SPSS에서 요인분석: 주성분 분석과 최대우도법

# 1. 주성분 분석 (Principal Component Analysis, PCA)

주성분 분석(PCA)은 요인분석의 일종으로, 데이터 내에서 가장 많은 분산을 설명할 수 있는 주성분(Principal Components)을 추출하는 방법입니다. PCA는 데이터의 차원을 축소하거나, 데이터의 구조를 파악하는 데 유용한 기법으로 사용됩니다.

## PCA의 주요 특징

- 분산 극대화: PCA는 각 성분이 가능한 한 많은 분산을 설명하도록 주성분을 추출합니다. 첫 번째 주성분은 가장 많은 분산을 설명하고, 두 번째 주성분은 남은 분산 중에서 최대한 많이 설명하도록 추출됩니다.

- 상관행렬/공분산행렬: PCA는 상관행렬 또는 공분산행렬을 기반으로 수행됩니다. 상관행렬을 사용하는 경우, 모든 변수는 동일한 중요도를 가지며, 공분산행렬을 사용하는 경우 각 변수의 분산 크기에 따라 가중치가 다르게 적용됩니다.

- 요인의 해석: PCA에서 추출된 요인은 원래의 변수들의 선형 결합으로 해석됩니다. 이러한 주성분은 일반적으로 실제 개념보다는 수학적 계산의 결과로 나타나며, 해석이 어려울 수 있습니다.

- 고유값(Eigenvalue): 각 주성분은 고유값을 가지며, 고유값이 큰 주성분은 데이터의 많은 부분을 설명합니다. 통상적으로 고유값이 1 이상인 주성분만을 요인으로 채택합니다.

- 데이터 축소: PCA는 고차원 데이터를 더 적은 차원으로 축소하는 데 주로 사용됩니다. 이를 통해 데이터의 주요 구조를 파악하면서도 노이즈나 불필요한 변수를 제거할 수 있습니다.

## PCA의 사용 목적

- 데이터의 차원 축소: 많은 변수들로 이루어진 데이터를 몇 개의 주성분으로 축소하여, 데이터의 핵심 정보를 보존하면서 분석을 간소화합니다.

- 데이터 시각화: 축소된 차원으로 데이터를 시각화하여, 변수 간의 관계를 더 명확하게 파악할 수 있습니다.

- 데이터 전처리: 머신러닝 모델에 데이터를 입력하기 전에 변수의 차원을 줄여 모델의 성능을 개선하거나, 과적합을 방지하는 데 사용됩니다.

# 2. 최대우도법 (Maximum Likelihood, ML)

최대우도법(ML)은 통계적 방법론에 기반한 요인분석 방법으로, 데이터가 특정 분포(주로 다변량 정규분포)를 따른다는 가정 하에, 관찰된 데이터가 주어진 모델에서 나올 확률(우도, Likelihood)을 최대화하는 요인을 추출합니다.

## ML의 주요 특징

- 통계적 추론: ML은 관찰된 데이터의 분포를 기반으로 모델을 구축하며, 이 모델이 데이터에 가장 적합하도록 요인들을 추출합니다. 이는 모델이 주어진 데이터를 가장 잘 설명할 수 있도록 요인을 추정하는 과정입니다.

- 가설 검정 가능: ML은 통계적 검정을 통해 모델의 적합성을 평가할 수 있습니다. 예를 들어, 요인의 개수를 결정할 때 카이제곱 검정 등을 사용하여 모델의 적합성을 검토할 수 있습니다.

- 공통성 추정: ML은 변수들 간의 공통 요인을 추정하며, 각 변수의 공통성을 직접 추정할 수 있습니다. 이로 인해 각 변수의 요인 부하량을 보다 정확하게 파악할 수 있습니다.

- 모수 추정: ML은 요인 부하량, 공통성, 요인 간 상관관계 등의 모수를 추정하는 데 사용되며, 이러한 추정값은 표준 오차와 함께 제공되어 추정의 신뢰성을 평가할 수 있습니다.

- 요인의 해석: ML로 추출된 요인은 해석 가능한 실제 개념일 수 있으며, 요인 간의 관계를 통계적으로 검정할 수 있습니다.

## ML의 사용 목적

- 통계적 가설 검정: 요인 수나 모델의 적합성을 검정하기 위해 ML을 사용합니다. 이는 요인분석 결과를 보다 객관적으로 평가할 수 있는 방법을 제공합니다.

- 공통성 추정: 변수들 간의 공통 요인을 추정하여, 변수들이 얼마나 많은 공통적인 정보를 공유하는지를 평가합니다.

- 모델 적합성 평가: 최대우도법을 통해 추정된 모델이 실제 데이터에 얼마나 적합한지를 통계적으로 평가할 수 있습니다.

# 요약

주성분 분석(PCA)은 데이터의 분산을 최대한 설명하는 주성분을 추출하여, 데이터를 축소하거나 구조를 파악하는 데 유용합니다. 이는 데이터의 차원을 축소하는 데 주로 사용되며, 요인 해석이 어렵지만 수학적으로 강력한 방법입니다. 최대우도법(ML)은 통계적 방법론을 기반으로 데이터의 분포에 적합한 요인을 추출합니다. ML은 통계적 검정과 공통성 추정이 가능하며, 요인 해석과 모델 적합성 평가에 강점을 가지고 있습니다.