

A Suggestion for Randomized Response Technique using Fuzzy Logic

Kyung Ho Choi¹⁾

Abstract

Randomized response technique is a survey technique for eliminating evasive answer bias. But this technique has a problem. This procedure suffers from linguistic expression in randomizing device. Thus for solving the problem, in this paper, we suggested the randomized response technique using fuzzy logic.

Keywords : Fuzzy set, α -cut set, Randomized Response Technique

1. 서 론

많은 사회조사의 수행 시 발생하는 오차 중, 비교적 통제가 어려운 부분은 거짓응답 등으로 인하여 유발되는 응답오차이다. 이러한 응답오차는 비표본오차 중에서 가장 취급하기 어려운 오차로 조사설문이 응답자의 명예나 사생활에 깊이 관련되어 있거나, 또는 개인 재산에 영향을 미치는 경우에 응답자가 응답을 회피하거나 거짓응답을 하게 됨으로써 발생된다. 부연하면 개인의 사생활과 관련된 민감한 사안 - 예컨대, A.I.D.S나 동성연애, 약물중독, 혼전성경험, 낙태여부 및 탈세 등 - 에 대한 조사시 직접질문(direct question)을 하게 되면 질문의 민감성 때문에 응답자는 응답을 회피하거나 거짓응답을 하는 경향이 있게되어, 결국 비표본오차의 증대를 가져와 추정의 신뢰성이 떨어진다. 따라서 이러한 민감한 사안에 대한 조사시에는 직접질문대신에 응답자의 조사에 대한 협력의 정도를 높일 수 있는 간접질문방식이 필요하다.

확률화응답기법(randomized response technique)이란 민감한 사안에 대한 조사시 거짓응답 등으로 인한 비표본오차를 줄이기 위하여 1965년에 Warner에 의하여 제안된 조사기법이다. 본질적으로 확률화응답기법이란 응답자에게 어떤 확률하에서 질문을 선택할 수 있는 기회를 부여하여 응답의 편의를 없애거나 줄일 수 있도록 고안된 간접질문방식이다.

한편 확률화응답기법은 조사과정에서 응답자의 신분보호를 위하여 확률장치가 도입되는 간접질문방식이기 때문에 직접질문에 비하여 정보의 손실(information loss)이 있게 된다. 그래서 확률장치에 기인한 이러한 손실을 줄여서 추정의 효율을 높이고 얻어진 정보를 좀 더 효율적으로 이용할 수 있는 새로운 기법에 관한 연구가 지속되어 온 바, Greenberg et al.(1969)의 무관질문기법,

1) Associate Professor, School of Information Technology, Jeonju University, Wansan-Gu, 560-759, Korea
E-mail : ckh414@jeonju.ac.kr

Horvitz et al.(1976)의 강요질문기법, Mangat와 Singh(1990)의 반복시행기법 등은 이에 대한 일환으로 고려할 수 있는 방법들이다.

그런데 이러한 장점에도 불구하고 확률화응답기법은 실제 적용면에서 다음과 같은 문제점을 안고 있다. Warner(1965)에 의하여 제안된 질적 확률화응답기법 형태의 조사 시, 구성된 확률장치로부터 응답을 얻음에 있어 질문의 언어적 표현에 불명확성이 발생할 수 있다. 즉, 응답자 자신이 민감집단에 속하는지 여부에 대한 명확한 기준이 모호하다. 예컨대 청소년 흡연상태에 관한 조사 시, 류제복 등(1993)에서와 같이 ‘당신은 흡연자 집단에 속하는가?’와 같은 질문에서 어디까지가 흡연자 집단인지에 대한 명확한 기준이 제시되지 않음으로 인하여 응답에 혼선이 초래될 수 있다.

이에 본 논문에서는 질적 확률화응답기법을 이용한 응답 획득 시, 기존의 crisp형태(“예” 또는 “아니오”)로 응답을 얻는 대신에 소속함수(membership function)로 표현되는 구체적인 기준을 제시해 주고 fuzzy형태(‘소속여부’와 ‘소속함수 값’)로 응답을 얻는, 퍼지논리를 이용한 확률화응답기법에 대해서 제안하고자 한다.

2. 퍼지집합과 α -수준집합

우리는 일상 생활에서 애매한 상태를 나타내는 말을 많이 사용한다. 예를 들어 ‘미녀들의 집합’, ‘키가 큰 사람들의 집합’, ‘흡연자 집단’ 등과 같이, 원소 x 가 집합에 속하는 정도를 0또는 1로는 나타낼 수 없는 집합개념을 흔히 사용한다. 위에서 예를 든 집합들은 다분히 주관적이고 집합표현에 애매한 언어가 사용되기 때문에 소속의 정도를 분명히 하기 어려우므로, 기존의 보통집합(crisp set) 개념으로는 다룰 수 없는 형태의 집합 개념이 된다. 이러한 형태의 집합개념을 다루기 위해 고안된 것이 Zadeh(1965)에 의하여 소개된 퍼지집합(fuzzy set)개념이다.

임의의 원소가 고려하는 집합 A 에 소속되는가 그렇지 않은가 하는 것은 소속함수(membership function)에 의해서 나타낼 수 있다. 그래서 보통집합의 경우 집합 A 가 주어졌다면, 원소 x 에 대한 소속함수 값 $\mu_A(x)$ 는 다음과 같다.

$$\begin{aligned} \mu_A(x) &= 1 && \text{if and only if } x \in A \\ &= 0 && \text{if and only if } x \notin A \end{aligned} \quad (2.1)$$

즉, 소속함수는 전체집합의 원소를 집합 $\{0,1\}$ 로 대응시킨다고 할 수 있다.

$$\mu_A : U \rightarrow \{0, 1\} \quad (2.2)$$

이에 반하여 소속함수의 값이 0과 1뿐만 아니라, 0과 1사이의 임의의 값을 가질 수 있는 퍼지집합을 고려할 수 있는데, 이에 대한 소속함수는 다음과 같다.

$$\mu_A : U \rightarrow [0, 1] \quad (2.3)$$

여기서 $[0, 1]$ 은 0과 1사이의 모든 실수(0과 1을 포함)를 나타낸다. 결국 보통집합에 비하여 퍼지집합은 집합의 경계가 애매한 집합이라고 할 수 있다.

한편 퍼지집합에 포함된 원소들 중에서 일정한 가능성(소속함수 값)이상을 갖는 원소들로만 구성된 보통집합을 만들 수 있다. 이것을 α -수준집합(α -cut set)이라고 부르는데, 이는 소속함수의 값이 α 이상인 원소들로 이루어진다.

$$A_\alpha = \{x \in U \mid \mu_A(x) \geq \alpha, 0 \leq \alpha \leq 1\} \quad (2.4)$$

한 예로 전체집합 U 를 청소년집단이라 하고 고려하는 집합 A 를 ‘흡연자 집단’이라고 해보자. 이 때 ‘흡연자’라는 용어는 정의하기에 따라서 애매한 표현이다. 일반적인 통념상 ‘흡연자’란 담배를 피우는 사람을 의미하지만, 예컨대 청소년 시기에 호기심 등으로 한 두 개비의 담배를 피우는 사람도 ‘흡연자’라고 분류하여 하루에 열 개비 이상 피우는 사람과 같은 흡연자집단에 속하는 것으로 평가해서는 안 된다. 이런 경우 다음과 같은 퍼지집합을 생각해 볼 수 있다.

x (하루에 피우는 담배 개비의 수)	0	1	3	5	7	10개비 이상
μ_A (‘흡연자’에 소속하는 정도)	0	0.1	0.3	0.5	0.7	1

이제 위의 퍼지집합 ‘흡연자’에서 $\alpha=0.3$ 으로 하여 만든 α -수준집합은 다음과 같다.

$$\text{흡연자}_{0.3} = \{3, 5, 7, 10\text{개비 이상}\}$$

한편 α 와 α' 에 대하여 두 개의 수준집합 A_α 와 $A_{\alpha'}$ 이 있을 때 $\alpha \leq \alpha'$ 의 관계가 있으면 $A_\alpha \supseteq A_{\alpha'}$ 의 관계가 성립한다. 그리고 소속함수의 값을 나타내는 α 는 구간 $[0, 1]$ 에 포함된 값으로서, α 로 이루어진 집합을 레벨집합(level set)이라 부른다. 즉,

$$A_\alpha = \{x \mid \mu_A(x) = \alpha, \alpha \geq 0, x \in U\} \tag{2.5}$$

이며, 따라서 앞의 퍼지집합 ‘흡연자’에 대한 레벨집합은 다음과 같다.

$$A_{\text{‘흡연자’}} = \{0, 0.1, 0.3, 0.5, 0.7, 1\}$$

3. 퍼지논리를 이용한 확률화응답기법

3.1 연구배경

앞에서도 언급했듯이 Warner(1965) 형태의 확률화응답기법이 갖는 문제점 중의 하나는 응답자에게 제시되는 질문의 언어적 표현의 애매함에 있다. 즉, 구성된 확률장치를 통하여 응답을 수행하는 과정에서 질문의 언어적 표현의 불명확성으로 인하여 “예” 나 “아니오”(보통집합 개념) 중 어느 하나로 응답하기 곤란한 경우가 생길 수 있는 바, 이는 결국 응답오차의 유발요인이 되어 전체적인 조사의 신뢰성을 떨어뜨리게 된다. 예를 들어 ‘당신은 흡연자 집단에 속하는가?’와 같은 질문에서 어디까지가 ‘흡연자 집단’인지에 대한 명확한 기준이 제시되지 않음으로 인하여 응답에 혼선이 초래될 수 있다.

사실 전체 조사대상 집합 중 ‘흡연자 집단’과 같은 집합은 앞 절에서 언급한 퍼지집합으로 보는 것이 타당하다. 따라서 보통집합의 개념으로 응답을 획득하는, Warner(1965)형태의 질적 확률화응답기법이 갖는 문제점을 해결하기 위한 방안으로 다음을 고려해 보자. 즉, 확률화응답기법을 통한 응답 획득 시, 소속함수로 표현되는 구체적인 기준을 제시해주는 (소속여부, 소속함수 값)의 형태로 응답을 얻는다. 그래서 ‘당신은 흡연자집단에 속하는가?’와 같은 민감질문에 단순히 “예”또는 “아니오”로 응답하게 하는 대신, (아니오, 0), (예, 0.1), (예, 0.3), (예, 0.5), (예, 0.7), (예, 1)과 같은 형태로 응답을 획득하는 방법을 고려할 수 있다. 여기서 0=담배 안 피움, 0.1=한개비, 0.3=3개비, 0.5=5개비, 0.7=7개비, 1=10개비 이상 등으로 미리 정한다.

참고로 한국 금연운동협의회의 조사자료(2000)에 따르면 1999년 현재 우리 나라 중학생의 흡연률은 다음과 같아, 청소년들에게 있어서 ‘흡연문제’가 민감한 사안으로 여겨지지 않을 수도 있으나

편의상 본 논문에서는 이 문제를 이용한 확률화응답기법을 고려하겠다.

중학교(남)				
5이하	6-10	11-15	15-20	20이상
1.1(%)	8.7	7.0	6.2	24.7

3.2 응답획득과 추정의 과정

본 논문에서 제시한 퍼지논리를 이용한 확률화응답기법을 통한 조사를 수행하기 위해서는 확률장치에 나타나는 질문의 구성에 주의를 기울일 필요가 있다. 이는 응답자에게 요구되는 응답이 질적응답(소속여부) 뿐만 아니라 양적응답(소속함수값)도 있기 때문이다. 그래서 수행하고자 하는 조사가 '청소년 흡연실태' 조사라면 다음과 같은 무관질문, 즉 Greenberg et al.(1969)이 제시한 형태의 질문을 고려할 수 있다.

<표 3-1> 확률장치

<p>Q1: 당신은 흡연집단에 속하는가? 그 정도는? (p)</p> <p>0=담배 안 피움, 0.1=한개비, 0.3=3개비, 0.5=5개비, 0.7=7개비</p> <p>1=10개비 이상</p> <p>Q2: 당신은 PC방 이용집단에 속하는가? 그 정도는? (1-p)</p> <p>이용 안 하는 경우에는 0, 이용집단에 속하는 경우에는 0.1, 0.3, 0.5, 0.7, 1 중 임의로 하나를 선택</p>
--

표본을 추출하고 응답을 얻는 과정은 기존의 무관질문기법과 동일하다. 다만 응답이 “예”나 “아니오”로만 얻어지는 것이 아니라, (“예”또는 “아니오”, 소속함수 값)의 형태로 얻어진다. 한 예로 질문 Q1을 선택한 응답자가 담배를 피우지 않는 경우라면 (“아니오”, 0)이라고 응답을 하고, Q2를 선택한 응답자가 PC방 이용집단에 속하는 경우라면 “예”와 0.1, 0.3, 0.5, 0.7, 1 중 임의로 하나 선택한 값을 응답하게 된다. 이렇게 함으로써 응답자가 선택된 질문에 대한 응답 시 느끼는 민감도를 줄이어 정직하게 응답하도록 할 수 있다.

한편 위의 과정을 통하여 얻어진 응답은 퍼지집합 형태이므로 이제 적절한 α 를 선택하여 α -수준집합을 구한 후 이에 대해서 민감집단(흡연집단)의 비율을 추정한다. 한 예로 위와 같이 설계된 경우라면 $\alpha=0.1$ 일 때의 α -수준집합을 이용한 민감집단의 비율추정은 기존의 무관질문기법과 동일하게 되어, 결국 본 논문에서 제시한 방법은 기존의 방법에 대한 일반화라 할 수 있다.

4. 예 제

본 논문에서 제시한 방법에 의하여 전체집단 내의 민감집단의 비율을 추정하는 과정을 예로 들어보자. 표 3-1과 같이 구성된 확률장치에 대하여 단순임의 추출된 $n=200$ 명의 응답자로부터 획득된 응답이 다음과 같다고 하자. 여기서 y 는 확률장치를 통한 응답이 “예”인 경우이고, n 은 “아니오”를 의미한다.

<표 4-1> n=200명에 대한 확률장치를 통한 응답

(n, 0)	(y, 1)	(y, 0.5)	(y, 1)	(y, 0.1)	(n, 0)	(y, 1)	(n, 0)	(n, 0)	(y, 1)
(y, 0.7)	(y, 0.1)	(y, 1)	(n, 0)	(y, 1)	(y, 0.3)	(n, 0)	(n, 0)	(y, 0.5)	(y, 1)
(y, 1)	(n, 0)	(y, 1)	(y, 0.3)	(y, 1)	(y, 1)	(y, 1)	(y, 0.7)	(y, 1)	(y, 0.1)
(y, 0.3)	(n, 0)	(n, 0)	(y, 1)	(y, 0.5)	(y, 0.7)	(n, 0)	(y, 1)	(y, 1)	(n, 0)
(y, 1)	(y, 0.1)	(y, 1)	(y, 0.5)	(y, 1)	(n, 0)	(n, 0)	(y, 1)	(y, 0.5)	(y, 1)
(y, 0.7)	(y, 1)	(n, 0)	(n, 0)	(y, 1)	(y, 0.1)	(n, 0)	(y, 1)	(n, 0)	(y, 1)
(y, 1)	(y, 0.1)	(y, 1)	(y, 0.3)	(n, 0)	(y, 0.5)	(y, 1)	(n, 0)	(y, 0.7)	(y, 0.3)
(n, 0)	(y, 1)	(n, 0)	(y, 1)	(y, 0.7)	(n, 0)	(n, 0)	(y, 1)	(n, 0)	(n, 0)
(y, 1)	(n, 0)	(y, 1)	(y, 1)	(n, 0)	(y, 0.5)	(y, 1)	(n, 0)	(y, 1)	(y, 1)
(y, 0.5)	(y, 0.1)	(n, 0)	(y, 1)	(y, 0.1)	(y, 0.7)	(n, 0)	(y, 0.5)	(y, 1)	(n, 0)
(y, 1)	(n, 0)	(y, 1)	(n, 0)	(y, 0.3)	(y, 1)	(n, 0)	(y, 1)	(n, 0)	(y, 0.3)
(n, 0)	(n, 0)	(y, 0.5)	(y, 0.3)	(y, 1)	(n, 0)	(y, 0.3)	(y, 1)	(y, 0.7)	(y, 1)
(y, 1)	(y, 1)	(n, 0)	(y, 1)	(n, 0)	(y, 1)	(y, 0.5)	(y, 1)	(y, 1)	(y, 1)
(y, 0.5)	(y, 1)	(y, 1)	(n, 0)	(y, 1)	(y, 0.1)	(y, 0.7)	(y, 0.5)	(n, 0)	(y, 0.3)
(n, 0)	(n, 0)	(y, 0.3)	(n, 0)	(n, 0)	(y, 1)	(y, 0.3)	(y, 1)	(n, 0)	(n, 0)
(y, 1)	(n, 0)	(y, 1)	(n, 0)	(y, 0.3)	(y, 1)	(y, 0.5)	(y, 0.1)	(n, 0)	(y, 1)
(n, 0)	(y, 1)	(y, 0.5)	(n, 0)	(y, 0.1)	(n, 0)	(y, 0.7)	(n, 0)	(y, 1)	(y, 0.5)
(y, 1)	(n, 0)	(y, 1)	(n, 0)	(y, 1)	(y, 1)	(n, 0)	(y, 1)	(y, 0.3)	(n, 0)
(y, 1)	(n, 0)	(y, 0.3)	(y, 1)	(n, 0)	(y, 0.7)	(y, 1)	(n, 0)	(y, 1)	(y, 0.3)
(y, 0.3)	(n, 0)	(y, 1)	(y, 0.5)	(y, 1)	(y, 0.3)	(n, 0)	(y, 0.1)	(n, 0)	(y, 1)

Greenberg et al.(1969)의 무관질문기법에 의하여 모집단 내의 민감집단의 비율을 추정할 경우에 대한 추정량은 다음과 같다.

$$\hat{\pi} = \frac{n_1 - (1-p)\pi_y}{p} \tag{4.1}$$

단 n_1 은 응답자 중 “예”라고 응답한 응답자의 수이며, π_y 는 무관집단의 모비율로 기지(known)이다.

한편 Greenberg et al.(1969)의 무관질문기법에 의하여 n=200명으로부터 획득된 응답이 표 4-1의 “예”나 “아니오”만으로 구성된 것과 같고, p=0.4이며 $\pi_y=0.6$ 이라면 식 (4.1)에 의한 모집단 내의 민감집단의 비율에 대한 추정치는 $\hat{\pi}=0.775$ 이다. 그런데 이 값은 단지 확률장치를 통한 “예”나 “아니오”만의 응답으로부터 구해지는 추정치로, 실제로 구하고자 하는 민감집단 즉 좀더 정확하게 정의된 민감집단에 대한 추정값보다 과추정된 결과일 수 있다. 나아가 이 결과는 응답을 얻는 과정에서 질문의 언어적 표현의 불명확성으로부터 유발되는 응답오차를 포함하고 있는 추정치여서, 신뢰성이 떨어지게 된다.

이제 청소년 흡연실태조사에서 구하고자 하는 민감집단이, 하루에 적어도 3개비 이상의 담배를 흡연하는 집단이라고 해보자. 그러면 이 경우 획득된 표 4-1의 응답으로부터 표 4-2와 같은 $\alpha = 0.3$ 인 α -수준집합을 고려하고, 식 (4.1)의 추정식을 이용하여 구하고자 하는 민감집단의 추정값을 구할 수 있다.

<표 4-2> $\alpha=0.3$ 인 α -수준집합에 대응하는 n=200명의 응답

(n, 0)	(y, 1)	(y, 0.5)	(y, 1)	(n, 0)	(n, 0)	(y, 1)	(n, 0)	(n, 0)	(y, 1)
(y, 0.7)	(n, 0)	(y, 1)	(n, 0)	(y, 1)	(y, 0.3)	(n, 0)	(n, 0)	(y, 0.5)	(y, 1)
(y, 1)	(n, 0)	(y, 1)	(y, 0.3)	(y, 1)	(y, 1)	(y, 1)	(y, 0.7)	(y, 1)	(n, 0)
(y, 0.3)	(n, 0)	(n, 0)	(y, 1)	(y, 0.5)	(y, 0.7)	(n, 0)	(y, 1)	(y, 1)	(n, 0)
(y, 1)	(n, 0)	(y, 1)	(y, 0.5)	(y, 1)	(n, 0)	(n, 0)	(y, 1)	(y, 0.5)	(y, 1)
(y, 0.7)	(y, 1)	(n, 0)	(n, 0)	(y, 1)	(n, 0)	(n, 0)	(y, 1)	(n, 0)	(y, 1)
(y, 1)	(n, 0)	(y, 1)	(y, 0.3)	(n, 0)	(y, 0.5)	(y, 1)	(n, 0)	(y, 0.7)	(y, 0.3)
(n, 0)	(y, 1)	(n, 0)	(y, 1)	(y, 0.7)	(n, 0)	(n, 0)	(y, 1)	(n, 0)	(n, 0)
(y, 1)	(n, 0)	(y, 1)	(y, 1)	(n, 0)	(y, 0.5)	(y, 1)	(n, 0)	(y, 1)	(y, 1)
(y, 0.5)	(n, 0)	(n, 0)	(y, 1)	(n, 0)	(y, 0.7)	(n, 0)	(y, 0.5)	(y, 1)	(n, 0)
(y, 1)	(n, 0)	(y, 1)	(n, 0)	(y, 0.3)	(y, 1)	(n, 0)	(y, 1)	(n, 0)	(y, 0.3)
(n, 0)	(n, 0)	(y, 0.5)	(y, 0.3)	(y, 1)	(n, 0)	(y, 0.3)	(y, 1)	(y, 0.7)	(y, 1)
(y, 1)	(y, 1)	(n, 0)	(y, 1)	(n, 0)	(y, 1)	(y, 0.5)	(y, 1)	(y, 1)	(y, 1)
(y, 0.5)	(y, 1)	(y, 1)	(n, 0)	(y, 1)	(n, 0)	(y, 0.7)	(y, 0.5)	(n, 0)	(y, 0.3)
(n, 0)	(n, 0)	(y, 0.3)	(n, 0)	(n, 0)	(y, 1)	(y, 0.3)	(y, 1)	(n, 0)	(n, 0)
(y, 1)	(n, 0)	(y, 1)	(n, 0)	(y, 0.3)	(y, 1)	(y, 0.5)	(n, 0)	(n, 0)	(y, 1)
(n, 0)	(y, 1)	(y, 0.5)	(n, 0)	(n, 0)	(n, 0)	(y, 0.7)	(n, 0)	(y, 1)	(y, 0.5)
(y, 1)	(n, 0)	(y, 1)	(n, 0)	(y, 1)	(y, 1)	(n, 0)	(y, 1)	(y, 0.3)	(n, 0)
(y, 1)	(n, 0)	(y, 0.3)	(y, 1)	(n, 0)	(y, 0.7)	(y, 1)	(n, 0)	(y, 1)	(y, 0.3)
(y, 0.3)	(n, 0)	(y, 1)	(y, 0.5)	(y, 1)	(y, 0.3)	(n, 0)	(n, 0)	(n, 0)	(y, 1)

표 4-2에서 음영으로 표시된 부분은 원래응답에서는 (“예”, 0.1)에 해당하는 응답이었으나, $\alpha = 0.3$ 인 α -수준집합을 이용하는 경우에는 (“아니오”, 0)으로 처리되는 응답을 나타낸다.

따라서 이 경우 구하고자 하는 민감집단의 모비율에 대한 추정치는 $\widehat{\pi}_{0.3} = 0.625$ 가 된다. 한편 다양한 α -수준집합에 대한 추정치를 구해보면 다음과 같다.

α	α 에 따른 n_1 의 수	추정치 $\widehat{\pi}$
Greenberg et al.	134	0.775
0.3	122	0.625
0.5	104	0.40
0.7	87	0.1875
1	76	0.05

5. 결론 및 토의

확률화응답기법은 응답자의 신분보호를 통하여 신뢰할만한 응답을 얻어낼 수 있는 장점에도 불구하고 실제 적용면에서 몇 가지 문제점을 안고 있다. 즉 Warner(1965)에 의하여 제안된 질적 확률화응답기법 형태의 조사 시, 구성된 확률장치로부터 응답을 얻음에 있어 질문의 언어적 표현에 불명확성이 발생할 수 있다. 예컨대 응답자 자신이 민감집단에 속하는지 여부에 대한 명확한 기준이 모호하다. 앞서 언급한대로 청소년 흡연실태에 관한 조사 시, 류제복 등(1993)에서와 같이 ‘당신은 흡연자 집단에 속하는가?’와 같은 질문에서 어디까지가 흡연자 집단인지에 대한 명확한 기준

이 제시되지 않음으로 인하여 응답에 혼선이 초래될 수 있다.

이에 본 논문에서는 확률화응답기법을 통한 응답 획득 시, 소속함수로 표현되는 구체적인 기준을 제시해주는 (소속여부, 소속함수 값)의 형태로 응답을 얻음으로써, 구하고자 하는 민감집단에 대한 좀더 구체적이고 정확한 추정을 할 수 있는 퍼지논리를 이용한 확률화응답기법에 대해서 알아보았다. 그 결과 앞의 예에서 확인 있듯이, 다양한 α -수준집합에 따라 좀 더 정확한 추정을 수행할 수 있음을 알 수 있었다.

마지막으로 본 논문에서 제안한 방법에 대한 장점과 단점은 다음과 같다. 먼저 장점으로 Warner(1965)형태의 기법에서 발생하는 언어적 표현의 애매성을 해결할 수 있어, 결국 응답오차로 인하여 야기되는 문제를 방지해서 추정의 신뢰성을 높일 수 있다. 나아가 다양한 α -수준집합에 대한 모비율 추정이 가능하여 한 번의 조사로 여러 집단에 대한 다각적인 추정 및 결론을 도출할 수 있다. 즉, 한번의 조사로 Abul-Ela et al.(1967)이 제시한 다지모형의 효과를 얻을 수 있다.

한편 문제점으로는 조사과정을 응답자에게 충분히 이해시키기 위해서는 기존의 방법에 비해 약간의 시간이 더 소요될 수 있으며, 따라서 조사비용의 증가가 우려되지만, 이는 다양한 α -수준집합에 대한 모비율의 추정이 가능한 점과 언어적 표현의 애매함이 해결될 수 있어 추정의 신뢰성을 높일 수 있는 점 등으로 충분히 상쇄될 수 있다.

참고문헌

- [1] 류제복, 홍기학, 이기성(1993). 「확률화응답모형」, 자유아카데미, 서울.
- [2] Abul-Ela, Abdel-Latif A., Greenberg, B. G. and Horvitz, D. G.(1967). Multiproportions randomized response model, *Journal of the American Statistical Association*, 62, 990-1008.
- [3] Greenberg, B. G., Abul-Ela, Abdel-Latif A., Simmons, W. R. and Horvitz, D. G.(1969). The unrelated question randomized response model : Theoretical framework, *Journal of the American Statistical Association*, 64, 520-539.
- [4] Horvitz, D. G., Shah, B.V. and Simmons, W. R.(1976). The unrelated question randomized response model, *American Statistical Association Proceedings of Social Statistics*, 65-72.
- [5] Mangat, N. S. and Singh, R.(1990). An alternative randomized response procedure, *Biometrika*, 77, 439-442.
- [6] Warner, S. L.(1965). Randomized response : A survey technique for eliminating evasive answer bias, *Journal of the American Statistical Association*, 60, 63-69.
- [7] Zadeh, L. A.(1965). Fuzzy sets, *Information and Control*, 8, 338-353.