

사회과학 연구를 위한 기초 지식

노성호, 구정화, 김상원. (2025). 사회과학 연구방법론. 사회과학 연구를 위한 기초 지식 (pp.70-87). 서울: 박영사

I. 개념과 가설

- 사회과학에서 **가정**은 다양한 **개념** 간의 **관계를** 진술하는 것
- **개념** → 사회과학 연구에 있어 가장 **기본적** 재료

- ▶ **구상개념**: 사람들이 특정 현상에 대해 **나름대로 생각하고 이미지화**하는 것
- ▶ **개념화**: 특정 용어가 가진 의미에 대해 **연구자들이 동의해가는 과정**
- ▶ **개념**: 개념화의 결과로 연구자들이 **동의할 수 있게 된 용어**

- **물화(구체화)**: 개념이라는 추상적 대상을 실제 존재하는 것으로 간주하는 것
- **물화가 필요한 이유**

: 어떤 개념이 물화된다 → **측정이 가능하다**

[예시] ‘비행 청소년’ → ‘학교 밖에서의 흡연 및 음주 경험이 있는 중·고등학생’

- **공리**: 이론이 기초를 두고 있는, **사실로 여겨지는 기본적인 주장**
- **명제**: 공리적 토대로부터 유래한 **개념들 사이의 관계에 대한 특별한 결론**

[예시]

공리1) “모든 사람은 물질적 안락을 추구한다.”

공리2) “합법적 수단을 통해 물질적 안락을 누릴 능력은 하층보다 상층의 사람이 크다.”

명제) “물질적 안락을 추구하기 위해 가난한 계층에 속한 청소년들은 부유한 계층의 청소년보다 법을 어길 가능성이 높다.”

- **가설**: 기본적으로 연구자가 가진 연구 질문에 대한 **잠정적인 답**

* 아무리 논리적으로 옳은 명제라도 **현실에서 그 타당성이 검증 혹은 입증되지 않으면** 합당한 이론으로 간주하지 않고, 단지 가설, 즉 그럴듯한 설명으로 취급하는 것이 관례

→ 가설은 개념과 개념의 관계를 설정하여 서술: “~한다면, ~하다.”

[예시] “물질적 안락을 추구하는 하층의 청소년은 상층의 청소년에 비해 비행을 더 많이 할 것이다.” → ‘물질적 안락’, ‘하층’, ‘상층’, ‘청소년’, ‘비행’의 관계 설정 및 개념화

II. 변수

- ‘개념’이 연구대상이 되면 ‘변수’가 되며, 모든 개념은 변수로 다뤄질 수 있음
- 변수를 연구 과정에서 구체적으로 측정하기 위해 연구자는 조작화 단계를 거치면서 개념의 변하는 값을 측정할 수 있는 **양화** 작업을 거치게 된다.

[예시] ‘청소년 비행’ → ‘중·고등학생의 지난 1학기 동안 5만 원 이하의 돈을 갈취한 경험’

- 개념을 조작적으로 정의하고 그것을 측정하게 되면 그 결과는 고정된 값이 아닌 변화되는 다양한 값으로 나타나는데, 이런 상태의 개념을 변수라고 한다.

- 독립변수: 가설에서 원인이 되는 변수
- 종속변수: 원인에 의해 발생하는 결과에 해당하는 변수

[예시] ‘소득 수준에 따라 생활만족도에 차이가 있다.’ → 소득 수준(독립변수) / 생활만족도(종속변수)

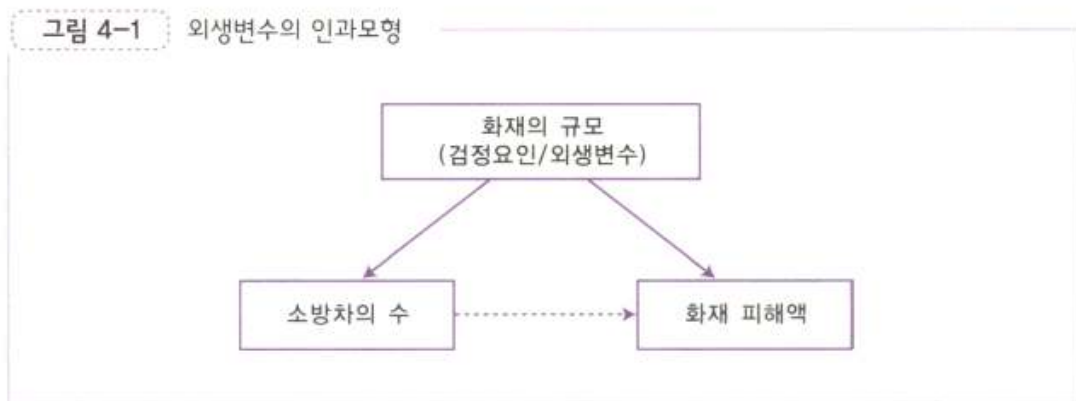
- ① 두 변수 사이에 공분산이 존재한다.
- ② 시간적으로 한 변수(독립변수)가 다른 변수(종속변수)보다 앞선다.
- ③ 두 변수 간의 관계가 제3의 요인에 의해 설명되지 않는다면, 두 변수 사이에는 인과관계가 성립한다.

- 검정요인(검정변수/통제변수): 두 변수에 공통으로 영향을 줄 가능성이 있는 제3의 요인
* 일반적으로 사회현상의 분석은 2개의 변수만으로 이루어지지 않고 다양한 변수 간의 관계를 설명함으로써 이루어진다.

● 검정요인의 종류

- ① 외생변수: 각 변수가 우연히 어떤 다른 변수와 연결됨으로써 관계가 있는 것처럼 보일 때, 영향을 주는 검정요인 * 이때 표면적으로 인과관계처럼 보였던 관계는 ‘허위관계’라고 부른다.

[예시]



- ② 매개변수: 독립변수의 결과인 동시에 종속변수의 원인이 되는 검정요인
[예시] 허쉬와 힌델랑의 연구(1977) “지능지수가 낮은 사람이 범죄를 저지를 가능성이 크다.”
→ 연구 도중 ‘지능지수는 학업 성적에 영향을 주고, 학업 성적이 범죄율에 영향을 준다’는 것을 알게 되었고, 이때 ‘학업 성적’은 매개변수가 된다.



③ **선행변수**: 독립변수에 앞서면서 독립변수에 대해 유의미한 영향을 주는 변수

- 1) 선행변수를 통제할 때, 독립변수와 종속변수의 관계가 사라져서는 안 된다.
- 2) 독립변수를 통제할 때, 선행변수와 종속변수와의 관계는 사라져야 한다.

[예시]

“학업성적이 낮은 사람이 범죄를 저지를 가능성이 크다.”라는 가설에서 독립변수인 ‘학업성적’에 영향을 주면서 종속변수인 ‘범죄율’과 직접적 관계가 없는 변수 → ‘지능지수’가 선행변수가 된다.

● **상관관계**: 두 변수의 변이가 함께 일어난다(=공분산) 관계

- ① **정(+)**적 상관관계: 두 변수가 같은 방향으로 변하는 관계
- ② **부(-)**적 상관관계: 두 변수가 다른 방향으로 변하는 관계

[예시]

“수입이 높을수록 생활만족도가 높아질 것이다.” → 정(+)적 상관관계

“여성의 사회적 지위가 높아질수록 출산율은 낮아질 것이다.” → 부(-)적 상관관계

* 이는 관계를 ‘설정하는’ 것으로 결과와는 다를 수 있다.

● **인과관계**: 두 변수가 직접적인 원인과 결과로 작용하는 관계

● **상관관계와 인과관계의 차이점**

[예시] “아이스크림 판매량이 올라갈수록 익사 사고 빈도도 늘어난다.”

→ 이는 상관관계에는 해당하지만, 인과관계에는 해당하지 않는다.

→ 독립변수 ‘**때문에**’ 종속변수의 변화가 일어나야 **인과관계**다.

→ 연구자가 궁극적으로 찾고자 하는 변수들 사이의 관계는 **인과관계**

Ⅲ. 변수의 측정 기준

● 변수는 하위 구성요소인 속성들의 논리적 집합이다.

[예시]

- 직업(변수)의 속성: 경찰관, 회사원, 농부 등 / 생산직, 서비스직, 전문직 등

- 계층(변수)의 속성: 상층, 중층, 하층

● 이러한 속성들을 구체적인 수치로 재는 것을 **측정**이라고 한다.

● **측정의 분류**

① **명목측정**: 속성이 단지 포괄성과 상호배타적 특성만을 가지는 변수

- 변수의 속성들에 숫자로 점수를 부여할지라도 구별 이외의 기능은 없음

[예시] ‘성별’을 남자와 여자로 구분하는 경우, 남자는 1, 여자는 2라는 값을 부여해도 그것으로 속성을 서열화할 수 없다.

② **서열측정**: 속성들을 논리적으로 서열화할 수 있는 변수

[예시] 사회계층을 ‘상’, ‘중’, ‘하’로 나누고 각각 3, 2, 1의 숫자를 부여한다면, 각 속성의 계층적 지위가 높고 낮음을 양적으로 표현할 수 있다.

- 그러나 서열화된 속성 간 간격이 같은지 여부는 알 수 없다.

③ **등간측정**: 서열화 + 속성 간 간격을 동일한 것으로 보는 변수

- 서열측정 중에서 점수 간의 간격이 동일한 측정

[예시] 나이를 구분할 때 1) 0~9세, 2) 10~19세, 3) 20~29세, ...로 측정 시 각 속성의 간격은 10세 간격으로 동일하다.

④ **비율측정**: 등간측정 중 비율을 적용할 수 있는 변수

- 절대 영점을 기준으로 할 수 있는 경우에 사용

[예시]

- 등간측정: 온도에서 화씨 25도와 26도의 거리는 67도와 68도의 거리와 같다. 그러나 화씨 80도가 40도보다 두 배만큼 더 뜨겁다고 할 수 없다. 화씨온도에서 0도는 임의적인 기준으로 0도가 실제로 열이 없다는 것을 의미하지는 않기 때문이다.

- 비율측정: 그러나 80cm는 40cm보다 두 배 길다고 할 수 있다. 0cm는 절대적인 의미가 있기 때문이다.

- 절대적인 성격이 있는 0을 절대 영점이라고 하며, 비율측정은 이 절대 영점을 기준으로 하기에 비율을 적용할 수 있다.

표 4-1 측정의 4가지 수준

구분	질적인 구별	서열과 순서	동일한 간격	비율
명목측정	○	×	×	×
서열측정	○	○	×	×
등간측정	○	○	○	×
비율측정	○	○	○	○

● **변수 측정 수준과 통계 측정 수준**

① 자료에 사용된 측정 수준보다 높은 수준의 통계 기법을 사용하면 논리적 오류를 범한다.

② 자료에 사용된 측정 수준보다 낮은 수준의 통계 기법을 사용하면 분석 과정에서 얻을 수 있는 정보를 누락시킬 수 있다.

IV. 관련 고려 사항

● **분석단위**: 조사연구에서 연구대상을 지칭

- 분석단위에 따라 선택되는 이론적 근거나 배경이 달라지며, 구체적인 연구방법과 자료수집 방법이 결정됨

① **개인**: 가장 전형적인 분석단위로, 개인에 대한 기술을 종합하고 분석하여 사회집단과의 상호작용을 기술하고 설명하려 함

[예시] '음주 교육 프로그램이 음주운전자의 재범발생률을 감소시키는데 대한 연구'에서 목적은 프로그램의 평가지만, 분석단위는 '개별적 음주운전자'이다.

② **사회집단**: 연구자가 어느 한 집단의 특징에 대해 관심을 가지는 경우

[예시] ‘특정 지역의 범죄율이 타 지역보다 높은 이유에 대한 연구’에서 해당 지역이 가진 특징적 요소들에 대한 연구가 필요하므로, 해당 지역이나 도시 자체가 연구 조사에서 분석단위가 된다.

● **분석단위 관련 오류**

① **생태학적 오류**: 분석단위를 집단으로 정하고 그 특성을 조사했지만, 해석은 개인의 특성으로 해석함으로써 발생하게 되는 오류

② **개인주의적 오류**: 분석단위를 개인으로 정하고 조사했지만, 해석을 집단의 특성으로 해석함으로써 발생하게 되는 오류

[예시] 공교육을 전혀 받지 못하고도 부자가 된 한두 명을 개인적으로 만났다고 해서, 그것을 일반적 경향성으로 해석할 수는 없다.

③ **환원주의적 오류**: 특정 현상을 제한적이거나 하위개념으로 설명하려 시도할 때 발생하는 오류

[예시] KBO에 속한 팀의 우승 예측을 할 때 각 팀의 개별 선수 역량만으로 예측한다면 환원주의적 오류다. 감독의 지도력, 팀워크, 재정, 시설 등도 포함해야 하기 때문이다.

● **모집단과 표본 선택**

① **모집단** → 전수 조사 → 모수

1) **모집단**: 연구대상이 되는 전체 집단

2) **전수 조사**: 모집단 전체를 대상으로 하는 조사

3) **모수**: 전수 조사에서 얻은 특정 변수에 대한 통계적 기술, 오차 없는 진짜 값

② **모집단** → 표집 → 표본 → 표본 조사 → 통계치: 추론통계학

1) **표집**: 모집단으로부터 그 일부를 추출하는 과정

2) **표본**: 표집 과정을 거쳐 얻은 모집단의 일부 요소

3) **표본 조사**: 표본을 통한 연구 조사

4) **통계치**: 표본 조사를 통해 얻은 표본의 변수에 대한 요약 기술, 모수를 추정하는 값

→ 중요한 것은 ‘통계치를 통한 모수의 추정이 얼마나 정확한가’

→ ‘모집단에서 표집 과정을 거쳐 추출한 표본이 얼마나 모집단과 닮아있는가’ → ‘대표성’

● **대표성**: 표본의 특성이 그 표본이 추출된 모집단의 특성과 동일한 특성을 가지는 것