

# 요인분석 (부교재 171쪽)

정의: 무수히 많은 변수 혹은 측정들을 상관관계의 원리를 이용해서 공통된 것을 묶어 인자를 추출해내는 분석방법.

예 1) 지능지수(IQ): 측정치 100개항목이 필요하다고 전제할 때 이 중에서 서로 같은 변수를 측정하는 경우가 많을 것임. -> 그러므로 기본적으로 특별한 몇 개의 질문(변수)이 없을까?

예 2) 참고서(179쪽, 4판은 142쪽)의 카탈로그의 특성= X1, X2, X3, X4, X5, X6, X7, X8, X9으로 측정 후 요인분석을 해서 **위험성**(X1, X2, X3), **경제성**(X4, X5, X6), **편리성**(X7, X8, X9)이라는 3개의 요인으로 축소하면 개념이해에 도움이 될 것임. (factor1.sav)

- X1: 카탈로그는 우송 시 상품의 분실가능성이 적음
- X2: 카탈로그는 이용 시 구매자와 제품에 관한 논의가 불필요
- X3: 견본과 똑같은 상품이 배달됨.
- X4: 카탈로그 이용 시 구매시간을 절약함.
- X5: 가격협상이 어렵지 않음
- X6: 구매 시 주문서에 기입하기가 용이함
- X7: 카탈로그의 상품가격은 저렴함
- X8: 집까지 배달됨
- X9: 판매원에 대한 지불이 불필요하기 때문에 가격이 저렴함

## <요인분석의 목적>

1. 변수의 수를 **축소**(variable reduction)해서 내용을 **단순화**시킨다.(개념이해와 설명용이)
2. 불필요한 변수(**중요도가 낮은 변수**)를 **제거**한다.
3. 변수들의 **특성을 파악**한다.(공통된 특성의 변수를 묶어 적절한 요인의 이름을 붙인다)
4. 측정항목의 **타당성(validity)**을 **평가**할 수 있다.

## <타당성의 유형>

1. **내용**타당성: 연구자가 **측정하고자 하는 내용(개념)**을 실제로 **측정**하였는가? 하는 문제.
2. **예측**타당성: 어떤 현상에 대한 측정(**현재**)이 미래의 또 다른 측정(**기대, 예측**)과 **관련성**이 높은가? 하는 문제.  
예) 어떤 제품의 구매의도가 높은 사람(현재응답)이 실제 구매가 높을 것임(미래의 측정)
3. **집중(수렴)**타당성: 하나의 개념을 측정하기 위한 여러 다른 측정방법을 사용했을 경우 그러한 측정치들 간에 **상관관계가 높은가?** 하는 문제(예, 전화면접방법 vs. 실제면접방법)
4. **판별**타당성: 서로 다른 개념을 측정하기 위해 사용한 측정치들 간에 **상관관계가 매우 낮은가?** 하는 문제. (예, 판매촉진 측정치 vs. 제품품질 측정치)

## <요인분석의 4단계 과정>

1. **상관관계 검토**: 변수들 간에는 적어도 하나 이상의 다른 변수와 **높은 상관관계**가 있어야 공통요인을 구할 수 있다.
2. **요인추출**: **주성분분석(PCA)**를 많이 사용한다. PCA사용이유: **정보의 손실을 최소화, 적은 수의 요인추출**에 유용.  
<요인 추출 시 공통요인 결정방법>
  - a) **고유치(eigenvalue) >1** (어떤 요인이라도 적어도 단일변수의 분산을 설명해야 함.)
  - b) **설명할 수 있는 분산의 비율** 고려(사회과학에선 **60%** 정도- Bartlette의 권고사항)
  - c) **스크리 도표 검토**(Scree test, 부교재 188쪽, 4판 150쪽): 추출된 요인의 도표를 기초로 곡선의 모양을 보고 **cut off의 점**을 정한다. -> **직선이 시작되는 첫 점**까지 요인의 수로 정함.

3. **요인회전**: 초기의 행렬을 해석하기 쉬운 행렬로 회전시킬 필요가 있음.(단순구조로 변화)
  - a) **배리맥스법**(varimax, **직교회전**): 각 요인이 다른 요인과 독립적(**요인 간 상관관계 0**)임을 전제. - 요인 간의 구분을 명확히 해서 해석이 쉽다.
  - b) **사교회전**(oblique rotation): **요인 간에 독립적이지 않**다는 것을 전제로 함.
4. **요인점수**: 여러 많은 변수 대신에 축소된 요인을 활용하여 추 후 **회귀분석**(regression)에 활용 할 수 있다.

<통계 용어>

**요인 적재치(값)**: 각 변수와 요인 간의 상관관계의 정도를 나타내는 지수

**공통성**: 추출된 요인들에 의해 설명되는 특정변수의 분산비율(변수에 대한 모든 요인 적재치를 제곱해서 합한 값) -일반적으로 공통성이 0.4 이하이면 낮다고 평가하며, 공통성이 낮은 변수는 요인분석에서 제외하는 것이 바람직하다.

**고유치(값)**: 특정 요인에 적재된 모든 변수의 적재치를 제곱하여 합한 값  
 요인추출 기준으로 지정한 고유값 1 이상인 요인만 추출됨.

**KMO**(Kaiser-Meyer-Oklin): 샘플링의 적합도를 측정하는 지수로 변수 쌍들 간의 상관관계가 다른 변수에 의해 잘 설명되는 정도를 나타냄. 값이 적으면 요인분석을 위한 변수선택이 좋지 못함을 의미한다. 1.0에 가까우면 매우 적합함을 의미함. 0.9: 상당히 좋음; 0.8이상 꽤 좋음 0.7이상은 적당함; 0.6이상은 평범함; 0.5이상은 바람직하지 못함;0.5미만은 받아들일 수 없음

**Bartlett의 구형성 검정치**(sphericity)는 '상관행렬이 단위행렬이다' (즉 변수들 간의 상관관계의 행렬표에서 대각선=1, 기타(off-diagonal=0)이라는 귀무가설(Ho)을 검정하기 위한 지수임. 그 값이 커서 유의확률(p<0.01, 혹은 p<0.05) 범위 안에 있으면(매우 작으면) 귀무가설(Ho)이 기각됨. 즉, 요인분석을 활용하는 것이 적합하며 공통요인이 존재한다라고 해석할 수 있음.

추출된 각 요인의 설명력(% 분산)= 요인(성분)의 [고유값/모든 변수의 수] x 100

이 값이 클수록 중요한 요인임.

추출된 모든 요인의 설명력은 위의 식에서 구한 각 요인의 설명력(%분산)을 모두 더한 값임