승법잡음모형을 이용한 가계조사 자료의 비밀보호

정동명 · 김경미

목 차

표 목 차

그림 목차

승법잡음모형을 이용한 가계조사 자료의 비밀보호

제1절 서 론

최근 들어 개인정보의 비밀보호가 사회적으로 이슈화되면서 통계청을 비롯한 국가통계 작성기관에서는 마이크로자료(micro data)나 매크로자료(macro data)를 외부에 제공할 때, 사전에 미리 응답자의 개인정보가 노출되지 않도록 적절한 비밀보호방법을 적용하고 있다. 예를 들어, 통계청에서는 통계이용자들을 위해 5년마다 실시하는 인구주택총조사 결과 중 2%를 마이크로자료 파일로 작성하여 그대로 제공하였으나, 지난 2005 인구주택총조사 결과부터는 일부 민감한 변수에 대해 비밀보호방법이 적용된 2% 마이크로자료 파일을 작성하여 제공하였다.

이러한 비밀보호방법은 1970년대 이후부터 영국이나 네덜란드, 미국 등 여러 통계선진국에서 꾸준히 연구되고 있는 분야이며, 자료의 특성에 따라 자료교환(data swapping)이나 그룹화(grouping), 반올림(rounding), 구간그룹화(grouping into intervals) 등 다양한 방법이 개발되어 현재 널리 활용되고 있다. 우리나라에서도 최근에 통계작성기관과 학계에서 이에 대한 연구가 점차 활성화되고 있는데, 특히 통계청에서는 2007년 통계법 개정으로 인한 마이크로자료 제공의무에 따라 응답자의 비밀보호를 위한 연구를 체계적으로 진행하고 있다. 정동명(2007) 등은 2005 인구주택총조사 결과에 그룹화 등의 방법을 적용하여 비밀보호된 마이크로자료 파일

을 작성하는 연구를 하였다. 또한 정동명(2008) 등은 우리나라 가구의 소득과 소비지출의 특성을 파악하는 가계조사의 결과에도 반올림과 구간그룹화, 승법잡음 등의 방법을 적용하여 분석하고 서로 비교해 보았다.

본 연구에서는 연속형 자료의 비밀보호방법에 효과적으로 적용할 수 있는 승법잡음모형에 대해 살펴보고, 이 모형을 가계조사에 적용하여 비밀보호된 마이크로자료를 작성하는 과정을 소개하고자 한다. 2절에서는 원자료의 변환에 이용되는 잡음의 모형을 소개하고, 특히 승법잡음모형에서 활용할 수 있는 4가지 분포와 이들의 특성을 제시한다. 3절에서는 제시된 분포별로 난수프로그램을 작성하여 잡음을 생성하고 그 결과를 비교해본다. 또한 가계조사에서 민감변수로 선정된 항목을 대상으로 자료를 변환하는 과정을 설명하고, 원자료와 승법잡음모형에 의해 변환된 자료를 분석하여 서로 비교해 본다. 마지막으로 본 연구의 결론은 4절에서 살펴보도록 한다.

제2절 잡음을 이용한 비밀보호

1. 변환모형

가. 가법잡음모형

주어진 원자료를 X라 하고, 이에 대응하는 잡음을 e라고 하자. 만약 원자료 X에 적절한 수준의 잡음을 더해주면 변환된 자료 Y가 얻어지는데, 이러한 변환모형을 가법잡음(additive noise)모형이라 하고 다음과 같이 나타낸다.

$$Y = X + e$$

여기서 잡음 e의 분산과 공분산은 원자료 X에 의존된다. 가법잡음모형은 연속형 자료(continuous data)인 경우 효과적으로 활용할 수 있으나, 원자료가 0인 경우 잡음 e의 값에 따라 변환된 Y값이 음수가 될 수 있기 때문에 비밀보호된 자료값이 항상 0이나 양수가 되어야 할 경

우에는 적용하기가 곤란하다. 가법잡음모형의 특성 등에 대한 자세한 내용은 Kim(1986)과 Fuller(1993) 등을 참고하기 바란다.

나. 승법잡음모형

가법잡음모형과 달리 원자료 X에 대응하는 잡음 e를 곱해주어 자료를 변환하는 것을 승법잡음(multiplicative noise)모형이라 하고, 이 때 변환된 자료 Y는 다음과 같이 나타낸다.

$$Y = X \cdot e$$

잡음 e는 일반적으로 1이 아니면서 1을 중심으로 양의 값을 가지는 것이 바람직하다. 이것은 e가 만약 음수가 되면 변환된 변수도 음의 값이 되고, e가 1이 되면 Y는 X와 같게 되어 변환의 의미가 없기 때문이다. 또한 e가 1에서 멀리 떨어진 값을 가지면, 원자료 X와 변환된 자료 Y가 너무 차이가 나서 정보의 손실이 크게 된다. 따라서 e는 1이 아니면서 1과 가까운 값을 가지는 것이 바람직하다.

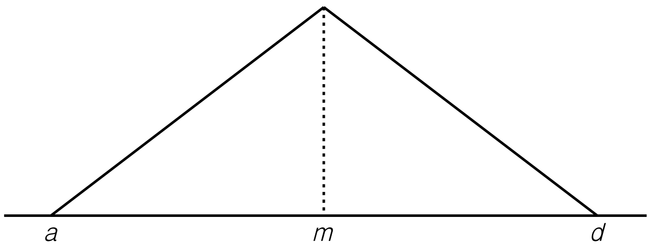
2. 잡음의 분포와 그 특성

잡음 e를 이용하여 자료를 변환시키고자 할 경우, e를 구하는 것이 매우 중요한 요인이 된다. 일반적으로 e를 구하기 위해서는 먼저 e에 대한 적절한 분포를 가정한 후, 컴퓨터를 이용해서 난수(random number)를 발생시키는 프로그램을 작성하여 e의 값을 구한다. 이러한 e에 대한 분포로는 모형에 따라 다양한 형태가 적용될 수 있으나, 이에 대한 연구는 많지 않다. 다만, 승법잡음모형인 경우, Kim & Winkler(2001)는 평균을 중심으로 좌우가 절단된 정규분포(truncated normal distribution)를 적용한 결과를 소개하였다. 또한 Kim(2003) 등은 절단된 삼각분포를 이용하여 승법잡음모형에 의한 자료변환과 이에 대한 특성을 분석하였다.

본 절에서는 승법잡음모형에서 잡음의 분포로 이용할 수 있는 삼각분포와 절단된 삼각분포, 사다리꼴분포, 이중삼각분포 등 모두 4가지 유형의 분포를 소개하고, 이들의 특성을 각각 살펴보고자 한다.

가. 삼각분포

삼각분포(triangular distribution)는 [그림 1]에 나타난바와 같이 m을 중심으로 삼각형 모양으로 이루어진 분포를 말한다. 여기서 a는 자료의 최소값, m은 최빈값(mode), d는 최대값이 된다.



[그림 1] 삼각분포

잡음 e가 최빈값 m을 중심으로 삼각분포를 따른다고 하면, 확률밀도함수(probability density function; pdf), $f^t(e)$는 다음과 같이 나타낼 수 있다.

$$f^t(e) = \begin{cases} \dfrac{2}{(m-a)(d-a)}(e-a), & a \le e < m \\ \\ \dfrac{2}{(d-m)(d-a)}(d-e), & m \le e < d \end{cases}$$

그리고 $f^t(e)$을 이용하여 잡음 e의 기대값 $E^t(e)$와 분산 $Var^t(e)$을 구하면 다음과 같이 나타낼 수 있다.

$$E^t(e) = \frac{(m+a+d)}{3} \tag{1}$$

$$Var^t(e) = \frac{[d^2 + a^2 + m^2 - m(a+d) - ad]}{18}$$

만약 잡음 e가 최빈값 m을 중심으로 좌우대칭인 삼각분포를 따른

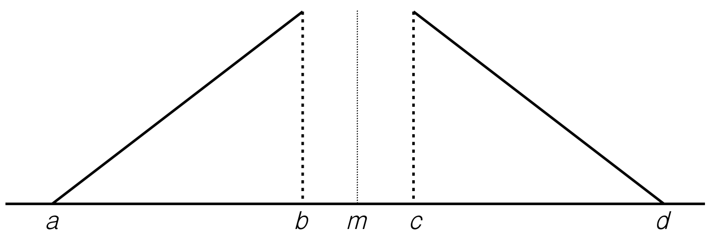
다면 $m-a=d-m$이고 $a+d=2m$이 되므로, 위 식 (1)에서 $E^t(e)$와 $Var^t(e)$는 다음과 같이 간단하게 표현할 수 있다.

$$E^t(e) = m$$

$$Var^t(e) = \frac{(d-m)^2}{6}$$

나. 절단된 삼각분포

절단된 삼각분포(truncated triangular distribution)는 [그림 1]의 삼각분포에서 최빈값 m을 중심으로 좌우로 일부분이 절단된 형태로 이루어진 분포를 말한다. 즉, 절단된 삼각분포는 [그림 2]에 나타난 바와 같이 m을 중심으로 절사점 b와 $c\,(c>b)$에서 각각 절단된 형태를 하고 있어 최빈값을 포함하지 않는다.



[그림 2] 절단된 삼각분포

잡음 e가 최빈값 m을 중심으로 절단된 삼각분포를 따른다면 확률밀도함수(pdf), $f^{tt}(e)$는 다음과 같이 나타낼 수 있다.

$$f^{tt}(e) = \begin{cases} \dfrac{2(d-m)}{(b-a)^2(d-m)+(d-c)^2(m-a)}(e-a), & a \le e < b \\[2ex] \dfrac{2(m-a)}{(b-a)^2(d-m)+(d-c)^2(m-a)}(d-e), & c \le e < d \end{cases}$$

그리고 위의 식 $f^{tt}(e)$ 을 이용하여 잡음 e 의 기대값 $E^{tt}(e)$ 와 분산 $Var^{tt}(e)$ 은 다음과 같이 구할 수 있다.

$$E^{tt}(e) = \frac{[(d-m)(b-a)^2(2b+a)+(m-a)(d-c)^2(2c+d)]}{3[(d-m)(b-a)^2+(m-a)(d-c)^2]} \tag{2}$$

$$\begin{aligned} Var^{tt}(e) = & \frac{1}{18[(b-a)^2(d-m)+(d-c)^2(m-a)]^2}\{(b-a)^6(d-m)^2 \\ & +(d-c)^6(m-a)^2+(b-a)^2(d-m)(d-c)^2(m-a) \\ & [(3b+a)^2+(3c+d)^2+2(a^2+d^2)-4(2b+a)(2c+d)]\} \end{aligned}$$

만약 최빈값 m을 중심으로 좌우대칭이면, $b-a=d-c$가 되므로 위의 식 (2)에서 $E^{tt}(e)$ 와 $Var^{tt}(e)$ 는 다음과 같이 간단히 나타낼 수 있다.

$$E^{tt}(e) = m$$

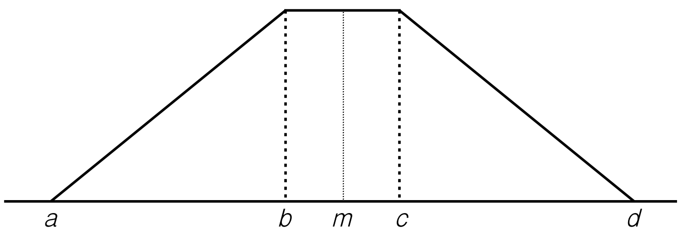
$$Var^{tt}(e) = \frac{(d+c)^2+2[(2m-c)^2-m(m+2d)]}{6}$$

절단된 삼각분포는 절단되지 않은 삼각분포와 기대값은 같지만 분산은 서로 다르다. 즉, 절단된 경우의 분산이 더 커지게 되는데, 이는 평균 주위의 자료들이 절단됨으로써 그만큼 더 나머지 자료들이 평균에서 멀리 떨어지기 때문일 것이다. 예를 들어, $a=0.5$, $b=0.75$, $m=1$, $c=1.25$, $d=1.5$ 라고 하면 절단된 삼각분포의 기대값은 1로서 절단되지 않은 삼각분포와 같다. 즉, $E^t(e)=1=E^{tt}(e)$ 이 된다. 그러나 절단된 삼각분포의 분산은 $Var^t(e)=0.5/18=0.028$ 이고 절단되지 않은 삼각분포의 분산은 $Var^{tt}(e)=8.25/72=0.115$ 이 된다. 따라서 이 예제에서는 절단된 경우가 절단되지 않은 경우에 비해 약 4배 정도 더 큰 것으로 나타났다.

다. 사다리꼴분포

사다리꼴분포(trapezoidal distribution)는 자료의 최빈값 m을 중심으로

[그림 3]과 같은 모양으로 이루어진 분포를 말한다. 즉, 사다리꼴분포는 [그림 2]의 절단된 삼각분포에서 두 절사점 b와 $c\,(c > b)$ 사이에 일정한 값을 가지는 형태로서, 절단된 삼각분포와는 달리 최빈값을 포함하고 있다.



[그림 3] 사다리꼴분포

만약 잡음 e가 최빈값 m을 중심으로 사다리꼴분포를 따른다고 한다면, 확률밀도함수(pdf), $f^{tr}(e)$ 는 다음과 같다.

$$f^{tr}(e) = \begin{cases} \dfrac{2k_1}{(d-a)(b-a)}(e-a)\,, & a \le e < b \\[2ex] \dfrac{2k_2}{(d-a)}\,, & b \le e < c \\[2ex] \dfrac{2k_3}{(d-a)(d-c)}(d-e)\,, & c \le e < d \end{cases}$$

여기서 k_1, k_2, k_3는 분포의 높이에 의존하는 값으로, 만약 높이가 같다면 이들 값은 모두 같게 된다. 즉, 두 절사점 b와 c에서 $f(b) = f(c)$가 된다면 다음과 같이 된다.

$$k_1 = k_2 = k_3 = \frac{(d-a)}{(d+c-b-a)}$$

따라서 사다리꼴분포에서 높이가 동일하다면 $f^{tr}(e)$ 는 다음과 같이

나타낼 수 있다.

$$f_{eq}^{tr}(e) = \begin{cases} \dfrac{2}{(d+c-b-a)}\dfrac{(e-a)}{(b-a)}, & a \le e < b \\[2ex] \dfrac{2}{(d+c-b-a)}, & b \le e < c \\[2ex] \dfrac{2}{(d+c-b-a)}\dfrac{(d-e)}{(d-c)}, & c \le e < d \end{cases}$$

위의 식 $f_{eq}^{tr}(e)$ 을 이용하여 잡음 e 의 기대값 $E^{tr}(e)$ 와 분산 $Var^{tr}(e)$ 을 구하면 다음과 같이 구할 수 있다.

$$E^{tr}(e) = \frac{(d^2+c^2-b^2-a^2+cd-ab)}{3(d+c-b-a)} \tag{3}$$

$$\begin{aligned} Var^{tr}(e) &= \frac{1}{18(d+c-b-a)} \\ &\times [3(d^3+c^3-b^3-a^3-a^2b-ab^2+c^2d+cd^2) \\ &\quad - \frac{2[(d+c-b-a)(d+c+b+a)+ab-cd]^2}{(d+c-b-a)}] \end{aligned}$$

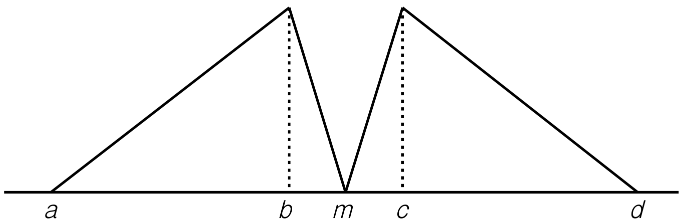
만약 잡음 e 가 최빈값 m 을 중심으로 좌우대칭이라고 한다면 $b+c = d+a = 2m$ 이 되므로, $a = 2m-d$ 와 $b = 2m-c$ 를 위 식 (3) 에 각각 대입하면 $E^{tr}(e)$ 와 $Var^{tr}(e)$ 은 다음과 같이 간단하게 나타낼 수 있다.

$$E^{tr}(e) = m$$

$$Var^{tr}(e) = \frac{[d^3+c^3-4m^3+(6m^2+cd)(d+c)-4m(d^2+c^2+cd)]}{6(d+c-2m)}$$

라. 이중삼각분포

이중삼각분포(double triangular distribution)는 [그림 4]에 나타난바와 같이 최빈값 m을 중심으로 좌우가 각각 삼각형 모양으로 이루어진 분포를 말한다. 즉, 최소값 a 에서 절사점 b를 중심으로 최빈값 m까지 하나의 삼각형 모양이 되고, 최빈값 m에서 절사점 c 를 중심으로 최대값 d 까지 또 하나의 삼각형 모양이 되는 분포이다. 이 분포는 두 절사점 b 와 c 사이에서 적절한 값을 갖지만 사다리꼴분포와는 달리 최빈값 m을 포함하지 않는다.



[그림 4] 이중삼각분포

잡음 e가 최빈값 m을 중심으로 이중삼각분포를 따른다면, 확률밀도함수(pdf), $f^{dt}(e)$ 는 다음과 같다.

$$f^{dt}(e) = \begin{cases} \dfrac{1}{(m-a)(b-a)}(e-a) \quad , & a \le e < b \\[2ex] \dfrac{1}{(m-a)(m-b)}(m-e), & b \le e < m \\[2ex] \dfrac{1}{(d-m)(c-m)}(e-m), & m \le e < c \\[2ex] \dfrac{1}{(d-m)(d-c)}(d-e) \quad , & c \le e < d \end{cases}$$

또한 $f^{dt}(e)$ 을 이용하여 잡음 e 의 기대값 $E^{dt}(e)$ 와 분산 $Var^{dt}(e)$ 을 계산하면 다음과 같이 된다.

$$E^{dt}(e) = \frac{(a+b+c+d+2m)}{6} \tag{4}$$

$$\begin{aligned} Var^{dt}(e) = \frac{1}{36}[&2a^2+2b^2+2c^2+2d^2+2m^2-m(a+b+c+d) \\ &-2(a+b)(c+d)+ab+cd] \end{aligned}$$

만약 잡음 e가 최빈값 m을 중심으로 좌우대칭이 된다면 $b+c=d+a=2m$이 되므로, $a=2m-d$ 와 $b=2m-c$ 를 위 식 (4)에 각각 대입하면 $E^{dt}(e)$와 $Var^{dt}(e)$은 다음과 같이 간단히 나타낼 수 있다.

$$E^{dt}(e) = m$$

$$Var^{dt}(e) = \frac{[(d-m)^2+(c-m)^2-m(d+c-m)+cd]}{6}$$

제3절 적용사례

1. 분석대상

가. 가계조사

앞 절에서 소개한 4개의 잡음분포를 이용한 승법잡음모형의 특성을 파악하기 위해 통계청에서 실시하는 가계조사에 적용해 보았다. 가계조사는 우리나라 가구에 대한 가계수지의 실태를 파악하여 국민의 소득과 소비수준 변화의 측정 및 분석 등에 필요한 자료를 제공하기 위한 조사이다. 2006년 가계조사의 대상가구는 약 8,000가구이며, 2005 인구주택총조사의 10% 표본조사구를 모집단으로 하여 전국을 25개로 층화한 후 확률비례추출법을 이용하여 선정하였다. 이렇게 선정된 전국의 표본가구를 대상으로 가구의 수입과 소비지출 관련 항목들을 응답가구에서 직

접 가계부에 매일 작성하도록 하고 있으며, 매월단위로 가계부의 내용을 집계, 정리한 후 그 결과를 분기별 및 연도별로 공표하고 있다.

2006년 가계조사의 대상가구는 <표 1>에 나타난바와 같이 동일한 가구의 중복을 제외하면 총 12,458가구이며, 이 중 서울이 1,638가구이고 서울이외지역이 10,820가구이다. 이 결과는 분기별 및 연별로 조사결과가 집계되어 공표되는 가구 수와 다른데, 이는 1년간 매월 조사에 응답한 동일가구는 12가구가 아닌 1가구로 처리하여 중복성을 제외하였기 때문이다.

〈표 1〉 2006년 가계조사의 표본규모

(단위: 가구)

지역	연간	1/4분기	2/4분기	3/4분기	4/4분기
전국	12,458	9,392	7,718	8,054	8,360
서울	1,638	1,259	947	965	987
기타(서울이외)	10,820	8,133	6,771	7,089	7,373

나. 분석항목

가계조사에서 가계부를 통해 조사되는 항목은 <표 2>에 주어져 있는 바와 같이 가구사항을 제외하면 총수입 항목과 총지출 항목으로 구분되는데, 총수입은 다시 소득과 기타수입, 그리고 전월이월금으로 나누어지고 총지출은 가계지출과 기타지출, 그리고 월말 현금잔고로 구분된다. 조사항목들을 세분화하여 분류한 후 이를 다시 품목별로 정리하면, 수입관련 품목이 47개이고 지출관련 품목이 510개로 모두 557개가 된다. 따라서 가계조사에서 조사되는 항목은 가구특성 62개와 조사품목 557개를 합해 총 619개 항목이 된다. 이에 대한 자세한 내용은 가계조사 지침서(2006)를 참고하기 바란다.

한편, 정동명(2008) 등은 가계조사 항목을 식별정보(identifying information)와 민감정보(sensitive information)로 각각 구분한 후, 이에 따라 비밀보호 방법을 달리 적용하였다. 그들은 62개의 가구특성 항목 중 응답자의 식별가능성이 높다고 판단되는 7개 항목을 식별변수(또는 key변수)로 선정하였다. 그리고 수입관련 품목 47개 중 31개, 지출관련 품목 510개 중 51개

등 총 82개 품목을 응답자들이 노출을 꺼려하는 민감변수로 각각 선정하였는데, 선정된 민감변수는 모두 최하단계의 품목들로 구성되어 있다.

〈표 2〉 가계수지 항목분류

- 총수입
  - 소득
    - 경상소득
      - 근로소득
      - 사업소득
      - 재산소득
      - 이전소득
    - 비경상소득
  - 기타 수입
    - 자산감소
    - 부채증가
  - 전월이월금
- 총지출
  - 가계지출
    - 소비지출
      - 식료품비
      - 주거비
      - 광열・수도
      - 가구・가사용품
      - 의류・신발
      - 보건・의료
      - 교육
      - 교양・오락
      - 교통・통신
      - 기타소비지출
    - 비소비지출
  - 기타지출
    - 자산증가
    - 부채감소
  - 월말현금잔고

그러나 본 연구에서는 최하단계 품목이 아닌 보다 상위단계에 있는 품목들을 민감변수로 새로 선정하였다. 이것은 최하단위의 품목이 너무 세분화 되어 있어 이용자들에게 제공할 경우 통계정보로서의 가치가 낮거나 전혀 의미없는 정보가 될 수도 있으며, 오히려 이들의 합의 형태로 구성된 상위단계의 품목이 통계정보로서 더 효용성이 높을 것이라고 판단했기 때문이다. 가령, 수입관련 품목 중 가구주급여소득과 가구주상여금을 합하면 가구주소득이 얻어지는데, 이용자들에게는 가구주급여소득이나 가구주상여금에 대한 것보다는 가구주소득에 대한 정보가 더 유용할 수 있을 것이다.

이러한 기준에 따라 최하단계의 품목은 이용자들에게 제공하지 않는 것으로 가정하고, 외부인(intruder)이 다른 경로를 통해 정보를 얻을 수 있을 것으로 판단되는 품목들을 분석한 후, 최종적으로 수입관련 9개 품목과 지출관련 21개 품목 등 총 30개 품목을 민감변수로 선정하였다. 이에 대한 결과는 <표 3>에 주어져 있다.

〈표 3〉 민감변수

구분	품 목		
총수입 (9개)	· 가구주소득	· 배우자소득	· 기타 가구원소득
	· 사업소득	· 재산소득	· 이전소득
	· 비경상소득	· 자산감소로인한수입	· 부채증가로인한수입
총지출 (21개)	· 월세	· 주택설비및수선비	· 기타주거
	· 수도료	· 전기료	· 연료
	· 공동주택난방비	· 의약품	· 보건의료용품기구
	· 보건의료서비스	· 납입금	· 교재비
	· 보충교육비	· 문구류	· 교통비
	· 통신비	· 조세	· 공적연금
	· 사회보험	· 기타비소비지출	· 월말현금잔고

2. 승법잡음모형에 의한 비밀보호

가. 잡음의 생성

자료의 비밀보호를 위해 승법잡음모형을 적용하기 위해서는 우선 잡음을 생성해야만 하는데, 여기서는 앞 절에서 소개한 4가지 분포에 따라 잡음을 생성하는 과정을 소개하고자 한다. 먼저 잡음 e 가 각각의 분포를 따른다는 가정하에 최빈값(m)을 1로 하고 자료의 범위가 $0.01 \le |e-1| \le 0.4$ 을 만족하도록 최소값(a)과 최대값(d), 두 절사점(b, c) 등을 정하였다. 이것은 두 절사점이 최빈값 1을 중심으로 너무 멀리 떨어지지 않도록 하기 위해서이다. 만약 절사점이 최빈값 1에서 너무 멀리 떨어지면 편향이 발생하고 분산이 커지게 된다. 이에 대한 자세한 내용은 정동명(2008) 등을 참고하기 바란다.

위의 조건이 만족되면 $a=0.6$, $d=1.4$, $b=0.99$, $c=1.01$ 이 되는데, 이 값들을 이용하여 각 분포별로 난수발생 프로그램을 SAS로 작성하였다. 그리고 2006년도 가계조사의 월별 표본규모를 고려하여 약 90,000개의 난수를 생성하였다.

각 분포별로 난수를 발생시킨 결과가 실제 이론에 의한 값과 어느 정도 차이가 나는지를 검토하기 위해 기초통계량을 계산하였으며, 그 결과를 <표 4>에 나타내었다. 이 표에 따르면 대부분의 경우에 큰 차이

가 나지 않음을 알 수 있는데, 이는 각 분포별로 난수가 잘 생성되었음을 보여주고 있다. 가령, 각 분포별로 난수의 평균값은 모두 1.0002로 나타나 이론적 평균값인 1과 거의 차이가 나지 않으며, 표준편차의 경우에도 거의 차이가 없음을 알 수 있다.

다만, 각 분포별로 이론적인 표준편차가 약간씩 다르게 나타나는데, 삼각분포가 가장 작고 이어서 사다리꼴분포와 이중삼각분포, 절단된 삼각분포의 순으로 나타났다. 이것은 두 절사점 사이의 정보량이 얼마나 더 많은지에 따른 차이로서, 두 절사점 사이에서 난수가 많이 발생되는 분포가 그렇지 못한 분포에 비해 정보량이 더 많아 표준편차가 적어지기 때문인 것 같다. 그러므로 두 절사점 사이에서 난수가 발생되지 않는 절단된 삼각분포의 표준편차가 가장 크게 나타난다고 할 수 있을 것이다.

〈표 4〉 생성된 난수의 결과비교

분포	구분	통계량			
		평균	표준편차	최소값	최대값
삼각	이론적(A)	1.0000	0.1633	0.6000	1.4000
	난수결과(B)	1.0002	0.1624	0.6041	1.3916
	차이(B-A)	0.0002	-0.0009	0.0041	-0.0084
절단된 삼각	이론적(A)	1.0000	0.1675	0.6000	1.4000
	난수결과(B)	1.0002	0.1666	0.6040	1.3918
	차이(B-A)	0.0002	-0.0009	0.0040	-0.0082
사다리꼴	이론적(A)	1.0000	0.1634	0.6000	1.4000
	난수결과(B)	1.0002	0.1624	0.6041	1.3916
	차이(B-A)	0.0002	0.0010	0.0041	-0.0084
이중삼각	이론적(A)	1.0000	0.1654	0.6000	1.4000
	난수결과(B)	1.0002	0.1645	0.6041	1.3917
	차이(B-A)	0.0002	0.0009	0.0041	-0.0083

나. 자료의 변환

원자료를 X_i, 잡음을 e_i 라 하면 승법잡음모형에 따라 변환된 자료 Y_i 는 다음과 같이 나타낼 수 있다.

$$Y_i = X_i \cdot e_i, \quad i = 1, \cdots, n$$

여기서 X_i 와 e_i 는 서로 독립이기 때문에 변환된 Y_i 의 기대값 $E(Y_i)$ 와 분산 $Var(Y_i)$ 은 다음과 같이 구할 수 있다.

$$E(Y_i) = E(X_i) \cdot E(e_i) \tag{5}$$

$$Var(Y_i) = Var(X_i)\ Var(e_i)$$
$$+ [E(e_i)]^2\ Var(X_i) + [E(X_i)]^2\ Var(e_i)$$

한편, 통계이용자들에게 변환된 자료를 제공할 경우 이용자들은 원자료 X_i 의 평균과 분산은 모르고 단지 변환된 자료 Y_i 와 잡음 e_i 의 평균과 분산만 알 수 있게 된다. 따라서 이용자들 중에는 추가적인 자료분석을 위해 원자료의 평균과 분산을 요구할 수도 있다. 이를 위해 원자료 X_i 의 평균과 분산을 구하는 추정식이 필요한데, 위의 식 (5)을 이용하여 X_i 의 기대값 $E(X_i)$ 와 분산 $Var(X_i)$ 은 다음과 같이 얻을 수 있다.

$$\hat{E}(X_i) = \hat{E}(Y_i) / E(e_i) = \hat{E}(Y_i)$$

$$\widehat{Var}(X_i) = \frac{\widehat{Var}(Y_i) - [\hat{E}(X_i)]^2\ Var(e_i)}{Var(e_i) + [E(e_i)]^2}$$

다. 소지역자료의 변환

소지역(domain) 자료의 변환을 위해 소지역을 나타내는 첨자를 s 라 하면, 승법잡음모형에 의해 변환된 소지역자료 Y_i^s 는 다음과 같이 나타낼 수 있다.

$$Y_i^s = X_i^s \cdot e_i, \quad i = 1, \cdots, n_s$$

여기서 n_s 는 소지역의 크기를 나타내며, 소지역의 원자료 X_i^s 와 e_i 는 서로 독립이기 때문에 변환된 Y_i^s 의 기대값 $E(Y_i^s)$ 와 분산 $Var(Y_i^s)$ 은 다음과 같이 구할 수 있다.

$$E(Y_i^s) = E(X_i^s) \cdot E(e_i) \tag{6}$$

$$Var(Y_i^s) = Var(X_i^s)\ Var(e_i) + [E(e_i)]^2\ Var(X_i^s) + [E(X_i^s)]^2\ Var(e_i)$$

따라서 위의 식 (6)을 이용하여 소지역 s 에서의 원자료 X_i^s 의 분산 $Var(X_i^s)$ 은 다음과 같이 얻을 수 있다.

$$\widehat{Var}(X_i^s) = \frac{\widehat{Var}(Y_i^s) - [E(X_i^s)]^2\ Var(e_i)}{Var(e_i) + [E(e_i)]^2}$$

라. 변환결과의 비교 · 분석

승법잡음모형의 자료변환에 대한 분석을 위해 2006년 가계조사 결과에 직접 적용해 보았다. 먼저 30개 민감변수 중 수입관련 품목의 가구주소득과 지출관련 품목의 월세를 대상으로 앞 절에서 언급한 4가지 잡음분포에 따라 난수를 생성한 후, 이 잡음을 가계조사자료에 곱하여 자료를 변환해 보았다. 즉, 가계조사의 연간자료를 원자료로 한 후 전국 및 지역별로 원자료와 변환된 자료의 평균과 표준편차를 각각 계산하고, 또한 원자료와 변환된 자료의 차이를 비교하기 위해 상대차이(relative difference)도 계산하였다. 이에 대한 결과가 <표 5>와 <표 6>에 각각 주어져 있다.

가구주소득에 대한 결과를 살펴보면, 먼저 전국단위의 경우 원자료와 변환된 자료의 평균값 차이는 약 2,200~2,300원 정도이고, 표준편차의 차이는 1,000원 내외인 것으로 나타나 이들의 차이가 크지 않음을 알 수 있다. 또한 원자료와 변환된 자료의 상대차이도 약 0.1% 정도로 나타나 각 분포별로 큰 차이가 없는 것으로 나타났다. 분포별로는 삼각분포를 이용하여 변환한 결과가 원자료와 가장 차이가 적고 이어서 사다리꼴분포와 이중삼각분포, 그리고 절단된 삼각분포의 순으로 나타났다.

이는 삼각분포나 사다리꼴분포에서 생성된 난수가 1을 포함하므로 자료 변환이 이루어지지 않은 경우가 있기 때문인 것 같다.

〈표 5〉 품목별 변환결과

(단위 : 원)

품목	통계량	지역	원자료	변환된 자료			
				삼각분포	절단된 삼각분포	사다리꼴분포	이중삼각분포
가구주 소득	평균	전국	2,258,159	2,255,944	2,255,851	2,255,942	2,255,894
		서울	2,295,665	2,299,081	2,298,817	2,299,078	2,298,942
		기타지역	2,252,439	2,249,366	2,249,299	2,249,364	2,249,330
	표준 편차	전국	1,636,192	1,635,247	1,635,132	1,635,209	1,635,181
		서울	1,559,763	1,574,642	1,574,811	1,574,607	1,574,723
		기타지역	1,647,478	1,644,212	1,644,055	1,644,173	1,644,124

품목	통계량	지역	원자료	변환된 자료			
				삼각분포	절단된 삼각분포	사다리꼴분포	이중삼각분포
월세	평균	전국	181,064	180,943	180,975	180,943	180,958
		서울	246,382	247,216	247,280	247,216	247,247
		기타지역	171,099	170,833	170,860	170,833	170,846
	표준 편차	전국	133,014	132,038	132,168	132,037	132,102
		서울	164,236	165,359	165,639	165,358	165,499
		기타지역	124,615	123,095	123,195	123,094	123,144

〈표 6〉 품목별 변환자료의 상대차이

품목	통계량	지역	상 대 차 이			
			삼각분포	절단된 삼각분포	사다리꼴분포	이중삼각분포
가구주 소득	평균	전국	0.00098	0.00102	0.00098	0.00100
		서울	0.00149	0.00137	0.00149	0.00143
		기타지역	0.00136	0.00139	0.00137	0.00138
	표준 편차	전국	0.00058	0.00065	0.00060	0.00062
		서울	0.00954	0.00965	0.00952	0.00959
		기타지역	0.00198	0.00208	0.00201	0.00204

품목	통계량	지역	상 대 차 이			
			삼각분포	절단된 삼각분포	사다리꼴분포	이중삼각분포
월세	평균	전국	0.00067	0.00049	0.00067	0.00058
		서울	0.00338	0.00365	0.00339	0.00351
		기타지역	0.00156	0.00140	0.00156	0.00148
	표준편차	전국	0.00733	0.00636	0.00734	0.00686
		서울	0.00684	0.00855	0.00684	0.00770
		기타지역	0.01220	0.01140	0.01221	0.01181

* 상대차이 = |원자료-변환자료|/원자료

비밀보호를 위해 실제 자료에 적용할 경우에는 잡음이 1이 포함되지 않는 것이 더 바람직하므로 절단된 삼각분포와 이중삼각분포만을 비교해 보면, 이중삼각분포를 이용하여 변환한 경우가 원자료와의 차이가 약간 더 적다. 이러한 결과는 절단된 삼각분포의 경우 두 절사점 b와 c 사이에서 난수가 전혀 생성되지 않지만, 이중삼각분포의 경우 최빈값을 중심으로 어느 정도 난수가 생성되어 그만큼 더 많은 정보가 이용되기 때문인 것 같다.

지역단위의 경우에도 원자료와 변환된 자료의 차이가 크지 않은 것으로 나타났으나, 서울지역은 절단된 삼각분포인 경우가 가장 차이가 적어 전국단위나 기타지역의 결과와 다르게 나타났다. 이것은 두 절사점 사이에서 난수가 발생하지 않아 정보량이 적으면 상대적으로 변환된 자료는 원자료와 차이가 크지만, 원자료에 어떤 값이 곱해지는가에 따라 차이가 크지 않을 수도 있다는 것이다. 다시 말하면, 난수발생을 위한 잡음의 분포가 자료의 변환에 중요한 요인이 될 수 있지만, 무엇보다도 원자료에 곱해지는 값이 가장 절대적인 요인이 된다는 것을 보여주고 있다.

월세에 대한 결과를 살펴보면, 전국단위의 경우 원자료와 변환된 자료의 평균값 차이는 100원 내외이고 표준편차의 차이는 840~980원 사이인 것으로 나타나 차이가 크지 않음을 알 수 있다. 그러나 분포별 변환결과는 가구주소득의 경우와 반대로 나타나는데, 전국단위와 기타지역의 경우에는 절단된 삼각분포를 이용하여 변환한 결과가 원자료와 가장

차이가 적고, 서울지역의 경우에는 삼각분포를 이용하여 변환한 결과가 가장 원자료와 차이가 적은 것으로 나타났다. 이러한 결과는 가구주소득에서 설명한 것과 유사하게 해석하면 될 것이다.

위에서 언급한 방법과 동일하게 모든 민감변수에 대해 자료를 변환하고 이들의 평균과 표준편차를 구한 후, 원자료와 어느 정도 차이가 나는지를 알아보기 위해 상대차이를 계산하여 그 결과를 <부록>에 수록하였다. 이 결과에 따르면 일부 품목을 제외한 대부분의 경우 삼각분포를 이용하여 변환한 결과가 원자료와 가장 차이가 적은 것으로 나타났다. 또한 잡음 e 가 1을 포함하지 않는 절단된 삼각분포와 이중삼각분포만을 비교해 보면, 두 절사점 사이의 정보를 어느 정도 이용하는 이중삼각분포가 더 효과적이라고 할 수 있다.

이상의 분석결과를 정리하면, 승법잡음모형을 이용하여 자료를 변환할 경우 원자료에 곱해지는 잡음의 값이 가장 중요한 요인이 되지만, 어떤 잡음분포를 이용해서 잡음을 생성해야 하는지도 중요한 요인이 될 수 있음을 보여주고 있다. 따라서 만약 승법잡음모형으로 자료를 변환하고자 할 경우 본 연구에서 제시한 이중삼각분포를 잡음의 분포로 이용하면 상당히 효과적이라고 할 수 있을 것이다.

제4절 결 론

본 연구에서는 최근 들어 연속형 자료의 비밀보호방법에서 널리 활용되고 있는 승법잡음모형을 소개하고, 2006년 가계조사 자료에 이 모형을 적용하여 비밀보호된 마이크로자료를 작성하는 과정을 설명하고 분석하였다. 이를 위해 먼저 승법잡음모형에서 요구되는 잡음의 분포에 대해 4종류의 분포를 제시하고, 이들 분포의 특성을 살펴보았다. 먼저 삼각분포를 기본으로 하여 자료의 최빈값을 중심으로 좌우가 절단된 삼각분포와 두 절사점간에 일정한 값을 갖는 사다리꼴분포, 그리고 두 삼각형의 형태로 이루어진 이중삼각분포에 대해 확률밀도함수를 각각 구하고 이를 이용하여 평균과 분산의 계산공식을 유도하였다.

가계조사 자료의 비밀보호를 위해 가구특성항목을 제외한 557개 조

사품목 중 수입관련 품목 9개와 지출관련 품목 21개 등 총 30개 품목을 민감변수로 선정한 후, 이들 변수별로 제시된 잡음분포에 따라 승법잡음모형을 적용하였다. 그 결과, 대부분의 항목에서 삼각분포를 가정하고 변환한 자료가 원자료와 가장 유사한 것으로 나타났는데, 이는 삼각분포를 가정할 경우 잡음 e 가 1을 포함하므로 원자료가 변환되지 않을 수도 있기 때문이다.

따라서 잡음 e 가 1이 포함하지 않는 절단된 삼각분포와 이중삼각분포만을 비교해보면, 대부분의 변수에서 이중삼각분포를 가정한 경우가 절단된 삼각분포를 가정한 경우보다 약간 더 효과적인 것으로 나타났다. 이러한 결과는 이중삼각분포가 두 절사점 사이의 정보를 이용하기 때문이라 할 수 있다. 이상에서 살펴본바와 같이 가계조사의 경우 응답가구에서 노출을 꺼려하는 민감한 정보들이 대부분 연속형 자료이므로 이중삼각분포를 가정한 승법잡음모형을 적용하는 것이 비밀보호에 상당히 효과적이라고 판단된다.

한편, 본 연구에서는 최하단계의 품목들을 제공하지 않는다는 가정하에, 상위단계의 품목들로 민감변수를 구성하여 분석하였으나 이에 대한 다른 견해도 있을 것이다. 하지만 가계조사의 경우 너무 세분화된 자료가 이용자들에게 제공됨으로써 오히려 정보의 질을 저하시킬 가능성도 있으므로, 수집된 자료를 일방적으로 모두 제공하기보다는 자료제공심의 등의 절차를 거쳐 자료제공의 범위와 수준을 정하고, 이에 따라 이용자들에게 제공하는 것이 더 바람직할 수도 있을 것이다.

향후에 본 연구의 결과를 실무에 적용하여 활용성을 높이기 위해서는 품목간의 연관성에 대한 분석 등 추가적인 연구가 필요하다. 외국의 경우에도 새로운 비밀보호방법의 실무적용을 위해서 연구자가 지속적인 연구가 수행되도록 다양한 지원을 하고 있다. 따라서 본 연구에서 제안한 승법잡음모형을 이용하여 비밀보호된 가계조사의 마이크로자료를 작성하고, 이용자들에게 제공하기 위해서는 체계적이고 지속적인 연구와 더불어 이와 관련된 부서에서도 꾸준한 관심과 적극적인 지원이 요구된다. 끝으로 본 연구를 계기로 응답자의 정보를 보호할 수 있는 다양한 연구가 활발히 진행되고, 그 결과를 바탕으로 자료제공의 신뢰성 확보와 함께 궁극적으로 조사의 응답률도 제고될 수 있기를 기대해 본다.

참고문헌

정동명, 김종익, 강동환(2007), "인구센서스자료의 비밀보호방법", 「통계연구」, 제12권 제1호, 95-120.

정동명, 정미옥(2008), "인구주택총조사 마이크로자료의 개인정보 노출제한방법", 「응용통계연구」, 제21권 제2호, 313-325.

정동명, 정남수, 한승훈(2008), "가계조사 마이크로데이터의 비밀보호", 연구보고서, 통계청.

통계청(2006), 「가계조사 조사지침서」.

Fuller, W. A. (1993), "Masking Procedures for Microdata Disclosure Limitation", *Journal of Official Statistics*, 9, 383-406.

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*, 370-374.

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.

Kim, J., Katzoff, M., Gonzalez Jr, J., and Williams, P. (2003), "Techniques for Masking Microdata", *National Center for Health Statistics internal memorandum*.

Kim, J. and Jeong, D. M. (2007), "Application of the Concept of Uniqueness for Creating Public Use Microdata Files", *in Proceedings of the Joint UNECE/Eurostat Work Session on Statistical Data Confidentiality*, UNECE, CD-ROM.

<부록> 품목별 변환결과의 상대차이

품목	통계량	지역	상 대 차 이			
			삼각분포	절단된 삼각분포	사다리꼴분포	이중삼각분포
가구주소득	평균	전국	0.00098	0.00102	0.00098	0.00100
		서울	0.00149	0.00137	0.00149	0.00143
		기타	0.00136	0.00139	0.00137	0.00138
	표준편차	전국	0.00058	0.00065	0.00060	0.00062
		서울	0.00954	0.00965	0.00952	0.00959
		기타	0.00198	0.00208	0.00201	0.00204
배우자소득	평균	전국	0.00116	0.00127	0.00116	0.00122
		서울	0.00053	0.00058	0.00053	0.00055
		기타	0.00145	0.00159	0.00145	0.00152
	표준편차	전국	0.00143	0.00125	0.00141	0.00134
		서울	0.00693	0.00692	0.00691	0.00694
		기타	0.00048	0.00026	0.00046	0.00037
기타 가구원소득	평균	전국	0.00242	0.00261	0.00242	0.00252
		서울	0.00251	0.00300	0.00251	0.00276
		기타	0.00240	0.00252	0.00240	0.00246
	표준편차	전국	0.00417	0.00451	0.00419	0.00435
		서울	0.00382	0.00396	0.00385	0.00390
		기타	0.00425	0.00463	0.00428	0.00444
사업소득	평균	전국	0.00286	0.00297	0.00286	0.00292
		서울	0.00095	0.00101	0.00095	0.00099
		기타	0.00313	0.00326	0.00314	0.00320
	표준편차	전국	0.01582	0.01628	0.01584	0.01605
		서울	0.00773	0.00745	0.00775	0.00761
		기타	0.01673	0.01726	0.01675	0.01700
재산소득	평균	전국	0.00892	0.00921	0.00893	0.00907
		서울	0.01429	0.01504	0.01430	0.01468
		기타	0.00764	0.00782	0.00765	0.00774
	표준편차	전국	0.05780	0.05887	0.05782	0.05835
		서울	0.06435	0.06696	0.06440	0.06571
		기타	0.05674	0.05739	0.05675	0.05707

* 상대차이= |(원자료-변환자료)|/원자료

품목	통계량	지역	상 대 차 이			
			삼각분포	절단된 삼각분포	사다리꼴분포	이중삼각분포
이전소득	평균	전국	0.00066	0.00058	0.00066	0.00062
		서울	0.01365	0.01419	0.01366	0.01393
		기타	0.00054	0.00069	0.00055	0.00062
	표준편차	전국	0.00271	0.00180	0.00271	0.00225
		서울	0.06537	0.06977	0.06544	0.06759
		기타	0.00672	0.00839	0.00673	0.00756
비경상소득	평균	전국	0.00077	0.00107	0.00077	0.00092
		서울	0.00863	0.00950	0.00864	0.00907
		기타	0.00185	0.00229	0.00186	0.00207
	표준편차	전국	0.00170	0.00455	0.00173	0.00314
		서울	0.06663	0.07272	0.06671	0.06970
		기타	0.00766	0.01125	0.00769	0.00947
자산감소로 인한 수입	평균	전국	0.00288	0.00301	0.00288	0.00295
		서울	0.00244	0.00279	0.00245	0.00262
		기타	0.00295	0.00304	0.00295	0.00300
	표준편차	전국	0.02813	0.02881	0.02815	0.02848
		서울	0.00280	0.00035	0.00275	0.00121
		기타	0.03255	0.03287	0.03256	0.03272
부채증가로 인한 수입	평균	전국	0.00032	0.00043	0.00032	0.00037
		서울	0.01471	0.01490	0.01472	0.01482
		기타	0.00231	0.00221	0.00231	0.00227
	표준편차	전국	0.01729	0.01809	0.01729	0.01768
		서울	0.07107	0.07284	0.07109	0.07201
		기타	0.00129	0.00083	0.00130	0.00109
월세	평균	전국	0.00067	0.00049	0.00067	0.00058
		서울	0.00338	0.00365	0.00339	0.00351
		기타	0.00156	0.00140	0.00156	0.00148
	표준편차	전국	0.00733	0.00636	0.00734	0.00686
		서울	0.00684	0.00855	0.00684	0.00770
		기타	0.01220	0.01140	0.01221	0.01181

* 상대차이= |(원자료-변환자료)|/원자료

품목	통계량	지역	상 대 차 이			
			삼각분포	절단된 삼각분포	사다리꼴분포	이중삼각분포
주택설비 및 수선비	평균	전국	0.00606	0.00607	0.00606	0.00607
		서울	0.02614	0.02782	0.02616	0.02699
		기타	0.00385	0.00368	0.00385	0.00377
	표준편차	전국	0.00305	0.00225	0.00304	0.00265
		서울	0.05203	0.05632	0.05209	0.05419
		기타	0.00046	0.00161	0.00047	0.00104
기타주거	평균	전국	0.00221	0.00229	0.00221	0.00225
		서울	0.00550	0.00591	0.00550	0.00570
		기타	0.00165	0.00167	0.00165	0.00166
	표준편차	전국	0.01778	0.01924	0.01781	0.01852
		서울	0.04892	0.05154	0.04896	0.05024
		기타	0.00078	0.00000	0.00076	0.00037
수도료	평균	전국	0.00054	0.00056	0.00054	0.00055
		서울	0.00117	0.00113	0.00117	0.00115
		기타	0.00082	0.00083	0.00082	0.00083
	표준편차	전국	0.00185	0.00192	0.00187	0.00189
		서울	0.01250	0.01273	0.01253	0.01262
		기타	0.00021	0.00025	0.00023	0.00024
전기료	평균	전국	0.00135	0.00139	0.00135	0.00138
		서울	0.00087	0.00104	0.00087	0.00096
		기타	0.00142	0.00144	0.00142	0.00143
	표준편차	전국	0.00486	0.00521	0.00488	0.00504
		서울	0.00852	0.00883	0.00854	0.00870
		기타	0.00449	0.00485	0.00452	0.00467
연료	평균	전국	0.00191	0.00199	0.00191	0.00195
		서울	0.00206	0.00229	0.00206	0.00218
		기타	0.00189	0.00194	0.00189	0.00191
	표준편차	전국	0.01155	0.01185	0.01157	0.01171
		서울	0.00833	0.00852	0.00835	0.00844
		기타	0.01184	0.01215	0.01185	0.01199

* 상대차이= |(원자료-변환자료)|/원자료

품목	통계량	지역	상 대 차 이			
			삼각분포	절단된 삼각분포	사다리꼴분포	이중삼각분포
공동주택 난방비	평균	전국	0.00131	0.00130	0.00131	0.00130
		서울	0.00034	0.00018	0.00034	0.00027
		기타	0.00172	0.00166	0.00172	0.00169
	표준편차	전국	0.00577	0.00602	0.00578	0.00590
		서울	0.00086	0.00161	0.00089	0.00123
		기타	0.00733	0.00743	0.00735	0.00739
의약품	평균	전국	0.00186	0.00195	0.00187	0.00191
		서울	0.00152	0.00172	0.00153	0.00163
		기타	0.00192	0.00198	0.00192	0.00195
	표준편차	전국	0.00825	0.00845	0.00826	0.00835
		서울	0.00159	0.00320	0.00160	0.00238
		기타	0.01079	0.01147	0.01081	0.01113
보건의료 용품기구	평균	전국	0.00159	0.00161	0.00159	0.00160
		서울	0.01345	0.01391	0.01346	0.01369
		기타	0.00018	0.00023	0.00018	0.00021
	표준편차	전국	0.00354	0.00354	0.00355	0.00353
		서울	0.10388	0.10784	0.10394	0.10588
		기타	0.01006	0.01056	0.01006	0.01032
보건의료 서비스	평균	전국	0.00097	0.00101	0.00097	0.00099
		서울	0.00502	0.00513	0.00502	0.00507
		기타	0.00183	0.00190	0.00183	0.00187
	표준편차	전국	0.01457	0.01579	0.01459	0.01519
		서울	0.01331	0.01387	0.01331	0.01359
		기타	0.01818	0.01964	0.01821	0.01892
납입금	평균	전국	0.00181	0.00160	0.00180	0.00170
		서울	0.01621	0.01673	0.01621	0.01648
		기타	0.00517	0.00501	0.00516	0.00509
	표준편차	전국	0.00072	0.00118	0.00071	0.00094
		서울	0.03409	0.03481	0.03409	0.03445
		기타	0.00576	0.00535	0.00576	0.00557

* 상대차이= |(원자료-변환자료)|/원자료

품목	통계량	지역	상 대 차 이			
			삼각분포	절단된 삼각분포	사다리꼴분포	이중삼각분포
교재비	평균	전국	0.00427	0.00441	0.00427	0.00434
		서울	0.01103	0.01158	0.01103	0.01130
		기타	0.00334	0.00342	0.00334	0.00338
	표준편차	전국	0.02348	0.02379	0.02349	0.02364
		서울	0.03212	0.03304	0.03214	0.03257
		기타	0.02283	0.02310	0.02285	0.02297
보충교육비	평균	전국	0.00041	0.00040	0.00041	0.00040
		서울	0.00052	0.00060	0.00052	0.00057
		기타	0.00058	0.00059	0.00058	0.00058
	표준편차	전국	0.00595	0.00702	0.00598	0.00650
		서울	0.02414	0.02760	0.02420	0.02589
		기타	0.00038	0.00013	0.00036	0.00025
문구류	평균	전국	0.00279	0.00289	0.00279	0.00284
		서울	0.00791	0.00830	0.00792	0.00810
		기타	0.00185	0.00190	0.00185	0.00188
	표준편차	전국	0.00395	0.00442	0.00394	0.00419
		서울	0.00826	0.00916	0.00827	0.00871
		기타	0.00227	0.00259	0.00227	0.00244
교통	평균	전국	0.00263	0.00264	0.00263	0.00264
		서울	0.00346	0.00333	0.00346	0.00339
		기타	0.00343	0.00342	0.00343	0.00343
	표준편차	전국	0.01030	0.01043	0.01031	0.01040
		서울	0.00722	0.00568	0.00719	0.00644
		기타	0.01304	0.01295	0.01305	0.01303
통신	평균	전국	0.00162	0.00166	0.00162	0.00164
		서울	0.00052	0.00060	0.00052	0.00056
		기타	0.00179	0.00182	0.00179	0.00181
	표준편차	전국	0.00390	0.00388	0.00392	0.00390
		서울	0.00143	0.00188	0.00141	0.00165
		기타	0.00476	0.00480	0.00478	0.00478

* 상대차이= |(원자료-변환자료)|/원자료

품목	통계량	지역	상 대 차 이			
			삼각분포	절단된 삼각분포	사다리꼴분포	이중삼각분포
조세	평균	전국	0.00530	0.00549	0.00530	0.00540
		서울	0.00203	0.00182	0.00203	0.00193
		기타	0.00645	0.00664	0.00645	0.00655
	표준편차	전국	0.01810	0.01864	0.01812	0.01837
		서울	0.01608	0.01602	0.01607	0.01605
		기타	0.02429	0.02492	0.02431	0.02461
공적연금	평균	전국	0.00146	0.00154	0.00146	0.00150
		서울	0.00164	0.00183	0.00164	0.00174
		기타	0.00143	0.00150	0.00143	0.00147
	표준편차	전국	0.00044	0.00005	0.00046	0.00025
		서울	0.00833	0.00881	0.00837	0.00857
		기타	0.00036	0.00084	0.00035	0.00060
사회보험	평균	전국	0.00167	0.00174	0.00167	0.00171
		서울	0.00062	0.00075	0.00062	0.00069
		기타	0.00185	0.00190	0.00185	0.00187
	표준편차	전국	0.00598	0.00612	0.00601	0.00606
		서울	0.00426	0.00412	0.00428	0.00419
		기타	0.00635	0.00654	0.00637	0.00645
기타 비소비지출	평균	전국	0.00400	0.00420	0.00400	0.00409
		서울	0.00354	0.00426	0.00355	0.00390
		기타	0.00406	0.00419	0.00406	0.00411
	표준편차	전국	0.02099	0.02320	0.02100	0.02188
		서울	0.03830	0.04431	0.03836	0.04133
		기타	0.01758	0.01902	0.01757	0.01804
월말 현금잔고	평균	전국	0.00140	0.00146	0.00140	0.00143
		서울	0.00040	0.00056	0.00041	0.00049
		기타	0.00154	0.00158	0.00154	0.00156
	표준편차	전국	0.01401	0.01338	0.01403	0.01377
		서울	0.00086	0.00093	0.00088	0.00090
		기타	0.01460	0.01393	0.01461	0.01434

* 상대차이= |(원자료-변환자료)|/원자료