

9강. 계통추출법

◆ 담당교수 : 이기재 교수

■ 주요용어

용어	해설
1/k계통추출법 (1-in-k systematic sampling)	모집단의 추출틀에서 처음의 k 개 단위들 중에서 랜덤하게 하나의 단위를 추출하고, 그 이후 매 번째 간격마다 하나씩의 단위를 표본으로 추출하는 방법
추출간격 (sampling interval)	크기 N 의 모집단에서 n 크기의 표본을 뽑을 때 N 을 n 으로 나누어 $k = N/n$ 를 추출간격으로 정함
분수간격법 (fractional interval method)	$N \neq nk$ 인 경우 추출간격 k 를 근사값으로 사용하지 않고, N/n (소수점 이하의 숫자 포함)의 값을 사용하는 추출방법
급내상관계수 ρ	동일한 계통표본 내에 있는 단위간의 급내상관관계 (intra-class correlation)를 나타내는 척도. 만일 값이 1에 가까우면 동일한 계통표본 내 단위들의 조사변수값들이 서로 매우 유사하다는 것을, ρ 값이 음수일 때는 동일한 계통표본내 조사값들이 매우 이질적이라는 것을 나타냄
랜덤모집단 (random population)	모집단에서 조사단위(또는 추출단위)의 배열이 조사 특성값과 아무런 관련성 없이 랜덤하게 배열되어 있는 모집단을 말함
순서모집단 (ordered population)	모집단에서 조사단위(또는 추출단위)의 특성값이 주기적으로 변동하는 모집단을 말한다. 주기모집단이라고도 함
순환모집단 (preodic population)	모집단에서 조사단위(또는 추출단위)의 배열이 조사 특성값의 크기 순서대로 나열되어 있는 모집단을 말함
반복계통추출법	하나의 계통표본 대신에 여러 개의 임의출발점을 택하여 여러 개의 계통표본을 뽑는 방법으로 모집단에 대한 특별한 가정 없이 추정량의 분산을 구할 수 있음

■ 실습하기

- 교재 166쪽 계통추출법에 의한 모평균 추정

* 모평균 추정, 표준오차, 신뢰구간 작성

■ 연습문제

이 과수원의 그루당 평균 사과수를 표본평균 \bar{y}_{11} 을 이용하여 추정하고자 한다. 추정량 \bar{y}_{11} 의 표준오차($\sqrt{\widehat{V}(\bar{y}_{11})}$)는 얼마인가?

※[1-2] 어느 과수원에서 $N = 1500$ 그루의 사과나무에 대해 1/10 계통표본을 사용하여 그루당 사과수를 추정하려 한다. 조사결과 얻어진 자료는 표본평균 $\bar{y}_{11} = 35.2$ 이고, 표본분산 $s^2 = 4.8$ 이다. (단, 모집단은 랜덤모집단으로 간주하시오.)

1.

- ① 0.47
- ② 0.37
- ③ 0.27
- ④ 0.17

정답 : ④

해설 :

$$\widehat{V}(\bar{y}_{11}) = \left(1 - \frac{n}{N}\right) \frac{s^2}{n} = \left(1 - \frac{150}{1500}\right) \times \frac{4.8}{150} = 0.029$$

2. 이 과수원의 전체 사과 총수에 대한 95% 신뢰구간은?

※[1-2] 어느 과수원에서 $N = 1500$ 그루의 사과나무에 대해 1/10 계통표본을 사용하여 그루당 사과수를 추정하려 한다. 조사결과 얻어진 자료는 표본평균 $\bar{y}_{11} = 35.2$ 이고, 표본분산 $s^2 = 4.8$ 이다. (단, 모집단은 랜덤모집단으로 간주하시오.)

- ① $52,800 \pm 1511$
- ② $52,800 \pm 511$
- ③ $35,200 \pm 1511$
- ④ $35,200 \pm 511$

정답 : ②

해설 :

$$\begin{aligned} \hat{\tau} &= N\bar{y}_{11} = 1,500 \times 35.2 = 52,800 \\ \widehat{V}(\hat{\tau}) &= N^2 \widehat{V}(\bar{y}_{11}) = 1,500^2 \times 0.029 = 652,500 \\ \text{따라서 구하고자 하는 95\% 신뢰구간은 다음과 같다.} \\ \hat{\tau} \pm z_{\alpha/2} \sqrt{\widehat{V}(\hat{\tau})} &\Leftrightarrow 52,800 \pm 2 \times \sqrt{652,500} \\ &\Leftrightarrow 52,800 \pm 511 \end{aligned}$$

3. 위의 표본추출방법에 대한 설명으로 옳지 않은 것은?

※ [3-4] 어느 지역의 총유권자는 $N = 5,450$ 명이다. 어느 후보자에 대한 이 지역

유권자들의 지지율을 알아보기 위하여 유권자 명부를 추출틀로 삼아, 1/10 계통표본을 뽑아 조사하였다. 조사 결과 298명이 지지한다고 응답하였다. 다음 물음에 답하시오.

- ① 이 조사의 추출간격은 10이다.
- ② 이 조사의 모집단은 랜덤모집단으로 간주할 수 있다.
- ③ 표본 크기는 545명이다.
- ④ 이 조사에서 계통표본의 급내상관계수는 1 근처의 값을 갖게 될 것이다.

정답 : ④

해설 : 이 조사에서 추출틀로 사용된 유권자명부는 후보자의 지지 여부에 대해서 랜덤하게 배열되어 있다고 가정할 수 있기 때문에 랜덤모집단으로 볼 수 있다. 따라서 급내상관계수는 0 근처의 값을 갖게 될 것이다.

4. 이 후보의 지지율에 대한 95% 신뢰구간을 구하면?

※ [3-4] 어느 지역의 총유권자는 $N=5,450$ 명이다. 어느 후보자에 대한 이 지역 유권자들의 지지율을 알아보기 위하여 유권자 명부를 추출틀로 삼아, 1/10 계통표본을 뽑아 조사하였다. 조사 결과 298명이 지지한다고 응답하였다. 다음 물음에 답하시오.

- ① 0.55 ± 0.04
- ② 0.55 ± 0.15
- ③ 0.77 ± 0.04
- ④ 0.77 ± 0.15

정답 : ①

해설 :

$$n = 5,450/20 = 545, \hat{p}_{10} = 298/545 = 0.547$$

$$\hat{V}(\hat{p}_{10}) = \left(1 - \frac{1}{10}\right) \frac{\hat{p}_{10} \hat{q}_{10}}{n-1} = 0.9 \times \frac{0.55 \times 0.45}{544} = 0.00041$$

오차의 한계는 $2\sqrt{\hat{V}(\hat{p}_{10})} = 0.04$ 이다.

5. 인구 및 주택 총조사는 전수조사와 표본조사를 병행하는데, 전체 조사구 중 10%를 표본으로 계통추출법으로 추출하여 해당 조사구에 대해서는 표본조사를 진행하고자 한다. 다음의 설명 중에서 옳지 않은 것은?

- ① 1/10계통추출법을 사용해야 한다.
- ② 표본 조사구의 선정이 단순임의추출법에 비해서 간편하다.
- ③ 표본추출간격은 10이다.

④ 조사원에 의한 선택오차(selection error)를 커져서 문제가 된다.

정답 : ④

해설 : 계통추출법은 표본추출이 간편해서 조사원에 의한 선택오차(selection error)를 줄일 수 있다.

6. 사업체를 대상으로 올해 매출액을 조사하고자 한다. 지난해 전체 사업체에 대해서 조사한 결과를 표본추출틀로 이용하고, 표본추출틀 내의 사업체들을 지난해 매출액 기준으로 정렬하여 계통추출법을 적용하고자 한다. 다음 설명 중에서 알맞은 것은?

- ① 이 경우는 순환모집단에 해당한다.
- ② 단순임의추출법을 적용하는 경우와 비교할 때 추정 효율이 크게 떨어질 것이다.
- ③ 이 경우에 급내상관계수는 거의 0 근처의 값을 보일 것이다.
- ④ 단순임의추출법을 적용하는 경우와 비교할 때 추정 효율이 좋아진다.

정답 : ④

해설 : 주어진 문제에서 전체 사업체들을 지난해 매출액을 기준으로 정렬하였으므로 순서모집단에 해당한다. 순서모집단의 경우에 계통추출법을 적용하면 추정의 효율이 좋아진다.

■ 정리하기

1. 추출틀로부터 처음의 k 개 단위들 중에서 랜덤하게 하나의 단위를 추출하고 그 이후 매 k 번째 간격마다 하나씩의 단위를 표본으로 추출하는 표본추출방법을 1 k 계통추출법(1-in-k systematic sampling)이라고 한다. 계통추출법의 장점은 표본의 추출이 간편하고 단순임의추출법의 대용으로 사용할 수가 있으며 대개의 경우는 표본이 모집단 전체를 잘 반영한다는 점이다.
2. 모평균 추정
① 추정량 : $\hat{\mu} = \bar{y}_{1..} = \frac{1}{n} \sum_{i=1}^k y_i$
② 추정량의 분산 : $V(\bar{y}_{1..}) = \frac{N-1}{N} \frac{S^2}{n} [1 - (n-1)\rho]$, 여기서 ρ 는 급내상관계수
$\hat{V}(\bar{y}_{1..}) = \frac{N-n}{N} \frac{s^2}{n}$, $s^2 = \frac{1}{n-1} \sum_{i=1}^k (y_i - \bar{y})^2$
③ 모평균에 대한 100(1- α)% 신뢰구간
$(\bar{y}_{1..} - z_{\alpha/2} \sqrt{\hat{V}(\bar{y}_{1..})}, \bar{y}_{1..} + z_{\alpha/2} \sqrt{\hat{V}(\bar{y}_{1..})})$

3. 모총계 추정

① 추정량 : $\hat{\tau}_{yz} = N \cdot \bar{y}_{yz}$

② 추정량의 분산 계산 : $\hat{V}(\hat{\tau}_{yz}) = N^2 \hat{V}(\bar{y}_{yz}) = N^2 \frac{N-n}{N} \frac{s^2}{n}$

③ 모총계에 대한 100(1- α)% 신뢰구간

$$(\hat{\tau}_{yz} - z_{\alpha/2} \sqrt{\hat{V}(\hat{\tau}_{yz})}, \hat{\tau}_{yz} + z_{\alpha/2} \sqrt{\hat{V}(\hat{\tau}_{yz})})$$

4. 모비율 추정

① 모비율의 추정량 : $\hat{p}_{yz} = \frac{1}{n} \sum_{i=1}^n y_i$

② 추정량의 분산 계산 : $\hat{V}(\hat{p}_{yz}) = \frac{N-n}{N} \frac{\hat{p}_{yz} \hat{q}_{yz}}{n-1}$ 단, $\hat{q}_{yz} = 1 - \hat{p}_{yz}$

③ 모비율에 대한 100(1- α)% 신뢰구간

$$(\hat{p}_{yz} - z_{\alpha/2} \sqrt{\hat{V}(\hat{p}_{yz})}, \hat{p}_{yz} + z_{\alpha/2} \sqrt{\hat{V}(\hat{p}_{yz})})$$

5. 계통추출법에서 모집단은 랜덤모집단(random population), 순서모집단(ordered population), 순환모집단(periodic population)

등으로 구분한다. 순서모집단에 계통추출법을 적용하면 단순임의추출법보다 효율적이지만, 순환모집단일 경우는 그 유효가 떨어진다.

■ 참고문헌

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