

---

2024 통계자료분석경진대회  
경기대학교 학생들의  
통학 환경 개선을 위한  
셔틀버스 노선 개설 지역 선정

---



제출일	2024.11.10	팀명	통통대장
학번	202111168	이름	김다빈
학번	202211293	이름	이예진
학번	202240322	이름	정연우
학번	202211362	이름	최세현

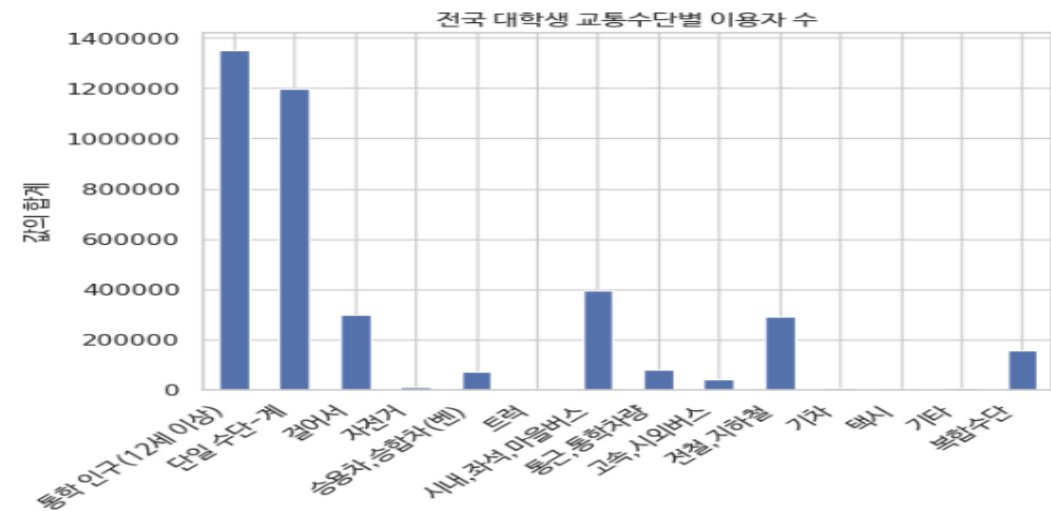
## - 목차 -

I . 서론 .....	3
II . 본론 .....	5
A. 이론적 배경 .....	5
B. 설문조사 .....	5
C. 자료 수집 .....	6
D. 데이터 전처리(data preprocessing) .....	8
E. TOPSIS 기법 .....	9
i . TOPSIS 기법의 개요.....	9
ii . TOPSIS 기법의 절차.....	10
iii . TOPSIS 기법의 적용.....	11
iv . 분석 결과.....	13
III . 결론.....	14
A. 세부 선정 셔틀버스 운영 지역.....	14
B. 셔틀버스 운영 관련 추가 제안 및 기대효과.....	15
C. 한계점.....	16
D. 결론.....	16
IV . 참고문헌.....	17
V . 데이터 출처.....	18

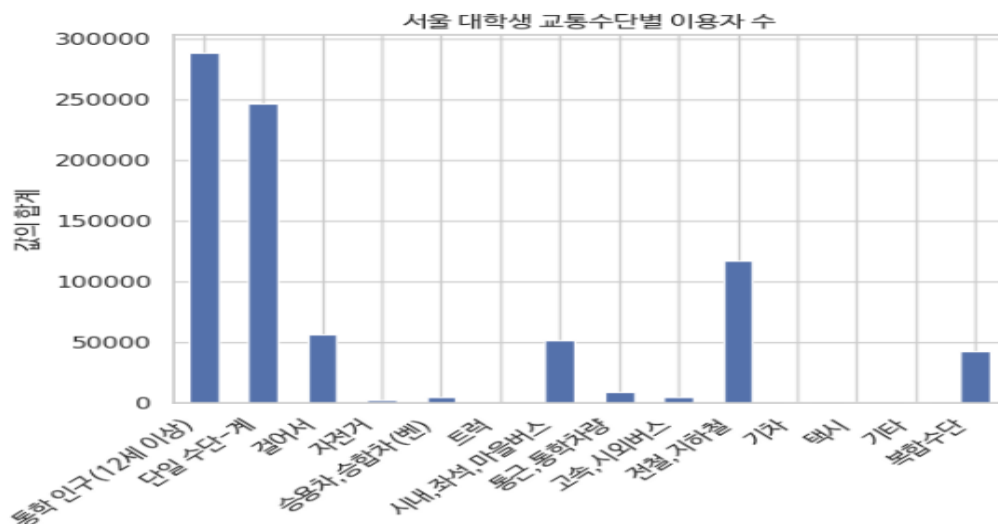
## I. 서론

최근 대학생들의 학업 성취도와 삶의 질에 대한 관심이 높아지면서, 대학 생활에서 통학 환경이 학생들에게 미치는 영향도 주목받고 있다. 이에 관한 근거로 여러 연구에서, 통학 시간이 길어질수록 학생들의 학업 몰입도가 낮아지고 이는 학업 중단으로 이어질 가능성이 높아진다는 명시되어 있다. 특히 불편한 통학 환경은 주말과 평일 생활 리듬을 깨트리며 학생들로 하여금 큰 스트레스를 유발하여 학업과 개인 생활 전반에 부정적인 영향을 미친다.

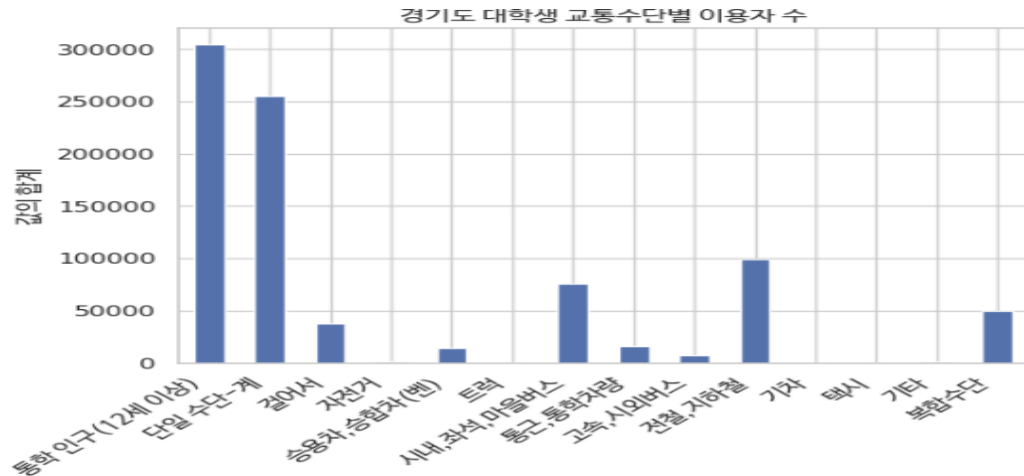
전국 및 서울·경기 지역 대학생들의 주요 통학 수단으로는 마을버스와 지하철이 가장 많이 이용되며 특히 서울과 인천 지역 20~29 세 사람들의 경우 평균 통학 시간이 80 분을 약간 넘는다. 실제 대중교통수단별 평균 통학 시간은 버스와 지하철 모두 약 80 분 내외로 장시간 통학에 따른 피로와 시간 부담이 큰 상황임을 보여준다.



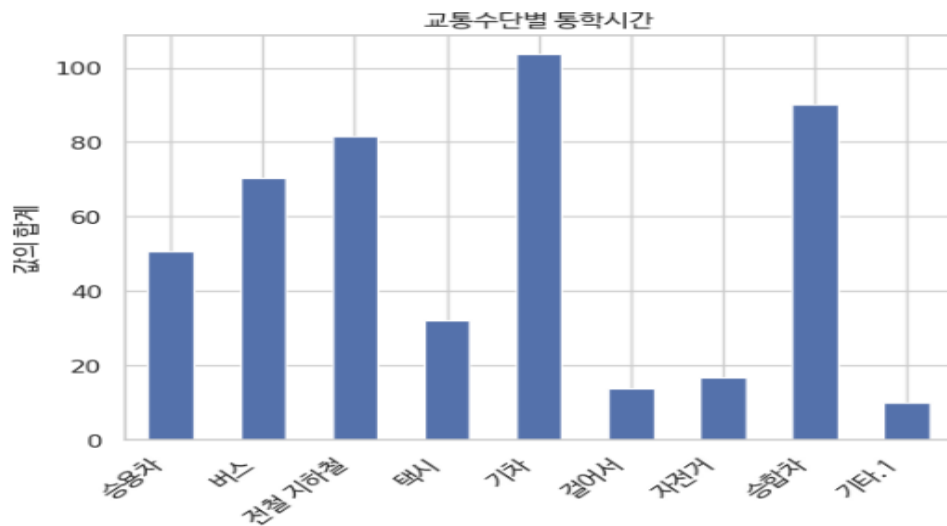
[그림 전국 대학생 교통수단별 이용자 수]



[그림 서울 대학생 교통수단별 이용자 수]



[그림 경기도 대학생 교통수단별 이용자 수]



[그림 교통수단별 통학 시간]

이와 관련해 경기대학교 학생들 사이에서도 통학에 대한 불편과 불만이 꾸준히 제기되고 있다. 학교 커뮤니티인 “에브리타임”에서 셔틀버스 승차권 부족으로 인해 버스 이용에 어려움을 겪거나, 높은 통학 비용에 부담을 느끼는 글을 쉽게 찾아볼 수 있다. 특히 특정 시간대에 통학 노선 혼잡도가 매우 높아 출퇴근 시간에 학생들이 버스를 타지 못해 지각하는 일도 빈번하다. 이와 관련한 예시로, 사당역에서 출발하는 7000번 광역버스의 경우 평균 재차 인원이 49명으로 좌석 수인 43석을 초과하며 수원역에서 학교로 향하는 버스도 평균 34명에서 최대 60명까지 혼잡해 심각한 불편을 유발하고 있다.

따라서 “경기대학교 학생들의 통학 환경 개선을 위한 통학 셔틀버스 노선 개설 지역 선정”을 통해 학생들이 겪는 통학의 불편함을 개선하고 더 나은 학업 환경을 제공함으로써 학업 몰입도 향상과 삶의 질 개선에 기여하고자 한다.

## II. 본론

“경기대학교 학생들의 통학 환경 개선을 위한 통학 셔틀버스 노선 개설 지역 선정”을 주제로 설정한 이후 자료 분석을 진행하였다. 자료 분석 이전에 선행연구를 통한 이론적 배경을 선행적으로 조사하였다.

### A. 이론적 배경

통학 또는 통근과 관련된 연구는 꾸준히 이루어져 왔다. 대학생들의 통학과 관련한 연구를 중심으로 분석하였고 그 중에서 통학 환경에 영향을 줄 수 있는 요소들을 찾고자 하였다. “대학생 통학”과 연관된 연구의 경우에는 강의 인식, 대학 만족도, 학업 만족도 등 통학으로 인하여 학업 지속에 영향을 주는지 판별하는 연구와, 통학 통행실태와 사회경제학적 특성을 바탕으로 한 통학 영향 분석 등의 연구가 이루어졌다. 이 중 통학 영향 분석과 통학(통근) 시간 및 행태를 결정하는 요인에 대한 연구를 위주로 확인하였다.

통학 시간 및 행태는 지역 및 개인적 특성에 의하여 영향을 받으며, 소득계층, 주거비용, 해당 지역의 특성까지 여러 요인이 해당 특성에 반영될 수 있다. 연구별로 영향을 미치는 변수를 아래 표 1에 정리하였다.

연구	영향을 주는 변수
장재민, 김태형(2016)	아파트 시세차익, 종사자수
이번송(1998)	소득, 주거지 소유 형태 등 통근자의 특성 거주지 이동, 통근 목적지와 주거지 특성
전명진, 이지현(2020)	접근성, 높은 주택 가격
송재민(2021)	인구밀도, 승용차, 대중교통 및 비동력 통행
Zheng et al.(2022)	도시 건조 환경(city built environment), 거주인구, 오락시설, 복합 개발 등 노동인구, 교통시설
윤대식(2001)	통행비용, 도보이동시간 등
신상영(2003)	도시성장관리 측면의 요인

[표 1 선행 연구]

선행 연구들을 종합해 보면, 통학(통근) 시간 및 거리의 증가와 같이 통학에 미치는 영향요인은 주거지와 통학 시에 발생하는 통행, 통학 목적지의 특성 등이 있다. 본 분석에서는 경기대학교 수원캠퍼스 재학생의 통학 문제를 해소하기 위해, 경기대학교의 접근성 및 도시 인프라적 특성을 반영하여 분석을 진행하였다.

### B. 설문조사

경기대학교 학생들이 겪고 있는 통학환경의 현황과 만족도를 구체적으로 파악하기 위해 2024년 9월 30일부터 10월 9일까지 10일간 설문조사를 실시하였다. 설문조사는 google form 을 통해 작성되었고, 대학교 커뮤니티인 “에브리타임”에 게시하여 참여를 유도하였다. 질문지법을 활용하여 학생들의 통학 환경에 대한 의견을 수집하였다. 설문지는 일반적인 질문에서 구체적인 질문으로 이어지도록 구성하여 응답의 논리성과 일관성을 높였고, 응답 항목이 중복되지 않도록 하였다.

총 103명이 설문조사에 참여하였고 이 중 유효한 표본 101개를 수집하였다. 설문 결과, 통학 환경에 대한 ‘매우 불만족’ 응답이 33.7%, ‘불만족’ 응답이 22.8%로 나타나, 상당수의 학생이 통학 환경에 불만족을 표하고 있어 통학 환경 개선을 원하고 있음을 알

[그림 설문조사 중 대중교통 이용 시 불편한 점]

통학 교통수단 개수에 대한 질문에서는 ‘2가지’라고 응답한 비율이 63.5%로 가장 높아, 다수의 학생이 최소 한 번 이상의 환승을 통해 통학하고 있는 것으로 나타났다.

### C. 자료 수집

앞서 살펴본 선행 연구와 설문조사를 바탕으로 자료를 수집하였다. 분석 대상은 경기대학교 수원 캠퍼스의 통학권 중 서울특별시, 경기도, 인천광역시의 지방자치단체인 시·군·구로 설정하였다. 자치단체 수는 서울특별시 자치구 25 개, 경기도 자치시 28 개 및 자치군 3개, 인천광역시 자치구 8개와 자치군 2개로 총 66개이다. 학생들의 거주환경과 통학 수요를 반영하는 것이 중요하다 판단하여, 경기대학교 학생들 대상의 설문조사 중 “1-5. 귀하의 거주지는 어디인가요? (시군구까지)”에 응답된 시·군·구를 최종 자료 분석 대상으로 설정하였다. 이에 따라 총 32 개의 시·군·구를 기준으로 통학과 관련한 자료를 수집하였다.

통학 환경에 영향을 줄 수 있는 요소들 중 접근성, 통학의 용이성, 교통 환경 등과 관련한 키워드를 선정한 후 교통 정보 플랫폼, 공공 데이터 포털 등을 통해 데이터를 수집하였다. 수집한 데이터 목록은 아래 표와 같다.

데이터 명	데이터 출처	수집 시점
대중교통 이용인원	교통카드빅데이터통합정보시스템	2024 년 9 월 9 일
이용량 지표	교통카드빅데이터통합정보시스템	2024 년 9 월 9 일
정류장공급도	교통카드빅데이터통합정보시스템	2024 년 9 월 9 일
행정동별 연령별 인구현황	행정안전부	2024 년 9 월
여객목적통행 OD	경기도 교통정보센터	2021 년
경기대학교까지 소요시간	네이버 지도	2024 년 11 월 4 일

[표 2 데이터 목록]

데이터 설명은 아래와 같다.

- 대중교통 이용 인원: 교통카드 이용 인원을 기준으로 하루 중 대중교통을 이용한 인원수를 나타낸다.



[그림 대중교통이용인원].

- 이용량 지표: 버스나 지하철을 탈 때마다 발생하는 승·하차 이용량으로 교통카드 데이터를 기반으로 측정되었다. 출근시간(07 시- 09 시), 12 시와 14 시 강의 시간표를 고려한 낮 시간대(10 시-14 시), 퇴근시간(17 시-19 시)의 평균 통행량을 계산하였다.

대중교통 이용 인원과 이용량 지표는 해당 지역의 통행량과 혼잡도를 추산하고자 수집하였다.

- 정류장 공급도: 지역별 시가화 면적 대비 정류장 수의 비율이다. 여기서 시가화 면적은 용도지역 중 주거지역, 상업지역, 공업지역 면적의 합으로 산출한다. 정류장은 버스정류장과 지하철역으로, 정류장이 많을수록 정류장 공급도의 수치가 크고, 해당 지역의 대중교통 접근성이 높음을 의미한다.
- 행정동별 연령별 인구현황: 각 시·군·구별 경기대학교 재학생 수를 반영하고자 수집 데이터로 설정하였다. 시·군·구별 경기대학교 재학생과 관련한 데이터를 학사 혁신처 측에 문의하였으나, 개인정보보호로 인해 접근이 어려워 대체 자료로 20 대 인구 데이터를 활용하였다. 남성은 20-26 세, 여성은 20-24 세로 대학생 연령대로 맞추어 설정하였다. 하지만 해당 인구가 경기대학교 재학생 수를 완전히 반영하기에는 한계가 있다고 판단하여 통계적 근거를 확인하고자 하였다. 이를 위해, 경기대학교 재학생을 반영할 수 있는 설문조사 데이터를 활용하여

시·군·구별 답변자 수를 측정하고, 20 대 인구 데이터와 해당 자료가 동일한 분포를 보이는지 가설검정을 통해 확인하였다.

분포의 동일성 검정은 Kolmogorov-Smirnov 검정(이하 K-S 검정)으로, 분포의 동일성에 대한 비모수(nonparametric) 검정이다. 두 샘플이 동일한 분포에서 나왔는지를 Kolmogorov-Smirnov 통계량을 통해 확인하며, 이는 두 표본의 경험적 분포 함수 간 거리를 정량화한 값이다. 검정통계량은 아래 식과 같다. 여기서,  $F_{1,n}$ 은 첫 번째 샘플의 경험적 분포 함수,  $F_{2,m}$ 은 두 번째 샘플의 경험적 분포함수이며,  $n, m$ 은 각각 첫 번째, 두 번째 샘플의 크기이다.

$$D_{n,m} = \sup |F_{1,n}(x) - F_{2,m}(x)|$$

먼저, 귀무가설과 대립가설을 설정하였다.

- 귀무가설(H0): 두 자료의 동일한 분포를 따른다.
- 대립가설(H1): 두 자료의 다른 분포를 따른다. (not H0)

다음으로, 신뢰수준은 99%로 설정하였다. 설문조사에서 유효 표본이 101 개로 경기대 전체를 온전히 반영하기에는 한계가 있다고 판단하여, 더 넓은 신뢰수준을 토대로 가설검정을 진행하였다. R의 ks.test() 메서드를 사용하여 검정통계량과 유의확률을 계산하였고, 양측검정(two-sided)을 수행하였다. 그 결과, 검정통계량  $D_{32,32}=0.375$ ,  $p\text{-value}=0.0139$ 로 유의수준 99%하에 귀무가설을 채택한다. 이를 통해 두 자료가 동일한 분포를 가진다고 판단할 수 있다. 해당 가설검정의 결과를 바탕으로, 행정동별로 나누어진 20 대 인구 데이터를 시·군·구 단위로 묶어 분석에 사용하기로 결론 내렸다.

- 여객목적통행 OD: 각 시·군·구에서 경기도 수원시 영통구까지의 통행량 중 여객 목적 중 “가정 기반 통학”인 통행량을 수집하였다.
- 경기대학교까지 소요시간: 각 시·군·구의 교통적 중심지에서 경기대학교 수원 캠퍼스까지 대중교통 이용할 때 소요 시간을 네이버 지도 기준, 11 월 4 일 월요일 오전 7 시로 설정하여 계산하였다. 예를 들어, 경기도 고양시는 정발산역, 인천광역시 연수구는 연수역을 교통적 중심지라고 설정하였으며, 중심지는 환승 및 접근 용이성, 지리적 중심 여부 등을 고려하여 선정하였다.

이러한 데이터를 바탕으로 최종 6 개의 변수와 32 개의 개체로 구성된 데이터 프레임을 형성하여 분석에 활용하였다.

#### D. 데이터 전처리(data preprocessing)

앞선 데이터 수집 단계에서 모은 32 개의 시·군·구별 데이터를 하나의 데이터 프레임(dataframe)으로 통합하였고, 데이터 클리닝(cleaning)과 데이터 통합 과정을 거쳐 전처리를 수행하였다. 데이터 프레임에서 각 변수의 이름은 아래와 같다.

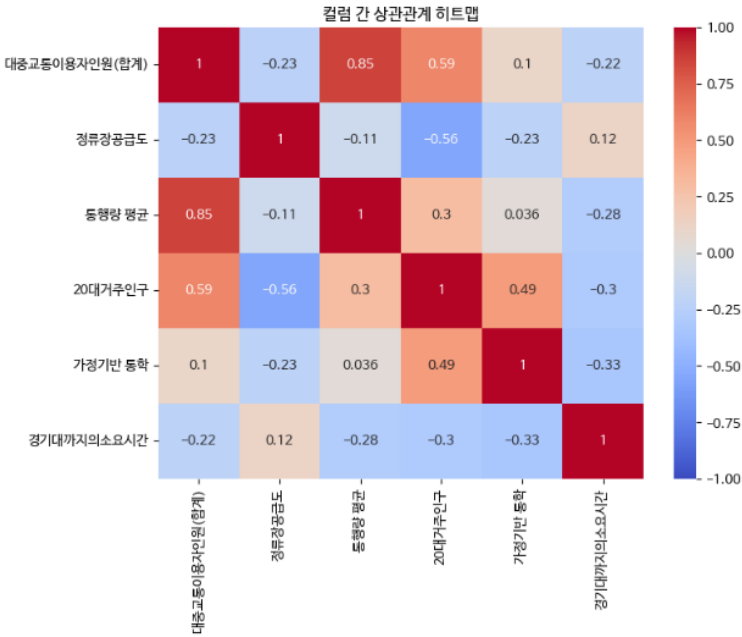
데이터 명	변수 명
대중교통 이용인원	대중교통이용자인원(합계)
정류장공급도	정류장공급도
이용량지표	통행량 평균
행정동별 연령별 인구현황	20 대거주인구
여객목적통행 OD	가정기반 통학
경기대학교까지 소요시간	경기대학교까지의소요시간

[표 3 데이터별 변수 명칭]



결측치(NaN)와 이상치 여부 확인, 변수 선정의 과정을 이어서 진행하였다. 전체적인 수치 확인과 시각화를 통해 결측치와 이상치가 없음을 확인하였다. 다음으로 변수 선정의 경우에는 변수 수집 목적의 중복성과 분석의 적합성 등을 우선적으로 고려하였다. 목적과 의미가 중첩되는 변수로는 통행량과 혼잡도 파악을 위한 “대중교통이용자인원(합계)”와 “통행량 평균”이 있었고, 두 변수 중 더 적합한 변수를 선택하고자 다중공선성 문제를 통해 확인하였다.

다중공선성 문제(Multicollinearity)는 회귀분석에서 독립 변수들 간에 강한 상관관계가 나타날 때 발생하는 문제이다. 비록 본 자료 분석은 회귀 분석은 아니지만, 강한 상관관계가 나타날 수 있는 변수 중 하나를 선택하고자 하는 목적에서 다중공선성 개념을 참고하여 변수 선택에 적용하였다. 우선, 변수들 간의 표본상관계수를 구하여 히트맵 시각화를 하였다.



[그림 변수 간 상관관계 히트맵]

결과를 보았을 때, “대중교통이용자인원(합계)”와 “통행량 평균”은 0.85로 강한 양의 상관관계를 가지고 있다. 다음으로, 분산팽창요인(VIF) 지수를 구하였고, 10 이 넘는다면 다중공선성의 문제가 있다고 판단하겠다.

변수	VIF 지수
대중교통이용자인원(합계)	8.161373
통행량 평균	5.716222

[표 4 대중교통이용자인원(합계)와 통행량 평균의 VIF]

산출된 분산팽창요인은 위 표와 같다. 두 수치 모두 10 이 넘지 않으나 두 변수 중 하나를 선택하고자 하였기 때문에 VIF 지수가 더 낮은 변수인 이용량 지표(이하 통행량 평균)를 최종 변수로 선정하였다.

최종적으로 선정된 변수는 ‘정류장공급도’, ‘통행량 평균’, ‘20 대거주인구’, ‘가정기반 통학’, ‘경기대학교까지의소요시간’, 총 5 개이다.

E. TOPSIS 기법

- i. TOPSIS 기법의 개요

TOPSIS란 The Technique for Order of Preference by Similarity to Ideal Solution의 준어로, 의역하면 “이상적인 해결에 대한 유사성에 의한 순위 결정 기법”이다. 다중 기준 의사 결정 분석 방법으로 1981년 Yoon, K이 처음 개발하였다. TOPSIS의 기본 아이디어는 기하학적 거리 개념을 기반으로 설명된다. 양의 이상적인 해(Positive Ideal Solution, 이하 PIS)로부터 가장 가까운 거리를 가지고, 음의 이상적인 해(Negative Ideal Solution, 이하 NIS)로부터 가장 긴 거리를 가져야 한다는 원리를 기반으로 한다. PIS와 NIS는 각 평가요소에 있어서 최선의 평가치와 최악의 평가치이다. TOPSIS 기법의 경우에는 상대적 거리를 기반으로 이상적인 대안을 선택하는 의사결정 방법론으로써, 절대적인 선택 기준은 존재하지 않으며 의사결정자의 상황에 따라 결과가 좌우된다.

TOPSIS 기법의 장점은 가중치를 생성하는 것보다 대상을 평가하는 과정에 초점을 두어 간단한 계산 과정으로 최적의 대상을 빠르게 찾을 수 있다는 점이다. 또한 최선책과 최악책을 동시에 고려하여 합리적 선택을 도출할 수 있고, 논리적이면서 표현이 용이하다.

본 자료 분석에서는 이상적인 최선의 해에 가까울수록 통학 셔틀버스 설치 지역으로 적합하다고 판단하였고, 반대로 이상적인 최악의 해에 가까울수록 통학 셔틀버스 설치 기역으로 부적합하다고 보았다. 최종적으로, 셔틀버스 설치 대상인 시·군·구를 기준으로 이상적인 최선의 해에 가까운 상위 5개 지역을 우선적으로 선정하고자 하였다.

## ii. TOPSIS 기법의 절차

1 단계: 평가 행렬 생성

m 개의 대안과 n 개의 기준으로 구성된 평가 행렬  $(x_{ij})_{m \times n}$ ,  $i = 1, 2, \dots, m$ ,  $j = 1, 2, \dots, n$ 을 생성한다.

2 단계: 정규화 행렬 형성

앞서 생성한 평가 행렬인  $(x_{ij})_{m \times n}$ 을 정규화 방법을 사용하여 정규화 행렬을 만든다.

$$R = (r_{ij})_{m \times n}$$

$$r_{ij} = \frac{x_{ij}}{\sqrt{\sum_{k=1}^m x_{kj}^2}}, i = 1, 2, \dots, m, j = 1, 2, \dots, n$$

3 단계: 가중치 계산

정규화 평가 행렬에 각 평가 항목에 대한 가중치를 곱하여 가중 정규행렬을 구한다. 가중치는 분석가와 연구자의 기준에 따라 설정되고, 본 자료 분석에서는 평가 항목의 쌍대비교행렬을 통하여 가중치를 계산하였다.

쌍대비교행렬(pairwise comparison matrix)은 계층화 분석법(AHP)에서 사용되는 정방행렬이며, 평가 기준이 다수이며 복합적인 경우에 항목 간 상대적 비교 값을 나타낸 행렬이다. 각 기준이 상대적으로 얼마나 중요한지 평가하기 위해 각 항목을 1~9 사이의 척도를 사용하여 비교하여 작성하였다. 홀수인 1, 3, 5, 7, 9는 수치가 상승할수록 중요도가 증가하고, 짝수인 2, 4, 6, 8은 중간의 값이나 세부적 평가에 사용된다. 상대적으로 반대의 경우에는 역수를 사용한다. 예를 들어, 요소 A가 요소 B에 대해 중요도가 3이라면, B에 대한 A의 값은 1/3이 된다. 이러한 과정을 통해

각각의 변수의 비율을 이용하여 행렬로 나타내고 기하평균을 계산한 후 정규화하여 가중치를 계산한다.

쌍대비교 행렬  $A_{n \times n}$  를 가정하면, 기하평균(Geometric Mean)은 다음과 같이 계산된다.

$$GM[i] = (A[i, 1] \times A[i, 2] \times \dots \times A[i, n])^{1/n}$$

$GM[i]$ 는 각 항목의 상대적 중요도를 나타내는 값이 된다. 가중치는 그 항목의 기하평균을 기하평균의 총합으로 나눠 구한다.  $i$ 번째 항목의 가중치  $W[i]$ 는 다음과 같이 계산된다.

$$W[i] = \frac{GM[i]}{\sum_{i=1}^n GM[i]}$$

계산된 가중치를 이용하여 가중 정규화 결정 행렬  $t_{m \times n}$ 을 계산한다.

$$t_{ij} = r_{ij} \cdot w_j, i = 1, 2, \dots, m, j = 1, 2, \dots, n$$

4 단계: 이상해(Positive Ideal Solution)과 비이상해(Negative Ideal Solution) 설정

PIS 을  $D^+$ , NIS 를  $D^-$ 일 때, 다음과 같이 결정한다.

$$D^+ = (t_1^+, t_2^+)$$

$$D^- = (t_1^-, t_2^-)$$

단,  $t_j^+ = \max(t_{ij}), t_j^- = \min(t_{ij}), i = 1, 2, \dots, m$

5 단계: 각 평가 대상과 PIS 의 유클리디안 거리  $D_i^+$ , NIS 의 유클리디안 거리  $D_i^-$  계산

$$D_i^+ = \sqrt{\sum_{j=1}^n (t_i^+ - t_{ij})^2}, i = 1, 2, \dots, m$$

$$D_i^- = \sqrt{\sum_{j=1}^n (t_{ij} - t_i^-)^2}, i = 1, 2, \dots, m$$

여기서,  $n$ 은 평가 기준의 수이다.

6 단계: 각 항목의 PIS 와의 유사성  $c_i$  계산

$$C_i = \frac{D_i^-}{D_i^+ + D_i^-}, i = 1, 2, \dots, m$$

이와 같이 정의되고, 0 과 1 사이의 값을 가진다.

7 단계: 순위 계산

순위를 매겨본 후 이상해와 가까운 항목을 최종 대안(Solution)으로 채택한다.

iii. TOPSIS 기법의 적용

1 단계: 평가행렬 생성

앞서 5 개의 변수와 32 개의 개체로 생성한 최종 데이터 프레임을 평가행렬로 선택하도록 하겠다. 여기서,  $m=32$ ,  $n=5$  이다.

$$(x_{ij})_{m \times n}, \quad i = 1, 2, \dots, 32, \quad j = 1, 2, \dots, 5$$

2 단계: 정규화 행렬 형성

데이터를 정규화하여 정규화 평가행렬  $R$ 을 만들어준다.

$$R_{i \times j}, i = 1, 2, \dots, 32, \quad j = 1, 2, \dots, 5$$

3 단계: 가중치 계산

가중치 계산을 위하여 쌍대비교행렬을 생성했고, 각 변수의 상대적 중요도를 평가하였다. 중요도를 평가하는 기준은 분석가나 의사결정자의 주관에 따라 달라질 수 있으나 본 분석에서는 자료와 객관적인 수치를 기반으로 한 결정을 내리고자 하였다. 이를 위해, 주성분분석(PCA)의 주성분들의 고유벡터(eigenvector)를 활용하여 상대적 중요도를 평가하였다.

주성분분석은 Principal Component Analysis로 P 개의 변수들을 선형변환 시켜 “주성분”이라는 새로운 인공 변수를 생성하는 방법이다. 주성분분석을 통해 차원의 축소, 변수들 간의 관계를 단순화할 수 있다. 주성분은 원 데이터의 변이, 즉 주성분의 분산의 크기에 의해 결정되고, 분산이 가장 큰 주성분(주성분 1)이 원 데이터의 흠어짐에 대한 정보를 가장 많이 가져간다. 이를 통해 다변량 데이터의 변이를 설명할 수 있으며, 본 분석에서도 이러한 방법을 활용하였다.

우선, 평가행렬로 선택한 최종 데이터 프레임을 원 데이터로 설정하였다. 해당 데이터 프레임은 표준화하여 표본상관행렬로 변환해주었다. 그 후, 표본상관행렬을 이용해 주성분을 추정하고, 가장 원 데이터의 변이를 잘 반영하는 주성분 1의 고유벡터(eigenvector)를 추출하였다. 해당 분석 결과로 나온 표본 추정 주성분과 고유벡터는 아래와 같다.

	주성분 1	주성분 2	주성분 3	주성분 4	주성분 5
Eigenvalue	2.2317666	1.0629892	0.93395465	0.59798916	0.33459071
Proportion	0.432404779	0.06482695	0.1158604	0.205954158	0.180953714
Cumulative (%)	43.2405	49.7232	61.3092	81.9046	100

[표 4 주성분 설명력]

	주성분 1	주성분 2	주성분 3
정류장 공급도	-0.4411	-0.44993672	-0.47638
통행량 평균	0.30490	-0.7030686	0.45701
20 대거주인구	0.5851	0.179282	0.16273
가정기반통학	0.4550	0.22223234	-0.5854
경기대까지의 소요시간	-0.40378	0.47087392	0.44153

[표 5 주성분별 고유벡터]

계산된 주성분 1을 활용하여 변수별 주성분 기여도를 구하였다. 기여도는 각 고유벡터 값의 제곱을 주성분 1의 고유벡터 값의 제곱의 합으로 나눈 값을 사용하였다. 기여도는 주성분 1에 해당 변수가 얼마나 영향을 미치는지를 나타내는 지표이다. 아래는 각 변수별 기여도를 정리한 표이다.

변수	정류장 공급도	통행량 평균	20 대거주인 구	가정기반통학	경기대까지의 소요시간
기여도(%)	19.45701	9.296533	34.236439	20.706555	16.303468

[표 6 주성분 1 를 이용한 변수별 기여도]

변수별 기여도를 통해 각 변수의 상대적 중요도를 1 에서 9 까지의 수치로 나타내어  
쌍대비교행렬을 생성하였다. 해당 행렬은 아래와 같다.

$$\begin{bmatrix} 1 & 2 & 1/3 & 1/5 & 3 \\ 1/2 & 1 & 1/5 & 1/7 & 1/3 \\ 3 & 5 & 1 & 1/2 & 7 \\ 5 & 7 & 2 & 1 & 8 \\ 1/3 & 3 & 1/7 & 1/8 & 1 \end{bmatrix}$$

다음으로, 쌍대비교행렬을 이용하여 변수별 가중치를 계산하였다. 각 행의  
기하평균을 기하평균의 총합으로 나누어 최종 가중치를 산출하였다. 변수별 가중치는  
아래 표와 같다.

변수	정류장 공급도	통행량 평균	20 대거주인 구	가정기반통 학	경기대까지 의 소요시간
가중치	0.11386	0.048062	0.298558	0.473371	0.06615

[표 7 변수별 가중치]

4 단계인 이상해 설정 후 5 단계 각 항목별 이상해와의 유클리디안 거리를 계산한다.  
그 후 각 항목의 PIS 와의 유사성을 계산해주고, 점수를 내림차순 정렬하여 순위를  
비교한 후 이상해와 가장 가까운 항목 5 가지를 대안으로 선정하였다.

#### iv. 분석 결과

TOPSIS 기법을 적용하여 선정한 지역은 5 개의 시 · 군 · 구로, 순위와 유사성 점수는  
아래 표와 같다.

순위	시 · 군 · 구	TOPSIS 유사성 점수
1	경기도 고양시	0.760914
2	인천광역시 서구	0.707545
3	인천광역시 연수구	0.660805
4	경기도 용인시	0.657799
5	경기도 하남시	0.624987

[표 8 TOPSIS 기법 최종 결과]

셔틀버스 설치 대상인 시 · 군 · 구를 기준으로 이상적인 최선의 해에 가까운 상위 5 개  
지역이 도출되었으나, 해당 대안을 실제로 적용할 수 있을지에 대한 검토가 필요하다.  
이를 위해, 셔틀버스 유무, 현재 운행 중인 셔틀버스와의 접근성 등을 고려하여 적절한  
대안을 고려해 보고자 하였다. 먼저, 지역 선정 결과 중 현재 셔틀버스를 운행 중인  
지역은 고양시가 유일하며, 고양시는 일산(정발산역) 노선을 통한 셔틀버스를 운행하고  
있다. 다음으로, 현재 운행 중인 셔틀버스와의 접근성을 살펴보았다. 인천광역시로 향하는  
셔틀버스는 부평 3 개, 주안 1 개, 송내(부평) 1 개로 총 5 개의 노선을 운영 중이며,  
경기도 용인시와 경기도 하남시는 인접한 셔틀버스 노선이 존재하지 않았다. 여기서  
“인접함”의 기준은 사전에 설정한 교통 중심지와 셔틀버스 탑승지까지의 소요 시간이  
20 분 이하라고 정의하였다.

인천광역시에는 이미 셔틀버스를 운영 중이나, 노선이 부평 중심으로만 4 개가 이루어져 있다. 또한, 11 월 11 일 월요일 오전 7 시 네이버 지도의 대중교통 이용을 기준으로 인천광역시 서구와 인천광역시 연수구의 교통 중심지인 서구청역과 연수역에서 부평 셔틀버스 탑승지까지 각각 19 분, 20 분이 소요된다.

19분   오전 7:20 도착   1,400원		19분   오전 7:23 도착   1,400원	
16분		2분	
인천2	서구청역	수인분당	연수역
07:01 운영행   가정역 방면		07:04 오이도행   원인재역 방면	
하차	주안역	인천1	원인재역
하차		하차	
인천시청역		인천시청역	

[그림 서구청역과 연수역의 셔틀버스 탑승지까지의 소요시간]

만약 실거주지에서 교통 중심지까지의 소요시간을 추가적으로 고려한다면 통학 시간은 증가할 것이다. 따라서 인천광역시 부평구에만 집중된 노선 외에도 인천광역시의 연수구와 서구를 고려한 노선을 기존 노선에서 대체하거나 추가 설치하는 방안을 제안하고자 한다.

### III. 결론

#### A. 세부 선정 셔틀버스 운영 지역

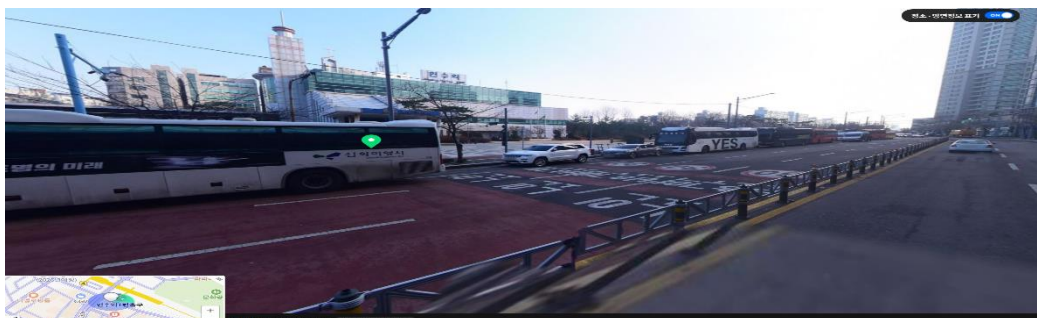
셔틀버스의 세부 정차 지역은 앞서 설정한 교통 중심지를 바탕으로 타 대학교의 셔틀버스 배치 지역이나 대형 기업의 셔틀버스 배치 지역을 참고하여 선정하였다.

인천광역시 서구의 경우, 근처 대학교가 없어 구광 인천 30 센터 주간조 셔틀 버스 노선을 참고하여 서구청역 4 번 출구를 셔틀 버스 정류장으로 최종 선정하였다.



[그림 서구청역 4 번 출구 버스 정류장]

인천광역시 연수구는 위와 같은 이유로 구광 인천 39 센터 주간조 셔틀 버스 노선을 참고하여, 연수역 1 번 출구를 셔틀 버스 정류장으로 최종 선정하였다.



[그림 연수역 1 번 출구]



경기도 용인시의 경우, 강남대학교와 장안대학교를 포함한 여러 대학의 셔틀 버스 정차 지점을 조사한 결과, 대부분의 셔틀버스가 4, 5, 6 번 출구에 정차하는 것을 확인하였다. 이 중 4 번 출구는 학교로 가기 위해 유턴을 해야 하는 점을 고려하여, 4 번 출구와 이를 보완할 수 있는 6 번 출구를 주요 정차 지점으로 최종 선정하였다.



[그림 기흥역 4 번 출구]



[그림 기흥역 6 번 출구]

경기도 하남시의 경우 중심지는 미사역이지만, 주변 학교들의 셔틀 버스 정류장을 고려하여 하남시청역 4 번 출구를 정차 지점으로 최종 선정하였다.



[그림 하남시청역 4 번 출구 도보 3 분거리 장지마을역 버스정류장]

## B. 셔틀버스 운영 관련 추가 제안 및 기대효과

학생들의 통학환경을 더욱 개선하기 위해 아래와 같은 추가적인 방안을 제안한다.

먼저, 정기권 도입에 대한 방안이다. 구체적으로, 주 1 회권, 주 3 회권, 주 5 회권 등 횟수를 다르게 하여 학생들의 통학 패턴에 맞는 선택지를 제공한다. 정기권을 이용하는 학생들에게는 기존 가격보다 저렴한 요금으로 제공하여 통학 비용에 대한 경제적 부담을 줄일 수 있다. 또한 정기권 횟수가 늘어날수록 할인 혜택을 추가적으로 제공하여 셔틀버스에 대한 이용 수요를 늘릴 수 있다.

다음으로는, 셔틀버스 실시간 위치 안내 서비스를 교내 애플리케이션에 제공하는 것이다. 셔틀버스가 기상 악화 또는 교통 체증으로 인해 제시간에 도착하지 않는 경우가 발생할 때 학생들이 셔틀버스 현재 위치에 대한 정보가 없어 불안감을 느낄 수 있다. 실시간 위치 안내 서비스가 교내 애플리케이션 내에 존재한다면, 학교 내 식당 메뉴, 강의 공지사항, 시간표 등을 한곳에서 확인할 수 있는 편리함을 제공할 뿐만 아니라, 학생들이 정확한 탑승 시간을 예측할 수 있어 불안감을 해소할 수 있게 된다.

이와 같은 방안이 실현된다면, 경기대학교 셔틀버스는 학생들의 통학 비용 부담을 줄이고, 통학 환경을 크게 개선하여 보다 효율적이고 편리한 통학 수단이 될 것으로 기대된다.

#### C. 한계점

이번 분석에서의 한계점은 개인정보 보호 문제로 인해 경기대학교 학생들의 정확한 거주지 데이터를 확보하지 못한 점이다. 가설검정을 통해 학생 전체 거주지 데이터를 20 대 거주지 데이터로 대체할 수 있다는 결론을 얻었으나, 이러한 대체 과정으로 인해 분석 결과가 경기대학교 학생들의 실제 통학 환경을 완전히 반영하지 못할 가능성이 있다. 그러나 향후 보다 정밀한 데이터를 확보한다면, 더욱 최적화된 노선 설계가 가능할 것이다.

#### D. 결론

학업 성취도와 삶의 질에 대한 관심이 높아지면서, 대학 생활에서 통학 환경의 중요성이 점차 강조되고 있다. 특히 효율적인 통학 환경을 위해 셔틀버스 노선이 중요한 요소로 떠오르고 있다. 본 자료분석에서는 학생들의 의견을 반영한 설문조사 결과와 통학 환경에 영향을 미치는 데이터를 바탕으로 경기대학교 셔틀버스 노선에 적합한 지역을 선정하는 과정을 다루었다.

먼저 수집된 데이터를 기반으로 데이터 프레임을 구성한 후, 주성분 분석(PCA)을 통해 상대적 중요도를 평가하였다. 이후 AHP 기법을 적용하여 가중치를 산출하고, TOPSIS 분석을 통해 경기도 고양시, 인천광역시 서구, 인천광역시 연수구, 경기도 용인시, 경기도 하남시를 셔틀버스 노선에 적합한 지역으로 선정하였다. 단, 고양시는 이미 정발산역에서 셔틀버스가 운영 중이므로 대상에서 제외하였다. 또한 기존 셔틀버스 노선이 운행 중인 지역을 고려하여 가장 적합한 정류장 위치를 선정하였다. 이 과정에서는 각 지역의 주요 셔틀버스 정류장을 분석하여, 인천광역시 서구는 서구청역 4번 출구 인근 버스정류장, 인천광역시 연수구는 연수역 1번 출구, 경기도 용인시는 기흥역 4번 및 6번 출구, 경기도 하남시는 장지마을역 버스 정류장을 최종 확정하였다.

또한, 학생들의 통학 환경을 더욱 개선하기 위한 추가 제안으로는 정기권 도입과 실시간 위치 안내 서비스 제공을 제안한다.

본 보고서의 분석을 통해 학생들이 실제로 필요로 하는 효율적이고 편리한 통학 환경을 제공할 수 있을 것으로 기대하며, 학업과 삶의 균형을 도모하는 데 기여할 수 있을 것으로 기대된다.



#### IV. 참고문헌

- 예제로 배우는 SAS 다변량 자료분석 입문(강헌철 외)
- 김민경, 강경식. (2016). TOPSIS 방법을 이용한 교육서비스품질 우선순위 선정에 관한 연구. 대한안전경영과학회지, 18(4), 195-209.
- 김종래, 김규태. (1997). TOPSIS 기법을 이용한 공급자 선정 방법. 경영과학, 14(2), 1-17. 쌍대비교행렬 분석 기법을 적용한 스마트 자동 인상 시스템의 성능 분석
- 이희연, 이제연. (2009). 수도권 대학생의 통학통행 흐름의 변화. 국토계획, 44(7), 137-152.
- 신혜숙 and 김미현. (2018). 대학생의 통학시간이 대학 몰입을 매개로 학업중단의사에 미치는 영향. 한국교육문제연구, 36(1), 25-44.
- 전용현, 노정현, 장준석. (2017). 접근성 변수를 반영한 통행발생 및 통행분포모형 개발. 한국산학기술학회 논문지, 18(6), 576-584
- 경기대학교 통학 셔틀버스 운영 현황:  
<https://www.kyonggi.ac.kr/www/contents.do?key=5749>
- Kolmogorov-Smirnov test wikipedia:  
[https://en.wikipedia.org/wiki/Kolmogorov%E2%80%93Smirnov\\_test](https://en.wikipedia.org/wiki/Kolmogorov%E2%80%93Smirnov_test)
- AHP Wikipedia: [https://en.wikipedia.org/wiki/Analytic\\_hierarchy\\_process](https://en.wikipedia.org/wiki/Analytic_hierarchy_process)
- TOPSIS Wikipedia: [https://en.wikipedia.org/wiki/TOPSIS#cite\\_note-Yoon-2](https://en.wikipedia.org/wiki/TOPSIS#cite_note-Yoon-2)
- 다중공선성 Wikipedia:  
<https://ko.wikipedia.org/wiki/%EB%8B%A4%EC%A4%91%EA%B3%B5%EC%84%A0%EC%84%B1>
- “통학 1 시간 이상’ 고등학생, 우울증 위험 1.6 배 높다”, 미디어데일, 2024 년 9 월 3 일 <http://www.mediadale.com/news/articleView.html?idxno=215394>

## V. 데이터 출처

### ① 행정안전부\_지역별(행정동) 성별 연령별 주민등록 인구수

출처: <https://www.data.go.kr/data/15097972/fileData.do>

데이터 제공자(제공기관): 행정안전부

발행연도: 2024 년

### ② 이용량 지표

출처: <https://stcis.go.kr/pivotIndi/wpsPivotIndicator.do?siteGb=P&indiCls=IC01>

데이터 제공자(제공기관): 교통카드빅데이터통합정보시스템

발행시점: 2024 년 9 월 9 일

### ③ 대중교통이용인원

출처: <https://stcis.go.kr/pivotIndi/wpsPivotIndicator.do?siteGb=P&indiCls=IC05>

데이터 제공자(제공기관): 교통카드빅데이터통합정보시스템

발행시점: 2024 년 9 월 9 일

### ④ 정류장공급도

출처: <https://stcis.go.kr/pivotIndi/wpsPivotIndicator.do?siteGb=P&indiCls=IC05>

데이터 제공자(제공기관): 교통카드빅데이터통합정보시스템

발행시점: 2024 년 9 월 9 일

### ⑤ 경기도까지의소요시간

출처: 네이버지도

### ⑥ 통학시 평균 소요 시간(편도 기준)

출처:

[https://kosis.kr/statHtml/statHtml.do?orgId=210&tblId=DT\\_20114\\_2023014\\_06&vw\\_cd=MT\\_ZTITLE&list\\_id=210\\_21004\\_004&seqNo=&lang\\_mode=ko&language=kor&obj\\_var\\_id=&itm\\_id=&conn\\_path=MT\\_ZTITLE](https://kosis.kr/statHtml/statHtml.do?orgId=210&tblId=DT_20114_2023014_06&vw_cd=MT_ZTITLE&list_id=210_21004_004&seqNo=&lang_mode=ko&language=kor&obj_var_id=&itm_id=&conn_path=MT_ZTITLE)

데이터 제공자(제공기관): KOSIS (국가통계포털)

발행연도: 2023 년

### ⑦ 성, 각급 학교 및 이용 교통수단별 통학 인구(12 세 이상)- 시도

출처:

[https://kosis.kr/statHtml/statHtml.do?orgId=101&tblId=DT\\_1PA2016&vw\\_cd=MT\\_ZTITLE&list\\_id=A11\\_2015\\_1\\_001\\_007&seqNo=&lang\\_mode=ko&language=kor&obj\\_var\\_id=&itm\\_id=&conn\\_path=MT\\_ZTITLE](https://kosis.kr/statHtml/statHtml.do?orgId=101&tblId=DT_1PA2016&vw_cd=MT_ZTITLE&list_id=A11_2015_1_001_007&seqNo=&lang_mode=ko&language=kor&obj_var_id=&itm_id=&conn_path=MT_ZTITLE)

데이터 제공자(제공기관): KOSIS (국가통계포털)

발행연도: 2020 년