

Exercise 2

Data Generating Experiments & Traditional Data Analysis

Helpful references: Chapter 1 of vdL&R (2011).

The actual observed data experiment is a crucial factor in determining what scientific questions can be answered and what scientific assumptions are needed to justify answers. We generally refer to the observed data of the experiment on n units as n independent and identically (i.i.d.) distributed copies of a random variable O .

1. What does it mean to say that O is a random variable?
2. Provide an example in which the n units of O are identically distributed but are not independent.
3. Provide an example in which the n units of O are independent but are not identically distributed.
4. Let the observed data be denoted $O = (W, A, Y)$ and we use
 W to denote _____,
 A to denote _____, and
 Y to denote _____.
We might also include Δ in O to denote _____.
5. In a randomized trial the assignment of _____ is controlled by the experiment there is no

unmeasured confounding for _____. What is unmeasured confounding?

6. Compare and contrast randomized control trials and observational studies. Give three distinctions.

7. List a few criticisms of the traditional approach to statistical data analysis (i.e. the practice that lets the type of data at hand dictate the scientific question of interest and the statistical model).