# 제 10 장  두 모집단의 추론

**예제.** $x$ ~ $N(100, 3^2)$

문제 1. $\mu_x$ =

문제 2. $\mu_x$를 10번 구하고, 이들의 평균과 분산을 구하시오.

Note: $\mu_x$는 상수이다.

문제 3. $n$ = 4인 표본을 5개 구한 후 통계처리한 결과이다.

| | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $\bar{x}$ |
|---|---|---|---|---|---|
| 1 | 99.89 | 96.68 | 102.40 | 102.85 | **100.45** |
| 2 | 101.06 | 99.45 | 100.07 | 93.57 | **98.54** |
| 3 | 94.64 | 95.68 | 102.83 | 97.63 | **97.70** |
| 4 | 103.92 | 97.21 | 104.25 | 100.06 | **101.36** |

$\bar{x}$는 상수이다. True, False

문제 4. $n$ = 4인 가능한 모든 표본을 구한 후, 각 “표본의 평균”의 평균은 무엇인가?

문제 5. $\mu_{\bar{x}}$의 평균과 분산은?

**예제.** 두 도시의 가구별 월소득 격차에 대해 조사를 하려한다. 편의상 첫 번째 도시는 1, 두 번째 도시는 2로 표기하자. 도시 1에서 $n_1$개의 가구별 월소득액을, 도시 2에서 $n_2$개의 가구별 월소득액을 구하였다.

문제 1. 도시 1에서 구한 $n_1$개의 월소득액의 평균을 기호로 표기하시오.

문제 2. 도시 2에서 구한 $n_2$개의 월소득액의 평균을 기호로 표기하시오.

문제 3. 문제 1과 2에서 구한 값의 차이를 기호로 표기하시오.

문제 4. 도시 1에서 $n_1$의 자료를 구하고 도시 2에서 $n_2$개의 자료를 구한 후, 이들의 차이를 구한다. 이를 반복한다고 하자. 이들 값의 차이는 상수인가 아니면 변수인가? 변수라면 이들의 평균과 분산은 어떻게 표기할 수 있는가?

문제 5. 아래 표에서 M은 충분히 큰 값을 의미한다. M이 충분히 크다면, 빈 칸의 값들은 어떤 값에 근접하는가?

| | $x_{11}$ | $x_{12}$ | $x_{13}$ | $x_{21}$ | $x_{22}$ | $\overline{x_1}$ | $\overline{x_2}$ | $\overline{x_1}-\overline{x_2}$ |
|---|---|---|---|---|---|---|---|---|
| 1 | 654.57 | 879.54 | 710.47 | 1,047.48 | 964.76 | **748.19** | **1,006.12** | -257.93 |
| 2 | 628.87 | 754.05 | 800.34 | 916.23 | 995.09 | **727.75** | **955.66** | -227.91 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| M | 896.15 | 525.96 | 877.68 | 949.74 | 940.57 | **766.60** | **945.16** | -178.56 |
| 평균 | | | | | | | | |
| 분산 | | | | | | | | |
| 표준편차 | | | | | | | | |

## 제1절 두 모집단 평균차의 추론: 독립표본

두 표본분포가 독립적이면

① $\mu_{\overline{X_1}-\overline{X_2}} = \mu_{\overline{X_1}} - \mu_{\overline{X_2}} = \mu_{X_1} - \mu_{X_2} = \mu_1 - \mu_2$

함수 형식으로 표현하면 $E(\overline{x_1}-\overline{x_2}) = E(\overline{x_1}) - E(\overline{x_2}) = E(x_1) - E(x_2)$

Note: $\mu_{\overline{X_1}+\overline{X_2}} = \mu_{\overline{X_1}} + \mu_{\overline{X_2}} = \mu_{X_1} + \mu_{X_2} = \mu_1 + \mu_2$

② $\sigma^2_{\overline{X_1}-\overline{X_2}} = \sigma^2_{\overline{X_1}} + \sigma^2_{\overline{X_2}} = \dfrac{\sigma^2_{X_1}}{n_1}+\dfrac{\sigma^2_{X_2}}{n_2}$

함수 형식으로 표현하면 $V(\overline{x_1}-\overline{x_2}) = V(\overline{x_1}) + V(\overline{x_2}) = \dfrac{V(x_1)}{n_1} + \dfrac{V(x_2)}{n_2}$

Note: $\sigma^2_{\overline{X_1}+\overline{X_2}} = \sigma^2_{\overline{X_1}} + \sigma^2_{\overline{X_2}} = \dfrac{\sigma^2_{X_1}}{n_1}+\dfrac{\sigma^2_{X_2}}{n_2}$

$\mu_{\overline{X_1}+\overline{X_2}} \neq \mu_{\overline{X_1}-\overline{X_2}}$이지만, $\sigma^2_{\overline{X_1}+\overline{X_2}} = \sigma^2_{\overline{X_1}-\overline{X_2}}$가 성립한다.

③ 두 모집단이 정규분포이거나 각 표본의 크기가 충분히 크면 ($\overline{X_1}$과 $\overline{X_2}$가 정규분포를 따르를 조건임)

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

예제 1. 두 모집단은 정규분포를 따르고, 각 표본은 서로 독립적이다.

$\mu_1 = 2{,}000,\ \sigma_1^2 = 100^2,\ n_1 = 40,\ \mu_2 = 1{,}800,\ \sigma_2^2 = 50^2,\ n_2 = 50$

① 두 모집단 표본 평균차는 어떤 분포인가?

두 모집단이 정규분포를 따르고, 각 표본은 서로 독립적이므로,

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

$\rightarrow \overline{X_1}-\overline{X_2} \sim N\left(2{,}000-1{,}800, \dfrac{100^2}{40}+\dfrac{50^2}{50}\right)$

② 두 모집단 표본 평균의 합은 어떤 분포인가?

두 모집단이 정규분포를 따르고, 각 표본은 서로 독립적이므로,

$\overline{X_1}+\overline{X_2} \sim N\left(\mu_1+\mu_2, \dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}\right)$이다.

$\rightarrow \overline{X_1}+\overline{X_2} \sim N\left(2{,}000+1{,}800, \dfrac{100^2}{40}+\dfrac{50^2}{50}\right)$

**모의실험 1.** $x_1 \sim N(5,4^2)$, $x_2 \sim N(-5,3^2)$, $n_1 = n_2 = 2$ (두 모집단은 서로 독립적)

| | 모집단 1 | | 모집단 2 | | $\overline{x_1}$ | $\overline{x_2}$ | $\overline{x_1}+\overline{x_2}$ | $\overline{x_1}-\overline{x_2}$ |
|---|---|---|---|---|---|---|---|---|
| | $x_{11}$ | $x_{12}$ | $x_{21}$ | $x_{22}$ | | | | |
| 1 | -0.79 | 10.74 | -5.18 | 2.50 | 4.97 | -1.34 | 3.64 | 6.31 |
| 2 | 10.71 | -2.76 | -9.57 | -12.13 | 3.97 | -10.85 | -6.88 | 14.82 |
| 3 | -1.95 | 3.88 | -7.78 | -8.66 | 0.96 | -8.22 | -7.25 | 9.18 |
| … | … | … | … | … | … | … | … | … |
| 100,000 | 8.54 | 2.62 | -7.30 | -2.40 | 5.58 | -4.85 | 0.73 | 10.43 |
| 평균 | 4.99 | 5.00 | -5.00 | -4.98 | 4.99 | -4.99 | -0.00 | 9.99 |
| 분산 | 16.04 | 16.11 | 8.96 | 8.97 | 8.05 | 4.47 | 12.54 | 12.49 |

**연습문제 1.**

문제 1. $x_{11}$의 평균과 분산은?

문제 2. $\overline{x_1}$의 평균과 분산은?

문제 3. $x_{21}$의 평균과 분산은?

문제 4. $\overline{x_2}$의 평균과 분산은?

문제 5. $E(\overline{x_1}+\overline{x_2})$와 $E(\overline{x_1}-\overline{x_2})$

문제 6. $V(\overline{x_1}+\overline{x_2})$와 $V(\overline{x_1}-\overline{x_2})$

문제 7. $\overline{x_1}+\overline{x_2}$와 $\overline{x_1}-\overline{x_2}$는 어떤 분포를 따르는가?

**모의실험 2.** $x \sim U[0,\ 20]$, $y \sim N(5,\ 3^2)$, $n_x\ =2$, $n_y\ =3$

| | 모집단 1 | | 모집단 2 | | | $\bar{x}$ | $\bar{y}$ | $\bar{x}-\bar{y}$ |
|---|---|---|---|---|---|---|---|---|
| | $x_1$ | $x_2$ | $y_1$ | $y_2$ | $y_3$ | | | |
| 1 | 1.89 | 18.17 | 8.59 | 1.25 | 6.08 | 10.03 | 5.31 | 4.72 |
| 2 | 14.15 | 17.57 | 7.25 | 3.08 | 6.95 | 15.86 | 5.76 | 10.10 |
| 3 | 10.24 | 8.41 | 4.63 | 6.28 | 1.67 | 9.32 | 4.19 | 5.13 |
| 4 | 10.31 | 6.38 | 9.38 | 4.02 | 5.68 | 8.34 | 6.36 | 1.98 |
| 5 | 3.73 | 6.13 | 5.27 | 6.50 | 5.19 | 4.93 | 5.65 | -0.72 |
| … | 5.46 | 5.94 | 4.45 | 3.95 | 7.99 | 5.70 | 5.47 | 0.24 |
| 10,000 | 5.72 | 8.68 | 8.10 | 6.95 | 6.20 | 7.20 | 7.08 | 0.12 |
| 평균 | 10.07 | 10.05 | 5.03 | 4.99 | 5.00 | 10.06 | 5.00 | 5.05 |
| 분산 | 33.76 | 33.27 | 8.96 | 8.89 | 9.15 | 16.61 | 3.05 | 19.72 |

**연습문제 2.**

문제 1. $E(\bar{x})$, $E(\bar{y})$, $V(\bar{x})$, $V(\bar{y})$

문제 2. $E(\bar{x}-\bar{y})$, $V(\bar{x}-\bar{y})$

문제 3. $\bar{x}-\bar{y}$는 어떤 분포를 따르는가?

문제 4. 위의 실험에서 $n_x\ =30$, $n_y\ =10$으로 변경시킨다고 하자. 이때 $E(\bar{x}-\bar{y})$, $V(\bar{x}-\bar{y})$는?

문제 5. 위의 실험에서 $n_x\ =30$, $n_y\ =10$으로 변경시킨다고 하자. $\bar{x}-\bar{y}$는 어떤 분포를 따르는가?

## 1. 두 모집단의 분산을 알고 있는 경우

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

… $X_1$과 $X_2$가 정규분포를 따르거나 또는 $n_1$과 $n_2$가 충분히 큰 경우(중심극한의 정리)이고 두 표본이 서로 독립적인 경우

**[참고] [$\overline{X_1} - \overline{X_2}$]를 하나의 변수로 생각합니다.**

① ($\overline{X_1} - \overline{X_2}$)의 평균과 분산의 표기는?

평균: $E$(변수) 또는 $\mu_{\text{변수}}$, 분산: $V$(변수) 또는 $\sigma^2_{\text{변수}}$

→ 현재 변수는 ($\overline{X_1} - \overline{X_2}$)이므로,

평균: $E(\overline{X_1}-\overline{X_2})$, $\mu_{\overline{X_1}-\overline{X_2}}$

분산: $V(\overline{X_1}-\overline{X_2})$, $\sigma^2_{\overline{X_1}-\overline{X_2}}$

② ($\overline{X_1} - \overline{X_2}$)의 평균과 분산을 $\mu_1$, $\mu_2$, $\sigma_1^2$, $\sigma_2^2$, $n_1$과 $n_2$로 표현하면? (단, 각 표본이 서로

독립적이란 전제가 필요하다.)

$\mu_{\overline{X}_1-\overline{X}_2} = \mu_1 - \mu_2$

$\sigma^2_{\overline{X}_1-\overline{X}_2} = \dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}$

③ 정규분포를 따르는 어떤 변수의 $z$값은?

$$z = \frac{\text{변수} - \text{모집단 평균}_{\text{변수}}}{\text{모집단 표준편차}_{\text{변수}}}$$

④ ③에 $(\overline{X}_1 - \overline{X}_2)$를 대입하면

$$z = \frac{(\overline{X}_1 - \overline{X}_2) - \mu_{\overline{X}_1-\overline{X}_2}}{\sigma_{\overline{X}_1-\overline{X}_2}} = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$$

## 1-1. 양측검정(Two-tailed Test)

(1) $H_0$: $\mu_1 - \mu_2 = 0$, $H_A$: $\mu_1 - \mu_2 \neq 0$

$\mu_1 - \mu_2 = 0$은 $\mu_{\overline{X}_1-\overline{X}_2} = 0$과 동일하며, $\mu_1 - \mu_2 \neq 0$은 $\mu_{\overline{X}_1-\overline{X}_2} \neq 0$과 동일하다.

(2) $(\overline{X}_1 - \overline{X}_2)$의 표준통계량

$$z = \frac{(\overline{X}_1 - \overline{X}_2) - \mu_{\overline{X}_1-\overline{X}_2}}{\sigma_{\overline{X}_1-\overline{X}_2}} = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$$ 이다.

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

$$(\overline{X}_1 - \overline{X}_2) \pm z_{\frac{\alpha}{2}} \cdot \sigma_{\overline{X}_1-\overline{X}_2} \rightarrow (\overline{X}_1 - \overline{X}_2) \pm z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

예제 2. 두 모집단은 정규분포를 따르고, 각 표본은 서로 독립적이다.

$\overline{X}_1 = 810$, $\sigma_1^2 = 100^2$, $n_1 = 40$, $\overline{X}_2 = 850$, $\sigma_2^2 = 50^2$, $n_2 = 50$, $\alpha = 0.05$

(1) $H_0$: $\mu_1 - \mu_2 = 0$, $H_A$: $\mu_1 - \mu_2 \neq 0$ ( → 양측 검정)

(2) Test Statistic: $\dfrac{(\overline{X}_1 - \overline{X}_2) - \mu_{\overline{X}_1-\overline{X}_2}}{\sigma_{\overline{X}_1-\overline{X}_2}} = \dfrac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$

(두 모집단이 정규분포를 따르고, 각 표본이 서로 독립적이며, 모집단 분산을 알고 있으므로 이 Test Statistic을 사용할 수 있다.)

[설명] Test Statistic에서 $\mu_1$-$\mu_2$의 값은 $H_0$에서 설정한 값을 사용하고, $\overline{X_1}$과 $\overline{X_2}$는 표본에서 구한 값을 사용한다.

(3) For $\alpha$ = 0.05,

Rejection Region: $Z > Z_{0.025}$ = 1.96 또는 $Z <$ -$Z_{0.025}$ = -1.96

(양측검정이므로 $z_{\frac{\alpha}{2}}$와 -$z_{\frac{\alpha}{2}}$가 사용된다.)

[설명] $H_0$가 사실일 때, -$Z_{0.025}$보다 작거나 또는 $Z_{0.025}$보다 큰 $Z$값이 나올 확률은 $\alpha$이다.

(4) Value of the Test Statistic:

$Z = \dfrac{-40-0}{17.32}$ = -2.31

[설명] 여기서, $\mu_1$-$\mu_2$=0은 바로 $H_0$이다.

(5) Conclusion:

$H_0$를 기각. 두 모집단의 평균의 차이가 유의하다.

[참고] p-value = 2・$P(z > |-2.31|)$ = 2・0.0104 = 0.0208

[참고] $\mu_1$-$\mu_2$의 95% 신뢰구간:

$(\overline{X_1} - \overline{X_2}) \pm z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}} \rightarrow$ -40 ± 1.96(17.32) = [-73.9, -6.11]

## 연습문제 3.

두 모집단은 모두 정규분포를 따른다. Test Statistic으로 $z = \dfrac{(\overline{X_1}-\overline{X_2})-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$를 사용할 있는 지를 살펴보고자 한다.

문제 1. 수식에 사용된 변수, 상수들 중 표본을 통해 알 수 있는 값을 모두 나열하시오.

문제 2. 가설로부터 인용하는 값을 모두 나열하시오.

문제 3. 표본으로부터 구할 수 없고 가설로부터도 인용할 수 없는 값(즉, 알려져 있어야 하는 값)을 모두 나열하시오.

**연습문제 4.**

두 모집단은 정규분포를 따르고, 각 표본은 서로 독립적이다. 두 모집단 평균이 같지 않다는 의심이 들어 조사한 결과 아래의 통계량을 구했다.

$\overline{X_1}$ = 800, $\sigma_1^2$ = $100^2$, $n_1$ = 10, $\overline{X_2}$ = 850, $\sigma_2^2$ = $50^2$, $n_2$ = 5,

문제 1. 가설검정을 수행하시오. 단, $\alpha$ = 0.05

문제 2. $\mu_1$-$\mu_2$의 95% 신뢰구간을 구하시오.

문제 3. 신뢰구간과 가설검정의 결론에 대해 논하시오.

문제 4. $p$-value를 구하시오.

문제 5. $p$-value와 가설검정의 결론에 대해 논하시오.

## 2. 두 모집단 분산을 모르는 경우

### 2.1 대표본인 경우

(1) 가설검정의 조건: $\overline{X_1}$- $\overline{X_2}$ ~ $N\left(\mu_1 - \mu_2, \dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}\right)$

$X_1$과 $X_2$가 정규분포를 따르거나 또는 $n_1$과 $n_2$가 충분히 큰 경우(중심극한의 정리)이고 두 표본이 서로 독립적인 경우

(2) $\sigma_1^2$과 $\sigma_2^2$의 추정

두 모집단의 분산이 알려져 있으면 $\overline{X_1}$-$\overline{X_2}$의 분산은 $\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}$으로 구할 수 있다. 대표본이므로 (각 표본에서 구하는) $S_1^2$과 $S_2^2$로 $\sigma_1^2$과 $\sigma_2^2$을 추정한다.

($\overline{X_1}$-$\overline{X_2}$)의 모집단 분산 = $\sigma^2_{\overline{X_1}-\overline{X_2}}$ = $\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}$

($\overline{X_1}$-$\overline{X_2}$)의 모집단 분산 추정치 = $S^2_{\overline{X_1}-\overline{X_2}}$ = $\dfrac{S_1^2}{n_1}+\dfrac{S_2^2}{n_2}$

(3) ($\overline{X_1}$- $\overline{X_2}$)의 표준통계량

$$z = \frac{(\overline{X_1}-\overline{X_2})-\mu_{\overline{X_1}-\overline{X_2}}}{S_{\overline{X_1}-\overline{X_2}}} = \frac{(\overline{X_1}-\overline{X_2})-(\mu_1-\mu_2)}{\sqrt{\dfrac{S_1^2}{n_1}+\dfrac{S_2^2}{n_2}}}$$ 이다.

($\sigma^2_{\overline{X_1}-\overline{X_2}}$을 모르는 경우 표준통계량으로 $t$를 사용한다. 그런데 $t$의 자유도가 커지면 $t$값은 $z$값에 근접하므로, $t$대신 $z$를 사용해도 차이가 별 차이가 없다. 물론 정확히

하려면 $t$를 사용한다.)

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

$$(\overline{X_1} - \overline{X_2}) \pm z_{\frac{\alpha}{2}} \cdot S_{\overline{X_1}-\overline{X_2}} \rightarrow (\overline{X_1} - \overline{X_2}) \pm z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}$$

예제 4. 두 표본은 서로 독립적이다.

$\overline{X_1}$ = 2.4, $S_1$ = 4.0, $n_1$ = 35, $\overline{X_2}$ = 4.5, $S_2$ = 6.2, $n_2$ = 30, $\alpha$ = 0.05

(1) $H_0$: $\mu_1 - \mu_2 = 0$, $H_A$: $\mu_1 - \mu_2 \neq 0$ ( → 양측 검정)

(2) Test Statistic: $t_{n_1+n_2-2} = \dfrac{(\overline{X_1}-\overline{X_2})-\mu_{\overline{X_1}-\overline{X_2}}}{S_{\overline{X_1}-\overline{X_2}}} = \dfrac{(\overline{X_1}-\overline{X_2})-(\mu_1-\mu_2)}{\sqrt{\dfrac{S_1^2}{n_1}+\dfrac{S_2^2}{n_2}}}$

(여기서, $t_{n_1+n_2-2}$ 대신 $z$를 사용해도 무방)

(3) For $\alpha$ = 0.05,

Rejection Region: $Z > Z_{0.025}$ = 1.96 또는 $Z <$ -$Z_{0.025}$ = -1.96

(4) Value of the Test Statistic:

$$z = \frac{(2.4-4.5)-0}{\sqrt{\dfrac{4.0^2}{35}+\dfrac{6.2^2}{30}}} = -1.59$$

(5) Conclusion:

$H_0$를 기각하는데 실패. 두 모집단의 평균의 차이가 유의하다 할 수 없다.

Note: ($\overline{X_1}$-$\overline{X_2}$)의 모집단 분산 추정치 = $\dfrac{S_1^2}{n_1}+\dfrac{S_2^2}{n_2} = \dfrac{4.0}{35}+\dfrac{6.2^2}{30}$

[참고] $p$-value = 2 · $P(z > |-1.59|)$ = 2 · 0.0559 = 0.1118

$p$-value가 $\alpha$보다 크므로, $H_0$를 기각할 수 없다.

[참고] $\mu_1$-$\mu_2$의 95% 신뢰구간:

$$(\overline{X_1} - \overline{X_2}) \pm z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}} \rightarrow -2.1 \pm 1.96(1.32) = [-4.69,\ 0.49]$$

양측검정이고, $H_0$: $\mu_1 - \mu_2$ = 0이다. 신뢰구간이 $H_0$에서 설정한 $\mu_1 - \mu_2$값인 0을 포함하고 있으므로, $H_0$를 기각할 수 없다.

## 연습문제 5.

문제 1. Test Statistic으로 $z = \dfrac{(\overline{X_1}-\overline{X_2})-(\mu_1-\mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}}$를 사용할 수 있는 조건을 나열하시오.

문제 2. $\sigma_1$과 $\sigma_2$를 모르는 경우, 표본으로부터 이들을 추정할 수 있는 통계량은?

문제 3. 어떤 표본을 정리하였다. {3.0, 5.0, 5.8, 6.2} $\overline{X}$, $s^2$, $s$, $s_{\overline{X}}$값을 구하시오.

문제 4. $\overline{X_1}$- $\overline{X_2}$이 정규분포를 따르지만 $\sigma_1^2$, $\sigma_2^2$은 알려져 있지 않다. $n_1$과 $n_2$는 크다.
이 경우 가설검정에서 사용 가능한 통계량은?

## 연습문제 6.

두 표본은 서로 독립적이다. 모집단 분산들은 알려져 있지 않다.

$\overline{X_1}$ = 3.5, $S_1$ = 4.0, $n_1$ = 25, $\overline{X_2}$ = 3.0, $S_2$ = 6.0, $n_2$ = 36, $\alpha$ = 0.05

문제 1. $S_{\overline{X_1}}$, $S_{\overline{X_2}}$, $S_{\overline{X_1}-\overline{X_2}}$을 구하시오.

문제 2. 양측 가설검정을 수행하시오. 단, $H_0$: $\mu_1-\mu_2$ = 0이다.

문제 3. $\mu_1-\mu_2$의 95% 신뢰구간을 구하시오.

문제 4. 신뢰구간과 가설검정의 결론에 대해 논하시오.

문제 5. $p$-value를 구하시오.

문제 6. $p$-value와 가설검정의 결론에 대해 논하시오.

## 2.2 소표본인 경우 [기초]

### 두 모집단의 분산이 같은 경우 표본으로부터 분산을 추정하는 방법

**모의실험 3.** $x_1 \sim N(100, 10^2)$, $x_2 \sim N(200, 10^2)$

| | $x_1$ | $(x_{1i} - \overline{x_1})^2$ | $x_2$ | $(x_{2i} - \overline{x_2})^2$ |
|---|---|---|---|---|
| 1 | 97.94 | 0.54 | 183.96 | 158.04 |
| 2 | 97.30 | 0.01 | 194.91 | 2.63 |
| 3 | 106.06 | 78.49 | 192.81 | 13.81 |
| 4 | 110.20 | 168.81 | 207.93 | 129.90 |
| 5 | 100.78 | 12.77 | 203.04 | 42.39 |
| 6 | 95.36 | 3.42 | | |
| 7 | 81.66 | 241.62 | | |
| 8 | 94.94 | 5.15 | | |
| 9 | 99.59 | 5.71 | | |
| 10 | 88.23 | 80.60 | | |
| 합계 | 972.05 | 597.12 | 982.65 | 346.77 |
| 분모 | 10 | 9 | 5 | 4 |
| 평균 | 97.21 | 66.35 | 196.53 | 86.69 |

**연습문제 7.**

문제 1. $S_1^2$과 $S_2^2$를 구하시오.

문제 2. 두 모집단의 분산은 동일한 것으로 알려져 있다. 위의 표에서 어떤 데이터들의 기대값이 같다는 것인가?

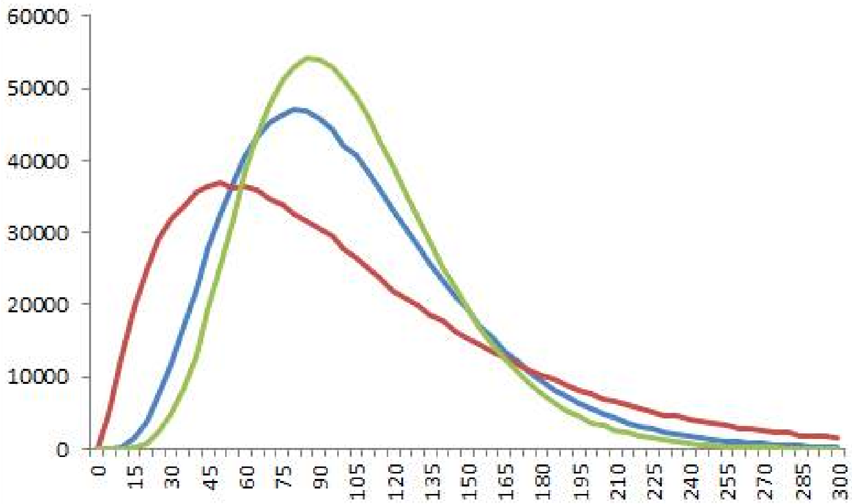
문제 3. 두 모집단의 분산이 동일하다면, $S_1^2$과 $S_2^2$으로부터 어떤 방법으로 모집단 분산을 추정할 수 있는가?

문제 4. 두 모집단의 분산이 동일하다고 한다. 각 표본들의 편차제곱의 합만 알려져 있다면, 어떤 방법으로 모집단 분산을 추정할 수 있는가?

문제 5. 문제 3과 문제 4에서 구한 값이 동일함을 보이시오.

**모의실험 4.** $x_1 \sim N(100, 10^2)$, $x_2 \sim N(200, 10^2)$, $n_1 = 10$, $n_2 = 5$

| | $S_1^2$ | $S_2^2$ | 가중평균 |
|---|---|---|---|
| 1 | 109.47 | 98.62 | 106.13 |
| 2 | 90.30 | 14.64 | 67.02 |
| 3 | 43.79 | 16.60 | 35.42 |
| 4 | 119.80 | 109.51 | 116.63 |
| 5 | 198.57 | 11.19 | 140.92 |
| 6 | 162.99 | 67.25 | 133.53 |
| 7 | 106.87 | 47.36 | 88.56 |
| 8 | 125.88 | 32.83 | 97.25 |
| 9 | 40.27 | 23.34 | 35.06 |
| 10 | 176.20 | 24.10 | 129.40 |
| … | … | … | … |
| 10,000 | 100.70 | 32.62 | 79.75 |
| 평균 | 99.56 | 100.51 | 99.85 |
| 분산 | 2,220.02 | 5,158.14 | 1,555.59 |




왼쪽부터 $S_2^2$, $S_1^2$, 가중평균이다.

## 2.2 소표본인 경우

(1) $X_1$과 $X_2$가 정규분포를 따르고, 서로 독립적이며, $\sigma_1 = \sigma_2$이고, $n_1$과 $n_2$는 충분히 크지 않다.

… 가설검정의 조건: $\overline{X_1} - \overline{X_2} \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right) \rightarrow \sim N\left(\mu_1 - \mu_2, \frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}\right)$

$\sigma_1 = \sigma_2$이므로 $\sigma_1$과 $\sigma_2$ 모두 $\sigma$로 표기한다.

(2) $\sigma_1^2 = \sigma_2^2 = \sigma^2$의 추정

$$S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{(n_1-1)+(n_2-1)} \leftarrow S_p^2\text{: pooled estimator of variance}$$

일종의 가중평균법이다.

$$\sigma^2_{\overline{X_1}-\overline{X_2}} = \frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2} = \frac{\sigma^2}{n_1}+\frac{\sigma^2}{n_2} \rightarrow (\text{추정})\ \frac{S_p^2}{n_1}+\frac{S_p^2}{n_2}$$

(3) $(\overline{X_1}-\overline{X_2})$의 표준통계량

$$t = \frac{(\overline{X_1}-\overline{X_2})-\mu_{\overline{X_1}-\overline{X_2}}}{S_{\overline{X_1}-\overline{X_2}}} = \frac{(\overline{X_1}-\overline{X_2})-(\mu_1-\mu_2)}{\sqrt{\frac{S_p^2}{n_1}+\frac{S_p^2}{n_2}}}\text{이다.}$$

자유도(degree of freedom, df) = $(n_1-1)+(n_2-1)$

($\sigma^2_{\overline{X_1}-\overline{X_2}}$을 모르는 경우 표준통계량으로 $t$를 사용한다. 그런데 $t$의 자유도가 커지면 $t$값은 $z$값에 근접하므로, $t$대신 $z$를 사용해도 차이가 별 차이가 없지만, 이번 경우는 자유도가 크지 않은 경우이다.)

Note: $S_p^2$은 두 모집단 분산의 추정치이고, $\frac{S_p^2}{n_1}+\frac{S_p^2}{n_2}$는 $(\overline{X_1}-\overline{X_2})$의 분산 추정치이다.

예제 5. 두 모집단은 독립적이고 정규분포를 이루며 두 모집단 분산은 같다. $\alpha = 0.10$

A: 85, 63, 92, 40, 76, 82, 85, 68, 80, 95

B: 98, 92, 60, 83, 85, 89, 70, 85, 53, 80

$\overline{X_1} = 76.6$, $S_1 = 16.18$, $n_1 = 10$, $\overline{X_2} = 78.5$, $S_2 = 14.20$, $n_2 = 10$

$$S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{(n_1-1)+(n_2-1)} = \frac{(10-1)16.18^2+(10-1)14.20^2}{(10-1)+(10-1)} = 231.72$$

(1) $H_0$: $\mu_1-\mu_2 = 0$, $H_A$: $\mu_1-\mu_2 \neq 0$ ( → 양측 검정)

(2) Test Statistic: $t = \frac{(\overline{X_1}-\overline{X_2})-\mu_{\overline{X_1}-\overline{X_2}}}{S_{\overline{X_1}-\overline{X_2}}} = \frac{(\overline{X_1}-\overline{X_2})-(\mu_1-\mu_2)}{\sqrt{\frac{S_p^2}{n_1}+\frac{S_p^2}{n_2}}}$

(3) For $\alpha = 0.10$,

Rejection Region: $t > t_{0.05,\,18} = 1.7341$ 또는 $t < -t_{0.05,\,18} = -1.7341$

(4) $t = \dfrac{(\overline{X_1}-\overline{X_2})-(\mu_1-\mu_2)}{\sqrt{\dfrac{S_p^2}{n_1}+\dfrac{S_p^2}{n_2}}} = \dfrac{(76.6-78.5)-0}{\sqrt{\dfrac{231.72}{10}+\dfrac{231.72}{10}}}$=-0.28

(5) $H_0$ 기각 실패

[참고] $p$-value = 2 · $P(t_{18} > |-0.28|)$ = 2 · 0.3913 = 0.7827

[참고] $\mu_1$-$\mu_2$의 90% 신뢰구간:

$$(\overline{X_1}-\overline{X_2}) \pm t_{0.05,\,18} \cdot \sqrt{\frac{S_p^2}{n_1}+\frac{S_p^2}{n_2}} \rightarrow -1.9 \pm 1.734(6.81) = [-13.71,\ 9.91]$$

### 연습문제 8.

두 모집단은 독립적이고 정규분포를 이루며 두 모집단 분산은 같다.

$\overline{X_1}$ = 95.6, $S_1$ = 16.18, $n_1$ = 16, $\overline{X_2}$ = 101.2, $S_2$ = 17.3, $n_2$ = 11

문제 1. 두 모집단의 분산을 추정하시오.

문제 2. 양측 가설검정을 수행하시오. 단, $H_0$: $\mu_1-\mu_2$ = 0이다.

문제 3. $\mu_1-\mu_2$의 95% 신뢰구간을 구하시오.

문제 4. 신뢰구간과 가설검정의 결론에 대해 논하시오.

문제 5. $p$-value를 구하시오.

문제 6. $p$-value와 가설검정의 결론에 대해 논하시오.

## 제2절 쌍체(matched pairs) 비교

| 번호 | $X_1$(전) | $X_2$(후) | $d(X_1$-$X_2)$ 차이 | $d$-$\overline{d}$ 편차 | $(d-\overline{d})^2$ 편차제곱 |
|---|---|---|---|---|---|
| 1 | 70 | 68 | 2 | -1 | 1 |
| 2 | 62 | 62 | 0 | -3 | 9 |
| 3 | 54 | 50 | 4 | 1 | 1 |
| … | … | … | … | … | 1 |
| 10 | 63 | 60 | 3 | 0 | 0 |
| 합계 | | | 30 | | 94 |

| 번호 | $X_1$(전) | $X_2$(후) | $d(X_1\text{-}X_2)$ 차이 | $d\text{-}\bar{d}$ 편차 |
|---|---|---|---|---|
| 1 | $X_{1,1}$ | $X_{1,2}$ | $d_1\ (=X_{1,1}-X_{1,2})$ | $d_1\text{-}\bar{d}$ |
| 2 | $X_{2,1}$ | $X_{2,2}$ | $d_2\ (=X_{2,1}-X_{2,2})$ | $d_2\text{-}\bar{d}$ |
| 3 | $X_{3,1}$ | $X_{3,2}$ | $d_3\ (=X_{3,1}-X_{3,2})$ | $d_3\text{-}\bar{d}$ |
| … | … | … | … | … |
| 10 | $X_{10,1}$ | $X_{10,2}$ | $d_{10}\ (=X_{10,1}-X_{10,2})$ | $d_{10}\text{-}\bar{d}$ |

$d$의 표본 평균과 표본 분산

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

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

$d$가 정규분포를 이루는 경우, $\mu_d$에 관한 inference를 제시할 수 있다.
현실적으로 $\sigma_d$는 잘 알려져 있지 않으므로 confidence interval이나 hypothesis test에서 $t$ -distribution을 이용한다.

(1) 검정통계량

$$t = \frac{\bar{d}-\mu_d}{S_d/\sqrt{n}}, \text{ degree of freedom} = n\text{-}1$$

(2) $\mu_d$의 (1-$\alpha$)100% 신뢰구간

$$\bar{d} \pm t_{\frac{\alpha}{2},\,n-1} \cdot S_d/\sqrt{n}$$

예) $\bar{d} = \frac{30}{10} = 3,\quad S_d^2 = \frac{94}{10-1} = 10.44,\ S_d = \sqrt{10.4444} = 3.2318,\ \alpha=0.05$

(1) $H_0$: $\mu_d = 0$, $H_A$: $\mu_d \neq 0$

(2) $d$의 검정통계량

$$t = \frac{\bar{d}-\mu_d}{S_d/\sqrt{n}}$$

(3) For $\alpha$=0.05,

Rejection Region: $t > t_{0.025,\,9} = 2.262$ 또는 $t < -t_{0.025,\,9} = -2.262$

(4) $t = \frac{\bar{d}-\mu_d}{S_d/\sqrt{n}} = \frac{3}{3.2318/\sqrt{10}} = 2.9355$

(5) Conclusion: Reject $H_0$.

[참고] p-value = 2・$P(t_9 > |2.9355|)$ = 2・0.0083 = 0.0166

[참고] $\mu_d$의 95% 신뢰구간

$\bar{d} \pm t_{\frac{\alpha}{2}, n-1} \cdot S_d / \sqrt{n}$ = 3 ± 2.262・3.2318 / $\sqrt{10}$ = [0.6883, 5.3117]

## 연습문제 정답

1. (1) $x_1 \sim N(5, 4^2)$이므로, 평균은 5, 분산은 $4^2$이다.

   (2) $x_1 \sim N(5, 4^2)$이고, $n_1$ = 2이므로, 평균은 5, 분산은 $\frac{4^2}{2}$이다.

   (3) $x_2 \sim N(-5, 3^2)$이므로, 평균은 -5, 분산은 $3^2$이다.

   (4) $x_2 \sim N(-5, 3^2)$이고, $n_2$ = 2이므로, 평균은 -5, 분산은 $\frac{3^2}{2}$이다.

   (5) $E(\overline{x_1} + \overline{x_2})$ = $E(\overline{x_1})$ + $E(\overline{x_2})$ = $E(x_1)$ + $E(x_2)$ = 5 + (-5) = 0

   $E(\overline{x_1} - \overline{x_2})$ = $E(\overline{x_1})$ - $E(\overline{x_2})$ = $E(x_1)$ -+ $E(x_2)$ = 5 - (-5) = 10

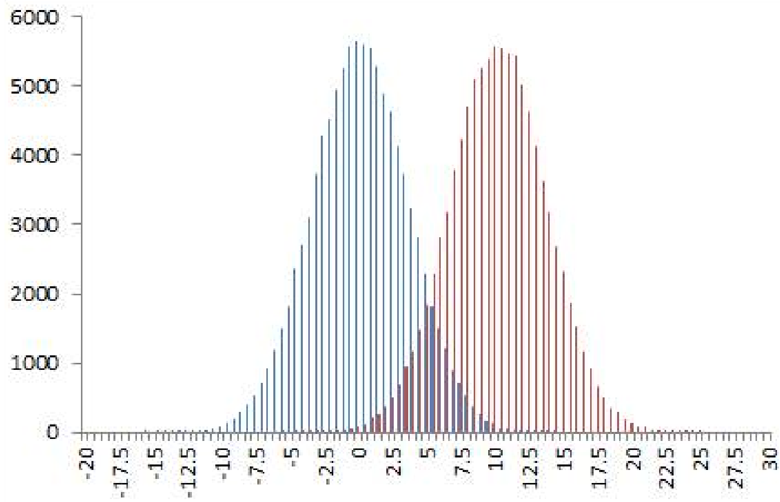
   (6) $V(\overline{x_1} + \overline{x_2})$ = $V(\overline{x_1})$ + $V(\overline{x_2})$ = $\frac{V(x_1)}{n_1}$ + $\frac{V(x_2)}{n_2}$ = $\frac{16}{2} + \frac{9}{2}$= 12.5

   $V(\overline{x_1} - \overline{x_2})$ = $V(\overline{x_1})$ + $V(\overline{x_2})$ = $\frac{V(x_1)}{n_1}$ + $\frac{V(x_2)}{n_2}$ = $\frac{16}{2} + \frac{9}{2}$= 12.5

   주의: $V(\overline{x_1} + \overline{x_2})$와 $V(\overline{x_1} - \overline{x_2})$는 동일하다. (단, 두 모집단이 서로 독립적이라면)

   (7) $x_1$과 $x_2$ 모두 정규분포를 따르므로, $\overline{x_1}$과 $\overline{x_2}$는 정규분포를 따른다.

   $\overline{x_1}$과 $\overline{x_2}$가 정규분포를 따르므로, $\overline{x_1} + \overline{x_2}$와 $\overline{x_1} - \overline{x_2}$는 모두 정규분포를 따른다.



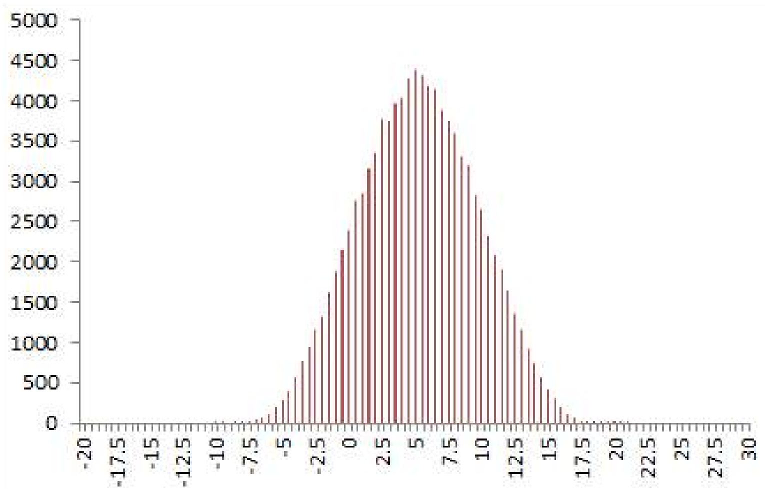

Note: 왼쪽 그래프는 $\overline{x_1}+\overline{x_2}$의 도수분포를, 오른쪽 그래프는 $\overline{x_1}-\overline{x_2}$의 도수분포이다. 두 그래프 모두 정규분포이다. 두 그래프의 중앙은 다르지만, 퍼져 있는 정도는 동일하다.

2. (1) $E(\overline{x})$ = $E(x)$ = $\dfrac{20-0}{2}$ = 10, $E(\overline{y})$ = $E(y)$ = 5

$V(\overline{x})$ = $\dfrac{V(x)}{n_x}$ = $\dfrac{(20-0)^2/12}{2}$ = 16.67, $V(\overline{y})$ = $\dfrac{V(y)}{n_y}$ = $\dfrac{3^2}{3}$ = 3

(2) $E(\overline{x}-\overline{y})$ = $E(\overline{x})$ - $E(\overline{y})$ = 5.0, $V(\overline{x}-\overline{y})$ = $V(\overline{x})$ + $V(\overline{y})$ = 19.67

(3) 잘 알려진 정규분포는 아님. (정규분포는 아님)



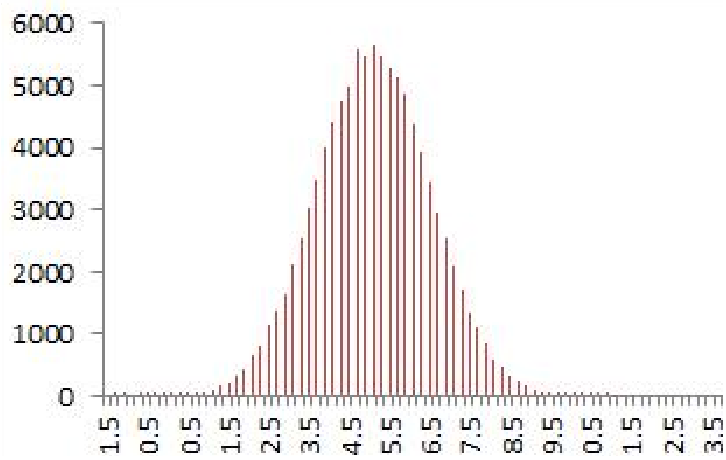

(4) $E(\overline{x}-\overline{y})$ = $E(\overline{x})$ - $E(\overline{y})$ = 5.0,

$V(\overline{x}-\overline{y})$ = $V(\overline{x})$ + $V(\overline{y})$ = $\dfrac{V(x)}{n_x}$ + $\dfrac{V(y)}{n_y}$ = $\dfrac{(20-0)^2/12}{30}$ + $\dfrac{3^2}{10}$ = 2.0111

(5) $n_x$ = 30로서 충분히 큰 값이므로 $\overline{x}$는 정규분포를 따른다.

$n_y$ = 10로서 충분히 큰 값이 아니지만 $y$가 정규분포를 따르므로 $\overline{y}$는 표본의 크기와 무관하게 정규분포를 따른다.

$\overline{x}$와 $\overline{y}$ 모두 정규분포를 따르므로 $\overline{x}$-$\overline{y}$도 정규분포를 따른다.




3. (1) $\overline{X_1}$, $\overline{X_2}$, $n_1$, $n_2$

   (2) $\mu_1$ - $\mu_2$ (이 값은 $H_0$에서 정한다.)

   (3) 표본으로부터 구할 수는 없는 값 (즉, 알려져 있어야 하는 값): $\sigma_1^2$, $\sigma_2^2$

4. (1) $H_0$: $\mu_1 - \mu_2 = 0$, $H_A$: $\mu_1 - \mu_2 \neq 0$

   Test Statistic: $z = \dfrac{(\overline{X_1} - \overline{X_2}) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$

   Rejection Region: $z > z_{0.025} = 1.96$ 또는 $z < -z_{0.025} = -1.96$

   Value of the Test Statistic: $z = \dfrac{(\overline{X_1} - \overline{X_2}) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} = \dfrac{-50 - (0)}{\sqrt{\dfrac{100^2}{10} + \dfrac{50^2}{5}}} = -1.29$

   Conclusion: Do not Reject $H_0$ .

   (2) $(\overline{X_1}$- $\overline{X_2}) \pm z_{\frac{\alpha}{2}} \cdot \sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}$ = -50 ± 1.96(38.73) = -50 ± 75.91 (-125.91, 25.91)

   (3) 신뢰구간에 $H_0$에서 설정한 $\mu_1$-$\mu_2$ 값인 0이 포함되므로, 양측 검정의 $H_0$는 기각 실패

   (4) 양측검정이므로 $p$-value = 2 × $P(z > |-1.29|)$ = 2×0.0985 = 0.1971

   (5) $p$-value가 $\alpha$보다 크므로, $H_0$를 기각할 수 없다.

5. (1) a. $\overline{X_1}$- $\overline{X_2}$이 정규분포를 따른다.

   ① $X_1$과 $X_2$가 정규분포를 따르거나

   ② $n_1$과 $n_2$가 충분히 크다.

   이 두 조건 중 최소 하나는 충족되어야 한다. (두 조건들은 모두 $\overline{X_1}$과 $\overline{X_2}$가

정규분포를 따를 조건이며, $\overline{X_1}$과 $\overline{X_2}$가 정규분포를 따르면 $\overline{X_1}$- $\overline{X_2}$도 정규분포를 따른다.)

b. $\sigma_1^2$, $\sigma_2^2$이 알려져 있다. ($\sigma_1^2$와 $\sigma_2^2$는 모수이므로, 표본으로부터 구할 수 없다.)

(2) $\sigma_1$과 $\sigma_2$은 표본으로부터 $s_1$과 $s_2$로 추정한다.

(3) $\overline{X} = \dfrac{3.0+5.0+5.8+6.2}{4} = 5$

$$s^2 = \frac{(3.0-5.0)^2+(5.0-5.0)^2+(5.8-5.0)^2+(6.2-5.0)^2}{4-1} = 2.0267$$

$$s = \sqrt{s^2} = 1.4236$$

$$s_{\overline{X}} = \frac{s}{\sqrt{n}} = \frac{1.4236}{\sqrt{4}} = 0.7118$$

(4) $z = \dfrac{(\overline{X_1}-\overline{X_2})-\mu_{\overline{X_1}-\overline{X_2}}}{S_{\overline{X_1}-\overline{X_2}}} = \dfrac{(\overline{X_1}-\overline{X_2})-(\mu_1-\mu_2)}{\sqrt{\dfrac{S_1^2}{n_1}+\dfrac{S_2^2}{n_2}}}$

($\sigma^2_{\overline{X_1}-\overline{X_2}}$을 모르는 경우 표준통계량으로 $t_{n_1+n_2-2}$를 사용한다. 그런데 $t$의 자유도가 커지면 $t$값은 $z$값에 근접하므로, $t$대신 $z$를 사용해도 차이가 별 차이가 없다.)

6. (1) $S_{\overline{X_1}} = \dfrac{S_1}{\sqrt{n_1}} = \dfrac{4.0}{\sqrt{25}} = 0.8$, $S_{\overline{X_2}} = \dfrac{S_2}{\sqrt{n_2}} = \dfrac{6.0}{\sqrt{36}} = 1$

$$S_{\overline{X_1}-\overline{X_2}} = \sqrt{S^2_{\overline{X_1}}+S^2_{\overline{X_2}}} = \sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}} = \sqrt{\frac{4^2}{25}+\frac{6^2}{36}} = \sqrt{1.64} = 1.2806$$

(2) $z$로 하는 경우

$H_0$: $\mu_1-\mu_2 = 0$, $H_A$: $\mu_1-\mu_2 \neq 0$

Test Statistic: $z = \dfrac{(\overline{X_1}-\overline{X_2})-\mu_{\overline{X_1}-\overline{X_2}}}{S_{\overline{X_1}-\overline{X_2}}} = \dfrac{(\overline{X_1}-\overline{X_2})-(\mu_1-\mu_2)}{\sqrt{\dfrac{S_1^2}{n_1}+\dfrac{S_2^2}{n_2}}}$

Rejection Region: $z > z_{0.025} = 1.96$ 또는 $z < -z_{0.025} = -1.96$

Value of the Test Statistic: $z = \dfrac{(3.5-4.0)-0}{1.2806} = -0.39043$

Conclusion: $H_0$를 기각하는 데 실패

$t$로 하는 경우

$H_0$: $\mu_1-\mu_2 = 0$, $H_A$: $\mu_1-\mu_2 \neq 0$

Test Statistic: $t_{n_1+n_2-2} = \dfrac{(\overline{X_1}-\overline{X_2})-\mu_{\overline{X_1}-\overline{X_2}}}{S_{\overline{X_1}-\overline{X_2}}} = \dfrac{(\overline{X_1}-\overline{X_2})-(\mu_1-\mu_2)}{\sqrt{\dfrac{S_1^2}{n_1}+\dfrac{S_2^2}{n_2}}}$

Rejection Region: $t > t_{59,\,0.025} = 2.0010$, 또는 $t < -t_{59,\,0.025} = -2.0010$)

Value of the Test Statistic: $t = \dfrac{(3.5-4.0)-0}{1.2806} = -0.39043$

Conclusion: $H_0$를 기각하는 데 실패

(3) $(\overline{X_1} - \overline{X_2}) \pm z_{\frac{\alpha}{2}} \cdot S_{\overline{X_1}-\overline{X_2}} = -0.5 \pm 1.96(1.2806)$

$t$로 한다면

$(\overline{X_1} - \overline{X_2}) \pm t_{59,\,0.025} \cdot S_{\overline{X_1}-\overline{X_2}} = -0.5 \pm 2.0010(1.2806)$

(4) 신뢰구간이 $H_0$: $\mu_1 - \mu_2$값인 0을 포함하므로, $H_0$를 기각하는 데 실패

(5) 양측이므로 $p$-value $= 2 \times P(z > |-0.39043|) = 2\times0.348109 = 0.696219$

$t$로 하면

$p$-value $= 2 \times P(t_{59} > |-0.39043|) = 2\times0.3488 = 0.6976$

(6) $p$-value가 $\alpha$보다 크므로, $H_0$를 기각할 수 없다.

7. (1) $S_1^2 = \dfrac{\sum_{i=1}^{10}(x_{1i}-\overline{x_1})^2}{10-1} = \dfrac{597.12}{9} = 66.35$, $S_2^2 = \dfrac{\sum_{i=1}^{5}(x_{2i}-\overline{x_2})^2}{5-1} = \dfrac{346.77}{4} = 86.69$

(2) $S_1^2$와 $S_2^2$의 기대값이 같다는 의미이며 또한 표본 1의 편차제곱값 0.54, 0.01, ...., 80.60과 표본 2의 편차제곱들 158.04, 2.63, ..., 42.39들의 기대값이 같다는 의미 (두 뜻은 동일한 의미이다.)

(3) 두 표본의 크기가 다르므로, 가중평균으로 구한다. 단, $S_1^2$을 구할 때는 10-1로 나누었고 $S_2^2$를 구할 때는 5-1로 나누었으므로, 의 가중치는 $\dfrac{9}{9+4}$, $S_2^2$의 가중치는 $\dfrac{4}{9+4}$로 한다.

이를 식으로 표현하면, 모집단 분산 추정치 = $\dfrac{9}{9+4}S_1^2 + \dfrac{4}{9+4}S_2^2$이 된다.

(4) 표본들의 편차제곱의 기대값은 모집단 분산이므로, 편차제곱들의 합을 적절한 수로 나누면 된다. 여기서 적절한 수란 (10-1)+(5-1)이다.

$S_1^2$을 구할 때는 10-1로, $S_2^2$를 구할 때는 5-1로 나누었음을 상기하자.

이를 수식으로 표현하면 $\dfrac{\sum_{i=1}^{10}(x_{1i}-\overline{x_1})^2 + \sum_{i=1}^{5}(x_{2i}-\overline{x_2})^2}{(10-1)+(5-1)}$이다.

(5) $\dfrac{9}{9+4}S_1^2 + \dfrac{4}{9+4}S_2^2 = \dfrac{9}{9+4}\dfrac{\sum_{i=1}^{10}(x_{1i}-\overline{x_1})^2}{10-1} + \dfrac{4}{9+4}\dfrac{\sum_{i=1}^{5}(x_{2i}-\overline{x_2})^2}{5-1}$

$= \dfrac{\sum_{i=1}^{10}(x_{1i}-\overline{x_1})^2}{(10-1)+(5-1)} + \dfrac{\sum_{i=1}^{5}(x_{2i}-\overline{x_2})^2}{(10-1)+(5-1)} = \dfrac{\sum_{i=1}^{10}(x_{1i}-\overline{x_1})^2 + \sum_{i=1}^{5}(x_{2i}-\overline{x_2})^2}{(10-1)+(5-1)}$

8. (1) $S_p^2 = \dfrac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{(n_1-1)+(n_2-1)} = \dfrac{(16-1)16.18^2+(11-1)17.3^2}{(16-1)+(11-1)} = 276.7914$

(2) $H_0$: $\mu_1 - \mu_2 = 0$, $H_A$: $\mu_1 - \mu_2 \neq 0$

Test Statistic: $t = \dfrac{(\overline{X}_1 - \overline{X}_2) - \mu_{\overline{X}_1 - \overline{X}_2}}{S_{\overline{X}_1 - \overline{X}_2}} = \dfrac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{S_p^2}{n_1} + \dfrac{S_p^2}{n_2}}}$

Rejection Region: $t > t_{0.025,\,25} = 2.38461$ 또는 $t < -t_{0.025,\,25} = -2.38461$

Value of the Test Statistic: $t = \dfrac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{S_p^2}{n_1} + \dfrac{S_p^2}{n_2}}}$

$= \dfrac{(95.6 - 101.2) - 0}{\sqrt{\dfrac{276.7914}{16} + \dfrac{276.7914}{11}}} = -0.8594$

Conclusion: $H_0$ 기각 실패

(3) $(\overline{X}_1 - \overline{X}_2) \pm t_{0.025,\,25} \cdot \sqrt{\dfrac{S_p^2}{n_1} + \dfrac{S_p^2}{n_2}} \rightarrow -5.6 \pm 2.38461(6.516312) = [-21.1389, 9.938864]$

(4) 신뢰구간이 $H_0$에서 설정한 $\mu_1 - \mu_2$의 값인 0을 포함하므로, $H_0$를 기각할 수 없다.

(5) 양측이므로 $p$-value $= 2 \cdot P(t_{25} > |-0.8594|) = 2 \cdot 0.1991 = 0.3983$

(6) $p$-value가 $\alpha$보다 크므로, $H_0$를 기각할 수 없다.