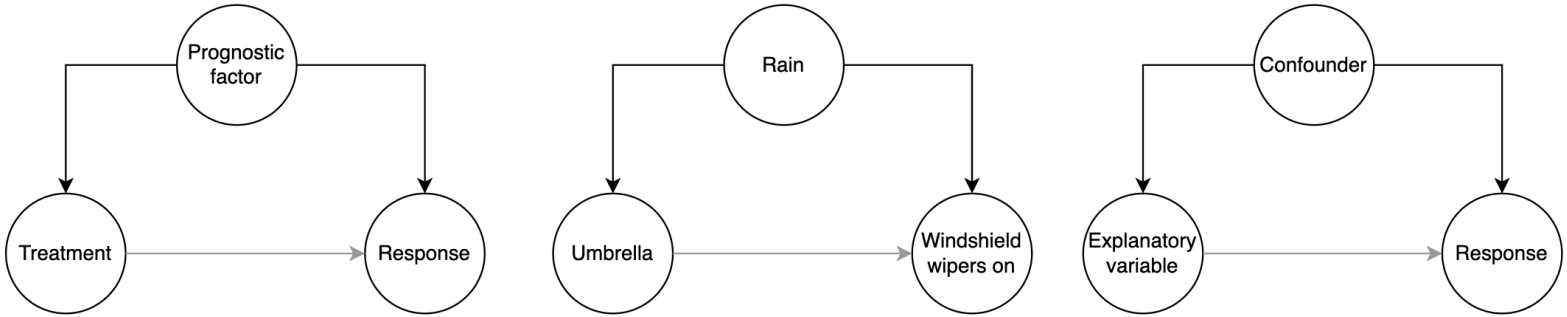
Ethics of Data Science – Part II

Confounders

What is a confounder?



A confounder is a common cause between two variables that, if ignored, introduces spurious correlation between those two variables. Treatment selection bias is a special case of such a situation.

Regression and confounders

It turns out that we can solve this problem sometimes with a simple regression.

First, a reminder of the potential outcomes framework, and the quantity we're trying to estimate:

$$\begin{aligned} Y_i &= Z_i Y_i^1 + (1 - Z_i) Y_i^0 \\ \text{ATE} &= \mathbb{E}[Y_i^1 - Y_i^0] \end{aligned}$$

However, we can only ever observe the following, which only agrees with ATE if treatment is randomized or at least completely independent of outcomes (or any other variable that correlates with outcomes):

$$\mathbf{E}[Y_i^1 \mid Z_i = 1] - \mathbf{E}[Y_i^0 \mid Z_i = 0]$$

Regression and confounders

It is actually relatively straightforward however to re-write the equation above into a simple regression:

$$Y_i = Z_i Y_i^1 + (1 - Z_i) Y_i^0$$
$$Y_i = Z_i (Y_i^1 - Y_i^0) + Y_i^0$$

Regression and confounders

It is actually relatively straightforward however to re-write the equation above into a simple regression:

$$
\begin{aligned}
Y_i &= Z_i Y_i^1 + (1 - Z_i) Y_i^0 \\
Y_i &= Z_i (Y_i^1 - Y_i^0) + Y_i^0 \\
Y_i &= d_i Z_i + b_i \\
&\text{where } d_i = Y_i^1 - Y_i^0, \text{ and } b_i = Y_i^0
\end{aligned}
$$

Although it seems we can estimate what we need using linear regression, you might want to recall the underlying assumptions of linear regression and observe one of them is violated by selection bias.

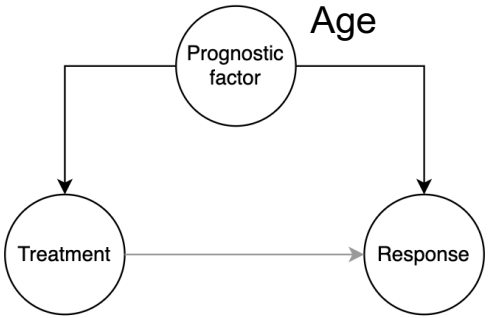
Regression and confounders

Imagine there is only one prognostic factor that clinicians use when deciding who to treat: age.

Treat with probability 0.8 if elderly, 0.2 otherwise

Intuitively, we might want to **control for** age in this situation – i.e., hold weight constant when we're estimating treatment effect.

Let's compare what happens when we do that, versus IPTW.



Regression and confounders

As before we consider two datasets, one where treatment is randomized, and another which is observational/real-world.
The real effect here is -0.5.

First, we compute the sample ATE, as expected, there is bias in RWD, not in RCT

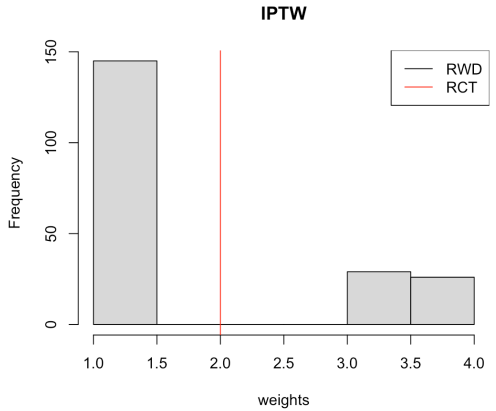
	RCT	RWD
Sample ATE	-0.489	-0.029

Regression and confounders

As before we consider two datasets, one where treatment is randomized, and another which is observational/real-world. ***The real effect here is -0.5.***

We then apply an IPTW correction. This has no effect on the RCT (the treatment variable there is actually independent of weight, hence the weights are all identical), but it does improve the RWD estimate.

	RCT	RWD
Sample ATE	-0.528	-0.045
IPTW sample ATE	-0.565	-0.507



Regression and confounders

As before we consider two datasets, one where treatment is randomized, and another which is observational/real-world. ***The real effect here is -0.5.***

We then consider a linear regression of response on treatment, without including age. The regression coefficient of treatment matches the sample ATE results.

	RCT	RWD
Sample ATE	-0.528	-0.045
IPTW sample ATE	-0.565	-0.507
Linreg (without age)	-0.528	-0.045

Regression and confounders

As before we consider two datasets, one where treatment is randomized, and another which is observational/real-world. ***The real effect here is -0.5.***

	RCT	RWD
Sample ATE	-0.528	-0.045
IPTW sample ATE	-0.565	-0.507
Linreg (without age)	-0.528	-0.045
IPTW linreg (without age)	-0.565	-0.507

We might think that weighting the regression itself using IPTW might fix the problem. Indeed, it does. Note that weights here are computed as 1 over the actual probability of the observed treatment value.

Regression and confounders

As before we consider two datasets, one where treatment is randomized, and another which is observational/real-world. ***The real effect here is -0.5.***

Finally, we add age as a covariate in the linear regression. RCT results are unaffected, because age is uncorrelated with treatment in that dataset. But RWD results improve, approximating IPTW ATE.

	RCT	RWD
Sample ATE	-0.528	-0.045
IPTW sample ATE	-0.565	-0.507
Linreg (without age)	-0.528	-0.045
IPTW linreg (without age)	-0.565	-0.507
Linreg (with age)	-0.565	-0.507

Matching

Yet another alternative is so-called “matching”. For example, we could match on age, creating a balanced cohort where for each treated patient, we sample from the dataset an untreated patient with the same age

	ind_treated	matched
[1,]	3	172
[2,]	5	198
[3,]	6	50
[4,]	8	154
[5,]	10	154
[6,]	13	54
[7,]	15	16
[8,]	18	125
[9,]	19	126
[10,]	20	62
[11,]	21	30

Matching

Yet another alternative is so-called “matching”. For example, we could match on age, creating a balanced cohort where for each treated patient, we sample from the dataset an untreated patient with the same age.

We then pretend that each patient’s “match” is a counterfactual observation for that same patient and compute the ATE directly.

This gets us to similar performance as IPTW.

	RCT	RWD
Sample ATE	-0.528	-0.045
IPTW sample ATE	-0.565	-0.507
Linreg (without age)	-0.528	-0.045
IPTW linreg (without age)	-0.565	-0.507
Linreg (with age)	-0.565	-0.507
Matching on age	-0.601	-0.476

Propensity matching

What if you suspect there are several factors that play into the treatment selection? How would you match then? For example, assume age and weight both play a role. Matching on multiple attributes can quickly get difficult.

One possibility is to still compute the propensity scores (note: probability of being treated, not of observed outcome of the treatment variable), and match on them.

Another possibility is to use propensity scores as covariates and allow regression to do our work for us.

	RCT	RWD
Sample ATE	-0.434	-0.244
IPTW sample ATE	-0.525	-0.498
Linreg (with age and weight)	-0.475	-0.497
Linreg (with propensity score)	-0.525	-0.501
Matching on propensity	-0.461	-0.490

Assumptions underlying propensity scores

Positivity

You will have noticed that we are taking the inverses of the probabilities of having been given the treatment a patient was actually given in a dataset. These values must therefore not be 0 or 1. We need at least a few (or even one!) observation of the counterfactual within each patient stratum to assign extra weight to.

$$0 < P(Z_i = 1 \mid X, Y_i^1, Y_i^0) < 1$$

No unmeasured confounders

The other core assumption is that we have measured and incorporated all relevant confounders (*no unmeasured confounders).* Unlike *positivity* which we can check, there is no way to check that latter condition. What if we added as many variables as we can, just in case they are confounders? We will see in the next lecture that this is unfortunately not a solution, and in fact it may do more harm than good.

$$P(Z_i \mid X, Y_i^1, Y_i^0) = P(Z_i \mid X)$$

Summary

We have seen several solutions to the problem of confounders:

- IPTW-adjusted sample ATE
- Controlling for confounders in a linear regression
- Matching on confounders and computing the sample ATE
- Controlling for propensity score in a linear regression
- Matching on propensity score and computing the sample ATE

But we need to pay attention to *positivity* (which we can easily check) and *no unmeasured confounders.* The latter is much harder to check, and conditioning on everything is not a good solution.