Problem 1. Modeling [5 points]

Consider the recent study of the connection between sleep quality and dementia presented here and discussed in class. 

(a) [2 points] Provide a structural causal diagram based on your understanding of the study and assumptions made.

(b) [1 point] Discuss the suitability of the different conclusions proposed by the study. You can focus on one or two statements.

(c) [2 points] The article talks about associations rather than causation. In your opinion, can they claim causation and under what assumptions.

Problem 2. Understanding the Model’s Granularity [10 points]

Consider the causal diagram $G$ below.

(a) [1 point] Determine whether the causal effect $P(y | do(x))$ is identifiable from $G$ and $P(V)$, where $V$ is the set of endogenous variables. If so, show how; otherwise, provide a counter-example.

(b) [2 points] Write an SCM $M$ that induces $G$ and a probability distribution $P(V)$, with $P(v) > 0$ for every $v$. You don’t need to show $P(V)$ in your answer.

Suppose that the same system (represented by the SCM) is investigated in another study. However, in this case, only the variables $V' = \{X, Y, B, C\}$ are measured.

(c) [3 points] Write a new SCM $M' = \langle V', U', F', P(u') \rangle$ corresponding to this different cut of reality, consistent with your answer to the previous question (i.e., departing from SCM $M$ written in (b)).

(d) [1 point] Draw the causal diagram $G'$ induced by $M'$.

(e) [3 points] Is the effect $P(y | do(x))$ identifiable from $P(V')$ and $G'$? Is there a back-door or front-door adjustment? Can it be solved with do-calculus?

Problem 3. Optimal Experiment Design [10 points]

An advertisement company is trying to identify the effect of a new campaign $X$ on the click through rate $Y$. They
have two hypotheses about how the strategy relates to a possibly measured set of covariates $Z$. The hypotheses are represented in the causal diagrams (a) and (b) shown below:

(a) [4 points] If it exists, find a minimal admissible set for adjustment in each of the graphs.

(b) [6 points] The company wants to minimize the measurement cost for identifying $P(y \mid do(x))$. Find the minimum cost ID expression based on the table (c) and justify your answer.

Problem 4. (*)[1 point]
Back-door Adjustment as a Substitute for the Direct Parents The causal effect of the intervention $do(X = x)$ on a variable $Y$ can be identified if all parents of $X$ are observed and is given by

$$P(y \mid do(x)) = \sum_{pa_X} P(y \mid x, pa_X) P(pa_X). \quad (1)$$

Based on this result, prove that if a set $Z$ satisfies the back-door criterion relative to $X$ and $Y$ in the graph $G$, it follows that

$$P(y \mid do(x)) = \sum_{z} P(y \mid x, z) P(z). \quad (2)$$

This question is asking you to leverage Eq. (1) to prove the backdoor identification formula in Eq. (2).

Problem 5. Many Paths Lead to ID [10 points]
Consider the following causal diagram.

Give three different functions of the observational distribution $P(V)$ that are equal to the effect $P(y \mid do(x))$. At least one answer should correspond to a front-door case and one to a back-door case. Justify each one of the expression showing its do-calculus derivation.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X$</td>
<td>2</td>
</tr>
<tr>
<td>$Y$</td>
<td>1</td>
</tr>
<tr>
<td>$Z_1$</td>
<td>4</td>
</tr>
<tr>
<td>$Z_2$</td>
<td>2</td>
</tr>
<tr>
<td>$Z_3$</td>
<td>4</td>
</tr>
<tr>
<td>$Z_4$</td>
<td>5</td>
</tr>
<tr>
<td>$Z_5$</td>
<td>5</td>
</tr>
<tr>
<td>$Z_6$</td>
<td>2</td>
</tr>
<tr>
<td>$Z_7$</td>
<td>1</td>
</tr>
</tbody>
</table>