



1인가구를 위한 공공주택 입지 선정

통계학과 1602006 권지혜
1678094 김지원
1729020 이수민

목차



01
분석 배경
&
목적

02
데이터 분석
-
①랜덤포레스트
&
로지스틱회귀

03
데이터 분석
-
②계층군집분석

04
기대효과
&
한계

05
참고문헌
&
출처

1. 분석 배경 및 목적 - 문제인식

[1인가구 시대, 어떤 정책 필요하나...](#) 농촌여성신문 | 2019.09.20. | [🔗](#)

[‘나 혼자 산다’ 170만가구... 주거공간 공유 급증세](#)

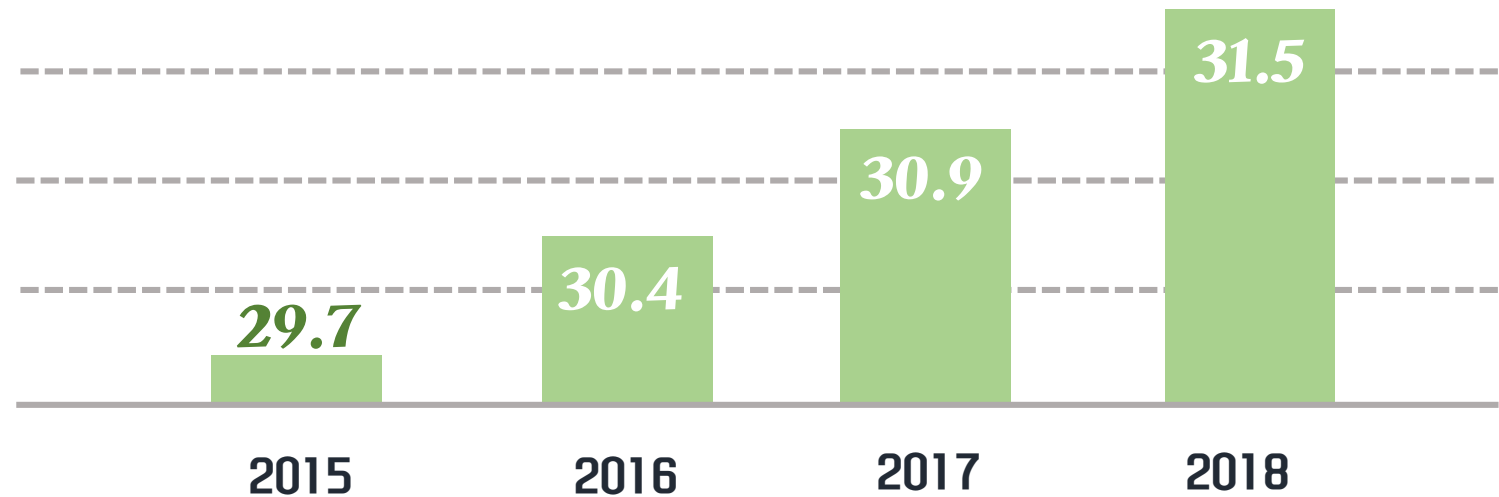
이데일리 | 1일 전 | 네이버뉴스 | [🔗](#)

변화하면서 1인가구의 수가 빠르게 증가하고 있다. 일명 혼밥, 혼술, 1인 임대 주택 등 1인가구에 맞춰... 이 사무관은 '국토교통부는 청년 1인가구 대상으로 공공임대주택, 셰어형 주택 공급을 하고 있고, 보건복지부는...

사는' 1인가구가 크게 증가하는 것으로 나타났다. 세대별로도 3세대 이상이 사는 대가족의 비중은 줄어드는 반면 부부들끼리만 거주하는 1세대가 늘고 있다. 특히 셰어하우스 등 주거를 공유하는 형태가 늘면서...

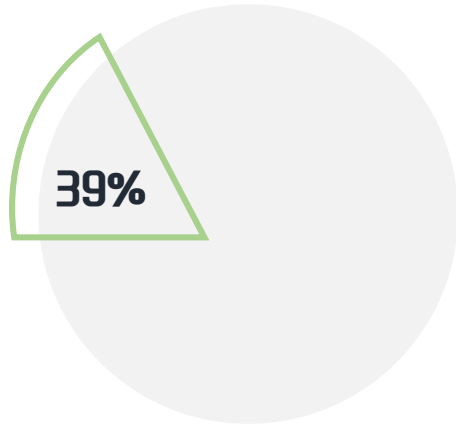
서울특별시 1인가구 비율(%)

출처 : 통계청, 「지역별고용조사」

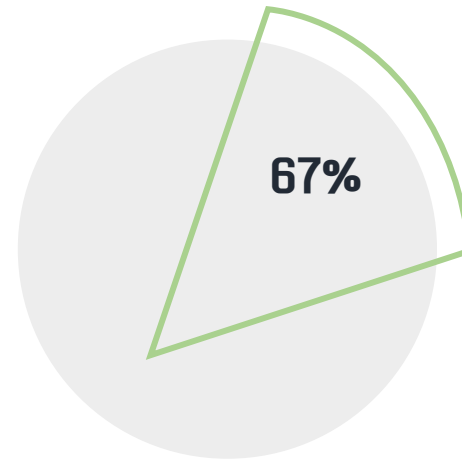


> 1인가구의 급증. 미래에도 꾸준히 늘어날 것으로 보인다.

1. 분석 배경 및 목적 - 문제인식



1인가구의 자가 주택 보유율



다인가구의 자가 주택 보유율

(2018년 기준)

【 다인가구 대비 1인가구의 자가 주택 보유율이 현저히 작은 것을 알 수 있다. 】

1. 분석 배경 및 목적

공공주택

**【 국가 또는 지자체가 주거 취약 계층을 위하여
공공 자금으로 건설하는 주택 】**

- › 늘어난 1인 가구를 위해 국가차원의 공공주택 설립이 필요하다.
- › 공공주택이 가장 필요한 **1인가구 세대주의 특징을 분석하여**
그를 위한 서울의 공공주택 설립의 최적 입지를 제시하고자 한다.

2.데이터 분석

※ 분석도구 – R studio

1단계

–
랜덤
포레스트

‘자가 소유 여부’와
‘1인 가구 여부’를
종속변수 y로 설정 후
유의한 변수를 확인

2단계

–
로지스틱
회귀

1단계에서 본
유의한 설명변수들로
로지스틱 회귀 적합하여
자가가 아닌 사람과 1인가구의
주요 특성을 확인, 타겟 설정

3단계

–
계층
군집분석

타겟 맞춤형 공공주택
입지 변수를 설정 후,
서울시 구별 데이터를 이용해
공공주택의 최적 위치를 선정.

2.데이터 분석

: 분석 데이터 소개

13차(2018) 한국 복지 패널 조사 자료

데이터 크기 : 6474개 * 1146(변수)

> 조사주체

: 한국보건사회연구원, 서울대학교 사회복지연구소

> 조사내용

: 가구일반사항, 경제활동상태, 소득, 지출, 부채, 사회보장 등

서울 열린 데이터 광장 - 구별 데이터



이 조사표에 기재된 내용은 통계법 제33조에 의하여 비밀이 보장됩니다.
작성기관: 한국보건사회연구원
서울대학교 사회복지연구소



2018년 한국복지패널조사
가구용[유형1]

안녕하십니까?

한국보건사회연구원과 서울대학교 사회복지연구소는 국무총리실과 보건복지부의 의뢰를 받아 한국복지패널조사를 실시하고 있습니다. 본 조사는 급격히 변화하는 사회환경 속에서 계층별, 연령별 인구집단의 생활실태와 사회복지 욕구를 역동적으로 파악하여 각종 복지정책 수립시 활용할 기초자료를 마련하는데 그 목적이 있습니다.

본 조사는 통계법 제25조에 의거하여 실시·관리되고 조사표에 기입되는 모든 응답내용은 통계법 제33조 및 제34조에 의거 통계목적에만 사용되고 그 비밀은 반드시 보장됩니다. 귀댁의 응답은 정부의 올바른 정책수립에 귀중한 기초자료로 이용되오니 시간을 내어 협조에 주시면 대단히 감사하겠습니다.

2018년 2월

< 문의 및 연락처 >
한국보건사회연구원 (☎ 044-267-8124, 8138, 8159, 8273, 8191, 8185)
서울대학교 사회복지연구소 (☎ 02-880-6320)

가구 패널 ID	가구생성자수		가구관리일련번호	
	1차 가구: 01 2차 가구: 02 3차 가구: 03 4차 가구: 04 5차 가구: 05 6차 가구: 06 7차 가구: 07 8차 가구: 08 9차 가구: 09 10차 가구: 10 11차 가구: 11	12차 가구: 12 13차 가구: 13 14차 가구: 14 15차 가구: 15 16차 가구: 16 17차 가구: 17 18차 가구: 18 19차 가구: 19 20차 가구: 20 21차 가구: 21 22차 가구: 22 23차 가구: 23 24차 가구: 24 25차 가구: 25 26차 가구: 26 27차 가구: 27 28차 가구: 28 29차 가구: 29 30차 가구: 30	1차 가구: 01 2차 가구: 02 3차 가구: 03 4차 가구: 04 5차 가구: 05 6차 가구: 06 7차 가구: 07 8차 가구: 08 9차 가구: 09 10차 가구: 10 11차 가구: 11 12차 가구: 12 13차 가구: 13 14차 가구: 14 15차 가구: 15 16차 가구: 16 17차 가구: 17 18차 가구: 18 19차 가구: 19 20차 가구: 20 21차 가구: 21 22차 가구: 22 23차 가구: 23 24차 가구: 24 25차 가구: 25 26차 가구: 26 27차 가구: 27 28차 가구: 28 29차 가구: 29 30차 가구: 30	1차 가구: 01 2차 가구: 02 3차 가구: 03 4차 가구: 04 5차 가구: 05 6차 가구: 06 7차 가구: 07 8차 가구: 08 9차 가구: 09 10차 가구: 10 11차 가구: 11 12차 가구: 12 13차 가구: 13 14차 가구: 14 15차 가구: 15 16차 가구: 16 17차 가구: 17 18차 가구: 18 19차 가구: 19 20차 가구: 20 21차 가구: 21 22차 가구: 22 23차 가구: 23 24차 가구: 24 25차 가구: 25 26차 가구: 26 27차 가구: 27 28차 가구: 28 29차 가구: 29 30차 가구: 30
※ 연호시점의 주치를 그대로 하기	※ 연호시점의 주치를 그대로 하기	※ 연호시점의 주치를 그대로 하기	※ 연호시점의 주치를 그대로 하기	
행정코드	시·도	구·시·군	동·읍·면	
상세주소	통·리	번지	호 (아파트) 동 층 호 () -	
가구주 성명	응답자 1		응답자 2	
	성명	가구원 인 호	성명	가구원 인 호
	휴대폰		휴대폰	
조사표원료 소요시간	총	분	총	회
1차방문	월 일 시 분	방문결과	① 완료 ② 미완료	※ 미원 사유코드
2차방문	월 일 시 분	방문결과	① 완료 ② 미완료	⑤ 미제정(관하) ① 높은 가격 ② 장거거리 ③ 무주택(원인비과)
3차방문	월 일 시 분	방문결과	① 완료 ② 미완료	④ 일부불량 미발 ⑤ 조사자부
4차방문	월 일 시 분	방문결과	① 완료 ② 미완료	⑥ 이상주소추적불가 ⑦ 사망으로 인한 가구소실
최종방문	월 일 시 분	방문결과	① 완료 ② 미완료	⑧ 기타 추적불가사유 ⑨ 기타
조사원 이름	지도원 확인		지도원	(인)

2018년 한국복지패널조사(가구용) :: 1

2.데이터 분석

: 변수

종속변수

자가 소유 유무 (house)		1인가구 여부 (one)	
0(자가0)	1(자가X)	0 (다인가구)	1(1인가구)
2721	3753	4380	2094

설명변수

연간 총 소득 (income)	재산 (asset)
평균 4183.76(만원)	평균 16368.24(만원)

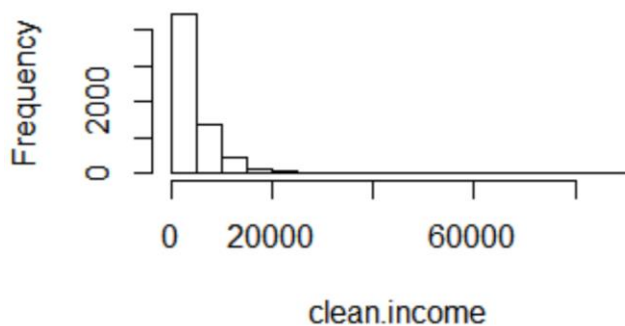
연령 (age.c)	성별 (sex.d)		대학졸업여부 (edu.d)		지역변수 (seoul.d)	
최빈값 40-60대	0(여)	1(남)	0	1(대학이상)	0(비서울)	1(서울)
	2188	4286	5288	1186	5590	884

2. 데이터 분석

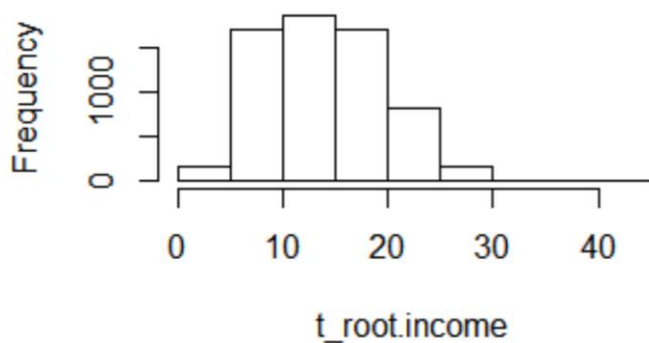
: 데이터 전처리 **income**

	h1308_114	h1308_122	h1308_aq2	h1308_160	h1308_166	h1308_172	h1308_aq9	h1308_aq10	h1308_aq11	h1308_116
1	NA	NA	NA	NA	NA	0	0	0	0	NA
2	NA	NA	NA	NA	NA	0	0	0	0	NA
3	NA	NA	NA	NA	NA	0	20	0	0	NA
4	NA	2054	NA	NA	NA	0	0	0	0	NA
5	NA	NA	3941	NA	NA	0	50	0	0	NA
6	NA	324	NA	NA	NA	0	0	0	0	NA
7	NA	NA	NA	NA	NA	0	0	0	0	NA
8	NA	NA	NA	NA	NA	0	70	360	0	NA
9	NA	NA	NA	NA	NA	0	0	0	0	NA
10	3868	NA	NA	NA	NA	0	0	0	0	NA
11	NA	NA	NA	NA	NA	0	0	950	0	NA
12	NA	NA	300	NA	NA	0	0	0	0	NA
13	NA	NA	NA	NA	NA	0	0	0	0	NA

Histogram of clean.income



Histogram of t_root.income



① income을 구성하는 문항들을 하나의 자료로 합침.
(연간 총 급여액, 임시일용자 총 급여액, 판매수입 등)

② 설문조사 특성 상 자신에게 해당하는 소득만을
체크하므로 결측치가 무조건 존재.

③ NA를 무시하고 모든 income을 더해서
clean.income을 만듦.

④ clean.income이 right-skewed한 분포를
보이기 때문에, 가장 정규분포에 가까운 자료인
 $(\text{clean.income})^{(1/3)}$ 을 용하기로 함.

> 최종 소득 변수 : t_root.income

2. 데이터 분석

: 데이터 전처리 **asset**

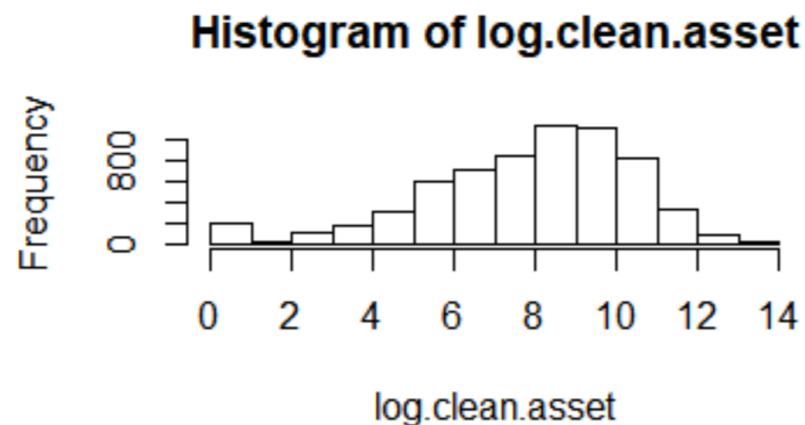
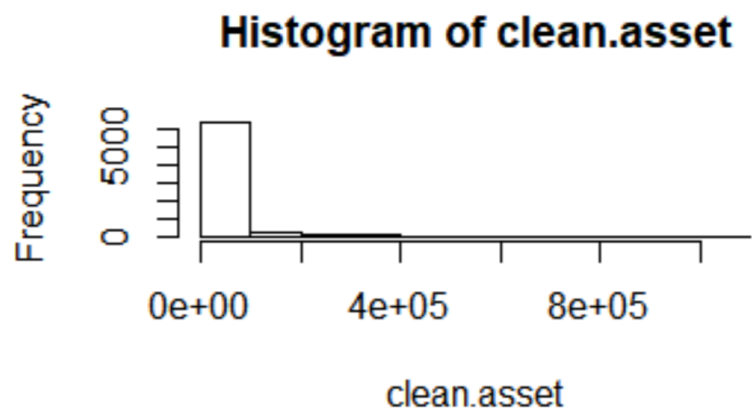
	h1310_aq1	h1310_aq2	h1310_aq3	h1310_aq4	h1310_aq5	h1310_aq6	h1310_aq7
1	0	0	0	0	0	200	0
2	0	0	0	0	0	500	0
3	0	0	0	0	0	2000	0
4	0	0	7000	0	0	50	0
5	0	0	0	0	0	4200	0
6	0	0	0	0	0	660	0
7	0	0	0	0	0	0	0
8	0	0	0	0	0	6000	0
9	0	0	0	0	0	0	0
10	0	0	0	0	0	600	1070

① asset을 구성하는 문항들을 하나의 자료로 합침.
(소유 부동산, 금융자산, 농축산물 등)

② clean.income과 동일한 결측값 처리

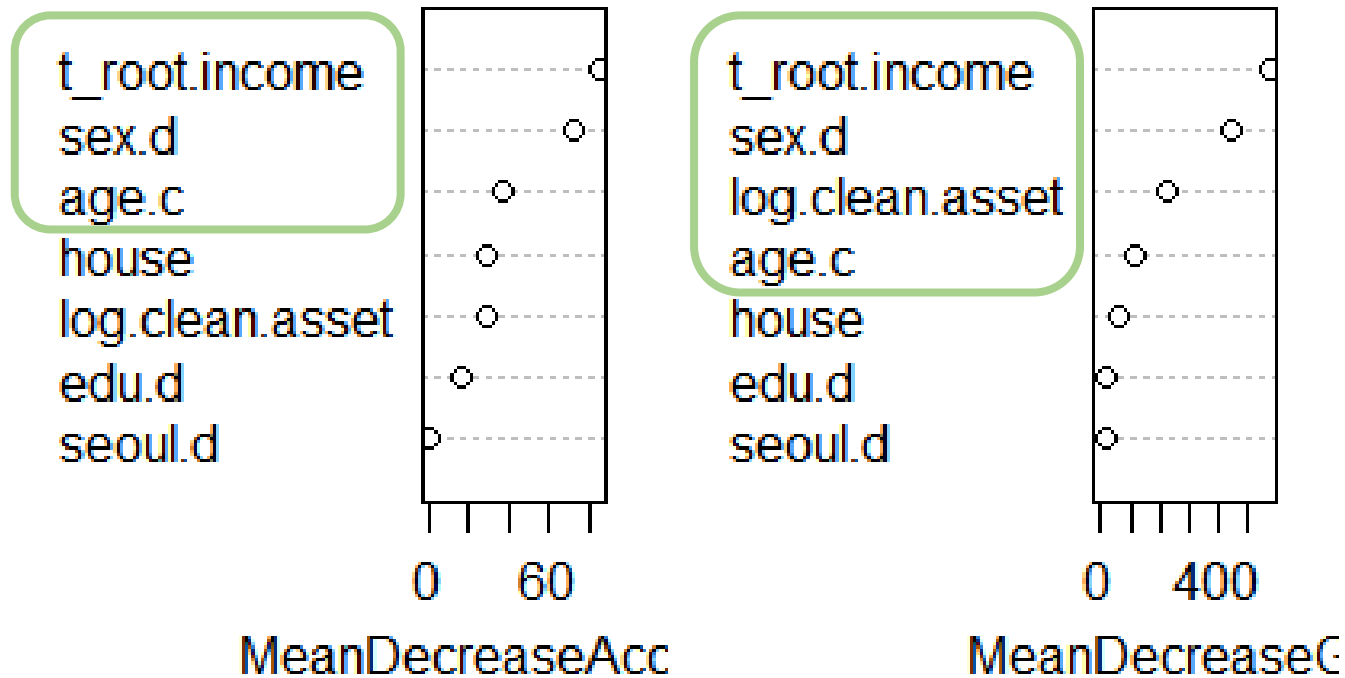
③ clean.asset이 right-skewed한 분포를 보이기 때문에, 로그를 씌워 정규분포에 가깝게 만들어 분석에 사용하기로 함.

> 최종 재산 변수 : **log.clean.asset**



2.데이터 분석 - (1) 랜덤포레스트

rf.fit1

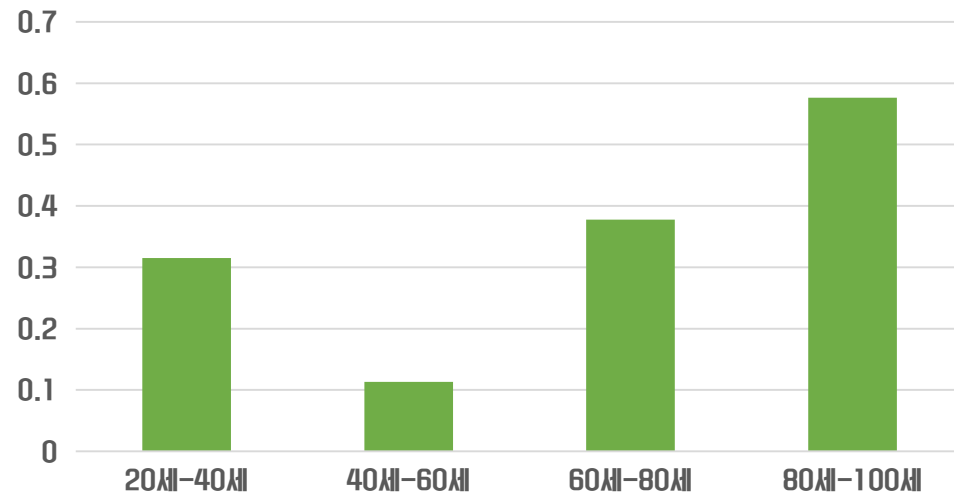


반응변수 : one (1인가구여부)

랜덤포레스트 결과, 중요변수는
Income , sex, age, asset 이다.

2.데이터 분석 - (2) 로지스틱회귀

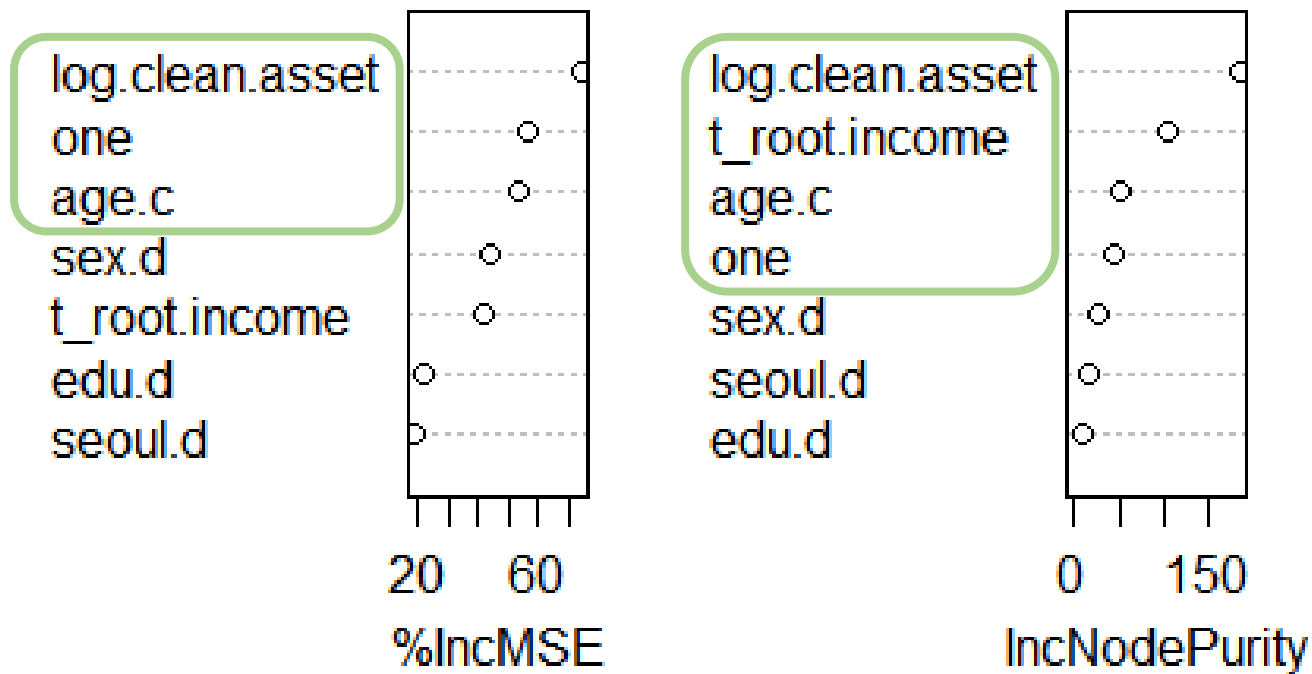
나이별 1인가구 비율



> 60세 이후의 노년층의 1인가구 비율이 상대적으로 높은 것을 확인 할 수 있다.

2.데이터 분석 - (1) 랜덤포레스트

rf.fit2



반응변수 : house(자가 소유 여부)

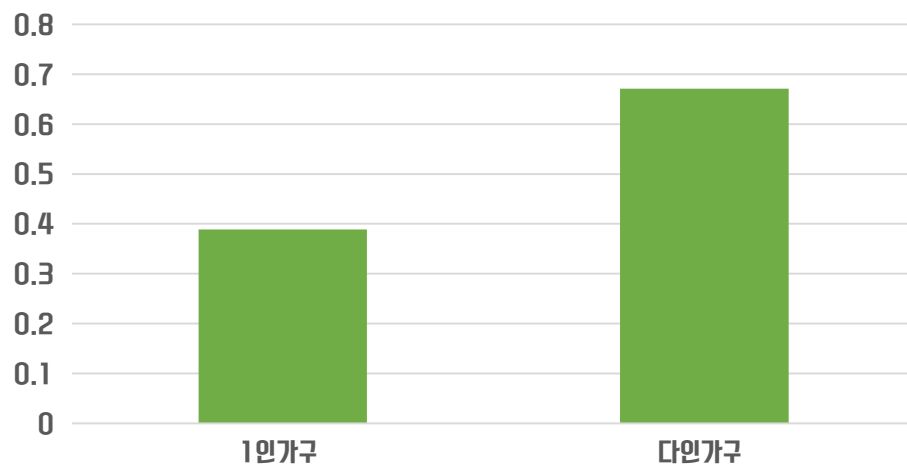
랜덤포레스트 결과, 중요변수는
asset, one, age, income이다.

2.데이터 분석 – (2) 로지스틱회귀

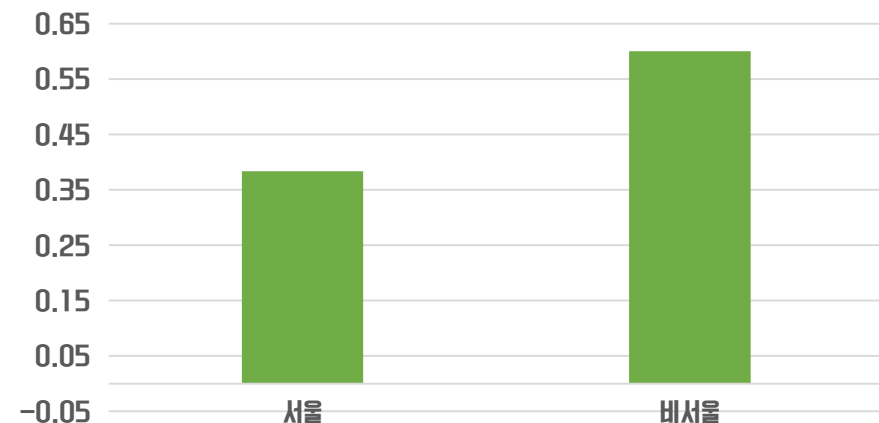
› logistic적합 결과,

자가 소유여부(house)에 1인가구/다인가구 여부, 나이, 성별, 서울/비서울 여부 변수가 유의한 영향을 줌을 알 수 있었다.

1인가구/다인가구 자가 소유 비율



서울/비서울 지역의 자가 소유 비율



그래프를 보면, 1인 가구일 수록, 서울지역에 거주할 수록 자가소유 비율이 낮다.

【 서울지역에 거주하는 60대 이상 노인을 대상으로 공공주택 최적입지를 분석하기로 결정 】

3.데이터 분석 - (2) 계층 군집분석

: 타겟 맞춤형 공공주택 입지 변수

노인 복지시설
및
교육시설 근접성

기존
노인 밀집 지역

의료시설과
산책로 등
근린시설 접근성

교통 편의성



3.데이터 분석 - (2) 계층 군집분석



: 타겟 맞춤형 공공주택 입지 변수

$$Z^1 (\text{독거노인밀집도}) = \frac{(\text{각 구 독거노인 수})}{(\text{각 구의 면적})}$$

$$Z^2 (\text{교통 접근성}) = (\text{각 구의 지하철 역 접근성})$$

$$Z^3 (\text{근린생활시설 접근성}) = 6.29 * \text{park} + 6.09 * \text{LQ}$$

$$\text{Park} = \frac{(\text{구별 공원 전체 면적})/(\text{구별 인구수})}{(\text{서울 공원 전체 면적})/(\text{서울 시 인구수})}, \text{LQ} = \frac{(\text{구별 의사수})/(\text{구별 고용자수})}{(\text{서울시 전체 의사수})/(\text{서울시 전체 고용자수})}$$

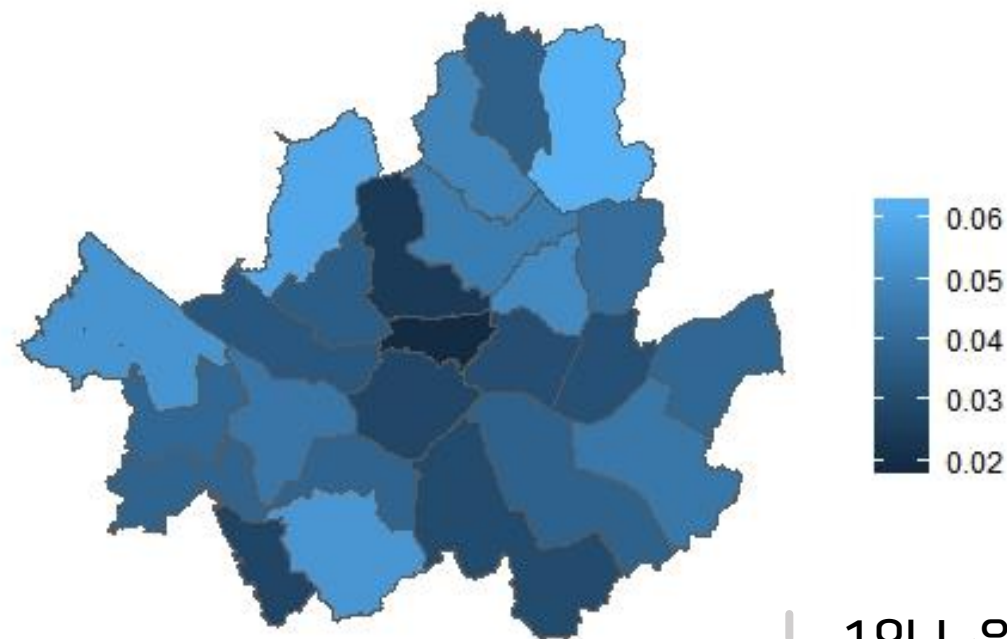
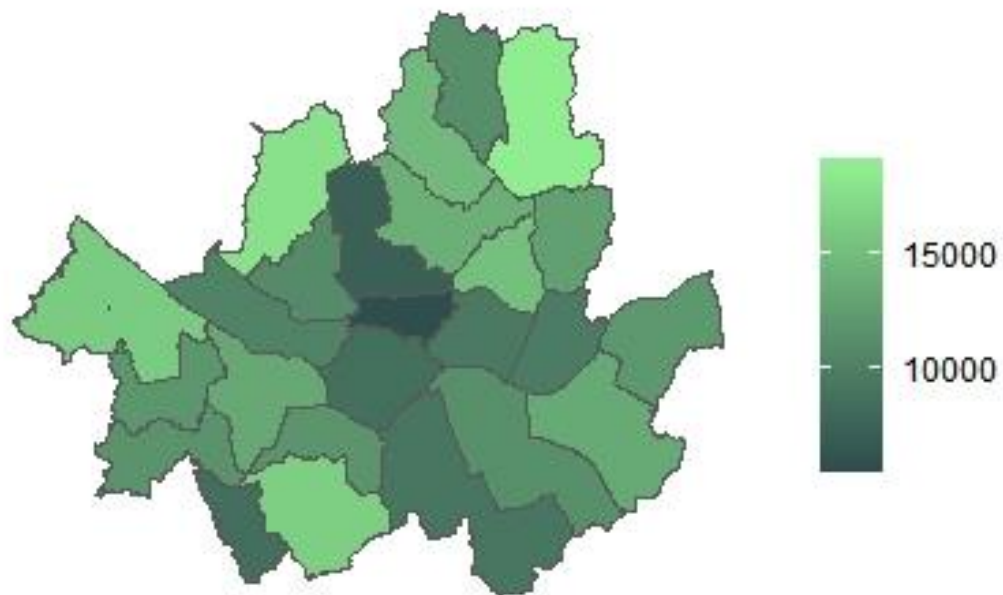
$$Z^4 (\text{사회활동 · 복지시설 근접성}) = 0.5 * \text{h.welfare} + 0.5 * \text{l.welfare}$$

$$\text{h.welfare} = \frac{(\text{구별 주거복지시설 전체 정원 수 합})/(\text{구별 노인 정원 수})}{(\text{서울시 주거복지시설 전체 정원 수 합})/(\text{서울 전체 노인수})}$$

$$\text{L.Welfare} = \frac{(\text{구별 여가복지시설 전체 정원 수 합})/(\text{구별 노인 정원수})}{(\text{서울시 여가복지시설 전체 정원 수})/(\text{서울시 전체 노인수})}$$

3.데이터 분석 – (2) 계층 군집분석

: 지도로 보는 EDA

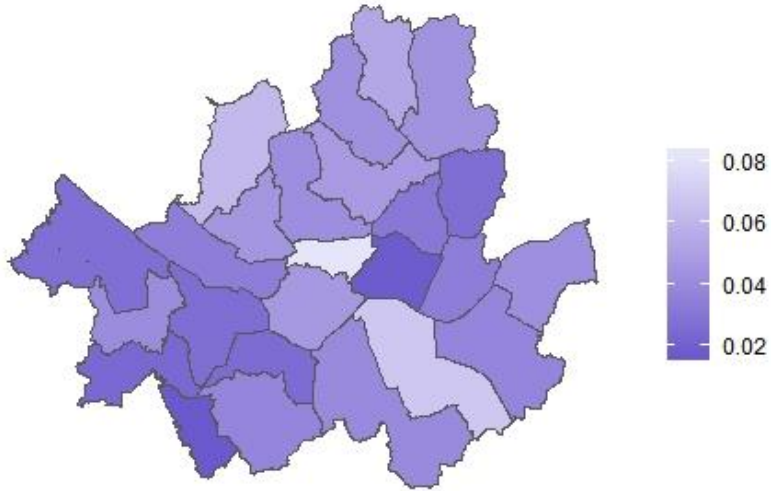


- 1위 노원구
- 2위 은평구
- 3위 관악구
- 4위 강서구
- 5위 동대문구

〉 독거노인 가구가 가장 많은 지역은 노원구이며,
단위 면적 당 독거노인 수가 가장 많은 지역도 노원구이다.

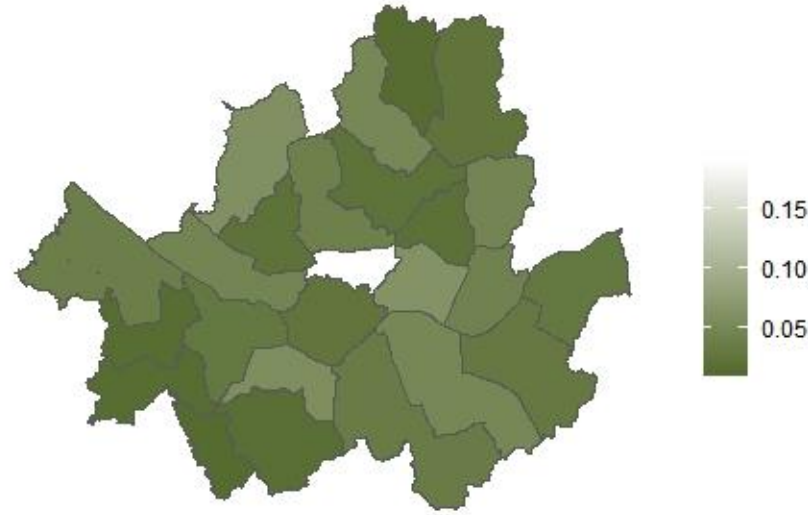
3.데이터 분석 - (2) 계층 군집분석

: 지도로 보는 EDA



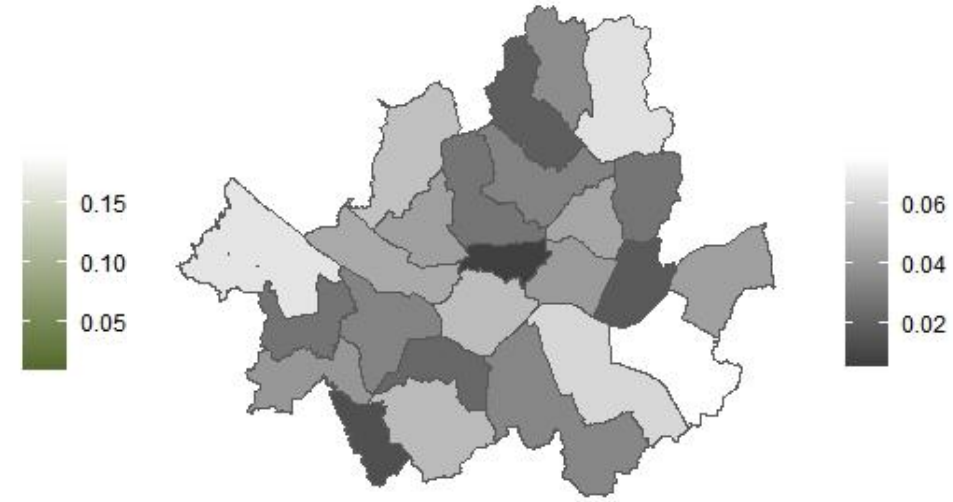
구별 근린시설 밀집도

- 1위 중구
- 2위 강남구
- 3위 은평구
- 4위 도봉구
- 5위 성북구



구별 대중교통 접근 편의성

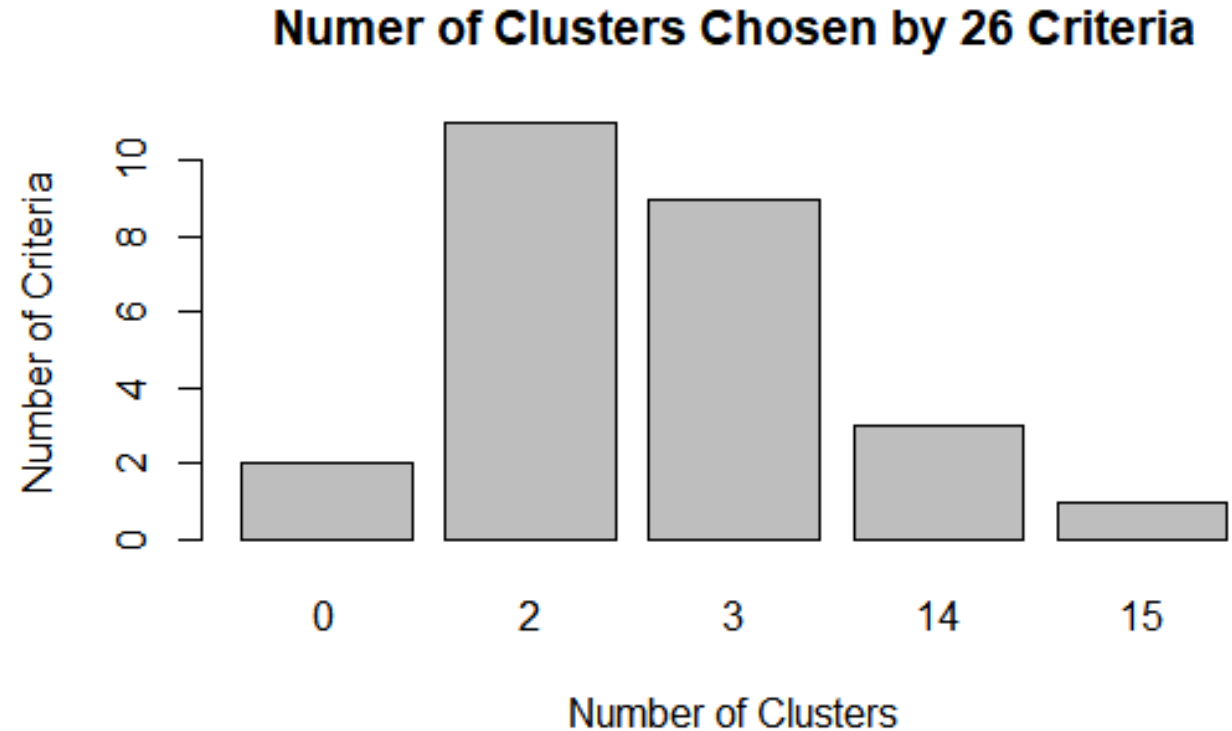
- 1위 중구
- 2위 성동구
- 3위 은평구
- 4위 동작구
- 5위 강남구



구별 노인복지 시설 밀집도

- 1위 송파구
- 2위 강서구
- 3위 노원구
- 4위 강남구
- 5위 은평구

3.데이터 분석 – (2) 계층 군집분석

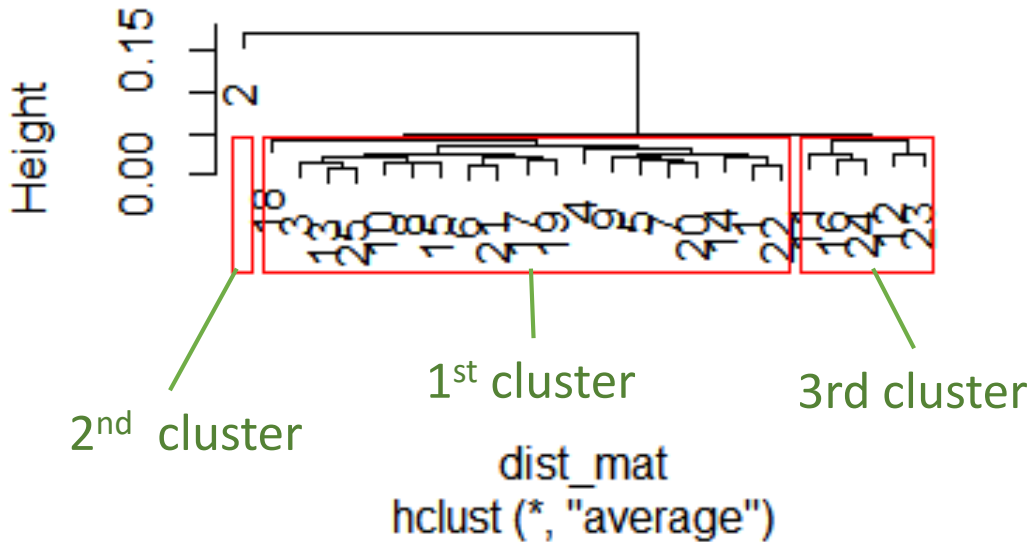


＞ 25개의 구를 2가지 군집으로 분류하는 것보다 3개로 분류하는 것이 지역별 차이를 보기 적합할 것이라고 판단하여 **3개**로 군집분석을 시행하였다.

3.데이터 분석 – (2) 계층 군집분석

: 군집분석 결과

Cluster Dendrogram



1st cluster

종로구 용산구 성동구 광진구 동대문구 중랑구 성북구
강북구 도봉구 서대문구 마포구 양천구 구로구 금천구
영등포구 동작구 관악구 서초구 강동구

2nd cluster

중구

3rd cluster

노원구 강서구 송파구 은평구 강남구

3.데이터 분석 - 분석결과



	지역	노인인구 밀집도	교통 접근성	근린 시설	사회 활동 · 복지	특성
c1	종로구 용산구 성동구 광진구 동대문구 중랑구 성북구 강북구 도봉구 서대문구 마포구 양천구 구로구 금천구 영등포구 동작구 관악구 서초구 강동구	0.0384	0.0293	0.0379	0.0347	교통 접근성이 상대적으로 취약하며, 근린시설이 가장 적은 지역
c2	중구	0.0188	0.1906	0.0824	0.0071	교통접근성, 근린시설 접근성은 매우 좋으나 노인들을 위한 복지시설 접근 성은 비교적 취약한 지역.
c3	노원구 강서구 송파구 은평구 강남구	0.0534	0.0381	0.0444	0.0658	기존 독거노인층 밀집도가 가장 높으 며, 사회활동, 복지시설이 가장 잘 갖추 어져 있는 지역

3.데이터 분석 – 분석결과 해석

: 주거 관련 변수 선호도

논문참고 : 이준민,신화경(2005)

	선호도	선호도 가중치 계산
교통접근성	5.87	$5.87/(6.18+5.87+4.97)=\mathbf{0.35}$
근린시설(공원, 병원 등)	6.18	$6.18/(6.18+5.87+4.97)=\mathbf{0.36}$
사회활동 · 복지	4.97	$4.97/(6.18+5.87+4.97)=\mathbf{0.29}$



선호도 가중치로 각 군집을 계산한 결과

$$c1 = 0.033876$$

$$c2 = 0.099515$$

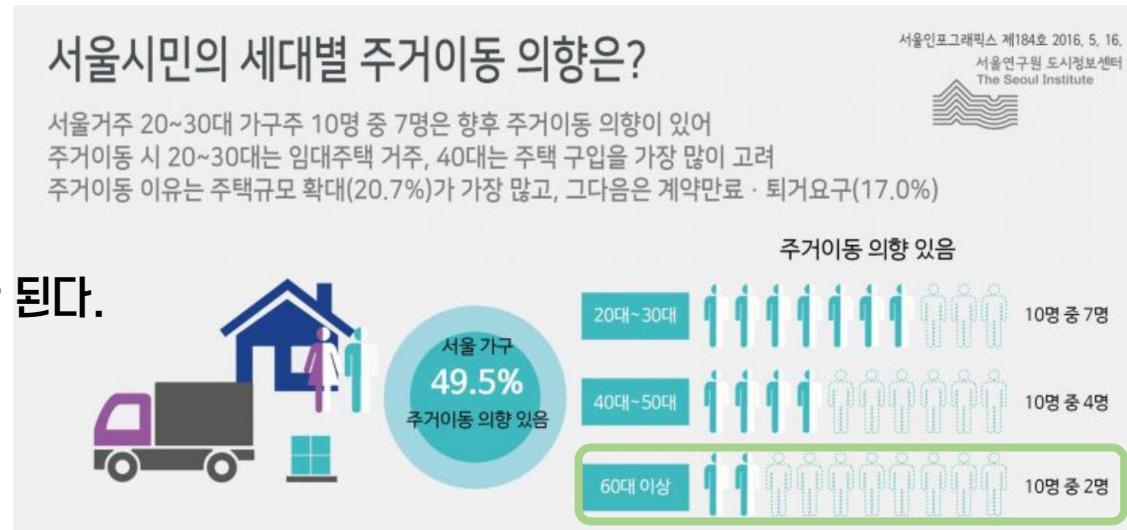
$$c3 = 0.048338$$

👉 $c2 > c3 > c1$ 순으로 높다.

3.데이터 분석 – 분석결과 해석

: 독거노인 인구 밀집도 고려

- 〉 환경적 변수만을 생각했을 때 중구가 교통 접근성 등 가장 좋은 입지가 된다.
그러나 노인들의 주거이동을 선호하지않는 경향성을 고려하였을 때
독거노인 인구밀집도에 가중치를 두어서 고려해야한다.



$$[0.2 * (\text{주거관련 선호도 고려한 수치}) + 0.8 * (\text{독거노인 인구밀집도})]$$

	주거관련 선호도 고려한 수치	독거노인 인구밀집도	최종
C1	0.033876	0.0384	$0.033876 * 0.2 + 0.0384 * 0.8 =$ 0.0374952
C2	0.099515	0.0188	$0.099515 * 0.2 + 0.0188 * 0.8 =$ 0.034943
C3	0.048338	0.0534	$0.048338 * 0.2 + 0.0534 * 0.8 =$ 0.0523876

선호도만 고려했을 때는 c2>c3>c1이었으나, 독거노인 인구밀집도까지 포함하면 최종적으로 c3>c1>c2 순서가 된다.



**1인가구 중 서울지역에 거주하는
60대 이상 노인을 위한 공공주택 최적 입지는
‘노원구’, ‘강서구’, ‘송파구’, ‘은평구’, ‘강남구’ 이다.**

4. 기대효과 및 한계

기대효과

독거노인 관련 정책에 활용

: 노인들의 선호도를 만족시킬 수 있는 환경적 변수를 고려하고, 노인 밀집도를 고려하였으므로 고령화 문제에 대비한 공공주택의 건설에 이용할 수 있을 것으로 보인다.

한계

1. 데이터의 한계

- 한국복지패널 데이터에서 1인 가구, 20세 이상으로 제한하면 표본이 2000여명으로 줄어든다.
- 분석 시 2018년의 횡단면자료만 사용했다. 추후 과거의 데이터를 포함하여 추가적인 분석을 한다면 좀더 정확한 결과를 낼 것으로 기대된다.
- 다인 가구의 경우, 가구원 중 한 명이 주택을 소유하고 있어도 자가가 있는 것으로 조사 응답을 하므로 소득과 자가소유 여부의 괴리가 있었다.

2. 정책성 한계

- 각 구의 환경적 요인으로 공공주택의 설립이 어려울 수 있음.

5. 참고문헌 및 출처

* 참고 문헌

김흥순, 정다운 (2010). 서울시 의료시설의 공간적 분포특성에 관한 연구. 도시행정학보, 23(1), 27-45

이준민, 신화경 (2005). 중년층이 선호하는 노인주거환경에 관한 연구. 자연과학연구, 14(0), 1-15

김진원, 윤병국, 윤성식 (2014). 공간통계기법을 이용한 도시공원 분포의 공간적 패턴탐색 연구-서울시를 사례로-, 관광학연구, 38(6), 209-232

신은진, 안건혁 (2010). 소득별 1인가구의 거주지 선택에 영향을 미치는 요인에 대한 연구. 국토계획, 45(4), 69-79

김옥연, 문영기 (2011). 가계 자산구조와 주택소유결정요인 분석. 주거환경, 9(1), 25-38

* 데이터 출처

13차 한국 복지패널 조사 자료

: 한국복지패널 (<https://www.koweps.re.kr:442>)

구별 독거노인 현황

구별 노인주거복지시설 현황

구별 노인복지관, 경로당 수

구별 의료기관 통계

구별 보건소 통계

구별 1인당 공원 면적 통계

구별 공원 수 통계

: 서울시 열린 데이터 광장 (<https://data.seoul.go.kr/>)

서울시민의 세대 이동 영향

: 서울연구원 (<https://www.si.re.kr/node/55133>)

