**데이터기반 행정으로 국민의 삶의 질을 개선하라!**

**데이턴십 해커톤 제 4회**

|  |
| --- |
| **부산시 노인 전문 체육시설**  **최적 입지 선정을 통한**  **노인 친화적 행정 실현** |

분석 결과보고서

|  |  |  |
| --- | --- | --- |
| 참여 조 : |  | 117조 |
| 참여자 : |  | 김경부 (조장) |
|  |  | 김민경 |
|  |  | 공다애 |
|  |  | 나다경 |
|  |  | 박병진 |
|  |  | 장현범 |

**목 차**

**1. 분석 개요 03**

**1.1. 분석 배경 및 개요 .............................................................. 03**

**1.2. 분석 목적 및 방향 .............................................................. 05**

**1.3. 분석 결과 활용 방안 ........................................................... 06**

**2. 분석 데이터 08**

**2.1. 분석 데이터 목록 ............................................................... 08**

**2.2. 데이터 상세 설명 ............................................................... 09**

**2.3. 데이터 정제 방안 ............................................................... 11**

**3. 분석 프로세스 14**

**3.1. 분석 프로세스 ................................................................... 14**

**3.2. 분석 내용 및 방법 .............................................................. 14**

**4. 분석결과 18**

**4.1. 회귀분석을 활용한 결과 ...................................................... 18**

**4.2. AHP 쌍대비교 분석을 통한 가중치 선정 ................................. 30**

**4.3. Q-GIS를 활용한 분석 ........................................................ 32**

**4.4. 결론 ................................................................................ 38**

**4.5. 한계 및 제언 ..................................................................... 43**

**5. 활용 방안 45**

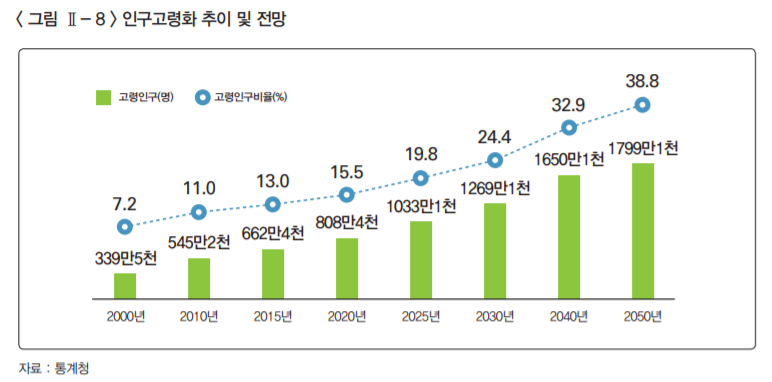
**5.1. 문제점 개선 방안 ............................................................... 45**

**5.2. 업무 활용 방안 .................................................................. 47**

**6. 부록 및 출처 51**

1. **분석 개요**
   1. **분석 배경 및 개요**

* 노인만을 위한 노인 체육 시설의 부족
* 2021년 6월 부산 기준 노인 인구는 전체 인구의 19.8%로 초고령 사회를 앞두고 있으며 2040년에는 노인 인구가 전체 인구의 32.9%에 달할 것으로 예상됨



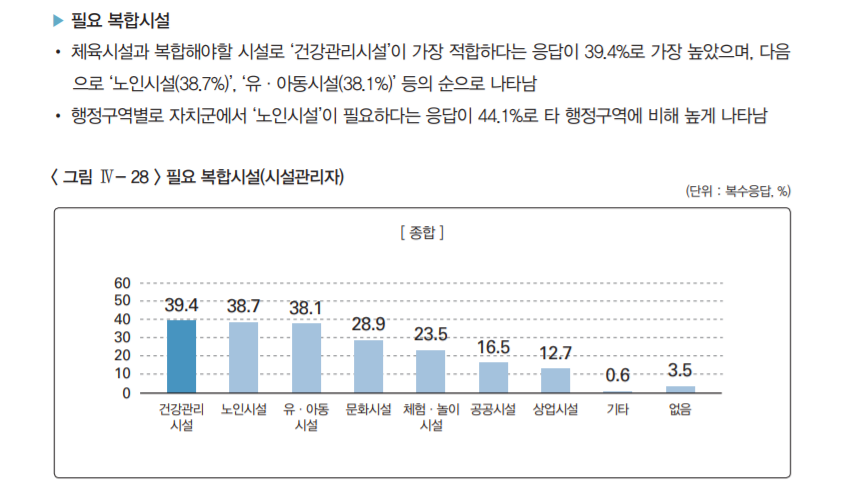
*그림 1고령화인구 추이 및 통계*

\* 고령화 사회 : 총 인구 중 65세 이상 인구가 차지하는 비율이 7% 이상인 사회

\* 고령 사회 : 총인구 중 65세 이상 인구가 차지하는 비율이 14% 이상인 사회

\* 초고령 사회 : 총인구 중 65세 이상 인구가 차지하는 비율이 20% 이상인 사회

* 노인 인구 증가에 따라 노인들의 건강에 대한 관심도가 높아지며, 노인 전문 체육시설의 수요도 증가할 것으로 예상됨
* 70세 이상 노년층도 걷기(58.1%), 체조(17.4%), 헬스(4.4%), 수영(3.6%)을 중심으로 체육활동이 이루어지고 있어 향후 노인 인구 증가에 대비한 충분한 시설이 필요함

**

*그림 2 필요 복합시설(시설관리자)*

* 체육시설과 복합해야 할 시설로 ‘건강관리시설’이 가장 적합하다는 응답이 39.4%로 가장 높았으며, 다음으로 ‘노인시설(38.7%)’, ‘유·아동시설(38.1%)’ 등의 순으로 나타남

* 현재 공공 체육 시설 조사 결과, 게이트볼장을 제외하고 노인만을 위한 노인 체육 시설이 따로 존재하지 않음.
* 노인 체육 정책의 모호성 및 저조한 실행률
* 국민 체육 진흥법 제10조 2에 따라 국가는 노인 체육 활성화에 필요한 시책을 마련하고 시설 및 비용을 지원해야 하나

\*국민 체육 진흥법 제 10조의2(노인 체육의 진흥)

① 국가와 지방자치 단체는 노인 체육 진흥에 필요한 시책을 마련하여야 한다.

② 국가와 지방자치단체는 노인 건강의 유지 및 증진을 위한 맞춤 체육 활동 프로그램을 운영하거나 그 운영에 필요한 비용 및 시설을 지원할 수 있다.

[본조신설 2020.06.09]

* 현재 노인체육 정책을 협소한 틀로 규정한 데다가 정책을 구체적으로 규율하기에도 미비한 상황
* 노인체육 정책을 문화체육관광부와 보건복지부를 중심으로 하고 있지만, 단체들이 협력하고 협업하기보다는 행정편의주의적으로 정책을 집행해 기존 정책을 반복하는 형태가 지속하고 있음
  1. **분석 목적 및 방향**

선행 연구로부터 공공 기관 입지 결정 과정에 영향을 미치는 주요 요인 선정

노인 시설의 수에 영향을 미치는 인자들에 대한 다중 선형 회귀분석 실시

유의한 변수를 대상으로 노인 전문 체육시설의 입지 선정을 위한 예측 모델 개발

분석된 데이터를 활용하여 Q-GIS 속성결합, 가중치 부여,

정규화, 시각화 과정을 거쳐 모델 실행

모델 평가 및 수정

최종적으로 기존 공공 체육 시설과 예측 모델로 선정된 입지를 비교하여 현실적으로 최적화된 입지를 선정

* 1. **분석 활용 방안**
* 차별 없는 공정한 체육 서비스 제공 가능
* 2018년부터 2021년 동안 노인 취업률은 큰 변화는 없는 상황에서 체육 활동을 하기 위한 비용이 발생한다면 체육 서비스에 격차가 생길 것으로 예상됨(헬스 1달 평균 비용 142,000). 노인만을 위한 체육 시설을 구축한다면 모두에게 공평한 체육 서비스를 제공할 수 있을 것으로 예상됨
* 적극적인 노인 복지를 추진 고령 친화도시 이미지 구축 가능
* 고령 친화 도시는 노인 세대의 건강과 사회 참여, 사회 보장에 관한 기회를 최적화하여 활기찬 고령화(Active Ageing)를 촉진하는 도시 환경을 갖춘 도시를 말함.
* 부산은 우리나라에서 서울시와 전라북도 정읍시, 경기도 수원에 이어 국내 도시 중 4번째 고령 도시로 선정됨. 노인만의 체육 시설 설립은 WHO에서 규정하는 '모든 연령대에서 나이가 들어감에 따라 자신들의 삶의 질을 향상할 수 있도록 ‘건강·참여·안전’에 대한 기회가 최적화된 ‘활기찬 고령화(Active Ageing)’를 추구할 수 있는 기반을 다질 수 있을 것
* 노인의 이용률 증가로 노인 건강에 긍정적인 효과를 기대
* 노인들은 건강 증진을 비롯한 신체 기능 상승효과를 위한 체육 활동이 필수적임. 노인의 체육 활동을 장려하여 병원 진료비 감소, 조기 사망 감소, 심혈관 질환 감소, 대사질환 감소, 근 골격 건강 증진, 정신 건강 증진 등의 효과를 기대할 수 있음

1. **분석 데이터**
   1. **분석 데이터 목록**

|  |  |  |  |
| --- | --- | --- | --- |
| 구분 | 분석데이터 | 기간 | 제공기관 |
| 인  구 | 부산시 구·군별  세대 및 인구 | 1995~  2018 | 통계청  (<https://kostat.go.kr/portal/korea/index.action>) |
| 접  근  성 | 부산광역시\_  버스 정류소 정보(SHP) | 2021 | 부산통계 : 공공데이터포털  (<https://data.busan.go.kr/index.nm>) |
| 공원 | 1995~ 2019 | 통계청  (<https://kostat.go.kr/portal/korea/index.action>) |
| 2021년 1/4분기  부산시 의료기관 현황 | 2021 | 부산광역시  (<https://www.busan.go.kr/index>) |
| 공공포털\_상가\_상권\_정보  \_2020-09 | 2020.  09 | 부산광역시 빅데이터 포털  (<https://bigdata.busan.go.kr/>) |
| 노인여가 복지시설 | 2005~ 2011 | 통계청  (<https://kostat.go.kr/portal/korea/index.action>) |

* 1. **데이터 상세 목록**

|  |  |  |  |
| --- | --- | --- | --- |
| 구분 | 분석데이터 | 데이터 형식 | 생성주기 |
| 공  공  데  이  터 | 부산시 구·군별 세대 및 인구 | XLSX | 연간 |
| 부산광역시\_버스 정류소 정보(SHP) | SHP | 연간 |
| 공원 | XLSX | 연간 |
| 2021년 1/4분기 부산시 의료기관 현황 | XLSX | 분기별 |
| 공공포털\_상가\_상권\_정보  \_2020-09 | TXT | 수시(특정주기 없음) |
| 노인여가 복지시설 | XLSX | 연간 |

* 부산시 구·군별 세대 및 인구[.XLSX]
* 통계청에서 제공하는 자료로 세대 수, 인구, 65세 이상 고령자 수 등의 내용이 포함되어 있음
* 부산광역시\_버스 정류소 정보[.SHP]
* 부산통계: 공공데이터 포털에서 제공하는 자료로 정류소 ID, 정류소 명, GPS좌표, 정류소 구분 등의 데이터를 담고 있음. 이 중 노인 전문 체육시설의 입지 선정을 위해 정류소 명과 GPS좌표를 선택, 동별 정류소 개수를 도출함
* 공원 [.XLSX]
* 통계청에서 제공하는 자료로 자연공원 그리고 어린이공원과 소공원, 근린공원 등을 다 아우르는 도시공원의 정보가 담긴 데이터
* 2021년 1/4분기 부산시 의료기관 현황[.XLSX]
* 부산광역시에서 제공하는 데이터로, 구별 의료기관 정보가 담긴 데이터

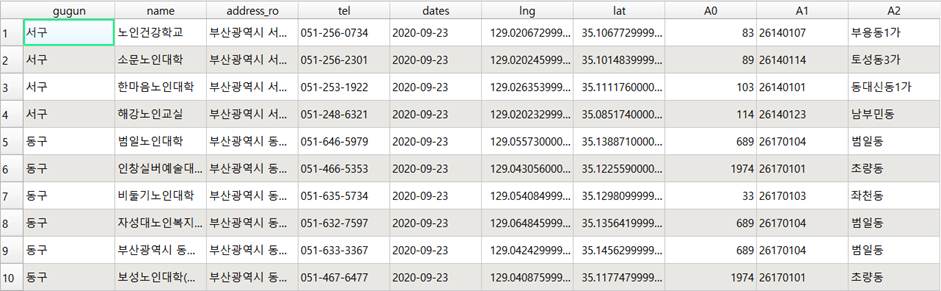
* 공공 포털\_상가\_상권\_정보\_2020-09[.TXT]
* 부산광역시 빅데이터 포털에서 제공하는 데이터로 부산 내의 상권 정보를 담고 있음. TXT파일은 가독성이 떨어지는 관계로 XLSX로 변환하여 사용함
* 노인 여가 복지시설[.XLSX]
* 통계청에서 제공하는 데이터로 부산시 구·군별 노인복지관, 경로당, 노인 교실, 노인휴양소 정보를 제공하고 있음. 이 중 노인복지관, 경로당, 노인 교실의 데이터를 사용함
  1. **데이터 정제 방안**
* 분석 배경 및 개요
* 노인 전문 체육시설 입지 선정에 영향을 미치는 요인을 확인하기 위해 회귀 분석을 사용하기로 계획함
* 초기 확보 데이터는 16개의 행을 가진 구별 시설 및 인수 수 데이터로, 행의 수가 회귀분석을 진행하기에 부족하다고 판단함. 본 문제 해결을 위해 Q-GIS 프로그램을 이용해 동 기준 데이터로 변환하는 방법을 사용하기로 함. Q-GIS 기준 부산의 동은 총 193개로 회귀분석을 진행하기에 충분함
* 국가 공간 정보 포털에서 제공하는 부산 지역 행정구역\_읍면동 (법정동) 경계도를 사용함

지도이(가) 표시된 사진

자동 생성된 설명

*그림 3 부산시 행정구역 읍면동(법정동) 경계도면*

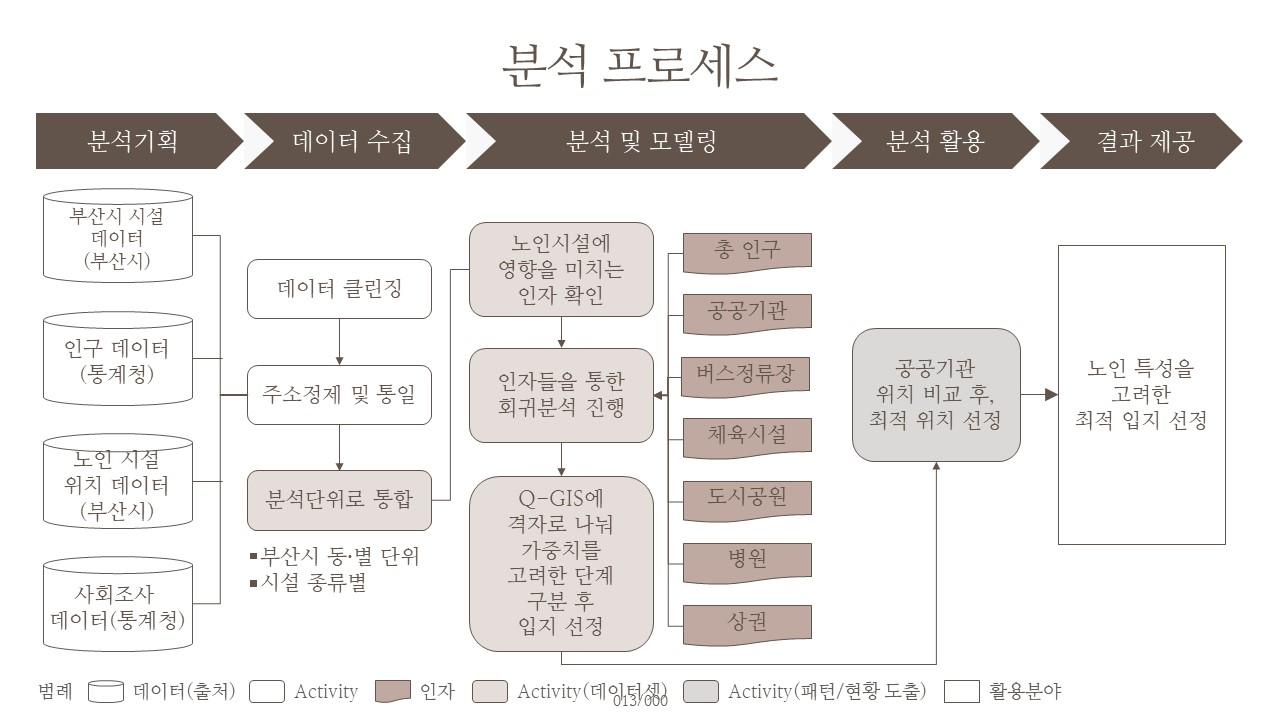
* 첫째, 좌표를 이용해 CSV 파일을 Q-GIS의 점 레이어로 불러옴. 둘째, 부산시 동 파일을 폴리곤 레이어로 불러옴. 셋째, 점 레이어와 폴리곤 레이어를 ‘위치에 따라 속성 결합’의 ‘교차’ 조건으로 결합함. 넷째, 결합한 파일을 CSV로 내보냄. 네 단계를 반복해 최종적으로 경로당, 공공기관, 버스정류장, 노인 교육 시설, 노인 요양 병원, 노인 복지관, 도시공원, 병원, 체육시설의 9개 변수에 동 정보를 입력함. <그림2>의 A2를 통해 각 구에 속한 동에 대한 정보를 알 수 있음



*그림 4 각 시설별 qgis를 이용하여 각 시설별 동 추출*

* 동을 기준으로 각 시설의 개수를 세기 위해 Q-GIS와 엑셀을 함께 활용함. 부산시 동 데이터 레이어 위에 9개 데이터를 점 레이어로 띄움. 다음 ‘폴리곤에 포함하는 포인트 개수 계산’을 통해 각 동에 있는 시설의 수를 추출함. 동별 시설 수를 CSV로 내보내기 하여 엑셀 파일로 만들고, 동 이름과 개수 정보만 모아 하나의 파일로 취합함. 본 파일을 사용하여 회귀분석을 실시함

1. **분석 프로세스**
   1. **분석 프로세스**



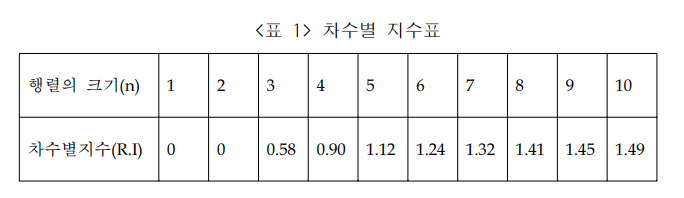
*그림 5 분석 프로세스*

* 1. **분석 내용 및 방법**
     1. 입지요인 선정을 위한 회귀분석
* 노인 전문 공공체육시설 입지 선정을 위해 고려해야 요인들을 알기 위해 회귀분석을 진행함. 총 7개의 독립변수를 가지고 현재 노인시설 입지에 영향을 미치는 요인을 파악해 공공체육시설에 접목하고, 다중 회귀분석을 이용해 분석을 진행하여 총인구, 공공기관, 버스정류장 변수가 속한 최종 모델을 구축함
  + 1. AHP 분석을 통한 가중치 선정
* AHP의 개념
* AHP(Analytic Hierarchy Process)는 상충하는 복수의 기준이 존재하는 다기준 의사결정(Multi-criteria Decision Making : MCDM)의 한 종류
* AHP는 복잡한 문제를 계층화하고 쌍별 비교함으로써 요인들의 중요도를 도출하여 의사결정을 진행함. 정량적, 정성적 정보도 비교할 수 있다는 점, 일관성 검증을 통해 신뢰도를 높이는 점, 그리고 그룹 의사 결정이 가능한 점 등의 장점이 있음
* AHP 의사 결정 절차
* 1단계 : 의사 결정 요소 간의 관계를 분석하여 의사 결정 계층구조 작성
* 2단계 : 각 계층 내의 의사결정 요소 간의 1:1 쌍별 비교를 통하여 쌍별 비교 행렬 산
* 3단계 : 고유치 방법(Eigenvalue Method)을 이용하여 쌍별 비교된 의사결정 요소 간의 상대적 가중치를 계산
* 4 단계 : 일관성 검증
* 가중치 설정과 일관성 검증
* 가중치 설정

요소 간의 가중치 설정을 위하여 두 요소씩 짝을 지어 상대적 중요도를 측정하는 쌍별 비교(pairwise comparison) 방법을 사용함. 이는 두 요소에 대한 상대적 가중치(relative importance)를 부여함으로써 의사결정에서의 유연성이 가중됨

* 일관성 검증

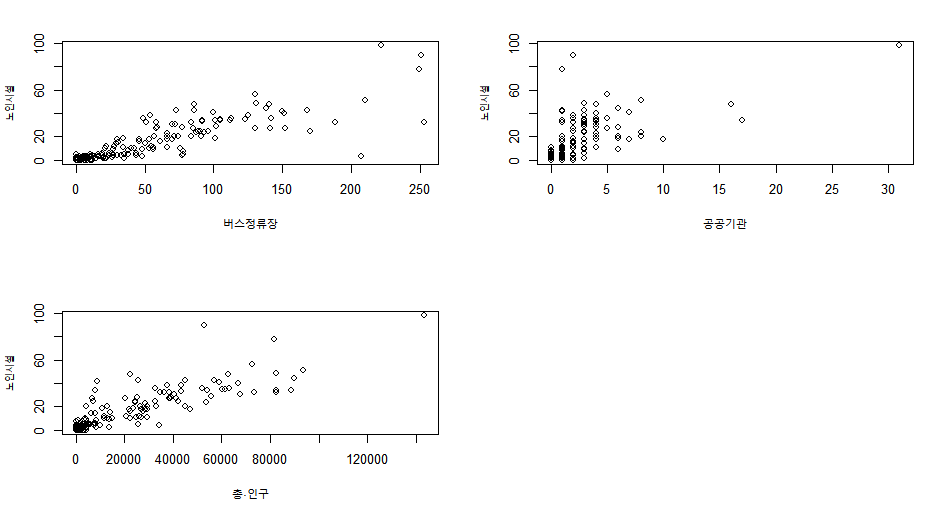
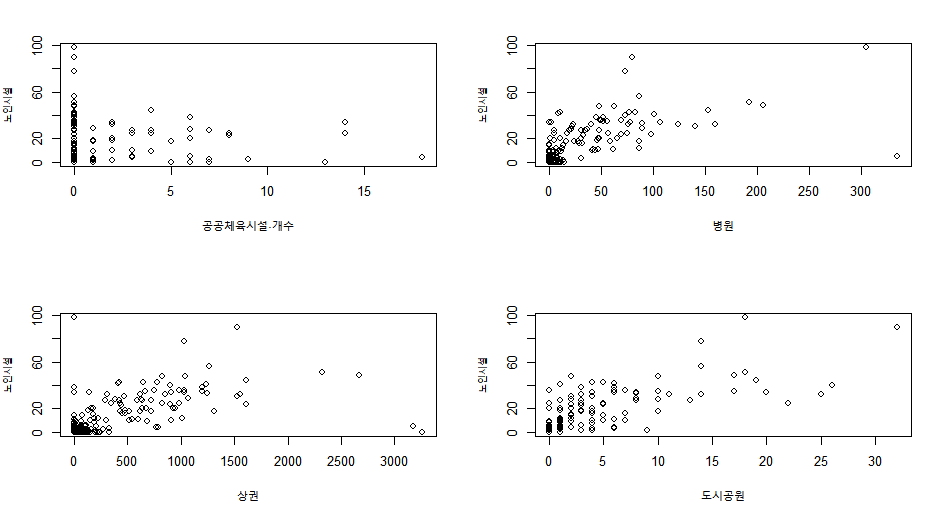
는 와 해당하는 차수의 의 비율을 일관성 비율 비로 정의하였으며이 이하이면 논리적으로 일관성이 있는 것으로 판단하여 사용 가능



*그림 6 차수별 지수표*

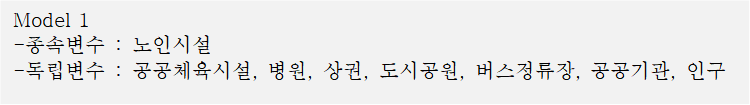
* AHP 분석의 의의
* 가중치를 결정하는 과정에 의사결정자가 다른 인자들을 고려하지 않고 오직 두 인자만을 쌍별 비교함으로써 AHP의 "상대적 가중치"를 도입한다면 의사결정자의 주관적 판단에 따른 오류를 범하는 것을 최소화할 수 있음
* ‘노인 전문 체육시설 최적 입지 선정’을 위해 회귀분석을 통해 선정된 변수에 “상대적 가중치” 도입을 통해 객관성을 부여함
  + 1. Q-GIS를 이용한 최종 입지 선정
* 3.2.1을 통해 선별된 변수들을 하나로 합친 후 Q-GIS의 단계 구분을 활용해 적합한 입지의 우선순위를 구별함. 변수 간 크기 차이를 최소화하고 각 변수의 중요도를 고려하기 위해 변수에 7개의 등급을 부여하고 가중치를 곱함
* 부산광역시의 격자 레이어와 속성 결합 후 단계 구분으로 시각화 함. 부산광역시의 공공 체육시설 데이터를 점 레이어로 불러온 후 두 레이어를 결합함. 부산의 동 이름과 구 이름이 포함된 데이터를 불러와 레이어와 결합함. 공공 체육시설을 중심으로 우선순위를 선정함

1. **분석 결과**
   1. **회귀분석을 활용한 결과**

* Summary
* 분석 대상인 데이터의 구조는 총 13개의 변수를 가지고 있으며 192개의 행이 있음. 데이터 타입은 동 이름은 character 타입이고 나머지 변수들은 모두 integer 타입임. 노인시설은 노인 교실, 경로당, 노인복지관을 모두 합한 새로운 column임 (부록 참조)
* 노인 시설과 각 변수 간의 연관성을 보기 위해 plot을 그린 결과, 다음과 같음

*그림 7 노인 시설과 각 변수 간의 연관성*

* Model Fitting



**Model 1**

**-종속변수 : 노인시설**

**-독립변수 : 공공체육시설, 병원, 상권, 도시공원, 버스 정류장, 공공기관, 인구**

텍스트이(가) 표시된 사진

자동 생성된 설명

*그림 8 모델1 회귀분석 결과*

* Model Selection
* 유의한 변수를 선택하기 위해 단계분석법을 사용하였음. 여러 독립변수 중에 다소 높은 설명력을 가지는 변수로 회귀모델을 구성하고자 함

텍스트, 신문, 영수증이(가) 표시된 사진

자동 생성된 설명

*그림 9 Model Selection*

* 대부분 변수들은 노인시설과 유의한 상관을 보이나, 공공체육시설이 0.0154로 상관이 없다고 판단되어 독립변수에서는 제외하고 분석을 진행함

**Model 2**

**모델 2는 상관성이 낮은 공공체육시설을 제외한 후 분석함**



텍스트이(가) 표시된 사진

자동 생성된 설명

*그림 10 모델2 회귀분석 결과*

텍스트이(가) 표시된 사진

자동 생성된 설명

*그림 11 Model 2*

* 단계선택 결과 버스정류장, 공공기관, 총인구가 독립변수로서 모델을 설명하는 것으로 나타남

텍스트이(가) 표시된 사진

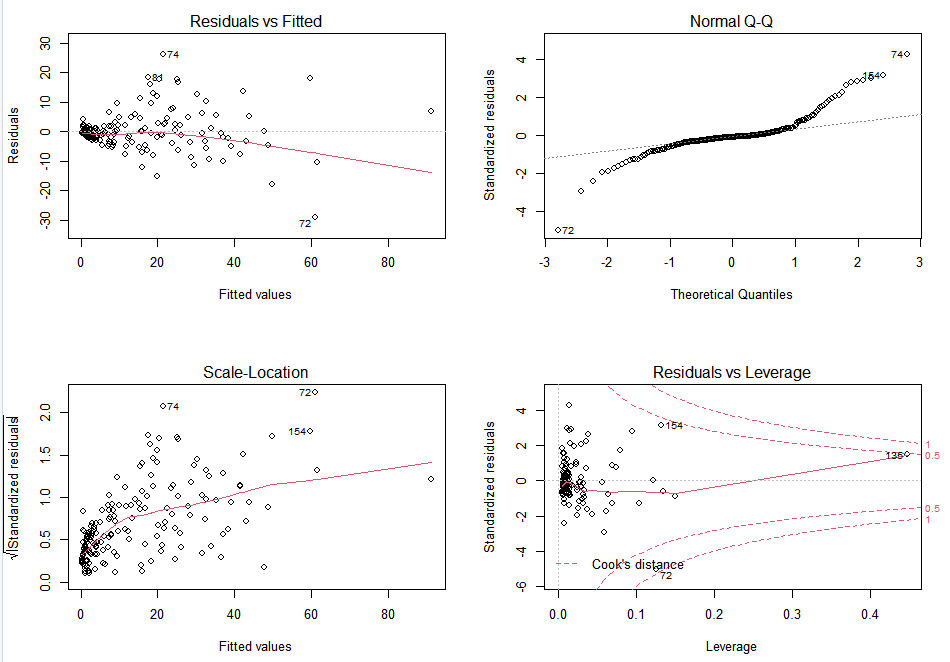
자동 생성된 설명

*그림 12 버스정류장, 공공기관, 총인구를 변수로 한 회귀분석*

* 모델이 3개의 독립변수(버스정류장, 공공기관, 총인구)로 구축되었을 때 R squared 값이 0.8359로 이 모델의 검정력이 83.59%임을 알 수 있음. Adjusted R-squared 값은 0.8333으로 모델이 적절함을 알 수 있음
* ANOVA 테이블을 보면 회귀모델의 유의확률이 2.2e-16으로 유의수준 0.05보다 작으므로 유의하다고 할 수 있음. 즉, 회귀 식이 종속변수를 설명하기에 유의하다고 볼 수 있음
* 또한, 각각의 계수의 유의확률을 보면 모두 유의수준 0.05보다 작으므로 모두 유의하다고 할 수 있음. 따라서 모든 독립변수는 종속변수에 유의한

영향을 미친다고 할 수 있음

* Model Diagnostics



*그림 13 모델 2 Diagnostics*

* 모델 진단을 하기 위해서 총 4가지 plot을 그려 확인함
* Residuals vs. Fitted plot 상에서는 fitted value가 커질수록 분산이 커져 잔차의 등분산성이 만족하지 않음을 알 수 있음
* 이로 인해, 회귀모델의 등분산성을 만족하지 않아서 종속변수에 로그와 루트를 씌워서 분석을 진행함. 하지만 종속변수를 변경함으로 인해 정규성 가정이 만족하지 않아 이상치를 제거해서 모델을 수정함

**Model 3**

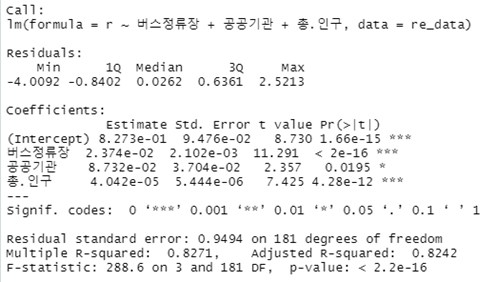
**모델 3은 종속변수에 로그를 씌우고 이상치를 제거하여 분석함**

텍스트, 영수증이(가) 표시된 사진

자동 생성된 설명

*그림 14 모델3 회귀분석 결과*

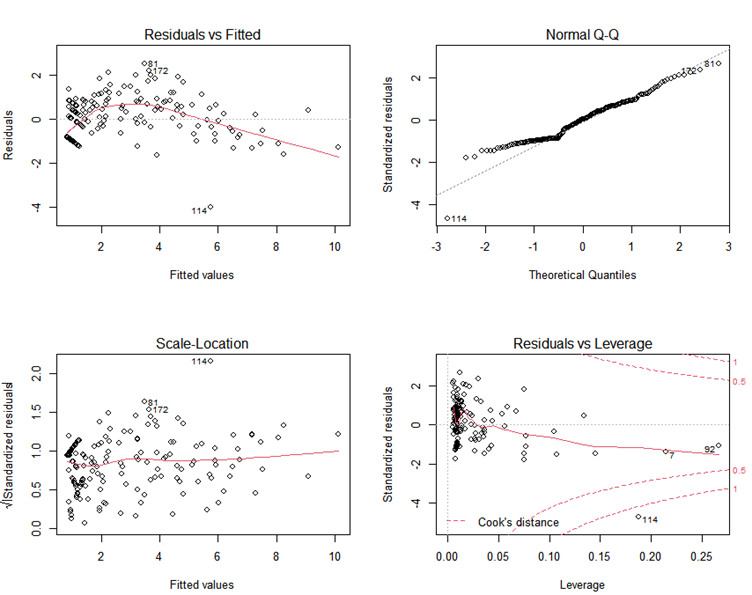
* 단계분석법을 통해 도시공원, 버스정류장, 공공기관, 총인구 변수가 선택됨. 하지만 이 모델에서는 도시공원이 유의하지 않았고, 정규성을 만족하지 않기 때문에 모델 4를 새로 구축함



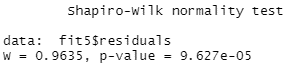
**Model 4**

**모델4는 종속변수에 루트를 씌우고 똑같이 이상치를 제거해서 분석을 진행함**

*그림 15 모델4 회귀분석 결과*

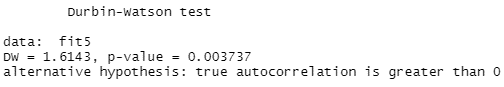
* 회귀분석 결과 상수를 포함해서 버스정류장, 공공기관, 총인구 모두 유의수준 0.05 하에 유의함

*그림 16 모델4 Diagnostics*



*그림 17 모델4 Shapiro-test 결과*

* Plot을 보더라도 오차의 등분산성이 만족함을 알 수 있음. Shapiro-test 결과 정규성 가정이 만족함을 알 수 있음
* 따라서 우리는 모델 4를 최종적으로 선택하고자 함



*그림 18 모델4 Durbin-watson test 결과*

* 더빈 왓슨 통계량이 1.5~2.5 사이에 있으며, 2에 가까우므로 잔차의 독립성이 있다고 판단할 수 있음
* 모델 4의 독립변수를 표준화한 베타계수는 다음과 같음



*그림 19 모델4의 독립변수 표준화한 베타계수*

* 다중공선성을 진단하는 VIF 값이 모두 10 미만이므로 다중공선성이 없다고 할 수 있음



*그림 20 모델4 VIF결과*

**Model 5**

**모델5는 leverage, influential point의 교집합을 제거하고 분석 진행**

* 텍스트이(가) 표시된 사진

  자동 생성된 설명Leverage, influential 포인트의 교집합인 7을 제거한 후 결괏값을 비교함

*그림 21 모델5 회귀분석 결과*

* 이상치를 제거한 결과, 병원 변수가 추가되었으나 유의하지 않기 때문에 이상치를 제거하지 않고 모델4를 최종적으로 선택함
* Final Model (Model4)

텍스트이(가) 표시된 사진

자동 생성된 설명

*그림 22 Model 4*

* 회귀계수는 모두 유의수준 0.05 하에 유의하고 Adjusted R-squared는 0.8242로 모델이 데이터에 잘 적합하다고 볼 수 있음
* 유의수준 0.05 하에 신뢰구간은 다음과 같음

텍스트이(가) 표시된 사진

자동 생성된 설명

*그림 23 Model 4 신뢰구간*

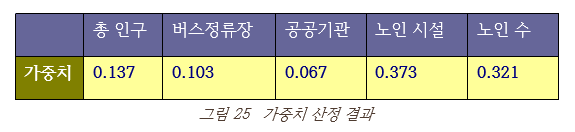
* Conclusion
* 노인 전문 공공체육시설의 입지를 선정하기 위해 7개의 독립변수(병원, 상권, 총인구, 도시공원, 버스정류장, 공공체육시설, 공공기관)를 활용해 회귀분석을 실시함
* 노인시설을 종속변수로 두어 노인시설의 입지에 영향을 미치는 요인을 파악해 공공체육시설 입지 선정에 활용하고자 함
* 첫 번째 모델은 상관성 분석에서 공공체육시설과의 연관성이 보이지 않아 제거했음
* 두 번째 모델은 단계선택법을 활용해서 선택한 결과 버스정류장, 공공기관, 총인구가 독립변수로 선정되었음. 하지만 오차항의 등분산성과 정규성을 만족하지 않아 종속변수에 로그를 씌운 모델3과 루트를 씌운 모델4를 새로 만들었음
* 모델 3은 유의하지 않은 변수(병원) 하나와 정규성을 만족하지 않아 선택하지 않음
* 모델 4는 유의수준 0.05 하에 총인구, 버스정류장, 공공기관이 유의한 변수라는 결과가 나왔음. 그리고 문제가 되었던 오차의 등분산성과 정규성 가정이 만족하였고 독립성도 만족하는 결과가 나왔음. 또한, 다중공선성이 없다는 결과를 통해 최종 모델4가 데이터에 잘 적합 하다고 할 수 있음
  1. **AHP 쌍대비교 분석을 통한 가중치 선정** 
     1. 쌍대비교 분석을 위한 상대적 중요도 조사

• 상대적 중요도 조사

텍스트, 테이블이(가) 표시된 사진

자동 생성된 설명

* + 1. 가중치 산출 결과
* 가중치 산정 결과

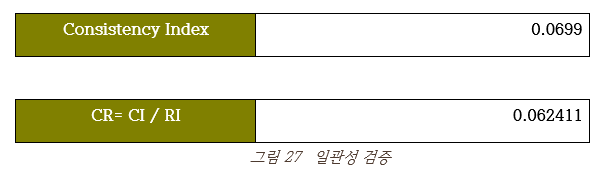


* 비교행렬

테이블이(가) 표시된 사진

자동 생성된 설명

* 일관성 검증



* 결과
* 이므로 위의 가중치 선정이 논리적임을 알 수 있음
* 따라서 총인구, 버스정류장, 공공기관, 노인시설, 노인 수 각 변수에 위의 가중치를 곱하여 진행함
  1. **Q-GIS를 활용한 분석**
     1. **Q-GIS 그래픽 모델 1**
* 과정
* 총 8개의 데이터(경로당, 공공기관, 버스정류장, 노인 교육 시설, 노인 복지관, 총인구 수, 총 노인 인구수, 상권)를 하나의 레이어로 합쳐 총 개수를 구한 후 최종 입지를 선정하는 방법으로 진행함

① 부산시 동 데이터를 폴리곤 레이어로 불러오기

② 변수 데이터를 점 레이어로 불러오기

③ ①과 ②를 ‘폴리곤에 포함하는 포인트 개수 계산’ 함수로 개수 정보 추출 후 한 레이어로 병합

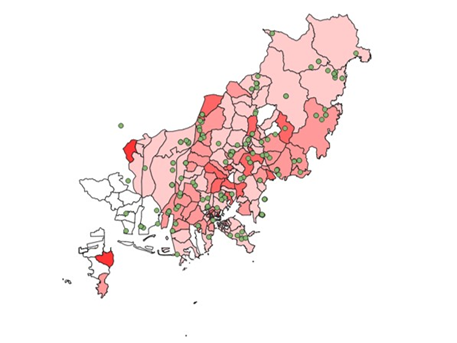
④ CSV로 내보낸 후 시설별 개수를 모두 합해 하나의 변수로 저장

⑤ ①과 ‘동 이름’을 기준으로 속성을 결합

⑥ 변수를 값으로 단계 구분, 내추럴 브레이크 모드로 총 6단계로 구분

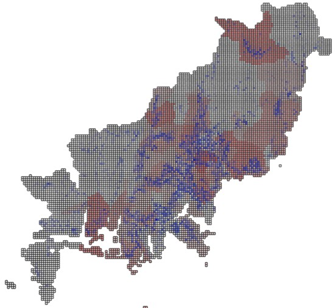
⑦ 부산광역시의 공공 체육시설 데이터를 점 레이어로 불러오기

* 결과



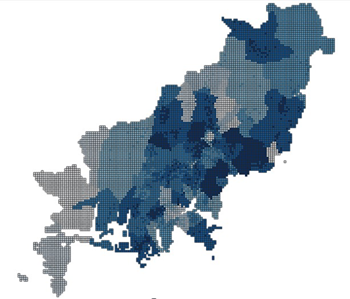
*그림 28 Graphic Model 1*

* 한계
* 노인 복지관, 노인 교실, 공공 기관은 각 동에 대체로 10개 이하이나 버스정류장은 최대 253개로 다른 변수에 비해 크기가 매우 큼. 즉 이 시설들을 단순히 합한 변수는 버스정류장의 수의 비중이 크다고 할 수 있음
* 따라서 단계 구분을 해 가장 붉게 나온 지역은 버스정류장의 수가 많은 지역일 확률이 높다는 한계가 있음. 또한 시설의 수만 고려했을 뿐, 노인 인구 수나 인구 수, 상가 정보 등이 고려되지 않았음
  + 1. **Q-GIS 그래픽 모델 2**
* 과정
* 각 변수의 크기를 고려하여 결합하기 위해 각 변수에 가중치를 곱한 다음 1점~10점으로 나누기로 함. 최종적으로 각 변수의 점수를 더한 후 격자 레이어에 시각화해 최적의 입지를 찾기로 함
* 각 변수를 부산시 격자 레이어와 폴리곤에 포함하는 포인트 개수 계산을 진행한 후 가중치가 고려된 각 시설의 정보를 격자 데이터에 저장함
* 가중치를 곱한 변수들의 값을 더해 단계 구분으로 격자 레이어에 나타냄. 격자 아래에는 총인구 데이터를 단계 구분한 폴리곤 레이어를 둠



*그림 29 최종 시각화 모델의 프로토타입*

* 변수들의 등급을 모두 더한 점수와 동 이름 정보를 포함한 하나의 CSV 파일을 만든 후 Q-GIS로 불러와 격자 데이터와 ‘레이어 속성’의 결합을 진행한 후 격자 좌표 정보를 부여
* 결과



*그 림 30 Graphic Model2*

* 한계
* 원본 데이터에 가중치를 곱하고 구간으로 나눴기 때문에 각 변수의 절대적인 수치가 결괏값에 크게 영향을 줌
  + 1. **Q-GIS 그래픽 모델 3**
* 과정
* 추출된 데이터를 7개의 구간으로 나누어 점수를 부여하고 가중치를 곱하여 수치가 보정된 새로운 모델을 만들고자 함. 체육 시설이 위치한 격자 점수가 높은 순서로 정렬한 후 입지의 우선순위 선정에 활용할 예정

① 모든 변수를 7단계로 구간 설정

② 산출된 구간 값에 ‘\* ‘함수를 사용해 가중치를 부여



*그림 31 가중치 테이블*

③ 가중치가 부여된 값들을 모두 합쳐 격자별 최종 점수 도출

④ 격자 정보가 포함된 최종 CSV와 격자 레이어를 속성 결합

⑤ 단계 구분의 값을 최종 점수로 설정하고 7 단계로 나누어 시각화

⑥ 공공 체육 시설 정보를 점 레이어로 불러온 후 위치로 속성 결합을 통해 결합

⑦ 점수가 높은 지역을 중심으로 노인 맞춤형 체육 시설 입지 선정

* 결과

텍스트, 장난감이(가) 표시된 사진

자동 생성된 설명

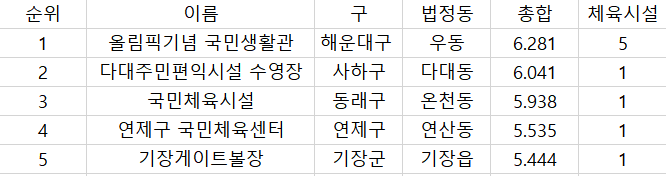
*그림 32 Graphic Model 3*

* 1. **결론**
     1. **현존하는 공공 체육시설 중 최적 입지 선정**

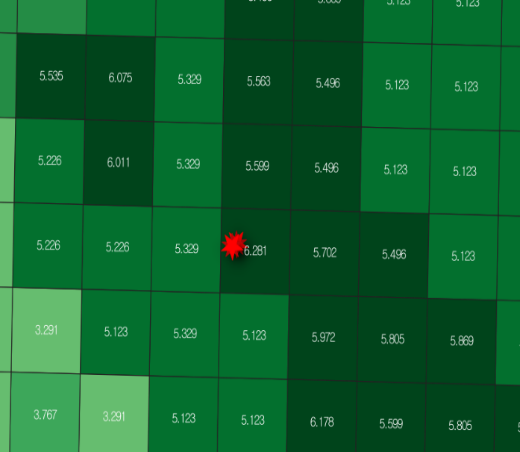
🢭 최종 점수와 체육 시설 및 동, 구 정보가 포함된 CSV를 활용하여 노인

맞춤형 체육 시설의 입지를 선정

* 노인 맞춤형 체육 시설을 신설하는 것은 비용 및 공간적 제약으로 현실성이 떨어짐. 따라서 현존하는 공공 체육시설에 노인 맞춤형 체육 시설 및 서비스를 제공하는 방안으로 현실성을 보완할 수 있음. 입지 점수가 높은 순서대로 노인 맞춤형 체육 시설의 최적 입지를 선정함
* 체육시설이 존재하는 지역 중 가장 점수가 높은 장소는 다음과 같음



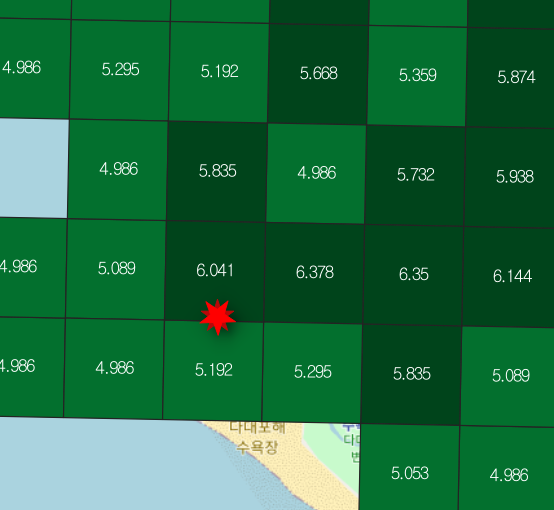
*그림 33 체육시설이 존재하는 지역 중 가장 점수가 높은 장소*

* 지도이(가) 표시된 사진

  자동 생성된 설명부산 사회 체육센터 : 부산 해운대구 우2동 1127-37

*그림 34 부산 사회 체육센터(OSM) 그림 35 부산 사회체육센터(격자)*

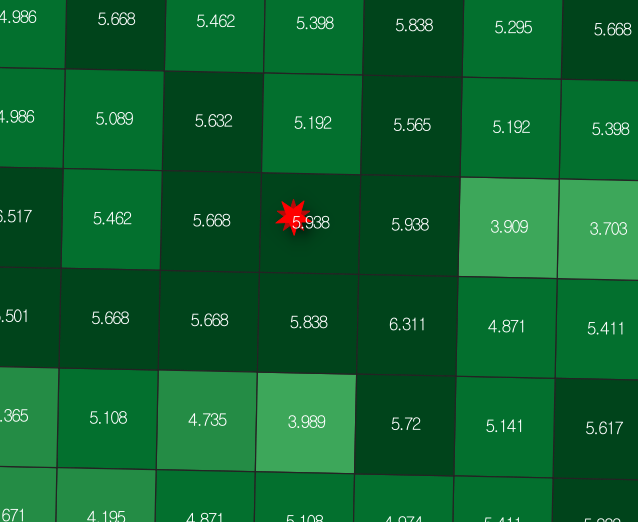
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 격자 ID | 점수 총합 | 체육시설 수 | 법정동 | 자치구 |
| 21440 | 6.281 | 5 | 우동 | 해운대구 |

* 지도이(가) 표시된 사진

  자동 생성된 설명다대 문화 체육 센터 : 부산 사하구 다대로 734

*그림 36 다대문화체육센터(OSM) 그림 37 다대문화체육센터(격자)*

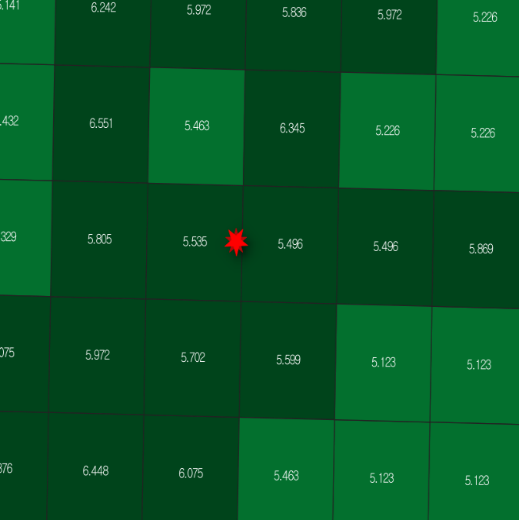
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 격자 ID | 점수 총합 | 체육시설 수 | 법정동 | 자치구 |
| 11467 | 6.041 | 1 | 다대동 | 사하구 |

* 지도이(가) 표시된 사진

  자동 생성된 설명동래구 국민체육센터 : 부산광역시 동래구 온천동 1448-2

*그림 38 동래구 국민체육센터(OSM) 그림39 동래구 국민체육센터(격자)*

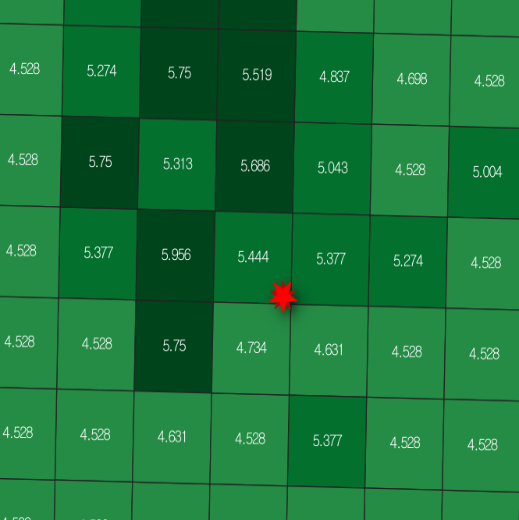
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 격자 ID | 점수 총합 | 체육시설 수 | 법정동 | 자치구 |
| 17458 | 5.938 | 1 | 온천동 | 동래구 |

* 연제구 국민체육센터 : 부산시 연제구 쌍미천로 132

*그림 40 연제구 국민체육센터(OSM) 그림 41 연제구 국민체육센터(격자)*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 격자 ID | 점수 총합 | 체육시설 수 | 법정동 | 자치구 |
| 18410 | 5.535 | 1 | 연산동 | 연제구 |

* 지도이(가) 표시된 사진

  자동 생성된 설명개나리 공원 : 부산광역시 기장군 기장읍 청강리 295

*그림 42 개나리 공원(OSM) 그림43 개나리 공원(격자)*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 격자 ID | 점수 총합 | 체육시설 수 | 법정동 | 자치구 |
| 25760 | 5.444 | 1 | 기장읍 | 기장군 |

* + 1. **노인 맞춤형 체육 시설의 *신설 우선순위 선정***
* 총점이 높지만 공공 체육 시설이 존재하지 않는 지역을 구별로 5개 선정함. 이는 노인 맞춤형 체육 시설을 신설할 때 우선으로 선정되어야 하는 5개 지역을 의미함
* 6.551점의 남구가 최우선 신설 입지로 선정되었고 수영구, 북구, 금정구, 영도구가 뒤를 이었음

텍스트, 낱말맞추기게임이(가) 표시된 사진

자동 생성된 설명텍스트, 장난감이(가) 표시된 사진

자동 생성된 설명

*그림44 최종모델*

* 1. **한계 및 제언**
* 데이터
* 교통 접근성 분석에서 체육시설과의 거리나 교통량 대신 버스정류장 수 데이터를 대체해 정교함이 떨어짐
* 인구 데이터를 제공하는 기준 동이 Q-GIS 법정동과 달라 일부 지역의 인구를 임의로 0명으로 설정함
* 분석
* 회귀분석

– 종속변수에 루트를 씌운 모델을 최종결정했으나 Q-GIS로 분석할 때는 독립변수를 변형하지 않고 같은 간격으로 나누는 과정을 통해 임의로 값을 부여한 한계가 있음

* AHP

- 상대적 중요도 선정 시 표본 데이터가 부족했으며 비교적 적은 변수로 인해 가중치 선정의 정확성이 떨어졌음

- 상대적 중요도 선정이 사람이 설정하는 기준과 주관에 의해 진행되므로 객관성이 떨어짐

* Q-GIS
* Q-GIS의 법정동이 현재의 법정동이 아니어서 외곽지역의 값이 병합되지 않고 NULL 값이 됨
* 격자 레이어 제작 시 지역이 사라지는 문제가 생김
* 동 이름이 같은 송정동의 경우 데이터 결합 시 데이터가 모두 NULL 값이 됨

1. **활용 방안**
   1. **문제점 개선 방안**

* 현재까지 존재하지 않던 노인 전문 체육 시설 구축
* 현시점, 노인들을 위한 체육 시설이 존재하지 않는 근본적인 문제를 해결할 수 있으며,
* 기존의 노인 시설을 기반으로 공간적 분포 특성을 분석하고 노인의 체육 활동 수요가 높은 최적 입지를 접근성 관점에서 다양한 변수들과 비교 분석을 통해 추출한 최적의 결과를 도출함
* 현재 최적의 위치는 기장읍, 온천3동, 다대1동, 연산1동, 우2동이 노인 전문 체육 시설이 구축된다면 노인 전문 체육 시설 확산의 시발점이 될 것으로 예상됨
* 공공 기관과 지자체의 체육 환경 조성 비용 문제 개선
* 노인만의 체육 시설 구축은 공공 기관 및 지자체로서 초기 비용 부담이 문제점으로 대두되고 있으나,
* 노인 전문 체육 시설 최적 위치로 선정된 체육 시설은 기존의 시설에서 필요한 부분이 추가되는 개념으로써 초기 비용이 절감된 채 시작하기에 많은 공공 기관과 지자체의 비용 문제를 해결함으로써 더 넓은 지역으로 쉽게 퍼질 수 있을 것으로 기대됨
* 부산 시청 홈페이지의 공공시설 건립 비용 정보에 따르면 부산 탁구 체육관 한 개를 건립하는데 약 41억9천만 원이 소요되었고, 반면에 금정구 경로회관 증축 공사로는 약 12억 원이 소요됨. 체육 시설 설치 비용은 단위면적당(㎡) 평균 289만 원이 추가된다는 대한민국 조달청의 자료를 근거로 경로회관 즉 노인 시설에 증축 공사 및 체육 시설을 도입하는 것은 탁구 체육관 1개 건립 비용과 비교하면 초기 비용을 아낄 수 있음이 증명되며(1000㎡ 초과 시 해당 사항 없음) 이는 큰 비용 문제를 개선할 수 있을 것으로 전망
* 또한 최적 입지 고려 사항으로 버스 정류장과 노인들의 도보 거리도 고려하여 진행하였기에 새로운 교통편을 마련하는 비용도 절감할 수 있을 것으로 기대됨
* 노인 체육 활동 증진으로 인한 의료비 증가 문제 개선
* 노인 전문 체육 시설 증설은 노인들이 체육 활동을 자주 접할 수 있는 환경을 제공할 수 있을 것이며 체육 활동을 통해 노인들의 심폐기능 상승, 근력 상승 등 질병 유발을 저하할 수 있음. 이는 2013년부터 늘어나고 있는 의료비의 감소에 큰 도움이 될 것을 전망됨

*그림 45 65세 이상 노인 진료비 추이* 테이블이(가) 표시된 사진

자동 생성된 설명

* 신체적 의료비 감소와 더불어 체육 활동을 통한 사회 활동으로 인해 노인 우울증 및 스트레스가 해소되어 정신 질환 발생 확률이 감소하였다는 결과를 토대로 신체적, 정신적 건강 모두 긍정적인 효과가 기대되며, 의료비 절감이란 결과를 도출해 낼 수 있음
  1. **업무 활용 방안**
* (부산 이미지 구축) 적극적인 노인 복지를 추진해 부산광역시에 고령 친화 도시 이미지를 구축 가능
* 유럽의 스포츠 클럽은 집 근처에 위치해 주기적인 운동 습관을 기르는데 기여함. 네덜란드의 80세 이상 고령자들은 은퇴 이후에도 습관적으로 운동을 하며, 운동 참여율은 50%에 달함
* 우리나라 70세 이상 인구의 8.4%만이 생활체육 동호회에 가입함. 노인 맞춤형 공공 체육 시설을 적합한 입지에 설치함으로써 유럽의 스포츠클럽 참여율만큼 부산의 운동 참여율 상승을 통한 부산광역시의 고령 친화 도시 이미지 구축 가능
* (노인 복지 만족도 상승) 부족했던 노인만을 위한 체육시설을 설립함으로써, 노인 복지 만족도에 이바지 가능
* 노인만을 위한 전문 체육시설인 만큼 노인의 특성을 고려한 운동기구를 설치는 물론, 그 외에 제공되는 서비스도 노인 맞춤형 서비스로 제공될 것이기에 노인 복지 만족도에 많은 이바지를 할 수 있음

*그림 46 노인 특성을 고려한 운동기구*텍스트, 기기이(가) 표시된 사진

자동 생성된 설명

* (일자리 창출) 노인 전문 체육시설 설립 후 관리나 운영자를 노인으로 지정하여 새로운 일자리 창출 가능
* 광주광역시 광진구의 노인 시설 관리자 채용 사례를 본받아 노인 일자리 창출 추진
* 광주 광산구 시설 관리 공단은 (사)한마음 금고 복지 센터(센터장 김민수)와 광산구 관내 저소득층 일자리 창출 및 사회활동 지원 사업을 위한 업무 협약을 체결함
* 센터 소속 어르신 근로자는 주 3회 공공체육시설의 소독, 체육시설 부식 방지를 위한 이음새 윤활 방청제 도포 작업, 고장·노후 시설 현황을 공단에 제보 등의 업무를 수행하는 관리 업무를 배정받음
* 위 사례를 바탕으로 시설 공단은 공공 체육 시설물 관리 협력을 통해 건강하고 근로 능력이 있는 노인들에게 근로의 기회를 제공하여 새로운 일자리를 창출하고, 업무로는 고장 신고된 시설물의 신속한 점검 및 유지 보수 등 안전한 체육시설 이용 환경을 조성 역할을 배당할 수 있음
* (취업률 증가) 노인 스포츠 지도자 인재 양성 및 취업률 증가
* 노인 전문 체육 시설을 운영하기 위해선 노인 스포츠 지도자가 절실히 필요함. 노인 스포츠 지도자는 사설 체육시설, 직장, 생활 체육 동호인 단체 등에서 전 국민을 대상으로 건강 유지, 여가 선용 등 노인 생활 체육 참여자의 목적을 충족시켜 주기 위하여 프로그램을 개발, 수집 및 지도하고 체육시설을 관리하는 전반적인 총괄 책임자 역할을 함. 하지만 이를 운영할 전문성을 가진 스포츠 지도사가 턱없이 부족한 실정임
* 노인 스포츠 지도자 인력 부족 문제를 해결하기 위해 청년 취업난과 묶어 업무 활용 방안을 구상함. 노인 스포츠 지도자 인재 양성 프로그램을 통해 지도자 자격증을 취득하도록 장려하여 노인 전문 체육 시설 개수에 적절한 노인 스포츠 지도사를 투입하는 것임. 청년층의 취업난을 해결하는 방안임은 물론, 노인 스포츠 지도사의 나이 제한을 두지 않는다면 중년층을 넘어, 같은 노인들이 서로 가르치고 배우며 함께 상생하는 사회까지 만들 수 있을 것으로 예상됨

1. **부록 및 출처**
   1. **분석데이터 출처**

- 부산광역시(공공체육시설 현황), 2020.05.10. <https://www.busan.go.kr/ghopen09/view?schCommand=List&schUpperClNo=277&schClNo=207&schBizNo=379&schIndx=12057&bizCurPage=1&cntPerPage=50&schSdate=&schEdate=&srchKey=&srchText=>

- 부산광역시(2021년 1/4분기 부산시 의료기관 현황), 2021.04.19. <https://www.busan.go.kr/ghopen09/view?schCommand=List&schUpperClNo=133&schClNo=176&schBizNo=240&schIndx=13666&bizCurPage=2&cntPerPage=15&schSdate=&schEdate=&srchKey=&srchText=>

- 잡 알리오 플러스(공공기관 정보), 2021.08.03. <http://www.alioplus.go.kr/organization/organByTypeList.do?pageNo=1&excelTitle=&schApbaGb=APBA_ALL&schForeignYn=&schApbaCate=&schApbaId=&schSiNa=%25EB%25B6%2580%25EC%2582%25B0%25EA%25B4%2591%25EC%2597%25AD%25EC%258B%259C&schSggNa=&schCont=>

- 부산광역시 빅데이터 포털(공공포털\_상가\_상권\_정보), 2021.01.07.

<https://bigdata.busan.go.kr/data/bigDataDetailView.do?menuCode=M00000000007>

- KOSIS(통계청, 도시 공원), 2021.05.13. <https://kosis.kr/statHtml/statHtml.do?orgId=202&tblId=DT_914&lang_mode=ko&vw_cd=MT_OTITLE&list_id=202_20203_I&conn_path=I4>

- KOSIS(통계청, 구·군별 세대 및 인구), 2020.05.27. <https://kosis.kr/statHtml/statHtml.do?orgId=202&tblId=DT_202&lang_mode=ko&vw_cd=MT_OTITLE&list_id=202_20203_B&conn_path=I4>

- KOSIS(통계청, 노인여가 복지시설(2005~2011)), 2016.12.09.

<https://kosis.kr/statHtml/statHtml.do?orgId=202&tblId=DT_1163A&lang_mode=ko&vw_cd=MT_OTITLE&list_id=202_20203_K3&conn_path=I4>

- 부산광역시 빅데이터 포털(공공포털\_상가\_상권\_정보), 2021.01.07.

<https://bigdata.busan.go.kr/data/bigDataDetailView.do?menuCode=M00000000007>

- 통계청(2019년 사회조사 결과(복지, 사회참여, 문화와 여가, 소득과 소비, 노동), 2019.11.25. <https://kostat.go.kr/portal/korea/kor_nw/1/6/3/index.board?bmode=read&bSeq=&aSeq=378876&pageNo=1&rowNum=10&navCount=10&currPg=&searchInfo=&sTarget=title&sTxt=>

-부산광역시(부산광역시\_버스 정류소 정보), 2021.07.06.

<https://data.busan.go.kr/dataSet/detail.nm?publicdatapk=15084251&contentId=10>

* 1. **참고 자료**

-조민제, 이한주, “부산, 국내도시 4번째 WHO 고령친화도시 네트워크 가입 부산과 지구촌 고령·도시화 문제 …국제사회 공동대응”, 부산시보, 2016.11.30., 1, <https://www.busan.go.kr/news/snsbusan01/view?dataNo=56969>

- 국민체육진흥공단, ‘노인스포츠지도사’ <https://www.insports.or.kr/examInfo/examInfo9.do>

- 주동석, “광산구시설관리공단, 노인일자리 창출 체육시설 건강지킴이 업무협약”, 이뉴스투데이, 2021.03.11

<http://www.enewstoday.co.kr/news/articleView.html?idxno=1459463>

- 양상훈 김종식, 야외운동기구를 이용한 운동프로그램이 노인들의 활동체력, 대사증후군 위험인자 및 염증인자에 미치는 영향, 2014, <https://www.ksep-es.org/upload/pdf/es-23-3-229.pdf>, 231p

- 신다정, “초고령사회 눈 앞, 노인 체육시설 따로 없다?”, 브라보마이라이프,

2021.08.02

<http://bravo.etoday.co.kr/view/atc_view.php?varAtcId=12560>

- 문화체육관광부, 생활밀착형 국민체육센터 모델개발 및 추진 방안 연구.pdf,38p,94p, <https://www.mcst.go.kr/kor/s_policy/dept/deptView.jsp?pSeq=1206&pDataCD=0417000000&pType=07>

- 강재훈, “갈 곳 없는 노인 스포츠, 유럽형 스포츠클럽 도입이 해법”,2019.07.05, <https://news.kbs.co.kr/news/view.do?ncd=4236565&ref=A>

* 1. **부록**
* 데이터 상세 목록

부산시 구·군별 세대 및 인구[.XLSX]

*테이블이(가) 표시된 사진

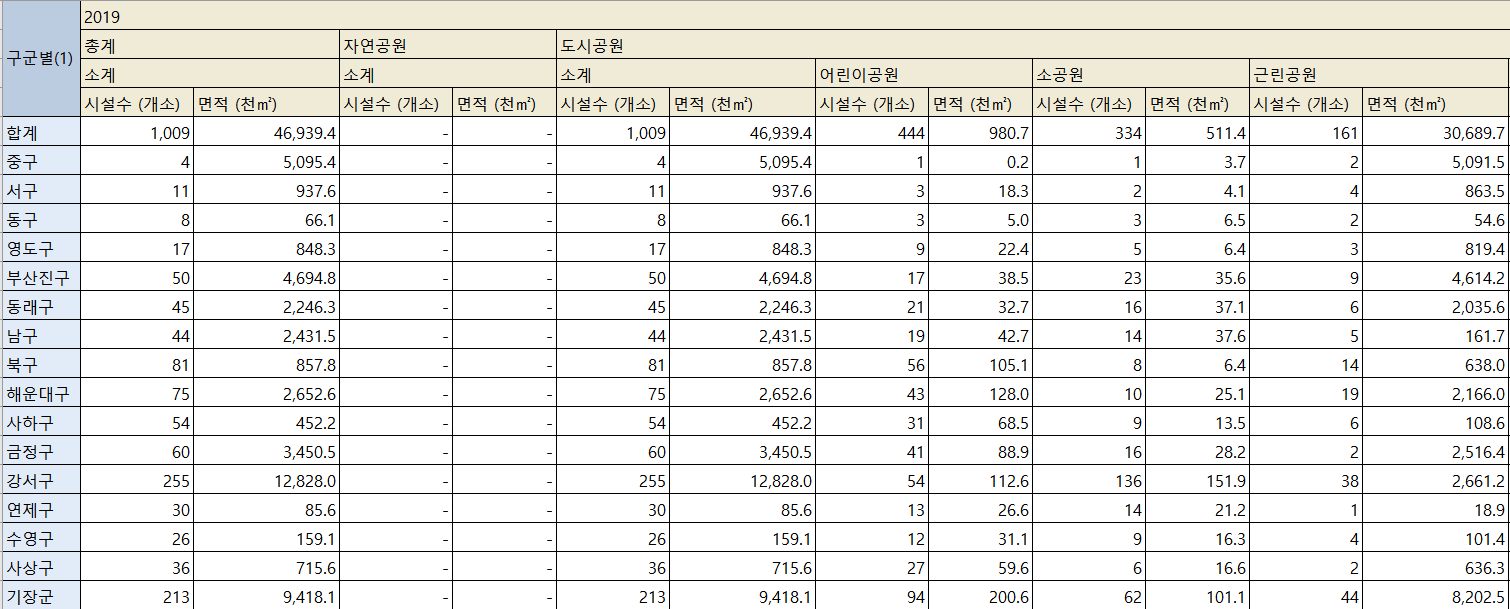
자동 생성된 설명*

*[부록] 그림 1 부산시 구·군별 세대 및 인구*

지도이(가) 표시된 사진

자동 생성된 설명부산광역시\_버스 정류소 정보[.SHP]

*[부록] 그림 2 부산광역시 버스 정류소 정보*

공원 [.XLSX]

*[부록] 그림 3 부산시 공원*

2021년 1/4분기 부산시 의료기관 현황[.XLSX]

텍스트, 캐비닛, 스크린샷이(가) 표시된 사진

자동 생성된 설명

*[부록] 그림 4 2021년 1/4분기 부산시 의료기관 현황*

공공포털\_상가\_상권\_정보\_2020-09[.TXT]

텍스트이(가) 표시된 사진

자동 생성된 설명

*[부록] 그림 5 공공포털 상가 상권 정보(2020.09. 기준)*

테이블이(가) 표시된 사진

자동 생성된 설명

*[부록] 그림 6 공공포털 상가 상권 정보(xlsx ver.)*

노인여가 복지시설[.XLSX]

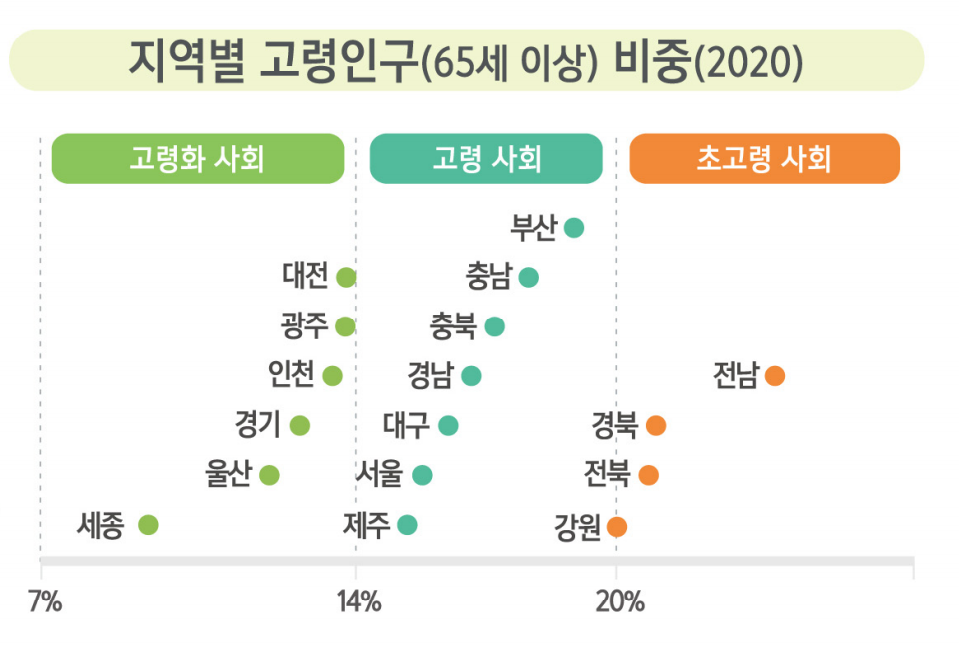
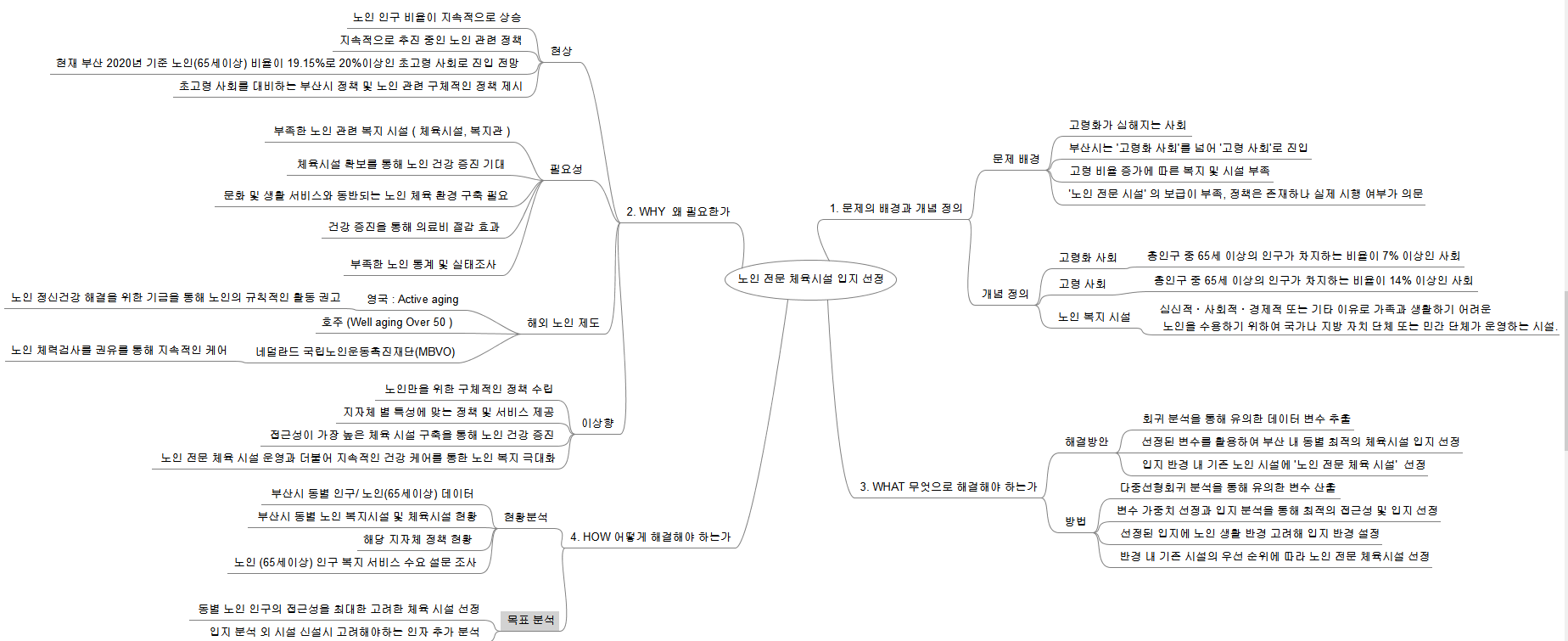
테이블이(가) 표시된 사진

자동 생성된 설명

*[부록] 그림 7 노인여가 복지시설*

<입지 분석 설계를 위한 논리 모델 마인드 맵>

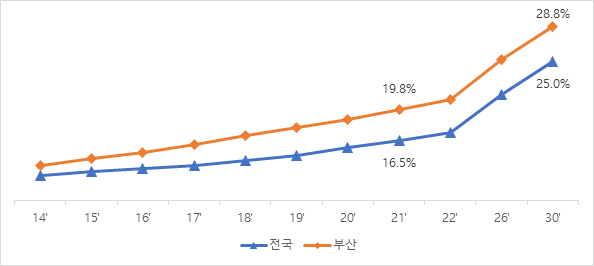
*[부록] 그림 8입지분석 논리모델을 위한 마인드맵*



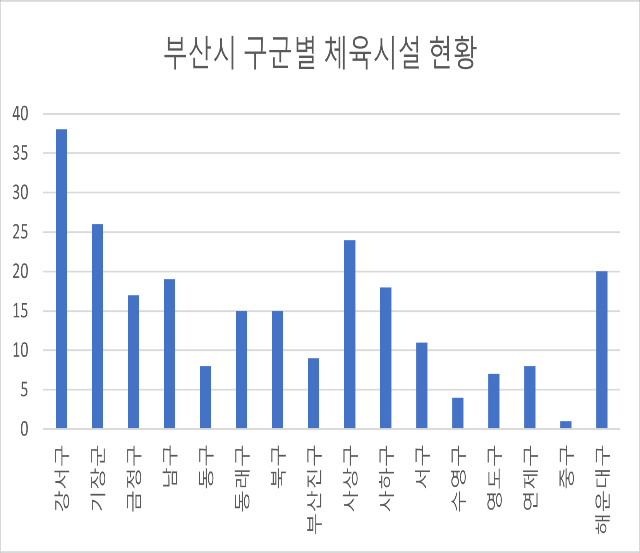
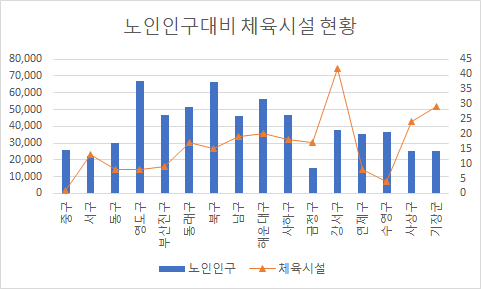
*[부록] 그림 9 지역별 고령인구 비중*

출처: 통계청

< 연도별 고령화 비율 그래프 >



*[부록] 그림 10 연도별 고령화 비율 그래프*

<노인인구 대비 체육 시설 현황>

출처: 통계청

출처: 부산시

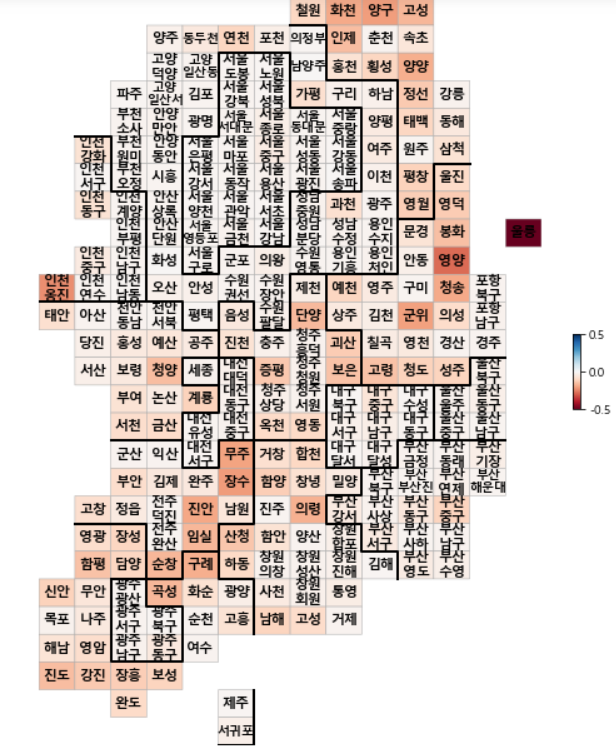
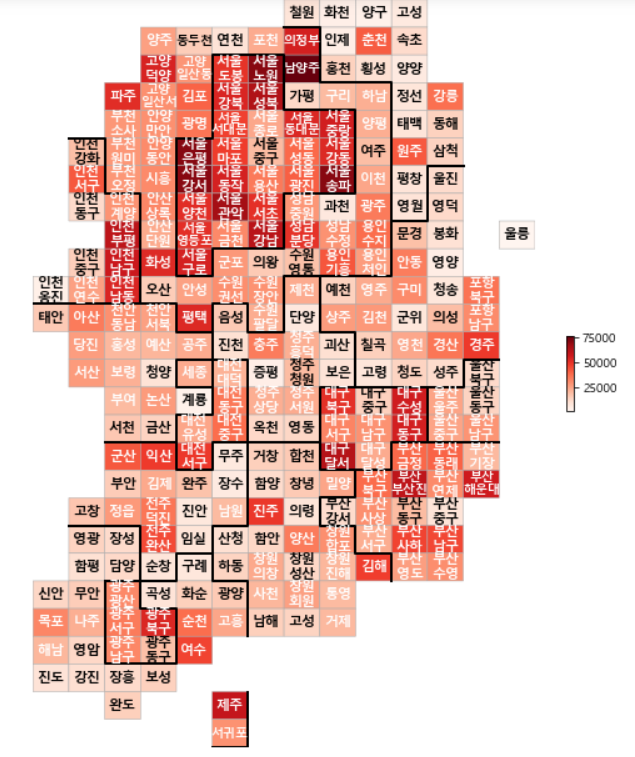
< 부산시 구군별 체육시설 현황>

*[부록] 그림 12 부산시 구군별 체육시설 현황*

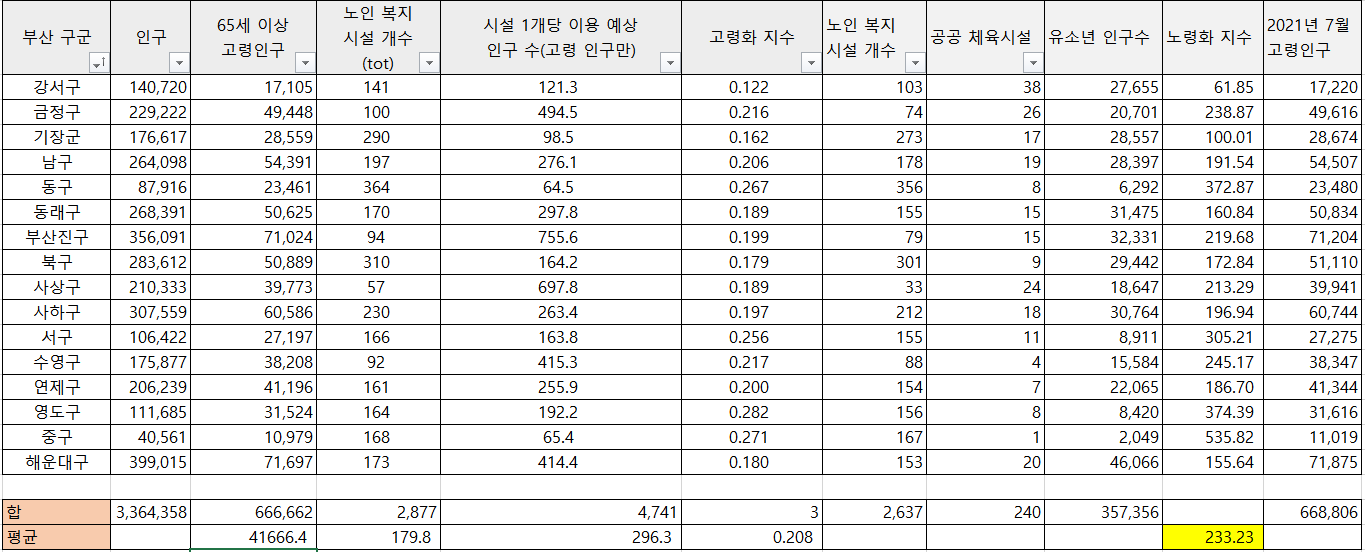
*[부록] 그림 11 노인인구 대비 체육시설 현황*

>

<시도별 인구 기반 Python 지도 시각화>

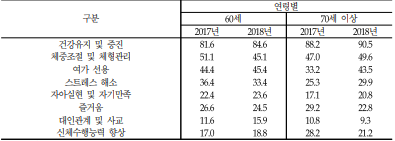


*[부록] 그림 13 시도별 고령화 지수 기반 Python지도 시각화*

<부산시 인구 및 노인 관련 수치 데이터>

*[부록] 그림 14 부산시 인구 및 노인 관련 수치 데이터*

<노인 체육 시설 참가 이유>

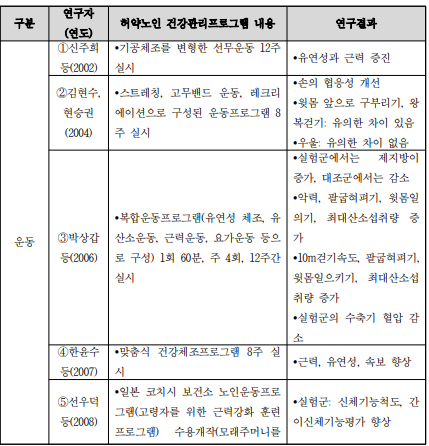


출처: 문화체육관광부

*[부록] 그림 15 노인 체육 시설 참가 이유*

<허약 노인 건강관리 관련 국내 연구 동향>

출처: 한국건강증진개발원



*[부록] 그림 16 허약 노인 건강관리 관련 국내 연구 동향*

출처: 문화체육관광부(2014)

<체육활동이 의료비 절감에 미치는 효과(65세 이상)>



*[부록] 그림 17 체육활동이 의료비 절감에 미치는 효과(65세 이상)*

* 선형 회귀 분석 코드

bigdata

rm(list=ls())  
data <- read.csv('C:/Users/user/bigdata/변수\_final.csv')

# summary

head(data)

## 동.이름 노인교실 경로당 노인복지관 병원 상권 총.인구 노인.인구 도시공원  
## 1 가야동 2 29 1 41 616 36332 8092 3  
## 2 감만동 2 25 0 19 294 20161 5389 3  
## 3 감전동 1 8 0 12 684 13543 3054 1  
## 4 감천동 2 26 0 21 387 25023 6687 2  
## 5 강동동 0 20 0 1 176 4332 1480 0  
## 6 개금동 2 41 0 83 777 57107 10762 4  
## 버스정류장 공공체육시설.개수 공공기관 노인시설  
## 1 51 0 2 32  
## 2 58 3 2 27  
## 3 56 4 6 9  
## 4 59 6 4 28  
## 5 91 6 1 20  
## 6 86 0 3 43

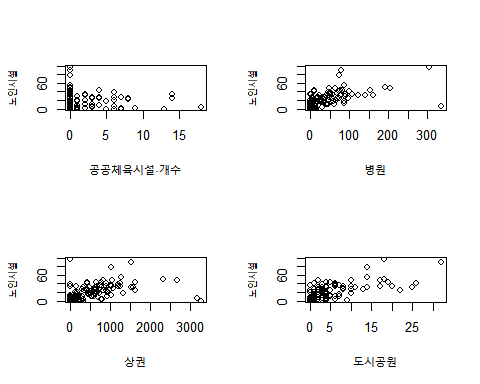
str(data)

## 'data.frame': 192 obs. of 13 variables:  
## $ 동.이름 : chr "가야동" "감만동" "감전동" "감천동" ...  
## $ 노인교실 : int 2 2 1 2 0 2 2 0 0 0 ...  
## $ 경로당 : int 29 25 8 26 20 41 46 0 0 0 ...  
## $ 노인복지관 : int 1 0 0 0 0 0 0 0 0 0 ...  
## $ 병원 : int 41 19 12 21 1 83 63 9 0 0 ...  
## $ 상권 : int 616 294 684 387 176 777 1042 333 240 250 ...  
## $ 총.인구 : int 36332 20161 13543 25023 4332 57107 62482 255 186 186 ...  
## $ 노인.인구 : int 8092 5389 3054 6687 1480 10762 11984 86 60 60 ...  
## $ 도시공원 : int 3 3 1 2 0 4 2 0 1 0 ...  
## $ 버스정류장 : int 51 58 56 59 91 86 140 0 0 0 ...  
## $ 공공체육시설.개수: int 0 3 4 6 6 0 0 7 0 0 ...  
## $ 공공기관 : int 2 2 6 4 1 3 16 0 0 0 ...  
## $ 노인시설 : int 32 27 9 28 20 43 48 0 0 0 ...

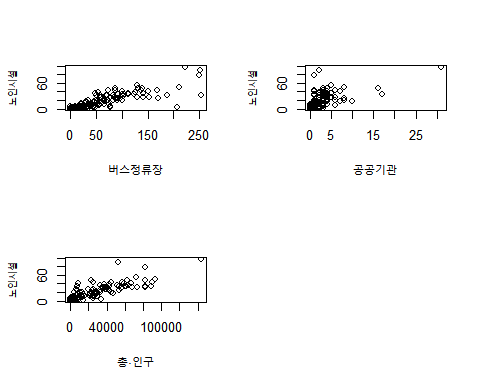
summary(data)

## 동.이름 노인교실 경로당 노인복지관   
## Length:192 Min. : 0.0000 Min. : 0.00 Min. : 0.0000   
## Class :character 1st Qu.: 0.0000 1st Qu.: 0.00 1st Qu.: 0.0000   
## Mode :character Median : 0.0000 Median : 4.00 Median : 0.0000   
## Mean : 0.8594 Mean :11.92 Mean : 0.3438   
## 3rd Qu.: 1.0000 3rd Qu.:20.00 3rd Qu.: 0.0000   
## Max. :11.0000 Max. :87.00 Max. :34.0000   
## 병원 상권 총.인구 노인.인구   
## Min. : 0.00 Min. : 0.00 Min. : 0.0 Min. : 0   
## 1st Qu.: 0.00 1st Qu.: 14.75 1st Qu.: 819.8 1st Qu.: 192   
## Median : 4.00 Median : 113.00 Median : 4123.5 Median : 1177   
## Mean : 25.78 Mean : 368.13 Mean : 17438.6 Mean : 3416   
## 3rd Qu.: 35.50 3rd Qu.: 605.25 3rd Qu.: 26963.8 3rd Qu.: 5798   
## Max. :335.00 Max. :3259.00 Max. :143442.0 Max. :29212   
## 도시공원 버스정류장 공공체육시설.개수 공공기관   
## Min. : 0.000 Min. : 0.00 Min. : 0.000 Min. : 0.000   
## 1st Qu.: 0.000 1st Qu.: 4.00 1st Qu.: 0.000 1st Qu.: 0.000   
## Median : 1.000 Median : 26.00 Median : 0.000 Median : 1.000   
## Mean : 3.151 Mean : 46.79 Mean : 1.198 Mean : 1.875   
## 3rd Qu.: 4.000 3rd Qu.: 73.25 3rd Qu.: 1.000 3rd Qu.: 2.250   
## Max. :32.000 Max. :253.00 Max. :18.000 Max. :31.000   
## 노인시설   
## Min. : 0.00   
## 1st Qu.: 1.00   
## Median : 4.50   
## Mean :13.12   
## 3rd Qu.:23.25   
## Max. :98.00

par(mfrow=c(2,2))  
plot(노인시설 ~ 공공체육시설.개수, data=data)  
plot(노인시설 ~ 병원, data=data)  
plot(노인시설 ~ 상권, data=data)  
plot(노인시설 ~ 도시공원, data=data)



plot(노인시설 ~ 버스정류장, data=data)  
plot(노인시설 ~ 공공기관, data=data)  
plot(노인시설 ~ 총.인구, data=data)  
attach(data)



# model fitting

m1 <- lm(노인시설 ~ 공공체육시설.개수 + 병원 + 상권 + 도시공원 + 버스정류장 + 공공기관 + 총.인구)  
summary(m1)

##   
## Call:  
## lm(formula = 노인시설 ~ 공공체육시설.개수 + 병원 + 상권 + 도시공원 +   
## 버스정류장 + 공공기관 + 총.인구)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -32.277 -2.095 -0.482 1.474 33.225   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 3.527e-01 7.165e-01 0.492 0.62313   
## 공공체육시설.개수 -2.226e-01 1.819e-01 -1.224 0.22266   
## 병원 -2.793e-02 2.108e-02 -1.325 0.18668   
## 상권 2.402e-04 1.478e-03 0.162 0.87111   
## 도시공원 1.909e-01 1.610e-01 1.186 0.23726   
## 버스정류장 1.380e-01 1.680e-02 8.216 3.65e-14 \*\*\*  
## 공공기관 6.634e-01 2.423e-01 2.738 0.00679 \*\*   
## 총.인구 3.078e-04 5.135e-05 5.994 1.06e-08 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 7.001 on 184 degrees of freedom  
## Multiple R-squared: 0.8402, Adjusted R-squared: 0.8341   
## F-statistic: 138.2 on 7 and 184 DF, p-value: < 2.2e-16

# model selection

data1 <- data[,2:13]  
cor(data1,method='pearson')

## 노인교실 경로당 노인복지관 병원  
## 노인교실 1.000000000 0.64333408 0.003207994 0.60191403  
## 경로당 0.643334084 1.00000000 0.017799397 0.65107182  
## 노인복지관 0.003207994 0.01779940 1.000000000 0.01968317  
## 병원 0.601914033 0.65107182 0.019683174 1.00000000  
## 상권 0.354946040 0.55814780 -0.003403288 0.73302901  
## 총.인구 0.683531712 0.82860625 0.253204588 0.78860528  
## 노인.인구 0.735513418 0.82510244 0.278052855 0.76905085  
## 도시공원 0.550546009 0.72289465 0.092177038 0.55333112  
## 버스정류장 0.533884494 0.85075771 0.149689663 0.60918409  
## 공공체육시설.개수 -0.059991321 -0.02815496 0.325933647 -0.03512814  
## 공공기관 0.597254911 0.60975115 0.106709340 0.62226652  
## 노인시설 0.685414675 0.98743924 0.160612862 0.66009450  
## 상권 총.인구 노인.인구 도시공원 버스정류장  
## 노인교실 0.354946040 0.6835317 0.73551342 0.55054601 0.53388449  
## 경로당 0.558147798 0.8286063 0.82510244 0.72289465 0.85075771  
## 노인복지관 -0.003403288 0.2532046 0.27805285 0.09217704 0.14968966  
## 병원 0.733029011 0.7886053 0.76905085 0.55333112 0.60918409  
## 상권 1.000000000 0.6335565 0.59355603 0.46775825 0.55968004  
## 총.인구 0.633556472 1.0000000 0.95679446 0.74348902 0.78891746  
## 노인.인구 0.593556028 0.9567945 1.00000000 0.68542147 0.73819086  
## 도시공원 0.467758246 0.7434890 0.68542147 1.00000000 0.75878628  
## 버스정류장 0.559680041 0.7889175 0.73819086 0.75878628 1.00000000  
## 공공체육시설.개수 0.025396920 0.0549426 0.05628476 0.02317645 0.05095992  
## 공공기관 0.334557093 0.6879713 0.75017542 0.37071409 0.51293300  
## 노인시설 0.548857213 0.8656704 0.87054420 0.73270838 0.85824286  
## 공공체육시설.개수 공공기관 노인시설  
## 노인교실 -0.059991321 0.597254911 0.68541468  
## 경로당 -0.028154963 0.609751151 0.98743924  
## 노인복지관 0.325933647 0.106709340 0.16061286  
## 병원 -0.035128138 0.622266515 0.66009450  
## 상권 0.025396920 0.334557093 0.54885721  
## 총.인구 0.054942604 0.687971318 0.86567044  
## 노인.인구 0.056284759 0.750175416 0.87054420  
## 도시공원 0.023176447 0.370714086 0.73270838  
## 버스정류장 0.050959920 0.512932997 0.85824286  
## 공공체육시설.개수 1.000000000 0.005587828 0.01547283  
## 공공기관 0.005587828 1.000000000 0.63383401  
## 노인시설 0.015472831 0.633834010 1.00000000

m2 <- lm(노인시설 ~ 병원 + 상권 + 도시공원 + 버스정류장 + 공공기관 + 총.인구) # 공공체육시설은 뺌  
summary(m2)

##   
## Call:  
## lm(formula = 노인시설 ~ 병원 + 상권 + 도시공원 + 버스정류장 +   
## 공공기관 + 총.인구)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -32.296 -2.473 -0.426 1.196 33.457   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.256e-01 6.930e-01 0.181 0.85635   
## 병원 -2.461e-02 2.093e-02 -1.176 0.24107   
## 상권 1.545e-04 1.478e-03 0.105 0.91687   
## 도시공원 2.000e-01 1.610e-01 1.242 0.21582   
## 버스정류장 1.376e-01 1.682e-02 8.183 4.35e-14 \*\*\*  
## 공공기관 6.688e-01 2.426e-01 2.757 0.00642 \*\*   
## 총.인구 3.012e-04 5.113e-05 5.891 1.78e-08 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 7.011 on 185 degrees of freedom  
## Multiple R-squared: 0.8389, Adjusted R-squared: 0.8337   
## F-statistic: 160.6 on 6 and 185 DF, p-value: < 2.2e-16

fit3 <- step(m2, direction = "both")

## Start: AIC=754.68  
## 노인시설 ~ 병원 + 상권 + 도시공원 + 버스정류장 + 공공기관 + 총.인구  
##   
## Df Sum of Sq RSS AIC  
## - 상권 1 0.5 9093.1 752.69  
## - 병원 1 68.0 9160.5 754.11  
## - 도시공원 1 75.8 9168.4 754.28  
## <none> 9092.6 754.68  
## - 공공기관 1 373.6 9466.1 760.41  
## - 총.인구 1 1705.7 10798.2 785.69  
## - 버스정류장 1 3291.3 12383.9 812.00  
##   
## Step: AIC=752.69  
## 노인시설 ~ 병원 + 도시공원 + 버스정류장 + 공공기관 + 총.인구  
##   
## Df Sum of Sq RSS AIC  
## - 도시공원 1 75.5 9168.5 752.28  
## - 병원 1 88.0 9181.0 752.54  
## <none> 9093.1 752.69  
## + 상권 1 0.5 9092.6 754.68  
## - 공공기관 1 408.9 9502.0 759.14  
## - 총.인구 1 1760.1 10853.2 784.67  
## - 버스정류장 1 3436.1 12529.2 812.24  
##   
## Step: AIC=752.28  
## 노인시설 ~ 병원 + 버스정류장 + 공공기관 + 총.인구  
##   
## Df Sum of Sq RSS AIC  
## - 병원 1 91.7 9260.3 752.19  
## <none> 9168.5 752.28  
## + 도시공원 1 75.5 9093.1 752.69  
## + 상권 1 0.2 9168.4 754.28  
## - 공공기관 1 343.2 9511.7 757.34  
## - 총.인구 1 2475.9 11644.4 796.18  
## - 버스정류장 1 4696.7 13865.3 829.69  
##   
## Step: AIC=752.19  
## 노인시설 ~ 버스정류장 + 공공기관 + 총.인구  
##   
## Df Sum of Sq RSS AIC  
## <none> 9260.3 752.19  
## + 병원 1 91.7 9168.5 752.28  
## + 도시공원 1 79.2 9181.0 752.54  
## + 상권 1 31.2 9229.0 753.54  
## - 공공기관 1 292.5 9552.7 756.16  
## - 총.인구 1 2737.5 11997.8 799.92  
## - 버스정류장 1 4729.1 13989.4 829.40

smfit3 <- summary(fit3)  
smfit3

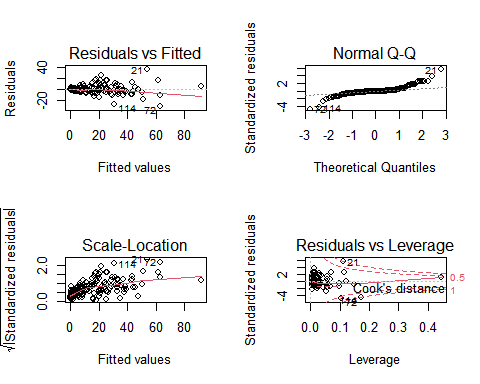
##   
## Call:  
## lm(formula = 노인시설 ~ 버스정류장 + 공공기관 + 총.인구)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -30.627 -2.017 -0.455 1.271 36.401   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.152e-01 6.704e-01 0.172 0.8637   
## 버스정류장 1.470e-01 1.500e-02 9.798 < 2e-16 \*\*\*  
## 공공기관 5.274e-01 2.165e-01 2.437 0.0158 \*   
## 총.인구 2.949e-04 3.956e-05 7.455 3.2e-12 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 7.018 on 188 degrees of freedom  
## Multiple R-squared: 0.8359, Adjusted R-squared: 0.8333   
## F-statistic: 319.3 on 3 and 188 DF, p-value: < 2.2e-16

# model diagnostics

attach(data)

## The following objects are masked from data (pos = 3):  
##   
## 경로당, 공공기관, 공공체육시설.개수, 노인.인구, 노인교실,  
## 노인복지관, 노인시설, 도시공원, 동.이름, 버스정류장, 병원, 상권,  
## 총.인구

n <- nrow(data)  
X <- model.matrix(fit3)  
p <- ncol(X)  
y.hat <- predict(fit3)  
res <- fit3$residuals  
sigma.hat <- smfit3$sigma  
ri <- rstandard(fit3)  
par(mfrow=c(2,2))  
plot(fit3)



detach(data)

re\_data = data[-c(135,143,74,43,72,53,77),]  
attach(re\_data)

## The following objects are masked from data:  
##   
## 경로당, 공공기관, 공공체육시설.개수, 노인.인구, 노인교실,  
## 노인복지관, 노인시설, 도시공원, 동.이름, 버스정류장, 병원, 상권,  
## 총.인구

l = log(re\_data$노인시설)  
l[is.infinite(l)] <- 0  
l

## [1] 3.4657359 3.2958369 2.1972246 3.3322045 2.9957323 3.7612001 3.8712010  
## [8] 0.0000000 0.0000000 0.0000000 3.4339872 2.3025851 3.7135721 0.0000000  
## [15] 3.1780538 2.3025851 3.5553481 3.4657359 2.3025851 1.3862944 4.4998097  
## [22] 2.6390573 2.7080502 2.3978953 2.4849066 0.0000000 0.0000000 0.0000000  
## [29] 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000 0.6931472 1.6094379  
## [36] 1.3862944 1.6094379 3.6888795 3.7612001 2.8903718 0.0000000 1.0986123  
## [43] 3.2958369 3.2188758 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000  
## [50] 1.0986123 0.6931472 1.0986123 3.4657359 2.8903718 0.0000000 0.0000000  
## [57] 0.0000000 0.0000000 0.0000000 1.6094379 1.7917595 1.3862944 3.2958369  
## [64] 1.0986123 1.7917595 3.5553481 3.2958369 2.3978953 2.8903718 3.3322045  
## [71] 0.0000000 2.8903718 4.0253517 3.1780538 1.7917595 3.5835189 2.1972246  
## [78] 0.6931472 0.0000000 1.3862944 0.0000000 0.6931472 0.6931472 0.0000000  
## [85] 1.0986123 2.0794415 3.5263605 0.0000000 0.0000000 0.6931472 3.4339872  
## [92] 0.0000000 0.6931472 1.6094379 0.0000000 0.0000000 0.0000000 0.0000000  
## [99] 3.3672958 1.6094379 1.0986123 0.6931472 0.6931472 2.6390573 2.9957323  
## [106] 0.0000000 1.0986123 1.0986123 1.0986123 1.3862944 2.3978953 2.3025851  
## [113] 2.9444390 1.0986123 1.0986123 0.6931472 1.6094379 0.0000000 0.0000000  
## [120] 0.0000000 0.0000000 2.9957323 1.9459101 0.0000000 2.1972246 3.2188758  
## [127] 2.3025851 3.2188758 2.7725887 2.4849066 0.0000000 0.0000000 0.0000000  
## [134] 1.6094379 2.4849066 0.0000000 2.0794415 3.5263605 3.8918203 2.9957323  
## [141] 3.7612001 3.2958369 3.7376696 2.9957323 3.5835189 3.6375862 4.3567088  
## [148] 3.7841896 2.9444390 3.5835189 1.0986123 1.3862944 3.4965076 0.0000000  
## [155] 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000 1.3862944  
## [162] 0.0000000 0.0000000 1.6094379 3.5263605 1.7917595 3.2188758 2.9957323  
## [169] 2.7725887 0.6931472 0.6931472 0.6931472 0.0000000 1.3862944 0.0000000  
## [176] 0.0000000 0.0000000 0.0000000 0.0000000 2.8903718 3.1354942 3.5263605  
## [183] 0.0000000 1.0986123 0.0000000

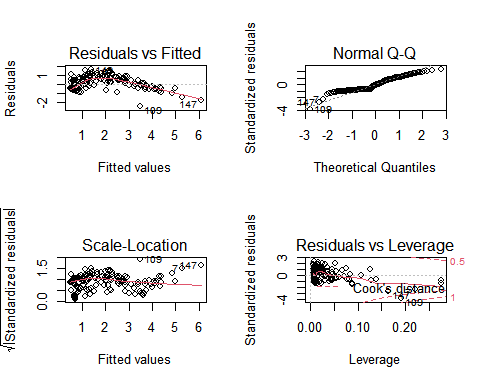
m4 <- lm(l ~ 병원 + 상권 + 도시공원 + 버스정류장 + 공공기관 + 총.인구)  
fit4 <- step(m4, direction = "both")

## Start: AIC=-121.92  
## l ~ 병원 + 상권 + 도시공원 + 버스정류장 + 공공기관 + 총.인구  
##   
## Df Sum of Sq RSS AIC  
## - 상권 1 0.001 88.738 -123.915  
## - 병원 1 0.134 88.871 -123.637  
## <none> 88.737 -121.916  
## - 도시공원 1 1.143 89.880 -121.549  
## - 공공기관 1 2.921 91.658 -117.925  
## - 총.인구 1 13.782 102.519 -97.207  
## - 버스정류장 1 34.787 123.524 -62.726  
##   
## Step: AIC=-123.91  
## l ~ 병원 + 도시공원 + 버스정류장 + 공공기관 + 총.인구  
##   
## Df Sum of Sq RSS AIC  
## - 병원 1 0.385 89.122 -125.114  
## <none> 88.738 -123.915  
## - 도시공원 1 1.170 89.908 -123.491  
## + 상권 1 0.001 88.737 -121.916  
## - 공공기관 1 2.922 91.660 -119.921  
## - 총.인구 1 14.288 103.025 -98.296  
## - 버스정류장 1 39.127 127.865 -58.335  
##   
## Step: AIC=-125.11  
## l ~ 도시공원 + 버스정류장 + 공공기관 + 총.인구  
##   
## Df Sum of Sq RSS AIC  
## <none> 89.122 -125.114  
## - 도시공원 1 1.124 90.246 -124.796  
## + 병원 1 0.385 88.738 -123.915  
## + 상권 1 0.251 88.871 -123.637  
## - 공공기관 1 2.852 91.974 -121.287  
## - 총.인구 1 15.838 104.960 -96.853  
## - 버스정류장 1 39.155 128.277 -59.740

smfit4 <- summary(fit4)  
smfit4

##   
## Call:  
## lm(formula = l ~ 도시공원 + 버스정류장 + 공공기관 + 총.인구)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2.3889 -0.5694 -0.0555 0.5680 1.5794   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 5.086e-01 7.027e-02 7.238 1.27e-11 \*\*\*  
## 도시공원 -2.467e-02 1.638e-02 -1.506 0.1337   
## 버스정류장 1.511e-02 1.699e-03 8.893 6.18e-16 \*\*\*  
## 공공기관 6.882e-02 2.867e-02 2.400 0.0174 \*   
## 총.인구 2.564e-05 4.533e-06 5.656 6.00e-08 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.7036 on 180 degrees of freedom  
## Multiple R-squared: 0.7616, Adjusted R-squared: 0.7563   
## F-statistic: 143.7 on 4 and 180 DF, p-value: < 2.2e-16

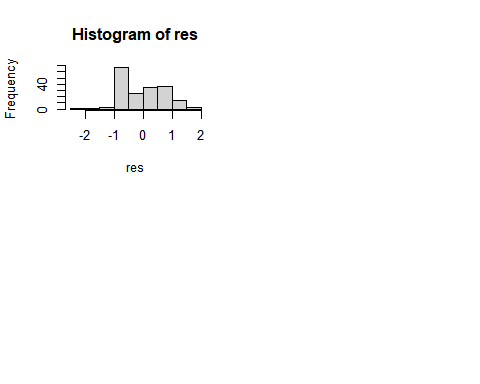
n <- nrow(re\_data)  
X <- model.matrix(fit4)  
p <- ncol(X)  
y.hat <- predict(fit4)  
res <- fit4$residuals  
sigma.hat <- smfit4$sigma  
ri <- rstandard(fit4)  
par(mfrow=c(2,2))   
plot(fit4)



shapiro.test(fit4$residuals)

##   
## Shapiro-Wilk normality test  
##   
## data: fit4$residuals  
## W = 0.94707, p-value = 2.318e-06

hist(res)  
detach(re\_data)



re\_data = data[-c(135,143,74,43,72,53,77),]  
  
attach(re\_data)

## The following objects are masked from data:  
##   
## 경로당, 공공기관, 공공체육시설.개수, 노인.인구, 노인교실,  
## 노인복지관, 노인시설, 도시공원, 동.이름, 버스정류장, 병원, 상권,  
## 총.인구

r = sqrt(re\_data$노인시설)  
r

## [1] 5.656854 5.196152 3.000000 5.291503 4.472136 6.557439 6.928203 0.000000  
## [9] 0.000000 0.000000 5.567764 3.162278 6.403124 0.000000 4.898979 3.162278  
## [17] 5.916080 5.656854 3.162278 2.000000 9.486833 3.741657 3.872983 3.316625  
## [25] 3.464102 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000  
## [33] 1.000000 1.414214 2.236068 2.000000 2.236068 6.324555 6.557439 4.242641  
## [41] 0.000000 1.732051 5.196152 5.000000 1.000000 0.000000 1.000000 1.000000  
## [49] 0.000000 1.732051 1.414214 1.732051 5.656854 4.242641 0.000000 0.000000  
## [57] 0.000000 0.000000 1.000000 2.236068 2.449490 2.000000 5.196152 1.732051  
## [65] 2.449490 5.916080 5.196152 3.316625 4.242641 5.291503 1.000000 4.242641  
## [73] 7.483315 4.898979 2.449490 6.000000 3.000000 1.414214 0.000000 2.000000  
## [81] 0.000000 1.414214 1.414214 1.000000 1.732051 2.828427 5.830952 0.000000  
## [89] 0.000000 1.414214 5.567764 1.000000 1.414214 2.236068 1.000000 0.000000  
## [97] 0.000000 1.000000 5.385165 2.236068 1.732051 1.414214 1.414214 3.741657  
## [105] 4.472136 1.000000 1.732051 1.732051 1.732051 2.000000 3.316625 3.162278  
## [113] 4.358899 1.732051 1.732051 1.414214 2.236068 0.000000 0.000000 0.000000  
## [121] 0.000000 4.472136 2.645751 0.000000 3.000000 5.000000 3.162278 5.000000  
## [129] 4.000000 3.464102 1.000000 1.000000 0.000000 2.236068 3.464102 1.000000  
## [137] 2.828427 5.830952 7.000000 4.472136 6.557439 5.196152 6.480741 4.472136  
## [145] 6.000000 6.164414 8.831761 6.633250 4.358899 6.000000 1.732051 2.000000  
## [153] 5.744563 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000  
## [161] 2.000000 0.000000 0.000000 2.236068 5.830952 2.449490 5.000000 4.472136  
## [169] 4.000000 1.414214 1.414214 1.414214 0.000000 2.000000 0.000000 0.000000  
## [177] 1.000000 0.000000 0.000000 4.242641 4.795832 5.830952 1.000000 1.732051  
## [185] 1.000000

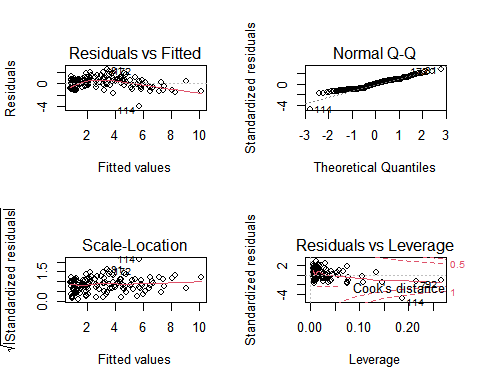
m5 <- lm(r ~ 병원 + 상권 + 도시공원 + 버스정류장 + 공공기관 + 총.인구,data = re\_data)  
fit5 <- step(m5, direction = "both")

## Start: AIC=-11.28  
## r ~ 병원 + 상권 + 도시공원 + 버스정류장 + 공공기관 + 총.인구  
##   
## Df Sum of Sq RSS AIC  
## - 상권 1 0.001 161.37 -13.279  
## - 도시공원 1 0.274 161.64 -12.967  
## - 병원 1 0.521 161.89 -12.685  
## <none> 161.37 -11.281  
## - 공공기관 1 4.176 165.55 -8.555  
## - 총.인구 1 38.062 199.43 25.897  
## - 버스정류장 1 89.340 250.71 68.230  
##   
## Step: AIC=-13.28  
## r ~ 병원 + 도시공원 + 버스정류장 + 공공기관 + 총.인구  
##   
## Df Sum of Sq RSS AIC  
## - 도시공원 1 0.284 161.66 -14.954  
## - 병원 1 1.539 162.91 -13.523  
## <none> 161.37 -13.279  
## + 상권 1 0.001 161.37 -11.281  
## - 공공기관 1 4.178 165.55 -10.550  
## - 총.인구 1 39.449 200.82 25.181  
## - 버스정류장 1 100.470 261.84 74.266  
##   
## Step: AIC=-14.95  
## r ~ 병원 + 버스정류장 + 공공기관 + 총.인구  
##   
## Df Sum of Sq RSS AIC  
## - 병원 1 1.494 163.15 -15.253  
## <none> 161.66 -14.954  
## + 도시공원 1 0.284 161.37 -13.279  
## + 상권 1 0.012 161.64 -12.967  
## - 공공기관 1 5.259 166.91 -11.031  
## - 총.인구 1 44.391 206.05 27.933  
## - 버스정류장 1 114.364 276.02 82.021  
##   
## Step: AIC=-15.25  
## r ~ 버스정류장 + 공공기관 + 총.인구  
##   
## Df Sum of Sq RSS AIC  
## <none> 163.15 -15.253  
## + 병원 1 1.494 161.66 -14.954  
## + 상권 1 0.910 162.24 -14.287  
## + 도시공원 1 0.239 162.91 -13.523  
## - 공공기관 1 5.009 168.16 -11.658  
## - 총.인구 1 49.689 212.84 31.933  
## - 버스정류장 1 114.916 278.07 81.388

smfit5 <- summary(fit5)  
smfit5

##   
## Call:  
## lm(formula = r ~ 버스정류장 + 공공기관 + 총.인구, data = re\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -4.0092 -0.8402 0.0262 0.6361 2.5213   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 8.273e-01 9.476e-02 8.730 1.66e-15 \*\*\*  
## 버스정류장 2.374e-02 2.102e-03 11.291 < 2e-16 \*\*\*  
## 공공기관 8.732e-02 3.704e-02 2.357 0.0195 \*   
## 총.인구 4.042e-05 5.444e-06 7.425 4.28e-12 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.9494 on 181 degrees of freedom  
## Multiple R-squared: 0.8271, Adjusted R-squared: 0.8242   
## F-statistic: 288.6 on 3 and 181 DF, p-value: < 2.2e-16

n <- nrow(data)  
X <- model.matrix(fit5)  
p <- ncol(X)  
y.hat <- predict(fit5)  
res <- fit5$residuals  
sigma.hat <- smfit5$sigma  
ri <- rstandard(fit5)  
par(mfrow=c(2,2))  
plot(fit5)



shapiro.test(fit5$residuals)

##   
## Shapiro-Wilk normality test  
##   
## data: fit5$residuals  
## W = 0.9635, p-value = 9.627e-05

detach(re\_data)

install.packages("lmtest",repos = "http://cran.us.r-project.org")

## 'C:/Users/user/Documents/R/win-library/4.1'의 위치에 패키지(들)을 설치합니다.  
## (왜냐하면 'lib'가 지정되지 않았기 때문입니다)

## package 'lmtest' successfully unpacked and MD5 sums checked

## Warning: cannot remove prior installation of package 'lmtest'

## Warning in file.copy(savedcopy, lib, recursive = TRUE): C:  
## \Users\user\Documents\R\win-library\4.1\00LOCK\lmtest\libs\x64\lmtest.dll를 C:  
## \Users\user\Documents\R\win-library\4.1\lmtest\libs\x64\lmtest.dll로 복사하는데  
## 문제가 발생했습니다: Permission denied

## Warning: restored 'lmtest'

##   
## The downloaded binary packages are in  
## C:\Users\user\AppData\Local\Temp\Rtmpi8nwzu\downloaded\_packages

library(lmtest)

## 필요한 패키지를 로딩중입니다: zoo

##   
## 다음의 패키지를 부착합니다: 'zoo'

## The following objects are masked from 'package:base':  
##   
## as.Date, as.Date.numeric

dwtest(fit5, alternative="greater")

##   
## Durbin-Watson test  
##   
## data: fit5  
## DW = 1.6143, p-value = 0.003737  
## alternative hypothesis: true autocorrelation is greater than 0

install.packages("lm.beta",repos = "http://cran.us.r-project.org") # 표준화 베타 계수

## 'C:/Users/user/Documents/R/win-library/4.1'의 위치에 패키지(들)을 설치합니다.  
## (왜냐하면 'lib'가 지정되지 않았기 때문입니다)

## package 'lm.beta' successfully unpacked and MD5 sums checked  
##   
## The downloaded binary packages are in  
## C:\Users\user\AppData\Local\Temp\Rtmpi8nwzu\downloaded\_packages

library(lm.beta)  
lm.beta(fit5)

##   
## Call:  
## lm(formula = r ~ 버스정류장 + 공공기관 + 총.인구, data = re\_data)  
##   
## Standardized Coefficients::  
## (Intercept) 버스정류장 공공기관 총.인구   
## 0.00000000 0.52195462 0.09431966 0.38782447

install.packages('car',repos = "http://cran.us.r-project.org") # 다중공선성 vif

## 'C:/Users/user/Documents/R/win-library/4.1'의 위치에 패키지(들)을 설치합니다.  
## (왜냐하면 'lib'가 지정되지 않았기 때문입니다)

## package 'car' successfully unpacked and MD5 sums checked  
##   
## The downloaded binary packages are in  
## C:\Users\user\AppData\Local\Temp\Rtmpi8nwzu\downloaded\_packages

library(car)

## 필요한 패키지를 로딩중입니다: carData

vif(fit5)

## 버스정류장 공공기관 총.인구   
## 2.237088 1.675930 2.856344

confint(fit5, level=0.95)

## 2.5 % 97.5 %  
## (Intercept) 6.402806e-01 1.014252e+00  
## 버스정류장 1.959056e-02 2.788750e-02  
## 공공기관 1.423101e-02 1.604157e-01  
## 총.인구 2.967685e-05 5.115981e-05

pii <- influence(fit5)$hat  
Ci <- cooks.distance(fit5)  
dfits <- dffits(fit5)  
abs(dfits)>=2\*sqrt(p/(n-p)) # 81,92,114,145,150,172,187,189

## 1 2 3 4 5 6 7 8 9 10 11 12 13   
## FALSE FALSE FALSE FALSE FALSE FALSE TRUE FALSE FALSE FALSE FALSE FALSE FALSE   
## 14 15 16 17 18 19 20 21 22 23 24 25 26   
## FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE   
## 27 28 29 30 31 32 33 34 35 36 37 38 39   
## FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE   
## 40 41 42 44 45 46 47 48 49 50 51 52 54   
## FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE   
## 55 56 57 58 59 60 61 62 63 64 65 66 67   
## FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE   
## 68 69 70 71 73 75 76 78 79 80 81 82 83   
## FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE TRUE FALSE FALSE   
## 84 85 86 87 88 89 90 91 92 93 94 95 96   
## FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE TRUE FALSE FALSE FALSE FALSE   
## 97 98 99 100 101 102 103 104 105 106 107 108 109   
## FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE   
## 110 111 112 113 114 115 116 117 118 119 120 121 122   
## FALSE FALSE FALSE FALSE TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE   
## 123 124 125 126 127 128 129 130 131 132 133 134 136   
## FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE   
## 137 138 139 140 141 142 144 145 146 147 148 149 150   
## FALSE FALSE FALSE FALSE FALSE FALSE FALSE TRUE FALSE FALSE FALSE FALSE TRUE   
## 151 152 153 154 155 156 157 158 159 160 161 162 163   
## FALSE FALSE FALSE TRUE TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE   
## 164 165 166 167 168 169 170 171 172 173 174 175 176   
## FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE TRUE FALSE FALSE FALSE FALSE   
## 177 178 179 180 181 182 183 184 185 186 187 188 189   
## FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE TRUE FALSE TRUE   
## 190 191 192   
## FALSE FALSE FALSE

pbar <- p/n  
lev.ind <- pii>2\*pbar  
lev.ind # 7,11,15,21,44,45,56,92,110,114,145,146,148,150,154,187,189

## 1 2 3 4 5 6 7 8 9 10 11 12 13   
## FALSE FALSE FALSE FALSE FALSE FALSE TRUE FALSE FALSE FALSE TRUE FALSE FALSE   
## 14 15 16 17 18 19 20 21 22 23 24 25 26   
## FALSE TRUE FALSE FALSE FALSE FALSE FALSE TRUE FALSE FALSE FALSE FALSE FALSE   
## 27 28 29 30 31 32 33 34 35 36 37 38 39   
## FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE   
## 40 41 42 44 45 46 47 48 49 50 51 52 54   
## FALSE FALSE FALSE TRUE TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE   
## 55 56 57 58 59 60 61 62 63 64 65 66 67   
## FALSE TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE   
## 68 69 70 71 73 75 76 78 79 80 81 82 83   
## FALSE FALSE FALSE FALSE FALSE FALSE FALSE TRUE FALSE FALSE FALSE FALSE FALSE   
## 84 85 86 87 88 89 90 91 92 93 94 95 96   
## FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE TRUE FALSE FALSE FALSE FALSE   
## 97 98 99 100 101 102 103 104 105 106 107 108 109   
## FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE   
## 110 111 112 113 114 115 116 117 118 119 120 121 122   
## TRUE FALSE FALSE FALSE TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE   
## 123 124 125 126 127 128 129 130 131 132 133 134 136   
## FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE   
## 137 138 139 140 141 142 144 145 146 147 148 149 150   
## FALSE FALSE FALSE FALSE FALSE FALSE FALSE TRUE TRUE FALSE TRUE FALSE TRUE   
## 151 152 153 154 155 156 157 158 159 160 161 162 163   
## FALSE FALSE FALSE TRUE TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE   
## 164 165 166 167 168 169 170 171 172 173 174 175 176   
## FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE   
## 177 178 179 180 181 182 183 184 185 186 187 188 189   
## FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE TRUE FALSE TRUE   
## 190 191 192   
## FALSE FALSE FALSE

infl.ind <- 1:n %in% order(Ci, decreasing=T)[1:2]  
names(infl.ind) <- names(lev.ind)  
infl.ind #7

## 1 2 3 4 5 6 7 8 9 10 11 12 13   
## FALSE FALSE FALSE FALSE FALSE FALSE TRUE FALSE FALSE FALSE FALSE FALSE FALSE   
## 14 15 16 17 18 19 20 21 22 23 24 25 26   
## FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE   
## 27 28 29 30 31 32 33 34 35 36 37 38 39   
## FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE   
## 40 41 42 44 45 46 47 48 49 50 51 52 54   
## FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE   
## 55 56 57 58 59 60 61 62 63 64 65 66 67   
## FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE   
## 68 69 70 71 73 75 76 78 79 80 81 82 83   
## FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE   
## 84 85 86 87 88 89 90 91 92 93 94 95 96   
## FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE   
## 97 98 99 100 101 102 103 104 105 106 107 108 109   
## FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE   
## 110 111 112 113 114 115 116 117 118 119 120 121 122   
## FALSE FALSE FALSE FALSE TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE   
## 123 124 125 126 127 128 129 130 131 132 133 134 136   
## FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE   
## 137 138 139 140 141 142 144 145 146 147 148 149 150   
## FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE   
## 151 152 153 154 155 156 157 158 159 160 161 162 163   
## FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE   
## 164 165 166 167 168 169 170 171 172 173 174 175 176   
## FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE   
## 177 178 179 180 181 182 183 184 185 186 187 188 189   
## FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE   
## 190 191 192 <NA> <NA> <NA> <NA> <NA> <NA> <NA>   
## FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

re\_data1 = data[-c(7,135,143,74,43,72,53,77),]  
  
attach(re\_data1)

## The following objects are masked from data:  
##   
## 경로당, 공공기관, 공공체육시설.개수, 노인.인구, 노인교실,  
## 노인복지관, 노인시설, 도시공원, 동.이름, 버스정류장, 병원, 상권,  
## 총.인구

r1 = sqrt(re\_data1$노인시설)  
r1

## [1] 5.656854 5.196152 3.000000 5.291503 4.472136 6.557439 0.000000 0.000000  
## [9] 0.000000 5.567764 3.162278 6.403124 0.000000 4.898979 3.162278 5.916080  
## [17] 5.656854 3.162278 2.000000 9.486833 3.741657 3.872983 3.316625 3.464102  
## [25] 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 1.000000  
## [33] 1.414214 2.236068 2.000000 2.236068 6.324555 6.557439 4.242641 0.000000  
## [41] 1.732051 5.196152 5.000000 1.000000 0.000000 1.000000 1.000000 0.000000  
## [49] 1.732051 1.414214 1.732051 5.656854 4.242641 0.000000 0.000000 0.000000  
## [57] 0.000000 1.000000 2.236068 2.449490 2.000000 5.196152 1.732051 2.449490  
## [65] 5.916080 5.196152 3.316625 4.242641 5.291503 1.000000 4.242641 7.483315  
## [73] 4.898979 2.449490 6.000000 3.000000 1.414214 0.000000 2.000000 0.000000  
## [81] 1.414214 1.414214 1.000000 1.732051 2.828427 5.830952 0.000000 0.000000  
## [89] 1.414214 5.567764 1.000000 1.414214 2.236068 1.000000 0.000000 0.000000  
## [97] 1.000000 5.385165 2.236068 1.732051 1.414214 1.414214 3.741657 4.472136  
## [105] 1.000000 1.732051 1.732051 1.732051 2.000000 3.316625 3.162278 4.358899  
## [113] 1.732051 1.732051 1.414214 2.236068 0.000000 0.000000 0.000000 0.000000  
## [121] 4.472136 2.645751 0.000000 3.000000 5.000000 3.162278 5.000000 4.000000  
## [129] 3.464102 1.000000 1.000000 0.000000 2.236068 3.464102 1.000000 2.828427  
## [137] 5.830952 7.000000 4.472136 6.557439 5.196152 6.480741 4.472136 6.000000  
## [145] 6.164414 8.831761 6.633250 4.358899 6.000000 1.732051 2.000000 5.744563  
## [153] 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 2.000000  
## [161] 0.000000 0.000000 2.236068 5.830952 2.449490 5.000000 4.472136 4.000000  
## [169] 1.414214 1.414214 1.414214 0.000000 2.000000 0.000000 0.000000 1.000000  
## [177] 0.000000 0.000000 4.242641 4.795832 5.830952 1.000000 1.732051 1.000000

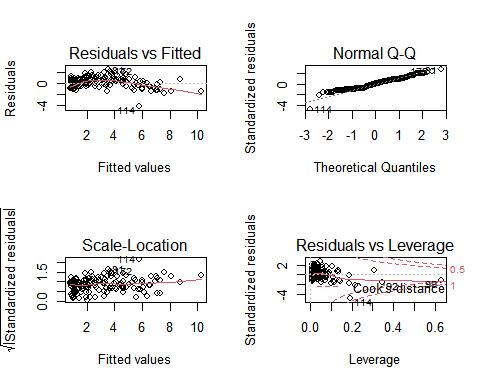
newfit <- lm(r1 ~ 병원 + 상권 + 도시공원 + 버스정류장 + 공공기관 + 총.인구,data = re\_data1) # 이상치 7 제거  
fit\_final <- step(newfit, direction = "both")

## Start: AIC=-12.61  
## r1 ~ 병원 + 상권 + 도시공원 + 버스정류장 + 공공기관 + 총.인구  
##   
## Df Sum of Sq RSS AIC  
## - 상권 1 0.001 159.23 -14.604  
## - 도시공원 1 0.433 159.66 -14.106  
## - 병원 1 0.634 159.86 -13.874  
## <none> 159.23 -12.605  
## - 공공기관 1 6.055 165.28 -7.738  
## - 총.인구 1 37.429 196.66 24.241  
## - 버스정류장 1 90.663 249.89 68.321  
##   
## Step: AIC=-14.6  
## r1 ~ 병원 + 도시공원 + 버스정류장 + 공공기관 + 총.인구  
##   
## Df Sum of Sq RSS AIC  
## - 도시공원 1 0.447 159.68 -16.088  
## <none> 159.23 -14.604  
## - 병원 1 1.879 161.11 -14.445  
## + 상권 1 0.001 159.23 -12.605  
## - 공공기관 1 6.058 165.29 -9.733  
## - 총.인구 1 38.791 198.02 23.512  
## - 버스정류장 1 101.911 261.14 74.423  
##   
## Step: AIC=-16.09  
## r1 ~ 병원 + 버스정류장 + 공공기관 + 총.인구  
##   
## Df Sum of Sq RSS AIC  
## <none> 159.68 -16.088  
## - 병원 1 1.802 161.48 -16.023  
## + 도시공원 1 0.447 159.23 -14.604  
## + 상권 1 0.016 159.66 -14.106  
## - 공공기관 1 7.171 166.85 -10.005  
## - 총.인구 1 42.633 202.31 25.456  
## - 버스정류장 1 115.086 274.76 81.779

smfit\_final <- summary(fit\_final)  
smfit\_final

##   
## Call:  
## lm(formula = r1 ~ 병원 + 버스정류장 + 공공기관 + 총.인구, data = re\_data1)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3.9907 -0.8163 0.0280 0.6580 2.5414   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 8.095e-01 9.532e-02 8.492 7.60e-15 \*\*\*  
## 병원 -3.507e-03 2.467e-03 -1.421 0.15694   
## 버스정류장 2.377e-02 2.093e-03 11.358 < 2e-16 \*\*\*  
## 공공기관 1.164e-01 4.105e-02 2.835 0.00511 \*\*   
## 총.인구 4.391e-05 6.351e-06 6.913 8.06e-11 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.9445 on 179 degrees of freedom  
## Multiple R-squared: 0.8274, Adjusted R-squared: 0.8235   
## F-statistic: 214.5 on 4 and 179 DF, p-value: < 2.2e-16

n <- nrow(re\_data)  
X <- model.matrix(newfit)  
p <- ncol(X)  
y.hat <- predict(newfit)  
res <- newfit$residuals  
sigma.hat <- newfit$sigma  
ri <- rstandard(newfit)  
par(mfrow=c(2,2))  
plot(newfit)



shapiro.test(fit5$residuals)

##   
## Shapiro-Wilk normality test  
##   
## data: fit5$residuals  
## W = 0.9635, p-value = 9.627e-05

detach(re\_data1)

Copyright ⓒ CSLEE Consortium

CSLEE Consortium의 사전 승인 없이 본 내용의 전부 또는 일부에 대한 복사, 배포, 사용을 금합니다.