연구논문

반복조사에서 모집단 크기 변동을 반영한 표본수 결정*

Sample Size Determination in Repeated Surveys with Varying Population Sizes

김규성[a)]
Kyu－Seong Kim

본 연구에서는 모집단의 크기가 변하는 반복조사에서 모집단의 크기 변동을 반영하는 표본수 결정방법을 탐구하였다. 기존의 상대표준오차를 이용하는 방법은 모집단 크기 변동을 반영하지 않기 때문에 이에 대한 개선이 요구되어 그 대안 방법으로 모집단 크기 비례방법과 가중평균 방법을 소개하고 설계효과를 이용한 방법을 제안하였다. 가중평균 방법은 상대표준오차 방법과 모집단 크기 비례방법을 합성한 방법이고 설계효과를 이용한 방법은 상대표준오차를 이용한 방법에 모집단 크기 변동을 반영하도록 고안된 방법이다. 기존의 방법과 제안된 방법을 농산물소득조사 데이터에 적용한 후 표본수를 구하여 제시하였다. 이를 통하여 설계효과를 이용한 방법을 사용하면 모집단의 크기가 줄어드는 경우에 표본수의 일부를 줄일 수 있음을 보였다.

주제어 : 농산물소득조사, 상대표준오차, 설계효과

I investigate sample size determination methods in repeated surveys with varying population sizes. The existing method using the coefficient of variation does not consider the population size, so modifications are needed. As alternatives to the existing method I introduce the method proportional to the population size and the weighted average method, then propose the method using the design effect, which are all devised to use varying population sizes. In addition, I apply these methods to the agricultural income data in order to get the expected sample

* 이 논문은 2012년도 농촌진흥청 공동연구사업(과제번호: PJ008714022012)의 지원에 의해 이루어진 것임.

a) 서울시립대학교 통계학과 교수 김규성.
E－mail: kskim@uos.ac.kr

sizes in the next survey. Through the numerical example I find that the method of using the design effect needs less sample size in the next survey when the population size is decreasing.

Key words: agricultural income survey, coefficient of variation, design effect.

Ⅰ. 문제제기

반복조사에서 표본재설계는 총조사 자료가 제공되는 해에 표본개편의 일환으로 이루어지는 것이 보통이다. 표본재설계에서는 현행 데이터의 오차분석을 통하여 현행 조사결과의 신뢰도를 평가하고 이를 바탕으로 다음 조사의 표본수를 결정하며 결정된 표본수만큼 총조사 자료에서 새로운 표본을 선정하게 된다.

표본수의 결정에 영향을 미치는 요인은 조사비용, 목표오차, 그리고 추정오차 등으로 비용의 범위가 사전에 정해지는 경우는 목표오차와 추정오차가 주요 요인이 된다. 반복조사에서는 현행 데이터 분석을 통하여 주요 모수의 추정오차를 계산할 수 있으므로 표본수를 결정에 이를 반영할 수 있다. 주요 항목별 혹은 주요 변수별로 모수추정량의 추정오차를 계산한 후에 추정오차가 크게 나오면 목표오차를 조금 낮게 잡아 표본수를 늘리는 방향으로 표본수를 결정하고, 반대로 추정오차가 작게 나오면 목표오차를 조금 크게 잡아 표본수를 감소시키는 방향으로 표본수를 결정하게 된다. 특히 다목적 조사(multi－purpose survey) 혹은 다항목 조사(multivariate survey)에서는 주요 변수의 오차를 관리하는 것이 중요하므로 표본수 조정을 통하여 주요 변수의 오차를 관리하는 것은 매우 좋은 조사관리 방안이 된다.

표본조사에서 비용을 고려하지 않을 때 표본수 결정방법으로 널리 쓰이는 방법은 허용오차를 정하고 이에 도달하는 표본수를 찾는 방법이다(Cochran 1977; Lohr 1999; Som 1996; Statistics Canada 2003). 허용오차와 표본수의 관계식을 만든 후 표본수를 허용오차의 함수로 만들어 주어진 허용오차에 대응하는 표본수를 찾는 것이다. 최초 조사에서는 허용오차가 주어지면 모집단 특성치를 적절히 추정하거나 대입하여 표본수를 산정할 수 있다. 반면 반복조사에서는 현행 조사결과가 있으므로 현행 표본수와

상대표준오차 값을 다음번 조사의 표본수 결정에 활용할 수 있다. 이러한 방법 중 널리 쓰이는 방법의 하나가 아래의 공식을 이용하는 것이다(박홍래 1989). 현행 조사로부터 상대표준오차를 계산하고 이를 사전에 정한 목표 상대표준오차(목표 CV)로 나누어 그 값의 제곱을 현행 표본수에 곱하여 기대 표본수 값을 얻는다.

$$(\text{기대 표본수}) = (\text{현행 표본수}) \times \left(\frac{\text{현행 } \widehat{CV}}{\text{목표 } CV} \right)^2 \tag{1.1}$$

여기에서 CV는 변동계수(coefficient of variation)의 약자로 상대표준오차를 뜻하며 (현행 $\widehat{CV}$)은 현행 표본을 이용하여 추정한 변동계수 값이다. 반복조사에서 총조사 자료를 기초로 표본재설계를 하게 되면 추출틀, 층화, 집락화 등이 기존의 것과 달라지므로 재설계에 기초하여 추정오차를 직접적으로 계산하기는 쉽지 않다. 대신 위의 식 (1.1)은 이에 대한 간편하면서도 유용한 대안이 된다. 이때 위의 식 (1.1)의 사용이 정당성을 갖는 경우는 표본수가 동일한 경우에 현행 설계의 상대표준오차와 재설계의 상대표준오차가 근사적으로 같다는 가정이 성립하는 경우이다(박홍래 1989). 만일 반복조사에서 급격한 모집단 변동이나 조사 변수의 급격한 상황 변동이 없다면 이러한 가정은 대체로 만족하므로 위의 식 (1.1)을 반복조사 표본재설계의 표본수 결정 과정에 활용하는 것은 최선책은 아닐 수 있으나 차선책은 된다고 할 수 있다.

이제 조금 구체적인 상황을 고려해 보자. 식 (1.1)에 의하면 목표오차와 현행 추정오차가 동일할 때 기대 표본수는 현행 표본수와 동일하게 계산된다. 그런데 만일 모집단의 크기가 지난번 조사 대비 20% 감소했다고 하자. 표본의 대표성 측면에서 보면 표본수를 20% 줄여야 포함확률이 동일하게 되기 때문에 표본수 감소 문제를 제기할 수 있다. 이러한 측면에서 보면 상대표준오차를 이용하는 식 (1.1)은 모집단 크기의 변동을 반영하지 못하는 약점을 보인다. 모집단의 크기가 표본수에 비해 상대적으로 클 때에는 모집단 크기의 변동이 상대표준오차에 큰 영향을 주지 않기 때문에 식 (1.1)을 사용하는 것이 크게 문제되지 않는다. 반면, 모집단이 크지 않고 표본추출률이 상대적으로 작지 않을 때에는 모집단 크기 및 표본추출률이 상대표준오차에 영향을 주기 때문에 식 (1.1)을 직접 사용하는 것은 그리 효과적이지 않을 수 있다.

본 연구에서 탐구하고자 하는 것은 반복조사 표본재설계에서 모집단 크기 변동을 반영하는 표본수 결정방법이다. 식 (1.1)을 개선하여 모집단 크기 변동까지 반영하는

표본수 결정방법을 찾는 것이 본 연구의 목적이다. 본 논문은 다음과 같이 구성되어 있다. 제2절에서는 반복조사에서 표본수 결정방법으로 모집단 크기 비례방법, 가중평균방법, 그리고 설계효과를 이용한 방법을 소개한다. 제3절에서는 농산물소득조사에 제안된 방법을 적용하여 현실적인 적용 가능성을 살펴본다. 마지막으로 제4절에서는 연구 내용을 요약한다.

Ⅱ. 모집단 크기를 반영하는 표본수 결정방법

반복조사에서 모평균 추정에 필요한 표본수를 결정하는 문제를 고려하자. 조사시점 k에서 n_k는 표본수, N_k는 모집단 크기, $\overline{y}_k$는 모평균 추정량이라 하고 추정량 $\overline{y}_k$의 상대표준오차를 $CV_k = CV(\overline{y}_k)$로 표현하자. k시점까지 조사가 이루어졌고 $(k+1)$시점 조사를 위한 표본수를 결정해야 한다고 하자. 그리고 이용 가능한 자료는 k시점의 표본자료와 모집단 크기 N_k, 그리고 $(k+1)$시점의 모집단 크기 N_{k+1}이라고 하자.

1. 상대표준오차를 이용한 방법

제1절에 소개한 식 (1.1)을 반복조사에 적용하여 표현하면 아래와 같이 된다.

$$n_{k+1} = n_k \left(\frac{\widehat{CV}_k}{CV_{k+1}} \right)^2 \tag{2.1}$$

여기에서 $\widehat{CV}_k$는 k시점의 상대표준오차 추정값이고 CV_{k+1}은 $(k+1)$시점의 목표 상대표준오차이다. k시점에서 n_k와 상대표준오차 추정치 $\widehat{CV}_k$를 계산할 수 있으므로 목표 상대표준오차 CV_{k+1}가 결정되면 이에 필요한 표본수 n_{k+1}을 계산할 수 있다. 이 방법이 타당성을 갖는 경우는 모집단의 크기가 충분히 커서 비복원 표본추출의 효과가 크지 않고 두 시점에서 모집단 평균과 표준편차가 비슷한 경우이다. 만일 모집단의 크기가 충분히 크지 않고 두 시점에서 모집단 크기에 변동이 생기면 위 공식은 사용하기 어렵다.

2. 모집단 크기 비례방법

모집단 크기에 비례하여 표본수를 정하는 것도 하나의 방법이 된다. 즉,

$$n_{k+1} = n_k \left(\frac{N_{k+1}}{N_k} \right) = n_k \lambda_k \tag{2.2}$$

여기에서 $\lambda_k = N_{k+1}/N_k$는 모집단 비이다. 만일 모집단 크기가 10% 감소하면 $\lambda = 0.9$가 된다. 이 방법이 타당성을 갖는 경우는 두 시점의 모집단 표준편차가 같은 경우이다. 즉, 시점 k의 모집단 표준편차를 S_k라고 하면 $S_k \approx S_{k+1}$인 경우에 위 방법은 타당성이 있다.

3. 가중평균 방법

상대표준오차와 모집단 크기 변동을 동시에 이용하는 직관적인 표본수 결정방법으로 두 방법에서 구한 표본수의 가중평균을 고려할 수 있다.

$$n_{k+1} = w \times n_k \left(\frac{\widehat{CV}_k}{CV_{k+1}} \right)^2 + (1-w) \times n_k \left(\frac{N_{k+1}}{N_k} \right) \tag{2.3}$$

여기에서 w는 두 방법을 연결하는 가중치로서 $0 \leqq w \leqq 1$의 값을 갖는다. 만일 $w = 1$이면 위의 식은 식 (2.2)와 같아지고 반대로 $w = 0$이면 식 (2.1)과 같아진다. 이 방법이 타당성을 갖기 위해서는 상대표준오차를 이용한 방법과 모집단 크기 비례방법이 동시에 타당해야 한다. 즉, $\overline{Y}_k \approx \overline{Y}_{k+1}$, $S_k \approx S_{k+1}$인 경우에 위 방법은 타당성이 있다.

위 방법이 타당하다고 하더라도 실제 활용되기 위해서는 가중값 w를 결정해야 하는 문제가 남아 있다. 통상적으로 합성추정량(composite estimator, 예를 들면 Cochran 1997, 354쪽)에서는 추정량의 분산을 최소로 하는 값을 w의 값으로 택한다. 그런데 식 (2.3)에서는 목표오차 CV_{k+1}, 모집단 크기 N_k, N_{k+1}가 주어진 상수이므로 합성추정량에서와 같은 가중값 결정방법은 적용되지 않는다. 그런데 시각을 바꾸어 보면 가중값 w는 모집단 변동 N_{k+1}/N_k과 상대표준오차 크기의 비 $(\widehat{CV}_k / CV_{k+1})^2$의 중요

도를 나타내는 수치임을 알 수 있다. 따라서 다소 주관적이기는 하지만 모집단 변동과 오차의 크기의 중요도를 고려하여 가중값 w를 정하는 것도 하나의 현실적인 대안이 될 수 있다.

4. 설계효과를 이용한 방법

이 소절에서는 모집단 크기 변동을 반영하는 표본수 결정방법으로 설계효과를 이용하는 방법을 소개한다. 2.3절에서 소개한 가중 평균 방법은 직관적으로 이해하기는 쉽지만 가중값 w을 구하는 문제가 명시적으로 해결되지 않았기 때문에 주관적인 판단에 의하여 표본수를 결정해야 하는 단점이 있다. 반면 이번 소절에서 소개할 설계효과를 이용한 표본수 결정방법은 유일한 표본수를 제공하는 방법이다.

반복조사에서 복합표본설계(complex sample design)를 이용하여 모평균을 추정할 때 시점 k에서 모평균 추정량 $\overline{y}_k$의 상대표준오차 CV_k와 설계효과 간에는 다음과 같은 관계가 있다.

$$CV_k^2 = \frac{S_k^2}{\overline{Y}_k^2}\left(\frac{1}{n_k} - \frac{1}{N_k}\right) \times DEFF_k \tag{2.4}$$

여기에서 $\overline{Y}_k$은 k시점의 모평균, S_k은 모집단 표준편차 그리고 $DEFF_k$은 설계효과이다.

이제 $(k+1)$시점 조사에 대하여 모평균 추정량의 목표 상대표준오차를 CV_{k+1}로 정하고 이 목표오차에 도달하는 표본수 n_{k+1}을 정하려 한다고 하자. 만일 두 시점의 모평균, 모 표준편차 그리고 설계효과가 유사하다고 가정하면 식 (2.4)로부터 다음을 얻을 수 있다.

$$CV_{k+1}^2 \approx \frac{S_k^2}{\overline{Y}_k^2}\left(\frac{1}{n_{k+1}} - \frac{1}{N_{k+1}}\right) \times DEFF_k \tag{2.5}$$

그리고 식 (2.4)를 식 (2.5)에 대입하여 정리한 후 CV_k에 추정값 $\widehat{CV}_k$을 대입하면 다음의 관계를 얻는다.

$$n_{k+1} \approx \frac{1}{\left(\frac{CV_{k+1}}{\widehat{CV}_k}\right)^2 \left(\frac{1}{n_k} - \frac{1}{N_k}\right) + \frac{1}{N_{k+1}}} \tag{2.6}$$

여기에서 CV_{k+1}은 $(k+1)$ 시점의 목표 상대표준오차이다. 따라서 목표 상대표준오차 CV_{k+1}에 도달하는 표본수는 위의 식을 만족시키는 자연수 n_{k+1}이 된다.

목표 상대표준오차 CV_{k+1}와 표본수 n_{k+1}, 그리고 모집단 크기 변동의 관계를 좀 더 이해하기 쉬운 형태로 바꾸어 살펴보자. 표현의 편리를 위하여 $f_k = n_k/N_k$, $\Delta_k = f_k\left(1 - \left(\widehat{CV}_k/CV_{k+1}\right)^2/\lambda_k\right)$이라고 하고 이를 식 (2.6)에 대입한 후 정리하면 아래의 식을 얻을 수 있다.

$$\begin{aligned} n_{k+1} &\approx \frac{1}{\frac{1}{n_k}\left(\frac{CV_{k+1}}{\widehat{CV}_k}\right)^2 (1-\Delta_k)} \\ &= n_k\left(\frac{\widehat{CV}_k}{CV_{k+1}}\right)^2 (1+\Delta_k+\Delta_k^2+\ldots) \\ &\approx n_k\left(\frac{\widehat{CV}_k}{CV_{k+1}}\right)^2 \left(1+f_k\left(1-\frac{1}{\lambda_k}\left(\frac{\widehat{CV}_k}{CV_{k+1}}\right)^2\right)\right) \end{aligned} \tag{2.7}$$

위 식 (2.7)의 간단한 경우를 살펴보자. 만일 목표 상대표준오차가 현행 상대표준오차와 동일하다고 가정하면, 즉 $\widehat{CV}_k = CV_{k+1}$이면 식 (2.7)은 다음과 같이 된다.

$$n_{k+1} \approx n_k\left(1 - f_k\left(\frac{N_k - N_{k+1}}{N_{k+1}}\right)\right) \tag{2.8}$$

즉, 모집단의 크기가 감소하면 상대변동비율 $(N_k - N_{k+1})/N_{k+1}$에 표본추출비율 f_k이 곱해진 비율만큼 표본수가 줄어든다.

〈표 1〉 **표본수 상대 비** (n_{k+1}/n_k)

$\dfrac{\widehat{CV}_k}{CV_{k+1}}$	f_k	λ_k					
		0.5	0.6	0.7	0.8	0.9	1.0
0.8	1/50	0.6364	0.6391	0.6411	0.6425	0.6437	0.6446
	1/100	0.6382	0.6395	0.6405	0.6412	0.6418	0.6423
	1/500	0.6396	0.6399	0.6401	0.6402	0.6403	0.6404
	1/1000	0.6398	0.6399	0.6400	0.6401	0.6401	0.6402
	1/3000	0.6399	0.6399	0.6400	0.6400	0.6400	0.6400
0.9	1/50	0.7999	0.8043	0.8074	0.8098	0.8116	0.8130
	1/100	0.8049	0.8071	0.8087	0.8099	0.8108	0.8115
	1/500	0.8090	0.8094	0.8097	0.8099	0.8101	0.8103
	1/1000	0.8095	0.8097	0.8098	0.8099	0.8100	0.8101
	1/3000	0.8098	0.8099	0.8100	0.8100	0.8100	0.8100
1.0	1/50	0.9800	0.9866	0.9914	0.9950	0.9977	1.0000
	1/100	0.9900	0.9933	0.9957	0.9975	0.9988	1.0000
	1/500	0.9980	0.9986	0.9991	0.9995	0.9997	1.0000
	1/1000	0.9990	0.9993	0.9995	0.9997	0.9998	1.0000
	1/3000	0.9996	0.9997	0.9998	0.9999	0.9999	1.0000
1.1	1/50	1.1756	1.1854	1.1923	1.1976	1.2016	1.2049
	1/100	1.1928	1.1977	1.2011	1.2038	1.2058	1.2074
	1/500	1.2065	1.2075	1.2082	1.2087	1.2091	1.2094
	1/1000	1.2082	1.2087	1.2091	1.2091	1.2095	1.2097
	1/3000	1.2094	1.2095	1.2097	1.2097	1.2098	1.2099
1.2	1/50	1.3858	1.3996	1.4095	1.4169	1.4227	1.4273
	1/100	1.4129	1.4198	1.4247	1.4284	1.4313	1.4336
	1/500	1.4345	1.4359	1.4369	1.4377	1.4382	1.4387
	1/1000	1.4372	1.4379	1.4384	1.4388	1.4391	1.4393
	1/3000	1.4391	1.4393	1.4394	1.4390	1.4397	1.4397

<표 1>에는 상대표준오차의 비 $\widehat{CV}_k/CV_{k+1}$, 표본추출률 f_k, 모집단 변동 비율 λ_k에 따른 표본수 비 n_{k+1}/n_k를 식 (2.8)에 의하여 계산한 근사값이 나타나 있다. <표 1>은 다음의 결과를 수치적으로 보여준다. 첫째, 표본수의 결정에 영향을 주는 세 요인 중 영향력이 가장 큰 것은 상대표준오차의 비, $\widehat{CV}_k/CV_{k+1}$이다. 잘 알려진 바와 같이 현행오차보다 목표오차를 작게 정하면 더 많은 표본수를 필요로 한다. 둘째, 모집단 크기가 감소하면($\lambda_k < 1$) 표본수를 줄여도 되며, 감소비율이 클수록 기대 표본수는 더 적어진다. <표 1>에서 모집단 크기 감소가 기대 표본수에 미치는 영향이 가장 큰 경우는 목표오차가 현행오차 보다 작고($\widehat{CV}_k/CV_{k+1} = 1.2$) 모집단 크기가 반으로 줄어든 경우($\lambda_k = 0.5$)이다. 이때 기대 표본수는 모집단 크기가 변하지 않은 경우에 비하여 $4.15\%\,((1.4273 - 1.3858) \times 100)$ 줄여도 된다. 셋째, 표본추출율 f_k의 효과는 상대표준오차의 비와 모집단 변동률에 따라 다르게 나타난다. <표 1>의 범위에서만 보면 목표오차가 현행오차보다 작을 때($\widehat{CV}_k/CV_{k+1} > 1$)에는 표본추출률 f_k이 클수록 기대 표본수는 적어진다. 반면 목표오차가 현행오차보다 큰 경우에는 모집단 크기가 감소하고 표본추출율이 클수록 기대 표본수는 적어진다.

본 연구의 관심사 중의 하나는 모집단 크기가 줄어들었을 때 기대 표본의 크기를 얼마나 줄일 수 있는가 하는 점인데 <표 1>은 대략적인 가이드라인을 제공해 준다. 즉, 목표오차가 현행오차보다 작아서 표본수를 늘려야 하는 경우에 모집단 크기가 작아지고 표본추출률이 크면 표본수는 어느 정도 줄여도 된다. 그러나 그 비율은 높지 않다.

Ⅲ. 농산물소득조사에 적용

1. 개요

농산물소득조사는 주요 농작물의 수입, 경영비, 소득을 조사하여 농업의 주요 지표 자료를 제공하는 조사로서, 2010년 조사에서는 전국작목 58작목과 지역작목 59작목을 조사하여 결과를 발표하였다(농촌진흥청 2011). 농산물소득조사는 해마다 조사하는 반복조사로서 농림어업총조사가 실시된 다음해에 총조사 자료를 기반으로 표본 개편을

주기적으로 실시하고 있다. 표본 개편에서는 현행 조사자료를 분석하고 이를 바탕으로 현행 조사를 평가한 후 미비점을 보완하며, 총조사 자료를 기초로 표본을 변경하게 된다. 여기에서는 표본 개편 과정 중 모집단 크기의 변화와 이에 따른 표본수 결정 문제만을 구체적으로 살펴본다.

아래의 <표 2>에는 농림어업총조사에서 조사된 작목 중 일부 작목의 농가수 및 재배면적이 나타나 있다(통계청 2012). 예를 들어 당근의 재배면적은 2005년 총조사에서는 286만ha로 조사되었고 2010년 총조사에서는 188만ha로 조사되어 2010년의 재배면적은 2005년 재배면적의 65.7%에 불과한 것으로 나타났다. 노지수박은 68.5%, 대파는 87.9%로 조사되었다. 재배농가수를 비교하면 당근의 경우, 2005년에 5,670호, 2010년에 5,235호가 재배한다고 조사되어 5년 새 435호가 줄어든 것으로 조사되었고 그 비율은 7.7%에 해당된다. <표 2>에 나타난 작목 중 재배 농가 수 비율이 가장 크게 줄어든 작목은 대파이다. 2010년의 대파 재배농가수는 2005년의 재배농가수의 26.5%에 불과하다.

<표 2> 작목별 총조사의 농가수 및 재배면적(작목 중 일부)

작목명		농가수(호)			면적(천 ha)		
		2005년	2010년	농가수 비	2005년	2010년	면적 비
식량작물	노지 풋옥수수	113,874	135,554	1.190	10,045.8	12,861.3	1.280
	고구마	298,439	306,303	1.026	18,779.0	24,443.9	1.302
노지채소	노지수박	5,852	4,989	0.852	3,419.1	2,342.4	0.685
	양배추	6,711	7,255	1.081	2,125.6	3,954.7	0.959
	대파	166,508	44,083	0.265	8,600.1	7,561.7	0.879
	당근	5,670	5,235	0.923	2,855.4	1,876.1	0.657
	노지시금치	24,622	30,501	1.239	3,227.3	3,729.4	1.156

* 농가수 비 = (2010년 농가수)/(2005년 농가수)
* 면적 비 = (2010년 면적)/(2005년 면적)

2. 표본수 결정

농산물소득조사의 전국 조사 작목 중에서 총조사 작목 분류와 일치하는 22개 작목을 대상으로 제2절에서 고찰한 표본수 결정방법을 적용하여 기대 표본수를 구하였다. 22개 작목에 대한 2010년 표본수, 상대표준오차 추정값, 표본추출률의 역수 $1/f$, 모집단 변동 비율 λ가 〈표 3〉에 제시되어 있다. 여기에서 상대표준오차는 작목별로 수입, 경영비, 소득상대표준오차를 구한 후 이를 평균한 값이다. 본 소절에서 세 변수의 상대표준오차의 평균값을 사용하여 표본수를 결정하였는데, 이 방법 이외에 세 변수별로 목표오차에 따른 표본수를 구하고 이렇게 구한 표본수의 최대값을 최종 표본수로 결정하는 방법이나 혹은 변수별로 목표오차를 정하고 이를 가중평균으로 결합하여 목표오차 가중평균을 만족하는 표본수를 구하는 방법 등을 표본수 결정방법으로 생각할 수 있다. 그런데 전자의 경우는 필요 표본수가 너무 커지는 단점이 있고, 후자의 경우는 세 변수의 목표오차의 결합을 주관적으로 결정해야 하는 문제가 남아 있다. 본 연구에서 채택한 방법은 후자의 방법 중에서 세 변수의 중요도를 1/3씩 부여한 경우로 볼 수 있고, 이 방법은 작목별 수입, 소득, 경영비의 세 변수를 동일한 중요도로 분석하려고 하는 경우에는 타당성을 부여받을 수 있다.

농산물소득조사의 22개 작목에 대하여 다음번 조사에서 사전에 정한 목표오차에 이를 것으로 기대되는 기대 표본수를 작목별로 표본수 결정방법에 따라 계산하였다. 표본수 결정방법은 앞 절에서 소개한 가중평균방법과 설계효과를 이용한 방법을 이용하였고 가중평균 방법에서 가중치는 $w=0,\ 0.6,\ 0.8,\ 1$을 고려하였으며 목표오차는 $4\%,\ldots,8\%$를 고려하였다. 가중평균방법 중에서 $w=0$인 경우는 모집단 크기에 비례한 방법이고 $w=1$인 경우는 상대표준오차를 이용한 방법이다. 〈표 4〉에 가중평균방법과 설계효과를 이용한 방법으로 구한 기대 표본수가 제시되어 있다. 여기에서 CV는 상대표준오차를 뜻한다.

〈표 3〉 작목별 표본수 및 평균 상대표준오차

작목분류	작목명	표본수	상대표준오차	$1/f$	λ
식량작물	노지풋옥수수	87	5.114	1,558.1	1.280
	고구마	106	6.325	2,889.7	1.302
노지채소	노지수박	74	6.616	43.4	0.685
	양배추	66	8.715	109.9	0.959
	대파	72	9.195	612.3	0.879
	당근	39	15.217	134,2	0.657
	노지시금치	33	10.990	924.3	1.156
시설채소	시설참외	108	5.193	71.4	0.877
	시설무	34	5.429	136.8	1.142
	시설배추	33	9.915	276.9	1.298
	시설시금치	33	9.704	230.2	1.207
	시설상추	87	6.754	123.1	1.138
노지과수	사과	143	3.878	283.8	1.071
	배	183	3.356	135.1	0.806
	복숭아	137	4.889	201.4	0.845
	노지포도	174	3.693	193.5	0.946
	노지감귤	45	6.186	485.2	0.959
	단감	106	4.274	308.0	0.893
시설과수	시설포도	49	9.382	98.6	1.214
	시설감귤	24	6.891	217.7	1.357
특약용	인삼	84	4.783	134.9	1.204
	참깨	48	12.850	2,690.0	0.493

〈표 4〉 표본수 결정방법에 따른 작목별 기대 표본수

작목명	현행 표본수	현행 cv	가중평균 방법										
			w=0	목표 cv(w=0.6)					목표 cv(w=0.8)				
				4%	5%	6%	7%	8%	4%	5%	6%	7%	8%
노지 풋옥수수	87	5.114	111	130	99	82	72	66	136	95	73	59	51
고구마	106	6.325	138	214	157	126	107	95	240	163	122	97	81
노지수박	74	6.616	51	142	98	74	60	51	172	114	82	63	51
양배추	66	8.715	63	213	146	109	87	72	263	173	124	95	75
대파	72	9.195	63	254	171	127	100	82	317	207	148	112	89
당근	39	15.217	26	349	227	161	121	95	457	294	206	153	118
노지 시금치	33	10.99	38	165	111	82	64	53	207	135	96	73	57
시설참외	108	5.193	95	147	108	86	74	65	165	112	84	66	55
시설무	34	5.429	39	53	40	32	28	25	58	40	30	24	20
시설배추	33	9.915	43	139	95	71	57	48	171	112	81	62	49
시설 시금치	33	9.704	40	132	91	68	54	45	163	107	77	59	47
시설상추	87	6.754	99	188	135	106	88	77	218	147	108	85	69
사과	143	3.878	153	142	113	97	88	81	138	99	78	66	58
배	183	3.356	147	136	108	93	84	78	133	95	75	63	55
복숭아	137	4.889	116	169	125	101	86	77	187	128	96	77	64
노지포도	174	3.693	165	155	123	105	95	88	152	109	86	72	63
노지감귤	45	6.186	43	82	59	46	38	33	95	64	47	37	30
단감	106	4.274	95	110	84	70	62	56	116	81	62	51	43
시설포도	49	9.382	59	186	127	96	77	64	228	150	108	82	66
시설감귤	24	6.891	33	56	40	32	27	24	63	43	32	25	21
인삼	84	4.783	101	113	87	72	64	58	116	82	63	52	44
참깨	48	12.85	24	307	200	142	107	84	401	258	181	134	104

(계속)

〈표 4〉 **표본수 결정방법에 따른 작목별 기대 표본수** (계속)

작목명	현행 표본수	현행 cv	가중평균 방법					설계효과를 이용한 방법				
			목표 $cv(w=1)$					목표 cv				
			4%	5%	6%	7%	8%	4%	5%	6%	7%	8%
노지 풋옥수수	87	5.114	142	91	63	46	36	142	91	63	46	36
고구마	106	6.325	265	170	118	87	66	265	170	118	87	66
노지수박	74	6.616	202	130	90	66	51	188	125	88	66	51
양배추	66	8.715	313	201	139	102	78	302	197	138	102	78
대파	72	9.195	380	243	169	124	95	377	242	169	124	95
당근	39	15.217	564	361	251	184	141	476	326	234	176	136
노지 시금치	33	10.990	249	159	111	81	62	248	159	110	81	62
시설참외	108	5.193	182	116	81	59	46	180	116	81	60	46
시설무	34	5.429	63	40	28	20	16	62	40	28	21	16
시설배추	33	9.915	203	130	90	66	51	200	129	90	66	51
시설 시금치	33	9.704	194	124	86	63	49	191	123	86	63	49
시설상추	87	6.754	248	159	110	81	62	245	158	110	81	62
사과	143	3.878	134	86	60	44	34	134	86	60	44	34
배	183	3.356	129	82	57	42	32	129	83	58	42	32
복숭아	137	4.889	205	131	91	67	51	204	131	91	67	51
노지포도	174	3.693	148	95	66	48	37	148	95	66	49	37
노지감귤	45	6.186	108	69	48	35	27	107	69	48	35	27
단감	106	4.274	121	77	54	40	30	121	77	54	40	30
시설포도	49	9.382	270	173	120	88	67	260	169	119	88	67
시설감귤	24	6.891	71	46	32	23	18	71	46	32	23	18
인삼	84	4.783	120	77	53	39	30	120	77	54	39	30
참깨	48	12.85	495	317	220	162	124	492	316	219	161	124

가중평균방법 중 $w=0$인 경우는 모집단 크기에 비례하여 구한 기대 표본수와 동일한 값이므로 노지풋옥수수나 고구마와 같이 2010년 총조사 크기가 2005년에 비하여 증가한 작목은 기대 표본수가 현행 표본수보다 크게 계산되었다. 이 방법은 앞에서 언급한 바와 같이 추정오차를 고려하지 않은 방법이므로 현실적으로 채택되기 힘든 방법이다. 가중평균방법 중 $w=1$인 경우의 표본수는 식 (2.1)의 상대표준오차를 이용하여 구한 표본수와 동일하다.

설계효과를 이용한 방법으로 계산한 표본수는 가중평균방법 중 $w=1$인 경우에 모집단 크기 변동을 반영하여 구한 표본수이다. 따라서 사과($\lambda=1.071$), 노지시금치($\lambda=1.156$) 등과 같이 모집단 변동이 크지 않는 작목은 설계효과를 이용한 방법의 기대 표본수나 가중평균 중 $w=1$인 표본수와 거의 비슷하다. 그러나 노지수박($\lambda=0.685$), 당근($\lambda=0.657$) 등과 같이 모집단의 크기가 줄어든 작목에서는 설계효과를 이용한 방법에서 제시하는 표본수가 가중평균 방법 중 $w=1$인 경우의 표본수보다 다소 작다. 노지수박의 경우 목표오차가 5%일 때 가중평균 방법($w=1$)에서 제시하는 표본수는 130 농가인 반면 설계효과를 이용한 방법에서 제시하는 표본수는 125 농가로서 5농가가 적다. 참깨인 경우 $\lambda=0.493$로서 모집단 크기가 대폭 줄어든 경우인데 〈표 4〉를 보면 가중평균 방법($w=1$)이나 설계효과를 이용한 방법의 표본수 크기가 크게 다르지 않다. 그 이유는 매우 작은 표본추출률 $f=1/2{,}690$ 때문이다. 표본추출률이 매우 작을 때에는 〈표 1〉에서도 나타났듯이 모집단의 크기의 변동이 표본수 크기 결정에 큰 영향을 미치지 못한다.

Ⅳ. 결론

본 연구에서는 모집단의 크기가 변하는 반복조사에서 모집단 크기의 변동을 반영하는 표본수 크기의 결정법을 고찰하였다. 기존의 방법인 상대표준오차를 이용하는 방법에 더하여 모집단 크기에 비례하는 방법과 가중평균방법을 소개하고 설계효과를 이용하는 방법을 제안하였다. 가중평균방법은 상대표준오차를 이용한 방법과 모집단 크기 비례방법을 합성한 방법이고, 설계효과를 이용한 방법은 상대표준오차를 이용한 방법을 개선하여 모집단 크기 변동을 반영할 수 있도록 고안된 방법이다. 기존의 방법인 상대표준오차를 이용하는 방법은 모집단 크기 변동을 반영하지 않는 방법인 반면에

모집단 크기 비례방법, 가중평균방법 그리고 설계효과를 이용한 방법은 모집단 크기 변동을 반영하기 때문에 반복조사에서 모집단 크기의 변화가 있는 경우에 사용하기 적절한 방법이다. 그리고 기존의 방법과 제안된 표본수 결정방법을 농산물소득조사 데이터에 적용하여 목표오차가 주어진 경우에 이에 대응하는 표본수를 수치적으로 구하였다. 본 연구에서 제안한 설계효과 방법을 사용하면 모집단의 크기가 줄어드는 경우에 표본추출률이 아주 작지 않으면 표본수의 일부를 줄여도 됨을 보였다.

참고문헌

농촌진흥청. 2011. ≪농업경영개선을 위한 농축산물소득 자료집≫.

박홍래. 1989. ≪통계조사론≫. 서울: 영지문화사.

통계청. 2012.. 2010 농림어업총조사. http://affcensus.go.kr/acensus 2012년 8월 15일 접속.

Cochran, W.G. 1977. *Sampling Techniques*. N.Y.: Wiley.

Lohr, S.L. 1999. *Sampling: Design and Analysis*. N.Y.: Duxbury press.

Som, R.K. 1996. *Practical Sampling Techniques*. N.Y.: Marcel Dekker.

Statistics Canada. 2003. *Survey Methods and Practices*. Ottawa: Statistics Canada.

<접수 2012/9/28; 수정 2012/11/7; 게재확정 2012/11/15>