# 계량분석

## Internal Reliability and Cronbach's Alpha

김현우, PhD[1]

[1] 충북대학교 사회학과 조교수

December 10, 2021

# 진행 순서

1. 주성분분석(계속)
2. 신뢰도
3. 크론바흐 알파

# 주성분분석(계속)

# 주성분분석(계속)

요인적재량의 해석을 돕기 위해 회전(rotation)을 수행할 수 있다.

- 회전은 사실 행렬과 삼각함수를 배워야 이해할 수 있는 개념이다. 이른바 회전행렬 (rotation matrix)이 있어 이것을 $\omega$들의 행렬에 곱하면 그래프의 축을 회전시킬 수 있다.
- 다음 페이지의 왼쪽 그림을 참고로 $X_1$ 의 대응하는 두 주성분의 값을 확인해보자. 이때 축을 뒤틀면 어떻게 될까? 다음 페이지의 오른쪽 그림에서는 축이 뒤틀려 대응하는 두 주성분의 값이 달라졌다. 이 과정을 회전이라고 부른다.
- 회전의 원리는 상당히 복잡하다. 그런데 X축과 y축이 본래 직교하듯(90도 관계), 직교의 제약을 여전히 유지하면서 x축과 y축을 뒤틀 수 있다. 이것을 직교 회전 (orthogonal rotation) 내지 대각 회전이라고 부르는데 가장 유명한 알고리즘은 베리맥스(varimax)이다. Stata의 디폴트도 이것이다.

충북대학교
CHUNGBUK NATIONAL UNIVERSITY

# 주성분분석(계속)

- 대각 회전의 치명적인 단점은 “주성분 사이에 상관관계가 없다”는 강한 제약을 유지한다는 것이다. 이 단점에도 불구하고 대각 회전은 수학적으로 간결하다 (그러므로 아름답다). 그리고 이 가정이 이론적으로 정당화될 수 있다면(혹은 되어야 한다면!) 괜찮다.
- 대각 회전 대신 사각 회전(oblique rotation)은 x축과 y축 사이의 직교 제약을 완화하여 따로따로 회전시킨다. 그리하여 주성분 사이에 상관관계를 허용한다.
- 이론과 무관하게 데이터 분석의 입장에서 본다면 사각 회전을 해보고 주성분 사이의 통계적으로 유의한 상관계수가 있을 때 사각 회전으로 갈 수 있다.
- 다만 사회통계학을 응용하는 입장에서는 실용적으로 둘 다 해보는 것이 바람직하다. 큰 차이가 있다면 사각 회전을 따라가는 것을 추천한다. 차이가 없다면 대각 회전을 유지하고 각주 등으로 사각 회전을 확인해보았지만 차이가 없었다고 언급해 둘 수 있다.

충북대학교
CHUNGBUK NATIONAL UNIVERSITY

# 주성분분석(계속)

Stata에서 주성분분석 이후 회전을 연습해보자(**[Stata 코드]** 참고).

- 다시 renpainters.dta를 열고 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.

충북대학교
CHUNGBUK NATIONAL UNIVERSITY

# 주성분분석(계속)

회전 이후에도 해석이 모호하다면 능동적으로 내용 구성을 바꾸어야 한다.

- 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 탈락시키는 것을 고려해 볼 수 있다.
- 반대로 둘 이상의 주성분에 대해 어중간하게 높은 요인적재량을 갖는 경우를 "교차적재량(cross-loading)이 높다"고 표현한다. 이것도 바람직하지 않은 상황이다. 이러한 변수 역시 탈락을 고려할 수 있다.
- 논문이나 보고서에 결과표를 제시할 때는 0.4 보다 작은 요인적재량은 아예 표기하지 않는 것도 보기 좋다.

충북대학교
CHUNGBUK NATIONAL UNIVERSITY

# 주성분분석(계속)

실제 데이터를 가지고 회전을 연습하자([Stata 코드] 참고).

- 아까 Y1_STD_EDU.csv로 데이터로 돌아가 먼저 대각 회전을 실시하자. 요인적재량은 어떻게 바뀌었는가? 요인적재량이 0.4보다 작은 것은 아예 나타나지 않게 하여 관찰해보자. 이 요인적재량에 근거하여 주성분을 생성한다면 어떨까?
- 요인적재량과 교차적재량을 면밀히 살펴가면서 어떤 문항을 제거하면 주성분분석을 통해 더 깔끔한 결과를 얻을 수 있을지 실험해보자.
- 최종적으로 주성분들을 도출한 뒤, 임의의 두 주성분 사이의 상관계수를 확인해보자. 이 값은 반드시 0이다(Why?).
- 만일 네 개의 주성분을 추출한다면 이것들은 전체 데이터 변량의 몇 퍼센트를 설명할까?
- 각각의 주성분에 이름과 이론적으로 실제 의미를 부여해보자. 각각의 변수들이 이론적으로 예상한 바대로 적절한 주성분 안으로 분류되었는가?
- 시험 점수를 종속변수로 하고 주성분들을 독립변수로 하는 회귀분석을 수행하고 그 결과를 해석해보자.

# 주성분분석(계속)

주성분분석이 필요한가 또는 적절한가 여부를 판단하는 (큰 의미를 갖지 않는) 절차도 있다.

- 주성분분석은 상관계수행렬(혹은 공분산계수행렬)을 사용하므로 원칙상 모든 변수는 숫자형 척도여야 한다. 그러나 관행상 거의 전 분야에서 리커트 척도가 아무렇지도 않게 사용된다.
- 요인분석이 적절한가를 확인하기 위해 다음의 두 가지 절차가 제시되고 있다.
- 상관계수행렬을 그려보고 상관계수가 0.4보다 큰 변수들이 충분히 많이 있는지 확인한다. 이 절차를 공식화한 것이 바렛의 구형성 검정(Barlett's Sphericity Test)와 카이저-마이어-올킨(Kaiser-Meyer-Olkin; KMO)의 표집적절성 점수(Measure of Sampling Adequacy)이다.
- 통계 컨설턴트에게 맡긴 경우 (맥락이 동떨어져 있는데도) 이런 지표들을 열심히 보고하는 경향이 있다. 그래서 학위논문에 이 지표들이 굉장히 적극적으로 보고되지만, 주성분분석이 사용될 때 전문적인 학술논문이나 연구보고서에서는 생각처럼 열심히 보고되지 않는다.

# 주성분분석(계속)

- 맥락을 되짚어보면 주성분분석을 수행할 때 우리는 이미 타당화된 지표를 사용한다. 다시 말해, 우리는 이미 기존 문헌에서 개발되고 확인적 요인분석을 거쳐 타당화가 완료된 개념과 그 측정항목을 사용한다.
- 그 측정이론(measurement theory)의 전공가가 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 하는 것이다. 이것은 실무자/사회통계학자 레벨에서 볼 때 충분히 납득할 만한 부분이다. 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. 물론 보고해도 전혀 잘못된 것은 아니고 리뷰어가 요구하면 해야 한다.

충북대학교
CHUNGBUK NATIONAL UNIVERSITY

# 주성분분석(계속)

Stata에서 적절성 평가를 연습하자(**[Stata 코드]** 참고).

- 이번에도 renpainters.dta를 가지고 연습해보자.
- 먼저 **factortest** 라는 user-written command를 설치해야 한다. 바렛의 구형성 검정을 수행하자. 영가설은 무엇인가? 이를 기각할 수 있는가?
- 다음으로 내부 명령어로 KMO를 확인해보자. 개별 변수에 대응하는 값과 종합 값이 0.5보다 크면 문제가 없는 것으로 알려져 있다. 결과는 어떠한가?
- 같은 테스트를 Y1_STD_EDU.csv에 대해서도 수행해보자.

충북대학교
CHUNGBUK NATIONAL UNIVERSITY

# 신뢰도

# 신뢰도

- 어떤 측정도구에 신뢰도가 확보되어 있는가를 확인하기 위한 몇 종류의 기본적 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).
- 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.

충북대학교
CHUNGBUK NATIONAL UNIVERSITY

# 신뢰도

- 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 $\alpha$)를 사용하는 것이 보통이다.
- 리커트 척도나 심지어 더미변수를 사용했을 때도 대체로 잘 작동한다. 방법론 전공자들 사이에서는 크론바흐 알파에 대한 비판적 연구가 이미 많이 나왔지만 (Sijtsma 2009), 여전히 널리 쓰이기 때문에 적어도 2021년 현재에도 크게 걱정할 필요는 없어 보인다.

충북대학교 CHUNGBUK NATIONAL UNIVERSITY

가장 먼저 신뢰도 비율($\rho$)의 논리를 이해해야 한다

- 경험과학의 맥락에서 어떤 사회 현상이나 그에 관한 개념은 보통 다차원적이다. 단일한 측정항목을 통해 이러한 다차원성을 한번에 포섭하기란 거의 불가능하다(e.g., “삶의 행복”같은 개념).
- 그러므로 여러 개의 측정항목들을 통해 서로 다른 차원들을 따로따로 측정한 뒤, 개별 측정항목들($X_1, X_2, ..., X_k$)을 모두 더해 합성지수(composite index)인 y를 만들 필요가 있다. 측정이론에서는 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.

# 신뢰도

- 모든 수학적 논리가 그러하듯 여기에도 몇가지 가정이 있다.

$$
\begin{aligned}
\mathrm{A1:}&\ X = T + e \\
\mathrm{A2:}&\ E(e) = 0 \\
\mathrm{A3:}&\ \rho_{eT} \equiv Cov(e, T) = 0 \\
\mathrm{A4:}&\ \rho_{e_i e_j} \equiv Cov(e_i, e_j) = 0
\end{aligned}
$$

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

충북대학교
CHUNGBUK NATIONAL UNIVERSITY

# 신뢰도

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

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

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

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

- 이제 신뢰도 비율($\rho$)을 아래와 같이 측정할 수 있다(Why?)

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

# 신뢰도

- 그런데 신뢰도 비율은 결국 관찰하는 값(X)과 진정한 값(T) 사이 상관계수의 제곱 ($r^2_{XT}$)이다(Why?).

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

- 수리적 증명 과정(강승호·김양분 2004)은 생략하고 신뢰도 비율의 실질적 의미를 반추해보자! 신뢰도란 결국 진정한 값에 대해 개별 문항들이 얼마나 높은 상관계수의 제곱을 가지는가를 측정한 것이다.
- 모든 신뢰도 평가법(재검사법, 동형검사법, 반분법 등)은 궁극적으로 상관관계로부터 신뢰도를 측정한다.

# 크론바흐 알파

크론바흐 알파(Cronbach's $\alpha$)의 논리를 직관적으로 살펴보자.

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

$$\rho = \frac{\sigma_{\mathrm{y}}^2 - \sum \sigma_{\mathrm{X_i}}^2}{\sigma_{\mathrm{y}}^2} = 1 - \frac{\sum \sigma_{\mathrm{X_i}}^2}{\sigma_{\mathrm{y}}^2}$$

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

충북대학교
CHUNGBUK NATIONAL UNIVERSITY

# 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.

$$\alpha = \left(\frac{k}{k-1}\right)\left(\frac{S_y^2 - \sum S_{X_i}^2}{S_y^2}\right)$$
$$= \left(\frac{k}{k-1}\right)\left(1 - \frac{\sum S_{X_i}^2}{S_y^2}\right)$$

- 분자에 들어간 $\sum S_X^2$ 가 $S_y^2$ 보다 상대적으로 커지면 $\alpha$는 감소한다(Why?).
- $\alpha$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 $\alpha$ 점수를 높일 필요가 있을 수도 있다.

충북대학교
CHUNGBUK NATIONAL UNIVERSITY

# 크론바흐 알파

- 앞의 공식에 나온 k는 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 중복되는(redundant) 변수를 억지로 많이 집어넣었을 때도 $\alpha$ 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.

충북대학교
CHUNGBUK NATIONAL UNIVERSITY

# 크론바흐 알파

eCampus에서 Fear_of_Statistics.xlsx를 다운받아 미국 대학생들의 "통계학에 대한 두려움"에 관한 합성지수를 계산하고 그것의 내적 일관성을 평가해보자.

- 이 변수들은 "통계학에 대한 두려움"이라는 개념의 다양한 차원(측면)을 측정하고 있다. 이들을 모두 더하여 하나의 합성지수를 만들어 그것을 잠재변수로 사용하면 좋을 것 같다. 문제는 이 잠재변수가 얼마나 내적 일관성이 있는가 아직 모른다는 점이다.
- 계산된 $\alpha$ 값에 근거할 때 새롭게 계산된 합성지수인 "통계학에 대한 두려움"은 내적 일관성을 가지고 있다고 결론내릴 수 있을까?
- (만약 그렇다면) 이제 모든 변수들을 합산하여 합성지수를 만들자. 새로운 합성지수의 히스토그램을 만들어보자(이때 width는 1로 조정하자). 실무나 연구 맥락에서 매우 중요한 절차는 "이 합성지수가 큰 값을 가지면 무엇을 뜻하는가?" 를 분명히 확인해 두는 것이다. 이 방향성과 일치하지 않는 변수는 사용하지 않거나 역코딩해야 한다.

충북대학교 CHUNGBUK NATIONAL UNIVERSITY

# 크론바흐 알파

- 통계분석 패키지가 자동적으로 역코딩을 수행한다. 그러나 이것이 정말로 내가 원하는 방향인지는 장담할 수 없으므로 항상 수동으로 이를 제어하자.
- 어떤 측정항목을 제거함으로써 $\alpha$ 값을 크게 높일수 있다면 적극적으로 제거를 고려하자. 그것은 어차피 신뢰도를 낮추는 요인일 뿐이다!
- $\alpha$ 값은 개별 측정항목을 정규화(normalization)했는지 여부에 따라 달라진다. 개별 측정항목들의 단위가 다른 경우라면 정규화 여부에 따라 $\alpha$ 값에 제법 큰 차이가 날 수도 있다. 편리하게 정규화하기 위해서 **std** 옵션을 사용할 수 있다.
- 여러 개의 더미 변수가 있을 때도 $\alpha$ 값을 계산할 수 있다.

# 크론바흐 알파

"아니 이제와서 좀 뜬금없지만, 타당도나 신뢰도 이런걸 대체 왜 따지지?"

- 지금까지 우리는 숫자형 척도를 따르는 변수를 분석하는 기법을 꽤 많이 배웠다. 그런데 문제는 ($\chi^2$ 검정을 제외한) 이 모든 기법들은 적어도 하나 이상 숫자형 변수가 있어야만 쓸 수 있다는 점이다.
- 문제는 일반적인 사회조사에서 그런 변수가 좀 드물다는 점이다. 어떤 변수들이 숫자형 변수일까? 소득(원 단위), 키, 몸무게, BMI, 또?
- 우리가 궁금한 수많은 사회학적 변수들은 대부분 범주형 변수다. 성역할 태도, 우리사회의 인지된 공정성, 직무 만족도/소외감, 인지된 기후위기 심각성, 최근 3일간 만나 이야기한 친구의 수(Why?) 등은 대체로 범주형 척도로 측정된다.
- 게다가 많은 사회학적 변수들은 다차원적이다. 결국 개별적으로 부각되는(salient) 차원들을 개별적인 (리커트) 측정문항으로 물어본 다음, 단일한 개념을 전달하는 숫자형 변수로 통합하기 위해 나중에 적절히 합칠 수 밖에 없는 것이다.

# 크론바흐 알파

- 회귀분석은 현대통계학의 핵심이다. 그러나 학부과정 및 석사과정 초년생 레벨에서 회귀분석에는 중대한 제약조건이 있다. 그것은 종속변수가 반드시 숫자형 변수여야 한다는 점이다.
- 아까 언급했듯, 일반적인 사회조사에서는 숫자형 척도로 측정된 변수가 좀 드물다. 그러므로 우리는 관심있는 사회학적 측정문항들을 합쳐 합성지수를 만들어낸 뒤, 이것이 (구성) 타당성이 있고 (내적 일관성) 신뢰성도 있는 지수임을 재확인해야 한다.
- 이렇게 만들어진 합성지수는 비록 진정한 의미에서 숫자형 척도라고는 할 수 없지만 그에 근접한 것으로 여겨진다.
- 독립변수의 경우 숫자형 변수를 넣을 수도 있지만, 설령 범주형 변수라도 더미 변수로 변환하여 넣을 수 있으니 애시당초 문제가 없다. 물론 필요에 따라 합성지수인 독립변수를 사용할 수도 있다.

충북대학교
CHUNGBUK NATIONAL UNIVERSITY