

標本配分에 관한 소고

김 중 호¹⁾

요 약

표본조사에 있어서 층화추출법은 모집단에 관한 예비정보를 필요로 하고 있다. 조사자가 표본설계시 층화와 표본배분의 문제를 막연히 추상적으로 처리함으로써 생기는 오류를 줄이기 위해서는 다윈적 입장에서 모집단에 대한 예비 정보를 정확하게 파악하고 이용하여 층화추출법의 효율을 올릴 수 있음을 지적하고 있다.

1. 서론

오늘날 많은 통계조사는 표본조사에 의하여 실시되고 있는 실정이다. 최근 두드러지게 나타나고 있는 사회조사에 있어서도 모두 표본조사기법에 따르고 있다. 정보화사회에 있어서 통계 정보 생산을 위한 표본조사기법은 필요 불가결한 도구로 되어 있고 가장 많은 응용분야를 갖고 있으며 과학적인 조사기법으로 인정을 받고 있는 것이다.

표본조사법은 표본추출의 문제와 추정의 문제로 대별하고 있다. 이 논문에서는 층화추출에 있어서 층화의 방법과 각 층에 표본배분에 관한 문제를 논하는 경우에 피하기 어려운 많은 현실적인 피리를 어떻게 해결하는 것이 좋은가에 대하여 알아 보려고 한다. 표본조사는 이론과 실제와의 적합으로 현실적 해결이 중요하기 때문에 현 상황을 도외시하면 오류를 많이 범하게 된다. 이러한 점을 감안하여 표본설계를 계획할 때에는 표본조사이론가와 조사를 실제 수행하는 통계전문가인 조사자와 통계를 이용하는 이용자가 같이 참여하여 모든 구조적인 문제점을 현실적으로 해결할 수 있는 방향에서 설계해야 한다.

2. 층화추출법

일반적으로 모집단은 매우 크고 구성요소가 복잡 다양하게 이루어져 있다. 이러한 모집단으로부터 보다 대표성 있는 표본을 추출하기 위하여 모집단을 층화(stratification)하고 각 층에 할당된 표본을 각 층으로부터 추출하는 방법이 층화추출법(stratified random sampling)이다. 이러한 층화추출법에 있어서 중요한 것은 층화를 어떻게 해야 하는가 하는 문제와 각 층에 정해진 표본을 어떻게 배분(allocation)해야 좋은가 하는 문제이다.

1) (100-715) 서울특별시 중구 필동 3가, 동국대학교 통계학과 교수

3. 총화의 문제

모집단을 총화하는데 여러 가지 목표가 있을 수 있다. 그 중에 제일 먼저 그리고 중요하게 생각해야 할 목표는 정도를 높이는 문제이다. 통계는 정확이 생명이기 때문에 정도를 높이기 위한 총화는 일차적 목표가 되어야 한다. 총화의 목적이 정도를 높이는데 있다면 총내에서 조사량간에 등질성(homogeneous)이 유지되게 총화하는 것이 필수조건이다. 총내에서 등질성을 측정하는 양은 총내분산이다.

우리가 현실적으로 조사를 하는 경우에 모집단을 구성하고 있는 조사단위를 모르고 있으며 그 조사단위를 알고자 하는 것이 조사인 것이다. 실제에 있어서는 조사량에 대한 지식이나 정보가 없기 때문에 총화를 한다는 것은 추상적인 이론으로 구체적인 어려움이 따른다. 물론 경험에 의하여 사전정보를 갖고 있다면 예외가 될 수 있다. 모집단에 관한 예비정보가 있다면 이 정보를 이용하여 정도를 높이기 위한 총화의 방법을 찾을 수 있다. 그러나 이 때에도 예비정보가 얼마나 정확하냐에 따라 총화의 효과는 달라진다. 그런대로 보가 있는 경우에는 정도를 높이기 위한 총화가 가능하다. 모집단에 대한 사전정보가 없는 경우에는 조사량과 상관이 있는 보조량의 정보를 갖고 이 정보를 이용하여 총화를 하는 방법을 제안하고 있다. 물론 조사량 x 와 상관관계가 높은 보조량 y 의 정보를 이용하여 조사량 x 를 조사하기 위한 총화를 하는 경우에 총화의 효과는 크게 나타나는 것이 일반적으로 알려져 있다. x , y 간의 상관의 정도가 높을수록 총화의 효율은 높다. 그러나 보조량 y 의 정보를 갖고 있다 해도 y 가 x 와 얼마나 상관이 있는가 하는 문제는 쉽게 판명할 수 있다. 보조량 y 가 x 와 상관이 높다고 하는 전제 속에서 총화를 할지라도 실제로 x , y 가 얼마나 상관이 있는가를 계산하고 조사량 x 를 총화하는 것은 역시 쉬운 일이 아니다.

모집단에 대한 정보가 없고 총화의 어떠한 방법을 찾지 못하면 다른 추출법을 생각해야 되지만 계층별 통계의 필요성에 따라 총화추출법을 이용해야 되는 경우가 있게 된다.

4. 표본배분의 문제

통계생산을 위한 조사에서 가장 먼저 중요하게 생각해야 할 사항은 정확성과 예산이다. 표본조사에서 정보를 최대한 높이면서 조사비용은 최소가 되게 하는 방법은 없다. 우리는 다음의 두가지 원칙 중 어느 하나에 따라야 한다.

- (1) 요구된 정도에 통계를 최소의 비용으로 조사한다.
- (2) 정해진 비용하에서 가장 정도 높은 통계를 작성한다.

일차적으로 총화를 잘하면 정도가 높아지는 것으로 되어 있지만 후속으로 정해진 표본의 크기 n 을 각 층에 어떻게 배분하는 것이 정도를 더 높일 수 있고 한편으로는 비용을 절약할 수 있는가와도 관련이 있다. 표본배분에는 여러 가지 방법이 있으나 (1), (2)의 조건하에서 최적배분법(optimum allocation)이 정도와 비용의 면에서 고려되는 방법이다. (1), (2)의 어느 조건하에서든 최적배분의 표본수 n_i 는 다음과 같이 구해진다.

$$n_i = \frac{N_i S_i / \sqrt{C_i}}{\sum_{i=1}^L (N_i S_i / \sqrt{C_i})} \cdot n \quad (i = 1, 2, \dots, L)$$

- 여기서 n : 표본의 크기
 L : 층의 수
 N_i : i 번째 층의 크기
 S_i^2 : i 번째 층의 분산
 C_i : i 번째 층의 조사단위 조사비용

물론 비례배분법이 이용하기 편리한 점이 있어서 현실적으로 비례배분법을 많이 이용하고는 있으나 비용과 정도의 문제를 충분히 고려해야 하는 경우에는 비례배분법은 문제가 된다. 따라서 비용과 정도가 전제되는 조사에서는 최적배분법을 이용해야 한다. 그런데 앞의 공식에서와 같이 n_i 를 구하자면 N_i 와 C_i 그리고 S_i^2 와 같은 정보가 필요하다. N_i 와 C_i 는 대체적으로 알수 있는 경우가 많으나 S_i^2 과 같은 층내분산은 아는 경우가 드물고 어떠한 방법으로든 추정해야 한다. 그리고 최적배분인 경우에 표본수 n 을 정하는 공식은 다음과 같다. ($D^2 = V(\bar{X}_{st})$).

$$n = \frac{(\sum_i^L N_i S_i \sqrt{C_i})(\sum_i^L N_i S_i / \sqrt{C_i})}{N^2 D^2 + \sum_i^L N_i S_i^2}$$

이 n 을 구하는 식에서도 n_i 를 구하는 식에서와 같이 N_i , C_i , S_i^2 과 같은 정보가 있어야 한다. n 과 n_i 를 구하기 위하여 예비정보로서 N_i , C_i , S_i^2 등이 알려져 있어야 한다.

5. 결론

표본조사를 층화추출법에 의한다고 할 때 층화나 표본배분을 하기 위하여 예비정보가 반드시 확보되어야 한다. 이론적 전개와 실제적 적용에 있어서 조사자가 사실적인 문제를 현실적으로 부각시키지 못하는 것은 실제과학인 통계가 형식논리에서 끝나 애매한 통계생산을 낳게 하는 원인을 제공하는 것이다.

우리는 적당한 방법으로 표본조사를 하고 있다. 표본이론가와 통계생산실무자와 그리고 이용자가 같은 차원에서 협력하여 현실적인 괴리를 극복한다고 할 때 표본조사법은 하나의 조사기법으로 통계생산에 기여할 것이고 바르게 정착될 것이다.

참고문헌

- [1] Levy P. S. & Lemeshow S. (1991). *Sampling of Populations (Method and Application)*, John Wiley & Sons, New York.
- [2] Hansen, M. H., Hurwitz, W. N. and Madow, W. G. (1953). *Sample Survey Methods and Theory*, Vol. I, II, John Wiley & Sons, New York.
- [3] Cochran, W. G. (1963). *Sampling Techniques*, John Wiley & Sons, New York.
- [4] Kish, L. (1965). *Survey Sampling*, John Wiley & Sons, New York.
- [5] Yates, F. (1960). *Methods for Censuses and Surveys*, Charles Griffin.
- [6] Yamane, T. (1967). *Elementary Sampling Theory*, Prentice Hall.
- [7] Deming, W. E. (1950). *Some Theory of Sampling*, John Wiley & Sons, New York.