Causal Data Analysis

2025/2026, 1st semester

Study Questions

Short questions. Answer the following questions. Always explain your answer. No points are given without explanation.

1. Is the statement true, false or it depends? “In experimental data ATE will always be equal to ITE.” (Assume that the randomization was done correctly.)
2. Is the statement true, false or it depends? “In experimental data, a simple regression conveys the same information as a simple mean comparison between the treated and control groups.”
3. Is the statement true, false or it depends? “Exact matching produces the same estimated effect as running an OLS regression with controls for the variables which were used in the exact matching.”
4. Is the statement true, false or it depends? “A simple OLS regression will always provide biased estimates.”

Long questions. Write a short essay and explain the problem asked in the question.

1. What is the ideal way to measure the causal effect of x on y (ITE), and how to do this in reality (ATE)? Define the two measures and provide the conditions when ATE=ITE (and explain the formulae). Give an example when they are not equal.
2. Explain the three types of endogeneity which can arise in observational data. Draw the causal map and give an example for each type.
3. An analyst runs the following regression:

   $$quality^{E} = \alpha + \beta \times family$$

   where quality=the quality management (1-worst…5-best), family is a dummy =1 if the firm is in family ownership (and 0 otherwise). The regression is run with simple OLS with control variables (β=-0.10**, N=8440), and with exact matching on the same controls (β=-0.16**, N=1207).

   - What is the meaning of β?
   - Why is the number of observations (N) different in the two regressions?
   - What can be the reason for β being different in the OLS and matched estimation?
   - What are the limits of matching estimation? Give an example when matching produces biased estimates.
4. Describe how to do propensity score matching. In your answer, explain the differences between this method and exact matching.