07 통계적 추정

확률과 통계 교과서 Review

문제 1

어느 회사에서 생산하는 음료수 한 병에 담긴 비타민 C 성분의 함유량은 정규분포를 따른다고 한다. 이 음료수 400병을 임의추출하여 비타민 C 성분의 함유량을 검사하였더니 평균이 17 mg, 표준편차가 2 mg이었다. 이 음료수 한 병에 담긴 비타민 C 성분의 평균 함유량 m에 대하여 다음을 구하여라.

(1) 신뢰도 95 %의 신뢰구간

(2) 신뢰도 99 %의 신뢰구간

문제 2

어느 지역의 가구당 한 달 의료비 X의 분포는 평균이 100000원, 표준편차가 60000원인 정규분포를 따른다고 한다. 이 중에서 다음과 같은 크기 n의 표본을 임의추출할 때, 표본평균 $\overline{X}$가 88000원 이상 112000원 이하일 확률을 각각 구하고, 이들을 비교하여 보자.

(1) $n=25$

(2) $n=64$

(3) $n=100$

문제 3

어느 공장에서 생산하는 테이프의 길이는 표준편차가 6 cm인 정규분포를 따른다고 한다. 이 테이프의 길이의 평균 m에 대한 신뢰도 95 % 신뢰구간의 길이를 2 cm이하로 하려면 표본의 크기를 얼마 이상으로 해야 하는지 구하여 보자.
(단, $\mathrm{P}(|Z| \leq 2)=0.95$로 계산한다.)

문제 4

어느 도시 주민의 60 %가 출퇴근 시 버스를 이용한다고 한다. 이 도시 주민 96명을 임의추출할 때, 출퇴근 시에 버스를 이용하는 사람이 48명 이상 60명 이하일 확률을 구하여라.

07 통계적 추정

확률과 통계 교과서 Review

문제 5

어느 지역의 실업률을 조사하기 위하여 이 지역의 주민 중에서 1600명을 임의추출하여 취업 여부를 조사하였더니 96명이 실업자였다. 이 지역의 실업률 p에 대한 신뢰도 95%의 신뢰구간을 구하여라.
(단, 반올림하여 소수 셋째 자리까지 구한다.)

문제 6

모비율 p에 대한 신뢰구간이 $\alpha \le p \le \beta$일 때, $\beta - \alpha$를 신뢰구간의 길이라고 한다. 모비율을 추정하여 자료를 분석할 때, 신뢰도는 높을수록, 신뢰구간의 길이는 짧을수록 자료 분석에 유용하다. 다음 물음에 답하여 보자.

(1) 표본의 크기가 일정할 때, 신뢰도가 높아질수록 신뢰구간의 길이는 어떻게 변하는지 말하여라.

(2) 신뢰도가 일정할 때, 신뢰구간의 길이를 짧게 하려면 어떻게 해야 하는지 말하여라.

문제 7

정규분포 $\mathrm{N}(12,\ 12^2)$을 따르는 모집단에서 임의추출한 크기가 n인 표본의 표본평균을 $\overline{X}$라고 하자. 이때 $\mathrm{P}(12 \le \overline{X} \le 15) = 0.4332$를 만족시키는 n의 값을 구하여라.

문제 8

어떤 정책에 대한 찬성률의 신뢰구간을 신뢰도 95 %로 추정하려고 한다. 신뢰구간의 길이가 0.08 이하인 신뢰구간을 얻기 위한 표본의 크기의 범위를 구하여라. (단, $\mathrm{P}(|Z| \le 2) = 0.95$로 계산한다.)

07 통계적 추정

확률과 통계 교과서 Review

문제 9 정규분포를 따르는 모집단에서 임의추출한 표본으로 모평균을 추정할 때, 옳은 것만을 보기에서 있는 대로 고른 것은?

보 기
ㄱ. 표본의 크기가 일정할 때, 신뢰도가 높아질수록 신뢰구간의 길이는 길어진다.
ㄴ. 신뢰도가 일정할 때, 표본의 크기가 커질수록 신뢰구간의 길이는 짧아진다.
ㄷ. 신뢰도를 높일 때, 신뢰구간의 길이를 일정하게 하려면 표본의 크기를 작게 하면 된다.

① ㄱ ② ㄱ, ㄴ ③ ㄱ, ㄷ ④ ㄴ, ㄷ ⑤ ㄱ, ㄴ, ㄷ

문제 10 정규분포 $\mathrm{N}(m,\ \sigma^2)$을 따르는 모집단에서 크기가 n인 표본을 임의추출하여 그 표본평균을 $\overline{X}$라고 하자. 모평균 m에 대한 신뢰도 $95\ \%$의 신뢰구간의 길이가 $\frac{1}{2}\sigma$ 이하일 때, 표본의 크기 n의 최솟값을 구하여라.

문제 11 어떤 제과점에서 만드는 과자 한 개의 무게는 평균이 $20\ \mathrm{g}$, 표준편차가 $2\ \mathrm{g}$인 정규분포를 따른다고 한다. 이 제과점에서는 과자 16개씩을 한 상자에 담아서 판매하는데, 16개의 과자를 담은 상자의 무게가 $306.88\ \mathrm{g}$ 이하이거나 $333.12\ \mathrm{g}$ 이상이면 반품된다고 한다. 어느 날 이 제과점에서 출하한 과자 상자 100개 중에서 반품된 상자가 7개 이하일 확률을 오른쪽 표준정규분포표를 이용하여 구하면? (단, 상자의 무게는 생각하지 않는다.)

z	$\mathrm{P}(0 \le Z \le z)$
1.0	0.34
1.64	0.45

① 0.16 ② 0.32 ③ 0.34 ④ 0.45 ⑤ 0.79

문제 12 정규분포 $\mathrm{N}(m,\ \sigma^2)$을 따르는 모집단에서 크기가 n인 표본을 임의추출하여 그 표본평균을 $\overline{X}$라고 하자. 모평균 m에 대한 신뢰도 $95\ \%$의 신뢰구간의 길이가 7.84일 때, $\mathrm{P}(\overline{X} \ge m+3.92)$는?

① 0.5 ② 0.475 ③ 0.35 ④ 0.1 ⑤ 0.025

07 통계적 추정

확률과 통계 교과서 Review

문제 13

어떤 도시 주민 중에서 n명을 임의추출하여 A 프로그램의 시청률을 조사할 때, 신뢰도 95 %로 추정하면 신뢰구간의 길이가 l이고, 신뢰도 α %로 추정하면 신뢰구간의 길이가 $\frac{1}{1.96}l$이다. 이때 오른쪽 표준정규분포표를 이용하여 α의 값을 구하면?

① 19.15 ② 47.5 ③ 68.26 ④ 86.64 ⑤ 95

z	$\mathrm{P}(0 \le Z \le z)$
0.5	0.1915
1.0	0.3413
1.5	0.4332
1.96	0.4750

문제 14

정규분포 $\mathrm{N}(m,\ 5^2)$을 따르는 모집단에서 크기가 25인 표본을 임의추출하여 신뢰도 x%로 추정한 신뢰구간이 $\alpha \le m \le \beta$일 때, $f(x)=\beta-\alpha$라고 하자. 상수 $x_1,\ x_2$에 대하여 $f(x_1)=2,\ f(x_2)=5$일 때, 위의 표준정규분포표를 이용하여 x_2-x_1의 값을 구하여라.

z	$\mathrm{P}(0 \le Z \le z)$
1.0	0.3413
1.5	0.4332
2.0	0.4772
2.5	0.4938

문제 15

정규분포를 따르는 모집단에서 크기가 n인 표본을 임의추출하여 모평균을 추정하려고 한다. 신뢰도 95 %로 추정할 때의 신뢰구간의 길이가 l, 신뢰도 a %로 추정할 때의 신뢰구간의 길이가 $\frac{3}{4}l$일 때, a의 값을 구하여라.

문제 16

어느 도시에서 새로운 안건에 대한 여론 조사를 하였더니 임의추출한 조사 대상자 중에서 64 %가 찬성하였다. 모비율을 신뢰도 95 %로 추정하였더니 모비율과 표본비율의 차가 0.02 이하가 되었다. 조사 대상자는 몇 명 이상인지 구하여라.

07 통계적 추정

확률과 통계 교과서 Review

〈정답 및 해설〉 확률과 통계 - 7단원. 통계적 추정

1. [정답] (1) $16.804 \leq m \leq 17.196$ (2) $16.742 \leq m \leq 17.258$

2. [정답] (1) 0.6826 (2) 0.8904 (3) 0.9544
[풀이]
(1) 표본평균 $\overline{X}$의 평균과 표준편차는

$$\mathrm{E}(\overline{X}) = 100000,\ \sigma(\overline{X}) = \frac{60000}{\sqrt{25}} = 12000$$

이므로 $\overline{X}$는 정규분포 $\mathrm{N}(100000,\ 12000^2)$을 따른다. 따라서 구하는 확률은

$$\begin{aligned}&\mathrm{P}(88000 \leq \overline{X} \leq 112000)\\&= \mathrm{P}(-1 \leq Z \leq 1) = 2\mathrm{P}(0 \leq Z \leq 1)\\&= 2 \times 0.3413 = 0.6826\end{aligned}$$

(2) 표본평균 $\overline{X}$의 평균과 표준편차는

$$\mathrm{E}(\overline{X}) = 100000,\ \sigma(\overline{X}) = \frac{60000}{\sqrt{64}} = 7500$$

이므로 $\overline{X}$는 정규분포 $\mathrm{N}(100000,\ 7500^2)$을 따른다.
따라서 구하는 확률은

$$\begin{aligned}&\mathrm{P}(88000 \leq \overline{X} \leq 112000)\\&= \mathrm{P}(-1.6 \leq Z \leq 1.6) = 2\mathrm{P}(0 \leq Z \leq 1.6)\\&= 2 \times 0.4452 = 0.8904\end{aligned}$$

(3) 표본평균 $\overline{X}$의 평균과 표준편차는

$$\mathrm{E}(\overline{X}) = 100000,\ \sigma(\overline{X}) = \frac{60000}{\sqrt{100}} = 6000$$

이므로 $\overline{X}$는 정규분포 $\mathrm{N}(100000,\ 6000^2)$을 따른다.
따라서 구하는 확률은

$$\begin{aligned}&\mathrm{P}(88000 \leq \overline{X} \leq 112000)\\&= \mathrm{P}(-2 \leq Z \leq 2) = 2\mathrm{P}(0 \leq Z \leq 2)\\&= 2 \times 0.4772 = 0.9544\end{aligned}$$

(1), (2), (3)에서 표본의 크기 n이 커질수록 표본평균 $\overline{X}$가 88000원 이상 112000원 이하일 확률은 커진다. 그 이유는 표본의 크기 n이 커질수록 표본평균의 표준편차인 $\dfrac{\sigma}{\sqrt{n}}$의 값이 작아지기 때문이다.

3. [정답] 144 이상
[풀이]
신뢰도 95 %의 신뢰구간은

$$\overline{x} - 2 \times \frac{6}{\sqrt{n}} \leq m \leq \overline{x} + 2 \times \frac{6}{\sqrt{n}}$$

이므로 신뢰구간의 길이는 $2 \times 2 \times \dfrac{6}{\sqrt{n}}$이다.
신뢰구간의 길이가 2 cm 이하이어야 하므로

$$2 \times 2 \times \frac{6}{\sqrt{n}} \leq 2,\ \sqrt{n} \geq 12$$

$$n \geq 144$$

따라서 표본의 크기를 144 이상으로 해야 한다.

07 통계적 추정

4. [정답] 0.6687

[풀이]

$$\begin{aligned}
&\mathrm{P}\left(\frac{48}{96} \le \hat{p} \le \frac{60}{96}\right)\\
&=\mathrm{P}(0.5 \le \hat{p} \le 0.625)\\
&=\mathrm{P}\left(\frac{0.5-0.6}{\sqrt{\frac{0.6\times0.4}{96}}} \le Z \le \frac{0.625-0.6}{\sqrt{\frac{0.6\times0.4}{96}}}\right)\\
&=\mathrm{P}(-2 \le Z \le 0.5)\\
&=\mathrm{P}(0 \le Z \le 2)+\mathrm{P}(0 \le Z \le 0.5)\\
&=0.4772+0.1915=0.6687
\end{aligned}$$

5. [정답] $0.048 \le p \le 0.072$

6. [정답] (1) 신뢰구간의 길이는 길어진다.

(2) 표본의 크기 n을 크게 한다.

[풀이]

(1) 표본비율을 $\hat{p}$이라고 하면 모비율 p에 대한 신뢰도 95 %의 신뢰구간은

$$\hat{p}-1.96\sqrt{\frac{\hat{p}\hat{q}}{n}} \le p \le \hat{p}+1.96\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

또, 모비율 p에 대한 신뢰도 99 %의 신뢰구간은

$$\hat{p}-2.58\sqrt{\frac{\hat{p}\hat{q}}{n}} \le p \le \hat{p}+2.58\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

따라서 신뢰도 95 %의 신뢰구간의 길이는

$$\begin{aligned}
&\hat{p}+1.96\sqrt{\frac{\hat{p}\hat{q}}{n}}-\left(\hat{p}-1.96\sqrt{\frac{\hat{p}\hat{q}}{n}}\right)\\
&=2\times1.96\sqrt{\frac{\hat{p}\hat{q}}{n}}
\end{aligned}$$

이고, 신뢰도 99 %의 신뢰구간의 길이는

$$\hat{p}+2.58\sqrt{\frac{\hat{p}\hat{q}}{n}}-\left(\hat{p}-2.58\sqrt{\frac{\hat{p}\hat{q}}{n}}\right)=2\times2.58\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

따라서 신뢰도가 높아질수록 신뢰구간의 길이는 길어진다.

(2) 신뢰도가 일정할 때, 신뢰구간의 길이는 $\sqrt{\frac{\hat{p}\hat{q}}{n}}$의 값과 관계가 있으므로 신뢰구간의 길이를 짧게 하려면 표본의 크기 n을 크게 해야 한다.

7. [정답] 36

[풀이]

표본평균 $\overline{X}$가 정규분포 $\mathrm{N}\left(12, \frac{12^2}{n}\right)$을 따르므로

$$\begin{aligned}
&\mathrm{P}(12 \le \overline{X} \le 15)\\
&=\mathrm{P}\left(\frac{12-12}{\frac{12}{\sqrt{n}}} \le \frac{\overline{X}-12}{\frac{12}{\sqrt{n}}} \le \frac{15-12}{\frac{12}{\sqrt{n}}}\right)=\mathrm{P}\left(0 \le Z \le \frac{\sqrt{n}}{4}\right)=0.4332
\end{aligned}$$

그런데 $\mathrm{P}(0 \le Z \le 1.5)=0.4332$이므로

$$\frac{\sqrt{n}}{4}=1.5,\ n=36$$

8. [정답] 625 이상
[풀이]
신뢰도 95 %의 신뢰구간의 길이가 $2\times2\sqrt{\dfrac{\hat{p}\hat{q}}{n}}$ 이므로

$$2\times2\sqrt{\frac{\hat{p}\hat{q}}{n}}\le 0.08,\ \ \frac{\hat{p}\hat{q}}{n}\le 0.0004$$

$$n\ge\frac{\hat{p}\hat{q}}{0.0004} \qquad \cdots\cdots ㉠$$

그런데 n이 어떤 $\hat{p}$에 대해서도 ㉠을 만족시키려면 $\hat{p}\hat{q}$의 최댓값보다 크거나 같아야 한다.

$$\hat{p}\hat{q}=\hat{p}(1-\hat{p})=-\hat{p}^2+\hat{p}=-\left(\hat{p}-\frac{1}{2}\right)^2+\frac{1}{4}$$

에서 $\hat{p}\hat{q}$의 최댓값은 $\dfrac{1}{4}=0.25$이므로

$$n\ge\frac{0.25}{0.0004}=625$$

따라서 표본의 크기를 625 이상으로 해야 한다.

9. [정답] ②
[풀이]
정규분포 $\mathrm{N}(m,\ \sigma^2)$을 따르는 모집단에서 크기가 n인 표본을 임의추출하여 추정한 모평균의 신뢰구간의 길이는 $2k\dfrac{\sigma}{\sqrt{n}}$ (k는 상수)이다.
한편, 신뢰도를 높일 때, 신뢰구간의 길이를 일정하게 하려면 표본의 크기를 크게 하면 된다.
그러므로 ㄷ은 옳지 않다.
따라서 옳은 것은 ㄱ, ㄴ이다.

10. [정답] 62
[풀이]
모평균 m에 대한 신뢰도 95 %의 신뢰구간은

$$\bar{x}-1.96\frac{\sigma}{\sqrt{n}}\le m\le\bar{x}+1.96\frac{\sigma}{\sqrt{n}}$$

이므로 신뢰구간의 길이는 $2\times1.96\times\dfrac{\sigma}{\sqrt{n}}$이다. 신뢰구간의 길이가 $\dfrac{1}{2}\sigma$ 이하이므로

$$2\times1.96\frac{\sigma}{\sqrt{n}}\le\frac{1}{2}\sigma$$

$$\sqrt{n}\ge 7.84,\ \ n\ge 61.4656$$

따라서 n의 최솟값은 62이다.

11. [정답] ①

[풀이]

한 상자에 들어 있는 과자 16개의 평균 무게를 $\overline{X}$라고 하면 $\overline{X}$는 정규분포 $\mathrm{N}\left(20,\ \frac{2^2}{16}\right)$, 즉

$\mathrm{N}\left(20,\ \left(\frac{1}{2}\right)^2\right)$을 따르므로 출하한 상자가 반품될 확률은

$\mathrm{P}(\overline{X} \le 19.18$ 또는 $\overline{X} \ge 20.82)$

$=\mathrm{P}(\overline{X} \le 19.18)+\mathrm{P}(\overline{X} \ge 20.82)$

$=\mathrm{P}(Z \le -1.64)+\mathrm{P}(Z \ge 1.64)$

$=2\mathrm{P}(Z \ge 1.64)$

$=2\{0.5-\mathrm{P}(0 \le Z \le 1.64)\}$

$=2\times(0.5-0.45)=0.1$

한편, 과자 상자 100개 중에서 반품되는 상자의 수를 확률변수 Y라고 하면 Y는 이항분포 $\mathrm{B}(100,\ 0.1)$을 따른다.

이때 n이 충분히 크므로 Y는 정규분포 $\mathrm{N}(10,\ 3^2)$을 따른다. 따라서 구하는 확률은

$\mathrm{P}(Y \le 7)=\mathrm{P}\left(\frac{Y-10}{3} \le \frac{7-10}{3}\right)$

$=\mathrm{P}(Z \le -1)$

$=0.5-\mathrm{P}(0 \le Z \le 1)$

$=0.5-0.34=0.16$

12. [정답] ⑤

[풀이]

모평균 m에 대한 신뢰도 95 %의 신뢰구간의 길이가 7.84이므로

$2\times 1.96\frac{\sigma}{\sqrt{n}}=7.84,\ \frac{\sigma}{\sqrt{n}}=2$

따라서

$\mathrm{P}(\overline{X} \ge m+3.92)$

$=\mathrm{P}\left(\frac{\overline{X}-m}{2} \ge \frac{m+3.92-m}{2}\right)$

$=\mathrm{P}(Z \ge 1.96)$

$=0.5-\mathrm{P}(0 \le Z \le 1.96)=0.5-0.475=0.025$

13. [정답] ③

[풀이]

모비율을 p라고 하면 표본의 크기는 n이므로 신뢰도 95 %의 신뢰구간의 길이 l은

$l=2\times 1.96\sqrt{\frac{\hat{p}\hat{q}}{n}}$ ⋯⋯ ㉠

또, $\mathrm{P}(-k \le Z \le k)=\frac{\alpha}{100}$라고 할 때, 신뢰도 α %로 추정한 신뢰구간의 길이가 $\frac{l}{1.96}$이므로

$\frac{l}{1.96}=2\times k\sqrt{\frac{\hat{p}\hat{q}}{n}}$ ⋯⋯ ㉡

㉠, ㉡에서

$2\times 1.96\sqrt{\frac{\hat{p}\hat{q}}{n}}=2\times 1.96\times k\sqrt{\frac{\hat{p}\hat{q}}{n}}$

이므로 $k=1$이다. 따라서

$\mathrm{P}(-1 \le Z \le 1)=2\mathrm{P}(0 \le Z \le 1)$

$=2\times 0.3413=0.6826$

이므로 $\alpha=100\times 0.6826=68.26$

14. [정답] 30.5

[풀이]

$\mathrm{P}(-k \leq Z \leq k)=\frac{x}{100}$ $(k>0)$라고 하면

$$f(x)=2k\frac{5}{\sqrt{25}}=2k$$

$f(x_1)=2$에서 $k=1$이므로

$$\frac{x_1}{100}=2\times 0.3413, \quad x_1=68.26$$

또 $f(x_2)=5$에서 $k=2.5$이므로

$$\frac{x_2}{100}=2\times 0.4938, \quad x_2=98.76$$

$$x_2-x_1=98.76-68.26=\mathbf{30.5}$$

15. [정답] 85.84

[풀이]

표준편차를 σ라고 하면 신뢰도 95 %로 추정한 신뢰구간의 길이가 l이므로

$$l=2\times 1.96\frac{\sigma}{\sqrt{n}}$$

$$\frac{3}{4}l=\frac{3}{4}\times 2\times 1.96\frac{\sigma}{\sqrt{n}}=2\times 1.47\frac{\sigma}{\sqrt{n}}$$

이므로 신뢰구간의 길이가 $\frac{3}{4}l$인 신뢰도는

$$\mathrm{P}(|Z|\leq 1.47)=2\mathrm{P}(0\leq Z\leq 1.47)$$

$=2\times 0.4292$

$=0.8584$

따라서 $a=85.84$이다.

16. [정답] 2213(명) 이상

[풀이]

모비율을 p, 표본의 크기를 n이라고 하면

$\hat{p}=0.64$이므로 모비율 p에 대한 신뢰도 95 %의 신뢰구간은

$$\hat{p}-1.96\sqrt{\frac{0.64\times 0.36}{n}}\leq p$$

$$\leq \hat{p}+1.96\sqrt{\frac{0.64\times 0.36}{n}}$$

$$|\hat{p}-p|\leq 1.96\sqrt{\frac{0.64\times 0.36}{n}}$$

이때 모비율과 표본비율의 차가 0.02 이하이므로

$$1.96\sqrt{\frac{0.64\times 0.36}{n}}\leq 0.02$$

$$\sqrt{n}\geq 47.04, \quad n\geq 2212.7616$$

따라서 조사 대상자는 2213명 이상이다.