

FS25 Homework1

For each problem, you should:

- Draw the DAG, clearly indicating the direction of presumed causal relationships. Use the abbreviations for each variable specified in the scenario text when creating the DAGs. You should use <https://www.dagitty.net/> or any related tool to draw the DAGs and identify the variables to include or exclude.
- Identify confounders that need to be controlled for.
- Note any potential colliders or other sources of bias and justify whether or not they should be included in the model based on the causal effect being estimated (direct or total).
- Create a simulation that demonstrates the posterior distribution of the coefficient of interest under a correct version of the model for the desired estimate versus an incorrect version (e.g., total vs direct causal effect). Include both the code (any programming language is fine) and a graph of the two distributions. **Set all effect coefficients to be 1 in your simulations.**

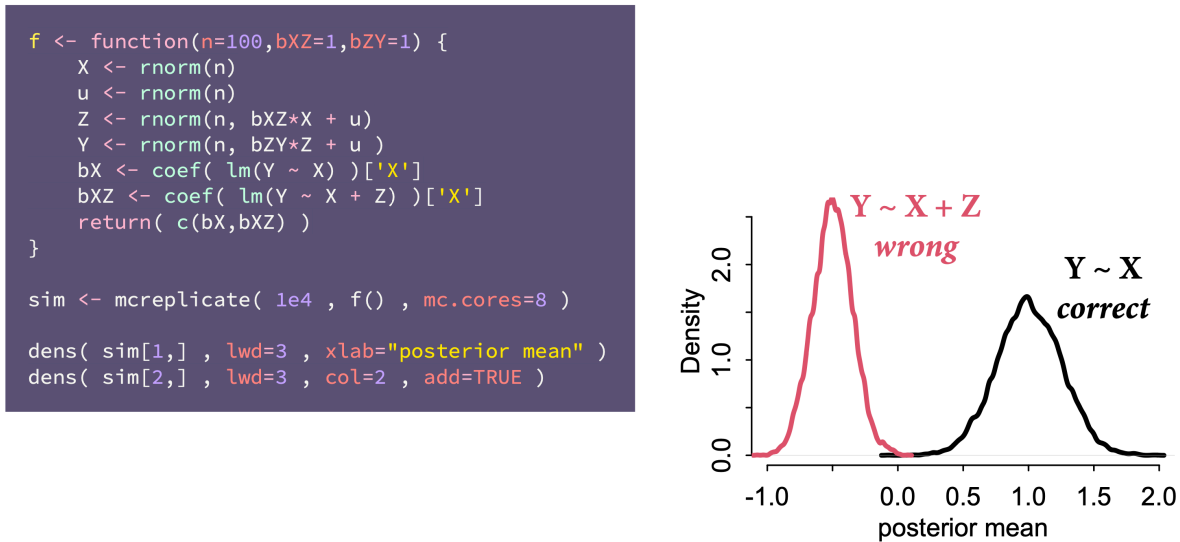


Figure 1: This is an example simulation and plot. You don't need to include the labels of the correct and incorrect distribution in the plot itself. Listing which is which in the figure caption is fine.

- Submit your work for all six problems in a single PDF or html file.

Problem 1: The Effect of Exercise on Weight Loss

Scenario: Researchers are interested in studying the direct effect of regular exercise (E) on weight loss (WL). They consider that diet (caloric intake; CI) might also affect weight loss and is influenced by exercise habits. Furthermore, they acknowledge that motivation (M) could influence both exercise habits and dietary choices.

Task: Draw a DAG for this scenario and identify the variables to include in the statistical model to estimate the **direct effect** of exercise on weight loss.

Problem 2: Education Level and Job Satisfaction

Scenario: A sociological study aims to investigate the direct effect of education level (EL) on job satisfaction (JS). The researchers think that the field of study (F) might influence job satisfaction and is determined by education level. They also think cultural environments (CE) could affect both education level and job satisfaction.

Task: Create a DAG representing the scenario. Decide which variables should be controlled for to estimate the **total effect** of education level on job satisfaction.

Problem 3: Medication Use and Patient Recovery Rate

Scenario: In a clinical study, the aim is to examine the effect of a new medication (M) on the recovery rate (RR) of patients with a specific illness. The researchers suspect that the severity of the illness (SI) influences both the likelihood of being prescribed the medication and the recovery rate. Additionally, they believe that age (A) might influence recovery rate and the severity of the illness.

Task: Sketch a DAG for this scenario. Identify the variables that should be included in the analysis to assess the **direct effect** of medication on recovery rate. Part 2: Show what happens if you control for age in the model. What changes and why?

Problem 4: Advertising and Sales

Scenario: A group of analysts wants to evaluate the effect of advertising levels (A) on sales (S). They consider that the current economic climate (EC) might influence both the budget (B) firms allocate to advertising and impact sales directly (customers buy more/less in good/bad economic climates). Additionally, they think that the quality (Q) of the product might affect sales (S) and also determine how much companies advertise it (A).

Task: Draw a DAG for this scenario. Determine which variables need to be controlled for to estimate the **total effect** of advertising on sales.

Problem 5: Social Media Use and Mental Health

Scenario: Psychologists are studying the relationship between social media use (SM) and mental health outcomes (MH). They hypothesize that the number of hours spent on social media platforms could directly affect mental health. They also consider that having a strong offline social network (OSN) might influence both social media use and mental health. Furthermore, they think that personal interests (PI) might lead to different patterns of social media use.

Task: Construct a DAG for this setup. Decide which variables should be controlled for and which should not, to accurately estimate the **direct effect** of social media use on mental health. Part 2: Show (with a plot) why we do/don't want include personal interests in the model.

Problem 6: Pesticide Exposure and Health Outcomes

Scenario

Researchers are studying the effect of pesticide exposure on developing a specific health condition. They are collecting data on individuals' exposure levels (PE), presence of the health

condition (HC), and two lifestyle factors. One of the lifestyle factors considered is dietary habits (D), which can influence both the level of pesticide exposure (e.g., due to consumption of non-organic foods) and health outcomes independently. A second lifestyle factor is the use of sunscreen (US). Among other things, the health condition makes skin very sensitive to UV rays and people with the condition must use sunscreen whenever they are outside.

Task: Construct a DAG for this setup. Decide which variables should be controlled for and which should not, to accurately estimate the **direct effect** of exposure levels on the health condition.