제 8장 두 집단 비교

8.1 독립된 두 집단 분산 비교

8.2 독립된 두 집단 평균차 추정

8.3 독립된 두 집단 평균차 검정

8.4 대응비교

❖ 자료구조 및 모형가정

▪ 자료구조

모집단 1 ⇨ $X_{11}, X_{12}, \cdots \cdots, X_{1n_1}$

모집단 2 ⇨ $X_{21}, X_{22}, \cdots \cdots, X_{2n_2}$

▪ 모형가정

- 각 그룹에서의 관측값은 각 모집단에서의 랜덤표본.
- 즉, 서로 다른 그룹에서의 관측값들은 독립적으로 관측된 것.

표본 : $\left.\begin{aligned} &X_{11}, X_{12}, \cdots \cdots, X_{1n_1} \sim iid\ N\ (\mu_1, \sigma_1^2) \\ &X_{21}, X_{22}, \cdots \cdots, X_{2n_2} \sim iid\ N\ (\mu_2, \sigma_2^2) \end{aligned}\right\}$ 서로 독립

❖ 자료의 요약

- **모집단 1의 표본 $\boldsymbol{X_{11}, X_{12}, \dots, X_{1n_1}}$에 대한 자료의 요약**
  - 표본의 크기 : n_1
  - 표본평균 : $\bar{X}_1 = \frac{1}{n_1}\sum_{j=1}^{n_1} X_{1j}$
  - 표본분산 : $S_1^2 = \frac{1}{n_1-1}\sum_{j=1}^{n_1}\left(X_{1j} - \bar{X}_1\right)^2$

❖ 자료의 요약

- **모집단 2의 표본 $\boldsymbol{X_{21}, X_{22}, \dots, X_{2n_2}}$에 대한 자료의 요약**
  - 표본의 크기 : n_2
  - 표본평균 : $\bar{X}_2 = \frac{1}{n_2}\sum_{j=1}^{n_2} X_{2j}$
  - 표본분산 : $S_2^2 = \frac{1}{n_2-1}\sum_{j=1}^{n_2}\left(X_{2j} - \bar{X}_2\right)^2$

❖ 두 모평균의 차이 구간 추정 (σ_1^2, σ_2^2을 알 때)

- **추정량 :** $\widehat{\mu_1 - \mu_2} = \bar{X}_1 - \bar{X}_2$
- **표본분포 :**

$\bar{X}_1 \sim N\left(\mu_1, \frac{\sigma_1^2}{n_1}\right)$, $\bar{X}_2 \sim N\left(\mu_2, \frac{\sigma_2^2}{n_2}\right)$ 과 정규분포의 성질로부터

$$\bar{X}_1 - \bar{X}_2 \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

❖ 두 모평균의 차이 구간 추정 (σ_1^2, σ_2^2을 알 때)

- **구간추정:**

$$P\left(-z_{\alpha/2} \le \frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{\sqrt{\sigma_1^2/n_1+\sigma_2^2/n_2}} \le z_{\alpha/2}\right) = 1-\alpha \text{ 로부터}$$

$$\left[(\bar{X}_1-\bar{X}_2) - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}, (\bar{X}_1-\bar{X}_2) + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\right]$$

❖ 예제 8.2

- **부록 1의 자료를 이용하여 우리나라 남자 노인과 여자 노인의 평균 신장 차이를 추정하고자 한다. 만일 남녀 신장은 각각 정규분포를 따르고 남녀 신장의 모분산이 각각 20, 36으로 알려져 있을 때, 평균 신장차이를 신뢰수준 95%에서 추정하여라.**

- **풀이**

① 남자의 평균신장 $\overline{x_1} = 166.1$, 여자의 평균신장 $\overline{x_2} = 151.5$

② $\overline{X_1} - \overline{X_2}$ 의 표준오차 : $s.e.(\overline{X_1} - \overline{X_2}) = 1.59$

③ ∴남녀신장의 차이에 대한 95% 신뢰구간 $(11.48, 17.72)$

❖ 두 모평균의 차이 구간추정 (σ^2을 모를 때, 단 $\sigma^2 = \sigma_1^2 = \sigma_2^2$)

- **분산이 같을 때(등분산)와 분산이 다를 때(이분산) 통계량의 분포가 다르므로 각각 정리**
- **추정량** : $\widehat{\mu_1 - \mu_2} = \bar{X}_1 - \bar{X}_2$
- **표본분포 :**

$$\bar{X}_1 \sim N\left(\mu_1, \frac{\sigma^2}{n_1}\right), \bar{X}_2 \sim N\left(\mu_2, \frac{\sigma^2}{n_2}\right), \bar{X}_1 - \bar{X}_2 \sim N\left(\mu_1 - \mu_2, \frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}\right)$$

❖ 두 모평균의 차이 구간추정 (σ^2을 모를 때, 단 $\sigma^2 = \sigma_1^2 = \sigma_2^2$)

- **통계량**

$$T = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{S_p\sqrt{1/n_1 + 1/n_2}} \sim t(n_1 + n_2 - 2)$$

단, 합동표본분산

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} = \frac{\sum_{j=1}^{n_1}\left(X_{1j} - \bar{X}_1\right)^2 + \sum_{j=1}^{n_2}\left(X_{2j} - \bar{X}_2\right)^2}{n_1 + n_2 - 2}$$

❖ 두 모평균의 차이 구간추정 (σ^2을 모를 때, 단 $\sigma^2 = \sigma_1^2 = \sigma_2^2$)

- <u>소표본에서</u> $\mu_1 - \mu_2$의 $100(1-\alpha)\%$ 신뢰구간 $\sim t(n_1 + n_2 - 2)$

$$P\left(-t_{\alpha/2}(n_1+n_2-2) \le \frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{S_p\sqrt{1/n_1+1/n_2}} \le t_{\alpha/2}(n_1+n_2-2)\right) = 1-\alpha \text{ 이므로}$$

$$\left[(\bar{X}_1-\bar{X}_2) - t_{\alpha/2}(n_1+n_2-2)S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}},\ (\bar{X}_1-\bar{X}_2) + t_{\alpha/2}(n_1+n_2-2)S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right]$$

- <u>대표본에서</u> $\mu_1 - \mu_2$의 $100(1-\alpha)\%$ 신뢰구간

$$\left[(\bar{X}_1-\bar{X}_2) - z_{\alpha/2}S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}},\ (\bar{X}_1-\bar{X}_2) + z_{\alpha/2}S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right]$$

❖ 예제 8.3

- 부록 1의 자료를 이용하여 우리나라 남자 노인과 여자 노인의 평균 신장 차이를 추정하고자 한다. 만일 남녀 신장은 각각 정규분포를 따르고 남녀 신장의 모분산이 같다는 사실이 알려져 있을 때, 신장의 차이를 신뢰수준 95%에서 추정하여라.

❖ 예제 8.3

- **풀이**

① 남자의 평균신장 $\overline{x_1} = 166.1$, 여자의 평균신장 $\overline{x_2} = 151.5$

② 합동분산 추정량 ${s_p}^2 = 26.35$

③ $\overline{X_1} - \overline{X_2}$ 의 표준오차의 추정값 : $\widehat{s.e}\,(\overline{X_1} - \overline{X_2}) = 1.47$

④ $t_{0.025}(48) = 2.01$을 이용

⑤ ∴남녀 신장의 차이에 대한 95%신뢰구간 (11.64,17.55)

❖ 예제 8.4

- **예제 8.3의 상황에서 두 표본의 크기가 충분히 크다고 간주할 수 있을 때, 정규분포를 이용하여 신뢰구간을 구하여라.**

- **풀이**

① 남자의 평균신장 $\overline{x_1} = 166.1$, 여자의 평균신장 $\overline{x_2} = 151.5$

② 합동분산추정량 ${s_p}^2 = 26.35$

③ $\overline{X_1} - \overline{X_2}$ 의 표준오차의 추정값 : $\widehat{s.e}\,(\overline{X_1} - \overline{X_2}) = 1.47$

④ $z_{0.025} = 1.96$을 이용

⑤ ∴남녀 신장의 차이에 대한 95%신뢰구간 (11.72,17.48)

❖ 두 모평균의 차이 구간추정 (σ_1^2 , σ_2^2 을 모를 때)

- **분산이 다를 때(이분산) 통계량의 분포는 정확한 t-분포를 따르는 것은 아니며 근사적으로 t-분포를 따른다.**
- **추정량** : $\widehat{\mu_1 - \mu_2} = \bar{X}_1 - \bar{X}_2$
- **표본분포**

$$\bar{X}_1 \sim N\left(\mu_1, \frac{{\sigma_1}^2}{n_1}\right), \bar{X}_2 \sim N\left(\mu_2, \frac{{\sigma_2}^2}{n_2}\right), \bar{X}_1 - \bar{X}_2 \sim N\left(\mu_1 - \mu_2, \frac{{\sigma_1}^2}{n_1} + \frac{{\sigma_2}^2}{n_2}\right)$$

❖ 두 모평균의 차이 구간추정 (σ_1^2 , σ_2^2 을 모를 때)

- **통계량**

$$T = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{S_1^2/n_1 + S_2^2/n_2}} \sim t(df), \quad df = \frac{\left(S_1^2/n_1 + S_2^2/n_2\right)^2}{\frac{\left(S_1^2/n_1\right)^2}{n_1 - 1} + \frac{\left(S_2^2/n_2\right)^2}{n_2 - 1}}$$

(새터스웨이트(Satterthwaite)의 근사식)

❖ 두 모평균의 차이 구간추정 (σ_1^2 , σ_2^2 을 모를 때)

- **소표본에서 $\mu_1 - \mu_2$의 $100(1-\alpha)\%$ 신뢰구간**

$$P\left(-t_{\alpha/2}(df) \le \frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{\sqrt{S_1^2/n_1+S_2^2/n_2}} \le t_{\alpha/2}(df)\right) = 1-\alpha \text{ 으로부터}$$

$$\Rightarrow \left[(\bar{X}_1-\bar{X}_2) - t_{\alpha/2}(df)\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}, (\bar{X}_1-\bar{X}_2) + t_{\alpha/2}(df)\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}\right]$$

df값은 정수가 아니므로 $t_{\alpha/2}(df)$의 값은
보간법으로 계산하거나 가장 가까운 정수 이용

❖ 두 모평균의 차이 구간추정 (σ_1^2, σ_2^2 을 모를 때)

- **대표본에서 $\boldsymbol{\mu_1 - \mu_2}$의 $\mathbf{100(1-\alpha)\%}$ 신뢰구간**

$$\left[(\bar{X}_1 - \bar{X}_2) - z_{\alpha/2}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}},\ (\bar{X}_1 - \bar{X}_2) + z_{\alpha/2}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}\right]$$

❖ 예제 8.5

- **부록 1의 자료를 이용하여 우리나라 남자 노인과 여자 노인의 평균 월 소득 차이를 추정하고자 한다. 유의수준 5%에서 남녀 월 소득의 모분산이 다름을 예제 8.1에서 확인하였다.**
  남녀 평균 월 소득 차이를 신뢰수준 95%에서 추정하여라.

- **풀이**

① 표본평균의 차이 : 110.04

② 표준오차 추정값 : 23.91

③ $df = 38.2$에서 $t_{0.025}(38.2) = 2.02$

④ ∴ 남녀 평균 월 소득 차이의 95% 신뢰구간 (61.74,158.34)

제 8장 두 집단 비교

8.1 독립된 두 집단 분산 비교

8.2 독립된 두 집단 평균차 추정

8.3 독립된 두 집단 평균차 검정

8.4 대응비교

❖ 두 모평균의 차이 가설검정 (σ_1^2, σ_2^2 을 알 때)

- **추정량 :** $\widehat{\mu_1 - \mu_2} = \bar{X}_1 - \bar{X}_2$

- **표본분포** :

$$\bar{X}_1 \sim N\left(\mu_1, \frac{\sigma_1^2}{n_1}\right), \bar{X}_2 \sim N\left(\mu_2, \frac{\sigma_2^2}{n_2}\right)$$

$$\bar{X}_1 - \bar{X}_2 \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

- **통계량 :**

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} \sim N(0,1)$$

❖ 두 모평균의 차이 가설검정 (σ_1^2, σ_2^2 을 알 때)

- **유의수준 α에서 모평균의 차에 대한 가설검정**

$$Z \sim N(0,1), P(Z > z_\alpha) = \alpha,$$

가설	$H_0 : \mu_1 - \mu_2 = \delta_0$ $H_1 : \mu_1 - \mu_2 \neq \delta_0$	$H_0 : \mu_1 - \mu_2 = \delta_0 \ (\leq \delta_0)$ $H_1 : \mu_1 - \mu_2 > \delta_0$	$H_0 : \mu_1 - \mu_2 = \delta_0 (\geq \delta_0)$ $H_1 : \mu_1 - \mu_2 < \delta_0$
검정 통계량 (관측값)	$z = \dfrac{(\bar{x}_1 - \bar{x}_2) - \delta_0}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$	$z = \dfrac{(\bar{x}_1 - \bar{x}_2) - \delta_0}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$	$z = \dfrac{(\bar{x}_1 - \bar{x}_2) - \delta_0}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$
임계값	$z_{\alpha/2}$	z_α	$-z_\alpha$
유의 확률	$2P(Z > \lvert z \rvert)$	$P(Z > z)$	$P(Z < z)$
H_0를 기각할 경우	$\lvert z \rvert > z_{\alpha/2}$ 또는 p-value< α	$z > z_\alpha$ 또는 p-value< α	$z < -z_\alpha$ 또는 p-value< α

❖ 예제 8.7

- 만일 우리나라 남자 노인과 여자 노인의 신장은 각각 정규분포를 따르고 남녀 신장의 모분산이 각각 $20, 36$일 때, 부록 1의 자료를 이용하여 남녀 평균 신장의 차이가 $10\ cm$인지를 유의수준 5%에서 검정하여라.

❖ 예제 8.7

- **풀이**

① 남자노인의 평균키 : μ_1 여자노인의 평균키 : μ_2

② 가설 $H_0 : \mu_1 - \mu_2 = 10$, $H_1 : \mu_1 - \mu_2 \neq 10$

③ 검정통계량 $\frac{14.6-10}{1.59} = 2.89$

④ Z=2.89 $> Z_{0.025} = 1.96 \Rightarrow H_0$ 기각

⑤ ∴ 남자와 여자의 키의 차이는 10cm라고 할 수 없다.

❖ 두 모평균의 차이 검정 (σ^2을 모를 때, 단 $\sigma^2 = \sigma_1^2 = \sigma_2^2$)

- **분산이 같을 때(등분산)와 분산이 다를 때(이분산) 통계량의 분포가 다르므로 각각 정리**
- **추정량** : $\widehat{\mu_1 - \mu_2} = \bar{X}_1 - \bar{X}_2$
- **표본분포** :

$$\bar{X}_1 \sim N\left(\mu_1, \frac{\sigma^2}{n_1}\right), \bar{X}_2 \sim N\left(\mu_2, \frac{\sigma^2}{n_2}\right), \bar{X}_1 - \bar{X}_2 \sim N\left(\mu_1 - \mu_2, \frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}\right)$$

❖ 두 모평균의 차이 검정 (σ^2을 모를 때, 단 $\sigma^2 = \sigma_1^2 = \sigma_2^2$)

- **통계량 :**

$$T = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{S_p\sqrt{1/n_1 + 1/n_2}} \sim t(n_1 + n_2 - 2)$$

단, 합동표본분산 $S_p^2 = \frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$

❖ 두 모평균의 차이 가설검정 (σ^2을 모를 때, 단 $\sigma^2 = \sigma_1^2 = \sigma_2^2$)

- (σ^2를 모르는 경우) 유의수준 α에서 모평균의 차에 대한 가설검정

$$T \sim t(n_1 + n_2 - 2), P(T > t_\alpha(n_1 + n_2 - 2)) = \alpha,\ s_p^2 = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}$$

가설	$H_0 : \mu_1 - \mu_2 = \delta_0$ $H_1 : \mu_1 - \mu_2 \neq \delta_0$	$H_0 : \mu_1 - \mu_2 = \delta_0\ (\leq \delta_0)$ $H_1 : \mu_1 - \mu_2 > \delta_0$	$H_0 : \mu_1 - \mu_2 = \delta_0(\geq \delta_0)$ $H_1 : \mu_1 - \mu_2 < \delta_0$
검정 통계량 (관측값)	$t = \dfrac{(\overline{x}_1 - \overline{x}_2) - \delta_0}{s_p\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$	$t = \dfrac{(\overline{x}_1 - \overline{x}_2) - \delta_0}{s_p\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$	$t = \dfrac{(\overline{x}_1 - \overline{x}_2) - \delta_0}{s_p\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$
임계값	$t_{\alpha/2}(n_1 + n_2 - 2)$	$t_\alpha(n_1 + n_2 - 2)$	$-t_\alpha(n_1 + n_2 - 2)$
유의 확률	$2P(T > \|t\|)$	$P(T > t)$	$P(T < t)$
H_0를 기각할 경우	$\|t\| > t_{\alpha/2}(n_1 + n_2 - 2)$ p-value$< \alpha$	$t > t_\alpha(n_1 + n_2 - 2)$ p-value$< \alpha$	$t < -t_\alpha(n_1 + n_2 - 2)$ p-value$< \alpha$

❖ 예제 8.8

- 만일 우리나라 남자 노인과 여자 노인의 신장은 각각 정규분포를 따르고 남녀 신장의 모분산은 같다는 사실이 알려져 있을 때, 부록 1의 자료를 이용하여 남녀 평균 신장의 차이가 10 cm인지를 유의수준 5%에서 검정하여라.

❖ 예제 8.8

▪ **풀이**

① 남자노인의 평균키 : μ_1 여자노인의 평균키 : μ_2

② 가설 $H_0 : \mu_1 - \mu_2 = 10$, $H_1 : \mu_1 - \mu_2 \neq 10$

③ $\overline{X_1} - \overline{X_2}$ 의 표준오차 추정값 : $\widehat{s.e.}(\overline{X_1} - \overline{X_2}) = 1.47$

④ 검정통계량 $\mathrm{t} = \frac{(14.6-10)}{1.47} = 3.13$

⑤ t=3.13 $> t_{0.025}(48) = 2.01 \Rightarrow H_0$ 기각

⑥ ∴ 남자와 여자의 키의 차이는 $10cm$라고 할 수 없다.

❖ 예제

- **나일론실 생산공장에서 새공법을 개발, 나일론실의 인장강도가 기존의 나일론실보다 더 높은지 확인하기 위해서 각각 10개의 표본을 추출하여 아래와 같이 관측하였다. 새 공법이 기존의 공법보다 우수하다고 할 수 있는지 유의수준 5%하에서 검정하라.**

새 공법	7.6	7.8	8.1	8.5	7.5	7.2	8.0	7.9	7.4	7.7
기존 공법	7.5	7.4	6.9	7.3	6.8	7.0	7.1	7.4	6.7	7.4

- **귀무가설과 대립가설의 설정**

μ_1 : 새 공법의 나일론실의 평균 인장강도

μ_2 : 기존 공법의 나일론실의 평균 인장강도

$$H_0 : \mu_1 - \mu_2 = 0 \quad H_1 : \mu_1 - \mu_2 > 0$$

❖ 예제 (계속)

- **검정통계량 계산**

$$\bar{x}_1 = \frac{1}{10}\sum_{i=1}^{n_1} x_{1i} = 7.77,\ s_1^2 = \frac{1}{n_1-1}\left(\sum_{i=1}^{n_1} x_{1i}^2 - n_1(\bar{x}_1)^2\right) = 0.1423$$

$$\bar{x}_2 = \frac{1}{10}\sum_{i=1}^{n_2} x_{2i} = 7.15,\ s_2^2 = \frac{1}{n_2-1}\left(\sum_{i=1}^{n_2} x_{2i}^2 - n_2(\bar{x}_2)^2\right) = 0.0828$$

$$s_p^2 = \frac{9(0.1423+0.0828)}{18} = 0.11255$$

$$t = \frac{\bar{x}_1-\bar{x}_2}{\sqrt{s_p^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{7.77-7.15}{\sqrt{0.11255\left(\frac{1}{10}+\frac{1}{10}\right)}} = 4.132 > t_{0.05}(18) = 1.734$$

유의수준 5%의 기각역에 포함되므로 유의수준 5%하에서 귀무가설을 기각한다. **즉, 유의수준 5%하에서 새 공법의 평균인장강도가 기존 공법 보다 우수하다고 판단한다**.

❖ 예제

- **마리화나의 주성분인 Δ^9THS (A) 와 11-OH- Δ^9THC (B)가 환각 효과에 미치는 영향의 차이는?**

  실험방법 : 건강상태가 비슷한 지원자 12명을 6명씩 랜덤추출하여 두 그룹으로 나눈 후 마리화나의 주성분을 정맥주사

  측정 : 환각효과가 느껴지기 시작하는 순간까지의 주사량을 체중 1kg당 10^{-6}gr 단위로 측정

  문제 : 환각에 필요한 두 주사량에 차이가 있는지 유의수준 5%하에서 검정하고 95% 신뢰구간을 구하라.

A	**19.54**	**14.47**	**16.00**	**24.83**	**26.39**	**11.49**
B	**15.95**	**25.89**	**20.53**	**15.52**	**14.18**	**16.00**

❖ 예제 (계속)

- **귀무가설과 대립가설의 설정**

μ_1 : A 성분으로 환각을 느끼는데 필요한 평균 주사량

μ_2 : B 성분으로 환각을 느끼는데 필요한 평균 주사량

$$H_0 : \mu_1 - \mu_2 = 0 \quad H_1 : \mu_1 - \mu_2 \neq 0$$

$$(\Leftrightarrow \; H_0 : \mu_1 = \mu_2 \quad H_1 : \mu_1 \neq \mu_2)$$

- **표본 평균과 표본 분산**

$$\bar{x}_1 = \frac{1}{6}\sum_{i=1}^{n_1} x_{1i} = 18.787, \quad s_1^2 = \frac{1}{n_1 - 1}\left(\sum_{i=1}^{n_1} x_{1i}^2 - n_1(\bar{x}_1)^2\right) = 34.908$$

$$\bar{x}_2 = \frac{1}{6}\sum_{i=1}^{n_2} x_{2i} = 18.012, \quad s_2^2 = \frac{1}{n_2 - 1}\left(\sum_{i=1}^{n_2} x_{2i}^2 - n_2(\bar{x}_2)^2\right) = 19.519$$

❖ 예제 (계속)

- **합동분산 및 검정통계량 계산**

$$s_p^2 = \frac{5(34.908+19.519)}{10} = 27.214$$

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_p^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{18.787-18.012}{\sqrt{27.214\left(\frac{1}{6}+\frac{1}{6}\right)}} = 0.257$$

양측검정, 기각역 $|t| > t_{0.025}(10) = 2.228, t = 0.257$은

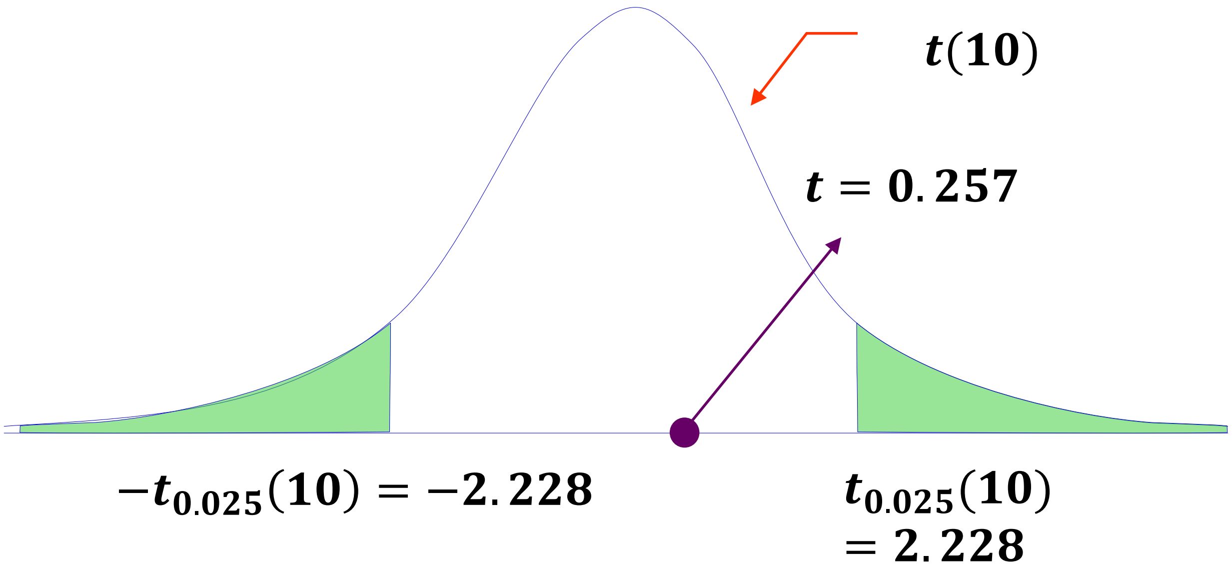
기각역에 포함되지 않으므로 귀무가설을 기각하지 못한다.

두 성분의 효과 다르다는 근거는 없다.

- **$\mu_1 - \mu_2$의 95% 신뢰구간**

$$[0.775 - 2.228(3.012), 0.775 + 2.228(3.012)] = [-5.935, 7.485]$$

❖ 예제 (계속)



- 기각역 : $|t| > 2.228$
- $|0.257| < 2.228$ 이므로 귀무가설을 기각 못함

❖ 두 모평균의 차이 가설검정 (σ_1^2 , σ_2^2 을 모를 때)

- **분산이 다를 때(이분산) 통계량의 분포는 근사적으로 t-분포**
- **추정량 :** $\widehat{\mu_1 - \mu_2} = \bar{X}_1 - \bar{X}_2$
- **표본분포 :**

$$\bar{X}_1 \sim N\left(\mu_1, \frac{\sigma^2}{n_1}\right), \bar{X}_2 \sim N\left(\mu_2, \frac{\sigma^2}{n_2}\right), \bar{X}_1 - \bar{X}_2 \sim N\left(\mu_1 - \mu_2, \frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}\right)$$

- **통계량 :**

$$T = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{S_1^2/n_1 + S_2^2/n_2}} \sim t(df), \qquad df = \frac{\left(S_1^2/n_1 + S_2^2/n_2\right)^2}{\frac{\left(S_1^2/n_1\right)^2}{n_1 - 1} + \frac{\left(S_2^2/n_2\right)^2}{n_2 - 1}}$$

(새터스웨이트(Satterthwaite)의 근사식)

❖ 예제 8.9

- **만일 우리나라 남자 노인과 여자 노인의 월 소득이 각각 정규분포를 따를 때, 부록 1의 자료를 이용하여 남녀 평균 월 소득의 차이가 100만원인지를 유의수준 5%에서 검정하여라.**

- **풀이**

① 예제 8.1의 결과에 따라 유의수준 5%에서 두 모집단의 분산이 다르므로 두 모분산이 다를 때 검정통계량 사용

② $t = \frac{(110.04-100)}{23.91} = 0.42$

③ $|t| = 0.42 > \; t_{0.025}(3.2) = 2.02 \quad \Rightarrow \quad H_0$채택

④ ∴ 남녀별 월 소득의 차이가 100만원이다.

❖ 두 모평균의 차이 가설검정 (σ_1^2 , σ_2^2 을 모를 때)

▪ (σ^2를 모르는 경우) 유의수준 α에서 모평균의 차에 대한 가설검정

$$T \sim t(df), P(T > t_\alpha(df)) = \alpha,\ df = \frac{\left(s_1^2/n_1 + s_2^2/n_2\right)^2}{\frac{\left(s_1^2/n_1\right)^2}{n_1 - 1} + \frac{\left(s_2^2/n_2\right)^2}{n_2 - 1}}$$

가설	$H_0 : \mu_1 - \mu_2 = \delta_0$ $H_1 : \mu_1 - \mu_2 \neq \delta_0$	$H_0 : \mu_1 - \mu_2 = \delta_0\ (\le \delta_0)$ $H_1 : \mu_1 - \mu_2 > \delta_0$	$H_0 : \mu_1 - \mu_2 = \delta_0 (\ge \delta_0)$ $H_1 : \mu_1 - \mu_2 < \delta_0$
검정 통계량 (관측값)	$t = \frac{(\bar{x}_1 - \bar{x}_2) - \delta_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$	$t = \frac{(\bar{x}_1 - \bar{x}_2) - \delta_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$	$t = \frac{(\bar{x}_1 - \bar{x}_2) - \delta_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$
임계값	$t_{\alpha/2}(df)$	$t_\alpha(df)$	$-t_\alpha(df)$
유의 확률	$2P(T > \lvert t \rvert)$	$P(T > t)$	$P(T < t)$
H_0를 기각할 경우	$\lvert t \rvert > t_{\alpha/2}(df)$ p-value$< \alpha$	$t > t_\alpha(df)$ p-value$< \alpha$	$t < -t_\alpha(df)$ p-value$< \alpha$

❖ 두 모평균의 차이 가설검정 (t-분포의 정규근사 이용)

- **유의수준 α에서 모평균의 차에 대한 가설검정**

$$Z \sim N(0,1), P(Z > z_{\alpha}) = \alpha$$

가설	$H_0 : \mu_1 - \mu_2 = \delta_0$ $H_1 : \mu_1 - \mu_2 \neq \delta_0$	$H_0 : \mu_1 - \mu_2 = \delta_0 \ (\leq \delta_0)$ $H_1 : \mu_1 - \mu_2 > \delta_0$	$H_0 : \mu_1 - \mu_2 = \delta_0 (\geq \delta_0)$ $H_1 : \mu_1 - \mu_2 < \delta_0$
검정 통계량 (관측값)	$z = \dfrac{(\overline{x}_1 - \overline{x}_2) - \delta_0}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$	$z = \dfrac{(\overline{x}_1 - \overline{x}_2) - \delta_0}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$	$z = \dfrac{(\overline{x}_1 - \overline{x}_2) - \delta_0}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$
임계값	$z_{\alpha/2}$	z_{α}	$-z_{\alpha}$
유의 확률	$2P(Z > \lvert z \rvert)$	$P(Z > z)$	$P(Z < z)$
H_0를 기각할 경우	$\lvert z \rvert > z_{\alpha/2}$ p-value$< \alpha$	$z > z_{\alpha}$ p-value$< \alpha$	$z < -z_{\alpha}$ p-value$< \alpha$

❖ 예제

- 두 지역 중학생의 학력 차이가 있는지 확인하기 위하여 각 지역에서 120명씩 랜덤하게 선발하여 같은 시험을 시행하였다. 주어진 정보로부터 두 지역의 학력이 다르다고 할 수 있는지 유의수준 5%하에서 검정하고 평균 차이에 대한 95% 신뢰구간을 구하여라.

	지역 A	지역 B
표본평균	$\bar{x}_1 = 82.3$	$\bar{x}_2 = 78.2$
표본표준편차	$s_1 = 13.6$	$s_2 = 10.4$

❖ 예제 (계속)

- **귀무가설과 대립가설의 설정**

  μ_1 : A 지역 학생의 평균 성적,

  μ_2 : B 지역 학생의 평균 성적

$$H_0 : \mu_1 = \mu_2 \quad H_1 : \mu_1 \neq \mu_2$$

- **표본 평균과 표본 표준편차**

$$\bar{x}_1 = 82.3, \bar{x}_2 = 78.2, s_1 = 13.6, s_2 = 10.4$$

❖ 예제 (계속)

- **검정통계량 계산**

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{82.3 - 78.2}{\sqrt{\frac{13.6^2}{120} + \frac{10.4^2}{120}}} = 2.62$$

양측검정의 기각역은 $|z| > z_{0.025} = 1.96$이므로 $z = 2.62$는 기각역에 포함되므로 귀무가설을 기각.

- **평균 성적 차이의 95% 신뢰구간**

$$\left[\bar{x}_1 - \bar{x}_2 - z_{0.025}\sqrt{s_1^2/n_1 + s_2^2/n_2}, \bar{x}_1 - \bar{x}_2 + z_{0.025}\sqrt{s_1^2/n_1 + s_2^2/n_2}\right] = [1.0, 7.2]$$