# Causal Inference for Health Data (Winter 2026)
# STATS C160/C260
# Homework 3

**Exercise 1. Counterfactual Quantities**

Consider an SCM $\mathcal{M}$ below.

$$\mathcal{M}^* = \begin{cases} \mathbf{U} & = \{U_1, U_2, U_3\}, \text{ all binary} \\ \mathbf{V} & = \{X, W, Y\} \\ \mathcal{F} & = \begin{cases} f_X(u_1) & = u_1 \\ f_W(x, u_1, u_2) & = (u_1 \wedge u_2) \oplus (\neg u_2 \wedge \neg x) \\ f_Y(x, w, u_3) & = (x \wedge w) \oplus u_3 \end{cases} \\ P(\mathbf{U}) & \text{defined such that } P(U_1 = 0) = 0.4, P(U_2 = 0) = 0.2, P(U_3 = 0) = 0.6, \end{cases}$$

We let $X$ be the treatment (whose effect we seek to assess), $Y$ be the outcome variable and $W$ for all the observed intermediate variables between X and Y (called mediators);

(a) Calculate the total variation $P(Y = 1 \mid X = 1)$ and the total effect $P(Y = 1 \mid do(X = 1))$.

(b) Calculate $ETT_{x,x'}(y)$ where $y = 1$, $x = 1$ and $x' = 0$.

(c) Calculate $PN/PS_{(x,y)(x',y')}(X; Y)$ where $x = 1, y = 0$ and $x' = 0, y' = 1$.

(d) Calculate the $NDE_{x_0,x_1}(y)$ and $NIE_{x_0,x_1}(y)$ where $x_0 = 0$, $x_1 = 1$, and $y = 1$.

(e) Calculate the counterfactual quantity $DE_{x_0,x_1}(y \mid x) := P(y_{x_1,W_{x_0}} \mid x) - P(y_{x_0}|x)$ where $x_0 = 0, x_1 = 1, x = 1, y = 1$.

**Exercise 2. Counterfactual Evaluation**

The team of data scientists is tasked with evaluating the psychological effects of a new health program. Consider the variable $X$, which represents whether the individual signed up for the gym membership in month 1 ($X = 1$ represents sign-up), $Z$ is the individual's body mass index (BMI) after month 6 ($Z = 1$ represents a healthy BMI), and $Y$ whether the individual ranks above 5 on a mood-scale after month 9 ($Y = 1$ represents a positive mood).

The true, underlying SCM is given as follows: Consider an SCM $\mathcal{M}$ below.

$$\mathcal{M}^* = \begin{cases} \mathbf{U} & = \{U_1, U_2, U_3, U_4, U_5\}, \text{ all binary} \\ \mathbf{V} & = \{X, Z, Y\} \\ \mathcal{F} & = \begin{cases} X & \leftarrow u_1 \wedge u_2 \\ Z & \leftarrow (u_1 \oplus u_3) \wedge ((x \vee u_2) \oplus u_4) \\ Y & \leftarrow (z \vee (\neg u_2) \vee (\neg x)) \oplus u_5 \end{cases} \\ P(\mathbf{U}) & \text{defined such that } P(U_1 = 1) = 0.5, \\ & P(U_2 = 1) = 0.5, P(U_3 = 1) = 0.7, P(U_4 = 1) = 0.4, P(U_5 = 1) = 0.2 \end{cases}$$

(a) First, write the counterfactual quantity that given an individual who signed up for gym membership, this same individual would not have reported a positive mood had they not reported a healthy BMI. Second, evaluate this quantity directly from $\mathcal{M}$.

(b) The team is asked about the necessity and sufficiency of $X$ regarding $Z$.

(i) Write and evaluate the quantity that describes how much the absence of $X$ is necessary to make $Z = 0$ given an individual who signed up for gym membership and reported a healthy BMI.

(ii) Write and evaluate the quantity that describes how much the presence of $X$ is sufficient to make $Z = 1$ given an individual who did not sign up for gym membership and reported an unhealthy BMI.

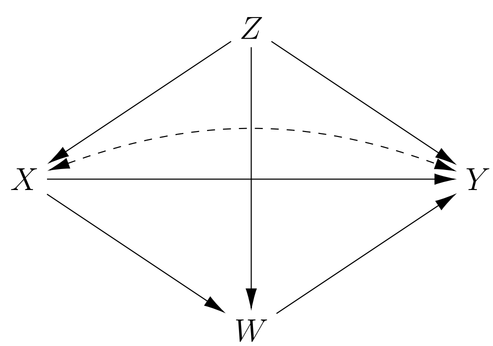
(c) The team is further asked how sign-up for the gym, directly and indirectly, affects the mood of the population ($Y$).

(i) Write in counterfactual notation the full expression for the natural direct effect, written $\text{NDE}_{X=0,X=1}(Y = 1)$, and evaluate it directly from $\mathcal{M}$.

(ii) Write in counterfactual notation the full expression for the natural indirect effect, written $\text{NIE}_{X=0,X=1}(Y = 1)$, and evaluate it directly from $\mathcal{M}$.

## Exercise 3. Network Construction

Consider the following graph.




(a) What is $An(Y_{xw}, W_{x'}, Z)$?

(b) Draw the ancestral multi-world network to check $Y_{xw} \perp W_{x'} \mid Z$.

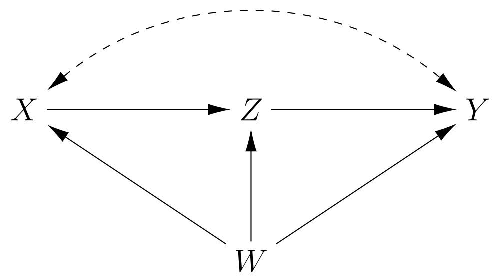
(c) Let's consider deriving the target query $P(y_{x,W_{x'}})$ given $P(\mathbf{V})$, $P(\mathbf{V} \mid do(x))$ and $P(\mathbf{V} \mid do(x'))$ and the causal diagram.

(i) Unnest the target query $P(y_{x,W_{x'}})$.

(ii) [Optional] Derive an expression for the target query $P(y_{x,W_{x'}})$. Suppose you have acces to both observational data $P(\mathbf{V})$, and also interventional data $P(\mathbf{V} \mid do(x))$, $P(\mathbf{V} \mid do(x'))$. Can the causal query be computed? *Hint: Use the results from (a)-(b).*

### Exercise 4. Counterfactual Identification

Consider the following graph.




(a) Compute the following counterfactual ancestors:

   (a) $An(Y_z)$
   (b) $An(Z_x)$
   (c) $An(W)$
   (d) $An(X)$

(b) Compute the ancestral multi-world network to evaluate the following counterfactual independencies:

   (a) $Y_z \perp Z_x \mid W, X$;
   (b) $Z_x \perp X \mid W$;
   (c) $Y_x \perp X \mid W$.

(c) Using consistency, exclusion, and independence rules, decide whether the query $P(y_x \mid x')$ is identifiable from $P(\mathbf{V})$ and the causal diagram above. If identifiable, provide the derivation step by step.

### Exercise 5. Counterfactual

(a) Consider the following SCM $M = \langle \mathbf{V}, \mathbf{U}, \mathcal{F}, P(\mathbf{u}) \rangle$:

$$\mathbf{V} = \{X, Y, Z, W\}$$
$$\mathbf{U} = \{U_1, U_2, U_3\}$$
$$\mathcal{F} = \begin{cases} f_Z = U_1 \oplus U_2, \\ f_W = Z \oplus U_3, \\ f_X = U_1 \oplus W, \\ f_Y = X \oplus U_2, \end{cases}$$
$$P(\mathbf{u}) = \left\{ P(U_1 = 1) = \frac{1}{2}, P(U_2 = 1) = \frac{1}{10}, P(U_3 = 1) = \frac{1}{4} \right\}$$

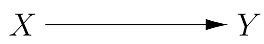
Compute following counterfactual queries using the three-step algorithm consisting of abduction, action and prediction.

(i) $P(Y_{x=0} = 1)$.

(ii) $P(Y_{x=0} = 1 | X = 1)$.

(iii) $P(Y_{x=0} = 1 | X = 1, W = 1, Z = 1)$.

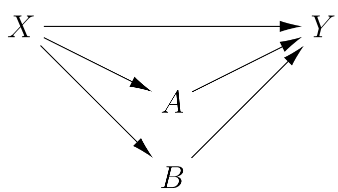
(b) Consider the following causal diagram




Is the distribution $P(y_x, y_{x'})$ for $x \neq x'$ identified from $P(x, y)$ and $P(y|do(x))$? If yes, provide a derivation. Otherwise, provide a counterexample.

**Exercise 6. Direct and Indirect Effects**

In this problem, we will investigate how counterfactual queries can allow more fine-tuned explanations of cause and effect. Consider the following graph $\mathcal{G}$ and observational data $P(\mathbf{V})$.




(a) Graph $\mathcal{G}$

| X | A | B | Y | P |
|---|---|---|---|---|
| 0 | 0 | 0 | 0 | $p_0 = 0.324$ |
| 0 | 0 | 0 | 1 | $p_1 = 0.036$ |
| 0 | 0 | 1 | 0 | $p_2 = 0.02$ |
| 0 | 0 | 1 | 1 | $p_3 = 0.02$ |
| 0 | 1 | 0 | 0 | $p_4 = 0.045$ |
| 0 | 1 | 0 | 1 | $p_5 = 0.045$ |
| 0 | 1 | 1 | 0 | $p_6 = 0.005$ |
| 0 | 1 | 1 | 1 | $p_7 = 0.005$ |
| 1 | 0 | 0 | 0 | $p_8 = 0.005$ |
| 1 | 0 | 0 | 1 | $p_9 = 0.005$ |
| 1 | 0 | 1 | 0 | $p_{10} = 0.009$ |
| 1 | 0 | 1 | 1 | $p_{11} = 0.081$ |
| 1 | 1 | 0 | 0 | $p_{12} = 0.004$ |
| 1 | 1 | 0 | 1 | $p_{13} = 0.036$ |
| 1 | 1 | 1 | 0 | $p_{14} = 0.036$ |
| 1 | 1 | 1 | 1 | $p_{15} = 0.324$ |

(b) Observational Data $P(\mathbf{V})$. Probability of $i$th row is labeled $p_i$.

(a) Identify the following expressions from observational data $P(\mathbf{V})$ and the graph $\mathcal{G}$.

(i) $P(Y_{X=x} = y)$

(ii) $P(Y_{X=x, A_{X=x'}, B_{X=x'}} = y)$

(b) Using the results from part (a), compute the following queries using the table in Figure (b).

(i) Average Treatment Effect (ATE): $P(Y_{X=1} = 1) - P(Y_{X=0} = 1)$

(ii) Natural Direct Effect (NDE): $P(Y_{X=1, A_{X=0}, B_{X=0}} = 1) - P(Y_{X=0} = 1)$

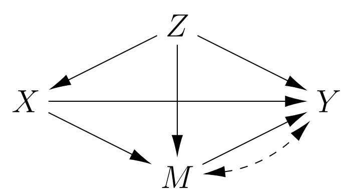
(iii) Natural Indirect Effect (NIE): $P(Y_{X=0, A_{X=1}, B_{X=1}} = 1) - P(Y_{X=0} = 1)$

(c) Query (b)(ii) can be interpreted as specifically the effect of intervening $X = 1$ on $Y$ through the direct path $X \to Y$, ignoring the indirect paths through $A$ and $B$. On the other hand, query (b)(iii) is the opposite: the effect of intervening $X = 1$ on $Y$ through the indirect paths through $A$ and $B$ and ignoring the direct path. Query (b)(i) aggregates the causal effect of $X = 1$ on $Y$ through all paths [1]. What query should be calculated if the goal is to get the causal effect of $X = 1$ on $Y$ specifically through the path $X \to A \to Y$, ignoring both the direct path $X \to Y$ and the indirect path $X \to B \to Y$?

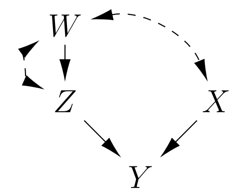
(d) Identify the terms of the query in part (c) and calculate its value using $P(\mathbf{V})$ from Table (b).

**Exercise 7. [Optional] Counterfactual Identification**

Consider the following graphs. Assume all variables have finite discrete domains (but not necessarily binary).




(a)




(b)

(a) Is $P(Y_{X=x_1} = y_1 \mid X = x_0)$ identifiable from $P(\mathbf{V})$ and the graph $\mathcal{G}$ in Figure (a)? If yes, then provide the identification expression (and show your work). If no, then provide a counterexample.

(b) Are the following queries identifiable from $P(\mathbf{V})$ and $P(X, Z, Y \mid do(W))$ and the graph $\mathcal{G}$ in Figure (b)? If yes, then provide the identification expression (and show your work). If no, then provide a counterexample.

(i) $P(Y_{X=x_0, W=w_0} = y_0, X_{W=w_1} = x_1)$

(ii) $P(Y_{X=x_0} = y_0, Z_{W=w_0} = z_0)$

[1] Despite this observation, it will typically not be the case that ATE = NDE + NIE (i.e., the sum of effects on all paths) because the causes are not necessarily disjoint.