creative commons

COMMONS DEED

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

dCollection

도시계획학 석사 학위논문

기계학습을 활용한 소득계층별 통근시간 및 통근거리 영향요인 차이에 관한 연구

Study on the Influencing Factors on Commuting Time and Commuting Distance by Income Class Using Machine Learning

2023년 2월

서울대학교 대학원

환경계획학과 도시 및 지역계획학 전공

김성주

기계학습을 활용한 소득계층별 통근시간 및 통근거리 영향요인 차이에 관한 연구

지도교수 송 재 민

이 논문을 도시계획학 석사 학위논문으로 제출함
2023년 2월

서울대학교 대학원
환경계획학과 도시 및 지역계획학 전공
김 성 주

김성주의 석사 학위논문을 인준함
2023년 2월

위 원 장 김 경 민 (인)

부위원장 김 태 형 (인)

위 원 송 재 민 (인)

국문초록

통근시간 및 통근거리를 줄이는 것은 직주근접의 형태를 지닌 압축도시로 나아가 탄소배출을 줄이기 위한 주요한 공간 전략이며, 근로자의 삶의 질을 높이기 위한 중요한 방안이다. 통근시간, 거리를 줄이기 위해서는 다양한 요인들이 고려되어야 한다. 그 중 주거비용은 통근비용 결정에 있어 큰 영향을 끼칠 수 있다. 주거비용과 통근비용이 상쇄관계를 지니고 있다는 것은 오랜 기간 도시경제 이론에서 설명되었으며, 대도시를 중심으로 한 부동산 가격의 상승은 근로자를 일자리로부터 떨어진 주택 가격이 낮은 지역으로 이동하게 만들어, 통근시간 및 거리가 길어지는 양상이 나타나고 있다. 이 뿐 아니라, 교육, 안전, 도시 어메니티 등 삶의 질에 대한 고려 또한 고용 접근성에 대해 상쇄관계를 지니기에, 거주지의 환경적 질을 나타내는 특성 또한 통근시간 및 거리 결정의 주요 요인으로 고려될 수 있다.

특히, 소득계층별로 주거비용 및 거주지 특성이 통근시간 결정에 미치는 영향력은 다를 수 있다. 왜냐하면, 저소득 계층은 한정된 재원으로 인해 주거비 및 교통비에 매우 민감할 가능성이 높으나, 고소득 집단은 주거지를 선택하는데 있어 대중교통 수단의 이용가능성을 덜 고려할 뿐만 아니라 다른 어메니티 요소를 고려할 가능성이 높기 때문이다. 그럼에도 주거비와 통근의 관계는 여전히 분명하지 않으며, 특히 소득계층별로 나누어 주거비용의 영향력 차이를 분석한 연구는 매우 부족하다.

따라서 거주지, 직장지 특성 및 개인 특성이 통근시간 및 통근거리에 미치는 영향에 대한 심도있는 연구가 필요하며, 이를 소득계층별로 나누어 파악하여 그 차이를 비교해볼 필요가 있다.

이와 같은 배경에서 본 연구는 수도권 통근자의 가구통행실태조사 자료를 이용하여 소득계층별 통근시간 및 통근거리 영향요인 차이를 세밀하게 비교 분석했다. 분석 방법론으로는 우수한 설명력을 지닌 기계학습 방법론인 XGBoost를 사용하였다. 주택가격, 거주지 특성, 직장지 특성 및 개인 변수와 같은 다양한 요인들을 활용하여 소득계층별 통근시간 및 통근거리 결정에 주요한 영향요인을 파악했고, 각 영향요인이 통근시간 및 통근거리에 어떤 비선형관계를 지니는지 확인했다.

연구 결과에 따르면, 고소득층의 통근시간 및 거리가 가장 길게 나타났다. 고소득층일수록 직장지가 고용 중심지에 가깝게 위치하고 있었는데, 직장지가 고용중심지에 가까울수록 통근시간은 길어졌다. 반면, 통근거리는 중, 고소득층의 경우 직장지가 서울 3도심에서 일정 이상 멀어지면 오히려 길어지는 것으로 나타났다. 거주지와 직장지 간 주택가격의 차이가 통근시간 및 거리에 미치는 영향을 보면, 0을 기준으로 U자형 그래프를 그리고 있었다. 특히, 고소득층으로 갈수록, 서울에 거주할수록 통근시간은 증가함에도 불가하고 직장지에 비해 비싼 거주지를 선택하는 경향이 심화되었다. 저소득층의 경우 대중교통 통근이 주를 이루고 거주지에서 지하철역까지의 거리가 통근시간 결정에 중요하게 작용하였으며, 거주지 및 직장지가 저밀화되어 있을수록 통근거리가 증가했다. 중, 고소득층에서는 근린환경 개선에 관한 요인이 통근시간 증가에 영향을 미치고 있었다.

연구 결과를 토대로, 통근시간 및 통근거리를 줄이기 위한 주거 정책의 차별화가 어떻게 이루어져야 하는지에 대한 시사점을 제안하였다. 본 연구는 전반적인 소득계층의 통근시간 감축을 위해서는 서울 3도심 및 이외 고용 중심지에 인접한 저렴한 주택 공급이 어느 정도 필요하며, 저소득층의 통근시간 감축을 위해서는 대중교통과 인접하

고, 고밀도 지역에 저렴한 주택 공급이 강조되어야 함을 시사한다. 반면, 중, 고소득층의 경우 통근시간 감축을 위해 직장지 인근의 주거환경 개선이 중요하고, 지나친 고용지의 교외화는 자가용 중심의 장거리 통근을 유발할 수 있음을 시사한다.

주요어 : 주택가격, 거주지 특성, 통근시간, 통근거리, 기계학습
학 번 : 2021-22630

목 차

제 1 장 서론 ····· 1

제 1 절 연구의 배경 및 목적 ····· 1

제 2 절 연구의 범위 및 방법 ····· 3

1. 연구의 대상 및 범위 ····· 3
2. 연구의 구성 ····· 4
3. 연구의 흐름 ····· 5

제 2 장 이론적 배경과 선행연구의 고찰 ····· 7

제 1 절 이론적 배경 ····· 7

1. 이론적 배경 검토 ····· 7
2. 통근비용과 주거비용의 관계 ····· 8

제 2 절 선행연구 검토 및 연구의 차별성 ····· 10

1. 통근행태 및 통근거리 결정요인 ····· 10
2. 소득계층별 통근행태 및 통근거리 결정요인 ····· 11
3. 연구의 차별성 ····· 13

제 3 장 자료 및 분석방법 ····· 15

제 1 절 분석자료 및 표본 ····· 15

제 2 절 연구가설 및 연구방법 ····· 16

제 3 절 변수설정 ····· 18

제 4 장 소득계층별 통근거리 영향요인 실증분석 ··· 23

제 1 절 기술통계 ····· 23

제 2 절 기계학습 모형 ····· 30
제 3 절 소득계층별 분석 결과 비교 ····· 34
1. 변수의 기여도 (Feature Importance) ····· 26
2. 소득계층별 영향요인의 비선형 관계 비교 ····· 43

제 5 장 결론 ····· 59
제 1 절 연구의 요약 및 시사점 ····· 59
제 2 절 연구의 의의 및 한계 ····· 63

참고문헌 ····· 65
부록 ····· 70
Abstract ····· 86

표 목 차

[표 1] 연구의 공간적 범위 ····· 4
[표 2] 분석에 사용된 변수 ····· 22
[표 3] 소득계층별 기초통계량 ····· 27
[표 4] 통근시간(분) 분석모형 평가 결과 ····· 30
[표 5] 통근거리(km) 분석모형 평가 결과 ····· 31
[표 6] 소득계층별 통근시간에 대한 실제값 및 예측값 ····· 32
[표 7] 소득계층별 통근거리에 대한 실제값 및 예측값 ····· 32
[표 8] 통근시간에 대한 XGBoost 모형에 사용된 매개변수 · 33
[표 9] 통근거리에 대한 XGBoost 모형에 사용된 매개변수 · 33
[표 10] 소득계층별 통근시간에 영향을 미치는 변수의
Global, Local Shapley Value ····· 35
[표 11] 소득계층별 통근거리에 영향을 미치는 변수의
Global, Local Shapley Value ····· 36
[표 12] 소득계층별 통근시간에 영향을 미치는
변수의 기여도 Shapley Value 순위 ····· 37
[표 13] 소득계층별 통근거리에 영향을 미치는
변수의 기여도 Shapley Value 순위 ····· 41
[표 14] 소득계층별 직장지 대비 거주지 주택가격 분포 ····· 50

그 림 목 차

[그림 1] 연구 흐름도 ····· 6
[그림 2] 시군구별 아파트 ····· 18
[그림 3] 시군구별 연립/다세대/다가구 ····· 18
[그림 4] 시군구별 오피스텔 ····· 18
[그림 5] 시군구별 단독주택 ····· 18
[그림 6] 변수 간 상관성 ····· 21
[그림 7] 고소득 응답자 거주지 분포 ····· 25
[그림 8] 고소득 응답자 직장지 분포 ····· 25
[그림 9] 중소득 응답자 거주지 분포 ····· 25
[그림 10] 중소득 응답자 직장지 분포 ····· 25
[그림 11] 저소득 응답자 거주지 분포 ····· 25
[그림 12] 저소득 응답자 직장지 분포 ····· 25
[그림 13] 소득계층별, 나이대별 통근수단 이용비율 ····· 39
[그림 14] 저소득층의 통근시간 영향요인에 대한 PDP ····· 45
[그림 15] 중소득층의 통근시간 영향요인에 대한 PDP ····· 47
[그림 16] 고소득층의 통근시간 영향요인에 대한 PDP ····· 49
[그림 17] 소득계층별 거주지와 직장지의 주택가격 차 분포 · 50
[그림 18] 저소득층의 통근거리 영향요인에 대한 PDP ····· 53
[그림 19] 중소득층의 통근거리 영향요인에 대한 PDP ····· 55
[그림 20] 고소득층의 통근거리 영향요인에 대한 PDP ····· 57

제 1 장 서론

제 1 절 연구의 배경 및 목적

통근시간 및 통근거리를 줄이는 것은 직주근접의 형태를 지닌 압축도시로 나아가 탄소배출을 줄이기 위한 주요한 공간 전략이다. 더불어, 통근시간 감축에 따른 개인 시간의 증가는 삶의 질 향상에 기여한다(장재민 외, 2019). 도시계획 분야에서 통근거리 및 통근시간을 줄이는 직주근접의 실현은 오랜 시간 공간계획의 주요한 목표로 남아있다.

통근시간 및 통근거리를 줄이기 위해서는 다양한 요인들이 고려되어야 한다. 다양한 요인들 중 주거비용은 통근비용 결정에 있어 큰 영향을 끼칠 수 있다. 어디서 살 것인가와 어디서 일할 것인가는 통합된 결정일 수 있기 때문이다(Molho, 1986). 오랜 기간 도시경제 이론에서는 주거비용과 통근비용이 상쇄관계를 지니고 있음을 주장해 왔다. 현대 도시경제이론의 토대가 되는 Alonso의 주거입지이론에서는 도시가구는 주거입지를 선택할 때 직장에 대한 접근성을 의미하는 교통비용과 주거비용의 상쇄관계(trade-off) 속에서 효용을 극대화하는 주거입지를 선택한다고 설명한다(전명진·안현주, 2016). 특히, 대도시를 중심으로 한 부동산 가격의 상승은 근로자를 일자리로부터 떨어진 주택 가격이 낮은 지역으로 이동하게 만들어 통근시간 및 거리가 더 길어지는 양상이 나타나고 있다 (Blumenberg and King, 2021).

또한, 주거입지를 설명하는 시장중심 접근방법 이론인 접근성-공간 맞교환 이론(Access-Space Trade-off)에 따르면, 가구의 주거입지 선택은 고용 접근성에 대한 고려와 삶의 질에 대한 고려의 상쇄관계에 의해 결정된다 (Phe and Wakely, 2000; Kim-J et al, 2005; Moos and Skaburskis, 2010; 김승남·안건혁, 2011). 따라서 주거비용 외에도 지역의 교육(Kim-J et al., 2005), 안전(Weisbrod et al., 1980), 도시 어메니티(Rapoport, 1980; Tyrvainen and Vaananen, 1998; Rouwendal and Meijer,

2001; Parkes et al.,2002) 등 거주지의 환경적 질을 나타내는 특성 또한 고용 접근성과 상쇄관계를 지니는 거주지 선택의 주요 요인으로 고려될 수 있다(김승남·안건혁, 2011).

이와 같이, 주거비용과 거주지의 환경적 질을 나타내는 거주지 특성은 고용접근성과 상쇄관계를 가진다. 특히, 소득계층별로 주거비용 및 거주지 특성이 통근시간 및 거리 결정에 미치는 영향력은 다를 수 있다. 왜냐하면, 저소득 계층은 한정된 재원으로 인해 주거비 및 교통비 상승에 매우 민감하게 반응하지만, 고소득층은 주거지 선택에 있어 자가용 이용으로 대중교통 접근성을 고려할 필요성이 낮고 다양한 어메니티 요소를 고려할 가능성이 높기 때문이다(Adair et al., 2000; 하재현 2020). 우리나라를 대상으로 진행한 이민주 & 박인권(2016) 연구에서도 고소득층보다 저소득층의 통근시간이 길게 나타나며, 주택가격이 높은 지역일수록 해당 지역으로 통근하는 저소득층과 고소득층의 통근시간 격차가 큰 것으로 나타났다.

한편, 저소득층보다 고소득층에서 주거비용이 통근시간 및 거리에 미치는 영향력이 더욱 크게 나타날 수도 있다. 이번송(1998)의 연구에서는 고소득자가 원거리 통근을 하고 있는 것으로 나타났으며, Park & Quercia (2015)는 미국 고소득 근로자는 통근시간의 감소보다 임대료의 상승이 더욱 가파른 경향이 있음을 보여주었다. 이는 고소득층의 직장이 고용중심지, 즉 지대가 더욱 비싼 지역에 위치할 가능성이 있어 주택가격이 저소득층에 비해 오히려 고소득층의 통근시간 결정에 있어 더욱 영향이 클 수도 있음을 의미하고 있다. 이와 같이, 주거비와 통근의 관계는 여전히 분명하지 않으며(Islam·Saphores, 2022), 특히 소득계층별로 나누어 주거비용을 포함한 거주지 특성이 통근시간에 미치는 영향 차이를 분석한 연구는 매우 부족하다.

이와 같은 배경에서 본 연구의 목적은 소득계층별 주택가격을 포함한 거주지 특성과 통근시간 및 통근거리의 관계와, 그 차이를 파악하는 것이다. 더불어, 기계학습 분석을 통해 전통적 회귀모형에 비해 설명력이 높은 모델을 구축하고 요인 간의 비선형 관계를 확인하고자 하였다. 본 연구에서는 수도권을 대상으로, 2016 가구통행실태조사 자료를 활용하여 기계학습 기법

중 하나인 XGBoost를 사용하여 주택가격 및 거주지 특성이 소득계층별 통근시간 및 통근거리에 영향 요인에 대하여 분석하였다. 분석 결과를 토대로, 주거비 및 거주지 특성이 통근시간에 미치는 영향의 상대적 크기를 실증적으로 파악하고, 각 소득계층별 통근시간 및 통근거리 결정에 대한 영향요인이 어떤 차이를 보이는지 비교해보고자 한다.

본 연구를 통해, 실제 수도권에서 통근시간 및 통근거리에 주거비 및 거주지 특성의 중요도가 얼마나 큰지를 경험적으로 입증할 수 있다. 소득계층별로 주택가격 및 거주지 특성이 통근시간 및 통근거리에 영향을 미치는 중요도를 비교하고, 우수한 설명력을 지닌 기계학습을 활용하여 비선형적 관계를 파악할 수 있다. 연구결과를 바탕으로, 직주근접을 위한 소득계층별 주거정책의 차별화에 중요한 시사점을 제시할 수 있다.

제 2 절 연구의 범위 및 방법

1. 연구의 대상 및 범위

본 연구는 2016년 가구통행실태조사를 활용하여, 통근 목적, 자택이 출발지이고 직장이 도착지인 응답자만을 대상으로 한다.

따라서, 시간적 범위는 2016년, 공간적 범위는 수도권 중 도시교통정비촉진법에서 지정한 서울특별시의 교통권역을 대상으로 한다. 도시교통정비촉진법 제2장 제4조에 따르면, 교통권역이란 도시교통정비지역 중 같은 교통생활권에 있는 둘 이상의 인접한 도시교통정비지역 간 연계된 교통관련 계획을 수립할 수 있도록 국토교통부장관이 지정·고시한 것으로, 서울특별시를 중심으로 교통 및 통행에 있어 밀접한 관계를 지닌 수도권 내 지역을 선정하는데 유용한 자료로서 활용될 수 있다고 파악하였다. 따라서 이에 기반하여 연구의 공간적 범위를 설정하였으며, 도서지역으로만 이루어진 옹진군은 연구 범위에서 제외하였다. 연구의 공간적 범위는 [표 1]과 같다.

[표 1] 연구의 공간적 범위

도시교통정비지역	교통권역 (*인천광역시 옹진군 제외)	
서울특별시	인천광역시	중구, 동구, 남구, 연수구, 남동구, 부평구, 계양구, 서구, 강화군
	경기도	성남시 수정구, 성남시 중원구, 성남시 분당구, 의정부시, 안양시 만안구, 안양시 동안구, 부천시 원미구, 부천시 소사구, 부천시 오정구, 광명시, 고양시 덕양구, 고양시 일산동구, 고양시 일산서구, 과천시, 구리시, 남양주시, 하남시, 김포시, 양주시

출처: 도시교통정비 촉진법

2. 연구의 구성

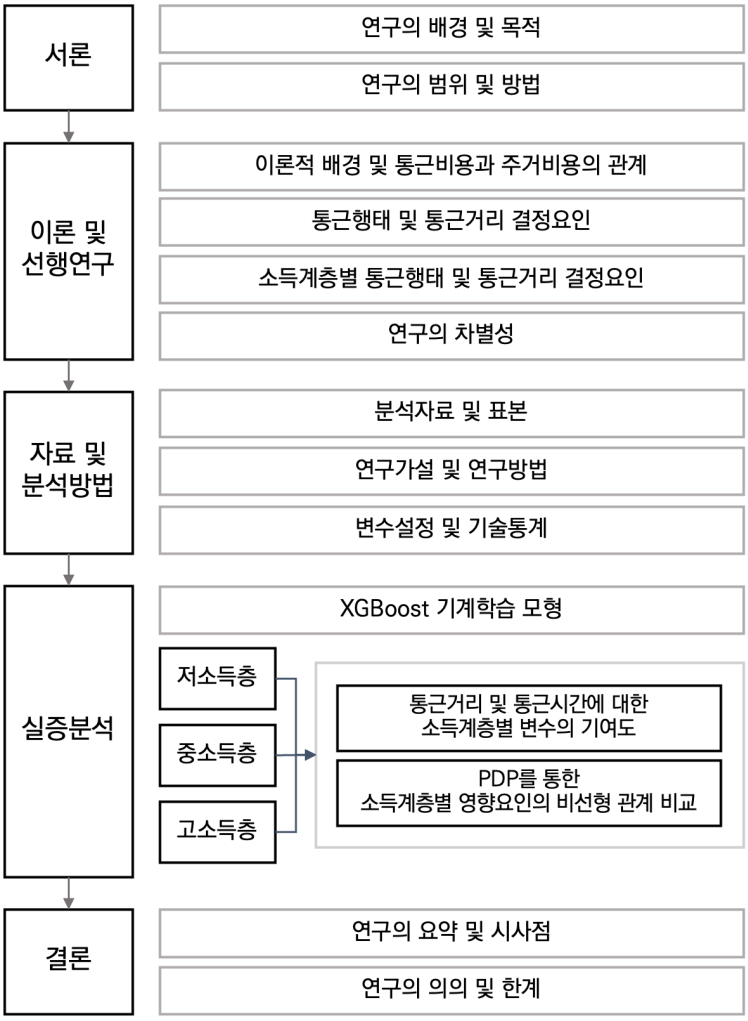
본 연구는 문헌연구와 실증연구로 구성된다. 문헌연구에서는 주택가격과 통근비용 간 관계를 상쇄관계를 지닌다는 지대이론 및 관련 이론들과, 해당 이론들을 바탕으로 한 선행연구들을 고찰한다. 이론 및 기존 연구들을 바탕으로 주택가격과 통근거리 간 관계가 어떠한지, 주택가격 이외 어떠한 요인들이 통근시간 및 통근거리 결정에 있어 중요하게 작용하는지, 소득계층별로 통근시간 및 통근거리 결정에 있어 작용하는 요인들이 어떻게 다른지를 살펴보았다.

다음으로 실증연구에서는 첫째, 비선형적 관계를 분석하기 위해 적합한 분석 방법 및 분석 모형에 필요한 자료를 선정 및 구축하였다. 보다 정확한 주택가격 측정을 위해 국토교통부 실거래가 자료에 기재된 주소를 바탕으로 행정동별 평균 거래가격을 구축하였으며, 주택 유형을 4가지, 즉, 아파트, 연립/다세대/다가구주택, 오피스텔, 단독주택으로 구분하여 개인의 주거비용이 최대한 현실적으로 반영되도록 하였다. 통근시간은 가구통행실태조사 자료에 기입된 응답자의 통근시간을 활용하였으며, 통근거리는 직선거리가 아닌 도로망에 기반한 거리를 산출하였다.

둘째, 구축된 자료를 바탕으로 소득계층별로 기계학습 기법 중 가장 우수한 성능을 보인 XGBoost를 활용하여 분석을 수행하였다. XGBoost를 통해 주택가격 및 거주지 특성이 소득계층별 통근시간 및 통근거리에 미치는 영향에 대해 분석을 진행하였다. 분석 결과를 토대로, 각 소득계층별 통근시간 및 통근거리에 대한 주택가격 및 거주지 특성의 중요도 및 영향관계가 어떤 차이를 보이는지 비교하였다.

3. 연구의 흐름

본 연구는 주택가격 및 거주지 특성이 소득계층별 통근시간 및 통근거리에 미치는 영향을 분석하고 그 차이를 파악하는데 목적이 있다. 이에 따라 다음과 같은 흐름으로 진행되었다. 제1장 서론에서는 연구의 배경 및 목적, 연구의 범위 및 방법을 제시한다. 제2장에서는 통근시간 및 통근거리와 주택가격 간 관계를 규명한 이론적 배경과 관련 선행연구를 검토하고, 본 연구의 차별성을 고찰한다. 제3장에서는 사용한 분석자료와 연구 방법을 설명하고, 이를 바탕으로 변수를 설정하고 이에 대한 기술통계를 살펴본다. 제4장에서는 기계학습 기법을 활용한 실증분석 결과를 바탕으로 소득계층별 변수의 중요도와 통근시간 및 통근거리에 미치는 영향 관계를 비교·분석한다. 마지막으로 제5장에서 본 연구의 의의와 한계를 제시한다. [그림 1]은 연구의 구성을 도식화한 연구의 흐름도이다.


서론
연구의 배경 및 목적
연구의 범위 및 방법
이론 및 선행연구
이론적 배경 및 통근비용과 주거비용의 관계
통근행태 및 통근거리 결정요인
소득계층별 통근행태 및 통근거리 결정요인
연구의 차별성
자료 및 분석방법
분석자료 및 표본
연구가설 및 연구방법
변수설정 및 기술통계
실증분석
XGBoost 기계학습 모형
저소득층
중소득층
고소득층
통근거리 및 통근시간에 대한 소득계층별 변수의 기여도
PDP를 통한 소득계층별 영향요인의 비선형 관계 비교
결론
연구의 요약 및 시사점
연구의 의의 및 한계

[그림 1] 연구 흐름도

제 2 장 이론적 배경과 기존연구의 고찰

제 1 절 이론적 배경

1. 이론적 배경 검토

고전 도시경제이론에서는 단핵도시라는 가정 하에, 도시의 중추기능과 핵심 업무지역을 담당하는 도심을 기준으로 거리가 멀어질수록 주거비용은 감소하게 되는 반면, 통근비용은 증가하게 되어 주거비용과 통근비용이 상쇄관계(Alonso, 1964)를 지닌다고 설명한다. 이에 따라 가구는 효용을 극대화하는 최적의 주거입지를 선택하게 된다(최준성·노승한, 2021). 상쇄 이론의 이론적 기반은 튀넨의 농업용 지대이론에서 시작되었으며, 뒤따라 Alonso(1964)에 의해 주거용 지대이론으로 발전되었고, Muth(1969), Mills(1972) 등에 의해 수정되고 정교화되었다(전명진·안현주, 2016).

폰 튀넨의 동심원 모형은 5가지 가정을 하고 있다. 첫째, 다른 지역과 교역이 전혀 없는 고립된 지역이며, 둘째, 하나의 점으로 표시되는 도시는 주변지역으로부터 농산물을 공급받으며, 해당 도시는 유일하게 시장을 담당하게 된다. 셋째, 해당 지역의 모든 토지는 지형의 차이가 없으며, 넷째, 농산물의 수송비용은 거리와 비례한다. 다섯째, 모든 농부는 이윤극대화를 추구한다. 5가지 가정이 충족되는 도시에서 생산되는 농산물의 단위면적당 연간 순이윤은 연간 총이윤에서 연간 수송비를 차감하여 계산하게 된다. 이에 따라 도심과 거리가 가까워질수록 수송비용은 감소하게 되고, 유일한 시장인 도심을 중심으로 동심원적인 농업용 토지이용이 일어나게 된다.

Alonso(1964)는 이를 토대로 입찰지대이론으로 발전시켰다. 이론은 3가지 전제조건을 기반으로 한다. 첫째, 농지를 임차하려는 수요자와 농지를 소유한 공급자가 모두 충분하며, 이에 따라 언제든지 대체가능한 완전경쟁시장

이다. 둘째, 토지가격 이외의 생산비는 모두 같다. 셋째, 도시의 중심과 시장이 하나인 단일도심으로부터 방사형으로 뻗어나가는 도시구조이다. 토지주는 완전경쟁시장 하에서 가장 높은 토지임대료를 제안하는 농부에 토지를 임대해주게 되고, 이에 따라 토지에서 발생하는 순수익은 토지임대료로 귀속되게 된다. 이에 따라 수송비용이 가장 적게 드는 도심에 가까워질수록 토지임대료가 높아지게 되고, 도심으로부터 멀어질수록 임대료는 줄어들게 된다. 더 나아가, 이러한 결론은 지가가 높은 것은 해당 지역의 생산성이 좋다는 것으로도 설명될 수 있다. 단, Alonso(1964)는 단핵중심모델을 기반으로, 모든 일자리가 한 중심지에 위치한다고 가정했다는데 실제 현대 도시와 다소 차이가 있다.

입찰지대이론을 바탕으로 주거비용과 통근비용의 관계를 살펴본 연구는 다음과 같다.

2. 통근비용과 주거비용의 관계

Islam & Saphores(2022)는 로스앤젤레스 카운티를 대상으로 Generalized Structural Equation Model을 사용하여 주거지와 직장지 주변의 가구 특성과 토지이용 특성을 통제하고, 통근거리에 주택가격이 미치는 영향을 살펴보았다. 연구 결과, 더욱 비싼 동네에 거주할 수 있는 가구는 그렇지 않은 가구보다 통근거리가 짧게 나타났다. 또한, 고소득 통근자는 저소득 통근자보다 빠르게 통근한다고 밝혔다.

Park & Quercia(2015)는 미국 8개 대도시권을 대상으로 주거비용과 통근거리의 관계를 파악하였다. 현대 대도시권의 고용중심지 다핵화와 분산에도 불구하고, 가구는 여전히 근본적인 주거비용과 통근비용의 상쇄관계가 존재한다고 밝혔다. 특히, 고소득 근로자는 통근시간의 감소보다 임대료의 상승이 더욱 가파른 것으로 나타나, 오히려 더 긴 통근시간을 가지고 있는 것으로 나타났다.

한편, Hass & Osland(2013)은 통근, 이주, 주택과 노동 시장의 복잡한

상호관계에 대한 문헌연구를 통해서 소득과 소득 전망은 주택비 부담과 밀접하게 관련이 있으며, 따라서 주택 가격 또는 소득 전망의 공간적 변동은 통근 거리를 결정할 수 있다고 설명하고 있다.

우리나라를 대상으로 한 주요 연구 결과는 다음과 같다. 전명진 & 안현주(2016)는 수도권 통근자를 대상으로 주거비와 통근비의 상쇄관계가 존재하는지 여부를 실증적으로 검증했다. 통근 시간에 대한 소득별 기회비용을 추정하고, 통근 시간에 대한 기회비용을 주거비용 결정 요인에 반영하였다. 분석 결과, 통근 비용이 한 단위 증가하면 주거비용이 3.5% 감소하는 것을 확인하여 주거비용과 통근비용 간의 상쇄관계가 존재함을 입증하였다.

최준성 & 노승한(2021)은 수도권 신도시를 대상으로 통근시간의 가치를 산출해 주거비 결정 요인모형의 주요 변수인 통근기회비용을 계산하였고, 이를 통해 통근비용과 임차료 간의 상쇄관계를 분석하였다. 연구자에 따르면 신도시의 경우 단핵도시 모형의 가정과 달리 직장은 도시 외부에 분포하고 거주지는 도시 내부에 위치한 경우가 많은데, 이러한 공간 특성을 가진 신도시에서는 통행비용과 주거비용 간의 상쇄관계가 존재하지 않는다는 점을 검증하였다. 상쇄효과가 나타나지 않았다는 것은 신도시 내에 밀집된 도심이 존재하지 않는다는 것을 의미하며, 이는 결국 신도시의 자족수준이 낮다는 것으로 해석할 수 있다고 설명했다.

박미선 외(2018)는 수도권을 대상으로 지역별, 주택유형별로 교통비를 포함한 주거비 부담을 산정하였다. 수도권 주거비 부담은 월소득 대비 19.6% 수준이었으나 교통비를 포함하는 경우 23.3%까지 상승하는 것을 확인하였다. 소득수준별로는 소득수준이 낮을수록 주거비 부담수준이 높았으며, 가구소득이 300만원 미만인 경우 주거교통부담 수준은 30%에 육박한다고 밝혔다.

제 2 절 선행연구 검토 및 연구의 차별성

1. 통근행태 및 통근거리 결정요인

통근행태 및 통근거리는 주거비용 외에도 지역 및 개인적 특성에 의해 영향을 받는다. 송재민(2021)은 정량 분석 기반의 선행연구를 통합적으로 분석하기 위해 메타분석을 실시하여 우리나라를 대상으로 도시형태 특성이 통행에 미치는 영향을 검토하였다. 분석 결과, 밀도, 혼합도, 대중교통 접근성, 교차로밀도 등의 변수는 다수의 연구에서 일반적인 경향성을 나타내고 있었다. 또한, 국내의 경우 도시형태 특성 중 인구밀도가 승용차 사용 및 대중교통, 비동력 통행에 가장 큰 영향을 미치는 것으로 나타났다.

장재민 & 김태형(2016)은 서울시 25개 자치구를 대상으로 10년 간의 통근시간 변화에 미치는 영향변수를 도출하였다. 아파트 시세차익과 종사자수는 통근시간 변화량에 영향을 미치는 변수로 도출되었으며, 이는 해당지역에 종사자수가 늘어 통근시간이 감소하여 통근시간의 시간가치가 아파트 시세에 반영된 것으로 해석하고 있다. 이와 같은 결과는 통근시간과 아파트 가격의 연관성을 시사하고 있다.

이번송(1998)은 소득, 주거지 소유 형태 등을 포함한 통근자의 특성, 거주지 이동, 통근목적지와 주거지의 특성이 통근거리 결정에 어떠한 영향을 미치는지를 연구하였다. 연구 결과, 고소득자가 원거리 통근을 하는 것으로 나타났으며, 도심 및 부도심에 속하는 서울 중심지에 거주하는 통근자의 통근거리가 비중심지에 거주하는 통근자에 비해 단거리임을 확인하였다. 또한, 서울시 내에서 이주한 서울 거주 통근자는 이주 후에 오히려 통근거리가 멀어지는 것으로 나타났다. 경기도 등 신도시로의 이주는 자가소유의 증대와 주택의 질 향상에는 도움이 되었으나, 통근거리를 현저히 상승시켰다.

전명진 & 이지현(2020)은 수도권을 대상으로 통근자의 직장 접근성과 주택 접근성이 통근시간에 미치는 영향을 분석하였다. 분석 결과, 서울에서 일

하는 통근자의 경우 실제 거주하는 주택 유형과 동일한 주택에 대한 접근성이 높음에도 직장에서 먼 곳에 주거지를 결정하는 것으로 나타났다. 반면, 경기 및 인천의 교외지역에서 일하는 통근자의 경우 직장 주변의 주택 접근성이 높으면 통근시간이 짧아지는 것으로 나타났다. 이와 같은 결과는 거주지와 직장지의 특성이 통근시간에 영향을 미칠 수 있으며, 특히 직장지의 높은 주택 가격은 실질적인 주택 접근성을 저해하는 것을 의미한다.

Zheng et al.(2022)은 도시의 건조 환경이 유입통근거리(work-based)와 유출통근거리(home-based)에 미치는 영향을 분석하였다. 분석 결과, 도시 건조 환경은 유입통근거리와 유출통근거리에 서로 다른 영향을 미치고 있는 것으로 나타났는데, 거주 인구, 오락 시설, 복합 개발은 유출통근거리를 증가시키는 요인이었던. 반면, 노동 인구, 교통 시설은 유출통근거리는 감소시켰으나 유입통근거리는 증가시키는 것으로 나타났다.

한편, 다수의 선행연구에서 통근행태와 도시의 건조 환경 간 비선형적 관계도 입증되었다. Liu et al.(2021)은 출발지와 도착지의 모든 건조 환경 변수는 통근과 쇼핑 목적의 실제 통행에서 비선형적 관계를 지니고 있음을 확인했다. Ding et al.(2018a)과 Ding et al.(2018b)의 연구에서도 연구 대상지는 워싱턴과 오슬로로 상이하나 두 연구 모두에서 건조환경이 통근 수단 및 통근 거리에 미치는 영향이 비선형적임을 보이고 있다.

2. 소득계층별 통근행태 및 통근거리 결정요인

이민주 & 박인권(2016)은 소득에 따른 직주불일치 현상을 확인하기 위해 개인 특성과 통근목적지의 지역 특성이 통근시간에 미치는 영향을 위계선형 모형을 활용하여 분석하였다. 분석 결과, 고소득층보다 저소득층의 통근시간이 길게 나타났으며, 주택가격이 높은 지역일수록 그 지역으로 통근하는 저소득층과 고소득층의 통근시간 격차가 커짐을 분석하였다. 이와 같은 연구 결과는 소득에 따라 통근행태가 실제로 다르게 나타나고 있으며, 주택 가격

이 통근시간에 미치는 영향이 소득계층에 따라 다름을 시사하고 있다.

하재현(2020)은 미국 대도시권 지역을 대상으로 도시공간구조가 소득수준별 통근거리에 미치는 영향을 연구하여, 고소득 통근자의 경우 자동차 의존도가 높아 긴 통행거리를 통근하고, 직장지나 주거지를 선택하는데 있어 입지 이외에 다양한 요인을 고려함을 보이고 있다. 그에 반해 중소득 이하의 통근자는 대중교통 수단에 대한 의존도가 높아, 도심 집중화 정도가 높을수록, 대중교통 결절지 주변으로 주거지를 결정하는 경향이 파악되었다. 본 연구의 결과는 도시공간구조 특성이 통근거리에 미치는 영향이 소득수준에 따라 차별적이라는 점을 실증하고 있다.

김희철 & 안건혁(2010)은 고밀개발, 용도혼합, 직주근접 등과 같은 압축도시 계획요소들이 통행거리에 미치는 영향이 소득계층별로 어떻게 차별적으로 나타나는지 분석하였다. 분석결과, 저소득계층에서는 밀도, 용도혼합, 고용접근성 변수가 통근거리 감소에 영향을 미치지만, 고소득계층에서는 고용접근성 변수만이 통근거리 감소에 영향을 미쳐 소득계층에 따라 압축도시 계획요소의 통근거리 감소효과의 차이가 존재함을 입증했다. 특히, 고용접근성의 통근거리 감소효과는 저소득계층보다 고소득계층에서 크게 도출되었다. 주거지의 고용접근성이 증가한다는 것은 주거지가 고용 중심지에 가까워지면서 통근거리가 감소하게 되는 것을 의미하는데, 이러한 관계를 주거비용, 통행비용, 개인의 시간가치 사이의 교환관계(trade-off)로 해석했다. 고용접근성의 통근거리 감소 효과가 고소득층에서 더욱 크게 나타났다는 것은 시간가치가 높은 개인일수록 고용 중심지 근처에 거주하여 주로 통행시간이나 통행거리로 나타나는 통행비용을 줄이려는 행동에서 기인한다고 이야기했다. 더 나아가 이는 고소득층이 고용중심지와 가까운 주거입지 선정에 있어서 보다 유리함을 시사한다.

3. 연구의 차별성

선행연구를 종합해보면, 통근비용과 주거비용의 상쇄가 명확하게 나타나지 않음을 확인할 수 있어, 통근비용으로 대변될 수 있는 통근시간과 주거비용과의 관계를 살펴볼 필요가 있으며, 또한 소득계층별 차등적 영향에 대한 심도깊은 연구가 필요함을 확인할 수 있다. 실제로 소득은 통근행태에 중요한 영향을 미치고 있었는데, 대체적으로 고소득 근로자가 저소득 근로자에 비해 통근시간 혹은 통근거리가 짧게 나타났다. 한편, 일부 연구에서는 고소득 근로자의 경우 직장의 위치로 인해 통근시간의 감소보다 임대료의 상승이 더욱 가파를 수 있을 수 있음을 확인했다. 실제로 전명진 & 이지현(2020)은 서울의 높은 주택가격은 오히려 서울에서 일하는 통근자가 직장에서 먼 곳을 주거지로 선택하는 경향이 있다고 밝혔다. 또한, 고소득 통근자가 직장지나 주거지를 선택하는데 있어 고려하는 요인이 더욱 많아 통근시간 및 거리가 오히려 늘어났을 가능성도 있다.

이와 같은 배경에서 본 연구는 다음과 같은 점에서 선행연구와 차별된다.

첫째, 통근시간 및 거리에 영향을 미치는 요인과 통근비용과 주거비용의 관계를 분석한 연구는 다수 있었으나, 소득계층별로 나누어 주거비용을 포함한 주거지 특성이 통근시간 및 거리에 미치는 영향력의 차이를 실증적으로 분석한 연구는 매우 부족했다. 선행연구에 따르면 소득은 통근시간 및 거리 결정에 중요한 영향을 미치며, 통근 목적지 지역 특성, 특히 통근 목적지 주거 비용은 통근시간, 통근거리와 밀접한 관계가 있는 것으로 나타났다. 이는 통근시간 및 거리 결정에 있어 소득계층별로 주거비용의 영향이 있음을 시사하고 있으나, 선행연구에서는 소득계층별로 영향을 미치는 요인이 어떻게 달라지는지, 또한 목적지와 주거지 간의 주거비 차이가 어떠한 영향을 미치는지에 대한 분석은 미비하다. 따라서, 소득계층을 구분하여 통근시간 및 거리에 미치는 영향 요인 차이를 식별하는 것은 직주근접에 대한 정책적 시사점을 도출하는 데에 매우 중요한 연구 주제이다.

둘째, 대부분의 선행연구에서는 주택가격과 통근시간 및 통근거리의 관계

를 전통적 회귀모형으로 분석해 왔다. 이는 주택가격과 통근시간 간의 선형 관계를 가정하는 것으로 요인들 간의 비선형 관계를 파악할 수 없다는 한계를 가진다. 본 연구에서는 기계학습 기법 중 하나인 XGBoost를 활용하여 주택가격과 통근시간 간 비선형적 관계를 파악하고 나아가 주택가격 및 거주지 특성이 통근시간에 영향을 미치는 임계치를 확인하고자 한다.

제 3 장 자료 및 분석방법

제 1 절 분석자료 및 표본

본 연구에서는 2016년 가구통행실태조사 응답자료 중, 연구의 공간적 범위 내 거주지와 직장지가 위치하며, 통근 목적의 출발지가 거주지이며, 도착지가 직장인 응답자만을 추출하였다. 결측치를 제외하고 총 82,416건의 자료를 분석하였다.

소득계층은 분석을 보다 용이하게 하기 위해 유사한 선행연구를 바탕으로 기존의 가구통행실태조사에서의 6개의 구분을 3개 계층으로 다시 구분하였다(오지예 외, 2019; 이혜승·이희연, 2009; 김호기, 유상균, & 정창무, 2010). 가구 전체의 월 소득 200만원 미만의 응답자는 저소득층, 200만원 이상 500만원 미만의 경우 중소득층, 500만원 이상의 응답자는 고소득층으로 분류하여 분석하였다.

분석자료 중 가구 전체의 월 소득이 200만원 미만인 저소득층 응답자는 9,435건이었으며, 그 중 서울 거주자는 약 39.70%, 경기 거주자는 약 31.75%, 인천 거주자는 약 28.54%였다. 가구 전체의 월 소득이 200만원 이상 500만원 미만인 중소득층 응답자는 49,731건이었으며, 그 중 서울 거주자는 약 49.41%, 경기 거주자는 약 30.06%, 인천 거주자는 약 20.54%를 차지했다. 마지막으로, 가구 전체의 월 소득이 500만원 이상인 고소득층 응답자는 23,250건이었으며, 이 중 서울 거주자는 50.21%, 경기 거주자는 32.82%, 인천 거주자는 16.97%를 차지하고 있었다.

제 2 절 연구가설 및 연구방법

이론과 선행연구 고찰을 바탕으로, 본 연구가 검증하고자 하는 가설은 다음과 같다.

[가설] 소득계층별로 통근시간 및 통근거리에 미치는 영향 요인 및 관계는 상이할 것이다.

선행연구를 통해 고소득층은 상대적으로 주택가격이 높은 고용중심지와 가까운 주거입지 선정에 있어서 보다 유리한 것을 확인할 수 있었다. 박미선 외(2018)는 소득수준이 낮을수록 주거비 부담수준이 높았으며, 가구소득이 300만원 미만인 경우 주거교통부담 수준은 30%에 육박한다고 밝혔다. 이는 저소득층의 월소득 대비 주거부담이 고소득층에 비해 높다는 것을 의미하며, 이는 고소득층보다 저소득층이 부담이 큰 주거비용을 포기하고 긴 통근거리를 감수할 가능성이 높음을 의미한다. 또는, 고소득 통근자가 직장지나 주거지 선택에 고려하는 요인이 더욱 많아 주택가격보다 중요한 도시 특성 요인이 존재할 수도 있다. 만약 이와 같을 경우, 저소득층보다 고소득층의 통근거리가 증가하겠지만, 주택가격이 다른 변수에 비해 중요한 요인으로 나타날 가능성은 낮을 것이다. 따라서 저소득층은 통근거리 결정에 있어 주택가격이 다른 변수에 비해 큰 영향을 미칠 것이라고 예상할 수 있다.

그러나 가설과 반대로 저소득층보다 고소득층에서 주거비용이 통근거리에 미치는 영향력이 더욱 크게 나타날 수도 있다. 이번송(1998)은 고소득자가 원거리 통근을 하고 있다고 밝혔으며, Park & Quercia (2015)는 미국 고소득 근로자는 통근시간의 감소보다 임대료의 상승이 더욱 가파른 것을 확인하였다. 이는 고소득층의 직장이 고용중심지, 즉 지대가 더욱 비싼 지역에 위치할 가능성이 있기 때문이다. 이로 인해 오히려 고소득층의 통근거리 결정에 있어 주택가격이 저소득층에 비해 더욱 큰 영향을 미치고 있을 가능성도 배제할 수 없다.

주택가격 이외에도 거주지 특성 등 소득계층별로 통근시간 및 통근거리 예측에 있어 중요하게 작용하는 변수는 다르게 나타날 수 있다. 또한, 같은

변수라도 통근시간 및 통근거리에 미치는 영향관계가 다를 것이라고 예상한다.

위 가설을 검증하기 위해 기계학습 기법 중 하나인 Extreme Gradient Boost (XGBoost)를 활용하였다. 머신러닝을 활용할 경우, 전통적인 통계학으로 분석하기 힘든 특정 의도 없이 수집되거나 다양한 형태를 가진 자료를 분석할 수 있으며, 비선형을 보이는 변수 사이의 복잡한 상호작용을 파악할 수 있게 된다(유진은, 2017; 박준상·이수기, 2022). 비선형적 관계를 파악하는 것은 특정 변수가 영향력을 미치는 임계치를 파악하여 정책 도입을 보다 효과적으로 제안할 수 있다는데 의의가 있다.

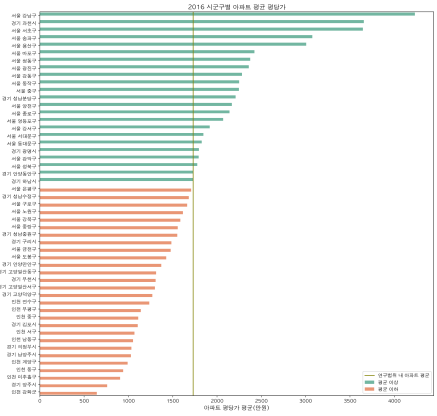
Gradient Boosting Decision Tree (GBDT)는 앙상블 기법으로서 먼저 다수의 단일 의사 결정 트리를 반복적으로 만든 다음, 모든 트리의 결과를 결합하여 최종 모델을 구축한다(Liu et al., 2021). 또한, 이는 경사 하강법을 활용하는 모형으로, 손실함수를 정량화하고 이러한 손실함수를 파라미터로 미분하여 기울기 값을 산출한 뒤, 이 손실(loss) 값이 작아지는 방향으로 파라미터를 움직인다(박준상·이수기, 2022). Gradient Boosting 모델은 다중공선성 문제를 보다 효율적으로 다룰 수 있으며, 강력한 예측력과 중요한 변수를 식별할 수 있는 능력의 신뢰도가 높아 교통 분야에서 활발히 사용되고 있다(Ma et al., 2017; Zhang and Haghani, 2015; Ding et al., 2016; Ding et al., 2019). XGBoost는 Gradient Boosting Decision Tree를 효율적으로 구현하기 위한 시스템으로, 속도를 크게 향상시킬 수 있도록 분산 컴퓨팅을 사용하여 GBDT에 비해 일반적으로 10배 이상 처리 속도가 빠르다(Chen and Guestrin, 2016; Liu et al., 2021).

따라서 통근거리에 영향을 미치는 요인들의 비선형적 관계를 고려할 수 있는 XGBoost를 적용하여 분석하였다. 또한, 본 연구에서는 과적합(overfitting)을 방지하고, 모델의 정확도를 높이기 위해 Optuna 프레임워크를 사용하여 하이퍼파라미터를 최적화하였다.

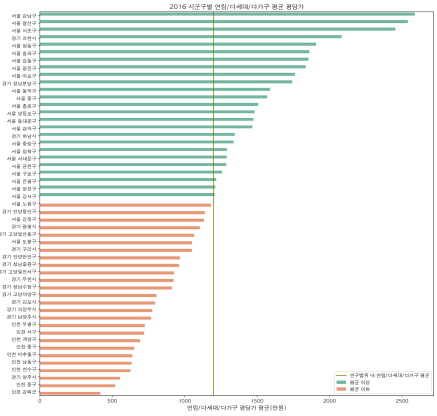
제 3 절 변수설정

주요 변수는 다음과 같이 구축하였다.

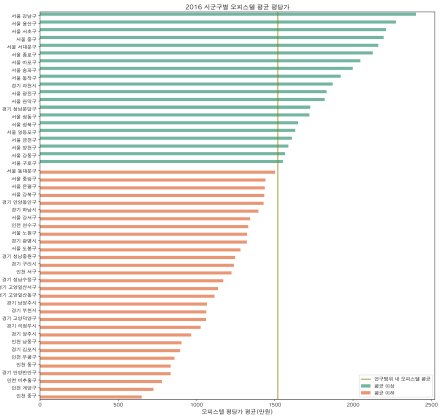
주택가격은 2016년 국토교통부 실거래가 자료를 바탕으로 측정하였다. 2016년 연구 범위 내 아파트 241,664건, 오피스텔 23,482건, 연립, 다세대 및 다가구주택 114,824건, 단독주택 13,663건의 거래가 있었다. 모든 거래건의 주소를 좌표로 변환하여 행정동별, 주택유형별 평균 평당(3.3㎡) 거래가격을 산출하였다. 행정동별 주택가격을 변수로 설정하기 앞서, 연구범위 내 시군구별, 주택유형별 평균 평당 거래가격을 살펴보았다.



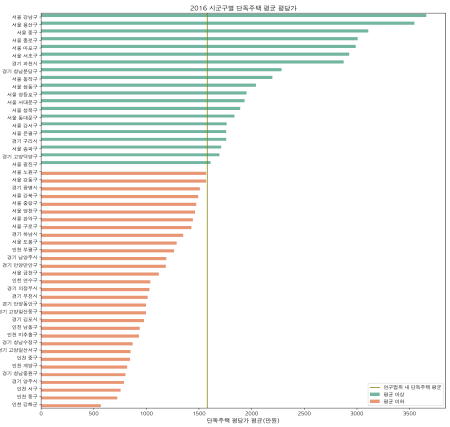
[그림 2] 시군구별 아파트



[그림 3] 시군구별 연립/다세대/다가구



[그림 4] 시군구별 오피스텔



[그림 5] 시군구별 단독주택

연구범위 내 아파트의 평균 평당 거래가격은 약 1,728만원, 연립, 다세대 및 다가구 주택의 평균 평당 거래가격은 약 1,198만원, 오피스텔의 평균 평당 거래가격은 약 1,519만원, 단독주택의 평균 평당 거래가격은 약 1,577만원으로 나타났다. 아파트의 평균 주택가격이 가장 높았으며, 단독주택, 오피스텔, 연립, 다세대 및 다가구 주택 순으로 평균 평당 거래가격이 낮아지고 있었다. 서울시 강남구, 서초구, 송파구, 용산구, 마포구, 성동구 등에서 상대적으로 높은 평균 아파트 거래가격을 나타냈으며, 경기도 과천시, 성남시 분당구 등에서도 평균 아파트 거래가격이 상위권에 위치한 것을 살펴볼 수 있었다. 연립, 다세대 및 다가구 주택과 오피스텔의 경우 아파트보다 거래가격은 낮았지만, 대체적으로 아파트와 비슷한 양상의 순위를 나타내고 있었다. 단, 오피스텔은 서울 중구, 서울 종로구에서 아파트에 비해 상대적으로 거래가격 상위권에 위치한 것으로 나타났으며, 이는 업무중심지와 인접한 지리적 특성에 의한 것이라고 파악할 수 있다. 단독주택은 서울시 강남구, 용산구, 중구, 종로구 순으로 높은 거래가격을 형성하고 있었다.

행정동별, 주택유형별 거래가격을 변수로서 활용하기 이전에, 연구범위 내 시군구별, 주택유형별 거래가격을 파악하여 전반적인 주택가격 양상을 확인하였다. 이를 토대로 분석에서는 행정동별, 주택유형별 거래가격을 변수로서 활용하였으며, 개별 응답자의 거주지 행정동과 주택유형에 맞추어 주택가격을 반영하여 연구의 정확성을 높이고자 하였다.

주택가격변화율의 경우, 2011년 대비 2016년에 아파트 매매가격이 얼마나 변화하였는지를 측정하였으며, 전희정 & 정수영(2021)의 평균주택가격 변화 측정 방법을 사용하였다. 해당 연구에서는 물가 차이 및 시기별 인플레이션으로 인한 주택가격 상승을 통제하고자 2010년과 2019년에 각 읍면동이 속한 7개 광역권의 ㎡당 평균주택가격에 대한 행정동의 ㎡당 평균주택가격의 상대적 비율을 계산하였다. 즉, 수도권 대비 행정동별 평균 주택가격을 기준으로, 두 시기의 상대적 비율의 변화를 평균주택가격 변화로 측정한 것이다. 예를 들어, 서울시 한 행정동의 2011년 수도권 대비 상대적 평균주택가격이 100%, 2016년 수도권 대비 상대적 평균주택가격이 120%일 경우

해당 근린은 상승한 것으로 설명할 수 있다. 마지막으로, 참고한 연구와 동일하게 2011년, 2016년의 상대적 비율의 변화에 로그를 취하여 분석하였다. 이를 식으로 나타내면 다음과 같다.

$$\log(y_{ij,t_2}/y_{ij,t_1})$$

y는 수도권j의 3.3㎡당 평균 아파트 매매가격에 대한 읍면동i의 3.3㎡당 평균 아파트 매매가격의 상대적 가격이며 t2는 2016년, t1은 2011년을 의미한다.

토지이용균형은 엔트로피로 측정하였으며, 국토교통부에서 제공하는 토지특성 자료를 활용하였다. 노태욱 & 강창덕(2009)의 엔트로피 산출식을 참고하여 4개 토지이용(주거, 상업, 주상용, 공업용)으로 엔트로피 값을 아래와 같이 구하였다.

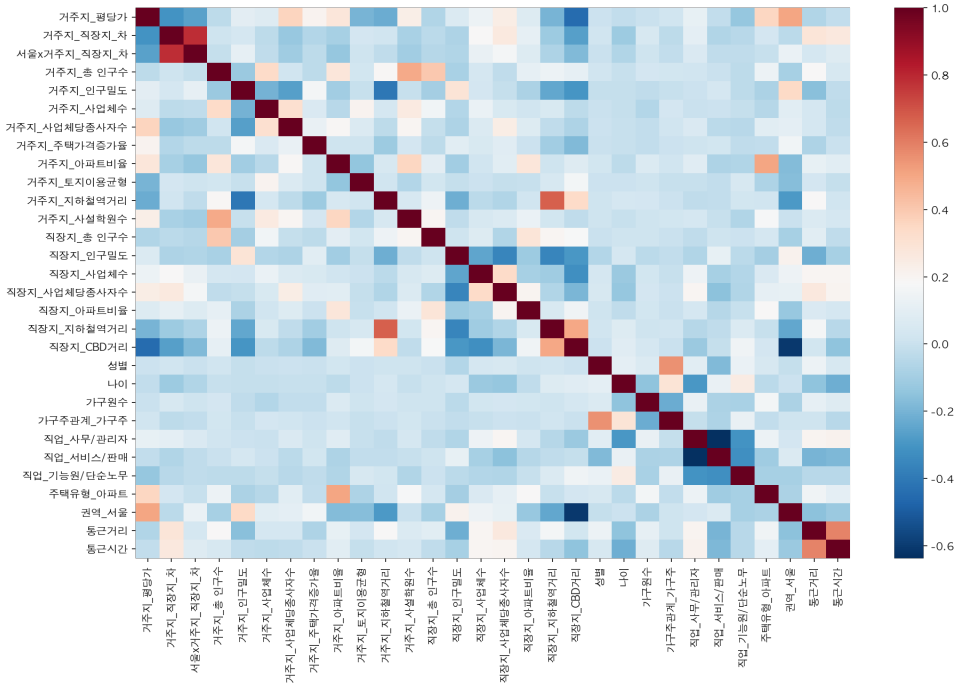
$$-\sum_{t=1}^{4}(p_i)\ln(p_i)/\ln(s)$$

P_i는 주거, 상업, 주상용, 공업용 건물의 연면적을 모두 더한 후 각 용도의 건물 연면적을 나눈 값이며, s는 토지이용의 개수로 4개가 된다(노태욱·강창덕, 2009). 값이 0에 가까울수록 단일 토지이용이 높은 반면, 1에 근접할수록 다양한 토지이용이 균등한 면적으로 분포하고 있음을 의미한다(김태형·고준호, 2016).

종속변수인 통근시간은 2016년 가구통행실태조사에 기재된 응답자의 출발시각과 도착시각을 바탕으로 산정하였다. 통근거리는 2016년 도로망을 바탕으로 개별 응답자의 출발지와 도착지의 도로 상 최단거리로 측정하였다. 직선거리가 아닌 도로망을 기준으로 측정하여, 실제 통근거리와 유사하게 측정하고자 하였다. 통근시간은 2016년 가구통행실태조사에 기재된 응답자의 출발시각과 도착시각을 바탕으로 산정하였다.

범주형 변수는 각 범주를 더미변수화하였으며, 모든 변수를 선택하지 않고 주요 범주형 변수만 채택하여 더미변수로 변환하였다. 다음으로, 독립변수 간 상관성이 지나치게 높은 변수는 최대한 제거하여 분석의 정확도를 높였다. 기계학습 기법은 다중공선성 문제를 다룰 수 있다는데 큰 강점이 있지

만, 각 독립변수와 종속변수와의 관계를 살펴볼 때에는 독립변수 간 상관성이 어느 정도 영향을 미칠 수 있기 때문이다. 따라서 분석에 최종적으로 사용한 변수는 [표 2]와 같이 선정하였으며, 변수 간 상관성을 그림으로 나타내면 [그림 6]과 같다.



[그림 6] 변수 간 상관성

[표 2] 분석에 사용된 변수

구분		데이터			출처
종속변수	통근시간	응답자의 통근시간 (분)			가구통행실태조사
	통근거리	도로망에 기반한 응답자의 통근거리 (km)			직접 구축
행정동별·주택유형별 주택가격	거주지 주택가격	응답자의 주택유형별, 거주지 행정동별 평균 평당 실거래가			국토교통부
	거주지와 직장지의 주택가격 차	(직장지의 주택가격) - (거주지의 주택가격)			
	interaction term	서울 dummy * {(직장지의 주택가격) - (거주지의 주택가격)}			
거주지 지역변수	인구	거주지의 총 인구 수			KOSIS
		거주지의 인구밀도			
	경제	거주지의 사업체수			
		거주지의 사업체당 종사자수			
		거주지의 주택가격증가율 (2011년 대비 2016년 아파트가격 증가율)			국토교통부
	물리적 환경	거주지의 아파트 비율			KOSIS
		거주지의 토지이용균형			국토교통부
		거주지 행정동 중심에서 가장 가까운 지하철역까지의 거리 (km)			직접 구축
		거주지의 사설학원 수			KEDI
직장지 지역변수	인구	직장지의 총 인구 수			KOSIS
		직장지의 인구밀도			
	경제	직장지의 사업체수			
		직장지의 사업체당 종사자수			
	물리적 환경	직장지의 아파트 비율			
		직장지 행정동 중심에서 가장 가까운 지하철역까지의 거리 (km)			직접 구축
		직장지 행정동 중심에서 가장 가까운 CBD(서울 3도심)로부터의 거리 (km)			
개인변수	성별	0: 여자	1: 남자		2016 가구통행 실태조사
	나이	응답자의 나이			
	가구원수	응답자의 가구원수			
	관계	1: 가구주	0: 기타 (부모, 자녀, 기타)		
	직업	관리자/사무종사자	1: 예	0: 아니오	
		서비스/판매종사자	1: 예	0: 아니오	
		기능원/단순노무종사자	1: 예	0: 아니오	
	주택유형	아파트	1: 예	0: 아니오	
	거주 권역	서울	1: 예	0: 아니오	

제 4 장 소득계층별 통근거리 영향요인 분석

제 1 절 기술통계

분석 결과에 앞서, 소득계층별 기술 통계량을 [표 3]에 요약하였다. 통근시간을 살펴보면, 저소득층에서 고소득층으로 갈수록 평균 통근시간이 증가하는 것을 확인할 수 있다. 거주지 주택가격 또한 고소득층으로 갈수록 커지고 있으나, 거주지와 직장지의 주택가격 차의 평균은 중소득층에서 가장 크게 나타났으며, 저소득층에서 가장 작게 나타났다.

거주지 지역변수의 경우, 사업체수, 아파트 비율, 사설학원 수는 고소득층으로 갈수록 증가하고 있었다. 반면, 행정동 중심지로부터 가까운 지하철역까지의 거리는 고소득층으로 갈수록 줄어들고 있었다. 이를 토대로 고소득층의 주거지역이 보다 주거환경이 정비되어 있는 지역일 가능성이 높음을 짐작할 수 있다. 거주지의 인구밀도는 중소득층에서 가장 높았으며, 저소득층에서 다소 낮게 나타난 것을 확인했다.

직장지 지역변수를 확인하면, 고소득층으로 갈수록 직장지의 인구밀도, 사업체수 및 사업체당 종사자수, 행정동 중심지로부터 가까운 지하철역 및 CBD까지의 거리는 고소득층으로 갈수록 줄어들고 있었다. 이는 고소득층의 직장지가 보다 서울 3도심에 밀집해있음을 의미한다.

개인적 특성 변수를 살펴보면, 저소득층으로 갈수록 서비스 및 판매종사자, 기능원 및 단순노무종사자, 농림어업 및 숙련종사자의 비율은 증가하고 있었다. 이와 반대로 전문가 및 관련종사자, 관리자 및 사무종사자의 비율은 고소득층으로 갈수록 증가하는 것으로 나타났다. 주택유형의 경우, 고소득층으로 갈수록 아파트 거주비율은 현저히 증가하는 반면, 연립, 다세대, 다가구주택 거주비율은 감소하고 있었다.

통근수단은 분석 변수로서 활용하지는 않았으나, 소득계층별 통근 행태에 대한 참고자료로서 기초통계량을 살펴보았다. 저소득층은 대중교통 이용비율

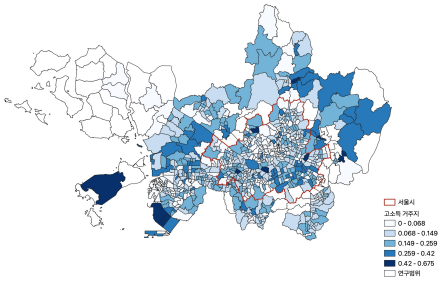
이 가장 높았고, 다음으로 도보 및 자전거, 승용차 및 오토바이 순으로 이용 비율이 낮아지고 있었으나, 고소득층으로 갈수록 승용차 및 오토바이 이용비율이 현저히 증가한 반면, 도보 및 자전거 이용률은 눈에 띄게 감소한 것으로 나타났다. 서울시의 교통권역 내 수도권에 거주하는 저소득층, 중소득층, 고소득층 모두 대중교통 이용 비율이 가장 높았다.

이에 더해, 소득계층별 응답자의 거주지 및 직장지 분포를 아래 지도로 나타내었다. 각 행정동에서 거주 또는 일하는 응답자의 수를 각 소득계층별 전체 응답자 수로 나누어 100을 곱한 값을 기준으로 범례를 표시했으며, 범례의 구간은 모든 소득계층에서 동일하다.

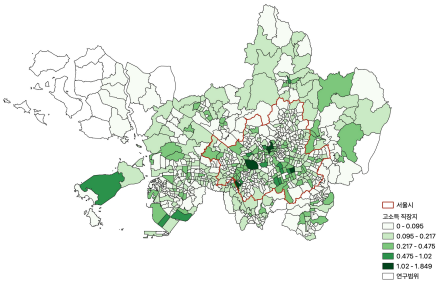
먼저, 고소득층의 거주지 분포를 살펴보면, 서울시 강남구, 서초구, 송파구, 광진구, 여의도 및 인근 지역, 이외 서울시와 인접한 인천광역시 계양구, 구리시에 집중적으로 분포되어 있었으며, 남양주시와 인천광역시 운서동, 송도동 또한 응답자의 거주 비율이 높았다. 이와 함께 고소득층의 직장지 분포를 살펴보면, 중구 및 종로구, 여의도, 강남구 및 서초구, 즉 서울 3도심에 집중적으로 포진하고 있는 것을 확인할 수 있었으며, 대부분의 직장지 분포가 서울시 및 인접지역에 집중되어 있었다. 이외 인천광역시 운서동, 송도동, 논현고잔동의 분포도 또한 높았다.

다음으로 중소득층의 거주지 분포를 확인하였다. 서울시와 인접한 경기도 구리시, 부천시, 남양주시, 인천광역시 계양구에 높은 분포도를 보이고 있는 것을 확인할 수 있었으며, 서울 3도심에 인접한 지역보다는 노원구 등 서울 외곽 주거집중지역에 보다 집중되어 있었다. 중소득층의 직장지 분포를 확인해보면, 고소득층과 마찬가지로 종로구, 여의도, 강남구에 직장이 다수 위치해 있었으나, 그 집중도가 고소득층에 비해서는 보다 낮았으며, 남양주시를 포함한 서울 외곽지역의 집중도가 보다 높아진 것으로 나타났다.

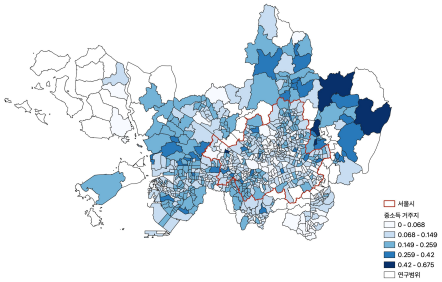
마지막으로 저소득층의 거주지 분포를 확인하였다. 서울 도심 인근에는 거주지 분포가 거의 없는 반면, 인천광역시 및 남양주시, 양주시, 구리시, 의정부시의 분포가 다른 소득계층에 비해 크게 증가한 것으로 드러났다. 서울시 내에서는 금천구 및 강동구에서 다른 소득계층에 비해 다소 집중된 분포



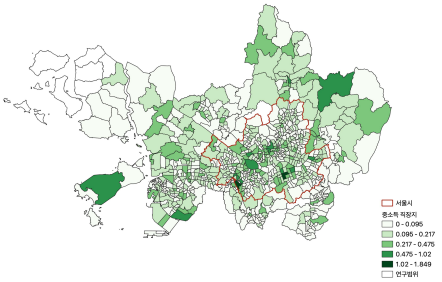
[그림 7] 고소득 응답자 거주지 분포



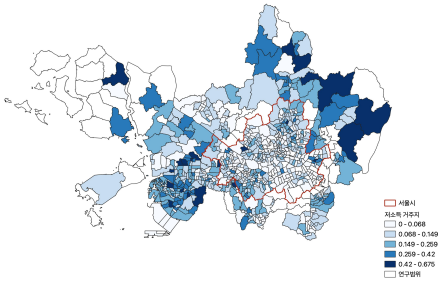
[그림 8] 고소득 응답자 직장지 분포



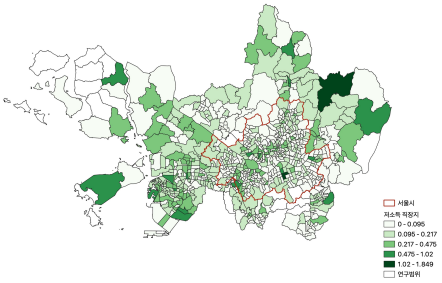
[그림 9] 중소득 응답자 거주지 분포



[그림 10] 중소득 응답자 직장지 분포



[그림 11] 저소득 응답자 거주지 분포



[그림 12] 저소득 응답자 직장지 분포

가 나타났다. 직장지 분포를 살펴보면, 여전히 서울 강남구 및 여의도 지역의 분포가 높긴 했지만, 다른 소득계층에 비하면 그 집중도는 오히려 다소 감소한 것으로 나타났다. 반면, 서울특별시 금천구, 인천광역시 중구, 연수구, 남동구 및 경기도 남양주시, 양주시, 의정부시, 구리시는 다른 소득계층의 직장지 분포에 비해 증가한 것을 확인할 수 있다.

종합해보면, 고소득층으로 갈수록 서울 3도심에 직장지가 집중되는 경향

이 두드러졌으며, 저소득층으로 갈수록 서울 3도심으로부터 떨어져 서울 외곽에 직장지 및 주거지 분포가 증가하고 있었다. 고소득의 경우 직장지는 서울 3도심 및 인근지역에 집중되어 있었으나, 거주지는 서울시 및 인접지역에 보다 고르게 분포해있는 것을 확인하였다. 중소득은 서울 3도심과 이외 경기, 인천 지역으로 직장지 분포가 뻗어나가고 있었으며, 거주지는 서울 외곽 지역에 보다 집중된 경향이 드러났다. 반면, 저소득의 경우 거주지와 직장지의 분포 집중 지역이 유사한 것으로 나타났다.

[표 3] 소득계층별 기초통계량

데이터		저소득층		중소득층		고소득층	
		mean	std.	mean	std.	mean	std.
obs.		9,435		49,731		23,250	
종속변수	통근거리 (km)	6.39	7.44	8.32	8.40	9.21	8.63
	통근시간 (분)	32.88	23.99	37.16	25.14	39.23	25.39
주택가격	거주지 주택가격	1151.10	670.31	1396.81	734.89	1680.08	955.22
	거주지와 직장지의 주택가격 차	117.76	517.47	203.71	704.29	155.22	866.40
	interaction term	49.10	395.92	94.83	561.81	35.07	715.98
거주지 지역변수	총 인구 수	25966.74	14080.93	26939.04	12504.39	27427.84	11558.69
	인구밀도	19264.98	12602.95	20272.47	12685.37	19910.14	12568.58
	사업체수	1807.04	1434.05	1826.69	1482.99	1838.07	1419.61
	사업체당 종사자수	4.67	2.21	4.94	2.53	5.45	3.04
	주택가격증가율	-0.0207	0.04	-0.0172	0.04	-0.0167	0.04
	아파트 비율	55.20	29.38	60.43	29.21	65.64	29.05
	토지이용균형	0.55	0.14	0.54	0.15	0.54	0.15
	지하철역까지의 거리	1.85	2.80	1.58	1.92	1.49	1.59
	사설학원 수	33.34	31.12	38.42	36.48	44.82	41.45

데이터		저소득층		중소득층		고소득층	
		mean, %	std.	mean, %	std.	mean, %	std.
직장지 지역변수	총 인구 수	26026.20	14132.77	25636.76	12421.48	25461.31	11801.49
	인구밀도	17780.75	12324.47	17537.48	11888.85	16993.21	11680.65
	사업체수	2594.42	2532.79	3030.94	3055.67	3322.88	3341.48
	사업체당 종사자수	5.54	3.05	6.30	3.71	7.00	4.18
	아파트 비율	56.08	28.72	57.50	28.72	59.94	28.84
	가장 가까운 지하철역까지의 거리	1.75	2.78	1.44	2.00	1.31	1.62
	가장 가까운 CBD까지의 거리	15.93	10.01	13.33	9.80	12.20	9.68
성별	남자	52.40%		64.11%		58.91%	
	여자	47.60%		35.89%		41.09%	
가구원 정보	나이	52.65	15.23	45.18	11.90	44.84	11.79
	가구원수	2.02	0.97	3.01	1.03	3.58	0.89
	가구주	73.18%		61.04%		47.50%	
	배우자	13.68%		21.17%		26.06%	
	자녀	10.83%		15.52%		23.52%	
	기타	1.58%		1.40%		1.57%	
직업	전문가 및 관련종사자	3.11%		5.61%		9.73%	
	서비스 및 판매종사자	47.64%		40.81%		35.01%	
	관리자 및 사무종사자	15.05%		38.44%		46.30%	
	기능원 및 단순노무종사자	29.33%		12.92%		7.00%	
	농림어업 및 숙련종사자	2.37%		0.63%		0.40%	
	기타	2.50%		1.58%		1.56%	

데이터		저소득층	중소득층	고소득층
		%	%	%
주택유형	아파트	42.05%	59.77%	74.38%
	연립, 다세대, 다가구	34.29%	27.26%	16.58%
	오피스텔	1.96%	1.33%	0.77%
	단독주택	21.71%	11.64%	8.27%
권역	서울	39.70%	49.41%	50.21%
	경기	31.75%	30.06%	32.82%
	인천	28.54%	20.54%	16.97%
통행수단	도보, 자전거	37.22%	21.49%	16.97%
	승용차, 오토바이	20.52%	37.23%	40.54%
	대중교통	40.49%	40.00%	41.56%
	택시	0.61%	0.47%	0.46%
	기타	1.16%	0.81%	0.47%

제 2 절 기계학습 모형

소득계층별 분석에서는 저소득층 9,435건, 중소득층 49,731건 고소득층 23,250건의 응답자료를 활용하였으며, 각 계층의 데이터를 학습 데이터와 평가 데이터로 분류하였다. 학습 표본은 소득계층별 표본 수의 70%, 평가 표본은 30%로 할당해 분할하였다. 기계학습 모형 중 Random Forest, Histogram-based Gradient Boost, Gradient Boost, CATBoost, XGBoost 모형의 평가 결과를 비교한 결과는 [표 4], [표 5]와 같다. 분석모형의 R-Squared, 표준절대오차(Mean Absolute Error), 표준제곱근오차(RMSE)를 통해 비교 평가하였으며, 그 중 모든 소득계층에서 가장 우수한 성능을 보인 XGBoost를 분석 모형으로 선정하였다. 동일한 변수를 사용하여 전통적으로 선형 관계를 가정하는 회귀분석 결과와 비교해보면, 변수 간 비선형 관계를 파악할 수 있는 기계학습 기법 중 하나인 XGBoost의 설명력이 매우 높다는 것을 확인할 수 있다. 이에 더해, [표 6], [표 7]을 통해 소득계층별 통근거리 및 통근시간에 대한 XGBoost 모형의 예측값과 실제값의 분포가 어떠한지를 대략적으로 나타낸 그래프를 첨부하였다. 녹색 점의 x축은 실제값, y축은 예측값이며, 파란 일직선은 x축, y축 모두 실제값으로, 실제 데이터에서 XGBoost 모형의 예측치가 얼마나 멀리 떨어져 있는지를 확인할 수 있다.

[표 4] 통근시간(분) 분석모형 평가 결과

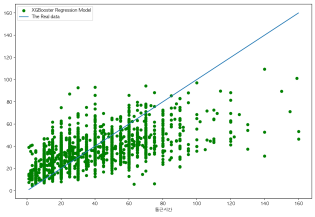
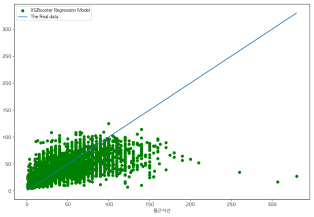
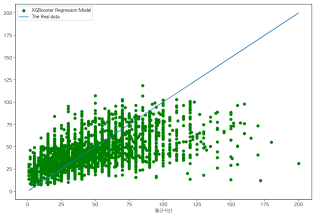
		저소득층	중소득층	고소득층
Linear Regression	R^2	0.180	0.222	0.189
	RMSE	21.502	21.934	23.108
	MAE	16.402	16.671	17.284
Random Forest	R^2	0.355	0.289	0.336
	RMSE	29.069	19.443	20.905
	MAE	13.694	13.871	14.711

Histogram-based Gradient Boost	R^2	0.375	0.385	0.342
	RMSE	18.774	19.502	20.804
	MAE	13.587	14.292	14.905
Gradient Boost	R^2	0.348	0.350	0.312
	RMSE	19.173	20.043	21.273
	MAE	14.114	14.803	15.287
CATBoost	R^2	0.382	0.419	0.366
	RMSE	18.663	18.955	20.429
	MAE	13.414	13.758	14.504
XGBoost	R^2	0.401	0.423	0.395
	RMSE	18.160	18.946	19.816
	MAE	13.283	13.384	14.210

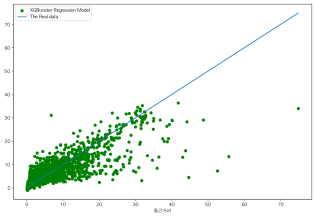
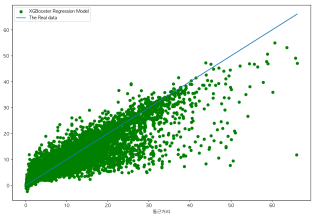
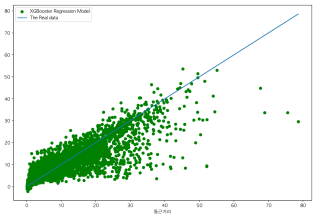
[표 5] 통근거리(km) 분석모형 평가 결과

		저소득층	중소득층	고소득층
Linear Regression	R^2	0.411	0.335	0.292
	RMSE	5.708	6.823	7.354
	MAE	3.961	4.999	5.348
Random Forest	R^2	0.683	0.741	0.691
	RMSE	4.218	4.308	4.812
	MAE	2.213	2.477	2.886
Histogram-based Gradient Boost	R^2	0.688	0.693	0.656
	RMSE	4.182	4.684	5.072
	MAE	2.363	3.066	3.341
Gradient Boost	R^2	0.625	0.605	0.570
	RMSE	4.586	5.316	5.670
	MAE	2.748	3.607	3.899
CATBoost	R^2	0.720	0.753	0.702
	RMSE	3.966	4.202	4.722
	MAE	2.195	2.650	3.018
XGBoost	R^2	0.742	0.819	0.756
	RMSE	3.779	3.561	4.312
	MAE	2.033	1.815	2.362

[표 6] 소득계층별 통근시간에 대한 XGBoost 모형 실제값 및 예측값

저소득층	중소득층	고소득층
		

[표 7] 소득계층별 통근거리에 대한 XGBoost 모형 실제값 및 예측값

저소득층	중소득층	고소득층
		

정확도를 높이고자 주요 매개변수를 [표 8], [표 9]와 같이 Optuna 프레임워크를 통해 조정하였다. max_depth는 의사 결정 트리의 최대 깊이, learning_rate는 학습률, colsample_bytree와 subsample은 트리가 과도하게 복잡하게 생성되지 않도록 과적합을 방지하기 위한 매개변수, alpha와 lambda는 과적합을 방지하기 위한 정규화 매개변수, min_child_weight는 의사 결정 트리에서 가지를 추가할지를 결정하기 위해 필요한 가중치의 총합을 의미한다.

[표 8] 통근시간에 대한 XGBoost 모형에 사용된 매개변수 (Hyper-parameter)

	저소득층	중소득층	고소득층
max_depth	11	12	10
learning_rate	0.0435	0.0362	0.0448
colsample_bytree	0.5236	0.5984	0.4672
subsample	0.6453	0.7810	0.7900
alpha	0.0488	0.0527	3.0302
lambda	0.0616	0.3174	2.0478
min_child_weight	29.1479	20.8974	10.0220

[표 9] 통근거리에 대한 XGBoost 모형에 사용된 매개변수 (Hyper-parameter)

	저소득층	중소득층	고소득층
max_depth	6	10	10
learning_rate	0.0439	0.0336	0.0437
colsample_bytree	0.5554	0.4281	0.5519
subsample	0.7508	0.5839	0.7552
alpha	0.1241	0.1815	0.1714
lambda	2.4505	0.0040	1.7059
min_child_weight	10.5592	15.2687	26.4055

제 3 절 소득계층별 분석 결과 비교

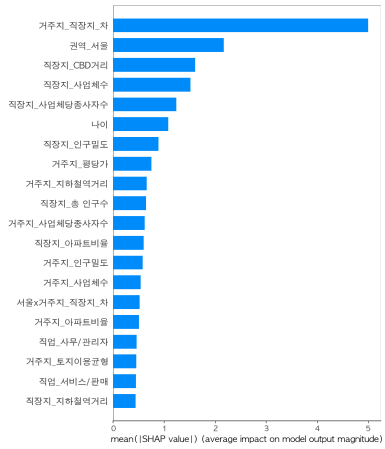
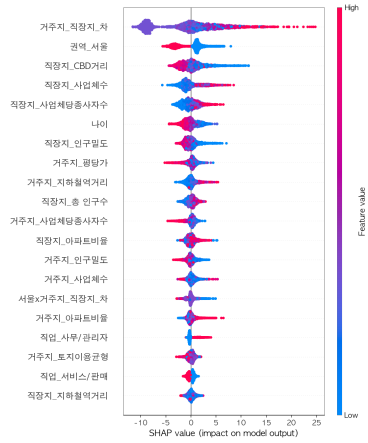
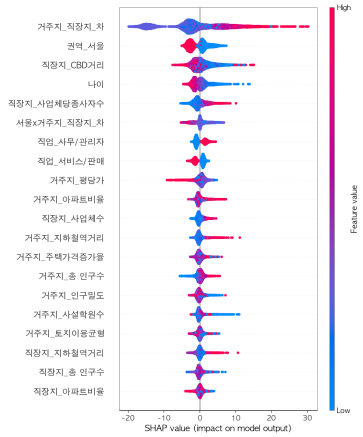
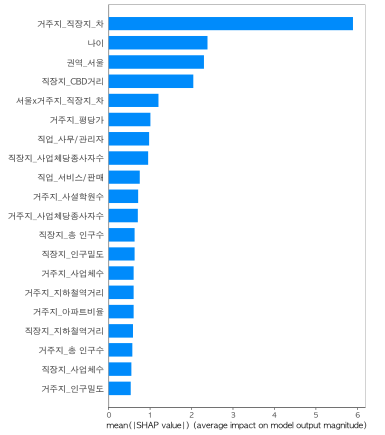
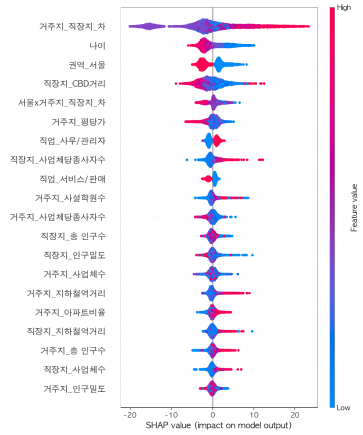
1. 변수의 기여도 (Feature Importance)

주택가격 및 거주지 특성이 소득계층별 통근거리 및 통근시간에 미치는 영향력을 파악하기 위해 XGBoost 분석 모형의 변수별 상대적 기여도(Feature Importance)를 살펴보았다. GBDT 모델에서 의사결정트리를 구축하기 위해 독립 변수는 반복적으로 무작위로 선택되며, 이 때 특정 변수가 선택된 횟수에 따라 상대적 중요도를 측정한다(Breiman et al., 1984; Friedman, 2001). 이에 따라 상대적 기여도가 더 높다는 것은 변수가 성공적인 예측에 기여한다는 것을 의미한다(Zhang and Haghani, 2015).

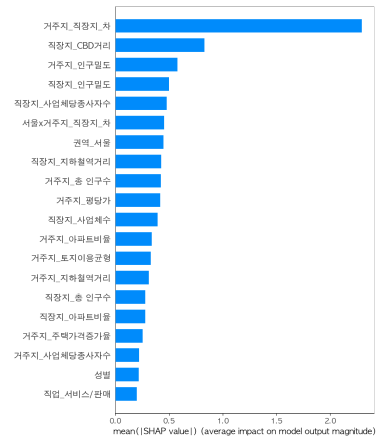
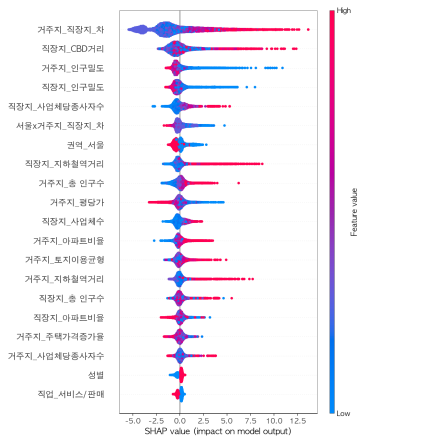
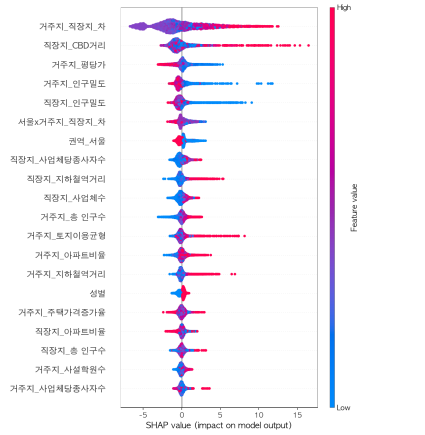
Shapley Value를 통해 변수의 기여도를 파악하였으며, 소득계층별 분석 모형의 Global Shapley Value와 Local Shapley Value를 나타내었다. Global Shapley Value는 각 변수에 대한 Shapley Value를 절댓값으로 나타낸 바 그래프이며, Local Shapley Value는 모든 데이터를 점으로 표현하여 어떤 영향을 나타내고 있는지를 표현한 그래프이다. Local Shapley Value 그래프 내에서 파란색에 가까울수록 변수의 값이 낮고, 빨간색에 가까울수록 변수의 값이 높으며, 0을 기준으로 오른쪽은 양(+)의 관계, 왼쪽은 음(-)의 관계를 나타낸다(박준상·이수기, 2022).

따라서 Shapley Value를 토대로 통근시간 및 통근거리에 있어 변수별 상대적 기여도를 파악하였다. 소득계층별 분석모형의 Global Shapley Value와 Local Shapley Value는 [표 10], [표 11]을 통해 확인할 수 있다. 더불어, Shapley Value에 따라 도출된 변수의 기여도 순위를 소득계층별로 비교하였으며, 통근시간에 대한 분석모형의 변수의 기여도 순위는 [표 12], 통근거리에 대한 분석모형의 변수의 기여도 순위는 [표 13]에서 살펴볼 수 있다.

[표 10] 소득계층별 통근시간에 영향을 미치는 변수의 Global, Local Shapley Value

	Global Shapley Value	Local Shapley Value
저소득층		
중소득층		
고소득층		

[표 11] 소득계층별 통근거리에 영향을 미치는 변수의 Global, Local Shapley Value

	Global Shapley Value	Local Shapley Value
저소득층		
중소득층		
고소득층		

1) 소득계층별 통근시간 영향요인의 변수의 기여도

[표 12] 소득계층별 통근시간에 영향을 미치는 변수의 기여도 Shapley Value 순위

순위	저소득층	중소득층	고소득층
1	거주지와 직장지의 주택가격 차	거주지와 직장지의 주택가격 차	거주지와 직장지의 주택가격 차
2	거주 권역-서울	거주 권역-서울	나이
3	직장지에서 가장 가까운 CBD까지의 거리	직장지에서 가장 가까운 CBD까지의 거리	거주 권역-서울
4	직장지의 사업체수	나이	직장지에서 가장 가까운 CBD까지의 거리
5	직장지의 사업체당 종사자수	직장지의 사업체당 종사자수	(서울 거주)*거주지와 직장지의 주택가격 차
6	나이	(서울 거주)*거주지와 직장지의 주택가격 차	거주지의 주택가격 (평당가)
7	직장지의 인구밀도	직업-사무/관리자	직업-사무/관리자
8	거주지의 주택가격 (평당가)	직업-서비스/판매	직장지의 사업체당 종사자수
9	거주지에서 가장 가까운 지하철역까지의 거리	거주지의 주택가격 (평당가)	직업-서비스/판매
10	직장지의 총 인구수	거주지의 아파트 비율	거주지의 사설학원 수
11	거주지의 사업체당 종사자수	직장지의 사업체수	거주지의 사업체당 종사자수
12	직장지의 아파트비율	거주지에서 가장 가까운 지하철역까지의 거리	직장지의 총 인구수
13	거주지의 인구밀도	거주지의 주택가격증가율	직장지의 인구밀도
14	거주지의 사업체수	거주지의 총 인구수	거주지의 사업체수
15	(서울 거주)*거주지와 직장지의 주택가격 차	거주지의 인구밀도	거주지에서 가장 가까운 지하철역까지의 거리
16	거주지의 아파트비율	거주지의 사설학원 수	거주지의 아파트 비율
17	직업-사무/관리자	거주지의 토지이용균형	직장지에서 가장 가까운 지하철역까지의 거리
18	거주지의 토지이용균형	직장지에서 가장 가까운 지하철역까지의 거리	거주지의 총 인구수
19	직업-서비스/판매	직장지의 총 인구수	직장지의 사업체수
20	직장지에서 가장 가까운 지하철역까지의 거리	직장지의 아파트비율	거주지의 인구밀도

소득계층별 통근시간에 대한 변수의 상대적 기여도를 살펴보면, 모든 소득계층에서 거주지와 직장지 간 주택가격의 차이가 통근시간 결정에 있어 미

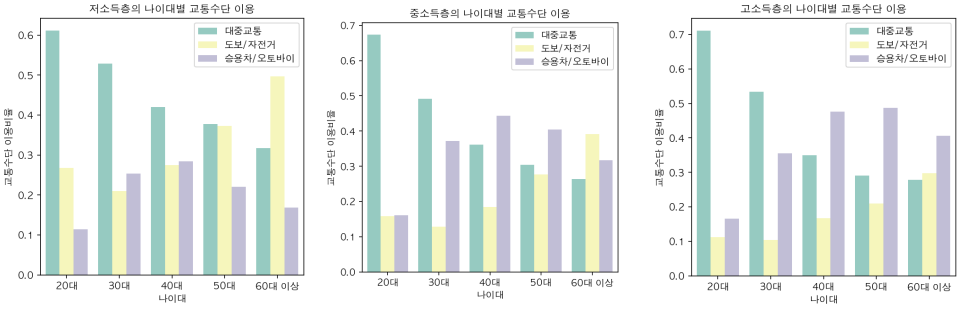
치는 영향력이 가장 큰 것으로 나타난다. 그러나, 거주 권역이 서울 여부인지를 나타내는 더미변수와의 상호작용 변수(interaction term)를 살펴보면 저소득층에서는 15번째 중요 변수로, 변수의 기여도가 미미한 반면, 중, 고소득층으로 갈수록 통근시간 결정에 미치는 영향력이 크게 나타났다. 이는 서울에 거주하는 저소득층에 비해 서울에 거주하는 중, 고소득층에서 직장지와 거주지의 주택가격 차이가 통근시간에 미치는 영향력이 높음을 의미한다.

거주지의 주택가격은 저소득층에서는 8번째, 중소득층에서는 9번째로 기여도 상위권에 있는 것을 확인했다. 반면, 고소득층에서는 6번째 중요변수로, 거주지의 주택가격이 다른 소득계층에 비해 통근시간 결정에 있어 보다 중요한 요인으로 파악되었다. 고소득으로 갈수록 평균통근 시간이 다소 길어지고 직장지에서 서울 3도심까지의 거리가 가까워지고 있었는데, 이는 고소득층의 직장이 고용중심지에 위치해 있어, 고소득 근로자는 통근시간의 감소보다 임대료의 상승이 더욱 가파르다는 Park & Quercia (2015)의 주장과 같은 이유에서 기인했다고 볼 수 있다. 이로 인해 오히려 고소득층의 통근시간 결정에 있어 주택가격이 중, 저소득층에 비해 더욱 큰 영향을 미친 것으로 사료된다.

거주 권역이 서울 여부인지를 나타내는 더미변수와 직장지에서 가장 가까운 CBD까지의 거리는 모든 소득계층에서 변수의 기여도 상위권에 위치해있었다. 거주지역이 서울인지, 직장지가 서울 3도심에서 가까운지에 대한 요인은 통근시간에 매우 중요한 요인으로 작용한다고 볼 수 있다. 나이는 저소득층에서 고소득층으로 갈수록 통근시간 결정에 주요한 요인을 미치고 있었으며, 특히 고소득층에서는 나이가 2번째로 높은 기여도 값을 보였다. 나이에 따른 통근시간의 변화는 나이대별 통근수단 변화와 밀접한 관련이 있을 것으로 추측된다.

실제 응답자의 소득계층별, 나이대별 이용 통근수단을 시각화한 [그림 13]을 보면, 모든 소득계층에서 나이대가 낮을수록 대중교통 이용비율이 높아지고 있다. 한편, 승용차 및 오토바이의 경우 고소득층에서는 나이대가 높아질수록 이용 비율이 현저히 증가하지만, 저소득층의 경우 증가율이 보다 미미

하며 40대 이후에는 다시 승용차 및 오토바이의 이용비율이 감소하는 것으로 나타났다. 40대 이상의 고소득층 응답자의 승용차 및 오토바이의 이용비율은 가장 높은 반면, 저소득층에서는 50대 이하의 모든 응답자는 대중교통 이용비율이 가장 높았다. 나이대가 높아질수록 모든 소득계층에서 도보 및 자전거 이용비율이 높아지고 있었으며, 특히 저소득층에서는 통근수단이 도보 및 자전거인 비율이 나이대가 높아질수록 현저히 증가하고 있었다.



[그림 13] 소득계층별(저소득층, 중소득층, 고소득층), 나이대별 통근수단 이용비율

이러한 통근수단과 나이대의 관계는 나이가 어릴수록 대중교통 이용비율이 높아 통근시간이 늘어날 가능성이 있다고 보여진다. 특히, 고소득층의 경우 나이대에 따른 대중교통 이용비율 및 승용차 및 오토바이의 이용비율의 현저한 증감을 보여주기에 나이가 통근시간 결정에 있어 굉장히 중요한 변수로 작용했을 것이라고 사료된다.

거주지의 대중교통 접근성을 나타내는 거주지에서 가장 가까운 지하철역까지의 거리는 저소득층에서 9번째 중요 변수, 중소득층에서는 12번째, 고소득층의 경우 15번째로 소득계층이 낮아질수록 중요 변수로 작용하고 있었다. 반대로, 직장지에서 가장 가까운 지하철역까지의 거리는 고소득층으로 갈수록 변수의 기여도 순위가 높아졌으나, 전반적으로 모든 소득계층에서 낮은 순위를 보여주었다.

직장지의 사업체당 종사자수는 저, 중소득층에서는 5번째, 고소득층에서는 8번째 중요 변수로 나타났으며, 직장지의 사업체 수는 저소득층에서는 변수의 기여도 상위권에 위치했으나, 고소득층으로 갈수록 변수의 중요도는 낮아

지고 있었다. 직장지에서 가장 가까운 지하철역까지의 거리는 고소득층으로 갈수록 중요도 순위가 올라가긴 했으나, 모두 하위권으로 중요도는 다소 저조하게 나타났다.

거주지의 아파트 비율은 다른 소득계층에 비해 중소득층에서 가장 중요한 요인으로 도출되었으며, 저, 고소득층은 16번째로 중요한 변수로 확인되었다. 거주지의 사설학원 수는 고소득층에서는 10번째 주요 변수로 확인된 반면, 중소득층에서는 16번째, 저소득층에서는 20개의 변수의 기여도 순위 내에 들지 않았다. 이는 교육과 관련한 거주지 특성이 고소득층으로 갈수록 통근시간 결정에 있어 영향력이 커짐을 보여준다.

응답자의 직업에 따른 통근거리 변화는 중, 고소득층에서 7, 8번째 중요 변수로 확인되었으며, 저소득층은 상대적으로 직업 변수는 변수 기여도 순위 하위권에 위치해있었다. 직업에 따른 통근시간의 변화는 소득계층별 직업 유형에 따른 직장지 분포가 상이하기 때문에 발생된다고 예상된다.

종합해보면, 전반적으로 거주지와 직장지의 주택가격 차, 서울 거주 여부, 직장지에서 가장 가까운 CBD까지의 거리, 나이, 직장지의 사업체당 종사자 수, 거주지의 주택가격은 모든 소득계층의 통근시간 결정에 있어 중요한 요인으로 파악되었다. 특히, 나이와 거주지의 주택가격은 고소득층으로 갈수록 그 중요도가 높아지고 있었으며, 서울에 거주하는 응답자의 거주지와 직장지의 주택가격 차(상호작용항)는 저소득층에서는 중요도가 저조하게 나타난 반면, 중, 고소득층에서는 중요도가 높게 도출되었다.

저소득층의 통근시간에는 직장지의 사업체수, 직장지의 인구밀도, 거주지에서 가장 가까운 지하철역까지의 거리가 다른 소득계층에 비해 보다 중요하게 나타났다. 중소득층의 통근시간의 경우, 직업, 거주지의 아파트비율이 다른 소득계층에 비해 중요하게 나타났으며, 거주지의 주택가격증가율의 경우 유일하게 13번째 변수로 20개의 주요변수 안에 포함되어 있는 것을 확인하였다. 고소득층의 통근시간에는 중소득층과 마찬가지로 직업이 중요한 변수로 나타났으며, 거주지의 사설학원 수는 다른 소득계층에 비해 높은 변수의 기여도 순위에 위치한 것을 볼 수 있었다.

2) 소득계층별 통근거리 영향요인의 변수의 기여도

[표 13] 소득계층별 통근거리에 영향을 미치는 변수의 기여도 Shapley Value 순위

순위	저소득층	중소득층	고소득층
1	거주지와 직장지의 주택가격 차	거주지와 직장지의 주택가격 차	거주지와 직장지의 주택가격 차
2	직장지의 인구밀도	직장지에서 가장 가까운 CBD까지의 거리	직장지에서 가장 가까운 CBD까지의 거리
3	직장지에서 가장 가까운 CBD까지의 거리	거주지의 인구밀도	거주지의 주택가격 (평당가)
4	거주지의 인구밀도	직장지의 인구밀도	거주지의 인구밀도
5	거주지의 주택가격 (평당가)	직장지의 사업체당 종사자수	직장지의 인구밀도
6	거주지의 총 인구수	(서울 거주)* 거주지와 직장지의 주택가격 차	(서울 거주)* 거주지와 직장지의 주택가격 차
7	거주 권역-서울	거주 권역-서울	거주 권역-서울
8	(서울 거주)* 거주지와 직장지의 주택가격 차	직장지에서 가장 가까운 지하철역까지의 거리	직장지의 사업체당 종사자수
9	직장지에서 가장 가까운 지하철역까지의 거리	거주지의 총 인구수	직장지에서 가장 가까운 지하철역까지의 거리
10	직장지의 총 인구수	거주지의 주택가격 (평당가)	직장지의 사업체수
11	직장지의 사업체수	직장지의 사업체수	거주지의 총 인구수
12	거주지의 사업체수	거주지의 아파트비율	거주지의 토지이용균형
13	나이	거주지의 토지이용균형	거주지의 아파트비율
14	성별	거주지에서 가장 가까운 지하철역까지의 거리	거주지에서 가장 가까운 지하철역까지의 거리
15	거주지에서 가장 가까운 지하철역까지의 거리	직장지의 총 인구수	성별
16	거주지의 아파트비율	직장지의 아파트비율	거주지의 주택가격증가율
17	직장지의 사업체당 종사자수	거주지의 주택가격증가율	직장지의 아파트비율
18	직업-사무/관리자	거주지의 사업체당 종사자수	직장지의 총 인구수
19	직장지의 아파트비율	성별	거주지의 사설학원 수
20	거주지의 토지이용균형	직업-서비스/판매	거주지의 사업체당 종사자수

다음으로, 통근거리에 대한 각 소득계층의 변수별 상대적 기여도를 살펴보았다. 통근시간과 마찬가지로, 모든 소득계층에서 거주지와 직장지 간 주

택가격의 차이가 통근거리 예측에 있어 미치는 영향력이 가장 큰 것으로 나타난다. 서울 더미변수와 거주지와 직장지 간 주택가격 차의 상호작용 변수를 살펴보면 저소득층에 비해 중, 고소득층에서 통근거리 예측에 미치는 영향력이 크게 나타났으며, 이는 서울에 거주하는 저소득층에 비해 서울에 거주하는 중, 고소득층에서 직장지와 거주지의 주택가격 차이가 통근거리에 미치는 영향력이 높음을 의미한다. 이는 통근시간과 비슷한 양상이나, 통근시간의 경우 상호작용 변수의 저소득층과 중, 고소득층 간 영향력 차이가 보다 크게 나타났다.

거주지의 주택가격에 대한 중요도는 고소득층에서 3번째로 다른 소득계층에 비해 가장 높게 나타났으며, 저소득층에서 5번째, 중소득층에서 10번째로 나타났다. 통근시간과 마찬가지로 고소득층에서 주택가격의 중요도 순위가 가장 높게 나타난 이유는 고소득층으로 갈수록 직장지가 서울 3도심에 가까워져 인근의 주택가격 또한 상승률이 높아져 보다 민감하게 작용했을 것으로 추론할 수 있다. 저소득층의 경우 중소득층에 비해 한정된 재원으로 인해 통근시간을 결정하는 거주지 선택에 있어 주택가격이 더욱 중요하게 영향을 미친 것으로 보인다.

직장지가 CBD, 즉 서울 3도심에서 가까운지에 대한 요인은 모든 소득계층에서 통근거리에 매우 중요한 요인으로 작용하고 있었다. 직장지의 사업체당 종사자수는 중, 고소득층에서 상대적으로 중요하게 나타난 반면, 저소득층에서는 17번째로 그 영향력이 미미하게 나타났다.

직장지에서 가장 가까운 지하철역까지의 거리는 통근거리에 있어 모든 소득계층에서 통근시간에 비해 높은 변수의 기여도 순위를 차지했다. 직장지에서 가장 가까운 지하철역까지의 거리는 직장이 위치한 행정동 중심점에서 가장 가까운 지하철역까지의 거리를 의미하는데, 이는 해당 행정동의 지하철역 등 대중교통 시설이 얼마나 잘 갖추어져 있는지를 의미한다. 따라서 지하철역이 가까이 갖춰져 있는 행정동일수록 도심에 위치해있을 확률이 높다. 직장지의 도심 여부 및 대중교통 접근성이 통근거리에 중요하게 작용하는 것으로 보인다.

통근시간에는 나이가 전반적으로 상위권에 위치해있었다면, 통근거리의 경우 전반적인 소득계층에서 거주지 및 직장지의 인구밀도, 총 인구 수의 중요도가 통근시간에 비해 크게 나타났다. 통근시간의 경우 대중교통 이용 등과 같은 통근수단의 영향이 크다면, 통근거리는 거주지 및 직장지의 인구 등과 관련한 특성이 크게 작용하는 것으로 보인다. 교외 지역 등 상대적으로 인구밀도가 낮은 지역의 경우 인구밀도가 높은 도시 지역보다 덜 혼잡한 경향이 있으며, 특히 자동차로 이동하는 지역의 경우 이동 속도가 증가하고 장거리 이동이 가능하다(Blumberg and King, 2019). 따라서 통근시간에는 인구밀도가 미치는 영향은 미미할 수 있으나, 통근거리에는 인구밀도에 따른 통근거리 변화가 확연히 드러나는 것으로 추론해볼 수 있다.

서울 거주여부는 통근거리에서도 여전히 모든 소득계층에서 7번째 중요변수로, 통근거리 결정에도 중요한 요인으로 도출되었다.

이외 변수의 기여도 자체가 높지는 않지만, 소득계층별로 변수의 기여도 순위가 차이를 보이는 변수는 다음과 같다. 거주지의 아파트비율은 중, 고소득층에서 저소득층에 비해 변수의 기여도가 높게 나타났다. 나이는 저소득층에서만 20개의 변수의 기여도 순위 내 위치했다. 거주지의 주택가격증가율은 고소득층에서 다른 소득계층에 비해 중요하게 나타났지만 16번째 중요변수로 그 영향력은 미미했으며, 저소득층의 경우 주택가격증가율은 20개의 변수의 기여도 순위 내에 들지 않았다. 거주지의 사설학원 수 또한 고소득층에서만 20개의 변수의 기여도 순위 내 위치했지만 19번째 중요변수로 그 영향력이 크지 않았다.

2. 소득계층별 영향요인의 비선형 관계 비교

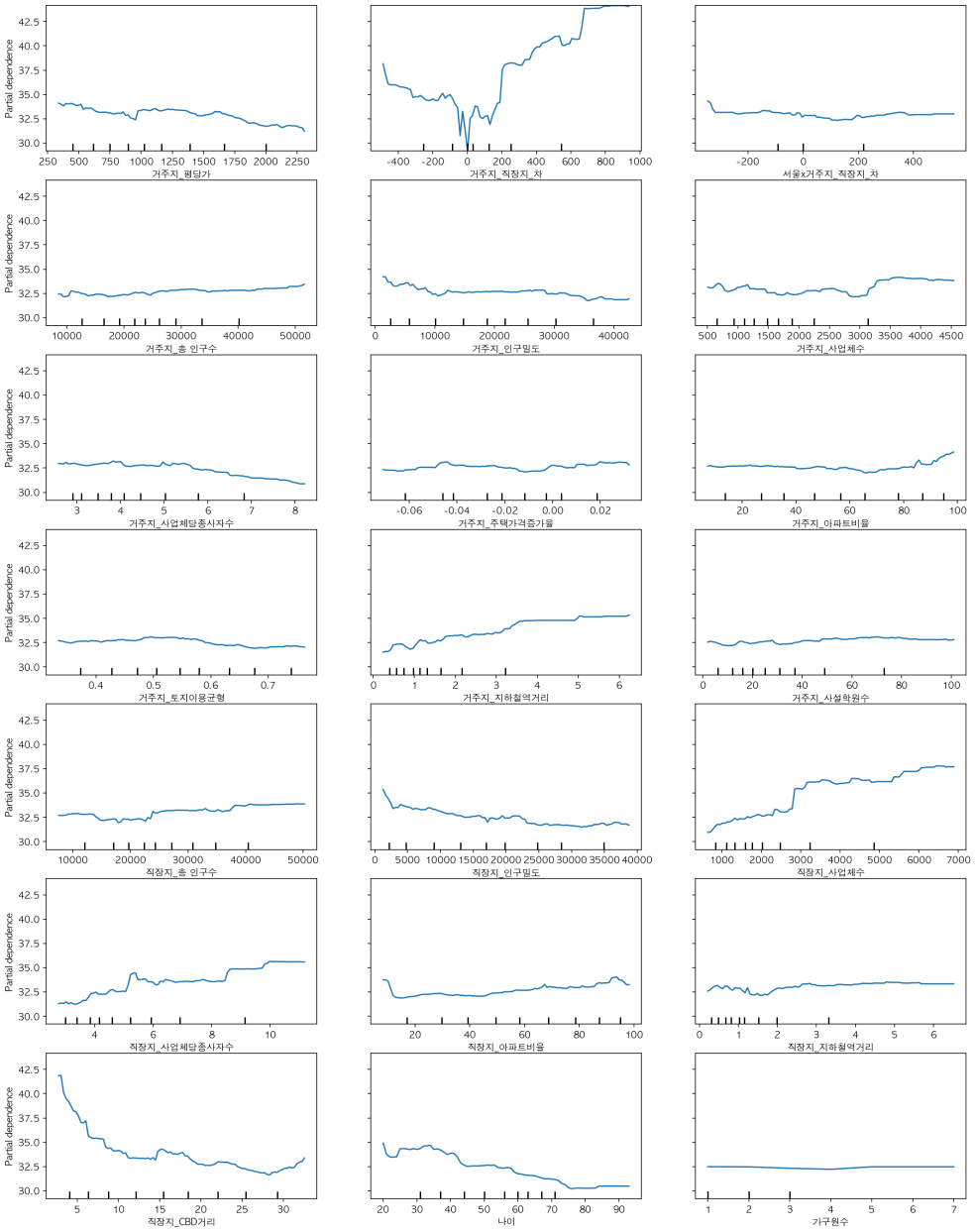
변수별 기여도를 통해 소득계층별 통근시간 및 통근거리 결정에 있어 중요도가 큰 주요 영향요인을 살펴보았다면, Partial Dependence Plot(PDP)를 통해 각 변수가 통근시간 및 통근거리에 어떠한 영향을 미치는지를 파악할 수 있다. Partial Dependence Plot(PDP)는 기계 학습 모델의 예측 결

과에 대해 하나 혹은 두 개의 변수가 갖는 한계 효과를 보여준다(Friedman, 2001). PDP를 통해, 각 변수가 통근거리에 어떠한 영향을 미치는지를 파악할 수 있으며, 이에 따라 변수와 통근거리 간 선형적 혹은 비선형적 관계가 있는지를 파악할 수 있다. PDP를 통해, 변수별 기여도가 높게 나타난 주요 영향요인이 소득계층별 통근시간 및 통근거리에 어떠한 영향을 미치고 있으며, 그 관계가 어떻게 다른지를 살펴보았다. 소득계층별 통근시간에 대한 XGBoost 모형의 PDP는 [그림 14], [그림 15], [그림 16], 소득계층별 통근거리에 대한 XGBoost 모형의 PDP는 [그림 18], [그림 19], [그림 20]에서 살펴볼 수 있다. PDP의 y축은 각각 통근시간 및 통근거리를 의미하며, x축은 각 변수의 실제값을 의미한다. 보다 용이한 비교를 위해 표준화된 PDP의 경우 부록에 첨부하였으며, y축은 각각 표준화된 통근시간 및 표준화된 통근거리를 의미하며, x축은 각 변수의 실제값을 의미한다.

1) 소득계층별 통근시간 영향요인의 Partial Dependence Plot

먼저, 저소득층 통근시간에 대한 PDP를 나타낸 [그림 14]를 보면, 변수의 기여도가 가장 높게 나타난 거주지와 직장지의 주택가격 차는 U자형 그래프를 그리고 있었다. 거주지와 직장지의 주택가격 차는 직장지의 평당 주택가격에서 거주지의 평당 주택가격을 뺀 값으로, x축이 음수일 경우 거주지가 직장지의 평당가보다 비싼 것이며, 양수일 경우 거주지가 직장지의 평당가보다 저렴하다는 것을 의미한다. 거주지가 직장지의 평당 주택가격보다 400만원 정도 비싼 경우 거주지와 직장지가 동일할 때에 비해 최대 7분 정도 시간이 증가했다. 반대로 거주지가 직장지의 평당 주택가격보다 저렴할수록 통근시간은 증가하는 것으로 보여졌다. 거주지의 주택 평당가격은 증가할수록 통근거리가 감소하고 있었으나, 감소폭은 크지 않은 것으로 확인되었다.

직장지에서 가장 가까운 CBD까지의 거리와 저소득층의 통근시간 간 PDP를 살펴보면, 직장지가 CBD에서 멀어질수록 통근시간은 급격히 감소하고 있는 것으로 나타났다. 직장지의 사업체당 종사자수와 직장지의 사업체수는 증가할수록 통근시간 또한 길어지고 있었다. 이는 종사자수가 많고, 직장



[그림 14] 저소득층의 통근시간 영향요인에 대한 Partial Dependence Plot

이 밀집되어 있는 지역일수록 해당 지역으로 통근하는 저소득층의 통근시간은 증가한다는 것을 의미한다. 반면, 직장지의 인구밀도는 높을수록 통근시간이 다소 감소하였다. 직장지의 업무시설 밀집과 반대로, 거주 밀집은 통근

시간 감소에 기여하는 것으로 해석가능하다.

응답자의 나이가 많을수록 통근시간은 감소했다. 앞서 [그림 13]에서 저소득층의 경우 응답자의 나이가 많을수록 대중교통 이용비율이 감소하고 도보 및 자전거 이용률이 증가하였는데, 이러한 통근행태가 나이 증가에 따른 통근시간 감소에 기여했을 것으로 보여진다.

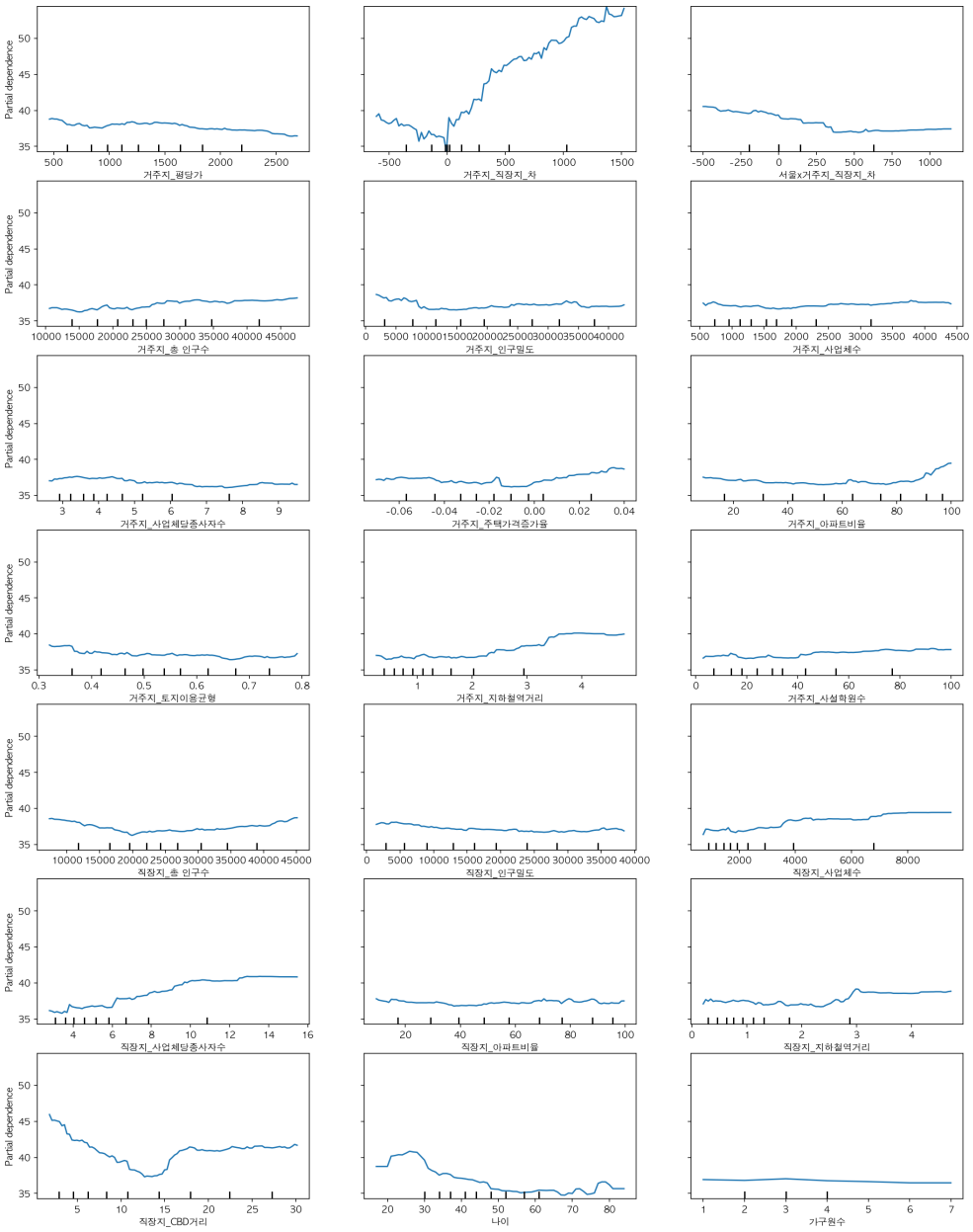
거주지에서 가장 가까운 지하철역까지의 거리가 길어질수록 통근시간은 증가했다. 거주지 행정동 중심지에서 가장 가까운 지하철역까지의 거리가 약 3km 이후부터는 통근시간 증가폭이 미미했다.

더미 변수인 거주 권역이 서울인지에 대한 여부는, 거주 권역이 서울일 경우 그렇지 않을 때보다 저소득층의 통근시간이 감소하는 것으로 나타났다.

다음으로, 중소득층의 통근시간 영향요인별 PDP를 나타낸 [그림 15]를 살펴보았다. 저소득층과 마찬가지로, 변수의 기여도가 가장 높게 나타난 거주지와 직장지의 주택가격 차는 U자형 그래프를 그리고 있었다. 단, 저소득층에 비해 x축의 구간이 더욱 넓어졌으며, 거주지와 주택가격 차로 인한 통근시간의 증가폭이 보다 커진 것을 확인할 수 있다.

반면, 서울에 거주하는 중소득층 응답자의 거주지와 주택가격 차로 인한 통근시간 변화를 보면, 거주지가 직장지의 주택가격보다 비쌀수록 통근시간은 다소 증가하였으며, 거주지가 직장지보다 저렴할 경우 통근시간의 변화는 굉장히 미미했다. 이러한 경향은 주거 환경 개선을 위해 통근시간 증가에도 불구하고 주택가격이 비싼 지역으로 이동했다고 볼 수 있으며, 수도권 거주자에 비해 서울 거주 응답자에게 보다 극명하게 나타나고 있었다. 이는 소득이 높을수록 주거지를 선택하는데 있어 대중교통 수단의 이용가능성을 덜 고려할 뿐만 아니라 다른 어메니티 요소를 고려할 가능성이 높다(Adair et al., 2000; 하재현 2020)는 선행연구의 내용으로 뒷받침될 수 있다.

직장지에서 가장 가까운 CBD까지의 거리를 보면, 약 15km까지는 직장지에서 서울 3도심에서의 거리가 멀어질수록 통근시간은 감소하나, 15km를 벗어날 경우 다시 통근시간이 다소 증가하다가 변화가 미미해지는 것을 나타난다. 저소득층과 마찬가지로 직장지의 사업체당 종사자수는 많을수록 통근



[그림 15] 중소득층의 통근시간 영향요인에 대한 Partial Dependence Plot

시간 또한 증가하였다. 중소득층 또한 업무중심지에 가까운 지역으로 통근할수록 통근시간은 다소 길어지는 것으로 보여진다.

중소득층 응답자의 나이대가 높아질수록 통근시간은 줄어들고 있었다. 단,

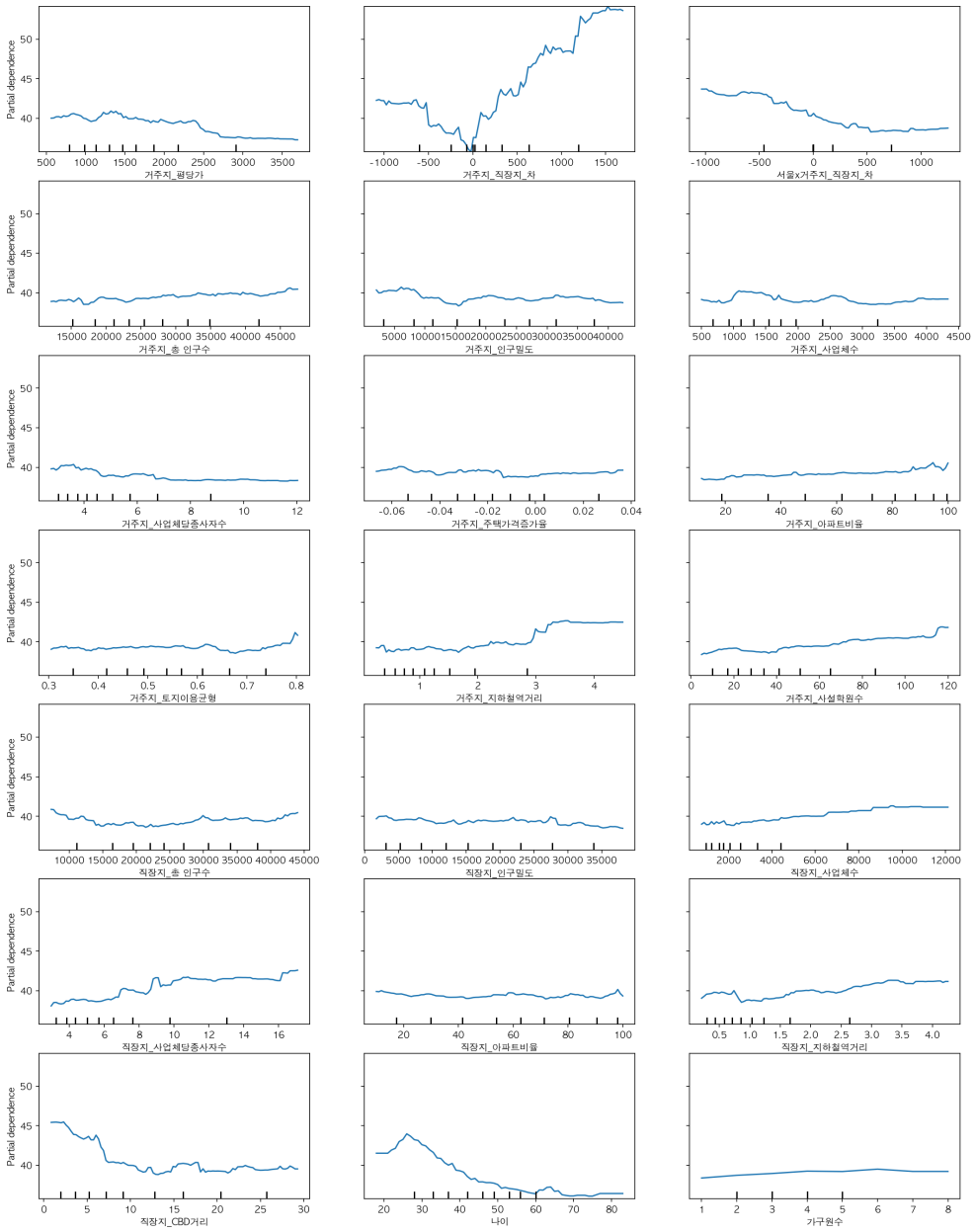
20대 구간에서는 20대 후반으로 갈수록 통근시간이 오히려 다소 증가하는 것을 확인했다. 중소득층 또한 나이가 어릴수록 대중교통 이용비율이 높아지기 때문에 나이에 따른 통근시간 변화가 일어난 것으로 보인다.

다른 소득계층에 비해 중소득층의 통근시간에 미치는 영향력이 높게 나왔던 거주지의 아파트 비율을 살펴보면, 거주지의 아파트비율이 증가할수록 통근시간은 다소 증가하는 양상을 나타냈다. 특히, 아파트비율이 80% 이상인 구간에서는 통근시간이 확연히 늘어나고 있었다. 이는 아파트비율이 지나치게 높은 거주지역일 경우 거주환경은 개선될 수 있어도, 통근시간에는 오히려 부정적인 영향을 미칠 수 있음을 보여준다.

이외에, 거주지에서 가장 가까운 지하철역까지의 거리를 살펴보면 거리가 멀수록 통근시간은 증가했다. 단, x축의 구간이 저소득층보다 짧았는데, 이는 중소득층의 거주지가 전반적으로 지하철역에 보다 가까운 곳에 분포해있다고 추론해볼 수 있다. 거주지의 주택가격증가율은 0을 기준으로 0보다 클 경우 2011년 대비 2016년의 해당 지역의 주택가격이 증가한 것을 의미하는데, 0 이상의 구간에서 증가율이 커질수록 통근시간이 함께 증가하는 것으로 나타났다. 통근시간이 보다 길어짐에도 불구하고 해당 거주지를 선택한 데에는 시간에 따른 근린환경 개선, 주택에 대한 투자가치 기대감 등이 작용했을 가능성이 있다.

저소득층과 마찬가지로, 중소득층에서도 거주 권역이 서울일 경우 그렇지 않을 때보다 통근시간이 줄고 있었다. 서비스/판매직에 종사할 경우 그렇지 않은 경우보다 통근시간이 감소한 반면, 사무/관리직에 종사할 경우 그렇지 않은 응답자에 비해 통근시간이 길어지고 있었다.

마지막으로, [그림 16]을 통해 고소득층의 통근시간에 대한 영향요인별 PDP를 확인하였다. 거주지의 주택가격이 직장지에 비해서 500~ 1000만원 정도 평당 가격이 높은 구간에서는 주택가격 차이 변화에 따른 고소득층 통근시간의 변화가 크지 않게 나타났다. 하지만 직장지의 주택가격이 거주지보다 높은 경우는 중소득층와 유사하게 주택가격차이가 커질수록 통근시간도 커지는 것으로 나타났다.



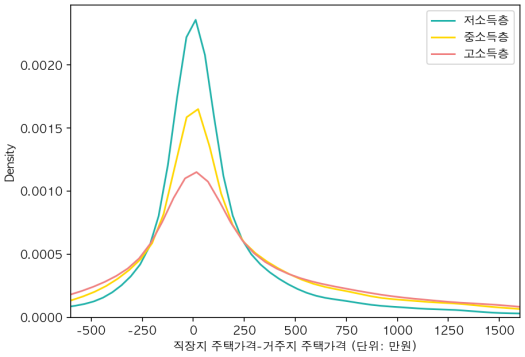
[그림 16] 고소득층의 통근시간 영향요인에 대한 Partial Dependence Plot

거주지와 직장지의 주택가격 차이가 통근시간에 미치는 영향은 실제 거주지와 직장지의 주택가격 분포와 연결시켜 이해할 필요가 있다. 소득계층별

거주지와 직장지의 주택가격 차 분포를 나타낸 [그림 17]을 보면, 저소득층의 경우 거주지와 직장지의 주택가격 차이가 크지 않은 경우가 대다수로 나타났다. 고소득층 또한 거주지와 직장지의 주택가격 차가 0에 가장 가까운 구간에 집중되어 있긴 하지만, 저, 중소득층에 비해 그 정도는 낮은 편이다. 대신, 이외 거주지가 직장지보다 비싸거나 저렴한 모든 구간에서 고소득층의 분포가 가장 높게 나타난다. 이는 고소득층이 다른 소득계층에 비해 거주지와 직장지의 주택가격 차이 분포가 가장 고르게 퍼져있음을 의미한다.

[표 14] 소득계층별 직장지 대비 거주지 주택가격 분포

주택가격	직장지 〈거주지	직장지 =거주지	직장지 〉주거지
저소득층	29.4%	26.4%	44.2%
중소득층	32.0%	15.8%	52.3%
고소득층	35.8%	12.5%	51.8%



[그림 17] 소득계층별 거주지와 직장지의 주택가격 차 분포

고소득층 또한 거주지와 직장지의 주택가격 차가 0에 가장 가까운 구간에 집중되어 있긴 하지만, 저소득층에 비해 그 정도는 낮은 편이다. 또한, 고소득층은 다른 소득계층에 비해 거주지와 직장지의 주택가격 차 분포가 상대적으로 넓게 분포하고 있다. 이는 저소득층의 경우 거주지 및 직장지가 다른 계층에 비해서 상대적으로 제한되어 있어 거주지와 직장지의 주택가격 차이 분포가 다른 계층에 비해 상대적으로 평균을 중심으로 몰려있는 것으로 이해될 수 있으며 저소득계층의 경우 주거지와 인접한 지역 일자리에 종사하는 경향이 높은 것으로 나타났다. 실제로 [표 14]에서 보는 것과 같이 거주지와 직장지의 행정동이 일치하는 경우가 저소득의 경우 26.4%로 나타났으나, 중소득층 및 고소득층에서는 각각 15.7%, 12.5%로 비율이 낮아지는 것으로 나타났다. 한편, 고소득층으로 갈수록 주거지 주택가격이 직장지 주택가격에 비해서 높은 곳에 거주하는 비율도 높아지지만, 동시에 직장지 주택가격이

주거지 주택가격에 비해서 높아지는 경우도 증가하고 있다. 이는 저소득층의 직장지와 주거지의 불일치(mismatch) 가설과는 달리 고소득층일수록 직장이 지대가 높은 고용중심지에 가까이 위치하기 때문에, 통근시간의 감소보다 임대료의 상승이 더욱 가파르기(Park and Quercia, 2015)때문이라고 해석가능하다. PDP 그래프와 연결하여 해석하면 이로 인해 통근시간이 저소득층에 비해서 길게 나타난 것으로 이해될 수 있다.

한편, 직장지가 거주지보다 저렴한 경우는 주거비용 외에도 지역의 교육(Kim-J et al., 2005), 안전(Weisbrod et al., 1980), 도시 어메니티(Rapoport, 1980; Tyrvainen and Vaananen, 1998; Rouwendal and Meijer, 2001; Parkes et al.,2002) 등 거주지의 환경적 질을 우선적으로 선택한 것으로 이해될 수 있다. 주거 선택에 있어서 고용 접근성 이외에 다양한 요인들로 인한 것으로 이와 같은 경향은 고소득층으로 갈수록 강하게 나타나고 있었다.

서울에 거주하는 고소득층 응답자의 거주지와 주택가격 차로 인한 통근시간 변화에서는 이러한 현상이 더욱 분명하게 나타난다. 거주지가 직장지보다 저렴할 경우 통근시간의 변화는 크지 않지만, 거주지가 직장지의 주택가격보다 비쌀 경우 통근시간은 증가하는 경향이 나타난다. 서울시 내에서 고소득층으로 갈수록 통근시간이 길어지는 것을 감수하고서라도 주택가격이 더욱 비싼 지역에 거주하려는 경향이 심화되고 있음을 보여준다. 이는 가구의 주거입지 선택이 고용 접근성에 대한 고려와 삶의 질에 대한 고려의 상쇄관계로 설명된다는(Phe and Wakely, 2000; Kim-J et al, 2005; Moos and Skaburskis, 2010; 김승남·안건혁, 2011) 접근성-공간 맞교환 이론(Access-Space Trade-off)과도 같은 맥락이며, 특히 본 연구에서는 이러한 경향이 고소득층으로 갈수록 심화되며, 서울 내에서 두드러진다는 것을 확인했다.

위와 동일한 이유로, 거주지의 사설학원 수는 다른 소득계층에 비해 고소득층에서 변수의 기여도가 높게 나타나기도 했는데, PDP를 살펴보면 거주지의 사설학원 수가 많을수록 통근시간은 증가하는 것으로 나타났다.

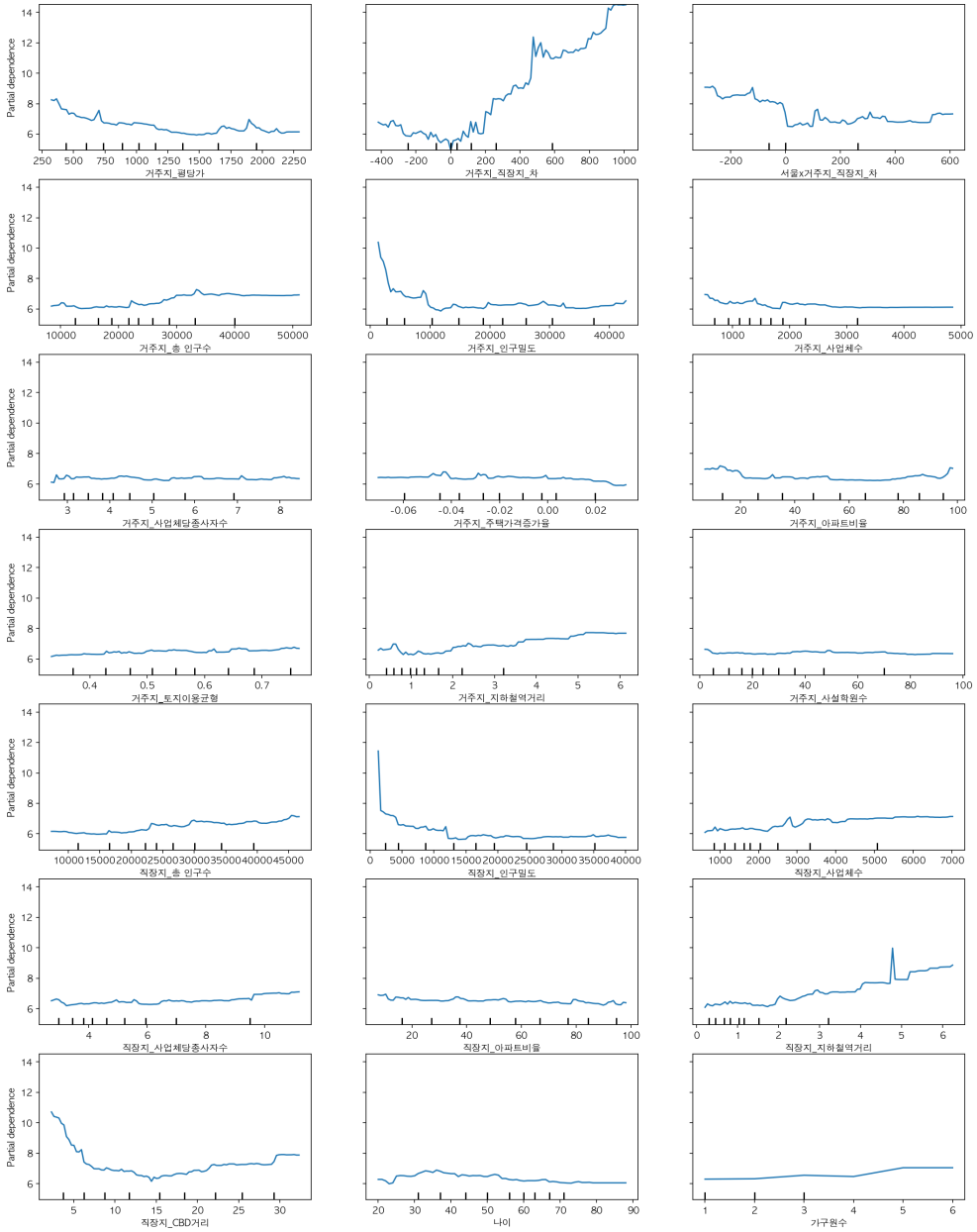
다른 소득계층과 마찬가지로 고소득층에서도 직장지에서 가장 가까운 CBD까지의 거리는 길어질수록, 사업체당 종사자수는 적을수록 통근시간이 짧았다. 거주지의 주택가격 또한 주택 평당가격이 비쌀수록 통근시간은 감소하였다. 나이의 경우, 30대 이후로는 나이가 많아질수록 고소득층의 통근시간은 선형적으로 감소하고 있었다. 고소득층 내에서 나이가 많아질수록 현저히 증가하는 승용차 이용률이 주 원인으로 보여진다. 단, 30대 이전까지는 중소득층과 마찬가지로 20대 후반으로 갈수록 통근시간이 오히려 다소 증가하는 것을 확인했다.

더미변수를 살펴보면, 거주 권역이 서울일 경우 그렇지 않은 응답자에 비해 통근시간이 줄고 있었다. 서비스/판매직에 종사할 경우 그렇지 않은 경우보다 통근시간이 감소하였으며, 사무/관리직에 종사할 경우 그렇지 않은 응답자보다 통근시간이 길어지고 있었다. 이러한 결과는 다른 소득계층과 동일했다.

2) 소득계층별 통근거리 영향요인의 Partial Dependence Plot

먼저, 저소득층의 통근거리에 대한 PDP를 [그림 18]을 통해 살펴보았다. 변수의 기여도가 가장 높게 측정된 거주지와 직장지의 주택가격 차는 0을 기준으로 직장지의 주택가격이 거주지의 주택가격보다 비쌀수록, 그 차이가 클수록 통근거리가 길어지는 것으로 나타났다. 직장지의 주택가격이 거주지보다 저렴할 때에는 통근거리가 증가하긴 했으나, 통근거리 증가폭이 약 1km 정도로 크지 않았다. 서울에 거주하는 응답자의 거주지와 직장지의 주택가격 차가 통근거리에 미치는 영향을 보면, 거주지의 주택가격이 직장지보다 저렴할 때에는 주택가격 차이에 따른 통근거리 증가폭이 크지 않지만, 오히려 거주지의 주택가격이 직장지에 비해 비쌀 경우 통근거리가 다소 증가하는 것으로 나타났다. 거주지의 주택 평당가격은 증가할수록 통근거리가 감소하고 있었다.

직장지에서 가장 가까운 CBD까지를 살펴보면, 직장지가 서울 3도심에서 멀어질수록 통근거리가 감소했다. 그러나 15km를 기준으로, 직장지가 서울 3도심에서 15km 이상 떨어져 있을 경우, 통근거리는 오히려 다소 증가하는



[그림 18] 저소득층의 통근거리 영향요인에 대한 Partial Dependence Plot

것으로 확인되었다.

거주지의 인구밀도는 높을수록 통근거리가 감소했으며, 거주지의 총 인구수는 증가할수록 통근거리가 다소 늘어났다. 인구밀도가 높을 경우 교외 지

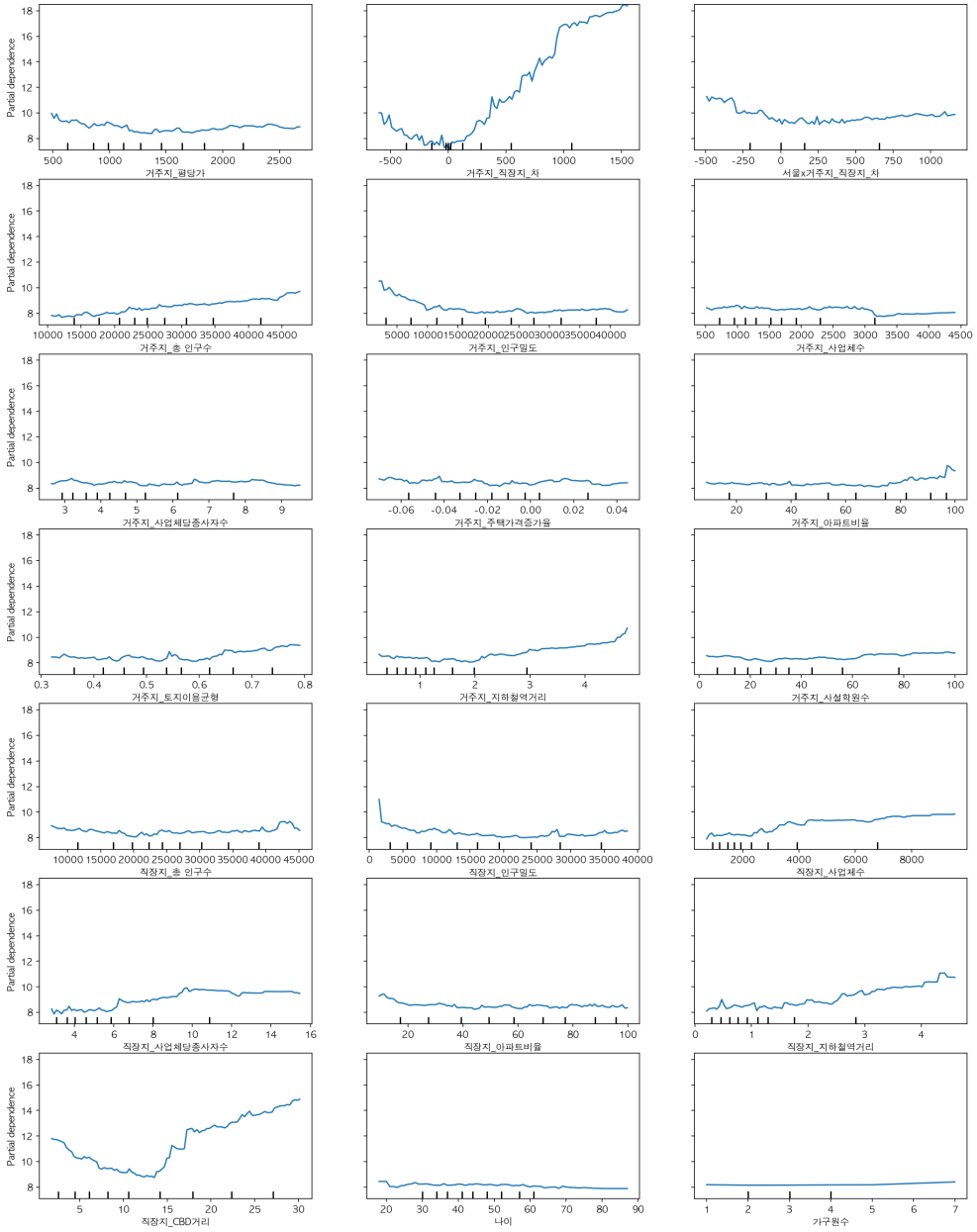
역보다는 도심부에 위치할 확률이 높으며, 이에 따라 통근거리가 줄어들었다고 볼 수 있다. 마찬가지로 직장지의 인구밀도 또한 증가할수록 통근거리가 급격히 감소하는 것을 확인할 수 있다. 직장지의 인구밀도가 증가하는 것은 직장지 인근의 주거 시설이 밀집해있다는 것을 의미하며, 교외지역보다는 도심에 가깝기 때문에 통근거리가 감소하는 것이라고 해석할 수 있다.

직장지에서 가까운 지하철역까지의 거리는 멀어질수록 통근거리가 증가하였다. 직장지에서 가장 가까운 지하철역까지의 거리는 직장이 위치한 행정동 중심점에서 가장 가까운 지하철역까지의 거리를 의미한다. 가장 가까운 지하철역까지의 거리가 짧을수록 지하철역이 가까이 갖춰져 있는 행정동에 위치한 직장지이며, 교외지역보다는 도심에 위치해있을 확률이 높다고 해석할 수 있다.

중요하게 나타난 더미 변수인 서울 거주 여부의 경우, 서울에 거주할 때 그렇지 않은 경우보다 통근거리가 감소하는 것으로 나타났다.

다음으로, 중소득층의 통근거리에 대한 PDP를 [그림 19]에 나타내었다. 저소득층과 마찬가지로, 거주지와 직장지의 주택가격 차는 직장지의 주택가격이 거주지의 주택가격보다 비쌀수록, 그 차이가 클수록 통근거리가 길어졌다. 직장지의 주택가격이 거주지보다 저렴할 때에는 통근거리가 증가했으며, 통근거리 증가폭이 약 2km 정도로, 저소득층에 비해 증가폭이 다소 커졌다. 그러나 저소득층과 마찬가지로 직장지가 거주지보다 저렴한 구간이 그렇지 않은 구간에 비해 작았다. 서울에 거주하는 응답자의 거주지와 직장지의 주택가격 차의 PDP에서는, 거주지의 주택가격이 직장지보다 저렴할 때에는 주택가격 차이에 따른 통근거리 변화가 크지 않지만, 거주지의 주택가격이 직장지에 비해 비쌀 경우 통근거리가 다소 증가하였다. 거주지의 주택가격은 증가할수록 통근거리가 감소했으나, 주택가격 증가에 따른 통근거리 감소 정도가 미미했다.

직장지에서 가장 가까운 CBD까지를 살펴보면, 직장지가 서울 3도심에서 멀어질수록 통근거리가 감소하다가, 약 15km를 기준으로 직장지가 서울 3도심에서 15km 이상 떨어져있을 경우 다시 통근거리가 증가하고 있었다.



[그림 19] 중소득층의 통근거리 영향요인에 대한 Partial Dependence Plot

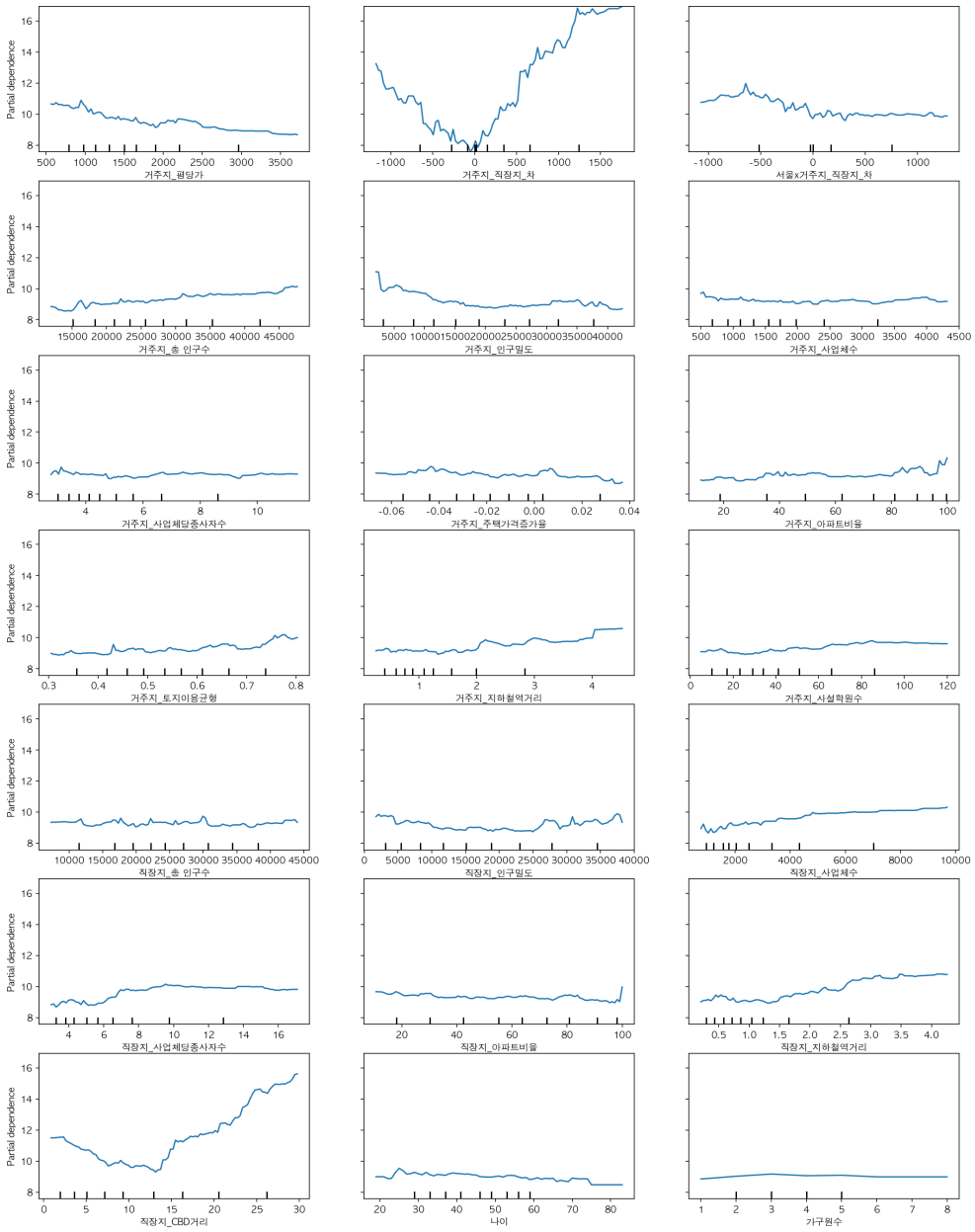
이는 직장의 교외화에 따른 역통근 현상이 발생했을 가능성이 있다. 대도시권의 교외지역이 고용 중심지로 성장하게 되면, 도심에 거주하는 도시민들이 교외지역의 일자리로 이동하게 된다(신학철·우명제, 2020). 특히, 주거지에

비해 직장지의 교외지역 이동 속도가 높기 때문에 이러한 현상은 지속적으로 증가할 확률이 높다(Davidson and Ryerson, 2018; Rosenbloom, 1992; Levine, 1992; Glaser et al., 2001; Aguilera et al., 2009; 신학철·우명제, 2020). 교외지역으로 분산된 고용지는 효율적인 교통서비스 제공이 어렵고, 고속도로 위주의 접근성으로 인해 자가용 중심의 통행을 유도한다(신학철·우명제, 2020). 저소득층에 비해 중소득층은 대중교통 의존도가 보다 낮기 때문에, 일자리 교외화가 이루어질 경우 통근거리가 길어짐에도 불구하고 서울 3도심에서 떨어진 직장지 인근으로 주거지를 옮기지 않을 가능성이 높다.

거주지의 인구밀도와 직장지의 인구밀도는 모두 증가할수록 통근거리는 감소하였다. 또한, 직장지에서 가장 가까운 지하철역까지의 거리가 멀어질수록 통근거리도 증가했다. 이는 저소득층과 비슷한 양상인데, 저소득층에 비해 거주지 및 직장지의 인구밀도, 직장지에서 가장 가까운 지하철역까지의 거리에 따른 통근거리 증감 정도는 다소 작아진 것으로 보인다. 직장지의 사업체당 종사자수는 증가할수록 통근거리가 다소 길어졌다. 이는 직장지가 업무중심지에 위치해있을수록 해당 지역에서 근무하는 통근자의 통근거리가 다소 길어진다는 것을 의미한다.

서울 거주 여부의 경우, 서울에 거주하는 응답자가 서울 외에 거주하는 응답자보다 통근거리가 짧았다. 거주지의 아파트비율에 따른 통근거리 변화는 매우 작았으나, 아파트 비율이 80% 이상일 때부터는 통근거리가 미미하게 증가했다.

마지막으로, 고소득층의 통근거리에 대한 PDP [그림 20]을 살펴보았다. 거주지와 직장지의 주택가격 차는 직장지의 주택가격이 거주지의 주택가격보다 비쌀수록, 그 차이가 클수록 통근거리가 길어졌다. 직장지의 주택가격이 거주지보다 저렴할 때에는 통근거리가 증가했으며, 통근거리 증가폭이 약 4km 이상으로, 저, 중소득층에 비해 굉장히 커졌다. 또한, 직장지의 주택가격이 거주지보다 저렴한 경우의 거주지와 직장지의 주택가격 차도 커진 것을 확인할 수 있다. 이는 앞서 설명했듯, 고소득층의 경우 대중교통 수단의 이



[그림 20] 고소득층의 통근거리 영향요인에 대한 Partial Dependence Plot

용 가능성을 덜 고려하여(Adair et al., 2000; 하재현 2020) 통근 거리 증가보다 거주지의 환경적 질을 우선시할 가능성이 있기 때문이라고 보인다. 서울에 거주하는 응답자의 거주지와 직장지의 주택가격 차의 PDP에서는, 다

른 소득계층과 마찬가지로 거주지의 주택가격이 직장지보다 저렴할 때에는 주택가격 차이에 따른 통근거리 변화가 크지 않지만, 거주지의 주택가격이 직장지에 비해 비쌀 경우 통근거리가 증가했다. 거주지의 주택가격은 증가할수록 통근거리가 감소했다.

직장지에서 가장 가까운 CBD까지를 살펴보면, 직장지가 서울 3도심에서 멀어질수록 통근거리가 감소하다가, 약 15km를 기준으로 직장지가 서울 3도심에서 15km 이상 떨어져있을 경우 다시 통근거리가 증가하고 있었다. 이는 중소득층과 마찬가지로 직장의 교외화에 따른 역통근 현상이 발생했을 가능성이 높다. 고소득층 또한 직장지가 서울 3도심에서 15km 이상 떨어진 경우 직장지 인근으로 주거지를 옮기기 보다는 도심에서 교외지역으로 승용차를 이용한 통근을 하고 있을 가능성이 높다고 보여진다. 거주지의 인구밀도는 증가할수록 통근거리가 감소했고, 직장지의 인구밀도에 따른 통근거리 변화는 크지 않았다. 다른 소득계층과 유사하게 직장지에서 가장 가까운 지하철역까지의 거리는 멀어질수록, 직장지의 사업체당 종사자수는 증가할수록 통근거리가 다소 길어졌으나, 그 변화가 크지 않았다.

다른 경우와 마찬가지로, 서울 거주 더미변수를 살펴보면 서울에 거주하는 응답자가 인천, 경기 지역에 거주하는 응답자에 비해 통근거리가 짧았다.

제 5 장 결론

제 1 절 연구의 요약 및 시사점

본 연구에서는 2016년 가구통행실태조사 자료를 활용해 수도권에 거주하는 통근 목적의 응답자를 소득계층별로 나누어 통근시간 영향요인을 비교분석하였다. 거주지의 주택가격, 거주지와 직장지의 주택가격 차 등 주거비용을 나타내는 변수, 응답자가 거주하는 지역의 인구, 사업체수, 주택가격증가율, 아파트 비율, 토지이용균형, 가장 가까운 지하철역까지의 거리, 사설학원수 등과 같은 거주지 특성 변수, 그 외 응답자의 직장지 특성 변수, 응답자 개인 특성 변수를 분석에 활용하였다. 이를 토대로, 통근시간과 주거비용 및 거주지의 환경적 질과의 상쇄관계를 소득계층별로 나누어 알아보고자 하였다. 분석방법론으로 기계학습 기법 중 하나인 XGBoost를 활용해 분석의 설명력을 높였다. 분석결과를 바탕으로 도출된 변수의 기여도와 통근시간과 변수 간 비선형관계를 파악하여 소득계층별 통근시간 영향요인 차이를 분석하였다.

주요 연구 결과는 다음과 같다.

첫째, 서울시 및 서울시 교통권역에 해당하는 수도권 내 거주자의 통근시간은 고소득층으로 갈수록 길게 나타났다. 통근에서 취약계층, 즉 본 연구에서 저소득계층의 공간 불일치(spatial mismatch)가 늘어나는 경향은 확인되지 않았다. 수도권 내 거주자의 평균 통근시간은 고소득층이 가장 높게 나타났다. 고소득층일수록 직장이 지대가 높은 고용중심지에 가까이 위치하므로, 오히려 직장에 인접한 주택 가격에 대한 부담이 증가해 통근시간이 저소득층에 비해 길게 나타났을 것으로 사료된다. 실제로 응답자의 직장이 서울 3도심에서 멀어질수록 통근시간은 감소했다. 이에 더해, 주거 선택에 있어서 고용접근성 이외에 도시 어메니티와 같은 지역 요인들을 보다 중시하는 경향 또한 고소득층으로 갈수록 강하게 나타나 통근시간이 오히려 길어졌을 가능성

도 배제할 수 없다. 통근수단과도 밀접한 관련이 있을 것으로 보인다. 저소득층의 승용차 및 오토바이 이용률은 다른 소득계층에 비해 낮은 반면, 대중교통 및 도보 통근이 높은 것으로 미루어 보아 오히려 저소득층의 낮은 교통 이동성이 그들의 거주지 선택권을 제한했을(Hu and Wang, 2017) 가능성이 있기 때문이다.

둘째, 변수의 기여도를 통해 각 소득계층의 요인별 영향력을 비교한 결과, 주택가격 및 서울 거주 여부, 직장지의 업무중심지적 특성, 응답자의 나이가 모든 소득계층의 통근시간 결정에 있어 주요한 요인으로 나타났다. 거주지와 직장지의 주택가격 차이는 모든 소득계층에서 가장 중요하게 나타났으며, 특히, 거주지의 주택가격은 고소득층에서 그 중요도가 가장 높았다. 주택가격이 고소득층의 통근시간에 더욱 민감한 이유는 고소득층으로 갈수록 직장지가 서울 3도심에 가까워지면서 통근시간의 감소보다 임대료의 상승이 더욱 가파르기(Park and Quercia, 2015) 때문이라고 보인다. 나이에 따른 통근시간의 변화는 통근수단과 밀접한 연관이 있을 것으로 보이는데, 실제로 대중교통 이용비율이 높은 젊은 응답자일수록 통근시간이 증가했다.

소득계층별로 중요도 차이가 발생한 영향요인도 존재했다. 저소득층의 경우 거주지의 대중교통 접근성이 높게 나타난 반면, 중, 고소득층은 저소득층에 비해 양호한 거주지 환경을 나타내는 요인들이 통근시간 결정에 중요하게 작용했다.

통근거리의 경우, 통근시간과 주요 변수가 대부분 비슷했으나, 거주지 및 직장지의 인구밀도가 통근거리 결정에 있어 변수의 기여도가 크게 나타났다. 교외 지역 등 상대적으로 인구 밀도가 낮은 지역은 인구밀도가 높은 도시지역보다 덜 혼잡한 경향이 있는데, 자동차로 이동할 경우 장거리 이동을 하게 됨에도 불구하고 이동 속도가 증가한다(Blumberg and King, 2019). 따라서 통근시간에는 인구밀도가 미치는 영향은 미미할 수 있으나, 거주지 및 직장지의 저밀화는 장거리 통근을 유발하는 것으로 보인다.

셋째, Partial Dependence Plot을 통해 소득계층별 통근시간과 영향요인의 비선형관계를 파악한 결과, 거주지의 주택가격과 통근시간은 상쇄관계

를 지니고 있었다. 반면, 거주지와 직장지의 주택가격 차는 0을 기준으로 U자형 그래프를 그리고 있었으며, 거주지와 직장지의 주택가격 차 분포는 고소득층으로 갈수록 넓어지고 있었다. 이는 두 가지 상반된 이유로 해석될 수 있다. 거주지가 직장지보다 저렴할 경우, 고소득층으로 갈수록 직장지가 서울 3도심에 가까이 위치해있어 직장지 인근의 임대료가 비싸지기 때문에 더욱 저렴한 주거지를 선택한 것이라고 볼 수 있다. 반대로 직장지가 거주지보다 저렴할 경우, 고소득층으로 갈수록 거주지의 환경적 질을 나타내는 특성을 우선시해 고용 접근성이 낮아지는 것을 감수하고서라도 비싼 거주지를 선택했다고 해석할 수 있다. 이는 가구의 주거입지 선택이 고용 접근성에 대한 고려와 삶의 질에 대한 고려의 상쇄관계로 설명된다는(Phe and Wakely, 2000; Kim-J et al, 2005; Moos and Skaburskis, 2010; 김승남·안건혁, 2011) 접근성-공간 맞교환 이론(Access-Space Trade-off)으로도 해석가능한데, 본 연구에서는 이러한 경향이 고소득층으로 갈수록 심화되며, 특히 서울 내에서 두드러진다는 것을 확인하였다. 서울에 거주하는 중, 고소득층의 경우 0을 기준으로 거주지가 직장지보다 저렴할 때에는 주택가격 차에 따른 통근시간의 변화가 크지 않았으나, 거주지가 직장지의 주택가격보다 비싼 경우 차이가 클수록 통근시간은 증가하는 경향이 극명하게 나타났다.

통근거리와 영향요인의 비선형관계를 살펴보면, 거주지 및 직장지의 인구밀도가 높아질수록 통근거리는 감소했다. 인구밀도가 높을 경우 교외 지역보다는 도심부에 위치할 확률이 높으며, 직장지 인근의 주거 시설이 밀집해있다는 것을 의미하기에, 거주지 및 직장지 모두 교외지역보다는 인구밀도가 높은 도심에 가까울 때 통근거리가 감소했다. 거주지 및 직장지의 인구밀도에 따른 통근거리 증감폭은 저소득층으로 갈수록 커졌다.

특히, 중, 고소득층에서는 직장지가 서울 3도심에서 멀어질수록 통근거리가 감소하다가, 서울 3도심에서 직장지가 약 15km 이상 떨어져있을 경우 다시 통근거리가 증가했다. 이는 직장의 교외화로 인해 역통근 현상이 발생한 것으로 해석가능하다. 교외지역으로 분산된 고용지는 효율적인 교통서비스 제공이 어렵고, 고속도로 위주의 접근성으로 인해 자가용 중심의 통행을

유도한다(신학철·우명제, 2020). 따라서 저소득층에 비해 중, 고소득층은 대중교통 의존도가 보다 낮기 때문에, 일자리 교외화가 이루어질 경우 자가용을 이용해 통근하며 서울 3도심에서 떨어진 직장지 인근으로 주거지를 옮기지 않을 가능성이 높다.

연구 결과를 토대로, 본 연구는 통근시간 및 통근거리 감축을 위한 정책적 시사점과 소득계층별 주거정책의 차별화 방안을 제안한다.

먼저 모든 소득계층에서 주택가격과 통근시간은 상쇄관계를 지니고, 직장지가 고용 중심지에 위치할수록 통근시간은 오히려 길어지는 양상이 나타났다. 이는 고용 중심지일수록 지가가 높을 가능성이 높고, 이에 따라 근로자는 고용접근성을 포기하고 보다 저렴한 거주지를 선택했을 가능성이 높다. 따라서 전반적인 소득계층의 통근시간 감축을 위해서는 서울 3도심 및 이외 고용 중심지에 인접한 저렴한 주택 공급이 어느 정도 필요함을 시사한다. 이는 반대로, 고용 중심지에서 지나치게 떨어진 곳에 주택 공급을 늘리는 것은 통근시간을 늘려 통근시간 감축에 부정적 영향을 줄 수 있음을 보여주기도 한다.

저소득층의 경우, 대중교통 및 도보 위주의 통근이 주를 이루고 있고, 거주지에서 가장 가까운 지하철역까지의 거리가 통근시간 결정에 중요하게 작용하였다. 따라서 저소득층의 통근시간 감축을 위해서는, 승용차 접근성을 강조한 신도시 혹은 주택 공급은 지양되어야 하며, 대중교통과 인접한 지역에 저렴한 주택 공급이 강조되어야 함을 의미한다. 실제로 거주지와 직장지 모두 저밀화되어있을 경우, 통근거리가 길어지고 있어 인구 저밀화를 유도하는 지나친 교외화 정책은 저소득층의 통근거리 감소에 부정적 영향을 끼칠 수 있음을 확인했다. 더불어, 모든 소득계층에서 나이가 젊을수록 대중교통 이용률이 증가한 것으로 보아, 이러한 주거정책은 청년들의 통근시간 감축 및 주거비 부담 완화에도 효과적일 것으로 보인다.

반면, 중, 고소득층의 경우 근린환경 개선에 관한 요인이 통근시간 증가에 영향을 미치고 있었다. 고소득층으로 갈수록 통근시간은 증가함에도 불가하고 직장지에 비해 비싼 거주지를 선택하는 경향이 심화되었다. 특히, 서울

내에서 그러한 양상이 두드러졌다. 이는 중, 고소득층으로 갈수록 통근시간 감소보다도 주거 환경 개선에 보다 관심이 많다는 것으로 해석될 수 있다.

마찬가지로, 직장지가 서울 3도심에서 약 15km 이상 떨어져 있을 경우, 고용지가 교외지역으로 분산되면서 중, 고소득층의 통근거리가 늘어나는 것을 확인했다. 교외로 이동할수록 자가용 중심의 교통체계를 지니기에 장거리임에도 불구하고 통근시간에는 큰 차이가 없을 수 있으나, 자가용을 이용한 장거리 통근 지양에 있어 지나친 고용지 교외화는 부정적 영향을 미칠 수 있음을 보여준다.

이를 방지하기 위해서는 직장지 인근의 주거 환경 개선이 중요하다고 사료된다. 따라서 주거 환경이 우수한 지역을 업무중심지에서 지나치게 떨어진 곳에 개발하는 것은 통근시간에 부정적 영향을 미칠 수 있으며, 인근 주거 환경을 고려하지 않은 업무중심지 개발 또한 오히려 통근시간 및 통근거리를 증가시킬 수 있음을 시사한다.

제 2 절 연구의 의의 및 한계

본 연구는 우수한 설명력을 지닌 XGBoost를 사용하여 소득계층별 통근시간 및 통근거리 영향요인 차이를 세밀하게 비교 분석했다는데 의의가 있다. 주택가격을 포함한 거주지 특성, 직장지 특성 및 개인 변수와 같은 다양한 요인들을 활용하여 소득계층별 통근시간 및 통근거리 결정에 주요한 영향요인을 파악했고, 각 영향요인이 통근시간 및 통근거리에 어떤 비선형관계를 지니는지 확인했다. 이에 더해, 연구의 결과가 통근시간 및 통근거리 감축을 위한 주거 정책에 실질적으로 활용될 수 있도록 소득계층별 주거정책의 차별화 방안을 제시하여 연구의 의의를 더했다.

한편, 본 연구는 다음과 같은 점에서 한계를 지닌다. 첫째, 본 연구에서 사용한 가구통행실태조사의 소득 자료의 경우 범주형으로 구성되어 있기 때문에 소득계층을 보다 세밀하게 구분하는데 한계가 있었다. 또한 가구 평균 월 소득으로 소득계층을 구분하기 때문에, 정기적으로 발생하는 소득이 아닌

가구의 자산 등을 계층 구분에 포함하지 못하였다. 이와 같은 이유로 평균 월 소득은 적으나 자산을 많이 보유하고 있는 일부 계층이 과소평가되었을 수 있다.

둘째, 가구통행실태조사 자료에 기입된 응답자의 정보는 행정동 단위로 거주지 및 직장지가 표시되어 있어 주택유형별, 행정동별 주택가격을 구분하였음에도 불구하고 개별 응답자의 주택가격 및 가장 가까운 지하철역까지의 거리와 같은 지역변수가 보다 일반화되어 적용되었을 수 있다.

셋째, XGBoost 계수의 신뢰수준을 보다 정확하게 확인하기 위해서는, 부트스트랩, 반응 변수의 조건부 분위값을 추정하는(Rodriguez and Yao, 2017; Nzarigema et al., 2022) 분위회귀모델 등이 제안된다. 그러나 본 연구에서는 해당 방안을 적용하여 계수의 신뢰수준을 확인하지 않았다는데 한계가 있다. 본 연구에서는 XGBoost 및 다른 기계학습 모형과의 비교, R-Squared, MAE, RMSE 등의 평가지표로 XGBoost 모형의 신뢰수준을 확인했다.

참 고 문 헌

김승남, 안건혁. (2011). 재택근무와 주거입지의 관계에 대한 실증연구: 수도권 거주 임금근로자를 중심으로. 국토계획, 46(7), 37-55.

김태형, 고준호. (2016). 대도시 토지이용 압축도 지표의 개발 및 적용: 서울시를 대상으로. 서울도시연구, 17(1), 1-21.

김호기, 유상균, & 정창무. (2010). 구매자 소득계층별 대형 상업시설의 선택요인 분석: 매장면적과 통행시간에 의한 영향력을 중심으로. 국토계획, 45(3), 77-92.

김희철, 안건혁. (2011). 압축도시 계획요소가 소득계층별 통근거리에 미치는 영향. 한국도시설계학회지 도시설계, 12(1), 55-70.

노승철, 이희연. (2009). 주거이동을 통한 주거 불안정성 변화에 관한 연구-저소득층을 대상으로 하여. 한국경제지리학회지, 12(4), 507-520.

박미선, 김호정, 강미나, 김동형, & 권기현. (2018). 교통비용을 고려한 주거부담 지표의 개발 및 활용방안 연구. 국토연구원

박준상, 이수기. (2022). 해석가능한 기계학습을 활용한 보행목적별 보행만족도 영향요인 분석. Journal of Korea Planning Association-Vol, 57(1), 26-41.

신학철, 우명제. (2020). 공간 미스매치를 고려한 역통근 현상에 관한 연구-비수도권 광역대도시권을 대상으로. 지역연구, 36(2), 3-12.

송재민. (2021). 도시형태가 통행행태에 미치는 영향 메타분석. 국토계획, 56(7), 103-114.

오인철. (2009). 지대이론과 토지이용에 대한 이론적 고찰. 부동산학보, 37, 264-277.

오지예, 이영호, 정예원, & 홍성연. (2019). 가구통행실태조사 자료를 활용한 소득계층별 공간적 분리 탐색. 교통연구, 26(1), 17-29.

유진은. (2017). 기계학습을 통한 TIMSS 2011 중학생의 수학 성취도 관련 변수 탐색. 교원교육, 33(1), 43-56.

이민주, 박인권. (2016). 지역 특성에 따른 소득별 직주불일치에 관한 연구. 지역연구, 32(1), 67-82.

이번송. (1998). 서울 거주자의 통근거리 결정요인 분석. 국토계획, 33(3), 241-263.

이혜승, 이희연. (2009). 서울시 대중교통체계 개편 이후 통근 교통수단 선택의 차별적 변화. 대한지리학회지, 44(3), 323-338.

장재민, 김태형. (2016). 통근시간 변화에 영향을 미치는 개인 및 지역변수 분석: 서울시 25 개 자치구를 중심으로 (2006 ~ 2014). 서울도시연구, 17(3), 105-116.

장재민, 이병호, & 고준호. (2019). 근로자의 통근시간 만족도 결정요인 연구: 경기도 거주자를 중심으로. 대한교통학회지, 37(4).

전명진, 안현주. (2016). 수도권 통근자의 주거비용과 통근비용의 상쇄관계에 대한 실증분석. 한국지역개발학회지, 28(2), 25-40.

전명진, 이지현. (2020). 수도권 통근자의 직장 및 주택접근성이 통근시간에 미치는 영향에 대한 분석. 한국지역개발학회지, 32(2), 119-135.

최준성. (2020). 수도권 신도시의 통근비용을 고려한 주거비용 분석에 관한 연구. 국내석사학위논문 건국대학교 부동산대학원, 서울

최준성, 노승한. (2021). 수도권 신도시의 통근비용을 고려한 주거비용 분석에 관한 연구. 주택연구, 29, 93-116.

하재현. (2020). 도시공간구조가 소득수준별 통근거리에 미치는 영향에 관한 연구 (Doctoral dissertation, 한양대학교).

홍성조, 안건혁. (2011). 소득계층별 주거이동과정에 관한 연구: 수도권내 아파트 거주자를 대상으로. 한국도시설계학회지 도시설계, 12(3), 91-100.

Adair, A., McGreal, S., Smyth, A., Cooper, J., & Ryley, T. (2000). House prices and accessibility: The testing of relationships within the Belfast urban area. Housing Studies, 15(5), 699-716.

Alonso, W. (1964). Location and Land Use: Toward a General Theory of Land Rent. Harvard University Press. Cambridge.

Blumenberg, E., & King, H. (2019). Low-income workers, residential location, and the changing commute in the United States. Built Environment, 45(4), 563-581.

Blumenberg, E., & King, H. (2021). Jobs-housing balance re-revisited. Journal of the American Planning Association, 87(4), 484-496.

Breiman, L., Friedman, J., Stone, C.J., Olshen, R.A., (1984). Classification and

Regression Trees. CRC Press, Florida.

Chen, T., Guestrin, C., (2016). Xgboost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785–794.

Ding, C., Cao, X., & Liu, C. (2019). How does the station-area built environment influence Metrorail ridership? Using gradient boosting decision trees to identify non-linear thresholds. Journal of Transport Geography, 77, 70-78.

Ding, C., Cao, X., & Wang, Y. (2018a). Synergistic effects of the built environment and commuting programs on commute mode choice. Transportation Research Part A: Policy and Practice, 118, 104-118.

Ding, C., Cao, X. J., & Næss, P. (2018b). Applying gradient boosting decision trees to examine non-linear effects of the built environment on driving distance in Oslo. Transportation Research Part A: Policy and Practice, 110, 107-117.

Ding, C., Wu, X.K., Yu, G.Z., & Wang, Y.P. (2016). A gradient boosting logit model to investigate driver's stop-or-run behavior at signalized intersections using high-resolution traffic data. Transp. Res. Part C-Emerg. Technol. 72, 225–238. https://doi.org/10.1016/j.trc.2016.09.016.

Friedman, J.H., (2001). Greedy function approximation: a gradient boosting machine. Ann. Stat. 1189–1232.

Haas, A., & Osland, L. (2014). Commuting, migration, housing and labour markets: Complex interactions. Urban Studies, 51(3), 463-476.

Islam, M. R., & Saphores, J. D. M. (2022). An LA story: The impact of housing costs on commuting. Journal of Transport Geography, 98, 103266.

Jean d'Amour Nzarigema, Jack Ngarambe, 조정훈, 윤근영. (2022). 건물의 에너지 벤치마크를 위한 LightGBM기반 분위 회귀 모델. 한국건축친환경설비학회 논문집, 16(5), 359-373.

Kim, J., F. Pagliara and J. Preston. 2005. "The Intention to Move and Residential Location Choice Behaviour". Urban Studies vol.42, no.9.

pp1621-1636.

Liu, J., Wang, B., & Xiao, L. (2021). Non-linear associations between built environment and active travel for working and shopping: An extreme gradient boosting approach. Journal of Transport Geography, 92, 103034.

Ma, X., Ding, C., Luan, S., Wang, Y., & Wang, Y. (2017). Prioritizing influential factors for freeway incident clearance time prediction using the gradient boosting decision trees method. IEEE Trans. Intell. Transp. Syst. 18 (9), 2303–2310.

Mills, E. S. (1972). Studies in the Structure of the Urban Economy. Resources for the Future, Inc.

Molho, I. (1986) Theories of migration: a review, Scottish Journal of Political Economy, 33(4), 396–419.

Muth, R. F. (1969). Cities And Housing; The Spatial Pattern Of Urban Residential Land Use. 49(3). 305-355.

Moos, M. and A. Skaburskis. 2007. "The Characteristics and Location of Home Workers in Montreal, Toronto and Vancouver". Urban Studies vol.44, no.9. pp1781-1808.

Parkes, A. A. Kearns and R. Atkinson. 2002. "What Makes People Dissatisfied with their Neighborhoods?". Urban Studies vol.39, no.13. pp2413-2438.

Park, K. A., & Quercia, R. (2015). Housing costs and commuting distance. Journal of the Center for Real Estate Studies.

Phe, H. H. and P. Wakely. 2000. "Status, Quality and the Other Trade-off: towards a New Theory of Urban Residential Location". Urban Studies vol.37, no.1. pp7-35.

Rapoport, A. 1980. "Environmental Preference, Habitat Selection and Urban Housing". Journal of Social Issues vol.36, no.3. pp118-134.

Rodriguez, R.N., Yao, Y. (2017, April). Five things you should know about quantile regression. In Proceedings of the SAS Global Forum 2017 Conference, Orlando, FL, USA, 2-5.

Rouwendal, J. and E. Meijer. 2001. "Preferences for Housing, Jobs, and Commuting: a Mixed Logit Analysis". Journal of Regional Science vol.41, no.3. pp475-505.

Tyrvainen, L. and H. Vaananen. 1998. "The Economic Value of Urban Forest Amenities: an Application of the Contingent Valuation Method". Landscape and Urban Planning vol.43, no.1-3. pp105-118.

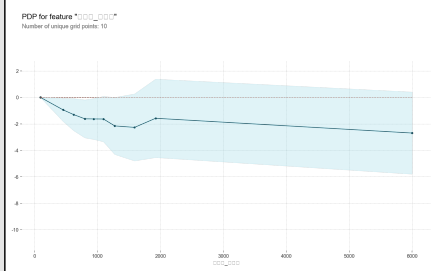
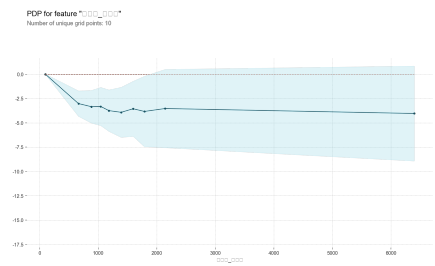
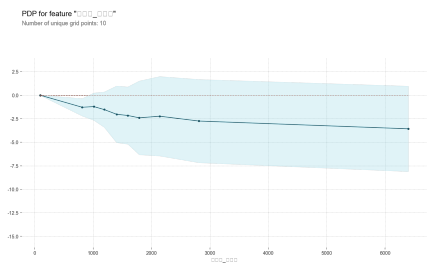
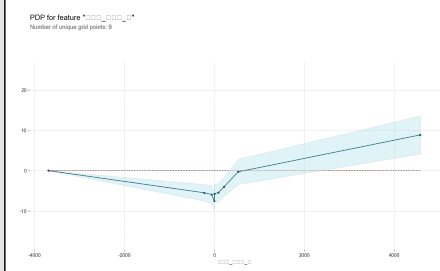
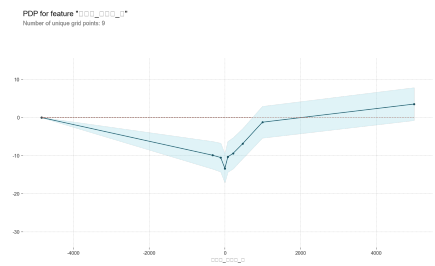
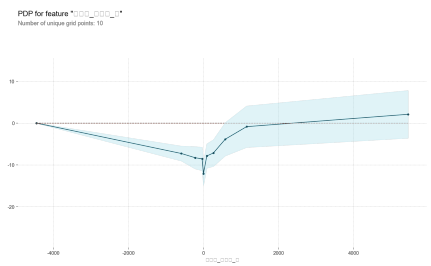
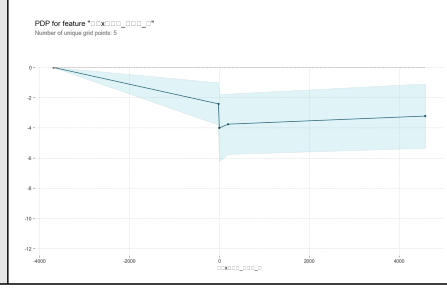
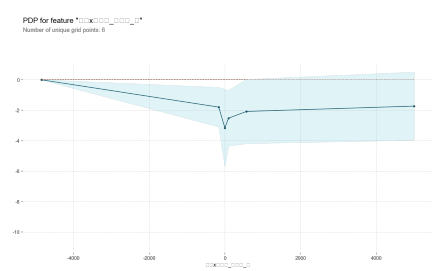
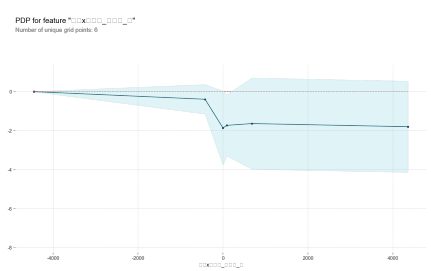
Weisbrod, G. E., S. R. Lerman and M. Ben-Akiva. 1980. "Tradeoffs in Residential Location Decisions: Transportation Versus Other Factors". Transport Policy and Decision Making vol.1, no.1. pp13-26.

Zhang, Y., & Haghani, A. (2015). A gradient boosting method to improve travel time prediction. Transp. Res. Part C: Emerg. Technol. 58, 308–324.

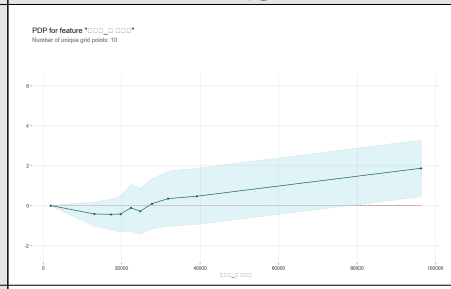
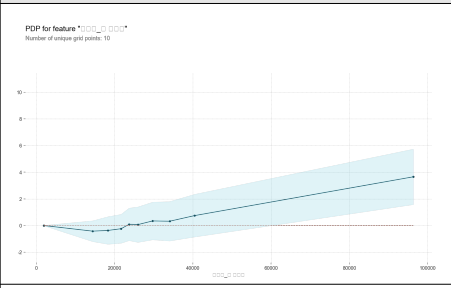
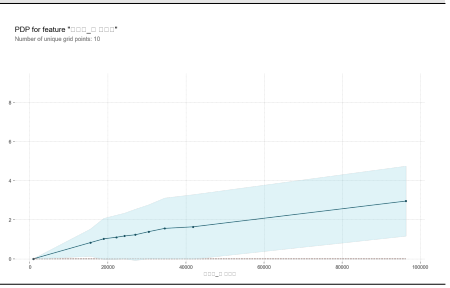
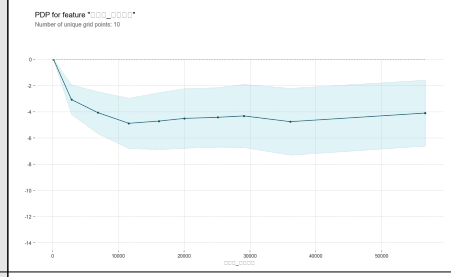
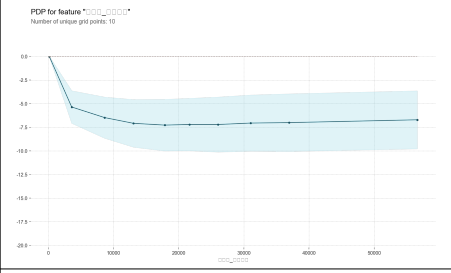
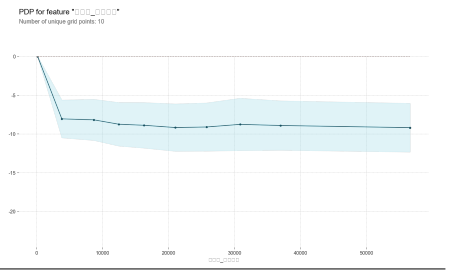
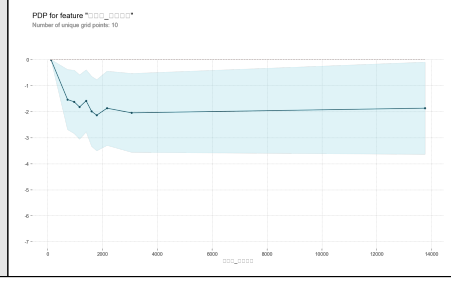
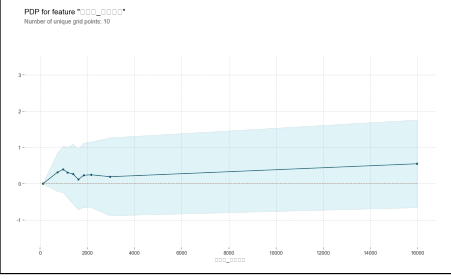
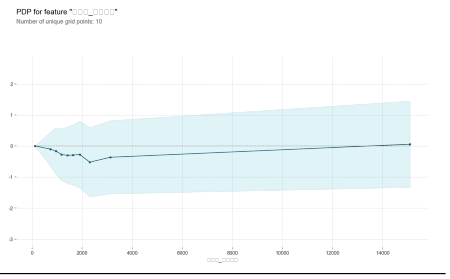
Zheng, Z., Zhou, S., & Deng, X. (2022). The spatially heterogeneous and double-edged effect of the built environment on commuting distance: Home-based and work-based perspectives. PloS one, 17(3), e0262727.

부록
Appendix

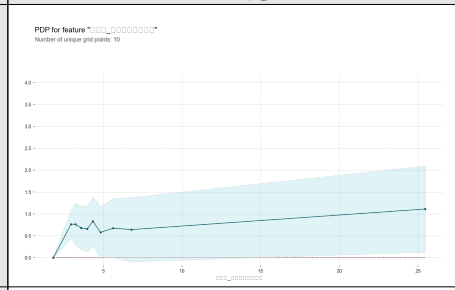
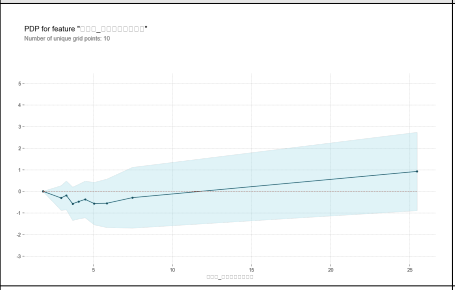
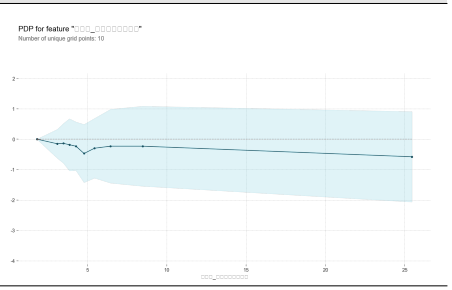
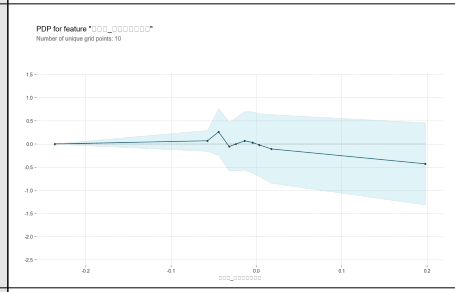
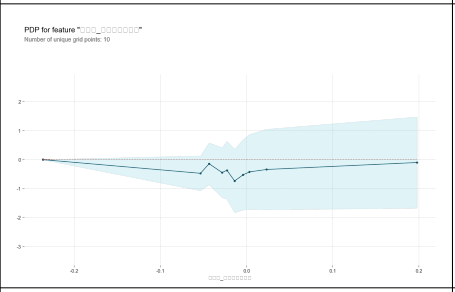
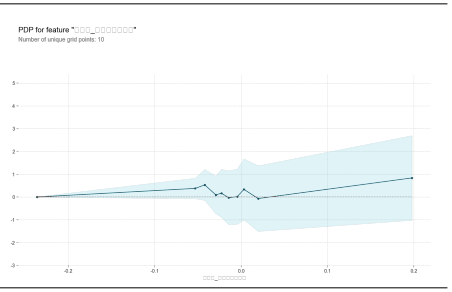
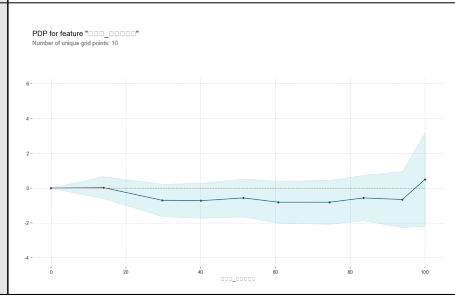
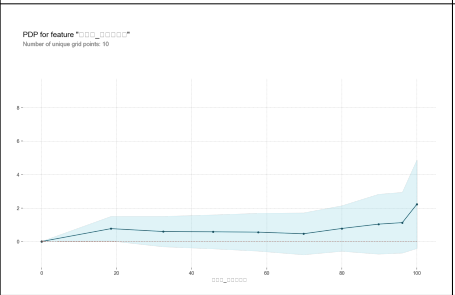
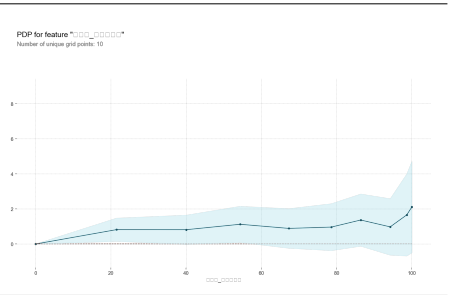
[표 1] 소득계층별 통근거리에 영향을 미치는 주택가격 변수 Partial Dependence Plot

	저소득층	중소득층	고소득층
거주지의 주택가격			
거주지와 직장지의 차			
interaction term			

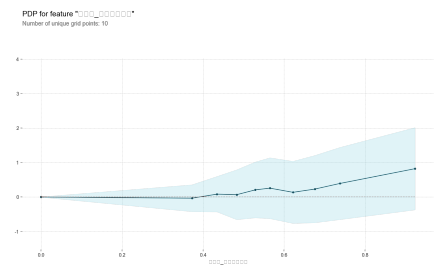
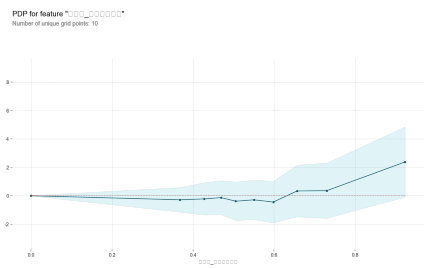
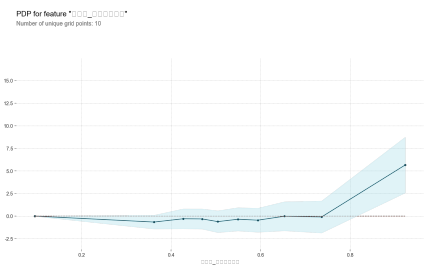
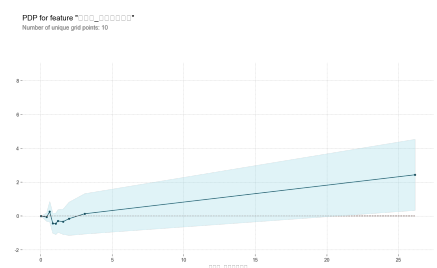
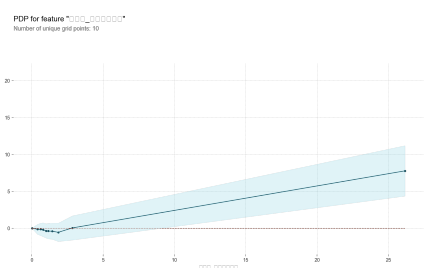
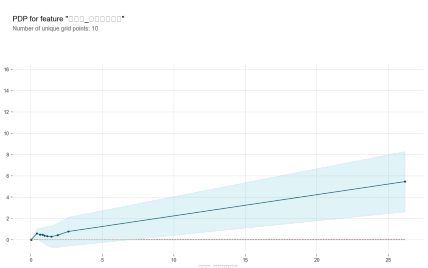
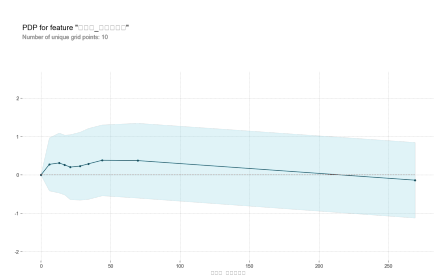
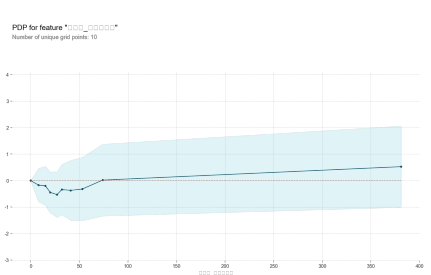
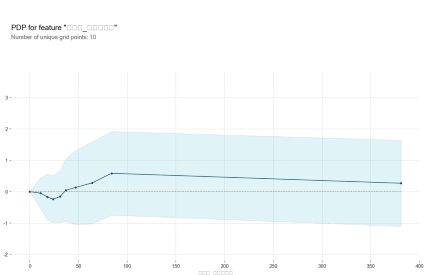
[표 2] 소득계층별 통근거리에 영향을 미치는 거주지역 변수 Partial Dependence Plot

	저소득층	중소득층	고소득층
거주지의 총 인구수			
거주지의 인구밀도			
거주지의 사업체수			

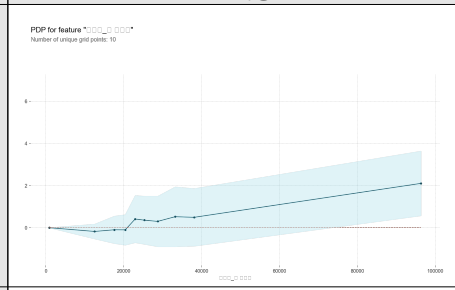
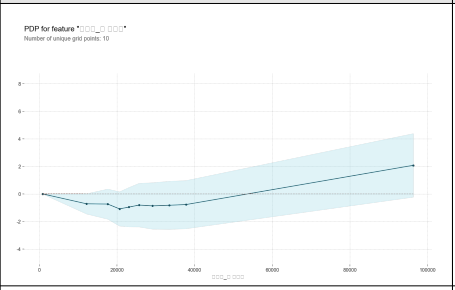
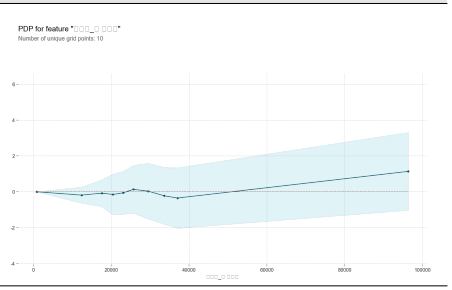
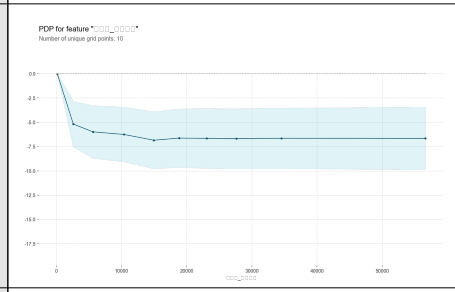
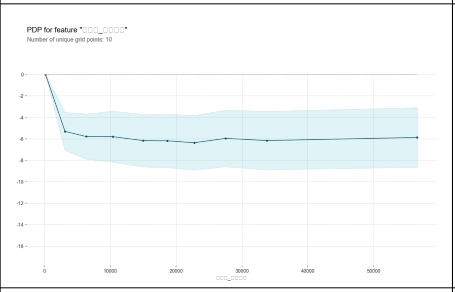
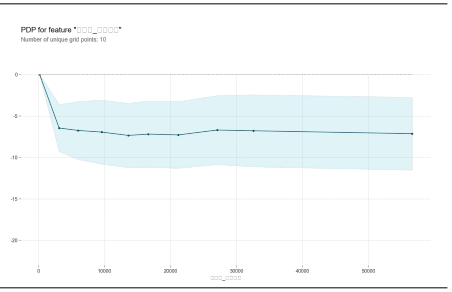
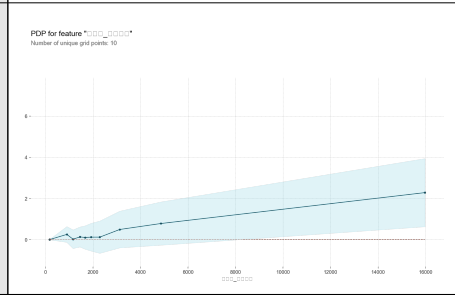
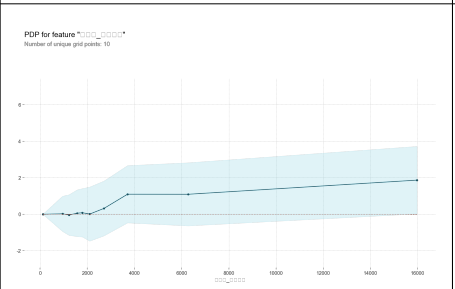
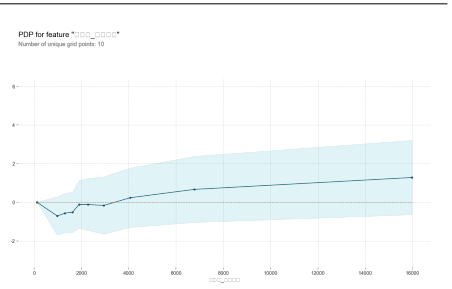
[표 3] 소득계층별 통근거리에 영향을 미치는 거주지역 변수 Partial Dependence Plot

	저소득층	중소득층	고소득층
거주지의 사업체당 종사자수			
거주지의 주택가격 증가율			
거주지의 아파트비율			

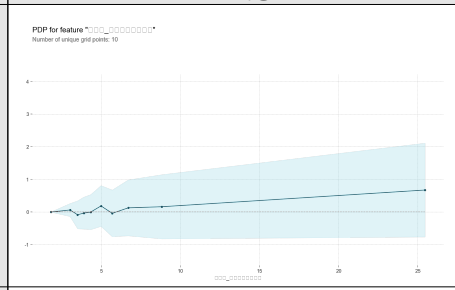
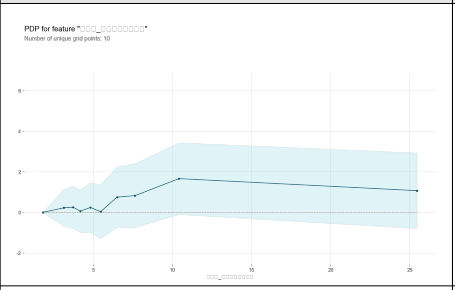
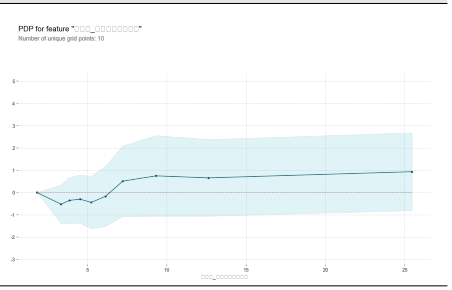
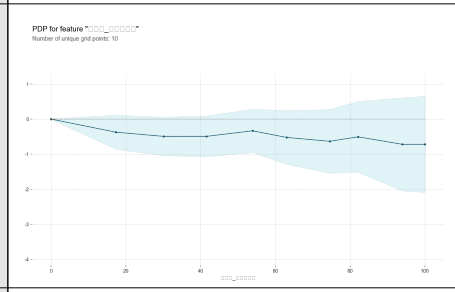
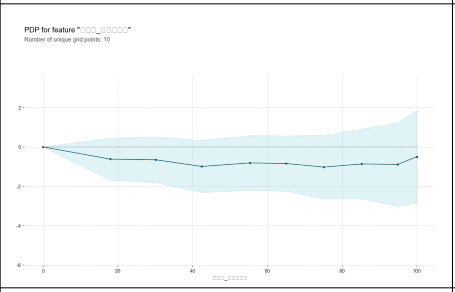
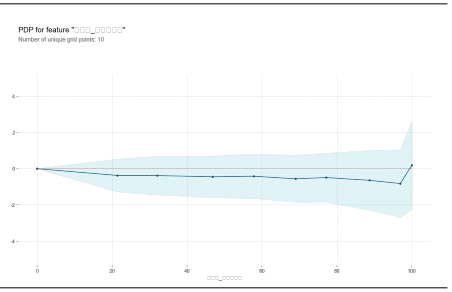
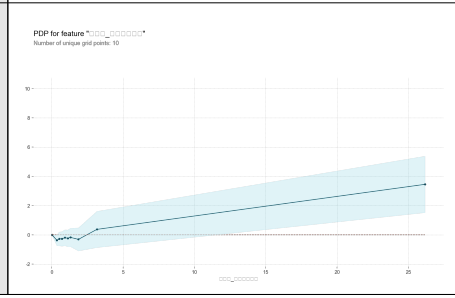
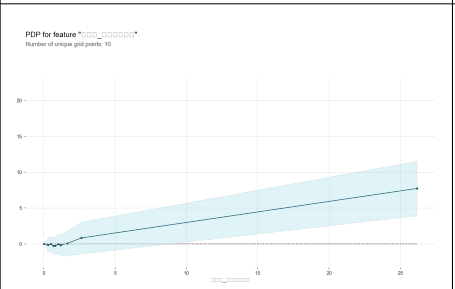
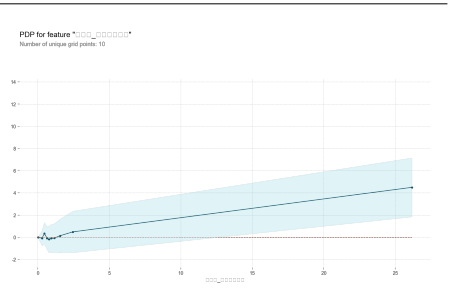
[표 4] 소득계층별 통근거리에 영향을 미치는 거주지역 변수 Partial Dependence Plot

	저소득층	중소득층	고소득층
거주지의 토지이용 혼합			
거주지에서 가장 가까운 지하철역 까지의 거리			
거주지의 사설학원 수			

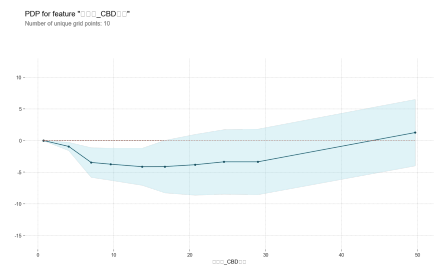
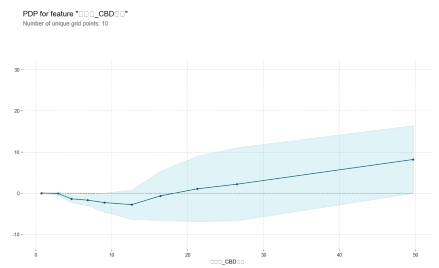
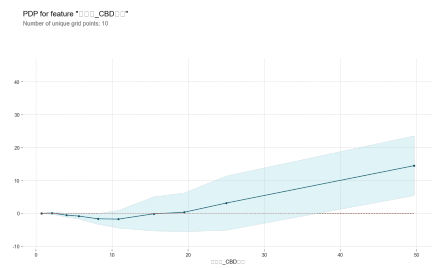
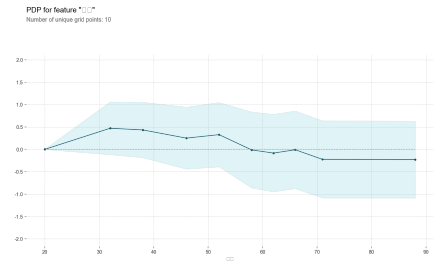
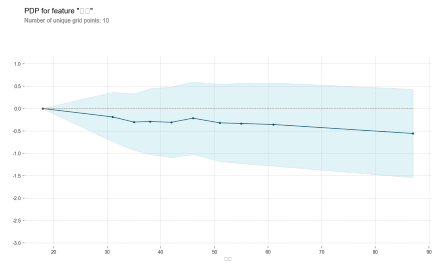
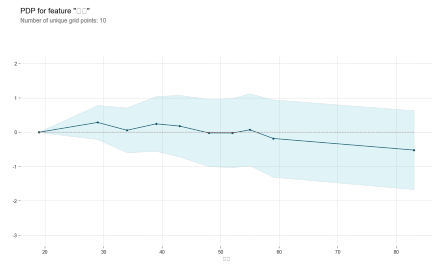
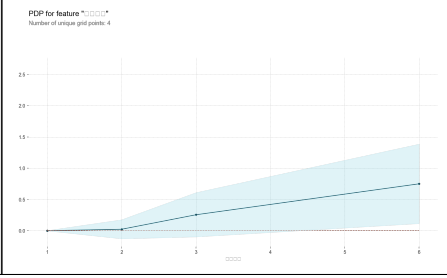
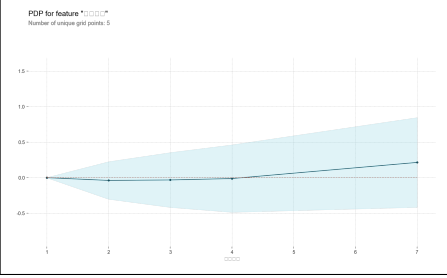
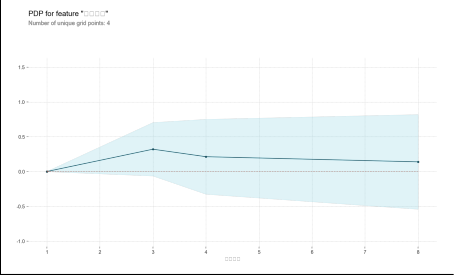
[표 5] 소득계층별 통근거리에 영향을 미치는 직장지역 변수 Partial Dependence Plot

	저소득층	중소득층	고소득층
직장지의 총 인구수			
직장지의 인구밀도			
직장지의 사업체수			

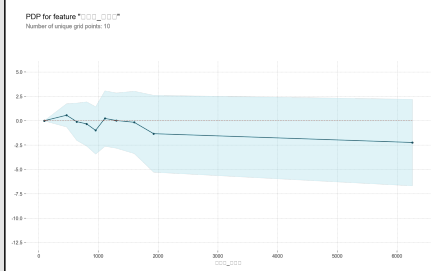
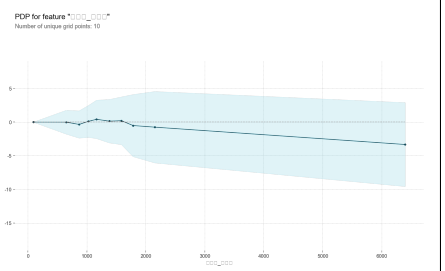
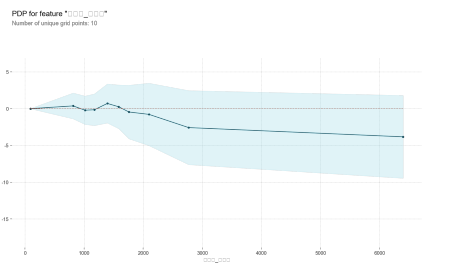
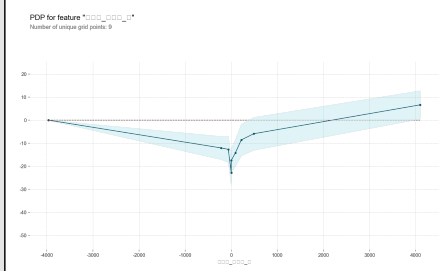
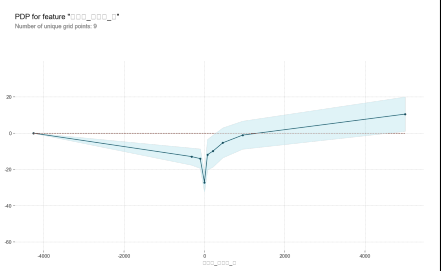
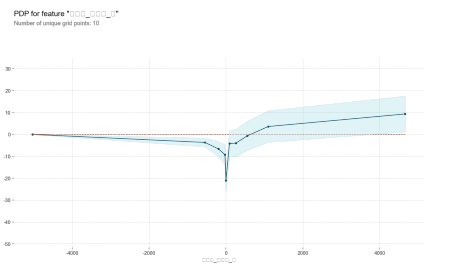
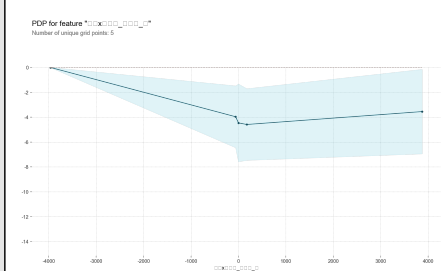
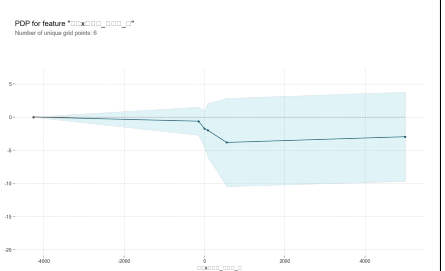
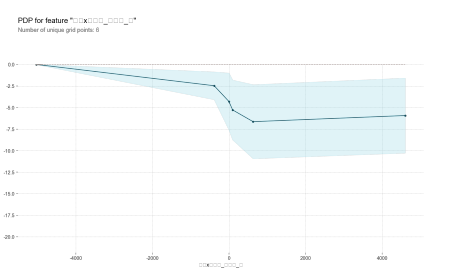
[표 6] 소득계층별 통근거리에 영향을 미치는 직장지역 변수 Partial Dependence Plot

	저소득층	중소득층	고소득층
직장지의 사업체당 종사자수			
직장지의 아파트비율			
직장지에서 가장 가까운 지하철역 까지의 거리			

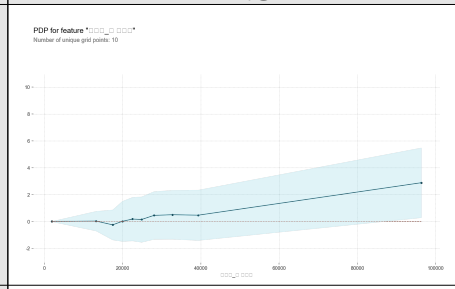
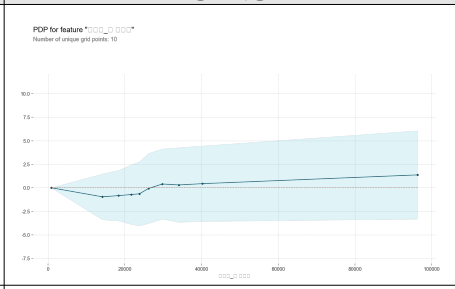
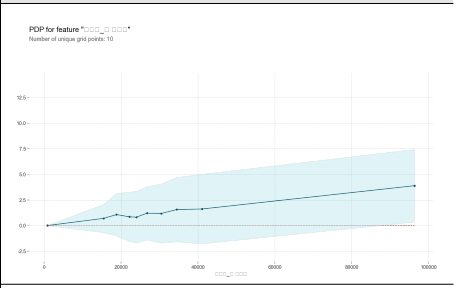
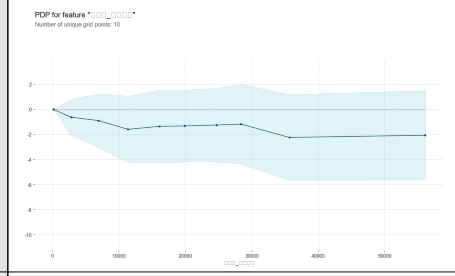
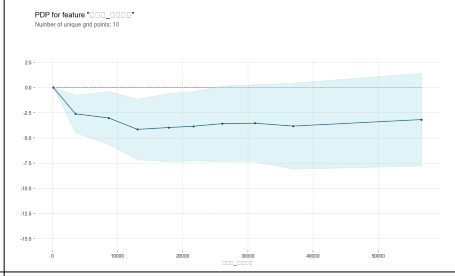
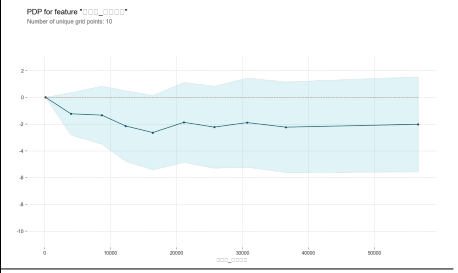
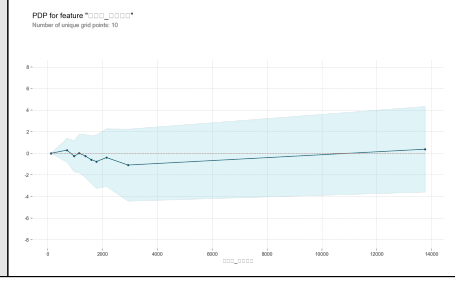
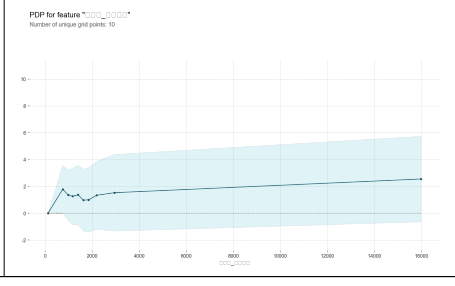
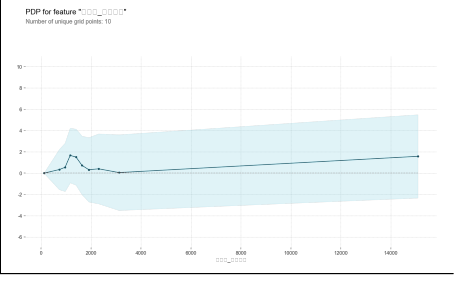
[표 7] 소득계층별 통근거리에 영향을 미치는 직장지역 및 개인 변수 Partial Dependence Plot

	저소득층	중소득층	고소득층
직장지에서 가장 가까운 CBD 로부터의 거리			
응답자의 나이			
응답자의 가구원수			

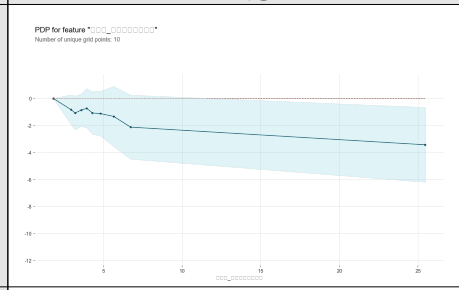
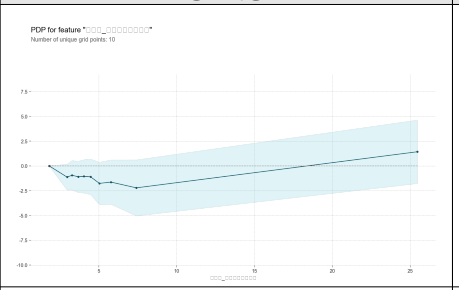
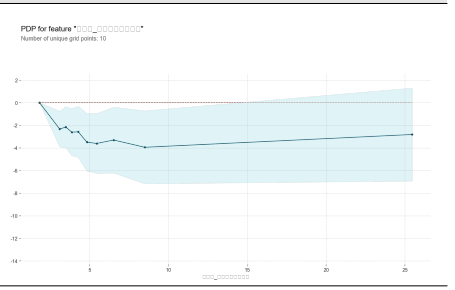
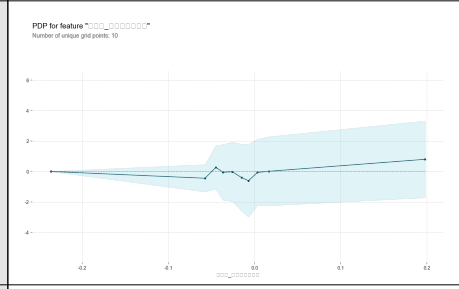
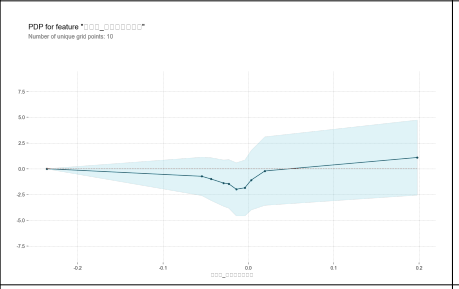
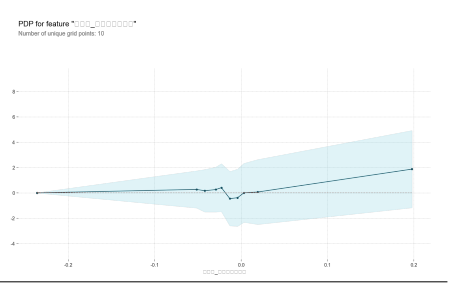
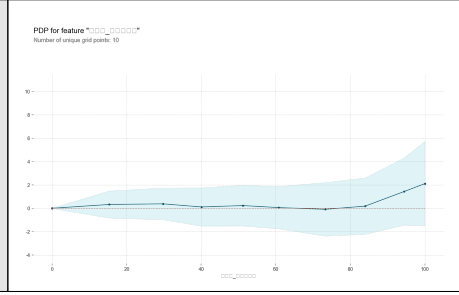
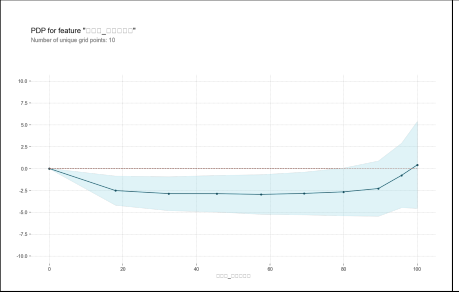
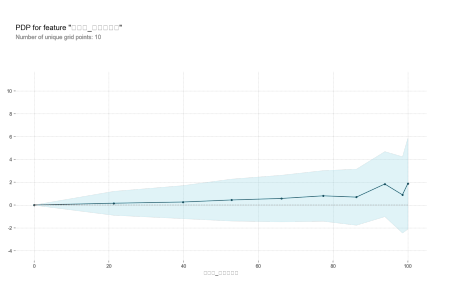
[표 8] 소득계층별 통근시간에 영향을 미치는 주택가격 변수 Partial Dependence Plot

	저소득층	중소득층	고소득층
거주지의 주택가격			
거주지와 직장지의 차			
interaction term			

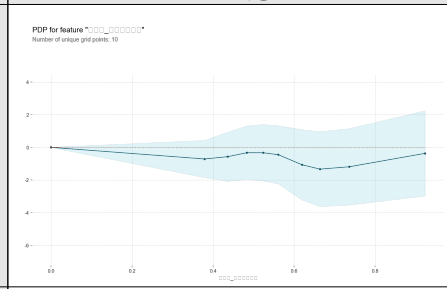
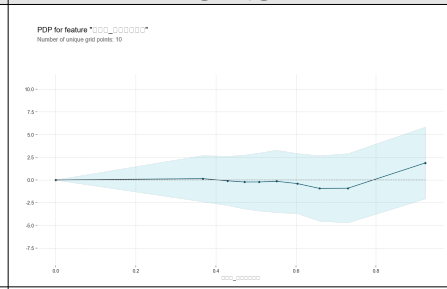
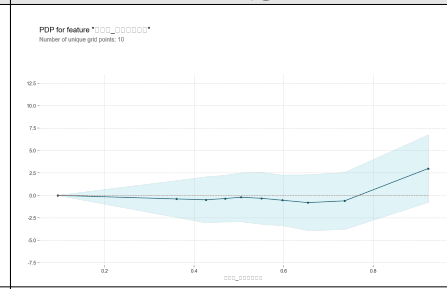
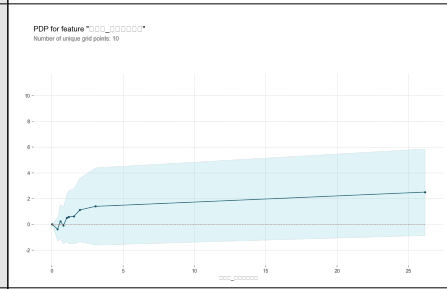
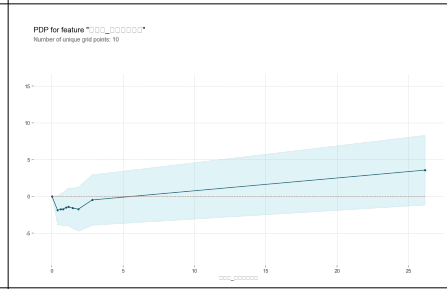
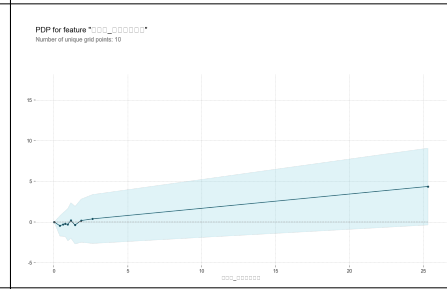
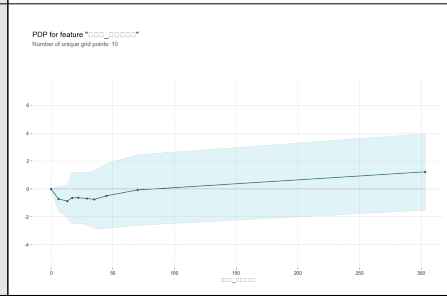
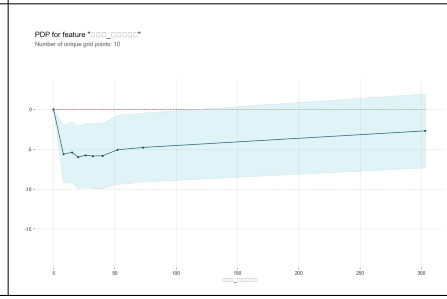
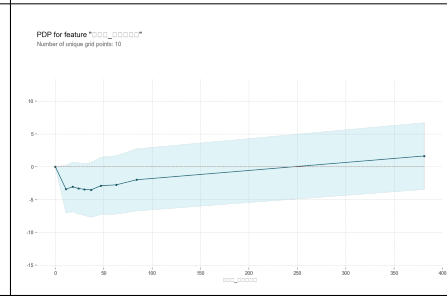
[표 9] 소득계층별 통근시간에 영향을 미치는 거주지역 변수 Partial Dependence Plot

	저소득층	중소득층	고소득층
거주지의 총 인구수			
거주지의 인구밀도			
거주지의 사업체수			

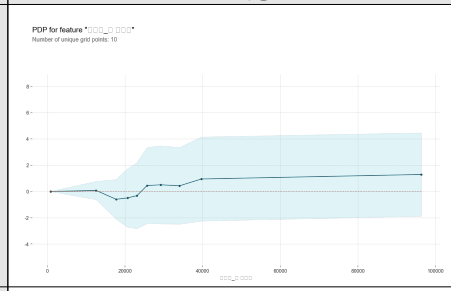
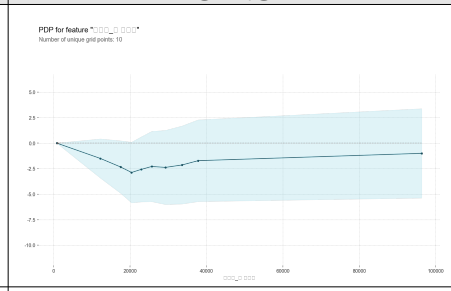
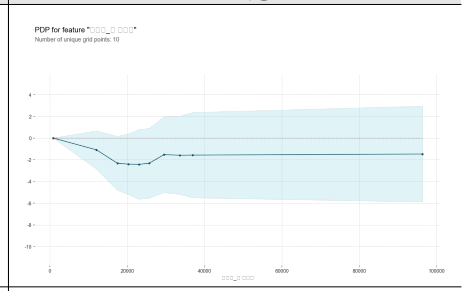
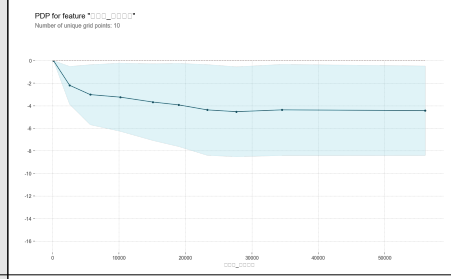
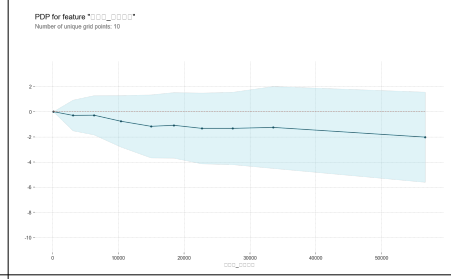
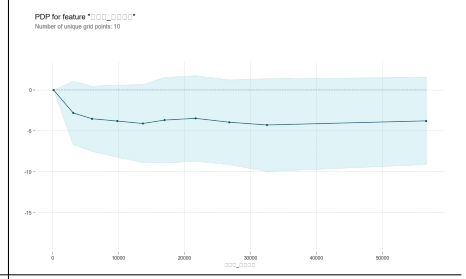
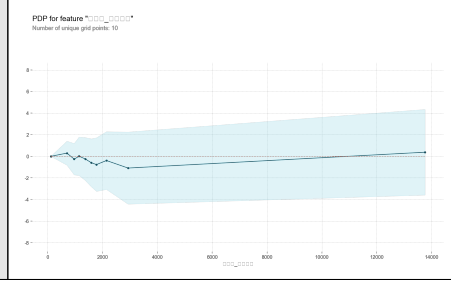
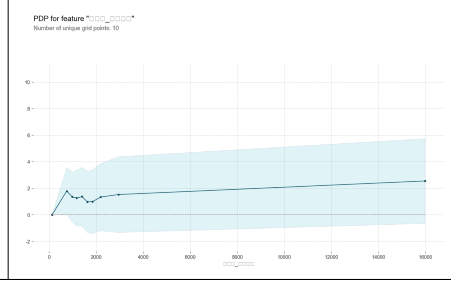
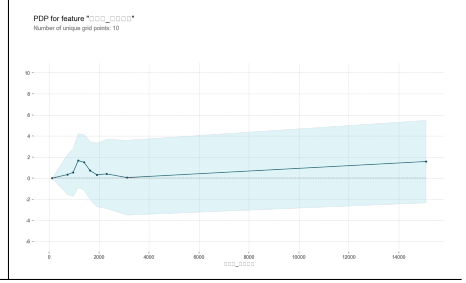
[표 10] 소득계층별 통근시간에 영향을 미치는 거주지역 변수 Partial Dependence Plot

	저소득층	중소득층	고소득층
거주지의 사업체당 종사자수			
거주지의 주택가격 증가율			
거주지의 아파트비율			

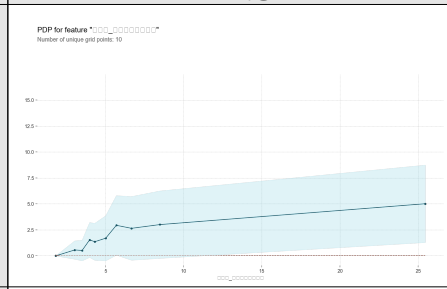
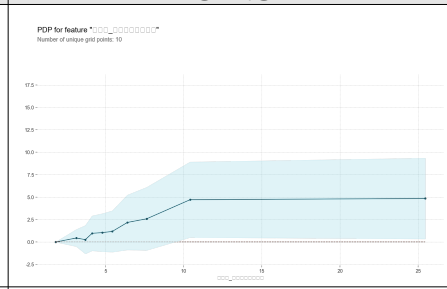
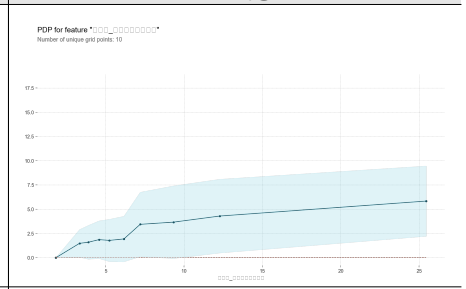
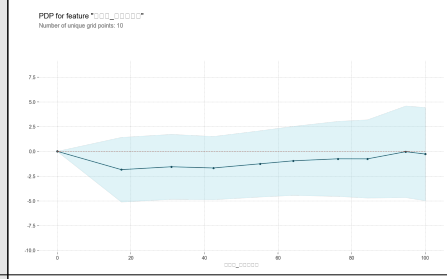
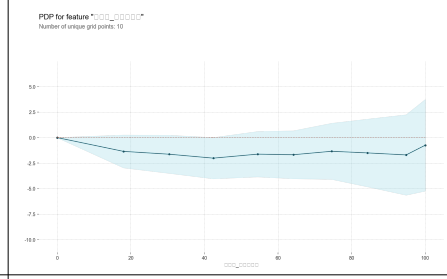
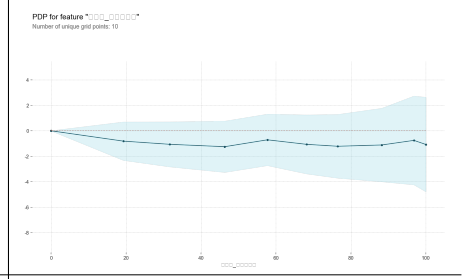
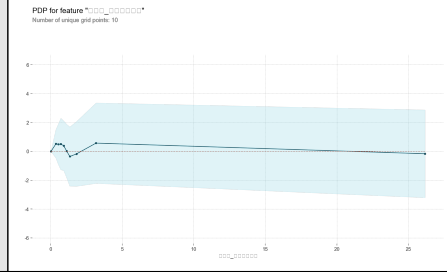
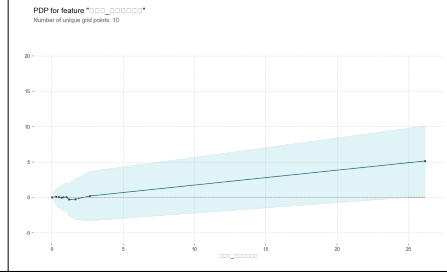
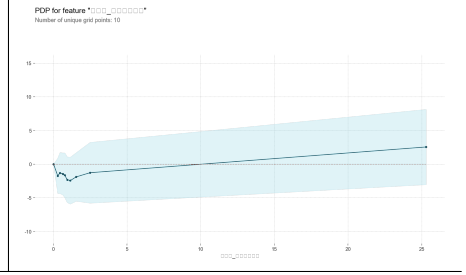
[표 11] 소득계층별 통근시간에 영향을 미치는 거주지역 변수 Partial Dependence Plot

	저소득층	중소득층	고소득층
거주지의 토지이용 혼합			
거주지에서 가장 가까운 지하철역 까지의 거리			
거주지의 사설학원 수			

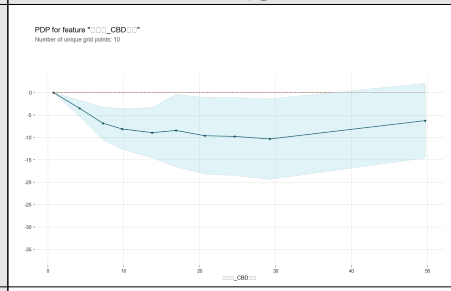
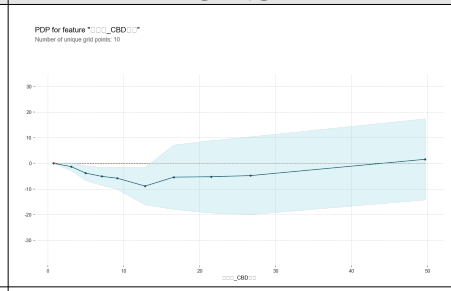
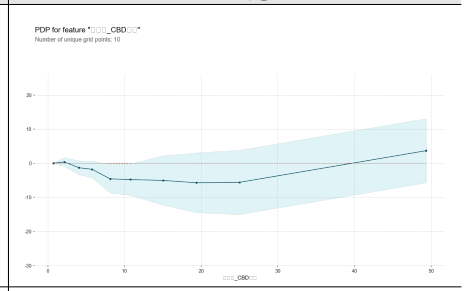
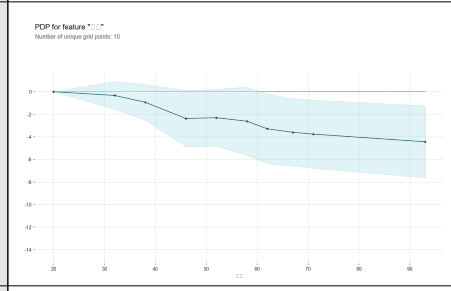
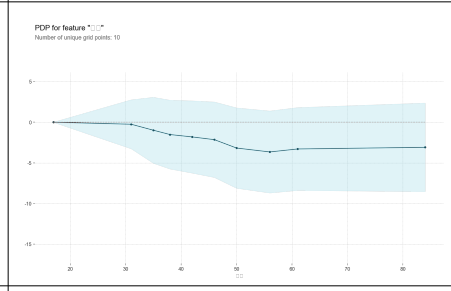
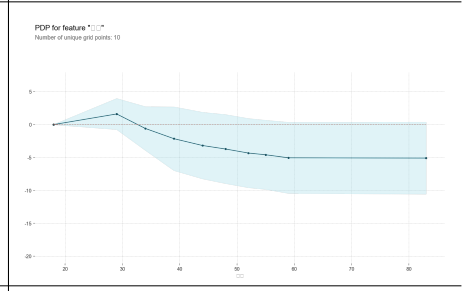
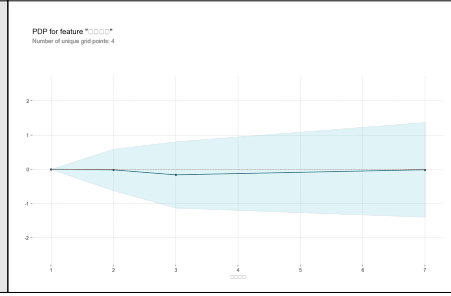
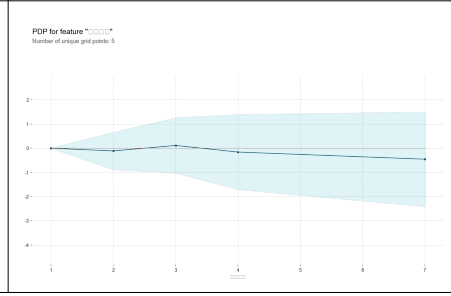
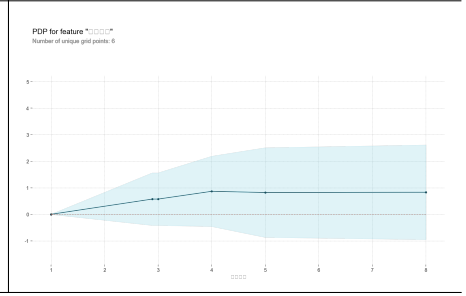
[표 12] 소득계층별 통근시간에 영향을 미치는 직장지역 변수 Partial Dependence Plot

	저소득층	중소득층	고소득층
직장지의 총 인구수			
직장지의 인구밀도			
직장지의 사업체수			

[표 13] 소득계층별 통근시간에 영향을 미치는 직장지역 변수 Partial Dependence Plot

	저소득층	중소득층	고소득층
직장지의 사업체당 종사자수			
직장지의 아파트비율			
직장지에서 가장 가까운 지하철역 까지의 거리			

[표 14] 소득계층별 통근시간에 영향을 미치는 직장지역 및 개인 변수 Partial Dependence Plot

	저소득층	중소득층	고소득층
직장지에서 가장 가까운 CBD 로부터의 거리			
응답자의 나이			
응답자의 가구원수			

Abstract

Study on the Influencing Factors on Commuting Time and Commuting Distance by Income Class Using Machine Learning

Sungjoo Kim
Urban and Regional Planning
Department of Environmental Planning
The Graduate School
Seoul National University

Reducing commuting time and distance is a major spatial strategy to reduce carbon emissions and an important way to improve the quality of life of workers. This study analyzes the differences in influencing factors of commuting time and distance by income class using Household Travel Survey Data in Seoul Metropolitan Area with machine learning methodology, XGBoost. Various factors such as housing price, residential characteristics, workplace characteristics, and personal variables were used to

identify major influencing factors in determining commuting time and distance by income class, and recognize nonlinear relationship between each influencing factor and commuting time and distance.

According to the results, housing prices and commuting time/distance have an trade-off relationship in all income classes, and commuting time is increased as the workplace is nearly located in the employment center. However, as the workplace moved away from the CBD in Seoul to a certain extent, the commuting distance between middle and high-income class is rather longer. In the case of low-income class, public transportation and the distance from the residence to the subway station played an important role in determining the commuting time. In middle and high-income groups, the influencing factors related to the improvement of the residential environment affect the increase in commuting time. The higher the income group lives in Seoul, the more likely to choose a residence more expensive than their workplace. Based on the results of the study, it suggests that the supply of affordable housing adjacent to the employment centers would be effective to reduce commuting time for the overall income class. To reduce commuting time for low-income level, low-cost housing supply should be emphasized in areas adjacent to high-density neighbourhoods and public transportation. In the case of middle and high-income level, it would be important to improve the residential environment near the workplace to reduce commuting time. Also, suburbanization of workplace could increase commuting distance using automobiles.

keywords : Housing Price, Residential Characteristics, Commuting Time, Commuting Distance, Machine Learning
Student Number : 2021-22630