군집분석

# 군집분석

## 1. 군집분석, 요인분석, 판별분석의 비교

* 요인분석 : 변인(평가항목)들을 요인에 따라 구분한다.
* 군집분석 : 대상의 유사한 특성을 토대로 그룹으로 구분한다.
* 판별분석 : 집단이 이미 나뉘어 있을 때, 집단들 간의 차이점을 분석한다.

|  |  |  |  |
| --- | --- | --- | --- |
| 종류 | 대상의 중복성 | 자료의 크기 | 설명 |
| 계층적 군집분석 | 없음 | 작음 | 대상 간 거리에 의해 군집을 형성한다. |
| 비계층적 군집분석 | 제한없음 | 군집의 수를 정하고, 군집의 중심으로부터 가까운 객체를 포함하면서 군집을 형성한다. |
| 중복 군집분석 | 있음 | 대상 간 상이한 군집 규칙을 적용하여 하나의 대상이 여러 군집에 포함될 수 있으나, 자료의 양이 많아 지면 복잡해질 수 있다. |

우리는 요인분석의 개념을 살펴보고, 분석을 진행하는 방법과 이를 해석하는 과정을 살펴보았다. 군집분석 역시 변수들을 내부 동질적이며 외부 이질적인 집단으로 구분한다는 점에서 요인분석과 같지만, 두 방법 간에는 명확한 차이가 있다.

요인분석인 R-type 요인분석과 Q-type 요인분석으로 나뉜다. R-type 요인분석은 평가항목을 동질적인 몇 개(요인의수)의 집단으로 만드는 반면, 군집분석과 유사한 Q-type 요인분석은 대상(응답자)들을 몇 개의 동질적인 집단으로 만든다. Q-type 요인분석은 케이스별로 상이한 특성을 가지는 개별 응답자들을 상호 동질적인 몇 개의 집단으로 구분하는 것이어서 계산하기가 복잡하다. 그래서 군집분석을 대안으로 활용한다.

다수의 대상에 대해 그들이 가지는 특성을 토대로 대상별 그룹을 나누고, 각 그룹을 ‘군집’이라 부르는 통계기법을 군집분석(clustering analysis)이라 한다. 군집분석은 변수에 대해 대상들의 점수를 측정하고 두 점의 직선거리를 뜻하는 유클리드 거리(Euclidean distance)를 기준으로 유사성을 측정한다.

## 2. 연구문제

사람마다 취향이 있듯이, 스마트폰 이용자마다(군집) 스마트폰을 구매할 때 고려하는 특성에 차이가 있다고 가정해 볼 수 있다. 이 가정을 확인하기 위해 스마트폰 이용자들을 대상으로 설문을 실시하고, 그 결과를 바탕으로 군집분석을 실시해 보자.

일반적으로 연구보고서나 논문에 쓰이는 자료(표본)의 크기는 보통 100개를 상회하는 경우가 많다. 군집분석을 실시하는데, 군집의 개수를 알지 못할 때는 먼저 ‘계층적 군집분석’을 통해 군집의 수를 미리 추출해 본다. 군집의 개수를 알지 못할 때는 먼저 ‘ 계층적 군집분석’을 통해 군집의 수를 미리 추출해 본다. 그 다음으로 연구자가 적절한 군집의 수를 선택한 후, 다시 ‘비계층적 군집분석’을 실시하여 군집의 수를 지정하면 된다.

우리는 표본의 개수를 325개를 설정하여 군집분석을 실시하였다.