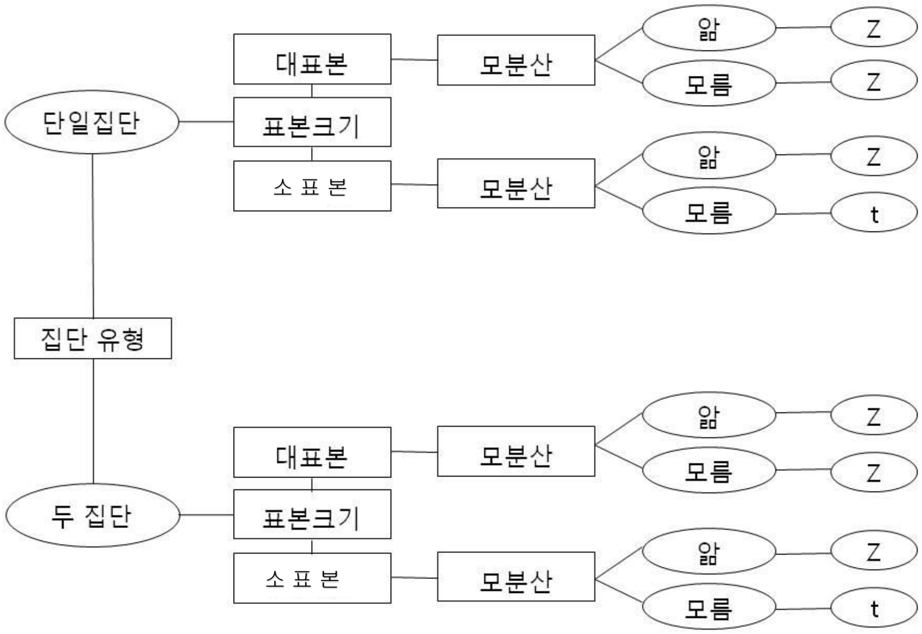
12-2 가설검정 2

12-2-1. 단일집단의 모평균에 대한 검정

모평균 가설검정을 위한 의사결정 트리



단일모평균 μ의 검정

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 $\mu < \mu_0$)

2. 귀무가설하에서의 검정통계량과 분포

(a) σ^2 을 아는 경우

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

예

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 조사해보니 표본평균이 1.3kg이었다.

"1일 1인당 생활폐기물량은 1kg이다"라는 귀무가설과
"1일 1인당 생활폐기물량은 1kg가 아니다."라는 대립가설을 세우고
5%의 유의수준에서 가설검정을 하라.

풀이

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

$$H_0 : \mu_0 = 1, H_1 : \mu_0 \neq 1$$

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.

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_0 을 기각하는 결정규칙에 의해 H_0 를 기각한다.

검정통계량 2.6833은 임계값 1.96보다 크므로 귀무가설을 기각한다.
즉, 5% 유의수준 하에서 도시민들의 1일 1인당 평균 생활폐기물량이 1kg은 아니라고 할 수 있다.

(b) σ^2 을 모르는 경우

(i) $n > 30$일 때

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

(ii) $n \le 30$일 때

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

예

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 0.05의 유의수준에서 평균 배당률이 10%가 아니라는 대립가설로 양측검정을 하라.

풀이

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

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

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

0.05의 유의수준에서 임계값은 $-z_{0.025}=-1.96$이다.

검정통계량(-1.4757)이 임계값(-1.96)보다 크므로

H_0 을 허용하는 결정규칙에 의해 귀무가설은 허용된다.

금융기관들의 평균 배당률이 10%가 아니라는 통계적 증거는 제시되지 못한다

예제

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.

그림 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 5%유의수준에서 양측검정하라.

풀이

검정하려는 귀무가설과 대립가설은 $H_0 : \mu = 34.5,\ H_1 : \mu \neq 34.5$
이며 주어진 정보와 표본자료에 대한 계산결과에 의하면
$\bar{x} = 33.798,\ S = 0.630,\ \mu_0 = 34.5,\ n = 15$ 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.

5% 유의수준에서 $\alpha = 0.05$이며 양측검정의 경우 $\frac{\alpha}{2} = 0.025$이고
임계값은 $-t_{(14,0.025)} = -2.145$이다.

$(\bar{x}-\mu_0)/(S/\sqrt{n}) = -4.314 <$ -2.145이므로 결정규칙에 의해 H_0 를 기각하게 된다.

따라서 체인점들의 평균 판매량 감소가 34.5%라는 주장은 부당하다고
결론지을 수 있다.

단일집단의 모분산 σ^2 에 대한 검정

1.가설의 설정

(a)양측검정 : $H_0 : \sigma^2 = \sigma_0^2,\ H_1 : \sigma^2 \neq \sigma_0^2$

(b)단측검정 : $H_0 : \sigma^2 = \sigma_0^2,\ H_1 : \sigma^2 > \sigma_0^2$(또는 $\sigma^2 < \sigma_0^2$)

2.검정통계량과 분포

$$T(X) = \frac{(n-1)S^2}{\sigma_0^2},\quad T(X) \sim \chi^2_{(n-1)}$$

3.검정

$\chi^2_{(n-1)}$에서 검정의 종류(단측검정 또는 양측검정)와 유의수준 α에 따라 기각역을 설정하고 T(X)가 기각역에 속하면 귀무가설을 기각

예

한 기술연구소에서 휴대전화 배터리 무게의 분산이 62g이라는 주장에 대한 양측검정을 하려고 한다.
휴대전화 7개를 무작위 선정하여 조사한 무게가 다음과 같으며, 휴대전화 배터리 무게는 정규분포한다고 하자.

표 휴대전화 배터리 무게(단위:g)

36	37	38	39	39	44	47

5%의 유의수준에서 휴대전화 배터리 무게의 분산이 62g이라는 귀무가설을 양측검정하라.

풀이

귀무가설과 대립가설을 설정하면 $H_0 : \sigma^2=62,\ H_1 : \sigma^2 \neq 62$

주어진 표본자료로부터

$$\bar{x} = \sum x/n = 280/7 = 40$$

$$S^2 = \sum (x_i - \bar{x})^2/(n-1) = 96/6 = 16$$

그리고 n=7이다. 이를 이용하여 검정통계량을 계산하면

$$\chi^2_{(n-1)} = \frac{(n-1)S^2}{\sigma_0^2} = \frac{6\times 16}{62} = 1.55$$

이다.

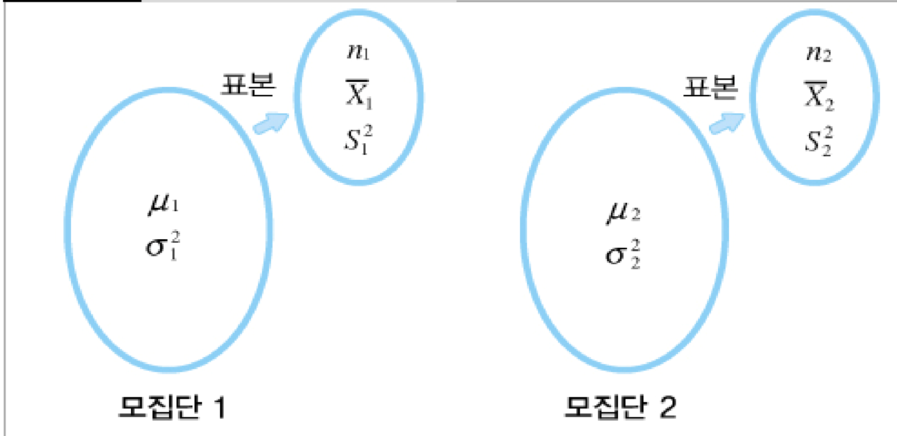
0.05의 유의수준에서 $\alpha/2=0.025$ 이고 자유도는 6이므로
임계값은 $\chi^2_{(n-1,\,1-\alpha/2)}=\chi^2_{(6,\,0.975)}=1.24$ 와 $\chi^2_{(n-1,\,\alpha/2)}=\chi^2_{(6,\,0.025)}=14.45$ 이다.
검정통계량 1.55가 임계값 1.24와 14.45사이에 놓이므로 결정규칙에 따라 H_0 을 채택한다. 따라서 휴대전화 배터리 무게의 분산은 62g이라고 할 수 있다.

12-2-3. 두 집단 모평균의 동일성에 대한 검정

표 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 두 모집단과 표본



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

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

(b) σ_1^2, σ_2^2이 알려져 있지 않는 경우

$(i)\ \sigma_1^2 = \sigma_2^2$인 경우

$$T(X) = \frac{\bar{X}_1 - \bar{X}_2}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \qquad S_p = \sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}}$$

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

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

예

대기업과 중소기업 근로자의 직장 만족도

표본 1 (중소기업 근로자, $n_1=10$)									
41	45	42	62	68	54	52	55	44	60
표본 2 (대기업 근로자, $n_2=17$)									
74 83	74 50	70 52	52 66	76 65	91 53	71 72	78	76	78

정규분포하는 두 모집단의 분산은 동일하다고 가정하자.

- 중소기업 근로자들의 평균 직장 만족도가 대기업 근로자와 같다는 귀무가설을 설정하고 5%유의수준에서 양측검정하라.

풀이

주어진 자료로부터 표본 통계량을 계산하면

$$\bar{X}_1 = 52.3, \quad S_1 = 9.23, \quad n_1 = 10$$

$$\bar{X}_2 = 69.471, \quad S_2 = 11.78, \quad n_2 = 17$$

$$S_p^2 = \frac{(9 \times 9.23^2) + (16 \times 11.78^2)}{10 + 17 - 2} = 119.48$$

$S_p = 10.9307$

풀이

중소기업 근로자들의 평균 직장 만족도가 대기업 근로자와 같다는 것은 두 모집단 평균 차 $\mu_1 - \mu_2$ 가 0이라는 것을 의미하므로 귀무가설과 대립가설은 $H_0: \mu_1 - \mu_2 = 0, H_1: \mu_1 - \mu_2 \neq 0$이 되며,

검정통계량은 $t_{25} = \dfrac{52.3 - 69.471}{10.9307 \times \sqrt{\dfrac{10+17}{10 \times 17}}} = -3.9418$ 이다.

임계값은 $t_{(25,0.025)} = -2.06$이므로 결정규칙에 의해 H_0 는 기각되며, 대기업 근로자들의 평균 직장 만족도와 중소기업 근로자들의 직장 만족도가 같지 않다고 할 수 있다.

(ii) $\sigma_1^{\,2} \neq \sigma_2^{\,2}$인 경우

$T(X) = \dfrac{\bar{X}_1 - \bar{X}_2}{\sqrt{\dfrac{S_1^{\,2}}{n_1} + \dfrac{S_2^{\,2}}{n_2}}}$ 이며, $T(X)$의 분포는 $n_1 + n_2 > 30$일 때

$T(X) \sim N(0,\, 1)$이다.

12-2-4. 짝진표본의 모평균에 대한 검정

표 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

N개의 짝진표본 관찰값이 주어져 있다고 할 때, 각 쌍에서 두 값의 차이 $d_i=x_i-y_i$, $i=1, 2, \ldots, n$을 구할 수 있음

n개의 쌍 (x_1,y_1), (x_2,y_2), …, (x_n,y_n)으로 관측된 표본에서 , $d_i=x_i-y_i$, $i=1, 2, \ldots, n$ 이라고 할 때 두 집단 평균 μ_X , μ_Y 의 동일성에 대한 검정은 $\mu_D = \mu_X - \mu_Y$ 이므로 다음과 같이 실시

1. 가설의 설정

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

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

2. 검정통계량과 분포

$$\bar{d} = \frac{1}{n}\sum_{i=1}^{n} d_i ,$$

$$S_d^2 = \frac{1}{n-1}\sum_{i=1}^{n}(d_i - \bar{d})^2$$

이라고 할 때, 귀무가설 하에서 검정통계량의 값은

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

이고, T(X)의 분포는 $T(X) \sim N(0,\ 1),\ n > 30$인 경우,
$T(X) \sim t_{n-1},\quad n \le 30$인 경우이다.

3. 검정

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

귀무가설하에서의 검정통계량의 값 T(X)가 기각역에 속하면 귀무가설을 기각하고 기각역에 속하지 않으면 귀무가설을 채택

예

통계학을 수강하는 경영학부 학생들을 대상으로 보충수업이 학생들에게 도움이 되는지 알아보기 위해 6명을 임의로 선정하였다.
보충수업 전에 시험을 보게 하고 보충수업을 수강한 후 다시 시험을 보게 하였으며, 그 결과는 다음과 같다고 가정하자.

표 보충수업 전과 후의 점수

학생	보충수업 전 (X_1)	보충수업 후 (X_2)	점수 차이$(d = X_1 - X_2)$
1	75	82	−7
2	71	73	−2
3	52	59	−7
4	46	48	−2
5	70	69	1
6	83	93	−10

보충수업이 학생들의 성적 향상에 도움이 되는지 5%유의수준에서 검정하라.

풀이

보충수업이 학생들의 성적을 향상시켰는지 여부에 관심이 있으므로 다음과 같이 가설 설정을 할 수 있다.

$$H_0 : \mu_1 - \mu_2 = 0, \ \ H_1 : \mu_1 - \mu_2 < 0$$

주어진 자료에 의하면

$$\bar{d} = -4.5, \quad S_d = 4.1352, \quad n = 6$$

귀무가설 $H_0 : \mu_1 - \mu_2 = 0$ 하에서 검정통계량의 값은

$$\frac{\bar{d}}{S_d/\sqrt{n}} = \frac{-4.5}{4.1352/\sqrt{6}} = -2.665$$

이다.

5%유의수준하에서 단측검정을 위한 임계값은 $-t_{(5,\ 0.05)} = -2.015$ 이므로 H_0는 기각된다. 검정결과 보충수업이 학생들의 시험성적을 향상시키는 효과가 있다고 할 수 있다.

두 모분산의 동일성 검정

1.가설의 설정

(a)양측검정 : $H_0 : \sigma_1^2 = \sigma_2^2,\ H_1 : \sigma_1^2 \neq \sigma_2^2$

(b)단측검정 : $H_0 : \sigma_1^2 = \sigma_2^2,\ H_1 : \sigma_1^2 > \sigma_2^2$(또는 $\sigma_1^2 < \sigma_2^2$)

2.검정통계량과 분포

$$F = \frac{S_1^2}{S_2^2} \quad F_{(n_1-1,\ n_2-1)}$$

3.검정

주어진 유의수준에서 기각역을 설정한 후
검정통계량이 기각역에 속하면 귀무가설을 기각

예

신용 평가 기관인 스탠더드-푸어(Standard and Poor's corp.)에서
우리 금융기관에 대해 평가한 신용등급에 따라 금융기관을 두 그룹으로 나누고
각 그룹의 주식에 대한 수익률을 서로 비교하였다. **높은** 신용 평가를 받은
첫 번째 그룹에서 7개의 주식 수익률을, **낮은** 신용 평가를 받은
두 번째 그룹에서 11개의 주식 수익률을 각각 독립적으로 선정하였다.
그리고 표본분산 $S_1^2=83.35$, $S_2^2=48.02$ 를 얻었다.

높은 신용 평가를 받은 그룹의 주식 수익률의 분산이 낮은 신용 평가를 받은 그룹의 주식 수익률의 분산보다 큰지 5%유의수준에서 검정하라.

풀이

높은 신용 평가를 받은 그룹의 주식 수익률의 분산이 낮은 신용 평가를 받은 그룹의 주식 수익률의 분산보다 큰 경우 $\sigma_1^2/\sigma_2^2>1$ 이므로 귀무가설과 대립가설은 다음과 같이 설정될 수 있다.

$$H_0 : \frac{\sigma_1^2}{\sigma_2^2}=1, \quad H_1 : \frac{\sigma_1^2}{\sigma_2^2}>1$$

$F_{(6,\ 10,\ 0.01)}$

검정통계량의 값은 $\frac{S_1^2}{S_2^2}=1.7357$ 이며

유의수준 5%에서 임계값 $F_{(v_1,\ v_2,\ \alpha)}$ 는 $F_{(6,10,0.05)} = 3.22$이다. 결정규칙에 따라 귀무가설이 채택되며, 높은 신용 평가를 받은 그룹의 주식 수익률의 분산이 낮은 신용 평가를 받은 그룹의 주식 수익률의 분산보다 크다는 통계적 증거는 없다.