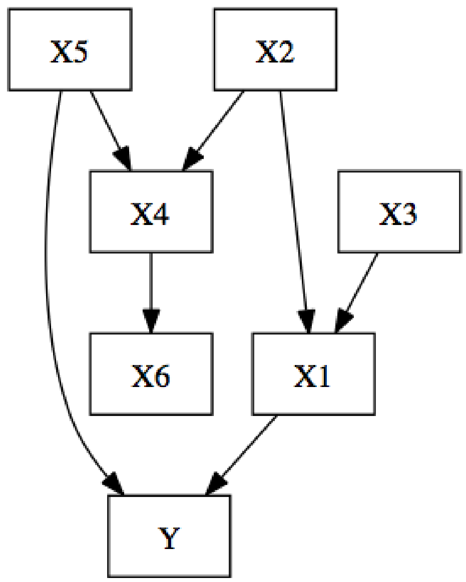
Causal Inference Midterm
Professor Richard Hahn

You may discuss this project with classmates, but all submitted work must be your own, including code.

Consider the following causal directed acyclic graph (CDAG), taken from *Advanced Data Analysis from an Elementary Point of View* by Cosma Shalizi.



The R script `midterm.dgp.R` generates data that follows the causal structure depicted in the graph above. Using the script and the graph, do the following.

1. Add a few additional lines of R code to generate the potential outcomes (Y^0, Y^1) corresponding to $do(x_3 = 1)$ and $do(x_3 = 0)$ (that is, let $d = x_3$ be the treatment variable). Each potential outcome will be a vector of length n, one for each individual.
2. Compute the true average treatment effect as $\mathrm{E}(Y^1 - Y^0)$.
3. Compute the estimated average treatment effect from the observed data using the naive estimator.

4. Do the two numbers match approximately? Why or why not?

5. Compute the true subgroup average treatment effect for those individuals with $x_6 = 1$ by calculating $\mathrm{E}(Y^1 - Y^0 \mid x_6 = 1)$.

6. Compute the estimated average treatment effect for the $x_6 = 1$ subgroup using the observed data.

7. Do the two previous numbers coincide? Why or why not?

8. Compute the true subgroup average treatment effect for those individuals with $x_1 = 1$.

9. Compute the subgroup average treatment effect for those individuals with $x_1 = 1$ using the observed data.

10. Do the previous two numbers coincide? Why or why not?

11. Now, let $d = x_1$ be the treatment variable. As before, write a few lines of R code to simulate the potential outcomes (Y^1, Y^0) for the new treatment variable x_1.

12. Again, compute the true average treatment effect as $\mathrm{E}(Y^1 - Y^0)$.

13. Compute the estimated average treatment effect from the observed data using the naive estimator.

14. Do the two numbers match (approximately) in this case? Why or why not?

15. Compute the true subgroup average treatment effect for those individuals with $x_4 = 1$: $\mathrm{E}(Y^1 - Y^0 \mid X_4 = 1)$.

16. Compute the subgroup average treatment effect for those individuals with $x_4 = 1$ using the observed data.

17. Do the previous two numbers coincide? Why or why not?

18. Next, consider the IV estimator using x_3 as an instrument for x_1 in the case where the estimand of interest is the subgroup average treatment effect for $x_4 = 1$. Is x_3 a valid instrument?

19. Calculate the IV estimate from the observed data. Does it match the true subgroup ATE computed above? Why or why not?

20. Check to see if monotonicity holds for this instrument, meaning that $D_i^1 \geq D_i^0$. To check this, write a few lines of R code to generate the potential outcomes for the treatment variable.

21. Use the generated potential outcomes, in both Y and $D = X_1$, to compute the (subgroup) local average treatment effect (LATE): $\mathrm{E}(Y^1 - Y^0 \mid D^1 - D^0 = 1, X_4 = 1)$. Does this match the IV estimate calculated above?

22. Estimate the subgroup average treatment effect (for $x_4 = 1$ as above) by blocking on x_2 as a control variable. That is, compute separate estimates of the ATE on the $x_2 = 1$ and $x_2 = 0$ data and combine them with weights proportional to the size of the subpopulation.

23. Does this match the true subgroup average treatment effect calculated above? Why or why not?