

한국의 사회경제적 집단 간 분리:
한국 기회불평등 데이터베이스를 이용한 탐색적 분석
온라인 부록

A: KIOD 표본의 구성

1. 행정구역 변동에 따른 시·군·구 통합

KIOD가 사회경제적 집단 간 분리 지수를 측정하기 위해 활용하는 데이터 중 센서스의 경우 2000년부터 2020년까지 다섯 개 연도 2% 표본 데이터를 활용했다. 이 기간 동안 있었던 시·군·구 행정구역 변동 및 가구통행실태조사에서 수집된 시·군·구 지역과의 미세한 차이를 조정하여 시계열을 유지하기 위해 아래와 같이 일부 지역을 통합했다. 통합 또는 분리된 지역들의 경우 통합된 단위를 취함으로써 2000-2020년 기간 동안 모든 지역의 시계열이 일관성있게 유지되도록 했다.

<표 A1> KIOD의 시·군·구 통합 및 변경

통합 전 (원래 행정구역)	통합 후 (새로운 행정구역)	통합 유형 (통합·변경 지역 수)
인천광역시 남구	인천광역시 미추홀구	명칭 변경 (1)
경기도 수원시 장안구, 권선구, 팔달구, 영통구	경기도 수원시	구 단위 통합 (4)
경기도 성남시 수정구, 중원구, 분당구	경기도 성남시	구 단위 통합 (3)
경기도 안양시 만안구, 동안구	경기도 안양시	구 단위 통합 (2)
경기도 부천시 원미구, 소사구, 오정구	경기도 부천시	구 단위 통합 (3)
경기도 안산시 상록구, 단원구	경기도 안산시	구 단위 통합 (2)
경기도 고양시 덕양구, 일산동구, 일산서구	경기도 고양시	구 단위 통합 (3)
경기도 용인시 기흥구, 처인구, 수지구	경기도 용인시	구 단위 통합 (3)
충청북도 청주시 상당구, 흥덕구, 충청북도 청원군	충청북도 청주시	시/군 통합 (2+1=3)
충청남도 천안시 동남구, 서북구	충청남도 천안시	구 단위 통합 (2)
전라북도 전주시 완산구, 덕진구	전라북도 전주시	구 단위 통합 (2)
경상북도 포항시 남구, 북구	경상북도 포항시	구 단위 통합 (2)
경상남도 창원시 의창구, 성산구, 진해구	경상남도 창원시	구 단위 통합 (3)
경상남도 창원시 마산합포구, 마산회원구	경상남도 창원시(구마산시)	구 단위 통합 (2)
경기도 여주군	경기도 여주시	군의 시 승격 (1)
충청남도 당진군	충청남도 당진시	군의 시 승격 (1)
충청남도 연기군	세종특별자치시	시 출범/군 폐지 (1)

이렇게 통합한 후의 전체 시·군·구 지역 수는 230개이며, 전체 목록은 아래와 같다.

<표 A2> 통합후 포함된 최종 시·군·구 목록

광역자치단체	통합된 시군구 개수	통합된 시군구 목록
서울특별시	25개	종로구, 중구, 용산구, 성동구, 광진구, 동대문구, 중랑구, 성북구, 강북구, 도봉구, 노원구, 은평구, 서대문구, 마포구, 양천구, 강서구, 구로구, 금천구, 영등포구, 동작구, 관악구, 서초구, 강남구, 송파구, 강동구
부산광역시	16개	중구, 서구, 동구, 영도구, 부산진구, 동래구, 남구, 북구, 해운대구, 사하구, 금정구, 강서구, 연제구, 수영구, 사상구, 기장군
대구광역시	8개	중구, 동구, 서구, 남구, 북구, 수성구, 달서구, 달성군
인천광역시	10개	중구, 동구, 연수구, 남동구, 부평구, 계양구, 서구, 미추홀구, 강화군, 옹진군
광주광역시	5개	동구, 서구, 남구, 북구, 광산구
대전광역시	5개	동구, 중구, 서구, 유성구, 대덕구
울산광역시	5개	중구, 남구, 동구, 북구, 울주군
세종특별자치시	1개	세종특별자치시
경기도	31개	수원시, 성남시, 의정부시, 안양시, 부천시, 광명시, 평택시, 동두천시, 안산시, 고양시, 과천시, 구리시, 남양주시, 오산시, 시흥시, 군포시, 의왕시, 하남시, 용인시, 파주시, 이천시, 안성시, 김포시, 화성시, 광주시, 양주시, 포천시, 여주시, 연천군, 가평군, 양평군, 이천시, 파주시, 평택시
강원특별자치도	18개	춘천시, 원주시, 강릉시, 동해시, 태백시, 속초시, 삼척시, 홍천군, 횡성군, 영월군, 평창군, 정선군, 철원군, 화천군, 양구군, 인제군, 고성군, 양양군
충청북도	11개	충주시, 제천시, 청주시, 보은군, 옥천군, 영동군, 진천군, 괴산군, 음성군, 단양군, 증평군
충청남도	15개	천안시, 공주시, 보령시, 아산시, 서산시, 논산시, 계룡시, 당진시, 금산군, 부여군, 서천군, 청양군, 홍성군, 예산군, 태안군
전라북도	14개	전주시, 군산시, 익산시, 정읍시, 남원시, 김제시, 완주군, 진안군, 무주군, 장수군, 임실군, 순창군, 고창군, 부안군
전라남도	22개	목포시, 여주시, 순천시, 나주시, 광양시, 담양군, 곡성군, 구례군, 고흥군, 보성군, 화순군, 장흥군, 강진군, 해남군, 영암군, 무안군, 함평군, 영광군, 장성군, 완도군, 진도군, 신안군
경상북도	23개	포항시, 경주시, 김천시, 안동시, 구미시, 영주시, 영천시, 상주시, 문경시, 경산시, 군위군, 의성군, 청송군, 영양군, 영덕군, 청도군, 고령군, 성주군, 칠곡군, 예천군, 봉화군, 울진군, 울릉군
경상남도	19개	진주시, 통영시, 사천시, 김해시, 밀양시, 거제시, 양산시, 창원시, 창원시(구마산시), 의령군, 함안군, 창녕군, 고성군, 남해군, 하동군, 산청군, 함양군, 거창군, 합천군
제주특별자치도	2개	제주시, 서귀포시
총계	230개	

*대구광역시 군위군은 현재 경상북도 군위군으로 표기되어 있음.

2. 활동공간 소재지 정보가 조사되지 않은 지역

가구통행실태 2010년 조사에서 활동공간(학교, 직장) 소재지 정보의 경우 일부 지역에서는 조사되지 않았다. 조사되지 않은 지역은 <표 A3>에 제시했다. 이들 지역을 제외한 136개 시·군·구 지역의 거주가구 가구원들의 활동공간 소재지가 분리 지수 측정에 활용되었다.

<표 A3> 활동공간 제외 시·군·구 목록

광역자치단체	전체 지역 수	추정 제외 지역 수	제외 지역
강원특별자치도	18	18	전 지역
충청북도	11	7	충주시, 제천시, 진천군, 괴산군, 음성군, 단양군, 증평군
충청남도	15	11	천안시, 보령시, 아산시, 서산시, 당진시, 부여군, 서천군, 청양군, 홍성군, 예산군, 태안군
전라북도	14	14	전 지역
전라남도	22	16	목포시, 여주시, 순천시, 광양시, 구례군, 고흥군, 보성군, 장흥군, 강진군, 해남군, 영암군, 무안군, 영광군, 완도군, 진도군, 신안군
경상북도	23	13	김천시, 안동시, 영주시, 상주시, 문경시, 의성군, 청송군, 영양군, 영덕군, 예천군, 봉화군, 울진군, 울릉군
경상남도	19	13	진주시, 통영시, 사천시, 거제시, 의령군, 함안군, 고성군, 남해군, 하동군, 산청군, 함양군, 거창군, 합천군
제주특별자치도	2	2	전 지역
총계	230	94	

B: 분리 지수 측정

분리 지수를 수식으로 표현하는 방법은 여러가지다. 여기서는 Reardon & Townsend (2018)가 제시한 방식을 따랐다.

1. 공통 기호 정의

수식의 이해를 돕기 위해 먼저 공통적으로 사용되는 기호를 다음과 같이 정의한다.

- i : 공간 단위 (예: 읍면동) ($i = 1, \dots, N$)
- N : 전체 공간 단위의 수 (예: 시군구 내 읍면동의 수)
- g_{1i} : 공간 단위 i 내에 속한 집단 1의 인구
- g_{2i} : 공간 단위 i 내에 속한 집단 2의 인구
- t_i : 공간 단위 i 내의 총 인구 ($t_i = g_{1i} + g_{2i} + \dots$)
- G_1 : 분석 대상 지역(예: 시군구) 전체의 집단 1 총 인구 ($G_1 = \sum_{i=1}^N g_{1i}$)
- G_2 : 분석 대상 지역(예: 시군구) 전체의 집단 2 총 인구 ($G_2 = \sum_{i=1}^N g_{2i}$)
- T : 분석 대상 지역(예: 시군구) 전체의 총 인구 ($T = \sum_{i=1}^N t_i$)

2. 상이성 지수 (Dissimilarity Index)

2.1. 개념

상이성 지수(D)는 분리 측정에서 가장 널리 사용되는 지수 중 하나로, 두 집단이 특정 지역 내 공간 단위들에 걸쳐 얼마나 고르게(evenly) 분포되어 있는지를 측정한다. 이 지수는 0과 1 사이의 값을 가지며, 0은 두 집단이 모든 공간 단위에서 전체 지역의 인구 구성비와 동일한 비율로 분포된 완벽한 통합 상태를 의미한다. 1은 두 집단이 동일한 공간 단위를 전혀 공유하지 않는 완벽한 분리 상태를 의미한다.

지수 값은 두 집단 중 한 집단의 구성원 중 몇 퍼센트가 다른 공간 단위로 이동해야만 지역 전체적으로 균등한 분포를 이룰 수 있는지를 나타낸다.

2.2. 산식

상이성 지수(D)는 각 공간 단위에서 집단 1이 차지하는 비율과 집단 2가 차지하는 비율 간의 차이를 절대값으로 합산하여 계산한다.

$$D = \frac{1}{2} \sum_{i=1}^N \left| \frac{g_{1i}}{G_1} - \frac{g_{2i}}{G_2} \right| \quad (1)$$

3. 노출 지수 (Exposure Index)

3.1. 개념

노출 지수(P*)는 한 집단(기준 집단)의 구성원이 다른 집단(대상 집단)의 구성원과 동일한 공간 단위 내에서 마주칠(encounter) 평균적인 확률을 측정한다. 이는 '상호작용의 잠재성'을 나타낸다.

3.2. 산식

집단 1의 집단 2에 대한 노출 지수($_{g_1}P_{g_2}^*$)는 집단 1의 구성원이 거주하는 평균적인 공간 단위에서 집단 2가 차지하는 비율의 가중평균으로 계산된다.

$$_{g_1}P_{g_2}^* = \sum_{i=1}^N \left(\frac{g_{1i}}{G_1} \right) \left(\frac{g_{2i}}{t_i} \right) \quad (2)$$

여기서 (g_{1i}/G_1) 항은 집단 1의 구성원이 i 단위에 거주할 확률(가중치)을, (g_{2i}/t_i) 항은 i 단위 내에서 임의의 사람을 만났을 때 그 사람이 집단 2일 확률을 의미한다.

4. 정규화 노출 지수 (Normalized Exposure Index)

4.1. 개념

정규화 노출 지수는 앞서 설명한 노출 지수를 특정 기준에 따라 조정한 값이다. 이 지수는 노출 지수($_{g_1}P_{g_2}^*$)를 전체 인구 대비 대상 집단(집단 2)의 비율(Q_2)을 이용해 정규화한다. 이 지수는 기준 집단(집단 1)이 무작위로 분포했을 때 기대되는 노출 수준(즉, Q_2)과 실제 노출 수준($_{g_1}P_{g_2}^*$)을 비교하는 척도로 해석할 수 있다.

4.2. 산식

먼저, 지역 전체 인구 대비 집단 2의 비율(Q_2)을 다음과 같이 정의한다.

$$Q_2 = \frac{G_2}{T} \quad (3)$$

정규화 노출 지수(N)는 이 Q_2 값과 노출 지수 $_{g_1}P_{g_2}^*$ 를 사용하여 다음과 같이 계산된다.

$$N = 1 - \frac{_{g_1}P_{g_2}^*}{Q_2} = 1 - \frac{\sum_{i=1}^N (g_{1i}/G_1)(g_{2i}/t_i)}{G_2/T} \quad (4)$$

이 산식에 따르면, 만약 집단 1의 집단 2에 대한 실제 노출($_{g_1}P_{g_2}^*$)이 무작위 기대치(Q_2)와 같다면 N 지수는 0이 된다. 반면, 실제 노출이 0에 가깝다면(즉, 고립도가 높다면) N 지수는 1에 가까워진다. 따라서 이 지수는 0(무작위 노출과 동일)에서 1(완전한 고립) 사이의 값을 가지는 '정규화된 고립(isolation)' 척도로 해석할 수 있다.

C: 분석 결과표

<표 C1> 가구 소득에 따른 거주지 분리와 지역 특성(<그림 1>)

	상이성		노출		정규화노출	
	모형 1	모형 2	모형 1	모형 2	모형 1	모형 2
대도시=1	-0.024 (0.021)	0.009 (0.023)	0.064*** (0.017)	-0.003 (0.011)	0.012 (0.017)	0.015 (0.020)
수도권=1	-0.077*** (0.020)	-0.029 (0.021)	0.127*** (0.023)	0.036** (0.011)	-0.016 (0.013)	-0.019 (0.016)
대도시=1 x 수도권=1	0.038 (0.031)	-0.028 (0.034)	-0.016 (0.036)	0.011 (0.018)	0.011 (0.025)	-0.025 (0.028)
평균 가구소득(2010년 가구통행조사)		-0.000 ⁺ (0.000)		0.001*** (0.000)		0.001** (0.000)
고학력전문직 부모 가구 비율(센서스2010)		0.005*** (0.001)		0.003*** (0.001)		0.001 (0.001)
소득불평등(지니계수, 가통 2010)		0.533* (0.262)		-0.246 ⁺ (0.145)		0.406 ⁺ (0.219)
아파트 거주 비율(센서스 2010)		-0.002*** (0.000)		0.001** (0.000)		-0.001 (0.000)
가구유형 거주지분리:고립지수(가통 2010)		0.323 (0.356)		0.232* (0.117)		0.245 (0.234)
Constant	0.436*** (0.011)	0.392** (0.128)	0.146*** (0.008)	-0.215** (0.082)	0.167*** (0.009)	-0.171 (0.111)
Observations	219	219	228	228	228	228
R ²	0.071	0.269	0.312	0.834	0.008	0.153

주: 노출 지수는 저소득층의 고소득층에 대한 노출을 의미. 괄호 안의 값은 표준편차임.

⁺ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

<표 C2> 아파트 거주 가구 비율과 상이성지수: 고정효과모형 결과(<그림 4>)

	모형1	모형2	모형3	모형4	모형5	모형6
아파트거주가구비율	0.068 (0.071)	-0.108 (0.079)	1.006*** (0.126)	1.652*** (0.256)	3.087*** (0.438)	3.319*** (0.689)
2005년	-0.051** (0.016)	-0.100*** (0.019)	-0.090*** (0.018)	-0.088*** (0.018)	-0.086*** (0.018)	-0.086*** (0.018)
2010년	-0.054** (0.019)	-0.111*** (0.022)	-0.086*** (0.021)	-0.087*** (0.020)	-0.085*** (0.020)	-0.084*** (0.020)
2015년	-0.056** (0.020)	-0.161*** (0.029)	-0.106*** (0.028)	-0.108*** (0.028)	-0.103*** (0.028)	-0.102*** (0.028)
2020년	-0.079** (0.025)	-0.219*** (0.037)	-0.127*** (0.036)	-0.132*** (0.036)	-0.118** (0.036)	-0.117** (0.036)
고학력전문직비율		0.262 (0.148)	0.450** (0.140)	0.448** (0.139)	0.414** (0.138)	0.410** (0.139)
저학력비전문직비율		-0.470*** (0.110)	-0.197 (0.106)	-0.177 (0.106)	-0.162 (0.105)	-0.161 (0.105)
아파트거주가구비율 ²			-1.247*** (0.115)	-2.928*** (0.592)	-9.868*** (1.824)	-11.654** (4.474)
아파트거주가구비율 ³				1.239** (0.428)	13.078*** (2.976)	18.274 (12.252)
아파트거주가구비율 ⁴					-6.572*** (1.635)	-12.964 (14.712)
아파트거주가구비율 ⁵						2.805 (6.417)
상수	0.339*** (0.027)	0.675*** (0.084)	0.315*** (0.086)	0.245** (0.089)	0.170 (0.090)	0.163 (0.092)
Observations	1135	1135	1135	1135	1135	1135
R ²	0.022	0.048	0.158	0.166	0.181	0.181

주: 괄호 안의 값은 표준편차임.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

<표 C3> 사회경제적수준 변수들의 상관관계 매트릭스(각주 3)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
(1) 평균가구소득(가통2010)	1								
(2) 고소득가구비율(가통2010)	0.853***	1							
(3) 가족소득백분위(GOMS)	0.676***	0.687***	1						
(4) 고학력부모비율(GOMS)	0.665***	0.760***	0.828***	1					
(5) 고학력전문직비율(센서스2000)	0.704***	0.799***	0.714***	0.859***	1				
(6) 고학력전문직비율(센서스2005)	0.717***	0.774***	0.782***	0.881***	0.894***	1			
(7) 고학력전문직비율(센서스2010)	0.717***	0.818***	0.737***	0.883***	0.891***	0.892***	1		
(8) 고학력전문직비율(센서스2015)	0.730***	0.814***	0.732***	0.876***	0.861***	0.866***	0.918***	1	
(9) 고학력전문직비율(센서스2020)	0.639***	0.743***	0.700***	0.826***	0.780***	0.820***	0.866***	0.905***	1

주: GOMS는 대졸자직업이동경로조사, 가통은 가구통행조사를 의미함.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

<표 C4> 가구 소득에 따른 학교 소재지 분리의 지역 간 차이(<그림 7>)

	초등학교	중학교	고등학교	대학교
수도권	0.095 (0.116)	0.089 (0.117)	0.215 ⁺ (0.121)	0.147 (0.127)
지방 대도시	0.269** (0.101)	0.192 ⁺ (0.109)	0.261* (0.118)	0.127 (0.131)
그외 읍면지역	0.255* (0.105)	0.253* (0.105)	0.137 (0.129)	0.039 (0.133)
Constant	0.680*** (0.095)	0.680*** (0.095)	0.560*** (0.101)	0.520*** (0.101)
Observations	135	133	137	132
R^2	0.087	0.055	0.043	0.015

주1: 생략된 준거집단은 서울.

주2: 종속변수는 거주지 분리 대비 학교 소재지 분리의 심화 여부(상이성 지수).

⁺ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

<표 C5> 지역의 사회경제적 특성과 가구 소득에 따른 학교 소재지 분리(<그림 7>)

	초등학교	중학교	고등학교	대학교
평균가구소득	-0.178*** (0.046)	-0.162** (0.054)	-0.145* (0.058)	-0.037 (0.077)
고소득자비율	-0.085+ (0.044)	-0.073 (0.059)	-0.087 (0.056)	0.044 (0.074)
고소득전문직 비율	0.085* (0.040)	0.065 (0.052)	0.024 (0.057)	-0.059 (0.072)
아파트비율	0.008 (0.042)	0.016 (0.040)	0.128** (0.046)	-0.027 (0.064)
지니계수	-0.074* (0.034)	-0.095** (0.034)	-0.102* (0.046)	-0.080 (0.090)
Constant	0.885*** (0.024)	0.848*** (0.032)	0.733*** (0.042)	0.607*** (0.060)
Observations	135	133	137	132
R^2	0.182	0.145	0.181	0.016

주: 종속변수는 거주지 분리 대비 학교 소재지 분리의 심화 여부(상이성 지수).

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

<표 C6> 지역의 사회경제적 특성과 가구 소득에 따른 직장 소재지 분리(<그림 9>)

	모형1	모형2
수도권	0.100 (0.093)	0.225 ⁺ (0.127)
지방 대도시	0.342** (0.104)	0.415** (0.135)
그외 읍면지역	0.605*** (0.091)	0.749*** (0.143)
평균가구소득		-0.079 ⁺ (0.047)
고소득자비율		-0.051 (0.064)
고소득전문직 비율		0.109 ⁺ (0.065)
아파트비율		-0.132** (0.050)
지니계수		-0.154*** (0.039)
Constant	0.120 ⁺ (0.066)	0.035 (0.113)
Observations	156	156
R ²	0.228	0.304

주1: 생략된 준거집단은 서울.

주2: 종속변수는 거주지 분리 대비 직장 소재지 분리의 심화 여부(상이성 지수).

⁺ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Reference

Reardon, S. F., and J. B. Townsend. 2018. “SEG: Stata module to compute multiple-group diversity and segregation indices.”