

R을 활용한 조건부 가치 측정법: 정보 가치 측정 사례 연구[†]

정병준¹ · 박노진²

^{1,2}단국대학교 정보통계학과

접수 2011년 8월 31일, 수정 2011년 9월 26일, 게재확정 2011년 10월 1일

요약

정보기술의 발달은 인간에게 많은 유익한 정보를 쉽게 제공하고 있으나 그 유익한 정보를 부적절한 사용자로부터 보호할 필요가 생겼다. 정보와 관련된 위험을 분석 및 관리하는 과정에서 정보에 대한 가치를 정확히 측정해야 한다. 한 가지 방법으로 조건부 가치 측정법에 대하여 고찰하려 한다. 공공재 또는 비시장재의 가치를 평가하는 방법인 조건부 가치 측정법을 통하여 관심 대상의 가치에 대한 지불용의액을 통계적인 방법으로 추정할 수 있다. 특별히, 지불 용의액에 대한 신뢰구간의 추정에 초점을 맞추어 보았다. 정보시스템에 존재하는 정보의 가치를 보전하기 위해 지불하고자 하는 금액으로부터 정보의 가치를 통계적으로 추정하는 방법을 R을 이용한 사례를 분석하여 소개한다.

주요용어: 델타 방법론, 로지스틱 회귀분석, 부스트랩, 조건부 가치 측정법, 크린스키-로브 방법론.

1. 서론

요즘 해킹과 같은 비정상적인 정보 행위로 인해 기업들이 고객 정보를 관리하는데 비상이 걸렸다. 정보 시스템의 위험을 관리함에 있어서 고객의 정보 자산을 파악하고 가치를 평가하는 것은 매우 중요하다. 고객의 정보들은 대부분 형태를 규정하기 어려운 무형자산들이고 그 가치의 평가는 매우 주관적일 수밖에 없다. 예를 들어, 유명한 포털 서비스 기업의 이메일 서버에 이상이 생겨 고객들의 이메일이 삭제되는 일이 발생하고 보상을 해야 되는 상황이 되었다고 하자. 과연 고객들의 이메일에 대한 금전적 가치를 어떻게 측정할 것인가 하는 것이 매우 현실적인 문제이다.

환경과 같은 공공재화에 대한 가치를 측정하기 위해 경제학 분야에서는 조건부 가치 측정법 (contingent valuation method, CVM)을 사용한다. 조건부 가치 측정법은 응답자에게 어떤 가상의 상황을 제시하고 그 상황을 대비하기 위해 응답자가 지불할 용의가 있는 최대 지불 용의액 (willingness to pay, WTP)을 통계적으로 추정하는 방법이다. 예를 들어, ‘한강의 수질을 1급수로 높이기 위한 사업에 얼마 정도를 지불하실 용의가 있습니까?’라는 질문에 대한 응답자들의 응답을 통해 한강 수질에 대한 가치를 추정하는 것이다.

본 연구는 먼저 조건부 가치 측정법에 대하여 간단히 설명을 하였다. 특별히 지불 용의액의 평균과 중간값에 대한 신뢰구간을 구하는 방법에 집중하였다. 과거 연구들을 보면 신뢰구간의 측정에 대한 개별적인 의견들이 있는데 본 연구에서 그것들을 모아 정리하여 보았다. 또한 R을 사용하여 실제로 계산하는 방법을 제시하였다. 사례 연구로서 정보서비스 중에서 SNS (social network service)에서의 개인정보에 대한 존재가치와 이용가치 중에서 이용가치를 평가하는 조건부 가치 측정법을 시도하여 보았다.

[†] 본 연구는 방위사업청과 국방과학연구소의 지원으로 수행되었습니다.

¹ (448-701) 경기도 용인시 수지구 죽전동, 단국대학교 정보통계학과, 석사과정.

² 교신저자: (448-701) 경기도 용인시 수지구 죽전동, 단국대학교 정보통계학과, 교수.

E-mail: rjpak@dankook.ac.kr

2. 조건부 가치 측정법

조건부 가치 측정법은 여가, 야생, 환경과 같은 비시장재의 가치를 측정하는 표준화된 방법들 중 하나이다. 조건부 가치 측정은 어떤 재화나 서비스에 대한 가치를 미래에 제공될 편익을 직접적으로 측정하는 한 가지 방법이다 (Hanemann, 1984). 어떤 재화의 가치를 시장가격이 아닌 개인의 효용으로 파악하고 있으며 그 가치는 WTP로 표시되는 지불 용의액으로 계산된다. WTP는 시장가격과 달리 소비자의 효용함수가 변하지 않는 한 변화하지 않는 개념이다. 효용극대화를 통해 구해진 간접효용함수를 $v(\cdot)$ 라고 할 때, 주어진 소득 M 과 경제적 상태 s^0 와 s^1 에 대하여 WTP는

$$v(s^0, M) = v(s^1, M - w)$$

을 만족하는 w 로 정의된다. 예를 들어, s^0 를 현재 시스템, s^1 을 새로운 시스템이라고 정의하면, 새로운 시스템을 위해 w 를 추가로 지불하여도 현재와 같은 효용을 얻을 수 있을 때 w 가 WTP가 된다. 개인들에게 어떤 상황의 향상을 담보하기 위해 B라는 액수를 지불할 용의가 있는지를 묻는 질문들을 통해서 WTP를 추정한다. 개인의 상황을 모수 θ 로 추상화하고 약간의 불확실성 (ϵ_0, ϵ_1)을 감안하여, B라는 제안에 대하여

$$v(s^0, M; \theta) + \epsilon_0 \leq v(s^1, M - B; \theta) + \epsilon_1$$

라면 제안을 받고 부등호가 반대라면 제안을 거절한다고 할 수 있다.

이제, $\eta = \epsilon_0 - \epsilon_1$, $\Delta v \equiv v(s^1, M - B; \theta) - v(s^0, M; \theta)$ 라고 할 때,

$$\Pr(\text{yes}) = \Pr(\eta \leq \Delta v) \equiv F_\eta(\Delta v)$$

라고 할 수 있다. 물론, $\Pr(\text{no}) = 1 - \Pr(\text{yes})$ 이다.

WTP의 누적확률함수를 $G_w(\cdot)$ 로 정의하면, WTP는 제안된 B보다는 크거나 같을 것이고

$$\Pr(\text{yes}) = \Pr(w \geq B) = 1 - \Pr(w \leq B) = 1 - G_w(B)$$

라고 하겠다. 따라서 $G_w(B) = 1 - F_\eta(\Delta v)$ 라는 관계가 성립한다.

Hanemann (1984, 1989) B와 Δv 에 대하여 $\Delta v = \alpha - \beta B$ 와 같은 선형관계를 제시했다. 더욱이, Hanemann (1984, 1989)은 일반적인 로지스틱 함수

$$G(B) = [1 + \exp\{\alpha - \beta(B)\}]^{-1} \quad (2.1)$$

를 제시하였다. Bishop과 Heberlein (1979)이 처음에 $G(\cdot)$ 를 지수-로지스틱 함수,

$$G(B) = [1 + \exp\{\alpha - \beta(\ln B)\}]^{-1}$$

를 제안하기도 하였다. 현재 식 (2.1)이 일반적으로 사용되고 있다.

이제, N명의 개인들에게 각각 $B_i, i = 1, \dots, N$,라는 제안을 했다고 하자. 만일 i 번째 개인의 응답이 '네' 이면 $d_i^y = 1$, 아니면 0, 한편 응답이 '아니오' 이면 $d_i^n = 1$, 아니면 0이라고 하면, $\theta = (\alpha, \beta)$ 에 대한 로그우도함수는

$$\ln L^s(\theta) = \sum_{i=1}^N d_i^y \ln[1 - G(B_i; \theta)] + d_i^n \ln G(B_i; \theta).$$

가 된다. 최대우도 추정량은 위의 로그우도함수를 모수 θ 에 대하여 최대화하는 $\hat{\theta}$, 즉, $\partial L^s(\hat{\theta}^s)/\partial \theta = 0$ 의 해가 된다.

$\theta = (\alpha, \beta)$ 에 대한 추정량이 구해지면 WTP의 기댓값은

$$C^+ = \int_0^{\infty} [1 - G(B)]dB \quad (2.2)$$

가 된다 (Hanemann, 1984, 1989). 이는 확률변수 B의 기댓값을 구하는 식

$$E[B] = \int_0^{\infty} [1 - G(B)]dB - \int_{-\infty}^0 G(B)dB$$

에서 B가 음수가 되는 구간의 적분을 제외한 결과이다. $G(B; \theta)$ 를 식 (2.1)과 같이 정의하면 WTP 평균은

$$C^+ = (1/\beta) \ln[1 + \exp(\alpha)]$$

이 되고 WTP의 중간값은 $G(B; \theta) = 1/2$ 이 되는

$$C^+ = \alpha/\beta$$

가 된다.

3. 조사 수행 방법

WTP을 추정하기 위한 질문 방법은 개방형질문법과 폐쇄형질문법으로 크게 나눌 수 있다 (박주현, 2000). 개방형질문법은 응답자에게 액수를 직접 기입하도록 하는 방법이다. 하지만, 응답자가 상황을 완전히 이해하지 못했다면 합리적 액수를 기대할 수 없고 무응답이 많은 경우가 발생한다. 폐쇄형질문법은 개방형질문법과 달리 응답에 필요한 보조 도구를 제공하거나 액수의 범주를 제공하여 응답을 유도하는 방법이다 (표희동 등, 2001; 도건우와 곽승준, 2004; 이주석 등, 2007). 폐쇄형 질문법은 방법에 따라 경매법, 지불카드법, 양분선택형법 등이 있다. 경매법은 응답자에게 초기 액수를 제시하고 물건을 경매하듯이 지불 의사액을 올려 가면서 액수를 정하는 방법이다. 지불카드법은 일련의 숫자가 적힌 카드를 응답자에게 제시하고 그 중에 적당한 액수가 적힌 카드를 선택하도록 하는 방법이다. 한편, 양분선택법은 사전에 준비된 액수를 제시하여 주어진 액수 이상을 지불할 용의가 있는지에 대해 ‘네’ 혹은 ‘아니요’로 답을 유도한다.

현재 많이 사용되는 방법은 먼저 양분선택을 위한 제시 액수를 사전 연구를 통해 그 범위와 범주를 미리 정하고 대상자들을 유사한 비율의 집단으로 나누고 각 집단에 무작위로 한 가지 액수를 제시하여 ‘네’ 또는 ‘아니요’로 응답하게 하고 있다. 그러나 실제 상황에서는 정해진 액수들이 기록된 카드를 준비하고 대상자들이 카드 하나를 무작위로 뽑아 주어진 액수에 대한 ‘네’ 또는 ‘아니요’로 응답하도록 하는 방법이 사용되고 있다 (이주석 등, 2007).

4. 신뢰구간 계산법

WTP의 정확한 신뢰구간을 구하기 위해

$$\text{var}(B) = \int_0^{\infty} 2B[1 - G(B) + G(-B)]dB - \mu_B^2$$

을 구해야하나 적분이 용의하지 않다. 따라서 간접적인 방법을 통해 분산을 구하고 신뢰구간을 계산하게 된다. WTP에 대한 신뢰구간을 구하는 방법은 1) 델타방법 2) Krinsky-Robb 방법 그리고 3) 부스트랩 방법이 있다. 지금까지 위의 세 가지 방법은 연구자에 따라 개별적으로 사용되어 왔다. 본 연구에서는 세 가지 방법을 간단하게 설명하고 예제를 통해 모두 사용하여 보겠다.

4.1. 델타방법

델타방법은 두 개 이상의 확률변수의 비선형함수 형태로 이루어진 통계량의 분산을 계산할 때 사용하는 방법으로 테일러급수의 1차 전개에 기반을 두어 구하게 된다 (Green, 2003). 먼저 \widehat{WTP} 는 WTP 의 추정량, WTP_α 와 WTP_β 는 WTP 의 α 와 β 에 대한 편미분을 의미한다고 하자. WTP 의 중간값에 대한 분산은

$$\begin{aligned} var(\widehat{WTP}) &= (\widehat{WTP}_\alpha)^2 var(\hat{\alpha}) + (\widehat{WTP}_\beta)^2 var(\hat{\beta}) + 2\widehat{WTP}_\alpha \widehat{WTP}_\beta cov(\hat{\alpha}, \hat{\beta}) \\ &= (1/\hat{\beta})^2 var(\hat{\alpha}) + (-\hat{\alpha}/\hat{\beta}^2)^2 var(\hat{\beta}) + 2(1/\hat{\beta})(-\hat{\alpha}/\hat{\beta}^2) cov(\hat{\alpha}, \hat{\beta}) \end{aligned}$$

가 된다. 한편, WTP 의 평균에 대한 분산은

$$\begin{aligned} var(\widehat{WTP}) &= (\widehat{WTP}_\alpha)^2 var(\hat{\alpha}) + (\widehat{WTP}_\beta)^2 var(\hat{\beta}) + 2\widehat{WTP}_\alpha \widehat{WTP}_\beta cov(\hat{\alpha}, \hat{\beta}) \\ &= (e^{\hat{\alpha}}/\hat{\beta}(1+e^{\hat{\alpha}}))^2 var(\hat{\alpha}) + (-\log(1+e^{\hat{\alpha}})/\hat{\beta}^2)^2 var(\hat{\beta}) + \\ &\quad 2(e^{\hat{\alpha}}/\hat{\beta}(1+e^{\hat{\alpha}}))(-\log(1+e^{\hat{\alpha}})/\hat{\beta}^2) cov(\hat{\alpha}, \hat{\beta}) \end{aligned}$$

가 된다. 따라서 $100(1-\alpha)\%$ 신뢰구간은

$$\widehat{WTP} \pm z_{\alpha/2} \sqrt{var(\widehat{WTP})}$$

가 되고 여기서 $z_{\alpha/2} = \Phi^{-1}(1 - \alpha/2)$, $\Phi(\cdot)$ 는 표준정규분포의 누적확률함수이다.

4.2. Krinsky-Robb 방법

Krinsky와 Robb (1986)는 델타방법의 대체 방법으로 다변량 정규분포에 기반을 두는 WTP 의 신뢰구간을 제시하였다. 방법은 먼저 계수 α, β 에 대한 평균과 분산-공분산 행렬을 추정하여 결합 정규 분포를 확정한다. 그 다음 결합 정규 분포에서 (α, β) 에 대한 모의 표본을 추출하여 WTP 를 계산하고 예를 들어 2.5와 97.5백분위수를 WTP 에 대한 95% 신뢰구간으로 정한다.

4.3. 부스트랩 방법

조사에서 확보한 표본을 모집단이라고 가정하고 다시 표본을 일정량 비복원 추출하여 WTP 을 계산한다. 계산된 WTP 의 분포로부터 예를 들어 2.5와 97.5백분위수를 WTP 에 대한 95% 신뢰구간으로 정한다 (Efron과 Tibshirani, 1993; Chung과 Han, 2009).

5. 조사와 분석

최근에 유명 SNS인 사이월드와 네이트가 해킹을 당하는 사건이 벌어졌다. 그 피해에 대한 손해 배상이 큰 이슈가 되고 있다. 따라서 과연 SNS가 포함하고 있는 정보의 가치가 얼마인가는 매우 현실적인 문제가 되었다. 본 연구에서는 그 가치를 앞서 설명한 조건부 가치 측정법을 통해 계산해 내는 간단한 조사를 예제로 들어 보았다. 가치의 유형에는 이용가치와 비이용가치가 있으며 조건부 가치 평가법은 이들 두 유형의 가치의 합인 총 가치를 측정할 수 있다. 본 연구는 이용가치에만 초점을 맞추어, 설문에서 SNS에 보관된 개인정보에 대한 이용가치에 대하여 질문하였다.

본 연구의 설문 조사는 2011년 5월 16일부터 30일까지 경기도 소재 D대학 학생들을 대상으로 실시하였고 총 375부의 응답을 사용하였다. 본 연구는 조건부 가치 평가법에 R을 활용하는 사례의 소개에 주된 연구목적이 있기 때문에 조사를 비교적 간편하게 수행할 수 있는 대학생에게 설문조사를 실시하였다. 그러나 조건부 가치 평가법을 엄밀히 수행하기 위해서는 경제활동을 하고 있으면서 정규 소득이 있는 응답자들로 한정해야 한다는 점을 주의할 필요가 있다.

사전 조사를 통하여 제시금액을 6000원부터 12000원까지 1000원 단위로 7가지 범주로 구분하여 범주에 따라 설문지를 만들고 대상자들에게 무작위로 설문지를 택하여 응답하도록 하였다. 설문지에는 성별, 학년, 연령, 소득, 인터넷 사용시간, SNS 사용시간, SNS 친구수, 가상공간 친구수, SNS 친구수에 대한 가상공간 친구수의 비율과 같은 인구통계학적 자료도 함께 수집하였다. 아래와 같은 질문을 통해 월 보험료에 대한 지불 용의액을 추정하였다. 사용된 변수들을 표 5.1에 정리하였다.

귀하께서 사용중인 소셜 네트워크 서비스의 해킹 등을 통해 개인정보가 유출되는 등의 피해를 보는 경우 이를 보상해주는 보험이 있다면, 귀하께서는 이 보험의 월 보험료로 *****원을 지불할 용의가 있습니까?

① 예

② 아니오

표 5.1 변수 설명

변수명 id	변수 유형	설명 고유번호	비고
Q1_1	범주형	사용하는 SNS1	1: 싸이월드 2: 페이스북 3: 트위터 4: 미투데이 5: 기타
Q1_2	범주형	사용하는 SNS2	상동
Q1_3	범주형	사용하는 SNS3	상동
Bid	연속형	제시금액	1,000원 12,000원 (1000원 단위로 12개 제시금액)
Response	범주형	응답, 지불의사	0: 아니오, 1: 예
Gender	범주형	성별	1: 남자, 2: 여자
Grade	범주형	학년	1: 1학년 2: 2학년 3: 3학년 4: 4학년
Age	연속형	연령	
Income	연속형	소득	단위: 만원 (용돈, 아르바이트 등 모두 포함) 하루에 1: 1시간 미만 2: 1시간 이상 2시간 미만 3: 2시간 이상 3시간 미만 4: 3시간 이상 4시간 미만 5: 4시간 이상
IntUse	범주형	인터넷 사용시간	1: 한 주 이상의 기간에 1회 방문 2: 한 주 동안 1 6회 (하루에 1회 미만) 방문 3: 매일 1회 정도 방문 4: 매일 2회 정도 방문 5: 매일 3회 이상 방문
SNSFreq	범주형	SNS사용빈도	1: 한 주 이상의 기간에 1회 방문 2: 한 주 동안 1 6회 (하루에 1회 미만) 방문 3: 매일 1회 정도 방문 4: 매일 2회 정도 방문 5: 매일 3회 이상 방문
Friend	연속형	SNS상의 친구의 수	
WebFrnd	연속형	웹상으로만 알고 있는 친구의 수	SNS상의 친구 중 웹상으로만 알고 있는 친구의 수 (실제로는 만난 적이 없는 친구)
WebPrct	연속형	친구 중 사이버 상으로만 알고 있는 친구의 비율	WebPrct = WebFrnd / Friend

5.1. 지불 용의액 계산을 위한 R-프로그램

기존의 다양한 통계소프트웨어 (Kim 등, 2004)를 이용하여 지불 용의액 계산이 가능하나 범용 소프트웨어인 R을 이용하여 계산하는 방법을 소개하겠다. 프로그램시 주의할 점은 R에서의 모형은 ' $\alpha + \beta \times$ 변수명'을 기본 형식으로 하나 지불 용의액 경우 모형이 ' $\alpha - \beta \times$ 변수명'임을 유념해야 한다.

0. 명령문 `glm(Response Bid, family=binomial)`과 하위 속성 `cov.unscale`을 사용하여 α 와 β 에 대한 추정치와 분산공분산 행렬 추정치를 구할 수 있다. 추정치들을 a 와 b , 분산을 va 와 vb 그리고 공분산을 vab 라고 하자.

1. 델타 방법에 의한 프로그램은 4.1에 제시된 공식을 사용하여

중간값의 분산: $(1/b)^2 * va + (-a/b^2)^2 * vb + 2*(1/b)*(-a/b^2)*vab$

평균의 분산: $(\exp(a)/(b*(1+\exp(a))))^2 * va + (-\log(1+\exp(a))/b^2)^2 * vb + 2*(\exp(a)/(b*(1+\exp(a))))*(-\log(1+\exp(a))/b^2)*vab$

을 계산한다. 신뢰구간은 위에서 구한 분산과 추정치들을 이용하여 4.1에 제시된 식을 이용하여 구한다.

2. Krinsky-Robb의 방법을 사용하기 위해서는 MABVAR이라는 library를 사용하여 α 와 β 에 대한 모의실험 값을 구하여 신뢰구간을 계산한다.

```
library(MSBVAR)
wtpmm <- NULL
for (i in 1:1000)
c <- rmultnorm(1,matrix(c(a,b),2,1), vmat=matrix(c(va,vab,vab,vb),2,2))
alpha <- c[1]
beta <- -c[2]
wtp <- (1/beta)*log(1+exp(alpha)) (평균)
wtp <- alpha/beta (중간값)
wtpmm[i] <- wtp
c(quantile(wtpmm,.025),quantile(wtpmm,.975))
```

3. 부스트랩 방법은 주어진 데이터 (bid)에서 부표본을 모의 추출하여 신뢰구간을 계산한다.

```
wtpm <- NULL
for (i in 1:1000)
bid1 <- bid[sample(1:375,100),]
attach(bid1)
alpha <- glm(Response Bid, family=binomial)coefficients[[1]]
beta <- -glm(Response Bid, family=binomial)coefficients[[2]]
wtp <- (1/beta)*log(1+exp(alpha)) (평균)
wtp <- alpha/beta (중간값)
wtpm[i] <- wtp
c(quantile(wtpm,.025),quantile(wtpm,.975))
```

5.2. 결과 분석

표 5.2를 보면 대체로 비슷한 비율로 범주들이 배당되었고 제시금액이 낮을수록 받아드리는 경향이 높고 반대로 제시금액이 높을수록 거절하는 경향이 높음을 알 수 있다. 제시금액이 4,000원이 되면서 거절과 수락의 비율이 역전되는 것을 볼 수 있다. 최대 지불 용의액이 4,000원 근방에서 형성될 것을 예측할 수 있다.

표 5.2 제시금액 * 응답 교차표

	응답		
	아니오(%)	네(%)	빈도(%)
1000	12.5	87.5	8.5
2000	35.5	64.5	8.3
3000	50.0	50.0	8.5
4000	61.3	38.7	8.3
5000	71.9	28.1	8.5
6000	71.0	29.0	8.3
7000	75.0	25.0	8.5
8000	79.3	20.7	7.8
9000	81.3	18.8	8.5
10000	86.7	13.3	8.0
11000	87.5	12.5	8.5
12000	90.3	9.7	8.3
합계	66.7	33.3	100

분석을 위해 사용한 모델은 지불 제안에 ‘예’라고 응답한 비율을 p 라고 하고 범주형 변수들은 해당하는 범주만큼 가변수를 만들어서 아래와 같은 로지스틱 회귀모형을 사용하였다.

$$\begin{aligned} \text{logit}(p) = & \alpha + \beta \text{bid} + \beta_{\text{gender}} \text{gender} + \sum_{i=1}^4 \beta_{\text{grade}_i} \text{grade}_i + \beta_{\text{income}} \text{income} \\ & + \sum_{i=1}^5 \beta_{\text{intuse}_i} \text{intuse}_i + \sum_{i=1}^5 \beta_{\text{snsfreq}_i} \text{snsfreq}_i + \beta_{\text{friend}} \text{friend} + \beta_{\text{webfrnd}} \text{webfrnd} \\ & + \beta_{\text{webprcnt}} \text{webprcnt} + \text{error} \end{aligned}$$

로지스틱회귀분석 (강명욱 등, 2010; Kahng, 2011)을 실시하면 아래 표 5.3과 같은 결과를 얻게 된다. 많은 변수들 중에서 Bid (제시액), Income (소득), Friend (친구수)와 WebPrnt (사이버 친구 비율)가 유의한 것으로 나타났다. 특히, 사이버 친구 비율의 승산비가 13.9로 매우 높게 계산되었는데, 이는 가상환경 속에만 보관되어 있는 친구들에 대한 정보의 손실에 대하여 응답자들이 매우 우려하고 있으며 따라서 사이버 친구수의 비율이 높을수록 제시된 보험료를 받아드릴 비율이 높아진다고 하겠다.

표 5.3 추정 통계량

	추정치	표준 오차	p-값	승산비
Bid	.000	.000	.000	1.000
Income	.057	.009	.000	1.058
Friend	.004	.001	.002	1.004
WebPrnt	2.633	1.036	.011	13.919
constant	-1.348	.404	.001	.260

지불 용의액에 따라 응답자를 나누고 로지스틱 모형을 추정하여 WTP의 평균과 중간값을 계산하여 그림 5.1를 그려 보았다. WTP의 평균은 제시액의 크기에 따라 증가하고 있으나 중간값은 제시액이 최대 10,000원일 때를 정점으로 감소함을 알 수 있다. 극단적이거나 비정상적인 응답이 있을 수 있다는 가정을 한다면 합리적인 최대 제시액을 10,000원으로 보는 것도 타당하다고 하겠다.

실제로 매우 중요한 입력변수로 판명된 사이버 친구의 비율을 다섯 범주로 나누어 WTP의 평균과 중간값을 계산하여 그림 5.2를 그려보았다. 사이버 친구의 비율이 10% 30%에 속한 응답자들의 WTP

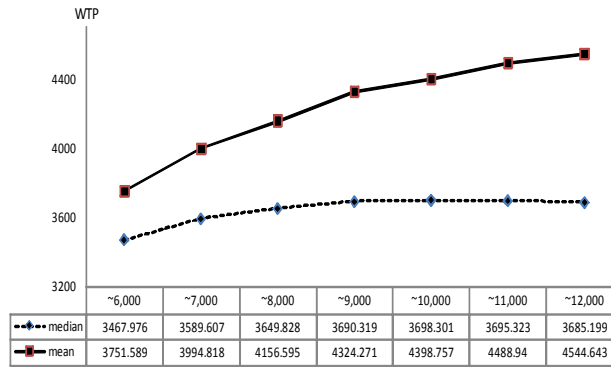


그림 5.1 최대 제시 금액에 따른 WTP

(평균, 중간값)가 가장 높은 것으로 나타났다. 이 범주에 속한 응답자들이 정보 손실에 대해 가장 심각하게 생각하고 있다고 할 수 있겠다.

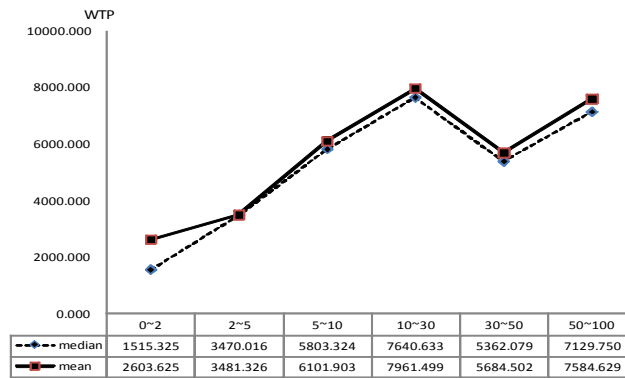


그림 5.2 사이버 친구 비율에 따른 WTP

전체 응답자들의 자료를 이용하여 제시금액을 입력변수로 하는 모델에 대한 추정값에 의하면 (지불용의한) 보험료의 평균은 4,787원, 중간값은 3,881원으로 계산되었다. (표 5.4). 신뢰구간을 고려하면 적게는 1,974원 많게는 5,686원 정도로 예상할 수 있다.

조사 결과에 따라 매달 보험료 4,787원씩 12개월 납입하고 연이율 3.25 (기준금리)에서 월복리로 계산하면 일 년간 총 보험료의 현금 가치는 58,465원이 된다. 즉, 연간 지불용의액이 평균 58,465원이라고 하겠다. 실제로, 보험료는 다음과 같은 다소 복잡한 식 ‘납입할 보험료 = 기본보험료×특약요율×가입자특성요율×특별요율×할인할증요율’을 통해 구해진다. 또한, 사고빈도와 사고금액의 크기 등을 고려해야 한다 (김명준과 김영화, 2009; 현정민과 차지환, 2010). 하지만, 본 논문에서는 간단히 지불용의액이 현재가치에 비례한다고 하고 요율을 2%로 가정하면 2,923,250원이 본 조사에 의한 SNS에 대한 연간 평균 정보 가치라고 하겠다. 비슷한 방법으로 중간값에 의한 정보 가치는 2,370,000원으로 계산된다. 신뢰구간 (평균)에 입각하여 계산한다면 SNS에 대한 정보 가치는 최소 2,218,550원에서 최대

표 5.4 관련 추정치

추정량	추정치		
$\hat{\alpha}$	1.1645		
$\hat{\beta}$.0003		
WTP	평균	중간값	
	4787.042	3881.666	
95%-CI	delta	(3954.803, 5134.483)	(2682.007, 4484.111)
	K-R	(3949.385, 5192.357)	(2946.566, 4676.766)
	bootstrap	(3633.134, 5686.633)	(1947.321, 4927, 211)

3,472,250원이라고 하겠다. 중간값의 신뢰구간을 고려한다면 최소 1,188,950원에서 최대 3,008,750원이라고 하겠다 (표 5.5).

표 5.5 정보 가치 통계량

WTP	기대값	하한	상한
평균	2,923,250	2,218,550	3,472,250
중간값	2,370,000	1,188,950	3,008,750

6. 결론

정보 자산의 가치를 측정하는 것은 오늘날 매우 중요한 과제이다. 조건부 가치 측정법에 대한 통계적 방법론을 정리하여 보았고 SNS에 관한 간단한 조사를 통해 구현을 하여 보았다. 그 과정에서 누구나 사용할 수 있는 R을 사용하여 계산하는 방법을 제시하였다. 지금 현재 조건부 가치 측정을 위한 사용되는 프로그램들이 대부분 상업용인 반면 R은 범용으로 누구나 사용할 수 있는 상황에서 손쉽게 조건부 가치 측정을 수행할 수 있으리라 생각된다. 하지만 실제로 사용된 사례의 결과를 일반화하기에는 모집단의 선정과 표본의 추출 그리고 설문 자체 등에 부족한 점이 있음이 연구의 한계라고 하겠다.

참고문헌

- 강명욱, 김부용, 홍주희 (2010). 로지스틱모형에서 그래픽을 이용한 회귀와 모형평가. <한국데이터정보과학회지>, **21**, 21-32.
- 김명준, 김영화 (2009). 다양한 모형을 통한 자동차 보험가격 산출. <한국데이터정보과학회지>, **20**, 515-526.
- 도건우, 곽승준 (2004). 조건부가치측정법을 이용한 환경마크 인증제품의 지불의사 분석. <재정정책논집>, **6**, 3-20.
- 박주현 (2000). <환경경제학>, 경문사, 서울.
- 이주석, 유승훈, 곽승준 (2007). 낙동강 수질개선의 편익추정-1.5경계 양분선택형 조건부 가치 측정법을 이용하여. <경제연구>, **25**, 111-129.
- 표희동, 유승훈, 곽승준 (2001). 이중경계 양자택일형의 조건부 가치측정법을 이용한 영산강유역 갯벌의 보존가치 추정. <지역연구>, **17**, 37-54.
- 현정민, 차지환 (2010). 중신보험에서의 영향 변수의 영향력 분석에 관한 연구. <한국데이터정보과학회지>, **21**, 71-86.
- Bishop, R. and Heberlein, T. (1979). Measuring values of extra-market goods: Are indirect measures biased? *American Journal of Agricultural Economics*, **61**, 926-30.
- Chung, H. and Han, C. (2009). Bootstrap confidence intervals for classification error rate in circular models when a block of observations is missing. *Journal of the Korean Date & Information Science Society*, **20**, 757-764.

- Efron, B. and Tibshirani, R. (1993). *An introduction to the bootstrap*, Chapman and Hall, New York.
- Green, W. H. (2003). *Econometric analysis*, 5th ed., Prentice Hall, Englewood Cliffs.
- Hanemann, M. (1984). Welfare evaluation in contingent valuation experiment with discrete responses. *American Journal of Agricultural Economics*, **66**, 332-41.
- Hanemann, M. (1989). Welfare evaluation in contingent valuation experiment with discrete responses data: Reply. *American Journal of Agricultural Economics*, **71**, 1057-1061.
- Kahng, M. W. (2011). A study on log-density ratio in logistic regression model for binary data. *Journal of the Korean Date & Information Science Society*, **22**, 107-113.
- Kim, S., Jeong, D. and Park, Y. (2004). Comparative study on statistical packages for analyzing logistic regression-Minitab, SAS, SPSS, STATA-. *Journal of the Korean Date & Information Science Society*, **15**, 367-378
- Krinsky, I. and Robb, A. L. (1986). On approximating the statistical properties of elasticities. *Review of Economics and Statistics*, **68**, 715-719.

Contingent valuation method implemented by R: Case study - measuring value of information[†]

Byung Joon Jung¹ · Ro Jin Pak²

¹²Department of Information & Statistics, Dankook University

Received 31 August 2011, revised 26 September 2011, accepted 1 October 2011

Abstract

The development of information technology provides us with more useful information but it arose to protect such information from inappropriate users. In the course of analyzing and managing the risks associated with information, it should be needed to accurately measure the value of information. We try to consider the contingent valuation method for this purpose. The contingent valuation method which is used to assess the value of public goods or nonmarket goods makes an statistical estimation for the willingness-to-pay. We show with an example how we can estimate the value of information by calculating the amount we are willing to pay the value of information that exists on the information system. Calculation is carried out by using R.

Keywords: Bootstrap, contingent valuation method, delta method, Krinsky-Robb method, logistic regression.

[†] This work was partially supported by Defense Acquisition Program Administration and Agency for Defense Development under the contract.

¹ Graduate student, Department of Information & Statistics, Dankook University, Yongin, Gyonggi 448-701, Korea.

² Corresponding author: Professor, Department of Information & Statistics, Dankook University, Yongin, Gyonggi 448-701, Korea. E-mail: rjpak@dankook.ac.kr