Journal of Basic Science, Dec. 1993
Basic Sicence Research Institite
Sungshin Women's University
Vol. 11. 109-119

표본크기 추정에 관한 연구

이 해 용
(성신여자대학교 자연과학대학 통계학과)

A Study on the Sample Size Estimation

Hae Yong Lee

(Dept. of Statistics, Sungshin Women's University, Seoul 136-742, Korea)

Abstract

The many variables measured in a survey and the multiple uses of survey data make the determination of a reasonable sample size to use for a sample survey an onerous task for the statistician. And sometimes we are faced with determining how many peoples will be selected for a survey of a public opinion. This paper reviews several methods to determine the sample size for estimating the parameters of a multinomial population and the survey of a public opinion and compares them.

Ⅰ. 서 론 (INTRODUCTION)

통계적 추론을 위해서는 추론의 목적에 부합되는 자료가 있어야 하며, 자료는 이미 주어진 경우와 주어져 있지 않은 경우가 있다. 자료가 주어진 경우에는 그 자료를 대상으로 원하는 통계분석을 실시하면 되겠지만, 자료가 주어지지 않은 경우에는 분석목적에 맞는 자료를 수집해야 한다. 표본의 크기추정 문제는 바로 표본이 주어져 있지 않은 경우에 원하는 통계분석을 위하여 몇개의 표본을 추출하여야 할 것인가를 결정하는 문제다. 이것은 사실 새로운 통계를 능동적으로 생산하는 첫 단계로 자료수집을 위하여 요구되는 자료수집에 소요되는 비용, 시간, 인력계획 등 제반 자료수집의 계획에 연관되어 있어 통계분석의 사활이 걸린 중요한 문제

일 뿐아니라 생산되는 통계의 질을 결정하는 문제이기도 하다. 따라서 이 문제는 오래 전부터 활발히 연구 되어 왔다. 그러나 Hansen & Hurwitz & Madow(1953) 등에 따르면 표본의 크기 추정문제는 주로 단일 모집단의 모수인 모집단 평균, 모집단 비율 등을 주 대상으로 연구되어 왔음을 확인할 수 있으며, 그후 Queensberry and Hurt(1964), Goodman(1965) 및 Tortora(1978) 그리고 Thomson(1987) 등은 실제 문제에 자주 접하게 되는 다항분포의 비율을 추정하기 위하여 표본크기 추정에 관한 연구를 한 바있다. 그리고 현재 사회과학에서 널리 사용되고 있는 설문지 조사를 위한 표본크기 추정은 비율 추정을 이용하거나, 아니면 Galtung(1967)이 제안한 방법을 사용하고 있는 형편이다.

본 연구에서는 자료수집을 통하여 다항분포를 가정한 비율추정 및 분할표(contingency table)의 독립성검정을 수행하려는 경우 표본크기 추정에 관한 제 방법을 살펴보고 그들이 앉고 있는 문제점과 활용방안을 검토한 다음, 각 경우에 따른 표본크기 추정을 좀더 정확히 하여 연구에 활용케 하고자 한다.

Ⅱ. 표본크기 추정의 기본이론

2.1 단일 모수추정을 위한 표본크기추정.

표본크기 추정의 기본 개념은 다음 식(2−1)과 같이 추론하고자 하는 모수를 θ라하고 그 모수의 추정치를 $\hat{\theta}$라 할 때 추정치와 모수의 차이 즉 표본오차(Sampling Error) d가 일정한 값 C 이내가 될 확률 값이 일정하게 되도록 표본의 크기를 결정하는 것을 원칙으로 하고 있다.

$$P\{|\hat{\theta}-\theta|\leq C\}=(1-\alpha) \quad \cdots\cdots (2-1)$$

예를들어 표본평균을 모집단평균의 추정치로 사용하기로 하는 경우 표본크기를 결정하는 문제를 보면 먼저 표본오차 d의 절대값이 모집단평균의 $100(1-\alpha)\%$ 구간추정의 1/2이 되도록 표본의 크기를 결정하면 표본오차는 C값의 이내가 되고 이때 표본오차가 모집단평균 구간추정치의 1/2이내에 있을 확율은$(1-\alpha)$가 되게 되어 표본크기 추정원칙을 만족하는 표본의 크기를 구할 수 있는 이론을 형성하게 된다. 즉,

$$P\{|\bar{X}-\mu|\leq C\}=(1-\alpha) \quad \cdots\cdots (2-2)$$

식 (2−2)를 다시 다음 식 (2−3)과 같이 변형할 수 있는데, 이는 이 식내에 구하고자 하는 표본의 크기 n이 내포되어 이를 구하면 곧 이 값이 표본크기추정 원칙을 따르는 값이 되기 때문이다.

$$P\{\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\leq Z_{\alpha/2}\}=(1-\alpha) \quad \cdots\cdots (2-3)$$

단 $Z_{\alpha/2}$ 는 표준정규분포의 값으로 α값에 따라서 계산되어질 수 있는 값이다. 따라서 식 (2−3)에서 $(\bar{X}-\mu)=d$라 놓고 식 (2−3)의 왼쪽 식의 내부를 정리하면 다음 식 (2−4)와 같이 정리된다.

$$d=Z_{\alpha/2}\times\sigma/\sqrt{n} \quad \cdots\cdots (2-4)$$

식 (2−4)에서 표본오차 d는 α로 표시되는 신뢰도 $Z_{\alpha/2}$와 추정치의 표준편차 $\sigma/\sqrt{n}$ 의 곱의 형태로 나타나는데 이는 곧 오차의 한계를 모수의 신뢰구간의 1/2이내로 하는 추정치를 구하는 조건을 만족하는 표본크기의 추정치를 구하는 것과 같다. 이식을 다시 n에 대해서 정리하면 구하고자 하는 표본크기 추정치가 얻어지게 되는데 이는 다음 식 (2−5)와 같다.

$$n=\left(\frac{Z_{\alpha/2}\times\sigma}{d}\right)^2 \quad \cdots\cdots (2-5)$$

결국 표본의 크기추정문제는 추정치의 분포와 신뢰수준(Significance Level) $(1-\sigma)$에 해당하는 모수의 구간추정치로 부터 구할 수 있다.

이상에서 살펴본 바와 같이 표본크기 추정문제는 추정하고자 하는 모수와 그 모수의 추정치에 대한 분포를 구한 다음 그 분포로 부터 추정치의 신뢰구간을 구하고 그 신뢰구간의 1/2과 표본오차 d를 동일하게 놓고 그 관계식에서 표본의 크기 n을 구하는 것으로 결론지을 수 있다. 그러나 이때 고려해야 할 사항은 추정치의 분포를 이용하여 표본의 크기를 추정한 결과 표본의 크기추정치가 표본의 크기의 함수로 표시되는 경우가 있게 되는데 이러한 경우에 이를 다루는 문제가 남게된다. 예를들어 추정치의 표준화식이 (n−1)의 자유도를 가지는 t−분포를 따른다고 할 경우 식 (2−5)에서 $Z_{\alpha/2}$ 대신 자유도 (n−1)인 t−분포에서 구한 값 $t_{\alpha/2}(n-1)$을 사용하게 되는데 이경우 표본의 크기추정치가 다시 표본의 크기에 따라 변하는 값으로 표시되기 때문이다.

2.2 다항 모수추정을 위한 표본크기 추정

다항분포의 경우에서 처럼 모수가 다수인 경우 이를 추정하기 위해서는 이를 동시에 고려하여 표본을 추출하여야 하고, 또한 다변량 분포에서 모수들의 차이를 추정한다거나 검정하고자할 경우에 요구되는 표본크기의 추정 문제나, 다수의 변수와 문항으로 구성된 설문지를 이용하여 자료를 구한다음 통계분석을 하려 할 때 조사대상으로 선정할 표본크기결정 문제 등도 기본적인 원칙은 단일모집단의 모수추정을 위하여 표본의 크기를 추정하는 것과 같이 행해지고 있다. 그러나 표본추출의 목적과 대상이 다르므로 경우에 따라서 사례별로 연구되어 오고 있으며 그 결과도 다양하다. k개의 모수를 동시에 추정하기 위한 표본크기를 추정한다고 하면 다음 식 (2−6)과 같이 k개의 모수를 모두 포함하는 동시 구간추정구간(simultaneous confidence interval)의 확률이 $(1-\alpha)$을 만족하도록 하는 표본을 구하는 것으로 단일모집단의 모수추정을 위한 경우를 확장하여 표본크기를 추정하는 것과 유사한 방법으로 하고 있다.

$$n = \min_{n \in I^+} \ni \left[\Pr\left(\bigcap_{i=1}^{k} [\pi_i^- \leq \pi_i \leq \pi_i^+] = 1-\alpha \right) \right] \cdots\cdots (2-6)$$

단, π_i^-는 π_i에 대한 구간추정치의 하한값이고, π_i^+는 π_i는 구간추정치의 상한값이며, I^+는 양의정수의 집합을 의미한다. 그러나 위 식 (2−6)의 식에서 결합확률을 구하기가 복잡하여 이산다변량으로 간주하여 표본크기 추정치를 다음 식 (2−7)과 같이 구하고 있다.

$$n = \min_{n \in I^+} \ni \left[\sum_{i=1}^{k} \alpha_i \leq \alpha \right] \cdots\cdots (2-7)$$

여기서 $\alpha_i = \Pr(\pi_i^- \leq \pi_i \leq \pi_i^+)$을 의미하고 있다.

Ⅲ. 다항분포의 모수추정을 위한 표본크기추정

다항분포의 모수추정을 위한 표본크기추정 문제에서 Tortora(1978)와 Thompson(1987)은 이들 구간추정치의 하한값 π_i^-과 상한값 π_i^+를 다음 식 (3−1)과 같이 정의하여 사용하고 있다.

$$\pi_i^- = \pi_i - Z_{1-\alpha/2k}\sqrt{\frac{\pi_i(1-\pi_i)}{n}}$$

$$\pi_i^+ = \pi_i - Z_{1-\alpha/2k}\sqrt{\frac{\pi_i(1-\pi_i)}{n}} \cdots\cdots (3-1)$$

여기서 π_i는 π_i의 최우추정치를 나타내고, $Z_{1-\alpha/2k}$는 표준정규분포에서 확률 $1-\alpha/2k$에 해당하는 값이다. 그러나 Tortora와 Thompson의 구체적인 차이점은 식 (2−7)에서 π_i의 값을 어떻게 사용하느냐에 달려 있다.

3.1 Tortora 방법

Tortora는 식 (2−7)을 근거로 하되 개별적인 모수의 구간추정치를 구하여 표본의 크기를 추정한 다음 표본의 크기 추정치가 최대인 값을 구하고자 하는 표본의 추정치로 결정하는 방법으로 이를 식으로 표현하면 다음 식 (3−2)와 같다.

$$n = \min \ni \left[Z_{1-\alpha/2k}\sqrt{\frac{\pi_i(1-\pi_i)}{n}} \leq d_i \text{ for } _{i=1,2\cdots k} \right] \cdots\cdots (3-2)$$

식 (3−2)에서 표본의 크기는 양의 정수이어야 하므로 다음 식 (3−3)과 같이 나타낼 수 있다.

$$n = 1 + \mathrm{int}\left(\max_{i \in 1,2\cdots k} \left[\frac{Z^2_{(1-(\alpha/2k))}\pi_i(1-\pi_i)}{d_i^2} \right] \right) \cdots\cdots (3-3)$$

위 식 (3－3)에서 π_i를 모르는 경우에는 표본의 크기를 최대로 하는(worst-case) 표본의 크기를 구하기 위하여 $\pi_i=1/2$로 하면 n은 다음 식 (3－4)와 같게 된다.

$$n=1+\mathrm{int}\left(\max_{i\in 1,2\cdots k}\left[\frac{0.25Z^2_{(1-(\alpha/2k))}}{d_i^2}\right]\right) \cdots\cdots (3-4)$$

따라서 식 (3－3)으로부터 a_i와 d_i 및 π_i 값에 따른 표본의 크기를 구하면 다음 표 (3－1)과 같다.

표(3－1) $d_i=0.05$로 놓은 경우 π_i 및 α 값에 따른 표본의 크기

α	k	n
0.5000	4	236
0.4000	4	271
0.3000	3	271
0.2000	3	335
0.1000	3	454
0.0500	3	574
0.0250	2	623
0.0200	2	664
0.0100	2	790
0.0050	2	922
0.0010	2	1245
0.0005	2	1307
0.0001	2	1518

3.2 Thompson 방법

Thompson 도 식 (2－7)을 근거로 표본의 크기를 추정하되 다만 Tortora와 달리 다음 식 (3－5)와 같이 각 추정치의 일정 구간에 있을 확률의 합이 α이하가 되도록 정의하여 표본크기를 추정하고 있다.

$$n=\min\ni[\Sigma\Pr(\pi_i-d_i\leq\pi_i\leq\pi_i+d_i)\leq\alpha] \cdots\cdots (3-5)$$

여기에서 $d_i=Z_{(1-\alpha/2)}\pi_i(1-\pi_i)/\sqrt{n}$이다. 그러나 Thompson의 방법은 동시에 다수의 모수에 대한 구간추정치를 구한다는 것이 어려워 실제로는 $d_i=d$라 놓고 모수 k개 중에서 0(zero)이 아닌 모수의 수를 m(<k)이라 할 때 m개의 모수를 동시에 고려하는 표본크기추정은 다음 식 (3－6)과 같이 하고 있다.

$$n=1+\mathrm{int}\left(\max_{m\in 3,4,\cdots k}\frac{Z^2_{(1-(\alpha/2m))}(1/m(1-1/m)}{d^2}\right) \cdots\cdots (3-6)$$

이 식으로부터 α와 m 및 d_i값에 따른 표본의 크기를 구하면 다음 표(3-2)와 같다.

표(3-2) d=0.05로 놓은 경우 α 및 m에 따른 표본의 크기

α	m	n
0.5000	4	177
0.4000	4	203
0.3000	3	241
0.2000	3	299
0.1000	3	403
0.0500	3	510
0.0250	2	624
0.0200	2	664
0.0100	2	788
0.0050	2	915
0.0010	2	1212
0.0005	2	1342
0.0001	2	1645

3.3 Bromaghin 방법

Bromaghin(1993)은 Tortora와 Thompson의 방법과 달리 각 모수의 구간추정치의 하한값과 상한값을 다음 식 (3-7)과 같이 정의하여 표본크기의 추정치를 구할 것을 제안하고 있다.

$$\pi_i = \frac{Z^2_{(1-(\alpha i/2))} + 2n_i - Z_{(1-(\alpha i/2))}\sqrt{Z^2_{(1-(\alpha i/2))} + 4n_i(\frac{n-n_i}{n})}}{2(n+Z^2_{(1-(\alpha i/2))})}$$

$$\pi_i = \frac{Z^2_{(1-(\alpha i/2))} + 2n_i + Z_{(1-(\alpha i/2))}\sqrt{Z^2_{(1-(\alpha i/2))} + 4n_i(\frac{n-n_i}{n})}}{2(n+Z^2_{(1-(\alpha i/2))})} \cdots\cdots\cdots\cdots (3-7)$$

따라서 구하고자 하는 표본크기의 추정치는 Tortora의 경우와 같은 방법으로 다음 식 (3-8)과 같이 얻을 수 있다.

$$n = \min_{n \in I^+} \ni \left[\frac{Z_{(1-(\alpha i/2))}\sqrt{Z^2_{(1-(\alpha i/2))} + 4n_i(\frac{n-n_i}{n})}}{2(n+Z^2_{(1-(\alpha i/2))})} \le d_i \text{ for } i=1,2,\cdots,k \right] \cdots\cdots (3-8)$$

이 식에서 n_i를 모를 때는 그 추정치로 $n\pi_i$을 대신 사용하면 위 식 (3-8)은 다음 식 (3-9)와 같이 변현됨을 알 수 있다.

$$n=\min_{n\in I^+}\ni\left[\frac{Z_{(1-(\alpha i/2))}\sqrt{Z^2_{(1-(\alpha i/2))}+4n\pi_i(1-\pi_i)}}{2(n+Z^2_{(1-(\alpha i/2))})}\le d_i \text{ for } i=1,2,\cdots,k\right] \quad \cdots\cdots\cdots\cdots (3-9)$$

이 식으로부터 주어진 하나의 π_i에 대하여 표본의 크기 n을 구하면 다음 식 (3-10)과 같다.

$$n=\left(\frac{Z^2_{(1-(\alpha i/2))}}{2d_i^2}\right)\left[\pi_i(1-\pi_i)-2d_i^2+\sqrt{\pi^2_i(1-\pi_i)^2-d_i^2[4\pi_i(1-\pi_i)-1]}\right] \quad \cdots\cdots\cdots\cdots (3-10)$$

위 식에서 표본의 크기 n을 구하기 위해서는 모수 π_i의 추정치를 알고 있어야 한다. 그러나 사전에 이 값을 아는 것은 어려우므로 실제로 표본의 크기를 추정하는 경우에는 표본의 크기가 가장 크게 되도록 $\pi_i=0.5$로 하여 다음 식 (3-11)과 같이 구하고 있다.

$$n=1+\text{int}\left(\max_{i\in\{1,2,\cdots k\}}\left[\frac{0.25Z^2_{(1-(\alpha i/2))}}{d_i^2}-Z^2_{(1-(\alpha i/2))}\right]\right) \quad \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots (3-11)$$

이 결과는 앞에서 언급한 Tortora(1978) 방법에 의해서 구한 표본크기의 추정값과 거의 유사하나 전체적으로 약간 작게 나타나고 있으며(참고문헌 (1)참조), 또한 $d_i=d$로 $a_i=a/k$으로 놓고 보면 식 (3-10)은 다음 식 (3-12)와 같이 됨을 알 수 있다.

$$n=n_T\left(\frac{0.25Z^2_{(1-(\alpha i/2k))}}{1/m(1-1/m)Z^2_{(1-(\alpha/2m))}}\right)-Z^2_{(1-(\alpha/2k))} \quad \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots (3-12)$$

단, 여기서 n_T는 Thompson(1987)이 제안한 표본크기 추정치를 나타내고 있다.

표(3-3) d=0.05인 경우 k와 α에 따른 표본의 크기

α　k	2	3	4	5	6	7	8	9	10
0.10	264	449	498	536	568	595	618	639	657
0.05	380	568	618	657	690	717	741	762	781
0.01	656	853	905	946	979	1007	1032	1053	1072

Ⅳ. 분할표 분석을 위한 표본의 크기추정

4.1 Galtung 방법

앞에서 검토된 표본크기 추정문제는 모두 한 변수에 대한 다수의 카테고리를 중심으로 논의된 것이다. 그러나 많은 경우에는 두변수 이상을 동시에 분석하는 다변량분석에 해당하는 경우가 허다하다. 그 예로 설문지 조사를 통하여 두 변수간의 독립성 검정여부를 따지는 검정을 실시하려 할 때 표본은 몇개로 할 것인가 하는 물음은 자주 접하는 문제이다. 설문지 조사의 경우에는 표본의 크기가 다름아닌 설문지 조사매수와 같다. 이와 같은 경우에 표본의 크기는 각 카테고리에 최소한 10개 혹은 20개 이상의 사례수가 있어야 신뢰도를 확보할 수 있다는 전자헹 Galtung(1967)은 변수의 수를 r이라하고 변수의 카테고리의 수를 C라 하면 표본의 크기추정치는 다음 식 (4－1) 혹은 식 (4－2)와 같이 추정하도록 권고하고 있다.

$$n = 10 \times C^r \quad \cdots\cdots (4-1)$$

$$n = 20 \times C^r \quad \cdots\cdots (4-2)$$

다음 표 (4－1)은 카테고리 수 C와 변수의 수 r에 따른 표본의 크기추정치를 보여주고 있다.

표(4－1) 변수의 수와 카테고리수에 따른 표본의 크기

		각 변수의 카테고리 수(C)			
		2개	3개	4개	5개
변수의 수 (r)	1개	20(40)	30(60)	40(80)	50(100)
	2개	40(40)	90(180)	160(320)	250(500)
	3개	80(40)	270(540)	640(1280)	1250(2500)
	4개	160(40)	810(1620)	2560(5120)	6250(12500)

4.2 Jai Won Choi 방법

Jai Won Choi(1987)은 Two-Way 분활표에서 도수를 이용한 자료분석을 하기 위한 표본의 크기를 추정하기 위하여 CV(coefficient of variation)를 이용 방법을 발표한 바 있다. 이에 따르면 Two-Way 분활표로 부터 카이자승검정을 위하여 소요되는 표본의 크기를 추정하는데 있어 Two-Way의 분활표의 각 칸(cell)의 비율의 최소값을 대상으로 표본의 크기를 추정하기 위하여 다음 식 (4－3)을 이용하도록 권고하고 있다.

$$n = \frac{(1-\pi_i)}{(CV_s)^2 \pi_i} \quad \cdots\cdots (4-3)$$

단, 여기에서 π_i는 주어진 분활표에서 칸(cell)의 비율이 최소인 값이고, CV는 변동계수(Coefficient of Variation)으로 다음 식 (4－4)와 같다.

$$CV = \frac{(1-\pi_i)}{\sqrt{n\pi_i}} \quad \cdots\cdots (4-4)$$

따라서 이 식에 의해서 CV와 π_i에 값에 따른 표본의 크기 n을 표로 정리한 것은 다음 표(4 −2)와 같다.

표(4−2) CV와 π_i 값에 따른 표본의 크기 n

		π_i 값							
		0.005	0.010	0.020	0.100	0.200	0.300	0.500	0.800
CV 값	0.01	1,999,000	990,000	490,000	90,000	40,000	23,333	10,000	2,500
	0.05	76,600	39,600	19,600	3,600	1,600	934	400	100
	0.25	3,184	1,584	784	144	64	38	16	4
	0.30	2,211	1,100	544	100	45	26	11	3
	0.50	796	396	196	36	16	10	4	1
	0.80	311	155	77	14	7	4	2	1
	0.90	246	122	71	11	5	3	2	1
	1.00	199	99	49	9	4	3	1	1

V. 결론 및 제언

위에서 살펴본 바와 같이 다항분포의 모수를 추정하기 위해서 어느 추정방법을 사용하여 표본의 크기를 구하느냐에 따라 그 차이가 나타나고 있음을 알 수 있다. 또한 그 차이의 정도는 유의수준 α, 허용오차 d_i, 신뢰수준 $100(1-\alpha)\%$ 및 범주(category) m의 수에 따라서 서로 다르게 된다. 그러나 한가지 사실은 위에서 설명한 다항분포의 모수를 추정하는 세가지 방법의 결과를 직접적으로 비교하는 것은 어렵지만 표(3−1)과 표(3−2) 및 표(3−3)에서 α, k 및 d가 동일한 조건 하에서 표본의 크기를 비교하면 Tortora Bromaghin Thompson 방법 순이다. 예를 들어 $\alpha=0.05$, $d_i=0.05$, $100(1-\alpha)=95\%$ 그리고 범주 $k=m=3$에서 각 방법의 경우 소요 표본추정치 n은 Tortora 방법의 경우 574개이고, Bromaghin 방법의 경우는 568개이며, hompson 방법의 경우는 510개이다. 따라서 표본의 크기를 작게 하려면 Thopson 방법을 사용하는 것이 바람직하나 그 추정 방법이 다른 방법에 비하여 조금 복잡하다는 것을 지적할 수 있다.

또한 분할표로 정리되어 분석될 것을 대비한 범주형자료분석을 위한 자료수집을 목적으로 하는 경우의 표본의 크기추정 문제에 대해서는 Galtung 방법과 Jai Won Choi 방법을 설명하였다. 그러나 살펴본 바와 같이 이 두가지 추정방법은 추정하는 기본방법이 서로 다르기 때문에 직접 비교할 수는 없다. Galtung 방법은 변수 r과 카테고리수 C만을 가지고 표본의 크기를 추정하고 있고, Jai Won Choi 방법은 성공확률 α_i 및 변동계수 CY 값을 이용하여 표본의 기를 추정하고 있다. 그리고 표본 크기추정의 식 (4−1) 혹은 (4−2)와 (4−3) 및 (4−4) 식으로 부터 구한 표본 크기의 추정치는 그 추정치의 편차가 크게 나타나고 있다.

결과적으로 표본의 크기추정을 하기 위해서는 무엇보다도 먼저 조사된 자료를 가지고 어떤 통계분석을 할 것인가를 먼저 확정하고 그 분석에 맞는 자료를 수집할 수 있도록 해야할 것이

다. 예를 들어 조사한 자료를 범주형으로 구분하여 변수간에 독립성여부를 조사하려 한다면 각 범주에 최소한 5개 이상의 자료가 있어야 한다는 권고사항이 있으므로 이와 같은 사실도 배려하여 표본의 크기를 추정하는 것이 바람직하다. 이 이외에도 표본의 크기추정은 이론만을 절대적으로 의존하기에는 어려움이 있다. 표본의 크기추정에서 얻은 결과는 곧 그 숫자만큼의 조사를 해야 하는데 이를 제약하는 주변여건이 많이 있기 때문이다. 즉 조사비용, 조사시간 및 조사인력구성 등은 표본의 크기추정 못지 않게 실무에서 중요한 요인이므로 이러한 요인을 감안하여 조사할 표본의 크기추정을 해야 할 것이다. 또한 중요한 사실은 0.01정도의 신뢰도를 높이기 위하여 표본을 몇 십 혹은 몇 백개를 더 추출하여야 하는 문제에 대해서도 한번쯤 생각할 문제이다.

국 문 요 약

조사연구를 위하여 다수의 변수를 조사할 경우 몇개의 표본을 추출하느냐하는 것은 중대한 문제이며, 이를 결정하는 문제 또한 간단한 것 만은 아니다. 본 연구에서는 이러한 문제를 해결하고자 지금까지 이러한 문제를 다루어온 여러가지의 이론을 정리하였고, 이들 이론의 특성을 살펴봄으로서 다변량 분포의 모수추정 및 다양한 여론조사에 있어 요구되는 최적의 표본의 크기추정 문제를 다루었다.

References

(1) Bromaghin, J. F.(1993), "Sample size determination for interval estimation of multinomial probabilities", *The A,erican Statistician*, 47, 203-206.

(2) Fitzpatrick S., and Scott A.,(1987), "Quick Simultaneous confidence interval for multinomial proportions", *Journal of the American Statistician Association*, Vol. 82, No. 399, 875-878

(3) Galtung, J.,(1967), *"Theory and Methods of Social Research"*, George Allen & Unurin Ltd.

(4) Goodman, Leo A.(1965), "On Simultaneous Confidence Intervals for Multinomial Proportions", *Technometrics*, 7, 247-254.

(5) Hansen, M.H., Hurwitz, W.N., and Madow, G.W.,(1953), *"Sample Survey Methods and Theory"*. Vol. Ⅰ, Ⅱ, New York: John Wiley & Sons, Inc.

(6) Jai Won Choi(1987), *"Use of coefficient of variation for a sample size of a two-way table"*, Analysis of categorical data from cluster data, National Center for Health Statistics, 245-255.

(7) Queensbury, C.P., and Hurst, d.c.,(1964), "Large Sample Simultaneous Confidence Intervals for Multinomial Proportions" *Technometrics*, 6, 191-195.

(8) Thompson, S.K.(1987), "Sample size for estimating multinomial proportions", *The American Statistician*, 41, 42–46.
(9) Thompson, S.K.(1992), *"Sampling"*, John Wilely & Sons Inc, 31–44.
(10) Tortora, R. D.(1978), "A note on sample size estimation for multinomial populations", *The American Statistician*, 32, 100–102.