

# 단계구분도의 계급구간 설정에 관해

손 일

경상대학교 사범대학 사회교육학부 교수

## 1. 단계구분도에 대한 개관

면단위로 통합된 자료를 여러 계급으로 나누고, 부여된 계급에 따라 해당되는 흑백 또는 컬러 기호로 각 면단위를 표시해 전체 지역의 공간적 패턴을 알아보는 것이 단계구분도의 기본적인 역할이다. 아래 식은  $N$ 개의 면단위와  $R$ 개의 계급수라는 조건하에서 만들어질 수 있는 단계구분도의 수를 계산하는 식이다. 실제로 200개의 면단위를 6개 계급으로 나눌 경우 2,472,258,789개의 단계구분도가 만들어질 수 있다. 따라서 어떤 특정 현상을 설명하기 위해 지리학 관련 문헌에 제시된 단계구분도는 이처럼 무수한 단계구분도 중에서 하나이므로, 그것을 선택함에 있어 지도제작자의 주관은 결코 무시할 수 없다.

$${}^{(N-1)}C_R = \frac{(N-1)!}{(N-1-R)! \cdot R!}$$

동일한 자료로부터 전혀 다른 설명이 가능한 지도가 만들어질 수 있기 때문에, 단계구분도의 기법이 이용되기 시작한 19세기초부터 단계구분도의 객관성을 확보하기 위한 여러 가지 방법이 도입되었다. 그 결과 이에 대한 연구도 60년대 이후 활발히 진행되었으며, 단계구분도 연구에 있어 가장 대표되는 학자로는 George F. Jenks을 들 수 있다. 계급수와 계급구간 설정 방법 이외에도 변수의 특성, 면단위의 수·크기·위치, 면단위의 범례 설정 등 고려해야 할 사항들이 무수히 많다. 따라서 단계구분도의 객관성을 확보하는 일은 대단히 중요하다고 판단된다.

본 발표에서는 계급구간 설정에 있어 통계적 인접성과 공간적 인접성이라는 두 가지 기준에 대해 간략히 살펴 본 후, 다양한 평가방법과 평가기준을 이용해서 특정 사례에 대한 단계구분도의 객관성에 대해 실험해 보았다. 또한 시각적 인지 한계를 기준으로 제시된 기존의 계급수 설정 기준과는 달리 보다 객관적으로 계급의 수를 선정할 수 있는 한 가지 방안을 제시해 보고, 마지막으로 단계구분도 제작시 적용할 수 있는 면기호의 범례 설정 특히 Color scheme의 설정 기준에 대한 예를 정리해 보았다.

## 2. 계급구간의 설정

계급구간 설정은 계급수의 결정과 함께 단계구분도 제작에 있어 핵심적인 요소인 동시에 지도제작자의 주관에 개입할 소지가 많다. 계급구간을 설정함에 있어 일반적으로 두 가지 기준을 적용하는데, 통계적 인접성(Statistical proximity)과 공간적 인접성(spatial proximity)이 그것이다.

통계적 인접성을 기준으로 계급구간을 설정할 경우 주로 면단위 자료 분포의 수치적, 통계적 특성에 따라 급간을 설정하는데, 일반적으로 제시되는 계급구간 설정 방법은 대부분 여기에 속한다.

등간격, 사분위, 누진법, 평균·분산, 시각적 구분법, 최적화 등이 대표적인 방법이며, 통계적 인접성을 기준으로 하는 계급구간 설정 방법에 대한 분류는 Evans(1977)에 잘 정리되어 있다.

한편, 공간적 인접성을 기준으로 계급구간을 설정할 경우, 자료 자체의 수치적 분포를 인정하면서도 가능한 한 이웃한 면단위가 유사한 통계치를 갖고 있다면 하나의 지도지역(map region)으로 통합시키는 방법이다. 이 경우 분할된 지도지역(map region)의 수, Jenks & Caspall(1971)의 BEI 지수, MacEachren(1982)의 map Complexity 지수 등이 평가 기준이 될 수 있다. 하지만 너무 공간적 인접성만을 강조할 경우 F-검정치가 감소하거나, 계급구간이 겹칠 수 있으며, 고려해야 할 변수가 너무 많아 최적화 프로그램을 만들기가 어렵다는 단점도 있다.

### 3. 최적화 기법(optimization)

통계적 인접성을 기준으로 계급구간을 설정할 경우, 지리학에서 중요시하는 면단위간의 공간적 인접성이나 연계와 같은 변수는 무시되고 단지 개별 면단위를 정해진 계급구간에 편입시킨다(typological regionalization). 실제로 자료의 분포 특성에 따라 적절한 설정 방법을 이용한다지만(예를 들어, 정규분포를 하고 있는 자료의 경우 평균·분산 방법을 이용), 결국 각각의 설정 방법에 대한 객관적 평가는 “계급구간내 동질성을 최대화하고, 계급구간 간에는 이질성을 최대화하는” 기준, 다시 말해 F-검정치나 GVF 값과 같은 통계치에 의존할 수밖에 없다.

따라서 Jenks & Caspall(1971)의 연구를 필두로 최적화 기법에 관한 연구가 진행되었다. 여기서 최적화란 주어진 면단위의 수와 계급수에서 만들어질 수 있는 모든 경우에 대하여 F-검정치 혹은 GVF 값을 계산한 후, 그 값이 최대가 되는 계급구간을 찾는 것을 말한다. Jenks & Caspall(1971)의 최적화 기법에서는 임의의 초기 계급구간을 설정해 주고 iteration의 수를 한정시켰기 때문에, global optimum 보다는 local optimum에 빠질 위험이 있다. 그후에 개발된 Jenks(1977)와 Lindberg(1990)의 기법에서는 Fisher(1958)의 알고리즘을 이용해 최적화를 시도하였다.

본인은 Lindberg의 최적화 프로그램의 결과가 모든 경우를 계산한 결과와 일치하는 지를 판단하기 위해, 아래와 같은 전수 계산 알고리즘을 이용해 그 결과를 비교해 보았다. 그 결과는 정확히 일치하였고 면단위의 수가 50개를 넘지 않을 경우 계산 시간은 짧아 큰 문제가 되지 않았으나, 그 이상인 경우 Lindberg의 프로그램은 전수를 계산하지 않고 최대값을 추적하기 때문에 계산 속도에 있어 본인의 것을 훨씬 능가하였다.

```
DO 100 I1=1,N-5
  T1=I1+1
DO 100 I2=T1,N-4
  T2=I2+1
DO 100 I3=T2,N-3
  T3=I3+1
DO 100 I4=T3,N-2
  T4=I4+1
DO 100 I5=T4,N-1
100 CONTINUE
```

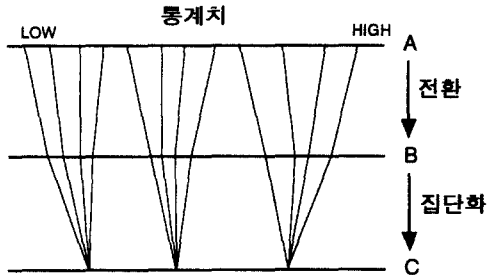
\* 이 전수 계산 알고리즘은 계급 수가 6개이고  
여기서 N는 면단위의 수를 의미한다

### 4. 공간적 인접성(Spatial proximity)을 기준으로 하는 계급구간 설정

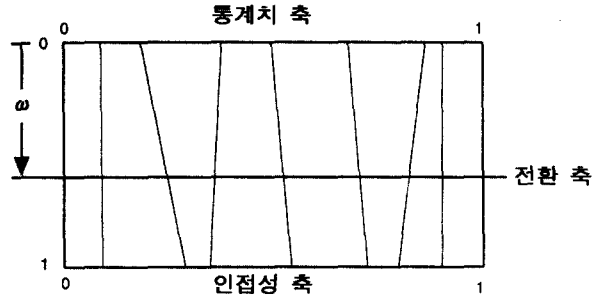
이 방법에서는 비공간적 과정을 거쳐 자료를 분류하는 최적화를 비롯한 여러 가지 통계적 방법과는 달리 면단위의 인접성과 연계성을 강조한다. 다시 말해 특정 현상의 유무 혹은 수치적 유사성에 따라 인접한 면단위를 하나로 묶는다(individual regionalization). 대표적인 것으로는 Monmonier(1968)의 “continuity-biased” 알고리즘과 Johnston(1968)의 “linkage method”가 있다.

Monmonier는 자료 크기에 따라 정리한 통계치를 해당 면단위 각각의 인접성을 기준으로 재배열한 후(그림 1의 1단계 전환), 여러 가지 방법으로 급간구간을 설정하였다(그림 1의 2단계 집단화). 그림 2는 1단계 전환 과정을 보다 상세히 설명한 것으로 공간적 인접성의 강조 정도에 따라  $\omega$ 의 값이 달라지고, 그 결과 급간구간 설정에 이용될 continuity-biased된 자료가 달라진다. 결국 여기에 지도제작자의 주관성이 개입될 수밖에 없고, 급간구간 설정 방법의 선택에서도 마찬가지이다.

<그림 1>



<그림 2>



실제 Monmonier의 연구 결과에 의하면  $\omega$  값을 증가시켜 공간적 인접성을 강조하더라도 그가 선택한 지도 효율성의 기준인 구획된 지도지역의 수는 그다지 감소하지 않았다. 또한 자료가 일정한 경향을 가진 채 분포하고 있을 경우 F-검정치가 조금 감소했으나, 자료가 불규칙하게 분포한 경우에는 F-검정치가 크게 줄어드는 결과를 초래하였다. 다시 말해 패턴의 단순성(pattern simplicity)과 집단내 등질성(within-group homogeneity) 사이에 trade-off 관계가 나타나 이 역시 지도제작자의 주관성이 개입될 여지가 있다.

## 5. 실험

실험에 이용되는 지도는 수도권 53개 시·군에 대한 1995년도 인구밀도를 나타낸 단계구분도이다. natural break, 등간격, n-tile, 최소 TEI, 최소 BEI, 최소 OEI 등, 모두 6가지 계급구간 설정 방법 각각에 대해 3~8개 계급 수의 단계구분도를 작성하여 총 36개의 단계구분도를 실험 자료로 이용하였다.

앞선 밝힌 전수 계산 알고리즘을 이용해 F-검정치 혹은 Jenks의 GVF 값이 최소가 되는 계급구간을 설정하는 것과 마찬가지로, 최소 BEI, 최소 OEI, 최소 지도지역이 나타나는 계급구간을 설정하기 위한 최적화 프로그램을 작성하였다. 단계구분도의 오차를 측정하는 기준으로는 Jenks & Caspall(1971)의 OEI, TEI, BEI와 MacEachren(1982)의 CF, CV, CL 이외에도 여러 다른 기준들이 있으나, 나머지는 이상의 6개 기준과 대동소이하므로 평가 기준을 6개로 한정하였다. 여기서 OEI와 TEI는 집단내 등질성을 추구하는 통계적 기준이고, BEI는 이웃한 면단위간의 통계치 차이가 큰 경계값이 지도지역을 구획하는 경계로 사용되는 정도를 나타내는 지수이며, CF, CV, CL는 지도 패턴의 단순성을 측정하는 지수이다.

지면 관계로 OEI, TEI, BEI, CF, CV, CL의 계산 방법(표 1) 및 실험 결과(표 2)는 발표 당일에 배포할 예정이다.

## 6. 계급수 결정에 대한 한 가지 방안

일반적으로 시각적 인지 능력의 한계 때문에 계급수는 몇 개 이상이 되어서는 곤란하다는 최대 한의 기준만을 제시하고 있다(예를 들어 흑백의 경우 5-7개, 색상과 명도를 이용한 경우 최대 9개). <그림 3>은 각 계급수에 대한 최적화의 결과 계급수의 증가에 따른 GVF값의 변화를 나타내고 있다. 계급수의 증가에 따라 GVF 값의 증가 정도가 급격히 감소하므로, 지도제작자는 이러한 그래프를 통해 적정 계급수를 판단할 수 있다. 한편 중학교 남녀 1학년생 344명에 대해 단계구분도의 인지특성을 연구한 문미옥(1998)에 의하면 계급수가 3개에서 5개로 증가함에 따라 인지 및 회상 능력이 급감하였다. 따라서 단계구분도의 계급수를 설정함에 있어 연령, 학력 등 지도이용자의 능력도 고려해야 한다.

## 7. 면기호의 범례 설정

### 1) Gray scale

Gray scale의 경우 계급수에 따라 등간격으로 인지되는 회색조의 범례를 만드는 것이 기본으로, 이때 두 가지 사항을 고려해야 한다. 첫째, 물리적으로 측정된 회색조의 명도와 시각적으로 인지된 회색조의 명도간의 비선형 관계를 밝히는 작업이 그 하나이고, 인쇄시 잉크과잉이나 잉크 부족에 따른 명도 변화를 고려해 회색조의 범례가 등간격으로 인지될 수 있도록 적절한 스크린톤을 선택하는 일이다. 이에 대한 기존의 연구로는 Jenks & Knos(1961), Kimerling(1975), Stevens((1975), Monmonier(1980), Williams(1982) 등을 들 수 있으며, Kimerling(1985)에 이들 연구 결과가 잘 요약되어 있다.

### 2) Color scheme

Color scheme의 경우 Gray scale과는 달리 두 가지 이상의 변수를 한꺼번에 표현할 수 있고 변수 특성을 고려해 다양한 기법의 색채이용이 가능하기 때문에 보다 세련된 방식으로 단계구분도를 제작할 수 있다. Brewer(1994)는 변수의 수, 정성적 변수와 정량적 변수, binary, diverging, sequential과 같은 변수의 특성을 조합한 Color scheme 목록을 만들고, 각각에 대해 실제 예를 들어 설명하고 있다. 이와 유사한 연구로는 Olson(1987), Mersey(1990) 등이 있으며, Slocum(1999)의 경우 Brewer의 Color scheme 각각에 대해 인쇄시 적용할 수 있는 CMYK 비율을 제시하고 있다.

## 참고문헌

Jenks, G.F. and Caspall, F.C., 1971, Error on choroplethic maps: Definition, measurement, reducton, *AAAG*, 61(2), 217-244.