MAT2410 (응용수학I)
과제 1-3

서강대학교 컴퓨터공학과 박수현 (20181634)

서강대학교 컴퓨터공학과

7 표본분포

7.1 모집단 분포와 표본분포

1. 전구를 생산하는 생산라인에서 하루 동안 발생하는 불량품의 수를 알아보기 위하여, 임의로 10일을 선택하여 발생한 불량품의 수를 조사한 결과 다음과 같았다. 이 때 표본평균과 표본분산을 구하라.

0	3	1	3	4	2	0	1	1	2

Solution. $\overline{X} = 1.7$, $s^2 = 1.79$

7.2 표본평균의 분포

2. 어느 회사에서 제조된 전구의 수명은 평균 516시간, 분산 185시간이라 한다. 이 회사에서 제조된 전구를 100개 구입했을 때, 이 전구들의 평균수명에 대한 평균과 분산을 구하고, 평균수명이 520시간 이상일 근사확률을 구하라.

Solution. $\mathrm{E}\left(\overline{X}\right) = \mu = 516,\ s^2 = \dfrac{\sigma^2}{100} = 1.85$이다. 표본의 크기가 100이면 대표본이므로 중심극한정리를 적용한다면 평균수명은 정규분포에 근사하게 된다. 따라서

$$\begin{aligned} P\left(\overline{X} > 520\right) &= P\left(\frac{\overline{X} - 510}{\sqrt{1.85}} > \frac{520 - 510}{\sqrt{1.85}}\right) \\ &= P(Z > 2.94) = 0.0016 \end{aligned}$$

이다.

6. 우리나라에서 생산되는 어떤 종류의 담배 한 개에 포함된 타르[tar]의 양이 평균 5.5mg 표준편차 2.5mg이라고 한다. 어느 날 판매점에서 임의로 수거한 500개의 담배를 조사했을 때 다음을 구하라.

(1) 평균 타르가 5.6mg 이상일 근사확률
(2) 평균 타르가 5.3mg 이하일 근사확률

Solution. $\mathrm{E}\left(\overline{X}\right) = \mu = 5.5,\ s^2 = \dfrac{\sigma^2}{500} = 0.0125$이고 중심극한정리에 의해 $\overline{X} \approx \mathrm{N}(5.5, 0.0125)$이다.

(1)

$$\begin{aligned} P\left(\overline{X} > 5.6\right) &= P\left(\frac{\overline{X} - 5.5}{\sqrt{0.0125}} > \frac{5.6 - 5.5}{\sqrt{0.0125}}\right) \\ &= P(Z < 0.89) = 0.1867 \end{aligned}$$

(2)

$$\begin{aligned} P\left(\overline{X} < 5.3\right) &= P\left(\frac{\overline{X} - 5.5}{\sqrt{0.0125}} < \frac{5.3 - 5.5}{\sqrt{0.0125}}\right) \\ &= P(Z < -1.79) = 0.0367 \end{aligned}$$

11. A 교수의 과거 경험에 따르면 학생들의 통계학 점수를 평균 77점 그리고 표준편차 15점이라고 한다. 현재 이 교수는 36명과 64명인 두 반에서 강의하고 있다.

(1) 두 반의 평균성적이 72점과 82점 사이일 근사확률을 각각 구하라.
(2) 36명인 반의 평균성적이 64명인 반보다 2점 이상 더 큰 근사확률을 구하라.

Solution. 중심극한정리에 의해, 36명 반에 대해서 $\mathrm{E}\left(\overline{X}\right) = \mu = 77$, $s_1 = \dfrac{\sigma}{\sqrt{36}} = 2.5$이고 $\overline{X} \approx \mathrm{N}\left(77, 2.5^2\right)$이다. 또한 64명 반에 대해서 $\mathrm{E}\left(\overline{Y}\right) = \mu = 77$, $s_2^2 = \dfrac{\sigma}{\sqrt{64}} = 1.875$이고 $\overline{Y} \approx \mathrm{N}\left(77, 1.875^2\right)$이다.

(1)

$$\begin{aligned} P\left(72 < \overline{X} < 82\right) &= P\left(\frac{72-77}{2.5} < \frac{\overline{X}-77}{2.5} < \frac{82-77}{2.5}\right) \\ &= P(-2 < Z < 2) = 0.9544 \end{aligned}$$

$$\begin{aligned} P\left(72 < \overline{Y} < 82\right) &= P\left(\frac{72-77}{1.875} < \frac{\overline{Y}-77}{2.5} < \frac{82-77}{1.875}\right) \\ &= P(-2.67 < Z < 2.67) = 0.9924 \end{aligned}$$

(2) $T = \overline{X} - \overline{Y}$라 하면 $T \approx \mathrm{N}\left(77-77, 2.5^2 + 1.875^2\right) = \mathrm{N}\left(0, 3.125^2\right)$이다. 따라서

$$\begin{aligned} P\left(\overline{X} > \overline{Y} + 2\right) &= P\left(\overline{X} - \overline{Y} > 2\right) \\ &= P(T > 2) \\ &= P\left(\frac{T}{3.125} > \frac{2}{3.125}\right) \\ &= P(Z > 0.64) = 0.2611 \end{aligned}$$

이다.

13. 모평균 $\mu_1 = 550$, $\mu_2 = 500$이고 모표준편차 $\sigma_1 = 9$, $\sigma_2 = 16$인 두 정규모집단에서 각각 크기 50과 40인 표본을 임의로 추출하였을 때, 두 표본평균의 차가 48과 52 사이일 확률을 구하라.

Solution. 중심극한정리에 의해, 크기 50인 표본 $\overline{X}$에 대해 $\overline{X} \approx \mathrm{N}\left(550, \dfrac{9^2}{50}\right)$이고 크기 40인 표본 $\overline{Y}$에 대해 $\overline{Y} \approx \mathrm{N}\left(500, \dfrac{16^2}{40}\right)$이다. 따라서 $T = \overline{X} - \overline{Y}$는 $T \approx \mathrm{N}\left(550-500, \dfrac{9^2}{50} + \dfrac{16^2}{40}\right) = \mathrm{N}(50, 8.02)$이다.

두 표본평균의 차가 48과 52 사이이려면 $-52 < T < -48$ 또는 $48 < T < 52$이어야 한다. 첫 번째 경우

$$\begin{aligned} P(-52 < T < -48) &= P\left(\frac{-52-50}{\sqrt{8.02}} < \frac{T-50}{\sqrt{8.02}} < \frac{-48-50}{\sqrt{8.02}}\right) \\ &= P(-36.02 < Z < -34.61) = 0.0000 \end{aligned}$$

이고, 두 번째 경우

$$\begin{aligned} P(48 < T < 52) &= P\left(\frac{48-50}{\sqrt{8.02}} < \frac{T-50}{\sqrt{8.02}} < \frac{52-50}{\sqrt{8.02}}\right) \\ &= P(-0.71 < Z < 0.71) = 0.5223 \end{aligned}$$

이므로 두 표본평균의 차가 48과 52 사이일 확률은 0.5223이다.

7.3 정규모집단에 관련된 분포

2. 다음 표는 고소공포증에 걸린 환자 32명을 각각 16명씩 기존의 치료법과 새로운 치료법에 의하여 치료한 결과로 점수가 높을수록 치료의 효과가 있음을 나타낸다. 이 때 두 방법에 의한 치료 결과는 정규분포를 이룬다고 한다.

기존 치료법	$\bar{x} = 44.33$	$s_X^2 = 101.666$
새 치료법	$\bar{y} = 47.67$	$s_Y^2 = 95.095$

(1) 두 모분산이 동일하게 100이라 할 때, 합동표본분산이 170보다 작을 근사확률을 구하라.
(2) 두 표본에 대한 합동표본분산을 구하라.
(3) 두 모평균이 동일하다 할 때, 새로운 방법에 의한 치료 결과가 기존의 치료법보다 7.56점 이상 높을 근사확률을 구하라.
(4) 두 모분산이 동일하게 100이라 할 때, $P\left(S_Y^2 \geq (2.4)\, S_X^2\right)$를 구하라.

Solution.

(1) $\sigma^2 = 100$이고 표본의 크기가 16으로 같으므로 합동표본분산 $\frac{16+16-2}{100} S_p^2 \sim \chi^2(16+16-2) \Rightarrow 0.3S_p^2 \sim \chi^2(30)$이다. 따라서

$$P\left(S_p^2 \leq 170\right) = P\left(0.3S_p^2 \leq 51\right) \approx 0.99$$

이다.

(2) $s_p^2 = \frac{1}{16+16-2}[(16-1)\times 101.666 + (16-1)\times 95.095] = 98.3805$

(3) 기존 방법으로 얻은 평균을 $\overline{X}$, 새 방법으로 얻은 평균을 $\overline{Y}$라 하면

$$\frac{\overline{Y}-\overline{X}}{\sqrt{98.3805}\sqrt{\frac{1}{16}+\frac{1}{16}}} = \frac{\overline{Y}-\overline{X}}{3.51} \sim \mathrm{t}(28)$$

이다. 따라서

$$P\left(\overline{Y}-\overline{X} \geq 7.56\right) = P\left(\frac{\overline{Y}-\overline{X}}{3.51} \geq 2.155\right) \approx 0.02$$

이다.

(4) $\dfrac{\frac{S_Y^2}{100}}{\frac{S_X^2}{100}} = \dfrac{S_Y^2}{S_X^2} \sim \mathrm{F}(15, 15)$이다. 따라서

$$\begin{aligned} \mathrm{P}\left(S_Y^2 \geq 2.4 S_X^2\right) &= \mathrm{P}\left(\frac{S_Y^2}{S_X^2} \geq 2.4\right) \\ &\approx 0.05 \end{aligned}$$

이다.

3. 시중에서 판매되고 있는 두 회사의 땅콩 잼에 포함된 카페인의 양을 조사한 결과, 다음 표를 얻었다. 이 때 두 회사에서 제조된 땅콩 잼에 포함된 카페인의 양은 동일한 분산을 갖는 정규분포에 따른다고 한다. 단위는 mg이다.

A 회사	$n = 15$	$\bar{x} = 78$	$s_X^2 = 30.25$
B 회사	$m = 13$	$\bar{y} = 75$	$s_Y^2 = 36$

(1) 두 모분산이 동일하게 35라 할 때, 합동표본분산이 12.4보다 작을 근사확률을 구하라.
(2) 두 표본에 대한 합동표본분산을 구하라.
(3) 두 모평균이 동일하다 할 때, A 회사에서 제조된 땅콩 잼의 평균이 B 회사에서 제조된 평균보다 3.7mg 이하일 근사확률을 구하라.
(4) $\sigma_A^2 = 30$, $\sigma_B^2 = 35$일 때, $P\left(S_Y^2 \geq 3S_X^2\right)$를 구하라.

Solution.

(1) 합동표본분산 S_p^2에 대해 $\dfrac{15+13-2}{35} S_p^2 \sim \chi^2(26)$이다. 따라서

$$\begin{aligned} P\left(S_p^2 \leq 12.4\right) &= P\left(\frac{26}{35} S_p^2 \leq 9.21\right) \\ &\approx 0.001 \end{aligned}$$

이다.

(2) $s_p^2 = \dfrac{1}{15+13-2}\left[(15-1)\times 30.25 + (13-1)\times 36\right] = 32.9$

(3)

$$\frac{\overline{X}-\overline{Y}}{\sqrt{32.9}\sqrt{\frac{1}{15}+\frac{1}{13}}} = \frac{\overline{X}-\overline{Y}}{2.17} \sim \mathrm{t}(26)$$

이다. 따라서

$$P\left(\overline{X}-\overline{Y}\le 3.7\right)=P\left(\frac{\overline{X}-\overline{Y}}{2.17}\ge 1.705\right)$$
$$\approx 0.95$$

이다.

(4) $\dfrac{\frac{S_Y^2}{30}}{\frac{S_X^2}{35}}=\dfrac{7S_Y^2}{6S_X^2}\sim \mathrm{F}(12,14)$이다. 따라서

$$\mathrm{P}\left(S_Y^2\ge 3S_X^2\right)=\mathrm{P}\left(\frac{S_Y^2}{S_X^2}\ge 3\right)$$
$$\approx 0.025$$

이다.

4. 두 정규모집단 A와 B의 모분산은 동일하고, 평균은 각각 $\mu_X=700$, $\mu_Y=680$이라 한다. 이 때 두 모집단으로부터 표본을 추출하여 다음과 같은 결과를 얻었다. 단위는 mg이다.

A 표본	$n=17$	$\bar{x}=704$	$s_X=39.25$
B 표본	$m=10$	$\bar{y}=675$	$s_Y=43.75$

(1) 합동표본분산의 관측값 s_p^2을 구하라.
(2) 두 표본평균의 차 $T=\overline{X}-\overline{Y}$에 대한 확률분포를 구하라.
(3) $P(T\ge t_0)=0.05$인 t_0을 구하라.

Solution.

(1) $s_p^2=\dfrac{1}{17+10-2}\left[(17-1)\times 39.25^2+(10-1)\times 43.75^2\right]=1675.0225$
(2)

$$\frac{\overline{X}-\overline{Y}-(\mu_X-\mu_Y)}{s_p\sqrt{\frac{1}{n}+\frac{1}{m}}}=\frac{T-20}{\sqrt{1675.0225}\sqrt{\frac{1}{17}+\frac{1}{10}}}$$
$$=\frac{T-20}{16.3105}\sim \mathrm{t}(17+10-2)=\mathrm{t}(25)$$

(3) $t_{0.05}(25)=1.708$이므로 $t_0=20+1.708\times 16.3105=47.8583$이다.

5. 모직 17묶음의 절단강도는 평균 452.4이고 표준편차는 12.3이고, 인조섬유 25묶음의 절단강도는 평균 474.6이고 표준편차는 5.50인 것으로 조사되었다. 그리고 두 종류 섬유의 절단강도는 동일한 분산을 갖는 정규분포를 이룬다고 한다.

(1) 합동표본분산의 측정값 s_p^2을 구하라.
(2) 모직의 평균 절단강도 $\overline{X}$와 인조섬유의 평균 절단강도 $\overline{Y}$의 차 $T = \overline{Y} - \overline{X}$에 대한 확률분포를 구하라.
(3) 표본의 측정값 $T_0 = \overline{y} - \overline{x}$를 이용하여 두 모집단 인조섬유와 모직에 대한 절단강도의 평균의 차이 μ_0에 대하여 $P(|T - \mu_0| \le t_0) = 0.95$인 μ_0의 범위를 구하라.

Solution.

(1) $s_p^2 = \dfrac{1}{17+25-2}\left[(17-1)\times 12.3^2 + (25-1)\times 5.50^2\right] = 78.666$
(2)

$$\frac{\overline{Y}-\overline{X}-(\mu_Y-\mu_X)}{s_p\sqrt{\dfrac{1}{n}+\dfrac{1}{m}}} = \frac{T-22.2}{\sqrt{78.666}\sqrt{\dfrac{1}{17}+\dfrac{1}{25}}} = \frac{T-22.2}{2.7882} \sim \mathrm{t}(17+25-2) = \mathrm{t}(40)$$

(3) $t_{0.025}(40) = 2.021$이다. 따라서 $t_0 = 2.021 \times 2.7882 = 5.635$이다. 또한 $T_0 = \overline{y} = \overline{x} = 22.2$이다. 따라서 μ_0은 $T_0 - t_0 \le \mu_0 \le T_0 + t_0$에서 범위 [16.565, 27.835]를 갖는다.

7.4 표본비율의 분포

6. 어느 도시의 시장 선거에서 A 후보자는 그 도시의 유권자를 상대로 53%의 지지율을 얻고 있다고 한다. 이때 400명의 유권자를 상대로 조사한 결과, 49% 이하의 유권자가 지지할 확률을 구하라.

Solution. $\hat{p} \approx \mathrm{N}\left(0.53, \dfrac{0.53\,(1-0.53)}{400}\right) = \mathrm{N}\left(0.7, 0.02495^2\right)$이다. 따라서

$$\begin{aligned} P\,(\hat{p} \geq 0.49) &= P\left(\frac{\hat{p}-0.53}{0.02495} \leq \frac{0.49-0.53}{0.02495}\right) \\ &= P\,(Z \leq -1.603) \approx 0.0548 \end{aligned}$$

이다.

8. 2005년 통계조사에 따르면 25세 이상 남성과 여성 중 대졸 이상은 각각 37.8%와 25.4%로 조사되었다. 남성 500명과 여성 450명을 표본조사한 결과, 남성과 여성의 비율의 차가 11.5% 이하일 확률을 구하라.

Solution. 남성에 대해

$$\hat{p_M} \approx \mathrm{N}\left(0.378, \frac{0.378\,(1-0.378)}{500}\right) = \mathrm{N}\,(0.378, 0.0004702)$$

이고, 여성에 대해

$$\hat{p_F} \approx \mathrm{N}\left(0.254, \frac{0.254\,(1-0.254)}{450}\right) = \mathrm{N}\,(0.254, 0.0004211)$$

이다. 따라서 $t = \hat{p_M} - \hat{p_F}$라 하면

$$t \approx \mathrm{N}\,(0.378-0.254, 0.0004702+0.0004211) = \mathrm{N}\left(0.124, 0.02985^2\right)$$

이다. 이를 통해 남성과 여성의 비율의 차가 11.5% 이하일 확률을 구하면

$$\begin{aligned} P\,(t < 0.115) &= P\left(\frac{t-0.124}{0.02985} \leq \frac{0.115-0.124}{0.02985}\right) \\ &= P\,(Z \leq -0.3015) \approx 0.3815 \end{aligned}$$

이다.

8 추정

8.1 점추정과 구간추정

3. $\mathrm{E}(X_1) = \mu$, $\mathrm{Var}(X_1) = 4$, $\mathrm{E}(X_2) = \mu$, $\mathrm{Var}(X_2) = 7$, $\mathrm{E}(X_3) = \mu$, $\mathrm{Var}(X_3) = 14$일 때, 다음 추정량을 이용하여 모평균 μ를 점추정하고자 한다.

$$\hat{\mu_1} = \frac{1}{3}(X_1 + X_2 + X_3), \qquad \hat{\mu_2} = \frac{1}{4}(X_1 + 2X_2 + X_3), \qquad \hat{\mu_3} = \frac{1}{3}(2X_1 + X_2 + 2X_3)$$

(1) 각 추정량의 편의를 구하라.
(2) 불편추정량과 편의추정량을 구하라.
(3) 불편추정량들의 분산을 구하고, 최소분산불편추정량을 구하라.
(4) $\mu = 1$일 때, 각 추정량들의 평균제곱오차를 구하라.

Solution.

(1)

$$\begin{aligned}
\mathrm{Bias}_\mu(\hat{\mu_1}) &= \frac{1}{3}(\mu + \mu + \mu) = 0 \\
\mathrm{Bias}_\mu(\hat{\mu_2}) &= \frac{1}{4}(\mu + 2\mu + \mu) = 0 \\
\mathrm{Bias}_\mu(\hat{\mu_3}) &= \frac{1}{3}(2\mu + \mu + 2\mu) = \frac{2}{3}\mu
\end{aligned}$$

(2) (1)에서 불편추정량 $\hat{\mu_1}$, $\hat{\mu_2}$를 얻고, 편의추정량 $\hat{\mu_3}$을 얻는다.
(3)

$$\begin{aligned}
\mathrm{Var}(\hat{\mu_1}) &= \left(\frac{1}{3}\right)^2 \mathrm{Var}(X_1 + X_2 + X_3) = 2.778 \\
\mathrm{Var}(\hat{\mu_2}) &= \left(\frac{1}{4}\right)^2 \mathrm{Var}(X_1 + 2X_2 + X_3) = 2.875 \\
\mathrm{Var}(\hat{\mu_3}) &= \left(\frac{1}{3}\right)^2 \mathrm{Var}(2X_1 + X_2 + 2X_3) = 8.778
\end{aligned}$$

이고 최소불편추정량 $\hat{\mu_1}$를 얻는다.

(4)

$$\begin{aligned}
\mathrm{MSE}(\hat{\mu}_1) &= \mathrm{Var}(\mu_1) + \left[\mathrm{Bias}_\mu(\hat{\mu}_1)\right]^2 = 2.778 \\
\mathrm{MSE}(\hat{\mu}_2) &= \mathrm{Var}(\mu_2) + \left[\mathrm{Bias}_\mu(\hat{\mu}_2)\right]^2 = 2.875 \\
\mathrm{MSE}(\hat{\mu}_3) &= \mathrm{Var}(\mu_3) + \left[\mathrm{Bias}_\mu(\hat{\mu}_3)\right]^2 = 9.222
\end{aligned}$$

5. $\mathrm{E}(X_1) = \mu$, $\mathrm{Var}(X_1) = 2$, $\mathrm{E}(X_2) = \mu$, $\mathrm{Var}(X_2) = 4$이다.

(1) $\hat{\mu} = \frac{1}{2}(X_1 + X_2)$의 분산을 구하라.
(2) $\hat{\mu}_1 = aX_1 + (1-a)X_2$의 분산이 최소가 되는 상수 a와 최소 분산을 구하라.

Solution.

(1) $\mathrm{Var}(\hat{\mu}) = \left(\frac{1}{2}\right)^2 \mathrm{Var}(X_1 + X_2) = \frac{3}{2}$
(2)

$$\begin{aligned}
\mathrm{Var}(\hat{\mu}_1) &= a^2\mathrm{Var}(X_1) + (a-1)^2\,\mathrm{Var}(X_2) \\
&= 2a^2 + 4(a-1)^2 \\
&= \frac{2}{3}(3a-2)^2 + \frac{4}{3}
\end{aligned}$$

이므로 분산이 최소가 될 때 $a = \frac{2}{3}$이고, 그 때의 분산은 $\frac{4}{3}$이다.

6. 크기 20인 표본을 조사한 결과 $\sum_{i=1}^{20} x_i = 48.6$과 $\sum_{i=1}^{20} x_i^2 = 167.4$인 결과를 얻었다. 이 결과를 이용하여 모평균과 모분산을 점추정하라.

Solution. 모평균의 점추정값은 $\frac{48.6}{20} = 2.43$이다.

표본분산은 $\frac{1}{19}\sum_{i=1}^{20}(x_i - \overline{x})^2$인데 이는 $\frac{1}{19}\left(\sum_{i=1}^{20} x_i^2 - 20\overline{x}^2\right)$로도 계산할 수 있다. 따라서 $s^2 = 2.595$를 얻고, 이 때 표본분산은 모분산에 대한 불편추정량이므로 모분산의 점추정값도 2.595이다.

7. 타이어를 생산하는 어느 회사에서 새로운 제조법에 의하여 생산한 타이어의 주행거리를 알아보기 위하여 7개를 임의로 선정하여 사용한 결과 다음과 같은 결과를 얻었다. 단, 단위는 1,000km이다.

59.2	60.6	56.2	62.0	58.1	57.7	58.1

(1) 이 회사에서 생산된 타이어의 평균 주행거리의 추정값을 구하라.
(2) 주행거리의 분산에 대한 추정값을 구하라.

Solution. 이 표본의 평균 $\bar{x} = 58.84$, 표준편차 $s = 1.94$이다.

(1) 표본평균은 모평균에 대한 불편추정량이므로 점추정값은 58.84이다.
(2) 표본분산은 모분산에 대한 불편추정량이므로 점추정값은 $1.94^2 = 3.76$이다.

9. $X_1, X_2, \cdots, X_{10}$은 $n = 10$이고 미지의 p를 모수로 갖는 이항분포로부터 추출된 확률표본이라 한다.

1	8	2	5	7	6	2	9	4	7

(1) $\hat{p} = \dfrac{\overline{X}}{10}$가 모수 p에 대한 불편추정량임을 보여라.
(2) 관찰된 표본이 위와 같을 때, 모수 p에 대한 불편추정값을 구하라.

Solution.

(1) $X_i \sim \mathrm{B}(10, p)$이다. 따라서 $\mathrm{E}(X_i) = 10p$이고, $\mathrm{E}\left(\overline{X}\right) = 10p$이다. 한편 $\mathrm{E}(\hat{p}) = \mathrm{E}\left(\dfrac{\overline{X}}{10}\right) = p$이므로, $\mathrm{Bias}_p(\hat{p}) = 0$이다. 따라서 $\hat{p}$는 p에 대한 불편추정량이다.
(2) 표본평균 $\bar{x} = 5.1$이므로 p의 불편추정값은 0.51이다.

8.2 모평균의 구간추정

1. 어느 회사에서 생산하는 비누 무게는 분산이 $\sigma^2 = 4$인 정규분포를 따른다고 한다. 25개의 비누를 임의로 추출하였을 때, 그 평균 무게의 값은 $\bar{x} = 97$이었다. 실제 평균 무게 μ에 대한 95% 신뢰구간을 구하라. 단위는 g이다.

Solution. $z_{0.025} = 1.96$이다. 따라서 95% 신뢰구간은

$$\left(\bar{x} - z_{0.025}\frac{\sigma}{\sqrt{n}}, \bar{x} + z_{0.025}\frac{\sigma}{\sqrt{n}}\right) = \left(97 - 1.96 \times \frac{2}{\sqrt{25}}, 97 + 1.96 \times \frac{2}{\sqrt{25}}\right)$$
$$= (96.216, 97.784)\ \text{g}$$

이다.

2. 문제 **1**에서 분산 σ^2를 모르지만 $s^2 = 4.25$라 할 때, 실제 평균 무게 μ에 대한 95% 신뢰구간을 구하라.

Solution. $t_{0.025}(25-1) = 2.064$이고 $s = \sqrt{4.25} = 2.062$이다. 따라서 95% 신뢰구간은

$$\left(\bar{x} - t_{0.025}(n-1)\frac{s}{\sqrt{n}}, \bar{x} + t_{0.025}(n-1)\frac{s}{\sqrt{n}}\right) = \left(97 - 2.064 \times \frac{2.062}{\sqrt{25}}, 97 + 2.064 \times \frac{2.062}{\sqrt{25}}\right)$$
$$= (96.149, 97.851)\ \text{g}$$

이다.

4. 다음 자료는 어느 직장에 근무하는 직원 20명에 대한 혈중 콜레스테롤 수치를 조사한 자료이다. 다음 각 조건 아래서 이 직장에 근무하는 직원들의 콜레스테롤 평균 수치에 대한 95% 신뢰구간을 구하라.

193.27	193.88	253.26	237.15	188.83	200.56	274.31	230.36	212.08	222.19
198.48	202.50	215.35	218.95	233.16	222.23	218.53	204.64	206.72	199.37

(1) 콜레스테롤 수치가 정규분포 $\mathrm{N}(\mu, 400)$에 따르는 경우
(2) 콜레스테롤 수치가 정규분포에 따르는 경우

Solution. 이 표본의 평균 $\bar{x} = 216.29$, 표준편차 $s = 21.53$이다.

(1) $z_{0.025} = 1.96$이다. 따라서 95% 신뢰구간은

$$\begin{aligned}\left(\bar{x} - z_{0.025}\frac{\sigma}{\sqrt{n}}, \bar{x} + z_{0.025}\frac{\sigma}{\sqrt{n}}\right) &= \left(216.29 - 1.96 \times \frac{20}{\sqrt{20}}, 216.29 + 1.96 \times \frac{20}{\sqrt{20}}\right) \\ &= (207.52, 225.06)\ \text{mg/dL}\end{aligned}$$

이다.

(2) $t_{0.025}(20-1) = 2.093$이다. 따라서 95% 신뢰구간은

$$\begin{aligned}&\left(\bar{x} - t_{0.025}(n-1)\frac{s}{\sqrt{n}}, \bar{x} + t_{0.025}(n-1)\frac{s}{\sqrt{n}}\right) \\ =&\left(216.29 - 2.064 \times \frac{21.53}{\sqrt{20}}, 216.29 + 2.064 \times \frac{21.53}{\sqrt{20}}\right) \\ =&(206.35, 226.23)\ \text{mg/dL}\end{aligned}$$

이다.

9. 강의실 옆의 커피 자판기에서 컵 한 잔에 나오는 커피의 양을 조사하기 위하여 101잔을 조사한 결과 평균 0.3리터, 표준편차 0.06리터이었다. 다음 각 조건 아래서 이 자판기에서 나오는 커피 한 잔의 평균 양에 대한 95% 신뢰구간을 구하라.

(1) 커피의 양은 정규분포에 따르며, 표준편차는 0.05리터로 알려져 있는 경우
(2) 커피의 양은 정규분포에 따르며, 표준편차를 모르는 경우

Solution.

(1) $z_{0.025} = 1.96$이다. 따라서 95% 신뢰구간은

$$\begin{aligned}\left(\bar{x} - z_{0.025}\frac{\sigma}{\sqrt{n}}, \bar{x} + z_{0.025}\frac{\sigma}{\sqrt{n}}\right) &= \left(0.3 - 1.96 \times \frac{0.05}{\sqrt{101}}, 0.3 + 1.96 \times \frac{0.05}{\sqrt{101}}\right) \\ &= (0.2902, 0.3098)\ \text{L}\end{aligned}$$

이다.

(2) $t_{0.025}(101-1) = 1.984$이다. 따라서 95% 신뢰구간은

$$\begin{aligned}\left(\bar{x} - t_{0.025}(n-1)\frac{s}{\sqrt{n}}, \bar{x} + t_{0.025}(n-1)\frac{s}{\sqrt{n}}\right) &= \left(0.3 - 1.984 \times \frac{0.06}{\sqrt{101}}, 0.3 + 1.984 \times \frac{0.06}{\sqrt{101}}\right) \\ &= (0.2881, 0.3119)\ \text{L}\end{aligned}$$

이다.

10. 어느 컴퓨터 제조회사에서 생산되는 컴퓨터의 내구연한이 정규분포를 이룬다고 한다. 10명의 소비자를 대상으로 설문 조사한 결과 다음 자료를 얻었다. 모평균 μ에 대한 90% 신뢰구간을 구하라. 단위는 년이다.

4.6	3.6	4.0	6.1	8.8	5.3	1.2	5.6	3.3	1.6

Solution. 이 표본의 평균 $\bar{x} = 4.41$이고 표준편차 $s = 2.23$이다. 모표준편차가 알려져 있지 않으므로 t-추정이 필요하다. $t_{0.05}(10-1) = 1.833$이다. 따라서 90% 신뢰구간은

$$\begin{aligned}\left(\bar{x} - t_{0.05}(n-1)\frac{s}{\sqrt{n}}, \bar{x} + t_{0.05}(n-1)\frac{s}{\sqrt{n}}\right) &= \left(4.41 - 1.833 \times \frac{2.23}{\sqrt{10}}, 4.41 + 1.833 \times \frac{2.23}{\sqrt{10}}\right) \\ &= (3.117, 5.703) \text{ 년}\end{aligned}$$

이다.

15. 모평균 μ_1, μ_2, 그리고 $\sigma_1 = 4$, $\sigma_2 = 5$인 두 정규모집단으로부터 각각 크기 12, 10인 표본을 임의로 추출하여 $\bar{x} = 75.5$, $\bar{y} = 70.4$인 결과를 얻었다. 모평균의 차 $\mu_1 - \mu_2$에 대한 95% 신뢰구간을 구하라.

Solution. $\bar{x} - \bar{y} = 5.1$이고 $\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}} = 1.9578$이다. 또한 $z_{0.025} = 1.96$이다. 따라서 95% 신뢰구간은

$$\begin{aligned}&\left(\bar{x} - \bar{y} - z_{0.025}\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}, \bar{x} - \bar{y} + z_{0.025}\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}\right) \\ &= (5.1 - 1.96 \times 1.9578, 5.1 + 1.96 \times 1.9578) \\ &= (1.2625, 8.9375)\end{aligned}$$

이다.

19. 모평균 μ_1, μ_2인 두 정규모집단으로부터 각각 크기 10, 15 표본을 임의로 추출하여 $\bar{x} = 485.5$, $\bar{y} = 501.4$와 $s_1 = 6$, $s_2 = 7$인 결과를 얻었다. 모평균의 차 $\mu_1 - \mu_2$에 대한 95% 신뢰구간을 구하라.

Solution. 합동표본분산 $s_p^2 = \frac{1}{10+15-2}\left[(10-1) \times 6^2 + (15-1) \times 7^2\right] = 43.913$이므로 합동표본표준편차 $s_p = 6.627$이고, $t_{0.025}(23) = 2.069$이므로 95% 신뢰구간은

$$\begin{aligned}&\left(\bar{x} - \bar{y} - t_{0.025}(n+m-2)\, s_p\sqrt{\frac{1}{n} + \frac{1}{m}}, \bar{x} - \bar{y} + t_{0.025}(n+m-2)\, s_p\sqrt{\frac{1}{n} + \frac{1}{m}}\right) \\ &= (-15.9 - 2.069 \times 6.627 \times 0.408, -15.9 + 2.069 \times 6.627 \times 0.408) \\ &= (-21.496, -10.303)\end{aligned}$$

이다.

21. 다음은 남자와 여자의 생존 연령을 조사한 자료이다.

남자	52	60	55	46	33	75	58	45	57	88
여자	62	58	65	56	53	45	56	65	77	47

(1) 남자와 여자의 평균 생존 연령에 대한 90% 신뢰구간을 구하라.
(2) 두 그룹의 모분산이 동일하다는 조건 아래서 합동표준편차를 구하라.
(3) (2)를 이용하여, 여자와 남자의 평균 생존 연령의 차이에 대한 90% 신뢰구간을 구하라.

Solution. 남자의 경우 평균 $\bar{x} = 56.9$, 표준편차 $s_1 = 15.51$이고, 여자의 경우 평균 $\bar{y} = 58.4$, 표준편차 $s_2 = 9.41$이다.

(1) $t_{0.05}(10-1) = 1.833$이다. 따라서 남자의 경우 90% 신뢰구간은

$$\begin{aligned}\left(\bar{x} - t_{0.05}(n-1)\frac{s_1}{\sqrt{n}}, \bar{x} + t_{0.05}(n-1)\frac{s_1}{\sqrt{n}}\right) &= \left(56.9 - 1.833 \times \frac{15.51}{\sqrt{10}}, 56.9 + 1.833 \times \frac{15.51}{\sqrt{10}}\right) \\ &= (47.91, 65.89) \text{ 세}\end{aligned}$$

이고 여자의 경우 90% 신뢰구간은

$$\begin{aligned}\left(\bar{y} - t_{0.05}(m-1)\frac{s_1}{\sqrt{m}}, \bar{y} + t_{0.05}(m-1)\frac{s_1}{\sqrt{m}}\right) &= \left(58.4 - 1.833 \times \frac{9.41}{\sqrt{10}}, 58.4 + 1.833 \times \frac{9.41}{\sqrt{10}}\right) \\ &= (52.95, 63.85) \text{ 세}\end{aligned}$$

이다.

(2) $s_p^2 = \dfrac{1}{10+10-2}\left[(10-1) \times 15.51^2 + (10-1) \times 9.41^2\right] = 164.55$, $s_p = 12.83$

(3) $t_{0.05}(18) = 1.734$이다. 따라서 90% 신뢰구간은

$$\begin{aligned}&\left(\bar{y} - \bar{x} - t_{0.05}(n+m-2)s_p\sqrt{\frac{1}{n} + \frac{1}{m}}, \bar{y} - \bar{x} + t_{0.05}(n+m-2)s_p\sqrt{\frac{1}{n} + \frac{1}{m}}\right) \\ &= (1.5 - 1.734 \times 12.83 \times 0.4472,\ 1.5 + 1.734 \times 12.83 \times 0.4472) \\ &= (-8.45,\ 11.45)\end{aligned}$$

이다.

8.3 모비율과 모분산의 구간추정

4. 다음은 어떤 제조회사에서 생산되는 음료수를 분석한 당분 함량의 자료이다. 이것을 이용하여 이 회사에서 생산되는 음료수의 당분 함량에 대한 분산과 표준 편차에 대한 95% 신뢰구간을 구하라.

15.1	13.4	16.5	14.6	14.4	14.0	15.4	13.8	14.6	14.3

Solution. 이 표본의 표준분산 $s^2 = 0.7854$이다. $(10-1)\dfrac{S^2}{\sigma^2} \sim \chi^2(10-1)$이고 $\chi^2_{0.975}(9) = 2.70$, $\chi^2_{0.025} = (19.02)$이므로 σ^2에 대한 95% 신뢰구간은

$$\begin{aligned}\left(\frac{9s^2}{\chi^2_{0.025}(9)}, \frac{9s^2}{\chi^2_{0.975}(9)}\right) &= \left(\frac{9\times 0.7854}{19.02}, \frac{9\times 0.7854}{2.70}\right)\\ &= (0.37, 2.62)\end{aligned}$$

이고, σ에 대한 95% 신뢰구간은 $(0.61, 1.62)$이다.

9. 2000년 4월 (사)한국청소년순결운동본부가 전국의 고등학생(남학생 256명, 여학생 348명)을 대상으로 청소년의 음주정도에 대한 무작위 표본조사를 실시한 결과, 남학생 83.9%, 여학생 59.2%가 음주 경험이 있는 것으로 조사되었다. 남학생과 여학생의 음주율 차에 대한 95% 신뢰구간을 구하라.

Solution. $z_{0.025} = 1.96$이다. 남학생에 대해 $\hat{p}_1 = 0.839$, 여학생에 대해 $\hat{p}_2 = 0.592$이다. 따라서 $\hat{p}_1 - \hat{p}_2 = 0.247$이고 표준 오차 $\sigma_{\hat{p}_1-\hat{p}_2} = \sqrt{\dfrac{\hat{p}_1(1-\hat{p}_1)}{256} + \dfrac{\hat{p}_2(1-\hat{p}_2)}{348}} = 0.035$ 이므로, $p_1 - p_2$에 대한 95% 신뢰구간은

$$\begin{aligned}&(\hat{p}_1 - \hat{p}_2 - z_{0.025}\sigma_{\hat{p}_1-\hat{p}_2}, \hat{p}_1 - \hat{p}_2 + z_{0.025}\sigma_{\hat{p}_1-\hat{p}_2})\\ =&(0.247 - 1.96\times 0.035, 0.247 + 1.96\times 0.035)\\ =&(0.178, 0.316)\end{aligned}$$

이다.

11. 두 종류의 약품 A, B의 효능을 조사하기 위하여 동일한 조건에 놓인 환자 200명에게 A를 투여하고, 다른 200명의 환자에게 B를 투여한 결과 각각 165명과 150명이 효과를 얻었다. 두 약품의 효율의 차이에 대한 95% 신뢰구간을 구하라.

Solution. $z_{0.025} = 1.96$이다. A에 대해 $\hat{p_A} = 0.825$, B에 대해 $\hat{p_B} = 0.750$이다. 따라서 $\hat{p_A} - \hat{p_B} = 0.075$ 이고 표준 오차 $\sigma_{\hat{p_A}-\hat{p_B}} = \sqrt{\dfrac{\hat{p_A}(1-\hat{p_A})}{200} + \dfrac{\hat{p_B}(1-\hat{p_B})}{200}} = 0.0407$이므로, $p_A - p_B$에 대한 95% 신뢰구

간은

$$
\begin{aligned}
&\left(\hat{p_{\mathrm{A}}} - \hat{p_{\mathrm{B}}} - z_{0.025}\sigma_{\hat{p_{\mathrm{A}}}-\hat{p_{\mathrm{B}}}},\, \hat{p_{\mathrm{A}}} - \hat{p_{\mathrm{B}}} + z_{0.025}\sigma_{\hat{p_{\mathrm{A}}}-\hat{p_{\mathrm{B}}}}\right)\\
=&\,(0.075 - 1.96 \times 0.0407,\, 0.075 + 1.96 \times 0.0407)\\
=&\,(-0.0048,\, 0.155)
\end{aligned}
$$

이다.

9 가설검정

9.2 모평균의 검정

4. 승용차에 사용되는 유리를 생산하는 회사에서 만든 유리의 두께는 정규분포에 따른다고 한다. 한편 이 회사에서 생산하는 유리의 두께는 평균 5mm라고 주장한다. 이러한 주장의 진위를 알아보기 위하여 41개의 유리를 표본조사한 결과 평균 4.96mm, 표준편차 0.124mm를 얻었다. 이 회사에서 주장하는 유리 두께의 평균과 표준편차에 대하여 유의수준 0.05와 0.01에서 양측검정하라.

Solution.

$$H_0 : \mu = 5 \qquad H_1 : \mu \neq 5$$

로 설정한다. 검정통계량 $T = \dfrac{\overline{X} - 5}{\frac{s}{\sqrt{41}}} \sim \mathrm{t}(41-1)$이고 관측값 $t_0 = \dfrac{4.96 - 5}{\frac{0.124}{\sqrt{41}}} = -2.066$이다.

- 유의수준 0.05에서 양측검정하면, $t_{0.025}(40) = 2.021$이므로 기각역 $R : |T| \geq 2.021$인데 $t_0 = -2.066$은 기각역에 들어가므로 귀무가설을 기각한다.
- 유의수준 0.01에서 양측검정하면, $t_{0.005}(40) = 2.704$이므로 기각역 $R : |T| \geq 2.704$인데 $t_0 = -2.066$은 기각역에 들어가지 않으므로 귀무가설을 기각할 수 없다.

8. 컴퓨터 회사에서 새로 개발한 소프트웨어는 초보자도 쉽게 사용할 수 있도록 만들었으며, 그 사용법을 능숙하게 익히는 데 3시간 이상 걸리지 않는다고 한다. 이를 확인하기 위하려 20명을 표본으로 선정하여 조사한 결과 다음과 같다. 유의수준 5%에서 이 회사의 주장에 대한 타당성을 조사하라. 단, 소프트웨어를 사용하는 데 걸리는 시간은 정규분포를 이룬다고 한다.

2.75	3.25	3.48	2.95	2.82	3.75	4.01	3.05	2.67	4.25
3.01	2.84	2.75	1.80	3.20	2.48	2.95	3.02	2.73	2.56

Solution. 표본의 평균 $\bar{x} = 3.016$, 표준편차 $s = 0.5501$이다.

$$H_0 : \mu \leq 3 \qquad H_1 : \mu > 3$$

으로 설정한다. 검정통계량 $T = \dfrac{\overline{X} - 3}{\frac{s}{\sqrt{20}}} \sim \mathrm{t}(20-1)$이고 관측값 $t_0 = \dfrac{3.016 - 3}{\frac{0.5501}{\sqrt{20}}} = 0.1301$이다.

유의수준 0.05에서 단측검정하면, $t_{0.05}(19) = 1.729$이므로 기각역 $R : T \geq 1.729$인데 $t_0 = 0.1301$은 기각역에 들어가지 않으므로 귀무가설을 기각할 수 없다. 따라서 회사의 주장은 타당하다고 생각할 수 있다.

10. 어느 진공관 제조회사의 주장에 따르면, 이 회사에서 생산되는 진공관의 수명의 수명이 2,550시간 이상 된다고 한다. 이를 확인하기 위하여 36개의 진공관을 임의로 추출하여 조사한 결과 평균 수명이 2,516시간이고 표준편차가 132시간이었다. 이 회사의 주장이 타당한지 유의수준 5%에서 검정하라. 단, 진공관의 수명은 정규분포에 따른다고 알려져 있다.

Solution.

$$H_0 : \mu \geq 2550 \qquad H_1 : \mu < 2550$$

으로 설정한다. 모분산이 알려져 있지 않은 대표본이므로 중심극한정리에 의해 검정통계량의 분포는 정규분포로 근사되며 이 때 $Z = \dfrac{\overline{X} - 2550}{\dfrac{s}{\sqrt{36}}} \approx \mathrm{N}(0, 1)$이고 관측값 $z_0 = \dfrac{2516 - 2550}{\dfrac{132}{\sqrt{36}}} = -1.5455$이다.

유의수준 0.05에서 단측검정하면, 기각역 $R : Z \leq -1.6$인데 $z_0 = -1.5455$는 기각역에 들어가지 않으므로 귀무가설을 기각할 수 없다. 따라서 회사의 주장은 타당하다고 생각할 수 있다.

12. 두 회사 A와 B에서 생산되는 타이어의 평균 수명에 차이가 있는지 조사하기 위하여, 각각 36개씩 타이어를 표본추출하여 조사한 결과 다음과 같았다. 두 회사에서 생산한 타이어의 평균 수명에 차이가 있는지 유의수준 5%에서 조사하라. 단위는 주행 km이다.

	표본평균	표본표준편차
A 회사 타이어	57,300	3,550
B 회사 타이어	56,100	3,800

Solution.

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

로 설정한다. 모분산이 알려져 있지 않은 대표본이므로 중심극한정리에 의해 검정통계량의 분포는 정규분포로 근사되며 이 때 $Z = \dfrac{\overline{X} - \overline{Y}}{\sqrt{\dfrac{s_1^2}{n} + \dfrac{s_2^2}{m}}} = \dfrac{\overline{X} - \overline{Y}}{866.7} \approx \mathrm{N}(0, 1)$이고 관측값 $z_0 = \dfrac{57300 - 56100}{866.7} = 1.38$ 이다.

유의수준 0.05에서 양측검정하면, 기각역 $R : |Z| \geq 1.96$인데 $z_0 = 1.38$는 기각역에 들어가지 않으므로 귀무가설을 기각할 수 없다. 따라서 두 회사에서 생산한 타이어의 평균 수명에는 차이가 없다고 생각할 수 있다.

13. 어떤 대학 병원에서 단기간 동안 이 병원에 입원한 남녀 환자들의 입원 기간을 조사하여 다음을 얻었다. 이 때 유의수준 5%에서 남자와 여자의 입원 기간에 차이가 있는지 검정하라. 단, 남자와 여자의 입원 기간은 각각 표준편차가 5.6일, 4.5일인 정규분포에 따른다고 한다.

남자	3	4	12	16	5	11	21	9	8	25
	17	3	8	6	13	7	30	12	9	10
여자	12	5	4	10	1	8	19	13	9	1
	13	13	7	9	15	8	28			

Solution. 남자의 경우 $\bar{x} = 11.45$, $s_1 = 7.258$이고 여자의 경우 $\bar{y} = 10.29$, $s_2 = 6.622$이다.

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

로 설정한다. 검정통계량 $Z = \dfrac{\overline{X} - \overline{Y}}{\sqrt{\dfrac{\sigma_1^2}{n} + \dfrac{\sigma_2^2}{m}}} = \dfrac{\overline{X} - \overline{Y}}{1.6443} \sim \mathrm{N}(0, 1)$이고 관측값 $z_0 = \dfrac{11.45 - 10.29}{1.6443} = 0.705$ 이다.

유의수준 0.05에서 양측검정하면, 기각역 $R : |Z| \geq 1.96$인데 $z_0 = 0.705$는 기각역에 들어가지 않으므로 귀무가설을 기각할 수 없다. 따라서 남자와 여자의 입원 기간에는 차이가 없다고 생각할 수 있다.

14. 서로 독립인 두 정규모집단에서 각각 크기 11과 16인 표본을 추출하여 다음 결과를 얻었다.

모집단	표본의 크기	표본평균	표본표준편차
1	11	704	1.6
2	16	691	1.2

(1) 합동표본분산 S_p^2을 구하라.
(2) 두 모분산이 같은 경우에 유의수준 0.1에서 $H_0 : \mu_1 = \mu_2$와 $H_1 : \mu_1 \neq \mu_2$를 검정하라.

Solution.

(1) $S_p^2 = \dfrac{1}{11 + 16 - 2}\left[(11-1) \times 1.6^2 + (16-1) \times 1.2^2\right] = 1.888$

(2) 검정통계량 $T = \dfrac{\overline{X} - \overline{Y}}{S_p\sqrt{\dfrac{1}{n} + \dfrac{1}{m}}} = \dfrac{\overline{X} - \overline{Y}}{0.538} \sim \mathrm{t}(11 + 16 - 2)$이고 관측값 $t_0 = \dfrac{704 - 691}{0.538} = 24.164$이다.
유의수준 0.1에서 양측검정하면, 기각역 $R : |T| \geq 1.316$인데 $t_0 = 24.164$는 기각역에 들어가므로 귀무가설을 기각한다.

15. A 고등학교 학생들의 주장에 따르면 자신들의 평균 성적이 B 고등학교 학생들보다 높다고 한다. 이를 확인하기 위하여 두 고등학교에서 각각 10명씩 임의로 추출하여 모의고사를 치른 결과 다음과 같은 점수를 얻었다. A 고등학교 학생들의 주장이 맞는지 유의수준 0.05에서 검정하라. 단, 두 학교 학생들의 표준편차는 거의 비슷하다고 한다.

A 고교	77	78	75	94	65	82	69	78	77	84
B 고교	73	88	75	89	54	72	69	66	87	77

Solution. A 고교의 경우 $\bar{x} = 77.9$, $s_1 = 7.951$이고 B 고교의 경우 $\bar{y} = 75.0$, $s_2 = 10.975$이다. 또한 $S_p^2 = \dfrac{1}{10+10-2}\left[(10-1)\times 7.951^2 + (10-1)\times 10.975^2\right] = 91.835$이다.

$$H_0 : \mu_1 \geq \mu_2 \qquad H_1 : \mu_1 < \mu_2$$

로 설정한다. 검정통계량 $T = \dfrac{\overline{X}-\overline{Y}}{S_p\sqrt{\dfrac{1}{n}+\dfrac{1}{m}}} = \dfrac{\overline{X}-\overline{Y}}{4.286} \sim \mathrm{t}(10+10-2)$이고 관측값 $t_0 = \dfrac{2.9}{4.286} = 0.677$ 이다.

유의수준 0.05에서 단측검정하면, 기각역 $R : T \leq 1.734$인데 $t_0 = 0.677$는 기각역에 들어가지 않으므로 귀무가설을 기각할 수 없다. 따라서 A 고교의 평균 성적은 B 고교의 평균 성적보다 높다고 생각할 수 있다.

16. 자동차 사고가 빈번히 일어나는 교차로의 신호체계를 바꾸면 사고를 줄일 수 있다고 경찰청에서 말한다. 이것을 알아보기 위하여 시범적으로 사고가 많이 발생하는 지역을 선정하여 지난 한 달 동안 발생한 사고 건수와 신호체계를 바꾼 후의 사고 건수를 조사한 결과 다음과 같았다. 유의 수준 5%에서 신호체계를 바꾸면 사고를 줄일 수 있는지 조사하라. 단, 사고 건수는 정규분포를 이룬다고 알려져 있다.

지역	1	2	3	4	5	6	7	8
바꾸기 전	5	10	8	9	5	7	6	8
바꾼 후	4	9	8	8	4	8	5	8

Solution. 바꾼 전후의 사고 건수의 차이를 계산하면 다음과 같다.

지역	1	2	3	4	5	6	7	8
바꾼 후 − 바꾸기 전	1	1	0	1	1	−1	1	0

$$H_0 : \mu_1 \le \mu_2 \qquad H_1 : \mu_1 > \mu_2$$

로 설정한다. 바꾼 전후의 사고 건수의 차이를 표본이라고 생각하면 평균 $\bar{x} = 0.50$, 표준편차 $s = 0.756$ 이다. 검정통계량 $T = \dfrac{\bar{d}}{\frac{s}{\sqrt{8}}} \sim \mathrm{t}(8-1)$이고 관측값 $t_0 = \dfrac{0.50}{\frac{0.756}{\sqrt{8}}} = 1.871$이다.

유의수준 0.05에서 단측검정하면, $t_{0.05}(7) = 1.895$이므로 기각역 $R : T \ge 1.895$인데 $t_0 = 1.871$은 기각역에 들어가지 않으므로 귀무가설을 기각할 수 없다. 따라서 신호체계를 바꾼다고 해서 사고가 줄어든다고 생각할 수는 없다.

9.3 모비율의 검정

2. 한국금연운동협의회에 따르면 전국의 20세 이상 성인 남자의 흡연율은 55.1%로 2001년 69.9%에 비해 14.8%p 감소했다고 주장하였다. 이러한 주장에 대한 진위 여부를 확인하기 위하여 850명의 20세 이상 성인 남자를 조사한 결과 503명이 흡연을 하는 것으로 조사되었다면, 흡연율이 55.1%라는 주장에 타당성이 있는지 p-값에 의하여 유의수준 1%에서 검정하라.

Solution.

$$H_0 : p = 0.551 \qquad H_1 : p \neq 0.551$$

으로 설정한다. Z-통계량 $Z = \dfrac{n\hat{p} - np_0}{\sqrt{np_0(1-p_0)}}$을 설정하면 관측값은 $Z = \dfrac{503 - 850 \times 0.551}{\sqrt{850 \times 0.551(1-0.551)}} = 2.39$이며 이 때의 p-값은 0.0168이다. 이는 유의수준 0.01보다 크므로 귀무가설을 기각하기에는 불충분하고, 따라서 흡연율이 55.1%라고 생각할 수 있다.

5. 두 표본분포 $X \sim \mathrm{B}(20, p_1)$과 $Y \sim \mathrm{B}(30, p_2)$에 대하여 성공이 각각 $x = 7$, $y = 12$로 관찰되었다.

(1) 두 표본비율 $\hat{p}_1$과 $\hat{p}_2$를 구하라.
(2) $\hat{p}_1 - \hat{p}_2$에 대한 99% 신뢰구간을 구하라.
(3) 가설 $H_0 : p_1 = p_2, H_1 : p_1 < p_2$를 유의수준 1%에서 검정하라.

Solution.

(1) $\hat{p}_1 = 0.35$, $\hat{p}_2 = 0.40$

(2) $z_{0.005} = 2.58$이다. 또한 $\sqrt{\dfrac{\hat{p}_1(1-\hat{p}_1)}{n} + \dfrac{\hat{p}_2(1-\hat{p}_2)}{m}} = 0.139$이다. 따라서 99% 신뢰구간은

$$\left(\hat{p}_1 - \hat{p}_2 - z_{0.005}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n} + \frac{\hat{p}_2(1-\hat{p}_2)}{m}},\ \hat{p}_1 - \hat{p}_2 + z_{0.005}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n} + \frac{\hat{p}_2(1-\hat{p}_2)}{m}}\right)$$
$$= (-0.05 - 2.58 \times 0.139,\ -0.05 + 2.58 \times 0.139) = (-0.41,\ 0.31)$$

이다.

(3) 합동표본비율 $\hat{p} = 0.38$이고 $z_0 = \dfrac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\dfrac{1}{n} + \dfrac{1}{m}\right)}} = -0.357$이므로 p-값은 0.7196이다. 이는 유의수준 0.1보다 크므로 귀무가설을 기각하기에는 불충분하다.

9. 어떤 단체에서 국영 TV의 광고방송에 대한 찬반을 묻는 조사를 실시하였다. 대도시에 거주하는 사람들 2,050명 중 1,250명이 찬성하였고, 농어촌에 거주하는 사람 800명 중 486명이 찬성하였다. 도시와 농어촌 사람들의 찬성 비율이 같은지 유의수준 5%에서 검정하라.

Solution. 대도시에서의 찬성률을 p_1, 농어촌에서의 찬성률을 p_2라 하자.

$$H_0 : p_1 = p_2 \qquad H_1 : p_1 \neq p_2$$

로 설정한다. 합동표본비율 $\hat{p} = 0.609$이고 $\hat{p}_1 = 0.610$, $\hat{p}_2 = 0.608$이다. 또한 $\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n} + \frac{\hat{p}_2(1-\hat{p}_2)}{m}} = 0.026$이다. 따라서 $z_0 = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n} + \frac{\hat{p}_2(1-\hat{p}_2)}{m}}} = 0.125$ 이다. 유의수준 0.05에서 $z_{0.025} = 1.96$이므로 기각역 $R : |Z| \geq 1.96$인데 z_0은 기각역에 들어가지 않으므로 귀무가설을 기각할 수 없다. 따라서 도시와 농어촌 사람들의 찬성 비율은 같다고 생각할 수 있다.

9.4 χ^2-검정과 모분산의 검정

2. 전자상가에 있는 10곳의 캠코더 판매점을 둘러본 결과, 동일한 제품의 캠코더 가격이 다음과 같이 다르게 나타났다. 판매상에 의하면 전체 캠코더 판매액이 정규분포에 따른다고 한다. 단위는 천원이다.

938.8	952.0	946.8	958.8	948.4	950.0	953.8	928.8	947.5	936.2

(1) 표본평균과 표본분산을 구하라.
(2) 모평균이 950인지 유의수준 5%에서 검정하라.
(3) 모표준편차가 11.2인지 유의수준 5%에서 검정하라.

Solution.

(1) $\bar{x} = 946.11$, $s^2 = 80.99$
(2)

$$H_0 : \mu = 950 \qquad H_1 : \mu \neq 950$$

으로 설정한다. 검정통계량 $T = \dfrac{\overline{X} - 950}{\frac{s}{\sqrt{10}}} \sim \mathrm{t}\,(10-1)$이고, 관측값 $t_0 = \dfrac{946.11 - 950}{\frac{9.00}{\sqrt{10}}} = -1.37$이다. 유의수준 0.05에서 양측검정하면, $t_{0.025}\,(9) = 2.262$이므로 기각역 $R : |T| \geq 2.262$인데 $t_0 = -1.37$은 기각역에 들어가지 않으므로 귀무가설을 기각할 수 없다. 따라서 모평균은 950이라고 생각할 수 있다.

(3)

$$H_0 : \sigma^2 = 125.44 \qquad H_1 : \sigma^2 \neq 125.44$$

로 설정한다. χ^2-통계량 $V = \dfrac{(10-1)\,S^2}{\hat{\sigma}^2} = \dfrac{S^2}{13.94}$를 선택할 때 관측값 $v_0 = \dfrac{80.99}{13.94} = 5.81$이다. 유의수준 0.05에서 양측검정하면, $\chi^2_{0.975}\,(9) = 2.7$, $\chi^2_{0.025}\,(9) = 19.02$에서 기각역 $R : (V \leq 2.7) \cup (V \geq 19.02)$인데 $v_0 = 5.81$은 기각역에 들어가지 않으므로 귀무가설을 기각할 수 없다.

3. 두 종류의 비료 A와 B를 각각 5개 지역에 사용하여 담위 면적당 쌀 수확량을 조사한 결과 다음을 얻었다. 이 때 쌀 수확량은 정규분포에 따른다고 한다.

A 지역	357	325	346	345	330
B 지역	335	328	335	344	326

(1) 두 종류의 비료에 의한 평균 쌀 수확량에 차이가 있는지 유의수준 5%에서 조사하라.
(2) 쌀 수확량의 분산에 차이가 있는지 유의수준 5%에서 조사하라.

Solution. A 지역의 경우 평균 $\bar{x} = 340.6$, 표준편차 $s_1 = 12.97$를, B 지역의 경우 평균 $\bar{y} = 333.6$, 표준편차 $s_2 = 7.09$를 각각 보였다. $S_p = \dfrac{1}{5+5-2}\left[(5-1)\times 12.97^2 + (5-1)\times 7.09^2\right] = 109.24$였다.

(1) 평균 수확량을 각각 μ_1, μ_2라 하자.

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

로 설정한다. 검정통계량 $T = \dfrac{\overline{X}-\overline{Y}}{S_p\sqrt{\dfrac{1}{n}+\dfrac{1}{m}}} = \dfrac{\overline{X}-\overline{Y}}{6.61} \sim \mathrm{t}(10-2)$이고 관측값 $t_0 = \dfrac{340.6-333.6}{6.61} = 1.06$이다.

유의수준 0.05에서 양측검정하면, $t_{0.025}(8) = 2.306$이므로 기각역 $R : |T| \geq 2.306$인데 $t_0 = 1.06$은 기각역에 들어가지 않으므로 귀무가설을 기각할 수 없다. 따라서 두 종류의 비료에 의한 평균 쌀 수확량에는 차이가 없다고 생각할 수 있다.

(2) 분산을 각각 σ_1, σ_2라 하자.

$$H_0 : \sigma_1 = \sigma_2 \qquad H_1 : \sigma_1 \neq \sigma_2$$

로 설정한다. 검정통계량 $F = \dfrac{S_1^2}{S_2^2}$이고 관측값 $f_0 = \dfrac{12.97^2}{7.09^2} = 3.35$이다.

유의수준 0.05에서 양측검정하면, $f_{0.025}(4,4) = 9.60$, $f_{0.975}(4,4) = \dfrac{1}{f_{0.025}(4,4)} = 0.104$에서 기각역 $R : (F \leq 0.104) \cup (F \geq 9.60)$인데 $f_0 = 3.35$는 기각역에 들어가지 않으므로 귀무가설을 기각할 수 없다. 따라서 두 종류의 비료에 의한 쌀 수확량의 분산에도 차이가 없다고 생각할 수 있다.

6. 어느 컴퓨터 공정라인에서 종사하는 남자와 여자의 작업 능률이 동일한지 알아보기 위하여 남·여 근로자를 각각 12명, 10명씩 임의로 추출하여 조사한 결과 남자 근로자의 표준편차는 2.3대이고, 여자 근로자의 표준편차는 2.0대였다. 남자와 여자가 생산한 컴퓨터의 대수의 모분산에 차이가 있는지 유의수준 10%에서 검정하라.

Solution. 분산을 각각 σ_1, σ_2라 하자.

$$H_0 : \sigma_1 = \sigma_2 \qquad H_1 : \sigma_1 \neq \sigma_2$$

로 설정한다. 검정통계량 $F = \dfrac{S_1^2}{S_2^2}$이고 관측값 $f_0 = \dfrac{2.3^2}{2.0^2} = 1.32$이다.

유의수준 0.1에서 양측검정하면, $f_{0.05}(11,9) = 3.1$, $f_{0.95}(11,9) = \dfrac{1}{f_{0.05}(11,9)} = 0.35$에서 기각역 $R : (F \leq 0.35) \cup (F \geq 3.1)$인데 $f_0 = 1.32$는 기각역에 들어가지 않으므로 귀무가설을 기각할 수 없다. 따라서 남자와 여자가 생산한 컴퓨터의 대수의 분산에는 차이가 없다고 생각할 수 있다.

12. 독립인 두 정규모집단으로부터 각각 크기 16과 21인 표본을 추출한 결과, 표본 A에서 표준편차 $s_1 = 5.96$, 표본 B에서 표준편차 $s_2 = 11.40$을 얻었다. 이 자료를 근거로 귀무가설 $H_0 : \sigma_1^2 = \sigma_2^2$를 유의수준 5%에서 검정하라.

Solution.

$$H_0 : \sigma_1 = \sigma_2 \qquad H_1 : \sigma_1 \neq \sigma_2$$

로 설정한다. 검정통계량 $F = \dfrac{S_1^2}{S_2^2}$이고 관측값 $f_0 = \dfrac{5.96}{11.40} = 0.523$이다.

유의수준 0.05에서 양측검정하면, $f_{0.025}(15, 20) = 2.57$, $f_{0.975}(15, 20) = \dfrac{1}{f_{0.025}(15, 20)} = 0.362$에서 기각역 $R : (F \leq 0.362) \cup (F \geq 2.57)$인데 $f_0 = 0.523$은 기각역에 들어가지 않으므로 귀무가설을 기각할 수 없다.