4강. 단순임의추출법(2)

◈ 담당교수 : 이기재 교수

■ 주요용어

용어	해설
목표정도(target precision)	표본설계 단계에서 그 조사를 통해서 얻고자 하는 추정량의 오차의 한계를 미리 정하게 되는데, 이것을 목표정도라고 함. 즉, 표본크기를 결정할 때 고려한 추정량의 목표 오차의 한계를 말함
절대오차의 한계	목표정도를 절대적인 값으로 나타낸 것으로 추정값의 허용 오차의 한계를 말함, 추정값의 크기에 대한 사전 지식이 있을 때 활용함
상대오차의 한계	목표정도를 상대적인 값으로 나타낸 것으로 추정값의 허용 오차의 한계를 추정모수로 나눈 값을 말함, 추정값의 크기에 대한 사전 지식이 부족할 때 활용함
상대표준오차(relative standard error)	추정량의 표준오차를 모수로 나눈 값으로써 일명 추정량의 변동계수(coefficient of variation: CV)라고도 함

■ 실습하기

- 교재 70쪽 문제 8번
  * 모총계 추정에 대한 추정값,
    표준오차, 신뢰구간 작성

- 교재 72쪽 심화문제 2번
  * 모총계 및 모비율 추정에 대한
    신뢰구간 작성

■ 연습문제

1. 크기가 인 모집단에서 단순임의추출법으로 크기 인 표본을 추출하였다. 다음 설명 중에서 잘못된 것은?

① $\hat{\tau} = N \cdot \bar{y}$ $V(\hat{\tau}) = N^2 \cdot V(\bar{y})$

② $\hat{\tau} = N \cdot \bar{y}$ $V(\hat{\tau}) = N \cdot V(\bar{y})$

③ $\hat{\tau} = \bar{y}/N$ $V(\hat{\tau}) = V(\bar{y})$

④ $\hat{\tau} = \bar{y}/N$ $V(\hat{\tau}) = N \cdot V(\bar{y})$

정답 : ①
해설 :

단순임의추출법에서 모총계 τ의 비편향추정량은 $\hat{\tau} = N\bar{y}$이다.
따라서 $V(\hat{\tau}) = V(N\bar{y}) = N^2 V(\bar{y}) = N^2 \times \frac{N-n}{N}\frac{S^2}{n}$ 이다.

2. 크기가 인 모집단으로부터 크기 인 단순임의표본을 추출하였다. 다음 설명 중에서 옳지 <u>않은</u> 것은?

① 표본비율 $\hat{p}$은 모비율 p에 대한 비편향추정량이다.

② 표본분산 s^2은 모분산 S^2을 대체로 크게 추정하는 편향추정량이다.

③ $\hat{\tau} = N \cdot \bar{y}$은 모종계에 대한 비편향추정량이다.

④ $\hat{V}(\bar{y}) = \frac{N-n}{N}\frac{s^2}{n}$이다.

정답 : ②
해설 :

표본분산 s^2은 모분산 S^2에 대한 비편향추정량이다.

3. 완공시기가 서로 다른 120가구의 주택을 건설 중인 한 대형 건설회사에서는 각 건설현장의 총 재고액을 추정하기 위하여 표본 크기가 12인 단순임의표본을 추출하여 조사한 결과 $\bar{y} = 324.83$, $s^2 = 1612.33$ 으로 나타났다. 120가구 전체의 총 재고액 추정에 대한 95% 신뢰수준에서 오차의 한계는?

① 4,639.2
② 3,639.2
③ 2,639.2
④ 1,639.2

정답 : ③
해설 :

$$V(\hat{\tau}) = V(N\bar{y}) = N^2 V(\bar{y}) = N^2 \times \frac{N-n}{N} \frac{S^2}{n}$$, 오차의 한계=$2 \times \sqrt{\hat{V}(\hat{\tau})}$

4. 이 후보자의 지지율 추정값과 95% 신뢰수준에서 오차의 한계를 구하면?

※ [4-6] 다음은 어느 일간지의 발표내용이다. 다음 물음에 답하시오.

총 유권자가 명인 어느 도시에서 시장선거에 출마하고자 하는 사람이 있다. 이들 중에서 500명의 유권자를 단순임의추출하여 조사한 결과 220명이 이 후보를 지지한다고 응답하였다. <참고>

$$\hat{V}(\hat{p}) = \frac{N-n}{N} \frac{\hat{p}(1-\hat{p})}{n-1}, \quad n = \frac{n_0}{1+\frac{n_0}{N}}, \quad n_0 = \frac{4p(1-p)}{B^2}$$

① 44%, 3.4%
② 44%, 4.4%
③ 54%, 5.4%
④ 54%, 6.4%

정답 : ②
해설 :

$$\hat{p} = 220/500 = 0.44, \quad \hat{V}(\hat{p}) = \frac{N-n}{N} \frac{\hat{p}\hat{q}}{n-1} = \frac{50000-500}{50000} \frac{0.44 \times 0.56}{489} = 0.000499,$$
$$\text{오차의 한계} = z_{\alpha/2} \sqrt{\hat{V}(\hat{p})} = 2 \times \sqrt{0.000489} = 0.0442$$

5. 95% 신뢰수준에서 지지율에 대한 오차의 한계가 0.025(2.5%) 이내에 있게 하려면 표본의 크기는 얼마로 해야 하는가?
(단, 문제 4에서 구한 지지율을 사전 정보로 활용하여 구하시오.)

① 약 1,530명
② 약 2,530명
③ 약 3,530명
④ 약 530명

정답 : ①
해설 :

$$n_0 = \frac{4p(1-p)}{B^2} = \frac{4\times0.44\times0.56}{0.025^2} = 1,577$$

$$n = \frac{n_0}{1+\frac{n_0}{N}} = \frac{1577}{1+1577/50000} = 1,530$$

6. 만약 이 후보자 지지율에 대한 사전정보가 전혀 주어져 있지 않은 경우라면 표본 크기는 얼마로 해야 하는가?

① 약 1,550명
② 약 2,550명
③ 약 3,550명
④ 약 550명

정답 : ①
해설 :

모비율에 대한 사전정보가 전혀 없는 경우에는 $p=0.5$로 가정하여 표본의 크기를 결정한다. $p=0.5$로 가정할 때 구하는 표본의 크기가 가장 크게 된다.

$$n_0 = \frac{4p(1-p)}{B^2} = \frac{4\times0.5\times0.5}{0.025^2} = 1,600$$

$$n_0 = \frac{4p(1-p)}{B^2} = \frac{4\times0.5\times0.5}{0.025^2} = 1,600$$

■ 정리하기

▪ 모총계의 추정량과 분산 추정량은 각각 다음과 같다.

- $\hat{\tau} = N\bar{y}$
- $\hat{V}(\hat{\tau}) = \hat{V}(N\bar{y}) = N^2\hat{V}(\bar{y}) = N^2\frac{N-n}{N}\frac{s^2}{n}$

▪ 모비율 p의 추정량은 $\hat{p} = \frac{1}{n}\sum_{i=1}^{n} y_i$ 로 주어지는데, 이 때 y_i는 0 또는 1의 값을 갖게 된다. 한편 이 추정량의 분산의 추정량은 다음과 같다.

$$\hat{V}(\hat{p}) = \frac{N-n}{N}\frac{\hat{p}\hat{q}}{n-1}, \quad \hat{q} = 1-\hat{p}$$

▪ 모평균

μ를 추정하는데 있어서 추정량에 대한 오차의 한계가 B 이하가 되도록 하는 표본크기 n은 다음의 식에 의해서 결정한다.

$$n = \frac{N(z_{\alpha/2}S)^2}{B^2 + (z_{\alpha/2}S)^2} = \frac{n_0}{1 + \frac{n_0}{N}}$$, 여기서 $n_0 = \frac{(z_{\alpha/2}S)^2}{B^2}$

■

상대오차의 한계 D가 주어지는 경우, 표본크기 n은 다음 식에 따라서 결정한다.

$$n = \frac{N(z_{\alpha/2}C)^2}{ND^2 + (z_{\alpha/2}C)^2} = \frac{n_0}{1 + \frac{n_0}{N}}$$, 여기서, $n_0 = (\frac{z_{\alpha/2}C}{D})$, $C = S/\mu$는 모집단의 변동계수이다.

■ 참고문헌

- 이계오, 박진우, 이기재, 표본조사론, 한국방송통대학교출판부, 2013. 제1장
- 통계청 홈페이지 : http://www.nso.go.kr