

A Comparative Study on Factor Recovery of Principal Component Analysis and Common Factor Analysis

Sunho Jung^{a,1} · Sangyun Seo^a

^aSchool of Management, Kyung Hee University

(Received August 23, 2013; Revised October 23, 2013; Accepted October 28, 2013)

Abstract

Common factor analysis and principal component analysis represent two technically distinctive approaches to exploratory factor analysis. Much of the psychometric literature recommends the use of common factor analysis instead of principal component analysis. Nonetheless, factor analysts use principal component analysis more frequently because they believe that principal component analysis could yield (relatively) less accurate estimates of factor loadings compared to common factor analysis but most often produce similar pattern of factor loadings, leading to essentially the same factor interpretations. A simulation study is conducted to evaluate the relative performance of these two approaches in terms of factor pattern recovery under different experimental conditions of sample size, overdetermination, and communality. The results show that principal component analysis performs better in factor recovery with small sample sizes (below 200). It was further shown that this tendency is more prominent when there are a small number of variables per factor. The present results are of practical use for factor analysts in the field of marketing and the social sciences.

Keywords: Principle component analysis, common factor analysis, sample size, communality, overdetermination, factor recovery.

1. 서론

탐색적 요인분석은 사회과학 연구에서 기본적인 분석도구로 사용되어 왔다 (Costello와 Osborne, 2005). 요인분석 사용자들은 탐색적 요인분석을 사용할 때 요인추출방법, 추출할 요인 수 결정, 요인 축의 회전방법에 대해 단계적으로 의사결정을 한다. 그런데 각 단계에서 다양한 방법들이 존재하고 특정 방법의 선택에 따라 요인분석 결과가 달라질 수 있다. 이와 같은 이유로 대표적인 요인추출방법인 주성분분석과 공통요인분석 중 어떤 방법을 선택해야 하는가에 대해 오랜 방법론적 논쟁이 있어왔다 (Velicer과 Jackson, 1990).

다양한 측면에서 요인추출방법을 비교할 수 있지만 요인적재량 추정 능력면에서 보면 다음과 같은 두 가지 기준에서 평가할 수 있다 (Preacher과 MacCallum, 2002). 첫 번째 기준은 정확성 관점(precision perspective)에서 요인추출방법이 모수값(population values)에 얼마나 가까운 추정치(parameter estimate)를 산출해 낼 수 있는가를 평가한다. 이 정확성 관점에서 수행된 연구들은 공통적으로 평균자승차

¹Corresponding author: Assistant Professor, School of Management, Kyung Hee University, 1st Hoeigi-Dong, Dongdaemoon-gu, Seoul 130-701, Korea. E-mail: sunho.jung@khu.ac.kr

이(mean squared difference) 등의 지수를 계산해서 샘플과 모집단의 요인적재량 값이 평균적으로 얼마나 일치하는지를 조사한다 (Snook과 Gorsuch, 1989; Velicer 등, 1982; Velicer과 Fava, 1998). 두 번째 기준은 타당성 관점(validity perspective)에서 요인추출방법이 얼마나 모집단의 요인구조(population factor structure)를 잘 복원할 수 있는가를 평가한다. 이 기준은 모집단과 표본 요인적재량의 패턴을 비교함으로써 요인구조의 유사성(similarity of factor structure)을 조사한다. 일반적으로 요인분석을 사용하는 연구자들은 요인적재량 값 그 자체 보다는 요인패턴에 더 관심을 가지게 된다. 이 요인패턴의 해석을 통해 어떤 측정변수가 어떤 요인에 의해 주로 설명되는지를 파악할 수 있기 때문이다. 따라서 실증 연구에서 두 요인추출방법을 통해 추정된 요인적재량의 값이 수치상 서로 다르더라도 요인패턴이 같다면 동일한 해석 결과를 얻게 된다.

선행연구들은 시뮬레이션 방법을 사용해서 공통요인분석과 주성분분석을 주로 정확성 관점에서 평가하였다. 예를 들면, Snook과 Gorsuch (1989)은 주성분분석과 공통요인분석 비교연구에서 공통요인분석이 더 정확한 요인적재량을 산출한다는 결과를 도출했다. 하지만 요인분석 사용자 입장에서는 올바른 요인해석을 뜻하는 타당성 관점에서 두 요인분석방법을 이해하는 것 또한 중요하다. 하지만 이 관점에서 두 방법을 체계적으로 비교 평가한 연구는 여전히 찾아보기 어렵다. 따라서 본 연구에서는 요인구조 복원에 중요한 영향을 미치는 조건을 제시하고 각 조건 하에서 주성분분석이 공통요인분석과 비교해서 얼마나 잘 모집단과 유사한 요인구조를 생성해 낼 수 있는지를 몬테카를로(Monte Carlo) 시뮬레이션 기법을 통해 체계적으로 평가해 보고자 한다. 이 연구 결과를 토대로 사회과학 연구자들은 어떤 상황에서 주성분분석이 올바른 요인 해석결과를 제공할 수 있는지를 이해할 수 있다.

2. 이론적 배경

유통업계 및 금융업계 등에서 관찰할 수 있는 데이터는 많은 수의 고객으로부터 발생하는 방대한 고객 정보와 많은 양의 구매 및 거래 정보로 이루어져 있으므로 그 규모가 매우 크다. 수많은 고객 정보를 다루는데 있어서 제기되는 중요한 통계적 이슈 중 하나는 과연 이 정보를 효과적으로 어떻게 요약할 것인지에 관한 것이다 (Witten 등, 2011). 다변량 통계 분석법 중 주성분분석은 이러한 문제에 대한 답을 쉽게 할 수 있다. 주성분분석은 차원축소를 통해 다변량 변수들을 요약하고 변수들 간의 구조를 분석할 수 있는 통계 기법이다. 그런데 공통요인분석과 달리 이 차원축소는 변수들의 선형결합을 통해서 이루어진다. 가중평균처럼 선형결합을 위해 필요한 것이 각 변수의 설명력을 나타내는 가중치(weight)인데 주성분분석은 데이터에 있는 정보를 바탕으로 이 가중치를 추정한다. 이 가중치 추정을 통한 선형결합의 결과물이 주성분(principal component)이다.

공통요인분석은 측정오차를 설명하기 위해 상관행렬 대각선에 있는 요소를 1대신 공통분산(common variance)으로 대체한 후 요인적재량을 추정하기 위해 특이값 분해법(singular value decomposition; SVD)을 적용한다. 고유분산을 추정하는 방법은 어떤 평가함수(criterion function)를 적용하는가에 따라 최대우도법과 주축분해법으로 구분된다 (Briggs와 MacCallum, 2003). 이에 반해 사회과학 연구자들이 주로 사용하는 통계소프트웨어인 SPSS 또는 SAS에 있는 요인분석의 디폴트 옵션으로 지정되어 있는 주성분분석은 SVD를 사용해서 상관행렬을 그대로 분해한다.

주성분분석은 상관행렬을 그대로 분석하기 때문에 축소된 상관행렬(reduced correlation matrix)을 분석하는 공통요인분석에 비해 요인적재량 패턴은 비슷해 보일 수 있지만 각각의 추정치는 상대적으로 큰 경향을 보이게 된다 (Velicer과 Jackson, 1990). Table 2.1은 이 점을 예증하고 있다. Table 2.1에 사용된 분석자료는 Hair 등 (2009)에서 요인분석 예제로 제시된 9개의 매장이미지 속성에 대한 상관행렬이다. 평행분석법을 사용해서 총 3개의 요인이 통계적으로 유의미하다는 결과를 얻었다 (O'Connor 2000). 이 자료에 주성분분석과 최대우도법을 이용한 공통요인분석을 적용하여 요인적재량 행렬을 구하

Table 2.1. The result of common factor analysis and principle component analysis on the image of store.

Variables	Factors			Components		
	1	2	3	1	2	3
Price Level	-.18	.15	.97	-.19	.16	.91
Sales Person	-.77	.29	.28	-.80	.28	.29
Repund Policy	-.85	.24	.09	-.88	.24	.08
Stock	-.25	.76	.31	-.25	.81	.30
Product Quality	-.17	.21	.73	-.14	.18	.91
Depth of Product Categories	-.21	.84	.09	-.19	.90	.07
Width of Product Categories	-.28	.77	.16	-.27	.84	.15
Service in Store	-.79	.16	.10	-.87	.14	.08
Mood in Store	-.79	.25	.19	-.84	.24	.18

고 이 행렬에 베리맥스(Varimax) 회전을 적용하여 Table 2.1에 나와있는 결과를 얻었다. Table 2.1을 보면 주성분분석을 통해 얻은 요인적재량이 공통요인분석 결과보다 거의 일관되게 크다는 것을 알 수 있다. 공통요인분석은 측정오차를 포함하는 고유분산이 상관행렬로부터 제거되기 때문에 전반적으로 주성분분석에 비해 모수에 더 가까운 값(즉, 작은 값)으로 이루어진 요인적재량 행렬을 산출한다. 그렇지만 요인 해석을 위해 요인적재량 0.40 컷오프포인트(Hair 등, 2009)를 사용하고 이에 따라 단순화된 요인패턴을 본다면 두 요인추출방법은 동일한 요인해석 결과를 도출한다.

3. Monte Carlo Study

사회과학 및 행동과학 연구자들은 요인과 측정변수 간의 구조적 관계를 나타내는 요인적재량의 구조를 해석하는 데 주된 관심을 가지고 있다. 따라서 요인추출방법의 요인구조 복원능력은 오랫동안 중요한 이슈였다(Acito와 Anderson, 1980; Gagne와 Hancock, 2006; Guilford, 1954; Pennell, 1968). 일반적으로 표본수는 요인구조 복원에 중요한 영향을 주는 것으로 알려져 있다(MacCallum 등, 1999). 그러나 요인구조 복원은 단순히 변수 개수 대 표본수의 비율로 정리할 수 있는 문제가 아니다. 최근 연구에 따르면 요인구조 복원력은 표본수뿐만 아니라 공통성(communality)과 요인 대 변수 비율(factor-to-variable ratio: overdetermination)에 의해서도 크게 영향을 받는다고 알려져 있다(Velicer과 Fava, 1998; MacCallum 등, 2001). 따라서 요인구조 복원에 관한 시뮬레이션 연구에서 검토해야 할 중요한 실험 변수는 공통성, 표본수, 요인 대 변수 비율이다.

본 연구에서 표본수(n)는 50, 100, 200, 300, 500으로 설정했다. 표본수 $n = 50$ 은 요인분석에서 허용될 수 있는 최소 표본수로 알려져 있다(Velicer과 Fava, 1998). 공통성은 다음과 같은 세 가지 수준으로 설정하였다: High (.6-.7), Low (.2-.4), Wide (.2-.8). 요인 대 변수 비율은 두 가지 수준으로 구분할 수 있다. 높은 요인 대 변수 비율(high overdetermination) 수준에서 요인 수(f)는 3개 그리고 변수 개수(p)는 20개로 설정했다($f : p = 1 : 7$). 낮은 요인 대 변수 비율(low overdetermination)에서 요인 수는 $f = 7$ 그리고 변수 개수는 $p = 20$ 이다($f : p = 1 : 3$). Tucker 등(1969)이 개발한 방식을 사용해서 6개의 모집단 상관행렬을 만들었다. 이는 공통성 세 수준(3) X 요인 대 변수 비율 두 수준(2)의 조합에 해당된다. 각각의 모집단 상관행렬에 Cholesky 분해법을 적용해서 상삼각행렬(upper triangular matrix)을 구하고 정규분포를 따르는 난수(random number)를 생성해서 이를 서로 곱해서 표본을 생성한다(Wijsman, 1959). 총 30개의 실험 조건(5 Sample Sizes x 3 levels of communality x 2 degrees of overdetermination)에서 각각 표본 상관행렬을 200개 생성하였다. 결과적으로 총 6000개의 표본 상관행렬이 생성되었다. 각 표본 상관행렬에 주성분분석과 공통요인분석의 최대우도법을 적용해서 추정

Table 3.1. The result of ANOVA of congruence coefficient

Source	Sum of Squares	df	F	Sig.	Eta Squared
Approach (A)	5.51	1	828.99	.00	.065
Communality (B)	12.43	2	935.41	.00	.135
Overdetermination (C)	24.42	1	3676.99	.00	.235
Sample Size (D)	33.81	4	1272.77	.00	.299
A*B	.55	2	41.46	.00	.007
A*C	5.17	1	777.88	.00	.061
A*D	9.91	4	372.98	.00	.111
B*C	3.31	2	249.33	.00	.040
B*D	4.08	8	76.73	.00	.049
C*D	14.99	4	564.27	.00	.159
A*B*C	.31	2	23.45	.00	.004
A*B*D	.44	8	8.22	.00	.005
A*C*D	8.95	4	337.04	.00	.101
B*C*D	.89	8	16.70	.00	.011
A*B*C*D	.36	8	6.81	.00	.005
Error	79.304	11940			

된 요인적재량 행렬을 구했다.

본 연구에서 요인구조 복원능력은 요인해석의 유사성(similarity of factor interpretations)을 의미한다. 이를 평가할 수 있는 척도로서 일반적으로 Tucker의 일치도 계수(congruence coefficient)를 사용한다. 이 계수를 활용해서 모집단과 표본의 요인적재량 행렬 패턴의 유사성을 평가한다. Lorenzo-Seva와 ten Berge (2006)는 이 계수의 해석기준을 제시했는데 일치도 계수 .85-.94를 만족할만한 수준의 유사성(fair similarity)으로 해석한다. 일치도계수가 .95 이상인 경우 모집단과 표본의 요인적재량 패턴은 서로 정확하게 일치한다고 말한다. 예를 들면, Table 2.1에 있는 두 요인적재량 행렬의 일치도 계수는 .997이다. 일반적으로 일치도계수 .85를 요인구조 복원의 최소 기준으로 사용한다 (Bijmolt와 Van de Velden, 2012; Rammstedt 등, 2010).

이 시뮬레이션 연구에서 세 가지 실험변수(즉, 공통성, 표본수, 요인 대 변수 비율)와 요인추출방법(주성분분석 vs. 공통요인분석-최대우도법)을 독립변수로, 일치도 계수를 종속변수로 설정한다. 여기에 분산분석(ANOVA)을 적용하여 각 독립변수가 종속변수에 어떤 영향을 주는지를 체계적으로 조사해볼 수 있다.

Table 3.1은 분산분석 결과를 보여준다. 모든 주효과(main effect)와 상호작용효과(interaction effect)는 통계적으로 유의미하다. 하지만 이 결과를 그대로 받아들이기 어렵다. 이 실험에서 생성된 6000개의 표본으로 인해 실질적으로 의미 없는 차이임에도 불구하고 통계적으로 유의미한 결과를 나타낼 수 있기 때문이다. 이런 경우 효과 크기(effect size)를 사용해서 각 독립변수들의 영향력을 평가해야 한다 (Paxton 등, 2001). 이 논문에서는 효과크기(eta square)가 .06 이상인 주효과 및 상호작용효과만 실질적으로 유의미하다고 간주한다 (Cohen, 1988). Table 3.1을 보면 모든 주효과는 중간 이상의 효과 크기를 가지고 있다. 다시 말해서, 각 독립변수의 수준 간에 실제적으로 유의미한 차이가 있음을 보여준다. 구체적으로, 주성분분석은 평균 일치도 계수가 .96이고 최대우도법은 .91이다. 이는 주성분분석이 평균적으로 모집단에 더 유사한 요인구조를 산출한다는 것을 뜻한다. 공통성은 High일 때 .97, Wide일 때 .94, Low일 때 .89이다. 높은 요인 대 변수 비율(요인 수 = 3)일 때 .98, 낮은 요인 대 변수 비율(요인 수 = 7)일 때 .89이다. 표본수가 50일 때 .84이고 500일 때는 .98에 달한다. 요인추출방법과 요인

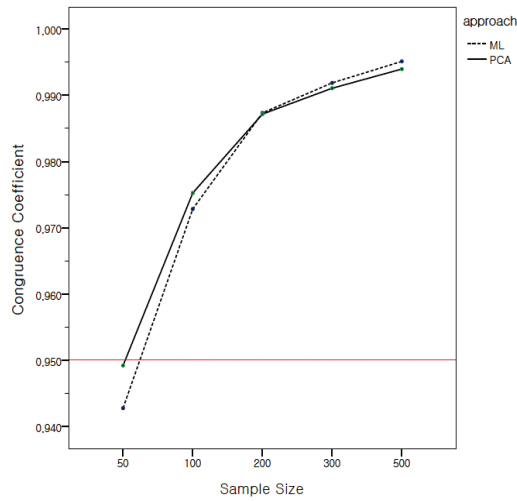


Figure 3.1. The congruence coefficient at high overdetermination of principle component analysis and maximum likelihood method

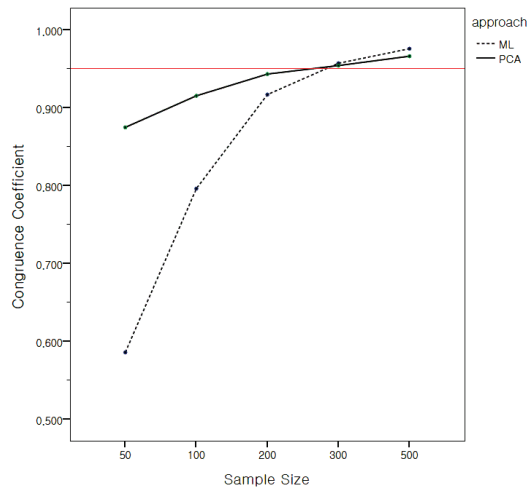


Figure 3.2. The congruence coefficient at low overdetermination

대 변수 비율 그리고 요인추출방법과 표본수 간의 이원 상호작용효과가 중간크기 이상의 효과 크기를 보여주었다. 또한 이 세 독립변수 간의 삼원 상호작용효과는 큰 크기의 효과 크기를 가지고 있다. 따라서 이 삼원 상호작용효과에 초점을 맞추어 요인추출방법과 표본수 간의 상호작용 패턴이 요인 대 변수 비율에 따라 어떻게 달라지는지 분석해보고자 한다.

Figure 3.1에서 높은 요인 대 변수 비율(요인 당 변수 7개: high overdetermination) 상황에서는 표본수와 상관없이 주성분분석과 최대우도법 모두 높은 수준의 일치도 계수를 보이고 있다. 다시 말해서 두 추출방법 모두 모집단 요인구조와 동일한 결과를 산출한다. 다음으로 Figure 3.2는 낮은 요인 대 변수 비율(요인 당 변수 3개: low overdetermination) 상황에서 두 요인추출방법의 요인구조 복원능력을 보여

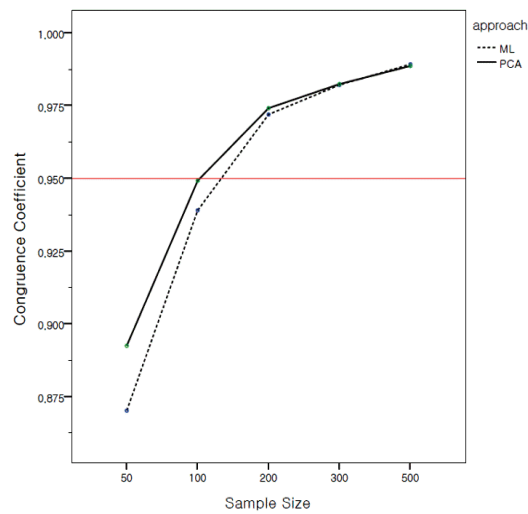


Figure 3.3. The congruence coefficient at high overdetermination and low communality condition

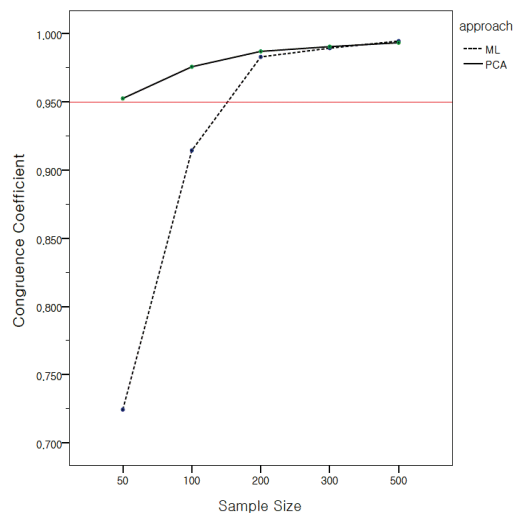


Figure 3.4. The congruence coefficient at low overdetermination and high communality condition

준다. 주성분분석은 표본수가 50인 경우에도 .85이상의 일치도 계수를 보이고 있어 모집단과 상당히 유사한 요인구조를 생성할 수 있음을 알 수 있다. 표본수가 200 이상인 경우 모집단 요인구조를 동일하게 복원할 수 있다. 반대로 최대우도법은 선행연구에서 밝혀졌듯이 표본수가 적은 경우 낮은 수준의 복원 능력을 보여준다. 일치도 계수가 .85 이하인 경우 모집단과 표본 요인구조 간에 전혀 유사성을 찾을 수 없다 (Lorenzo-Seva와 ten Berge, 2006). 따라서 요인 당 변수 개수가 상대적으로 적은 경우 표본수가 100 보다 작으면 최대우도법은 왜곡된 요인해석 결과를 도출 수 있다.

Table 3.1에서 공통성은 다른 독립변수들과 유의미한 상호작용효과를 보이지 않는다. 하지만 공통성은 표본수 다음으로 높은 수준의 효과 크기를 가진다. 즉, 공통성의 세 가지 수준에서 일치도 계수는 서로

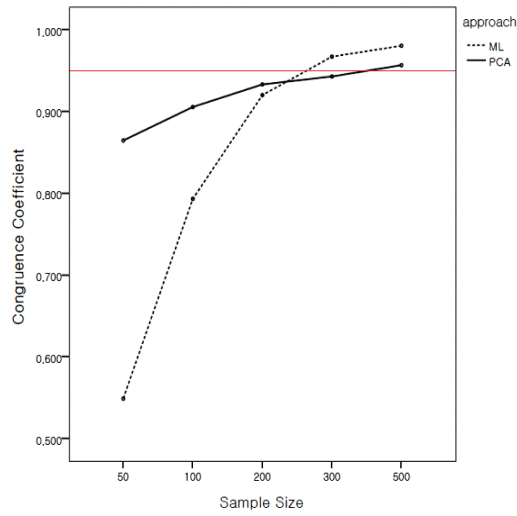


Figure 3.5. The congruence coefficient at low overdetermination and wide communality condition

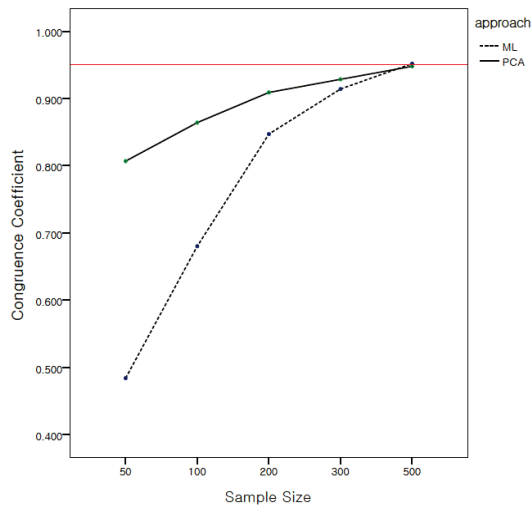


Figure 3.6. The congruence coefficient at low overdetermination and low communality condition

큰 차이를 가지고 있다. 또한 앞서 설명한 바와 같이 많은 선행연구에서 공통성은 요인구조 복원에 큰 영향을 미치고 있음을 보여왔다. 따라서 공통성 수준에 따라 Figure 3.1과 Figure 3.2에서 설명한 일치도 계수 패턴에 어떤 변화가 생기는지 알아보려고 한다.

먼저, 요인 대 변수 비율이 높은 상황에서 공통성 수준이 일관되게 높거나(High) 큰 폭으로 퍼져있는(Wide) 경우 표본수에 관계없이 두 방법 모두 .95 이상의 일치도 계수를 보였다. 하지만 Figure 3.3에서 보듯이 공통성 수준이 일관되게 낮은(low) 경우 주성분분석은 표본수가 100 이상일 때 모집단과 동일한 요인해석 결과를 얻을 수 있다.

끝으로 요인 대 변수 비율이 낮은 상황에서 공통성의 수준에 따라 요인추출방법의 요인구조 복원 패턴

이 어떻게 달라지는지 살펴보자. Figure 3.4에서 보듯이 공통성의 수준이 높다면 주성분분석은 표본수가 50인 경우에도 모집단과 같은 요인해석 결과를 도출할 수 있다. 이에 반해 공통성이 높고 낮은 수준으로 서로 섞여 있을 때 Figure 3.5가 보여주듯이 최대우도법이 주성분분석보다 더 적은 수의 표본으로 .95 수준에 먼저 도달할 수 있다. 주성분분석에서 상관행렬의 대각선 요소들은 모두 1로 이루어져 있다. 이는 공통성 수준이 모두 동일하다는 가정을 한 것과 같다. 다시 말해서 공통성이 낮고 큰 다양한 값으로 이루어져 있는 경우 이 가정에 위배되므로 상대적으로 낮은 요인구조 복원력을 보여주게 되는 것이다. 그렇지만 주성분분석은 표본수가 50인 경우에 이미 일치도 계수가 .85이므로 표본수가 적은 경우 여전히 더 강력한 성능을 보인다. Figure 3.6에서 공통성 및 요인 대 변수 비율 둘 다 낮은 경우 두 추출 방법 모두 상대적으로 약한 요인구조 복원력을 보여준다. 둘 다 표본수가 500일 때 .95 일치도 계수에 도달할 수 있다. 그런데 주성분분석은 표본수 100인 경우에 이미 만족할만한 수준인 .85에 도달한다.

4. 결론 및 시사점

요인분석에서 요인을 추출하는 방법은 크게 전체변수의 분산을 이용하는 주성분분석과 변수들의 공통분산을 이용하는 공통요인분석방법으로 구분되는데, 공통요인분석은 변수의 측정오차를 고려한다는 측면에서 보다 타당한 방식으로 간주할 수 있다. 그렇지만 요인적재량 추정이 용이하다는 이유에서 주성분분석이 사회과학 및 행동과학 연구에서 더 빈번하게 사용되어 왔다 (Conway와 Huffcutt, 2003).

그런데 주성분분석이 요인구조 복원 측면에서 어떤 능력을 가지고 있는지에 대해서는 체계적인 평가가 이뤄지지 않았다. 따라서 본 연구에서는 공통요인분석과 비교해서 어떤 수준의 요인구조 복원능력을 가지고 있는지를 시뮬레이션 기법을 통해서 체계적으로 평가하고자 하였다. 이를 위해 요인구조 복원에 중요한 영향을 미치는 요인 대 변수 비율, 공통성, 표본수를 실험변수로 하여 각 조건하에서 주성분분석의 요인구조복원 능력을 공통요인분석과 비교 평가하였다.

이 연구를 통해 다음과 같은 중요한 결과를 얻었다. 첫 번째, 표본수가 200보다 적은 경우 주성분분석이 공통요인분석보다 일관되게 요인구조 복원력이 떨어졌다. 두 번째, 요인 당 변수 수가 높은 조건에서는 표본수가 50인 경우에도 주성분분석과 공통요인분석 모두 모집단과 동일한 수준의 요인패턴을 산출할 수 있었다. 만약 이 조건에서 공통성 수준이 낮아지면 두 요인추출방법 모두 모집단과 동일한 요인구조를 산출하기 위해 100 이상의 표본수를 필요로 했다. 세 번째, 요인 대 변수 수가 낮은 조건에서 주성분분석은 공통요인분석 훨씬 더 우수한 요인구조 복원력을 보여주었다. 공통요인분석은 표본의 크기에 따라 일치도 계수의 변화가 큰 반면에, 주성분분석은 표본크기에 관계없이 안정적인 일치도 계수 패턴을 보여주었다. 이 조건에서 주성분분석은 모집단과 동일한 요인구조 복원을 위해 평균적으로 최소한 200 개의 표본수를 필요로 한다. 그런데 주성분분석은 표본수 50인 경우에도 만족할만한 수준의 복원력을 보여주었다. 구체적으로 공통성 수준이 높으면 주성분분석의 경우 모집단과 동일한 요인구조 복원을 위해 필요한 표본수는 50으로 줄어든다. 이에 반해 공통성이 넓은(wide) 분포를 갖는 경우, 주성분분석과 공통요인분석은 동일한 요인구조 복원을 위해 300 이상의 표본수를 필요로 한다. 그런데 주성분분석은 표본수 50인 경우에도 만족할만한 수준의 복원력을 보여주었다. 마지막으로 요인 대 변수 비율과 공통성 수준이 모두 낮은 경우에는 두 방법 모두 모집단과 동일한 요인구조 복원을 위해 500 이상의 표본수를 필요로 하지만, 주성분분석은 100개의 표본에서도 만족할만한 수준의 요인복원능력을 보여주었다.

결론적으로 주성분분석은 전반적으로 공통요인분석보다 안정적이며 우수한 요인복원능력을 보여주었으며, 최대우도법을 이용한 공통요인 분석은 표본크기가 작은 경우 요인복원능력이 상당히 저조한 것으로 나타났다. 본 연구결과는 사회과학 및 행동과학에서 주로 사용하는 주성분분석의 타당한 사용에 관한 가이드라인을 올바른 요인해석 측면에서 제시해 준다.

References

- Acito, F. and Anderson, R. D. (1980). A monte carlo comparison of factor analytic methods, *Journal of Marketing Research*, **17**, 228–236.
- Bijmolt, T. H. A. and Van de V., M. (2012). Multiattribute perceptual mapping with idiosyncratic brand and attribute sets, *Marketing Letters*, **23**, 585–601.
- Briggs, N. E. and MacCallum, R. C. (2003). Recovery of weak common factors by maximum likelihood and ordinary least squares estimation, *Multivariate Behavioral Research*, **38**, 25–56.
- Carroll, J. D. and Green, P. E. (1997). Psychometric methods in marketing research: Part 2, multidimensional scaling, *Journal of Marketing Research*, **34**, 193–204.
- Cohen, J. (1988). *Statistical Power for the Behavioral Sciences*, Hillsdale, Lawrence Erlbaum, NJ.
- Conway, J. M. and Huffcutt, A. I. (2003). A review and evaluation of exploratory factor analysis practices in organizational research, *Organizational Research Methods*, **6**, 147–168.
- Costello, A. and Osborne, J. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis, *Practical Assessment, Research and Evaluation*, **10**, 1–9.
- Gagné, P. and Hancock, G. R. (2006). Measurement model quality, sample size, and solution propriety in confirmatory factor models, *Multivariate Behavioral Research*, **41**, 65–83.
- Guilford, J. P. (1954). *Psychometric Methods*, McGraw Hill, New York.
- Hair, J. F., Black, W. C., Babin, B. J. and Anderson, R. E. (2009). *Multivariate Data Analysis*, Upper Saddle River, Prentice-Hall, NJ.
- Lorenzo-Seva, U. and ten Berge, J. M. F. (2006). Tucker's congruence coefficient as a meaningful index of factor similarity, *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, **2**, 57–64.
- MacCallum, R. C., Widaman, K. F., Zhang, S. and Hong, S. (1999). Sample size in factor analysis, *Psychological Methods*, **4**, 84–99.
- MacCallum, R. C., Widaman, K. F., Preacher, K. and Hong, S. (2001). Sample size in factor analysis: The role of model error, *Multivariate Behavioral Research*, **36**, 611–637.
- O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test, *Behavior Research Methods, Instruments, and Computers*, **32**, 396–402.
- Paxton, P., Curran, P. J., Bollen, K., Kirby, J. and Chen, F. (2001). Monte Carlo Experiments: Design and implementation, *Structural Equation Modeling*, **8**, 287–312.
- Pennell, R. (1968). The influence of communality and N on the sampling distributions of factor loadings, *Psychometrika*, **33**, 423–439.
- Preacher, K. J. and MacCallum, R. C. (2002). Exploratory factor analysis in behavior genetics research: Factor recovery with small sample sizes, *Behavior Genetics*, **32**, 153–161.
- Rammstedt, B., Goldberg, L. R. and Borg, I. (2010). The measurement equivalence of big-five factor markers for persons with different levels of education, *Journal of Research in Personality*, **44**, 53–61.
- Snook, S. C. and Gorsuch, R. L. (1989). Component analysis versus common factor analysis: A Monte Carlo study, *Psychological Bulletin*, **106**, 148–154.
- Tucker, L. R., Koopman, R. F. and Linn, R. L. (1969). Evaluation of factor analytic research procedures by means of simulated correlation matrices, *Psychometrika*, **34**, 421–459.
- Velicer, W. F. and Fava, J. L. (1998). Effects of variable and subject sampling on factor pattern recovery, *Psychological Methods*, **3**, 231–251.
- Velicer, W. F., Peacock, A. C. and Jackson, D. N. (1982). A comparison of component and factor patterns: A Monte Carlo approach, *Multivariate Behavioral Research*, **17**, 371–388.
- Velicer, W. F. and Jackson, D. N. (1990). Component analysis vs. Common factor analysis: Some issues in selecting an appropriate procedure, *Multivariate Behavioral Research*, **25**, 1–28.
- Wijsman, R. A. (1959). Applications of a certain representation of the Wishart matrix, *Annals of Mathematical Statistics*, **30**, 597–601.
- Witten, I. H., Frank, E. and Hall, M. A. (2011). *Data Mining: Practical Machine Learning Tools and Techniques*, Morgan Kaufmann, Burlington, MA.

주성분분석과 공통요인분석에 대한 비교연구: 요인구조 복원 관점에서

정선호^{a,1} · 서상윤^a

^a경희대학교 경영대학

(2013년 8월 23일 접수, 2013년 10월 23일 수정, 2013년 10월 28일 채택)

요약

본 연구에서는 시뮬레이션 방법을 사용해서 다양한 조건에서 주성분분석이 얼마나 잘 요인 구조를 복원할 수 있는지를 공통요인분석과 비교하여 체계적으로 평가하였다. 이 연구에서 요인 대 변수 비율, 공통성, 그리고 표본크기를 실험변수로 설정하였다. 주성분분석은 표본의 크기가 200개 이하인 경우 공통적으로 공통요인분석에 비해 더 우수한 요인구조의 복원력을 보여주었다. 특히, 요인 당 변수 수가 적은 경우, 주성분분석은 50개의 표본에서도 만족할 만한 수준의 요인복원능력을 보여주었다. 이와 더불어 공통성 수준 또한 낮은 경우 필요한 표본수는 100개로 늘어난다. 본 연구결과는 요인추출방법으로서 주성분분석의 선택의 근거를 제시하고 타당한 사용에 관한 가이드라인을 제시해 준다.

주요용어: 주성분분석, 공통요인분석, 표본의 수, 공통성, 요인 대 변수 비율, 요인 복원.

¹교신저자: (130-701) 서울특별시 동대문구 회기동 1번지, 경희대학교 경영대학, 조교수.
E-mail: sunho.jung@khu.ac.kr