

한국 기회불평등 데이터베이스
코드북 및 사용자 가이드
Korea Inequality of Opportunity Database
Codebook and User Guide

v.20250911

KIOD 연구진
2025 년 9 월 11 일

연세대학교 클리오사회발전연구소
연세대학교 사회학과

업데이트 내역

2024.4.5. 파일럿 버전 공개.

2025.9.11. 2 년차 v.20250911 버전 업데이트. 가구통행실태조사 2010 년 데이터 바탕으로 학교 소재지 분리 및 직장 소재지 분리 지수 추가.

Codebook for Korea Inequality Opportunity Database (KIOD):

이 코드북은 한국 기회불평등 데이터베이스(KIOD)의 시군구 단위 집계 데이터의 주요 기회 불평등 지표들에 대한 소개 및 추정 과정에 대해 설명한다. 한국의 기회 불평등 현황 분석을 위해 본 연구에서는 대졸자직업이동경로조사(GOMS), 센서스 마이크로데이터, 가구통행실태조사의 3가지 데이터를 활용하였다. 이들 데이터는 시군구 단위의 지역 정보를 제공하고 있으며, 서로 다른 목적으로 수행되어 왔기 때문에 자료마다 포함하고 있는 변수들의 정보에서 조금씩 차이가 있다. 각 자료마다 다른 자료에서 포함하지 않는 변수를 활용하여 개별 자료의 한계를 보완하고, 지역 단위에서 나타나는 불평등의 현황을 가능한 한 종합적으로 파악하는 데 적절하게 활용될 수 있다. GOMS 자료의 경우, 대졸 청년들을 대상으로 한 고등교육 및 노동시장 이행에서의 기회 불평등의 분석에 활용되었으며, 센서스 마이크로데이터와 가구통행실태조사는 사회경제적 지위와 가구유형에 따른 거주지 분리 현황을 파악하는 데 활용되었다. 각 자료에서 제공하는 지역 정보는 모두 통계청에서 제공하는 시군구 단위로 통일하였다.

데이터 출처 1: 대졸자직업이동경로조사 (GOMS)

1. 자료소개

대졸자직업이동경로조사 (Graduate Occupational Mobility Survey, 이하 GOMS)는 한국고용정보원 주관으로 매년 18,000여 명의 4년제, 교육대학 및 전문대학 졸업자들을 대상으로 수행되고 있는 횡단면 표본 조사이다. 조사 항목은 대학 진학 이전 응답자의 가족 배경, 그리고 진학한 대학의 특성 및 대학 졸업 이후 노동시장 진입에서의 성과 등을 포괄한다. GOMS 자료가 갖는 강점은 표본의 규모가 크고, 대졸자의 청소년기 성장지역에 대한 시군구 단위 정보를 제공하고 있어, 지역 단위로 가족 배경에 따른 자녀의 학력 및 노동시장 성과에서의 격차를 분석하기에 용이하다는 데에 있다. 본 연구에서는 GOMS 자료와, 별도의 신청 절차를 통해 제공받을 수 있는 접근제한 자료인 응답자의 출신 대학명 자료를 연계하고, 2008년부터 2019년까지 총 12년 동안의 자료를 통합하여 분석에 활용하였다. GOMS 조사는 졸업한 지 1년이 지난 대졸자들을 대상으로 이루어진다. 이를테면 2008년 대졸자는 2009년에 조사 대상이 되며, 2019년 대졸자는 2020년에 조사 대상이 된다.

2. 변수의 측정

GOMS 자료에서 응답자의 지역 정보는 출생 지역, 출신 고교 소재지, 출신 대학 소

재지 등이 제공되고 있다. 이 가운데 출신 고교 소재지에서 시군구 단위의 지역 정보를 제공하며, 이를 활용하여 청소년기 성장지로 설정하고 이들 지역 단위에서 나타나는 기회의 불평등을 추정하는 접근을 택하였다. 단 일부 시군구 지역에서 분석에 활용 가능할 정도의 충분한 표본 수가 확보되지 않는 사례들이 있는데, 이런 경우에는 인근 지역들과 병합하여 지역 표본을 구성하였다. 여기서 분석에 활용한 유효 표본은 일반계고 졸업자에 국한하였다. GOMS 자료에서는 총 233개의 시군구 분류를 제공하고 있으며, 표본이 부족한 일부 시군구 표본을 병합하여 재구성한 결과 최종적으로 167개의 시군구 표본을 구성하였다.

기회의 불평등을 측정 및 평가하는 지표는 출신 대학의 특성, 전공, 그리고 대학 졸업 후 노동시장 성과 등이 있다. GOMS 자료와 출신 대학명 자료를 연계하면 소위 명문대 진학 여부를 기회의 불평등을 나타내는 지표로 구성할 수 있다. 대학 서열의 구성은 2000년대 이후의 대학입시 배치표 및 대학 유형과 소재지를 기준으로 이루어졌으며, 개별 대학에 대한 직접적인 정보가 유출되지 않도록 복수의 대학들을 하나의 집단으로 묶어 처리하였다. 배치표를 기준으로 한 대학 서열 변수는 이항 변수로 구성되었는데, (1) 배치표상 최상위 20위권 대학 진학 여부, (2) 최상위 대학 및 지역 거점국립대 진학 여부, (3) 최상위 대학, 지역거점국립대 및 교육/사범대 진학 여부, (4) 최상위, 지역거점국립대, 교육/사범대 및 수도권 중상위권 대학 진학 여부로 총 4개 유형의 변수를 구성하고 여기에 서울 소재 4년제 대학 여부를 추가로 구성하여 총 5개의 변수를 구성하였다. 대학명을 바탕으로 한 중상위권 이상 대학 진학 여부에 대한 변수 구성은 아래 제시된 표에 요약되어 있다. 전공의 경우 상대적으로 더 많은 소득이 보장되는 전공계열과 그렇지 않은 전공계열의 선택 여부가 기회의 불평등을 나타내는 지표가 될 수 있다. 이 연구에서는 의약학 전공과 공학계열 전공을 선택했는지 여부로 전공을 측정하였다.

대학군 세부 정보(구성 대학명)

교육 기회 변수 (대학 서열)	대학 정보
(1) 최상위 대학 (20위권)	서울대, 연세대, 고려대, 서강대, 성균관대, 포항공대, 한국과학기술원, 한양대, 경희대, 한국외대, 중앙대, 서울시립대, 이화여대, 건국대, 동국대, 홍익대, 숙명여대, 인하대, 아주대, 가톨릭대
(2) (1) + 지역 거점국립대	강원대, 충남대, 충북대, 전북대, 전남대, 경북대, 경상국립대, 부산대, 제주대, 울산과학기술원, 광주과학기술원, 대구경북과학기술원
(3) (2) + 교육대/사범대	한국교원대, 경인교대, 공주교대, 대구교대, 부산교대, 서울교대, 전주교대, 진주교대, 청주교대, 춘천교대, 광주교대, 제주교대

(4) (3) + 수도권 중상위권 대학	세종대, 국민대, 인천대, 공주대, 단국대, 성신여대, 숭실대, 상명대, 명지대, 가천대, 한성대, 한신대, 경기대, 서울여대, 동덕여대, 광운대, 삼육대, 덕성여대
-----------------------	--

대학 졸업 이후 노동시장 성과는 취업한 일자리의 특성과 근로소득으로 측정하였다. 취업한 일자리의 경우, 일반적으로 좋은 직장의 요건으로 여겨지는 대기업 정규직 및 공기업 취업 여부로 측정하였다. 여기서 대기업 여부는 사업장 근로자 수가 300인 이상인 경우를 대기업으로 정의하였으며, 정규직 여부는 별도의 근로기간 계약을 하지 않은 상용직 근로자를 정규직으로 정의하였다. 근로소득의 경우, 조사 당시 취업하여 수입활동을 하는 응답자의 월평균 근로소득 정보를 바탕으로 구성하였다. 조사 자료가 다년간의 횡단면 자료를 통합한 것이기 때문에, 연도에 따른 물가변화를 고려하여 소비자물가지수 (consumer price index, CPI) 조정을 거친 후 근로소득 백분위 변수를 구성하여 분석에 활용하였다.

그 외에 통계적 추정에 이용되지는 않았으나, 지역 단위에서 기회 불평등을 나타낼 수 있는 지표로서 일련의 변수들을 선별하여 지역 단위에서의 기술통계를 구하였다. 해당 변수들로는 대학 재학 당시 근로 경험 여부, 어학연수 여부, 공무원 시험 응시여부, 영어점수 취득 여부, 자격증 취득 여부, 재수 여부가 있다.

기회 불평등의 추정은 가족 배경을 비롯한 귀속적 특성의 차이가 교육 및 노동시장 성과의 차이와 얼마나 관련되어 있는가와 관련되어 있다. 가족 배경은 크게 두 가지 변수로 구성하였는데, 하나는 대학 입학 당시 부모의 소득이며, 다른 하나는 부모의 학력이다. GOMS 자료에서 부모의 소득은 아래와 같은 범주형 항목으로 조사되었는데, 이를 연속형 변수로 변환하여 분석에 이용하였다. GOMS 자료에서는 응답자의 대학 입학연도에 대한 정보를 제공하는데, 이를 바탕으로 하여, 각 응답 범주의 중간값을 할당한 후 입학연도 당시의 소비자물가지수 조정을 거쳐 연속형 소득변수로 구성한 뒤 백분위를 구하였다. 예를 들어 대학입학 당시 부모님의 소득이 200만원 이상 299만원 미만이고 대학에 입학한 시기가 2008년인 응답 사례에는 아래와 같이 250만원의 소득 중간값을 할당하고 2008년 당시의 물가지수로 조정하여 소득이 책정된다.

GOMS 자료의 부모소득 설문 문항

	대학입학 당시 부모님의 소득	소득 중간값
(1)	100 만원 미만	50 만원
(2)	100 만원 이상 200 만원 미만	150 만원
(3)	200 만원 이상 300 만원 미만	250 만원

(4)	300 만원 이상 400 만원 미만	350 만원
(5)	400 만원 이상 500 만원 미만	450 만원
(6)	500 만원 이상 700 만원 미만	600 만원
(7)	700 만원 이상 1000 만원 미만	850 만원
(8)	1,000 만원 이상	1,250 만원

한편 부모 학력의 경우, 아버지와 어머니 양친의 학력을 모두 고려하는 방식으로 변수를 구성하였는데, 양친의 학력이 모두 고졸 이하의 저학력인 경우, 양친 중 한 명이라도 전문대 이상 학력인 경우, 양친이 모두 4년제 대졸 이상의 고학력인 경우의 3가지 범주로 구분하여 분석에 활용하였다.

3. 분석 방법 및 전략

이 연구의 목적은 기회의 불평등을 지역 단위에서 파악하고 그 정보를 제공하는 것에 있다. 일차적으로는 자료를 바탕으로 구성한 일련의 변수들, 즉 부모소득과 부모학력과 같은 가족배경 변수와, 대학서열 및 전공계열 선택, 노동시장 성과 변수들의 평균을 시군구 단위에서 구하여, 기회 불평등을 기술적인(descriptive) 수준에서 살펴보았다. 그 다음으로는 각각의 시군구 표본을 대상으로 다변량 분석을 적용하여 가족배경과 성별에 따른 기회 불평등의 지역별 양상을 살펴보았다. 앞서 설명했듯이, 기회 불평등을 나타내는 지표들이 근로소득 백분위를 제외하면 모두 이분형 변수이기 때문에, 아래와 같은 일련의 로지스틱 회귀모형을 적용하였다.

$$\text{logit}(p) = \alpha_1 + \beta_1 \text{Fam_rank} + \gamma_1 \text{Female} + \delta_1 \text{Year} + \zeta_1 \text{Year}^2 \quad (1)$$

$$\text{logit}(p) = \alpha_2 + \sum_{j=1}^2 \beta_{2,j} \text{Pareduc} + \gamma_2 \text{Female} + \delta_2 \text{Year} + \zeta_2 \text{Year}^2 \quad (2)$$

$$\text{logit}(p) = \alpha_3 + \beta_3 \text{Female} + \sum_{j=1}^2 \lambda_j \text{Pareduc} + \eta \text{Fam_rank} + \delta_3 \text{Year} + \zeta_3 \text{Year}^2 \quad (3)$$

수식 (1)에서 (3)은 교육 성취 혹은 노동시장 성취 여부를 추정하기 위해 이 연구에서 적용한 일련의 이항 로지스틱 회귀모형들을 나타낸다. 수식 (1)은 부모의 소득 백분위 변화에 따른 기회의 불평등을 추정하는 모형으로서, α 는 상수항, Female 은 성별, Year 와 Year^2 는 각각 조사연도(연속형) 및 조사연도의 제곱항으로서 통제변수로 고려되고 있다. 그리고 Fam_rank 는 이 모형의 주요 독립변수인 부모의 소득 백분위이며, 회귀계수 β_1 은 소득 백분위의 단위 변화에 따른 종속변수의 로그 오즈에 대한 효과를 나타낸

다. 수식 (2)는 수식 (1)과 동일한 통제변수가 포함된 상태에서 부모소득 대신 부모의 학력 효과를 추정한 모형이다. 마지막으로 수식 (3)은 성별 효과를 추정한 모형으로서 부모 학력, 부모소득 백분위, 연령, 연령의 제곱항이 모두 통제된 상태에서 여성이 겪는 성취 기회의 (불)이익을 추정한다. 더불어 유일한 연속형 변수인 근로소득 백분위에 대해서도 로지스틱 회귀모형에서 선형 회귀모형으로 분석 방법만 달라졌을 뿐 수식 (1)에서 수식 (3)에 적용된 모형을 채택하여 추정이 이루어졌다. 모든 분석은 시군구 단위 표본들을 대상으로 이루어졌으며, GOMS 자료에서 제공하는 표본 가중치를 적용하여 전체 대졸자 인구집단을 대표할 수 있도록 하였다.

로지스틱 회귀모형의 회귀계수는 선형 회귀모형의 회귀계수처럼 기회의 불평등을 나타내는 격차로 해석할 수 없으며, 서로 다른 표본집단 간 회귀계수의 크기를 직접적으로 비교할 수 없다(최성수·이수빈 2018; Breen Karlson Holm 2018). 시군구 표본별로 독립변수의 효과를 비교하기 위해 일반적으로는 독립변수들이 취하는 값에 따른 조건부 기대확률을 구해 확률에서의 격차를 계산하는 방법이 있다. 이는 기회 불평등의 절대적 격차를 구하는 방법인데, 지역별 대학 진학 및 노동시장 성과의 평균 분포의 차이를 고려하지 못하는 한계가 있다. 예를 들어, 수도권을 비롯한 대도시 및 인근 지역에서는 상위권 대학 진학자의 비율이 상대적으로 높을 것이라 예상할 수 있으며, 반대로 지방 중소 도시에서는 상위권 대학 진학자의 비율이 낮을 것이라 예상할 수 있다. 절대적 격차는 이러한 한계 분포상에서의 차이를 제대로 고려하지 못하는 한계가 있다. 본 연구에서는 이러한 한계를 보완하기 위해, 절대적 격차 외에도 Karlson과 Jann(2023)이 제안한 한계 승산비(marginal odds ratio)를 시군구별로 구함으로써 부모의 SES와 성별에 따른 상대적 격차를 구하였다. 한계 승산비는 상대적인 격차를 추정하면서도, 절대적 격차와 마찬가지로 서로 다른 지역 표본들에서 얻은 추정치들과 비교 가능한 성격을 지니고 있다는 점에서, 지역 단위에서 나타나는 기회 불평등을 파악하기에 적합하다고 할 수 있다.

데이터 출처 2: 인구주택총조사 (census) 마이크로데이터

1. 자료소개

센서스 자료는 통계청의 주관 하에 5년 주기로 이루어지는 국가기본통계조사로서, 인구의 규모와 분포, 주택의 특성에 관한 정보들을 수집한다. 본 연구에서는 2000년부터 2020년도까지 5년 단위로 수집된 센서스 인가용 마이크로데이터를 바탕으로 지역별 사회경제적 지위와 가구 유형에 따른 거주지 분리현상을 파악하기 위해 활용하였다. 데이터의 분석은 만 18세 이하의 미성년자 자녀가 있는 가구만을 대상으로 이루어졌으며, 한 부모 가정도 포함되었다.

2. 변수 측정 및 조작화

센서스 마이크로데이터에서는 부모의 학력과 직업, 아파트 거주 여부, 주택 거주 여부에 관한 정보를 제공한다. 부모의 SES는 학력과 직업 정보를 조합하여 구성하였는데, 지역별로 양친이 모두 대졸 이상 고학력이면서 부모 가운데 한 명이라도 전문사무직에 종사하는 비율과, 양친이 모두 고졸 이하의 저학력이면서 전문사무직 종사자가 아닌 경우의 비율을 구하였다. 여기서 주의해야 할 부분은 연도별로 센서스 데이터의 학력 구분과 직업코드에 다소 차이가 있다는 점이다. 학력의 경우, 2005년 데이터는 전문대 졸업자와 4년제 대졸자의 구분이 없고 전문대 졸업자가 모두 대졸자로 포함되어 있어, 다른 연도에 비해 고학력 부모의 비율이 과대 추산되었을 가능성이 있다. 그리고 직업의 경우에도 2000년도와 2005년도 데이터에서 제공하는 직업분류코드가 2010년부터 2020년도 데이터에서 제공하는 직업분류와 다소 차이가 있어, 결과 해석에 있어 주의를 요한다.

한편, 거주유형은 일차적으로 지역별 아파트 거주 비율과 주택 거주 비율을 구하였으며, 시도/시군구 상이성지수(index of dissimilarity)와 노출지수(exposure index)를 구성하였다. 여기서 상이성지수(D)는 거주지역에서 서로 다른 SES를 가진 집단이 얼마나 균질적 혹은 이질적으로 분포해 있는지를 평가하는 데 활용되며, 최소값 0과 최대값 1 사이의 값을 취한다. 만약 상이성지수가 0이라면 그 지역에서 상이한 SES를 가진 집단들이 완전히 균질적으로 분포하고 있음을 의미하며, 1에 가까워질수록 SES에 따른 거주지 분리 정도가 큼을 뜻한다.

노출지수는 상호작용 지수라고도 불리기도 하는데, 특정 SES에 속한 집단이 거주지역 내에서 다른 SES 집단에 속한 사람과 상호작용할 확률을 나타낸다. 상이성지수와 마찬가지로 노출지수도 0에서 1까지의 값을 갖지만 상이성지수와는 달리 0에 가까울수록 거주지 분리가 심하고 1에 가까울수록 상호작용 수준이 높음을 의미한다. 본 연구에서는 상이성지수와 노출지수를 2가지 유형으로 구성하였는데, 시군구 내 아파트/주택 간 가구 상이성지수, 시군구 내 아파트/주택 간 가구 노출지수와 시군구 간 가구 상이성지수와 시군구 간 가구 노출지수를 구하였다.

데이터 출처 3: 가구통행실태조사

1. 자료소개

가구통행실태조사는 국토교통부 주관으로 4~5년 주기로 실시되며, 인구구조 및 사회경제적 여건의 변화 및 교통체계 변화로 인한 국민통행행태의 변화 현황을 파악하는 것에 목적을 두고 있다. 본 연구에서는 2006년과 2010년 실태조사 자료를 활용하여 사회경제적 지위 및 가구유형에 따른 거주지 분리현황 파악에 이용하였다. 초기에 수집된

2006년 조사 데이터는 수도권 중심의 교통실태 현황 파악에 목적을 두고 있었기 때문에, 수도권(서울, 경기, 인천)에 한해서만 지역코드를 제공하고 있으나, 이후 시행된 2010년과 2016년 조사 데이터는 전국 단위의 시군구 지역코드를 제공하고 있다. 그리고 16년도 가구통행조사 데이터는 이전 조사와는 달리 미취학 아동 가구를 포함하고 있지 않다.

2. 변수 측정 및 조작화

가구통행조사에서는 가구 유형과 아파트 거주 여부 외에도 센서스 데이터에서는 제공하지 않는 가구소득 정보를 제공한다. 가구소득은 아래 표와 같이 전체 월 가구 총소득을 범주형으로 측정하고 있는데, Von Hippel 등(2016)이 제안한 방법을 활용하여 시군구 단위에서의 소득 불평등을 추정하였다. 소득 불평등을 나타내는 지표로는 평균소득(mean), 지니 계수(Gini index), 타일 지수(Theil index) 등의 불평등 지수 등이 측정되었다.

가구통행실태조사의 가구소득 문항

	월 전체 가구소득	최소값	최대값
(1)	100 만원 미만	-	100 만원
(2)	100 만원 이상 200 만원 미만	100 만원	199 만원
(3)	200 만원 이상 300 만원 미만	200 만원	299 만원
(4)	300 만원 이상 500 만원 미만	300 만원	499 만원
(5)	500 만원 이상 1000 만원 미만	500 만원	999 만원
(6)	1,000 만원 이상	1,000 만원	-

가구통행실태조사에서도 센서스 데이터와 마찬가지로 지역 단위의 상이성지수와 노출지수를 구하였다. 가구유형 분리 분석을 제외한 나머지 지수 구성은 만 19세 미만의 미성년 자녀가 있는 가구만을 대상으로 분석하였다. 지표는 크게 가구유형 노출지수, 가구소득 상이성지수, 가구소득 노출지수를 구하였다. 가구소득 노출지수의 경우 저소득가구가 고소득가구에 얼마나 노출되는지 보여주는 것과 고소득가구가 저소득가구에 얼마나 노출되는지를 보여주는 것을 각각 측정했다. 또한 상이성지수(D)와 달리 노출지수(X)는 지역별 집단(고소득, 저소득 집단)의 비율에 따라 민감하게 달라지므로(예를 들어, 고소득 가구가 많은 시군구 지역의 경우 저소득층의 고소득층에 대한 노출지수가 높아짐), 이렇게 구성적 차이로 인한 부분을 조정해주는 정규화노출지수(N)도 측정했다.

가구유형의 경우, 2006년도 조사와 2010년도 조사 데이터는 1인 가구, 무자녀 부부 가구, 미취학자녀 가구, 미성년 취학 자녀 가구 및 기타 가구들로 구성되어 있어 이를 바탕으로 가구유형 노출지수를 구하였으며, 2016년도 데이터에는 미취학 아동 가구의 미

포함으로 인해 해당 유형을 제외한 나머지 가구 유형들에 기반한 노출지수를 구하였다. 그리고 가구소득 상이성지수와 노출지수는 모두 시군구 간 지수와 시군구 내 지수 각각을 따로 구하였다.

2025년 2년차 데이터에서는 2010년 가구통행실태조사에서 가용한 이동 소재지 정보를 활용해 활동공간 소재지의 가구소득 분리 지표들을 측정, 추가 제공한다. 가구원들 중 학생(초등학생, 중학생, 고등학생, 대학생)의 경우 통학하는 읍/면/동 학교소재지의 가구소득에 따른 분리지수들을, 직장인들은 통근하는 읍/면/동 직장소재지의 가구소득에 따른 분리지수들을 측정했다. 이 활동공간 소재지 정보의 경우 서울, 경기, 부산, 대구, 인천, 울산, 대전 등 주요 대도시 거주가구만을 대상으로 수집이 되었기 때문에 155개의 시/군/구 지역의 분리지수만 측정이 되었다. 학교 및 직장소재지 분리는 상이성지수와 노출지수로 측정했다. 학교소재지의 경우 모든 학생들을 대상으로 측정하기도 했지만, 기본적으로 초등학교, 중학교, 고등학교, 대학교 등 학교급을 기준으로 나눠서 측정했다.

아래 제시된 표는 상술한 데이터 출처 및 측정 방법을 적용함으로써 구성한 시군구 집계 데이터가 포함하는 주요 변수명과 이들의 특성에 대한 설명을 요약하여 정리한 것이다.

KIOD 기회불평등 데이터베이스 변수 설명

변수명	설명
데이터 출처: 대졸자직업이동경로조사	
region	시군구
geocodenum	지역코드번호 (5자리): 통계청 제공 지역코드 앞의 2자리는 광역시도를, 뒤의 3자리는 시군구 코드를 나타냄
lowedu	지역별 대졸자 가운데 양친 모두가 고졸 이하 저학력인 비율
hiedu	지역별 대졸자 가운데 양친 모두가 대졸 이상 고학력인 비율
faminc_rank	지역별 대학입학 당시 부모소득 백분위 평균
tier1	지역별 대졸자 가운데 최상위 대학 진학률 - 서울대, 연세대, 고려대, 서강대, 성균관대, 한국과학기술원(KAIST), 포항공대 (본교 캠퍼스)
tier2	지역별 대졸자 가운데 tier1 + 상위권 대학 진학률 - 한양대, 경희대, 중앙대, 서울시립대, 한국외대, 이화여대, 건국대, 동국대, 홍익대,

	숙명여대, 인하대, 아주대, 가톨릭대 (본교 캠퍼스)
tier3	지역별 대졸자 가운데 tier2 + 거점국립대 진학률 - 강원대, 충남대, 충북대, 전남대, 경북대, 경상대, 부산대, 제주대, 울산과학기술원 (UNIST), 광주과학기술원(GIST), 대구경북과학기술원(DGIST)
tier4	지역별 대졸자 가운데 tier3 + 교대/사범대 진학률 - 한국교원대, 경인교대, 공주교대, 대구교대, 부산교대, 서울교대, 전주교대, 진주교대, 청주교대, 춘천교대, 광주교대, 제주교대
tier5	지역별 대졸자 가운데 tier4 + 수도권 중상위대 진학률 - 세종대, 국민대, 인천대, 공주대, 단국대, 성신여대, 송실대, 상명대, 명지대, 가천대, 한성대, 한신대, 경기대, 서울여대, 동덕여대, 광운대, 삼육대, 덕성여대
seoul4y	지역별 대졸자 가운데 서울 소재 4년제 대학 진학률
medstem	지역별 의약학/공학 전공 선택률
goodjob	지역별 대졸 취업자 가운데 대기업 (사업체 종사자 300인 이상) 정규직 & 공기업 취업자 비율
incom_rank	지역별 취업 대졸자의 근로소득 백분위 평균
workexp	지역별 대학 재학 당시 근로 경험이 있다고 응답한 비율
langabroad	지역별 대졸자 가운데 어학연수 경험자 비율
exam	지역별 대졸자 가운데 공무원시험 응시자 비율
license	지역별 대졸자 가운데 자격증 취득자 비율
repeat	지역별 대졸자 가운데 재수자 비율
englishscore	지역별 영어성적 (TOEIC 등) 취득비율
<i>[outcome]_[predictor]_[estimator]</i>	지역별 기회불평등 추정량 - <i>outcome</i> : 대학 특성 및 노동시장 성과 <ul style="list-style-type: none"> ■ tier2 ■ tier3 ■ tier4 ■ tier5 ■ seoul4y ■ medstem ■ goodjob

	<ul style="list-style-type: none"> ■ incom (= incom_rank): 근로소득 백분위 - <i>predictor</i>: 가족 배경 및 젠더 <ul style="list-style-type: none"> ■ pct: 부모소득 백분위 ■ pedu: 부모 학력 ■ gender: 성별 (남성 대비 여성의 성취 격차) - <i>estimator</i>: 추정량의 유형 및 특성 <ul style="list-style-type: none"> ■ gap: 절대적 격차 ■ mor: 한계승산비 (marginal odds ratio), 지역 간 비교 가능한 상대적 격차 추정량
데이터 출처: 인구주택총조사 마이크로데이터	비교: <i>year</i> = 2000, 2005, 2010, 2015, 2020
<i>parproper_{year}</i>	지역별 대졸 이상 전문사무직 가구비율 (양친)
<i>parnonproper_{year}</i>	지역별 고졸 이하 비전문사무직 가구비율 (양친)
<i>aptp_{year}</i>	지역별 아파트에 거주하는 가구 비율
<i>hspr_{year}</i>	지역별 주택에 거주하는 가구 비율
<i>d_{par}_{hs_{year}}</i>	시군구 내 아파트와 주택 간 고학력전문사무직 가구와 저학력비전문직 가구의 상이성지수 (dissimilarity, 0~1 사이)
<i>x_{par}_{hs_{year}}</i>	시군구 내 아파트와 주택 간 대졸전문사무직 비율의 노출지수 (exposure, 0~1 사이)
<i>n_{par}_{hs_{year}}</i>	시군구 내 아파트와 주택 간 고학력전문사무직 비율의 정규화 노출지수 (normalized exposure)
<i>d_{par}_{sid_{year}}</i>	광역시도별 시군구 간 부모가 고학력전문직인 가구와 저학력비전문직 가구의 상이성지수
<i>x_{par}_{hs_{low_{year}}}</i>	시군구 내 아파트와 주택 간 저학력비전문직 가구의 고학력전문직 가구에 대한 노출지수
<i>x_{par}_{sido_{low_{year}}}</i>	광역시도별 시군구 간 저학력비전문직 가구의 고학력전문직 가구에 대한 노출지수
데이터 출처: 가구통행실태조사	비교: <i>year</i> = 2006, 2010
<i>p_{hs}_{sigungu_{year}}</i>	시군구 내 (within) 읍면동 간 가구유형 노출지수 (normalized)
<i>p_{hs}_{sido_{year}}</i>	시도 내 시군구 간 (between) 가구유형 노출지수 (normalized)
<i>hspl_{year}</i>	시군구별 가구 수
<i>mean_{year}</i>	지역별 가구소득 평균
<i>gini_{year}</i>	지역별 가구소득의 지니계수 (gini coefficient)

<i>theilyear</i>	지역별 가구소득의 타일지수 (Theil index)
<i>lowincpro2year</i>	시군구 저소득 가구 비율 (월 200만원 미만)
<i>highincpro2year</i>	시군구 고소득 가구 비율 (월 500만원 이상)
<i>d_inc_sigunguyear</i>	시군구 내 읍면동 간 가구소득 상이성지수
<i>x_inc_sigungu_lowyear</i>	시군구 내 읍면동 간 저소득층 가구의 고소득층 가구에 대한 노출지수
<i>n_inc_sigungu_lowyear</i>	시군구 내 읍면동 간 저소득층 가구의 고소득층 가구에 대한 정규화노출지수
<i>x_inc_sigungu_highyear</i>	시군구 내 읍면동 간 고소득층 가구의 저소득층 가구에 대한 노출지수
<i>n_inc_sigungu_highyear</i>	시군구 내 읍면동 간 고소득층 가구의 저소득층 가구에 대한 정규화노출지수
<i>d_inc_sidoyear</i>	시도 내 시군구 간 가구소득 상이성지수
<i>x_inc_sido_lowyear</i>	시도 내 시군구 간 저소득층 가구의 고소득층 가구에 대한 노출지수
<i>n_inc_sido_lowyear</i>	시도 내 시군구 간 저소득층 가구의 고소득층 가구에 대한 정규화노출지수
<i>x_inc_sido_highyear</i>	시도 내 시군구 간 고소득층 가구의 저소득층 가구에 대한 노출지수
<i>n_inc_sido_highyear</i>	시도 내 시군구 간 고소득층 가구의 저소득층 가구에 대한 정규화노출지수
<i>d_inc_sigungu_sch2010</i>	시군구 내 학교 소재 읍면동 간 가구소득 상이성 지수, 2010년
<i>x_inc_sigungu_low_sch2010</i>	시군구 내 학교 소재 읍면동 간 저소득층 가구의 고소득층 가구에 대한 노출지수, 2010년
<i>n_inc_sigungu_high_sch2010</i>	시군구 내 학교 소재 읍면동 간 저소득층 가구의 고소득층 가구에 대한 정규화노출지수, 2010년
<i>d_inc_sigungu_es2010</i>	시군구 내 초등학교 소재 읍면동 간 가구소득 상이성지수, 2010년
<i>x_inc_sigungu_low_es2010</i>	시군구 내 초등학교 소재 읍면동 간 저소득층 가구의 고소득층 가구에 대한 노출지수, 2010년
<i>n_inc_sigungu_high_es2010</i>	시군구 내 초등학교 소재 읍면동 간 저소득층 가구의 고소득층 가구에 대한 정규화노출지수, 2010년
<i>d_inc_sigungu_ms2010</i>	시군구 내 중학교 소재 읍면동 간 가구소득 상이성지수, 2010년
<i>x_inc_sigungu_low_ms2010</i>	시군구 내 중학교 소재 읍면동 간 저소득층 가구의 고소득층 가구에 대한 노출지수, 2010년

n_inc_sigungu_high_ms2010	시군구 내 중학교 소재 읍면동 간 저소득층 가구의 고소득층 가구에 대한 정규화노출지수, 2010년
d_inc_sigungu_hs2010	시군구 내 고등학교 소재 읍면동 간 가구소득 상이성지수, 2010년
x_inc_sigungu_low_hs2010	시군구 내 고등학교 소재 읍면동 간 저소득층 가구의 고소득층 가구에 대한 노출지수, 2010년
n_inc_sigungu_high_hs2010	시군구 내 고등학교 소재 읍면동 간 저소득층 가구의 고소득층 가구에 대한 정규화노출지수, 2010년
d_inc_sigungu_cl2010	시군구 내 대학교 소재 읍면동 간 가구소득 상이성지수, 2010년
x_inc_sigungu_low_cl2010	시군구 내 대학교 소재 읍면동 간 저소득층 가구의 고소득층 가구에 대한 노출지수, 2010년
n_inc_sigungu_high_cl2010	시군구 내 대학교 소재 읍면동 간 저소득층 가구의 고소득층 가구에 대한 정규화노출지수, 2010년
d_inc_sigungu_com2010	시군구 내 직장 소재 읍면동 간 가구소득 상이성지수, 2010년
x_inc_sigungu_low_com2010	시군구 내 직장 소재 읍면동 간 저소득층 가구의 고소득층 가구에 대한 노출지수, 2010년
n_inc_sigungu_high_com2010	시군구 내 직장 소재 읍면동 간 저소득층 가구의 고소득층 가구에 대한 정규화노출지수, 2010년

참고문헌

최성수·이수빈. 2018. "한국에서 교육 기회는 점점 더 불평등해져 왔는가?: 부모 학력에 따른 자녀 최종학력 격차의 출생 코호트 추세." *한국사회학* 52(4): 77-113.

Breen, Richard, Kristian Bernt Karlson, and Anders Holm. 2018. "Interpreting and Understanding Logits, Probits, and Other Nonlinear Probability Models." *Annual Review of Sociology* 44: 39-54

Karlson, Kristian Bernt, and Ben Jann. 2023. "Marginal Odds Ratios: What They Are, How to Compute Them, and Why Sociologists Might Want to Use Them." *Sociological Science* 10: 332-347.

Von Hippel, Paul T., Samuel V. Scarpino, and Igor Holas. 2016. "Robust Estimation of Inequality from Binned Incomes." *Sociological Methodology* 46(1): 212-251.