

# 사회통계연습

통계분석의 이해

김현우, PhD<sup>1</sup>

<sup>1</sup>충북대학교 사회학과 부교수



## ① 자료분석의 첫 출발

## 자료분석의 첫 출발

# 자료분석의 첫 출발

연습 1. 통계 분석 프로그램을 구동하여 hsb2.sav 파일을 불러오자.

- Jamovi 또는 SPSS를 기동하기 위해서는 (1) 해당 어플리케이션 아이콘을 더블클릭하거나, (2) 데이터 파일을 더블 클릭하거나, (3) 윈도우 메뉴에서 Jamovi를 찾아 선택한다.
- 각 윈도우가 무엇을 의미하는지 살펴보자.
- 이 자료에서 데이터가 무엇이고 관찰단위가 무엇인지 파악해보자.
- 변수와 관측치는 각각 몇 개인가?
- 이 자료는 누구에게서 수집한건가? 왜 수집했나? 이것을 가지고 무엇을 할 수 있나?

# 자료분석의 첫 출발

자료와 관찰단위 그리고 분석단위의 개념을 되짚어보자.

- 자료(data): 분석 대상의 속성(attributes)을 기록한 수열(series).
- “자료의 관찰”은 사람을 단위로 이루어졌으나, 나중에 “자료의 분석”은 각 사람이 속한 지역 또는 국가 단위로 이루어질 수도 있다.
- 관찰단위(unit of observations): 자료의 관찰과 수집이 이루어지는 단위
- 분석단위(unit of analysis): 분석이 수행되고 발견이 일반화(generalization)될 수 있는 단위

# 자료분석의 첫 출발

열은 변수가 되고, 행은 관측치가 된다.

- 앞서 데이터는 수열이라고 말했지만, 사실 여러 변수(variables)를 갖기 때문에 각각의 변수가 하나씩 수열을 이루어서 결국 행렬(matrix)이 된다. 즉 데이터는 사각형이다.
- 행렬은 행과 열의 조합이다. 행은 가로고 관측치(observations)가 된다. 열은 세로고 변수(variables)가 된다.
- 변수는 변(vary)할 수 있는(able) 숫자이고, 관측치는 해당 변수에 실제 채워진 값들이다.
- 데이터를 남에게 설명할 때 “변수는 몇 개, 관측치는 몇 개”와 같은 식으로 표현할 수 있다.

# 자료분석의 첫 출발

각각의 변수마다 자료유형이 다를 수 있다.

- 자료유형(data type)이 다르다는 것은 보다 구체적으로 측정의 척도(scales of measurement)가 다르다는 것을 뜻한다.
- 측정의 척도란 변수가 정의(define)되고 유형화(categorize) 되는 방식을 뜻한다.
- 보통 네 종류의 척도를 거론할 수 있다: (1) 명목(nominal), (2) 서열(ordinal), (3) 등간(interval), (4) 비율(ratio).

# 자료분석의 첫 출발

네 종류의 척도에는 위계적인 관계가 있다.

- **명목척도(nominal scale)**: 범주(category)가 존재할 뿐, 그들 사이에 우열이나 대소가 없음.
- **서열척도(ordinal scale)**: 범주들 사이에 서열 혹은 순서(order)가 있음.
- **등간척도(interval scale)**: 범주들 사이에 우열이나 대소가 있고, 그 간격은 동등함(등간; equal interval).
- **비율척도(ratio scale)**: 범주들 사이에 우열이나 대소가 있고, 그 간격도 동일하며, 절대영(absolute zero) 또한 의미를 가짐.



# 자료분석의 첫 출발

연습 2. hsb2.sav 파일에서 사용된 변수의 척도를 파악해보자.

변수	척도	변수	척도
id		read	
female		write	
race		math	
ses		science	
schtyp		socst	
prog			

# 자료분석의 첫 출발

통계 분석 프로그램에서 자료유형을 확인하고 변환해보자.

- SPSS나 Jamovi의 한국어 번역은 통계용어규정을 제대로 반영하고 있지 않음에 주의하자.
- 측도는 measure의 번역어이고, **nominal**, **ordinal**, **scale**은 각각 **명목형**, **순서형**, **척도**로 번역되었다.
- 여기서 척도(scale)는 결국 비율척도와 등간척도를 뭉뚱그린 것이다. 사실 이렇게 종종 비율척도와 등간척도는 뭉뚱그려 분석에 사용한다(Why?).
- 명목척도와 서열척도를 **질적변수(qualitative variable)**로, 등간척도와 비율척도를 **양적변수(quantitative variable)**로 묶을 수도 있다.
- 자료를 잘 살펴보고 SPSS나 Jamovi에서 설정된 측도를 옳게 바꾸어보자.
- 척도 재설정은 종종 건너뛰어도 된다(하지만 가끔씩 필수적이다).

# 자료분석의 첫 출발

연습 3. poleff.sav를 불러와 살펴보자.

- **정치적 효능감(political efficacy)**이란 한마디로 정치 영역에서의 “I can do it!”이다.
- 만일 하나의 문항만으로 정치적 효능감을 측정해야 한다면 그 문항은 어떤 것일까?
- 정치적 효능감이 높은 사람은 이 문항에 대해 전반적으로 어떻게 대답할지 예상해보자.

# 자료분석의 첫 출발

- 자료에서 주어진 각각의 문항은 어떤 척도로 볼 수 있을까?
- 종종 리커트 척도(Likert scale)는 본래 성질상 서열척도이지만, 편의상 대체로 등간(approximately interval)이라고 부르며 양적척도처럼 분석하기도 한다.
- 각 문항에 대한 응답 총점은 어떤 척도로 볼 수 있을까?
- 같은 개념에 대해 여러 개의 리커트 척도를 활용하여 측정한 뒤, 이를 모두 더하거나 평균을 구한 점수를 “대체로 등간척도”로 보기도 한다.

# 자료분석의 첫 출발

자료유형의 판별은 기초적이지만 매우 필수적이다.

- 거의 모든 통계학이나 데이터분석 수업이 자료유형 식별로부터 출발한다.
- 특히 숙제와 복습을 통해 명목, 서열, 등간, 비율, 또는 대체로 등간인지 여부를 반복숙달해서 익혀야 한다.