Causal inference HW1

1. Find a real-life example in which the Simpson's paradox arises.

2. With potential outcomes $\{Y_i(1), Y_i(0)\}_{i=1}^n$ for n units under the treatment and control, the difference in means equals the mean of the ITEs:

$$\frac{\sum_{i=1}^n Y_i(1)}{n} - \frac{\sum_{i=1}^n Y_i(0)}{n} = n^{-1} \sum_{i=1}^{n} [Y_i(1) - Y_i(0)]$$

Therefore, the ATE is a linear causal estimand.

Other estimands may not be linear. For instance, we can define the median treatment effect as

$$\delta_1 = \text{median}\{Y_i(1)\}_{i=1}^n - \text{median}\{Y_i(0)\}_{i=1}^n$$

which is, in general, different from the median of the individual treatment effect

$$\delta_2 = \text{median}\{Y_i(1) - Y_i(0)\}_{i=1}^n.$$

a) Give numerical examples which have $\delta_1 = \delta_2$, $\delta_1 > \delta_2$, and $\delta_1 < \delta_2$.
b) Which estimand makes more sense, δ_1 or δ_2? Why? Use examples to justify your conclusion. If you feel that both δ_1 and δ_2 can make sense in different applications, you can also give examples to justify both estimands.