

## 제3장

# 농업총조사 무응답 대체기법 연구 |

최필근

## 제1절 서론

### 1. 연구배경 및 목적

농업총조사는 5년 주기로 조사되는 전수조사로서 농가 및 농가인구의 규모와 분포, 농업의 경영구조특성 등을 파악하여 농업정책의 기본 방향을 수립하고, 각종 농업통계 개선을 위한 모집단 자료를 확보함과 동시에 지방화시대에 요구되는 소지역 자료의 생산 및 농업부문에 대한 국가경쟁력 향상에 기여하는 것을 목적으로 하고 있다. 이러한 목적을 충족하기 위해서는 표본의 설계에서부터 조사과정 및 분석이 완료되기까지의 일련의 과정이 정확하게 이루어져야 한다. 그러나 모든 과정에서 의도하지 않은 오차가 발생하기 마련이며, 그 중 하나가 무응답으로 인해서 발생하게 된다.

통계조사에서 모든 조사대상으로부터 필요한 정보를 완벽하게 얻는 것은 쉬운 일이 아니다. 하지만 이렇게 발생된 무응답은 조사결과에 상당한 영향을 줄 수도 있어 조사의 신뢰도를 높이기 위해서는 무응답률을 낮추는 것이 매우 중요하다. 많은 노력에도 불구하고 시간적·경제적 여건으로 인하여 무응답은 계속적으로 발생되고 있으며, 특히 최근에는 1인·노인 및 맞벌이 가구가 증가와 사생활을 중시하는 의식이 높아져 점점 더 무응답의 비율이 증가할 것으로 보인다. 농업총조사의 경우 현재는 조사특성상 조사담당자의 반복적인 방문 및 조사현장에서의

## 2 무응답 처리를 위한 방법론 연구 1

대체로 인해서 무응답 없는 완전한 자료가 작성되고 있으나, 변화하는 농촌의 환경과 임업총조사의 이관 등으로 조사담당자의 업무부담뿐만 아니라 응답자의 응답부담도 더욱 가중되어 향후 무응답에 대한 문제를 심각하게 고려해야 할 것이다. 따라서 무응답에 대한 근본적인 문제해결이 쉽지 않기 때문에 불가피하게 발생하는 무응답에 대하여 사후적으로 처리하는 방법을 체계적이고 지속적으로 연구하여 농업총조사의 무응답 발생 시 정확성이 높은 대체를 할 수 있도록 대비해야 할 필요성이 있을 것이다.

무응답 발생은 조사내용, 자료수집 방법, 응답자의 가구구조 및 태도 등에 많이 좌우된다. 무응답의 형태를 나누어 보면 단위 무응답(unit nonresponse)과 항목 무응답(item nonresponse)으로 정의할 수 있다. 단위 무응답은 조사자로부터 얻은 정보가 하나도 없는 것을 의미하고, 항목 무응답은 응답을 해야 할 항목에 대해서 응답을 하지 않거나 질문과는 무관한 응답을 함으로써 불필요한 자료가 되는 것을 의미한다. 이러한 무응답이 발생하면 효과적인 통계분석을 위해서는 적절한 방법을 통해 무응답을 처리해야 하는데 단위 무응답의 경우는 가중치 조정 방법을 사용하고, 항목 무응답의 경우는 적절한 값을 채워 넣기 위한 여러 가지 대체법을 이용하게 된다. 농업총조사의 무응답 처리는 전수조사라는 특성상 항목 무응답을 대체하는 방법 연구로 이루어져야 할 것으로 판단되며, 조사자별로 많은 항목이 무응답이거나 또는 단위 무응답일 경우에는 최소한의 기본정보를 이용한 mass imputation을 사용하면 될 것으로 본다. 따라서 무응답 대체를 통해서 무응답에 의한 편향을 줄일 수 있으며, 제조사의 과정을 생략함으로써 조사시간의 단축 및 조사비용의 절감이라는 효과를 얻을 수 있을 것으로 판단된다.

무응답 대체방법은 여러 가지가 있으며, 개별문제에 따라 적절한 대체 방법을 채택하여 이용하는 것이 통상적인 방법이다. 일반적인 사회조사에서 주로 관심의 대상이 되는 범주형 항목의 경우 핫덱대체와 이와 유사한 최근방 기증자(donor) 대체방법이 가장 적합한 대체방법으로 알려져 있으며, 경제관련조사와 같이 연속형 항목들이 주를 이루는 경우는 주로 회귀대체, 최근방대체, 평균대체 방법이 적용되고 있다. 본 연구에서 사용할 농업총조사 자료는 범주형과 연속형 항목들이 혼합되

어 있으나 적절하게 응용한 핫덱 방법을 사용하면 될 것으로 본다. 핫덱을 사용하여 효율적인 결측값 대체를 하기 위해서는 각 항목에 가장 적합한 대체군을 구성해야 한다. 이를 위해서 본 연구에 앞서 연관성 분석 및 대체군 개발연구가 사전에 진행이 되었다. 이 연구의 내용은 『농업총조사 항목 간 연관성 분석 및 대체군(보조변수) 개발(최필근, 2008)』을 참조하면 자세하게 알 수 있을 것이다.

본 연구는 농업총조사의 항목별로 개발된 대체군(보조변수)을 이용하여 각 항목에 맞는 대체방법을 제시하고자 함에 있다. 지금까지 통계청에서 실시하는 조사들에 대하여 인구주택총조사를 제외하면 대체군 개발 및 각 항목에 대한 적절한 대체방법 연구가 상당히 미흡한 실정이다. 그러므로 본 연구를 통해서 체계적이고 정확도가 높은 농업총조사 무응답 대체방법의 기반으로 마련하고, 더 나아가 다른 조사통계의 무응답 처리를 위한 선행연구로서도 활용하고자 한다.

## 2. 연구내용 및 방법

본 연구에서 가장 핵심적으로 다루어지는 내용은 2005년 농업총조사 자료를 이용하여 가구원 및 가구부분의 항목에 대한 대체기법의 개발이다. 이를 위해 농업총조사의 항목 간 연관성 분석을 통하여 개발된 대체군을 이용하여 각 항목에 대하여 가장 적절하게 대체할 수 있는 방법을 제시하고자 한다. 그러나 선행연구에서 밝혔듯이 과수, 작물, 가축부분에 대한 연구는 데이터파일 설계상의 이유로 대체군 개발이 완성된 후에 무응답 대체방법을 제시할 것이다. 이러한 대체군(보조변수)을 사용하는 근본적인 이유는 대체하고자 하는 항목과 연관성이 높은 항목들의 정보를 이용하여 보다 정확한 대체를 하기 위함이다.

본 연구에서는 개발된 대체군을 이용하되 새로운 내용을 추가하도록 한다. 대체군에 속한 모든 항목들에 대하여 목표변수와의 관계를 고려하여 가중치를 부여하는 것이다. 이러한 가중치는 앞서 연구된 CHAID 알고리즘을 이용한 항목 간 연관성 분석의 결과에서 제시된 목표변수와의 연관도(중요도)의 값을 사용하면 될 것으로 판단된다. 가중치를 적용함으로써 대체에 이용되는 가장 적절한 도너(donor)를 찾는 새로운 함수

#### 4 무응답 처리를 위한 방법론 연구 I

로 사용할 수 있을 것이다. 즉, 대체군의 모든 항목의 값들과 완전히 일치하는 도너가 가장 높은 점수를 얻게 되며 일치하지 않는 항목이 많을수록 낮은 점수를 가지게 되지만 각 항목에 부여된 가중치를 고려하여 점수의 총 합계가 결정되게 하는 것이다. 따라서 이 과정을 거쳐 총점이 가장 높은 도너의 값이 대체가 되는 것이며, 이를 위해서 랜덤 핫덱대체 방법을 응용하여 사용하고자 한다.

또 하나의 중요한 사항은 대체를 한 후의 값이 기존의 내검규칙에 위배가 되는지 검토가 되어야 한다. 예를 들면 주종사분야 항목의 경우 나이가 16세 이상이면 그 항목에 기입을 하고, 미만이면 공란으로 두게 된다. 그런데 주종사분야를 대답하지 않은 16세 이상의 사람을 대체하기 위해 찾은 도너가 16세 미만일 경우가 있다. 그러면 대체값을 부여하지 않고 공란으로 두게 되어 문제가 발생하게 된다. 이러한 도너가 선택될 수 있는 이유는 연관성 분석의 결과 주종사분야에 대한 대체군에 나이 항목이 들어가 있지 않기 때문이다. 비록 주종사분야와 나이는 연관성이 없으나 대체후의 결과가 내검규칙에 위배되지 않으려면 나이를 대체군에 포함시켜야 할 것으로 본다. 따라서 모든 항목에 대하여 특별히 대체군에 추가해야 할 항목이 있는지 검토가 되어야 할 것이다. 하지만 이러한 작업은 몇 개의 내검규칙을 만족하기 위한 수단일 뿐 모든 내검규칙에 적합한 값을 한 번에 대체하기는 어려울 것으로 판단된다. 본 연구는 제시된 대체방법의 정확성을 검토하기 위한 것이므로 내검과는 연계하지 않을 것이지만 향후 실제 대체를 할 때에는 대체가 된 후에 다시 내검을 해야 하는 과정을 거쳐야 하므로 이에 관련된 연구도 전문 연구자에 의하여 향후 진행되어야 할 것으로 생각된다.

마지막으로 제시된 대체방법을 이용하여 각 항목별로 모의실험을 실시하고, 이 결과로 대체방법의 정확성을 검토하고자 한다. 2005년 농업총조사의 항목들은 목표변수가 범주형과 연속형 항목이 존재하므로 범주형의 경우는 각 범주의 분포변화 및 정분류율(일치율)을 검토하고, 연속형의 경우는 반복을 통한 표본평균의 차이정도와 항목값을 범주화한 후의 분포변화를 비교할 것이다.

본 연구를 위하여 제2절에서는 무응답을 대체하는 여러 가지 방법에 대하여 설명한다. 제3절에서는 본 연구에서 사용할 농업총조사에 대한 조

사개요와 주요국의 현황에 대하여 간략하게 소개한다. 제4절에서는 본 연구에서 사용하고자 하는 응용된 랜덤 핫덱대체 방법의 내용을 예를 들어 자세하게 설명을 하고, 각 항목에 대한 모의실험을 통하여 본 연구에서 제시한 방법의 효율성을 다양하게 검토할 것이다. 마지막으로 제5절에서는 연구의 최종적인 결론과 더불어 향후 연구되어야 할 내용들을 제시하고자 한다.

## 제2절 무응답 대체방법

무응답 대체방법은 무응답 항목의 대체값으로 한 개의 값을 부여하는 단일 대체방법(single imputation)과 여러 개의 값을 대체하는 다중 대체방법(multiple imputation, Rubin(1987))으로 구분되며, 단일 대체방법은 무응답 항목에 유일하게 결정된 대체값을 대입하는 결정적 대체방법(deterministic imputation)과 대체값을 확률적으로 결정하여 대입하는 확률적 대체방법(stochastic imputation)으로 구분된다. 이 절에서는 일반적으로 쓰이고 있는 단일 대체방법에 대하여 간략하게 소개하고자 한다.

### 1. 결정적 대체방법

결정적 대체방법에서의 대체값은 실제 조사된 응답값 중의 하나를 선택할 수도 있고, 다른 보조변수를 이용하여 대체값을 만들어 적용할 수도 있다. 이때 대체되는 목표변수의 값이 연속형인 경우는 어느 방법을 이용해도 무방하지만 범주형인 경우에는 최종 대체값이 이산형이 되도록 보정을 해 주어야 하는 과정이 더 필요하다.

#### 가. 연역적 대체방법(deductive imputation)

논리적인 제약조건이나 다른 기록에 의하여 확실하게 대체값을 지정하여 무응답을 대체하는 방법이다. 이 방법은 내검과정에서 응답값이 논리적으로 타당하지 않을 때에나 합계를 구성하고 있는 하나의 항목이 무응답일 경우 주로 이용된다.

### 나. 시기적 대체방법(historical imputation)

반복조사(repeated survey)에서 매우 유용하게 사용되는데, 만일 동일 항목의 응답값이 조사시점에 따라 안정된 값을 보이고 전회 조사값과 금회 조사값의 상관관계가 높으면 금회 조사의 무응답 항목에 전회 조사의 응답값을 대체하는 방법이다. 이 방법은 특히 응답패턴으로 인한 편향의 영향을 받지 않는 장점이 있다.

### 다. 평균 대체방법(mean imputation)

무응답 항목에 목표변수의 전체 평균을 대입하거나 또는 대체군 내의 평균을 대입하는 방법이다. 즉,

$$y_k^* = \begin{cases} y_k, & k \in R, & \text{응답값} \\ \bar{y}_R, & k \in R^c, & \text{대체값} \end{cases}$$

여기서  $R$ 은 응답 집합,  $R^c$ 은 무응답 집합이며  $\bar{y}_R$ 은 응답값의 평균이다. 이 방법은 간단하여 이용되기 쉬운 장점이 있으며, 항목이 양적 변수이고 구하고자 하는 통계량이 평균일 때 유용하다. 그러나 대체 후의 값들은 평균값의 빈도수가 지나치게 많아져 응답값들의 분포가 왜곡되고, 중위수나 백분위수와 같은 평균이 아닌 통계량을 구할 때에는 효율이 저하되는 단점이 있다.

### 라. 축차 핫덱 대체방법(sequential hot-deck imputation)

데이터 파일을 구성할 때 직전에 응답한 단위의 항목값으로 대체하는 방법이다. 즉,

$$y_k^* = \begin{cases} y_k, & k \in R, & \text{응답값} \\ y_i, & k \text{ 직전의 } i \in R, & k \in R^c, & \text{대체값} \end{cases}$$

이 방법은 사회·인구통계조사에서 유용하다. 표본은 사회·인구통계적 지표에 의해서 자연스럽게 대체군으로 구분되며, 대체군 내의 항목값은 서로 유사할 가능성이 높다. 또한 이러한 조사는 지리적인 연속

성을 가지므로 무응답이 발생하면 직전의 응답값으로 대체하는 것이 타당할 것이다. 하지만 동일한 값을 여러 번 사용하게 될 수 있다는 위험성이 있으며, 결측값을 할당하는 데 있어서 확률구조가 아닌 자료파일 순서에 의존한다는 단점이 있다.

#### 마. 비대체 방법(ratio imputation)

이용 가능한 보조변수가 있을 때 목표변수와 보조변수의 관계를 이용하여 무응답을 대체하는 방법이다. 만일 목표변수와 보조변수가 비례 관계이면 다음과 같다.

$$y_k^* = \begin{cases} y_k, & k \in R, & \text{응답값} \\ \left(\frac{\bar{y}_R}{\bar{x}_R}\right) x_k, & k \in R^c, & \text{대체값} \end{cases}$$

이 방법은 목표변수와 보조변수가 원점을 지나는 직선관계이며, 분산이 보조변수에 비례하는 경우 효과적인 것으로 나타났다. 유한 모집단에서 예측이론에 의하면 비대체방법이 매우 우수한 것으로 알려져 있으나 항목들이 양적변수일 때에만 사용가능하다.

#### 바. 회귀대체 방법(regression imputation)

비대체방법과 유사하게 사용되며, 목표변수와 보조변수의 관계가 절편이 있는 직선 관계이고 목표변수의 분산이 동일할 때 유용한 방법이다. 즉,

$$y_k^* = \begin{cases} y_k, & k \in R, & \text{응답값} \\ \bar{y}_R + b(x_k - \bar{x}_R), & k \in R^c, & \text{대체값} \end{cases}$$

여기서  $b = \frac{\sum_{k \in R} (x_k - \bar{x}_R)(y_k - \bar{y}_R)}{\sum_{k \in R} (x_k - \bar{x}_R)^2}$  이다. 이 방법은

미국의 인구조사(CPS: current population survey)에서 발생하는 결측값을 대체하기 위해서 이용하였으며, 그 결과 대체된 값과 실제값의 평균절대편차를 비교할 때 다른 대체 방법에 비해 매우 적절함을 보였다.

### 사. 최근방 대체방법(nearest-neighbor imputation)

보조변수를 이용하여 응답하지 않은 개체와 가장 유사한 응답개체를 찾아 대응되는 항목값을 대체하는 방법으로 현실적으로 이용 가능성이 높은 방법이다. 즉,

$$y_k^* = \begin{cases} y_k, & k \in R, & \text{응답값} \\ y_i, i: \min_{i \in R} \|x_k - x_i\|, & k \in R^c, & \text{대체값} \end{cases}$$

여기서 거리가 가장 가까운 응답값을 찾는 방법은 여러 가지가 있으며, 주로 절대거리를 많이 이용한다. 이 방법은 활용 가능성은 높지만 주어진 보조변수가 적절치 않거나 무응답률이 높은 경우 추정에 큰 편향을 가져올 수 있다.

## 2. 확률적 대체방법

결정적 대체방법은 대체값을 유일하게 결정하기 때문에 목표변수의 변동을 줄이는 경향이 있다. 이러한 단점을 보완하기 위하여 확률적 대체방법이 제안되었으며, 이 방법은 대체값에 확률적인 변동을 부여해 줌으로써 결정적 대체방법의 단점을 보완할 수 있다.

### 가. 랜덤 핫덱 대체방법(random hot-deck imputation)

대체군 내의 응답값 중에서 하나를 임의로 선택하여 대체하는 방법이다. 즉,

$$y_k^* = \begin{cases} y_k, & k \in R, & \text{응답값} \\ y_i, \text{적당한 } i \in R, & k \in R^c, & \text{대체값} \end{cases}$$

랜덤 핫덱 대체를 이용하면 무응답 대체 후에도 표본의 분포가 그대로 유지될 수 있다는 장점이 있다. 또한 평균 대체나 회귀대체 등과는 달리 표본 분포가 유지되므로 통계량의 형태에 무관하게 이용될 수 있다. 그러나 응답 패턴이 목표변수와 무관한 경우에 더 적합하며 경제관련 조사보다는 사회관련 조사에 주로 사용된다.



### 나. 가중 핫덱 대체방법(weighted hot-deck imputation)

랜덤 핫덱 방법과 유사하나 응답값 중에서 하나를 선정할 때 가중값을 두어 선정하는 방법이다. 이 방법은 복합표본에서 층화나 집락화로 인하여 서로 다른 추출확률에 의하여 표본이 선정될 때 랜덤 핫덱 방법 대신에 주로 이용된다.

### 다. 랜덤 비대체방법(random ratio imputation)

기존의 비대체값에 확률오차를 포함시켜 대체값으로 이용하는 방법이다.

$$y_k^* = \begin{cases} y_k, & k \in R, & \text{응답값} \\ \left(\frac{\bar{y}_R}{\bar{x}_R}\right) x_k + e_k, & k \in R^c, & \text{대체값} \end{cases}$$

여기에 더해주는 확률오차  $e_k$ 는 대체로 인한 변동의 감소를 보정해주는 효과가 있으며, 그 크기는 대체 전 변동과 대체 후 변동이 동일하게 되도록 변동의 폭을 계산하여 포함시킨다. 랜덤 회귀대체방법 역시 기존의 대체값에 확률오차의 효과를 주어 대체한다.

## 제3절 농업총조사 개요 및 주요국 현황

이 절에서는 본 연구에서 사용되는 농업총조사의 개요와 주요국의 현황에 대하여 간략하게 정리하고자 한다.

### 1. 농업총조사 개요

농업총조사는 조사기준시점(현재 2005. 12. 1)에서 농가정의에 해당하는 규모이상의 농업을 직접 경영하는 모든 농가의 현황과 경영실태를 파악하기 위해 실시하는 전수조사이다. 현재 농가의 정의는 조사 시기

별로 농가구조의 변화를 반영하는 방향으로 수정되는데, 2000·2005년 농업총조사의 경우 이전에 비하여 단순화되어 조사기준 현재 경지 10a (약 300평) 이상을 직접 경작하는 가구와 연간 농축산물의 판매금액이 50만원 이상으로 농업을 계속하는 가구(단, 판매금액이 50만원 미만이라도 조사기준시점 현재 50만원 이상의 가축을 사육하는 가구는 포함)로 경작면적과 농업수입만을 기준으로 정하고 있다.

우리나라 농업통계조사는 조선시대 말기인 1908년부터 농가인구, 파종면적과 수확량 등을 조사한 것이 시초가 되었으며, 지금과 같이 농업총조사를 시작한 것은 1960년 세계농업센서스에 처음 참가하면서부터이고 농업전반에 대해 세밀한 자료를 파악하여 농업통계의 기준을 마련하고 아울러 농업통계 개선에 획기적인 전환점을 마련하였다. 1961년 ‘제1차 농업국세조사’를 시작으로 2000년까지 10년 주기로 실시하여 왔으며, 실시기간이 길어서 발생하는 자료의 부족함을 채우기 위하여 1975년과 1985년에는 간이농업센서스를 실시하였다. 1995년에는 세계무역기구(WTO)체제의 출범과 지방자치제의 실시 등으로 대내외적인 농업여건이 급변함에 따라 제5차 농업총조사를 특별히 5년 만에 실시하였다. 농업총조사는 1995년까지 농림부가 주관하여 지방자치단체를 통해 실시하였으나, 1998년 정부조직법의 개정으로 업무가 통계청에 이관됨에 따라 2000년부터 통계청이 주관하여 지방자치단체를 통해 실시하였으며, 특히 2005년부터는 시의성 있는 자료제공을 위해 총 조사주기를 5년으로 단축하여 조사를 실시하고 있다.

본 조사는 조사원이 조사구 내 모든 농가를 직접 방문하여 면접조사를 하고 있으며, 주택총조사와 연계하고 농림어업총조사를 동시에 조사하고 있다. 조사에서 발생하는 무응답의 처리는 이전의 조사 자료를 참조하고, 조사원들이 계속적으로 재방문하거나 전화를 이용하여 보완 작업이 이루어지므로 현재 특별한 무응답 처리기법을 적용하여 무응답을 대체하고 있지는 않은 실정이다.

농업총조사에 대한 자세한 연혁은 이전에 연구되었던 『농업총조사 항목 간 연관성 분석 및 대체군(보조변수) 개발(최필근, 2008)』에서 내용을 발췌하여 <표 3-1>에 제시하고자 한다.

〈표 3-1〉 시기별 농업총조사의 명칭과 의의

| 명칭         | 조사기준일        | 의의 및 특징  |
|------------|--------------|--|
| 제1차 농업국세조사 | 1961. 2. 1   | · 제3회 세계농업센서스에 처음 참여<br>· 농업통계를 획기적으로 재정비  |
| 제2차 농업센서스  | 1970. 12. 1  | · 조사결과를 읍·면별로 집계공표   |
| 간이농업센서스    | 1975. 12. 1  | · 표본조사로 실시하여 농업자료를 보완  |
| 제3차 농업조사   | 1980. 12. 11 | · 조사원을 지역실정과 농업부문에 정통한 사람으로 선발   |
| 간이농업센서스    | 1985. 12. 1  | · 농수산통계조사 공무원이 조사하여 조사상의 미숙으로 인한 오차를 줄임  |
| 제4차 농업총조사  | 1990. 12. 1  | · 영농형태, 농지임차 및 위탁 등에 관한 사항을 추가   |
| 제5차 농업총조사  | 1995. 12. 1  | · 농가정의를 판매금액 기준으로 변경<br>· 농축산물 판매금액 및 방법 등 항목 추가   |
| 제6차 농업총조사  | 2000. 12. 1  | · 1998. 7. 작성기관을 농림부에서 통계청으로 변경<br>· 통계청으로 이관 후 처음 실시<br>· 어업총조사와 통합 실시<br>· 친환경농업, 농가의 정보화 현황 등 미래지향적 항목 추가 |
| 제7차 농업총조사  | 2005. 12. 1  | · 임업 및 어업총조사와 통합 실시<br>· 혼인상태, 농업관련사업 등 농가인구 특성, 소득창출을 위한 사업 파악을 위한 항목 추가                                    |

## 2. 주요국의 농업총조사 현황

통계청에서 실시하고 있는 농업총조사처럼 많은 나라에서도 이와 같은 조사를 하고 있다. 몇몇 주요 국가들에 대하여 조사방법 등 기본적인 현황을 정리하고자 한다.

### 가. 미국

미국은 농무성(United States Department of Agriculture)에서 농업총조사를 담당하고 있다. 1840년 처음 조사를 실시하여 10년 단위로 조사를 해오다 현재는 끝자리 숫자가 2, 7인 연도에 5년 단위로 조사하고 있으며, 가장 최근의 조사는 2007년의 조사로 조사 기준일은 2007년 12월 31일이다. 조사대상은 농가 단위가 아닌 농장, 목장 등 경영체를 조사대상으로 삼고 있으며 전체의 80%는 전수조사 항목, 20%는 표본조사 항목으로 조사된다. 주로 우편, 전화, 면접조사를 단계별로 적용하여 조사가 실시되나, 2007년 조사에서는 우편과 인터넷조사를 병행하여 조사되었다. 조사에서 발생하는 무응답에 대해서는 과거자료를 이용한 대체 및 유사 특성을 보유한 인접 경영체 자료를 찾아 무응답 경영체에 대체하고 있다.

### 나. 캐나다

캐나다는 통계청(Statistics Canada)에서 농업총조사를 담당하고 있으며 1871년 처음 조사를 실시하여 10년 단위로 조사를 해오다 1956년부터 끝자리 숫자가 1, 6인 연도에 5년 단위로 조사하고 있다. 최근의 조사는 2006년 농업총조사로 조사 기준일은 2006년 5월 16일이다. 조사대상은 미국과 같은 농장, 목장 등 경영체이며 전수조사를 실시하고 있다. 설문지는 도시는 우편으로, 농촌은 직접 배부를 하고 있으며 회수는 사생활 보호를 목적으로 우편으로 받고 있다. 또한 경우에 따라 인터넷조사를 병행하여 조사한다. 무응답에 대해서는 전화를 통한 사후 보완을 실시하고, 이것으로도 완전하지 않을 때에는 무응답 경영체와 특성이 가장 가까운 경영체를 조사된 데이터로부터 찾아서 자동적으로 대체를

하게 된다.

#### 다. 호주

호주도 캐나다처럼 통계청(Australian Bureau of Statistics)에서 농업총 조사를 담당하고 있으며 1850년 처음 조사를 실시하였으며 현재는 끝자 리 숫자가 1, 6인 연도에 5년 단위로 조사하고 있다. 최근의 조사는 2006년 농업총조사로 조사 기준일은 2006년 6월 30일이다. 조사대상은 미국, 캐나다와 유사한 농업 사업체이며 전수조사를 실시하고 있다. 조 사는 주로 우편에 의하여 실시된다. 무응답에 대해서는 미국과 같이 과 거자료를 이용한 대체 및 유사 특성을 보유한 인접 사업체 자료를 찾아 내어 무응답 사업체에 대체하고 있다.

#### 라. 영국

영국은 통계청(UK Statistics Authority)의 환경·식품·농촌개발부에서 농업관련조사를 담당하고 있다. 1866년 처음 조사를 실시하여 10년 단위로 조사되고 있으며 최근의 조사는 2000년 농업·원예조사로 조사 기준일은 2000년 6월 1일이다. 조사대상은 등록 경작권(holdings)을 대상으로 우편으로 전수조사를 실시하고 있다. 또한 매년 표본조사를 실시 하고 있는데 210,000개의 등록 경작권 중 표준 노동 요구에 따라 층화하여 랜덤으로 61,000개의 표본을 뽑아 우편과 온라인을 병행하여 표본조사 항목을 조사하고 있다. 무응답에 대해서는 주로 전화를 통한 사후 보완을 실시하고 있다.

#### 마. 일본

일본은 농림수산성(Ministry of Agriculture, Forestry and Fisheries)에서 농업총조사를 담당하고 있으며 1960년 처음 조사가 시작되어 현재는 끝 자리 숫자가 0, 5인 연도에 5년 단위로 조사하고 있다. 최근의 조사는 2005년이며 조사 기준일은 2005년 2월 1일이다. 조사대상은 농가 및 관련 사업체이며 전수조사를 실시하고 조사원이 직접 면접조사를 한다.

무응답 처리는 한국과 유사하게 재방문, 전화조사 및 이웃으로부터의 정보를 이용하여 보완하는 것으로 알려져 있다.

## 바. 대만

대만은 행정원(Council of Agriculture)에서 농업총조사를 담당하고 있으며 1956년 처음 조사가 시작되어 현재는 끝자리 숫자가 0, 5인 연도에 5년 단위로 조사하고 있다. 최근의 조사는 2005년이며 조사 기준일은 2005년 12월 31일이다. 조사대상은 농가 및 관련 사업체이며 전수조사를 실시하고 조사원이 직접 면접조사를 한다. 한국, 일본, 대만의 농업총조사는 유사한 형태를 보인다.

## 제4절 농업총조사 무응답 대체방법

이 절에서는 본 연구에서 사용하고자 하는 응용 핫덱 대체방법에 대하여 자세하게 설명하고, 개발된 대체군을 이용하여 항목별로 모의실험을 실시하여 대체의 정확도를 검토하고자 한다. 2005년 농업총조사의 항목들은 목표변수가 범주형과 연속형 항목이 존재하므로 범주형의 경우는 각 범주의 분포변화 및 정분류율을 검토하고, 연속형의 경우는 반복을 통한 표본평균의 차이정도를 비교할 것이다.

### 1. 응용 핫덱 대체방법

제2절에서 다양한 무응답 대체방법을 소개하였다. 이러한 무응답 대체방법들은 일반적으로 자료의 특성(형태)에 의해서 적용되는 방법들이 있다. 목표변수가 연속형인 경우에는 회귀대체, 최근방대체, 평균대체 방법을 주로 사용하고, 범주형인 경우에는 확률적 대체방법, 핫덱 대체방법, 순차적(hierarchical) 핫덱 대체방법 등을 주로 이용하게 된다. 그러나 본 연구에서 적용하고자 하는 농업총조사 자료는 연속형 항목과 범주형 항목이 혼합되어 있으며 개발된 각 항목별 대체군에도 두 종류

의 형태가 섞여있다. 따라서 기존의 방법들을 바로 적용하기에는 여러 측면에서 한계가 있을 것으로 판단된다. 그러므로 기존의 핫덱 대체방법을 응용하여 모든 항목에 대하여 일괄적으로 적용할 수 있는 새로운 방법을 제시하고자 한다.

핫덱 대체방법을 응용해서 사용하고자 하는 주 이유는 다음과 같다. 첫째, 적절한 대체군에 속한 모든 항목의 내용을 만족한다는 가정 하에서는 핫덱 대체방법은 확률적 대체방법보다는 대체의 정확도가 높을 것이며 순차적 핫덱 대체방법과는 같은 결과가 나올 것이다. 하지만 핫덱 대체의 단점인 응답자가 무응답자 수보다 커야 한다는 조건을 만족하지 않을 경우 순차적 핫덱 대체방법을 사용하게 된다. 그러나 순차적 핫덱 대체방법에도 단점이 존재하는데 이는 대체군에서 빠져나간 항목은 다시 사용되지 않는다는 점이다. 이러한 이유 때문에 도너의 선택 시 제약이 있을 수가 있다고 본다. 따라서 본 연구에서는 각각의 단점을 보완하여 두 방법이 혼합된 새로운 방법을 제시하고자 한다.

둘째, 농업총조사의 양적자료 특성 때문이다. 다시 말하면 농업총조사 항목들 중에서 대부분의 양적자료들은 일정한 부분만 값이 존재하고 나머지는 0의 값을 갖는다. 이는 다른 조사에서의 자료와는 많은 차이가 있다고 할 수 있다. 예를 들면 키나 몸무게처럼 대부분의 개체가 적절한 값을 가지는 변수는 회귀대체나 비대체를 하는 것이 합당할 것이다. 하지만 논면적이나 목초지면적 등 농업총조사의 양적변수들은 소유하고 있지 않은 경우 이 값들은 모두 0으로 주어져 있다. 따라서 회귀대체를 한다면 소유하지 않은 가구에 대해서 비록 작은 값이라고 할지라도 많은 경우 0보다 큰 값이 대체될 것이다. 이러한 대체결과는 아마도 구성비 측면에서 큰 문제가 될 수도 있을 것이다. 비대체 역시 0의 값이 많아 사용하기에 적절하지 않을 것이다. 따라서 대체후의 값에 대해서도 고려하여 작물면적을 소유하지 않은 가구는 0으로 대체를 하고 소유한 가구에 대해서는 적절한 값을 대체할 수 있는 방법이 필요하다.

셋째, 대체한 후의 분포가 일정하게 하는 것이 매우 중요하다고 할 수 있다. 핫덱 대체방법은 대체 후에도 표본의 분포가 일정하게 유지될 수 있다는 큰 장점이 있어 핫덱 대체방법을 응용하여 사용하는 것이 농업총조사 자료에 가장 적절할 것으로 판단된다.

### 가. 설명을 위한 자료

제시하고자 하는 응용 핫덱 대체방법의 내용을 쉽게 이해할 수 있게 하기 위하여 적당한 예를 가지고 설명한다. 설명을 위한 항목은 성별을 선택했으며 사전의 연관성 분석 연구에서 개발된 대체군과 목표변수와의 연관도(중요도)는 <표 3-2>와 같다.

<표 3-2> 성별 항목에 대한 대체군

| 성별      |          |        |          |
|---------|----------|--------|----------|
| 1차 대체군  | 연관도(중요도) | 2차 대체군 | 연관도(중요도) |
| 경영주와의관계 | 100      | 농업종사경력 | 100      |
| 혼인상태    | 62       | 혼인상태   | 90       |
| 농업종사경력  | 23       | 교육정도   | 42       |
|         |          | 나이     | 25       |

성별 항목이 무응답일 경우 경영주와의관계, 혼인상태, 농업종사경력의 정보를 가지고 대체를 하며 만일 경영주와의 관계도 무응답이면 2차 대체군인 농업종사경력, 혼인상태, 교육정도, 나이를 이용하는 것이다. 이 때 각 항목에 대한 가중치로써 연관도(중요도) 값을 부여하여 가장 좋은 도너를 찾기 위한 함수로 사용할 것이다. 하지만 농업총조사의 경우 자료의 수가 매우 방대하여 모든 대체군의 조건을 만족하는 도너가 대부분의 경우에 존재할 것으로 판단되기 때문에 큰 의미가 없을 수도 있으나, 이 방법은 향후 자료의 수가 작은 조사에서도 이용될 수 있으므로 연관도의 사용은 매우 중요할 것으로 보인다.

다음의 <표 3-3>은 실제 농업총조사 자료 중에서 일부를 발췌하여 작성한 것으로 설명을 위한 자료로 사용하고자 한다. 총 자료의 수는 3,448,057명이지만 임의로 30명의 자료만 가지고 설명을 할 것이다.



〈표 3-3〉 2005년 농업총조사 자료일부

| 개체<br>번호  | 성별          | 경영주와<br>의관계 | 혼인상태 | 농업종사<br>경력 | 교육정도 | 나이 |
|-----------|-------------|-------------|------|------------|------|----|
| 1         | 1           | 1           | 2    | 3          | 2    | 64 |
| 2         | 2           | 2           | 2    | -1         | 4    | 64 |
| 3         | 2           | 3           | 1    | -1         | 6    | 35 |
| 4         | 1           | 1           | 2    | 5          | 4    | 66 |
| 5         | 2           | 2           | 2    | -1         | 3    | 59 |
| 6         | 1           | 1           | 2    | 5          | 2    | 65 |
| 7         | 1           | 3           | 1    | -1         | 6    | 27 |
| 8         | 2           | 1           | 3    | 40         | 4    | 76 |
| 9         | 1           | 1           | 2    | 10         | 6    | 51 |
| 10        | 2           | 3           | 1    | -1         | 6    | 22 |
| 11        | 1           | 3           | 1    | -1         | 4    | 19 |
| 12        | 1           | 1           | 2    | 13         | 4    | 72 |
| 13        | 2           | 2           | 2    | -1         | 4    | 68 |
| 14        | 2           | 2           | 2    | -1         | 4    | 48 |
| <b>15</b> | <b>1(A)</b> | <b>1(B)</b> | 2    | 30         | 2    | 62 |
| 16        | 2           | 2           | 2    | -1         | 2    | 57 |
| 17        | 2           | 3           | 1    | -1         | 4    | 24 |
| 18        | 2           | 1           | 3    | 29         | 3    | 64 |
| 19        | 1           | 1           | 2    | 29         | 2    | 60 |
| 20        | 2           | 2           | 2    | -1         | 3    | 51 |
| 21        | 2           | 3           | 1    | -1         | 6    | 25 |
| 22        | 2           | 3           | 1    | -1         | 4    | 23 |
| 23        | 2           | 3           | 1    | -1         | 4    | 16 |
| 24        | 2           | 7           | 3    | -1         | 2    | 68 |
| 25        | 1           | 1           | 2    | 31         | 6    | 51 |
| 26        | 2           | 2           | 2    | -1         | 6    | 46 |
| 27        | 2           | 3           | 1    | -1         | 4    | 18 |
| 28        | 1           | 3           | -1   | -1         | 2    | 11 |
| 29        | 1           | 3           | 1    | -1         | 4    | 26 |
| 30        | 2           | 3           | 1    | -1         | 4    | 16 |

자료 중에서 (-1)의 값을 하고 있는 자료가 존재한다. 이것은 연구자가 임의로 만든 값으로 허용되는 무응답을 표시하기 위함이다. 즉, 혼인상태 항목에서는 16세 미만이면 내용을 기입을 하지 않도록 되어 있다. 28번 자료의 값을 보면 (-1)이며 실제 나이를 보면 11세로 되어 있음을 알 수 있다. 또한 농업종사경력 항목은 경영주만이 기입을 하는 것으로 경영주와의관계가 1인 경우에만 자료의 값이 존재하게 된다. 주어진 자료에서 보면 경영주와의관계가 1이 아닌 경우는 (-1)을 부여하여 허용하는 무응답임을 표시하고 있다.

#### 나. 기본 알고리즘

<표 3-3>은 임의의 30명의 자료이며 완전하게 조사가 된 내용들이다. 무응답 대체과정을 설명하기 위하여 15번째 사람이 성별에 대하여 무응답을 했다고 가정하자. 따라서 (A)항목은 무응답임을 표시하기 위해 (-2)의 값을 부여하게 된다. 하지만 실제 이 사람의 성별은 남자(1)이다.

**step 1** : 첫 번째 가구원부터 성별 항목의 값이 (-2)인지 아닌지를 체크한다. 만일 (-2)가 아니면 다음 사람으로 이동하고 맞으면 성별과 가장 연관성이 높은 경영주와의관계 항목과의 일치여부를 판단하여 점수화한다. 이때 경영주와의관계 항목(B)이 무응답이면 **step 2-1** 로 이동한다.

- 경영주와의관계가 1(경영주)인 개체에는 100점을 부여한다.
- 경영주와의관계가 1이 아닌 개체에는 0점을 부여한다.

step 1의 과정이 끝나고 나면 <표 3-4>와 같이 가장 점수가 높은 총 9명의 예비 도너가 선택된다. 현재의 점수는 100점이 되며 경영주와의관계가 1인 것만이 선택된 것을 볼 수 있다. 또한 경영주와의관계 항목이 성별과 높은 연관성을 가지므로 선택된 예비도너가 적절하다는 것을 알 수 있다.

〈표 3-4〉 step 1 절차 후의 예비도너(donor)

| 개체<br>번호 | 성별   | 경영주와<br>의관계 | 혼인상태 | 농업종사<br>경력 | 교육정도 | 나이 | 점수  |
|----------|------|-------------|------|------------|------|----|-----|
| 1        | 1    | 1           | 2    | 3          | 2    | 64 | 100 |
| 4        | 1    | 1           | 2    | 5          | 4    | 66 | 100 |
| 6        | 1    | 1           | 2    | 5          | 2    | 65 | 100 |
| 8        | 2    | 1           | 3    | 40         | 4    | 76 | 100 |
| 9        | 1    | 1           | 2    | 10         | 6    | 51 | 100 |
| 12       | 1    | 1           | 2    | 13         | 4    | 72 | 100 |
| 15       | 1(A) | 1(B)        | 2    | 30         | 2    | 62 |     |
| 18       | 2    | 1           | 3    | 29         | 3    | 64 | 100 |
| 19       | 1    | 1           | 2    | 29         | 2    | 60 | 100 |
| 25       | 1    | 1           | 2    | 31         | 6    | 51 | 100 |

**step 2 :** 첫 번째 절차가 끝난 후 두 번째로 성별과 연관성이 높은 혼인상태 항목과의 일치여부를 판단하여 이전에 획득한 점수와 합산한다.

- 혼인상태가 2(배우자 있음)인 개체에는 62점을 부여한다.
- 혼인상태가 2가 아닌 개체에는 0점을 부여한다.

〈표 3-5〉 step 2 절차 후의 예비도너(donor)

| 개체<br>번호 | 성별   | 경영주와<br>의관계 | 혼인상태 | 농업종사<br>경력 | 교육정도 | 나이 | 점수  |
|----------|------|-------------|------|------------|------|----|-----|
| 1        | 1    | 1           | 2    | 3          | 2    | 64 | 162 |
| 4        | 1    | 1           | 2    | 5          | 4    | 66 | 162 |
| 6        | 1    | 1           | 2    | 5          | 2    | 65 | 162 |
| 9        | 1    | 1           | 2    | 10         | 6    | 51 | 162 |
| 12       | 1    | 1           | 2    | 13         | 4    | 72 | 162 |
| 15       | 1(A) | 1(B)        | 2    | 30         | 2    | 62 |     |
| 19       | 1    | 1           | 2    | 29         | 2    | 60 | 162 |
| 25       | 1    | 1           | 2    | 31         | 6    | 51 | 162 |

step 2의 과정이 끝나고 나면 <표 3-5>와 같이 합산한 점수가 가장 높은 개체가 다시 선택된다. 이 예제에서는 남은 도너의 성별이 모두 남자(1)로 정확한 대체가 가능할 것이다. 실제 많은 자료를 가지고 도너를 구성하여도 많은 개체가 1이고 2는 많지 않을 것이다. 이러한 현상은 연관성 분석의 결과 성별과 경영주와의관계, 혼인상태는 매우 높은 관련이 있기 때문이다.

**step 3** : 두 번째 절차가 끝난 후 세 번째로 성별과 연관성이 높은 농업종사경력 항목과의 일치여부를 판단하여 이전에 획득한 점수와 합산한다.

앞의 두 절차와는 달리 농업종사경력은 연속형 항목이다. 따라서 일치여부를 같은 값으로 판단하기에는 어려움이 있을 것으로 본다. 특히, 논면적이나 밭면적과 같은 값들은 매우 다양한 값들이 있어 도너를 선택할 때 일치되는 것이 나오지 않을 수도 있을 것이다. 그래서 여기에 신뢰구간의 개념을 도입하기로 한다. 농업종사경력이 30년과 같은 개체는 없으나 이와 비슷한 값은 존재한다. 따라서 이 값으로부터 2.5% 또는 5%의 오차 범위 안에 있다면 같은 값으로 간주하는 것이다. 즉,

농업종사경력이  $(30) \times (0.95) = 28.5$  보다 많고  $(30) \times (1.05) = 31.5$  보다 적으면 30년과 같은 값으로 인정을 할 것이다. 그리고 이 비율(%)의 크기는 항목에 따라 다르게 적용할 것이다. 결국 이러한 규칙을 적용함으로써 범주형이나 연속형 자료에 관계없이 이 방법을 사용할 수 있을 것으로 판단된다. 이로부터 농업종사경력의 일치여부를 판단하여 23 점을 부여한다. 이 과정이 끝나고 나면 <표 3-6>과 같이 합산한 점수가 가장 높은 개체가 최종적으로 선택된다.

<표 3-6> step 3 절차 후의 예비도너(donor)

| 개체 번호 | 성별   | 경영주와의관계 | 혼인상태 | 농업종사 경력 | 교육정도 | 나이 | 점수  |
|-------|------|---------|------|---------|------|----|-----|
| 15    | 1(A) | 1(B)    | 2    | 30      | 2    | 62 |     |
| 19    | 1    | 1       | 2    | 29      | 2    | 60 | 185 |
| 25    | 1    | 1       | 2    | 31      | 6    | 51 | 185 |

**step 4** : 최종적으로 선택된 도너의 성별 값을 무응답 가구원의 성별에 대체한다.

- 최종적으로 선택된 도너가 1개이면 그 값을 대체한다.
- 최종적으로 선택된 도너가 2개 이상이면 랜덤하게 뽑아서 선택된 도너의 값을 대체한다.

도너의 값을 대입하는 과정에서 한 가지 주의할 것이 있다. 선택된 도너의 성별이 무응답일 경우가 있으므로 대입과정에 이 여부를 체크하는 것이 필요하다. 따라서 선택된 도너의 성별이 무응답이면 다시 랜덤하게 뽑는 과정을 반복하여 대입을 진행해야 한다. 본 예에서는 <표 3-7>과 같이 25번이 랜덤하게 뽑혔으므로 이것을 최종도너로 사용하게 되며 대체하는 성별이 남자(1)로 정확한 대체가 되었음을 알 수 있다.

<표 3-7> step 4 절차 후의 최종도너(donor)

| 개체 번호 | 성별   | 경영주와의 관계 | 혼인상태 | 농업종사 경력 | 교육정도 | 나이 | 점수  |
|-------|------|----------|------|---------|------|----|-----|
| 15    | 1(A) | 1(B)     | 2    | 30      | 2    | 62 |     |
| 25    | 1    | 1        | 2    | 31      | 6    | 51 | 185 |

**step 5** : 성별 항목의 값이 (-2)가 나오지 않을 때까지 **step 1** 로 돌아가 계속적으로 반복한다.

### < 2차 대체군 사용을 위한 알고리즘 >

**step 2-1** : 경영주와의관계 항목을 사용할 수 없을 경우 2차 대체군을 이용하여 앞서와 같은 방법으로 대체한다. 성별과 가장 연관성이 높은 농업종사경력 항목과의 일치여부를 판단하여 점수화한다.

- 농업종사경력이 28.5년에서 31.5년 사이인 개체에는 100점을 부여한다. 그 외에는 0점을 부여한다.
- 그 외의 값을 가진 개체는 0점을 부여한다.

step 2-1의 과정이 끝나고 나면 <표 3-8>과 같이 가장 점수가 높은 총 3명의 예비 도너가 선택된다.

<표 3-8> step 2-1 절차 후의 예비도너(donor)

| 개체<br>번호 | 성별   | 경영주와<br>의관계 | 혼인상태 | 농업종사<br>경력 | 교육정도 | 나이 | 점수  |
|----------|------|-------------|------|------------|------|----|-----|
| 15       | 1(A) | 1(B)        | 2    | 30         | 2    | 62 |     |
| 18       | 2    | 1           | 3    | 29         | 3    | 64 | 100 |
| 19       | 1    | 1           | 2    | 29         | 2    | 60 | 100 |
| 25       | 1    | 1           | 2    | 31         | 6    | 51 | 100 |

**step 2-2** : 다음으로 두 번째로 성별과 연관성이 높은 혼인상태 항목과의 일치여부를 판단하여 이전에 획득한 점수와 합산한다.

- 혼인상태가 2(배우자 있음)인 개체에는 90점을 부여한다.
- 혼인상태가 2가 아닌 개체에는 0점을 부여한다.

step 2-2의 과정이 끝나고 나면 <표 3-9>와 같이 합산한 점수가 가장 높은 개체가 다시 선택된다.

<표 3-9> step 2-2 절차 후의 예비도너(donor)

| 개체<br>번호 | 성별   | 경영주와<br>의관계 | 혼인상태 | 농업종사<br>경력 | 교육정도 | 나이 | 점수  |
|----------|------|-------------|------|------------|------|----|-----|
| 15       | 1(A) | 1(B)        | 2    | 30         | 2    | 62 |     |
| 19       | 1    | 1           | 2    | 29         | 2    | 60 | 190 |
| 25       | 1    | 1           | 2    | 31         | 6    | 51 | 190 |

**step 2-3** : 다음으로 세 번째로 성별과 연관성이 높은 교육정도와의 일치여부를 판단하여 이전에 획득한 점수와 합산한다.

- 교육정도가 2(초등학교)인 개체에는 42점을 부여한다.
- 교육정도가 2가 아닌 개체에는 0점을 부여한다.

step 2-3의 과정이 끝나고 나면 <표 3-10>과 같이 합산한 점수가 가

장 높은 개체가 다시 선택된다. 이 예에서는 하나의 개체만 남았지만 실제는 많은 개체가 선택된다.

<표 3-10> step 2-3 절차 후의 예비도너(donor)

| 개체<br>번호 | 성별   | 경영주와<br>의관계 | 혼인상태 | 농업종사<br>경력 | 교육정도 | 나이 | 점수  |
|----------|------|-------------|------|------------|------|----|-----|
| 15       | 1(A) | 1(B)        | 2    | 30         | 2    | 62 |     |
| 19       | 1    | 1           | 2    | 29         | 2    | 60 | 232 |

**step 2-4** : 마지막으로 네 번째로 성별과 연관성이 높은 나이 항목과의 일치여부를 판단하여 이전에 획득한 점수와 합산한다.

앞의 경우처럼 나이도 연속형 항목이므로 이 값으로부터 2.5% 또는 5%의 오차 범위 안에 있다면 같은 값으로 간주하기로 한다. 즉,

나이가  $(62) \times (0.95) = 58.9$ 세 보다 많고  $(62) \times (1.05) = 65.1$ 세 보다 적으면 62세와 같은 값으로 인정을 할 것이다. 이로부터 나이의 일치여부를 판단하여 25점을 부여한다. 이 과정이 끝나면 <표 3-11>과 같이 합산한 점수가 가장 높은 개체가 최종적으로 선택된다.

<표 3-11> step 2-4 절차 후의 최종도너(donor)

| 개체<br>번호 | 성별   | 경영주와<br>의관계 | 혼인상태 | 농업종사<br>경력 | 교육정도 | 나이 | 점수  |
|----------|------|-------------|------|------------|------|----|-----|
| 15       | 1(A) | 1(B)        | 2    | 30         | 2    | 62 |     |
| 19       | 1    | 1           | 2    | 29         | 2    | 60 | 257 |

**step 2-5** : **step 4** 로 다시 복귀한다. 본 예에서는 19번 개체의 성별 값이 존재하므로 최종도너로 사용하게 되며 대체하는 성별이 남자(1)로 정확한 대체가 되었음을 알 수 있다.

#### 다. 특별 대체군의 적용

본 연구에서는 각 항목들에 대하여 연관성 분석을 통해서 얻어진 대체군을 이용한다. 제시된 대체군들은 매우 좋은 정보를 가지고 있는 보

조변수임에 틀림없다. 하지만 한 가지 고려해야 할 사항이 있다. 무응답 항목의 대체가 끝나면 대체된 값이 허용되는 값인지 검토할 필요성이 있다. 즉, 내검규칙에 합당한 것인지를 살펴보아야 한다.

주종사분야 항목을 예로써 설명하고자 한다. 주종사분야의 경우 나이가 16세 미만이면 내용을 기입하지 않는다. 그러므로 이 항목을 대체한 후에는 이러한 내용을 꼭 확인해야 한다. 왜냐하면 <표 3-12>와 같은 현상이 일어날 수 있기 때문이다.

<표 3-12> 주종사분야의 대체를 위한 최종도너(donor)

| 주종사분야   | 농업종사기간  | 농업외종사기간 | 경영주와의관계 | 나이 |
|---------|---------|---------|---------|----|
| -1      | -1      | -1      | 3       | 14 |
| missing | missing | missing | 3       | 18 |

주종사분야의 대체군은 농업종사기간, 농업외종사기간, 경영주와의 관계이다. 이때 대체군 중에서 농업종사기간과 농업외종사기간의 정보를 알 수 없다면 경영주와의관계 항목으로만 도너를 찾아야 한다. 그러므로 <표 3-12>와 같은 도너가 최종적으로 선택될 수가 있다. 이 결과로 주종사분야의 무응답 값으로 (-1)이 대체가 되지만 실제 이 사람의 나이는 18세이기 때문에 임의 결측값 (-1)이 아닌 적당한 값이 들어가야 내검규칙에 적합하다고 할 수 있다. 이 경우 새로운 도너를 찾아 다시 대체를 하여야 한다. 하지만 항상 대체한 후에 이러한 내용을 검토하는 것보다는 처음부터 내검규칙에 합당한 값의 도너가 선택되도록 하는 것이 매우 중요하다고 할 수 있다.

따라서 본 연구에서는 특별 대체군을 사용하기로 한다. 특별 대체군은 연관성 분석에서는 선택되지 않았으나 필요에 의하여 임의로 대체군에 편입시키는 보조변수를 지칭할 것이다. 예제의 경우 주종사분야의 특별 대체군으로 나이를 추가한다. 그러면 <표 3-12>에서 선택된 도너는 나이가 같지 않아 제외될 것이며, 나이가 같은 새로운 도너가 선택되어져 내검규칙에 만족하는 값이 대체될 것이다. 이런 이유로 각 항목들에 대하여 특별 대체군이 필요할 경우에는 추가하여 내검규칙에 합당한 대체가 되도록 할 것이다.



그러나 특별 대체군의 사용은 몇 개의 내검규칙을 만족하게 하려는 하나의 수단에 불과하다. 실제 농업총조사 자료에는 개개의 항목 그리고 항목들간에 여러가지 내검규칙들이 존재한다. 따라서 모든 내검규칙에 적합한 값을 한번에 대체하기는 상당히 어려울 것으로 생각된다. 그러므로 내검이 필요한 항목은 대체를 하고난 후 내검을 실시하여 적절하지 못한 값은 또다시 대체가 되어야한다. 그러나 본 연구의 모의실험은 제시한 대체방법의 정확성을 검토하기 위한 것이므로 내검과는 연계하지 않을 것이지만 향후 실제 농업총조사 자료를 대체할 때에는 모든 자료가 내검규칙에 만족할 때까지 내검 ⇒ 대체 ⇒ 내검 ⇒ 대체 과정을 반복해야 할 것이다. 따라서 농업총조사 내검관련 연구도 전문 연구자에 의해 빠른 시일 내에 연구가 진행되었으면 하는 바람이다.

## 2. 각 항목에 대한 모의실험

앞에서 제시한 무응답 대체기법을 적용하여 각 항목들에 대하여 모의실험을 실시한다. 실험에서 사용할 2005년 농업총조사 자료는 가구원 부문 3,448,057명, 가구부문 1,272,908가구가 있으며 이들 자료전체를 사용하여 대체기법의 정확성을 검토한다. 실험을 위해 각 항목에 대하여 임의로 5,000개의 무응답을 발생시킨다. 이 때 목표변수만을 대상으로 무응답을 발생시키는 것이 아니라 모든 항목에 대하여 무응답을 동시에 발생시켜 현실에 맞는 실험을 하고자 한다. 그리고 항목별로 5,000개의 무응답을 모두 대체를 하고 난 후에 목표변수가 범주형의 경우는 각 범주의 분포변화 및 정분류율을 검토하고, 연속형의 경우는 임의로 범주화된 분포변화와 반복을 통한 표본평균의 차이정도를 비교할 것이다. 대체를 위한 도너는 무응답 항목이 있는 개체 주위로부터 50,000개를 비교하여 가장 적절한 것을 선택하도록 한다. 농업총조사 자료의 양이 너무 많아 모든 개체를 고려하는 것은 시간적인 한계로 가능하지 않을 것으로 판단되지만, 실제로 50,000개도 상당히 많은 개체이므로 모든 대체군의 조건을 만족하는 도너는 많이 선택될 수 있을 것이다. 따라서 모든 개체를 고려하여 도너를 찾는 것과 큰 차이가 나지 않을 것이라고 확신한다.

## 가. 가구원에 관한 사항

### 1) 성별

성별에 대한 연관성 분석의 결과 경영주와의관계, 혼인상태, 농업종사경력(100)을 성별의 대체군으로 사용하였으며, 경영주와의관계 정보를 사용할 수 없을 경우 교육정도와 나이가 추가된 2차 대체군을 적용하였다. 앞에서 설명한 알고리즘을 이용하여 총 5,000명을 대체한 결과 대체의 정확도는 81.84%로 높은 적중률을 보였다. <표 3-13>에 정리된 대체결과를 살펴보면 남자는 2,464명 중에서 2,003명, 여자는 2,536명 중에서 2,089명을 정확하게 찾아내었다. 성별 항목의 대체에는 자녀 및 손자녀 등 어린 아이들에 대해서는 뚜렷한 구별의 규칙이 없어 정확도의 손실이 발생하는 것으로 판단된다.

<표 3-13> 성별에 대한 대체군 및 대체결과

| 대체군(1차)                                |             | 대체군(2차)                                     |             |
|--|-------------|---|-------------|
| 경영주와의관계(100)<br>혼인상태(62)<br>농업종사경력(23) |             | 농업종사경력(100)<br>혼인상태(90)<br>교육정도(42), 나이(25) |             |
| 대체결과_대체정확도 (81.84%)                    |             |   |             |
|  | 남(대체)       | 여(대체)                                       | 합계(구성비(%))  |
| 남                                      | 2003        | 461   | 2464(49.28) |
| 여                                      | 447         | 2089  | 2536(50.72) |
| 합계                                     | 2450(49.00) | 2550(51.00)                                 | 5000        |

한 가지 더 고려할 사항은 대체 전후의 항목분포의 변화이다. 대체결과에서 보면 남자는 실제 인원이 2,464명(49.28%)에서 2,450명(49.00%), 여자는 2,536명(50.72%)에서 2,550명(51.00%)으로 대체가 되어 본 연구에서 사용한 대체방법은 대체로 인하여 일어날 수 있는 분포의 왜곡 문제는 거의 없는 것으로 판단이 되며, 대체군의 사용도 매우 적절한 것으로 보인다. 만약 대체의 정확도를 더 높이기 위해서는 성별과 연관성이 높은 새로운 조사항목을 개발하면 가능할 수도 있을 것이다.

## 2) 나이

나이에 대한 무응답 대체는 경영주와의관계, 교육정도, 교육상태, 농업종사경력, 혼인상태를 대체군으로 사용하였으며, 경영주와의관계 정보를 사용할 수 없을 경우 추가되는 항목이 없는 2차 대체군을 적용하였다. 성별과는 다르게 나이는 연속형 항목이므로 반복을 통한 표본평균의 차이정도를 실험을 통하여 살펴보았다. 즉, 임의로 발생시킨 5000명의 표본평균과 대체한 후의 표본평균이 얼마나 일치하는지 10번을 반복하여 비교를 하였다. <표 3-14>의 실험결과를 보면 실제표본평균과 대체한 후의 표본평균의 차이는 0.01세에서 0.09세 정도로 나왔으며, 10번의 평균은 0.045세로 거의 차이가 나지 않음을 알 수 있다.

<표 3-14> 나이에 대한 대체군 및 표본평균 차이정도

| 대체군(1차)  |        | 대체군(2차)                                     |      |
|--|--------|---|------|
| 경영주와의관계(100)<br>교육정도(35), 교육상태(27)<br>농업종사경력(22), 혼인상태(15) |        | 혼인상태(100), 교육정도(39)<br>농업종사경력(17), 교육상태(11) |      |
| 표본평균 절대차이의 평균 (0.045세)                                     |        |   |      |
| 실험   | 실제표본평균 | 대체후 표본평균                                    | 절대차이 |
| 1  | 50.20  | 50.21                                       | 0.01 |
| 2  | 50.31  | 50.40                                       | 0.09 |
| 3  | 49.34  | 49.30                                       | 0.04 |
| 4  | 50.31  | 50.25                                       | 0.06 |
| 5  | 50.10  | 50.01                                       | 0.09 |
| 6  | 50.24  | 50.21                                       | 0.03 |
| 7  | 50.14  | 50.15                                       | 0.01 |
| 8  | 49.99  | 49.94                                       | 0.05 |
| 9  | 50.07  | 50.10                                       | 0.03 |
| 10   | 50.23  | 50.19                                       | 0.04 |

그리고 나이는 연속형 항목이므로 직접적으로 분포변화를 고려하기 보다는 연령대로 범주화를 시켜 확인하는 것이 더 적절할 것으로 판단

된다. <표 3-15>에서처럼 대체 전후의 연령대분포의 변화를 보면 0-9세는 209명(4.18%)에서 205명(4.10%), 10-19세는 455명(9.10%)에서 447명(8.94%), 20-29세는 432명(8.64%)에서 454명(9.08%) 등으로 대체를 한 후에도 연령대에 따른 분포의 변화는 일어나지 않는 것으로 판단된다.

<표 3-15> 나이에 대한 대체전후의 분포변화

| 나이                | 0-9           | 10-19         | 20-29         | 30-39         | 40-49          | 50-59          | 60-69           | 70-79          | 80-89         | 90-          |
|-------------------|---------------|---------------|---------------|---------------|----------------|----------------|-----------------|----------------|---------------|--------------|
| 실제인원<br>(구성비(%))  | 209<br>(4.18) | 455<br>(9.10) | 432<br>(8.64) | 388<br>(7.76) | 622<br>(12.44) | 865<br>(17.30) | 1057<br>(21.14) | 762<br>(15.24) | 184<br>(3.68) | 26<br>(0.52) |
| 대체후인원<br>(구성비(%)) | 205<br>(4.10) | 447<br>(8.94) | 454<br>(9.08) | 375<br>(7.50) | 645<br>(12.90) | 851<br>(17.02) | 1051<br>(21.02) | 743<br>(14.86) | 199<br>(3.98) | 30<br>(0.60) |

### 3) 경영주와의관계

경영주와의관계에 대한 연관성 분석의 결과 나이, 성별, 혼인상태, 농업종사기간을 대체군으로 사용하게 되나 이 경우 농업종사경력을 특별 대체군으로 추가하여 대체를 실시하였다. 그 이유는 경영주만이 농업종사경력을 기입을 하게 되므로 농업종사경력의 값으로 경영주를 알 수 있기 때문이다. 따라서 실제관계가 경영주인 경우는 대부분 정확하게 대체할 수 있을 것으로 판단된다. 또한 가구원번호도 특별 대체군으로 사용하였다. 많은 경우 조사표 상에서 첫 번째는 경영주, 두 번째는 배우자를 기입하는 경향이 있으므로 이 항목은 경영주와의관계를 추정하는 데 중요한 정보임에 틀림없다. 그리고 나이 정보를 사용할 수 없을 경우 2차 대체군을 이용하지만 이후에 실제 대체적용을 할 때에는 나이 항목을 먼저 대체하고자 함으로 1차 대체군을 이용할 수 있을 것으로 판단된다. 본 연구의 알고리즘에 의하여 총 5,000명을 대체한 결과 <표 3-16>에서 보듯이 대체의 정확도는 93.58%로 상당히 높게 나타났다. 또한 특별대체군의 사용으로 경영주는 모두 정확한 대체가 되었으며 배우자, 자녀, 부모의 경우도 매우 좋은 대체가 되었음을 보여준다. 하지만 그 외의 항목은 실제로 구성비가 상당히 작아 연관성 분석에 의한 연관 규칙이 잘 나타나지 못하고 구성비가 큰 항목으로 묻혀버리는 현상이 일어나기 때문에 좋은 대체는 되지 못하는 것을 알 수 있다. 예를 들면

손자녀는 자녀로 대체되는 경우가 많으며, 형제나 친인척의 경우는 특별한 분류규칙이 없기 때문에 주로 나이와 성별을 보고 판단하여 자녀로 대체되는 경우가 많아짐을 알 수 있다. 하지만 전체적으로 볼 때에는 그 수가 상당히 작기 때문에 오분류의 효과는 미미하게 나타난다.

대체 전후의 항목분포의 변화를 살펴보면 경영주는 1,781명 모두 정확한 대체가 되었고, 배우자는 실제 인원이 1,394명(27.88%)에서 1,410명(28.20%), 자녀는 1,254명(25.08%)에서 1,269명(25.38%), 부모는 316명(6.32%)에서 300명(6.00%), 손자녀는 209명(4.18%)에서 195명(3.90%) 등으로 대부분의 항목들이 대체로 인하여 발생할 수 있는 분포의 왜곡 문제는 없는 것으로 판단된다.

<표 3-16> 경영주와의관계에 대한 대체군 및 대체결과

| 대체군(1차)  |                 | 대체군(2차)  |                 |               |               |             |              |              |             |             |
|--|-----------------|--|-----------------|---------------|---------------|-------------|--------------|--------------|-------------|-------------|
| 나이(100), 성별(94)<br>혼인상태(69), 농업종사기간(47)<br>농업종사경력(특별대체군)<br>가구원번호(특별대체군) |                 | 혼인상태(100), 성별(82)<br>농업종사기간(32)<br>농업종사경력(특별대체군)<br>가구원번호(특별대체군) |                 |               |               |             |              |              |             |             |
| 대체결과_대체정확도 (93.58%)  |                 |  |                 |               |               |             |              |              |             |             |
|  | 경영주<br>(대체)     | 배우자<br>(대체)  | 자녀<br>(대체)      | 부모<br>(대체)    | 손자녀<br>(대체)   | 조부모<br>(대체) | 형제<br>(대체)   | 친인척<br>(대체)  | 고용인<br>(대체) | 합계          |
| 경영주  | 1781            | 0  | 0               | 0             | 0             | 0           | 0            | 0            | 0           | 1781(35.62) |
| 배우자  | 0               | 1368   | 20              | 2             | 1             | 0           | 2            | 0            | 1           | 1394(27.88) |
| 자녀   | 0               | 22   | 1124            | 0             | 84            | 0           | 10           | 13           | 1           | 1254(25.08) |
| 부모   | 0               | 18   | 0               | 290           | 0             | 3           | 1            | 4            | 0           | 316(6.32)   |
| 손자녀  | 0               | 0  | 99              | 0             | 109           | 0           | 0            | 1            | 0           | 209(4.18)   |
| 조부모  | 0               | 0  | 0               | 2             | 0             | 2           | 0            | 0            | 0           | 4(0.08)     |
| 형제   | 0               | 2  | 14              | 3             | 0             | 0           | 4            | 0            | 0           | 23(0.46)    |
| 친인척  | 0               | 0  | 11              | 3             | 1             | 0           | 1            | 1            | 0           | 17(0.34)    |
| 고용인  | 0               | 0  | 1               | 0             | 0             | 0           | 1            | 0            | 0           | 2(0.04)     |
| 합계   | 1781<br>(35.62) | 1410<br>(28.20)  | 1269<br>(25.38) | 300<br>(6.00) | 195<br>(3.90) | 5<br>(0.10) | 19<br>(0.38) | 19<br>(0.38) | 2<br>(0.04) | 5000        |

#### 4) 교육정도

교육정도에 대한 연관성 분석의 결과 교육상태, 나이, 혼인상태, 농업종사경력, 성별을 대체군으로 사용하였으며, 교육상태 정보를 사용할 수 없을 경우 농업종사경력이 제외되고 성별이 추가된 2차 대체군을 적용하였다.

<표 3-17> 교육정도에 대한 대체군 및 대체결과

| 대체군(1차)                                   |              | 대체군(2차)                     |             |              |              |              |                |
|---|--------------|-----------------------------|-------------|--------------|--------------|--------------|----------------|
| 교육상태(100), 나이(91)<br>혼인상태(52), 농업종사경력(33) |              | 혼인상태(100)<br>나이(95), 성별(42) |             |              |              |              |                |
| 대체결과_대체정확도 (70.18%)                       |              |                             |             |              |              |              |                |
|   | 안받았음<br>(대체) | 초등학교<br>(대체)                | 중학교<br>(대체) | 고등학교<br>(대체) | 3년대학<br>(대체) | 4년대학<br>(대체) | 합계<br>(구성비(%)) |
| 안받았음                                      | 961          | 0                           | 0           | 0            | 0            | 0            | 961(19.22)     |
| 초등학교                                      | 0            | 1395                        | 198         | 145          | 14           | 14           | 1766(35.32)    |
| 중학교                                       | 0            | 206                         | 304         | 158          | 11           | 14           | 693(13.86)     |
| 고등학교                                      | 0            | 131                         | 163         | 549          | 47           | 101          | 991(19.82)     |
| 3년대학                                      | 0            | 10                          | 17          | 60           | 94           | 49           | 230(4.60)      |
| 4년대학                                      | 0            | 17                          | 26          | 76           | 34           | 206          | 359(7.18)      |
| 합계  | 961(19.22)   | 1759(35.18)                 | 708(14.16)  | 988(19.76)   | 200(4.00)    | 384(7.68)    | 5000           |

앞의 실험에서와 같이 총 5,000명을 대체한 결과 <표 3-17>에서 알 수 있듯이 대체의 정확도는 70.18%로 나타났다. 앞의 항목에 비해서는 다소 정확도가 떨어지나 이 정도의 일치율도 높은 것으로 보아야 할 것이다. 다른 항목에 비하여 정확도가 낮은 궁극적인 이유는 교육정도(특히, 초·중·고)를 세부적으로 분리할 만한 보조변수가 현재의 조사항목에는 부족하기 때문이다. 결과에서 보듯이 중·고등학교, 3년제대학 부분이 대체의 정확도 측면에서 다른 항목에 비해 다소 많이 떨어지고 있다. 따라서 교육정도와 연관성이 높은 항목을 조사표에 추가하지 않는 한 대체의 정확도를 더 높이기에는 한계가 있을 것으로 판단된다. 왜

냐하면 지금의 대체군이 현재로서는 가장 좋은 보조변수임에 틀림없기 때문이다. 만약 공부(과외)시간과 관련된 항목이 있다면 정확도를 높일 수 있는 매우 좋은 정보가 될 수 있을 것이다.

대체 전후의 항목분포의 변화를 보면 교육 ‘안받았음’은 정확한 대체가 되고, 초등학교는 실제 인원이 1,766명(35.32%)에서 1,759명(35.18%), 중학교는 693명(13.86%)에서 708명(14.16%), 고등학교는 991명(19.82%)에서 988명(19.76%) 등으로 대체의 정확도에 비해 대체로 인한 분포의 변화는 크지 않은 것으로 보인다.

### 5) 혼인상태

혼인상태에 대한 연관성 분석의 결과 나이, 경영주와의관계, 교육상태, 농업종사경력을 대체군으로 사용하였으며, 나이 정보를 사용할 수 없을 경우 농업종사기간 항목이 추가된 2차 대체군을 적용하였다. 하지만 이후에 실제 대체적용을 할 때에는 나이 항목이 먼저 대체되므로 1차 대체군을 이용할 수 있을 것으로 판단된다. 총 5,000명을 대체한 결과 대체의 정확도는 84.10%로 매우 높은 정확도를 보였다. <표 3-18>에 정리된 대체결과를 살펴보면 미혼은 732명 중에서 617명, 배우자있음은 3,184명 중에서 2,795명을 정확하게 찾아내었다. 사별과 이혼은 미혼과 배우자있음에 비해 상대적으로 오분류율이 높게 나타나고 있다는 것을 볼 수 있다. 이런 결과는 경영주와의관계의 몇몇 항목처럼 구성비가 작고 연관성 분석에 의한 연관규칙이 잘 나타나지 못하여 불가피하게 일어나는 현상임을 알 수 있다. 허용결측란은 결측값으로 대체되어야 하는 값들을 의미하는데 여기에서는 공백이 아닌 (-1)의 값이 부여되어야 한다. 즉, 혼인상태는 16세 이상인 경우에만 기입을 하게 되므로 나이를 고려하여 16세 미만인 경우에는 (-1)의 값으로 대체가 되어야 하는 것이다. 대체군에 나이가 포함되어 있기 때문에 대체결과 허용결측을 제대로 판별하고 있음을 알 수 있다.

〈표 3-18〉 혼인상태에 대한 대체군 및 대체결과

| 대체군(1차)                                      |            | 대체군(2차)                              |            |            |              |                |
|--|------------|--------------------------------------|------------|------------|--------------|----------------|
| 나이(100), 경영주와의관계(43)<br>교육상태(41), 농업종사경력(24) |            | 경영주와의관계(100)<br>교육상태(40), 농업종사기간(23) |            |            |              |                |
| 대체결과_대체정확도 (84.10%), 허용결측 제외한 정확도(82.31%)    |            |                                      |            |            |              |                |
|  | 미혼<br>(대체) | 배우자있음<br>(대체)                        | 사별<br>(대체) | 이혼<br>(대체) | 허용결측<br>(대체) | 합계<br>(구성비(%)) |
| 미혼   | 617        | 106                                  | 6          | 3          | 0            | 732(14.64)     |
| 배우자있음  | 105        | 2795                                 | 250        | 34         | 0            | 3184(63.68)    |
| 사별   | 4          | 237                                  | 282        | 2          | 0            | 525(10.50)     |
| 이혼   | 11         | 32                                   | 5          | 5          | 0            | 53(1.06)       |
| 허용결측   | 0          | 0                                    | 0          | 0          | 506          | 506(10.12)     |
| 합계   | 737(14.74) | 3170(63.40)                          | 543(10.86) | 44(0.88)   | 506(10.12)   | 5000           |

대체 전후의 항목분포의 변화를 보면 미혼은 실제 인원이 732명(14.64%)에서 737명(14.74%), 배우자있음은 3,184명(63.68%)에서 3,170명(63.40%), 사별은 525명(10.50%)에서 543명(10.86%), 이혼은 53명(1.06%)에서 44명(0.88%)으로 변화는 거의 발생하지 않음을 볼 수 있다.

## 6) 주종사분야

주종사분야에 대한 연관성 분석의 결과 농업종사기간, 농업외종사기간, 경영주와의관계를 대체군으로 사용하게 되나 이 경우 나이를 특별 대체군으로 추가하여 대체를 실시하였다. 그 이유는 혼인상태에서처럼 주종사분야는 나이가 16세 이상인 경우에만 기입을 하게 되므로 나이에 맞는 대체가 이루어져야 하기 때문이다. 따라서 나이가 16세 미만인 경우는 허용결측으로, 16세 이상이면 적절한 대체값으로 대체가 되어야 한다. 그러므로 나이를 대체군에 추가함으로써 대체 후 내검규칙에 합당한 대체가 될 것이라 판단된다. 농업종사기간 정보를 사용할 수 없을 경우에는 추가항목이 없는 2차 대체군을 적용하였다. <표 3-19>에서와 같이 총 5,000명을 대체한 결과 대체의 정확도는 90.06%로 상당히 높은



정확도를 보였다. 주종사분야가 농업, 종사안함, 허용결측은 대부분 정확한 대체가 되었고, 그 외의 항목들은 많은 경우 기타산업으로 대체가 되었음을 알 수 있다. 이는 실제 구성비가 상당히 작아 연관성 분석에 의한 연관규칙이 잘 나타나지 못하여 기타산업 항목으로 묻혀버리는 현상이 일어나기 때문인 것으로 보인다. 그러나 전체적으로 그 수가 상당히 작기 때문에 오분류의 효과는 상당히 감소됨을 알 수 있다. 또한 농업과 종사안함이 많은 부분을 차지하지만 다른 항목의 경우와 마찬가지로 대체로 인한 항목분포의 변화는 발생하지 않아 적절한 대체가 되었다고 판단된다.

〈표 3-19〉 주종사분야에 대한 대체군 및 대체결과

| 대체군(1차)   |            | 대체군(2차)                                  |            |             |             |                  |                  |                  |                  |                  |      |
|---|------------|--|------------|-------------|-------------|------------------|------------------|------------------|------------------|------------------|------|
| 농업종사기간(100)<br>농업외종사기간(67)<br>경영주와의관계(9)<br>나이(특별대체군) |            | 농업외종사기간(100)<br>경영주와의관계(82)<br>나이(특별대체군) |            |             |             |                  |                  |                  |                  |                  |      |
| 대체결과_대체정확도 (90.06%), 허용결측 제외한 정확도(89.05%)             |            |  |            |             |             |                  |                  |                  |                  |                  |      |
|   | 농업<br>(대체) | 임업<br>(대체)                               | 어업<br>(대체) | 제조업<br>(대체) | 건설업<br>(대체) | 도소<br>매업<br>(대체) | 숙박<br>음식<br>(대체) | 기타<br>산업<br>(대체) | 종사<br>안함<br>(대체) | 허용<br>결측<br>(대체) | 합계   |
| 농업  | 2957       | 3  | 9          | 8           | 12          | 9                | 6                | 47               | 10               | 0                | 3061 |
| 임업  | 6          | 7  | 1          | 2           | 1           | 1                | 0                | 2                | 0                | 0                | 20   |
| 어업  | 10         | 1  | 16         | 3           | 1           | 3                | 2                | 6                | 0                | 0                | 42   |
| 제조업   | 6          | 4  | 1          | 19          | 2           | 5                | 4                | 45               | 0                | 0                | 86   |
| 건설업   | 3          | 0  | 1          | 5           | 10          | 0                | 1                | 27               | 0                | 0                | 47   |
| 도소매업  | 7          | 1  | 2          | 2           | 2           | 13               | 4                | 25               | 0                | 0                | 56   |
| 숙박음식  | 3          | 0  | 2          | 4           | 2           | 4                | 10               | 15               | 0                | 0                | 40   |
| 기타산업  | 47         | 2  | 3          | 51          | 16          | 19               | 24               | 282              | 1                | 0                | 445  |
| 종사안함  | 12         | 0  | 0          | 1           | 0           | 0                | 0                | 1                | 729              | 0                | 743  |
| 허용결측  | 0          | 0  | 0          | 0           | 0           | 0                | 0                | 0                | 0                | 460              | 460  |
| 합계  | 3051       | 18                                       | 35         | 95          | 46          | 54               | 51               | 450              | 740              | 460              | 5000 |

### 7) 농업종사기간

농업종사기간에 대한 연관성 분석의 결과 주종사분야, 경영주와의관계를 대체군으로 사용하게 되나 이 경우도 나이를 특별 대체군으로 추가하여 대체를 실시하였다. 그 이유는 이전의 주종사분야 항목처럼 나이가 16세 이상인 경우에만 기입을 하는 항목이기 때문이다. 그리고 주종사분야 정보를 사용할 수 없을 경우 2차 대체군을 이용하지만 이후에 실제 대체적용을 할 때에는 주종사분야 항목을 먼저 대체하므로 1차 대체군을 이용할 수 있을 것으로 보인다.

<표 3-20> 농업종사기간에 대한 대체군 및 대체결과

| 대체군(1차)                                    |             | 대체군(2차)                   |               |               |               |              |                |
|--|-------------|---------------------------|---------------|---------------|---------------|--------------|----------------|
| 주종사분야(100)<br>경영주와의관계(62)<br>나이(특별대체군)     |             | 경영주와의관계(100)<br>나이(특별대체군) |               |               |               |              |                |
| 대체결과_ 대체정확도 (84.76%), 허용결측 제외한 정확도(83.01%) |             |                           |               |               |               |              |                |
|  | 없음<br>(대체)  | 1개월미만<br>(대체)             | 1-3개월<br>(대체) | 3-6개월<br>(대체) | 6개월이상<br>(대체) | 허용결측<br>(대체) | 합계<br>(구성비(%)) |
| 없음   | 884         | 46                        | 49            | 20            | 10            | 0            | 1009(20.18)    |
| 1개월미만                                      | 51          | 20                        | 7             | 4             | 3             | 0            | 85(1.70)       |
| 1-3개월                                      | 38          | 7                         | 65            | 48            | 37            | 0            | 195(3.90)      |
| 3-6개월                                      | 20          | 3                         | 43            | 157           | 150           | 0            | 373(7.46)      |
| 6개월이상                                      | 17          | 1                         | 47            | 161           | 2598          | 0            | 2824(56.48)    |
| 허용결측                                       | 0           | 0                         | 0             | 0             | 0             | 514          | 514(10.28)     |
| 합계   | 1010(20.20) | 77(1.54)                  | 211(4.22)     | 390(7.80)     | 2798(55.96)   | 514(10.28)   | 5000           |

<표 3-20>에서와 같이 총 5,000명을 대체한 결과 대체의 정확도는 84.76%로 높게 나타나고 있다. 하지만 농업종사기간이 없음, 6개월이상, 허용결측은 정확도가 상당히 높게 나타났으나, 그 외의 항목들은 오분류율이 높아짐을 보였다. 이러한 현상은 농업종사기간이 없음과 6개월 이상에 비해서 다른 항목의 구성비가 작아 일어나는 불가피한 상황이라

고 말할 수 있으며, 1개월미만은 없음으로, 3-6개월은 6개월이상으로 많은 부분 대체가 되고 있음을 보여준다. 또 하나의 중요한 이유는 연속형 항목을 범주화시킴으로써 대체를 할 때 서로 인접하는 기간에서는 당연히 오분류가 크게 일어날 수밖에 없기 때문이다. 즉, 1-3개월과 3-6개월에서는 3개월 근처에서, 3-6개월과 6개월 이상에서는 6개월 근처에서 서로 어긋나게 대체가 되는 것들이 증가하고 있음을 보여준다. 따라서 <표 3-20>과 같은 정확도의 결과는 큰 의미가 없을 수도 있다. 그리고 농업종사기간과 연관성이 높은 대체군 항목이 적은 것도 하나의 이유가 될 수 있을 것이다.

대체 전후의 항목분포의 변화를 보면 없음은 실제 인원이 1,009명(20.18%)에서 1,010명(20.20%), 1개월미만은 85명(1.70%)에서 77명(1.54%), 1-3개월은 195명(3.90%)에서 211명(4.22%), 3-6개월은 373명(7.46%)에서 390명(7.80%), 6개월이상은 2,824명(56.48%)에서 2,798명(55.96%)으로 대체가 되어 분포변화로 인한 문제는 고려하지 않아도 될 것으로 본다.

## 8) 농업외종사기간

농업외종사기간에 대한 연관성 분석의 결과 주종사분야 하나의 항목만을 대체군으로 사용하지만 농업종사기간과 마찬가지로 나이를 특별 대체군으로 추가하여 대체를 실시하였다. 그리고 주종사분야 정보를 사용할 수 없을 경우 2차 대체군을 이용하지만 실제 대체적용을 할 때에는 주종사분야 항목을 먼저 대체하고자 하므로 1차 대체군을 이용할 수 있을 것으로 판단된다. <표 3-21>의 대체결과를 보면 총 5,000명에 대한 대체의 정확도는 83.38%로 높게 나타나고 있다. 또한 허용결측을 제외하고도 81.75%로 대체의 정확도는 높다. 하지만 농업외종사기간이 없음, 6개월이상, 허용결측은 다소 정확도가 높게 나타났으나, 그 외의 항목들은 오분류율이 높아짐을 보였다. 이러한 현상은 농업종사기간의 경우와 같다고 볼 수 있다. 농업외종사기간의 '없음' 항목의 구성비는 대략 전체의 70%를 차지하고 있다. 그러므로 다른 항목들은 구성비가 큰 항목으로 묻혀버리는 현상이 일어나기 때문에 이 부분에서 좋은 대체가

일어나지 않는다. 대체결과에서 보듯이 실제 1개월미만, 1-3개월, 3-6개월, 6개월이상인데도 불구하고 농업외종사기간이 없음이라고 대체된 경우가 많다는 것을 알 수 있다. 또한 연속형 항목을 범주화시킴으로써 일어나는 인접효과도 대체의 정확도를 떨어지게 하며, 무엇보다도 농업외종사기간과 연관성이 높은 보조변수가 부족하다는 것이 가장 큰 이유라고 볼 수 있을 것이다. 따라서 ‘없음’ 이외의 내용들을 세부적으로 분리할 만한 항목을 조사표에 추가하지 않는 한 현재의 결과보다 대체의 정확도를 더 높이기에는 한계가 있을 것으로 판단된다.

<표 3-21> 농업외종사기간에 대한 대체군 및 대체결과

| 대체군(1차)                                   |             | 대체군(2차)                           |               |               |               |              |              |
|---|-------------|-----------------------------------|---------------|---------------|---------------|--------------|--------------|
| 주종사분야(100)<br>나이(특별대체군)                   |             | 교육상태(100)<br>농업종사기간(94)<br>나이(55) |               |               |               |              |              |
| 대체결과_대체정확도 (83.38%), 허용결측 제외한 정확도(81.74%) |             |                                   |               |               |               |              |              |
|   | 없음<br>(대체)  | 1개월미만<br>(대체)                     | 1-3개월<br>(대체) | 3-6개월<br>(대체) | 6개월이상<br>(대체) | 허용결측<br>(대체) | 합계<br>(구성비%) |
| 없음  | 3060        | 37                                | 94            | 91            | 121           | 0            | 3403(68.06)  |
| 1개월미만                                     | 29          | 16                                | 3             | 2             | 0             | 0            | 50(1.00)     |
| 1-3개월                                     | 95          | 3                                 | 31            | 14            | 9             | 0            | 152(3.04)    |
| 3-6개월                                     | 98          | 1                                 | 6             | 41            | 34            | 0            | 180(3.60)    |
| 6개월이상                                     | 135         | 4                                 | 16            | 39            | 572           | 0            | 766(15.32)   |
| 허용결측                                      | 0           | 0                                 | 0             | 0             | 0             | 449          | 449(8.98)    |
| 합계  | 3417(68.34) | 61(1.22)                          | 150(3.00)     | 187(3.74)     | 736(14.78)    | 449(8.98)    | 5000         |

대체 전후의 항목분포의 변화를 보면 ‘없음’은 실제 인원이 3,403명 (68.06%)에서 3,417명(68.34%), 1개월미만은 50명(1%)에서 61명(1.22%), 1-3개월은 152명(3.04)에서 150명(3.00%), 3-6개월은 180명(3.60%)에서 187명(3.74%), 6개월이상은 766명(15.32%)에서 736명(14.78%)으로 대체가 되어 분포변화로 인하여 발생하는 문제는 거의 없는 것으로 보인다.

### 9) 농업종사경력

농업종사경력에 대한 연관성 분석의 결과 나이, 교육정도를 대체군으로 사용하게 되나 이 경우 경영주와의관계를 특별 대체군으로 추가하여 대체를 실시하였다. 그 이유는 경영주만이 농업종사경력을 기입을 하므로 경영주에 대해서만 대체를 하고 그 외에는 허용결측으로 두어야 하기 때문에 경영주와의관계 정보가 필요하다. 그리고 나이 정보를 사용할 수 없을 경우 2차 대체군을 이용하지만 이후에 실제 대체적용을 할 때에는 나이 항목을 먼저 대체하므로 1차 대체군을 이용할 수 있을 것으로 보인다.

〈표 3-22〉 농업종사경력에 대한 대체군 및 표본평균 차이정도

| 대체군(1차)                               |        | 대체군(2차)  |             |
|---------------------------------------|--------|--|-------------|
| 나이(100)<br>교육정도(31)<br>경영주와의관계(특별대체군) |        | 교육정도(100)<br>교육상태(27)<br>농업외종사기간(24)<br>경영주와의관계(특별대체군) |             |
| 표본평균 절대차이의 평균 (0.17년)                 |        |  |             |
| 실험                                    | 실제표본평균 | 대체후 표본평균   | 절대차이        |
| 1                                     | 33.67  | 33.82  | <b>0.15</b> |
| 2                                     | 33.57  | 33.28  | <b>0.29</b> |
| 3                                     | 33.71  | 33.82  | <b>0.11</b> |
| 4                                     | 33.85  | 33.99  | <b>0.14</b> |
| 5                                     | 33.67  | 33.89  | <b>0.22</b> |
| 6                                     | 33.63  | 33.85  | <b>0.22</b> |
| 7                                     | 33.84  | 33.95  | <b>0.11</b> |
| 8                                     | 33.75  | 33.92  | <b>0.17</b> |
| 9                                     | 33.58  | 33.76  | <b>0.18</b> |
| 10                                    | 33.85  | 33.73  | <b>0.12</b> |

농업종사경력은 나이처럼 연속형 항목이므로 임의로 발생시킨 5000명의 표본평균과 대체한 후의 표본평균이 얼마나 일치하는지 10번을 반복하여 비교를 하였다. <표 3-22>의 실험결과를 보면 실제표본평균과 대체한 후의 표본평균의 차이는 0.11년에서 0.29년 정도로 나왔으며, 10번의 평균은 0.17년으로 거의 차이가 나지 않음을 알 수 있다. 그리고 농업종사경력에 대한 분포변화를 고려하기 위하여 경력기간을 범주화하여 확인하였다. <표 3-23>에서처럼 대체 전후의 농업종사경력의 변화를 보면 0-9년은 482명(9.64%)에서 463명(9.26%), 10-19년은 472명(9.44%)에서 497명(9.94%), 20-29년은 702명(14.04%)에서 715명(14.30%), 30-39년은 955명(19.10%)에서 936명(18.72%) 등으로 나타나 대체를 한 후에도 농업종사경력에 따른 분포의 변화는 일어나지 않는 것으로 판단된다.

<표 3-23> 농업종사경력에 대한 대체전후의 분포변화

| 농업종사경력            | 0-9           | 10-19         | 20-29          | 30-39          | 40-49           | 50-59           | 60-69         | 70-         |
|-------------------|---------------|---------------|----------------|----------------|-----------------|-----------------|---------------|-------------|
| 실제인원<br>(구성비(%))  | 482<br>(9.64) | 472<br>(9.44) | 702<br>(14.04) | 955<br>(19.10) | 1173<br>(23.46) | 1014<br>(20.28) | 195<br>(3.90) | 7<br>(0.14) |
| 대체후인원<br>(구성비(%)) | 463<br>(9.26) | 497<br>(9.94) | 715<br>(14.30) | 936<br>(18.72) | 1156<br>(23.12) | 1048<br>(20.96) | 180<br>(3.60) | 5<br>(0.10) |

## 나. 경지에 관한 사항

### 1) 논면적

논면적에 대한 무응답 대체는 일모작논, 경지정리논, 논갈이방법을 대체군으로 사용하였으며, 일모작논의 정보를 사용할 수 없을 경우 2차 대체군을 적용하였다. 논면적은 경지에 관한 항목들 중 가장 중요한 항목으로 제일 먼저 대체가 이루어져야 할 것으로 판단된다. 그리고 논면적은 자기논과 남의논, 일모작논과 이모작논, 경지정리논과 경지미정리논의 합계이기 때문에 이들의 값을 알 수 있다면 연역적 대체방법을 사용하여 대체를 하면 되므로 본 연구에서는 연역적 대체를 할 수 없을 경우의 대체를 응용 핫덱 알고리즘을 통하여 실시하였다. 경지에 관한

사항은 모두 연속형 항목이므로 반복을 통한 표본평균의 차이정도와 면적을 범주화한 후 대체전후 분포를 실험을 통하여 살펴보았다. 임의로 발생시킨 5000가구의 표본평균과 대체한 후의 표본평균의 값은 <표 3-24>에 주어져 있다. 10번 반복된 실험결과를 보면 실제표본평균과 대체한 후의 표본평균의 차이는 8.15평에서 31.19평 정도로 나왔으며, 10번의 평균은 20.55평으로 적절한 대체가 되었음을 알 수 있다.

<표 3-24> 논면적에 대한 대체군 및 표본평균 차이정도

| 대체군(1차)                             |         | 대체군(2차)  |       |
|-------------------------------------|---------|--|-------|
| 일모작논(100)<br>경지정리논(21)<br>논갈이방법(15) |         | 트랙터(100), 영농형태(72)<br>논갈이(57), 판매금액구분(55)<br>모내기방법(26) |       |
| 표본평균 절대차이의 평균 (20.55평)              |         |  |       |
| 실험                                  | 실제표본평균  | 대체후 표본평균   | 절대차이  |
| 1                                   | 2187.08 | 2156.32  | 30.76 |
| 2                                   | 2263.72 | 2249.35  | 14.37 |
| 3                                   | 2159.75 | 2128.56  | 31.19 |
| 4                                   | 2235.39 | 2206.20  | 29.19 |
| 5                                   | 2238.99 | 2217.39  | 21.60 |
| 6                                   | 2265.91 | 2243.73  | 22.18 |
| 7                                   | 2268.34 | 2254.13  | 14.21 |
| 8                                   | 2194.45 | 2186.30  | 8.15  |
| 9                                   | 2286.35 | 2265.13  | 21.22 |
| 10                                  | 2259.29 | 2246.62  | 12.67 |

논면적에 대한 분포변화를 고려하기 위하여 논면적을 범주화하여 확인한 결과가 <표 3-25>에 주어져 있다. 0-499평은 1515가구(30.30%)에서 1525가구(30.50%), 500-999평은 664가구(13.28%)에서 651가구(13.02%), 1000-1499평은 637가구(12.74%)에서 642가구(12.84%), 4500평 이상은 660가구(13.20%)에서 659가구(13.18%) 등으로 나타나 대체로 인한 분포의 변화는 일어나지 않음을 알 수 있다.

〈표 3-25〉 논면적에 대한 대체전후의 분포변화

| 논면적               | 0-499           | 500-999        | 1000-1499      | 1500-1999     | 2000-2499     | 2500-2999     | 3000-3499     | 3500-3999     | 4000-4499     | 4500-          |
|-------------------|-----------------|----------------|----------------|---------------|---------------|---------------|---------------|---------------|---------------|----------------|
| 실제가구<br>(구성비(%))  | 1515<br>(30.30) | 664<br>(13.28) | 637<br>(12.74) | 393<br>(7.86) | 394<br>(7.88) | 189<br>(3.78) | 271<br>(5.42) | 107<br>(2.14) | 170<br>(3.40) | 660<br>(13.20) |
| 대체후가구<br>(구성비(%)) | 1525<br>(30.50) | 651<br>(13.02) | 642<br>(12.84) | 406<br>(8.12) | 390<br>(7.80) | 184<br>(3.68) | 265<br>(5.30) | 118<br>(2.36) | 160<br>(3.20) | 659<br>(13.18) |

## 2) 자기논

자기논에 대한 무응답 대체는 논면적, 경지정리논을 대체군으로 사용하였으며, 이후에 실제 대체적용을 할 때에는 논면적 항목을 먼저 대체하고자 하므로 1차 대체군을 이용할 수 있을 것으로 판단된다. 그리고 자기논은 논면적과 남의논 면적을 알고 있으면 연역적 대체가 가능하므로 남의논 면적을 알지 못할 경우로 한정하여 모의실험을 실시하였다. 임의로 발생시킨 5000가구의 표본평균과 대체한 후의 표본평균의 값은 <표 3-26>에 주어져 있다. 10번 반복된 실험결과를 보면 실제표본평균과 대체한 후의 표본평균의 차이는 9.40평에서 41.30평 정도로 나왔으며, 10번의 평균은 34.07평으로 대체 후에도 자기논의 면적은 큰 차이가 나지 않음을 알 수 있다.

자기논 면적에 대한 분포변화를 고려하기 위하여 자기논 면적을 범주화하여 확인한 결과가 <표 3-27>에 주어져 있다. 0-299평은 1829가구(36.58%)에서 1845가구(36.90%), 300-599평은 328가구(6.56%)에서 318가구(6.36%), 600-899평은 475가구(9.50%)에서 457가구(9.14%), 900-1199평은 431가구(8.62%)에서 453가구(9.06%), 2700평 이상은 778가구(15.56%)에서 746가구(14.92%) 등으로 변화되나 분포상의 문제는 거의 일어나지 않는다고 판단된다.



〈표 3-26〉 자기논에 대한 대체군 및 표본평균 차이정도

| 대체군(1차)                |         | 대체군(2차)                                     |       |
|------------------------|---------|---|-------|
| 논면적(100)<br>경지정리논(48)  |         | 이양기(100), 판매금액구분(75)<br>논갈이방법(68), 영농형태(60) |       |
| 표본평균 절대차이의 평균 (34.07평) |         |   |       |
| 실험                     | 실제표본평균  | 대체후 표본평균                                    | 절대차이  |
| 1                      | 1399.31 | 1380.82                                     | 18.49 |
| 2                      | 1356.08 | 1397.38                                     | 41.30 |
| 3                      | 1411.19 | 1384.03                                     | 27.16 |
| 4                      | 1345.54 | 1371.04                                     | 25.50 |
| 5                      | 1358.92 | 1374.74                                     | 15.82 |
| 6                      | 1347.83 | 1357.23                                     | 9.40  |
| 7                      | 1347.54 | 1332.29                                     | 15.25 |
| 8                      | 1325.93 | 1350.33                                     | 24.40 |
| 9                      | 1335.88 | 1357.41                                     | 21.53 |
| 10                     | 1397.90 | 1427.28                                     | 29.38 |

〈표 3-27〉 자기논에 대한 대체전후의 분포변화

| 자기논               | 0-299           | 300-599       | 600-899       | 900-1199      | 1200-1499     | 1500-1799     | 1800-2099     | 2100-2399    | 2400-2699     | 2700-          |
|-------------------|-----------------|---------------|---------------|---------------|---------------|---------------|---------------|--------------|---------------|----------------|
| 실제가구<br>(구성비(%))  | 1829<br>(36.58) | 328<br>(6.56) | 475<br>(9.50) | 431<br>(8.62) | 357<br>(7.14) | 246<br>(4.92) | 306<br>(6.12) | 86<br>(1.72) | 164<br>(3.28) | 778<br>(15.56) |
| 대체후가구<br>(구성비(%)) | 1845<br>(36.90) | 318<br>(6.36) | 457<br>(9.14) | 453<br>(9.06) | 373<br>(7.46) | 260<br>(5.20) | 308<br>(6.16) | 93<br>(1.86) | 147<br>(2.94) | 746<br>(14.92) |

### 3) 남의논

남의논의 경우는 앞에서 논면적과 자기논의 면적을 알 수 있으므로 연역적 대체방법(남의논=논면적-자기논)으로 바로 대체가 가능하다.

#### 4) 일모작논

일모작논에 대한 무응답 대체는 논면적 하나의 항목으로 대체를 실시해야하나 논면적은 사전에 대체가 되므로 문제가 없을 것으로 판단된다. 일모작논은 논면적과 이모작논 면적을 알고 있으면 연역적 대체가 가능하므로 이모작논 면적을 알지 못할 경우로 한정하여 모의실험을 실시하였다. <표 3-28>의 결과를 살펴보면 실제표본평균과 대체한 후의 표본평균의 차이는 11.65평에서 36.37평 정도로 나왔으며, 10번의 평균은 19.73평으로 적절한 대체가 되었다고 생각된다.

<표 3-28> 일모작논에 대한 대체군 및 표본평균 차이정도

| 대체군(1차)                |         | 대체군(2차)                                      |       |
|------------------------|---------|--|-------|
| 논면적(100)               |         | 이양기(100), 콤바인(72)<br>판매금액구분(67)<br>논갈이방법(46) |       |
| 표본평균 절대차이의 평균 (19.73평) |         |  |       |
| 실험                     | 실제표본평균  | 대체후 표본평균                                     | 절대차이  |
| 1                      | 2072.41 | 2087.04                                      | 14.63 |
| 2                      | 2055.23 | 2091.60                                      | 36.37 |
| 3                      | 2016.13 | 2039.84                                      | 23.71 |
| 4                      | 2035.46 | 2053.11                                      | 17.65 |
| 5                      | 2017.31 | 2032.39                                      | 15.08 |
| 6                      | 2045.92 | 2073.54                                      | 27.62 |
| 7                      | 2067.19 | 2081.95                                      | 14.76 |
| 8                      | 2031.53 | 2046.43                                      | 14.90 |
| 9                      | 2058.42 | 2079.33                                      | 20.91 |
| 10                     | 2024.08 | 2035.73                                      | 11.65 |

일모작논 면적에 대한 대체전후의 분포는 <표 3-29>에 주어져 있다. 0-499평은 1744가구(34.88%)에서 1727가구(34.54%), 500-999평은 602가구(12.04%)에서 606가구(12.12%), 1000-1499평은 610가구(12.20%)에서

635가구(12.70%), 4500평 이상은 599가구(11.98%)에서 602가구(12.04%) 등으로 나타나 대체로 인한 분포의 변화는 일어나지 않음을 알 수 있다.

〈표 3-29〉 일모작논에 대한 대체전후의 분포변화

| 일모작논              | 0-499           | 500-999        | 1000-1499      | 1500-1999     | 2000-2499     | 2500-2999     | 3000-3499     | 3500-3999     | 4000-4499     | 4500-          |
|-------------------|-----------------|----------------|----------------|---------------|---------------|---------------|---------------|---------------|---------------|----------------|
| 실제가구<br>(구성비(%))  | 1744<br>(34.88) | 602<br>(12.04) | 610<br>(12.20) | 352<br>(7.04) | 397<br>(7.94) | 172<br>(3.44) | 263<br>(5.26) | 111<br>(2.22) | 150<br>(3.00) | 599<br>(11.98) |
| 대체후가구<br>(구성비(%)) | 1727<br>(34.54) | 606<br>(12.12) | 635<br>(12.70) | 362<br>(7.24) | 385<br>(7.70) | 166<br>(3.32) | 265<br>(5.30) | 109<br>(2.18) | 143<br>(2.86) | 602<br>(12.04) |

### 5) 이모작논

이모작논의 경우는 앞에서 논면적과 일모작논의 면적을 알 수 있으므로 연역적 대체방법(이모작논=논면적-일모작논)으로 바로 대체가 가능하다.

### 6) 경지정리논

경지정리논에 대한 무응답 대체는 일모작논과 같이 논면적을 이용하여 대체를 실시하면 될 것으로 판단된다. 이전에도 설명을 하였지만 경지정리논이나 일모작논은 논면적과 상당히 큰 연관성을 가지고 있기 때문에 좋은 대체가 될 수 있다. 경지정리논은 논면적과 경지미정리논 면적을 알고 있으면 연역적 대체가 가능하므로 경지미정리논 면적을 알지 못할 경우로 한정하여 모의실험을 실시하였다. <표 3-30>의 결과를 살펴보면 실제표본평균과 대체한 후의 표본평균의 차이는 15.46평에서 33.54평 정도로 나왔으며, 10번의 평균은 24.31평으로 적절한 대체가 되었다고 생각한다. 경지정리논에 대한 대체전후의 분포는 <표 3-31>에 주어져있다. 0-299평은 2553가구(51.06%)에서 2544가구(50.88%), 600-899평은 313가구(6.26%)에서 320가구(6.40%) 등으로 변화되나 분포상의 문제는 거의 일어나지 않는다고 판단된다. 0-299평의 비율이 상당히 높은 이유는 경지정리논이 0평인 가구가 절반 정도이기 때문이다.

〈표 3-30〉 경지정리논에 대한 대체군 및 표본평균 차이정도

| 대체군(1차)                |         | 대체군(2차)   |       |
|------------------------|---------|---|-------|
| 논면적(100)               |         | 이양기(100)<br>판매금액구분(95)<br>영농형태(84)<br>논갈이방법(46) |       |
| 표본평균 절대차이의 평균 (24.31평) |         |   |       |
| 실험                     | 실제표본평균  | 대체후 표본평균  | 절대차이  |
| 1                      | 1503.92 | 1526.27   | 22.35 |
| 2                      | 1532.22 | 1564.61   | 32.39 |
| 3                      | 1504.66 | 1538.20   | 33.54 |
| 4                      | 1568.34 | 1542.38   | 25.96 |
| 5                      | 1539.95 | 1524.49   | 15.46 |
| 6                      | 1524.98 | 1542.95   | 17.97 |
| 7                      | 1558.38 | 1532.95   | 25.43 |
| 8                      | 1519.24 | 1538.57   | 19.33 |
| 9                      | 1562.88 | 1535.69   | 27.19 |
| 10                     | 1547.97 | 1524.48   | 23.49 |

〈표 3-31〉 경지정리논에 대한 대체전후의 분포변화

| 경지정리논             | 0-299           | 300-599       | 600-899       | 900-1199      | 1200-1499     | 1500-1799     | 1800-2099     | 2100-2399    | 2400-2699     | 2700-          |
|-------------------|-----------------|---------------|---------------|---------------|---------------|---------------|---------------|--------------|---------------|----------------|
| 실제가구<br>(구성비(%))  | 2553<br>(51.06) | 170<br>(3.40) | 313<br>(6.26) | 253<br>(5.06) | 240<br>(4.80) | 155<br>(3.10) | 246<br>(4.92) | 85<br>(1.70) | 128<br>(2.56) | 857<br>(17.14) |
| 대체후가구<br>(구성비(%)) | 2544<br>(50.88) | 185<br>(3.70) | 320<br>(6.40) | 240<br>(4.80) | 251<br>(5.02) | 150<br>(3.00) | 241<br>(4.82) | 90<br>(1.80) | 135<br>(2.70) | 844<br>(16.88) |

### 7) 경지미정리논

경지미정리논의 경우도 (경지미정리논=논면적-경지정리논)의 관계를 이용한 연역적 대체를 실시하면 될 것이다.

## 8) 발면적

발면적에 대한 무응답 대체는 자기밭, 영농형태, 판매금액구분 항목을 대체군으로 사용하였으며, 자기밭의 정보를 사용할 수 없을 경우 2차 대체군을 적용하였다. 그리고 발면적은 자기밭과 남의밭의 합계이기 때문에 이들의 값을 알 수 있다면 연역적 대체방법을 사용하여 대체를 하면 되므로 본 연구에서는 남의밭 면적을 알지 못할 경우로 한정하여 모의실험을 실시하였다. <표 3-32>의 결과를 살펴보면 실제표본평균과 대체한 후의 표본평균의 차이는 8.31평에서 26.83평 정도로 나왔으며, 10번의 평균은 17.62평인 것으로 나타났다.

<표 3-32> 발면적에 대한 대체군 및 표본평균 차이정도

| 대체군(1차)                            |         | 대체군(2차)                            |       |
|------------------------------------|---------|------------------------------------|-------|
| 자기밭(100)<br>영농형태(35)<br>판매금액구분(26) |         | 판매금액구분(100)<br>영농형태(73)<br>관리기(14) |       |
| 표본평균 절대차이의 평균 (17.62평)             |         |                                    |       |
| 실험                                 | 실제표본평균  | 대체후 표본평균                           | 절대차이  |
| 1                                  | 1315.96 | 1289.13                            | 26.83 |
| 2                                  | 1346.38 | 1332.21                            | 14.17 |
| 3                                  | 1315.21 | 1335.69                            | 20.69 |
| 4                                  | 1321.58 | 1312.44                            | 9.14  |
| 5                                  | 1320.72 | 1339.35                            | 18.63 |
| 6                                  | 1317.43 | 1325.74                            | 8.31  |
| 7                                  | 1349.69 | 1331.22                            | 18.47 |
| 8                                  | 1323.11 | 1347.53                            | 24.42 |
| 9                                  | 1341.83 | 1327.01                            | 14.82 |
| 10                                 | 1353.64 | 1332.94                            | 20.70 |

발면적에 대한 대체전후의 분포는 <표 3-33>에 주어져있다. 0-299평은 1356가구(27.12%)에서 1373가구(27.46%), 300-599평은 1034가구

(20.68%)에서 1020가구(20.40%), 600-899평은 582가구(11.64%)에서 561가구(11.22%), 2700평 이상은 665가구(13.30%)에서 650가구(13.00%) 등으로 변화되나 분포상의 문제는 거의 일어나지 않는다고 판단된다.

<표 3-33> 발면적에 대한 대체전후의 분포변화

| 발면적               | 0-299           | 300-599         | 600-899        | 900-1199      | 1200-1499     | 1500-1799     | 1800-2099     | 2100-2399    | 2400-2699    | 2700-          |
|-------------------|-----------------|-----------------|----------------|---------------|---------------|---------------|---------------|--------------|--------------|----------------|
| 실제가구<br>(구성비(%))  | 1356<br>(27.12) | 1034<br>(20.68) | 582<br>(11.64) | 470<br>(9.40) | 223<br>(4.46) | 253<br>(5.06) | 273<br>(5.46) | 52<br>(1.04) | 92<br>(1.84) | 665<br>(13.30) |
| 대체후가구<br>(구성비(%)) | 1373<br>(27.46) | 1020<br>(20.40) | 561<br>(11.22) | 453<br>(9.06) | 242<br>(4.84) | 279<br>(5.58) | 269<br>(5.38) | 67<br>(1.34) | 86<br>(1.72) | 650<br>(13.00) |

## 9) 자기발

자기발에 대한 무응답 대체는 발면적, 영농형태, 남의논을 대체군으로 사용하였으며, 이후에 실제 대체적용을 할 때에는 발면적 항목을 먼저 대체하고자 함으로 1차 대체군을 이용할 수 있을 것으로 판단된다. 그리고 자기발은 발면적과 남의발 면적을 알고 있으면 연역적 대체가 가능하므로 남의발 면적을 알지 못할 경우로 한정하여 모의실험을 실시하였다. 임의로 발생시킨 5000가구의 표본평균과 대체한 후의 표본평균의 값은 <표 3-34>에 주어져 있다. 10번 반복된 실험결과를 보면 실제표본평균과 대체한 후의 표본평균의 차이는 7.85평에서 29.27평 정도로 나왔으며, 10번의 평균은 18.31평으로 대체 후에도 자기발의 면적은 큰 차이가 나지 않음을 알 수 있다.

자기발 면적에 대한 분포변화를 고려하기 위하여 자기발 면적을 범주화하여 확인한 결과가 <표 3-35>에 주어져 있다. 0-199평은 1756가구(35.12%)에서 1730가구(34.60%), 200-399평은 595가구(11.90%)에서 604가구(12.08%), 400-599평은 570가구(11.40%)에서 578가구(11.56%), 600-799평은 339가구(6.78%)에서 341가구(6.82%), 1800평 이상은 755가구(15.10%)에서 772가구(15.44%) 등으로 변화되나 분포상의 문제는 거의 일어나지 않는다고 판단된다.

〈표 3-34〉 자기밭에 대한 대체군 및 표본평균 차이정도

| 대체군(1차)                         |        | 대체군(2차)                 |       |
|---------------------------------|--------|-------------------------|-------|
| 발면적(100)<br>영농형태(19)<br>남의논(18) |        | 판매금액구분(100)<br>영농형태(91) |       |
| 표본평균 절대차이의 평균 (18.31평)          |        |                         |       |
| 실험                              | 실제표본평균 | 대체후 표본평균                | 절대차이  |
| 1                               | 886.65 | 894.50                  | 7.85  |
| 2                               | 862.65 | 887.13                  | 24.48 |
| 3                               | 883.17 | 912.44                  | 29.27 |
| 4                               | 855.85 | 870.91                  | 15.06 |
| 5                               | 874.68 | 892.38                  | 17.70 |
| 6                               | 863.72 | 875.39                  | 11.67 |
| 7                               | 857.41 | 879.73                  | 22.32 |
| 8                               | 856.63 | 873.98                  | 17.35 |
| 9                               | 852.27 | 878.35                  | 26.08 |
| 10                              | 861.48 | 872.84                  | 11.36 |

〈표 3-35〉 자기밭에 대한 대체전후의 분포변화

| 자기밭               | 0-199           | 200-399        | 400-599        | 600-799       | 800-999       | 1000-1199     | 1200-1399     | 1400-1599     | 1600-1799    | 1800-          |
|-------------------|-----------------|----------------|----------------|---------------|---------------|---------------|---------------|---------------|--------------|----------------|
| 실제가구<br>(구성비(%))  | 1756<br>(35.12) | 595<br>(11.90) | 570<br>(11.40) | 339<br>(6.78) | 231<br>(4.62) | 328<br>(6.56) | 160<br>(3.20) | 188<br>(3.76) | 78<br>(1.56) | 755<br>(15.10) |
| 대체후가구<br>(구성비(%)) | 1730<br>(34.60) | 604<br>(12.08) | 578<br>(11.56) | 341<br>(6.82) | 229<br>(4.58) | 343<br>(6.86) | 168<br>(3.36) | 168<br>(3.36) | 67<br>(1.34) | 772<br>(15.44) |

### 10) 남의밭

남의밭의 경우는 앞에서 발면적과 자기밭의 면적을 알 수 있으므로  
연역적 대체방법(남의밭=발면적-자기밭)으로 바로 대체가 가능하다.

### 11) 목초지면적

목초지면적에 대한 무응답 대체는 영농형태, 판매금액구분, 발면적, 시도, 자기발 항목을 대체군으로 사용하였으며, 영농형태의 정보를 사용할 수 없을 경우 2차 대체군을 적용하였다. <표 3-36>의 결과를 살펴보면 실제표본평균과 대체한 후의 표본평균의 차이는 1.35평에서 3.34평 정도로 나왔으며, 10번의 평균은 2.22평인 것으로 나타났다. 그러나 목초지를 소유하고 있는 가구는 1% 미만이며 99% 이상이 목초지면적이 0평이므로 대체전후의 분포는 고려하지 않아도 될 것으로 판단된다. 왜냐하면 대부분이 실제값과 대체값이 0평이므로 분포변화는 나타나지 않는 것으로 보이기 때문이다.

<표 3-36> 목초지면적에 대한 대체군 및 표본평균 차이정도

| 대체군(1차)   |        | 대체군(2차)                                |      |
|---|--------|--|------|
| 영농형태(100)<br>판매금액구분(55)<br>발면적(55), 시도(49)<br>자기발(13) |        | 승용차보유(100)<br>컴퓨터활용여부(80)<br>화물차보유(68) |      |
| 표본평균 절대차이의 평균 (2.22평)                                 |        |  |      |
| 실험  | 실제표본평균 | 대체후 표본평균                               | 절대차이 |
| 1   | 19.86  | 22.67                                  | 2.81 |
| 2   | 22.22  | 20.87                                  | 1.35 |
| 3   | 20.16  | 21.89                                  | 1.73 |
| 4   | 19.11  | 22.45                                  | 3.34 |
| 5   | 23.11  | 25.97                                  | 2.86 |
| 6   | 20.45  | 22.18                                  | 1.73 |
| 7   | 19.68  | 21.95                                  | 2.27 |
| 8   | 19.78  | 21.85                                  | 2.07 |
| 9   | 21.75  | 20.13                                  | 1.62 |
| 10  | 19.45  | 21.83                                  | 2.38 |



## 다. 작물시설에 관한 사항

### 1) 자동화비닐하우스

자동화비닐하우스 면적에 대한 무응답 대체는 판매금액구분, 농업용 난방기, 비닐하우스, 영농형태, 판매처구분 항목을 대체군으로 사용하였으며, 판매금액구분 정보를 사용할 수 없을 경우 2차 대체군을 적용하였다. <표 3-37>의 결과를 살펴보면 5000가구의 실제표본평균과 대체한 후의 표본평균의 차이는 0.22평에서 0.46평 정도로 나왔으며, 10번의 평균은 0.32평인 것으로 나타났다. 자동화비닐하우스도 목초지면적처럼 소유하고 있는 가구가 1% 미만이기 때문에 대체전후의 분포는 변화되지 않을 것으로 보인다.

<표 3-37> 자동화비닐하우스에 대한 대체군 및 표본평균 차이정도

| 대체군(1차)   |        | 대체군(2차)                           |      |
|---|--------|-----------------------------------|------|
| 판매금액구분(100)<br>농업용난방기(77)<br>비닐하우스(67)<br>영농형태(41), 판매처구분(35) |        | 영농형태(100)<br>판매처구분(80)<br>밭면적(77) |      |
| 표본평균 절대차이의 평균 (0.32평)   |        |                                   |      |
| 실험  | 실제표본평균 | 대체후 표본평균                          | 절대차이 |
| 1   | 10.84  | 11.23                             | 0.39 |
| 2   | 9.86   | 10.10                             | 0.24 |
| 3   | 10.23  | 10.56                             | 0.33 |
| 4   | 10.19  | 10.65                             | 0.46 |
| 5   | 10.98  | 11.34                             | 0.36 |
| 6   | 10.45  | 10.67                             | 0.22 |
| 7   | 10.53  | 10.88                             | 0.35 |
| 8   | 10.32  | 10.56                             | 0.24 |
| 9   | 10.05  | 10.31                             | 0.26 |
| 10  | 10.43  | 10.79                             | 0.36 |

## 2) 비닐하우스

비닐하우스 면적에 대한 무응답 대체는 채소관련활동, 판매금액구분, 영농형태, 시도, 자동화비닐하우스, 판매처구분 항목을 대체군으로 사용하였으며, 채소관련활동 정보를 사용할 수 없을 경우 2차 대체군을 적용하였다. <표 3-38>의 결과에서 실제표본평균과 대체한 후의 표본평균의 차이는 0.42평에서 3.59평 정도로 나왔으며, 10번의 평균은 2.49평인 것으로 나타났다. 비닐하우스도 소유하고 있는 가구가 많지 않아 대체 전후 분포의 변화는 나타나지 않을 것으로 보인다.

<표 3-38> 비닐하우스에 대한 대체군 및 표본평균 차이정도

| 대체군(1차)  |        | 대체군(2차)                              |      |
|--|--------|--------------------------------------|------|
| 채소관련활동(100)<br>판매금액구분(66)<br>영농형태(60), 시도(26)<br>자동화비닐하우스(24)<br>판매처구분(35) |        | 판매금액구분(100)<br>영농형태(76)<br>판매처구분(46) |      |
| 표본평균 절대차이의 평균 (2.49평)  |        |                                      |      |
| 실험   | 실제표본평균 | 대체후 표본평균                             | 절대차이 |
| 1  | 110.45 | 113.52                               | 3.07 |
| 2  | 113.23 | 111.09                               | 2.41 |
| 3  | 111.33 | 114.24                               | 2.91 |
| 4  | 108.24 | 111.83                               | 3.59 |
| 5  | 109.18 | 112.45                               | 3.27 |
| 6  | 109.29 | 108.87                               | 0.42 |
| 7  | 110.28 | 112.47                               | 2.19 |
| 8  | 111.53 | 113.84                               | 2.31 |
| 9  | 109.76 | 113.24                               | 3.48 |
| 10   | 113.56 | 112.03                               | 1.53 |

### 3) 유리온실

유리온실 면적에 대한 무응답 대체는 판매금액구분, 농업용난방기수, 발면적, 채소관련활동 항목이 대체군으로 사용되며, 판매금액구분 정보가 없을 경우 2차 대체군이 적용된다. 유리온실은 소유하고 있는 가구가 0.1% 정도로 매우 적으며 대체전후의 값도 대부분 0평으로 나타나기 때문에 대체로 인한 영향은 거의 없을 것으로 판단된다. <표 3-39>의 결과에서 보듯이 실제표본평균과 대체한 후의 표본평균의 차이가 평균 0.02평으로 거의 나타나지 않음을 알 수 있다.

<표 3-39> 유리온실에 대한 대체군 및 표본평균 차이정도

| 대체군(1차)               |        | 대체군(2차)     |      |
|-----------------------|--------|-------------|------|
| 판매금액구분(100)           |        | 농업종사경력(100) |      |
| 농업용난방기(58)            |        | 컴퓨터보유여부(72) |      |
| 발면적(51)               |        | 컴퓨터활용여부(71) |      |
| 채소관련활동(13)            |        | 농업정보활용(51)  |      |
| 표본평균 절대차이의 평균 (0.02평) |        |             |      |
| 실험                    | 실제표본평균 | 대체후 표본평균    | 절대차이 |
| 1                     | 0.79   | 0.76        | 0.03 |
| 2                     | 0.71   | 0.70        | 0.01 |
| 3                     | 0.78   | 0.83        | 0.05 |
| 4                     | 0.74   | 0.76        | 0.02 |
| 5                     | 0.77   | 0.75        | 0.02 |
| 6                     | 0.73   | 0.72        | 0.01 |
| 7                     | 0.79   | 0.76        | 0.03 |
| 8                     | 0.80   | 0.76        | 0.04 |
| 9                     | 0.75   | 0.76        | 0.01 |
| 10                    | 0.74   | 0.76        | 0.02 |

### 4) 기타시설

기타시설 면적에 대한 무응답 대체는 판매금액구분, 영농형태, 건조기, 농업종사경력, 발면적, 판매처구분, 농업용난방기 항목이 대체군으

로 사용되며, 판매금액구분 정보가 없을 경우 2차 대체군을 적용하게 된다. 기타시설은 유리온실과 마찬가지로 소유하고 있는 가구가 0.1% 정도로 매우 적으며 대체전후의 값도 대부분 0평으로 나타나기 때문에 대체로 인한 영향은 거의 없을 것으로 판단된다. <표 3-40>의 결과에서 보듯이 실제표본평균과 대체한 후의 표본평균의 차이가 0.06평에서 0.14 평 정도로 거의 나타나지 않음을 알 수 있다. 작물시설에 관한 항목들은 소유하고 있는 가구가 매우 작기 때문에 전국 가구로의 결과보다는 해당가구만을 고려한 결과를 보는 것도 필요할 것으로 생각된다.

<표 3-40> 기타시설에 대한 대체군 및 표본평균 차이정도

| 대체군(1차)  |        | 대체군(2차)                                      |      |
|--|--------|--|------|
| 판매금액구분(100)<br>영농형태(95), 건조기(54)<br>농업종사경력(50), 발면적(46)<br>판매처구분(41)<br>농업용난방기(12) |        | 화물차보유(100)<br>영농형태(83)<br>발면적(63)<br>관리기(56) |      |
| 표본평균 절대차이의 평균 (0.10평)  |        |  |      |
| 실험   | 실제표본평균 | 대체후 표본평균                                     | 절대차이 |
| 1  | 3.04   | 3.12   | 0.08 |
| 2  | 3.15   | 3.01   | 0.14 |
| 3  | 3.02   | 2.91   | 0.11 |
| 4  | 3.11   | 3.23   | 0.12 |
| 5  | 3.06   | 3.15   | 0.09 |
| 6  | 2.95   | 3.04   | 0.09 |
| 7  | 3.03   | 3.15   | 0.12 |
| 8  | 3.14   | 3.04   | 0.10 |
| 9  | 3.09   | 3.03   | 0.06 |
| 10   | 2.98   | 3.08   | 0.10 |

## 라. 농기계 및 논벼 농사방법에 관한 사항

### 1) 경운기

경운기 대수에 대한 무응답 대체는 농약살포방법, 이앙기, 관리기, 판매금액구분, 밭면적, 시도 항목을 대체군으로 사용하였으며, 농약살포 방법 정보를 사용할 수 없을 경우 2차 대체군을 적용하였다. <표 3-41>의 결과에서 보듯이 5000가구의 실제표본평균과 대체한 후의 표본평균의 차이는 평균 0.0064대로 거의 없는 것으로 판단된다.

<표 3-41> 경운기에 대한 대체군 및 표본평균 차이정도

| 대체군(1차)  |        | 대체군(2차)  |       |
|--|--------|--|-------|
| 농약살포방법(100)<br>이앙기(49), 관리기(46)<br>판매금액구분(43)<br>밭면적(16), 시도(12) |        | 모내기방법(100)<br>판매금액구분(67)<br>벼묘판작업방법(37)<br>논면적(15), 시도(12) |       |
| 표본평균 절대차이의 평균 (0.0064대)  |        |  |       |
| 실험   | 실제표본평균 | 대체후 표본평균   | 절대차이  |
| 1  | 0.598  | 0.608  | 0.010 |
| 2  | 0.614  | 0.618  | 0.004 |
| 3  | 0.609  | 0.606  | 0.003 |
| 4  | 0.614  | 0.601  | 0.013 |
| 5  | 0.581  | 0.583  | 0.002 |

경운기 수는 연속형 항목이지만 각각의 대수를 범주로 간주하여 대체전후의 분포변화를 고려할 수 있다. 0대를 소유하고 있는 가구는 2251가구(45.02%)에서 2228가구(44.56%), 1대는 2508가구(50.16%)에서 2519가구(50.38%), 2대는 220가구(4.40%)에서 235가구(4.70%), 3대는 20가구(0.40%)에서 18가구(0.36%) 등으로 변화되나 분포상의 문제는 나타나지 않는다고 판단된다. 또한 개개의 가구에 대해서도 얼마나 정확한 대체가 이루어지고 있는지 살펴보았다. 정확한 대수를 추정된 가구는

총 5000가구 중에서 3564가구(71.28%), 1대의 차이가 난 가구는 1349가구(26.98%), 2대는 81가구(1.62%), 3대는 6가구(0.12%)로 나타났다. 자세한 실험결과는 <표 3-42>를 참조하기 바란다.

<표 3-42> 경운기에 대한 대체전후의 분포변화 및 정확도 검토결과

| 대수                | 0대              | 1대              | 2대            | 3대           | 4대이상        |
|-------------------|-----------------|-----------------|---------------|--------------|-------------|
| 실제가구<br>(구성비(%))  | 2251<br>(45.02) | 2508<br>(50.16) | 220<br>(4.40) | 20<br>(0.40) | 1<br>(0.02) |
| 대체후가구<br>(구성비(%)) | 2228<br>(44.56) | 2519<br>(50.38) | 235<br>(4.70) | 18<br>(0.36) | 0<br>(0.00) |
| 대수차이              | 일치              | 1대차이            | 2대차이          | 3대차이         | 4대이상        |
| 가구<br>(구성비(%))    | 3564<br>(71.28) | 1349<br>(26.98) | 81<br>(1.62)  | 6<br>(0.12)  | —           |

## 2) 트랙터

트랙터 대수에 대한 무응답 대체는 논갈이방법, 논면적, 판매금액구분, 콤바인 항목을 대체군으로 사용하였으며, 논갈이방법 정보를 사용할 수 없을 경우 2차 대체군을 적용하였다. <표 3-43>의 결과에서 보듯이 5번의 반복실험에서 5000가구의 실제표본평균과 대체한 후의 표본평균의 차이는 평균 0.0014대로 거의 없는 것으로 판단된다.

트랙터도 경운기와 마찬가지로 각각의 대수를 범주로 간주하여 대체전후의 분포변화를 고려할 수 있다. <표 3-44>에 주어진 결과를 보면 0대를 소유하고 있는 가구는 4165가구(83.30%)에서 4155가구(83.10%), 1대는 780가구(15.60%)에서 798가구(15.86%), 2대는 49가구(0.98%)에서 40가구(0.80%), 3대는 4가구(0.08%)에서 6가구(0.12%), 4대 이상은 2가구(0.04%)에서 1가구(0.02%)로 변화되나 대체로 인한 분포상의 문제는 나타나지 않는다고 판단된다. 그리고 개개의 가구에 대한 정확도 검토결과 정확한 대수를 추정한 가구는 총 5000가구 중에서 4397가구(87.94%), 1대의 차이가 난 가구는 584가구(11.68%), 2대는 18가구(0.36%), 3대는 1가구(0.02%)로 나타나 정확도가 높은 대체가 이루어졌음을 알 수 있다.

〈표 3-43〉 트랙터에 대한 대체군 및 표본평균 차이정도

| 대체군(1차)  |        | 대체군(2차)  |       |
|--|--------|--|-------|
| 논갈이방법(100)<br>논면적(58)<br>판매금액구분(37)<br>콤바인(20) |        | 모내기방법(100)<br>논면적(72)<br>판매금액구분(37)<br>벼베기탈곡방법(44) |       |
| 표본평균 절대차이의 평균 (0.0014대)                        |        |  |       |
| 실험   | 실제표본평균 | 대체후 표본평균   | 절대차이  |
| 1  | 0.179  | 0.178  | 0.001 |
| 2  | 0.181  | 0.180  | 0.001 |
| 3  | 0.180  | 0.180  | 0.000 |
| 4  | 0.179  | 0.176  | 0.003 |
| 5  | 0.179  | 0.177  | 0.002 |

〈표 3-44〉 트랙터에 대한 대체전후의 분포변화 및 정확도 검토결과

| 대수                | 0대              | 1대             | 2대           | 3대          | 4대이상        |
|-------------------|-----------------|----------------|--------------|-------------|-------------|
| 실제가구<br>(구성비(%))  | 4165<br>(83.30) | 780<br>(15.60) | 49<br>(0.98) | 4<br>(0.08) | 2<br>(0.04) |
| 대체후가구<br>(구성비(%)) | 4155<br>(83.10) | 798<br>(15.86) | 40<br>(0.80) | 6<br>(0.12) | 1<br>(0.02) |
| 대수차이              | 일치              | 1대차이           | 2대차이         | 3대차이        | 4대이상        |
| 가구<br>(구성비(%))    | 4397<br>(87.94) | 584<br>(11.68) | 18<br>(0.36) | 1<br>(0.02) | —           |

### 3) 콤바인

콤바인 대수에 대한 무응답 대체는 벼베기탈곡방법, 트랙터, 논면적, 이앙기, 논갈이방법 항목을 대체군으로 사용하였으며, 벼베기탈곡방법 정보를 사용할 수 없을 경우 2차 대체군을 적용하였다. <표 3-45>에 주어진 5번의 반복실험의 결과 5000가구의 실제표본평균과 대체한 후의 표본평균의 차이는 평균 0.0012대로 거의 나타나지 않는다.

〈표 3-45〉 콤바인에 대한 대체군 및 표본평균 차이정도

| 대체군(1차)  |        | 대체군(2차)               |       |
|--|--------|-----------------------|-------|
| 벼베기탈곡방법(100)<br>트랙터(57), 논면적(17)<br>이앙기(17), 논갈이방법(15) |        | 논갈이방법(100)<br>논면적(97) |       |
| 표본평균 절대차이의 평균 (0.0012대)                                |        |                       |       |
| 실험   | 실제표본평균 | 대체후 표본평균              | 절대차이  |
| 1  | 0.072  | 0.071                 | 0.001 |
| 2  | 0.070  | 0.072                 | 0.002 |
| 3  | 0.075  | 0.074                 | 0.001 |
| 4  | 0.075  | 0.076                 | 0.001 |
| 5  | 0.071  | 0.070                 | 0.001 |

〈표 3-46〉 콤바인에 대한 대체전후의 분포변화 및 정확도 검토결과

| 대수                | 0대              | 1대            | 2대          | 3대          |
|-------------------|-----------------|---------------|-------------|-------------|
| 실제가구<br>(구성비(%))  | 4642<br>(92.84) | 355<br>(7.10) | 2<br>(0.04) | 1<br>(0.02) |
| 대체후가구<br>(구성비(%)) | 4646<br>(92.92) | 349<br>(6.98) | 5<br>(0.10) | 0<br>(0.00) |
| 대수차이              | 일치              | 1대차이          | 2대차이        | 3대차이        |
| 가구<br>(구성비(%))    | 4827<br>(96.54) | 172<br>(3.44) | 1<br>(0.02) | —           |

〈표 3-46〉에 주어진 콤바인 수에 대한 대체전후 분포변화를 살펴보면 0대를 소유하고 있는 가구는 4642가구(92.84%)에서 4646가구(92.92%), 1대는 355가구(7.10%)에서 349가구(6.98%), 2대는 2가구(0.04%)에서 5가구(0.10%) 등으로 큰 변화는 나타나지 않는다. 그리고 개개의 가구에 대한 정확도 검토결과 정확한 대수를 추정한 가구는 총 5000가구 중에서 4827가구(96.54%), 1대의 차이가 난 가구는 172가구(3.44%), 2대는 1가구(0.02%)로 경운기와 트랙터에 비해서 상당히 정확한 대체가 되었음을 알 수 있다.



#### 4) 관리기

관리기 대수에 대한 무응답 대체는 판매금액구분, 경운기, 시도, 밭면적, 트랙터, 비닐하우스 항목을 대체군으로 사용하였으며, 판매금액구분 정보를 사용할 수 없을 경우 2차 대체군을 적용하였다. <표 3-47>에 주어진 5번의 반복실험의 결과 5000가구의 실제표본평균과 대체한 후의 표본평균의 차이는 평균 0.0084대인 것으로 나타나 대체로 인한 평균의 차이는 보이지 않는다.

<표 3-47> 관리기에 대한 대체군 및 표본평균 차이정도

| 대체군(1차)   |        | 대체군(2차)   |       |
|---|--------|---|-------|
| 판매금액구분(100)<br>경운기(58), 시도(58)<br>밭면적(51), 트랙터(24)<br>비닐하우스(19) |        | 밭면적(100), 시도(49)<br>농약살포방법(43)<br>논갈이방법(35)<br>영농형태(28) |       |
| 표본평균 절대차이의 평균 (0.0084대)   |        |   |       |
| 실험  | 실제표본평균 | 대체후 표본평균  | 절대차이  |
| 1   | 0.277  | 0.274   | 0.003 |
| 2   | 0.287  | 0.274   | 0.013 |
| 3   | 0.280  | 0.273   | 0.007 |
| 4   | 0.284  | 0.270   | 0.014 |
| 5   | 0.286  | 0.281   | 0.005 |

<표 3-48>에 주어진 관리기 수에 대한 대체전후 분포변화를 살펴보면 0대를 소유하고 있는 가구는 3610가구(72.20%)에서 3643가구(72.86%), 1대는 1349가구(26.98%)에서 1310가구(26.20%), 2대는 35가구(0.70%)에서 42가구(0.84%), 3대는 6가구(0.12%)에서 5가구(0.10%)로 대체로 인한 분포변화는 보이지 않는다. 그리고 개개의 가구에 대한 정확도 검토결과 정확한 대수를 추정한 가구는 총 5000가구 중에서 3595가구(71.90%), 1대의 차이가 난 가구는 1372가구(27.44%), 2대는 31가구(0.62%), 3대는 2가구(0.04%)로 콤바인에 비해서는 다소 정확도가 떨어지나 일반적으로 적절한 대체가 되었다고 판단된다.

〈표 3-48〉 관리기에 대한 대체전후의 분포변화 및 정확도 검토결과

| 대수                | 0대              | 1대              | 2대           | 3대          |
|-------------------|-----------------|-----------------|--------------|-------------|
| 실제가구<br>(구성비(%))  | 3610<br>(72.20) | 1349<br>(26.98) | 35<br>(0.70) | 6<br>(0.12) |
| 대체후가구<br>(구성비(%)) | 3643<br>(72.86) | 1310<br>(26.20) | 42<br>(0.84) | 5<br>(0.10) |
| 대수차이              | 일치              | 1대차이            | 2대차이         | 3대차이        |
| 가구<br>(구성비(%))    | 3595<br>(71.90) | 1372<br>(27.44) | 31<br>(0.62) | 2<br>(0.04) |

### 5) 건조기

건조기 대수에 대한 무응답 대체는 이앙기, 콤바인, 발면적, 시도, 관리기, 논면적, 경운기, 트랙터 항목을 대체군으로 사용하였으며, 이앙기 정보를 사용할 수 없을 경우 2차 대체군을 적용하였다. <표 3-49>에 주어진 5번의 반복실험의 결과 5000가구의 실제표본평균과 대체한 후의 표본평균의 차이는 평균 0.0038대로 대체를 한 후에도 평균의 차이는 보이지 않는다는 것을 알 수 있다.

〈표 3-49〉 건조기에 대한 대체군 및 표본평균 차이정도

| 대체군(1차)  | 대체군(2차)   |          |       |
|--|---|----------|-------|
| 이앙기(100), 콤바인(63)<br>발면적(54), 시도(52)<br>관리기(44), 논면적(40)<br>경운기(27), 트랙터(24) | 모내기(100), 논면적(70)<br>발면적(67), 시도(45)<br>벼베기탈곡방법(38)<br>농약살포방법(18) |          |       |
| 표본평균 절대차이의 평균 (0.0038대)  |   |          |       |
| 실험   | 실제표본평균  | 대체후 표본평균 | 절대차이  |
| 1  | 0.176   | 0.171    | 0.005 |
| 2  | 0.176   | 0.169    | 0.007 |
| 3  | 0.165   | 0.169    | 0.004 |
| 4  | 0.168   | 0.167    | 0.001 |
| 5  | 0.172   | 0.170    | 0.002 |

<표 3-50>에 주어진 건조기 수에 대한 대체전후 분포변화를 살펴보면 0대를 소유하고 있는 가구는 4236가구(84.72%)에서 4258가구(85.16%), 1대는 683가구(13.66%)에서 658가구(13.16%), 2대는 57가구(1.14%)에서 60가구(1.20%), 3대는 19가구(0.38%)에서 17가구(0.34%), 4대 이상은 5가구(0.10%)에서 7가구(0.14%)로 대체 후에도 분포는 변화가 없는 것으로 보인다. 또한 개개의 가구에 대한 정확도 검토결과 정확한 대수를 추정한 가구는 총 5000가구 중에서 4126가구(82.52%), 1대의 차이가 난 가구는 823가구(16.46%), 2대는 42가구(0.84%), 3대는 7가구(0.14%), 4대 이상은 2가구(0.04%)로 나타났다. 건조기를 많이 가지고 있는 가구는 10대 이상을 가지고 있어 대체의 차이가 4대 이상 나는 경우도 발생하는 것으로 판단된다.

<표 3-50> 건조기에 대한 대체전후의 분포변화 및 정확도 검토결과

| 대수                | 0대              | 1대             | 2대           | 3대           | 4대이상        |
|-------------------|-----------------|----------------|--------------|--------------|-------------|
| 실제가구<br>(구성비(%))  | 4236<br>(84.72) | 683<br>(13.66) | 57<br>(1.14) | 19<br>(0.38) | 5<br>(0.10) |
| 대체후가구<br>(구성비(%)) | 4258<br>(85.16) | 658<br>(13.16) | 60<br>(1.20) | 17<br>(0.34) | 7<br>(0.14) |
| 대수차이              | 일치              | 1대차이           | 2대차이         | 3대차이         | 4대이상        |
| 가구<br>(구성비(%))    | 4126<br>(82.52) | 823<br>(16.46) | 42<br>(0.84) | 7<br>(0.14)  | 2<br>(0.04) |

## 6) 이양기

이양기 대수에 대한 무응답 대체는 모내기방법, 논면적, 경운기, 논갈이방법 항목을 대체군으로 사용하였으며, 모내기방법 정보를 사용할 수 없을 경우 2차 대체군을 적용하였다. <표 3-51>에 주어진 5번의 반복 실험의 결과 5000가구의 실제표본평균과 대체한 후의 표본평균의 차이는 평균 0.0018대로 대체를 한 후에도 평균의 차이는 보이지 않는다는 것을 알 수 있다. 건조기 수에 대한 대체전후 분포변화를 살펴보면 0대를 소유하고 있는 가구는 3849가구(76.98%)에서 3846가구(76.92%), 1대는 1132가구(22.64%)에서 1135가구(22.70%), 2대는 19가구(0.38%)에서

18가구(0.36%) 등으로 대체 후에도 분포는 변화가 없는 것으로 보인다. 그리고 개개의 가구에 대한 정확도 검토결과 정확한 대수를 추정한 가구는 총 5000가구 중에서 4508가구(90.16%), 1대의 차이가 난 가구는 486가구(9.72%), 2대는 5가구(0.10%), 3대는 1가구(0.02%)로 나타났다. 이양기는 다른 농기계에 비해 대체의 정확도가 높은 것으로 판단된다. 자세한 내용은 <표 3-52>를 참조하기 바란다.

<표 3-51> 이양기에 대한 대체군 및 표본평균 차이정도

| 대체군(1차)                                   |        | 대체군(2차)  |       |
|---|--------|--|-------|
| 모내기방법(100), 논면적(31)<br>경운기(22), 논갈이방법(11) |        | 논면적(100), 논갈이방법(65)<br>농약살포방법(50)<br>벼베기탈곡방법(38) |       |
| 표본평균 절대차이의 평균 (0.0018대)                   |        |  |       |
| 실험  | 실제표본평균 | 대체후 표본평균   | 절대차이  |
| 1   | 0.234  | 0.235  | 0.001 |
| 2   | 0.235  | 0.239  | 0.004 |
| 3   | 0.238  | 0.237  | 0.001 |
| 4   | 0.241  | 0.242  | 0.001 |
| 5   | 0.239  | 0.241  | 0.002 |

<표 3-52> 이양기에 대한 대체전후의 분포변화 및 정확도 검토결과

| 대수                | 0대              | 1대              | 2대           | 3대          |
|-------------------|-----------------|-----------------|--------------|-------------|
| 실제가구<br>(구성비(%))  | 3849<br>(76.98) | 1132<br>(22.64) | 19<br>(0.38) | 0<br>(0.00) |
| 대체후가구<br>(구성비(%)) | 3846<br>(76.92) | 1135<br>(22.70) | 18<br>(0.36) | 1<br>(0.02) |
| 대수차이              | 일치              | 1대차이            | 2대차이         | 3대차이        |
| 가구<br>(구성비(%))    | 4508<br>(90.16) | 486<br>(9.72)   | 5<br>(0.10)  | 1<br>(0.02) |

## 7) SS분무기

SS분무기 대수에 대한 무응답 대체는 영농형태, 밭면적, 시도, 판매금액구분, 트랙터, 농업특성조사구 항목을 대체군으로 사용하였으며, 영농형태 정보를 사용할 수 없을 경우 2차 대체군을 적용하였다. <표 3-53>에 주어진 5번의 반복실험의 결과 5000가구의 실제표본평균과 대체한 후의 표본평균의 차이는 평균 0.0016대인 것으로 나타나 대체로 인한 평균의 차이는 보이지 않는다.

<표 3-53> SS분무기에 대한 대체군 및 표본평균 차이정도

| 대체군(1차)  |        | 대체군(2차)  |       |
|--|--------|--|-------|
| 영농형태(100), 밭면적(75)<br>시도(57), 판매금액구분(31)<br>트랙터(29), 농업특성조사구(11) |        | 자기밭(100), 과수관련활동(73)<br>농업특성조사구(47)<br>밭면적(22) |       |
| 표본평균 절대차이의 평균 (0.0016대)  |        |  |       |
| 실험   | 실제표본평균 | 대체후 표본평균                                       | 절대차이  |
| 1  | 0.032  | 0.029  | 0.003 |
| 2  | 0.028  | 0.029  | 0.001 |
| 3  | 0.030  | 0.029  | 0.001 |
| 4  | 0.029  | 0.031  | 0.002 |
| 5  | 0.031  | 0.032  | 0.001 |

SS분무기는 많은 가구에서 보유하고 있지 않는 농기계로서 대부분이 0대로 대체되고 있음을 알 수 있다. 대체전후 분포변화를 살펴보면 0대를 보유하고 있는 가구는 4846가구(96.92%)에서 4846가구(66.92%), 1대는 148가구(2.96%)에서 149가구(2.98%), 2대는 3가구(0.06%)에서 3가구(0.06%) 등으로 대체전후의 분포가 거의 같게 나타났다. 또한 개개의 가구에 대한 정확도 검토결과를 보면 정확한 대수를 추정된 가구는 총 5000가구 중에서 4844가구(96.88%)로 매우 정확도가 높음을 보여준다. 1대의 차이가 난 가구는 155가구(3.10%), 2대는 1가구(0.02%)로 나타났다. 자세한 내용은 <표 3-54>를 참조하기 바란다.

〈표 3-54〉 SS분무기에 대한 대체전후의 분포변화 및 정확도 검토결과

| 대수                | 0대              | 1대            | 2대          | 3대          |
|-------------------|-----------------|---------------|-------------|-------------|
| 실제가구<br>(구성비(%))  | 4846<br>(96.92) | 148<br>(2.96) | 3<br>(0.06) | 3<br>(0.06) |
| 대체후가구<br>(구성비(%)) | 4846<br>(96.92) | 149<br>(2.98) | 3<br>(0.06) | 2<br>(0.04) |
| 대수차이              | 일치              | 1대차이          | 2대차이        | 3대차이        |
| 가구<br>(구성비(%))    | 4844<br>(96.88) | 155<br>(3.10) | 1<br>(0.02) | —           |

### 8) 농업용난방기

농업용난방기 대수에 대한 무응답 대체는 판매금액구분, 비닐하우스, 자동화비닐하우스, 화훼관련활동, 영농형태, 시도 항목을 대체군으로 사용하였으며, 판매금액구분 정보를 사용할 수 없을 경우 2차 대체군을 적용하였다. <표 3-55>에 주어진 5번의 반복실험의 결과 5000가구의 실제표본평균과 대체한 후의 표본평균의 차이는 평균 0.0032대로 대체를 한 후에도 평균의 차이는 보이지 않는다는 것을 알 수 있다.

〈표 3-55〉 농업용난방기에 대한 대체군 및 표본평균 차이정도

| 대체군(1차)   | 대체군(2차)   |          |       |
|---|---|----------|-------|
| 판매금액구분(100), 비닐하우스(85)<br>자동화비닐하우스(77),<br>화훼관련활동(51)<br>영농형태(51), 시도(24) | 판매처구분(100), 영농형태(77)<br>컴퓨터활용여부(58)<br>비닐하우스(41), 발면적(22) |          |       |
| 표본평균 절대차이의 평균 (0.0032대)   |   |          |       |
| 실험  | 실제표본평균  | 대체후 표본평균 | 절대차이  |
| 1   | 0.079   | 0.083    | 0.004 |
| 2   | 0.078   | 0.075    | 0.003 |
| 3   | 0.079   | 0.075    | 0.004 |
| 4   | 0.081   | 0.083    | 0.002 |
| 5   | 0.079   | 0.082    | 0.003 |

<표 3-56>에 주어진 농업용난방기 대수에 대한 대체전후 분포변화를 살펴보면 0대를 소유하고 있는 가구는 4815가구(96.30%)에서 4810가구(96.20%), 1대는 102가구(2.04%)에서 107가구(2.14%), 2대는 36가구(0.72%)에서 39가구(0.78%), 3대는 11가구(0.22%)에서 13가구(0.26%), 4대 이상은 36가구(0.72%)에서 31가구(0.62%)로 대체 후에도 분포는 변화가 없는 것으로 보인다. 또한 개개의 가구에 대한 정확도 검토결과 정확한 대수를 추정한 가구는 총 5000가구 중에서 4748가구(94.96%), 1대의 차이가 난 가구는 171가구(3.42%), 2대는 45가구(0.90%), 3대는 33가구(0.66%), 4대 이상은 3가구(0.06%)로 나타났다. 농업용난방기도 건조기와 마찬가지로 대부분이 소유하고 있지는 않지만 소유하고 있는 가구는 많은 대수를 보유하고 있어 대체의 차이가 4대 이상 나는 경우도 발생되는 것으로 판단된다. 하지만 전체적으로 대체의 정확도는 상당히 높다는 것을 보여주고 있다.

<표 3-56> 농업용난방기에 대한 대체전후의 분포변화 및 정확도 검토결과

| 대수                | 0대              | 1대            | 2대           | 3대           | 4대이상         |
|-------------------|-----------------|---------------|--------------|--------------|--------------|
| 실제가구<br>(구성비(%))  | 4815<br>(96.30) | 102<br>(2.04) | 36<br>(0.72) | 11<br>(0.22) | 36<br>(0.72) |
| 대체후가구<br>(구성비(%)) | 4810<br>(96.20) | 107<br>(2.14) | 39<br>(0.78) | 13<br>(0.26) | 31<br>(0.62) |
| 대수차이              | 일치              | 1대차이          | 2대차이         | 3대차이         | 4대이상         |
| 가구<br>(구성비(%))    | 4748<br>(94.96) | 171<br>(3.42) | 45<br>(0.90) | 33<br>(0.66) | 3<br>(0.06)  |

### 9) 비묘판작업

비묘판작업에 대한 연관성 분석의 결과 농약살포방법, 논갈이방법, 모내기방법, 시도, 읍면동을 대체군으로 사용하게 되나 이 경우 논면적을 특별 대체군으로 추가하여 대체를 실시하였다. 그 이유는 논벼 농사방법 항목은 논벼를 소유하고 있는 가구만을 대상으로 기입을 하게 하므로 논면적에 맞는 대체가 이루어져야 하기 때문이다. 따라서 논면적이 없는 경우는 허용결측으로, 있는 경우는 적절한 대체값으로 대체

가 되어야 한다. 그러므로 논면적을 대체군에 추가함으로써 대체 후 내 검규칙에 합당하게 될 것이라 판단된다. 그리고 농약살포방법 정보를 사용할 수 없을 경우에는 2차 대체군을 적용하였다. <표 3-57>에서와 같이 총 5,000가구를 대체한 결과 대체의 정확도는 82.88%로 높은 정확도를 보였다. 허용결측은 논면적을 특별 대체군으로 사용함으로써 1303가구 모두를 정확히 대체하였다. 그러나 허용결측을 제외한 비묘판작업의 순수 정확도는 76.84%로 조금 낮아지는 것을 알 수 있다. 자신이 직접 한 경우는 높은 정확도를 보여준 반면 타인이 일부 또는 전부 한 경우는 다소 오분류율이 높아진 것으로 보여진다. 이는 구성비가 작은 항목들은 연관규칙이 잘 나타나지 못하거나 구성비가 큰 항목으로 묻혀버리는 현상이 일어나기 때문인 것으로 판단된다.

<표 3-57> 비묘판작업에 대한 대체군 및 대체결과

| 대체군(1차)  |              | 대체군(2차)                                  |              |              |              |
|--|--------------|--|--------------|--------------|--------------|
| 농약살포방법(100), 논갈이방법(47)<br>모내기방법(33), 시도(28)<br>읍면동(17), 논면적(특별대체군) |              | 경운기(100), 수도형태(31)<br>시도(25), 논면적(특별대체군) |              |              |              |
| 대체결과_대체정확도 (82.88%), 허용결측 제외한 정확도(76.84%)                          |              |  |              |              |              |
|  | 자신직접<br>(대체) | 타인일부<br>(대체)                             | 타인전부<br>(대체) | 허용결측<br>(대체) | 합계<br>(구성비%) |
| 자신직접   | 2249         | 125                                      | 251          | 0            | 2625(52.50)  |
| 타인일부   | 121          | 142                                      | 56           | 0            | 319(6.38)    |
| 타인전부   | 246          | 57                                       | 450          | 0            | 753(15.06)   |
| 허용결측   | 0            | 0  | 0            | 1303         | 1303(26.06)  |
| 합계   | 2616(52.32)  | 324(6.48)                                | 757(15.14)   | 1303(26.06)  | 5000         |

대체 전후의 항목분포의 변화를 살펴보면 자신이 직접 한 경우는 실제 가구가 2625가구(52.50%)에서 2616가구(52.32%), 타인일부는 319가구(6.38%)에서 324가구(6.48%), 타인전부는 753가구(15.06%)에서 757가구(15.14%)로 변화되어 대체로 인하여 발생할 수 있는 분포의 왜곡 문제는 나타나지 않음을 알 수 있다.



## 10) 논갈이방법

논갈이방법에 대한 무응답 대체는 모내기방법, 트랙터, 벼베기탈곡방법, 시도, 농약살포방법 항목을 대체군으로 사용하였으며, 벼묘판작업과 같은 이유로 논면적을 특별 대체군으로 추가하여 대체를 실시하였다. 그리고 모내기방법 정보를 사용할 수 없을 경우에는 2차 대체군을 적용하였다. <표 3-58>에서와 같이 총 5,000가구를 대체한 결과 대체의 정확도는 86.74%이며, 허용결측을 제외한 논갈이방법의 순수 정확도는 82.27%로 나타났다. 자신이 직접 한 경우와 타인이 전부 한 경우에는 높은 정확도를 보여준 반면 타인이 일부 한 경우는 다소 오분류율이 높아진 것으로 보여진다. 이는 앞의 경우처럼 구성비가 상당히 작아 발생하는 현상으로 이해하면 될 것이다.

<표 3-58> 논갈이방법에 대한 대체군 및 대체결과

| 대체군(1차)  |              | 대체군(2차)                |              |              |                |
|--|--------------|------------------------|--------------|--------------|----------------|
| 모내기방법(100), 트랙터(60)<br>벼베기탈곡방법(33), 시도(19)<br>농약살포방법(11), 논면적(특별대체군) |              | 트랙터(100)<br>논면적(특별대체군) |              |              |                |
| 대체결과 대체정확도 (86.74%), 허용결측 제외한 정확도(82.27%)                            |              |                        |              |              |                |
|  | 자신직접<br>(대체) | 타인일부<br>(대체)           | 타인전부<br>(대체) | 허용결측<br>(대체) | 합계<br>(구성비(%)) |
| 자신직접   | 1147         | 56                     | 203          | 0            | 1406(28.12)    |
| 타인일부   | 43           | 150                    | 57           | 0            | 250(5.00)      |
| 타인전부   | 235          | 69                     | 1780         | 0            | 2084(41.68)    |
| 허용결측   | 0            | 0                      | 0            | 1260         | 1260(25.20)    |
| 합계   | 1425(28.50)  | 275(5.50)              | 2040(40.80)  | 1260(25.20)  | 5000           |

대체 전후의 항목분포의 변화를 살펴보면 자신이 직접 한 경우는 실제 가구가 1406가구(28.12%)에서 1425가구(28.50%), 타인일부는 250가구(5.00%)에서 275가구(5.50%), 타인전부는 2084가구(41.68%)에서 2040가구(40.80%)로 변화되어 분포변화는 발생하지 않는다고 볼 수 있다.

### 11) 모내기방법

모내기방법에 대한 무응답 대체는 이앙기, 논갈이방법, 벼베기탈곡방법, 농약살포방법 항목을 대체군으로 사용하였으며, 논면적을 특별 대체군으로 추가하여 대체를 실시하였다. 또한 이후에 실제 대체적용을 할 때에는 이앙기 항목을 먼저 대체하고자 하므로 1차 대체군을 이용할 수 있을 것으로 판단된다. <표 3-59>의 결과를 보면 총 5,000가구를 대체한 결과 대체의 정확도는 88.86%이며, 허용결측을 제외한 모내기방법의 순수 정확도는 84.97%로 높은 일치율을 보인다. 대체의 내용은 논갈이방법과 거의 유사한 형태를 보이고 있음을 알 수 있다.

<표 3-59> 모내기방법에 대한 대체군 및 대체결과

| 대체군(1차)  |              | 대체군(2차)                         |              |              |                |
|--|--------------|---------------------------------|--------------|--------------|----------------|
| 이앙기(100), 논갈이방법(60)<br>벼베기탈곡방법(30)<br>농약살포방법(11), 논면적(특별대체군) |              | 트랙터(100), 경운기(67)<br>논면적(특별대체군) |              |              |                |
| 대체결과_대체정확도 (88.86%), 허용결측 제외한 정확도(84.97%)                    |              |                                 |              |              |                |
|  | 자신직접<br>(대체) | 타인일부<br>(대체)                    | 타인전부<br>(대체) | 허용결측<br>(대체) | 합계<br>(구성비(%)) |
| 자신직접   | 1224         | 75                              | 152          | 0            | 1451(29.02)    |
| 타인일부   | 60           | 184                             | 66           | 0            | 310(6.20)      |
| 타인전부   | 135          | 69                              | 1741         | 0            | 1945(38.90)    |
| 허용결측   | 0            | 0                               | 0            | 1294         | 1294(25.88)    |
| 합계   | 1419(28.38)  | 328(6.56)                       | 1959(39.18)  | 1294(25.88)  | 5000           |

대체 전후의 항목분포의 변화를 살펴보면 자신이 직접 한 경우는 실제 가구가 1451가구(29.02%)에서 1419가구(28.38%), 타인일부는 310가구(6.20%)에서 328가구(6.56%), 타인전부는 1945가구(38.90%)에서 1959가구(39.18%)로 나타나 대체로 인한 분포변화는 발생하지 않는다고 판단된다.

## 12) 농약살포방법

농약살포방법에 대한 무응답 대체는 벼묘판작업, 논갈이방법, 경운기, 시도, 모내기방법 항목을 대체군으로 사용하였으며, 논면적을 특별 대체군으로 추가하여 대체를 실시하였다. 또한 이후에 실제 대체적용을 할 때에는 벼묘판작업 항목을 먼저 대체하고자 하므로 1차 대체군을 이용할 수 있을 것으로 판단된다. <표 3-60>의 결과를 보면 총 5,000가구를 대체한 결과 대체의 정확도는 82.64%이며, 허용결측을 제외한 농약살포방법의 순수 정확도는 76.59%로 다른 논벼 농사방법 항목에 비해서는 다소 낮은 것으로 나타났다. 타인일부 항목에서 대체의 정확도가 많이 떨어지나 전체적으로 볼 때에는 그 수가 상당히 작기 때문에 오분류로 인한 효과는 미미하게 나타난다.

<표 3-60> 농약살포방법에 대한 대체군 및 대체결과

| 대체군(1차)   |              | 대체군(2차)                                    |              |              |              |
|---|--------------|--|--------------|--------------|--------------|
| 벼묘판작업(100), 논갈이방법(61)<br>경운기(51), 시도(18)<br>모내기방법(14), 논면적(특별대체군) |              | 경운기(100), 가구원수합계(28)<br>시도(19), 논면적(특별대체군) |              |              |              |
| 대체결과 대체정확도 (82.64%), 허용결측 제외한 정확도(76.59%)                         |              |  |              |              |              |
|   | 자신직접<br>(대체) | 타인일부<br>(대체)                               | 타인전부<br>(대체) | 허용결측<br>(대체) | 합계<br>(구성비%) |
| 자신직접  | 2152         | 101  | 270          | 0            | 2523(50.46)  |
| 타인일부  | 101          | 120  | 52           | 0            | 273(5.46)    |
| 타인전부  | 289          | 55   | 568          | 0            | 912(18.24)   |
| 허용결측  | 0            | 0  | 0            | 1292         | 1292(25.84)  |
| 합계  | 2542(50.84)  | 276(5.52)                                  | 890(17.80)   | 1292(25.84)  | 5000         |

대체 전후의 항목분포의 변화를 살펴보면 자신이 직접 한 경우는 실제 가구가 2523가구(50.46%)에서 2542가구(50.84%), 타인일부는 273가구(5.46%)에서 276가구(5.52%), 타인전부는 912가구(18.24%)에서 890가구(17.80%)로 변화되어 분포변화는 거의 발생하지 않는다고 볼 수 있다.

### 13) 벼베기탈곡방법

벼베기탈곡방법에 대한 무응답 대체는 콤바인, 모내기방법, 논갈이방법, 농약살포방법 항목을 대체군으로 사용하였으며, 논면적을 특별 대체군으로 추가하여 대체를 실시하였다. 또한 이후에 실제 대체적용을 할 때에는 콤바인 항목을 먼저 대체하고자 하므로 1차 대체군을 이용할 수 있을 것으로 판단된다. <표 3-61>의 결과를 보면 총 5,000가구를 대체한 결과 대체의 정확도는 90.40%이며, 허용결측을 제외한 농약살포방법의 순수 정확도는 86.87%로 높은 일치율을 보인다. 타인일부 항목에서 대체의 정확도가 많이 떨어지나 전체적으로 볼 때에는 그 수가 상당히 작기 때문에 오분류로 인한 효과는 미미하게 나타난다.

<표 3-61> 벼베기탈곡방법에 대한 대체군 및 대체결과

| 대체군(1차)  |              | 대체군(2차)                         |              |              |                |
|--|--------------|---------------------------------|--------------|--------------|----------------|
| 콤바인(100), 모내기방법(42)<br>논갈이방법(23), 농약살포방법(13)<br>논면적(특별대체군) |              | 트랙터(100), 이앙기(42)<br>논면적(특별대체군) |              |              |                |
| 대체결과_대체정확도 (90.40%), 허용결측 제외한 정확도(86.87%)                  |              |                                 |              |              |                |
|  | 자신직접<br>(대체) | 타인일부<br>(대체)                    | 타인전부<br>(대체) | 허용결측<br>(대체) | 합계<br>(구성비(%)) |
| 자신직접   | 385          | 24                              | 134          | 0            | 543(10.86)     |
| 타인일부   | 25           | 114                             | 87           | 0            | 226(4.52)      |
| 타인전부   | 137          | 73                              | 2677         | 0            | 2887(57.74)    |
| 허용결측   | 0            | 0                               | 0            | 1344         | 1344(26.88)    |
| 합계   | 547(10.94)   | 211(4.22)                       | 2898(57.96)  | 1344(26.88)  | 5000           |

대체 전후의 항목분포의 변화를 살펴보면 자신이 직접 한 경우는 실제 가구가 543가구(10.86%)에서 547가구(10.94%), 타인일부는 226가구(4.52%)에서 211가구(4.22%), 타인전부는 2887가구(57.74%)에서 2898가구(57.96%)로 나타나 대체로 인한 분포변화는 발생하지 않는다고 판단된다.

## 마. 농업경영에 관한 사항

### 1) 전·겸업수입구분

전·겸업수입구분에 대한 무응답 대체는 가구원수합계, 판매금액구분, 농업특성조사구, 승용차보유여부, 농업종사경력 항목을 대체군으로 사용하였다. 그리고 가구원수합계는 일반적으로 알 수 있는 정보이므로 2차 대체군은 필요하지 않을 것으로 판단된다. 총 5,000가구를 대체한 결과 <표 3-62>에서 알 수 있듯이 대체의 정확도는 66.08%로 나타났다. 다른 항목들에 비해서는 다소 정확도가 떨어지는 것을 볼 수 있다. 가장 궁극적인 이유는 전·겸업수입구분은 농업이외종사기간 항목과 매우 밀접한 관계가 있는데 이 항목을 사용하지 못하기 때문에 일어난다고 볼 수 있다. 따라서 가구부문에 농업이외의 일을 하는 가구원수 항목을 추가한다면 지금보다는 더 정확한 대체가 이루어질 것으로 확신한다. 하지만 현재로서는 대체의 정확도를 더 높이기에는 한계가 있을 것으로 판단된다.

<표 3-62> 전·겸업수입구분에 대한 대체군 및 대체결과

| 대체군(1차)   |                | 대체군(2차)                               |                 |             |
|---|----------------|---------------------------------------|-----------------|-------------|
| 가구원수합계(100), 판매금액구분(79)<br>농업특성조사구(64), 승용차보유(32)<br>농업종사경력(14) |                | 승용차보유(100), 판매금액구분(75)<br>농업특성조사구(62) |                 |             |
| 대체결과_대체정확도 (66.08%)   |                |                                       |                 |             |
|   | 농업수입뿐임<br>(대체) | 농업수입많음<br>(대체)                        | 농업외수입많음<br>(대체) | 합계(구성비(%))  |
| 농업수입뿐임  | 2377           | 324                                   | 410             | 3111(62.22) |
| 농업수입많음  | 330            | 194                                   | 123             | 647(12.94)  |
| 농업외수입많음   | 394            | 115                                   | 733             | 1242(24.84) |
| 합계  | 3101(62.02)    | 633(12.66)                            | 1266(25.32)     | 5000        |

대체 전후의 항목분포의 변화를 살펴보면 농업수입뿐임은 실제 가구가 3111가구(62.22%)에서 3101가구(62.02%), 농업수입많음은 647가구

(12.94%)에서 633가구(12.66%), 농업외수입많은은 1242가구(24.84%)에서 1266가구(25.32%)로 변화되어 대체의 정확도와는 달리 대체로 인한 분포변화의 문제는 전혀 발생되지 않음을 알 수 있다.

## 2) 영농형태

영농형태에 대한 연관성 분석의 결과 논면적, 밭면적, 과수활동여부, 화훼활동여부, 농업특성조사구, 판매처구분 항목을 대체군으로 사용하게 되나 이 경우 영농형태상세 항목을 특별 대체군으로 추가하여 대체를 실시하였다. 그 이유는 특정 영농형태 항목만 영농형태상세 항목을 조사하므로 이러한 규칙에 맞는 대체가 이루어져야 하기 때문이다.

<표 3-63> 영농형태에 대한 대체군 및 대체결과

| 대체군(1차)  |                 | 대체군(2차)   |                  |                |              |                   |               |                  |              |
|--|-----------------|---|------------------|----------------|--------------|-------------------|---------------|------------------|--------------|
| 논면적(100), 밭면적(45)<br>과수활동(44), 화훼활동(39)<br>농업특성조사구(32), 판매처구분(22)<br>영농형태상세(특별대체군) |                 | 논갈이방법(100)<br>농업특성조사구(85), 밭면적(25)<br>영농형태상세(특별대체군) |                  |                |              |                   |               |                  |              |
| <b>대체결과_대체정확도 (87.68%)</b>   |                 |   |                  |                |              |                   |               |                  |              |
|  | 논벼<br>(대체)      | 과수<br>(대체)  | 특용<br>작물<br>(대체) | 채소<br>(대체)     | 화훼<br>(대체)   | 일반<br>밭작물<br>(대체) | 축산<br>(대체)    | 양잠<br>기타<br>(대체) | 합계<br>(구성비%) |
| 논벼   | 2441            | 0   | 23               | 0              | 2            | 62                | 0             | 3                | 2531(50.62)  |
| 과수   | 0               | 492   | 0                | 49             | 0            | 0                 | 22            | 0                | 563(11.26)   |
| 특용작물   | 36              | 0   | 34               | 0              | 4            | 40                | 0             | 1                | 115(2.30)    |
| 채소   | 0               | 53  | 0                | 817            | 0            | 0                 | 59            | 0                | 929(18.58)   |
| 화훼   | 4               | 0   | 3                | 0              | 12           | 11                | 0             | 1                | 31(0.62)     |
| 밭작물  | 75              | 0   | 43               | 0              | 5            | 367               | 0             | 12               | 502(10.04)   |
| 축산   | 1               | 38  | 0                | 57             | 0            | 0                 | 219           | 0                | 315(6.30)    |
| 양잠기타   | 8               | 0   | 2                | 0              | 1            | 1                 | 0             | 2                | 14(0.28)     |
| 합계   | 2565<br>(51.30) | 583<br>(11.66)                                      | 105<br>(2.10)    | 923<br>(18.46) | 24<br>(0.48) | 481<br>(9.62)     | 300<br>(6.00) | 19<br>(0.38)     | 5000         |

따라서 영농형태상세를 기입한 경우는 과수, 채소, 축산 중에서 대체가 되어야 한다. 또한 이후에 실제 대체적용을 할 때에는 논면적이 먼저 대체되므로 1차 대체군만을 이용할 수 있을 것으로 판단된다. <표 3-63>에서와 같이 총 5,000가구를 대체한 결과 대체의 정확도는 87.68%로 높은 정확도를 보였다. 구성비가 큰 논벼, 과수, 채소는 상당히 정확한 대체가 되고 있으며, 특용작물과 화훼는 대체의 정확도가 다소 떨어지나 전체적으로 볼 때에는 그 수가 상당히 작기 때문에 오분류로 인한 효과는 미미하게 나타난다고 할 수 있다.

대체 전후의 항목분포의 변화를 살펴보면 논벼는 실제 가구가 2531가구(50.62%)에서 2565가구(51.30%), 과수는 563가구(11.26%)에서 583가구(11.66%), 채소는 929가구(18.58%)에서 923가구(18.46%), 축산은 315가구(6.30%)에서 300가구(6.00%) 등으로 나타나 대체로 인한 분포변화는 발생하지 않는다고 판단된다.

### 3) 판매금액구분

판매금액구분 대한 무응답 대체는 판매처구분, 논면적, 트랙터, 발면적, 영농형태, 영농형태상세 항목을 대체군으로 사용하였으며, 판매처구분 정보를 사용할 수 없을 경우 추가되는 항목이 없는 2차 대체군을 적용하였다. 판매금액은 연속형의 항목이지만 현재 조사에서는 범주화시켜서 해당 금액에 기입을 하고 있다. 따라서 이 항목의 특성상 완전한 범주형 자료처럼 대체의 정확도를 측정한다는 것은 적절하지 못하다. 예를 들면 실제 판매금액이 1000만원인데 999만원으로 대체가 되었다면 상당히 정확한 대체라고 할 수 있다. 그러나 이것을 범주화하여 값을 넣으면 다른 범주에 속하기 때문에 잘못 추정된 결과가 된다. 이런 이유로 본 항목에서는 대체의 정확도는 큰 의미가 없으므로 대체전후의 분포변화에 대해서만 살펴보기로 한다.

대체 전후의 항목분포의 변화가 <표 3-64>에 주어져있다. 판매금액이 없는 가구는 실제 가구가 467가구(9.34%)에서 469가구(9.38%), 50만원 미만은 288가구(5.76%)에서 314가구(6.28%), 100-200만원은 500가구(10.00%)에서 505가구(10.10%), 1000-2000만원은 680가구(13.60%)에서

643가구(12.86%), 1-2억은 41가구(0.82%)에서 35가구(0.70%) 등으로 변화되어 대체로 인한 분포변화의 문제는 전혀 발생되지 않음을 알 수 있다.

<표 3-64> 판매금액구분에 대한 대체군 및 대체전후의 분포변화

| 대체군(1차)   |               |               |               |                | 대체군(2차)  |                |                |               |               |               |              |              |
|---|---------------|---------------|---------------|----------------|--|----------------|----------------|---------------|---------------|---------------|--------------|--------------|
| 판매처구분(100), 논면적(56)<br>트랙터(40), 밭면적(39)<br>영농형태(21), 영농형태상세(20) |               |               |               |                | 논면적(100), 밭면적(87)<br>트랙터(51), 영농형태(36)<br>영농형태상세(25) |                |                |               |               |               |              |              |
| 판매금액<br>구분  | 판매<br>없음      | 50만<br>미만     | 50-<br>100    | 100-<br>200    | 200-<br>500  | 500-<br>1000   | 1000-<br>2000  | 2000-<br>3000 | 3000-<br>5000 | 5000-<br>1억   | 1억-<br>2억    | 2억<br>이상     |
| 실제가구<br>(구성비(%))  | 467<br>(9.34) | 288<br>(5.76) | 403<br>(8.06) | 500<br>(10.00) | 924<br>(18.48)                                       | 840<br>(16.80) | 680<br>(13.60) | 395<br>(7.90) | 287<br>(5.74) | 155<br>(3.10) | 41<br>(0.82) | 20<br>(0.40) |
| 대체후가구<br>(구성비(%))   | 469<br>(9.38) | 314<br>(6.28) | 414<br>(8.28) | 505<br>(10.10) | 909<br>(18.18)                                       | 869<br>(17.38) | 643<br>(12.86) | 401<br>(8.02) | 272<br>(5.44) | 154<br>(3.08) | 35<br>(0.70) | 15<br>(0.30) |

#### 4) 판매처구분

판매처구분에 대한 연관성 분석의 결과 판매금액구분, 영농형태, 시도, 영농형태상세, 직판장직거래 항목을 대체군으로 사용하게 되나 이 경우 친환경작물판매처 항목을 특별 대체군으로 추가하여 대체를 실시하였다. 그 이유는 두 항목에서 일치하는 세부 항목들이 있어 이를 이용하면 조금이라도 정확도를 높일 수 있을 것이라고 판단되었기 때문이다. 또한 이후에 실제 대체적용을 할 때에는 판매금액구분이 먼저 대체되므로 1차 대체군만을 이용할 수 있을 것으로 생각된다. <표 3-65>에서와 같이 총 5,000가구를 대체한 결과 대체의 정확도는 43.68%로 다른 항목들에 비해서 매우 낮은 정확도를 보였다. 정확도가 떨어지는 가장 큰 원인은 판매처구분을 세밀하게 분리할 수 있는 보조변수가 부족하며, 실제 가구마다 판매하는 데 있어서의 패턴이 다른 항목들에 비해서 뚜렷하지 못한 것으로 보인다. 예를 들면 판매금액이나 영농형태가 같은 가구들도 판매하는 장소가 각기 다른 경우가 많이 나타나고 있다. <표 3-65>의 대체 결과를 보더라도 실제로 판매하는 장소가 다양하다는 것을 알 수 있다. 그러므로 이 항목은 대체의 정확도를 높이기 위해서 보조 항목을 추가하는 것보다는 세부 항목의 범주를 더 간소화(그룹화)



하는 것이 더 효율적이지 않을까 하는 생각이 든다.

대체 전후의 항목분포의 변화를 살펴보면 도매시장에서의 판매는 실제 가구가 286가구(5.72%)에서 292가구(5.84%), 농협·농업법인은 1221가구(24.42%)에서 1194가구(23.88%), 개인소비자는 883가구(17.66%)에서 880가구(17.60%), 재래시장은 162가구(3.24%)에서 150가구(3.00%) 등으로 변화되어 대체의 정확도와는 달리 대체로 인한 분포변화의 문제는 거의 발생되지 않음을 알 수 있다.

<표 3-65> 판매처구분에 대한 대체군 및 대체결과

| 대체군(1차)  |               | 대체군(2차)  |                 |                |                |                |              |             |               |               |               |              |
|--|---------------|--|-----------------|----------------|----------------|----------------|--------------|-------------|---------------|---------------|---------------|--------------|
| 판매금액구분(100), 영농형태(53)<br>시도(17), 영농형태상세(16)<br>직판장직거래(10)<br>친환경작물판매처(특별대체군) |               | 영농형태(100)<br>논면적(84), 경운기(55)<br>친환경작물판매처(특별대체군) |                 |                |                |                |              |             |               |               |               |              |
| 대체결과_대체정확도 (43.68%)  |               |  |                 |                |                |                |              |             |               |               |               |              |
|  | 1<br>(대체)     | 2<br>(대체)  | 3<br>(대체)       | 4<br>(대체)      | 5<br>(대체)      | 6<br>(대체)      | 7<br>(대체)    | 8<br>(대체)   | 9<br>(대체)     | 10<br>(대체)    | 11<br>(대체)    | 합계<br>(구성비%) |
| 1  | 93            | 28   | 60              | 13             | 49             | 29             | 1            | 0           | 8             | 5             | 0             | 286(5.72)    |
| 2  | 35            | 51   | 38              | 5              | 29             | 16             | 1            | 0           | 3             | 2             | 0             | 180(3.60)    |
| 3  | 32            | 45   | 520             | 217            | 164            | 139            | 0            | 2           | 73            | 29            | 0             | 1221(24.42)  |
| 4  | 15            | 11   | 185             | 287            | 57             | 90             | 1            | 1           | 84            | 11            | 1             | 743(14.86)   |
| 5  | 57            | 23   | 148             | 64             | 189            | 109            | 4            | 1           | 49            | 26            | 0             | 670(13.40)   |
| 6  | 31            | 17   | 140             | 77             | 108            | 419            | 2            | 1           | 44            | 43            | 1             | 883(17.66)   |
| 7  | 0             | 0  | 2               | 2              | 7              | 1              | 5            | 0           | 1             | 0             | 0             | 18(0.36)     |
| 8  | 1             | 0  | 1               | 1              | 0              | 1              | 0            | 0           | 0             | 0             | 0             | 4(0.08)      |
| 9  | 17            | 5  | 77              | 76             | 36             | 19             | 2            | 3           | 102           | 4             | 1             | 342(6.84)    |
| 10   | 11            | 6  | 23              | 10             | 20             | 56             | 1            | 0           | 5             | 30            | 0             | 162(3.24)    |
| 11   | 0             | 0  | 0               | 0              | 1              | 1              | 0            | 0           | 1             | 0             | 488           | 491(9.82)    |
| 합계   | 292<br>(5.84) | 186<br>(3.72)                                    | 1194<br>(23.88) | 752<br>(15.04) | 660<br>(13.20) | 880<br>(17.60) | 17<br>(0.34) | 8<br>(0.16) | 370<br>(7.40) | 150<br>(3.00) | 491<br>(9.82) | 5000         |

\* 1(도매시장), 2(산지공판장), 3(농협·농업법인), 4(정부예판매), 5(수집상), 6(개인소비자)  
7(대형유통업체), 8(대량수요처), 9(농축산물 가공업체), 10(재래시장), 11(기타, 판매없음)

## 바. 친환경농업에 관한 사항

### 1) 친환경수확작물유무

친환경수확작물유무에 대한 연관성 분석의 결과 친환경작물판매처 항목만이 연관이 있어 이를 대체군으로 사용하였으며, 친환경작물판매처의 정보를 사용할 수 없을 경우에는 다른 연관성 항목이 없으므로 대체군 없이 무작위 대체를 할 수밖에 없을 것으로 보인다. 하지만 지역(시도, 시군구, 읍면동)에 관한 정보를 사용해도 큰 효과는 없을 것으로 판단된다. <표 3-66>에서와 같이 총 5,000가구를 대체한 결과 대체의 정확도는 99.12%로 다른 항목들에 비해서 매우 높은 정확도를 보였다. 대체결과를 살펴보면 친환경작물이 있는 경우는 335가구 중에서 319가구, 없는 경우는 4665가구 중에서 4637가구를 정확하게 찾아내었다. 이렇게 대체의 정확도가 높은 이유는 친환경작물판매처의 정보가 친환경수확작물유무를 결정하는 열쇠와도 같기 때문이다. 그러나 친환경작물판매처의 정보가 없을 경우에는 대체의 정확도는 현저하게 떨어질 것으로 보인다. 따라서 결국 대체의 정확도는 주어진 보조정보(대체군)에 의하여 좌우된다고 해도 과언은 아닐 것이다.

<표 3-66> 친환경수확작물유무에 대한 대체군 및 대체결과

| 대체군(1차)             |           | 대체군(2차)     |             |
|---------------------|-----------|-------------|-------------|
| 친환경작물판매처(100)       |           | 없음          |             |
| 대체결과_대체정확도 (99.12%) |           |             |             |
|                     | 유(대체)     | 무(대체)       | 합계(구성비(%))  |
| 유                   | 319       | 16          | 335(6.70)   |
| 무                   | 28        | 4637        | 4665(93.30) |
| 합계                  | 347(6.94) | 4653(93.06) | 5000        |

대체 전후의 항목분포의 변화를 살펴보면 친환경작물이 있는 경우는 실제 가구가 335가구(6.70%)에서 347가구(6.94%), 없는 경우는 4665가구(93.30%)에서 4653가구(93.06%)로 나타나 대체로 인한 분포변화는 발생하지 않는다고 판단된다.

## 2) 친환경작물판매처

친환경작물판매처에 대한 연관성 분석의 결과 판매처구분 항목만을 대체군으로 사용하게 되나 이 경우 친환경수확작물유무 항목을 특별 대체군으로 추가하여 대체를 실시하였다. 그 이유는 친환경작물판매처 항목은 친환경작물을 수확한 가구만을 대상으로 기입하므로 이에 맞는 대체가 이루어져야 하기 때문이다. 따라서 친환경수확작물이 없는 경우는 허용결측으로, 있는 경우는 적절한 대체값으로 대체가 되어야 한다. 그러므로 친환경수확작물유무를 대체군에 추가함으로써 대체 후 내검규칙에 합당하게 될 것이라 판단된다. 또한 이후에 실제 대체적용을 할 때에는 판매처구분이 먼저 대체되므로 1차 대체군만을 이용할 수 있을 것으로 생각된다.

〈표 3-67〉 친환경작물판매처에 대한 대체군 및 대체결과

| 대체군(1차)                                    |               | 대체군(2차)  |                |              |               |               |              |               |                  |                |
|--|---------------|--|----------------|--------------|---------------|---------------|--------------|---------------|------------------|----------------|
| 판매처구분(100)<br>친환경수확작물유무(특별대체군)             |               | 영농형태(100), 논면적(64)<br>판매금액구분(46), 시도(36)<br>직판장직거래(35)<br>친환경수확작물유무(특별대체군) |                |              |               |               |              |               |                  |                |
| 대체결과_대체정확도 (97.48%), 허용결측 제외한 정확도 (62.49%) |               |  |                |              |               |               |              |               |                  |                |
|  | 1<br>(대체)     | 2<br>(대체)  | 3<br>(대체)      | 4<br>(대체)    | 5<br>(대체)     | 6<br>(대체)     | 7<br>(대체)    | 8<br>(대체)     | 9<br>(대체)        | 합계<br>(구성비(%)) |
| 1  | 228           | 7  | 51             | 4            | 27            | 16            | 3            | 16            | 0                | 352(0.70)      |
| 2  | 8             | 114  | 48             | 4            | 8             | 13            | 1            | 8             | 0                | 204(0.40)      |
| 3  | 77            | 43   | 1192           | 21           | 70            | 128           | 4            | 57            | 0                | 1592(3.18)     |
| 4  | 6             | 2  | 23             | 16           | 9             | 16            | 2            | 4             | 0                | 78(0.15)       |
| 5  | 13            | 8  | 72             | 6            | 68            | 26            | 12           | 14            | 0                | 219(0.43)      |
| 6  | 34            | 16   | 122            | 18           | 22            | 378           | 6            | 40            | 0                | 636(1.27)      |
| 7  | 4             | 2  | 8              | 4            | 12            | 4             | 27           | 2             | 0                | 63(0.12)       |
| 8  | 9             | 8  | 62             | 6            | 16            | 34            | 3            | 75            | 0                | 213(0.42)      |
| 9  | 0             | 0  | 0              | 0            | 0             | 0             | 0            | 0             | 46643            | 46643(93.28)   |
| 합계   | 379<br>(0.75) | 200<br>(0.40)  | 1578<br>(3.15) | 79<br>(0.15) | 232<br>(0.46) | 615<br>(1.23) | 58<br>(0.11) | 216<br>(0.43) | 46643<br>(93.28) | 50000          |

\* 1(도매시장), 2(산지공판장), 3(농협·농업법인), 4(소비자단체), 5(친환경농산물 전문유통업체), 6(개인소비자), 7(대형유통업체), 8(기타), 9(허용결측)

<표 3-67>에서와 같이 총 50,000가구를 대체한 결과 허용결측을 제외한 대체의 정확도는 62.49%로 다른 항목들에 비해서는 다소 낮음을 알 수 있다. 정확도가 떨어지는 가장 큰 원인은 판매처구분 항목과 유사하게 친환경작물판매처 구분을 세밀하게 분리할 수 있는 보조변수가 부족하며, 실제 가구마다 판매하는 데 있어서의 패턴이 다른 항목들에 비해서 뚜렷하지 못하기 때문으로 보인다.

대체 전후의 항목분포의 변화를 살펴보면 도매시장에서의 판매는 실제 가구가 352가구(0.70%)에서 379가구(0.75%), 농협·농업법인은 1592가구(3.18%)에서 1578가구(3.15%), 개인소비자는 636가구(1.27%)에서 615가구(1.23%) 등으로 변화되어 대체의 정확도와는 달리 대체로 인한 분포변화의 문제는 거의 발생되지 않음을 알 수 있다.

## 사. 정보화현황에 관한 사항

### 1) 컴퓨터보유여부

컴퓨터보유여부 대한 연관성 분석의 결과 컴퓨터활용여부 항목만이 연관이 있어 이를 대체군으로 사용하였으며, 친환경작물판매처의 정보를 사용할 수 없을 경우에는 2차 대체군을 적용하였다.

<표 3-68> 컴퓨터보유여부에 대한 대체군 및 대체결과

| 대체군(1차)             |             | 대체군(2차)                                  |             |
|---------------------|-------------|--|-------------|
| 컴퓨터활용여부(100)        |             | 가구원수합계(100)<br>승용차보유여부(35)<br>농업종사경력(23) |             |
| 대체결과_대체정확도 (97.90%) |             |  |             |
|                     | 보유(대체)      | 미보유(대체)                                  | 합계(구성비(%))  |
| 보유                  | 1742        | 44                                       | 1786(35.72) |
| 미보유                 | 61          | 3153                                     | 3214(64.28) |
| 합계                  | 1803(36.06) | 3197(63.94)                              | 5000        |

<표 3-68>에서와 같이 총 5,000가구를 대체한 결과 대체의 정확도는

97.90%로 매우 높은 정확도를 보였다. 대체결과를 살펴보면 컴퓨터를 보유한 경우는 1786가구 중에서 1742가구, 보유하지 않은 경우는 3214가구 중에서 3153가구를 정확하게 찾아내었다. 이렇게 대체의 정확도가 높은 이유는 컴퓨터활용여부의 정보가 컴퓨터보유여부를 결정하는 중요 항목이기 때문이다. 그러나 컴퓨터활용여부의 정보가 없을 경우에는 대체의 정확도는 지금보다 많이 떨어질 것으로 보인다. 그러므로 대체의 정확도는 주어진 보조정보와 밀접한 관계가 있음을 알 수 있다.

대체 전후의 항목분포의 변화를 살펴보면 컴퓨터를 보유하고 있는 경우는 실제 가구가 1786가구(35.72%)에서 1803가구(36.06%), 보유하지 않은 경우는 3214가구(64.28%)에서 3197가구(63.94%)로 나타나 대체로 인한 분포변화는 발생되지 않는다고 판단된다.

## 2) 컴퓨터활용여부

컴퓨터활용여부에 대한 연관성 분석의 결과 홈페이지개설 항목만을 대체군으로 사용하게 되나 이 경우 컴퓨터보유여부 항목을 특별 대체군으로 추가하여 대체를 실시하였다. 그 이유는 컴퓨터활용여부 항목은 컴퓨터를 보유한 가구만을 대상으로 기입하므로 이에 맞는 대체가 이루어져야 하기 때문이다. 따라서 컴퓨터가 없는 가구는 허용결측으로, 있는 가구는 적절한 대체값으로 대체가 되어야 한다. 그러므로 컴퓨터보유여부를 대체군에 추가함으로써 대체 후 내검규칙에 합당하게 될 것이라 판단된다. 그리고 홈페이지개설 정보를 사용할 수 없을 경우에는 2차 대체군을 적용하였다.

<표 3-69>에서와 같이 총 10,000가구를 대체한 결과 대체의 정확도는 99.90%이며 허용결측을 제외한 대체의 정확도도 99.73%로 상당히 높게 나타나고 있다. 대체결과를 살펴보면 대부분의 가구가 정확하게 대체된 것을 볼 수 있다. 이는 홈페이지가 개설이 되었다는 것이 곧 컴퓨터를 활용하고 있다고 것을 의미한다고 볼 수 있을 것이다. 그러나 홈페이지개설에 관한 정보가 없을 경우에는 대체의 정확도가 현저하게 떨어질 수도 있다는 것도 생각해야 할 것이다.

〈표 3-69〉 컴퓨터활용여부에 대한 대체군 및 대체결과

| 대체군(1차)                                    |             | 대체군(2차)   |             |             |
|--|-------------|---|-------------|-------------|
| 홈페이지개설(100)<br>컴퓨터보유여부(특별대체군)              |             | 농업정보수집활용(100)<br>기타농업활용(36)<br>농축산물판매활용(18)<br>컴퓨터보유여부(특별대체군) |             |             |
| 대체결과_대체정확도 (99.90%), 허용결측 제외한 정확도 (99.73%) |             |   |             |             |
|  | 활용(대체)      | 미활용(대체)   | 허용결측(대체)    | 합계(구성비(%))  |
| 활용   | 1062        | 3   | 0           | 1065(10.65) |
| 미활용  | 7           | 2686  | 0           | 2693(26.93) |
| 허용결측                                       | 0           | 0   | 6242        | 6242(62.42) |
| 합계   | 1069(10.69) | 2689(26.89)   | 6242(62.42) | 10000       |

대체 전후의 항목분포의 변화를 살펴보면 컴퓨터를 활용하고 있는 경우는 실제 가구가 1065가구(10.65%)에서 1069가구(10.69%), 미활용의 경우는 2693가구(26.93%)에서 2689가구(26.89%)로 나타나 대체로 인한 분포변화는 거의 없는 것으로 보인다.

### 3) 홈페이지개설

홈페이지개설에 대한 무응답 대체는 농축산물판매활용, 친환경수확작물유무, 시군구 항목을 대체군으로 사용하였으며, 컴퓨터활용여부와 같은 이유로 컴퓨터보유여부와 활용여부를 특별 대체군으로 추가하여 대체를 실시하였다. 그리고 농축산물판매활용 정보는 주어진 정보를 그대로 사용하고자 하므로 1차 대체군을 이용할 수 있을 것으로 생각된다.

<표 3-70>에서와 같이 총 20,000가구를 대체한 결과 대체의 정확도는 99.68%이며 허용결측을 제외한 대체의 정확도도 96.93%로 상당히 높게 나타나고 있다. 대체결과를 살펴보면 허용결측을 제외하면 대부분의 가구가 미개설 상태이며 이는 매우 정확하게 대체가 되고 있음을 알 수 있다. 하지만 개설에 대한 대체는 절반 정도를 맞추고 있는데 구성비가 작아서 일어나는 현상으로 정확도의 손실은 받아들여야 할 것이다.

〈표 3-70〉 홈페이지개설여부에 대한 대체군 및 대체결과

| 대체군(1차)   |          | 대체군(2차)                          |              |              |
|---|----------|----------------------------------|--------------|--------------|
| 농축산물판매활용(100)<br>친환경수확작물유무(13)<br>시군구(10)<br>컴퓨터보유여부(특별대체군)<br>컴퓨터활용여부(특별대체군) |          | 컴퓨터보유여부(특별대체군)<br>컴퓨터활용여부(특별대체군) |              |              |
| 대체결과_대체정확도 (99.68%), 허용결측 제외한 정확도 (96.93%)                                    |          |                                  |              |              |
|   | 개설(대체)   | 미개설(대체)                          | 허용결측(대체)     | 합계(구성비(%))   |
| 개설  | 34       | 34                               | 0            | 68(0.34)     |
| 미개설   | 30       | 1990                             | 0            | 2020(10.10)  |
| 허용결측  | 0        | 0                                | 17912        | 17912(89.56) |
| 합계  | 64(0.32) | 2024(10.12)                      | 17912(89.56) | 20000        |

대체 전후의 항목분포의 변화를 살펴보면 홈페이지를 개설한 가구는 실제 68가구(0.34%)에서 64가구(0.32%), 개설하지 않은 경우는 2020가구(10.10%)에서 2024가구(10.12%)로 나타나 대체로 인한 분포변화는 발생되지 않는다고 판단된다.

## 아. 생활여건에 관한 사항

### 1) 동사무소이동수단

동사무소이동수단에 대한 무응답 대체는 금융기관이동수단, 병원이동수단 항목을 대체군으로 사용하였으며, 금융기관이동수단 정보를 사용할 수 없을 경우에는 2차 대체군을 적용하였다. <표 3-71>에서와 같이 총 5,000가구를 대체한 결과 대체의 정확도는 91.34%로 매우 높은 정확도를 보였다. 대체결과를 살펴보면 걸어서 이동하는 경우는 1092가구 중에서 909가구, 자동차로 이동하는 경우는 3092가구 중에서 2920가구를 정확하게 찾아내었다. 일반적으로 대체의 정확도가 높은 이유는 동사무소, 금융기관, 병원의 위치가 비슷해서 이동수단이 같은 가구가 많기 때문인 것으로 판단된다.

〈표 3-71〉 동사무소이동수단에 대한 대체군 및 대체결과

| 대체군(1차)                     |             | 대체군(2차)                  |            |             |
|-----------------------------|-------------|--------------------------|------------|-------------|
| 금융기관이동수단(100)<br>병원이동수단(17) |             | 수도형태(100)<br>농업특성조사구(67) |            |             |
| 대체결과_대체정확도 (91.34%)         |             |                          |            |             |
|                             | 걸어서(대체)     | 자동차(대체)                  | 기타(대체)     | 합계(구성비(%))  |
| 걸어서                         | 909         | 149                      | 34         | 1092(21.84) |
| 자동차                         | 142         | 2920                     | 30         | 3092(61.84) |
| 기타                          | 36          | 42                       | 738        | 816(16.32)  |
| 합계                          | 1078(21.56) | 3111(62.22)              | 802(16.04) | 5000        |

대체 전후의 항목분포는 걸어서 이동하는 경우 실제 가구가 1092가구(21.84%)에서 1078가구(21.56%), 자동차로 이동하는 경우는 3092가구(61.84%)에서 3111가구(62.22%), 기타는 816가구(16.32%)에서 802가구(16.04%)로 변화되어 대체전후의 분포변화는 일어나지 않음을 알 수 있다.

## 2) 동사무소소요시간

동사무소소요시간에 대한 무응답 대체는 금융기관소요시간 하나의 항목만을 대체군으로 사용하였으며, 금융기관소요시간 정보를 사용할 수 없을 경우에는 2차 대체군을 적용하였다. 금융기관소요시간은 동사무소소요시간과 상관관계가 매우 높기 때문에 좋은 대체가 되지만 이 정보를 알 수 없을 경우에는 대체의 정확도는 많이 낮아질 수도 있을 것이다. 임의로 발생시킨 5000가구의 표본평균과 대체한 후의 표본평균의 값은 <표 3-72>에 주어져 있다. 5번 반복된 실험결과를 보면 실제표본평균과 대체한 후의 표본평균의 차이는 0.04분에서 0.13분 정도로 나왔으며, 5번의 평균은 0.088분으로 대체 후에도 동사무소소요시간은 차이가 거의 나지 않음을 알 수 있다.



〈표 3-72〉 동사무소소요시간에 대한 대체군 및 표본평균 차이정도

| 대체군(1차)                |        | 대체군(2차)                           |      |
|------------------------|--------|-----------------------------------|------|
| 금융기관소요시간(100)          |        | 승용차보유(100)<br>화물차보유(92), 수도형태(58) |      |
| 표본평균 절대차이의 평균 (0.088분) |        |                                   |      |
| 실험                     | 실제표본평균 | 대체후 표본평균                          | 절대차이 |
| 1                      | 13.98  | 14.11                             | 0.13 |
| 2                      | 14.10  | 14.01                             | 0.09 |
| 3                      | 14.05  | 14.12                             | 0.07 |
| 4                      | 13.95  | 13.99                             | 0.04 |
| 5                      | 14.02  | 14.13                             | 0.11 |

동사무소소요시간에 대한 분포변화를 고려하기 위하여 범주화하여 확인한 결과가 <표 3-73>에 주어져 있다. 0-4분은 176가구(3.52%)에서 174가구(3.48%), 5-9분은 1146가구(22.92%)에서 1138가구(22.76%), 10-14분은 1584가구(31.68%)에서 1590가구(31.80%), 25-29분은 88가구(1.76%)에서 96가구(1.92%), 40-44분은 92가구(1.84%)에서 106가구(2.12%), 45분 이상은 76가구(1.52%)에서 79가구(1.58%) 등으로 변화되나 대체전후의 분포변화 문제는 일어나지 않는다고 판단된다.

〈표 3-73〉 동사무소소요시간에 대한 대체전후의 분포변화

| 소요시간              | 0-4           | 5-9             | 10-14           | 15-19          | 20-24          | 25-29        | 30-34         | 35-39        | 40-44         | 45-          |
|-------------------|---------------|-----------------|-----------------|----------------|----------------|--------------|---------------|--------------|---------------|--------------|
| 실제가구<br>(구성비(%))  | 176<br>(3.52) | 1146<br>(22.92) | 1584<br>(31.68) | 679<br>(13.58) | 783<br>(15.66) | 88<br>(1.76) | 366<br>(7.32) | 10<br>(0.20) | 92<br>(1.84)  | 76<br>(1.52) |
| 대체후가구<br>(구성비(%)) | 174<br>(3.48) | 1138<br>(22.76) | 1590<br>(31.80) | 690<br>(13.80) | 758<br>(15.16) | 96<br>(1.92) | 362<br>(7.24) | 7<br>(0.14)  | 106<br>(2.12) | 79<br>(1.58) |

### 3) 금융기관이동수단

금융기관이동수단에 대한 무응답 대체는 동사무소이동수단, 병원이동수단 항목을 대체군으로 사용하였으며, 이후에 실제 대체적용을 할

때에는 동사무소이동수단이 먼저 대체되므로 1차 대체군을 이용할 수 있을 것으로 생각된다. <표 3-74>에서와 같이 총 5,000가구를 대체한 결과 대체의 정확도는 93.86%로 상당히 높은 정확도를 보였다. 대체결과를 살펴보면 걸어서 이동하는 경우는 1195가구 중에서 1047가구, 자동차로 이동하는 경우는 3027가구 중에서 2891가구를 정확하게 찾아내었다. 대체 전후의 항목분포는 걸어서 이동하는 경우 실제 가구가 1195가구(23.90%)에서 1186가구(23.72%), 자동차로 이동하는 경우는 3027가구(60.54%)에서 3022가구(60.44%), 기타는 778가구(15.56%)에서 792가구(15.84%)로 대체전후의 분포변화는 일어나지 않음을 알 수 있다.

<표 3-74> 금융기관이동수단에 대한 대체군 및 대체결과

| 대체군(1차)                     |             | 대체군(2차)                  |            |             |
|-----------------------------|-------------|--------------------------|------------|-------------|
| 동사무소이동수단(100)<br>병원이동수단(18) |             | 수도형태(100)<br>농업특성조사구(63) |            |             |
| 대체결과_대체정확도 (93.86%)         |             |                          |            |             |
|                             | 걸어서(대체)     | 자동차(대체)                  | 기타(대체)     | 합계(구성비(%))  |
| 걸어서                         | 1047        | 122                      | 26         | 1195(23.90) |
| 자동차                         | 125         | 2891                     | 11         | 3027(60.54) |
| 기타                          | 14          | 9                        | 755        | 778(15.56)  |
| 합계                          | 1186(23.72) | 3022(60.44)              | 792(15.84) | 5000        |

#### 4) 금융기관소요시간

금융기관소요시간에 대한 무응답 대체는 동사무소소요시간, 병원소요시간 항목을 대체군으로 사용하였으며, 동사무소소요시간 항목이 먼저 대체되므로 1차 대체군만을 이용할 수 있을 것이다. 임의로 발생시킨 5000가구의 표본평균과 대체한 후의 표본평균의 값은 <표 3-75>에 주어져 있다. 5번 반복된 실험결과를 보면 실제표본평균과 대체한 후의 표본평균의 차이는 0.04분에서 0.15분 정도로 나왔으며, 5번의 평균은 0.086분으로 대체 후에도 금융기관소요시간은 차이가 거의 나지 않음을 알 수 있다.

〈표 3-75〉 금융기관소요시간에 대한 대체군 및 표본평균 차이정도

| 대체군(1차)                     |        | 대체군(2차)                           |      |
|-----------------------------|--------|-----------------------------------|------|
| 동사무소소요시간(100)<br>병원소요시간(23) |        | 승용차보유(100)<br>화물차보유(92), 수도형태(63) |      |
| 표본평균 절대차이의 평균 (0.086분)      |        |                                   |      |
| 실험                          | 실제표본평균 | 대체후 표본평균                          | 절대차이 |
| 1                           | 13.58  | 13.64                             | 0.06 |
| 2                           | 13.96  | 14.11                             | 0.15 |
| 3                           | 14.07  | 13.96                             | 0.11 |
| 4                           | 13.86  | 13.90                             | 0.04 |
| 5                           | 13.79  | 13.86                             | 0.07 |

금융기관소요시간에 대한 분포변화를 고려하기 위하여 범주화하여 확인한 결과가 <표 3-76>에 주어져 있다. 0-4분은 177가구(3.54%)에서 182가구(3.64%), 5-9분은 1154가구(23.08%)에서 1136가구(22.72%), 10-14분은 1623가구(32.46%)에서 1632가구(32.64%), 40-44분은 97가구(1.94%)에서 112가구(2.24%), 45분 이상은 78가구(1.56%)에서 89가구(1.78%) 등으로 변화되나 대체전후의 분포변화는 일어나지 않는다고 판단된다.

〈표 3-76〉 금융기관소요시간에 대한 대체전후의 분포변화

| 소요시간            | 0-4           | 5-9             | 10-14           | 15-19          | 20-24          | 25-29        | 30-34         | 35-39        | 40-44         | 45-          |
|-----------------|---------------|-----------------|-----------------|----------------|----------------|--------------|---------------|--------------|---------------|--------------|
| 실제가구<br>(구성비%)  | 177<br>(3.54) | 1154<br>(23.08) | 1623<br>(32.46) | 639<br>(12.78) | 764<br>(15.28) | 76<br>(1.52) | 380<br>(7.60) | 12<br>(0.24) | 97<br>(1.94)  | 78<br>(1.56) |
| 대체후가구<br>(구성비%) | 182<br>(3.64) | 1136<br>(22.72) | 1632<br>(32.64) | 630<br>(12.60) | 760<br>(15.20) | 77<br>(1.54) | 369<br>(7.38) | 13<br>(0.26) | 112<br>(2.24) | 89<br>(1.78) |

### 5) 병원이동수단

병원이동수단에 대한 무응답 대체는 금융기관이동수단, 금융기관소요시간, 병원소요시간 항목을 대체군으로 사용하였으며, 이후에 실제 대체적용을 할 때에는 금융기관이동수단이 먼저 대체되므로 1차 대체군을

이용할 수 있을 것으로 생각된다. <표 3-77>에서와 같이 총 5,000가구를 대체한 결과 대체의 정확도는 90.74%로 매우 높은 정확도를 보였다. 대체결과를 살펴보면 걸어서 이동하는 경우는 768가구 중에서 640가구, 자동차로 이동하는 경우는 3479가구 중에서 3261가구를 정확하게 찾아내었다. 앞의 경우와 마찬가지로 대체의 정확도가 높은 이유는 동사무소, 금융기관, 병원의 위치가 비슷해서 이동수단이 같은 가구가 많기 때문인 것으로 판단된다. 하지만 이러한 정보가 없을 때에는 대체의 정확도가 상당부분 떨어질 수도 있을 것이다.

<표 3-77> 병원이동수단에 대한 대체군 및 대체결과

| 대체군(1차)                                     |            | 대체군(2차)                  |            |             |
|---|------------|--------------------------|------------|-------------|
| 금융기관이동수단(100)<br>금융기관소요시간(26)<br>병원소요시간(25) |            | 수도형태(100)<br>농업특성조사구(72) |            |             |
| 대체결과_대체정확도 (90.74%)                         |            |                          |            |             |
|   | 걸어서(대체)    | 자동차(대체)                  | 기타(대체)     | 합계(구성비(%))  |
| 걸어서   | 640        | 110                      | 18         | 768(15.36)  |
| 자동차   | 114        | 3261                     | 104        | 3479(69.58) |
| 기타  | 24         | 93                       | 636        | 753(15.06)  |
| 합계  | 778(12.80) | 3464(69.28)              | 758(15.16) | 5000        |

대체 전후의 항목분포는 걸어서 이동하는 경우 실제 가구가 768가구 (15.36%)에서 778가구(12.80%), 자동차로 이동하는 경우는 3479가구 (69.58%)에서 3464가구(69.28%), 기타는 753가구(15.06%)에서 758가구 (15.16%)로 대체전후의 분포변화는 발생되지 않음을 알 수 있다.

## 6) 병원소요시간

병원소요시간에 대한 무응답 대체는 금융기관소요시간, 읍면동, 동사무소소요시간 항목을 대체군으로 사용하였으며, 금융기관소요시간 항목이 먼저 대체되므로 1차 대체군만을 이용할 수 있을 것이다. 임의로

발생시킨 5000가구의 표본평균과 대체한 후의 표본평균의 값은 <표 3-78>에 주어져 있다. 5번 반복된 실험결과를 보면 실제표본평균과 대체한 후의 표본평균의 차이는 0.01분에서 0.24분 정도로 나왔으며, 5번의 평균은 0.138분으로 대체 후에도 병원소요시간은 차이가 거의 나지 않음을 알 수 있다.

<표 3-78> 병원소요시간에 대한 대체군 및 표본평균 차이정도

| 대체군(1차)  |        | 대체군(2차)   |      |
|--|--------|---|------|
| 금융기관소요시간(100)<br>읍면동(32), 동사무소소요시간(23)<br>병원이동수단(12) |        | 수도형태(100), 승용차보유(67)<br>화물차보유(55), 농업종사경력(43)<br>거처종류(29) |      |
| 표본평균 절대차이의 평균 (0.138분)                               |        |   |      |
| 실험   | 실제표본평균 | 대체후 표본평균  | 절대차이 |
| 1  | 19.72  | 19.96   | 0.24 |
| 2  | 19.64  | 19.65   | 0.01 |
| 3  | 19.61  | 19.84   | 0.23 |
| 4  | 19.78  | 19.85   | 0.07 |
| 5  | 19.70  | 19.84   | 0.14 |

병원소요시간에 대한 분포변화를 고려하기 위하여 범주화하여 확인한 결과가 <표 3-79>에 주어져 있다. 5-9분은 681가구(13.62%)에서 669가구(13.38%), 10-14분은 1244가구(24.88%)에서 1224가구(24.48%), 40-44분은 316가구(6.32%)에서 309가구(6.18%), 45분 이상은 228가구(4.56%)에서 231가구(4.62%) 등으로 변화되나 대체전후의 분포변화는 일어나지 않는다고 판단된다.

<표 3-79> 병원소요시간에 대한 대체전후의 분포변화

| 소요시간              | 0-4          | 5-9            | 10-14           | 15-19          | 20-24          | 25-29         | 30-34          | 35-39        | 40-44         | 45-           |
|-------------------|--------------|----------------|-----------------|----------------|----------------|---------------|----------------|--------------|---------------|---------------|
| 실제가구<br>(구성비(%))  | 80<br>(1.60) | 681<br>(13.62) | 1244<br>(24.88) | 579<br>(11.58) | 894<br>(17.88) | 147<br>(2.94) | 785<br>(15.70) | 46<br>(0.92) | 316<br>(6.32) | 228<br>(4.56) |
| 대체후가구<br>(구성비(%)) | 89<br>(1.78) | 669<br>(13.38) | 1224<br>(24.48) | 578<br>(11.56) | 910<br>(18.20) | 154<br>(3.08) | 798<br>(15.96) | 38<br>(0.76) | 309<br>(6.18) | 231<br>(4.62) |

## 자. 주거에 관한 사항

### 1) 거처종류

거처종류에 대한 연관성 분석의 결과 난방연료종류, 건축년도, 시도 항목을 대체군으로 사용하였으며, 이 경우 지역(시군구, 읍면동, 조사구) 항목들을 특별 대체군으로 추가하여 대체를 실시하였다. 그 이유는 주거에 관한 사항은 다른 항목들과는 달리 근처의 가구와 많은 유사성이 있을 것으로 판단된다. 실제로 분석을 한 결과 더 정확한 대체가 되었음을 확인하였다. 따라서 주거에 관한 항목들은 지역에 대한 정보를 추가하여 모의실험을 실시할 것이다. 그리고 난방연료종류 정보를 사용할 수 없을 경우에는 지역이 포함된 2차 대체군을 적용하였다.

<표 3-80> 거처종류에 대한 대체군 및 대체결과

| 대체군(1차)   |              | 대체군(2차)                               |              |             |              |              |                |
|---|--------------|---------------------------------------|--------------|-------------|--------------|--------------|----------------|
| 난방연료종류(100)<br>건축년도(54), 시도(12)<br>지역(특별대체군)_시군구, 읍면동,<br>조사구 |              | 건축년도(100)<br>농업특성조사구(94)<br>지역(특별대체군) |              |             |              |              |                |
| 대체결과_대체정확도 (95.88%)   |              |                                       |              |             |              |              |                |
|   | 단독주택<br>(대체) | 아파트<br>(대체)                           | 연립주택<br>(대체) | 다세대<br>(대체) | 비거주용<br>(대체) | 주택이외<br>(대체) | 합계<br>(구성비(%)) |
| 단독주택  | 4539         | 23                                    | 17           | 7           | 30           | 17           | 4633(92.66)    |
| 아파트   | 24           | 196                                   | 3            | 5           | 1            | 1            | 230(4.60)      |
| 연립주택  | 10           | 5                                     | 16           | 1           | 2            | 0            | 34(0.68)       |
| 다세대   | 7            | 2                                     | 0            | 10          | 0            | 0            | 19(0.38)       |
| 비거주용  | 32           | 4                                     | 2            | 0           | 7            | 0            | 45(0.90)       |
| 주택이외  | 13           | 0                                     | 0            | 0           | 0            | 26           | 39(0.78)       |
| 합계  | 4625(92.50)  | 230(4.60)                             | 38(0.76)     | 23(0.46)    | 40(0.80)     | 44(0.88)     | 5000           |

<표 3-80>에서와 같이 총 5,000가구를 대체한 결과 대체의 정확도는 95.88%로 상당히 높음을 알 수 있다. 이는 단독주택이 90%이상을 차지

하고 있고, 단독주택에 대한 대체의 정확도가 매우 높기 때문인 것으로 보인다. 아파트 역시 정확도가 높으나, 다른 항목들은 좋은 대체가 이루어지지 않는다. 이전에도 언급했듯이 단독주택을 제외하면 구성비가 매우 작기 때문에 구성비가 큰 쪽으로 묻혀버리는 현상이 일어남을 알 수 있다. 하지만 전체적으로 볼 때에는 그 수가 상당히 작기 때문에 오분류의 효과는 미미하게 나타난다.

대체 전후의 항목분포의 변화를 살펴보면 단독주택은 실제 가구가 4633가구(92.66%)에서 4625가구(92.50%), 아파트는 230가구(4.60%)에서 230가구(4.60%), 연립주택은 34가구(0.68%)에서 38가구(0.76%) 등으로 변화되어 대체로 인한 분포변화의 문제도 거의 발생되지 않음을 알 수 있다.

## 2) 건축년도

건축년도에 대한 무응답 대체는 화장실종류, 난방연료종류, 부엌형태, 거처종류, 수도형태 항목을 대체군으로 사용하였으며, 지역(시군구, 읍면동, 조사구)항목들을 특별 대체군으로 추가하여 대체를 실시하였다. 그리고 화장실종류 정보를 사용할 수 없을 경우 2차 대체군을 적용하였다. 건축년도는 이전의 판매금액처럼 항목의 특성상 완전한 범주형 자료처럼 대체의 정확도를 측정한다는 것은 적절하지 못하다. 예를 들면 실제 건축년도가 1999년인 집이 있다고 하자. 이때 대체된 값이 2000년이면 상당히 비슷하게 추정되었음에도 불구하고 오분류로 판단되고, 반대로 대체값이 1995년이면 차이가 있음에도 불구하고 정확한 대체가 되었다고 판단한다. 이런 이유로 본 항목에서는 대체의 정확도는 큰 의미가 없으므로 대체전후의 분포변화에 대해서만 살펴보고자 한다.

대체 전후의 항목분포의 변화가 <표 3-81>에 주어져있다. 2002년에 지어진 집은 실제 105가구(2.10%)에서 94가구(1.88%), 2000년은 177가구(3.54%)에서 188가구(3.76%), 1995-1999년은 971가구(19.42%)에서 992가구(19.84%), 1980-1984년은 424가구(8.48%)에서 413가구(8.26%), 1959년 이전은 651가구(13.02%)에서 671가구(13.42%) 등으로 변화되어 대체로 인한 분포변화의 문제는 발생되지 않음을 알 수 있다.

<표 3-81> 건축년도에 대한 대체군 및 대체전후의 분포변화

| 대체군(1차)   |               |              |              |               |              | 대체군(2차)  |                |                |               |               |                |               |                |
|---|---------------|--------------|--------------|---------------|--------------|--|----------------|----------------|---------------|---------------|----------------|---------------|----------------|
| 화장실종류(100), 난방연료종류(45)<br>부엌형태(36), 거처종류(32)<br>수도형태(25), 지역(특별대체군) |               |              |              |               |              | 난방연료종류(100), 농업종사경력(53)<br>거처종류(44), 지역(특별대체군) |                |                |               |               |                |               |                |
| 년도  | 2005<br>년     | 2004<br>년    | 2003<br>년    | 2002<br>년     | 2001<br>년    | 2000<br>년                                      | 1995-<br>1999  | 1990-<br>1994  | 1985-<br>1989 | 1980-<br>1984 | 1970-<br>1979  | 1960-<br>1969 | 1959<br>이전     |
| 실제가구<br>(구성비(%))  | 108<br>(2.16) | 98<br>(1.96) | 85<br>(1.70) | 105<br>(2.10) | 92<br>(1.84) | 177<br>(3.54)                                  | 971<br>(19.42) | 866<br>(17.32) | 458<br>(9.16) | 424<br>(8.48) | 624<br>(12.48) | 341<br>(6.82) | 651<br>(13.02) |
| 대체후가구<br>(구성비(%))   | 111<br>(2.22) | 98<br>(1.96) | 96<br>(1.92) | 94<br>(1.88)  | 88<br>(1.76) | 188<br>(3.76)                                  | 992<br>(19.84) | 837<br>(16.74) | 465<br>(9.30) | 413<br>(8.26) | 584<br>(11.68) | 363<br>(7.26) | 671<br>(13.42) |

### 3) 난방연료종류

난방연료종류에 대한 무응답 대체는 거처종류, 부엌형태, 목욕시설, 시군구, 시도, 건축년도 항목을 대체군으로 사용하였으며, 지역(읍면동, 조사구) 항목들을 특별 대체군으로 추가하여 대체를 실시하였다. 그리고 실제 대체적용을 할 때에는 거처종류가 먼저 대체되므로 1차 대체군을 이용할 수 있을 것으로 생각된다. <표 3-82>에서와 같이 총 5,000가구를 대체한 결과 대체의 정확도는 71.12%로 다른 항목들에 비해서는 다소 낮게 나타났다. 이는 세부항목들이 매우 다양하며 난방연료종류를 확실하게 구분을 해줄 수 있는 보조변수가 없는 것으로 보인다. 특히 기름보일러의 구성비가 상당히 높기 때문에 다른 난방들이 기름보일러로 대체가 되는 현상이 일어남을 알 수 있다. 하지만 일반적으로 볼 때 70%가 넘는 정확도는 낮은 것이 아님을 밝혀두고자 한다.

대체 전후의 항목분포의 변화를 살펴보면 도시가스보일러를 사용하는 가구는 실제 가구가 218가구(4.36%)에서 208가구(4.16%), 기름보일러는 3561가구(71.22%)에서 3590가구(71.80%), 전기보일러는 663가구(13.26%)에서 643가구(12.86%), 연탄보일러는 213가구(4.26%)에서 233가구(4.66%) 등으로 변화되어 대체의 정확도와는 달리 대체로 인한 분포 변화의 문제는 거의 발생되지 않음을 알 수 있다.



〈표 3-82〉 난방연료종류에 대한 대체군 및 대체결과

| 대체군(1차)   |              | 대체군(2차)                                   |               |                 |              |                |               |              |               |              |             |
|---|--------------|---|---------------|-----------------|--------------|----------------|---------------|--------------|---------------|--------------|-------------|
| 거처종류(100), 부업형태(59),<br>목욕시설(51), 시군구(31)<br>시도(30), 건축년도(17)<br>지역(특별대체군)_읍면동, 조사구 |              | 목욕시설(100), 부업형태(76)<br>지역(특별대체군)_읍면동, 조사구 |               |                 |              |                |               |              |               |              |             |
| 대체결과_대체정확도 (71.12%)   |              |   |               |                 |              |                |               |              |               |              |             |
|   | 1<br>(대체)    | 2<br>(대체)                                 | 3<br>(대체)     | 4<br>(대체)       | 5<br>(대체)    | 6<br>(대체)      | 7<br>(대체)     | 8<br>(대체)    | 9<br>(대체)     | 10<br>(대체)   | 합계          |
| 1   | 19           | 1   | 1             | 0               | 0            | 0              | 0             | 0            | 0             | 0            | 21(0.42)    |
| 2   | 2            | 18  | 0             | 0               | 0            | 0              | 0             | 0            | 0             | 0            | 20(0.40)    |
| 3   | 2            | 2   | 175           | 24              | 4            | 11             | 0             | 0            | 0             | 0            | 218(4.36)   |
| 4   | 1            | 0   | 20            | 2947            | 13           | 364            | 111           | 15           | 34            | 56           | 3561(71.22) |
| 5   | 0            | 0   | 3             | 23              | 40           | 2              | 0             | 0            | 0             | 0            | 68(1.36)    |
| 6   | 0            | 0   | 7             | 374             | 3            | 229            | 30            | 1            | 8             | 11           | 663(13.26)  |
| 7   | 0            | 0   | 2             | 101             | 0            | 18             | 70            | 3            | 16            | 3            | 213(4.26)   |
| 8   | 0            | 0   | 0             | 15              | 0            | 1              | 2             | 4            | 0             | 0            | 22(0.44)    |
| 9   | 0            | 0   | 0             | 42              | 1            | 6              | 14            | 3            | 47            | 6            | 119(2.38)   |
| 10  | 0            | 0   | 0             | 64              | 0            | 12             | 6             | 2            | 4             | 7            | 95(1.90)    |
| 합계  | 24<br>(0.48) | 21<br>(0.42)                              | 208<br>(4.16) | 3590<br>(71.80) | 61<br>(1.22) | 643<br>(12.86) | 233<br>(4.66) | 28<br>(0.56) | 109<br>(2.18) | 83<br>(1.66) | 5000        |

\* 1(중앙난방), 2(지역난방), 3(도시가스보일러), 4(기름보일러), 5(프로판가스보일러)  
6(전기보일러), 7(연탄보일러), 8(연탄아궁이), 9(재래식아궁이), 10(기타)

#### 4) 부업형태

부업형태에 대한 무응답 대체는 난방연료종류, 목욕시설, 건축년도, 화장실종류 항목을 대체군으로 사용하였으며, 지역(시도, 시군구, 읍면동, 조사구) 항목들을 특별 대체군으로 추가하여 대체를 실시하였다. 그리고 실제 대체적용을 할 때에는 난방연료종류가 먼저 대체되므로 1차 대체군을 이용할 수 있을 것으로 생각된다.

<표 3-83> 부업형태에 대한 대체군 및 대체결과

| 대체군(1차)  |             | 대체군(2차)                        |         |             |
|--|-------------|--------------------------------|---------|-------------|
| 난방연료종류(100), 목욕시설(37)<br>건축년도(16), 화장실종류(12)<br>지역(특별대체군)_시도, 시군구,<br>읍면동, 조사구 |             | 지역(특별대체군)_시도, 시군구,<br>읍면동, 조사구 |         |             |
| 대체결과_대체정확도 (95.20%)  |             |                                |         |             |
|  | 입식(대체)      | 재래식(대체)                        | 없음(대체)  | 합계(구성비(%))  |
| 입식   | 4668        | 124                            | 2       | 4794(95.88) |
| 재래식  | 112         | 89                             | 1       | 202(4.04)   |
| 없음   | 1           | 0                              | 3       | 4(0.08)     |
| 합계   | 4781(95.62) | 213(4.26)                      | 6(0.12) | 5000        |

<표 3-83>에서와 같이 총 5,000가구를 대체한 결과 대체의 정확도는 95.20%로 매우 높은 정확도를 보였다. 대체결과를 살펴보면 부업형태가 입식인 경우는 4794가구 중에서 4668가구, 재래식인 경우는 202가구 중에서 89가구를 찾아내었다. 부업형태에 대한 대체의 정확도가 높은 이유는 대부분의 가구가 입식인 형태이고 대체는 구성비가 큰 입식으로 되기 때문인 것으로 판단된다. 따라서 구성비가 작은 재래식인 경우는 다소 정확도가 떨어지는 것을 볼 수 있다. 하지만 전체적으로 볼 때에는 그 수가 상당히 작기 때문에 오분류의 효과는 없어 보인다.

대체 전후의 항목분포를 살펴보면 입식인 경우는 실제 가구가 4794가구(95.88%)에서 4781가구(95.62%), 재래식인 경우는 202가구(4.04%)에서 213가구(4.26%), 없음은 4가구(0.08%)에서 6가구(0.12%)로 대체전후의 분포변화는 거의 발생되지 않음을 알 수 있다.

### 5) 수도형태

수도형태에 대한 무응답 대체는 읍면동, 시도, 거처종류, 난방연료종류, 동사무소이동수단 항목을 대체군으로 사용하였으며, 지역(시군구,

조사구) 항목들을 특별 대체군으로 추가하여 대체를 실시하였다. 수도 형태의 경우는 가장 연관성이 높은 읍면동 항목이 행정구역이므로 무응답이 될 수 없어 2차 대체군은 없어도 될 것이다.

<표 3-84> 수도형태에 대한 대체군 및 대체결과

| 대체군(1차)  |             | 대체군(2차)       |              |             |             |
|--|-------------|---------------|--------------|-------------|-------------|
| 읍면동(100), 시도(89)<br>거처종류(41), 난방연료종류(28)<br>동사무소이동수단(28)<br>지역(특별대체군)_시군구, 조사구 |             | 필요없음          |              |             |             |
| 대체결과_대체정확도 (82.50%)  |             |               |              |             |             |
|  | 상수도<br>(대체) | 간이상수도<br>(대체) | 자가수도<br>(대체) | 없음<br>(대체)  | 합계(구성비(%))  |
| 상수도  | 1523        | 70            | 42           | 109         | 1744(34.88) |
| 간이상수도  | 70          | 1262          | 55           | 139         | 1526(30.52) |
| 자가수도   | 34          | 61            | 156          | 27          | 278(5.56)   |
| 없음   | 94          | 145           | 29           | 1184        | 1452(29.04) |
| 합계   | 1721(34.42) | 1538(30.76)   | 282(5.64)    | 1459(29.18) | 5000        |

<표 3-84>에서와 같이 총 5,000가구를 대체한 결과 대체의 정확도는 82.50%로 높은 정확도를 보였다. 세부항목들은 비슷한 구성비를 가지며 대체의 정확도도 비슷하게 나타나고 있다.

대체 전후의 항목분포의 변화를 살펴보면 수도형태가 상수도인 가구는 실제 가구가 1744가구(34.88%)에서 1721가구(34.42%), 간이상수도는 1526가구(30.52%)에서 1538가구(30.76%), 자가수도는 278가구(5.56%)에서 282가구(5.64%), 없음은 1452가구(29.04%)에서 1459가구(29.18%)로 변화되어 대체로 인한 분포변화의 문제는 발생되지 않음을 알 수 있다.

## 6) 화장실종류

화장실종류에 대한 무응답 대체는 목욕시설, 건축년도, 시도, 부엌형태, 항목을 대체군으로 사용하였으며, 지역(시군구, 읍면동, 조사구) 항목들을 특별 대체군으로 추가하여 대체를 실시하였다. 그리고 목욕시설 정보를 사용할 수 없을 경우 2차 대체군을 적용하였다.

<표 3-85> 화장실종류에 대한 대체군 및 대체결과

| 대체군(1차)   |             | 대체군(2차)   |         |             |
|---|-------------|---|---------|-------------|
| 목욕시설(100), 건축년도(64)<br>시도(44), 부엌형태(21)<br>지역(특별대체군)_시군구, 읍면동,<br>조사구 |             | 건축년도(100), 시도(58)<br>지역(특별대체군)_시군구, 읍면동,<br>조사구 |         |             |
| 대체결과_대체정확도 (77.26%)   |             |   |         |             |
|   | 수세식(대체)     | 재래식(대체)   | 없음(대체)  | 합계(구성비(%))  |
| 수세식   | 3085        | 583   | 1       | 3669(73.38) |
| 재래식   | 552         | 777   | 1       | 1330(26.60) |
| 없음  | 0           | 0   | 1       | 1(0.02)     |
| 합계  | 3637(72.74) | 1360(27.20)                                     | 3(0.06) | 5000        |

<표 3-85>에서와 같이 총 5,000가구를 대체한 결과 대체의 정확도는 77.26%로 나타났다. 많은 부분을 차지하고 있는 수세식의 경우는 80% 이상의 대체 정확도를 보여주고 있으나, 재래식의 경우는 다소 정확도가 떨어짐을 알 수 있다. 이러한 이유는 목욕시설, 건축년도 등으로 수세식과 재래식을 완전하게 나눌 수 없고 겹치는 부분이 20~30%가 있기 때문인 것으로 판단된다. 따라서 재래식의 경우 구성비가 높은 수세식으로 많은 부분이 대체가 되고 있음을 알 수 있다.

대체 전후의 항목분포의 변화를 살펴보면 화장실이 수세식인 경우는 실제 가구가 3669가구(73.38%)에서 3637가구(72.74%), 재래식은 1330가구(26.60%)에서 1360가구(27.20%) 등으로 대체전후의 분포변화는 거의 발생되지 않음을 알 수 있다.

## 7) 목욕시설

목욕시설에 대한 무응답 대체는 난방연료종류, 화장실종류, 부엌형태, 시도, 거처종류 항목을 대체군으로 사용하였으며, 지역(시군구, 읍면동, 조사구) 항목들을 특별 대체군으로 추가하여 대체를 실시하였다. 그리고 실제 대체적용을 할 때에는 난방연료종류가 먼저 대체되므로 1차 대체군을 이용할 수 있을 것으로 생각된다.

<표 3-86> 목욕시설에 대한 대체군 및 대체결과

| 대체군(1차)  |             | 대체군(2차)                        |           |             |
|--|-------------|--------------------------------|-----------|-------------|
| 난방연료종류(100), 화장실종류(90)<br>부엌형태(48), 시도(12), 거처종류(11)<br>지역(특별대체군)_시군구, 읍면동,<br>조사구 |             | 지역(특별대체군)_시도, 시군구,<br>읍면동, 조사구 |           |             |
| 대체결과_대체정확도 (89.64%)  |             |                                |           |             |
|  | 온수시설(대체)    | 비온수시설(대체)                      | 없음(대체)    | 합계(구성비(%))  |
| 온수시설   | 4221        | 17                             | 203       | 4441(88.82) |
| 비온수시설  | 22          | 28                             | 24        | 74(1.48)    |
| 없음   | 235         | 17                             | 233       | 485(9.70)   |
| 합계   | 4478(89.56) | 62(1.24)                       | 460(9.20) | 5000        |

<표 3-86>에서와 같이 총 5,000가구를 대체한 결과 대체의 정확도는 89.64%로 높은 정확도를 보였다. 대체결과를 살펴보면 목욕시설이 온수시설인 경우는 4441가구 중에서 4221가구, 비온수시설인 경우는 74가구 중에서 28가구, 없음은 485가구 중에서 233가구를 찾아내었다. 목욕시설은 90%정도의 가구가 온수시설이고 대체는 구성비가 큰 온수시설로 많은 부분 이루어지기 때문에 대체의 정확도는 높게 나타나고 있다. 하지만 비온수시설 및 없음은 구성비가 작아 다소 정확도가 떨어지는 것을 볼 수 있으나 전체적으로 볼 때에는 문제가 되지 않을 것으로 판단된다.

대체 전후의 항목분포를 살펴보면 온수시설인 경우는 실제 가구가

4441가구(88.82%)에서 4478가구(89.56%), 비온수시설인 경우는 74가구(1.48%)에서 62가구(1.24%), 없음은 485가구(9.70%)에서 460가구(9.20%)로 대체전후의 분포변화는 거의 발생되지 않음을 알 수 있다.

### 3. 각 항목별 모의실험결과 요약

각 항목에 대하여 적절한 설계하에서 모의실험을 실시하였다. 대체의 정확도가 95%가 넘는 항목도 있으며, 60%대로 나타난 것도 있었다. 그러나 많은 경우 상당히 좋은 무응답 대체가 되는 것을 확인하였다. 그리고 대체의 정확도는 대체군(보조변수)의 정보를 얼마나 많이 사용할 수 있는지의 여부에 따라 결정된다고 해도 과언이 아닐 것이다. 이상으로 앞에서 실시한 모의실험 결과를 간단하게 요약하고 본 연구를 마무리 하고자 한다.

〈표 3-87〉 가구원 항목들에 대한 모의실험결과 요약

| 항목      | 대체의 정확도(질적변수) | 표본평균 차이정도(양적변수) |
|---------|---------------|-----------------|
| 성별      | 81.84(%)      | —               |
| 나이      | —             | 0.045(세)        |
| 경영주와의관계 | 93.58(%)      | —               |
| 교육정도    | 70.18(%)      | —               |
| 혼인상태    | 82.31(%)      | —               |
| 주종사분야   | 89.05(%)      | —               |
| 농업종사기간  | 83.01(%)      | —               |
| 농업외종사기간 | 81.74(%)      | —               |
| 농업종사경력  | —             | 0.17(년)         |

〈표 3-88〉 경지 항목들에 대한 모의실험결과 요약

| 항목    | 표본평균 차이정도(양적변수) |
|-------|-----------------|
| 논면적   | 20.55(평)        |
| 자기논   | 34.07(평)        |
| 일모작논  | 19.73(평)        |
| 경지정리논 | 24.31(평)        |
| 밭면적   | 17.62(평)        |
| 자기밭   | 18.31(평)        |
| 목초지면적 | 2.22(평)         |

〈표 3-89〉 작물시설 항목들에 대한 모의실험결과 요약

| 항목       | 표본평균 차이정도(양적변수) |
|----------|-----------------|
| 자동화비닐하우스 | 0.32(평)         |
| 비닐하우스    | 2.49(평)         |
| 유리온실     | 0.02(평)         |
| 기타시설     | 0.10(평)         |

〈표 3-90〉 농기계, 논벼농사방법 항목들에 대한 모의실험결과 요약

| 항목     | 대체의 정확도(질적변수) | 표본평균 차이정도(양적변수) |
|--------|---------------|-----------------|
| 경운기    | —             | 0.0064(대)       |
| 트랙터    | —             | 0.0014(대)       |
| 콤바인    | —             | 0.0012(대)       |
| 관리기    | —             | 0.0084(대)       |
| 건조기    | —             | 0.0038(대)       |
| 이앙기    | —             | 0.0018(대)       |
| SS분무기  | —             | 0.0016(대)       |
| 농업용난방기 | —             | 0.0032(대)       |
| 벼묘판작업  | 76.84(%)      | —               |
| 논갈이    | 82.27(%)      | —               |
| 모내기    | 84.97(%)      | —               |
| 농약살포   | 76.59(%)      | —               |
| 벼베기탈곡  | 86.87(%)      | —               |

<표 3-91> 농업경영 항목들에 대한 모의실험결과 요약

| 항목       | 대체의 정확도(질적변수) |
|----------|---------------|
| 전·겸업수입구분 | 66.08(%)      |
| 영농형태     | 87.68(%)      |
| 판매금액구분   | 의미없음(본문참조)    |
| 판매처구분    | 43.68(%)      |

<표 3-92> 친환경농업 항목들에 대한 모의실험결과 요약

| 항목        | 대체의 정확도(질적변수) |
|-----------|---------------|
| 친환경수확작물유무 | 99.12(%)      |
| 친환경작물판매처  | 62.49(%)      |

<표 3-93> 정보화현황 항목들에 대한 모의실험결과 요약

| 항목       | 대체의 정확도(질적변수) |
|----------|---------------|
| 컴퓨터보유여부  | 97.90(%)      |
| 컴퓨터활용여부  | 99.73(%)      |
| 홈페이지개설여부 | 96.93(%)      |

<표 3-94> 생활여건 항목들에 대한 모의실험결과 요약

| 분석항목     | 대체의 정확도(질적변수) | 표본평균 차이정도(양적변수) |
|----------|---------------|-----------------|
| 동사무소이동수단 | 91.34(%)      | —               |
| 동사무소소요시간 | —             | 0.088(분)        |
| 금융기관이동수단 | 93.86(%)      | —               |
| 금융기관소요시간 | —             | 0.086(분)        |
| 병원이동수단   | 90.74(%)      | —               |
| 병원소요시간   | —             | 0.138(분)        |



〈표 3-95〉 주거 항목들에 대한 모의실험결과 요약

| 항목     | 대체의 정확도(질적변수) |
|--------|---------------|
| 거처종류   | 95.88(%)      |
| 건축년도   | 의미없음(본문참조)    |
| 난방연료종류 | 71.12(%)      |
| 부엌형태   | 95.20(%)      |
| 수도형태   | 82.50(%)      |
| 화장실종류  | 77.26(%)      |
| 목욕시설   | 89.64(%)      |

## 제5절 결론 및 향후 연구계획

본 연구에서는 농업총조사의 무응답 항목을 효율적으로 대체하기 위하여 핫텍 대체방법을 응용하여 농업총조사 항목에 적합한 대체기법을 개발하였다. 무응답 대체는 자료의 형태가 양적자료와 질적자료 여부에 따라서 주로 사용되는 대체방법이 있으나 농업총조사는 두 가지의 형태가 적절하게 섞여 있으므로 기존의 방법을 그대로 적용하기란 쉽지 않다. 따라서 본 연구에서는 기존의 질적자료 무응답 대체에 주로 이용되는 핫텍 대체방법을 응용하여 양적자료에도 사용할 수 있게 함으로써 이러한 문제점이 극복되었다고 본다.

핫텍 대체방법을 응용한 이유는 본문에서 설명한 내용과 같이 농업총조사 자료에는 이 방법이 가장 적절할 것으로 판단되기 때문이다. 자료의 형태라든지 양적자료의 특성, 대체후의 값 그리고 대체전후의 분포를 고려한다면 핫텍 대체방법을 응용하여 이용하는 것이 매우 정확도가 높을 것이라고 판단된다. 비록 구체적으로 모의실험을 통한 비교를 하지는 않았지만 직관적으로 알 수 있는 부분이라고 생각된다. 하지만 향후 비교연구를 통하여 본인의 주장을 뒷받침하고자 한다.

본 연구의 모의실험에서는 2005년 농업총조사 가구원부문 3,448,057명, 가구부문 1,272,908가구 전체를 자료로 이용하였다. 실험을 위해 목표변수에 대해서 임의로 5000개의 무응답을 발생시켰으며, 다른 변수들

도 동시에 일정비율만큼 무응답을 발생시켰다. 그리고 항목별로 5,000개의 무응답을 모두 대체하고 난 후에 목표변수가 범주형(질적변수)의 경우에는 각 범주의 분포변화 및 정분류율을 검토하였고, 연속형(양적변수)의 경우는 임의로 범주화된 분포변화와 반복을 통한 표본평균의 차이정도를 비교하였다. 모의실험의 결과를 살펴보면 95% 이상의 정확도를 가지는 항목도 있으며 60% 정도의 정확도도 있었다. 그러나 전체적으로 볼 때 매우 좋은 대체가 이루어지고 있음을 알 수 있었다. 그리고 표본평균의 차이정도는 거의 없는 것으로 보인다. 또한 대체후의 분포는 대체전과 거의 유사하여 대체로 인한 분포변화 문제는 고려하지 않아도 될 것으로 판단된다. 하지만 여기서 한 가지 고려해야 할 사항이 있다. 본 연구에서의 실험은 완전한 자료에 대해서 임의로 무응답을 발생시킨 것으로 실제 결측치가 있는 농업총조사 자료와는 차이가 있을 것으로 판단된다. 왜냐하면 모든 항목들을 일정비율로 결측을 시켰으나, 실제 자료는 단위 무응답에 가까운 개체들이 많이 존재할 것으로 보인다. 이러한 경우에는 연구된 대체군을 사용할 수가 없어 대체의 정확도는 상당히 떨어질 것이기 때문이다. 따라서 본 연구에서의 결과와 실제 자료를 대체한 후의 결과는 다소 차이가 있을 것으로 보인다. 하지만 본 연구에서 제시된 대체군과 대체방법은 현재의 농업총조사 자료를 적절하게 대체하기 위한 하나의 방법임에 틀림없을 것으로 확신한다. 모든 항목들에 대한 자세한 모의실험의 결과는 제4절을 참조하기 바란다.

본 연구에서는 2005년 농업총조사의 항목들 중에서 가구원과 가구부문 55개 항목에 대하여 제안된 대체기법을 적용하여 모의실험을 실시하였다. 따라서 향후 과수, 작물, 가축부문에 대한 연구가 계속적으로 진행되어야 할 것이다. 현재 무응답 대체가 가능한 데이터파일로 바꾸는 작업을 진행 중에 있으며 2009년 상반기에는 과수, 작물, 가축부문의 연구가 수행될 것으로 생각된다. 또한 2010년 농업총조사에 적용 가능하도록 추가연구를 실시해야 할 것이다. 2005년 조사에 비해서 조사표 항목이 다소 바뀔 것으로 예상되며, 이러한 항목들에 대한 대체군 및 대체 적용 연구가 수행되어야 할 것으로 판단된다.

통계조사의 환경이 계속적으로 악화되고 있고, 최근에는 농촌의 환경도 점차 도시화가 되어 무응답률이 높아지고 있는 현실이다. 미국, 캐

나다, 호주와 같은 통계 선진국들은 이미 10여년 전부터 무응답 처리에 관한 연구를 장기적으로 계획을 세워 진행해오고 있다. 물론 한국에 비해 국토면적이 상당히 크기 때문에 많은 조사에서 우편조사가 실시되어 무응답률이 상당히 높은 것은 사실이다. 하지만 한국도 시대적 변화에 의해서 발생될 수 있는 높은 무응답률에 대비한 연구가 지속적으로 진행이 되어야 할 것이다. 이러한 측면에서 본 연구는 향후 농업총조사의 품질을 향상시키는 데 많은 부분 도움이 될 것이며, 또한 다른 조사에서의 무응답 처리연구에도 일조할 수 있을 것으로 기대한다.

## 참고문헌

- 김규성(2000), “무응답 대체 방법과 대체 효과”, 「조사연구」, 제1권 2호, pp.1-14.
- 김규성(2000), “표본 대체 방법과 대체자료의 합리적 이용”, 한국은행 지원논문
- 김규성 · 이기재 · 김진(2005), “농어가경제조사에서 가중하택 무응답 대체방법의 활용”, 「응용통계연구」, 제18권 2호, pp.311-328.
- 김영원 · 이주원(2003), “CART를 활용한 결측값 대체방법: 인구주택총조사 혼인상태 항목을 중심으로”, 「조사연구」, 제4권 2호, pp.1-21.
- 김영원 · 조선경(1996), “표본조사에서 항목 무응답 대체 방법”, 「한국통계학회논문집」, 제3권 3호, pp.145-159.
- 김재광 · 한근식 · 윤연옥(2004), “가계조사 무응답 처리기법 연구”, 통계청, 「통계연구」, 제9권 1호, pp.79-102.
- 김진(2004), “농가경제조사에 대한 대체법 비교”, 통계청, 「통계연구」, 제9권 2호, pp.133-145.
- 송순관(2005), 「2005 인구주택총조사 무응답 처리방법 연구 및 읍면동 통계작성 가능성 검토」, 통계청 인구조사과.
- 이진희 · 김진 · 이기재(2006), “표본조사에서 공간 변수를 이용한 결측 대체의 효율성 비교”, 「응용통계연구」, 제19권 1호, pp.57-67.
- 조사통계연구회(2000), 「무응답 오차」, 자유아카데미.
- 최통진(2006), “농림어업총조사를 위한 무응답 보정에 관한 연구”, 석사학위논문.
- 최필근(2008), “농업총조사 항목 간 연관성 분석 및 대체군(보조변수) 개발”, 연구보고서, 통계개발원.
- 통계교육원(2005), 「무응답 처리 실무론」.
- 통계청(2005), 「2005 농림어업총조사 조사지침서」.
- \_\_\_\_\_ (2007), 「농림어업총조사 조사항목 변천 자료집」.
- Affifi, A. A. and R. M. Elashoff(1966), “Missing Observations in Multivariate Statistics I : Review of the Literature”, J. Am. Statist. Assoc., Vol. 61, pp.595-604.

- Agresti, A.(1990), *Categorical Data Analysis*, A Wiley-Interscience Publication.
- Berry, M. J. A. and G. S. Linoff(1997), *Data Mining Techniques*, John Wiley & Sons, New York.
- Kalton, G. and D. Kasprzyk(1986), “The Treatment of Missing Survey Data”, *Survey Methodology*, Vol. 12, pp.1-16.
- Kass, G.(1980), “An Exploratory Technique for Investigating Large Quantities of Categorical Data”, *Applied Statistics*, Vol. 29, No. 2, pp.119-127.
- Lessler and Kalsbeek(1992), *Nonsampling Error in Surveys*, John Wiley & Sons, New York.
- Quinlan, J. R.(1986), “Induction of Decision Tree”, *Machine Learning*, 1, pp.81-106.
- Rubin, D. B. and J. A. Little(1986), *Statistical Analysis with Missing Data*, John Wiley & Sons, New York.
- Sande, I. G.(1979), “A Personal View of Hot Deck Imputation Procedures”, *Survey Methodology*, Vol. 5, pp.238-258.