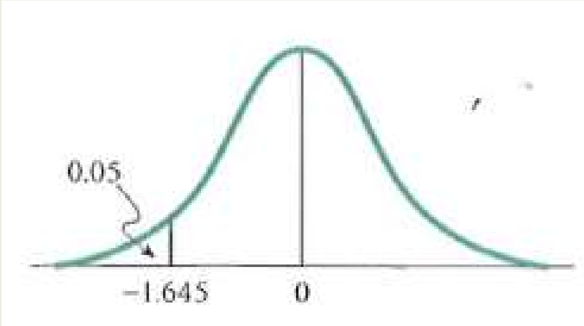
⑵ 검정통계량과 분포 : 모비율 P의 추정량의 값과 귀무가설 하에서의 검정통계량의 값은 다음과 같이 계산된다.

$$\hat{P} = \frac{10}{300} = \frac{1}{30} = 0.033$$

$$T(X) = \frac{\hat{P} - P_0}{\sqrt{\frac{P_0(1-P_0)}{n}}} = \frac{0.033 - 0.05}{\sqrt{\frac{0.05 \times 0.95}{300}}} = -1.35$$

$T(X)$는 근사적으로 $N(0, 1)$을 따른다.

(3) 검정 : 유의수준 $\alpha = 0.05$일 때 왼쪽단측검정에 대한 $N(0,1)$에서의 기각역은 부록 Ⅲ의 [표 3]에 의하여 아래 그림의 빗금친 부분과 같다.



검정통계량의 값 $T(X) = -1.35$가 기각역에 속하지 않으므로 귀무가설은 기각되지 않는다. 즉, 위에 주어진 표본관측결과에 의할 때 그 회사제품의 불량률이 5%보다 적다고 할 수 없다.

예8-10.

최근에 한 조사기관에서는 대학졸업자 중에서 약 20%가 자신의 전공을 살릴 수 있는 직장에 입사한다고 발표하였다. 이 발표가 사실인가를 알아보기 위하여 400명의 대학졸업자를 임의로 선발하여 조사한 결과 100명이 전공을 살릴 수 있는 직장에서 일하고 있음을 알 수 있었다. 조사기관의 발표가 타당한가를 5% 유의수준하에서 검정하라.

(sol)

(1) 가설의 설정 : 모비율 P가 20%인가 또는 그와 상당히 다른가에 대하여 관심이 있으므로 H_0와 H_1은 각각 다음과 같이 설정할 수 있다.

$$H_0 : P = 0.20$$

$$H_1 : P \neq 0.20$$

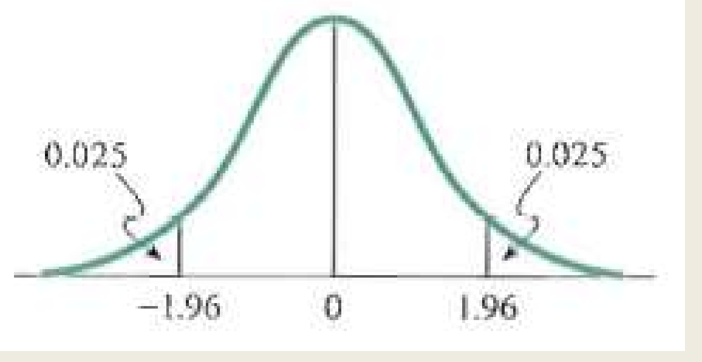
⑵ 검정통계량과 분포 : 표본비율과 귀무가설 하에서의 검정통계량 값은 다음과 같이 계산된다.

$$\hat{P} = \frac{100}{400} = 0.25$$

$$T(X) = \frac{\hat{P} - P_0}{\sqrt{\frac{P_0(1 - P_0)}{n}}} = \frac{0.25 - 0.20}{\sqrt{\frac{0.20 \times 0.80}{400}}} = 2.5$$

$T(X)$은 근사적으로 $N(0,1)$을 따른다.

⑶ 검정 : 유의수준 5%하에서 양측검정에 대한 $N(0,1)$의 기각역은 부록 V의 [표 3]에 의하여 아래 그림의 빗금친 부분과 같다.



귀무가설 하에서의 검정통계량의 값 $T(X) = 2.5$가 기각역에 속하므로 귀무가설은 기각된다. 즉 위의 표본관측결과에 의할 때 대학졸업생 중에서 20%가 자기 전공을 살릴 수 있는 직장에 취직한다는 주장을 받아들일 수 없다.

끝~~ ♥♥