

「2023년 통계데이터 활용대회」 데이터분석 보고서

신청자명	소속/직위	고려대학교/학부생	성명	고범석
	휴대전화	010-8012-9253	전자우편	rhrh8429@naver.com
신청자명	소속/직위	고려대학교/학부생	성명	이준서
	휴대전화	010-9519-6784	전자우편	lake2102@gmail.com
신청자명	소속/직위	고려대학교/학부생	성명	홍유찬
	휴대전화	010-5108-3555	전자우편	remnant0709@naver.com
제출일	2023.07.21.			

제목	인구 감소에 따른 국회의원 선거의 선거구 획정 모델 제안
----	------------------------------------

접수번호	
------	--

1. 배경

□ 주제선정(소제목)

KOSIS 시나리오별 추계인구 정보에 따르면, 대한민국은 2020년 인구 수가 최고치를 기록한 뒤 자연 감소하고 있으며, 높은 출산율을 가정하더라도 2050년까지 2020년 인구 수의 6.06% 가량 감소할 것으로 예상된다. 그러나 헌법 제41조와 공직선거법 제21조에 따라 정해지는 국회의원 정수는 현행 헌법이 시행된 이후 1997년 금융위기에 의한 경우를 제외하고는 300여 석으로 유지되어 왔으므로, 개헌 혹은 법률의 개정 없이는 의석 수의 큰 변동이 없을 것이다.

한 표가 동등한 가치를 가져야 하는 민주선거에서, 각 지역구 국회의원은 최대한 비슷한 수의 국민을 대표해야 한다. 따라서 헌법재판소 역시 1995년, 2001년, 2014년 각각 선거구 간 인구비율을 각각 “4:1”, “3:1”, “2:1”로 줄여온 경향 역시 선거구 간 표의 등가성을 제고하기 위함이다. 이러한 배경 하에서 대한민국의 인구 감소는 권역별 선거구의 수, 즉 할당 의석수의 변동을 발생시킨다.

이러한 변동이 선거구 획정이라는 중대한 문제에 큰 영향을 미칠 수 있다는 점으로 미루어 보아 국회의원 선거구를 최소한의 요소를 기반으로 자동적으로 획정하는 모델의 필요성을 느껴 본 연구를 진행하였다.

□ 분석 필요성(문제점) 및 전략

1) 분석 필요성

선거구 획정은 각 정당의 이해관계가 첨예하게 대립하는 영역이므로 제20대 및 제21대 국회의원 선거 등 최근 선거에서 꾸준히 정당 간 의견 충돌이 발생하였다. 특히 제21대 국회의원 선거의 경우, 선거가 치러지기 39일 전에야 선거구 재획정안이 가결되었고, 춘천시를 여러 동으로 나누어 주변 행정구역과 임의 결합한 선거구는 ‘누더기 선거구’라는 불명예를 낳았다. 이러한 게리맨더링의 문제를 방지하기 위해 선거구 획정에 개입되는 요소는 최소화되어야 하고, 선거구 간 표의 등가성은 최대화해야 한다.

2) 전략

(1) 선거구 획정 요소 선택

선거구는 유사한 사회적 분위기에서 생활하는 사람들끼리 구획되는 것이 타당하므로 ‘지리적 요소’가 고려되어야 하기에 ‘각 지역 별 위치’ 및 ‘두 지역 간 거리’가 중요하게 다루어져야 한다. 따라서 각 지역의 위치 정보를 기반으로 두 기초자치단체 간 구역을 계산하여 광역자치단체를 기준으로 행정구역의 경계가 인접한 지역끼리만 하나의 선거구가 될 수 있도록 한다.

선거구 간 표의 등가성을 고려하기 위해서는 각 기초자치단체 별 인구 정보가 필수적이므로, 각 단체 별 인구를 토대로 각 선거구의 인구 편차가 최소화되도록 한다.

따라서 ‘지역 간 거리’와 ‘인구’만을 선거구 획정에 있어 고려되는 요소로 선택하며, 각 거리 정보 및 인구 정보를 추출하여 산정 및 파악한다.

(2) 선거구 획정 실시

각 지역의 위치 정보를 토대로 군집화를 실시하여 거리가 비슷한 지역의 군집을 생성한다. 이 군집 중 사전에 설정된 인구 제약 조건(여기서는 헌법재판소가 설정한 최대 인구 편차인 2:1)을 만족하는 군집의 집합만을 선택한다. 이 과정을 n 회 반복한다.

(3) 선거구 간 표의 등가성 계산

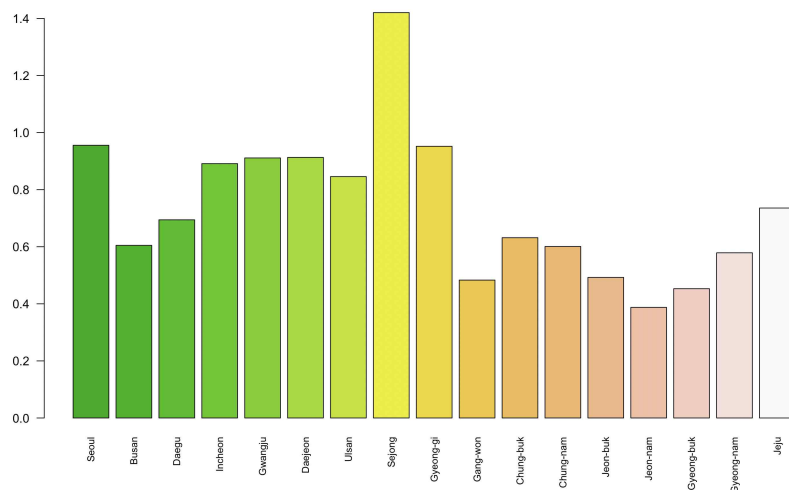
(2)의 과정을 통해 확정된 선거구가 실제로 선거구 간 표의 등가성을 강화하는 집합인지를 일정 지표를 이용하여 계산한다.

(4) 인구추계를 이용하여 선거구 모델 적용

실제로 인구 감소에 대응할 수 있는 모델인지 확인하기 위하여 인구 추계를 바탕으로 할당 의석 수와 인구를 조정하여 (2), (3)의 과정을 반복한다.

2. 데이터 분석

□ 데이터 선정(사용한 데이터 및 이유 등)



본 연구를 진행하기에 앞서, 전국을 대상으로 연구를 실시하기보다는 실제로 지역이 소멸할 위기가 심각한 지역을 선정하여 해당 지역을 대상으로 선거구 확정 모델을 고안하고자 하였다. 인구의 증가율을 토대로 지역을 선정하고자 하였으나 서울특별시, 경기도 등이 증가율에 있어 하위를 기록하였다.

이러한 지역에 대한 연구는 현재 대한민국의 인구 감소와 지역 소멸 위기를 배경으로 하는 본 연구의 취지와는 어긋난다고 판단하여 통계청에서 발표하는 시군구별 인구추계 자료가 아닌, 지방소멸위험지수를 토대로 결정된 인구감소지역을 사용하였다. 지방소멸위험지수는 0에 가까울수록 소멸 위험이 높음을 의미한다.

$$\text{지방소멸위험지수} = \frac{\text{20~39세 여성 인구}}{\text{65세 이상 인구}}$$

KOSIS 연령 및 성별 인구 참고, 2020년 인구 기준, 지방소멸위험지수가 가장 0에 가까운 ‘전라남도’를 연구 대상 지역으로 선정하였다.

□ 데이터 분석(분석 프로세스, 분석방법, 접근방법 등)

(변수 생성 및 계산: x, y좌표 이용 방법, 인구 계산 방법 등)

인구가구통계등록부(인구)로부터 전라남도의 각 시군별 인구 수 자료만을 추출한다. 소지역코드_정보로

부터 각 격자 단위별 x좌표 및 y좌표를 추출한 후 평균을 구해 전라남도 각 시군별 무게중심의 좌표를 산출한다.

(K-means 클러스터링 사용)

k-means 클러스터링은 비지도 학습의 일종으로, 여러 종류의 데이터를 k개의 군집으로 분류하는 클러스터링 기법이다. 각 데이터끼리의 유사도에 따라 군집화를 실시하는 알고리즘이다. 선거구 획정에 있어서, n개의 시군을 q개의 선거구로 구획할 때의 경우의 수는 집합의 분할 문제와 유사한 형태를 띠게 되고 이는 NP 문제이다. 또한 경우의 수 각각이 지리적 인접 조건이 항상 충족되는 것은 아니기 때문에 brute force로는 확인할 수 없고, 선별적으로 해를 찾아내는 과정이 필요하다고 판단했다. 이와 관련된 선행연구로 기존의 Hess et al 모델이 있다. 하지만 이를 이용해 선거구 구획을 실시한다고 하더라도 그 최적의 해가 항상 지리적 인접도를 충족하는 것은 아님을 확인했다.(용인시 중심 사례) 따라서 격자점 데이터를 활용해 추출한 각 시군별 좌표를 바탕으로 K-means 클러스터링을 진행해 지리적 인접도를 만족하는 선거구를 먼저 구한다. 이후 인구 제약조건을 설정함으로써 지리적 조건과 민주적인 선거를 위한 인구 조건을 충족시킬 수 있는 선거구들을 설정했다.

(클러스터 제약 식 사용)

인구가구통계등록부(인구) 데이터로부터 전국 인구수와 전라남도의 각 시군별 인구 수 자료를 계산한 결과, 전국 인구 51,829,136명, 전라남도의 2020년 추계인구는 1,788,807명으로 집계되었다. 2020년 실시된 제 21대 국회의원 선거에서 전라남도의 선거구는 전국 인구를 전라남도 인구로 나누어 $\frac{1788807}{51829136} \times 253 = 8.731926 \approx 9$ 석을 얻었다. 따라서 인구 비례 할당 의석은 실제보다 1석 적은 9석을 할당하여 연구를 진행하였다.

(인구 min max)

인구 제약조건을 위해 각 선거구 간 최소 선거구 인구와 최대 선거구 인구의 비율이 1:2를 넘지 않아야 한다는 조건 하에서 연구를 진행하였다. 각 선거구별 평균 인구에 대해 $\sqrt{2}$ 를 곱한 값이 선거구의 가능한 최대 인구 수, $\frac{1}{\sqrt{2}}$ 을 곱한 값이 선거구의 가능 최소 인구수로 책정되었다. 선거구별 평균 인구 $51829136 \div 253 = 204858.2450 \approx 204858$ 에 대해 최소 인구 1,425,000명, 최대 인구 2,850,000명으로 진행하였고, 평균 인구보다 많은 단일 시군에 대해서는 단일 시군으로 설정하여 클러스터링에선 배제하였다.

(코사인 제곱 식)

본 연구에서 선거구 집단의 실효성 평가를 위해 코사인제곱 값을 적용한다. 코사인제곱 값은 인구 비중 벡터와 의석수 비중 벡터를 통해 구한 코사인 유사도를 제곱한 값이다. 인구 비중 벡터의 요소는 전체 전남 인구 대비 해당 선거구에 포함되는 인구수의 합이고, 의석수 비중 벡터는 해당 선거구가 차지하는 의석수의 비중으로 전체 의석수에 대한 해당 선거구의 의석수 값을 의미한다.

$$\vec{v_{\text{인구 비중}}} = (a_{\text{갑}}, a_{\text{을}}, \dots, a_{\text{임}}), \vec{v_{\text{선거구 비중}}} = (b_{\text{갑}}, b_{\text{을}}, \dots, b_{\text{임}})$$

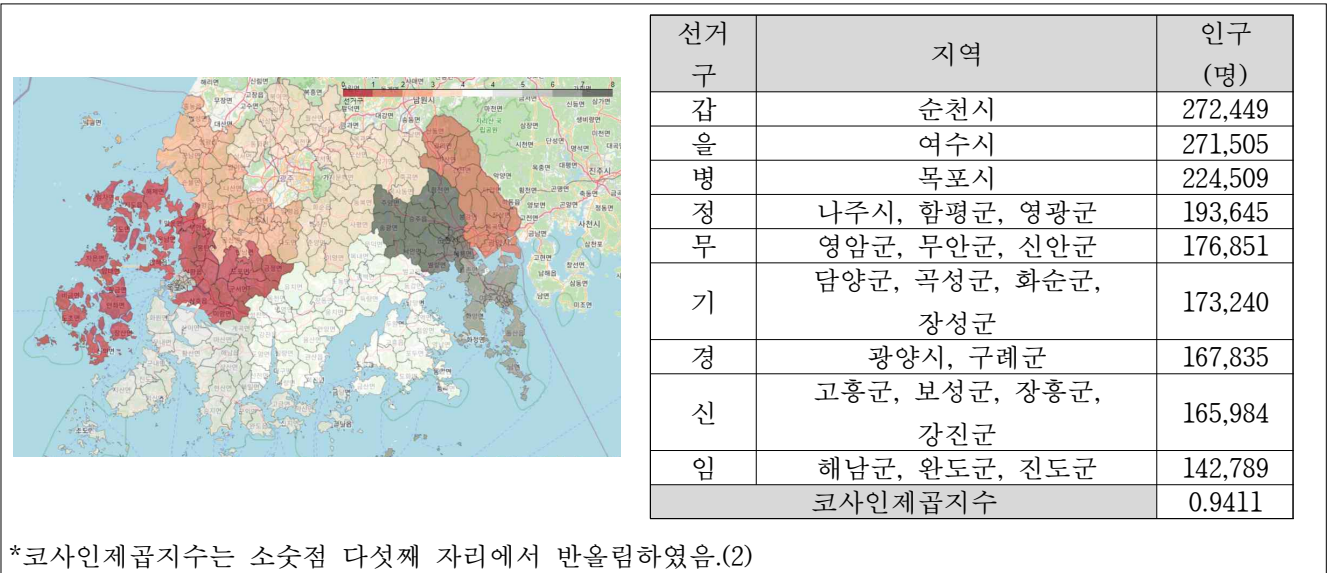
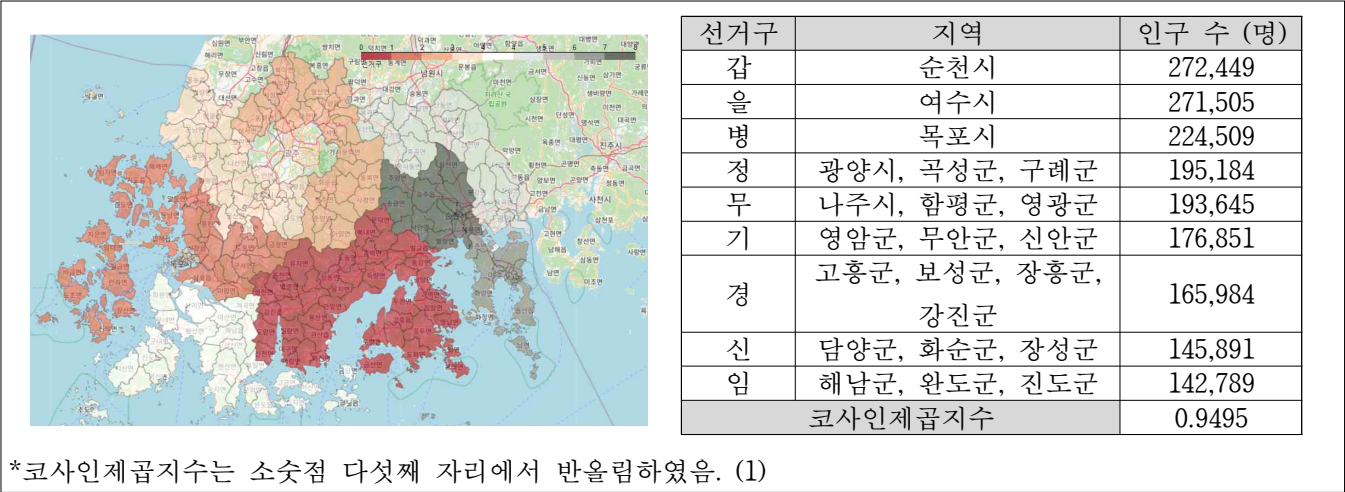
$$(a_{\text{갑}} = \frac{\text{갑 선거구 인구 수}}{\text{전라남도 인구 수}}, b_{\text{갑}} = \frac{\text{갑 선거구 의석 수}}{\text{전라남도 의석 수}} = \frac{1}{9})$$

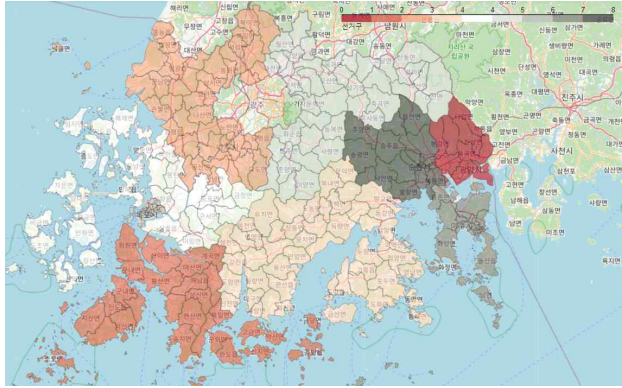
$$\cos(\theta) = \frac{\vec{v_{\text{인구 비중}}} \cdot \vec{v_{\text{선거구 비중}}}}{\|\vec{v_{\text{인구 비중}}}\| \|\vec{v_{\text{선거구 비중}}}\|}$$

전라남도 기초자치단체의 구를 기반으로 제21대 국회의원 선거에서 적용한 모델과 동일한 모델을 적용

하여 획정한 선거구는 다음과 같다.

(도출된 클러스터 시각화)



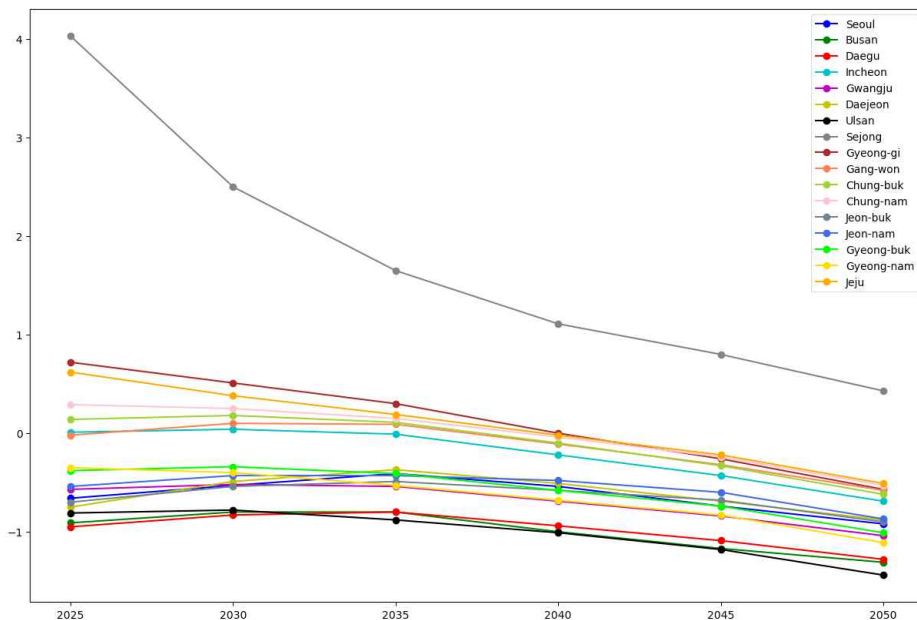


선거구	지역	인구 수 (명)
갑	순천시	272,449
을	여주시	271,505
병	목포시	224,509
정	광양시	143,928
무	나주시, 함평군, 영광군, 장성군	235,088
기	영암군, 무안군, 신안군	176,851
경	고흥군, 보성군, 장흥군, 강진군	165,984
신	담양군, 곡성군, 구례군, 화순군	155,704
임	해남군, 완도군, 진도군	142,789
코사인제곱지수		0.9526

*코사인제곱지수는 소숫점 다섯째 자리에서 반올림하였음. (3)

□ 분석 결과 및 해석

2020년 제21대 국회의원 선거를 기준으로 본 연구에서 도출한 전반적인 선거구 획정 결과나 선거구 별 코사인제곱지수 계산 결과 0.9495, 0.9411, 0.9526은 실제 21대 국회의원 전라남도 선거구의 지수값인 0.9589와 크게 다르지 않았다. 다만, 21대 국회의원 선거에서 여주시와 순천시는 가 강제로 분할되어 7 사회적 동질성에 저하가 있었으나 본 연구에서 제시한 선거구 집합은 이러한 시·군의 분할을 최소화하여 선거구의 인구 왜곡을 비슷한 수준으로 유지함과 동시에 사회적 동질성 역시 극대화하였다.

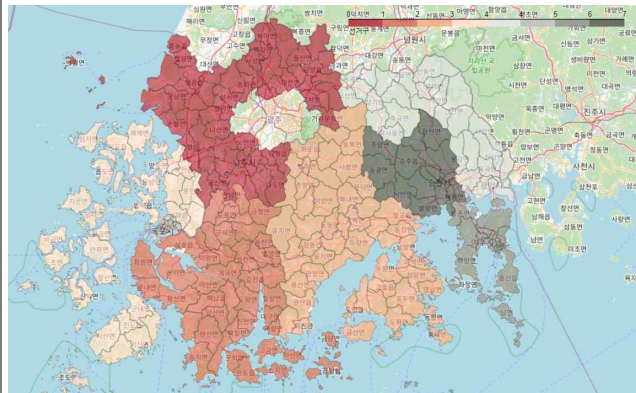


(인구 추계 그래프)

KOSIS 추계인구(시/군/구) 인용, 전라남도의 2040년 추계인구는 1,625,140명으로, 대한민국의 2040년 추계

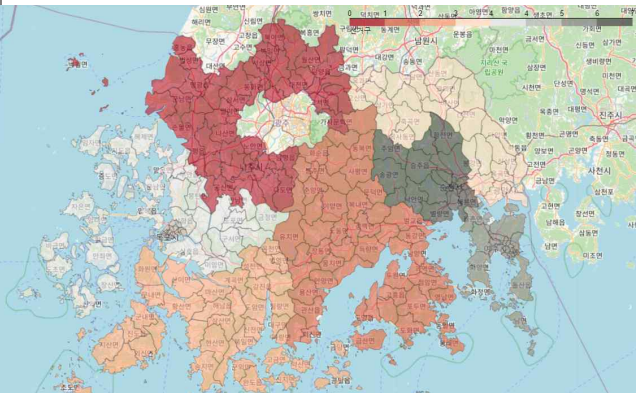
인구 50,193,281명에 대비하여 인구 비례 할당 의석은 기존 9석에서 $\frac{1625140}{50193281} \times 253 = 8.191543 \approx 8$ 석으로 줄어든다. (지역구 국회의원의 의원정수는 253석으로 유지된다고 가정한다.) 따라서 2040년 예상 선거구를 확정할 때는 인구감소를 반영하여 8개의 선거구로 확정한다. (K=8)

전라남도 기초자치단체의 추계인구를 기반으로 제20대 국회의원 선거에서 적용한 모델과 동일한 모델을 적용하여 확정된 선거구는 다음과 같다.



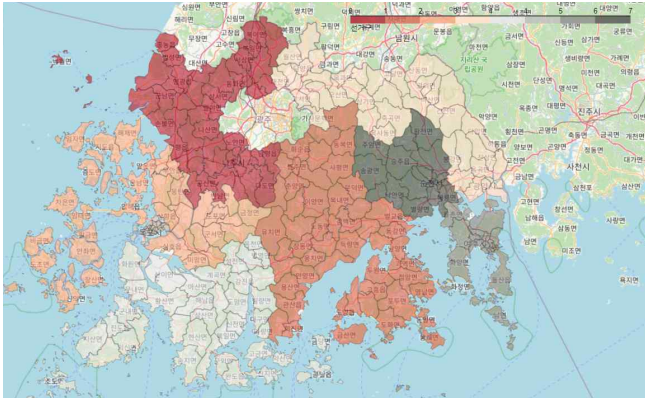
선거구	지역	인구 수 (명)
갑	순천시	249,308
을	여수시	228,815
병	목포시	180,888
정	나주시, 담양군, 함평군, 영광군, 장성군	279,855
무	광양시, 곡성군, 구례군	185,740
기	고흥군, 보성군, 화순군, 장흥군	178,521
경	강진군, 해남군, 영암군, 완도군	175,806
신	무안군, 진도군, 신안군	146,207
코사인제곱지수		0.9591

*코사인제곱지수는 소숫점 다섯째 자리에서 반올림하였음. (1)



선거구	지역	인구 수 (명)
갑	순천시	249,308
을	여수시	228,815
병	목포시	180,888
정	나주시, 담양군, 함평군, 영광군, 장성군	279,855
무	광양시, 곡성군, 구례군	185,740
기	고흥군, 보성군, 화순군, 장흥군	178,521
경	영암군, 무안군, 신안군	170,777
신	강진군, 해남군, 완도군, 진도군	151,236
코사인제곱지수		0.9598

*코사인제곱지수는 소숫점 다섯째 자리에서 반올림하였음. (2)



선거구	지역	인구 수 (명)
갑	순천시	249,308
을	여주시	228,815
병	목포시	180,888
정	나주시, 함평군, 영광군, 장성군	234,254
무	광양시, 담양군, 곡성군, 구례군	231,341
기	고흥군, 보성군, 화순군, 장흥군	178,521
경	영암군, 무안군, 신안군	170,777
신	강진군, 해남군, 완도군, 진도군	151,236
코사인제곱지수		0.9723

*코사인제곱지수는 소숫점 다섯째 자리에서 반올림하였음. (3)

□ 분석 결과 및 해석

2040년에는 전라남도의 추계 인구가 전국 추계 인구에 대비하여 할당 의석 수를 줄여야 하는 수준에 이른다. 따라서 지역구 국회의원의 의석 수, 국회의원 선거제도(준연동형 비례대표제, 소선거구제), 법정 제약(선거구 간 인구 편차 최대 2:1, 시군의 분할은 최소화, 경계가 인접한 지역끼리 한 선거구로 통합 가능) 등 다른 조건은 모두 동일하다는 전제 하에 인구 조건만을 달리하여 2040년 시행될 제26대 국회의원 선거의 전라남도 선거구를 확정하였다. 그 결과, 코사인제곱지수가 최대 1.34%p 상승한 0.9723의 결과로 선거구 집합을 도출하였다.

3. 분석 활용 전략

□ 기대효과

본 연구에서 제시한 모델은 오로지 기초자치단체 간 거리와 인구만을 확정 요소로 한다. 도출되는 여러 선거구 집합 중 가장 코사인제곱지수가 높은 선거구 집합을 선택함으로써 선거구의 인구 왜곡을 최소화할 수 있다. 따라서 매 선거마다 불거지는 게리맨더링의 논란으로부터 한층 자유로울 수 있으며, 정치적으로 중립적이고 정치지형의 왜곡을 최소화한 모델을 단 시간에 선택하여 선거 운영에의 어려움을 완화할 수 있다.

□ 방향제시

모든 경우에 대해 고려할 필요 없이 효율적으로 지리적 인접도를 만족하는 선거구들의 경우의 수를 확인하고 이에 대한 코사인 유사도 등의 다양한 평가지표를 편리하게 계산해볼 수 있다. 현재 연구의 경우 제약조건에 인구 관련 데이터만 포함되어 있지만, 이외에도 다양한 데이터 등을 고려함으로써 더욱 민주적이고 공정한 선거구 획정이 가능할 것으로 기대된다.

참고문헌