Practice on Causal Inference

Lúa Arconada and Alejandro Macías

2024-01-06

Introduction (A)

The data set contains information about some individuals of a given species of animals. More specifically, it contains three variables: their weight (in pounds), their illness status (sick or not sick), and their chronological age (in years).

Our objective is to provide an estimation of the causal effect of a relevant variable on a suitable response variable using the third variable as possible confounding using the library `Matching`.

Method (M)

We are going to use `Match` with the variable `weight` as the response variable, `sick` as the intervention variable and `age` as the potential confounder. In summary, we are going to explore the question: How does being sick influence an individual's weight?

This setup allows us to explore how being sick influences the weight of the animals while controlling for the potential confounding effect of their age. The basic idea is that if we match treated and untreated by all confounders, the difference could be ascribed to a causal effect due to sickness. Afterwards, we are going to check if matching is appropiately balanced for the confounders.

Results (A^{-1})

First, we carry out the matching process and obtain the following output:

- Estimate: 69.659.

This is the estimated Average Treatment Effect on the Treated (ATT). In the context of our data, it measures the average change in the outcome variable (`weight`) for those sick compared to the control group after matching on the covariate `age`. It indicates that, on average, those who are `sick` have an increase of approximately 69.659 units in the `weigth` variable compared to the control group.

- AI SE (Asymptotic Influence Standard Error): 31.2.

This is the standard error associated with the estimate. In simpler terms, it tells us how much the estimate might vary if we were to repeat the study. It is 31.2, suggesting considerable uncertainty around the estimated treatment effect.

- T-stat (t-statistic): 2.2327.

The t-statistic is calculated by dividing the estimated treatment effect by its standard error. It is used to test the null hypothesis that the true treatment effect is zero. A t-statistic close to zero suggests that the estimated effect is not significantly different from zero. Here, a t-statistic of 2.2327 indicates that the estimated effect is about 2.2327 times the standard error away from zero.

- p.val (p-value): 0.025572.

The p-value is associated with the t-statistic. It indicates the probability of observing a t-statistic as extreme as the one calculated, assuming the null hypothesis (no treatment effect) is true. A small p-value suggests evidence against the null hypothesis. Here, the p-value is 0.025572, meaning that the effect of being sick on the weight of the animals is statistically significant at the conventional 0.05 significance level.

- Original number of observations: 10.

The total number of observations in our original dataset.

- Original number of treated obs: 6.

The number of observations that received the treatment (e.g., 'sick' = 1) in our dataset. In this case, there are 6 observations that are 'sick' in our dataset.

- Matched number of observations: 6.

The number of observations used in the matched sample after propensity score matching.

- Matched number of observations (unweighted): 6.

The number of observations in the matched sample without considering weights. In matching, sometimes weights are assigned to observations, but here, the unweighted number is also provided.

In summary, the estimated effect suggests an increase in the 'weight' variable for those who are sick compared to healthy individuals. The effect caused by sickness is statistically significant at the conventional significance level of 0.05.

And secondly, we check if the matching is balanced, that is, if there exists a sufficient number of treated and untreated (sick and healthy) units in the dataset with similar weights. This is important since it is desirable to have for each sick unit at least a healthy unit of similar age, in order to best infer whether the effect on the weight is due to the sickness status or some other covariate. Through both the naive and bootstrap Kolmogorov-Smirnov tests p-values (0.048 and 0.04), as well as through the T-test p-value (0.026), we can see that the dataset does not present a good balance in terms of units to match according to age, at least at the usual 0.05 significance level. That is, more units would be need for this causal analysis of the effect of sickness status on weight to be more secure.

Bibliography

Lin, L., Sperrin, M., Jenkins, D.A. et al. A scoping review of causal methods enabling predictions under hypothetical interventions. Diagn Progn Res 5, 3 (2021).

Appendix

```
A = Match(df$weight, Tr=df$sick, X=df$age, estimand='ATT', M=1)
summary(A)

##
## Estimate...  69.659
## AI SE......  31.2
## T-stat.....  2.2327
## p.val......  0.025572
##
## Original number of observations..............  10
## Original number of treated obs...............  6
## Matched number of observations...............  6
## Matched number of observations  (unweighted).  6

MB = MatchBalance(df$sick ~ df$age, match.out=A, nboots=500)

##
## ***** (V1) df$age *****
##                         Before Matching        After Matching
## mean treatment........     51.597                 51.597
## mean control..........      32.99                 44.029
## std mean diff.........     270.57                 110.05
##
## mean raw eQQ diff.....     16.619                 8.5608
## med  raw eQQ diff.....     15.634                 9.2663
## max  raw eQQ diff.....     20.958                 14.25
##
## mean eCDF diff........     0.45833                0.38095
## med  eCDF diff........      0.5                   0.33333
## max  eCDF diff........     0.83333                0.83333
##
## var ratio (Tr/Co).....     0.45213                   Inf
## T-test p-value........     0.025567               0.03178
## KS Bootstrap p-value..      0.042                  0.01
## KS Naive p-value......     0.047619               0.015152
## KS Statistic..........      0.83333                0.83333
```