

서울 강남 아파트와 강원도 춘천 아파트의 특성 탐색

조 : B조 강현욱, 김예지, 김주영

이변량 분석 담당 및 탐색

1. 서론

2. 데이터 및 변수 소개

3. 강남 아파트 특성

- 1) 분포탐색
- 2) 로버스트 회귀
- 3) 이변량 자료분석
- 4) 다변량 자료분석

4. 춘천 아파트 특성

- 1) 분포탐색
- 2) 로버스트 회귀
- 3) 이변량 자료분석
- 4) 다변량 자료분석

5. 시계열 분석

- 1) 강남 아파트 가격
- 2) 춘천 아파트 가격

6. 결론

- 결론 및 논의

1. 서론

강남 아파트는 우리나라에서 가장 비싼 값으로 거래된다. 아파트 가격에 영향을 미치는 변수는 학군, 교통편 등 많은 영향이 있을 것이며, 그런 점에서 강남 아파트가 비싼 것은 당연해 보인다.

그러나, 본 조는 아파트가 비싸다면 어떤 요인이 이 비싼 가격을 결정했는지 의문이 들었다. 따라서, 강남 아파트와 춘천 아파트의 특성을 비교하여 강남 아파트에서 비싼 가격을 결정한 요인들이 타 지역인 춘천 아파트에도 이 요인들이 동일하게 영향을 행사하는지 알아보고자 한다.

※ 자료 분석 방법

본 연구는 자료의 통계분석을 위해 수집된 자료를 가공 및 처리하였으며, 기본 분포탐색, 로버스트회귀, 이변량 자료분석, 다변량 자료분석 방법을 이용해 분석하였다. 또한, R 및 Excel 2010을 이용하여 분석하였다.

첫째, 각 지역의 아파트 매매가격으로 분석을 시행하였다.

둘째, 연 단위로 자료를 수집하였으며, 기간은 2023년 1월부터 2024년 1월 자료를 사용하였다.

셋째, 시공사 규모는 중소기업 기본법을 기준으로 판단하였다.

2. 데이터 및 변수 소개

데이터는 국토교통부 아파트 실거래가 데이터를 사용하였다.

전처리가 크게 필요 없으며, 원하는 기간에 맞춰 데이터를 손쉽게 추출할 수 있기 때문에 분석에 사용하기 좋다고 생각하였다.

변수는 다음과 같다. 이 중, 시군구1과 시군구2는 데이터 나누기를 사용해 나눈 것이고, 시공사, 시공사 규모는 해당 아파트의 정보를 참고해 추가했으며, 평당 가격은 1평(약 $3m^2$)에 맞게 변환해준 값이다.

시

시군구1

시군구2

번지

본번

부번

단지명

전용면적(m^2)

계약년월

계약일

거래금액(만원)

동

층

매수자

매도자

건축년도

도로명

해제사유발생일

거래유형

중개사소재지

등기일자

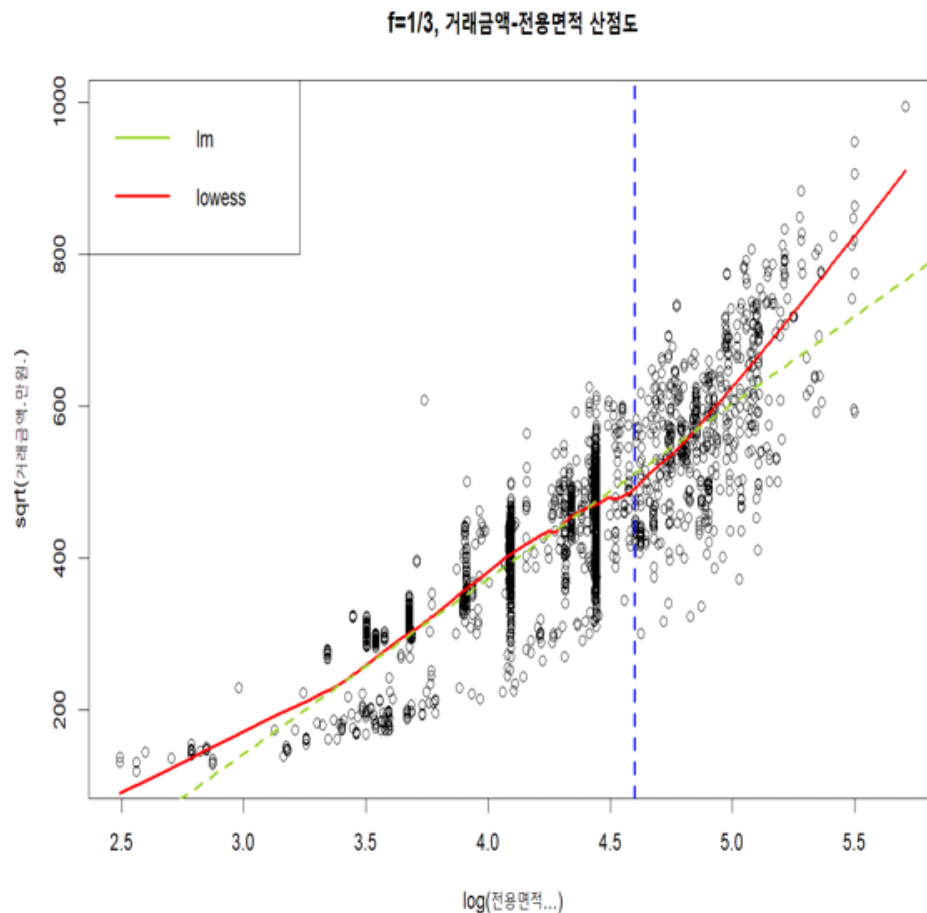
시공사, 시공사 규모, 평당 가격

강남 아파트 이변량 자료분석

3-1. 강남 아파트 실거래가(매매) 자료 분석

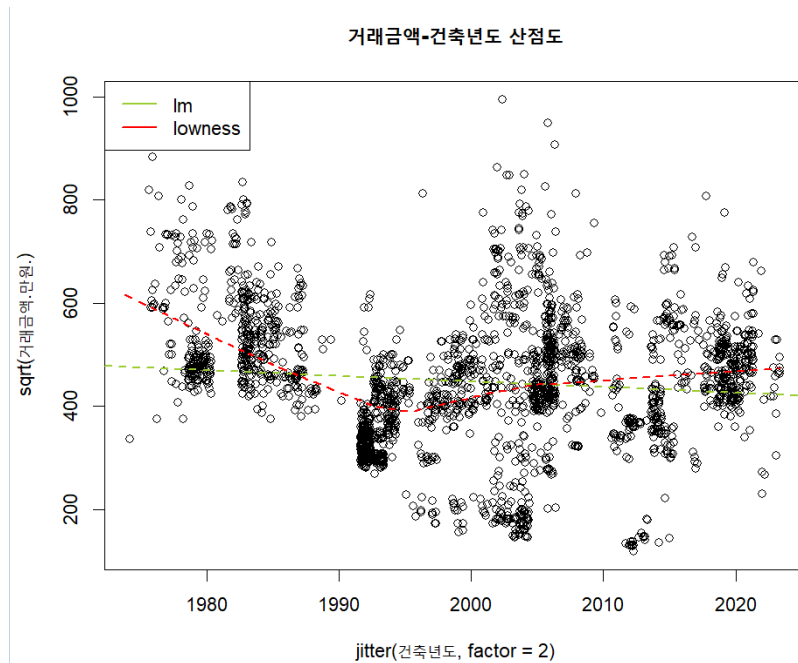
3-1-1. 산점도

(1) 거래금액(만원) - 전용면적(m^2) 산점도



다음은 건축년도에 호트림효과를 주어 거래금액-건축년도 간 산점도를 그리고 그 위에 최소제곱 회귀직선(녹색), 평활너비를 1/3으로 설정한 조각별 선형회귀를 적용했다. 우상향하는 추세가 보인다. 아파트의 전용면적이 커질수록 아파트의 거래금액이 커지는 것을 확인할 수 있다. 전용면적 $90m^2$ 부터 로웨스 곡선이 더 가파르게 상승한다.

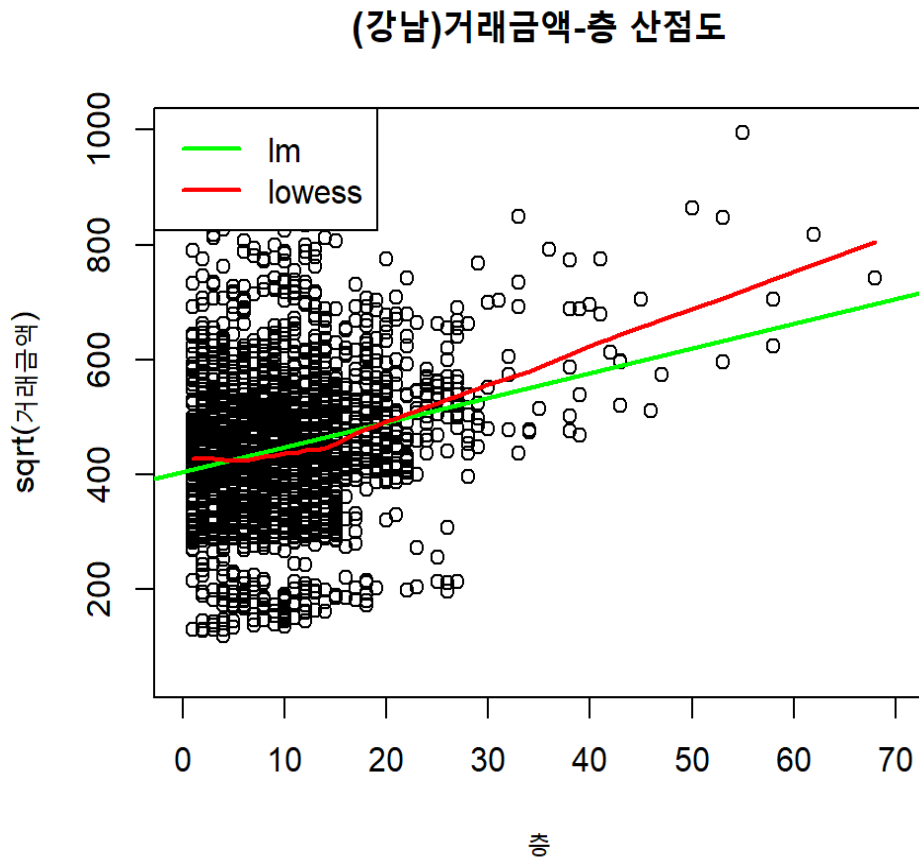
(2) 거래금액(만원) - 건축년도 산점도



다음은 건축년도에 호트림효과를 주어 거래금액-건축년도 간 산점도를 그리고 그 위에 최소제곱 회귀직선(녹색), 평활너비를 2/3으로 설정한 로웨스 곡선을 그린 것이다. 최소제곱직선과 로웨스 곡선을 보면, 신식 아파트일수록 비쌀 것이라는 일반적인 짐작과는 다르게, 구식 아파트가 더 비싸다는 결과가 나온다.

로웨스 곡선은 아파트의 건축연도별 거래금액의 추세를 보다 자세히 보여준다. 1980년대 건설된 아파트가 가장 비싸고, 1990년대 이후 건설된 아파트가 가장 저렴하며, 2000년대 이후 건설된 아파트는 가격변화 없이 상수에 가까운 가격으로 거래되고 있다.

(3) 거래금액(만원) - 층 산점도

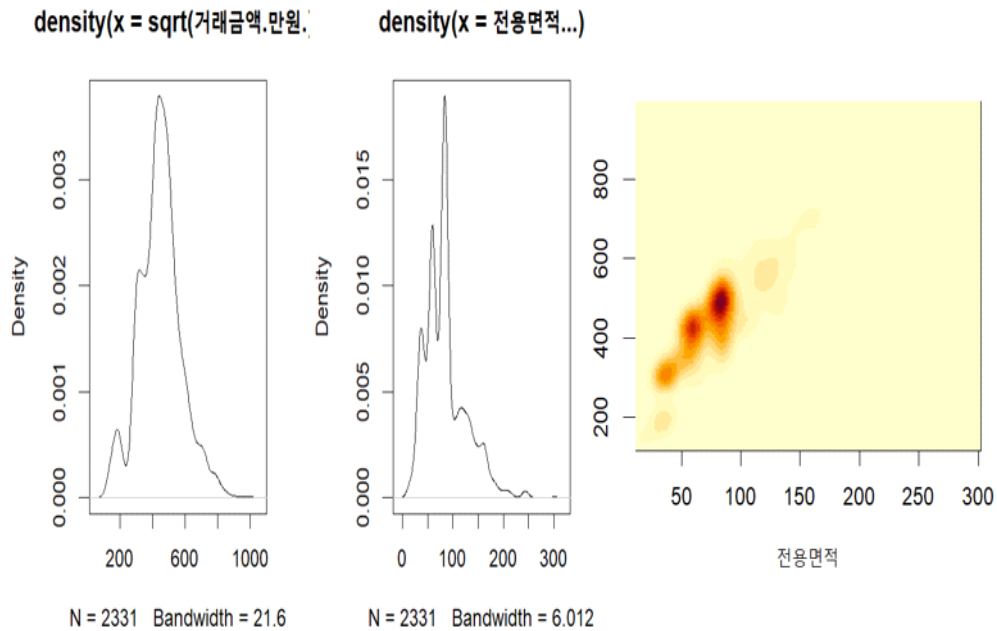


위의 그림은 건축년도와 층의 산점도 위에 최소제곱 회귀직선(녹색)과 평활너비를 2/3으로 한 조각별 선형회귀직선을 그린 것이다. 산점도로는 15층 이하에 자료점이 몰려있어 분포의 특징을 찾기 힘들지만, 회귀직선 자체만 보았을 때 우상향하는 결과를 볼 수 있으므로 층의 높이가 올라갈수록 아파트의 거래가격도 올라간다고 할 수 있다.

로웨스 곡선은 15층까지는 상수로 유지되다 15층 이후 선형회귀모형보다 가파른 기울기로 상승한다. 로웨스 곡선을 볼 때, 고층 아파트의 거래금액 상승은 확실해 보인다. 다만 고층 아파트의 거래 표본이 적고, 전체적인 자료의 분산이 커 추세를 확신하기는 어렵다. 지금의 24년 4월 자료가 아니라 표본이 더 많은 연단위 자료를 분석하면 확실한 결과를 얻을 수 있으리라 기대된다.

3-1-2. 커널 밀도추정 및 등고선 그림

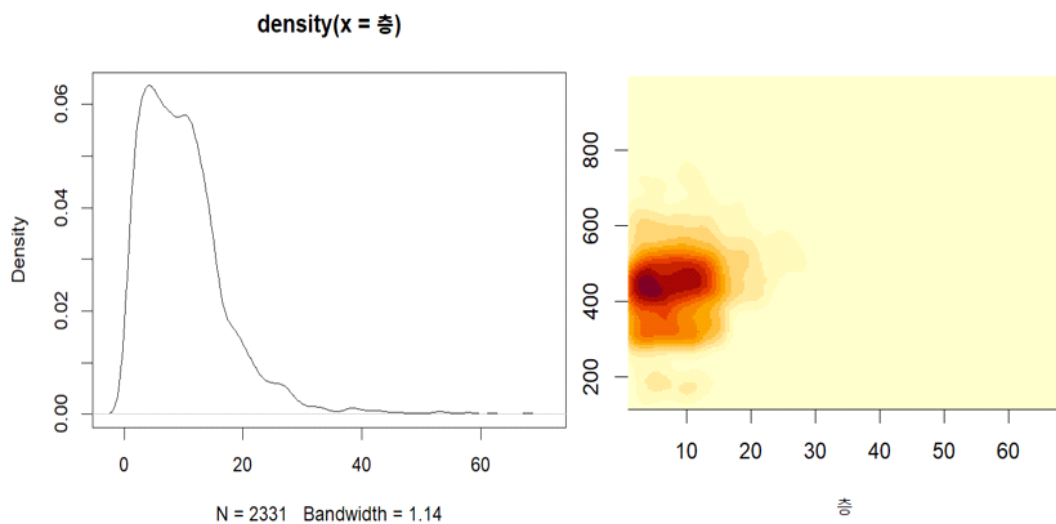
(1) 거래금액(만원)과 전용면적(m²)



왼편의 그림은 거래금액의 밀도함수이고, 중앙 그림은 전용면적의 밀도함수이다. 거래금액을 대칭분포로 맞추고 금액의 파악을 쉽게 하기 위해 제곱근 변환한 값을 사용했다. 거래금액의 밀도함수는 큰 봉우리 하나, 매우 작은 봉우리가 두 개 있어 한 개의 군집이 있으리라 예상된다. 한편 전용면적은 큰 봉우리 두 개, 작은 봉우리 한 개가 있어 두 개 혹은 세 개의 군집이 있으리라 예상된다. 가장 높은 봉우리는 각각 거래금액 20억, 면적 90m²에 자리 잡고 있다.

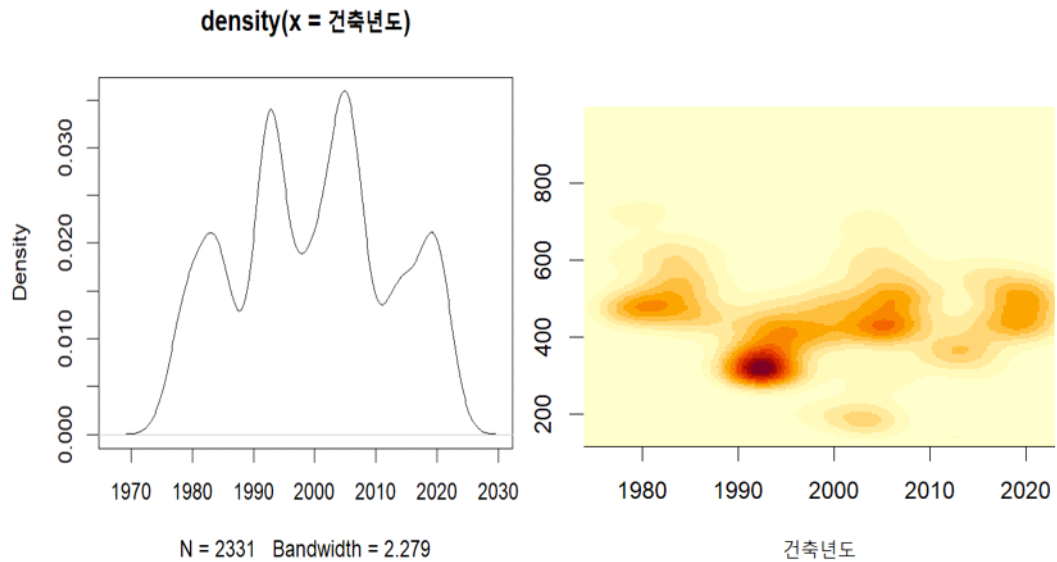
오른편의 그림은 거래금액(만원)의 제곱근 변환한 값들과 전용면적 간의 등고선 그림이다. 세 개의 군집을 관찰할 수 있다. 세 군집은 전용면적이 커질수록 거래금액이 비싸지는 정비례 관계에 있다. 커널 밀도함수에서도 확인할 수 있었듯, 거래금액 대략 20억, 전용면적 대략 90m²부분이 가장 짙은 색을 띠므로 가장 자료가 많이 밀집되어 있다는 것을 알 수 있다. 해당 군집의 색이 가장 진하기 때문이다.

(2) 거래금액(만원)과 층



왼편은 층에 대한 커널 밀도함수이다. 거래된 매물이 대략 1-10층 사이에 가장 많음을 확인할 수 있다. 밀도함수의 오른쪽 부분이 끊겨있으니, 최댓값을 이상치로 의심할 수 있다. 전체적으로 보았을 때, 봉우리는 매우 큰 봉우리 하나가 존재한다. 따라서 한 개 군집이 있으리라 기대할 수 있다. 오른쪽은 제공된 변환을 취한 거래금액과 층의 등고선 그림이다. 한 개의 군집이 존재한다. 거래금액 약 20억, 층수 약 5층에 자료가 밀집해 있다.

(3) 거래금액(만원)과 건축년도



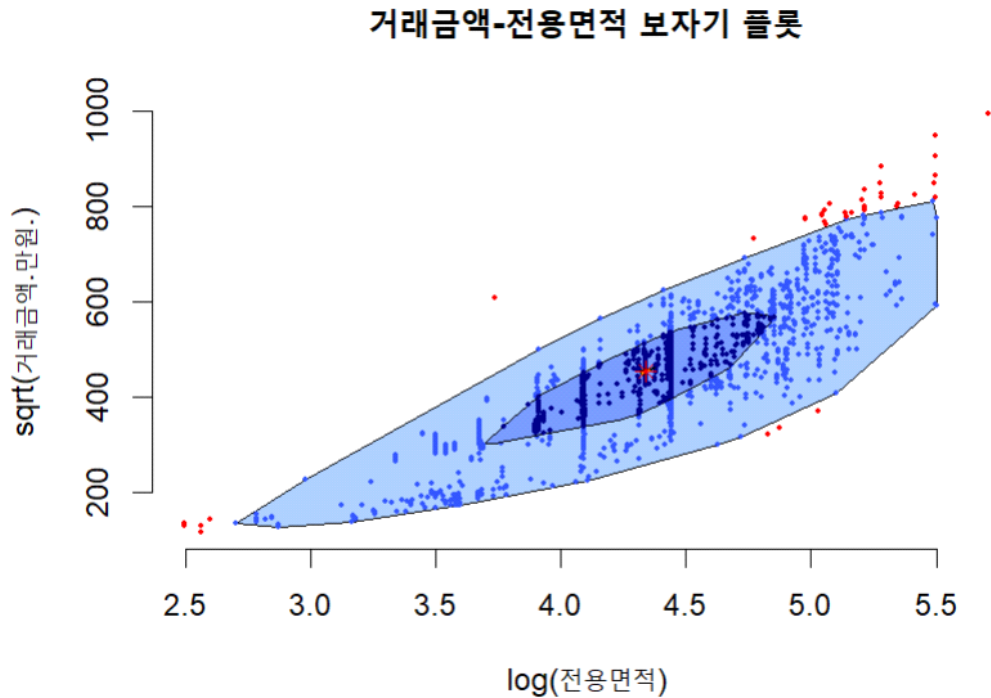
왼편은 건축년도의 커널 밀도함수 그림이다. 네 개 봉우리가 보인다. 1980, 1990, 2005, 2020에 건축된 아파트의 빈도가 높다. 따라서 해당 연도에 네 개 군집이 있으리라 짐작할 수 있다.

등고선 그림에서도 주된 4개의 군집이 보인다. 왼쪽부터 차례로 (거래금액 25억, 건축년도 1980년대), (거래금액 9억, 건축년도 1990년대 초), (거래금액 20억, 건축년도 2005년대), (거래금액 20억, 건축년도 2020년대 초)로 볼 수 있다. 그 외에 2000년대 초 거래금액이 4억 정도인 낮은 밀도의 군집이 하나 더 있다. 이것은 연도의 밀도함수의 특징이 아닌 거래금액 밀도함수의 특징이 반영된 결과로 보인다.

특이하게도 1990년도 초에 건축된 아파트의 빈도가 매우 높은데, 이러한 현상의 원인은 이후 결론 부분에서 살펴보겠다.

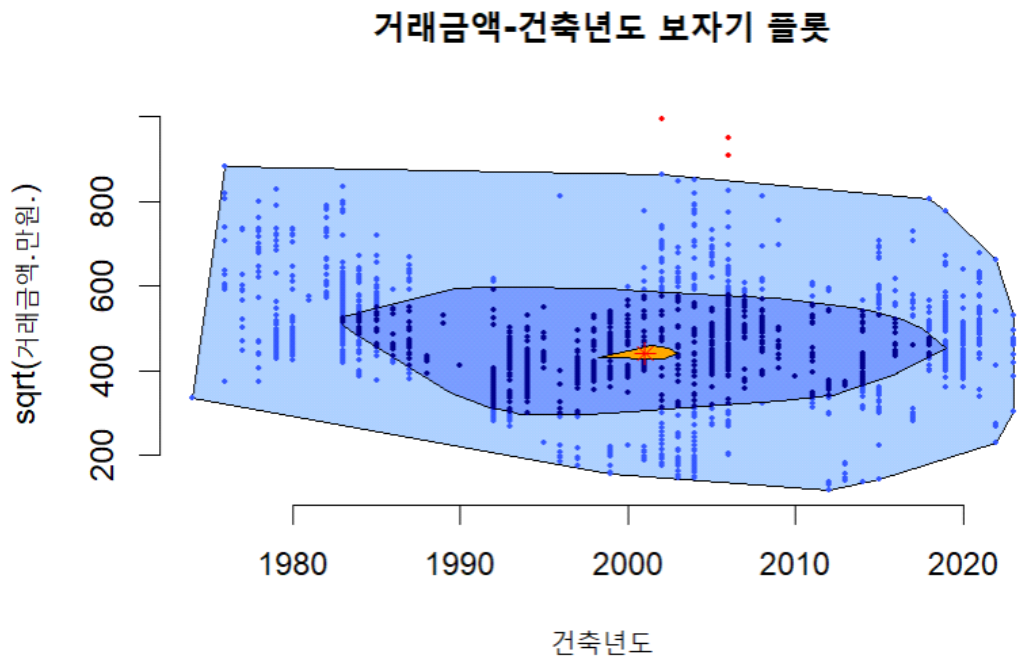
3-1-3. 보자기플롯 (bagplot)

(1) 거래금액(만원)과 전용면적(m²)



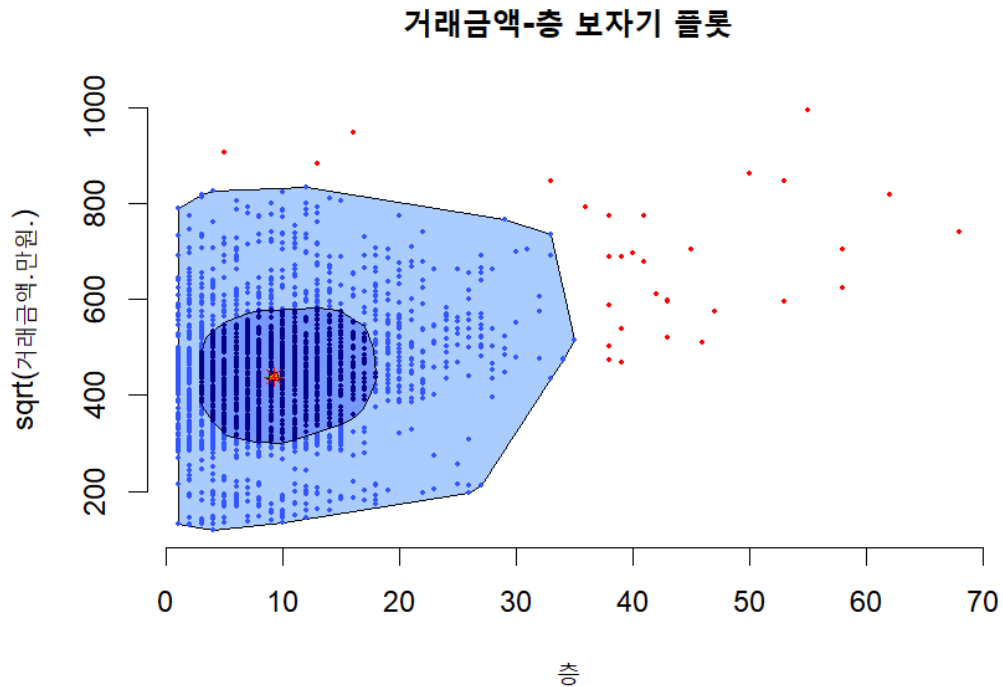
전체적으로, 우상향하는 추세를 확인할 수 있다. x=전용면적, y=거래금액 으로, 전용면적이 커질수록 거래금액이 커지는 경향을 확인 가능하다. 또한 전체적으로 이상점이 많이 나타나는 것을 알 수 있지만 이는 이상점으로 간주하고 삭제하기는 어려워 보인다. 앞서 언급했듯, 아파트의 가격결은 요인은 매우 다양한 요인을 통하여 신중하게 결정되기 때문이다.

(2) 거래금액(만원)과 건축년도



분포를 보았을 때 큰 특징을 찾을 수 없다.

(3) 거래금액(만원)과 층

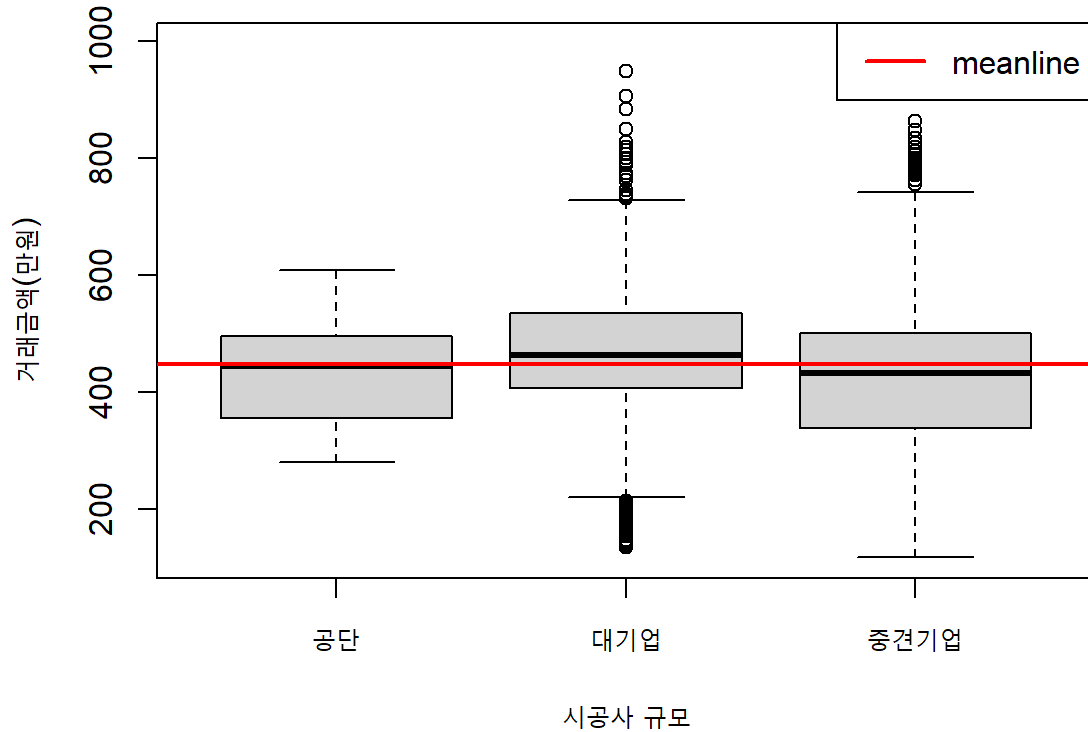


외부점을 포함하여 분포를 보았을 때 미세하게 우상향하는 특징을 찾을 수 있다. 따라서 층 수가 올라가면 아파트의 거래금액이 조금은 올라간다고 예상할 수 있다. 그러나 자료의 분산이 매우 크고 고층의 자료가 적어 우상향의 추세가 유효하리라 (신뢰할 수 있다고) 확신하기는 어려워 보인다.

3-1-4. 상자그림

(1) 시공사 규모와 거래금액(만원)

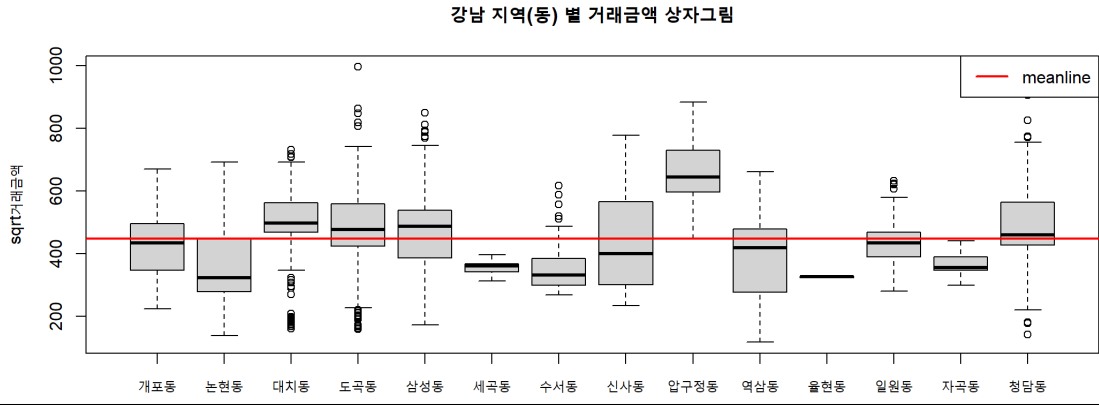
강남 시공사별 거래금액 상자그림



공단에서 시공한 아파트의 상자그림은 꼬리가 짧다. 따라서 공단 시공한 아파트는 비교적 균일한 가격에 공급되리라 기대할 수 있다. 공단은 정부의 정책사업을 위수탁 받아 자사의 수익보다 국가 행정의 효율성을 목표하는 특수법인이다. 따라서 일정한 가격에 아파트를 시공했다고 짐작된다.

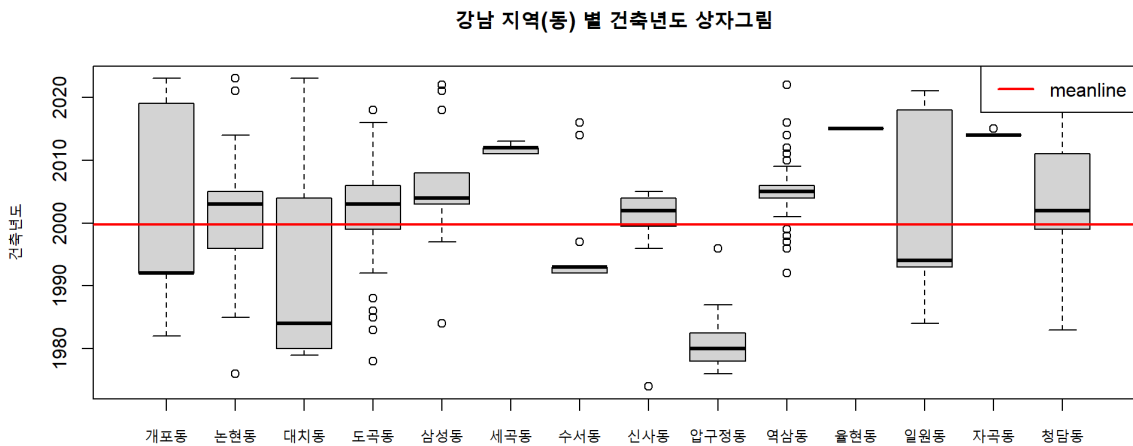
반면, 대기업과 중견기업의 상자그림은 꼬리가 길다. 따라서 이들이 시공한 아파트의 가격은 편차가 크리라 예상할 수 있다. 대기업과 중견기업의 경우, 이상점도 많다. 이들은 이익을 중요시하는 사기업이므로 가능한 고가의 아파트를 짓고자 할 것이다. 따라서 거래금액이 높은 이상점들의 존재는 쉬이 예측할 수 있다. 그러나 낮은 가격에 거래된 이상점들은 앞선 논리로 설명하기 어렵다. 이들의 경우, 다양한 요인에 의해 아파트의 가격이 대기업의 예상보다 낮게 형성되어 사업이 실패했으리라 의심할 수 있다. 또는, 대기업의 사업이 성공적이었다면, 최초 분양비는 비쌌으나 거래금액이 차차 떨어져 지금과 같은 낮은 거래금액이 형성되었을 가능성도 있다.

(2) 지역(동)과 거래금액(만원)



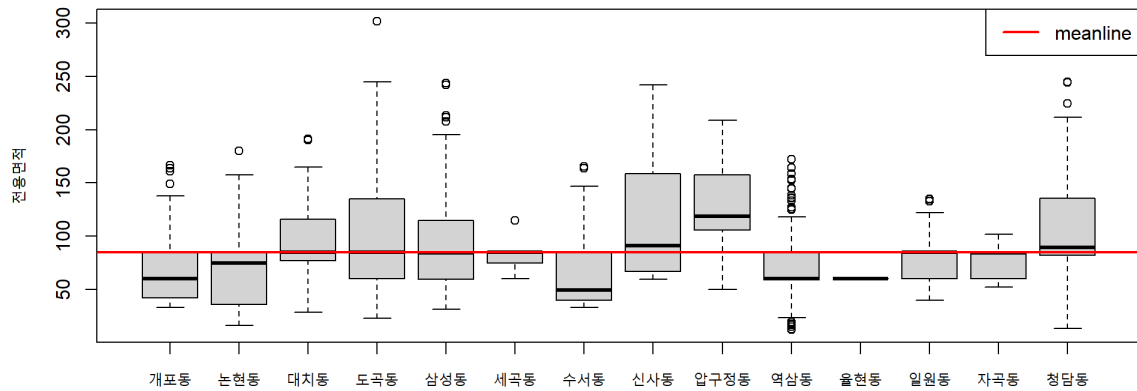
압구정동의 아파트 거래금액(만원)이 높게 나타나고, 논현동, 세곡동, 수서동, 자곡동이 비교적 낮게 나타나는 것을 확인할 수 있다. 또한, 역삼동의 경우 분산이 상대적으로 매우 큰 것을 알 수 있고, 세곡동의 경우 비교적 거래금액의 분산이 상대적으로 작은 것을 알 수 있다. 이어서 압구정동의 아파트 거래금액이 가장 높게 나타나는 이유를 앞서 파악한 이변량 자료분석법들과 연계하여 해석해보겠다.

아래의 건축년도의 상자그림을 보면, 압구정동에서 거래된 아파트들이 가장 오래된 아파트임을 것을 볼 수 있다. 등고선 그림에서 파악한 강남지역 아파트의 연도 별 거래금액 군집을 상기해보면, 건축년도가 가장 오래된 80년도 아파트 군집의 거래금액이 가장 높았다. 아래 상자그림에서 보듯, 압구정동은 80년도에 지어진 아파트가 많으며, 다른 지역들에 비해 건축년도가 가장 오래되었다. 이는 압구정동의 거래금액이 높은 현상의 이유 중 하나일 것이다.

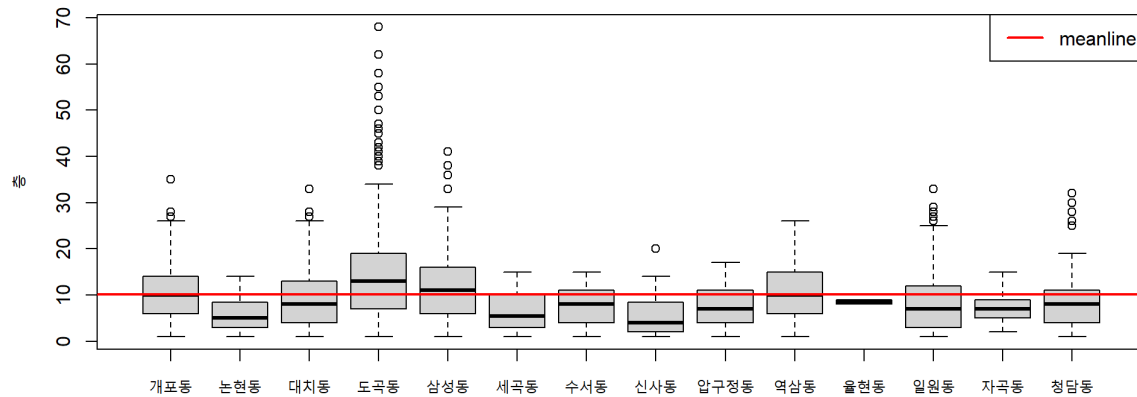


이어서 앞서 살펴본 보자기플롯과 로웨스 곡선들을 참고하면, 압구정동 아파트의 면적이 넓고 층이 높다고 의심할 수 있다. 해당 변수로 상자그림을 그려 분포를 확인해 보겠다.

강남 지역(동) 별 전용면적 상자그림



강남 지역(동) 별 층 상자그림



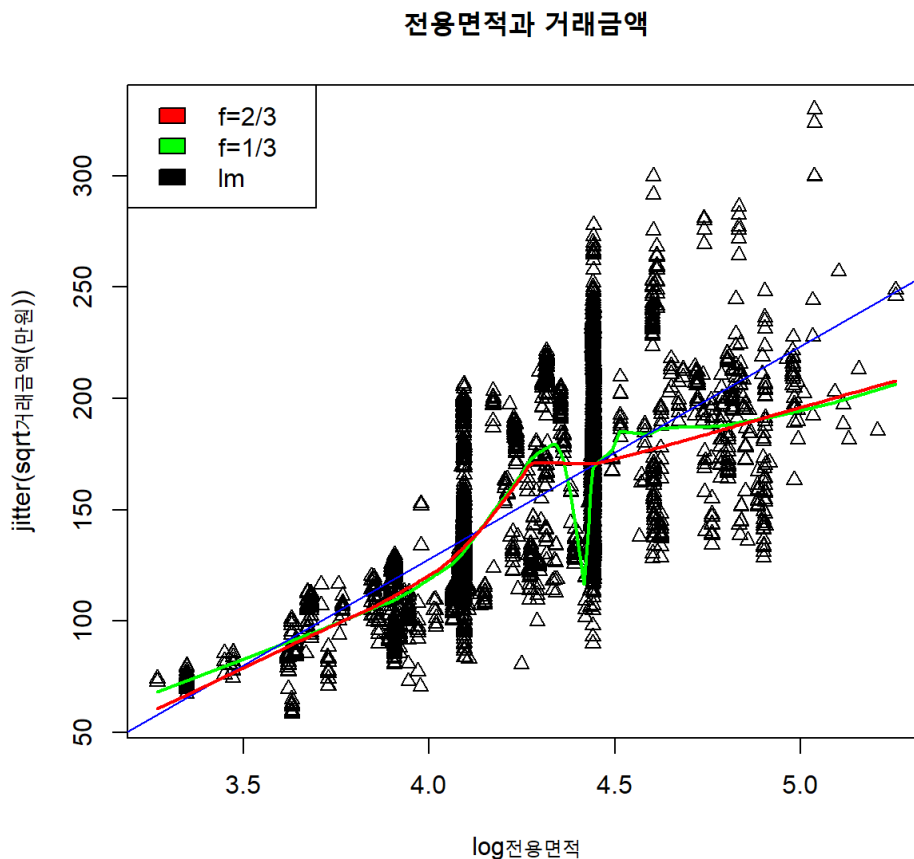
위의 상자그림에서 압구정동의 위치를 보면, 압구정동은 층수는 평균보다 낮지만, 면적은 가장 넓음을 알 수 있다. 앞서 보자기 그림으로 파악한 추세선을 기억해 보면, 층수의 기울기는 작지만, 면적의 기울기는 컸다. 그러므로 압구정동의 넓은 면적이 낮은 층수보다 거래금액에 큰 영향을 주리라 짐작할 수 있다.

3. 춘천 아파트 이변량 자료분석

3-2. 춘천 아파트 실거래가(매매) 자료 분석

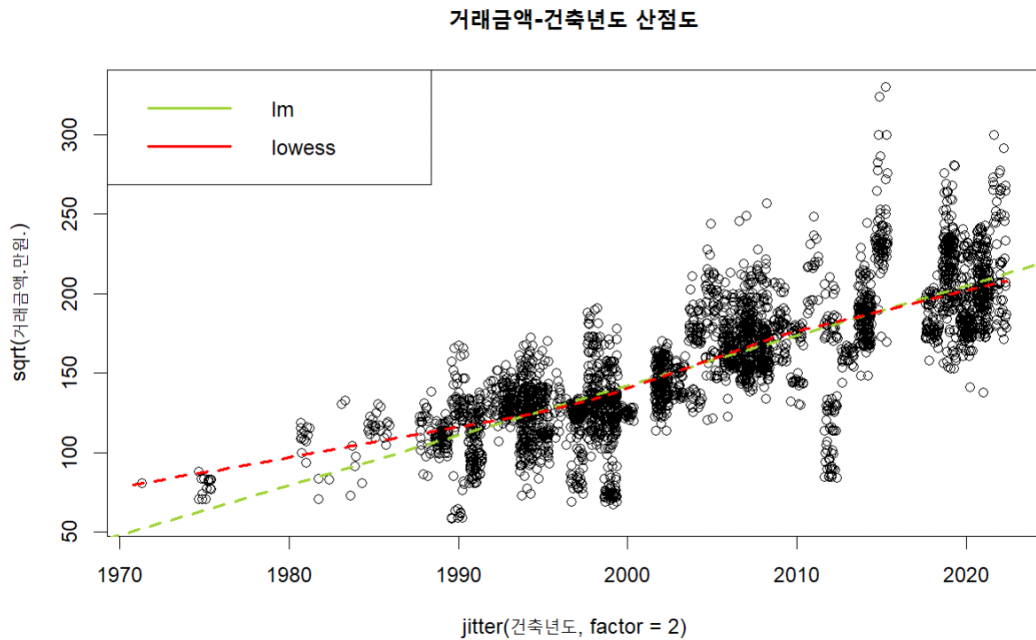
3-2-1. 산점도

(1) 전용면적(m²)과 거래금액(만원)



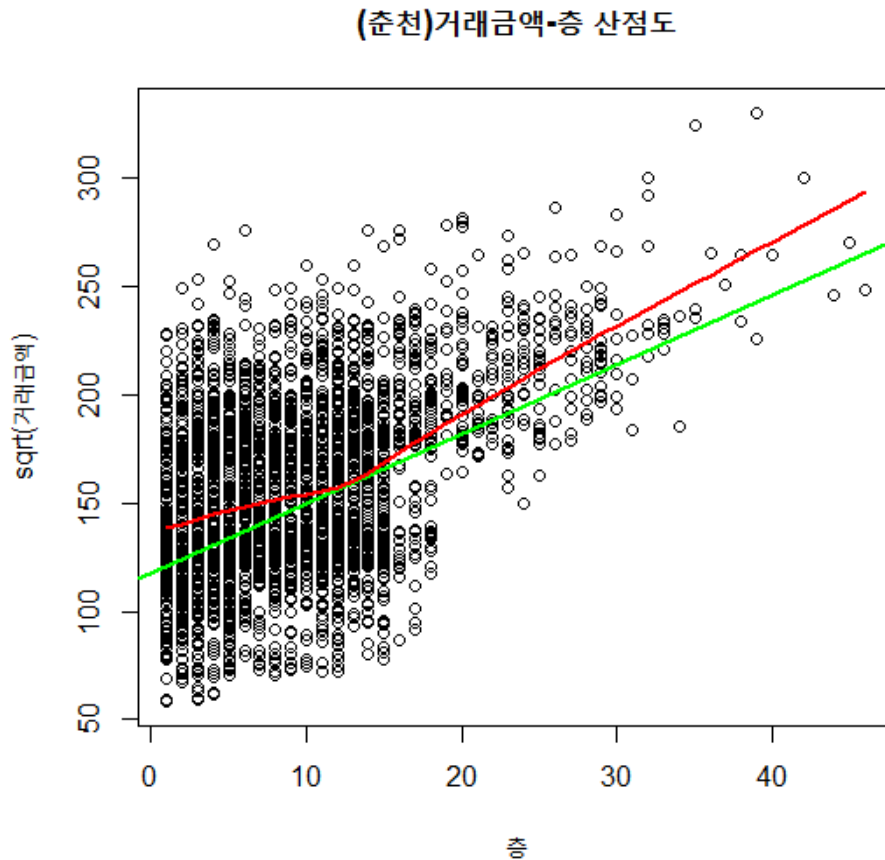
위쪽은 거래금액-전용면적 간 산점도를 그리고 그 위에 최소제곱 회귀직선(푸른 색), 평활너비를 2/3로 설정한 로웨스 곡선(붉은색), 평활너비 1/3인 로웨스 곡선(초록색)을 그린 그림이다. 두 변수를 대칭분포로 만들어주기 위하여 거래금액은 제곱근변환, 전용면적은 로그변환을 했다. 이후 자연스러운 분포를 위하여 거래금액에 jitter 처리를 해주었다. 과대적합된 녹색 로웨스 곡선을 제외하면, 공통적으로 우상향하는 비례관계가 보인다. 따라서 아파트의 전용면적이 커질수록 아파트의 거래금액이 커진다고 생각할 수 있다.

(2) 건축년도와 거래금액(만원)



위는 춘천지역 아파트의 제곱근변환 거래금액과 건축년도 산점도에 추세선을 그은 그림이다. 자연스러운 분포를 위하여 건축년도에 jitter 처리를 해주었다. 최소제곱 회귀직선과 로웨스 곡선($f=2/3$)을 비교해보면, 1995년 이전의 로웨스 곡선은 최소제곱 회귀직선보다 기울기가 완만하다. 1995년 이후 로웨스 곡선의 기울기가 최소제곱 회귀직선과 맞추어짐을 알 수 있다. 따라서 노후화된 아파트의 실제 감가가 최소제곱 회귀직선이 예측한 회귀계수보다 적을 것이라 예상할 수 있다.

(3) 거래금액(만원)과 층

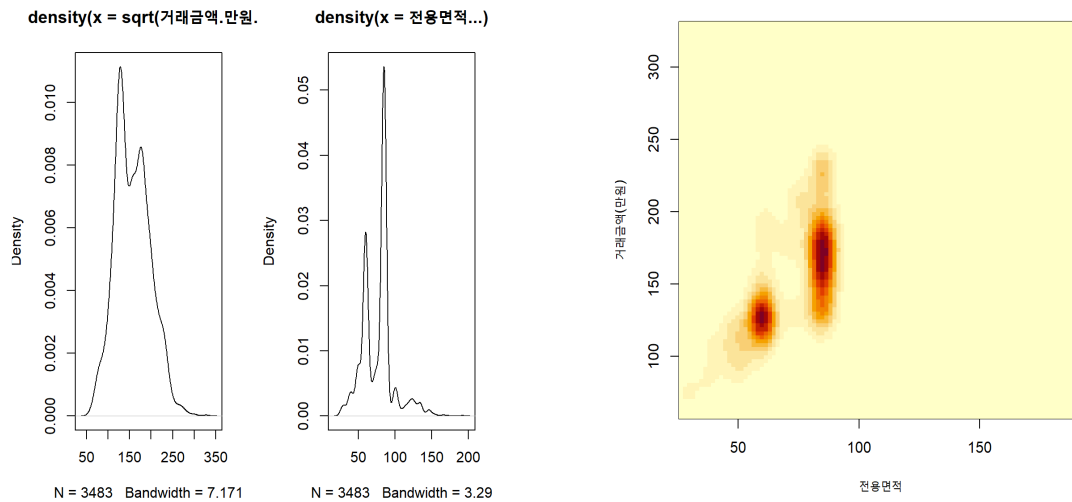


위의 그림은 춘천지역 아파트의 제곱근변환 거래금액과 층수의 산점도에 추세선을 그은 그림이다. 연두색은 최소제곱 회귀직선, 붉은색은 평활너비가 2/3인 로웬스 곡선이다. 공통적으로 고층으로 갈수록 가격이 비싸지는 경향이 있다. 대략 12층 정도에서부터 기울기가 가팔라진다. 12층 정도부터 가격이 비싸진다고 볼 수 있다.

앞서 탐색한 강남의 자료보다는 고층의 점이 많아 보이지만, x축 범위가 다르므로 확신하기 어렵다. 다만 강남과 마찬가지로, 고층으로 갈수록 아파트의 거래 빈도가 낮아지는 모습이 있다. 그러므로 전체적인 자료의 분산이 크고 고층 아파트의 거래 표본이 적다는 문제도 같다.

3-2-2. 커널 밀도추정과 등고선 그림

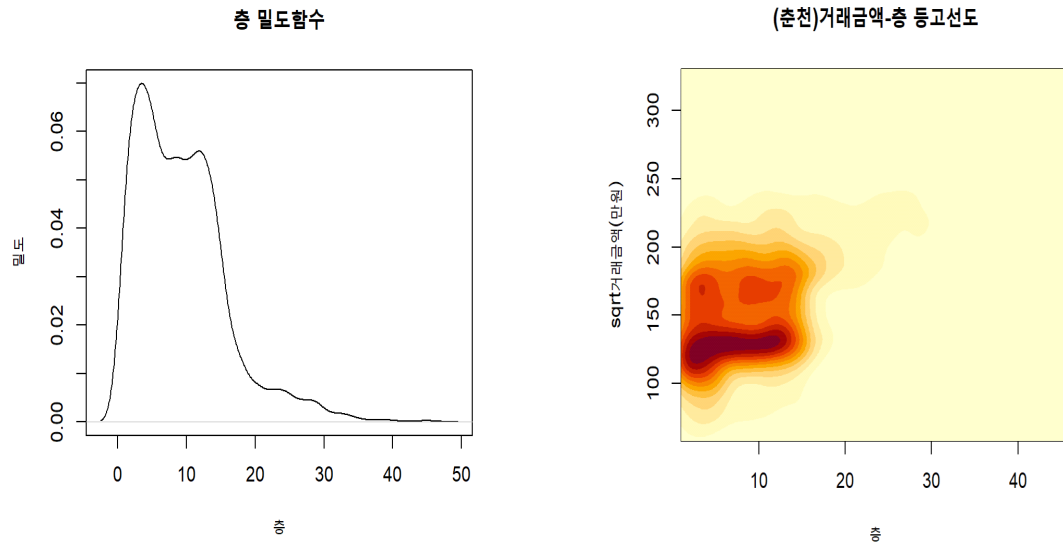
(1) 거래금액(만원)과 전용면적(m²)



왼편의 그림은 거래금액의 밀도함수이고, 중앙 그림은 전용면적의 밀도함수이다. 거래금액을 대칭분포로 맞추고 금액의 파악을 쉽게 하도록 제곱근 변환한 값을 사용했다. 거래금액의 밀도함수는 큰 봉우리 하나, 작은 봉우리 한 개가 있어, 전체적으로는 한 개의 군집이 있으리라 예상된다. 한편 전용면적은 큰 봉우리 두 개, 미세한 봉우리 여러 개가 있어 두 개의 군집이 있으리라 예상된다. 가장 높은 봉우리는 각각 거래금액 1.5억, 면적 90m²에 자리 잡고 있다.

오른편의 그림은 거래금액(만원)의 제곱근 변환한 값들과 전용면적 간의 등고선 그림이다. 두 개의 군집을 관찰할 수 있다. 두 군집은 전용면적이 커질수록 거래금액이 비싸지는 정비례 관계에 있다. 색이 같으므로 밀도 차이는 없어 보이지만, 크기 차이가 존재한다. 커널 밀도함수에서 거래금액 1.5억 근방에 밀도가 가장 높음을 확인했으나 실제 군집은 거래금액 3.5억 근방의 군집이 더 크게 나타났다. 이는 거래금액 봉우리의 폭 차이로 인해 발생한 듯 하다. 전용면적 대략 60m², 90m²부분이 가장 짙은 색을 띠므로 가장 자료가 많이 밀집되어 있다는 것을 알 수 있다.

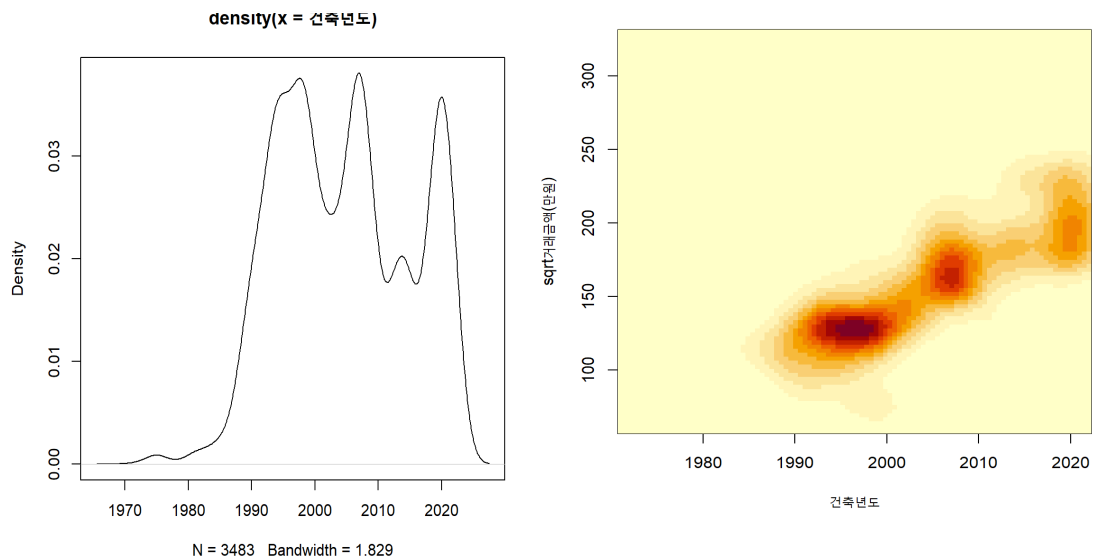
(2) 거래금액(만원)과 층



왼편은 층에 대한 커널 밀도함수이다. 거래된 매물이 대략 1-15층 사이에 가장 많음을 확인할 수 있다. 전체적으로 보았을 때, 봉우리는 매우 큰 봉우리 하나가 존재한다. 따라서 한 개 군집이 있으리라 기대할 수 있다. 오른쪽은 제공된 변환한 거래금액과 층의 등고선 그림이다. 한 개의 군집이 존재한다. 다만 층 봉우리의 폭이 크기 때문인지, 자료가 밀집한 구역이 층수 1~15층에 걸쳐 넓게 나타난다. 밀도가 높은 구역의 거래금액은 마찬가지로 약 1.5억으로 매우 저렴하다.

춘천 아파트 거래금액은 강남 아파트의 거래금액보다 저렴하다. 따라서 춘천 거래금액~층 등고선도와 강남 거래금액~층 등고선도를 비교하면 춘천 등고선도는 납작하게 눌러서 나올 것이다. 이러한 거래금액의 차이를 고려하면, 해당 군집의 거래금액 분산은 매우 작고, 층의 분산은 매우 넓은 군집이라고 짐작할 수 있다. 즉, 현재 등고선 이미지에서 춘천 군집은 동글게 나오지만, 강남과 비교하면 상대적으로 가로로 긴 형태일 것이다.

(3) 거래금액(만원)과 건축년도

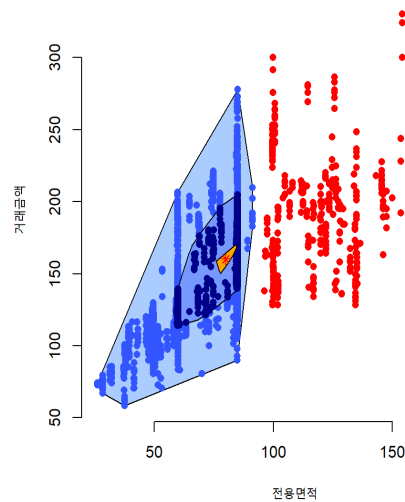


왼편은 춘천 아파트의 건축년도 커널 밀도함수이다. 1995년, 2005년, 2020년에 세 개의 봉우리가 있으니 세 개의 군집이 있으리라 예측할 수 있다. 오른쪽은 제곱근 변환한 거래금액과 건축년도의 등고선 그림이다. 마찬가지로 3개의 군집이 보인다. 왼편 아래의 군집이 가장 밀도가 높고, 중앙의 군집이 다음으로 밀도가 높으며, 맨 오른쪽 위의 군집이 가장 밀도가 낮다. 각 군집은 (1995년, 1.7억), (2007년, 2.5억), (2020년, 4억)에 위치한다.

강남의 경우와 다르게 춘천의 세 군집은 건축년도가 최신일수록 거래금액이 증가하는 정비례 관계를 보인다. 이는 춘천의 경우 아파트의 노후화가 가격 결정에 부정적인 영향을 미친다는 뜻이다. 따라서 춘천의 아파트 가격을 결정하는 회귀모형 등에서는 건축년도 변수가 유의미한 영향을 주리라 기대된다.

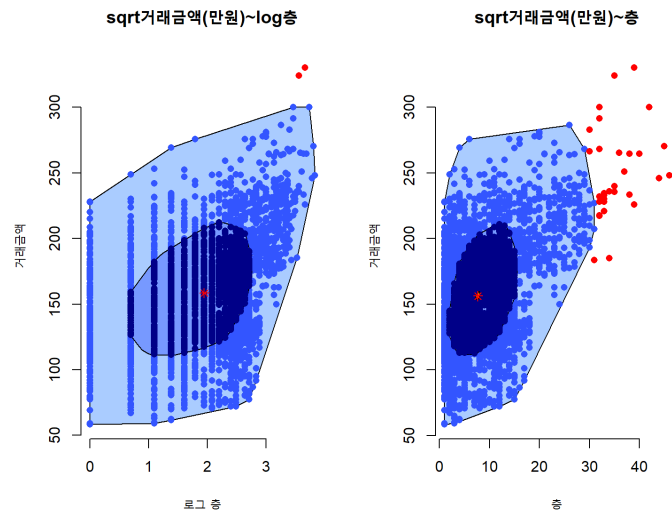
3-2-3. 보자기 플롯

(1) 거래금액(만원)과 전용면적(m²)



jitter sqrt 거래금액과 전용면적의 보자기플롯. 중앙점은 전용면적 80m², 거래금액 2.2억에 위치한다. 전용면적 100 이상 부분은 전부 외부점이 되었다. 이 점들을 이상치로 단언하고 삭제하기에는 것은 위험해 보인다. 왜냐하면 아파트의 거래금액은 여러 요인과 신중한 의사결정을 통해 결정되기 때문이다. 따라서 이상점보다는 영향점일 가능성이 커 보인다. 또한 앞서 선형회귀 결과 전용면적이 클수록 아파트 가격이 커지는 경향성을 확인했다. 고가의 아파트 거래량은 저가의 아파트 거래량보다 적으므로, 상대적으로 밀도가 낮은 전용면적이 넓고 거래금액이 높은 아파트들이 외부점으로 처리되었을 수 있다.

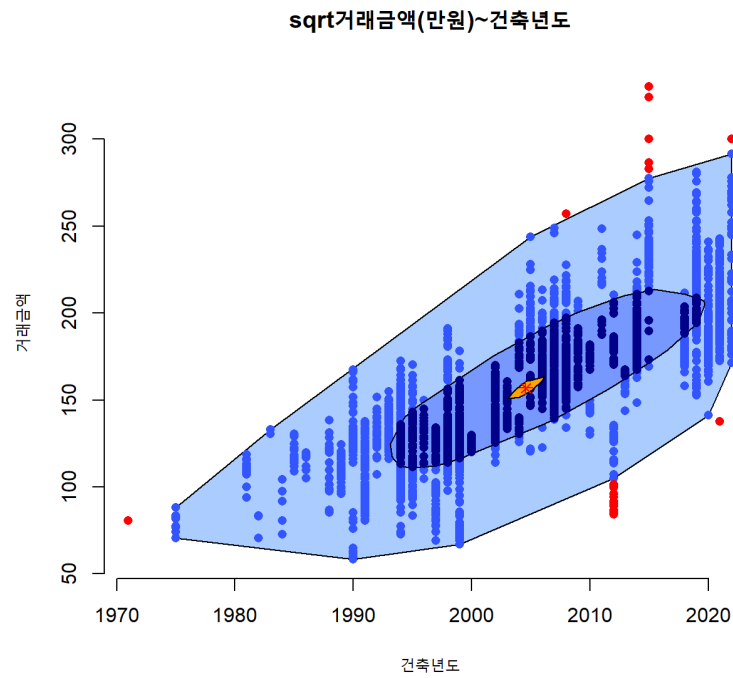
(2) 거래금액(만원)과 층



제곱근 변환 후 jitter를 적용한 거래금액과 층의 보자기플롯. 앞서 층의 히스토그램이 왼쪽으로 치우친 분포임을 확인했다. 로그 변환으로 대칭분포를 만든 결과 확실히 이상치가 줄어든 것이 보인다. 다만 층의 범위가 1~60정도로 작으므로 로그 변환을 했을 때 왼쪽의 분포가 매우 부자연스럽게 나온다는 문제가 있다.

앞서 강남의 경우는 층과 거래금액의 보자기 플롯이 기울기가 거의 없게 나와 양의 상관성을 확신하기 어려웠다. 반면 춘천의 보자기 플롯은 확실하게 고층으로 갈수록 거래금액이 낮은 아파트들이 줄어드는 경향을 보여준다.

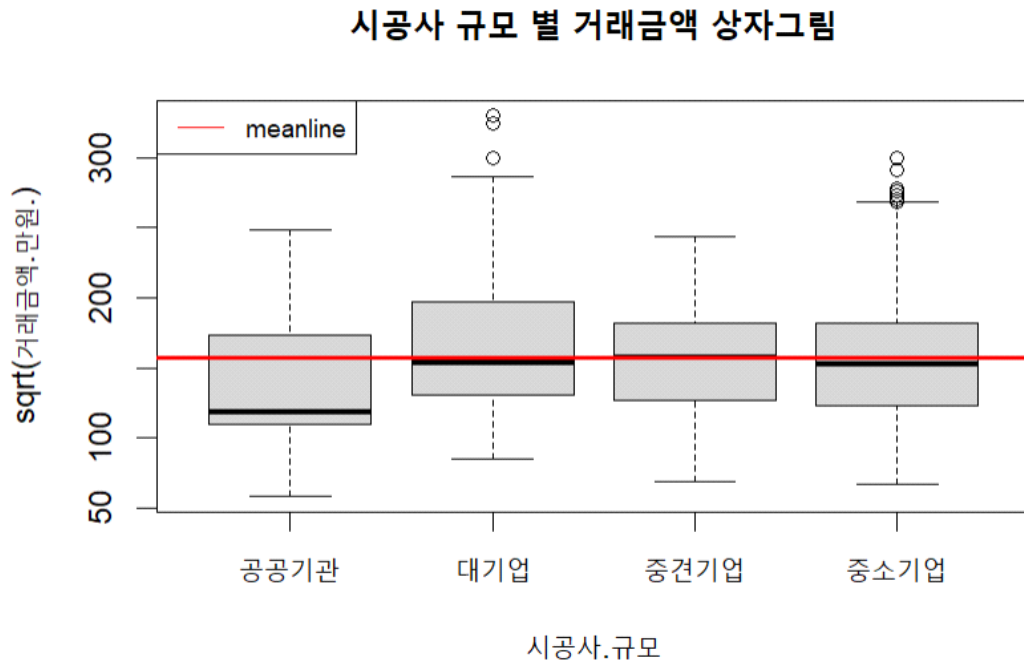
(3) 거래금액(만원)과 건축년도



춘천의 경우 건축년도와 거래금액의 선형성이 있다.

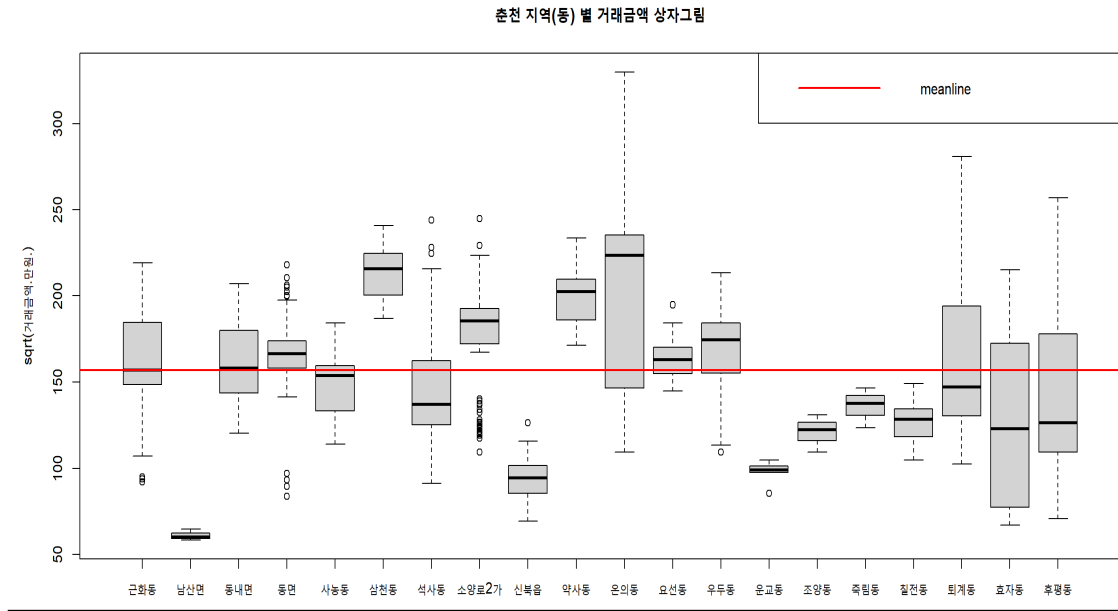
3-1-4. 상자그림

(1) 시공사 규모와 거래금액(만원)

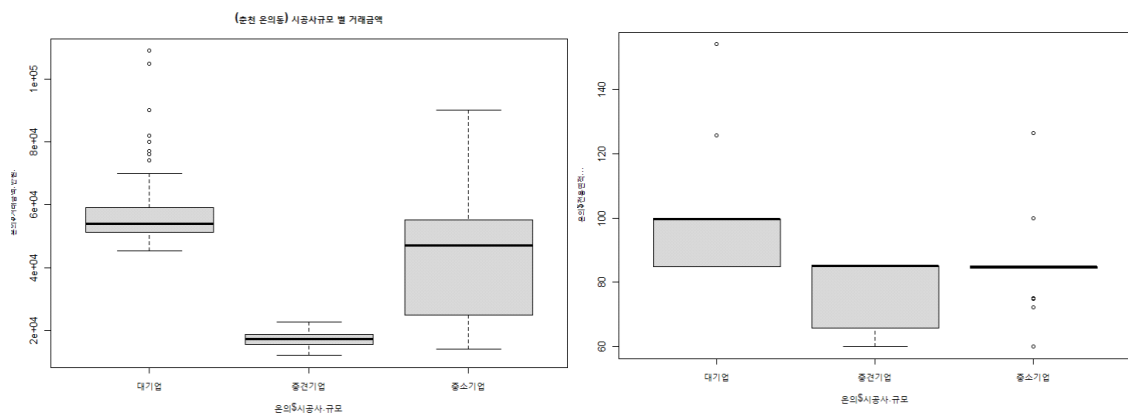


공공기관에서 시공한 아파트는 비교적 평균 거래금액(만원)이 낮다. 공공기관 범주는 수서주택조합과 주택공단으로 이루어져 있다. 이들은 국민들의 주거안정을 위해 아파트를 건축하므로 때문에 사기업의 아파트에 비해 분양가가 저렴하게 나오는 특징이 있다. 낮은 분양가 때문에 거래금액도 낮게 형성되었다고 짐작할 수 있다. 혹은 공공기관은 사기업보다 기술경쟁력이 부족하여 공공기관 아파트의 안전성, 디자인 등 품질이 떨어지므로 수요가 적어지고 낮은 거래금액이 형성되었다고 의심할 수 있다.

(2) 지역(동)과 거래금액(만원)



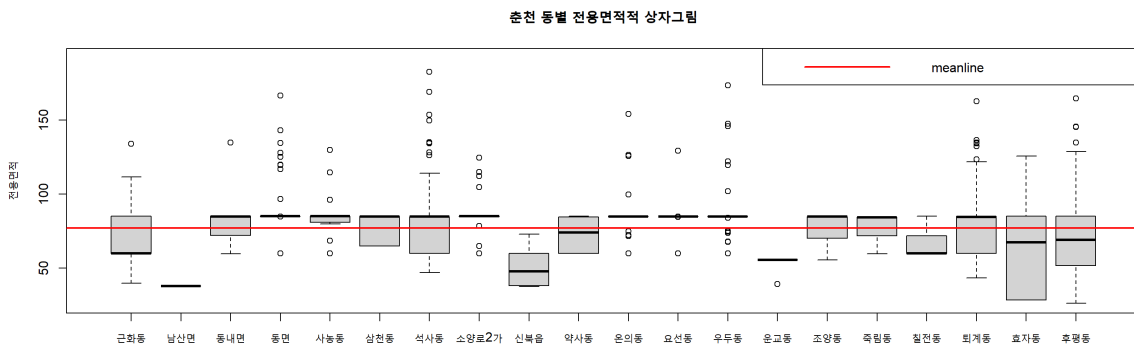
삼천동, 소양로2가, 약사동의 아파트 거래금액(만원)이 높게 나타나고, 남산면, 신북읍, 운교동의 아파트가 비교적 낮게 나타나는 것을 확인할 수 있다. 그중 온의동의 아파트 거래금액의 중앙값이 가장 비싸고, 오른쪽 꼬리가 길며, 그 분산도 크게 나타난다. 이러한 분포적 특징의 원인을 앞서 파악한 이변량 자료분석법들과 연계하여 해석해 보겠다.



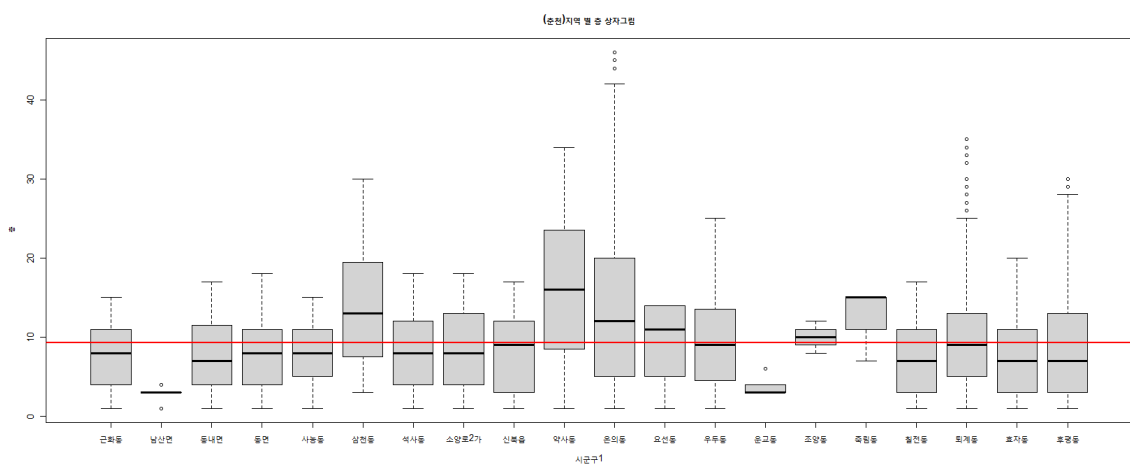
원편은 온의동에서 거래된 아파트들의 시공사 규모 별 거래 금액의 상자 그림이다. 각각의 빈도는 대기업 72, 중견기업 35, 중소기업 136으로 중소기업의 시공수가 가장 많고, 가장 다양한 가격으로 거래되었음을 확인할 수 있다. 반면, 공공기관의 시공수는 0건이었다. 거래금액이 가장 낮은 공공기관 시공 아파트가 없으므로, 온의

동 아파트의 거래금액이 다른 지역보다 높으리라 기대할 수 있다. 또한, 대기업이 시공한 아파트의 거래금액 이상치들도 온의동 전체 아파트의 가격을 올리는 데 영향을 미쳤을 것이다.

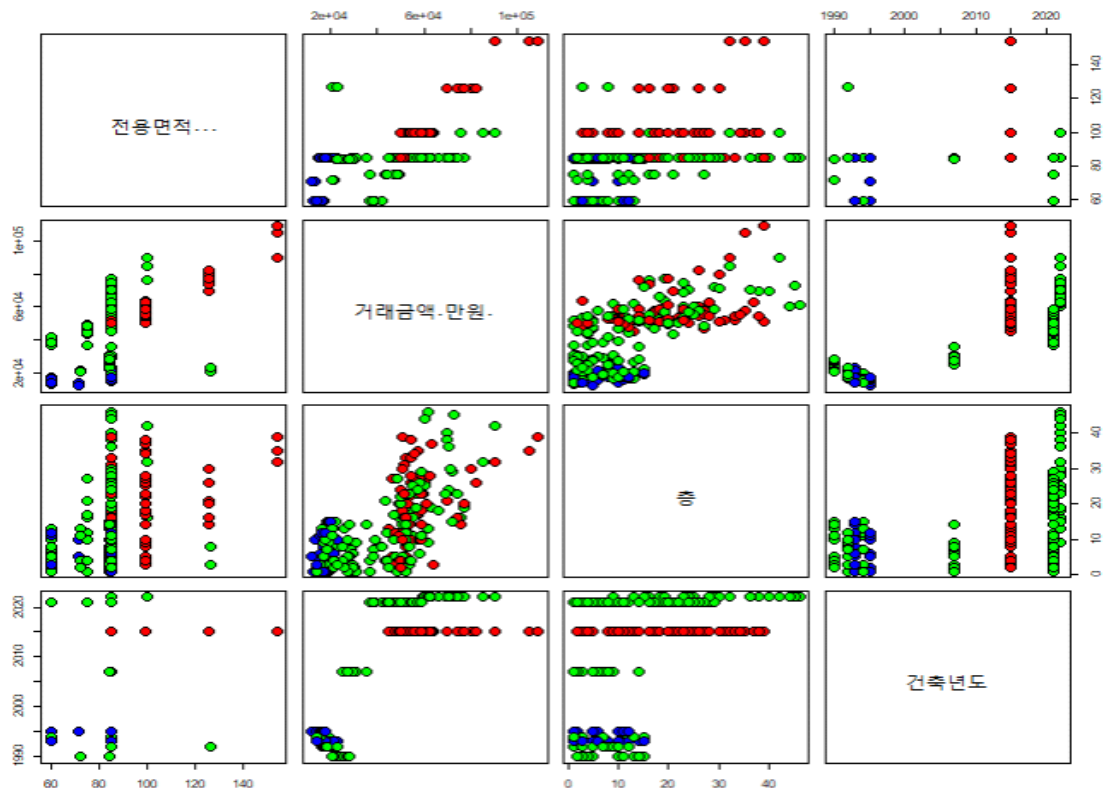
오른편은 온의동 아파트 대상 시공사 규모별 전용면적의 상자 그림이다. 표본이 적어 분포가 정규성을 따지는 못하나, 중소기업이 시공한 아파트의 면적이 거의 90 m²에 몰려 있음을 확인할 수 있었다. 대기업 시공의 아파트 중 거래금액과 전용면적이 극단적으로 크게 나타난 점이 있어 이에 관한 추가적인 조사의 필요성이 있다.



온의동과 다른 동들의 전용면적을 비교하면 위와 같다. 온의동 아파트의 전용면적 중앙값은 다른 동에 비해 높음을 확인했다. 앞서 확인한 면적과 가격의 정비례관계에 미루어 볼 때, 이 또한 온의동 아파트의 높은 거래금액의 원인이라 예상할 수 있다. 하지만, 시공사와 전용면적의 자료로는 온의동 아파트의 높은 가격의 이유를 짐작할 수 있을 뿐, 온의동 아파트 거래금액의 큰 분산을 설명하기는 어렵다. 오히려 온의동의 전용면적의 분산은 매우 작으므로, 해당 변수만을 보았을 때, 온의동 아파트의 거래금액 분산도 낮으리라고 짐작이 된다.



위의 상자 그림은 춘천 지역 동별 아파트의 층에 대한 상자 그림이다. 온의동이 층의 분포가 가장 넓은 것을 알 수 있다. 따라서 온의동이 거래금액의 분산이 큰 이유는 온의동 아파트 층의 분산에 영향을 받았으리라고 짐작할 수 있다.



온의동 자료의 산점도 행렬로 설명을 보충하겠다. 위의 산점도 행렬에서 초록 점은 중소기업, 푸른 점은 중견기업, 붉은 점은 대기업의 자료점이다. 이중 거래금액과 층의 산점도를 보면 층이 높아질수록 거래금액이 비싸지는 선형성을 관찰할 수 있다. 따라서 온의동의 아파트 거래금액과 층이 양의 상관관계가 있음을 확인할 수 있다.

정리하자면 다음과 같다. 온의동 아파트가 높게 거래된 세 가지 원인 추측이 가능하다. 첫째, 온의동에서 공공기업이 시공한 아파트가 거래되지 않았기 때문이다. 둘째, 온의동에서 거래된 아파트들의 면적이 평균보다 컸기 때문이다. 셋째, 온의동에서 거래된 고층 아파트들이 많았기 때문이다. 한편, 온의동 아파트의 거래금액 분산이 큰 이유는 다음과 같이 추측할 수 있다. 온의동 아파트가 가장 고층 아파트가 많이 거래되었는데, 산점도 행렬에서 보듯 층과 거래금액 사이에는 양의 상관성이 존재하기 때문이다. 고층 아파트가 많으므로 고가에 거래된 아파트도 많이 있을 것이고, 고가 아파트의 빈도가 높아지므로 분산이 크고 오른쪽 꼬리가 긴 분포가 나타났으리라는 논리이다.

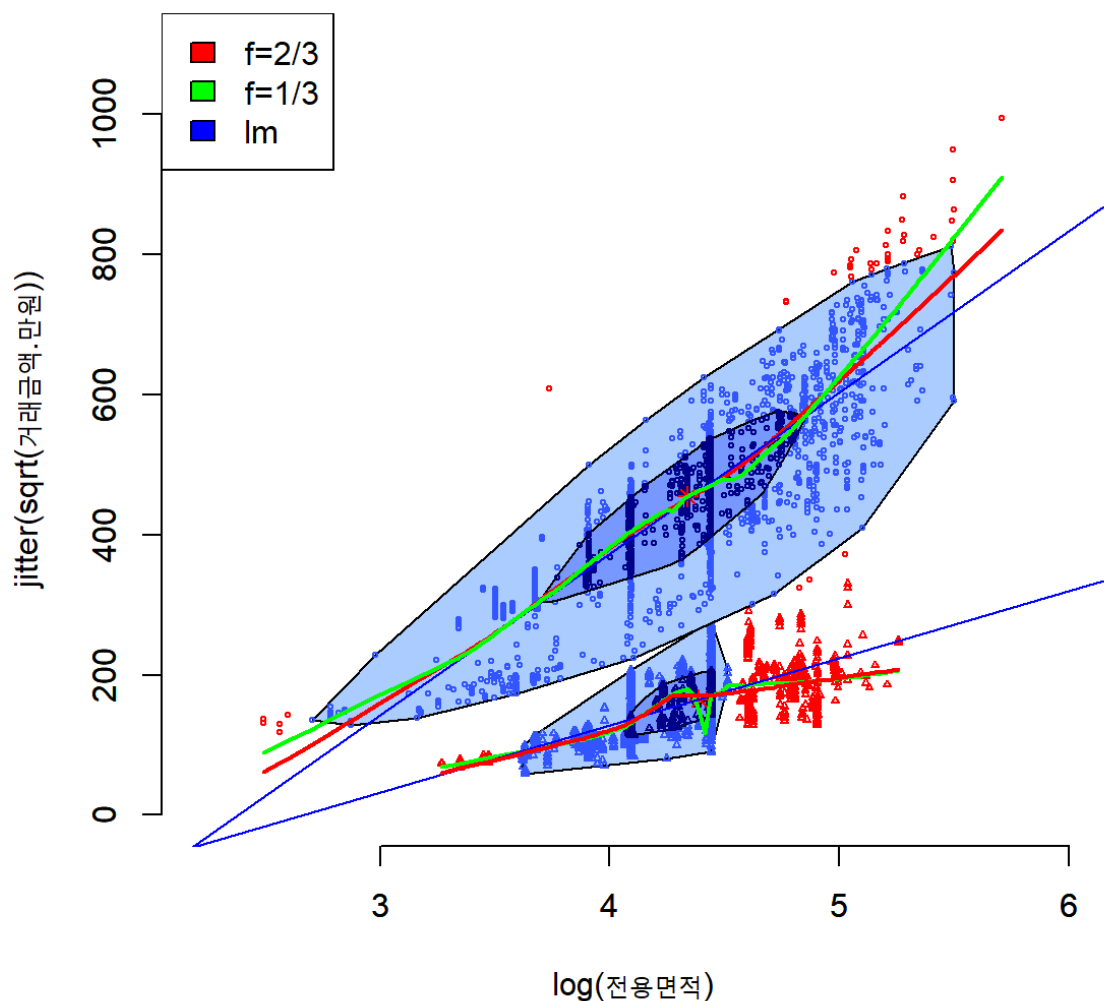
6. 결론

1. 춘천과 강남의 데이터 비교

1) 전용면적에 따른 거래금액 비교

다음은 춘천과 강남의 전용면적, 거래금액의 보자기 플롯이다. 자료의 형태를 정규분포에 유사하게 맞추기 위하여 전용면적은 로그 변환하고 거래금액은 제곱근변환했다. 거래금액은 만 단위로 요약된 수치형 변수이므로 변수간의 자연스러운 관계를 보기 위하여 jitter 처리를 추가했다. 푸른 실선은 최소제곱법을 활용한 단순선형회귀직선이다. 붉은 실선은 평할너비를 2/3으로 한 로웬스 회귀곡선이고, 초록 실선은 평할너비를 1/3으로 한 로웬스 회귀곡선이다. 원형 점들은 강남의 자료이고, 삼각형 점들은 춘천의 자료이다.

거래금액-전용면적 보자기 플롯
(춘천, 강남 비교)



먼저 보자기 그림을 보자. 두 자료의 중간점의 전용면적 값이 $80m^2$ 정도인 것은 유사하다.

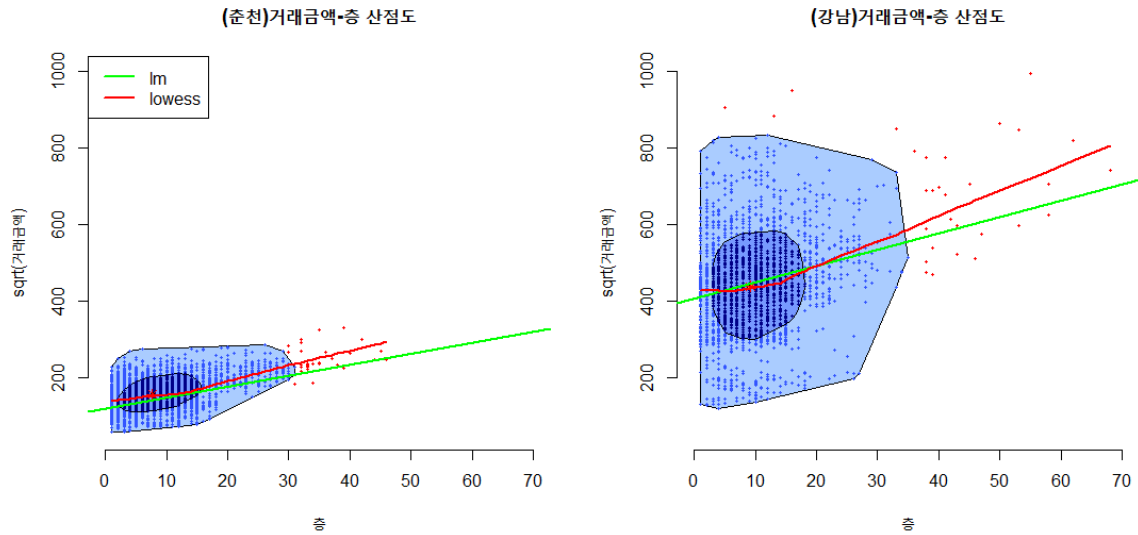
그러나 강남 아파트는 전용면적이 2~6까지 넓은 분포를 보이지만, 춘천의 경우는 3.2~5.5에 면적이 몰려있다. 강남의 중간점의 경우 거래금액 값이 20억 정도지만, 춘천의 경우는 4억도 되지 않는다. 점들의 분포도 강남의 경우 최대 1000까지로 거래금액의 분포가 다양하고 비싸지만, 춘천의 경우는 100에서 400 사이에 몰려있다. 강남의 경우 외부점이 적으나, 춘천의 경우 전용면적이 큰 자료점들이 대거 외부점으로 분류되었다. 하지만 아파트의 구매는 신중한 의사결정 과정을 통하여 이루어지므로, 해당 자료들을 이상치로 의심할 뿐, 제거하기는 어려워 보인다.

자료의 분포가 선형성을 띠므로, 유효한 회귀적합을 기대할 수 있다. 단순선형회귀직선은 전용면적과 거래금액간에 양의 상관성이 있음을 시사한다. 춘천과 강남 아파트는 전용면적이 커질수록 비싸게 거래된다. 그러나, 강남의 회귀직선의 기울기가 춘천의 회귀직선의 기울기보다 크다. 즉, 전용면적이 동일하게 증가할 때, 강남의 아파트 거래금액 증가량이 춘천의 아파트 거래금액 증가량보다 크다. 로웬스 곡선은 전용면적의 증가에 따른 거래금액의 변화를 보다 상세하게 보여준다. 강남의 경우, 면적이 좁으면 거래금액이 완만하게 상승하나, 면적이 커지면 거래금액이 가파르게 상승한다. 이와 반대로 춘천의 경우, 면적이 좁으면 거래금액이 가파르게 상승하나, 면적이 커지면 거래금액이 완만하게 상승한다.

금액이 수요와 공급에 따라 결정된다는 통념에 미루어 볼 때, 강남의 경우 작은 면적의 아파트의 수요가 적고 공급이 많으며, 넓은 면적의 아파트의 수요가 많고 공급이 적다고 짐작할 수 있다. 반면, 춘천의 경우는 작은 면적의 아파트의 수요가 많고, 넓은 면적의 아파트의 수요가 적다고 짐작할 수 있다. 실제로, 강남의 면적은 39km²인 반면에 춘천의 면적은 1116km²이다. 강남은 면적이 작으므로 넓은 면적의 아파트의 공급이 적고, 춘천은 면적이 넓으므로 넓은 면적의 아파트 공급이 충분하리라 짐작할 수 있다.

위의 자료에는 면적이 넓어질수록 거래금액의 분산이 커지는 특징이 있다. 지역별로 전용면적이 3.5일 때 거래금액의 분산과 전용면적이 5일 때 거래금액의 분산을 보면 직관적으로 이해할 수 있다. 따라서 면적이 좁으면 회귀모형의 설명력이 높지만, 면적이 크면 회귀모형의 설명력이 낮으리라 예상된다. 이는 아파트의 가격이 비싸질수록, 면적과 독립적인 미지의 요인들이 거래금액에 큰 영향력을 행사함을 시사한다. 즉, 사람들은 고가의 아파트를 구매할수록 면적 외의 다양한 조건을 요구한다. 일조권, 조망권, 인테리어 여부, 시공업체, 교통, 학군, 소음, 인프라 등이 기타 조건이라 짐작이 간다. 그러나 연구 자료는 유의미해 보이는 변수로 건축년도, 시공사, 층 정보만을 포함하므로 그 자료들에 관한 탐색을 이어가겠다.

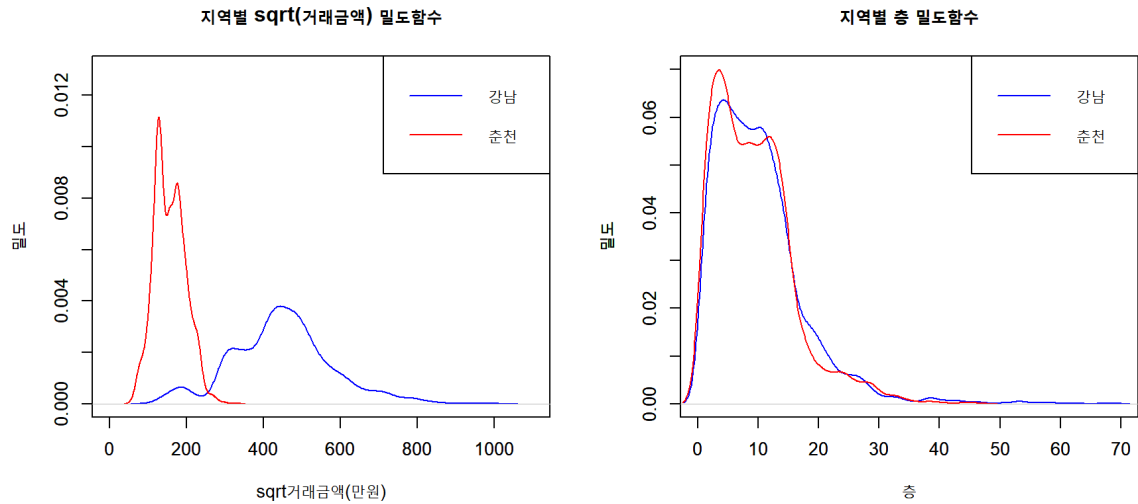
2) 층에 따른 거래금액 비교



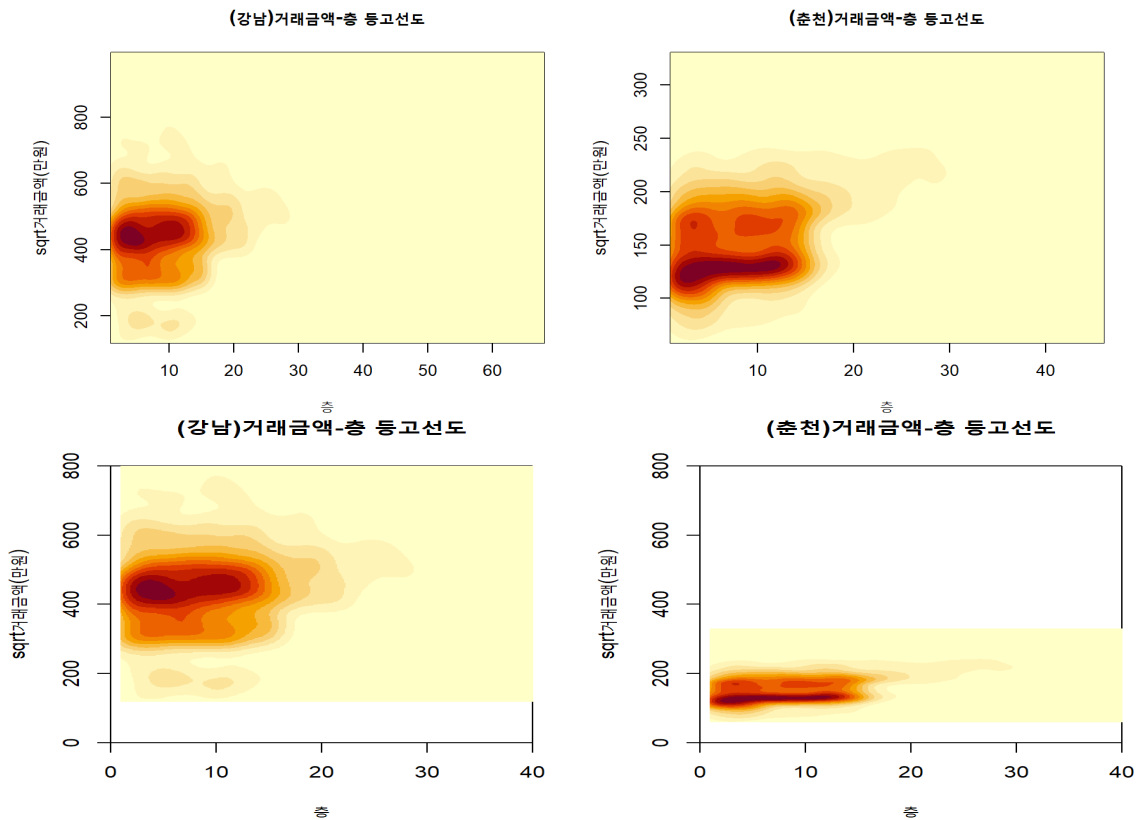
다음은 춘천과 강남의 거래금액-층의 보자기 플롯이다. 자료의 형태를 정규분포에 유사하게 맞추기 위하여 거래금액은 제곱근변환했다. 층은 왼쪽으로 치우친 분포이지만, 값이 작아 로그변환을 하지 않았다. 층을 로그변환한다면 자료의 분포가 자연스럽지 못한 형태를 띠기 때문이다. 초록 실선은 최소제곱법을 활용한 단순선형회귀직선이다. 붉은 실선은 평활너비를 2/3으로 한 로웨스 회귀곡선이다.

두 자료의 중간점의 층 값은 10층에 조금 못 미치는 정도로 별 차이가 없다. 그러나 중간점의 거래금액 값은 춘천 약 1억, 강남 약 16억으로 큰 차이를 보인다. 앞서 살펴보았듯 춘천 아파트의 거래금액이 강남 아파트의 거래금액보다 낮은 경향이 있다. 또한, 강남 지역의 아파트 거래금액의 분산이 춘천 지역의 아파트 거래금액의 분산보다 크다. 두 자료 모두 적지 않은 외부점을 가지고 있으나, 이는 영향점으로 판단하겠다.

단순선형회귀직선에 따르면, 층과 거래금액은 작은 양의 상관을 가진다. 두 지역 모두 낮은 층에서는 거래금액이 상수로 유지되지만, 15층을 기점으로 로웨스 회귀곡선의 기울기가 가팔라진다. 이러한 현상의 원인은 아파트의 고층 선호가 많기 때문일 수 있다. 주변 소음이 없고 경관이 좋다는 점에서 높은 층을 선호하는 사람이 많기 때문이다. 또는, 단순히 15층 이상의 층수를 가지는 고층 아파트가 15층 미만의 층수를 가지는 저층 아파트보다 비싸기 때문일 수 있다. 고층 아파트는 저층 아파트보다 최근에, 비싼 토지에서, 높은 건설 기술을 활용해 지어지기 때문이다. 한편, 층이 높아질수록 거래 빈도가 낮아지는 경향이 있다. 그러나 산점도에는 여러 점이 표현되어 있어 구체적인 빈도를 파악하기는 어렵다.



위는 지역별 거래금액과 층의 밀도함수이다. 강남의 경우, 제곱근변환한 거래금액의 밀도함수를 보면 16억에서 25억 사이에 봉우리가 한 개 있다. 층의 밀도함수에서는 0에서 20층의 넓은 폭을 지니는 큰 봉우리가 한 개 존재한다. 따라서 강남의 층과 거래금액 자료에는 한 개의 군집이 있으리라 예상할 수 있다. 한편, 춘천의 경우, 제곱근변환한 거래금액의 밀도함수를 보면 1억과 4억 근방에서 큰 봉우리와 작은 봉우리가 있다. 이는 춘천 자료에 군집이 두 개 있음을 시사한다. 층의 밀도함수에는 강남과 마찬가지로 0에서 20까지의 폭을 지니는 큰 봉우리 한 개가 존재하는 듯 보인다.



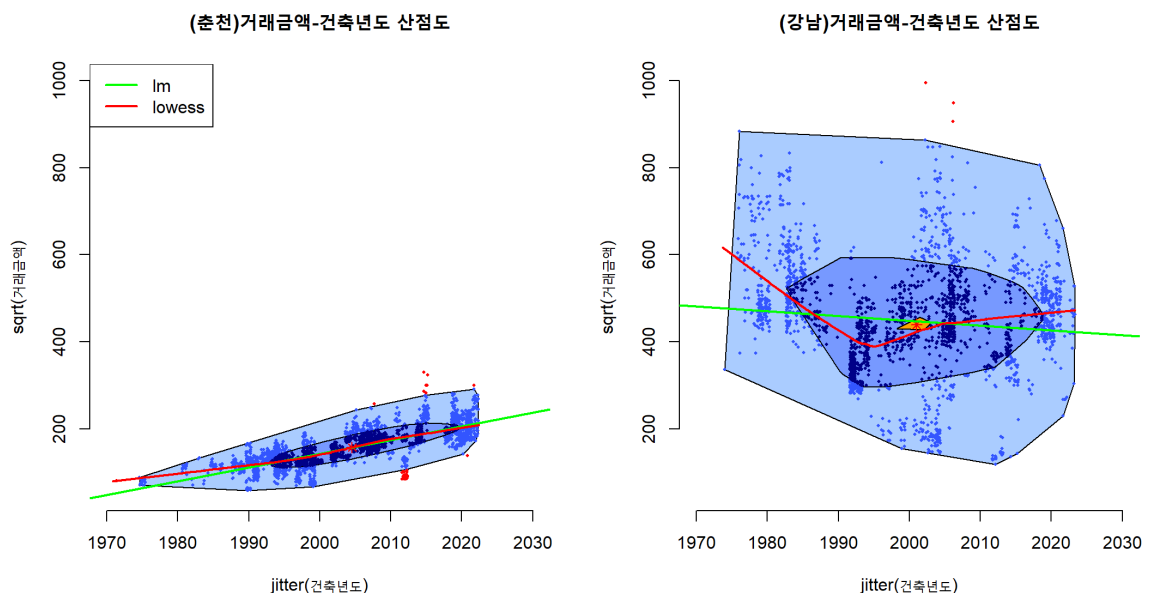
밀도함수를 등고선 이미지로 나타내면 위와 같다. 위쪽 두 개는 x축 y축 범위를 맞추지 않은 것이고, 아래쪽 두 개는 범위를 맞춘 것이다.

강남 아파트의 경우, 20층 이하에 거래금액이 18억정도인 위치에 자료가 몰려있다. 특히 10층 이하 좁은 지역에 색이 매우 진하여 밀도가 높음을 알 수 있다. 춘천 아파트 자료의 경우에도 10층 이하의 위치에 자료가 몰려있지만, 거래금액은 1억 5천만원 정도로 강남에 비해 낮다. 또한, 진한 색의 지역도 0층에서 12층가량까지 긴 구간에 걸쳐 있다. 두 지역 모두 군집의 개수는 한 개이다. 이 밀도의 시각화 자료를 통하여 밀도가 높은 위치 조건의 아파트가 거래량이 많으리라고 짐작할 수 있다.

층에 따른 거래량 차이로 미루어 다음의 사실들을 예상할 수 있다. 첫째, 15층 이하 아파트의 수요와 공급이 원활하다. 앞서 아파트 입주자들은 저층보다는 중간층 이상을 선호하는 경향이 있음을 언급했다. 그럼에도 자료의 밀도는 저층 아파트의 거래가 활발함을 보여준다. 특히 강남의 경우 그 경향이 확연하다. 따라서 고층 선호 정도가 저가 선호 정도보다 작다고 짐작할 수 있다.

둘째, 20층 이상 아파트의 공급이 없을 수 있다. 만약 최고 층수가 20인 아파트들이 대부분이라면, 20층 이상 아파트의 거래량은 20층 이하 아파트의 거래량보다 적으리라 기대된다. 또는 20층 이상 아파트의 소유자들이 보유한 아파트를 판매하지 않을 수 있다. 이 경우 로웬스 회귀곡선의 기울기 증가도 설명되는데, 고층 아파트에 대한 수요는 크지만 공급은 적으므로 거래금액이 상승했다고 보인다.

3) 건축년도에 따른 거래금액 비교



다음은 춘천과 강남의 거래금액-건축년도의 보자기 플롯이다. 자료의 형태를 정규분포에 유사하게 맞추기 위하여 거래금액은 제곱근 변환했다. 건축 연도는 연 단위 수치형 변수로 정리되어 있었기에, 자료의 자연스러운 분포를 보여주기 위하여 jitter 처리를 했다. 초록 실선은 최

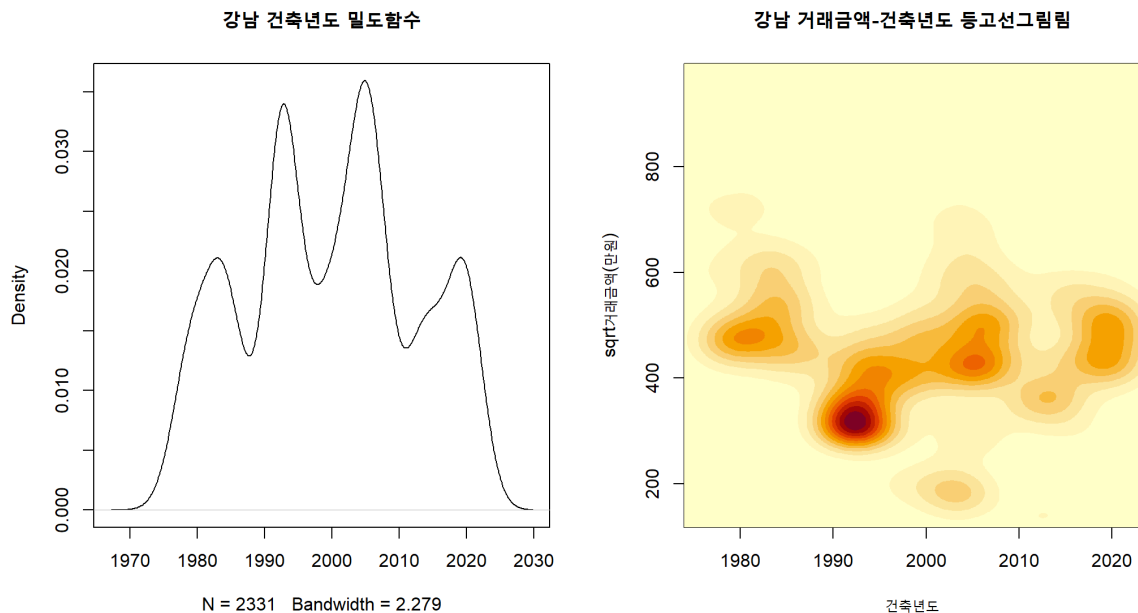
소제곱법을 활용한 단순선형회귀 직선이다. 붉은 실선은 평활 너비를 2/3로 한 로웬스 회귀곡선이다.

자료점들의 분포를 볼 때, 춘천의 자료점들은 선형성을 가진다. 반면 강남의 경우 자료점들의 선형관계를 판단하기는 어렵고, 일정 구역에 뭉쳐 있는 모습을 보인다. 두 자료 모두 외부점은 많지 않다. 춘천 아파트의 경우 건축년도와 거래금액의 중앙점은 약 2005년, 2억에 위치한다. 강남 아파트의 경우 중앙점은 약 2000년, 28억에 위치한다.

단순선형회귀직선을 보면, 춘천 아파트의 건축년도와 거래금액 사이에는 양의 상관성이 있다. 즉, 춘천 지역은 신축 아파트일수록 가격이 비싸다. 반면 강남 아파트의 경우, 건축년도와 거래금액 사이에 음의 상관성이 있다. 그러나 자료의 분산이 크므로 설명력은 낮으리라 예상된다.

로웬스 회귀곡선을 보면 보다 자세한 추세를 파악할 수 있다. 춘천 지역 아파트는 약 1995년까지는 건축년도가 상승에 따라 거래가격이 완만히 증가하지만, 1995년 이후부터는 거래가격이 비교적 가파르게 상승한다. 반면 강남 지역의 아파트는 1975년~1995년 구간에서는 건축년도와 거래가격이 강한 음의 상관관계를 가지고, 1995년~2005년 구간에서는 양의 상관관계를, 2005년~2025년 구간에서는 약한 양의 상관관계를 가진다. 강남 지역 로웬스 곡선의 모습을 볼 때, 강남 지역 아파트의 건축년도 변수는 조각별 선형회귀를 하는 것이 적절하다.

강남 지역 아파트 가격의 특이한 추세는 아파트의 연식이 아닌 다른 변수들이 아파트 가격 결정에 주요한 요인임을 시사한다. 추가적인 분석을 위하여 강남 지역 자료를 등고선 그림으로 나타내었다.



강남 건축년도의 밀도함수를 보면, 봉우리가 네 개니, 각 위치별로 네 군집이 있으리라 예상된다. 건축년도와 거래금액의 등고선도를 보면, 다섯 개의 군집이 보인다. 1990년 9억 근방에 자료의 밀도가 가장 높은 군집이 하나 있다. 2005년 16억 근방에 두 번째 군집이 있다. 1980

년 25억 근방에 세 번째 군집이 있고, 2020년 25억 근방에 네 번째 군집이 있다. 특이하게도 2005년 4억 근방에 작은 군집이 추가로 존재한다. 이는 등고선 그림이 강남 밀도함수의 특성 뿐 아니라 거래금액 밀도함수의 특성도 반영하기 때문이다.

강남 아파트는 건축년도의 분포가 고르지 않고 네 구간에 밀집되어 있음을 확인했다. 특히 1990년대 초에 건축된 아파트가 많은데, 이 원인을 한국 부동산 역사를 참고하여 해석하여 보겠다. 89년에 토지공개념¹⁾이 제정되어 신도시 개발과 같은 국책사업에서 토지문제가 해결되었다. 이후 수도권에 5대 신도시가 개발됐는데, 이를 효율적으로 진행하기 위하여 93년에 국토이용계획이 개편²⁾된다. 95년에는 건축법을 개정하여 건축절차를 간소화하고 건축편의를 도모했다. 이러한 사실에 미루어 볼 때, 해당 군집의 발생 원인은 1990년도 건설 붐 때문이라고 추측된다.

특이하게도 1980년대 위치한 군집의 거래금액이 매우 높다. 이는 아파트의 노후화로 인한 감가를 웃도는 다른 금액 결정 요인들이 해당 군집의 아파트에 존재함을 시사한다. 1980년대 강남에 건설된 아파트들은 삼성로 은마아파트, 개포동 개포주공아파트, 대치동에 한보미도맨션 등으로 고가의 토지에 건설되었다. 따라서 해당 아파트들의 가격에는 토지의 가격, 인프라, 학군, 교통편의성 등이 아파트의 건축년도 보다 영향력 있는 변수로 작용했다고 짐작된다.

1) 토지를 공공재의 개념으로 보아 국가적인 필요가 있을 시, 국가가 개인에게 적절한 보상을 해주고 그 토지를 사용하는 것.

2) 용도지역을 단순화하여 개발을 막아둔 지역의 비율을 줄이고, 개발이 가능한 용도지역을 확대했다.