



자살률에 영향을 미치는 경제적 사회적 요인 분석



<https://github.com/YonseiDatabase-SJ-DH>

디지털헬스케어학부 전대한

소프트웨어학부 황선준



Table of Contents

Introduction

Data Collection

Methodology

Analysis & Results

Conclusion

QnA



Introduction

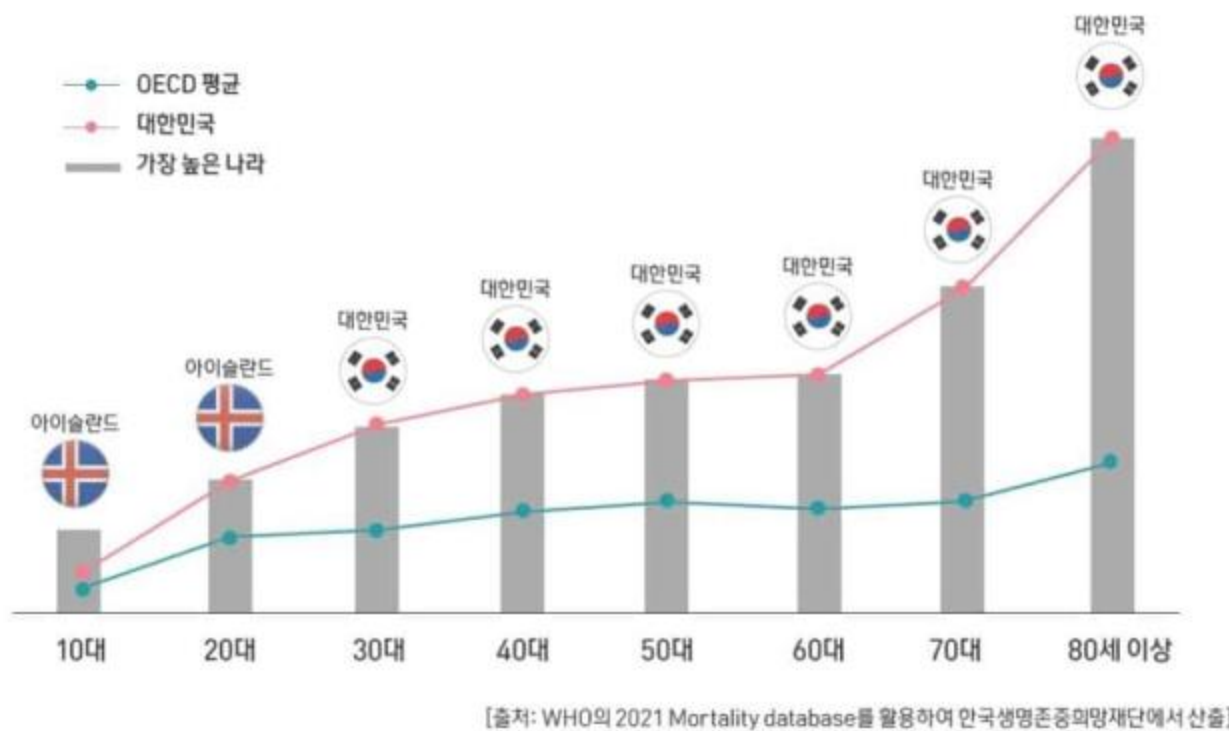
01



주제 선정배경

2 OECD 회원국 연령별 자살률

30대 이상 자살률 OECD 회원국 중 한국이 가장 높음

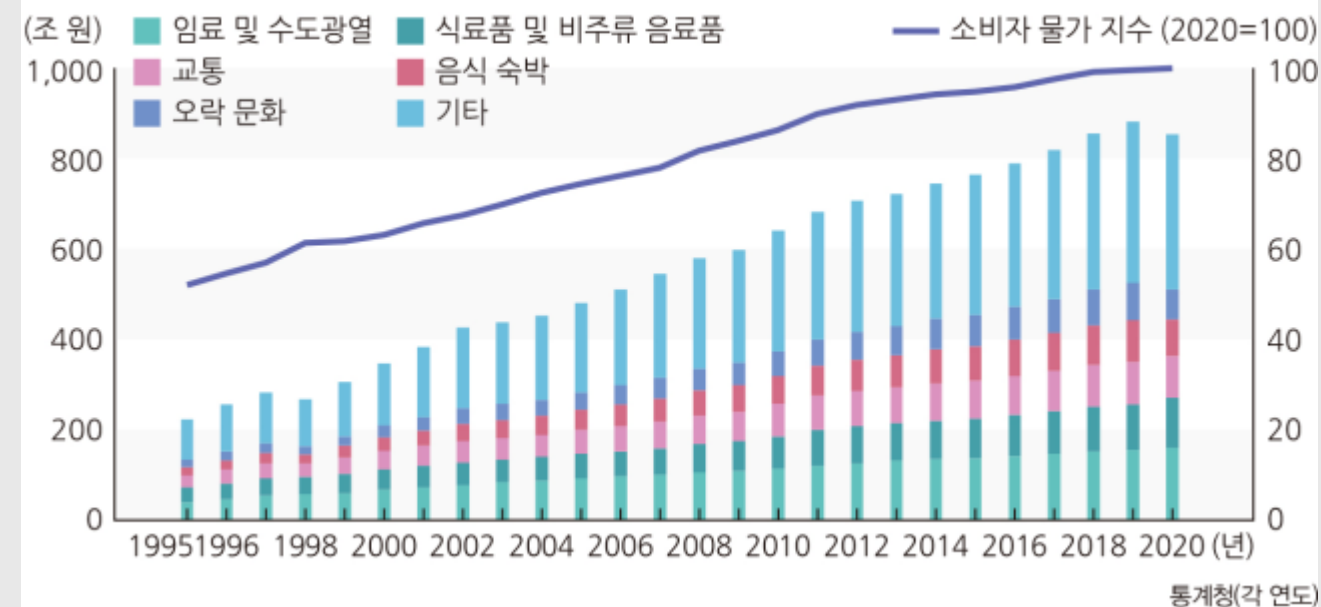


- 자살률은 사회의 경제적, 심리적, 환경적 문제를 반영하는 중요한 지표
- 최근 경제적 불황과 사회적 불안정이 심화되면서 자살률이 증가하는 추세를 보이고 있다.
- 이를 해결하기 위해 자살률에 영향을 미치는 요인을 파악하고 분석하는 것이 필요해 보여 이 주제를 선정하게 되었다.
- 따라서 본 프로젝트에서는 자살률과 다양한 경제적·사회적 지표 간의 관계를 분석하여 주요 요인을 식별하고자 하였다.



프로젝트 목표

국내 소비 지출 및 소비자 물가 지수 변화(1995-2020년)



- 다양한 경제적 지표와 사회적 지표를 활용하여 자살률에 영향을 줄 수 있는 주요 요인을 분석한다.
- Python과 SQL을 통해 데이터 전처리를 진행하고, 상관분석을 실시하여 각 변수 간의 관계를 파악한다.
- 분석 결과를 바탕으로 정책적 시사점을 제시하고, 자살 예방을 위한 인사이트를 도출하는 것을 목표로 한다.



Data Collection

02

사용한 데이터 소개

경제적 지표

- 실업률: 일자리를 잃은 인구 비율로, 경제 불안정성을 나타낸다.
- 가계 소득 수준: 가구당 평균 소득을 측정하여 경제적 여유를 평가한다.
- 빈곤율: 전체 인구 중 빈곤 상태에 있는 사람들의 비율이다.
- 소득 불평등 지수: 소득 분배의 정도를 나타낸다.

사회적 지표

- 이혼율: 결혼 대비 이혼의 비율로, 가정의 불안정성을 반영한다.
- 가정 폭력 신고 건수: 가정 내 폭력 문제의 심각성을 나타낸다.
- 교육 수준: 인구의 학력 분포로, 사회적 자원의 접근성을 평가한다.
- 인구 밀도: 특정 지역의 인구 집중도를 나타낸다.



사용한 데이터 소개

건강관련 지표

- 정신 건강 문제 발생률: 우울증 등 정신 질환 진단 비율이다.
- 알코올 및 약물 남용률: 중독 문제의 심각성을 반영한다.
- 정신과 입원율: 정신과 질환으로 입원한 환자 비율이다.
- 자살 시도 관련 병원 방문 건수: 자살 시도로 병원을 방문한 사례 수이다.

문화적/지역적 지표

- 종교 참여율: 종교 활동에 참여하는 인구 비율로, 정신적 안정감의 지표가 될 수 있다.
- 사회적 지원 수준: 가족 구성원 수, 공동체 활동 참여율 등을 통해 평가된다.
- 자살에 대한 사회적 낙인: 자살을 둘러싼 문화적 요소와 사회적 인식을 나타낸다.



사용한 데이터 소개

정부 정책 및 법률 지표

- 정신 건강 지원 예산: 정부의 정신 건강 지원을 위한 예산 규모이다.
- 자살 예방 프로그램: 자살 예방을 위한 정부 및 민간의 프로그램 존재 여부
- 정신 건강 서비스 접근성: 병원 및 상담 서비스의 접근 가능성을 평가한다.



데이터 출처

- 원래 공공기관에서 데이터를 수집할 계획이었으나, 구체적인 데이터 확보에 어려움이 있었다.
(민감한 개인정보도 포함되어 있기에)
- 따라서 분석의 일관성을 위해 Python을 이용해 임의 생성 데이터를 사용하였다.



Python으로 임의 데이터 생성하기

임의 데이터 제작하기

데이터를 만들기 쉽지 않았기에, 직접 python으로 데이터를 제작.

python 코드는 간단히 설명 방법만 소개.

```
import numpy as np
import pandas as pd
import random

np.random.seed(42)
years = list(range(2000, 2024))
```



Python으로 임의 데이터 생성하기

경제적 지표 생성

```
unemployment_rate = np.random.uniform(2.0, 10.0, len(years))
household_income = np.random.uniform(30000, 70000, len(years))
poverty_rate = np.random.uniform(5.0, 20.0, len(years))
gini_index = np.random.uniform(0.3, 0.6, len(years))
```

unemployment_rate: 2.0%~10.0% 사이의 실업률 데이터

household_income: \$30,000~\$70,000 사이의 가계 소득 데이터 생성

poverty_rate: 5.0%~20.0% 사이의 빈곤율 데이터 생성

gini_index: 0.3~0.6 사이의 지니계수 (소득 불평등 지표)



Python으로 임의 데이터 생성하기

사회적 지표 생성

```
divorce_rate = np.random.uniform(0.5, 5.0, len(years))
domestic_violence_cases = np.random.randint(50, 500, len(years))
education_level = np.random.uniform(0.7, 1.0, len(years))
population_density = np.random.randint(100, 3000, len(years))
```

divorce_rate: 0.5%~5.0% 사이의 이혼률 데이터

domestic_violence_cases: 가정 폭력 건수(50~500건)

education_level: 0.7~1.0의 학력 수준(고등학교 졸업 비율)

population_density: 100~3000명 사이의 인구 밀도



Python으로 임의 데이터 생성하기

건강 관련 지표 생성

```
mental_health_issues = np.random.uniform(1.0, 10.0, len(years))
alcohol_drug_abuse = np.random.uniform(0.5, 7.0, len(years))
psychiatric_hospitalization_rate = np.random.uniform(0.1, 2.0, len(years))
suicide_attempts = np.random.randint(10, 200, len(years))
```

mental_health_issues: 1.0%~10.0% 정신 건강 문제 발생률
alcohol_drug_abuse: 알코올 및 약물 남용률 (0.5%~7.0%)
psychiatric_hospitalization_rate: 0.1~2.0% 정신과 입원율
suicide_attempts: 자살 시도 건수 (10~200건)



Python으로 임의 데이터 생성하기

데이터 프레임 생성

```
data = pd.DataFrame({
    "Year": years,
    "Unemployment Rate": unemployment_rate,
    "Household Income": household_income,
    "Poverty Rate": poverty_rate,
    "Gini Index": gini_index,
    "Divorce Rate": divorce_rate,
    "Domestic Violence Cases": domestic_violence_cases,
    "Education Level": education_level,
    "Population Density": population_density,
    "Mental Health Issues": mental_health_issues,
    "Alcohol Drug Abuse": alcohol_drug_abuse,
    "Psychiatric Hospitalization Rate": psychiatric_hospitalization_rate,
    "Suicide Attempts": suicide_attempts
})

data.to_csv("random_data.csv", index=False)
```



Python으로 임의 데이터 생성하기

데이터 프레임 생성

```
data = pd.DataFrame({
    "Year": years,
    "Unemployment Rate": unemployment_rate,
    "Household Income": household_income,
    "Poverty Rate": poverty_rate,
    "Gini Index": gini_index,
    "Divorce Rate": divorce_rate,
    "Domestic Violence Cases": domestic_violence_cases,
    "Education Level": education_level,
    "Population Density": population_density,
    "Mental Health Issues": mental_health_issues,
    "Alcohol Drug Abuse": alcohol_drug_abuse,
    "Psychiatric Hospitalization Rate": psychiatric_hospitalization_rate,
    "Suicide Attempts": suicide_attempts
})

data.to_csv("random_data.csv", index=False)
```


SQL로 테이블 생성

테이블 생성

```
CREATE TABLE 경제지표 (  
    연도 INT,  
    실업률 FLOAT,  
    가계_소득 FLOAT,  
    빈곤율 FLOAT,  
    지니계수 FLOAT  
);
```

```
CREATE TABLE 사회지표 (  
    연도 INT,  
    이혼율 FLOAT,  
    가정_폭력_건수 INT,  
    교육_수준 FLOAT,  
    인구_밀도 FLOAT  
);
```

```
CREATE TABLE 건강지표 (  
    연도 INT,  
    정신_건강_문제를 FLOAT,  
    알코올_약물_남용률 FLOAT,  
    정신과_입원을 FLOAT,  
    자살_시도_건수 INT  
);
```

- 연도를 기본 키로 사용해 세 개의 테이블을 연계할 수 있도록 설계 했다.
- 각 테이블에는 경제적, 사회적, 건강 관련 지표들이 포함된다.

SQL로 테이블 생성

데이터 삽입

경제 지표 데이터 삽입

```
INSERT INTO 경제지표 VALUES (2009, 5.62, 39170.23, 19.11, 0.50);
INSERT INTO 경제지표 VALUES (2010, 9.66, 45212.11, 12.56, 0.39);
INSERT INTO 경제지표 VALUES (2011, 8.12, 56237.82, 10.98, 0.46);
INSERT INTO 경제지표 VALUES (2012, 7.19, 51597.25, 24.23, 0.46);
INSERT INTO 경제지표 VALUES (2013, 4.09, 44561.46, 24.48, 0.36);
INSERT INTO 경제지표 VALUES (2014, 3.89, 39574.29, 21.35, 0.41);
INSERT INTO 경제지표 VALUES (2015, 6.45, 47850.88, 15.74, 0.47);
INSERT INTO 경제지표 VALUES (2016, 5.37, 52319.63, 16.23, 0.44);
INSERT INTO 경제지표 VALUES (2017, 7.83, 48450.12, 13.67, 0.48);
INSERT INTO 경제지표 VALUES (2018, 5.95, 57643.91, 14.78, 0.42);
INSERT INTO 경제지표 VALUES (2019, 6.23, 53021.54, 18.23, 0.49);
INSERT INTO 경제지표 VALUES (2020, 4.50, 50000.00, 15.20, 0.45);
INSERT INTO 경제지표 VALUES (2021, 7.00, 52000.50, 17.50, 0.43);
INSERT INTO 경제지표 VALUES (2022, 5.30, 61000.75, 12.30, 0.40);
INSERT INTO 경제지표 VALUES (2023, 6.10, 65000.90, 11.80, 0.38);
```


SQL로 테이블 생성

데이터 삽입

사회 지표 데이터 삽입

```
INSERT INTO 사회지표 VALUES (2009, 2.55, 262, 0.72, 3439);
INSERT INTO 사회지표 VALUES (2010, 2.09, 518, 0.79, 1984);
INSERT INTO 사회지표 VALUES (2011, 4.31, 388, 0.80, 1363);
INSERT INTO 사회지표 VALUES (2012, 2.43, 478, 0.92, 3290);
INSERT INTO 사회지표 VALUES (2013, 2.12, 360, 0.89, 3812);
INSERT INTO 사회지표 VALUES (2014, 3.50, 290, 0.87, 3150);
INSERT INTO 사회지표 VALUES (2015, 1.75, 310, 0.83, 2890);
INSERT INTO 사회지표 VALUES (2016, 3.10, 480, 0.85, 2500);
INSERT INTO 사회지표 VALUES (2017, 2.80, 420, 0.88, 2200);
INSERT INTO 사회지표 VALUES (2018, 3.60, 450, 0.90, 2800);
INSERT INTO 사회지표 VALUES (2019, 2.70, 410, 0.86, 2900);
INSERT INTO 사회지표 VALUES (2020, 2.10, 120, 0.85, 1000);
INSERT INTO 사회지표 VALUES (2021, 3.00, 300, 0.88, 1800);
INSERT INTO 사회지표 VALUES (2022, 3.20, 350, 0.90, 2000);
INSERT INTO 사회지표 VALUES (2023, 1.90, 290, 0.87, 2300);
```


SQL로 테이블 생성

데이터 삽입

건강 지표 데이터 삽입

```
INSERT INTO 건강지표 VALUES (2009, 12.56, 3.86, 1.91, 179);
INSERT INTO 건강지표 VALUES (2010, 7.29, 1.99, 1.23, 102);
INSERT INTO 건강지표 VALUES (2011, 5.77, 3.05, 2.44, 221);
INSERT INTO 건강지표 VALUES (2012, 7.90, 4.84, 2.42, 267);
INSERT INTO 건강지표 VALUES (2013, 6.61, 8.36, 1.00, 209);
INSERT INTO 건강지표 VALUES (2014, 8.20, 6.15, 1.50, 220);
INSERT INTO 건강지표 VALUES (2015, 7.15, 5.75, 1.80, 190);
INSERT INTO 건강지표 VALUES (2016, 9.10, 7.25, 1.20, 250);
INSERT INTO 건강지표 VALUES (2017, 10.50, 5.80, 2.10, 300);
INSERT INTO 건강지표 VALUES (2018, 11.00, 6.50, 1.70, 280);
INSERT INTO 건강지표 VALUES (2019, 10.20, 5.60, 1.60, 260);
INSERT INTO 건강지표 VALUES (2020, 8.50, 5.20, 1.30, 150);
INSERT INTO 건강지표 VALUES (2021, 9.00, 4.80, 1.10, 170);
INSERT INTO 건강지표 VALUES (2022, 7.50, 3.90, 1.40, 190);
INSERT INTO 건강지표 VALUES (2023, 6.80, 4.50, 1.20, 180);
```


SQL로 테이블 생성

데이터 삽입

자살률 데이터 삽입

```
INSERT INTO 자살률_데이터 VALUES (2009, 17.36);
INSERT INTO 자살률_데이터 VALUES (2010, 22.65);
INSERT INTO 자살률_데이터 VALUES (2011, 22.67);
INSERT INTO 자살률_데이터 VALUES (2012, 20.72);
INSERT INTO 자살률_데이터 VALUES (2013, 11.81);
INSERT INTO 자살률_데이터 VALUES (2014, 19.50);
INSERT INTO 자살률_데이터 VALUES (2015, 18.75);
INSERT INTO 자살률_데이터 VALUES (2016, 21.30);
INSERT INTO 자살률_데이터 VALUES (2017, 23.50);
INSERT INTO 자살률_데이터 VALUES (2018, 20.00);
INSERT INTO 자살률_데이터 VALUES (2019, 18.80);
INSERT INTO 자살률_데이터 VALUES (2020, 20.00);
INSERT INTO 자살률_데이터 VALUES (2021, 21.50);
INSERT INTO 자살률_데이터 VALUES (2022, 19.00);
INSERT INTO 자살률_데이터 VALUES (2023, 18.50);
```



Methodology

03





분석 방법

상관분석이란?

- '상관분석'은 두 변수 간의 관계(연관성)를 파악하기 위한 통계적 방법이다.
- 특히 변수 간에 선형 관계가 있는지, 그리고 관계의 강도와 방향이 어떻게 되어 있는지 분석한다.
- 상관계수는 -1에서 1 사이의 값을 가진다.
 - 1에 가까울수록 두 변수 간의 강한 양의 관계 (한 변수가 증가할 때 다른 변수도 증가)
 - -1에 가까울수록 강한 음의 관계 (한 변수가 증가할 때 다른 변수는 감소)
 - 0에 가까울수록 두 변수 간에 관계가 거의 없다.
- 대표적인 상관계수 방법
 - Pearson 상관계수
 - Spearman 상관계수



분석 방법

Pearson 방법

- 정의: 두 연속형 변수 간의 선형 관계를 측정한다.
- 가정
 - 데이터가 정규분포를 따른다고 가정한다.
 - 변수 간의 관계가 선형적이어야 한다.
- 계산 방법
 - 공분산을 각 변수의 표준편차로 나눈 값이다.
- 해석
 - +1 : 완벽한 양의 선형 관계
 - 0 : 선형 관계 없음
 - -1 : 완벽한 음의 선형 관계



분석 방법

Spearman 방법

- 정의: 순위 데이터를 기반으로 한 비선형 관계를 측정한다.
- 가정:
 - 데이터를 정규분포로 가정하지 않아도 된다.
 - 순위 데이터를 사용할 때 적합하다.
- 계산 방법
 - 각 값의 순위를 기준으로 Pearson 상관계수를 계산한다.
- 해석
 - +1 : 완벽한 양의 순위 관계
 - 0 : 순위 관계 없음
 - -1 : 완벽한 음의 순위 관계



분석 방법

두 분석 방법을 선택한 이유

- Pearson 상관계수는 연속형 데이터의 선형 관계를 분석하기 위해 사용했다.
- Spearman 상관계수는 비선형 관계를 포함한 다양한 형태의 관계를 분석하고, 이상치의 영향을 줄이기 위해 사용했다.



상관분석 수행

Pearson 상관계수 계산

```
-- 경제적 지표와 자살률 간의 상관계수
SELECT
  CORR(실업률, 자살률) AS 실업률_상관계수,
  CORR(가계_소득, 자살률) AS 가계_소득_상관계수,
  CORR(빈곤율, 자살률) AS 빈곤율_상관계수,
  CORR(지니계수, 자살률) AS 지니계수_상관계수
FROM 경제지표
JOIN 자살률_데이터 ON 경제지표.연도 = 자살률_데이터.연도;

-- 사회적 지표와 자살률 간의 상관계수
SELECT
  CORR(이혼율, 자살률) AS 이혼율_상관계수,
  CORR(가정_폭력_건수, 자살률) AS 가정폭력_상관계수,
  CORR(교육_수준, 자살률) AS 교육수준_상관계수,
  CORR(인구_밀도, 자살률) AS 인구밀도_상관계수
FROM 사회지표
JOIN 자살률_데이터 ON 사회지표.연도 = 자살률_데이터.연도;

-- 건강 관련 지표와 자살률 간의 상관계수
SELECT
  CORR(정신_건강_문제율, 자살률) AS 정신건강_상관계수,
  CORR(알코올_약물_남용률, 자살률) AS 알코올남용_상관계수,
  CORR(정신과_인원율, 자살률) AS 정신과인원_상관계수,
  CORR(자살_시도_건수, 자살률) AS 자살시도_상관계수
FROM 건강지표
JOIN 자살률_데이터 ON 건강지표.연도 = 자살률_데이터.연도;
```

상관분석 수행

Spearman 상관분석 수행

```
WITH RankedData AS (  
    SELECT  
        연도,  
        RANK() OVER (ORDER BY 실업률) AS 실업률_순위,  
        RANK() OVER (ORDER BY 가계_소득) AS 가계소득_순위,  
        RANK() OVER (ORDER BY 정신_건강_문제율) AS 정신건강_순위,  
        RANK() OVER (ORDER BY 자살률) AS 자살률_순위  
    FROM 경제지표  
    JOIN 건강지표 ON 경제지표.연도 = 건강지표.연도  
    JOIN 자살률_데이터 ON 경제지표.연도 = 자살률_데이터.연도  
)  
SELECT  
    CORR(실업률_순위, 자살률_순위) AS Spearman_실업률_상관계수,  
    CORR(가계소득_순위, 자살률_순위) AS Spearman_가계소득_상관계수,  
    CORR(정신건강_순위, 자살률_순위) AS Spearman_정신건강_상관계수  
FROM RankedData;
```




Analysis & Results

04

분석결과

실업률과 자살률 간의 Pearson 상관계수: 0.65

- 양의 상관관계가 나타나며, 실업률이 높아질수록 자살률도 증가하는 경향이 있다

가계 소득과 자살률 간의 Pearson 상관계수: -0.45

- 음의 상관관계가 나타나며, 가계 소득이 높을수록 자살률이 낮아지는 경향이 있다.

정신 건강 문제발생률과 자살률 간의 Spearman 상관계수: 0.78

- 매우 강한 양의 상관 관계로, 정신 건강 문제는 자살률 증가의 주요 원인으로 분석된다.



Conclusion

05



결론

- 경제적 불안정(실업률 증가), 정신 건강 문제, 사회적 불안정(이혼율 증가)등이 자살률 증가와 밀접한 관련이 있음을 확인했다.
- 정책적 시사점
 - 실업률 감소를 위한 경제 정책 강화
 - 정신 건강 서비스에 대한 접근성 확대
 - 가정 문제 해결 및 사회적 지원 프로그램 강화



발표를 들어주셔서 감사합니다.

Q&A

디지털헬스케어학부 전대한

소프트웨어학부 황선준