

공학석사 학위논문

공간 빅데이터를 활용한 소지역 상권
매출에 영향을 미치는 요인분석에 관한 연구

A Study on An Analysis on Factors Influencing
the Sales of Sub-region Commercial Area
using Geo-Spatial Big Data

2016년 12월

안양대학교 대학원

도시정보공학과

이 명 호

홍상기 교수 지도
공학석사 학위논문

공간 빅데이터를 활용한 소지역 상권
매출에 영향을 미치는 요인분석에 관한 연구

A Study on An Analysis on Factors Influencing
the Sales of Sub-region Commercial Area using
Geo-Spatial Big Data

이 논문을 이 명 호의 공학석사 학위논문으로 제출함

2016년 12월

안양대학교 대학원


도시정보공학과


이 명 호


인 준 서

이 논문을 이 명 호의 공학석사 학위논문으로 인준함

2016년 12월

위원장 신 증 빈 

위 원 안 종 욱 

위 원 홍 상 기 

ABSTRACT

As a result of the development of information and communication technology, there is accumulating massive amounts of Geo-Spatial Big Data. Geo-Spatial Big Data is the technology to extract meaningful information from Geo-spatial fusion data and support decision making responding with rapidly changing activities by analysing with almost realtime solutions while efficiently collecting, storing and managing structured, semi-structured or unstructured big data. Geo-Spatial Big Data can also be used to analyze local commercial areas.

In this study, Geo-Spatial Big Data construction, processing and analysis methods were presented to analyze the factors affecting sales of small area commercial area using Geo-Spatial Big Data. For this purpose, the analytical methods such as analytical methods were established through the precedent research review. Regression analysis was performed for each sub-region and Eup-Myeon-Dong unit using the Geographical Weighted Regression Model. Through this, the explanatory power of the research model was evaluated through the Local-R-Squares and the unit analysis of the small area and the Eup-Myeon-Dong. In addition, we visualized regression coefficients of major variables and measured regional influence. Finally, this study presented the limitations of the application and the research utilizing the analysis process and results.

국문 초록

정보통신기술의 발전으로 지역개발사업을 비롯한 다양한 분야에서 공간 빅데이터가 축적되고 있다. 공간 빅데이터는 정형·반정형·비정형 공간빅데이터를 효율적으로 수집·저장·관리하는 동시에 공간정보와 융합된 다양한 속성정보에 대해 실시간·통합 분석을 수행하여 의미 있는 정보를 추출함으로써 미래에 대응할 수 있는 기술이라 할 수 있다. 이러한 공간 빅데이터는 지역상권의 분석에서도 유용하게 활용될 수 있다.

본 연구는 공간 빅데이터를 활용하여 소지역 상권매출에 미치는 영향요인을 분석하고자 공간 빅데이터 구축 및 처리, 그리고 분석 방법을 제시하였다. 이를 위해 먼저 선행연구 검토를 통해 분석 방법과 이론을 확립하였다. 분석을 위한 데이터는 카드매출 및 유동 인구 등 공간 빅데이터와 관련 데이터를 수집하여 소지역 단위로 데이터를 구축하였다. 이후 공간가중회귀모형을 적용한 소지역 및 읍면동 단위의 회귀분석을 실시하였다. 이를 통해 국지적 R Square를 활용하여 연구모형의 설명력을 평가하였으며 소지역 분석결과와 읍면동 단위의 분석결과를 비교·검토하였다. 또한 주요 변수별 회귀계수를 시각화하여 지역별 영향력을 측정하였다. 마지막으로 본 연구의 분석과정 및 결과, 연구의 한계 등 결론을 제시하였다.

<목 차>

Abstract	i
국문초록	ii
제1장 서론	1
제1절 연구의 배경 및 목적	1
제2절 연구내용 및 방법	4
제2장 선행연구 검토 및 이론적 고찰	6
제1절 선행연구 검토 및 차별성	6
(1) 선행연구 검토	6
(2) 선행연구와의 차별성	11
제2절 이론적 고찰	14
(1) 공간 빅데이터의 개념	14
(2) 상권의 개념	16
(3) 상권의 매출액 영향요인	18
(4) 분석이론	23
제3장 분석모형 및 자료의 구축	29
제1절 분석변수 및 자료의 구축	29
(1) 분석변수 및 데이터 선정을 위한 기준요소	29
(2) 데이터의 구축	31
제2절 분석방법론	41
제3절 상권매출 영향력 분석 모형	45
제4장 매출액 영향요인 실증분석	47
제1절 분석의 개요	47
제2절 매출 영향력 분석	48

(1) 국지적 R Square	48
(2) 독립변수의 영향력 측정	51
제3절 분석결과의 종합	59
제4절 분석모형의 활용방안	61
 제5장 결론	 62
 참고문헌	 64

〈표 차 례〉

<표 1> 주요 선행연구 검토 및 차별성 요약	13
<표 2> 빅데이터의 정의	14
<표 3> 공간 빅데이터의 정의	15
<표 4> 상권의 유형	17
<표 5> 매출액 결정요인	18
<표 6> 요인별 설명력	19
<표 7> 상권결정요인	19
<표 8> 공간 빅데이터의 데이터 유형, 종류 및 수집기술	27
<표 9> 변수 및 데이터 선정을 위한 기준요소	29
<표 10> 매출에 미치는 영향요인 변수	30
<표 11> 공간 분석단위 비교 · 검토	32
<표 12> 공간 빅데이터 구축 데이터	38
<표 13> 수요·공급요인 구축 데이터	40

〈그 립 차 례〉

<그림 1> 연구 흐름도	5
<그림 2> 상권정보시스템 홈페이지	9
<그림 3> 우리마을가게 상권분석시스템 홈페이지	10
<그림 4> 빅데이터 분석의 유형	25
<그림 5> 공간 빅데이터 분석기법 및 모형 유형	26
<그림 6> 공간 빅데이터 분석 프로세스	26
<그림 7> 소지역 상권 매출에 영향요인분석을 위한 데이터 구축절차	31
<그림 8> 법정동 행정구역 경계	33
<그림 9> 기초단위구 경계	34
<그림 10> 집계구 경계	35
<그림 11> 공간 데이터 분석 단위	36
<그림 12> 매출 영향력 분석 모형	46
<그림 13> 소지역 단위의 국지적 R 스퀘어 추정결과	49
<그림 14> 읍면동 단위의 국지적 R 스퀘어 추정결과	49
<그림 15> 소지역 단위의 주거인구 회귀계수 지역편차	52
<그림 16> 읍면동 단위의 주거인구 회귀계수 지역편차	52
<그림 17> 소지역 단위의 유동인구 회귀계수 지역편차	53
<그림 18> 읍면동 단위의 유동인구 회귀계수 지역편차	53
<그림 19> 소지역 단위의 직장인구 회귀계수 지역편차	55
<그림 20> 읍면동 단위의 직장인구 회귀계수 지역편차	55
<그림 21> 소지역 단위의 평균소득 회귀계수 지역편차	56
<그림 22> 읍면동 단위의 평균소득 회귀계수 지역편차	57
<그림 23> 소지역 단위의 지역 내 상가업소 수 회귀계수 지역편차	58
<그림 24> 읍면동 단위의 지역 내 상가업소 수 회귀계수 지역편차	58

제1장 서론

제1절 연구의 배경 및 목적

오늘날 지역 상권¹⁾은 과거에 비해 그 정착과 확장 양상이 보다 가속화, 다핵화하고 있다. 상권 형성은 빠르고 다양해지는 현대 소비자의 요구를 반영하고 있으며 이를 위해 소비자의 정보를 수집·활용하는 중요성이 부각되고 있다. 과거 상권정보는 입소문과 주요언론, 일부 기관에서 구축, 공유, 확산되었다면 근래에는 일부 민간 대기업에서 다양한 목적으로 상권, 소비자에 관한 부가가치 높은 정보를 생산하고 있다. 이러한 정보는 전략적인 신규상권의 입점, 마케팅 등의 분석을 위한 필수적인 요소로 인식되면서 관련 공공에서의 부가가치 창출, 민간에서의 매출향상 등 다양한 목적으로 활용될 것으로 예상된다. 상권정보의 중요성에도 불구하고 이러한 정보는 권리주체가 일부 대기업으로 한정되어 있으며 일반 시민이 구득할 수 없거나 많은 비용을 지불해야 하는 한계가 있다. 또한 상권정보의 활용 측면에서는 각종 서비스 혹은 시스템 등에 활용하기 위한 분석기반이 미비할 뿐만 아니라 상권, 소비자 정보를 구축하고 있는 기업, 일부 컨설팅 업체에 의해 제공되는 한계가 있다. 이처럼 지역상권의 다양한 정보와 이를 활용한 지역상권에 관한 분석요구가 증가하고 있지만 이를 지원하는 기초 데이터와 분석 기반의 부족으로 성공적인 신규점포의 개발, 상권정착에 어려움이 있다. 상권에 관한 체계부족은 전략적인 상권개발에 있어 약

1) 일정한 지역을 중심으로 재화와 용역의 유통이 이루어지는 공간적 범위를 의미하며 협의의 상권은 상품이 유통되는 일정한 지역을 의미하며 상세권이 라고도 한다. 즉 상행위의 영향이 미치는 범위로써 특정 점포를 기준으로 할 때 그 점포에서 고객을 흡입할 수 있는 지역 및 공간적인 범위를 의미한다.

영향을 미칠 뿐만 아니라 단순 상권의 양적인 팽창과 생애주기 수명을 단축시키고 있으며 이를 해결하기 위해 기술지원 등 질적인 개선이 필수적인 상황이다. 이에 근래에 들어 복잡한 소비자 행태와 상권변화 양상을 예측하는 다양한 분석들이 시도되고 있다. 하지만 이 또한 기존의 단순한 확산·예측분석기법을 적용한 지역상권분석으로 공간현상, 보다 복잡한 사회현상이 고려되지 못하고 있으며 가치창출에도 한계가 있다. 특히 일반적인 통계자료를 근거로 단순 확산 모형을 적용함으로써 공간적인 상호작용에 대한 고려가 미흡한 실정이다. 따라서 과거 상권 분석 패러다임에서 벗어나 상권에 관한 다양한 외부요인을 적용하고 체계적으로 분석할 수 있는 새로운 분석 기법이 반드시 필요하다.

한편 21세기 스마트 혁명에 따른 빅데이터 등장은 다양한 분야의 정보를 융합·활용하여 부가가치를 창출할 수 있는 획기적인 기술로 인식되고 있다. 빅데이터는 물리공간, 네트워크 공간 등에서 생성되는 광범위한 데이터를 기반으로 개인, 기업, 공공분야와 같은 사회 전반에 걸쳐 다양하게 응용할 수 있다. 특히 빅데이터가 사회적 이슈로 부각되면서 공간현상에 대한 효율적인 진단, 처방 등이 가능한 공간정보 기술과 융합하여 활용할 수 있는 공간 빅데이터²⁾가 새로운 이슈로 대두되고 있다. 데이터의 약 80%가 위치 또는 공간과 관련되어 있듯이 빅데이터의 80%는 공간정보와 결합하여 지도로 변환할 수 있다(국토연구원, 2013). 따라서 공간 빅데이터의 활용을 통해 다양한 미래 트렌드와 패턴에 대한 공간적인 맥락을 이해할 수 있다. 급변하는 지역여건에 신속하게 대응하기 위해서는 공간 빅데이터 중심의 새로운 분석기법이 요구된다. 하지만 기존 연구에서 빅데

2) 공간 빅데이터란 데이터의 양, 속도, 다양성 때문에 합리적인 노력으로 관리, 분석, 활용 등을 위해 현재 컴퓨터의 능력을 초과하는 공간데이터 세트를 의미한다(Shashi, 2012).

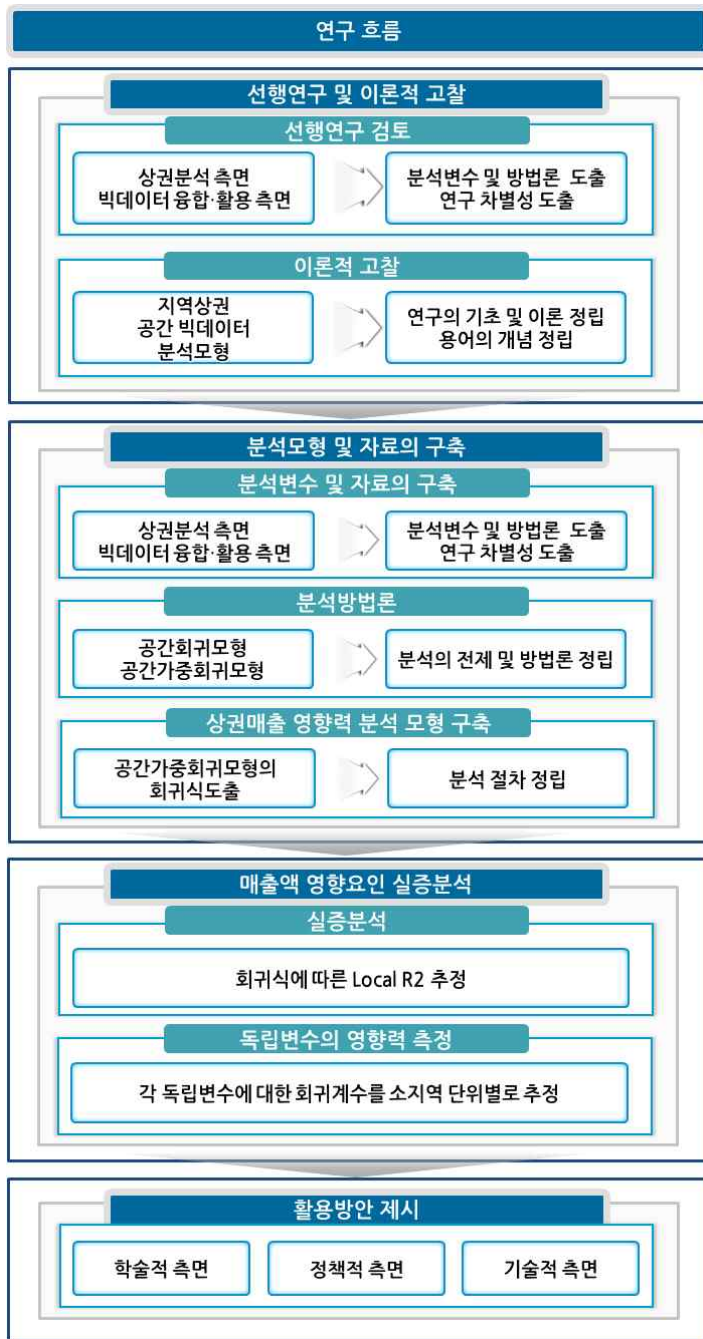
이터를 기반으로 지역 활성화 부문에 적용한 연구는 미흡한 실정이며 아직까지 그 파급력에 비해 관련 연구가 상당부분 미지의 영역으로 존재한다. 따라서 지역상권에 관한 여건과 공간적 요소들을 고려하여 기존에는 알 수 없었던 미래 변화를 신속·정확하게 예측하기 위해서는 공공, 민간에서 생성되는 정형, 비정형, 반정형 공간 빅데이터를 활용·분석할 수 있는 분석기반이 필요하다.

이러한 맥락에서 본 연구는 공공, 민간에서 구축·생성되는 카드매출정보, 유동인구 등 공간 빅데이터를 활용하여 지역상권에 관한 인구, 고용, 소득 등의 매출요인을 도출하고 이를 공간적으로 연계하여 소지역 상권매출에 관한 영향력 분석모형을 설계·분석하고 그 결과물을 통해 다양한 활용방안을 제시하는데 목적이 있다.

제2절 연구내용 및 방법

본 연구는 선행연구 및 이론적 고찰, 분석모형 및 자료의 구축, 지역상권의 매출영향력 분석 등 크게 4가지 부문으로 구분하여 연구를 수행하였다.

먼저 선행연구는 ‘상권 분석’ 과 ‘빅데이터 융합·활용’ 에 관한 두 가지 측면으로 구분하여 선행연구를 검토하였다. 상권 분석 측면에서는 GIS(Geographic Information System) 공간분석 및 상권분석에 선행된 연구를 검토하였으며 지역상권의 매출 영향력 분석과 관련한 변수와 분석기법 등을 중점적으로 검토하였다. 빅데이터 융합 활용 측면에서는 다양한 부문에서 생성되는 빅데이터의 특성과 이를 활용하여 공간정보와의 융합·분석에 필요한 각종 기법을 검토하였다. 선행연구의 검토를 종합하여 시사점을 도출하고 본 연구에서의 차별성을 제시하였다. 다음으로 이론적 고찰에서는 지역상권, 빅데이터, 공간 빅데이터, 공간통계분석, 공간가중회귀 분석 등 본 연구의 기초가 되는 이론적 배경과 용어의 개념 등에 대해 고찰하였다. 분석모형 및 자료의 구축에서는 지역상권의 매출 영향력 분석을 위한 매출영향요인 및 요인 간 상호관계 등을 파악하고 분석변수 및 자료 구축, 분석 방법론 검토, 본 연구의 분석모형 등을 설계한다. 이후 실증분석에 사용될 대상지 내 통계정보와 빅데이터 등을 검토, 수집하고 공간정보와 연계하여 데이터를 구축한다. 실증분석에서는 분석결과의 파급성, 효과성 등을 고려하여 분석의 공간적 범위를 서울특별시로 선정하였으며 구체적으로는 서울시 내에 지역 간 비교가 가능하도록 전체 행정구역을 분석범위로 하며 세부적인 지역의 매출영향력 도출을 위해 소지역 단위의 분석단위로 세분화하였다. 앞서의 실증자료를 공간가중회귀, 공간통계분석 등을 연계한 본 연구의 모형에 적용하여 분석하고 지역 간 매출 영향력을 해석하였다. 이를 토대로 다양한 부문의 활용방안을 제시하였다.



<그림 2> 연구 흐름도

제2장 선행연구 검토 및 이론적 고찰

제1절 선행연구 검토 및 차별성

(1) 선행연구 검토

공간 빅데이터 기반 지역상권의 매출 영향력 분석을 위해 ‘상권 분석’, ‘빅데이터 융합·활용’ 등 크게 두 가지 측면으로 구분하여 관련 연구, 보고서, 유사 분석 사례 및 서비스, 시스템 등을 검토하였다.

(가) 상권 분석 측면

상권 분석은 전통적으로 많은 연구가 이루어진 분야이다. 상권분석에 관한 연구분야에서는 GIS와 통계적 방법론에 대한 연구에 힘입어 더욱더 발전하여 보다 효과적이고 실용적인 방법론이 제시되고 있다.

GIS 공간분석기법을 적용한 상권분석 연구에서 정대석 외(2015)는 상권의 업종별 분포와 매출에 미치는 영향요인을 경기도 31개 시군을 대상으로 분석하였다. 이를 위해 음식, 소매, 생활서비스, 관광 및 여가, 스포츠 등 5개 업종에 대한 회귀분석 모형을 구축하였으며 종속변수는 해당업종의 매출액, 독립변수는 공동주택평균기준시가, 업종별 시설 수, 주요시설 수, 집객시설 수 등을 활용하였다. 회귀분석 결과 음식, 생활서비스, 스포츠 업종의 경우 공동주택기준시가 변수가 매출에 가장 큰 영향력을 나타내었으며 해당 업종별 업소 수 변수는 음식 및 생활 서비스 업종에서 양의 요인으로 나타났다.

이경주 외(2015)는 강원도지역의 지역상권 현황진단과 상권활성화 및 발전방안 마련을 위해 상권 활성화에 큰 영향을 주는 매출액을 기반으로 공간가중회귀모형을 적용하여 지역상권의 현황을 진단하고 활성화를 위한 정책적 시사점을 제시하였다. 분석을 위해 먼저 매출에 영향을 주는 수요·공급 요인 등을 선정하였으며 이를 공간가중회귀모형에 적용하여 분석하였다. 분석결과 수요·공급 요인으로 도출된 각 독립변수의 영향력이 지역별로 뚜렷한 편차를 보이고 있음을 확인할 수 있었다.

정은애 외(2015)는 의류 소매업에서 유발되는 매출액에 대한 영향요인에 공간적인 요소를 고려하여 체계적인 접근을 시도하였다. 서울특별시를 중심으로 의류 소매업 매출에 영향을 주는 요인들을 인구특성, 토지이용특성, 접근성특성, 입지특성으로 구분하여 파악하였다. 각 특성에 따라 최소제곱법모형, 일반최소제곱법모형, 공간계량모형 등을 구축하였으며 이들 모형 간의 비교·분석을 실시하였다. 이러한 분석결과는 상권에 대한 매출영향요인분석에서 기존에 통계요소를 활용한 한계를 극복하고 공간요소를 고려하여 체계적으로 분석한데 큰 의의를 갖는다.

박완서 외(2012)는 근대기 서울의 공간구조와 상권의 변화에 대해 심층적으로 고찰하였다. 먼저 근대 서울의 공간구조를 분석하기 위해 공간구문론(Space Syntax)에 기반한 DeptMap³⁾를 이용하여 당시 서울의 도로 접근성을 분석하였다. 또한 이를 활용하여 접근성 수치와 상점의 분포수와의 상관관계를 회귀분석을 통해 밝히고 기존의 유통 구조적인 변화, 상품구매 관습 등 사회·경제적인 측면뿐만 아니라 공간구조의 변화에도 영향을 받는다는 사실을 알아냈다.

3) DepthMap은 UCL의 VR Centre에서 제공하는 공간네트워크분석 소프트웨어로써 공간구문론의 방법론을 따라 정량적으로 공간구조를 분석할 수 있는 도구이다.

정대영 외(2009)는 지리적 공간상에서 공간객체 간의 상호의존성과 상호작용 및 통계적 상관분석을 이용하여 서비스업종간의 상관분석법을 제시하고자 하였으며 근린상권의 업종 간 상관관계를 통하여 공간특성을 분석하였다. 이를 통해 서비스업, 요식업 등의 분류별로 Moran I 계수와 국지적 상관지수인 Getis-Ord General G 지수를 산출하였다. 또한 이러한 지수 공간객체에 입력하여 지도상에 핫스팟으로 표현하였다. 분석 결과 교육관련 서비스업, 숙박관련서비스업, 오락관련서비스업, 양의 상관관계를 보이는 업종과 음의 상관관계를 보이는 업종을 도출하였다.

(나) 빅데이터 융합·활용 측면

국내에서 빅데이터 융합·활용에 관한 연구는 아직 초기단계로써 다양한 분야의 데이터를 융합, 분석하여 새로운 인사이트(Insights)를 발견할 수 있는 연구가 증가하고 있다.

이승주(2015)는 기존 통계적 분석에 한계를 극복하고 새로운 접근방법의 모색을 위해 주성분 분석을 이용한 빅데이터분석 모형을 제시하였다. 기존 많은 분야에서 활용되는 다변량 통계분석기법인 주성분분석을 응용하여 빅데이터 분석을 위한 새로운 표본추출방법을 제안하였으며 통계적인 모의실험으로 그 성능을 평가하였다.

구자용(2015)은 공간 빅데이터의 공간적 분포를 통해 지역별로 발생하는 현상이나 특성 등을 파악하였다. 실증분석을 위해 서울시 SNS 데이터를 수집하였으며 공간정보와의 연계과정을 통해 지도 시각화를 수행하였다. 이후 지역별로 위치한 SNS 내용을 통해 분포특성을 파악하였으며 실제 장소에 대한 정보량을 비교하여 분석모형에 객관성을 확보하였다.

허자연(2014)은 SNS(Social Network Services)상의 정보량이 지역 방

문자 수에 양(+)의 영향을 미치는 것을 파악함으로써 SNS 게시물 규모와 실제 공간상에 상관관계를 밝혔다.

김관호 외(2013)는 지하철 이동패턴을 효과적으로 분석할 수 있는 기법 개발을 위해 서울시 스마트카드 빅데이터를 활용하여 특정 관측지점에서 다른 지점으로 이동을 설명하는 클러스터링 기반 패턴 분석기법을 제안하였다. 제안된 분석기법을 이용하여 서울시 내 실제 지하철 데이터를 비교·분석하여 분석기법의 실효성을 검증하였다.

전상권(2012)은 다양한 부문에서 생성되는 이기종 미디어의 빅데이터를 활용한 분석 플랫폼을 개념적으로 설계하여 실증분석을 통해 그 효과를 입증하였으며 이를 바탕으로 활용방안을 제시하였다.

상권 정보 및 소비자 정보 등의 빅데이터를 융합·분석한 서비스 사례로는 2006년 중소기업청에서 개발한 상권정보시스템이 있다. 이는 빅데이터 및 공간정보를 기반으로 최초 10여종의 데이터로 20개의 정보(메뉴)를 제공하였으며, 2015년 기준, 27종 데이터를 활용하여 49개의 분석 정보와 점포평가, 점포이력, 창업과밀지수 등을 서비스하고 있다. 상권정보시스템의 분석은 크게 업종분석, 매출분석, 인구분석, 지역분석을 제공하며 모두 지도상에 표출할 수 있는 통계추이, 현황 등에 지도 서비스를 제공하고 있다.



<그림 3> 상권정보시스템 홈페이지

또한 서울시에서는 소상공인 지원을 위한 우리마을가게상권분석 서비스를 제공하고 있다. 분석 서비스에는 점포, 소비패턴, 구매력, 집객시설, 유동인구, SNS 트렌드 등의 상권 관련 빅데이터를 가공하여 창업 위험 지표, 업종 관련 지표, 매출 트렌드, 경쟁현황, 고객/인구 현황, 배후지 특성 등 상권분석에 유용하게 활용할 수 있는 정보를 제공한다.



<그림 4> 우리마을가게 상권분석시스템 홈페이지

FNC 코오롱은 소비자 맞춤형 마케팅을 위하여 본 사의 고객 구매 데이터 및 코오롱 온라인 로그데이터 등 자체 빅데이터를 활용하여 고객중심의 새로운 전략을 수립했다. 마찬가지로 신한카드는 고객 맞춤형 서비스를 위해 2200만 고객의 정보를 모아 빅데이터 센터를 구축했다. 기존에 다수를 대상으로 하던 매스마케팅을 벗어나 개별적인 특성을 강조한 마케팅에 초점을 두어 추진하였다. 이를 적용한 하이포인트 카드, 빅플러스 카드 등 7종의 신용카드에서 신한러브카드, 에스초이스 카드 등 2종류의 체크카드에도 빅데이터 기능이 적용되어 카드 활용 폭이 획기적으로 증가하게 되었다. 또 현대백화점은 고객 맞춤형 매장의 물건 배치를

위해 고객의 특성, 상품정보 등을 수집하여 활용함으로써 보다 세밀한 고객 마케팅을 가능하게 했다. 이 밖에 공공측면으로 기상청에서는 기상 빅데이터 분석을 통해 위험기상을 예측하였으며, 문화체육관광부에서는 빅데이터 분석을 통해 2013년 국민의 인식과 트렌드 변화를 분석하여 관 정책수립에 활용하였다. 또 안전행정부는 스마트빅보드⁴⁾를 통해 각종 기상, 재난이력 및 국내외 재난정보의 정보 공유를 통해 재난현장 상황 정보를 취할 수 있었다. 스마트 빅보드에서 활용하는 빅데이터는 재난관리에 활용성이 높은 트위터 정보, 과거 재난이력 및 원인분석결과, 위성영상, 시뮬레이션, 관련 웹사이트 정보 등이 해당된다. 이 밖에 빅데이터를 활용한 국내 사례로는 11번가(SK플래닛)의 스마트 추천 알고리즘 시스템, 올레TV(KT)의 실시간 시청정보 및 콘텐츠 제공, 두산중공업의 발전소 고장예방 효율화 사례, 더존비즈온의 빅데이터 기반 회계 관리 시스템 등이 있다. 또한 국외 사례로는 노스페이스의 기계학습을 통한 지식 기반 상품 추천 서비스, 스타캐스트의 빅데이터 기반 MLB 중계 시스템, 오므론의 빅데이터 분석을 이용한 제품 생산성향상 등 다양한 시스템 및 서비스, 마케팅의 목적으로 활용되고 있다.

(2) 선행연구와의 차별성

선행연구를 검토한 결과 기존 전통적인 상권분석은 GIS 공간분석기법을 적용하여 그 정확도를 향상하는 연구들이 다양하게 진행되고 있음을 알 수 있었다. 또한 지역상권의 활성화 분석 및 매출분석 등 동태적 분석을 다루는 연구에서는 소비자 설문을 통하여 만족도 분석을 실시하거

4) 스마트빅보드는 재난 및 안전사고 시 현장 중심의 모든 가용한 정보 네트워크를 가동하여 위험상황을 분석하는 최첨단 스마트 재난 상황실을 말한다.

나 소비행태(재방문의사, 지출액 등)를 종속변수로 하는 회귀분석을 실시하여 매출의 영향요인을 밝히는 연구들이 진행되어 왔다. 그러나 기존 상권분석 및 활성화분석과 관련된 선행연구는 과학적 분석을 위해 공간 분석과 통계분석을 적용하고 있으나 수학적 모델에 의존적이다. 또한 자료구득의 한계에 따라 대리변수를 사용하거나 시계열자료의 구축 미흡, 공간에 대한 고려부족 등의 한계를 갖는다. 반면 빅데이터 연구에서는 실시간 웹 정보, 민간의 빅데이터, SNS 비정형정보 등을 활용할 뿐만 아니라 공간참조 데이터와 GIS 공간분석기법과 연계하여 실제 공간상에 미치는 상호작용을 입증하였으며 빅데이터 기반의 새로운 방법론을 적용함으로써 새로운 가치를 발견할 수 있는 가능성을 시사했다. 따라서, 과거 지역상권의 제한된 분석모형의 틀에서 벗어나 공간적인 요소와 함께 빅데이터 기반의 다양한 영향요인을 고려할 수 있는 공간 빅데이터 분석이 필요하다.

본 연구에서는 유동인구, 카드매출액 정보 등 공간 빅데이터를 활용하여 이에 영향을 미치는 요인들을 분석한다는 측면에서 타 연구와 근본적인 차이를 가진다. 또한 기존 행정경계 등 단위의 분석단위보다 규모가 작은 소지역 분석단위를 통해 보다 세분화된 지역을 분석함으로써 분석과정 및 결과의 일반화 및 활용도가 높다. 본 연구는 유동인구, 카드매출 등의 공간 빅데이터와 인구, 고용 등의 통계정보를 활용하여 지역상권 매출의 영향요인을 도출하고 공간 빅데이터 처리, 분석 모형을 구축·실증하여 서울특별시 소지역 단위의 매출영향력을 도출한다는 점에서 연구의 차별성이 있다.

<표 1> 주요 선행연구 검토 및 차별성 요약

구분		선행연구와의 비교	
		연구목적	연구내용
상 권 분 석	1	<ul style="list-style-type: none"> • 논문명 : 상권 업종별 분포 및 매출 영향요인 분석 - 경기도 31개 시군을 대상으로(정대석 외, 2015) • 연구목적 : 상권 업종별 매출영향요인 도출 및 상관관계 분석 	<ul style="list-style-type: none"> • 선행연구 검토를 통해 매출영향요인 선정 및 회귀분석 모형 구축 • 실증분석을 통한 업종 간 상관관계 측정
	2	<ul style="list-style-type: none"> • 논문명 : 공간가중회귀모형을 이용한 강원도지역상권 현황진단 및 발전방안 연구(이경주 외, 2015) • 연구목적 : 강원도 지역상권의 현황진단 및 발전방안 제시 	<ul style="list-style-type: none"> • 도소매업의 매출액 영향요인 분석 • 주요 요인별 공간가중회귀모형 구축 • 실증분석을 통해 지역별 매출영향력 측정 및 정책방향 제시
	3	<ul style="list-style-type: none"> • 논문명 : 공간자기상관성을 고려한 의류 소매업 매출액 영향요인 분석(정은애 외, 2015) • 연구목적 : 의류 소매업에 미치는 매출영향요인 실증분석 	<ul style="list-style-type: none"> • 의류 소매업 매출액 영향요인 분석 • 주요 요인별 회귀모형 구축 • 실증분석을 통한 주요 요인별 매출영향력 측정
빅 데 이 터 융 합 활 용	1	<ul style="list-style-type: none"> • 논문명 : 주성분 분석을 이용한 빅데이터 분석(이승주, 2015) • 연구목적 : 주성분 분석을 적용한 빅데이터 분석 모형 제시 	<ul style="list-style-type: none"> • 주성분 분석 적용방안 검토 • 주성분 분석 기반 표본추출 모형구축 • 실증분석 및 통계적 유의검증
	2	<ul style="list-style-type: none"> • 논문명 : 소셜네트워크와 상권 활성화에 관한 연구(허자연 외, 2014) • 연구목적 : SNS정보와 실제 상권의 상호관계를 실증분석 	<ul style="list-style-type: none"> • SNS의 비정형 빅데이터를 활용한 분석 사례 분석 및 이를 통한 모형구축 • 실증분석을 통해 SNS 게시물 수와 상권 방문자간의 유의성 도출
	3	<ul style="list-style-type: none"> • 논문명 : 스마트카드 빅데이터를 이용한 서울시 지하철 이동패턴 분석(김관호, 2013) • 연구목적 : 빅데이터 기반 지하철 이동패턴분석 모형제시 및 이동특성 분석 	<ul style="list-style-type: none"> • 이동패턴 분석에 관한 지표 분석 • 분석지표 및 방법론 선정 • 실증분석을 통한 지역별 이동패턴 분석
본 연 구		<ul style="list-style-type: none"> • 본 연구는 유동인구, 카드매출정보 등 공간 빅데이터와 인구, 고용 등을 활용하여 서울특별시 내 지역상권의 따라 발생하는 매출 영향력을 실증 분석함으로써 다양한 활용방안을 제시 	<ul style="list-style-type: none"> • 선행연구 및 유사분석 사례 등의 검토를 통한 변수 및 분석모형 설정 • 실증분석을 통한 매출요인 간 영향력을 도출하고 실제 매출자료와의 비교를 통해 검증 • 분석결과에 해석 및 시사점 도출, 다양한 분야의 활용방안 제시

제2절 이론적 고찰

(1) 공간 빅데이터의 개념

(가) 빅데이터

공간 빅데이터의 정의를 위해 먼저 빅데이터에 관한 정의가 필요하다. 빅데이터에 관한 선행연구를 검토한 결과, 빅데이터 정의는 아직까지 통일된 정의가 부재하며 여러 기관들에 따라 다양하게 정의하고 있다. 특히 근래 스마트기기의 보급 및 인터넷 확산에 따른 방대한 양의 데이터가 생성·축적되면서 빅데이터의 개념이 등장하였으며, 다양한 분야에서 여러 가지 관점으로 정의하고 있다.

<표 2> 빅데이터의 정의

구분	정의
IDC(2011)	• 다양한 데이터로 구성된 방대한 데이터로부터 고속 캡처, 데이터 탐색 및 분석을 통해 경제적으로 필요한 가치를 추출할 수 있도록 디자인된 차세대 기술 및 아키텍처
가트너(2012)	• 대량의 데이터가 실시간으로 끊임없이 다양한 형태로 들어오는 것
맥켄지(2013)	• 전통적인 데이터베이스 소프트웨어로는 수입, 저장, 관리, 분석이 어려운 정도의 큰 규모의 데이터
삼성경제연구소 (2012)	• 기존의 관리 및 분석 체계로는 감당할 수 없을 정도의 거대한 데이터의 집합
(구)국가정보화 전략위원회(2011)	• 대용량 데이터를 활용·분석하여 가치있는 정보를 추출하고, 생성된 지식을 바탕으로 능동적으로 대응하거나 변화를 예측하기 위한 정보화 기술

해외에서는 IDC, 가트너 등 여러 기관에서 빅데이터를 기존의 시스템이 수집, 저장, 분석할 수 있는 특정 규모 이상의 방대한 데이터, 또는 이를 활용한 미래 핵심 가치를 추출할 수 있는 새로운 기술 등으로 정의하였다. 국내의 경

우 삼성경제연구소 및 국가정보화전략위원회에서 정의하였는데, 마찬가지로 데이터 양적측면이나 이를 처리하기 위한 기술 등으로 정의하였다.

(나) 공간 빅데이터

공간 빅데이터의 개념 또한 빅데이터와 마찬가지로 명확한 정의가 부재하며 일부 대학 및 학자에 의해 정의되고 있다. 선행연구에서는 빅데이터의 특성을 반영하여 공간 빅데이터를 개념적으로 정의하거나 공간정보의 특성을 포함하는 개념으로 정의하고 있다.

<표 3> 공간 빅데이터의 정의

구분	정의
Shashi(2011)	• 데이터의 양, 속도, 다양성 때문에 합리적인 노력으로 관리, 분석, 활용 등을 하기 위해 현재 컴퓨터의 능력을 초과하는 공간데이터세트
김미정(2013)	• 공공·민간에서 생성되는 빅데이터와 공간정보를 융·복합한 데이터를 말하거나 세상에 존재하는 빅데이터를 공간 형태로 표현, 또는 가능한 모든 자료
안종욱(2013)	• 공간정보의 특성과 빅데이터의 특성을 고려하여 데이터의 관리, 처리, 분석에 있어 데이터의 양, 생성속도, 형태의 다양성이 현재 시스템으로 수용이 어려운 공간데이터

먼저 Shashi(2012)는 데이터의 양, 속도, 다양성 때문에 합리적인 노력으로 관리, 분석, 활용 등을 하기 위해 현재 컴퓨터의 능력을 초과하는 공간데이터세트로 정의하였다. 김미정(2013)은 공공·민간에서 생성되는 빅데이터와 공간정보를 융·복합한 데이터를 말하거나 세상에 존재하는 빅데이터를 공간 형태로 표현, 또는 가능한 모든 자료로 정의하였다. 또한 안종욱 외(2013)의 연구에서는 공간 빅데이터는 공간정보의 특성과 빅데이터의 특성을 고려하여 데이터의 관리, 처리, 분석에 있어 데이터의 양(Volume), 데이터 생성속도(Velocity), 형태의 다양성(Variety)이 현재 시스템으로 수용이

어려운 공간데이터라 정의하였다.

선행연구 검토를 통해 살펴본 빅데이터 및 공간 빅데이터의 개념은 초기 단순 대용량 데이터 집합에서 점차 관리, 분석, 활용을 위한 새로운 기술 체계로 변화하고 있다. 즉, 공간 빅데이터의 개념은 다양한 형태의 모든 빅데이터와 공간정보를 융복합한 데이터 집합이나 이러한 공간 빅데이터를 관리, 처리, 분석, 활용하는 동시에 미래 새로운 가치를 추출할 수 있는 기술체계의 확장된 의미로 해석할 수 있다. 본 연구에서는 공공에서 제공하는 인구, 고용 등의 통계정보와 민간에서 생성되는 유동인구, 스마트카드 데이터 등의 빅데이터를 기반으로 공간정보와 융합·활용하여 기존에 알 수 없었던 세분화된 분석을 수행하는데 초점을 둔다.

(2) 상권의 개념

(가) 상권의 정의 및 유형

상권(Trading area, Market area)이란 점포와 고객을 흡인하는 지리적 영역이며, 모든 소비자의 공간선호(Space Preference)의 범위를 의미하기도 한다. 따라서 상권은 판매액의 비율을 고려하여 생각할 수 있는데, 대표적으로 상품 판매액의 약 70%를 차지하는 지역을 1차 상권, 다음 25%를 차지하는 지역을 2차 상권, 그 나머지를 3차 상권이라 말한다. 또한 상권은 의존형태에 따라 주거지형 상권, 도심지형 상권, 복합형 상권으로 정의되고 있다. 이러한 상권의 변화와 이동원인으로는 신규 지하철 노선개설, 버스터미널의 형성 및 이전, 버스정류장 이전 등 교통수단에서의 변화가 있으며 대형 할인점 및 백화점 개발 등 대형업체의 등장에 민감하게 반응한다. 또 공공 및 민간기관들의 이전, 대기업의 입지, 이전이

상권의 변화와 이동에 영향을 주고 배후지역의 소득수준 변화가 큰 요인으로 작용한다. 상권의 발전단계별 특성은 성장기, 성숙기, 쇠퇴기 등 크게 3가지로 구분된다. 먼저 성장기 지역의 상권은 상권발전에 영향을 주는 요소가 상승하는 상권으로 지가가 상승중이며, 토지투기현상, 젊고 높은 교육수준의 주민 층이 분포하는 특징이 있다. 성숙기 지역의 상권은 상권발전 요인들이 충분히 성숙한 상권으로 지역개발이 완료·활성화 되었거나 부동산의 가격, 기능 등이 최고조를 보이며 배후지역의 소득수준이 상대적으로 매우 높은 특징을 보인다. 쇠퇴기 상권의 경우 각종 상권 기능이 점차 쇠퇴하는 상권으로 건물들이 점차 노후하기 시작하여 지가가 낮아지기 시작하고 주민 층도 노령화 추세를 보이는 특징이 있다.

<표 4> 상권의 유형

구분	정의
1차 상권	<ul style="list-style-type: none"> • 점포 매출액의 60~70%를 포함하는 범위 • 점포에서 반경 약 500미터 이내의 지점
2차 상권	<ul style="list-style-type: none"> • 점포 매출액의 15~25%를 포함하는 범위 • 점포에서 반경 약 1Km 이내의 지점
3차 상권	<ul style="list-style-type: none"> • 1, 2차 상권 이외의 매출지역을 포함하는 범위
주거지형 상권	<ul style="list-style-type: none"> • 지역 밀착 형상 업 지역으로 상주인구를 중심으로 고객을 흡수하는 상권
도심지형 상권	<ul style="list-style-type: none"> • 도시중심상권으로써 유동인구 내 포함되어 있는 고객을 흡수하는 상권
복합형 상권	<ul style="list-style-type: none"> • 인근지역 상주인구와 유동인구를 동시에 흡수하는 상권

출처 : 한국프랜차이즈협회(2004), 프랜차이즈 경영가이드 총서

(3) 상권의 매출액 영향요인

(가) 상권결정요인

지역상권의 매출 영향력 분석을 위해 먼저 상권의 결정요인 살펴보고 이후 서울특별시에 적용가능한 지역상권의 매출 영향요인을 도출하고자 한다.

Mejia & Benjamin(2002)은 점포의 결정요인을 크게 공간적 요인과 비공간적 요인으로 구분하였다. 공간적 요인에는 상권, 부지, 건물이 요인으로 포함되며 비공간적 요인에는 점포이미지와 상품 구성이 요인으로 포함된다.

<표 5> 매출액 결정요인

구분	결정요인
공간적 요인 (Spatial Factors)	상권(Market)
	부지(Site)
	건물(Building)
비공간적 요인 (Non-Spatial Factors)	점포이미지(Retail Image)
	상품 구성(Retail Mix)

출처 : Mejia & Benjamin(2002)

이러한 요인은 본 연구에서 특정한 점포를 대상으로 하지 않고 상권 전체의 결정을 대상으로 하고 있으므로 상점의 부지, 건물, 점포이미지, 상품 구성과 같은 요인은 본 연구에서 제외한다. 또한, 국내의 상권분석 연구에서는 상권의 결정요인 중 공간적 요인의 영향력이 비공간적 요인에 비하여 매우 큰 것으로 나타났다(김성문, 2014).

<표 6> 요인별 설명력

구분	설명력
공간적 요인	0.815
비공간적 요인	0.033

출처 : 김성문(2014), 소매업태의 매출액 결정요인에 관한 연구

따라서 상품구성과 같은 비공간적 요인을 연구에서 제외하는 것은 연구목적 달성에 큰 영향을 미치지 않을 것으로 판단된다.

매출 영향력 분석 중에서 본 연구의 주요 관심이 되는 것은 상권결정요인이다. 상권결정요인은 소비자와 관련된 수요요인과 판매자와 관련된 공급요인으로 나눌 수 있는데 수요요인은 인구, 소득수준, 인구통계적 특성을 세부요인으로 포함하며, 공급요인은 경쟁, 집적, 거리를 세부요인으로 포함한다.

<표 7> 상권결정요인

구분		설명
수요 요인	인구	<ul style="list-style-type: none"> • 상권의 잠재구매력을 측정하는 기본적인 필수적인 지표 • 인구수는 점포결정과 매출액에 양의 영향을 미치는 것으로 알려짐
	소득수준	<ul style="list-style-type: none"> • 지역의 경제적 수준을 나타내는 변수 • 소매점 매출액을 결정하는 필수 지표
	인구통계적 특성	<ul style="list-style-type: none"> • 교육수준, 직업, 문화적 성향 등의 인구통계적 특성 • 소비자의 기호를 설명하는 변수로 활용가능
공급 요인	경쟁	<ul style="list-style-type: none"> • 대형점포가 상권범위 내에 존재
	집적	<ul style="list-style-type: none"> • 상점들이 지리적으로 근접하게 입지하는 것 • 집적효과에 따라 상권을 결정하는데 긍정적 영향을 줌
	거리	<ul style="list-style-type: none"> • 상점과 소비자 간 물리적 거리는 소비자의 이동비용에 영향 • 매출액은 소비자와 상점 간 거리에 영향을 받음

(나) 수요요인

상권의 결정에 있어 소비자와 관련된 수요요인으로서는 크게 소득, 인구, 인구통계적 특성이라는 세 가지 요인이 있다(김성문, 2014). 지역의 배후인구는 지역상권을 결정하는 가장 중요한 요인이라는 것이 일반적인 통념이다.

배후지역의 인구 및 세대수를 주요변수로 하여 점포결정 및 매출 등의 연구사례로 Ingene & Yu(1981)은 미국의 SMSA(Standard Metropolitan Statistical Area)를 분석단위로 하여 소매업의 점포결정요인, 매출액 결정요인을 분석하였는데 SMSA의 총인구가 지역의 점포결정과 매출액에 양(+)의 영향을 미치는 가장 유의한 변수로 나타났다. 김도현 외(2014)이 수행한 대형할인점의 단위면적효율을 종속변수로 한 분석에서는 지역 인구수가 유의미한 양(+)의 영향을 미치는 것으로 나타났다. 이상민(2013)이 수행한 나들가게⁵⁾를 대상으로 한 연구에서 역시 배후상권의 세대수가 매출과 점포결정에서 정(+)의 영향을 미치는 것으로 나타났다.

지역의 소득수준은 지역의 구매력을 나타내는 중요한 변수로 지역상권의 매출액에 영향을 미치는 요인으로 널리 알려져 있다. 이에 대한 선행연구로 Ferber(1959)는 소매점이 입지한 지역의 소득수준이 높을수록 점포의 수요가 높아지는 것을 분석하였다. 김성문 외(2014)는 백화점의 결정요인을 분석하였는데, 지역의 소득수준을 대표하는 변수인 세대별 지방세 징수액과 공시지가가 백화점의 결정에 유의미한 양(+)의 영향을 나타내는 것으로 나타났다.

연령구성, 성별구성, 교육수준, 문화적 특성 등의 인구 통계적 특성은 소비자의 기호에 영향을 미치고 이에 따라 지역상권의 결정에 영향을 미치는 중요한 요소로 여겨진다. Liu(1970)는 대졸 이상의 고학력의 소비자 비율이 높을수록 소매점의 매출액과 점포결정이 비율이 높아지는 것을 분석하였다. 정

5) 나들가게란 중소기업청이 추진해 온 스마트샵 육성지원사업의 명칭이다.

은애(2015)는 서울시 지역상권에서 의류 소매업의 영향을 미치는 요인을 분석하였다. 분석결과 의류 소매업이라는 특징에 맞게 인구 통계적 특성 중 하나인 여성인구수와 직장인수가 유의미한 양(+)의 영향을 나타내었다.

(다) 공급요인

상권 내에 유사한 점포가 입지하게 되면 경쟁관계가 발생하며, 유사한 점포가 동종업태와 이종업태인 경우 모두 경쟁관계가 발생한다. 경쟁관계와 매출액 사이의 관계는 경쟁이 선행하여 매출액에 영향을 미친다는 연구와 지역의 매출액 수준이 새로운 경쟁자의 출현을 유도한다는 연구관점이 서로 대립하고 있다(김성문, 2014). 그러나 상권분석에 관한 많은 연구에서는 대형점포와의 경쟁을 주로 고려하고 있다. 신우진 외(2010)는 SSM⁶⁾(Super Super Market)의 입지로 인한 기존 슈퍼마켓의 상권잠식의 정도를 분석하였는데 SSM 입점으로 인한 상권잠식의 정도가 큰 것으로 나타났다. 문소연(2010)은 전통시장이 아닌 슈퍼마켓을 연구의 대상으로 하여 SSM과의 경쟁으로 인한 소비와 매출액 감소를 분석하였다. 이 연구에서는 허프 모형(Huff model)을 활용하여 SSM의 입점으로 인하여 감소하는 슈퍼마켓의 점포결정 및 매출액을 추정하였다.

집적은 다수의 업체들이 지역 내에 지리적으로 근접하여 입지하는 것으로 이를 통하여 소비자들에 대한 집객효과를 기대할 수 있다. 이는 경쟁과는 반대의 효과로 지역 내에 소매점의 개수가 많을수록 긍정적인 효과가 나타날 수 있음을 의미한다. 국내외 많은 연구들은 집객효과에 있어서 소매점의 집적이 긍정적인 영향을 미친다는 분석결과를 도출하고 있다. 지리적으로 인접한 지역에 집적되어 있는 상점군은 고객의 다양한 수요욕구를 충족시켜 줄 가능성이 높기 때문에 전체 상점군의 매출에 긍정적인 요인을 보이는 것으로 판단된다(김성문, 2014).

소비자의 거주지와 상점 간 지리적 거리는 소비자의 이동비용을 발생시키기 때문에 거리가 멀수록 소비자의 이용은 감소된다. 상점과 소비자 사이의

6) Super Super Market은 ‘기업형 슈퍼마켓’으로 불리는 것으로, 대형마트보다 작고 일반 동네 슈퍼마켓보다 큰 유통매장을 지칭한다.

거리는 해당 상점의 선택하는데 주요한 결정요인이다. 이와 같은 가정은 Reilly의 소매 인력의 법칙⁷⁾의 근간을 이루고 있으며 이를 발전시킨 허프 모형에서 역시 중요한 핵심전제로 널리 알려져 있다. 허프 모형은 정교한 상권분석에 많이 활용되고 있다.

(4) 분석이론

(가) 통계적 공간분석

일반적 의미에서 공간분석이란 공간 자료를 분석 목적에 따라 각기 다른 형태로 가공하여 이로부터 추가적인 의미를 추출 할 수 있는 기술이라고 정의 할 수 있다(Baliley, 1994). 이는 자료의 관리, 지도학적 모델링, 네트워크 분석 등과 같은 GIS의 일반적인 분석 기능들을 모두 포함하는 광의적 의미를 담고 있다. 반면에 보다 학문적인 의미에서의 공간 분석은 자료에서 발견된 패턴과 관계의 확률론적인 속성에 관한 연구로서의 통계적 공간분석을 의미한다.

GIS의 통계적 공간분석은 다음의 네 가지 과정으로 구분할 수 있다. 데이터베이스로부터 분석 대상을 추출하거나 분석의 공간적 스케일을 설정하고 공간 질의를 수행하는 “자료의 선택” 과정과, 단위 지역별 집계 및 분할, 또는 위상관계의 생성 등과 같은 일련의 작업이 수행되는 “자료의 처리” 과정은 GIS의 통계적 공간분석을 수행하기 위한 가장 기본적인 기능에 해당한다. 또한 GIS가 주로 대용량의 자료를 다루며 분석가의 수

7) 소매 인력의 법칙이란 W. J Reilly가 주장한 법칙으로 두 개의 상권 사이에 존재하고 있는 소비자에 대한 영향력을 인력관계로 설명한 이론이다. 두 도시간의 고객흡입인력은 두 도시의 인구규모에 비례하고 두 도시의 분기점으로부터의 거리의 제곱에 반비례한다는 특징이 있다.(매일경제용어사전).

준에 관계없이 제공되어야 한다는 점을 고려할 때, 자료의 패턴 탐색이나 시·공간적 군집 여부의 조사, 공간적 상관관계 등을 분석하는 “탐색적 공간자료 분석(ESDA)⁸⁾” 과정 및 이를 통해 습득된 사전 정보를 바탕으로 연구 지역에 대한 예측모형을 정의·추정·평가할 수 있는 “확정적 공간자료 분석(CSDA)⁹⁾” 과정의 중요성은 점차 증대되고 있다(Anselin et al., 1992).

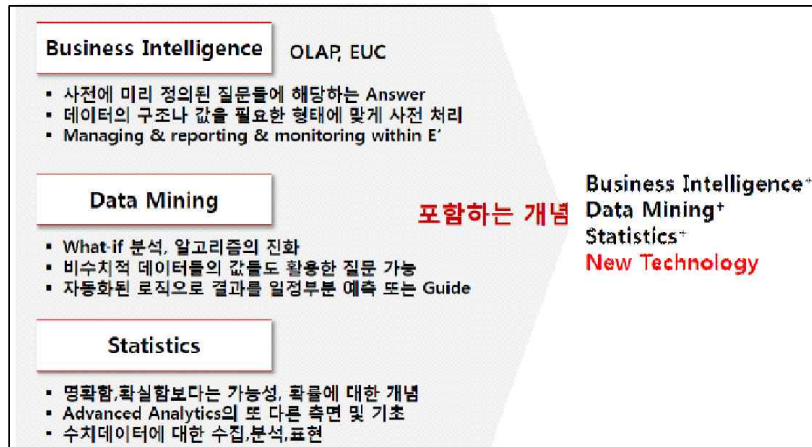
그러나 현재 대부분의 GIS는 자료의 선택과 자료의 처리에 해당하는 분석 기능만을 제공하고 있으나, 입체적 분석이 가능한 GIS 관련 상용소프트웨어의 새로운 모듈들이 추가되어 있어 응용 및 활용에 있어서 끊임없이 개선되고 있다. ESDA나 CSDA의 경우는 지리학, 지질학, 생태학, 의학, 경제학, 역학 등과 같은 학문을 중심으로 이론적인 전개와 발전을 지속적으로 보완하고 있는 추세이다.

(나) 빅데이터 분석

빅데이터 분석은 기존의 BI(Business Intelligence)분석, 데이터마이닝, 통계분석 등을 포함하는 분석기법으로 데이터에 내포하고 있는 속성정보들을 통해 미래를 예측한다. 빅데이터 분석의 유형으로는 예측 분석, 콘텐츠 분석, 실시간 분석, 고급 분석, 감성 분석, 소셜 분석 등으로 구분할 수 있다. 이러한 분석기법들은 상호 조합, 결합하고 적합한 시스템 및 인프라를 통해 설명, 예측, 의사결정을 위한 측정도구로 활용될 수 있다.

8) 탐색적 공간 데이터 분석(Exploratory Spatial Data Analysis, ESDA)은 불규칙적인 위치나 공간적인 특정지역의 공간패턴을 발견하기 위하여 다양한 기술적, 그래픽적 통계방법을 이용하여 공간정보를 효율적으로 제시하고 요약하는 방법이다(윤은혜 외, 1998).

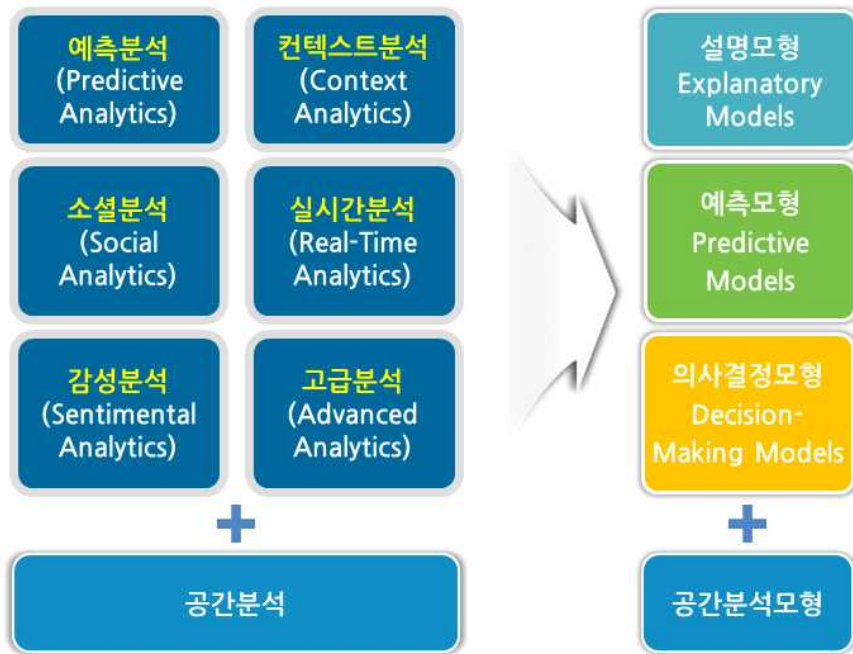
9) 확정적 공간 데이터 분석(Confirmatory Spatial Data Analysis)은 탐색적 공간 데이터 분석 과정 및 이를 통해 습득된 사전정보를 바탕으로 연구 지역에 대한 예측모형을 정의·추정·평가할 수 있는 데이터 분석방법이다(윤은혜 외, 1998).



<그림 5> 빅데이터 분석의 유형(출처 : 임동진, 2013)

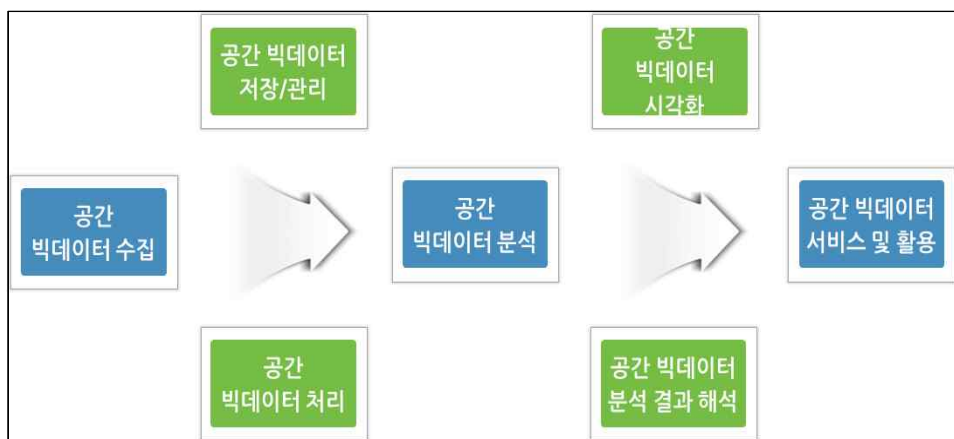
(다) 공간 빅데이터 분석

공간 빅데이터 분석은 앞서 검토된 개념에 따라 공간정보 특성을 반영한 빅데이터 분석이라 해석할 수 있다. 빅데이터 분석유형이 예측분석, 컨텍스트분석, 소셜분석, 감성분석, 실시간분석, 감성분석, 고급분석으로 구분된다면 공간 빅데이터 분석은 공간현상을 분석하는 GIS분석기법을 포괄하는 분석범위로 설명할 수 있다. 즉, 각종 공간분석 기법 및 모형의 개념과 빅데이터 분석기법이 결합되어 이를 적정 시각화 방법에 의해 표출함으로써 공간적 상호작용을 탐색하고 측정하는 의미로 해석 할 수 있다.



<그림 6> 공간 빅데이터 분석기법 및 모형 유형

공간 빅데이터 분석은 크게 수집, 저장/관리, 처리, 분석, 시각화, 분석결과 해석, 서비스 및 활용으로 구분된다.



<그림 7> 공간 빅데이터 분석 프로세스

공간 빅데이터 수집의 경우 정형 데이터 및 비정형, 반정형 데이터를 포함하며 단순한 데이터수집·저장뿐만 아니라 필요한 정보를 수집 정보를 분석에 활용하기 위한 데이터 형식의 변환과 가공 과정이 필요하다. 특히, 공간 빅데이터는 위치 및 공간요소를 포괄하는 개념으로 데이터 내부의 속성정보를 적절하게 활용해야 하므로 이를 위해 데이터 정제과정을 거친다. 이렇게 공간 빅데이터의 데이터 유형과 수집기술 등을 정리하면 다음과 같다.

<표 8> 공간 빅데이터의 데이터 유형, 종류 및 수집기술

데이터 유형	데이터 종류	수집기술
정형 데이터	RDB, 스프레드 시트	ETL, FTP, Open API
반정형 데이터	HTML, XML, JSON, 웹문서, 웹로그, 센서 데이터	Crawling, RSS, Open API, FTP
비정형 데이터	소셜 데이터, 문서(워드, 한글), 이미지, 오디오, 비디오, IoT	Crawling, RSS, Open API, Streaming, FTP
공간 데이터	행정경계, 건물, 도로망, 항공/위성영상, 등고선/DEM	Open API, Download 등

공간 빅데이터 분석의 경우, 앞서 살펴본 바와 같이 예측분석, 컨텍스트 분석, 소셜분석, 감성분석, 실시간분석, 감성분석, 고급분석등 통계학과 기계학습, 데이터마이닝 분야에서 사용되는 분석기법들을 확장하여 적용하고 있다. 최근에는 소셜미디어 등 비정형 데이터를 분석하는 소셜 분석, 감성분석, 군집분석 등이 주목받고 있다. 공간 빅데이터 분석 기법은 아직까지 그 유형의 구분이 명확하지 않으나 대표적인 분석기법으로 공간정보 기반의 실시간 분석, 통계분석, 데이터마이닝, 예측분석, 최적화, 평판분석, 소셜 분석 등이 있다.

공간 빅데이터 분석 가시화의 경우 빅데이터 방대한 자료속에서 의미있는 정보와 가치를 추출하여 공간정보 표준으로 가공하여 사람들이 직관적으로 알 수 있도록 지도 등의 매개체를 통해 표현하는 기술로 이해할 수 있다. 즉 일반적으로 활용하는 그래프, 차트 등과 네트워크 및 지도시각화가 포함 될 수 있으며 최근에는 지오인포그래픽이라는 새로운 시각화 기술이 적용되고 있다.

제3장 분석모형 및 자료의 구축

제1절 분석변수 및 자료의 구축

(1) 분석변수 및 데이터 선정을 위한 기준요소

본 연구의 상권매출 분석을 위해 먼저 주요 선행연구 및 이론적 배경 검토를 통해 수요요인, 공급요인 등 상권결정요인 검토하였다. 도출된 상권결정요인을 토대로 다음 표와 같이 대표성, 방향성, 이론적 근거, 자료구득가능성의 충족 여부를 검토하여 최종변수를 선정하였다.

<표 9> 변수 및 데이터 선정을 위한 기준요소

기준 요소	세부 내용
대표성	상권 매출에 대한 영향력을 판단하기 위해 특정 변수가 중복되어 선정되지 않도록 포괄적으로 판단할 수 있는 대표성을 지닌 변수를 선정
방향성	변수별 매출요인에서 나타나는 패턴 등이 일반화된 결과를 나타낼 수 있는지 여부를 나타냄
이론적 근거	기존의 매출요인과 관련한 선행연구에서 이론적 근거가 존재하며 중요도가 높다고 언급되는 변수를 선정
자료의 구득가능성	공간참조데이터가 존재해야하며 연계·분석이 가능한 공간단위 변수이고 실제 구축되어 구득 가능한 데이터이어야 함

출처 : 유선철 외(2015), 기후변화에 대응한 도시재생 지표개발 및 활성화 지역 선정연구 재구성

이러한 기준요소에 따라 본 연구에서 선정된 주요변수는 다음 표와 같다. 앞서 2장에서의 선행연구 및 이론적 고찰과 기준요소를 종합적으로 검토한 결과, 지역상권 매출액에 영향을 주는 주요 변수를 선정하였다.

<표 10> 매출에 미치는 영향요인 변수

구분		변수		자료의 구축방법 및 출처
독립 변수	수 요 요 인	인구	주거인구수	<ul style="list-style-type: none"> 아파트, 빌라 등의 거주인구 밀집시설물의 정보와 주택으로 추정되는 건물정보를 취합하여 거주인구 수 100m격자 중심점 데이터를 산정(2011년 기준) 출처 : Biz-GIS_XsDB 서비스
			유동인구 (빅데이터)	<ul style="list-style-type: none"> SPSS의 케이스 선택 툴을 통한 소지역 단위의 데이터 합계(2014년 1월~2015년 12월 기준) Arc GIS의 공간연산을 활용한 월평균 유동인구 수를 산정하고 소지역 중심점(Central Point) 데이터 생성 출처 : 공간 빅데이터 연구단(국토 공간정보의 관리, 분석 서비스 플랫폼 기술 개발) 내부자료
		고용	직장 인구수	<ul style="list-style-type: none"> 거주인구 밀집시설물의 정보와 통계청 집계구별 사업체정보를 통해 직장인구 수를 산정 출처 : Biz-GIS_XsDB 서비스
		소득 수준	지역의 평균소득	<ul style="list-style-type: none"> 공동주택(아파트, 빌라)과 오피스텔의 공시지가, 표준지공시지가를 이용한 토지가격 및 고 전-월세 비율 등을 기준으로 소득분위 100m 격자 중심점 데이터를 산정(2011년 기준) 출처 : 소득분위추정 10분위-Biz-GIS_XsDB, 소득분위별 평균소득-통계청, 전국경제인연합회(2015)
	공 급 요 인	집적	지역상권 상점의 수	<ul style="list-style-type: none"> SPSS의 케이스 선택 툴을 통해 본 연구의 분석대상지인 서울특별시 내 상권에 대한 POI를 추출하여 Arc GIS에 점 데이터 표출 출처 : 중소기업청 상권정보시스템
		거리	모형에 포함	-
종속 변수	매출		카드매출 (빅데이터)	<ul style="list-style-type: none"> SPSS의 케이스 선택 툴을 통한 소지역 단위의 데이터 합계(2014년 1월~2015년 12월 기준) Arc GIS의 공간연산을 활용한 월평균 카드매출액을 산정하고 소지역 중심점(Central Point) 데이터 생성 출처 : 공간 빅데이터 연구단(국토 공간정보의 관리, 분석 서비스 플랫폼 기술 개발) 내부자료

(2) 데이터의 구축

(가) 데이터 구축 절차

소지역 상권의 매출액에 미치는 영향력 분석을 위한 데이터는 다음 그림과 같은 데이터 구축절차에 따라 데이터 수집, 데이터 가공 등을 통해 분석 데이터를 구축하였다.



<그림 8> 소지역 상권 매출에 영향요인분석을 위한 데이터 구축절차

앞서 선정된 분석변수를 기준으로 각각 수요요인, 공급요인 등 공간 빅데이터, 통계정보 등을 수집하였다. 데이터 수집은 공간 빅데이터에 해당하는 카드매출, 유동인구, 상점의 수와 기존 통계정보에 해당하는 주거인구, 직장인구, 평균소득(소득분위 및 소득분위별 평균소득) 등을 수집하였다. 다음으로 좌표정의 및 속성매칭에서는 기존의 시군구, 법정동 행정경계 보다 세분화된 분석단위인 소지역 경계를 기준으로 각각의 데이터를 집계하기 위해 공간 데이터 타입을 모두 점(Point) 형태로 구축하였다. 주거인구, 직장인구, 지역의 평균 소득의 경우 100M 격자 중심점 데이터를 활용하며 유동인구, 지역 상점의 수는 POI(Point of Interest)를 활용한다. 카드매출의 경우 본래의(Original) 데이터가 소지역 기준 거래내역으로 이루어지는 텍스트이며 공간참조 식별자는 소지역 ID로 구성

되어 있다. 이러한 소지역 경계를 기준으로 거래내역 데이터를 집계하게 되며 이를 위한 방법은 Arc GIS를 활용하여 테이블 조인을 수행하여 최종적인 소지역 기준 매출 집계 데이터를 구축하였다.

(나) 공간 분석 단위

본 연구는 시군구, 법정동 등 한정된 분석단위에서 벗어나 소지역(Micro Zone) 상권에서의 매출액 영향요인을 분석하는데 의의가 있다. 공간 분석 단위는 GIS 공간분석의 수행에 있어 매우 중요한 요소로 작용하며 기초구역의 단위에 따라 분석과정과 방법이 서로 다르게 적용되고 이에 따라 분석결과 또한 다양하고 상이하게 해석될 수 있다. 따라서 기존의 법정동경계, 통계구역(집계구), 국가기초구역과 소지역의 활용과 구역화 방법에 대한 비교, 검토가 필요하다.

행정경계를 기반으로 가능하면 변화가 적은 도로, 하천, 능선 등의 지형 지물 경계로 구분하고 구역 번호를 할당하여 사용하고 있다. 이렇게 구분된 전국단위 권역(Zone)은 각각의 행정부서에서 활용 시 업무에 따라 권역을 더 분할하거나 몇 개의 권역을 합하여 사용되고 있다.

<표 11> 공간 분석단위 비교·검토

구분	법정동경계	통계구역	국가기초구역	소지역
목적	법으로 정한 구역으로 모든 정부기관의 공부나 재산권 및 관리행사의 법률행위 구역	인구·주택·가구 및 농림어업, 사업체센서스 자료를 토대로 지도위에서 소지역 단위 통계를 제공	도로명주소를 기반으로 국토를 읍·면·동의 면적보다 작게 일정한 경계를 정하여 나눈 구역	기존 공통주택정보, 유동인구, 직장인구, 거주인구, 외감기업 등의 각종 정보를 소지역별 통계를 제공
계층구조	시도 시군구 읍면동 리	시도 시군구 읍면동 기초단위구 집계구	시도 시군구 국가기초구역	시도 시군구 행정동 기초정보(소지역)

구분	법정동경계	통계구역	국가기초구역	소지역
구역 화 방법	지번 필지를 이용하여 구역을 설정 구역 단위로 필지에 코드(숫자) 부여	기초단위:지형지물을 이용하여 구역화 집계구:최적인구수, 토지이용 등 사회적 요인을 반영하여 구획	지형지물, 행정경계, 도시계획, 인구, 건물분포 등 생활권 중심으로 구획	국가기초구역 및 집계구 경계와 연속지적도의 지번을 합계하여 주요 상권지역을 대상으로 세분화
구축 기관	국토교통부	통계청	행정자치부	민간 통신사
유형	면(shp)	면(shp)	면(shp)	면(shp)

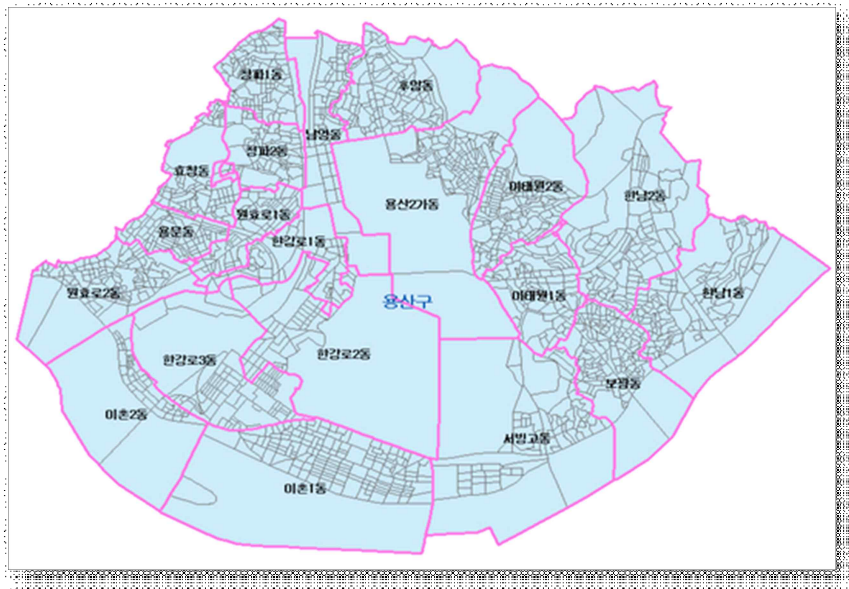
출처 : 통계청(2008), 통계지역획정

먼저 법정동경계는 법률로 지정된 행정구역 단위의 경계, 즉 법(法)으로 정(定)한 동(洞)에 해당되며 그 스케일을 표현하면 다음 그림과 같다.



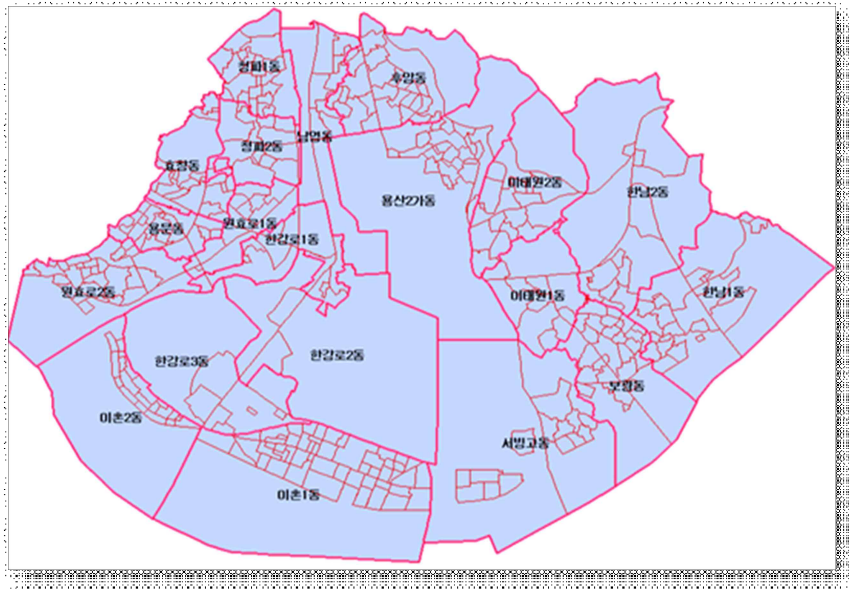
<그림 9> 법정동 행정구역 경계(출처 : 통계청, 2008)

다음으로 통계구역 경계는 통계청에서 구축·관리되는 경계로써 행정구역 경계(읍면동, 법정동) 이 외에도 기초단위구, 집계구 등을 활용하여 센서스 정보뿐만 아니라 다양한 기관에서 생성되는 정보를 결합하여 사용할 수 있는 기초경계 자료로 사용되고 있다. 이러한 경계는 준항구적(準恒久的)인 지형지물을 기준으로 구축되기 때문에 변화가 적다는 장점을 가지고 있다. 기초단위구는 도로, 하천 등 준 항구적인 지형지물을 경계로 하여 획정한 구역으로 조사구 설정 또는 근린지역 통계서비스를 위한 최하 단위구역을 말하며 그 스케일을 표현하면 다음 그림과 같다.



<그림 10> 기초단위구 경계(출처 : 통계청, 2008)

또한 집계구는 기초단위구를 결합시켜 만든 구역으로 기초단위구에 통계적 특성인 동질성, 동량성을 부가하여 획정한 근린지역 통계서비스 구역을 말하며 그 스케일은 다음 그림과 같다.



<그림 11> 집계구 경계(출처 : 통계청, 2008)

본 연구의 소지역단위는 다음 그림과 같이 법정동, 집계구, 기초단위구 경계 보다 더욱 세분화된 공간단위이다. 소지역 경계의 가장 큰 특징은 각종 마이크로 데이터를 취합하여 행정구역 단위 보다 정확한 실제 상점, 배후지 등에 위치정보를 획득할 수 있다는 점이다. 소지역 경계는 통계청의 통계구역과 다르게 명확한 정의나 통일되는 작업지침이 부재하나 일반적으로 집계구 경계를 기준으로 실제 건축물, 지번, 필지 등을 참조하여 ‘블록’ 단위로 세분화하여 구성한다. 공간단위 세분화 방법에 따라 서울시 기준 19,940개의 소지역 경계가 구축되었다. 이러한 소지역 경계는 본 연구에서 선정한 분석변수인 주거인구, 유동인구, 직장인구, 평균 소득, 상점의 수, 카드매출 등을 집계하여 각종 분석에 활용할 수 있다. 다음 그림은 행정동, 집계구, 소지역 등 공간 데이터 분석 단위의 차이를 나타낸다.



<그림 12> 공간 데이터 분석 단위(강남구 논현동 일대)

<그림 11>은 강남구 논현동 일대를 중심으로 (Top)은 행정동 경계와 집계구 경계, (Down)은 행정동 경계와 소지역 경계를 나타낸 것이다.

(다) 공간 빅데이터 구축

본 연구에서 활용하는 공간 빅데이터는 카드매출액과 유동인구자료를 활용한다. 여기서 종속변수는 지역의 매출액이다. 이 자료는 민간통신사의 자료를 취합하여 구축하고 있는 공간 빅데이터 연구단¹⁰⁾의 내부 자료를 활용하였다. 먼저 카드매출데이터는 2014년 1월~2015년 12월 간의 월별 소지역에서 발생한 데이터를 Text형태로 구축하였다. 분석을 위한 데이터구축을 위해 대용량 데이터에 읽기, 쓰기 및 정렬 등이 가능한 IBM 사의 SPSS Statistics 24를 활용하였다. SPSS를 통해 공간 참조가 가능한 소지역 ID를 기준으로 특정 케이스 선택을 위한 질의어(Query)를 입력하게 되면 이에 만족하는 서울시 소지역에 대한 결과를 반환받을 수 있다. 케이스 선택을 통한 결과는 데이터베이스 파일형태(DBF, dBASE Database File)로 저장하게 되며 이러한 결과는 GIS 소프트웨어를 통해 다시 소지역 경계에 매핑(Mapping)할 수 있다. 여기서 GIS 소프트웨어는 ERSI 사의 Arc GIS 10.1을 활용하였으며 GIS 공간분석과 연계 가능하도록 좌표를 설정하고 시간대별, 월별 합계 및 평균 등으로 속성을 매칭하여 구축하였다. 다음으로 유동인구 정보의 경우 카드매출정보와 마찬가지로 SPSS와 Arc GIS를 활용하였으며 X, Y를 소지역 속성에 매칭하였다. 카드매출정보와 유동인구 정보의 처리 전·후 데이터 건수 차이가 상이하게 나타난다. 그 이유로 먼저 카드매출데이터의 경우 공간참조 데이터가 소지역 ID로 정의하여 대부분 소지역을 기준으로 중복되어 집계되는 부분과 누락되는 곳이 일부 존재하기 때문이다. 즉, 1개의 소지역 경계에 중복되는 매출 속성이 다수가 존재하며 이러한 매출 속성을 집계하면 그 건수가 감소하고 또한 소지역 경계에 매출 속성이 포함

10) 공간 빅데이터 연구단은 국토교통부 국토공간정보연구사업 '국토공간정보의 빅데이터 관리, 분석 및 서비스 플랫폼 기술개발'(16NSIP-B081011-03)을 말한다.

되지 않는다면 이를 제외하여 나타난다. 반면 유동인구는 X, Y점(Point) 형태의 정보를 소지역으로 공간조인한 결과로 이를 통해 시간대별, 월별 유동인구 자료를 구축하였기 때문에 소지역 코드로 집계된 Polygon데이터와 유동인구의 Point 데이터의 건수가 다소 차이가 있다. 다음 집적은 공급요인에 해당하는 변수로써 본 연구의 대상인 서울특별시 전체의 상점의 수를 집적으로 판단한다. 데이터의 구축을 위해 SPSS를 활용하여 데이터베이스 파일을 생성하였으며 X, Y 좌표를 기준으로 Arc GIS에 시각화하였다. 분석·연계를 위해서 격자별로 전체 상점의 수의 총합을 집계하여 변수로 활용하였다. 공간 빅데이터에 대한 구축 정보를 정리하면 다음 <표 12>와 같다.

<표 12> 공간 빅데이터 구축 데이터

데이터명	원 데이터 건수	처리 후 데이터 건수 (서울시 기준)	처리 후 데이터 건수 (소지역 기준)	출처
카드매출 데이터	7,525,969	-	14,432 (소지역 코드로 매핑)	공간 빅데이터 연구단 내부자료
유동인구 데이터	3,674,082	165,361	19,940 (소지역 기준 Spatial Join)	공간 빅데이터 연구단 내부자료
상가업소 데이터	2,505,817	530,623	19,940 (소지역 기준 Spatial Join)	중소기업청

(나) 수요·공급요인 데이터구축

다음으로 독립변수의 수요요인은 앞서 언급한 유동인구, 지역의 주거인구수, 직장인구수, 지역의 소득수준이 해당된다. 본 연구에서는 공간 빅데이터 기반의 세분화된 지역을 분석단위로 설정하고 있으므로 이와 동일하거나 연계분석이 가능한 형태로 자료가 구축되어야 한다. 이를 위해 먼저 지역의 주거 인

구수는 행정자치부 새주소 사업DB 중 건물의 위치정보와 밀집시설물의 정보를 통해 추정된 격자 중심점(100M) 주거 인구를 Biz-GIS 사의 XsDB 서비스에서 다운로드하여 활용하였다. 직장인구수 또한 XsDB 서비스에서 집계구별 사업체 정보를 통해 추정된 점 데이터를 활용하였다. 기준연도는 주거인구, 직장인구 모두 2011년이다. 본 연구 시간적 범위와 상이함에도 불구하고 이를 활용하는 이유는 행정동 기준으로 집계한 공간단위의 분석이 아닌 격자 단위의 분석을 위해서 집계구나 소지역단위의 점 데이터가 필요하기 때문이다. 따라서 본 연구에서의 매출 영향력 분석모형에서 독립변수에 설명력을 면밀하게 살펴볼 필요가 있다. 소득수준의 경우 공동주택, 오피스텔, 공시지가를 통해 추정된 추정소득분위자료와 2015년 통계청, 전국경제인연합회에서 제공하는 소득분위별 평균소득 자료를 활용하여 지역별 인구수에 따른 평균소득을 구축하였다. 이를 위해 활용한 수식은 다음과 같다.

$$g.INcome(i) = \frac{\sum (k) \times Pop(k)}{Pop(k)} \quad <수식 1>$$

위 식에서 $Avg.INcome(i)$ 는 단위지역인 100m에서의 평균소득을 의미한다. 구체적으로는 100m 크기의 정방형 격자 k에서의 평균소득($come(k)$)을 인구수($Pop(k)$)를 가중치로 적용한 인구가중 평균소득에 해당한다.

<표 13> 수요·공급요인 구축 데이터

데이터명	원 데이터 건수	처리 후 데이터 건수 (서울시 기준)	처리 후 데이터 건수 (소지역 기준)	출처
주거인구수	959,660	29,638	11,237 (소지역 기준 Spatial Join)	Biz-GIS XsDB 서비스
직장인구수	421,514	24,388	13,038 (소지역 기준 Spatial Join)	Biz-GIS XsDB 서비스
평균소득	959,660	29,638	13,038 (소지역 기준 Spatial Join)	Biz-GIS, 통계청 전국경제인연합회

거리는 경쟁업체와의 거리뿐만 아니라 상점과 소비자와의 거리를 모두 포함하는 개념이다. 본 연구에서 활용하는 공간가중회귀모형에는 모형자체에 거리의 영향력이 포함되기 때문에 따로 분석변수로 거리관련 변수를 포함하지 않았다.

제2절 분석방법론

공간가중회귀분석(GWR, Geographically Weighted Regression)은 (Fotheringham 외, 1998)가 제안한 회귀분석방법론의 하나이다. 경제학을 포함한 사회과학 분야에서 일반적으로 널리 활용되는 회귀분석은 OLS(Ordinary Least Square, 이후 OLS)로 지칭되는데 이는 회귀분석 방법론의 근본적 구성원리는 매우 유사하다. 즉, 잔차(residual) 제곱의 합을 최소로 함으로써 가장 설명력이 높은 회귀계수를 도출한다는 의미를 담고 있는 용어인 셈이다. OLS는 종속변수와 독립변수 간 관계성 (relationship)을 파악하기 위한 보편적인 수단으로 활용되는데 일반적으로 표현되는 회귀계수가 이러한 관계성을 정량화한 결과이다. OLS는 잔차 제곱의 합을 최소화하는 방식으로 회귀계수를 추정한다. 잔차 제곱의 합의 최소화란 결국 종속변수 값이 형성되는데 독립변수가 기여한 정도 (혹은 설명력)를 주어진 조건 하에서 가장 잘 설명할 수 있는 회귀계수 추정을 위한 정량적 기준인 셈이다. 주어진 조건이란 회귀계수 추정에 이용할 수 있는 자료의 범위나 표본의 수를 의미한다. 회귀계수는 일반적으로 아래의 <수식 2>와 같이 정의한다.

$$= +B_1X_1+ +B_iX_i+ +B_mX_m+\varepsilon \quad \text{<수식 2>}$$

OLS로부터 추정한 회귀계수는 독립변수가 종속변수에 미치는 영향력 혹은 설명력이 동일하다는 특징을 가지고 있다. 그러나 이러한 가정을 전제로 공간자료(spatial data)에 OLS를 적용하여 도출한 회귀계수에는 상당한 오류의 가능성이 내포되어 있다. 공간자료라 함은 지리적 공간상에서 점의 위치 (point location)를 결정하거나(occupying) 일정한 지리적 범위를 포괄하는

(covering) 지역(region)을 설명하는 자료를 의미한다. 예를 들어, 동 단위는 행정구역 체계를 정의하는 공간적 위계 속에서 일정한 범위 혹은 크기를 차지하는 지리적 영역을 지칭한다. 따라서 동 단위는 지역이라는 의미를 내포하고 있다. OLS에서는 공간자료들이 가지는 표본 값이 통계적으로 독립적인 방식으로 추출된 것임을 가정한다. 즉, 표본들 간 독립성을 전제로 분석을 수행하는데, 상권분석을 위한 자료는 많은 경우 시군구나 읍면동과 같은 단위지역으로 집계하여 수집되는 것이 일반적이기 때문에 인접한 단위지역들 간 공간적 자기상관성(Spatial Autocorrelation)이 발생할 가능성이 크다. 공간적 자기상관성이란 지리적으로 인접할수록 비슷한 특성을 가질 가능성이 높음을 의미하는 명제적 전제이다. 따라서 표본자료들 간 인접한 정도에 따라서 통계정보로서의 유사성은 증가하고 독립성 가정이 위배될 가능성 역시 커지게 된다. 결과적으로 추정한 회귀계수는 신뢰하기 어려운 왜곡된 결과일 여지가 크다. 지리적으로 인접한 단위지역들에서 수집한 표본자료에 OLS를 적용하여 추정한 결과를 바탕으로 잔차도(residual map)를 그려보면 공간적 자기상관성이 뚜렷하게 나타난다. 즉, 특정 지역을 중심으로는 실제 값보다 과하게 추정되어 양의 값을 가지는 잔차가 지리적으로 군집해있고 반대로 또 다른 특정 지역에서는 실제 값보다 작은 값이 추정된 결과 음의 값을 가지는 잔차가 모이는 형태가 되는 것이다. 이는 회귀계수 값이 어느 지역에서나 동일하게 적용될 수 있다는 OLS에 내포된 전제를 심각하게 위배하는 것이다. 이러한 결과는 OLS가 자료들 간 지리적 거리 즉, 인접성에 상관없이 통계적으로 독립적인정보를 가진다는 가정에서 분석한 결과이지만 실제로는 가까운 표본자료들일수록 공간적 자기상관성으로 인하여 정보의 유사성이 증가하고 전혀 독립적이지 않은 자료임에도 불구하고 독립성 가정을 적용하여 회귀계수를 추정한 것에 기인한다. 이는 잔차의 공간적 패턴이 뚜렷하게 나타나는 결과로 귀결된다. OLS가 가지는 이러한 한계를 보완하기 위하여 지역적으로 인접한

자료들 간 공간적자기상관성을 반영할 수 있도록 공간회귀모형(spatial regression model)이 새롭게 대두되었다. 공간회귀모형은 표본자료들 간 공간 자기상관성으로 인하여 발생하는 잔차를 모형에서 별도의 독립변수로 포함하여 오류를 최소화하는 것을 목적으로 한다. GWR은 종속변수에 대한 독립변수의 영향력이 표본지역의 위치마다 달라질 수 있음을 전제한다. 즉, 독립변수와 종속변수의 관계성에 있어서 공간적 이질성(spatial heterogeneity)을 반영하여 표본지역마다 다른 값을 가지는 회귀계수를 추정하는 것이다. 따라서 GWR은 회귀계수의 지역 간 편차를 이해하기에 편리한 분석도구이다. 아래의 <수식 4>는 GWR 모형을 수학적으로 정의한 것이다.

$$(i) = (i) + B(i)_1X_1 + \dots + B(j)_iX_i + \dots + B(j)_mX_m + \varepsilon \quad \text{<수식 3>}$$

$$\beta(j) = (X^T W(j) X)^{-1} X^T W(j) Y \quad \text{<수식 4>}$$

$$W_{ij} = \begin{bmatrix} W_{ij} & 0 \\ 0 & W_{im} \end{bmatrix} \quad \text{<수식 5>}$$

$$W_{ij} = \frac{1}{2\pi\sigma} e^{-\frac{d_{ij}^2}{2\sigma}} \quad \text{<수식 6>}$$

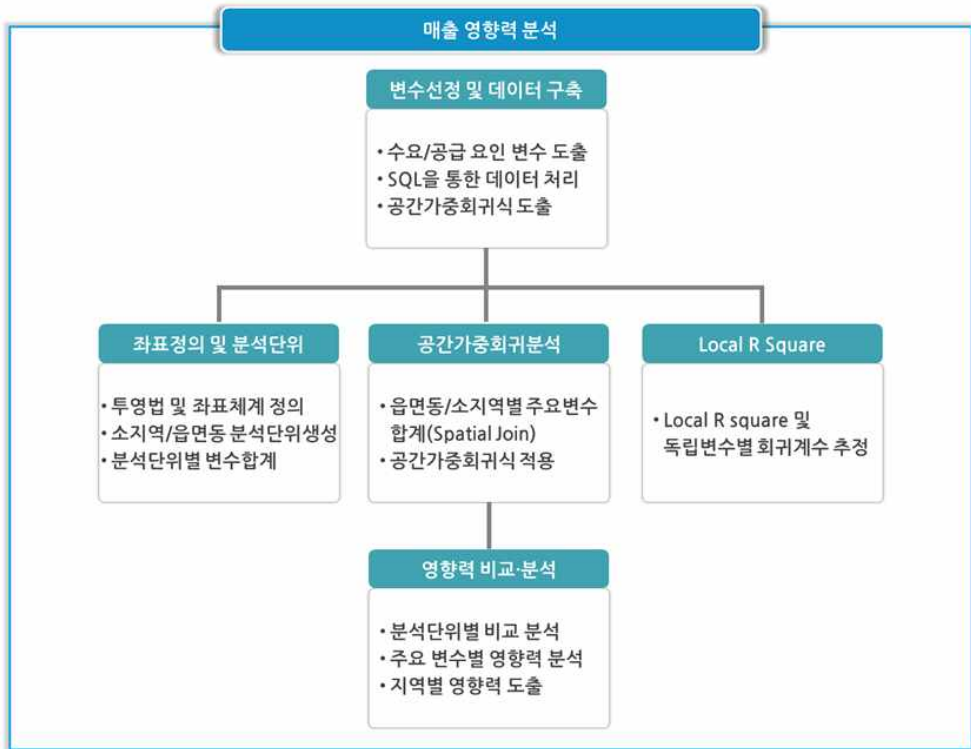
<수식 3>으로 정의되는 GWR과 OLS를 구별하는 핵심요소는 <수식 5>에서 정의한 공간가중행렬(spatial weight matrix)이다. 공간가중행렬의 대각요소들은 각 행과 열에 해당하는 단위지역들 간 거리 증가에 따라 감소하는 거리조락함수(distance-decay function)로 추정한다. 여기서 거리조락함수는 거리증가에 따른 관계성 약화 또는 감소폭을 의미한다. 거리조락함수에서 핵심이 되는 요소는 대역폭(Bandwidth)이며 <수식

6>에서 표시를 의미한다. 대역폭이 클수록 멀리 떨어진 지점에 상대적으로 큰 공간가중치가 부여된다. 큰 공간 가중치가 먼 지역에 부여된다는 것은 변수들 간 관계성이 먼 지역에서도 가중치에 비례하여 그 유사성이 증가한다는 것을 의미한다. 반대로 대역폭이 작으면 가까이 있는 지점에 부여되는 공간가중치는 큰 반면 거리가 증가함에 따라 공간가중치의 감소폭이 더 크다. 정리하면 대역폭은 한 지점으로부터 거리가 증가함에 따라 공간적 유사성이 감소하는 폭을 조정하기 위한 파라미터이다.

제3절 상권매출 영향력 분석 모형

본 연구는 공공, 민간에서 생성되는 공간 빅데이터를 활용하여 서울시 내 지역상권의 매출액에 영향을 주는 변수를 선정하고 그 영향력을 파악·검증하여 다양한 분야에 활용방안을 제시하는데 목적이 있다. 이를 위해 본 연구에서는 카드매출, 유동인구, 상가업소 수 등 공간 빅데이터와 매출액 영향을 주는 주거인구, 평균소득, 직장인구 데이터를 구축하였다. 매출액 영향력 추정을 위해 활용하는 방법론을 그 간 사회과학분야에 활용되어온 회귀분석 이론을 지리학법칙과 응용한 GWR을 활용한다. 또한 GIS 공간분석 수행에 있어 발생하는 가변적공간단위문제(MAUP, Modifiable Areal Unit Problem)를 해소하고 보다 세분화된 분석이 가능한 소지역 분석단위를 활용하였다. 이를 위해 모든 독립변수는 점 데이터 형태로 구축하였으며 각각의 독립변수는 소지역 단위에 합계되어 분석된다.

앞서 살펴보았듯이, GWR 모형에 적용하기 위하여 설정한 종속변수는 공간 빅데이터의 카드매출 데이터이다. GWR의 독립변수로는 1) 주거인구, 2) 유동인구, 3) 직장인구, 4) 평균소득, 5) 상가업소 수 등 5개 변수를 설정하였다. 본 연구의 분석모형을 도식화하면 다음 <그림 12>와 같다.



<그림 13> 매출 영향력 분석 모형

제4장 매출액 영향요인 실증분석

제1절 분석의 개요

본 장에서는 앞서 2장, 3장에서 도출된 분석변수와 데이터, 분석모형을 토대로 공간분석 소프트웨어 중 하나인 Arc GIS 10.1 활용하여 서울특별시 전체 지역상권의 매출액에 미치는 영향요인을 분석한다. 서울특별시는 대한민국의 수도로써 한반도 중앙에 있으며, 한강을 사이에 두고 남북으로 펼쳐져 있다. 행정 구분은 총 25구로 구분되어 있으며, 한반도의 0.28%(남한 면적의 0.61%)에 해당하는 넓이로, 남북 간 연장거리 30.30km, 동서 간 연장거리 36.78km이다. 본 연구의 분석모형적용을 위해 서울특별시를 기준으로 분석단위를 생성하였다. 앞서 3장에서의 변수를 참고하여 종속변수와 독립변수에 대한 실증자료는 소지역 단위로 구축하였다. 최종적으로 선정된 종속변수는 공간 빅데이터의 카드매출데이터이며, 독립변수는 종속변수에 영향을 주는 주거인구 수, 유동인구 수, 직장인구 수, 지역상권 내 상점의 수로 선정하였다. 모든 변수들에 대하여 소지역 단위에 해당되는 지역으로 실증자료를 취합하며 공간가중회귀 모형에 적용하여 회귀계수 등을 추정하는 실증분석을 수행하였다. 이를 통해 도출한 분석결과는 각각의 독립변수별로 구체적인 영향력 수준을 해석하였다. 또한 해석결과를 통해 시사점을 도출하고 본 연구에서의 분석모형을 검증하기 위해 Local R Square와 각 독립변수의 회귀계수를 지역별로 시각화하여 비교·분석하였다.

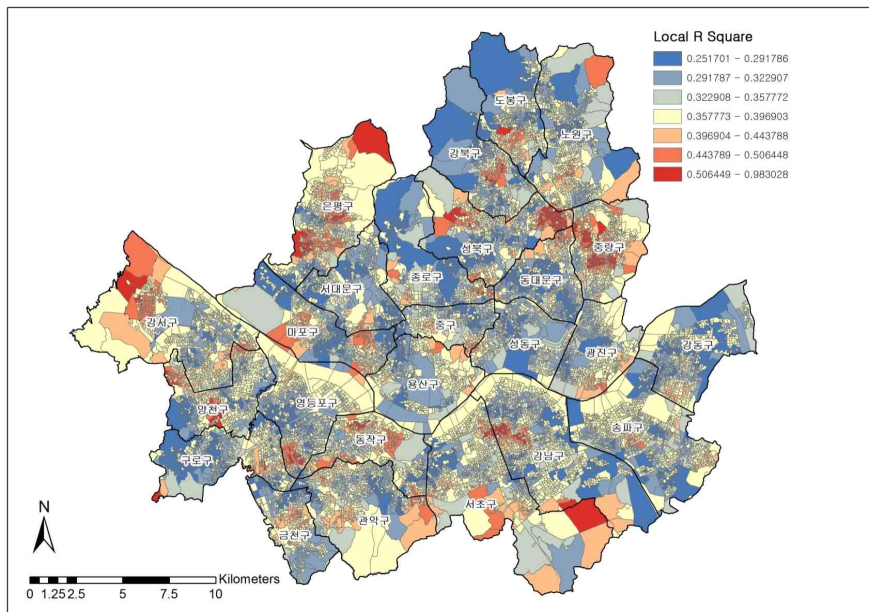
제2절 매출 영향력 분석

(1) 국지적 R Square

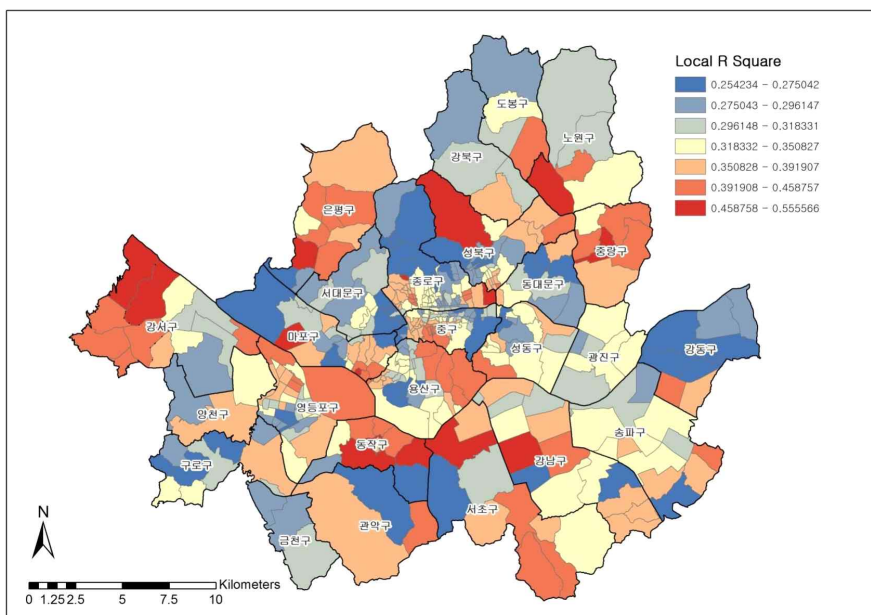
본 연구는 종속변수 매출액에 대한 영향력을 분석을 위해 매출영향요인에 해당되는 독립변수 주거인구, 유동인구, 직장인구, 평균소득, 지역상권 내 상점의 수 등을 소지역 단위로 합계하여 공간가중회귀분석을 실시하였다. 또한 분석결과의 직관적인 확인이 가능하도록 읍면동 단위 합계분석을 추가하여 비교하였다. 공간가중회귀분석모형은 독립변수마다의 회귀계수를 분석하며 그 결과를 통해 종속변수인 매출액에 미치는 영향력을 지역별로 파악하는데 활용 할 수 있다. 다음 <수식 7>은 상권영역에서의 회귀계수를 추정하기 위한 GWR 모형을 나타낸다.

$$\begin{aligned} \text{출액 } j) = & (j) + B(j)_1 X_{\text{거인구}} + B(j)_2 X_{\text{유동인구}} \\ & + B(j)_3 X_{\text{직장인구}} + B(j)_4 X_{\text{평균소득}} + \\ & B(j)_5 X_{\text{지역의 점포수}} \end{aligned} \quad \text{<수식 7>}$$

서울특별시를 소지역으로 구분한 결과 19,440개의 분석단위가 생성되었으며 19,440개의 회귀식을 통해 국지적 R Square를 분석하였다. 즉 본 연구의 회귀식이 모든 소지역 단위 19,440개에 적용되어 소지역별 회귀계수와 국지적 R Square가 산출되는 것이다. 여기서 국지적 R Square는 0에 가까울수록 설명력이 낮은 값을 의미하고 반대로 1에 가까울수록 설명력이 높은 값을 나타낸다. 다음 그림은 국지적 R Square 값의 공간적 분포를 나타낸다. 국지적 R Square 값이 붉은색일수록 1에 가까우므로 매출액에 미치는 영향력이 높고 푸른색일수록 0에 가깝게 나타나므로 매출액 미치는 영향력이 낮은 지역임을 확인 할 수 있다.



<그림 14> 소지역 단위의 국지적 R 스퀘어 추정결과



<그림 15> 읍면동 단위의 국지적 R 스퀘어 추정결과

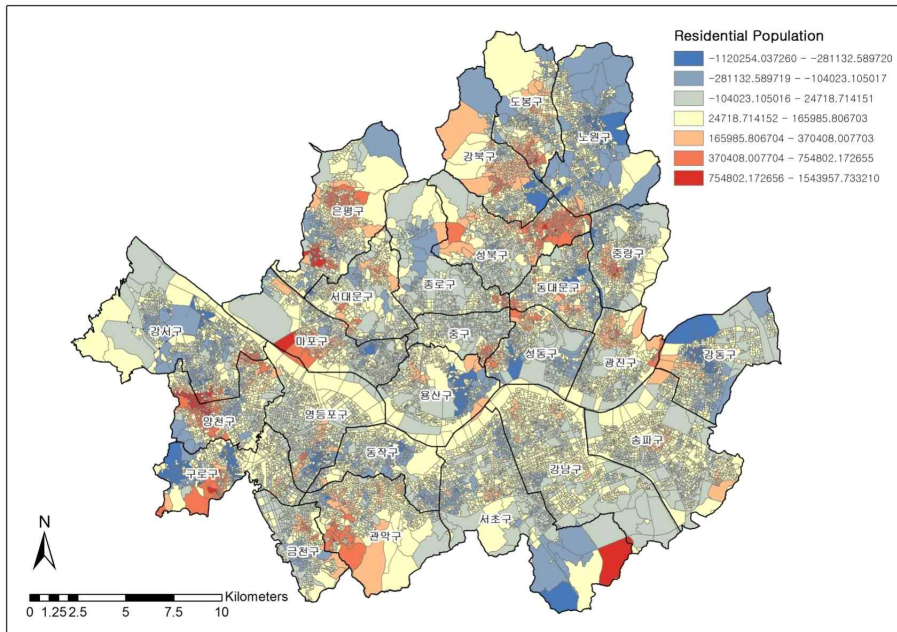
<그림 13>을 보면 용산구, 마포구 일대, 영등포구 일대, 동대문구 일대 등에서 R Square값이 상대적으로 낮은 것을 알 수 있다. 이는 일정 인구규모 이상의 지역에서의 매출액이 연구에서 설정한 5개의 독립변수 이외에도 다른 요인들에 의해 영향을 받을 가능성이 높다는 것을 의미한다. 예를 들어 연구변수에 포함되지 않은 인터넷 쇼핑, 홈쇼핑의 활용빈도 등의 소비행태, 백화점의 이용, 대형마트와 같은 경쟁업소 등의 영향이 나타날 수 있다.

반면 붉은 색으로 표시된 지역은 양천구 일대, 은평구, 강남구 일대, 강북구 일대, 동작구 등에 속해 있는 상권으로 나타났다. 이러한 지역은 본 연구에서 선정한 5개의 독립변수의 영향력이 적절히 반영된 것을 시사한다. <그림 14>를 살펴보면 마찬가지로 용산구, 마포구 일대, 영등포구 등이 약 0.40 이상의 R Square값으로 나타내어 본 연구의 독립변수가 큰 영향을 미치는 것으로 나타났다. 이와 대조적으로 서대문구, 강동구, 동대문구 등에서는 전반적으로 낮은 R Square 값을 나타내어 본 연구의 독립변수 이외에도 다양한 변수에 의한 영향이 높음을 알 수 있었다. 또한 소지역의 분석과의 차이점은 동작구, 영등포구 등과 같이 특정 지역의 값이 합계되어 나타나는 가변적공간단위문제(MAUP, Modifiable Areal Unit Problem)가 존재하는 것을 알 수 있다. 즉, 집계 단위가 행정구역 읍, 면, 동을 기준으로 전반적인 경향을 측정하는 데에는 효과적일 수 있으나 보다 지역 내 세부지역의 영향력을 측정하는 데에는 한계가 있다. 소지역 분석의 이점은 이러한 한계를 보완하여 보다 정밀한 분석이 가능하고 원인을 발견하여 진단할 수 있다는 점이다. 다만 전반적인 경향을 파악하거나 적절한 시각화 등 분석 목적과 데이터 스케일 등을 고려하여 이에 부합하는 분석단위 선정이 필요할 것으로 판단된다.

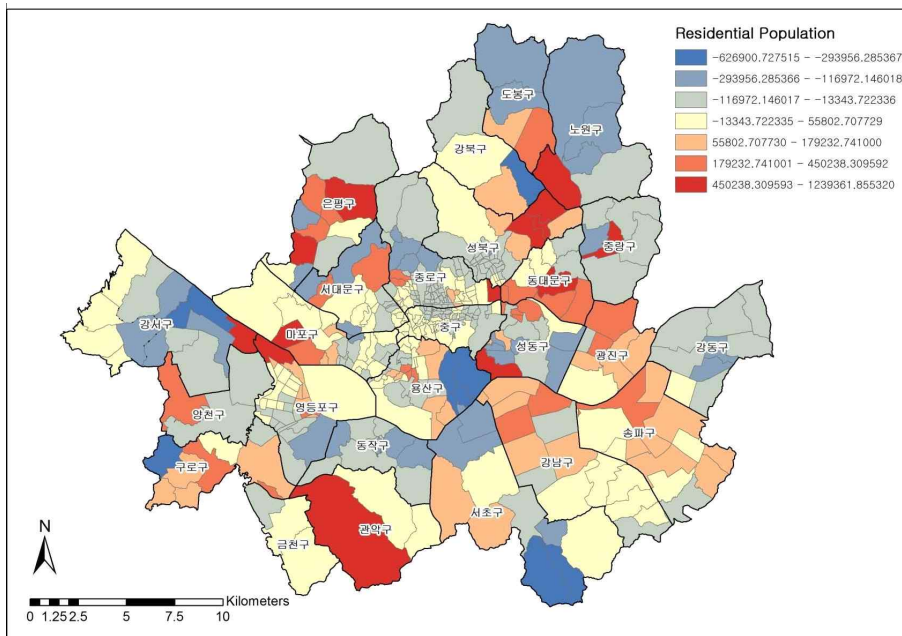
(2) 독립변수의 영향력 측정

소지역 상권의 매출에 미치는 영향력 분석에서 주거인구, 유동인구, 직장인구, 지역의 평균소득, 상점의 수 등 개별 독립변수의 영향력을 분석하였다. 분석결과 공간가중회귀모형의 적용을 통해 독립변수 마다의 회귀계수가 추정되었다. 공간가중회귀모형의 회귀계수는 소지역 및 읍면동 등 분석단위의 개수만큼 추정되므로 추정된 모든 회귀계수 값을 제시하는 것보다는 지역적인 차이를 비교하는 것이 효과적이다(김진희 외, 2011). 따라서 대표적인 지도 시각화 방법 중 단계구분도(Choropleth Map)로 표현하며 급간구분방법은 Natural Breaks를 적용하였다. 이를 통해 도출된 분석결과를 최소 값, 중위 값, 최대 값의 영향을 적절히 배분하여 표현하였다.

다음 <그림 15>는 잠재 수요변수 중 하나인 주거인구를 소지역으로 구분하여 회귀계수를 시각화한 것이다. 서울특별시 매출액에 미치는 영향력의 지역 차이를 확인해 보면 주거인구의 영향력이 가장 큰 지역은 마포구, 양천구 등이다. 이는 해당 지역의 인구가 매출액에 크게 영향을 주는 것을 의미하며, 지역민이 아닌 외부요인의 영향이 다른 지역과 비교적으로 적다는 것을 의미한다. 반면 영향력이 낮은 지역은 구로구 일대, 용산구 일대로 나타났다. 이는 지역 내 배후인구보다 제3의 요인으로 인한 결과일 수 있으며 또한 지역 내 인구의 수요가 적은 지역으로 해석할 수 있다. <그림 16>은 읍면동단위 분석결과를 나타내는 것으로 대체적으로 소지역 단위분석과 비슷한 결과를 나타냈으나 동대문구, 성동구, 송파구 등에서 가변적공간단위문제가 발생하였다.

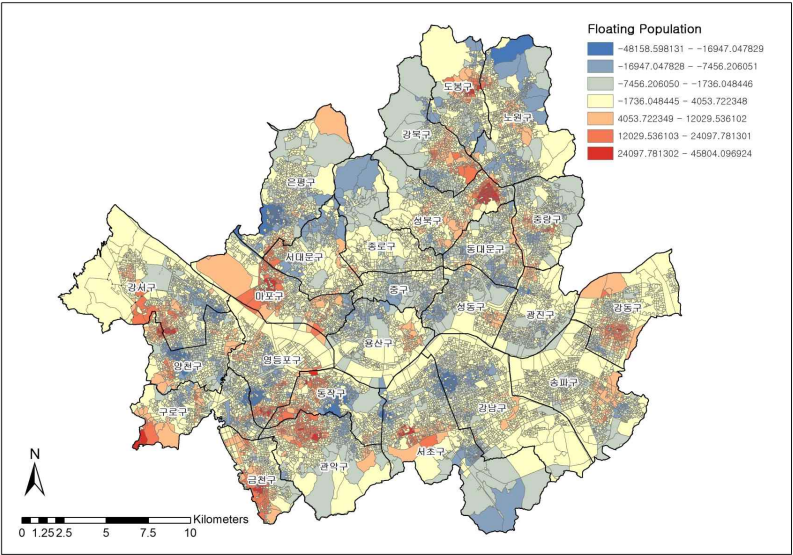


<그림 16> 소지역 단위의 주거인구 회귀계수 지역편차

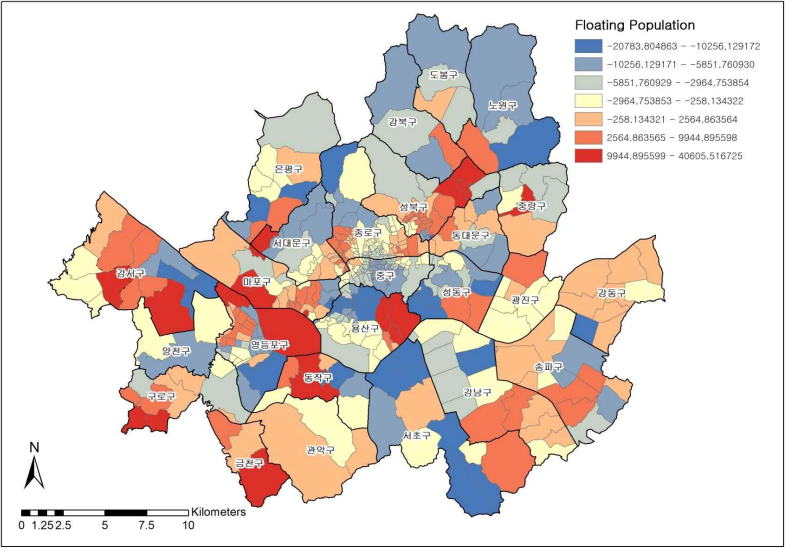


<그림 17> 읍면동 단위의 주거인구 회귀계수 지역편차

다음 그림은 유동인구에 대한 지역별 회귀계수를 나타낸 것으로 유동인구가 지역 매출액에 미치는 영향력에 대한 단위 지역별 차이를 확인 할 수 있다.



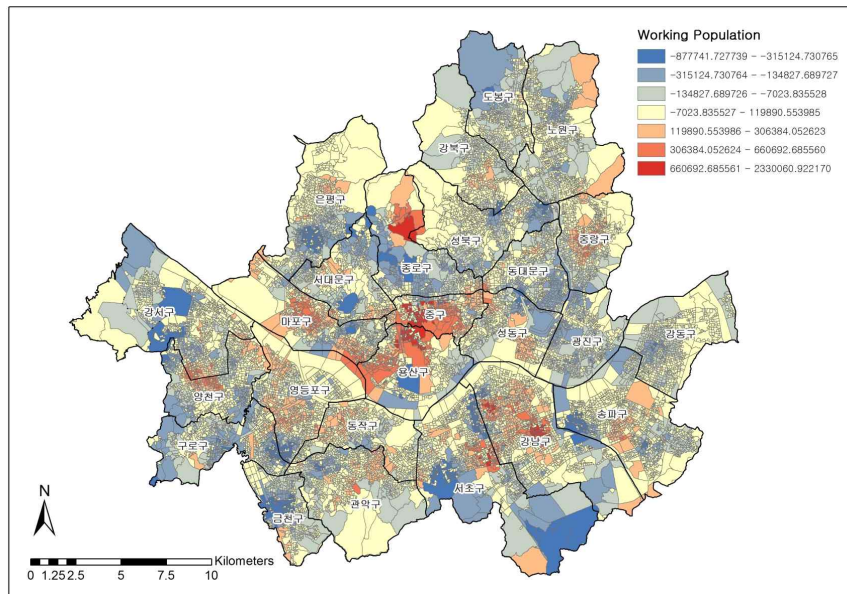
<그림 18> 소지역 단위의 유동인구 회귀계수 지역편차



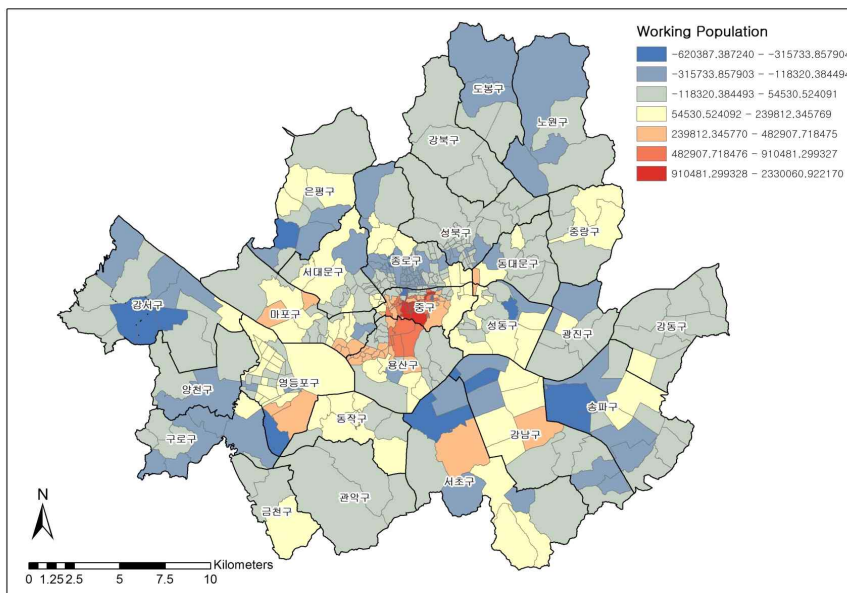
<그림 19> 읍면동 단위의 유동인구 회귀계수 지역편차

먼저 유동인구의 영향력이 가장 크게 나타나는 지역은 성북구이다. 이는 해당 지역의 매출액이 특정 지역의 유동인구에 크게 영향을 받는다는 것을 의미하며 직·간접적인 잠재수요가 높은 지역임을 뜻한다. 반면 양천구, 은평구 일대에서는 상대적으로 낮은 영향력을 보이는데 이는 해당 지역의 유동인구가 타 상점으로 흡수되었을 가능성이 높거나 실제 유동인구가 많더라도 상점 매출액에 미치는 영향은 상대적으로 낮은 지역으로 해석할 수 있다. 은평구 지역은 음(-)의 영향력이 가장 크게 나타나며 이는 인접지역인 서대문구의 일부에서 수요를 흡수하여 나타난 결과로 해석할 수 있다. 또한 강남의 경우 양(+) 영향력과 음(-) 영향력의 지역이 상존하며 양, 음의 영향력이 서로 대립하는 특징이 있다. 일반적으로 강남구 유동인구는 타 지역보다 상대적으로 많으나, 다양한 음(-)에 요인으로 인해 대체적으로 영향력이 낮은 지역으로 도출되었다. 읍면동 단위 분석결과와 비교하여 매출액에 대한 유동인구의 영향력이 양(+)을 나타내는 지역으로 영등포구, 동작구, 강서구 일대, 용산구 일대 등으로 도출되었다. 이러한 결과는 인접지역과 특정지역의 회귀계수가 전체를 경향을 나타내는 가변적공간단위문제로 인한 결과로 해석된다.

<그림 19>는 직장인구 회귀계수에 대한 지역차이를 나타낸다. 직장인구는 주거인구와 함께 지역 내 매출을 발생시키는 잠재인구로 해석된다. 지역 매출액에 대해 높은 영향력이 미치는 지역은 주로 지역중심에 위치하고 있는 중구와 용산구 일대, 마포구, 강남구 등이 해당된다. 이는 해당지역에서 직장인구가 매출액에 큰 영향을 주는 요인임을 의미한다. 반면 강서구, 금천구, 도봉구 등의 외곽지역과 강남, 서초 등의 일부지역에서도 매출액에 음(-)의 영향력을 미치는 것으로 나타났다. <그림 20>과 같은 읍면동단위 분석에서는 마포구, 강남구, 서초구 일대 등에서 영향력이 다소 낮게 측정되었으며 양(+) 영향을 주는 지역이 모두 합계되어 대체적으로 낮은 지역이 나타났다.

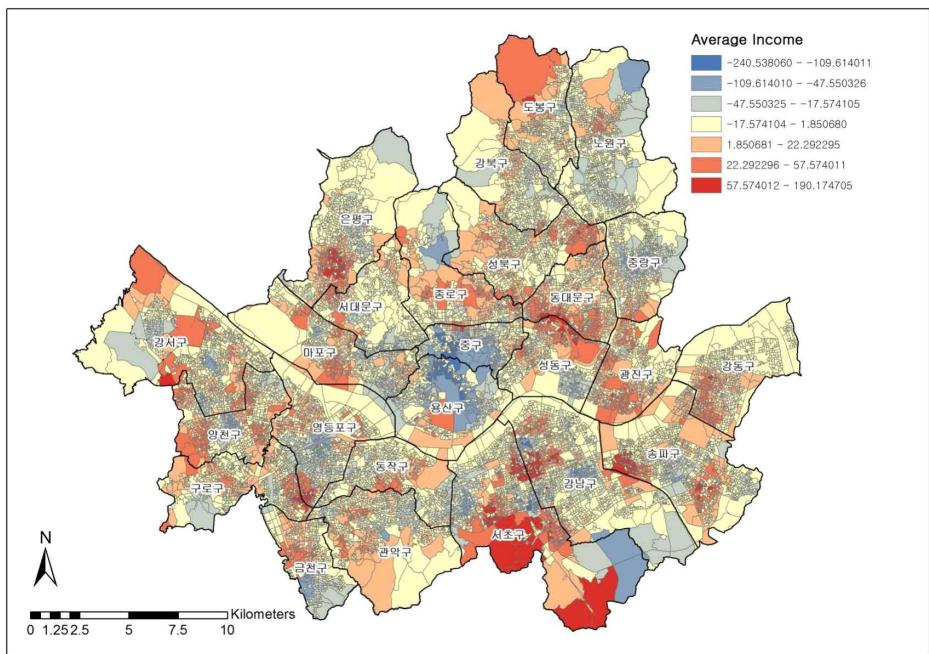


<그림 20> 소지역 단위의 직장인구 회귀계수 지역편차

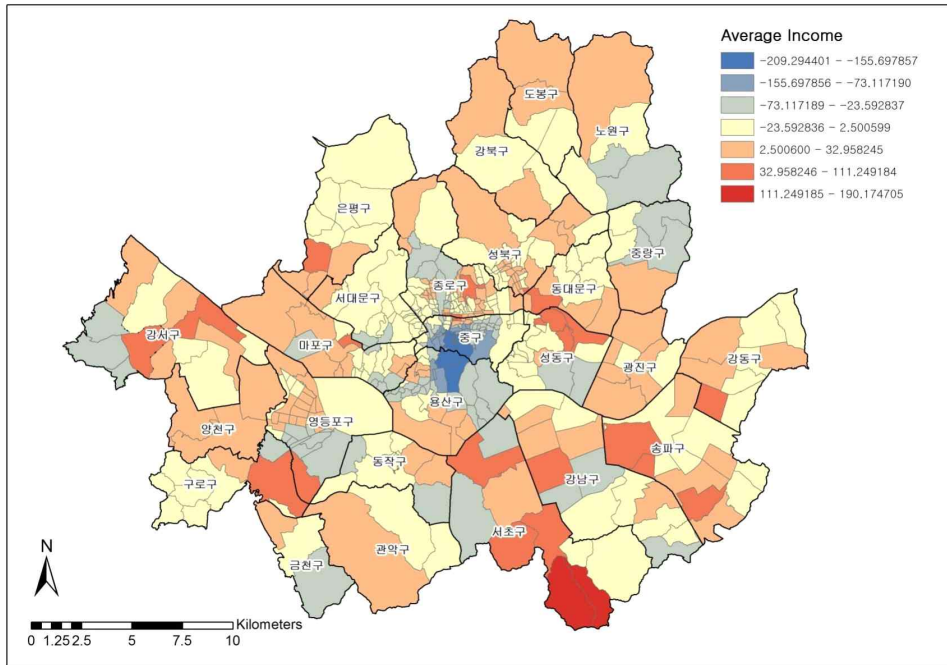


<그림 21> 읍면동 단위의 직장인구 회귀계수 지역편차

<그림 21>은 평균소득의 회귀계수를 지역별로 나타낸 것이다. 평균소득 변수는 지역상권 매출에 대체적으로 양(+)의 영향을 주는 것으로 나타났다. 먼저 양(+)의 영향력이 가장 뚜렷한 지역은 강남구와 서초구, 동대문구와 성북구 등 행정구역이 맞닿은 일대지역이었으며 지역 내에 영향력 또한 모두 높은 것으로 분석되었다. 또한 동대문구, 광진구, 송파구 등 일대에서 높은 영향력을 보였다. 반면 중랑구와 중구 그리고 용산구의 일부지역에서 낮은 음(-)의 영향력을 나타냈다. <그림 22>와 같은 읍면동단위 분석에서는 중구, 용산구 일대를 제외하고 대체로 매출액에 양(+)의 영향을 주었으며, 양(+), 음(-)의 회귀계수가 합계되어 비교적 완만한 값이 측정되었다.

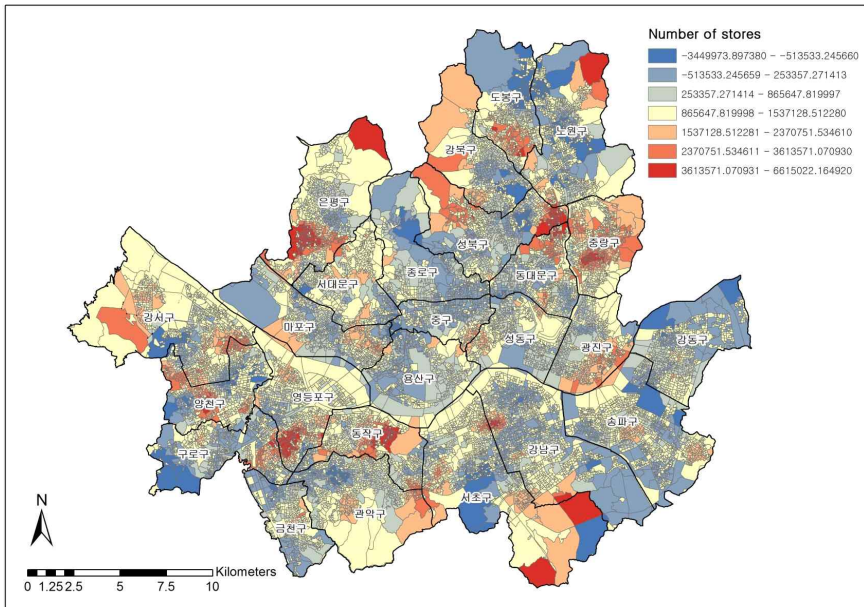


<그림 22> 소지역 단위의 평균소득 회귀계수 지역편차

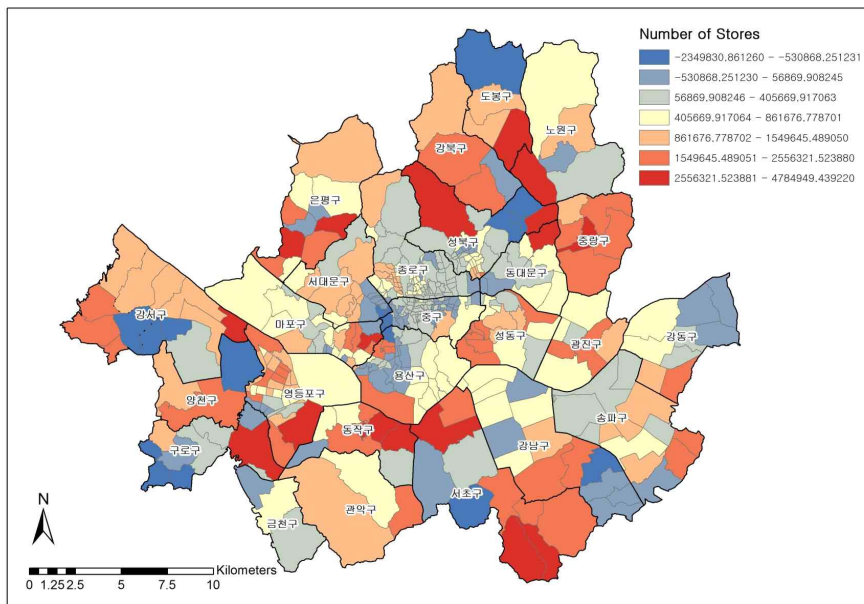


<그림 23> 읍면동 단위의 평균소득 회귀계수 지역편차

<그림 23>은 지역 내 상가업소 수에 대한 회귀계수를 지역별로 나타낸 것이다. 이는 지역상권의 집적효과를 측정하기 위한 변수로써 특정 지역의 특성을 나타내고 있다. 먼저 매출에 대한 집적효과를 통해 양(+)의 영향력이 가장 뚜렷한 지역은 동대문 북부지역이었으며 이와 인접한 중랑구에서도 높은 영향력을 나타냈다. 또한 광진구, 동작구, 강남구의 일부지역, 양천구 일대에서 집적효과가 지역상권 매출액에 높게 기여하고 있는 것으로 나타났다. 반면 용산구, 중구, 종로구, 구로구 등에서 음(-)의 영향을 나타냈으며 송파구, 관악구, 서대문구, 마포구 등에서는 양의 영향과 음의 영향이 함께 나타난 것을 알 수 있다. 읍면동단위 분석 결과의 경우 일부지역의 값에 따라 매출에 대한 영향력이 대부분 양(+)의 지역으로 가변적공간단위문제가 많은 곳에서 발생했음을 알 수 있다.



<그림 24> 소지역 단위의 지역 내 상가업소 수 회귀계수 지역편차



<그림 25> 읍면동 단위의 지역 내 상가업소 수 회귀계수 지역편차

제3절 분석결과의 종합

본 연구에서는 서울특별시 지역상권의 매출 영향력 분석을 위해 선행 연구 검토 및 이론고찰을 통해 지역상권에서의 매출 영향 요인을 도출하였으며 기준요소의 검토를 통해 최종적인 분석변수를 선정하였다. 종속변수의 경우 공간 빅데이터의 카드매출정보로 활용하였으며 이어 매출액에 영향을 주는 독립변수는 주거인구 수, 유동인구 수, 지역의 평균소득, 상가업소 수가 해당된다. 변수선정을 토대로 실증분석에 활용하기 위한 데이터를 수집하였으며 빅데이터 처리와 공간분석 연계·분석을 위한 응용프로그램은 SPSS와 Arc GIS를 활용하였다.

종속변수 매출에 대해 영향을 주는 5개 독립변수는 소단위 지역 19,940개에 적용되어 19,940개의 회귀계수를 추정한다. 또한 동 단위 분석에서는 각 독립변수를 서울특별시 읍면동 단위의 467개에 적용하여 회귀계수를 추정하였다. 먼저 국지적 R Square 추정결과는 5개의 독립변수의 영향력 종합을 통해 얻어지는 모형의 설명력을 나타낸다. 분석을 통해 가장 영향력 있는 지역은 양천구, 은평구, 강남구, 강북구, 동작구 등에 지역으로 나타났다. 반면 용산구, 마포구, 영등포구, 동대문구에서의 영향력이 낮은 것으로 분석되었다. 각 독립변수별 영향력 분석에서는 해당 변수별로 회귀계수를 추정하였다. 주거인구 변수에서는 영향력이 가장 큰 지역이 마포구, 양천구 등으로 분석되었으며 반면 영향력이 낮은 지역은 구로구 일대, 용산구 일대로 나타났다. 다음 유동인구에서 가장 영향력이 높은 지역으로 성북구, 동작구 일대 등이 분석되었으며 영향력이 낮은 지역은 시 외곽에 위치한 강북구, 노원구 일대와 양천구, 은평구 일대로 분석되었다. 직장인구의 경우 양(+)의 영향력을 보이는 지역으로 서울 중심에 위치하고 있는 중구와 용산구 일대, 마포구, 강남구 등이 분석되었

다. 반면 강서구, 금천구, 도봉구 등의 시 외곽지역은 매출액에 음(-)의 영향력을 미치는 것으로 나타났다. 매출액에 대한 평균소득은 양(+)의 영향력이 가장 뚜렷한 지역은 강남구와 서초구, 동대문구, 광진구, 송파구 등 일대에서 높은 영향력을 보였다. 반면 중랑구와 중구 그리고 용산구의 일부지역에서 다소 낮은 음(-)의 영향력을 나타냈다. 다음으로 상가업소 수에 대한 영향력 분석에서는 양(+)영향력이 가장 뚜렷한 지역으로 동대문, 중랑구, 광진구, 동작구, 강남구의 일부지역, 양천구 일대에서의 집적효과가 지역상권 매출에게 높게 기여하고 있는 것으로 나타났다. 반면 용산구, 중구, 종로구, 구로구 등에서 음(-)의 영향을 나타냈으며 송파구, 관악구, 서대문구, 마포구 등에서는 양의 영향과 음의 영향이 함께 나타난 것을 알 수 있다. 또한 읍면동 단위 분석결과는 대체적으로 비슷한 경향을 나타냄을 알 수 있다. 그러나 소지역 단위 분석과 다르게 세부지역에 회귀계수 및 R Square 값이 합계되어 정확한 분석이 불가능했다. 특히 특정지역(소지역)의 값이 매우 높은 경우 읍면동으로 집계되어 분석단위 전체가 영향력이 높은 지역으로 측정되거나 반대로 분산·합계되어 낮은 지역으로 나타나는 결과를 확인할 수 있다. 이는 그간 GIS 및 공간정보 분야에서 분석수행 시 많이 발생하는 대표적인 문제라고 할 수 있다. 소지역 단위 분석에서는 이러한 오류를 최소화 할 수 있으며 동단위 분석에서는 전체적인 경향분석에 활용 할 수 있다. 반면 단점으로 세밀한 분석단위는 전체 경향을 보는데 방해가 될 수 있으며 동단위 분석에서는 집계로 인한 오류가 발생한다. 따라서 이러한 특징을 인지하고 분석목적에 부합하는 적절한 활용방안이 필요할 것으로 판단된다.

제4절 분석모형의 활용방안

본 연구의 분석과정과 결과 그리고 분석모형을 통해 학술적, 정책적, 기술적 측면으로 구분하여 다음과 같은 활용방안을 기대할 수 있다.

먼저 학술적 활용방안으로는 기존 지역상권 분석모형에서 다양한 요인을 고려하지 못한 점, 공간적 이질성과 가변적 공간단위 문제 등 그동안 지속적으로 지적되어온 자료구축과 활용에 대한 문제점을 최소화할 수 있는 방법과 가능성을 제공한다는 측면에서 큰 의의가 있다. 또한 상권 활성화 등과 관련된 학술연구를 보다 더 확장시켜 공간 빅데이터 처리 및 분석 모형을 적용하거나 응용함으로써 지역개발과 기타 관련 학술 분야에서 일종의 가이드라인 형태로 활용될 수 있을 것으로 기대된다.

다음으로 정책적 측면에서는 공간 빅데이터 처리를 통해 세부화된 지역의 모니터링체제로 활용가능하다. 공간 빅데이터의 세밀한 자료는 쇠퇴지역 및 정책 사각지역을 관리할 수 있도록 지원하거나 시계열 자료의 구축을 통해 새로운 패턴을 탐색할 수 있다. 이를 통해 정책지원의 세분화, 관리방안 등에 실질적으로 활용될 수 있다. 또한 공간정보 및 빅데이터 기반의 데이터 구축을 통해 다양한 분석이 가능한 기초DB로 활용가능 할 것으로 예상된다. 분석 측면에서는 분석의 파급성과 효과성, 시급성 등을 고려하여 포괄적인 분야의 데이터를 축적하고 시스템 및 플랫폼으로 개발하여 고도화된 서비스가 가능할 것으로 기대된다. 기술적 측면에서는 공간 빅데이터 기반 처리, 분석기법과 기존 사회과학분야에서 활용되는 통계분석 원리 등을 응용하여 범용적이고 확장성이 뛰어난 기술로 활용 될 것으로 기대된다. 또한 대용량 처리, 분석뿐만 아니라 공간분석의 원리를 기반으로 관련 분야의 응용기술 발전에도 기여될 것으로 기대된다.

제5장 결론

본 연구는 공공, 민간에서 구축·생성되는 카드매출정보, 유동인구 등 공간 빅데이터를 활용하여 지역상권에 관한 인구, 고용, 소득 등의 매출요인을 도출하고 이를 공간적으로 연계한 매출 영향력 분석모형을 설계·분석하고 그 결과물을 통해 다양한 활용방안을 제시하는데 목적으로 수행하였다. 이를 위해 본 연구에서는 서울특별시를 중심으로 공간 빅데이터를 기반 공간가중회귀모형(GWR)을 적용하여 기존 상권분석에서 도출하기 어려운 소지역 단위별 회귀계수와 국지적 R^2 를 추정하였다. 소지역 단위 분석은 지역상권의 매출액 영향요인의 공간적 이질성을 효과적으로 파악하고 소지역별 차이와 상권분석과 관련된 논의의 폭을 확장하고자 하는 시도이다. 또한 기존 지역상권 분석연구가 개별 대상지나 시군구, 읍면동과 같은 특정 지역을 중심으로 이루어진 문제와 대리변수, 집계 데이터의 활용에 따른 한계를 벗어나 세분화된 지역을 도출하고자 하였다. 아울러 다양한 데이터의 처리 및 분석을 통해 도출한 결과를 지도상에 시각화함으로써 서울특별시의 매출 요인의 영향력 수준을 파악하고 실제 상권의 세분화된 분포특성 등을 실증적으로 제시하였다. 실증분석 결과 매출액에 미치는 변수는 지역에 따라 상이하게 나타났으며 본 연구에서 선정된 변수 이외에도 제 3의 변수가 존재함을 알 수 있었다. 특히 읍면동 단위의 비교를 통해 기존 공간 데이터의 분석단위에 따라 발생하는 오류를 최소화할 수 있는 결과를 도출하였다. 그간 정통적인 상권분석과 GIS 분석이 가지고는 있는 한계를 극복할 수 있는 소지역 단위는 보다 세분화된 지역에서의 지역상권 매출 영향력을 체계적으로 분석할 수

있다. 이러한 분석모형은 대용량 데이터에 내포하고 있는 패턴을 발견하여 새로운 가치를 생성하는 가능성을 제공한다는 점에서 다양한 분야에서 활용될 것으로 기대된다.

이러한 이점에도 불구하고 본 연구의 한계는 먼저 다양한 변수와 분석방법을 고려하지 못한데 있다. 이는 세분화된 자료의 구축 가능성이 크게 작용한 것으로 이러한 데이터를 구축하기 위해서는 많은 시간과 노력이 필요하다. 아직까지는 공공 및 민간 일부기관에서 구축, 관리되고 있지만 최근 사회전반에서 정보 인프라와 관련 시스템이 확대되고 있는 시점에서 향후 다양한 기관에서 이러한 데이터를 추가하여 보다 많은 변수를 고려할 수 있을 것으로 예상된다. 또한 SNS와 같은 비정형 데이터와 실시간 데이터의 수집과 이를 처리·분석할 수 있는 방안이 필요할 것이다. 다음으로 주거 인구, 직장인구 데이터와 카드매출, 유동인구 등 시계열 데이터가 서로 상이하여 실제 시점에서의 공간현상을 파악하는 등 분석결과의 일반화에는 한계가 있다. 이는 앞서 언급된 데이터의 구축 가능성과 밀접한 관련이 있는데, 마이크로 데이터의 구축 가능성에 초점을 두어 수집했기 때문에 모든 데이터의 시점을 동일하게 하지는 못하였다. 하지만 분석절차 및 방법, 경향파악 및 분석단위의 비교 등에는 유의미한 해석이 가능함으로 향후 추가적인 연구를 통해 다양한 변수의 고려와 시점의 통일 같은 오류를 해소할 수 있을 것으로 판단된다.

본 연구는 기존 공간정보를 포함하고 있는 공간 빅데이터를 활용하여 대용량 데이터 내에 포함하고 있는 정보를 세분화하여 처리하고 그 결과를 시각화하여 공간적으로 연계한 시범적인 연구이다. 본 연구의 분석과정과 결과 등을 통해 향후 개발되는 다양한 공간 빅데이터 기술을 접목하여 새로운 분석시도가 지속되길 기대해본다.

참고문헌

국외문헌

- Anselin, L and Getis, A, 1992, Spatial statistical analysis and geographical information system, The Annals of Regional Science 26, 19-33
- Bailey, Trevor C. and Anthony C. Gatrell, 1995, Interactive Spatial Data Analysis, Longman Scientific & Technical
- Batty, M. 2012, Editorial: Smart Cities, big data, Environment and Planning B: Planning and Design, 39:191-193
- Ferber, R., 1958, Variation in Retail Sales between Cities, Journal of Marketing, 22(3), pp. 295-303
- Fotheringham, A.S., Brunson, C., and Charlton, M.E. 1998, Geographically Weighted Regression : a natural evolution of the expansion method for spatial data analysis, Environment and Planning A 30, pp. 1905-1927
- Gartner, 2012, Gartner IT Glossary
- IDC, 2011., Extracting Value from Chaos
- Ingene, C. and E. Yu, 1981, Determinants of Retail Sales in SMSAs, Regional Science and Urban Economics, 11(4), pp. 529-549
- kwat et al, 2010, what is Twitter, a social network or a news media? Proceedings of the 19th international conference on world wide web

- Liu, B. C., 1956, Determinants of Retail Sales in Large Metropolitan Areas, *Journal of the American Statistical Association*, 65(332) pp.1460-1473
- McKinsey, Company, 2013, Big data : What's your plan
- Mejia, Luis C. and John, D. Benjamin, 2002, What do we know about the determinants of shopping center sales? Spatial vs. Non-Spatial Factors, *Journal of Real Estate Literature* 10(1), pp. 1-26
- Shashi Shekhar, 2012, Spatial Big Data, AAG-NIH Symp on Enabling a National Geospatial cyberinfrastructure for Health Research
- Zukin, S. 2009, *Naked city: The death and die of authentic urban places*. 1st ed. Oxford University Press

국내 문헌

- 구자용, 2015, 공간정보 빅데이터의 지도화와 공간적 분포 특성에 관한 연구, 국토지리학회지 제49권 3호 pp.349-360
- 국토연구원, 2013, 공간 빅데이터로 똑똑하고 신뢰받는 정부 구현방안
- 국토연구원, 2013, 과학적 국토정책을 위한 공간 빅데이터 활용방안
- 김관호 외, 2013, 스마트카드 빅데이터를 이용한 서울시 지하철 이동 패턴 분석, 한국전자거래학회지, 제18권 3호 pp.211-222
- 김도현 외, 2014, 집합 대형할인점의 매출에 미치는 입지결정요인, 집합건물법학, 제13집, pp.123-152
- 김성문, 2014, 소매업태의 매출액 결정요인에 관한 연구, 건국대학교 박사학위논문
- 김진희 외, 2011, 공간가중회귀분석을 이용한 통행발생모형, 대한교통학회지 제29권 2호 pp.101-109
- 문소연, 2010, SSM 입점에 따른 슈퍼마켓의 피해범위 및 피해를 결정에 관한 연구, 건국대학교 석사학위논문
- 미래창조과학부, 2014, 실무자를 위한 빅데이터 업무절차 및 기술활용 매뉴얼
- 박완서 외, 2012, 근대기 서울의 공간구조와 상권의 변화에 관한 연구, 대한건축학회 논문집, 계획계 제28권 10호 pp.291-299
- 신우진 외, 2010, 기업형 슈퍼마켓의 상권 잠식 강도에 관한 연구, 국토계획 제 45권 7호 pp.207-218
- 안종욱 외, 2013, 공간 빅데이터 개념 및 체계 구축방안 연구, 한국공간정보학회지 제21권 5호 pp.43-51

- 유선철 외, 2015, 기후변화에 대응한 도시재생 지표개발 및
활성화지역 선정 연구 : 서울특별시 사례를 중심으로,
한국도시행정학회 제28권 4호 pp.77-99
- 이경주 외, 2015, 공간가중회귀모형을 이용한 강원도 지역상권 현황
진단 및 발전방안 연구, 한국은행 강릉본부
- 이기영, 2014, 트위터 기반 고객의 관심도 분석을 통한 마케팅 조언
시스템의 설계 및 구현, 한국인터넷방송통신학회지 제14권 3호
pp.185-190
- 이상민, 2013, 소매업 입지요인과 경영성과에 관한 연구 : 서울지역
나들가게 중심으로, 한성대학교 석사학위논문
- 이승주, 2015, 주성분 분석을 이용한 빅데이터 분석, 한국지능시스템학
회 논문지, 제25권 6호 pp.592-599
- 이영민 외, 2014, 공간통계분석기법을 이용한 소셜 네트워크 유력지
역 탐색기법 연구, 한국지형공간정보학회지 제22권 4호 pp.21-30
- 이우평, 2002, Basic 고교생을 위한 지리 용어사전
- 임동진, 2013, Big Data Analytic과 기업경쟁력
- 전상권, 2012, 이기종 미디어에서의 비정형 빅데이터 융합적 분석 연
구, 한세대학교 일반대학원 박사학위 논문
- 정대영 외, 2009, 공간자기상관기법을 이용한 근린상권의 공간특성분
석, 대한국토·도시계획학회지, 국토계획 제50권 5호 pp.215-231
- 정윤영 외, 2014, 유동인구 자료를 이용한 서울시 도시공간구조 분석
연구, 한국지역개발학회지 제26권 3호 pp.139-158
- 정은애 외, 2015, 공간자기상관성을 고려한 의류 소매업 매출액 영향
요인분석, 대한국토·도시계획학회지 국토계획, 제50권 5호 pp.215-231

- 중소기업청 시장경영지원센터, 2006, 대형마트 출점이 지역중소유통업에 미치는 영향
- 한국정보화진흥원 빅데이터 분석활용센터, 2012, 빅데이터로 진화하는 세상 - Big Data 글로벌 선진사례
- 한국정보화진흥원 빅데이터 분석활용센터, 2013, 더 나은 미래를 위한 데이터 분석 - Big Data 글로벌 선진사례Ⅱ
- 한국프랜차이즈협회, 2004, 프랜차이즈 경영가이드 총서
- 허자연 외, 2014, 소셜네트워크서비스와 상권 활성화에 관한 연구, 관광경영학회지, 제18권 4호 pp.517-534

홈페이지 및 웹사이트

- biz-gis 홈페이지 <http://www.biz-gis.com>
- 공공데이터포털, <http://www.data.go.kr>
- 국가통계포털, <http://www.kosis.kr>
- 네이버 지식백과, <http://terms.naver.com>
- 상권정보시스템 홈페이지 <http://sg.sbiz.or.kr>
- 우리마을가게 상권분석시스템 홈페이지 <http://golmok.seoul.go.kr>
- 통계청 홈페이지, <http://www.kostat.go.kr>

감사의 글

제가 무사히 석사생활을 마치기까지는 주변의 많은 도움의 손길이 있었기에 가능하였습니다. 이 짧은 지면으로나마 도움의 손길을 주신 분들께 감사의 마음을 전합니다.

먼저 언제나 연구에 대한 아낌없는 지도를 해주신 지도교수님이신 홍상기 교수님께 깊은 감사의 뜻을 전합니다. 연구에 대해서 아무것도 모르던 저를 교수님께서서는 연구방법뿐만 아니라 인생에 필요한 지혜 등 많은 것들을 가르쳐 주셨습니다. 또한 실수가 있어도 넓은 마음으로 받아주시고 연구자의 길을 몸소 가르쳐주신 신동빈 교수님 진심으로 감사드립니다. 덕분에 다양한 연구에 참여하여 많은 것들을 배울 수 있었습니다. 그리고 언제나 학업에 있어 따뜻한 조언과 격려 때로는 매서운 지적을 아끼지 않으신 안종욱 교수님 감사합니다. 학부시절부터 항상 깊은 관심과 응원을 아끼지 않으셨던 이미숙 교수님께도 감사드립니다.

이 외에도 대학원 수업으로 새로운 지식을 알려주신 손경환 교수님 감사드립니다. 학부시절 많은 도움을 주신 김주현 교수님, 장준호 교수님, 정일훈 교수님, 김성희 교수님, 신언교 교수님, 김동윤 교수님, 이제선 교수님께도 감사드립니다.

함께 일하는 스마트도시공간연구소 선·후배님들께도 많은 도움을 받았습니다. 특히, 같은 팀으로 함께 일하면서 어떤 업무에도 끝까지 믿어주시고 격려해주신 유선철 박사님, 최원욱 형님으로부터 정말 많은 것을 배울 수 있었습니다. 혹독한 트레이닝으로 제게 깨달음을 주신 김병선 박사님께도

감사드립니다. 진로에 대해 고민할 때 먼저 다가와 상담해 주신 경주형님, 날카로운 눈으로 제 부족함을 지적해주신 경희누님, 민경누님께도 감사합니다. 그리고 선배이자 동료인 지송이형, 준민이형 동기인 상은이, 다은이, 민우, 수정누나와 함께 지내면서 즐거운 학업, 연구소 생활을 할 수 있었고 그들의 연구 활동을 보며 많은 자극을 받을 수 있었습니다. 이 외에도 연구에 몰두할 수 있도록 행정 및 회계 관련 업무를 해주시는 헤미누나, 세라누나, 윤주에게도 많은 도움을 받았습니다. 모두 진심으로 감사드립니다.

“공간정보 융복합 핵심인재 양성사업”을 지원해주신 공간정보산업진흥원 관계자 분들에게도 감사의 말씀을 드립니다.

더하여 세상에서 하나밖에 없는 동생에게 고마움을 전합니다. 끝으로 언제나 변함없이 아들을 믿고 지원해주시는 아버지, 어머니께 감사의 인사를 드립니다. 자주 찾아뵙지 못하여 전화 음성으로 안부를 여쭙어 볼 때마다 오히려 저를 걱정해주시고 아낌없는 성원을 보내주신 그분들이 있었기에 저는 힘든 상황에 굴하지 않고 이겨낼 수 있었습니다. 정말 감사합니다.

이 분들의 도움이 더욱 빛나도록 앞으로도 최선을 다하겠습니다. 감사합니다.