

발 간 등 록 번 호

11-1240000-000448-10

마이크로자료의 활용도제고를 위한  
**비 밀 보 호 방 법**

2006. 12.



통계개발원

---

## 제 출 문

---

본 보고서를 2006년도 연구과제인 「마이크로자료의 활용도제고를 위한 비밀보호방법」의 연구결과 보고서로 제출합니다.

2006년 12월

연구기획실장 최 성 욱

---

---

연구자 : 정 동명, 강 동환

# 목 차

I. 서론 .....	1
II. 비밀보호방법 .....	3
1. 노출과 식별 .....	3
2. 비밀보호방법 .....	4
3. Uniqueness .....	7
III. 마이크로자료의 Masking .....	10
1. 분석대상 자료 .....	10
2. 마이크로자료 파일의 작성 .....	12
IV. 결론 .....	23
참고문헌 .....	25
부록 .....	27
A. 2005 인구주택총조사 결과(충청남도) .....	27
B. Key 변수의 Entropy .....	30
C. Key 변수의 Uniqueness .....	38

# ·I 서 론

대부분의 통계작성기관에서는 각종 통계조사를 통해 다양한 자료들을 수집하고 있는데, 이렇게 수집된 통계자료는 원자료(raw data)와 마이크로자료(microdata), 그리고 매크로자료(macrodatta)로 구분할 수 있다. 원자료란 통계조사 자료에서 최초 입력한 전산화일 자료로서 입력오류, 조사오류 등이 걸러지기 이전 단계의 자료를 말한다. 마이크로자료란 원자료에서 입력오류 등을 제거하여 공표 통계표 작성 등 자료가공의 기초자료로 사용되는 자료를 말하며 통계원시자료라고도 한다. 매크로자료란 마이크로자료를 임의의 기준에 따라 집계한 자료로서, 집계의 정도에 따라 세분화된 자료에서 통합된 자료까지 다양하게 제공할 수 있다. 따라서 통계작성기관에서 자료를 제공할 경우에는 주로 매크로자료를 제공하고 있으며, 일부 제한적인 경우에만 마이크로자료를 제공하고 있다. 그러나 매크로자료를 이용해서 갈수록 복잡하고 다양해지는 경제·사회현상을 심층적으로 분석하기에는 어느 정도 한계가 있다. 그러므로 여러 분야의 통계이용자들은 통계를 작성하는 기관에서 매크로자료뿐만 아니라 마이크로자료도 제공할 것을 요구하고 있으며, 이러한 요구는 점차 증가하고 있는 추세이다.

마이크로자료에는 응답자의 개인 사생활에 대한 정보들이 모두 수록되어 있기 때문에 마이크로자료를 그대로 제공할 경우 응답한 개인의 특성을 나타내는 정보들이 그대로 노출될 뿐만 아니라, 심지어는 특정 응답자에 대한 식별도 가능하게 되어 심각한 개인 사생활의 침해가 발생할 위험이 있다. 그러므로 마이크로자료를 제공할 경우에는 사전에 미리 응답자의 개인정보가 노출되지 않도록 통계적으로 보호방법을 적용하는 등 각별한 노력을 기울여야만 한다.

외국의 통계작성기관에서는 1970년대부터 마이크로자료의 제공에 대한 요구와 개인정보의 비밀보호라는 두 가지 상반된 요인을 동시에 고려하여 적절한 수준의 자료를 제공할 수 있는 방법을 찾기 위해 노력해 왔으며, 1980

년대에는 이에 대한 다양한 연구논문들이 발표되었다. 미국 상무부 센서서국에서는 마이크로자료를 제공할 때 통계적 비밀보호방법을 활용하여 개인의 정보를 최대한 보호할 있는 방법들을 연구하였는데, 특히 Cox 등(1986)은 미국의 개인비밀보호와 관련된 자료제공에 대해 언급하였으며, Kim(1986)은 통계적 기법을 적용한 비밀보호방법을 연구하였다. 네덜란드 중앙통계국(1987)에서는 스웨덴과 네덜란드간의 협력세미나를 통해 개인사생활의 개방과 보호에 대해 언급하였다. Bethlehem 등(1990)은 자료파일에서 개인의 식별(identification)과 노출(disclosure)의 문제에 대해 언급하고, 이를 해결하기 위한 방법으로 모집단 uniqueness의 추정을 위한 통계적 모형을 설정하고 예제를 통해 제시하였다. Marsh 등(1991)은 영국의 1991년 센서스에서 개인 비밀보호를 위해 익명화된 마이크로자료 파일작성의 필요성과 그 효과 등을 설명하였다. 이외에도 Dalenius(1977), Fuller(1993) 등 여러 학자들이 노출의 위험성과 비밀보호에 대한 방법을 연구하였다.

우리나라의 경우 2006년 10월 1일 현재 통계청을 포함해서 153개의 통계작성기관에서 673종의 국가통계가 작성되고 있다. 이 중 통계청에서 실시하는 인구주택총조사(census)는 0과 5로 끝나는 연도를 기준으로 매 5년마다 전국을 대상으로 실시하고 있다. 인구주택총조사에서 수집된 결과에는 우리나라의 인구·사회현상을 심층적으로 파악하고 분석할 수 있는 좋은 정보들이 많이 있어, 학계나 연구기관 등 여러 분야의 통계이용자들이 마이크로자료의 제공을 요구하고 있는 실정이다. 이에 부응코자 통계청에서는 자료이용자가 인터넷을 통해 시스템에 접속하여 원하는 자료를 직접 가공 및 이용할 수 있도록 지원하는 one-stop 서비스 시스템인 MDSS(Micro Data Service System)를 2006년부터 운영하고 있다(<http://mdss.nso.go.kr>). MDSS는 기존의 주문형 위탁처리 방식에서 탈피하여 마이크로자료를 이용하는 이용자가 직접 자료를 가공할 수 있도록 지원함으로써 보다 빠르고 간편한 절차에 따라 자료를 제공받을 수 있도록 하고 있다. 가령, 자료이용자가 인구주택총조사의 마이크로자료를 원할 경우 MDSS에 접속하면 10% 표본조사 자료 중 2%의 표본자료를 일정한 절차에 따라 제공받을 수 있다.

그러나 개인의 정보가 식별되고 노출되는 민감한 항목들은 일부 제거한 후 제공하고 있기 때문에 자료의 분석 및 활용측면에서 다소 미흡한 점이 있는 것이 사실이다. 이런 점을 고려하여 통계청에서는 개인의 비밀을 보호하면서 마이크로자료를 제공할 수 있는 방법을 개발코자 수년전부터 통계적 보호기법에 대한 연구에 노력하였다. 통계청 통계연구과에서 박원환 등(2004)이 통계자료의 비밀보호에 대해 연구하였다.

본 논문에서는 개인의 비밀을 보호하는 방법들을 살펴보고, 이러한 방법을 적용하여 mask된 마이크로자료 파일을 직접 작성하고자 한다. 2장에서는 노출과 식별에 대해 알아보고, 식별의 측도로 사용되는 uniqueness의 개념을 설명한다. 또한 지금까지 연구되어진 비밀보호의 방법을 이산형(discrete) 변수인 경우와 연속형(continuous) 변수인 경우로 나누어 소개한다. 3장에서는 2005 인구주택총조사 일부 지역의 표본조사 결과자료에 대해 grouping 방법을 적용하여 마이크로 자료파일을 작성해 보고, 이에 대한 내용을 살펴보고자 한다. 마지막으로 4장에서는 본 연구의 최종적인 결론과 이를 토대로 앞으로 고려해야 할 사항들을 언급하고 있다.

## . 비밀보호방법

### II

#### 1. 노출과 식별

노출(disclosure)이란 통계작성기관이 자료를 수집·정리하여 다양한 형태의 통계정보(통계간행물, 테이프, 디스크 등)로 제공할 경우, 이를 통해서 응답자의 특성이 파악되는 경우를 말한다. 따라서 노출에 대한 문제는 통계작성기관이 통계목적으로 응답자에게 자료를 수집한다는 가정하에서 발생하게 되는데, 어떤 경우에도 노출은 발생되지 않도록 하는 것이 바람직하다. 이에 대해 Bethlehem 등(1990)은 노출이 바람직하지 않은 것은 개인의 사생활이 보호되어야 하는 법적이고 윤리적인 이유뿐만 아니라 조사의 응답률에 영향

을 줄 수 있다는 실질적인 이유가 있기 때문이며, 가능하면 마이크로자료에서 노출을 제한해야 한다고 하였다. 노출은 개인의 식별(identification)과 서로 깊은 연관성을 가지고 있다. 즉, 개인의 민감한 정보가 노출된다면 이는 개인에 대한 식별이 가능하게 되고, 반대로 주어진 통계정보에서 개인의 식별이 가능하다면 노출의 가능성도 더 커지게 된다.

노출을 제한하는 가장 기본적인 방법인 간단한 방법은 많은 마이크로자료가 공표되지 않도록 하는 것이지만, 이는 자료수집의 목적에 위배되는 것이며 현실적으로도 마이크로자료의 제공에 대한 요구가 날로 증가하는 추세이므로 이 방법은 적절하지 않다. 노출과 식별에 영향을 주는 핵심요소는 자료이용자들의 사전 정보이다. 즉, 만약 누군가가 특정 개인에 대한 정보를 알 수 없다면 노출과 식별은 불가능하다. 따라서 노출은 자료이용자들이 사전에 가지고 있는 유용한 정보의 성격과 양에 따라 좌우된다고 할 수 있다. 그러므로 노출을 제한하기 위해서는 제공되는 마이크로자료의 정보와 자료이용자들이 가지고 있는 사전정보가 서로 일치하지 않도록 자료파일을 구성하면 될 것이다.

## 2. 비밀보호방법

마이크로자료에서 개인의 비밀을 보호하는 방법은 자료의 형태가 이산형인 경우와 연속형인 경우에 따라 달리 적용된다. 이산형 자료인 경우에는 Data swapping, Coding approach, Grouping 등의 방법이 널리 활용되고 있고, 연속형 자료인 경우에는 Rounding, Grouping into Intervals, Micro aggregation 등의 방법이 활용되고 있다.

이산형 변수의 경우 성별 등과 같이 단 2개의 범주로 구분되는 변수를 Dichotomous variable이라 하고, 교육정도나 혼인상태 등과 같이 3개 이상의

범주로 구분되는 변수를 Polychotomous variable이라 하는데, 이들 특성에 따라 다음과 같은 방법들이 있다.

### 1) Data Swapping

Data Swapping은 Dalenius(1979)가 제안한 방법으로서, 마이크로자료내의 민감한 항목에 대한 자료노출을 방지하기 위하여 동일한 Key 변수 조합을 갖는 record간에 자료값을 상호 교환하는 방법이다. 가령, HIV 상태 등과 같은 민감한 항목에 대하여 인종이나 성별, 나이, 수입 등과 같은 항목을 Key 변수로 사용하여 값을 적절하게 교환하는 것이다. 이 방법은 masking 후에도 각 Key 변수별로 민감한 항목의 빈도수는 swapping하기 전의 빈도수와 동일하게 되어 자료의 분석이 용이하고, Dichotomous variable과 Polychotomous variable인 경우에도 모두 적용이 가능하다.

### 2) Coding Approach

이 방법은 원 자료에 적절한 값(noise)을 추가하여 다른 자료값으로 변형하는 방법이다. 가령,  $X_i$  와  $Y_i$  , 그리고  $e_i$  를 0 또는 1을 갖는 Dichotomous variable라고 가정하자. 만약  $X_i$  를 원자료,  $Y_i$  를 와 서로 독립이고 동일한 분포를 갖는 noise, 그리고  $e_i$  를 mask된 자료라고 한다면,  $Y_i = (X_i + e_i)_{\text{mod}2}$  는 다음과 같이 나타낼 수 있으며,

$Y_i$

$X_i$	$e_i$	$Y_i$
1	1	$0 = (1+1)_{\text{mod}2}$
1	0	$1 = (1+0)_{\text{mod}2}$
0	1	$1 = (0+1)_{\text{mod}2}$
0	0	$0 = (0+0)_{\text{mod}2}$

### 3) Grouping

자료파일에서 어떤 변수들은 특성상 노출되기가 쉬운 범주로 구성된 경우가 있다. 가령, 나이 변수에서 100세인 사람이 단 1명뿐이라면 누구인지 쉽게 식별할 수가 있을 것이다. 이러한 경우에 인근의 다른 범주들과 통합함으로써 쉽게 식별이 되지 않도록 할 수 있는데, 이러한 방법을 Grouping이라고 한다. 이 방법은 Polychotomous variable인 경우에 적용하면 매우 효과적이다.

#### 나. 연속형 변수

##### 1) Rounding

주어진 자료를 적당한 몫과 나머지의 형태로 나타내는 방법을 Rounding이라고 한다. 예를 들어,  $r$ 를 정수,  $B$ 를 base,  $r \sim U(0, B-1)$ 을 나머지라고 하면,  $r$ 의 형태로 나타낼 수 있다. 만약,  $E(r)$  이라고 한다면,  $r$ 의 확률분포와 기대값 및 분산은 다음과 같다.

	0	1	2	...	B-1	계
$P(r)$	$1/B$	$1/B$	$1/B$	$1/B$	$1/B$	1
$E(r)$	$\sum_{r=0}^{B-1} r P(r) = \frac{B-1}{2}$					
$Var(r)$	$\frac{1}{12} (B^2 - 1)$					

Rounding 방법에서 나머지를 처리하는 방법은 크게 4가지로 구분이 되는데, 이에 대한 자세한 내용은 Kim(2003)을 참고하기 바란다.

##### 2) Grouping into Intervals

이 방법은 연속형 자료를 적당한 구간(interval)으로 그룹화하여 각 구간을 대표하는 code값으로 원 자료를 대체하는 방법으로서, 구간을 넓게 하면 자료의 노출 방지에는 용이하다. 그러나 회귀모형 또는 상관계수 등의 자료분

석시에 다소 제약이 있으며, 특히 구간평균이 제공되지 않기 때문에 구간의 중앙값(median)을 사용하는 것이 적절하다.

### 3) Micro aggregation

자료 파일내에 많은 record들은 선정된 하나의 기준변수에 따라 정렬될 수 있으며, 이를 이용하여 보통 3~4개의 관측값을 그룹화하고 그룹내의 각 자료값은 그룹평균으로 대체할 수 있는데, 이러한 방법으로 노출을 방지하는 것을 Micro aggregation이라고 한다.

### 4) 기타

위에서 언급된 방법 이외에도 이산형 변수의 경우에 적용되는 Data swapping 방법을 동일하게 적용할 수 있으며, Kim(1986, 2001)에 의해 체계화된 Adding noise 방법과 Multiplicative noise 방법 등도 있다.

## 3. Uniqueness

### 가. Uniqueness

제공되는 마이크로자료 파일에서 식별이 어떻게 이루어지는지를 파악할 수 있다면 어느 정도 노출을 제한할 수가 있는데, 이것은 uniqueness를 통해서 가능하다. Uniqueness란 전체 자료파일에서 조사단위의 특성이 유일하게 존재하는 것을 말하며 어떤 조사단위가 식별될 가능성을 나타내는 척도로 사용된다. 가령, 100명으로 구성된 모집단에서 만약 나이가 100세인 사람이 단 한명 있다면 그 사람은 모집단에서 unique하다고 하며, 그 사람은 다른 사람들에 비해 자료에서 식별될 가능성이 매우 높게 된다. 주어진 자료에서 uniqueness는 하나의 변수만으로도 파악할 수 있고 여러 개의 변수들을 조합하여 파악할 수도 있다. 가령, 위의 예에서 100명 중 나이가 60세인 사람이 3명 있다고 하더라도 직업이라는 변수를 포함하여 사원이면서 나이가 60

세인 사람은 단 한명일 수도 있다. 따라서 고려할 변수가 많아질수록 자료의 uniqueness는 점점 더 커지게 된다.

## 나. 노출위험

노출위험(disclosure risk)은 모집단에서 unique한 경우로 한정되는지 아닌지에 따라 좁은 의미와 넓은 의미로 정의할 수 있는데, 여기에서는 좁은 의미에서의 노출위험에 대해서 다루기로 한다. 일반적으로 노출은 다음의 3가지 조건을 만족하는 경우에 발생한다.

### < 노출의 발생조건 >

- ① 어떤 사람이 특정 변수에 대해 모집단에서 unique하다.
- ② 그 사람은 어떤 조사에서 마이크로자료 파일에 포함되어 있다.
- ③ 그 사람은 외부인이 작성한 또 다른 자료파일에도 포함되어 있다.

따라서 어떤 사람이 위의 조건에서 언급한 자료파일들 중 어느 하나에도 나타나지 않는다면 노출은 일어나지 않는다.

한편, 위에서 언급한 노출의 발생조건 하에서 노출위험은 확률적 모형으로 표현될 수 있으며, 이를 위해 다음과 같이 기호를 정의하기로 한다.

- $S_1$
  - $S_2$  : 관심의 대상인 사람
  - $U_P$  : 통계작성기관의 마이크로자료로 구성된 파일 1
  - $U_S$  : 외부인(intruder)에 의해 구성된 파일 2
  - : 모집단의 unique class
  - : 표본의 unique class
- A

만약, 조사과정에서 측정오차가 없다고 가정하고 어떤 사람( )이 모집단에서 unique하다면, 그 사람은 또한 표본에서도 unique할 것이다. 즉, 이것을 확률적으로 표현하면 다음과 같이 나타낼 수 있다.

$$\begin{aligned} \Pr[(A \in U_S) \cap (A \in U_P)] &= \Pr(A \in U_P) \Pr[(A \in U_S) | (A \in U_P)] \\ &= \Pr(A \in U_P) \end{aligned}$$

1) 외부인이  $S_2$  파일에 특정인이 포함된 것을 모르고 있을 경우

만약 금융기관이나 이웃주민 등과 같은 외부인이 그들이 작성한 파일( )  
에 관심있는 특정 사람( )이 포함되어 있다는 것을 모르고 있다면, 위의 관계식  
으로부터 의 노출위험 은 다음과 같이 정의할 수 있다.

$$DR(A) = \Pr[(A \in S_1) \cap (A \in S_2) \cap (A \in U_P)] \quad (1)$$

그리고 조건부 확률공식과 가 모집단에서 unique한 사건은 가 표본에  
추출되거나 추출되지 않거나 독립이라는 사실을 이용하여 위의 식 (1)은 다  
시 다음과 같이 표현할 수 있다.

$$= \Pr[(A \in S_1) \cap (A \in S_2)] \Pr(A \in U_P)$$

$$A \quad S_1 \quad A \quad S_2$$

$$\Pr[(A \in S_1) \cap (A \in S_2)] = \Pr[(A \in S_1)] \times \Pr[(A \in S_2)]$$

반적으로 가 파일 1( )에 있다는 사건은 가 파일 2( )에 있다는 사건  
과 서로 독립이 되므로, 이 된

다. 그러나 어떤 경우에는 두 사건이 서로 독립이 되지 않는 경우도 있다.  
가령, 통계청에서 실시하는 가계조사는 경제활동인구조사의 표본 중 일부만  
을 표본으로 사용하기 때문에 서로 독립이라고 할 수 없다. 만약 이 경제  
활동인구조사에서 작성된 파일이고 가 가계조사에서 작성된 파일이라고 한  
다면

가 될 것이다. 따라서 노출위험은 두  
사건이 서로 독립인지 아닌지에 따라 다음과 같이 나타낼 수 있다.

$$\begin{cases} \Pr(A \in S_1) \Pr(A \in S_2) \Pr(A \in U_P), & \text{독립인 경우} \\ \Pr(A \in S_2) \Pr(A \in U_P), & \text{중속인 경우} \end{cases}$$

(2)

## 2) 외부인이 $S_2$ 파일에 특정인이 포함된 것을 알고 있을 경우

만약 외부인이 파일( $S_2$ )에 관심있는 특정 사람이 포함되어 있다는 것을 이미 알고 있다고 한다면,  $\Pr(A \in S_2) = 1$  이 될 것이다. 따라서 노출위험은 다음과 같이 된다.

$$DR(A) = \Pr(A \in S_1) \Pr(A \in U_p) \quad (3)$$

실제로 은행이나 대출 회사에서는 어떤 채무자에 대한 자세한 정보를 가지고 있다고 하고, 회사의 고용(원역) 마이크로자료 파일에서 그 채무자를 식별하기를 시도한다고 가정하면, 이기 때문에 그 채무자의 노출위험은 위의 식 (3)을 사용하여 계산할 수 있다.

$\Pr(A \in U_p)$  사람이 전체 모집단에서 unique하다는 것을 외부인이 알고 있다면 이 될 것이므로,  $DR(A) = \Pr(A \in S_1)$  다음과 같이 나타낼 수 있다.

## . 마이크로자료의 Masking

### III

#### 1. 분석대상 자료

본 연구에서 마이크로자료 파일을 작성하기 위해 사용한 자료는 지난 2005년 11월에 통계청에서 실시한 “2005 인구주택총조사”의 표본조사 결과의 충청남도 자료이다. 인구주택총조사의 조사표는 크게 전수조사표와 표본조사표로 구분하여 가구원과 가구, 그리고 주택에 관한 사항들을 질의하도록 구성되어 있다. 전수조사표는 기본적인 특성과약을 위해 가구원에 관한 사항 8개, 가구에 관한 사항 7개, 주택에 관한 사항 6개 등 총 21개 항목으로 구성되어 있다. 표본조사표는 전수조사항목을 포함하여 보다 세부적인 특성을 파악코자 가구원에 관한 사항 24개, 가구에 관한 사항 11개, 주택에 관한 사

항 6개 등 총 41개 항목으로 구성되어 있으며, 이외에 추가로 16개 시·도별로 각각 서로 다른 조사항목 3개가 포함되어 전체적으로는 44개 조사항목으로 구성되어 있다. 이에 대한 자세한 내용은 2005 인구주택총조사 조사지침서(2005)를 참고하기 바란다.

전수조사에 대한 집계결과는 각 분야별로 올해 세 번(5월, 6월, 7월)에 걸쳐 공표되었으며, 표본조사에 대한 결과는 1차적으로 통근 및 통학, 인구이동에 대한 사항들이 9월에 공표되었고, 나머지 분야에 대한 집계결과는 올 연말까지 순차적으로 공표될 예정이다. 따라서 본 연구의 시작시점이 8월인 것을 감안하여 분석에서는 전수조사결과와 아울러 표본조사의 잠정결과를 사용하였으며, 대상지역은 우선적으로 충청남도 지역만을 고려하였다. 또한 조사의 편의를 위해 설정된 6종류의 조사구들 중 기숙시설조사구, 특수사회시설조사구, 관광호텔 및 외국인거주지역 조사구는 제외하고 아파트조사구, 보통조사구, 섬조사구를 대상으로 분석하였는데, 이는 자료의 활용측면에서 자료이용자들이 집단가구보다는 개별가구의 특성과악에 더 관심이 많을 것으로 판단하였기 때문이다.

2005 인구주택총조사 결과에 의하면 <표 1>에 나타난바와 같이 충청남도는 총 2,460개 일반 및 아파트조사구에서 가구원 1,798,397명과 660,526가구 및 586,757호의 주택이 있는 것으로 나타났다. 10% 표본의 경우 조사구는 745개이고 가구원은 189,505명, 가구는 71,091가구, 주택은 65,398호인 것으로 각각 나타났다.

표 1> 2005 인구주택총조사 결과(충청남도)

조사	조사구	가구원	가 구	주 택
전수	2,460	1,798,397	660,526	586,757
10% 표본	745	189,505	71,091	65,398

(단위 : 개, 명, 가구, 호)

## 2. 마이크로자료 파일의 작성

### 가. Key 변수의 선정

비밀보호방법을 적용하여 마이크로자료 파일을 작성하기 위해서는 반드시 주요 Key 변수를 선정하여야 한다. 이에 따라 Key 변수의 선정을 위해 2005 인구주택총조사의 조사항목을 검토하였는바, 조사항목의 수가 다소 많은 편이고 특히 개인의 특성에 대한 민감한 질문들이 포함되어 있어 다소 어려움이 예상되었다. 따라서 전체 표본조사의 항목들 중 지역관련 조사항목 3개와 표본조사 세부항목 19개는 제외하고 전수조사항목 21개만을 우선적으로 검토하였다.

주요 Key 변수는 다음과 같은 3단계의 과정을 거쳐 선정하고자 하였다. 먼저 1단계로는 통계청 이외의 외부기관에서 보유하고 있는 자료들 중 인구주택총조사의 조사항목과 중복되어 식별될 가능성이 높다고 판단되는 항목을 우선 선정하였다. 여기서 외부기관이란 통계를 작성하는 기관이나 중앙행정기관, 지방자치단체, 민간기관, 그리고 이웃주민 등이 될 수 있으며 자금을 대출해주는 은행 등의 금융기관도 포함된다. 2단계에서는 1단계에서 식별가능성이 높다고 판단되는 항목에 대해서 빈도분석(frequency analysis)을 실시하여 각 항목의 빈도수와 분포형태 등 보다 자세한 자료의 특성을 파악하였다. 마지막 3단계에서는 각 항목별로 상대적 정보의 양을 나타내는 측도인 엔트로피(entropy)를 다양한 조합별로 각각 계산한 후, 최종적으로 Key 변수를 선정하였다.

엔트로피란 원래 물리학에서 불확실성의 정도를 나타내는데 사용되고 있으나, Shannon(1948)이 엔트로피를 확률적 방식으로 일반화시켜 정보의 양을 나타내는 측도로 활용하도록 하였다. 가령, 변수  $X$ 가 주어져 있을때,  $X$ 의 엔트로피  $H(X)$ 는 다음과 같이 정의된다.

$$\begin{aligned}
 H(X) &= - \sum_x P(x) \log_2[P(x)] \\
 &= \sum_x P(x) \log_2 \left[ \frac{1}{P(x)} \right]
 \end{aligned}$$

단,  $P(x) = \Pr(X=x)$  이다. 또한 변수  $X_1, X_2, \dots, X_n$  의 joint 엔트로피는 다음과 같이 계산할 수 있다.

$$\begin{aligned}
 H(X_1, \dots, X_n) &= - \sum_{x_1} \dots \sum_{x_n} P(x_1, \dots, x_n) \log_2[P(x_1, \dots, x_n)] \\
 &= \sum_{x_1} \dots \sum_{x_n} P(x_1, \dots, x_n) \log_2 \left[ \frac{1}{P(x_1, \dots, x_n)} \right]
 \end{aligned}$$

$$P(x_1, \dots, x_n) = \Pr(X_1 = x_1, \dots, X_n = x_n)$$

단, 이다.

인구주택총조사의 조사항목을 검토한 후, 위에서 언급한 3단계의 과정을 거쳐 Key 변수를 선정된 결과는 다음과 같다.

### 1) 가구원에 관한 사항

가구원에 관한 사항에서는 총 8개의 전수조사 항목 중 5개 항목(성별, 나이, 가구주와의 관계, 교육정도, 혼인상태)을 외부에서 식별가능성이 높은 항목으로 선정하였다. 이렇게 선정된 5개 항목에 대해 10% 표본조사의 결과를 바탕으로 빈도분석을 실시하여 자료의 특성을 파악하였는데, 성별은 남녀별로 서로 비슷하게 분포되어 있다. 그러나 나이 항목의 경우 30~40대가 전체의 약 30%를 차지하고 있으며, 100세 이상인 고령자도 24명이 있는 것으로 나타났다. 가구주와의 관계 항목에서는 가구주와 배우자, 자녀와 그 배우자가 전체의 90.7% 정도로 대부분을 차지하고 있다. 교육정도 항목에서는 고등학교 졸업이 전체의 약 22.6%로 가장 많았으며, 혼인상태 항목의 경우에는 미혼이 약 20.7%이고 사별과 이혼을 포함해 약 13.7%가 배우자가 없는 것으로 나타났다.

각 항목의 다양한 조합별로 엔트로피를 계산한 결과는 부록 B의 <부표 B-1>에 주어져 있다. 먼저 개별 항목을 살펴보면, 성별의 경우 카테고리가 2개로 엔트로피가 약 0.999이고, 나이는 총 105개의 카테고리에 약 6.398, 가구주와의 관계는 총 14개 카테고리에 약 2.126, 교육정도는 총 32개 카테고리에 약 3.323, 그리고 혼인상태는 결측값을 제외하고 총 4개 카테고리에 약 1.355의 엔트로피 값이 각각 나타났다. 항목들 간의 조합에 의한 경우를 살펴보면, 항목이 점차 추가될수록 엔트로피 값이 더 높아지는데 이는 그만큼 더 정보의 양이 많아짐을 의미한다. 가령, 성별의 경우 엔트로피는 약 0.999이지만 성별과 나이를 조합한 경우의 엔트로피는 약 7.390으로 더 높아진다. 이러한 결과를 바탕으로 가구원에 관한 사항에서는 처음 5개의 항목 중 교육정도를 제외하고 성별과 나이, 가구주와의 관계, 혼인상태 등 4개 항목을 최종적인 Key 변수로 선정하였다.

## 2) 가구에 관한 사항

가구에 관한 사항에서는 총 7개의 전수조사 항목 중 4개 항목(가구구분, 거주층, 점유형태, 주인가구 및 주택소유여부)을 외부에서 식별가능성이 높은 항목으로 선정하였다. 10% 표본조사의 결과를 바탕으로 4개 항목에 대해 빈도분석을 실시하여 특성을 파악한 결과, 가구구분의 항목에서는 가족으로 이루어진 가구가 전체의 74.7%를 차지하고 있으며, 거주층 항목에서는 지상이 99.6%로 가장 많은 빈도수를 나타내고 있다. 점유형태 항목의 경우 전체의 약 65.6%가 주거전용이면서 자기집인 경우이며, 주인가구 및 주택소유여부를 묻는 항목에서는 약 61.9%가 주인가구이면서 다른 곳에 주택미소유인 것으로 나타났다. 4개 항목에 대한 각 범주별 빈도수와 엔트로피를 계산한 결과가 부록 B의 <부표 B-2>에 주어져 있다. 이 결과를 바탕으로 가구구분, 점유형태, 주인가구 등 3개 항목을 Key 변수로 최종 선정하였는데, 특히 주인가구 및 주택소유여부 항목에서 주택소유여부에 대한 내용은 중요성이 상대적으로 낮아 제외하고 주인가구 여부만 Key 변수로 선정하였다.

### 3) 주택에 관한 사항

주택에 관한 사항에서는 총 6개 항목 중 거처의 종류, 연건평, 대지면적 등 3개 항목을 외부에서 식별가능성이 높은 항목으로 선정하였다. 따라서 10% 표본조사의 결과를 바탕으로 이들 3개 항목에 대해 빈도분석을 실시하여 특성을 파악한 결과, 거처의 종류 항목에서는 단독주택이 59.9%, 아파트가 32.5%로 대부분을 차지하고 있다. 연건평의 항목에서는 전체의 약 80.9%가 20~30평대이며, 200평대가 넘는 경우도 0.06% 정도가 있다. 대지면적의 경우에는 100평대가 약 13.3%로 가장 많고, 400평대 이상도 약 1.1%가 있는 것으로 나타났다.

한편, 연건평과 대지면적의 조사결과를 자세히 살펴보면, 5평이나 10평단위로 조사된 결과가 많이 나타남을 알 수 있다. 이는 응답자들이 자세한 면적을 응답하지 않고 5평 또는 10평단위로 응답한 결과로서, 조사과정에서 발생하는 응답자나 조사원에 의한 측정오차(measurement error)라 할 수 있다. 따라서 연건평과 대지면적에 대한 결과분석은 각 조사단위별로 하지 않고 10평단위로 범주화하여 실시하였으며, 응답자들의 응답속성을 감안하여 연건평이나 대지면적이 25~34평인 것을 30평대, 35~44평인 것을 40평대라고 범주화하였다. 또한 대지면적의 경우 다가구주택 등으로 인해 20평대 이하의 값들도 있는데, 이에 대한 처리는 추후에 별도로 해야 할 것이다. 각 항목별 엔트로피에 대한 결과는 부록 A의 <부표 3>에 주어져 있으며, 이를 바탕으로 주택에 관한 사항에서는 연건평과 대지면적을 제외하고 거처의 종류만 **최종표본의 Key 변수**로 선정하였다.

현재 통계청에서 제공하고 있는 인구주택총조사의 마이크로자료는 전체 10% 표본조사 중 2%에 해당하는 표본조사 결과자료이다. 따라서 본 연구에서도 비밀보호방법을 적용한 마이크로자료 파일의 작성을 위해 같은 규모인 2%의 표본을 추출하기로 하였다. 15 -

## 1) 추출틀과 추출단위

2% 표본의 추출을 위한 추출틀(sample frame)은 충청남도 표본조사구내에 있는 모든 가구명부로 하였으며, 추출단위(sampling unit)는 가구로 하였다. 추출단위를 가구로 한 것은 인구주택총조사의 표본조사구 특성을 감안할 때, 조사구보다는 가구가 개인의 식별가능성이 줄이는데 더 바람직 할 것으로 판단하였기 때문이다. 즉, 현행 인구주택총조사에서는 60~80가구를 하나의 조사구로 설정한 후, 이 중 10%를 표본조사구로 추출하고 표본조사구내 모든 가구는 표본조사표를 작성하도록 하고 있다. 이러한 방법은 조사의 편리성은 있으나 응답자의 식별가능성이 높아 노출의 위험이 그만큼 더 커지게 된다는 것을 의미한다. 가령, 아파트 표본조사구내에 포함되어 응답한 어떤 가구는 자신들의 이웃집도 같은 표본조사구로 조사에 응답했다는 사실을 쉽게 알 수 있다. 따라서 마이크로자료가 공표될 경우 이웃집의 정보를 식별할 가능성은 상당히 높게 될 것이다. 이러한 이유로 개인의 식별 가능성을 최대한 줄이기 위해 가구를 추출단위로 하였다.

## 2) 추출방법과 규모

위에서 언급한 바와 같이 마이크로자료 파일의 작성을 위해 새로 추출하는 2% 표본은 가급적 개인의 식별 가능성을 최소화하기 위해 계통추출법을 적용하여 추출하였다. 즉, 16개 시·군·구별로 정렬한 후 조사구 및 거처순으로 다시 정렬을 하고 최종적으로 가구원들의 성별 및 나이순으로 각각 정렬하여 추출틀을 작성하였다. 충청남도 전체의 2% 표본은 기존의 10% 표본의 20%에 해당되므로 SAS 프로그램을 이용하여 최종적인 2% 표본을 계통추출 하였으며, 그 결과가 <표 2>에 주어져 있다. 이렇게 추출한 표본이 모집단을 얼마나 잘 대표하는지를 알아보기 위해 가구원의 성별을 비교한 결과, 모집단(전수조사 결과)과 10% 표본의 구조와 크게 다르지 않음을 알 수 있었다.

〈표 2〉 표본추출 현황

구분	전수(A)	10% 표본(B)		2% 표본(C)		
		인수	구성비(B/A)	인수	구성비(C/A)	구성비(C/B)
가구원	1,798,397	189,505	10.54%	38,027	2.11%	20.07%
남	892,044	93,023	10.43%	18,644	2.09%	20.04%
녀	906,353	96,482	10.65%	19,383	2.14%	20.09%
가 구	660,526	71,091	10.76%	14,218	2.15%	20.00%
주 택	586,757	65,398	11.15%	13,038	2.22%	19.94%

다. 노출위험의 계산

마이크로자료 파일의 작성을 위해 새로 추출한 2% 표본에서 노출과 식별의 가능성이 어느 정도인지를 파악하고자 uniqueness와 노출위험을 계산하기로 하고,  $S_A$  이를 위해 계산에 필요한 기호를 다음과 같이 나타내었다.

- $S_1$  : 관심의 대상인 사람
- $S_2$  : 새로 추출한 2% 표본파일
- $U_P$  : 외부인에 의해 구성된 파일
- $U_S$  : 모집단(전수조사결과)의 unique class
- : 2% 표본의 unique class

Uniqueness와 노출위험의 계산을 위해 먼저, 앞 절에서 최종 선정한 8개의 Key 변수에 대한 다양한 조합을 고려하였다. 8개 변수로 얻을 수 있는 경우의 수는 1개 변수를 이용한 조합부터 8개 변수를 모두 이용한 조합까지 합하여 총 247개가 된다.

충청남도의 전수조사결과를 모집단으로 하고 새로 추출한 2% 표본조사결과를 표본으로 한 후, Key 변수의 각 조합별로 빈도수가 1이 되는 경우를 파악한 결과가 부록 C의 <부표 C-1>과 <부표 C-2>에 주어져 있다. 여러 조합들 중 8개 변수를 모두 고려한 조합이 uniqueness의 발생이 최대가 되는 경우이므로, 결국 uniqueness는 8개 변수를 모두 고려한 조합의 발생 빈도수로 파악할 수가 있는데, 모집단에서는 총 9,664개가 있으며, 표본에서는 1,810개가 있는 것으로 각각 나타났다. 따라서 모집단에서 어떤 사람( )이 unique할 확률은 약 0.54%가 되고,

$$\Pr(A \in U_P) = \frac{9,664}{1,798,397} = 0.00537$$

A

같은 방법으로 표본에서 어떤 사람( )이 unique할 확률은 약 4.76%가 된다.

$$\Pr(A \in U_S) = \frac{1,810}{38,027} = 0.04759$$

즉, 표본에서의 uniqueness가 모집단에서의 uniqueness보다 약 8.8배 이상 더 높게 나타나고 있음을 알 수 있다. 이는 모집단보다 표본에서 unique한 사람이 더 많이 발생한다는 것을 의미하지만, 모집단에서 unique하지 않은 사람도 표본에서는 unique할 수가 있기 때문에 표본에서 unique한 사람이 반드시 모집단에서도 unique한 것은 아니다. 만약 모집단에서 unique한 어떤 사람이 표본으로 추출되었다면 이 사람의 노출 가능성은 상대적으로 높을 것이다. 충청남도의 경우 전체 모집단에서 unique한 9,664명 중 216명이 다시 2% 표본으로 추출되어 unique한 것으로 나타났다.

			$U_P$	$U_S$	$U_P \cap U_S$
					(단위 : 명, %)
조사	모집단	2% 표본			
개수	1,798,397	38,027	9,664	1,810	216
구성비			0.537	4.760	0.01

이상의 uniqueness에 대한 결과를 이용하여 충청남도의 노출위험을 계산해 보았다. 노출위험의 발생 가능성이 최대가 되는 가장 극단적인 경우를 고려하여 외부인이 직접 작성한 파일( $S_1$ )에 어떤 특정사람이 포함되어 있다는 것을 이미 알고 있다고 가정하였다. 이 가정하에서는  $\Pr(A \in S_2) = 1$  이 되므로 2장에서 정의한 수식 (3)을 이용하여 노출위험을 계산하면, 어떤 특정 사람(A)이 모집단에서 unique하면서 인구주택총조사의 2% 표본에 포함될 확률은 약 0.01%가 된다. 즉,

$$DR(A) = \Pr(A \in S_1) \Pr(A \in U_p) = 0.02 \times \frac{9,664}{1,798,397} = 0.00011$$

$$\Pr(A \in S_1) = 0.02$$

여기서  $\Pr(A \in U_p)$  인데, 이는 제공하는 마이크로자료의 규모가 2%이기 때문이다. 위의 결과는 외부인이 가지고 있는 파일에 어떤 사람이 포함되어 있다는 것을 알고 있다고 가정하고 2%를 마이크로자료 파일로 제공할 경우 10,000명 중 1명 정도가 노출될 가능성이 있다는 것을 의미한다.

#### 라. 자료의 Grouping

위에서 2% 표본의 uniqueness에 대해 살펴본 결과 총 38,027명 중 약 4.76%인 1,810명이 unique한 것으로 나타났다. 만약 추출된 2%의 표본에 비밀보호방법을 적용하면 unique한 사람은 더 감소할 것이며, 이는 결국 mask된 자료파일을 제공할 경우 개인의 식별과 노출에 대한 위험이 훨씬 더 줄어들 것이다. 비밀보호를 위한 방법은 자료의 형태에 따라 다양한 방법이 있는데, 여기서는 grouping방법을 적용하여 mask된 자료파일을 만들고자 하였다. Grouping 방법은 주로 이산형 자료로 구성된 경우 적용이 간편하고 매우 효과적인 방법이다.

본 연구에서 선정한 8개의 Key 변수들은 나이 변수를 제외하고는 모두 이산형 변수들이고 나이변수도 범주형 자료로 변환할 수가 있으므로, 각 변수들의 빈도분석의 결과를 이용하여 적절한 수준의 범주로 grouping하였다. 먼저, 가구원에 관한 사항에서는 4개의 Key 변수가 사용되었는데 성별 항목은

남녀별 범주를 그대로 사용하고, 나이 항목은 99세까지는 5세별로 20개의 범주로 하고, 100세 이상은 하나의 범주로 하여 총 21개 범주로 grouping하였다. 가구주와의 관계 항목은 14개의 범주를 총 4개(가구주, 배우자, 자녀 및 자녀의 배우자, 기타)의 범주로 축소하였다. 혼인상태 항목은 4개 범주 중 사별과 이혼을 합하여 총 3개(미혼, 기혼 및 배우자 있음, 기혼 및 배우자 없음)의 범주로 grouping하였다.

가구에 관한 사항에서는 Key 변수가 3개인데, 가구구분 항목은 5개 범주에서 가족이 아닌 남남끼리 사는 5인 이하 가구와 가족이 아닌 남남끼리 사는 6인 이상 가구를 합쳐 기타 범주로 하여 총 4개 범주로 grouping하였다. 점유형태는 원래 12개의 범주가 있는데, 주거전용과 영업전용은 그대로 두고 4개(자가, 전세, 월세, 무상) 범주로 하여 총 8개의 범주로 grouping하였다. 주인가구 및 주택소유여부 항목은 주인가구와 세들어 사는 가구로 크게 2개 범주로 축소하였다. 한편, 주택에 관한 사항에서는 거처의 종류 항목이 Key 변수로 사용되었는데, 기존의 12개 범주를 4개(단독주택, 아파트, 공동주택, 기타)의 범주로 grouping하였다. 각 항목에 대한 grouping의 결과가 <표 4>에 주어져 있다.

2% 표본에서 8개 Key 변수를 <표 4>에 주어진 범주별로 그룹화한 후 uniqueness를 살펴보면, 총 553명이 unique한 것으로 나타났는데, 이는 그룹화하기 전의 1,810명보다 1,257명이 줄어들어 69.4% 정도 감소하였다. 따라서 그룹화를 한 후 2% 표본을 마이크로자료 파일로 제공해도 노출위험이 적기 때문에 적절하다고 할 수 있을 것이다.

<표 4> Key 변수의 grouping 결과

Key 변수	1인 가구		2인 이상 가구	
	개수	범주	개수	범주
성별	2	① 남자 ② 여자	2	① 남자 (1) ② 여자 (2)
나이	105	각 세별	21	①~⑪ (각 5세별) ⑫ 100세 이상
가구주와의 관계	14	① 가구주 ② 배우자 ③ 자녀 ④ 자녀의 배우자 ⑤ 가구주의 부모 ⑥ 배우자의 부모 ⑦ 손자녀, 그의 배우자 ⑧ 증손자녀, 그의 배우자 ⑨ 조부모 ⑩ 형제자매, 그 배우자 ⑪ 형제자매의 자녀 ⑫ 부모의 형제자매	4	① 가구주 (1) ② 배우자 (2) ③ 자녀 및 자녀의 배우자 (3+4) ④ 기타 (5+ ... +14)
혼인상태	4	⑬ 기타 친인척 ⑭ 기타동거인  ① 미혼 ② 배우자 있음 ③ 사별 ④ 이혼	3	① 미혼 (1) ② 기혼 및 배우자 있음 (2) ③ 기혼 및 배우자 없음 (3+4)
가구구분	5	① 가족으로 이루어진 가구 ② 가족과 가족이외의 사람이 함께사는 가구 ③ 1인 가구 ④ 가족이 아닌 남남끼리 사는 5인 이하 가구 ⑤ 가족이 아닌 남남끼리 사는 6인 이상 가구	4	① 가족으로 이루어진 가구 (1) ② 가족과 가족이외의 사람이 함께사는 가구 (2) ③ 1인 가구 (3) ④ 기타 (4+5)

(단위 : 개)

<표 4> Key 변수의 grouping 결과(계속)

Key 변수	주거형태		주거형태	
	개주	범주	개주	범주
점유형태	12	① 주거전용 + 자기집 ② 주거전용 + 전세 ③ 주거전용+보증금있는 월세 ④ 주거전용+보증금없는 월세 ⑤ 주거전용 + 사글세 ⑥ 주거전용 + 무상 ⑦ 영업전용 + 자기집 ⑧ 영업전용 + 전세 ⑨ 영업전용+보증금있는 월세 ⑩ 영업전용+보증금없는 월세 ⑪ 영업전용 + 사글세 ⑫ 영업전용 + 무상	8	a) 주거전용 + 자기집 (①) b) 주거전용 + 전세 (②) c) 주거전용 + 월세 (③+④+⑤) d) 주거전용 + 무상 (⑥) e) 영업전용 + 자기집 (⑦) f) 영업전용 + 전세 (⑧) g) 영업전용 + 월세 (⑨+⑩+⑪) h) 영업전용 + 무상 (⑫)
주인가구	6	① 주인가구 + 다른곳에 주택소유 ② 주인가구 + 다른곳에 주택미소유 ③ 대표가구 + 다른곳에 주택소유 ④ 대표가구 + 다른곳에 주택미소유	2	a) 주인가구 (①+②) b) 세들어 사는가구 (③+ ... +⑥)
거처의 종류	12	⑤ 세들어 사는가구 + 다른곳에 주택소유 ⑥ 세들어 사는가구 + 다른곳에 주택미소유  ① 단독주택 ㉠ 일반단독주택 ㉡ 다가구단독주택 ㉢ 영업겸용단독주택 ② 아파트 ③ 연립주택 ④ 다세대주택 ⑤ 비거주용 건물 ⑥ 오피스텔 ⑦ 호텔, 여관내 객실	4	a) 단독주택 (①) b) 아파트 (②) c) 공동주택 (③+④) d) 기타 (⑤+ ... +⑩)
		⑧ 기숙사 및 특수시설 ⑨ 판잣집 등 ⑩ 기타		

## IV 결 론

앞에서 살펴본 바와 같이 본 연구에서는 마이크로자료 파일을 공표할 경우 개인의 비밀을 보호하는 다양한 형태의 masking 방법을 소개하였고, 마이크로자료가 제공될 때 노출과 개인의 식별의 발생원인과 문제점 등을 살펴 보았다. 또한 식별의 측도로 사용되는 uniqueness의 개념도 소개하였으며 특히, 노출의 위험을 통계적 확률로 정리하여 계산함으로써 쉽게 이해할 수 있도록 설명하였다.

비밀보호방법을 적용한 마이크로자료 파일을 작성하기 위해 2005 인구주택총조사의 표본조사 결과자료 중 충청남도 지역의 자료를 이용하였다. 먼저 masking에 필요한 Key 변수로 8개 변수를 선정하였다. 즉, 가구원에 관한 사항에서는 4개 항목(성별, 나이, 가구주와의 관계, 혼인상태)을 선정하였고, 가구에 관한 사항에서는 3개 항목(가구구분, 점유형태, 주인가구)을, 그리고 주택에 관한 사항에서는 1개 항목(거처의 종류)을 각각 선정하였다. 통계이용자들에게 제공할 마이크로자료 파일은 현행 통계청에서 제공하는 것과 같은 규모인 2% 표본조사결과로 정하고, 충청남도의 10% 표본조사결과들 중 20%를 계통추출법으로 추출하여 전체적으로 2% 표본을 구성하였다.

한편, 자료파일의 uniqueness와 노출의 위험을 알아보기 위해 전수조사결과와 2% 표본조사 결과에서 8개 Key 변수의 각 조합별로 빈도분석을 통해 빈도수가 1개인 경우를 파악하여 보았다. 그 결과, 모집단(전수조사결과)에서는 1,798,397명 중 9,664명(약 0.54%)이 unique한 것으로 나타났고, 2% 표본에서는 38,027명 중 1,810명(약 4.76%)이 unique한 것으로 나타났다. 그리고 이에 따른 노출 위험은 약 0.01%가 됨을 알 수 있었다. 8개 Key 변수에 대해 grouping 방법을 적용하여 2% 표본을 다시 구성한 결과 unique한 경우는 553명으로 grouping하기 전의 1,810명에 비해 약 69.4%가 감소하였다. 따라서 grouping을 한 후 2% 표본을 마이크로자료 파일로 제공해도 노출위험이 적기 때문에 적절하다고 할 수 있다.

일반적으로 Key 변수의 선정은 연구자에 따라 다소 주관적이며, 일정하지 않을 수도 있다. 즉, 본 연구에서는 여러 과정을 거쳐 최종적으로 8개를 선정하였으나, 외부인(금융기관, 이웃주민 등)들이 파악할 수 있는 항목이 8개 이외에도 더 있다면 이 항목을 Key 변수의 선정에 고려해야 할 것이다. 또한 자료를 grouping할 때에도 자료제공의 범위 등에 따라 다양하게 할 수도 있다. 가령, 나이 항목의 경우 본 연구에서는 5세 간격으로 하고 100세 이상은 하나의 범주로 grouping하였으나, 경우에 따라서는 10세 간격이나 20세 간격으로 할 수도 있을 것이다. 따라서 본 연구에서 선정한 8개 Key 변수와 그에 따른 grouping방법이 최적의 방법이라고 할 수는 없으며, 자료의 특성이나 제공의 범위에 따라 적절한 방법을 적용하면 될 것이다.

본 연구에서는 마이크로자료 파일을 작성할 때 인구주택총조사의 질문항목이 주로 이산형 형태로 구성된 것을 감안하여 이산형 변수의 비밀보호기법인 grouping방법을 적용하였으나, 인구주택총조사의 일부 항목들은 연속형인 경우도 있다. 가령, 주택에 관한 사항에서 연건평이나 대지면적 등의 항목은 연속형 자료로 얻어진다. 따라서 rounding방법 등과 같은 연속형 변수에 대한 기법을 적용하는 것이 더 타당할 것이다. 그러나 여기서는 연속형 변수에 대한 방법은 적용하지 않았기 때문에, 위의 항목에 대한 자료를 제공할 경우에는 원래자료를 제공하는 것보다는 이산형 형태로 범주화시켜 제공하는 것이 더 바람직할 것으로 판단된다.

향후에는 본 연구의 결과를 바탕으로 인구주택총조사의 표본조사결과에 대해 16개 시·도를 모두 포함한 전국 단위의 마이크로자료 파일을 작성할 수 있는 연구가 계속해서 추진되어야 할 것이다. 아울러 여기서 언급한 grouping방법 이외에 다른 방법들을 적용하는 연구뿐만 아니라, 통계청에서 생산하고 있는 각종 통계들 중 마이크로자료의 제공을 많이 요구하는 통계 조사를 대상으로 비밀보호방법을 적용하는 연구도 병행해서 진행되어야 할 것이다. 또한 자료이용자들의 요구가 발생할 경우 마이크로자료의 제공범위를 어느 정도로 할 것인가에 대한 고민과 함께 얼마만큼 신속하게 제공할 수 있는가에 대한 검토도 필요하다. 끝으로 이번 연구를 계기로 노출과 개인 식별에 대한 올바른 이해와 비밀보호방법에 대한 관심이 고조되어 다양한 방법들이 개발되고 활용되어지길 기대한다.

## 참고문헌

- [1] 박 원환, 황 조연(2004), “통계자료의 특성과 비밀보호방법에 관한 연구”, 통계청, 통계연구보고서.
- [2] 통계청(2005), 2005 인구주택총조사 조사지침서.
- [3] Bethlehem, J. G., Keller, W. J., and Pannekoek J. (1990), “Disclosure Control of Microdata”, *Journal of the American Statistical Association*, 85, 38-45.
- [4] Cox, L. H., McDonald, S., and Nelson, D. (1986), “Confidentiality Issues at the United States Bureau of the Census”, *Journal of Official Statistics*, 2, 135-160.
- [5] Dalenius, T., and Reiss, S. P. (1982), “Data Swapping: A Technique for Disclosure Control”, *Journal of Statistical Planning and Inference*, 6, 73-85.
- [6] Fuller, W. A. (1993), “Masking Procedures for Microdata Disclosure Limitation”, *Journal of Official Statistics*, 9, 383-406.
- [7] Kim, J. (1986), “A Method for Limiting Disclosure in Microdata Based on Random Noise and Transformation”, in *Proceedings of the Section on Survey Research Methods, American Statistical Association*, pp. 370-374.
- [8] Kim, J. (1987), “A Further Development of the Randomized Response Technique for Masking Dichotomous Variables”, in *Proceedings of the Section on Survey Research Methods, American Statistical Association*, pp. 239-244.
- [9] Kim, J. (1990), “Subdomain Estimation for the Masked Data”, in *Proceedings of the Section on Survey Research Methods, American Statistical Association*, pp. 456-461.

- [10] Kim, J., and Winkler, W. E. (1995), "Masking Microdata Files", in *Proceedings of the Section on Survey Research Methods, American Statistical Association*, pp. 114-119.
- [11] Kim, J., and Winkler, W. E. (2001), "Multiplicative Noise for Masking Continuous Data", in *Proceedings of the Section on Survey Research Methods, American Statistical Association*, CD-ROM.
- [12] Kim, J., Katzoff, M., Gonzalez Jr, J., and Williams, P. (2003), "Techniques for Masking Microdata", *National Center for Health Statistics internal memorandum*.
- [13] Marsh, C., Skinner, C., Arber, S., Penhale, B., Openshaw, S., Hobcraft, J., Lievesley, D., and Walford, N. (1991), "The Case for Samples of Anonymized Records from the 1991 Census", *Journal of the Royal Statistical Society A*, 154, 305-340.
- [14] Netherlands Central Bureau of Statistics (1987), *Proceedings of the Seminar on Openness and Protection of Privacy in the Information Society*, Voorburg: Author.
- [15] Shannon, C. E. (1948), "A Mathematical Theory of Communication", *The Bell System Technical Journal*, 27, 379-423, 623-656.
- [16] Skinner, C., Marsh, C., Openshaw, S., and Wymer, C. (1994), "Disclosure Control for Census Microdata", *Journal of Official Statistics*, 10, 31-51.
- [17] Skinner, C., and Holmes, D. (1998), "Estimating the Re-identification Risk Per Record in Microdata", *Journal of Official Statistics*, 14, 361-372.

## 부 록

### A. 2005 인구주택총조사 결과(충청남도)

#### <부표 A-1> 전수조사 결과

(단위 : 개, 명, 가구, 호)

지역	조사구	가구원	가구	주택
전체	2,460	1,798,397	660,526	586,757
천안시	281	502,515	174,753	143,630
공주시	173	117,880	44,012	38,141
보령시	79	94,071	35,111	33,033
아산시	148	187,336	68,406	60,845
서산시	105	139,239	50,420	44,565
논산시	151	117,293	44,382	40,733
계룡시	110	30,476	9,768	9,357
금산군	129	50,648	20,057	18,714
연기군	239	75,479	28,369	23,600
부여군	148	70,647	27,732	26,764
서천군	96	55,993	22,892	21,821
청양군	64	30,592	12,430	12,068
홍성군	223	80,165	30,372	27,118
예산군	195	78,911	29,737	28,220
태안군	150	57,492	21,758	20,808
당진군	169	109,660	40,327	37,340

<부표 A-2> 10% 표본조사 결과

(단위 : 개, 명, 가구, 호)

지역	조사구	가구원	가구	주택
전체	749	189,505	71,091	65,398
천안시	139	43,233	15,098	12,878
공주시	51	13,554	5,098	4,513
보령시	35	10,598	4,023	3,858
아산시	68	18,324	6,732	6,009
서산시	52	13,300	4,771	4,365
논산시	57	12,777	4,788	4,568
계룡시	16	2,877	955	891
금산군	34	6,866	2,840	2,694
연기군	37	7,418	2,772	2,492
부여군	36	11,203	4,598	4,494
서천군	32	7,277	3,155	3,045
청양군	21	5,917	2,486	2,432
홍성군	44	9,275	3,587	3,468
예산군	42	9,374	3,609	3,505
태안군	34	6,677	2,519	2,476
당진군	51	10,835	4,060	3,710

<부표 A-3> 2% 표본조사 결과

(단위 : 개, 명, 가구, 호)

지역	조사구	가구원	가구	주택
전체	749	38,027	14,218	13,038
천안시	139	8,592	3,020	2,569
공주시	51	2,719	1,019	894
보령시	35	2,166	805	773
아산시	68	3,722	1,346	1,195
서산시	52	2,707	954	870
논산시	57	2,544	958	906
계룡시	16	562	191	177
금산군	34	1,384	568	537
연기군	37	1,493	554	495
부여군	36	2,251	920	898
서천군	32	1,422	631	608
청양군	21	1,193	497	487
홍성군	44	1,869	718	693
예산군	42	1,867	721	701
태안군	34	1,371	504	495
당진군	51	2,165	812	740

## B. Key 변수의 Entropy

### <부표 B-1> 성별의 Entropy

(단 위 : 명 , % )

성 별	빈 도 수	구 성 비	Entropy
계	1 8 9 , 5 0 5	1 0 0 . 0	0 . 9 9 9 7 6
남	9 3 , 0 2 3	4 9 . 1	0 . 5 0 3 9 2
여	9 6 , 4 8 2	5 0 . 9	0 . 4 9 5 8 4

### <부표 B-2> 나이의 Entropy

(단 위 : 세 , %)

나이	빈도수	구성비	Entropy	나이	빈도수	구성비	Entropy
계	189,505	100.0	6.39830	19	2,035	1.1	0.07024
1	1,338	0.7	0.05045	20	1,585	0.8	0.05772
2	1,692	0.9	0.06078	21	1,471	0.8	0.05441
3	1,879	1.0	0.06600	22	1,360	0.7	0.05112
4	1,903	1.0	0.06666	23	1,960	1.0	0.06821
5	2,194	1.2	0.07447	24	2,224	1.2	0.07526
6	2,438	1.3	0.08080	25	2,256	1.2	0.07610
7	2,408	1.3	0.08003	26	2,340	1.2	0.07828
8	2,493	1.3	0.08220	27	2,249	1.2	0.07592
9	2,592	1.4	0.08469	28	1,995	1.1	0.06916
10	2,534	1.3	0.08323	29	2,150	1.1	0.07331
11	2,743	1.4	0.08844	30	2,247	1.2	0.07586
12	2,656	1.4	0.08629	31	2,240	1.2	0.07568
13	2,657	1.4	0.08632	32	2,654	1.4	0.08624
14	2,604	1.4	0.08499	33	2,642	1.4	0.08594
15	2,528	1.3	0.08308	34	2,697	1.4	0.08731
16	2,316	1.2	0.07766	35	3,056	1.6	0.09602
17	2,176	1.1	0.07400	36	2,832	1.5	0.09063
18	2,081	1.1	0.07147	37	2,616	1.4	0.08529

(단위 : 세, %)

나이	빈도수	구성비	Entropy	나이	빈도수	구성비	Entropy
38	2,945	1.6	0.09336	70	2,782	1.5	0.08940
39	2,636	1.4	0.08579	71	2,220	1.2	0.07516
40	2,944	1.6	0.09334	72	2,242	1.2	0.07573
41	2,719	1.4	0.08785	73	2,188	1.2	0.07431
42	2,777	1.5	0.08928	74	1,933	1.0	0.06748
43	3,166	1.7	0.09863	75	1,666	0.9	0.06004
44	2,978	1.6	0.09416	76	1,593	0.8	0.05795
45	3,007	1.6	0.09485	77	1,291	0.7	0.04903
46	3,399	1.8	0.10405	78	1,405	0.7	0.05246
47	2,995	1.6	0.09457	79	1,011	0.5	0.04028
48	3,093	1.6	0.09690	80	1,083	0.6	0.04258
49	2,957	1.6	0.09365	81	862	0.5	0.03539
50	2,618	1.4	0.08534	82	759	0.4	0.03190
51	2,644	1.4	0.08599	83	709	0.4	0.03016
52	2,644	1.4	0.08599	84	663	0.3	0.02855
53	1,987	1.0	0.06895	85	516	0.3	0.02320
54	2,658	1.4	0.08634	86	378	0.2	0.01789
55	1,868	1.0	0.06569	87	332	0.2	0.01604
56	2,162	1.1	0.07363	88	275	0.1	0.01368
57	2,316	1.2	0.07766	89	219	0.1	0.01128
58	2,171	1.1	0.07387	90	211	0.1	0.01092
59	2,578	1.4	0.08434	91	162	0.1	0.00871
60	2,041	1.1	0.07040	92	113	0.1	0.00639
61	1,626	0.9	0.05890	93	92	0.0	0.00534
62	1,924	1.0	0.06723	94	69	0.0	0.00416
63	1,927	1.0	0.06731	95	56	0.0	0.00346
64	2,343	1.2	0.07836	96	25	0.0	0.00170
65	2,660	1.4	0.08639	97	33	0.0	0.00217
66	1,909	1.0	0.06682	98	15	0.0	0.00108
67	2,263	1.2	0.07628	99	12	0.0	0.00088
68	2,554	1.3	0.08374	100~	24	0.0	0.00193
69	2,216	1.2	0.07505				

<부표 B-3> 가구주와의 관계의 Entropy

(단위 : 명, %)

가 구 주 와 의 관 계	빈 도 수	구 성 비	Entropy
계	189,505	100.0	2.1258
1	71,091	37.5	0.5306
2	45,889	24.2	0.4954
3	53,422	28.2	0.5150
4	1,515	0.8	0.0557
5	6,715	3.5	0.1707
6	491	0.3	0.0223
7	4,624	2.4	0.1307
8	41	0.0	0.0026
9	123	0.1	0.0069
10	1,368	0.7	0.0514
11	394	0.2	0.0185
12	36	0.0	0.0023
13	457	0.2	0.0210
14	3,339	1.8	0.1027

<부표 B-4> 혼인상태의 Entropy

(단위 : 명, %)

혼 인 상 태	빈 도 수	구 성 비	Entropy
계	154,415	100.0	1.3547
1	32,020	20.7	0.4707
2	101,159	65.5	0.3997
3	17,610	11.4	0.3572
4	3,626	2.3	0.1271

<부표 B-5> 교육정도의 Entropy

(단 위 : 명 , % )

교육 정도	빈 도 수	구 성 비	Entropy
계	189,505	100.0	3.3231
10	31,976	16.9	0.4332
21	32,768	17.3	0.4378
22	15,464	8.2	0.2950
24	2	0.0	0.0002
25	1,738	0.9	0.0621
31	15,725	8.3	0.2980
32	7,343	3.9	0.1817
34	2	0.0	0.0002
35	698	0.4	0.0298
41	42,892	22.6	0.4851
42	6,251	3.3	0.1624
44	22	0.0	0.0015
45	770	0.4	0.0323
51	9,915	5.2	0.2227
52	1,459	0.8	0.0541
53	232	0.1	0.0118
54	504	0.3	0.0228
55	588	0.3	0.0259
61	12,716	6.7	0.2615
62	4,154	2.2	0.1208
63	144	0.1	0.0079
64	629	0.3	0.0273
65	688	0.4	0.0294
71	1,628	0.9	0.0590
72	352	0.2	0.0169
73	319	0.2	0.0155
74	26	0.0	0.0018
75	23	0.0	0.0016
81	322	0.2	0.0156
82	65	0.0	0.0039
83	81	0.0	0.0048
84	9	0.0	0.0007

<부표 B-6> 가구구분의 Entropy

(단 위 : 가 구 , % )

가 구 구 분	빈 도 수	구 성 비	Entropy
계	71,091	100.0	0.9630
1	53,085	74.7	0.3146
2	266	0.4	0.0302
3	16,054	22.6	0.4848
4	1,624	2.3	0.1245
5	62	0.1	0.0089

<부표 B-7> 거주층의 Entropy

(단 위 : 가 구 , % )

거 주 층	빈 도 수	구 성 비	Entropy
계	71,091	100.0	0.0382
1	230	0.3	0.0268
2	70,822	99.6	0.0054
3	39	0.1	0.0059

<부표 B-8> 주인가구여부의 Entropy

(단 위 : 가 구 , % )

주 인 가 구 여 부	빈 도 수	구 성 비	Entropy
계	71,091	100.0	1.6045
11	4,086	5.7	0.2369
12	43,983	61.9	0.4286
21	2,665	3.7	0.1776
22	14,664	20.6	0.4698
30	5,693	8.0	0.2917

<부표 B-9> 점유형태의 Entropy

(단 위 : 가 구 , % )

점 유 형 태	빈 도 수	구 성 비	Entropy
계	71,091	100.0	1.8059
11	46,661	65.6	0.3987
12	8,095	11.4	0.3569
13	5,851	8.2	0.2965
14	1,984	2.8	0.1441
15	1,467	2.1	0.1155
16	4,463	6.3	0.2507
21	1,408	2.0	0.1121
22	229	0.3	0.0267
23	429	0.6	0.0445
24	165	0.2	0.0203
25	55	0.1	0.0080
26	284	0.4	0.0318

<부표 B-10> 거처의 종류의 Entropy

(단 위 : 가 구 , % )

거 처 종 류	빈 도 수	구 성 비	Entropy
계	65,398	100.0	1.5397
11	37,928	58.0	0.4558
12	913	1.4	0.0860
13	365	0.6	0.0418
2	21,277	32.5	0.5270
3	1,999	3.1	0.1538
4	1,430	2.2	0.1206
5	1,095	1.7	0.0988
6	123	0.2	0.0170
7	38	0.1	0.0062
8	17	0.0	0.0031
9	50	0.1	0.0079
10	163	0.2	0.0216

<부표 B-11> 연건평의 Entropy

(단위 : 평대, %)

연건평	빈도수	구성비	Entropy
계	65,130	100.0	1.8682
10	6,779	10.4	0.3397
20	29,070	44.6	0.5194
30	23,618	36.3	0.5307
40	3,315	5.1	0.2187
50	1,176	1.8	0.1046
60	513	0.8	0.0550
70	201	0.3	0.0257
80	92	0.1	0.0134
90	79	0.1	0.0118
100	55	0.1	0.0086
110	22	0.0	0.0039
120	43	0.1	0.0070
130	23	0.0	0.0040
140	26	0.0	0.0045
150	37	0.1	0.0061
160	7	0.0	0.0014
170	7	0.0	0.0014
180	13	0.0	0.0025
190	13	0.0	0.0025
200	24	0.0	0.0042
210 ~	17	0.0	0.0031

<부표 B-12> 대지면적의 Entropy

(단 위 : 평 대 , % )

대 지 면 적	빈 도 수	구 성 비	Entropy
계	65,130	100.0	1.8682
10	160	0.4	0.0324
20	955	2.4	0.1305
30	2,530	6.5	0.2551
40	2,799	7.1	0.2719
50	4,830	12.3	0.3722
60	2,663	6.8	0.2635
70	2,729	7.0	0.2676
80	2,879	7.3	0.2767
90	1,383	3.5	0.1702
100	5,220	13.3	0.3873
110	909	2.3	0.1259
120	1,566	4.0	0.1856
130	960	2.4	0.1310
140	523	1.3	0.0831
150	2,339	6.0	0.2426
160	456	1.2	0.0747
170	460	1.2	0.0752
180	665	1.7	0.0998
190	299	0.8	0.0536
200	2,382	6.1	0.2455
210	170	0.4	0.0340
220	150	0.4	0.0307
230	143	0.4	0.0295
240	102	0.3	0.0223
250	337	0.9	0.0590
260	69	0.2	0.0161
270	101	0.3	0.0222
280	110	0.3	0.0238
290	59	0.2	0.0141
300	551	1.4	0.0865
310	25	0.1	0.0068
320	45	0.1	0.0112
330	28	0.1	0.0075
340	22	0.1	0.0061
350	67	0.2	0.0157
360	18	0.0	0.0051
370	27	0.1	0.0072
380	27	0.1	0.0072
390	19	0.0	0.0053
400	186	0.5	0.0366
410 ~	243	0.6	0.0455

## C. Key 변수의 Uniqueness

### <부표 C-1> 모집단(전수조사)의 Uniqueness

Key 변수의 조합			총청남도		천안시	
			1인 경우	유효조합 수	1인 경우	유효조합 수
1개 변수	1	S	-	2	-	2
	2	A	2	111	1	105
	3	R	-	14	-	14
	4	M	-	4	-	4
	5	HT	-	5	-	5
	6	PT	-	12	-	12
	7	HO	-	3	-	3
	8	PJ	-	10	-	10
2개 변수	1	S×A	5	217	3	206
	2	S×R	-	28	-	28
	3	S×M	-	10	-	10
	4	S×HT	-	10	-	10
	5	S×PT	-	24	-	24
	6	S×HO	-	6	-	6
	7	S×PJ	-	20	-	19
	8	A×R	65	1,037	101	916
	9	A×M	11	335	16	303
	10	A×HT	29	414	31	360
	11	A×PT	63	963	77	848
	12	A×HO	10	274	11	253
	13	A×PJ	71	717	77	566
	14	R×M	-	55	-	53
	15	R×HT	-	5	-	5
	16	R×PT	-	12	-	12
	17	R×HO	-	3	-	3
	18	R×PJ	-	10	-	10
	19	M×HT	-	22	-	21
	20	M×PT	1	54	3	51
	21	M×HO	-	15	1	14
	22	M×PJ	2	43	2	39
	23	HT×PT	-	60	4	58
	24	HT×HO	-	15	-	15

Key 변수의 조합			충청남도		전안시	
			1인 경우	유효조합 수	1인 경우	유효조합 수
2개 변수	25	HT×PJ	1	47	-	43
	26	PT×HO	-	22	-	22
	27	PT×PJ	9	116	10	102
	28	HO×PJ	-	20	1	20
3개 변수	1	S×A×R	167	1,917	206	1,631
	2	S×A×M	30	647	34	576
	3	S×A×HT	72	778	67	655
	4	S×A×PT	147	1,841	207	1,551
	5	S×A×HO	22	532	24	489
	6	S×A×PJ	151	1,282	136	996
	7	S×R×M	2	108	5	102
	8	S×R×HT	-	10	-	10
	9	S×R×PT	-	24	-	24
	10	S×R×HO	-	6	-	6
	11	S×R×PJ	-	20	-	19
	12	S×M×HT	-	44	1	39
	13	S×M×PT	6	106	3	99
	14	S×M×HO	2	30	1	26
	15	S×M×PJ	2	84	3	72
	16	S×HT×PT	5	120	6	109
	17	S×HT×HO	-	30	-	29
	18	S×HT×PJ	3	91	5	80
	19	S×PT×HO	-	44	-	44
	20	S×PT×PJ	22	213	19	175
	21	S×HO×PJ	-	40	1	38
	22	A×R×M	349	2,337	379	1,823
	23	A×R×HT	29	414	31	360
	24	A×R×PT	63	963	77	848
	25	A×R×HO	10	274	11	253
	26	A×R×PJ	71	717	77	566
	27	A×M×HT	154	1,170	139	939
	28	A×M×PT	330	2,843	441	2,222

Key 변수의 조합			충청남도		전안시	
			1인 경우	유효조합 수	1인 경우	유효조합 수
3개 변수	29	A×M×HO	57	888	62	782
	30	A×M×PJ	300	1,911	274	1,437
	31	A×HT×PT	544	3,157	514	2,339
	32	A×HT×HO	117	1,069	105	895
	33	A×HT×PJ	473	2,221	387	1,604
	34	A×PT×HO	139	1,667	181	1,412
	35	A×PT×PJ	1,029	4,102	863	2,803
	36	A×HO×PJ	162	1,285	160	988
	37	R×M×HT	-	22	-	21
	38	R×M×PT	1	54	3	51
	39	R×M×HO	-	15	1	14
	40	R×M×PJ	2	43	2	39
	41	R×HT×PT	-	60	4	58
	42	R×HT×HO	-	15	-	15
	43	R×HT×PJ	1	47	-	43
	44	R×PT×HO	-	22	-	22
	45	R×PT×PJ	9	116	10	102
	46	R×HO×PJ	-	20	1	20
	47	M×HT×PT	24	238	22	207
	48	M×HT×HO	3	66	2	59
	49	M×HT×PJ	16	172	12	142
	50	M×PT×HO	1	94	3	91
	51	M×PT×PJ	55	386	50	304
	52	M×HO×PJ	3	83	6	74
	53	HT×PT×HO	3	109	7	101
	54	HT×PT×PJ	64	396	69	317
	55	HT×HO×PJ	6	90	6	80
	56	PT×HO×PJ	9	116	10	102
4개 변수	1	S×A×R×M	713	4,034	685	3,039
	2	S×A×R×HT	72	778	67	655
	3	S×A×R×PT	147	1,841	207	1,551
	4	S×A×R×HO	22	532	24	489

Key 변수의 조합			충청남도		전안시	
			1인 경우	유효조합 수	1인 경우	유효조합 수
4개 변수	5	S×A×R×PJ	151	1,282	136	996
	6	S×A×M×HT	304	2,054	282	1,595
	7	S×A×M×PT	836	4,987	887	3,600
	8	S×A×M×HO	120	1,657	134	1,450
	9	S×A×M×PJ	601	3,204	500	2,343
	10	S×A×PT×PT	1,086	5,370	996	3,818
	11	S×A×PT×HO	248	1,956	248	1,603
	12	S×A×PT×PJ	840	3,646	694	2,581
	13	S×A×PT×HO	338	3,129	439	2,494
	14	S×A×PT×PJ	1,833	6,645	1,450	4,324
	15	S×A×HO×PJ	330	2,259	294	1,727
	16	S×R×M×HT	-	44	1	39
	17	S×R×M×PT	6	106	3	99
	18	S×R×M×HO	2	30	1	26
	19	S×R×M×PJ	2	84	3	72
	20	S×R×HT×PT	5	120	7	107
	21	S×R×HT×HO	-	30	-	29
	22	S×R×HT×PJ	3	91	5	80
	23	S×R×PT×HO	-	44	-	44
	24	S×R×PT×PJ	22	213	19	175
	25	S×R×HO×PJ	-	40	1	38
	26	S×M×HT×PT	54	441	46	367
	27	S×M×HT×HO	9	129	4	110
	28	S×M×HT×PJ	37	312	34	250
	29	S×M×PT×HO	6	186	4	179
	30	S×M×PT×PJ	108	661	103	509
	31	S×M×HO×PJ	14	161	15	138
	32	S×HT×PT×HO	12	212	19	191
	33	S×HT×PT×PJ	119	680	129	524
	34	S×HT×HO×PJ	11	168	17	144
	35	S×PT×HO×PJ	22	213	19	175
	36	A×R×M×HT	154	1,168	139	939

Key 변수의 조합			충청남도		전안시	
			1인 경우	유효조합 수	1인 경우	유효조합 수
4개 변수	37	A×R×M×PT	330	2,843	441	2,222
	38	A×R×M×HO	57	888	62	792
	39	A×R×M×PJ	300	1,911	274	1,437
	40	A×R×HT×PT	544	3,157	514	2,339
	41	A×R×HT×HO	117	1,069	105	895
	42	A×R×HT×PJ	473	2,221	387	1,604
	43	A×R×PT×HO	139	1,667	181	1,412
	44	A×R×PT×PJ	1,029	4,102	863	2,803
	45	A×R×HO×PJ	162	1,285	160	988
	46	A×M×HT×PT	1,766	7,107	1,484	4,706
	47	A×M×HT×HO	480	2,825	454	2,192
	48	A×M×HT×PJ	1,309	4,774	1,012	3,204
	49	A×M×PT×HO	782	4,661	821	3,409
	50	A×M×PT×PJ	2,679	8,747	1,979	5,359
	51	A×M×HO×PJ	592	3,291	519	2,421
	52	A×HT×PT×HO	958	4,864	931	3,523
	53	A×HT×PT×PJ	3,022	9,043	2,209	5,445
	54	A×HT×HO×PJ	853	3,591	673	2,500
	55	A×PT×HO×PJ	1,029	4,102	863	2,803
	56	R×M×HT×PT	24	238	22	207
	57	R×M×HT×HO	3	66	2	59
	58	R×M×HT×PJ	16	172	12	142
	59	R×M×PT×HO	1	94	3	91
	60	R×M×PT×PJ	55	386	50	304
	61	R×M×HO×PJ	3	83	6	74
	62	R×HT×PT×HO	3	109	7	101
	63	R×HT×PT×PJ	64	396	69	317
	64	R×HT×HO×PJ	6	90	6	80
65	R×PT×HO×PJ	9	116	10	102	
66	M×HT×PT×HO	44	401	49	342	
67	M×HT×PT×PJ	231	1,093	220	797	
68	M×HT×HO×PJ	32	305	33	252	

Key 변수의 조합			충청남도		전안시	
			1인 경우	유효조합 수	1인 경우	유효조합 수
4개 변수	69	M×PT×HO×PJ	55	386	50	304
	70	HT×PT×HO×PJ	64	396	69	317
5개 변수	1	S×A×R×M×HT	304	2,054	282	1,595
	2	S×A×R×M×PT	836	4,987	887	3,600
	3	S×A×R×M×HO	120	1,657	134	1,450
	4	S×A×R×M×PJ	601	3,204	500	2,343
	5	S×A×R×HT×PT	1,086	5,370	996	3,818
	6	S×A×R×HT×HO	248	1,956	248	1,603
	7	S×A×R×HT×PJ	840	3,646	694	2,581
	8	S×A×R×PT×HO	338	3,129	439	2,494
	9	S×A×R×PT×PJ	1,833	6,645	1,450	4,324
	10	S×A×R×HO×PJ	330	2,259	294	1,727
	11	S×A×M×HT×PT	3,182	11,114	2,492	7,002
	12	S×A×M×HT×HO	965	4,812	835	3,601
	13	S×A×M×HT×PJ	2,167	7,383	1,705	4,778
	14	S×A×M×PT×HO	1,672	7,789	1,570	5,307
	15	S×A×M×PT×PJ	4,467	12,902	3,095	7,475
	16	S×A×M×HO×PT	1,171	5,419	981	3,831
	17	S×A×HT×PT×HO	1,888	8,112	1,709	5,571
	18	S×A×HT×PT×PJ	4,933	13,421	3,384	7,736
	19	S×A×HT×HO×PJ	1,477	5,804	1,203	3,959
	20	S×A×PT×HO×PJ	1,833	6,645	1,450	4,324
	21	S×R×M×HT×PT	54	441	46	367
	22	S×R×M×HT×HO	9	129	4	110
	23	S×R×M×HT×PJ	37	312	34	250
	24	S×R×M×PT×HO	6	186	4	179
	25	S×R×M×PT×PJ	108	661	103	509
	26	S×R×M×HO×PJ	14	161	15	138
	27	S×R×HT×PT×HO	12	212	19	191
	28	S×R×HT×PT×PJ	119	680	129	524
	29	S×R×HT×HO×PJ	11	168	17	144
	30	S×R×PT×HO×PJ	22	213	19	175

Key 변수의 조합			충청남도		전안시	
			1인 경우	유효조합 수	1인 경우	유효조합 수
5개 변수	31	S×M×HT×PT×HO	92	726	94	599
	32	S×M×HT×PT×PJ	407	1,755	370	1,226
	33	S×M×HT×HO×PJ	71	544	75	432
	34	S×M×PT×HO×PJ	108	661	103	509
	35	S×HT×PT×HO×PJ	119	680	129	524
	36	A×R×M×HT×PT	1,766	7,107	1,484	4,706
	37	A×R×M×HT×HO	480	2,825	454	2,192
	38	A×R×M×HT×PJ	1,309	4,774	1,012	3,204
	39	A×R×M×PT×HO	782	4,661	821	3,409
	40	A×R×M×PT×PJ	2,679	8,747	1,979	5,359
	41	A×R×M×HO×PJ	592	3,291	519	2,421
	42	A×R×HT×PT×HO	958	4,864	931	3,523
	43	A×R×HT×PT×PJ	3,022	9,043	2,209	5,445
	44	A×R×HT×HO×PJ	853	3,591	673	2,500
	45	A×R×PT×HO×PJ	1,029	4,102	863	2,803
	46	A×M×HT×PT×HO	2,925	10,372	2,354	6,624
	47	A×M×HT×PT×PJ	6,455	16,014	4,002	8,663
	48	A×M×HT×HO×PJ	2,178	7,477	1,649	4,818
	49	A×M×PT×HO×PJ	2,679	8,747	1,979	5,359
	50	A×HT×PT×HO×PJ	3,022	9,043	2,209	5,445
	51	R×M×HT×PT×HO	44	401	49	342
	52	R×M×HT×PT×PJ	231	1,093	220	797
	53	R×M×HT×HO×PJ	32	305	33	252
	54	R×M×PT×HO×PJ	55	386	50	304
	55	R×HT×PT×HO×PJ	64	396	69	317
	56	M×HT×PT×HO×PJ	231	1,093	220	797
6개 변수	1	S×A×R×M×HT×PT	3,182	11,114	2,492	7,002
	2	S×A×R×M×HT×HO	965	4,812	835	3,601
	3	S×A×R×M×HT×PJ	2,167	7,383	1,705	4,778
	4	S×A×R×M×PT×HO	1,672	7,789	1,570	5,307
	5	S×A×R×M×PT×PJ	4,467	12,902	3,095	7,475
	6	S×A×R×M×HO×PJ	1,171	5,419	981	3,831

Key 변수의 조합			충청남도		전안시		
			1인 경우	유효조합 수	1인 경우	유효조합 수	
6개 변수	7	S×A×R×HT×PT×HO	1,888	8,112	1,709	5,571	
	8	S×A×R×HT×PT×PJ	4,933	13,421	3,384	7,736	
	9	S×A×R×HT×HO×PJ	1,477	5,804	1,203	3,959	
	10	S×A×R×PT×HO×PJ	1,833	6,645	1,450	4,324	
	11	S×A×M×HT×PT×HO	5,099	15,718	3,773	9,521	
	12	S×A×M×HT×PT×PJ	9,664	22,110	5,719	11,461	
	13	S×A×M×HT×HO×PJ	3,667	11,339	2,709	7,009	
	14	S×A×M×PT×HO×PJ	4,467	2,902	3,095	7,475	
	15	S×A×HT×PT×HO×PJ	4,933	13,421	3,384	7,736	
	16	S×R×M×HT×PT×HO	92	726	94	599	
	17	S×R×M×HT×PT×PJ	407	1,755	370	1,226	
	18	S×R×M×HT×HO×PJ	71	544	74	432	
	19	S×R×M×PT×HO×PJ	108	661	103	509	
	20	S×R×HT×PT×HO×PJ	119	680	129	524	
	21	S×M×HT×PT×HO×PJ	407	1,755	370	1,226	
	22	A×R×M×HT×PT×HO	2,925	10,372	2,354	6,624	
	23	A×R×M×HT×PT×PJ	6,455	16,014	4,002	8,663	
	24	A×R×M×HT×HO×PJ	2,178	7,477	1,649	4,818	
	25	A×R×M×PT×HO×PJ	2,579	8,747	1,979	5,359	
	26	A×R×HT×PT×HO×PJ	3,022	9,043	2,209	5,445	
	27	A×M×HT×PT×HO×PJ	6,455	16,014	4,002	8,663	
	28	R×M×HT×PT×HO×PJ	231	1,093	220	797	
	7개 변수	1	S×A×R×M×HT×PT×HO	5,099	15,718	3,773	9,521
		2	S×A×R×M×HT×PT×PJ	9,664	22,110	5,719	11,461
		3	S×A×R×M×HT×HO×PJ	3,667	11,339	2,708	7,009
		4	S×A×R×M×PT×HO×PJ	4,467	12,902	3,095	7,475
		5	S×A×R×HT×PT×HO×PJ	4,933	13,421	3,384	7,736
		6	S×A×M×HT×PT×HO×PJ	9,664	22,110	5,719	11,461
7		S×R×M×HT×PT×HO×PJ	407	1,755	370	1,226	
8		A×R×M×HT×PT×HO×PJ	6,455	16,014	4,002	8,663	
8개 변수	1	S×A×R×M×HT×PT×HO×PJ	9,664	22,069	5,719	11,461	

주) S(성별), A(나이), R(가구주와의 관계), M(결혼상태), HT(가구구분), PT(점유형태), HO(주택소유), PJ(거처종류)

<부표 C-2> 2% 표본자료의 Uniqueness

Key 변수의 조합			충청남도		전안시	
			1인 경우	유효조합 수	1인 경우	유효조합 수
1개 변수	1	S	-	2	-	2
	2	A	1	101	3	96
	3	R	-	14	1	14
	4	M	-	4	-	4
	5	HT	-	5	-	5
	6	PT	-	12	1	12
	7	HO	-	3	-	3
	8	PJ	-	10	-	6
2개 변수	1	S×A	1	196	4	183
	2	S×R	1	28	2	28
	3	S×M	-	8	-	8
	4	S×HT	-	10	2	10
	5	S×PT	1	24	2	21
	6	S×HO	-	6	-	6
	7	S×PJ	1	20	-	12
	8	A×R	123	630	120	426
	9	A×M	20	259	35	219
	10	A×HT	49	262	30	184
	11	A×PT	102	582	115	349
	12	A×HO	12	223	24	189
	13	A×PJ	79	377	69	266
	14	R×M	6	46	3	34
	15	R×HT	-	5	-	5
	16	R×PT	-	12	1	12
	17	R×HO	-	3	-	3
	18	R×PJ	-	10	-	6
	19	M×HT	2	20	3	17
	20	M×PT	2	46	8	40
	21	M×HO	-	12	-	12
	22	M×PJ	6	37	1	23
	23	HT×PT	5	49	6	36
	24	HT×HO	-	14	1	13

Key 변수의 조합			충청남도		전안시	
			1인 경우	유효조합 수	1인 경우	유효조합 수
2개 변수	25	HT×PJ	6	38	3	21
	26	PT×HO	1	22	2	21
	27	PT×PJ	12	65	9	42
	28	HO×PJ	-	17	-	11
3개 변수	1	S×A×R	242	1,012	217	647
	2	S×A×M	41	481	76	371
	3	S×A×HT	78	439	67	308
	4	S×A×PT	250	974	199	528
	5	S×A×HO	34	422	62	331
	6	S×A×PJ	145	605	154	411
	7	S×R×M	11	82	10	62
	8	S×R×HT	-	10	2	10
	9	S×R×PT	1	24	2	21
	10	S×R×HO	-	6	-	6
	11	S×R×PJ	1	20	-	12
	12	S×M×HT	4	36	7	31
	13	S×M×PT	7	87	12	65
	14	S×M×HO	-	24	-	24
	15	S×M×PJ	10	60	6	45
	16	S×HT×PT	12	86	13	60
	17	S×HT×HO	1	27	6	25
	18	S×HT×PJ	15	64	5	37
	19	S×PT×HO	2	43	4	35
	20	S×PT×PJ	24	110	17	70
	21	S×HO×PJ	1	33	-	22
	22	A×R×M	293	991	224	590
	23	A×R×HT	49	262	30	184
	24	A×R×PT	102	582	115	349
	25	A×R×HO	12	223	24	187
	26	A×R×PJ	79	377	69	266
	27	A×M×HT	124	568	140	365
	28	A×M×PT	398	1,198	283	603

Key 변수의 조합			충청남도		전안시	
			1인 경우	유효조합 수	1인 경우	유효조합 수
3개 변수	29	A×M×HO	89	595	139	422
	30	A×M×PJ	225	772	206	472
	31	A×HT×PT	388	1,158	259	579
	32	A×HT×HO	128	579	114	389
	33	A×HT×PJ	269	748	182	430
	34	A×PT×HO	220	833	163	440
	35	A×PT×PJ	474	1,196	272	551
	36	A×HO×PJ	133	607	127	391
	37	R×M×HT	2	20	3	17
	38	R×M×PT	2	46	8	40
	39	R×M×HO	-	12	-	12
	40	R×M×PJ	6	37	1	23
	41	R×HT×PT	5	49	6	36
	42	R×HT×HO	-	14	1	13
	43	R×HT×PJ	6	38	3	21
	44	R×PT×HO	1	22	2	21
	45	R×PT×PJ	12	65	9	42
	46	R×HO×PJ	-	17	-	11
	47	M×HT×PT	29	142	30	88
	48	M×HT×HO	10	51	8	39
	49	M×HT×PJ	24	95	16	60
	50	M×PT×HO	6	80	14	62
	51	M×PT×PJ	40	176	34	102
	52	M×HO×PJ	6	61	5	43
	53	HT×PT×HO	10	80	11	56
	54	HT×PT×PJ	53	166	31	89
	55	HT×HO×PJ	8	58	7	34
	56	PT×HO×PJ	12	65	9	42
4개 변수	1	S×A×R×M	473	1,478	355	819
	2	S×A×R×HT	78	439	67	307
	3	S×A×R×PT	250	974	199	528
	4	S×A×R×HO	34	422	62	330

Key 변수의 조합			충청남도		천안시	
			1인 경우	유효조합 수	1인 경우	유효조합 수
4개 변수	5	S×A×R×PJ	145	605	154	410
	6	S×A×M×HT	242	903	237	513
	7	S×A×M×PT	699	1,669	425	785
	8	S×A×M×HO	274	979	265	592
	9	S×A×M×PJ	381	1,094	326	599
	10	S×A×PT×PT	675	1,673	419	791
	11	S×A×PT×HO	243	949	241	587
	12	S×A×PT×PJ	387	1,061	296	578
	13	S×A×PT×HO	464	1,299	279	643
	14	S×A×PT×PJ	774	1,657	402	730
	15	S×A×HO×PJ	260	947	237	560
	16	S×R×M×HT	4	36	7	30
	17	S×R×M×PT	7	87	12	65
	18	S×R×M×HO	-	24	-	24
	19	S×R×M×PJ	10	60	6	45
	20	S×R×HT×PT	12	86	13	60
	21	S×R×HT×HO	1	27	6	25
	22	S×R×HT×PJ	15	64	5	37
	23	S×R×PT×HO	2	43	4	35
	24	S×R×PT×PJ	24	110	17	70
	25	S×R×HO×PJ	1	33	-	22
	26	S×M×HT×PT	59	229	40	133
	27	S×M×HT×HO	15	87	12	68
	28	S×M×HT×PJ	36	151	33	99
	29	S×M×PT×HO	20	144	22	100
	30	S×M×PT×PJ	75	269	60	155
	31	S×M×HO×PJ	12	100	10	75
	32	S×HT×PT×HO	26	136	19	90
	33	S×HT×PT×PJ	81	254	48	131
	34	S×HT×HO×PJ	21	99	13	59
	35	S×PT×HO×PJ	24	110	17	70
	36	A×R×M×HT	124	568	140	365

Key 변수의 조합			충청남도		전안시	
			1인 경우	유효조합 수	1인 경우	유효조합 수
4개 변수	37	A×R×M×PT	398	1,198	283	603
	38	A×R×M×HO	89	595	139	422
	39	A×R×M×PJ	225	772	206	472
	40	A×R×HT×PT	388	1,158	259	579
	41	A×R×HT×HO	128	579	114	389
	42	A×R×HT×PJ	269	748	182	430
	43	A×R×PT×HO	220	833	164	440
	44	A×R×PT×PJ	474	1,196	272	551
	45	A×R×HO×PJ	133	607	127	391
	46	A×M×HT×PT	862	1,861	464	799
	47	A×M×HT×HO	369	1,143	307	620
	48	A×M×HT×PJ	515	1,251	347	613
	49	A×M×PT×HO	624	1,507	363	708
	50	A×M×PT×PJ	939	1,911	459	782
	51	A×M×HO×PJ	395	1,194	307	621
	52	A×HT×PT×HO	609	1,492	343	695
	53	A×HT×PT×PJ	922	1,804	429	739
	54	A×HT×HO×PJ	404	1,096	285	584
	55	A×PT×HO×PJ	474	1,196	272	551
	56	R×M×HT×PT	29	142	30	88
	57	R×M×HT×HO	10	51	8	39
	58	R×M×HT×PJ	24	95	16	60
	59	R×M×PT×HO	6	80	14	62
	60	R×M×PT×PJ	40	176	34	102
	61	R×M×HO×PJ	6	61	5	43
	62	R×HT×PT×HO	10	80	11	56
	63	R×HT×PT×PJ	53	166	31	89
	64	R×HT×HO×PJ	8	58	6	28
65	R×PT×HO×PJ	12	65	9	42	
66	M×HT×PT×HO	52	213	40	126	
67	M×HT×PT×PJ	131	353	77	174	
68	M×HT×HO×PJ	30	151	27	92	

Key 변수의 조합			충청남도		전안시	
			1인 경우	유효조합 수	1인 경우	유효조합 수
4개 변수	69	M×PT×HO×PJ	40	176	34	102
	70	HT×PT×HO×PJ	53	166	31	89
5개 변수	1	S×A×R×M×HT	242	903	237	513
	2	S×A×R×M×PT	699	1,669	425	785
	3	S×A×R×M×HO	274	979	265	592
	4	S×A×R×M×PJ	381	1,094	326	599
	5	S×A×R×HT×PT	675	1,673	419	791
	6	S×A×R×HT×HO	243	949	241	587
	7	S×A×R×HT×PJ	387	1,061	296	578
	8	S×A×R×PT×HO	464	1,299	279	643
	9	S×A×R×PT×PJ	774	1,657	402	730
	10	S×A×R×HO×PJ	260	947	237	560
	11	S×A×M×HT×PT	1,262	2,386	617	961
	12	S×A×M×HT×HO	689	1,643	460	785
	13	S×A×M×HT×PJ	799	1,661	467	728
	14	S×A×M×PT×HO	988	2,001	526	896
	15	S×A×M×PT×PJ	1,280	2,323	588	904
	16	S×A×M×HO×PT	678	1,600	435	749
	17	S×A×HT×PT×HO	1,003	2,034	536	911
	18	S×A×HT×PT×PJ	1,256	2,274	576	893
	19	S×A×HT×HO×PJ	630	1,534	411	737
	20	S×A×PT×HO×PJ	774	1,657	402	730
	21	S×R×M×HT×PT	59	229	40	133
	22	S×R×M×HT×HO	15	87	12	68
	23	S×R×M×HT×PJ	36	151	33	99
	24	S×R×M×PT×HO	20	144	22	100
	25	S×R×M×PT×PJ	75	269	60	155
	26	S×R×M×HO×PJ	12	100	10	75
	27	S×R×HT×PT×HO	26	136	19	90
	28	S×R×HT×PT×PJ	81	254	48	131
	29	S×R×HT×HO×PJ	21	99	13	59
	30	S×R×PT×HO×PJ	24	110	17	70

Key 변수의 조합			충청남도		전안시	
			1인 경우	유효조합 수	1인 경우	유효조합 수
5개 변수	31	S×M×HT×PT×HO	98	334	62	190
	32	S×M×HT×PT×PJ	213	497	112	241
	33	S×M×HT×HO×PJ	52	237	44	142
	34	S×M×PT×HO×PJ	75	269	60	155
	35	S×HT×PT×HO×PJ	81	254	48	131
	36	A×R×M×HT×PT	862	1,861	464	799
	37	A×R×M×HT×HO	369	1,143	307	620
	38	A×R×M×HT×PJ	515	1,251	347	613
	39	A×R×M×PT×HO	624	1,496	363	708
	40	A×R×M×PT×PJ	939	1,911	459	782
	41	A×R×M×HO×PJ	395	1,194	308	621
	42	A×R×HT×PT×HO	609	1,492	343	695
	43	A×R×HT×PT×PJ	922	1,804	429	739
	44	A×R×HT×HO×PJ	404	1,096	285	584
	45	A×R×PT×HO×PJ	474	1,196	272	551
	46	A×M×HT×PT×HO	1,153	2,184	566	906
	47	A×M×HT×PT×PJ	1,442	2,465	609	900
	48	A×M×HT×HO×PJ	836	1,755	464	757
	49	A×M×PT×HO×PJ	939	1,911	459	782
	50	A×HT×PT×HO×PJ	922	1,804	429	739
	51	R×M×HT×PT×HO	52	213	40	126
	52	R×M×HT×PT×PJ	131	353	77	174
	53	R×M×HT×HO×PJ	30	151	27	92
	54	R×M×PT×HO×PJ	40	176	34	102
	55	R×HT×PT×HO×PJ	53	166	31	89
	56	M×HT×PT×HO×PJ	131	353	77	174
6개 변수	1	S×A×R×M×HT×PT	1,262	2,386	617	961
	2	S×A×R×M×HT×HO	689	1,643	460	785
	3	S×A×R×M×HT×PJ	799	1,661	467	728
	4	S×A×R×M×PT×HO	988	2,001	526	896
	5	S×A×R×M×PT×PJ	1,280	2,323	588	904
	6	S×A×R×M×HO×PJ	678	1,600	435	749

Key 변수의 조합			충청남도		전안시		
			1인 경우	유효조합 수	1인 경우	유효조합 수	
6개 변수	7	S×A×R×HT×PT×HO	1,003	2,034	536	911	
	8	S×A×R×HT×PT×PJ	1,256	2,274	576	893	
	9	S×A×R×HT×HO×PJ	630	1,534	411	737	
	10	S×A×R×PT×HO×PJ	774	1,657	402	730	
	11	S×A×M×HT×PT×HO	1,563	2,693	576	893	
	12	S×A×M×HT×PT×PJ	1,810	2,873	709	995	
	13	S×A×M×HT×HO×PJ	1,196	2,194	570	863	
	14	S×A×M×PT×HO×PJ	1,280	2,323	588	904	
	15	S×A×HT×PT×HO×PJ	1,256	2,274	726	1,063	
	16	S×R×M×HT×PT×HO	98	334	62	190	
	17	S×R×M×HT×PT×PJ	213	497	112	241	
	18	S×R×M×HT×HO×PJ	52	237	44	142	
	19	S×R×M×PT×HO×PJ	75	269	60	155	
	20	S×R×HT×PT×HO×PJ	81	254	48	131	
	21	S×M×HT×PT×HO×PJ	213	497	112	241	
	22	A×R×M×HT×PT×HO	1,153	2,184	566	906	
	23	A×R×M×HT×PT×PJ	1,442	2,465	609	900	
	24	A×R×M×HT×HO×PJ	836	1,755	464	757	
	25	A×R×M×PT×HO×PJ	939	1,911	459	782	
	26	A×R×HT×PT×HO×PJ	922	1,804	429	739	
	27	A×M×HT×PT×HO×PJ	1,442	2,465	609	900	
	28	R×M×HT×PT×HO×PJ	131	353	77	174	
	7개 변수	1	S×A×R×M×HT×PT×HO	1,563	2,693	726	1,063
		2	S×A×R×M×HT×PT×PJ	1,810	2,873	709	995
		3	S×A×R×M×HT×HO×PJ	1,196	2,194	570	863
		4	S×A×R×M×PT×HO×PJ	1,280	2,323	588	904
		5	S×A×R×HT×PT×HO×PJ	1,256	2,274	576	893
		6	S×A×M×HT×PT×HO×PJ	1,810	2,873	709	995
7		S×R×M×HT×PT×HO×PJ	213	497	112	241	
8		A×R×M×HT×PT×HO×PJ	1,442	2,465	609	900	
8개 변수	1	S×A×R×M×HT×PT×HO×PJ	1,810	2,873	709	995	

주) S(성별), A(나이), R(가구주와의 관계), M(결혼상태), HT(가구구분), PT(점유형태), HO(주택소유), PJ(거처종류)

<부표 C-3> Grouping한 2% 표본의 Uniqueness

Key 변수의 조합			충청남도		천안시	
			1인 경우	유효조합 수	1인 경우	유효조합 수
1개 변수	1	S	-	2	-	2
	2	A	-	21	-	20
	3	R	-	4	-	4
	4	M	-	4	-	4
	5	HT	-	4	-	4
	6	PT	-	8	-	8
	7	HO	-	2	-	2
	8	PJ	-	8	-	4
2개 변수	1	S×A	1	42	2	40
	2	S×R	-	8	-	8
	3	S×M	-	8	-	8
	4	S×HT	-	8	-	8
	5	S×PT	-	16	-	14
	6	S×HO	-	4	-	4
	7	S×PJ	-	8	-	8
	8	A×R	3	70	5	66
	9	A×M	1	46	7	46
	10	A×HT	2	59	5	49
	11	A×PT	6	110	20	90
	12	A×HO	1	34	1	31
	13	A×PJ	2	64	5	58
	14	R×M	-	12	1	12
	15	R×HT	-	4	-	4
	16	R×PT	-	8	-	8
	17	R×HO	-	2	-	2
	18	R×PJ	-	4	-	4
	19	M×HT	-	12	-	12
	20	M×PT	-	24	2	23
	21	M×HO	-	6	-	6
	22	M×PJ	-	12	-	12
	23	HT×PT	2	32	3	26
	24	HT×HO	-	8	-	8

Key 변수의 조합			충청남도		천안시	
			1인 경우	유효조합 수	1인 경우	유효조합 수
2개 변수	25	HT×PJ	-	15	1	16
	26	PT×HO	-	8	-	8
	27	PT×PJ	2	29	4	27
	28	HO×PJ	-	8	-	8
3개 변수	1	S×A×R	5	133	10	122
	2	S×A×M	3	90	10	84
	3	S×A×HT	9	109	17	89
	4	S×A×PT	25	204	37	145
	5	S×A×HO	3	67	4	61
	6	S×A×PJ	6	122	17	105
	7	S×R×M	-	24	1	23
	8	S×R×HT	-	8	-	8
	9	S×R×PT	-	16	-	14
	10	S×R×HO	-	4	-	4
	11	S×R×PJ	-	8	-	8
	12	S×M×HT	1	23	4	23
	13	S×M×PT	1	48	4	39
	14	S×M×HO	-	12	-	12
	15	S×M×PJ	-	24	-	24
	16	S×HT×PT	3	57	8	43
	17	S×HT×HO	-	16	2	16
	18	S×HT×PJ	1	30	3	29
	19	S×PT×HO	-	16	-	14
	20	S×PT×PJ	8	55	5	44
	21	S×HO×PJ	-	16	-	16
	22	A×R×M	5	134	-	12
	23	A×R×HT	2	59	5	49
	24	A×R×PT	6	110	20	90
	25	A×R×HO	1	34	1	31
	26	A×R×PJ	2	64	5	58
	27	A×M×HT	18	132	22	100
	28	A×M×PT	27	234	48	167

Key 변수의 조합			충청남도		천안시	
			1인 경우	유효조합 수	1인 경우	유효조합 수
3개 변수	29	A×M×HO	3	82	8	77
	30	A×M×PJ	12	145	24	122
	31	A×HT×PT	66	289	47	171
	32	A×HT×HO	9	109	12	80
	33	A×HT×PJ	30	184	31	128
	34	A×PT×HO	6	110	20	90
	35	A×PT×PJ	52	285	58	186
	36	A×HO×PJ	11	120	18	102
	37	R×M×HT	-	12	-	12
	38	R×M×PT	-	24	2	23
	39	R×M×HO	-	6	-	6
	40	R×M×PJ	-	12	-	12
	41	R×HT×PT	2	32	3	26
	42	R×HT×HO	-	8	-	8
	43	R×HT×PJ	-	15	1	16
	44	R×PT×HO	-	8	-	8
	45	R×PT×PJ	2	29	4	27
	46	R×HO×PJ	-	8	-	8
	47	M×HT×PT	13	83	16	57
	48	M×HT×HO	2	24	2	21
	49	M×HT×PJ	5	44	9	41
	50	M×PT×HO	-	24	2	23
	51	M×PT×PJ	9	76	14	61
	52	M×HO×PJ	-	24	-	24
	53	HT×PT×HO	2	32	3	26
	54	HT×PT×PJ	17	89	20	65
	55	HT×HO×PJ	1	30	5	27
	56	PT×HO×PJ	2	29	4	27
4개 변수	1	S×A×R×M	17	253	33	206
	2	S×A×R×HT	9	109	17	89
	3	S×A×R×PT	25	204	37	145
	4	S×A×R×HO	3	67	4	61

Key 변수의 조합			충청남도		천안시	
			1인 경우	유효조합 수	1인 경우	유효조합 수
4개 변수	5	S×A×R×PJ	6	122	17	105
	6	S×A×M×HT	34	224	41	162
	7	S×A×M×PT	75	389	75	250
	8	S×A×M×HO	11	154	25	134
	9	S×A×M×PJ	35	251	58	196
	10	S×A×PT×PT	128	460	80	265
	11	S×A×PT×HO	29	193	33	143
	12	S×A×PT×PJ	70	306	57	201
	13	S×A×PT×HO	25	204	37	145
	14	S×A×PT×PJ	106	473	97	277
	15	S×A×HO×PJ	23	221	40	172
	16	S×R×M×HT	1	23	4	23
	17	S×R×M×PT	1	48	4	39
	18	S×R×M×HO	-	12	-	12
	19	S×R×M×PJ	-	24	-	24
	20	S×R×HT×PT	3	57	8	43
	21	S×R×HT×HO	-	16	2	16
	22	S×R×HT×PJ	1	30	3	29
	23	S×R×PT×HO	-	16	-	14
	24	S×R×PT×PJ	8	55	5	44
	25	S×R×HO×PJ	-	16	-	16
	26	S×M×HT×PT	31	141	22	89
	27	S×M×HT×HO	5	45	6	38
	28	S×M×HT×PJ	8	79	18	69
	29	S×M×PT×HO	1	48	4	39
	30	S×M×PT×PJ	18	132	20	98
	31	S×M×HO×PJ	-	47	-	44
	32	S×HT×PT×HO	3	57	8	43
	33	S×HT×PT×PJ	25	148	28	98
	34	S×HT×HO×PJ	4	56	9	47
	35	S×PT×HO×PJ	8	55	5	44
	36	A×R×M×HT	18	132	22	100

Key 변수의 조합			충청남도		천안시	
			1인 경우	유효조합 수	1인 경우	유효조합 수
4개 변수	37	A×R×M×PT	27	234	48	167
	38	A×R×M×HO	3	82	8	77
	39	A×R×M×PJ	12	145	24	122
	40	A×R×HT×PT	66	289	47	171
	41	A×R×HT×HO	9	109	12	80
	42	A×R×HT×PJ	30	184	31	128
	43	A×R×PT×HO	6	110	20	90
	44	A×R×PT×PJ	52	285	58	186
	45	A×R×HO×PJ	11	120	18	102
	46	A×M×HT×PT	141	495	100	282
	47	A×M×HT×HO	42	226	40	159
	48	A×M×HT×PJ	74	337	82	217
	49	A×M×PT×HO	27	234	48	167
	50	A×M×PT×PJ	129	532	111	298
	51	A×M×HO×PJ	35	262	47	191
	52	A×HT×PT×HO	66	289	47	171
	53	A×HT×PT×PJ	197	600	113	296
	54	A×HT×HO×PJ	67	305	60	193
	55	A×PT×HO×PJ	52	285	58	186
	56	R×M×HT×PT	13	83	16	57
	57	R×M×HT×HO	2	24	2	21
	58	R×M×HT×PJ	5	44	9	41
	59	R×M×PT×HO	-	24	2	23
	60	R×M×PT×PJ	9	76	14	61
	61	R×M×HO×PJ	-	24	-	24
	62	R×HT×PT×HO	2	32	3	26
	63	R×HT×PT×PJ	17	89	20	65
	64	R×HT×HO×PJ	1	30	5	27
65	R×PT×HO×PJ	2	29	14	61	
66	M×HT×PT×HO	13	83	16	57	
67	M×HT×PT×PJ	40	191	43	122	
68	M×HT×HO×PJ	7	78	17	64	

Key 변수의 조합			충청남도		천안시	
			1인 경우	유효조합 수	1인 경우	유효조합 수
4개 변수	69	M×PT×HO×PJ	2	29	14	61
	70	HT×PT×HO×PJ	17	89	20	65
5개 변수	1	S×A×R×M×HT	34	224	41	162
	2	S×A×R×M×PT	75	389	75	250
	3	S×A×R×M×HO	11	154	25	134
	4	S×A×R×M×PJ	35	251	58	196
	5	S×A×R×HT×PT	128	460	80	265
	6	S×A×R×HT×HO	29	193	33	143
	7	S×A×R×HT×PJ	70	306	57	201
	8	S×A×R×PT×HO	25	204	37	145
	9	S×A×R×PT×PJ	106	473	97	277
	10	S×A×R×HO×PJ	23	221	40	172
	11	S×A×M×HT×PT	255	742	170	405
	12	S×A×M×HT×HO	85	373	83	252
	13	S×A×M×HT×PJ	165	525	139	312
	14	S×A×M×PT×HO	75	389	75	250
	15	S×A×M×PT×PJ	252	775	174	404
	16	S×A×M×HO×PT	86	428	101	290
	17	S×A×HT×PT×HO	128	460	80	265
	18	S×A×HT×PT×PJ	335	855	182	415
	19	S×A×HT×HO×PJ	131	485	99	288
	20	S×A×PT×HO×PJ	106	473	97	277
	21	S×R×M×HT×PT	31	141	22	89
	22	S×R×M×HT×HO	5	45	6	38
	23	S×R×M×HT×PJ	8	79	18	69
	24	S×R×M×PT×HO	1	48	4	39
	25	S×R×M×PT×PJ	18	132	20	98
	26	S×R×M×HO×PJ	-	47	-	44
	27	S×R×HT×PT×HO	3	57	8	43
	28	S×R×HT×PT×PJ	25	148	28	98
	29	S×R×HT×HO×PJ	4	56	9	47
	30	S×R×PT×HO×PJ	8	55	5	44

Key 변수의 조합			충청남도		천안시	
			1인 경우	유효조합 수	1인 경우	유효조합 수
5개 변수	31	S×M×HT×PT×HO	31	141	22	89
	32	S×M×HT×PT×PJ	85	306	65	178
	33	S×M×HT×HO×PJ	15	138	27	104
	34	S×M×PT×HO×PJ	18	132	20	98
	35	S×HT×PT×HO×PJ	25	148	28	98
	36	A×R×M×HT×PT	141	495	100	282
	37	A×R×M×HT×HO	42	226	40	159
	38	A×R×M×HT×PJ	74	337	82	217
	39	A×R×M×PT×HO	27	234	48	167
	40	A×R×M×PT×PJ	129	532	111	298
	41	A×R×M×HO×PJ	35	262	47	191
	42	A×R×HT×PT×HO	66	289	47	171
	43	A×R×HT×PT×PJ	197	600	113	296
	44	A×R×HT×HO×PJ	67	305	60	193
	45	A×R×PT×HO×PJ	52	285	58	186
	46	A×M×HT×PT×HO	141	495	100	282
	47	A×M×HT×PT×PJ	363	903	200	419
	48	A×M×HT×HO×PJ	161	536	123	303
	49	A×M×PT×HO×PJ	129	532	111	298
	50	A×HT×PT×HO×PJ	197	600	113	296
	51	R×M×HT×PT×HO	13	83	16	57
	52	R×M×HT×PT×PJ	40	191	43	122
	53	R×M×HT×HO×PJ	7	78	17	64
	54	R×M×PT×HO×PJ	9	76	14	61
	55	R×HT×PT×HO×PJ	17	89	20	65
	56	M×HT×PT×HO×PJ	40	191	43	122
6개 변수	1	S×A×R×M×HT×PT	255	742	170	405
	2	S×A×R×M×HT×HO	85	373	83	252
	3	S×A×R×M×HT×PJ	165	525	139	312
	4	S×A×R×M×PT×HO	75	389	75	250
	5	S×A×R×M×PT×PJ	252	775	174	404
	6	S×A×R×M×HO×PJ	86	428	101	290

Key 변수의 조합			충청남도		천안시	
			1인 경우	유효조합 수	1인 경우	유효조합 수
6개 변수	7	S×A×R×HT×PT×HO	128	460	80	265
	8	S×A×R×HT×PT×PJ	335	855	182	415
	9	S×A×R×HT×HO×PJ	131	485	99	288
	10	S×A×R×PT×HO×PJ	106	473	97	277
	11	S×A×M×HT×PT×HO	255	742	170	405
	12	S×A×M×HT×PT×PJ	553	1,197	298	534
	13	S×A×M×HT×HO×PJ	286	779	208	415
	14	S×A×M×PT×HO×PJ	252	775	174	404
	15	S×A×HT×PT×HO×PJ	335	855	182	415
	16	S×R×M×HT×PT×HO	31	141	22	89
	17	S×R×M×HT×PT×PJ	85	306	65	178
	18	S×R×M×HT×HO×PJ	15	138	27	104
	19	S×R×M×PT×HO×PJ	18	132	20	98
	20	S×R×HT×PT×HO×PJ	25	148	28	98
	21	S×M×HT×PT×HO×PJ	85	306	65	178
	22	A×R×M×HT×PT×HO	141	495	100	282
	23	A×R×M×HT×PT×PJ	363	903	200	419
	24	A×R×M×HT×HO×PJ	161	536	123	303
	25	A×R×M×PT×HO×PJ	129	532	111	298
	26	A×R×HT×PT×HO×PJ	197	600	113	296
27	A×M×HT×PT×HO×PJ	363	903	200	419	
28	R×M×HT×PT×HO×PJ	40	191	43	122	
7개 변수	1	S×A×R×M×HT×PT×HO	255	742	170	405
	2	S×A×R×M×HT×PT×PJ	553	1,197	298	534
	3	S×A×R×M×HT×HO×PJ	287	779	208	415
	4	S×A×R×M×PT×HO×PJ	252	775	174	404
	5	S×A×R×HT×PT×HO×PJ	335	855	182	415
	6	S×A×M×HT×PT×HO×PJ	553	1,197	298	534
	7	S×R×M×HT×PT×HO×PJ	85	306	65	178
	8	A×R×M×HT×PT×HO×PJ	363	903	200	419
8개 변수	1	S×A×R×M×HT×PT×HO×PJ	553	1,197	298	534

주) S(성별), A(나이), R(가구주와의 관계), M(결혼상태), HT(가구구분), PT(점유형태), HO(주택소유), PJ(거처종류)