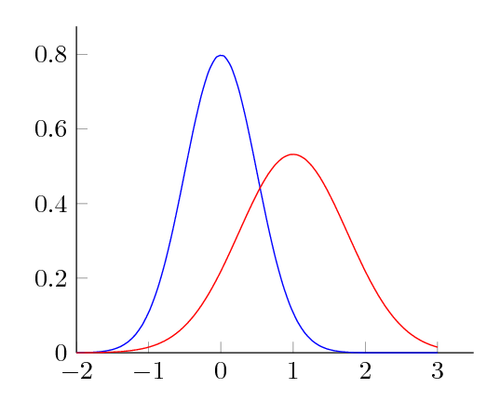
Propensity Score-Based Methods for Causal Inference

Module 6: Creating the Pseudo Population



**I. Module Objectives**

Since the assignment mechanism in observational studies can be affected by many confounders, and there may also be significant differences in the distribution of outcome predictors, the distributions of propensity scores are often very different between treatment groups or exposure levels.

The basic idea of propensity score-based methods is then to create a new “pseudo population” where those populations (i.e. the populations receiving experimental versus standard treatment, or populations exposed versus unexposed) that emulate a randomized trial. Those new pseudo populations can then be analyzed as if they were from a randomized trial and provide a less biased estimate of the treatment effect.

There are several different main approaches for creating the pseudo population, including matching, stratifying, inverse probability weighting, and covariate adjustment. **Matching** on the propensity score (PS) may be 1:1, or m:n, and may use any number of algorithms for identifying matched sets. **Weighting** on the (inverse of the) PS uses the same approach as survey sampling, where one group was over-sampled, and we weight observations by the inverse probability to obtain representative estimates. In the case of PSs, we have over or under-sampling of different PSs within exposure levels, and weight by 1/PS in the exposed (where the PS = the probability of being exposed or treated) and by 1/(1-PS) in the unexposed. In this way, the pseudo population provides a less biased reflection of what the PSs would be in a randomized trial. **Stratifying** on the PS provides subgroups of the data (usually based on quintiles of the PS) which are then assumed to be homogenous in terms of confounders and predictors.

Regardless of the method used, one should then assess the subsequent balance of predictors and confounders in the pseudo population. The next module discusses the ways to analyze these different pseudo populations to estimate the treatment or exposure effect.

By the end of this module, you will be able to:

1. Identify and describe the different types of pseudo populations that can be created from the propensity score
2. Describe possible strengths and limitations of each approach for creating the pseudo population
3. Assess the covariate balance in the post-propensity score adjusted pseudo population

**II. Module Assignments**

**Required Assignments: (~35 minutes of videos + 25 pages to read)**

For an introduction to the propensity score approach and some methods for creating the pseudo population, watch slides 1-14 of Lesson 15 (**~17 minutes**) from the [UC-Davis CER Lessons](https://cpeonline.ucdavis.edu/courses/1874/pages/Lessons).

For a further description of the different approaches for using the propensity score to create a pseudo population (and assess subsequent covariate balance), and an example using matching, watch Module 6a (**~12.5 minutes**) and Module 6b **(~6 minutes**) of the [Category 8 videos (on causal inference) from the PCORI Methodology Standards Academic Curriculum](https://www.pcori.org/research-results/about-our-research/research-methodology/methodology-standards-academic-curriculum-5).

Finally, to review many of the key components of propensity score-based models, and discuss the topics of average treatment effect (ATE) versus average treatment effect on the treated (ATT), read the following article: Austin, P.C., 2011. An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate behavioral research*, *46*(3), pp.399-424. The discussion of ATE versus ATT is first described in paragraphs 2 and 3 under “The Potential Outcomes Framework and Average Treatment Effects” and is continued throughout the article specific to different methods for creating the pseudo population.

**Optional Assignments: (~38 pages to read)**

The following articles provide further mathematical details on using different propensity score-based approaches to form the pseudo population.

1. Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*(1), 41-55.
2. Rosenbaum, P. R., & Rubin, D. B. (1984). Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association*, *79*(387), 516-524.
3. Rosenbaum, P. R., & Rubin, D. B. (1985). The bias due to incomplete matching. *Biometrics*, 103-116.

**III. Project Exercises**

Create a copy of this Google Doc or download the Module onto your computer and review the material offered above under Module Assignments before beginning these workbook exercises.

Thinking about what you learned in this module so far, begin developing the analysis plan for your project by answering the following questions:

1. Selecting your approach for creating the pseudo population depends in part on which effect (the ATE or ATT) you wish to estimate.

Describe whether and why the ATT or ATE is of more relevance for your research question. See the Austin article for more detail and an example of where one is possibly more relevant than the other. If you are more interested in the ATE, you will need to use weighting or stratification (rather than matching).

1. Selecting your approach for creating the pseudo population also partially depends on overlap, or the area of common support, in the propensity score distributions (between the exposed and unexposed).

If there are subjects with a very low probability of exposure or a very high probability of exposure, you may want to exclude them (i.e. exclude outside the area of common support) or employ matching. Weighting and stratification can be problematic when considering subjects with very low or very high probabilities of exposure (i.e. where there is not clinical indecision).

Comment on whether this concern is relevant for your research question and how that might affect your choice for the pseudo population method.

[Link to go back to the Course Overview Document](https://docs.google.com/document/d/1UDTkp3rbhqdun7jvSvktaZmTtoUWOz_VUDQw3HIsElg/edit?usp=sharing)