

농어촌 지역의 다양한 특성을 반영한 유형화 방법

김서영 · 정동명 · 김경미

목 차

제 1절 서 론	1
1. 농어촌 유형화 배경	1
2. 연구목적 및 필요성	2
3. 연구범위와 방법	4
4. 연구의 한계	4
제 2절 통계적 분석 방법	5
1. 요인분석	6
2. 군집분석	8
가. 계층적 군집	9
나. 분할적 방법	10
3. 군집 개수 선정 기준	12
가. CCC	12
나. pseudo F	12
제 3절 농어촌지역 유형 분석	13
1. 자료와 변수 설명	13
2. 분석절차	14
3. 분석결과	16
가. 도시와 농어촌 구분	16
나. 요인분석 결과	17
다. 군집분석 결과	19

라. 군집에 의한 유형화 결과 해석	30
제 4 절 결론 및 논의	35
< 부 록 >	39
1. K-means 결과 mapping : 4개 군집	39
2. 계층적 군집 결과 mapping : 5개 군집	40

표 목 차

<표 3- 1> 변수 유형별 변수 설명	3
<표 3- 2> 도시화율에 의한 도시와 농어촌의 구분	17
<표 3- 3> 요인특성	18
<표 3- 4> 주성분방법에 의한 요인분석 결과	19
<표 3- 5> 군집방법별 CCC와 psF	20
<표 3- 6> 군집방법별 군집분포	22
<표 3- 7> 군집1에 할당된 910개 지역들의 각 요인 구간별 분포	24
<표 3- 8> 군집2에 할당된 301개 지역들의 각 요인 구간별 분포	25
<표 3- 9> 군집3에 할당된 39개 지역들의 각 요인 구간별 분포	25
<표 3-10> 군집4에 할당된 50개 지역들의 각 요인 구간별 분포	26
<표 3-11> 군집1에 할당된 194개 지역들의 각 요인 구간별 분포	28
<표 3-12> 군집2에 할당된 430개 지역들의 각 요인 구간별 분포	28
<표 3-13> 군집3에 할당된 483개 지역들의 각 요인 구간별 분포	29
<표 3-14> 군집4에 할당된 145개 지역들의 각 요인 구간별 분포	29
<표 3-15> 군집5에 할당된 48개 지역들의 각 요인 구간별 분포	30
<표 3-16> K-means에 의한 4개 군집별 평균, 표준편차	31
<표 3-17> 계층적 방법의 5개 군집별 평균, 표준편차	32
<표 3-18> K-means 군집들에 대한 군집 특성 요약	33
<표 3-19> 계층적 군집들에 대한 군집 특성 요약	34

그림 목차

[그림 2-1] 군집방법 분류도	8
[그림 2-2] 계층적 방법의 군집화 과정	9
[그림 2-3] 비계층적 방법의 군집화 과정	11
[그림 3-1] 계층적 군집방법의 덴드로그램	21

제 1절 서론

1. 농어촌 유형화 배경

국가적 차원에서 작은 지역(localities)단위의 경제·사회적 성과가 점점 세분화 되어 나타나고 있다. 오늘날 지역(또는 영역)은 그 지역만의 잠재적 발전 경로를 형성하게 되면서 지역간에 서로 다른 특성을 지니게 되었다. 따라서 정책을 집행함에 있어 이러한 지역별 다양한 조건들이 반영되어야 한다는 인식이 확산되고 있다. 작은 지역 단위를 겨냥한 정책과 관련하여 한 가지 중요한 것은 각 지역 형태를 고려하여 발전 전략과 정책이 어느 정도 수행가능한지의 여부이다. 결과적으로 이는 동일 지역(homogeneous regions)들의 구조와 특성에 관한 많은 문제점들을 낳게 하였다(Alasia, 2004).

특히 WTO 체제 하에서 농산물 시장이 개방되고 전통적인 농업보호 정책이 축소됨에 따라 농업 선진 국가들의 농정이 분야(sector)별 관점에서 지역(territory)별 관점으로 바뀌어가고 있다(임석희, 2005). 우리나라는 FTA 체결 등으로 인해 농어업 및 농어촌을 둘러싼 환경이 크게 변하고 있는 상황에서 농수산 정책 또한 농어업과 관련하여 많은 변화를 꾀하고 있다.

최근 농수산 정책이 농어업 구조 개선에서 농어업 또는 농어촌 종합 정책으로 전환되면서 이를 뒷받침하기 위한 새로운 농어촌 지표 개발에 대한 필요성이 증대되고 있다. 현행 농어업관련 통계는 농어가 및 작물 생산 중심의 통계로서 농어업 및 농어가의 실태를 파악하는 데 유용하게 활용되고 있다. 그러나 이러한 통계들은 복합적이고 다양한 농어촌 지역의 실태를 파악하는 데는 다소 부족하다는 것이 사용자와 농어촌 전문가들의 일반적인 견해이다. 물론 농촌의 생활환경적 측면을 반영한 통계로서 농촌진흥청에서 작성되고 있는 농촌생활지표가 있다. 그러나 이 또한 전국 평균적인 도시·농촌 지표만을 제공하고 있어 다양한 농어촌의 특성을 파악하는 데는 여전히 한계가 있다는 지적이다. 예컨대

농어촌으로 통칭되는 읍면지역들은 도시 접근성, 자연·지리적 환경 및 타 산업 발달정도 등 여러 요인들에 의해 복합적인 속성을 지니기 때문에 지금과 같이 획일적인 속성에 의해 농어촌을 이해하는 것은 바람직한 현상이라 할 수 없다.

이처럼 농어촌으로 통칭되는 지역들이 사실상 지역별로 다양한 특성을 지니고 있음에도 불구하고, 행정구역에 의한 도시·농촌 구분과 같은 단순 구분만으로는 실질적 농업정책 입안 및 실행 등 탄력적 운영면에서 비효율적인 측면이 있다. 따라서 농어촌의 지표별 격차와 추이를 다양하게 파악하고, 지역사회 발전을 위한 정부정책이 보다 효과적으로 실현될 수 있도록 하기 위해서 일률적으로 적용되었던 농어촌을 세분화하고, 이 세분화된 유형별로 농어촌 특성을 반영하는 지표를 제공할 필요성이 제기되고 있다.

2. 연구목적 및 필요성

다른 나라의 경우 대규모적 차원에서는 동일한 특성을 가진 영역이라 하더라도 동일 지역 내의 다양한 특성을 반영하고 정책 지향적인 정보를 제공하기 위해 대규모 형태의 지역을 보다 세분화된 몇 개의 유형으로 구분하여 탄력적으로 운용하고 있다. 우리나라의 농어촌 지역은 현재 행정구역단위에 의해 통상 읍면지역을 농어촌으로 구분하고 있다. 그러나 실제 읍면 지역은 단순한 농어촌의 특성 이외에도 복합적인 특성을 지니고 있기 때문에 행정구역에 의한 도시·농어촌의 2분법적 구분만으로는 다양한 정책 수립 및 집행 상의 한계에 직면하고 있는 것이 현실이다. 따라서 사회 기반시설이 취약하다거나 고령 인구 증가 등 취약한 농어촌 지역에 대한 각종 정책이 보다 효율적으로 구현되기 위해서는 다양한 농어촌의 특성을 반영할 수 있는 농어촌 구분하여, 이를 바탕으로 세분화된 유형별로 필요한 통계를 작성하여 제공할 필요성이 제기되고 있다.

각 국가별로 농어촌을 세분화하는 방법은 다양하다. 미국은 경제적, 정책적 관점에서 6개 또는 7개 범주로 미국 전역을 카운티 단위로 나누고 있다(통계청, 2007). 프랑스는 칸통(canton)¹⁾을 기준으로 비도시지역

의 공간 변화 추이 및 지역별 차이를 평가하여 취약 지역들과 발전 지역들에 관해서 세분화를 수행하였다. 캐나다는 농촌의 정의 및 지역유형 분류는 OECD의 정의 및 분류 체계를 적용하여 기초행정 단위에서 인구밀도가 km^2 당 150명 이하인 지역을 농촌으로 규정하고, 캐나다 통계구역인 CD(Census Division: 주정부와 자치시의 중간정도) 수준에서 도시와 농촌지역 비율에 따라 유형화하고 있다(통계청, 2007).

한편, 농촌을 유사한 특성을 갖는 세부 단위로 유형화하는 작업은 농어촌 전문가들에게 있어서 중요한 관심사이며 지금까지 많은 연구들이 있었다. 많은 연구에서 지역의 다양한 특성을 파악하기 위해 통계적 기법이 자주 활용되었다 (이원섭,2003; 임석희,2005; 성주인과 송미령, 2003). 그러나 이처럼 다변량 통계분석기법을 활용하여 지역을 세부 유형으로 분류할 경우, 분석에 사용된 변수와 분석 기법 선택에 따라 유형화 결과가 달라질 수 있기 때문에 매우 신중해야 함을 미리 언급해 둔다. 이런 이유에서인지 국가 통계기관에서 이러한 통계적 분석 기법에 의해 농어촌을 유형화한 사례는 아직까지 없다. 단, 캐나다 통계청의 경우, OECD기준에 따라 농어촌을 몇 개의 유형으로 분류한 후, 이들 유형들의 속성을 파악하기 위해 통계적 요인분석(factor analysis)결과를 통계청 내부 자료로 활용한 바 있다(Alasia,2004).

최근 통계청에서도 전국 농어촌 지역에 대해서 보다 작은 지역 단위의 통계를 제공할 목적으로 농어촌 유형화에 착수하였다. 현재 우리나라의 경우 농어촌 유형화에 대한 특정한 기준이 마련되어 있지 않기 때문에 다양한 전문가의 의견과 경험적 기술과 통계적 방법을 통해 농어촌 유형화를 시도하였다. 그 과정에서 유형화 분석에 사용된 통계적 절차에 대한 과학적 진단이 필요하게 되었다. 따라서 본 연구의 주요 목적은 다변량 통계분석방법을 활용하여 전국 농어촌 지역을 몇 개의 작은 단위로 세분화하는 데 있다. 본 연구는 연구 목적 달성을 위해 통계적 분석기법을 적용하는데 있어서 핵심적으로 고려되어야 할 점들을 체계적으로 설명하는데 초점을 둔다. 특히, 농어촌을 유형화하는 다양한 방

1) 프랑스 행정구역의 일종으로 프랑스 전역에서 4,000개 정도이며 *departement* 의 지방의회 선거시에만 선거단위로써 주로 활용됨. 간동의 평균 인구는 13,998명(연제기준)이며 *commune*과 *departement*의 중간 규모를 지니고 있음

법 가운데 통계적 방법론에 관해서 중점적으로 검토하고 방법론의 특성을 설명함으로써 실무적용에 도움을 주고자 하는 것이 본 연구의 최종 목적이라 하겠다.

3. 연구범위와 방법

본 연구의 주요 내용은 통계적 방법을 통하여 우리나라 전국적으로 분포되어 있는 농어촌에 대해서 다양한 농어촌의 특성을 반영할 수 있는 몇 개의 유형으로 세분화 하는 것이다. 따라서 본 연구는 농어촌 유형화를 위한 방법론적인 측면에서 접근하되, 세분화 결과에 의한 농어촌 특성을 규명하고 해석하는 것은 고려대상으로 삼지 않았다. 또한 본 유형화 결과는 실무자의 판단에 도움을 주기 위한 연구자 개인의 판단에 의한 것으로서 통계청의 의견과는 무관함을 미리 밝혀두고자 한다.

농어촌 지역의 유형화 대상 지역은 2005년 행정구역 구분에 의한 전국 3,573개 동읍면으로, 이들의 지역적 다양한 특성을 고려하기 위해 인구, 가구, 지역, 산업 및 생활여건 부문에서 통계청 조사를 통해 얻을 수 있는 변수 범위 내에서 모두 52개 변수를 고려하였다. 52개 관련 변수들은 전문가 의견 및 통계적 분석 결과를 토대로 농어촌 유형에 깊은 연관성이 있을 것으로 판단되는 20개 변수를 최종 분석 변수로 사용하였다. 분석방법은 1차적으로 전국 3,573개 동읍면에 대해서 2000년 도시화율을 기준으로 도시와 농촌으로 구분하고, 이 구분에 의한 농어촌 지역을 다시 통계적 군집방법에 의해 몇 개의 농어촌 지역 단위로 세분화하였다. 즉, 본 연구는 주어진 자료에 기반한 탐색적 연구로 그 범위를 제한하고자 한다.

4. 연구의 한계

본 연구는 크게 두 가지 측면에서 근본적인 한계를 지니고 있다. 첫째는 우리나라 행정구역 단위인 동읍면과 같이 매우 작은 지역 단위를 사용하여 전국의 농어촌 지역을 유형화한다는 것이다. 일반적으로 지역 유형을 세분화하는 작업은 분석의 공간 단위가 작을수록 더욱 정밀하기

는 하지만, 분석 자료의 획득이 어렵다는 특징이 있다. 둘째는 분석에 사용될 변수가 상당히 제한적이라는 것이다. 통계적 방법을 이용하여 보다 작은 단위로 세분화하는 데 있어서 가장 중요한 것은 분석에 유용한 변수를 선정하는 것이다. 통계적 방법에 의한 유형화 작업의 경우, 유형화 목적에 가장 부합되는 변수가 사용되었을 때, 그 결과에 대한 신뢰를 확보할 수 있다. 결국 이는 통계분석을 통한 지역의 유형화는 변수 선정과 밀접한 관련이 있음을 나타낸다. 그러나 본 연구는 통계청 조사를 통해 얻어지는 변수를 활용해야 한다는 측면에서 통계청 조사 외의 변수가 지니는 농어촌의 특성은 반영하기 어렵다는 제약이 있다. 특히, 농어촌의 경제적 특성을 파악할 만한 변수가 사용되지 않았다는 점은 매우 아쉬운 점 중의 하나라 할 수 있다.

제 2절 통계적 분석 방법

본 장에서는 연구에 사용된 자료와 분석 방법을 설명하였다. 목적은 이 연구에 사용된 분석적 논리와 통계적 개념을 명확히 하는데 있다. 요인분석과 군집분석과 같은 다변량 통계분석 (multivariate statistical analysis)에 관한 보다 자세한 내용은 Donald(2005) 등을 참고할 수 있다.

연구의 접근 방법은 농어촌의 기능적 관계보다는 지형적(territorial)단위의 특성과 관련하여 유사한 지역으로 구분하는 데 있다. 즉, 유형화의 넓은 의미는 보다 넓은 시각에서 인구·사회적 지표와 통계 분석의 탐색적 속성을 활용하자는 것이다. 접근방법에 대해서는 다음 절에서 간략하게 설명하였다.

먼저, 이 연구에서 강조하는 것은 행정구역 상 동읍면들의 인구·사회적 수행능력을 비교하는데 있다. 이것은 분석 결과가 받아들일만한 동읍면의 수행능력인지 아닌지에 관한 표준을 정하지는 못한다는 것을 의미하기도 한다. 분석 결과를 비교·해석하는데 있어서 사용된 지표들의 범위는 각 지표(indicator)들이 갖는 변동 범위를 기초로 구분한 것으로 사용된 지표들에 대해 사전 경계값을 미리 정하고 있는 것은 아니다. 이러한 의미에서 보면 이 결과를 이용한 비교는 읍면동 단위에서만 의

미가 있을 뿐, 시도 등의 보다 넓은 영역단위에서 수행된 분석결과와는 다른 결과를 갖기 때문에 시도 단위별 비교에는 사용될 수 없다.

또한 통계적 분석 방법을 적용한 농어촌 유형 구분이라는 점에서 이 분석 결과는 자료에 기반한 탐색적 분석결과를 의미한다. 따라서 자료에 따라 또는 사용된 방법 및 정의 기준에 따라서 약간씩 차이를 보일 수 있다. 그렇지만 사전에 알려진 도시·농어촌간 구분 및 다양한 농어촌 지역의 세분화에 대한 정의가 어려운 상황에서 통계적 접근에 의한 유형화 시도가 향후 지역 특성을 반영한 세분화 연구에 많은 도움이 될 것으로 기대한다.

1. 요인분석

요인분석은 크게 두 가지 목적에서 사용되고 있다. 하나는 관측 가능한 변수로부터 관측 가능한 변수들의 변동을 설명할 수 있는 소수의 관측 불가능한 잠재적 변수(요인)를 찾는 것이다. 다른 하나는 서로 상관관계가 높은 관측 가능한 변수들에 대해서 서로 독립적인 잠재 변수를 찾고자 하는 데 있다. 경우에 따라서는 상관관계가 높은 변수들이 분석 변수로 적당하지 않은 때가 있다. 대표적인 예가 회귀분석에서 설명변수들끼리 상관관계가 높을 때 발생하는 다중공선성(multicollinearity) 문제를 들 수 있다. 이런 경우에 요인분석을 통해 서로 독립적인 요인들을 찾은 후, 이를 회귀분석의 설명변수로 활용함으로써 다중공선성의 문제를 극복할 수 있다. 그렇다 하더라도 원 변수의 정보는 사용된 요인에 그대로 함축되어 있기 때문에 분석하는 데는 지장이 없다.

요인모형에서 관측변수들은 각 요인들과 어느 정도의 상관성을 유지해야 한다. 통계적 절차는 서로 강한 상관관계를 갖는 변수그룹과 그렇지 않은 그룹들로 나누는 방법이다. 이러한 일반적 수리적 모형은 다중 회귀방정식(multiple regression)과 유사하다. 각 변수는 요인들의 선형결합으로 표현될 수 있고, 이때 요인들은 관측 불가능한 것으로 구체적인 식은 다음과 같다.

$$x_{ik} = a_{i1}f_{1k} + a_{i2}f_{2k} + \dots + a_{in}f_{nk} + \delta_i$$

여기서, x_{ik} 는 k 번째 관측(우리의 경우 ‘읍면’)에 대한 i 번째 변수 값으로 표준화된 값을 나타낸다. f_{jk} 는 k 번째 관측에 대한 j 번째 요인 값으로 보통 요인점수(factor score)를 나타낸다. a_{ij} 는 j 번째 공통요인에 대한 i 번째 변수의 표준화 회귀 계수(standardized regression coefficient)로서 보통 요인적재(factor loading)값이라 부른다. 마지막으로 δ_i 는 변수 i 에 대해 유일한 요인 특성으로 그 변수가 갖는 오차를 의미한다.

본 연구에서 사용된 20개 변수를 x_1, \dots, x_k , $k = 20$ 라 하면 요인모형은 다음과 같다.

$$x_k = a_1(\text{factor1})_k + a_2(\text{factor2})_k + \dots + a_n(\text{factorn})_k + \delta_x$$

이때 요인점수, *factor score*는 평균 0, 분산 1이 되도록 표준화되었고, 그 값의 범위는 약 -3에서 +3의 값을 갖는다. 따라서 요인점수는 요인이라는 잠재적 변수의 상대적 능력을 나타내고, 요인과 연관된 관측 변수들의 행위(behavior)를 보여준다. 요인적재값 a_i 는 관측변수와 요인간의 밀접한 관계 정도를 설명한다. 변수들은 모두 표준화되었기 때문에 요인 적재값은 각 요인에 할당된 가중치(weight)가 어느 정도인가를 나타낸다. 특히, 본 연구에서는 요인분석 방법으로 주성분 방법을 사용하였다. 주성분 방법에 의해 요인 적재값을 계산할 경우, 요인적재값은 요인과 변수간의 상관성을 나타내고, 요인 적재값의 절대값이 클수록 더 밀접한 상관관계를 갖는다.

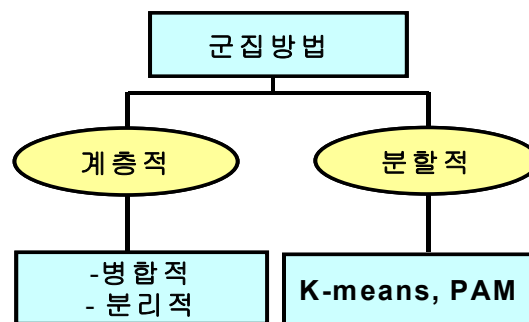
요인 개수를 선정하는 방법은 요인분석에서 계산되는 고유값(eigenvalue)을 이용할 수 있다. k 개의 관측 변수들의 상관계수 행렬로부터 k 개의 고유값이 얻어지는데 크기 순서대로 e_1, e_2, \dots, e_k 라고 하면 $e_1 + e_2 + \dots + e_k = k$ 가 된다. 즉, 가장 작은 고유값 e_1 은 정수 1 기준으로 아주 큰 값이고, 가장 작은 e_k 는 0에 가까운 아주 작은 값을 취한다. 이때 1보다 큰 고유값의 개수 ($r \leq k$)만큼의 요인들을 선정하면 이 요인들을 이용하여 전체 k 개의 변수가 지니는 변수들의 정보를 설명하는데 큰 문제가 되지 않는다는 원리이다. 또 한편으론 추출된 요인들은 전체 자료들의 $(e_1 + \dots + e_r)/k$ 만큼의 정보를 유지하는데 이 비율이 0.7이

상이 되면 통계 분석하는데 크게 문제가 되지 않는다고 볼 수 있다. 물론 0.7은 이론적 근거보다는 통계 전문가들의 경험적 근거에 의한 것이고, 분야에 따라서는 다른 기준을 적용할 수도 있다. 본 연구에서는 고유값 1 기준에 의해 1보다 큰 요인의 개수를 선정하였다.

2. 군집분석

통계적 군집방법은 넓게는 [그림 2.1]과 같이 계층적 방법(hierarchical clustering)과 분할적 방법(partitioning clustering)으로 분류되고 있다. 계층적 방법은 다시 병합적 방법(bottom-up)과 분리적 방법(top-down)으로 나눈다. 병합적 방법에는 최단연결법, 최장연결법, 워드(ward)방법 등 다양한 방법이 있고, 분리적 방법에는 Diana 방법 등이 있다. 분할적 방법에는 K-means, Pam(partitioning around medoids), Fuzzy 방법 등이 있다.

[그림 2.1] 군집방법 분류도



이 중에서 상용패키지(SAS, SPSS 등)를 통해서 통상적으로 널리 활용되고 있는 방법은 계층적 방법으로 병합적 방법, 분할적 방법으로 K-means를 들 수 있다. 본 연구에서도 방법의 상용성과 사용자의 용이성 측면에서 ward 방법에 의한 계층적 방법과 K-means에 의한 분할적 방법을 사용하였다. 각 방법들은 군집을 할당하는 방법에 있어서 구조적으로 군집 알고리즘이 다르기 때문에 방법에 따라 다른 결과를 제시

하고 있다. 따라서 어떤 방법이 확일적으로 좋다고 단언할 수 있는 문제는 아니고, 주어진 자료의 성격이나 연구 목적 및 연구자의 사전 경험에 따라 군집 방법이 선택될 수 있다. 그렇다고 무작정 선택하는 것은 아니고 각 방법에 대한 충분한 검토와 장단점이 파악된 후 분석에 사용될 자료에 가장 적합하다고 판단되는 방법을 선택하는 것이 바람직하다 하겠다.

가. 계층적 군집

계층적 군집 방법 중 ward 방법을 사용하였다. 계층적 군집 방법은 아래의 그림의 덴드로그램과 같이 단계적으로 군집을 형성해 가는 방법이다.

[그림 2.2] 계층적 방법의 군집화 과정



ward 방법은 계층적 군집방법 가운데 가장 보편적으로 사용되는 방법 중 하나이다. 자료의 구조가 특별한 경우가 아니라면 ward 방법은 안정적인 군집 결과를 주는 것으로 알려져 있다. 위 그림에서와 같이 ward 방법은 각 개체를 병합하는 단계에서 잔차제곱합(Error Sum of Squares)의 증분을 최소화 하도록 군집들을 병합시키는 방법이다. 이때 개체들 간의 거리는 유클리드(euclid)거리를 사용하였다. 예를 들어 i 번째 군집 내의 잔차 제곱합을 ESS_i 라 하면, 그 식은 아래와 같이 표현된다.

$$ESS_i = \sum_{j=1}^{n_i} \sum_{k=1}^p (x_{ijk} - \overline{x_{ik}})^2, \quad \overline{x_{ik}} = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ijk}$$

여기서 j 는 개체수, k 는 변수를 나타낸다. 그러면 전체 c 개 군집에 대한 잔차제곱합은 $ESS = \sum_{i=1}^c ESS_i$ 과 같다.

군집이 병합될 때마다 잔차제곱합은 증가하며 ward 방법에서 두 군집 C_K 과 C_L 가 병합될 때에 잔차제곱합의 증분은 다음과 같이 두 군집 사이의 거리를 나타낸다.

$$\text{군집간거리} = \|\overline{\mathbf{x}_K} - \overline{\mathbf{x}_L}\|^2 / \left(\frac{1}{N_K} + \frac{1}{N_L} \right)$$

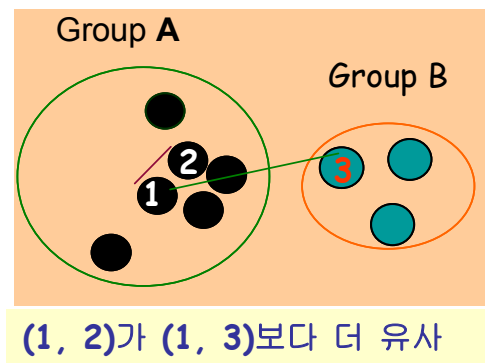
ward방법을 포함한 계층적 군집방법은 개체를 군집에 할당함에 있어서 반복 계산을 하지 않기 때문에 한번 임의의 군집에 할당된 개체는 다음 단계에서 바뀌지 않고 고정된다는 단점이 있다. 또한 일반적인 계층적 군집방법의 경우, 단계별로 개체 또는 군집을 병합하는 과정에서 개체수가 작은 군집들을 병합하는 경향이 있기 때문에 개체수가 비슷한 군집을 상위 단계의 군집으로 병합하는 과정에서 강한 편이가 발생할 수 있다. 또한 개체 간 또는 군집간 거리정보를 이용하여 군집을 병합하기 때문에 이상치(outlier)에 민감하다는 특성을 지니고 있다(Milligan 1980).

나. 분할적 방법

분할적 방법은 비계층적 방법으로서 단계적 방법에 의해 군집을 형성하지 않는다. 즉 계층적 방법과 같이 덴드로그램을 작성하지 않는다. 분할적 방법에서는 보통 개체들을 몇 개의 군집으로 나누기 위해 주어진 판정 기준을 최적화하는 최적분리기법을 사용한다. 분할적 방법에서는 보통 군집의 개수와 군집의 반경을 사전에 미리 정하고 다음과 같이 군집을 형성한다. 먼저 초기 군집의 군집 중심값에 의해 초기 군집을 만

들고, 각 개체를 군집 중심으로부터 가까운 군집에 할당한다. 그 다음 각 군집의 일부 또는 전부를 판정기준에 의해 재할당하는 방법을 사용한다. 개체들 간의 거리는 유클리드 거리를 사용한다. 초기 군집은 관측값들을 기초로 군집의 중심을 선택하고, 각 개체를 중심이 가까운 군집으로 할당한다 [그림 2.3].

[그림 2.3] 비계층적 방법의 군집화 과정



초기 군집의 개체를 재할당하는 방법은 각 군집의 중심을 다시 계산하고, 군집내 제곱합(within sum of squares)이 최소가 되도록 각 개체를 가까운 중심으로 재할당한다. 이러한 재할당에 의한 변화가 적어서 군집이 안정될 때까지 이 과정을 반복한다. 재할당 기준인 군집내 제곱합 W 는 다음과 같다.

$$W = \sum_i \sum_j (x_{ij} - \bar{x}_i)(x_{ij} - \bar{x}_i)'$$

K-means 방법은 100개 이상의 개체수가 많은 경우의 군집분석에 유리한 반면에 개체수가 작을 경우에는 오히려 개체들의 순서에 매우 민감하게 작용하는 단점을 지니고 있다. 또한 사전의 군집의 개수가 미리 정해져야 하고, 최종 군집결과는 초기 군집중심 설정에 상당히 의존하는 경향이 있다.

3. 군집 개수 선정 기준

군집개수 선정은 CCC(Cubic Clustering Criteria; Sarle,1983)와 pseudo F(Milligan and Cooper, 1985)값을 사용하였다. 이 방법은 SAS 패키지의 'proc cluster' 와 'proc fastclust' 를 통해 쉽게 구할 수 있다.

가. CCC

CCC는 자료가 균등분포(uniform distribution)를 따른다는 가정 하에서 이 자료들의 기대분포와 군집들의 편향 정도를 비교하는 측도이다.

$$CCC = \ln \left[\frac{1 - E(R^2)}{1 - R^2} \right] \times k$$

여기서 $E(R^2)$ 은 R^2 의 기댓값, R^2 는 관측된 결정계수, K 는 분산안정화 변환(Sarle, 1983)을 나타낸다. CCC가 양의 큰 값을 가질 때 보다 좋은 군집결과를 나타낸다. 더 큰 값의 CCC는 군집이 1개인 균등분포와 해당 군집 개수일 때 군집 간 분리가 더 큰 차이를 보인다는 것을 의미한다. 보통 CCC는 $CCC > 2$ 이면서 국소 최대값(local maximum value)을 취할 때의 군집의 개수를 선택한다. 그러나 CCC는 군집분석에 사용된 변수들 간에 강한 상관관계가 있을 때에는 부정확할 수 있기 때문에 사용하는데 주의해야 한다.

나. pseudo-F

pseudo F(psF)는 군집의 밀집 정도를 측정하는 통계량으로서, 군집 간 분산과 군집 내 분산의 비를 사용한다(Milligan and Cooper, 1983). psF는 다음과 같이 계산된다.

$$psF = \frac{R^2 / (C - 1)}{1 - R^2 / (n - C)}$$

여기서, C 는 군집개수, R^2 는 결정계수로서 군집 간 분산을 의미하고, n 은 총 개체수를 나타낸다. 이때 psF 값이 클수록 좋은 군집결과를 나타낸다.

제 3절 농어촌지역 유형 분석

1. 자료와 변수 설명

본 연구는 우리나라 2005년 행정구역 구분에 의한 전국 3,573 동읍면을 지역 유형화의 대상으로 하였다. 농어촌 지역 유형화에 필요한 변수를 총 57개 변수를 고려하여, 최종적으로 20개 변수를 분석대상으로 하였다. 20개 변수에 대한 설명은 <표3.1> 과 같다.

<표 3.1> 변수 유형별 변수 설명

변수 명	변수 설명	
인구	인구밀도	읍면동별 인구(명)/읍면동별 면적(km^2)
	노령화 지수	(읍면동별 65세 이상 인구/0~14세 인구)*100
	연평균 인구 증가율	((2005년/2000년) ^{1/5} -1)*100
	농림어업인구비율	읍면동별 농림어업인구/읍면동별 총 인구
가구	농가율	읍면동별 농가 수/읍면동별 총 가구 수
	전업농가율	읍면동별 전업농가 수/읍면동별 농가 수
	2종겸업농가율	읍면동별 2종겸업농가 수/읍면동별 농가 수
	어가율	읍면동별 어가 수/읍면동별 총 가구 수
	전업어가율	읍면동별 전업 어가 수/읍면동별 총 가구 수
	2종 겸업 어가율	읍면동별 2종 겸업 어가 수/읍면동별 어가 수

〈표 3.1〉 변수 유형별 변수 설명(계속)

변수 명	변수 설명
지형	평야지대 비율
	산간지역 비율
	해안지역 비율
	섬지역 비율
산업	제조업종사자수 비율
	도소매·음식숙박업종 사자수 비율
생활	수세식화장실수 비율
여건	아파트수 비율
	노후주택 비율
	전작가구수 비율

2. 분석절차

전국 3,573 동읍면 지역을 대상으로 먼저 도시와 농어촌을 구분하였다. 이들 중에서 농어촌으로 분류된 지역에 대해 유사한 농어촌 특성을 갖는 몇 개의 농어촌 지역으로 유형화하였다. 분석 절차를 설명하면 다음과 같다.

Step1. 2005년 행정구역 단위를 기준으로 3,573 전국 동읍면에 대해서 우선 도시와 농촌을 구분하였다. 이때 구분 변수로 2000년 기준 도시화율²⁾을 사용하였다. K-means 군집 방법에 의해 2개의 군집

2) GIS도시화율이란, 해당지역 전체 인구에서 도시화지역 인구가 차지하는 비율로서 해당지역 (도시화지역 인구/해당지역 전체인구)×100.

- 도시화지역: 동읍면내 기초단위구가 ①인구밀도 3,000명/km²이상 ② 지목기준 도시

으로 나누고 읍면지역이 많이 포함된 그룹을 농어촌 지역 그룹이라 하였다. 이때 농어촌 그룹에서 ‘동’에 해당되는 지역은 따로 구분하여 농어촌에 준하는 그룹으로 명명하고, 유사한 개념으로 도시그룹에 포함된 읍면지역 또한 준도시적 성격을 지닌 그룹으로 명명한다. 이렇게 해서 최종적으로 1,300개의 읍면지역을 순수 농어촌 지역으로 구분하고 이를 다음 단계의 분석대상으로 하였다.

- Step2. 분석에 사용된 변수는 2단계에 걸쳐 선정하였다. 1차적으로 원래 총 52개 변수에 대해서 전문가와 실무담당자의 의견을 반영하여 25개를 선정하였다. 이 25개 변수를 Step1에서 선정된 1300개 분석대상에 대해서 요인분석을 수행한 후, 각 변수 간 상관관계와 각 변수의 요인들에 대한 통계적 기여도(공통성 0.5이상)를 고려하여 농어촌지역을 세분화하는데 중요도가 높을 것으로 예상되는 20개 변수를 최종적으로 선정하였다.
- Step3. 선정된 20개 변수는 서로 상관관계가 높은 변수들으로써 변수 간에 서로 중복된 정보를 포함하고 있다. 따라서 20개의 변수 정보를 최대한 활용하되, 서로 중복된 정보 활용을 피하기 위해 요인분석을 수행하여 서로 독립된 새로운 변수를 구성하였다. 요인분석 결과에 의해 고유값 1 기준으로 총 5개의 요인을 선택하고, 이를 추후 군집분석의 입력변수로 활용하였다.
- Step4. Step3의 5개 요인에 대해 1,300개 읍면지역을 대상으로 군집분석을 수행하였다. K-means 와 계층적 방법을 적용하여 각각 4개와 5개의 군집을 최종 군집 개수로 정하고, 각 방법에 의한 군집결

토지 이용면적 비율이 50% 이상 중 최소한 어느 하나를 충족 시키면서 공간상에서 인접한 기초 단위구들에 둘러싸인 구역을 포함하되 총 인구가 3,000명 이상인 지역을 말한다.

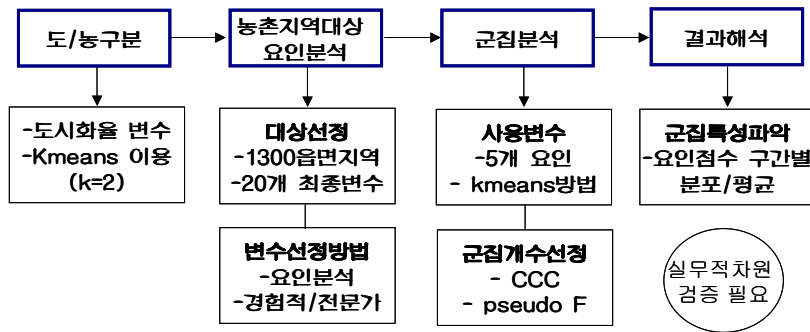
- 기초단위구란 지도상에 준항구적인 지형지물(도로, 하천, 산능선 등과 같이 변화가 적고 식별이 확실한 인공물 또는 자연물)을 이용하여 구획한 최소 단위구역을 말한다 (단, 뚜렷한 지형지물이 없는 경우는 통/리 경계를 보조적으로 활용한다)

과를 비교하였다. 군집 선정기준은 CCC와 pseudo F값을 사용하였다.

Step5. 마지막으로, 군집 결과는 각 군집에 대한 요인점수의 구간별 분포(7개 구간)와 20개 변수의 평균을 이용하여 군집의 특성을 설명하고자 하였다. 또한 각 군집에 포함된 지역을 mapping 하여 2차적으로 분류된 결과에 대한 타당성을 확인하였다.

위의 Step1에서 Step5 까지의 절차를 요약하면 [그림 3.1]과 같다.

[그림 3.1] 농어촌 유형화를 위한 분석 과정



3. 분석결과

가. 도시와 농어촌 구분

2005년 기준 행정구역 단위 3,573 동읍면 지역에 대해서 도시와 농어촌 두 그룹으로 분류하기 위해 군집의 수를 2로 하여 K-means 군집 방법을 수행하였다. 2000년 도시화율 변수를 사용하여 분석한 결과가 <표 3.2>과 같다. <표 3.2>에서 두 그룹을 구분 짓는 도시화율의 경계값은 농어촌 지역 그룹은 최대값 0.4861이고, 도시지역 그룹은 최소값 0.4908

이다. 도시화율 경계값에 의해 2분법적 도시·농어촌을 적용하면 1,432개 농어촌과 2141 도시지역으로 구분된다. 그러나 2분법적 도시와 농어촌 구분은 다양한 지역적 특성을 반영하기 어렵다는 실무자와 전문가 의견을 반영하여 4분법적 구분에 의한 도시·농어촌 구분법을 적용하였다. <표 1>에 의하면 농어촌지역 1,432개 중에서 읍면 지역은 1,300개, 행정구역상 도시로 분류되어진 ‘동’은 132개이다. 도시지역으로 분류된 2,141개 동읍면지역 가운데 ‘동’은 2,031개이고, ‘읍면’은 110개로 나타났다. 따라서 농어촌지역의 1,300개 읍면을 ‘순수 농어촌’과 132개 동을 ‘준 농어촌’이라 편의상 명명한다. 도시지역도 마찬가지로 110개 읍면을 ‘도시적 농어촌’, 2031개 동을 ‘순수 도시’라 명명한다.

4분법 구분에 의한 1,300개 읍면지역인 ‘순수 농어촌(이하 농어촌)’을 농어촌 유형의 분석대상으로 하였다.

<표 3.2> 도시화율에 의한 도시와 농어촌의 구분

2분법적 구분	4분법적 구분	지역 분포	도시화율		
			최소값	최대값	
농어촌 지역	순수농어촌	읍	119	0	0.4861
		면	1181		
	준농어촌	동	132		
	총	1432			
도시 지역	도시적 농어촌	읍	91	0.4908	1
		면	19		
	순수 도시	동	2031		
	총	2141			

나. 요인분석 결과

요인분석 변수로서 농어촌유형 분석 대상이 되는 52개 변수를 우선적으로 고려하였다. 이 중에서 실무적 특성과 또는 농어촌 지역 분석 연구자들의 의견을 반영하여 35개의 변수를 1차적으로 선정하고, 이를 통계적 요인분석에 의해 요인추출을 위한 영향력 정도를 고려하여 최종

20개 변수를 선정하였다. 요인분석에서는 초기 공통성(community)이 약 0.5이상이면서 요인적재값이 0.5이상인 변수를 대상으로 하였다.

20개 변수에 대해서 주성분(principal component)방법에 의한 요인분석을 수행하고, 베리맥스(varimax) 회전하여 고유값 1 기준에 의해 요인 5개를 최종 선정하였다. 요인 5개는 전체 분산에 대해 약 72.4%의 설명력을 갖는다. 이 72.4%의 의미는 요인분석에 사용된 20변수들의 72.4%가 5개의 요인에 의해 설명되어진다는 것이다. 5개 요인은 주어진 자료를 에 대한 결과는 <표 3.4>와 같다.

<표 3.4>는 요인적재값의 절대값이 각 요인별로 0.5이상인 경우만 나타내었다. 요인명은 각 요인들에 강하게 연관된 변수들을 기초로 명명되어지게 된다. 따라서 요인에 명명된 요인명은 다소 주관성이 개입될 수는 있지만, 그 요인명은 각 요인에 강하게 연관되어진 원 변수들의 특성을 반영하는 경향이 있다. 5개 요인명은 다음과 같다.

<표 3.3> 요인특성

요인	요인 특성
요인 1	고령전업농촌
요인 2	해안겸업농어촌
요인 3	섬중심전업어촌
요인 4	산간전작농어촌
요인 5	도소매숙박업종사자 비율이 높은 지역

일반적으로 요인은 해석상 측면에서 보면, 개념적으로 일치해야 하고, 의미있고 적절한 특성을 나타낼 수 있어야 한다. 각 요인은 몇 개의 변수와 매우 강한 연관성을 갖고 있지만 여전히 나머지 변수들에 대해서도 상당한 효과를 포함하고 있다. 이러한 변수들의 결합 때문에 어떤 요인들은 설명되어지는 변수들이 바람직하든 그렇지 않든 또는 국소적 특성이거나 지시적이라는 점에서 진단에 더 많이 도움을 주게 되고, 다른 요인들은 설명되어지는 변수들이 강한 함축성을 갖지 않는다는 의미에서 더 탐색적인 경향이 있다.

〈표 3.4〉 주성분방법에 의한 요인분석 결과

변수	요인1	요인2	요인3	요인4	요인5
농리어업인구비율	0.91				
농가율	0.89				
노령화지수	0.81				
노후주택율	0.66				
2종겸업농가율	-0.64				
연평균증가율	-0.68				
수세식화장실비율	-0.72				
인구밀도	-0.73				
아파트수 비율	-0.81				
해안지역 비율		0.83			
2종겸업어가율		0.76			
어가율		0.70	0.65		
전업농가율	0.52	-0.56			
섬지역			0.90		
전업어가율			0.79		
산간지역				0.81	
전작생산가구비율				0.54	
평야지대비율				-0.83	
도소매숙박업 종사자수비율					0.87
제조업종사자 비율					-0.85
고유값	5.78	2.54	2.47	1.99	1.69
총분산비율(%)	30.4	19.2	10.7	6.6	5.5

다. 군집분석 결과

계층적 방법과 K-means 방법에 의해 군집 분석을 수행하고 두 결과를 비교하였다. 군집분석에서 가장 우선적으로 결정되어야 할 것은 군집의 개수이다. 군집의 개수가 결정되어지면, 각 개체는 군집중심에 가까운 군집으로 할당되게 된다. 각 군집방법에 따른 CCC와 psF 값은 <표

3.5>과 같다. K-means 방법은 CCC와 psF의 큰 값 기준에 의해 군집 수를 4개로 선정하였다. 계층적 방법은 CCC기준으로 5개, psF로 5개 또는 6개가 가능하다. 계층적 방법의 psF값을 값을 보면, 군집 수가 5일때는 300, 6일때는 309로 그 증가폭이 그다지 크지 않다. 따라서 분석에 사용된 변수가 제한적이라는 점을 고려하여 가능하면 적은 수의 군집 수를 선택하는 방법을 사용하였다. 최종적으로 K-means는 4개, 계층적 방법은 5개를 최종 군집 개수로 선정하였다.

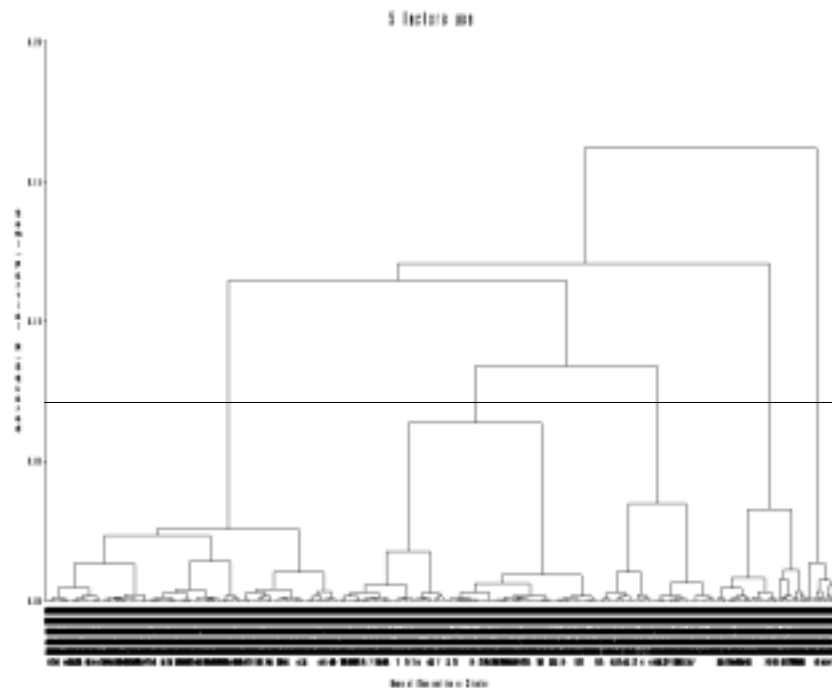
〈표 3.5〉 군집방법별 CCC와 psF, $c \leq 6$

군집방법	군집수	CCC	psF
K-means	2	-11.279	162.30
	3	2.347	250.00
	4	7.540	283.70
	5	-15.377	196.85
	6	-5.195	253.00
계층적	2	3.020	250.00
	3	3.240	255.00
	4	7.680	284.00
	5	8.830	300.00
	6	7.830	309.00

[그림 3.1]은 계층적 방법에 의해 각 개체 간 또는 군집 간의 병합과정을 보여주는 덴드로그램이다. [그림 3.1]에서 x축은 개체(지역), y축은 해당 군집수에 대해 결정계수값을 나타낸다. 일반적으로 군집 개수를 정하는 데 있어서 덴드로그램을 참고하는 경우가 많다. 덴드로그램에 의존하여 군집개수를 정하는 것은 매우 시각적이고 주관적일 수 있기 때문에 사용에 있어서 주의를 기울일 필요가 있다. 그렇지만 덴드로그램은 군집형성에 대한 탐색적 기능을 한다는 점에서 군집의 개수를 정하는데 좋은 참고가 될 수 있다. 본 자료의 경우 계층적 방법에 의해 군집

개수를 5개로 정할 때 아래의 점선의 형태로 구분지을 수 있으며 각 그룹에 할당되는 개체들의 규모를 예측할 수 있다.

[그림 3.1] 계층적 군집방법의 덴드로그램



K-means와 계층적 방법에 의한 군집별 개체 분포는 <표 3.6>와 같다. <표 3.6>에서 군집1부터 군집5라는 군집명 자체는 군집방법과는 무관하고, 단지 군집의 개수가 4개 또는 5개로 나뉘었다는 것만을 의미할 뿐이다. 각 군집의 특성을 고려한 작명과 관련한 내용은 뒤에서 설명하기로 한다. K-means는 군집 개수가 4개로 전체 1,300읍면 지역 중 910개(70%) 지역이 군집1에 할당되었다. 즉, 910개 지역은 서로 유사한 특성을 지닌 지역들로 이 군집에 할당된 농어촌 지역들은 5개 요인에 대해서 매우 유사한 특성을 갖는다는 것을 볼 수 있다. 군집 수를 5개를 사용한 계층적 방법은 군집 개수가 4개인 K-means 보다는 한 군집에 집중되는 양상은 나타나지 않았다. K-means와 계층적 방법은 군집을 형성하는 알고리

준 자체가 다르기 때문에 군집의 분포만으로 군집 결과에 대해 직접적인 비교는 할 수 없다. 군집 수 또는 군집 방법 선택에 있어서의 타당성 문제는 추후 다양한 검증 작업을 거쳐 판단하는 것이 바람직할 것이다.

〈표 3.6〉 군집방법별 군집분포

군집방법	지역개수(%)				
	군집1	군집2	군집3	군집4	군집5
K-means, c=4	910(70.0)	301(23.2)	39(3.0)	50(3.8)	-
계층적, c=5	194(14.9)	430(33.1)	483(37.2)	145(11.1)	48(3.7)

군집 명은 각 군집에 포함된 지역들의 특성을 반영할 수 있어야 한다. 또한 이러한 명명 작업은 자료 분석 결과를 토대로 놓어준 전문가들의 전문적 지식을 바탕으로 이루어져야 한다. 그럼에도 불구하고 본 연구는 통계적 해석에 초점을 두었기 때문에 놓어준 전문가들의 의견은 반영하지 않기로 한다. 가장 일반적인 접근 방법은 분석에 사용된 변수와 군집 간의 연관성 정도를 고려할 수 있다. 게다가 각 군집별로 원 변수들의 정보(평균을 포함한 기초통계)를 이용하는 것도 군집의 특성을 파악하는데 도움이 될 수 있다. 이 연구에서는 20개의 원 변수들을 5개의 요인으로 축약하여 분석에 사용하였기 때문에 우선적으로 이 요인들과 군집에 속한 개체들과의 관련성을 파악하고, 원 변수들의 평균을 추가적으로 해석해 보았다.

이를 위해 5개의 요인점수(양적변수)를 7개의 범주를 갖는 구간으로 변환하였다. 이때 요인점수는 평균 0, 분산 1이 되도록 표준화되었기 때문에 약 +3과 -3사이의 값을 취한다. 이때 요인점수를 표준화하는 이유는 요인별 상대적 비교를 할 수 있다는 장점이 있기 때문이다. 또한 요인점수를 범주화하는 특별한 기준은 없다. 다만, 양적자료인 요인점수를 세분화시킴으로써 요인점수 구간별 놓어준지역의 분포를 파악하고자 하였다. 이때, 7개 구간³⁾에 대한 근거는 경험적 판단에 따르되, 7개보다

3) 구간범위: 구간1 \leq -3, -3 $<$ 구간2 \leq -2, -2 $<$ 구간3 \leq -1, -1 $<$ 구간4 \leq 0, 0 $<$ 구간5 \leq 1, 1 $<$ 구간6 \leq 2, 2 $<$ 구간7

더 많으면 구간화의 의미가 상쇄될 수 있고, 더 작은 경우 구간별 분포를 파악하는데 한계가 있다는 점을 고려하였다. 구간번호가 클수록 요인점수가 높다.

<표 3.7>에서 <표 3.10>는 K-means 군집 결과를 4개 군집별로 각 요인구간에 해당하는 지역분포를 나타낸 것이다.

- 군집1: 이 군집에 속한 지역들의 63.63%가 요인 1인 고령전업농촌의 특성을 강하게 반영함과 동시에 산간전작 요인과 도소매숙박종사자 요인을 끌고루 지냈다. 반면에 해안겸업관련요인과 섬전어업요인에 대해서는 전체 지역 중 약 71.54%와 75.38%가 요인점수 구간 4에 분포되어 이 두 요인의 성격은 대체로 낮게 반영되었음을 알 수 있다.
- 군집2: 대부분의 지역들은 도소매숙박종사자특성이 강한 요인에 대해서는 요인점수 구간에 고루 분포하면서 나머지 요인들의 특성에 대해서는 대부분 낮은 구간에 분포되어 있다. 따라서 이 군집은 농어촌의 특성이 약한 지역들이 포함되었을 가능성이 높다.
- 군집3: 군집 3에 속한 지역들의 100%가 섬지역전어업요인에 대해 가장 높은 요인구간에 속하면서 고령전업농촌요인과 산간전작농촌 및 도소매숙박종사요인의 특성에 대해서 50%이상지역이 5이상 구간에 분포되었다. 따라서 이 군집은 어업을 주로 하는 섬지역 특성과 농촌과 도소매음식숙박업이 다소 발달되어 있는 지역들의 특성을 반영한다고 볼 수 있다.
- 군집4: 이 군집에 속한 100% 지역이 요인 2인 해안겸업요인에 대해 요인점수 구간 7에 포함되었고, 대부분은 지역들은 고령전업농촌요인과 도소매숙박종사요인에 대해서 요인구간 5이상에

속하였다. 반면에 섬전어업요인과 산간전작요인에 대해서는 이 군집에 속한 지역들의 50% 이상이 낮은 요인구간(구간4이하)에 포함되었다. 이에 군집4는 해안지역 특성이 강하고 농업과 어업을 겸하는 지역으로서 도소매음식숙박업 등 서비스업이 발달된 지역 특성을 가졌다고 볼 수 있다.

그러나 요인 구간별 분포만에 의해서 군집의 특성을 규명하는 것은 정확성 측면에서 신뢰성이 축소될 가능성이 있기 때문에 이를 보완하기 위해 각 군집별로 분석에 사용된 20개 변수들의 평균값(표준편차)를 살펴해보았다.

〈표 3.7〉 군집1에 할당된 910개 지역들의 각 요인 구간별 분포

	요인점수 구간						
	1	2	3	4	5	6	7
요인1 고령전업	0 (0.0)	0 (0.0)	0 (0.0)	189 (20.77)	579 (63.63)	139 (15.27)	3 (0.33)
요인2 해안겸업	0 (0.0)	0 (0.0)	19 (2.09)	651 (71.54)	187 (20.55)	45 (4.95)	8 (0.88)
요인3 섬전어업	0 (0.0)	0 (0.0)	1 (0.11)	686 (75.38)	213 (23.41)	1 (0.11)	9 (0.99)
요인4 산간전작	0 (0.0)	13 (1.43)	176 (19.34)	283 (31.10)	257 (28.24)	156 (17.14)	25 (2.75)
요인5 도소매숙박	0 (0.0)	35 (3.85)	110 (8.46)	286 (31.43)	385 (42.31)	89 (9.78)	5 (0.55)

〈표 3.8〉 군집2에 할당된 301개 지역들의 각 요인 구간별 분포

	요인점수 구간						
	1	2	3	4	5	6	7
요인1	15	44	142	100	0	0	0
고령전업	(4.98)	(14.62)	(47.18)	(33.22)	(0.0)	(0.0)	(0.0)
요인2	0	0	5	214	64	18	0
해안겸업	(0.0)	(0.0)	(1.66)	(71.10)	(21.26)	(5.98)	(0.0)
요인3	0	0	1	215	82	3	0
섬전어업	(0.0)	(0.0)	(0.33)	(71.43)	(27.24)	(1.0)	(0.0)
요인4	0	0	25	141	90	38	7
산간전작	(0.0)	(0.0)	(8.31)	(46.84)	(29.90)	(12.62)	(2.33)
요인5	0	8	62	66	85	65	15
도소매숙박	(0.0)	(2.66)	(20.60)	(21.93)	(28.24)	(21.59)	(4.98)

〈군집 3.9〉 군집3에 할당된 39개 지역들의 각 요인 구간별 분포

	요인점수 구간						
	1	2	3	4	5	6	7
요인1	0	0	1	15	20	3	0
고령전업	(0.0)	(0.0)	(2.56)	(38.46)	(51.28)	(7.69)	(0.0)
요인2	0	0	7	12	11	5	4
해안겸업	(0.0)	(0.0)	(17.95)	(30.77)	(28.21)	(12.82)	(10.26)
요인3	0	0	0	0	0	0	39
섬전어업	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(100)
요인4	0	0	0	7	24	8	0
산간전작	(0.0)	(0.0)	(0.0)	(17.95)	(61.54)	(20.51)	(0.0)
요인5	0	0	3	11	22	3	0
도소매숙박	(0.0)	(0.0)	(7.69)	(28.21)	(56.41)	(7.69)	(0.0)

<표 3.10> 군집4에 할당된 50개 지역들의 각 요인 구간별 분포

	요인점수 구간						
	1	2	3	4	5	6	7
요인1 고령전업	0 (0.0)	0 (0.0)	0 (0.0)	16 (32.0)	33 (66.0)	1 (2.0)	0 (0.0)
요인2 해안겸업	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	50 (100)
요인3 섬전어업	0 (0.0)	0 (0.0)	2 (4.0)	31 (62.0)	12 (24.0)	4 (8.0)	1 (2.0)
요인4 산간전작	0 (0.0)	0 (0.0)	6 (12.0)	25 (50.0)	17 (34.0)	2 (4.0)	0 (0.0)
요인5 도소매숙박	0 (0.0)	2 (4.0)	6 (12.0)	7 (14.0)	22 (44.0)	10 (20.0)	3 (6.0)

<표 3.11>에서 <표 3.15>는 계층적 법에 의한 5개 군집결과를 나타낸다.

- 군집1: 군집1에 할당된 194개 지역은 5개 요인 모두에 대해서 대체로 낮은 요인점수 구간에 대부분의 지역들이 분포되어 있다. 따라서 군집1은 일반적인 농어업의 특성을 반영하고 있지 않은 지역으로 비농어업적 지역특성을 반영한다고 볼 수 있다.
- 군집2: 군집2에 할당된 430개 지역은 고령전업농촌요인과 도소매음식숙박업종사비율이 높은 요인에 대해서는 50%이상의 지역들이 요인점수가 높은 구간에 속하는 반면, 해안겸업과 섬전어업요인에 대해서는 대부분의 지역들이 구간4인 낮은 구간에 분포되어 있다. 따라서 이 지역은 산촌과 어촌의 특성을 보이지 않는 전형적인 평야지대를 많이 포함하는 고령농촌의 특성을 반영하는 경향이 있다.
- 군집3: 군집3에 할당된 483개 지역은 고령전업농촌요인과 산간전작요인에 대해서 70%이상 지역들이 요인점수 구간5에 포함되

었다. 반면에 해안겸업과 섬전어업요인에 대해서는 각각 70% 또는 80%이상이 구간 4이하에 포함되면서 이 요인들의 특성을 낮게 반영하고 있다. 도소매음식숙박업종사비율이 높은 요인에 대해서는 전 구간에 걸쳐 지역들이 골고루 분포되어 있다. 따라서 이 군집은 전형적인 산간지역을 포함한 전형적인 농촌의 특성을 반영하는 경향이 있다.

- 군집4: 군집4에 할당된 145개 지역은 해안겸업요인에 대해 약 80% 지역들이 요인점수 구간 6과 7에 속하면서 이 요인의 특성을 매우 높게 반영하고, 또한 고령전업농촌과 도소매숙박종사요인의 특성도 어느 정도 반영되어 있다. 반면에 섬전어업과 산간전작요인에 대해서는 낮은 요인점수 구간에 많은 지역들이 분포함으로써 이 요인들의 특성은 대체로 낮게 반영하고 있다. 이 군집은 해안지역중심으로 농업과 어업을 겸하거나 도소매음식숙박업 등이 발달한 지역일 가능성이 높다.
- 군집5: 군집5에 할당된 48개 지역은 섬전어업요인에 대해서 100%에 해당하는 지역들이 요인점수 구간 7에 포함되어 있고, 나머지 요인들의 특성도 어느 정도는 반영하고 있는 것으로 나타났다. 따라서 이 군집은 섬 지역으로 농업과 어업 및 서비스업 등이 발달한 지역들이 포함되었을 가능성이 있다.

<표 3.11> 군집1에 할당된 194개 지역들의 각 요인 구간별 분포

	요인점수 구간						
	1	2	3	4	5	6	7
요인1 고령전업	14 (7.22)	30 (15.46)	46 (23.71)	47 (24.23)	55 (28.35)	2 (1.03)	0 (0.0)
요인2 해안겸업	0 (0.0)	0 (0.0)	1 (0.52)	134 (69.07)	56 (28.87)	3 (1.55)	0 (0.0)
요인3 섬전어업	0 (0.0)	0 (0.0)	0 (0.0)	140 (72.16)	54 (27.84)	0 (0.0)	0 (0.0)
요인4 산간전작	0 (0.0)	0 (0.0)	12 (6.19)	118 (60.82)	56 (28.87)	8 (4.12)	0 (0.0)
요인5 도소매숙박	0 (0.0)	29 (14.95)	110 (56.70)	52 (26.80)	3 (1.55)	0 (0.0)	0 (0.0)

<표 3.12> 군집2에 할당된 430개 지역들의 각 요인 구간별 분포

	요인점수 구간						
	1	2	3	4	5	6	7
요인1 고령전업	1 (0.23)	13 (3.02)	80 (18.60)	90 (20.93)	204 (47.44)	42 (9.77)	0 (0.0)
요인2 해안겸업	0 (0.0)	0 (0.0)	4 (0.93)	385 (89.53)	41 (9.53)	0 (0.0)	0 (0.0)
요인3 섬전어업	0 (0.0)	0 (0.0)	0 (0.0)	284 (66.05)	146 (33.95)	0 (0.0)	0 (0.0)
요인4 산간전작	0 (0.0)	12 (2.79)	164 (38.14)	195 (45.35)	43 (10.0)	13 (3.02)	3 (0.7)
요인5 도소매숙박	0 (0.0)	0 (0.0)	3 (6.25)	15 (31.25)	26 (54.17)	3 (6.25)	1 (2.08)

〈표 3.13〉 군집4에 할당된 483개 지역들의 각 요인 구간별 분포

	요인점수 구간						
	1	2	3	4	5	6	7
요인1 고령전업	0 (0.00)	0 (0.00)	11 (2.28)	116 (24.02)	262 (54.24)	91 (18.84)	3 (0.62)
요인2 해안겸업	0 (0.00)	0 (0.00)	12 (2.48)	344 (71.22)	124 (25.67)	3 (0.62)	0 (0.00)
요인3 섬전어업	0 (0.00)	0 (0.00)	0 (0.00)	423 (87.58)	60 (12.42)	0 (0.00)	0 (0.00)
요인4 산간전작	0 (0.00)	0 (0.00)	1 (0.21)	61 (12.63)	220 (45.55)	172 (35.61)	29 (6.00)
요인5 도소매숙박	0 (0.00)	11 (2.28)	46 (9.52)	155 (32.09)	199 (41.20)	66 (13.66)	6 (1.24)

〈표 3.14〉 군집4에 할당된 145개 지역들의 각 요인 구간별 분포

	요인점수 구간						
	1	2	3	4	5	6	7
요인1 고령전업	0 (0.00)	1 (0.69)	5 (3.45)	52 (35.86)	83 (57.24)	4 (2.76)	0 (0.00)
요인2 해안겸업	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	30 (20.69)	57 (39.31)	58 (40.0)
요인3 섬전어업	0 (0.00)	0 (0.00)	4 (2.76)	85 (58.62)	47 (32.41)	8 (5.52)	1 (0.69)
요인4 산간전작	0 (0.00)	1 (0.69)	30 (20.69)	67 (46.21)	44 (30.34)	3 (2.07)	0 (0.00)
요인5 도소매숙박	0 (0.00)	4 (2.76)	9 (6.21)	15 (10.34)	80 (55.17)	28 (19.31)	9 (6.21)

<표 3.15> 군집5에 할당된 48개 지역들의 각 요인 구간별 분포

	요인점수 구간						
	1	2	3	4	5	6	7
요인1 고령전업농업	0 (0.00)	0 (0.00)	1 (2.08)	15 (31.25)	28 (58.33)	4 (8.33)	0 (0.00)
요인2 해안겸업	0 (0.00)	0 (0.00)	14 (29.17)	14 (29.17)	11 (22.92)	5 (10.42)	4 (8.33)
요인3 섬전업어업	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	48 (100)
요인4 산간전작	0 (0.00)	0 (0.00)	0 (0.00)	15 (31.25)	25 (52.08)	8 (16.67)	0 (0.00)
요인5 도소매숙박	0 (0.00)	0 (0.00)	3 (6.25)	15 (31.25)	26 (54.17)	3 (6.25)	1 (2.08)

라. 군집에 의한 유형화 결과 해석

<표 3.16>과 <표 3.17>는 각 군집에 포함된 읍면지역들의 20개 변수들에 대한 평균과 표준편차를 나타낸다. 그리고 두 표에서 1,300개 전체 지역에 대한 각 변수별 평균값을 기준값으로 설정하고, 각 군집에서 전체 평균보다 높으면 그 군집에서 해당 변수의 특성이 높게 반영되는 것으로 판단한다. 이는 변수 특성에 대한 반영 정도를 측정할 만한 기준을 정하기가 어렵고, 각 군집별 특성이 전체 평균을 기준으로 높거나 낮으면 어느 정도 그 변수의 특성을 반영한다고 볼 수 있기 때문이다. 또한 군집별 평균값 수준이 대체로 높은 경우는 각 군집의 평균값이 전체 평균값과 동일한 수준이라도 어느 정도의 특성을 지니는 것으로 해석하였다.

지금까지 살펴본 요인과 군집 특성과의 관계 그리고 분석에 사용된 20개 변수와 군집 특성과의 관계를 정리하면 <표 3.18>과 <표 3.19>와 같다. 이들 결과들은 각 군집의 특성을 명명하는데 중요한 정보가 될 수 있다. 이미 앞에서 언급한 바와 같이 본 연구는 농어촌 지역 유형화를 위한 통계적 측면을 다루는데 초점을 두었고, 정확한 명명을 위해서는 농어촌 지역 연구에 대한 전문적 지식을 요하기 때문에 각 군집에 대한 정확한 명명은 하지 않을 것이다.

〈표 3.16〉 K-means 에 의한 4개 군집별 평균, 표준편차

분석변수	전체평균 (표준편차)	각 군집별 평균 (표준편차)			
		군 집 1 (910개)	군 집 2 (301개)	군 집 3 (39개)	군 집 4 (50개)
인구밀도	95.7 (121.5)	58.3 (36.7)	210.4 (204.6)	89.3 (55.2)	91.0 (35.4)
노령화지수	2.9 (1.7)	3.5 (1.6)	1.1 (0.5)	3.2 (1.6)	2.6 (0.1)
연평균증가율	-2.6 (2.8)	-3.4 (1.5)	0.1 (4.1)	-3.5 (2.4)	-3.2 (1.6)
농림어업인구율	0.5 (0.2)	0.6 (0.1)	0.3 (0.1)	0.6 (0.1)	0.6 (0.1)
농가율	0.5 (0.2)	0.6 (0.1)	0.3 (0.1)	0.4 (0.2)	0.5 (0.1)
전업농가율	0.7 (0.2)	0.7 (0.1)	0.6 (0.1)	0.4 (0.2)	0.4 (0.2)
2종겸업농가율	0.2 (0.1)	0.1 (0.1)	0.3 (0.1)	0.4 (0.2)	0.3 (0.1)
어가율	0.0 (0.1)	0.0 (0.0)	0.0 (0.0)	0.3 (0.2)	0.3 (0.1)
전업어가율	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.1 (0.1)	0.0 (0.0)
2종겸업어가율	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.1 (0.1)	0.1 (0.1)
제조업종사자율	0.3 (0.2)	0.3 (0.2)	0.3 (0.3)	0.1 (0.1)	0.2 (0.2)
도소매종사자율	0.3 (0.1)	0.3 (0.1)	0.3 (0.2)	0.3 (0.2)	0.4 (0.2)
평야지대비율	0.4 (0.3)	0.4 (0.3)	0.3 (0.2)	0.0 (0.0)	0.2 (0.2)
산간지역비율	0.3 (0.3)	0.4 (0.3)	0.2 (0.2)	0.0 (0.0)	0.1 (0.1)
해안지역비율	0.1 (0.1)	0.0 (0.1)	0.0 (0.1)	0.0 (0.1)	0.6 (0.2)
섬지역비율	0.0 (0.2)	0.0 (0.1)	0.0 (0.0)	1.0 (0.1)	0.0 (0.1)
수세식화장식비율	0.7 (0.2)	0.6 (0.2)	0.8 (0.1)	0.8 (0.1)	0.8 (0.1)
아파트수비율	0.1 (0.1)	0.0 (0.0)	0.3 (0.2)	0.0 (0.0)	0.0 (0.1)
노후주택수비율	0.3 (0.1)	0.4 (0.1)	0.2 (0.1)	0.5 (0.2)	0.3 (0.1)
전작생산가구비율	0.1 (0.1)	0.1 (0.1)	0.1 (0.1)	0.3 (0.3)	0.1 (0.1)

〈표 3.17〉 계층적 방법의 5개 군집별 평균, 표준편차

분석변수	전체평균 (표준편차)	각 군집별 평균 (표준편차)				
		군집1 (194)	군집2 (430)	군집3 (483)	군집4 (145)	군집5 (48)
인구밀도	95.7 (121.5)	226.2 (236.7)	99.9 (77.3)	42.0 (29.8)	91.8 (61.9)	83.7 (51.6)
노령화지수	2.9 (1.7)	1.6 (1.2)	2.8 (1.6)	3.5 (1.9)	2.7 (1.1)	3.1 (1.5)
연평균증가율	-2.6 (2.8)	-0.4 (4.2)	-2.6 (2.8)	-3.2 (1.6)	-3.2 (1.8)	-3.5 (2.2)
농림어업인구율	0.5 (0.2)	0.4 (0.2)	0.5 (0.2)	0.6 (0.1)	0.6 (0.1)	0.6 (0.1)
농가율	0.5 (0.2)	0.4 (0.2)	0.5 (0.2)	0.6 (0.1)	0.5 (0.2)	0.4 (0.2)
전업농가율	0.7 (0.2)	0.6 (0.1)	0.7 (0.1)	0.7 (0.1)	0.6 (0.2)	0.5 (0.2)
2종겸업농가율	0.2 (0.1)	0.3 (0.1)	0.2 (0.1)	0.2 (0.1)	0.2 (0.1)	0.4 (0.2)
어가율	0.0 (0.1)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.2 (0.1)	0.3 (0.2)
전업어가율	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.1 (0.1)
2종겸업어가율	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.1 (0.1)	0.1 (0.1)
제조업종사자율	0.3 (0.2)	0.6 (0.1)	0.2 (0.2)	0.2 (0.2)	0.2 (0.2)	0.1 (0.1)
도소매종사자율	0.3 (0.1)	0.1 (0.1)	0.3 (0.1)	0.3 (0.1)	0.3 (0.1)	0.3 (0.2)
평야지대비율	0.4 (0.3)	0.5 (0.2)	0.6 (0.3)	0.2 (0.2)	0.3 (0.2)	0.0 (0.0)
산간지역비율	0.3 (0.3)	0.2 (0.2)	0.2 (0.2)	0.6 (0.2)	0.1 (0.1)	0.0 (0.0)
해안지역비율	0.1 (0.1)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.4 (0.2)	0.0 (0.1)
섬지역비율	0.0 (0.2)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.1)	1.0 (0.1)
수세식화장식비율	0.7 (0.2)	0.8 (0.1)	0.7 (0.2)	0.6 (0.2)	0.8 (0.1)	0.7 (0.1)
아파트수비율	0.1 (0.1)	0.2 (0.2)	0.1 (0.1)	0.0 (0.0)	0.0 (0.1)	0.0 (0.0)
노후주택수비율	0.3 (0.1)	0.2 (0.1)	0.3 (0.1)	0.4 (0.1)	0.3 (0.1)	0.5 (0.2)
전작생산가구비율	0.1 (0.1)	0.1 (0.1)	0.1 (0.0)	0.1 (0.1)	0.1 (0.1)	0.3 (0.2)

〈표 3.18〉 K-means 군집들에 대한 군집 특성 요약

특성/군집		K-means 방법 : 4개 그룹	
		군집1	군집2
요인특성		- 고령농촌형 지역	- 농어촌특성미약
변수 특성	인구	- 인구밀도 매우 낮음 - 노령지수/연평균인구 감소율 높음 - 농림어업인구 비율 매우 높음	- 인구밀도/연평균인구 증가율 매우 높음 - 노령지수 매우 낮음
	가구	- 농가율/전업농가율 높음	- 2종 겸업농가율 높음
	지형	- 평야지대율 보통 - 산간지대율 약간높음	- 특색없음
	산업	- 제조업, 도소매 보통	- 특색없음
	생활 여건	- 노후주택비율 높음	- 수세식/아파트수비율 매우 높음
지역 수		910	301

〈표 3.18〉 K-means 군집들에 대한 군집 특성 요약(계속)

특성/군집		K-means 방법 : 4개 그룹	
		군집3	군집4
요인특성		-섬/어업/농업/관광	-해안/농어업겸업/관광
변수 특성	인구	-노령지수/연평균인구감소율 매우 높음	-연평균인구감소율 매우 높음 -농림어업인구비율 높음
	가구	-겸업농가율/어가율 /전업어가율 매우 높음 -전작가구율 높음	-전업농가율 낮음 -겸업농가율/어가율 높음
	지형	-섬비율 100%	-평야/산간지대율 낮음 -해안지대율 매우 높음
	산업	-제조업종사비율 낮음	-제조업 낮음 -도소매숙박종사 높음
	생활 여건	-수세식화장실율/노후 주택비율 높음	-수세식화장실율 높음
지역 수		39	50

〈표 3.19〉 계층적 군집들에 대한 군집 특성 요약

특성/군집		계층적 방법 : 5개 그룹		
		군집1	군집2	군집3
요인특성		-비농업적 특성	-평야 고령 농촌	-산간형 고령 농촌
변수 특성	인구	-인구밀도/인구증가 감소율 낮음	-인구밀도 약간 높음	-인구밀도 매우 낮음 -노령화지수/연평균 인구감소율 매우 높음 -농림어업인구비율 높음
	가구	-겸업 농가율 높음	-특색없음	-농가율 높음
	지형	-평야지대율 약간 높음	-평야지대율 높음 -산간/해안비율 낮음	-산간지역 비율 매우 높음
	산업	-제조업종사율 매우 높음	-특색없음	-특색없음
	생활 여건	-수세식화장실/아파트 비율 높음 -노후주택율 낮음	-특색없음	-노후주택율 높음 -수세식화장실비율 낮음
지역 수		194	430	483

〈표 3.19〉 계층적 군집들에 대한 군집 특성 요약(계속)

특성/군집		계층적 방법 : 5개 그룹	
		군집4	군집5
요인특성		-해안 농어업겸업/관광	-섬 /전어업/관광
변수 특성	인구	-연평균 인구 감소율 매우 높음 -농림어업인구율 높음	-노령화지수/농림어업인구비율 높음 -연평균인구감소율 매우 높음
	가구	-어가율/겸업어가율 높음	-겸업농가율/어가율 /겸업어가율 매우 높음 -전작가구비율 매우 높음
	지형	-해안지역비율 매우 높음	-섬비율 100%
	산업	-특색없음	-제조업비율 낮음
	생활 여건	-수세식화장실비율 높음	-노후주택비율 매우 높음
지역 수		145	48

제 4절 결론 및 논의

지금까지 전체 3,573 동읍면 중심의 농어촌 지역을 대상으로 통계적 기법을 적용하여 유사한 특성을 갖는 소수개의 농어촌 그룹으로 분류해 보았다. 전체적인 연구의 흐름이나 분석 과정은 통계학적 이론 측면보다는 실무차원의 의견을 충분히 반영하고자 하였다.

전체 3,573 동읍면은 도시화율 기준에 의해 도시와 농촌을 구분하고, 최종적으로 행정지역 구분을 고려한 순수 농어촌 지역인 1,300 읍면을 분류대상으로 삼았다. 통계적 분류 방법으로 K-means 와 계층적 방법을 사용하였다. 군집 수 결정은 CCC와 pseudo F를 사용하였고, 이 기준에 의해 K-means와 계층적 방법은 각각 4개와 5개를 최적 군집 수로 선택하였다.

본 연구에서는 유형화 분석 결과를 토대로 몇 가지 사실을 찾을 수 있다.

첫째, 분석에 사용된 20개 변수 중에서 수세식화장실비율이나 농림어업인구비율은 농어촌 구분 변수로서 크게 의미가 없어 보인다. 수세식화장실비율의 경우, 현재 우리나라 농어촌의 수세식 화장실 비율이 높기 때문에 농어촌 구분 변수로서 크게 변별력이 없는 것으로 예상된다. 또한 현재 사용된 변수는 통계청에서 전수조사가 가능한 변수로만 한정했기 때문에 그 외의 농어촌 유형화에 필요한 중요한 변수들이 상당부분 배제되었을 가능성이 높다. 실제로, 농어촌의 경제적 수준을 측정할 수 있는 변수라든가 발전정도 면에서 전체 농어촌 평균 수준보다 낙후된 지역을 구분할 수 있는 변수들은 농어촌 유형 구분에 상당한 영향력을 미칠 수 있다. 통계학적 측면에서도 분석결과가 사용된 변수에 따라 다를 수 있다는 점을 감안하면 변수 선정이 유형화에 있어서 얼마나 중요한지를 짐작할 수 있다.

둘째, K-means 군집에서 군집은 지형적 요소에 크게 영향을 받는다는 것을 알 수 있다. 4개의 군집은 지형적 측면에서 보면, 평야산간형(내륙), 해안(내륙과 섬의 경계), 섬, 그리고 도시근교 또는 소재지 인접 지역 형태로 분류된다. 더구나 4개 군집에 속한 지역들이 매우 편중되

어 나타난다. 평야·산간지역의 고령 농촌형은 전체 1,300개 읍면지역 중 70%(910개)를 포함하고 있다. 이 유형이 우리나라 대부분의 농촌의 성격을 잘 반영한다고 짐작할 수 있다. [그림 4.1]에서 알 수 있듯이 K-means 군집들은 지형적 특성을 매우 강하게 반영하는 섬·해안과 내륙 형태로 확실하게 양분화 되는 경향이 있다. 따라서 내륙 특성을 강하게 지닌 전형적인 농촌지역이라 할 수 있는 910개 군집과 301개 군집에 대해 세분화하는 것이 오히려 농촌유형화에 유익한 작업이 될 것이다. 이는 910개 또는 301개 지역을 분류할 수 있는 새로운 변수 선정이 중요한 출발점이 되어야 한다. 순수 농촌의 특성을 변별할 수 있는 포괄적인 변수 선정만이 농촌이라는 유사한 특성을 지닌 지역들을 제대로 분류할 수 있을 것이다.

셋째, 계층적 방법도 K-means와 마찬가지로 지형적 요인이 5개 유형을 분류하는데 크게 작용하였다. K-means방법과 군집화 방법론적 측면에서 다르기 때문에 두 방법의 직접적인 비교는 의미가 없다. 단, 독자들의 이해를 돕기 위해서 굳이 비교 설명을 한다면, [그림 4.2]에 나타난 바와 같이 계층적 방법의 5개 군집은 K-means의 910개 유형이 다시 재분류 되는 효과를 갖는다고 볼 수 있다. 전형적인 농촌 유형이 평야지대 비율이 높거나 농촌 특성이 강한 유형과 산간지대 비율이 높거나 농촌 특성을 지닌 유형으로 구분된다. 그러나 두 유형에 대한 구분점이 명확하지 않기 때문에 이 유형 구분에 대해서는 신중하게 검토할 필요가 있다. 또한 해안·겸업 유형으로 분류된 지역들도 현재 사용된 변수 정보만으로는 그 경계점을 찾기가 어렵기 때문에 위험한 구분 결과를 가져올 수 있다는 것을 인지할 필요가 있다.

본 연구에서는 요인분석을 통해 얻어진 요인들을 군집변수의 입력변수로 사용한 만큼, 유형화에 유리한 정보, 즉 변수를 선정하는 것이 무엇보다 중요하다 하겠다. 이를 위해서는 우선 지역을 유형화하는 목적이 명확해야 하고, 전문가적 또는 실무적 차원에서 그 목적에 맞는 타당한 변수를 선정할 수 있어야 한다. 또한 통계적 군집방법을 사용하여 개체를 분류할 경우, 어떤 분류 방법을 사용하느냐는 전적으로 연구자의 몫이다. 그렇다고 전혀 기준이 없는 것은 아니다. 연구목적, 자료크기, 사용할 변수의 특성에 대한 정확한 이해는 통계 분류 방법 선택에 중요

한 요인이 된다.

결론적으로, 연구자의 판단은 방법론적 측면에서는 우리나라 농어촌 지역을 유형화 하는데 있어서 K-means방법이 계층적 방법보다 더 적합한 군집 결과를 준다고 볼 수 있다. 이론적 측면에서도 분류할 개체수가 많을 경우에는 계층적 방법보다 분할적 방법이 더 유리한 것으로 알려져 있다. 또한 분석에 사용된 변수가 다소 제한적이고, 특정 지표(지형, 인구 등)에 치우쳐 있다는 것을 감안하면 매우 타당한 결과라고 판단된다.

한편, 본 연구에 사용된 변수만을 활용하여 지역을 더 세분화하고자 하는 것에 대해서는 좀더 신중할 필요가 있다. 우선 K-means 방법으로 4개 이상의 군집으로 분류하는 것은 통계적 측면에서 바람직하지 않다. 또한 계층적 방법으로 세분화하고자 할 경우, 세분화는 가능하지만 주어진 변수의 정보가 매우 제한적이기 때문에 군집결과를 해석하는데 매우 신중해야 한다. 특히, 계층적 방법은 한번 어떤 군집으로 분류가 되면 절대 바뀔 수 없다는 원리를 사용하기 때문에 많은 개체를 분류하고자 할 경우 상당히 불리하게 작용할 수 있다. 분석에 사용된 변수가 지형적 특성과 가구 특성이 매우 강하기 때문에 주어진 변수로는 최소한의 유형화를 시도하는 것이 바람직할 것으로 생각된다.

더 세분화가 필요할 경우, 1차 유형화 결과를 토대로 2차적으로 유형화에 적합한 농어촌 특성들을 활용하여 재분류하는 것도 하나의 대안이 될 수 있겠다. 그렇지 않으면 통계적 방법이 아닌, 농어촌 유형에 필요한 객관적 기준을 마련하여 몇 개의 유형으로 구분하는 것도 방법이 될 수 있을 것이다. 객관적 기준에 의해 유형을 구분하고, 구분된 유형은 구체적 속성은 통계적 요인분석 등을 통해 파악하는 것도 좋은 대안이 될 수 있다. 마지막으로 지역유형 결과가 국가 정책의 기반이 되거나 장기적으로 활용될 가능성이 있다면, 농어촌 유형화에 사용된 변수들은 시간에 따라 가변적일 수 있기 때문에 농어촌 지역 유형화는 최소한의 시간 간격을 두고 새로운 기준에 의해 재분석이 이루어져야 할 것이다.

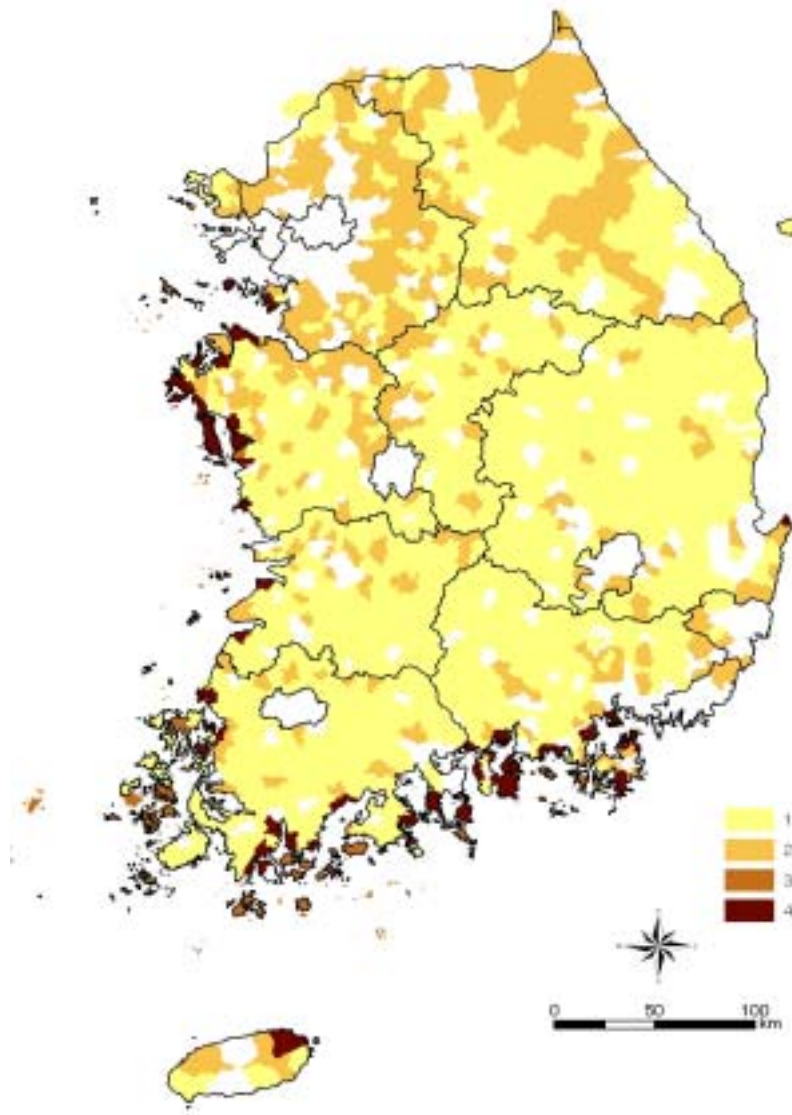
이에 본 연구는 유형화할 대상이 많고, 변수선정 및 통계분석 방법 선택 기준이 없는 경우에 좋은 참고자료가 될 수 있을 것으로 기대한다.

참고문헌

- [1] Alasia, A. (2004). Mapping the socio-economic diversity of rural Canada: A multivariate analysis, Agriculture and Rural Working Paper Series Working Paper No. 67.
- [2] Milligan, G.W. and Cooper, M.C. (1985). An examination of procedures for determining the number of clusters in a data set, Psychometrika, 50, 159-179.
- [3] Sarle, W.S. (1983). Cubic clustering criterion, SAS Technical Report A-108, Cary, NC: SAS Institute Inc.
- [4] 이원섭 (2003). 국가자원의 지역간 차등화를 위한 지역유형 구분 방안에 관한 연구, 국토연구 36권, 19-35.
- [5] 임석희(2005). 농촌지역의 유형화와 특성 분석, 한국지역지리학회지 11권 2호, pp. 211-232.
- [6] 성주인, 송미령(2003). 지역유형 구분과 농촌지역의 유형별 특성, 농촌경제 26권, 1-22.
- [7] 통계청 (2007). 농어촌 지표개발 추진 결과.

부록 1. K-means 결과 mapping : 4개 군집

군집	개수	군집 특성
1(노랑)	910	전형적 고령 농촌 지역
2	301	도시근교/소재지 인근 지역 또는 발달된 농촌 지역
3	39	섬지역/전어업 비율이 높은 지역
4(갈색)	50	해안지역/겸업비율이 높은 지역



부록 2. 계층적 군집 결과 mapping: 5개 군집

군집	개수	군집 특성
1(노랑)	194	비농업특성이 강한 지역 또는 도시근교/소재지 인근
2	430	평야지역 또는 고령 농촌 특성이 강한 지역
3	483	산간지대 및 고령 농촌이 강한 지역
4	145	해안지역/겸업 비율이 높거나 서비스업 발달 지역
5(진갈)	48	섬지역의 전어업 비율이 높은 지역

