Problem Set 03
Directed acyclical graphs(DAGs)

TAs: Aida Hatami, Sam Fathinejad, Fatemeh Salehi, Reza Sahour

February 28, 2025

1 - Causal Analysis of Fast Food Consumption and Heart Disease Risk

Part 1: Drawing the Causal Diagram (DAG) and Selecting Control Variables

1. Draw a Directed Acyclic Graph (DAG) to analyze the causal effect of fast food consumption on heart disease risk.
   - Include key variables such as **fast food consumption, physical activity, diet quality, BMI, blood pressure, age, gender, and genetics**.
   - Use arrows to indicate the causal relationships among these variables.
2. Based on your DAG, determine:
   - Which variables should be **controlled** (confounders)? Explain why.
   - Which variables should **not be controlled** (bad controls such as mediators and colliders)? Explain why.

Part 2: Regression Analysis and Bias Interpretation

Using the provided simulated dataset, conduct the following regression analyses in **Stata or Python** and interpret the results.

1. **Full Specification: Controlling for All Variables** Estimate the following regression model, where all available variables are included as controls:

   ```
   regress heart_disease fast_food exercise healthy_diet age bmi
   blood_pressure genetic_risk gender
   ```

   - Report the estimated coefficient for **fast food consumption** along with its standard error and p-value.
   - What potential issues arise from controlling for all variables?
2. **Correct Specification and Bias Estimation** Based on your DAG from Part 1, run a regression that includes only the appropriate control variables (i.e., the correct set of confounders).

- Write down the correct regression specification.
- Report the estimated coefficient for **fast food consumption** in this correct model.
- Compare the results from the full specification and the correct specification.
- Estimate the approximate bias in the full model by comparing it with the correct model. Discuss the direction and magnitude of the bias and explain why it occurs.

2 - DAG Construction for Causal Analysis

Task:

For this section, you are required to construct Directed Acyclic Graphs (DAGs) for **four** out of the six causal questions listed below. For each selected question, complete the following tasks:

1. Construct a DAG illustrating the potential causal relationships.
2. Identify at least one variable that you believe should be controlled (confounders) and justify your selection.
3. Identify at least one variable that should **not** be controlled (mediators or colliders) and explain why.

Select four of the Following Topics:

1. The effect of attending art classes in childhood (before age 15) on future earnings.
2. The impact of opening a new museum on neighborhood housing prices.
3. The effect of having a degree in fine arts on financial success in the art market.
4. The impact of attending a gifted school on university admission rates.
5. The effect of air pollution on a person's academic performance.
6. The effect of smoking on lung cancer.

3 - Backdoor Path Identification in Causal Diagrams

1. Given the following graph that illustrates the concept of backdoor paths, identify all possible backdoor paths between the variables X and Y.
2. For each backdoor path, explain why it qualifies as a backdoor path and how it could potentially bias the estimated effect between X and Y.
3. According to the back-door criterion, write down the sufficient set. (A sufficient set is a set of nodes whose control blocks all backdoor paths.)

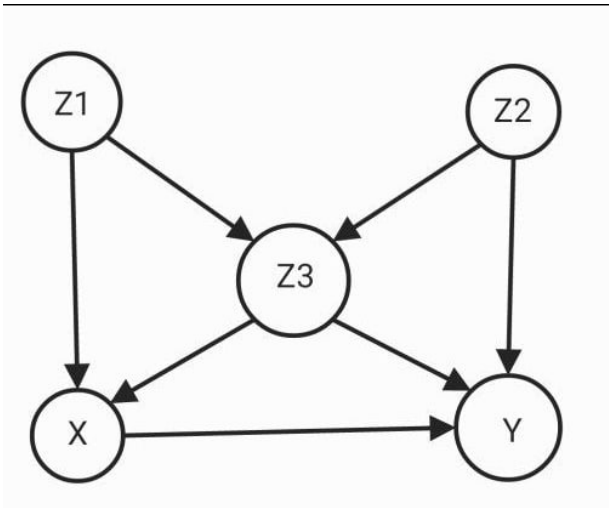

Figure 1: Causal diagram illustrating back-door paths.

Guidance for Your Analysis:

- Be clear in defining your variables and their potential causal relationships.
- For the second part, select topics where you can effectively justify causal relationships and understand how bias may arise.
- Consider how omitted variable bias might influence estimates if key confounders are not controlled.
- Ensure that your explanation of colliders and mediators reflects their role in causal inference.