

사회조사방법론

연구 모형과 가설의 제시

김현우, PhD¹

¹충북대학교 사회학과 조교수

March 28, 2025



진행 순서

- 1 연구 모형의 가치
- 2 가설의 제시
- 3 연구 가설의 구체화
- 4 선행연구 고찰

연구 모형의 가치

연구 모형의 가치

연구 질문이 구체화됨에 따라 연구 모형을 구성할 수도 있다.

- 연구 주제와 연구 질문이 무엇(what)에 관한 문제라면 연구 모형은 그 무엇들의 관계에 관한 것이다.
- 연구 모형을 나타내는 가장 기초적인 방식 중 하나는 **인과지도(causal diagram)**를 그리는 것이다.
- 인과지도는 개념들(보다 구체적으로는 변수들) 사이의 관계를 간단히 요약하는 그림이다.
- 이것은 필수적인 것은 아니다. 그러나 생각의 정리에 상당히 유용하다!
- 아주 단순한 인과지도에서 시작하여 조금씩 확장하면서 연구주제를 발전시켜 나갈 수 있다.

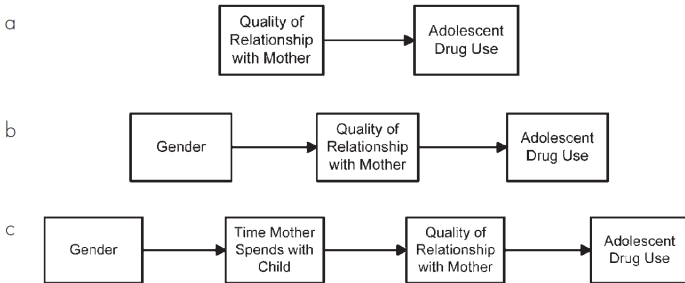


연구 모형의 가치

- 가령 어머니와의 관계가 청소년의 약물남용과 어떻게 연결되어 있는지를 사례로 살펴보자.
- 처음에는 어머니와의 관계의 질(quality of relationship with mother)이라는 변수가 청소년의 약물남용(adolsecent drug use)이라는 변수와 어떻게 연결되어 있는가를 그린다.
- 좀 더 나아가 혹시 청소년의 성별이 어머니와의 관계의 질에 영향을 미치고, 궁극적으로 약물남용으로 이어지지 않는가를 고민한다.
- 만약 아들은 딸보다 어머니와 함께 보내는 시간이 적고, 이것이 어머니와의 관계의 질에 악영향을 미치고, 다시 청소년의 약물남용으로 이어지지 않는지 의심한다.

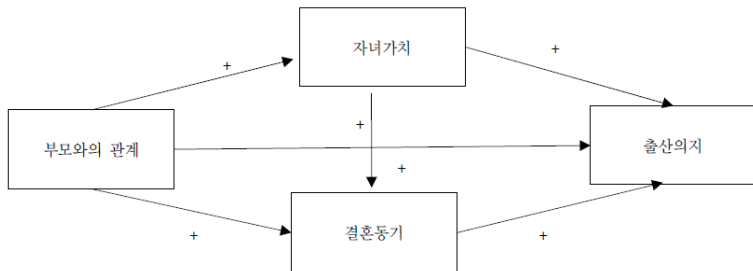


연구 모형의 가치



연구 모형의 가치

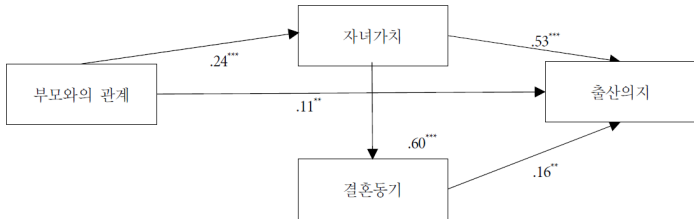
- 김은진 외(2002)의 연구처럼 부모와의 관계, 자녀가치, 결혼동기, 출산의지 등 여러 개념들의 복합적인 관계를 인과지도로 나타낼 수 있다.



김은진, 정혜정, 정유진. 2022. “대학생이 지각한 자녀가치와 결혼동기 및 부모와의 관계가 출산의지에 미치는 영향: 성별에 따른 다집단 경로분석.” 『가정과삶의질연구』 40(1): 27-39.

연구 모형의 가치

- 연구가 끝나고 나서 각각의 화살표 위에 **표준화 회귀계수(standardized regression coefficients)**를 써넣는 방식으로 연구 결과 요약에 활용할 수 있다.



** $p < .01$, *** $p < .001$

모든 경로계수는 표준화계수임



연구 모형의 가치

연구 모형은 여러 연구 질문들을 반영한다.

- 연구 모형에서 하나의 화살표는 하나의 연구 질문에 대응한다. 단 강박적으로 화살표를 연결할 필요는 없다!
- 두 변수간 연관성에만 매몰되지 않고, “내가 어떤 큰 그림을 그리고 있지?”와 “내가 이 변수간 관계를 왜 보고 있지?”를 끊임없이 반추해야 한다.
- 연구 모형은 (개별적 질문들을 넘어) 여러 연구 질문들의 종합적 관계를 총체적으로 살펴볼 수 있도록 도와준다.
- 이를 곰곰이 살펴보다보면 새로운 연구 질문이 나오기도 한다! 이것이 연구 모형의 **발견적(heuristic) 유용성**인 셈이다.



가설의 제시

가설의 제시

연구 가설은 연구 질문보다도 구체적이다.

- 연구 가설(research hypothesis)은 연구 질문에 대한 임시적인(tentative) 대답이다.
- 그러나 가설은 변수 간 연관성을 직접적으로 다룬다는 점에서 매우 구체적으로 서술되어야 한다.
- “개념 X 와 개념 Y 간에는 이러저러한 관계가 있지 않을까?” 수준이 연구 질문이라면, 연구 가설에서는 “변수 x 가 증가하면 변수 y 는 감소할 것이다”와 같은 구체성을 갖추어야 한다.



가설의 제시

가설이 유도되기까지 충분한 이론적 근거가 있어야 한다.

- 가설을 아무렇게나 추측해서 근거없이 내던지면 안된다! 종종 그렇게 가설을 제시하는 사람들이 있다.
- 첫 작업 가설을 그렇게 잡았더라도, 기존 문헌을 철저하게 살펴보고 제대로 이론적 근거를 찾아내야 한다!
- 문제는 종종 구체적인 가설의 **방향(direction)**을 예상할 만큼 (기존의) 이론이 풍부하지 못한 경우도 많다는 점이다.
- 그런 경우 그저 “차이가 있을 것이다” 하고 예상하는 정도에 그치는 경우가 있다(가령 “도시 지역 1인가구 거주자의 주관적 건강상태는 농촌 지역 1인가구 거주자의 그것과 차이가 있을 것이다”).



가설의 제시

- 해당 연구문제에 관해 풍부한 기존 문헌이 주어져 있고 우리가 이 논문에서 개발한 이론도 철저하다면 방향을 구체적으로 예측해 볼 수 있다.
- 그렇다면 가설 또한 구체적이고 능동적으로 제시하는 것이 바람직하다(가령 “도시 지역 1인가구 거주자의 주관적 건강 상태가 농촌 지역 1인가구 거주자의 그것보다 우수할 것이다”).
- 이것은 양측검정(two-tailed test)과 단측검정(one-tailed test)의 차이를 반영하기도 하지만, 동시에 가설을 유도하기 위한 이론적 자원이 얼마나 풍부한가와도 이어져 있다(Why?).



가설의 제시

좋은 연구 질문과 가설에는 나름의 규칙이 있다.

- 연구 질문과 가설은 가능한 간단명료하게 진술한다. 필요하다면 연구 보고서의 제목에 이를 답아도 좋다!
- 연구 질문과 가설은 학술적인 용어와 표현을 사용하여 진술해야 한다. 함부로 은유나 과장과 같은 문학적 수사를 사용해선 안된다.
- 연구 질문과 가설은 충분히 선행 연구를 살펴보고 이를 근거로 하여 제시되어야 한다.
- 연구 보고서(특히 서론 말미)에 명시적으로 연구 질문과 가설을 꼭 써야하는 것은 아니다. 그러나 최종 보고서에서는 연습삼아 명시적으로 쓰도록 하자.
- 참고로 **귀무가설(null hypothesis)**와 연구 가설을 혼동하지 말자. 귀무가설을 쓸 필요는 없다.



연구 가설의 구체화

연구 가설의 구체화

이주 노동자에 대한 원주민의 차별적 태도를 사례로 생각해보자.

- 우리나라의 원주민의 이주 노동자에 대한 **외집단 편견(out-group prejudice)**을 연구하기로 결심하였다고 하자.
- 기존 문헌을 통해 **민족 경쟁(ethnic competition)** 그리고 **지위 위협(status threat)** 개념으로 외집단 편견을 설명할 수 있다고 배우게 되었다.
- 이제 연구 주제는 “원주민의 지위 위협 인식과 이주 노동자에 대한 외집단 편견 간 연관성”으로 설정된다.



연구 가설의 구체화

- 탐구를 계속하면서 “이주 노동자의 수가 증가하면 원주민들이 점점 차별적인 태도를 발전시키면서 높은 수준의 외집단 편견을 가지게 되지 않을까?” 라는 연구 질문을 세웠다.
- 이제 “거주 지역의 이주 노동자 수”와 “이주민에 대한 외집단 편견” 두 변수 사이의 관계를 살펴볼 필요가 있다.
- 연구 가설이 무엇인지 생각해보자!



연구 가설의 구체화

- 회귀분석(regression analysis)을 통해 두 변수 사이의 연관성을 검정해 볼 수 있다.

$$y = b_0 + b_1x_1 + e$$

- 이때 독립변수 x_1 는 거주 지역의 이주 노동자 수이고 y 는 외집단 편견이 된다.
- 물론 실제 연구에서는 여러 통제 변수(control variables)가 포함되어야 한다(Why?).
- 만일 $b_1 = 0$ 이라는 통계적으로 유의하게 귀무가설을 기각한다면 이 연구가설을 받아들일 수 있게 된다.
- 이렇게 끝맺을수도 있지만 그러면 다소 뻔한 질문에 뻔한 대답에 그치고 만다(Why?).



연구 가설의 구체화

여기서 멈추지 않고 계속해서 질문을 던져보자!

- “객관적인 이주 노동자 수를 누가 어떻게 알 수 있겠어? 그보다는 인지된 이주 노동자 수가 더 중요하지 않을까?”
- 센서스 자료에서 수집된 (객관적인) 이주 노동자 수 대신 설문조사 대상자에게 “당신이 살고 있는 지역의 이주 노동자 수는 몇 명이라고 생각하십니까?”에 대한 답을 사용할 수도 있다.
- 이제 인지된 이주 노동자 수(x_2)와 외집단 편견(y) 간 연관성에 초점을 두고 분석한다.

$$y = b_0 + b_1x_1 + x_2 + e$$



연구 가설의 구체화

- 여기서 멈추지 않고 계속해서 질문을 던져보자!
- “그보다는 좀 더 미묘한 관계가 있지 않을까? 혹시 과장스럽게 인지한 이주 노동자 수가 더 중요한 것 아닐까?”
- (여러가지 방법이 있을수 있겠지만) 한가지 접근법은 인지된 이주 노동자 수(x_2)와 실제 이주 노동자 수(x_1)의 차이를 독립변수로 삼아 같은 분석을 반복해 보는 것이다.

$$y = b_0 + b_1(x_2 - x_1) + e$$

- 만일 이 회귀계수가 통계적으로 유의하다면 무엇을 의미할까?



연구 가설의 구체화

- “(객관적인) 이주 노동자 수가 외집단 편견과 꼭 **선형적 관계(linear relationship)**를 가지고 있을까?”
- 이주 노동자 수가 전반적으로 적을 때에는 이주 노동자 수가 증가할 때, 조금씩 원주민과의 상호작용이 증가하면서 외집단 편견이 오히려 깨질 수 있다.
- 그러므로 “일정 수준까지는” 두 변수 사이에 음(-)의 연관성이 나타날 수 있다 (편견개선 효과).
- 그러나 이주 노동자 수가 너무 많아진다면 실질적인 위협으로 느껴져 양(+)의 연관성이 나타날 수 있다(편견악화 효과).



연구 가설의 구체화

- 다음과 같이 회귀분석에서 이차항(quadratic term)을 포함하면 이러한 비선형적 관계(non-linear relationship)를 간단히 검정할 수 있다.

$$y = b_0 + b_1x_1 + b_2x_1^2 + e$$

- 만약 추정 결과 $b_2 > 0$ 이고, $b_2 = 0$ 이라는 귀무가설을 통계적으로 유의하게 기각할 수 있다면 이 연구 가설을 수용할 수 있다(Why?).
- 만약 $b_2 < 0$ 이고, $b_2 = 0$ 이라는 귀무가설을 통계적으로 유의하게 기각할 수 있다면 무엇을 뜻할까?



연구 가설의 구체화

여전히 식상하다고 느껴진다면 좀 더 고민해보자.

- “이주 노동자들로 쉽게 대체될 수 있는 직종의 사람들이 더 강한 외집단 편견을 지니는 것은 아닐까?”
- 물론 “쉽게 대체될 수 있는 직종”이라는 개념에 대한 조작적 정의(operational definition)가 필요하다.

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_1x_2 + e$$

- 이때 x_1 는 (실제 또는 인지된) 이주 노동자 수, x_2 는 해당 응답자가 대체 가능한 직종인지 여부를 묻는 **가변수(dummy variable)**이다. 그리고 x_1x_2 는 두 변수를 곱한 변수가 된다(Why?).
- 만약 추정 결과 $b_3 > 0$ 이고, $b_{12} = 0$ 이라는 귀무가설을 통계적으로 유의하게 기각할 수 있다면 이 연구 가설을 지지하는 것이다(Why?).



연구 가설의 구체화

질문과 답의 연쇄를 연구자의 내면 속에서 끊임없이 이어나간다.

- 처음에는 작고 식상한 주제로 시작하였더라도 점차 그 내용을 깊게 파고 들어가다보면 풍성한 분석이 가능하다.
- 적어도 주목할 만한 가설이 두세 개 정도는 나올 수 있도록 하자. 너무 적거나 너무 많아도 곤란하다.
- 이 과정에서 특히 **악마의 주창(Devil's Advocate)**을 활용한 토론이 필요하다(Why?).
- 수많은 가설을 세우고 테스트해 보아야 한다. 이 과정이 상당히 소모적이라고 느껴질 수도 있지만, 꼭 필요한 과정이다. 그 중 정말로 살아남은 가설들에 주목해보자.



선행연구 고찰

이민에 대한 개인적 태도는 어떻게 형성될까?

- 경제학적 관점에서 보면, 자본이나 상품의 이동 만큼이나 사람의 이동(immigration)은 모두에게 좋은 것이다(Why?).
- Enrico Moretti에 따르면 이동하는 사람은 더 나은 기회를 누릴 수 있고, 이민을 받는 곳의 사람들 역시 저렴하고 창의적인 노동력을 확보하게 된다.
- 그러나 현실에서 이민자는 좀처럼 환영받지 못할 뿐더러, 격렬한 정치적·사회적 갈등의 원천이 되기도 한다.
- 왜 모두에게 이익이 되는 일을 어떤 사람은 거부할까? 이것은 사회과학적으로 설명되어야 한다!
- O'Rourke and Sinnott (2006)은 이민에 대한 태도를 경제학적 관점에 따라 설명하고 있다(이것은 합리적 선택 설명이다).

O'Rourke, Kevin H. and Richard Sinnott. 2006. "The Determinants of Individual Attitudes towards Immigration." *European Journal of Political Economy* 22(4): 838-861.

- “The first asked if the number of immigrants to their economy should be increased a lot (1), a little (2), remain the same (3), be reduced a little (4) or reduced a lot (5). The second asked if refugees should be allowed to stay in the country; responses ran from agree strongly (1) to disagree strongly (5).”

Table 1
Average sentiment regarding immigrants and refugees

Country	Anti-immigrant		Anti-refugee		
	Mean	S.D.	Mean	Rank	S.D.
Hungary	4.402	0.817	2.838	8	1.077
E. Germany	4.338	0.871	1.961	24	0.879
W. Germany	4.226	0.910	2.049	23	1.022
Bulgaria	4.219	0.990	2.661	13	1.379
Latvia	4.182	0.884	3.757	1	1.312
Czech Rep.	4.158	0.880	2.463	15	1.143
Italy	4.151	0.900	2.846	7	1.269
Britain	4.052	0.962	2.820	9	1.100
Slovakia	4.004	0.911	3.021	4	1.258
Sweden	3.961	1.017	2.275	20	1.074
Slovenia	3.939	0.868	3.565	3	1.103
Poland	3.888	1.060	2.535	14	1.144
USA	3.873	1.044	2.748	11	1.098
Norway	3.847	0.982	2.340	19	0.990
Netherlands	3.826	0.924	2.366	18	1.044
Austria	3.804	0.933	2.095	22	1.111
Philippines	3.796	1.102	3.708	2	1.000
Australia	3.768	1.042	2.954	6	1.202
New Zealand	3.742	1.053	2.807	10	1.075
Russia	3.717	0.971	2.698	12	1.242
Spain	3.401	0.813	2.460	16	1.036
Japan	3.391	1.008	3.014	5	1.296
Canada	3.317	1.135	2.404	17	1.129
Ireland	3.071	0.829	2.163	21	0.911

Source: Data from ISSP National Identity Survey 1995.



선행연구 고찰

- 연구자들은 높은 직무 능력을 갖고 쉽사리 대체되지 않는 직종에 속한다면 이민에 대해 관용적일 것으로 가설을 세웠다.
- 여러 나라에서 공통적으로 사용하는 국제노동기구(ILO) 국제 표준 직업 분류 체계 (ISCO88)에서 (1) 단순 작업 종사자, (2) 설비·기계의 운전·조립공, (3) 기능공·관련직 종사자, (4) 전문직, (5) 관리직 중에 3, 4, 5번에 속하는 경우를 나타내는 변수 (Skill345)를 만들었다.
- “Skill345 직종에 있는 이들은 반이민 태도가 낮을 것이다.”
- “부유한 나라일수록 Skill345의 반이민 태도는 덜할 것이다(Why?).”
- “빈부격차가 심할수록 Skill345의 반이민 태도는 심할 것이다.”
- “이 모든 효과는 세계 여러 나라에 걸쳐 일관될 것이며, 심지어 수많은 다른 변수를 집어넣더라도 여전히 나타날 것이다!”



Determinants of anti-immigrant preferences (ordered probit) (dependent variable: anti-immigrant)

	(1)	(2)	(3)	(4)
<i>Panel A: trade theory</i>				
Skill345	-0.3189* [0.0881]	-0.3179* [0.0895]	-0.2772* [0.0939]	-0.3025* [0.0941]
Skill345 * GDP per capita	-0.0105* [0.0023]	-0.0109* [0.0024]	-0.0065* [0.0025]	-0.0058** [0.0025]
Skill345 * inequality	0.0070* [0.0026]	0.0076* [0.0026]	0.0070** [0.0028]	0.0080* [0.0028]
<i>Panel B: other economic considerations</i>				
Age		0.0087* [0.0024]	0.0073* [0.0025]	0.0062** [0.0025]
Age ²		-0.0000***	-0.0001** [0.0000]	-0.0000*** [0.0000]
Unemployed		0.0371 [0.0336]	0.0382 [0.0346]	0.0299 [0.0346]
National mobility		-0.0639* [0.0159]	-0.0127 [0.0171]	-0.0113 [0.0171]
<i>Panel C: cultural variables</i>				
Patriotism			0.0805* [0.0114]	0.0623* [0.0115]
Chauvinism			0.3192* [0.0112]	0.2822* [0.0114]
International mobility			-0.0803* [0.0214]	-0.0697* [0.0214]
Never lived abroad			0.1224* [0.0213]	0.1105* [0.0213]
Native			0.1802* [0.0484]	0.1823* [0.0484]
Native parents			0.1964* [0.0430]	0.1964* [0.0430]
Catholic			-0.0239 [0.0206]	-0.0282 [0.0206]
Female			0.0354** [0.0151]	0.0121 [0.0152]
Protectionism				0.1231* [0.0075]
No. of observations	23,714	22,697	21,347	21,317
Log likelihood	-30,208.22	-28,831.3	-26,055.4	-25,874.95
Pseudo-R ²	0.04	0.04	0.08	0.08

Robust standard errors in brackets. All regressions include country dummies; coefficients not reported.

