10. 두 모집단의 비교

10.1

(1) $\sqrt{\dfrac{8\times1.3^2+11\times1.5^2}{9+12-2}}=1.419$

(2) $\bar{x}\pm t_{0.025,8}\,s/\sqrt{m}=5.8\pm2.306\times1.3/\sqrt{9}=$[4.801, 6.799]

(3) 가설: $H_0:\sigma_A^2\geq\sigma_B^2$ vs $H_1:\sigma_A^2<\sigma_B^2$

- 검정통계량: $f=1.3^2/1.5^2=0.751>F_{0.95,8,11}=0.302$이므로 5% 유의수준에서 $H_0$ 기각할 수 없음
  ⇒ A의 분산이 B의 분산보다 작다고 할 수 없음

(4) $\mu$: 평균진통효과시간

- 가설: $H_0:\mu_A\leq\mu_B$ vs $H_1:\mu_A>\mu_B$
- 검정통계량: $t=\dfrac{5.8-5}{1.419\sqrt{1/9+1/12}}=1.2785<t_{0.05,19}=1.729$이므로 5% 유의수준에서 $H_0$ 기각할 수 없음 ⇒ A의 진통효과가 B의 진통효과보다 있다고 할 수 없음

10.2:

(1) 새민련: $\bar{x}=\dfrac{520.73}{9}=57.859$, $s_x^2=\dfrac{1}{8}(30989.169-9\times57.859^2)=95.037$

⇒ $\bar{x}\pm t_{0.025,8}\sqrt{s_x^2/m}=57.859\pm2.306\sqrt{95.037/9}=$[50.365, 65.352]

새누리: $\bar{y}=\dfrac{468.95}{8}=58.619$, $s_y^2=\dfrac{1}{7}(28115.465-8\times58.619^2)=89.4575$

⇒ $\bar{y}\pm t_{0.025,7}\sqrt{s_y^2/n}=58.619\pm2.365\sqrt{89.4575/8}=$[50.712, 66.526]

(2) 가설: $H_0:\mu_x=\mu_y$ vs $H_1:\mu_x\neq\mu_y$

- $s_y^2/s_x^2=0.9413$으로 분산이 같다고 할 수 있으므로 합동분산추정치는 $s_p^2=\dfrac{8\times s_x^2+7\times s_y^2}{9+8-2}=92.433$
- $|t|=\left|\dfrac{58.619-57.859}{\sqrt{92.433(1/9+1/8)}}\right|=0.163<t_{0.025,15}=2.131$이므로 5% 유의수준에서 $H_0$ 기각할 수 없음 ⇒ 두 당의 평균득표율에 차이가 있다고 보기 어려움

10.3:

(1) 가설 $H_0:\sigma_x^2=\sigma_y^2$ vs $H_1:\sigma_x^2\neq\sigma_y^2$

- $m=18, n=16, s_x^2=0.2392, s_y^2=0.1307$
- $s_x^2/s_y^2=1.8312<F_{0.025,17,15}=2.813$이므로 5% 유의수준에서 $H_0$ 기각할 수 없음
  ⇒ 두 기간의 최저방어률 분산에 차이가 없음

(2) 문제수정: 유의수준 ⇒ 신뢰구간

(1)의 결과 분산이 차이가 없는 것으로 나타났기 때문에 $\mu_x-\mu_y$에 대한 95% 신뢰구간은

$\bar{x}-\bar{y}\pm t_{0.025,m-n-2}\,s_p\sqrt{\dfrac{1}{m}+\dfrac{1}{n}}=1.6178-2.5138\pm2.037\times0.4892\sqrt{1/18+1/16}=$[-1.1997,-0.5922]

(3) 등분산성 검정 결과: $s_x^2/s_y^2=2.8338>F_{0.025,17,15}=2.8128$ ⇒ 분산이 같다고 할 수 없음

- 가설: $H_0 : \mu_x = \mu_y$ vs $H_1 : \mu_x \neq \mu_y$
- $\overline{x}$ =0.3561, $s_x^2$ =0.0006, 이후: $\overline{y}$ =0.3529, $s_y^2$ =0.0002
- 분산이 다름 ⇨ 자유도 $\min(m-1, n-1)$ 사용
- 검정통계량: $|t| = \left| \dfrac{0.3561 - 0.3529}{\sqrt{0.0006/18 + 0.0002/16}} \right| = 0.4815 < t_{0.025,15} = 2.131$이므로 5% 유의수준에서 $H_0$ 기각할 수 없음 ⇒ 두 기간에 따라 최고타율의 평균에 차이가 있다고 할 수 없음

(4) **프로그램 "Exercise10.R" 참조**

10.4 **늘어난 수면 시간 및 ⓑ-ⓐ의 분포는 정규분포를 따른다고 가정함**

(1) $m = 10$, $\overline{x} = 0.66$, $s_x^2 = 3.4516$, $s_x = 1.8578$, $t_{0.025,9} = 2.262$

- $\overline{x} \pm t_{\alpha/2, m-1} s_x / \sqrt{n} = 0.66 \pm 2.262 \times 1.8578 / \sqrt{10} = [-0.669, 1.989]$

(2) 가설: $H_0 : \mu_B \leq 0$ vs $H_1 : \mu_B > 0$

- $n = 10$, $\overline{y} = 2.33$, $s_y^2 = 4.009$, $s_y = 2.0022$
- 검정통계량: $t = \dfrac{2.33}{2.0022/\sqrt{10}} = 3.6799 > t_{0.05,9} = 1.833$이므로 5% 유의수준에서 $H_0$ 기각할 수 있음 ⇒ 처치 ⓑ를 했을 때 수면시간이 늘었다고 할 수 있음

(3) $\delta = \mu_y - \mu_x$라고 하면 $H_0 : \delta \leq 0$ vs $H_1 : \delta > 0$

- $\overline{d} = 1.67$, $s_d = 1.1304$
- 검정통계량: $t = \dfrac{1.67}{1.1304/\sqrt{10}} = 4.672 > t_{0.05,9} = 1.833$이므로 5% 유의수준에서 $H_0$ 기각 ⇒ 처치 ⓑ가 유의하게 처치 ⓐ보다 수면시간을 늘려주는 효과가 있다고 할 수 있음

(4) **16.6을 17.6으로 수정**

① $\delta = \mu_y - \mu_x$라고 하면 $H_0 : \delta \leq 0$ vs $H_1 : \delta > 0$

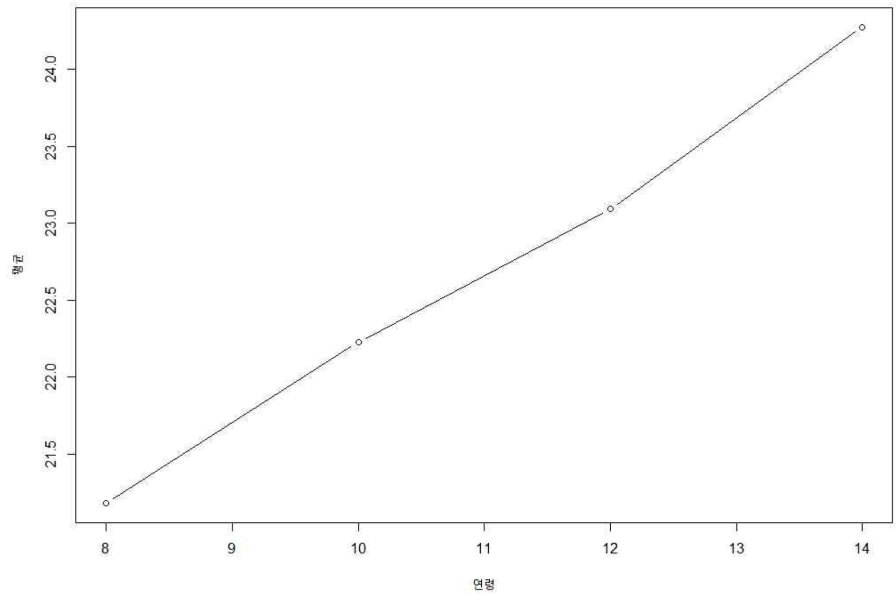
- $\overline{d} = 2.97$, $s_d = 5.1616$
- 검정통계량: $t = \dfrac{2.97}{5.1616/\sqrt{10}} = 1.8196 < t_{0.05,9} = 1.833$이므로 5% 유의수준에서 $H_0$ 기각 못함 ⇒ 처치 ⓑ가 유의하게 처치 ⓐ보다 수면시간을 늘려주는 효과가 있다고 할 수 없음

② 차가 모두 양수임에도 불구하고 유의하지 않은 것은 17.6이라는 극단적인 이상치가 자료에 포함되었기 때문. 이 이상치는 검정통계량의 분자인 평균보다 분모에 있는 표준편차의 값이 상태적으로 커져 유의하지 않게 만듬

(5) 상자그림 결과 9번째 자료가 4.6일 때도 이상치로 판정되었고 자료의 크기도 크지 않아 비모수적인 방법을 적용하는 것이 적절함

10.5: 가정추가: 거리는 정규분포를 따름

(1) ①




② $d_i = x_{10,i} - x_{8,i}$라고 하면

- $n$ = 11, $\overline{d}$ =1.04545, $s_d$ = 1.1928, $t_{0.025,10}$ =2.228
- $\overline{d} \pm t_{0.025,n-1}\, s_d / \sqrt{n}$ = [0.2441, 1.8468]

③ $\delta = \mu_{14} - \mu_8$ 라고 하면 가설: $H_0 : \delta \le 0$ vs $H_1 : \delta > 0$

- $\overline{d}$ = 3.0909, $s_d$ = 1.3194
- 검정통계량: $t = \dfrac{3.0909}{1.3194/\sqrt{11}} = 7.77 > t_{0.05,9} = 1.812$이므로 5% 유의수준에서 $H_0$ 기각

  ⇒ 14세 때 거리가 8세 때보다 늘었다고 할 수 있음

(2) $\delta = \mu_{14} - \mu_8$ 라고 하면 $H_0 : \delta \le 0$ vs $H_1 : \delta > 0$

- $\overline{d}$ = 4.5938, $s_d$ = 2.6722
- 검정통계량: $t = \dfrac{4.4938}{2.6722/\sqrt{16}} = 6.876 > t_{0.05,15} = 1.753$이므로 5% 유의수준에서 $H_0$ 기각

  ⇒ 14세 때 거리가 8세 때보다 늘었다고 할 수 있음

(3) $\sigma_x^2$ : 8세와 14세 때의 여자 길이차이 분산, $\sigma_y^2$: 8세와 14세 때의 남자 길이차이 분산

① $x_i = x_{i,14} - x_{i,8}$ , $y_i = y_{i,14} - y_{i,8}$ 라고 하면 $s_x^2 = 1.7409$ , $s_y^2 = 7.1406$ ⇨ $f = s_x^2 / s_y^2 = 0.2438$

- $\sigma_x^2/\sigma_y^2$의 95% 신뢰구간: $\left[ \dfrac{f}{F_{0.025,10,15}}, \dfrac{f}{F_{0.975,10,15}} \right]$=[0.0797, 0.8586]

② $\mu_x$ : 8세와 14세 때의 여자 길이차이 평균, $\mu_y$: 8세와 14세 때의 남자 길이차이 평균

- 가설: $H_0 : \mu_x = \mu_y$ vs $H_1 : \mu_x \ne \mu_y$
- 두 분산이 다르기 때문에 검정통계량: 자유도 $\min(m-1, n-1)$ 사용

  $|t| = \left| \dfrac{3.0909 - 4.5936}{\sqrt{\dfrac{1.7409}{11} + \dfrac{7.1407}{16}}} \right| = 1.9328 < 2.228 = t_{0.025,10}$ 이므로 5% 유의수준에서 $H_0$ 기각 못함

  ⇒ 14세와 8세 때의 거리차이 평균은 남녀간에 다르다고 할 수 없음

(4) $\mu_x$ : 14세 때의 여자 평균거리, $\mu_y$: 14세 때의 남자 평균거리

- 가설: $H_0 : \mu_x = \mu_y$ vs $H_1 : \mu_x \neq \mu_y$
- 분산비의 95% 신뢰구간이 [0.4560, 4.9139]로 차이가 없다고 할 수 있음
- 검정통계량: $|t| = 3.6359 > t_{0.025,25} = 2.060$ 이므로 5% 유의수준에서 $H_0$ 기각

  ⇒ 14세 때의 남녀 간 거리평균에 차이가 있다고 할 수 있음

10.6 추가가정: 시험점수는 정규분포를 따름

| 번호 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 중간점수(x) | 65 | 72 | 93 | 87 | 77 | 55 |
| 기말점수(y) | 69 | 70 | 99 | 91 | 88 | 62 |
| x-y | -4 | 2 | -6 | -4 | -11 | -7 |

- 가설: $H_0 : \mu_x = \mu_y$ vs $H_1 : \mu_x < \mu_y$
- $\bar{d} = -5$, $s_d^2 = (242 - 6 \times (-5)^2)/5 = 18.4$
- 검정통계량: $t = \dfrac{-5}{\sqrt{18.5/6}} = -2.855 < -2.015 = -t_{0.05,5}$ 이므로 5% 유의수준에서 $H_0$ 기각

  ⇒ 기말시험 점수가 중간시험 점수보다 높다고 할 수 있음

10.7 표본크기가 50, 100으로 중심극한정리를 사용할 수 있음

(1) $\mu_A - \mu_B$에 대한 95% 신뢰구간: $155 - 151 \pm 1.96\sqrt{\dfrac{49}{50} + \dfrac{36}{100}}$ =[1.731, 6.269]

(2) $H_0 : \mu_A = \mu_B$ vs $H_1 : \mu_A > \mu_B$

- 검정통계량: $\dfrac{155 - 151}{\sqrt{\dfrac{49}{50} + \dfrac{36}{100}}} = 3.455 > 1.645 = z_{0.05}$ 이므로 5% 유의수준에서 $H_0$ 기각

  ⇒ A회사의 초코파이 평균열량이 B회사의 것보다 높다고 할 수 있음

10.8

(1) 봄·겨울: 962명, 여름·가을: 900

① $\theta$: 봄·겨울 출생비율

- 가설: $H_0 : \theta \le 0.5$ vs $H_1 : \theta > 0.5$
- $n = 1862$, $p = 0.5166$ ⇨ 검정통계량: $z = \dfrac{0.5166 - 0.5}{\sqrt{\dfrac{0.5 \times 0.5}{1862}}} = 1.437 < 1.645 = z_{0.05}$ 이므로 5% 유의수준에서 $H_0$ 기각못함 ⇒ 봄과 겨울의 출생비율이 여름과 가을의 출생비율보다 높다고 할 수 없음

② 봄: 480 ⇨ $p = 480/1862 = 0.2578$

- $p \pm z_{0.025}\sqrt{\dfrac{p(1-p)}{n}} = 0.2578 \pm 1.96\sqrt{\dfrac{0.2578(1-0.2578)}{1862}}$ =[0.2379, 0.2777]

(2) $\theta_1$ : A 지역 여름출생비율, $\theta_2$ : B지역 여름출생비율

- $p_1 = 454/1862 = 0.2458$, $p_2 = 53/188 = 0.2819$

- $p_1 - p_2 \pm z_{0.025}\sqrt{\frac{p_1(1-p_2)}{n_1}+\frac{p_2(1-p_2)}{n_2}}$ =[-0.1053, 0.0291]

(3) $\theta_1$ : A 지역 겨울출생비율, $\theta_2$ : B지역 겨울출생비율

- 가설: $H_0 : \theta_1 = \theta_2$ vs $H_1 : \theta_1 \neq \theta_2$
- $n_1 = 1862,\ p_1 = 482/1862 = 0.2589,\ n_2 = 188,\ p_2 = 40/188 = 0.2128,\ p = 522/2050 = 0.2546$
- 검정통계량: $z = \frac{0.2589 - 0.2128}{\sqrt{0.2546(1-0.2546)\left(\frac{1}{1862}+\frac{1}{88}\right)}} = 1.3828$ ⇨ $|z| < 1.96 = z_{0.025}$이므로 5% 유의수준에서 $H_0$ 기각못함 ⇒ A 지역 겨울출생비율과 B지역 겨울 출생비율에 유의한 차이가 없음

10.9

(1) $\sqrt{\frac{\theta_1(1-\theta_1)}{n_1}}$ , $\sqrt{\frac{\theta_2(1-\theta_2)}{n_2}}$

(2) $E(P_1 - P_2) = E(P_1) - E(P_2) = \theta_1 - \theta_2$ ,

$$Var(P_1 - P_2) = Var(P_1) + Var(P_2) = \frac{\theta_1(1-\theta)}{n_1} + \frac{\theta_2(1-\theta_2)}{n_2}$$

$$\Rightarrow SE(P_1 - P_2) = \sqrt{\frac{\theta_1(1-\theta)}{n_1} + \frac{\theta_2(1-\theta_2)}{n_2}}$$

(3) $Var(P) = \frac{1}{(n_1+n_2)^2}\{Var(X) + Var(Y)\} = \frac{1}{(n_1+n_2)^2}\{n_1\theta(1-\theta) + n_2\theta(1-\theta)\} = \frac{\theta(1-\theta)}{n_1+n_2}$

$$\Rightarrow SE(P) = \sqrt{\frac{\theta(1-\theta)}{n_1+n_2}}$$

10.10

(1) $p_1 = \frac{119}{11037} = 0.0108,\ p_2 = \frac{98}{11034} = 0.0089$

- $\theta_1$: $p_1 \pm 1.96\sqrt{\frac{p_1(1-p_1)}{11037}}$ =[0.0089, 0.0127]
- $\theta_2$: $p_2 \pm 1.96\sqrt{\frac{p_2(1-p_2)}{11034}}$ =[0.0071, 0.0107]

(2) $0.0108 - 0.0089 \pm 1.96\sqrt{\frac{0.0108(1-0.0108)}{11037}+\frac{0.0089(1-0.0089)}{11034}}$ =[-0.0007, 0.0045]

(3) 가설: $H_0 : \theta_1 = \theta_2$ vs $H_1 : \theta_1 > \theta_2$

- 귀무가설 하에서의 $\theta_1 = \theta_2 = \theta$ 추정 : $p = (119+98)/(11037+11034) = 0.0098$
- 검정통계량: $\frac{0.0108 - 0.0089}{\sqrt{0.0098(1-0.0098)}\sqrt{\frac{1}{11037}+\frac{1}{11034}}} = 1.433 < 1.645 = z_{0.05}$이므로 5%

유의수준에서 $H_0$ 기각못함 ⇒ 아스피린을 복용한 사람들이 위약을 복용한 사람들보다 뇌졸중에 걸릴 확률이 높다고 할 수 없음

10.11

(1) $\mu_x$: 민간부분 대졸 남성정규직 평균, $\mu_y$: 공공부분 대졸 남성정규직 평균

- 가설: $H_0 : \mu_x = \mu_y$ vs $H_1 : \mu_x \neq \mu_y$
- $m = 1081$, $\overline{x} = 2266$, $s_x^2 = 562^2$, $n = 235$, $\overline{y} = 2610$, $s_y^2 = 529^2$
- 검정통계량: $z = \dfrac{2266 - 2610}{\sqrt{\dfrac{562^2}{1081} + \dfrac{529^2}{235}}} = -8.933$ ⇨ $|z| > z_{0.025} = 1.96$이므로 5% 유의수준에서 $H_0$ 기각

  ⇒ 민간부분과 공공부분의 대졸 남성정규직 평균연봉 간에 유의한 차이가 있음

(2) $\mu_x$: 노조사업장 대졸 남성정규직 평균, $\mu_y$: 비노조사업장 대졸 남성정규직 평균

- 가설: $H_0 : \mu_x = \mu_y$ vs $H_1 : \mu_x \neq \mu_y$
- $m = 829$, $\overline{x} = 2693$, $s_x^2 = 531^2$, $n = 576$, $\overline{y} = 2449$, $s_y^2 = 488^2$
- 검정통계량: $z = \dfrac{2693 - 2449}{\sqrt{\dfrac{531^2}{829} + \dfrac{488^2}{576}}} = 8.897$ ⇨ $|z| > z_{0.025} = 1.96$이므로 5% 유의수준에서 $H_0$ 기각

  ⇒ 노조사업장과 비노조사업장의 대졸 남성정규직 평균연봉 간에 유의한 차이가 있음