

교육통계 프로젝트 분석 보고서

:사교육비 증가와 고교 유형별 진학을 및 청소년 비행·비만율·우울감의 상관관계

박준식·이찬녕·이형석·임유하

프로젝트 기반 빅데이터 서비스 개발자 양성 과정 (4기)

-----<요약>-----

이 연구에서는 2018년부터 2022년도까지 지역별 중학생 졸업 후 상황 데이터를 분석하고, 각 지역의 사교육비 지출이 청소년 비행 및 학업성취도에 미치는 영향을 실증하고자 한다.

연도별 고교 진학을 및 청소년 비행률과 사교육비 지출을 통계 내 스피어만 상관계수를 통해 두 변수 사이의 상관관계를 분석하였다.

분석 결과 진학과 사교육비 지출의 p-value는 0.05 이상으로 상관관계가 없다는 결과를 보였고, 비행률과 사교육비 지출의 p-value는 0.05 미만으로 상관관계가 통계적으로 유의한 결과를 나타내었다.

주제어 : 사교육비 지출, 고등학교 진학, 청소년 비행, spearman 상관분석

I. 도입

기존 연구들에서는 지방보다 도시지역의 사교육비 지출이 더 크다는 결과를 보인다. (김현진, 2004) 즉, 서울지역에 거주하는 가구일수록 사교육비 지출이 더 많은 것을 의미한다. (우천식, 2004). 또한, 특목고 진학계획이 있는 집단은 그렇지 않은 집단보다 도시지역 학생일 확률이 높다. (서은경, 2018)

지역 규모가 클수록 사교육비는 유의하게 증가하는 경향이 있고(김현진, 2004; 김현진, 박균달, 2008), 특목고에 진학하려는 학생은 일반고에 진학하려는 학생보다 사교육을 더 많이 받거나 비싼 사교육을 받는다(박현정, 이준호, 2009). 그렇다면, 사교육비 지출이 고교 진학률에 큰 영향을 준다는 결과를 나타내지 않을까.

고교 진학 관련 뉴스에 따르면 가구소득, 사교육비 지출이 높을수록 특목고, 자사고 진학률이 높다고 한다(신현성, 2021). 이와 관련한 지역별 사교육비 지출에 따른 진학률 차이는 청소년들이 사회경제적 차이를 느낄 수 있다. 많은 선행연구에서 부모의 사회경제적 지위는 청소년 자녀의 불안, 우울과 같은 정신건강 문제 등의 내재화 문제와 청소년 비행, 알코올중독, 공격적인 행동 등의 외현화 문제에 영향을 미친다는 연구 결과를 제시하였다(Vollebergh et al, 2006; Wight et al, 2006).

불균등한 사교육비 지출은 청소년의 주관적인 사회경제적 지위에 대한 인지에 영향을 미쳐 청소년들의 자존감이 낮아지는 결과를 초래할 수 있고, 이는 신체적 및 정신적인 문제로 이어질 수 있다. (Brunner, 1997; Chen & Paterson, 2006).

따라서 지역별 교육 기회의 불평등성 및 청소년 비행 문제 해소를 위해 지역별 중학생 졸업 후 상황 데이터, 사교육비 지출 데이터와 청소년 비행 데이터를 분석하고, 각 변수 사이의 상관관계를 실증하고자 한다.

II. 방법

가설 1 : 사교육비 지출이 커질수록 청소년 학업성취도를 향상해 특목고와 자율고의 진학률이 상승할 것이다.

가설 2 : 사교육비 지출이 커질수록 청소년 비행이 줄어들 것이다.

가설 3 : 사교육비 지출이 커질수록 청소년 비만율과 우울감이 늘어날 것이다.

i) 데이터 수집 및 전처리

통계청과 시도 유초중등 교육통계 사이트를 통해 데이터 세트를 수집했다. 수집한 데이터 세트는 아래와 같다.

- 2018~2022 년 지역별 중학교 졸업 후 상황(고교 유형별 진학률)
- 2018~2022 년 지역별 중학교 사교육비 지출
- 2018~2022 년 지역별 청소년 음주율, 흡연율(청소년 비행)
- 2018~2022 년 지역별 청소년 우울감 경험률
- 2018~2022 년 지역별 청소년 비만율

데이터별로 데이터 형태가 다르기 때문에 시도 유초중등 교육통계 사이트에서 수집한 중학생의 졸업 후 상황 데이터를 기준으로 전처리 과정을 거쳐서 한 개의 데이터프레임으로 통합했다.

ii) 시각화

각종 그래프와 지리적인 시각화를 통해 대시보드를 만들었다. 대시보드는 상관관계를 보여주는 산점도 그래프를 중심으로 생성했다. 그리고 산점도 그래프의 x 축과 y 축 값을 연도별로 볼 수 있는 꺾은선 그래프 2 개를 보여주도록 만들었다.

지도 시각화는 plotly 의 `express` 를 활용했으며, 데이터의 지역별 시각화는 지오서비스웹(GEOSERVICE-WEB)이라는 곳에서 제공하는 행정구역도 데이터를 JSON 파일로 변환해 사용했다. JSON 파일 내 'CTP_KOR_NM'을 수집한 데이터의 지역명과 일치시켜 줌으로써 위치 데이터를 받아올 수 있었다.

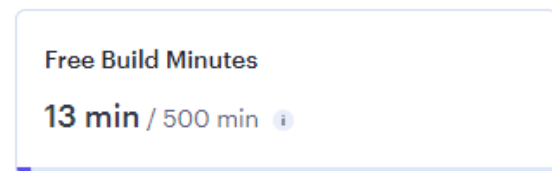
최종적으로 Plotly 의 Dash 와 합치기 위해 html 로 변환했고 같은 경로 내 assets 폴더를 생성 후 코드 실행 시 참조될 수 있게 했다.

데이터 분포를 지역별로 표시하여 비교하고자, plotly 의 `graph_obj` 를 활용해 Box plot 을 추가했다. Dash 를 활용한 것도 처음이었고 callback 함수 구조가 상당히 익숙하지 않아 의도대로 그리기가 쉽지 않았지만, ChatGPT 의 도움을 받아 컬럼별 dropdown 에 상호작용할 수 있게 구현할 수 있었다.

iii) 호스팅

단순히 로컬에서 구현만 하는 것이 아쉬워 배포를 꼭 시도해 보고 싶었다. Plotly의 공식 youtube 채널을 참고하였고, render 라는 사이트를 활용했다. render 사이트의 호스팅 구조는 필요한 라이브러리를 명시한 requirements.txt 파일과 .py 파일을 git repository에 업로드 해 놓으면 되는 간단한 방식이었다.

여러 번의 배포 실패와 이번 프로젝트에서는 총 두 번의 배포를 성공했다. render의 무료 서비스는 최대 500분의 build 제한 시간이 있지만 처음 사용해 보기에 충분한 듯 보인다.



iv) 분석

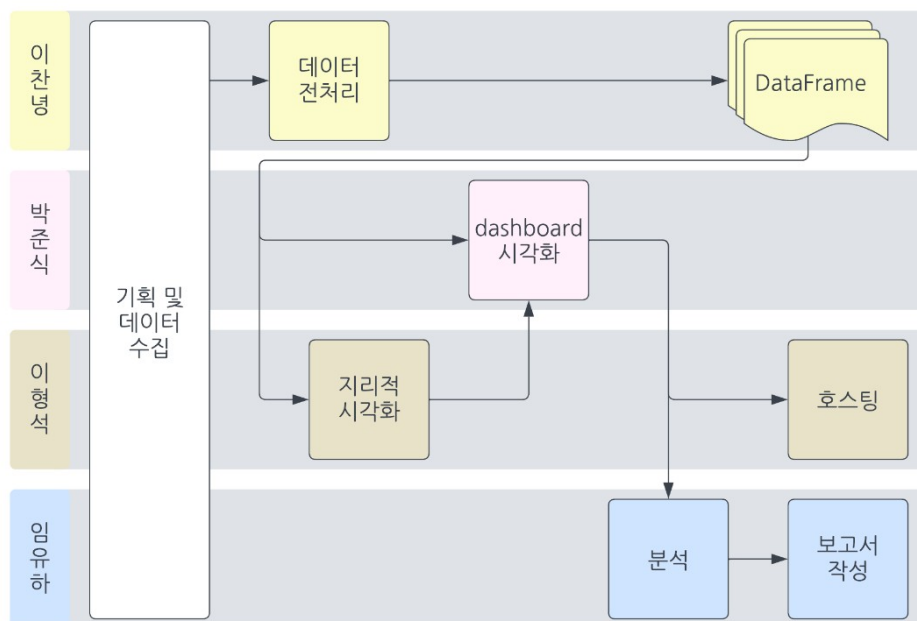
데이터 간의 상관관계를 분석하기에 앞서서 데이터들이 정규성을 만족하는지 검정하기 위해 shapiro-wilk test를 진행했다. 검정은 데이터가 정규성을 만족한다는 귀무가설과 데이터가 정규성을 만족하지 못한다는 대립가설로 유의수준 0.05에서 진행한다.

정규성 검정 결과가 귀무가설을 채택한다면 pearson 상관분석을 진행하며, 귀무가설을 기각하고 대립가설을 채택한다면 spearman 상관분석을 통해 상관관계를 분석한다.

상관분석은 유의수준 0.05 하에 두 변수 간의 상관관계가 있다는 귀무가설과, 두 변수 간의 상관관계가 없다는 대립가설로 분석을 진행한다.

분석 진행 후 사교육비와 다른 변수들의 상관관계의 시각화로 산점도와 선그래프로 작성한다.

사교육비와 고교 진학 비율 및 청소년 비행 데이터, 우울감 등의 변수들을 분석하고 그래프를 작성한다.



[Figure 1] 교육 통계 프로젝트 Workflow 프로젝트의 흐름과 구성원의 역할. 모든 구성원은 전 과정에 종합적으로 기여했으나 주도적으로 맡은 역할을 각자의 프레임에 넣어 나타냈다.

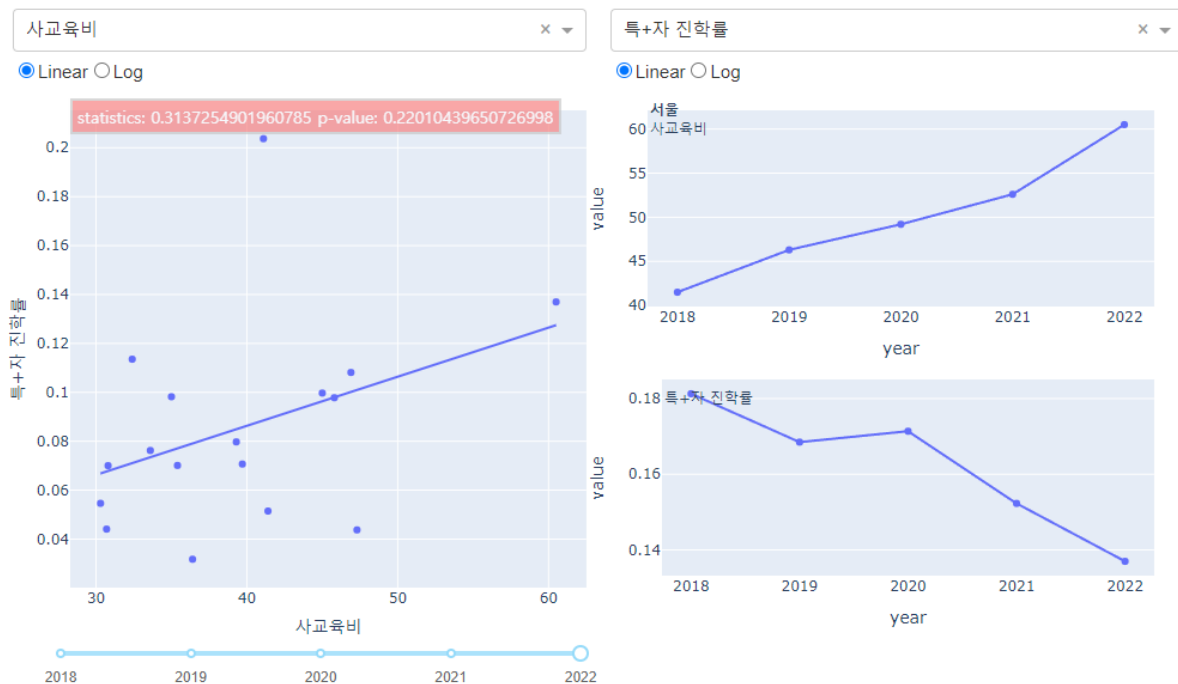
III. 결과

대시보드는 (URL: <https://first-dashboard.onrender.com>)에서 확인할 수 있다.

총 9 개의 데이터 중에서 5 개의 데이터의 정규성 검정 결과 p-value 가 유의수준인 0.05 보다 작기 때문에 귀무가설을 기각하고 대립가설을 채택해서 데이터가 정규성을 만족하지 못한다고 주장할 수 있다. 따라서 pearson 상관분석이 아니라 spearman 상관분석으로 상관관계를 분석했다.

특성화고진학률 0.03166132792830467
특목고진학률 0.006314253434538841
자율고진학률 7.290593906361664e-09
특+자 진학률 1.7305550500168465e-05
사교육비 0.0028868576046079397
음주율 0.08020992577075958
우울감경험률 0.4732091724872589
흡연율 0.14145895838737488
비만율 0.14381158351898193

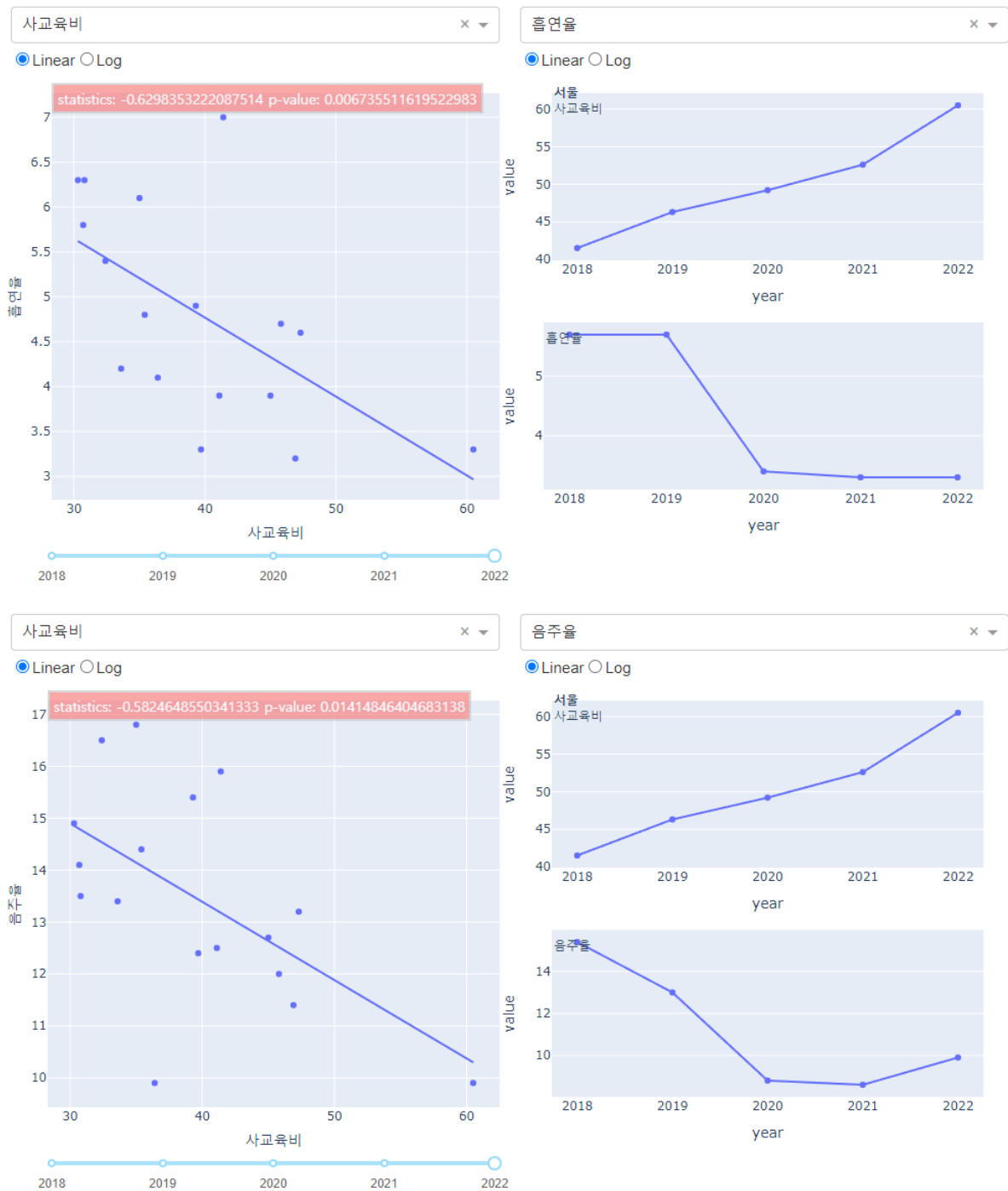
i) 사교육비와 고교 진학률과의 상관관계



[Figure 2] 2022 년 서울의 사교육비와 특목고, 자율고 진학비율의 스피어만 상관계수 분석 그래프

사교육비와 진학률의 상관관계는 2018 년부터 2022 년까지의 17 개 지역 모두 p-value 가 0.05 이상인 결과를 보인다. 따라서 귀무가설을 기각할 수 없어 통계적으로 유의하다고 할 수 없다.

ii) 사교육비와 청소년 비행률과의 상관관계



[Figure 3] 2022 년 서울의 사교육비와 흡연율의 스피어만 상관계수 분석 그래프

사교육비와 청소년 비행률의 상관관계는 2018 년부터 2022 년까지의 도합 10 개 년도 중 7 개 년도가 p-value 가 0.05 이하의 값을 가져 통계적으로 유의하다는 결과를 나타냈다.

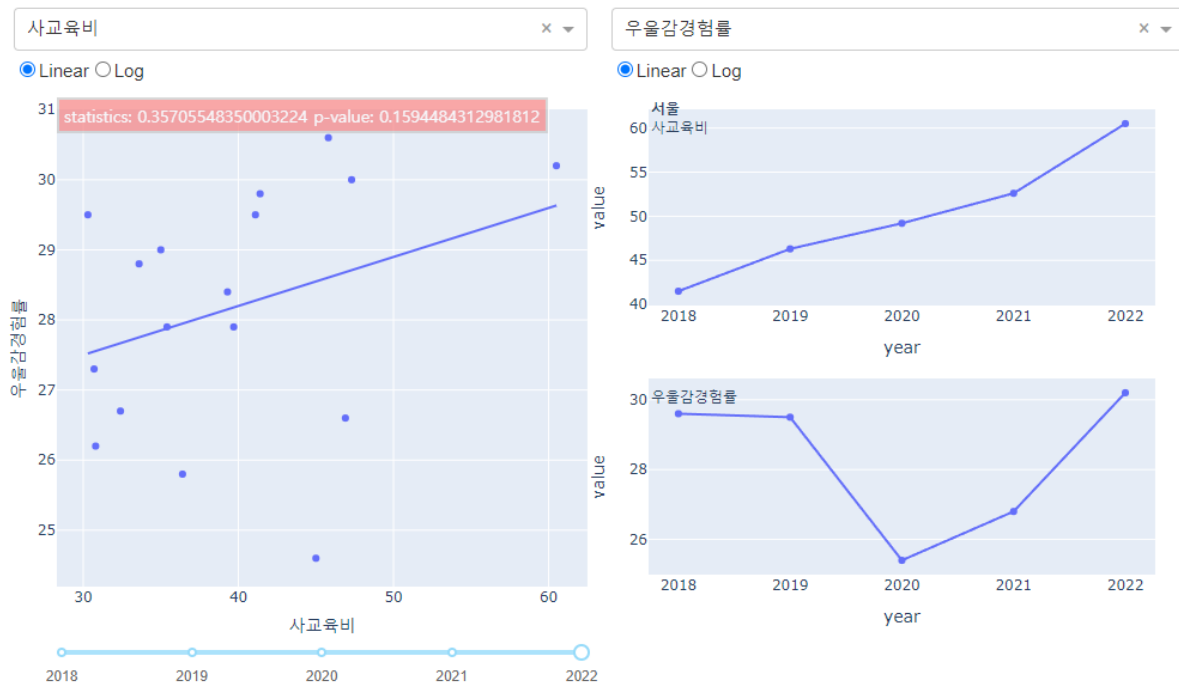
iii) 사교육비와 청소년 비만율과의 상관관계



[Figure 4] 2022 년 서울의 사교육비와 비만율의 스피어만 상관관계 분석 그래프

사교육비와 비만율의 상관관계는 2020 년을 제외하면 모두 p-value 가 0.05 이하의 값을 가져 통계적으로 유의하다는 결과를 나타냈다.

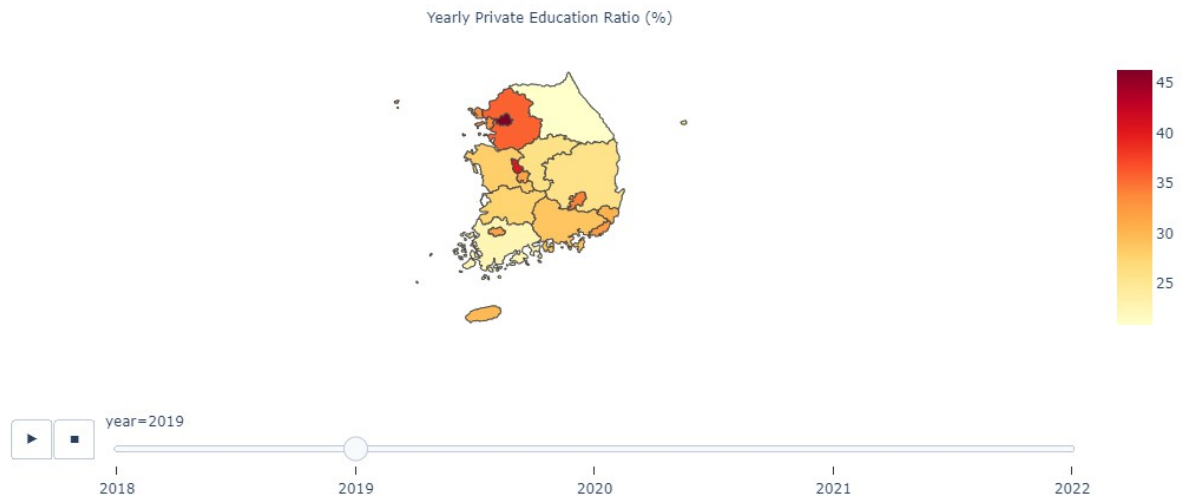
iv) 사교육비와 청소년 우울감 경험과의 상관관계



[Figure 5] 2022 년 서울의 사교육비와 우울감 경험률의 스피어만 상관관계 분석 그래프

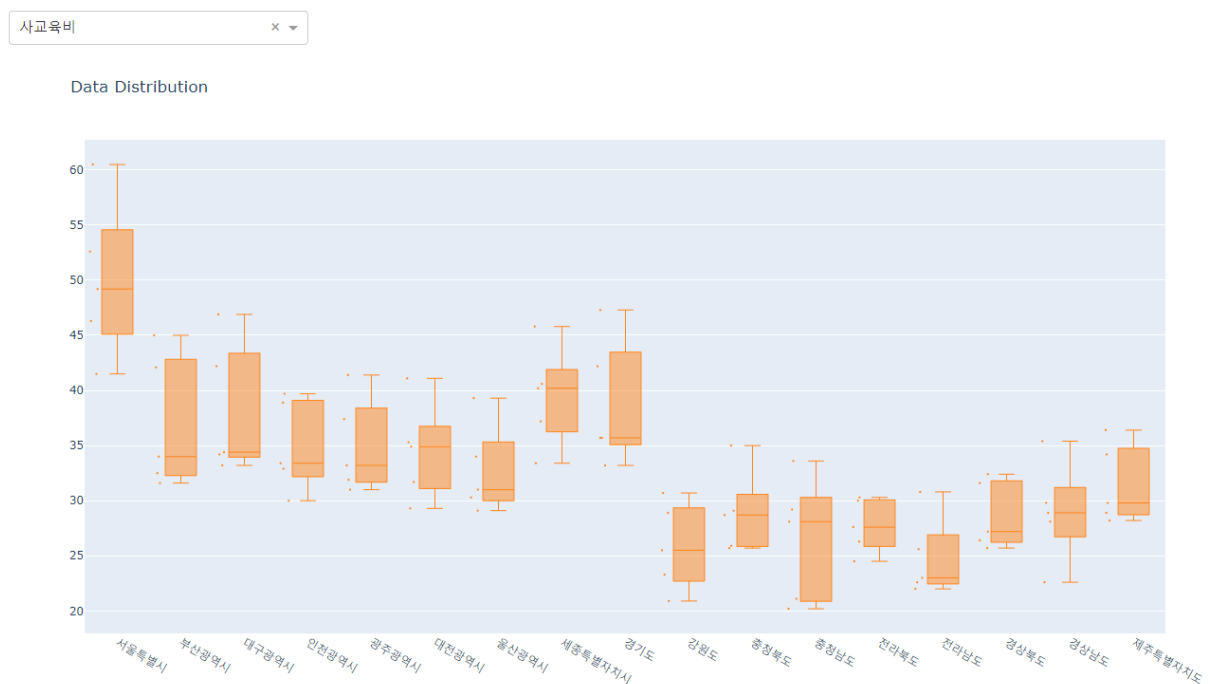
사교육비와 우울감 경험률은 2018~2022 년 모두 p-value 가 0.05 이하로 통계적으로 유의하지 않았다.

결론적으로 사교육비와 진학률 및 우울감 경험률은 상관관계가 없으며, 사교육비와 청소년 비행률 및 비만율과는 상관관계가 있다는 결과를 얻을 수 있다. 사교육비 지출이 높아질수록 청소년 비행률과 비만율이 유의하게 감소하는 경향을 보인다.



[Figure 6] 사교육비 지출 정도를 지도를 이용하여 시각화한 그래프 (2018~2022)

지방보다 도심지역에 가까울수록 사교육비 지출이 더 높아진다. 그래프를 통해 서울지역에 가까워질수록 사교육비 지출이 높아지는 양상을 확인할 수 있다.



[Figure 7] 지역별 변수 값에 따른 box plot (2018~2022)

각 지역별로 드롭다운 메뉴에 선택된 변수의 값과 년도에 따른 분포를 볼 수 있다. 사교육비의 경우에는 서울과 세종, 경기도가 높은 것을 확인할 수 있다.

IV. 논의

i) 사교육비와 특목고 진학률에는 유의미한 상관관계가 없었다. 고등학교 진학을 타 지역으로 가는 경우는 대학 진학에 비해 흔하지 않은 만큼 사교육비나 학생들의 성취도보다는 지역 내의 특목고 개수가 더 많은 영향을 끼치는 것으로 보인다. 실제로 지역의 특목고 진학 비율 중 도내 진학 비율이 도외 진학비율 보다 더 많이 차지한다.

ii) [Figure 3]를 보면 사교육비와 흡연율, 음주율 등의 청소년 비행이 유의미한 상관관계를 나타냈다. 음주율의 경우 2019~2022년, 흡연율의 경우 2018, 2020, 2022년이 p-value 0.05 이하의 음의 상관관계를 나타냈다. 또한 [Figure 4]의 사교육비와 비만율도 2020년을 제외하면 모두 p-value 0.05 이하로 유의미한 음의 상관관계를 나타냈다. 일반적인 상식으로는 사교육비의 증가를 학습량의 증가 및 운동량의 하락으로 연결 지을 수 있다고 생각했지만, 결과는 반대로 나타났다. 위의 두 결과를 통해 높은 사교육비를 부모의 관심 및 좋은 생활 환경과 연관 지어 생각해 볼 수 있을 것이다.

iii) 사교육을 청소년의 우울감이나 자살위험과 연결 짓는 경우가 많아 사교육비와 우울감 경험률의 연관관계를 살펴보았으나 5개년 모두 연관관계가 없었다. 사교육은 수많은 우울 유발 요인 중의 하나일 뿐인 데다 ii)의 논의와 마찬가지로 사교육비가 높다는 것은 자녀에 대한 관심이나 부모의 높은 수입, 좋은 생활 환경을 기대할 수 있기 때문으로 보인다. 다만 "사교육비 지출이 청소년 자녀의 우울과 신체증상에 미치는 영향."(이성림, 김진숙, 2022)의 연구에서는 음의 상관관계가 보고되었는데, 이는 수집한 데이터의 차이로 인한 결과로 보인다.

iv) 사교육비가 올라갈수록 청소년 비행과 비만 등의 부정적인 수치가 내려간다는 결과를 통해 사교육을 긍정적인 가치와 연결 지을 수 있다는 의의가 있다. 교육을 제공하는 회사의 입장에서 사교육이 연일 ‘불평등’이나 ‘카르텔’, ‘학대’ 등의 부정적인 단어와 엮이는 것이 불편할 수 있다. 그러나 이번 분석을 통해 사교육을 부모의 관심이나 사랑, 노력 등의 긍정적 가치와 연결 짓는다면 대중을 설득하고 나아가 매출을 증대하는 데 도움이 될 수 있다.

v) 개인이나 각 가정의 데이터가 아닌 시·도별로 뭉뚱그려진 데이터를 사용했기 때문에 표본의 수가 적고 변수를 통제할 수 없다는 한계가 있다. 예를 들어 고교별 진학률같은 경우 각 도시 내에서 개인별로 사교육비 지출과 특목고 진학률의 상관관계를 알아봤다면 도시 내 특목고의 입학 정원이라는 변수를 통제할 수 있고 가설 1이 채택됐을 가능성도 있을 것이다.

V. 코드

프로젝트를 수행하기 위해 만든 데이터프레임과 코드는 다음의 Github에서 확인할 수 있다.
(URL: <https://github.com/bigdata4th-first-line/edustats.git>)

VI. 사용 라이브러리

dash Python package (version 2.9.3)
Flask Python package (version 2.2.2)
pandas Python package (version 1.5.3)
plotly Python package (version 5.9.0)
scipy Python package (version 1.10.0)
statsmodels Python package (version 0.14.0)
gunicorn Python package (version 21.2.0)

VII. 참고문헌

박현정, 이준호 "중학생의 특수목적고등학교 진학계획이 사교육 참여 및 사교육비 지출에 미치는 영향 분석" 아시아교육연구 10.3 pp.213-238 (2009) : 213.

이덕난, 유지연 "초·중·고교 사교육비 변화 추이 분석 및 향후 과제" NARS 현안분석 247 호, 국회입법조사처, 2022.4.28.

박선영, 마강래 "지역의 교육환경이 사교육비 지출에 미치는 영향에 관한 연구" 지역연구 31. 3 pp.3-17 (2015) : 3.

서은경, 김현철, 황수진, 이철원 "특목고 진학계획의 사교육비 지출에 대한 인과효과 분석" 교육학연구 56.2 pp.161-190 (2018) : 161.

하준경 "특목고가 가계의 사교육비 지출에 미치는 영향" 경제분석 16.3 pp.159-194 (2010) : 159.

박소영, 민병철 "특수목적고 진학 결정요인 분석" 한국교육 36.4 pp.29-48 (2009) : 29.

이성림, 김진숙. "사교육비 지출이 청소년 자녀의 우울과 신체증상에 미치는 영향." Human Ecology Research(HER) 60.2 (2022): 289-302.

신현성, “가구소득, 사교육비 높을수록 특목고, 자사고 진학률↑” 서울뉴스통신, 2021.10.7 s
nakorea.com/news/articleView.