Quizzes

I. Causality

1. Consider an individual i who receives a binary treatment D_i (1 for treated, 0 for not treated). Let Y be the outcome of interest, with Y_i(1) and Y_i(0) the elements of the potential outcome.

- **Describe the causal question;**
  The causal question is: does the treatment D cause a change in the outcome Y?
- **Describe what Y_i(1) and Y_i(0) mean. Can they be both observed?**
  Y_i(1) means the outcome we would see if a person is treated.
  Y_i(0) means the outcome we would see if a person is not treated.
  For every person, only one of these is visible in reality. We never see both.
  This is called the Fundamental Problem of Causal Inference.
- **What is the distinction between the individual treatment effect and the average treatment effect?**
  The individual treatment effect (ITE) is the effect for one person, calculated as Y(1) minus Y(0).
  The average treatment effect (ATE) is the average of all these individual effects across the whole population.

2. Explain in words the meaning of the "simple difference in mean outcomes" (SDO). Why is it not a good measure of the effect of a treatment?
SDO is the difference between the average outcome of the treated group and the average outcome of the untreated group.
It is often biased and is called the naive average treatment effect.
It is not a good measure because the treated and untreated groups may already be different even before the treatment starts.
The SDO thus mixes the true effect with selection bias.

3. Under which condition(s) are SDO and ATE equivalent?
SDO equals ATE only if treatment is assigned independently of potential outcomes.
This happens when treatment is randomly assigned (Randomization).
Randomization makes the treated and untreated groups statistically comparable, eliminating selection bias.

4. Explain in words the meaning of SUTVA.
SUTVA stands for Stable Unit Treatment Value Assumption.
It has two parts:

- Everyone in the treatment group receives the exact same version (dose) of the treatment.

- There are no externalities or interference; one person's treatment status does not affect the outcome of any other person.

This assumption is critical for isolating the individual effect.

5. Why would one include control variables in a regression even if randomizing a treatment?

Even with perfect randomization, including control variables helps to improve precision.
Controls absorb non-treatment-related variation (noise) in the outcome.
This reduces the residual error and makes the standard errors smaller, leading to more reliable estimates.
Crucially, since treatment is randomized, adding controls does not change the causal interpretation of the treatment effect.
Controls are also necessary if the randomization was done conditionally (stratified randomization).

II. Randomization

1. Internal vs. External Validity

- Internal Validity: This is the most important type of validity for causal inference. It means the observed effect on the outcome (dependent variable) is truly due to the treatment and not something else, like a lurking variable or pre-existing difference.
- External Validity: This is about generalizability. It means the sample used in the study accurately represents the larger society or population you want to draw conclusions about. If a study is internally valid but externally invalid, the results are true only for the small group studied.

2. Law of Large Numbers (LLN)

LLN in simple terms: If we repeat an experiment many times, the average result of the sample will get closer and closer to the true average of the population.

Application to RCTs: This is the mathematical rule that guarantees randomization works. It says that if the sample size is large enough (usually 30 or more), the average of the sample will approach the true population mean. This ensures that the control and treatment groups are balanced on average.

3. Two-Stage Randomization

Concept:

- Stage 1 (Sampling): First, you take a randomized sample from the total population. This helps ensure external validity.
- Stage 2 (Assignment): Then, you randomly assign the individuals in that sample into the control group or the treatment group. This is the key step for internal validity.

Why it is a good strategy:

It eliminates selection bias by making the groups comparable.

It also eliminates the heterogeneous treatment problem (in terms of observable and unobservable characteristics).

It assures the Independence Assumption (treatment assignment is independent of potential outcomes).

4. Potential Pitfalls of Randomization

Non-compliance: The ideal assignment is broken. People who were assigned to get the treatment (eligible) don't want to get it. This requires using methods like Instrumental Variables (IV) to fix the estimate.

Attrition: Participants drop out of the study. It becomes difficult to observe the outcome of the treatment for these units. This is a problem if the people who drop out are not random (e.g., the sickest or poorest people leave).

Hawthorne effect: A participant (unit) who is being observed changes their behavior simply because they know they are part of a study. This affects the treatment group.

John Henry effect: The control group knows they are missing out on the treatment, so they try to work harder to compensate. This makes the measured effect of the treatment seem smaller than it really is.

5. RCTs for Different Interventions

Effectiveness of vaccines: RCTs are common in medical research and are the gold standard.

Issue: Externalities exist because vaccination protects others besides the recipient (herd immunity). This is a violation of the SUTVA assumption.

Use of seat-belts:

Issue: In practice, this is not feasible or ethical to randomize. Also, non-compliance would be very high, as people would simply ignore the assignment. Other methods are needed.

Use of parachutes:

Issue: This is the classic example where randomization is impossible. You cannot ethically assign people to jump without parachutes. It illustrates the limit of RCTs in questions where the risk is too high.

Effectiveness of masks:

Issue: RCT is possible by randomizing who gets masks. However, there is a risk of non-compliance and Hawthorne effects. Also, masks create externalities (protecting others by reducing transmission), meaning SUTVA may fail.

III. RDD

1. Why RDD provides LATE instead of ATE

LATE (Local Average Treatment Effect): RDD estimates the effect only at the specific cutoff point (the threshold).

The design assumes that individuals just above and just below the threshold are nearly identical - as if they were randomly assigned to treatment or control.

Because the estimation is only valid in this local neighborhood around the threshold, the resulting causal effect is local.

Therefore, RDD gives us the LATE, which is the average treatment effect for those units whose treatment status was determined by crossing the cutoff. This cannot be generalized to the entire sample (which would be ATE).

2. Difference between Sharp and Fuzzy Discontinuity

- Sharp Discontinuity (Sharp RD):

The cutoff precisely determines the treatment status.

Example: Those above the threshold must receive the treatment, and those below must not receive the treatment.

There is full compliance with the assignment rule.

The effect is estimated as the simple jump in the outcome at the cutoff.

- Fuzzy Discontinuity (Fuzzy RD):

The cutoff does not totally determine whether a person receives treatment or not.

Example: Some people above the threshold might not take the treatment, or some below might find a way to get it (non-compliance).

The cutoff is used as an Instrumental Variable (IV) to estimate the causal effect.

3. Why RDD would suffer more from power problems

Statistical Power: The ability to detect a true effect if one exists.

RDD estimation is done only around the threshold.

To maintain the assumption that groups are comparable (like randomization), we must use a very narrow bandwidth (only data points closest to the cutoff).

This approach reduces the number of observations used for the main estimation compared to the total sample size.

Fewer observations mean the model has less statistical power, making it harder to find a significant result.

4. How one would implement RDD (Regression)

Running Variable: continuous variable (the eligibility index) that determines the cutoff.

Implementation Steps:

- Construct a dummy variable (index): Assign a value of 1 if the individual is above the cutoff (e.g., income higher than the threshold) and 0 if they are not.
- Estimate the regression: Estimate the effect this index has on the study's outcome variable.
- Method: This is typically done using Local Linear Regression or a polynomial function, focusing on observations close to the threshold. The coefficient on the index

represents the difference between the control group (CG) and treatment group (TG) at the cutoff point.

5. Which type of data would one use to implement RD?

One must use continuous data for the eligibility index (the running variable).

The running variable must be continuous so that there is a smooth change right up to the discontinuity point.

Examples: income, age, test scores, or a geographical rating.

IV. DD

1. Explain in words what DD does.

DD is a quasi-experiment. It uses a natural break or policy change (an "abruption") as a substitute for true randomization.

It isolates the causal effect by comparing the change in the outcome for the treated group to the change in the outcome for the control group over the same time period.

The method constructs the counterfactual outcome (what would have happened to the treated group without the treatment) by assuming they would have followed the trend of the control group.

2. Explain the importance of the common trends assumption.

The treatment group and the control group must have followed a parallel trajectory (common trends) in the outcome variable during the period before the intervention.

This is the most crucial assumption. It ensures that any difference in outcomes after the treatment is due only to the treatment itself.

If the groups had diverging trends beforehand, the DD estimate will be biased because it incorrectly attributes that pre-existing difference to the treatment effect.

3. Which type of data would one use to implement DD?

DD requires data observed at multiple points in time (longitudinal data) for both the treatment and control groups.

The primary data structures are Panel Data (tracking the same units over time) or Repeated Cross-Sections (samples from the same population at different times).

You must be able to observe the outcome before and after the intervention.

4. Give an example of a placebo test.

A placebo test (or falsification test) is used to check the validity of the Common Trends Assumption.

- Example 1 (Fake Intervention Date): Run the DD analysis using only the pre-treatment data and set a "fake" intervention date during that period. If you find a significant treatment effect, the Common Trends Assumption is violated.
- Example 2 (Fake Treatment Group): Use a group in the population that you know was not affected by the policy as the "treatment group" in the DD model. If you find a significant effect, the common trends assumption is likely false, and the model is flawed.

V. Matching

1. Explain in words what matching does.

Problem: We cannot observe the counterfactual outcome for treated individuals (what would have happened if they hadn't been treated).

Solution (Matching): This is a statistical technique used to estimate this missing counterfactual.

Mechanism: It finds a control unit that is "closest" or most similar to a treated unit based on their observable characteristics (like age, income, education) that existed before the treatment started.

Goal: To measure the treatment effect by comparing the outcome of the treated unit with the outcome of its closely matched control unit.

2. Explain the two key assumptions under which matching is an appropriate technique to use.

Matching relies on two fundamental assumptions:

- Unconfoundedness (or Conditional Independence Assumption - CIA):
  This assumes that the selection into the treatment group is based only on characteristics that we can observe.
  In other words, once we control for all the observed characteristics, the treatment assignment is independent of the potential outcomes.
  If this holds, any remaining differences between the treated and matched control groups are due to the treatment.
- Overlap (or Common Support):
  This assumes there must be a sufficient overlap in the observed characteristics between the treated and control groups.
  You must be able to find control units with the same set of characteristics as the treated units.
  If this fails, you cannot find a suitable match for some treated individuals, and their effect cannot be estimated.

3. Explain when matching would not be an appropriate technique to use.

Matching is not appropriate when:

- Unobservables are Key: When selection into treatment is heavily influenced by factors that we cannot observe or measure (e.g., motivation, innate ability, future expectations). Since matching only controls for observables, unobservable differences will still cause selection bias.
- Insufficient Data Richness: When the observed characteristics are not rich enough to control for all the heterogeneity (differences) between the treated and control groups.

4. Explain in words what Propensity Score Matching (PSM) does and on which assumption(s) it relies.

What PSM is a probabilistic approach to matching. It simplifies the matching process by converting a complex, multi-dimensional set of characteristics (e.g., 10 different variables) into a single, one-dimensional score.

The PS is a number between 0 and 1 that represents the probability that an individual will enroll in (or receive) the program, given their observed pre-treatment characteristics.
Mechanism: Instead of matching individuals on all the separate characteristics, PSM matches individuals based on this single PS.
It relies on the same two key assumptions as standard matching: Unconfoundedness and Overlap.

5. Explain why ex post matching is not an appropriate method.

"Ex Post" Meaning: Refers to using data (characteristics) that were observed or collected after the treatment has already begun.
Using ex post characteristics is dangerous because these characteristics might have already been influenced by the treatment itself.
If a characteristic used for matching has been changed by the treatment, you are effectively matching based on a result of the treatment, which distorts the control group and leads to biased estimates of the treatment effect.
Rule: Matching must always be based on characteristics observed ex ante (before the treatment).

VI. IV

1. Why use IV within the context of programme evaluation?

In real-world programs, we often face non-compliance. People assigned to the treatment group might refuse the treatment, and people assigned to the control group might find a way to get the treatment.

This non-compliance violates the core assumption of randomization. The simple comparison of outcomes (SDO) becomes biased.

The Solution: IV is used to correct this bias caused by non-compliance, restoring the ability to find a causal effect when treatment take-up is voluntary or imperfect.

2. ITT (Intention-to-Treat) and TOT (Treatment-on-the-Treated)

ITT (Intention-to-Treat): Compares the outcomes of groups based on their random assignment (who was offered the treatment), regardless of whether they actually took it. This is the effect randomization naturally gives.

TOT (Treatment-on-the-Treated): This is the true causal effect for those individuals who actually received the treatment. This is the parameter we want to estimate. TOT is usually larger than ITT.

When they coincide: ITT and TOT are equivalent only under the condition of full compliance.

Explanation: If everyone assigned to treatment takes it, and no one in the control group takes it, then the group assigned to treatment is identical to the group that actually received it.

3. Problems of generating an instrument ex post

Ex Post Meaning: Using a characteristic or variable that was measured or collected after the treatment had already begun.

The Requirement: A valid Instrumental Variable (IV) must influence the probability of receiving the treatment, but must not directly affect the outcome (Exclusion Restriction), except through the treatment itself.

The Problem: If an instrument is generated ex post, there is a high risk that the instrument itself was influenced by the treatment or the outcome.

Consequence: Using an ex post instrument violates the Exclusion Restriction and leads to biased (wrong) estimates. The instrument must always be Ex Ante (measured before treatment).

4. Explain the instrument of Randomized Promotion

What it is: A randomized promotion (or encouragement) is an instrument that randomly assigns an incentive, information, or persuasion to a sub-group of participants to increase their likelihood of taking up the program.

Purpose: It is used to correct for non-compliance in programs where individuals can choose to enroll.

Key Issues (for Validity):

1. Relevance Condition: The promotion must substantially increase the actual take-up of the program.

2. Exclusion Restriction: The promotion itself must not directly affect the final outcome, except through the participant's decision to enroll in the program.

5. Why the IV strategy yields LATE but not ATE?

IV is a local effect (LATE). This is the average treatment effect only for a specific sub-group called "Compliers".

Who are Compliers?: Compliers are those individuals whose decision to participate is changed by the instrument (i.e., they only participated because they received the promotion/instrument).

Limitation: IV cannot estimate the effect for the entire population (ATE) because it focuses only on the behavior of those units who were "induced" to change their behavior by the instrument.

Caution: The results from IV cannot be easily extrapolated to the whole population, as the effect is localized to the Compliers group. Different instruments will identify LATE for different groups of Compliers.

VII. Exam question example

Question 1: Randomization

(a) Describe a two-stage randomization procedure. Discuss how it will imply internal and external validity of the results.

A formal two-stage randomization procedure follows a specific sequence to ensure the scientific rigor of an evaluation. In the first stage, a random sample of units is selected from a larger, well-defined eligible population. In the second stage, this sample is randomly assigned into a treatment group and a comparison (control) group.

U can make a scheme instead of text

This procedure is designed to achieve two distinct types of validity:

• External Validity: The first stage ensures that the results obtained from the study sample can be generalized to the broader population of eligible units within a defined level of sampling error.

• Internal Validity: The second stage ensures that the observed effect on the dependent variable is truly due to the treatment rather than confounding factors. By randomly assigning the sample, the treatment and comparison groups become equivalent in both observed and unobserved characteristics.

(b) Back to the Randomization lab, how would you test for balance in a baseline covariate?

To test for balance in a baseline covariate, you compare the mean values of that variable (e.g., age, income, or previous test scores) between the treatment and control groups before the intervention begins.

• In practice, this is done by running a regression where the baseline covariate is the dependent variable and the treatment assignment is the independent variable.

• If the randomization was successful, there should be no statistically significant difference between the groups. A high p-value (typically greater than 0.05) indicates that the groups are balanced and comparable.

Question 2: Compare the sharp and fuzzy discontinuity RD designs.

The primary difference between Sharp and Fuzzy Regression Discontinuity (RD) designs lies in the nature of the treatment assignment rule and the level of compliance.

• Sharp RD Design: In this design, treatment assignment is a deterministic function of the running variable. There is full compliance, meaning the discontinuity at the threshold precisely determines the treatment: everyone above the threshold receives it, and everyone below it does not. The causal effect is identified by the direct "jump" in the outcome variable at the cutoff.

• Fuzzy RD Design: In a fuzzy design, the threshold does not totally determine whether an individual receives treatment, often due to non-compliance or marginal differences in eligibility. Instead, the probability of treatment changes discontinuously at the cutoff.

• Estimation Differences: While Sharp RD provides an estimate of the effect for those at the threshold, Fuzzy RD requires an Instrumental Variables (IV) approach, using

the threshold as an instrument for actual participation. Both designs typically yield a Local Average Treatment Effect (LATE) rather than an Average Treatment Effect (ATE) for the whole sample, because the estimation is performed only around the threshold.

U can write it in a table instead of text

Question 3: State and explain the unconfoundedness assumption. What does it imply for the propensity score?

The unconfoundedness assumption (also known as the Conditional Independence Assumption) is a cornerstone of matching methods.

• Explanation: It assumes that the potential outcomes are independent of the treatment assignment, provided we control for a set of observable characteristics. This means that once we account for all visible factors that might influence both participation and the outcome, the selection into the treatment group is essentially as good as random.

• Implication for the Propensity Score: If the unconfoundedness assumption holds based on a vector of observable characteristics, it also holds when conditioned only on the propensity score. The propensity score is a single number (from 0 to 1) representing the probability of enrollment based on those characteristics.

• Simplification: This is significant because it allows researchers to convert a complex, multidimensional matching problem (e.g., comparing 10 different variables) into a one-dimensional process, matching treated and control units based solely on their propensity scores.

Question 4: Assume no full compliance in a programme. What are the desired properties of an instrument?

In the specific setting where participants choose whether to enroll (non-compliance), an instrument must possess four key properties to be valid:

1. Relevance: The instrument must be correlated with program participation. For example, in a "randomized promotion" design, the promotion campaign must substantially increase the take-up of the program.
2. Exclusion Restriction: The instrument must affect the outcome of interest only through its impact on program participation. It cannot have a direct effect on the outcome itself.
3. Independence (Ex-Ante): To avoid being affected by the outcome of the study, the instrument should ideally be determined ex-ante (before the treatment starts) and be as good as randomly assigned.
4. Monotonicity: The instrument should affect the likelihood of participation in the same direction for everyone (e.g., the promotion should make everyone more likely to join, not some more and some less).

Question 5: Focusing on Matching, state the variables used and compared. Alternatively, the variation used to identify the treatment effect.

Matching is a statistical technique used to estimate the missing counterfactual by pairing individuals from a treatment group with "close" individuals from a control group.

• Variables Used: Matching relies entirely on observable pre-treatment characteristics. These variables must be "rich enough" to control for all relevant heterogeneity between the treated and control groups. Importantly, these variables must be collected ex-ante (before the treatment) so they are not influenced by the treatment itself.
• Variables Compared: The method compares the outcomes of the treated units with the outcomes of their matched counterparts in the control group.
• Variation for Identification: The treatment effect is identified through cross-sectional variation between matched pairs. It assumes that by balancing all observable differences between the two groups, any remaining difference in their outcomes can be attributed solely to the treatment. Unlike methods that use time-series variation, matching identifies the effect by comparing similar units at a specific point in time.