



## 저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

博 士 學 位 論 文

통신 빅데이터 기반 공간상관성을 고려한  
통행발생모형 개발

嶺南大學校 大學院

都市工學科

交通工學 專攻

金 根 昱

指導教授 鄭 然 植

2019年 2月



博 士 學 位 論 文

통신 빅데이터 기반 공간상관성을 고려한  
통행발생모형 개발

指導教授 鄭 然 植

이 論文을 博士學位 論文으로 提出함.

2019年 2月

嶺南大學校 大學院

都市工學科 交通工學 專攻

金 根 昱



金根昱의 博士學位 論文을 認准함.

審査委員\_\_\_\_\_ (引)

審査委員\_\_\_\_\_ (引)

審査委員\_\_\_\_\_ (引)

審査委員\_\_\_\_\_ (引)

審査委員\_\_\_\_\_ (引)

2019年 2月

嶺南大學校 大學院



## 감사의 글

저에게도 이렇게 감사의 글을 쓸 기회가 주어졌습니다. 돌이켜보면 2006년부터였던 것 같습니다. 이 분야에 대해 전혀 알지도 못했고, 전혀 의도하지도 않았고 전혀 기대하지도 않은 연구실 생활에 발을 디디게 된 그 순간부터 2019년이 된 지금에 이르기까지 근 10여 년 동안 제가 이렇게 박사학위까지 공부하게 될 줄은 몰랐습니다. 그리고 이 학위를 계기로 앞으로도 계속 '교통'에서 일을 하며 보람을 찾고, 가족을 부양하며, 사회에 보탬이 되는 일을 조금 더 잘 해내고자 합니다.

논문을 쓰는 과정은 처음부터 끝까지 제가 그동안 얼마나 부족했는지를 알아가는 과정이었다고 해도 과언이 아닌 것 같습니다. 비록 부족하고 부끄러운 결과물이지만 이렇게 세상에 내보여서 조금이라도 후학을 위해, 사회를 위해 가치 있게 다루어진다면 그것만으로도 정말 감사한 일이 될 것 같습니다.

이 학위를 마치는 것을 포함하여 제가 지금까지 해왔던 연구원, 공무원, 엔지니어로서 만들어 온 작은 성과와 결과물들은 결코 저 혼자 힘으로 이루어진 것이 아님을 잘 알고 있습니다. 그래서 그동안 저를 지켜봐주신 교수님, 선배님, 동료들, 후배님들에게 더욱 더 감사한 마음이 큼니다.

먼저, 저를 학문의 길로 인도해주신 김갑수 교수님과 일면식도 없던 저를 기꺼이 받아주신 정연식 교수님께 정말 진심으로 감사드립니다. 그리고 논문이 완성되게끔 세심히 지도해주시고 마음 써주신 윤대식 교수님께도 감사의 말씀을 드립니다. 더불어, 부족한 제 논문을 성심성의껏 지도·편달해주신 이종달 교수님, 우용한 교수님과 정우현 박사님께도 깊은 감사의 인사를 드립니다.



감사드려야 할 분들이 너무나도 많습니다만, 이게 끝이 아님을 잘 알고 있기에 앞으로 살아가면서 도움 주신 모든 분들께 조금이나마 은혜를 갚으며, 받은 모든 도움 다시 보답할 수 있는 삶을 살겠습니다.

이제부터 다시 시작임을 잘 알고 있습니다. 제 경력에 아주 작은 점 하나를 찍고 다음의 걸음을 걷고자 합니다. 앞으로 조금 더 겸손하고 낮은 자세로 조금이라도 사회에 기여하고 세상이 나아지는데 있어 제가 이 분야에서 할 수 있는 일이 무엇인가에 대해 끊임없이 질문하며 살아가고자 합니다.

무엇보다도 아들이 가는 길을 묵묵히 지켜봐주시고 지지해주신 아버지, 어머니 낳아주시고 키워주셔서 감사합니다. 두 분께 배운 대로 행하는 사람이 되겠습니다.

못난 사위지만 항상 응원해주시고 챙겨주시는 장인어른, 장모님. 두 분 덕분에 제가 마음껏 공부할 수 있었습니다. 항상 건강하시고 감사합니다.

그리고 내가 제일 아끼는 우리 동생네와 처남네, 조카 연아, 현준, 태린이 귀여운 조카들 고맙고 사랑한다.

마지막으로 적지 않은 나이에 안정된 직장까지 그만두고 제가 하고 싶은 거 마음대로 하며 살게 해준 사랑하는 제 아내 신애주씨에게 진심으로 감사드리며, 아빠가 언제나 행복했으면 좋겠다고 얘기해주는 우리 아들 김주원에게 이 작은 성과를 바칩니다.

2019년 2월

김 근 욱



# 목 차

<b>제1장 서론</b>	<b>1</b>
1-1. 연구의 배경 및 목적	1
1-2. 연구의 범위	4
1-2-1. 시·공간적 범위	4
1-2-2. 내용적 범위	6
1-3. 연구의 수행절차	6
<b>제2장 이론 및 선행연구 검토</b>	<b>9</b>
2-1. 이론적 배경	9
2-1-1. 4단계 수요추정모형	9
2-1-2. 통행발생모형의 추정	12
2-1-3. 공간계량모형	15
2-2. 선행연구 검토	34
2-2-1. 통행발생모형 개선 관련 연구	34
2-2-2. 공간분석을 이용한 통행발생모형 관련 연구	38
2-3. 연구의 차별성	48
<b>제3장 자료의 구성 및 연구방법론</b>	<b>51</b>
3-1. 자료의 구성	51
3-1-1. 자료의 수집	51
3-2. 데이터 전처리 (Data Preprocessing)	53
3-3. 변수별 데이터 개념정의 및 구축	56
3-3-1. 종속변수	56
3-3-2. 설명변수	57



3-4. 연구방법론 정립 .....	64
3-4-1. 공간계량모형의 적용 .....	64
3-4-2. 연구의 진행과정 .....	68
3-4-3. 분석과정의 정립 .....	70
<b>제4장 통신 빅데이터 및 O/D 상관성 분석 .....</b>	<b>73</b>
4-1. 분석개요 .....	73
4-1-1. 비교 데이터 구성 .....	73
4-2. 분석결과 .....	74
4-2-1. 통행량 기준 통행목적별 상관성 분석 (행정동 기준) .....	75
4-2-2. 공간상관성 기준 통행목적별 상관성 분석 (행정동 기준) .....	96
4-3. 소결 .....	101
<b>제5장 공간회귀를 통한 통행발생모형 구축 .....</b>	<b>105</b>
5-1. 분석자료 .....	105
5-1-1. 변수의 선정 .....	105
5-1-2. 기초통계분석 .....	107
5-2. 통행발생 선형회귀(OLS) 기준모형 구축 .....	108
5-2-1. CDR 미적용 OLS 통행발생모형 .....	108
5-2-2. CDR 적용 OLS 통행발생모형 .....	120
5-3. 통행발생 지리적 가중회귀분석(GWR) 모형 도출 .....	136
5-3-1. 공간상관성 검증 .....	137
5-3-2. CDR 적용 GWR 통행발생모형 .....	140
5-4. 모형의 검증 .....	157
5-4-1. 모형 비교분석결과 .....	157
5-4-2. 데이터 오차검증 .....	159
5-4-3. 소결 .....	162





<b>제6장 결론 및 연구의 한계점 .....</b>	<b>165</b>
6-1. 결론 및 정책적 시사점 .....	165
6-2. 연구의 한계와 향후 연구과제 .....	167
 참고문헌 .....	 169
 Abstract .....	 179



## 표 목 차

<표 1-1> 자료의 구성 .....	5
<표 2-1> 통행발생모형 개선관련 선행연구 .....	37
<표 2-2> 공간계량분석을 이용한 통행발생모형 관련 선행연구 .....	39
<표 2-3> 국내 통행발생 모형적용 및 변수자료 .....	41
<표 2-4> 목적별 모형정산 결과(대구광역시권) .....	44
<표 2-5> 해외 도시별 통행발생모형 변수 .....	46
<표 3-1> 자료의 수집 .....	52
<표 3-2> 종속변수의 개념정의 .....	56
<표 3-3> 통신자료의 구성 .....	57
<표 3-4> 통신량 자료(CDR)의 개념적 정의 .....	59
<표 3-5> 사회경제지표 변수의 선정 .....	60
<표 3-6> 토지이용지표 변수의 개념 정립 및 선정 .....	62
<표 3-7> 물리적지표 변수의 개념 정립 및 선정 .....	63
<표 3-8> O/D 및 CDR 검증자료 구분 .....	69
<표 4-1> 데이터 비교분석 개념 정의 .....	74
<표 4-2> 행정동 기준 CDR과 O/D 총통행발생량 비교 .....	75
<표 4-3> 행정동 기준 CDR과 O/D 총통행유입량 비교 .....	81
<표 4-4> 기술통계량 .....	87
<표 4-5> CDR과 O/D 총통행발생·유입량 상관관계분석 종합 .....	87
<표 4-6> CDR 및 O/D 총통행발생·유입통행량 대응표본 검정결과 .....	88
<표 4-7> 기술통계량 .....	89



<표 4-8> CDR 평일 오전첨두시 필수적 통행 통행발생·유입량 상관관계분석 종합	90
<표 4-9> CDR 평일 오후첨두시 필수적 통행 통행발생·유입량 상관관계분석 종합	91
<표 4-10> 대응표본 검정	92
<표 4-11> 기술통계량	93
<표 4-12> CDR 주말오전첨두시 선택적 통행 통행발생량 상관관계분석 종합	93
<표 4-13> CDR 주말오후첨두시 선택적 통행 통행유입량 상관관계분석 종합	94
<표 4-14> 대응표본 검정	95
<표 4-15> 통행량 기준 결과종합	101
<표 4-16> 국지적 공간상관성 기준 결과종합	103
<표 5-1> 종속·설명변수의 선정	106
<표 5-2> 기초통계분석	107
<표 5-3> 모형요약 (O/D 총통행발생·유입모형)	108
<표 5-4> 회귀계수 추정결과 (O/D 총통행발생모형)	110
<표 5-5> 회귀계수 추정결과 (O/D 총통행유입모형)	112
<표 5-6> 모형요약 (O/D 필수적통행발생·유입모형)	112
<표 5-7> 회귀계수 추정결과 (O/D 필수적통행발생모형)	114
<표 5-8> 회귀계수 추정결과 (O/D 필수적통행유입모형)	116
<표 5-9> 모형요약 (O/D 선택적통행발생·유입모형)	117
<표 5-10> 회귀계수 추정결과 (O/D 선택적통행발생모형)	118
<표 5-11> 회귀계수 추정결과 (O/D 선택적통행유입모형)	119
<표 5-12> 모형요약 (CDR적용 OLS 총통행발생·유입모형)	121
<표 5-13> 회귀계수 추정결과 (CDR적용 O/D 총통행발생모형)	122
<표 5-14> 회귀계수 추정결과 (CDR적용 O/D 총통행유입모형)	123



<표 5-15> O/D total OLS기준모형의 오차의 공간자기상관 검증결과	124
<표 5-16> 모형요약 (CDR적용 O/D Mand PA)	126
<표 5-17> 회귀계수 추정결과 (CDR적용 O/D 필수적통행발생모형)	127
<표 5-18> 회귀계수 추정결과 (CDR적용 O/D 필수적통행유입모형)	129
<표 5-19> O/D Mand OLS기준모형의 오차의 공간자기상관 검증결과	130
<표 5-20> 모형요약 (CDR적용 O/D Opt PA)	131
<표 5-21> 회귀계수 추정결과 (CDR적용 O/D 선택적통행발생모형)	133
<표 5-22> 회귀계수 추정결과 (CDR적용 O/D 선택적통행유입모형)	134
<표 5-23> O/D Opt OLS기준모형의 오차의 공간자기상관 검증결과	135
<표 5-24> O/D total 종속변수의 전역적 공간상관성 검증	138
<표 5-25> O/D Mand 종속변수의 전역적 공간상관성 검증	139
<표 5-26> O/D Opt 종속변수의 전역적 공간상관성 검증	140
<표 5-27> 모형요약 (CDR적용 GWR 총통행발생 · 유입모형)	141
<표 5-28> GWR분석 회귀계수 추정결과 (CDR적용 O/D 총통행발생모형)	141
<표 5-29> GWR분석 회귀계수 추정결과 (CDR적용 O/D 총통행유입모형)	142
<표 5-30> O/D total GWR기준모형의 오차의 공간자기상관 검증결과	144
<표 5-31> 모형요약 (CDR적용 GWR 필수적통행발생 · 유입모형)	147
<표 5-32> GWR분석 회귀계수 추정결과 (CDR적용 O/D 필수적통행발생모형)	147
<표 5-33> GWR분석 회귀계수 추정결과 (CDR적용 O/D 필수적통행유입모형)	148
<표 5-34> O/D Mand GWR기준모형의 오차의 공간자기상관 검증결과	149
<표 5-35> 모형요약 (CDR적용 GWR 필수적통행발생 · 유입모형)	152
<표 5-36> GWR분석 회귀계수 추정결과 (CDR적용 O/D 선택적통행발생모형)	152
<표 5-37> GWR분석 회귀계수 추정결과 (CDR적용 O/D 선택적통행유입모형)	153



<표 5-38> O/D Opt GWR기준모형의 오차의 공간자기상관 검증결과 .....	155
<표 5-39> 모형간 설명력( $R^2$ ) 및 적합도(AICc) 비교 .....	158
<표 5-40> 모형간 오차의 공간상관성(Moran's I) 비교 .....	159
<표 5-41> 데이터 검증(RMSE, SRMSE) .....	160



## 그 립 목 차

<그림 1-1> 연구의 수행절차 .....	8
<그림 2-1> 4단계 교통수요 예측과정 .....	10
<그림 2-2> The classic four-stage transport model .....	11
<그림 2-3> 선형회귀의 일반적 예시 .....	12
<그림 2-4> 결정계수 적합성 척도 .....	13
<그림 2-5> 모집단에서 Y의 확률분포 .....	14
<그림 2-6> 공간자료에서 심슨의 역설이 나타나는 예(Fotheringham, 2002) ...	19
<그림 2-7> Moran's I값과 공간적 분포 개념도(김진희, 2011) .....	20
<그림 2-8> 모란 산점도의 예(정우현, 2011) .....	23
<그림 2-9> 공간인접성의 수량화(정건섭, 2011) .....	26
<그림 2-10> Rook방식 공간연속성 가중치형렬 구성의 예(Lesage, 1999) .....	26
<그림 3-1> 데이터 전처리의 형태 .....	54
<그림 3-2> ArcGIS 10.4, Spatial join .....	55
<그림 3-3> MSSQL Query .....	55
<그림 3-4> 통신자료의 구성(ArcGIS .shp) .....	58
<그림 3-5> Kernel function(Fotheringham et al., 2002) .....	67
<그림 3-6> 연구 진행과정 및 방법론 .....	71
<그림 4-1> CDR vs O/D Production 비교 (행정동 기준) .....	80
<그림 4-2> CDR vs O/D Attraction 비교 (행정동 기준) .....	86
<그림 4-3> CDR total & O/D PA 산점도 .....	88
<그림 4-4> CDR Mand Mpeak & O/D P A Mand 산점도 .....	90



<그림 4-5> CDR Mand Mpeak & O/D P A Mand 산점도 .....	91
<그림 4-6> CDR Mand Mpeak & O/D P A Mand 산점도 .....	94
<그림 4-7> CDR Mand Mpeak & O/D P A Mand 산점도 .....	95
<그림 4-8> CDR total & O/D PA 국지적 공간상관성 분석결과 .....	97
<그림 4-9> CDR Mand & O/D Mand PA 국지적 공간상관성 분석결과 .....	99
<그림 4-10> CDR Opt & O/D Opt PA 국지적 공간상관성 분석결과 .....	100
<그림 5-1> CDR total OLS기준 모형의 국지적인 공간자기상관 분석결과 ..	125
<그림 5-2> CDR Mand OLS기준 모형의 국지적인 공간자기상관 분석결과 ...	131
<그림 5-3> CDR Opt OLS기준 모형의 국지적인 공간자기상관 분석결과 ...	136
<그림 5-4> CDR total GWR모형의 Local R2 분포 .....	143
<그림 5-5> CDR total GWR기준 모형의 국지적인 공간자기상관 분석결과	145
<그림 5-6> O/D total GWR모형의 CDR Mand Morning 분포 .....	146
<그림 5-7> CDR Mand GWR모형의 Local R2 분포 .....	149
<그림 5-8> CDR Mand GWR기준 모형의 국지적인 공간자기상관 분석결과 ...	150
<그림 5-9> CDR Mand GWR모형의 Local R2 분포 .....	151
<그림 5-10> CDR Opt GWR모형의 Local R2 분포 .....	154
<그림 5-11> CDR Opt GWR기준 모형의 국지적인 공간자기상관 분석결과 ...	155
<그림 5-12> CDR Opt GWR모형의 Local R2 분포 .....	156
<그림 5-13> 모형별 관측치와 예측치의 비교 그래프 .....	162







# 제1장 서론

## 1-1. 연구의 배경 및 목적

지난 수십여 년에 걸쳐 교통계획 분야에서는 교통수요예측을 위해 전통적 4단계 수요추정기법인 통행발생(Trip generation), 통행분포(Trip distribution), 수단선택(Modal choice), 통행배정(Trip assignment)의 순차적 단계수요추정 방법을 지속적으로 발전시켜왔다. 이 기법은 분석방법론의 이해가 비교적 용이하고 이를 대체할 만한 이론적 토대가 정립되어 있지 못한 등의 이유로 교통수요분석의 정설(定說)로서 광범위하게 이용되고 있다.

그러나 이 방법은 각 단계별로 구분된 교통수요가 상호 독립적인 선택으로 간주되어 순차적인 과정을 거쳐 예측됨으로 인하여 각 단계에서 예측을 위해 사용되는 파라미터와 변수 값 등이 각 단계 간에 일치하지 않는 문제점이 있다(윤대식, 2001). 또한 순차적으로 앞선 단계의 추정결과(output)가 다음 단계의 투입요소(input)가 되는 과정을 가짐에 따라 이전 단계에서 발생한 오차 역시 다음 단계로 누적되어 전이(error propagation)되는 문제가 발생되어 최종 결과치의 과대·과소 수요추정에 영향을 주는 등 방법론상의 근본적인 한계를 가지고 있다.

구체적으로 언급되고 있는 수요예측 실패 사례로 박경아 외(2010)는 민간투자사업인 인천공항고속도로와 천안-논산고속도로, 대구-포항고속도로, 우면산 터널 사업 등을 대상으로 수요 추정 문제를 분석하였는데 이때 장래 교통수요 예측치가 대부분 기준연도의 실제 통행량에 비해 20~50%까지 과대추정(overestimation)된 문제를 보이고 있다고 지적하였다. 이와 같은 교통 수요추



정의 불확실성 문제는 수요 예측시 관련 개발계획의 반영 문제, 사회적 여건 변화와 같은 외부적 요인도 있을 수 있으나 근본적으로 교통 수요예측을 위한 기초데이터(origin-destination: O/D)의 정확성 및 4단계 수요예측 방법론이 가진 문제점 등의 내재적 요인이 그 원인으로 크게 지적받아오고 있는 상황이다.

최근 들어 교통계획 분야에서 ‘빅 데이터(Big data)<sup>1)</sup>’ 개념의 대용량 오픈 소스(Open source)<sup>2)</sup> 데이터를 활용하여 이러한 수요분석 분야의 단점을 개선시키려는 연구가 진행되고 있다. 그러나 아직까지 빅 데이터는 활용 목적별로 사용될 데이터의 명확한 정립이나 개인정보 보호 등의 법·제도적 정비 마련이 미흡한 부분이 있다. 일례로 국내·외 빅 데이터를 이용한 수요예측 관련 연구에서는 대체적으로 모바일(mobile) 데이터를 이용하여 통신 자료의 특정 좌표와 체류시간 가정을 통해 집과 근무지 간의 통근통행 등과 관련한 기종점(O/D)을 추정하는 연구가 주로 진행되고 있다(김중학 외, 2015; Šćepanović, 2015; Xian 외, 2017).

이처럼 빅 데이터를 활용하여 교통수요예측 방법을 개선시키기 위한 연구는 앞서 언급한 내재적 요인 중 기초데이터(O/D)의 신뢰성 확보를 위한 차원에 중점을 두어 진행되고 있으며, 향후 장래 수요예측 분야로의 광범위한 활용을 위해서는 더 많은 연구가 필요하다고 판단된다. 빅 데이터를 활용한 기초데이터 정확성 확보를 위한 연구 이외에도 수요추정 방법론 개선을 위한

- 
- 1) 일반적으로 사용되는 데이터 수집, 관리 및 처리 소프트웨어의 수용 한계를 초과하는 분량의 데이터를 말하며, 양(volume), 생성 속도(velocity), 다양성(variety), 가치(value), 복잡성(complexity)의 특성으로 요약된다.
  - 2) 소프트웨어 혹은 하드웨어의 제작자의 권리를 지키면서 원시코드를 누구나 열람할 수 있도록 한 소프트웨어 혹은 오픈 소스 라이선스에 준하는 모든 통칭을 일컫는다. (출처 : 위키백과)



측면에서 최근 공간분석기법을 통해 모형의 추정력을 개선시키고자 하는 연구도 많이 진행되고 있다. 대표적으로 정우현(2011)은 서울시 통근통행생성·유입량을 대상으로 통행발생모형의 집계오차(aggregation error) 문제를 개선하기 위해 공간상관성(spatial correlation) 개념을 통행발생량 추정에 적용한 공간계량분석(spatial econometrics model) 기법을 이용하여 통행발생모형을 도출하고 기존 모형과의 비교·분석을 통한 추정력 개선에 관한 연구를 진행하였다. 그러나 해당 연구는 다양한 공간계량분석기법을 모두 적용하여 모형의 추정력 개선 여부를 살펴본 연구로서 본 연구에서는 이 중 국지적 회귀모형식의 도출이 가능하고 공간종속성 및 이질성을 모두 고려할 수 있는 지리적 가중 회귀분석(Geographically Weighted Regression: GWR) 방법을 주로 이용하여 통행발생모형을 구축해보고자 한다.

이러한 배경들 하에서 교통 수요추정의 정확도 향상을 위한 거시적 목적을 달성하기 위한 연구의 일환으로 기존 수요예측 단계에서 사용되지 않았던 모바일 데이터라는 새로운 빅 데이터 변수를 통행발생모형 단계에 적용해 보고자한다. 또한 모형의 추정력 개선을 위해 통행의 공간적 상호작용과 집계모형으로서 가지는 통행발생모형의 한계를 고려한 공간상관성 분석 기법을 이용하여 수요분석방법론 개선을 그 목적으로 한다.

이를 위한 본 연구의 세부적인 목적은 다음과 같다.

첫째, 통행발생모형을 추정함에 있어 기존 방식에서 설명변수로 사용되고 있는 인구, 종사자수 등과 같은 사회경제지표 이외에도 용도지역별 건축물연면적, 토지이용혼합도 등의 토지이용변수와 유형별 교통유발시설, 버스(지하철)정류장수와 같은 물리적 변수를 고려한 통행발생 기준모형 구축으로 통행발생의 보다 구체적인 영향요인에 대해 알아보하고자 한다.



둘째, 통행발생모형의 주요 변수로 모바일 데이터를 활용함에 있어 실제 개인의 모바일 통신사용이 ‘통행’의 개념을 내포하고 있는지에 대해 O/D와 통신량 자료(Call Detail Record: CDR) 간의 상관관계를 분석해봄으로서 그 여부를 살펴보고자 한다.

셋째, 이를 바탕으로 통행발생모형의 주요 변수로 CDR을 모형의 설명변수로 적용시켜보고 대안별 모형의 추정력 개선 여부를 비교·분석한다.

넷째, 기존 선형회귀분석이 반영하지 못하는 공간상관성 문제를 고려하기 위하여 O/D 통행발생·유입량의 공간적 상관성을 우선 파악해보고 CDR을 설명변수로 적용한 공간상관성이 감안된 지리적 가중회귀분석(GWR)의 통행발생모형을 구축하여 기존 선형회귀분석과의 비교·분석을 통해 모형의 개선 여부를 파악해보고자 한다.

## 1-2. 연구의 범위

### 1-2-1. 시·공간적 범위

본 연구의 공간적 범위는 대구광역시 139개의 행정동 단위의 소존을 대상으로 하였다. 종속변수로 사용된 O/D자료의 경우 2016년도를 기준하여 가정, 비가정기반 통근·통학·쇼핑·업무·기타통행 자료의 통행발생(유입)량을 통행목적별로 성격을 구분하여 집계화한 데이터를 사용하였다.

시간적 범위는 O/D 통행발생(유입)량 자료의 경우 2016년을 기준으로 하였고 본 연구에서 주요 설명변수로 다루어질 CDR은 2017년 4월 8~9일(주말)과 2017년 4월 12~13일(평일)의 2일간 자료로서 2017년도를 기준으로 한 데이터를 사용하였다. 이 CDR 자료는 O/D 통행발생(유입)량과 직접 비교·분석을



위하여 해당 통신사 점유율을 바탕으로 1일 단위로 변환하는 데이터 핸들링 과정을 거쳤다. 여타 다른 설명변수 자료들 경우에는 가능한 한 종속변수와 시간적 범위를 동일시하는 자료를 취득하려 노력하였으나, 통계자료의 배포 및 수집연도가 상이한 측면이 있어 연구의 수행 시점에서 구득이 가능한 최신연도의 자료를 활용하였다. 본 연구에서 사용될 자료와 관련하여 세부적인 데이터의 구성과 전처리(preprocessing) 과정 및 데이터들의 기술통계 분석은 제3장에서 상세히 제시하였다.

<표 1-1> 자료의 구성

구 분	자료명	내 용	기준 연도	공간 범위	출처
종속 변수	기종점 통행량 (O/D)	<ul style="list-style-type: none"> <li>· 대구광역시 기준연도O/D</li> <li>- 통행목적별 O/D자료(P/A기반)</li> <li>- 가정기반 통근·통학</li> <li>- 비가정기반 쇼핑·여가·업무·기타</li> </ul>	2016	대구시	KTDB
설명 변수	통신량 자료 (CDR)	<ul style="list-style-type: none"> <li>· 모바일 통신량 자료(CDR)</li> <li>- 평일 오전(07~09), 오후침두(17~19)</li> <li>- 주말 오전(09~12), 오후침두(14~16)</li> </ul>	2017	대구시	SKT
	사회 경제 지표	<ul style="list-style-type: none"> <li>· 총인구수, 취업자수, 총종사자수, 3차산업종사자수, 학원종사자수, 초중고학생수, 연령대별 인구수 등</li> </ul>	2015	행정동	‘16년 교통관련 기초조사 (Ⅱ)
	토지 이용 지표	<ul style="list-style-type: none"> <li>· 용도지역별 건축물연상면적</li> <li>- 주거·상업·공업지역별</li> <li>· 토지이용복합도(Land Use Mix Diversity)</li> <li>· 직주균형계수(Job-Housing Balance Index)</li> </ul>	2016	대구시	공공 데이터 포털
	물리적 지표	<ul style="list-style-type: none"> <li>· 유형별 교통유발시설</li> <li>- 판매, 의료, 교통, 교육, 문화관람집회, 행정기관</li> <li>· 버스정류장수, 철도(지하철)역수</li> </ul>	2015	대구시	KTDB



### 1-2-2. 내용적 범위

우선 본 연구에서는 통행발생모형의 정확성과 신뢰성을 개선시키기 위하여 CDR을 적용하기에 앞서 해당 데이터가 가진 특성을 우선 파악하고, 기존 O/D 통행발생·유입량과 공간적 상관관계의 비교·분석을 통해 자료의 활용적 측면에서 의의를 먼저 살펴보고자 한다. 다음으로는 이러한 CDR을 통행발생모형의 설명변수로 적용해보고 이를 미적용한 통행발생·유입모형과의 비교·분석을 통해 CDR의 적용시 모형의 개선여부를 파악하고자 한다.

따라서 향후 이러한 모바일 관련 빅 데이터를 기존 통행발생모형의 개선을 위한 활용적 측면에서 가능성을 제시하고자 한다. 또한 통행발생모형에 영향을 주는 요인으로 CDR의 적용뿐만 아니라 선행 연구에서 검토된 설명변수와 다양한 변수들을 적용시켜 보다 실제 상황의 정확한 모사가 가능한 통행발생 영향요인을 알아보는 연구도 진행하였다. 마지막으로 통행발생모형을 구축함에 있어 공간자기상관의 고려가 가능한 공간회귀분석 중에서 국지적으로 회귀계수의 추정이 가능한 지리적 가중회귀분석(GWR) 방법을 이용하여 기존 선형회귀분석(OLS)을 통한 통행발생모형과의 비교·분석으로 모형의 추정력 개선과 관련한 결과를 살펴보고자 한다.

### 1-3. 연구의 수행절차

본 연구는 앞서 언급한 연구의 배경에 따라 기존 통행발생모형 및 공간계량모형 분석방법론에 대한 이론적 고찰을 실시하고 이를 바탕으로 선행연구에서 통행발생모형의 개선을 위한 다양한 설명변수의 적용 및 방법론 개선 연구 사례들을 검토한다. 이를 통해 기존 선행 연구들로부터 본 연구의 차별



성에 대해 서술하는 과정을 거치고자 한다.

다음으로는 분석에 사용되는 빅 데이터인 CDR을 본 연구의 목적에 부합되어 분석이 가능하도록 MSSQL과 ArcMAP 10.4 Spatial Join을 활용하여 데이터 마이닝(data mining) 및 데이터 전처리 과정(data preprocessing)을 진행한다. 종속 및 설명변수들의 기술통계량과 상관관계 및 회귀분석은 SPSS Statistics 23 및 STATA를 사용하고 공간상관성 분석은 GeoDa, ArcGIS Spatial statistics tools, 지리적 가중회귀분석(GWR)은 GWR4.0 프로그램을 이용하여 분석을 시행하였다. 그리고 설명변수로 적용된 CDR의 적용 및 미적용 통행발생모형 간의 비교·분석과 선형회귀(OLS) 및 지리적 가중회귀분석(GWR)을 통해 통행발생모형 간 모형의 설명력( $R^2$ ,  $adjR^2$ ) 및 적합도(AICc), 데이터 검증(RMSE, SRMSE)을 통해 비교·분석결과를 제시하였다.

요약하자면 제1장에서는 서론 부분으로서 연구의 배경 및 목적, 연구의 범위에 대해 서술하고 제2장에서는 관련 이론적 고찰 및 선행 연구의 검토를 통한 연구의 차별성을 도출한다. 제3장에서는 자료의 구성 및 연구방법론의 세부적인 이론적 고찰을 제시하고 데이터 전처리 과정에 대해 기술한다. 제4장에서는 통행발생모형의 정확성과 신뢰성을 개선시키기 위해 CDR을 적용함에 있어 해당 데이터가 가진 특성을 파악하고 기존 O/D 통행발생(유입)량과 공간적 상관관계의 비교·분석을 통해 자료의 활용적 측면에서 의의를 살펴본다. 제5장에서는 통행발생모형을 구축함에 있어 CDR을 설명변수로 미적용한 것과 적용한 모형 간의 비교를 통해 모바일 빅 데이터를 이용한 모형의 추정력 개선에 관한 연구를 진행하고 기존 통행발생모형에서 다루어지지 못한 공간상관성을 감안한 지리적 가중회귀(GWR)분석을 통한 통행발생모형과의 비교를 통해 모형의 개선여부와 데이터 검증 분석을 실시한다. 마지막으로





로 제6장에서는 설명변수로 CDR이 적용된 선형회귀(OLS) 기준 통행발생모형과 지리적 가중회귀분석(GWR) 기준 통행발생모형 비교 분석에 관한 주요 결과를 요약하고 연구의 정책적 시사점 및 향후 연구과제에 대해 기술한다.



<그림 1-1> 연구의 수행절차





## 제2장 이론 및 선행연구 검토

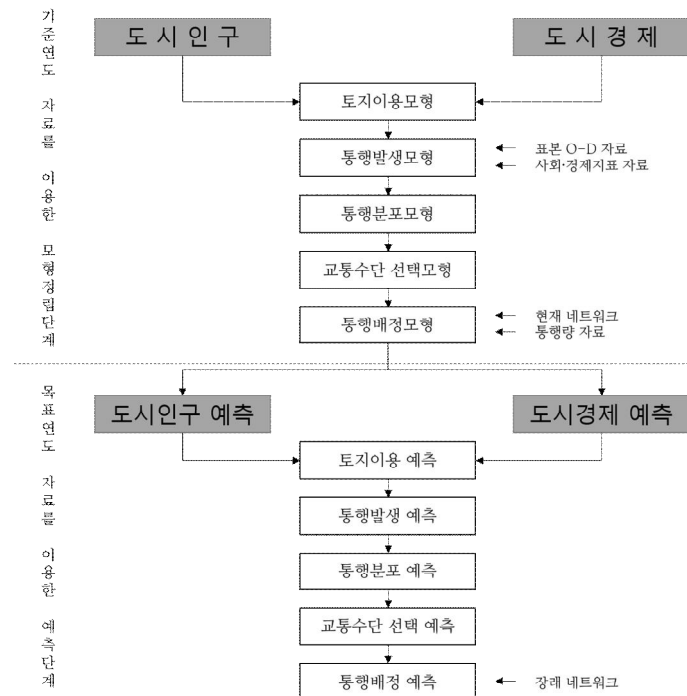
### 2-1. 이론적 배경

#### 2-1-1. 4단계 수요추정모형

교통수요를 예측하는 일반적이고 대중적인 4단계 교통수요 예측모형은 통행발생(Trip generation), 통행분포(Trip distribution), 수단선택(Modal choice), 배정(Assignment)으로 이루어진다. 이러한 전통적 기법 이외에도 통행수요를 예측하는 데 있어 개별통행자의 활동을 비집계화(Disaggregate)된 자료를 바탕으로 모형화함으로써 통행행태를 추정하는 활동기반모형(Activity-based model)이 있지만 본 연구에서는 이 두 가지 방식 중 전통적 4단계 교통수요 예측기법 중의 통행발생모형에 중점을 두어 연구하였다.

Weiner(1997)에 의하면 4단계 모형을 실제 교통계획에 최초로 적용한 것은 1962년 CATS(Chicago Area Transportation Study)로 알려져 있다(정우현, 2011). 4단계 교통수요예측 기법은 통행자의 통행선택 행위를 순차적인 과정을 가정하여 모형화하는 것으로 통행발생에서는 통행여부의 선택, 통행분포에서는 목적지의 선택, 수단선택에서는 통행수단의 선택, 노선배정에서는 통행경로의 선택 4단계로 구성된다. 각 단계별 추정결과(output)는 다음 단계의 추정을 위한 투입요소(input)가 되는데, 이러한 교통수요 예측과정은 아래 그림2-1과 같다.





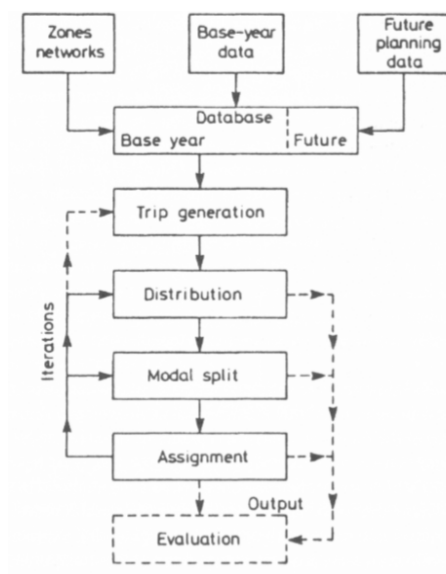
자료 : 교통계획(교통수요 이론과 모형), 윤대식, 2018, 박영사

<그림 2-1> 4단계 교통수요 예측과정

Ortúzar and Willumsen(2011)은 전통적 수요예측모형을 아래 그림2-2와 같이 설명하고 있다. 이 접근방식은 존(zone)의 설정 및 네트워크 체계, 데이터의 수집과 코딩(Coding), 정산(Calibration), 검증(Validation)로부터 시작된다고 하였다. 이러한 데이터들은 연구 대상범위(Study area)내 각 존으로부터 기준연도의 인구뿐만 아니라 사회경제적인 지표 즉, 고용, 상업지역, 교육 및 여가시설 등의 자료를 포함하여야 한다. 통행발생단계에서 이러한 데이터들은 결국 연구대상 범위 내 각 존에 의한 통행발생과 집중의 총량을 추정하는 자료로 사용된다. 다음 단계는 특정한 목적지를 향하는 통행을 공간상에 할당하는



통행배분(Trip distribution)으로서 Trip matrix를 생성하는 과정이다. 수단 선택 (Modal split)은 수단분담 유추하여 다양한 수단별로 Matrix 상에 통행을 할당 하는 과정이다. 마지막 단계는 통행배분(Assignment) 단계로 각 수단별 통행을 네트워크에 배정하게 된다(Ortúzar et al., 2011).



자료 : Modeling Transport (4<sup>th</sup> edition), Juan de Dios Ortuzar et al, 2011

<그림 2-2> The classic four-stage transport model

본 연구에서 중점적으로 다루어질 통행발생모형은 전통적 기법의 4단계 교통수요예측 과정의 첫 번째 단계로서 도시 전역 또는 도시 내에서 구획된 교통존(Zone)에서 발생하는 사람 또는 차량의 통행량을 예측한다. 통행발생은 보통 1일 통행량을 통행의 목적별로 추정하며 통행유출과 통행유입으로 구분 된다(윤대식, 2018).

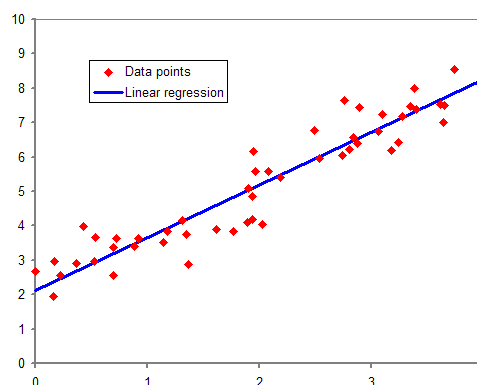


Williams(1977)는 전통적 4단계 추정모형의 문제점에 대해 일반적으로 통행 결정(Travel Decision)은 실제 이러한 순차적인 유형에 의해 발생되지 않는다고 보았다. 단계별 수요추정에서 각 하위 모형(Sub-model)의 ‘위치(Location)’가 앞선 모든 통행에 영향을 주는 것으로 각 단계별 모형에서 가정된 효용함수의 형태에 의존하는 것은 문제가 있다고 주장했다. 또한, 이러한 4단계 모형은 단지 통행자의 응답에 의한 제한적인 범위 내에서 추정되는 점도 문제점이라고 판단했다.

## 2-1-2. 통행발생모형의 추정

### (1) 선형회귀분석 개념

통행발생모형에서 사용하고 있는 선형회귀분석은 통행량과 통행에 영향을 줄 수 있는 영향요인 변수들 간의 인과관계를 하나의 방정식 형태로 작성하고, 일반화최소자승법(Ordinary Least Square: OLS) 등의 회귀분석기법을 이용하여 이들 변수들 간의 연관정도를 규명함으로써 얻어진 모형을 의미한다(노정현, 1999).



<그림 2-3> 선형회귀의 일반적 예시



회귀분석을 위한 인과관계의 함수적 표현은 회귀모형으로 나타나는데 일반적으로 통행발생을 나타낼 때 여러 가지의 설명변수를 통한 회귀식을 사용하기 때문에 여기에서는 다중선형회귀모형을 기준으로 하여 설명한다.

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki} + e_i \quad \dots\dots\dots (식2.1)$$

$(i = 1, 2, \dots, n)$

여기에서  $e_i$ 는 오차항으로서 평균이 0, 분산이  $\sigma^2$ 인 정규분포를 따르는 서로 독립인 확률변수이다. 즉, 일반적인 선형회귀분석은 다음과 같은 몇 가지 가정을 내포하고 있다.

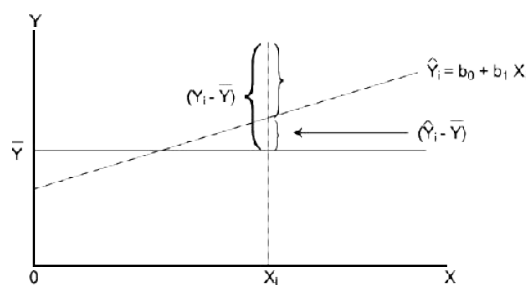
첫째, 회귀분석은 설명변수와 종속변수 간 일정한 크기로 함께 변화하는 선형성을 가지고 있어야 하는 선형성 조건이다.

둘째, 종속변수들은 서로 독립적이어야 한다는 잔차의 독립성을 전제로 한다.

셋째, 설명변수의 모든 값들에 대해 오차들의 분산이 일정한 등분산성을 가지고 있어야 한다.

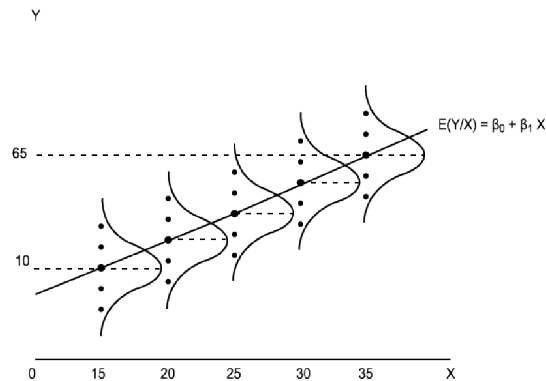
넷째, 관측치들의 잔차끼리는 상관이 없어야 하는 비상관성을 가져야 한다.

다섯째, 잔차 항이 정규분포를 이루어야 한다는 가정을 기본으로 한다.



<그림 2-4> 결정계수 적합성 척도





<그림 2-5> 모집단에서 Y의 확률분포

## (2) 통행발생모형에서의 선형회귀분석의 이용

통행발생모형의 구축은 일반적으로 원단위법, 교차분류분석법, 회귀분석법, 성장률법 등이 주로 사용되고 있다. 이 중 가장 광범위하게 이용되고 있는 방법은 회귀분석법으로서 종속변수인 통행량과 설명변수인 사회경제지표 간의 선형관계를 기반으로 하여 도출한 모형을 의미한다.

McNally(2000)는 통행발생모형의 개념으로 ‘가구 또는 존의 일일 총 통행의 양을 통행목적(또는 활동)별로 정의하는 것’이라 하였다. 이는 각 분석단위  $i$  와 통행목적  $p$  의 통행생성량( $P_i^p$ )과 통행유인량( $A_i^p$ )을 다음과 같이 정의하는 것으로 통행생성량과 유인량은 활동시스템 특성을 변수로 하는 모형이라는 것이라고 하였다. 여기서 활동시스템 특성(Activity-system characteristics)이란 분석 존의 사회경제적, 인구학적, 토지이용 데이터를 의미한다(정우현, 2011).

$$P_i^p = f_p^p(\text{activity-system characteristics } P) \quad \dots\dots\dots (\text{식2.2})$$

$$A_i^p = f_A^p(\text{activity-system characteristics } A)$$



정우현(2011)의 연구에 의하면, 회귀분석을 기반으로 하는 통행발생모형은 식2-3, 2-4와 같이 통행생성모형과 통행유인모형을 별도로 구축하는 것이 일반적이라고 설명하고 있다. 다음의 식2-3, 2-4는 통행생성과 유인량을 추정하는 회귀모형의 일반식이다. 여기에서  $a$ 와  $b$ 는 회귀계수를 의미하며,  $X$ 는 설명변수로서 각종 사회경제지표 등이 사용된다.

$$P_i = a_0 + \sum_k a_k X_{ik} \quad \dots\dots\dots (식2.3)$$

$$A_i = b_0 + \sum_k b_k X_{ik} \quad \dots\dots\dots (식2.4)$$

통행발생모형에서의 회귀분석 역시 일반적으로 계량모형에서 이용하는 동일한 방식의 기법과 추정식이 이용되는 것으로 회귀방정식이 가진 다중공선성의 문제나 최소자승법, 오차의 정규성 및 등분산성 등의 가정을 그대로 따르는 것이 일반적이다.

## 2-1-3. 공간계량모형

### (1) 이론적 배경

이성우 외(2006)는 공간계량모형의 이론적 배경과 관련하여 ‘인간의 모든 활동은 언제나 시간과 공간이라는 두 가지 요소를 포함하여 일어나게 되며, 사회과학의 모든 분야는 궁극적으로 시간과 공간이 만든 여러 가지 복합적 요인들이 인간의 행동과 어떠한 영향을 주고받는 지를 규명하고 객관화하는 것을 그 목표로 한다.’고 하였다. 계량경제학적 방법에서는 이 두 요소를 포함할 수 있는 여러 가지 모형들에 대한 연구가 지속적으로 이루어져 왔는데



시간의 경우에 있어서는 시계열 분석(Time Series Analysis)이 이를 직접적으로 다루고 있는 영역이며, 또한 모든 자료는 특정 시간에 이루어진 결과를 설명하고 있기 때문에 횡단면 자료(Cross Sectional Data)를 이용한 연구까지도 시간을 포함하고 있다고 볼 수 있다. 계량경제학에서는 시간에 대한 개념을 고려할 경우를 시계열 모형이라고 하며, 공간적인 요소들에 대해서 고려하는 기법을 공간계량모형이라고 한다. 즉, 공간계량모형은 인간의 활동에 의한 사회경제적 현상을 모형화함에 있어 현상들이 공간적으로 상관관계가 존재함을 경우를 통계적으로 표현하는 모형이라고 할 수 있다.

이론적으로 공간적 상관관계는 공간 종속성(Spatial Dependency)과 공간 이질성(Spatial Heterogeneity)의 두 가지 개념으로 나뉜다. 이 개념을 기반으로 공간적 상관관계는 기존 선형회귀 모형을 보정하는 형태로 이루어진 여러 가지 모형들 중에서 가장 기본이 되는 3~4가지 모형 즉, 공간시차모형(SLM; Spatial Lag model), 공간오차모형(SEM; Spatial Error Model), 일반공간모형(General Spatial Model), 지리적 가중회귀모형(Geographically Weighted Regression)으로 구성된다(정우현, 2011).

## **(2) 공간상관성 (spatial correlation)**

### **1) 공간종속성 (spatial dependency)**

Lesage(1999)와 이성우 외(2006)에서 정의한 공간 종속성(Spatial dependency)이란 종속변수끼리 상관관계를 가지게 되는 경우를 의미하며, Tobler(1970)의 지리학 제1법칙인 ‘모든 것은 그 밖에 다른 모든 것과 관련되어 있지만, 인접해 있는 것들이 멀리 있는 것들보다는 더 높은 관련성을 보인다.’라는 명제에서 가장 잘 설명되어지고 있다. Fotheringham 외(2002)는 공간종속성이 한 지





역의 속성이 인접한 다른 지역의 속성에 영향을 받는 것이라고 정의한 바 있다. 즉,  $i$  지역의 변수  $y_i$ 가 다음의 관계를 가지고 있음을 뜻한다.

$$y_i = f(y_j) \quad \text{단, } i = 1, 2, \dots, n \quad i \neq j \quad \dots\dots\dots \text{(식2.5)}$$

따라서, 식2.5는 한 지역의 변수가 다른 지역의 변수에 의해 관계가 규명되는 형태로서 종속변수가 설명변수처럼 작용하고 있다는 것을 의미한다. 정우현(2011)은 이를 통행생성모형(trip production model)에 대입할 경우 다음과 같은 가설을 세울 수 있다고 주장하였다.

- 가설 : 존  $i$ 의 통행생성량  $P_i$ 는 인접한 다른 존  $j$ 의 통행생성량  $P_j$ 와 상관관계가 있다. ( $P_i = f(P_j), i \neq j$ )

Lesage(1998)에 의하면 공간종속성의 발생원인은 크게 두 가지로 구분할 수 있다. 첫째, 자료수집의 공간단위에 의해 발생한다는 것이다. 연구자가 사용하게 될 자료의 지역 간 단위는 흔히 행정구역별 또는 경도-위도 좌표 등으로 구성되나 이러한 지역 간 구분이 수행될 연구의 바람직한 구분 단위인지는 확인하기가 쉽지 않다. 따라서 수집된 자료의 지역 간 단위가 수집된 지역과 일치하지 않는 경우에 발생된다. 둘째, 인구사회학적인 요인에 의해서 발생되는데 인간의 행위는 특정 공간에 고정된 채 이루어지는 것이 아니라 공간들을 이동하면서 이루어지는 행위가 대부분이다. 이러한 인구·사회학적 요인은 공간 종속성 발생의 가장 중요한 이유인 동시에 필연적인 성격을 지니고 있다는 점에서 매우 중요한 개념으로 다루어질 필요가 있다.



## 2) 공간이질성 (spatial heterogeneity)

공간이질성(spatial heterogeneity)은 공간이 종속변수의 결정에 미치는 영향이 균일하게 나타나지 않음을 뜻한다(이성우외, 2006). 따라서 다음과 같은 관계가 형성된다.

$$y_i = f_i(X_i\beta_i + \epsilon_i) \quad \dots\dots\dots (식2.6)$$

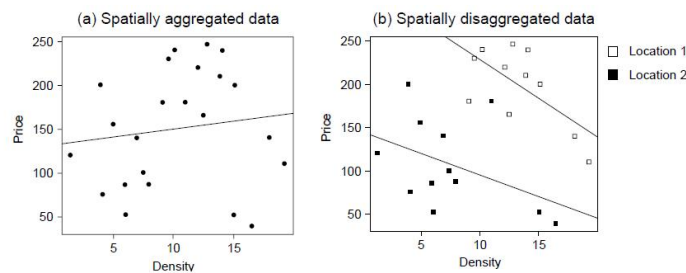
이 모형식은 일반 선형회귀식인  $y_i = X_i\beta_i + \epsilon_i$ 와 달리 일반선형회귀식의 또 다른 함수관계를 형성하고 있음을 뜻한다. 즉, 매우 유사한 설명변수를 가진 지역이라 할지라도 지역에 따라 다른 종속변수의 값을 가지게 됨을 뜻한다.

김광구(2003)는 공간이질성의 사례로서 중심지와 배후지의 존재, 선도 및 낙후지역(leading and lagging regions)의 존재 등 동종 제조업체들의 특정 지역에 집중되어 분포되어 있는 경우와 이종의 업체들도 특정 지역에만 집중되는 경우 공간이질성이 있는 것으로 보았다. 변필성(2007)은 공간이질성이 공간데이터로 측정된 변수들 간의 관계가 공간상에 일정하지 않음을 의미한다고 하여 지역을 도시와 농촌으로 구분하는 변수와 지역별 대학졸업자 수라는 변수 간에 일어날 수 있는 관계, 즉 도시지역일수록 인구 1천명 당 대학졸업자가 많고, 농촌일수록 그 수치가 적다는 공간패턴을 사례로 설명하였다. 또한 어떤 도시의 북쪽 근린에는 평당 아파트가격이 집중적으로 높게 나타나고 남쪽 근린에는 해당 가격이 집중적으로 낮게 나타난다고 가정해 봤을 때 이를 근린이라는 공간스케일에서 보면 평당 아파트가격에 있어 공간적 자기상관이 일어나지만, 스케일을 확장해서 해당 도시 전체를 놓고 볼 경우 남쪽과 북쪽 그런 간의 공간적 이질성이 나타난다고 간주할 수 있다고 보았다.



### (3) 공간상관성의 진단

정우현(2011)은 공간상관성이 분석 대상지역의 물리적 범위에 따라 전역적으로 나타날 수도 있고 국지적으로 일부 지역에서만 나타날 수도 있다고 하였다. Fotheringham(2002)은 두 개의 서로 다른 지역의 데이터를 하나로 집계할 경우 평균의 논리에 의해 이른바 심슨의 역설(Simpson's Paradox)이 나타날 수 있다고 보았다. 예를 들어 두 개의 지역 모두 거주밀도와 지가 사이에 음의 상관관계를 갖고 있음에도 불구하고 이를 집계할 경우 양의 상관관계가 생기는 역설적인 결과가 나타날 수 있다. 이는 곧 국지적으로 상관성을 갖는 여러 집단이 존재할 수 있으나, 전역적인 상관성 지표로는 측정할 수 없는 경우가 생길 수 있음을 의미하는 것이다.



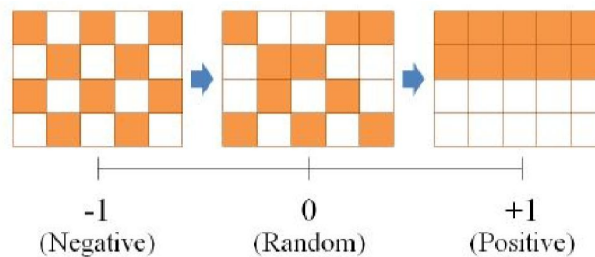
<그림 2-6> 공간자료에서 심슨의 역설이 나타나는 예(Fotheringham, 2002)

#### 1) 전역적 공간상관성

전역적 공간상관성은 분석대상지역 전체에 대하여 특정 변수값들이 공간적인 배치에 따라 상호 연관성을 갖는 것을 뜻한다. 공간변수의 전역적 공간상관성을 검정하기 위한 대표적인 통계량은 Moran's I(Moran 1950), Geary's C(Geary 1954) 등이 대표적이며, 이 중 가장 널리 사용되는 통계량은 Moran's I 통계량이다.



Moran's I 통계량은 공간자기상관을 파악하기 위한 측정척도로서 인접해 있는 공간단위(Neighboring spatial unit)들이 갖는 값들의 크기가 공간적으로 유사하게 분포되어 있는 정도를 의미하여, 공간단위들이 갖는 값을 비교하여 계수를 산출한다. 즉, 공간단위의 값들이 유사한 존들이 인접하게 분포되어 있으면 Moran's I의 값은 1에 가깝게 산출되고, 무작위 패턴을 가지고 있으면 0에 가깝게 산출된다. 또한 -1에 가깝게 산출되면 부(-)의 공간적 상관성을 갖는 것으로 해석하며, 이는 서로 다른 값들이 공간적으로 인접하고 위치하는 양상으로 분포되는 것을 의미한다.



<그림 2-7> Moran's I값과 공간적 분포 개념도(김진희, 2011)

다음의 식2-7에서 변수  $X$ 는 검정대상 공간설명변수를 나타내고,  $w_{ij}$ 는 공간가중치를 의미한다. 즉 통계량을 계산함에 있어 공간가중치의 값은 고정된 조사치인 설명변수의 값보다 큰 영향을 미치게 된다.

$$Moran's\ I = \frac{N}{\sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij} (X_i - \bar{X})(X_j - \bar{X})}{\sum_i (X_i - \bar{X})^2} \dots\dots\dots (식2.7)$$



여기서,  $X_i, X_j$  : 공간  $i$ 와 공간  $j$ 의 값,

$\bar{X}$  :  $X$ 의 평균 값,  $w_{ij}$ : 공간가중치,  $N$  : 공간수

또 다른 통계량인 Geary's C는 Moran's I에 비해 보다 국지적인 자기상관성에 민감한 것으로 알려져 있다. Geary's C는 식2-8과 같이 구할 수 있다.

$$C = \frac{(N-1) \sum_i \sum_j w_{ij} (X_i - X_j)^2}{2 \sum_i \sum_j w_{ij} (X_i - \bar{X})^2}, \quad 0 \leq C \leq 2 \quad \dots\dots\dots (\text{식2.8})$$

여기서,  $X_i, X_j$  : 공간  $i$ 와 공간  $j$ 의 값,

$\bar{X}$  :  $X$ 의 평균 값,  $w_{ij}$ : 공간가중치,  $N$  : 공간수

본 연구에서는 전역적 공간상관성 분석을 위해 Moran's I를 이용하여 공간상관성을 진단해 보았다.

## 2) 국지적 공간상관성

전역적 공간상관성으로 인해 연구 대상범위의 지역적인 내부의 국지적 공간적 연관성을 규명하기는 어렵다. 특히 분석대상지역 내에서 여러 개의 지역 중심(Local Center)이 존재할 가능성이 높은 대도시지역이나 대도시권역(Metropolitan Area)의 경우 다수의 국지적인 공간종속성이 측정될 가능성이 높다.

Anselin(1995)은 Moran's I를 수정하여 국지적인 공간종속성을 측정할 수 있는 지표인 LISA(local indicator of spatial Association)을 제시하였다. 이 지표는



Local Moran's I라고 불리어지기도 하며, 지역  $i$ 의 Local Moran's I는 식 2-9와 같이 정의된다.

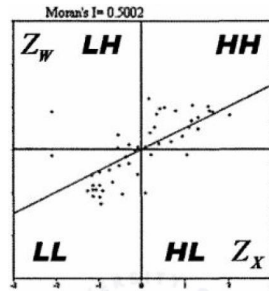
$$Local\ Moran's\ I = (X_i - \bar{X}) \sum_j w_{ij} (X_j - \bar{X}) \dots\dots\dots (식2.9)$$

여기서,  $X_i, X_j$  : 공간  $i$ 와 공간  $j$ 의 값,

$\bar{X}$  :  $X$ 의 평균 값,  $w_{ij}$ : 공간가중치,  $N$  : 공간수

이 통계량은 특정 지역의 값과 인접한 주변 지역의 값들이 갖는 가중 평균이 유사할 경우 양의 값을 반대로 가중평균값들이 상이할 경우 음의 값을 나타낸다. LISA분석은 지역적으로 어떠한 공간적 군집패턴이 발생하는지를 파악하기 위해 실시하는 분석이다(정경석 외., 2009). 아래 그림2-3은 표준화된 Local Moran's I 값을 좌표평면에 나타낸 모란산점도(Moran scatter plot)의 예이다. X축은 검정의 대상이 되는 변수의 표준화되는 값이며, Y축은 표준화된 공간가중치이다. 1사분면과 3사분면에 위치한 지역의 Local Moran's I는 양(+)의 값을 가지며, 2사분면과 4사분면에 위치한 지역의 Local Moran's I는 음(-)의 값을 갖게 된다. 1사분면의 경우 해당지역을 중심으로 높은 통계량이 존재하는 HH(High-High) 유형으로, 3사분면은 모두 낮은 통계량이 존재하는 LL(Low-Low)유형으로 분류한다. 2사분면의 경우 해당지역의 통계량은 낮은데 비해 주변 값이 높은 유형으로 LH(Low-High)로, 4사분면은 그 반대의 경우로 해당지역의 통계량은 높은데 주변 값은 낮은 유형인데 HL(High-Low)로 분류한다.





<그림 2-8> 모란 산점도의 예(정우현, 2011)

또 다른 국지적 공간 상관성 통계량으로는 Getis & Ord(1995)에 의해 제시된 Local  $G_i^*$ 가 있다. 이 통계량은 Local Moran's I와는 달리 평균보다 주변지역의 변수값이 높을 경우 양의 값을, 반대일 경우 음의 값을 갖는 일종의 공간적 군집분석으로 Hot spot analysis라고 불리기도 한다.

$G_i^*$ 는 식2-10을 통해 산출할 수 있으며,  $G_i^*$ 는 값 자체가 Z값이기 때문에 추가적인 표준화 과정은 필요하지 않다.

$$\text{Getis-Ord } G_i^* = \frac{\sum_j w_{ij} X_j - \bar{X} \sum_j w_{ij}}{S \sqrt{\frac{[N \sum_j w_{ij}^2 - (\sum_j w_{ij})^2]}{N-1}}} \quad \dots\dots\dots (\text{식2.10})$$

여기서,  $S = \sqrt{\frac{\sum_j X_j^2}{N} - (\bar{X})^2}$ ,  $X_i, X_j$  : 공간  $i$ 와 공간  $j$ 의 값,

$\bar{X}$  :  $X$ 의 평균 값,  $w_{ij}$ : 공간가중치,  $N$  : 공간수

### 3) 공간이질성

공간이질성은 이분산성이 공간적으로 나타내지는 특성이다. 일반적인 회귀



모형에서 이분산성에 대한 진단은 Breusch-Pagan 검정에 의한다.

Breusch-Pagan 검정법은 회귀모형의 잔차 자승을 종속변수로 하는 회귀모형을 구하고, 여기서 얻어진 결정계수를 통해 라그랑지 승수 검정(Lagrange Multiplier test)으로 공간적으로 균질하다는 귀무가설을 검정하는 기법이다 (Heij & De Boer, 2004). 이 검정통계량은 식2-11과 같다.

$$LM = nR_{e^2}^2 \sim \chi_k^2 \quad \dots\dots\dots (식2.11)$$

여기서,  $R_{e^2}^2$ 는 잔차제곱을 종속변수로 하는 회귀모형의 결정계수를 의미하며,  $n$ 은 데이터의 수,  $k$ 는 설명변수의 수,  $\chi_k^2$ 는 자유도가  $k$ 인 카이제곱분포를 의미한다.

#### (4) 공간가중치행렬

공간계량모형에서 공간적 특성을 갖는 변수들 간의 공간적 관계를 정량적으로 표현하는 개념이 바로 공간가중치이다. 이 공간가중치는 행렬의 형태로서 분석대상지역 내의 모든 개별지역  $i$ 와  $j$ 간의 관계를 표현하게 되며, 이 행렬이 바로 공간가중치 행렬로서, 앞서 살펴본 Moran's I나 Local Moran's I, Getis-Ord  $G_i^*$ 를 계산하는 산식에 포함된  $w_{ij}$ 가 바로 이것이다. 이 공간가중치 행렬은 매우 중요한 개념으로 Anselin(1992)은 이를 어떻게 구성하느냐에 따라 모형의 성패가 좌우된다고 하였다.

공간가중치는 크게 두 가지 형태로 구분된다. 먼저, 분석 존의 물리적 인접성(연속성)에 따라 0 또는 1의 값을 부여하는 연속성(Contiguity) 기반의 가중





치와 실제 존 간의 거리에 따라 가중치를 부여하는 거리(Distance) 기반의 가중치로 구분된다.

### 1) 연속성기반 가중치

공간적 인접성을 고려하여 인접한 경우 1, 이격된 경우 0을 부여하는 것을 의미한다. 이는 공간적 연속성을 고려하는 것으로 공간 연속성 행렬(Spatial Contiguity Matrix)로 불리기도 한다. 이는 공간계량모형에서 가장 기본적으로 사용되는 공간가중치라 할 수 있다. Lesage(1999)는 지리적 인접여부를 활용한 행렬 구성으로 다음과 같이 4가지로 구분하여 가중치 부여방법론을 제시하였다.

- Linear contiguity : 특정 지역의 경계선에 공유된 다른 지역의 경계선이 완전히 공유되는 경우 행렬의 원소  $w_{ij} = 1$ 을 다른 경우에는  $w_{ij} = 0$ 을 부여하는 방법이다.
- Rook contiguity : 두 지역의 경계선이 공유되기만 하면  $w_{ij} = 1$ , 그렇지 않으면  $w_{ij} = 0$ 을 부여하는 방법이다.
- Bishop contiguity : 두 지역이 한 모서리를 공유하면  $w_{ij} = 1$ , 그렇지 않으면  $w_{ij} = 0$ 을 부여하는 방법이다.
- Queen contiguity : 두 지역이 한 면이나 모서리를 공유하면  $w_{ij} = 1$ , 그렇지 않으면  $w_{ij} = 0$ 을 부여하는 방법으로 Rook와 Bishop 방식의 합집합을 의미한다.

Anselin(2003)은 실제 공간단위에서 수집된 데이터를 분석할 때 Rook contiguity의 인접성 가중치가 보편적으로 사용된다고 하였다. 본 연구에서는 분석에 사용되는 종속변수의 연속성 및 거리기반 인접성 공간가중치 방식을



모두 적용시켜 전역적 Moran's I 통계량을 바탕으로 가장 공간상관성을 잘 나타내는 형태의 인접성 가중치 방식을 도출하여 선정하였다.

0	0	0
1	A	1
0	0	0

Linear

0	1	0
1	A	1
0	1	0

Rook

1	0	1
0	A	0
1	0	1

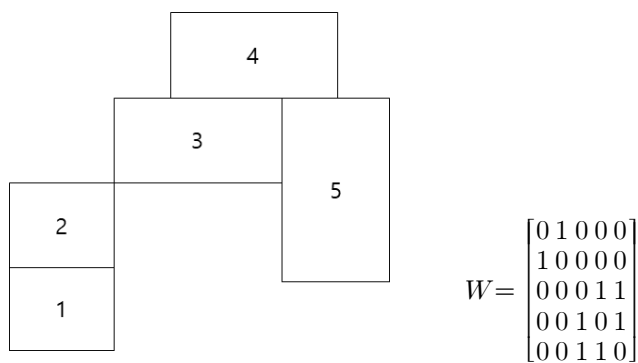
Bishop

1	1	1
1	A	1
1	1	1

Queen

<그림 2-9> 공간인접성의 수량화(정건섭, 2011)

Lesage(1999)가 제시한 Rook contiguity에 대해 5개 존으로 구성된 간단한 예시를 통해 설명을 하면 아래 그림2-10과 같다. 이를 식으로 표현하면 식 2-0과 같이 존1의 경우 존2에만 인접하고 있으므로 1행 2열의  $w_{12}$ 는 1의 값을 가지며, 나머지 행렬을 0의 값을 가지게 된다. 존3의 경우에는 존4와 존5와 인접하여 있으므로  $w_{34}$ 와  $w_{35}$ 에서 1의 값을 가지며, 반대의 경우도 마찬가지이므로 이 행렬은 대칭의 형태를 보인다.



<그림 2-10> Rook방식 공간연속성 가중치행렬 구성의 예(Lesage, 1999)



반면 이러한 공간가중치 행렬을 그대로 모형에 적용시킬 경우 모든 지역에 대해 인접성의 여부에 따라 동일한 크기의 가중치가 있다는 해석상의 중대한 오류를 범할 수 있다는 문제가 발생한다. 이에 이러한 문제를 해결하기 위해 공간가중치 행렬을 표준화(standardization)시키는 과정이 필요하다. 표준화의 기본적인 원칙은 많은 지역과 연계되어 있는 지역은 그 지역들에서 동일한 영향을 받고 있으며 그 가중치의 합이 1이 된다는 것에서부터 출발한다(이성우 외, 2006). 표준화된 공간가중치 행렬 구성의 사례를 살펴보면, 상기 그림 2-10에서의 예와 같이 존3과 존4, 존5는 상호 연계되어 있으며 존3의 값은 1이 아니라 1/2 즉 0.5가 된다. 따라서 존3의 행은 (0, 0, 0, 0.5, 0.5)가 된다. 이 방법을 통해 전체 행에 적용시켜 표준화한 식은 다음 식2-12와 같이 도출된다(홍지연, 2013).

$$\text{standardization } W = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.5 & 0.5 \\ 0 & 0 & 0.5 & 0 & 0.5 \\ 0 & 0 & 0.5 & 0.5 & 0 \end{bmatrix} \dots\dots\dots \text{(식2.12)}$$

## 2) 거리기반 가중치

거리기반 가중치는 실제거리를 기반으로 두 지역의 접근성(Accessibility)에 대한 척도로 해석이 가능하며, 공간적 상호작용(Spatial interaction)의 크기가 적용될 수도 있다. 거리의 경우 가까운 곳일수록 큰 가중치를 가져야 하므로 거리의 역수가 가장 많이 사용된다. 가장 간단한 예를 수식으로 표현하면 식2-13, 식2-14과 같다(홍지연, 2013).



$$w_{ij} = \frac{1}{d_{ij}} \dots\dots\dots (\text{식2.13})$$

$$w_{ij} = e^{-d_{ij}} \dots\dots\dots (\text{식2.14})$$

여기서,  $w_{ij}$ 는 두 지역  $i$ 와  $j$ 의 거리기반 가중치로  $d_{ij}$ 는 두 지역  $i$ 와  $j$ 사이의 거리이다. 거리를 정의하는 방법에 따라 두 지역의 중심점(Centroid) 사이의 유클리드거리(Euclidian Distance)가 사용될 수도 있고, 실제 도로망을 이용한 네트워크거리(Network Distance)가 사용될 수도 있다(정우현, 2011).

## (5) 공간회귀모형(spatial regression model)

### 1) 전역적 공간회귀모형(Global spatial regression model)

공간회귀모형은 일반선형회귀모형에 공간상관 관계, 즉 공간종속성, 공간이질성을 나타내는 공간가중치행렬의 선형합(Linear sum)으로 일반적인 모형식은 식2-15와 같다(이성우 외, 2006).

$$y = \rho W_1 y + X\beta + u \dots\dots\dots (\text{식2.15})$$

$$u = \lambda W_2 u + \epsilon, \epsilon \sim N(0, \sigma^2 I_n) \dots\dots\dots (\text{식2.16})$$

여기서,  $y$  : 종속변수

$X$  :  $n \times k$ 의 설명변수 행렬

$\beta$  :  $k \times 1$ 의 설명변수의 계수 벡터

$W_1, W_2$  : 공간가중치 행렬

$\rho, \lambda$  : 공간가중치 행렬의 계수

$I_n$  :  $n$ 차의 단위 행렬



여기에서,  $W_1$ 과  $W_2$ 는 공간적 상관성을 나타내는 공간가중행렬을 의미하며,  $\rho$ 와  $\lambda$ 는 각 공간가중행렬의 회귀계수이다. Lesage(1999)의 분류에 의하면, 모형에서  $X$ 와  $W_2$ 가 0이면, 그 모형은 FAR(First Order Spatial AR model)이 되고,  $W_2$ 만 0인 경우는 공간시차모형(SLM; Spatial Lag Model) 또는 공간자기회귀모형(SAR; Spatial Autoregression Model)이 되며,  $W_1$ 이 0이면 공간오차모형(SEM; Spatial Error Model)이 된다. 또한,  $W_1$ 과  $W_2$ 가 모두 0이 아니면 일반공간모형(SAC; General Spatial Model)으로 SAR과 SEM을 모두 혼합한 모형이 된다(홍지연, 2013). 이 중에서 공간시차모형과 공간오차모형이 가장 널리 사용되는 공간회귀모형이라고 할 수 있다.

#### (a) 공간시차모형(SLM)

공간시차모형은 종속변수 자체가 공간적 자기상관을 갖고 있다고 가정하는 모형으로 상기 식2-16에서  $W_2$ 가 0인 경우로 FAR모형에 설명변수 행렬  $X$ 를 더해준 형태의 식이며, 일반식은 다음 식2-17과 같다.

$$y = \rho W_1 y + X\beta + \epsilon, \quad \epsilon \sim N(0, \sigma^2 I_n) \quad \dots\dots\dots (\text{식2.17})$$

여기서,  $y$  : 종속변수

$X$  :  $n \times k$ 의 설명변수 행렬

$\beta$  :  $k \times 1$ 의 설명변수의 계수 벡터

$W$  : 공간가중치 행렬

$\rho$  : 공간가중치 행렬의 계수

$I_n$  :  $n$ 차의 단위 행렬



여기서,  $W$ 는 공간가중치행렬로 주로 표준화된 1차 Rook contiguity matrix가 사용된다.  $\rho$ 는 종속변수인  $y$ 의 spatial lag항의 계수가 되고,  $X$ 는  $(n \times k)$  설명변수행렬이 되며,  $\beta$ 는  $(k \times 1)$ 의 설명변수의 계수벡터가 된다. 이 모형의 계수 추정은 OLS 대신 로그우도(log likelihood)함수에 의해 이루어진다(이성우 외, 2006).

#### (b) 공간오차모형(SEM)

공간오차모형은 오차항에 공간적인 종속성이 포함되는 모형이다. 따라서 모형식은 식2-18과 같이 구성된다. 마찬가지로  $X$ 는  $(n \times k)$  설명변수 행렬이 되며,  $\beta$ 는  $(k \times 1)$ 의 설명변수의 계수벡터가 된다. 공간오차모형의 특성을 파악하기 위해 식2-19와의 차이점을 살펴보면, spatial lag항이 설명변수 항이 아니라 오차항에 포함되어 있음을 알 수 있다. 이는 공간가중치행렬이 오차항 부분에 속해 있음에도 spatial lag의 개념을 공간자기상관으로 해석하고자 하는 의도를 포함하고 있다는 것을 뜻한다. 자기상관은 회귀분석을 위한 가정 중 오차항끼리는 상관관계를 가지지 않아야 한다는 가정에 위배된 경우이다. 그러나 SEM모형에서는 이러한 가정이 Spatial lag에 의해 위배되고 있음을 전제하고 이를 해결하기 위하여 공간가중치 행렬을 오차항에서 다시 정의하고 있다(이성우 외, 2006).

$$y = X\beta + u \quad \dots\dots\dots (\text{식2.18})$$

$$u = \lambda Wu + \epsilon, \quad \epsilon \sim N(0, \sigma^2 I_n) \quad \dots\dots\dots (\text{식2.19})$$

여기서,  $y$  : 종속변수

$X$  :  $n \times k$ 의 설명변수 행렬



$\beta$  :  $k \times 1$ 의 설명변수의 계수 벡터

$W$  : 공간가중치 행렬

$\lambda$  : 공간가중치 행렬의 계수

$I_n$  :  $n$ 차의 단위 행렬

## 2) 국지적 공간회귀모형(Locally linear spatial model)

앞서 설명한 바와 같이 전역적 공간회귀모형은 공간시차모형과 공간오차모형으로 구분된다고 하면 국지적 공간회귀모형으로는 지리적 가중회귀모형(GWR)으로 명명된다. GWR은 Fortheringham 등(1999)에 의해 개발된 공간통계 분석 기법으로 공간상에서 서로 다른 위치에 존재하는 데이터들에 대하여 각각 다른 회귀모형을 적용하여 공간적 비정상성(non-stationarity)을 탐색하는 방법이다(김익희, 2005).

GWR은 분석대상지역내에서 개별 관측치에 대해 회귀분석 기준점을 선정하여 그 중심점으로부터 기준거리 내에 있는 관측치들을 대상으로 수행하는 회귀분석으로 지역별로 각각 다른 계수를 추정함으로써 변수들의 국지적인 영향을 파악할 수 있는 결과를 제시한다(홍지연, 2013). GWR이 국지적으로 분석범위를 나누어 시행하는 일반적인 회귀분석방법과 다른 점은 기존의 회귀분석에서 고려할 수 없는 인접지역과의 공간적 상호작용을 반영한다는 것이다. GWR은 회귀지점에서 멀어질수록 낮아지는 차별적인 가중치를 데이터들에 적용함으로써 Tobler의 지리학 제1법칙을 반영하고 있다.

아래 식2-20은 GWR의 일반식을 나타낸다.

$$Y_i = \sum_k \beta_k(u_i, u_j) x_{ik} + \epsilon_i \quad \dots\dots\dots (\text{식}2.20)$$



여기서  $(u_i, u_j)$ 는 존  $i$ 의 중심점의 좌표를 의미하며,  $\beta_k(u_i, u_j)$ 는 존  $i$ 의  $k$ 번째 설명변수의 계수를 의미한다.

$$\hat{\beta}(u_i, v_j) = (X^T W(u_i, v_i) X)^{-1} X^T W(u_i, v_i) Y \quad \dots\dots\dots (\text{식2.21})$$

여기서,  $W(u_i, v_i) = c \begin{bmatrix} W_{i1} & 0 & \dots & 0 \\ 0 & W_{i2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & W_{i\in} \end{bmatrix},$

$$W_{ij} = \exp(-\gamma d_{ij}^2) \quad \dots\dots\dots (\text{식2.22})$$

여기서,  $d_{ij}$  : 지역  $i$ 의 중심점에서 지역  $j$ 의 중심점까지의 거리

$\gamma$  : kernel bandwidth

이 bandwidth는 회귀분석의 중심점이 영향을 미치는 국지적 공간상관범위를 뜻한다. bandwidth값을 구하는 대표적인 방법은 AIC(Akaike Information Criterion)를 최소화하는 값을 찾는 AICc(AIC with a small sample correction) 방법과 CV(Cross Validation)를 이용한 방법이 있으며, 각각에 대한 설명은 다음과 같다. AICc는 AIC통계량을 한정된 샘플크기에 보정한 값으로 식2-23과 같다.

$$AIC_C = 2n \ln(\hat{\sigma}) + n \ln(2\pi) + n \left( \frac{n + \text{tr}(S)}{n - 2 - \text{tr}(S)} \right) \quad \dots\dots\dots (\text{식2.23})$$

여기서,  $\hat{\sigma}$  : 회귀분석으로부터 표준화된 잔차제곱합

$S$  :  $\hat{y} = Sy$ 를 만족시키는 hat matrix





$tr(S)$  :  $S$ 의 대각행렬 원소의 합

여기서,  $\hat{\sigma}$ 는 회귀분석으로부터 표준화된 잔차제곱합을 의미하며,  $S$ 는  $\hat{y} = Sy$ 를 만족시키는 hat matrix이며,  $tr(S)$ 는  $S$ 의 대각행렬원소의 합을 의미한다. CV를 이용한 방법론은 식2-24와 같이 관측치와 추정치의 차이 제곱합을 CV로 정의하고, bandwidth를 변화시키면서 CV가 감소하다 증가하는 변곡점에서 Bandwidth를 결정하게 되는 방법이다.

$$CV = \sum_{i=1} [y_i - \hat{y}_i(b)]^2 \dots\dots\dots (식2.24)$$

여기서,  $\hat{y}_i(b)$ 는 기준거리  $b$ 에 있는 관측지점에 대한 종속변수의 추정치를 의미한다(Charlton & Forheringham, 2003). 통행발생단계에서 주로 이용되는 선형회귀분석모형은 일반적으로 OLS에 의해 추정된다. 이는 종속변수인 각 존의 통행발생량이 서로 독립적임을 가정한다. 따라서 전통적 4단계 수요추정 기법에서의 통행발생모형은 실제 통행발생과 높은 상관관계가 있는 공간적 영향을 반영할 수 없다는 단점이 있고 만약 공간적 영향이 존재한다는 사실을 검증한 이후에도 계속해서 선형회귀분석을 이용한 통행발생모형의 사용은 사실상 왜곡된 결과를 나타낼 수도 있다. 따라서 통행발생모형의 구축 시에는 현재 본 연구의 대상범위 내에서 공간적 상관성이 있는지에 대한 여부를 우선적으로 검증·파악한 이후에 공간적 영향을 고려할 수 있는 통행발생모형을 구축하는 것이 바람직하다고 판단된다.



## 2-2. 선행연구 검토

### 2-2-1. 통행발생모형 개선 관련 연구

#### (1) 통행발생모형 개선 관련

통행발생모형을 여러 가지 변수와 방법론을 적용시켜 개선해 나가기 위한 선행연구는 국내외에서 많은 연구가 이루어져 왔다.

유완 외(1999)는 1997년 서울시 교통센서스 자료에 수록된 일주일간 가구통행실태 조사 자료를 이용하여 가구의 통행행태를 요일별·통행목적 및 통행수단별로 분석한 선형회귀모형으로 각각의 통행 수요 집단에 대해 총통행수 및 통행/인을 종속변수로 통행발생 예측모형을 정립하였다. 그 결과 서울시 가구의 요일별 통행패턴은 수요일 통행이 많은 수요 집단과 평일 통행이 많은 주중집단, 주말 통행이 많은 주말집단의 3개 집단으로 분류되었으며, 집단별 특성을 고려한 통행발생예측모형은 가구의 총통행수를 위주로 하여서는 집단간 차이가 크지 않았으나 일주일간의 통행패턴에서는 3집단에 차이가 있으므로 연평균 일일 통행발생량을 예측하기 위해서는 집단별 특성을 감안한 변수의 선정이 필요하다고 보았다. 또한 가구 총통행수를 종속변수로 하였을 때 가구원 평균 통행수를 이용한 경우와 비교해 모형의 설명력이 훨씬 높아진 것으로 나타나 이를 이용하는 것이 통행발생모형을 구축하는데 더 나은 결과를 보였다고 하였다.

장수은 외(2000)는 제1차 전국물류현황 조사 자료를 바탕으로 기존 대표적 통행발생모형인 회귀모형과 신경망 이론에 근거한 통행발생모형을 비교·분석함으로써 통행발생모형에 새로운 방법론을 제시하였다. 신경망 이론은 복잡하고 부정확한 데이터나 특별한 사전정보가 없는 경우를 모형화하는데 우수



한 능력을 발휘하는 것으로서 역전파(Backpropagation) 모형으로 불리는 역전파 학습 알고리즘을 이용한 다층전파 네트워크의 경우 적용이 간편하고 여러 연구에서 모형의 우수성이 확인되고 있다는 사실에 근거하여 새로운 분석방법론으로서의 그 가치를 정립하고자 하였다. 그 결과 선형 및 비선형회귀분석 결과값과 비교하여 신경망 모형의 우수성이 실증적으로 입증되어 향후 통행발생모형에서 이의 적용가능성에 대해 높이 평가하였다.

정지교(2001)는 1996년 가구통행실태조사 자료를 바탕으로 용도별 면적비율을 이용하여 서울시 행정동을 주거중심지역, 상업중심지역, 상업·업무중심지역으로 군집분석을 통해 구분한 후 사회경제지표를 이용하여 통행목적별 개별 통행발생예측모형을 구축하였다.

김진자 외(2004)에서는 수도권지역의 통근·등교 통행발생률을 종속변수로 하여 카테고리분석법과 선형회귀모형을 활용하여 예측치와 관측값에 대한 비교분석을 시도하였다. 이를 통해 수도권에서의 장래 통근과 등교통행발생량 예측에서 카테고리분석법이 회귀분석법보다 예측력에서 우수하다는 결과를 도출하였다.

김태호 외(2010)는 통행발생모형에서 각종 사회경제지표와 통행발생량의 관계가 선형임을 전제로 한 선형회귀모형의 한계를 지적하며 다양한 존의 특성을 반영할 수 있는 변수에 대한 시장분할을 토대로 새로운 유형별 통행발생모형을 개발하였다. 즉 존의 다양한 특성(토지이용, 사회경제적 등)을 고려하여 시장분할 방법론으로서 통행 발생률을 기반으로 한 CART(Classification & regression tree)방법과 회귀분석을 이용하였다. 그 결과, CART분석을 통한 존 특성분석 결과, 유출통행은 사회경제적 요인에 영향을 받고 있으며, 유입통행은 토지이용, 사회·경제적 요인에 영향을 받고 있는 것으로 나타났다. 또한



2002년 가구통행실태조사 자료를 활용하여 실측·검증하였을 때 유출·유입의 경우 기존 모형보다 적합도가 높아진 것으로 나타났다. 즉 사회경제적, 토지이용 요인들을 활용한 시장분할 결과를 반영하여 유형별 통행발생모형을 개발한 결과가 기존 연구보다 우수한 것으로 나타났다고 제시하였다.

송재인 외(2011)에서는 기존 수도권 통행발생모형의 선형회귀 모형식에서 토지이용 특성을 반영한 통행발생모형을 구축하였다. 이를 통한 검증을 위해 기존 모형식과 RMSE값을 비교 분석하였고 그 결과 기존 모형은 주거 특성이 강한 지역의 추정에서 오차범위가 적었으나, 2·3차 산업비중이 높은 지역에서는 설명력이 떨어지는 것으로 나타났다. 또한 각 목적별 모형에서 전반적으로 기존 모형보다 오차가 적은 것으로 나타나 사회경제지표 및 토지이용 변수를 활용한 지역별 모형 추정결과가 기존 연구보다 우수한 것으로 나타난 결과를 보였다.

Washington & Wolf(1997) 연구에서는 계층적 구조 기반의 회귀분석(Hierarchical Tree-Based Regression Models)을 이용하여 기존 선형회귀를 통한 통행발생모형을 보완하는 이론적 연구를 진행하였다.

Liya et al.(2008) 연구에서는 기존 통행발생모형에서 각 존의 개인특성을 고려하지 못한 불확실한 평균적 통행 발생률로 추정되는 점을 개선시키기 위하여 비집계화(Disaggregate) 이론을 토대로 목적지선택모형으로 조합된 Trip efficiency와 통행존의 효과를 추정하는 방법을 제시하였다. 그 결과 교통존의 통행유인력이 고려된 목적지 선택모형이 비집계 모형을 토대로 한 효용이론을 이용하여 구축되었다. 교통존 유인력은 통행발생량에 영향을 주었고 유인력은 연령, 수입, 목적, 차량보유여부, 지불수단, 시간과 비용 변수에 영향을 받는 것으로 나타났다. 교통존 유인력은 목적지 유인력의 고려하여 개선



된 비집계 통행발생모형의 정확성을 보여준 최종 예측결과에서 존별로 추정될 수 있는 것으로 나타났다.

Chang et al.(2014) 연구에서는 서울시 가정기방 통근통행 자료를 이용하여 회귀모형, 토빗모형, 포아송모형, 순서형로짓모형, 교차분류분석, 다중분류분석법을 이용하여 통행발생모형 추정결과를 비교분석한 결과를 제시하였다. 그 결과 카테고리 유형의 모델이 전체적으로 가장 좋은 결과를 보이는 것으로 나타났다.

<표 2-1> 통행발생모형 개선관련 선행연구

연구자	주요내용	분석모형
유완 외 (1999)	1997년 서울시 가구통행실태조사 자료를 이용하여 요일별 통행수와 가구특성변수를 군집분석 후 회귀모형을 통한 통행발생모형을 구축	선형회귀모형
장수은 외 (2000)	신경망 이론을 이용하여 물동량 자료를 바탕으로 통행발생모형에 적용하여 선형·비선형 모형과의 비교분석을 통한 우수성 제시	신경망모형
정지교 (2001)	사회경제적 특성을 고려하여 서울시 행정동을 주거중심지역, 상업중심지역, 상업·업무중심지역으로 구분 후 목적별 통행발생예측모형 정립	선형회귀모형
김진자 외 (2004)	수도권의 통근·등교통행 발생률을 카테고리모형과 선형회귀모형을 이용하여 예측하고 관측값과 비교하여 카테고리모형의 예측력이 우수하다는 결과를 제시	카테고리, 선형회귀모형
김태호 외 (2010)	사회경제지표와 통행발생량과의 선형관계의 한계를 지적하고 존 특성을 감안한 변수를 시장분할을 통한 CART 분석과 회귀분석을 이용하여 검증한 결과 시장분할 결과를 반영한 통행발생모형이 우수하다고 제시	CART분석 선형회귀모형
송재인 외 (2011)	기존 통행발생모형에서 사용된 설명변수들의 한계점을 개선하기 위해 수도권을 대상으로 지역별 토지이용특성을 반영한 O/D와 P/A기반 통행발생모형을 구축하고 추정오차 비교분석을 통한 통행발생모형 추정의 적합한 접근방법 모색	선형회귀모형



연구자	주요내용	분석모형
Washington & Wolf (1997)	계층적 구조 기반의 회귀분석(Hierarchical Tree-Based Regression Models)을 이용하여 기존 선형회귀를 통한 통행발생모형을 보완하는 이론적 연구	계층적회귀 분석모형
Liya et al. (2008)	통행자의 일반특성 및 통행특성과 토지이용면적, CBD거리 통행서비스를 이용하여 개별행태모형을 추정하고 도착지의 유인력을 거리, 토지이용면적비 등을 정량화시킨 후 해당지역이 갖는 상대적 효용을 산출하여 통행발생량을 설명	카테고리, 선형회귀 모형
Chang et al (2014)	서울시의 HBW통행자료를 이용하여 회귀모형, 토빗모형, 포아송모형, 순서형로짓모형, 교차분류분석, 다중분류분석법을 이용하여 통행발생모형을 비교분석한 결과를 제시	회귀, 토빗, 포아송, 순서형로짓, 교차분류분석법 등

## 2-2-2. 공간분석을 이용한 통행발생모형 관련 연구

### (1) 공간계량분석을 이용한 통행발생모형 관련 연구

김진희 외(2011)에서는 대구광역권을 대상으로 인구, 고용자수, 학생수를 이용한 지리적 가중회귀모형(GWR)을 통해 통행발생모형을 구축하였다. 그 결과, 일반적인 회귀모형과 적합도, RMSE등을 비교 분석함으로써 국지적 공간상관성이 존재함을 확인하고 공간상관성으로 인한 통행발생모형의 오차가 크게 개선되는 점을 제시하였다.

정우현(2011)은 서울시 통행조사 자료를 이용하여 집계화된 통근통행발생량의 공간중속성과 이질성을 확인하여 그 원인을 분석하고 이를 변수로 하는 공간계량모형을 구축하였다. 그 결과, 통근통행발생량의 공간상관성과 이질성은 분석대상지역의 국지적인 토지이용특성과 지리적 입지에 기인하며, 그 결과는 추정된 회귀모형에 의한 오차의 공간적 자기상관으로 나타난다는 결과



를 도출하였다. 또한 지리적 가중회귀모형의 구축을 통한 계량화된 모형은 일반 다변량회귀모형과 비교하여 우수한 설명력을 가지는 것으로 나타났다.

진창중 외(2012)은 통행발생단계에서 선형회귀모형이 반영할 수 없는 공간적 자기상관을 통제하기 위한 방법인 공간회귀분석을 사용하여 통행발생 유출·유입모형을 구축하였다. 또한 기존 회귀분석에서 사용되는 사회경제지표 이외에도 지역적 특성을 나타내는 토지이용지표, 물리적 지표를 설명변수로 사용하여 다양하게 시도해보았다. 그 결과, 통행목적별 유입·유출량을 종속변수로 하여 기타목적 유입·유출모형과 통학통행 유출모형을 제외한 9개의 모형이 공간종속성 검정결과에 따라 공간회귀모형으로 선택되었으며, 회귀모형과의 비교를 통해 설명력이 상승하였음을 확인하였다.

Amavi et al.(2014) 연구에서는 개선된 통행발생·집중모형의 도출을 위하여 스페인 Santander지역을 대상으로 전역적인 공간자기상관을 고려한 공간회귀분석기법인 SLM, SEM분석을 적용하여 개선된 통행발생집중모형을 제시하였다.

<표 2-2> 공간계량분석을 이용한 통행발생모형 관련 선행연구

연구자	주요내용	분야	비고
김진희 외 (2011)	대구광역권의 가정기반 기타통행과 비가정기반 기타통행 자료를 바탕으로 국지적인 공간상관성을 확인하고 GWR분석을 이용하여 통행발생모형을 구축	지리가중 회귀분석 (GWR)	HBO, NHBO
정우현 (2011)	통근통행생성유인량 및 밀도를 종속변수로 하여 기존 선형회귀분석의 공간적 자기상관 문제를 통제하기 위한 공간회귀분석 기법을 적용한 통근통행발생 모형의 추정력 개선 연구	공간회귀 분석 (SLM, SEM, WLS, GWR)	HBW





연구자	주요내용	분야	비고
진창중 외 (2012)	통행발생단계에서 선형회귀모형의 공간자기상관을 통제하기 위한 방법인 공간회귀분석을 사용하여 통행발생 유출·유입모형을 구축하고 기존 회귀분석에서 적용된 사회경제지표 이외에도 지역적 특성을 나타내는 토지이용, 물리적 지표를 설명변수로 사용	공간회귀 분석 (SLM, SEM)	충통행, 목적별 통행(통근, 통학, 업무, 쇼핑, 기타)
Amavi et al. (2014)	개선된 통행발생·집중모형의 도출을 위하여 스페인 Santander지역을 대상으로 전역적인 공간자기상관을 고려한 공간회귀분석기법인 SLM, SEM분석을 적용하여 진보된 통행발생 집중모형을 제시	공간회귀 분석 (SLM, SEM)	오전 첨두시 통행발생 유입

## (2) 통행발생모형 적용 관련 연구사례

### 1) 국내 사례

통행발생모형을 이용한 수요추정은 기종점통행량(O/D)을 종속변수로 한 선형회귀분석을 적용하여 사용하는 것이 가장 일반적이다. 이때 회귀분석에서 사용되었던 설명변수는 인구, 종사자수, 거주학생수, 수용학생수, 승용차 보유대수 등 주로 사회경제지표를 기준으로 하여 분석된다.

주요 국내 통행발생모형과 관련된 실무 연구에서 통행목적별 통행발생 및 유입별로 사용된 분석기법과 설명변수들은 다음 표2-3과 같이 나타났다. 표에서 보듯이 통행발생모형에서 주로 사용되는 설명변수들은 통행목적별로 출근, 등교와 같은 필수적 통행(Mandatory trip)에 있어서는 인구, 연령별 인구수 등의 변수가 주로 사용되었고 쇼핑, 여가, 기타와 같은 선택적 통행(Optional trip)에 있어서는 1, 2, 3차 산업종사자수, 수용학생수 등의 사회경제지표 관련 변수가 주로 사용된 것을 알 수 있다.





<표 2-3> 국내 통행발생 모형적용 및 변수자료

구 분	모 형		목 적	통행발생/생성	통행도착/유인
				설명변수	
수도권 및 5대 광역권 여객통행량 분석	카테고리 분석법		출근	성별·연령별인구	-
			등교		
	회귀 분석법		출근	-	종사자수
			등교		수용학생수
			귀가	종사자수, 수용학생수	인구
			업무	종사자수	인구, 종사자수
			업무 후 복귀	업무의 도착	업무의 출발
			기타	인구, 종사자수, 수용학생수	인구, 종사자수
교통수요원단 위분석	가정 기반	카테고리 분석법	출퇴근	성별·연령별인구	-
			통학		
			기타		
		회귀 분석법	출퇴근	취업가능인구 (21~65세)	종사자수
			통학	취업가능인구(5~25세)	학생수
			기타	인구, 3차종사자수	3차종사자수
	비가정 기반	카테고리 분석법	모든통행	성별·연령별인구	-
		회귀 분석법		인구, 종사자수	3차종사자수
서울시 장래교통 수요예측 및 대응방안	가정 기반	카테고리 분석법	출근	성별·연령별·지역별 인구	-
			등교		
		회귀 분석법	출근	-	종사자수
			등교		수용학생수
			학원	12~17세 인구	승용차보유대수, 쇼핑관련종사자수
			쇼핑	인구	쇼핑관련종사자수, 초중고 수용학생수



구 분	모 형		목적	통행발생/생성	통행도착/유인
				설명변수	
	비가정 기반	회귀 분석법	기타	인구	3차산업 종사자수, 초중고 수용학생수
			업무	종사자수	종사자수
			쇼핑	3차산업 종사자수, 22~64세 인구수	쇼핑 관련 종사자수, 승용차보유대수
			기타	3차산업 종사자수, 수용학생수	3차산업 종사자수, 수용학생수
수도권 장래교통 수요예측 및 대응방안 연구	가정 기반	회귀 분석법	출퇴근	취업자수	1,2차산업 종사자수, 3차산업 종사자수
			등하교	6~7세 인구, 18~21세 인구	초중고 수용학생수, 대학교 수용학생수
			학원	6~21세 인구(학령인구)	학원관련 종사자수
			쇼핑	15세이상 인구, 천가구당 자동차보유대수	3차산업 종사자수
	비가정 기반		기타	15세이상 인구, 천가구당 자동차보유대수	3차산업 종사자수
			업무	2차산업 종사자수, 3차산업 종사자수	2차산업 종사자수, 3차산업 종사자수
			쇼핑	2·3차산업 종사자수, 총수용학생수	3차산업 종사자수
			기타	2·3차산업 종사자수, 총수용학생수	3차산업 종사자수, 총수용학생수
2007년 광역권 여객기종점 통행량 전수화 대구광역권	회귀 분석법	출근	취업자수	총종사자수	
		등교	5~29세 인구수	초중고 학생수, 대학생수	
		학원	5~59세 인구수	학원관련 종사자수	
		업무	총종사자수	총종사자수	



구 분	모 형	목 적	통행발생/생성	통행도착/유인
			설명변수	
기 준		쇼핑	20세 이상 인구	쇼핑관련 종사자수
		기타	총인구	총인구
		귀가	총인구	총인구
2011년 전국 여객O/D전수 화 및 장래수요예측 광역권 기준	회귀 분석법	출근, 등교, 쇼핑, 기타	취업자수	총취업자수
		업무	총종사자수	총종사자수
		귀가	총종사자수	총인구수
		여가	20세 이상 인구수	총종사자수
2014년 전국여객 O/D보완갱신 연구 대도시권 기준 국가교통DB	회귀분석법	기타 (출근, 등 교, 쇼핑, 기타)	총인구수	총인구수
		업무	총종사자수	총종사자수
		귀가	총종사자수	총인구수
		여가	총인구수	총인구수

자료 : 1) 2002년 국가교통DB구축사업 ‘수도권 및 5대 광역권 여객통행량 분석’, 2003.2.  
교통개발연구원

- 2) 2003년 국가교통DB구축사업 ‘교통수요원단위분석’, 2004. 3. 교통개발연구원
- 3) 서울시 장래교통수요예측 및 대응방안연구, 2004. 11. 서울특별시
- 4) 수도권 장래교통수요예측 및 대응방안연구, 2009. 2. 수도권교통본부
- 5) 2007년 광역권 여객기종점 통행량 전수화, 2008. 4. 국토해양부
- 6) 2011년 국가교통수요조사 및 DB구축사업, ‘전국 여객O/D전수화 및 장래수요예측’,  
2012. 4. 한국교통연구원
- 7) 2014년 국가교통수요조사 및 DB구축사업, ‘전국여객O/D 보완갱신연구’, 2015. 4.  
한국교통연구원

현재 KTDB가 주관이 되어 각 광역권별로 교통기초조사를 통한 여객 O/D  
현행화에서 추정하고 있는 통행발생모형식을 살펴보면, 아래 식2-25와 같은



선형회귀식을 사용하여 통행발생(유입)을 추정하고 있는 것으로 나타났다.

$$Y^{p_{ki}} = \beta_1 \cdot x_{i1} + \beta_2 \cdot x_{i2} + \dots \quad \dots\dots\dots (식2.25)$$

여기서,  $Y^{p_{ki}}$ 는 지역  $k$ 에 속하는 존 $i$ 의 통행목적 $p$ 별 통행량  
 $x_{i1}, x_{i2}, \dots$ 는  $i$ 존의 설명변수 값,  $\beta_1, \beta_2, \dots$ 은 추정계수를 의미한다.

이때 사용된 목적별 모형정산 결과는 아래 표2-4와 같다. 통근통행 및 가정 기반 등하교 통행과 같은 필수적 통행(Mandatory trip)에서 주로 사용된 변수는 통행생성에서 5~19세 인구와 취업자 수 등의 변수가 높은 상관관계를 가지는 것으로 나타났고 통행유입에서는 취업자 수, 종사자수, 수용 학생 수 등의 변수가 높은 상관성을 가지는 것으로 나타났다. 반면 가정기반 쇼핑 및 기타통행, 비가정기반 업무·쇼핑·기타통행과 같은 선택적 통행(Optional trip)에 있어서는 통행생성에서 5~19세 인구, 15세 이상 인구, 3차 산업 종사자수 등의 변수가 높은 상관성을 나타내고 통행유입에서는 학원 관련종사자수, 3차 산업 종사자수 등의 변수가 높은 상관관계를 나타내는 결과를 보였다. 결국 통행목적별로 통행발생량에 영향을 미치는 사회경제지표 변수들은 통행의 발생지와 도착지에서 서로 다른 영향을 미치는 것으로 볼 수 있고 높은 상관관계를 가지는 것으로 나타나 모형의 신뢰성은 상당 부분 증명되었다고 볼 수 있다.

<표 2-4> 목적별 모형정산 결과(대구광역시권)

구 분	모형	목적	변수	계수	t값	P-value	$R^2$
통근통행	생성	대구시	취업자수	1.988	69.381	.000a	0.972



구 분	모형	목적	변수	계수	t값	P-value	$R^2$
		경상북도	취업자수	1.831	50.299	.000a	0.938
	유인	대구시	1,2차산업종사자	1.493	18.248	.000a	0.971
			3차산업종사자	2.700	44.869		
		경상북도	1,2차산업종사자	2.696	48.638	.000a	0.977
			3차산업종사자	2.359	33.798		
가정기반 등하교통행	생성	대구시	5세~19세	1.837	55.796	.000a	0.957
		경상북도	5세~19세	1.863	44.142	.000a	0.922
	유인	대구시	수용 학생 수	1.841	53.368	.000a	0.954
		경상북도	수용 학생 수	1.902	72.490	.000a	0.969
가정기반 학원통행	생성	대구시	5세~19세	0.688	31.444	.000a	0.877
		경상북도	5세~19세	0.583	23.976	.000a	0.808
	유인	대구시	학원 관련 종사자	9.570	25.296	.000a	0.823
		경상북도	학원 관련 종사자	11.785	31.767	.000a	0.883
가정기반 쇼핑통행	생성	대구시	15세 이상 인구	0.148	21.842	.000a	0.774
		경상북도	15세 이상 인구	0.164	9.837	.000a	0.375
	유인	대구시	학원 관련 종사자	0.680	14.263	.000a	0.599
		경상북도	학원 관련 종사자	0.735	11.188	.000a	0.442
가정기반 기타통행	생성	대구시	15세 이상 인구	0.560	23.501	.000a	0.800
		경상북도	15세 이상 인구	0.519	25.763	.000a	0.800
	유인	대구시	3차 산업 종사자	2.446	19.735	.000a	0.738
		경상북도	3차 산업 종사자	2.659	34.200	.000a	0.876
비가정기반 업무통행	생성	대구시	3차 산업 종사자	0.727	30.261	.000a	0.868
		경상북도	3차 산업 종사자	0.325	20.821	.000a	0.724
	유인	대구시	3차 산업 종사자	0.805	19.063	.000a	0.725
		경상북도	3차 산업 종사자	0.310	19.026	.000a	0.686
비가정기반 쇼핑통행	생성	대구시	3차 산업 종사자	0.091	16.504	.000a	0.669
		경상북도	3차 산업 종사자	0.067	10.244	.000a	0.469
	유인	대구시	3차 산업 종사자	0.112	11.275	.000a	0.508
		경상북도	3차 산업 종사자	0.077	10.729	.000a	0.494
비가정기반 기타통행	생성	대구시	3차 산업 종사자	0.341	18.918	.000a	0.862
			수용 학생 수	0.096	5.837		
		경상북도	3차 산업 종사자	0.442	15.607	.000a	0.796
			수용 학생 수	0.102	4.947		
	유인	대구시	3차 산업 종사자	0.439	16.636	.000a	0.667



구 분	모형	목적	변수	계수	t값	P-value	R <sup>2</sup>
		경상북도	3차 산업 종사자	0.552	24.130	.000a	0.778

자료 : 2016년 교통 관련 기초조사(Ⅱ) -여객기종점통행량(O/D) 현행화 및 장래 수요예측 -,  
2016. 12, 대구광역시

## 2) 국외 사례

국외에서 사용된 통행발생모형의 생성·유입별 변수 및 분석법과 관련해서는 주로 미국의 사례를 중심으로 검토하였다. 미국의 통행발생모형의 경우 통행생성에서는 대부분 P/A접근법에 기초한 카테고리 분석법을 적용하고 있으나 통행유입에서는 회귀분석을 이용한 사례가 많은 것으로 나타났다. 디트로이트의 경우 가정기반 출근통행에서는 생성과 유입 모두에서 회귀분석을 사용하고 있으나 나머지 통행에서는 교차분류분석과 회귀분석을 생성과 유입 별로 나누어 사용하고 있다. 피츠버그의 경우에는 생성에서는 교차분류분석법을 사용하고 있으나 유입에서는 회귀분석이 아닌 원단위법을 주로 사용하고 있는 것으로 나타났다. 달라스-포트워스(Dallas-Fort Worth)의 경우 생성과 유입에서 모두 교차분류분석법을 주로 사용하고 있는 것으로 나타났으며 유입에서는 교차분류에 따른 원단위법을 적용하고 있다. 미국의 통행발생모형 적용변수는 국내 사례와 마찬가지로 주로 소득수준, 가구크기, 고용자수, 종사자수 등의 사회경제지표가 많이 이용되고 있으며, 일부 토지이용형태별 지역구분과 같은 토지이용 측면의 변수를 활용하고 있는 것으로 보인다.

<표 2-5> 해외 도시별 통행발생모형 변수

구 분	통행목적	생성	유입	모형형태
디트로이트	가정기반 출근통행	소득, 생활주기, 인구를 근거로 추정한 취업자수	총고용자수	생성/유입 : 회귀모형



구 분	통행목적	생성	유입	모형형태
	가정기반 쇼핑통행	가구크기, 가구별 자동차 보유수준	총고용자수, 도매업 고용자수, 소매업 고용자수, 인구	생성 : 교차분류분석 유입 : 회귀모형
	가정기반 등교통행	가구크기, 가구별 생활주기	초·중·고·대 학교 등록자수	생성 : 교차분류분석, 유입 : 회귀모형
	가정기반 기타통행	가구크기, 가구별 자동차보유대수	비기초산업 고용자수, 인구	생성 : 교차분류분석, 유입 : 회귀모형
	비가정기반통행		총고용자수, 도매업 고용자수, 소매업 고용자수, 인구	유입 : 회귀모형
	트럭 통행		기초산업 고용자수, 비기초산업 고용자수, 도매업 고용자수, 소매업 고용자수, 인구	유입 : 회귀모형
피츠버그	가정기반 출근통행	가구별 소득수준, 가구별 승용차보유수준, 저밀도와 고밀도의 지역구분	가구수 및 직종별 종사자수	생성 : 교차분류분석, 유입 : 원단위법
	가정기반 기타통행	가구별 소득수준, 가구별 승용차보유수준, 저밀도와 고밀도의 지역구분	가구수 및 직종별 종사자수	생성 : 교차분류분석, 유입 : 원단위법
	비가정기반 통행 도착		가구수 및 직종별 종사자수	유입 : 원단위법



구 분	통행목적	생성	유입	모형형태
	트럭통행 발생/도착	가구수 및 직종별 종사자수, 소형·중형·대형 의 트럭분류	가구수 및 직종별 종사자수, 소형·중형·대 형의 트럭분류	교차분류 원단위법
달라스- 포트워스	가정기반 출근통행	가구크기, 소득수준	소득수준, 직종별 고용자수, 토지이용형태별 지역구분	생성/유입 : 교차분류분석
	기정기반 기타통행	가구크기, 소득수준	직종별 고용자수, 토지이용형태별 지역구분	생성/유입 : 교차분류분석
	비가정기반 통행	가구크기, 소득수준	직종별 고용자수, 토지이용형태별 지역구분	생성 : 교차분류분석, 유입 : 교차분류 원단위법
	기타통행	직종별 고용자수, 토지이용형태별 지역구분	직종별 고용자수, 토지이용형태별 지역구분	생성 : 교차분류 원단위법, 유입 : 교차분류원단위법

### 2-3. 연구의 차별성

앞선 절에서는 본 연구의 이론적 배경 및 선행 연구를 검토하여 기존 통행 발생모형을 개선시키기 위한 적용 변수 및 방법론의 다양한 시도를 고찰해보았다. 통행발생모형을 개선하기 위한 변수의 다양화한 측면에 연구에서는 기존 사회경제변수 이외에 토지이용 변수인 용도지역면적, 용도지역별 연상면적, 토지이용복합도 등의 변수와 물리적 변수 등을 적용한 통행목적별 통행 발생예측의 정확도를 개선한 연구가 주를 이루었다. 또한 분석기법의 다양화





측면에서는 주로 카테고리분석법, 교차분류분석법, 신경망모형 등을 이용하여 모형의 정확도를 개선시키고자 한 연구가 있었다.

본 연구에서 주요 분석방법론으로 다루어지는 공간계량모형과 관련한 연구에서는 기존 선형회귀분석이 가진 통행발생모형의 통계적 가정의 한계 및 통행의 공간적 상관성을 미반영한 문제를 개선시키기 위하여 다양한 공간회귀분석 기법을 활용한 통행발생모형을 구축하여 기존 선형회귀를 통한 통행발생모형과 비교하여 나은 추정력의 개선결과를 제시하였다. 이와 같은 선행연구를 바탕으로 본 연구에서는 통행발생에 영향을 주는 변수들을 보다 다양화하고 빅 데이터를 활용해보는 측면에서 CDR을 주요 설명변수로 적용한 통행발생모형을 구축해보고자 한다. 또한 이러한 CDR 변수를 고려하여 기존 선형회귀분석의 통행발생모형에서 고려하지 못한 통행의 공간상관성을 감안한 GWR 기법을 이용하여 새로운 통행발생모형을 개발하여 모형의 신뢰성 개선을 위한 연구를 진행하고자 한다.

앞서 언급된 본 연구의 차별성에 대해 요약하면 다음과 같다.

첫째, 선행 연구에서 통행발생 영향변수로 다루지 않은 다양한 영향변수들을 적용해보고 그 중에서도 특히 CDR을 통행발생모형의 주요 설명변수로 적용하여 본다.

둘째, 일반적으로 통행발생모형에서 사용되는 선형회귀분석(OLS) 기법이 가진 통행의 공간적 상관성 문제를 통제하기 위해 국지적으로 회귀분석이 가능한 GWR을 이용하여 공간성이 감안된 통행발생모형을 구축해보고자 한다. 이때 앞서 검증된 CDR을 이러한 공간상관성에 기반한 통행발생모형의 주요 변수로 적용해봄으로서 향후 교통수요추정에 있어 이러한 빅 데이터의 활용성 측면에 정책적 시사점을 제시하고자 한다.





## 제3장 자료의 구성 및 연구방법론

### 3-1. 자료의 구성

#### 3-1-1. 자료의 수집

본 연구에서는 통행발생에 영향을 주는 요인들을 광범위하게 탐색해보기 위하여 취득 가능한 변수의 다양한 빅데이터를 수집하여 분석에 이용하였다. 종속변수로 사용되는 O/D는 KTDB에서 제공하는 2016년 기준 대구광역시 139개 행정동을 기준으로 한 전수화된 통행목적별 통행발생(유입)량 데이터를 수집하여 분석에 활용하였다.

주요 설명변수로 이용되는 CDR은 SKT 통신사의 대구광역시 요일별(주중, 주말) 각 2일간 24시간 데이터로서 10~60대 이상의 10대 급간의 연령별 자료를 기준으로 하였다. 설명변수로 통행발생에 영향을 미치는 영향요인을 다양하게 알아보기 위하여 토지이용지표 관련 변수의 경우 공공데이터포털 사이트를 이용하여 GIS건물정보제공서비스, 개별공시지가, 용도지역·지구, 개별주택 가격정보, 공동주택가격정보 데이터로부터 용도지역·지구, 건물대지면적, 건물연면적, 용적율, 건폐율, 공시가격 등의 세부 속성데이터를 수집하였다.

또한 사회경제지표 관련 변수자료는 KTDB에서 O/D 전수화 과정에 실제 이용하는 인구수(연령별), 취업자수, 총종사자수, 3차산업종사자수, 학원종사자수, 총학생수 등 실제 O/D 전수화 및 현행화에 직접 사용하였던 데이터를 수집하였다. 다음으로 물리적지표 관련 변수 자료는 대구광역시 내부자료인 교통유발시설(시장, 병원, 백화점, 터미널 등) 현황과 KTDB에서 배포된 네트워크 자료인 도로폭원, 도로등급별 링크길이, 신호등수, 버스정류장, 지하철(철



도)역 수 등과 같은 대중교통 결절점(node) 등의 자료를 수집하였다.

사용된 종속·설명변수 자료들은 실제로 우리나라 전체의 자료를 다양한 형태의 속성과일로 제공하고 있으나 본 연구에서는 분석을 위해 연구 대상의 공간적 범위인 대구광역시를 기준으로 데이터를 한정하였으며, 분석에 필요한 공간적 단위(행정동)로 데이터 전처리(data preprocessing) 과정을 통해 데이터마이닝과 집계화(aggregation)하는 과정을 거쳤다. 실제 CDR과 각종 설명변수 관련 빅데이터 자료들은 UTM X-Y좌표계 기준으로 제공되어 훨씬 세분화된 분석이 가능함에도 불구하고 본 연구에서 사용된 변수별 빅데이터의 공간적 단위를 소존(행정동) 규모로 집계화한 것은 기존 통행발생모형 도출에 공간적 집계단위인 교통존(TAZ)이 일반적으로 행정동 단위로 설정됨에 따라 이를 비교·분석하기 위한 분석의 용이성 차원에서 이를 집계화하여 사용하였음을 밝힌다.

<표 3-1> 자료의 수집

구분	자료명	자료구성	세부내용	제공 연도	출처
종속 변수	O/D	기준연도 목적O/D 통행발생 · 유입량	<ul style="list-style-type: none"> <li>· 대구광역시</li> <li>· 가정기반 통근,통학</li> <li>· 가정/비가정기반 학원,쇼핑, 업무, 기타</li> </ul>	2016	KTDB
설명 변수	통신량 자료 (CDR)	요일별(주중,주말) 시간대별(24시간) 연령별(10~60대이상)	<ul style="list-style-type: none"> <li>· '17.4.8~9 / 4.12~13</li> <li>· 1시간단위</li> <li>· 10~60대 이상 (10대 급간)</li> </ul>	2017	SKT



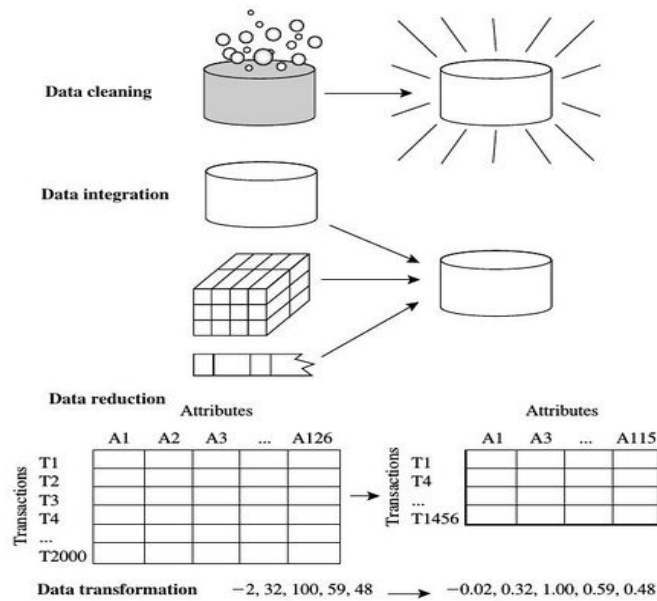
구분	자료명	자료구성	세부내용	제공 연도	출처
	사회 경제 지표	인구 종사자수 학생수 등	<ul style="list-style-type: none"> <li>· 인구수(총·연령 별), 취업자수</li> <li>· 총종사자수, 3차산업종사자구, 학원종사자수</li> <li>· 초중고 수용학생수, 총학생수</li> </ul>	2015	센서스
	토지 이용 지표	GIS건물정보 개별공시지가 용도지역, 용도지구 건물연면적, 용적율 개별주택가격정보 공동주택가격정보	<ul style="list-style-type: none"> <li>· 건물대지면적, 건물연면적, 용적율, 건폐율, 건물높이, 지상층수, 지하층수, 총주차수, 총주차장, 총주차장면적, 공동주택 평균공시가격, 개별주택 가격</li> <li>· 용도지역·지구별 면적(대구시)</li> </ul>	2017	공공 데이터 포털
	물리적 지표	교통유발시설 도로등급 철도역·버스정류장	<ul style="list-style-type: none"> <li>· 시장, 대형마트, 준종합병원, 백화점, 초중고대학교, 관공서 등</li> <li>· 도로종별, 도로폭원, 버스정류장, 철도역 등 대중교통결절점</li> </ul>	2017	대구시 내부 자료, KTDB

### 3-2. 데이터 전처리 (Data Preprocessing)

데이터 전처리 과정은 데이터를 탐색하고 데이터 마이닝을 위해 적당한 형태로 변환시켜 주는 과정을 의미하는 것으로 전처리 과정은 노이즈(noise) 데이터나 중복된 데이터를 제거하거나 분산되어 있는 데이터들을 통합하는 과정을 포함하여 필요한 경우 정규화, 이산형화 또는 수학적 변환과정을 의미한다. 즉, 데이터를 분석이 가능한 형태로 추출하거나 차원을 변환하는 작업



을 통칭하는 의미로서 빅데이터를 다룰 때 필수적으로 거쳐야 하는 중요한 과정이다.

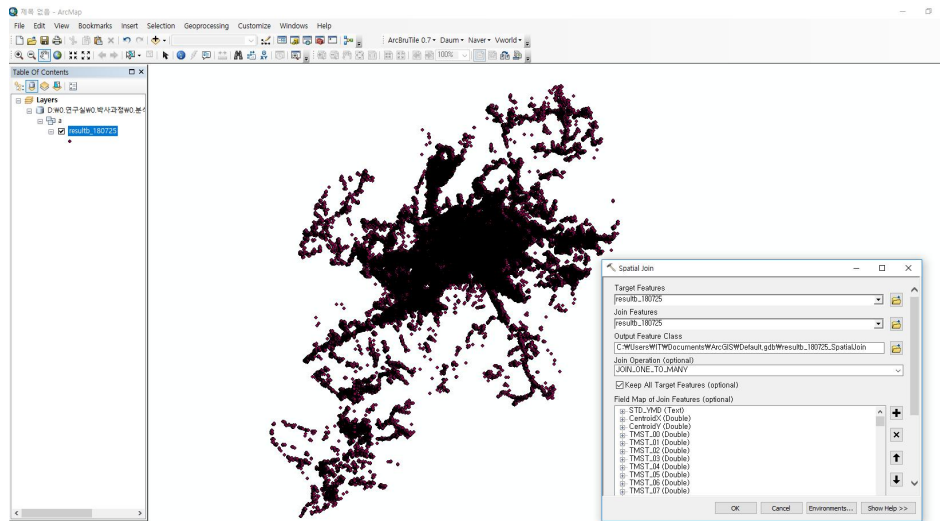


<그림 3-1> 데이터 전처리의 형태<sup>3)</sup>

본 연구에서 이러한 데이터 전처리과정이 필요한 이유는 변수로 사용되는 대용량의 CDR과 토지이용변수 가운데 GIS건물정보, 개별 및 공동주택가격정보와 같은 빅데이터들을 분석이 가능한 수준으로 집계화 또는 변환하는 과정이 필요하기 때문이다. 이에 CDR과 해당 변수들에 대한 데이터를 ArcMAP 10.4의 Spatial join과 MSSQL을 이용하여 데이터 분석이 가능하도록 전처리 과정을 수행하였다. 먼저 CDR 데이터의 XY좌표 기준으로 건물정보, KTDB의 네트워크 정보를 ArcGIS상에서 Spatial Join을 통해 자료를 결합하였다.

3) 데이터 마이닝 개념과 기법(제2판), 지아웨이 한 외, 2016





<그림 3-2> ArcGIS 10.4, Spatial join

다음으로 이렇게 결합된 자료들로부터 MSSQL을 이용하여 본 연구에서 분석이 가능하도록 행정동 소준 단위로 집계화하고 각 변수들을 적용 가능하도록 추출하고 계산하는 과정을 거쳤다.

```

SELECT TOP (100) PERCENT 기준.DISTRICT_N, 기준.구분일, 기준.age, DATA_ALL.X_COORDINA,
DATA_ALL.Y_COORDINA, DATA_ALL.오전월동TMS, DATA_ALL.오후월동TMS, DATA_ALL.공시지가,
DATA_ALL.면적, DATA_ALL.건물대지면적, DATA_ALL.건물건축면적, DATA_ALL.건물연면적,
DATA_ALL.용적율, DATA_ALL.건폐율, DATA_ALL.건물높이, DATA_ALL.종주차수, DATA_ALL.종주차장면적,
DATA_ALL.건물산정연면적, DATA_ALL.과년도1주택가격, DATA_ALL.평균공시가격, DATA_ALL.전체공시가격,
DATA_ALL.과년도1평균공시가격, DATA_ALL.평균링크걸이, DATA_ALL.최대링크걸이, DATA_ALL.평균도로폭,
DATA_ALL.평균차선수, DATA_ALL.평균신호등수, DATA_ALL.평균신호등밀도, DATA_ALL.전체인구,
DATA_ALL.인구15세이하, DATA_ALL.인구5_19세, DATA_ALL.인구15_19세, DATA_ALL.인구5_24세,
DATA_ALL.인구20_29세, DATA_ALL.취업자, DATA_ALL.종종사자수, DATA_ALL.3차산업종사자,
DATA_ALL.학원종사자, DATA_ALL.초중고, DATA_ALL.대학생, DATA_ALL.중학생, DATA_ALL.교통유발시설1,
DATA_ALL.교통유발시설2, DATA_ALL.교통유발시설3, DATA_ALL.교통유발시설4, DATA_ALL.교통유발시설5,
DATA_ALL.교통유발시설6, DATA_ALL.교통유발시설7, DATA_ALL.교통유발시설8, DATA_ALL.교통유발시설9,
DATA_ALL.교통유발시설10, DATA_ALL.교통유발시설11, DATA_ALL.교통유발시설12,
DATA_ALL.교통유발시설13, DATA_ALL.교통유발시설14, DATA_ALL.교통유발시설15,
DATA_ALL.교통유발시설16, DATA_ALL.정류장수, DATA_ALL.고속도로, DATA_ALL.도시고속화도로,
DATA_ALL.일반국도, DATA_ALL.특별/광역시도, DATA_ALL.국가차원지방도, DATA_ALL.지방도,
DATA_ALL.시군도, DATA_ALL.연결램프, DATA_ALL.미구분지역, DATA_ALL.근린상업지역,
DATA_ALL.보전녹지지역, DATA_ALL.생산녹지지역, DATA_ALL.유통상업지역, DATA_ALL.일반공업지역,
DATA_ALL.일반상업지역, DATA_ALL.자연녹지지역, DATA_ALL.제1종일반주거지역,
DATA_ALL.제1종전용주거지역, DATA_ALL.제2종일반주거지역, DATA_ALL.제2종전용주거지역,
DATA_ALL.제3종일반주거지역, DATA_ALL.종공업지역, DATA_ALL.준주거지역, DATA_ALL.중심상업지역,
DATA_ALL.철도, DATA_ALL.도시철도, DATA_ALL.경위도_X, DATA_ALL.경위도_Y
FROM (SELECT DISTINCT A.DISTRICT_N, B.구분일, C.age
FROM (SELECT DISTINCT DISTRICT_N FROM dbo.t_sk) AS A CROSS JOIN
(SELECT CASE WHEN STD_YMD IN ('20170408', '20170409')
THEN '주말' ELSE '평일' END AS 구분일
FROM dbo.t_sk AS t.sk_4) AS B CROSS JOIN
(SELECT DISTINCT age
FROM dbo.t_sk AS t.sk_3) AS C) AS 기준 LEFT OUTER JOIN

```

<그림 3-3> MSSQL Query



### 3-3. 변수별 데이터 개념정의 및 구축

#### 3-3-1. 종속변수

##### (1) 통행목적별 O/D

본 연구에서 종속변수로 사용되는 O/D를 통행목적의 성격별로 가정기반 통근, 통학과 같은 비교적 고정적이고 반복적인 성격의 통행을 필수적 통행(Mandatory trip)으로 구분하고 이와 반대의 속성을 가진다고 볼 수 있는 가정, 비가정기반 학원, 쇼핑, 업무, 기타와 같은 비교적 고정적이고 불규칙적인 성격을 갖는 통행을 선택적 통행(Optional trip)으로 크게 2가지로 구분하였다. 이는 주요 설명변수로 사용되는 CDR데이터가 주중과 주말, 시간대별 자료만을 가지고 있는 속성의 한계에 따라 통행목적별 기존 O/D의 통근, 통학, 등교, 귀가, 쇼핑, 여가, 기타 통행목적 등과 같이 세부적으로 구분하지 못하는 이유에 기인하여 통행의 성격을 상식적 수준에서 필수적 통행(Mandatory trip)과 선택적 통행(Optional trip)으로 크게 2가지로 구분하여 범주화시킨 통행발생모형을 구축하고자 하기 위함이다.

<표 3-2> 종속변수의 개념정의

구분	통행목적	내용	비고
O/D	총통행 (Total Trip)	총통행발생유입량	KTDB '16년 기준
	필수적 통행 (Mandatory Trip)	가정기반 통근, 통학통행 발생유입량 (HBW, HBSC)	
	선택적 통행 (Optional Trip)	가정기반 학원, 쇼핑, 기타 비가정기반 쇼핑, 기타통행 발생유입량 (HBO, HBAC, HBSH, NHBSH, NHBO)	





### 3-3-2. 설명변수

#### (1) 통신량 자료(CDR)

본 연구에서 주요 변수로 사용될 통행량인 CDR은 SKT에서 취득한 데이터로서 대구광역시를 기준으로 연령대별(10~60대 이상), 시간대별(24시간), 주중·주말 2일간 자료로 구성되어 있다. 이 데이터의 기준 시점인 2017년 4월의 SKT는 전국 기준 총 가입자수 55,332,064명 중 24,511,118명으로 전체 44.3%의 점유율을 가지고 있는 것으로 나타났다.<sup>4)</sup>

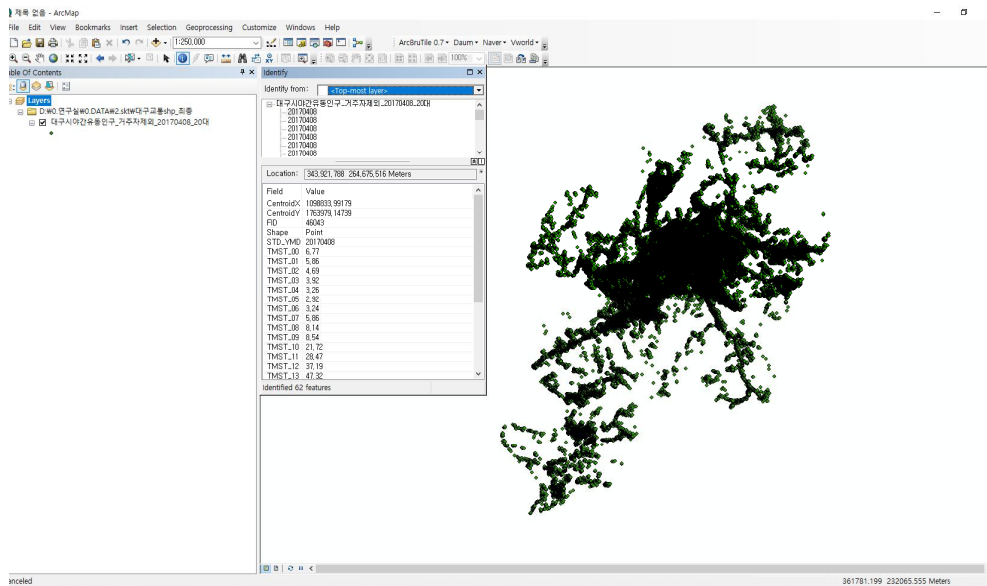
통신데이터는 ArcGIS에서 사용 가능하도록 .shp파일로 제공되어졌으며, 50×50m cell의 point로 각 연령대별 통신량 데이터가 속성값으로 내포되어 있는 형태이다. 세부적인 데이터 설명은 아래 표3-3과 같다.

<표 3-3> 통신자료의 구성

구분	명칭	유형	세부내용	예시	비고
1	std ymd	string	연월일 (주말 2017.4.8.~4.9 주중 2017.4.12.~4.13)	20170408	
2	X좌표·Y좌표	Number	UTM-K Coordinate system	1077483.991 79	50×50m cell point
3	TMST 00~23	Number	1시간대별 통신량	0.12	소수점 2자리

4) 미래창조과학부, 무선서비스 통계현황(2017년도), 이동전화 용도별 회선수(2017.4월 기준)





<그림 3-4> 통신자료의 구성(ArcGIS .shp)

CDR은 SKT에서 설정한 50m×50m 단위의 셀(cell)인 pCell기준으로 구성되어 있다. pCell은 개별 이동통신 단말기에 관여하는 기지국 데이터로부터 위치를 측정하여 정의된 pCell 중 가장 가까운 값에 사상(mapping)하여 생성된다. 개인식별코드가 없이 각 셀에 누적 정보가 기록되는 방식이므로, 셀 간의 이동정보는 직접 추출할 수 없는 구조이다(김경태 외, 2016).

각 셀에 기록된 통신량은 개인의 통신기록으로서 음성(통화), 문자, 데이터(인터넷)을 이용한 경우가 포함되며, 푸시(push) 서비스 등 수동적으로 발생하는 통신도 이에 포함이 된다. 이 통신량은 하루 중 해당 셀에서 이동통신 서비스를 이용한 기록이 있는 이용자의 통계수치라고 볼 수 있고 정확히는 회선에 대한 정보이지만 대체적으로 회선과 이용자가 동일하다는 가정 하에서 통신 자료를 통행이나 이동의 개념으로 다룰 수 있다고 생각된다.



따라서 본 연구에서 기준한 통신량 데이터는 1일 단위로 해당 셀에서 통신한 사람의 수로서 엄밀히 말해 각 셀에서 통신(문자, 통신 수·발신, 데이터 통신 기록 기준이며, wifi통신은 제외)한 휴대전화의 수라고 보면 된다. 사실상 이러한 CDR은 통행의 방향성 즉, 통행경로를 추적하기에는 다소 어려움이 있어 본 연구에서는 기존 O/D의 통행목적별 통행발생유입량을 기준으로 CDR과의 비교·분석을 통해 통행발생모형 구축에 활용하였다.

<표 3-4> 통신량 자료(CDR)의 개념적 정의

구분	통행목적	내용	비고
통신량 자료 (CDR)	총통행 관련 (Total)	<ul style="list-style-type: none"> <li>• 총통신량</li> <li>• 평일 오전·오후첨두(07~09, 17~19)</li> <li>• CDR total</li> </ul>	
	필수적 통행 관련 (Mandatory Trip)	<ul style="list-style-type: none"> <li>• 평일 오전첨두시 통신량</li> <li>• 평일 오전첨두(07~09)</li> <li>• CDR Mand MPeak</li> </ul>	
		<ul style="list-style-type: none"> <li>• 평일 오후첨두시 통신량</li> <li>• 평일 오후첨두(17~19)</li> <li>• CDR Mand EPeak</li> </ul>	
	선택적 통행 관련 (Optional Trip)	<ul style="list-style-type: none"> <li>• 주말 오전첨두시 통신량</li> <li>• 주말 오전첨두(09~12)</li> <li>• CDR Opt MPeak</li> </ul>	
		<ul style="list-style-type: none"> <li>• 주말 오후첨두시 통신량(CDR Opt)</li> <li>• 주말 오후첨두(14~16)</li> <li>• CDR Opt EPeak</li> </ul>	

## (2) 사회경제지표 변수

본 연구의 사회경제지표 변수의 경우 KTDB에서 O/D 전수화시 사용하는 자료와 동일한 데이터를 수집하여 이용하였다. 기존 선행연구에서 통행발생모형에 주로 사용된 사회경제지표에 대해 고찰하고 연령대별 인구와 학생수



등의 변수들 간의 다중공선성 문제 등을 적절히 고려하여 각 변수의 특성을 가장 잘 나타내는 지표를 선정하여 분석에 적용하였다.

<표 3-5> 사회경제지표 변수의 선정

구분	명칭	세부내용	예시	비고
1	인구	<ul style="list-style-type: none"> <li>총인구</li> <li>연령대별 인구 (15~19세, 20~29세)</li> </ul>	967	
2	종사자수	<ul style="list-style-type: none"> <li>총종사자수</li> <li>3차산업 종사자수</li> <li>학원종사자수</li> <li>취업자수</li> </ul>	13,487	
3	학생수	<ul style="list-style-type: none"> <li>초중고 학생수</li> <li>대학생수</li> </ul>	1,049	

### (3) 토지이용지표 변수

토지이용과 통행발생과의 관계는 상호 불가분의 관계를 가지며, 통행발생에 토지이용이 미치는 영향은 절대적이라고 볼 수 있다(윤대식, 2001). 토지이용은 접근성 뿐만 아니라 토지이용밀도, 용도지역·지구 등과 같은 요소들에 의해 통행발생을 유발 또는 집중시키는 아주 중요한 요인으로 작용되는 변수임은 이론의 여지가 없다. 결국 통행발생(유입)은 토지이용적 측면에서 도시의 복잡다단한 구조와 확장, 빠르게 변화되는 도시구조의 공간성을 감안하여 단순히 사회경제적인 변수만이 아닌 복합적인 공간적 영향에 작용을 받는 것은 사실이다. 이에 본 연구에서는 현재 국가적으로 공공데이터로 관리·구축되어져 일반에게 널리 공개되어 있는 토지이용에 관련된 건축물 또는 용도지역 자료를 수집하여 이를 GIS상에서 Spatial Join 및 MSSQL을 통해 주거·상업·업무시설의 용도지역별 건축물연상면적을 도출하였다. 또한 토지이용복



합도(Land Use Mix Diversity) 변수는 Baht & Guo(2007) 연구에서 제안된 아래 식3.1에 근거하여 본 연구에서 수집된 토지이용 용도별 면적 데이터를 활용하여 가공하여 계산하였다.

$$LI = 1 - \frac{\left| \frac{r}{L} - \frac{1}{4} \right| + \left| \frac{o}{L} - \frac{1}{4} \right| + \left| \frac{c}{L} - \frac{1}{4} \right| + \left| \frac{ot}{L} - \frac{1}{4} \right|}{5/4RIGHT} \quad \dots\dots (\text{식3.1})$$

여기서,  $L = r + o + c + ot$

$r$  : 주거용 건축물연상면적 (용도지역별 주거지역 면적( $m^2$ ))

$o$  : 업무용 건축물연상면적 (용도지역별 공업지역 면적( $m^2$ ))

$c$  : 상업용 건축물연상면적 (용도지역별 상업지역 면적( $m^2$ ))

$ot$  : 기타용도 건축물연상면적 (용도지역별 녹지지역 면적( $m^2$ ))

직주균형계수 변수의 경우, 직주균형상태를 나타내는 지표로서 Zhao et al.(2005)를 참고하여 정우현(2011)의 연구에서 사용된 지표의 계산식인 식3-2를 일부 수정하여 분석하였다.

$$JH_i = \frac{TPopW_i - TotWor_i}{TPopW_i + TotWor_i} \quad \dots\dots\dots (\text{식3.2})$$

여기서,  $TPopW_i$  :  $i$ 지역의 22~65세이하 인구

$TotWor_i$  :  $i$ 지역의 총고용자수

본 연구에서는 해당 존의 취업자수와 총인구수의 차이와 합계를 이용하는 개념으로서 상기 식3-2를 일부 수정하여 식3-3의 방식으로 직주균형계수를 도



출하였으며, 이 값이 양의 값을 가지거나 1에 가까운 존의 경우 고용이 상대적으로 적은 지역으로서 주거의 성격을 많이 가지는 성격을 보인다고 할 수 있고 반대의 경우 고용이 많은 지역으로서 상업·업무 성격을 많이 가지는 지역의 상태를 의미한다고 볼 수 있다.

$$JH_i = \frac{Pop_i - Emp_i}{Pop_i + Emp_i} \dots\dots\dots (식3.3)$$

여기서,  $Pop_i$  :  $i$  지역의 총인구수

$Emp_i$  :  $i$  지역의 총취업자수

<표 3-6> 토지이용지표 변수의 개념 정립 및 선정

구분	통행목적	내용	비고
토지 이용 변수 (Land Use)	토지이용 용도별 건축물 연면적(Land Use Floor Area)	<ul style="list-style-type: none"> <li>· 주거지역 건축물연상면적 (㎡)</li> <li>· 상업지역 건축물연상면적 (㎡)</li> <li>· 공업지역 건축물연상면적 (㎡)</li> </ul>	
	토지이용복합도 (Land Use Mix Diversity)	<ul style="list-style-type: none"> <li>· 토지이용복합도 지표(LI)</li> <li>· 지표값이 높을수록, 다양한 토지이용이 지역적으로 존재</li> <li>· Baht &amp; Guo(2007)</li> </ul>	
	직주균형계수 (Job-Housing Balance Index)	<ul style="list-style-type: none"> <li>· 해당 존의 직장과 주거지역의 존재</li> <li>· 자족도 개념으로 해석</li> <li>· Zhao et al.(2005), 정우현(2011)</li> </ul>	

#### (4) 물리적지표 변수

일반적으로 통행발생과 관련되어 주로 이용되는 사회경제지표 및 토지이용



지표 이외에도 본 연구에서는 물리적 지표로서 국가교통DB에서 제공하고 있는 도로망 및 철도망 GIS 네트워크 DB자료와 본 연구에서 구축한 대구광역시 교통유발시설 지점 자료 등을 이용하여 물리적인 교통유발시설 및 대중교통 이용 관련 시설과 같은 변수들을 도출하였다. 우선 교통유발시설의 경우 대구광역시 내부자료를 바탕으로 재래시장, 백화점, 대형마트, 준·종합·대학병원, 공항, 철도역, 버스터미널, 초·중·고등학교, 극장, 음식점, 관람시설 및 시청을 비롯한 행정기관 시설의 위치를 GIS상 점(point)으로 맵핑(mapping)하고 행정동 단위로 교통유발시설을 집계화한 데이터를 추출하였다. 교통유발시설을 정의함에 있어 유발시설의 특징별로 시장, 백화점, 대형마트 등을 판매시설로 병원을 의료시설, 공항·철도역·버스터미널을 교통시설, 초·중·고·대학교를 교육시설로 구분하는 등 총 6개의 유형별 교통유발시설을 구분하여 데이터를 구분하였다. 또한 대중교통과 관련한 시설로 버스정류장과 철도·지하철역수를 마찬가지로 행정동 단위로 집계화하여 분석에 활용하였다.

<표 3-7> 물리적지표 변수의 개념 정립 및 선정

구분	통행목적	내용		비고
물리적 변수 (Facility Factor)	유형별 교통유발 시설 (Derived Fac)	<ul style="list-style-type: none"> <li>· 판매시설(Dsales)</li> <li>· 의료시설(Dhospital)</li> <li>· 교통시설(Dterminal)</li> <li>· 교육시설(DMidHighUni)</li> <li>· 문화관광집회시설(DCultu)</li> <li>· 행정기관시설(Doffice)</li> </ul>	<ul style="list-style-type: none"> <li>· 시장, 백화점, 대형마트</li> <li>· 준종합 대학병원</li> <li>· 공항, 역, 버스터미널</li> <li>· 중고대학교</li> <li>· 극장, 음식점, 종교시설</li> <li>· 시청, 구군청 등</li> </ul>	
	버스 정류장수 (BusStop)	· 해당 존의 버스정류장 수 합계		
	철도 지하철역수 (railSub Station)	· 해당 존의 철도, 지하철역수 합계		



### 3-4. 연구방법론 정립

#### 3-4-1. 공간계량모형의 적용

앞서 설명한 바와 같이, 공간계량모형은 일반선형모형에서 공간가중치행렬을 부가한 형태로 일반선형모형의 확장된 형태라고 볼 수 있다. 결국 공간계량모형에서 가장 중요한 것은 공간가중치행렬을 어떻게 적용할 것인지가 핵심이라고 볼 수 있는데 그에 앞서 일반적으로 연구에서 다루어지는 종속변수인 O/D 통행발생(유입)량이 실제로 공간적으로 자기상관이 존재하는지에 대한 여부를 먼저 판단하여야 한다. 이때 공간자기상관은 전역적 Moran's I 방법을 이용하여 검증하고 이를 통해 변수의 공간종속성이나 이질성에 관련된 통계량 결과에 따라 분석을 시행하였다.

#### (1) 지리적 가중회귀모형(Geographically Weighted Regression)

지리적 가중회귀모형(GWR)은 Fortheringham 외(1999)에 의해 개발된 국지적인 공간지역 회귀모델의 집합체를 나타낸 것으로서 각 공간마다 회귀계수값을 기술하는 회귀분석방법이다. GWR은 공간 상에서 서로 다른 위치에 존재하는 데이터들에 대하여 각각 다른 회귀모형을 적용하여 공간적 비정상성을 탐색하는 방법이라고 하였다(김익희, 2005).

일반적으로 통행발생모형에서 사용되는 선형회귀분석(OLS)은 전체 연구대상 범위에서 변수별로 하나의 계수(coefficient)만을 제시하므로 특성에 관계없이 그 변수의 영향은 연구지역 전체에서 동일한 것을 해석하는 반면, GWR은 분석대상지역 내에서 개별 관측치에 대해 회귀분석 기준점을 선정하여 그 중심점으로부터 기준 거리 내에 있는 관측치들을 대상으로 수행하는 회귀분석





으로 지역별로 각각 다른 계수를 추정함으로써 변수들의 국지적인 영향력을 파악할 수 있는 장점이 있다. GWR이 일반 OLS와 가장 큰 차이점은 기존의 회귀분석에서 고려할 수 없는 인접 지역과의 공간적 상호작용을 반영한다는 점이다. GWR은 회귀지점에서 멀어질수록 값이 낮아지는 차별적인 가중치를 데이터들에 적용함으로써 ‘Tobler의 지리학 제1법칙’을 반영하고 있다. GWR의 식은 식3-4와 같은데, 이는 다중선형회귀분석에 좌표를 부여함으로써 모형을 지역적인 구조로 확장시키는 것을 의미한다.

$$Y_i = \sum_k \beta_k(u_i, v_i) x_{ik} + \epsilon_i \quad \dots\dots\dots (식3.4)$$

여기에서  $(u_i, v_i)$ 는 지역  $i$ 의 중심점의 좌표를 의미하며,  $\beta_k(u_i, v_i)$ 는 지역  $i$ 의  $k$ 번째 독립변수의 계수를 의미한다. 따라서 각 독립변수 별로 모든 회귀분석의 중심점에 대하여 각각 다른 계수들이 추정되며, 추정치는 식3-5에 의해 계산된다.

$$\hat{\beta}(u_i, v_i) = (X^T W(u_i, v_i) X)^{-1} X^T W(u_i, v_i) Y \quad \dots\dots\dots (식3.5)$$

$$\text{여기서, } W(u_i, v_i) = c \begin{bmatrix} W_{i1} & 0 & \dots & 0 \\ 0 & W_{i2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & W_{in} \end{bmatrix}$$

여기에서  $X$ 는  $(n \times k)$ 행렬,  $Y$ 는  $(n \times 1)$ 벡터이고  $W(u_i, v_i)$ 는  $(n \times n)$ 행렬이다.  $W(u_i, v_i)$ 의 대각선에 존재하는 값들은 좌표  $(u_i, v_i)$ 를 갖는 지역  $i$ 에 대해

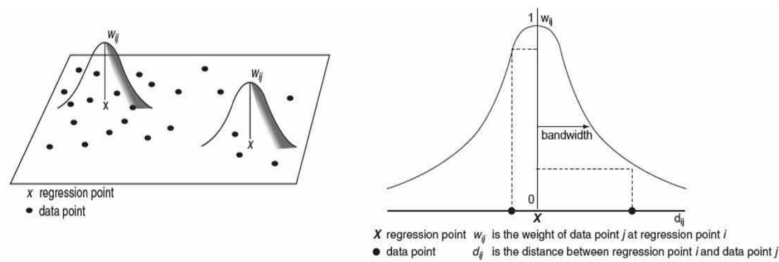


주변  $n$ 개 관측치의 지리적 가중치를 의미한다. Moran's I를 산출하기 위해 적용하는 공간가중치 행렬과 유사한 개념이지만 행렬의 형태가 다를 수 유의하여야 한다. 또한 분석시 공간상관성을 분석하기 위한 공간가중치행렬은 연구자가 외생적으로 정의 하는 것이지만, GWR에서의 공간가중치행렬은 내생적으로 추정된다(김진희 외, 2011). 공간가중치  $W_{ij}$ 는 식 3-6과 같이 Kernel함수로 정의할 수 있다.

$$W_{ij} = \exp(-\gamma d_{ij}^2) \dots\dots\dots (식3.6)$$

여기에서  $d_{ij}$ 는 지역  $i$ 의 중심점에서 지역  $j$ 의 중심점까지의 거리를 의미하며  $\gamma$ 는 Kernel Bandwidth를 의미한다. 이때 이 대역폭(bandwidth)은 회귀분석의 중심점이 영향을 미치는 국지적 공간상관범위를 뜻한다. 즉, 가중치  $W_{ij}$ 는 회귀점에서 주변의 데이터 점과의 거리  $d_{ij}$ 차이에 따라 결정되며, 인접할수록 큰 값이 부여된다. 이러한 커널함수를 사용하면 연속적이면서도 부드러운 모습을 띄는 가중치 값들을 얻을 수 있게 되고 실제의 현상을 잘 반영한다고 볼 수 있다. Kernel은 고정식(fixed)과 적응식(adaptive)의 2가지 방식이 있는데 가중치를 만드는 대역폭이 고정되어 있는 경우는 고정식 kernel이고, 사례 수에 따라 서로 다른 대역폭을 적용하는 경우는 적응식 Kernel이라고 한다.





<그림 3-5> Kernel function(Fotheringham et al., 2002)

모형의 부합도는 이 대역폭에 영향을 많이 받는데 연구지역에서 표본점들이 규칙적으로 분포하는 경우에는 고정식 kernel을 사용해도 되지만 불규칙한 경우에는 적응식 kernel을 사용하는 것이 좋으며 이 대역폭을 구하는 대표적인 방법은 AIC(Akaike Information Criterion)를 최소화하는 값을 찾는 AICc(AIC with a small sample correlation)방법과 CV(Cross-Validation)를 이용한 방법이 있으며, 각각에 대한 식은 다음 식3-7과 같다.

$$AICc = 2n \ln(\hat{\sigma}) + n \ln(2\pi) + n \left( \frac{n + \text{tr}(S)}{n - 2 - \text{tr}(S)} \right) \quad \dots\dots\dots (\text{식3.7})$$

여기에서  $\hat{\sigma}$ 는 회귀분석으로부터 표준화된 잔차제곱합을 의미하며  $S$ 는  $\hat{y} = Sy$ 를 만족시키는 hat matrix이며,  $\text{tr}(S)$ 는  $S$ 의 대각행렬원소의 합을 의미한다.

$$CV = \sum_{i=1} [y_i - \hat{y}_{-i}(b)]^2 \quad \dots\dots\dots (\text{식3.8})$$



CV를 이용한 방법론은 식3-8과 같이 관측치와 추정치의 차이 제곱합을 CV로 정의하고, 대역폭을 변화시키면서 CV가 감소하다 증가하는 변곡점에서 대역폭을 결정하게 하는 방법이다. 여기서,  $\hat{y}_{*1}(b)$ 는 기준거리  $b$ 에 있는 관측지점에 대한 종속변수의 추정치를 의미한다(Charlton & Fotheringham, 2003).

### 3-4-2. 연구의 진행과정

#### (1) O/D 및 CDR의 상관관계 분석

본 연구는 O/D를 바탕으로 통행목적별 통행발생모형을 구축함에 있어 CDR이 어떠한 특성을 가지고 있는지에 대해 살펴볼 필요가 있다고 판단하였다. 이를 위해 CDR이 실제 사람들이 이용하는 모바일 통신에 따른 이동성을 내포하고 있는 자료임을 증명하고 기존 O/D와 어느 정도의 유사성이 있는지를 살펴보는 것이 중요한 과정이라고 판단된다. 이에 두 자료의 직접적인 비교를 위해 기준연도 O/D자료와 같이 CDR을 기준 2일 단위 제공치에서 1일 단위 자료로 변환하고 자료 취득시점의 전국단위 44.3%의 통신사 점유율을 고려하여 보간법을 통해 100%로 전수화를 실시하였다. 또한 통행 단위의 통일성을 기하기 위해서 O/D의 통행/일 단위를 그대로 사용하되 통행목적별로 통행유출(production)·통행유입(attraction) 자료를 모두 활용하여 비교·분석하였다. CDR은 가상의 셀(pCell) 단위에서 집계되는 통신량이므로 이것이 통행발생인지 통행유입인지에 대한 정의가 모호함에 따라 이를 모두 적용시켜 비교결과를 살펴보고자 한다. 통행목적별 목적 O/D를 CDR과 비교하는 이유는 수단 O/D를 사용하였을 경우 재차인원 또는 교통량 등을 고려한 값이어야 하나, 실제 CDR은 사람의 이동에 바탕에 둔 개인의 모바일 통신 자료라는 의미에서 목적 O/D와의 비교가 보다 합리적일 것이라고 판단에 따른 것이다.



<표 3-8> O/D 및 CDR 검증자료 구분

구분	자료명	자료구성	세부내용	비고
Data	기종점 통행량 (O/D)	<ul style="list-style-type: none"> <li>· 기준연도 목적O/D</li> <li>· 읍면동 (소존단위)</li> </ul>	<ul style="list-style-type: none"> <li>· 총목적통행</li> <li>· 유출·유입통행</li> <li>· 가정·비가정기반 통근, 통학, 학원, 쇼핑, 업무, 기타</li> </ul>	2016
	통신량 자료 (CDR)	<ul style="list-style-type: none"> <li>· 평일, 주말</li> <li>· 읍면동 (소존단위)</li> </ul>	<ul style="list-style-type: none"> <li>· 1일 통행 변환</li> <li>· 점유율(44.3%) 고려 전수화</li> </ul>	2017

## (2) 선형회귀분석(OLS)을 통한 통행발생영향요인 및 기준모형 구축

본 연구에서 CDR을 활용한 통행발생모형을 도출함에 앞서 통행발생에 영향을 미치는 영향변수를 모두 적용한 선형회귀분석을 통해 기준모형을 구축할 필요가 있다. 다시 말해, 통행목적별 O/D 통행발생유입량을 종속변수로 하고 CDR을 미적용한 통행발생 기준모형을 도출하여 추후 CDR을 모형에 적용하였을 경우 모형의 설명력 및 개선정도에 대한 파악과 비교가 가능하다. 따라서 CDR을 미적용한 선형회귀분석으로 도출된 통행목적별 통행발생모형을 기준모형으로 도출하여 CDR을 적용시킨 대안별 통행발생모형과의 비교를 통해 CDR의 통행발생모형의 적용성에 대해 알아보고자 한다.

## (3) 지리적 가중회귀분석(GWR) 통한 통행발생모형 구축

통행발생·유입량에 대한 공간적 영향은 전역적(global)으로 존재할 수 있고 국지적(local)으로 존재할 수 있다. 본 연구에서 분석하고자 하는 교통존 단위의 통행발생·유입과 같은 경우 상식적으로 도시 전체에 전반적인 현상이라고



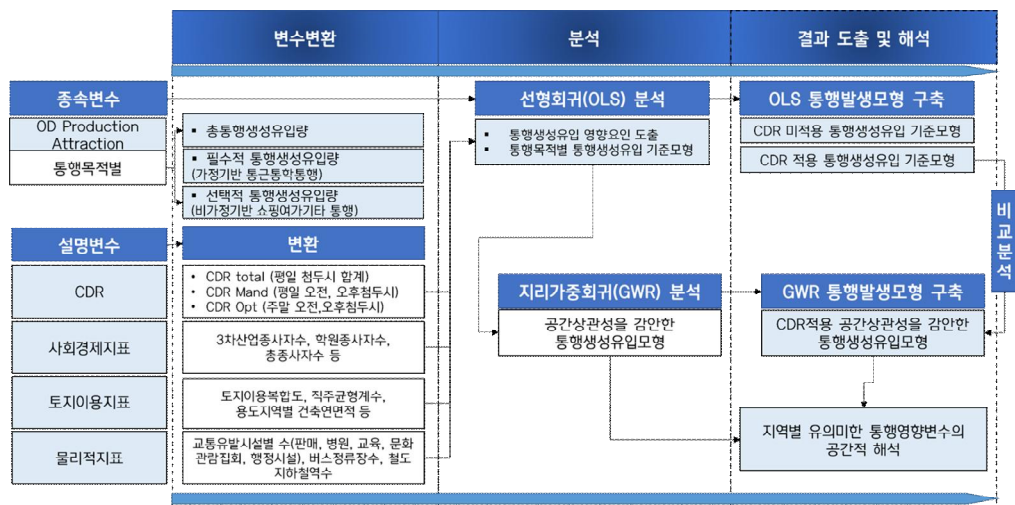
보기 보다는 통행발생에 영향을 주는 지역적이고 국지적 공간인—중심상업업무지구(C.B.D)나 부도심의 베드타운(bedtown)과 같은—특정지역 내에서 나타나는 현상으로 보는 것이 타당하다고 생각된다. GWR을 이용한 통행발생모형을 도출함에 있어 앞서 CDR이 주요 설명변수로 적용된 GWR에 기반한 통행발생모형으로부터 전체적으로 모형의 개선여부를 살펴보고 통행목적별로 CDR이 지역적 공간 내에서 유의미한 계수가 어떠한 공간적 분포를 가지고 있는지를 살펴봄으로서 실제 사람의 이동성이 대구시의 공간 구조상에 어떻게 변화하고 있는지에 대해서도 알아본다.

### 3-4-3. 분석과정의 정립

상기 언급된 연구 방법론, 타 연구와의 차별성 검토에 입각하여 본 연구의 분석과정은 다음과 같은 4단계의 과정으로 구분하여 요약한다.

첫 번째 단계에서는 CDR을 비롯한 다양한 설명변수를 통행발생모형의 변수로 활용하기 위하여 데이터를 집계화하고 변환하는 과정을 거친다. 두 번째 단계에서는 종속변수인 통행목적별 O/D통행발생유입량을 기준으로 CDR이 미적용된 선형회귀분석을 통해 통행발생 기준모형을 도출한다. 세 번째 단계에서는 앞선 단계의 통행발생 기준모형에 CDR을 적용하여 모형을 구축하여 보고 모형의 설명력의 적합도 개선결과를 살펴본다. 네 번째 단계에서는 CDR이 적용된 통행발생모형을 국지적 공간회귀모형인 GWR분석을 통한 통행발생모형을 구축하여 비교·분석결과에 대한 해석과 최적 모형을 도출한다. 마지막 다섯 번째 단계에서는 도출된 GWR분석 결과를 바탕으로 CDR이 연구 대상범위인 대구광역시의 공간 구조 내에서 어떠한 유의미한 분포를 보여주는 지에 대해 살펴보고 정책적 시사점을 논의해보고자 한다.





<그림 3-6> 연구 진행과정 및 방법론







## 제4장 통신 빅데이터 및 O/D 상관성 분석

### 4-1. 분석개요

#### 4-1-1. 비교 데이터 구성

본 절에서는 CDR을 통행발생모형에 적용하기에 앞서 O/D 자료와의 비교·검증을 통해 모바일 통신량이 사람들이 ‘통행’이나 이동성과 관련된 자료로서의 데이터 특성을 우선 살펴보고자 한다. 이에 CDR 자료의 속성 상 그 자체의 통행목적별 알 수 없기 때문에 기존 O/D의 통행목적별 구분을 요일별 CDR과의 비교를 통해 행정동 기준으로 집계된 통행량(量)적인 측면에서 상관분석과 대응표본 t-test를 실시하고 CDR과 O/D의 공간적 분포에 대한 비교를 위해 Local Moran's I와 Getis-Ord G\*의 국지적 공간성 분석을 통해 두 자료의 상관관계를 알아보하고자 한다. 구체적으로 통행목적별 데이터의 비교·분석은 CDR에서 필수적 통행(Mandatory Trip) 범주에 속성을 가진다고 볼 수 있는 평일기준 오전·오후첨두시 데이터와 선택적 통행(Optional Trip)의 속성을 가진다고 볼 수 있는 주말기준 오전·오후 첨두시 데이터를 가지고 O/D의 가정기반 통근·통학과 같은 필수적 통행(Mandatory Trip)과 가정·비가정기반 업무, 쇼핑, 여가, 기타 등의 선택적 통행(Optional Trip)과의 비교·분석을 시행하였다.



<표 4-1> 데이터 비교분석 개념 정의

구분	CDR	O/D	
총통행량	<ul style="list-style-type: none"> <li>• 통신량(CDR total)</li> <li>• 평일 오전·오후첨두 (07~09, 17~19)</li> <li>• CDR total</li> </ul>	통행발생(O/D P)	
		통행유입(O/D A)	
필수적 통행 (Mandatory Trip) 관련	<ul style="list-style-type: none"> <li>• 통신량(CDR Mand)</li> <li>• 평일 오전첨두(07~09)</li> <li>• CDR Mand MPeak</li> </ul>	발생 (O/D P Mand)	가정기반 통근통학 (HBW, HBSC)
	<ul style="list-style-type: none"> <li>• 통신량(CDR Mand)</li> <li>• 평일 오후첨두(17~19)</li> <li>• CDR Mand EPeak</li> </ul>	유입 (O/D A Mand)	
선택적 통행 (Optional Trip) 관련	<ul style="list-style-type: none"> <li>• 통신량(CDR Opt)</li> <li>• 주말 오전첨두(09~12)</li> <li>• CDR Opt MPeak</li> </ul>	발생 (O/D P Opt)	가정기반 학원, 쇼핑, 기타, 비가정기반 쇼핑, 기타 등 (HBO, HBAC, HBSH, NHBSH, NHBO)
	<ul style="list-style-type: none"> <li>• 통신량(CDR Opt)</li> <li>• 주말 오후첨두(14~16)</li> <li>• CDR Opt EPeak</li> </ul>	유입 (O/D A Opt)	

## 4-2. 분석결과

본 절에서는 앞서 언급한 CDR의 특성을 파악하기 위하여 O/D 자료와의 통행량(量)과 국지적 공간적 분포를 직접 비교하기 위한 분석을 실시하고자 한다. CDR과 O/D의 통행량(量) 비교의 경우, 행정동을 기준으로 집계화한 CDR의 요일별 집계량과 O/D의 총 통행, 필수적통행, 선택적통행 발생·유입량과의 상관관계 분석과 동일 모집단의 사전·사후점수의 평균을 비교할 때 주로 사용되는 대응표본 t-test를 실시하여 두 데이터의 차이성 검증에 대해 살펴본다.



#### 4-2-1. 통행량 기준 통행목적별 상관성 분석 (행정동 기준)

##### (1) 총 통행량(Total Trip) 기준

행정동 단위로 집계화된 CDR과 O/D 총통행발생량 합계를 비교해본 결과, CDR은 6,137,942통행, O/D는 5,994,680통행으로 약 2.3% 차이를 보이는 것으로 나타났다. 구·군단위로 집계된 결과를 살펴보면 수성구, 남구, 중구의 경우 44.8~49.7%의 높은 오차율을 보였으나 북구와 동구, 달성군의 경우 10.1~19.4%로 비교적 낮은 오차율을 나타내었다. 특히 북구 국우동, 수성구 범물1·2동, 고산3동과 같은 경우에는 O/D 총통행발생량이 CDR보다 훨씬 많은 것으로 나타나나 남구 이천동, 대명2동의 경우 CDR이 O/D 총통행발생량보다 많은 것으로 나타났다. 이러한 결과는 주로 대규모 공동주택단지가 밀집한 지역일수록 O/D 총통행발생량이 CDR보다 높은 결과를 보이고 반대로 공동주택이나 주거지 밀집이 낮은 지역인 경우 O/D 총통행발생량이 CDR보다 낮은 결과를 보이는 것을 알 수 있다.

<표 4-2> 행정동 기준 CDR과 O/D 총통행발생량 비교

구분		CDR total (A)	O/D P (B)	차이 (A-B)	오차율 ((A-B/A)* 100))
중존 (시군구)	소존 (읍면동)				
중구	동인동	73,538	85,034	-11,495	-15.6%
중구	삼덕동	80,025	53,840	26,184	32.7%
중구	성내1동	248,787	105,982	142,805	57.4%
중구	성내2동	202,649	107,186	95,463	47.1%
중구	성내3동	41,783	16,540	25,244	60.4%
중구	대신동	90,099	31,902	58,197	64.6%
중구	남산1동	46,517	16,583	29,933	64.3%
중구	남산2동	47,847	23,606	24,241	50.7%
중구	남산3동	14,554	15,830	-1,276	-8.8%



구분		CDR total (A)	O/D P (B)	차이 (A-B)	오차율 ((A-B/A)* 100))
중존 (시군구)	소존 (읍면동)				
중구	남산4동	37,226	17,681	19,545	52.5%
중구	대봉1동	78,051	57,349	20,702	26.5%
중구	대봉2동	8,593	3,639	4,953	57.6%
중구	소계	969,670	535,173	434,497	44.8%
동구	신암1동	26,794	22,632	4,162	15.5%
동구	신암2동	25,749	15,527	10,222	39.7%
동구	신암3동	27,883	19,078	8,806	31.6%
동구	신암4동	56,783	48,202	8,581	15.1%
동구	신암5동	12,900	15,220	-2,320	-18.0%
동구	신천1·2동	28,440	21,110	7,331	25.8%
동구	신천3동	51,839	50,209	1,630	3.1%
동구	신천4동	90,035	37,716	52,319	58.1%
동구	효목1동	26,311	25,757	554	2.1%
동구	효목2동	66,191	35,963	30,229	45.7%
동구	도평동	4,547	6,860	-2,312	-50.9%
동구	불로·봉무동	35,461	64,961	-29,500	-83.2%
동구	지저동	10,505	16,382	-5,877	-55.9%
동구	동춘동	14,678	23,657	-8,979	-61.2%
동구	방춘동	21,395	40,450	-19,054	-89.1%
동구	해안동	13,560	25,397	-11,837	-87.3%
동구	공산동	27,178	28,949	-1,772	-6.5%
동구	안심1동	40,492	99,232	-58,740	-145.1%
동구	안심2동	18,386	24,740	-6,355	-34.6%
동구	안심3·4동	65,224	121,535	-56,310	-86.3%
동구	소계	664,352	743,576	-79,224	-11.9%
서구	내당1동	31,405	22,984	8,422	26.8%
서구	내당2·3동	26,076	28,988	-2,912	-11.2%
서구	내당4동	71,459	37,451	34,008	47.6%
서구	비산1동	6,585	7,578	-993	-15.1%
서구	비산2·3동	15,643	18,068	-2,425	-15.5%
서구	비산4동	16,040	17,401	-1,361	-8.5%



구분		CDR total (A)	O/D P (B)	차이 (A-B)	오차율 ((A-B/A)* 100))
중존 (시군구)	소존 (읍면동)				
서구	비산5동	15,281	8,046	7,235	47.3%
서구	비산6동	11,009	7,486	3,523	32.0%
서구	비산7동	74,195	38,736	35,459	47.8%
서구	평리1동	9,360	20,876	-11,517	-123.0%
서구	평리2동	4,515	12,635	-8,121	-179.9%
서구	평리3동	26,532	28,716	-2,184	-8.2%
서구	평리4동	34,077	38,673	-4,597	-13.5%
서구	평리5동	12,688	12,416	271	2.1%
서구	평리6동	18,635	13,742	4,893	26.3%
서구	상중이동	133,484	87,173	46,311	34.7%
서구	원대동	48,206	20,735	27,471	57.0%
서구	소계	555,189	421,704	133,485	24.0%
남구	이천동	67,284	22,810	44,474	66.1%
남구	봉덕1동	41,452	24,127	17,325	41.8%
남구	봉덕2동	48,679	22,571	26,108	53.6%
남구	봉덕3동	87,300	36,303	50,997	58.4%
남구	대명1동	15,826	21,007	-5,181	-32.7%
남구	대명2동	80,181	28,726	51,455	64.2%
남구	대명3동	69,600	39,319	30,280	43.5%
남구	대명4동	38,307	36,613	1,694	4.4%
남구	대명5동	52,077	32,993	19,084	36.6%
남구	대명6동	24,533	16,199	8,334	34.0%
남구	대명9동	55,105	30,106	24,999	45.4%
남구	대명10동	44,286	20,036	24,250	54.8%
남구	대명11동	69,493	28,003	41,490	59.7%
남구	소계	694,123	358,814	335,309	48.3%
북구	고성동	37,329	12,537	24,792	66.4%
북구	칠성동	161,232	126,611	34,621	21.5%
북구	침산1동	20,061	28,374	-8,312	-41.4%
북구	침산2동	67,983	32,582	35,401	52.1%
북구	침산3동	83,206	51,381	31,825	38.2%



구분		CDR total (A)	O/D P (B)	차이 (A-B)	오차율 ((A-B/A)* 100))
중존 (시군구)	소존 (읍면동)				
북구	산격1동	11,277	15,071	-3,794	-33.6%
북구	산격2동	77,670	79,871	-2,200	-2.8%
북구	산격3동	56,634	73,344	-16,709	-29.5%
북구	산격4동	17,236	21,096	-3,861	-22.4%
북구	북현1동	26,733	16,346	10,387	38.9%
북구	북현2동	50,913	59,533	-8,621	-16.9%
북구	대현동	42,162	33,054	9,108	21.6%
북구	검단동	22,323	38,979	-16,657	-74.6%
북구	무태조야동	32,484	63,011	-30,527	-94.0%
북구	태전2동	25,937	37,798	-11,861	-45.7%
북구	관문동	51,628	52,671	-1,043	-2.0%
북구	읍내동	18,510	48,460	-29,950	-161.8%
북구	관음동	8,165	24,099	-15,933	-195.1%
북구	태전1동	30,919	50,308	-19,389	-62.7%
북구	구암동	17,764	58,772	-41,008	-230.8%
북구	노원동	128,172	72,468	55,704	43.5%
북구	동천동	28,136	90,781	-62,644	-222.6%
북구	국우동	13,437	47,210	-33,773	-251.3%
북구	소계	1,029,911	1,134,357	-104,446	-10.1%
수성구	범어1동	45,029	61,590	-16,561	-36.8%
수성구	범어2동	47,049	77,253	-30,205	-64.2%
수성구	범어3동	39,075	55,508	-16,432	-42.1%
수성구	범어4동	23,679	47,484	-23,805	-100.5%
수성구	만촌1동	100,938	66,713	34,226	33.9%
수성구	만촌2동	17,451	33,236	-15,785	-90.5%
수성구	만촌3동	38,659	50,915	-12,256	-31.7%
수성구	수성1가동	31,088	35,380	-4,293	-13.8%
수성구	수성2·3가동	43,909	48,210	-4,301	-9.8%
수성구	수성4가동	48,598	29,680	18,918	38.9%
수성구	황금1동	9,512	41,403	-31,891	-335.3%
수성구	황금2동	28,014	35,517	-7,503	-26.8%

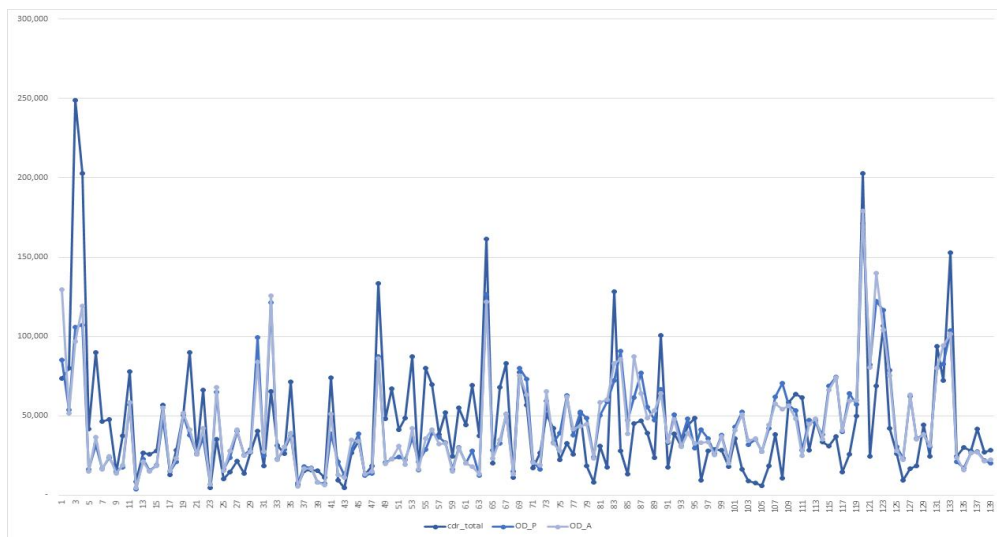


구분		CDR total (A)	O/D P (B)	차이 (A-B)	오차율 ((A-B/A)* 100))
중존 (시군구)	소존 (읍면동)				
수성구	중동	28,809	26,205	2,604	9.0%
수성구	상동	28,206	37,735	-9,529	-33.8%
수성구	파동	18,269	21,356	-3,087	-16.9%
수성구	두산동	35,528	42,843	-7,315	-20.6%
수성구	지산1동	16,168	52,664	-36,496	-225.7%
수성구	지산2동	8,985	32,026	-23,041	-256.4%
수성구	범물1동	7,922	35,457	-27,535	-347.6%
수성구	범물2동	5,884	27,661	-21,777	-370.1%
수성구	고산1동	18,310	42,050	-23,740	-129.7%
수성구	고산2동	38,196	61,791	-23,595	-61.8%
수성구	고산3동	10,723	70,599	-59,876	-558.4%
수성구	소계	690,000	1,033,275	-343,275	-49.7%
달서구	성당동	58,626	56,829	1,798	3.1%
달서구	두류1,2동	63,511	53,152	10,359	16.3%
달서구	두류3동	61,434	28,254	33,180	54.0%
달서구	본리동	28,418	47,254	-18,836	-66.3%
달서구	감삼동	47,371	45,519	1,853	3.9%
달서구	죽전동	33,756	38,083	-4,327	-12.8%
달서구	장기동	30,780	68,965	-38,185	-124.1%
달서구	용산1동	37,036	74,295	-37,259	-100.6%
달서구	용산2동	14,826	40,053	-25,227	-170.2%
달서구	이곡1동	25,884	63,960	-38,075	-147.1%
달서구	이곡2동	49,892	57,223	-7,331	-14.7%
달서구	신당동	202,694	171,424	31,270	15.4%
달서구	월성1동	24,456	82,224	-57,768	-236.2%
달서구	월성2동	68,790	122,353	-53,563	-77.9%
달서구	진천동	106,844	116,854	-10,010	-9.4%
달서구	상인1동	42,301	78,884	-36,583	-86.5%
달서구	상인2동	26,151	30,403	-4,251	-16.3%
달서구	상인3동	9,475	23,324	-13,850	-146.2%
달서구	도원동	16,611	62,403	-45,792	-275.7%





구분		CDR total (A)	O/D P (B)	차이 (A-B)	오차율 ((A-B/A)* 100))
중존 (시군구)	소존 (읍면동)				
달서구	송현1동	18,559	35,732	-17,172	-92.5%
달서구	송현2동	44,118	38,264	5,854	13.3%
달서구	본동	24,712	30,716	-6,004	-24.3%
달서구	소계	1,036,247	1,366,165	-329,918	-31.8%
달성군	화원읍	93,940	80,610	13,330	14.2%
달성군	논공읍	72,402	82,436	-10,034	-13.9%
달성군	다사읍	152,740	103,914	48,826	32.0%
달성군	가창면	24,472	20,919	3,553	14.5%
달성군	하빈면	30,073	17,127	12,946	43.0%
달성군	옥포면	27,794	26,604	1,190	4.3%
달성군	현풍면	41,743	27,745	13,998	33.5%
달성군	유가면	27,078	22,006	5,073	18.7%
달성군	구지면	28,207	20,255	7,952	28.2%
달성군	소계	498,449	401,616	96,833	19.4%
합계		6,137,942	5,994,680	143,262	2.3%



<그림 4-1> CDR vs O/D Production 비교 (행정동 기준)





행정동 단위로 집계화된 CDR과 O/D 총통행유입량을 구·군단위로 집계된 결과는 수성구, 남구의 경우 44.4~48.5%의 높은 오차율을 보였으나 북구와 동구의 경우 7.2~14.1%로 비교적 낮은 오차율을 나타내었다. 특히 수성구 범물1·2동, 고산3동과 같은 경우에는 O/D 총통행유입량이 CDR보다 훨씬 많은 것으로 나타났으나 남구 이천동, 대명11동의 경우 O/D총통행유입량이 CDR보다 적은 것으로 나타났다. 이러한 결과는 앞서 O/D 총통행발생량과 유사한 지역으로 대규모 공동주택단지 등 거주인구가 많은 지역일수록 O/D 총통행발생유입량이 CDR보다 높은 결과를 보이고 있는 것을 알 수 있다.

<표 4-3> 행정동 기준 CDR과 O/D 총통행유입량 비교

구분		CDR total (A)	O/D A (B)	차이 (A-B)	오차율 ((A-B/A)* 100))
중촌 (시군구)	소촌 (읍면동)				
중구	동인동	73,538	129,518	-55,980	-76.1%
중구	삼덕동	80,025	51,531	28,494	35.6%
중구	성내1동	248,787	96,946	151,842	61.0%
중구	성내2동	202,649	119,151	83,498	41.2%
중구	성내3동	41,783	14,899	26,884	64.3%
중구	대신동	90,099	36,587	53,512	59.4%
중구	남산1동	46,517	16,134	30,382	65.3%
중구	남산2동	47,847	24,595	23,252	48.6%
중구	남산3동	14,554	13,969	585	4.0%
중구	남산4동	37,226	18,928	18,298	49.2%
중구	대봉1동	78,051	58,408	19,642	25.2%
중구	대봉2동	8,593	4,592	4,001	46.6%
중구	소계	969,670	585,259	384,411	39.6%
동구	신암1동	26,794	21,066	5,727	21.4%
동구	신암2동	25,749	15,097	10,652	41.4%
동구	신암3동	27,883	18,502	9,381	33.6%
동구	신암4동	56,783	55,146	1,636	2.9%



구분		CDR total (A)	O/D A (B)	차이 (A-B)	오차율 ((A-B/A)* 100))
중존 (시군구)	소존 (읍면동)				
동구	신암5동	12,900	15,531	-2,630	-20.4%
동구	신천1·2동	28,440	23,044	5,397	19.0%
동구	신천3동	51,839	51,509	330	0.6%
동구	신천4동	90,035	41,497	48,539	53.9%
동구	효목1동	26,311	25,679	633	2.4%
동구	효목2동	66,191	42,272	23,919	36.1%
동구	도평동	4,547	6,739	-2,192	-48.2%
동구	불로·봉무동	35,461	67,890	-32,429	-91.4%
동구	지저동	10,505	15,284	-4,778	-45.5%
동구	동촌동	14,678	27,988	-13,310	-90.7%
동구	방촌동	21,395	41,457	-20,061	-93.8%
동구	해안동	13,560	24,982	-11,423	-84.2%
동구	공산동	27,178	27,663	-485	-1.8%
동구	안심1동	40,492	83,898	-43,406	-107.2%
동구	안심2동	18,386	27,333	-8,947	-48.7%
동구	안심3·4동	65,224	125,608	-60,383	-92.6%
동구	소계	664,352	758,183	-93,831	-14.1%
서구	내당1동	31,405	22,203	9,202	29.3%
서구	내당2·3동	26,076	30,678	-4,601	-17.6%
서구	내당4동	71,459	39,245	32,214	45.1%
서구	비산1동	6,585	5,451	1,134	17.2%
서구	비산2·3동	15,643	16,975	-1,332	-8.5%
서구	비산4동	16,040	16,769	-728	-4.5%
서구	비산5동	15,281	8,005	7,276	47.6%
서구	비산6동	11,009	6,302	4,707	42.8%
서구	비산7동	74,195	51,307	22,888	30.8%
서구	평리1동	9,360	13,013	-3,654	-39.0%
서구	평리2동	4,515	11,292	-6,778	-150.1%
서구	평리3동	26,532	35,018	-8,486	-32.0%
서구	평리4동	34,077	33,784	292	0.9%
서구	평리5동	12,688	13,205	-517	-4.1%



구분		CDR total (A)	O/D A (B)	차이 (A-B)	오차율 ((A-B/A)* 100))
중존 (시군구)	소존 (읍면동)				
서구	평리6동	18,635	15,015	3,620	19.4%
서구	상중이동	133,484	86,038	47,446	35.5%
서구	원대동	48,206	19,574	28,632	59.4%
서구	소계	555,189	423,874	131,316	23.7%
남구	이천동	67,284	22,999	44,285	65.8%
남구	봉덕1동	41,452	31,112	10,340	24.9%
남구	봉덕2동	48,679	19,341	29,338	60.3%
남구	봉덕3동	87,300	42,142	45,158	51.7%
남구	대명1동	15,826	16,399	-573	-3.6%
남구	대명2동	80,181	35,792	44,389	55.4%
남구	대명3동	69,600	41,146	28,454	40.9%
남구	대명4동	38,307	32,448	5,859	15.3%
남구	대명5동	52,077	32,810	19,266	37.0%
남구	대명6동	24,533	15,089	9,444	38.5%
남구	대명9동	55,105	29,569	25,537	46.3%
남구	대명10동	44,286	20,414	23,873	53.9%
남구	대명11동	69,493	18,095	51,398	74.0%
남구	소계	694,123	357,356	336,767	48.5%
북구	고성동	37,329	13,439	23,889	64.0%
북구	철성동	161,232	121,647	39,585	24.6%
북구	침산1동	20,061	23,313	-3,252	-16.2%
북구	침산2동	67,983	34,655	33,328	49.0%
북구	침산3동	83,206	51,282	31,924	38.4%
북구	산격1동	11,277	13,173	-1,896	-16.8%
북구	산격2동	77,670	75,422	2,248	2.9%
북구	산격3동	56,634	63,079	-6,444	-11.4%
북구	산격4동	17,236	20,787	-3,551	-20.6%
북구	북현1동	26,733	19,079	7,654	28.6%
북구	북현2동	50,913	65,277	-14,365	-28.2%
북구	대현동	42,162	32,879	9,283	22.0%
북구	검단동	22,323	28,001	-5,678	-25.4%



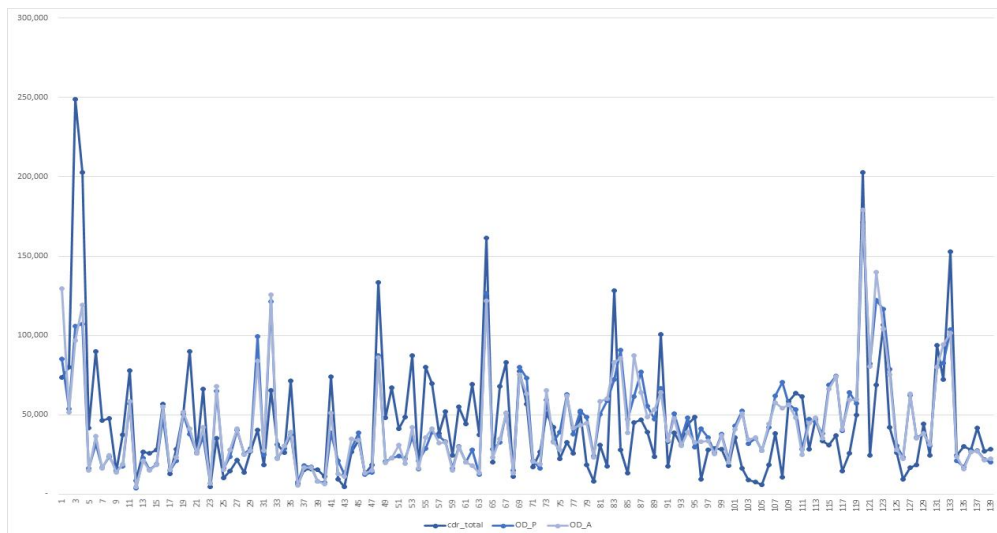
구분		CDR total (A)	O/D A (B)	차이 (A-B)	오차율 ((A-B/A)* 100))
중존 (시군구)	소존 (읍면동)				
북구	무태조야동	32,484	62,039	-29,555	-91.0%
북구	태전2동	25,937	41,655	-15,718	-60.6%
북구	관문동	51,628	43,452	8,175	15.8%
북구	읍내동	18,510	44,820	-26,310	-142.1%
북구	관음동	8,165	23,425	-15,260	-186.9%
북구	태전1동	30,919	58,739	-27,819	-90.0%
북구	구암동	17,764	60,110	-42,345	-238.4%
북구	노원동	128,172	83,039	45,133	35.2%
북구	동천동	28,136	85,757	-57,621	-204.8%
북구	국우동	13,437	38,899	-25,462	-189.5%
북구	소계	1,029,911	1,103,968	-74,057	-7.2%
수성구	범어1동	45,029	87,289	-42,259	-93.8%
수성구	범어2동	47,049	64,036	-16,987	-36.1%
수성구	범어3동	39,075	48,785	-9,710	-24.8%
수성구	범어4동	23,679	53,420	-29,741	-125.6%
수성구	만촌1동	100,938	63,945	36,994	36.6%
수성구	만촌2동	17,451	34,124	-16,673	-95.5%
수성구	만촌3동	38,659	47,830	-9,171	-23.7%
수성구	수성1가동	31,088	30,541	547	1.8%
수성구	수성2·3가동	43,909	39,179	4,730	10.8%
수성구	수성4가동	48,598	32,501	16,097	33.1%
수성구	황금1동	9,512	32,917	-23,404	-246.0%
수성구	황금2동	28,014	33,407	-5,393	-19.3%
수성구	중동	28,809	25,394	3,414	11.9%
수성구	상동	28,206	37,090	-8,884	-31.5%
수성구	파동	18,269	20,342	-2,073	-11.3%
수성구	두산동	35,528	40,919	-5,391	-15.2%
수성구	지산1동	16,168	50,835	-34,667	-214.4%
수성구	지산2동	8,985	34,323	-25,338	-282.0%
수성구	범물1동	7,922	35,873	-27,951	-352.8%
수성구	범물2동	5,884	27,492	-21,608	-367.2%



구분		CDR total (A)	O/D A (B)	차이 (A-B)	오차율 ((A-B/A)* 100))
중존 (시군구)	소존 (읍면동)				
수성구	고산1동	18,310	44,215	-25,905	-141.5%
수성구	고산2동	38,196	57,528	-19,332	-50.6%
수성구	고산3동	10,723	54,201	-43,478	-405.5%
수성구	소계	690,000	996,183	-306,184	-44.4%
달서구	성당동	58,626	56,166	2,461	4.2%
달서구	두류1,2동	63,511	48,075	15,436	24.3%
달서구	두류3동	61,434	24,876	36,558	59.5%
달서구	본리동	28,418	44,822	-16,404	-57.7%
달서구	감삼동	47,371	48,138	-766	-1.6%
달서구	죽전동	33,756	35,470	-1,714	-5.1%
달서구	장기동	30,780	66,431	-35,651	-115.8%
달서구	용산1동	37,036	74,078	-37,041	-100.0%
달서구	용산2동	14,826	41,015	-26,189	-176.6%
달서구	이곡1동	25,884	59,320	-33,436	-129.2%
달서구	이곡2동	49,892	59,937	-10,044	-20.1%
달서구	신당동	202,694	179,115	23,579	11.6%
달서구	월성1동	24,456	80,682	-56,226	-229.9%
달서구	월성2동	68,790	140,066	-71,276	-103.6%
달서구	진천동	106,844	104,258	2,586	2.4%
달서구	상인1동	42,301	75,115	-32,814	-77.6%
달서구	상인2동	26,151	28,131	-1,980	-7.6%
달서구	상인3동	9,475	22,216	-12,741	-134.5%
달서구	도원동	16,611	63,435	-46,824	-281.9%
달서구	송현1동	18,559	35,434	-16,875	-90.9%
달서구	송현2동	44,118	38,085	6,033	13.7%
달서구	본동	24,712	31,270	-6,558	-26.5%
달서구	소계	1,036,247	1,356,134	-319,887	-30.9%
달성군	화원읍	93,940	80,122	13,817	14.7%
달성군	논공읍	72,402	94,337	-21,936	-30.3%
달성군	다사읍	152,740	101,716	51,024	33.4%
달성군	가창면	24,472	23,728	744	3.0%



구분		CDR total (A)	O/D A (B)	차이 (A-B)	오차율 ((A-B/A)* 100))
중존 (시군구)	소존 (읍면동)				
달성군	하빈면	30,073	16,084	13,989	46.5%
달성군	옥포면	27,794	26,806	988	3.6%
달성군	현풍면	41,743	26,981	14,763	35.4%
달성군	유가면	27,078	21,719	5,360	19.8%
달성군	구지면	28,207	22,230	5,977	21.2%
달성군	소계	498,449	413,723	84,726	17.0%
합계		6,137,942	5,994,680	143,262	2.3%



<그림 4-2> CDR vs O/D Attraction 비교 (행정동 기준)

### 1) 상관관계 분석결과

CDR 총통행량과 O/D 총통행발생·유입량의 기술통계량 분석결과는 표4-4와 같다. CDR 총통행량의 평균은 44,157통행이고 O/D 총통행발생·유입량의 평균은 43,127통행으로 나타났다.



<표 4-4> 기술통계량

	평균	표준편차	N
CDR TOTAL	44157.8345	39116.16919	139
O/D P	43127.2158	28688.66272	139
O/D A	43127.2302	29879.45240	139

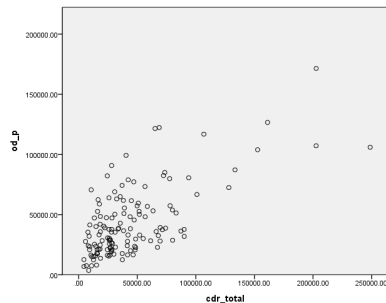
이를 바탕으로 각 데이터별 상관관계를 분석해본 결과, CDR과 O/D 총통행 발생량은 0.666의 상관관계를 보였고 O/D 총통행유입량과는 0.673의 상관관계를 나타내었다.

<표 4-5> CDR과 O/D 총통행발생·유입량 상관관계분석 종합

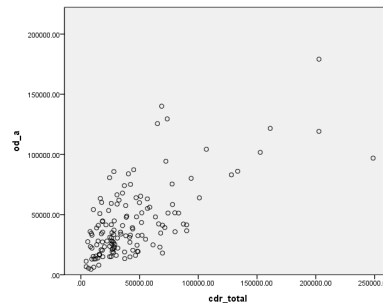
		CDR TOTAL	O/D P	O/D A
CDR TOTAL	Pearson 상관	1		
	유의확률 (양측)			
	N	139		
O/D P	Pearson 상관	.666**	1	
	유의확률 (양측)	.000		
	N	139	139	
O/D A	Pearson 상관	.673**	.975**	1
	유의확률 (양측)	.000	.000	
	N	139	139	139

\*\* . 상관관계가 0.01 수준에서 유의합니다(양측).





CDR total & O/D P



CDR total & O/D A

<그림 4-3> CDR total & O/D PA 산점도

## 2) 대응표본 t-test

동일 모집단의 사전·사후점수의 평균을 비교할 때 사용되는 대응표본 t-test 분석결과, CDR총통행량과 O/D 총통행발생·유입량의 유의확률은 각각 0.679와 0.677로 나타나 유의수준 0.005보다 크므로 귀무가설인 ‘두 변수간 평균 차이가 없다.’가 기각되지 아니하여 CDR총통행량과 O/D통행발생·유입량의 두 집단 간 평균 차이는 없는 것으로 나타났다.

- 귀무가설  $H_0 : \mu_1 - \mu_2 = 0$  : CDR과 O/D간에는 평균 차이가 없다.
- 대립가설  $H_1 : \mu_1 - \mu_2 \neq 0$  : CDR과 O/D간에는 평균 차이가 있다.

<표 4-6> CDR 및 O/D 총통행발생·유입통행량 대응표본 검정결과

	대응차			t	자유도	유의확률 (양측)
	평균	표준편차	평균의 표준오차			
CDR TOTAL & O/D P	1030.6	29288.6	2484.2	.415	138	.679
CDR TOTAL & O/D A	1030.6	29148.1	2472.3	.417	138	.677





## (2) 필수적 통행(Mandatory Trip) 기준

필수적 통행(Mandatory Trip)과 관련하여 CDR 평일 오전·오후침두시와 O/D 가정기반 통근·통학통행 발생·유입량을 비교하여 CDR의 요일별 데이터의 통행목적별 특성에 대해 분석해보았다.

### 1) 상관관계 분석결과

CDR 평일 오전침두시 통행량과 O/D 가정기반 통근·통학통행 발생·유입량의 기술통계 결과는 표4-7과 같다. CDR 평일 오전침두시 통행량의 평균은 18,774통행이고 O/D 가정기반 통근통학 통행발생·유입량의 평균은 각각 22,239통행, 22,111통행으로 나타났다.

<표 4-7> 기술통계량

	평균	표준편차	N
CDR Mand Mpeak	18774.1655	16105.25657	139
O/D P Mand	22239.9137	14958.21674	139
O/D A Mand	22111.6475	15110.61846	139

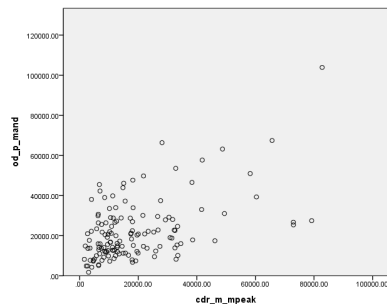
이를 바탕으로 각 데이터별 상관관계를 분석해본 결과, CDR 평일 오전침두시와 O/D 가정기반 통근·통학통행발생·유입량은 모두 0.512의 상관관계를 나타내었다.



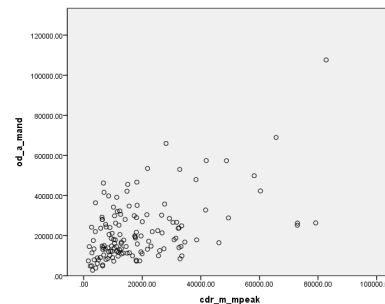
<표 4-8> CDR 평일 오전첨두시 필수적 통행 통행발생·유입량 상관관계분석 종합

		cdr m mpeak	O/D p mand	O/D a mand
CDR Mand Mpeak	Pearson 상관	1		
	유의확률 (양측)			
	N	139		
O/D P Mand	Pearson 상관	.512**	1	
	유의확률 (양측)	.000		
	N	139	139	
O/D A Mand	Pearson 상관	.512**	.995**	1
	유의확률 (양측)	.000	.000	
	N	139	139	139

\*\* . 상관관계가 0.01 수준에서 유의합니다(양측).



CDR Mand Mpeak & O/D P Mand



CDR Mand Mpeak & O/D A Mand

<그림 4-4> CDR Mand Mpeak & O/D P A Mand 산점도

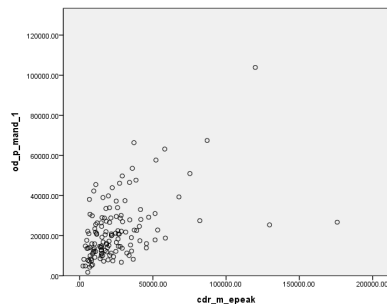
다음으로 CDR 평일 오후첨두시와 O/D 가정기반 통근·통학통행 발생·유입량은 각각 0.481과 0.479로 CDR 평일 오전첨두시 통행발생·유입량보다는 상대적으로 낮은 상관관계를 나타냈다.



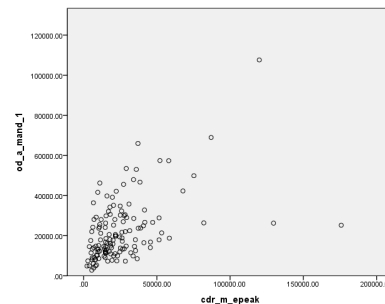
<표 4-9> CDR 평일 오후침두시 필수적 통행 통행발생·유입량 상관관계분석 종합

		cdr m Epeak	O/D p mand	O/D a mand
CDR Mand Epeak	Pearson 상관	1		
	유의확률 (양측)			
	N	139		
O/D P Mand	Pearson 상관	.481**	1	
	유의확률 (양측)	.000		
	N	139	139	
O/D A Mand	Pearson 상관	.479**	.995**	1
	유의확률 (양측)	.000	.000	
	N	139	139	139

\*\* . 상관관계가 0.01 수준에서 유의합니다(양측).



CDR Mand Epeak & O/D P Mand



CDR Mand Epeak & O/D A Mand

<그림 4-5> CDR Mand Mpeak & O/D P A Mand 산점도

## 2) 대응표본 t-test

대응표본 t-검정 결과, CDR 평일 오전침두시와 O/D 가정기반 통근·통학통행 발생·유입량의 유의확률은 0.009~0.012로 나타나 유의수준 0.05보다 작아 귀무가설인 ‘두 변수간 평균 차이가 없다.’가 기각되어 두 통행량은 차이가



있는 것으로 나타났다. 반면 CDR 평일 오후침두시와 O/D 가정기반 통근·통학통행발생·유입량의 유의확률은 0.082~0.071로 나타나 유의수준 0.05보다 커 귀무가설인 ‘두 변수간 차이가 없다.’가 기각되지 아니하여 두 통행량은 차이가 없는 것으로 나타났다.

- 귀무가설  $H_0 : \mu_1 - \mu_2 = 0$  : CDR과 O/D간에는 평균 차이가 없다.
- 대립가설  $H_1 : \mu_1 - \mu_2 \neq 0$  : CDR과 O/D간에는 평균 차이가 있다.

<표 4-10> 대응표본 검정

	대응차			t	자유도	유의 확률 (양측)
	평균	표준편차	평균의 표준오차			
CDR Mand Mpeak & O/D P Mand	-3465.7	15372.1	1303.8	-2.658	138	.009
CDR Mand Mpeak & O/D A Mand	-3337.5	15450.6	1310.5	-2.547	138	.012
CDR Mand Epeak & O/D P Mand	3143.8	21144.6	1793.5	1.753	138	.082
CDR Mand Epeak & O/D A Mand	3272.1	21208.2	1798.9	1.819	138	.071

### (3) 선택적 통행(Optional Trip) 기준

선택적 통행(Optional Trip)과 관련하여 CDR 주말 오전·오후침두시와 O/D 가정기반 학원·쇼핑·기타 및 비가정기반 쇼핑·여가·기타 통행발생·유입량을 비교하여 CDR의 요일별 데이터의 통행목적별 특성에 대해 분석해보았다.

#### 1) 상관관계 분석결과



CDR 주말 오전첨두시 통행량과 O/D 가정·비가정기반 학원·쇼핑·여가·기타통행발생량의 기술통계 결과는 표4-11과 같다. CDR 주말 오전첨두시 평균은 36,033통행이고 O/D 가정·비가정기반 학원·쇼핑·여가·기타통행발생·유입량의 평균은 22,906통행으로 나타났다.

<표 4-11> 기술통계량

	평균	표준편차	N
CDR Opt Mpeak	36033.8273	30481.72211	139
O/D P Opt	22906.1727	18148.79298	139
O/D A Opt	22906.2014	19533.04963	139

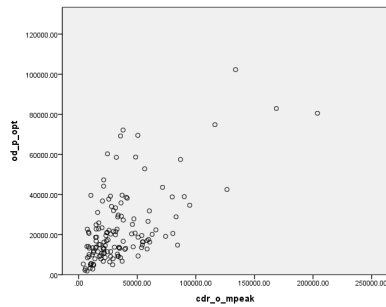
이를 바탕으로 각 데이터별 상관관계를 분석해본 결과, CDR 주말 오전첨두시와 가정·비가정기반 학원·쇼핑·여가·기타통행발생·유입량은 각각 0.620과 0.614의 높은 상관관계를 나타내었다.

<표 4-12> CDR 주말오전첨두시 선택적 통행 통행발생량 상관관계분석 종합

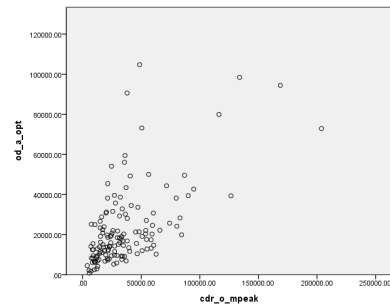
		cdr m mpeak	O/D p mand	O/D a mand
CDR Opt Mpeak	Pearson 상관 유의확률 (양측)	1		
	N	139		
O/D P Opt	Pearson 상관 유의확률 (양측)	.620**	1	
	N	139	139	
O/D A Opt	Pearson 상관 유의확률 (양측)	.614**	.940**	1
	N	139	139	139

\*\* . 상관관계가 0.01 수준에서 유의합니다(양측).





CDR Opt Mpeak & O/D P Opt



CDR Opt Mpeak & O/D A Opt

<그림 4-6> CDR Mand Mpeak & O/D P A Mand 산점도

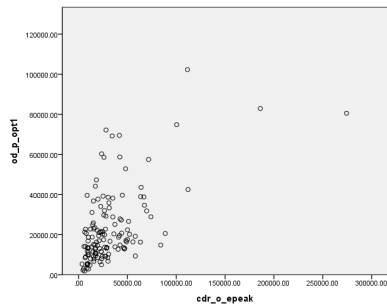
다음으로 CDR 주말 오후첨두시와 O/D 가정·비가정기반 학원·쇼핑·여가·기타통행발생·유입량은 각각 0.589과 0.572로 CDR 주말 오전첨두시 통행발생·유입량보다는 상대적으로 낮은 상관관계를 나타냈다.

<표 4-13> CDR 주말오후첨두시 선택적 통행 통행유입량 상관관계분석 종합

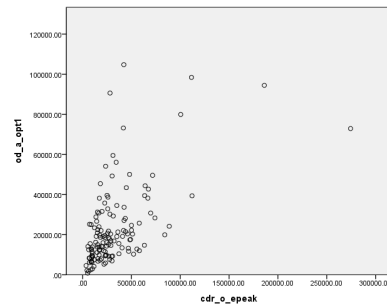
		cdr m epeak	O/D p mand	O/D a mand
CDR Opt Epeak	Pearson 상관	1		
	유의 확률 (양측)			
	N	139		
O/D P Opt	Pearson 상관	.589**	1	
	유의 확률 (양측)	.000		
	N	139	139	
O/D A Opt	Pearson 상관	.572**	.940**	1
	유의 확률 (양측)	.000	.000	
	N	139	139	139

\*\* . 상관관계가 0.01 수준에서 유의합니다(양측).





CDR Opt Epeak & O/D P Opt



CDR Opt Epeak & O/D A Opt

<그림 4-7> CDR Mand Mpeak & O/D P A Mand 산점도

## 2) 대응표본 t-test

대응표본 t-test 분석결과, CDR 주말 오전·오후침두시와 O/D 가정·비가정기 반 학원·쇼핑·여가·기타통행발생·유입량의 유의확률은 모두 0.000으로 나타나 유의수준 0.005보다 작으므로 귀무가설인 ‘두 변수간 평균 차이가 없다.’가 기각되어 두 통행량은 차이가 있는 것으로 나타났다.

- 귀무가설  $H_0 : \mu_1 - \mu_2 = 0$  : CDR과 O/D간에는 평균 차이가 없다.
- 대립가설  $H_1 : \mu_1 - \mu_2 \neq 0$  : CDR과 O/D간에는 평균 차이가 있다.

<표 4-14> 대응표본 검정

	대응차			t	자유도	유의확률 (양측)
	평균	표준편차	평균의 표준오차			
CDR Opt Mpeak & O/D P Opt	13127.7	23919.6	2028.8	6.471	138	.000
CDR Opt Mpeak & O/D A Opt	13127.6	24070.6	2041.6	6.430	138	.000
CDR Opt Epeak & O/D P Opt	9322.1	26428.9	2241.7	4.159	138	.000
CDR Opt Epeak & O/D A Opt	9322.1	26804.3	2273.5	4.100	138	.000



#### 4-2-2. 공간상관성 기준 통행목적별 상관성 분석 (행정동 기준)

앞선 절에서 통행목적별 CDR과 O/D의 양(量)적인 측면에서 두 데이터의 상관성과 t-test결과를 살펴보았고 본 절에서는 존 별로 CDR과 O/D가 통행목적별로 공간상에서 어떠한 공간적인 분포와 상관성에 가지는지에 대하여 살펴보고자 한다. CDR과 O/D의 공간적 상관성을 알아보기 위한 국지적 공간상관성의 검증은 제2장에서 살펴본 2개의 지표인 Local Moran's I와 Getis-Ord  $G^*$ 를 이용하였다.

##### (1) 국지적 공간상관성

###### 1) 총통행량 기준

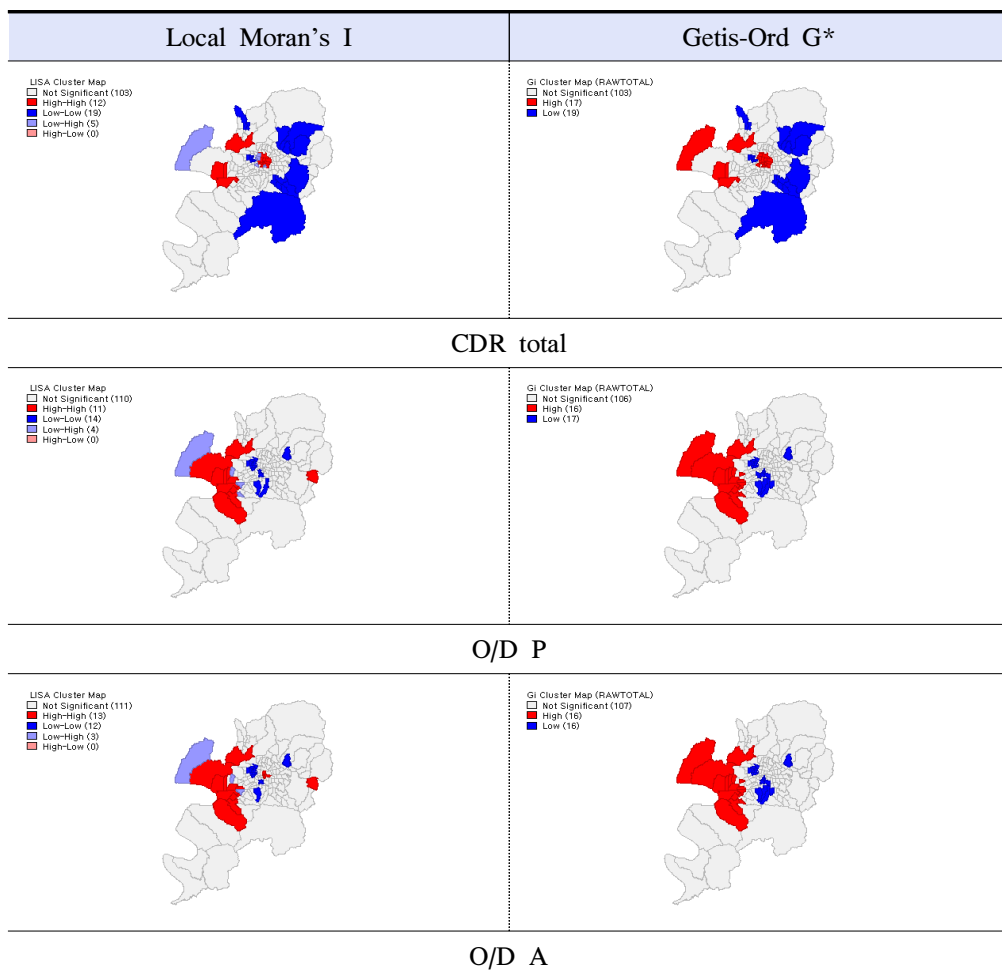
CDR 총통행량의 Local Moran's I 검정결과는 그림4-8과 같이 대구 도심지역인 동인동, 성내1·2동, 삼덕동, 북구 칠성동 등의 지역과 부도심지역인 달서구 신당동, 이곡1동, 월성2동에서 H-H 상관성을 나타냈다. 반면 동구 불로·봉무동, 도평동, 읍내동, 동천동, 지저동, 해안동 등과 수성구 고산2동, 지산1·2동, 범물1·2동, 달성군 가창면이 L-L상관성을 나타냈다. 또한 고성동, 성내3동, 남산4동, 하빈면은 주변지역보다 낮은 값(L-H)을 갖는 것으로 나타났다. Getis-ord  $G^*$  검정결과는 Local Moran's I과 유사한 군집을 보이고 공간 상관성이 높은 지역도 유사하게 나타났다. 다만, 하빈면은 Getis-ord  $G^*$ 에서 평균보다 높은 국지적 공간상관성을 나타냈다.

O/D 총통행발생·유입량의 Local Moran's I 검정결과는 대구 주거단지 밀집지역인 달서구 신당동, 월성1·2동, 진천동, 관문동, 신당동과 달성군 다사읍, 화원읍 등의 지역과 수성구 고산동 등지에서 H-H 상관성을 나타내었다. 반대





로 서구 평리동, 비산동, 내당동, 남구 대명6동과 동구 동촌동이 L-L상관성을 나타냈다. 또한 하빈면, 용산동 등은 주변지역보다 낮은 값(L-H)을 갖는 것으로 나타났다. Getis-ord  $G^*$  검정결과는 Local Moran's I과 유사한 군집을 보이고 공간 상관성이 높은 지역도 유사하게 나타났으나 L-H상관성을 나타낸 지역은 Getis-ord  $G^*$ 에서도 높은 국지적 공간상관성을 나타냈다.



<그림 4-8> CDR total & O/D PA 국지적 공간상관성 분석결과



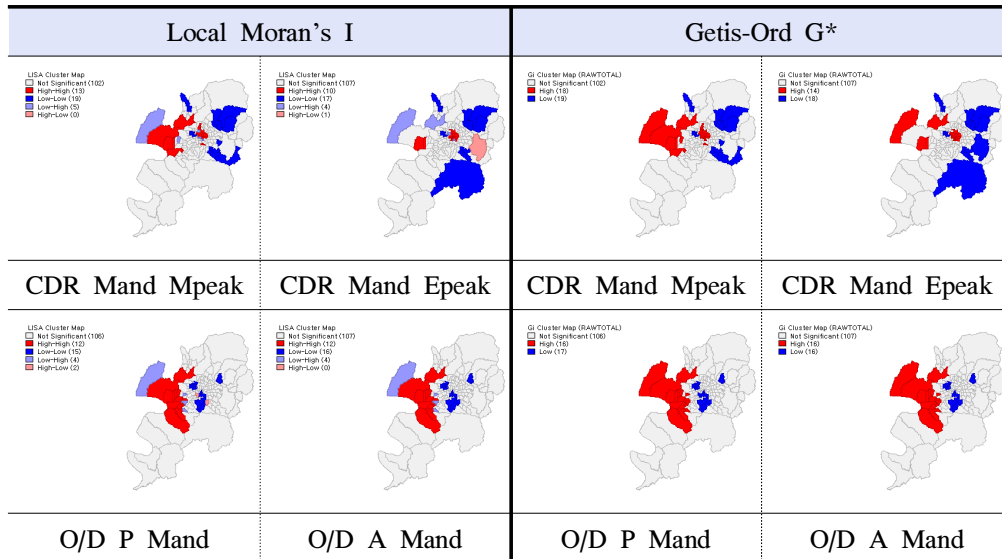
## 2) 필수적 통행 기준

CDR 평일 오전첨두시의 Local Moran's I 검정결과는 그림4-9와 같이 대구 도심지역인 동인동, 성내동, 삼덕동 등과 북구 침산동, 칠성동 지역과 부도심 지역인 달서구 관문동, 신당동, 이곡동, 달성군 다사읍에서 H-H 상관성을 나타냈다. 반면 서구 비산동, 평리동, 동구 불로·봉무동, 도평동, 동천동, 지저동, 해안동과 수성구 만촌동, 지산동 고산동, 범물동 일대는 L-L상관성을 나타냈다. 또한 고성동, 침산동, 남산동, 대봉동, 용산동, 하빈면은 주변지역보다 낮은 값(L-H)을 갖는 것으로 나타났다.

CDR 평일 오후첨두시의 경우 대구 도심지역에서 앞서 오전첨두시의 대구 도심 집중과 유사하게 나타났으나 부도심지역인 신당동, 이곡2동 등지에서 H-H상관성을 나타나 전체적으로 지역적 범위가 작아진 것을 알 수 있다. O/D 가정기반 통근·통학통행발생량은 달서구 용산, 이곡, 신당동, 장기동, 달성군 다사읍, 화원읍 등의 주거밀집지역에서 H-H상관성을 나타내고 도시 중심부에서는 L-L상관성을 나타내어 CDR 평일 오전첨두시와 비교적 유사한 결과치를 나타내는 것을 알 수 있으나 O/D 가정기반 통근·통학통행유입량은 CDR 평일 오후첨두시와 다른 양상을 보이는 것으로 나타났다.

Getis-ord  $G^*$  검정결과는 CDR과 O/D 필수적 통행 모두에서 Local Moran's I과 유사한 군집을 보이고 공간 상관성이 높은 지역도 유사하게 나타났으나, 두 통행량 모두에서 L-H상관성을 나타낸 지역은 Getis-ord  $G^*$ 에서도 높은 국지적 공간상관성을 나타냈다.





<그림 4-9> CDR Mand & O/D Mand PA 국지적 공간상관성 분석결과

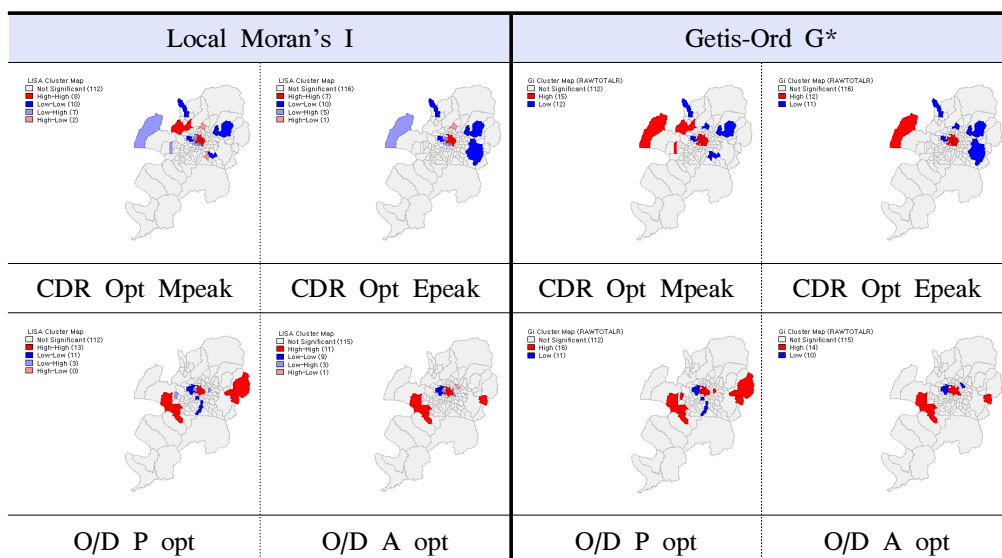
### 3) 선택적 통행 기준

CDR 주말 오전첨두시의 Local Moran's I 검정결과와 그림4-10과 같이 대구 도심지역인 동인동, 성내동, 남산동, 삼덕동, 대봉동 등의 지역과 부도심지역인 달서구 관문동에서 H-H 상관성을 나타냈다. 반면 동구 동촌동, 해안동, 북구 읍내동, 동천동과 서구 평리동, 비산동 등은 L-L상관성을 나타냈다. 또한 고성동, 성내3동, 남산1·4동, 대봉2동, 이곡2동, 하빈면은 주변지역보다 낮은 값(L-H)을 갖는 것으로 나타났고 북구 산격2동, 수성구 두산동은 주변지역 평균보다 높은 값(H-L) 상관성을 나타내었다. CDR 주말 오후첨두시는 오전첨두시와 비교적 유사하게 나타났으나 L-L상관성 지역으로 고산2동이 추가되고 관문동이 제외되는 결과를 나타내었다. 반면 O/D 가정·비가정기반 학원·쇼핑·기타 통행발생량은 달서구 신당동, 월성동, 진천동, 동구 안심지구 등의 주거



밀집지역에서 H-H상관성을 나타냈으나 서구 비산동, 노원동 남구 이천동, 봉덕동 등지에서는 L-L상관성을 보이는 것으로 나타났다. 또한 O/D 가정·비가정기반 학원·쇼핑·기타 통행유입량은 앞선 O/D 가정·비가정기반 학원·쇼핑·기타 발생량과 유사한 공간적 상관성을 보이는 것으로 나타났으나, 동구 안심지구와 남구 일부 지역의 상관성은 제외되는 것으로 나타났다.

Getis-ord  $G^*$  검정결과는 CDR과 O/D 선택적 통행 모두에서 Local Moran's I과 유사한 군집을 보이고 공간 상관성이 높은 지역도 유사하게 나타났으나, 두 통행량 모두에서 L-H상관성을 나타낸 지역은 Getis-ord  $G^*$ 에서도 높은 국지적 공간상관성을 나타냈다.



<그림 4-10> CDR Opt & O/D Opt PA 국지적 공간상관성 분석결과



### 4.3. 소결

본 장에서는 CDR 및 O/D 데이터를 통행목적별 구분을 통한 통행량(量)적인 측면의 비교·분석 및 공간적 측면의 국지적 공간상관성 분석인 Local Moran's I와 Getis-Ord G\*의 분석을 통해 두 데이터가 가진 상관성에 대해 알아보았다. 통행목적별 구분을 통한 통행량(量)적 측면에서 모형의 상관성을 나타내는  $R^2$ 와 대응표본t-test p-value를 비교하여 가장 유사한 대안을 비교해 본 결과, CDR 총통행량은 상대적으로 O/D 총통행발생량보다는 O/D 총통행유입량과의 관계가 높은 것으로 나타났으며, 통행목적별로 필수적 통행(Mandatory Trip)에서는 CDR의 평일 오전첨두시와 O/D 통행발생·유입량과의 관계성이 입증된 것으로 보이나, CDR의 평일 오후첨두시와의 관계성은 상대적으로 낮은 것으로 판단된다. 선택적 통행(Optional Trip)에서는 CDR 주말 오전첨두시와 O/D 통행발생량과의 관계성이 있는 것으로 보이나 다른 통행패턴과는 특별한 관계성을 나타내 보이지는 못한 것으로 판단된다.

<표 4-15> 통행량 기준 결과종합

구분	통행목적	$R^2$	P-value
총목적통행	CDR total & O/D P	0.666	.679
	CDR total & O/D A	0.673	.677
필수적통행 (Mandatory Trip)	CDR Mand Mpeak & O/D p mand	0.512	.009
	CDR Mand Mpeak & O/D a mand	0.512	.012
	CDR Mand Epeak & O/D p mand	0.481	.082
	CDR Mand Epeak & O/D a mand	0.479	.071
선택적통행 (Optional Trip)	CDR Opt Mpeak & O/D p opt	0.620	.000
	CDR Opt Mpeak & O/D a opt	0.614	.000
	CDR Opt Epeak & O/D p opt	0.589	.000
	CDR Opt Epeak & O/D a opt	0.572	.000



다음으로 공간적 측면에서 국지적 공간상관성 분석인 Local Moran's I와 Getis-Ord G\* 분석 결과 CDR의 총통행량의 국지적 공간상관성 분포와 O/D 총통행발생·유입량의 공간적 분포는 CDR의 경우 대구 도심내부와 부도심 중 주거지 밀집지역인 달서구 신당동, 이곡동, 월성동에서 H-H 공간적 군집을 보였으나 O/D 총통행발생·유입의 경우 달서구, 수성구, 달성군의 주거지역 밀집지역에서만 H-H상관성을 나타내었다. 특히 O/D 총통행유입량의 경우 Local Moran's I에서 일부 대구 도심 집중 현상을 나타내어 CDR총통행량과 유사한 공간적 분포를 보였다. 이는 O/D의 경우 전수화 과정에서 가정기반의 사회경제지표에 근거한 전수화계수의 적용에 따른 주거지 통행발생·유입량의 추정에 따른 결과라고 생각된다.

필수적 통행과 관련하여 CDR 평일 오전첨두시 통행발생량과 O/D 가정기반 통근통학통행발생·유입량이 비교적 유사한 공간적 분포를 보이는 것으로 나타났으나, CDR 평일 오후첨두시통행량은 서로 다른 양상을 보이는 것으로 나타났다. 선택적 통행과 관련한 결과에서는 CDR 주말 오전·오후첨두시 통행량과 O/D 가정·비가정기반 학원·쇼핑·기타 통행발생유입량은 대구 도심내부지역에 일부 집중되는 부분만 유사한 공간적 분포를 나타내었고 다른 지역에서는 서로 상이한 공간적 군집을 가지는 것으로 나타났다.



<표 4-16> 국지적 공간상관성 기준 결과종합

구분	통행목적	Local MI		Getis-Ord G*	
		CDR	O/D	CDR	O/D
충목적 통행	CDR total & O/D P				
	CDR total & O/D A				
필수적 통행 (Mandatory Trip)	CDR Mand Mpeak & O/D p mand				
	CDR Mand Mpeak & O/D a mand				
	CDR Mand Epeak & O/D p mand				
	CDR Mand Epeak & O/D a mand				
선택적 통행 (Optional Trip)	CDR Opt Mpeak & O/D p opt				
	CDR Opt Mpeak & O/D a opt				
	CDR Opt Epeak & O/D p opt				
	CDR Opt Epeak & O/D a opt				





요약하자면 통행량(量)과 국지적 공간상관성을 기준으로 한 CDR과 O/D의 상관성은 총통행량에서 O/D 총통행발생량과 유입량 중 유입량에 상대적으로 높은 상관성이 있었으며, 필수적 통행에서는 통행량 기준에서 CDR 평일 오전첨두시와 O/D 가정기반 통근·통학통행발생·유입량 모두에서 다소 상관성이 있는 것으로 나타났고 선택적 통행은 CDR 주말 오전첨두시와 O/D 가정·비가정기반 학원, 업무, 쇼핑, 기타통행발생량에서 다소 관계성이 있는 것으로 나타났다.

국지적 공간상관성 기준에서는 CDR 총통행량은 O/D 총통행발생유입량과 유사한 공간적 분포를 보이는 것으로 나타났고 특히, O/D 총통행유입량은 대구 도심에 집중되고 있는 분포를 보여주어 CDR총통행량과 유사한 결과를 나타내었다. 또한 필수적 통행과 관련한 CDR 평일 오전·오후첨두시 통행량 분포는 O/D 가정기반 통근·통학통행 발생유입량과 공간적으로 어느 정도 유사한 분포를 보이는 것으로 나타났으나 CDR 평일 오후첨두시 통행량은 다른 형태를 나타내었다.

선택적 통행은 CDR 주말 오전·오후첨두시 통행량과 O/D 가정·비가정기반 학원, 쇼핑, 기타, 통행발생유입량에서 일부 대구 도심내부 지역의 공간적 군집을 제외하면 다른 지역에서는 특별한 상관성을 나타내지는 못하였다.

종합해보면 총통행량과 필수적 통행과 같이 비교적 많은 비중을 차지하는 통행량을 기준하였을 때 CDR은 O/D의 통행발생유입량 중 O/D 통행유입량과 상대적으로 높은 상관성을 보이는 것으로 나타났으며 공간적으로 살펴보았을 때 O/D의 특징적 성격인 주거지 밀집지역에서 O/D 통행발생과 유입량이 집중되는 공간적 패턴과 CDR이 가진 모바일 통신의 이동성에 대한 특성상 O/D와는 상당히 다른 통행패턴과의 차이를 확연히 드러내는 결과를 보였다.





## 제5장 공간회귀를 통한 통행발생모형 구축

### 5-1. 분석자료

#### 5-1-1. 변수의 선정

본 장에서는 통신 빅 데이터 및 공간상관성을 고려한 통행발생모형 추정을 위하여 기존 OLS 분석의 통행발생모형을 기준모형으로 하여 CDR의 요일별·시간대별로 구분된 데이터를 통행목적별 통행발생모형의 설명변수로 적용해보았다. 또한 통행발생에 영향을 주는 다양한 영향요인을 살펴보기 위해 선행연구 고찰을 통해 도출된 토지이용 및 사회경제변수, 물리적 변수를 적용하여 보다 현실적인 통행발생모형 도출을 위한 자료로 이용하였다. 구체적으로 각 통행목적별로 O/D 통행발생·유입량을 종속변수로 하여 CDR의 평일·주말기준 오전·오후침두시 통신량 데이터를 각 통행목적별로 대안별로 적용하였다. 토지이용변수의 경우, 앞서 2장에서 언급한 공공데이터포털의 대구광역시 건축물정보 관련 빅 데이터를 집계·분석하여 각 행정동의 토지이용 용도별 건축물 연상면적의 합계 데이터를 사용하였다. 또한, 토지이용복합도 및 직주균형계수의 경우, 선행연구 고찰을 통해 도출된 계산식으로부터 본 연구의 목적에 활용하기 알맞은 지표로 가공하여 사용하였다. 사회경제변수의 경우, 기존 O/D 전수화 과정에서 사용된 변수들 중 종속변수에 유의미한 영향을 준다고 판단된 기초데이터를 사용하였다. 물리적변수의 경우, 대구광역시내 교통유발시설의 유형을 총 6개의 카테고리로 나누어 유사한 용도별 데이터를 집계화하여 사용하였고, 버스정류장 및 철도·도시철도역의 경우 행정동별 합계 데이터를 사용하였다.



<표 5-1> 종속 · 설명변수의 선정

구분		구분기호		내용	비고
종속 변수	O/D 통행 발생 유입량	· O/D total · O/D Production · O/D Attraction		· 총통행발생유입량 · 가정기반 통근통학 · 가정기반 학원, 쇼핑, 기타 / 비가정기반 쇼핑, 기타 등	
	CDR	· CDR total · CDR Mand · CDR Opt		· 평일 오전·오후첨두시 · 평일 오전첨두(07~09) / 평일 오후첨두(17-19) · 주말 오전첨두(09~12) / 주말 오후첨두(14-16)	
설명 변수	Land Use	Land Use Floor Area	Residential Floor area	주거지역 건축물연상면적 (㎡)	
			Commercial Floor area	상업지역 건축물연상면적 (㎡)	
			Industrial Floor area	공업지역 건축물연상면적 (㎡)	
	Land Use	LI	Land Use Diversity Index	· 토지이용복합도 지표 · LI가 높을수록 다양한 토지이용 지역이 존재 · Baht and Guo(2007)	
		JH	Job-Housing balance index	· 직주균형계수 지표 · 해당 존의 직장파 주거지의 존재, 자족도의 개념 · Zhao et al.(2005), 정우현(2011)	
	Social Factor	ServiceWorkers		· 해당 존의 3차산업종사자수	
		AcaWorkers		· 해당 존의 학원종사자수	
		totalPop		· 해당 존의 총인구수	
		EmployN		· 해당 존의 취업자수	
		Nworkers		· 해당 존의 총종사자수	
		age	age15~19	· 해당 존의 15~19세 인구	
	age20~29		· 해당 존의 20~29세 인구		
	Facility Factors	교통 유발 시설	Dsales	· 판매시설(시장, 백화점, 대형마트)	
			Dhospital	· 의료시설(준종합, 대학병원)	
			Dterminal	· 교통시설(공항, 역, 버스터미널)	
			DMidHighUniN	· 교육시설(중·고·대학교)	
			DCultur	· 문화관광시설(영화관, 박물관, 공연장 등)	
			DOffice	· 행정시설(구군청, 시청 등)	
		버스 정류장수	BusStop	· 해당 존의 버스정류장수 합계	
		철도· 지하철역수	railSubStation	· 해당 존의 철도지하철역수 합계	



### 5-1-2. 기초통계분석

본 연구에서 설명변수로 사용되는 지표의 기초통계량은 표5-2와 같다. 각 변수들은 대구광역시 행정동 기준 139개(N=139)를 대상으로 분석하였다. 대표적으로 CDR 총통행량의 경우 최소값은 4,514.6통행이고 최대값은 248,787.5통행, 평균과 표준편차는 각각 44,157.9와 39,116.2로 나타났다. 모형에 사용된 설명변수들은 각 데이터별 Q-Q Plot의 정규분포 확인을 통해 모형분석의 정확도를 위하여 해당되는 변수들에 한해 Log변환을 시행하였다.

해당 변수들의 기초통계량 분석 결과는 아래 표5-2와 같다.

<표 5-2> 기초통계분석

구분	단위	Mean	std	Min	Max	비고
CDR total	통행	44,157.9	39,116.2	4,514.6	248,787.5	
CDR Mand Morning	통행	18,774.1	16,105.3	1,605.9	82,777.9	
CDR Mand Evening	통행	25,383.7	23,781.4	2,282.5	175,837.9	
CDR Opt Morning	통행	36,033.8	30,481.7	3,950.0	203,819.4	
CDR Opt Evening	통행	32,228.3	32,665.9	3,512.6	274,388.1	
Residential Floor area	m <sup>2</sup>	2,609,997.8	3,405,104.5	0.0	21,128,856.4	
Commercial Floor area	m <sup>2</sup>	886,968.4	3,994,119.1	0.0	38,730,816.6	
Industrial Floor area	m <sup>2</sup>	105,937.0	846,171.9	0.0	9,682,719.2	Log
LI	-	1.3305	0.3143	0.31	1.88	
JH	-	0.3183	0.0317	0.26	0.63	Log
ServiceWorkers	인	4,567.20	3,147.10	88.2	16,148.10	
AcaWorker	인	210	204.4	13.9	1,284.30	
totalPopu	인	15,228.80	8,852.70	3,270.10	51,208.30	
EmployN	인	7,966.60	4,888.00	1,648.10	28,010.70	
Nworkers	인	6,397.90	5,496.50	810.1	34,243.30	
age15 19	인	1,121.60	902.1	123.3	4,072.40	
age20 29	인	2,314.80	1,438.60	384.2	7,729.60	
EleMiddHighN	인	2,097.70	2,233.20	0	9,791.30	
UniStudN	인	0.6375	1.1529	0.3700	4.4500	Log



구분	단위	Mean	std	Min	Max	비고
DsalesFac	수	0.28	0.58	0	4	
Dhospital	수	0.22	0.52	0	3	
Dterminal	수	0.1	0.37	0	2	
DMidHighUniN	수	1.7	1.96	0	9	
DCultuF	수	0.27	0.57	0	2	
DOfficeF	수	0.14	0.49	0	2	
BusStop	수	1.2114	0.3474	0.3000	2.1200	Log
railSubStation	수	0.63	0.94	0	5.00	

## 5-2. 통행발생 선형회귀(OLS) 기준모형 구축

### 5-2-1. CDR 미적용 OLS 통행발생모형

#### (1) 총통행(Total Trip)발생 · 유입모형

##### 1) 모형 추정결과

O/D 총통행발생 · 유입량을 종속변수로 하여 F통계량 0.05를 기준으로 단계 선택법(stepwise)을 적용하여 다중선형회귀모형을 분석하였다. 분석결과, O/D 총통행발생모형의 결정계수( $R^2$ )는 0.693이며, 수정된 결정계수( $adjR^2$ ) 0.679, 총통행유입모형의  $R^2$ 는 0.672, 수정된 결정계수( $adjR^2$ ) 0.657으로 나타났다.

<표 5-3> 모형요약 (O/D 총통행발생 · 유입모형)

구분	$R^2$	$adjR^2$	AICc	F-value	pvalue
O/D total P	0.693	0.679	3113.35	49.018	0.000
O/D total A	0.672	0.657	3131.26	44.424	0.000



O/D 총통행발생 OLS모형에서 선택된 변수는 총 6개로서 공업지역 건축물 연상면적, 직주균형계수(JH), 초중고학생수, 3차산업종사자수, 판매시설, 버스정류장수로 도출되었다. 이 중 직주균형계수는 존 내의 직장과 주거지가 근 거리에 있는 경우의 지역내 자족도를 의미하는 것으로서 통행생성량에 음(-)의 영향을 주는 것으로 나타났다. 즉, 통행거리가 짧은 존 내부통행이 많은 경우에는 총 통행발생량에 음(-)의 영향을 주는 것을 알 수 있다. 구체적으로 모형의 회귀계수 추정결과는 표5-4와 같고 표준화계수(beta) 추정결과에 따라 변수별 모형의 영향력 정도를 살펴보면, 3차산업종사자수가 O/D 총통행발생량을 증가시키는 영향력이 가장 큰 것으로 나타난 반면 직주균형계수는 통행량 감소에 미치는 영향력이 가장 큰 결과를 보였다. 6개의 설명변수 모두 t-통계량이 1.96이상으로 종속변수를 추정하는데 통계적으로 유의한 것을 알 수 있고 다중공선성 문제에 있어서도 모두 분산팽창인자(VIF) 값이 모두 10 이하로 나타나 다중공선성의 발생 가능성도 크지 않은 것으로 보인다.

추정된 모형식은 다음과 식5-1과 같다.

$$OD_{total}^P = -89524.586 + 3929.669 \text{Industrial Floor Area} - 16968.474 \text{JH} \\ + 2.452 \text{EleMiddHighN} + 4.424 \text{ServiceWorkers} + 5580.765 \text{DsalesFac} \\ + 16737.480 \log(\text{Busstop}) \quad \dots\dots\dots \text{(식5.1)}$$

추정된 모형식으로부터 O/D 총통행발생모형에 대한 독립변수의 회귀계수 값에 대하여는 다음과 같이 해석할 수 있다.

첫째, O/D 총통행발생에 공업지역 건축물연상면적과 초중고학생수, 3차산업종사자수, 판매시설수, 버스정류장수는 양(+)의 상관관계를 나타내며, 이 변수



들이 증가함에 따라 O/D 총통행발생량도 역시 증가함을 나타낸다. 다른 모든 변수들을 통제하였을 때 O/D 총통행발생량은 회귀계수의 중요도가 가장 높은-표준화계수 베타값이 가장 큰-3차산업종사자수의 4.424배에 비례하고 다음으로 버스정류장수에 16737.48배에 비례한다고 볼 수 있다.

둘째, 직주균형계수의 경우 음(-)의 상관관계를 보였으며, 이 변수가 증가함에 따라 O/D 총통행발생량은 감소하는 것을 나타낸다. 이는 앞서 언급된 동일한 설명변수를 갖는 존이라 할지라도 직주균형계수가 높은 존은 상대적으로 O/D 총통행발생량이 적다는 것을 의미한다. 따라서, 존 내부의 통근통행 비율이 높을 경우 총통행발생량에는 음(-)의 영향을 주어 통행발생의 추정에 있어 존 크기가 중요한 변수가 될 수 있음을 의미한다.

<표 5-4> 회귀계수 추정결과 (O/D 총통행발생모형)

Category	Variables	Model (optimal)				
		Coefficient	beta	t-value	p-value	VIF
Production	(Constant)	-89524.59		-2.412	0.017	
Land Use	Industrial Floor area	3929.669	0.194	3.799	0	1.104
	JH	-166968.5	-0.117	-2.194	0.03	1.211
Social	EleMiddHighN	2.452	0.19	3.537	0.001	1.223
Factor	ServiceWorker	4.424	0.479	8.311	0	1.408
Facility	DsalesFac	5580.765	0.112	2.114	0.036	1.199
Factors	log(BusStop)	16737.48	0.202	3.638	0	1.301
F-value	49.018					
Prob>F	0					
$R^2$	0.693					
$adjR^2$	0.679					
AICc	3113.35					



O/D 총통행유입 OLS모형에서 선택된 변수 역시 총 6개로서 공업지역 건축물연상면적, 초중고학생수, 3차산업종사자수, 판매시설, 행정기관, 버스정류장수로 도출되었다. 구체적으로 모형의 회귀계수 추정결과는 표5-5와 같고 변수별 모형의 영향력 정도를 살펴보면, 총통행발생모형과 마찬가지로 3차산업종사자수가 O/D 총통행유입량을 증가시키는 영향력이 가장 큰 것으로 나타난 반면 판매시설수는 영향력이 가장 낮은 결과를 보였다. 설명변수 모두 t-통계량이 1.96이상으로 종속변수를 추정하는데 통계적으로 유의하고 VIF 값이 모두 10 이하로 나타나 다중공선성의 발생 가능성도 크지 않았다.

추정된 모형식은 다음과 식5-2와 같다.

$$OD_{total}^A = -10035.61 + 4875.776 \text{Industrial Floor Area} + 2.704 \text{EleMiddHighN} \\ + 4.623 \text{ServiceWorkers} + 5692.236 \text{DsalesFac} + 8545.344 \text{DOfficeF} \\ + 17799.813 \log(\text{Busstop}) \quad \dots\dots\dots (\text{식}5.2)$$

추정된 모형식으로부터 O/D 총통행유입모형에 대한 독립변수의 회귀계수 값에 대하여는 다음과 같이 해석할 수 있다.

첫째, O/D 총통행유입에 공업지역 건축물연상면적과 초중고학생수, 3차산업종사자수, 판매시설수, 행정기관수, 버스정류장수는 양(+)의 상관관계를 나타내며, 이 변수들이 증가함에 따라 O/D 총통행유입량도 역시 증가함을 나타낸다. 다른 모든 변수들을 통제하였을 때 O/D 총통행유입량은 회귀계수의 중요도가 가장 높은-표준화계수 베타값이 가장 큰-3차산업종사자수의 4.623배에 비례하고 다음으로 공업지역 건축물연상면적에 4875.776배에 비례한다고 볼 수 있다.





<표 5-5> 회귀계수 추정결과 (O/D 총통행유입모형)

Category	Variables	Model (optimal)				VIF
		Coefficient	beta	t-value	p-value	
Attraction	(Constant)	-10035.61		-1.78	0.077	
Land Use	Industrial Floor area	4875.776	0.231	4.416	0	1.083
Social	EleMiddHighN	2.704	0.201	3.668	0	1.191
Factor	ServiceWorkers	4.623	0.48	8.001	0	1.43
Facility Factors	DsalesFac	5692.236	0.11	1.986	0.049	1.216
	DOfficeF	8545.344	0.133	2.579	0.011	1.05
	log(BusStop)	17799.813	0.206	3.62	0	1.28
F-value	44.424					
Prob>F	0					
$R^2$	0.672					
$adjR^2$	0.657					
AICc	3131.26					

## (2) 필수적 통행(Mandatory Trip)발생 · 유입모형

### 1) 모형 추정결과

O/D 필수적통행발생 · 유입량을 종속변수로 하여 F통계량 0.05를 기준으로 단계선택법을 적용하여 다중선형회귀모형을 분석하였다. 분석결과, O/D 필수적 통행발생모형의 결정계수( $R^2$ )는 0.738이며, 수정된 결정계수( $adjR^2$ ) 0.723, 필수적통행유입모형의  $R^2$ 는 0.732, 수정된 결정계수( $adjR^2$ ) 0.713으로 나타났다.

<표 5-6> 모형요약 (O/D 필수적통행발생 · 유입모형)

구분	$R^2$	$adjR^2$	AICc	F-value	pvalue
O/D Mand P	0.738	0.724	2881.72	51.929	0.000
O/D Mand A	0.732	0.713	2890.88	38.548	0.000





O/D 필수적통행발생 OLS모형에서 선택된 변수는 총 7개로서 공업지역 건축물연상면적, 주거지역 건축물연상면적, 직주균형계수, 3차산업종사자수, 대학생수, 초중고학생수, 버스정류장수로 도출되었다. 이 중 직주균형계수는 총 통행발생모형과 마찬가지로 통행생성량에 음(-)의 영향을 주는 것으로 나타났다. 구체적으로 모형의 회귀계수 추정결과는 표5-7과 같고 변수별 모형의 영향력 정도를 살펴보면, 초중고학생수가 O/D 필수적통행발생량을 증가시키는 영향력이 가장 큰 것으로 나타난 반면 직주균형계수는 통행량 감소에 미치는 영향력이 가장 큰 결과를 보였다. 7개의 설명변수 모두 통계적으로 유의한 것을 알 수 있고 VIF 값도 모두 10 이하로 나타나 다중공선성의 발생 가능성도 크지 않은 것으로 보인다.

추정된 모형식은 다음과 식5-3과 같다.

$$OD_{Mand}^P = 48007.274 + 2399.465 \log(\text{Industrial Floor Area}) + 0.001 \text{Resid Floor Area} \\ - 134239.724 JH + 4.424 \text{ServiceWorkers} + 0.848 \text{UniStudN} + 1.735 \text{EleMiddHighN} \\ + 143.647 \text{Busstop}) \dots\dots\dots (\text{식5.3})$$

추정된 모형식으로부터 O/D 필수적통행발생모형에 대한 독립변수의 회귀계수 값에 대하여는 다음과 같이 해석할 수 있다.

첫째, O/D 필수적통행발생에 공업·주거지역 건축물연상면적과 3차산업종사자수, 대학생수, 초중고학생수, 버스정류장수는 양(+)의 상관관계를 나타내며, 이 변수들이 증가함에 따라 O/D 필수적통행발생량도 역시 증가함을 나타낸다. 다른 모든 변수들을 통제하였을 때 O/D 필수적통행발생량은 회귀계수의 중요도가 가장 높은-표준화계수 베타값이 가장 큰-초중고학생수의 1.735배



에 비례하고 다음으로 공업지역 건축물연상면적에 2399.465배 비례한다고 볼 수 있다.

둘째, 직주균형계수의 경우 총통행발생모형과 마찬가지로 음(-)의 상관관계를 보였으며, 이 변수가 증가함에 따라 O/D 필수적통행발생량은 감소하는 것을 나타낸다. 이는 앞서 언급된 동일한 설명변수를 갖는 존이라 할지라도 직주균형계수가 높은 존은 상대적으로 O/D 필수적통행발생량이 적다는 것을 의미한다. 따라서, 존 내부의 통근통학통행 비율이 높을 경우 필수적 통행발생량에는 음(-)의 영향을 주어 총통행발생과 마찬가지로 추정에 있어 존 크기가 중요한 변수가 될 수 있음을 의미한다.

<표 5-7> 회귀계수 추정결과 (O/D 필수적통행발생모형)

Category	Variables	Model (optimal)				
		Coefficient	beta	t-value	p-value	VIF
Production	(Constant)	48007.274		3.079	0.003	
Land Use	log(IndusFloorArea)	2399.465	0.251	5.346	0	1.082
	ResidFloorArea	0.001	0.182	3.04	0.003	1.764
	JH	-134239.724	-0.143	-2.796	0.006	1.292
Social Factor	ServiceWorkers	0.927	0.213	4.166	0	1.282
	UniStudN	0.848	0.225	4.819	0	1.071
	EleMiddHighN	1.735	0.284	5.058	0	1.558
Facility Factors	BusStop	143.647	0.222	4.504	0	1.196
F-value		51.929				
Prob>F		0				
$R^2$		0.738				
$adjR^2$		0.724				
AICc		2881.72				



O/D 필수적통행유입 OLS모형에서 선택된 변수는 총 9개로서 공업지역 건축물연상면적, 직주균형계수, 20~29세인구수, 대학생수, 초중고학생수, 총종사자수, 교육시설수, 행정기관수, 버스정류장수로 도출되었다. 구체적으로 모형의 회귀계수 추정결과는 표5-8과 같고 변수별 모형의 영향력 정도를 살펴보면, 공업지역 건축물연상면적이 O/D 필수적통행유입량을 증가시키는 영향력이 가장 큰 것으로 나타난 반면 직주균형계수는 영향력이 가장 낮은 결과를 보였다. 9개의 설명변수 모두 통계적으로 유의한 것을 알 수 있고 VIF 값이 모두 10 이하로 나타나 다중공선성의 발생 가능성도 크지 않았다.

추정된 모형식은 다음과 식5-4와 같다.

$$\begin{aligned}
 OD_{Mand}^A = & 49183.593 + 2223.43 \log(\text{IndustrialFloorArea}) - 142850.654JH \\
 & + 1.648age20\sim29 + 0.663UniStudN + 1.356EleMiddHighN + 0.494Nworkers \\
 & + 1075.013DMiddHighUniFac + 3163.073DOfficFac + 119.984BusStop \\
 & \dots\dots\dots \text{(식5.4)}
 \end{aligned}$$

추정된 모형식으로부터 O/D 필수적통행유입모형에 대한 독립변수의 회귀계수 값에 대하여는 다음과 같이 해석할 수 있다.

첫째, O/D 필수적통행유입에 공업지역 건축물연상면적과 20~29세인구수, 대학생수, 초중고학생수, 총종사자수, 교육시설수, 행정기관수, 버스정류장수는 양(+)의 상관관계를 나타내며, 이 변수들이 증가함에 따라 O/D 필수적통행유입량도 역시 증가함을 나타낸다. 다른 모든 변수들을 통제하였을 때 O/D 필수적통행유입량은 회귀계수의 중요도가 가장 높은-표준화계수 베타값이 가장 큰-공업지역 건축물연상면적 2223.43배에 비례하고 다음으로 초중고학생수의 1.356배에 비례한다고 볼 수 있다.



<표 5-8> 회귀계수 추정결과 (O/D 필수적통행유입모형)

Category	Variables	Model (optimal)				VIF
		Coefficient	beta	t-value	p-value	
Production	(Constant)	49183.593		2.762	0.007	
Land Use	log(IndusFloor Area)	2223.430	0.229	4.668	0.000	1.145
	JH	-142850.654	-0.151	-2.650	0.009	1.532
Social Factor	age20 29	1.648	0.173	2.235	0.027	2.845
	UniStudN	0.663	0.174	3.419	0.001	1.223
	EleMiddHighN	1.356	0.220	2.794	0.006	2.929
	Nworkers	0.494	0.198	3.529	0.001	1.493
Facility Factors	DMidHighUni Fac	1075.013	0.153	2.349	0.020	2.020
	DOfficFac	3163.073	0.107	2.256	0.026	1.068
	BusStop	119.984	0.183	3.440	0.001	1.345
F-value	38.548					
Prob>F	0.000					
$R^2$	0.732					
$adjR^2$	0.713					
AICc	2890.88					

### (3) 선택적 통행(Optional Trip)발생 · 유입모형

#### 1) 모형 추정결과

O/D 선택적통행발생 · 유입량을 종속변수로 하여 F통계량 0.05를 기준으로 단계선택법을 적용하여 다중선택회귀모형을 분석하였다. 분석결과, O/D 선택적통행발생모형의 결정계수( $R^2$ )는 0.682이며, 수정된 결정계수( $adjR^2$ ) 0.670, 선택적통행유입모형의  $R^2$ 는 0.636, 수정된 결정계수( $adjR^2$ ) 0.619으로 나타났다.



<표 5-9> 모형요약 (O/D 선택적통행발생 · 유입모형)

구분	$R^2$	$adjR^2$	AICc	F-value	pvalue
O/D Opt P	0.682	0.670	2976.15	56.186	0.000
O/D Opt A	0.636	0.619	3019.01	37.853	0.000

O/D 선택적통행발생 OLS모형에서 선택된 변수는 총 5개로서 상업 · 주거 · 공업지역 건축물연상면적, 3차산업종사자수, 판매시설수로 도출되었다. 구체적으로 모형의 회귀계수 추정결과는 표5-10과 같고 변수별 모형의 영향력 정도를 살펴보면, 3차산업종사자수가 O/D 선택적통행발생량을 증가시키는 영향력이 가장 큰 것으로 나타난 반면 판매시설수는 영향력이 가장 작은 결과를 보였다. 5개의 설명변수 모두 VIF 값이 모두 10 이하로 나타나 다중공선성의 발생 가능성도 크지 않았다.

추정된 모형식은 다음과 식5-5와 같다.

$$\begin{aligned}
 OD_{Opt}^P = & 3626.894 + 0.001CommerFloorArea + 0.001ResidFloorArea \\
 & + 2379.617\log(IndusFloorArea) + 2.81ServiceWorkers + 5269.05DsalesFac \\
 & \dots\dots\dots (식5.5)
 \end{aligned}$$

추정된 모형식으로부터 O/D 선택적통행발생모형에 대한 독립변수의 회귀계수 값에 대하여는 다음과 같이 해석할 수 있다.

첫째, O/D 선택적통행발생에 상업 · 공업 · 주거지역 건축물연상면적과 3차 산업종사자수, 판매시설수는 양(+)의 상관관계를 나타내며, 이 변수들이 증가함에 따라 O/D 선택적통행발생량도 역시 증가함을 나타낸다. 다른 모든 변수



들을 통제하였을 때 O/D 선택적통행발생량은 회귀계수의 중요도가 가장 높은-표준화계수 베타값이 가장 큰-3차산업종사자수의 2.81배에 비례하고 다음으로 상업지역 건축물연상면적에 0.001배 비례한다고 볼 수 있다.

<표 5-10> 회귀계수 추정결과 (O/D 선택적통행발생모형)

Category	Variables	Model (optimal)				
		Coefficient	beta	t-value	p-value	VIF
Production	(Constant)	3626.894		2.215	0.028	
	CommerFloor Area	0.001	0.275	5.116	0.000	1.190
Land Use	ResidFloorArea	0.001	0.206	3.903	0.000	1.145
	log(IndusFloor Area)	2379.617	0.185	3.735	0.000	1.016
Social Factor	ServiceWorkers	2.810	0.481	8.296	0.000	1.384
Facility Factors	DSalesFac	5269.050	0.168	3.007	0.003	1.281
F-value	56.186					
Prob>F	0.000					
$R^2$	0.682					
$adjR^2$	0.670					
AICc	2976.15					

O/D 선택적통행유입 OLS모형에서 선택된 변수는 총 6개로서 상업·공업·주거지역 건축물연상면적, 3차산업종사자수, 행정기관수, 판매시설수로 도출되었다. 구체적으로 모형의 회귀계수 추정결과는 표5-11과 같고 변수별 모형의 영향력 정도를 살펴보면, 3차산업종사자수가 O/D 선택적통행유입량을 증가시키는 영향력이 가장 큰 것으로 나타난 반면 주거지역 건축물연상면적의



영향력이 가장 낮은 결과를 보였다. 6개의 설명변수 모두 통계적으로 유의한 것을 알 수 있고 VIF값이 모두 10 이하로 나타나 다중공선성의 발생 가능성도 크지 않았다.

추정된 모형식은 다음과 식5-6와 같다.

$$OD_{Opt}^A = 2382.096 + 0.001CommerFloorArea + 282.277\log(IndusFloorArea) + 0.001ResidFloorArea + 3.036ServiceWorkers + 6970.381DOfficFac + 5175.248DSalesFac \dots\dots\dots (식5.6)$$

추정된 모형식으로부터 O/D 선택통행유입모형에 대한 독립변수의 회귀계수에 대하여는 다음과 같이 해석할 수 있다.

첫째, O/D 선택적통행유입에 상업·공업·주거지역 건축물연상면적과 3차산업종사자수, 행정기관수, 판매시설수는 양(+)의 상관관계를 나타내며, 이 변수들이 증가함에 따라 O/D 선택적통행유입량도 역시 증가함을 나타낸다. 다른 모든 변수들을 통제하였을 때 O/D 선택적통행유입량은 회귀계수의 중요도가 가장 높은-표준화계수 베타값이 가장 큰-3차산업종사자수의 3.036배에 비례하고 다음으로 상업지역 건축물연상면적의 0.001배에 비례한다고 볼 수 있다.

<표 5-11> 회귀계수 추정결과 (O/D 선택적통행유입모형)

Category	Variables	Model (optimal)				
		Coefficient	beta	t-value	p-value	VIF
Attraction	(Constant)	2382.096		1.258	0.211	
	CommerFloorArea	0.001	0.234	4.051	0.000	1.195



Category	Variables	Model (optimal)				
Attraction		Coefficient	beta	t-value	p-value	VIF
	log(IndusFloor Area)	2862.277	0.207	3.866	0.000	1.026
	ResidFloorArea	0.001	0.136	2.403	0.018	1.149
Social Factor	ServiceWorkers	3.036	0.483	7.645	0.000	1.423
Facility	DOfficFac	6970.381	0.166	3.047	0.003	1.054
Factors	DSalesFac	5175.248	0.153	2.540	0.012	1.295
F-value	37.853					
Prob>F	0.000					
$R^2$	0.636					
$adjR^2$	0.619					
AICc	3019.01					

### 5-2-2. CDR 적용 OLS 통행발생모형

본 연구에서는 CDR을 주요한 설명변수로 적용하여 통행발생모형의 추정력 개선여부를 살펴보기 위하여 각 통행목적별로 CDR의 주중·주말 침두시 데이터를 모두 적용해보고 F통계량 0.05를 기준으로 단계선택법을 통해 유의미한 CDR 데이터를 도출하여 본다.

#### (1) 총통행(Total Trip)발생·유입모형

##### 1) 모형 추정결과

O/D 총통행발생·유입량을 종속변수로 CDR 데이터를 적용하여 F통계량 0.05를 기준으로 단계선택법을 통해 다중선형회귀모형을 분석하였다.

분석결과, O/D 총통행발생모형의 결정계수( $R^2$ )는 0.727이며, 수정된 결정계





수( $adjR^2$ ) 0.710, 총통행유입모형의  $R^2$ 는 0.730, 수정된 결정계수( $adjR^2$ ) 0.712으로 나타났다.

<표 5-12> 모형요약 (CDR적용 OLS 총통행발생 · 유입모형)

구분	$R^2$	$adjR^2$	AICc
O/D total P	0.727	0.710	3085.74
O/D total A	0.730	0.712	3097.62

O/D 총통행발생 OLS모형에서 선택된 변수는 CDR평일 오전침두시 통행량 데이터를 포함하여 총 5개로서 공업지역 건축물연상면적과 3차산업종사자수, 초중고학생수, 버스정류장수로 도출되었다. 기존 CDR 미적용 OLS 총통행발생모형에 있던 직주균형계수와 판매시설수 변수는 CDR평일 오전침두시 통행량이 해당 변수보다 모형에 유의미한 영향을 주어 p값이 0.10보다 커진 설명변수로서 제외되었다.

구체적으로 모형의 회귀계수 추정결과는 표5-13과 같고 t값을 바탕으로 CDR 평일 오전침두시 통행량과 3차산업종사자수, 초중고학생수의 사회경제지표가 총통행발생모형에 큰 영향을 주는 것으로 나타났다.

추정된 모형식은 다음 식5-7과 같다.

$$\begin{aligned}
 CDR\ OD_{total}^P = & -8184.483 + 0.593CDRMandMpeak + 3176.876\log(IndusFloorArea) \\
 & + 3.058ServiceWorkers + 3.275EleMiddHighN \\
 & + 11189.255\log(BusStop) \dots\dots\dots (식5.7)
 \end{aligned}$$



추정된 모형식으로부터 CDR이 포함된 O/D 총통행발생모형에 대한 독립변수의 회귀계수 값에 대하여는 다음과 같이 해석할 수 있다.

O/D 총통행발생에 CDR평일 오전첨두시 통신량과 공업지역 건축물연상면적, 3차산업종사자수, 초중고학생수, 버스정류장수는 양(+)의 상관관계를 나타내며, 이 변수들이 증가함에 따라 O/D 총통행발생량도 역시 증가함을 나타낸다. 다른 모든 변수들을 통제하였을 때 O/D 총통행발생량은 CDR 평일 오전첨두시 통행량의 0.593배에 비례하고 3차산업종사자수는 3.058배, 초중고학생수는 3.275배에 비례한다고 볼 수 있다.

<표 5-13> 회귀계수 추정결과 (CDR적용 O/D 총통행발생모형)

Category	Variables	Model (OLS)		
		b	Std.error	t-value
Production	(Constant)	-8184.483	20388.486	-0.401
CDR	CDR Mand Morning	0.593	0.105	5.666
Land Use	IndustrialFloorArea	3176.876	1011.492	3.141
Social Factor	ServiceWorker	3.058	0.566	5.400
	EleMiddHighN	3.275	0.650	5.038
Facility Factors	log(BusStop)	11189.255	4291.267	2.607
$R^2$		0.727		
$adjR^2$		0.710		
AICc		3085.74		

O/D 총통행유입 OLS모형에서 선택된 변수는 CDR평일 오전첨두시 통신량 데이터를 포함하여 총 6개로서 공업지역 건축물연상면적과 3차산업종사자수, 초중고학생수, 행정기관수, 버스정류장수로 도출되었다. 기존 CDR 미적용 총



통행유입모형에 있던 직주균형계수와 판매시설수는 CDR이 해당 변수보다 모형에 유의미한 영향을 주어 p값이 0.10보다 커진 설명변수로서 제외되었다.

구체적으로 모형의 회귀계수 추정결과는 표5-14와 같고 t값을 바탕으로 CDR 평일 오전침두시 통행량과 3차산업종사자수, 초중고학생수의 사회경제지표가 총통행발생모형에 큰 영향을 주는 것으로 나타났다.

추정된 모형식은 다음 식5-8과 같다.

$$CDR OD_{total}^A = -19895.821 + 0.674CDRMandMpeak + 3532.677\log(IndusFloorArea) + 2.842ServiceWorkers + 3.308EleMiddHighN + 8980.907DOOfficeF + 10663.421\log(BusStop) \dots\dots\dots (식5.8)$$

추정된 모형식으로부터 CDR이 포함된 O/D 총통행유입모형에 대한 독립변수의 회귀계수 값에 대하여는 다음과 같이 해석할 수 있다.

O/D 총통행유입에 CDR평일 오전침두시 통행량과 공업지역 건축물연상면적, 3차산업종사자수, 초중고학생수, 행정기관수, 버스정류장수는 양(+)의 상관관계를 나타내며, 이 변수들이 증가함에 따라 O/D 총통행유입량도 역시 증가함을 나타낸다. 다른 모든 변수들을 통제하였을 때 O/D 총통행유입량은 CDR 평일 오전침두시 통행량의 0.674배에 비례하고 3차산업종사자수는 2.842배, 초중고학생수는 3.308배에 비례한다고 볼 수 있다.

<표 5-14> 회귀계수 추정결과 (CDR적용 O/D 총통행유입모형)

Category	Variables	Model (OLS)		
		b	Std.error	t-value
Production	(Constant)	-19895.821	22175.416	-0.897



Category	Variables	Model (OLS)		
		b	Std.error	t-value
Production				
CDR	CDR Mand Morning	0.674	0.109	6.196
Land Use	IndustrialFloorArea	3532.677	1052.180	3.357
Social Factor	ServiceWorker	2.842	0.600	4.737
	EleMiddHighN	3.308	0.676	4.896
Facility Factors	DOfficeF	8980.907	2955.987	3.038
	log(BusStop)	10663.421	4458.424	2.392
$R^2$		0.730		
$adjR^2$		0.712		
AICc		3097.62		

## 2) 모형의 잔차분석

O/D 총통행량 기준 CDR이 적용된 OLS 통행발생·유입모형 도출에 따라 종속변수의 등분산 가정을 증명하기 위하여 표준화잔차(Std.Resid)에 대하여 전역적 공간자기상관(Moran's I)을 분석해보았다. 그 결과, Moran's I값은 0.115996, 유의확률은 0.013510으로 나타나 잔차가 공간적을 무작위하게 퍼져 있다는 귀무가설을 기각하여 잔차의 공간자기상관이 존재함을 확인하였다.

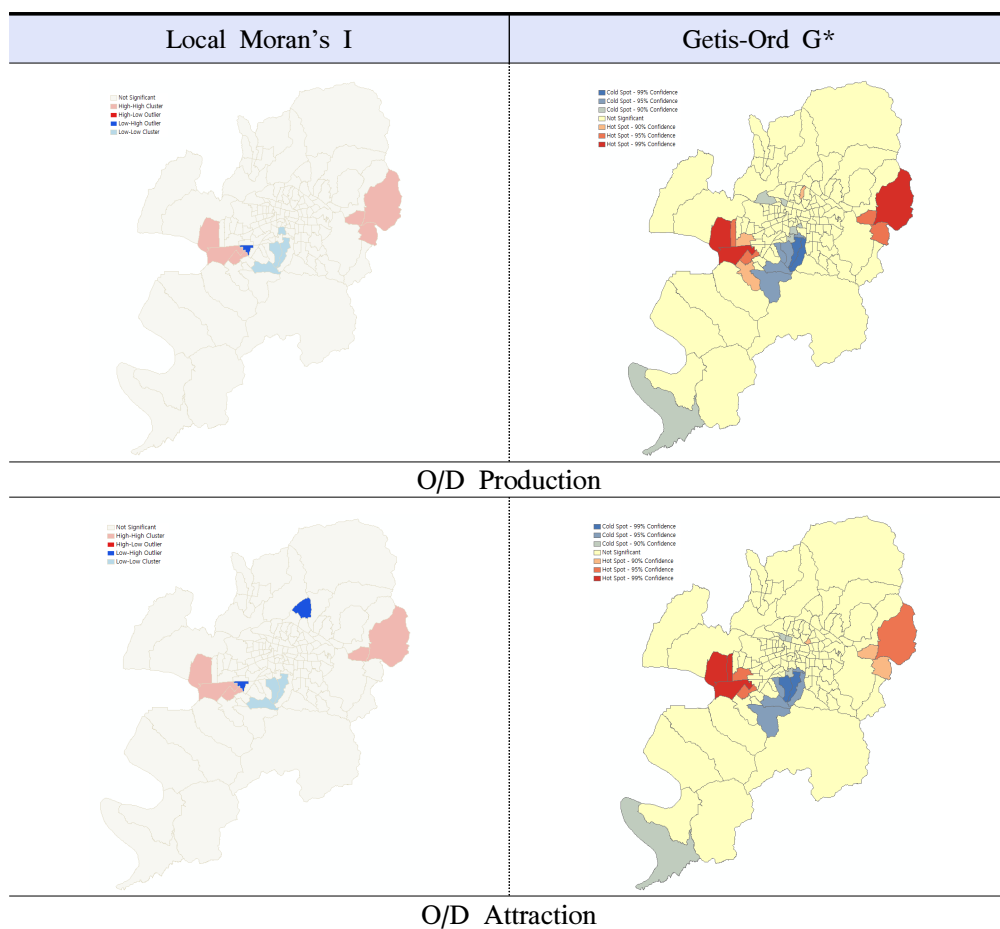
- 귀무가설 H0 : 모형의 표준화잔차는 공간적 자기상관이 없다.
- 대립가설 H1 : 모형의 표준화잔차는 공간적 자기상관이 있다.

<표 5-15> O/D total OLS기준모형의 오차의 공간자기상관 검증결과

구분	Moran's I	ZScore	Pvalue
O/D total P	0.152801	3.194219	0.001402
O/D total A	0.108184	2.308632	0.020964



국지적 공간상관성 탐색을 위하여 Local Moran's I와 Getis-Ord G\* 방법을 통해 공간적 군집을 확인해본 결과, 달서구 신당동, 월성1·2동, 진천동 일원과 동구 안심동 일원, 수성구 고산동 등에서 총통행발생·유입량의 추정치는 관측치보다 높은 값을 보이거나, 남구 대명동, 서구 봉덕동, 비산동, 원대동 일원 등에서는 추정치가 관측치보다 낮은 값을 보이는 것으로 나타났다.



<그림 5-1> CDR total OLS기준 모형의 국지적인 공간자기상관 분석결과



## (2) 필수적 통행(Mandatory Trip)발생 · 유입모형

### 1) 모형 추정결과

O/D 필수적통행발생 · 유입량을 종속변수로 CDR 데이터를 적용하여 단계 선택법을 통해 다중선형회귀모형을 분석하였다. 분석결과, O/D 필수적통행발생모형의 결정계수( $R^2$ )는 0.781이며, 수정된 결정계수( $adjR^2$ ) 0.766, 필수적통행유입모형의  $R^2$ 는 0.791, 수정된 결정계수( $adjR^2$ ) 0.774으로 나타났다.

<표 5-16> 모형요약 (CDR적용 O/D Mand PA)

구분	$R^2$	$adjR^2$	AICc
O/D Mand P	0.781	0.766	2848.99
O/D Mand A	0.791	0.774	2849.80

O/D 필수적통행발생 OLS모형에서 선택된 변수는 CDR평일 오전침두시 통신량 데이터를 포함하여 총 6개로서 공업 · 주거지역 건축물연상면적과 대학생수, 초중고학생수, 버스정류장수로 도출되었다. 기존 CDR 미적용 필수적통행발생모형에 있던 직주균형계수와 3차산업종사자수는 CDR이 해당 변수보다 모형에 유의미한 영향을 주어 p값이 0.10보다 커진 설명변수로서 제외되었다. 구체적으로 모형의 회귀계수 추정결과는 표5-17과 같고 t값을 바탕으로 CDR 평일 오전침두시 통행량과 대학생수, 초중고학생수의 사회경제지표가 필수적통행발생모형에 큰 영향을 주는 것으로 나타났다.

추정된 모형식은 다음 식5-9와 같다.



$$CDR OD_{Mand}^P = 9483.46 + 0.269CDRMandMpeak + 2013.53\log(IndusFloorArea) \\ + 0.001ResidFloorArea + 0.828UniStudN + 1.861EleMiddHighN \\ + 111.838BusStop \dots\dots\dots (식5.9)$$

추정된 모형식으로부터 CDR이 포함된 O/D 필수적통행발생모형에 대한 독립변수의 회귀계수 값에 대하여는 다음과 같이 해석할 수 있다.

O/D 필수적통행발생에 CDR평일 오전침두시 통행량과 공업·주거지역 건축물연상면적, 대학생수, 초중고학생수, 버스정류장수는 양(+)의 상관관계를 나타내며, 이 변수들이 증가함에 따라 O/D 필수적통행발생량도 역시 증가함을 나타낸다. 다른 모든 변수들을 통제하였을 때 O/D 필수적 통행발생량은 CDR 평일 오전침두시 통행량의 0.269배에 비례하고 대학생수는 0.828배, 초중고학생수는 1.861배에 비례한다고 볼 수 있다.

<표 5-17> 회귀계수 추정결과 (CDR적용 O/D 필수적통행발생모형)

Category	Variables	Model (OLS)		
		b	Std.error	t-value
Production	(Constant)	9483.460	6356.555	1.492
CDR	CDR Mand Morning	0.269	0.044	6.106
Land Use	log(IndustrialFloorArea)	2013.530	416.471	4.835
	ResidFloorArea	0.001	0.000	4.931
Social Factor	UniStudN	0.828	0.160	5.162
	EleMiddHighN	1.861	0.313	5.937
Facility Factors	log(BusStop)	111.838	29.507	3.790
$R^2$		0.781		
$adjR^2$		0.766		
AICc		2848.99		



O/D 필수적통행유입 OLS모형에서 선택된 변수는 CDR평일 오전침두시 통신량 데이터를 포함하여 총 7개로서 공업지역 건축물연상면적과 20~29세 인구수, 대학생수, 초중고학생수, 행정기관수, 버스정류장수로 도출되었다. 기존 CDR 미적용 필수적통행유입모형에 있던 직주균형계수와 총종사자수, 교육시설수는 CDR이 해당 변수보다 모형에 유의미한 영향을 주어 p값이 0.10보다 커진 설명변수로서 제외되었다. 구체적으로 모형의 회귀계수 추정결과는 표 5-18과 같고 t값을 바탕으로 CDR평일 오전침두시 통행량과 20~29세 인구수, 공업지역 건축물연상면적이 필수적통행유입모형에 큰 영향을 주는 것으로 나타났다.

추정된 모형식은 다음 식5-10과 같다.

$$CDR OD_{Mand}^A = 8637.661 + 0.28CDRMandMpeak + 2010.84\log(IndusFloorArea) + 3.004age20\sim29 + 0.662UniStudN + 1.217EleMiddHighN + 3346.595DOOfficeF + 106.189BusStop \dots\dots\dots (식5.10)$$

추정된 모형식으로부터 CDR이 포함된 O/D 필수적통행유입모형에 대한 독립변수의 회귀계수 값에 대하여는 다음과 같이 해석할 수 있다.

O/D 필수적통행유입에 CDR평일 오전침두시 통신량과 공업지역 건축물연상면적, 20~29세 인구수, 대학생수, 초중고학생수, 행정기관수, 버스정류장수는 양(+)의 상관관계를 나타내며, 이 변수들이 증가함에 따라 O/D 필수적통행유입량도 역시 증가함을 나타낸다. 다른 모든 변수들을 통제하였을 때 O/D 필수적통행유입량은 CDR 평일 오전침두시 통행량의 0.28배에 비례하고 20~29세인구수는 3.004배, 공업지역 건축물연상면적은 2010.84배에 비례한다고 볼 수 있다.





<표 5-18> 회귀계수 추정결과 (CDR적용 O/D 필수적통행유입모형)

Category	Variables	Model (OLS)		
		b	Std.error	t-value
Production	(Constant)	8637.661	6524.678	1.324
CDR	CDR Mand Morning	0.280	0.042	6.731
Land Use	log(IndustrialFloorArea)	2010.840	416.744	4.825
Social Factor	age20~29	3.004	0.573	5.238
	UniStudN	0.662	0.170	3.893
	EleMiddHighN	1.217	0.420	2.899
Facility Factors	DOfficeF	3346.595	1211.962	2.761
	BusStop	106.189	30.686	3.461
	$R^2$	0.791		
	$adjR^2$	0.774		
	AICc	2849.80		

## 2) 모형의 잔차분석

O/D 필수적 통행량 기준 CDR이 적용된 OLS 통행발생·유입모형의 표준화 잔차 전역적 공간자기상관을 분석결과, Moran's I값은 0.115996, 유의확률은 0.013510으로 나타나 잔차가 공간적을 무작위하게 퍼져있다는 귀무가설을 기각하지 아니하여 잔차의 공간자기상관이 존재하지 않는 것으로 나타났다.

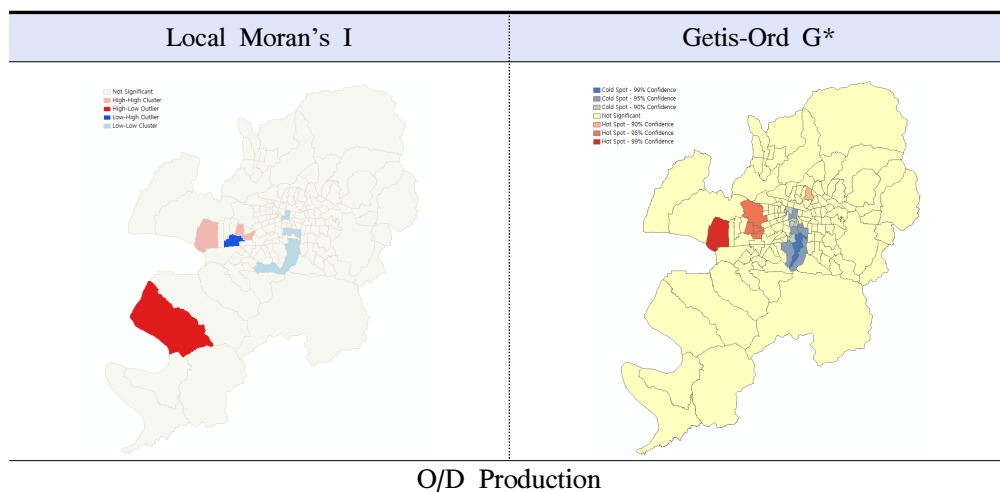
- 귀무가설 H0 : 모형의 표준화잔차는 공간적 자기상관이 없다.
- 대립가설 H1 : 모형의 표준화잔차는 공간적 자기상관이 있다.

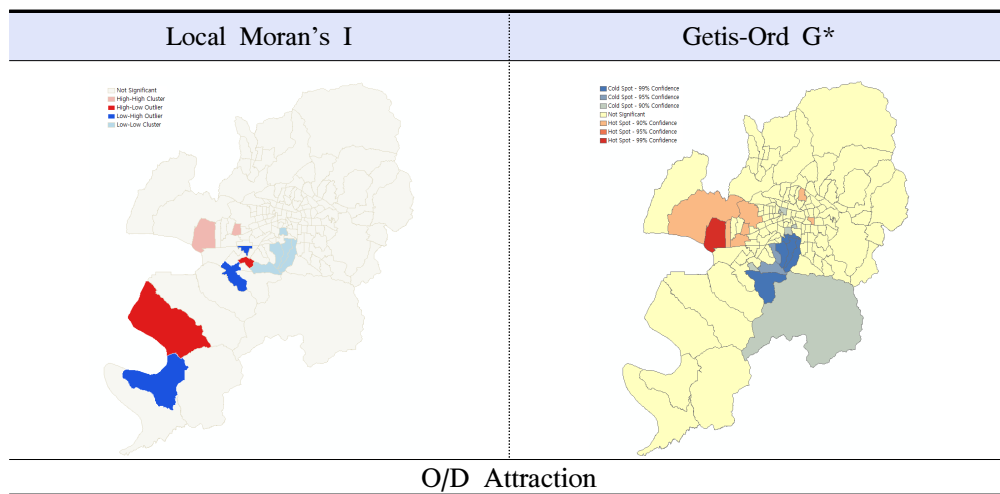


<표 5-19> O/D Mand OLS기준모형의 오차의 공간자기상관 검증결과

구분	Moran's I	ZScore	Pvalue
O/D Mand P	0.053402	1.222969	0.22142
O/D Mand A	0.061481	1.391294	0.164136

국지적 공간상관성 탐색을 위하여 Local Moran's I와 Getis-Ord G\* 방법을 통해 공간적 군집을 확인해본 결과, 달성군 논공읍, 다사읍, 달서구 신당동, 용산동, 감삼동, 상인1동동 일원과 북구 산격3동, 수성구 범어3동 등에서 필수적 통행발생유입량의 추정치는 관측치보다 높은 값을 보이거나, 달서구 장기동, 진천동, 상인3동, 중구 성내2동, 남구 대명2·9동, 이천동, 달성군 가창면 일원 추정치가 관측치보다 낮은 값을 보이는 것으로 나타났다.





<그림 5-2> CDR Mand OLS기준 모형의 국지적인 공간자기상관 분석결과

### (3) 선택적 통행(Optional Trip)발생 · 유입모형

#### 1) 모형 추정결과

O/D 선택적통행발생 · 유입량을 종속변수로 CDR 데이터를 적용하여 단계 선택법을 통해 다중선행회귀모형을 분석하였다. 분석결과, O/D 선택적통행발생모형의 결정계수( $R^2$ )는 0.701이며, 수정된 결정계수( $adjR^2$ ) 0.685, 선택적통행유입모형의  $R^2$ 는 0.6571, 수정된 결정계수( $adjR^2$ ) 0.636으로 나타났다.

<표 5-20> 모형요약 (CDR적용 O/D Opt PA)

구분	$R^2$	$adjR^2$	AICc
O/D Opt P	0.701	0.685	2968.82
O/D Opt A	0.657	0.636	3010.43



O/D 선택적통행발생 OLS모형에서 선택된 변수는 CDR주말 오전침두시 통신량 데이터를 포함하여 총 6개로서 상업·주거·공업지역 건축물연상면적과 3차산업종사자수, 판매시설수로 도출되었다. 기존 CDR 미적용 총통행발생모형에 있던 직주균형계수와 3차산업종사자수는 CDR이 해당 변수보다 모형에 유의미한 영향을 주어 p값이 0.10보다 커진 설명변수로서 제외되었다. 구체적으로 모형의 회귀계수 추정결과는 표5-21과 같고 t값을 바탕으로 3차산업종사자수와 상업·주거지역 건축물연상면적 변수가 선택적통행발생모형에 큰 영향을 주는 것으로 나타났다.

추정된 모형식은 다음 식5-11과 같다.

$$\begin{aligned}
 CDR\ OD_{Opt}^P = & 2641.544 + 0.119CDROptMpeak + 0.001CommerFloorArea \\
 & + 0.001ResidFloorArea + 2109.506\log(IndusFloorArea) \\
 & + 2.154ServiceWorkers + 4087.556DsalesFac \\
 & \dots\dots\dots (식5.11)
 \end{aligned}$$

추정된 모형식으로부터 CDR이 포함된 O/D 선택적통행발생모형에 대한 독립변수의 회귀계수 값에 대하여는 다음과 같이 해석할 수 있다.

O/D 선택적통행발생에 CDR주말 오전침두시 통신량과 상업·주거·공업지역 건축물연상면적, 3차산업종사자수, 판매시설수는 양(+)의 상관관계를 나타내며, 이 변수들이 증가함에 따라 O/D 선택적통행발생량도 역시 증가함을 나타낸다. 다른 모든 변수들을 통제하였을 때 O/D 선택적 통행발생량은 CDR주말 오전침두시 통행량의 0.119배에 비례하고 3차산업종사자수는 2.154배에 비례한다고 볼 수 있다.



<표 5-21> 회귀계수 추정결과 (CDR적용 O/D 선택적통행발생모형)

Category	Variables	Model (OLS)		
		b	Std.error	t-value
Production	(Constant)	2641.544	1612.411	1.638
CDR	CDR Opt Morning	0.119	0.039	3.069
	CommerFloorArea	0.001	0.000	4.547
Land Use	ResidFloorArea	0.001	0.000	4.530
	log(IndusFloorArea)	2109.506	622.185	3.390
Social Factor	ServiceWorkers	2.154	0.376	5.727
Facility Factors	DSalesFac	4087.556	1737.636	2.352
	$R^2$	0.701		
	$adjR^2$	0.685		
	AICc	2968.82		

O/D 선택적통행유입 OLS모형에서 선택된 변수는 CDR주말 오전첨두시 통  
신량 데이터를 포함하여 총 7개로서 상업·공업·주거지역 건축물연상면적과  
3차산업종사자수, 행정기관수, 버스정류장수로 도출되었다. 구체적으로 모형  
의 회귀계수 추정결과는 표5-22와 같고 t값을 바탕으로 3차산업종사자수와 상  
업·공업지역 건축물연상면적, CDR 주말 오전첨두시 통행량이 선택적통행  
유입모형에 큰 영향을 주는 것으로 나타났다.

추정된 모형식은 다음 식5-12와 같다.

$$\begin{aligned}
 CDR OD_{Opt}^A = & 1006.217 + 0.146CDROptMpeak + 0.001CommerFloorArea \\
 & + 2519.668\log(IndusFloorArea) + 0.001ResidFloorArea \\
 & + 2.262ServiceWorkers + 6598.717DOOfficeFac + 3677.17DSalesFac \\
 & \dots\dots\dots (식5.12)
 \end{aligned}$$



추정된 모형식으로부터 CDR이 포함된 O/D 선택적통행유입모형에 대한 독립변수의 회귀계수 값에 대하여는 다음과 같이 해석할 수 있다.

O/D 선택적통행유입에 CDR주말 오전첨두시 통신량과 상업·공업·주거지역 건축물연상면적, 3차산업종사자수, 행정기관수, 판매시설수는 양(+)의 상관관계를 나타내며, 이 변수들이 증가함에 따라 O/D 선택적통행유입량도 역시 증가함을 나타낸다. 다른 모든 변수들을 통제하였을 때 O/D 선택적통행유입량은 CDR 주말 오전첨두시 통행량의 0.146배에 비례하고 3차산업종사자수는 2.262배에 비례한다고 볼 수 있다.

<표 5-22> 회귀계수 추정결과 (CDR적용 O/D 선택적통행유입모형)

Category	Variables	Model (OLS)		
		b	Std.error	t-value
Production	(Constant)	1006.217	1869.293	0.538
	CDR			
	CDR Opt Morning	0.146	0.045	3.264
Land Use	CommerFloorArea	0.001	0.000	3.426
	log(IndusFloorArea)	2519.668	722.816	3.486
	ResidFloorArea	0.001	0.000	3.057
Social Factor	ServiceWorkers	2.262	0.437	5.173
Facility Factors	DOfficFac	6598.717	2076.843	3.177
	DSalesFac	3677.170	2018.446	1.822
$R^2$		0.657		
$adjR^2$		0.636		
AICc		3010.43		



## 2) 모형의 잔차분석

O/D 선택적 통행량 기준 CDR이 적용된 OLS 통행발생·유입모형의 표준화 잔차(Std.Resid) 전역적 공간자기상관을 분석결과, Moran's I값은 0.115996, 유의확률은 0.013510으로 나타나 잔차가 공간적을 무작위하게 퍼져있다는 귀무가설을 기각하지 아니하여 잔차의 공간자기상관이 존재하지 않는 것으로 나타났다.

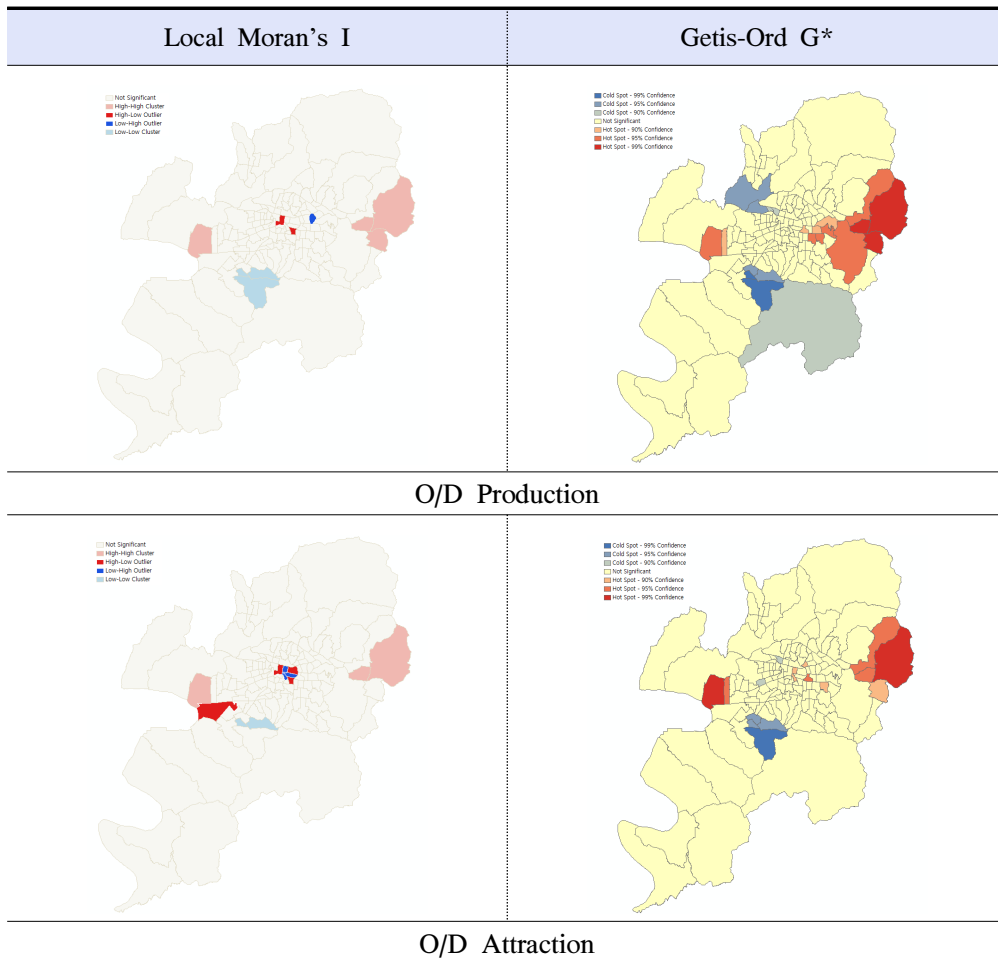
- 귀무가설  $H_0$  : 모형의 표준화잔차는 공간적 자기상관이 없다.
- 대립가설  $H_1$  : 모형의 표준화잔차는 공간적 자기상관이 있다.

<표 5-23> O/D Opt OLS기준모형의 오차의 공간자기상관 검증결과

구분	Moran's I	ZScore	Pvalue
O/D Opt P	0.074529	1.605898	0.108296
O/D Opt A	0.002362	0.190314	0.849063

국지적 공간상관성 탐색을 위하여 Local Moran's I와 Getis-Ord  $G^*$  방법을 통해 공간적 군집을 확인해본 결과, 달서구 신당동, 안심동 일원, 북구 고산동, 수성구 범어동, 만촌동, 중구 동인동, 대봉1동, 성내2동 등에서 선택적 통행발생유입량의 추정치는 관측치보다 높은 값을 보이나, 달성구 상인동, 도원동, 관문동, 서구 비산동, 원대동, 내당4동, 중구 삼덕동, 성내1동 일원에서는 추정치가 관측치보다 낮은 값을 보이는 것으로 나타났다.





<그림 5-3> CDR Opt OLS기준 모형의 국지적인 공간자기상관 분석결과

### 5-3. 통행발생 지리적 가중회귀분석(GWR) 모형 도출

공간성을 내포하고 있는 데이터는 일반적으로 공간자기상관의 관점에서 공간종속성과 공간이질성이라는 개념을 가지고 있다. 본 연구의 주요 분석대상인 통행량의 경우 공간종속성은 통행량에 영향을 미치는 토지이용, 교통결절





점, 주거·상업지역의 밀집 등의 공간적인 의존성을 가지는 변수에 영향을 받는 것 즉, 종속변수인 통행량이 해당 존의 속성이 인접한 다른 존의 속성에 영향을 받는다는 것으로 이해할 수 있다. 이와 반대로 공간이질성은 불균등한 공간적 영향으로 해당 존이 종속변수인 통행량의 결정에 미치는 영향이 균일하게 나타나지 않는 것을 의미한다. 이는 선형회귀모형에 이분산성과 유사한 의미를 가지는 것으로 회귀모형의 집계단위 즉, 존의 공간이질성과 이분산성은 같은 의미를 가진다(정우현, 2011).

이를 해결하기 위해 본 절에서는 실제 통행발생·유입량이 이러한 공간종속성 또는 공간이질성의 공간자기상관을 가지고 있는지에 대해 우선 검증하고 이를 바탕으로 국지적 공간회귀모형인 지리적 가중회귀분석(GWR)을 이용하여 공간성이 감안된 통행발생모형을 새롭게 구축하고자 한다. 이때 사용되는 GWR 분석방법은 공간종속성 뿐만 아니라 공간이질성까지 반영하여 개별 존마다 각기 다른 회귀식이 산출되어 분석대상지역내 존별 회귀계수를 도출함으로써 변수들의 국지적인 영향을 파악할 수 있는 결과를 제시하는 장점이 있다.

### 5-3-1. 공간상관성 검증

#### (1) 총통행발생·유입량 기준

O/D 총통행발생·유입량과의 전역적 공간종속성 검정 결과, 공간가중치 행렬을 선정하는 다양한 가중치방식 중에서 Queen방식의 Z-score가 가장 높게 산출되어 인접성(contiguity) 기반 가중치가 설명력이 좋음을 확인할 수 있다. 구체적으로 O/D총통행발생·유입량의 Moran's I 값은 0.265543, 0.304381로 가장 큰 것으로 나타났으며, 유의수준(Pvalue) 0.05에서 '존의 통행발생·유입량



은 공간종속성이 없다.’는 귀무가설을 기각하여 공간상관성이 존재하는 것으로 나타나 OLS의 기본 가정인 오차항의 독립성 가정을 위배하므로 공간분석이 필요하다고 볼 수 있다.

- 귀무가설 H0 : 존의 통행발생 · 유입량은 공간적 자기상관이 없다.
- 대립가설 H1 : 존의 통행발생 · 유입량은 공간적 자기상관이 있다.

<표 5-24> O/D total 종속변수의 전역적 공간상관성 검증

가중치 방식	종속변수	Moran's I	ZScore	Pvalue
Rook	O/D P	0.304015	6.116231	0.000000
	O/D A	0.264715	5.35640	0.000000
<b>Queen</b>	<b>O/D P</b>	<b>0.304381</b>	<b>6.228419</b>	<b>0.000000</b>
	<b>O/D A</b>	<b>0.265543</b>	<b>5.464836</b>	<b>0.000000</b>
거리기반 역수 ( $d_{ij}^{-1}$ )	O/D P	0.132159	4.828630	0.000001
	O/D A	0.109526	4.054069	0.000005
거리기반 역수 제곱( $d_{ij}^{-2}$ )	O/D P	0.181720	5.220465	0.000000
	O/D A	0.156381	4.530917	0.000006

## (2) 필수적 통행발생 · 유입량 기준

O/D 필수적 통행발생·유입량과의 전역적 공간종속성 검정 결과 역시 앞선 총통행발생유입량 공간가중치 행렬방식과 마찬가지로 Queen방식의 Z-score가 가장 높게 산출되었다. 구체적으로 O/D 필수적통행발생 · 유입량의 Moran's I 값은 0.346674, 0.337366으로 가장 큰 것으로 나타났으며, 유의수준(Pvalue) 0.05에서 ‘존의 통행발생 · 유입량은 공간종속성이 없다.’는 귀무가설을 기각하



여 공간상관성이 존재하는 것으로 나타나 OLS의 기본 가정인 오차항의 독립성 가정을 위배하므로 공간분석이 필요하다고 볼 수 있다.

- 귀무가설  $H_0$  : 존의 통행발생 · 유입량은 공간적 자기상관이 없다.
- 대립가설  $H_1$  : 존의 통행발생 · 유입량은 공간적 자기상관이 있다.

<표 5-25> O/D Mand 종속변수의 전역적 공간상관성 검증

가중치 방식	종속변수	Moran's I	ZScore	Pvalue
Rook	O/D Mand P	0.341952	6.972412	0.000000
	O/D Mand A	0.332425	6.795582	0.000000
<b>Queen</b>	<b>O/D Mand P</b>	<b>0.346674</b>	<b>7.187849</b>	<b>0.000000</b>
	<b>O/D Mand A</b>	<b>0.337366</b>	<b>7.012638</b>	<b>0.000000</b>
거리기반 역수 ( $d_{ij}^{-1}$ )	O/D Mand P	0.185586	6.786671	0.000000
	O/D Mand A	0.186214	6.822175	0.000000
거리기반 역수 제곱( $d_{ij}^{-2}$ )	O/D Mand P	0.227666	6.594385	0.000000
	O/D Mand A	0.22774	6.610694	0.000000

### (3) 선택적 통행발생 · 유입량 기준

O/D 선택적 통행발생·유입량과의 전역적 공간종속성 검정 결과, 공간가중치 행렬방식 중 인접성 기반의 Rook방식의 Z-score가 가장 높게 산출되었다. 구체적으로 O/D 선택적통행발생 · 유입량의 Moran's I 값은 0.258510, 0.215575으로 가장 큰 것으로 나타났으며, 유의수준(Pvalue) 0.05에서 ‘존의 통행발생 · 유입량은 공간종속성이 없다.’는 귀무가설을 기각하여 공간상관성이 존재하는 것으로 나타나 OLS의 기본 가정인 오차항의 독립성 가정을 위배하므로 공간



분석이 필요하다고 볼 수 있다.

- 귀무가설  $H_0$  : 존의 통행발생 · 유입량은 공간적 자기상관이 없다.
- 대립가설  $H_1$  : 존의 통행발생 · 유입량은 공간적 자기상관이 있다.

<표 5-26> O/D Opt 종속변수의 전역적 공간상관성 검증

가중치 방식	종속변수	Moran' I	ZScore	Pvalue
<b>Rook</b>	<b>O/D Opt P</b>	<b>0.258510</b>	<b>5.238295</b>	<b>0.000000</b>
	<b>O/D Opt A</b>	<b>0.215575</b>	<b>4.414292</b>	<b>0.000010</b>
Queen	O/D Opt P	0.252313	5.203861	0.000000
	O/D Opt A	0.210564	4.388996	0.000011
거리기반 역수 ( $d_{ij}^{-1}$ )	O/D Opt P	0.082642	3.123149	0.001789
	O/D Opt A	0.055961	2.207241	0.027297
거리기반 역수 제곱( $d_{ij}^{-2}$ )	O/D Opt P	0.13114	3.834920	0.000126
	O/D Opt A	0.103746	3.091432	0.001992

### 5-3-2. CDR 적용 GWR 통행발생모형

#### (1) 총통행(Total Trip)발생 · 유입모형

##### 1) 모형 추정결과

O/D 총통행발생 · 유입량을 종속변수로 CDR 데이터를 적용하여 GWR을 실시하였다. 분석결과, O/D 총통행발생모형의 결정계수( $R^2$ )는 0.727이며, 수정된 결정계수( $adjR^2$ ) 0.710, 총통행유입모형의  $R^2$ 는 0.730, 수정된 결정계수( $adjR^2$ ) 0.712으로 나타났다.



<표 5-27> 모형요약 (CDR적용 GWR 총통행발생 · 유입모형)

구분	$R^2$	$adjR^2$	AICc
O/D total P	0.786	0.752	3068.37
O/D total A	0.794	0.757	3078.65

GWR분석으로 O/D 총통행발생 · 유입모형을 분석한 결과, GWR분석에서는 각 존별로 서로 다른 회귀계수가 도출되는 관계로 전체 존에 대한 평균 및 최대, 최소 결과치에 대해 일반적으로 아래 표5-28과 같이 표현된다.

<표 5-28> GWR분석 회귀계수 추정결과 (CDR적용 O/D 총통행발생모형)

Production		Model (GWR)				
Category	Variables	Mean	STD	Min	Max	Range
	(Constant)	-3652.953	27022.099	-60170.157	31032.888	91203.045
CDR	CDR Mand Morning	0.453	0.127	0.3	0.728	0.428
Land Use	log(IndusFloor Area)	2414.712	979.666	974.989	4243.07	3268.081
Social Factors	ServiceWorkers	3.424	0.836	1.79	4.51	2.72
	EleMiddHighN	2.469	0.382	1.877	3.354	1.476
Facility Factors	log(BusStop)	13679.525	2588.309	10651.736	19930.743	9279.006
$R^2$		0.786				
$adjR^2$		0.752				
AICc		3068.37				

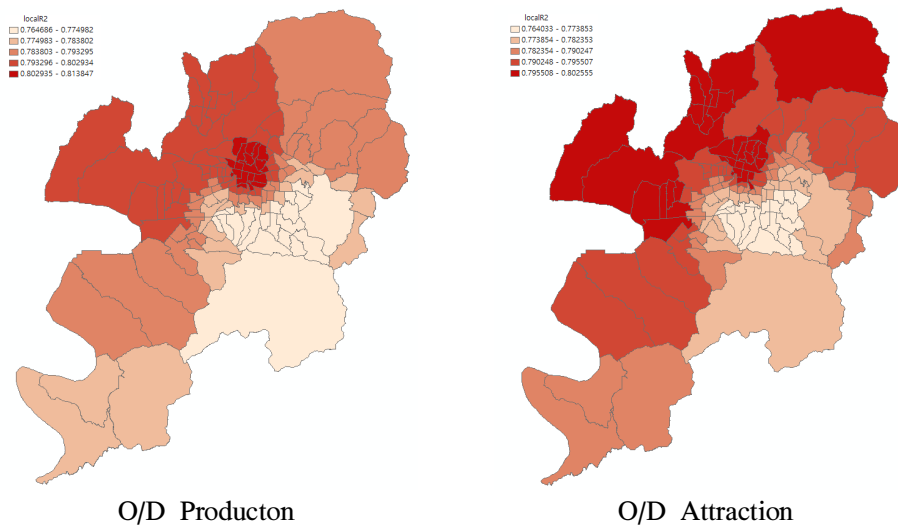


<표 5-29> GWR분석 회귀계수 추정결과 (CDR적용 O/D 총통행유입모형)

Attraction		Model (GWR)				
Category	Variables	Mean	STD	Min	Max	Range
	(Constant)	-4830.261	23021.698	-50217.460	25908.848	76126.308
CDR	CDR Mand Morning	0.528	0.127	0.383	0.798	0.415
Land Use	log(IndusFloor Area)	2653.823	1339.754	748.112	5072.860	4324.747
Social Factors	ServiceWorkers	3.478	0.980	1.578	4.581	3.003
	EleMiddHighN	2.593	0.397	1.960	3.492	1.532
Facility Factors	DOfficeF	6807.957	1506.780	5361.871	10238.785	4876.914
	log(BusStop)	12457.065	2376.268	9634.905	18060.187	8425.282
$R^2$		0.794				
$adjR^2$		0.757				
AICc		3078.65				

GWR은 각 존별로 국지적인 모형의 결정계수(Local  $R^2$ )를 산출하기 때문에 추정된 GWR모형의 국지적 결정계수의 분포를 살펴볼 필요가 있다. 아래 그림5-4와 같이 O/D 총통행발생모형의 경우, 대구시 북구 침산동, 칠성동, 성내동, 산격동 일원으로 대구서 서북부지역에서 높은 Local  $R^2$  값을 가지고, O/D총통행유입모형의 경우에는 북구 칠곡지구와 달성군 다사읍 지역 등으로 총통행발생모형보다 상대적으로 넓은 지역에 높은 Local  $R^2$  값을 가지는 것으로 나타났다. 반면 두 모형 모두 남구 대명동을 비롯한 남부지역과 수성구 지역에서는 상대적으로 낮은 Local  $R^2$  값을 나타내어 대구 도심 내부지역으로 갈수록 국지적 결정계수가 작아지는 형태를 보였다.





<그림 5-4> CDR total GWR모형의 Local  $R^2$  분포

## 2) 모형의 잔차 분석

O/D 총통행량 기준 CDR이 적용된 GWR 통행발생·유입모형 도출에 따라 종속변수의 등분산 가정을 증명하기 위하여 표준화잔차(Std.Resid)에 대하여 전역적 공간자기상관을 분석해보았다. 그 결과, 총통행발생모형의 Moran's I 값은 0.107755, 유의확률은 0.021572로 나타나 잔차가 공간적으로 무작위하게 퍼져있다는 귀무가설을 기각하여 잔차의 공간자기상관이 존재함을 확인하였다.

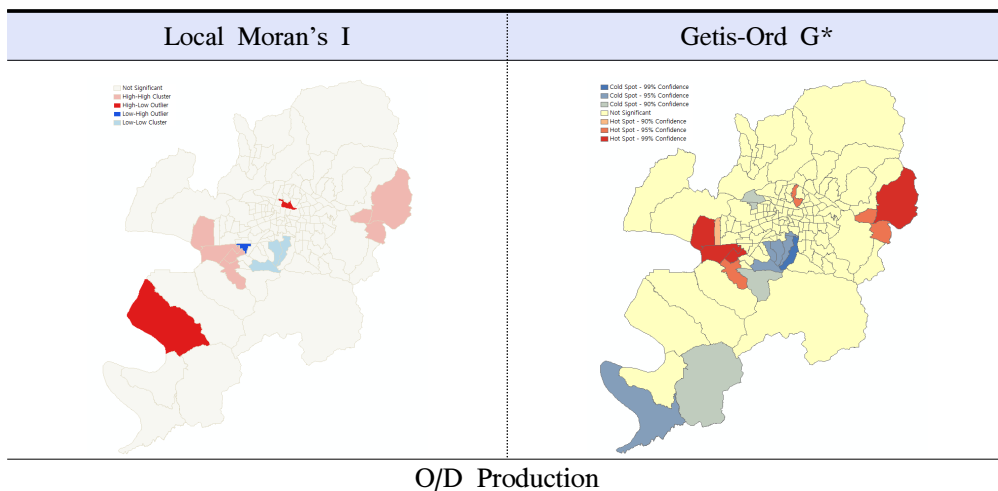
반면, 총통행유입모형의 Moran's I 값은 0.05544, 유의확률은 0.209958로 나타나 잔차가 공간적으로 무작위하게 퍼져있다는 귀무가설을 기각하지 아니하여 잔차의 공간자기상관이 존재하지 않는 것을 확인하였다.

- 귀무가설  $H_0$  : 모형의 표준화잔차는 공간적 자기상관이 없다.
- 대립가설  $H_1$  : 모형의 표준화잔차는 공간적 자기상관이 있다.

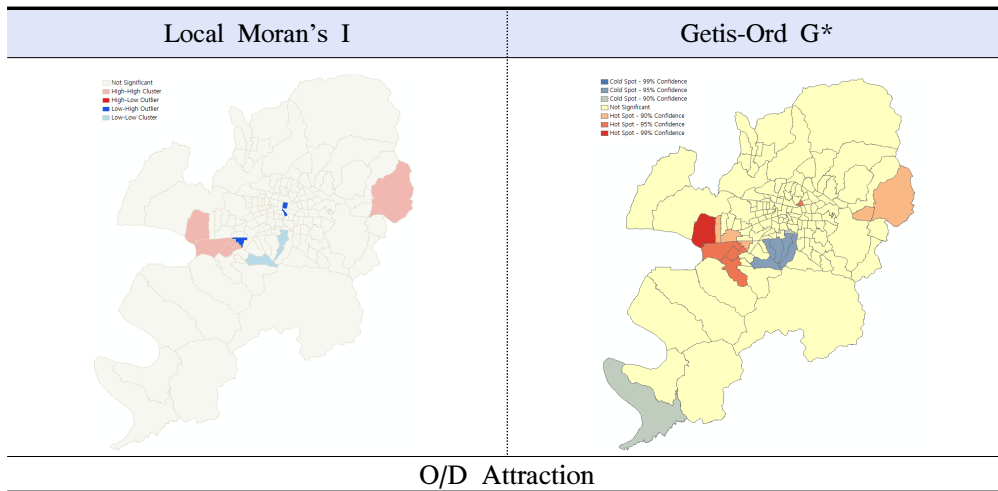
<표 5-30> O/D total GWR기준모형의 오차의 공간자기상관 검증결과

구분	Moran's I	ZScore	Pvalue
O/D total P	0.107755	2.297821	0.021572
O/D total A	0.055440	1.253680	0.209958

국지적 공간상관성 탐색을 위하여 Local Moran's I와 Getis-Ord G\* 방법을 통해 공간적 군집을 확인해본 결과, OLS 총통행발생 · 유입모형과 비교하여 통행발생모형에서는 잔차의 공간적 분포가 유사한 것으로 나타났으나 통행유입모형에서는 상대적으로 잔차의 국지적 공간상관성 군집이 줄어든 것으로 나타나 모형이 개선된 것으로 나타났다.







<그림 5-5> CDR total GWR기준 모형의 국지적인 공간자기상관 분석결과

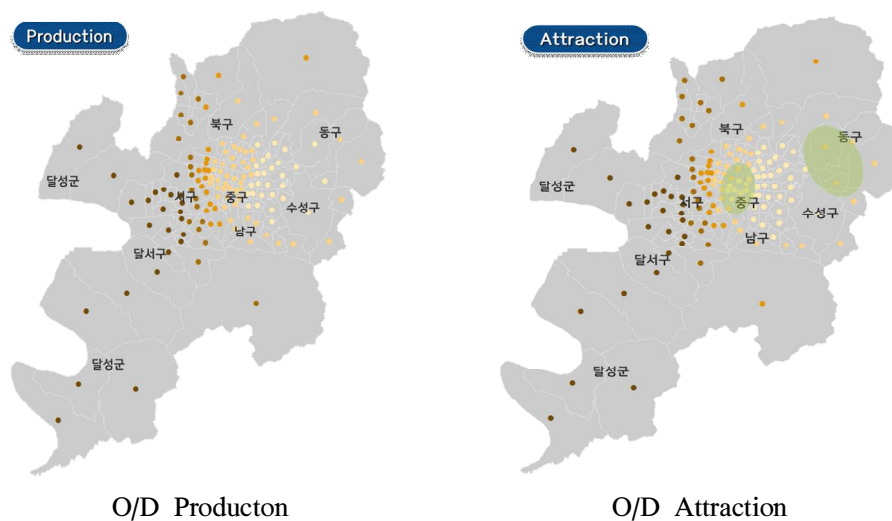
### 3) CDR변수의 공간적 해석

앞서 언급했듯이 GWR분석은 전체 연구대상지역의 하나의 회귀식을 나타내는 일반적 OLS와 달리 분석대상 존별 각각의 다른 계수값들을 나타냄으로써 변수들의 국지적인 영향력을 파악할 수 있는 장점이 있다. 이에 유의미한 독립변수의 존별 계수값 분포에 따른 공간적 분포를 살펴보고 이를 해석하는 과정이 중요하다고 볼 수 있다. 특히 본 연구에서 통행발생모형에 주요한 설명변수로 적용된 CDR의 유의미한 공간적 분포를 살펴보는 것은 실제 사람들의 통행행태 추이를 추정하고 향후 빅 데이터 연구의 활용적 측면의 시사점을 제시할 수 있다는 점에서 중요한 과정이라고 판단된다. 이에 본 연구에서는 각 통행목적별 통행발생·유입모형별로 적용된 CDR 변수의 유의미한 공간적 분포를 살펴보고, 이에 대한 공간적 해석을 통해 실제 사람들의 통행이 통행발생에 어떠한 영향을 미치는 지에 대해 세부적으로 살펴보고자 한다.

O/D총통행발생·유입모형에서 CDR 평일 오전첨두시 통신량이 유의미한



영향을 미치는 공간적 분포를 살펴본 결과, 달서구와 달성군 방향의 서남부 축으로 CDR 평일 오전첨두시 통행량 회귀계수의 값이 높게 형성되어 있는 분포를 보여 이 지역에서 통행발생과 유입에 영향을 미치는 지역적 통행 활성화 도시구조가 형성된 것을 알 수 있다. 통행유입모형에 있어서도 통행발생과 유사한 CDR의 공간적 분포를 나타냈으나 도심 내부의 중구 일부지역에서 상대적으로 낮은 회귀계수를 나타냈고 동구 안심동 일원에서 상대적으로 약간 높은 회귀계수의 차이를 보였다. 이러한 결과로부터 총통행발생·유입 모형에 있어 대구광역시 달서구, 달성군 서구, 북구 일부지역의 서남부축 지역이 전체 도시공간 구조 상에서 통행발생·유입에 사람들의 통행 활동성이 높은 축을 형성하고 있는 것으로 추정되어 해당 지역의 도시 인프라 구축이나 대중교통 서비스의 공급 확대 등의 정책 시행이 해당 지역을 중심으로 필요할 것으로 판단된다.



<그림 5-6> O/D total GWR모형의 CDR Mand Morning 분포



## (2) 필수적통행(Mandatory Trip)발생 · 유입모형

### 1) 모형 추정결과

O/D 필수적통행발생 · 유입량을 종속변수로 CDR 데이터를 적용하여 GWR을 실시하였다. 분석결과, O/D 필수적통행발생모형의 결정계수( $R^2$ )는 0.825이며, 수정된 결정계수( $adjR^2$ ) 0.802, 필수적통행유입모형의  $R^2$ 는 0.836, 수정된 결정계수( $adjR^2$ ) 0.808로 나타났다.

<표 5-31> 모형요약 (CDR적용 GWR 필수적통행발생 · 유입모형)

구분	$R^2$	$adjR^2$	AICc
O/D Mand P	0.825	0.802	2828.28
O/D Mand A	0.836	0.808	2829.90

GWR분석으로 O/D 필수적통행발생 · 유입모형을 분석한 결과는 아래 표 5-32와 같다.

<표 5-32> GWR분석 회귀계수 추정결과 (CDR적용 O/D 필수적통행발생모형)

Production		Model (GWR)				
Category	Variables	Mean	STD	Min	Max	Range
	(Constant)	6332.878	1509.367	5061.449	9306.181	4244.732
CDR	CDR Mand Morning	0.202	0.053	0.115	0.309	0.193
Land Use	log(IndusFloor Area)	2110.844	204.308	1717.663	2517.514	799.851
	ResidFloorArea	0.001	0.000	0.001	0.001	0.000
Social	UniStudN	0.922	0.078	0.767	1.038	0.271



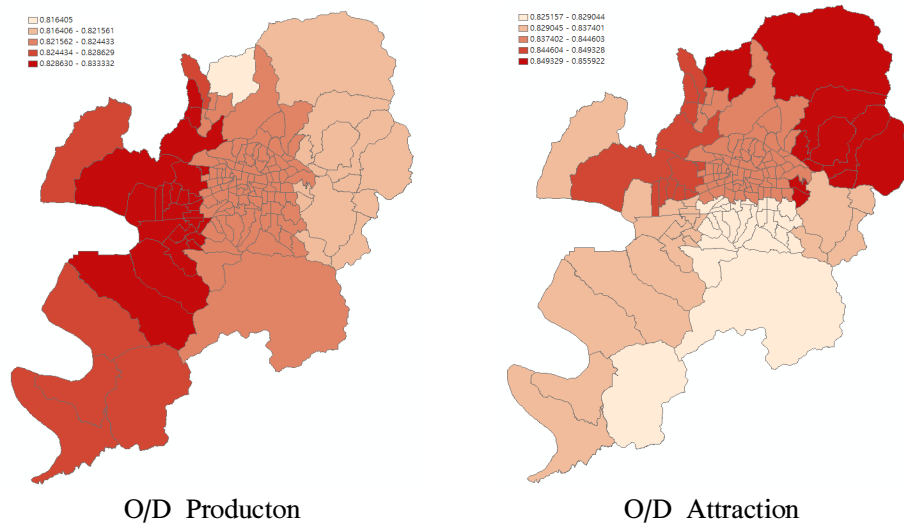
Production		Model (GWR)				
Category	Variables	Mean	STD	Min	Max	Range
Factors	EleMiddHighN	1.803	0.106	1.521	1.983	0.461
Facility Factors	BusStop	138.975	19.386	74.289	176.070	101.781
$R^2$		0.825				
$adjR^2$		0.802				
AICc		2828.28				

<표 5-33> GWR분석 회귀계수 추정결과 (CDR적용 O/D 필수적통행유입모형)

Attraction		Model (GWR)				
Category	Variables	Mean	STD	Min	Max	Range
	(Constant)	6218.772	1247.095	3925.862	8721.642	4795.780
CDR	CDR Mand Morning	0.231	0.039	0.152	0.306	0.154
Land Use	log(IndusFloor Area)	2010.665	338.819	1154.441	2625.004	1470.562
Social Factors	age20 29	2.948	0.210	2.620	3.688	1.068
	UniStudN	0.782	0.099	0.585	0.950	0.364
	EleMiddHighN	1.218	0.144	0.836	1.482	0.646
Facility Factors	DOfficFac	2618.578	480.239	1948.669	3822.601	1873.931
	BusStop	124.688	20.868	63.798	162.339	98.541
$R^2$		0.836				
$adjR^2$		0.808				
AICc		2829.90				

그림5-7과 같이 O/D 필수적통행발생모형의 경우, 대구시 달서구와 달성군 다사읍, 북구 태전동, 국우동 일원으로 대구 서북부지역에서 높은 Local  $R^2$  값을 가지고, O/D 필수적통행유입모형의 경우에는 북구 칠곡지구와 동구 공산동, 안심동 일원으로 대구 북부지역에 높은 Local  $R^2$  값을 가지는 것으로 나타났다.





<그림 5-7> CDR Mand GWR모형의 Local R<sup>2</sup> 분포

## 2) 모형의 잔차 분석

필수적통행발생 · 유입모형의 Moran's I값은 각각 0.036938, 0.015962, 유의확률은 0.369651, 0.636269로 나타나 잔차가 공간적으로 무작위하게 퍼져있다는 귀무가설을 기각하지 아니하여 잔차의 공간자기상관이 존재하지 않는 것을 확인하였다.

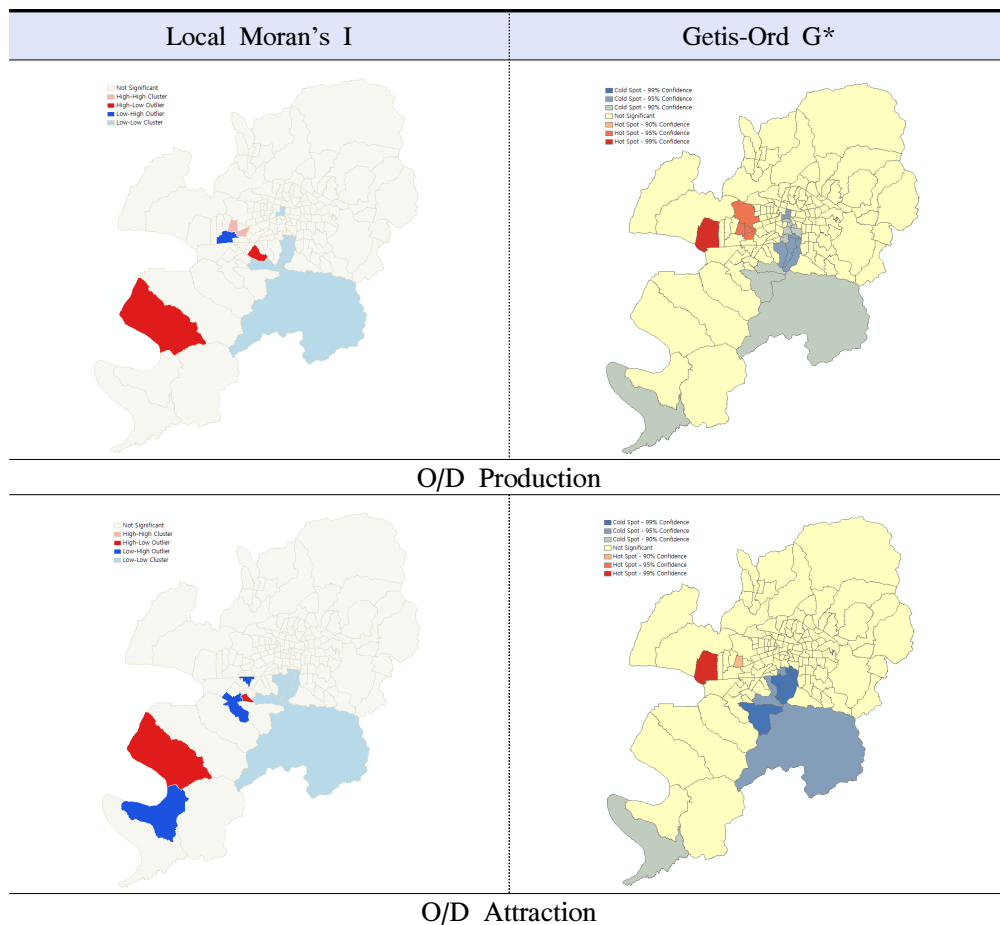
- 귀무가설 H0 : 모형의 표준화잔차는 공간적 자기상관이 없다.
- 대립가설 H1 : 모형의 표준화잔차는 공간적 자기상관이 있다.

<표 5-34> O/D Mand GWR기준모형의 오차의 공간자기상관 검증결과

구분	Moran's I	ZScore	Pvalue
O/D Mand P	0.036938	0.897127	0.369651
O/D Mand A	0.015962	0.472922	0.636269



Local Moran's I와 Getis-Ord G\* 분석 결과, OLS 필수적통행발생 · 유입모형과 비교하여 통행발생모형에서는 잔차의 공간적 분포가 유사한 것으로 나타났으나 통행유입모형에서는 상대적으로 잔차의 국지적 공간상관성 군집이 줄어든 것으로 나타나 모형이 개선된 것으로 나타났다.

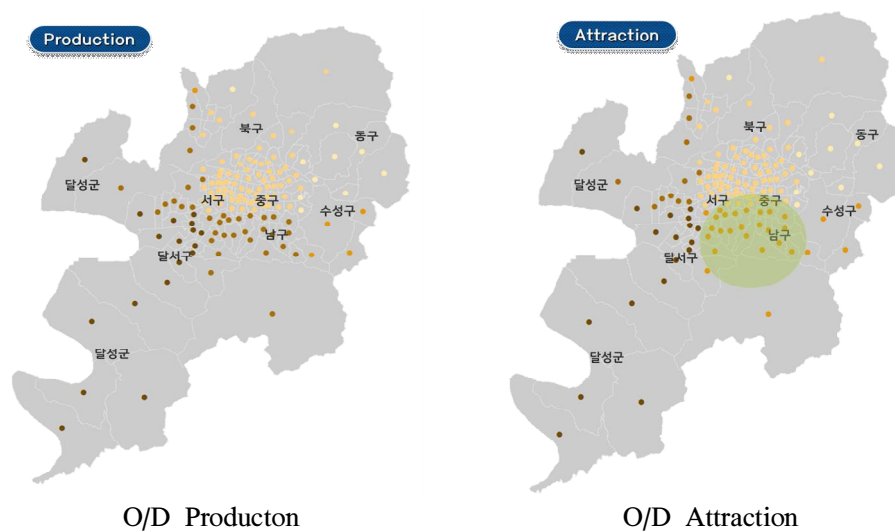


<그림 5-8> CDR Mand GWR기준 모형의 국지적인 공간자기상관 분석결과



### 3) CDR변수의 공간적 해석

O/D 필수적통행발생·유입모형에서 CDR 평일 오전첨두시 통신량이 유의미한 영향을 미치는 공간적 분포를 살펴본 결과, 달서구와 달성군 방향과 남부, 수성구 일부지역의 서남부축으로 CDR 평일 오전첨두시 통행량 회귀계수의 값이 높게 형성되어 있는 분포를 보여 이 지역에서 통행발생과 유입에 영향을 미치는 지역적 통행 활성화 도시구조가 형성된 것을 알 수 있다. 통행유입모형에 있어서도 통행발생과 유사한 CDR의 공간적 분포를 나타냈으나 남구와 수성구 일부지역에서 상대적으로 낮은 회귀계수를 나타냈고 동구 안심동, 도평동 일원에서는 전체적으로 낮은 회귀계수의 차이를 보였다. 이러한 결과로부터 필수적통행발생·유입모형은 총통행발생·유입모형과 비교하여 대구광역시 서남부축이 공통적으로 도시 활성화 지역으로 나타났고 남구와 수성구 일부지역이 필수적 통행과 관련하여 보다 넓은 지역에서 CDR 통행활동성이 높은 지역으로 추정되었다.



<그림 5-9> O/D Mand GWR모형의 CDR Mand Morning 분포



### (3) 선택적통행(Optional Trip)발생 · 유입모형

#### 1) 모형 추정결과

O/D 선택적통행발생 · 유입량을 종속변수로 CDR 데이터를 적용하여 GWR을 실시하였다. 선택적 통행발생 · 유입모형에 채택된 CDR은 주말 오전첨두시 통행량으로 총통행 및 필수적통행목적모형과는 다른 CDR 변수가 채택되었다. 분석결과, O/D 선택적통행발생모형의 결정계수( $R^2$ )는 0.739이며, 수정된 결정계수( $adjR^2$ ) 0.715으로 선택적통행유입모형의  $R^2$ 는 0.703, 수정된 결정계수( $adjR^2$ ) 0.671로 나타났다.

<표 5-35> 모형요약 (CDR적용 GWR 필수적통행발생 · 유입모형)

구분	$R^2$	$adjR^2$	AICc
O/D Mand P	0.739	0.715	2956.02
O/D Mand A	0.703	0.671	2997.92

GWR분석으로 O/D 선택적통행발생 · 유입모형을 분석한 결과는 아래 표 5-36과 같다.

<표 5-36> GWR분석 회귀계수 추정결과 (CDR적용 O/D 선택적통행발생모형)

Production		Model (GWR)				
Category	Variables	Mean	STD	Min	Max	Range
	(Constant)	2324.594	457.272	1735.770	3028.084	1292.314
CDR	CDR Opt Morning	0.108	0.024	0.061	0.146	0.085
Land Use	CommerFloorArea	0.001	0.000	0.001	0.001	0.000





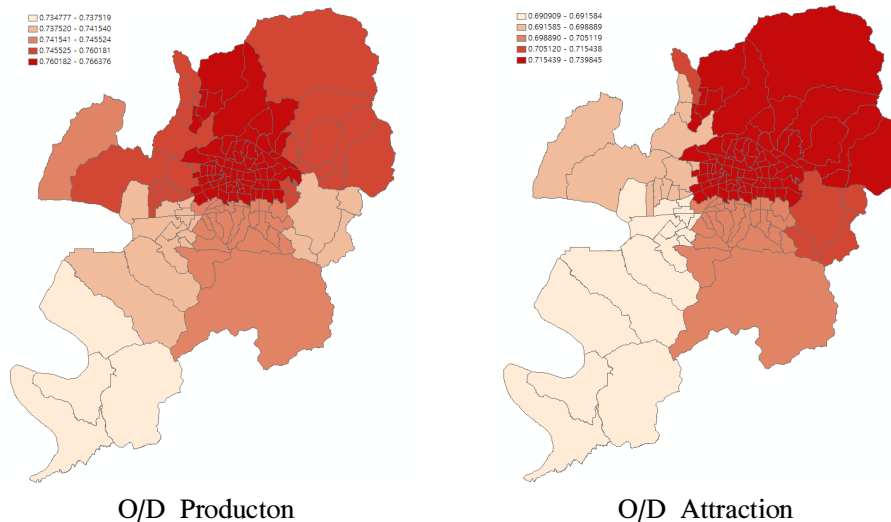
Production		Model (GWR)				
Category	Variables	Mean	STD	Min	Max	Range
	ResidFloorArea	0.001	0.000	0.001	0.001	0.000
	log(IndusFloorArea)	1946.143	575.385	944.436	2767.052	1822.616
Social Factors	ServiceWorkers	2.357	0.405	1.688	3.126	1.439
Facility Factors	DSalesFac	3758.505	178.495	3622.150	4177.079	554.929
$R^2$		0.739				
$adjR^2$		0.715				
AICc		2956.02				

<표 5-37> GWR분석 회귀계수 추정결과 (CDR적용 O/D 선택적통행유입모형)

Attraction		Model (GWR)				
Category	Variables	Mean	STD	Min	Max	Range
	(Constant)	531.913	616.970	-253.530	1540.618	1794.148
CDR	CDR Opt Morning	0.135	0.023	0.089	0.173	0.084
Land Use	CommerFloorArea	0.001	0.000	0.001	0.001	0.000
	log(IndusFloorArea)	2220.231	749.075	936.624	3321.919	2385.294
	ResidFloorArea	0.001	0.000	0.001	0.001	0.000
Social Factors	ServiceWorkers	2.526	0.437	1.776	3.337	1.561
Facility Factors	DOfficFac	6071.301	595.207	5086.006	7407.061	2321.055
	DSalesFac	3282.274	210.812	3116.940	3777.356	660.416
$R^2$		0.703				
$adjR^2$		0.671				
AICc		2997.92				



그림5-10과 같이 O/D 선택적통행발생모형의 경우, 대구시 북구 일원으로 높은 Local  $R^2$  값을 가지고, O/D 선택적통행유입모형의 경우에도 유사한 지역에서 높은 Local  $R^2$  값을 가지는 것으로 나타났다.



<그림 5-10> CDR Opt GWR모형의 Local  $R^2$  분포

## 2) 모형의 잔차 분석

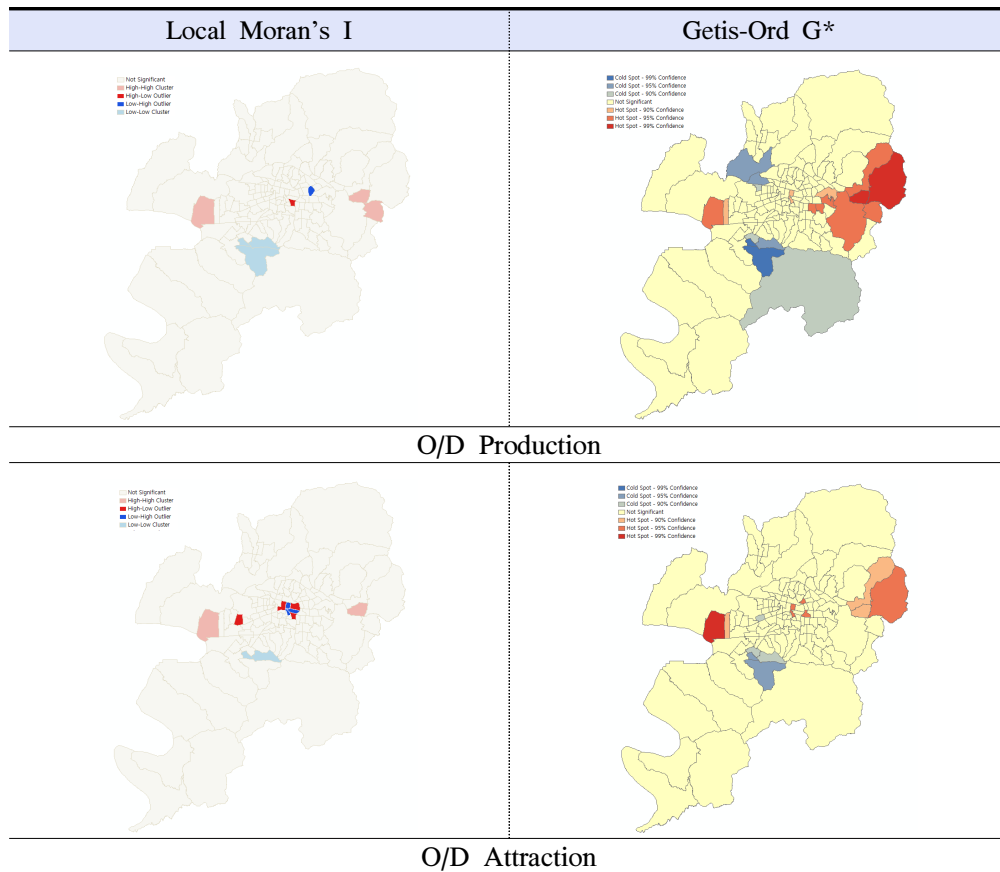
선택적통행발생 · 유입모형의 Moran's I값은 각각 0.034683, -0.030420, 유의 확률은 0.410864, 0.645490으로 나타나 잔차가 공간적일 무작위하게 퍼져있다는 귀무가설을 기각하지 아니하여 잔차의 공간자기상관이 존재하지 않는 것을 확인하였다.

- 귀무가설  $H_0$  : 모형의 표준화잔차는 공간적 자기상관이 없다.
- 대립가설  $H_1$  : 모형의 표준화잔차는 공간적 자기상관이 있다.

<표 5-38> O/D Opt GWR기준모형의 오차의 공간자기상관 검증결과

구분	Moran's I	ZScore	Pvalue
O/D Opt P	0.034683	0.822375	0.410864
O/D Opt A	-0.030420	-0.460036	0.645490

Local Moran's I와 Getis-Ord G\* 분석 결과, OLS 선택적통행발생·유입모형과 비교하여 통행발생모형에서는 잔차의 공간적 분포가 유사한 것으로 나타났다. 통행유입모형에서는 상대적으로 잔차의 국지적 공간상관성 군집이 줄어든 것으로 나타나 모형이 개선된 것으로 나타났다.

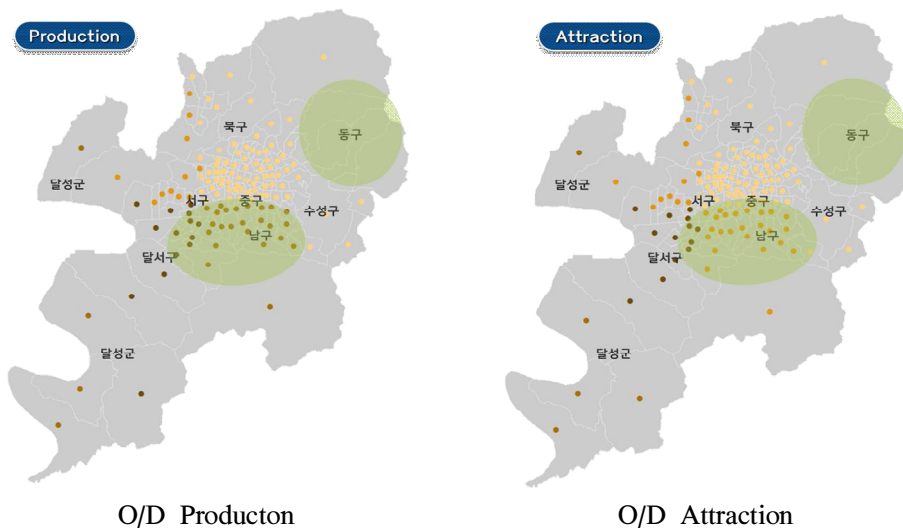


<그림 5-11> CDR Opt GWR기준 모형의 국지적인 공간자기상관 분석결과



### 3) CDR변수의 공간적 해석

O/D 선택적통행발생 · 유입모형에서 CDR 주말 오전첨두시 통신량이 유의미한 영향을 미치는 공간적 분포를 살펴본 결과, 필수적 통행발생에 높은 공간적 영향을 보인 달성군, 달서구, 남구, 수성구 일부지역으로 CDR 주말 오전첨두시 회귀계수의 값이 유사하게 형성되어 있는 분포를 보였으나, 동구 지역의 도평동, 해안동, 안심동 등 다소 인구밀도가 낮고 도시 개발도가 낮은 지역의 경우, CDR 주말 오전첨두시 통행량이 선택적 통행발생 · 유입에 유의하지 않은 관계를 보이는 것으로 나타났다.



<그림 5-12> O/D Opt GWR모형의 CDR Opt Morning 분포

## 5-4. 모형의 검증

### 5-4-1. 모형 비교분석결과

본 절에서는 지금까지 O/D 총통행발생·유입량을 비롯한 필수적·선택적 통행량을 종속변수로 한 OLS를 기준모형으로 하여 GWR의 결과와의 비교분석을 통해 모형의 추정력 개선에 대해 알아보고자 한다. 이에 모형별 설명력을 나타는 결정계수( $R^2$ ,  $adjR^2$ ) 및 모형의 적합도(AICc) 결과치 대한 비교와 오차의 공간자기상관 검증을 위한 Moran's I 결과 및 모형의 관측치와 예측치 간의 비교를 위한 오차의 평균치에 대한 검정(RMSE, SRMSE)결과를 살펴보았다.

#### (1) 모형추정력 개선결과 요약

통행발생·유입량 기준 OLS분석법에 따른 통행목적별 통행발생모형을 기준모형으로 하여 공간성을 감안한 GWR분석방법에 따른 통행발생모형과의 모형의 설명력( $R^2$ ,  $adjR^2$ ) 및 모형적합도(AICc) 통계량 분석결과를 비교해 본 결과는 아래 표5-39와 같다. 분석결과, 전반적으로 CDR이 포함되었을 경우 GWR분석을 이용한 모형식의 설명력이 모두 높아지는 것으로 나타났으며 특히 총통행유입모형에서는 CDR이 미적용된 OLS모형과 비교해 CDR이 포함된 GWR 통행유입모형의 경우  $adjR^2$ 가 0.100정도의 차이를 보여 설명력이 가장 많이 개선된 결과를 나타내었다. 일반적으로 최적 모형을 추정함에 있어 모형의 설명력( $R^2$ )을 판단기준으로 사용하지만 이 값은 설명변수의 수가 많아질수록 그 값이 증가되는 경향을 보이기 때문에 모형의 설명력과 함께 절약성을 고려해야 한다. 모형의 절약성은 복잡한 모형보다는 간결한 모형이



선호되어야 한다는 것으로 추정계수의 수가 많은 것에 대해 벌칙을 가함으로써 고려될 수 있다. 이러한 두 판단기준을 함께 고려한 선택기준이 바로 정보기준 접근법으로 Akaike(1974)에 의하여 개발된 Akaike Information Criterion(AIC)이다.

모형의 적합도를 나타내는 AICc 통계량의 경우 일반적으로 4이상의 차이가 나면 모형이 개선되었다고 볼 수 있는데 CDR이 적용된 GWR 필수적 통행유입모형에서 CDR 미적용 OLS기준모형보다 60.9이상 감소된 것으로 나타나 모형의 적합도 측면에서 가장 많은 개선이 이루어진 것으로 나타났다.

<표 5-39> 모형간 설명력( $R^2$ ,  $adjR^2$ ) 및 적합도(AICc) 비교

구분		총통행		필수적통행		선택적통행	
		발생	유입	발생	유입	발생	유입
CDR 미포함 OLS	$R^2$	0.693	0.672	0.738	0.732	0.682	0.636
	$adjR^2$	0.679	0.657	0.724	0.713	0.670	0.619
	AICc	3113.35	3131.26	2881.72	2890.88	2976.15	3019.01
CDR 포함 OLS	$R^2$	0.727	0.730	0.781	0.791	0.701	0.657
	$adjR^2$	0.710	0.715	0.766	0.774	0.685	0.636
	AICc	3085.74	3097.62	2848.99	2849.80	2968.82	3010.43
CDR 포함 GWR	$R^2$	0.785	0.794	0.825	0.836	0.739	0.703
	$adjR^2$	0.752	0.757	0.802	0.808	0.715	0.671
	AICc	3068.37	3078.65	2828.28	2829.90	2956.02	2997.92

## (2) 잔차의 공간상관성 (Moran's I)

OLS분석의 통계적 기본 가정인 오차의 등분산 가정에 대해 이를 증명하기



위한 방법으로 각 통행발생모형의 표준화잔차에 대한 전역적 공간상관성 (Moran's I) 분석을 시행하였다. 이에 공간자기상관분석을 실시한 결과는 아래 표5-40과 같다. 분석결과, 전반적으로 GWR을 통행발생유입모형에서는 오차의 Moran's I의 값이 OLS 통행발생모형에 비해 0에 가깝고 Pvalue 역시 유의수준 0.05를 넘어 귀무가설을 기각하지 아니하여 오차의 공간자기상관문제가 해결되어 오차의 분포가 무작위하게 퍼져 있다는 사실을 알 수 있다. 다만, 필수적 통행 및 선택적 통행과 관련하여 OLS와 GWR모형 모두 유의수준 0.05에서 귀무가설이 기각되지 아니하여 잔차가 무작위하게 분포된 결과를 나타내었으나, 상대적으로 GWR모형에서 잔차의 무작위분포에 대한 통계량이 보다 높은 것으로 나타났다.

<표 5-40> 모형간 오차의 공간상관성(Moran's I) 비교

구분		총통행		필수적통행		선택적통행	
		발생	유입	발생	유입	발생	유입
CDR 포함 OLS	MI	0.152801	0.108184	0.053402	0.061481	0.074529	0.002362
	ZScore	3.194219	2.306320	1.222969	1.391294	1.605898	0.190314
	PValue	0.001402	0.020964	0.221342	0.164136	0.108296	0.849063
CDR 포함 GWR	MI	0.107755	0.055440	0.036938	0.015962	0.034683	-0.030420
	ZScore	2.297821	1.253680	0.897127	0.472922	0.822375	-0.460036
	PValue	0.021572	0.209958	0.369651	0.636269	0.410864	0.645490

## 5-4-2. 데이터 오차검증

도출된 모형의 오차검증을 위하여 본 연구에서는 RMSE와 SRMSE를 산출



하여 비교해보고자 한다. 모형의 관측치와 예측치를 대상으로 오차를 검증해 보는 척도로서 RMSE(Root Mean Squared Error)와 SRMSE(Standardized Root Mean Square Error)는 아래 식과 같다.

$$RMSE = \sqrt{\frac{1}{N} \sum_i^N |e_i^2|} \quad \dots\dots\dots (식5.13)$$

$$SRMSE = \frac{\sqrt{\frac{1}{N} \sum_i^N |e_i^2|}}{Y_{obs,i}} \quad \dots\dots\dots (식5.14)$$

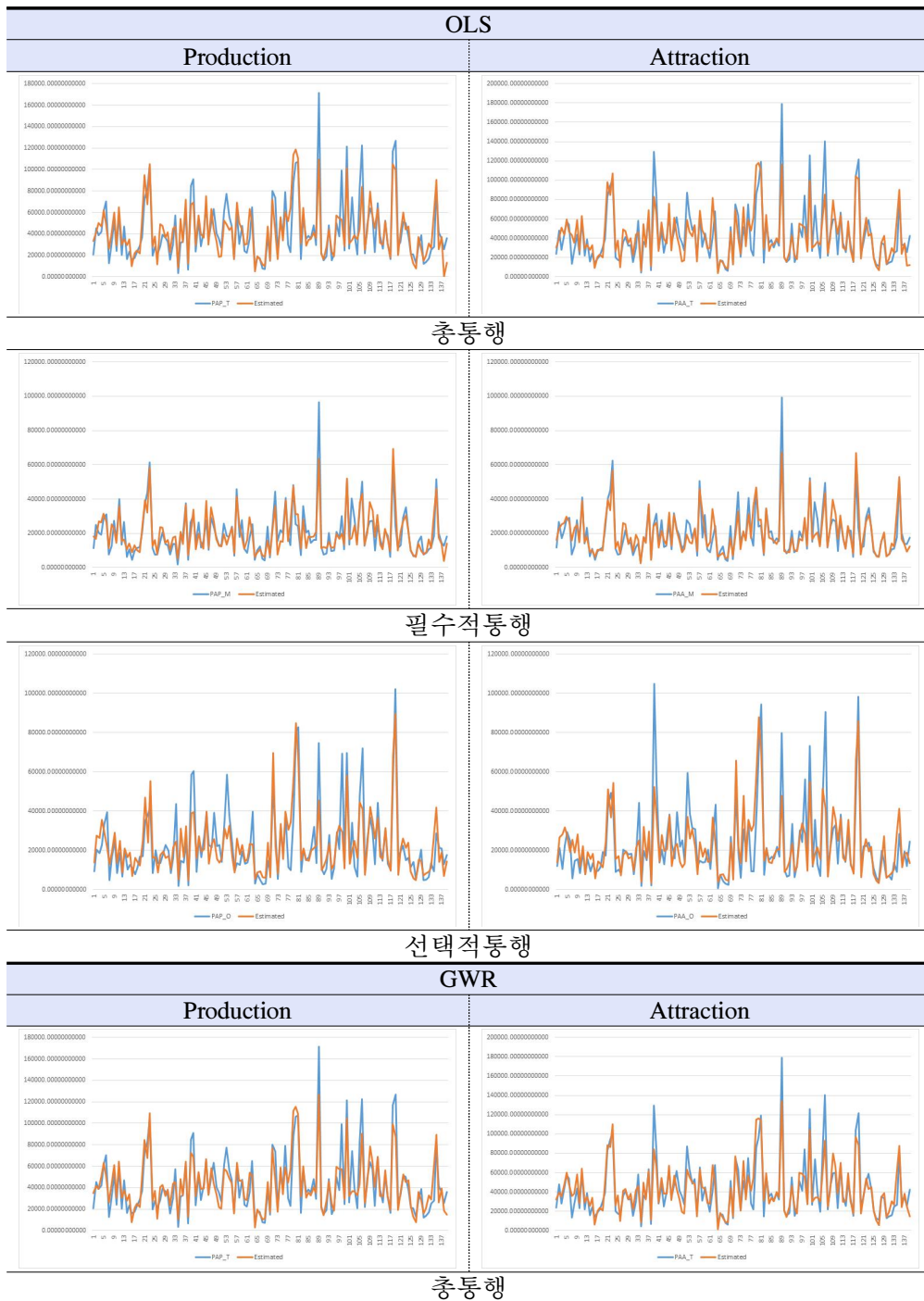
이때, N은 관측수이고  $Y_{obs}$ 는 관측치,  $Y_{model}$ 은 모형에 따른 예측치,  $e$ 는 관측치와 예측치의 차이를 의미한다. 결국 이 검증치들의 값이 작을수록 모형의 신뢰성을 증가된다고 볼 수 있다. SRMSE의 경우 일반적으로 0.08이하인 경우 매우 좋은 모형으로 간주되며, 0에 가까울수록 완벽한 모형으로 해석된다(정우현, 2011). 분석결과, GWR을 이용한 모형식이 전반적으로 오차검증에서 개선된 결과를 보여 OLS 보다 우수한 모형임을 확인할 수 있다.

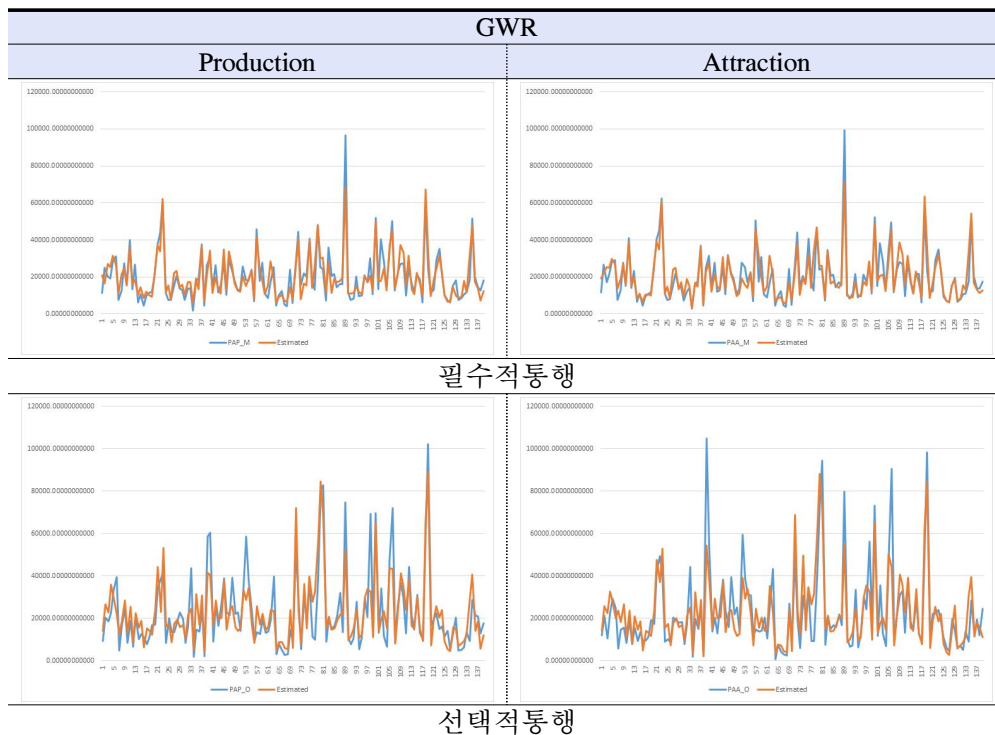
<표 5-41> 데이터 검증(RMSE, SRMSE)

구분		총통행		필수적통행		선택적통행	
		발생	유입	발생	유입	발생	유입
CDR 포함 OLS	RMSE	10823.7	10687.7	4553.5	3983.0	7045.0	7658.4
	SRMSE	0.303074	0.289294	0.274276	0.222818	0.405459	7172.6
CDR 포함 GWR	RMSE	9605.6	9429.1	4092.4	3652.2	6636.2	0.433712
	SRMSE	0.283070	0.265164	0.256025	0.211117	0.380645	0.400877









<그림 5-13> 모형별 관측치와 예측치의 비교 그래프

### 5-4-3. 소결

본 장에서는 기존 통행발생모형에서 주로 사용하던 OLS를 기준모형으로 하여 통신 빅 데이터를 모형의 추정력 개선을 위한 변수로 활용하기 위해 요일별·시간대별 통신량 데이터(CDR)를 통행발생모형의 변수로 적용시킨 대안별 통행발생모형을 추정하여 모형의 추정력 개선에 대해 살펴보았다. 또한 통행의 공간적 상관성을 감안한 GWR을 이용한 공간성을 감안한 통행발생모형을 도출하여 비교·분석하였다. 분석 결과, CDR을 미적용한 OLS 통행목적별 통행발생모형을 기준으로 모든 모형에서 CDR 변수를 설명변수로 적용하였을 때 상대적으로 높은 설명력( $adjR^2$ )을 가진 통행발생모형이 도출되었다.



구체적으로 O/D 총통행발생·유입과 필수적통행발생·유입모형의 경우, CDR 평일 오전침두시 통행량 데이터가 주요한 설명변수로 채택되어 모형의 설명력을 높이는 결과를 나타내었고 특히, O/D 선택적통행발생·유입모형의 경우에는 CDR 주말 오전침두시 통행량 데이터가 주요한 설명변수로 적용되었다. 일반적으로 선택적 통행목적 성격의 통행발생·유입모형의 경우 기존 연구에서는 주로 사회경제지표와 관련된 특정한 변수-예를 들어, 3차산업종사자수-로서 다른 통행목적에 비해 비교적 낮은 상관관계의 변수를 바탕으로 추정되었는데, 본 연구의 결과에 따라 향후 선택적 통행과 관련한 통행발생모형 도출시 CDR의 주말 데이터가 유의미하게 활용될 수 있을 것으로 기대된다.

또한 종속변수인 통행발생·유입량의 공간상관성을 감안한 GWR 통행발생모형 구축 결과, CDR이 포함된 모든 통행발생모형에서 OLS보다 높은 설명력( $adjR^2$ )을 나타내었고 모형의 절약성을 나타내는 AICc값도 4이상의 차이를 보여 줄어든 것으로 나타나 모형의 추정력이 개선된 결과를 보였다. OLS모형의 잔차의 무작위 분포 가정 문제 검증을 위한 표준화잔차의 공간성 검증 결과에서도 GWR 통행발생모형에서 상대적으로 잔차가 0에 가깝고 잔차가 무작위하게 분포되어 있다는 귀무가설에 기각되지 않는 결과를 보여 모형이 개선된 것으로 나타났다. 마지막으로 각 통행목적별 추정된 OLS 및 GWR 통행발생모형의 결과로 나타난 관측치와 예측치간의 오차 검증 결과 GWR 분석에서 RMSE, SRMSE 통계량이 모두 줄어든 것으로 나타나 GWR 통행발생모형이 상대적으로 우수한 모형임을 확인하였다.





## 제6장 결론 및 연구의 한계점

### 6-1. 결론 및 정책적 시사점

본 연구는 기존 전통적 4단계 수요예측방법 중 통행발생단계의 추정력 개선을 위하여 최근 많이 논의되고 있는 빅 데이터 가운데 모바일 통신량 자료(Call Detail Record: CDR)을 이용하여 모형을 개선하고 기존 선형회귀분석에 따른 통행의 공간적 상관성을 미반영 문제를 고려한 공간계량분석 기법 중 지리적 가중회귀분석(GWR)을 이용한 통행발생모형을 추정하여 모형의 개선에 따른 결과를 비교·분석하였다.

우선 통행발생모형 추정의 개선을 위한 빅 데이터의 활용적 관점에서 각 통행목적별로 CDR을 적용해본 결과, 모든 통행목적에서 CDR이 주요한 변수로 채택되었으며, 특히 총통행 및 필수적통행발생·유입모형에서는 CDR 평일 오전첨두시 통행량이 주요한 변수로 채택되었으나 선택적통행발생·유입모형에서는 CDR 주말 오전첨두시 통행량이 유의미한 변수로 선정되어 모형의 설명력과 적합도가 개선된 결과를 나타냈다. 일반적으로 비가정기반 쇼핑·여가·기타 통행과 같은 비고정적이고 불규칙한 통행성격을 가지는 선택적 통행발생·유입모형과 관련해서는 기존 사회경제지표의 3차산업종사자수나 취업자수 등의 변수를 주로 이용하여 상대적으로 낮은 설명력의 모형을 도출하였는데, 향후에는 본 연구의 결과로 나타난 CDR 주말 오전첨두시 통행량과 같은 데이터를 이용하여 선택적 목적통행발생모형의 유의미한 변수로의 활용이 가능할 것으로 판단된다.

또한 기존 선형회귀분석에서 고려하지 못하는 통행의 공간적 상호작용을



감안하여 GWR분석을 통한 통행발생모형 구축결과, 모든 통행목적에서 기존 OLS 통행발생모형보다 설명력( $adjR^2$ )과 적합도(AICc)가 개선되는 것으로 나타났다으며, 특히 통행유입모형에서 그 값이 크게 개선된 결과를 보였다. 더불어, OLS와 GWR기법을 통한 통행발생모형으로 도출된 관측치와 예측치 간의 데이터 오차검증(RMSE, SRMSE) 결과에서도 GWR분석 통행발생모형의 오차검증 통계량이 더 적어지는 것으로 나타나 GWR 통행발생모형의 추정력이 상대적으로 우수한 것으로 나타났다.

마지막으로 통행목적별로 유의미한 변수로 채택된 CDR은 데이터의 특성상 해당 지역에서 통신이 이루어지는 경우로서 통행량이나 사람의 이동 즉, 활동인구 등의 개념으로 해석이 가능한데 이러한 CDR의 통행목적별 통행발생모형으로부터 유의미한 계수값에 대한 공간적 분포를 살펴보았을 때, 대구광역시내의 경우 달서구, 달성군, 북구의 일부지역의 지역적 활성화가 많이 이루어지고 있는 것으로 나타났고 필수적통행과 선택적통행과 같은 통행목적이 세분화될 경우 남구나 수성구 일부지역이 포함되는 도시의 구조를 나타내었다. 이를 바탕으로 통행목적별로 유의미한 활동력을 가진 해당 도시지역에 따라 보다 집중된 도시 인프라의 구축이나 대중교통을 비롯한 교통 서비스의 탄력적인 제공 등이 필요할 것으로 판단된다.

이를 근거로 본 연구의 결과를 요약하면 다음과 같다.

첫째, CDR을 설명변수로 적용하여 통행발생모형을 비교분석한 결과 모든 통행목적의 통행발생모형에서 설명력과 적합도가 개선된 결과를 나타내었다.

둘째, 통행이라는 공간적 데이터를 반영할 수 있는 공간계량분석을 통한 통행발생모형의 구축으로 모형의 설명력 및 적합도가 개선된 결과를 보였고, 데이터 오차검증 결과에서도 공간상관성을 고려한 GWR 통행발생모형이 우



수한 결과를 나타내었다.

셋째, 통행발생모형의 주요한 설명변수로 채택된 CDR의 경우, 통행목적별로 유의미한 계수값의 공간적 분포를 통해 도시의 지역적 활동성에 대해 알아볼 수 있으며 이를 통해 도시의 교통인프라 구축 및 탄력적인 교통 서비스의 제공 측면에서의 활용이 가능할 것으로 기대된다.

## 6.2. 연구의 한계와 향후 연구과제

본 연구는 통행발생모형 구축에 있어 CDR을 바탕으로 공간적 상호작용을 고려한 통행발생모형을 구축하였다는 의의가 있으나, 개인정보 보호 관련 법·제도적 한계에 따라 통신 빅 데이터의 가장 큰 장점인 ‘전수에 가까운 개인의 이동성을 내포한 모바일 데이터’가 가진 통행 기·종점 및 경로를 추정하는 데 한계가 있어 빅 데이터의 적극적 활용측면에서는 다소 아쉬움이 있다. 또한 CDR이 가진 광범위한 공간 상에 분포된 빅 데이터를 통행발생모형에 한정하여 직접적 비교와 분석의 편의, 설명변수들의 적용상의 문제로 인해 공간상에서 집계화할 수 밖에 없었다는 점에서는 분명한 한계가 있다.

다음으로 취득한 CDR 데이터의 자료 시점과 여타 설명변수들의 기준연도가 불일치한 문제가 있어 실무적인 적용보다는 연구 목적으로의 분석이 이루어졌다는 한계가 존재한다. 향후에는 이러한 문제점에 대해 심층적으로 고찰한 연구가 필요하며, 공간적 집계단위의 세분화에 따라 모바일 자료 뿐만 아니라 광범위한 빅 데이터 자료의 적극적 활용이 이루어질 필요가 있다고 생각한다. 또한 본 연구에서 공간상관성 감안을 위해 각 설명변수들의 국지적 영향을 살펴본 GWR 분석기법에서 더 나아가 적용된 설명변수들의 공간적 영향이 전역적인지 국지적인지 판별이 가능한 혼합 지리가중회귀분석(Mixed



Geographically Weighted Regression: MGWR) 기법을 이용한 각 통행발생영향  
요인들의 공간적 영향과 함께 새로운 통행발생모형의 구축을 시도해 볼 가치  
가 있다고 판단된다.





## 참고문헌

1. 진창중, 이향숙, 추상호(2012), “공간회귀분석을 활용한 통행발생모형 추정 연구 : 서울특별시를 중심으로”, 국토연구, 제73권, pp131~143.
2. 이연수, 진창중, 추상호(2012), “공간계량분석을 이용한 대중교통 이용에 영향을 미치는 공간적 특성요인 분석에 관한 연구-서울시 행정동을 중심으로-”, 서울도시연구, 제13권 제4호, pp97~111.
3. 김진희, 박일섭, 정진혁(2011), “공간가중회귀분석을 이용한 통행발생모형”, 대한교통학회지, 제29권 제2호.
4. 변필성(2007), “공간계량경제모델링 : 지리학의 제1법칙과 공간회귀모델”, 월간국토, 2007년 2월호, pp111~119.
5. 이경아(2016), “공간계량모형을 활용한 교통사고 유형별 발생 특성 분석 - 서울시를 대상으로-”, 서울대학교 박사학위논문.
6. 홍지연(2013), “공간계량분석을 통한 도시교통사고 예측모형”, 서울시립대학교 박사학위논문.
7. 이석환(2014), “공간의존성에 대한 이해와 공간회귀분석의 활용”, 한국정책과학회보 18(3), pp39~79.
8. 김광구(2003), “공간자기상관의 탐색과 공간회귀분석의 활용”, 정책분석평가학회보, 제13권 제1호, pp.273~294.
9. 정우현(2011), “공간종속성·이질성을 고려한 통근통행발생모형 개발”, 아주대학교 박사학위논문.
10. 박신형(2010), “고속도로 사고취약구간 판별을 위한 공간분석기법 연구”, 서울대학교 박사학위논문.



11. 김경태, 이인묵, 민재홍, 곽호찬(2015), “무선통신 자료를 활용한 통행발생량 분석”, 한국철도학회논문집, 제18권 제5호, pp.481~488.
12. 심준석, 김호용, 남광우, 이성호(2013), “지리가중회귀모델을 이용한 역세권 공간구조 특성 분석”, 한국지리정보학회, 제16권 제1호, pp.67~79.
13. 엄진기, 송지영, 이광섭(2014), “2호선 혼잡구간 OD수요 분리유도를 통한 혼잡도 개선 방안-교통카드 빅데이터 분석을 중심으로”, 한국철도학회논문집, 제17권 제6호, pp.457~465.
14. 김경태, 이인묵, 곽호찬, 민재홍(2015), “유동인구 추정 시 통신 자료의 활용에 관한 연구”, 서울도시연구, 제16권 제3호, pp.177~187.
15. 성명준, 엄진기, 최명훈, 이준(2010), “스마트 카드 데이터를 이용한 통행패턴 분석”, 한국철도학회 학술발표대회논문집, pp.2076~2087.
16. 김감영(2011), “GWR과 공간 군집 분석 기법을 이용한 중심지 식별 : 대구광역시를 사례로”, 한국도시지리학회지, 제14권 제3호, pp.73~86.
17. 전해정(2016), “GWR모형과 GIS를 이용한 주택가격 추정에 관한 연구”, 한국주거환경학회, 제14권 제1회, pp.1~11.
18. 송용철, 박현수(2012), “공간계량경제 접근방법을 이용한 농지가격추정에 관한 연구-수도권 근교농지를 중심으로”, 국토연구, pp.121~140.
19. 박현수, 정수연, 노태욱(2003), “공간계량경제모형을 이용한 아파트가격과 공간효과분석”, 국토계획, 제38권 제5호, pp.115~125.
20. 전제호(2009), “공간적 집계수준에 따른 교통수요 예측의 영향 분석(서울시 네트워크를 중심으로)”, 서울대학교 석사학위논문.
21. 최기주, 이성욱(1997), “교통분석론 구성체계의 개발”, 대한교통학회지, 제15권 제4호, pp.93~112.



22. 국토연구원(2004), 교통수요 예측을 위한 기준 및 절차 지침 연구.
23. 임용택, 강민구, 아창훈(2008), “교통수요예측시 O/D존 및 네트워크 집계수준에 따른 영향분석”, 대한교통학회지, 제26권 제2호, pp.149~156.
24. 이광섭, 엄진기, 성명언, 유소영, 민재홍, 이준(2017), “통신량 빅데이터 활용성에 관한 연구: 의정부시 내외부 활동 및 이동패턴과 도시철도망 연계성을 중심으로”, 대한국토·도시계획학회지, 제52권 제5호, pp.113~130.
25. 심준석, 김호용, 남광우, 이성호(2013), “지리가중회귀모델을 이용한 역세권 공간구조 특성 분석”, 한국지리정보학회지 제16권 제1호, pp.67~79.
26. 최경호(2003), “교통존 규모가 교통수요예측에 미치는 영향”, 명지대학교 석사학위논문.
27. 김정인, 김익기(2015), “대존 세분화에 따른 내부 소존 간의 O-D 통행량 추정 방법”, 대한교통학회지, 제33권 제6호, pp.575~583.
28. 한국교통연구원(2010), 대도시 권역의 교통수요 분석 가이드라인 기초연구.
29. 한명주, 성홍모, 백승한, 임용택, 이영인(2011), “존 체계 구축이 교통수요 추정에 미치는 영향에 관한 연구”, 대한교통학회지, 제29권 제1호, pp.17~27.
30. 유영근, 이주환(2001), “존내통행의 특성과 Network배분 방법에 관한 연구”, 대한토목학회논문집, 제21권 제2-D호, pp.153~164.
31. 서울연구원(2016), 빅데이터와 교통정책의 연계 방향.
32. 김주영(2018), “2018년 국가교통빅데이터 연구방향과 과제”, 월간교통, pp.30~34.
33. 서울특별시(2013), 민관융합 빅데이터와 공공데이터, 서울특별시 정보기획단.
34. 성지은, 박기량(2014), “빅데이터를 활용한 정책 사례 분석과 시사점”, 과학기술정책, 제24권 제2호, pp.94~106.
35. 국토연구원(2014), 공간 빅데이터를 활용한 국토도시 정책방안 연구.



36. 엄진기(2015), “교통분야의 빅데이터 활용 활성화 방향”, 국토, pp.6~11.
37. 윤서연(2015), “교통분야의 빅데이터 활용과 위치정보 보호”, 국토, pp12~17.
38. 진상규, 김황배(2016), “교통카드 빅 데이터를 활용한 철도역의 대중교통 연계영향권 설정을 위한 GIS 분석 기법 연구”, 대한토목학회논문집, 제36권 제6호, pp1093~1099.
39. 엄진기, 민재홍(2016), “도시철도 노선계획을 위한 모바일폰 데이터 활용방안”, 한국도시철도학회논문집, 제4권 제3호, pp.551~555.
40. 길재현(2015), “개인 및 가구 속성을 고려한 통행발생원단위 산정 모형”, 서울대학교 석사학위논문.
41. 김시곤, 유병석, 강승필(2005), “휴대폰 기지국 정보를 이용한 O/D 추정기법 연구”, 대한교통학회지, 제23권 제1호, pp.93~102.
42. 국토연구원(2018), 모바일 빅데이터를 활용한 기종점 통행량 구축 및 활용방안 연구
43. 김태호, 노정현, 김영일, 오영택(2010), “Zone특성 분할을 통한 유형별 통행발생 모형 개발”, 한국도로학회 논문집, 제12권 제4호, pp.93~100.
44. 김진자, 이종호(2004), “수도권지역의 통행발생모형의 검증(회귀모형과 카테고리모형을 중심으로)”, 대한교통학회지, 제22권 제3호, pp.49~58.
45. 장수은, 김대현, 임강원(2000), “신경망 이론을 이용한 통행발생 모형연구(선형/비선형 회귀모형과의 비교)”, 대한교통학회지, 제18권 제4호, pp.95~105.
46. 송재인, 나승원, 추상호(2011), “토지이용 특성을 반영한 통행발생모형 추정 연구”, 한국ITS학회논문지, 제10권 제6호, pp.126~139.
47. 윤대식(2011), 도시모형론(제4판) 분석기법과 적용, 서울: 기문사.
48. 이성우, 윤성도, 박지영, 민성희(2006), 공간계량모형응용, 박영사.



49. 지아웨이 한, 미셸린 캄버, 지안 페이(2015), 데이터 마이닝 개념과 기법, 서울: 에이콘출판.
50. 이학식, 임지훈(2013), SPSS 20.0 매뉴얼, 서울: 집현재.
51. 우재남(2016), 이것이 MySQL이다, 서울: 한빛미디어.
52. 국토연구원(2004), 공간분석기법, 파주; 한울.
53. 대구광역시(2016), 2016년 교통 관련 기초조사(Ⅱ)-여객기종점통행량(O/D) 현행화 및 장래 수요예측-
54. 이혜승, 이희연(2009), “서울시 대중교통체계 개편 이후 통근 교통수단 선택의 차별적 변화”, 대한지리학회, 제44권 제3호, pp.323~338.
55. 이원규(2013), “빅 데이터(Big Data)를 활용한 스마트 도시관리”, BDI정책포커스 제190호, pp.1~12.
56. 김병석, 이동석, 손동글(2017), “공간계량분석을 이용한 도시특성요인이 지역 인구에 미치는 영향에 관한 연구-인천광역시를 중심으로-”, 지역연구, 제33권 제3호, pp.21~30.
57. 박일수, 김은주, 홍성옥, 강성홍(2013), “지리적 가중회귀모형을 이용한 지역별 심정지 발생요인에 관한 연구”, 보건사회연구, 제33권 제3호, pp.237~257.
58. 김경태, 오동규, 이인묵, 민재홍(2016), “이동통신 자료를 활용한 가정기반 OD 구축 및 분석”, 한국철도학회논문집, 제19권 제5호, pp.656~662.
59. 엄진기(2008), “시·공간 활동인구 추정에 의한 통행수요 예측”, 대한교통학회지, 제26권 제5호, pp.163~174.
60. 한국교통연구원(2016), 2016년 국가교통조사 및 DB구축사업 전국 여객 기종점 통행량 조사.



61. Anselin, L.(2013), Spatial econometrics: methods and models, Springer Science & Business Media.
62. LeSage, J. and R. K. Pace(2009), Introduction to spatial econometrics, Chapman and Hall/CRC.
63. Fischer, M. M. and A. Getis(2009), Handbook of applied spatial analysis: software tools, methods and applications, Springer Science & Business Media.
64. Tobler, W. R.(1970), "A computer movie simulating urban growth in the Detroit region." Economic geography, Vol. 46, No.(sup1), pp. 234-240.
65. McMillen, D. P. (2004). Geographically weighted regression: the analysis of spatially varying relationships, Oxford University Press.
66. Geary, R. C.(1954), "The contiguity ratio and statistical mapping." The incorporated statistician, Vol. 5, No.(3), pp.115-146.
67. Anselin, L.(1995), "Local indicators of spatial association—LISA." Geographical analysis, Vol. 27, No.(2), pp.93-115.
68. LeSage, F.(1999), Spatial Econometrics, Web Books,<http://www.mri.wvu.edu/WebBook/LeSage/spatial/spaia1.html>, University of Toledo.
69. Ord, J. K. and A. Getis(1995), "Local spatial autocorrelation statistics: distributional issues and an application.", Geographical analysis, Vol. 27, No.(4), pp.286-306.
70. Charlton, M., A. S. Fotheringham.(2003), Geographically Weighted Regression: White Paper, National Centre for Geocomputation, National University of Ireland.
71. LeSage, J. P. and R. K. Pace(2008), "Spatial econometric modeling of origin destination flows.", Journal of Regional Science, Vol. 48, No.(5), pp.941-967.
72. Nakaya, T.(2001), "Local spatial interaction modelling based on the geographically



- weighted regression approach.", *GeoJournal*, Vol. 53, No.(4), pp.347-358.
73. Kim, J., J. Chung, I. Kim, and U. Lim.(2009), "Spatial Dependence in Trip Generation Models: A Spatial Econometric Perspective.", in *Proceeding of 88th Annual Meetings of Transportation Research Board*, Washington D.C.
  74. Kim, J., J. Chung, et al.(2009)“. Spatial Dependence in Trip Generation Models: Spatial Econometric Perspective.". in *Proceeding of 88th Annual Meetings of Transportation Research Board*, Washington D.C.
  75. Niemeier, D. A. and S. Bai.(2009), "Urban Travel Demand Modeling." *Transportation Planning Handbook*, 3rd Edition, Institute of Transportation Engineers, pp.123-150.
  76. Balzotti, C., A. Bragagnini, et al.(2018), "Understanding Human Mobility Flows from Aggregated Mobile Phone Data.", *IFAC PaperOnLine*, 51(9), pp.25~30.
  77. Widhalm, P., Nitsche et al.(2012), "Transport Mode Detection with Realistic Smartphone Sensor Data.", *ICPR*, pp.573~576.
  78. Louail, T., M. Lenormand, et al.(2014), "From mobile phone data to the spatial structure of cities." *Scientific reports*, Vol. 4, pp.5276.
  79. Bowman, J. L. and M. E. Ben-Akiva(2001), "Activity-based disaggregate travel demand model system with activity schedules.", *Transportation research part a: policy and practice*, Vol. 35, No.(1), pp.1-28.
  80. Çolak, S., Lauren, P., Bernardo, G., Shomik, R. and Marta C. González.(2015), "Analyzing Cell Phone Location Data for Urban Travel-Current Methods, Limitations, and Opportunities.", *Transportation Research Record: Journal of Transportation Research Board*, No.2526, pp.126~135.
  81. Çolak, S., L. P. Alexander, et al.(2015), "Analyzing cell phone location data for urban travel: current methods, limitations, and opportunities.", *Transportation research record:*



- Journal of the transportation research board, No.(2526), pp.126-135.
82. Reddy, S., M. Mun, et al.(2010), "Using mobile phones to determine transportation modes.", *ACM Transactions on Sensor Networks(TOSN)*, Vol. 6, No.(2), pp.13.
  83. Gadziński, J.(2018), "Perspectives of the use of smartphones in travel behaviour studies: Findings from a literature review and a pilot study.", *Transportation Research Part C: Emerging Technologies*, Vol. 88, pp.74-86.
  84. Sohn, K. and D. Kim(2008), "Dynamic origin - destination flow estimation using cellular communication system.", *IEEE Transactions on Vehicular Technology*, Vol. 57, No.(5), pp.2703-2713.
  85. Páez, A. and D. M. Scott(2005), "Spatial statistics for urban analysis: a review of techniques with examples.", *GeoJournal*, Vol. 61, No.(1), pp.53-67.
  86. Mei, C-L(2005), "Geographically weighted regression technique for spatial data analysis." School of Science Xi'an Jiaotong University.
  87. Chen, C., J. Ma, et al.(2016), "The promises of big data and small data for travel behavior (aka human mobility) analysis.", *Transportation Research Part C: Emerging Technologies*, Vol. 68, pp.285-299.
  88. Yao, L., H. Guan, et al.(2008), "Trip generation model based on destination attractiveness.", *Tsinghua Science and Technology*, Vol. 13, No.(5), pp.632-635.
  89. Chang, J. S., D. Jung, et al.(2014), "Comparative analysis of trip generation models: results using home-based work trips in the Seoul metropolitan area.", *Transportation Letters*, Vol. 6, No.(2), pp.78-88.
  90. Golob, T. F.(2000), "A simultaneous model of household activity participation and trip chain generation.", *Transportation Research Part B: Methodological*, Vol. 34, No.(5), pp.355-376.





91. Amavi, A. A., J. P. Romero, et al.(2014), "Advanced trip generation/attraction models.," *Procedia-Social and Behavioral Sciences*, Vol. 160, pp.430-439.
92. Alexander, L., S. Jiang, et al.(2015), "Origin - destination trips by purpose and time of day inferred from mobile phone data.", *Transportation Research Part C: Emerging Technologies*, Vol. 58, pp.240-250.
93. Toole, J. L., S. Colak, et al.(2015), "The path most traveled: Travel demand estimation using big data resources.", *Transportation Research Part C: Emerging Technologies*, Vol. 58, pp.162-177.
94. Hadayeghi, A., A. S. Shalaby, et al.(2010), "Development of planning level transportation safety tools using Geographically Weighted Poisson Regression.", *Accident Analysis & Prevention*, Vol. 42, No.(2), pp.676-688.
95. Epperlein, J., J. Legierski, et al.(2018), *The Use of Presence Data in Modelling Demand for Transportation*.
96. Ahmed, E., I. Yaqoob, et al.(2018), "Recent Advances and Challenges in Mobile Big Data.", *IEEE Communications Magazine*, Vol. 56, No.(2), pp.102-108.
97. Sanja Šćepanović et al.(2015), "Mobile Phone Call Data as a Regional Socio-Economic Proxy Indicator.", *PLoS ONE* 10(4): e0124160.
98. Do, C. X. and M. Tsukai(2017), "Exploring Potential Use of Mobile Phone Data Resource to Analyze Inter-regional Travel Patterns in Japan.", *International Conference on Data Mining and Big Data*, Springer.





Ph.D. Thesis

Development of Trip Generation Models considering Spatial  
Autocorrelation based on Mobile phone data

Keun Uk Kim

Department of Urban Planning and Engineering

Graduate School

Yeungnam University

(Supervised by Professor Younshik Chung)

Abstract

The objective of this dissertation is to develop trip generation models for improving the estimation accuracy of trip generation as the first fundamental step in the travel demand estimation by using mobile big data, called call detail records (CDR), and by using a geographically weighted regression (GWR) method, which explores spatial non-stationarity in the parameters estimated for more particular localities. Due to the relative newness of CDR data on the travel demand modeling area, the first part of this dissertation introduces its characteristics and preliminary data analysis for GWR analysis.



Historically, O/D trips have been estimated based on an ordinary least squares (OLS) method with traditional survey data. However, since the availability of CDR data poses new challenges to the travel demand modeling, specifically in the trip generation step, it is required to identify the relationship between CDR and traditional origin-destination (O/D) trips. Therefore, their relationships are statistically analyzed. Based on the results, two types of trip generation models are estimated with various independent variables including socioeconomic characteristics as well as CDR data: 1) an OLS-based model, and 2) a GWR-based model.

In general, the validity of estimated trip generation models has been assessed by using various goodness-of-fit (GoF) measures such as the small-sample-size corrected version of Akaike information criterion (AICc) and the adjusted  $R^2$ . Additionally, their accuracy has been compared by using the performance index of prediction models such as the root-mean-square error (RMSE) and the standardized root mean square error (SRMSE). In this study, two estimated trip generation models are compared in terms of the validity and accuracy.

Although the proposed trip generation model is at an early stage and includes various limitations such as privacy issues for the use of CDR data and difficulties for identifying individual activities or trajectories and for combining demographics and CDR data, the advancement of mobile information and communication technology (ICT) will play an important role in reducing such limitations. Then, this study can serve as a stimulus and a basis to address in a more rationale way the use of CDR data in the travel demand modeling area.

