사회통계연습
What is Empirical Science?

김현우, PhD[1]

[1] 충북대학교 사회학과 조교수

September 3, 2021

충북대학교
CHUNGBUK NATIONAL UNIVERSITY

진행 순서

1. 경험과학으로의 초대
2. 자료유형(Data Types)
3. 척도의 성질
4. 설문지에서 리커트 척도의 활용

경험과학으로의 초대

1. 경험과학(empirical science)이란 무엇인가?
   - 사회학의 맥락에서 경험과학이란 "사회 현상을 체계적으로 관찰하여 얻어낸 자료(data)를 토대로 분석하여 법칙이나 원리 따위를 개발하는 학문"이다.
   - 넓게 보면 이렇게 "증거를 수집하고 정리하는 체계적인 절차"를 개발하는 것도 경험과학의 일부로 볼 수 있다.
   - 이렇게 "증거를 수집 정리하는 체계적인 절차"를 흔히 방법(method)이라고 하며, 방법론(methodology)이란 방법(method)에 대한 과학이다.
   - 여러분은 사회조사방법론과 사회조사실습을 3학년에 배운다.

충북대학교
CHUNGBUK NATIONAL UNIVERSITY

경험과학으로의 초대

1. 너무나도 당연한 말이지만, 경험과학은 측정도구를 필요로 한다!
   - 측정(measurement)이란 "어떤 현상을 관찰한 후 일정한 규칙에 따라 수치를 부여하는 것"이다.
   - 이런 행위에는 현상을 계량화(quantify)할 수 있다는 철학적 세계관을 내포하고 있다. 어떤 사람은 이런 관념을 배척한다.
   - 경험과학의 개념화는 사실 매우 복잡한 역사를 가지고 있다. 논쟁 끝에 우리가 지금 아는 경험과학이 정립되었다. 어떤 학자들(몇몇 과학철학자들)은 여전히 이 개념화에 반대한다.

충북대학교
CHUNGBUK NATIONAL UNIVERSITY

경험과학으로의 초대

1. 일단 교과서대로라면 경험과학적 연구는 아래와 같은 순서대로 진행된다.
   1. 문제인식: 엄격한 정의, 왜 연구가 필요한가, 기존 연구 속에 위치
   2. 가설설정: 잠정적 진술의 꼴
   3. 연구설계: 누구/무엇을, 언제, 어떻게 연구?
   4. 자료수집: 설문조사? 실험?
   5. 자료분석: 이른바 통계적 분석(statistical analysis)
   6. 가설검정 및 일반화: 가설의 입증 또는 기각
   7. 결과발표: 논문 또는 포스터

2. 물론 현실은 종종 다르다.

충북대학교
CHUNGBUK NATIONAL UNIVERSITY

경험과학으로의 초대

1. 경험과학은 자료(data)를 토대로 한다.
   - 주장에는 증거(evidence)가 필요하고, 증거는 결국 자료를 토대로 한다.

2. 자료(data)에 대해 여러분이 알아야 할 모든 상식!
   - 발음은 데이타(o), 다타(x), 데타(x)
   - 어원은 라틴어 dare -〉 주다(to give).
   - data는 복수. datum은 단수. 일상 생활에서 datum이란 단어를 누가 들어본 적이나 있을까? 그러므로 “Your data are trash.” 라고 해야하지만 who cares?
   - 딱 하나의 숫자는 데이터라고 하지 않음. 그런건 정의상 datum이지만...
   - 일단 data가 주어졌다면 이것은 ”많은 관측치(many observations)”를 담고 있다.

3. 그런데 관측치가 많다보니 아무래도 이걸 요약해야 할 필요가 있겠지?
   - 많은 관측치들을 요약한 한두개의 숫자를 요약통계량(summary statistics)이라고 한다.
   - 요약통계와 같은 의미로 데이터를 기술(記述)하는 것을 기술통계(descriptive statistics)라고 한다.
   - 다음 주에는 요약통계/기술통계에 대해 공부한다.

충북대학교
CHUNGBUK NATIONAL UNIVERSITY

경험과학으로의 초대

1. 통계학의 정의는 "자료(data)를 수집, 요약, 분석하는 이론과 방법"이다.
   - "A body of techniques and procedures dealing with the collection, organization, analysis, interpretation, and presentation of information that can be stated numerically."

     Kuzma, J. W. 1984. Basic Statistics for the Health Sciences. Palo Alto, CA: Mayfield.

2. 사회통계연습에서는 data가 일단 주어졌다고 전제하고 시작한다.
   - 데이터 수집의 설계, 수행, 관리, 보관 등의 이슈는 다른 수업에서 다루어야 한다.
   - 우리 과에서는 사회조사방법론과 사회조사실습이 있지만, 본격적인 서베이방법론, 데이터베이스(SQL 등), 클라우드 컴퓨팅 등은 (적어도 현 시점) 우리 과에서 개설되지 않으므로 다른 과에서 들어야 한다.

3. 우리 과에서는 다음의 순서대로 여러분을 경험과학자로 훈련시킨다.
   - 사회통계(2학년1학기) → 사회통계연습(2학년2학기) → 사회조사방법론(3학년1학기) → 사회조사실습(3학년2학기)

충북대학교
CHUNGBUK NATIONAL UNIVERSITY

자료유형(Data Types)

ID	female	IQ	income	socialclass
1	1	135	250	1
2	0	110	310	2
3	1	128	1500	3
4	0	98	122	2
5	1	106	450	2
6	0	102	190	1

1. 데이터는 사각형이다!
   - 좀 더 수학적으로 표현하자면 행렬(matrix) 꼴이다.
   - 행렬은 행과 열의 조합이다. 행은 가로고, 열은 세로다.
   - 행은 rows (=극장 따위의 좌석 줄)라고 쓰고, 열은 columns (=기둥)이라고 쓴다.

충북대학교
CHUNGBUK NATIONAL UNIVERSITY

자료유형(Data Types)

ID	female	IQ	income	socialclass
1	1	135	250	1
2	0	110	310	2
3	1	128	1500	3
4	0	98	122	2
5	1	106	450	2
6	0	102	190	1

1. 통계학의 맥락에서 "분석적으로 의미있는" 열(columns)은 변수(variables)가 되고, 행(rows)은 관측치(observations; records)가 된다.

- 변수(variables)는 변(vary)할 수 있는(able) 숫자다!
- 위 예제에서 변수(variables)는 모두 4개다.
- 위 예제에서 관측치(observations; records)는 모두 6개다.
- 그러므로 데이터를 남에게 설명할때 이렇게 변수가 몇 개, 관측치가 몇 개 이런 식으로 표현할 수 있다.

충북대학교
CHUNGBUK NATIONAL UNIVERSITY

자료유형(Data Types)

ID	female	IQ	income	socialclass
1	1	135	250	1
2	0	110	310	2
3	1	128	1500	3
4	0	98	122	2
5	1	106	450	2
6	0	102	190	1

1. 각각의 열(columns) 또는 변수(variables)마다 자료유형(data type)이 다르다.
   - 이것을 두고 척도(scale)가 다르다고 말한다.
   - 아까 경험과학에서는 측정(measurement)을 필요로 한다고 했다.
   - 측정의 척도(scales of measurement)란 변수(variables)가 정의(define)되고 유형화(categorize) 되는 방식을 뜻한다.
   - 네 종류의 척도: 명목(nominal), 서열(ordinal), 등간(interval), 비율(ratio).

충북대학교 CHUNGBUK NATIONAL UNIVERSITY

척도의 성질

ID	female	IQ	income	socialclass
1	1	135	250	1
2	0	110	310	2
3	1	128	1500	3
4	0	98	122	2
5	1	106	450	2
6	0	102	190	1

1. 네 종류의 척도에는 위계적인 관계가 있다.
   - 명목척도(nominal scale)에는 우열이나 대소가 없고, 단지 분류(category)만이 의미를 가진다.
   - 서열척도(ordinal scale)에는 우열이나 대수가 있지만, 순서(order)만 의미를 가진다.
   - 등간척도(interval scale)에는 우열이나 대수가 있고, 그 간격은 동일하다(등간; 동등한 간격; equal interval).
   - 비율척도(ratio scale)에는 우열이나 대수가 있고, 그 간격은 동일하며, 영(zero)도 의미를 가진다.

충북대학교
CHUNGBUK NATIONAL UNIVERSITY

척도의 성질

1. 네 종류의 척도는 각각이 가진 정보량에 차이가 있다.
   - 정보량의 크기: 비율 〉 등간 〉 서열 〉 명목

Measure	Category	Order	Equal interval	Absolute zero
명목(nominal)	O	X	X	X
서열(ordinal)	O	O	X	X
등간(interval)	O	O	O	X
비율(ratio)	O	O	O	O

Stevens, S. S. 1946. "On the theory of scales of measurement." Science 103, 677–680.

충북대학교
CHUNGBUK NATIONAL UNIVERSITY

척도의 성질

1. 명목척도(nominal scale)에는 우열이 없다.
   - 남자=1, 여자=0이라고 입력되었을때 남자의 숫자가 크니까 우월한 게 아니다.

2. 서열척도(ordinal scale)로는 산술연산을 아예 할 수 없다.
   - 왜일까? 예컨대 1+2=3이 성립하려면 2가 1보다 정확히 두 배 커야 하기 때문이다.

3. 등간척도(interval scale)로는 덧셈 뺄셈은 할 수 있지만 곱셈 나눗셈은 못한다.
   - 왜일까? 대수학적으로 절대영(absolute zero)이 정의되지 않기 때문이다.

4. 비율척도(ratio scale)로는 뭐든지 할 수 있다!

충북대학교
CHUNGBUK NATIONAL UNIVERSITY

척도의 성질

ID	female	IQ	income	socialclass
1	1	135	250	1
2	0	110	310	2
3	1	128	1500	3
4	0	98	122	2
5	1	106	450	2
6	0	102	190	1

1. 척도 간에는 변환(recoding)이 가능하지만 오로지 일방향으로만 가능하다!

- 정보량에 차이가 있기 때문에 변환하는 순간 "사라진 정보"는 복원 불가능하게 된다.
- 예를 들면 월평균 소득은 ratio로 측정되었지만 다음과 같이 ordinal로 변환할 수 있다.
- e.g., 100만 미만 $\rightarrow$ 1; 100만–300만 $\rightarrow$ 2; 300만–500만 $\rightarrow$ 3; 500만 이상 $\rightarrow$ 4

충북대학교
CHUNGBUK NATIONAL UNIVERSITY

척도의 성질

1. 근데 현실에서는 숫자형(numerical) 또는 범주형(categorical)으로 대충 나누는 경우가 많다.
   - 비율(ratio)과 등간(interval)은 숫자형(numerical)이다.
   - 서열(ordinal)과 명목(nominal)은 범주형(categorical)이다.

2. Numerical은 좀 더 나누어 연속형(continuous과)과 이산형(discrete)으로 나뉜다.
   - 곧 배우니까 당장은 신경쓰지 않아도 좋다.

3. Numerical에서 categorical로는 변환할 수 있지만, categorical에서 numerical로는 변환할 수 없다.

충북대학교
CHUNGBUK NATIONAL UNIVERSITY

척도의 성질

1. 연습문제
   - 모 대학의 학칙에 따르면 학점 90점 이상=A; 80–89점=B; 70–79점=C; 60–69점=D; 60점 미만=F으로 letter grade가 부여된다. Letter grade의 척도는 뭐냐?
   - 그것은 숫자형인가? 범주형인가?
   - 그러면 학점의 원점수(100점 만점) 척도는 무엇인가?
   - 그것은 숫자형인가? 범주형인가?
   - 원점수에서 letter grade로 변환(recoding)할 수 있나?
   - Letter grade에서 원점수로 변환(recoding)할 수 있나?

충북대학교
CHUNGBUK NATIONAL UNIVERSITY

척도의 성질

ID	female	IQ	income	socialclass
1	1	135	250	1
2	0	110	310	2
3	1	128	1500	3
4	0	98	122	2
5	1	106	450	2
6	0	102	190	1

1. 빈도분포표(frequency distribution table)

- 범주형(categorical) 자료가 주어져 있으면 즉각 빈도분포표로 요약하여 나타낼 수 있다.

Sex of Respondent	Number of Respondent	Proportion
0	3	0.5
1	3	0.5

충북대학교
CHUNGBUK NATIONAL UNIVERSITY

척도의 성질

ID	female	IQ	income	socialclass
1	1	135	250	1
2	0	110	310	2
3	1	128	1500	3
4	0	98	122	2
5	1	106	450	2
6	0	102	190	1

1. 빈도분포표(frequency distribution table)

- 숫자형(numerical) 자료가 주어져 있다면 먼저 구간(intervals) 단위로 나누고 빈도분포표로 요약하여 나타낼 수 있다.

Income of Respondent	Number of Respondent	Proportion
$x \leq 200$	2	0.33
$200 < x < 300$	1	0.16
$300 < x < 400$	1	0.16
$400 \leq x$	2	0.33

충북대학교
CHUNGBUK NATIONAL UNIVERSITY

설문지에서 리커트 척도의 활용

1. 사회통계학의 데이터는 설문조사(questionnaire survey)를 통해 수집되는 경우가 압도적으로 많다.
   - 최근 동향을 보면 데이터 수집방식이 훨씬 더 다원화되었지만 여전히...

※ 먼저 우리 사회가 얼마나 믿고 살만한지에 대해서 여쭈어 보겠습니다.

RELIAB03

1. 우리 사회가 "매우 믿을 수 없다" 고 할 경우 1점을 주고 "매우 믿을 수 있다" 고 할 경우 10점을 준다면 귀하는 1점과 10점 사이에 몇 점을 주시겠습니까? (해당 점수 앞에 ∨표하여 주십시오)

매우 믿을 수 없다				중간					매우 믿을 수 있다	
___1점	___2점	___3점	___4점	___5점	___6점	___7점	___8점	___9점	___10점	

HELPFUL

2. 귀하는 사람들이 대체로 남을 도우려 한다고 생각하십니까, 아니면 자신만을 위한다고 생각하십니까?

___① 남을 도우려고 한다 ___(3) 경우에 따라 다르다 ☐

___② 자신만을 위한다 ___(8) 모르겠다

FAIR

3. 대부분의 사람들이 기회만 있으면 귀하를 이용하려들 것이라고 생각하십니까, 아니면 공정하게 대할 것이라고 생각하십니까?

___① 이용하려 들 것이다 ___(3) 경우에 따라 다르다 ☐

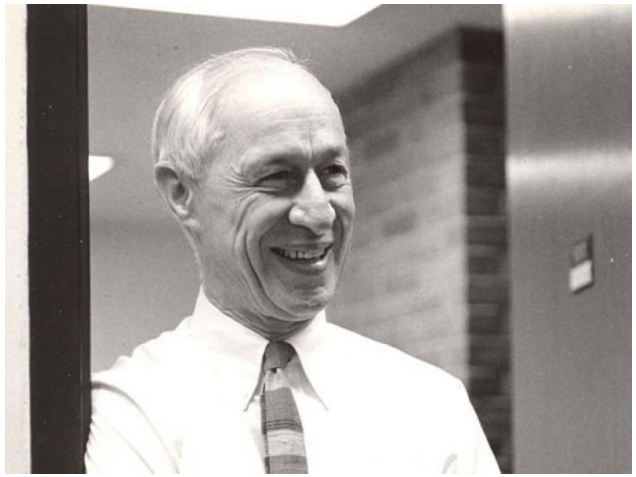
___② 공정하게 대할 것이다 ___(8) 모르겠다

KGSS 설문지 예제

충북대학교 CHUNGBUK NATIONAL UNIVERSITY

설문지에서 리커트 척도의 활용

1. 워낙 익숙하기 때문에 주의를 기울이지 않았겠지만...
   - 아까 설문의 중요한 특징은 바로 (주어진 문항에 대하여) "동의하는 정도를 미리 설정한 카테고리 중에 고르도록 유도했다"는 것이다.
   - 이런 식으로 응답을 유도하는 척도를 특별히 리커트 척도(Likert scale)이라고 부른다.
   - 렌시스 리커트(Rensis Likert)는 미시건대 사회학 학부를 나와 컬럼비아대에서 바로 이 척도에 대한 심리학 박사논문을 썼다.



Rensis Likert (1903-1981)

충북대학교
CHUNGBUK NATIONAL UNIVERSITY

설문지에서 리커트 척도의 활용

1. 엄밀히 말하면 리커트 척도는 서열척도(ordinal scale)다.
   - 아까 말했듯 원칙적으로 서열척도로는 산술연산이 불가능하다.
   - 그러나 현실 속 많은 연구에서는 리커트 척도를 마치 등간척도(interval scale)처럼 사용한다.
   - 범주형(categorical)이더라도 범주의 숫자가 아주 많아지면 사실상 연속형(approximately continuous)이 된다.
   - 스키마1: 40점 미만 $\rightarrow$ 1; 40점-60점 $\rightarrow$ 2; 60점-80점 $\rightarrow$ 3; 80점 이상 $\rightarrow$ 4
   - 스키마2: 10점 미만 $\rightarrow$ 1; 10점-20점 $\rightarrow$ 2; ... 80점-90점 $\rightarrow$ 9; 90점-100점 $\rightarrow$ 10

2. 학기말 무렵이면 리커트 척도를 좀 더 멋진 방식으로 활용할 수 있다.

충북대학교
CHUNGBUK NATIONAL UNIVERSITY

끝!

와~ 숙제다~

충북대학교
CHUNGBUK NATIONAL UNIVERSITY