

서울시 우울증 데이터 분석을 통한 정신건강 개선 방안 연구

김지은¹, 김의준², 김관빈³, 이가영⁴, 송병진⁵
¹ 한국폴리텍대학 원주 캠퍼스 의료공학과 학부생
² 한국폴리텍대학 원주 캠퍼스 의료공학과 학부생
³ 한국폴리텍대학 원주 캠퍼스 의료공학과 학부생
⁴ 한국폴리텍대학 원주 캠퍼스 의료공학과 학부생
⁵ 한국폴리텍대학 원주 캠퍼스 의료공학과 교수

rlawldms6775@naver.com, soakaeofhq@naver.com, gg86968773@gmail.com, dlrkud99@naver.com, bjsong@kopo.ac

Analyzing depression data in Seoul to study ways to improve mental health.

Jieun Kim¹, Uijun Kim², Gwanbin Kim³, GaYoung Kim⁴, Byung-Jin Song⁵

¹Dept. of Medical Engineering, KOREA POLYTECHNICS

²Dept. of Medical Engineering, KOREA POLYTECHNICS

³Dept. of Medical Engineering, KOREA POLYTECHNICS

⁴Dept. of Medical Engineering, KOREA POLYTECHNICS

⁵Dept. of Medical Engineering, KOREA POLYTECHNICS

요 약

본 연구는 서울시의 우울증 진단 경험률 통계 데이터를 분석하여 지역별 우울증 발생 패턴과 인구 통계적 특성을 파악하고자 하였습니다. 결측치 처리와 이상치 조정을 통한 데이터 전처리 후, 연령대와 성별, 교육 수준 등 다양한 변수에 따른 우울증 진단률의 분포를 탐색적 데이터 분석을 통해 시각화하였습니다. 특히 여성과 고령 인구에서 높은 우울증 경험률을 관찰하였으며, 이를 통해 맞춤형 정신건강 개선 방안을 제시하고자 하였습니다. 본 연구는 정책 입안자와 보건 전문가들에게 유용한 인사이트를 제공하고, 효과적인 우울증 관리 및 예방 전략 개발에 조금이라도 기여할 것으로 기대됩니다.

1. 서론

연구의 배경 및 목적

본 연구는 현대사회에서 우울증이 증가하는 추세를 고려할 때, 우울증의 치료도 중요하지만 우울증 발생 이전의 예방이 중요하여[1], 서울시 우울증 진단 경험률 통계와 2014년 이후 서울시 우울증 경험률 통계를 통합한 데이터를 기반으로 수행되었습니다. 이 데이터셋은 서울시의 다양한 구별로 구분되어 있으며, 연령대, 교육 수준, 직업군별로 세분화되어 우울증 진단 경험률을 포함하고 있습니다. 데이터 전처리 과정을 통해 결측치의 처리, 이상치의 조정 및 필요한 데이터 변환 작업이 이루어졌습니다. 이는 분석의 정확도를 높이고 데이터 간의 비교를 용이하게 하기 위함입니다. 대한민국은 우울증과 같은 정신적인 질환의 발병률이 OECD 국가 중 최고 수준이고[2], 우울증은 전 세계적으로 많은 사람들이 겪고 있는 정신 건강 문제 중 하나로, 적절한 조치와 정책이 요구됩니다. 본 데이터 분석은 우울증 진단 경험률의 지역적 분포와 인구 통계적 특성을 이해함으로써, 보다 효과적인 정책 수립과 자원 배분에 기여할 수 있을 것입니다. 또한, 우울증 예방 및 관리를 위한 기초자료로 활용될 수 있습니다. 본 보고서는 이 데이터를 깊이 있게 분석하여, 서울시 내에서 우울증 진단 경험률에 영향을 미치는 주요 요인들을 밝히고자 합니다. 이를 통해 우울증 관리 및 예방을 위한 실질적인 방안을 도출할 예정입니다.

데이터 전처리

데이터 전처리는 분석의 품질을 결정하는 중요한 단계입니다. 본 연구에서 사용된 서울시 우울증 진단 및 경험률 데이터의 전처리 과정은 다음과 같습니다.

결측치 처리

데이터셋을 검토한 결과, 일부 변수에서 결측치가 발견되었습니다. 결측치 처리 방법으로는 다음과 같은 접근이 이루어졌습니다:

- 제거: 소수의 결측치가 있는 경우, 해당 데이터 포인트를 전체 데이터셋에서 제거하였습니다. 이 방법은 데이터의 양이 충분히 많을 때 유용합니다.

- 대체: 특정 변수의 결측치가 비교적 많거나 중요한 정보를 포함하는 경우, 중앙값, 평균값 또는 최빈값 등으로 결측치를 대체하였습니다. 특히, 연령대나 지역과 같이 카테고리 명칭이 명확한 경우 최빈값 대체가 적절할 수 있습니다.

이상치 탐지 및 처리

이상치는 데이터 분석 결과를 왜곡할 수 있으므로 적절한 처리가 필요합니다. 다음과 같은 방법으로 이상치를 탐지하고 처리하였습니다:

데이터 변환 및 정규화

모델의 성능을 최적화하고 변수 간의 비교를 용이하게 하기 위해 다음과 같은 데이터 변환 및 정규화 작업을 수행하였습니다:

- 로그 변환: 우울증 진단률과 같이 편향된 분포를 가진 데이터에 대해 로그 변환을 적용하여 분포의 왜곡을 줄였습니다.

- 표준화 (Z-점수 변환): 각 변수의 값을 평균 0, 표준편차 1 이 되도록 변환하여, 변수 간의 스케일 차이를 제거하였습니다. 이는 특히 다변량 분석에서 중요한 전처리 단계입니다. 이러한 전처리 과정을 통해 데이터의 질을 향상시키고, 분석 결과의 신뢰성을 높일 수 있습니다.

2. 탐색적 데이터 분석(EDA)

기술 통계

기술 통계는 데이터의 중심 경향성 및 분산성을 측정하여 전반적인 데이터의 특성을 파악하는 데 도움을 줍니다.

데이터 분포

- 연령대별, 교육 수준별, 직업별 우울증 지표 분포: 히스토그램, 박스 플롯 등을 사용하여 각 분류 기준에 따른 우울증 진단률의 분포를 시각적으로 파악합니다.

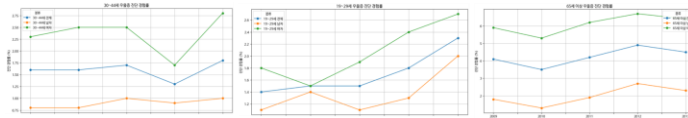
상관 분석

- 변수 간의 상관 관계 분석: 데이터셋의 여러 변수 간의 상관계수를 계산하여 어떤 변수들이 서로 강한 관계를 가지는지 분석합니다. 이는 피어슨 상관계수 또는 스피어만 순위 상관계수를 통해 수행될 수 있습니다.

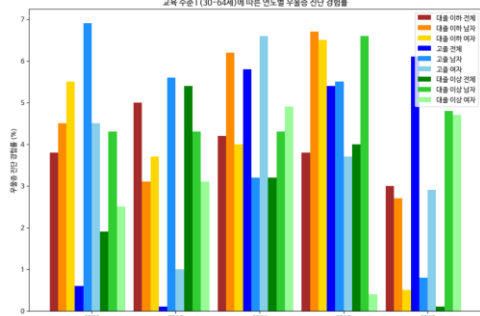
지역별 분석

- 서울시 내 구별 우울증 진단률 차이 분석: 각 구별 우울증 진단률을 비교하여 지역적 차이를 파악합니다. 지도 시각화 또는 바 차트를 활용하여 각 구의 우울증 진단률을 시각적으로 비교할 수 있습니다.

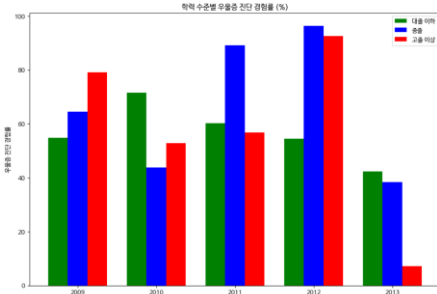
이러한 분석을 통해 데이터 내에서 숨겨진 패턴이나 인사이트를 발견하고, 이를 바탕으로 보다 구체적인 연구 질문을 설정하거나 가설을 수립하는 데 도움이 됩니다. 데이터의 이해를 돕기 위한 시각적 자료와 통계적 수치를 제공함으로써, 복잡한 데이터셋에서 유의미한 정보를 추출할 수 있습니다.



(그림 1) 19~29세, 30~44세, 45~64세 남자, 여자, 전체 우울증 진단 경험률



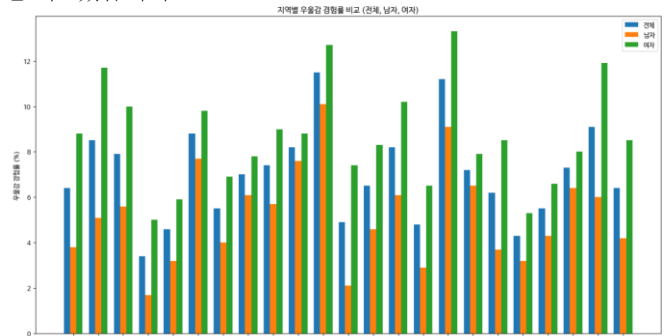
(그림 2) 교육 수준 I (30~64세)에 따른 연도별 우울증 진단 경험률



(그림 3) 학력 수준별 우울증 진단 경험률 (%)

3. 결론

그림1과 그림2, 그림3에서 알 수 있듯이 여성이 남성에 비해 우울증 진단 경험률이 현저히 높습니다. 이는 여성이 남성보다 우울증을 경험할 가능성이 더 높음을 시사하며, 이는 다양한 원인에 기인할 수 있습니다. 예를 들어, 사회적, 호르몬적, 생물학적 요인 등이 여성의 우울증 진단률을 높일 수 있습니다. 연구결과로도 여성이 남성보다는 더 사회적으로 억압받고 왜곡된 현실을 반영하고 있다고 할 수 있다[3]. 연도별 추세를 보면 우울증 진단 경험률은 연도별로 변동이 있는데, 특히 2012년에 전체적으로 가장 높은 경험률을 보여주고 있습니다. 이러한 변동은 우울증 진단 기준의 변화, 인식의 개선, 경제적 혹은 사회적 변화 등 여러 외부 요인의 영향을 받을 수 있습니다. 개인에 따라서는 우울증 유형과 심각도 및 지속기간도 사람에 따라 다르다[4]. 연령대별 비교에서는 “65세 이상”의 데이터만으로 볼 때 연령이 높아질수록 우울증에 더 취약할 수 있음을 암시합니다[5]. 고령화와 관련된 여러 건강 문제나 사회적 고립, 경제적인 문제 등이 우울증 발생에 영향을 줄 수 있습니다.



(그림 4) 지역별 우울감 경험률 비교 (전체, 남자, 여자)

그림 4는 2021년 기준으로 각 지역과 연령대별 우울감 경험률이 제공되고 있습니다. 전체적으로 여성의 우울감 경험률이 남성보다 높은 경향이 있는 것으로 나타나지만, 이러한 차이가 인과적 관계를 의미하는지 아니면 단순 상관관계인지는 명확하지 않습니다. 인과관계를 정확히 파악하기 위해서는 추가적인 변수를 포함한 다변량 분석이 필요합니다. 본 연구는 서울시 우울증 진단률 데이터를 분석하여 여성과 고령자에서 높은 우울증 비율을 확인했습니다. 이 결과는 정책 개발에 있어 특정 집단에 초점을 맞춘 맞춤형 정신건강 지원이 필요함을 시사합니다. 연구는 우울증 관리와 예방을 위한 실질적인 방안 제시에 기여할 것으로 기대됩니다.

참고문헌

[1] Eun-seon Joo, Young-im Cho, Dan-bi Kim, Yu-jin Kang. (2017). Qualitative study on depression experience and recovery process. Journal of the Korea Contents Association, 17(7), 505-526.

[2] Da-Young Shin, Ha-Young Oh, “Analysis of Chatbot Research Trends for Diagnosis and Treatment of Depression,” Journal of the Korea Institute of Information and Communication, Vol. 27, No. 3, pp. 337-347, March 2023./jkiice.2023.27.3.337 (Korean).

[3] Soojin Kim, “Changes in Household Medical Cost Burden and Implications of Strengthening Health Insurance Coverage,” Health and Welfare Forum, No. 289, November 2020, pp. 38-52 (15 pages in total), Korea Institute for Health and Social Research. (Korean).

[4] Choi, Seung Eun, “Development of a cognitive behavioral program to alleviate depression in the elderly,” Graduate School of Interdisciplinary Theology, Hannam University, 2010.

[5] Jeon Junho, “The effect of community environment satisfaction on depression among older adults: focusing on the moderating effects of age and gender,” Graduate School of Social Work, Gwangju National University, 2023.