

탐색적 요인분석 방법

□ 요인분석의 목적

- 변수 또는 문항 간의 유사성(similarity)을 근거로 요인 또는 개념(construct)를 탐색적으로 추출하는 과정
- 요인구조의 타당도(타당성, validity)을 검증하는 방법에는 크게 내용타당도(content validity), 준거타당도(criterion-related validity), 구성타당도(개념타당도, construct validity)가 존재함. 일반적으로 탐색적 요인분석은 내용타당도와 구성타당도 측면에서 요인구조의 타당도를 입증하고자 하는 것임.

□ 탐색적 요인분석 프로세스

➤ 탐색적 요인분석상에서의 적정 표본의 크기

- 탐색적 요인분석 상에서 적절 표본의 크기는 크게 관점이 2가지가 있다. 첫 번째는 절대적 표본크기에 대한 기준이며, 두 번째는 조사대상자의 수와 측정변수의 비율에 근거한 기준이다.
- 첫 번째 기준의 경우, 절대적 표본의 크기를 기준으로 논의하였음.
 - Hair, Anderson, Tatham과 Black(1995)은 표본크기가 100보다는 적어도 커야 한다고 제안하였고,
 - Guilford(1956)는 적어도 200개의 표본은 되어야 한다고 했으며,
 - Tabachnick와 Fidell(2007)은 적어도 300개의 표본이 필요하다고 제안함.
 - Comrey와 Lee(1992)를 인용한 여러 문헌에서는 표본크기가 100이면 나쁜 편이며, 200개 정도는 괜찮은 편이고, 300개 정도는 좋은 편이며, 500개 정도는 아주 좋은 편, 1,000개 이상은 훌륭하다고 서술하고 있음.
- 두 번째 기준의 경우, 조사대상자의 수와 측정변수의 비율에 근거한 이론들이 있음.
 - 20:1(Hair et al., 1995),
 - 10:1(Nunnally, 1978),
 - 5:1(Tabachnick & Fidel, 2007),
 - 2:1(Guilford, 1956)이상이면 된다고 다양하게 제안하고 있음.

- Tak(2007)은 조사대상자 수가 적어도 200개 이상이거나 **대상자 수와 측정변수의 비율이 5:1이상이면 안정권이라고 하였음**(이 부분은 상황에 맞게 적용하십시오.)

→ 참고문헌

- Comrey,A.I., & Lee,H.B.(1992). A first course in factor analysis(2nd edition). Hillsdale, NJ: Erlbaum.
- Guilford,J.P.(1956). Psychometric method(2nd edition). New York, NY: McGraw-Hill.
- Hair,J.E Jr., Anderson,R.E., Tatham,R.L., & Black,W.C.(1995). Multivariate data analysis(4th edition). Upper Saddle River, NJ: Prentice Hall.
- Nunnally,J.C.(1978). Psychometric theory(2nd edition). New York, NY: McGraw-Hill.
- Tabachnick,B.G., & Fidell,L.S.(2007). Using multivariate statistics(5th edition). Boston, MA: Pearson/ Allyn & Bacon.
- Tak,J.K.(2007). Psychological testing: An understanding of development and evaluation method(2nd edition). Seoul: Hakjisa Publisher.

➤ **요인추출방법**

→ 주성분법(principal component method)

- 초기 문항개발 단계에서 선형적으로 진행하는 요인추출법
- 오차항을 가정하지 않는 수학적 모형으로 요인을 추출함.
- Principal component analysis

→ 공통요인법(common factor method)

- 이미 개발된 요인구조의 타당성에 대해서 탐색적 요인분석을 다시 시도할 때 적용을 많이 하는 방법
- 오차항을 가정한 확률적 모형으로 요인을 추출함.
- 주축분해법(principal axis factoring), 최대우도법(maximum likelihood method) 등등

➤ **요인의 개수 선정**

→ Kaiser's Rule(eigenvalues greater than one rule: Kaiser, 1960): 각 요인의 고유치 1이상인 곳에서 요인의 개수를 선정하는 방법

- 단점: 요인의 개수가 과대 추정되는 위험요소가 있음
 - 장점: 가장 일반적인 방법으로 논의되고 있음
- 연구자의 주관에 따라 선정할 수 있음
- 조건: 1개의 요인을 지지하는 변수의 개수가 적어도 3개 이상이어야 함
- 가장 안정적인 요인개수 설정방법
- 카이저 법칙에 따라 고유치가 1이상인 곳에서 요인을 추출
 - Hair 등(1995)이 제안한 사회과학에서 누적분산비율(누적설명분산)이 적어도 50~60% 정도가 되어야 한다는 제안에 적합한 지 검토해야 함.
 - Crawford(1975), Zwick & Velicer(1986)의 연구에 비추어 초기 요인분석의 결과에서 “요인을 지지하는 변수가 적어도 3개 이상”인지를 확인함.
 - 누적분산비율이 50%인지, 한 개의 요인을 구성하는 측정변수가 3개 이상인지 종합적으로 확인하여 요인의 개수를 최종적으로 설정하게 됨.
- 전통적으로 카이저(1960)의 법칙(Kaiser's rule)에 의거하여 고유치 1이상인 곳에서 요인의 개수를 정해 왔다. 요인의 개수를 결정하는 방법은 카이저의 법칙(Scree chart도 동일한 결과 도출)이 거의 유일하고 절대적인 방법처럼 사용되어 왔으나, Zwick와 Velicer(1986) 그리고 Lee(2010)는 선행연구가 없는 초기 연구에서 탐색적으로 요인추출하게 될 때 카이저의 법칙에 의거하여 요인을 추출할 경우 요인이 과대 추출되는 현상을 많이 보인다고 하였다. 특히, Cliff(1988)는 고유치가 1이하인 경우에도 내용타당성에 문제가 없다면 요인을 해석가능하다고 하였으며, Zwick와 Velicer(1986, p441)는 “카이저의 법칙은 절대적인 준거로 사용하는 것을 지지할 증거가 없으며, (중략) 이 법칙을 더 이상 사용해서는 안 된다.”고 까지 제안한 바 있다.
- 따라서, 본 연구에서는 요인의 개수를 해석가능성, 요인(개념)을 지지하는 변수의 개수, 누적분산비율을 종합적으로 고려하여 결정하게 되었다. 요인들을 해석함에 있어, 적어도 3개 이상의 변수를 기초로 해야 의미있는 요인이 될 수 있다는 가정은 오래되고 중요한 지침이다(Thurstone, 1947; Crawford, 1975; Zwick & Velicer, 1986) 라는 주장을 반영하였으며, Hair 등(1995)이 제안한 사회과학에서 누적분산비율(누적설명분산)이 적어도 50~60% 정도가 되어야 한다는 주장을 반영하였다.

- 본 연구에서는 먼저 초기 요인분석을 카이저 법칙에 따라 요인을 추출하였으며, 도출된 요인구조가 과대 추출되었는지 여부를 확인하였다.
- Crawford(1975), Zwick & Velicer(1986)의 연구에 비추어 초기 요인분석의 결과에서 “요인을 지지하는 변수가 적어도 3개 이상”인지를 확인하였다. 만약 카이저 법칙에 근거하여 초기 요인의 개수가 7개 도출되었다면, 요인을 지지하는 변수가 적어도 3개 이상인 요인이 몇 개인지 확인하고 2차 요인분석부터는 요인의 개수를 그에 맞게 조정하게 된 것이다.

→ 참고문헌

- Hair, J.E. Jr., Anderson, R.E., Tatham, R.L., & Black, W.C. (1995). *Multivariate data analysis* (4th edition). Upper Saddle River, NJ: Prentice Hall.
- Lee, S.M. (2010). *A fundamental of factor analysis* (2nd edition). Paju: Kyoyookbook Publisher.
- Cliff, N. (1988). The Eigenvalues-Greater-Than-One Rule and the Reliability of Components. *Psychological Bulletin*, 103, pp276-279.
- Crawford, C.B. (1975). Determining the Number of Interpretable Factors. *Psychological Bulletin*, 82, pp226-237.
- Thurstone, L.L. (1947). *Multiple factor analysis*. Chicago, IL: University of Chicago Press.
- Zwick, W.R. & Velicer, W.F. (1986) Comparison of Five Rules for Determining the Number of Components to Retain. *Psychological Bulletin*, 99, pp432-442.

➤ 회전방법(rotation method)

→ 직각회전(orthogonal rotation)

- 가장 일반적이고 대표적인 방법으로, 베리맥스(Varimax) 기법이 있음
- 요인간의 상관을 최대한 배제한 회전방법
- 판별타당성(discriminant validity)에 친화적인 회전방법, 요인간의 독립성을 최대한 인정하는 방법

→ 사각회전(oblique rotation)

- 요인간의 상관을 가정한 회전방법
- Quartimax, Oblimin 등등이 있음.

➤ 문항제거 기준

- 요인추출 기준은 요인부하량(요인적재량: 요인적재치, factor loading) 0.4 이상, 공통성(공통분산, communality) 0.5 이상으로(Hair et al., 1995) 설정하여, 요인부하량이 아주 낮거나 여러 요인들에 모호하게 분산된 문항들이 제거되도록 제안하였으나, 이 기준은 절대적이지 않으며 **보수적인 방법으로 알려져 있다.**
- Kang(2013)은 높은 적재값의 기준은 다양하며, 요인부하량이 0.3~0.5정도인 경우 변수를 요인에 포함한다고 하였으며,
- Lee(1995, 2010)는 요인구조의 타당성은 내용타당도 기준에서 접근해야 하며 통계적 수치는 절대적일 수는 없다고 가정하였다.
- Song(2016)는 공통성이 0.4이하면 문항을 제거할 것을 권고하였으며, 요인부하량은 0.3이상이면 문제는 없지만 보수적인 기준에서 0.4이상으로 선택할 것을 제안하고 있다.
- 중도적인 방법은 요인부하량이 0.4미만 또는 0.4에 근접할 경우 변수를 제거하고, 또는 공통성이 0.4미만일 때 변수를 제거할 것을 권고함.

➔ 가장 중도적인 문항제거 방법

- **요인부하량이 0.4 미만 또는 0.4에 근접할 경우**
- **공통성이 0.4 미만인 경우**
- **위의 2가지 조건 중 하나라도 위배되면 제거함.**

➔ 참고문헌

- Hair, J.E. Jr., Anderson, R.E., Tatham, R.L., & Black, W.C. (1995). Multivariate data analysis (4th edition). Upper Saddle River, NJ: Prentice Hall.
- Kang, H.C. (2013). A Guide on the Use of Factor Analysis in the Assessment of Construct Validity. Journal of Korean Academy of Nursing, Vol.43(5), 592. <http://dx.doi.org/10.4040/jkan.2013.43.5.587>
- Lee, S.M. (1995). Factor analysis I (2nd edition). Seoul: Hakjisa Publisher.
- Lee, S.M. (2010). A fundamental of factor analysis (2nd edition). Paju: Kyoyookbook Publisher.
- Song, J.J. (2016). Statistical analysis method on SPSS/AMOS (2nd edition). Paju: 21 Century Book Publisher.
- 이순목 (1995). 요인분석 I (2nd edition). 서울: 학지사.



- 이순목(2010). 요인분석의 기초(2nd edition). 파주: 교육과학사.
- 송지준(2016). SPSS/AMOS 통계분석방법(2nd edition). 파주: 21세기사.
- **공통성, 고유치, 분산(설명량)비율, 누적분산비율에 대한 제시**
 - 공통성은 특정 변수가 여러 요인에 대해 설명되는 정도임. 특정 변수가 가지는 여러 요인의 요인부하량의 제곱합으로 정의됨.
- **KMO와 Bartlett 구형성 검정**
 - Kaiser-Meyer-Olkin(KMO)의 Kaiser-Meyer-Olkin Measure of Sampling Adequacy는 변수들 간의 상관관계가 다른 변수에 의해 잘 설명되는 정도는 나타내는 값으로써,
 - 일반적으로 KMO수치가 .90이면 상당히 좋은 것이고, .80~.89면 꽤 좋은 편, .70~.79면 적당한 편, .60~.69면 평범한 편, .50~.59면 바람직하지 못한 편, .50미만이면 받아들일 수 없는 수치로 판단된다.