

# 실업자 수와 자살자 수의 연관성 파악

홍지민

충남대학교 우주지질학과

과거 실업이 자살에 미치는 영향을 ARIMA 회귀분석을 통해 분석한 연구에서는 뚜렷한 경향이 보이지 않았지만 '특정 시기와 사건으로 인한 단기적인 변화가 보다 자살률에 영향을 많이 미치고 있다'고 서술 했습니다. 또한 실업자의 자살 생각 영향 요인을 분석한 연구에서는 1998년, 2003년, 2009년에 각각 IMF 경제위기, 세계 금융위기 등과 같은 경제적 위기가 발생했고 자살률이 급증 했다는 연구 결과도 있습니다. 이와 같은 연구를 바탕으로 저는 실업과 자살이 연관이 있을 것이라 생각하고 분석을 진행 했습니다. 기존의 연구들보다 더 긴 기간 동안의 데이터(1997-2020)를 사용하여 연령별, 지역별로 구분하여 분석을 진행 했습니다. 실업과 자살이 양의 상관관계를 가질 것이라는 예상과 다르게 20대에선 실업자가 많을수록 자살자가 감소하는 음의 상관관계를 보였습니다. 반대로 주 경제활동 인구인 30대, 40대, 50대에선 예상대로 양의 상관관계를 보여줍니다. 지역에 따라서도 다른 경향을 보입니다. 부산의 경우, 실업자 수는 가장 적지만 그에 비해 자살자 수는 꽤 많은 것으로 해석됩니다. 경남, 경북은 실업자 수는 많은 편은 아니지만 많은 자살자 수를 보이므로 좀 더 세부적인 분석이 필요합니다. 몇몇 지역을 제외하고 대체적으로 실업자 수가 증가하면 자살자 수도 증가하는 모습을 보여줍니다.

**키워드:** 실업자 수, 자살자 수, 회귀분석, 군집분석

## I. 서론

### 연구의 목적

2019년 발병한 코로나 바이러스 감염증이 확산되면서 경제가 침체되고 실업자 수가 크게 늘었습니다. 유동인구의 감소로 자영업자의 수입이 감소하고, 폐업을 하는 경우가 많아졌습니다. 2021년 1월에는 코로나 바이러스 감염증 확산 방지를 위해 노인 일자리 등 신규 공공 일자리 사업을 진행하지 않았고, 그 결과 실업자 수가 157만 명으로 통계 작성을 시작한 1999년 6월 이후 최대치를 기록 했습니다. 특히, 매달 증가세를 보이던

60대 이상의 취업자 수는 1만5000명 줄었으며, 이는 노인 일자리 사업을 진행하지 않거나 규모가 줄어들었기 때문입니다.

사회적 거리 두기가 장기화 되면서 답답함과 우울함을 호소하는 사람이 늘었고, '코로나 블루'라는 신조어가 생겼습니다. 의학적인 질병이라기보다는 사회현상에 따른 심리적 증상에 가까운 코로나 블루는 가슴 답답, 두통, 어지러움, 이명, 우울감을 호소합니다. 코로나 블루로 발생한 우울증 때문에 삶의 질과 기능을 포함한 모든 측면에서 부정적인 영향을 미치며 우울증 환자의 자살 위험은 일반인에 비해 20배나 높고, 심지어 인간의 평균 기대수명을 약 10년가량 단축시킨다는 연구 결과도 있습니다. 특히, 20-30대의 정신건강이 두드러지게 악화된 것을 확인할 수 있으며, 코로나 바이러스 감염증 발병 이전인 2018년에 비해서는 약 6배 증가한 우울 위험군 비율을 보여줍니다.

코로나 바이러스 감염증에 의한 실업자 수의 증가가 자살자 수의 증가에 영향을 미치는 것으로 생각되며 이는 과거 경제위기에서도 나타날 것이라 생각됩니다. 실업이 1년 이상 장기화 되면 자살로 이어지는 확률이 높다는 선행 연구 결과도 있으며, 이를 바탕으로 본 프로젝트에서는 실업자 수와 자살자 수의 상관관계를 파악하고 '빅데이터통계기법'에서 배운 방법론을 적용하는 것입니다.



그림. 1. 2021년 1분기 「코로나19 국민정신 건강 실태조사」 결과, 20대와 30대에서 다른 연령대에 비해 높은 우울 위험군 비율 (각각 30%, 30.5%)을 보이고 있습니다.

## II. 연구방법 및 구성

### 연구방법

본 프로젝트에선 과거 20년 동안의 데이터를 사용하여 '회귀분석' 및 '군집분석'을 실시했습니다. 자료는 통계청의 국가통계포털(KOSIS) 및 마이크로데이터(MDIS)에서 이용할

수 있으며 필요시에 자료를 가공하여 이용하기 쉽게 핸들링 하였습니다. 자살자 데이터는 연도별 사망자 데이터를 얻은 후, 사망원인을 고의적 자해(자살)로 필터링을 진행합니다. 연도에 따라 자살자 수를 집계하고 연령별, 지역별로 나눠 핸들링 데이터를 생성합니다. 실업자 데이터는 KOSIS에서 쉽게 핸들링 하여 다운로드 받을 수 있었습니다. Python의 Numpy, Scipy, Scikit-Learn등의 수학 및 머신 러닝 라이브러리를 사용하여 분석을 진행 했습니다. 연령별로 1차 선형 회귀 및 다항 회귀를 진행하고, 지역에 따라 거리 기반, 밀도 기반, 계층적 세 가지의 군집분석을 진행 하였습니다.

1)

사망자 주소(시도)	성별	사망연월일	사망연령(5세단위)	56항목 사망원인
11	1	19970101	10	25
11	1	19970101	10	35
11	1	19970101	10	42
11	1	19970101	10	50
11	1	19970101	10	50

2)

연도	10대	20대	30대
1997	379	1243	1431
1998	476	1531	1922
1999	380	1135	1556
2000	274	942	1352
2001	233	931	1371
2002	234	1030	1652
2003	301	1233	1951

표. 1. 1) 사망자 데이터를 받아 사망원인을 고의적 자해(자살)로 필터링을 실시 후, 2) 연령별로 범주를 나눠 연도별로 수를 집계합니다. 추가로 사망자의 주소 데이터를 활용하여 지역별로도 분석을 진행했습니다.

### Ⅲ. 자료 분석 및 결과

#### 1) 분석자료

자살자의 데이터는 통계청의 MDIS를 이용하여 다운로드 했습니다. MDIS에서 제공하는 보건-사망원인통계-사망\_연간자료\_B형을 이용 했습니다. 1997년부터 2019년까지 자료를 제공하고 있으며 사망 연도, 사망자 주소(시도), 성별, 사망 연령 등과 같은 데이터를 제공합니다. 표. 1. 1)과 같은 데이터를 다운로드하여 사망원인에 따라 필터링을 진행합니다. 56항목 사망원인 중 고의적 자해(자살)는 코드 번호가 55입니다. 사망 연령 코드는 1은 0세, 2는 1~4세, 3은 5~9세로 5세 단위로 19까지 주어집니다. 사망 연령 코드 19의 경우 85세 이상을 나타내며, 99의 경우 연령 미상을 나타냅니다. 사망자 주소 코드는 11은 '서울특별시', 21은 '부산광역시', 22는 '대구광역시'를 나타내며 이렇게 총 17개의 시도를 나타냅니다. 성별은 1이 남자, 2가 여자를 나타냅니다.

Python의 Pandas 라이브러리를 이용하여 자살로 필터링 된 데이터를 연도-연령에 따라서 개수를 집계합니다. 핸들링 된 표. 1. 2)를 바탕으로 데이터 분석을 진행했으며 같은 방법으로 연도-지역의 데이터도 생성합니다. 각각 'count(age)\_자살자', 'count(locate)\_자살자'로 csv파일을 생성합니다.

실업자의 데이터는 통계청의 KOSIS를 이용 했습니다. 노동-경제활동인구조사-실업자에

서 다운로드하였으며, KOSIS에서 제공하는 데이터 핸들링 UI를 이용하여 쉽게 연도-연령, 연도-지역의 데이터를 얻을 수 있었습니다. 역시나 각각 'count(age)\_실업자', 'count(locate)\_실업자'로 csv파일을 생성 했습니다. 월이나 분기로 나눠서 분석하는 방법도 있었지만 실업자 데이터의 경우, 공채 상반기, 하반기와 같은 분기에 따른 요소가 있기 때문에 연도별로 나눠서 분석을 진행 했습니다.



그림. 2. 통계청-KOSIS에서 이용할 수 있는 실업자 데이터의 목록입니다.

표. 1에서 알 수 있듯이 각 연령, 지역에 따라 값의 차이가 있기 때문에 Scikit-Learn에서 제공하는 MinMaxScaler 라이브러리를 통해 최소/최댓값이 각각 0, 1로 스케일링을 진행합니다. 즉, 위의 표에서는 자살로 가장 많은 사람이 사망한 연도의 값을 1로 스케일링하고 자살로 가장 적은 사람이 사망한 연도의 값을 0으로 스케일링 합니다. 또한, 위와 같은 방법을 실업자 데이터에 대해서도 적용합니다.

## 2) 기초분석

우선, 연도에 따라서 실업자 수의 변화를 살펴봤습니다. 1998년 IMF 경제위기의 여파로 2000년에도 높은 실업자 수를 기록하고 있으며, 특히 주 경제활동 인구인 30대, 40대에서 가장 높은 실업자 수를 보여줍니다. 이후 실업자 수는 감소추세를 보이거나 2003-2005년에 전 연령대에서 크게 증가하는 모양을 보입니다. 이는 2003년 신용카드 대란이 원인으로 생각되는데, 무분별한 카드 발급으로 많은 사람들이 신용불량자로 전락하고 기업에게도 많은 부채가 생기는 등의 경제에 큰 타격을 입었기 때문입니다. 시간이 지나 차츰 회복하던 경제도 2007-2009년 세계 금융 위기를 맞게 됩니다. 2007년에 발생한 서브프라임 모기지(Subprime mortgage) 사태를 시작으로 2008년 리먼 브라더스

(Lehman Brothers)까지 파산을 하면서 전 세계적인 금융 위기가 시작됩니다. 수출입 규모가 크게 감소하고 주가가 하락하며, 기업들은 채용규모를 줄이기 시작합니다. 이에 따라서 실업자 수가 증가한 것으로 보입니다. 경기를 회복한 후에도 20대 청년층의 실업자 수는 계속 증가하고 이때부터 본격적으로 청년 실업률이 사회적 문제로 대두되기 시작했습니다. 또한, 50대의 중년층 실업자 수도 같이 증가하는 모습을 볼 수 있는데, 사회가 고령화가 됨에 따라서 일을 하고 싶어 하는 중년층은 많아지지만 그만큼 충족할 수 있는 일자리는 줄어들다는 것으로 해석할 수 있습니다.

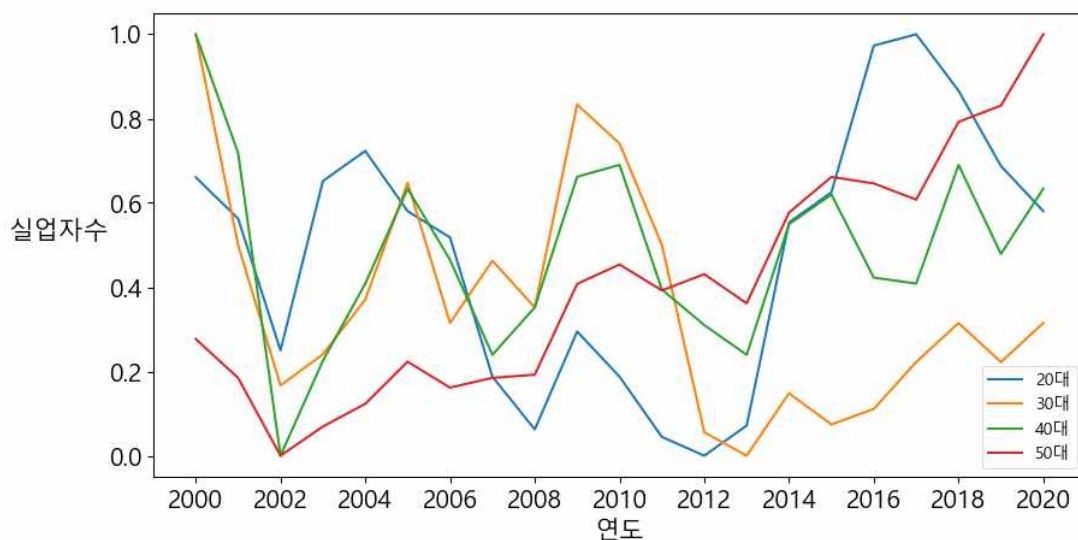


그림. 3. 연도에 따른 실업자 수의 변화를 보여줍니다. 2010년 이전에는 경제 상태에 따라서 실업자 수가 증감하지만 이후에는 인구 고령화에 따른 청년 실업률, 노인 실업률의 영향으로 보입니다.

자살자 수는 과거보다 최근 10년에서 높은 수치를 기록하고 있습니다. 2020년 기준 대한민국은 OECD 국가 중 자살률 1위의 불명예스러운 타이틀을 가지고 있습니다. 전 세계를 기준으로 해도 3~4위정도로 상위권에 위치하고 있습니다. 경제위기가 있을 때마다 자살자 수의 증감이 나타나지만 실업자 수만큼 큰 증감은 보이지 않습니다. 또한 과거 20-30대의 자살자 수가 많았던 것과 다르게, 최근 40-50대의 자살이 사회적 문제로 떠오르고 있습니다. 물론 과거에 비해 전 연령에서 높은 자살자 수를 보이는 것도 특징입니다.

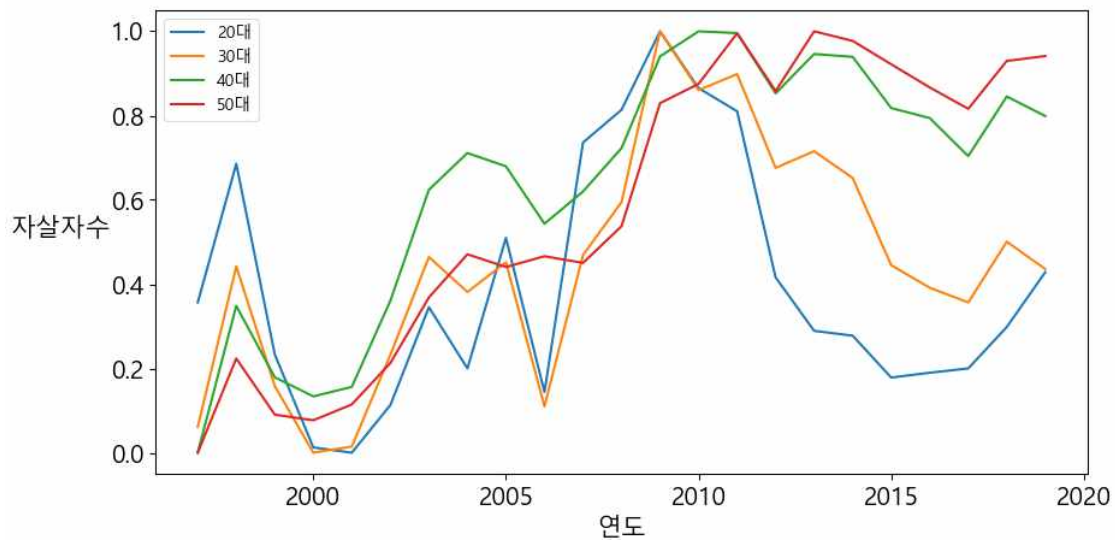


그림. 4. 연도에 따른 자살자 수의 변화를 보여줍니다. 과거 20-30대의 자살자 수가 많았지만 시간이 지남에 따라 40-50대의 중장년층의 자살자 수가 크게 증가합니다.

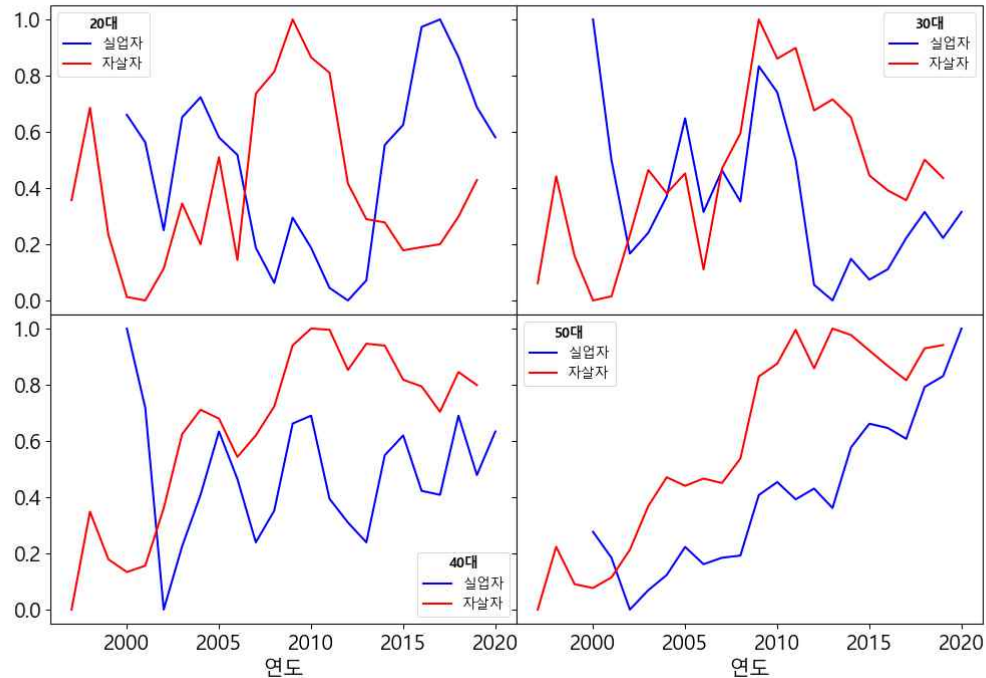


그림. 5. 연도에 따른 실업자 수와 자살자 수를 연령에 따라 나눈 그래프입니다. 파란 실선은 실업자의 수, 빨간 실선은 자살자의 수로 MinMaxScaler로 0-1까지 스케일링 했습니다. 50대에선 실업자가 증가함에 따라 자살자가 늘어나고 있음을 명확히 볼 수 있습니다. 나머지 연령대에서도 유사한 패턴을 보이지만 뒤에 서술한 회귀를 통해 좀 더 심층적으

로 분석 해보겠습니다.

추가로 지역에 따라서 취업자 수와 자살자 수를 분석 했습니다. '부산'에서는 낮은 실업자 수를 기록하고 있음에도 높은 자살자 수를 보이고 있으며, '인천'과 '경기'에서는 실업자가 증가함에 따라 자살자도 증가하는 좋은 상관관계를 보여주고 있습니다. '서울'에서는 대체적으로 실업자의 수가 큰 폭으로 증감하고 있으며 이는 대한민국의 경제의 중심지라고 생각되기 때문입니다. '경북'과 '경남'에서는 낮은 실업자 수에 비해서 높은 자살자 수를 기록하고 있는데 이는 40-50대의 비율이 높기 때문에 발생하는 현상이라고 생각합니다. 또한, 전 지역에서 2008-2010년에 자살자 수의 큰 증가가 나타나는데 위에서 언급된 세계 금융 위기의 영향인 것으로 추측 됩니다.

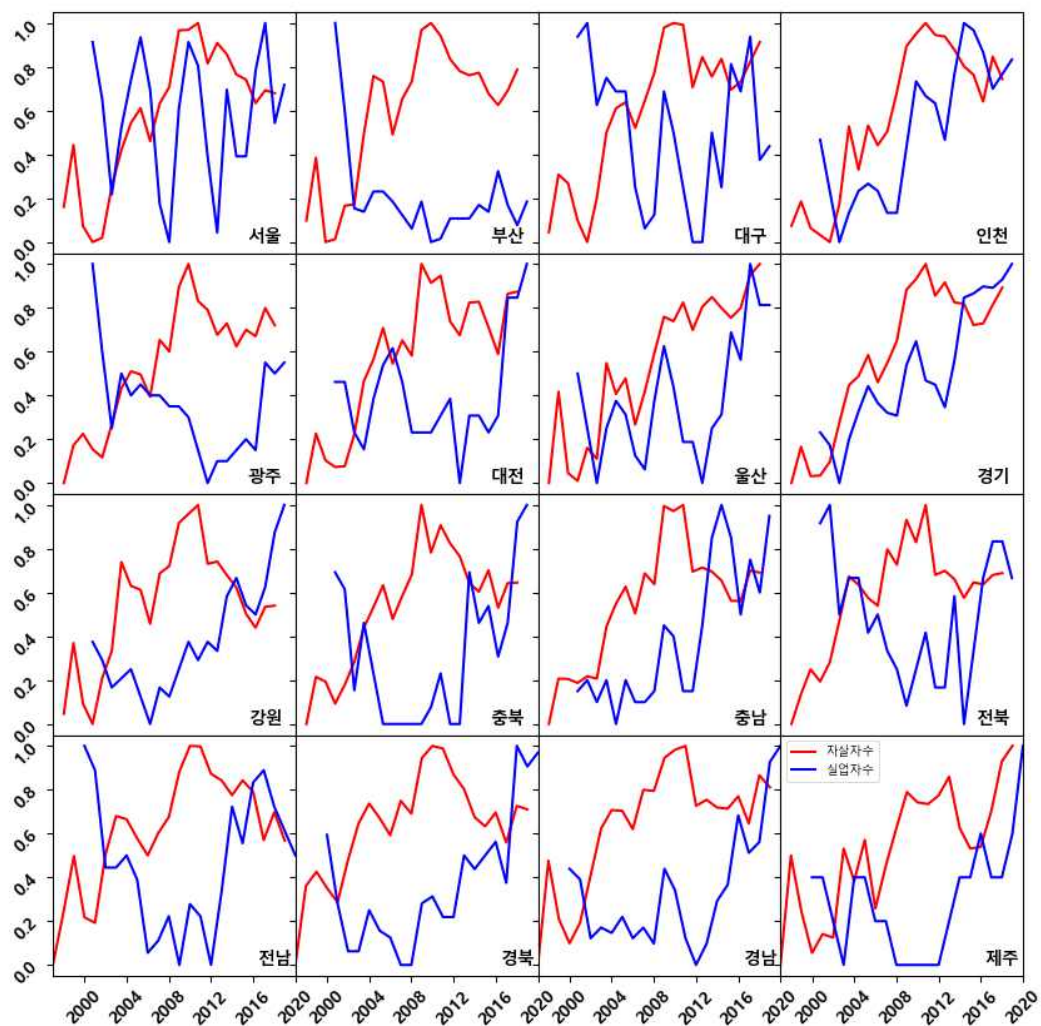


그림. 6. 지역별 연도에 따른 실업자 수 및 자살자 수의 변화를 보여줍니다. 파란 실선은 실업자의 수, 빨간 실선은 자살자의 수로 MinMaxScaler로 0-1까지 스케일링 했습니다.



‘부산’은 명확하게 실업자 수는 낮지만 자살자 수는 높은 것으로 보입니다. ‘인천’과 ‘경기’에서 실업자가 증가함에 따라 자살자 수도 증가하는 양의 상관관계로 해석됩니다.

### 3) 실증분석

우선 연령에 따라서 실업과 자살과의 관계를 분석합니다. IMF 외환위기를 겪은 직후인 1999-2001의 데이터는 IMF 전후를 비교할 수 없기 때문에 제외하고 회귀분석을 진행했습니다. 20대를 제외한 전 연령에서 실업자 수가 증가하면 자살자 수가 증가하는 관계를 보여줍니다. 특히 40-50대에서 강한 양의 상관관계를 보이는 것으로 중장년층의 자살은 실업과 밀접한 관련이 있는 것으로 해석됩니다. 30대에선 40-50대보단 아니지만 여전히 양의 상관관계를 보이며, 실업과 다른 복합적인 요인의 결과로 보입니다. 20대에선 특이하게 실업자가 감소하면 자살자가 증가하는 경향이 보입니다. 실업자가 감소하는 것은 반대로 취업자가 증가한다는 뜻이기도 합니다. 취업자가 증가하면 자살자가 증가한다는 해석은 취업으로 인한 주변의 스트레스나 사내 따돌림과 같은 정신적인 요인이 원인이 될 수 있다는 뜻으로 해석이 가능할 수 있습니다.

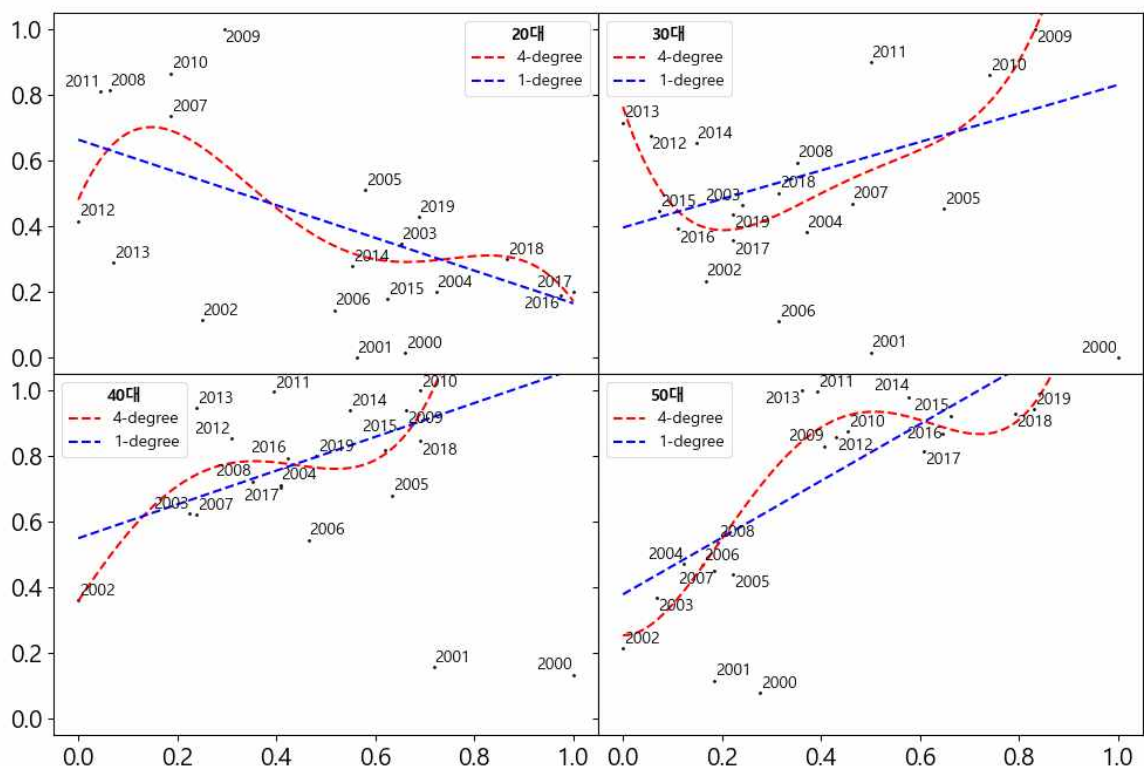


그림. 7. x축은 실업자 수를 y축은 자살자 수를 나타낸 연령별 그래프입니다. 각 점에 해당하는 연도를 표기 했으며 빨간 점선은 4차 다항 회귀의 결과를, 파란 점선은 1차 선형



회귀의 결과를 보여줍니다. 20대를 제외한 전 연령에서 좋은 양의 상관관계를 보이며 이는 실업이 자살의 원인으로 꽤나 큰 영향을 미치는 것으로 해석할 수 있습니다.

실제로 2018년 경찰청에서 발표한 변사자통계에서 10-30세의 자살 동기는 '정신적 어려움'이라고 발표한 기사도 찾아볼 수 있었습니다. 즉, 20대의 자살은 단순한 실업에서 이어지는 것이 아닌 취업을 하더라도 주변 환경이나 직장에서 오는 스트레스, 정신적인 괴로움이 원인이 될 수 있다는 것입니다. 최근 문제되고 있는 청년 고독사가 단순히 장기적인 실업 때문에 발생하는 것이 아니라 긴 실업기간, 스트레스, 정신적인 괴로움 등에서 발생하는 것으로 추측됩니다.

## 연령대에 따라 다른 자살 동기



[출처: 경찰청, 2018년 변사자통계]

그림. 8. 경찰청에서 2018년에 발표한 변사자 통계

두 번째는 지역에 따라서 군집분석을 실시했습니다. '빅데이터통계기법' 수업에서 배운 것과 같이 거리 기반, 밀도 기반, 계층적으로 분석을 실시했습니다. MinMaxScaler 이용해서 0-1로 스케일링한 데이터를 이용해서 각 지역에 해당하는 연도를 전부 더하는 것으로 실업 지수와 자살 지수를 구했습니다. 즉, 이 지수가 높다면 실업자 수와 자살자 수가 높았던 연도가 많다는 것을 뜻합니다. 실업자 수와 자살자 수가 모두 있는 2000-2019년 20개의 데이터를 사용했습니다. 군집은 4개로 설정하여 분석 했으며, 모든 알고리즘에서 괜찮게 구분하는 것으로 보이며 같은 군집으로 잘 분류합니다.

'부산'은 실업 지수가 약 4로 현저히 낮은 실업과 다르게 자살 지수는 약 13.5정도로 비교적 높은 편입니다. '경남', '경북'은 평균적인 실업 지수를 보이지만 자살 지수는 약 14.5로 굉장히 높은 값을 보입니다. 이 경향은 여러 가지로 해석될 수가 있는데 지역에 따른 연령대의 비율 차이 때문에 발생하는 것으로 추측하고 있습니다. 즉, '경남', '경북' 지역은 40-50대의 비율이 높고 최근 높아진 자살자 수가 많기 때문에 이러한 경향을 반

영하는 것으로 생각합니다. 비슷한 도시로 '전남', '대구'가 있지만 이 두 도시는 비교적 실업자 수도 높은 편으로 보입니다. '서울'의 경우 극단적으로 높은 실업 지수를 보이며 이는 2010년부터 사회현상으로 떠오른 청년 실업, 노인 실업의 영향으로 보입니다. '부산', '경남', '경북'을 제외한 나머지 도시에서는 실업 지수가 증가할수록 자살 지수가 높아지는 경향이 보입니다.

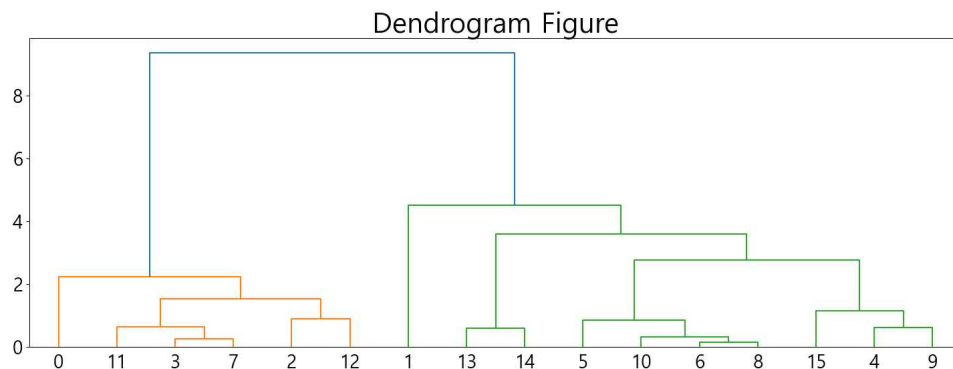


그림. 9. 계층적 분석을 하기 전에 덴드로그램을 작성 했습니다.

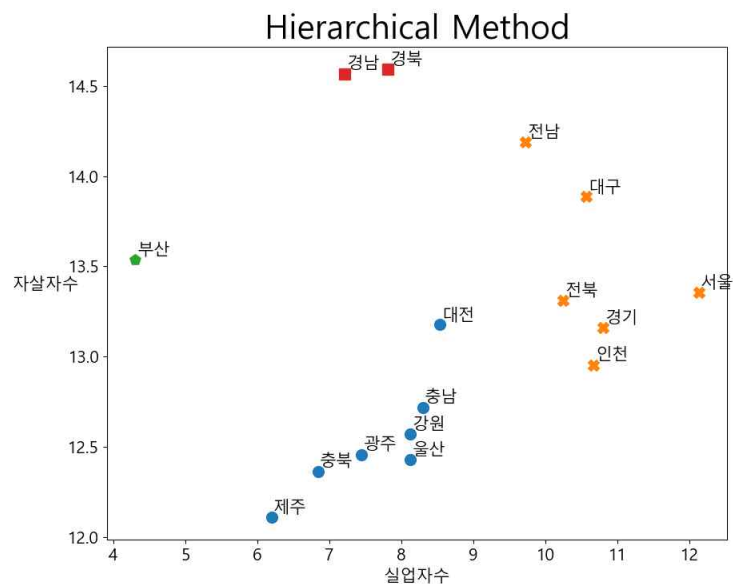


그림. 10. 계층적 군집분석, 파란 동그라미는 비교적 실업 지수가 낮고 자살 지수도 낮은 군집, 노란 엑스는 실업 지수가 높고 자살 지수도 높은 도시를 나타냅니다.

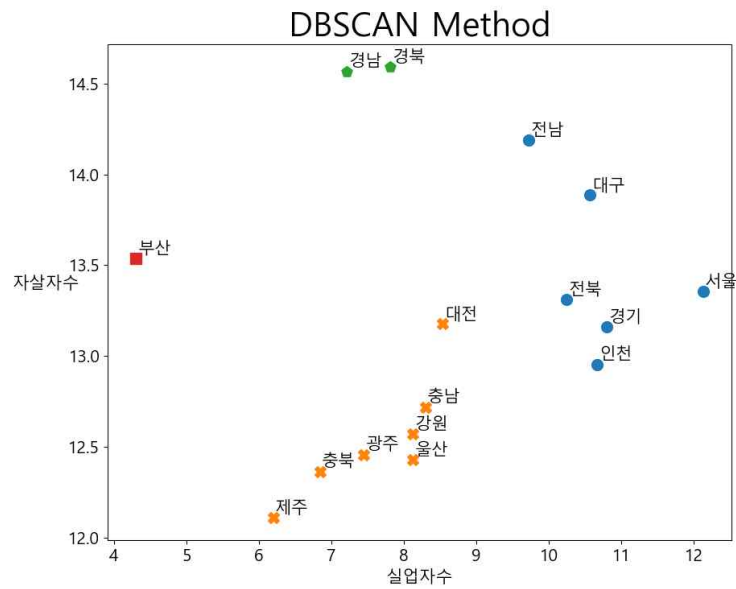


그림. 11. 밀도 기반 군집분석, 위의 그림과 똑같이 군집화 했으며 잘 분류한 것으로 해석됩니다.

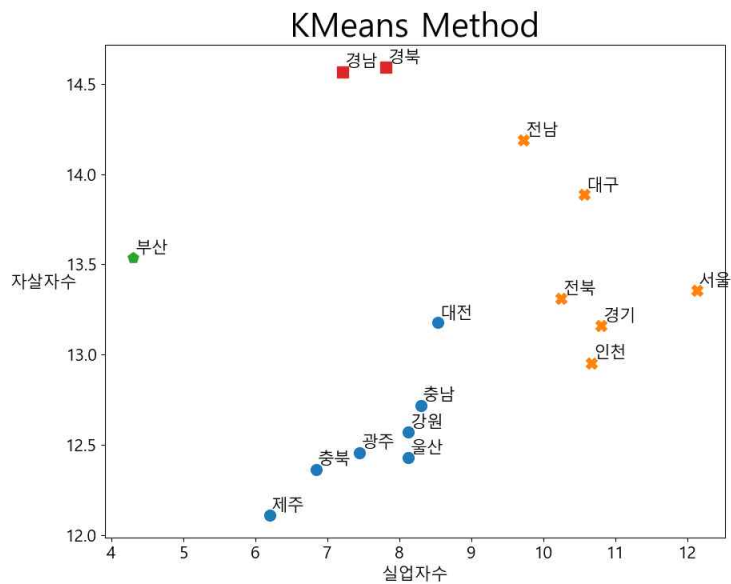


그림. 12. 거리 기반 군집분석, 여전히 똑같이 분류하는 것으로 보이며 각 도시에 따라서 확실히 다른 경향을 보이는 것으로 생각합니다.

#### IV. 결론

본 연구에서는 실업과 자살과의 관계를 파악하고 지역에 따라서 군집분석을 실시했습니다. 20대를 제외한 나머지 연령대에서 실업과 자살은 유의미한 상관관계를 보여줍니다. 현 사회문제로 높은 자살률이 대두되는 가운데 이러한 문제를 해결하기 위해서 경제적인 문제를 해결하는 것도 하나의 방법이 될 수 있을 것 같습니다. 20대의 자살에는 복합적인 요인이 있으며 그 중에 하나가 될 수 있는 것이 정신적인 요인이 될 수 있다는 것입니다.

‘부산’에서는 낮은 실업 지수를 보이지만 비교적 높은 자살 지수를 보여줍니다. 이러한 경향은 다른 도시에서 볼 수 없는 특이한 경향이며 원인을 분석하기 위해서 좀 더 심층적인 분석이 필요해 보입니다. ‘경남’, ‘경북’은 평균적인 실업 지수에 비해서 높은 자살 지수를 보여주는데 이것은 40-50대의 비율이 높은 지역이기 때문에 최근 증가한 자살자수를 반영하는 것으로 보입니다. 세 지역을 제외한 곳은 비교적 실업 지수와 자살 지수가 양의 상관관계를 보이며, 그림 .12에서 파란 동그라미는 낮은 실업 지수와 낮은 자살 지수를 보이는 지역이며, 노란 엑스는 높은 실업 지수와 높은 자살 지수를 보이는 지역으로 보입니다. 거리 기반, 밀도 기반, 계층적 군집분석 모두 잘 분류하는 것으로 판단되며 좋은 성능을 보여줍니다.

좀 더 좋은 분석결과를 위해서는 자살자의 정신건강 상태나 불안 지수와 같은 좀 더 많은 변수가 필요할 것으로 생각됩니다. 하지만 공공 데이터 특성상 개인정보나 민감한 정보에는 접근하기 힘들고 데이터 또한 실업 데이터나 자살 데이터와 같이 많은 기간에 걸쳐서 수집한 데이터가 아니기 때문에 정확한 분석을 하기가 힘들었습니다. 또한, 실업자 데이터의 경우 기업에서의 채용 규모에 따라서 달라질 수도 있고, 현재 코로나와 같은 특수한 상황도 영향을 받을 수 있기 때문에 여러 가지 변수를 제하는 것이 어렵습니다.

하지만 현재 분석을 통해 실업이 자살의 한 요인이 될 수 있으며, 나아가 현재 사회문제로 떠오른 청년 고독사, 청년 실업률, 노인 실업률, 독거노인 문제 등을 연쇄적으로 해결할 수 있는 열쇠 역할을 할 수 있지 않을까 생각 했습니다.

## V. 출처 및 부록

- '1월 실업자 157만명, 실업률 사상 최고 5.7% 쇼크...'세금주도일자리'의 민낯', 조선일보, 2021, [https://biz.chosun.com/site/data/html\\_dir/2021/02/10/2021021001009.html](https://biz.chosun.com/site/data/html_dir/2021/02/10/2021021001009.html)
- '사회적 거리 두기 장기화에 "코로나 블루" 급증...정신과 전문의가 알려준 극복 비결', 중앙일보, 2020, [https://jhealthmedia.joins.com/article/article\\_view.asp?pno=21636](https://jhealthmedia.joins.com/article/article_view.asp?pno=21636)
- '2030 세대 실업 장기화되면 자살 위험 높아져', 대구신문, 2020, <https://www.idaegu.co.kr/news/articleView.html?idxno=330398>
- '여성.젊은층에 더 짙게 드리우는 "코로나 블루"의 그늘', 의사신문, 2020, <http://www.doctorstimes.com/news/articleView.html?idxno=213187>
- '코로나 19 장기화 우울 위험군, 자살 생각을 증가...20-30대 정신건강 악화', 후생신보, 2021, <http://www.whosaeng.com/126907>
- '80년 vs 2010년 비교한 자살 연령층의 변화', 헬스미디어, 2017, <http://www.healthmedia.co.kr/news/articleView.html?idxno=59229>
- '"2020 자살예방백서" 발간[파일 별첨]', 이엠디, 2020, <https://mdon.co.kr/mobile/article.html?no=27665>
- '실업과 소득격차가 자살에 미치는 영향검증: 1995년에서 2008년까지의 시계열 자료 분석', 윤우석, 2011
- '실업자의 자살생각 영향요인', 김혜영, 김희숙, 2017

```
import sys, os, glob, natsort # 데이터를 불러오기 위한 라이브러리
import pandas as pd # 데이터를 전처리하기 위한 라이브러리
from matplotlib import pyplot as plt # 그림을 그리기 위한 라이브러리
import numpy as np # 수리적인 계산 라이브러리
from matplotlib import font_manager, rc # 그림의 폰트를 지정해주는 라이브러리
from sklearn.preprocessing import MinMaxScaler # 데이터를 정규화하는 라이브러리
from sklearn.cluster import KMeans, DBSCAN, AgglomerativeClustering # 각각 거리기반, 밀도기반, 계층적 군집분석 라이브러리
from adjustText import adjust_text # 텍스트 겹침 방지 라이브러리
from scipy.cluster.hierarchy import dendrogram, linkage # 덴드로이드 라이브러리
from sklearn.preprocessing import PolynomialFeatures # 다항 회귀를 위한 라이브러리
from sklearn.linear_model import LinearRegression # 계수를 얻은 후 선형회귀를 하기 위한 라이브러리
```

그림. 13. 사용한 라이브러리의 목록 및 간단한 설명

```

poly_features = PolynomialFeatures(degree=4,include_bias=False) # 다항회귀 실시
linear_features = PolynomialFeatures(degree=1,include_bias=False) # degree=1인 경우 1인 선형회귀
ddt = (dt_scaled.dropna(axis=0)).drop(index=[2000,2001]) # 2000,2001년 데이터 제외

for age,aax in zip(col_1,ax_list):
    X = np.array(ddt['실업자수'][age])

    X_poly = poly_features.fit_transform(X.reshape(-1,1)) # 다항회귀 실시
    X_poly_1 = linear_features.fit_transform(X.reshape(-1,1)) # 선형회귀 실시
    y = ddt['자살자수'][age]
    lin_reg = LinearRegression();lin_reg_1 = LinearRegression() # 선형회귀 라이브러리를 이용하여 계수를 얻기
    lin_reg.fit(X_poly,y)
    lin_reg_1.fit(X_poly_1,y)

    X_new = np.linspace(0,1,100).reshape(-1,1)
    X_new_poly = poly_features.transform(X_new)
    X_new_poly_1 = linear_features.transform(X_new)
    y_new = lin_reg.predict(X_new_poly) # 다항회귀 예측
    y_new_1 = lin_reg_1.predict(X_new_poly_1) # 선형회귀 예측
    ax[aax].plot(X_new,y_new,'r--',label="4-degree") # 4차 다항회귀 선
    ax[aax].plot(X_new,y_new_1,'b--',label="1-degree") # 1차 선형회귀 선
    plt.setp(ax[aax].get_xticklabels(),fontsize=15)
    plt.setp(ax[aax].get_yticklabels(),fontsize=15)
    leg = ax[aax].legend()
    leg.set_title(str(age),prop={'weight':'heavy'})

```

그림. 14. 실업자 수와 자살자 수와의 회귀분석 코드

```

cluster = AgglomerativeClustering(n_clusters=4) # 군집분석 라이브러리를 바꿔가면서 분석
ddtdt['cluster'] = cluster.fit_predict(ddtdt) # KMeans, DBSCAN, AgglomerativeClustering
# 클러스터를 만들어서 군집 분류
fig,ax = plt.subplots(1,1,figsize=(10,8))

for city in col_1:
    plt.annotate(city,(ddtdt['실업자수'].loc[city]+0.025,ddtdt['자살자수'].loc[city]+0.025),fontsize=15)

for f in [0,1,2,3]: # 판독도를 높여서 군집에 따라서 마커와 색깔 변경
    if f == 0:
        s = 'o'
    elif f == 1:
        s = 'X'
    elif f == 2:
        s = 'p'
    else:
        s = 's'
    plt.scatter(ddtdt[ddtdt['cluster'] == f]['실업자수'],ddtdt[ddtdt['cluster'] == f]['자살자수'],s=100,marker=s)

plt.xlabel("실업자수",fontsize=15);plt.ylabel("자살자수",fontsize=15,rotation=0,labelpad=20)
plt.setp(ax.get_xticklabels(),fontsize=15);plt.setp(ax.get_yticklabels(),fontsize=15)
plt.title("Hierarchical Method",fontsize=30)

```

그림. 15. 지역에 따른 군집 분석 코드