

계량분석

Internal Reliability and Cronbach's Alpha

김현우, PhD¹

¹충북대학교 사회학과 조교수

November 28, 2024



진행 순서

- 1 요인회전
- 2 신뢰도
- 3 크론바흐 알파
- 4 실제 응용
- 5 연습문제

요인회전

요인적재량의 해석을 돕기 위해 요인회전을 수행할 수도 있다.

- 회전(rotation)은 사실 삼각함수(trigonometric functions)의 행렬을 배워야 이해할 수 있는 개념이다.
- 이른바 회전행렬(rotation matrix)이 있고 이것을 ω 들의 행렬에 곱하면 그래프의 축을 회전시킬 수 있다.
- x 축과 y 축이 본래 90° 로 직교하듯, 직교의 제약을 여전히 유지하면서 x 축과 y 축을 뒤틀 수 있다.
- 이것을 직교 회전(orthogonal rotation) 내지 대각 회전이라고 부르는데, 우리는 베리맥스(varimax)를 사용한다.



요인회전

- 주성분 간에는 서로 직교하므로 상관계수는 0이다(직접 확인해보자!). 이것은 수학적으로 아름답지만 실질적으로 심각한 제약이 된다(Why?).
- 대각 회전의 치명적인 단점은 “주성분 사이에 상관관계가 없다”는 강한 제약이 그대로 유지된다는 것이다.
- 이 가정이 이론적으로 정당화될 수 있다면(혹은 되어야 한다면!) 물론 납득할 수 있다.
- 대각 회전 대신 **사각 회전(oblique rotation)**은 x 축과 y 축 사이의 직교 제약을 완화하여 따로따로 회전시킨다. 그리하여 주성분 사이에 상관관계를 허용한다.



실용적으로 둘 다 해보는 것이 바람직하다!

- 이론과 무관하게 데이터 분석의 입장에서 본다면, 먼저 사각 회전을 해보고 주성분 사이의 통계적으로 유의한 상관관계수가 있을 때 사각 회전으로 갈 수 있다.
- 큰 차이가 있다면 사각 회전을 따라가는 것을 추천한다(Why?).
- 별 차이가 없다면 대각 회전을 유지하고, 각주에서 사각 회전을 확인해보았지만 차이가 없었다고 언급할 수 있다.



Stata에서 주성분분석 이후 회전을 연습해보자.

- 다시 renpainters에서 모든 변수에 대해 주성분분석을 일차적으로 수행하자.
요인적재량을 들여다보고 주성분을 도출해보자.
- 대각 회전을 실시하자. 요인적재량은 어떻게 바뀌었는가? 요인적재량이 0.4보다 작은 것은 아예 나타나지 않게 하여 관찰해보자.
- 이번에는 promax(3) 사각 회전을 실시하자. 요인적재량은 어떻게 바뀌었는가?
요인적재량이 0.4보다 작은 것은 아예 나타나지 않게 하여 관찰해보자.



회전 이후에도 해석이 모호하다면 능동적으로 내용 구성을 바꿀 수 있다.

- 어떤 변수의 요인적재량이 0.4 이하라면 해당 변수와 주성분 간에는 상관관계가 약한 셈이다. 만일 어떤 주성분에 대해서도 0.4를 넘지 못한다면 주성분분석에서 이 변수를 아예 탈락시킬 수 있다.
- 둘 이상의 주성분에 대해 0.3 이상의 요인적재량을 갖는 경우를 “교차적재량 (cross-loading)이 높다”고 표현한다. 이것도 바람직하지 않으므로 이러한 변수를 제거할 수 있다(Why?).
- 논문이나 보고서에 결과표를 제시할 때는 0.4 보다 작은 요인적재량은 아예 표기하지 않는 것도 보기 좋다(직접 찾아볼 것).
- 이 모든 문제는 타당화된 측정도구를 사용하면 거의 발생하지 않는다(Why?).



엄밀히 말해 주성분분석에서 회전은 무의미하거나 심지어 나쁠 수 있다.

- 주성분분석은 본래 주어진 변수들의 분산을 최대화하는 주성분들을 찾아내는 목적을 가지고 있다.
- 그런데 회전을 하면 주성분들의 분산이 더이상 최대라는 보장이 없어지므로 본래 목적을 해칠 위험이 생긴다.
- 왜 그럼에도 불구하고 회전을 할까? 이는 해석상의 편리함 때문이다.
- 해석상의 편리함을 추구한다면 주성분분석보다 확인적 요인분석이 이론적으로 옳다. 그럼에도 불구하고 관행적으로 사회과학자들은 주성분분석 이후 회전을 한다.



신뢰도

신뢰도

- 어떤 측정도구에 신뢰도가 확보되어 있는가를 확인하기 위한 몇 종류의 기본적인 유형 내지 방법이 알려져 있다: (1) 재검사 신뢰도(test-retest reliability), (2) 반분 신뢰도(split-half reliability), (3) 내적 일관성 신뢰도(internal consistency reliability), (4) 채점자간 신뢰도(inter-coder reliability) 등.
- 신뢰도 높은 문항은 엄격한 조작적 정의와 구체적인 측정문항 및 절차를 전제로 하므로 사회학적 개념의 풍부한 의미와 뉘앙스를 박탈하는 것처럼 보인다(Babbie 2021).



신뢰도

- 재검사 신뢰도는 같은 측정도구를 한 번 더 시험해서 앞서 얻은 결과와 얼마나 일관되어 있는지 확인하는 방법 내지 그렇게 확인할 수 있는 유형의 신뢰도이다.
- 반분 신뢰도는 측정도구의 문항들을 임의로 나누어 시험한 뒤, 두 세트의 결과가 얼마나 일관되어 있는지 확인하는 방법 내지 그렇게 확인할 수 있는 유형의 신뢰도이다. 강승호·김양분(2004)은 이를 내적 일관성의 일부로 본다.
- 내적 일관성 신뢰도는 문항들 사이에서 상관관계가 높다면 이에 근거한 측정도구의 (내적) 일관성이 높음을 확인하는 방법 내지 그렇게 확인할 수 있는 유형의 신뢰도이다.
- 채점자간 신뢰도(혹은 코더간 신뢰도)는 조금 다른 맥락(내용분석 등)에서 여러 채점자 간 혹은 코더 간 내용을 평가한 뒤, 이 일관성을 확인하는 방법 내지 그렇게 확인할 수 있는 유형의 신뢰도이다.



- 방법론 자체를 연구하는 경우 “시험을 위한 시험”을 치를 의향이 있으므로 다양한 방식을 고려할 수 있지만, 실제 연구에서는 내적 일관성 점수인 **크론바흐 알파 (Cronbach's α)**를 사용하는 것이 보통이다.
- 방법론 전공자들 사이에서는 크론바흐 알파에 대한 비판적 연구가 이미 많이 나왔지만(Sijtsma 2009), 여전히 널리 쓰이기 때문에 적어도 2024년 현재에도 크게 걱정할 필요는 없어 보인다.
- 크론바흐 알파는 원칙적으로 양적 변수에 대해 사용하지만, 실제로는 리커트 척도나 심지어 가변수(dummy variable)를 사용했을 때도 대체로 잘 작동한다.



가장 먼저 신뢰도 비율 ρ 의 논리를 이해해야 한다

- 어떤 복합적인 개념(e.g., 삶의 행복)을 단일 문항으로 측정하기 어려워, 여러 문항들을 사용한다고 하자.
- 개별 문항들(X_1, X_2, \dots, X_k)을 모두 더해 **합성지수(composite index)**인 Y 를 만들었다.
- 측정 이론(measurement theory)**에서는 이렇게 만들어지는 변수를 **잠재변수(latent variable)**라고 부른다(단 합성지수가 잠재변수와 동의어인 것은 아니다!).
- (조사방법론의 맥락에서) 내적 일관성은 바로 잠재변수의 신뢰도를 계산하는 원리이자 방법을 의미한다. 따라서 잠재변수를 따로 구성해 만들지 않았다면 내적 일관성 이야기는 (조사방법론의 맥락에서) 나올 이유가 없다.



- 모든 수학적 논리가 그러하듯 여기에도 몇가지 가정이 있다.

$$A1: X = T + e$$

$$A2: E(e) = 0$$

$$A3: \rho_{e,T} \equiv Cov(e, T) = 0$$

$$A4: \rho_{e_i, e_j} \equiv Cov(e_i, e_j) = 0$$

- 이 가정들의 함의를 살펴보면 **불편성(unbiasedness)**, 즉 $E(X) = E(T)$ 이 전제되어 있다.
- 이 맥락에서 신뢰도 계산은 일단 측정지표의 타당성을 전제로 한다(Why?).



신뢰도

- 우리가 관찰하는 값(X)에는 본래 측정하고자 의도했던 진정한 값(T)에 오차(e)가 끼어있다.

$$X = T + e \quad (A1)$$

- 과녁판의 비유에서 살펴보았듯 신뢰도는 분산의 문제이다. 그러므로 양변의 분산을 씌우면 신뢰도의 정도를 살펴볼 수 있다.

$$\begin{aligned} \text{Var}(X) &= \text{Var}(T + e) \\ &= \text{Var}(T) + \text{Var}(e) + 2\text{Cov}(T, e) \\ &= \text{Var}(T) + \text{Var}(e) \quad (A3) \end{aligned}$$

- 이제 **신뢰도 비율**(ρ)을 아래와 같이 측정할 수 있다(Why?)

$$\rho = \frac{\text{Var}(T)}{\text{Var}(X)} = 1 - \frac{\text{Var}(e)}{\text{Var}(X)}$$

신뢰도

- 그런데 신뢰도 비율 ρ 은 결국 관찰하는 값 X 과 진정한 값 T 사이 상관계수의 제곱 r_{XT}^2 이다.

$$\rho = \frac{Var(T)}{Var(X)} = \frac{Var^2(T)}{Var(X)Var(T)} = \frac{Cov^2(X, T)}{Var(X)Var(T)} = r_{XT}^2$$

- 수리적 증명 과정(강승호·김양분 2004)은 생략하고 신뢰도 비율의 실질적 의미를 살펴보자.
- 신뢰도란 결국 진정한 값에 대해 개별 문항들이 얼마나 높은 상관계수의 제곱을 가지는가를 측정한 것이다.

크론바흐 알파

크론바흐 알파

Cronbach's α 의 논리를 직관적으로 살펴보자.

- (과녁판의 비유에서 알 수 있듯) 개별 측정항목의 높은 분산은 곧 해당 측정항목의 높은 불확실성을 의미한다. 개별 측정항목들의 분산합은 결국 사용된 측정항목들의 전반적인 불확실성을 반영한다.
- 그러므로 만약 잠재변수의 전체 분산 σ_X^2 에서 개별 측정항목들의 분산합 $\sum \sigma_{x_i}^2$ 의 비중이 크다면 이는 “일관되지도 않고 재생산되지도 않는” 개별 측정항목들로 그 잠재변수를 만든 것이므로 (그 잠재변수는) 신뢰도가 낮다고 할 수 있다.

$$\rho = \frac{\sigma_X^2 - \sum \sigma_{x_i}^2}{\sigma_X^2} = 1 - \frac{\sum \sigma_{x_i}^2}{\sigma_X^2}$$

- 만일 $\sum \sigma_{x_i}^2 = \sigma_y^2$ 이면 $\rho = 0$ 이고, $\sum \sigma_{x_i}^2 = 0$ 이면 $\rho = 1$ 이다(Why?).



크론바흐 알파

크론바흐 알파는 변수가 늘어나는 것에 약간 패널티를 준 신뢰도 비율이다.

- 합성지수 혹은 잠재변수를 구성하는데 사용한 변수의 수 k 로 적절히 가중된 신뢰도 비율을 계산에 사용한다.

$$\begin{aligned}\alpha &= \left(\frac{k}{k-1} \right) \left(\frac{S_X^2 - \sum S_{x_i}^2}{S_X^2} \right) \\ &= \left(\frac{k}{k-1} \right) \left(1 - \frac{\sum S_{x_i}^2}{S_X^2} \right)\end{aligned}$$

- 분자에 들어간 $\sum S_{x_i}^2$ 가 S_X^2 보다 상대적으로 커지면 α 는 감소한다(Why?).
- α 는 0에서 1사이에 놓인 값을 갖는다. 1에 가까울수록 내적 일관성이 높다(Why?). 관행상 0.7보다 높으면 대체로 문제가 없는 것으로 판단한다.
- 경우에 따라 새로운 측정문항을 추가하거나 기존 측정문항을 빼서 α 점수를 높일 필요가 있을 수도 있다.



크론바흐 알파

- 앞의 공식에 나온 k 는 합성지수를 계산하기 위해 사용된 변수의 수이다. 상식적으로 생각해서 $k = 2$ 인 경우 $2/1 = 2$ 으로 상당히 큰 가중치를 부여하게 되지만, $k = 5$ 인 경우 $5/4 = 1.25$ 로 훨씬 작은 가중치를 부여하게 된다.
- 한편 $k = 10$ 인 경우 $10/9 = 1.11$ 로 변수가 두 배 많이 들어갔는데 줄어드는 정도가 그만큼 크지 않다. 따라서 쓸데없는(redundant) 변수를 억지로 많이 집어넣었을 때도 α 값이 커지는 단점이 있다.
- 말할 필요도 없지만 측정항목이 하나라면 그것만 가지고 합성지수를 만들 수 없다. 최소한 두 개 이상은 필요하기 마련이다.
- 크론바흐 알파는 오로지 내적 일관성 신뢰도를 측정하기 위한 방법일 뿐이므로 신뢰도 일반을 완전하게 측정하는 지표는 아니다.



Stata에서 실습해보자.

- renpainters.dta에서 모든 변수들을 토대로 하나의 합성지수를 만들고자 한다.
- 계산된 α 값에 근거할 때 새롭게 계산된 합성지수인 “미적 평가”는 내적 일관성을 가지고 있다고 결론내릴 수 있을까?
- (만약 그렇다면) 이제 모든 변수들을 합산하거나 주성분분석을 통해 합성지수를 만들자. 새로운 합성지수의 히스토그램을 만들어보자.



크론바흐 알파

- Stata는 자동적으로 역부호화(reverse-coding)를 수행한다.
- 어떤 측정항목을 제거함으로써 α 값을 크게 높일수 있다면 제거를 고려해 볼 수 있다.
- α 값은 개별 측정항목을 정규화(normalization)했는지 여부에 따라 달라진다.
- 개별 측정항목들의 단위가 다른 경우라면 정규화 여부에 따라 α 값에 제법 큰 차이가 날 수도 있다. 편리하게 정규화하기 위해서 std 옵션을 사용할 수 있다.
- 여러 개의 더미 변수가 있을 때도 α 값을 계산할 수 있다.



“아니 이제와서 좀 뜬금없지만, 내적 일관성 이런걸 대체 왜 하지?”

- 지금까지 우리는 통계분석 기법을 꽤 많이 배웠다. 그런데 χ^2 검정을 제외한) 이 모든 기법들은 양적변수가 있어야만 쓸 수 있다!
- 문제는 일반적인 사회조사에서 그런 변수가 좀 드물다는 점이다. 어떤 사회학적 변수들이 양적변수일까? 소득(원 단위), 키, 몸무게, BMI, 그리고 또?
- 우리가 궁금한 수많은 사회학적 변수들은 대부분 질적변수다. 성역할 태도, 우리사회의 인지된 공정성, 직무 만족도/소외감, 인지된 기후위기 심각성 등은 대체로 질적척도로 측정된다.
- 게다가 많은 사회학적 개념들은 다차원적이다. 결국 몇 개의 리커트 척도로 물어보고 이들을 합산하여 단일한 개념을 표상하는 양적변수로 만들 수 밖에 없다.



실제 응용

주성분분석을 하려면 독립변수들이 등간척도 혹은 비율척도여야 한다.

- 앞서 공부했듯 주성분 PC 은 요인적재량 ω 및 변수 X 와 선형결합(linear combination)되어 있다.
- 그러므로 독립변수가 명목척도 또는 서열척도인 경우 **비선형 주성분분석(nonlinear principal component analysis)**을 사용하는 것이 원칙이다.
- 그러나 현실적으로 사회과학 연구에서는 리커트 척도 변수들을 사용하는 경우가 무척 흔하다.



주성분분석이 필요한가 또는 적절한가 여부를 판단하는 절차도 있다.

- 상관계수행렬을 그려보고 상관계수가 0.4보다 큰 변수들이 충분히 많이 있는지 확인해보면 된다(Why?).
- 이 아이디어를 공식화한 것이 (1) 구형성 검정(Bartlett's Sphericity Test) 그리고 (2) Kaiser-Meyer-Olkin (KMO)의 표집 적절성 점수(measure of sampling adequacy)이다.
- KMO는 변수간 상관관계와 부분상관관계(partial correlation)를 비교하는 지표로 이해하자.



실제 응용

- 통계 컨설턴트에게 맡긴 경우 (맥락이 동떨어져 있는데도) 이런 지표들을 열심히 보고하는 경향이 있다.
- 왜 그다지 필요가 없을까? 주성분분석을 수행할 단계에서 우리는 이미 기존 문헌에서 개발되고 타당화까지 완료된 개념과 그 측정항목을 사용하기 때문이다.
- 측정 이론과 척도 개발(scale development) 분야 전공자가 아닌 이상 새로운 측정도구를 함부로 고안/개발하려고 시도하는 것은 너무 위험하다.
- 기존 문헌/설문지 자료 등을 꼼꼼히 읽고 거기에서 연구자들이 식별해 놓은 측정도구를 그대로 사용해야 한다(다행히 모두가 궁금해 할 만한 대부분의 사회학적 현상들은 이미 개발된 측정도구가 있다).
- 바로 이 맥락 때문에 앞서 언급한 구형성 검정이나 KMO는 주성분분석을 수행할 때 그렇게 심각한 의미는 갖지 못하게 된다.



- 기존 척도 연구에서 타당도가 여러 차례 입증된 측정도구(e.g., 직무만족도)라면 구성 문항들로 주성분분석을 수행했을 때 하나로만 묶인다.

나 14 현재 _____님께서 주로 하시는(맡고 계시는) 일에 대한 _____님의 평소 생각은 어떠하십니까? 다음의 각 항목에 대해서 얼마나 동의하시는지 혹은 동의하지 않으시는지 정도를 표시해 주십시오.

| 현재 하시는 일(업무, 직무)에 대한 _____님의 생각 | 전혀 그렇지 않다 | 그렇지 않은 편이다 | 보통이다 | 그런 편이다 | 아주 그렇다 |
|--|-----------------|------------------|------|-----------|-----------|
| (1) 나는 현재 하고(맡고) 있는 일에 만족하고 있다 | 1 | 2 | 3 | 4 | 5 |
| (2) 나는 현재 하고(맡고)있는 일을 열정적으로 하고 있다 | 1 | 2 | 3 | 4 | 5 |
| (3) 나는 현재 하고(맡고)있는 일을 즐겁게 하고 있다 | 1 | 2 | 3 | 4 | 5 |
| (4) 나는 현재 하고(맡고)있는 일을 보람을 느끼면서 하고 있다 | 1 | 2 | 3 | 4 | 5 |
| (5) 별다른 일이 없는 한 현재 하고(맡고)있는 일을 계속하고 싶다 | 1 | 2 | 3 | 4 | 5 |

Stata에서 적절성 평가를 연습하자.

- 먼저 `factortest` 라는 사용자 작성 명령어(user-written command)를 설치해야 한다. 바렛의 구형성 검정을 수행하자. 영가설은 무엇인가? 이를 기각할 수 있는가?
- 다음으로 내부 명령어로 KMO를 확인해보자.
- 개별 변수에 대응하는 값과 종합(overall)이 0.5보다 크면 문제가 없는 것으로 알려져 있다. 0.6 이상이 바람직하다.



연습문제

연습문제

연습 1. 다음 SPSS 자료([링크])를 사용하여 주성분분석을 수행하시오 (import spss 명령어를 사용하여 이 자료를 Stata에서 열 수 있다).

- 계산된 요인적재량과 고윳값에 대해 설명하시오.
- 우울감 측정도구인 C1에서 C20까지 문항들을 몇 개의 주성분으로 추출하는 것이 좋은가? 근거는 무엇인가?
- 적절한 회전을 수행하고 표로 나타낸 뒤, 각 요인에 대해 의미를 부여해보자. 각각의 경우 주성분 간의 상관계수행렬은 어떠한가?
- 회귀기반 점수로 잠재변수를 만든 뒤, 단순합 합성지수인 CESD와는 어떻게 다른지 평가해보자.
- 각각의 주성분 안에서 내적 일관성 점수는 어떠한지 평가하시오. 이를 토대로 주성분분석을 달리 해야 한다면 다시 수행하시오.



연습 2. candidates.csv를 사용하여 주성분분석을 수행하시오(import delimited 명령어를 사용하여 이 자료를 Stata에서 열 수 있다).

- 계산된 요인적재량과 고윳값에 대해 설명하시오.
- 지원자의 인성(personality) 특정 차원을 측정하는 여러 변수들을 이제 몇 개의 주성분으로 추출하는 것이 좋은가? 근거는 무엇인가?
- 적절한 회전을 수행하고 표로 나타낸 뒤, 각 요인에 대해 의미를 부여해보자. 각각의 경우 주성분 간의 상관계수행렬은 어떠한가?
- 회귀기반 점수로 잠재변수를 만든 뒤, 단순합 합성지수와 어떻게 다른지 평가해보자.
- 각각의 주성분 안에서 내적 일관성 점수는 어떠한지 평가하시오. 이를 토대로 주성분분석을 달리 해야 한다면 다시 수행하시오.

