Causal statistics for treatment models
Ecole Normale Suprieure–PSL, L3 Economics and Cogmaster, 2025

Exercise on manipulating counterfactuals

In the public employment service (PES), job-seekers may benefit from a job-search program. The outcome of interest is a dummy for being on a job after 6 months (or not), and we call it Y, with $Y(1)$ the counterfactual when the job-seeker receives the program and $Y(0)$ the counterfactual when she doesn't. Y is observed for all job-seekers.

The PES is wondering whether it should rely on counselor's expertise to allocate job-seekers to the program, or whether it should use an algorithm that generates a score based on job-seekers characteristics. (Nota: we are not directly interested here in evaluating the impact of the program on job finding, but rather on the best allocation rule).

A score has been produced for all job-seekers **and the score recommendation is observed**. We note $S = 1$ when according to the score the job-seeker should be sent to the program and $S = 0$ otherwise. BUT, the decision to actually send job-seekers to the program has been left to the counselors, with the following rule: when $S = 1$ the counselor must send the job-seeker to the program; but when $S = 0$ she can decide to overrule the algorithm, and send the job-seeker to the program nevertheless. We note $C = 1$ if the job seeker has been sent to the program, and $C = 0$ if she hasn't. C **is also observed**.

Here are the proportions of job-seekers sent to the program for the four possible combinations of algorithm and counselor decision:

	$S = 0$	$S = 1$
$C = 0$	0.2	0
$C = 1$	0.2	0.6

1. Comment this table; is it consistent with the description of the process? Do counselors overrule a lot the algorithm?
2. Compute the empirical values of $P(S = 1)$, $P(C = 1)$, $P(S = 0|C = 1)$, $P(S = 1|C = 1)$, $P(C = 1|S = 1)$, $P(C = 1|S = 0)$ and $P(C = 0|S = 0)$ from this table. I recommend you to detail every time your Bayes formula.
3. The objective of the PES is to maximize the average value of:

$$W = Y - \alpha T$$

where α is some cost parameter set by the PES and T is a dummy for being treated or not. Interpret this formula, why would the PES take it as an objective function?

4. We call $E_S(W)$ the expected value of W in the population of job-seekers **if the allocation rule followed strictly the score-based decision**, with no intervention from counselors. Show that:

$$E_S(W) = E(Y(1)|S=1)P(S=1)+E(Y(0)|S=0)P(S=0)-\alpha P(S=1)$$

5. In this expression, which terms can be directly observed in the data from the actual working of the program, and which cannot?

6. We call $E_C(W)$ the expected value of W in the population of job-seekers under the allocation rule that was implemented. Derive a similar expression of that quantity, but as a function of $P(C=1)$ and $P(C=0)$. Which terms can be directly observed in the data?

   Assume that the treatment effect is constant, i.e.

$$Y(1) = Y(0) + \Delta$$

   where Δ is a scalar (a number) common to all job-seekers.

7. Derive the expression for $E_S(W) - E_C(W)$. Which of the two decision rules would generate the maximum value of the PES objective? What would be the optimal decision rule?

Now, no longer assume that the treatment effect is constant, and allow for any form of heterogeneity.

The average values of the observed outcomes Y, depending on algorithm and counselor decisions are as follows:

	$S=0$	$S=1$
$C=0$	0.1	-
$C=1$	0.1	0.3

8. Compute the empirical value of $E_C(W)$ as a function of α.

9. Can you compute the empirical value of $E_S(W)$ as a function of α?