Kelsey Hoekstra
Math Major
2017

Taking the "con" out of econometrics

Over a year ago, John Oliver did a segment on scientific studies (link here). He called out studies for being sponsored with an agenda, small sample size, and inability to be replicated. This podcast had a similar purpose, although it focused on a different aspect. This interview focused on the creation of the control groups, data mining, and distinguishing between causal variable and characteristics.

One thing I hadn't thought about regarding analytics was context. It makes sense that higher wages in a city would correlate to successful companies and higher cost of living. That means when analyzing things like minimum wages, analysts can't compare San Francisco to a small town in Iowa. Instead, San Francisco should be compared to Portland, Oregon and Seattle, Washington. Susan Athley also pointed out the counterfactual is not static over time, so the control group could be a Frankenstein-like combination of other similar cities. The cities could be weighted to match characteristics of the city of study. The risk there lies in what aspects to control. Machine learning algorithms could try out multiple combinations.

On another end of the "con" spectrum, small studies, like psychology experiments cannot be replicated due to small sample size. Super large studies can be analyzed until smaller subgroups yield significant results (five thirty-eight did a great piece on p hacking (link here). One way to better "prove" that a subgroup had great treatment effects, like in a medicine study, would be to do a pre-analysis plan. One drawback is that it's hard to know in advance how things will break down. Also, if medical biases come into play during the experiment, they'll stay in the data for the analysis. Athley listed some things to remember like always keeping the causal effect in mind. Other things like regularization and creating pseudo-test and pseudo-train data splits can help. These splits are not causal, but they can help machine learning algorithms fit the best model. All these precautions can help take the "con" out of econometrics.