INTERNATIONAL JOURNAL OF STATISTICS AND MEDICAL INFORMATICS

Vol. 12 No.1/2019, pp: 1-6

Application of Propensity Score – an overview with the help of r software package

1. Introduction

Experimental studies or randomized clinical trail in health care setting are usually the preferred type of research when we want to compare to treatments or two groups or test procedures. When the experimental or clinical trials are not feasible or unethical to conduct then observational studies are preferred over the experimental studies to compare the two groups out of which one is received the treatment under consideration and another one is not the received the treatment or placebo group. One of the main drawbacks of the observational studies is that the experimental units are not randomized among the two groups which results in imbalance in the base line covariates which ultimately affect the outcome under consideration which is not the case in experimental or clinical trials. Propensity score[1,2] helps us to overcome the imbalance in the two groups by assigning propensity scores to treatment groups and non-treatment groups and the members of the two groups with equal propensity scores are matched which makes the two groups comparable with respect to the covariates. This paper provides an overview of propensity score, its application and computing the propensity score with the use of R statistical package which is an open source statistical package.

Keywords: Propensity Score, matching, observational studies, r package

2. Propensity Score

Propensity score is computed as a conditional probability of experiment unit is assigned to treatment group given the covariates. The covariates from the study are selected for calculating the propensity score which have an effect on the outcome. The covariate selection can be done by the referring to the similar studies [3].

After calculating the propensity score, matching of experimental units in the two groups with respect to the propensity score can be done by the following algorithms [1, 4,5]:

a. Nearest Neighbor algorithm
b. Manhattan algorithm
c. Caliper matching algorithm

Vol. 12 No.1/2019, pp: 1-6

After matching the two groups, the outcome measure is computed to find out the treatment effect [6,7,8].

3. Advantages of Propensity score matching [9,10,11,12**]**

1. To helps to balance the treatment and non-treatment group and to study treatment effect on the outcome
2. Reduces selection or allocation bias
3. Reduces bias in the outcome estimation

4. Disadvantages of Propensity score matching [1,2**]**

1. It considers only observed covariates while computing the propensity score

5. Propensity Score computation example

Let us consider a hypothetical observational study which aims to study to observe the effect of treatment (patients who took aspirin) and non-treatment group (who have not taken aspirin) on the outcome of heart attack. Here the two groups of patients are not randomly assigned and hence the outcome heart attach might be affected by the covariates such as age, gender, food habits, diabetic status, hypertension status, smoking and alcohol status.

The propensity score is computing by calculated by conditional probability of patient being in the aspirin group given the above covariates like age, gender etc.

After matching the two groups of patients, the outcome measure is computed to find out the effect of Aspirin on heart attack.

The data set used in the study can be downloaded from the site: **www.ijsmi.com/book.php**

6. R software package for computing Propensity score

MatchIt R software[13] can be used to compute the Propensity score. The following code needs to be executed directly in the environment or in the R Studio[14] environment which is Integrated Development Environment for R package.

INTERNATIONAL JOURNAL OF STATISTICS AND MEDICAL INFORMATICS

Vol. 12 No.1/2019, pp: 1-6

R code for computing Propensity Score

```
#install the required package using install.package statement
Install.packages("MatchIt")
#call the MatchIt package into R studio environment  using library statement
library(MatchIt)
#call the readxl  package using library statement to import excel data sheet
library(readxl)
#Data clinical1.xlsx can be downloaded from website www.ijsmi.com/book.php
clinical1 <- read_excel("C:/test/clincial1.xlsx")
#we use the following covariates age, diet, bilirubin and creatine for the test purpose
ps = matchit(group ~ age+ diet +  bilirubin + creatinine,   data = clinical1, method = "nearest")
# we generate the output using summary() and plot() statements
summary(ps)
plot(ps)
newclinial1 <- match.data(ps)
```

7. Results

The following output is obtained using the above code

Table-1 : Summary of Balance for all the data

	Means Treated	Means Control	SD Control	Mean Diff	eQQ Med	eQQ Mean	eQQ Max
distance	0.5642	0.4077	0.1681	0.1564	0.1656	0.1651	0.2516
age	42.9655	36.129	10.5916	6.8365	8	7.4483	12
diet	1.6552	1.5161	0.508	0.139	0	0.1379	1
bilirubin	2.8931	2.3871	1.3423	0.506	0.6	0.6379	1.5
creatinine	1.0138	0.9161	0.582	0.0977	0.1	0.1483	0.5

Table-2: Summary of balance for matched data

	Means Treated	Means Control	SD Control	Mean Diff	eQQ Med	eQQ Mean	eQQ Max
distance	0.5642	0.4256	0.1586	0.1386	0.1432	0.1394	0.2195
age	42.9655	36.6552	10.6176	6.3103	8	6.4483	11
diet	1.6552	1.5517	0.5061	0.1034	0	0.1034	1
bilirubin	2.8931	2.4241	1.3179	0.469	0.5	0.5655	1.4
creatinine	1.0138	0.9552	0.5804	0.0586	0.1	0.1276	0.4

Vol. 12 No.1/2019, pp: 1-6

Table-3: Percent Balance Improvement

	Mean Diff.	eQQ Med	eQQ Mean	eQQ Max
distance	11.4298	13.5016	15.5269	12.773
age	7.6961	0	13.4259	8.3333
diet	25.6	0	25	0
bilirubin	7.3203	16.6667	11.3514	6.6667
creatinine	39.9772	0	13.9535	20

Sample sizes:

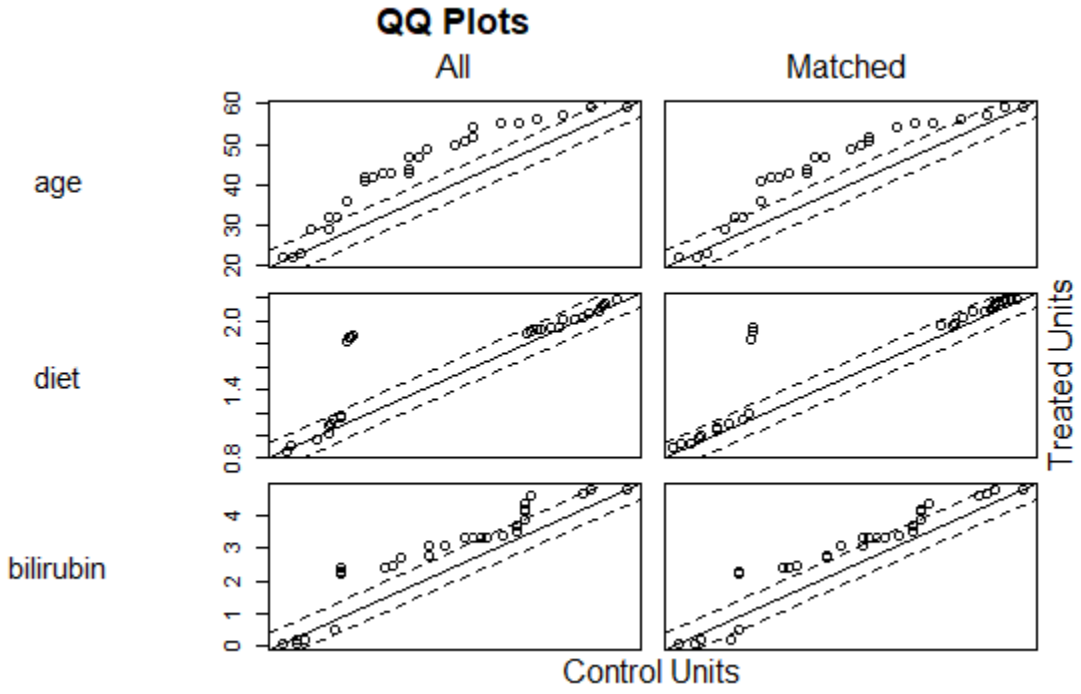
	Control	Treated
All	31	29
Matched	29	29
Unmatched	2	0
Discarded	0	0

Here we can see that only 2 patients are matched and no patient is discarded with respect to the covariates and the balance between the two experiment groups is introduced. The mean difference between the treatment group and non-treatment group (control group) was higher before the balancing and after balancing it was improved as shown in the Table-1, Table-2 and Table-3.

For example the mean difference between the treatment group and control group was initially 6.83 and after balancing it was reduced to 6.31 and the percentage improvement was 7.69.

The following plot is obtained from the above output

Vol. 12 No.1/2019, pp: 1-6



The balanced data set can be used for further analysis using match.data statement

Conclusion

The paper provided an overview of provided an overview of propensity score with the help of R software package using hypothetical example. The propensity matching introduced balance between the treatment and the control group.

Vol. 12 No.1/2019, pp: 1-6

References

1. 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.
2. Guo, S., & Fraser, M. W. (2015). Propensity score analysis. Sage.
3. Brookhart, M. A., Schneeweiss, S., Rothman, K. J., Glynn, R. J., Avorn, J., & Stürmer, T. (2006). Variable selection for propensity score models. *American journal of epidemiology*, *163*(12), 1149-1156.
4. Austin, P. C. (2014). A comparison of 12 algorithms for matching on the propensity score. Statistics in medicine, 33(6), 1057-1069.
5. Abadie, A., & Imbens, G. W. (2016). Matching on the estimated propensity score. Econometrica, 84(2), 781-807.
6. Imbens, G. W. (2000). The role of the propensity score in estimating dose-response functions. Biometrika, 87(3), 706-710.
7. Lapointe-Shaw, L., Bell, C., Austin, P., Abrahamyan, L., Ivers, N. M., Li, P., ... & Dolovich, L. (2019). Community pharmacy medication review, death and re-admission after hospital discharge: a propensity score-matched cohort study. BMJ Quality & Safety, bmjqs-2019.
8. Austin, P. C., & Fine, J. P. (2019). Propensity-score matching with competing risks in survival analysis. Statistics in medicine, 38(5), 751-777.
9. Caliendo, M., & Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. Journal of economic surveys, 22(1), 31-72.
10. Hirano, K., Imbens, G. W., & Ridder, G. (2003). Efficient estimation of average treatment effects using the estimated propensity score. Econometrica, 71(4), 1161-1189.
11. Austin, P. C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. Multivariate behavioral research, 46(3), 399-424.
12. D'Agostino Jr, 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.
13. https://cran.r-project.org/
14. https://www.rstudio.com/