

이학석사 학위논문

부산광역시 사회조사 표본설계
개선에 관한 연구



2011년 2월

부경대학교 대학원

통 계 학 과

조 한 나

이학석사 학위논문

부산광역시 사회조사 표본설계
개선에 관한 연구



2011년 2월

부경대학교 대학원

통 계 학 과

조 한 나

조한나의 이학석사 학위논문을 인준함.

2011년 2월



주 심 이학박사 장 대 흥 인

위 원 이학박사 윤 민 인

위 원 이학박사 노 맹 석 인

목 차

I. 서 론.....	1
II. 사회통계조사.....	4
1. 2008년 이전까지의 표본설계 방식.....	4
2. 새로운 표본설계의 필요성.....	6
III. 사회조사.....	7
1. 제안된 새로운 표본설계 방식.....	7
2. 제안된 구·군별 새로운 표본설계 결과.....	14
3. 제안된 추정방법 및 표본관리 지침.....	17
IV. 연구결과 및 해석.....	28
1. 새로운 표본설계 제안에 대한 평가.....	28
2. 구·군별 통계치 및 부산광역시 통계치 추정.....	41
V. 요약 및 결론.....	60
참고문헌.....	62

표 차례

<표 2-1> 2008년 이전까지의 표본추출방법	4
<표 2-2> 2008년 이전까지 방식에서 조사대상 제외가구	5
<표 3-1> 2005년 인구주택총조사 및 2006~2007년 신축아파트 조사구 기준 구·군별 조사구 현황	9
<표 3-2> 2005년 인구주택총조사 및 2006~2007년 신축아파트 조사구 기준 구·군별 조사대상 가구 수 및 가구원 수	10
<표 3-3> 분류지표의 표본조사구 추출 활용	11
<표 3-4> 3가지 분류지표에 따른 층화코드	14
<표 3-5> 층화코드에 따른 부산시 구·군 배분 현황	15
<표 3-6> 구·군별 가구, 가구원 수 및 구·군별 가중치	22
<표 3-7> 조사지침서 개선사항	26
<표 4-1> 2009년도 구·군별 응답가구의 가구원 수, 불응률	29
<표 4-2> 주택유형에 따른 주요항목(월 평균 소득)에 대한 비율	31
<표 4-3> 가구주 직업별(농림어가 비율)에 따른 주요항목(월 평균 소득)에 대한 비율	32
<표 4-4> 가구주 학력에 따른 주요항목(월 평균 소득)에 대한 비율	33
<표 4-5> 가구주 연령에 따른 주요항목(월 평균 소득)에 대한 비율	34
<표 4-6> 주택자가소유 여부에 따른 주요항목(월 평균 소득)에 대한 비율	35
<표 4-7> 5개의 분류지표 χ^2/df	36
<표 4-8> 구·군별 대체조사구 수, 대체율	37
<표 4-9> 3개 층화지표별 대체조사구 수, 대체율	38
<표 4-10> 표본조사구에서 대체된 조사구번호, 가구 수, 가구원 수	39
<표 4-11> 기장군의 설계 가중치 예시	42

<표 4-12> 구·군별 합계추정치(가구단위 조사).....	43
<표 4-13> 구·군의 총 가구 수 추정.....	44
<표 4-14> 구·군의 월 평균소득에 대한 평균(±표준오차), 효율 추정치.....	48
<표 4-15> 구·군별 합계추정치(조사원단위 조사).....	51
<표 4-16> 구·군의 총 가구원 수 추정.....	53
<표 4-17> 구·군의 부산시 지역경제 견해에 대한 평균(±표준오차), 효율 추정치.....	54
<표 4-18> 구·군별 가중치(가구단위, 가구원단위).....	57
<표 4-19> 부산광역시 총 평균(±표준오차), 효율 추정치(가구단위 조사).....	58
<표 4-20> 부산광역시 총 평균(±표준오차), 효율 추정치(가구원단위 조사).....	59



그림 차례

<그림 3-1> 표본조사구 내 조사가구 선정 예시.....	13
<그림 4-1> 주택유형에 따른 주요항목(월 평균 소득)에 대한 평균.....	32
<그림 4-2> 직업에 따른 주요항목(월 평균 소득)에 대한 평균.....	33
<그림 4-3> 가구주 학력에 따른 주요항목(월 평균 소득)에 대한 평균.....	33
<그림 4-4> 가구주 연령에 따른 주요항목(월 평균 소득)에 대한 평균.....	34
<그림 4-5> 주택 자가소유에 따른 주요항목(월 평균 소득)에 대한 평균.....	35



A study on improvement of sampling design for Busan metropolitan
social survey

Hanna Jo

*Department of Statistics, The Graduate School,
Pukyong National University*

Abstract

This study conducted the 『2009 Busan Social Survey』 to the samples who were chosen for a new sampling design, analyzed the result, found the derived problems, and estimated it. The purpose of the study is to prepare for a new sampling design of 『2012 Busan Social Survey』 which will be done by the Busan and which will be based on the 2010 Busan population and housing censuses 100% enumeration district as a sampling fraction.

The result of 『2009 Busan Social Survey』, the sample size as 1,068 households has $\pm 3.0\%$ point margin of error within general reliability 95% of country statistics under confidence level. It is good for Busan to use quinquennial population and housing census enumeration district and two-stage cluster sampling according as stratification variable within gugun like National Statistical Office. The result from comparing the 3 variables except Housing types and agriculture and forestry and fishing household proportion in 5 classification variables, it is right to use a householder level of education as the third stratification variable like sample modeling suggestion. Using one question in a household unit and a member of household unit, the mean estimate was similar to a sample mean, but the standard error estimate was large. However, the result compared with efficiency, the sample modeling was right to have efficiencies up to 80%.

The Busan should extract new sampling design which can reflect the suggested matters and solve the problems for the correct result.

Keyword : Busan Social Survey, sampling design, two-stage cluster sampling, stratified sampling, margin of error

I. 서론

『사회통계조사』는 시민생활의 질적 측면과 전반적인 복지 정도를 측정하고, 이를 지역사회 개발 정책의 기초자료로 활용함으로써 보다 나은 삶이 실현될 수 있도록 하기 위해 시행되는 조사로서, 『통계청의 사회통계조사』는 2005년에 실시된 『인구주택총조사』의 10% 표본조사구 결과자료를 모집단(population)의 기초자료로 사용하여 우리나라 전체 국민의 삶의 질과 사회적 관심사 및 주관적 의식에 관한 사항을 전국적으로 매년 조사하고 있다(통계청, 2009, 사회통계조사보고서).

지방자치단체가 중심이 되어 생산되는 지역통계(regional statistics) 작성조직이 정부의 위임통계 위주로 구성되어 계량적인 지역실정 통계를 작성하는데 어려움을 가지고 있다(안정용 등, 2005). 이에 지역여론을 능동적으로 수집하며 지역시민들의 의식과 삶의 질을 파악하기 위해 사회통계조사가 고려되며(최경호와 이강진, 2008) 지방자치단체들이 거의 매년 정기적으로 사회통계조사를 개별적으로 시행하고 있다.

부산광역시는 1996년부터 통계청과 중복되는 문항 없이 기존의 각종 조사에서 생산되고 있지 않는 각 부분에 대한 시민들의 주관적 의식 및 사회적 관심사와 삶의 질에 관련된 항목들을 조사시점에 부산광역시 내에 거주하고 있는 만 15세 이상 가구원을 대상으로 『부산의 사회통계조사』를 매년 실시하고 있으며(부산광역시, 2009, 부산의 사회지표) 통계라는 단어의 생소함으로 인해 조사시 거부감과 부담감을 느끼는 조사대상 가구원이 많아 2009년부터 『부산의 사회조사』로 명칭을 변경하여 실시하였다.

2008년도까지 실시된 『부산의 사회통계조사』는 주민등록 세대단위를 기준으로 하기 때문에, 조사시점 모집단(부산광역시 거주 15세이상 가구원)을 파악하는데 장기간의 검토시간이 소요되며, 『부산의 사회조사』는 전체 항목의 약 30% 정도를 구·군별로 특성화된 항목을 추가로 반영하여 16개 구·군별 표본설계가 필요하다. 또한 2008년도까지 적용된 표본설계 방식은 부산광역시만의 자체적인 방식이므로, 통계청이 생산하는 국가통계의 기본틀을 사용함으로써 부산광역시 통계의 표준화를 달성할 필요가 있다(부산광역시, 2009, 부산의 사회통계조사 표본설계 최종보고서).

마지막으로 『부산의 사회통계조사』는 『농어촌 주민의 삶의 질 만족도 조사』(승인번호 : 11438)와 같이 국가승인통계이나 표본설계에서 표본들이 불명확하고 추정 과정이 제시되어 있지 않으며 조사 결과의 정확성이 떨어질 것으로 판단된다(농림부, 2007, 농촌통계조사 표본설계). 이러한 이유로 『부산의 사회조사』 대상을 추출하는 새로운 표본설계를 필요로 하게 되었고, 부산광역시는 2009년도 2월에 과거 표본설계 방식과 『통계청 사회통계조사』에 적용된 표본설계 방식을 비교 검토한 후 개선 방향을 도출하고, 이러한 개선방향을 토대로 통계청이 생산하는 국가통계의 기본틀을 유지하면서, 새로운 표본설계를 수행하였다.

새로운 표본설계에서는 표본크기(sample size)를 구·군별 동일하게 1,068가구, 부산광역시 전체 17,088가구로 제안하고 있으며, 표본추출률(sampling fraction)은 통계청에서 사용하고 있는 2005년 기준 인구주택총조사의 부산광역시 20,065개 조사구에 2006~2007년도 신축아파트 조사구 586개를 추가한 총 20,651개 조사구를 사용하였다. 『경북인의 생활과 의식조사』의 경우 표본추출률로 인구주택총조사의 10% 표본조사구를(경상북도, 2007, 경북인의 생활과 의식조사 표본설계) 제안하고 있으며, 우리나라에서 대부분의 전국단위 또는 대형 가구조사의 경우 통계청에서 제공하는 인구주택총조사 90% 자료를 표본추출률로 사용하고 있다(강현철 등, 2009). 그러나 모집단의 편의를 방지하고자 새로운 표본설계에서는 부산광역시 조사구 100%를 표본추출률로 활용할 것을 제안하고 있다.

조사모집단(survey population)은 조사대상 가구의 만 15세 이상 가구원만을 제안하고 있으며, 조사대상 제외가구의 정의는 『통계청의 사회통계조사』의 정의를 따른다. 분류지표는 조사구를 주택유형에 따라 4분류, 농림어가 비율(강서구, 기장군에 국한)에 따라 2분류, 가구주 학력에 따라 2분류, 경제활동가능 가구주 연령에 따라 2분류로 분류하고, 주택의 자가소유 여부에 따라 2분류로 제안하고 있다. 인구주택총조사 결과 생산된 조사구 100%와 신축아파트조사구를 활용하여, 16개 구·군별로 107개의 표본조사구를 층화코드 분류에 따라 층내 조사구 수에 비례하여 배분한 후, 층내에서는 읍·면·동 행정구역 순서로 부여되어 있는 조사구 번호의 순서에 따라 표본조사구를 계통추출한다. 새로운 표본설계에서 조사구내 조사대상 가구 수는 『통계청의 사회통계조사』의 20가구보다 적은 10가구를 추출할 것을 제안하고 있으며, 철거 혹은 철거

예정과 같은 조사구 교체사유에 대비하여, 구·군별로 107개의 표본조사구에 1대1로 대응되는 107개의 예비표본조사구를 추출된 표본조사구의 동일 층화코드 내에서 추출하였다(부산광역시, 2009, 부산의 사회통계조사 표본설계 최종보고서).

새로운 표본설계방식의 표본추출틀로 2005년도 인구주택총조사의 조사구를 활용하였으므로, 통계청의 2010년 인구주택총조사의 결과공표 기한이 2011년일 때, 새로운 표본설계 방식은 2009년부터 2011년도까지 3년간 사용할 수 있으며, 2015년 인구주택총조사의 결과공표가 있기 전까지인 2012~2016년 5년간은 2010년도 인구주택총조사의 조사구 자료를 활용해서 표본설계를 해야한다.

따라서 본 연구의 목적은 2012년도부터 부산광역시에서 자체적으로 수행해야 할 사회조사 표본설계에 대비하여 기존의 표본설계를 설명하고 새로운 표본설계 방식을 제시하며, 실제 조사 후 나타난 결과를 여러 가지 방법을 통해 비교 분석하여 향후 계획 될 표본설계에 활용하는데 있다.

본 논문의 구성은 다음과 같다. 제 2장에서는 사회통계조사로 2008년 이전까지의 『부산의 사회통계조사』 조사개요와 표본설계 방식을 나타내며, 제 3장에서는 사회조사로 제안된 새로운 표본설계 방식과 세부적인 사항 및 새로운 표본설계 하에서 모집단의 특성을 추정하기 위해 주요항목 통계치에 대한 표준오차 등 특성분석 및 모수 추정절차를 위한 추정식을 나타내며, 제안된 표본관리 지침을 나타내었다. 특히 조사에서 무응답으로 인한 결측값이 있는 불완전한 자료인 경우 조사결과의 정확한 분석과 예측이 어렵게 되므로(도세록과 이관제, 2006) 결측값 발생시 처리방법 또한 제시되어 있다. 제 4장에서는 연구결과 및 해석으로 새로운 표본설계로 추출된 표본을 대상으로 실제 조사 후 변경된 사항과 조사결과를 추정한다. 제 5장에서는 본 연구의 결론 부분으로 연구 결과를 종합하고, 본 연구에 나타난 조사과정, 문제점 및 향후 시사점을 제시한다.

II. 사회통계조사

『부산의 사회통계조사』는 통계법 제15조, 제17조 및 제18조의 규정에 의거하여 승인된 일반통계(승인번호 : 제20205호)로써 사회지표 작성을 위하여 기존의 각종 조사에서 생산되고 있지 않는 각 부분에 대한 시민들의 주관적 인식 및 사회적 관심사와 삶의 질에 관련된 항목들을 조사하여 사회상태의 변동을 파악하고 사회개발정책의 기초자료를 제공하기 위한 목적으로 1996년부터 매년 실시되고 있는 조사이다(부산광역시, 2009, 부산의 사회지표).

본 장에서는 2008년 이전까지의 『부산 사회통계조사』에 적용되었던 표본설계 내용을 검토하고, 이를 토대로 새로운 표본설계 방식의 필요성을 도출하고자 한다.

1. 2008년 이전까지의 표본설계 방식

2008년 이전까지의 표본추출 방식으로 표본추출은 층화임의추출법(stratified random sampling method)과 계통추출법(systematic sampling method)을 활용하며, 다음 <표 2-1>과 같이 4단계로 이루어진다.

<표 2-1> 2008년 이전까지의 표본추출방법

단계	표본추출방법
1단계	<ul style="list-style-type: none"> • 16개 구·군별로 352가구로 동일하게 할당
2단계	<ul style="list-style-type: none"> • 구·군별로 「반(班)」을 거주지 특성에 따라 3개 층(주거지역, 아파트지역, 기타 지역(농업·어업·상업·공업 지역 등))으로 층화 • 88개 「반(班)」을 층 세대수에 비례하여 각 층에 할당
3단계	<ul style="list-style-type: none"> • 행정구역 순서(읍·면·동⇒통(統)⇒반(班))에 따라서 88개 표본 「반(班)」을 계통추출
4단계	<ul style="list-style-type: none"> • 구·군별로 총 88개의 표본 「반(班)」 내에서 4가구를 임의추출

가. 표본크기

2008년 이전까지의 표본설계 방식에서 조사대상 가구수는 각 구·군별로 동일하게 352가구, 부산광역시 전체에서 조사대상 가구수는 총 5,632가구(=16개 구·군×352가구)를 추출하고 있으며, 단순임의추출법(simple random sampling method)이라고 가정하였을 때, 95% 신뢰수준 하에서 구·군별 통계치의 오차한계는 ±5.2% 포인트, 부산광역시 전체 통계치의 오차한계는 ±1.3% 포인트이다.

나. 표본추출틀

표본추출틀로 조사시점에서 주민등록 세대단위를 기준으로 행정구역상 「반(班)」을 이용하였으며, 구·군별로 표본가구 수는 352가구(=88개 표본 「반(班)」 × 4가구)이다.

다. 조사모집단

조사모집단(survey population)은 조사시점(2008년 11월 10일~11월 20일)에 부산광역시에 거주하는 만 15세 이상 가구원이다. 이때, 가구원이란 함께 살고 있는 가구의 구성원을 말하며, 혈연관계가 없는 사람도 가구원이 될 수 있다.

조사대상 가구는 혈연가구와 혈연가족과 혈연관계가 없는 5인 이하의 남남끼리 한 집에서 생활을 하는 가구이며, 조사대상에서 제외되는 가구는 다음 <표 2-2>와 같다.

<표 2-2> 2008년 이전까지 방식에서 조사대상 제외가구

조사대상 제외가구	세부사항
1인가구	• 1인이 독립적으로 취사 취침 등 생계를 유지하고 있는 가구, 조사대상 가구원(15세 이상)수가 1인인 가구
비혈연 6인 이상 가구	• 혈연관계가 없는 6인 이상의 사람이 가구의 형태를 이루어 살고 있는 가구
집단시설 가구	• 기숙사, 교아원, 양로원, 수녀원, 대사찰 등 특정시설 내에서 혈연관계가 없는 사람들을 집단적으로 수용하여 생활을 같이하는 가구
외국인 가구	• 혈연관계와는 상관없이 외국인(외국 국적 소지인)들로만 구성된 가구

라. 조사항목

조사항목은 매년 공통된 조사 항목으로 주거형태, 주택점유형태의 표지사항 2개와 인적사항 및 경제활동사항에 관련된 18개 항목이며, 소득·소비, 교육, 환경, 주거·교통 등 3년 단위 조사항목으로 구성되어 있다.

마. 조사방법

조사방법은 조사원이 조사대상 가구를 직접 방문하여 면접 타계식 조사를 원칙으로 조사대상 가구의 요청에 따라 자계식 조사를 병행하고 있다.

2. 새로운 표본설계의 필요성

2008년 이전까지의 표본방식은 수집된 정보의 한계로 인하여, 「반(班)」의 거주지 특성만을 분류지표로 사용하여 거주지 특성을 3개층(주거지역, 아파트지역, 기타지역)으로 분류하여, 표본 「반(班)」을 계통추출하고 있어 통계치의 효율이 낮은 문제점이 있다. 따라서, 주택유형 뿐만 아니라 조사구별 다양한 특성을 표본추출에 반영함으로써 표본의 대표성을 높여 통계치의 효율을 최대화할 필요가 있다. 또한 2008년도까지 실시된 『부산의 사회통계조사』는 주민등록 세대단위를 기준으로 하기 때문에 조사시점 모집단인 가구원을 파악하는데 장기간의 검토시간이 소요되어 가구단위의 조사구 선정이 필요하다. 현행 방식은 부산광역시만의 자체적인 표본설계 방식이므로, 통계청이 생산하는 국가통계의 기본 틀을 사용함으로써 부산광역시 통계의 표준화를 달성할 필요가 있다.

따라서, 위와 같은 새로운 표본설계의 필요성으로 인해 2009년 2월 부산광역시는 새로운 표본설계를 수행하였다.

Ⅲ. 사회조사

2008년도까지 『부산 사회통계조사』라는 명칭으로 실시된 조사는 통계라는 단어의 생소함으로 인해 조사 시 거부감과 부담감을 느끼는 조사대상 가구원이 많아 2009년부터 『부산의 사회조사』로 명칭을 바꾸어 실시하였다.

본 장에서는 제안된 새로운 표본설계 방식에 대해 설명하고, 구·군별 새로운 표본설계 결과를 나타내며, 새로운 표본설계 하에서 모집단의 특성을 추정하기 위해서 주요항목 통계치에 대한 표준오차 등 특성분석 및 모수 추정절차 및 결과 작성을 위한 추정식을 나타낸다. 이를 위해 추정에 사용된 가중치를 계산하며, 추정오차를 계산하기 위한 분산추정공식을 나타내고, 마지막으로 제안된 표본관리 지침을 설명한다.

1. 제안된 새로운 표본설계 방식

제안된 새로운 표본설계에 따르면 통계청으로부터 2005년 『인구주택총조사』 자료인 ‘MDSS(Micro Data Service System) 가구’, ‘MDSS 가구원’과 ‘부산광역시 행정동 코드’, ‘외부제공불가조사구’, ‘신축아파트조사구’ 5개의 자료를 제공받아 표본조사구 및 예비표본조사구를 추출하였다.

새로운 표본설계 방식, 절차 및 각 절차별 세부사항으로 다음 7단계로 이루어진다.

- (1) 16개 구·군을 1,068가구, 107개 조사구로 구·군별로 층화한다.
- (2) 구·군별 조사구를 분류지표에 따라 층화코드를 부여 후 107개 조사구를 비례배분한다.
- (3) 할당된 층내 조사구를 계통추출한다.
- (4) 추출된 조사구의 가구수를 확인하여 40가구 미만인 조사구이면 인접조사구로 대체한다.
- (5) 추출된 조사구와 통계청 조사와의 중복 확인 후 중복이면 인접조사구로 대체한다.
- (6) 기타 조사구 교체사유가 발생시 인접조사구로 대체한다.
- (7) 표본조사구 거쳐번호 부여 후 시작거처로부터 10가구를 조사한다.(1,068가구를

추출하기 위해 107개 조사구 중 1개 조사구는 8가구를 조사하거나, 2개 조사구는 9가구를 조사한다)

가. 표본크기

일반적으로 표본크기의 결정을 위해서는 추정량의 $100(1-\alpha)\%$ 오차한계를 만족하는 표본크기를 결정하게 되는데(Yamane, 1967; Cochran, 1977), 구·군별 통계치의 오차한계를 모든 구·군에서 동일하게 설정하기 위하여 구·군별로 1,068가구, 부산광역시 전체 17,088가구로 제안되었다. 95% 신뢰수준 하에서 통계치의 오차한계는 구·군별로는 $\pm 3.0\%$ 포인트, 부산광역시 전체로는 $\pm 0.75\%$ 포인트가 된다. 구·군별 통계치 $\pm 3.0\%$ 포인트 오차한계는 통계청이 생산하는 국가통계 통계치의 일반적인 신뢰도(통계청, 2007, 통계품질가이드라인)로 각 구·군별로 1,068가구는 일반적인 사회조사에서 신뢰성을 확보할 수 있는 최소 수준의 표본크기이다.

(1) 구·군별 무한모집단일 때, 95% 신뢰수준 하에서 표본크기가 n 이라고 하였을 때, 구·군별 통계치의 오차한계(=d) 구하는 공식(성내경, 2003)은 다음과 같다.

$$d = 1.96 \times \sqrt{\frac{p(1-p)}{n}} = 1.96 \times \sqrt{\frac{1}{4 \times n}} \quad (\text{단, 표본크기 산출시 } p=0.5 \text{라 둔다})$$

따라서 $n = 1,068$ 가구 일 때 구·군별 통계치의 오차한계(=d)는 $\pm 3.0\%$ 포인트가 된다.

$$d = 1.96 \times \sqrt{\frac{1}{4 \times 1,068}} = 0.03$$

나. 표본추출률

표본추출률(sampling fraction)로 통계청에서 사용하고 있는 2005년 기준 『인구주택총조사』의 부산광역시 조사구 20,065개에 2006~2007년도 신축아파트조사구 586개를 추가하여 총 20,651개 조사구를 활용한다.

전체 조사구 수는 부산진구가 2,408개로 가장 많았고, 해운대구와 사하구의 조사구 수가 각각 2,355개, 2,011개로 부산진구 다음으로 많았다. 조사구 수가 적은 구·군으로는 강서구가 268개, 중구가 355개, 기장군이 448개였다.

<표 3-1> 2005년 인구주택총조사 및 2006~2007년 신축아파트조사구 기준 구·군별 조사구 현황

구 · 군	인구주택	신축아파트	전체	구 · 군	인구주택	신축아파트	전체
	총조사	조사구			총조사	조사구	
중 구	353	2	355	해운대구	2,266	89	2,355
서 구	813	4	817	사 하 구	1,972	39	2,011
동 구	642	23	665	금 정 구	1,475	83	1,558
영 도 구	885	1	886	강 서 구	257	11	268
부산진구	2,326	82	2,408	연 체 구	1,202	61	1,263
동 래 구	1,472	65	1,537	수 영 구	1,024	26	1,050
남 구	1,665	63	1,728	사 상 구	1,500	16	1,516
북 구	1,769	17	1,786	기 장 군	444	4	448
합 계					20,065	586	20,651

다. 조사모집단

제안된 조사모집단(survey population)은 조사대상 가구의 만 15세 이상 가구원만을 대상으로 하며, 조사대상 제외가구의 정의는 『통계청의 사회통계조사』의 정의를 따른다. 조사대상 가구는 이전까지 『부산의 사회통계조사』의 대상가구에서 1인 가구만 추가되었다.

새로운 표본설계의 부산광역시 조사대상 가구 수는 총 1,224,081가구이며, 조사대상 가구원은 총 3,553,749명(비혈연 6인 이상 가구원, 집단시설 가구원 포함)이다. 2005년 『인구주택총조사』 자료인 MDSS에서는 조사대상 가구인 일반가구의 가구 수와 가구원 수는 기재되어 있지만, 조사모집단인 15세 이상 가구원 수는 집단가구원이 포함된 15세 이상 가구원 수가 기재되어 있다. 통계청에서 제공받은 신축아파트조사구의 가구원수 정보는 결측으로 처리되어 있기 때문에, 가구원 단위 통계치 생산을 위해 신축아파트조사구와 가장 유사할 것으로 판단되는 신축아파트가 속해있는 구·군

의 아파트 규모에 따라 가구당 평균 가구원수를 가지고 지역적, 공간적 정보를 활용한 평균대체법(mean imputation method)을 활용하여 가구원을 처리하였다(Little과 Rubin, 2002).

각 구·군별 2005년 인구주택총조사 및 신축아파트 조사구 기준 조사대상 가구수와 조사대상 가구원 수는 다음 <표 3-2>와 같다. 2005년 『인구주택총조사』 결과에서는 총 가구원 수는 3,428,470명이고 집단가구원이 포함된 15세 이상 가구원 수는 2,909,331명으로 나타나, 2005년 기준으로 조사모집단 규모는 약 290만명으로 추측된다.

<표 3-2> 2005년 인구주택총조사 및 2006~2007년 신축아파트조사구 기준 구·군별 조사대상 가구 수 및 가구원 수

구 · 군	2005년 인구주택총조사 기준			2006~2007 신축아파트 기준		합계	
	가구	가구원	15세 이상 가구원 (집단가구원 포함)	가구	가구원	가구	가구원
중 구	21,010	51,478	45,633	169	477	21,179	51,955
서 구	47,668	128,781	113,861	252	759	47,920	129,540
동 구	37,450	101,310	89,532	1,440	4,573	38,890	105,883
영 도 구	53,668	151,798	135,291	42	128	53,710	151,926
부산진구	138,263	388,638	337,364	5,366	18,336	143,629	406,974
동 래 구	87,021	256,565	220,731	4,054	13,483	91,075	270,048
남 구	98,986	286,390	246,331	4,202	13,875	103,188	300,265
북 구	105,754	323,000	263,032	1,097	3,759	106,851	326,759
해운대구	131,106	385,617	317,075	5,732	19,228	136,838	404,845
사 하 구	117,935	350,680	294,493	2,605	8,455	120,540	359,135
금 정 구	87,109	249,353	218,809	5,179	17,026	92,288	266,379
강 서 구	15,315	43,121	39,852	671	2,388	15,986	45,509
연 제 구	70,836	206,574	178,037	4,006	13,571	74,842	220,145
수 영 구	59,857	168,880	146,743	1,645	4,969	61,502	173,849
사 상 구	90,051	266,157	224,882	996	3,452	91,047	269,609
기 장 군	24,349	70,117	37,665	247	811	24,596	70,928
부 산 시	1,186,378	3,428,470	2,909,331	37,703	125,279	1,224,081	3,553,749

라. 분류지표

적절한 분류지표를 설정하여 표본배분에 반영되는 층화변수 및 표본추출 결과의 대표성을 확인할 수 있는 체크지표로 활용한다면, 부산광역시 전체와 구·군별 특성을 반영하고 실효성 있는 통계를 산출할 수 있으며 이는 표본조사에서 매우 중요한 과정이다.

『통계청의 사회통계조사』에서는 주택유형, 농림어가비율, 가구주 학력, 경제활동가능 가구주 연령 비율을 분류지표로 사용하고 있다(통계청, 2009, 사회통계조사보고서). 따라서, 부산광역시 새로운 표본설계에서는 주택유형, 농림어가 비율, 가구주 학력, 경제활동가능 가구주 연령 비율, 주택의 자가소유 비율 총 5개의 지표를 분류지표로 활용하고, 조사구를 주택유형에 따라 4분류, 농림어가 비율(강서구, 기장군에 국한)에 따라 2분류, 가구주 학력에 따라 2분류, 경제활동가능 가구주 연령에 따라 2분류, 주택의 자가소유 여부에 따라 2분류로 분류한다.

2005년도 인구주택총조사 결과에서 생산된 조사구를 주택유형, 농림어가비율, 가구주 학력에 따라 강서구, 기장군은 총 16개, 강서구, 기장군 외의 구는 8개 기준으로 층화코드를 부여하고, 경제활동가능 가구주 연령과 주택의 자가소유 여부는 표본추출된 결과를 전체 조사구와 비교하여 모집단의 대표성을 확인할 수 있는 체크 변수로 활용한다.

<표 3-3> 분류지표의 표본조사구 추출 활용

분류지표		분류지표별 코드	표본조사구 추출활용
1차분류	주택유형	1: 단독주택 2: 아파트 26평 미만 3: 아파트 26평 이상 4: 기타주택(연립, 다세대, 오피스텔, 기타)	제1층화변수로 활용
2차분류	농림어가 비율	1: 농림어가 비율이 평균 미만 2: 평균이상 강서구(평균 36.2%), 기장군(평균 16.0%)	제2층화변수로 활용
3차분류	가구주 학력	1: 대졸자 비율이 평균 미만 2: 평균이상	제3층화변수로 활용

분류지표		분류지표별 코드	표본조사구 추출활용
4차분류	경제활동 가능 가구주 연령	1: 가구주 연령 15~64세 비율이 평균 미만 2: 평균 이상	제1체크변수로 활용
5차분류	주택의 자가소유 여부	1: 주택의 자가소유 비율이 평균 미만 2: 평균 이상	제2체크변수로 활용

마. 표본조사구 선정방법

인구주택총조사 결과 생산된 100% 조사구 20,065개와 2006~2007년 신축아파트 조사구 586개를 활용하여, 다음과 같은 순서로 16개 구·군별로 표본조사구를 추출한다.

- (1) 107개의 표본조사구를 층화코드 분류에 따라 층내 조사구 수에 비례하여 배분한다.
- (2) 층내에서는 읍·면·동 행정구역 순서로 부여되어 있는 조사구 번호의 순서에 따라 표본조사구를 계통추출한다.
- (3) 계통추출된 표본조사구가 제외조사구 해당여부를 확인하고, 해당되는 경우에는 동일 층화코드 내에서 조사구 번호 순으로 바로 뒤 번호의 조사구로 대체한다.
- (4) 위 3단계에서 계통추출된 표본조사구가 가장 마지막 번호의 조사구이고 제외 조사구에 해당되면, 바로 뒤 번호에 해당되는 조사구가 없게 되므로, 이 경우 동일 층화코드 내에서 조사구 번호 순으로 바로 앞 번호의 조사구로 대체한다.

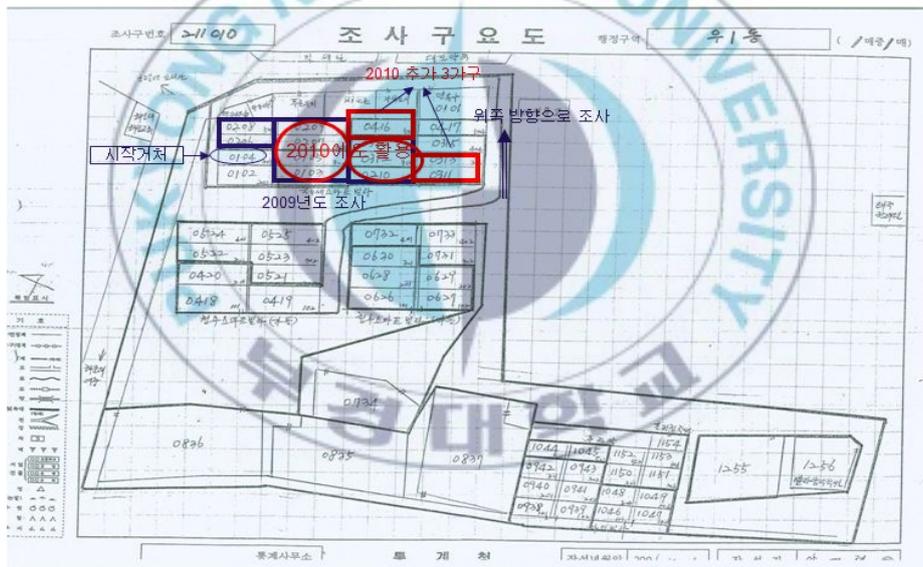
바. 조사구내 조사대상 가구 수

새로운 표본설계에서는 표본조사구내 10가구를 추출할 것을 제안하고 있다. 조사구 내 5가구를 조사할 경우에는 구·군별 표본가구 수 1,068가구를 기준으로 214개의 많은 표본조사구가 필요하며, 선정된 표본조사구는 기타 통계조사 및 신규조사에 활용될 예정으로, 일반적으로 조사불응 가구의 비율은 30% 정도 내외로 보고되고 있어, 조사구내 15가구 이상을 조사할 경우에는 표본조사구의 총 가구 수보다 많은 수의 가구가 조사대상 가구로 선정될 우려가 있어 조사대상 가구의 중복의 문제점이 발생할 수 있다.

사. 표본조사구 내 조사가구 선정

확정된 최종 표본조사구별로 랜덤으로 부여된 시작 거처부터 단독주택 지역이면 시작거처가 홀수번호면 오른쪽, 시작거처가 짝수번호면 왼쪽으로 10가구를 조사한다. 아파트 혹은 기타주택(연립, 다세대, 오피스텔 등) 지역의 경우에는 시작거처가 홀수번호면 시작거처부터 위쪽 방향, 시작거처가 짝수번호면 아래쪽 방향으로 10가구를 조사한다.

새로운 표본설계는 구·군별 총 107개의 조사구를 대상으로 1,068가구를 조사하기 때문에, 각 조사구당 10가구씩을 조사하고 1가구는 8가구를 조사하거나, 2가구는 9가구를 조사해야 한다.



<그림 3-1> 표본조사구 내 조사가구 선정 예시

아. 예비표본조사구

예비표본조사구란 표본조사구와 동일한 구·군내 및 층코드를 가지면서 조사구 번호가 가장 인접한 조사구를 말하며, 철거 혹은 철거예정과 같은 조사구 교체사유에 대비하여, 구·군별로 107개의 표본조사구에 1대1로 대응되는(one-to-one correspondence)

예비표본조사구를 동일한 층화코드 내에서, 선정된 표본조사구의 조사구 번호 순으로 바로 뒤 번호의 조사구를 추출한다.

2. 제안된 구·군별 새로운 표본설계 결과

이번 절에서는 1절에서 설명한 제안된 새로운 표본설계 방식에 따라 추출된 표본 조사구의 각 구·군별 층화코드에 따른 배분결과를 나타내고자 한다.

가. 층화코드 설명

주택유형이 단독주택이면 층화코드 1, 아파트 26평 미만이면 층화코드 2, 아파트 26평 이상은 층화코드 3, 기타주택은 층화코드 4로 주어지며, 농림어가 비율이 16% 미만이면 층화코드 1, 16% 이상이면 층화코드 2로 주어진다. 부산광역시에서 강서구와 기장군을 제외한 나머지 14개 구는 농림어가 비율이 16% 미만으로 층화코드 1밖에 주어지지 않았다. 가구주 학력 대졸비율이 평균 미만이면 층화코드 1, 평균 이상이면 층화코드 2로 주어져 3가지 분류지표(주택유형, 농림어가 비율, 가구주학력 대졸 비율)를 층화코드로 나타낸 것은 다음 <표 3-4>와 같다.

<표 3-4> 3가지 분류지표에 따른 층화코드

주택유형	농림어가 비율	가구주학력 대졸비율	층화코드
단독주택	16% 미만	평균 미만	111
단독주택	16% 미만	평균 이상	112
단독주택	16% 이상	평균 미만	121
단독주택	16% 이상	평균 이상	122
아파트 26평미만	16% 미만	평균 미만	211
아파트 26평미만	16% 미만	평균 이상	212
아파트 26평미만	16% 이상	평균 미만	221
아파트 26평미만	16% 이상	평균 이상	222
아파트 26평이상	16% 미만	평균 미만	311
아파트 26평이상	16% 미만	평균 이상	312
아파트 26평이상	16% 이상	평균 미만	321
아파트 26평이상	16% 이상	평균 이상	322
기타주택	16% 미만	평균 미만	411
기타주택	16% 미만	평균 이상	412
기타주택	16% 이상	평균 미만	421
기타주택	16% 이상	평균 이상	422

나. 각 구·군별 층화코드에 따른 배분결과

강서구와 기장군을 제외한 나머지 14개 구는 농림어가 비율 코드가 1밖에 없어 가능한 모든 층은 8개의 층이며, 배분 결과 중구(층화코드 '311'), 서구(층화코드 '311'), 동구(층화코드 '311'), 북구(층화코드 '412'), 해운대구(층화코드 '112')는 구 이름 옆에 기재한 층화코드에 따른 조사구 수가 적어 가장 인접한 층화코드와 합쳐져 7개의 층으로 배분되었다. 강서구의 경우 농림어가 비율이 높은 지역이면서, 2005년도에는 아파트 조사구가 없었으며, 거의 대부분 단독주택으로 구성되어 있었다. 다만 2006년도 이후 신축 아파트조사구 11개가 추가되었다. 농림어가 비율이 고려되기 때문에 총 16개의 층이 최대로 가능하나, 강서구의 특성으로 인하여 6개의 층이 존재한다. 기장군 또한 최대 16개의 층에서 12개의 층이 존재한다.

2005년 인구주택총조사 및 2006~2007년 신축아파트 조사구 추가 기준으로 부산시 총 조사구 수는 20,651개이며 부산진구가 2,408개로 가장 많고, 그 다음으로 해운대구 2,355개, 사하구 2,011개 순으로 조사구가 많으며, 강서구가 268개로 조사구가 가장 적었다.

<표 3-5> 층화코드에 따른 부산시 구·군 배분 현황

구	군	층	층화 코드	조사구 수	표본 조사구 수	구	군	층	층화 코드	조사구 수	표본 조사구 수
중	구	1	111	124	37	서	구	1	111	450	59
		2	112	52	16			2	112	147	19
		3	211	25	8			3	211	16	2
		4	212	30	9			4	212	52	7
		5	311+312	16	5			5	311+312	88	11
		6	411	54	16			6	411	21	3
		7	412	54	16			7	412	43	6
합	계	-	-	355	107	합	계	-	-	817	107
동	구	1	111	341	55	영	도	1	111	420	51
		2	112	172	28			2	112	81	10
		3	211	33	5			3	211	118	14
		4	212	41	7			4	212	151	18
		5	311+312	36	6			5	311	8	1

구	군	층	층화 코드	조사구 수	표본 조사구 수	구	군	층	층화 코드	조사구 수	표본 조사구 수
동	구	6	411	15	2	영	도	6	312	73	9
		7	412	27	4			7	411	19	2
								8	412	16	2
합	계	-	-	665	107	합	계	-	-	886	107
부산진구		1	111	1,116	49	동	래	1	111	688	48
		2	112	62	3			2	112	76	5
		3	211	158	7			3	211	130	9
		4	212	410	18			4	212	193	13
		5	311	21	1			5	311	16	1
		6	312	423	19			6	312	283	20
		7	411	137	6			7	411	99	7
		8	412	81	4			8	412	52	4
합	계	-	-	2,408	107	합	계	-	-	1,537	107
남	구	1	111	732	45	북	구	1	111	429	26
		2	112	83	5			2	112	14	1
		3	211	122	7			3	211	406	24
		4	212	198	12			4	212	382	23
		5	311	32	2			5	311	28	2
		6	312	332	21			6	312	336	20
		7	411	170	11			7	411+412	191	11
		8	412	59	4						
합	계	-	-	1,728	107	합	계	-	-	1,786	107
해운대구		1	111+112	611	28	사	하	1	111	705	37
		2	211	353	16			2	112	67	4
		3	212	437	20			3	211	269	14
		4	311	61	3			4	212	336	18
		5	312	602	27			5	311	29	1
		6	411	193	9			6	312	390	21
		7	412	98	4			7	411	163	9
합	계	-	-	2,355	107	합	계	-	-	2,011	107
금정구		1	111	659	45	강	서	1	111	72	29
		2	112	164	11			2	112	62	25
		3	211	68	5			3	121	80	32
		4	212	162	11			4	122	40	16
		5	311	12	1			5	312	11	4
		6	312	258	18			6	412	3	1
		7	411	120	8						
		8	412	115	8						
합	계	-	-	1,558	107	합	계	-	-	268	107

구	군	층	층화 코드	조사구 수	표본 조사구 수	구	군	층	층화 코드	조사구 수	표본 조사구 수
연 제 구		1	111	621	53	수 영 구		1	111	439	45
		2	112	18	1			2	112	33	3
		3	211	72	6			3	211	48	5
		4	212	198	17			4	212	131	13
		5	311	8	1			5	311	17	2
		6	312	238	20			6	312	206	21
		7	411	81	7			7	411	123	13
		8	412	27	2			8	412	53	5
합	계	-	-	1,263	107	합	계	-	-	1,050	107
사 상 구		1	111	567	40	기 장 군		1	111	58	14
		2	112	41	3			2	112	5	1
		3	211	231	16			3	121	134	32
		4	212	320	23			4	122	5	1
		5	311	15	1			5	211	29	7
		6	312	231	16			6	212	91	22
		7	411	90	6			7	221+222	2	1
		8	412	21	2			8	311	6	1
합	계	-	-	1,516	107	합	계	-	-	448	107
		부산시 합계		-	-			20,651	1,712		

3. 제안된 추정방법 및 표본관리 지침

본 절에서는 새로운 표본설계에 맞게 제안된 각 구·군별 추정방법에 대해서 정리한 후 이를 추계한 부산광역시 전체 통계치 추정방법에 대해서 설명하고자 한다. 또한 조사항목 무응답 발생시 제안된 처리 방법을 제시하며, 끝으로 표본조사구내 조사구역 선정, 표본조사구 교체, 표본설계 내 조사가구 선정 등과 같은 표본관리에서 나타나는 다양한 고려사항들에 대한 제안을 나타내하고자 한다.

가. 추정방법

부산광역시에서 16개 구·군별로 층화하였기 때문에 각 구·군별 통계치를 가중합(weighted sum)하여 부산광역시 전체 통계치를 산출할 수 있다. 새로운 표본설계에서는

통계청 방식대로 구·군내에서 층화코드에 따른 2단 집락추출법을 사용하였다. 2단 집락추출(two-stage cluster sampling)은 집락표집의 개념을 확장하여, 두 번의 확률표집이 사용되는 추출방법으로, 일단계에서는 집락에 대한 확률표본을 얻고 이단계에서는 일단계에서 표본으로 추출된 각 집락으로부터 조사단위들에 대한 확률표본을 추출하는 랜덤 표집법이다(Cochran, 1977). 집락추출에서는 조사구가 psu(일차추출단위: primary sampling unit)가 되고 가구는 ssu(이차추출단위: secondary sampling unit)가 된다. 관측값은 표본 가구 내 15세 이상 모든 가구원의 응답값이며, 가구 단위로 집계할 수 있다. 관심있는 모수(interesting parameter)는 특정 항목의 부산광역시 전체 모집단(가구 혹은 가구원)의 총계(total sum) 혹은 총 평균(total mean), 구·군별 합계(sum) 혹은 평균(mean)이 된다.

(1) 구·군별 통계치 추정방법

새로운 표본설계에 맞게 제안된 각 구·군별 가구 혹은 가구원 단위 통계치 산출 방법에 대해서 정리한다.

(가) 첨자의 정의

구·군별, 층별, 조사구별에 따른 첨자(index)들을 다음과 같이 정의한다. 예를 들어 $ghki$ 는 g 구·군에서 h 번째 층의 k 번째 표본조사구에서의 i 번째 가구를 말한다.

g : g 구·군 (g =중구, 서구, ..., 사상구, 기장군)

h : 층화코드에 따른 h 번째 층 ($h=1, 2, \dots$)

k : k 번째 표본조사구 ($k=1, 2, \dots$)

i : i 번째 가구 ($i = 1, 2, \dots$)

(나) 구·군별 설계 가중치(g 구·군에서 h 번째 층의 k 번째 표본조사구에서의 설계 가중치)

새로운 표본설계에 따른 설계 가중치는 다음과 같다. 설계 가중치는 구·군의 층내에서 전체 및 표본조사구 수와 표본조사구 내 전체 및 표본 가구 수에 의해 결정된다.

$$w_{ghk} = \frac{N_{gh}}{n_{gh}} \times \frac{M_{ghk}}{m_{ghk}}$$

N_{gh} : g 구·군에서 h 번째 층의 총 조사구 수

n_{gh} : g 구·군에서 h 번째 층의 표본조사구 수

M_{ghk} : g 구·군에서 h 번째 층의 k 번째 표본조사구에서 총 가구 수

m_{ghk} : g 구·군에서 h 번째 층의 k 번째 표본조사구에서 표본 가구 수

설계 가중치는 일차추출단위 추출확률과 이차추출단위 추출확률 곱의 역수로 산출된다.

$$w_{ghk} = \frac{1}{P(ppu)P(ssu)}, \quad (\text{단, } P(ppu) = \frac{n_{gh}}{N_{gh}}, P(ssu) = \frac{m_{ghk}}{M_{ghk}})$$

(다) 관측값 y_{ghki}

g 구·군에서 h 번째 층의 k 번째 표본조사구에서의 i 번째 가구 혹은 가구원의 응답값 합계가 관측값이 된다.

(라) 구·군별 합계 추정치

구·군별 설계가중치를 고려한 구·군별 합계 추정치 산출공식은 다음과 같다.

$$\hat{Y}_g = \sum_h^{H_g} \sum_k^{n_{gh}} \sum_i^{m_{ghk}} w_{ghk} y_{ghki}$$

\hat{Y}_g : g 번째 구·군의 합계 추정치

H_g : g 번째 구·군의 층 크기

(마) 구·군별 평균 추정치

구·군별 설계가중치를 고려한 구·군별 평균 추정치 산출공식은 다음과 같다.

$$\bar{\hat{Y}}_g = \frac{\hat{Y}_g}{K_g} = \frac{\sum_h^{H_g} \sum_k^{n_{gh}} \sum_i^{m_{ghk}} w_{ghk} y_{ghki}}{K_g}$$

\widehat{Y}_g : g 번째 구·군의 평균 추정치

K_g : g 번째 구·군의 총 가구(가구단위 조사결과 일 때) 혹은 가구원
(가구원 단위 조사결과 일때) 수

(바) 구·군별 합계 추정치의 표준오차

구·군별 표본설계 방식은 층화 이단 집락추출방법을 사용하였기 때문에, 이에 맞는 산출 공식을 활용한 분산추정치 식은 다음과 같다. 표준오차 추정치는 분산 추정치의 제곱근을 사용하면 된다.

1) 분산 추정치

$$\widehat{var}(\widehat{Y}_g) = \sum_{h=1}^{H_g} \left[N_{gh}^2 \frac{N_{gh} - n_{gh}}{N_{gh}} \frac{s_{gh}^2}{n_{gh}} + \frac{N_{gh}}{n_{gh}} \sum_{k=1}^{n_{gh}} M_{ghk}^2 \frac{M_{ghk} - m_{ghk}}{M_{ghk}} \frac{s_{ghk}^2}{m_{ghk}} \right]$$

$$s_{gh}^2 = \frac{1}{n_{gh} - 1} \sum_{k=1}^{n_{gh}} (\widehat{Y}_{ghk} - \widehat{Y}_{gh})^2$$

: 구·군별 h 번째 층에서의 합계 추정치의 표본분산

$$s_{ghk}^2 = \frac{1}{m_{ghk} - 1} \sum_{i=1}^{m_{ghk}} (y_{ghki} - \bar{y}_{ghk})^2$$

: 구·군별 h 번째 층의 k 번째 표본조사구에서의 합계 추정치의 표본분산

$$\widehat{Y}_{ghk} = \sum_{i=1}^{m_{ghk}} M_{ghk} \frac{y_{ghki}}{m_{ghk}}, \quad \widehat{Y}_{gh} = \frac{1}{n_{gh}} \sum_{k=1}^{n_{gh}} \widehat{Y}_{ghk}, \quad \bar{y}_{ghk} = \frac{1}{m_{ghk}} \sum_{i=1}^{m_{ghk}} y_{ghki}$$

2) 표준오차 추정치

$$\widehat{se}(\widehat{Y}_g) = \sqrt{\widehat{var}(\widehat{Y}_g)}$$

(사) 구·군별 평균 추정치의 표준오차

1) 분산 추정치

$$\widehat{var}(\widehat{Y}_g) \approx \frac{\widehat{var}(\widehat{Y}_g)}{\widehat{K}_g^2}$$

2) 표준오차 추정치

$$\widehat{se}(\widehat{Y}_g) = \sqrt{\widehat{var}(\widehat{Y}_g)}$$

(2) 부산광역시 전체 통계치 추정방법

부산광역시 전체 통계치 추정방법은 다음과 같다.

(가) 부산광역시 전체 총계 추정치

부산광역시 전체 총계 추정치는 구·군별 합계 추정치의 구·군별 합계를 통해서 다음과 같이 산출한다.

$$\widehat{Y} = \sum_{g=\text{중구}}^{\text{기장군}} \widehat{Y}_g$$

\widehat{Y}_g : g 번째 구·군별 합계 추정치

(나) 구·군별 가중치

새로운 표본설계는 각 구·군별 크기에 상관없이 동일하게 1,068가구를 대상으로 조사하였기 때문에, 구·군별 가중치(= W_g)를 계산해야 한다. 가중치는 다음 식과 같이 구·군별 총 가구 혹은 가구원 수에서 부산광역시 전체 가구 혹은 가구원 수를 나눈 값이다.

$$W_g = \frac{\widehat{K}_g}{\sum_{g=\text{중구}}^{\text{기장군}} \widehat{K}_g}$$

(다) 구·군별 총 가구 및 가구원 수

구·군의 총 가구, 가구원을 구하는 방법은 두 가지가 있는데 첫 번째는 2005년 인구주택총조사 및 2006~2007년 신축아파트 조사구 기준으로 산출하는 것이며, 두 번째는 실제 조사 응답결과를 활용한 방법이 있다.

1) 2005년 인구주택총조사 및 2006~2007년 신축아파트 조사구 기준 방법

새로운 표본추출틀인 2005년 인구주택총조사 100% 조사구와 2006~2007년 신축아파트 조사구를 추가하여 각 구·군별로 전체 가구 혹은 가구원을 산출할 수 있다. 모집단의 큰 변동이 없다면, 정밀한 통계치 산출이 가능하나 매년 모집단의 변동이 클 것이기 때문에 통계치를 산출하는데 있어 변동을 반영하지 않으면 추정치의 편이(bias)가 발생할 수 있다. 2005년 인구주택총조사 및 신축아파트 조사구 기준으로서의 부산광역시 전체 가구 수는 1,224,081가구로 부산시 16개 구·군 중 부산진구의 가구 수가 143,629가구로 가장 많고, 강서구의 가구 수가 15,986가구로 가장 적다.

<표 3-6> 구·군별 가구, 가구원 수 및 구·군별 비중치
(인구주택총조사 및 신축아파트조사구 기준)

구 · 군	전체(= K_g)		구·군별 비중치(= W_g)	
	가구 수	가구원 수	가구 기준	가구원 기준
중 구	21,179	51,955	1.7%	1.5%
서 구	47,920	129,540	3.9%	3.6%
동 구	38,890	105,883	3.2%	3.0%
영 도 구	53,710	151,926	4.4%	4.3%
부산진구	143,629	406,974	11.7%	11.5%
동 래 구	91,075	270,048	7.4%	7.6%
남 구	103,188	300,265	8.4%	8.4%
북 구	106,851	326,760	8.7%	9.2%
해운대구	136,838	404,845	11.2%	11.4%
사 하 구	120,540	359,135	9.8%	10.1%
금 정 구	92,288	266,379	7.5%	7.5%
강 서 구	15,986	45,509	1.3%	1.3%
연 제 구	74,842	220,145	6.1%	6.2%
수 영 구	61,502	173,848	5.0%	4.9%
사 상 구	91,047	269,609	7.4%	7.6%
기 장 군	24,596	70,928	2.0%	2.0%
합 계	1,224,081	3,553,749	100.0%	100.0%

2) 응답결과를 활용한 방법

조사결과에서 나타난 표본가구 내 가구원 수를 이용하여 다음과 같이 K_g 를 추정하는 방법이 있다. 이 방법은 최근 조사결과를 반영하기 때문에 편의가 작을 수 있으나, 일부 표본만을 대상으로 하여 추정치의 분산이 큰 문제점이 발생할 수 있다.

$$\text{구·군별 가구 수 추정 : } \widehat{K}_g = \sum_h \sum_k \sum_i w_{ghk}$$

$$\text{구·군별 가구원 수 추정 : } \widehat{K}_g = \sum_h \sum_k \sum_i w_{ghk} x_{ghki}$$

x_{ghki} : g 구·군에서 h 번째 층의 k 번째 표본조사구에서 표본 가구내에서의 조사결과 나타난 가구원 수

(라) 부산광역시 총평균 추정치

산출된 구·군별 가중치를 이용한 부산광역시 총평균 추정치는 다음과 같다.

$$\widehat{Y} = \sum_{g=\text{중구}}^{\text{기장군}} W_g \widehat{Y}_g$$

(마) 부산광역시 총계 추정치의 표준오차

분산 추정치는 구·군별 합계 추정치의 합으로부터 다음 식과 같이 구할 수 있다. 표준오차 추정치는 분산 추정치의 제곱근을 사용하면 된다.

1) 분산 추정치

$$\widehat{var}(\widehat{Y}) = \sum_{g=\text{중구}}^{\text{기장군}} \widehat{var}(\widehat{Y}_g)$$

2) 표준오차 추정치

$$\widehat{se}(\widehat{Y}) = \sqrt{\widehat{var}(\widehat{Y})}$$

(바) 부산광역시 총평균 추정치의 표준오차

1) 분산 추정치

$$\widehat{\text{var}}(\widehat{Y}) = \sum_{g=\text{중구}}^{\text{기장군}} W_g^2 \widehat{\text{var}}(\widehat{Y}_g)$$

2) 표준오차 추정치

$$\widehat{se}(\widehat{Y}) = \sqrt{\widehat{\text{var}}(\widehat{Y})}$$

나. 조사항목 무응답 발생시 처리 방법

조사통계의 이론적 측면은 무응답이 없다는 것을 전제로 하고 있다. 그러나 선진국들의 경우 수십 년 전 부터 가구 대상 조사에서의 무응답이 증가해왔다(De Leeuw과 De Heer, 2001). 실제 조사를 실시하는 과정에서 표본조사구내 가구들의 변동이 발생할 수 있고, 응답을 했다 하더라도 불성실한 응답을 한 경우, 어느 문항에 대해 이해 부족으로 응답해야할 문항에 응답을 하지 못한 경우 등의 문제가 발생한다.

조사항목에서 결측값이 생기는 경우 표본교체(substitution), 가중치조정(weighting adjustment) 또는 대체방법(imputation)이 널리 사용된다(김영원과 조선경, 1996).

새로운 표본설계에서는 통계분석 담당자가 이해하고 사용하기 편리한 가중치를 조정하는 방법을 사용할 것을 제안되었으며, 가구단위에서 조사항목 무응답이 발생하면 표본가구 수 m_{ghk} 를 조사항목 무응답을 제외한 가구 수로 변경하면 가중치를 조정하는 효과가 나타난다. 또한, 가구원 단위에서 조사항목 무응답이 발생하면 표본가구 내 가구원 수 x_{ghki} 를 조사항목 무응답을 제외한 가구원 수로 변경한다.

(1) 표본교체

조사항목에서 결측이 발생한 표본을 다른 표본으로 교체하는 방법으로 교체된 표본을 다시 조사하는 방법이다.

(2) 가중치조정

결측단위(가구 혹은 가구원의 조사항목 무응답)를 다루는 가장 일반적인 방법으로 결측 단위에 대해 별도로 다른 값을 대체해주지 않은 대신에 응답한 단위들의 가중치를 조정해주어 결측으로 인한 영향을 줄여주는 방법이다.

(3) 대체방법

다른 관측된 조사항목의 정보를 활용하여 결측된 조사항목을 특정 값으로 대체하는 방법이다.

다. 표본관리 지침

표본조사구내 조사구역 선정, 표본조사구 교체, 표본설계내 조사가구 선정 등과 같은 표본관리에서 나타나는 다양한 고려사항들에 대한 제안된 지침을 나타내고자 한다.

(1) 표본조사구 내 조사구역 선정

통계청 방식인 ① 조사구 경계확인 및 요도(要圖)보완, ② 거처번호 부여 및 가구 명부 작성, ③ 조사 전 가구 확인 작업 순으로 진행할 것을 제안되었다.

(2) 표본조사구 교체

조사구역이 선정된 표본조사구에서 교체사유가 발생한 경우 표본조사구와 1대 1로 대응되는 예비표본조사구로 대체하여야 한다.

통계청에서 정의하고 있는 표본조사구 교체사유는 다음과 같다.

- ① 조사구 내 가구가 당해 년에 전체 철거 또는 철거 예정 조사구의 경우
- ② 행정구역 변동에 의한 해당 조사구 관할이 서로 다른 경우
- ③ 지형지물의 심한 변동으로 조사구 식별이 불가능한 경우
- ④ 조사구내 위험이나 조상장애가 생겨 조사 수행이 불가능한 경우
- ⑤ 기타 기관 등에서 현 조사구에서 조사를 실시하고 있는 경우

(3) 조사항목 무응답이 많은 가구의 처리지침

부산광역시 통계담당관이 판단하기에 중요한 조사항목의 무응답이 많은 가구의 경우에는 조사불능 가구 처리지침에 따라 동일 표본조사구내 인접 가구로 대체한다.

표본가구 대체를 허용하는 적절한 기준을 설정하여 이런 기준에 부합되는 경우에 한하여 표본가구를 대체하는 것이 바람직하다.

(4) 사회조사 조사지침서 변경

새로운 표본설계 및 표본관리 지침에 맞게 「사회조사 조사지침서」에서 일부 수정이 필요한 내용에 대한 2008년 이전 방식 및 개선사항을 다음 <표 3-7>에 나타냈다. <표 3-7>는 통계청에서 사용하였던 방식을 적용, 제안된 내용으로 향후 「부산 사회조사」의 현실적인 상황 및 특성에 맞게 다시 재조정할 필요가 있다.

<표 3-7> 조사지침서 개선사항

항목	2008년 이전 방식	개선
표본설계	<ul style="list-style-type: none"> 표본추출방법으로 층화3단 추출법을 택하고, 각 구를 3개층으로 층화한 다음 1차 추출단위로서 통을 반수에 비례하는 확률로 확률비례 추출하고 추출된 통에서는 1개의 반(2차 추출단위)을 등확률로 추출 표본 반에서는 총 추출확률이 동일하게 되게끔 4가구를 계통추출하여 자체 가중 표본 선정 16개 구·군별 352가구씩 총 5,632가구를 확률적으로 추출하여 표본가구내의 모든 조사대상자(만15세이상 가구원)를 조사 	<ul style="list-style-type: none"> 표본추출방법으로 구·군별 조사구를 3개의 층화변수(주택 유형, 농림어가 비율, 가구주 학력)를 이용하여, 층별 조사구수에 비례하여 표본을 할당 층내에서 읍·면·동 행정구역 순서로 부여되어 있는 조사구 번호의 순서에 따라 표본조사구를 계통추출, 제외조사구일 경우 동일 층 내에서 조사구 번호가 인접한 조사구로 대체 표본조사구 내 부여된 시작거처로부터 인접한 10가구를 조사대상 가구로 선정 16개 구·군별 1,068가구씩 총 17,088가구가 조사대상 가구이며, 표본가구내 모든 조사대상자(만15세이상 가구원)를 조사

항목	2008년 이전 방식	개선
응답거부 가구 조사	<ul style="list-style-type: none"> • 응답을 계속 거부할 경우에는 지도공무원(구·군 직원)에게 보고하여 조치를 받는다. 	<ul style="list-style-type: none"> • 응답을 2회 불응시에는 지도공무원(구·군 직원)에게 보고하여, 3회째 방문시 슈퍼바이저(구·군 및 시청 통계담당관)를 동행한다. • 3회 방문한 후 조사를 계속 불응시, 동일 표본조사구내 인접 가구로 대체 조사한다.
조사대상 가구	<ul style="list-style-type: none"> • 혈연가구 : 혈연가족 관계로 구성된 가구 • 혈연가족과 혈연관계가 없는 5인 이하의 남남끼리 한 집에서 생활을 같이하는 가구 	<ul style="list-style-type: none"> • 가족으로 이루어진 가구 • 1인가구 • 가족이 아닌 남남끼리 함께 사는 5인 이하의 가구
조사대상 제외 가구	<ul style="list-style-type: none"> • 1인가구 • 비혈연 6인~19인 이하 가구 • 집단시설가구 • 외국인 가구 	<ul style="list-style-type: none"> • 가족이 아닌 남남끼리 함께 사는 6인 이상의 가구 • 보육원 등 집단가구 • 양로원 등 집단가구 • 장애인 복지시설에 살고 있는 집단가구 • 기타 복지시설에 살고 있는 집단가구 • 외국인 가구

IV. 연구결과 및 해석

본 장에서는 III장에서 제시한 새로운 표본설계에서 추출된 구·군별 1,068가구, 107개 표본조사구, 부산광역시 전체 16,088가구, 1,712개 표본조사구를 대상으로 2009년 9월 10일 ~ 9월 28일까지 실시된 『2009 부산사회조사』 주요 결과를 이용하여, 새로운 표본설계에서 제시한 사항들 중에서 실제 조사 수행 과정 속에서 변경된 주요 사항 및 구·군별 통계치 및 부산광역시 전체 통계치 추정 등을 나타내고자 한다.

1. 새로운 표본설계 제안에 대한 평가

새로운 표본설계에서 제안한 표본크기, 표본추출틀, 조사모집단, 분류지표, 조사구내 조사대상 가구 수 등에 대해 실제 조사 결과를 토대로 적절한 제안인가를 확인하고자 한다.

가. 표본크기

2008년도 조사에선 부산광역시 전체 조사원 128명이 조사기간 11일간 5,632가구를 조사 시 1인 기준 일일 조사대상 가구 수가 4가구였다. 표본크기가 2008년도에 비해 구·군별 352가구에서 1,068가구로 약 3배 이상의 표본이 늘어났지만, 충분한 예산의 확보로 조사원의 일일 업무량 부담감을 줄이기 위해 조사원을 236명으로 늘리고 조사기간을 19일로 현행보다 늘려 17,088가구를 대상으로 1인 기준 일일 조사대상 가구 수가 3.8가구로 예년조사보다 조사상의 큰 어려움은 없었다.

제안한 표본크기인 구·군별 1,068가구는 비록 구·군별 가구규모의 차이를 고려하지 않고 무한모집단을 가정하여 계산한 결과이지만, 구·군별 실질 오차한계는 $\pm 2.9\% \sim \pm 3.0\%$ 포인트이다. 새로운 표본설계에서 책정한 구·군별 $\pm 3.0\%$ 포인트 오차한계는 통계청이 생산하는 통계의 일반적인 신뢰도로 각 구·군별로 일반적인 사회통계조사에서 신뢰성을 확보할 수 있는 최소 수준의 표본크기이므로 예산상의 어려움이 없다면 각 구·군별 1,068가구의 표본크기를 유지하는 것이 바람직하다고 생각된다. 2009년도 조사 결과 구·군 각 107개 조사구, 1,068가구를 대상으로 구·군별 응답가구

에서 만 15세 이상 가구원 수는 2,794~2,991명으로 부산광역시 전체 응답가구의 만 15세 이상 가구원 수는 총 45,634명이다. 구·군별 가구원 수는 동래구가 2,991명으로 가장 많았고, 사상구 2,943명, 금정구 2,904명의 순으로 나타났다. 반면 부산진구는 2,795명, 중구는 2,794명으로 가장 적었다. 이는 구·군별로 가구당 만 15세 이상 가구원의 차이를 보인다. 불응 가구원수는 부산광역시 전체 383명으로 불응률은 0.8%이며, 서구가 50명, 1.7%로 불응률이 가장 높았고, 동래구, 기장군이 각각 7명, 0.2%로 불응률이 가장 낮았다.

2005년 인구주택총조사 및 2006~2007년 신축아파트 조사구 기준 가구수원 수 추정에서 자료수집 한계로 15세 미만 가구원 수가 포함된 전체 가구원 수를 대상으로 구·군별 가중치가 추정되었는데, 실제 조사 결과 구·군별 15세 이상 조사대상 구성비(%)의 차가 크지 않아, 추정상 문제가 발생하진 않았지만, 향후 표본설계에서는 수정되어야 할 부분이다. 조사결과 2009년도 구·군별 응답가구원 수, 불응률은 다음 <표 4-1>과 같다.

<표 4-1> 2009년도 구·군별 응답가구의 가구원 수, 불응률 (단위 : 명, %)

구 · 군	총 가구원수	15세 미만 가구원수	15세 이상 조사대상 가구원수	조사대상 구성비(%)	불응 가구원수	응답 가구원수	불응률
중 구	3,210	398	2,812	87.6	18	2,794	0.6%
서 구	3,257	394	2,863	87.9	50	2,813	1.7%
동 구	3,289	408	2,881	87.6	24	2,857	0.8%
영 도 구	3,306	408	2,898	87.7	43	2,855	1.5%
부산진구	3,260	448	2,812	86.3	17	2,795	0.6%
동 래 구	3,486	488	2,998	86.0	7	2,991	0.2%
남 구	3,416	557	2,859	83.7	31	2,828	1.1%
북 구	3,475	642	2,833	81.5	11	2,822	0.4%
해운대구	3,437	607	2,830	82.3	29	2,801	1.0%
사 하 구	3,407	564	2,843	83.4	26	2,817	0.9%
금 정 구	3,398	463	2,935	86.4	31	2,904	1.1%
강 서 구	3,178	268	2,910	91.6	27	2,883	0.9%
연 체 구	3,328	468	2,860	85.9	20	2,840	0.7%
수 영 구	3,377	510	2,867	84.9	31	2,836	1.1%
사 상 구	3,544	590	2,954	83.4	11	2,943	0.4%
기 장 군	3,413	551	2,862	83.9	7	2,855	0.2%
합계	53,781	7,764	46,017	85.6	383	45,634	0.8%

나. 표본추출틀

새로운 표본추출틀로 통계청에서 사용하고 있는 2005년 기준 인구주택총조사의 부산광역시 100% 조사구와 2006~2007년 신축아파트조사구의 활용이 제안되었는데, 현행 표본 「반(反)」 조사시 일일이 방문 확인을 통해서 조사구역 확인 후 요도(要圖)를 그려 조사 때 사용하는 번거로움이 있었는데, 조사구의 활용은 현행 보유하고 있는 요도(要圖)를 사용하면서 변경사항만 수정하면 되기 때문에 업무적 부담감이 현저히 줄었다.

2005년 인구주택총조사 이후부터 최근까지 조사구 변동이 발생할 수 있어 그 변동부분을 반영하여 2006~2007년도 신축아파트조사구를 추가하여 표본추출에 반영하였는데, 새로운 표본추출틀의 문제점으로는 2005년 이후 재개발로 인한 조사구 삭제와 같이 현재까지 없어진 조사구는 반영이 되지 않았으면서 586개의 조사구가 추가로 표본추출틀에 포함됨으로 Ⅲ.3절에 제시한 응답결과를 활용한 구·군별 총 가구원 수 추정에서 부산광역시 15세 이상 가구원 수가 3,307,203명이라는 실제 예측과 많은 차이가 있는 결과가 나타났다.(〈표 4-16〉 참조).

만약 매년 변동되는 조사구 정보를 활용할 수 있다면, 매년 변동을 감안한 표본 설계를 수정하는 것이 바람직하다. 그러나, 현실적으로 어려운 부분이 있기 때문에, 통계청과 같은 방식으로 매년 조사구의 변동사항이 있지만 조사구를 추가하지 않고 『인구주택총조사』의 부산광역시 100% 조사구만을 표본추출틀로 사용하는 것이 가구원 수의 과대추정 문제를 해결할 수 있을 것이다. 그러나, 신축조사구를 조사모집단에서 제외하는 것은 부산광역시 가구원의 정확한 의사반응을 반영하지 못할 수 있으므로 표본추출틀로는 통계청 100% 조사구에 신축아파트조사구를 포함하고 가중치는 통계청의 100% 조사구만을 이용할 수 있는 방안도 검토할 수 있다.

다. 조사모집단

조사모집단은 조사시점에 부산에 거주하는 표본가구 내 만 15세 이상 가구원으로 현행 방식과 달라진 점은 『통계청의 사회통계조사』 방식을 준수한다는 제안방식으로 1인 가구가 포함되었다. 그러나 구·군 통계담당관으로부터 1인 가구의 조사상 어려

움에 대한 의견을 반영하여 현행과 마찬가지로 『2009 부산 사회조사』 조사모집단에 몇몇 1인 가구는 대체되어 조사가 실시되었다.

라. 분류지표

새로운 표본설계 분류지표로 3개의 층화변수(주택유형, 농림어가 비율, 가구주 학력)와 2개의 체크변수(경제활동가능 가구주 연령, 주택의 자가소유 여부)를 활용하였는데, 적절한 분류지표를 설정하였는가를 가구주 응답 주요 항목(가구의 월 평균 총 소득)에 대해 비교를 통해 알아보려고 한다.

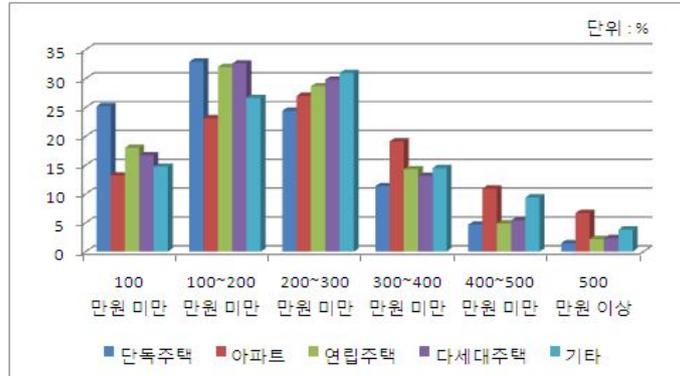
또한 주택유형, 농림어가 비율을 제외한 나머지 3개 분류지표에서 2012년도부터 부산광역시에서 2010년 인구주택총조사 조사구를 가지고 설계될 표본설계에서 제 3 층화변수로 적절한 분류지표가 무엇인지 알아보려고 한다.

1차 분류지표인 주택유형은 「2009 부산사회조사」 조사표에서 단독주택, 아파트, 연립주택, 다세대주택, 기타 다섯 개의 항목으로 구분되어 있다. 아파트의 경우 월 평균 총 소득이 500만원 이상인 비율이 6.68%로 다른 주거유형에 비해 1.75배~4.63 배나 높았고, 300~400만원 미만, 400~500만원 미만의 경우에도 아파트의 경우 월 평균 총 소득 비율이 높음을 알 수 있다. 단독주택의 경우 100만원 미만과 100~200만원 미만 비율의 합이 58.13%로 단독주택에 거주하는 절반 이상의 가구가 월 평균 200만원 미만의 소득임을 알 수 있다. 따라서, 주택유형에 따라 가구 월 평균 소득에 대한 차이를 보여 1차 분류지표로 사용함이 적절하다고 할 수 있다.

<표 4-2> 주택유형에 따른 주요항목(월 평균 소득)에 대한 비율

(단위 : %)

구 분	주요항목(월 평균 총 소득)					
	100만원 미만	100-200 만원 미만	200-300 만원 미만	300-400 만원 미만	400-500 만원 미만	500만원 이상
주택 유형						
단독주택	25.21	32.92	24.41	11.34	4.67	1.44
아파트	13.19	23.10	27.01	19.09	10.92	6.68
연립주택	18.01	31.99	28.67	14.29	4.87	2.17
다세대주택	16.71	32.61	29.80	13.14	5.41	2.33
기타	14.72	26.65	30.96	14.47	9.39	3.81



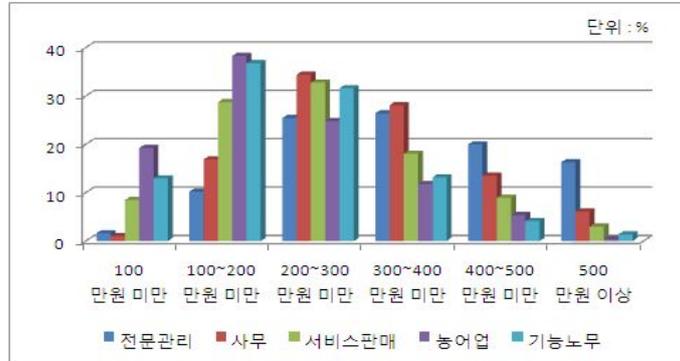
<그림 4-1> 주택유형에 따른 주요항목(월 평균 소득)에 대한 평균

2차 분류지표인 농림어가비율은 「2009 부산사회조사」 조사표에서 가구주의 직업을 통해 유추할 수 있다. 농림어업에 종사하는 가구주의 경우 월 평균 총 소득으로 100~200만원 미만이 38.30%로 가장 높았고, 200~300만원 미만이 24.84%로 두 번째로 높았다. 가구주가 농림어업 종사자의 경우 다른 직업에 비해 월 평균 총 소득 500만원 이상의 비율이 현저히 낮음을 알 수 있다.

제 2층화변수로 농림어가 비율을 사용함은 강서구, 기장군의 경우 조사구 특성을 반영하는데 중요하며, 실제 조사 결과에서 알 수 있듯이 가구 월 평균 소득에 대한 차이를 보여 2차 분류지표로 사용함이 적절하다고 할 수 있다.

<표 4-3> 가구주 직업별(농림어가 비율)에 따른 주요항목(월 평균 소득)에 대한 평균
(단위 : %)

구 분	주요항목(월 평균 총 소득)					
	100만원 미만	100-200만원 미만	200-300만원 미만	300-400만원 미만	400-500만원 미만	500만원 이상
직업별 전문관리	1.61	10.23	25.44	26.42	20.01	16.29
(농림어 사무	1.02	16.89	34.40	28.09	13.50	6.11
가비율) 서비스판매	8.51	28.73	32.76	18.06	8.96	2.99
농림어업	19.25	38.30	24.84	11.80	5.38	0.41
기능노무	12.95	36.80	31.59	13.15	4.13	1.38

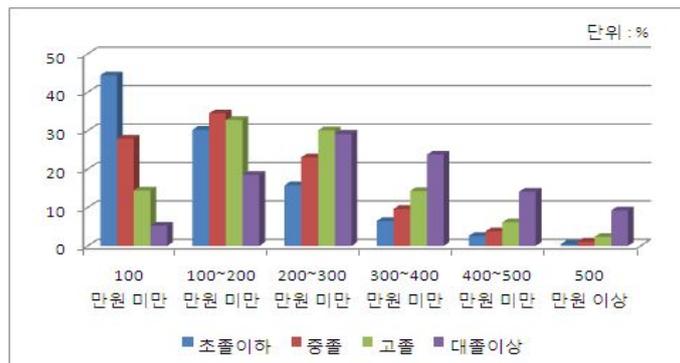


<그림 4-2> 직업에 따른 주요항목(월 평균 소득)에 대한 평균

3차 분류지표인 가구주 학력은 「2009 부산사회조사」 조사표에서 초졸이하, 중졸, 고졸, 대졸이상 4항목으로 구성되어 있다. 가구주 학력이 높아질수록 가구의 월 평균 총 소득 또한 높음을 알 수 있다. 따라서, 가구주 학력에 따라 가구 월 평균 소득에 대한 차이를 보여 3차 분류지표로 사용함이 적절하다고 할 수 있다.

<표 4-4> 가구주 학력에 따른 주요항목(월 평균 소득)에 대한 평균 (단위 : %)

구 분	주요항목(월 평균 총 소득)					
	100만원 미만	100-200만원 미만	200-300만원 미만	300-400만원 미만	400-500만원 미만	500만원 이상
가구주 학력 초졸이하	44.41	30.23	15.78	6.48	2.59	0.50
중졸	27.94	34.52	23.07	9.60	3.80	1.07
고졸	14.42	32.75	30.08	14.29	6.12	2.34
대졸이상	5.24	18.52	29.17	23.76	14.10	9.21

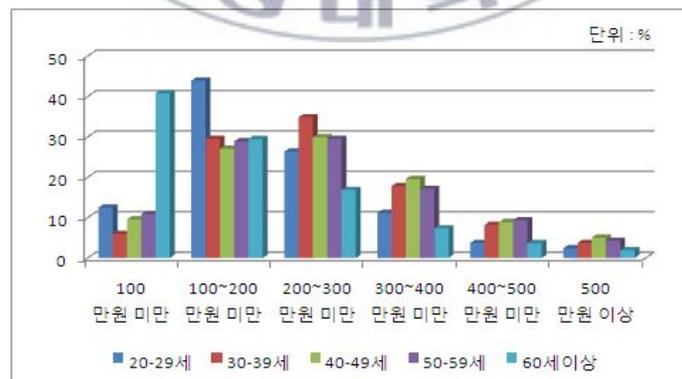


<그림 4-3> 가구주 학력에 따른 주요항목(월 평균 소득)에 대한 평균

4차 분류지표인 가구주 연령은 「2009 부산사회조사」 조사표에서 20-29세, 30-39세, 40-49세, 50-59세, 60세 이상 5항목으로 구성되어 있다. 가구주 연령이 20-29세인 경우, 월 평균 총 소득이 100~200만원의 비율이 43.92%로 가장 높았고, 200~300만원의 비율이 26.35%로 두 번째로 높았다. 이와 반대로 30-39세, 40세-49세, 50세-59세인 경우는 200~300만원의 비율이 가장 높았고, 100~200만원의 비율이 두 번째로 높았다. 가구주 연령이 60세이상의 경우 월 평균 총 소득이 100만원 미만의 비율이 40.81%로 가장 높아 다른 연령대에 비해 월 평균 총 소득의 차이가 나타남을 알 수 있다. 따라서 가구주 연령을 4차 분류지표로 사용함이 적절하다고 할 수 있다.

<표 4-5> 가구주 연령에 따른 주요항목(월 평균 소득)에 대한 평균
(단위 : %)

구 분	주요항목(월 평균 총 소득)					
	100만원	100-200	200-300	300-400	400-500	500만원
	미만	만원 미만	만원 미만	만원 미만	만원 미만	이상
연령 20-29세	12.50	43.92	26.35	11.15	3.72	2.36
30-39세	5.94	29.51	34.85	17.82	8.20	3.68
40-49세	9.56	27.05	29.90	19.55	8.90	5.03
50-59세	10.88	28.91	29.49	17.09	9.37	4.25
60세이상	40.81	29.46	16.84	7.31	3.66	1.92



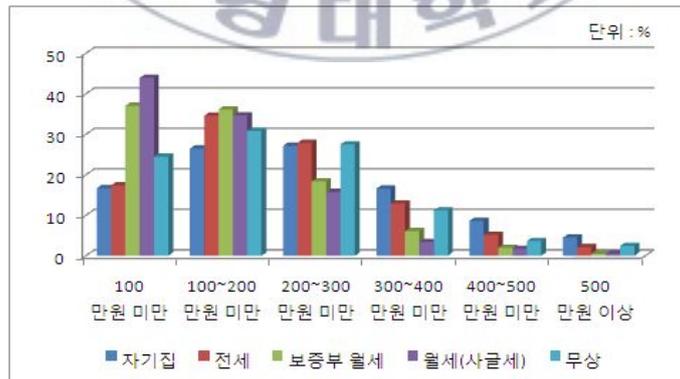
<그림 4-4> 가구주 연령에 따른 주요항목(월 평균 소득)에 대한 평균

5차 분류지표인 주택의 자가소유 여부는 「2009 부산사회조사」 조사표에서 자기집, 전세, 보증부 월세, 월세(사글세), 무상 5항목으로 구성되어 있다.

주택자가소유 여부가 보증부 월세, 월세(사글세)인 경우 월 평균 총 소득이 100만원 미만, 100~200만원 미만 비율의 합이 70% 이상을 차지하였으며, 400만원 이상은 2%대의 낮은 비율을 보였다. 그에 반해 주택 자가소유 여부가 자기집, 전세, 무상인 경우 월 평균 소득이 100만원대와 200만원대의 비율이 높았고, 400만원 이상의 비율은 각각 13%, 7%, 6%대의 비율을 나타냈다. 따라서, 주택의 자가소유 여부에 따라 월 평균 소득의 차이를 보여 5차 분류지표로 사용함이 적절하다고 할 수 있다.

<표 4-6> 주택의 자가소유 여부 형태에 따른 주요항목(월 평균 소득)에 대한 평균 (단위 : %)

구 분	주요항목(월 평균 총 소득)					
	100만원	100-200	200-300	300-400	400-500	500만원
	미만	만원 미만	만원 미만	만원 미만	만원 미만	이상
주택 자기집	16.67	26.48	27.14	16.57	8.64	4.49
자가소유 전세	17.37	34.60	27.86	12.91	5.13	2.13
여부 보증부 월세	37.04	36.11	18.37	6.07	1.95	0.47
월세(사글세)	43.94	34.65	15.77	3.38	1.69	0.56
무상	24.47	30.82	27.49	11.18	3.63	2.42



<그림 4-5> 주택 자가소유 여부 형태에 따른 주요항목(월 평균 소득)에 대한 평균

주택유형, 농림어가 비율은 조사상 분류지표로 반드시 필요하다. 이 2개 분류지표를 제외한 나머지 3개 분류지표에서 χ^2/df 값이 가구주 학력이 223.71로 경제활동 가능 가구주 연령 108.31, 주택 자가소유 여부 51.12보다 큰 값이므로 향후 사회조사에서도 기존과 같이 제 3층화변수로 가구주 학력을 사용하는 것이 적당하다.

<표 4-7> 5개의 분류지표 χ^2/df

변수	χ^2/df
주택유형	52.92
농림어가 비율	154.01
가구주 학력	223.71
경제활동 가능 가구주 연령	108.31
주택 자가소유 여부	51.12

마. 표본조사구 선정방법

인구주택총조사 결과 생산된 조사구 100%와 신축아파트조사구를 활용하여, 107개의 표본조사구를 층화코드 분류에 따라 층내 조사구 수에 비례하여 배분한 후 층내에서는 읍·면·동 행정구역 순서로 부여되어 있는 조사구 번호의 순서에 따라 표본조사구를 계통추출하며, 계통추출된 표본조사구가 제외조사구에 해당되면 동일 층화코드 내에서 조사구 번호 순으로 바로 뒤 번호의 조사구로 대체하는 표본조사구 선정방법은 읍·면·동 행정단위별로 골고루 표본조사구가 선정되었다고 생각된다.

바. 조사구내 조사대상 가구 수

조사구 내 조사 가구수를 5의 배수로 하는 것이 좋다고 제안되어 있는데 조사원 1인이 하루에 조사할 수 있는 가구는 대략 4~6가구 정도이므로, 책정된 예산과 조사기간, 조사원 업무량을 기준으로 조사구내 조사대상 가구 수로 10가구가 적당하다고 생각된다.

특히, 표본조사구는 『관광실태조사』, 『안전의식조사』 등 부산광역시에서 필요로 하는 다양한 통계조사에 적용될 수 있기 때문에 10가구 이상을 하게되면 중복조사의 문제점이 발생할 수 있다.

사. 예비표본조사구

새로운 표본설계 후 추출된 구·군별 107개 표본조사구에서 실제 조사 실시 때 대체사유로 변경된 조사구는 중구 6개, 서구 10개, 동구 3개, 영도구 1개, 부산진구 6개, 동래구 7개, 남구 2개, 북구 4개, 해운대구 4개, 사하구 2개, 금정구 6개, 강서구 11개, 연제구 6개, 수영구 3개, 사상구 1개, 기장군 8개로 부산시 전체 1,712개 조사구 중에서 80개의 조사구가 변경되었다.

대체사유로는 사하구 다대1동(층화코드 '211', '212')의 경우 재개발로 인한 아파트 철거로 해당 표본조사구에 가구가 살고 있지 않았으며, 강서구 천가동(층화코드 '121')의 경우 지역 위치상 배를 타고 조사를 나가야 하는 조사의 어려움으로 인해 동일한 층화코드를 가진 강서구의 다른 동으로 조사구를 변경해 주는 등 합당한 대체사유가 있을 시 예비표본조사구 목록에서 조사구를 대체하였다.

<표 4-8> 구·군별 대체조사구 수, 대체율

구 군	조사구 수	표본 조사구 수	대체 조사구 수	대체율
중 구	355	107	6	5.61%
서 구	817	107	10	9.35%
동 구	665	107	3	2.80%
영 도 구	886	107	1	0.93%
부산진구	2,408	107	6	5.61%
동 래 구	1,537	107	7	6.54%
남 구	1,728	107	2	1.87%
북 구	1,786	107	4	3.74%
해운대구	2,355	107	4	3.74%
사 하 구	2,011	107	2	1.87%
금 정 구	1,558	107	6	5.61%
강 서 구	268	107	11	10.28%
연 제 구	1,263	107	6	5.61%
수 영 구	1,050	107	3	2.80%
사 상 구	1,516	107	1	0.93%
기 장 군	448	107	8	7.48%
부산시 전체	20,651	1,712	80	4.67%

3개의 층화변수별 대체조사구 수와 대체율은 다음 <표 4-9>과 같다.

주택유형의 경우 기타주택의 표본조사구 수 191개에서 대체된 조사구는 25개로 13.09%의 대체율을 보였고, 두 번째로 단독주택이 대체율 4.68%였다. 기타주택과 단독주택의 경우 재개발 등으로 인해 조사구 대체가 발생할 경우가 많으므로 예비 표본조사구 또한 많이 필요할 것으로 생각된다.

농림어가 비율의 경우 농림어가 비율이 평균 이상인 표본조사구 수 85개에서 대체된 조사구는 11개로 12.94%의 대체율을 보였는데, 이는 강서구 천가동의 7개 표본조사구를 모두 다른 동으로 대체하였으므로 대체율이 높게 나타났다.

가구주 학력의 경우 대졸자 비율 평균미만의 대체율 5.56%, 평균이상의 대체율 3.41%였다.

<표 4-9> 3개 층화지표별 대체조사구 수, 대체율

층화지표	구분	표본조사구 수	대체조사구 수	대체율
주택유형	단독주택	877	41	4.68%
	아파트 26평 미만	377	12	3.18%
	아파트 26평 이상	267	2	0.75%
	기타주택	191	25	13.09%
농림어가 비율	16% 미만	1,627	69	4.24%
	16% 이상	85	11	12.94%
가구주 학력	평균 미만	1,008	56	5.56%
	평균 이상	704	24	3.41%

대체사유 발생시 『부산의 사회통계조사 표본설계 최종보고서』 부록3에 제시되어 있는 예비표본조사구 목록을 보고 동일한 층화코드 내에서 처음 표본 조사구와 인접한 조사구를 대체 한 후, 변경 전 조사구번호, 가구 수 및 가구원 수와 변경 후 조사구 번호, 가구 수 및 가구원 수는 다음 <표 4-10>과 같다.

새로운 표본설계 제안 상 동일한 층화코드 내에서 선정된 표본조사구의 조사구 번호 순으로 바로 뒤 번호의 조사구를 대체조사구로 추출해야하는데 재개발로 인한 철거같은 경우 인접조사구 또한 철거된 경우나 그 외 적절한 사유에 대해선 다른 동으로 대체된 경우도 있다. 각 조사구 당 살고 있는 가구 수와 가구원 수가 다르기 때문에 어느 구는 대체 전 총 가구 수보다 적어진 구도 있고, 또한 대체 전 총 가구원보다 많아진 구도 있다.

<표 4-10> 표본조사구에서 대체된 조사구번호, 가구 수, 가구원 수

구	군	층화 코드	읍·면·동	조사구번호	가구 수	가구원 수	읍·면·동	변경된 조사구번호	가구 수	가구원 수
중	구	111	부평동	01056038	48	107	부평동	01056036	78	193
		211	부평동	01056040	44	94	부평동	01056041	60	129
		411	중앙동	01051015	43	89	중앙동	01051004	51	122
		411	부평동	01056039	58	108	보수동	01054083	52	154
		411	남포동	01058004	48	89	대청동	01053017	57	173
		412	중앙동	01051017	47	57	대청동	01053004	63	177
서	구	111	서대신3동	02056033	51	144	서대신4동	02057003	48	143
		111	남부민2동	02066019	62	173	남부민2동	02066029	71	214
		112	서대신3동	02056035	52	145	서대신4동	02057017	64	154
		312	충무동	02064009	41	128	충무동	02064012	76	211
		411	충무동	02064056	124	166	서대신4동	02057002	66	211
		412	동대신3동	02053002	48	62	동대신3동	02053003	61	116
		412	동대신3동	02053017	58	96	동대신3동	02053019	72	164
		412	서대신3동	02056011	41	73	서대신3동	02056040	53	147
		412	암남동	02068072	68	92	초장동	02063005	59	193
		412	암남동	02068136	44	49	암남동	02068121	52	90
동	구	111	수정2동	03057025	43	111	수정3동	03058003	45	115
		112	수정3동	03058006	50	133	수정2동	03057003	78	216
		412	초량1동	03051024	51	143	초량1동	03051029	77	154
영도	구	111	동삼3동	04067019	57	130	동삼3동	04067011	66	192
부산진	구	111	부전2동	05052036	40	68	부전2동	05052045	67	132
		111	범전동	05053028	62	133	범전동	05053029	62	175
		111	범천1동	05077035	46	104	범천2동	05078001	45	126
		111	범천4동	05079027	52	130	범천4동	05079052	62	136
		412	부전2동	05052059	47	64	연지동	05054122	49	82
		412	범천4동	05079060	53	87	가야1동	05071055	63	161
동래	구	111	명륜1동	06053030	61	175	명륜1동	06053014	73	179
		111	명륜2동	06054027	53	165	명륜2동	06054048	52	151
		111	명륜2동	06054061	53	143	온천3동	06057042	74	233
		312	온천1동	06055059	74	187	온천2동	06056081	59	184
		411	온천3동	06057179	52	78	온천3동	06057061	56	138
		412	온천1동	06055104	60	84	온천3동	06057024	54	141
		412	온천2동	06056091	46	80	사직2동	06059040	65	231
남	구	111	감만1동	07062084	53	151	감만1동	07062086	63	179
		411	대연3동	07053092	94	116	대연6동	07056018	46	130
북	구	211	화명1동	08055127	65	141	화명1동	08055046	60	193
		211	화명1동	08055155	78	187	덕천1동	08056072	67	198
		211	덕천1동	08056023	49	119	덕천1동	08056053	69	139
		212	화명1동	08055142	66	148	화명1동	08055061	66	248

구	군	층화 코드	읍·면·동	조사구번호	가구 수	가구원 수	읍·면·동	변경된 조사구번호	가구 수	가구원 수		
해운대구		211	중2동	09054031	43	126	우2동	09052109	60	183		
		211	중2동	09054059	53	120	반여3동	09059055	73	215		
		212	중2동	09054051	56	142	우2동	09052073	59	174		
		412	중1동	09053114	48	53	중2동	09054119	53	119		
사하구		211	다대1동	10062175	58	185	다대1동	10062094	67	237		
		212	다대1동	10062173	50	139	다대1동	10062087	69	221		
금정구		111	부곡2동	11058008	45	84	부곡2동	11058009	66	185		
		111	장전2동	11062029	43	75	장전3동	11063023	64	145		
		111	장전2동	11062052	40	121	장전3동	11063047	54	159		
		111	구서1동	11069053	66	173	구서2동	11070129	48	143		
		112	장전1동	11061025	65	208	장전1동	11061031	58	125		
		412	구서1동	11069044	72	226	구서2동	11070130	63	201		
강서구		111	대저1동	12051069	40	112	대저2동	12052001	101	312		
		112	녹산동	12056032	63	182	녹산동	12056021	55	169		
		121	녹산동	12056003	75	216	명지동	12054027	77	253		
		121	녹산동	12056006	53	136	명지동	12054031	59	206		
		121	천가동	12057001	45	121	대저1동	12051011	63	181		
		121	천가동	12057005	46	109	대저1동	12051027	50	153		
		121	천가동	12057006	60	149	대저1동	12051070	58	136		
		121	천가동	12057009	48	88	대저2동	12052018	73	204		
		121	천가동	12057011	69	151	대저2동	12052021	50	160		
		121	천가동	12057017	48	98	대저2동	12052027	88	244		
		121	천가동	12057018	53	115	강동동	12053028	73	198		
		연제구		111	거제2동	13052007	60	168	거제2동	13052008	55	139
				111	연산1동	13055062	63	156	연산1동	13055021	49	132
111	연산1동			13055074	60	168	연산1동	13055091	68	159		
111	연산8동			13062005	61	178	연산8동	13062044	61	162		
111	연산8동			13062019	59	163	연산8동	13062056	65	180		
212	거제1동			13051128	78	121	거제1동	13051129	63	82		
수영구		111	민락동	14060099	59	148	민락동	14060115	65	204		
		211	망미1동	14054115	65	86	망미2동	14055028	80	249		
		411	광안1동	14056021	61	104	광안1동	14056037	59	177		
사상구		412	패법동	15057144	52	71	주례2동	15061002	86	185		
기장군		111	일광면	31031045	49	110	일광면	31031046	48	132		
		111	기장읍	31011165	43	57	기장읍	31011121	60	169		
		121	정관면	31032021	65	212	정관면	31032028	56	141		
		121	철마면	31033003	66	185	철마면	31033005	61	186		
		411	기장읍	31011166	57	65	기장읍	31011049	62	201		
		411	기장읍	31011171	40	46	기장읍	31011077	48	157		
		411	기장읍	31011174	55	68	기장읍	31011134	42	137		
		411	일광면	31031002	47	94	일광면	31031052	50	135		

2. 구·군별 통계치 및 부산광역시 통계치 추정

「2009 부산사회조사」 결과를 가지고 위 III장 3절의 추정방법에서 제시한 추정식을 이용하여 가구단위 문항(월 평균 총 소득)과 가구원단위 문항(부산시 경제에 관한 견해)에 대한 구·군별 및 부산광역시 전체 평균추정치, 표준오차, 효율 등을 기장군의 예시를 통해 살펴보고자 한다.

가. 구·군별 통계치 추정

(1) 구·군별 설계 가중치

구·군별 설계 가중치를 구하는 방법은 다음과 같다.

$$w_{ghk} = \frac{N_{gh}}{n_{gh}} \times \frac{M_{ghk}}{m_{ghk}}$$

기장군의 경우를 살펴보면 다음 <표 4-11>과 같다.

조사구번호가 31011069인 첫 번째 표본조사구의 설계 가중치는

$$w_{16,1,1} = \frac{N_{16,1}}{n_{16,1}} \times \frac{M_{16,1,1}}{m_{16,1,1}} = \frac{58}{14} \times \frac{44}{10} = 18.23 \text{ 이다.}$$

조사구번호가 31011118인 15번째 표본조사구의 설계 가중치는

$$w_{16,2,15} = \frac{N_{16,2}}{n_{16,2}} \times \frac{M_{16,2,15}}{m_{16,2,15}} = \frac{5}{1} \times \frac{62}{10} = 31.00 \text{ 이다.}$$

조사구번호가 31011043인 101번째 표본조사구의 설계 가중치는

$$w_{16,11,101} = \frac{N_{16,11}}{n_{16,11}} \times \frac{M_{16,11,101}}{m_{16,11,101}} = \frac{25}{6} \times \frac{77}{10} = 32.08 \text{ 이다.}$$

조사구번호가 31011017인 마지막 107번째 표본조사구의 설계 가중치는

$$w_{16,12,107} = \frac{N_{16,12}}{n_{16,12}} \times \frac{M_{16,12,107}}{m_{16,12,107}} = \frac{6}{1} \times \frac{55}{10} = 33.00 \text{ 이다.}$$

<표 4-11> 기장군의 설계 가중치 예시

조사구번호	조사구(k)	층(h)	N_{gh}	n_{gh}	M_{ghk}	m_{ghk}	W_{ghk}
31011069	1	1	58	14	44	10	18.23
...
31011118	15	2	5	1	62	10	31.00
...
31011002	16	3	134	32	49	10	20.52
...
31033021	48	4	5	1	40	10	20.00
...
31011043	101	11	25	6	77	10	32.08
...
31011017	107	12	6	1	55	10	33.00

(2) 가구 단위의 조사(2009 부산사회 조사표-소득에 대한 문항)

2009 부산사회조사 문항 중 가구의 월 평균 총 소득에 대한 문항을 주요 항목으로 선택하여 모집단의 특성을 추정하고자 한다.

(가) 구·군별 합계추정치

$$\hat{Y}_g = \sum_h^{H_g} \sum_k^{n_{gh}} \sum_i^{m_{ghk}} w_{ghk} y_{ghki}$$

기장군의 경우를 살펴보면 다음과 같다.

$$\begin{aligned} \hat{Y}_{16} = & W_{16,1,1} Y_{16,1,1,1} + W_{16,1,1} Y_{16,1,1,2} + \dots + W_{16,1,1} Y_{16,1,1,10} + \dots \\ & + W_{16,12,107} Y_{16,12,107,1} + \dots + W_{16,12,107} Y_{16,12,107,10} \end{aligned}$$

6개의 문항 중 소득에 대한 문항값이 1번인 경우를 예로 들어 기장군의 합계 추정치를 구해보면 다음과 같다.

$$\hat{Y}_{16} = 18.23 \times 0 + 18.23 \times 0 + \dots + 18.23 \times 0 + \dots + 33.00 \times 0 + \dots + 33.00 \times 1 = 6060.12$$

<표 4-12> 구·군별 합계추정치

구 · 군	소득에 대한 합계추정치					
	100만원 미만	100-200만원미만	100-200만원미만	100-200만원미만	100-200만원미만	500만원 이상
중 구	4399.54	7228.24	5625.29	2807.93	982.73	748.58
서 구	11377.25	13870.93	12677.93	6040.91	3075.34	987.71
동 구	8492.79	13162.86	9904.62	4909.81	2666.80	477.16
영 도 구	11204.77	13846.75	15588.19	8616.94	3563.91	1476.50
부산진구	32953.74	42006.29	35208.21	18463.43	8700.26	5598.81
동래구	14599.68	26817.52	22683.24	14252.56	9339.65	4532.98
남 구	18764.60	28942.33	27328.1	17110.23	9142.64	5875.78
북 구	15729.21	29699.54	31651.59	20801.66	7932.59	2650.47
해운대구	23423.84	32862.10	36505.97	21286.57	12967.02	8885.10
사하구	21866.68	30663.44	31322.25	21151.48	11212.86	4825.2
금정구	19354.91	23876.28	22645.31	14092.02	8806.48	6274.8
강서구	3224.47	5538.52	4367.00	2361.50	922.65	265.89
연제구	14523.25	23253.22	19231.62	10067.73	5310.78	3532.44
수영구	10584.64	18218.25	16941.25	9433.13	5383.07	3488.24
사상구	12017.18	29200.45	29347.24	14128.48	4943.99	1967.24
기장군	6060.12	8132.78	6593.15	3472.10	1359.68	271.08

(나) 구·군별 평균추정치

산출된 구·군별 합계추정치를 구·군의 총 가구(가구단위 조사결과 일 때) 혹은 가구원(가구원 단위 조사결과 일 때)으로 나누어 구·군별 합계추정치를 구할 수 있다.

1) K_g (g 번째 구·군의 총 가구) 수 추정

구·군의 총 가구, 가구원을 구하는 방법은 두 가지가 있는데 첫 번째는 2005년 인구주택총조사 및 2006~2007년 신축아파트 조사구 기준으로 산출하는 것이며, 두 번째는 실제 조사 응답결과를 활용한 방법이 있다.

가) 2005년 인구주택총조사 및 신축아파트 조사구 기준에서의 가구 수 추정

2005년 인구주택총조사 및 신축아파트 조사구 기준에서의 부산광역시 전체 가구 수는 1,224,081가구로 부산시 16개 구·군 중 부산진구의 가구 수가 143,629가구로 가장 많고, 강서구의 가구 수가 15,986가구로 가장 적다.

나) 응답결과를 활용한 가구 수 추정

$$\text{구·군별 가구 수 추정 : } \hat{K}_g = \sum_h^{H_g} \sum_k^{n_{gh}} \sum_i^{m_{ghk}} w_{ghk}$$

응답결과로 \hat{K}_g 를 추정한 결과 부산광역시 전체 가구 수는 1,240,682가구로 K_g 모집단 활용 가구 수보다 16,601가구 더 많게 나타났다. 그 결과를 나타내면 아래 <표 4-13>과 같다.

<표 4-13> 구·군의 총 가구 수 추정

(단위 : 개)

구	군	K_g 모집단활용 (인구주택총조사 기준 및 신축아파트조사구 기준)	\hat{K}_g 추정 (응답결과 기준)
		가구수	가구수
중	구	21,179	21,792
서	구	47,920	48,030
동	구	38,890	39,614
영	도	53,710	54,297
부	산	143,629	142,931
동	래	91,075	92,226
남	구	103,188	107,164
북	구	106,851	108,465
해	운	136,838	135,931
사	하	120,540	121,042
금	정	92,288	95,049
강	서	15,986	16,680
연	제	74,842	75,919
수	영	61,502	64,049
사	상	91,047	91,605
기	장	24,596	25,889
합	계	1,224,081	1,240,683

2) 구·군별 평균 추정치 산출

$$\hat{Y}_g = \frac{\hat{Y}_g}{K_g} = \frac{\sum_h^{H_g} \sum_k^{n_{gh}} \sum_i^{m_{ghk}} w_{ghk} y_{ghki}}{K_g}$$

\widehat{Y}_g : g 번째 구·군의 평균 추정치

K_g : g 번째 구·군의 총 가구(가구단위 조사결과 일 때) 혹은 가구원
(가구원 단위 조사결과 일때) 수

그런데 2009년 실제조사 자료를 가지고 산출한 설계가중치를 대입해 구한 합계 추정치를 5년 전 조사된 모집단의 수로 나누어 평균 추정치를 구한 결과 매년 모집단의 변동이 생김으로 인해 평균 추정치의 합이 95~109%사이의 값을 나타내 올바른 결과를 도출할 수 없어 \widehat{K}_g (응답결과 기준)을 사용하여 평균 추정치를 구하고자 한다.

$$\widehat{Y}_g = \frac{\widehat{Y}_g}{\widehat{K}_g} = \frac{\sum_h \sum_k \sum_i w_{ghk} y_{ghki}}{\widehat{K}_g}$$

(다) 구·군별 합계 추정치의 분산 추정치 산출

$$\widehat{var}(\widehat{Y}_g) = \sum_{h=1}^{H_g} \left[N_{gh}^2 \frac{N_{gh} - n_{gh}}{N_{gh}} \frac{s_{gh}^2}{n_{gh}} + \frac{N_{gh}}{n_{gh}} \sum_{k=1}^{n_{gh}} M_{ghk}^2 \frac{M_{ghk} - m_{ghk}}{M_{ghk}} \frac{s_{ghk}^2}{m_{ghk}} \right]$$

$\widehat{var}(\widehat{Y}_g)$: 구·군별 합계 추정치의 분산 추정치

$$s_{gh}^2 = \frac{1}{n_{gh} - 1} \sum_{k=1}^{n_{gh}} (\widehat{Y}_{ghk} - \widehat{Y}_{gh})^2$$

: 구·군별 h 번째 층에서의 합계 추정치의 표본분산

$$s_{ghk}^2 = \frac{1}{m_{ghk} - 1} \sum_{i=1}^{m_{ghk}} (y_{ghki} - \bar{y}_{ghk})^2$$

: 구·군별 h 번째 층의 k 번째 표본조사구에서의 합계 추정치의 표본분산

$$\widehat{Y}_{ghk} = \sum_{i=1}^{m_{ghk}} M_{ghk} \frac{y_{ghki}}{m_{ghk}}, \quad \widehat{Y}_{gh} = \frac{1}{n_{gh}} \sum_{k=1}^{n_{gh}} \widehat{Y}_{ghk}, \quad \bar{y}_{ghk} = \frac{1}{m_{ghk}} \sum_{i=1}^{m_{ghk}} y_{ghki}$$

6개의 문항 중 소득에 대한 문항 값이 1번인 경우를 예로 들어 기장군의 분산 추정치를 구해보면 다음과 같다.

$$\textcircled{1} \hat{Y}_{ghk} = \sum_{i=1}^{m_{ghk}} M_{ghk} \frac{y_{ghki}}{m_{ghk}} \text{ 산출 예시}$$

$$\hat{Y}_{\text{기장군,1,1}} = \sum_{i=1}^{10} 44 \times \frac{Y_{\text{기장군,1,1},i}}{10} = 44 \times \frac{0}{10} + 44 \times \frac{0}{10} + \dots + 44 \times \frac{1}{10} = 13.2$$

...

$$\hat{Y}_{\text{기장군,12,107}} = \sum_{i=1}^{10} 55 \times \frac{Y_{\text{기장군,12,107},i}}{10} = 55 \times \frac{0}{10} + 55 \times \frac{0}{10} + \dots + 55 \times \frac{0}{10} = 0$$

$$\textcircled{2} \hat{Y}_{gh} = \frac{1}{n_{gh}} \sum_{k=1}^{n_{gh}} \hat{Y}_{ghk} \text{ 산출 예시}$$

$$\begin{aligned} \hat{Y}_{\text{기장군,1}} &= \frac{1}{14} \sum_{k=1}^{14} \hat{Y}_{\text{기장군,1},k} \\ &= \frac{1}{14} \times 13.2 + \frac{1}{14} \times 9.6 + \dots + \frac{1}{14} \times 14.4 + \frac{1}{14} \times 28.8 = 15.98 \end{aligned}$$

...

$$\begin{aligned} \hat{Y}_{\text{기장군,12}} &= \frac{1}{1} \sum_{k=1}^1 \hat{Y}_{\text{기장군,12,107}} \\ &= \frac{1}{1} \times 0 = 0 \end{aligned}$$

$$\textcircled{3} \bar{y}_{ghk} = \frac{1}{m_{ghk}} \sum_{i=1}^{m_{ghk}} y_{ghki} \text{ 산출 예시}$$

$$\bar{Y}_{\text{기장군,1,1}} = \frac{1}{10} \sum_{i=1}^{10} Y_{\text{기장군,1,1},i} = \frac{1}{10} (0+0+1+0+0+1+1+0+0+0) = 0.3$$

...

$$\bar{Y}_{\text{기장군,5,52}} = \frac{1}{10} \sum_{i=1}^{10} Y_{\text{기장군,5,52},i} = \frac{1}{10} (1+1+1+0+1+0+1+1+1+1) = 0.8$$

$$\textcircled{4} s_{gh}^2 = \frac{1}{n_{gh} - 1} \sum_{k=1}^{n_{gh}} (\hat{Y}_{ghk} - \widehat{Y}_{gh})^2 \text{ 산출 예시}$$

$$\begin{aligned} S_{\text{기장군},1}^2 &= \frac{1}{14-1} \sum_{k=1}^{14} (\hat{Y}_{\text{기장군},1,k} - \widehat{Y}_{\text{기장군},1})^2 \\ &= \frac{1}{13} ((13.20 - 15.99)^2 + (9.60 - 15.98)^2 + \dots + (28.80 - 15.98)^2) = 86.81 \end{aligned}$$

$$\textcircled{5} s_{ghk}^2 = \frac{1}{m_{ghk} - 1} \sum_{k=1}^{m_{ghk}} (y_{ghki} - \bar{y}_{ghk})^2 \text{ 산출 예시}$$

$$\begin{aligned} S_{\text{기장군},1,1}^2 &= \frac{1}{10-1} \sum_{k=1}^{10} (Y_{\text{기장군},1,1,i} - \bar{Y}_{\text{기장군},1,k})^2 \\ &= \frac{1}{9} ((0-0.3)^2 + (0-0.3)^2 + \dots + (1-0.3)^2 + (0-0.3)^2) \\ &= 0.01 + 0.01 + 0.05 + 0.01 + 0.01 + 0.05 + 0.05 + 0.01 + 0.01 + 0.01 = 0.23 \end{aligned}$$

$$\textcircled{6} \sum_{h=1}^{H_g} \left[N_{gh}^2 \frac{N_{gh} - n_{gh}}{N_{gh}} \frac{s_{gh}^2}{n_{gh}} + \frac{N_{gh}}{n_{gh}} \sum_{k=1}^{n_{gh}} M_{ghk}^2 \frac{M_{ghk} - m_{ghk}}{M_{ghk}} \frac{s_{ghk}^2}{m_{ghk}} \right] \text{ 산출 예시}$$

$$\begin{aligned} \widehat{Var}(\hat{Y}_{\text{기장군}}) &= (15824.91 + 2764.14) + (0.00 + 286.58) + \dots \\ &= (78161.48 + 1175.85) + (0.00 + 264.00) = 308240.56 \end{aligned}$$

(라) 구·군별 평균 추정치의 분산 추정치, 표준오차 추정치, 효율 산출

1) 구·군별 평균 추정치의 분산 추정치

$$\widehat{var}(\widehat{Y}_g) \approx \frac{\widehat{var}(\hat{Y}_g)}{\widehat{K}_g^2}$$

2) 구·군별 평균 추정치의 표준오차 추정치

$$\widehat{se}(\widehat{Y}_g) = \sqrt{\widehat{var}(\widehat{Y}_g)}$$

기장군 예시처럼 구·군별 합계추정치를 응답결과를 고려하여 구한 \widehat{K}_g 로 나눈 평균 추정치와 합계 분산추정치로 \widehat{K}_g^2 로 나눈 평균 추정치의 표준오차를 구해보면

다음 <표 2-8>과 같다. 각 구·군별 월 평균 소득에 대한 표본평균과 평균 추정치가 비슷한 수치를 가지고 있으나, 응답결과를 활용하여 구한 \hat{K}_g 를 사용한 결과 이 방법은 최근 조사결과를 반영하기 때문에 편의가 작을 수 있지만 일부 표본만을 대상으로 하여 추정치의 대부분의 표준오차가 표본평균의 표준오차보다 큰 경향을 보였다.

단순임의추출이라고 가정 한다면

$$\text{표본평균} = \frac{1}{N_g} y_g, \quad y_g = N_g \text{명 중에서 각 항목별 응답원 수}$$

$$\text{표준오차} = \sqrt{\frac{1}{N_g} (\hat{p}_g(1-\hat{p}_g))} \quad \text{와 같이 계산할 수 있다.}$$

(표본평균)

설계 가중치를 반영한 평균추정치의 효율은

$$\text{효율} = \frac{\text{표본평균의 표준오차(단순임의 추출 가정)}}{\text{설계 가중치 반영 표준오차}} \quad \text{와 같이 정의할 수 있다.}$$

만약 효율이 1보다 크다면, 설계 가중치 반영 표준오차의 효율이 좋다고 할 수 있다. 효율 추정치를 계산한 결과 표본평균의 표준오차(단순임의 추출 가정) 대비 이단집락 추출법의 효율이 대개 0.8 이상으로 나타나 표본설계가 잘 이루어졌다고 할 수 있다.

<표 4-14> 구·군 가구의 월 평균소득에 대한 평균(±표준오차), 효율

구·군	월 평균 소득	표본평균 (±표준오차)	설계 가중치 반영 평균추정치 (±표준오차)	효율
중구	100만원미만	20.79%(±1.24%)	20.19%(±1.32%)	0.94
	100-200만원미만	33.15%(±1.44%)	33.17%(±1.58%)	0.91
	200-300만원미만	25.56%(±1.33%)	25.81%(±1.41%)	0.94
	300-400만원미만	12.73%(±1.02%)	12.88%(±1.07%)	0.95
	400-500만원미만	4.49%(±0.63%)	4.51%(±0.80%)	0.79
	500만원이상	3.28%(±0.54%)	3.44%(±0.63%)	0.86
서구	100만원미만	23.97%(±1.31%)	23.27%(±1.43%)	0.92
	100-200만원미만	29.12%(±1.39%)	28.74%(±1.58%)	0.88
	200-300만원미만	26.40%(±1.35%)	26.92%(±1.67%)	0.81
	300-400만원미만	12.17%(±1.00%)	12.34%(±1.13%)	0.88
	400-500만원미만	6.27%(±0.74%)	6.56%(±0.75%)	0.99
	500만원이상	2.06%(±0.43%)	2.16%(±0.48%)	0.90

구·군	월 평균 소득	표본평균 (±표준오차)	설계 가중치 반영 평균추정치 (±표준오차)	효율
동구	100만원미만	21.16%(±1.25%)	20.90%(±1.55%)	0.81
	100-200만원미만	32.96%(±1.44%)	32.55%(±1.78%)	0.81
	200-300만원미만	25.00%(±1.32%)	25.59%(±1.84%)	0.72
	300-400만원미만	12.55%(±1.01%)	12.66%(±1.21%)	0.83
	400-500만원미만	7.02%(±0.78%)	7.01%(±0.93%)	0.84
	500만원이상	1.31%(±0.35%)	1.29%(±0.34%)	1.03
영도구	100만원미만	20.51%(±1.24%)	20.46%(±1.56%)	0.79
	100-200만원미만	25.66%(±1.34%)	25.67%(±1.56%)	0.86
	200-300만원미만	28.84%(±1.39%)	28.89%(±1.52%)	0.91
	300-400만원미만	15.64%(±1.11%)	15.62%(±1.29%)	0.86
	400-500만원미만	6.65%(±0.76%)	6.60%(±0.78%)	0.97
	500만원이상	2.72%(±0.50%)	2.76%(±0.57%)	0.88
부산진구	100만원미만	22.94%(±1.29%)	21.95%(±1.35%)	0.96
	100-200만원미만	28.93%(±1.39%)	28.72%(±1.58%)	0.88
	200-300만원미만	24.44%(±1.31%)	24.81%(±1.66%)	0.79
	300-400만원미만	13.30%(±1.04%)	13.48%(±1.03%)	1.01
	400-500만원미만	6.27%(±0.74%)	6.61%(±0.94%)	0.79
	500만원이상	4.12%(±0.61%)	4.43%(±0.63%)	0.97
동래구	100만원미만	15.82%(±1.12%)	15.59%(±1.10%)	1.02
	100-200만원미만	29.40%(±1.39%)	28.89%(±1.58%)	0.88
	200-300만원미만	25.00%(±1.32%)	24.65%(±1.39%)	0.95
	300-400만원미만	14.89%(±1.09%)	15.13%(±1.28%)	0.85
	400-500만원미만	9.93%(±0.91%)	10.21%(±1.15%)	0.79
	500만원이상	4.96%(±0.66%)	5.52%(±1.22%)	0.54
남구	100만원미만	17.04%(±1.15%)	16.26%(±1.29%)	0.89
	100-200만원미만	26.87%(±1.36%)	26.00%(±1.47%)	0.93
	200-300만원미만	25.66%(±1.34%)	26.09%(±1.66%)	0.81
	300-400만원미만	16.01%(±1.12%)	16.54%(±1.37%)	0.82
	400-500만원미만	8.80%(±0.87%)	9.33%(±0.97%)	0.90
	500만원이상	5.62%(±0.70%)	5.79%(±0.97%)	0.72
북구	100만원미만	14.79%(±1.09%)	15.09%(±1.76%)	0.62
	100-200만원미만	27.62%(±1.37%)	27.43%(±1.72%)	0.80
	200-300만원미만	28.84%(±1.39%)	28.89%(±1.69%)	0.82
	300-400만원미만	19.01%(±1.20%)	18.95%(±1.25%)	0.96
	400-500만원미만	7.30%(±0.80%)	7.23%(±0.98%)	0.82
	500만원이상	2.43%(±0.47%)	2.41%(±0.54%)	0.87

구·군	월 평균 소득	표본평균 (±표준오차)	설계 가중치 반영 평균추정치 (±표준오차)	효율
해운대구	100만원미만	17.32%(±1.16%)	16.96%(±1.31%)	0.89
	100-200만원미만	24.16%(±1.31%)	24.11%(±1.48%)	0.89
	200-300만원미만	26.97%(±1.36%)	26.91%(±1.39%)	0.98
	300-400만원미만	15.73%(±1.11%)	16.11%(±1.16%)	0.96
	400-500만원미만	9.46%(±0.90%)	9.49%(±0.93%)	0.97
	500만원이상	6.37%(±0.75%)	6.43%(±1.07%)	0.70
사하구	100만원미만	17.79%(±1.17%)	17.65%(±1.71%)	0.68
	100-200만원미만	25.37%(±1.33%)	25.20%(±1.37%)	0.97
	200-300만원미만	25.84%(±1.34%)	25.75%(±1.78%)	0.75
	300-400만원미만	17.60%(±1.17%)	17.80%(±1.22%)	0.96
	400-500만원미만	9.08%(±0.88%)	9.24%(±1.01%)	0.87
	500만원이상	4.31%(±0.62%)	4.37%(±0.82%)	0.76
금정구	100만원미만	21.07%(±1.25%)	20.68%(±1.58%)	0.79
	100-200만원미만	25.09%(±1.33%)	25.29%(±1.64%)	0.83
	200-300만원미만	23.50%(±1.30%)	23.85%(±1.46%)	0.89
	300-400만원미만	14.61%(±1.08%)	14.66%(±1.24%)	0.87
	400-500만원미만	9.18%(±0.88%)	9.01%(±0.98%)	0.90
	500만원이상	6.55%(±0.76%)	6.51%(±1.07%)	0.71
강서구	100만원미만	19.10%(±1.20%)	18.55%(±1.38%)	0.87
	100-200만원미만	32.87%(±1.44%)	32.49%(±1.40%)	1.03
	200-300만원미만	26.31%(±1.35%)	26.71%(±1.56%)	0.87
	300-400만원미만	14.70%(±1.08%)	15.19%(±1.27%)	0.85
	400-500만원미만	5.52%(±0.70%)	5.61%(±0.74%)	0.95
	500만원이상	1.50%(±0.37%)	1.45%(±0.34%)	1.09
연제구	100만원미만	18.91%(±1.20%)	18.24%(±1.30%)	0.92
	100-200만원미만	30.43%(±1.41%)	30.07%(±1.64%)	0.86
	200-300만원미만	25.47%(±1.33%)	25.40%(±1.47%)	0.90
	300-400만원미만	13.30%(±1.04%)	13.74%(±1.06%)	0.98
	400-500만원미만	7.21%(±0.79%)	7.50%(±0.93%)	0.85
	500만원이상	4.68%(±0.65%)	5.04%(±0.86%)	0.76
수영구	100만원미만	17.23%(±1.16%)	17.02%(±1.26%)	0.92
	100-200만원미만	28.65%(±1.38%)	28.29%(±1.55%)	0.89
	200-300만원미만	26.40%(±1.35%)	26.02%(±1.40%)	0.96
	300-400만원미만	14.61%(±1.08%)	14.77%(±1.23%)	0.88
	400-500만원미만	7.96%(±0.83%)	8.32%(±0.79%)	1.05
	500만원이상	5.15%(±0.68%)	5.58%(±1.03%)	0.66

구·군	월 평균 소득	표본평균 (±표준오차)	설계 가중치 반영 평균추정치 (±표준오차)	효율
사상구	100만원미만	13.11%(±1.03%)	13.51%(±1.51%)	0.68
	100-200만원미만	31.65%(±1.42%)	31.76%(±1.49%)	0.95
	200-300만원미만	32.02%(±1.43%)	31.80%(±1.57%)	0.91
	300-400만원미만	15.54%(±1.11%)	15.26%(±1.36%)	0.82
	400-500만원미만	5.62%(±0.70%)	5.51%(±0.75%)	0.93
	500만원이상	2.06%(±0.43%)	2.17%(±0.51%)	0.84
기장군	100만원미만	23.13%(±1.29%)	22.89%(±2.14%)	0.60
	100-200만원미만	30.99%(±1.42%)	30.88%(±2.03%)	0.70
	200-300만원미만	25.84%(±1.34%)	26.18%(±1.97%)	0.68
	300-400만원미만	13.48%(±1.05%)	13.55%(±1.25%)	0.84
	400-500만원미만	5.43%(±0.69%)	5.46%(±0.75%)	0.92
	500만원이상	1.12%(±0.32%)	1.04%(±0.29%)	1.10

(3) 가구원 단위의 조사(2009 부산사회통계 조사표-부산시 지역 경제에 대한 견해 문항)

가구원 단위의 조사인 부산시 지역 경제에 대한 견해 문항의 응답을 가지고 구·군별, 부산시 전체 추정을 실시하고자 한다.

(가) 구·군별 합계추정치

구·군별 합계추정치 계산 결과는 다음 <표 4-15>와 같다.

<표 4-15> 구·군별 합계추정치

구 · 군	부산시 지역경제 견해에 대한 구·군별 합계추정치				
	매우 좋다	좋은 편이다	그저 그렇다	어려운 편이다	매우 어렵다
중 구	117.43	1703.94	16011.92	30367.95	8887.88
서 구	112.91	2167.63	39338.15	63173.71	22059.27
동 구	38.44	1547.22	27693.89	57988.71	18566.75
영 도 구	60.12	2956.27	40710.00	81647.14	19643.49
부산진구	109.32	9544.49	102141.30	200810.95	60316.77
동 래 구	1618.78	10819.19	79026.66	130871.29	35859.09

구 · 군	부산시 지역경제 견해에 대한 구·군별 합계추정치				
	매우 좋다	좋은 편이다	그저 그렇다	어려운 편이다	매우 어렵다
남 구	661.18	6901.86	82783.45	140318.28	51895.13
북 구	570.75	10488.22	86142.49	147918.59	41322.39
해운대구	1340.38	16607.65	122526.55	166498.96	49296.40
사 하 구	348.61	7436.38	103563.11	171426.42	36792.02
금 정 구	535.89	8190.81	76368.14	133044.82	39683.70
강 서 구	50.13	1056.50	14170.47	22669.25	7061.03
연 제 구	401.82	8339.94	65751.15	94725.66	32561.34
수 영 구	364.66	4602.27	48398.77	91817.50	24989.30
사 상 구	982.20	9969.42	76113.28	123250.59	42440.07
기 장 군	292.39	2516.07	25400.10	33510.87	7195.26

(나) 구·군별 평균추정치

산출된 구·군별 합계추정치를 구·군의 총 가구원(가구원 단위 조사결과일 때)으로 나누어 구·군별 합계추정치를 구할 수 있다.

1) K_g (g 번째 구·군의 총 가구원) 수 추정

구·군의 총 가구원을 구하는 방법으로 가구 단위의 조사에서 언급한 이유로 인해 두 번째 방법인 실제 조사 응답결과를 활용하여 계산해 나가 고자 한다. 추정 결과 표본추출틀로 2005년 인구주택총조사 부산광역시 100%조사구에 2006~2007년도 신축아파트 조사구를 추가함으로 부산광역시 15세 이상 가구원이 3,307,203명으로 실제 예상과 차이가 나는 추정 결과가 나타났다.

가) 응답결과를 활용한 가구원 수 추정

$$\text{구·군별 가구원 수 추정} : \hat{K}_g = \sum_h \sum_k \sum_i^{n_{gh} m_{ghk}} w_{ghk} x_{ghki}$$

x_{ghki} : g 구·군에서 h 번째 층의 k 번째 표본조사구에서 표본 가구내에서의 조사결과 나타난 가구원 수

<표 4-16> 구·군의 총 가구원 수 추정

(단위 : 명)

구	군	\widehat{K}_g 추정 (응답결과 기준) 15세 이상 가구원	구	군	\widehat{K}_g 추정 (응답결과 기준) 15세 이상 가구원	
중	구	57,089	해운대	구	356,270	
서	구	126,852	사	하	구	319,567
동	구	105,835	금	정	구	257,823
영	도	구	강	서	구	45,007
부	산	진	연	제	구	201,780
동	래	구	수	영	구	170,173
남	구	282,560	사	상	구	252,756
북	구	286,442	기	장	군	68,914
		부산광역시			3,307,203	

2) 구·군별 평균 추정치 산출

$$\widehat{Y}_g = \frac{\widehat{Y}_g}{\widehat{K}_g} = \frac{\sum_h \sum_k \sum_i w_{ghk} y_{ghki}}{\widehat{K}_g}$$

\widehat{Y}_g : g 번째 구·군의 평균 추정치

K_g : 응답된 결과를 이용한 g 번째 구·군의 추정된 총 가구원 수

(다) 구·군별 합계 추정치의 분산 추정치 산출

$$\widehat{var}(\widehat{Y}_g) = \sum_{h=1}^{H_g} \left[N_{gh}^2 \frac{N_{gh} - n_{gh}}{N_{gh}} \frac{s_{gh}^2}{n_{gh}} + \frac{N_{gh}}{n_{gh}} \sum_{k=1}^{n_{gh}} M_{ghk}^2 \frac{M_{ghk} - m_{ghk}}{M_{ghk}} \frac{s_{ghk}^2}{m_{ghk}} \right]$$

$\widehat{var}(\widehat{Y}_g)$: 구·군별 합계 추정치의 분산 추정치

$$s_{gh}^2 = \frac{1}{n_{gh} - 1} \sum_{k=1}^{n_{gh}} (\widehat{Y}_{ghk} - \widehat{Y}_{gh})^2$$

: 구·군별 h 번째 층에서의 합계 추정치의 표본분산

$$s_{ghk}^2 = \frac{1}{m_{ghk} - 1} \sum_{k=1}^{m_{ghk}} (y_{ghki} - \bar{y}_{ghk})^2$$

: 구·군별 h 번째 층의 k 번째 표본조사구에서의 합계 추정치의 표본분산

$$\hat{Y}_{ghk} = \sum_{i=1}^{m_{ghk}} M_{ghk} \frac{y_{ghki}}{m_{ghk}}, \quad \hat{Y}_{gh} = \frac{1}{n_{gh}} \sum_{k=1}^{n_{gh}} \hat{Y}_{ghk}, \quad \bar{y}_{ghk} = \frac{1}{m_{ghk}} \sum_{i=1}^{m_{ghk}} y_{ghki}$$

(라) 구·군별 평균 추정치의 분산 추정치, 표준오차 추정치, 효율 산출

1) 구·군별 평균 추정치의 분산 추정치

$$\widehat{var}(\hat{Y}_g) \approx \frac{\widehat{var}(\hat{Y}_g)}{\hat{K}_g^2}$$

2) 구·군별 평균 추정치의 표준오차 추정치

$$\widehat{se}(\hat{Y}_g) = \sqrt{\widehat{var}(\hat{Y}_g)}$$

기장군 예시처럼 구·군별 합계추정치를 응답결과를 고려하여 구한 \hat{K}_g 로 나눈 평균 추정치와 합계 분산추정치를 \hat{K}_g^2 로 나눈 평균 추정치의 표준오차를 구해보면 다음 <표 4-17>과 같다. 가구단위의 조사와 마찬가지로 대부분의 표준오차 추정치가 표본평균의 표준오차보다 큰 경향을 보이며, 효율 또한 대부분 80% 이상이었던 가구단위와는 달리 60~70% 이상의 효율을 보이고 있다.

<표 4-17> 구·군 가구원의 부산시 지역경제 견해에 대한 평균(±표준오차), 효율

구·군	지역경제 견해	표본평균 (±표준오차)	설계 가중치 반영 평균추정치 (±표준오차)	효율
중구	매우 좋다	0.21(±0.09%)	0.21(±0.09%)	1.00
	좋은 편이다	2.93(±0.32%)	2.98(±0.40%)	0.80
	그저 그렇다	28.02(±0.85%)	28.05(±1.32%)	0.64
	어려운 편이다	53.26(±0.94%)	53.19(±1.62%)	0.58
	매우 어렵다	15.57(±0.69%)	15.57(±1.04%)	0.66
서구	매우 좋다	0.07(±0.05%)	0.09(±0.09%)	0.56
	좋은 편이다	1.71(±0.24%)	1.71(±0.26%)	0.92
	그저 그렇다	31.03(±0.87%)	31.01(±1.53%)	0.57
	어려운 편이다	49.70(±0.94%)	49.80(±1.71%)	0.55
	매우 어렵다	17.49(±0.72%)	17.39(±1.41%)	0.51

구·군	지역경제 견해	표본평균 (±표준오차)	설계 가중치 반영 평균추정치 (±표준오차)	효율
동구	매우 좋다	0.04(±0.03%)	0.04(±0.04%)	0.75
	좋은 편이다	1.44(±0.22%)	1.46(±0.31%)	0.71
	그저 그렇다	25.97(±0.82%)	26.17(±1.35%)	0.61
	어려운 편이다	54.74(±0.93%)	54.79(±1.78%)	0.52
	매우 어렵다	17.82(±0.72%)	17.54(±1.32%)	0.55
영도구	매우 좋다	0.04(±0.04%)	0.04(±0.04%)	1.00
	좋은 편이다	2.03(±0.26%)	2.04(±0.37%)	0.70
	그저 그렇다	28.09(±0.84%)	28.07(±1.42%)	0.59
	어려운 편이다	56.46(±0.93%)	56.30(±1.53%)	0.61
	매우 어렵다	13.38(±0.64%)	13.55(±1.01%)	0.63
부산진구	매우 좋다	0.04(±0.04%)	0.03(±0.03%)	1.33
	좋은 편이다	2.50(±0.30%)	2.56(±0.37%)	0.81
	그저 그렇다	27.41(±0.84%)	27.39(±1.43%)	0.59
	어려운 편이다	53.63(±0.94%)	53.85(±1.97%)	0.48
	매우 어렵다	16.42(±0.70%)	16.17(±1.06%)	0.66
동래구	매우 좋다	0.64(±0.15%)	0.63(±0.24%)	0.63
	좋은 편이다	4.08(±0.36%)	4.19(±0.63%)	0.57
	그저 그렇다	30.83(±0.84%)	30.61(±1.18%)	0.71
	어려운 편이다	50.69(±0.91%)	50.69(±1.41%)	0.65
	매우 어렵다	13.77(±0.63%)	13.89(±0.87%)	0.72
남구	매우 좋다	0.25(±0.09%)	0.23(±0.10%)	0.90
	좋은 편이다	2.48(±0.29%)	2.44(±0.28%)	1.04
	그저 그렇다	28.64(±0.85%)	29.30(±1.45%)	0.59
	어려운 편이다	49.65(±0.94%)	49.66(±1.58%)	0.59
	매우 어렵다	18.99(±0.74%)	18.37(±1.44%)	0.51
북구	매우 좋다	0.21(±0.09%)	0.20(±0.09%)	1.00
	좋은 편이다	3.72(±0.36%)	3.66(±0.39%)	0.92
	그저 그렇다	30.16(±0.86%)	30.07(±1.21%)	0.71
	어려운 편이다	51.31(±0.94%)	51.64(±1.69%)	0.56
	매우 어렵다	14.60(±0.66%)	14.43(±0.87%)	0.76
해운대구	매우 좋다	0.39(±0.12%)	0.38(±0.13%)	0.92
	좋은 편이다	4.71(±0.40%)	4.66(±0.46%)	0.87
	그저 그렇다	34.49(±0.90%)	34.39(±1.27%)	0.71
	어려운 편이다	46.66(±0.94%)	46.73(±1.44%)	0.65
	매우 어렵다	13.75(±0.65%)	13.84(±1.13%)	0.58
사하구	매우 좋다	0.11(±0.06%)	0.11(±0.08%)	0.75
	좋은 편이다	2.31(±0.28%)	2.33(±0.35%)	0.80
	그저 그렇다	32.52(±0.88%)	32.41(±1.32%)	0.67
	어려운 편이다	53.60(±0.94%)	53.64(±1.63%)	0.58
	매우 어렵다	11.47(±0.60%)	11.51(±0.90%)	0.67

구·군	지역경제 견해	표본평균 (±표준오차)	설계 가중치 반영 평균추정치 (±표준오차)	효율
금정구	매우 좋다	0.21(±0.08%)	0.21(±0.08%)	1.00
	좋은 편이다	3.17(±0.33%)	3.18(±0.40%)	0.83
	그저 그렇다	29.72(±0.85%)	29.62(±1.34%)	0.63
	어려운 편이다	51.52(±0.93%)	51.60(±1.67%)	0.56
	매우 어렵다	15.39(±0.67%)	15.39(±1.03%)	0.65
강서구	매우 좋다	0.10(±0.06%)	0.11(±0.06%)	1.00
	좋은 편이다	2.25(±0.28%)	2.35(±0.33%)	0.85
	그저 그렇다	31.22(±0.86%)	31.48(±1.41%)	0.61
	어려운 편이다	50.33(±0.93%)	50.37(±1.63%)	0.57
	매우 어렵다	16.09(±0.68%)	15.69(±1.02%)	0.67
연제구	매우 좋다	0.21(±0.09%)	0.20(±0.08%)	1.13
	좋은 편이다	4.08(±0.37%)	4.13(±0.61%)	0.61
	그저 그렇다	32.64(±0.88%)	32.59(±1.45%)	0.61
	어려운 편이다	46.97(±0.94%)	46.95(±1.44%)	0.65
	매우 어렵다	16.09(±0.69%)	16.14(±1.09%)	0.63
수영구	매우 좋다	0.21(±0.09%)	0.21(±0.10%)	0.90
	좋은 편이다	2.64(±0.30%)	2.70(±0.33%)	0.91
	그저 그렇다	28.21(±0.85%)	28.44(±1.36%)	0.63
	어려운 편이다	53.81(±0.94%)	53.96(±1.58%)	0.59
	매우 어렵다	15.13(±0.67%)	14.68(±0.87%)	0.77
사상구	매우 좋다	0.41(±0.12%)	0.39(±0.13%)	0.92
	좋은 편이다	3.87(±0.36%)	3.94(±0.70%)	0.51
	그저 그렇다	29.90(±0.84%)	30.11(±1.44%)	0.58
	어려운 편이다	48.93(±0.92%)	48.76(±1.29%)	0.71
	매우 어렵다	16.89(±0.69%)	16.79(±1.01%)	0.68
기장군	매우 좋다	0.42(±0.12%)	0.42(±0.15%)	0.80
	좋은 편이다	3.54(±0.35%)	3.65(±0.46%)	0.76
	그저 그렇다	37.16(±0.90%)	36.86(±1.35%)	0.67
	어려운 편이다	48.44(±0.94%)	48.63(±1.62%)	0.58
	매우 어렵다	10.44(±0.57%)	10.44(±0.74%)	0.77

나. 부산광역시 전체 통계치 추정

구·군별 가중치, 총 평균 추정치 등을 추정하여 부산광역시 전체 통계치 추정을 실시하고자 한다.

(1) 구·군별 가중치

새로운 표본설계는 각 구·군별 크기에 상관없이 동일하게 1,068가구를 대상으로 조사하였기 때문에, 구·군별 가중치(= W_g)를 계산해야 한다. 가중치는 다음 식과 같이 구·군별 총 가구 혹은 가구원 수에서 부산광역시 전체 가구 혹은 가구원 수를 나눈 값이다.

<표 4-18> 구·군별 가중치(가구 단위, 가구원 단위)

(단위 : 개, 명, %)

구 · 군	전체 가구 수 (= \hat{K}_g)	구·군별 가중치 (= W_g)	전체 가구원 수 (= \hat{K}_g)	구·군별 가중치 (= W_g)
중 구	21,792	1.8%	57,089	1.7%
서 구	48,030	3.9%	126,852	3.8%
동 구	39,614	3.2%	105,835	3.2%
영 도 구	54,297	4.4%	145,017	4.4%
부산진구	142,931	11.4%	372,923	11.3%
동 래 구	92,226	7.4%	258,195	7.8%
남 구	107,164	8.6%	282,560	8.5%
북 구	108,465	8.7%	286,442	8.7%
해운대구	135,931	11.0%	356,270	10.8%
사 하 구	121,042	9.8%	319,567	9.7%
금 정 구	95,049	7.7%	257,823	7.8%
강 서 구	16,680	1.3%	45,007	1.4%
연 제 구	75,919	6.1%	201,780	6.1%
수 영 구	64,049	5.2%	170,173	5.1%
사 상 구	91,605	7.4%	252,756	7.6%
기 장 군	25,889	2.1%	68,914	2.1%
합 계	1,240,683	100.0%	3,307,203	100.0%

(2) 부산광역시 총 평균 추정치

산출된 구·군별 가중치를 이용한 부산광역시 총 평균 추정치는 다음과 같다.

$$\text{부산광역시 총 평균 추정치} : \hat{Y} = \sum_{g=\text{중구}}^{\text{기장군}} W_g \hat{Y}_g$$

하나의 예로 가구의 월 평균 총소득이 100만원 미만인 비율의 부산광역시 총 평균추정치는 위 공식을 이용하여 계산된 값은 18.11%이다.

(3) 부산광역시 총 평균 추정치의 표준오차

산출된 구·군별 가중치를 이용한 부산광역시 총 평균 추정치의 표준오차는 다음과 같다.

$$\text{분산 추정치} : \widehat{var}(\widehat{Y}) = \sum_{g=\text{중구}}^{\text{기장군}} W_g^2 \widehat{var}(\widehat{Y}_g)$$

$$\text{표준오차 추정치} : \widehat{se}(\widehat{Y}) = \sqrt{\widehat{var}(\widehat{Y})}$$

<표 4-19>, <표 4-20>에서와 같이 가구 단위 문항(월 평균 총 소득)과 가구원 단위 문항(부산시 지역 경제에 대한 견해)의 표본평균과 설계 가중치 반영 평균추정치의 비율은 비슷한 값을 가지며, 효율을 추정한 결과, 구·군별 효율과 같이 가구 단위 문항은 80% 내외의 값을 보였으며, 가구원 단위 문항은 60% 내외의 값을 보였다.

<표 4-19> 부산광역시 총 평균(±표준오차) 추정치, 효율
(가구 단위-월 평균 총 소득)

구분	표본평균 (±표준오차)	설계 가중치 반영 평균추정치(±표준오차)	효율
100만원미만	19.04(±0.30%)	18.11(±0.41%)	0.73
100-200만원미만	28.93(±0.35%)	27.72(±0.43%)	0.81
200-300만원미만	26.38(±0.34%)	26.49(±0.44%)	0.77
300-400만원미만	14.74(±0.27%)	15.40(±0.34%)	0.79
400-500만원미만	7.26(±0.20%)	7.88(±0.27%)	0.74
500만원 이상	3.64(±0.14%)	4.37(±0.24%)	0.58

<표 4-20> 부산광역시 총 평균(±표준오차) 추정치, 효율
(가구원 단위-부산시 지역 경제에 대한 견해)

구분	표본평균 (±표준오차)	실제 가중치 반영 평균추정치(±표준오차)	효율
매우 좋다	0.22(±0.02%)	0.23(±0.03%)	0.67
좋은 편이다	2.97(±0.08%)	3.17(±0.12%)	0.67
그저 그렇다	30.37(±0.22%)	30.43(±0.37%)	0.59
어려운 편이다	51.23(±0.23%)	51.10(±0.45%)	0.51
매우 어렵다	15.21(±0.17%)	15.07(±0.29%)	0.59



V. 요약 및 결론

부산 시민의 의식을 조사하는 『부산 사회조사』는 2008년도 이전까지의 표본 설계 방식에서는 조사시점 모집단(부산광역시 거주 15세 이상 가구원)을 파악하는데 장기간의 검토시간이 필요하며, 자료수집의 한계로 인해 거주지 특성 하나의 변수만을 증화지표로 사용하여 통계치의 효율이 떨어지는 등의 여러 가지 문제점으로 인해 새로운 표본설계가 필요하였다. 따라서, 2009년 2월 부산광역시에서는 새로운 표본설계를 수행하였고, 본 연구는 새로운 표본설계로 추출된 표본을 대상으로 『2009 부산사회조사』 실시 후 결과 및 주요 개선사항을 도출하고, 구·군별 및 부산광역시 통계치를 추정하여, 『2010 인구주택총조사』 결과가 공표되는 결과를 활용할 수 있는 2012년 사회조사부터 『2010 인구주택총조사』 부산광역시 100% 조사구를 표본추출틀로 하여 부산광역시에서 자체적으로 수행해야 할 표본설계에 대비하는데 목적이 있다.

제안된 새로운 표본설계에서 표본크기로 구·군별 동일하게 1,068가구를 대상으로 하는 것은 정해진 예산과 조사업무 관리, 조사원의 수행량 하에 적당한 크기이며, 표본추출틀로 매년 변화하는 모집단에 대비하여 2005년 인구주택총조사 부산광역시 100%조사구에 2006~2007년 신축아파트조사구를 추가하였지만, 삭제되는 조사구는 반영되지 못해 실제 구·군별 총 가구 및 가구원 수 추정시 예상보다 많은 가구원 수가 추정되는 문제점이 생겼다. 따라서 통계청의 방식과 동일하게 5년 주기의 인구주택총조사 조사구만을 표본추출틀로 활용해야 한다고 판단된다.

조사모집단으로 조사대상 가구의 만 15세 이상 가구원만을 조사모집단으로 활용해야 한다는 제안은 적절하며, 조사대상인 1인가구를 실제 조사 때 조사 어려움으로 인해 몇몇 다른 가구로 대체해 주었는데 올바른 추정을 위해 1인가구 또한 반드시 조사가 이루어져야 할 것이다.

분류지표로 3개의 층화변수(주택유형, 농림어가 비율, 가구주 학력)와 2개의 체크변수(경제활동가능 가구주 연령, 주택의 자가소유 비율) 또한 조사결과 적절하며, 표본조사구 선정방법 또한 읍·면·동별로 표본이 골고루 배분되었다. 조사구내 조사대상 가구 수로 5가구를 조사할 경우 214개의 많은 조사구가 필요하며 15가구를 조사할 경우 다른 조사와 대상이 중복될 가능성이 높아 제안된 10가구가 가장 적당하며, 예비표본조사구로 구·군별로 107개 표본조사구에 1대 1로 대응되는 107개 예비표본조사구를 추출함으로써 조사시 조사구가 대체되더라도 대체 조사구가 부족한 경우가 발생하지 않았다. 또한 실제 조사에서 가장 중요한 점은 조사원의 자질과 그 역할로 새로운 표본설계의 전체적인 개념을 잘 주시시켜 교체 조사구가 발생하는 경우 시에 보고 없이 임의적으로 조사구를 교체해 조사결과 추정시 문제가 발생하지 않도록 각별히 주의해야 한다.

2012년부터 부산광역시에서 자체 수행해야 할 새로운 표본설계에 위에서 제시한 사항들을 반영하여 표본의 대표성 및 신뢰성을 높일 수 있는 표본을 추출하여 더욱 정도가 높은 통계치가 생산되기를 기대한다.

참고문헌

- [1] 강현철, 박승열, 김지연, 김인수, 이동수, 황재일, 박민규 (2009). 2008년 서울서베이 추출틀 구축 및 표본추출 사례연구, *조사연구*, Vol. 10, No. 3, pp. 157-172.
- [2] 경상북도 (2007). *경북인의 생활과 의식조사 표본설계 최종보고서*.
- [3] 김영원, 류제복, 박진우, 홍기학 (2006). *표본조사의 이해와 활용*, 톱스커피레이션코리아.
- [4] 김영원, 조선경 (1996). 표본조사에서 항목 무응답 대체 방법, *한국통계학회논문집* 제3권 3호, pp. 145-159.
- [5] 농림부 (2007). *농촌통계조사 표본설계*.
- [6] 도세록, 이관제 (2006). 국민건강 검진조사의 무응답 대체에 관한 연구, *Journal of the Korean Data Analysis Society*, Vol. 7, No. 3, pp. 785-795.
- [7] 박홍래 (2000). *통계조사론*, 개정판, 영지문화사.
- [8] 부산광역시 (2008). *부산의 사회조사 조사지침서*.
- [9] 부산광역시 (2009). *부산의 사회조사보고서*.
- [10] 부산광역시 (2009). *부산의 사회지표*.
- [11] 부산광역시 (2009). *부산의 사회통계조사 표본설계 최종보고서*.
- [12] 성내경 (2003). *표본조사방법론*, 자유아카데미.
- [13] 안정용, 김대경, 최경호 (2005). 지역통계의 문제점과 개선방안, *Journal of the Korean Data Analysis Society*, Vol. 7, No. 6, pp. 2037-2047.
- [14] 최경호, 이강진 (2008). 전라북도 사회통계 소개 및 제언, *Journal of the Korean Data Analysis Society*, Vol. 10, No. 6, pp. 3435-3442.

- [15] 통계청 (2008). *구미시 사회통계조사 표본설계*.
- [16] 통계청 (2009). *사회통계조사보고서*.
- [17] 통계청 (2008). *충북 사회통계조사 표본설계*.
- [18] 통계청 (2007). *통계품질가이드라인*.
- [19] Cochran, W. G. (1997). *Sampling Techniques. 3rd edition*, Wiley, New York.
- [20] De Leeuw, E. and De Heer, W. (2001). *Trends in Households Survey Non-Response, A Longitudinal and International Comparison*, In *Survey Non-Response*, Wiley & Sons, New York.
- [21] Little, R. J. A. and Rubin, D. B. (2002). *Statistical Analysis with Missing Data. 2nd edition*, Wiley, New York.
- [22] SAS (1999). *SAS/STAT User's Guide, Version 8*, SAS Institute Inc., Cary, North Carolina.
- [23] Yamane, T. (1967). *Elementary sampling theory*, Prentice-Hall, Inc., Englewood Cliffs, N.J.