

# **1. What is regression? 처음부터 A. Simple regression 까지**

202304708 한서현

# 순서

1. 회귀분석이란?
2. 단순회귀분석
3. 회귀모형의 구성요소와 변수 분류
4. 산점도와 데이터 시각화 & 선형 관계와 직선의 방정식
5. 노이즈 항의 가정
6. 최소제곱법

# 회귀분석이란?

변수들 간의 관계를 정량화하는 통계적 방법

ex) 노동시장에서 소득을 결정하는 요인을 파악하고 측정하고자 할 때 사용

# 단순회귀분석

하나의 독립 변수가 종속변수에 미치는 영향을 추정하는 방법

## 교육과 소득

교육 수준이 높을수록 소득이 높아지는 경향

## 인과관계

교육 → 소득 관계가 고득 → 교육보다 더 타당성이 높음

## 가설

조건 동일 시, 교육 수준 ↑ → 소득 수준 ↑

# 회귀모형의 구성요소와 변수 분류

$$I = \alpha + \beta E + \varepsilon$$

01

## 기본 소득, 상수항 ( $\alpha$ )

교육이 전혀 없어도 얻는 기본적인 소득 수준을 나타내는 상수항입니다.

02

## 교육의 효과, 계수 ( $\beta$ )

추가적인 1년의 교육이 소득에 미치는 달러 단위의 효과로, 양수일 것으로 가정합니다.

03

## 노이즈 항, 오차항 ( $\varepsilon$ )

소득에 영향을 미치는 다른 요인들을 반영하는 "노이즈" 항입니다.

04

## I(종속변수)

내생변수라고도 함

05

## E(독립변수)

- 설명변수 또는 외생변수라고도 함

# 산점도와 데이터 시각화 & 선형 관계와 직선의 방정식

- 개인들의 교육연수( $E$ )와 연간 소득( $I$ ) 데이터로 2차원 산점도
- 교육이 많을수록 소득이 높아짐
- 관계는 완벽하지 않으며, 교육만으로 소득을 정확히 예측하기 어려움
- 노이즈 항( $\varepsilon$ )을 제외하면  $I = \alpha + \beta E$ 라는 직선 방정식이 성립
- $\alpha$ 는 소득의 절편,  $\beta$ 는 교육 1년 증가 시 소득의 증가분
- 산점도의 각 점과 이 직선 간 오차( $\varepsilon$ )를 최소화하는 선이 회귀선

$E$

# 노이즈 항의 가정

1

## 체계적 편향 없음

노이즈 항이 대체로 큰 음수나 양수가 아닌,  
평균적으로 0과 같다고 가정

2

## 직선의 위치

데이터의 중간 정도에 위치하는 직선을 추정  
하며, 일부 관측치는 위에, 일부는 아래에 위  
치

3

## 최적 직선 선택

여러 가능한 직선 중에서 특정 기준에 따라  
하나의 직선을 선택

# 최소제곱법



## SSE 최소화

추정오차의 제곱합이 최소가 되는 직선을 선택하는 최소제곱오차(minimum SSE) 기준을 사용



## 계산의 편의성

SSE 기준은 계산상 매우 사용하기 쉽고, 컴퓨터로 쉽게 평가할 수 있는  $\alpha$ 와  $\beta$ 의 표현식을 제공



## 통계적 특성

노이즈 항에 대한 그럴듯한 가정 하에서 매력적인 통계적 특성