



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

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

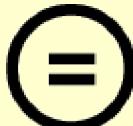
다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원 저작자를 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 권리와 책임은 위의 내용에 의하여 영향을 받지 않습니다.

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

[Disclaimer](#)



경북대학교 공학석사학위논문

빅데이터 기반의 아파트 가격
형성 요인 분석 및 가격 예측을 위한
다중회귀모델 연구
-대구광역시 법정동 중심으로-

대학원 건축학과 건축계획전공

김 선 현

2021년 12월

경북대학교 대학원

빅데이터 기반의 아파트 가격
형성 요인 분석 및 가격 예측을 위한
다중회귀모델 연구
-대구광역시 법정동 중심으로-

이 논문을 공학석사 학위논문으로 제출함

대학원 건축학과 건축계획전공

김 선 현

지도교수 추 승 연

김선현의 공학석사 학위논문을 인준함

2021년 12월

위원장 _____

경북대학교 대학원위원회

■ 목 차

제 1 장 서 론

1.1. 연구의 배경 및 목적	2
1.2. 연구의 방법 및 범위	7

제 2 장 다중회귀분석과 아파트 가격에 대한 선행연구

2.1. 다중회귀분석의 개념 및 선행연구 분석	10
2.1.1 다중회귀분석에 대한 개념	10
2.1.2 다중회귀분석에 대한 선행연구 분석	13
2.2 아파트 가격 형성 요인 선행연구 분석	17

제 3 장 가격 형성 요인 빅데이터 수집 및 변수 검증

3.1 가격 형성 요인 선정과 빅데이터 수집	23
3.1.1 독립변수 선정 및 빅데이터 수집	23
3.1.2 대구광역시 공시가격 및 실거래가격 빅데이터 수집	27
3.2 데이터 분류 및 변수 검증	28

제 4 장 아파트 가격 형성 요인 분석 및 가격 예측

4.1 공시가격의 요인 분석 및 통계학적 검증	32
4.1.1 공시가격과 독립변수의 상관관계 분석	32
4.1.2 공시가격의 다중회귀분석 및 요인 도출	35
4.2 실거래가격의 요인 분석 및 통계학적 검증	38
4.2.1 실거래가격과 독립변수의 상관관계 분석	38
4.2.2 실거래가격의 다중회귀분석 및 요인 도출	40
4.3 다중회귀모델의 검증 및 가격 예측	44

제 5 장 결 론

■ 참고문헌	58
■ 영문초록	59

■ 그 림 목 차

[그림 1-1] 2019.12-2020.12월 전국 아파트 매매가격 지수	3
[그림 1-2] 2011.-2020년 1인당 GNI(국민총소득)	3
[그림 1-3] 2015-2020년 7대 특·광역시 아파트 증가율	4
[그림 1-4] 현재 거주중인 아파트를 선택한 이유 설문 결과	5
[그림 1-5] 연구의 흐름도	8
[그림 4-1] 상관계수 값에 따른 두 변수의 분포	33
[그림 5-1] 공시가격과 실거래가격 영향도 분석 그래프	54

■ 표 목 차

[표 2-1] 다중회귀분석의 변수 투입 방법(추정 방법)	12
[표 2-2] 다중회귀분석의 선행연구	16
[표 2-3] 선행연구의 분석방법과 중요 작용 변수	21
[표 3-1] 선정된 변수 및 출처	26
[표 3-2] 공시가격과 실거래가격의 출처	27
[표 3-3] 수정된 독립변수 및 종속변수	29
[표 3-4] 최종 변수의 기술적 통계 자료	30
[표 4-1] 상관계수와 상관관계	33
[표 4-2] 공시가격과 변수의 pearson 상관관계 분석	34
[표 4-3] 공시가격 모형 요약	36
[표 4-4] 공시가격 ANOVA 분산분석	37
[표 4-5] 공시가격 회귀계수, p값, 공선성 통계량 검토	37
[표 4-6] 실거래가격과 변수의 pearson 상관관계 분석	39
[표 4-7] 실거래가격 모형 요약	41
[표 4-8] 실거래가격 ANOVA 분산분석	41
[표 4-9] 실거래가격 회귀계수, p값, 공선성 통계량 검토	42
[표 4-10] 예상가격과 실제가격의 오차율	44
[표 4-11] 수정된 공시가격의 모형 요약	46
[표 4-12] 수정된 공시가격의 ANOVA 분산분석	46
[표 4-13] 수정된 공시가격의 회귀계수, p값, 공선성 통계량 검토	47
[표 4-14] 수정된 실거래가격의 모형 요약	49
[표 4-15] 수정된 실거래가격의 ANOVA 분산분석	49
[표 4-16] 수정된 실거래가격의 회귀계수, p값, 공선성 통계량 검토	50

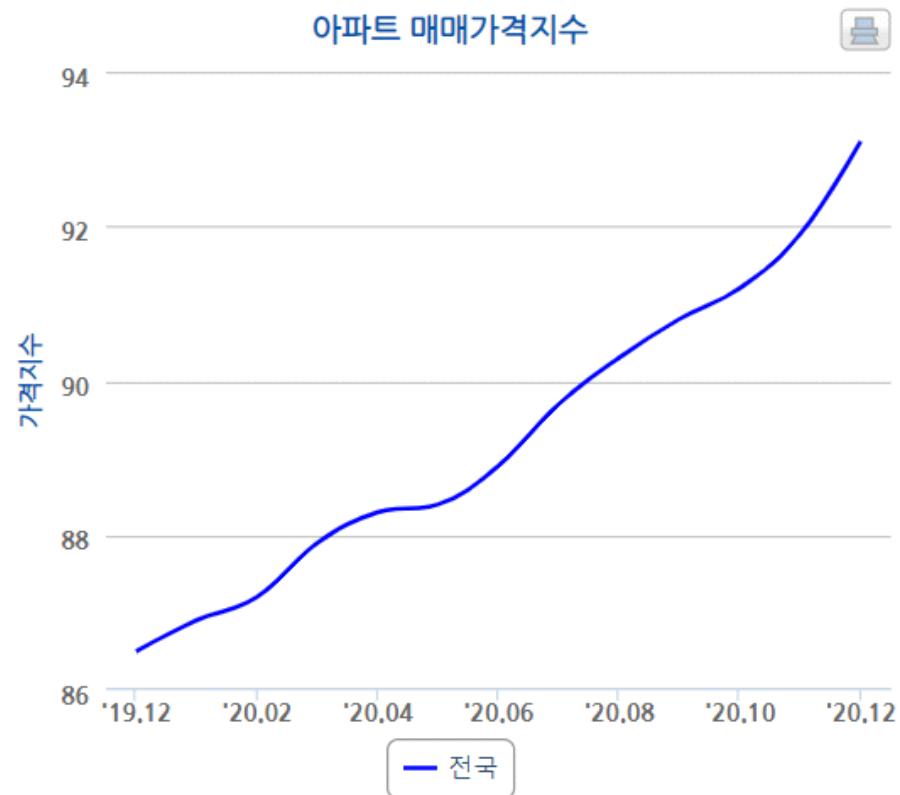
[표 4-17] 수정된 예상가격과 실제가격의 오차율	51
[표 5-1] 최종 회귀모델 표준화 계수 값 비교	53
[표 5-2] 세대 당 사설 학원 수 상위 10위	55

제 1 장 서 론

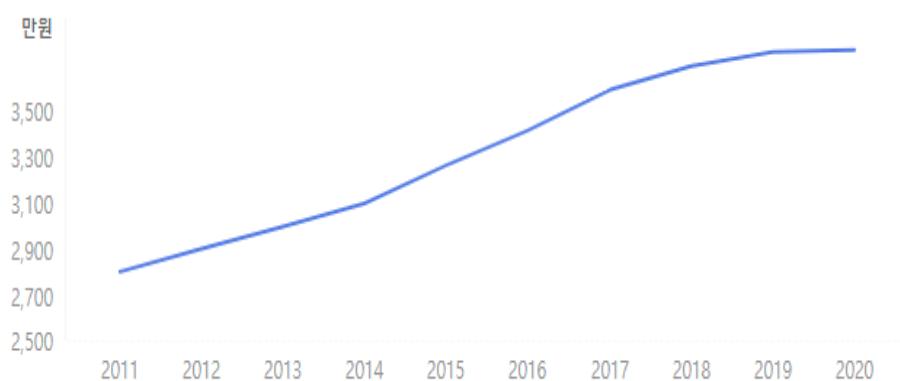
- 1.1. 연구의 배경 및 목적
- 1.2. 연구의 방법 및 범위

1.1. 연구의 배경 및 목적

2020년 코로나 시대를 맞이하면서 부동산 자산 가격의 상승으로 인해 대다수의 국민들이 아파트에 대한 관심이 뜨겁다. 이와 더불어 가계 부채가 급격하게 증가하면서 여러 문제가 제기 되고 있다. 주택 가격 안정화를 위해 여러 부동산 대책을 실시하였지만, 아파트 가격은 지속적으로 상승하였다. 한국부동산원(2021)의 데이터 [그림 1-1]을 보면 전국 아파트 매매가격지수는 2019년 12월 기준 86.5에서 2020년 12월 기준 93.1로 상승하였다. 이는 최근 5년 이내 최대로 상승했으며, 전례 없는 아파트 가격 상승세를 보여주었다. 국가통계 포털(2020)의 데이터 [그림 1-2] 의하면 국민총소득은 2019년 3,753.9 만원에서 2020년 3,762.1만원으로 약 8만원 상승하면서 최근 10년 동안 최저 성장을 보여주고 있다. 아파트 가격은 크게 상승한 반면 소득수준의 증가는 미미하게 되면서 구매에 대한 부담은 더욱 크게 증가되었다. 최근 치솟는 아파트 가격으로 인해 아파트 구매자들은 더 상승하기 전에 서둘러 구매자하자는 조급함이 생겼으며, 대출 금액은 더 증가하게 되었다. ‘영끌’, ‘빚투’라는 신조어를 만들면서 MZ 세대가 과도한 대출을 통해 아파트를 구매하는 모습을 기사에서 흔히 볼 수 있다. 금융감독원 자료에 따르면 MZ세대의 2021년 6월 말 기준 가계대출은 271조원에 달하며, 1년 전 226조원 보다 약 20% 폭증하였다. 반면 40대의 증가율은 8.8%, 50대는 5.3%였다. 대출을 받아서 구매한 아파트는 가계부채 상승과 직결되면서 높은 증가 추이를 보여주고 있다. 아파트 가격의 결정요인을 분석하고 합리적 구매를 통해 주거문제에 대한 해결이 필요한 시점이다.

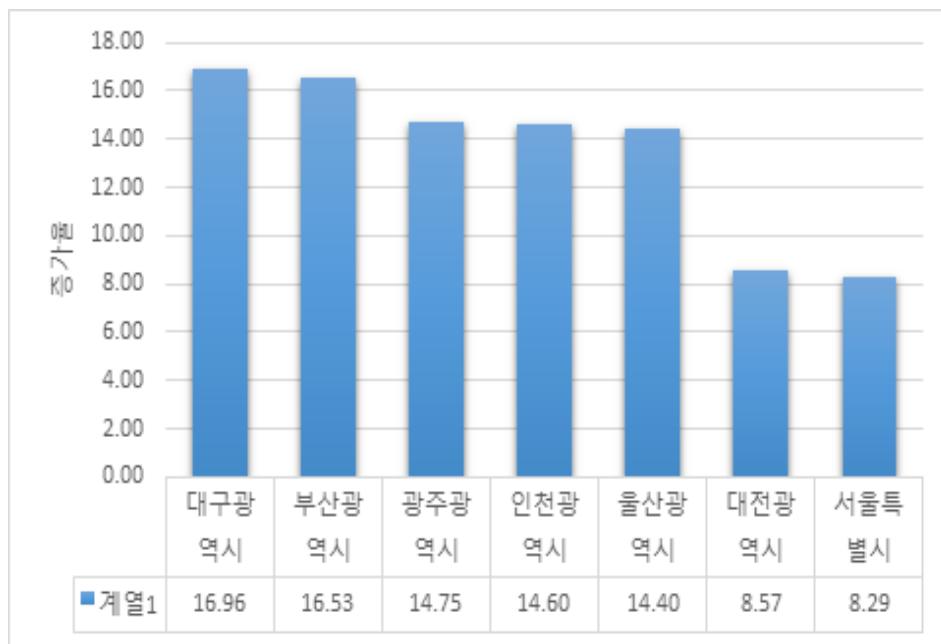


[그림 1-1] 2019.12–2020.12월 전국 아파트 매매가격 지수



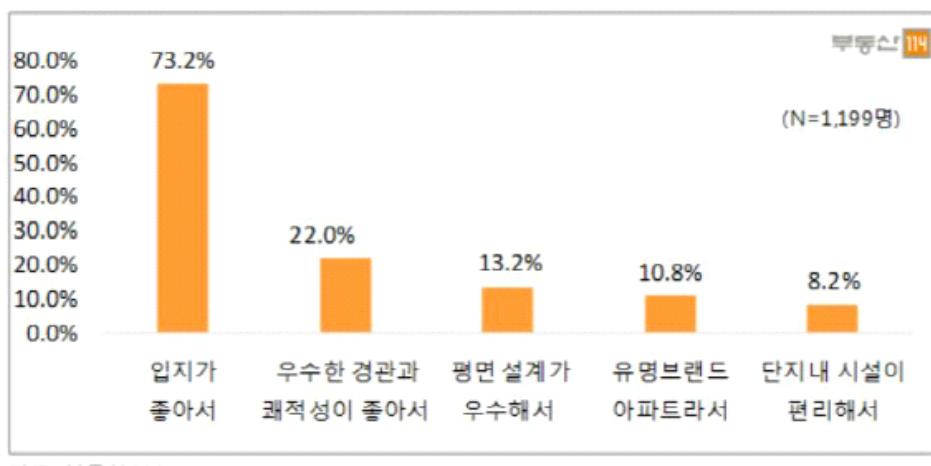
[그림 1-2] 2011—2020년 1인당 GNI(국민총소득)

국가통계포털(2020)에 의하면 전국 총 주택 수 18,525,844호 중 아파트는 11,661,851호이며, 이는 약 63%를 차지한다. 대구광역시는 총 주택 수 809,802호 중 아파트는 595,392호이다. 이는 전체의 약 73%를 차지하며, 전국 평균보다 훨씬 웃돌고 있다. 대구광역시는 광주광역시와 세종특별자치시에 이어 전국 아파트 비중이 3위이다. 아파트 증가율은 1위를 차지하고 있다. 국가통계포털 자료를 기반으로 만든 [그림 1-3]을 보면 2015-2020년 아파트 증가율이 7대 특·광역시에서 대구광역시가 1위다. 대구광역시는 2015년 509,068호에서 2020년 595,392호로 약 86,000호가 증가하였다. 약 17% 증가하면서 짧은 시간 가장 많이 증가했음을 보여주고 있다. 또한 대구광역시는 2020년 전용면적 84m² 기준 아파트값이 주택담보대출 금지선인 15억 원을 넘긴 사례가 서울을 제외하고 처음 나온 곳이며, 특히 대구광역시 수성구는 투기과열지구로 뮤여있을 정도로 부동산 열기가 뜨겁다.



[그림 1-3] 2015-2020년 7대 특·광역시 아파트 증가율

종합부동산포털 ‘부동산114’에서 전국 성인남녀 약 1,500명을 대상으로 실시한 설문조사에 따르면 현재 살고 있는 아파트를 선택한 이유 중 가장 높은 항목은 전체의 약 73%인 입지이며 [그림 1-4]와 같다. 입지란 인간이 경제 활동을 영위하기 위해 선택하는 장소이며, 부동산업계에서 말하는 동네, 위치가 등과 같은 의미라고 볼 수 있다. 그러므로 수요자들의 아파트 선택 기준이라고 볼 수 있는 입지(동네, 위치)를 중심으로 분석을 할 필요가 있다.



[그림 1-4] 현재 거주중인 아파트를 선택한 이유 설문 결과

도수관 외 2인의 연구¹⁾에서는 대구 지상철 3호선 건설에 따른 아파트 가격 변동에 대한 연구를 하였으며, 이를 통해 착공시작시점이 공사완료시점보다 역세권 아파트 가격에 더 큰 영향을 준다는 결론을 도출하였다. 장동현 외 1인의 연구²⁾에서는 공간계량모형을 활용하여 대구시 수성구를 중심으로 아파트가격 영향요인을 분석하였다. 이처럼 기존 연구는 일부 변수 효과에 관심을 두거나, 특정 구나 특정

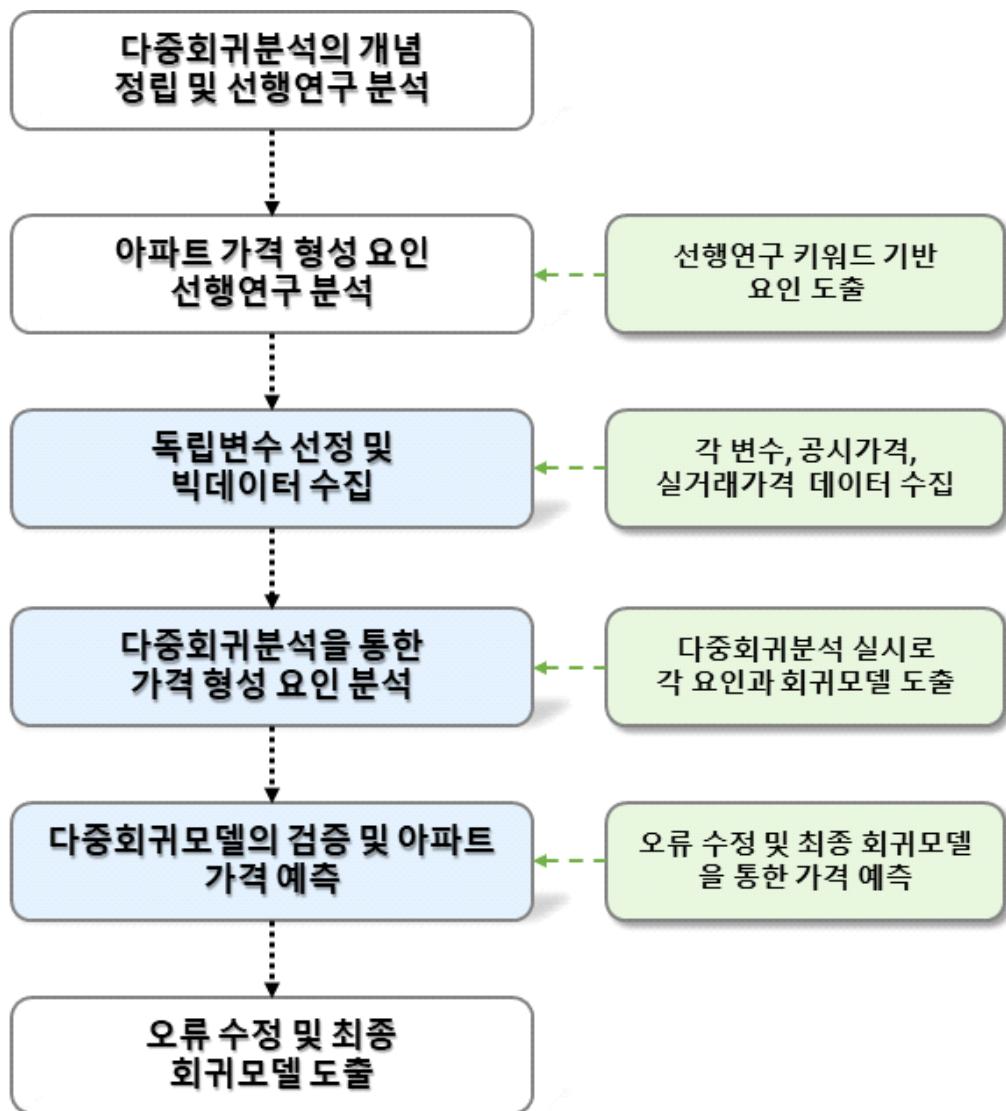
1) 도수관, 이규태, 김은지. “도시철도 역세권과 아파트 가격간의 관계 분석.” 한국정부학회 학술발표논문집 . (2015): 231-249.
 2) 장동현, 김한수. “공간계량모형을 활용한 아파트가격 영향요인 분석 연구 – 대구시 수성 구를 중심으로 –.” 한국주거학회논문집 31(1) (2020): 79-86.

동을 한정하여 분석하고 있다. 하지만 보다 신뢰도 높은 분석을 위해
서는 전체 지역의 특성을 파악하여 입지라는 범위를 광범위하게 조
사해야 하고, 넓은 범위에서 매물간의 관계를 비교 분석하는 것이 필
요하다. 따라서 본 연구에서는 범률로 지정된 행정구역 단위인 법정
동 중심으로, 대구광역시 아파트 가격에 영향을 미치는 요인을 다중
회귀분석을 통해 찾으려 한다.

1.2 연구의 방법 및 범위

본 연구의 흐름은 [그림 1-5]와 같다. 본 연구에서 활용하고자 하는 다중회귀분석의 개념에 대해 우선적으로 살펴보고, 이를 활용한 선행연구들은 분석한다. 동시에 아파트 가격에 영향을 미치는 요인에 대해 선행연구 분석을 실시한다. 이를 통해 본 연구에 사용하기 위한 요인을 추출 및 선정한다. 이어 아파트 가격 형성 요인에 필요한 빅데이터를 수집하고 이를 법정동 단위로 정리를 하여 법정동 별로 분석을 할 수 있도록 데이터를 구축한다. 구축한 데이터로 아파트 가격 다중회귀분석을 실시하고 검증을 한다. 최종적으로 도출된 다중회귀모델을 통해 아파트 가격에 대한 예측을 하며, 연구의 시사점과 활용방안을 결론에서 도출한다.

부동산에 대한 관심이 높은 현 시점에서, 총 주택 중 아파트에 대한 비중이 서울특별시 보다 높고, 아파트 증가율 또한 1위를 하고 있으며, 주택담보대출 금지선 15억원을 초과할 만큼 아파트에 대한 열기가 뜨거운 대구광역시를 본 연구에서의 공간적 범위로 선정하였다. 수성구에만 고가의 아파트들이 집중적으로 몰리면서 대구광역시 내에서 지역적으로 불균형도 있다고 판단되며 본 연구의 범위로 충분한 가치가 있다. 대구광역시 전체 법정동에 속한 모든 아파트 빅데이터를 사용하는 것으로 하며, 이를 위해 공공데이터를 활용한다. 데이터 중 전수 조사가 가능한 데이터만 사용하며, 공공데이터가 없는 경우는 직접 데이터를 수집한다. 시간적 기준은 2021년 1월이며, 기준 날짜 중 가장 최신의 공공데이터를 사용하는 것으로 한다.



[그림 1-5] 연구의 흐름도

제 2 장 다중회귀분석과 아파트 가격에 대한 선행연구

- 2.1 다중회귀분석의 개념 및 선행연구에 대한 분석
- 2.2 아파트 가격 형성 요인 선행연구 분석

2.1 다중회귀분석의 개념 및 선행연구에 대한 분석

2.1.1 다중회귀분석에 대한 개념

회귀분석은 매개변수 모델을 이용하여 통계적으로 변수들 사이의 관계를 추정하는 분석방법이다. 주로 독립변수가 종속변수에 미치는 영향을 확인하고자 할 때 사용하는 분석방법이다. 즉 한 변수의 변화가 원인이 되어 다른 변수에 어느 정도 영향을 미치는지를 측정하는 통계적인 분석 방법이다. 두 변수 간의 인과관계를 분석 할 때 주로 사용하며, 독립변수가 1개일 때는 단순회귀분석, 독립변수가 2개 이상 일 때는 다중회귀분석이라고 한다. 종속변수를 Y 로, n 개의 독립 변수들을 X_1, X_2, \dots, X_n 로 표기할 때 다중회귀모형식은 다음과 같다.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_n X_n$$

Y : 종속변수

X : 독립변수

β : 회귀계수

$\beta_0, \beta_1, \dots, \beta_n$ 은 회귀계수라 하고 β_1 은 독립변수 X_1 의 회귀계수라고 한다. 이 회귀계수는 회귀분석을 통해 얻어진다. 본 연구에서는 종속변수로는 아파트 가격을 투입하고, 독립변수로는 다양한 요인들을 투입하여 아파트 가격에 영향을 미치는 요인을 추출하고, 회귀계수를 도출해서 각 요인이 얼마나 가격에 영향을 미치는지 조사하고자 한다.

회귀분석은 많은 독립변수 중에 종속변수에 영향을 주는 변수를

선택하여 회귀식에 포함시킨다. 회귀식에서 변수가 유의적인 기여를 하는지의 여부에 따라 선택하는 변수 투입 방법이 달라진다. 변수 투입 방법은 입력, 전진, 후진, 제거, 단계선택으로 [표 2-1]처럼 5가지이다. ‘입력’ 방법은 독립변수 모두를 한 번에 투입하여 회귀모형을 추정하는 방법이다. 모든 변수를 투입하게 되므로 유의미한 변수와 유의하지 않은 변수에 대한 모든 정보가 산출된다. ‘전진’ 방법은 중요도가 가장 높은 독립변수부터 하나씩 추가해 나가는 방법이며, 더 이상 중요한 변수가 없을 때 중단된다. 반대로 ‘후진’은 중요도가 가장 낮은 독립변수부터 하나씩 제거되면서 분석이 진행되는 방법이다. ‘제거’ 방법은 연구자가 선택한 독립변수가 하나씩 제거되어 분석이 되는 방법이다. 마지막 ‘단계선택’ 방법은 회귀분석에 투입되는 독립변수들 가운데 설명력이 가장 높은 변수들로 회귀모델을 구성하는 방법이다. 첫 번째 단계에서는 종속변수 간 상관관계가 가장 높은 변수를 투입하고, 두 번째 단계에서는 종속변수 간 편상관관계가 있는 변수들을 투입한다. 세 번째 단계는 각 단계별로 설명력이 높은 변수에 대한 유의성 검증을 실시하며, 유의하지 않는 변수는 제거한다. 이는 종속변수에 독립변수가 영향을 미치지만 제3의 독립변수가 다른 독립변수에 많은 영향을 미친다면 이 변수를 제거하면서 진행한다는 뜻이다. 상관관계가 높다면 각각의 변수가 독립성을 띠어야 하는 독립변수의 성질과 어긋난다고 판단하여 본 연구에서는 ‘단계선택’ 방법을 채택하여 상관관계가 높은 독립변수는 제거하면서 유의미한 결과를 도출하고자 한다.

변수 투입 방법 (추정방법)	내용
입력	<ul style="list-style-type: none"> 독립변수 모두를 한 번에 투입하여 회귀모형 추정 모든 변수 투입으로 유의미한 변수와 유의하지 않은 변수에 대한 모든 정보 산출
전진	<ul style="list-style-type: none"> 중요도가 높은 독립 변수부터 하나씩 추가하면서 분석 진행 중요 변수가 없다고 판단될 때 분석 중단
후진	<ul style="list-style-type: none"> 중요도가 낮은 독립 변수부터 하나씩 제거하면서 분석 진행 중요도 낮은 변수가 없다고 판단 될 때 분석 중단
제거	<ul style="list-style-type: none"> 분석자가 선택한 독립변수가 하나씩 제거 되면서 분석 진행
단계선택	<ul style="list-style-type: none"> 설명력이 높은 변수들로 회귀모델을 구성하는 방법 종속변수 간 상관관계가 높은 변수부터 투입, 두 번째는 편상관관계가 있는 변수들을 투입, 마지막으로 설명력 높은 변수에 대한 유의성 검증 실시 및 유의하지 않은 변수 제거

[표 2-1] 다중회귀분석의 변수 투입 방법(추정 방법)

2.1.2 다중회귀분석에 대한 선행연구 분석

회귀분석을 활용하여 특정 요인들을 분석한 기존 연구를 살펴보면, 이병훈 외 3명의 연구³⁾에서는 회귀분석을 통해 부동산투자회사(REITs) 오피스빌딩의 수익률 영향요인을 분석하였다. 부동산투자회사 13개를 선정하고 10가지 요인을 바탕으로 배당액과 배당률 2가지 요소로 구분하여 분석을 실시하였다. 그 결과로 배당액에 영향을 미치는 요인으로 기업규모요인, 임대수익요인으로 도출이 되었으며, 기업규모요인으로는 자본금, 자본총계, 영업현금흐름, 투자현금흐름, 재무현금흐름이 있으며, 임대수익요인으로는 임대면적, 전용률, m^2 당 관리비, m^2 당 임대료가 있는 것으로 나타났다. 기업규모요인의 경우에는 기업의 규모가 클수록 보다 안정적인 투자대상을 발굴할 수 있기 때문에 지속적이고 안정적인 수익을 기대할 수 있다고 하였고, 이에 기업의 규모가 클수록 배당 가능액이 증가하게 되어 배당액 또한 증가하는 것으로 판단하였다. 또한 배당률에 영향을 미치는 요인으로는 지역적 특성 요인으로 꼽았으며, 지역적 특성요인으로 전년대비공시지가 상승률, 최고층, 최하층, 내용연수 등을 꼽았다. 전년대비공시지가 상승률의 경우는 해당 투자부동산의 공시지가가 상승하거나 최고, 최하층의 경우 대형오피스 밀집지역의 이점 등이 배당률에 영향을 미쳤다고 분석하였다.

김진원 외 3명의 연구⁴⁾에서는 회귀분석을 통해 교육시설의 공간계획에 따른 공사비 예측 모델을 도출했다. 종속변수를 총 공사비로 놓고 독립변수는 사업규모 그룹과 공간계획그룹으로 구분하였다. 사업

3) 이병훈, 이백래, 김주형, 김재준. "회귀분석을 통한 부동산투자회사(REITs) 오피스빌딩의 수익률 영향요인 분석." 대한건축학회 논문집 – 계획계 27.9 (2011): 293–300.

4) 김진원, 이백래, 김주형, 김재준. "회귀분석을 이용한 교육시설의 공간계획에 따른 공사비 예측 모델에 관한 연구." 대한건축학회 논문집 – 계획계 27.10 (2011): 103–110.

규모 그룹에는 학급 수, 학생 수용인원, 대지면적, 연면적으로 정의하였고, 공간계획 그룹에는 보통교실, 특별교실, 특수활동교실, 관리실, 보건위생실, 체육관, 기타시설, 공유면적으로 두고 각 변수의 면적 데이터를 사용하였다. 그 결과로 보통교실 면적, 특별교실 면적, 특수활동교실 면적, 체육관 면적, 기타시설 면적, 공유 면적을 요인으로 도출하였고, 이 요인을 통한 회귀식을 통해 기본설계단계에서 공사비를 예측 할 수 있다고 하였고 그 평균 오차율은 11.61%로 나왔다. 이 예측을 통해서 발주자나 시공사의 입장에서 적정한 공사비를 미리 예측 할 수 있다고 하였고, 경기도 교육청에서 발주한 BTL 사업을 대상으로만 분석을 하였기에 범위가 한정적이라는 점을 한계점으로 꼽았다.

강남규 외 3인의 연구⁵⁾에서는 국토교통부와 한국부동산원(舊 한국감정원)의 ‘2018년 주거용 건물에너지 소비량 제공 자료’를 통해 우리나라 주거용 건물의, 용도별, 지역별 에너지 소비량과 원단위(Total energy unit intensity) 특성을 분석하였다. 독립변수를 시·도 지역, 연면적, 지역구분(중부1, 중부2, 남부, 제주), 위도와 경도로 선정하였다. 종속변수는 단독주택 239개 지역, 아파트 215개 지역으로 선정하였다. 이 결과로 우리나라에서 에너지를 가장 많이 사용하는 주거 유형으로 약 59%를 차지하는 아파트를 꼽았고, 에너지원은 도시가스가 전체 약 54%로 가장 높다고 하였다. 또 시지역과 도지역을 비교 했을 때 시지역이 도지역보다 원단위가 높은 경향을 보이고 있다고 하였다.

양성필 외 3인의 연구⁶⁾에서는 한반도에 영향을 미친 태풍을 타입별로 분류를 하였고, 이를 다중회귀분석을 통해 태풍 타입별 건물피

5) 강남규, 김순호, 강은혜, 최정민. "주거용 건물의 에너지 소비량 특성 및 선형 회귀모형 개발에 관한 연구 - 2018년 데이터를 중심으로 -." 대한건축학회논문집 36.7 (2020): 115-122.

6) 양성필, 손기영, 이경훈, 김지명. "태풍타입별 피해 분석 및 다중회귀분석을 활용한 태풍 피해예측모델 개발 연구." 한국건축시공학회지, 16(5), (2016): 437-445.

해율 예측모델을 개발하였다. 종속변수는 건물피해율이며, 독립변수는 태풍 이동속도, 최대풍속, 강풍반경, 일최다강수량, 하천개소수, 완전개수율, 공원녹지비율, 댐, 절토사면, 고층건물, 태풍타입 코드로 선정하였다. 회귀분석의 결과로 건물 피해율에는 영향을 미치는 요소로 최대풍속, 강풍반경, 일최다강수량, 절토사면과 고층건물 총 5가지를 꼽았다. 회귀모델의 설명력을 나타내는 $adj\ R^2$ 값은 0.552로 낮게 나왔지만 자연재해에 대한 기존 연구 강원지역 대설 관련 피해 추정 모델은 $adj\ R^2$ 값이 0.112로 도출, 베트남의 풍수해 관련에 대한 국제적 원조를 위한 지리적 기초자료를 제공하기 위해 회귀분석에 대한 $adj\ R^2$ 값이 0.517로 도출을 언급하였다. 이처럼 자연재해에 대한 예측은 매우 어렵다고 언급하면서 기존의 연구보다 높은 설명력을 도출하였다. 사회현상에 대한 분석은 설명력이 35%이하로 낮게 나오는 경향이 있기 때문에 양성필의 연구에서는 유의미한 수치로 볼 수 있다.

회귀분석에 대한 선행 연구 자료를 정리하면 [표 2-2]와 같다. 이처럼 회귀분석은 다양한 요인들 사이에서 종속변수에 영향을 미치는 요인을 논리적으로 찾을 수 있다. 또한 특정 분야에 얹혀있지 않고 사회 다양한 분야에서 사용할 수 있으며 신뢰도 또한 높게 인정받고 있다. 이에 본 연구에서는 다중회귀분석을 이용하여 공시가격과 실거래가격을 종속변수로 두고, 각 독립변수들을 추출하여 아파트 가격에 영향을 미치는 요인을 찾고자 한다.

[표 2-2] 다중회귀분석의 선행연구

저자 (년도)	구분	내용
이병훈 외 3인 (2011)	제목	회귀분석을 통한 부동산투자회사(REITs) 오피스빌딩의 수익률 영향요인 분석
	키워드	부동산투자회사, 오피스빌딩, 다중회귀분석, 배당액, REITs
김진원 외 3인 (2011)	제목	회귀분석을 이용한 교육시설의 공간계획에 따른 공사비 예측 모델에 관한 연구
	키워드	교육시설, 공간계획, 공사비 예측, 회귀분석
강남규 외 3인 (2020)	제목	주거용 건물의 에너지 소비량 특성 및 선형 회귀모형 개발에 관한 연구 - 2018년 데이터를 중심으로
	키워드	주거용 건물, 에너지, 소비량, 원단위, 선형회귀모형
양성필 외 3인 (2016)	제목	태풍타입별 피해 분석 및 다중회귀분석을 활용한 태풍피해예측모델 개발 연구
	키워드	태풍, 다중회귀분석, 건물 피해율, 리스크 평가

2.2 아파트 가격 형성 요인 선행연구 분석

아파트 가격 결정요인에 관한 연구는 다양한 분석방법을 이용하여 이루어지고 있다. 헤도닉가격모형과 다중회귀분석, 분위수 회귀분석, 비교분석, 상관분석, 요인분석 등의 분석방법을 통해서 아파트 가격의 결정요인을 분석하였다. 가격 결정요인에 대한 법적인 근거나 가이드 라인의 부재로 선행연구에서 요인을 찾는 것이 신뢰도가 높다고 판단하였다. 본 연구에서는 2001년부터 2021년도까지의 연구를 대상으로 ‘아파트 가격’, ‘주택 가격’ 키워드를 가진 43편을 찾았으며, 이중 가격 요인을 도출하지 않고 특정 요인만 상대로 분석한 39편의 연구를 제외하고 4편을 대상으로 선행연구 분석을 실시하였다.

이강 외 1인의 연구⁷⁾에서는 헤도닉 가격모형을 활용한 주택가격 결정요인 분석을 위해 서울 용산구 이촌동 지역을 대상으로 26개 아파트단지를 조사하였다. 아파트 가격 결정의 요인을 크게 ‘주택특성’, ‘이웃특성’, ‘접근성’으로 분류했다. 첫 번째 ‘주택특성’에는 공급면적, 전용면적, 방수, 욕실 수, 주차대수, 경과 연수로 변수를 구성하였다. 두 번째 ‘이웃특성’은 세대수, 최고층, 평균관리비, 건설사 지명도, 한강까지 거리, 국립중앙박물관과의 거리로 변수를 구성하였다, 세 번째 ‘접근성’은 초등학교와의 거리, 중학교와의 거리, 고등학교와의 거리, 지하철역과의 거리, 강변북로와의 거리로 변수를 구성하였다. 그 결과 주택특성에서 주차대수와 욕실수가 양의

7) 이강, 최근희. "헤도닉 가격모형을 활용한 주택가격 결정요인에 관한 연구." 한국도시행정학회 학술발표대회 논문집 . (2016): 317-333.

영향을 갖는 변수였고, 음의 영향을 갖는 변수는 없었다. 이웃특성에
서는 세대수, 건설사 지명도가 양의 영향을 가졌고, 국립박물관과의
거리가 음의 영향을 나타냈다. 접근성에서는 고등학교, 지하철역과의
거리가 양의 영향을 나타냈고, 초등학교와의 거리는 음의 영향을 나
타냈다.

박운선 외 2인의 연구⁸⁾에서는 지역유행이 주택가격에 미친 영향을
분석하기 위해서 헤도닉 가격모형을 활용했다. 서울시와 부산시 중
가격이 가장 높게 형성되는 행정구와 가장 낮게 형성되는 행정구를
공간적 범위로 선정하여 조사하였다. 서울시 강남구와 부산시 해운대
구를 가격이 높게 형성되는 행정구로 선정했다. 반대로 낮은 가격 행
정구는 서울시 금천구와 강북구를 선정했고 부산시는 중구와 사상구
를 선택했다. 아파트 가격 결정의 요인으로는 ‘세대특성’, ‘단지
특성’, ‘입지특성’과 ‘교육특성’, ‘환경특성’으로 구분하였다.
첫 번째 ‘세대특성’은 방수로 구성하였다. 두 번째 ‘단지특성’은
총세대수, 세대 당 주차수로 구성하였다. 세 번째 ‘입지특성’은 도
심까지 거리, 지하철역까지 거리, 버스정류장까지 거리, IC까지 거리,
종합병원까지 거리를 선정했다. 네 번째 ‘교육특성’에는 초등학교
까지 거리, 고등학교 수를 선정했다. 마지막 ‘환경특성’으로는 녹
지 및 공원까지의 거리를 선정했다. 헤도닉 가격모형을 활용한 결과
로 입지특성의 도심까지 거리는 서울시 고가지역의 실거래가 모형을
제외하고는 모두 예상값과 상반된 결과값이 나왔으며 이는 도심과
거리가 멀어질수록 가격이 상승함을 알 수 있다. 이것을 통해 수요자
들이 도심 외곽의 주거환경이 우수한 주택을 선호하고 있다는 결과
를 도출했다. 세대특성과 단지특성인 방수, 총세대수, 세대 당 주차
수와 교육특성인 고등학교 수는 지역에 상관없이 아파트 가격에 궁

8) 박운선, 임병준. “헤도닉 가격모형을 활용한 아파트 가격결정요인 분석.” 대한부동산학회
논문집 . 28.2 (2010): 245-271.

정적 영향을 주는 중요한 특성요인으로 도출되었다.

김희호 외 1인의 연구⁹⁾에서는 서울 강남구와 노원구를 대상으로 주택가격 결정요인을 찾고자 했다. 주택가격의 조건부 분포에 따라 각 분위수 별로 해도닉가격모형을 추정하기 위하여 분위수 회귀분석을 하였다. 가격 결정 요인으로는 건축 경과 연수, 전용면적, 총 층 수, 거주층수, 남향여부 더미, 지하철 더미, 고등학교 더미, 산과 강의 풍경이 보이는 조망권 더미 등을 사용하였으며, 경과 연수와 전용면적의 제곱 변수를 추가적으로 사용하였다. 분위수 회귀분석 결과 고가 아파트는 3~5년의 기간이 경과하면 가격이 상승추세로 반전되는 것을 발견할 수 있었다. 면적에서는 면적의 계수는 양의 값, 면적의 제곱의 계수는 음의 값을 보이면서 면적이 클수록 아파트 가격이 상승하다가, 일정면적을 초과하면 가격이 하락하는 결과를 도출했다. 또 지하철 인접성은 가격결정에 통계적으로 유의적인 영향을 미치지 못하거나 음의 영향을 미치는 것으로 나타났다. 이를 통해 서울 지하철망이 잘 갖추어진 것을 그 원인으로 유추했다.

송호창 외 2인의 연구¹⁰⁾에서는 주상복합아파트의 주택규모별 가격 결정요인을 분석하기 위해 서울시 주상복합아파트 228개를 대상으로 조사를 진행하였다. CART 분석, 독립변수 상관분석, 요인분석을 실시하였고. 이를 통해 모형개발의 유형을 확정하고, 다중회귀모형을 도출했다. 가격결정요인은 ‘주택특성’, ‘입지특성’, ‘복합기능 특성’ 3개의 부분으로 나누었다. ‘주택특성’은 경과 연수, 대지면적, 연면적, 용적률, 세대수, 평균층수, 전용면적, 전용률, 주동 수, 세대 당 주차대수, 건설사지명도를 선정했다. 두 번째 ‘입지특성’은 접도수, 접근도로 폭, 지하철까지 거리, 초교 수, 중고교 수, 공원 .

9) 김희호, 박세운. "서울 주택가격의 결정요인: 분위수 회귀분석." 한국주택학회 주택연구 논문집 21.2 (2013): 141-168.

10) 송호창, 김태호, 이주형. "주상복합아파트의 주택규모별 가격결정요인 분석." 서울도시 연구 9.3 (2008): 79-92.

하천 수, 공원 · 하천 거리, 개발유형으로 선정했다. 세 번째 ‘복합 기능 특성’은 주거대비 상업기능비율, 주거대비 오피스기능 비율, 단지 내 녹지 유/무로 선정했다. 다중회귀분석을 통한 결과로 주상복합시장에서 단지규모, 개발밀도, 단지 내 공원이나 균린공원 · 하천, 건설사 지명도가 가격에 영향을 미치는 주요한 요인으로 분석되었다. 주상복합아파트의 경우 순수 주거기능 외 상업과 오피스기능의 비율이 높으면 가격에 마이너스 요인으로 작용한다는 결과를 도출했다. 이를 통해 거주자의 삶의 질과 쾌적한 주거환경을 고려한 개발 및 공급이 필요하다고 제시하였다. 선행연구의 분석방법과 중요 작용 변수를 정리하면 [표 2-3]과 같다.

현재까지 아파트 가격에 대한 선행연구들은 특정 동이나 특정 구에 한해서 제한적 범위의 연구를 하거나, 특정 아파트만을 상대로 하는 연구가 주를 이뤘다. 특정 시 전체를 연구한 시도는 미비한 실정이다. 기존 연구의 대다수는 수도권에 편중되어 있으며 지방권에 대한 연구가 많이 부족하다. 전체 자료를 수집하는 것에 대한 시간적 한계도 있고, 데이터를 취득하는 것이 어려운 경우 직접 일일이 취득해야하는 경우도 존재하기 때문이다. 특정 시 전체를 살펴보고 이를 통해 시의 특성을 분석해 볼 필요가 있다.

[표 2-3] 선행연구의 분석방법과 중요 작용 변수

	이강 외 1인	박운선 외 1명	김희호 외 1명	송호창 외 2명
논문 제목	해도닉 가격 모형을 활용한 주택가격 결정요인에 관한 연구: 서울 이촌동 지역을 중심으로	해도닉 가격 모형을 활용한 아파트 가격 결정요인 분석	서울 주택가격의 결정요인: 분위수 회귀분석	주상복합아파트의 주택 규모별 가격 결정요인 분석
분석 방법	해도닉 가격모형	해도닉 가격모형	분위수 회귀분석	CART 분석, 상관관계분석, 요인분석, 회귀분석
중요 작용 변수	주차대수, 욕실 수, 세대 수, 건설사 지명도, 국립중앙박물관까지 거리, 초등학교, 고등학교, 지하철까지 거리	도심까지 거리, 방수, 세대 수, 세대 당 주차 대수, 고등학교 수	전용면적, 전축경과연수 제곱, 전용면적 제곱, 지하철 더미	주동 수, 공원 및 하천의 수, 건설사 지명도, 주거 대비 상가 비율, 주거 대비 오피스 비율

제 3 장 가격 형성 요인 빅데이터 수집 및 변수 검증

- 3.1 가격 형성 요인 선정과 빅데이터 수집
- 3.2 데이터 분류 및 변수 검증

3.1 가격 형성 요인 선정과 빅데이터 수집

3.1.1 독립변수 선정 및 빅데이터 수집

선행연구를 통해서 아파트 가격에 영향을 미치는 요인 중 가장 많이 나온 키워드를 통해 독립변수를 추출하고자 하였다. 전체 키워드 중 2번 이상 나온 키워드로 추출하였고, 한강 또는 국립중앙박물관처럼 서울특별시에 국한 되는 키워드는 제외하였다. 그 결과로 전용면적, 주차대수, 세대수, 방수, 층수, 경과 연수, 건설사 지명도, 초·중·고 학교, 지하철, 공원으로 10개의 키워드로 요약이 되었다. 이를 아파트 ‘내부적 특성’과 ‘외부적 특성’, ‘기타 변수’로 구분하였다.

‘내부적 특성’은 세대수, 경과 연수, 주차대수, 건설사 지명도(상위브랜드, 공공임대아파트), 층수, 전용면적으로 변수를 구성하였다. 앞서 나온 키워드에서 방수는 제외하였는데, 일반적으로 방의 개수가 늘어나면 전용면적이 커지기 때문에 전용면적과 직접적인 상관관계가 존재하게 된다. 독립변수끼리의 상관관계를 줄이고 독립성을 유지하기 위해서는 제거하는 것이 타당하다고 판단하였다. 건설사 지명도에서 상위브랜드와 공공임대아파트를 구분하였다. 상위브랜드는 한국기업평판연구소 브랜드평판지수 1위에서 20위까지의 아파트 브랜드로 구성하였고, 공공임대아파트는 LH, 도시공사가 사업 주체인 아파트로 구성하였다. 선행연구에서는 상위브랜드에 초점을 맞춰서 독립변수를 구성했지만, 본 연구에서는 공공임대아파트도 추가하였다. 일반적으로 공공임대주택이 주변 주택 가격을 하락시킬 것이라는 예상

과 달리 이재영 외 1인의 연구¹¹⁾에 따르면 공공임대주택이 매매시장에서 주택가격의 상승을 시킨다는 결과를 도출하였다. 전세시장에서는 전세 가격 인하효과가 있다고 하였지만 매매시장에서는 반대로 작용했는데, 그 이유로 전세임대나 매입임대의 경우 기존 주택을 매입 혹은 전세로 활용하여 매매가격을 자극할 가능성을 꼽았다. 이에 본 연구에서도 공공임대아파트가 대구광역시 아파트 가격에 미치는 영향을 분석해 볼 필요가 있다고 판단하였다.

‘외부적 특성’은 공원면적, 초·중·고 학교 수 및 학생 수, 역세권, 사설 학원수로 변수를 구성하였다. 여기서 사설 학원수를 추가하였는데, 학교 뿐 아니라 사설 학원 또한 교육환경의 한 요소이며 아파트 가격에 영향을 미치기 때문이다. 정수연의 연구¹²⁾에 따르면 서울지역 전체모형을 분석한 결과 사설학원이 아파트 가격에 유의한 영향을 미친다는 결과를 도출하였다. 이에 교육환경 측면에서 학교뿐만 아니라 사설학원도 추가하여 분석하고자 한다.

‘기타 변수’는 절대적인 수치 외에 상대적 수치를 계산하기 위한 변수들이며, 법정동의 면적과 인구수로 구성하였다.

‘내부적 특성’은 세움터 기반의 건축데이터개발 서비스의 건축물대장을 활용하여 19,973동의 대구광역시 아파트 정보를 수집하였다. ‘외부적 특성’의 공원은 대구광역시청 홈페이지에서 [2020 공원유원지 농지 현황]을 통해 도시공원, 유원지, 자연공원 3가지 구분으로 806개소의 공원 위치와 면적을 취득하였다. 학교 수와 사설학원 수는 대구광역시 교육청 홈페이지에서 [학교현황], [학원 및 교습소 현황]을 통해 457개의 학교와 주소가 없는 학원을 제외한 5,162개의 학원 정보를 취득하였다. 초·중·고 학생 수는 국토지리정보원을 통해

11) 이재영, 박태원. "공공임대주택이 주택 매매 및 전세 가격에 미치는 영향 – 서울 지역을 중심으로 –." *부동산분석* 2.1 (2016): 53–65.

12) 정수연. "교육요인이 서울아파트가격에 미치는 영향에 관한 연구." *국토계획* 41.2 (2006): 153–166.

법정동 단위로 집계된 학생 수로 수집하였다. 역세권 면적은 데이터의 부재로 네이버 지도 서비스를 이용하여 직접 수집하였고, 지하철 1,2호선과 지상철 3호선 역을 중심으로 반경 500m안에서 도보로 이용할 수 있는 권역으로 설정하고 면적을 측정하였다. ‘기타 변수’에서 법정동 면적과 법정동 인구수는 국토지리정보원을 통하여 수집하였다. 선정된 변수와 그 출처를 정리하면 [표 3-1]과 같다.

[표 3-1] 선정된 변수 및 출처

카테고리	변수	출처
내부적 특성	세대 수	
	경과 연수	
	주차 대수	
	건설사 지명도(상위브랜드 아파트, 공공임대아파트)	세움터 건축 데이터 개방 (건축물 대장)
	최고 층수	
	전용 면적(m^2)	
외부적 특성	공원 면적(m^2)	대구광역시 홈페이지 (2020 공원유원지 녹지현황)
	초중고 학교 수	대구광역시 교육청 홈페이지 (학교 현황, 학원 및 교습소 현황)
	사설 학원 수	
	초중고 학생 수	국토지리정보원
기타 변수	역세권 면적(m^2)	네이버 지도 서비스
	법정동 면적(m^2)	
	법정동 인구	국토지리정보원

3.1.2 대구광역시 공시가격 및 실거래가격 빅데이터 수집

아파트 가격인 종속변수는 실거래가격과 공시가격 2가지로 구성하였다. 공시가격은 정부가 조사 및 산정해서 공시하는 가격이다. 과세기준이 되는 가격이며 종합부동산세 등 보유세 과표로 사용되고 있다. 이에 공시가격은 세금과 직결되어 있어 실제 매매가격을 의미하는 실거래가격과는 성격이 다르기 때문에, 두 가격 간에 영향을 미치는 요인에도 차이가 있을 것으로 예상된다.

공시가격은 전축데이터개방 서비스의 전축물대장에 공개 되어 있는 약 57만호의 가격데이터를 수집하였고, 실거래가격은 국토교통부 실거래가 공개시스템을 이용하여 최근 3년(2018~2020) 간의 거래 내용 110,328개를 수집하였다. 실거래가격의 1년간의 데이터 즉 2020년의 거래를 중점적으로 분석해볼 필요가 있지만, 2020년 아파트 가격의 상승으로 인해 실거래가 이루어지지 않은 법정동이 많이 존재했다. 실거래가 없는 법정동을 제외하면 표본으로 들어가는 법정동이 100개 미만으로 줄어들게 되면서 표본수가 굉장히 많이 감소하게 된다. 표본수의 감소는 통계적으로 논리성이 떨어질 수도 있기에 100개 이상을 유지하기 위하여 3년으로 시간적 범위를 늘려서 수집을 하였다. 공시가격과 실거래가격의 출처는 [표 3-2]와 같다.

[표 3-2] 공시가격과 실거래가격의 출처

카테고리	변수	출처
종속변수	공시가격(원)	세움터 건축 데이터 개방 (건축물 대장)
	실거래가격(원)	국토교통부 실거래가 공개시스템

3.2 데이터 분류 및 변수 검증

법정동 단위로 분류하기에 앞서 구 단위로 남구, 달서구, 달성군, 동구, 북구, 서구, 수성구, 중구로 7구 1군으로 우선 분류하였다. 총 204개의 법정동 중 아파트가 존재하지 않는 법정동을 제외하고 128 개의 법정동으로 분류하여 데이터를 구축하였다. 데이터 특성상 읍·면·동 단위로 분류된 데이터가 많고 리 단위로의 데이터 분류는 대부분 제공하지 않기에, 달성군은 읍 또는 면 단위로 구축하였다.

회귀분석에 들어가기 전에 앞서 변수들을 단순히 절대적 수치로 분석하는 것에는 한계가 있다고 판단하였다. 법정동 면적이 가장 작은 곳은 약 0.02km^2 로 중구 수동과 장관동이고, 가장 넓은 곳은 달성군의 읍·면을 제외하고 약 10.9km^2 로 동구의 지묘동이다. 면적을 비교 했을 때 약 500배 이상 차이가 난다. 이처럼 면적이 작은 법정동은 아파트나 학교, 학원, 공원 등의 시설이 토지의 물리적 한계로 적게 들어설 수밖에 없다. 거주 인구 또한 물리적으로 적기 때문에, 각 변수들을 상대적인 값으로 변환해서 형평성을 갖출 필요가 있다. 이에 본 연구에서는 면적, 인구, 세대 단위로 나눈 상대적 변수로 변경하였고 최종적 변수는 [표 3-3]과 같이 독립변수 12개와 종속변수 2개로 구성되었다. [표 3-4]는 각 독립변수와 종속변수의 기술적 통계를 정리한 것이다. 종속변수의 실거래가격에서 동구 동내동은 2018~2020년 동안 아파트 실제 거래가 이루어지지 않았고 데이터의 부재로 표본에서 제외되었으며, 실거래가격의 표본수(N)는 127개가 되었다.

[표 3-3] 수정된 독립 변수 및 종속변수

카테고리	변수	수정된 변수
내부적 특성	세대 수	법정동 면적당 세대 수
	경과 연수	경과 연수
	주차 대수	세대 당 주차 대수
	건설사 지명도	상위 브랜드 비율
		공공임대브랜드 비율
	최고 층수	최고 층수
	전용 면적(m^2)	전용 면적
외부적 특성	공원 면적(m^2)	1인당 공원 면적
	초·중·고 학교 수	세대 당 초·중·고 학교 수
	사설 학원 수	세대 당 사설 학원 수
	초·중·고 학생 수	법정동 인구 내 초·중·고 학생 수 비율
	역세권 면적(m^2)	법정동 면적 내 역세권 면적 비율
종속 변수	공시가격(원)	공시가격
	실거래가격(원)	실거래가격

[표 3-4] 최종 변수의 기술적 통계 자료

변 수	평균값	표준 편차	표본수
법정동 면적당 세대 수	2263.90	2208.22	128
경과 연수	17.81	9.10	128
세대 당 주차 대수	0.93	0.36	128
상위 브랜드 비율	10.90	19.74	128
공공임대브랜드 비율	17.67	27.78	128
최고 층수	14.88	5.27	128
전용 면적(m^2)	77.63	17.56	128
1인당 공원 면적	39.29	136.01	128
세대 당 초·중·고 학교 수	0.0016	0.0052	128
세대 당 사설 학원 수	0.0134	0.5016	128
법정동 인구 내 초·중·고 학생 수 비율	9.54	4.38	128
법정동 면적 내 역세권 면적 비율	33.88	33.99	128
공시가격(원)	193,456,809	113,638,601	128
실거래가격(원)	296,697,165	152,714,704	127

제 4 장 아파트 가격 형성 요인 분석 및 가격 예측

- 4.1 공시가격의 요인 분석 및 통계학적 검증
- 4.2 실거래가격의 요인 분석 및 통계학적 검증
- 4.3 다중회귀모델의 검증 및 가격 예측

4.1 공시가격의 요인 분석 및 통계학적 검증

4.1.1 공시가격과 독립변수의 상관관계 분석

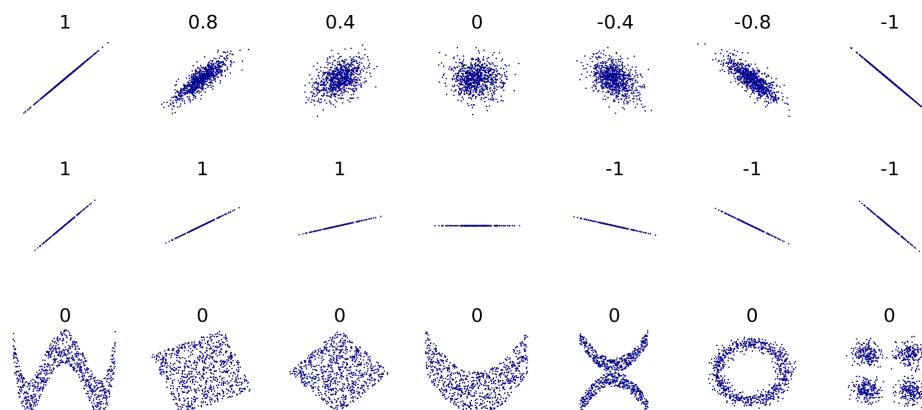
상관관계 분석은 원인과 결과가 되는 두 변수 간의 밀접한 정도나 어떤 선형적 또는 비선형적 관계를 갖는지 분석하는 기법을 말한다. 변수는 서로 독립적인 관계이거나 상관적 관계일 수 있으며 이때 두 변수 간에 관계의 강도를 상관관계라고 한다. 상관관계의 연관된 정도를 상관계수라고 하며 수치로 표시된다. 상관계수가 양(+)의 값을 가질 때는 정적상관, 음(-)의 값을 가질 때는 부적상관이라고 하며 상관계수가 0일 때는 대상 간에 관련성이 없음을 의미한다. 부호에 상관없이 숫자의 절댓값이 클수록 관련성이 크다. [표 4-1]은 상관계수에 따른 상관관계 정도를 나타낸다. [그림 4-1]은 상관계수 값에 따른 두 연속형 변수의 대략적인 분포를 나타낸다. 상관관계가 강할수록 직선적인 형태를 띠고, 상관관계가 낮을수록 비선형 분포에 가깝다. [표 4-2]는 공시가격 회귀분석에 들어가기에 앞서 Pearson 상관관계 분석¹³⁾을 통하여 도출된 상관계수이다. 이를 이용하여 변수 상호간의 관계를 알아볼 수 있다. 표 내용에서는 대부분의 변수가 공시가격과 양의 상관관계를 가지고 있으며, 세대 당 사설 학원 수, 1인당 공원 면적, 공공임대 브랜드 비율, 경과 연수가 음의 상관관계를 가지고 있다. 독립변수끼리의 상관관계도 중요한데 B~M이 독립변

13) 피어슨의 곱적률 상관(Pearson's product-moment correlation)은 두 변수의 선형 관계가 존재할 경우 그 관계가 얼마나 강한지 알 수 있는 값이며, 두 변수가 연속형 양적 변수일 경우에 사용가능한 방법, 상관계수는 공분산으로부터 유도되며, 이를 제곱하는 경우 단순선형회귀의 결정계수와 같다.

수이며 이들끼리의 상관관계가 높게 나온다면 독립변수의 독립성을 보장할 수 없다. 일반적으로 ± 0.7 이하로 나오면 상관관계가 있지만 높지는 않다고 판단을 한다. 독립변수끼리의 상관계수가 모두 ± 0.7 미만이며 최고 상관계수가 0.549로 독립변수 사이에 상관도가 높지 않다고 판단하였다.

[표 4-1] 상관계수와 상관관계

상관계수	상관관계
± 0.9 이상	상관관계가 아주 높다
$\pm 0.7\sim 0.9$	상관관계가 높다
$\pm 0.4\sim 0.7$	상관관계가 있다
$\pm 0.2\sim 0.4$	상관관계가 있으나 낮다
± 0.2 미만	상관관계가 거의 없다.



[그림 4-1] 상관계수 값에 따른 두 변수의 분포

[표 4-2] 공시가격과 변수의 pearson 상관관계 분석

	A	B	C	D	E	F	G	H	I	J	K	L	M
A	1												
B	-.001	1											
C	.007	.060	1										
D	-.116	-.054	-.045	1									
E	.181	.221	0.45	-.210	1								
F	.350	-.054	-.101	-.059	-.073	1							
G	-.200	-.071	-.065	-.001	-.096	.059	1						
H	.530	-.162	-.168	.143	-.056	.471	.023	1					
I	.454	-.011	-.203	.089	.030	.298	-.135	.549	1				
J	.247	.106	-.190	-.240	.363	.234	.057	.257	.110	1			
K	-.300	.245	.021	-.229	.323	-.316	-.030	-.538	-.382	.118	1		
L	.788	-.120	.086	-.121	.189	.217	-.308	.339	.468	.112	-.137	1	
M	.420	-.073	-.118	.031	-.300	.324	-.009	.501	.426	.280	-.308	.273	1

※ 변수 목록

A : 공시가격, B : 세대 당 사설 학원 수, C : 세대 당 초·중·고 학교 수,
D : 1인당 공원 면적, E : 법정동 면적 내 역세권 면적 비율,
F : 상위 브랜드 비율, G : 공공임대 브랜드 비율, H : 최고 층수,
I : 세대 당 주차 대수, J : 법정동 면적 당 세대 수, K : 경과 연수,
L : 전용 면적, M : 법정동 인구 내 초·중·고 학생 수 비율

4.1.2 공시가격의 다중회귀분석 및 요인 도출

독립변수가 총 12개로 회귀분석의 설명력이 높아질 수 있지만 입력 변수사이에 상관관계를 가지는 다중공선성¹⁴⁾의 문제를 일으킬 수 있다. 이를 위해 단계선택 방법을 통하여 상관관계가 높은 변수는 제거하면서 유의미한 결과를 얻고자 하였다. 선형회귀모형의 통계적 유의성을 검증하기 위해 ANOVA(Analysis of variance) 분산분석¹⁵⁾을 이용하였으며, F-검정에 의한 유의확률 p값으로 검정하였다. 단계선택 방법으로 산정된 공시가격 예측 회귀식은 다음과 같다.

$$Y_1 = -265,968,595 + 4,421,887.299 \times X_1 \\ + 5,701,261.462 \times X_2 + 298,074,080.5 \times X_3 \\ + 2,855,275.946 \times X_4$$

Y_1 = 공시가격
 X_1 = 전용면적
 X_2 = 최고층수
 X_3 = 세대 당 사설 학원 수
 X_4 = 법정동 인구 내 학생 수 비율

공시가격에 영향을 미치는 요인은 전용면적과, 최고층수, 세대 당 사설 학원 수, 법정동 인구 내 학생 수 비율로 도출이 되었다. [표 4-3]의 모형요약을 보면 회귀식의 설명력을 나타내는 수정결정계수 값인 adj R^2 값이 0.716으로 높게 나타났다. adj R^2 값은 일반적으로 학계에서는 0.6이상, 마케팅조사 실무에서는 0.3이상이면 의미가 있는 수치로 해석을 한다. 이 수치가 의미하는 것은 원인변수가 결과변수

14) 회귀 분석에서의 일부 예측 변수가 다른 예측 변수와 상관 정도가 높아, 부정적 영향을 미치는 현상

15) 분산분석(ANOVA)은 관측 자료가 몇 개의 그룹으로 구분된 경우 그룹 평균 간 차이를 그룹 내 변동에 비교하여 살펴보는 데이터 분석 방법

를 설명하는 정도이다. DW(Durbin-Watson)¹⁶⁾ 값도 1.871로 2에 가까운 수치를 나타냈다. 일반적으로 DW값은 2에 가깝고, 0 또는 4와 가깝지 않으면 잔차들 간에 상관관계가 없어 회귀모형이 적합다고 본다. [표 4-4]의 ANOVA 분산분석 결과 또한 F-검정에 의한 유의확률 p값이 0.05보다 작은 0.001이 나왔다. 다중공선성의 문제를 다룬 분산팽창계수 VIF(variation inflation factor)값을 확인한 결과 [표 4-5]와 같이 변수 모두 10미만이며, 이는 변수간의 상관관계가 낮음을 의미한다. 회귀계수 β 값에 대한 유의성 검정을 위해 t분포에 의한 p값을 확인한 결과 모두 0.05 미만으로 유의미한 결과가 도출되었다.

[표 4-3] 공시가격 모형 요약

모형	R	R^2	$adj R^2$	추정값의 표준오차	Durbin-Watson
1	0.788	0.620	0.617	70,310,051.65	
2	0.836	0.698	0.694	62,902,366.80	
3	0.846	0.716	0.709	61,322,964.10	
4	0.851	0.725	0.716	60,593,402.76	1.871

1. 예측자 : (상수), 전용 면적
2. 예측자 : (상수), 전용 면적, 최고 층수
3. 예측자 : (상수), 전용 면적, 최고 층수, 세대 당 사설 학원 수
4. 예측자 : (상수), 전용 면적, 최고 층수, 세대 당 사설 학원 수, 법정동 인구 내 초·중·고 학생 수 비율

* 종속변수 : 공시가격

16) Durbin-Watson Test : 회귀분석 후 잔차의 독립성을 확인할 때 사용하며, 잔차끼리 자기상관성 유무를 판단하는 검증방법, Durbin-Watson 통계량 = d , $0 < d < 4$ 사이의 값을 가진다.

[표 4-4] 공시가격 ANOVA 분산분석

모형		제곱합	자유도	평균제곱	F	유의확률 p
4	회귀	1.188E+18	4	2.971E+17	80.922	0.001
	잔차	4.516E+17	123	3.672E+15		
	전체	1.640E+18	127			

1. 예측자 : (상수), 전용 면적
2. 예측자 : (상수), 전용 면적, 최고 층수
3. 예측자 : (상수), 전용 면적, 최고 층수, 세대 당 사설 학원 수
4. 예측자 : (상수), 전용 면적, 최고 층수, 세대 당 사설 학원 수, 법정동 인구 내 초·중·고 학생 수 비율

※ 종속변수 : 공시가격

[표 4-5] 공시가격 회귀계수, p값, 공선성 통계량 검토

모형	비표준화 계수			표준화 계수	t(p)	공선성 통계량	
	β	표준화 오류	베타			공차	VIF
상수	-265,968,595	26,152,638.48			-10.170 ***		
전용 면적	4,421,887.299	328,869.706	0.683	13.446 ***	0.866	1.154	
최고 층수	5,701,261.462	1,223,969.753	0.265	4.658 ***	0.694	1.441	
세대 당 사설 학원 수	298,074,080.5	108,897,963.6	0.132	2.737 **	0.969	1.032	
법정동 인구 내 초·중·고 학생 수 비율	2,855,275.946	1,426,930.983	0.110	2.001 *	0.737	1.357	

*p<0.05, **p<0.01, ***p<0.001

4.2 실거래가격의 요인 분석 및 통계학적 검증

4.2.1 실거래가격과 독립변수의 상관관계 분석

[표 4-6]은 실거래가격 회귀분석에 들어가기에 앞서 Pearson 상관관계 분석을 통하여 도출된 상관계수이다. 대부분의 변수가 실거래가격과 양의 상관관계를 가지고 있으며, 세대 당 초·중·고 학교 수, 1인당 공원 면적, 공공임대 브랜드 비율, 경과 연수가 음의 상관관계를 가지고 있다. 실거래가격은 동구 동내동의 3년 실거래가 없었으므로 표본수가 127개이며, 이에 공시가격의 상관계수와 미묘한 차이가 있지만 그 차이는 크지 않은 것으로 판단하였다. 공시가격과 실거래가격 모두 최고총수와 세대 당 주차 대수, 경과 연수, 법정동 인구 내 초·중·고 학생 수 비율이 절댓값 0.5이상으로 도출되었다. 상관관계가 높은 수치는 아니므로 독립변수의 독립성이 보장되었다고 볼 수 있다고 판단하였고, 회귀분석을 진행하였다.

[표 4-6] 실거래가격과 변수의 pearson 상관관계 분석

	A	B	C	D	E	F	G	H	I	J	K	L	M
A	1												
B	.046	1											
C	-.018	.059	1										
D	-.113	-.051	-.041	1									
E	.199	.220	.042	-.197	1								
F	.354	-.055	-.103	-.050	-.078	1							
G	-.144	-.073	-.067	.010	-.101	.057	1						
H	.544	-.163	-.169	.152	-.059	.470	.021	1					
I	.434	-.013	-.207	.109	.022	.295	-.141	.549	1				
J	.263	.105	-.193	-.228	.359	.231	.052	.256	.103	1			
K	-.365	.244	.018	-.212	.316	-.324	-.037	-.546	-.397	.109	1		
L	.735	-.132	.081	-.070	.172	.212	-.338	.344	.463	.091	-.178	1	
M	.401	-.079	-.126	.070	-.324	.320	-.021	.504	.418	.269	-.339	.232	1

※ 변수 목록

- A : 실거래가격, B : 세대 당 사설 학원 수,
- C : 세대 당 초·중·고 학교 수, D : 1인당 공원 면적,
- E : 법정동 면적 내 역세권 면적 비율, F : 상위 브랜드 비율,
- G : 공공임대 브랜드 비율, H : 최고 층수, I : 세대 당 주차 대수,
- J : 법정동 면적 당 세대 수, K : 경과 연수, L : 전용 면적,
- M : 법정동 인구 내 초·중·고 학생 수 비율

4.2.2 실거래가격의 다중회귀분석 및 요인 도출

공시가격 회귀분석에 이어 실거래가격을 종속변수로 설정하고 회귀분석을 실시하였을 때 예측 회귀식은 다음과 같다.

$$\begin{aligned}Y_2 = & -201,884,355 + 5,643,560.797 \times X_1 \\& + 6,642,713.004 \times X_2 + 603,628,400.3 \times X_3 \\& - 3,586,637.530 \times X_4 + 8,628.711 \times X_5 \\& - 126,203.620 \times X_6\end{aligned}$$

Y_2 = 실거래가격

X_1 = 전용면적

X_2 = 최고 층수

X_3 = 세대당 사설 학원 수

X_4 = 경과 연도

X_5 = 법정동 면적 당 세대수

X_6 = 1인당 공원면적

실거래가격에 영향을 미치는 요인은 전용면적, 층수, 세대 당 사설 학원 수, 경과 연도, 법정동 면적 당 세대 수, 1인당 공원면적으로 도출이 되었다. [표 4-7]의 모형요약을 보면 $adj R^2$ 값이 0.702로 높게 나타났고, DW 값도 1.890으로 2에 가까운 수치를 나타냈다. [표 4-8]과 같이 ANOVA 분산분석 결과 또한 F-검정에 의한 유의확률 p값이 0.05보다 작은 0.001로 도출되었다. 분산팽창계수 VIF 값은 [표 4-9]와 같이 변수 모두 10미만이며, 회귀계수 β 값에 대해 t분포에 의한 p값을 확인한 결과 모두 0.05 미만으로 유의미한 결과가 도출되었다.

[표 4-7] 실거래가격 모형 요약

모형	R	R^2	$adj R^2$	추정값의 표준오차	Durbin-Watson
1	0.735	0.540	0.536	103,984,722.9	
2	0.798	0.636	0.630	92,839,190.07	
3	0.819	0.670	0.662	88,735,798.76	
4	0.828	0.686	0.676	86,965,619.34	
5	0.840	0.706	0.694	84,520,438.38	
6	0.847	0.717	0.702	83,302,441.07	1.890

1. 예측자 : (상수), 전용 면적
2. 예측자 : (상수), 전용 면적, 최고 층수
3. 예측자 : (상수), 전용 면적, 최고 층수, 세대 당 사설 학원 수
4. 예측자 : (상수), 전용 면적, 최고 층수, 세대 당 사설 학원 수, 경과 연수
5. 예측자 : (상수), 전용 면적, 최고 층수, 세대 당 사설 학원 수, 경과 연수, 법정동 면적 당 세대수
6. 예측자 : (상수), 전용 면적, 최고 층수, 세대 당 사설 학원 수, 경과 연수, 법정동 면적 당 세대수, 1인당 공원 면적

* 종속 변수 : 실거래가격

[표 4-8] 실거래가격 ANOVA 분산분석

모형	제곱합	자유도	평균제곱	F	유의확률 p
6	회귀	2.106E+18	6	3.510E+17	50.577
	잔차	8.327E+17	120	6.939E+15	
	전체	2.939E+18	126		

6. 예측자 : (상수), 전용 면적, 최고 층수, 세대 당 사설 학원 수, 경과 연수, 법정동 면적 당 세대수, 1인 당 공원 면적

* 종속 변수 : 실거래가격

[표 4-9] 실거래가격 회귀계수, p값, 공선성 통계량 검토

모형	비표준화 계수			표준화 계수 t(p)	공선성 통계량	
	β	표준화 오류	베타		공차	VIF
상수	-201,884,355	48,649,927.53		-4.150 ***		
전용면적	5,643,560.797	473,629.760	0.625	11.916 ***	0.859	1.164
최고층수	6,642,713.004	1,913,635.532	0.230	3.471 ***	0.537	1.861
세대당사설학원수	603,628,400.3	153,471,012.6	0.199	3.933 ***	0.922	1.084
경과연수	-3,586,637.530	1,039,819.820	-0.213	-3.449 ***	0.617	1.619
법정동면적당세대수	8,628.711	3,791.051	0.125	2.276 *	0.786	1.273
1인당공원면적	-126,203.620	59,072.778	-0.111	-2.136 *	0.879	1.138

*p<0.05, **p<0.01, ***p<0.001

실거래가격에서는 전용면적, 최고층수, 세대당사설학원수는 공시가격과 일치하는 변수이며, 경과연도, 법정동면적당세대수, 1인당공원면적이 새로운 변수로 추가되었다. 공시가격은 모두 양의 회귀계수를 띠는 반면에, 경과연도와 1인당공원면적이 음의 회귀계수를 띠고 있다. 경과연도는 건축물의 노후도로 볼 수 있으며, 노후

도의 증가가 아파트 가격에 반대로 작용할 것이라는 예상이 일반적
이기 때문에 음의 계수를 띠는 것을 예상 할 수 있었다. 1인당 공원
면적은 일반적으로 인근 거주민들의 삶의 질과 연결 되며 아파트 가
격도 상승된다는 예상과 달리 회귀계수가 음의 값을 띠면서 반대의
결과가 도출되는 것을 알 수 있었다.

영향을 미치는 요소와 반대로, 세대 당 주차대수, 상위브랜드비율,
공공임대브랜드비율, 세대 당 학교 수, 법정동 면적 당 역세권 면적
은 공시가격과 실거래가격 모두에 영향을 미치지 않는 변수라는 결
과를 도출했다.

4.3 다중회귀모델의 검증 및 가격 예측

‘단계 선택’ 방식으로 추정된 다중회귀모델을 검증하기 위해서, 도출된 회귀식을 사용하여 공시가격과 실거래가격의 예상값을 계산 했으며 이를 실제값과 비교하였다. 선정된 법정동 128개를 본 논문에 모두 기재하기에는 지면관계상 어려움이 있어 각 구별로 평균 오차율을 구했을 때 [표 4-10]과 같이 나타났다. 실제 공시가격과 회귀식으로 예상된 공시가격의 평균 오차율은 21.14%로 나타났다. 실제 실거래가격과 회귀식으로 예상된 실거래가격의 평균 오차율은 19.43%로 나타났다. 공시가격과 실거래가격 모두 동구와 서구에서 가장 큰 오차율을 보였고, 오차가 가장 큰 법정동을 분석한 결과 각 법정동에 아파트 동 전체 개수가 1개인 동들이 많았다. 1개 동의 데이터가 법정동 전체 값을 대변하게 되면서 오차가 증가한 것으로 보인다.

[표 4-10] 예상가격과 실제가격의 오차율

구, 군	오차율(%)	
	공시가격	실거래가격
남구	8.24	12.74
달서구	16.65	16.05
달성군	25.63	14.94
동구	34.11	21.98
북구	15.94	17.61
서구	29.85	35.42
수성구	18.32	15.48
중구	20.41	21.20
평균 오차율	21.14	19.43

이 오차를 보완하기 위해 전체 아파트 동 개수가 1개인 법정동 11개를 제거하고 117개의 법정동으로 다중회귀분석을 재설시하였다. 공시가격 다중회귀분석 결과 도출된 수정 회귀식은 다음과 같다.

$$Y_1 = -256,992,797 + 5,271,278.887 \times X_1 \\ + 4,335,491.776 \times X_2 + 273,987,128.6 \times X_3 \\ - 75,416.919 \times X_4 - 1,910,633.27 \times X_5 \\ + 5,179.301 \times X_6$$

Y_1 = 공시가격
 X_1 = 전용면적
 X_2 = 최고 층수
 X_3 = 세대당 사설 학원수
 X_4 = 1인당 공원면적
 X_5 = 경과 연도
 X_6 = 법정동 면적당 세대수

adj R² 값은 0.762로 증가하였고, DW값은 2.024, ANOVA 분산분석 p값 0.001, VIF값 모두 10미만, 회귀계수 β 의 p값 모두 0.05미만으로 유의미한 값이 나왔다. 회귀분석 데이터 자료는 [표 4-11], [표 4-12], [표 4-13]과 같다.

[표 4-11] 수정된 공시가격 모형 요약

모형	R	R^2	$adj R^2$	추정값의 표준오차	Durbin-Watson
1	0.818	0.668	0.666	63,158,588.98	
2	0.856	0.733	0.729	56,902,882.16	
3	0.864	0.747	0.740	55,649,133.39	
4	0.870	0.757	0.748	54,842,756.06	
5	0.875	0.766	0.755	54,035,191.12	
6	0.880	0.775	0.762	53,239,762.75	2.024

1. 예측자 : (상수), 전용 면적
2. 예측자 : (상수), 전용 면적, 최고 층수
3. 예측자 : (상수), 전용 면적, 최고 층수, 세대 당 사설 학원 수
4. 예측자 : (상수), 전용 면적, 최고 층수, 세대 당 사설 학원 수, 1인당 공원 면적
5. 예측자 : (상수), 전용 면적, 최고 층수, 세대 당 사설 학원 수, 1인당 공원 면적, 경과 연수
6. 예측자 : (상수), 전용 면적, 최고 층수, 세대 당 사설 학원 수, 1인당 공원 면적, 경과 연수, 법정동 면적 당 세대수

* 종속변수 : 공시가격

[표 4-12] 수정된 공시가격 ANOVA 분산분석

모형	제곱합	자유도	평균제곱	F	유의 확률 p
4	회귀	1.072E+18	6	1.786E+17	63.014
	잔차	3.118E+17	110	2.834E+15	
	전체	1.383E+18	116		

6. 예측자 : (상수), 전용 면적, 최고 층수, 세대 당 사설 학원 수, 1인당 공원 면적, 경과 연수, 법정동 면적 당 세대수

* 종속변수 : 공시가격

[표 4-13] 수정된 공시가격 회귀계수, p값, 공선성 통계량 검토

모형	비표준화 계수			표준화 계수 t(p)	공선성 통계량	
	β	표준화 오류	베타		공차	VIF
상수	-256,992,797	34,266,952.08		-7.500 ***		
전용 면적	5,271,278.887	347,800.068	0.738	15.156 ***	0.865	1.156
최고 층수	4,335,491.776	1,306,120.739	0.202	3.319 ***	0.554	1.803
세대 당 사설 학원 수	273,987,128.6	103,099,390.6	0.128	2.658 **	0.881	1.135
1인당 공원 면적	-75,416.919	38,631.754	-0.095	-1.952 *	0.863	1.159
경과 연수	-1,910,633.270	745,468.556	-0.149	-2.563 *	0.602	1.660
법정동 면적 당 세대 수	5,179.301	2,485.696	0.106	2.084 *	0.787	1.271

*p<0.05, **p<0.01, ***p<0.001

실거래가격 또한 [표 4-14]와 같이 $adj\ R^2$ 값이 0.709로 약간 증가하였고, Durbin-Watson값 1.906으로 도출되었다. [표 4-15]와 같이 ANOVA 분산분석 p값 0.001, VIF값 모두 10미만으로 도출되었고, [표 4-16]과 같이 회귀계수 β 의 p값 모두 0.05미만으로 유의미한 값이 나왔다. 실거래가격의 수정된 회귀식은 다음과 같다.

$$\begin{aligned}Y_2 = & -212,540,332 + 6,190,382.796 \times X_1 \\& + 5,398,221.178 \times X_2 + 502,806,743 \times X_3 \\& - 4,116,450.391 \times X_4 + 9,135.435 \times X_5 \\& - 133,167.771 \times X_6\end{aligned}$$

Y_2 = 실거래가격

X_1 = 전용면적

X_2 = 최고 층수

X_3 = 세대당 사설 학원수

X_4 = 경과 연수

X_5 = 법정동 면적당 세대수

X_6 = 1인당 공원면적

[표 4-14] 수정된 실거래가격 모형 요약

모형	R	R^2	$adj R^2$	추정값의 표준오차	Durbin-Watson
1	0.747	0.558	0.554	95,701,172.45	
2	0.801	0.641	0.635	86,564,899.48	
3	0.816	0.666	0.657	83,928,127.36	
4	0.827	0.684	0.673	81,929,685.92	
5	0.843	0.710	0.697	78,867,976.67	
6	0.851	0.724	0.709	77,269,114.65	1.906

1. 예측자 : (상수), 전용 면적
2. 예측자 : (상수), 전용 면적, 최고 층수
3. 예측자 : (상수), 전용 면적, 최고 층수, 세대 당 사설 학원 수
4. 예측자 : (상수), 전용 면적, 최고 층수, 세대 당 사설 학원 수, 경과 연수
5. 예측자 : (상수), 전용 면적, 최고 층수, 세대 당 사설 학원 수, 경과 연수, 법정동 면적 당 세대수
6. 예측자 : (상수), 전용 면적, 최고 층수, 세대 당 사설 학원 수, 경과 연수, 법정동 면적 당 세대수, 1인당 공원면적

※ 종속변수 : 실거래가격

[표 4-15] 수정된 실거래가격 ANOVA 분산분석

모형	제곱합	자유도	평균제곱	F	유의 확률 p
4	회귀	1.724E+18	6	2.874E+17	48.139
	잔차	6.568E+17	110	5.971E+15	
	전체	2.381E+18	116		

6. 예측자 : (상수), 전용 면적, 최고 층수, 세대 당 사설 학원 수, 경과 연수, 법정동 면적 당 세대수, 1인당 공원면적

※ 종속변수 : 실거래가격

[표 4-16] 수정된 실거래가격 회귀계수, p값, 공선성 통계량 검토

모형	비표준화 계수			표준화 계수 t(p)	공선성 통계량	
	β	표준화 오류	베타		공차	VIF
상수	-212,540,332	49,733,073.79		-4.274 ***		
전용 면적	6,190,382.796	504,776.917	0.660	12.264 ***	0.865	1.156
최고 층수	5,398,221.178	1,895,628.154	0.191	2.848 **	0.554	1.803
세대 당 사설 학원 수	502,806,743.0	149,632,496.9	0.179	3.360 **	0.881	1.135
경과 연수	-4,116,450.391	1,081,929.978	-0.245	-3.805 ***	0.602	1.660
법정동 면적 당 세대 수	9,135.435	3,607.595	0.143	2.532 *	0.787	1.271
1인당 공원 면적	-133,167.771	56,067.895	-0.128	-2.375 *	0.863	1.159

*p<0.05, **p<0.01, ***p<0.001

각 회귀식의 변수들이 일치하면서 공시가격과 실거래가격 모두 같은 요인으로 가격이 결정되는 결과가 도출되었다. 새로운 회귀식으로 다시 오차율을 구해보았을 때 그 값은 [표 4-17]과 같다. 공시가격은 17.43%로 3.71%가 감소하였고, 실거래가도 17.20%로 2.23%가 감소하였다. 특히 서구와 동구에서 많이 감소하면서 전체 오차율이 많이 낮아졌다. 두 가격 회귀모델 모두 정확도가 82%이상으로 높은 수준으로 나왔다.

[표 4-17] 수정된 예상가격과 실제가격의 오차율

구, 군	오차율(%)	
	공시가격	실거래가격
남구	13.01	13.69
달서구	20.91	17.21
달성군	12.50	14.76
동구	24.36	21.62
북구	17.18	18.40
서구	15.55	17.09
수성구	15.15	14.16
중구	20.79	20.64
평균 오차율	17.43	17.20

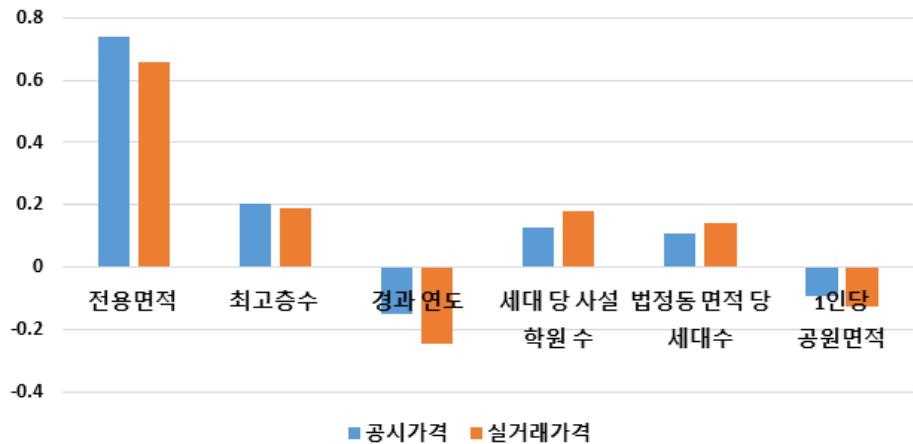
제 5 장 결 론

본 연구에서는 빅데이터와 직접 수집한 데이터를 활용하여 다중회귀분석을 통해 대구광역시 법정동 단위로 공시가격과 실거래가격에 영향을 미치는 요인을 분석하였다. 도출된 요인을 수정된 두 회귀모델 표준화 계수 값은 [표 5-1]과 같으며, 이를 보기 쉽게 그래프로 나타나면 [그림 5-1]과 같다. 이를 통해 대구광역시의 특성을 분석한 결과는 다음과 같다.

[표 5-1] 최종 회귀모델 표준화 계수 값 비교

변 수	표준화 계수 값	
	공시가격	실거래가
전용 면적	0.738	0.660
최고 층수	0.202	0.191
세대 당 사설 학원 수	0.128	0.179
경과 연도	-0.149	-0.245
법정동 면적 당 세대수	0.106	0.143
1인당 공원 면적	-0.095	-0.128

공시가격과 실거래가격 영향도 분석 그래프



[그림 5-1] 공시가격과 실거래가격 영향도 분석 그래프

첫 번째, 대구광역시 전체를 보면 전용면적, 최고층수, 경과 연도의 표준화계수 절댓값이 가장 높으므로 가격에 가장 높은 비중을 차지하고 있다. 면적, 높이, 시간의 개념인 물리적인 아파트 수치가 가격에 가장 큰 영향을 미치는 것이다.

두 번째, 세대 당 사설학원수가 많은 곳이 가격이 더 높다는 것과 교육 시설이 지하철과 같은 교통시설보다 아파트 가격에 더 영향을 미치는 것을 알 수 있었다. 대구광역시 수성구는 학군으로도 유명한데, 실거래가 상위 10위권에 1~8위까지의 법정동이 모두 수성구이다. 하지만 [표 5-2]처럼 세대 당 사설학원 수의 상위 10위권에 수성구는 수성동2가, 범어동 2개의 법정동만 존재하며, 중구, 동구, 달서구, 북구, 서구의 법정동이 고르게 10위권에 분포되어 있다. 이것으로 수성구뿐만 아니라 대구광역시 전체에서 교육환경이 좋은 곳이 가격이 높다고 볼 수 있으며, 법정동 별로 교육환경에 불균형이 존재한다는 의미이기도하다.

[표 5-2] 세대 당 사설 학원 수 상위 10위

순위	법정동	세대 당 사설 학원 수
1	중구 도원동	0.550
2	수성구 수성동2가	0.129
3	동구 서호동	0.074
4	서구 원대동3가	0.060
5	중구 봉산동	0.051
6	달서구 대천동	0.042
7	중구 삼덕동2가	0.037
8	수성구 범어동	0.022
9	북구 학정동	0.021
10	수성구 만촌동	0.020

세 번째, 1인당 공원면적이 클수록 가격이 떨어지는 원인을 분석해봤을 때, 대구의 지역적 특성이 원인인 것으로 예상된다. 대구는 사방이 산으로 둘러싸인 분지로 도심 외곽지역이 대부분 산이다. 도심 외곽에 위치한 법정동이 산과 접하게 되면서 공원면적이 압도적으로 증가하게 된다. 도심과 멀어진 법정동은 상대적으로 도심에 비해 아파트 가격이 낮기 때문에 발생한 요인으로 예상된다. 산과 접한 대공원을 제외하고, 소공원만을 데이터로 수집했을 때는 다른 결과를 도출 할 수도 있을 것이다.

네 번째, 법정동 면적 당 세대수가 많을수록 가격이 증가되는 것을 분석해봤을 때, 법정동 내에서 세대 밀도라고 볼 수 있는 이 값이 높으면 아파트 가격이 높게 측정되었다고 볼 수 있다. 세대 밀도가 높을수록 병원, 학교, 마트, 편의점 등 각종 주변시설이 들어오게 되며 생활환경이 좋아지기 때문이다. 이를 통해 대구광역시는 세대 밀도에 따른 주변시설이 잘 갖춰져 있다는 것을 알 수 있다.

본 연구에서는 12가지 요인을 통해 다중회귀분석으로 대구광역시의 공시가격과 실거래가격을 분석해보았다. 이를 통해 최종적으로 공통적인 6가지 요인을 도출 할 수 있었고, 이를 바탕으로 대구광역시의 아파트 가격 특징을 분석하였다. 추후에 본 연구를 통해 아파트 실수요자들이 아파트 선택에 있어서 적정가격을 예측할 수 있을 것이며, 이를 통해 입지(동네, 위치)를 선택함에 있어서 가격에 대한 근거를 마련할 수 있을 것으로 사료 된다. 12개의 요인 외에 각 실의 향, 각 실의 조망, 외부 및 내부의 마감, 방의 수, 욕실의 수, bay수 등등 건축적인 내부 요소 또한 중요한 변수로 작용 할 수 있지만, 기존 데이터나 빅데이터의 부재로 직접 추출하기에는 매우 광범위하다. 본 논문의 범위가 대구광역시 전수조사임을 고려하면 이를 한계점으로 꼽을 수 있다. 아파트 가격에 영향을 미치는 모든 요소들은 굉장히 다양하면서 복잡하고 주관적인 의견 또한 많이 들어가기에 100% 예측은 어렵다고 볼 수 있다. 또한 도시, 하천, 숲을 볼 수 있는 조망 또한 지역적인 위치에 따라 수요자들의 성향도 달라질 것이며, 이에 대한 형평성 또한 고려해야 한다. 하지만 본 논문과 같은 연구를 통해 객관적인 요소들을 찾을 수 있으며, 공통적인 요소들로 추릴 수 있을 것이다. 추후에 데이터의 다양화 또는 인공지능을 이용해 다양하고 복잡한 모든 요인들에 대한 데이터를 추출할 수 있다면 향후의 연구방향으로 설정하여 가격 요소에 더욱 자세하게 요인 분석을 진행 할 수도 있을 것이다. 이를 통해 아파트 가격에 대한 요인들이 더

속 다양해 질 수도 있고 추출된 요인이 더욱 논리적일 것이다.

참 고 문 헌

■ 학회지 및 학술발표 논문

1. 도수관, 이규태, 김은지, 도시철도 역세권과 아파트 가격간의 관계 분석, 한국정부학회 학술발표논문집 231-249, 2015
2. 장동현, 김한수, 공간계량모형을 활용한 아파트가격 영향요인 분석 연구 - 대구 시 수성구를 중심으로, 한국주거학회논문집 31(1), 79-86, 2020
3. 이병훈, 이백래, 김주형, 김재준, 회귀분석을 통한 부동산투자회사(REITs) 오피스 빌딩의 수익률 영향요인 분석, 대한건축학회 논문집 - 계획계 27.9, 293-300, 2011
4. 김진원, 이백래, 김주형, 김재준, 회귀분석을 이용한 교육시설의 공간계획에 따른 공사비 예측 모델에 관한 연구, 대한건축학회 논문집 - 계획계, 27.10, 103-110, 2011
5. 강남규, 김순호, 강은혜, 최정민, 주거용 건물의 에너지 소비량 특성 및 선형 회귀모형 개발에 관한 연구 - 2018년 데이터를 중심으로, 대한건축학회논문집 36.7, 115-122, 2020
6. 양성필, 손기영, 이경훈, 김지명, 태풍타입별 피해 분석 및 다중회귀분석을 활용한 태풍피해예측모델 개발 연구, 한국건축시공학회지, 16(5), 437-445, 2016
7. 이강, 최근희, 해도닉 가격모형을 활용한 주택가격 결정요인에 관한 연구, 한국도시행정학회 학술발표대회 논문집, 317-333, 2016
8. 박운선, 임병준, 해도닉 가격모형을 활용한 아파트 가격결정요인 분석, 대한부동산학회 논문집, 28.2, 245-271, 2010
9. 김희호, 박세운, 서울 주택가격의 결정요인: 분위수 회귀분석, 한국주택학회 주택연구논문집 21.2, 141-168, 2013
10. 송호창, 김태호, 이주형, 주상복합아파트의 주택규모별 가격결정요인 분석, 서울도시연구 9.3, 79-92, 2008
11. 이재영, 박태원, 공공임대주택이 주택 매매 및 전세 가격에 미치는 영향 - 서울

울 지역을 중심으로, 부동산분석 2.1, 53-65, 2016

12. 정수연, 교육요인이 서울아파트가격에 미치는 영향에 관한 연구, 국토계획
41.2, 153-166, 2006

■ 보고서 및 단행본

1. 김규태, 권민석, 최경민, 학위논문과 통계분석, 2016
2. 김지형, Excel+R 누구나 하는 데이터 분석, 2020
3. 김권현, R로 하는 빅데이터 분석 :데이터전처리와 시각화 :개념적 기초에서 심층 활용까지, 2019
4. 김진화, 박성택, 이성원, 빅데이터 R point:빅데이터 분석 기본, 2018
5. 노경섭, 제대로 알고 쓰는 논문 통계분석 SPSS & AMOS, 2019

A study on multiple regression models for analyzing factors that influence apartment prices and predicting prices based on big-data*

- Focusing on legal district(dong), Daegu Metropolitan City -

Seonhyeon Kim

*Department of Architecture
Graduate School, Kyungpook National University
Daegu, Korea
(Supervised by Professor Seungyeon Choo)*

(Abstract)

In Korea, real estate prices have risen to their highest level in the past five years. Income levels are not supported, but tend to rush to buy apartments through excessive loans. It is necessary to find factors that affect apartment prices for reasonable purchases by consumers. However, previous studies have limited interest in some variable effects or limited analysis of specific regions. In this paper, we investigated all the apartments in legal district(dong), Daegu Metropolitan City. Excluding legal district(dong) where there are no apartments, it was classified into 128 legal district(dong). and we tried to find factors that affect the public price and actual transaction price of apartments through multiple regression analysis. It was found that the error rate of the analysis was high when the number of apartments was one in the legal district(dong). Regression analysis was conducted again except for legal district(dong), which has one apartment building in legal district(dong). As a result, we found six factors that affect apartment prices. Specifically, physical factors such as the

* A thesis submitted to the Council of Graduate School of Kyungpook National University in partial fulfillment of the requirements for the degree of Master of Science in December 2021

exclusive area of apartments, elapsed years, and the number of floors, the number of private academies per household, the number of households per legal district(dong) area, and the park area per person were confirmed. As the park area per person increased, the results were different from those expected to decrease in apartment prices.