

Report

지역사회의 향후주거의향에 미치는 요인 분석

- <안양시 사회조사 데이터>를 바탕으로

2022.11.17

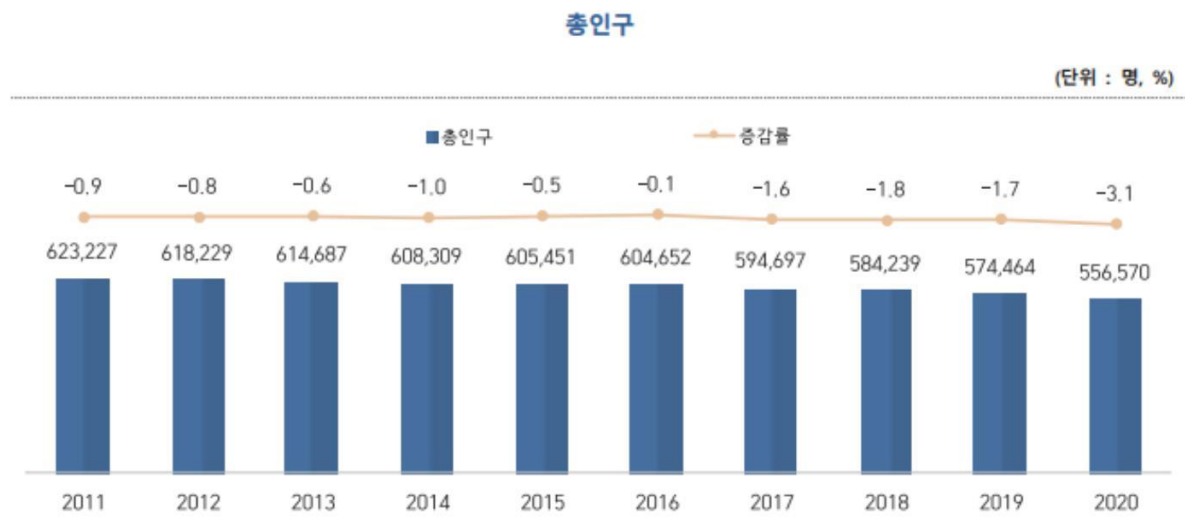
목차

1. 분석의 동기(Problem motivation)	2
2. 문헌 연구(Literature review)	4
3. 분석의 목적(Research objective)	5
4. 자료 및 방법론 설명(Data and method explanation)	6
5. 분석 결과 정리 보고(Analysis result)	8
5-1. 회귀분석 모델.....	8
5-2. 다중공선성 및 VIF 분석	9
5-3. BP test(Breusch-Pagan 검정).....	9
5-4. DW test(Durbin-Watson 검정).....	10
5-5. Normal Q-Q Plot.....	11
5-6. Shapiro-wilk normality test (샤피로 윌크 검정).....	11
5-7. Leverage Point(이상치 판단).....	12
5-8. 결론	13
6. 참고 문헌(Reference list)	14

1. 분석의 동기(problem motivation)

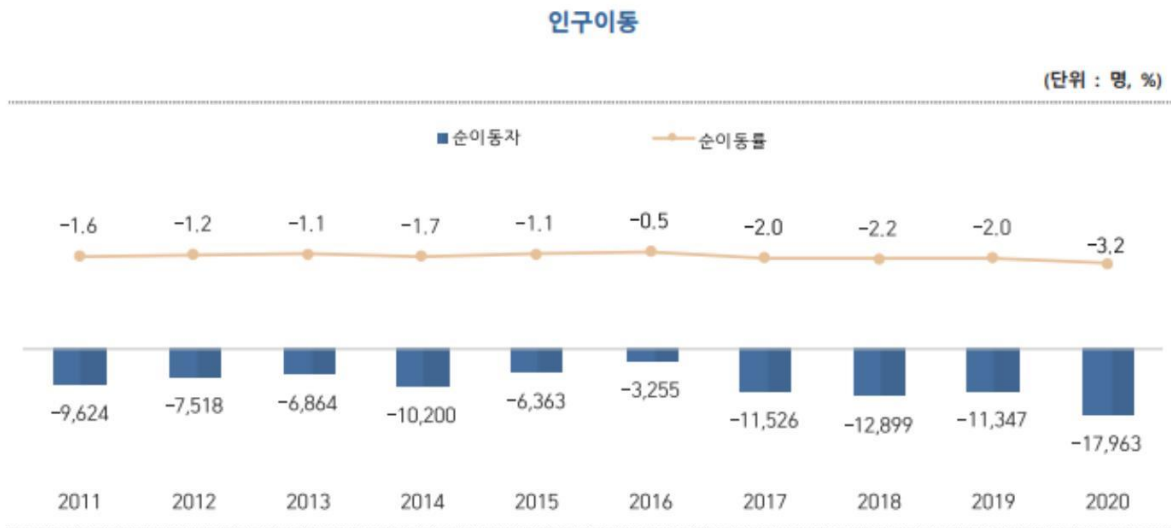
현재 우리나라는 세계에서 가장 빠른 속도의 초저출산과 고령화를 겪고 있다. 20 년 동안 우리나라는 1.3 명 이하 초저출산이 계속되고 있고, 2018년에는 합계출산율이 0.98 명으로 OECD 회원국 중에서 합계출산율이 1.0 미만인 유일한 나라가 됐다. 2005 년부터 우리나라는 저출산·고령사회대책을 국가정책으로 추진해오고 있지만 저출산 대책에만 집중하여 지역의 사회적 인구감소 문제에 대해서는 대책이 미흡한 편이다. 지역의 인구감소 문제는 결국 일자리, 교육, 복지, 교통, 행정 등이 포함되는 격차문제, 균형 발전 문제로 귀착된다. 저출산 문제는 국가의 주요한 해결과제로써 많은 연구가 이루어지고 있으나, 상대적으로 지역 인구감소의 원인에 대해서는 포커스가 맞춰지지 않고 있다. 하여 본 리포트에서는 지역이동에 의한 인구감소에 대해 다루고자 한다.

<표 1> 안양시 총인구



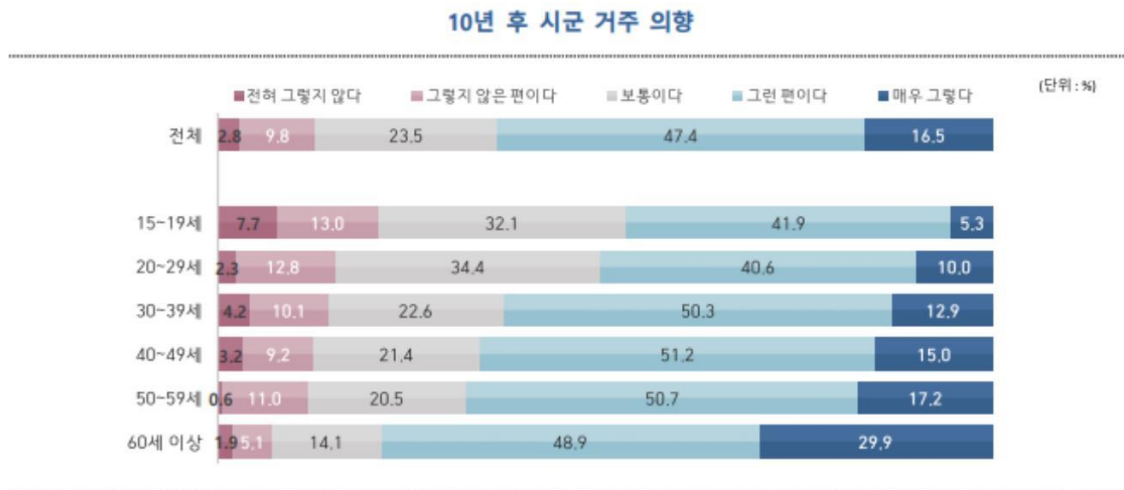
보고자가 현재 거주하고 있는 안양시의 2021 년 인구조사결과를 살펴보면(표 1), 안양시의 총인구는 매년 감소하고 있다. 특히 매년 1% 미만의 감소율을 보이던 반면, 2020 년부터는 3%라는 수치로 급격히 감소되었다.

<표 2> 안양시의 인구이동



그 중, 인구 이동으로 인한 인구 변화 지표를 보자면 2020 년 한 해에만 전체 안양시 인구의 3.2%인 17,963 명의 순유출이 있었다. 이는 지난 10 년 동안의 수치 중 가장 큰 변화로, 상당인구가 안양시에서 전출되고 있음을 단적으로 보여주고 있다.

<표 3> 10 년 후 안양시 거주 의향



10 년 후 안양시 거주 의향 지표를 보면 안양시 전체 인구의 63.9%가 10 년 후 안양시에 거주하고 싶다고 답했다. 특히 10 대인구, 즉 젊은 층의 향후 거주의향은 47.2%에 불과해, 미래 안양시 인구의 고령화에 기여할 것으로 보인다.

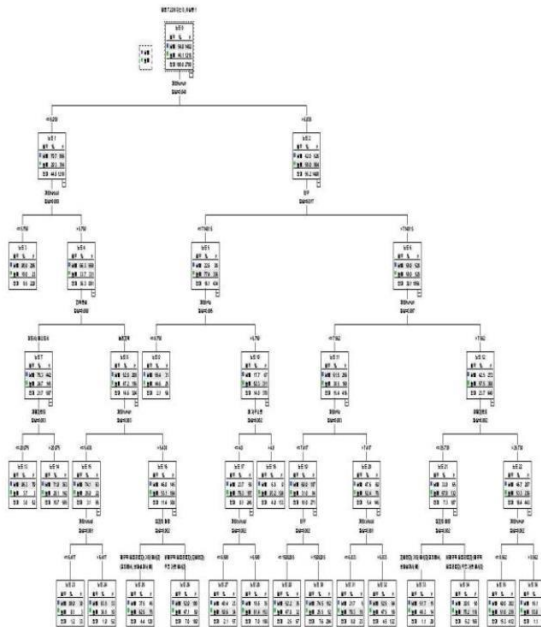
꾸준히 인구가 감소하는 가운데 2020 년을 기준으로 큰 폭의 변화가 있었기 때문에 향후 빠른 인구 유출이 예상되는 바, 현재 안양시에 거주하고 있는 인구의 향후 거주의향에 영향을 미치는 요인을 파악함으로써 관련 정책 수립에의 의견을 제안하는 등, 보고자가 현재 거주하고 있고 앞으로도 거주를 희망하고 있는 지역인 안양시를 좀 더 살기 좋은 지역으로 만들기 위한 노력에 기여하고자 본 리포트를 보고하게 되었다.

2. 문헌연구(literature review)

기존의 다른 선행연구들이 거주의향 요인으로 제시하고 있는 영향요인들은 연령, 소득수준, 결혼상태, 학력, 지역의 수준, 직장까지의 교통소요 시간, 직장까지의 교통 비용, 슈퍼마켓까지의 교통 비용, 인구밀도, 주거지역, 주거지 인근 학교 수준, 지역의 인구규모, 주택가격, 성별, 함께 사는 어린이 수 등 다양하다. 본 리포트는 "지역거주 의향에 미치는 영향요인들의 조건결합"¹ 연구를 참고하였다. 해당 연구에서는 우리나라 지역 주민 2700명을 대상으로 하여 주민들의 거주의향에 영향을 미치는 변수들을 설정하고 (개인적 배경 변수, 지역특성 변수, 정책적 변수) 이러한 변수들 중에서 어떠한 변수들이 영향을 미치는 데 있어서 중요한 변수인지, 그리고 거주의향이 높은 그룹의 특성이 무엇인지를 파악해 내는 것을 연구목적으로 하였다. 현재 살고 있는 지역에 대한 계속 거주의향 영향요인을 확인하기 위하여 의사결정나무 분석을 수행하였고, 그 결과 인본적 자본 변수(공공도서관, 평생교육, 지역내 학교수준, 건강수준, 의료, 노인, 장애인, 보육 등 8개)의 값이 높으며, 인구가 작은 자치단체에 살고 인프라 자본이 높고, 월가구 소득이 500 만원 이상 되는 주민 그룹이 1위를 차지하였다.

¹ 최영철. 2019, "지역거주 의향에 미치는 영향요인들의 조건결합 - 의사결정나무분석 활용 -", 한국지방행정학보, vol.16, no.1 pp.253-276.

<그림1> 거주의향 수준 의사결정나무 모형



특히 개인적 배경변수, 지역특성 변수, 그리고 정책적 변수들을 확인한 결과, 지역 주민들의 거주주의향에 대해서는 정책적 변수들이 크게 영향을 미치는 것으로 나타났다. 정책적 변수들 중에서도 인본적 자본변수 (공공도서관, 평생교육, 지역내 학교수준, 건강수준, 의료, 노인, 장애인, 보육 등 8개)들이 다른 정책변수들 보다도 상대적으로 더 중요한 영향을 미치고 있다는 점이다. 이런 결과들에 대해 지방자치단체들은 자원배분이나 정책결정과정에서 크게 관심을 가져야 할 것이다. “안양시”에 대해서도 다양한 변수들을 통해 확인이 필요한 부분이나, 시간관계 상 여의치 않아 개인적 배경변수에 한하여 영향도를 분석해본다.

3. 분석의 목적(research objective)

본 리포트는 가구원수계, 나이, 성별, 교육정도, 혼인 상태, 지역거주기간, 통근 시간, 월평균가구소득과 같은 독립변수를 사용하여 다중 회귀모형을 구축하고 각각의 변수들이 종속변수인 향후거주의향에 통계적으로 유의한 영향을 미치는 지 살펴보고자 한다. 선행연구에서 현재 살고있는 지역의 거주주의향이 높은 그룹의 조건으로 월가구 소득이 500만원 이상인 주민그룹이 상위 3개 그룹에 포함되었으므로, 본 리포트에서도 경제적 상황에 따른 거주 희망도에 집중하고자 한다. 하여 독립변수 중 월평균가구소득에 집중해 “월평균가구소득이 높을수록 안양시에 거주하려는 의향이 높을 것이다”는 대립가설을 검증하고 그 영향력을 알아보고자 한다. 이를 토대로 균형적 지역개발을 위한 장기적 정책수립에의 합리적인 기초자료로 활용되기를 기대해 본다.

4. 자료 및 방법론 설명(data and method explanation)

4-1. 사용 데이터

대한민국 통계청의 <2021년 제7회 안양시 사회조사>² 데이터를 사용하였다. 해당 데이터는 2020년 9월 1일부터 2021년 8월 31일까지 1년 간 안양 시민 총 인구인 556,570명을 대상으로 주거지 중심의 평소 생활과 만족도를 조사하여 시민 생활의 양적, 질적 수준을 종합적으로 측정하고자 한 결과이다. 본 리포트에서는 전체 모집단 중 안양시 향후 거주意向 문항에 응답한 안양시민 1200명의 표본 데이터를 가지고 회귀분석을 진행하였다.

4-2. 변수 소개

4-2-1. 종속변수

1) future_residency (향후 거주意向)

현재 거주하는 안양시에 10년 후에도 살意向이 있는지 응답자로부터 확인한 결과이며, 5점 Likert 척도로 측정되었다. (1:전혀 그렇지 않다, 2:그렇지 않은 편이다, 3:보통이다, 4:그런 편이다, 5:매우 그렇다)

4-2-2. 독립변수

1) num_family (가구원수계)

응답자가 속한 가구의 총 가구원수를 나타낸다. 수치형 변수로 수집되었다.

2) age(나이)

응답자의 만 나이를 나타낸다. 연속형 변수이다.

3) gender(성별)

응답자의 성별을 나타내며 본 분석에서 dummy변수로 사용되었다.

4) education(교육정도)

응답자의 교육정도를 나타내는 변수로, 본 분석에서 dummy변수로 사용되었다. 설문조사 raw data에서의 구분은 다음과 같다. (1:안받았음, 2:초등학교, 3:중학교, 4:고등학교, 5:대학(4년제 미만), 6:대학(4년제 이상), 7:대학원석사, 8:대학원박사과정)

² 데이터 출처 : https://mdis.kostat.go.kr/ofrData/selectOrgOfrData.do?curMenuNo=UI_POR_P9220

6) marriage(혼인상태)

응답자의 응답 당시 혼인상태를 나타내며, 본 분석에서 dummy변수로 사용되었다.

(1:미혼, 2:배우자 있음, 3:사별, 4:이혼)

7) residency(지역 거주기간(년))

응답자의 안양시 거주기간을 의미하며, 연 단위로 수집되어 이산형 변수이다.

8) commute_time (통근시간(분))

응답자의 평균 통근시간을 나타내는 변수로, 분 단위로 수집되었다.

9) income(월평균가구소득)

아래와 같이 10단위로 응답을 받아 범주형 변수이므로 본 리포트에서는 dummy변수로 사용되었다. 설문조사 raw data에서의 구분은 다음과 같다.

<표4> 월평균가구소득 변수명

응답항목	응답 문항	변수명
1	50만원 미만	Under50
2	50 ~ 100만원 미만	50_99
3	100 ~ 200만원 미만	100_199
4	200 ~ 300만원 미만	200_299
5	300 ~ 400만원 미만	300_399
6	400 ~ 500만원 미만	400_499
7	500 ~ 600만원 미만	500_599
8	600 ~ 700만원 미만	600_699
9	700 ~ 800만원 미만	700_799
10	800만원 이상	over800

5. 분석결과 정리 보고(analysis result)

5-1. 회귀분석 모델

<표5> Rstudio를 통한 회귀분석

```
Call:
lm(formula = future_residency ~ num_family + age + factor(gender) +
    factor(education) + factor(marriage) + residency + commute_time +
    factor(income), data = datay)

Residuals:
    Min       1Q   Median       3Q      Max
-3.0556 -0.4762  0.1493  0.5602  1.8082

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   3.455808   0.239545  14.427 < 2e-16 ***
num_family    -0.090742   0.032680  -2.777  0.00558 **
age            0.005714   0.002714   2.105  0.03548 *
factor(gender)male  0.098962   0.073226   1.351  0.17681
factor(education)high -0.124808   0.114828  -1.087  0.27730
factor(education)master -0.060422   0.152006  -0.397  0.69107
factor(education)middle  0.079994   0.132311   0.605  0.54557
factor(education)none -0.156979   0.226724  -0.692  0.48883
factor(education)phd    -0.139359   0.222003  -0.628  0.53030
factor(education)university2 -0.116055   0.133612  -0.869  0.38525
factor(education)university4 -0.143475   0.125217  -1.146  0.25211
factor(marriage)married -0.088529   0.105567  -0.839  0.40186
factor(marriage)single -0.121696   0.123174  -0.988  0.32336
factor(marriage)widowed  0.022875   0.127353   0.180  0.85748
residency      0.016043   0.001925   8.335 < 2e-16 ***
commute_time   -0.003025   0.002885  -1.049  0.29462
factor(income)200_299 -0.066344   0.096174  -0.690  0.49044
factor(income)300_399  0.004943   0.100844   0.049  0.96092
factor(income)400_499  0.170324   0.111206   1.532  0.12589
factor(income)50_99    0.010987   0.113115   0.097  0.92264
factor(income)500_599  0.173076   0.123346   1.403  0.16083
factor(income)600_699  0.141770   0.141924   0.999  0.31804
factor(income)700_799  0.186828   0.158093   1.182  0.23754
factor(income)over800  0.115357   0.125663   0.918  0.35881
factor(income)under50  0.223897   0.180557   1.240  0.21521
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.8915 on 1175 degrees of freedom
Multiple R-squared:  0.1368,    Adjusted R-squared:  0.1191
F-statistic: 7.756 on 24 and 1175 DF, p-value: < 2.2e-16
```

Rstudio를 통한 초기 회귀분석 결과는 위와 같다. 본 리포트의 주요변수인 income은 통계적으로 유의하지 않은 것으로 나타났다. 'Residency'는 유의수준 0.001에서 강한 통계적 유의성을 갖는다. 안양시에 오래 거주했을수록 향후 거주지향에 긍정적이다. 'num_family'는 유의수준 0.01에서 통계적으로 유의하였다. 가족 구성원 수가 많을수록 향후 안양시 거주지향이 있는 것으로 나타났다.

다. 'age'는 유의수준 0.1에서 통계적으로 유의하였다. 나이가 많을수록 안양시에의 거주향이 높은 것으로 볼 수 있다.

5-2. 다중공선성 및 VIF 분석

<표6> 다중공선성 test 결과

```
> vif(out1)
```

	GVIF	Df	GVIF^(1/(2*Df))
num_family	2.259822	1	1.503270
age	2.425440	1	1.557382
factor(gender)	1.564143	1	1.250657
factor(education)	2.466515	7	1.066611
factor(marriage)	4.112742	3	1.265771
residency	1.302031	1	1.141066
commute_time	1.035813	1	1.017749
factor(income)	2.908730	9	1.061112

해당 데이터의 독립변수들 간 다중 공선성이 있는지 확인하는 테스트이다. VIF(Variance Inflation Factor) 값이 1이면 다중 공선성이 없음을 의미하고, 10 이상부터 독립변수 사이에 강한 다중공선성이 존재한다고 말할 수 있다. 만약 다중공선성이 존재한다면 상관관계가 높은 독립변수 중 하나 또는 일부를 제거하여 해결해야 하며, 경우에 따라 변수를 변형시키거나 새로운 관측치를 이용하는 방법이 있고 주성분 분석을 이용해 설명력이 높은 변수를 선택하는 등 다중공선성은 해결해야 하는 선결과제이다. 위의 결과와 같이 VIF값이 대부분 1~3 사이의 값으로 심각한 다중공선성은 없는 것으로 보인다.

5-3. BP test(Breusch-Pagan 검정)

<표7> BP test 결과

```
> #BP TEST
> bptest(out1)
```

studentized Breusch-Pagan test

data: out1

BP = 42.709, df = 24, p-value = 0.01074

모델이 등분산성을 만족하는지 확인하는 테스트이다. 해당 테스트의 귀무가설은 '모델이 등분산성을 만족한다'이고, '모델이 등분산성을 만족하지 않아 이분산성을 갖는다'는 대립가설이 된다. 만약 귀무가설을 기각한다면 등분산성을 만족하지 않아 회귀분석으로는 정확도가 떨어질 수 있다. 위의 결과와 같이 p-value는 0.01074로 유의수준 0.05에서 통계적으로 유의하므로 귀무가설을 기

각한다. 즉 해당 모델이 이분산성을 갖는다는 문제가 있다.

5-4. DW test(Durbin-Watson 검정)

<표8> DW test 결과

```
> #DW TEST
> dwtest(out1)

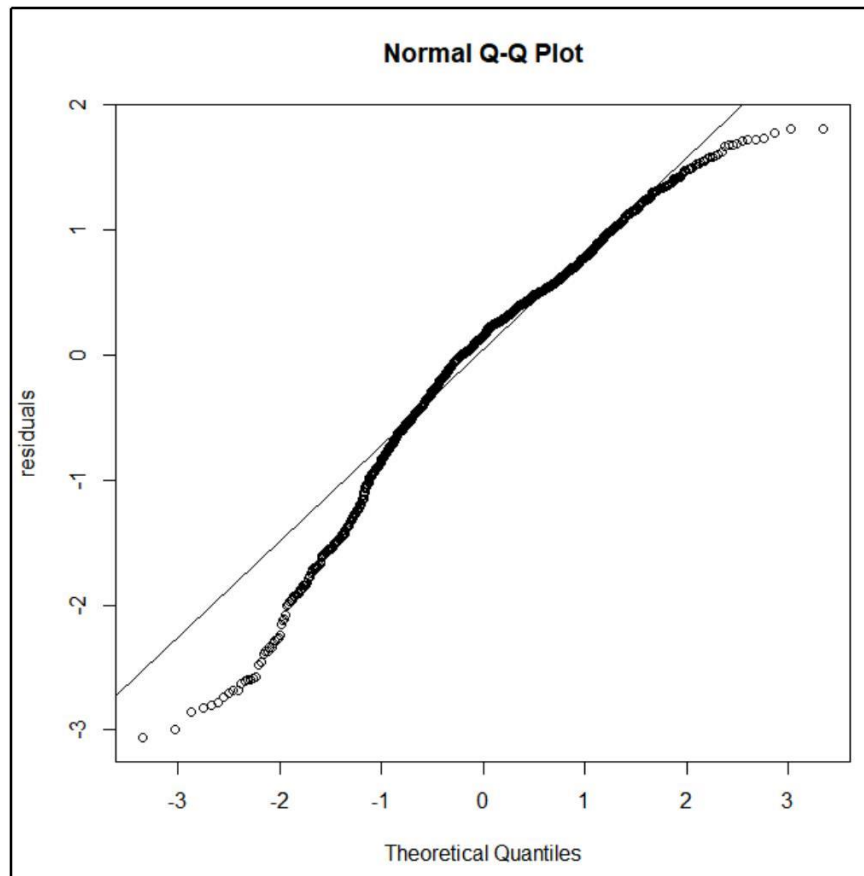
Durbin-watson test

data: out1
DW = 2.0585, p-value = 0.723
alternative hypothesis: true autocorrelation is greater than 0
```

DW test는 회귀분석 후 잔차의 독립성을 확인할 때 사용하는 테스트이다. 잔차들 사이에 자기상관성이 있는지 없는지를 판단한다. 귀무가설은 '자기상관성이 0이다'이고, 대립가설은 '자기상관이 0보다 크다'이다. 위의 결과와 같이 p-value가 유의수준 0.05보다 크기 때문에 귀무가설을 기각할 수 없고 이는 이 모델의 자기상관성이 0이라는 해석을 할 수 있다.

5-5. Normal Q-Q Plot

<그림2> Normal Q-Q Plot 결과



Normal Q-Q Plot은 해당모델의 잔차가 정규분포를 따르는지를 보여준다. 산포도의 점이 Q-Q line 선상에 있으면 정규분포를 따른다고 임의적으로 판정한다. x축이 커질수록 선상에서 떨어지는데 '정규분포가 아니다' 라고 얘기할 수 있는 수준일까? Q-Q Plot은 그림이기 때문에 명확하게는 답을 낼 수 없다. 하지만 Normal Q-Q Plot의 산포도 모양으로 분포의 종류를 추론 가능하다.

x축의 중간 값은 0이고, 이를 중심으로 좌측과 우측으로 구분하여 위의 결과를 보면, 좌측에서 산포도 점들이 Q-Q line보다 아래 쪽에 있으므로 분포 좌측은 long-tail 형태, 우측에서 산포도 점들이 Q-Q line보다 아래 쪽에 있으므로 분포 우측은 short-tail 형태라고 할 수 있다.

5-6. Shapiro-wilk normality test (샤피로 윌크 검정)

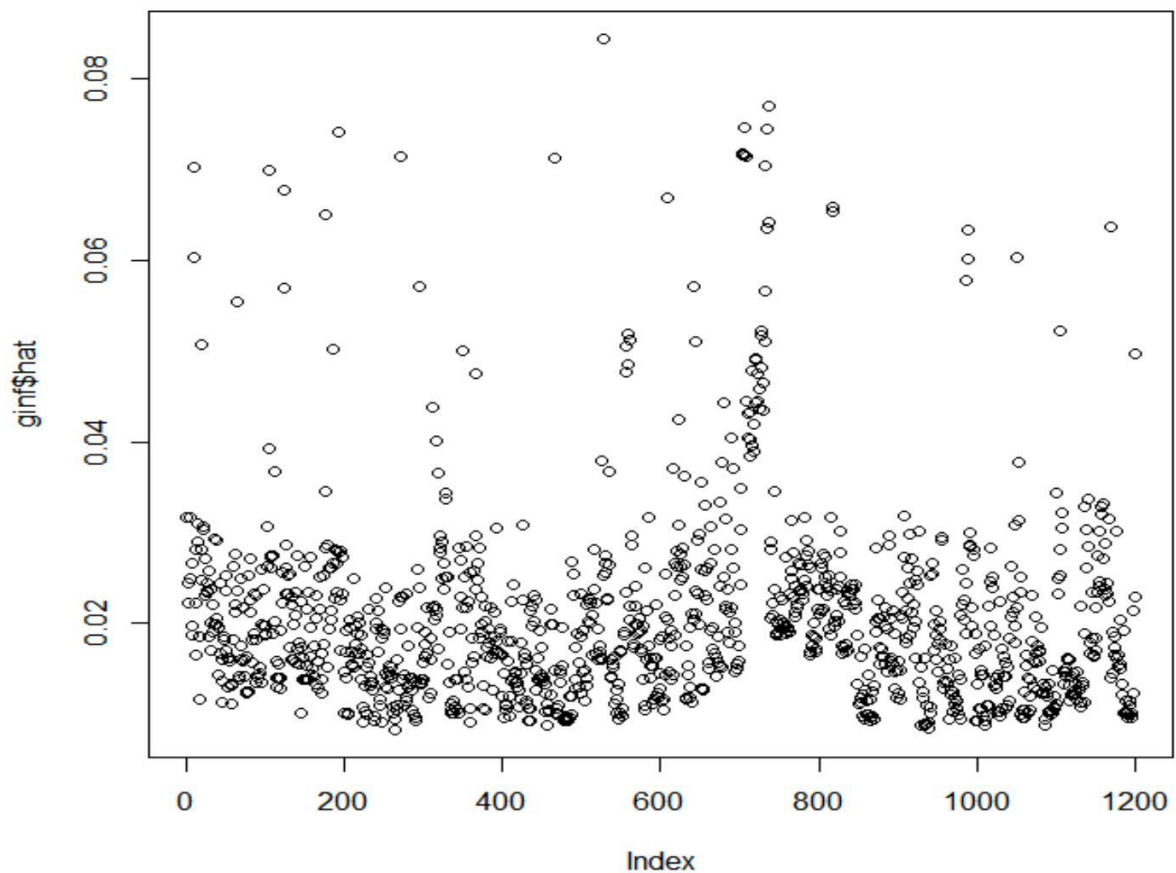
정확도가 높은 정규성 테스트이다. 정규분포만을 위한 테스트이며 귀무가설은 '해당 모델의 잔차가 정규분포를 따른다'이다. 결과를 보면 p-value가 유의수준 0.05보다 매우 작으므로 귀무가설을 기각해야한다. 즉 해당 데이터는 정규분포를 따르지 않는다는 것을 의미한다.

<표9> Shapiro-wilk normality test 결과

```
> shapiro.test(residuals(out1))  
  
      shapiro-wilk normality test  
data:  residuals(out1)  
W = 0.96105, p-value < 2.2e-16
```

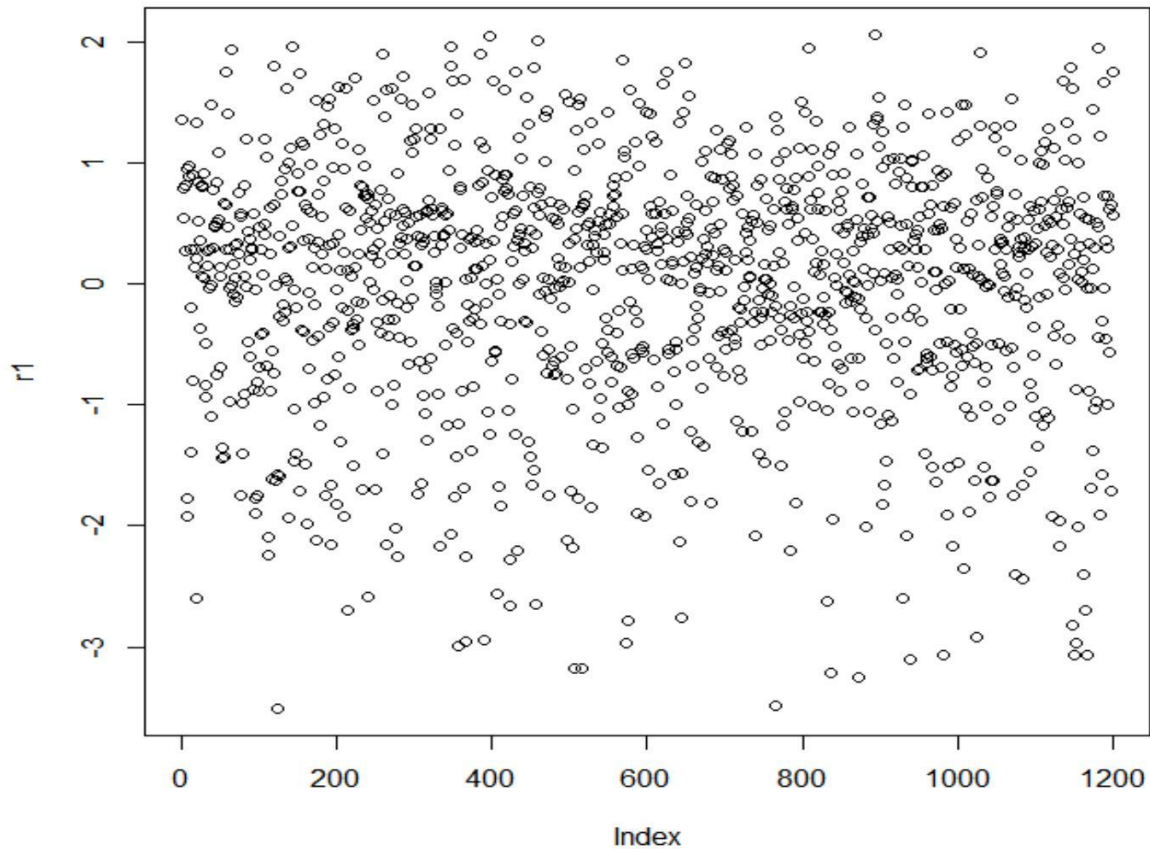
5-7. Leverage Point(이상치 판단)

<그림3> Leverage Point 확인 결과



Leverage Point를 구해 X방향의 이상치를 확인해보았다. 다중회귀분석의 경우 Leverage가 $2(k+1)/n$ 보다 크면 이상치라고 할 수 있다. (k =독립변수 개수) 위 결과를 보면 기준 0.048 기준으로 0.048을 초과하는 값들은 X방향 이상치로 볼 수 있다.

<그림4> sdr 확인 결과



위 결과는 Y방향의 이상치를 보여준다. Sdr(studentized deleted residual)의 절대값이 2 이상이면 이상치로 볼 수 있다.

5-8. 결론

이 분석은 안양시에 현재 거주하고 있는 시민들의 향후 거주 의향과 월평균 가구소득과의 상관 관계를 살펴보고 하였으나, 통계적으로 유의한 상관성을 발견하지 못하였다. 다른 독립변수인 가구원수가 많을 수록, 지역에 거주한 기간이 길 수록, 나이가 많을수록 향후 안양시에 계속 거주 할 것이라는 답변을 한 것으로 나타난다. 독립변수들 간에 다중공선성이 없다고는 하나 일반적으로 나이가 많을수록 지역에 거주한 기간이 긴 경향성이 있고, 거주지는 쉽게 결정하여 바꿀 수 없는 것이기 때문에 오래 거주한 사람일수록 향후에도 거주 의사가 있을 것으로 쉽게 예상되는 점이 본 분석의 아쉬운 점이다. 앞으로 예상되는 안양시 인구 유출을 예방하기 위해 매년 시행되는 안양시 사회조사 데이터를 들여다 보며 관심을 가지고 좀 더 다양한 변수 선택을 통한 추가 연구를 통해 본 연구의 부족한 점을 보완하면 좋을 것 같다.

6. 참고문헌(reference list)

박진경, 김도형. 2020, 인구감소대응 지방자치단체 청년유입 및 정착정책 추진방안, 한국지방행정연구원

이경영, 정문기, 정예은, 2018, '지속거주의향의 영향 요인에 관한 연구 : 서울시 5 대 생활권 비교를 중심으로', 한국행정연구, vol. 27,no. 2, pp. 179-210.

최영출. 2019, "지역거주 의향에 미치는 영향요인들의 조건결합 - 의사결정나무분석 활용 -", 한국지방행정학보, vol.16, no.1 pp.253-276.