

인과추론 한눈에 보기 : 정의와 필수 개념

1. 인과추론 정의

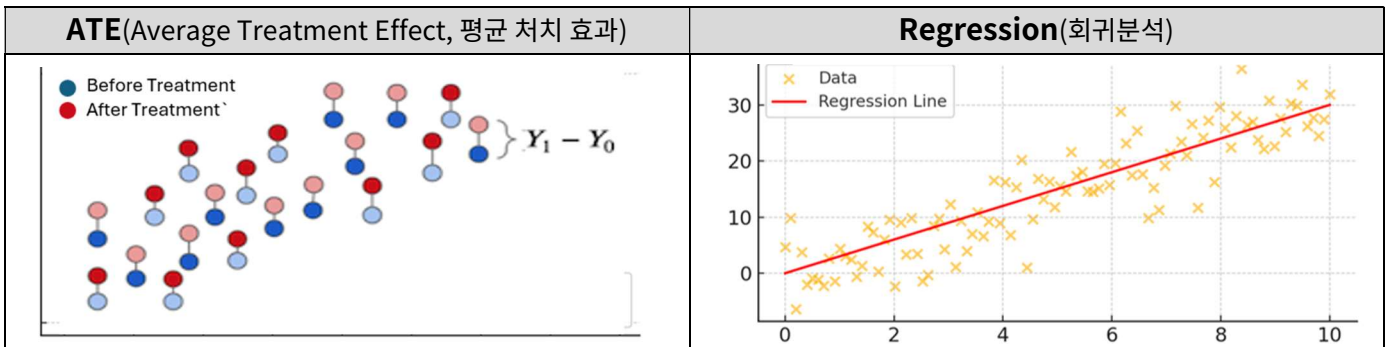
인과추론(Causal Inference)은 변수 간 상관관계가 아닌 **인과관계를 식별하고 측정하는 분석 방법**입니다. 이는 특정 행동, 정책, 또는 처치(Treatment)가 결과에 미치는 영향을 정량적으로 파악하는 데 초점을 맞춥니다. 예를 들어 광고 캠페인이 매출에 미치는 영향이나 정책 시행이 시민 행동에 미치는 영향을 평가할 때 활용됩니다.

2. 인과추론 필요성: 기존 상관관계 분석과 차이점

기존 데이터 분석은 변수 간 상관관계를 탐구하지만, **교란요인(Confounder)**을 통제하지 못해 **원인과 결과를 구분하기 어렵습니다**. 예를 들어, 아이스크림 판매와 익사 사고는 상관관계가 있지만 온도라는 교란요인의 영향을 받습니다. 인과추론은 교란요인을 통제해 특정 요인이 결과에 미치는 순수한 효과를 측정합니다.

3. 주요 개념

- **Treatment(처치)**: 분석에서 특정 효과를 발생시키는 주요 변수
- **Control(통제군)**: 처치를 받지 않은 비교 그룹으로, 처치의 효과를 평가하기 위한 기준
- **Confounder(교란요인)**: 처치와 결과에 영향을 미칠 수 있는 외부요인으로 통제되지 않으면 왜곡된 결론 초래 가능
- **RCT (Randomized Controlled Trials, 무작위대조실험)**: 무작위로 처치군, 통제군 배정하여 인과관계 식별하는 실험
- **IV(Instrumental Variable, 도구 변수)**: 교란요인 통제하기 위해 외부적 요인을 도구 변수로 활용하는 방법
- **ATE(Average Treatment Effect, 평균 처치 효과)**: 처치군과 통제군 간의 결과 차이를 분석하여 효과 추정하는 방법
- **Regression(회귀분석)**: 결과 변수에 대한 독립 변수들의 영향을 추정하여 관계를 분석하는 방법



4. 주요 분석 방법

- **DiD(Difference-in-Differences, 이중차분법)**: 처치 시행 전후 처치군, 통제군 간 변화량 비교하여 효과 추정하는 방법
- **RD(Regression Discontinuity, 회귀불연속분석)**: 특정 기준점을 기준으로 결과변수의 변화 분석하여 효과 추정하는 방법

