

에너지원별 냉방기기 표본설계

강 용 태[†], 이 덕 주, 김 의 경*, 전 호 철*

경희대학교 기계·산업 시스템 공학부, *에너지관리공단

Sample design of cooling systems for each energy source

Yong Tae Kang[†], Deok Joo Lee, Euy Kyung Kim*, Ho Cheol Jeon*

Department of Mechanical Engineering, Kyung Hee University, Gyeong-gi 449-701, Korea

*Korea Energy Management Corporation

ABSTRACT: The objectives of this study are to collect the population of each cooling system for gas and electric driven systems, and propose sample design for five cooling systems; ice storage systems, system air-conditioning system, turbo system as electric driven cooling systems, and absorption system and Gas driven Heat Pump (GHP) system as gas driven cooling systems. The sample design are carried out based on types of business, capacity, installation region and year. This study proposes criterion of the sample design for cooling systems for each energy source.

Key words: Population of cooling systems, Sample design, Gas and electric driven systems, Ice storage, System air-conditioning, Turbo, Absorption, GHP

기 호 설 명

p : 모비율
 α : 유의수준
 \hat{p} : 표본비율
 n : 표본수
 $z_{\alpha/2}$: 기각치
 e : 오차
 N : 모집단수

성 분석을 위한 실태조사 자료가 부족한 실정이다. 따라서 경제성장과 함께 지속적으로 증가하는 하계 냉방전력수요를 분담하고, 계절별 천연가스 수요평준화를 위한 천연가스 냉방기기의 보급촉진 정책을 수립하기 위하여, 에너지원별 냉방기기에 대한 실태조사를 수행하고 각 냉방기기에 대한 경제성을 비교분석할 필요가 있다. 이를 위하여 본 연구에서는 그 첫 단계로 각 냉방기기별 설치 및 운용실태조사를 위한 표본설계를 수행하고자 한다.

1. 서론

국가적 에너지이용의 합리화차원에서 천연가스 냉방의 중요성이 부각되고 있으나 보급촉진을 위한 문제점 도출 및 에너지원별 냉방기기의 경제

2. 표본설계의 기본방향

2.1 표본설계의 목적

통계조사 특히 대규모의 표본조사의 효율성과 신뢰수준을 높이기 위하여 우선 과학적이고 합리적인 표본추출이 이루어져야 하며 이에 따른 적절한 추정방법을 선택해야 한다.

[†] Corresponding author

Tel.: +82-31-201-2990 FAX: +82-31-202-3260

E-mail address: ytkang@khu.ac.kr

Table 1 표본설계 모집단 선정 기준^(1,2)

구분		세부 분류
에너지 원별	전기 구동식	빙축열시스템, 시스템에어컨, 터보냉동기
	가스 구동식	흡수식 시스템, GHP
업종별	가정용	단독, 다세대, 아파트
	산업용	제조업, 섬유/종이출판, 석유/화학 및 기타
	상업용	사무실, 보조 및 사회복지, 도/소매업 등
용량별	150RT급 이하	
	150RT급 이상 - 150RT급 이하	
	300RT급 이상	
지역별	수도권	
	중부권	
	남부권	

본 표본설계는 주어진 조사비용 하에서 최적의 표본크기를 산정, 조사의 효율성과 정도를 제고시키고, 에너지원별 각 냉방기기들의 보급 및 에너지소비현황 실태조사, 각 냉방기기들의 경제성 및 운용현황 조사 분석이 가능하도록 표본 설계하는데 목적이 있다.

2.2 표본설계의 기본 방향

2.2.1 모집단과 표본틀 선정

모집단이란 표본을 설계할 때 조사하고자 하는 대상인 기본단위의 전체집단을 의미한다. 그러나 조사대상을 모집단에서 직접 추출하기 어렵거나 비용이 많이 드는 경우가 있다. 이때 표본추출의 편의를 위해 추출단위를 설정하게 되며 이 추출단위들로 구성되어 실제로 표본추출이 이루어지는 리스트를 표본틀이라고 한다.

2.2.2 표본규모의 결정

표본설계에서 가장 중요한 문제 중 하나는 표본의 크기를 결정하는 것으로 표본의 크기가 필요 이상으로 크면 예산이 낭비가 되고, 반대로 너무 작으면 표본의 정도가 떨어지게 될 수 있다. 따라서 표본의 크기는 목표정도를 충족시킬 수 있는 범위 내에서 가능한 예산의 범위 내에서 최

적의 규모를 산정해야한다.

표본조사에서 목표정도란 추정 값의 허용오차를 통계이용가치의 관점에서 정하는 것으로 본 표본설계의 허용오차는 상대표준오차(변이계수, CV)를 이용한다. CV값은 시간적, 공간적으로 변동이 작으며, 거의 일정한 값을 가지는 특징이 있다. CV값이 일정한 값으로 안정되어 있다는 것은 허용오차가 주어지면 모수를 추정할 때 필요한 표본크기를 구하는데 도움이 된다.

표본오차에 중요한 요인이 되는 표본조사에서의 표본규모산정은 조사결과에 매우 중요한 역할을 한다. 본 표본설계에서 우선적으로 표본규모를 산정하게 되며, 일반적으로 조사결과와의 정도(허용오차 혹은 상대표준오차)와 조사비용을 고려하여 제안된 표본규모를 고려한다.

2.2.3 표본추출 및 추정방법 산정

일반적으로 사업체통계조사^(1,2)에서 이용되고 있는 절사법과 네이만 배분법을 이용하여 규모를 정한 후 냉방기기별 보급 및 에너지소비현황 등의 비중을 감안하여 부모집단에서 설치대수(용량)가 일정 기준 이상이면 전수층으로, 기준 미만인 사업체를 표본층으로 층화한 후 전수층은 전수조사하고 표본층에서 표본사업체를 설치대수(용량) 순으로 나열한 후 계통추출 한다.

3. 각 냉방기기별 모집단 정의

에너지원별 냉방기기 보급 및 소비현황 실태조사를 위한 표본설계를 위한 모집단 선정은 Table 1과 같은 기준으로 분류한다.

본 논문에서는 그 대표로 빙축열시스템에 대한 모집단 정의 및 표본설계의 결과를 제시한다.

3.1 빙축열 시스템의 업종별 설치대수 (Fig. 1)

3.2 빙축열 시스템의 용량별 설치대수 분류 (Table 2)

3.3 빙축열 시스템의 지역별 설치대수 분류 (Table 3)

3.4 빙축열 시스템의 연도별 설치대수 분류 (Table 4)

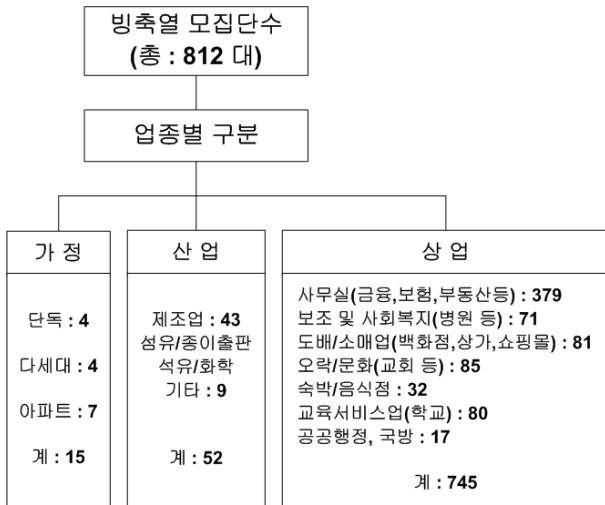


Fig. 1 빙축열 모집단 정의 기준^(1,2)

Table 2 빙축열 시스템의 용량별 설치대수^(1,2)

	가 정	산 업	상 업
150RT급 이하	10	23	508
150RT~300RT	5	9	136
300RT급 이상	-	20	100

Table 3 빙축열 시스템의 지역별 설치대수^(1,2)

구 분	가 정	산 업	상 업
수도권	12	16	393
중부권	-	29	241
남부권	2	6	113
계	14	51	747

4. 표본규모 선정방법

4.1 문제의 정의

모집단으로부터 추출되는 표본은 극단적으로 한 개의 표본추출 단위(sampling units)로 구성될 수도 있고 또는 모집단에서 한 개를 제외한 전체 표본추출 단위로 구성될 수도 있다. 이 때 어느 정도의 크기로 표본을 추출하는 것이 적절한가 하는 문제는 표본조사에 소요되는 최소의 비용으

Table 4 빙축열 시스템의 연도별 설치대수^(1,2)

연도	가 정	산 업	상 업	계
1990	-	-	7	7
1991	-	-	2	2
1992	-	-	5	5
1993	-	-	20	20
1994	-	-	29	29
1995	-	-	45	45
1996	1	-	42	43
1997	-	-	39	39
1998	-	-	46	46
1999	1	1	44	46
2000	1	3	47	51
2001	1	4	77	82
2002	-	6	102	108
2003	3	6	63	72
2004	3	6	79	88
2005	2	14	60	76
2006	1	10	42	53

로 원하는 수준의 정확성을 확보할 수 있는 표본의 크기가 얼마인가 하는 문제로 정리할 수 있다.

사실 적절한 표본의 크기를 정밀하게 규정한다는 것은 대단히 어려운 문제이며, 누구나 동의할 수 있는 특정 이론이 존재하고 있는 것도 아니기 때문에 실제로 표본조사를 행함에 있어서 주먹구구식으로 해결하고 있는 경우를 종종 발견할 수 있다. 표본의 크기를 결정하기 위한 기준으로 흔히 잘못 생각되고 있는 것에는 다음과 같은 것들이 있다.

첫째 모집단의 일정비율 이상을 표본으로 해야 한다는 생각이다. 이러한 오해 중 가장 많이 사용되고 있는 것이 1/10 법칙인데, 이것은 조사자가 통계적 분석을 행하고자 하는 모집단의 10%를 표본으로 선정하는 것이 바람직하다는 생각이다. 그러나 이러한 법칙은 모집단의 크기에 따라 많은 수정이 필요하다. 모집단이 대단히 큰 경우에는 10% 이하의 크기로도 충분히 원하는 수준의 정확성을 확보할 수도 있으며, 반대로 모집단의 크기가 대단히 적은 경우에는 표본의 크기를 10%이상 해야만 하는 경우도 있을 수 있는 것이다.

두 번째로는 표본의 크기는 항상 일정크기 이상이어야 한다는 생각으로, 예를 들어 표본은 2,000단위 혹은 5,000단위 이상은 되어야 한다는 생각이다. 그러나 표본은 항상 일정한 크기 이상

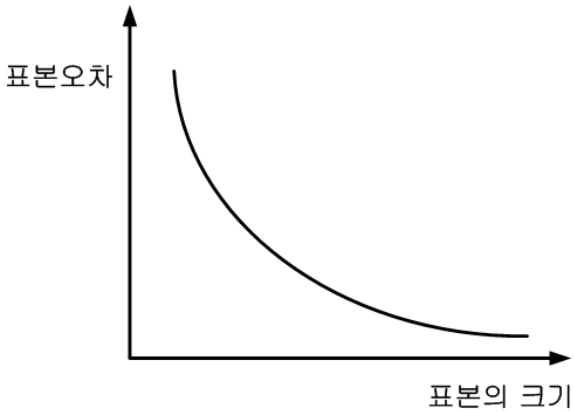


Fig. 2 표본크기와 오차의 관계

이어야 한다는 법칙은 없으며 표본크기가 가질 수 있는 범위는 매우 넓다고 하겠다. 따라서 이만한 조사를 하려면 적어도 5천개의 사례, 또는 2천개의 사례는 확보해야 한다고 하는 것은 잘못된 생각이며 통계이론을 잘 이해하지 못하는 소치라고 밖에 할 수 없을 것이다. 다만 표본크기가 30 이상이면 표본분포가 항상 정규분포를 따른다는 사실을 감안하면, 분석의 편의를 위해서 표본의 크기는 가능하다면 30 이상일 필요성은 있다고 볼 수 있다.

4.2 오차의 표본크기와의 관계

모비율의 추정의 경우, 만일 모비율 p 가 $(1-\alpha)100\%$ 신뢰구간의 중앙에 위치한다면 표본비율 \hat{p} 는 오차 없이 모비율을 추정하게 되나, 대부분의 경우에 \hat{p} 는 정확히 모비율을 추정하지 못한다. 즉, 일반적으로 표본조사를 통한 모수의 추정에는 오차가 수반되기 마련인데, 앞에서 논의된 신뢰구간을 이용하면 이러한 오차의 한계를 명확하게 제시해 줄 수 있다.

오차의 크기는 모비율과 표본비율의 차이의 절대값, 즉 $|p - \hat{p}|$ 로 표현할 수 있으며, 신뢰구간의 정의에 의하면 오차크기의 한계는 신뢰구간의 절반을 넘지 않게 된다. 따라서 모비율의 추정시 오차의 한계는 다음과 같이 나타낼 수 있다.

$$\begin{aligned} \text{오차한계} &= \pm z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \quad (\text{무한모집단}) \\ &= \pm z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n} \frac{N-n}{N-1}} \quad (\text{유한모집단}) \end{aligned}$$

Table 5 오차의 크기

표본수 \ 표본비율	150	300	600
5% 또는 95%	3.5%	2.5%	1.7%
10% 또는 90%	4.8%	3.4%	2.4%
20% 또는 80%	6.4%	4.5%	3.2%
30% 또는 70%	7.3%	5.2%	3.7%
40% 또는 60%	7.8%	5.5%	3.9%
50%	8.0%	5.7%	4.0%

위 식을 보면 표본오차의 크기는 표본크기 n 과 표본비율 \hat{p} 의 값의 크기에 관계가 있음을 알 수 있다. 우선 표본오차는 표본크기와 반비례의 관계임을 알 수 있는데, 정확히는 표본크기의 제곱근 (\sqrt{n})에 반비례한다. 이러한 관계를 Fig. 2에 나타내었다.

이는 표본크기가 클수록 오차는 감소한다는 사실을 의미한다. 따라서 같은 수준의 신뢰도를 유지하면서 추정의 정확도를 높이기(즉, 신뢰구간의 길이를 줄이기) 위해서는 표본의 크기를 늘리는 방법이 있을 수 있다. 구체적으로 오차를 반으로 줄이기 위해서는 표본의 크기를 4배 늘려야 하는 것이다.

한편 오차의 크기는 표본비율의 값과도 관계가 있는데, 위 식을 자세히 살펴보면 표본비율의 값이 0.5일 때 오차가 가장 크게 나타나며, 0과 1의 값에 가까울수록 오차가 작아짐을 알 수 있다.

위 관계를 예를 들어 살펴보도록 하자. 신뢰수준을 95%로 하면 $z_{\alpha/2} = 1.96$ 이 되며, 표본크기와 값의 몇 가지 조합에 대해서 오차의 크기를 계산해보면 Table 5와 같다.

계산 결과를 보면, 하나의 표본수에 대해서는 표본비율이 5%부터 시작해서 점차 커짐에 따라 오차의 크기가 증가하다 50%에 이르렀을 때 최대 값을 갖고 이 후로는 다시 감소하고 있음을 알 수 있다. 또한 같은 표본비율에 대해서는 표본의 크기가 커질수록 오차가 적어지며, 특히 표본수가 150에서 600으로 4배가 늘어나면 오차가 절반이 된다는 사실을 확인할 수 있다.

4.3 표본크기의 결정 원리

앞에서 살펴본 바와 같이 오차의 크기 즉, 신뢰구간의 폭은 표본수의 제곱근 \sqrt{n} 에 반비례한다. 따라서 오차의 크기를 줄이기 위해서는 표본수를 증가시키는 것이 좋으나, 반면에 표본수가 지나치게 커지게 되면 표본추출에 따르는 경비나 자료의 처리시간 등 표본조사와 관련된 비용이 커질 수밖에 없다는 문제점이 발생된다. 그러므로 일반적으로 통계적 표본조사를 실시하는 경우에는 조사의 목적이나 결과가 어떻게 사용될 것인가 등을 고려하여 모수 추정에 요구되는 오차의 한계(신뢰구간의 길이)를 미리 정하여 놓고, 이를 만족시키는 최소의 표본수 n 을 결정하는 것이 바람직한 것으로 인식되고 있다. 이에 모비율을 추정하는 경우에 표본수를 결정하는 방법에 대해서 살펴보도록 하자.

먼저 무한모집단의 경우 신뢰수준 $(1-\alpha)100\%$ 의 확신으로 오차한계가 이내가 되도록 하려면 다음과 같은 부등식이 성립되어야 하며,

$$z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq e \quad (1)$$

유한모집단의 경우에는 다음과 같은 부등식이 성립되어야 한다.

$$z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n} \frac{N-n}{N-1}} \leq e \quad (2)$$

위 두식을 표본수 n 에 대하여 풀면 다음과 같은 식을 얻게 된다.

$$z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n} \frac{N-n}{N-1}} \leq e \text{ (무한모집단)} \quad (3)$$

$$n \geq \frac{\hat{p}(1-\hat{p})}{\left[\frac{e}{z_{\alpha/2}}\right]^2 + \frac{\hat{p}(1-\hat{p})}{N}} \text{ (유한모집단)} \quad (4)$$

표본의 크기는 정수이므로 위 부등식의 우변항 값보다 큰 정수로서 가장 작은 값을 표본수로 선택하면 된다. 그러나 표본수는 실제로 표본조사를 실시하기 이전에 결정하여야 하는 변수인데, 표본

조사를 실시하기 이전에 표본비율 \hat{p} 를 알 수 없기 때문에 위 식을 그대로 사용할 수 없다. 따라서 모든 가능한 표본비율 값에 대해서 오차한계가 e 이내가 되도록 하는 표본수를 찾아야 하는데, 이는 $\hat{p}(1-\hat{p})$ 의 최대값이 0.25라는 사실을 이용하면 해결될 수 있다.

즉, 모든 표본비율 \hat{p} 에 대해서 오차한계가 e 이내가 되도록 하는 표본수는 다음과 같은 부등식을 만족하는 최소의 정수로 선정하면 된다.

$$n \geq \frac{1}{4} \left[\frac{z_{\alpha/2}}{e} \right]^2 \text{ (무한모집단)} \quad (5)$$

$$n \geq \frac{1}{4 \left[\frac{e}{z_{\alpha/2}} \right]^2 + \frac{1}{N}} \text{ (유한모집단)} \quad (6)$$

5. 결론

어떤 조사상황에서 표본의 크기가 만족할 만한 것이 되려면 분석자에 대한 상당히 높은 수준의 경험과 능력이 요구되게 된다. 따라서 표본의 크기를 결정할 때에는 수학적 공식에만 의존하기 보다는 다음과 같은 여러 가지 사항들을 충분히 고려해서 결정하는 것이 바람직하다고 볼 수 있다.

5.1 비용과 시간

비용과 시간만을 고려할 때는 표본의 크기를 크게 할수록 많은 비용이 소요되기 때문에 되도록이면 적은 규모의 표본이 유리하다. 그런데 우리가 충분한 예산과 시간을 갖고 있다면, 비록 표본크기가 증가할수록 오차가 감소하므로 되도록이면 큰 규모의 표본을 조사하는 것이 좋을 것이다. 그러나 현실적으로 볼 때 예산과 시간은 모든 요구를 충족시킬 만큼 그렇게 충분하지 않은 경우가 대부분이다. 따라서 전 조사과정에 대해서 충분하지 않은 예산과 시간을 적절히 배분할 필요가 있는 것이다. 다시 말해서 신뢰도와 정확도의 수준이 높으면 높을수록 이를 만족시키기 위해서는 매우 큰 규모의 표본과 비용이 소모되므로, 일정한 수준의 신뢰도와 정확도만 확보하게 하고 추가로 이들을 높이기 위하여 소요될 비용

은 조사의 설계, 자료의 처리, 분석 등에 사용함으로써 비 표본오차를 줄이도록 하는 것이 결국 조사 전체의 오차를 줄일 수 있는 바람직한 방법이라 할 수 있겠다.

5.2 모집단의 동질성

조사에서 필요로 하는 표본의 크기는 모집단이 어느 정도 동질적인가에 달려있다. 즉 모집단이 동질적이면 동질적일수록 적은 수의 표본으로도 조사에서 요구하는 신뢰도를 확보할 수 있는 것이다. 극단적인 경우에는 모든 구성원이 조사하려는 특성에 있어 똑같다고 한다면 표본의 크기는 하나로 충분할 것이다. 가령 어떤 사람의 피를 검사하려고 할 경우 그 사람의 피를 전부 다 추출해서 검사하는 것이 아니라 한 방울의 혈세포만 표본으로 선정하여 검사하면 충분한 것과 같은 의미이다.

그런데 모집단이 이질적인 경우라도 몇 개의 층 또는 부분으로 나누면 필요한 표본의 크기를 줄일 수 있다. 즉 각 층과 층간은 서로 이질적이라도 층을 하나의 하위모집단으로 하여 이의 요소가 동질적이라면 이 층은 적은 규모의 표본으로도 층을 잘 대표할 수 있게 되는 것이다.

그런데 문제는 이렇게 모집단을 층화할 수 있는 것으로 보이는 변수와 조사하려는 특성이 관련되어 있는가 하는 점이다. 가령 냉방기기에 대한 조사에서 설치면적을 기준으로 층화할 수 있으나 이때 설치면적이 냉방기기(빙축열 시스템, 터보형 등)에 대한 선호에 관련이 있는지를 확신하기가 어렵다는 것이다. 따라서 이때는 소규모의

예비조사를 실시하여 관련성을 파악하는 것이 좋다. 만약 관련성이 없는데도 이를 기준으로 층화하면 조사하고자 하는 변수로 볼 때는 층 내의 동질성이 전혀 확보되지 않게 되고 따라서 층화의 의미가 없게 되는 것이다.

5.3 분석변수 및 분석점주의 수

표본의 크기를 층화로 나누어서 결정하려면 분석되는 범주의 수를 고려해야 하는데, 범주의 수가 많아질수록 신뢰도를 높이는데 필요한 전체 표본크기는 증가한다. 즉 전체 표본크기가 일정하다고 할 경우 범주의 수가 많아지면 많아질수록 각 범주에 속하게 되는 표본수는 점점 줄어들게 되는 것이다. 이때 한 두 개의 표본이 그 범주 전체를 대표한다고 한다면 신뢰도에 의문을 제기하는 사람이 많아질 것이다. 이와 같이 자료가 세밀하게 분류될수록 모집단의 한 원소가 각 칸(cell)에 포함될 가능성은 점점 작아지고 따라서 적은 수의 표본을 가지고 계산한 추정치는 신뢰도가 떨어지게 되기 마련이다. 따라서 범주의 수가 많아질수록 그만큼 표본의 크기가 증대되어야만 신뢰도를 보장할 수 있게 된다.

참고문헌

1. 산업자원부 에너지관리공단, 2006, 국가에너지 종합분석보고서 (가정/상업부문)
2. 산업자원부 에너지관리공단, 2006, 산업부문 에너지소비량 표본조사 및 분석보고서 (광업/제조업부문)

