

2강. 표본조사의 기본 개념

◆ 담당교수 : 이기재 교수

■ 주요용어

용어	해설
모수(parameter)	모집단의 특성값인데, 표본조사에서 다루는 대표적인 모수로 모평균, 모비율, 모총계, 모분산, 모표준편차 등이 있다.
추정량(estimator)	표본을 통해 관측이 가능한 확률변수들의 함수로서 모수에 대한 추측값이다. 대표적인 추정량으로 표본평균, 표본표준편차 등을 들 수 있다.
표본분포 (sampling distribution)	확률표본으로부터 얻어지는 추정량은 표본추출변동으로 인하여 그 값이 변하지만 표본추출을 반복하면 어떤 분포를 따르게 되는데 이 분포를 표본분포라고 한다.
오차의 한계	모수 θ 의 추정량 $\hat{\theta}$ 의 $100(1-\alpha)\%$ 신뢰구간을 구할 때 계산되는 범위를 나타내는 값으로서 $B = z_{\alpha/2} \sqrt{V(\hat{\theta})}$ 로 표현된다.
상대표준오차 (relative standard error : RSE)	추정량의 정도를 나타내는 상대적 기준으로 사용되는 값이며, 흔히 추정량의 변동계수(coefficient of variation : CV)라고도 하는데 그 공식은 다음과 같다. $RSE(\hat{\theta}) = \frac{\sqrt{V(\hat{\theta})}}{\hat{\theta}} \times 100(\%)$

■ 실습하기

- 모평균 및 모분산 계산
- * 0, 1, 2, ..., 9에 각각 1/10의 확률이 주어지는 이산형 균등분포
- * 엑셀을 이용해서 모평균과 모분산 계산
- 이산형 균등분포로부터 표본크기 10인 표본을 100회 반복해서 추출
- * 표본에 대해서 표본평균 계산
- * 표본추출변동, 표본분포, 비편향성 개념 이해

■ 연습문제

1. 오차의 한계, 상대표준오차 등의 값을 통해서 알 수 있는 것은?

- ① 표본조사 결과의 비표본오차 크기 정도
- ② 표본조사 결과의 통계적 정확도
- ③ 모집단 추출틀의 정확도
- ④ 표본조사 결과의 편향 정도

정답 : ②

해설 : 추정량의 오차의 한계, 상대표준오차 등을 통해서 표본조사 결과의 통계적 정확도 또는 표본오차의 크기를 가늠해 볼 수 있다.

2. 표본조사는 모수에 대한 추정값뿐만 아니라 오차의 한계(또는 표준오차)를 함께 제시하게 된다. 다음의 설명 중에서 옳은 것은?

- ① 오차의 한계가 작을수록 모수 추정의 신뢰구간의 폭은 넓게 된다.
- ② 오차의 한계가 작을수록 점추정량의 정확도는 높다.
- ③ 오차의 한계는 표준오차가 클수록 줄어든다.
- ④ 오차의 한계는 추정값의 편향 정도를 나타낸다.

정답 : ②

해설 : 오차의 한계가 작을수록 추정의 정확도는 높아지고, 신뢰구간의 폭은 줄어든다. 오차의 한계는 추정량의 표준오차가 작을수록 줄어든다.

3. 다음은 표본조사에서 널리 사용되고 있는 추정량의 상대표준오차에 대한 설명이다. 옳지 않은 것은?

- ① 추정량(θ)의 정도(精度)를 나타내기 위한 척도이다.

②
$$RSE(\hat{\theta}) = \frac{\sqrt{V(\hat{\theta})}}{\theta} \times 100(\%)$$

- ③ 추정량의 상대표준오차가 작으면 추정량의 정확도는 떨어진다.
- ④ 추정값의 표준오차가 작을수록 추정량의 상대표준오차는 작게 된다.

정답 : ③

해설 : 추정량의 상대표준오차(Relative Standard Error : RSE)는

$$RSE(\hat{\theta}) = \frac{\sqrt{V(\hat{\theta})}}{\theta} \times 100(\%)$$
로 정의된다. 추정량의 상대표준오차는 추정량의 표준오차가 작을수록 작아진다. 추정량의 상대표준오차가 작을수록 추정량의 정확도는 높다.

4. 다음의 괄호 안에 알맞은 용어는 무엇인가?

모수 θ 에 대한 추정량으로 $\hat{\theta}$ 를 이용하는 경우에 추정량 $\hat{\theta}$ 의 표준오차는 ()로 정의되고, $100(1-\alpha)\%$ 오차의 한계는 ()이다.

① $\sqrt{V(\hat{\theta})} = z_{\alpha/2} \sqrt{V(\hat{\theta})}$

② $z_{\alpha/2} \sqrt{V(\hat{\theta})} = \sqrt{V(\hat{\theta})}$

③ $\frac{\sqrt{V(\hat{\theta})}}{\hat{\theta}} = \sqrt{V(\hat{\theta})}$

④ $\sqrt{V(\hat{\theta})} = \frac{\sqrt{V(\hat{\theta})}}{\hat{\theta}}$

정답 : ①

해설 : $\sqrt{V(\hat{\theta})} = z_{\alpha/2} \sqrt{V(\hat{\theta})}$

5. 본조사에서 추정량을 구할 때 요구하는 가장 중요한 두 가지 성질은 무엇인가?

- ① 효율성과 편향성
- ② 효율성과 비편향성
- ③ 일치성과 편향성
- ④ 일치성과 효율성

정답 : ②

해설 : 비편향성은 모수를 추정할 때 모수를 표본추출에 따라 과대하게 또는 과소하게 추정하지 않음을 의미한다. 추정량의 효율성이란 비편향성을 만족하는 여러 개의 추정량 중에서는 추정량의 분산이 작은 것을 사용할 때 효율적이라는 의미이다.

■ 정리하기

- 표본조사의 궁극적인 목적은 모집단을 대표하는 표본을 조사하여 모집단 특성값인 모평균이나 모총계, 모비율 등의 모수에 대한 효율적인 추정량을 얻는데 있다.
- 동일한 모집단에서 같은 표본추출방법으로 같은 크기의 표본을 추출할지라도 각 표본에서 계산된 추정량의 값은 표본마다 달라진다. 이렇게 추정량의 값이 표본마다 달라지는 것을 표본추출변동이라고 한다. 확률표본으로부터 얻어지는 추정량은 표본추출변동으로 인하여, 그 값이 변하지만 표본추출에 따

른 분포를 따르게 되는데 이 분포를 표본분포(sampling distribution)라고 한다.

- 표본오차(sampling error)는 표본에서 구한 결과와 센서스의 결과(모수)의 차이로 모집단의 일부를 표본추출하여 조사하여 추정함으로써 발생하는 우연적 오차를 말한다.
- 바람직한 추정량이 되기 위해 요구되는 두 가지 중요한 성질은 비편향성과 효율성이다.
- 표본조사의 결과에 대한 정확도를 평가하는 방법은 다음과 같다.

① 추정량의 표준오차(standard error) : $\sqrt{V(\hat{\theta})}$ (추정량의 분산의 제곱근)

② 추정량의 오차의 한계 : $z_{\alpha/2} \sqrt{V(\hat{\theta})}$

③ 추정량의 상대표준오차 : $CV(\hat{\theta}) = \frac{\sqrt{V(\hat{\theta})}}{\hat{\theta}} \times 100$

추정량의 표준오차가 작을수록 추정량의 상대표준오차도 작아진다.

상대표준오차가 작을 때 추정량의 정도가 높다고 하며 그러한 값들은 믿을 만한 값으로 받아들여진다.

■ 참고문헌

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