

III_3. 통계적 추정

[12확통03-05] 모집단과 표본의 뜻을 알고
표본추출의 원리를 이해한다.

[12확통03-06] 표본평균과 모평균의 관계를 이해하고
설명할 수 있다.

[12확통03-07] 모평균을 추정하고, 그 결과를 해석할 수 있다.

□ 1 모집단과 표본

- (1) 통계 조사에서 조사의 대상이 되는 집단 전체를 ‘모집단’이라고 하고, 조사를 하기 위하여 모집단에서 뽑는 일부분을 ‘표본’이라고 한다. 또한 모집단에서 표본을 뽑는 것을 ‘추출’이라고 한다.
- (2) 통계 조사에서 모집단 전체를 조사하는 것을 ‘전수조사’라고 하고, 모집단의 일부분, 즉 표본을 조사하는 것을 ‘표본조사’라고 한다. 또한 표본에 포함된 대상의 개수를 ‘표본의 크기’라고 한다.
- (3) 모집단에서 표본을 추출할 때, 모집단에 속하는 각 대상이 같은 확률로 추출되도록 하는 방법을 ‘임의추출’이라고 한다.

② 모평균과 표본평균 ①

- (1) 어떤 모집단에서 조사하고자 하는 특성을 나타내는 확률변수를 X 라 할 때, X 의 평균, 분산, 표준편차를 각각 ‘모평균’, ‘모분산’, ‘모표준편차’라 하고, 기호로 각각 m , σ^2 , σ 와 같이 나타낸다.
- (2) 모집단에서 임의추출한 크기가 n 인 표본을 X_1, X_2, \dots, X_n 이라 할 때, 이 표본의 평균, 분산, 표준편차를 각각 ‘표본평균’, ‘표본분산’, ‘표본표준편차’라 하고, 기호로 \overline{X} , S^2 , S 와 같이 나타낸다. 이때 \overline{X} , S^2 , S 는 다음과 같이 구한다.

② 모평균과 표본평균 ②

$$\textcircled{1} \quad \overline{X} = \frac{X_1 + X_2 + \dots + X_n}{n}$$

$$\textcircled{2} \quad S^2 = \frac{1}{n-1} \left\{ (X_1 - \overline{X})^2 + (X_2 - \overline{X})^2 + \dots + (X_n - \overline{X})^2 \right\} \quad (\text{단, } n \geq 2)$$

$$\textcircled{3} \quad S = \sqrt{S^2}$$

☑ 표본분산 S^2 을 구할 때, 표본평균 \overline{X} 를 구할 때와 달리 $n-1$ 로 나누는 것은 표본분산과 모분산의 차이를 줄이기 위함이다.

③ 표본평균의 확률분포 ①

모평균 m 은 고정된 상수이지만 표본평균 \bar{X} 는 임의추출된 표본에 따라 여러 가지 값을 가질 수 있으므로 확률변수이다. 따라서 \bar{X} 의 확률분포를 구할 수 있다.

예) 모집단의 확률변수 X 의 확률분포가 오른쪽 표와 같을 때, 이 모집단에서

| X | 1 | 2 | 3 | 합계 |
|----------|---------------|---------------|---------------|----|
| $P(X=x)$ | $\frac{1}{3}$ | $\frac{1}{3}$ | $\frac{1}{3}$ | 1 |

임의추출한 크기가 2인 표본을 (X_1, X_2) 라 하면

(X_1, X_2) 와 그 표본평균 $\bar{X} = \frac{X_1 + X_2}{2}$ 는 다음과 같다.

③ 표본평균의 확률분포 ②

| (X_1, X_2) | (1, 1) | (1, 2) | (1, 3) | (2, 1) | (2, 2) | (2, 3) | (3, 1) | (3, 2) | (3, 3) |
|--------------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| \bar{X} | 1 | 1.5 | 2 | 1.5 | 2 | 2.5 | 2 | 2.5 | 3 |

따라서 확률변수 \bar{X} 의 확률분포를 표로 나타내면 오른쪽과 같다.

| \bar{X} | 1 | 1.5 | 2 | 2.5 | 3 | 합계 |
|----------------------|---------------|---------------|---------------|---------------|---------------|----|
| $P(\bar{X}=\bar{x})$ | $\frac{1}{9}$ | $\frac{2}{9}$ | $\frac{1}{3}$ | $\frac{2}{9}$ | $\frac{1}{9}$ | 1 |

$$\begin{aligned}
 P(\bar{X} = 1.5) &= P(X = 1) \times P(X = 2) + P(X = 2) \times P(X = 1) \\
 &= \frac{1}{3} \times \frac{1}{3} + \frac{1}{3} \times \frac{1}{3} = \frac{2}{9}
 \end{aligned}$$

④ 표본평균의 평균, 분산, 표준편차 ①

모평균이 m , 모분산이 σ^2 인 모집단에서 크기가 n 인 표본을 임의추출할 때, 표본평균 \bar{X} 에 대하여 다음이 성립한다.

$$(1) E(\bar{X}) = m \quad (2) V(\bar{X}) = \frac{\sigma^2}{n} \quad (3) \sigma(\bar{X}) = \frac{\sigma}{\sqrt{n}}$$

☑ 모집단의 확률변수 X 의 확률분포가 오른쪽 표와 같을 때, 모평균 m ,

| X | 1 | 2 | 3 | 합계 |
|----------|---------------|---------------|---------------|----|
| $P(X=x)$ | $\frac{1}{3}$ | $\frac{1}{3}$ | $\frac{1}{3}$ | 1 |

모분산 σ^2 , 모표준편차 σ 는 각각 다음과 같다.

$$m = \frac{1}{3} \times (1 \times 1 + 2 \times 1 + 3 \times 1) = \frac{6}{3} = 2$$

④ 표본평균의 평균, 분산, 표준편차 ②

$$\sigma^2 = \frac{1}{3} \times (1^2 \times 1 + 2^2 \times 1 + 3^2 \times 1) - 2^2 = \frac{2}{3}$$

$$\sigma = \sqrt{\frac{2}{3}} = \frac{\sqrt{6}}{3}$$

이 모집단에서 임의추출한 크기가 2인 표본을 (X_1, X_2) 라

할 때, 그 표본평균 $\bar{X} = \frac{X_1 + X_2}{2}$ 가 갖는 값은 1, 1.5,

2, 2.5, 3이고 확률변수 \bar{X} 의 확률분포를 표로 나타내면 오른쪽과 같다.

| \bar{X} | 1 | 1.5 | 2 | 2.5 | 3 | 합계 |
|----------------------|---------------|---------------|---------------|---------------|---------------|----|
| $P(\bar{X}=\bar{x})$ | $\frac{1}{9}$ | $\frac{2}{9}$ | $\frac{1}{3}$ | $\frac{2}{9}$ | $\frac{1}{9}$ | 1 |

④ 표본평균의 평균, 분산, 표준편차 ③

$$E(\bar{X}) = \frac{1}{9} \times (1 \times 1 + 1.5 \times 2 + 2 \times 3 + 2.5 \times 2 + 3 \times 1) = 2$$

$$V(\bar{X}) = \frac{1}{9} \times (1^2 \times 1 + 1.5^2 \times 2 + 2^2 \times 3 + 2.5^2 \times 2 + 3^2 \times 1) - 2^2 = \frac{1}{3}$$

$$\sigma(\bar{X}) = \sqrt{\frac{1}{3}} = \frac{\sqrt{3}}{3}$$

이때 표본의 크기가 $n = 2$ 이므로

④ 표본평균의 평균, 분산, 표준편차 ④

$$\frac{\sigma^2}{n} = \frac{2}{3} \times \frac{1}{2} = \frac{1}{3}, \quad \frac{\sigma}{\sqrt{n}} = \frac{\sqrt{6}}{3} \times \frac{1}{\sqrt{2}} = \frac{\sqrt{3}}{3}$$

따라서 다음이 성립한다.

$$\textcircled{1} E(\bar{X}) = 2 = m$$

$$\textcircled{2} V(\bar{X}) = \frac{1}{3} = \frac{\sigma^2}{n}$$

$$\textcircled{3} \sigma(\bar{X}) = \frac{\sqrt{3}}{3} = \frac{\sigma}{\sqrt{n}}$$

⑤ 표본평균의 분포 ①

모평균이 m , 모표준편차가 σ 인 모집단에서 임의추출한 크기가 n 인 표본의 표본평균 \bar{X} 에 대하여 다음이 성립한다.

(1) 모집단이 정규분포 $N(m, \sigma^2)$ 을 따르면

표본평균 \bar{X} 는 정규분포 $N\left(m, \frac{\sigma^2}{n}\right)$ 을 따른다.

(2) 모집단이 정규분포를 따르지 않을 때에도 표본의 크기 n 이 충분히 크면 표본평균 \bar{X} 는 근사적으로 정규분포

$N\left(m, \frac{\sigma^2}{n}\right)$ 을 따른다.

⑤ 표본평균의 분포 ②

예 ① 정규분포 $N(12, 6^2)$ 을 따르는 모집단의 확률변수를 X 라 하고, 이 모집단에서 크기가 9인 표본을 임의추출할 때 표본평균을 \bar{X} 라 하면

$$m = E(X) = 12, \sigma^2 = V(X) = 6^2 = 36$$

이므로

$$E(\bar{X}) = E(X) = 12, V(\bar{X}) = \frac{V(X)}{n} = \frac{36}{9} = 4$$

따라서 표본평균 \bar{X} 는 정규분포 $N(12, 2^2)$ 을 따른다.

5 표본평균의 분포 ③

② 정규분포 $N(12, 6^2)$ 을 따르는 모집단에서 크기가 9인 표본을 임의추출할 때 표본평균 \bar{X} 는 정규분포 $N(12, 2^2)$

을 따르므로 확률변수 $Z = \frac{\bar{X} - 12}{2}$ 는 표준정규분포

$N(0, 1)$ 을 따른다. 따라서 $P(\bar{X} \geq 15)$ 의 값을 오른쪽 표준정규분포표를 이용하여 구하면

$$\begin{aligned} P(\bar{X} \geq 15) &= P\left(Z \geq \frac{15 - 12}{2}\right) \\ &= P(Z \geq 1.5) \\ &= 0.5 - P(0 \leq Z \leq 1.5) \\ &= 0.5 - 0.4332 = 0.0668 \end{aligned}$$

| z | $P(0 \leq Z \leq z)$ |
|-----|----------------------|
| 0.5 | 0.1915 |
| 1.0 | 0.3413 |
| 1.5 | 0.4332 |
| 2.0 | 0.4772 |
| 2.5 | 0.4938 |

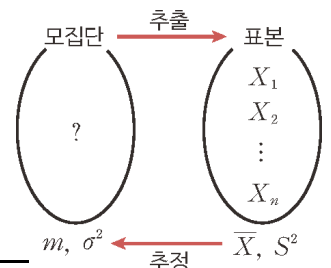
6 모평균의 추정 ①

(1) 모집단에서 추출한 표본에서 얻은 자료를 이용하여 모집단의 어떤 성질을 확률적으로 추측하는 것을 ‘추정’이라고 한다.

(2) 정규분포 $N(m, \sigma^2)$ 을 따르는 모집단에서 크기가 n 인 표본을 임의추출하여 구한 표본평균 \bar{X} 의 값이 \bar{x} 일 때, 모평균 m 에 대한 신뢰구간은 다음과 같다.

① 신뢰도 95%의 신뢰구간

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



② 신뢰도 99%의 신뢰구간

$$\bar{x} - 2.58 \times \frac{\sigma}{\sqrt{n}} \leq m \leq \bar{x} + 2.58 \times \frac{\sigma}{\sqrt{n}}$$

6 모평균의 추정 ②

☑ ① 모집단이 정규분포 $N(m, \sigma^2)$ 을 따를 때, 크기가 n 인

표본을 임의추출하면 표본평균 \bar{X} 는 정규분포 $N\left(m, \frac{\sigma^2}{n}\right)$

을 따르고, 확률변수 $Z = \frac{\bar{X} - m}{\frac{\sigma}{\sqrt{n}}}$ 은 표준정규분포

$N(0, 1)$ 을 따른다. 이때 표준정규분포표에서

$P(-1.96 \leq Z \leq 1.96) = 0.95$ 이므로

$$P\left(-1.96 \leq \frac{\bar{X} - m}{\frac{\sigma}{\sqrt{n}}} \leq 1.96\right)$$

6 모평균의 추정 ③

$$= P\left(-1.96 \times \frac{\sigma}{\sqrt{n}} \leq \bar{X} - m \leq 1.96 \times \frac{\sigma}{\sqrt{n}}\right)$$

$$= P\left(-1.96 \times \frac{\sigma}{\sqrt{n}} \leq m - \bar{X} \leq 1.96 \times \frac{\sigma}{\sqrt{n}}\right)$$

$$= P\left(\bar{X} - 1.96 \times \frac{\sigma}{\sqrt{n}} \leq m \leq \bar{X} + 1.96 \times \frac{\sigma}{\sqrt{n}}\right) = 0.95$$

따라서 모집단으로부터 크기가 n 인 표본을 임의추출하여 구한 표본평균 \bar{X} 의 값을 \bar{x} 라 할 때,

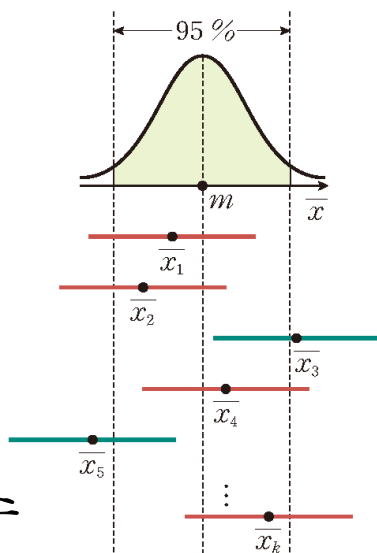
$$\bar{x} - 1.96 \times \frac{\sigma}{\sqrt{n}} \leq m \leq \bar{x} + 1.96 \times \frac{\sigma}{\sqrt{n}}$$

를 ‘모평균 m 에 대한 신뢰도 95%의 신뢰구간’이라고 한다.

6 모평균의 추정 ④

모평균 m 의 신뢰도 95%의 신뢰구간

② 표본평균 \bar{X} 는 확률변수이므로 추출되는 표본에 따라 표본평균 \bar{X} 의 값 \bar{x} 가 달라지고 그에 따라 신뢰구간도 달라진다. 이와 같은 신뢰구간 중에는 오른쪽 그림과 같이 모평균 m 을 포함하는 것과 포함하지 않는 것이 있을 수 있다. 따라서 모평균 m 에 대하여 신뢰도 95%의 신뢰구간이란 크기가 n 인 표본을 여러 번 임의추출하여 신뢰구간을 각각 구하면 그 중에는 95%는 모평균 m 을 포함할 것으로 기대된다는 것을 의미한다.



6 모평균의 추정 ⑤

예 정규분포 $N(m, 12^2)$ 을 따르는 모집단에서 크기가 36인 표본을 임의추출하여 구한 표본평균 \bar{X} 의 값이 70일 때, 모평균 m 에 대한 신뢰도 95%의 신뢰구간을 구해 보자.

$\sigma = 12, n = 36, \bar{X} = 70$ 이므로

$$70 - 1.96 \times \frac{12}{\sqrt{36}} \leq m \leq 70 + 1.96 \times \frac{12}{\sqrt{36}}$$

$$66.08 \leq m \leq 73.92$$

☑ 모평균에 대한 신뢰구간을 구할 때 모표준편차 σ 의 값을 알 수 없는 경우가 많다. 이 경우 표본의 크기 n 이 충분히 크면 모표준편차 σ 대신 표본표준편차 s 를 이용하여 모평균에 대한 신뢰구간을 구할 수 있다는 것이 알려져 있다.

☆ 신뢰도 $\alpha\%$ 의 모평균 m 의 신뢰구간

$$(1) P(\underline{|Z| \leq k}) = \frac{\alpha}{100} \Rightarrow \left| \frac{\bar{X} - m}{\frac{\sigma}{\sqrt{n}}} \right| \leq k$$

$$\Leftrightarrow |\bar{X} - m| \leq k \times \frac{\sigma}{\sqrt{n}} \Leftrightarrow |m - \bar{X}| \leq k \times \frac{\sigma}{\sqrt{n}}$$

$$(2) m \text{의 신뢰구간} : \bar{X} - k \times \frac{\sigma}{\sqrt{n}} \leq m \leq \bar{X} + k \times \frac{\sigma}{\sqrt{n}}$$

$$(3) \text{신뢰구간의 길이} : l = 2k \times \frac{\sigma}{\sqrt{n}}$$

$$(4) \text{오차 } |m - \bar{X}| \text{의 한계} : k \times \frac{\sigma}{\sqrt{n}}$$