

# 1. 과학적 연구와 이론, 인과관계와 연구설계

## 정치와 데이터분석

박상훈 ([sh.park.poli@gmail.com](mailto:sh.park.poli@gmail.com))

강원대학교

# 오늘의 목표

**10:10-10:50**

리딩 자료를 바탕으로 간단한 과학으로서의 정치학(political science)에 대한 내용들 소개

**11:00-11:50**

더 구체적으로 우리가 왜 '통계모형'을 배우고 '데이터'를 '분석'하는지에 대한 설명

**12:00-12:45**

실습과제 1 풀이 및 실습과제 2 설명

기타 R/RStudio/Quarto 설치 및 작동에 관한 질의응답

# 정치학의 과학적 연구

## 정치학의 과학적 접근

정치학은 단순히 정치적 '사실'만을 나열하는 학문이 아니라, 정치 현상의 원인에 대한 인과적 설명을 수립하고 검증하는 과학적 연구를 지향

- 이론적 접근은 복잡하고 끊임없이 변화하는 정치 세계를 체계적으로 이해하는 틀을 제공

## 과학적 지식으로 가는 길

과학적 지식은 다음과 같은 체계적인 과정을 통해 축적

- **인과 이론** → **가설** → **경험적 검증** → **가설의 평가** → **이론의 평가** → **과학적 지식**
- 모든 주장은 잠정적인 것으로 간주하고, 끊임없이 새로운 경험적 근거를 통해 검증
- 자신의 주장을 '증명'하려는 변호사와는 달리, 가장 엄격한 방법으로 이론이 틀릴 수 있음을 보이려 노력

# 정치학의 과학적 연구

## 변인과 인과관계로 생각하기

정치 현상을 이해하기 위해 세상을 변인(variable)로 파악

- 종속변인(dependent variable): 결과가 되는 변인
  - 내생변인(endogenous variable)이나 결과변인(outcome variable) 등으로 부르기도 함.
- 설명변인(explanatory variable): 결과를 설명하는 변인
  - 독립변인(independent variable), 외생변인(exogenous variable), 혹은 예측자(predictors) 등으로 부르기도 함.

**본 강의에서는 종속변인-설명변인(혹은 예측변인)이라고 함.**

# 정치학의 과학적 연구

## 변인과 인과관계로 생각하기

정치 현상을 이해하기 위해 세상을 변인(variable)로 파악

**이론(Theory):** 설명변인이 종속변인의 변화를 야기한다는 잠정적 추측

- 경제 상황이 대통령 선거 결과에 영향을 미친다.

**가설(Hypothesis):** 이론을 현실 세계에서 측정가능한 변인들 간의 관계로 구체화한 진술

- 1인당 실질 경제 성장률이 높을수록, 현직 여당의 득표율도 높을 것이다.

## 사회과학에서 이론이란 잠정적으로 검증된 위상의 가설

- 어떠한 이론도 완전무결하지 않음. 깨어질 수 있으며(반증가능성), 깨어진 이론은 다시 가설의 위치로 돌아가게 됨.

# 정치학의 과학적 연구

## 과학적 연구의 기본 규칙

1. 인과성(Causality): 단순한 상관관계(covariation)가 아니라, 무엇이 무엇을 야기하는지에 대해 집중. Why Y?
2. 이론이 데이터를 앞서야 함: 데이터에서 보이는 패턴만으로 이론을 도출하지 않음.
3. 경험적 증거(empirical evidence) 기반: 현실 세계의 관찰을 통해서만 이론을 검증
  - 우리는 어디까지나 연역적으로 이론을 도출하여 귀납적으로 그러한 이론적 기대에 반하는 혹은 상응하는 경험적 결과를 확인할 뿐임.
4. 규범적(normative) 진술 지양: '세상은 어떠해야 한다'는 주관적 가치 판단을 가능한 배제
5. 일반성(generality)와 간결성(parsimony) 추구: 이론은 가능한 한 넓은 범위의 현상을, 가장 단순하게 설명해야 함.

# 이론 구축의 기술

## 좋은 이론이란 무엇인가?

흥미로운 연구 질문에 대한 답을 제공하는 이론

세상을 이해하는 우리의 관점을 변화시키는 이론

경험적 검증을 통과하여 과학적 지식에 기여하는 이론

이론을 어떻게 만드는데 대한 정해진 공식은 없지만, 좋은 이론을 만들 확률을 높이는 전략은 존재

- 단순히 번개를 기다리는 것이 아니라, 번개가 칠 만한 곳으로 가서 피뢰침을 드는 것과 같음.

# 이론 구축의 기술

## 이론 구축의 출발점: 흥미로운 '변이' 찾기

설명하고자 하는 현상, 즉 종속변인에서 나타나는 흥미로운 변이를 발견하는 것은 이론 구축의 좋은 시작점

- 횡단면적(cross-sectional) 변이: 특정 시점의 여러 공간 단위(국가, 개인 등) 간 차이
  - 2005년 기준, 왜 어떤 국가는 다른 국가보다 GDP 대비 군비 지출 비율이 더 높을까?
- 시계열적(time-series) 변이: 하나의 공간 단위 내에서 시간의 흐름에 따른 변화
  - 미국에서 대통령 지지율은 왜 시간이 지남에 따라 오르내리는가?

# 이론 구축의 기술

## 독창적인 이론 구축을 위한 세 가지 전략

1. 새로운 종속변인 제시: 기존에 주목하지 않았던 새로운 현상을 정의하고, 그 원인을 탐구
2. 기존 종속변인에 대한 새로운 원인 제시: 이미 많이 연구된 현상에 대해 새로운 설명, 즉 새로운 독립변인을 제안
3. 새로운 조절변인 제시: 기존에 알려진 인과관계 ( $X \rightarrow Y$ )가 특정 조건 ( $Z$ )에 따라 어떻게 달라지는지를 설명

## 지식 활용 및 문헌 연구

구체적인 것에서 일반적인 것으로: 특정 사건이나 사례에서 출발하되, 일반화된 이론으로 발전

- 9.11 테러 직후 부시 대통령 지지율 급등에서 '국제 분쟁은 국가 지도자 지지율을 단기적으로 상승시킨다'라는 인과적 설명으로 발전

# 이론 구축의 기술

## 독창적인 이론 구축을 위한 세 가지 전략

1. 새로운 종속변인 제시: 기존에 주목하지 않았던 새로운 현상을 정의하고, 그 원인을 탐구
2. 기존 종속변인에 대한 새로운 원인 제시: 이미 많이 연구된 현상에 대해 새로운 설명, 즉 새로운 독립변인을 제안
3. 새로운 조절변인 제시: 기존에 알려진 인과관계 ( $X \rightarrow Y$ )가 특정 조건 ( $Z$ )에 따라 어떻게 달라지는지를 설명

## 지식 활용 및 문헌 연구

구체적인 것에서 일반적인 것으로: 특정 사건이나 사례에서 출발하되, 일반화된 이론으로 발전

내 아이디어가 정말 새로운 것인지 기존 연구들을 체계적으로 검토하고 기존 연구들이 놓친 다른 원인은 없는지, 다른 맥락에도 적용 가능한지 등의 비판적 검토

# 인과관계 평가하기

## 인과관계의 이해

상관관계  $\neq$  인과관계: 정치학의 과학적 연구는 두 변인이 단순히 함께 움직이는 것(상관관계)을 넘어, 하나가 다른 하나의 원인이 되는지(인과관계)를 밝히는 것

결정론적 관계 vs. 확률적 관계

- 결정론적: 원인이 발생하면 결과가 반드시 발생(자연과학, 법칙)
- 확률적: 원인이 발생하면 결과가 발생할 가능성이 높아짐(사회과학, 통칙)
  - "부유한 사람은 반드시 감세를 선호한다" (X)  
→ "부유한 사람은 감세를 선호할 경향이 있다." (O)

사회과학의 연구 대상이 되는 '인간' 혹은 '인간의 집단'은 자연과 달리 매우 가변적

# 인과관계 평가하기

## 인과관계 수립을 위한 관문

$X$ 가  $Y$ 의 원인이라고 주장하기 위해 몇 가지 논리적 관문을 통과해야 함. 이 관문들은 교재에 따라서, 연구자에 따라서 그 개수가 조금 다르기는 하지만 일맥상통함.

1. 인과 기제(causal mechanism):  $X$ 가  $Y$ 를 어떻게, 왜 변화시키는지에 대한 **설득력 있는 이야기**가 존재하는가?
2. 역인과관계(reverse causality) 혹은 내생성의 문제: 혹시  $Y$ 가  $X$ 의 원인은 아닌가? 이 가능성을 배제할 수 있는가?
3. 공변규칙성(covariation):  $X$ 와  $Y$ 는 실제로 함께 규칙적으로 움직이는가?
4. 허위적 관계(spurious relationship): 제3의 변인  $Z$ 가  $X$ 와  $Y$  모두에 영향을 주어, 둘 사이에 가짜 관계가 있는 것처럼 보이는 것은 아닌가?

# 인과관계 평가하기

예시: "사립고등학교에 가면 SKY에 갈 가능성이 높다"

주장: 사립고등학교 소속 여부 ( $X$ )  $\rightarrow$  SKY 입학 ( $Y$ )

1. 메커니즘: ○ (소규모 특별반, 더 나은 교사 등 그럴듯한 설명 존재)
2. 역인과관계: 학생들이 SKY에 입학한 다음에 다시 고등학생이 될 수는 없으므로 배제 가능
3. 공변: ○ (실제로 사립고 출신들이 평균적으로 SKY 입학 비율이 더 높음)
4. 허위관계: **주의!** '부모의 소득' ( $Z$ )이 사립고등학교 입학과 SKY 진학 모두에 영향을 미칠 수 있음. 이 ( $Z$ )를 통제하지 않으면  $X$ 와  $Y$ 의 관계는 허위일 가능성이 높음.

# 인과관계 평가하기

## 허위 관계와 혼란변인

혼란변인(confounding variable,  $Z$ ): 설명변인과 종속변인 모두와 관련이 있어 이 둘의 관계를 왜곡시킬 수 있는 제3의 변인

통제의 중요성: 혼란변인을 통제하지 못하는 것은 인과적 추론에서 저지를 수 있는 가장 심각한 실수

- 아이스크림 판매량과 범죄율 사이의 강한 양의 상관관계는 기온이라는 혼란변인에 의해 발생하는 허위 관계
- 우리는 이 혼란변인을 모두 찾아낼 수도, 혹은 잘못 식별할 수도 있음. 둘 중 어떤 것이 더 우리의 추론을 '왜곡'할 가능성이 클까?

# 통계모형과 현실

"모든 모델은 틀렸다(wrong), 다만 몇몇 모델만이 유용할 뿐이다" (Box 1979: 202).

- 복잡한 현실을 특정한 모형으로 모두 설명하는 것은 불가능
- 사회 구조 혹은 다른 체계적 요인들이 개인의 선택을 조건짓고, 제약
- 특정한 개인을 예측하는 것이 아니라 일반적 경향성을 포착하고자 함.

사회과학에는 수많은 '추상적인' 이론들이 존재

- 추상적인 사회이론들이 만약 복잡한 현실세계를 이해하는 데 유용한가?
- 만약 유용하다면, 이와 같은 이론(논리)과 현실(경험)을 어떻게 연결할 수 있을까?

# 통계모형과 현실

앞서 언급했던 바와 같이 이론은 특정 현상들을 '정의'(define)하고, 그 정의된 현상들 간의 관계를 서술, 설명, 및 예측

우리는 이 추상적인 이론을 구성하는 각 부분들이 무엇인지 설명하고, 평가하거나 측정하는 과정을 통해 이론과 현실을 연결(bridging)

- 교육 수준이 상승할 시, 소득 수준도 상승한다.
  - 교육 수준과 소득 수준을 어떻게 정의할 것인가?
  - 교육 수준과 소득 수준의 관계를 어떻게 설명할 것인가?
  - 교육 수준과 소득 수준을 실제로 어떻게 측정할 것인가?
  - 이 관계에 영향을 미치는 다른 요인은 존재하는가? 존재한다면 무엇인가?

# 통계모형과 현실

통계모형: 복잡한 현실을 단순하게(*simplified*) 서술한 것 (Fox 2016, 2)

- 고용노동자들에 대한 Large-N 표본의 설문조사 자료
- 교육 수준, 성별, 인종, 거주지역 등과 같은 특성들로 개인의 소득에 대한 회귀분석을 시행하였다고 가정
  - 과연 위의 모델로 개개인의 소득 수준을 완벽하게 설명할 수 있는가?
- 효과(effect): 관측하여 변인으로 만들어 모델에 포함한 것들이 종속변인(소득)의 '체계적'(systematic) 변화에 미치는 영향
- 잔차(residuals): 관측한 것들로 설명하지 못하는 불확실성, 변인들의 효과로 설명하지 못하는 종속변인의 변화

# 통계모형과 현실

통계모델의 잔차는 클 수도, 작을 수도 있음.

- 하지만 잔차가 아무리 작다라고 하더라도 모델이 완벽하게 현실을 설명, 예측할 수 있다고 할 수 없음.
- 통계모델은 근본적으로 서술적(*descriptive*) (Fox 2016, 3)
  - 어디까지나 논리적인 기대(이론)와 일치하는 방향으로 경험적인 관측(데이터)가 배열되어 있는지를 보여줄 뿐임.
  - 따라서 통계모델은 항상 불확실성을 내재하지만 그럼에도 변인들의 체계적 효과, 유의미한 측면을 보여줌.

# 통계모형과 현실

정리하자면, 통계모델은 그 자체로 현실을 직접적으로 보여주는 것이 아니라 간단하게 유의미한 특성들만을 구조화하여 보여줄 뿐

- 구조적 조응성을 가지지만 기능적 조응성을 가지지는 않음.
- 모형 비행기를 통해 실제 비행기의 구조를 이해하는 데 도움을 얻을 수 있지만, 그렇다고 모형 비행기가 실제로 날 수 있는 것은 아님.

통계모델의 실용성(practice)과 정확성(accuracy)

- 데이터(현실)을 정확하게 보여주지 못하는 모델은 쓸모없지만 정확하게 현실을 보여주는 모델이라고 하더라도 모두 쓸모 있는 것은 아님.
  - 통계모델 자체는 굉장히 정교하고 정확하더라도 실제 현실을 설명하는 데 아무 도움이 되지 않을 수 있음.
  - 모델 이전에 이론 수립(theory-building)이 중요한 이유

# 연구설계의 핵심: 비교(comparison)

우리는 기본적으로 인과관계를 밝히기 위해, 다른 모든 조건들이 동일한 상태에서 오직 설명변인 ( $X$ )의 값만 다른 두 집단을 비교하고자 함.

따라서 좋은 모형은 이러한 '순수한 비교'가 가능하도록 하여, 다른 요인( $Z$ )의 개입을 최소화하는 모형

이러한 모형을 만들기위한 설계에는 크게 두 가지가 있음

1. 실험연구(experimental study)
2. 관찰연구(observational study): 횡단면 연구, 시계열 연구

# 연구설계의 핵심: 비교(comparison)

## 실험연구설계(experimental design)

연구자가 처치(treatment,  $X$ )의 값을 통제(control)하고, 참가자를 여러 집단에 무작위로 할당(random assignment)하는 연구 설계

- 무작위 할당을 통해 우리가 알고 있는 요인뿐 아니라, 알지 못하는 모든 잠재적 혼란변인 ( $Z$ )까지 통계적으로 동일하게 만들어 줌.
- 여기서 중요한 개념은 바로 **무작위화(randomization)**
- 인과추론에 대한 확신, 즉 내적 타당도(internal validity)가 매우 높음.
  - 네거티브 광고의 투표 의향에 대한 효과를 알아보기 위해, 참가자들을 무작위로 광고 시청 집단(실험군)과 비시청 집단(통제군)으로 나눈 후 결과를 비교

# 연구설계의 핵심: 비교(comparison)

## 실험연구설계(experimental design)

하지만 실험연구의 경우에는 성별, 인종, 국가 체제 등 사회과학의 많은 중요 변인들을 인위적으로 조작하거나 무작위 할당하는 것이 불가능하다는 한계(조작 불가능성)

낮은 외적 타당도(low external validity)

- 실험실이라는 인위적 환경이 현실 세계와 달라 결과의 일반화가 어려울 수 있음.
- 대부분 편의 표본(sample of convenience)으로 대학생 설문 등을 사용하므로, 연구 결과를 전체 인구에 적용하기 어려움.
  - 하지만 최근 [Krupnikov, Nam, and Style \(2021\)](#) 등의 연구는 여러가지 우려에도 불구하고 편의표본도 신뢰할 수 있는 대표성 있는 표본으로 기능한다는 결과를 제시

윤리적 문제: 연구 참여자에게 잠재적 해를 끼칠 수 있는 실험은 수행할 수 없음.

# 연구설계의 핵심: 비교(comparison)

## 관찰연구설계(experimental design)

연구자가 설명변인을 통제하지 않고, 현실에 자연적으로 존재하는 데이터를 '관찰'하여 인과관계를 추론하는 연구설계

- 앞서 언급한 바와 같이 횡단면 연구, 시계열 연구가 대표적
- 혹은 횡단면의 공간 단위들을 여러 시간에서 반복 관찰하는 교차사례-시계열(cross-sectional time-series) 연구나 동일한 개체를 시간에 따라 반복하여 관측하는 패널 자료도 존재
  - CSTS: 개체  $i$ 가 동일한 개체는 아님. 2000년부터 2024년 미국의 GDP를 측정한다고 했을 때, 미국이라는 국가는 개념 상 동일하지만 우리가 측정한 GDP는 완벽히 동일한 개체로부터 얻어낸 것이 아님.
  - 패널: 설문조사를 할 때 2000년부터 2024년까지 동일한 사람을 추적하여 응답을 측정

# 연구설계의 핵심: 비교(comparison)

## 관찰연구설계(experimental design)

관찰연구설계는 인과관계(causality)을 담보하지 않음. 이에 따라 King, Keohane, and Verva (1994)는 관찰자료를 사용한 연구방법을 준실험설계(quasi-experimental design)이라고 부르기도 함.

- 적절한 변인을 모델에서 누락할 수 있음.
  - 관찰되지 않은 혼란변인(confounding variables)이 존재할 경우
  - 만약 종속변인에 영향을 미치는 변인인데 모델에 들어가지 않았다면 통제가 충분하지 못했다는 것
  - 만약 누락된 변인이 다른 설명변인들에 영향을 미치는 변인이라면 간접적인 방식으로 종속변인에 영향을 줄 수 있으므로 마찬가지로 통제가 충분히 이루어지지 않았다는 것

# 연구설계의 핵심: 비교(comparison)

## 관찰연구설계(experimental design)

관찰연구설계는 인과관계(causality)을 담보하지 않음. 이에 따라 King, Keohane, and Verva (1994)는 관찰자료를 사용한 연구방법을 준실험설계(quasi-experimental design)이라고 부르기도 함.

- 적절한 변인을 모델에서 누락할 수 있음.
- 원인과 결과 간에 존재하는 무수히 많은 잠재적인 영향들을 배제할 수 없음.
- 설문조사와 같은 경우, 문항들은 약간의 시간 차는 있겠지만 거의 동일한 시점에서 측정된 것. 원인과 결과 간의 시간적 순차를 확인할 수 없음.

# 연구설계의 핵심: 비교(comparison)

## 실험연구설계와 관찰연구설계

관찰자료를 이용할 경우, 종속변인에 가장 우선적인 원인이 되는 설명변인과 그 설명변인-종속변인 관계에 끼어드는 변인을 구별하는 것이 중요

사회과학에서 "원인"(cause)이라는 개념은 조금 느슨하게 사용됨.

# 연구설계의 핵심: 비교(comparison)

## 실험연구설계와 관찰연구설계

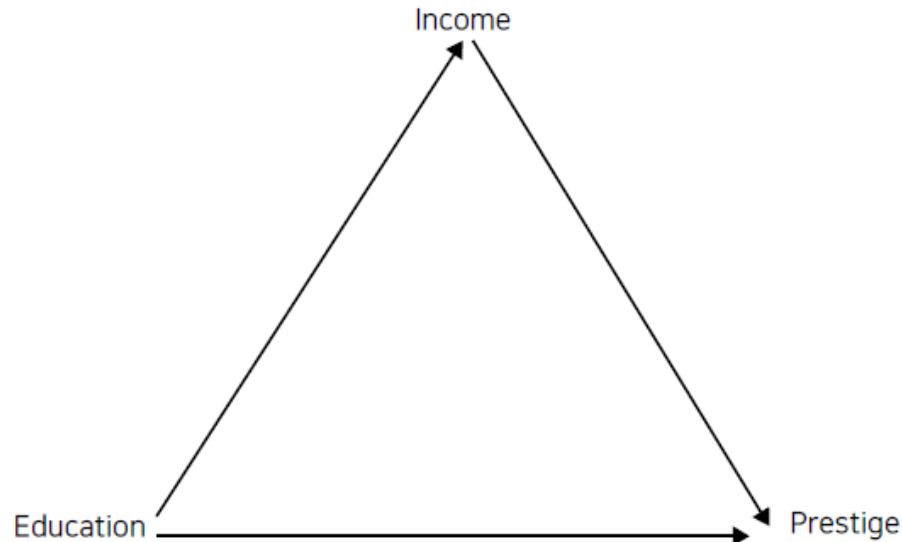
관찰연구설계는 통제가 제대로 이루어졌다는 확신을 할 수 없기 때문에 관측자료를 바탕으로 한 통계분석을 인과적으로 해석하는 데 한계가 존재

- 상대적으로 실험연구설계는 관찰연구설계에 비해 통제가 용이
- 그러나 설명변인의 식별이라는 문제에 있어서는 마찬가지로 불확실성이 내재
- 어떠한 방법으로도 *모든* 적절한 변인들이 통제되었다고 확신할 수는 없음.
  - 연구 대상인 사회현상과 인간 행태가 가지는 본연적 불확실성

**Effect of Cause vs. Cause of Effect**의 문제도 존재

# 연구설계의 핵심: 비교(comparison)

## 실험연구설계와 관찰연구설계



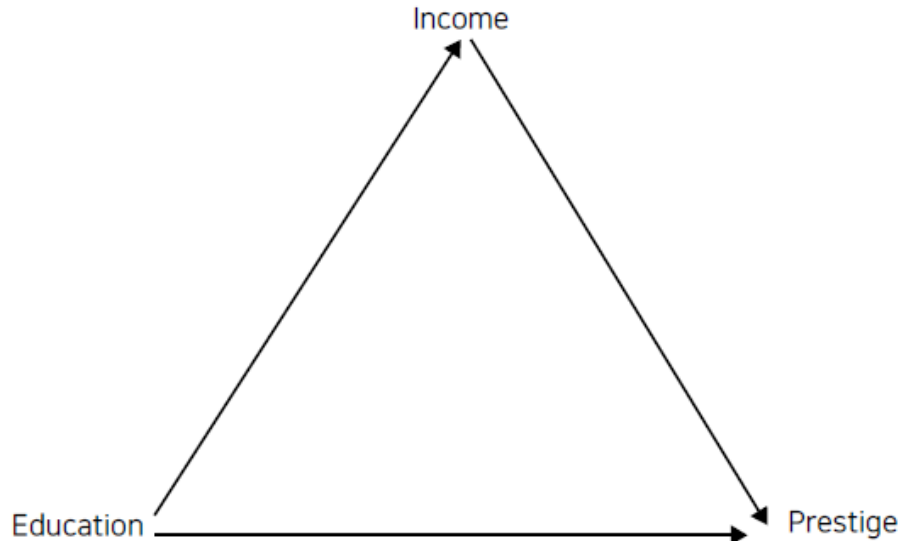
교육, 소득, 직업에 관한 인과모델

소득 수준은 직업 수준에만 영향을 미치지 만, 교육 수준은 소득 수준과 직업 수준 모두에게 영향을 미침.

- 직업 수준과 소득 수준의 관계는 어느 정도 허위적
  - 교육 수준을 통제했을 때, 소득 → 직업의 효과는 감소할 것: 허위적 요소의 제거

# 연구설계의 핵심: 비교(comparison)

## 실험연구설계와 관찰연구설계



교육, 소득, 직업에 관한 인과모델

소득 수준은 직업 수준에만 영향을 미치지  
만, 교육 수준은 소득 수준과 직업 수준 모  
두에게 영향을 미침.

- 반대로 교육 수준과 직업 수준의 관계  
는 소득 수준에 의해 어느 정도 매개됨  
(mediated).
  - 소득 수준을 통제했을 때, 교육 →  
직업의 효과가 감소하였다면, 그 감  
소한 만큼이 교육이 소득을 경유해  
직업 수준에 미치는 '간접적 효과'

# 모집단과 표본

**연구 문제:** 고등학교 3학년 수험생이 공부를 오래할수록 모의고사 성적을 더 잘 받을까?

- 모집단: 전체 고등학교 3학년 수험생
- 표본: 모의고사 시험을 친 고등학교 3학년 수험생 집단

# 모집단과 표본

**연구 문제:** 고등학교 3학년 수험생이 공부를 오래할수록 모의고사 성적을 더 잘 받을까?

이때, 표본은 전체 고등학교 3학년 수험생에 대한 대표성을 지닌, 그보다 작은 규모의 집단을 의미

- 우리가 관심을 가지고 있는 것은 '일반화할 수 있는' 모집단
- 하지만 현실에서 모집단을 관측하기란 불가능
- 따라서 우리는 모집단을 대표할 수 있는 표본의 특성을 통해 모집단 또한 그러할 것이라는 추론(inference)을 하게 됨.
  - 기술추론(descriptive inference): 표본의 생김새를 보고 모집단도 어떻게 생겼을 것이라는 추론
  - 인과추론(causal inference): 표본에서 어떠한 변인들 간의 관계가 모집단 수준에서도 그러할 것이라 추론

## 모집단과 표본

통계 추론은 어느 정도 불확실성을 내재한 채로 수집된 자료들이 나타내는 경향성(pattern)의 안정성에 관한 것

- 실험 자료를 통해서 우리는 그 관계가 안정적일 때, 인과성(causation)을 얘기할 수 있음.
- 마찬가지로 불확실성을 내재한 자료들을 통해 모집단에 관한 일반화(generalization)를 할 수 있지만, 어디까지나 가치판단의 문제와 직결

# 데이터

변인의 측정 수준, 시각화, 그리고 기초통계

- 앞으로 다룰 여러 모델들은 데이터의 특성에 대해서 서로 다른 가정을 가지고 있음.
- 다음 주는 다른 유형의 데이터와 변인들의 관계를 통계적 개념으로 이해하는 것에 목적이 있음.

사회과학자들은 연구과정에 데이터 그 자체를 들여다보는 것에 시간을 할애하지 않는 경향 → 왜 데이터를 보아야 할까?

- 첫째, 데이터는 우리에게 현실 세계가 어떻게 돌아가고 있는지를 알려주는 단서임.
- 둘째, 연구자로 하여금 현실에 불가피하게 존재하는 오차(errors)를 확인할 수 있게 함.
- 셋째, 데이터를 통해 이론화되지 않은 무언가를 파악할 수 있음.

# 데이터

주어진 시간에 대해 하나 이상의 변인들을 분석하는 것은 연구에서 매우 중요

- 데이터 탐색(exploration), 모델 특정(specification), 진단(diagnostics)

Example data: Anscombe (1973)

x	y1	y2	y3	x4	y4
10	8.04	9.14	7.46	8	6.58
8	6.95	8.14	6.77	8	5.76
13	7.58	8.74	12.74	8	7.71
9	8.81	8.77	7.11	8	8.84
11	8.33	9.26	7.81	8	8.47
14	9.96	8.10	8.84	8	7.04

# 데이터

---

R Code

---

Data Table

Explanation

```
anscombe |> psych::describe() |>
  dplyr::select(0bs. = n, Mean = mean, Std. = sd, Min = min, Max = max) |>
  knitr::kable(
    caption = "Descriptive Statistics of Anscombe (1973)" |>
    kableExtra::kable_styling() |>
    kableExtra::column_spec(c(1:5), width = "5em")
```

# 데이터

---

R Code	Plot	Explanation
--------	------	-------------

---

```
lm1 <- lm(y1 ~ x, data = anscombe)
lm2 <- lm(y2 ~ x, data = anscombe)
lm3 <- lm(y3 ~ x, data = anscombe)
lm4 <- lm(y4 ~ x4, data = anscombe)




with(anscombe, {
  par(mfrow = c(2, 2), mar = c(3, 3, 3, 3))
  plot(x, y1, main = "(a)",
       xlim=c(0, 20), ylim=c(0, 15)) + abline(coefficients(lm1))
  plot(x, y2, main = "(b)",
       xlim=c(0, 20), ylim=c(0, 15)) + abline(coefficients(lm2))
  plot(x, y3, main = "(c)",
       xlim=c(0, 20), ylim=c(0, 15)) + abline(coefficients(lm3))
  plot(x4, y4, main = "(d)",
       xlim=c(0, 20), ylim=c(0, 15)) + abline(coefficients(lm4))
})
```

## **실습과제 1 검토 및 실습과제 2 설명, 질의응답**

## 감사합니다!

궁금한 것이 있으면 언제든지 연락하세요.

강사 연락처

연락처	박상훈
	<a href="mailto:sh.park.poli@gmail.com">sh.park.poli@gmail.com</a>
	<a href="http://sanghoon-park.com/">sanghoon-park.com/</a>
	영상바이오관 405