



# Causal Graph

2025 Spring

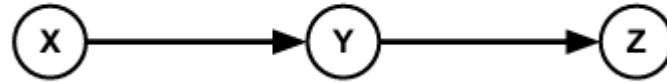
## 10. Causal Graph

- Causal Graph (인과 그래프)
  - 변수들 간의 인과관계를 화살표로 연결한 다이어그램
  - 변수 간의 인과 연결 구조를 시각화
  - 주요 특징
    - 통계 모델(statistical model)을 시각적으로 표현
    - 변수 간 인과 방향성을 명확히 표시
    - 다양한 연구 설계에 대한 교환성(exchangeability) 확보에 도움

## 10. Causal Graph

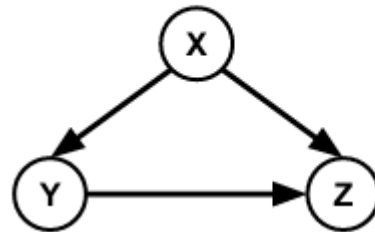
- Causal Graph (인과 그래프)의 예

**Adjacent nodes:** X and Y, Y and Z

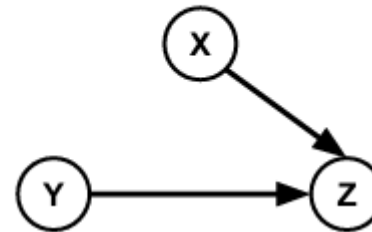


**Non-adjacent nodes:**  
X and Z

**Complete:**

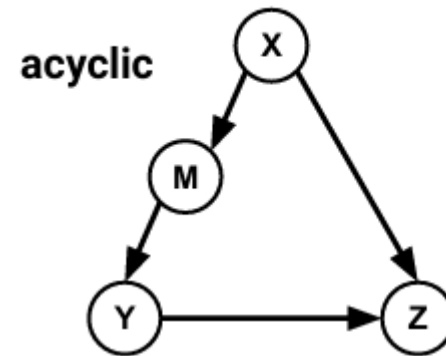


**Incomplete:**



## 10. Causal Graph

- Causal Graph (인과 그래프)의 예2



**Paths:**

$X \rightarrow M \rightarrow Y$

**directed path**

$X \rightarrow Z \leftarrow Y$

**undirected path**



**Cycle:**

$X \rightarrow Y \rightarrow X$

## 10. Causal Graph

- Causal Graph (인과 그래프)

- Directed Acyclic graphs(DAG)

- 노드: 변수

- 엣지: 시간에 따른 흐름

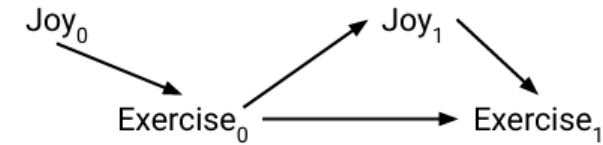
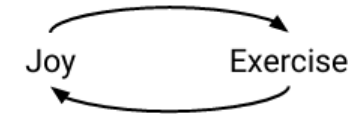
- 사이클 없음

- Feedback loop

- Cycle이 없음

- Joy, exercise는 positive feedback loop

- joy, exercise를 시간을 고려해 분리: time varyii



## 10. Causal Graph

- **Causal Graph (인과 그래프)**
  - 통계 모형으로서의 causal graph
    - 인과 그래프는 단순한 그림이 아니라 통계적 모델
    - 노드(Node): 확률 변수(Random Variable)를 의미
    - 엣지(Edge): 어떤 변수들이 다른 변수들의 함수로 표현될 수 있음을 의미
      - $A \rightarrow B$  : B는 A에 조건부로 의존함
    - 자식 노드의 분포는 부모 노드의 값에 조건부로 의존
    - 통계적 의존 구조를 명확하게 표현 가능

## 10. Causal Graph

- **Causal Graph (인과 그래프)**

- 인과 그래프를 통계 모델로 해석

- 교환성: 두 집단(예: 치료군 vs 대조군)의 결과 분포가 같을 것이라고 가정할 수 있을 때 이 두 집단은 교환 가능(exchangeable), 치료 여부 외에는 다른 차이가 없음

- 교환성의 결여 (Lack of Exchangeability)가 발생하는 이유 1

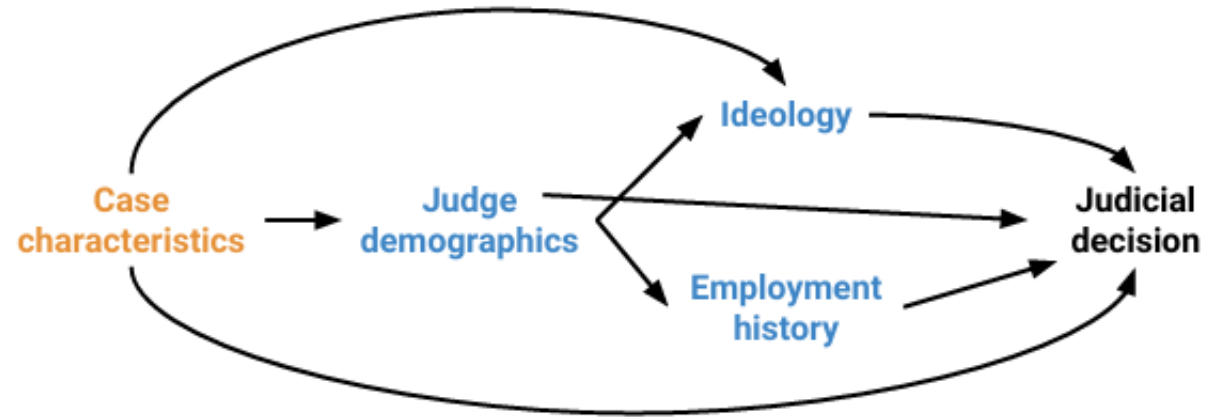
- 혼란 변수(confounder) : 나이, 건강 상태, 소득 등 영향을 주는 다른 변수, 예를 들어, 나이가 많은 사람이 치료를 더 많이 받았고, 또 나이 때문에 사망률이 높다면 결과(사망률)가 치료 때문인지 나이 때문인지 구분이 안 됨
      - 선택 편향(selection bias): 치료나 노출이 무작위가 아닌 경우, 예를 들어, 스스로 병원에 온 사람만 연구에 포함되면 일반 인구와 다름

## 10. Causal Graph

- **Causal Graph (인과 그래프)**
  - 인과 그래프를 통계 모델로 해석
    - 교환성의 결여 (Lack of Exchangeability)가 발생하는 이유 2
      - 측정 오류(measurement error): 변수 측정이 부정확할 경우, 진짜 관계를 잘 나타내지 못하거나, 혹은 시간적 순서가 모호하면, 원인과 결과가 뒤섞이거나 동시에 발생되어 인과 방향 추정 어려움
    - 조건부 교환성(Conditional Exchangeability)
      - 혼란 요인 통제
      - 최종 목적:인과 효과(Causal Effect)를 추정



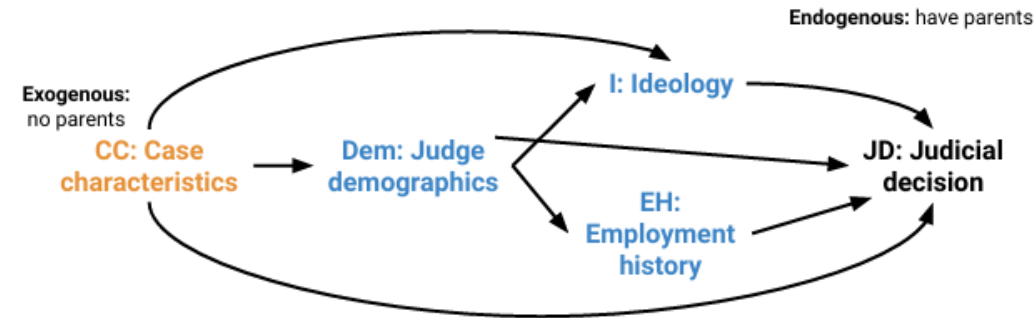
## 10. Causal Graph



$X \rightarrow Y$  means  $Y$  is a function of  $X$

- $X \rightarrow Y$ 의 의미:  $Y$ 가  $X$ 의 함수. 즉,  $Y$ 는  $X$ 에 따라 값이 달라짐
- 위의 예에서, 판사 인구통계가 사건 특성의 영향을 받는 것처럼 보일 수 있음 → 실제로는 특정 사건이 특정 판사에게 자주 배정되기 때문임
  - 사건 특성 → 판사 인구통계 → 특정 사건 유형이 특정 판사에게 배정되는 경향
  - 판사 인구통계 → 이념 / 경력 → 사법 결정 → 판사의 배경이 이념 형성과 경력, 그리고 판결에 영향
  - 사건 특성 → 이념 / 사법 결정 → 사건 자체의 성격이 판사의 판단에 직접적 영향 줄 수 있음

## 10. Causal Graph

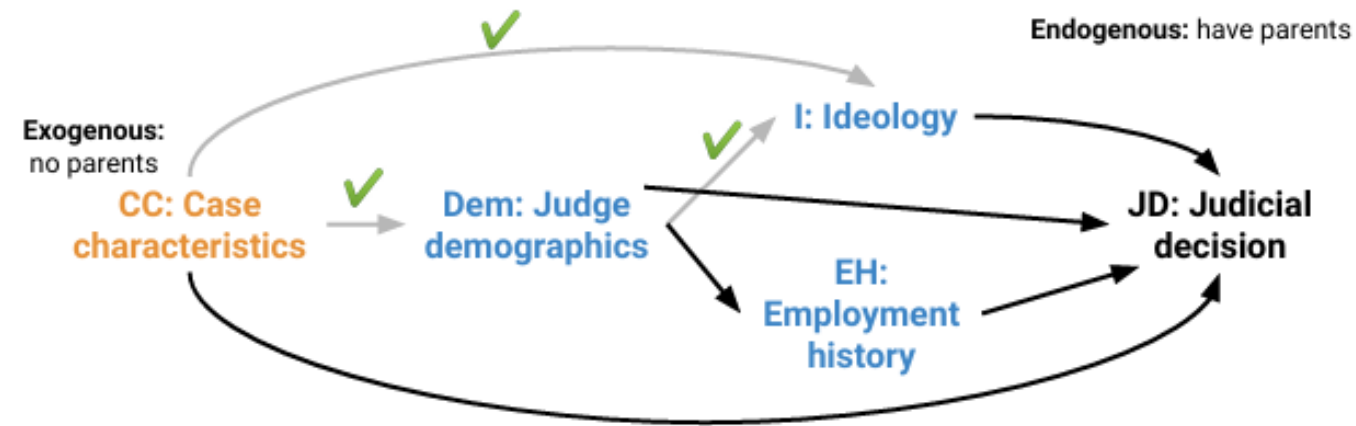


$$P(CC, Dem, I, EH, JD) = P(CC) P(Dem | CC)$$

### 인과 그래프 기반 생성적 접근

- 결합분포는 시간 순서에 따라 앞선 변수부터 시작, 이후 변수들을 생성하는 방식으로 표현할 수 있음
  - Case characteristics(CC)는 부모 노드 없음 → exogenous variable로 간주 → 시스템 외부에서 주어진 변수
  - 전체 분포는  $P(CC)$  부터 시작해서 순차적으로 구성 가능
  - $P(CC, Dem, I, EH, JD) = P(CC)$  로 시작
- 구조 무시하고 표현할 경우:  $P(CC)P(Dem|CC)P(I|Dem, CC)P(EH|I, Dem, CC)P(JD|EH, I, Dem, CC)$ 
  - 그래프 구조를 반영하면 조건부 독립을 활용해 더 간단하게 표현 가능
  - CC가 "원인 없음"이라는 뜻은 아님 → 모델 내에서는 외부 요인 취급함

## 10. Causal Graph



Endogenous V:  
Exogenous V와 다른  
endogenous V에 의해  
결정되는 변수

$$P(CC, Dem, I, EH, JD) =$$

$$P(CC) P(Dem | CC) P(I | Dem, CC) P(EH | Dem)$$

I는 Dem과 CC의 함수인 점을 반영

EH는 Dem의 함수인 점 반영

Ignoring graph structure:

$$P(CC) P(Dem | CC) P(I | Dem, CC) P(EH | I, Dem, CC) P(JD | EH, I, Dem, CC)$$

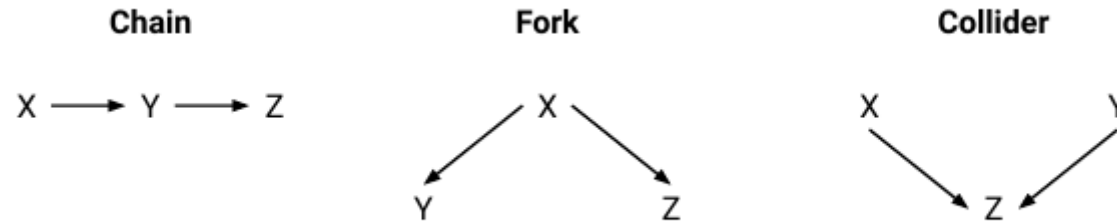
## 10. Causal Graph

- 조건부 확률의 식: 결합분포 표현을 위해 필요
  - 지금까지는 확률 구조만 정의함
  - 각 조건부 확률이 어떤 form인지도 정의해야 함

### Example

- 사건 난이도(CC)의 확률질량함수(PMF):
  - $P(CC = \text{hard}) = 0.9$
  - $P(CC = \text{mild}) = 0.08$
  - $P(CC = \text{easy}) = 0.02$
- 고용이력(EH)은 Dem에 따라 선형 회귀로 표현 가능:
  - $E[EH|Dem] = \beta_0 + \beta_1 \cdot \text{DemCategory2} + \beta_2 \cdot \text{DemCategory3}$
  - 나머지 변수들도 같은 방식으로 각 부모 변수에 대해 종속된 함수로 표현 가능
- 구조방정식모형 (Structural Equation Model)
  - 각 변수가 부모 변수에 따라 어떻게 결정되는지를 나타내는 함수 집합
  - 인과 그래프만으로는 결합분포 완전 정의 불가능
  - 인과 그래프 + 구조방정식모형이 결합되어야 전체 분포 명확히 명시 가능

## 10. Causal Graph

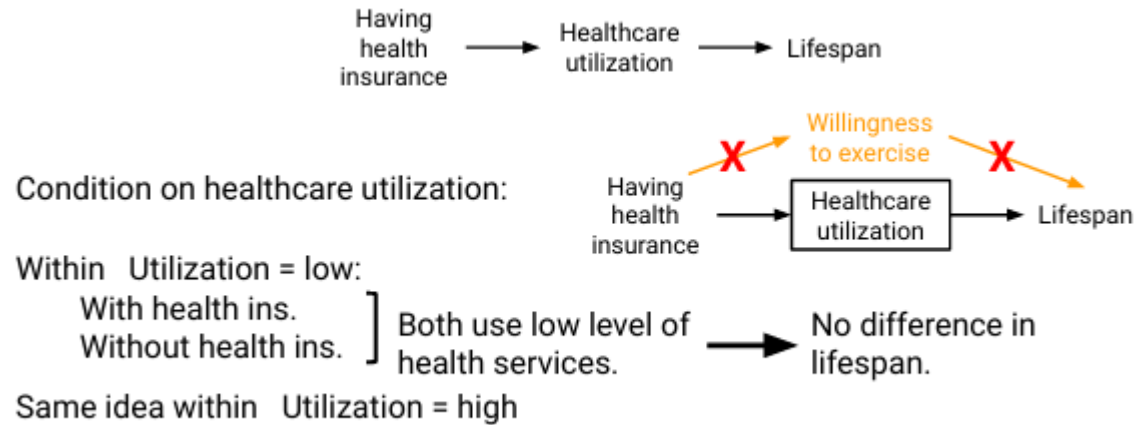


- 모든 인과 그래프(DAG)는 아래 3가지 구조로 구성됨:
- **체인 (Chain):**  $X \rightarrow Y \rightarrow Z$ 
  - 한 변수가 다른 변수를 거쳐 또 다른 변수에 영향 줌
  - $X$ 가  $Y$ 를,  $Y$ 가  $Z$ 를 원인으로 함
- **포크 (Fork):**  $X \rightarrow Y$  그리고  $X \rightarrow Z$ 
  - 한 변수가 두 변수의 **공통 원인**
  - $Y$ 와  $Z$ 는  $X$ 를 통해 연결됨
- **콜라이더 (Collider):**  $X \rightarrow Z \leftarrow Y$ 
  - 두 변수가 하나의 변수에 화살표로 모이는 구조
  - $Z$ 는  $X$ 와  $Y$ 의 **공통 결과(효과)**
- 이 세 가지 구조가 인과 그래프의 기본 구성요소 → 조건부 독립성과 의존성의 패턴도 이 구조들로부터 유도됨

## 10. Causal Graph

- 체인 구조:  $X \rightarrow Y \rightarrow Z$ 
  - 기본 의존 관계
    - X와 Y는 의존적: Y는 X의 값에 따라 달라짐
    - Y와 Z도 의존적: Z는 Y의 값에 따라 달라짐
    - X와 Z도 의존적: X가 Y에, Y가 Z에 영향을 주므로 X-Z도 연결됨
  - 조건부 독립성 (Conditional Independence)
    - Y를 알고 있으면, X와 Z는 조건부 독립
    - Y를 고정하면 X의 변화가 Z에 더 이상 영향 안 줌

## 10. Causal Graph



- 예시: 건강보험 → 의료 이용 → 수명
- 건강보험 있음 → 의료 서비스 이용 증가
- 의료 이용 증가 → 수명 증가
- 건강보험 있음 → 수명 증가

- 조건부 독립 예시
  - "의료 이용"을 고정했을 경우
  - 의료 이용 수준이 동일한 집단 내
    - 건강보험 유무와 수명 간 차이 없음
    - 보험 유무와 무관하게 동일한 수준의 치료 받기 때문
  - 의료 이용 수준 고정 시 보험 여부 ⊥ 수명 | 의료 이용

- 건강보험이 오직 의료 이용을 통해서만 수명에 영향 준다고 가정
- 만약 "운동 의지" 같은 다른 경로가 있다면 의료 이용을 고정해도 보험과 수명은 조건부 독립 아님

## 10. Causal Graph

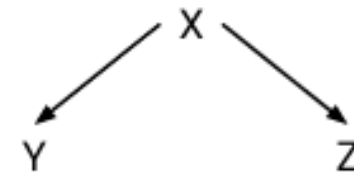
- 체인에 있는 어떤 두 변수들은 주변적으로 의존적(marginally dependent)
- 두 변수 사이에 위치한 변수에 대해 조건부 → 조건부 독립(Conditionally Independent)
- 체인은 처치(Treatment)에서 결과(Outcome)로 이어지는 인과 경로를 의미함
  - 중간 경로를 통해 원인이 결과에 영향을 미침, 인과 메커니즘을 나타냄
- 예시:
  - $A \rightarrow X \rightarrow B \rightarrow C \rightarrow Y$
  - X와 Y는 다음 조건 중 하나만 만족해도 조건부 독립:
    - $Z = \{B\}$
    - $Z = \{C\}$
    - $Z = \{B, C\}$
  - → 즉, X와 Y 사이에 위치한 변수 중 하나라도 알고 있으면, X와 Y는 조건부 독립



## 10. Causal Graph

- Fork 구조 (Confounder의 역할)

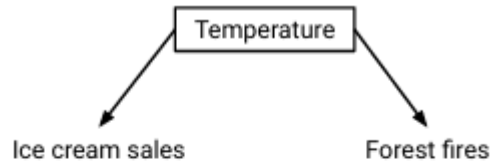
- X와 Y는 의존적 : Y는 X에 의해 결정됨
- X와 Z도 의존적 : Z도 X에 의해 결정됨
- Y와 Z도 의존적 : Y와 Z는 각각 X의 영향을 받기 때문에, 변화가 함께 나타날 수 있음
- 조건부 독립성: Y와 Z는 X가 주어졌을 때 조건부 독립, X의 값을 고정하면, Y와 Z는 더 이상 연결되지 않음,  $Y \perp Z | X$
- X가 원인이라서 Y, Z가 동시에 반응하는 것처럼 보이지만 X 값을 고정하면 Y와 Z는 단지 각자 노이즈일 뿐, 연결되어 있지 않음



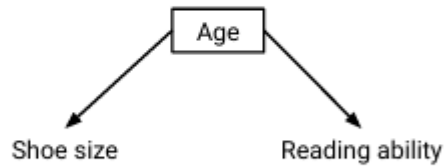
X is called a **common cause** of Y and Z.

**혼란(confounding)** 상황에서 매우 중요하며, 공통 원인(X)을 통제해야 정확한 인과 해석 가능

## 10. Causal Graph



At a fixed temperature, no longer a relationship between ice cream sales and forest fires.



At a fixed age, no longer a relationship between shoe size and reading ability.

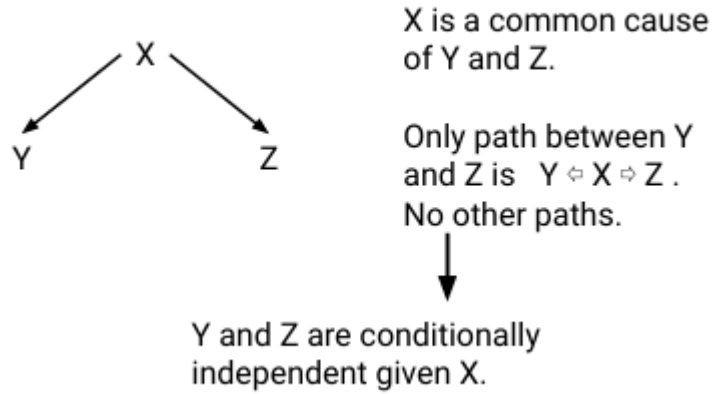


Forks involving treatment and outcome create lack of exchangeability.

- 예시 1:
  - Temperature → Ice cream sales
  - Temperature → Forest fires
  - 온도는 아이스크림 판매량과 산불 발생의 공통 원인
  - 온도가 변하면 두 변수 모두 변하므로 서로 상관 있어 보임
  - 하지만 온도를 고정하면 → 아이스크림과 산불은 조건부 독립
- 예시 2:
  - Age → Shoe size
  - Age → Reading ability
  - 나이는 신발 사이즈와 독해 능력의 공통 원인
  - 나이를 고정하면 → 신발 사이즈와 독해 능력 간 상관관계 사라짐

- 포크 구조는 "상관 ≠ 인과"를 보여주며, 중간 노드(X)는 겉보기 상관관계를 만들어내는 구조
- 포크 구조는 confounding의 근본 구조, 특히 포크 구조가 처치와 결과(treatment & outcome) 사이에 있을 경우 → 교환성(Exchangeability) 결여를 초래
  - 교환성: treated group, control group이 통계적으로 동일한 조건에서 비교될 수 있는 상태
  - 교환성 결여: 처치 여부 외 결과에 영향을 미치는 다른 요인(confounder)이 있어 두 집단이 비교 불가능

## 10. Causal Graph



- 조건부 독립
  - X를 조건화하면, Y와 Z는 조건부 독립
  - $Y \perp Z | X$ 
    - X를 고정하면 Y와 Z는 더 이상 연관되지 않음

- 포크(Fork) 구조
  - X는 Y와 Z의 공통 원인(common cause)
  - Y와 Z 사이의 경로는 오직:  $Y \leftarrow X \rightarrow Z$
  - 다른 경로 없음
- 교환성 확보
- 포크의 중간 노드(X)를 조건화
  - 조건부 교환성(Conditional Exchangeability) 확보 가능
  - 혼란(confounding) 제거의 핵심 기법임

## 10. Causal Graph



- Collider
  - Z는 두 화살표가 충돌(collision) 하는 지점
  - X와 Y는 독립: 서로 원인이 아니고, 공통 원인도 없음
  - X와 Z는 의존적: X는 Z에 영향 줌
  - Y와 Z도 의존적: Y도 Z에 영향 줌
- Collider는 조건부 의존성을 만들어냄
  - 조건화하면 오히려 의존이 생김 → confounding과 정반대
  - 따라서 Collider는 조건화하지 않아야 하며, 불필요한 조건화는 거짓된 인과 생성 가능
- 조건부 의존성 (Collider의 핵심)
  - Z를 조건화하면, X와 Y는 조건부 의존:  $X \perp Y | Z$
  - Z의 값을 고정하면, X가 바뀔 때 Y도 그에 맞게 바뀌어야 Z가 유지됨
- 예
  - X: 첫 번째 주사위 눈
  - Y: 두 번째 주사위 눈
  - Z: 두 눈의 합
  - 평소에는 X와 Y는 독립, 하지만 "합이 8이 나왔다"는 정보를 알게 되면, X와 Y는 서로 얽힘

## 10. Causal Graph

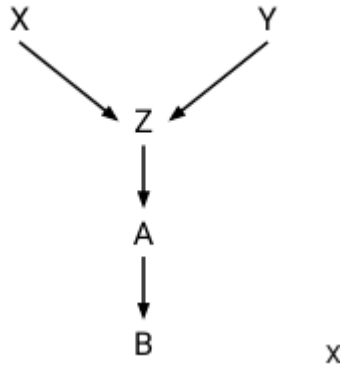


- 예시
  - 음악 재능과 보훈 자녀 여부는 일반적으로 서로 독립임
  - 둘 다 장학금(Scholarship) 수여에 영향을 줌 → Z는 Collider
  - 조건화하지 않았을 때는 독립
  - 콜라이더에 조건화하면 → 두 원인이 조건부로 의존
    - 콜라이더를 조건화하면 거짓된 상관관계 발생 가능

- 조건부 의존성
  - 장학금을 받았다는 조건(Scholarship = "yes")을 고정:
    - 음악 재능이 없다고 가정할 경우
    - 장학금을 유지하려면 → 보훈 자녀여야 함
    - 둘 중 하나가 바뀌면 다른 하나도 바뀌어야 조건(Z)이 유지됨

콜라이더에 조건을 걸면, 원래 독립이었던 두 변수가 의존적으로 바뀌기에, Collider 조건화는 주의해야 함

## 10. Causal Graph



- 주요 특성
- Z는 X와 Y의 공통 결과 (common effect)
- X와 Y 사이 경로는 오직  $X \rightarrow Z \leftarrow Y$ 
  - X와 Y는 주변적으로 독립(marginally independent)
- 조건부 의존성 발생
  - Z에 조건화하면 X와 Y는 조건부 의존(conditionally dependent)
  - Z의 자손(A, B 등)에 조건화해도 X와 Y는 조건부 의존됨
  - $X \perp Y | Z, A, B$ , 또는 그 조합
- 콜라이더 또는 그 자손에 조건화하면 교환성 결여
  - 불필요한 상관관계 발생, 잘못된 인과 해석 초래 가능
- 콜라이더 노드를 조건화하면 안 됨, 교환성 깨짐

## 10. Causal Graph

$$\text{Observed data associations} = \text{Causal associations} + \text{Noncausal associations}$$

Ensure that these are removed / not present

- Causal: 인과 메커니즘에 의해 생긴 상관
- Noncausal: 혼란(confounding), collider 조건화 등으로 생긴 잘못된 상관

- 인과 그래프는 통계적 상관의 원인 분석 도구로 사용됨
- 비인과 상관(noncausal association)을 제거 또는 통제해야 인과 효과를 올바르게 추론할 수 있음
- 인과 추론에서는 "관측된 상관관계 = 인과관계"가 아님

## 10. Causal Graph

### Chain

$X \longrightarrow Y \longrightarrow Z$

### Fork

$X \longleftarrow Y \longrightarrow Z$

### Collider

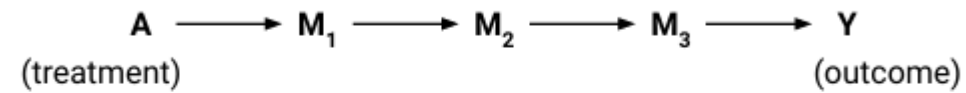
$X \longrightarrow Y \longleftarrow Z$

- Path: 상관의 흐름
  - Chain, Fork, Collider 구조는 상관관계를 전달할 수 있는 경로 (path) 역할을 함
  - 이 경로를 통해 association(상관)이 흐름
  - 어떤 노드에 조건화하느냐에 따라 이 흐름이 막히거나 열림
- 구조별 흐름 요약
  - Chain ( $X \rightarrow Y \rightarrow Z$ )
    - X와 Z 사이에 상관 있음
    - Y에 조건화하면 흐름이 차단됨 (block)  $X \perp Z \mid Y$
  - Fork ( $X \leftarrow Y \rightarrow Z$ )
    - X와 Z 사이에 상관 있음 (공통 원인 Y 때문)
    - Y에 조건화하면 흐름이 차단됨 (block)  $X \perp Z \mid Y$
  - Collider ( $X \rightarrow Y \leftarrow Z$ )
    - X와 Z는 원래 독립
    - Y에 조건화하면 흐름이 열림 (unblock/open)  $X \not\perp Z \mid Y$
- Blocking과 Unblocking
  - 경로를 차단(block) = 독립으로 만듦
  - 경로를 여는 것(unblock) = 상관을 발생시킴
  - 조건화를 통해 상관 흐름을 제어함 → DAG 해석의 핵심

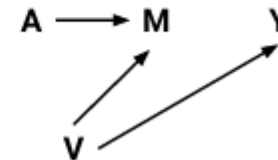
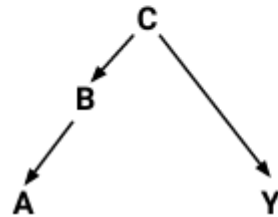


## 10. Causal Graph

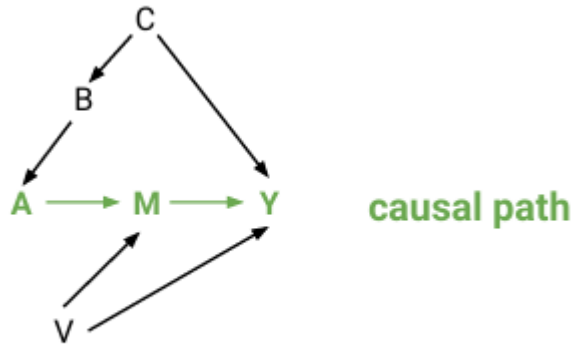
**Causal path:** *directed* path from treatment to outcome



**Noncausal path:** any path from treatment to outcome that is not causal (necessarily an *undirected* path)

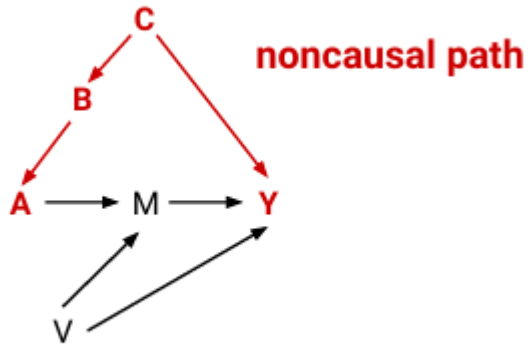


## 10. Causal Graph

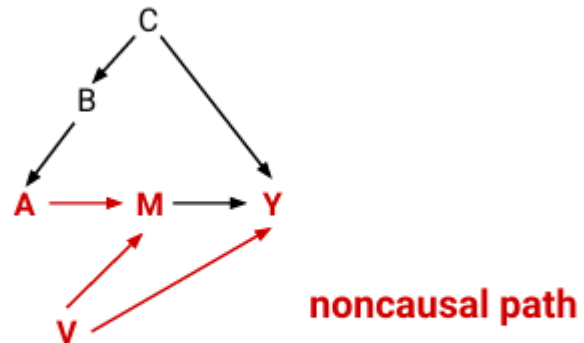


- 현실 데이터의 문제: 인과 경로 vs 비인과 경로
- 실제 데이터에서는 인과 경로와 비인과 경로에서 발생한 상관이 섞여 있어서, 구분 (disentangle) 필요
- 예시:
  - 인과 그래프에서  $A \rightarrow M \rightarrow Y$  경로는 진짜 인과 경로(causal path)
  - 이 경로를 통해 처치(A)가 결과(Y)에 인과적 영향을 주는 것을 발견
  - 관심 있는 인과 효과를 제대로 보기 위해서는 비인과 경로를 차단하고, 인과 경로만 추적해야 함

## 10. Causal Graph



- 인과 그래프 내 비인과 경로
  - 그래프에는 두 가지 경로가 존재함:
  - 인과 경로:  $A \rightarrow M \rightarrow Y$
  - 비인과 경로(noncausal path):  $A \leftarrow B \leftarrow C \rightarrow Y$ 
    - 체인:  $A \leftarrow B \leftarrow C$
    - 포크:  $B \leftarrow C \rightarrow Y$
- A와 Y 사이의 연관성은 단순한 인과 효과가 아니며, 비인과 경로에서도 상관이 발생할 수 있음
- 이런 경로를 차단하지 않으면, A의 인과 효과를 잘못 해석



- $A \rightarrow M \leftarrow V \rightarrow Y$ 
  - 단순한 체인·포크보다 더 복잡한 혼합 구조임
  - 구조적 특징:
    - Collider:  $A \rightarrow M \leftarrow V$
    - Fork:  $M \leftarrow V \rightarrow Y$
  - A와 Y 사이에 잘못된 상관관계(noncausal association)를 만들어 낼 수 있음
  - 인과 효과와 섞이면 해석이 왜곡

# 10. Causal Graph

구조 유형	차단(Blocking) 방법	예시 경로	조건화 결과
Chain	중간 노드 조건화	$A \rightarrow B \rightarrow C$	$A \perp C \mid B$
Fork	공통 원인 조건화	$A \leftarrow B \rightarrow C$	$A \perp C \mid B$
Collider	조건화 금지! (조건화 시 오히려 상관 생김)	$A \rightarrow B \leftarrow C$	$A \not\perp C \mid B$ (의존 생김)



## Industrial Data Science Lab

Contact:

won.sang.l@gwnu.ac.kr

<https://sites.google.com/view/idslab>