Science Table plots

Rachael Caelie (Rocky) Aikens

6/7/2021

Set up

A quantity inherently related to the prognostic score is the true potential outcome under the control assignment. Note that these are not the same. By it's informal definition, the prognostic score is the conditional expectation of the control potential outcome, conditional on the observed covariates, $E[Y(0)|X]$. Contrastingly, the control potential outcomes are the actual individual values $\{Y_i(0)\}_{i=1}^n$. While a treatment-focused prognostic score, for example $E[Y(1)|X]$, is less common, we might make a similar comparison between this quantity and the treated potential outcomes, $\{Y_i(1)\}_{i=1}^n$. The full suite of both treated and control potential outcomes in all study individuals is often said to make up the "Science Table."

Here, we consider what can be learned from considering these potential outcomes visually, as "Science Table plots."

A visual representation of individual treatment effects

Here we have the basic Science Table:
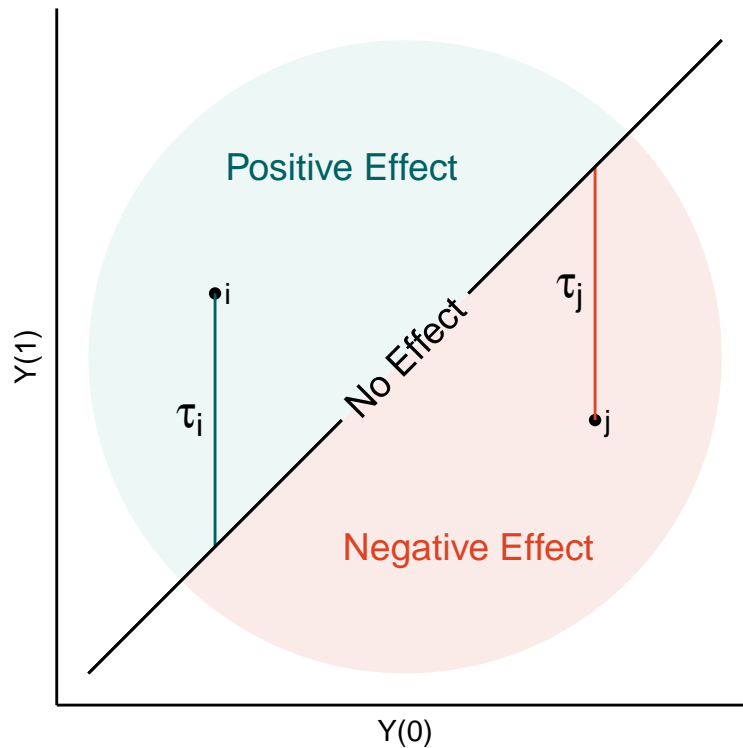
An individual's treatment effect is visualized by the difference between their horizontal and vertical coordinates on a Science Table plot. Individuals on the diagonal have no treatment effect. Individuals above and below the diagonal have positive and negative treatment effects, respectively, and a greater distance from the diagonal corresponds to a greater response to treatment. An interesting observation is that the individual treatment effect for an observation is the signed vertical distance to the diagonal (or, equivalently, the negative signed horizontal distance to the diagonal).
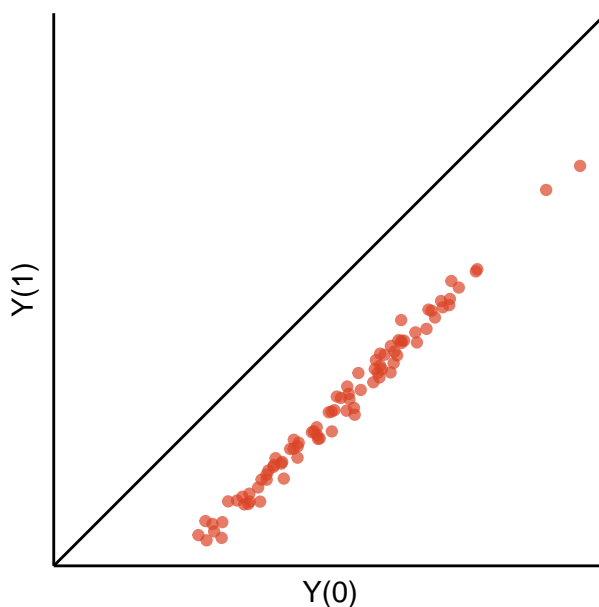
Treatment Effects on a Science Table plot

We can now re-interpret the sample average treament effect, visually, as the average signed distance between all observations and the diagonal. For example, the plots below show a fairly constant positive and negative treatment effects.
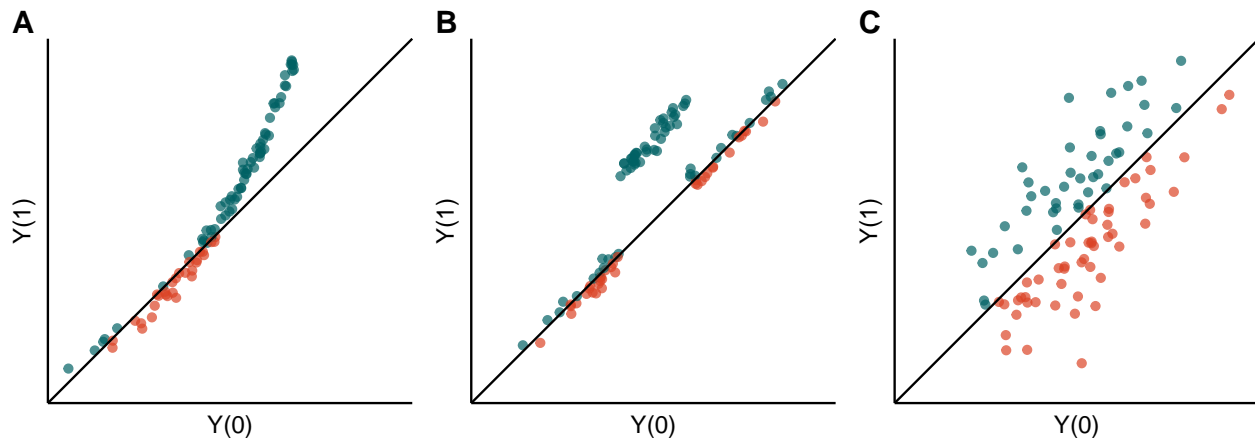
A
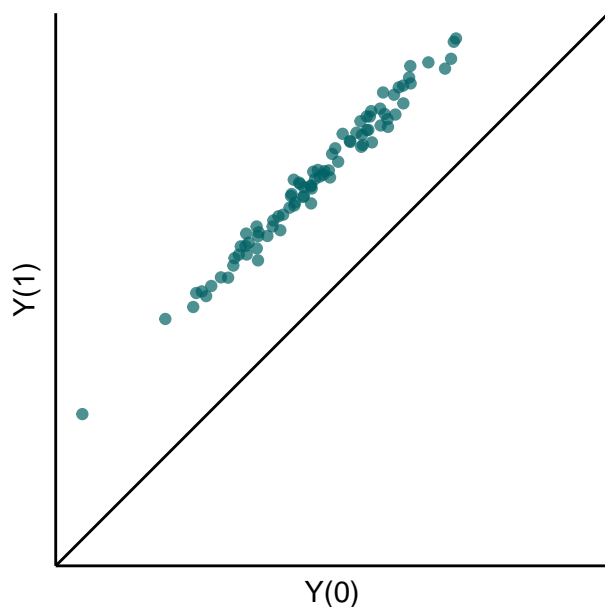
B

Treatment effect heterogeneity

A major focus in current causal inference literature is treatment effect heterogeneity. A Science Table plot allows us to illustrate different kinds of heterogeneity in treatment effect.

Panel A depicts a scenario in which individuals with higher $Y(0)$ tend to respond better to treatment. The opposite relationship is also possible: in which individuals with lower $Y(0)$ are more responsive. B describes an interesting scenario in which there is a treatment "sweet spot" of $Y(0)$ at which certain individuals benefit (Erin Craig and Rob Tibshirani are working on methods for "sweet spot" treatment effect estimation).

What about panel 3? Does panel 3 show treatment effect heterogeneity? Arguably yes, it does: the individual level treatment effects are variable. The difference between C and the two prior panels is that treatment response in panel C is not correllated with the control potential outcome, $Y(0)$. Perhaps the heterogeneity in treatment response in panel C can be explained by some other baseline covariate independent of $Y(0)$. Perhaps this covariate is measurable - and thus potentially actionable. Or perhaps it is unmeasurable, and so simply acts to add random noise in treatment response.

Variability in the treatment effect response

I incluse this last example just to point out that there is more to treatment effect heterogeneity than what I've walked through above. In the plot below, response to treatment is highly variable, and there is a correllation between $Y(0)$ and treatment effect variability. This points out that more than the average treatment effect - or even the conditional average treatment effect - can be important. In the scenario below, individuals with low $Y(0)$ consistently respond moderately well to treatment, while individuals with high $Y(0)$ sometimes respond very well and sometimes very badly. The average treatment effect, however, is the same at level sets of $Y(0)$. Thus, even if we knew the conditional average treatment effect with respect to, for example, a very good prognostic score, this might not tell the entire story.