**빅데이터 분석 결과 보고서**

**“응급실 혼잡도 분석을 통한 의료서비스 극대화”**

**목차**

1. **분석 배경 및 기획의도**
   1. **분석 배경**
   2. **기획 의도**
2. **분석 목표**
3. **분석 데이터**
   1. **데이터 수집**
   2. **데이터 정합성 검정**
   3. **데이터 전처리**
4. **분석 과정 및 결과**

**4.1 인구 수 별 필요 응급실 개수**

**4.2 응급실의 수와 응급실 혼잡도의 관계**

**4.3 서울시의 미래 인구 변화율**

1. **활용방안**
2. **참고문헌**

**1. 배경 및 기획 의도**

**1.1 분석 배경**

올해에도 ‘응급실 뺑뺑이’로 인해 제대로 된 치료를 받지 못하고 사망한 사고가 여러 건 발생했다. 이러한 사고가 발생하는 가장 큰 원인은 응급실 과밀화이다. 응급실의 역량을 초과하는 다수의 환자가 몰려 응급실의 혼잡도가 높아졌고, 이는 응급실에 도착한 환자들의 긴 대기시간, 응급실로 이송된 응급환자에 대한 진료 거부로 이어졌다.

정부는 이러한 응급실 과밀화를 해소하기 위해 권역 응급진료센터를 확대하고, ‘응급의료에 관한 법률’ 개정안을 시행하는 등의 노력을 기울였다. 그러나 여전히 응급실의 과밀화는 심각한 상황이다.

응급실의 과밀화를 해결하기 위해서는 각 응급실의 혼잡도를 분석하여 적절히 의료 자원이 분배될 수 있도록 해야 한다. 하지만 혼잡도를 측정하는 기준이 명확하지 않아서, 혼잡도를 낮추기 위해서는 어느 정도의 의료 자원이 더 필요한가에 대해 결정을 내리는 데 큰 어려움을 겪고 있다.

**1.2 기획의도**

응급실의 혼잡도를 분석하여 그 혼잡도에 영향을 미치는 요인을 파악한다. 이후, 혼잡도를 낮추기 위해 필요한 최소한의 의료 자원의 수치를 계산하여, 얼마나 혼잡도를 감소시키는데 도움이 되는지 확인한다.

**2. 분석 목표**

과거부터 지금까지의 인구 데이터를 통한 인구 수 변화율 데이터, 인구 수 비례 실제 필요 응급실 개수 데이터를 통해 인구 변화율에 따라 실제로 필요한 응급실의 수를 계산한다. 그리고 응급실 이용 환자 수 데이터, 응급실 평균 재실 시간, 기존 응급실 병상수, 등을 통해 혼잡도를 측정한 뒤, 실제 응급실의 수가 변화하였을 때 응급실의 혼잡도를 낮추는 데 영향을 미치는지 확인한다. 만일 응급실 혼잡도를 낮추는데 응급실의 수가 영향을 미친다면, 인구 수의 변화율에 따라 응급실을 더 개원해야 할지, 현상을 유지해야 할지 예측해 본다.

**3. 분석 데이터**

**3.1 데이터 수집**

**-인구 수 비율별 필요 응급실 개수 관련 데이터**

국내에는 인구 수 비율별 필요 응급실 개수에 대한 기준이 없기 때문에 톈진시의 사례를 참조하였다. 또한 서울시의 필요 응급실의 수와 실제 응급실의 수를 비교하면 서울시 응급실의 현황을 쉽게 이해할 수 있을 거라 판단해서, 실제 서울시의 응급의료시설 개수도 조사했다. 톈진시의 사례의 경우 PubMed Central에 있는 논문을 참조했고, 서울시의 응급의료시설 개수 데이터는 응급의료포털에 있는 응급실 병상 현황 데이터를 CSV 형태로 다운받아 저장했다.

•Urban–sub-urban–rural variation in the supply and demand of emergency medical services

\_PubMed Central

데이터 비율: 인구 60,000: 응급실 1(단위 인구: 명, 응급실: 개)

•응급실 병상 현황(시도별, 연도별)\_응급의료통계포털

데이터 지역: 서울

데이터 기간: 2016년 ~ 2022년(7년)

**- 응급실의 수와 응급실 혼잡도의 관계 관련 데이터**

응급실 혼잡도를 구하기 위해 응급실평가의 적시성 평가 기준인 ‘병상포화지수’를 참고하였다. ‘병상포화지수’를 구하기 위해서는 내원 환자 수 데이터나 응급실 평균 재실 시간, 기준 병상수 데이터가 필요했다. 그래서 응급의료통계포털에 있는 시도별 응급실 이용 현황, 응급실 재실 시간, 응급실 병상 현황 데이터를 CSV 형태로 다운받아 저장했다.

•시도별 응급실 이용 현황(성별, 연령별) \_응급의료통계포털

•응급실 재실 시간(시도별) \_응급의료통계포털

•응급실 병상 현황(시도별, 연도별) \_응급의료통계포털

데이터 지역: 서울

데이터 기간: 2016년~2022년(7년)

**- 서울시의 미래 인구 변화율 관련 데이터**

미래의 인구 변화율을 예측하기 위해서는 인구 수 데이터, 출생아 수 데이터, 사망자 수 데이터, 인구 변화율 데이터가 필요하다고 판단했다. 대한민국에서 인구 수가 가장 많은 곳은 서울이기 때문에 서울의 미래 인구 변화율을 예측해 보기로 했다. 서울 열린데이터 광장에 있는 서울\_인구밀도 통계 데이터를 CSV형태로 다운받아 저장했다. 출생아 수, 사망자 수는 월.분기.연간 인구동향(출생, 사망, 혼인, 이혼)에서 csv 형태로 다운받아 저장했다. 서울시의 미래 인구 변화율을 예측하기 위해서는 미래의 인구수, 출생아 수, 사망자 수 데이터가 필요하다고 판단했고, 이는 KOSIS 국가통계포털의 자료를 조사해 2023년부터 2037년까지 연도별로 데이터를 저장했다.

•서울\_인구밀도(구별) 통계 \_서울 열린데이터 광장

•월.분기.연간 인구동향(출생, 사망, 혼인, 이혼) \_KOSIS 국가통계포털

데이터 지역: 서울

데이터 기간: 2008~2022년(15년)

•총인구 (추계 기준) \_인구상황판 | 인구로 보는 통계(KOSIS 국가통계포털)

•장래 출생아 수 \_인구상황판 | 인구로 보는 통계(KOSIS 국가통계포털)

•장래 사망자 수 \_인구상황판 | 인구로 보는 통계(KOSIS 국가통계포털)

데이터 지역: 서울

데이터 기간: 2023~2037년(15년)

**3-2 데이터 정합성 검정**

데이터 정합성 검증은 주어진 데이터가 정확하고 일관되게 유지되는지를 확인하고 보증하는 프로세스이다. 이러한 데이터 정합성을 검증하기 위해 데이터 유효성 검사, 데이터 완전성 검사, 데이터 중복성 검사를 실시했다. 먼저 특정 테이블에서 필수적인 열(컬럼)에 값이 비어 있는지 확인해봤다. 이때 사용한 쿼리문은 아래와 같다.

SELECT \*

FROM 테이블명

WHERE 필수열1 IS NULL OR 필수열2 IS NULL OR ...;

이후 테이블1의 키에 해당하는 값이 테이블 2에 없는 경우를 찾아내어 데이터의 완전성을 검사했다. 이때 사용한 쿼리문은 아래와 같다.

SELECT \*

FROM 테이블1

LEFT JOIN 테이블2 ON 테이블1.키 = 테이블2.키

WHERE 테이블2.키 IS NULL;

마지막으로 중복된 데이터가 있는지 확인했다. 이때 사용한 쿼리문은 아래와 같다.

SELECT 열1, 열2, COUNT(\*)

FROM 테이블명

GROUP BY 열1, 열2

HAVING COUNT (\*) > 1

**3.3 데이터 전처리**

시도별 응급실 이용 현황 데이터는 서울을 포함한 전국 18개의 시도 데이터를 전부 포함하고 있었다. 그 중 서울의 데이터만 필요했기 때문에 다른 데이터들은 전부 삭제했다. 또한 성별, 연령별 데이터가 아닌 응급실을 이용한 총원의 데이터만 필요했기 때문에, 총합 인원 데이터인 계를 제외한 나머지 행들을 전부 삭제했다. 시도별 응급실 재실 시간 데이터 역시 서울을 포함한 전국 18개의 시도 데이터를 전부 포함하고 있었기 때문에 서울을 제외한 나머지 데이터는 삭제했다. 그리고 재실 시간 데이터가 2시간 미만, 2~4시간 미만, 4~6시간 미만, 6~8시간 미만, 8~12시간 미만, 12~24시간 미만, 미상으로 구간이 나뉘어 각 구간에 어느 정도의 데이터가 있는지 알 수 있었다. 하지만 미상에 속하는 데이터의 경우, 어떠한 구간에도 속하지 않는 값들의 모임이기 때문에 특이값이라고 판단해 제거했다.

응급실 재실 시간 데이터를 통해 연도별 평균재실시간 데이터를 계산하여 구하고자 했다. 평균재실시간을 구하기 위해 아래의 식을 사용하였다.

여기서 계급값은 계급을 대표하는 값으로, 각 구간의 중앙값을 의미한다. 도수는 계급에 속한 변량의 개수이다. 변량은 여러 자료를 수량으로 나타낸 것을 의미한다. 그래서 각 구간별 계급값의 중앙값을 계산한 뒤, 연도별 도수와 곱해 총합을 구한 뒤, 도수의 총합을 구해 나눠 연도별 평균 응급실 재실 시간을 구하였다.

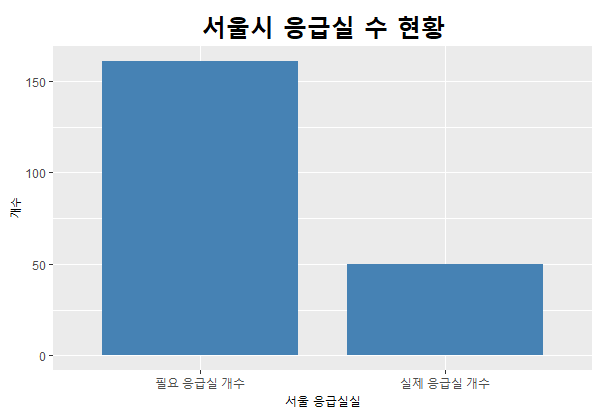
응급실 병상 현황 데이터에는 전국 18개의 시도의 응급실의 수, 총병상수, 권역응급의료센터, 지역응급의료센터, 지역응급실의 수 데이터가 존재했다. 이 중 실제로 필요한 데이터는 서울의 총병상수 데이터와 서울의 응급실의 개수였다. 시도가 서울인 응급실의 개수를 추출해 저장했고, 서울의 총 병상수 데이터는 연도별로 구분해 저장했다.

서울시 인구밀도 통계 데이터의 열행에는 동별(1), 동별(2), 동별(3), 그리고 2022, 2021, 2020… 각 연도가 2007년까지 나열되어 있었다. 동별(1)은 서울시 전체의 인구 수, 동별(2)에는 서울의 속하는 구의 데이터가, 동별(3)에는 구를 이루는 동에 대한 데이터가 있었다. 각 연도에는 인구 수, 면적, 인구밀도의 열이 있었다. 이 중 연도별 서울시 전체의 인구 수 데이터만 필요했기 때문에, 다른 데이터들은 전부 제거했다. 이후 서울시 전체의 인구수를 연도별로 구분해 저장했다. 월, 분기, 연간 인구동향(출생, 사망, 혼인, 이혼)의 조회 설정에서 행정구역은 서울을 선택했고, 종류에서는 출생아 수(명), 사망자수(명)를, 시점은 2016년부터 2022년까지로 지정했다. 조회 조건을 충족하는 출력된 데이터는 csv 형태로 다운 받아 저장했다. 2023년부터 2030년 까지의 서울 인구 수, 출생아 수, 사망자 수는 인구상황판 | 인구로 보는 통계(KOSIS 국가통계포털)에서 데이터를 갖고 와

**4. 분석 과정 및 결과**

**4.1 인구 수 별 필요 응급실 개수**

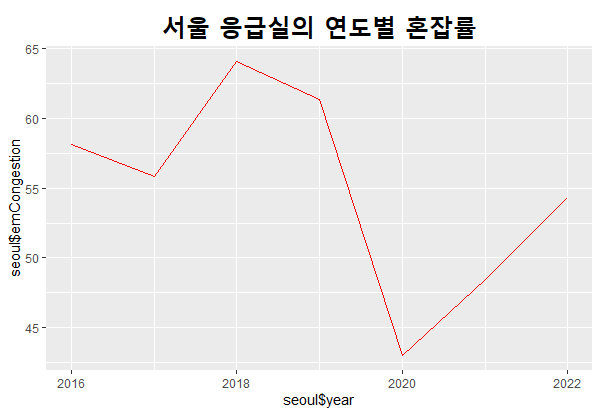
우선 톈진시의 인구 비율을 기준으로 실제 서울시에 얼마나 응급실이 부족한지 알아봤다. 응급 의료기관 병상 현황 데이터를 통해 알아본 결과, 서울시에는 현재 50개의 응급실이 존재했다. 톈진시의 인구수별 필요 응급실 개수의 비율인 인구 60,000명: 응급실 1개로 계산한 결과, 2022년 기준 서울의 인구 수는 9,667,669명 이므로 , 총 161개의 병상이 필요했다. 서울시의 경우, 약 3배의 응급실이 더 필요하다는 것을 알 수 있었다.



**4.2 응급실의 수와 응급실 혼잡도의 관계**

서울의 연도별 혼잡도를 구하기 위해 ‘병상포화지수’를 구하는 방법을 참조했다. ‘병상포화지수’는 병상 이용률을 나타내는 데이터로, 응급실의 혼잡도를 구할 때 사용된다. 병상포화지수를 구하는 식은 아래와 같다.

다만 시도별 응급실 이용 현황, 응급실 재실 시간, 응급실 병상 현황 데이터의 단위가 연 단위이기 때문에 월별일자 수가 아닌 연별 일자 수로 바꿔 응급실의 혼잡도를 구했다. 병상포화지수로 계산한 서울 응급실의 연도별 혼잡도는 아래의 그래프와 같다.

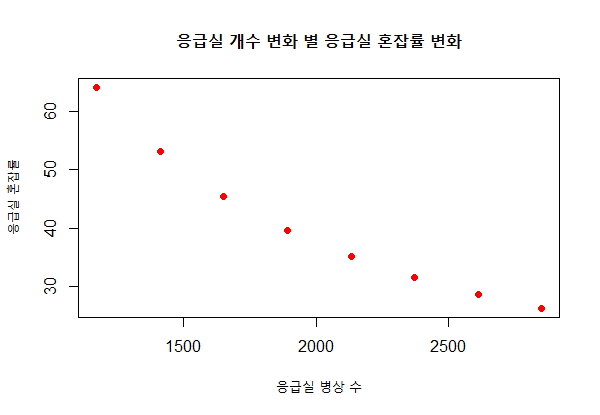


2019년 12월 말부터 2023년 6월 1일까지의 경우, 코로나19 바이러스로 인해 전국의 응급실 이용이 통제되었기 때문에 2020년부터 2022년까지의 혼잡도 데이터는 유의미한 데이터가 될 수 없다고 판단했다. 그래서 이후 응급실 수의 변화에 따른 응급실의 혼잡도 변화를 확인할 때는, 2020년부터 2022년까지의 데이터는 제외하였다.

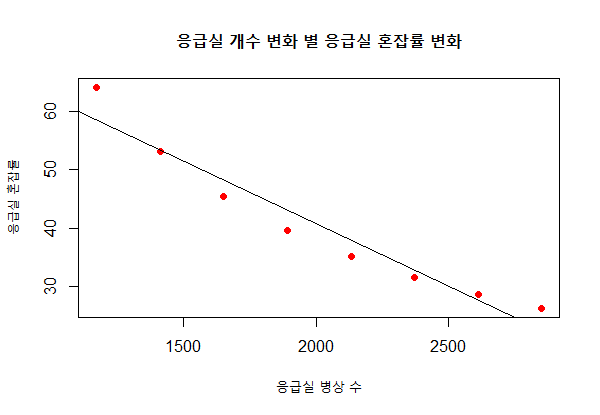
응급실 수의 변화에 따른 혼잡도의 변화를 알기 위해서는 응급실 수를 제외한 나머지 데이터인 응급실 내원 환자 수, 평균 재실 시간은 같아야 한다고 판단했다. 그래서 응급실 혼잡도가 가장 높았던 2018년을 기준으로, 그 연도의 응급실 내원 환자 수, 평균 재실 시간은 그대로 두고, 응급실 수의 변화에 따른 혼잡도의 차이를 알아보고자 했다.

응급실의 수가 변화한다는 것은 결국 응급실의 기준 병상 수가 변화한다는 것과 같은 의미라고 판단했다. 그래서 기준 병상 수를 점점 추가하면서 응급실의 혼잡도가 어떻게 변화하는지를 분석했다. 처음에는 서울에 위치한 응급의료기관 중 가장 적은 응급실 병상 수를 기준으로 점진적으로 병상을 추가하려고 했다. 하지만 응급의료기관들의 기준 응급실 병상 수의 격차가 심했다. 그래서 응급의료기관의 기존 응급실 병상 수의 평균값인 24개를 기준으로, 응급실이 10개씩 추가될 때마다 응급실의 혼잡도가 어떻게 변화하는지 확인했다.

먼저 응급실의 개수와 응급실의 혼잡도, 두 변수로 구성된 자료의 분포를 산점도를 통해 그래프로 나타내어 두 변수 사이의 연관성을 알아봤다.



응급실 병상의 수가 증가하면 응급실의 혼잡도는 감소하는 추세를 볼 수 있는데, 추세의 모양이 선이어서 응급실 병상 수와 응급실 혼잡도는 ‘선형적 관계’에 있다고 할 수 있다. 이 선형적 관계가 얼마나 선형성을 보이는지 수치상으로 표현하기 위해 회귀식을 이용하여 회귀선을 그렸다.



마지막으로, 선형성의 정도를 나타내는 상관계수값을 구했더니 -0.9680188가 나왔다. 상관계수값이 0.5보다 크거나 -0.5보다 작으면 두 변수 사이의 상관성이 높다고 판단하기 때문에, 응급실 병상 수와 응급실의 혼잡도는 상관성이 높다고 할 수 있다.

**4.3 서울시의 미래 인구 변화율**

미래의 데이터를 예측하기 위해서는 독립 변수에 해당하는 변수와 종속 변수에 해당하는 자료를 모아 관계를 분석하고 이를 예측에 사용할 수 있도록 예측 모델을 만들어야 한다. 여기서 독립 변수는 예측 모델에 영향을 미치는 요인들을 의미하고, 이러한 독립 변수에 따라 값이 결정되는 것을 종속 변수라고 한다. 서울시의 미래 인구 변화율을 예측하기 위해 서울시 인구수, 출생아 수, 사망자 수를 독립변수로 정했고, 인구 변화율을 종속변수로 지정했다. 그 후 lm()함수를 통해서 회귀식을 생성해 예측 모델을 만들었다. 이 모델을 summary()함수를 사용해 평가해 보았다.

Residuals:

Min 1Q Median 3Q Max

-0.197177 -0.074537 -0.004371 0.095269 0.162584

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -6.560e+00 7.055e+00 -0.930 0.372

deaths -2.678e-06 2.862e-06 -0.936 0.370

childBirth -2.662e-06 8.135e-06 -0.327 0.750

population 7.744e-07 6.871e-07 1.127 0.284

Residual standard error: 0.1242 on 11 degrees of freedom

Multiple R-squared: 0.8449, Adjusted R-squared: 0.8026

F-statistic: 19.97 on 3 and 11 DF, p-value: 9.304e-05

모델을 평가할 때는 p-value와 Adjusted R-squared의 값을 살펴본다. 일반적으로 p-value의 값이 0.05보다 작으면 이 모델의 신뢰수준이 95% 이상으로 신뢰할 수 있는 모델이라고 할 수 있다. Summary()함수의 결과 p-value의 값이 9.304e-05으로 이 모델은 신뢰할 만한 모델이다.

Adjusted R-squared 값은 모델의 설명력을 나타내며 0~1 사이의 값을 갖는다. 이 값이 클수록 구한 예측 모델이 현실을 잘 설명할 수 있다는 뜻이 된다. 회귀모델이 현실을 잘 설명할 수 있다는 것은 회귀모델에 의한 예측 값과 실제 측정값 사이의 오차가 적다는 뜻이다. 즉, 이 모델의 경우 현실을 80.26% 정도 설명할 수 있다는 것을 의미한다. 결과적으로 이 모델로 예측한 데이터는 신뢰할 만한 데이터라고 할 수 있다.

이 모델을 통해 2016년부터 2030년까지의 서울시 인구수 변화율을 예측하여 그래프로 그린 결과는 다음과 같다.

텍스트, 도표, 라인, 그래프이(가) 표시된 사진

자동 생성된 설명

위의 그래프를 보면 알 수 있듯이, 서울시의 인구수는 앞으로 계속 감소할 것이다. 하지만 위에서 언급 했듯이, 서울의 실제 응급실의 수는 인구수 대비 필요 응급실의 수보다 현격히 부족하다. 따라서 앞으로의 인구 수 변화율에 맞게 점진적으로 응급실의 수를 늘려가야 한다.

**5. 활용 방안**

특정 지역에 응급실을 개원하고자 할 때, 그 지역의 인구 변화율과 인구 수 대비 실제 필요한 응급실의 수를 계산하여 몇 개의 응급실이 필요할지 예측해 볼 수 있다. 또한 인구 수가 밀집되어 응급실 혼잡도가 높은 지역에 혼잡도를 낮추기 위해서는 최소한 몇 개의 응급실이 더 추가로 필요할지 계산할 때도 유용하게 사용할 수 있다.

**6. 참고 문헌**

[1] Urban–sub-urban–rural variation in the supply and demand of emergency medical services \_Li, Li, Geng, Liu, Liu, Fan 및 Cao

[2] 2023년 응급의료기관 평가기준집\_보건복지부

[3] 인구 성장률 그래프\_국가지표체계

[4] 응급실 병상 현황(시도별, 연도별)\_응급의료통계포털

[5] 응급실 재실 시간(시도별) \_응급의료통계포털

[6] 시도별 응급실 이용 현황(성별, 연령별) \_응급의료통계포털

[7] 서울\_인구밀도(구별) 통계 \_서울 열린데이터 광장

[8] 장래 출생아 수 \_인구상황판 | 인구로 보는 통계(KOSIS 국가통계포털)

[9] 장래 사망자 수 \_인구상황판 | 인구로 보는 통계(KOSIS 국가통계포털)

[10] 총인구 (추계 기준) \_인구상황판 | 인구로 보는 통계(KOSIS 국가통계포털)

[11] 월.분기.연간 인구동향(출생, 사망, 혼인, 이혼) \_KOSIS 국가통계포털

[12] 장래인구추계 \_통계청