---
title: "The role of propensity scores in observational study "
authors: " Jennifer Hill, Ray Lu, & Zarni Htet"
output: PDF

YOU MAY WORK IN PAIRS FOR THIS ASSIGNMENT ONLY

Objective
This assignment will give you the opportunity to practice several different propensity score approaches to causal inference. In addition you will be asked to interpret the resulting output and discuss the assumptions necessary for causal inference.

R Packages
You will need to use an R package that you may not already have installed, arm.

Problem Statement
In this assignment will use data from a constructed observational study. The data and an associated data dictionary are available in this folder.
The treatment group for the study that the data are drawn from is the group of children who participated in the IHDP intervention discussed in class. The research question of interest focuses on the effect of the IHDP intervention on age 3 IQ scores for the children that participated in it. The data for the comparison sample of children was pulled from the National Longitudinal Study of Youth during a similar period of time that the data were collected for the IHDP study.

Question 1: Load the data and choose confounders (Step 1)
Load the data. You can use the load command since the data are in a .Rdata file; this will create a data frame called hw4. Choose the covariates you want to use as confounders. To make life easier you many want to choose binary indicators of unordered categorical variables (rather than a variable labeled e.g. as 1, 2, 3 for different levels of a categorical variable).

Create a new data frame for analysis that includes the outcome in the 1st column, the treatment indicator in the 2nd column, and the covariates in the remaining columns. Be thoughtful about your choices with respect to the nature of the covariates (e.g. is an unordered categorical being represented as such) and timing (don't control for post-treatment variables!). Provide your code and a list of the variable names for the confounder variables chosen.

Also reduce that data frame to include only observations for children whose birth weight is less than 3000 grams.

Question 2: Estimate the propensity score (Step 2)

Estimate the propensity score. That is, fit a propensity score model and save the predicted scores.

Question 3: Restructure your data through matching. [Or at least create the weights variable that will let you to do so in the following steps] (Step 3)

(a) The first thing you need to be clear on before restructuring your data is the estimand. Given the description above about the research question, what is the estimand of interest?

(b) First please perform *one-to-one nearest neighbor matching with replacement* using your estimated propensity score from Question 2. Perform this matching using the matching command in the arm package. The "cnts" variable in the output reflects the number of times each control observation was used as a match (the length is equal to the number of control observations). Use the output of this function to create a weight variable that 1) equals one for treated observations and 2) equals the number of times used as a match for non- treated observations.

[Note: If you use matching without replacement in a later context it will not provide the “cnts” variable in the output. Instead you will use the "matched" variable to create the weights variable described above (these are not substitutes for each other – you’ll have to create a different command to use this information).]

Question 4: Check overlap and balance. (Step 4)

(a) Examining Overlap. Check overlap on the raw data (that is without imposing the matched structure) using some diagnostic plots. Check overlap for the propensity scores as well as two other covariates. Note that it may be necessary to exclude some observations from the plots if they are being obscured in ways similar to the example discussed in class on 10/5.

(b) Interpreting Overlap. What do these plots reveal about the overlap required to estimate our estimand of interest?

(c) Examining Balance. You will build your own function to check balance! This function should take as inputs (at least) the data frame created in Question 1, the vector with the covariate names chosen in Question 1, and the weights created in Question 2. It should output the following:

1) Mean in the unmatched treatment group
2) Mean in the unmatched control group
3) Mean in the matched treatment group*
4) Mean in the matched control group
5) Unmatched mean difference (standardized for continuous variables, not standardized for binary variables)
6) Matched mean difference (standardized for continuous variables, not standardized for binary variables)

7) Ratio of standard deviations across unmatched groups (control/treated)
8) Ratio of standard deviations across matched groups (control/treated)

Note that the standardized mean differences should be calculated by dividing by the standard deviation in the inferential group. So, for instance, if you're are making inferences about the treatment group (ATT) then you would divide by the standard deviation in the treatment group.

I provide a "unit test" of this function in (e) to help ensure that you are doing the right thing. Also note that the Hmisc package has a useful weighted variance function.

What to report. Show us your function and the resulting balance, rounded to two decimal places, for this initial match.

* This will only differ from column (1) if you restrict your dataset to observations with common support.

(d) How do you interpret the resulting balance? In particular what are your concerns with regard to covariates that are not well balanced (Write about 5 or 6 sentences).

(e) Unit test. Show the results of your balance function on a simple example with the same sample as above (that is, limited to children with birth weight less than 3000) where the propensity score is fit using logistic regression on "bw" and "b.marr" and the matching is performed using 1-1 nearest neighbor matching with replacement. The output of your balance function should match the following (when rounded to 3 decimal places):

```
            mn1      mn0    mn1.m    mn0.m   diff diff.m ratio ratio.m
bw     2008.648 2629.482 2008.648 2001.838 -2.191  0.024 1.175   1.044
b.marr    0.431    0.595    0.431    0.486 -0.164 -0.055 0.000   0.000
```

Question 5: Repeat steps 2-4 within the matching framework.

What to do: It is rare that your first specification of the propensity score model or choice of matching method is the best. Try at least *3* new approaches. Try to achieve better balance! For continuous variables strive for standardized mean differences less than .1. Try to get ratios of standard deviation closer to 1 than they are for the unmatched data (it may be difficult for some covariates to get the ratio close to 1). For binary variables strive for difference in means (equivalently difference in percentages) less than .05.

Ideas for trying something new in Step 2. You could try a new propensity score specification and then find the corresponding matched sample and calculate balance and overlap. For instance, you could change the inputs to the model (add quadratic

terms, transformed versions of variables, or interactions, or delete predictors) or the model/algorithm used to estimate propensity scores (try probit or GAM or GBM or something else!). Alternately you could try a different matching method. A simple switch would be to switch from matching without replacement to matching with replacement. You could try k-1 matching or caliper matching or optimal matching though this will require using another package such as MatchIt. You could also try eliminating observations from the dataset. Importantly though if you eliminate observations from the group that we are trying to make inferences about you will need to profile those who have been removed. If you remove control observations from the comparison group (for instance those in states not represented by the IHDP observations) you do not need to do this. Save your results (weights and balance) for reporting later.

What to report. For this question write a sentence or two *briefly* describing each of the methods attempted. Include 1) whether any observations were dropped before estimating the propensity score and why, 2) the model used to estimate the propensity score, 3) the type of matching performed.

Question 6: Repeat steps 2-4, but this time using IPTW.
Save your results (weights and balance) -- do not display them here. Make sure that you use weights specific to the effect of the treatment on the treated. In this section report only your code for estimating the propensity scores and your code for creating the IPTW weights.

Question 7: Comparative balance table
Create a table with columns 6 and 8 from your function for each of the matching and weighting methods performed above. Which approach would you choose and why? (1-2 paragraphs at most)

Question 8: Estimate the treatment effect for the restructured datasets implied by Questions 4-6 (Step 5)
Estimate the effect of the treatment on the treated for each of your five approaches by fitting a regression with weights equal to the number of times each observation appears in the matched sample (that is, use your weights variable from above) or using IPTW weights. Report the treatment effect and standard error for each approach.

Question 9: Assumptions
What assumptions are necessary to interpret the estimates from the propensity score approaches causally? List and describe *briefly*.

Question 10: Causal Interpretation
Provide a causal interpretation of *one* of your estimates above. Remember to specify the counterfactual and to be clear about whom you are making inferences. Also make sure to use causal (counterfactual) language.

Question 11: Comparison to linear regression

Fit a regression of your outcomes to the treatment indicator and covariates.
(a) Report your estimate and standard error.
(b) Interpret your results non-causally.
(c) Why might we prefer the results from the propensity score approach to the linear regression results in terms of identifying a causal effect?