

사회조사방법론

실험설계

김현우, PhD¹

¹충북대학교 사회학과 조교수

April 4, 2025



진행 순서

- 1 실험설계의 기초 개념
- 2 실험설계의 타당성
- 3 사회학에서 실험하기

실험설계의 기초 개념

실험설계의 기초 개념

실험은 보통 자연과학의 방법으로 취급된다.

- 그러나 사회과학에서도 얼마든지 가능하며 대단히 흥미진진한 실험 연구의 결과가 있다.
- 무엇보다 실험설계(design of experiment: DoE)는 과학적 방법론 전반과 조사방법론의 주요개념과 이론을 학습하는데 있어 매우 중요한 토픽이 된다.
- 사회조사분석사 자격증에서도 실험설계 관련 내용이 살짝 언급된다. 그다지 수준이 높지는 않다(Why?).
- 나는 사회과학자들에게도 실험설계에 관한 깊이 있는 지식이 유용하다고 본다. 오늘날 실험이 생각보다 매우 자주 쓰이기 때문이다!



실험설계의 기초 개념

고전적 실험설계의 기초적인 개념들을 먼저 살펴보자.

- 독립변수(independent variable)와 종속변수(dependent variable)
- 사전조사(pre-test)와 사후조사(post-test)
- 실험집단(experimental group)과 통제집단(control group)



실험설계의 기초 개념

독립변수가 종속변수에 영향을 미친다.

- “어떤 사람들에게 인종차별을 고발하는 다큐멘터리 영화를 보여주었고, 그 결과 이 사람들의 인종편견(racial prejudice)이 감소하였다.”
- 이때 다큐멘터리 영화 관람은 독립변수가 된다.
- 이것은 일종의 **처리(treatment)**이자, **개입(intervention)**이자, 특정 자극에 대한 **노출(exposure)**이 된다.
- 이때 인종편견은 종속변수가 된다.
- 이것은 처리로 인해 나타난 **결과(outcome)**이다.



실험설계의 기초 개념

사전조사와 사후조사는 조사의 타이밍에 관한 문제이다.

- 어떤 사람은 인종 편견을 “처음부터” 덜 가지고 있을 수 있다(반대일수도 있다). 따라서 처리 전후로 “얼마나 변화했는가”를 알려면 처음에 어떤 수준으로 인종 편견을 가지고 있었는지 미리 알아야 한다.
- 그러므로 “실험 전에” 종속변수를 미리 측정하고(사전조사), “실험 후에” 종속변수를 한 번 더 측정해야 한다(사후조사).
- 이 맥락에서 본다면, 사후조사에서 측정된 인종 편견 수준을 사전조사에서 측정된 인종 편견 수준로 뺀 값이 다큐멘터리의 효과 크기라고도 볼 수 있다(Why?).

$$TE = E(Y_{post} - Y_{pre})$$



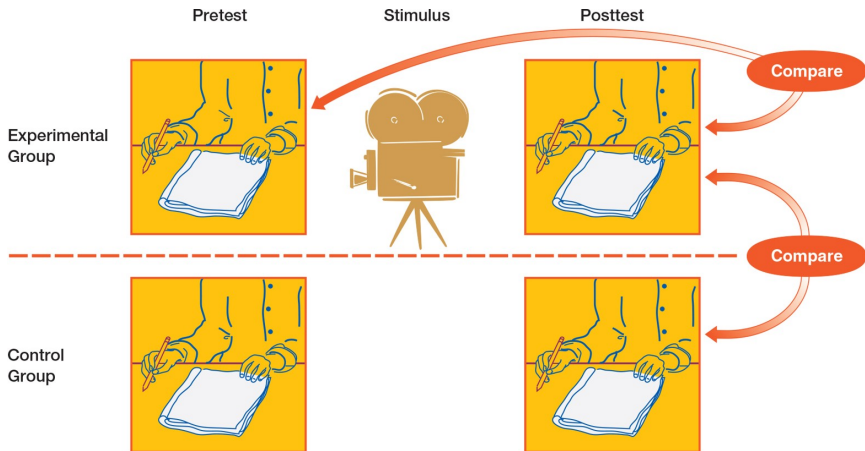
실험설계의 기초 개념

마지막으로 실험집단과 통제집단은 실험설계의 핵심이다!

- 오로지 실험집단에 대해서만 처리를 가한다. 통제집단은 그러한 처리를 하지 않는다.
- 때때로 하나 이상의 실험집단이 있을 수 있다. 이 경우 처리 내용을 달리한다(e.g., 다큐멘터리, 토의, 다큐멘터리+토의).
- 실험집단 내(within)에서, 사후점수와 사전점수를 비교하여 **처리효과(treatment effect)**의 크기를 알 수 있다.
- 실험집단과 통제집단 간(between)에서 각 점수를 비교하여 처리효과의 크기를 알 수 있다.



실험설계의 기초 개념



실험설계의 기초 개념

이렇게 보면 실험설계 방식은 크게 세 종류로 나눌 수 있다.

- 첫째, **피험자내 설계(within-subjects design)**는 각각의 실험 참가자(subjects)가 여러 개의 처리에 노출되는 설계방식이다. 이 경우 처리효과는 피험자 “내” 반응의 차이에 따른다.
- 둘째, **피험자간 설계(between-subjects design)**는 각각의 실험 참가자가 하나의 처리에만 노출되는 설계방식이다. 이 경우 처리효과는 피험자 “간” 반응의 차이에 따른다.
- 셋째, **요인설계(factorial design)**는 실험 참가자가 둘 이상의 처리에 대하여 그만큼 집단을 나누어 “가능한 모든 조합에서” 실험이 이루어지도록 하는 설계방식이다.



실험설계의 타당성

실험설계의 타당성

불완전한 실험설계를 살펴보면 바람직한 실험의 조건을 알 수 있다.

- Donald Campbell과 Julian Stanley가 1963년에 출판한 <Experimental and Quasi-Experimental Designs for Research>은 이 분야의 아주 유명한 고전이다.
- 세 종류의 원시실험 연구설계(pre-experimental research designs)를 살펴보자.
- 가장 먼저 단일집단 사례조사(one-group design)에서는 사전조사를 하지 않고 통제집단도 없다.
- 실무적으로 보면 “모든” 구성원에 대해서 프로그램을 집행하고 그 결과를 기록한 다음, 정책평가를 위해 연구자를 찾는 경우가 생각보다 흔하다(Why?).
- 이 설계에는 사전조사도 통제집단도 없으므로 안타깝지만 잘못된 수집된 자료이다. 실험설계 이전에 연구자와 미리 상의를 해야 한다.



실험설계의 타당성

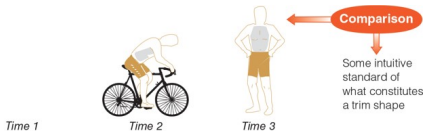
- 다음으로 **단일집단 사전사후 설계(one-group pretest-posttest design)**에는 사전조사와 사후조사를 하지만 통제집단이 없다.
- 통제집단이 없으므로 처리효과를 가져오는 그 원인이 처리로 인한 것이라는 확증을 얻을 수 없다(Why?).
- 마지막은 **정적집단 비교(static-group comparison)**에는 사전조사와 사후조사는 없지만 통제집단은 있다.
- 이것은 충분히 괜찮은 것처럼 들리지만 사전조사가 없었기 때문에 **초기값(initial value)**을 알 수 없다는 한계를 갖는다.



실험설계의 타당성

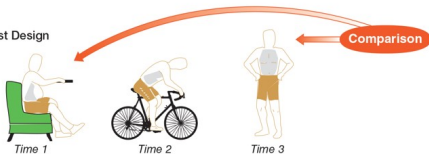
One-Shot Case Study

A man who exercises is observed to be in trim shape



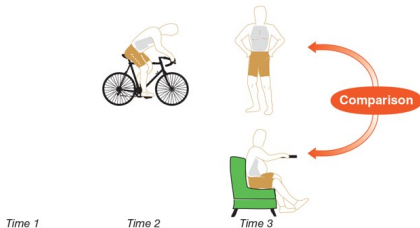
One-Group Pretest–Posttest Design

An overweight man who exercises is later observed to be in trim shape



Static-Group Comparison

A man who exercises is observed to be in trim shape while one who doesn't is observed to be overweight



실험설계의 타당성

실험설계는 타당해야 한다. 이 타당하는 말은 특별한 의미를 갖는다.

- 타당성(validity)은 대단히 복잡한 개념이지만, 가장 단순하게 내적(internal) 타당성과 외적(external) 타당성으로 구분할 수 있다.
- 내적 타당성은 인과효과(causal effect)에 관한 것이다. 즉 실험에서 가해진 처리(treatment)의 결과(outcome)로서 효과가 나타나야 “내적으로” 타당한 실험설계라고 할 수 있다.
- 외적 타당성은 일반화(generalization)에 관한 것이다. 즉 실험에서 얻은 발견(finding)이 실험 참가자를 넘어 모집단 전체로 확장될 수 있어야 “외적으로” 타당한 실험설계라고 할 수 있다.



실험설계의 타당성

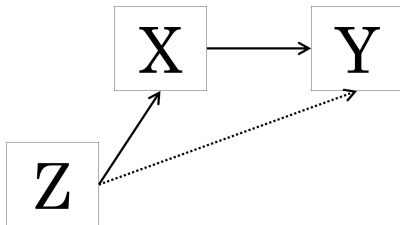
내적 타당성을 먼저 다루고 그것이 위반되는 상황을 살펴보자.

- 실험설계에서는 실험집단과 통제집단을 구분한다고 하였다.
- 먼저 Ball-Rokeach, Grube, and Rokeach (1981)의 드라마 <뿌리(Roots)> 시청효과에 관한 연구를 읽어보자(Babbie 2021: 378-379).
- 이 사례가 설명하듯 자기선택으로 인한 왜곡(self-selection bias)은 실험설계의 내적 타당성을 치명적으로 훼손한다.
- 실험에 참여하기로 “선택하는 것”은 실험 참가자의 몫이다. 다시 말해, “드라마를 보기로 한 것”도 그 실험 참가자의 한 몫이다.
- 만약 이 선발단계에서 이미 체계적인 차이가 나타났다면(즉 드라마 시청자와 비시청자 사이에 차이가 있다면) 실험에 의한 처리효과(treatment effect)와 자기선택효과는 뒤섞이고 만다(Why?).



실험설계의 타당성

- 문제는 일상에서 볼 수 있는 사회적 규칙성에 대한 관찰은 대부분 자기선택으로 인한 왜곡에 의해 오염된다는 사실이다(Why?).
- 최종학력이 소득에 미치는 영향을 본다고 하자. 최종학력은 자기선택의 결과이다. 그러므로 최종학력의 효과는 자기선택의 결과이기도 하고 학교 교육의 결과이기도 하다.
- 그렇기 때문에 우리는 최종학력과 고소득 사이에는 **상관관계(correlation)**만 있다고 말할 수 있으며, **인과관계(causality)**가 있다고는 할 수 없다.



실험설계의 타당성

- 실험설계에서는 일단 실험 참가자가 들어오면 일단 실험집단과 통제집단으로 나눈다. 그리고 이 둘을 나눌 때는 무작위화(randomization)를 사용한다.
- 동전던지기를 해서 1/2의 확률로 실험집단에, 1/2의 확률로 통제집단에 돌린다. 그러면 $Z \rightarrow X$ 에 해당하는 자기선택으로 인한 왜곡은 마술처럼 사라지고 만다!
- 아까 예로 되돌아가, 진짜 최종학력과 고소득 사이의 인과관계를 보려면 강제로 동전 던지기를 시켜 자기선택으로 인한 왜곡을 0으로 만들어야 한다!



실험설계의 타당성

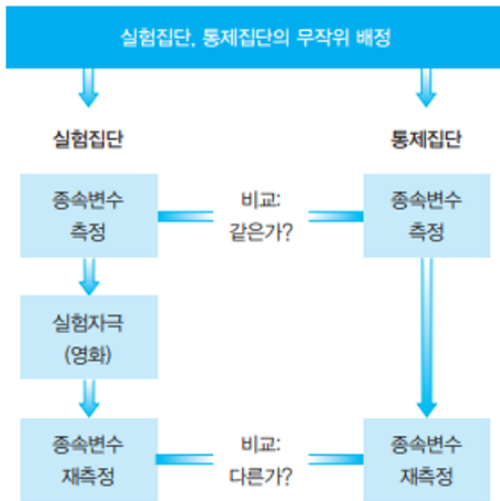
무작위화는 완벽한 실험설계의 꽃이다.

- 무작위화를 통해 마술처럼 실험설계에 내적 타당성을 확보할 수 있다.
- 만일 실험집단과 통제집단을 무작위(random)로 나누었다면 심지어 사전조사조차 필요하지 않다. 무작위화를 통해 (사전조사 없이도) 두 집단이 이미 사전에 동등하다는 것을 알 수 있기 때문이다(Why?).
- 무작위하게 두 집단을 나누었다는 전제 하에서, 실험집단과 통제집단 간 점수 차이는 곧 처리효과의 크기를 반영한다.

$$TE = E(Y_{treated} - Y_{control})$$



실험설계의 타당성



실험설계의 타당성

- 때때로 연구 주제상 무작위화가 불가능한 경우가 있다. 가령 민권운동(civil rights movement)이 사회복지액 증가에 미치는 효과를 살펴본다고 하자.
- 이런 경우 대안으로 **짝맞추기(matching)**를 사용할 수 있다. 즉 “처리를 받았다는 사실을 제외하고” 나머지 요소가 최대한 비슷한 대상을 어떻게든 찾아내 실험집단과 짝짓는다.
- 나머지 조건은 비슷한데 오로지 처리 여부만이 달랐으므로 종속변수의 변화(사회복지액 증가)는 처리(민권운동) 탓으로 귀속시킬 수 있다(Why?).



실험설계의 타당성

- 하지만 모든 조건이 완벽히 일치하는 짝을 찾아내기란 매우 어렵다.
- 가령 흑인 비율, 주 의회(state legislature) 민주당원 비율, 경찰 인력 등 다양한 조건을 몇 개 정도 일치하는 경우야 찾을 수 있을지도 모른다.
- 아무리 그래도 내가 모르는 조건도 있어 그것이 맞지 않을 수도 있다(Why?).
- 무작위만은 못하지만, 짝맞추기는 차선으로서 나름 의미가 있긴 있다.



실험설계의 타당성

보통 실험설계는 외적 타당성 측면에서 근본적인 한계가 있다.

- 연구자로서 교수는 수업을 듣는 대학생에게 실험 참여를 요청한다. 실제로 과거의 수많은 심리학적 발견은 미국, 백인, 남자 대학생 실험 대상으로 삼아 얻어진 것이다.
- 피험자 풀(subject pool)이 그토록 제한적인데 이들로부터 얻은 발견을 모든 사람에게 대해 일반화(generalization) 할 수 있을까?
- 설령 대학 밖을 벗어나 지하철 유리창에 실험 참여자 모집 광고를 싣는다고 해도 이 문제는 극복되지 않는다.
- (발견이 일반화될 대상인) 모든 사람이 “동등하게 실험에 참여하게 될 확률”을 갖는 것은 아니다.
- 다시 말해, 실험 참가자는 모집단을 대표(representative)하지 못한다.



사회학에서 실험하기

사회학에서 실험하기

솔직히 말해 사회학 분야에서 실험연구는 사실 주류는 아니다.

- 실험실적 실험연구(laboratory experiment studies)는 심리학이 주도해 왔다. 그 뒤를 한참 멀리서 따르고 있는 건 교육학과 경제학 정도이다.
- 그러나 사회학에서도 실험실 바깥에서 실험에 가까운 연구가 종종 있어왔다. Correll, Benard, and Paik (2007)의 연구에 대해 읽어보자(Babbie 2021: 374-375).
- 여러분도 20-30명 정도 실험 참가자를 모집하여 간단한 실험을 수행하고 그 의미를 해석한다면 그 자체로 훌륭한 사회학적 연구를 수행한 것이다.
- 아쉽게도 연구 윤리 문제가 까다롭기 때문에, 이 수업에서는 추천하지 않는다. 꼭 한다면 빠른 시일 내에 상담이 필요하다.

Correll, Shelley J., Stephen Benard and In Paik. 2007. "Getting a Job: Is There a Motherhood Penalty?" American Journal of Sociology 112(5): 1297-1338.



사회학에서 실험하기

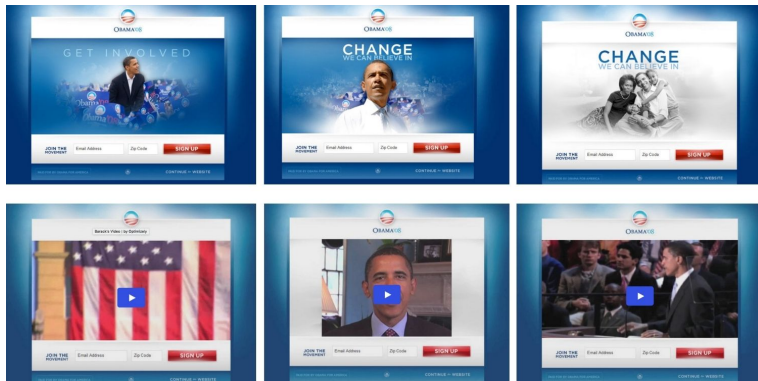
실험설계는 웹기반 자료수집 및 분석과 관련하여 최근 매우 주목받고 있다.

- 이제 웹이나 어플리케이션을 매순간 이용하지 않고 살아가기란 매우 어렵다. 그런데 웹과 어플은 그 자체로 훌륭한 실험실이 된다.
- 사람이 웹이나 어플을 사용하는 순간 실험집단과 통제집단으로 임의화(randomization)하여 나누면 즉각 실험을 수행할 수 있기 때문이다.
- 잡마켓에서 이른바 A/B Test가 데이터 분석가의 직무수행 필요능력으로 종종 거론된다.
- 이 분야는 이른바 인간-컴퓨터 상호작용(human-computer interaction; HCI)에서 활발히 연구되고 있다.



사회학에서 실험하기

- Barack Obama는 2010년 미 대선 당시 웹 실험을 통해 선거운동 자금 6천만 달러를 추가로 모을 수 있었다.



사회학에서 실험하기

자연실험 개념도 사회학자에게 매우 중요한 연구 방법이다.

- 자연실험(natural experiment)은 실험실 밖에서 (즉 인간의 의도나 통제 바깥의 영역에서) 우연히 발생한 사건을 토대로 실험처럼 그 함의를 해석할 수 있게 되는 상황이다.
- 우리나라의 독특한 맥락에서는 동성학교(same-sex school)의 학업효과 측정을 자연실험할 수 있었다(Why?).
- 때로는 자연재해가 예측치 못하게 자연실험의 계기를 가져다주기도 한다. 허리케인 카트리나는 미국 New Orleans에서 수많은 이재민을 발생시켰다. David Kirk (2009)는 특히 범죄우발지역에 살던 전과자들의 피난 이후 돌아올 곳이 있던 이들과 그렇지 못했던 이들을 비교하여, 범죄우발지역에 사는 것의 재범률에 대한 인과 효과를 찾아냈다.

Kirk, David S. 2009. "A Natural Experiment on Residential Change and Recidivism: Lessons from Hurricane Katrina." American Sociological Review 74(3): 484-505.

사회학에서 실험하기

실험에서 윤리문제는 언제나 민감한 주제이다.

- 많은 윤리 이슈가 실험방법과 관련하여 대두되었음을 떠올리자(Stanley Milgram과 Philip Zimbardo).
- 취약한 실험대상자 배려(vulnerable participants), 고지된 동의(informed consent), 개인정보 보호(privacy and confidentiality)가 키워드이다. 앞글자를 따면 VIP이다.

