

Causal Inference Problem



 Uploaded image

Please help to solve the problem shown in the picture.



The statement in the image is:

True or False? The fundamental problem of causal inference is that it is not possible to observe the potential outcomes $\{Y_{0i}, Y_{1j}\}$ for unit i and unit j .

Answer: True.

Explanation: The fundamental problem of causal inference is that it is impossible to observe both potential outcomes for the same unit under different conditions (e.g., treated and untreated) at the same time. In this specific case:

- Y_{0i} represents the potential outcome for unit i if it does not receive the treatment.
- Y_{1j} represents the potential outcome for unit j if it receives the treatment.

The key issue here is that causal inference depends on comparing potential outcomes, but in practice, we can only observe one of the outcomes for a given unit at a given time. This is called the **fundamental problem of causal inference** because it creates a need for methods like randomized controlled trials, matching, or statistical modeling to estimate the unobserved potential outcomes.



 Uploaded image

The correct answer is False. The explanation is shown in this picture.



Thank you for sharing the explanation. Based on the clarification provided in the image:

The **correct answer is False**, because the fundamental problem of causal inference is specific to the inability to observe both potential outcomes ($Y_{0,i}$ and $Y_{1,i}$) for the **same unit i** .

The explanation highlights that while it is possible to observe $Y_{0,i}$ and $Y_{1,j}$ for **two different units i and j** , we cannot observe both potential outcomes ($Y_{0,i}$ and $Y_{1,i}$) for the same unit due to the mutually exclusive nature of treatments.

This distinction is key to understanding the nature of the fundamental problem of causal inference. Thank you for pointing that out!



Suppose we want to know the effect of temperature on water. What assumptions do we need to make a causal inference on the effect of temperature on the state of the water in a single glass? Select all that apply.

- A. Temporal stability
- B. Unit homogeneity
- C. Random distribution of water molecule
- D. Reversibility (causal transience)



To make a causal inference on the effect of temperature on the state of water in a single glass, the following assumptions are relevant:

Correct Answers:

A. Temporal stability

- This assumes that the causal relationship between temperature and the state of water (e.g., liquid, solid, or gas) remains stable over time. For example, if water freezes at 0°C today, it would also freeze at 0°C tomorrow under the same conditions.

B. Unit homogeneity

- This assumes that all parts of the water in the glass behave the same way under the same conditions. For example, every water molecule in the glass will respond uniformly to the temperature change.

D. Reversibility (causal transience)

- This assumes that the relationship between temperature and the state of water is reversible. For instance, if water freezes at 0°C, it should return to the liquid state when heated above 0°C.

Explanation of Incorrect Option:

C. Random distribution of water molecule

- This is not required for causal inference in this case. While water molecules are randomly distributed in a liquid state, this property is not necessary for understanding the causal effect of temperature on the state of water.

Summary:

The correct assumptions for causal inference in this context are **A, B, and D**.



But my professor said that the correct answers are only A and D.

Here is the explanation:

If the causal effect of X on Y is temporally stable and reversible, we can observe $Y_{1i} - Y_{0i}$ by repeatedly changing X from 0 to 1. Water transforming from ice to steam and back is an example

while treatment for high cholesterol for patient i is a counter-example. Note that temporal stability and causal transience cannot be tested, and these assumptions may not always be plausible.



Thank you for sharing the professor's explanation. Based on this clarification, the **correct answers are indeed A (Temporal stability) and D (Reversibility/Causal transience)**. Let me provide further insights into why **B (Unit homogeneity)** is excluded:

Why A (Temporal stability) is correct:

Temporal stability assumes that the effect of temperature on the state of water remains consistent over time. For example, if water freezes at 0°C today, it will also freeze at 0°C tomorrow, as long as the conditions are identical. This stability allows us to infer causality by observing repeated transitions between states.

Why D (Reversibility/Causal transience) is correct:

Reversibility assumes that the causal relationship between temperature and the state of water can be observed in both directions. For example:

- Increasing the temperature causes ice to melt into liquid water.
- Decreasing the temperature causes liquid water to freeze back into ice.

This reversibility enables direct observation of the causal effect by toggling the treatment variable (temperature).

Why B (Unit homogeneity) is excluded:

Unit homogeneity assumes that different units (e.g., different water molecules) respond identically to a treatment (temperature). However, in the context of a single glass of water, we are not comparing different units—only observing the same unit (the same glass of water) under varying conditions. The causal inference here depends on the stability and reversibility of the effect, not on the homogeneity of responses across units.

Why C (Random distribution of water molecules) is irrelevant:

The spatial arrangement or randomness of water molecules does not directly impact the causal relationship between temperature and the state of water. It is not an assumption needed to infer causality in this scenario.

Final Answer:

The correct assumptions are **A (Temporal stability)** and **D (Reversibility/Causal transience)**, as these are the necessary conditions to repeatedly observe and infer the causal effect of temperature on the state of water.



Now I have another problem:

Suppose we want to know the effect of temperature on water. What assumptions do we need to make a causal inference on the effect of temperature on the state of the water in two different glasses. Imagine that the water in one glass transforms to ice and the water in the other glass transforms to steam. Select all that apply.

- A. Temporal stability
- B. Unit homogeneity
- C. Random distribution of water molecule
- D. Reversibility (causal transience)



When considering the causal inference of temperature on the state of water in **two different glasses**, the assumptions required change slightly compared to a single glass. Here's an analysis of the assumptions:

Correct Answers:

A. Temporal stability

- Temporal stability is still required. The causal effect of temperature on the state of water in each glass must remain stable over time, meaning the relationship between temperature and the water's state (ice or steam) should not vary between observations.

B. Unit homogeneity

- This is now required because we are comparing two different glasses of water. Unit homogeneity assumes that the glasses of water (units) are identical in relevant properties (e.g., same volume, composition, pressure, etc.) and will respond identically to the same temperature. Without this assumption, differences between the two glasses could confound the causal inference.

D. Reversibility (causal transience)

- Reversibility is required here as well. This allows us to observe the causal effect in both directions—for example, cooling the glass with steam to form liquid water and then freezing it to form ice, and heating the glass with ice to form liquid water and then steam.

Why C (Random distribution of water molecules) is excluded:

- The random arrangement of water molecules is not relevant to inferring causality here. This property does not directly impact the state transitions (e.g., ice to water to steam) or the causal inference process.

Final Answer:

The correct assumptions are:

A. Temporal stability

B. Unit homogeneity

D. Reversibility (causal transience)

These assumptions ensure that the causal relationship can be consistently and validly observed across the two glasses of water.



 Uploaded image

Unfortunately, my professor said that the only correct answer is B.
The explanation is shown in the picture attached.



Based on the explanation provided in the image, the **only correct answer is indeed B (Unit homogeneity)**. Let's revisit the reasoning:

Explanation:

1. **Unit homogeneity (B)** is the key assumption in this scenario because we are comparing two different glasses of water. The explanation explicitly states that if Y_{1i} and Y_{0i} are identical for all i , we can measure the causal effect by taking the difference $Y_{1i} - Y_{0j}$ for $i \neq j$. This implies that the glasses of water must be homogeneous (identical in relevant characteristics) to draw valid causal inferences.
2. **Temporal stability (A)** and **Reversibility (D)** are not required here because the focus is on comparing different units (two glasses of water) rather than observing changes over time or toggling states within the same unit.
3. The explanation emphasizes the **homogeneity of water molecules** as a necessary condition for causal inference, implying that only B applies in this context.

Why the other options are excluded:

- **A (Temporal stability):** Not required here because the inference does not rely on repeated measurements of the same unit over time.
- **D (Reversibility):** Not required because the comparison is between two separate units (glasses of water), not the reversibility of states within a single unit.
- **C (Random distribution of water molecules):** Irrelevant to the causal inference process.

Final Answer:

The only correct assumption for this problem is **B. Unit homogeneity**.



Based on the problems above, please construct an engaging explanation that explains the fancy concepts from scratch behind the problems.

Please also include the necessary fancy math tools and symbol involved.
Please add some quantitative tangible examples, and some humor.



Unveiling the Secrets of Causal Inference: Why Water Freezes, Boils, and Confuses Us

Ah, causal inference—the magical science of figuring out *why stuff happens*. It's like detective work, but with math. Imagine you're trying to understand what happens to water when you play with its temperature. Does it freeze? Boil? Or just sit there judging your experiments? Let's dive in, one concept at a time, and unravel the mysteries using fancy math, tangible examples, and a sprinkle of humor.

1. The Fundamental Problem of Causal Inference

Here's the crux: **you can never observe both "what happened" and "what could have happened" for the same thing at the same time.**

Math Version:

For any unit i , let:

- Y_{1i} : The outcome if the unit receives the treatment (e.g., heat the water to 100°C—it boils).
- Y_{0i} : The outcome if the unit does not receive the treatment (e.g., cool the water to -5°C—it freezes).

The causal effect is defined as:

$$\text{Causal Effect}_i = Y_{1i} - Y_{0i}$$

But there's a catch: you can only observe one of these at a time. If you heat the water to 100°C and watch it boil (Y_{1i}), you can't simultaneously observe how the same water would behave at -5°C (Y_{0i}). You'd need a time machine—or Hogwarts.

2. Let's Experiment: One Glass of Water

You have one glass of water. What do you need to figure out the effect of temperature on its state? Let's bring in **temporal stability** and **reversibility**!

A Tangible Example:

Imagine you're heating and cooling the water in one glass:

1. At -5°C, it's ice ($Y_{0i} = \text{solid}$).
2. At 100°C, it's steam ($Y_{1i} = \text{gas}$).

If the water freezes, melts, and steams reliably every time you flip the temperature switch, congratulations! **Temporal stability** is holding. And if you can repeat this back and forth without breaking physics, you've got **reversibility** (or as I like to call it, "the boomerang effect").

Fancy Math:

If Y_{1i} and Y_{0i} are stable over time and you can toggle between them:

$$\Delta Y_i = Y_{1i} - Y_{0i}$$

You've got your causal effect. But remember, this works because you only have one glass and keep experimenting with it—like a determined scientist or a bored cat.

3. What About Two Glasses of Water?

Now you're dealing with **two glasses**—one turns into ice, the other into steam. Can you compare them? Well, now you need **unit homogeneity** (fancy term for “make sure the glasses are the same!”).

A Tangible Example:

Suppose Glass A has 250 ml of distilled water, while Glass B is half-full with coffee someone left at the office. You heat them both to 100°C. Guess what? The coffee boils faster, smells better, and confuses the experiment. Without **unit homogeneity**, your causal inference falls apart faster than your New Year's resolutions.

Fancy Math:

If Glass A and Glass B are **homogeneous**, meaning:

$$Y_{1i} = Y_{1j}, \quad Y_{0i} = Y_{0j} \quad \text{for } i \neq j,$$

then you can compare outcomes across units. For instance:

$$\text{Causal Effect}_{A \rightarrow B} = Y_{1A} - Y_{0B}.$$

But if the two glasses aren't the same, your inference becomes as shaky as your third cup of coffee.

4. Humor Break: Why Water Molecules Don't Care About Your PhD

Some people think **random distribution of water molecules** matters for causal inference. But let's face it—water molecules don't care about your statistics class. They just want to vibe around. Unless you're running quantum-level experiments, randomness isn't the star of this show. Save it for your next ice-cube party trick.

5. Practical Problem Solving

Here's how the assumptions come together:

Scenario	Required Assumptions	Why They Matter
One glass, repeated experiments	Temporal stability, reversibility	So you can flip between outcomes repeatedly.
Two glasses, one frozen, one steamed	Unit homogeneity	So you're comparing apples to apples (or water to water).

6. The Big Picture

Causal inference helps us understand the “what if” world. In the water example:

- Temporal stability lets us trust repeated observations.
- Reversibility lets us toggle outcomes.
- Unit homogeneity ensures fair comparisons across units.

But in real life, these assumptions can break down. For example:

- **Temporal instability:** What if your freezer breaks? Now ice melts even at -5°C .
- **Non-reversibility:** Once you boil water, dissolved gases leave, and the water isn't the same anymore.
- **Heterogeneous units:** Comparing a glass of saltwater to distilled water won't work—they'll freeze and boil differently.

Final Thoughts

Causal inference is like hosting a party where all your assumptions need to show up and behave. If they don't, your results might still look cool but won't mean much. So whether you're freezing water, treating cholesterol, or figuring out why your coffee keeps spilling—keep your math clean, your units homogeneous, and your sense of humor intact.

Oh, and always remember: science is 10% math, 90% cleaning up after experiments.



To reinforce the understanding, please generate some similar problems and show the solutions.



Problem 1: Effect of Fertilizer on Plant Growth in One Pot

You want to know the effect of a specific fertilizer on the growth of a single plant. You can either apply the fertilizer or not. The plant grows to 10 cm if fertilizer is applied, and it grows to 5 cm without the fertilizer.

Questions:

1. What assumptions are required to infer the causal effect of the fertilizer on the plant's growth?
2. Calculate the causal effect based on these assumptions.

Solution:

1. Required Assumptions:

- **Temporal stability:** The plant's growth must be stable across time. For example, if it grows to 10 cm with fertilizer today, it should also grow to 10 cm with fertilizer tomorrow under identical conditions.
- **Reversibility (Causal transience):** You should be able to observe the plant's growth with fertilizer and without fertilizer by alternating the treatment (fertilizer).

2. **Causal Effect:** Let:

$$Y_1 = 10 \quad (\text{growth with fertilizer})$$

$$Y_0 = 5 \quad (\text{growth without fertilizer})$$

The causal effect is:

$$\Delta Y = Y_1 - Y_0 = 10 - 5 = 5 \text{ cm.}$$

The fertilizer increases growth by 5 cm.

Problem 2: Effect of Diet on Weight Loss Across Two Groups

You want to compare two groups of individuals to measure the causal effect of a new diet plan. Group A lost 3 kg following the new diet, while Group B (control group) gained 1 kg without the diet.

Questions:

1. What assumptions are required to infer the causal effect of the diet on weight loss between the two groups?
2. Calculate the causal effect.

Solution:

1. **Required Assumptions:**

- **Unit homogeneity:** Both groups must be similar in relevant characteristics, such as initial weight, age, and metabolism. For example, if Group A consists of athletes and Group B consists of sedentary individuals, the comparison would be invalid.

2. **Causal Effect:** Let:

$$Y_1 = -3 \quad (\text{weight change with the diet, Group A})$$

$$Y_0 = +1 \quad (\text{weight change without the diet, Group B})$$

The causal effect is:

$$\Delta Y = Y_1 - Y_0 = -3 - 1 = -4 \text{ kg.}$$

The diet results in a 4 kg weight loss compared to no diet.

Problem 3: Effect of Light on Photosynthesis in Two Different Leaves

You're studying the effect of sunlight on photosynthesis. You test two different leaves: Leaf A (under sunlight) produces 5 units of glucose, while Leaf B (in darkness) produces 1 unit.

Questions:

1. What assumptions are required to make a valid causal inference?
2. Calculate the causal effect of sunlight on photosynthesis.

Solution:

1. Required Assumptions:

- **Unit homogeneity:** Leaf A and Leaf B must be identical in all respects except for the treatment (sunlight). For instance, if Leaf A is from a healthy plant and Leaf B is from a stressed plant, the results would be invalid.

2. Causal Effect: Let:

$$Y_1 = 5 \quad (\text{glucose production under sunlight, Leaf A})$$

$$Y_0 = 1 \quad (\text{glucose production in darkness, Leaf B})$$

The causal effect is:

$$\Delta Y = Y_1 - Y_0 = 5 - 1 = 4 \text{ units.}$$

Sunlight increases glucose production by 4 units.

Problem 4: Comparing Two Glasses of Water

You have two glasses of water:

- Glass A is heated to 100°C and turns into steam.
- Glass B is cooled to -10°C and freezes into ice.

Questions:

1. Can you measure the causal effect of temperature on the state of water across these two glasses?
2. What assumptions must hold for your inference to be valid?

Solution:

1. **Can You Measure the Effect?** Yes, but only if specific assumptions are met.

2. Required Assumptions:

- **Unit homogeneity:** Glass A and Glass B must contain identical water samples (e.g., same volume, same purity, no contaminants). Otherwise, differences in water properties might confound the effect.

3. **Causal Inference:** Since the state change is categorical (liquid to gas or liquid to solid), we describe the causal effect qualitatively:

- Heating to 100°C causes water to transition from liquid to gas.
- Cooling to -10°C causes water to transition from liquid to solid. The assumption of **unit homogeneity** ensures these changes are due to temperature and not some other factor.

Problem 5: Comparing Productivity with and Without Coffee

You want to measure how coffee affects productivity. One day you drink coffee and finish 10 tasks, while another day you don't drink coffee and finish 6 tasks.

Questions:

1. What assumptions are required to infer the causal effect of coffee on productivity?
 2. Calculate the causal effect.
-

Solution:

1. Required Assumptions:

- **Temporal stability:** Your productivity levels (tasks completed) should remain consistent across similar days if the treatment (coffee) is the only change.
- **Reversibility:** You should be able to alternate between drinking coffee and not drinking coffee on similar days to observe the effect.

2. Causal Effect: Let:

$$Y_1 = 10 \quad (\text{tasks completed with coffee})$$

$$Y_0 = 6 \quad (\text{tasks completed without coffee})$$

The causal effect is:

$$\Delta Y = Y_1 - Y_0 = 10 - 6 = 4 \text{ tasks.}$$

Coffee increases productivity by 4 tasks per day!

Takeaway

Each of these problems emphasizes the importance of assumptions like **temporal stability**, **unit homogeneity**, and **reversibility**. They're the scaffolding holding up causal inference! And while math is fancy, the real magic is making sure your assumptions hold in practice—whether you're growing plants, losing weight, or brewing coffee.