The Korean Communications
in Statistics Vol. 8, No. 3, 2001
pp. 753-765

A Conditional Unrelated Question Model with Quantitative Attribute[1)]

Gi Sung Lee[2)], Ki Hak Hong[3)]

Abstract

We suggest a quantitative conditional unrelated question model that can be used in obtaining more sensitive information. For whom say "yes" about the less than sensitive question B we ask only about the more sensitive variable X. We extend our model to two sample case when there is no information about the true mean of the unrelated variable Y. Finally we compare the efficiency of our model with that of Greenberg et al.'s.

Keywords : quantitative conditional unrelated question model, two sample unrelated question model, Greenberg et al's model.

1. 서론

민감한 조사에서 응답자들이 응답을 회피하거나 정직하게 응답하지 않는 질문들에 대하여 응답자의 신분이나 비밀을 노출시키지 않고서 민감한 질문에 대한 정보를 이끌어 내기 위하여 Warner(1965)는 확률장치를 이용하여 간접 응답을 하도록 하는 확률화응답모형(randomized response model ; RRM)을 처음으로 제시하였다. Warner는 응답자들에게 민감한 질문과, 민감한 질문과 배반되는 질문으로 구성된 확률장치를 이용하는 관련질문모형을 제안하였고, Greenberg 등(1971)은 민감한 질문과 배반되는 질문 대신에 민감한 질문과 무관한 질문을 사용하는 무관질문모형(unrelated question model)을 제안하여 양적속성의 무관질문모형으로 확장하기도 하였다. 또한, Loynes (1976)는 Warner모형의 민감한 질문과 배반이 되는 질문 대신에 "예"라고 응답하도록 강요하는 강요질문모형(forced answer model)을 제안하였다. 특히, Carr 등(1982)은 덜 민감한 속성 B와 강요질문으로 구성된 확률장치를 통해 "예"라고 응답한 사람들에게만 Loynes의 강요질문모형을 사용하도록 하여 민감한 속성 A에 대한 모비율을 추정해 내는 조건부 확률화응답모형

---

1) This paper was supported by Woosuk University
2) Associate Professor, Department of Computer Science & Statistics, Woosuk University, Wanju-gun, Jeonbuk, 565-701, Korea.
E-mail : gisung@core.woosuk.ac.kr
3) Associate Professor, Department of Computer Science, Dongshin University, Daeho-dong, Naju, Chonnam, 520-714, Korea.
E-mail : khhong@blue.dongshinu.ac.kr

을 제안하였다.

본 논문에서는 덜 민감한 속성 B와 강요질문으로 구성된 확률장치를 통해 "예"라고 응답한 사람들에게만 Greenberg 등의 양적속성의 무관질문모형을 사용하도록 하여 민감한 변수 X에 대한 정보를 얻을 수 있는 양적속성의 조건부 무관질문모형을 제안하고자 한다. 그리고, 제안한 양적속성의 조건부 무관질문모형을 무관한 변수를 모르고 있을 때 두 개의 독립표본을 이용하는 양적속성의 이표본 조건부 무관질문모형으로 확장하고자 한다. 또한, 제안한 양적속성의 조건부 무관질문모형과 Greenberg 등의 양적속성의 무관질문모형과의 효율성을 비교해 보고자 한다.

2. 양적속성의 조건부 무관질문모형

이 장에서는 덜 민감한 속성 B와 강요질문으로 구성된 확률장치를 통해 "예"라고 응답한 사람들에게만 민감한 변수 X와 무관한 변수 Y를 포함하고 있는 Greenberg 등의 양적속성의 무관질문모형을 사용하도록 하여 민감한 변수 X에 대한 정보를 얻을 수 있는 양적속성의 조건부 무관질문모형을 제안하고자 한다.

Greenberg 등의 양적속성의 무관질문모형을 사용하는데 있어서 민감한 변수 X가 연속인 밀도함수 $g(\cdot)$를 갖는다고 가정하고, Y를 밀도함수 $h(\cdot)$를 갖는 무관한 변수라 하자. 이 때, 무관한 변수 Y의 모평균 μ_y를 알고 있다고 가정하며, 응답자가 Z라고 응답한다고 가정하자. 그리고, Z의 모평균과 모분산을 각각 μ_z, σ_z^2이라 하자.

첫 번째 단계에서는 단순임의추출된 n명의 응답자에게 다음과 같은 2개의 질문으로 구성된 Loynes의 강요질문모형의 확률장치 R_1를 통해 선택된 질문에 대하여 응답하도록 한다.

< 첫 번째 단계의 확률장치 R_1 >

	질문내용	선택확률
질문 1	당신은 덜 민감한 속성 B를 가지고 있습니까?	p
질문 2	"예"라고 응답하시오.	$1-p$

이 때, 응답자들은 확률장치에 의해서 질문 1이 선택되면 "예" 또는 "아니오"라고 응답하고, 질문 2가 선택되면 "예"라고 응답하게 된다.

첫 번째 단계에서 응답자가 "예"라고 응답할 확률을 λ라 하면, 그 확률은 다음과 같다.

$$\lambda = p\pi_B + (1-p)\,. \tag{2.1}$$

여기서, π_B은 덜 민감한 속성 B에 대한 모비율이다.

이 때, n명의 응답자들 중에서 "예"라고 응답한 사람의 수를 n'명이라 하면, $\hat{\lambda} = \frac{n'}{n}$ 이 되

므로 덜 민감한 속성 B에 대한 모비율 π_B의 추정량 $\hat{\pi}_B$는 다음과 같다.

$$\begin{aligned}\hat{\pi}_B &= [\hat{\lambda}-(1-p)]/p \\ &= \frac{1}{np}[n'-(1-p)n]. \end{aligned} \tag{2.2}$$

두 번째 단계에서는 첫 번째 단계에서 "예"라고 응답한 n'명의 응답자들만이 다음과 같은 2개의 질문으로 구성된 Greenberg 등의 양적속성의 무관질문모형의 확률장치 R_2를 이용하게 된다.

<두 번째 단계의 확률장치 R_2>

	질문내용	선택확률
질문 1	당신의 민감한 변수 X에 대한 값은 얼마입니까?	p
질문 2	당신의 무관한 변수 Y에 대한 값은 얼마입니까?	$1-p$

그러므로, 응답자가 첫 번째 단계에서 "예"라고 응답했다는 조건하에서 두 번째 단계에서 Z라고 응답할 때 기대할 수 있는 응답의 평균 즉, Z의 모평균 μ_z는 다음과 같다.

$$\mu_z = p\frac{\mu_x}{\lambda} + (1-p)\mu_y. \tag{2.3}$$

첫 번째 단계에서 "예"라고 응답한 n'명의 응답자들이 두 번째 단계에서 $Z_i(i=1,2,\cdots,n')$라고 응답했을 때, 민감한 변수 X에 대한 모평균 μ_x의 추정량 $\hat{\mu}_{x1}$는 다음과 같다.

$$\begin{aligned}\hat{\mu}_{x1} &= \hat{\lambda}\left[\frac{\hat{\mu}_z-(1-p)\mu_y}{p}\right] \\ &= \frac{1}{np}[Z_0-(1-p)\mu_y n'].\end{aligned} \tag{2.4}$$

여기서, $Z_0 = n'\hat{\mu}_z$, $\hat{\mu}_z = \frac{1}{n'}\sum_{i=1}^{n'} Z_i$, $E(\hat{\mu}_z \mid n') = \mu_z$, $Var(\hat{\mu}_z \mid n') = \frac{\sigma_z^2}{n'}$ 이다.

<정리 1> 추정량 $\hat{\mu}_{x1}$는 모평균 μ_x의 비편향추정량이다.

(증명)

$n' \sim b(n,\lambda)$이고, $E(Z_0) = n\lambda\mu_z$이므로, $\hat{\mu}_{x1}$의 기대값을 구해보면 다음과 같다.

$$
\begin{aligned}
E(\hat{\mu}_{x1}) &= E\left[\frac{1}{np}\{Z_0-(1-p)\mu_y n'\}\right] \\
&= \frac{1}{np}[E(Z_0)-(1-p)\mu_y E(n')] \\
&= \frac{1}{np}[n\lambda\mu_z-(1-p)\mu_y n\lambda] \\
&= \frac{\lambda}{p}\left[p\frac{\mu_x}{\lambda}+(1-p)\mu_y-(1-p)\mu_y\right] \\
&= \mu_x .
\end{aligned}
$$

<정리 2> 추정량 $\hat{\mu}_{x1}$ 의 분산은 다음과 같다.

$$
Var(\hat{\mu}_{x1}) = \frac{\lambda\,\sigma_z^2}{np^2}+\frac{\mu_x^2(1-\lambda)}{n\lambda} . \tag{2.5}
$$

(증명)

$$
\begin{aligned}
Var(\hat{\mu}_{x1}) &= Var\left[\frac{1}{np}\{Z_0-(1-p)\mu_y n'\}\right] \\
&= \frac{Var(Z_0)+(1-p)^2\,\mu_y^2\,Var(n')-2(1-p)\mu_y\,Cov(n',Z_0)}{(np)^2}
\end{aligned} \tag{2.6}
$$

에서 $n' \sim b(n,\lambda)$이고 $E(Z_0)=n\lambda\mu_z$이며 $E(Z_0 \mid n')=n'\mu_z$이므로, 다음을 얻을 수 있다.

$$
\begin{aligned}
Var(n') &= n\lambda(1-\lambda), \\
Var(Z_0) &= E[Var(Z_0 \mid n')]+Var[E(Z_0 \mid n')] \\
&= E[n'\,\sigma_z^2]+Var(n'\mu_z) \\
&= \sigma_z^2 E(n')+\mu_z^2\,Var(n') \\
&= \sigma_z^2 n\lambda+\mu_z^2 n\lambda(1-\lambda) \\
&= n\lambda[\sigma_z^2+(1-\lambda)\mu_z^2], \\
Cov(n',Z_0) &= E(n'Z_0)-E(n')E(Z_0) \\
&= E[n'E(Z_0 \mid n')]-(n\lambda)(n\lambda\mu_z) \\
&= E(n'n'\mu_z)-n^2\lambda^2\mu_z \\
&= \mu_z E(n'^2)-n^2\lambda^2\mu_z
\end{aligned}
$$

$$= \mu_z Var(n') + \mu_z[E(n')]^2 - n^2 \lambda^2 \mu_z$$
$$= \mu_z n\lambda(1-\lambda) + n^2 \lambda^2 \mu_z - n^2 \lambda^2 \mu_z$$
$$= n\lambda(1-\lambda)\mu_z .$$

따라서, 위의 세 식을 식(2.6)에 대입하여 $Var(\hat{\mu}_{x1})$를 구해보면 식(2.5)를 얻을 수 있다.

3. 양적속성의 이표본 조건부 무관질문모형

이 장에서는 2장에서 제안한 양적속성의 조건부 무관질문모형을 무관한 변수 Y의 모평균 μ_y를 모르고 있는 경우에 사용할 수 있는 이표본 조건부 무관질문모형으로 확장해 보고자 한다.

첫 번째 단계에서는 2장에서와 마찬가지로 단순임의추출된 n명의 응답자에게 2개의 질문으로 구성된 Loynes의 강요질문모형의 확률장치를 통해 선택된 질문에 대하여 응답하도록 한다.

무관한 변수 Y의 모평균 μ_y를 모르고 있는 경우에는 두 개의 서로 다른 독립표본이 필요하게 된다. 따라서, 두 번째 단계에서는 첫 번째 단계에서 "예"라고 응답한 n'명의 응답자들로부터 단순임의복원으로 n_1'과 n_2'명의 임의표본을 추출한다. 그리고, $n_i'(i=1,2)$명의 응답자들은 각각 다음과 같은 확률장치 $R_2(i)$를 이용하여 응답하도록 한다.

<두 번째 단계의 확률장치 $R_2(i)$>

	질문내용	선택확률
질문 1	당신의 민감한 변수 X에 대한 값은 얼마입니까?	p_i
질문 2	당신의 무관한 변수 Y에 대한 값은 얼마입니까?	$1-p_i$

응답자가 첫 번째 단계에서 "예"라고 응답했다는 조건하에서 두 번째 단계에서 Z라고 응답할 때 기대할 수 있는 응답의 평균 즉, Z의 모평균 μ_{zi} $(i=1,2)$는 다음과 같다.

$$\mu_{zi} = E(\overline{Z}_i) = p_i \frac{\mu_x}{\lambda} + (1-p_i)\mu_y . \tag{3.1}$$

첫 번째 단계에서 "예"라고 응답한 n'명의 응답자들이 두 번째 단계에서 $Z_{ij}(i=1,2,\ j=1,2,\cdots,n_i')$라고 응답했을 때, 민감한 변수 X에 대한 모평균 μ_x의 추정량 $\hat{\mu}_{x2}$는 다음과 같다.

$$\begin{aligned}\hat{\mu}_{x2} &= \hat{\lambda}\left[\frac{(1-p_2)\,\overline{Z}_1-(1-p_1)\,\overline{Z}_2}{p_1-p_2}\right] \\ &= \frac{(1-p_2)Z_{10}-(1-p_1)Z_{20}}{n(p_1-p_2)}\ ,\quad p_1 \neq p_2\ .\end{aligned} \tag{3.2}$$

여기서, $Z_{10}=n'\ \overline{Z}_1$, $Z_{20}=n'\ \overline{Z}_2$, $\overline{Z}_1=\frac{1}{n_1'}\sum_{j=1}^{n_1'}Z_{1j}$, $\overline{Z}_2=\frac{1}{n_2'}\sum_{j=1}^{n_2'}Z_{2j}$이다.

<정리 3> 추정량 $\hat{\mu}_{x2}$는 모평균 μ_x의 비편향추정량이다.

(증명)

$n' \sim b(n,\lambda)$이므로 $E(Z_{10})=n\lambda\mu_{z1}$, $E(Z_{20})=n\lambda\mu_{z2}$이므로 $\hat{\mu}_{x2}$의 기대값을 구해보면 다음과 같다.

$$\begin{aligned}E(\hat{\mu}_{x2}) &= E\left[\frac{(1-p_2)Z_{10}-(1-p_1)Z_{20}}{n(p_1-p_2)}\right] \\ &= \frac{(1-p_2)n\lambda\mu_{z1}-(1-p_1)n\lambda\mu_{z2}}{n(p_1-p_2)} \\ &= \frac{(1-p_2)n\lambda\left[p_1\frac{\mu_x}{\lambda}+(1-p_1)\mu_y\right]-(1-p_1)n\lambda\left[p_2\frac{\mu_x}{\lambda}+(1-p_2)\mu_y\right]}{n(p_1-p_2)} \\ &= \frac{(1-p_2)np_1\mu_x-(1-p_1)np_2\mu_x}{n(p_1-p_2)} \\ &= \mu_x\ .\end{aligned}$$

<정리 4> 추정량 $\hat{\mu}_{x2}$의 분산은 다음과 같다.

$$Var(\hat{\mu}_{x2}) = \frac{\lambda\left[\{(1-\lambda)+n\lambda\}\left\{(1-p_2)^2\frac{\sigma_{z1}^{\ 2}}{n_1'}+(1-p_1)^2\frac{\sigma_{z2}^{\ 2}}{n_2'}\right\}\right]}{n(p_1-p_2)^2}+\frac{\mu_x^{\ 2}(1-\lambda)}{n\lambda}\ . \tag{3.3}$$

(증명)

$$Var(\hat{\mu}_{x2}) = Var\left[\frac{(1-p_2)Z_{10}-(1-p_1)Z_{20}}{n(p_1-p_2)}\right]$$

$$= \frac{(1-p_2)^2 Var(Z_{10}) + (1-p_1)^2 Var(Z_{20}) - 2(1-p_1)(1-p_2)Cov(Z_{10}, Z_{20})}{n^2(p_1-p_2)^2} \quad . \quad (3.4)$$

에서 $n' \sim b(n, \lambda)$이고 $E(Z_{10}) = n\lambda\mu_{z1}$, $E(Z_{20}) = n\lambda\mu_{z2}$이며, $E(Z_{10} \mid n') = n'\mu_{z1}$, $E(Z_{20} \mid n') = n'\mu_{z2}$이므로 다음을 얻을 수 있다.

$$\begin{aligned}
Var(Z_{10}) &= E[Var(Z_{10} \mid n')] + Var[E(Z_{10} \mid n')] \\
&= E\left[n'^2 \frac{\sigma_{z1}^2}{n_1'} \right] + Var(n'\mu_{z1}) \\
&= \frac{\sigma_{z1}^2}{n_1'} E(n'^2) + \mu_{z1}^2 Var(n') \\
&= \frac{\sigma_{z1}^2}{n_1'} [Var(n') + \{E(n')\}^2] + \mu_{z1}^2 n\lambda(1-\lambda) \\
&= \frac{\sigma_{z1}^2}{n_1'} [n\lambda(1-\lambda) + (n\lambda)^2] + \mu_{z1}^2 n\lambda(1-\lambda) \\
&= n\lambda\left[\frac{\sigma_{z1}^2}{n_1'} \{(1-\lambda) + n\lambda\} + (1-\lambda)\mu_{z1}^2 \right],
\end{aligned}$$

$$Var(Z_{20}) = n\lambda\left[\frac{\sigma_{z2}^2}{n_2'} \{(1-\lambda) + n\lambda\} + (1-\lambda)\mu_{z2}^2 \right],$$

$$\begin{aligned}
Cov(Z_{10}, Z_{20}) &= E(Z_{10}Z_{20}) - E(Z_{10})E(Z_{20}) \\
&= E[n'^2 E(Z_{10}Z_{20} \mid n')] - n\lambda\mu_{z1} n\lambda\mu_{z2} \\
&= \mu_{z1}\mu_{z2} E(n'^2) - n^2\lambda^2\mu_{z1}\mu_{z2} \\
&= \mu_{z1}\mu_{z2} Var(n') + \mu_{z1}\mu_{z2}[E(n')]^2 - n^2\lambda^2\mu_{z1}\mu_{z2} \\
&= \mu_{z1}\mu_{z2} n\lambda(1-\lambda) + n^2\lambda^2\mu_{z1}\mu_{z2} - n^2\lambda^2\mu_{z1}\mu_{z2} \\
&= n\lambda(1-\lambda)\mu_{z1}\mu_{z2} .
\end{aligned}$$

따라서, 위의 세 식을 식(3.4)에 대입하여 $Var(\hat{\mu}_{x2})$를 구해보면 식(3.3)을 얻을 수 있다.

한편, 식(3.3)의 분산은 p_1과 p_2의 선택에 따라 그 크기가 크게 차이가 나타나게 된다. 이에 Greenberg 등은 $p_1 + p_2 = 1$이 되도록 하며 p_2를 p_1으로부터 멀리 떨어진 값으로 선택하는 것

이 바람직하다고 주장하였고, Moors(1971)는 $p_2 = 0$가 가장 바람직하는 것을 제시하였다.

4. 효율성 비교

이 장에서는 제안한 양적속성의 조건부 무관질문모형과 Greenberg 등의 양적속성의 무관질문모형과의 효율성을 분산을 이용하여 비교해 보고자 한다.

4.1 제안한 양적속성의 조건부 무관질문모형과 Greenberg 등의 양적속성의 무관질문모형과의 비교

응답자들로부터 응답을 얻는 과정이 제안한 양적속성의 조건부 무관질문모형의 경우는 2단계로 이루어지고, Greenberg 등의 양적속성의 무관질문모형은 1단계로 이루어지고 있다. 따라서, 제안한 양적속성의 조건부 무관질문모형의 경우에 민감한 변수 X에 대한 정보를 얻기 위해 사용된 표본의 크기는 n'명이지만, Greenberg 등의 양적속성의 무관질문모형에서 사용된 표본의 크기는 n명이 된다. 근본적으로 두 모형간에는 이와 같은 차이점을 가지고 있는 것이 사실이다. 하지만, 제안한 양적속성의 조건부 무관질문모형의 경우에 전체적인 응답절차에서 사용된 표본의 크기가 n명이고 효율성을 비교하는 데 사용된 분산이 n에 의존하고 있으므로 두 모형의 효율성을 분산 비교를 통해 실시해 보고자 한다.

Greenberg 등의 양적속성의 무관질문모형에서 민감한 변수 X의 추정량 $\hat{\mu}_{g1}$의 분산은 다음과 같다.

$$Var(\hat{\mu}_{g1}) = \frac{\sigma_z^2}{np^2} . \tag{4.1}$$

여기서, σ_z^2은 응답자들이 Z라고 응답한다고 할 때, Z의 모분산이다.

제안한 양적속성의 조건부 무관질문모형의 분산 식(2.5)와 Greenberg 등의 양적속성의 무관질문모형의 분산 식(4.1)에서 $Var(\hat{\mu}_{g1}) - Var(\hat{\mu}_{x1}) > 0$을 이용하여

$$Var(\hat{\mu}_{g1}) - Var(\hat{\mu}_{x1}) = \frac{(1-\lambda)(\lambda\,\sigma_z^2 - p^2\,\mu_x^2)}{\lambda n p^2} > 0$$

으로부터 $Var(\hat{\mu}_{x1}) < Var(\hat{\mu}_{g1})$를 만족하는 식을 μ_x의 기준으로 정리하여 나타내 보면 다음과 같다.

$$\mu_x < \frac{\sqrt{p\pi_B + (1-p)}\,\sigma_z}{p} . \tag{4.2}$$

식(4.2)는 제안한 양적속성의 조건부 무관질문모형의 효율성이 기본적으로 μ_x의 값이 상대적으로 σ_z의 값에 비하여 작으면 작을수록 더 좋아진다는 것을 나타내고 있다. 또한, p와 π_B의 값에도 의존하고 있으므로 궁극적으로는 제안한 양적속성의 조건부 무관질문모형의 효율성이 p와 π_B 그리고 σ_z의 값에 의존한다는 것을 보여주고 있다.

한편, 식(4.2)에서 μ_x/σ_z를 C라고 하면 다음과 같이 달리 표현할 수 있다.

$$Cp^2 + p(1-\pi_B) < 1 . \tag{4.3}$$

식(4.3)으로부터 제안한 양적속성의 조건부 무관질문모형의 효율성이 p와 π_B보다 C의 값에 많은 영향을 받고있다는 사실을 알 수 있다. 따라서, C의 값이 작을 때, 제안한 양적속성의 조건부 무관질문모형이 Greenberg 등의 양적속성의 무관질문모형보다 효율성이 증대되고, C의 값이 클 때 효율성이 떨어짐을 알 수 있다.

다음으로 C의 값이 고정되어 있다고 가정하고 π_B와 p의 크기에 따라 효율성이 어떻게 변하는지를 살펴보기 위하여 식(2.5)와 식(4.1)을 이용하여 제안한 양적속성의 조건부 무관질문모형과 Greenberg 등의 양적속성의 무관질문모형과의 효율성을 수치적으로 비교해 보고자 한다. 두 모형의 효율성을 비교하는 데 있어서 $n = 100$일 때 $\mu_x = 2$, $\sigma_z = 4(C = 1/2)$라고 가정하고, π_B를 0.3에서 0.9까지, p를 0.1에서 0.9까지 0.1간격으로 변화시켜가면서 두 모형의 분산비를 계산해 본 결과, 다음 <표 1>을 얻었다. 물론, μ_x와 σ_z가 어떻게 주어지느냐에 따라 효율성이 달라질 수 있지만 <표 1>로부터 π_B와 p의 변화에 따른 효율성의 경향을 찾을 수는 있다.

<표 1> 제안한 양적속성의 조건부 무관질문모형과 Greenberg 등의 양적속성의 무관질문모형과의 효율성 비교

p \ π_B	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	1.075	1.063	1.052	1.041	1.030	1.020	1.010
0.2	1.160	1.134	1.109	1.085	1.063	1.041	1.020
0.3	1.256	1.212	1.171	1.132	1.096	1.062	1.030
0.4	1.359	1.294	1.234	1.179	1.129	1.082	1.039
0.5	1.462	1.375	1.297	1.226	1.161	1.102	1.048
0.6	1.549	1.447	1.353	1.268	1.190	1.120	1.057
0.7	1.593	1.495	1.396	1.302	1.215	1.136	1.064
0.8	1.553	1.497	1.415	1.323	1.233	1.148	1.070
0.9	1.398	1.433	1.397	1.326	1.242	1.156	1.075

<표 1>에서 1보다 큰 값은 제안한 양적속성의 조건부 무관질문모형이 Greenberg 등의 양적속성의 무관질문모형보다 더 효율적임을 나타낸다. <표 1>로부터 p값이 클수록 π_B값이 작을수록 제안한 양적속성의 조건부 무관질문모형이 Greenberg 등의 양적속성의 무관질문모형보다 더 효율적임을 알 수 있다. 한편, π_B값이 1에 근접할수록 두 모형의 효율이 비슷하게 됨을 알 수 있었고, π_B가 1인 경우에는 덜 민감한 속성 B를 가지고 있는 응답자 모두가 Greenberg 등의 양적속성의 무관질문모형을 사용하는 결과가 되므로 제안한 양적속성의 조건부 무관질문모형이 Greenberg 등의 양적속성의 무관질문모형과 일치하게 된다.

4.2 제안한 양적속성의 이표본 조건부 무관질문모형과 Greenberg 등의 양적속성의 이표본 무관질문모형과의 비교

Greenberg 등의 양적속성의 이표본 무관질문모형에서 민감한 변수 X의 추정량 $\hat{\mu}_{g2}$의 분산은 다음과 같다.

$$Var(\hat{\mu}_{g2}) = \frac{[(1-p_2)^2\ \sigma_{z1}{}^2/n_1] + [(1-p_1)^2\ \sigma_{z2}{}^2/n_2]}{(p_1-p_2)^2}\ ,\ p_1 \neq p_2\ . \tag{4.4}$$

제안한 양적속성의 이표본 조건부 무관질문모형의 분산 식(3.3)과 Greenberg 등의 양적속성의 이표본 무관질문모형의 분산 식(4.4)에서 $Var(\hat{\mu}_{g2}) - Var(\hat{\mu}_{x2}) > 0$을 이용하여

$$\begin{aligned} Var(\hat{\mu}_{g2}) - Var(\hat{\mu}_{x2}) = & \frac{1}{n(p_1-p_2)^2}[\ (1-p_2)^2\ \sigma_{z1}{}^2[n/n_1 - \lambda\{(1-\lambda)+n\lambda\}/n_1{}'] \\ & + (1-p_1)^2\ \sigma_{z2}{}^2[n/n_2 - \lambda\{(1-\lambda)+n\lambda\}/n_2{}']] \\ & - \frac{\mu_x{}^2(1-\lambda)}{n\lambda} > 0 \end{aligned}$$

으로부터 $Var(\hat{\mu}_{x2}) < Var(\hat{\mu}_{g2})$를 만족하는 조건을 구하보면 다음과 같다.

$$\mu_x < \frac{\lambda}{(1-\lambda)(p_1-p_2)^2}[\ (1-p_2)^2\ \sigma_{z1}{}^2 A + (1-p_1)^2\ \sigma_{z2}{}^2 B]\ . \tag{4.5}$$

여기서, $A = [n/n_1 - \lambda\{(1-\lambda)+n\lambda\}/n_1{}']$, $B = [n/n_2 - \lambda\{(1-\lambda)+n\lambda\}/n_2{}']$이다.

식(4.5)에서 A와 B가 양수이므로 제안한 양적속성의 이표본 조건부 무관질문모형의 효율성은 기본적으로 μ_x의 값이 상대적으로 $\sigma_{z1}{}^2$, $\sigma_{z2}{}^2$에 비하여 작으면 작을수록 더 좋아진다는 것을 알 수 있다. 또한, λ와 p_1, p_2의 값에도 의존하는데 λ는 p와 π_B의 함수이므로 궁극적으로는 제

안한 양적속성의 이표본 조건부 무관질문모형의 효율성이 p와 π_B 그리고 p_1, p_2와 σ_{z1}^2, σ_{z2}^2에 의존한다는 것을 나타내고 있다.

이러한 사실을 구체적으로 살펴보기 위하여 식(3.3)과 식(4.4)를 이용하여 제안한 양적속성의 이표본 조건부 무관질문모형과 Greenberg 등의 양적속성의 이표본 무관질문모형과의 효율성을 수치적으로 비교해 보고자 한다. 이 때, 두 이표본 모형의 효율성을 비교하는 데 있어서 π_B와 p의 크기에 따라 효율성이 어떻게 변하는지를 살펴보기 위하여 n, n_1, n_2, p_1, p_2 및 μ_x와 σ_{z1}^2, σ_{z2}^2가 주어졌다고 가정하자. 이 때, p_1과 p_2의 선택은 Greenberg 등이 제안한 $p_1 + p_2 = 1$이 되는 조건에 따라 $p_1 = 0.7$, $p_2 = 0.3$으로 두었고, $\sigma_{z1}^2 = \sigma_{z2}^2$으로 두었다. 따라서, $n = 100$, $n_1 = 60$, $n_2 = 40$일 때, $\mu_x = 2$, $\sigma_{z1}^2 = \sigma_{z2}^2 = 4$라고 가정하고, π_B를 0.3에서 0.9까지, p를 0.1에서 0.9까지 0.1간격으로 변화시켜가면서 두 모형의 분산비를 계산해 본 결과 <표 2>를 얻었다. 이 때, $n' \sim b(n, \lambda)$이므로 p와 π_B의 변화에 따라 얻어진 λ를 이용하여 $E(n')$을 구한 후 n'의 크기로 사용하였으며, n_1'과 n_2'은 $n_1' = 0.6 \times n'$, $n_2' = 0.4 \times n'$이 되도록 배분하였다.

물론, p_1, p_2 및 μ_x와 σ_{z1}^2, σ_{z2}^2이 어떻게 선택되느냐에 따라 효율성이 달라질 수 있지만 <표 1>과 마찬가지로 <표 2>로부터 π_B와 p의 변화에 따른 효율성의 경향을 찾을 수는 있다.

<표 2> 제안한 양적속성의 이표본 조건부 무관질문모형과 Greenberg 등의 양적속성의 이표본 무관질문모형과의 효율성 비교

p \ π_B	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	1.061	1.052	1.043	1.034	1.025	1.016	1.008
0.2	1.128	1.108	1.089	1.070	1.052	1.034	1.016
0.3	1.200	1.168	1.138	1.108	1.079	1.052	1.025
0.4	1.277	1.233	1.189	1.148	1.108	1.070	1.034
0.5	1.358	1.300	1.244	1.189	1.138	1.089	1.043
0.6	1.437	1.369	1.300	1.233	1.168	1.108	1.052
0.7	1.509	1.437	1.358	1.277	1.200	1.128	1.061
0.8	1.559	1.500	1.415	1.323	1.233	1.148	1.070
0.9	1.567	1.548	1.470	1.369	1.266	1.168	1.079

<표 2>에서 1보다 큰 값은 제안한 양적속성의 이표본 조건부 무관질문모형이 Greenberg 등의 양적속성의 이표본 무관질문모형보다 더 효율적임을 나타낸다. <표 2>로부터 p값이 클수록 π_B값이 작을수록 제안한 이표본 양적속성의 조건부 무관질문모형이 Greenberg 등의 양적속성의 이

표본 무관질문모형보다 더 효율적임을 알 수 있다. 한편, π_B값이 1에 근접할수록 두 모형의 효율이 비슷하게 됨을 알 수 있었고, 이러한 결과는 <표 1>의 내용과 같다.

5. 결론

본 논문에서는 첫 번째 단계에서 덜 민감한 속성 B와 강요질문으로 구성된 확률장치를 통해 "예"라고 응답한 사람들에게만 두 번째 단계에서 Greenberg 등의 양적속성의 무관질문모형을 사용하도록 하여 민감한 변수 X에 대한 정보를 얻을 수 있는 양적속성의 조건부 무관질문모형을 제안하였다. 한편, 제안한 양적속성의 조건부 무관질문모형은 무관한 변수를 모르고 있을 때는 민감한 변수에 대한 정보를 얻기가 어려운 점이 있어 이러한 점을 해결하기 위하여 첫 번째 단계에서 "예"라고 응답한 사람들로부터 두 개의 독립표본을 추출하여 민감한 변수에 대한 정보를 얻을 수 있는 양적속성의 이표본 조건부 무관질문모형으로 확장하였다. 또한, 제안한 양적속성의 조건부 무관질문모형과 Greenberg 등의 양적속성의 무관질문모형과의 효율성을 분산측면에서 비교해 본 결과, p값이 클수록 π_B값이 작을수록 제안한 양적속성의 조건부 무관질문모형이 Greenberg 등의 양적속성의 무관질문모형보다 더 효율적임을 알 수 있었다. 한편, π_B가 1인 경우에는 덜 민감한 속성 B를 가지고 있는 응답자 모두가 Greenberg 등의 양적속성의 무관질문모형을 사용하는 결과가 되므로 제안한 양적속성의 조건부 무관질문모형이 Greenberg 등의 양적속성의 무관질문모형과 일치하게 된다는 사실을 확인할 수 있었다.

물론, 제안한 양적속성의 조건부 무관질문모형이 분산비교 측면에서 Greenberg 등의 양적속성의 무관질문모형보다 효율적이지만, 두 번의 확률장치를 사용한다는 점에서 응답자들에게 번거로움을 줄 수 있으므로 응답자들의 협조가 더 필요하다고 생각된다. 하지만, 민감한 사항과 전혀 관계가 없을지도 모르는 사람들에게 바로 확률장치를 사용하여 응답을 얻는 것보다는 조금 번거롭더라도 덜 민감한 질문을 통해 민감한 사항과 관계가 있을 만한 응답자들에게 응답을 얻는 것이 바람직하다고 생각할 수 있다.

참고문헌

[1] 류 제복, 홍 기학, 이 기성 (1993). 「확률화응답모형」, 자유아카데미, 서울.

[2] Carr, J. W. and Marascuilo, L. A. (1982). Optimal Randomized Response Models and Methods for Hypothesis Testing, *Journal of Educational Statistics*, Vol. 7, 295- 310.

[3] Loynes, R. M. (1976). Asymptotically Optimal Randomized Response Procedures, *Journal of the American Statistical Association*, Vol. 71, 924-928.

[4] Chaudhuri, A. and Mukerjee, R. (1988). *Randomized Response : Theory and Techniques*, Marcel Dekker, Inc., New York.

[5] 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*, Vol. 64, 520-539.

[6] Greenberg, B. G., Kubler, R. R., Abernally, J. R., and Horvitz, D. G. (1971). Applications of the Randomized Response Technique in Obtaining Quantitative Data, *Journal of the American Statistical Association*, 66, 243-250.

[7] Warner, S. L. (1965). Randomized Response ; A Survey Technique for Eliminating Evasive Answer Bias, *Journal of the American Statistical Association*, Vol. 60, 63-69.