

절사법과 추정

Cut off Sampling and Estimation

한근식*

사업체조사에서 절사법이 흔히 이용되고 있다. 이 경우, 절사되는 사업체가 모집단에서 차지하는 비중이 모집단 전체에 비하여 매우 작다. 본 연구에서는 사례를 이용하여 모집단을 절사층, 표본층, 전수층으로 구분하고 절사층을 추정하는 문제를 살펴보고자 한다.

In business survey, cut off sampling is usual. The contribution from cut off part of the population is at least small in comparison with the remaining population. In this study, the population have been separated as three stratum, cut-off, take-some, take-all, and we will estimate cut-off part.

1. 개요

사업체관련 표본설계시 소규모 사업체(장)를 추출틀에서 제외하는 경우가 흔하다. 즉, 절사되는 사업체들이 모집단 전체에 미치는 영향이 무시할만한 경우 절사법(cut off method)을 흔히 이용하고 있다. 이와 같은 방법은 소규모 사업체들에게 응답에 대한 부담을 줄여주는 효과도 있다. 그러나 절사법은 절사되는 사업체들에게 추출확률을 0으로 제한한다는 면에서는 이론위주의 연구자들에게 환영받는 방법이 아님은 확실하다. 그럼에도 불구하고 이 방법을 이용한다면 “절사점(cut off threshold)은 어떻게, 어느 곳을 선택해야 하는가?” 라는 문제에 직면하게 된다.

절사법과 관련한 이론적, 방법론적 연구는 극히 제한적이나 Sarndal, Swensson, Wretman(1992)은 부분적으로나마 다양한 설명을 하고 있으며, 한근식(1996)등은 일반적인 내용을 다양한 자료를 이용해 보여주고 있다. 한편, 절사점의 선택에 대해서는 이론적이기 보다 경험적 사례가 발표되고 있는데 Statistics Canada(2001)에서 출간되는 “Data Quality Statements”를 보면 Canadian Monthly Survey of Manufacturing의 경우 각 주(province)에서 생산액의 2%에 해당하는 최하위 사업체를 제외한 사업체들을 목표모집단(target population)으로 설정하고 있다. 스웨덴에서 실시한 어느 광공업(mining and

* 한신대학교 정보과학대학 정보시스템학과 교수

quarrying and manufacturing)실태조사는 종업원 수 10인 이하는 절사한 후 조사하는 것으로 보고 되고 있다.

통계청에서도 다수의 표본조사를 절사법을 이용하여 시행하고 있으며, 공식통계로 지정 받은 많은 조사가 절사법에 의해 시행되고 있는 것이 현실이다. 비용의 측면에서 볼 때 조사 의뢰자는 절사점을 증가시키고 싶을 것이고 실제 절사점은 증가되고 있는 것이 사실이다. 그렇다면 “절사점을 어느 선까지 증가시킬 것인가?”라는 문제를 만나게 된다.

II. 절사법

여기에서 이용할 자료는 2002년도 광업, 제조업 조사자료의 일부로서 한국표준산업분류상 2자리 숫자에 해당하는 대분류 산업이며 모집단의 크기는 N=9211 사업체이다. 이 모집단에서 생산액(단위: 1000원)과 종업원수(단위: 명)가 중요한 변수가 되며 생산액을 추정하는 것이 표본조사의 목적에 해당한다. 이 자료에서 9,211개 사업체의 총 생산액은 91,429,079(천)원이다. 이 모집단에서 허용오차를 5%로 하는 표본조사를 절사법을 이용하여 시행하는 경우 총 표본의 크기는 $n = 555$ 개사업체이다. 이 중 전수층은 366개 사업체이며 8,845개로 구성된 표본층에서 189개의 사업체를 추출하면 된다. 아래 <표 1>은 허용오차를 5%로 주었을 때 절사점의 증가에 따른 표본의 크기와 목표모집단의 변화를 보여 주고 있다.

<표 1> 허용오차 5%에서 절사점의 이동에 따른 표본의 크기

| 절사점(단위: 1000원) | 목표모집단 크기 | 총표본 크기 | 전수층 크기 | 표본층 크기 | 모집단 총 계(생산액) | 목표모집단의 비율(%) |
|-------------------|-------------|-----------|-----------|-----------|-----------------|-----------------|
| 0 | 9,211 | 555 | 366 | 189 | 91,429,079 | 100 |
| 500 | 5,193 | 428 | 282 | 146 | 90,618,278 | 99.11 |
| 1,000 | 3,592 | 361 | 241 | 120 | 89,464,981 | 97.85 |
| 2,000 | 2,367 | 296 | 203 | 93 | 87,738,319 | 95.96 |
| 3,000 | 1,826 | 262 | 184 | 78 | 86,408,260 | 94.51 |
| 4,000 | 1,493 | 240 | 177 | 63 | 85,250,835 | 93.24 |
| 5,000 | 1,265 | 222 | 159 | 63 | 84,233,513 | 92.13 |
| 6,000 | 1,093 | 207 | 146 | 61 | 83,291,202 | 91.09 |

위 <표 1>에서 보는바와 같이 생산액이 500이하인 사업체를 제외한 후의 목표모집단은 원래 목표모집단의 99.11%를 반영하고 있어 절사법을 이용한다하더라도 목표모집단에서의

손실되는 부분은 미미하다. 반면에 목표모집단의 크기는 9,211에서 5,193으로 약 45%가량 줄어 들었다. 생산액이 작은 4,018개의 사업체를 제외함으로써 표본조사에서의 비표본오차를 상당부분 줄였다. 이는 전년도 조사에서 무응답한 사업체들을 분석한 결과 표본에 추출된 사업체 중 생산액이 500 이하인 사업체의 응답율이 20%에 미치지 못하였다는 것으로 알 수 있다. 이제 생산액이 1,000이하인 사업체를 절사한 후 표본조사를 시행한다고 생각하자. 이때 의 목표모집단은 원 모집단의 97.85%를 반영하게 된다. 그리고 목표모집단의 크기는 3,592개 사업체로 원 모집단 크기의 39%로 줄어들었다. 그 만큼 비표본오차가 줄어들 것이다. 이와 같이 절사점을 증가시키면 6,000이하 사업체를 절사하는 경우, 목표모집단의 크기는 9211에서 1093으로 줄어들게 되고 이때 목표모집단은 원 모집단의 91.09%를 반영하게 된다.

III. 절사층의 추정

요즈음 표본조사에서 가장 큰 문제는 무응답을 줄이는 것이다. 표본설계자는 무응답을 줄이기 위해 답례품의 질을 높이고, 조사원에 대한 교육을 강화하고, 설문 문항을 줄이는 등 응답자의 부담을 줄이려고 노력하고 있다.

같은 방법으로 조사연구자의 부담을 줄이는 방법은 없을까? 만약 위 <표 1>에서 보는 바와 같이 6,000이하를 절사하는 방법으로 표본조사를 시행하고 6,000이하의 부분을 추정할 수는 없을까? 활용 가능한 적절한 보조변수가 있다면 부분적으로 추정이 가능하다. 아래 <표 2>는 절사점의 이동에 따른 목표모집단내에서의 생산액과 종업원수간의 상관계수를 보여주고 있다.

<표 2> 절사점의 이동에 따른 생산액과 종업원 수간의 상관계수

| 절사점 | 0 | 500 | 1,000 | 2,000 | 3,000 | 4,000 | 5,000 | 6,000 |
|------|-------|-------|-------|-------|-------|-------|-------|-------|
| 상관계수 | 0.500 | 0.495 | 0.496 | 0.488 | 0.485 | 0.478 | 0.472 | 0.465 |

<표 2>에서 보는바와 같이 모집단의 생산액과 종업원간의 상관계수는 0.500이며 생산액이 500이하인 사업체를 제외한 목표모집단에서의 생산액과 종업원과의 상관계수는 0.495, 생산액이 1,000 이하인 사업체를 제외한 목표모집단에서의 생산액과 종업원과의 상관계수는 0.496 등이다.

<표 3> 절사점의 이동에 따른 모집단의 절사비율

| | | | | | | | | |
|----------|---|------|-------|-------|-------|-------|-------|-------|
| 절사점 | 0 | 500 | 1,000 | 2,000 | 3,000 | 4,000 | 5,000 | 6,000 |
| 모집단의절사비율 | 0 | 0.89 | 2.15 | 4.04 | 5.49 | 6.76 | 7.87 | 8.91 |

위 <표 3>에서 보는바와 같이 절사점의 이동에 따라 모집단의 절사 비율이 달라진다. 500이하를 절사하는 경우 절사법을 이용한 추정치에 문제가 없으나 2,000이하, 3,000이하를 절사한 후 추정치는 원 모집단 총계의 4.04%, 5.49%가 각각 제외된 것이다. 5,000이하, 6,000이하를 절사한 후의 추정치는 원 모집단 총계의 7.87%, 8.91%가 각각 제외된 것이다. 따라서 절사점이 큰 경우, 제외된 부분에 대한 추정을 고려할 필요가 있다.

<그림 1> 절사층을 포함한 모집단의 표현

| | | |
|--------------|----------------|---------------|
| 절사층(cut-off) | 표본층(take-some) | 전수층(take-all) |
|--------------|----------------|---------------|

위 그림에서 보는바와 같이 표본층과 전수층의 생산액에 대한 추정은 절사법으로 가능하다. 이제 절사층의 추정은 <표 2>의 상관관계를 이용하여 비 추정, 회귀추정 등의 방법을 이용하여 추정이 가능하다.

참고문헌

- 한근식·김용철.1996. “왜도가 심한 모집단에서의 절사법 효과에 관한 연구.” 『응용통계연구』 9.
- 통계청. 2002. 『광업, 제조업 통계조사보고서』.
- Sarnda, C. E., Swensson, and J, Wretmen. 1992. *Model Assisted Survey sampling*. New York: Springer-Verlag.
- Statistics Canada. 2001. “Monthly Survey of Manufacturing.” *Statistical Data Documentation System*.