

제4장

집세조사의 표본설계방안

정미옥 · 박소현

제4장

제1절 서론

1. 연구배경 및 필요성

집세조사에서는 소비자물가조사 대상 38개 도시에서 표본으로 선정된 임차가구들을 방문하여 이들이 거주생활을 위해 지출하는 전·월세 금액에 대한 조사를 수행하고 있다. 조사된 자료는 소비자물가지수에서 품목별 상대적 중요도가 약 9.75%를 차지하고 있는 집세항목을 작성하는 데 활용된다. 또한 16개 시도에 대해서는 전세와 월세지수를 작성하여 공표하고 있으며, 38개 도시에 대해서는 전·월세를 구분하지 않은 집세지수로 공표하고 있다.

집세조사의 표본은 별도의 표본설계과정을 거치지 않고, 통계청의 가구대상 표본조사를 동일한 표본에서 실시할 수 있도록 설계된 다목적 표본에서 임차가구만을 표본으로 선정하고 있으며, 표본수가 부족한 일부 도시에 대해서만 표본을 보완해 주고 있다. 즉, 집세조사의 조사지역은 38개 소비자물가 조사도시의 경제활동인구조사 조사구역이며, 이 조사구역 내 주택에 거주하는 모든 전·월세 가구에 대해 조사가 수행되고 있다. 서로 다른 조사목적에 가지고 있는 여러 표본조사를 모두 만족시키면서 대표성 있는 표본설계를 하는 것은 현실적으로 매우 어렵다. 따라서 다목적 표본은 각 조사별 특성반영이 다소 미흡할 가능성이 있다.

본 연구에는 아파트 단지 및 시세정보로 주 부동산뱅크의 자료를 사용하였다. 원활하게 연구가 수행될 수 있도록 협조해 주신 주 부동산뱅크에 감사드린다.

()
()

또한 표본으로 선정된 가구는 여러 조사에 모두 응답해야 하기 때문에 표본가구에 응답부담이 가중될 우려가 있다. 그럼에도 불구하고 다목적 표본을 이용하는 가장 큰 이유는 제한된 예산과 인력 내에서 효율적으로 조사를 수행하기 위함이며, 동일한 가구에서 여러 조사가 수행되므로 표본관리가 용이하다는 장점도 있다.

현행 통계청의 다목적 표본은 가구부문 표본조사 중 표본규모가 가장 큰 경제활동인구조사 중심으로 경제활동과 상관성이 높은 특성들이 우선적으로 고려되어 설계되어 있으며, 주택특성이나 집세변동과 관련이 높은 항목들은 표본설계과정에 반영되지 못하였다. 또한 표본으로 선정된 조사구역 내에 거주하는 모든 임차가구를 조사대상으로 하고 있기 때문에 특정 조사구역에서 집중적으로 표본이 선정되어 다양한 지역적 특성들이 반영되지 못할 가능성이 있다. 따라서 정도 높은 집세지수의 작성을 위해 조사 본연의 취지에 맞는 대표성 있는 표본선정이 필요한 상황이다.

소비자물가는 단순히 물가의 변동정도만을 제시하여 물가정책에만 사용하는 경제지표가 아니라 다음과 같이 파급효과가 매우 큰 경제지표이다. 예를 들어 1996년 발표한 미국의 보스킨위원회 보고서에 따르면, 미국의 소비자물가는 평균 1년에 1.1%p 정도 과대평가되고 있으며, 이런 추세를 10년 동안 지속하는 경우 10년 후인 2006년 적자에 약 1,480억 달러, 국가부채에 6,910억 달러를 추가시키는 영향력을 소비자물가가 가지고 있다고 했다. 또한 소비자물가의 하위분류들은 GDP 디플레이터로 사용되고 있어 기준년 기준 GDP에도 소비자물가는 영향력을 미치고 있다. 이와 같이 소비자물가의 정확성 정도가 국가 경제에 미치는 영향은 매우 크며, 특히 소비자물가지수 작성대상 489개 품목(2005년 기준)중 집세에 해당하는 전세와 월세는 2가지 품목만으로 소비자물가지수 내에서 차지하는 비중이 9.75%로 가장 높다. 따라서 집세 변동의 정확성 정도는 소비자물가지수의 정확성에 미치는 영향이 매우 크며, 나아가서 국가 경제에 미치는 영향도 매우 큼을 알 수 있다.

집세조사가 갖고 있는 중요성에 비하여 그동안 집세조사에 대한 연구는 그렇게 활발하지 못하였으며, 그 방법론에 대한 기초연구가 최근 시도되고 있다. 강제춘(2006)은 현행 집세조사가 독립표본이 아니어서

지역별, 주택유형별의 표본 수 편차가 심하여 지수 산출에 대표성이 떨어지는 단점이 있다고 언급하였다. 이를 개선하기 위한 방안으로 독립 표본을 추출하는 방법이 최선이지만, 예산이나 인력과 같은 현실을 감안하여 지역별, 주택유형별 표본규모를 보완하는 방법을 차선책으로 제안하였다. 이 논문에서는 표본부문에 대한 문제점 뿐 아니라 가격반영 방법에 대한 문제점도 살펴보았으며, 주택유형에 따른 전·월세 지수의 세분화와 전세 가격 반영방법 등에 대한 개선안을 제시하였다. 또한 전백근(2006)도 집세조사의 표본은 다목적 표본으로 경제활동인구 조사구에서 부수적으로 조사가 이루어지고 있어 집세 표본조사구가 집세조사 목적과 부합되지 않아 정확한 집세 가격변동 파악에 어려움이 있다고 지적하였다. 이를 개선하기 위해 집세조사를 위한 독립표본이 필요하지만, 이를 실행하기는 현실적으로 어려워 집세지수의 정도 제고를 위한 통계적 추정방법을 대안으로 제시하였다.

집세조사의 표본에 대한 대표성이 낮다는 문제점은 몇몇 연구로부터 꾸준히 거론되어 왔지만, 모집단 구조와 표본구조 간 차이에 대한 근거는 구체적으로 제시되지 못하였다. 이에 정미옥(2009)은 모집단의 지역별 각 속성들이 표본에 얼마만큼 반영되고 있는지를 파악하기 위해 집세조사의 표본구조를 실증적으로 분석하였다. 2005년 인구주택총조사의 10% 표본조사 자료와 집세조사 자료를 연결하여 전세와 월세 가구별 주택유형의 구성비를 비교하였는데, 16개 시도보다 38개 도시별 분포에서 구성비에 많은 차이가 나타남을 확인하였다. 특히 대상 자료로부터 집세의 수준이 높을수록 집세의 변동률도 높게 나타나는 경향이 있음을 파악하였으나, 지역에 따라 집세 수준의 분포가 모집단과 차이가 있어 지역별 변동률이 과소 혹은 과대 반영될 우려가 있다고 지적하였다.

이렇듯 집세조사가 본연의 목적에 따라 설계된 표본이 아닌 다목적 표본을 사용하는 구조적인 한계로 인하여 정확한 집세의 변동을 파악하기 힘들다는 주장이 지속적으로 제기되고 있다. 물론 선행연구에서 차선책으로 제시되었던 가격반영방법 개선 및 사후층화조정 등의 후속조치를 통하여 통계의 정확성을 다소 향상시킬 수는 있겠지만, 이 방법으로는 다목적표본에 의해 발생하는 문제점을 근본적으로 해결할 수 없다. 또한 사후층화조정을 위해서는 임차가구의 모집단 구조를 파악할

수 있는 자료가 필요한데, 이용 가능한 자료로는 인구주택총조사가 유일하며 이 조사는 5년에 한 번씩 수행되고 있어 변화하는 모집단의 구조를 추정과정에 정확히 반영하기 어렵다. 따라서 표본의 구조적인 한계를 개선하기 위하여 실제로 집세의 변동에 영향을 미치는 요인을 고려하여 어떻게 표본설계에 반영할 수 있는지에 대한 연구가 필요하다.

2. 연구목적 및 내용

서울시 주택을 대상으로 임차료 가격과 변동에 영향을 미치는 요인을 분석하고, 이를 기초로 집세조사의 목적에 부합하는 표본설계 방법을 제시하는 것이 본 연구의 목적이다. 집세조사는 전국 38개 도시에 수행되고 있어 이들 도시를 모두 고려해야겠지만, 주택임대시장에서 서울이라는 도시의 비중과 영향력, 집세지수에서의 상대적 중요도, 연구기간, 데이터 수집 가능여부 등의 상황을 고려하여 서울시로 그 분석범위를 제한하도록 하겠다. 또한 이 연구에서 이용되는 가구와 주택의 정의, 주택특성의 구분기준 등은 2005 인구주택총조사의 정의를 따르며, 이에 대한 자세한 내용은 조사지침서를 참고하기 바란다. 그리고 전세, 보증금이 있는 월세, 보증금이 없는 월세, 사글세와 같이 다양한 점유형태의 가격을 비교하기 위해 국민은행에서 발표하는 2005년 11월 기준 지역별 월세이율¹⁾을 근거로 모두 전세금액으로 환산²⁾하여 분석하였으며, 이 환산금액을 이하 집세로 정의한다.

이 연구의 주요한 내용은 크게 두 가지로 요약될 수 있다. 첫 번째는 서울시 임차가구에 대해서 모집단 분석을 통해 집세의 변동에 영향을 미치는 요인을 탐색하는 것이다. 주택의 매매·전세가격 결정요인과 관련된 문헌연구를 통하여 선행연구에서 검증된 주택의 물리적 특성, 입지적 조건에 대한 요인들을 주요변수로 선정하여 분석하였다. 이때 주택의 유형에 따라 집세의 변동에 영향을 미치는 요인이나 집세변동의

지역별 월세이율	서울 울산	부산 경기	대구 기타지방	인천	광주	대전	
1)	:	(0.88) (1.14)	(0.96) (0.97)	(0.91) (1.16)	(1.17)	(1.13)	(1.01)
2)	전세보증금 = 월세금보증금 + $\frac{\text{월세금}}{\text{월세이율}} * 100$						

추세 및 크기 등이 상이하기 때문에 아파트와 아파트 이외의 주택으로 구분하여 분석하였다. 여기서 아파트 이외의 주택이란 단독주택, 연립주택, 다세대주택을 일컬으며, 이하 일반주택으로 지칭한다. 또한 아파트의 경우는 부동산정보업체의 자료를 이용하여 분석하였고, 일반주택은 집세의 변동을 파악할 수 있는 자료가 없어 집세 수준에 대해 2005 인구주택총조사 자료로 분석하였다.

두 번째는 집세변동에 영향을 미치는 유의미한 요인들을 기초로 하여 서울시 아파트와 일반주택에 대한 표본을 설계하는 것이다. 이를 위해 먼저 목표모집단을 정의하고, 조사가능 여부 및 현실적인 여건 등을 감안하여 조사모집단과 추출틀을 설정하였다. 다음으로 선정된 추출틀에 대하여 모집단을 분석하고, 층화방법 및 표본의 크기를 결정하였다. 그리고 각 층별로 일정 기준에 의해 표본을 배분하고 추출방법을 결정한 후, 층별로 집락과 집락 내 최종표본단위를 추출하는 등 일련의 표본설계 과정을 따랐다. 또한 서울시의 주택을 아파트와 일반주택으로 분리하여 독립적으로 표본 설계하였으며, 추정과정을 단순하게 하기 위해 층 내에서 추출단위의 추출률이 동일하게 유지될 수 있도록 하는 자체 가중설계가 되도록 하였다.

2절에서는 집세변동에 영향을 미치는 요인을 분석하고, 3절에서는 그 요인에 근거하여 일반주택에 대해서 표본설계를 하였다. 4절에서는 아파트에 대한 표본설계를 하였으며, 5절에서는 이 연구를 통해서 확인한 시사점 및 결론을 제시한다. 본 연구는 집세조사의 조사목적에 맞는 표본선정방법을 제안함으로써 대표성 있는 표본을 추출을 통해 집세의 변동을 현실적으로 반영해주고, 결과적으로 집세지수의 정확도 향상에 기여할 수 있을 것이라 기대한다.

제2절 집세변동 설명요인 분석

이 절에서는 집세의 변동을 결정짓는 요인을 실증적으로 분석하고자 한다. 먼저 선행연구들에서 주택의 매매가격과 임차가격의 결정에 유의미한 영향을 미치는 것으로 선택된 요인들이 무엇인지를 살펴보고, 해당 변수들을 기초로 하여 분석하였다. 아파트의 경우 부동산정보업체에 아파트 단지정보 뿐 아니라 매매 및 전세가격에 대한 정보가 시계열로 구축되어 있어 가격의 변동요인에 대한 분석이 가능하다. 그러나 일반 주택의 경우 모집단을 파악할 수 있는 자료로 인구주택총조사 자료가 유일한데, 인구주택총조사는 5년마다 수행되고 있어 특정 시점의 집세 가격 자료만 확인할 수 있을 뿐이며 집세가격의 변동을 파악할 수 없는 어려움이 있다. 따라서 일반주택의 경우 집세의 변동을 설명하는 요인을 분석하는 대신 이용 가능한 자료 중 가격변동과 상관성이 높은 집세의 수준을 반응변수로 대체하여 분석하고자 한다.

1. 선행연구

주택시장과 관련되어 그동안 활발히 연구되어온 주제로는 가격결정 요인에 관한 연구, 주택하위시장 및 권역설정에 관한 연구, 주택 매매와 임대시장 간 인과관계에 관한 연구 등이 있다. 그 중 가격결정요인에 관한 연구는 여러 시점에 걸쳐 다양한 지역을 대상으로 많은 연구가 수행되어왔다. <표 4-1>에는 이와 같은 선행연구들에 의해 주택의 매매가 혹은 전세가를 결정하는 요인으로 유의미하게 선택된 변수들과 이용되었던 분석방법이 요약되어 있다. 결정요인 분석에는 주택의 물리적인 특성과 입지적 특성, 주변 환경적 특성들이 변수로 고려되었으며, 주로 회귀분석과 상관분석이 방법론으로 이용되고 있다.

특히 임차료에 관점을 두고 수행된 연구로는 염돈민 외(1991)는 서울시에서 학군을 주택하위시장으로 간주하고 임대료에 영향을 주는 주택 특성 요인을 분석하였으며, 각 학군별로 가격함수의 구조가 다르게 나타나고 있음을 확인하였다. 그리고 정성훈 외(2002)는 수원시의 동일한

평형과 매매가격을 보이는 아파트에 대해 전세가격의 차이를 발생시키는데 많은 영향을 미치는 요인으로 전용면적, 방수, 현관방식, 소음을 꼽았다. 김현재(2003)는 서울시 아파트를 대상으로 동일한 변수에 대해 매매가격모형과 전세가격모형을 별도로 구축하고 두 모형을 비교함으로써 매매와 전세를 결정하는 요인을 도출하고자 하였다. 평수, 향, 층, 도심과의 거리, 교통시설과의 거리, 학교, 학군, 공원 등이 주택가격 결정의 주요인이며, 재건축, 재개발의 여부와 주거 면적이 매매가와 전세가 모형에서 서로 차이를 보이고 있다고 분석하였다. 또한 이정화(2008)는 지역계정에서 주거서비스 산출액을 추계하기 위해 16개 시도별로 임차료에 영향을 미치는 변수를 선택하였는데, 임차료에 많은 영향을 미치는 변수로 연건평, 사용방수, 난방시설, 건축년도를 선정하고, 이를 기초로 지역별 주거서비스 회귀모형을 구축한 바 있다.

본 연구에서는 이와 같이 다양한 주택 매매 및 임차료 결정요인들 중 표본설계를 위해 현실적으로 표본틀로부터 얻을 수 있는 변수만을 고려하였다. 일반주택의 경우 주택유형, 점유형태, 사용방수, 연건평, 대지면적, 거주 층, 건축년도, 난방시설형태, 주택 내 총 거주가구 수, 학군이 고려되었다. 아파트의 경우 공급면적, 난방방식, 입주년도, 총동수, 층수, 주차대수, 평형별 세대수, 방수, 욕실수, 재건축여부, 주상복합여부, 현관구조, 건설사 도급순위와 같은 단지별 특성정보와 부도심과의 거리, 지하철역과의 거리, 자연환경(공원, 산, 강)과의 거리, 학교(초, 중, 고)와의 거리, 생활편의시설과의 거리와 같은 입지적 조건 및 주변 환경정보를 결정요인 분석을 위한 변수로 선택하였다.

<표 4-1> 선행연구에서 채택한 주요 가격결정요인

연구자	사례지역	종속 변수	설명변수 가격결정요인	분석방법
엄든민 외 (1991)	서울시 아파트	매매가 전세가	방수 면적, 시설, 화장실 등 토심거리, 지가(地價), 교통, 전철, 도로율, 편의시설, 공해정도	회귀분석
우경 외 (2002)	수도권 주택	매매가	서울시와 물리적 거리, 교통축, 경부고속도로, 경의선, 전철역, 시청, 시장, 백화점, 초등학교, 중학교, 차진료기관, 종합병원, 근린공원, 직장과의 거리, 오염물질 배출시설수, 인구밀도, 사업체수, 사용면적, 사용방수, 건축년도, 주택시설, 부엌, 화장실, 목욕시설, 난방 방식, 출입구	회귀분석
정성훈 외 (2002)	수원시 아파트	전세가	전용면적, 총세대수, 방수, 현관방식, 입주년도, 아파트동수, 난방방식, 난방연료, 인동거리, 수원역까지의 거리, 건설회사, 평형, 동수, 시공업체, 준공일, 버스노선	상관분석
함중범 (2002)	서울시 아파트	매매가	수, 방수, 현관, 종합병원 수, 고등학교 수, 학군, 난방방식, 총세대수, 층수, 구별 면적, 공원면적, 규모, 건축연령, 총세대수, 총평수, 총세대	회귀분석
김현재 (2003)	서울시 아파트	매매가 전세가	수 재개발여부, 전망, 명성, 주상복합여부, 난방방식, 난방연료, 병상침실수, 공원면적, 도로현황, 공기오염도, 지역, 전용면적, 방수, 욕실수, 현관구조, 난방	회귀분석
신동훈 (2004)	춘천시 아파트	매매가	방식, 입주년도, 주차대수, 비율, 중심상업지거리, 평형, 층수, 향, 현관구조, 난방방식, 준공	상관분석
장한섭 (2006)	일산 아파트	매매가	년도, 학교, 사설학원 수, 역세권여부, 조망권여부	상관분석 회귀분석
정관호 외 (2009)	서울시 아파트	매매가	방수, 층상난방방식, 노후년수, 노후년수, 년 경과여부, 지하철역까지의 토보치간거리, 20	공간자기회귀분석

2. 일반주택

집세조사에서는 임차료의 평균수준보다는 수준의 변동에 관심이 있기 때문에 표본 설계변수를 선정하기 위해서는 임차료의 변동에 영향을 미치는 요인을 탐색하는데 관점을 두어야 할 것이다. 그러나 일반주택의 경우 월별 혹은 분기별 임차료 수준과 같은 시계열 자료가 없기 때문에 집세의 변동을 파악할 만한 모집단 자료가 없다. 정미옥(2009)은 집세조사 자료를 이용하여 집세 금액대별 변동률의 차이를 분석하였는데, 집세의 수준에 따라 변동률에 차이가 나타나고 있음을 보였다. 이에 본 연구에서는 집세 변동에 대한 자료가 없는 일반주택에 대해 집세의 변동과 상관성이 높은 집세로 대체하여 결정요인분석을 수행하였다.

<표 4-2>는 2005년 인구주택총조사 자료의 서울시 일반주택³⁾에 거주하는 일반가구⁴⁾ 중 점유형태가 주거전용인 임차가구를 대상으로 집세와 연속형 변수들 간에 상관분석을 수행한 결과이다.

집세는 가구의 총방수를 집세로 나눈 방1개당 집세와의 상관계수가

<표 4-2> 주택특성들 간 상관행렬

상관행렬	집세	방 개당 집세	가구 총방수	주택 총방수	주택 연건평	주택 대지면적	주택내 거주가구수
집세							
방 개당 집세	1.00						
가구 총방수	0.54	1.00					
주택 총방수	0.43	-0.39	1.00				
주택 연건평	-0.04	0.13	-0.19	1.00			
주택 대지면적	0.07	0.31	-0.22	0.80	1.00		
주택내 거주가구수	0.22	0.32	-0.10	0.52	0.57	1.00	
	-0.06	0.33	-0.38	0.80	0.77	0.45	1.00

일반주택 : 단독주택, 아파트, 연립주택, 다세대주택
 일반가구 : 가구 구분어 가족으로 이루어진 가구, 가족과 가족 이외의 사람이 함께 사
 3) 는 가구 : 인 가구, 가족이 아닌 남남끼리 사는 인 이하의 가구
 4) : , 1 , 5

0.54로 가장 높고, 가구의 총방수와는 0.43, 주택 대지면적과는 0.22의 순으로 양의 상관관계에 있다. 주택의 총방수와 연건평, 대지면적, 주택 내 거주가구수 간에는 매우 높은 양의 상관관계가 있는데, 연건평과 대지면적, 거주가구수가 주택의 총방수보다 집세에 기여하는 바가 더 클 것으로 생각되어 주택의 총방수는 뒤의 회귀모형에서 제외시켰다.

임차가구의 거주형태에 따라 집세에 차이가 있는지를 확인해보기 위해 <표 4-3>과 같이 주택유형과 점유형태 별로 집세의 평균과 중위수를 구해보았다. 집세의 평균과 중위수 모두 일반단독과 다가구단독주택에 비해 연립주택과 다세대주택이 높는데, 연립·다세대주택의 평균 전세금은 약 6,100만원 정도로 단독주택이 4,200~4,400만원에 비해 약 2,000만원 가량 높음을 알 수 있다. 또한 동일 주택유형 내에서는 월세가구보다 전세가구의 집세수준이 높으며, 중위수가 평균보다 금액이 낮은 것으로 보아 집세는 오른쪽으로 꼬리가 긴 분포의 형태를 띠고 있는 것으로 보인다.

이와 같이 집세평균은 주택유형과 점유형태에 따라 집세 평균에는 차이가 있는 것으로 보이는데, 이러한 차이가 통계적으로 유의미한지를 검증하기 위해 분산분석을 실시하였다. <표 4-4>의 분산분석 결과를 보

<표 4-3> 임차가구 거주 주택유형 및 점유형태에 따른 집세

구분	일반영입겸		다가구		연립주택		다세대주택		
	용 단독주택		단독주택		전세		전세		
	전세	월세	전세	월세	전세	월세	전세	월세	
	가구	가구	가구	가구	가구	가구	가구	가구	
평균									
집세	중위수	4,408	3,799	4,284	4,010	6,171	4,266	6,126	5,638
		4,000	3,409	4,000	3,709	5,300	3,641	5,500	5,114

<표 4-4> 집세에 대한 주택유형 및 점유형태 분산분석 결과

주택유형 점유형태	DF	Anova SS	Mean Square	F Value	Pr > F
*	7	5.79E+10	8.27E+09	1447.9	<.0001

면, p-값이 0.0001보다 작은 값으로 유의수준 0.05하에서 주택유형 및 점유형태에 따라 평균 집세에는 유의미한 차이가 있는 것으로 확인되었다. 이렇듯 주택유형과 점유형태에 의해 발생하는 집세수준의 차이는 배제시키고, 주택특성만의 효과를 파악하기 위해 다음의 회귀모형은 주택유형과 점유형태에 따라 모형을 개별적으로 구축하였다. 이때 주택유형이 단독주택인 경우 일반 단독주택과 다가구 단독주택, 영업겸용 단독주택으로 구분할 수 있는데, 영업겸용 단독주택을 일반 단독주택에 포함시켜 일반단독주택(영업겸용포함)과 다가구단독주택의 두 유형으로 구분하였다.

집세를 반응변수로 하는 다중회귀모형을 구축하기 위한 설명변수로는 선행연구들에 의해 검토된 변수들과 상관분석을 통해 검증된 변수들을 바탕으로 선정하였다. 총사용방수, 주택의 연건평, 주택의 대지면적, 거주층구분, 건축년도, 난방시설형태, 주택내 총거주가구수, 학군을 모형에 포함시킨 후, 변수선택의 방법 중 단계별 회귀방법(stepwise regression)에 의해 최종 변수를 선정하였다. 단계별 회귀방법이란 전진선택법에 후진제거법을 가미한 방법으로 새로운 변수가 추가되면서 이미 선택된 변수가 중요성을 상실하여 제거될 수 있는지를 매 단계별로 검토하여 설명변수를 선택함으로써 최적의 다중회귀모형을 찾아가는 방법이다. 또한 집세와 연건평, 대지면적 변수는 분산안정화 및 선형관계로 변환을 위해 자연로그변환을 취해주었다. 이들 변수에 자연로그변환을 취한 뒤 모형의 잔차의 형태를 살펴보았는데 변환을 통하여 나온 결과가 좀 더 정상적인 형태를 나타내고 있어 자연로그변환이 바람직하다고 판단되었다. 그리고 설명변수가 범주형인 건축년도, 난방시설형태, 거주층 구분, 학군의 경우 주어진 자료값이 특성을 갖고 있으면 1, 아니면 0의 값을 부여하는 가변수(dummy variable) 형태로 만들어 모형에 포함시켰다. 여기서 거주층 구분 변수는 거주하는 층이 지상이면 1, 지하, 반지하, 옥탑, 옥상이면 0의 값을 갖는다.

이와 같은 방법에 의해 최종 채택된 회귀모형의 결과는 <표 4-5>와 같다. 일반단독 전세가구에 대한 모형은 총 20개의 설명변수 중에서 1학군여부와 2학군여부가 단계별 변수선택단계에서 제거되었으며, 나머지 변수들은 모두 유의수준 0.05하에서 유의미한 것으로 나타났다. 그

중 총사용방수(0.30), 지상여부(0.37), 00년대건축여부(0.43), 도시가스여부(0.51), 8학년여부(0.37)의 회귀계수가 높게 추정되었다. 즉, 해당되는 특성을 가지고 있거나, 특성 값이 증가할수록 집세수준이 높아짐을 알 수 있다. 또한 모형의 설명력을 나타내는 R^2 값이 0.5204로 반응변수의 총변동의 약 52%가 이러한 설명변수들이 도입됨으로써 설명되었다고 할 수 있다. 나머지 7개 모형에 대해서도 동일한 방식으로 해석할 수 있다.

〈표 4-5〉 주택유형 및 점유형태별 다중회귀모형 분석 결과
 일반 영영경용 단독주택 다가구 단독주택

설명변수	전세 [†]		월세		전세		월세	
	Parameter Estimate	Pr> t						
	절편							
총사용방수	5.57	<.0001	6.52	<.0001	5.83	<.0001	6.57	<.0001
연건평	0.30	<.0001	0.22	<.0001	0.28	<.0001	0.18	<.0001
ln대지면적	0.13	<.0001	0.09	<.0001	0.11	<.0001	0.09	<.0001
개상여부)	0.08	<.0001	0.08	<.0001	0.07	<.0001	0.08	<.0001
년대건축여부	0.37	<.0001	0.18	<.0001	0.36	<.0001	0.19	<.0001
00년대건축여부	0.43	<.0001	0.35	<.0001	0.40	<.0001	0.36	<.0001
90년대건축여부	0.21	<.0001	0.20	<.0001	0.18	<.0001	0.20	<.0001
80년대건축여부	0.15	<.0001	0.11	<.0001	0.06	<.0001	0.10	<.0001
도시가스여부	0.09	<.0001	0.07	<.0001	0.03	0.0002	0.06	<.0001
총거주가구수	0.51	<.0001	0.24	<.0001	0.39	<.0001	0.23	<.0001
1학년여부	-0.01	<.0001	-0.01	<.0001	removed		removed	
2학년여부	removed		removed		removed		-0.03	0.0011
3학년여부	removed		removed		0.06	<.0001	0.08	<.0001
4학년여부	-0.05	0.0016	-0.12	<.0001	-0.08	<.0001	-0.12	<.0001
5학년여부	-0.14	<.0001	-0.11	<.0001	-0.10	<.0001	-0.06	<.0001
6학년여부	0.20	<.0001	0.16	<.0001	0.20	<.0001	0.18	<.0001
7학년여부	0.13	<.0001	0.10	<.0001	0.17	<.0001	0.09	<.0001
8학년여부	-0.08	<.0001	-0.11	<.0001	-0.08	<.0001	-0.08	<.0001
9학년여부	0.37	<.0001	0.28	<.0001	0.44	<.0001	0.37	<.0001
10학년여부	0.05	0.0001	0.03	0.0155	0.13	<.0001	0.03	0.0009
11학년여부	0.11	<.0001	0.10	<.0001	0.18	<.0001	0.11	<.0001
R-Square	0.5204		0.4057		0.5262		0.4618	
관측값 수 Adj R-Sq	0.5197		0.4046		0.5260		0.4615	
	12,231		9,909		36,449		30,871	

〈표 4-5〉 주택유형 및 점유형태별 다중회귀모형 분석 결과(계속)

설명변수	연립주택				다세대주택			
	전세		월세		전세		월세	
	Parameter Estimate	Pr> t						
절편								
총사용방수	5.76	<.0001	7.03	<.0001	6.49	<.0001	7.04	<.0001
연건평	0.19	<.0001	0.15	<.0001	0.16	<.0001	0.12	<.0001
차량(여부)	0.43	<.0001	0.24	<.0001	0.29	<.0001	0.14	<.0001
70년대건축여부	0.44	<.0001	0.34	<.0001	0.36	<.0001	0.23	<.0001
80년대건축여부	0.46	<.0001	0.24	<.0001	0.47	<.0001	0.49	<.0001
90년대건축여부	0.33	<.0001	0.21	<.0001	0.26	<.0001	0.25	<.0001
80년대건축여부	0.24	<.0001	0.10	0.0166	0.15	<.0001	0.14	<.0001
90년대건축여부	0.22	<.0001	removed		0.08	<.0001	0.09	0.0032
총거주가구수	0.44	<.0001	0.15	<.0001	0.23	<.0001	0.18	<.0001
학군여부	-0.25	<.0001	-0.23	<.0001	-0.23	<.0001	-0.09	<.0001
학군여부	0.10	0.0038	removed		0.07	0.0003	0.08	0.0145
학군여부	removed		-0.12	0.0022	0.10	<.0001	0.10	<.0001
학군여부	removed		-0.23	<.0001	removed		removed	
학군여부	-0.52	<.0001	-0.19	<.0001	-0.04	0.026	removed	
학군여부	0.16	<.0001	0.12	0.0182	0.21	<.0001	0.23	<.0001
학군여부	0.24	<.0001	removed		0.41	<.0001	0.30	<.0001
학군여부	removed		-0.18	0.0012	removed		removed	
학군여부	0.48	<.0001	0.33	<.0001	0.53	<.0001	0.55	<.0001
9학군여부	0.14	<.0001	-0.15	0.0086	0.26	<.0001	0.16	<.0001
10	0.38	<.0001	removed		0.20	<.0001	0.17	<.0001
R-Square	0.5734		0.6686		0.5746		0.5197	
관측값 수	3,166		967		11,427		3,696	
Adj R-Sq	0.5713		0.6634		0.5740		0.5177	

이와 같이 추정된 개별모형들을 비교해보면, 먼저 주택유형과 점유 형태에 따라 8개의 모형에서 추정된 절편의 크기나 각 설명변수의 회귀 계수 추정값이 다르다. 그러나 일반단독주택과 다가구단독주택 간에, 그리고 연립주택과 다세대주택 간에 추정된 모형의 회귀계수의 값이나 각 모형에서 제거된 변수 등이 다소 유사하다. 총거주가구수는 다가구단독 주택 모형에서는 모두 제거되었으며, 일반단독주택에서 회귀계수 추정 값도 매우 작게 나타나 단독주택의 집세가격에 크게 영향을 미치는 요 인은 아닌 것으로 보이는 반면, 연립주택과 다세대 주택에서는 유의미

한 변수로 선택되었다. 또한 건축년도의 경우 최근년대에 지어질수록 회귀계수가 크게 추정되어 집세도 올라가는 구조를 보이고 있음을 확인할 수 있었다.

마지막으로 8개의 회귀모형을 구축하기 위해 세웠던 오차의 등분산성, 모형의 선형성, 오차의 정규성과 같은 가정에 문제점이 있었는지 진단해보았다. 이러한 가정들의 타당성을 알 수 있는 가장 보편화된 방법은 스튜던트화잔차를 t 축으로 하고 e 축으로 하는 산점도를 그려보는 것인데, 잔차의 형태로부터 모형의 가정이 바람직한지 여부를 알 수 있다(강명욱 외, 1996). 각각의 모형에 대한 잔차산점도를 그려보았는데, 지면 관계상 개별 산점도는 신지 않고 결과만 기술한다. 잔차산점도의 모양이나 t 축의 스튜던트화잔차의 흩어짐의 정도를 보았을 때 오차의 등분산성이 t 정규성 가정을 어느 정도 만족하고 있어 추정된 모형이 적합하다고 판단되었다.

개별모형으로부터 집세에 영향을 미치는 것으로 나타난 공통요인을 비교해보면, 모든 모형에서 총사용방수, 지상여부, 건축년도, 도시가스여부, 연건평이 설명변수로 선택되었으며, 추정된 회귀계수 값의 크기도 집세수준에 영향력이 있는 것으로 나타났다. 특히 학군의 경우 일부 모형에서 몇몇의 학군여부가 제거되었으나 8학군여부의 경우 모든 모형에서 추정된 회귀계수 값도 매우 크고, 모형 내에서 상당한 설명력을 가지고 있다. 학군은 행정구역 단위의 2~3개의 구가 지리적인 인접성에 의해 설정되어있는데, 이는 지리적 요소뿐 아니라 해당 지역의 생활수준이나 주거환경 등을 동시에 포괄하는 의미도 가지고 있기 때문에 모형 내 설명력이 높게 나타난 것으로 보인다.

3. 아파트

아파트의 경우 부동산 정보업체에 가격의 변동을 파악할 수 있는 시계열 자료 및 아파트 단지 별 특성정보가 잘 구축되어 있기 때문에 부동산 정보업체로부터 자료를 제공받아 분석하였다. 본 절에서 이용한 분석자료는 서울시 아파트에 대해서 2008년 12월 기준 평형별 단지의

특성 및 입지조건 정보와 그에 해당하는 2005년 1월부터 2008년 12월까지 48개월 간의 전세가격에 대한 부동산뱅크의 자료를 이용하였다. 이 자료는 부동산뱅크와 협력관계를 맺은 부동산중개업소에서 직접 가격 정보를 입력하는 방식으로 수집되며, 아파트 전세가격은 해당 아파트에 입주자가 시작되고 시세가 형성된 시점부터 매월 조사된다. 특히 부동산뱅크에는 2006년 기준의 지하철, 학교, 편의시설, 자연환경까지의 거리와 같은 입지조건 요인에 대한 자료가 구축되어 있어 분석변수에 추가적으로 포함시킬 수 있었다. 그러나 이는 최단거리에 대한 정보이며, 주변에 초등학교가 여러 개가 있더라도 단지로부터 가장 가까운 학교에 대한 거리를 의미하는 것으로 제한적으로 해석 가능하다.

전세가격은 상한가와 하한가가 조사되는데 이들의 평균을 구하여 전세금액으로 사용하였다. 또한 전세금액의 변동은 평형 단지별로 전월 대비 금월 가격의 변동비율(이하 변동률)을 구하여 사용하였다. 또한 전세금액을 공급면적(m^2)으로 나누어 주어 면적에 대해 전세금액을 표준화한 공급면적당 전세금액의 변수를 이용하였다. 다음으로 현관구조는 계단식, 복도식, 타원식, 혼합식, 기타로 분류되어 있는데, 이 중 계단식이 전체 평형단지의 약 70% 이상을 차지하고 있어서 계단식이면 1, 아니면 0의 값을 갖는 계단식여부로 지시변수화 하였다. 난방방식도 개별난방, 지역난방, 중앙난방으로 구분되어 있던 자료를 개별난방이면 1, 아니면 0으로 표현되는 개별난방 여부로 변환하였다. 또한 공공임대세대가 섞여 있는 아파트 단지에 대해서는 공공임대 세대수를 총 세대수로 나누어 아파트 단지별 공공임대세대 혼합률을 구하여 설명변수로 사용했다.

수집된 48개월의 자료는 전세금액과 변동만이 시간에 따라 변하고, 이를 설명하기 위한 단지 특성과 입지조건들은 시간이 지나도 변하지 않는다. 이러한 자료의 형태를 고려하여 아파트 집세의 변동에 영향을 미치는 요인을 탐색하기 위해 본 연구에서는 다음과 같은 두 가지 분석 방법을 이용하였다. 먼저 독립변수인 평형단지 전세금액 변동률에 대해 월별로 회귀모형을 각각 구축하여 설명변수의 유의성을 확인하고, 48개의 회귀모형에서 각 설명변수가 유의미한 변수로 선택된 횟수를 정리하였다. 이는 특정 시점에 발생된 변동에 대한 개별 요인들의 효과를 각

시점마다 확인하여 전체 기간 동안 지속적으로 영향을 미친 요인을 탐색하는 방법이다. 48개 모형에 대한 분석결과를 바탕으로 각 설명변수가 단계별 회귀분석에 의해 총 48개 개별모형에서 유의미한 요인으로 선택된 횟수가 <표 4-6>에 요약되어 있다. 유의수준은 0.05이다.

개별 모형마다 유의한 것으로 선택된 설명변수는 분석자료의 시점마다 다르다. 이 중 가장 많은 모형에서 선택된 변수는 공급면적당 전세금액으로 48개 모형에서 총 35번 유의한 변수로 선택되었다. 다음으로 학군 변수가 선택되었는데, 강남구와 서초구로 이루어진 8학군 여부가 30회, 노원구와 도봉구의 4학군 여부가 29회, 종로구, 중구, 용산구의 5학군 여부가 28회, 강동구와 송파구의 6학군 여부가 26회 집계 변동에 유의한 영향을 미친 것으로 나타났다. 또한 단지 특성을 나타내는 변수 중 주장복합여부가 24회, 개별난방여부가 22회, 입주경과년수가 21회 선택되었다. 다음으로 입지적 조건을 의미하는 거리에 대한 변수들 중 초등학교까지의 거리가 21회로 많이 선택되었고, 고등학교와 중학교, 전철역 거리는 모두 18회로 다른 변수에 비하여 중요성이 상대적으로 낮다.

<표 4-6> 월별 변동률에 대한 48개 다중회귀모형 분석결과

설명변수	유의한		설명변수	유의한	
	횟수	*		횟수	*
공급면적당 전세금액 만원		*	중학교거리		*
입주경과년수 ()	35		고등학교거리(m)		18
계단식현관여부	21		편의시설거리 (m)		18
개별난방여부	17		1학군여부 (m)		16
공공임대세대혼합률	22		2학군여부		15
방 수	10		3학군여부		22
육질 수	12		4학군여부		21
주장복합여부	14		5학군여부		29
세대 수	24		6학군여부		28
세대별 주차대수	20		7학군여부		26
공급면적	18		8학군여부		21
전철역거리	16		9학군여부		30
자연환경거리 (m)	18		10학군여부		14
초등학교거리 (m)	13		유의수준		25
(m)	21		*	(0.05)	

다음은 평형단지별로 48개월 동안의 평균 변동률을 구함으로써 각 단지별 평균 변동수준을 측정한 후, 평형단지별 평균 변동률을 독립변수로 놓고 동일한 설명변수를 대상으로 단계별 회귀분석하였다. 이는 각각의 시점에서 일시적인 변동에 대해 지속적으로 영향을 미치는 요인을 탐색한 앞의 방법과는 달리 분석대상 기간 동안 평균적으로 발생한 변동률 수준을 설명하는 변수를 탐색하는 것이다. 즉, 48개월 동안 형성된 평형단지별 변동률 수준의 차이를 일으키는 데 영향을 미치는 요인이 무엇인지를 하나의 회귀모형을 구축하여 파악한 것이다. 이때 설명변수인 공급면적당 전세금액은 48개월 동안 매월 변하는 값이기 때문에 48개월의 평균값으로 바꾸고, 입주경과년수도 입주년도로 바꾸어 모형에 포함시켰다.

평균 변동률에 대한 회귀분석을 실시한 결과 단계별 변수선택법에 의해 유의수준 0.05에서 선택된 변수와 추정된 회귀계수가 <표 4-7>에 정리되어 있다. 공급면적당 평균전세금액과 입주년도의 회귀계수가 양의 값으로 추정되었으며, 나머지 변수에 대해서는 모두 음의 값으로 추정되었다. 특히 6학군 여부의 회귀계수는 -0.2923으로 6학군인 강동구와 송파구에 포함된 평형단지의 경우 평균 변동률 모형의 절편이 95.6865로 다른 지역보다 낮게 추정됨을 알 수 있다. 또한 학군여부의 계수의 크기가 다른 설명변수에 비해 큰데, 아파트 단지가 속해있는 지리적인 위치가 평균 변동률 수준을 결정하는 데 큰 영향을 미치고 있다고 해석할 수 있다.

이 절에서 수행한 분석 결과를 정리해 보면, 이 두 방법은 단계적 회귀방법에 의해 다중회귀모형을 동일하게 구축하였지만, 독립변수를 다르게 설정하여 모형의 결과를 해석하는 데 있어 차별적이다. 앞의 방법은 일시적인 변동에, 뒤의 방법은 일정기간 동안의 평균적인 변동에 초점을 둔 분석 방법이었다. 이 두 방법에는 나름의 장·단점이 있겠지만, 두 방법에서 공통적으로 유의하게 선택된 변수를 아파트 전세금액 변동률에 영향을 미치는 요인으로 선택하도록 하겠다. 이에 따라 공급면적당 평균전세금액, 입주경과년수(입주년도), 난방방식, 주상복합여부, 초등학교까지의 거리, 학군(지역)과 같은 변수가 변동에 영향을 미치는 주요한 요인인 것으로 보인다.

〈표 4-7〉 48개월 평균 변동률에 대한 다중회귀모형 분석결과

설명변수	회귀계수 추정값	Pr> t
결편		
공급면적당 평균전세금액	95.97880	<0.0001
입주년도	0.00036	0.00028
개별난방여부	0.00223	0.00046
방 수	-0.02910	0.00385
주상복합여부	-0.03361	<0.0001
세대별 주차대수	-0.09601	<0.0001
자연환경거리	-0.03702	<0.0001
초등학교거리	0.00004	<0.0001
고등학교거리	0.00012	<0.0001
학군여부	-0.00006	<0.0001
학군여부	-0.09340	<0.0001
학군여부	-0.06244	<0.0001
학군여부	-0.09020	<0.0001
학군여부	-0.29230	<0.0001
학군여부	-0.05538	<0.0001
8학군여부	-0.19440	<0.0001
10	-0.21780	<0.0001
R-Square	0.099664	

제3절 일반주택 표본설계

이 절에서는 서울시 일반주택에 거주하는 임차가구의 표본추출을 위해 다양한 층화 기준을 설정해보고, 각 기준에 의해 추출된 표본의 추정 정도와 효율을 비교하여 최적의 표본설계를 제안하고자 한다. 단, 추출 틀에는 집세의 변동에 대한 정보가 없기 때문에 추출된 표본에 대해 변동 대신 집세수준을 추정하고, 추출된 표본과 추출틀의 구조를 비교함으로써 제안한 방법의 효과를 분석하는 것으로 연구의 범위를 제한한다.

1. 모집단 정의

일반주택의 집세 조사를 위한 모집단은 2005년 11월 실시된 인구주택 총조사 결과의 서울시 주거전용 일반주택에 거주하는 일반가구 중에서

총 1,208,855의 임차가구이다. 다음 <표 4-8>은 일반주택 표본설계를 위한 모집단을 정리한 것이다.

<표 4-8> 서울시 일반주택 임차가구 모집단

총 임차가구 수	일반 영업겸용		다가구				다세대주택	
	거주		거주		거주		거주	
	전세 가구	월세 가구	전세 가구	월세 가구	전세 가구	월세 가구	전세 가구	월세 가구
1,208,855	143,102	104,423	420,265	332,428	36,785	9,033	126,513	36,306

2. 표본설계

가. 표본추출틀 구성

서울시 일반주택의 집세조사 표본설계를 위한 표본추출틀(sampling frame)은 2005년 인구주택총조사 10% 표본 보통조사구 중에서 현실적으로 임차가구에 대한 조사가 가능한 조사구 내 10가구 이상의 임차가구가 거주하는 3,113개 조사구로 구성하였다.

<표 4-9> 구별 조사 대상 조사구 및 임차가구 수 현황

구	조사구수	임차가구수	구	조사구수	임차가구수
서울시			서대문구		
종로구	3,113	107,982	마포구	139	4,500
중구	66	1,997	양천구	132	4,675
용산구	44	1,368	강서구	114	3,670
성동구	85	2,731	구로구	150	4,273
광진구	102	3,928	금천구	129	4,149
동대문구	160	6,225	영등포구	96	3,280
중랑구	142	4,971	동작구	118	4,253
성북구	155	5,578	관악구	139	5,055
강북구	163	5,046	서초구	221	8,702
도봉구	119	3,856	강남구	89	3,416
노원구	80	2,389	송파구	115	4,846
은평구	58	1,796	강동구	170	6,701
	197	5,615		130	4,962

정확한 표본선정 및 추정을 위해 자료의 오류로 판단되는 연건평이 주택총방수보다 더 작은 가구는 제외시키고 이를 조사모집단으로 정의한다. 이와 같은 정의에 의해 설정된 조사모집단에서의 구별 조사 대상 조사구 및 임차가구 현황은 <표 4-9>와 같다. 또한 조사의 편의성과 효율성을 위해 1차 표본추출단위(primary sampling unit: PSU)는 조사구로 하고, 2차 표본추출단위(secondary sampling unit: SSU)는 개별 가구로 하였다.

나. 층화

현행 집세조사에서는 서울시 전체에 대한 집세지수를 통계 공표 단위로 발표하고 있다. 따라서 이러한 통계 공표단위에 근거하여 서울시 전체에 대한 집세를 추정하기 위한 표본설계를 수행하였다. 서울시의 집세는 지역별 편차가 매우 심하고, 표본추출시 다양한 서울시의 특성을 반영해주기 위하여 다양한 기준에 의해 서울시를 여러 층(strata)으로 나누어 표본을 추출하고자 하였다. 첫 번째로는 '구'라는 행정구역 단위로 지역을 층화하였으며, 두 번째는 군집분석을 통하여 조사구를 층화하였다. 세 번째는 누적도수제공근 방법에 의해 조사구를 층화하였다.

1) 지리적 요인에 의한 층화 : 일반주택 층화1

우리는 앞에서 행정구역 단위인 구에 따라서 집세와 방1개당 집세에 차이가 있음을 확인하였다. 또한 이러한 2~3개의 구를 통합한 학군 단위에 대해서도 집세의 차이를 확인한 바 있다. 이렇게 지역에 따라 발생하는 집세수준의 차이를 감안하여 표본을 추출하기 위해 서울시를 총 25개 구로 층화를 하고, 각 층 내에서 1차추출단위를 추출하고자 한다. 이러한 층화작업을 통해 지리적으로 인접한 지역 내에서의 동질성은 가정하되 모든 지역에서 나타나는 특성과 변동을 설계에 모두 반영해줄 수 있다.

25개 구 단위로 형성된 각 층별로 1차추출단위인 조사구를 추출하기 위해 조사구의 특성에 따라 정렬순서를 지정해주었다. 이렇게 정렬을 하는 이유는 어떠한 특성에 의해 지나치게 치우친 표본이 추출될 가능

성을 줄이고 표본의 대표성을 제고하기 위함이다. 개별 조사구의 특성을 잘 설명할 수 있는 분류지표로 조사구별 방1개당 집세의 평균값과 조사구내 주택유형의 비율을 선택하였다. 집계된 조사구의 평균 방1개당 집세에 따라 각 층내에서 4분위수를 구하여 각 층별로 조사구를 4개의 그룹으로 분류하였다. 두 번째로 조사구내에서 단독주택 비율이 높은지에 대한 여부로 분류하였다. 각 층별로 설정된 분류기준에 의해 <표 4-10>과 같이 조사구명부를 배열하였다.

<표 4-10> 층내 분류기준 현황

차	집세	차	단독주택비율	차	조사구번호
1	:	조사구내	단독주택	3	:
조사구별 방 개당 집세 평균값 ≤ 각 층내 사분위수 1		비율	연립 다세대비율		:
		조사구내	단독주택		
조사구별 방 개당 집세 평균값 ≤ 각 층내 사분위수 2		비율	연립 다세대비율		:
		조사구내	단독주택		
조사구별 방 개당 집세 평균값 ≤ 각 층내 사분위수 3		비율	연립 다세대비율		:
		조사구내	단독주택		
조사구별 방 개당 집세 평균값 각 층내 사분위수 > 3		비율	연립 다세대비율		:
		조사구내	단독주택		
		비율	연립 다세대비율		:
		<	+		:

2) 군집분석을 이용한 조사구의 층화 : 일반주택 층화2

이번에는 지리적인 인접성에 근거하기보다는 개별 조사구에 대한 특성을 좀 더 고려하여 유사한 특성이 있는 조사구들을 군집분석을 통해 묶어 층을 구성해보고자 한다. 군집분석은 어떤 개체나 대상들이 가지고 있는 다양한 특성에 의하여 동질성을 지닌 군집으로 집단화하는 방법이다(성웅현, 1997). 본 연구에서는 조사구의 특성을 자세하게 파악하기 위해 여러 개의 층화변수를 고려하였다. Jarque(1981), 박진우 외

(2007), 박진우 외(2008)는 층화변수가 여러 개 있을 때 다변량 층화를 위해 군집분석(cluster analysis)을 이용하여 층화하는 것을 제안하였다. 또한 주택하위시장을 분석하는 연구에서 매매가격, 전세가격, 가격 변동률을 군집분석(cluster analysis)하여 하위시장을 구분하는 연구가 서울시 및 수도권을 중심으로 김범태(2000), 이성규 외(2004), 방창석(2007) 등에 의해 진행되었다.

조사구의 특성을 결정짓는 요인으로는 조사구의 평균 방1개당 집세와 조사구내 단독주택의 비율, 조사구내 2000년대에 지어진 신축주택의 비율을 선택하였다. 이와 같은 특성이 유사한 조사구들을 묶어 층을 형성하기 위해 k-평균군집분석을 실시하였다. k-평균군집방법(k-means clustering)은 계보적 군집방법(hierarchical clustering)에 비하여 초기에 잘못 배정된 개체를 반복을 통해 다시 군집화 하는 과정이 있다는 장점이 있다. 여기서의 k는 군집의 수를 의미하는데, 우선 최대 군집 수를 10개로 설정하였다. 이러한 방법으로 분석된 결과에는 10개의 군집 중 할당된 조사구의 수가 매우 적은 군집이 형성되었다. 이에 따라 개별 군집의 특성과 각 군집에 할당된 조사구 수 등을 고려하여 크기가 매우 작은 군집을 거리가 가장 가까운 군집에 통합시키는 과정을 거쳐 최종적으로 5개의 군집을 형성하였다.

이렇게 조사구를 최종 5개의 층으로 구분한 결과가 다음 <표 4-11>과 같다. 각 층의 특성을 살펴보면, 1번 층에는 982개의 조사구가 배정되었으며, 이 층에 배정된 가구들의 평균 방1개당 집세는 1,111만원이다. 또한 이 층에는 2000년대 건축된 주택에 거주하는 가구의 비율이 4%로 가장 낮으며, 단독주택에 거주하는 가구의 비율은 90%로 가장 높다. 또한 층번호가 올라갈수록 방1개당 집세의 평균은 높아지고, 2000년대 건축된 주택에 거주하는 비율이 높아지고, 단독주택의 비율은 낮아지면서 차등적으로 군집의 특성이 형성되었음을 알 수 있다. 이렇게 형성된 5개의 층에 어떠한 구들이 분포해있는지를 <표 4-12>에서 확인해보면, 강남구의 경우 집세수준이 가장 높은 층인 5번 층에 집중되어 있으며, 구내 변동이 가장 큰 관악구의 경우 5개 층에 골고루 분포하고 있다.

〈표 4-11〉 군집분석에 의해 형성된 5개 조사구 층

층 번호	조사구 수	방 개당 집세평균	가구 수	단위 개 만원)			
				1년대 건축주택 00 가구수	비율	단독주택 가구수	비율
1	982	1,111	32,149	1,298	0.04	28,884	0.90
2	948	1,407	31,920	2,274	0.07	28,745	0.90
3	574	1,756	19,292	2,396	0.12	17,167	0.89
4	286	2,216	10,356	2,653	0.26	8,263	0.80
5	323	3,309	14,265	6,023	0.42	11,775	0.83

〈표 4-12〉 구별 층 배정 현황

구 \ 층 번호	1	2	3	4	5
중로구					
중구	10	20	7	13	16
용산구	3	14	12	10	5
성동구	12	33	9	24	7
광진구	21	28	7	42	4
동대문구	6	61	35	51	7
중랑구	40	25	8	62	7
성북구	90	17	1	47	0
강북구	71	20	12	50	10
도봉구	89	3	1	26	0
노원구	63	3	0	14	0
은평구	30	6	2	18	2
서대문구	96	14	6	81	0
마포구	47	25	9	45	13
양천구	14	38	14	44	22
강서구	75	6	5	28	0
구로구	57	17	3	72	1
금천구	63	13	2	51	0
영등포구	51	12	3	29	1
동작구	36	25	4	53	0
관악구	25	43	11	42	18
서초구	56	42	20	59	44
강남구	1	14	31	2	41
송파구	1	7	20	2	85
강동구	7	38	51	36	38
	18	50	13	47	2

3) 누적도수제공 방법을 이용한 조사구의 층화 : 일반주택 층화3

층화변수가 일변량인 경우 층의 경계점을 정하는 방법으로 누적도수 제공근(the cumulative)법이 가장 널리 알려져 있다. 조사구별 평균 방1개당 집세를 층화변수로 선택하고, 층의 개수는 집세에 대해서는 앞의 층의 개수와 동일하게 5개로 구성하였다. 먼저 <표 4-13>에는 조사구 방1개당 집세 구간에 대한 조사구 수가 나와 있다. 근사적인 최적 층화방법은 누적제공근 값을 등간격으로 분리하는 것인데 $(181.78)/5=36.36$ 이므로 첫 번째 층경계는 가능한 36.36이 되어야 하고, 두 번째는 $36.36*2=72.72$ 가 되어야 한다. 실제 누적제공근을 보면 27.44가 36.36에 가장 가깝고, 66.11이 72.71에 가장 가까우므로 이러한 규칙에 의해 층의 경계를 정해서 5개의 층으로 조사구를 최종 구분하였다.

<표 4-13> 누적도수제공 방법에 의한 층 경계 현황

조사구 평균 방 개당 집세	조사구 수	도수제공근	누적도수제공근	층 경계
1	()			
0~1000	159	12.61	12.61	1
1000~1100	220	14.83	27.44	
1100~1200	393	19.82	47.27	2
1200~1300	355	18.84	66.11	
1300~1400	334	18.28	84.38	3
1400~1600	508	22.54	106.92	
1600~1800	327	18.08	125.01	4
1800~2000	217	14.73	139.74	
2000~2500	277	16.64	156.38	5
2500~3000	150	12.25	168.63	
3000	173	13.15	181.78	

1차추출단위인 조사구를 추출하기 위해 각 층내에서 1990년 이후 건축주택의 비율이 그 이전 건축된 비율보다 높은지 여부와 조사구번호에 따라 정렬하였다.

다. 표본크기 결정

새로운 표본설계를 위한 적정 표본규모의 산정방식은 크게 두 가지로 구분될 수 있다. 첫 번째로 계속 조사를 하는 표본조사의 경우 기존의 추정량에 대한 상대표준오차(CV_1)와 기존 표본규모(n_1), 그리고 새로운 설계에서 허용할 수 있는 목표오차(CV_2)를 고려하여 아래와 같은 관계식에 의해 새로운 표본규모(n_2)를 산정할 수 있다.

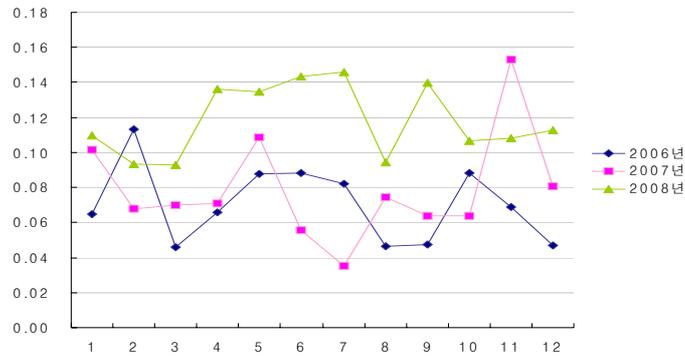
$$n_1 CV_1^2 = n_2 CV_2^2$$

두 번째로는 기존 조사가 없이 새로운 표본을 설계하는 경우 추정량에 대한 목표오차와 신뢰도를 이용하여 각 표본설계방법에 적합한 표본규모를 결정하는 방법이다. 층화추출의 경우 추정값이 확률 약 0.95로서 모집단 평균으로부터 추정 오차의 한계 B단위 이내에 놓여있어야 한다고 가정하면 모집단 평균을 추정하기 위한 근사 표본 크기는 아래와 같이 구할 수 있다. 여기서 N_i 는 i 번째 층의 크기, σ_i^2 은 i 번째 층의 모집단 분산을 나타낸다.

$$n = \frac{\sum_{i=1}^L N_i \sigma_i^2}{\frac{NB^2}{4} + \frac{1}{N} \sum_{i=1}^L N_i \sigma_i^2}$$

집세조사의 경우 기존의 조사가 수행되고 있기 때문에 앞서 설명했던 방법 중 첫 번째 방법에 의해 적정 표본규모를 구하도록 하겠다. 2006년 1월부터 2008년 12월까지 집세조사 자료를 이용하여 36개월에 대한 월별 집세 변동률의 상대표준오차를 분석하였다. 여기서 변동률이란 가구별 금월 집세금액을 전월 집세금액으로 나누고 100을 곱한 값을 의미한다. 기존 집세조사에서 집세 평균 변동률 추정값에 대한 상대표준오차의 크기를 [그림 4-1]에서 보면, 월별로 상대표준오차 크기의 차이가 매우 심하기 때문에 기존조사에서 월별 추정값에 대한 상대표준오차의 대표값을 하나로 정하기가 힘들다. 월별로 다르게 나타나는 상대

표준오차에 맞춰 매월 표본 규모를 정할 수는 없기 때문에 목표오차 내에서 안정적인 추정값을 구하기 위해 상대표준오차의 연평균과 3년 동안의 평균값을 구해보았다. <표 4-14>에서와 같이 2006년과 2007년의 평균 상대표준오차는 각각 0.71%, 0.78%이며 2008년에는 이보다 높은 약 1.12%의 수준을 보이고 있어 해마다 다른 수준을 보이는 것을 알 수 있다. 이에 따라 기존표본에 대한 상대표준오차의 크기를 3년 평균인 0.89%로 놓고, 서울시 일반주택에 대한 표본크기를 구하였다.



[그림 4-1] 일반주택의 월별 변동률 평균의 상대표준오차(%)

<표 4-14> 일반주택의 월별 평균 변동률의 상대표준오차(%) 연평균 값

구분	2006	2007	2008
연평균			
연최고치	0.0705	0.0787	0.1181
년평균	0.1130	0.1530	0.1459
3	0.0891		

<표 4-15> 일반주택의 목표 상대표준오차에 따른 예상 표본규모

예상표본규모	0.1	0.125	0.15	0.2	0.3
CV(%)					
예상표본조사구수	1,136	728	504	280	128
	142	91	63	35	16

2008년 집세조사에서는 일반주택에 대해 매월 약 1,429가구를 조사 하였으며, 목표오차의 수준을 다르게 하였을 때 예상되는 표본의 수를

<표 4-15>와 같이 구해보았다. 목표오차를 0.1%로 하였을 때, 예상 표본 가구 수는 1,136가구이며 0.15%로 하였을 때는 504가구의 표본가구가 필요하다. 목표오차를 설정하는 명확한 기준은 없으나, 통계청의 다른 가구대상 조사들에서는 전국의 목표정도는 1% 이하, 시·도 단위에서는 서울과 경기 2% 내외, 광역시는 4% 내외, 기타 도는 5% 내외로 설정한다. 그러나 집세조사는 조사대상 월에 계약갱신이 일어나는 표본 가구 중 집세금액에 변화가 발생해야 변동에 반영되는 집세조사만의 특수한 상황으로 인하여 변동률의 상대표준오차가 매우 낮게 나타나고 있어 일반적인 기준을 적용하기가 어렵다. 이에 따라 본 연구에서는 현행 집세조사의 표본규모와 기존의 추정량에 대한 상대표준오차의 수준 등을 감안하여 서울시에 대한 목표정도를 0.1%로 설정하고 표본규모를 산정하였다.

일반적으로 집락 안에는 이질적인 특성을 갖는 개체들로 구성이 되었을 경우 집락추출의 효과를 기대할 수 있다. 그러나 본 설계에서 조사의 편의성을 위해 구성된 조사구라는 집락은 주택의 특성과 집세의 수준과 변동 등에 있어서 유사한 특성을 보인다. 따라서 설계 효율이 떨어지므로 조사구 내 표본 수를 조절하여 다양한 특성의 조사구를 추출함으로써 설계의 효율을 높이고자 한다. 그러나 조사구의 수를 무조건 늘리면 조사에 투입해야 하는 비용과 시간이 더 많이 늘어나는 문제점이 발생하므로 이들을 적절한 수준에서 조정해 줄 필요가 있다. 현행 집세조사의 일반주택에 대해서 조사구당 평균 10가구 정도 조사가 되고 있는데, 이와 같은 사실을 감안하여 각 조사구별로 8가구만을 조사하는 것으로 설정하고 서울시 전체에 대하여 142개의 조사구에서 1,136가구를 조사하는 것으로 최종 결정하였다. 조사구 하나를 추가하였을 때 증가되는 실질적인 비용을 고려하여 한 조사구 내에서 조사되는 가구의 수가 어느 수준이 적정한지에 대해서는 추가적으로 검토될 필요성이 있다.

이렇게 서울시에 대해 결정된 표본수를 각 층별로 배분해주기 위해 각 층의 분산이 동일하다는 가정 하에서 층별 조사단위의 수에 비례하게 표본을 배정하는 비례배분법을 3가지 층화방법에 모두 적용하였다. 각 층화방법에 대해서 층내 임차가구의 수에 비례하여 표본을 배정하였으며, 각 층화방법에 대한 층별 표본 배분 현황이 다음 <표 4-16>과 같다.

〈표 4-16〉 일반주택 층별 표본 배분 현황

층화 1: 비례배분				층화 2: 비례배분	
층	조사구수	층	조사구수	층	조사구수
전체		서대문구			
종로구	142	마포구	6	1	45
중구	3	양천구	6	2	43
용산구	2	강서구	5	3	26
성동구	4	구로구	6	4	13
광진구	5	금천구	5	5	15
동대문구	8	영등포구	4	층화 3: 비례배분	
중랑구	7	동작구	6	층	조사구수
성북구	7	관악구	7	1	17
강북구	7	서초구	11	2	34
도봉구	5	강남구	4	3	39
노원구	3	송파구	6	4	25
은평구	2	강동구	9	5	27
	7		7		

라. 표본 추출

1차단위인 조사구 추출을 위해 먼저 각 층화방법에서 결정된 정렬방법에 의해 조사구를 정렬시킨다. 이어 각 층에 배정된 표본 수만큼의 조사구를 조사구별 임차가구수의 크기에 비례하는 확률로 확률비례계통추출(PPS systematic sampling)하였다. 이렇게 선정된 조사구에서는 점유형태(전세/월세), 사용방수, 거주층구분(지상/옥탑·지하)으로 가구를 정렬한 후, 각 조사구에서 8가구씩을 계통추출(systematic sampling)하여 표본가구를 최종 선택하였다.

3. 추정

일반주택 표본은 자체가중(self-weighting)이 되도록 설계하였기 때문에 각 층 내에서 표본 가구의 기본 가중값은 동일하다. 각 가구에 대한 설계가중값은 다음과 같이 구해진다.

$$w_{hij} = \left(n_h \times \frac{M_{hi}}{M_h} \times \frac{8}{M_{hi}} \right)^{-1} = \left(\frac{8 \times n_h}{M_h} \right)^{-1}$$

여기서, n_h 는 h 번째 층의 표본 조사구 수이고, M_h 는 h 번째 층의 임차가구 총수이며, M_{hi} 는 h 번째 층의 i 번째 조사구내의 임차가구 총수이다.

집세조사의 표본설계는 집세변동을 추정하기 위함이 목적이지만, 추출된 표본에 대한 시계열 자료가 없기 때문에 변동을 추정할 수 없다. 집세변동에 영향이 큰 집세의 평균을 추정하고, 변동에 영향을 미치는 것으로 보이는 요인들에 대해서 모집단의 구조와 유사한 구조를 보이는지를 확인하는 것으로 표본설계에 대한 타당성 및 효율성을 확인해보고자 한다.

먼저 각 방법에 의해 조사모집단으로부터 추출된 표본으로 평균 집세에 대해 추정해보도록 하겠다. <표 4-17>에 조사모집단의 평균 집세와 3가지 층화방법과 단순임의추출방법(SRS)에 의해 선정된 표본의 평균 집세 추정값과 표준오차를 구해보았다. 조사모집단의 집세 평균은 4,458만원이며 층화1~3번 방법 모두 95%신뢰구간 안에 조사모집단의 평균을 포함하고 있다. 또한 SRS로 추출된 표본의 추정값이 모집단의 값에 가장 근사하게 추정되었으며, 표준오차도 가장 낮음을 알 수 있다. 층화1~3번 방법은 모두 2단집락추출법을 적용하여 1차추출단위로 조사구를, 2차추출단위로 가구를 선정하였기 때문에 1차추출단위를 가구로 바로 선정한 단순임의추출에 비해 효율이 낮을 수밖에 없다. 각 층화 방법들의 효율을 비교해보면, 층화1번 방법의 표준오차가 112로 나머지 두 방법에 비해 설계효과가 낮게 나타나고 있다. 이는 층화2번과 3번 방법은 집세수준을 층화변수로 이용하여 층을 나누었기 때문인 것으로 보인다.

제4장

<표 4-17> 일반주택의 각 추출방법에 따른 집세 평균 추정값

구분	표본수	평균	신뢰구간 (95%)		표준오차	deff(%)
			하한	상한		
조사모집단						
층화	107,982	4,458				
층화 1	1,136	4,575	4,354	4,797	112	1.58
층화 2	1,136	4,339	4,136	4,541	103	1.45
3	1,136	4,379	4,194	4,564	93	1.32
SRS	1,136	4,455	4,316	4,593	71	

<표 4-18> 일반주택의 각 추출방법에 따른 표본의 분포 비교

항목 구분		조사 모집단	층화 1	층화 2	층화 3	SRS	집세 조사
점유 형태	전세	58.1	59.6	58.4	58.6	57.4	57.6
	월세	41.9	40.4	41.6	41.4	42.6	42.4
주택 유형	일반단독	20.3	21.4	19.8	22.0	19.7	85.7
	다가구단독	62.2	60.4	64.7	63.6	62.8	
	연립	3.6	2.6	4.9	2.9	4.0	14.3
	다세대	13.9	15.7	10.6	11.5	13.6	
가구 방수	1	10.8	10.6	12.3	9.7	10.8	20.1
	2	21.2	21.1	21.6	21.6	22.4	48.2
	3	43.2	43.1	42.8	43.6	41.2	21.8
	4 및 이하	24.8	25.2	23.3	25.1	25.6	9.9
집세	2800만 이하	25.4	24.5	26.0	26.2	26.7	18.7
	4000만 이하	28.5	28.2	30.9	26.9	28.2	34.3
	5500만 초과	21.7	22.8	21.8	23.4	20.1	20.8
	5500대	24.4	24.4	21.3	23.5	25.1	26.2
건축 년대	00년대	13.6	14.9	10.6	13.8	14.2	-
	90년	50.6	50.4	54.1	47.9	48.9	
도시가스	도시가스	35.8	34.7	35.3	38.3	36.9	
여부	그외	3.8	3.6	5.7	4.8	4.4	
집세조사	년 월 자료	96.2	96.4	94.3	95.2	95.6	-

(2005 11)

다음으로 <표 4-18>은 조사모집단의 구조와 각 층화방법에 의해 추출된 표본과, 2005년 11월 집세조사 표본의 구조를 비교한 것이다. 층화 1, 2, 3번 방법에 의한 구조가 조사모집단의 구조와 각 특성별로 유사한 구조를 보이고 있다. 또한 단순임의추출에 의해 선정된 표본의 조사모집단과 유사한 구조로 추출된 것으로 보인다. 집세조사의 표본 역시 점유형태나 주택유형에서 조사모집단과 유사한 구조를 보이지만, 집세변동과 관련이 높은 가구의 방수나 집세수준에 있어서는 조사모집단과는 차이가 있음을 확인할 수 있다. 2005년 11월 집세조사의 표본가구는 방

수가 1~2개인 주택에 거주하는 가구가 집중적으로 많이 분포하고 있으며, 집세가 2800만원 이하로 집세의 수준이 낮은 그룹의 표본 비율이 조사모집단에 비해 적게 분포하고 있다. 관심변수에 영향을 미치는 중요한 변수의 구조가 조사모집단에 비해 치우쳐져 있다는 것은 집세변동의 편향 발생 가능성을 의심해 볼 수 있다.

4. 38개 도시별 표본규모 검토

집세조사 결과에 대한 공표수준은 16개 시도에 대해서는 전세지수와 월세지수로, 38개 도시별로는 전세와 월세를 구분하지 않고 집세지수로 공표되고 있다. 이에 따라 표본규모를 결정하기 위해서는 38개 도시별로 표본규모를 검토해 봐야 한다. 본 연구에서는 서울시를 연구범위로 설정하여 표본설계과정을 기술하였지만, 실제로는 소비자물가조사 대상 전 도시에 대해 표본설계가 수행되어야 하기 때문에, 38개 도시별로 필요한 표본규모가 어느 정도인지를 확인해보고자 한다.

앞서 언급했던 바와 같이 표본규모를 산출하는 방법에는 크게 두 가지로 구분할 수 있는데, 기존 조사가 있는 경우엔 기존 조사결과를 이용하여 추정값의 상대표준오차를 기준으로 산정할 수 있으며, 기존 조사가 없는 경우엔 모집단의 분산과 목표오차를 기준으로 표본크기를 결정할 수 있다.

먼저 38개 도시에 대해 2006년 1월부터 2008년 12월까지 집세조사에서 일반주택 월별 평균 변동률의 상대표준오차를 구하고, 36개월의 평균을 구하여 기존조사에서의 도시별 평균 변동률에 대한 상대표준오차의 수준을 파악해보았다. <표 4-19>를 보면, 서울의 상대표준오차는 0.089%로 특·광역시 중 가장 낮은 수준이며, 울산을 제외한 다른 광역시는 약 0.12~0.18%의 범위 수준에 있다. 그 밖의 31개 도시 중 군산, 남원, 김해, 서귀포는 0.1% 이하로 매우 낮으며, 그 이외의 도시에서는 0.1~0.27%정도의 수준을 보이고 있었다. 기존 조사의 집세조사 평균변동률의 상대표준오차를 참고하여 목표 상대표준오차를 0.1~0.3으로 달리할 경우 요구되는 표본의 수를 <표 4-19>와 같이 구하였다.

〈표 4-19〉 38개 도시별 평균 변동률의 상대표준오차에 의해 산출한 표본규모

도시	집계조사평 균변동률의 상대표준오 차 년평균 (%)	2008 년 월별 평균 집세표본 수	표본 상대표준오차에 따른 예상 표본수				
			0.1	0.125	0.15	0.2	0.3
합계		6,388	15,599	9,983	6,933	3,900	1,733
서울	0.089	1,429	1,135	726	504	284	126
부산	0.120	474	686	439	305	172	76
대구	0.128	518	844	540	375	211	94
인천	0.183	355	1,185	758	527	296	132
광주	0.150	311	697	446	310	174	77
대전	0.127	293	473	303	210	118	53
울산	0.248	297	1821	1165	809	455	202
성남	0.276	81	613	392	272	153	68
의정부	0.169	106	303	194	135	76	34
안양	0.274	83	623	399	277	156	69
부천	0.191	98	358	229	159	89	40
안산	0.221	86	420	269	187	105	47
고양	0.288	99	823	527	366	206	91
춘천	0.173	113	339	217	151	85	38
원주	0.202	99	402	257	179	100	45
강릉	0.188	102	358	229	159	90	40
청주	0.147	70	151	97	67	38	17
충주	0.169	108	309	198	137	77	34
천안	0.200	68	274	175	122	69	30
보령	0.233	86	463	296	206	116	51
서산	0.200	51	204	131	91	51	23
전주	0.205	38	160	102	71	40	18
군산	0.184	75	254	163	113	63	28
담원	0.035	41	5	3	2	1	1
목포	0.089	95	74	48	33	19	8
여수	0.150	46	104	67	46	26	12
순천	0.131	56	97	62	43	24	11
	0.119	58	82	52	36	20	9

포항	0.169	86	246	157	109	61	27
경주	0.147	144	312	200	139	78	35
안동	0.124	70	107	68	48	27	12
구미	0.110	176	215	137	95	54	24
창원	0.192	109	399	255	177	100	44
마산	0.167	114	317	203	141	79	35
진주	0.181	104	339	217	151	85	38
제주	0.077	73	43	28	19	11	5
서귀포	0.155	130	315	202	140	79	35
	0.098	51	48	31	22	12	5

앞서 살펴본 바와 같이 기존의 집세조사 결과에서 38개 도시별로 상대표준오차의 크기가 매우 상이하여 도시별 일정수준의 목표오차 기준을 명확하게 설정하기가 어렵다. 특히 현행 집세조사의 상대표준오차가 다른 도시에 비해 매우 작은 군산의 경우 다른 도시의 목표정도와 동일한 기준을 적용하게 된다면 표본의 수가 매우 작게 계산이 된다. 목표정도를 0.1로 하였을 때 군산의 표본수는 5가구가 되는데, 이러한 표본규모에 의해 조사된 결과는 신뢰할 수 없을 것이다. 결국, 현행 집세조사의 추정값의 상대표준오차를 기준으로 표본크기를 산정하였을 때, 새로운 표본설계의 표본규모가 현행 표본 수와 조사결과에 매우 의존하게 된다. 현행 집세조사에서 상대표준오차가 다른 도시들에 비해 매우 낮은 몇몇 도시들의 경우 이들 지역에서 실제로 3년 동안 집세의 변동이 없었는지, 아니면 현재의 표본의 수가 매우 작아서 변동이 있는 가구를 포함할 수 없었는지의 여부는 명확하게 구분할 수 없기 때문에 이러한 정보만을 가지고 표본수를 결정하는 데는 다소 문제가 있을 수 있다.

또한 변동률의 경우 특정한 시점이나 외부요인 효과에 의해 특정 지역에서 일시적으로 상대표준오차가 커지거나 작아지는 현상이 발생할 수 있으며, 이러한 경우 안정적인 표본규모를 계산할 수가 없다. 따라서 본 연구에서는 그에 대한 대안으로 2005 인구주택총조사의 결과로 38개 도시별 평균집세의 오차의 한계를 이용하여 표본규모를 구해보았다. 집세의 변동과 수준 간에는 매우 높은 상관관계가 있고, 집세조사에서 집세항목을 조사하여 변동을 파악하므로 직접적인 목표변수가 아니더라도

도 개연성은 충분하다고 여겨졌다. <표 4-20>에 조사모집단에서의 도시별 가구 수와 집세평균, 그리고 목표오차를 100만원부터 300만원까지 다르게 설정하였을 때 필요한 표본수를 구하였다. 평균집세의 오차의 한계에 의해 표본 수를 구하였을 때는 앞의 경우와는 달리 38개 도시별로 일정 규모 이상의 표본수를 구할 수 있었다.

<표 4-20> 38개 도시별 집세평균 오차의 한계에 의해 산출한 표본규모
단위 만원

도시	조사모집단		목표 오차의 한계에 따른 예상 표본수 :)				
	가구수	평균집세 *	100	150	200	250	300
합계		-	100	150	200	250	300
서울	275,535		19,052	9,508	5,650	3,729	2,639
부산	107,540	4,463	2,337	1,052	594	381	265
대구	27,283	2,366	651	293	166	106	74
인천	17,907	2,811	643	291	165	106	74
광주	16,982	2,357	405	183	103	66	46
대전	6,788	2,219	421	194	110	71	49
울산	10,795	2,533	608	279	159	102	71
수원	6,582	2,278	544	254	145	94	65
성남	9,949	2,985	584	268	153	98	68
의정부	9,947	3,735	1,730	851	497	324	227
안양	2,685	2,825	422	205	120	78	54
부천	3,901	3,247	693	342	200	130	91
안산	6,574	3,059	778	370	214	138	97
고양	8,144	2,724	548	253	144	93	65
춘천	4,305	3,697	1,217	641	386	255	181
원주	1,949	2,142	259	124	72	47	33
강릉	1,399	2,134	266	132	77	51	36
청주	1,265	2,093	232	115	67	44	31
충주	4,410	2,220	459	217	125	81	56
전안	859	1,711	218	113	67	44	31
보령	2,424	2,520	359	174	101	66	46
서산	155	1,783	97	66	46	33	24
	395	2,282	249	170	118	84	63

전주	3,065	2,054	367	175	101	65	46
군산	663	1,639	417	285	198	142	105
남원	169	1,612	115	83	59	43	33
목포	814	1,802	308	173	108	72	52
여수	588	1,764	270	161	103	70	51
순천	620	1,989	230	129	80	54	38
포항	2,171	1,990	334	163	95	61	43
경주	919	2,023	387	225	141	96	69
안동	380	2,103	209	134	89	62	46
구미	1,932	2,480	395	198	117	76	54
창원	3,476	2,652	576	282	164	107	75
마산	2,835	2,064	341	162	94	61	42
진주	2,622	2,589	617	315	187	123	87
김해	1,759	2,113	315	155	91	59	42
제주	1,124	2,839	314	165	99	66	46
평균집세	1,601	2,495	413	218	141	97	65

* :
:

그러나 이와 같은 결과를 바탕으로 최적의 표본규모를 제안하기 위해서는 지역별로 목표정도를 다르게 책정해야 하는지, 그때의 목표정도는 어떠한 기준을 적용할 것인지 등에 대한 세밀한 검토가 필요하다. 또한 이론상 계산된 표본의 크기 뿐 아니라 비용, 인력, 시간과 같은 현실적인 측면을 종합적으로 검토할 필요가 있다.

제4절 아파트 표본설계

본 절에서는 아파트에 거주하는 임차가구의 집세를 조사하기 위한 표본설계방안을 제시하고자 한다. 아파트에 대한 표본설계도 일반주택의 설계과정과 유사하며, 층화2단집락추출을 하였다. 단, 아파트는 일반주택과는 달리 추출틀에 집세의 변동에 대한 정보가 있어 이를 활용하였으며 다양한 층화기준에 의해 추출된 표본에 대해서 전월대비 변동률

을 추정하고 각 방법에 따른 효율을 비교하였다. 또한 임차료에 대한 시세 형성이 전세에 대해서 이루어지고 있고, 월세의 경우 시세 데이터의 양이 부족하여 전세금액으로 집세를 정의한다.

1. 모집단 정의

아파트 집세조사를 위한 모집단은 부동산 정보업체의 최신 모집단 자료를 이용하기 위해 2008년 이전 입주한 서울시 주거전용 아파트에 거주하는 일반가구 중 임차가구로 정의하였다. 해당 시점에 대한 모집단의 크기를 헤아릴 수 있는 정보는 없지만, 2005년 11월 실시한 인구주택총조사의 결과를 이용하여 아파트 거주 임차 가구수를 파악해보면 다음 <표 4-21>과 같다.

<표 4-21> 서울시 아파트 임차가구 모집단

총 임차가구 수	전세가구	월세가구
430,291	308,145	122,146

2. 표본설계

가. 표본추출틀 구성

아파트 표본설계에는 부동산뱅크의 평형단지별 최신의 자료가 있기 때문에 인구주택총조사에서 설정된 조사구단위보다는 평형단지별 추출이 합리적이라 판단되었다. 따라서 아파트 추출틀로는 2008년 이전 입주한 아파트 평형단지 중 임차가구에 대한 조사가 가능할 것으로 보이는 평형단지별 세대수가 10세대 이상인 9,640개 평형단지를 표본추출틀로 구성하고, 이를 조사모집단으로 정의한다. 서울시 구별 아파트 조사대상 평형단지 수 및 세대수 현황은 <표 4-22>와 같다. 강남구, 서초구, 노원구 순으로 평형단지의 수가 많으며, 이를 세대수로 비교해보면, 노원구, 강남구, 송파구 순이다. 부동산정보업체의 자료만으로는 각 구별 임차가구의 규모를 파악할 수 없기 때문에 2005년 인구주택총조사 자료

로부터 각 구별 임차가구의 비율을 구하여 총 세대수에 임차비율을 곱하여 임차가구 수를 예상해보았다. 이 추출틀로부터 기대할 수 있는 임차가구 수는 약 418,802가구 정도이다. 아파트의 표본설계에서는 1차표본추출단위는 평형단지이며, 2차표본추출단위는 개별 가구로 하였다.

〈표 4-22〉 구별 조사 대상 단지 및 총 세대수 현황

구	평형 총		임차비율 *	예상임차가구수 **
	단지수	세대수		
종로구				
중구	140	8,927	0.33	2,938
용산구	66	10,762	0.52	5,583
성동구	332	25,413	0.39	9,787
광진구	277	39,501	0.37	14,434
동대문구	239	25,449	0.25	6,489
중랑구	301	37,236	0.29	10,969
성북구	313	35,631	0.29	10,293
강북구	340	50,439	0.30	15,218
도봉구	137	19,717	0.45	8,913
노원구	388	61,402	0.23	13,961
은평구	658	131,045	0.39	50,540
서대문구	221	16,978	0.23	3,859
마포구	264	29,145	0.31	9,000
양천구	395	39,985	0.42	16,601
강서구	654	63,351	0.35	22,456
구로구	548	62,609	0.45	28,183
금천구	455	56,082	0.22	12,488
영등포구	135	19,703	0.34	6,639
동작구	485	55,653	0.27	14,991
관악구	356	40,579	0.31	12,495
서초구	292	37,929	0.39	14,784
강남구	725	66,744	0.36	23,874
송파구	841	102,701	0.48	48,878
강동구	563	100,773	0.34	33,765
총합계	515	58,266	0.37	21,665
인구주택총조사 주거전용 아파트가구 일반가구 중 임차가구의 비율				
세대 가구 거주 가정	9,640	1,196,020		418,802

*2005

**1 1

나. 총화

아파트의 표본설계에서도 마찬가지로 서울시의 다양한 특성을 반영할 수 있도록 다양한 기준에 의해 서울시를 여러 개의 층으로 구분하였다. 첫 번째로 '구'라는 행정구역 단위로 지역을 층화하였고, 두 번째는 아파트 평형단지의 특성에 대한 군집분석을 수행하여 단지를 층화하였다. 세 번째는 평형단지의 규모에 의해 층화하고, 이러한 세 가지 층화 기준에 의해 각각의 표본을 추출하여 각 층화방법의 효율성과 다양성을 비교하고자 한다.

1) 지리적 요인에 의한 층화 : 아파트 층화1

앞에서 아파트 전세금액 변동 영향요인 분석결과에서 확인한 바와 같이 학군은 변동률을 설명하는 데 매우 중요한 변수이다. 즉 지리적 위치에 따라 발생하는 집세변동의 차이를 효과적으로 표본설계에 반영하기 위해 서울시 평형단지를 25개 구별로 층을 나누었다.

각 층내에서 1차추출단위인 평형단지를 추출하기 위해 층내 정렬 기준은 다음과 같은 변수를 이용하였다. 먼저 평형단지를 개별 단지의 평수에 따라 그룹을 5개5로 나눈 면적규모에 의해 정렬하고, 주상복합여부와 입주년도를 4개6의 그룹으로 나눈 입주년대에 의해 정렬한 후, 마지막으로 초등학교까지의 거리에 의해 명부를 정렬하였다.

2) 군집분석을 이용한 평형단지의 층화 : 아파트 층화2

아파트 평형단지를 유사한 특성이 있는 그룹으로 묶어주기 위해 공급면적, 공급면적당 전세금액, 입주경과년수, 주상복합여부의 변수를 이용하여 k-평균군집분석을 실시하였다. 먼저 최대군집 수를 10개로 설정하여 1차 군집을 형성하였으며, 10개의 군집 중 할당된 평형단지의 수가 매우 적은 군집은 가장 거리가 가까운 군집에 통합시키는 과정을 거

평 평 평 평 평
 —년대 이전 —년대 —년대 —년대
 5) ~18 , 19~27 , 28~35 , 36~42 , 43 ~
 6) 70 , 80 , 90 , 2000

쳐 최종적으로 5개의 평형단지 그룹을 형성하였다.

이렇게 형성된 5개의 평형단지 그룹이 층이 되며, 각 층에 대한 특성은 다음의 <표 4-23>과 같다. 1번 층에 속해있는 평형단지들의 평균 공급면적이 가장 작으며, 평균 입주경과년수 역시 16년으로 가장 오래 되고, 공급면적당 평균전세금액도 127만원 수준으로 다른 층에 비하여 가장 낮다. 또한 2번 층과 3번 층은 평균 공급면적이나 평균 입주경과년수의 수준이 비슷하지만, 공급면적당 평균전세금액이 3번층이 2번층에 비해 약 50만원 이상 높은 특성의 차이를 보이고 있다. 또한 5번 층은 공급면적당 평균전세금액이 약 277만원으로 다른 층에 비해 가장 높고, 주상복합의 비율이 0.29로 높다는 차별적인 특성을 보인다.

<표 4-23> 군집분석에 의해 형성된 5개 아파트 평형단지 층

층 번호	단지수	세대수	단위 만원 m ²			
			공급면적당 평균전세금액	평균 공급면적	평균입주 경과년수	(주상복합, 비율)
1	2,224	312,360	125.7	88.6	16.0	0.022
2	2,737	358,494	166.0	95.4	10.2	0.072
3	1,591	244,208	213.6	97.9	10.3	0.124
4	1,211	122,280	149.0	147.5	11.6	0.081
5	1,877	158,678	277.3	157.5	9.2	0.293

이렇게 다른 특성을 지닌 5개의 층들이 각 구별로 어떻게 분포하고 있는지를 <표 4-24>에서 확인해보면, 성동구와 동대문구, 성북구는 2번 층에, 도봉구는 1번 층에 노원구와 마포구, 강서구, 구로구는 1~2번 층에 많이 분포하고 있다. 반면, 용산구는 5번 층에, 서초구와 강남구는 5번 층과 3번 층에 집중되어 있음을 알 수 있다. 즉, 층을 형성하는 데 지역을 고려하지 않았지만, 층의 특성에 따라 자연스럽게 지역적 안배가 된 것으로 미루어 보아 층의 구분이 타당한 것으로 보인다.

1차 추출단위인 평형단지를 추출하기 위한 각 층내 정렬기준은 행정 구역단위인 구 코드와 입주년대, 초등학교까지의 거리를 이용하였다.

<표 4-24> 구별 아파트 평형단지 층 배정 현황

구	1	2	3	4	5
종로구	16	14	45	14	51
중구	10	9	30	5	12
용산구	37	47	68	22	158
성동구	6	119	82	42	28
광진구	9	54	103	25	48
동대문구	38	154	32	71	6
중랑구	137	127	2	47	0
성북구	74	133	57	73	3
강북구	83	30	3	21	0
도봉구	255	67	1	60	5
노원구	321	205	44	80	8
은평구	74	81	7	59	0
서대문구	77	70	43	57	17
마포구	35	142	118	47	53
양천구	158	138	149	33	176
강서구	222	208	27	85	6
금천구	192	151	30	74	8
영등포구	96	24	0	14	1
동작구	85	192	80	58	70
관악구	15	125	114	68	34
서초구	63	101	68	55	5
강남구	5	55	174	21	470
송파구	16	53	176	14	582
강동구	54	187	109	80	133
	146	251	29	86	3

3) 면적에 의한 평형단지의 층화 : 아파트 층화3

아파트 평형단지들을 면적에 의해 구분하고자 다음 <표 4-25>와 같이 5개로 층화하였다. 조사모집단에서 18평 이하의 평형단지는 505개로

층 세대수는 약 10만이며, 18평 이하의 평형단지 당 평균 세대수는 약 208세대이다. 또한 5번 층인 43평 이상의 평형단지는 1,974개로 이 평형 단지에는 약 16만8천 세대가 있고, 평형단지 당 평균 세대수는 약 85세대이다.

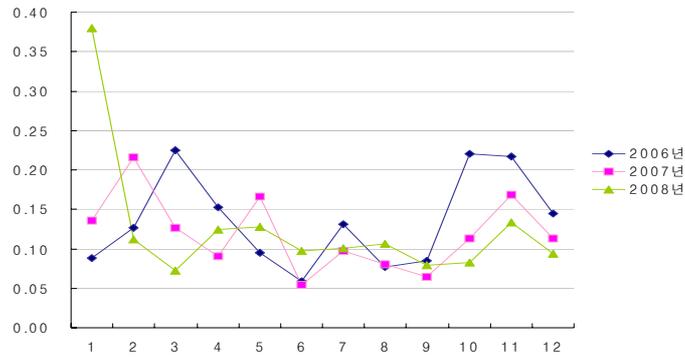
이렇게 규모에 의해 구분된 5개의 층 내에서 1차추출단위인 평형단지를 선정 시 관심변수들이 치우치지 않고 고르게 추출될 수 있도록 행정구역단위인 구 코드와 주상복합여부, 입주년대 그리고 초등학교까지의 거리에 의해 정렬한 후 1차추출단위를 추출하였다.

〈표 4-25〉 면적에 의한 층 경계 현황

층 번호	층경계 평 평()	단지수	세대수
1	~18	505	105,255
2	19~27	2,522	335,013
3	28~37	3,684	489,755
4	38~42	955	98,009
5	43 ~	1,974	167,988

다. 표본크기 결정

표본의 크기를 구하기 위하여 2006년부터 2008년까지 집세조사 자료를 이용하였다. 아파트의 월별 집세 변동률 평균의 상대표준오차(%)를 구하였는데, 여기서 변동률이란 가구별 금월 집세금액을 전월 집세금액으로 나누고 100을 곱한 값을 의미한다. [그림 4-2]를 보면, 월별로 상대표준오차의 크기에 차이가 있음을 알 수 있다. 월별 표본규모를 다르게 정할 수 없기 때문에, 월별 상대표준오차의 크기에 대한 연평균과 3년 평균값을 구해보니, <표 4-26>과 같았다. 2006년에는 0.135%, 2007년에는 0.1189%, 2008년에는 0.1257%로 해마다 차이가 있어 안정적인 표본규모의 산정을 위해 3년 평균값을 기존 조사에서의 상대표준오차로 이용하였다.



[그림 4-2] 아파트의 월별 변동률 평균의 상대표준오차(%)

<표 4-26> 아파트의 월별 평균 변동률의 상대표준오차(%) 연평균 값

연평균	2006	2007	2008
연최고치	0.1350	0.1189	0.1257
년평균	0.2249	0.2159	0.3796
3		0.1265	

<표 4-27> 아파트의 목표 상대표준오차에 따른 예상 표본규모

목표 상대표준오차(%)	0.1	0.125	0.15	0.2	0.3
예상표본크기(가구수)	800	512	356	200	88
예상표본단지수	200	128	89	50	22

2008년 집세조사에서는 아파트에 대해 매월 약 498가구가 평균적으로 조사되었다. 목표 상대오차의 값을 달리하면서 예상표본규모를 구해 보았다. <표 4-27>을 보면, 목표상대오차가 0.1%일 때 표본크기는 800가구였으며, 현행수준과 비슷한 0.125%로 목표상대오차를 정하면 표본크기는 512가구이다. 본 연구에서는 추정치의 정도를 제고하기 위해 기존의 수준보다 낮은 목표오차를 0.1로 설정하였다. 아파트의 경우 평형단지내에서는 주택의 특성이나 가격의 변동이 유사하므로 한 평형단지내에서 많은 가구를 추출하기보다 평형단지내 가구수는 제어하면서 다양한 특성과 변동패턴을 보이는 여러 평형단지를 추출할 필요가 있다. 따

라서 각 평형단지별로 4가구만을 조사하는 것으로 가정하고 단지 규모를 구하였으며, 200개 평형단지에서 800가구를 조사하는 것으로 최종 결정하였다.

다음으로 서울시에 대한 200개 평형단지를 각 층화 방법에 따라 층별로 배정해줘야 한다. 아파트 층화 1번 방법에 대해서는 25개 층별 임차가구 비율을 인구주택총조사로부터 구하여 예상임차가구수를 유추할 수 있었다. 이에 따라 층화 1번 방법은 비례배분법을 이용하여 각 층별 예상임차가구수에 비례하도록 표본수를 층별로 배분하였다. 또한 층화 2, 3번 방법에 대해서도 비례배분법을 적용하였는데, 예상임차가구수를 확인할 수 없었기 때문에 각 층별 세대수에 비례하도록 배분하였다. 각 층화방법에 대한 층별 표본 배분 현황은 다음 <표 4-28>에서 확인할 수 있다.

<표 4-28> 아파트 층별 표본 배분 현황

층화 1: 비례배분				층화 2: 비례배분	
층	단지수1	층	단지수	층	단지수
전체		서대문구			
종로구	200	마포구	4	1	52
중구	2	양천구	8	2	60
용산구	3	강서구	11	3	41
성동구	5	구로구	14	4	20
광진구	7	금천구	6	5	27
동대문구	3	영등포구	3	층화 3: 비례배분	
중랑구	5	동작구	7	층 3: 조사가구수	
성북구	5	관악구	6		
강북구	7	서초구	7	1	18
도봉구	4	강남구	11	2	56
노원구	7	송파구	23	3	82
은평구	24	강동구	16	4	16
	2		10	5	28

라. 표본 추출

1차추출단위인 평형단지를 추출하기 위해 먼저 각 층화방법에서 결정된 정렬방법에 의해 평형단지를 정렬시켰다. 이어 각 층에 배정된 표

본규모 만큼의 평형단지를 확률비례계통추출하였다. 다음으로 1차 선정된 평형단지 내에서 2차추출단위인 가구를 추출하기 위해 점유형태(전세/월세), 거주층 수로 정렬한 후, 각 평형단지에서 4가구를 계통추출하여 최종 표본을 선택하였다. 그러나 아파트 시세정보를 이용하는 현재의 추출틀에서는 평형단지내 전세금액은 동일하다고 가정하고 있어 각 세대별 정보는 모두 같다. 따라서 추출된 200개의 평형단지 내에서 4가구씩을 위의 방법으로 계통추출하였다고 가정하고, 추출된 평형단지별로 4가구씩 동일한 정보를 부여해 주어 총 800가구의 정보를 생성하였다.

3. 추정

아파트 표본은 층내 표본가구의 가중값이 동일한 자체가중설계이며, 각 가구에 대한 설계가중값은 다음과 같이 구해진다.

$$w_{hij} = \left(n_h \times \frac{M_{hi}}{M_h} \times \frac{4}{M_{hi}} \right)^{-1} = \left(\frac{4 \times n_h}{M_h} \right)^{-1}$$

여기서, n_h 는 h 번째 층의 표본 평형단지 수이고, M_h 는 h 번째 층의 예상임차가구수(총 세대수)이며, M_{hi} 는 h 번째 층의 i 번째 평형단지의 예상임차가구수(총 세대수)이다.

2008년 1월부터 12월까지 조사모집단에서 구한 평균 변동률과 3가지 층화방법과 단순임의추출방법에 의해 추출된 표본의 과거 시계열 자료를 이용하여 평균 변동률과 표준오차를 <표 4-29>와 같이 추정하였다. 먼저 층화 1~ 3번 방법은 모두 평형단지를 1차추출단위로 하여 집락추출을 하였기 때문에 집락추출하지 않은 SRS보다는 표준오차가 약 2배 가까이 높게 나타났다. 또한 층화방법에 따라 월별 표준오차의 수준이 다른데, 층화1번 방법이 표준오차의 수준이 다른 방법에 비해 평균적으로 가장 낮은 수준이며, 평균에 대한 95% 신뢰구간 추정값에 조사모집단의 참값을 가장 많이 포함하고 있는 것으로 나타났다.

〈표 4-29〉 2008년 월별 평균 변동률 추정값

월	조사									
	모집단	층화 1			층화 2		층화 3		SRS	
	평균 변동률	평균 변동률	표준 오차	평균 변동률	표준 오차	평균 변동률	표준 오차	평균 변동률	표준 오차	
1	100.49	100.59	0.174	100.65	0.204	100.73	0.212	100.56	0.088	
2	100.37	100.42	0.160	100.56	0.155	100.45	0.156	100.47	0.073	
3	100.66	100.83	0.208	100.46	0.140	100.56	0.143	100.50	0.068	
4	100.59	100.58	0.174	100.73	0.231	100.85	0.195	100.73	0.112	
5	100.45	100.36	0.098	100.49	0.135	100.29	0.126	100.70	0.096	
6	100.33	100.25	0.134	100.09	0.115	100.16	0.116	100.25	0.075	
7	100.35	100.34	0.179	100.46	0.244	100.39	0.203	100.22	0.069	
8	100.24	100.25	0.150	100.15	0.094	100.30	0.121	100.09	0.059	
9	100.04	99.95	0.083	100.07	0.159	99.89	0.107	100.10	0.079	
10	99.79	99.43	0.151	99.82	0.224	99.53	0.223	99.55	0.126	
11	99.46	99.34	0.167	99.38	0.158	99.27	0.181	98.98	0.113	
12	98.81	98.01	0.200	99.12	0.188	98.50	0.212	98.00	0.160	

를기 추정값을 그린 [그림 4-3]을 보면, 조사모집단에서의 월별 평균변동률과 층화 1~3번 표본에 의해 추정된 평균 변동률이 나와 있다. 세 방법에 의해 추출된 표본으로부터 구한 추정값이 전체적으로 조사모집단의 참값의 추세와 유사하게 구해지며, 월별 다소 차이는 있지만 전반적으로 층화1번이 조사모집단의 값에 가장 가까운 것으로 보인다.

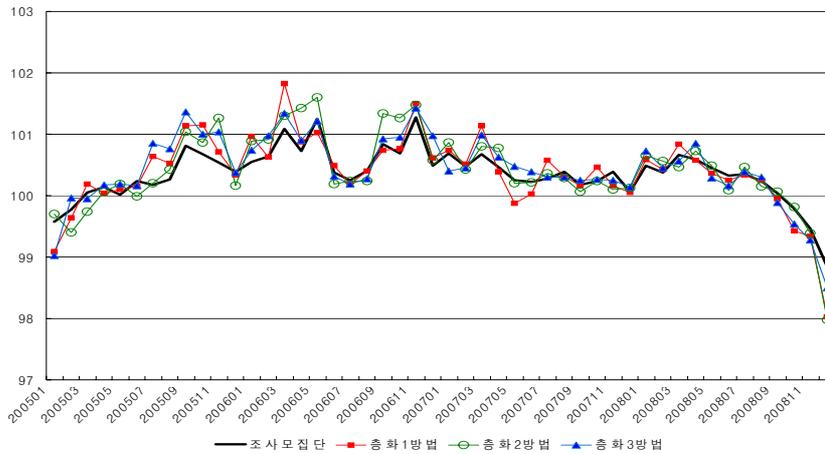
다음으로 월별 평균 전세금액을 추정하였다. 2005년 1월부터 12월까지의 월별 평균 전세금액의 12개월 평균값을 기준가격으로 설정하고, 이 기준가격과 월별 평균 전세금액과의 비를 구하여 월별 지수를 아래와 같은 식에 의해서 작성해 보았다.

$$I_t = \frac{t\text{시점의 전세 평균금액 추정값}}{2005\text{년 전세 평균금액 추정값(기준가격)}} * 100$$

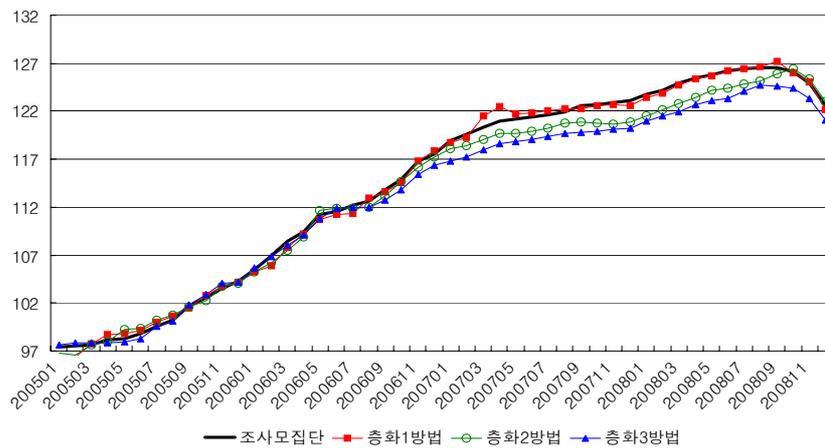
[그림 4-4]에서 보는 바와 같이 전체적인 추세는 각 층화방법 모두

$$7) \text{금월 변동률}(\%) = \frac{\text{전월 집세금액}}{\text{금월 집세금액}} * 100$$

유사한 경향을 보이고 있다. 이 중 증화1번 방법이 조사모집단의 지수를 가장 근사하게 추정하고 있으며, 증화2번 방법은 조사모집단보다 다소 높게, 증화 3번 방법은 다소 낮게 추정되고 있음을 확인할 수 있다.



[그림 4-3] 2005년 1월부터 2008년 12월까지의 월별 평균 변동률 추정값



[그림 4-4] 2005년 1월부터 2008년 12월까지의 월별 지수 추이

마지막으로 각 방법에 의해 표본으로 선정된 평형단지들의 평균적인 변동률의 분포를 조사모집단과 비교하고자 한다. 먼저 조사모집단의 모든 평형단지별로 2005년 1월부터 2008년 12월까지의 월별 평균 변동률을 구하고, 48개월 평균값을 구하였다. 이에 대한 서울시 전체의 4분위수를 기준으로 개별 평형단지를 4개의 그룹으로 구분하여 각 평형단지의 변동수준을 결정하였다. <표 4-30>에서 보는 바와 같이 1번 그룹에는 48개월 동안의 변동수준이 가장 낮은 평형단지들이 속해있으며, 4번 그룹의 평형단지들이 변동 수준이 가장 높다.

본 연구에서 선정된 표본의 타당성을 확인해보기 위해 층화 1~3번 방법과 SRS에 의해 표본으로 선정된 평형단지들의 변동 수준의 분포를 구하였다. 또한 2008년 12월에 집세조사에서 조사된 아파트 가구자료에 대해 아파트의 이름과 면적 정보를 이용하여 해당 단지를 조사모집단에서 찾아 집세조사의 표본 평형단지에 변동수준을 부여해주었다. <표 4-30>을 보면, 조사모집단에는 각 변동수준 그룹별로 평형단지들이 25%씩 고르게 분포하고 있다. 층화 1번 방법에 의해 선정된 평형단지들이 다른 층화방법에 비하여 선정된 표본들보다 각 변동수준 그룹에 가장 고르게 퍼져있음을 알 수 있다. 즉, 우리가 관심이 있는 변동수준에 대해서 어느 한쪽으로 치우치지 않고 고루 잘 뽑혀있다는 것은 현행 설계방법이 합리적으로 수행되었다는 것을 뜻한다. 반면, 집세조사에서 선정된 표본 평형단지의 경우 변동수준이 2~3번 그룹에 많이 분포하며, 평균 변동률이 가장 낮은 1번 그룹에는 약 4%만이 분포하고 있어 다소 치우친 분포를 보이고 있음을 알 수 있다.

<표 4-30> 변동수준에 의해 구분된 평형단지 그룹의 분포 비교

변동 수준	평형단지 구분기준	조사 모집단	총화			SRS	조사 *
			1	2	3		
1	평형단지의 개월 평균 변동률 ≤ 48	25.4	22.2	18.5	16.0	20.2	3.9
2	평형단지의 개월 평균 변동률 ≤ 48	25.3	25.3	22.5	21.0	26.3	40.0
3	평형단지의 개월 평균 변동률 ≤ 48	24.9	29.3	34.0	30.0	26.9	28.5
4	평형단지의 개월 평균 변동률 ≤ 48	24.5	23.2	25.0	33.0	26.6	18.5

단위 (집계: %)

* : 9.0%

4. 38개 도시별 표본규모 검토

현행 집계조사의 지수 공표는 주택유형에 따라 지수가 구분되어 있지 않다. 그러나 세분화된 통계작성 측면에서 도시별 아파트 지수를 작성하기 위한 표본수를 검토하고자 한다. 일반주택에서 살펴본 것처럼 38개 도시별 2006년 1월부터 2008년 12월까지 집계조사에서 아파트의 월별 평균 변동률의 상대표준오차를 구하고, 이를 36개월 평균하여 현행조사에서의 도시별 아파트 평균 변동률에 대한 상대표준오차의 수준을 파악하였다. <표 4-31>에서 도시별로 비교해보면, 울산은 다른 특·광역시와 비교하여 상대표준오차 수준보다 높은 0.201%이며, 나머지 특·광역시들은 0.1~0.16% 정도이다. 도시들은 대체로 0.2~0.3% 수준을 보이고 있으나, 안산, 서산, 전주, 군산, 여수, 안동, 마산, 김해, 서귀포에서 0.1% 미만이다. 또한 남원의 경우 2008년에는 아파트에 거주하는 표본가가 한 가구도 표본으로 선정되지 않은 것으로 나타났다. 즉, 아파트에서도 앞의 일반주택에서 살펴보았던 문제처럼 새로운 표본설계가 기존 조사 결과에 대한 의존도가 매우 높아 적절한 표본의 분배가 되지 않을 가능성이 크며, 이에 대한 좀 더 면밀하고 구체적인 해결방안이 연구되어야 할 것으로 보인다.

〈표 4-31〉 38개 도시별 평균 변동률의 상대표준오차에 의해 산출한 표본규모

도시	집세조사평 균변동률의 상대표준오 차 년평균 (%)	2008 년 월별 평균 집세표본 수	목표 상대표준오차 (%)				
			0.1	0.125	0.15	0.2	0.3
합계		2,982	8,070	5,165	3,587	2,017	897
서울	0.127	498	798	511	355	199	89
부산	0.127	242	389	249	173	97	43
대구	0.158	96	241	154	107	60	27
인천	0.164	187	502	322	223	126	56
광주	0.152	216	500	320	222	125	56
대전	0.101	225	227	145	101	57	25
울산	0.201	97	391	250	174	98	43
수원	0.248	58	357	228	159	89	40
성남	0.319	29	291	186	129	73	32
의정부	0.301	32	291	186	129	73	32
안양	0.348	50	601	384	267	150	67
부천	0.217	43	200	128	89	50	22
안산	0.000	13	0	0	0	0	0
고양	0.190	102	367	235	163	92	41
춘천	0.196	58	223	143	99	56	25
원주	0.099	86	84	54	38	21	9
강릉	0.123	55	83	53	37	21	9
청주	0.077	74	44	28	20	11	5
충주	0.279	61	477	305	212	119	53
천안	0.170	98	283	181	126	71	31
보령	0.371	30	414	265	184	104	46
서산	0.081	53	35	23	16	9	4
전주	0.074	36	20	13	9	5	2
군산	0.073	42	22	14	10	6	2
남원	-	0	0	0	0	0	0
목포	0.125	25	38	25	17	10	4
여수	0.098	70	67	43	30	17	7
순천	0.143	52	105	67	47	26	12
포항	0.123	36	55	35	24	14	6
경주	0.128	13	20	13	9	5	2
안동	0.047	44	10	6	4	2	1

구미	0.135	101	183	117	81	46	20
창원	0.120	64	91	58	41	23	10
마산	0.019	8	0	0	0	0	0
진주	0.482	21	487	312	217	122	54
김해	0.098	33	31	20	14	8	3
제주	0.228	27	139	89	62	35	15
서귀포	0.012	12	0	0	0	0	0

제5절 결론

1. 연구결과 요약

본 연구에서는 집세조사의 표본을 기존의 다목적 표본이 아닌 전용 표본으로 선정하기 위한 표본설계방법에 대해서 제안하였다. 먼저 집세 변동을 설명할 수 있는 요인이 무엇인지 실증분석을 통해 확인하고, 임차가구의 거주 주택유형에 따라 아파트와 아파트 이외의 일반주택으로 모집단을 분리하여 표본을 독립적으로 설계하였다. 효과적인 표본설계를 위해 집세변동과 관련이 높은 것으로 확인된 요인들을 각각의 설계 과정에 최대한 반영해주고자 하였다.

먼저 집세변동 설명요인 분석에서 일반주택의 경우 모집단에 대한 시계열 자료가 없기 때문에 인구주택총조사의 자료를 이용하여 변동 대신 집세수준을 설명할 수 있는 요인에 대해 분석하였다. 주택유형과 점유형태에 따라 8개의 회귀모형을 구축하고, 개별 모형으로부터 집세에 영향을 미치는 공통적 요인을 파악하였다. 총사용방수, 거주층구분, 건축년도, 도시가스여부, 연건평, 학군이 주요한 요인으로 나타났으며, 특히 임차가구의 거주 지역을 구분할 수 있는 학군변수가 회귀모형 내에서 높은 설명력을 갖고 있는 것으로 확인할 수 있었다.

아파트는 부동산뱅크의 아파트 평형단지의 특성정보 및 시세정보를 이용하여 아파트 전세금액의 변동을 설명할 수 있는 요인을 탐색하였다. 48개월의 자료에 대해서 매월마다의 변동률을 종속변수로 하여 48개의 회귀모형을 구축하고, 각 개별모형으로부터 유의성이 확인된 공통

요인을 파악하였다. 이렇게 각 시점마다의 일시적인 변동에 대한 분석 뿐 아니라 일정기간 동안 평균적으로 발생해온 변동에 대한 분석도 실시하였다. 평형단지별로 48개월 동안 발생해온 월별 변동의 평균수준을 종속변수로 하여 회귀분석을 수행하였다. 그 결과, 공급면적당 평균전세 금액, 입주경과년수, 난방방식, 주상복합여부, 초등학교까지의 거리, 학군이 변동에 영향을 미치는 주요한 요인으로 선택되었다.

다음으로 표본설계부문에서는 일반주택과 아파트에 대해 별도의 추출틀을 마련하고 독립적으로 표본을 설계하였지만, 설계원칙 및 층화 및 추출방법 등은 유사하게 진행하였다. 일반주택과 아파트 모두 층화2 단집락추출방법을 적용하였으며, 층내 자체가중을 만족할 수 있도록 설계하였다. 우선 서울시를 여러 개의 층으로 구분한 후, 각 층내에서 1차 추출단위인 집락을 확률비례계통추출하여, 각 집락내에서 2차추출단위인 가구를 사전에 정해진 일정 표본 수만큼 계통추출하였다. 이때, 1차 추출단위는 일반주택의 경우엔 조사구가 되며, 아파트의 경우는 평형단지가 된다. 층을 나누는 방법도 아파트와 일반주택 설계에서 실질적으로 이용된 변수는 다소 차이가 있지만, 지역적 요인에 의해 층을 구분하는 방법(층화1)과 여러 개의 층화변수를 고려하여 집락을 나누는 군집 분석을 통해 층을 구분하는 방법(층화2), 핵심이 되는 1개 변수만을 이용하여 층을 구분하는 방법(층화3)을 이용하였다. 이러한 층화방법에 의해 각각의 표본을 선정하여 추정값과 표준오차를 구하여 각 방법에 대한 설계효과와 추정정도, 분포구조 등을 비교하였다.

마지막으로 기존 조사 추정값에 대한 상대표준오차와 기존의 표본규모 및 목표오차를 이용하여 집세조사대상 38개 도시에 대한 표본규모를 산출하였다.

2. 시사점

층화추출에서는 층을 나누는 기준 변수나 층의 경계를 설정하는 등과 같이 어떠한 층화방법을 사용하였는가 하는 문제는 표본설계의 효율 향상과 추정값의 정도 제고에 기여하는 바가 크다. 본 연구에서는 우리가 관측하고자 하는 목표 변수와 상관성이 높은 요인들을 분석하여 각

기 다른 증화변수를 활용하여 3가지 증화방법을 제안하였다. 일반주택의 경우 지역적 층을 나눈 증화1번 방법이 집세를 증화변수로 활용한 증화 2, 3번 방법에 비하여 집세평균 추정시 낮은 효율을 보였다. 그러나 전세금액의 평균 변동률을 추정한 아파트에서는 지역적 증화방법이 다른 방법에 비해 높은 효율을 확인할 수 있었다.

이와 같은 결과는 일반주택의 집세의 평균 수준을 추정하는 데 있어서 목표변수가 증화변수로 직접 사용되었기 때문이기도 하지만, 변동률을 반영하는 데 있어서 다른 어떠한 변수보다 지역적 요소가 더 우선적으로 고려되어야 하기 때문으로 보인다. 따라서 지리적으로 인접한 층을 먼저 나누고, 그 안에서 관심변수들을 제어하는 것이 설계의 효율을 높일 수 있는 방법이라고 판단되며 증화1번 방법을 제안한다.

집세조사의 자료를 이용하여 일반주택과 아파트의 월별 평균 변동률을 분석하였는데, 일반주택의 변동률에 대한 상대표준오차가 아파트보다 낮게 추정되었다. 일반주택과 아파트의 목표정도가 0.1%로 동일하게 설정된다면, 일반주택은 현행 표본규모보다 적게, 아파트에서는 더 많이 조사되어야 한다.

본 연구를 수행하면서 모집단의 자료로 활용할 수 있는 자료를 수집하는 데 어려움이 있었다. 특히 일반주택의 경우 변동을 파악할 수 있는 모집단 자료가 없기 때문에 집세변동 대신 수준에 대한 설명요인을 분석하였고, 이를 표본설계 과정에 증화변수나 정렬기준 변수로 활용하였다. 집세의 수준과 변동 간에는 높은 상관성이 있기 때문에 표본설계에는 크게 무리가 없는 것으로 여겨지지만, 모집단 변동을 정확하게 파악할 수 없고, 일반주택에 대해서는 추출된 표본의 효율을 정확하게 비교할 수 없었다는 점은 이 연구의 한계점이라고 할 수 있다.

또한 집세조사대상 38개 도시에 대한 표본규모를 검토하는 과정에서 도시별로 3년 동안 변동이 거의 발생하지 않는 지역이 있어 기존조사의 상대표준오차만으로는 새로운 표본설계를 위한 적정 표본규모를 설정하는 데 무리가 있었다. 현행 표본에서 상대표준오차가 낮게 나타나는 지역이 해당 지역의 표본규모가 작아서 변동이 거의 발견되지 않은 것인지, 해당 지역에서 실제로 변동이 거의 발생하지 않았기 때문인지를 명확하게 구분할 수 없었다. 따라서 표본의 규모가 매우 작게 산출된 경

우, 그 지역의 수준의 평균에 대한 오차의 한계 등을 이용하여 기본적으로 유지되어야 하는 최소표본규모를 산정해주는 것이 하나의 방안이 될 수 있을 것으로 보인다. 또한 집세조사의 특성상 다른 일반적인 조사에 비해 상대표준오차의 크기가 작게 나타나고 있는데, 적정 표본규모를 산정하기 위해 필요한 목표오차의 수준을 어떠한 기준에 의해 설정할 것인지는 예산과 인력의 문제를 함께 고려하여 면밀하게 검토되어야 할 것으로 보인다.

마지막으로 본 연구에서는 설계가중값만을 이용하여 추정결과를 확인하였지만, 향후 추정값의 정도를 높이기 위한 다양한 추정방법이 검토되어야 할 것이다. 또한 집세조사의 계약갱신 현황 등을 검토하여 조사주기를 조정할 수 있는 방법과 조정된 조사주기에 따른 지수반영방법도 연구되어야 한다. 그리고 아파트와 일반주택의 지수를 별도로 작성할 수 있는 방법과 현행 집세지수 작성 방법에 대한 개선방안 등도 검토 될 필요가 있다.

참고문헌

- 강명욱 김영일 안철환 이용구 「회귀분석 모형개발과 진단」
 강재춘 (1996), 소비자물가통계에 있어 집세지수에 대한 연구 소비자
 물가지수에서 집세지수 개선 방안, 학위논문 (2006),
 김범태 서울시 주택하위시장의 권역설정, 학위논문 (2000),
 김현재 서울시 아파트 매 및 전세가격 결정요인의 분석, 「부동
 산학보」 제 22집 (2003),
 박진우 윤석훈 김진흠 정형철 군집분석을 이용한 다목적 조사
 의 증화에 관한 연구 「응용통계연구」 제 20권 제 2호 (2007),
 박진우 김영원 이석훈 신지은 다목적 표본조사를 위한 다변량
 증화 어업비계통생산량조사를 위한 표본설계 사례 「조사연
 구」 제 20권 제 2호 (2008),
 방창석 서울시 주택하위시장에 관한 연구 아파트시장을 중심
 으로, 학위논문 (2007),
 성응현 「응용 다변량분석 이론과 활용」
 신동훈 (1997), -의 회귀분석을 활용한 가격결
 정요인에 관한 연구, 학위논문 (2004),
 영민 김혜승 「셋집」 임대료 결정요인과 주거소비행태 「국
 토연 (1994)토연구원」, (2004),
 우경 김주경 1-10, 수도권 주택하위시장 분석에 관한 연구 「국토연
 구」 제 20권 제 2호 (2004),
 이성규 홍성언 4박수홍 평균 군집화 기법을 이용한 공시지가
 유사가격권의 설정(2004), 학위논문 (2004),
 이정화 지역계정에서의 주거서비스 산출액 추계방법에 관한 연
 구(2008)년 아파트 추계방법에 대한 연구를 중심으로 학위논문
 장한섭 공동주택가격결정요인에 관한 연구 일산을 중심으로
 학위논문 (2006),
 전백근 집세조사 방법론 추정법을 중심으로 고찰 「년도
 연구결과」 「모음집」 통계개발원 (2006), 2006
 , pp.275-321,

정근오, 조무상 (2009), 패널자료를 이용한 서울지역 아파트의 가격 결정요인 분석, 2009 경제학 공동학술대회 발표 논문집

정미옥 (2009), 집세조사의 표본 구조분석, 「국가 통계의 분류 수집 공표체계 품질향상을 위한 연구」, 통계개발원, pp.69-161

정성훈, 강준모 (2002), 아파트 전세 가격 결정요인 연구 수원시를 사례로, 「한국지역개발학회지」 제 14 권 제 2 호

함종범 (2002), 서울특별시 아파트단지의 특성이 아파트 가격에 미치는 영향분석, 학위논문

Jarque, C. M.(1931), "A Solution to the Problem of Optimum Stratification in Multivariate Sampling", Applied Statistics 30, pp.163-169.