

사회통계연습

합성지수의 신뢰도

김현우, PhD¹

¹충북대학교 사회학과 부교수



진행 순서

- 1 측정오차
- 2 신뢰도와 내적 일관성
- 3 크론바흐 알파 연습



측정오차



측정오차

오차는 경험과학적 연구에 있어서 필연적인 요소이다.

- (정치, 종교, 예언 등과는 달리) 경험과학적 연구는 틀릴 수 있다는 가능성에서 출발하여 오차(error)가 일어나는 조건과 크기에 주목한다.
- 과학적 활동으로서의 모든 관찰(observation)에는 어느 정도 측정오차(measurement error)가 뒤따르기 때문이다.
- 다양한 원인과 맥락에서 사람이 측정하고자 하는 참 값(true score)과 측정된 값(observed score) 사이에 괴리가 발생한다.

$$X = T + \epsilon$$

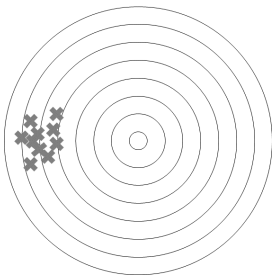
- 오차로 말미암아 측정도구(instrument)의 타당도(validity)와 신뢰도(reliability)가 훼손된다.



측정오차

타당도는 재고자 한 개념을 얼마나 충실하게/제대로 측정했는가의 문제다.

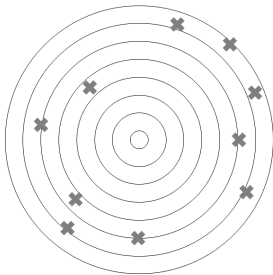
- “the degree to which it measures what it is supposed to measure”
- 고래가 체중계에 올라섰는데 계속 40kg이 나온다면 타당도가 낮은 측정이다.



측정오차

신뢰도는 재고자 한 개념을 일관성있게 측정했는가의 문제다.

- 고래가 체중계에 올라설 때마다 다른 값이 나온다면 신뢰도가 낮은 측정이다.



측정오차는 여러가지 맥락에서 기인한다.

- 조사자 또는 측정도구의 맥락(e.g., 설문 내용의 모호함 등)
- 응답자 맥락(e.g., 질문에 대한 귀찮음, 오해, 무응답 등)
- 검사 과정/환경의 맥락(e.g., 장소가 지나치게 산만함 등)
- 조사자-응답자 상호작용(e.g., 인터뷰어-인터뷰이 권력거리/성차 등)



측정오차

- 오차로 인해 **사회조사의 질(quality)**은 결정적으로 영향받는다!
- 사회조사 단계에서 **선행조사(pilot test)**, **피드백(feedback)**, **조사원 훈련** 등에 많은 노력을 기울여야 한다.
- (조사전문가 커리어와 관련하여) **표본조사설계(Survey Research Design)** 수업이 아쉽게도 우리 학교에서는 개설되지 않는다.
- 물론 실무에 직접 뛰어들어 노하우를 쌓는 것이 아무래도 제일 중요하다.



측정오차는 체계적 오차와 무선오차로 분해된다.

- 체계적 오차(systematic error)는 일정한 방향으로 치우친(bias) 오차다.
- '일정한 방향'이라는 점에서 이런 오차는 (잘하면) 설명될 수도 있다.
- 측정도구가 일관성 있게 잘못 작동하여 수집된 자료가 참 자료와 일정하게 어긋나 있는 경우에 해당한다(e.g., 체중계의 영점 조정이 안되어 일정하게 5kg씩 높게 측정함).
- 체계적 오차로 인해 측정된 값과 참 값 사이에는 평균(expected value)이 다르게 된다.
- 측정의 체계적 오차로 인해 타당도의 문제가 나타난다.



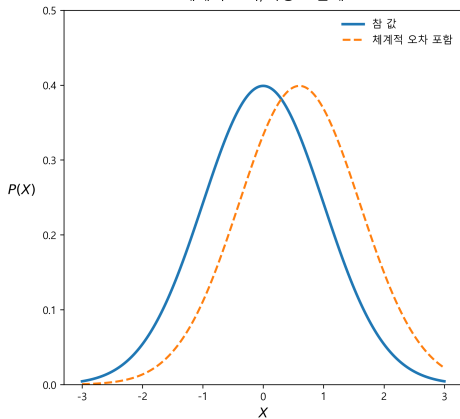
측정오차

- 무선오차(random error)는 임의(random)의 방향으로 흩어진(variance) 오차다.
- 이것은 ‘임의적’이므로 설명될 수 없다.
- 이때 무선(無線)이란 선을 그릴 수 없다는 의미에서 만들어진 일본어다(wireless와는 다르다!).
- 전통적인 사회과학에서는 비교적 무해하고 자연스러운 것으로 여겨졌다(e.g., 디지털 체중계의 전기적 오류로 $\pm 1\text{kg}$ 씩 다르게 측정함).
- 무선 오차로 인해 측정된 값과 참 값 사이에는 분산(variance)이 다르게 된다.
- 측정의 체계적 오차로 인해 신뢰도의 문제가 나타난다.

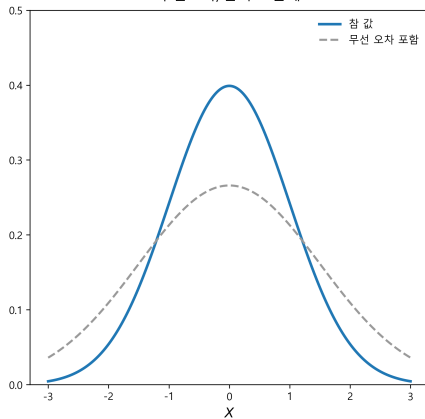


측정오차

체계적 오차/타당도 문제



무선오차/신뢰도 문제



타당도와 신뢰도 사이의 관계는 미묘하다!

- 측정오차 = 체계적 오차 + 무선오차.
- 당연히 타당도와 신뢰도 둘 다 높을수록 좋다.
- 현실적으로 양자는 **교환관계(trade-off)** 관계에 놓여 있으므로, 양자 간에 균형을 맞추는 것이 중요하다.
- 경우에 따라 경제학이나 정책과학 등 몇몇 사회과학적 관점에서는 종종 타당성의 훼손, 즉 **편의(bias)**를 더 줄이기 위해 신뢰도의 훼손, 즉 **분산(variance)**을 감내하기도 한다.



참 값이란 당연히 알 수 없다!

- 측정없이 참을 안다는 것은 인간으로서는 불가능하다. 애시당초 그걸 알면 구태여 “측정된 값”을 구할 필요조차 없다.
- 그러므로 보통 (1) **측정된 값(X)-참 값(T)**을 계산할 수 없어 측정오차 ϵ 이 얼마나 큰지 알 수 없고, (2) 측정오차 안에서 체계적 오차와 무선오차가 각각 얼마나 큰지도 알 수도 없다.
- **측정이론(measurement theory)**에서는 측정도구의 타당도와 신뢰도를 평가하는 여러가지 통계학적 논리와 기술들을 발전시켜 왔다.
- 그 중 오늘 우리가 공부하는 내용은 신뢰도를 평가하는 기법이다.



신뢰도와 내적 일관성



신뢰도와 내적 일관성

제대로 된 사회조사라면 일정 이상의 신뢰도를 반드시 갖춰야 한다.

- 사회과학에서 측정도구는 적어도 둘 이상의 측정항목으로 이루어져 있는 것이 보통이다.
- 하나의 질문만으로 특정 개념의 **다차원성(multi-dimensionality)**을 포괄하기 어렵기 때문이다(Why?).
- 가령 사랑은 어떻게 측정할 수 있을까? 종교성(religiosity)은? 직장만족도는?
- 결국 다문항 측정도구에 얼마나 **일관성(consistency)** 혹은 **재생산가능성(reproducibility)**이 있는가가 사회조사에서 신뢰도의 문제가 된다.



신뢰도와 내적 일관성

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

〈한국노동패널〉 2020년도 자료 직장만족도 문항



신뢰도와 내적 일관성

측정도구의 신뢰도를 확인하는 세 종류의 방법이 있다.

- **재검사 신뢰성(test-retest reliability)**은 같은 측정도구를 한 번 더 시험해서 앞서 얻은 결과와 얼마나 일관되어 있는지 확인하는 방법 내지 그렇게 확인할 수 있는 종류의 신뢰도이다.
- **반분 신뢰성(split-half reliability)**은 측정도구의 문항들을 임의로 나누어 시험한 뒤, 두 세트의 결과가 얼마나 일관되어 있는지 확인하는 방법 내지 그렇게 확인할 수 있는 종류의 신뢰도이다.
- **내적 일관성(internal consistency)**은 문항들 사이에서 상관관계가 높다면 이에 근거한 측정도구의 (내적) 일관성이 높음을 확인하는 방법 내지 그렇게 확인할 수 있는 종류의 신뢰도이다.



신뢰도와 내적 일관성

- (1) 재검사법과 (2) 반분법은 시험을 위한 시험을 치러야 하므로 방법론 자체의 연구가 아니면 보통 선호되지 않는다.
- 실제 연구와 조사에서는 (3) 내적 일관성 점수인 **크론바흐 알파(Cronbach's α)**를 흔히 사용한다.
- 여러 비슷한 문항들이 정말로 신뢰성 높다면 그것들의 측정 결과 각각은 내적 일관성이 높을 것이다.
- 크론바흐 알파가 높다면, 내적 일관성에 대해 안심하고 개별 문항들 x_1, x_2, \dots, x_k 을 모두 더해 **합성지수(composite index)** X 를 만들 수 있다.



신뢰도와 내적 일관성

크론바흐 알파에 앞서 신뢰도 비율 ρ 을 이해해야 한다

- (과녁판의 비유에서 살펴보았듯) 개별 측정항목의 높은 분산 σ^2 은 곧 해당 측정항목의 높은 불확실성을 의미한다. 그러므로 개별 측정항목들의 분산합 $\sum \sigma_{x_i}^2$ 은 결국 사용된 측정항목들의 전반적인 불확실성을 반영한다(Why?).
- 만약 합성지수의 분산 σ_X^2 에서 개별 측정항목들의 분산합 $\sum \sigma_{x_i}^2$ 의 비중이 크다면, 이는 “일관되는지 불확실한” 개별 측정항목들로 그 합성지수를 만든 것이므로 (그 합성지수는) 신뢰도가 낮다고 할 수 있다.

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

- 만일 $\sum \sigma_{x_i}^2 = \sigma_X^2$ 이면 $\rho = 0$ 이고, $\sum \sigma_{x_i}^2 = 0$ 이면 $\rho = 1$ 이다.



신뢰도와 내적 일관성

크론바흐 알파는 변수가 늘어나는 것에 약간 패널티를 준 신뢰도 비율이다.

- 합성지수를 구성하는데 사용한 변수의 수 k 로 적절히 가중된 신뢰도 비율을 계산에 사용한다.

$$\alpha = \left(\frac{k}{k-1} \right) \left(1 - \frac{\sum s_{x_i}^2}{s_X^2} \right)$$

- 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.
- 새로운 측정문항을 추가하거나 기존 측정문항을 빼서 α 점수의 변동을 살펴보아야 한다!



크론바흐 알파 연습



크론바흐 알파 연습

연습 1. FEAR.SAV에서 통계학에 대한 두려움이라는 합성지수를 구성하고 그 기술통계를 살펴본 뒤, 분포를 시각화하시오. 이 합성지수의 신뢰도를 보고하고 내적 일관성 여부를 판정하시오.



연습 2. Robert Sampson과 그의 동료들에 따르면 집단효능감 (collective efficacy)은 “collective efficacy [is] defined as social cohesion among neighbors combined with their willingness to intervene on behalf of the common good”이라고 정의된다. CRIME.SAV 자료를 활용하여 집단효능감(collective efficacy)에 관한 합성지수를 구성하고 그 기술통계를 살펴본 뒤, 분포를 시각화하시오. 이 합성지수의 신뢰도를 보고하고 내적 일관성 여부를 판정하시오.



“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.
- 결국 몇 개의 리커트 척도로 물어보고 이들을 합산하여 단일한 개념을 표상하는 양적변수로 만들어야 하는 경우가 많다.



“그러면 새로운 개념은 어떻게 척도화하지?”

- 그건 묻지 말자! 이것은 (더이상 신뢰도가 아니라) 타당도의 문제가 되고 주성분 분석(principal component analysis)을 배워야 하기 때문이다.
- 주성분분석은 일단 고유값(eigenvalue)과 고유벡터(eigenvector) 개념을 이해하면 어이없을 만큼 쉽지만, 막상 두 개념을 알기 위해서는 선형대수학(linear algebra) 수업을 한 학기 들어야 한다.



크론바흐 알파 연습

- 그러므로 여러분은 개념들의 측정도구를 함부로 고안/개발하려고 해서는 안된다.
- 기존 문헌/설문지 자료 등을 꼼꼼히 읽고 거기에서 연구자들이 식별해 놓은 측정도구를 그대로 사용하자. 이것은 실무자 레벨에서 볼 때 충분히 납득할 만한 부분이다.
- 다행히 여러분이 궁금해 할 만한 대부분의 사회학적 현상들은 이미 개발된 측정도구가 있다(아까 본 직장만족도 문항처럼 행렬식 질문을 할 때가 많다!).



크론바흐 알파 연습

- 여러분이 합성지수의 신뢰도를 계산해서 .7보다 크다는 것을 보이면, 이제 내적 일관성을 확인한 것이다.
- 따라서 합성지수를 따로 구성해 만들지 않았다면 내적 일관성 이야기는 (조사방법론의 맥락에서) 나올 이유가 없다.
- (말할 필요도 없지만) 측정항목이 하나 뿐이라면 그것만 가지고 합성지수를 만들 수 없다. 최소한 두세 개 이상의 측정항목이 필요하다!



게다가 우리는 곧 회귀분석을 연습한다!

- 회귀분석이야말로 현대통계학의 척추와도 같다! 그러나 학부과정 및 석사과정 초년생 레벨에서 회귀분석에는 중대한 제약조건이 있다. 그것은 **종속변수가 반드시 양적변수여야 한다**는 점이다.
- 아까 언급했듯, 일반적인 사회조사에서는 숫자형 척도로 측정된 변수가 다소 드물다. 그러므로 우리는 종종 관심있는 사회학적 측정문항들을 합쳐 합성지수를 만들어낸 뒤, 이것이 내적 일관성이 있는 지수임을 재확인하게 한다.
- 이렇게 만들어진 합성지수는 비록 진정한 의미에서 양적 척도라고는 할 수 없지만 그에 근접한 것으로 여겨진다.

