PROPENSITY SCORES

THE PROBLEM

* Non-randomized studies are being used more frequently to estimate treatment effects.
* Because treatment assignment is not randomized, there may be systematic differences in baseline characteristics that influence treatment effect.

SOLUTION: PROPENSITY SCORES

* Measure the probability of subject being assigned to treatment given the subject’s observed baseline characteristics.
* Scores are generally estimated using logistic regression.
* Once estimated, propensity scores can be incorporated into study design to account for systematic differences between treatment groups at baseline.

MOTIVATING EXAMPLE (*R* ANALYSIS)

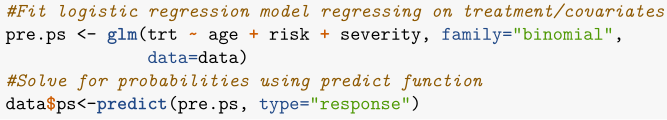
* Data retrieved from:<http://web.hku.hk/~bcowling/data/propensity.csv>
* Data contains 400 observations for male heart attack patients, some treated with a new drug (trt=1) and others receiving standard care (trt=0). Outcome variable is 30-day mortality (death=1) and covariates of interest are age, admission severity, and risk score.

METHOD



1. Select Baseline Covariates
   * Need to choose covariates to use when estimating propensity scores.
   * Only include variables measured at baseline (post-baseline covariates might be influenced by treatment).
   * Best to use variables associated with the outcome or variables associated with both treatment assignment and the outcome
   * EXAMPLE: Age, admission severity, and risk score are all likely associated with mortality.
   * Unknown if these variables are associated with treatment assignment.
2. Estimate Propensity Scores
   * Estimated propensity score *e*for subject *i* where *Z* is an indicator variable for treatment assignment (i.e. Z=1: treatment, Z=0: control) and **X** is a matrix of selected baseline covariates:
   * Typically estimate propensity score using logistic regression with the log odds of *e­i* as the outcome and the baseline characteristics as covariates.
   * Can also use classification and regression trees, random forests, and other machine learning or data mining techniques instead of logistic regression.

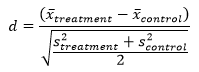
EXAMPLE: Estimating propensity score R code:



1. Apply Propensity Score Method
   * Various methods for incorporating propensity scores into study design including the four detailed below.

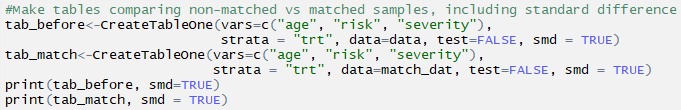
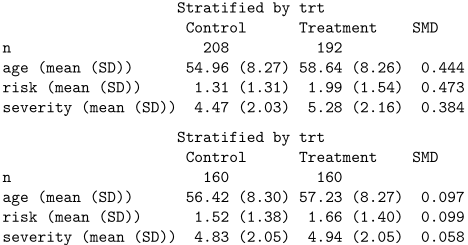
|  |  |
| --- | --- |
| **Method** | **Description** |
| Matching | * Usually match treated and control subjects into pairs by propensity score to create new sample. * Matching can be done with or without replacement, using greedy or optimal methods, or with nearest neighbor caliper matching. * Treatment effect estimated by comparing outcome between groups in matched sample. * EXAMPLE: Creating matched sample in *R*: |
| Stratification | * Subjects split into groups according to propensity scores (commonly five equal groups). * Treatment effect estimated for each strata and then pooled for overall treatment effect. * EXAMPLE: Creating stratified groups in *R:* |
| Inverse Probability of Treatment Weight (IPTW) | * Calculate weights for each subject equal to the inverse probability of receiving the treatment actually received. * Treatment effect estimated by comparing average weighted outcomes between treated group and control group. * EXAMPLE: Creating weights for each subject |
| Covariate Adjustment | * Outcome is regressed on propensity score and treatment indicator variable using linear regression (continuous outcome) or logistic regression (binary outcome). * Treatment effect is estimated from model coefficient of the treatment indicator variable. * EXAMPLE: Logistic regression adjusting for propensity score: |

1. Check Balance Diagnostics
   * To ensure that the propensity score model was specified adequately, need to make sure that there aren’t systematic differences in covariates when conditioning on propensity score.
   * For matching, stratification, and IPTW, check balance diagnostics between treatment groups for matched sample, individual strata or weighted sample.
   * There are different methods of assessing goodness of fit for propensity score adjusted regression models. See Austin (2008) for more details.
   * Can assess using the standardized difference of covariate:

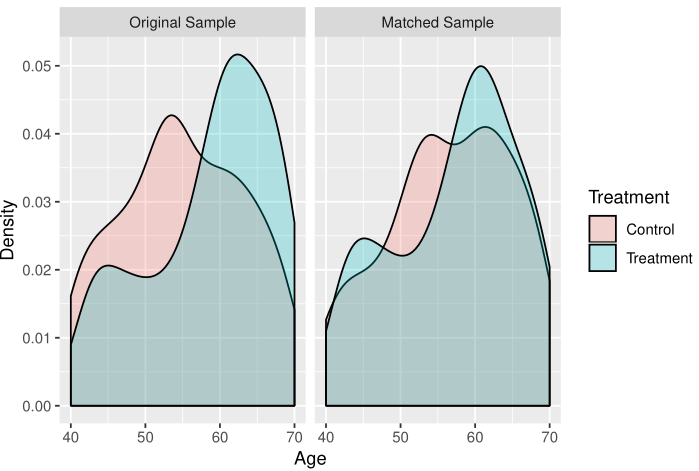
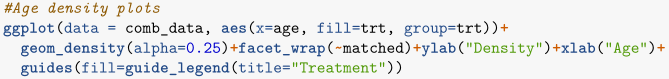


Where represents are the sample means for the covariate in each group and are the variances of the covariates for each group (shoot for d<0.1).

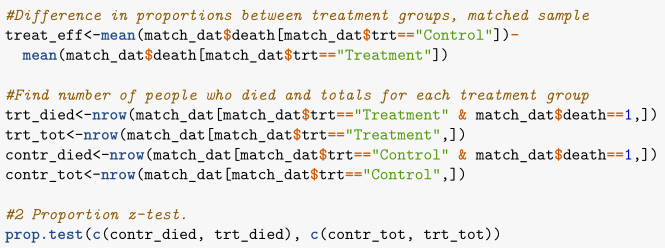
* Can also assess balance by comparing plots of the covariariates for each treatment group..
* EXAMPLE: Balance Diagnostics for matched sample



* Standardized mean differences from resulting table all <0.1, but age and risk are close.
* Plot age distribution:



1. Estimate Treatment Effects
   * See table for details on finding treatment effect for each method.
   * Additional significance tests and confidence intervals can be calculated using standard statistical methods.
   * There is controversy over if matched data should be treated as independent or not (see Schafer & Kang, 2008, Austin, 2011b).
   * EXAMPLE: Treatment effect analysis for matched sample:



* Using our matched sample, we find that the proportion who survived was 7% higher in the untreated than the treated group, but the confidence interval is (-0.01, 0.15) so evidence that treatment works is inconclusive.

ALTERNATE METHOD: REGRESSION COVARIATE ADJUSTMENT

* Could use regression and adjust for baseline covariates in the model.
* Several limitations to this approach when compared with propensity scores:
  + Harder to check goodness of fit diagnostics for regression model than to check propensity score balance diagnostics.
  + With regression the design of the study and analysis of treatment effect are integrated leading to possible bias in design. It is also less effective when the outcome under study is rare or the treatment is common.

LIMITATIONS IN PROPENSITY SCORE METHODS

* Doesn’t always work well with small samples.
* If one covariate is missing, propensity score will also be missing.
* Only controls for measured covariates.

References

Austin, P. C. (2008). Goodness-of-fit diagnostics for the propensity score model when

estimating treatment effects using covariate adjustment with the propensity score. *Pharmacoepidemiol Drug Saf, 17*(12), 1202-1217.

doi:10.1002/pds.1673

Austin, P. C. (2011a). An Introduction to Propensity Score Methods for Reducing the

Effects of Confounding in Observational Studies. *Multivariate Behav Res, 46*(3), 399-424. Retrieved from<https://www.ncbi.nlm.nih.gov/pubmed/21818162>

doi:10.1080/00273171.2011.568786

Austin, P. C. (2011b). A Tutorial and Case Study in Propensity Score Analysis: An

Application to Estimating the Effect of In-Hospital Smoking Cessation Counseling on Mortality. *Multivariate Behav Res, 46*(1), 119-151. Retrieved from<https://www.ncbi.nlm.nih.gov/pubmed/22287812>

doi:10.1080/00273171.2011.540480

Austin, P. C. (2014). A comparison of 12 algorithms for matching on the propensity score.

*Stat Med, 33*(6), 1057-1069. Retrieved from<https://www.ncbi.nlm.nih.gov/pubmed/24123228>

doi:10.1002/sim.6004

Braitman, L. E., & Rosenbaum, P. R. (2002). Rare outcomes, common treatments: Analytic

strategies using propensity scores. *Annals of Internal Medicine, 137*(8), 693-695. doi:10.7326/0003-4819-137-8-200210150-00015

Caliendo, M., & Kopeinig, S. (2008). Some Practical Guidance for the Implementation of

Propensity Score Matching. *Journal of Economic Surveys, 22*(1), 31-72. Retrieved from<https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-6419.2007.00527.x>. doi:10.1111/j.1467-6419.2007.00527.x

Cowling, B. (April 19, 2017). Propensity Score Analysis. Retrieved from

<http://web.hku.hk/~bcowling/examples/propensity.htm>

d'Agostino, R. B. (1998). Propensity score methods for bias reduction in the comparison

of a treatment to a non‐randomized control group. *Statistics in medicine, 17*(19), 2265-2281.

Gu, X. S., & Rosenbaum, P. R. (1993). Comparison of Multivariate Matching Methods:

Structures, Distances, and Algorithms. *Journal of Computational and Graphical Statistics, 2*(4), 405-420. Retrieved from<https://amstat.tandfonline.com/doi/abs/10.1080/10618600.1993.10474623>

doi:10.1080/10618600.1993.10474623

Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2011). MatchIt: Nonparametric Preprocessing for

Parametric Causal Inference. *Journal of Statistical Software, 42*(8).

doi:<https://doi.org/10.18637/jss.v042.i08>

Lee, B. K., Lessler, J., & Stuart, E. A. (2010). Improving propensity score weighting using

machine learning. *Stat Med, 29*(3), 337-346. Retrieved from<https://www.ncbi.nlm.nih.gov/pubmed/19960510>

doi:10.1002/sim.3782

Olmos, A., & Govindasamy, P. (2015). Propensity Scores: A Practical Introduction Using R.

*Journal of MultiDisciplinary Evaluation, 11*(25), 68-88.

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. doi:10.1093/biomet/70.1.41

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). Retrieved from<https://www.tandfonline.com/doi/abs/10.1080/01621459.1984.10478078>. doi:10.1080/01621459.1984.10478078

Schafer, J. L., & Kang, J. (2008). Average causal effects from nonrandomized studies: a

practical guide and simulated example. *Psychol Methods, 13*(4), 279-313. Retrieved from<https://www.ncbi.nlm.nih.gov/pubmed/19071996>.

doi:10.1037/a0014268

Setoguchi, S., Schneeweiss, S., Brookhart, M. A., Glynn, R. J., & Cook, E. F. (2008).

Evaluating uses of data mining techniques in propensity score estimation: a simulation study. *Pharmacoepidemiol Drug Saf, 17*(6), 546-555. Retrieved from<https://www.ncbi.nlm.nih.gov/pubmed/18311848>

doi:10.1002/pds.1555