

# 1.1 통계학이란 무엇인가

- 1) 왜 통계학을 배워야 하나
- 2) 자료

# 1) 왜 통계학을 배워야 하나

## 통계학의 유용성

경제학, 경영학, 사회과학, 인문과학, 법학, 의학, 공학, 자연과학 등 분야를 막론하고 통계학을 쓴다. 우리는 통계에 묻혀 산다.

- 10% 담뱃값 인상이 청소년층의 흡연을 어느 정도 줄이나?
- 직업훈련은 재 취업률을 높이는가? 임금률에 미치는 효과는?
- 외국인 직접투자가 늘면 경제성장률이 제고되는가?
- 특정 광고가 매출증가 효과를 가져왔는가?
- 노동시장에서 여성에 대한 차별이 존재하는가?

# 1) 왜 통계학을 배워야 하나

## 월별 출생률

Births are not uniformly distributed over a year.

- The typical United States pattern for births is an April-May trough.
- April-May births correspond to July-August conceptions, hottest months in the States.
- This trend is more pronounced in the southern hot states of Georgia and Louisiana than in New York.
- Less visible April-May trough as air-conditioning spreads.

# 1) 왜 통계학을 배워야 하나

## 통제된 실험에서 두 비율의 비교

‘인간의 합리성’ 가정을 검증하기 위한 실험.

의사집단 1: 서식A를 제공. 의사집단 2: 서식B를 제공.

[서식 A] 수술환자 100명 중에서 10명은 수술 도중에 죽고, 32명은 1년 이내에 죽으며, 66명은 5년 이내에 죽는다. 방사선 치료를 받는 100명의 환자 중에서는 아무도 치료 도중에 죽지 않고, 23명이 1년 이내에 죽으며, 78명이 5년 이내에 죽는다.

[서식 B] 수술환자 100명 중에서 90명은 살아서 수술 기간을 넘기고, 68명은 일년 이상 살아남으며, 34명은 5년 이상 살아남는다. 방사선 치료를 받는 100명의 환자 중에서는 모두 다 살아서 치료기간을 넘기고, 77명이 1년 이상 살며, 22명이 5년 이상 산다.

서식을 읽은 후 의사들은 각자 폐암 환자에게 추천할 치료법을 제시.

- 의사집단 1: 80명 중 40명(50%)이 수술을 추천.
- 의사집단 2: 87명 중 73명(84%)이 수술을 추천.

# 1) 왜 통계학을 배워야 하나

## 영터리 통계의 예

- “딸만 둘 낳았으니 이제 아들 낳을 차례다.” - 시어머니 말씀 중 -
- “주식은 저점에서 사고, 고점에서 팔아야...” - 투자 전문가 조언 중
- “원칙을 지키며 펀더멘털에 따라 투자한 결과 3개월 만에 800%가 넘는 높은 수익률을 올렸습니다.” - 대학생 모의주식투자 대회 1등 팀에 대한 전문가 평가 중 -
- “수능 성적과 입학 후 학점간에는 상관관계가 낮습니다. 수능 성적은 입학 후 성취를 예측하는데 별로 도움이 되지 못합니다.” - 모 대학 교무처 발표 중 -

# 1) 왜 통계학을 배워야 하나

영어리 통계의 예



각 타석은 독립입니다.

류근관. (2013). 통계학, 제 3 판. 서울: 법문사. P. 237

# 1) 왜 통계학을 배워야 하나

## 영터리 통계의 예

### 아동학대, 친부모 '80%' 계부모 '10%미만'(조선일보, 2003.05.05)

- 보건복지부가 전국 17개 시,도 아동학대 예방센터와 신고전화 '1391'을 통해 접수한 아동학대 사례는 모두 4,111건이며, 이 중 2,478건이 실제 아동학대.
- 복지부에 신고된 학대 사례들도 사망 등 극한 상황으로 가는 특수한 경우가 아니라, 자기 집에서 친부모에게 거의 매일 학대를 당한 사례가 주종을 이뤘다. 가해자의 80%는 친부모이며, 계부모와 양부모는 각각 10% 미만이었다.

**잘못의 지적:** 친부모가 계부모보다 자식을 8배 더 학대하는 경향이 있는 게 아니라 대부분의 아이가 친부모 밑에서 자라고 있음.

**올바른 분석:** 친부모 슬하의 자녀들 집단, 계부모 슬하의 자녀들 집단 등 두 집단간 매맞는 아이의 비율을 비교하는 이른바 two-sample analysis해야 타당함.

# 1) 왜 통계학을 배워야 하나

## 영터리 통계의 예

### '속도의 유혹', "곧은 길서 사고 더 많아요" (중앙일보, 2006.10.9)

- 곧게 뻗은 직선도로에서 속도의 유혹을 이기지 못하는 바람에 발생하는 교통사고가 전체 교통사고 10건 중 9건이나 된다고 합니다. 건설교통부와 경찰청이 지난해 발생한 교통사고 21만4000여 건을 분석한 결과입니다.
- 쪽 뻗은 직선 도로에 올라서면 가슴이 뻥 뚫리는 듯한 느낌을 받으신 적이 있을 겁니다. 차도 별로 많지 않다면 자연스레 액셀러레이터를 밟는 발에 힘이 들어가게 됩니다. 악마의 유혹. 사고의 위험성이 순간적으로 높아지는 순간입니다.

**잘못의 지적:** 전체 도로 중 대부분이 곧은 도로이고 굽은 도로나 오르막/내리막 길은 상대적으로 적음.

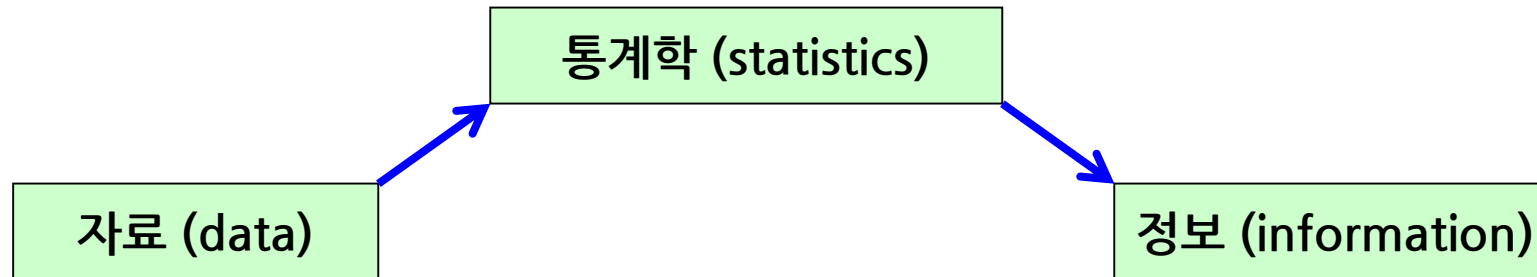
**올바른 분석:** 쪽 뻗은 직선 도로 주행 100,000km 당 사고 발생률과 그렇지 않은 도로 주행 100,000km 당 사고발생률을 상호 비교해 보아야 함.



# 1) 왜 통계학을 배워야 하나

통계학이란 무엇인가?

세상은 자료들로 가득 차 있다.



통계학은 자료를 정리/분석해 유용한 정보를 얻기 위한 언어이자 도구임.

# 1) 왜 통계학을 배워야 하나

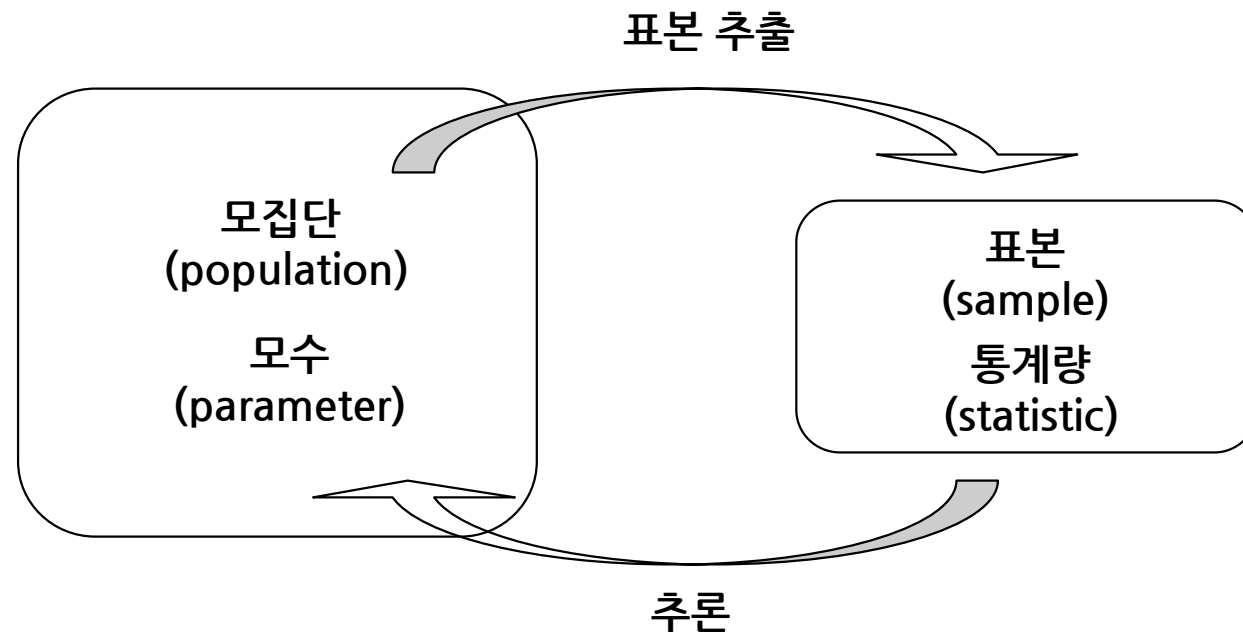
## 통계학의 분류

통계학은 크게 두 가지로 분류할 수 있다.

- 기술통계학 (descriptive statistics)
  - 자료를 변수 별로 따로따로 또는 관계되는 변수끼리 묶어서 요약.
- 추론통계학(inferential statistics)
  - 정리된 자료에 담긴 의미를 해석하여 미지의 세계에 대해 추론.

## 2) 자료

모집단과 표본 사이의 관계



통계학은 표본의 자료를 수집, 정리, 요약하고 나아가 요약된 자료를 토대로 그 자료의 모태가 되는 모집단에 대해 짐작, 추측해 보는 작업을 포함.

## 2) 자료

### 자료의 종류

횡단면 자료(cross-sectional data)

- 한 시점에서 여러 개체를 관측한 자료.

시계열 자료(time-series data)

- 한 개체를 여러 시점에 걸쳐 관측한 자료.

패널 자료 (panel data) 또는 종적 자료 (longitudinal data)

- 횡단면과 시계열의 특성을 결합하여 여러 개체를 여러 시점에 걸쳐 관측한 자료.

## 2) 자료

### 국내외 패널자료 예시 1

- 한국노동패널
- PSID, NLS, NLSY
- BHPS
- JPSC: Japanese panel survey on consumers
- Compustat, CRSP
- Scanner data (마케팅 분야)

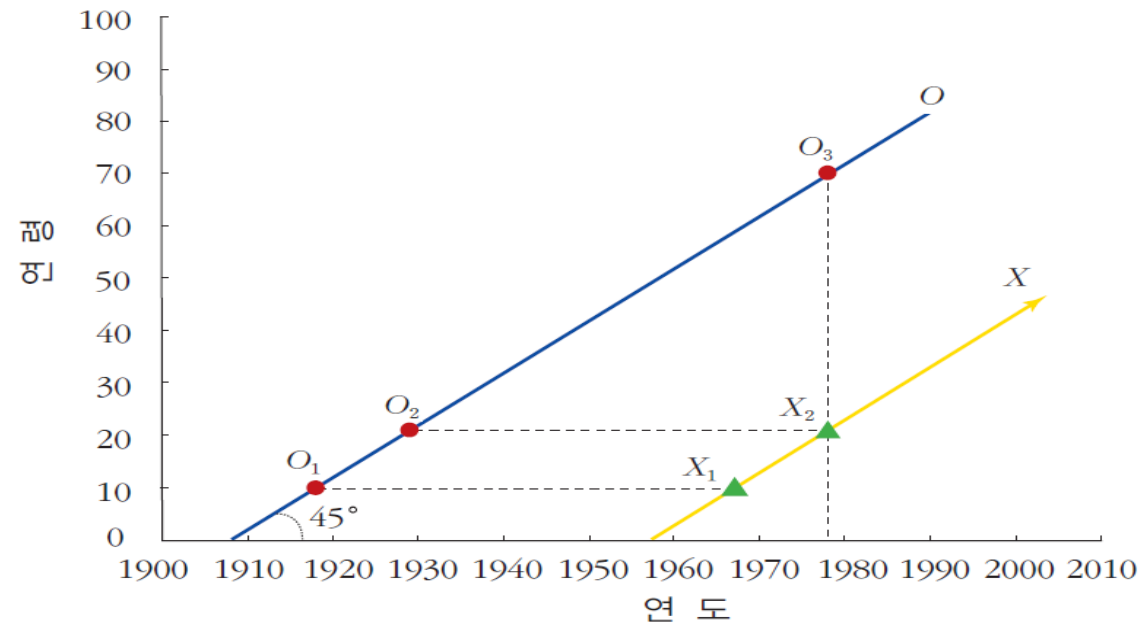
## 2) 자료

### 국내외 패널자료 예시 2

- 신용카드 보유자의 월별 업종별 지출액 패널 자료
- 통신회사(전화, 인터넷)가 보유한 개인별 통신 사용자료
- 전기 등 각종 공과금의 월별 지출액 자료
- 은행의 기업/가계별 대출 및 사후 관리 자료
- 기업의 신용등급 변화 자료
- 결혼정보회사의 개인별 matching history 자료

## 2) 자료

### 렉시스 도표



주:  $O_2$ 와  $X_2$ 는 같은 나이를,  $O_3$ 와  $X_2$ 는 같은 시대를 나타낸다.

류근관. (2013). 통계학, 제 3 판. 서울: 법문사. P. 7

종적 분석은 여러 개체를 여러 시점에 걸쳐 다양하게 비교함으로써 보다 풍부한 분석을 가능하게 해줌.