Cumhuriyet Science Journal

e-ISSN: 2587-246X
ISSN: 2587-2680

Cumhuriyet Sci. J., 41(2) (2020) 369-376
http://dx.doi.org/10.17776/csj.677510

Comparison of greedy matching methods on cigarette usage of individuals in Turkey

Ezgi NAZMAN[1,*] **Hülya OLMUŞ**[2]

[1]Sivas Cumhuriyet University, *Department of Statistics and Computer Sciences, Sivas/ TURKEY*

[2]Gazi University, *Department of Statistics, Ankara/ TURKEY*

Abstract

There has been a growing interest in using observational studies to estimate treatment effects on outcomes where treatment selection is often influenced by covariates. Recently, propensity score matching (PSM) method has increasingly being used to reduce bias in estimated treatment effect for observational studies. Greedy Matching (GM), one of the PSM methods, is widely preferred in many studies because of the calculation simplicity of the method. However, GM is still open to be evaluated in terms of bias reduction and classification performances. For this purpose, data including cigarette usage of 17242 individuals in Turkey were used for the comparison of nearest neighbor, caliper, stratification, Mahalanobis metric, and combined propensity score and Mahalanobis metric matching methods in terms of average standardized bias, bias reduction, and accuracy rate. The stratification-matching method should be preferred for not only low standardized bias and high bias reduction, but also high accuracy rate.

Article info

History:
Received: 20.01.2020
Accepted: 28.05.2020
Keywords:
Bias reduction, Greedy matching methods, Standardized bias, Cigarette usage

1. Introduction

The aim of randomized study is to eliminate the effects of confounding and treatment selection bias when using observational data ensuring groups are comparable. However, observational studies cause treatment selection bias. Therefore, propensity score matching (PSM) is a statistical approach to reduce bias in estimated treatment effects. Opposed to randomized study, a subject's probability of receiving a treatment is not known; however, it depends on observed and unobserved covariates in nonrandomized study. The propensity score was first proposed by [1] as a conditional probability of receiving a treatment given that subject's observed covariates [2]. The studies including large data sets are generally observational rather than experimental. Thus, PSM is sensible and more reliable to try to estimate the effects of treatments from such large data sets [3].Adjustment was indicated based on PSM which can change results if age, sex, education, and income are used as predictors on smoking status [4].

Greedy matching (GM) methods are the most commonly used PSM methods for observational studies due to the application easiness of the methods. Propensity score calculations on cigarette using have been utilized in individual health from different point of views in literature for the observational studies. The effect of in-patient smoking cessation counselling on mortality in patients hospitalized with a heart attack was estimated by [5]. Association between smoking cessation and change in mental health before and after PSM was examined by [6]. The effect of maternal smoking during pregnancy on offspring's initial reactions to cigarettes and alcohol usage was studied by [7]. Multiple potential risk factors for non-smoking related lung cancer among Asian ethnic groups using PSM was investigated by [8]. Smoking's independent contribution to the risk of short-term complication after total joint arthroplasty was studied by [9]. Association of cigarette smoking to cerebral atherosclerosis along with other risk factors was studied by [10].

This study first aims to assess the association of cigarette usage with gender, age group, education level, marital status, employment, income level, chronic disease, Body Mass Index (BMI), tobacco exposure,

---

*Corresponding author. *Email address: ezgicabuk@cumhuriyet.edu.tr.*

and alcohol usage in Turkey. In addition, the study compares performances of the greedy matching methods in terms of average standardized bias, bias reduction, and accuracy rate. It is the largest study on cigarette usage of individuals in Turkey up to now that propensity score matching has been applied to reduce bias of the covariates using GM methods.

2. Propensity Matching Method

The propensity score is defined as the conditional probability of assigning a subject to a particular treatment situation when the observed covariates are considered [1, 5]. The propensity score e(x_i) is given in Eq. (1) which can be obtained by using the probit model, discriminant and cluster analysis in addition to the logistic regression model [11].

$$e(x_i) = P(z_i = 1|x_i), \qquad 0 < e(x_i) < 1 \tag{1}$$

where z_i is the treatment (1) and control (0) groups, respectively and x_i is the observed covariates vector for ith subject. The propensity score when more than two treatment groups $(z_1, \dots, z_n)$ exist for covariates vector $(x_1, \dots, x_n)$ as in Eq.(2).

$$P(z_1, \dots, z_n|x_1, \dots, x_n) = \prod_{i=1}^{N} e(x_i)^{z_i}\big(1 - e(x_i)\big)^{1-z_i} \tag{2}$$

Binary logistic regression is the most commonly used method to estimate the propensity score. The dependent variable is the probability of receiving a particular treatment using logit link function considering all covariates in the propensity score model as independent variables. Then, a propensity score for each subject in the study can be calculated by using the fitted model to estimate the probability of receiving the treatment given that subject's covariates. Once a propensity score for each subject has been estimated, subjects are matched using the propensity scores to create a balanced sample. Logistic regression model considering logit link function is given in Eq. (3).

$$ln\left(\frac{e(x_i)}{1-e(x_i)}\right)= ln\left(\frac{P(z_i=1|x_i)}{1-P(z_i=1|x_i)}\right) = \beta_0 + \beta_1 x_1 + \cdots + \beta_N x_N \tag{3}$$

The propensity scores of ith subject can be calculated by using as in Eq. (4):

$$e(x_i) = \frac{e^{\beta_0+\beta_1 x_1+\cdots+\beta_N x_N}}{1+e^{\beta_0+\beta_1 x_1+\cdots+\beta_N x_N}} = \frac{1}{1+e^{-(\beta_0+\beta_1 x_1+\cdots+\beta_N x_N)}} \tag{4}$$

2.1. Greedy matching methods

In this section, nearest neighbor, caliper, stratification, Mahalanobis, and combined propensity score and Mahalanobis metric matching methods are explained.

2.1.1. Nearest neighbor matching

The control and treatment units are randomly ordered in this method. Then, the smallest absolute value of the difference between the propensity score of the first treatment unit and the control unit are matched [12]. This process is continuous until all units in the treatment group have been generally matched as $1{:}N$, $N{:}1$ and $N{:}N$. Nearest neighbor matching was considered as 1:1 in this study. Having more units in the control group than treatment group allows better predictions [13]. In general, matching is described as follow:

$$C(P_i) = min\left|P_i - P_j\right| \tag{5}$$

where $C(P_i)$ is the smallest absolute difference between ith treatment group unit matching the jth control group unit. P_i and P_j are the calculated propensity scores for ith treatment and jth control groups, respectively.

2.1.2. Caliper matching

In this method, units of the control and treatment groups are matched according to the absolute value of the difference between propensity scores and a standard error between predetermined interval values which is often 0.25 as first quantile. Then, the values outside of this range are removed. The fewer the standard error is, the fewer matching units will be [14]. Caliper matching method is as follow:

$$|P_i - P_j| < e \tag{6}$$

where e is the predetermined standard error.

2.1.3. Stratification matching

This method separates the units of treatment and control groups according to their propensity scores and calculates the effect within each interval by taking into account the mean difference between the groups. Within each propensity score stratum, treatment and control units will have roughly similar values of the propensity score [1]. One of the most important problems with this method is to specify number of strata. [15] showed that five strata are often sufficient since it is often enough to reduce 95% of the bias associated with one single covariate.

2.1.4. Mahalanobis metric matching

This method is based on the Mahalanobis metric distance which calculates a multi-dimensional space distances [16]. Mahalanobis metric distance between the treatment and control group units is calculated by starting from the first treatment group unit. Mahalanobis metric distance is shown in Eq. (7):

$$D_{ij} = (x_i - x_j)^T S^{-1}(x_i - x_j) \tag{7}$$

where S^{-1} is the inverse of the covariance matrix for the control units. x_i and x_j are covariate matrices of ith treatment and jth control group subjects. The treatment and control units are matched according to the smallest Mahalanobis metric distance. This process continues until the treatment group matches and unmatched control units are removed from the study [17].

2.1.5. Combined propensity score and Mahalanobis metric matching (Mah-Ps)

This method combines the propensity score and Mahalanobis metric distance for matching [17]. After propensity scores are calculated for all units, these scores are added to the data. Then, units in the treatment group are randomly ordered and Mahalanobis metric distances of the control and treatment groups are calculated by using the combination of covariates and propensity scores [16]. Matching is conducted according to the calculated distance of new Mahalanobis metric.

2.2. Evaluation of matching methods

Inferences from bias reduction (BR) performance of GM methods are calculated to give information on the degree of bias. Percent standardized bias (SB_k) comparing the distribution of each of the variables is defined as in Eq.(8) *k*th continuous variable.

$$SB_k = \frac{\bar{x}_i - \bar{x}_j}{\sqrt{\frac{s_i^2 + s_j^2}{2}}} \times 100 \tag{8}$$

where $\bar{x}_i$ and $\bar{x}_j$ are the sample means in the treatment and control groups, respectively, and s_i^2 and s_j^2 are the corresponding sample variances. In addition, percent SB_k is defined as in Eq.(9) for each dichotomous variable.

$$SB_k = \frac{\hat{p}_i - \hat{p}_j}{\sqrt{\frac{\hat{p}_i(1-\hat{p}_i)+\hat{p}_j(1-\hat{p}_j)}{2}}} \times 100 \tag{9}$$

where $\hat{p}_i$ and $\hat{p}_j$ denote the prevalence of the dichotomous variable in the treatment and control groups, respectively. [18] mentioned SB_k with more than 10% of a covariate is substantial. The percent BR_k on the *k*th covariate to evaluate the effectiveness of matching was also calculated as in Eq.(10):

$$BR_k = \frac{|B_{k,beforematching}| - |B_{k,aftermatching}|}{|B_{k,beforematching}|} \times 100 \tag{10}$$

where $B_k = \bar{x}_i - \bar{x}_j$. The percent *BR* value with more than 80% is acceptable to evaluate effectiveness of GM methods on the covariates [19].

3. Case Study

Turkey Health Survey 2016 administered by TurkStat [20] many indicators on health which have been periodically conducted by TurkStat since 2008 for every two years. Thus, this data set was obtained from TurkStat by special permission. Health Survey 2016 was carried out with 23606 individuals aged 75 and younger. The study was restricted on cigarette using of individuals; thus, the children, who are aged between 0-14, was out of the study. The remained sample included 17242 individuals ignoring children. Gender, age group, education level, marital status, employment, income level, chronic disease, BMI, tobacco exposure, and alcohol usage were specified as the covariates that may related with cigarette using. MASS, Matching, MatchIt, optmatch packages in RStudio were used for data analyses. Individuals are specified as control and treatment groups, respectively as no (non-smokers including never and quit smoking individuals) and yes (smokers including every day and sometimes smoking individuals). Baseline characteristics before matching of the two groups were given in Table1. Categorical variables were explained with the frequencies and percentages whereas BMI was given with mean and standard deviation.

Table 1: Baseline characteristics between control and treatment groups before matching

Variables	Cigarette using		*p*-value	$\lvert SB_k \rvert$(%)
	No (*N*=12167)	Yes (N=5075)		
Female	7797 (64.0)	1777 (35.0)	0.001*	60.8
≥45 year	5964 (49.0)	1923 (37.9)	0.001*	22.6
Higher Education	1749 (14.4)	815 (16.1)	0.001*	4.7
Married	8255 (67.9)	3657 (72.1)	0.001*	9.2
Working	3667 (30.1)	2790 (55.0)	0.001*	51.9
≥2541TL	3953 (32.5)	1759 (34.7)	0.536	4.6
Chronic Disease	6165 (50.7)	2294 (45.2)	0.018*	11.0
Tobacco Exposure	812 (6.7)	1095 (21.6)	0.001*	43.8
Alcohol	2008 (16.5)	2209 (43.7)	0.001*	62.1
BMI	26.7 ± 5.3	25.8 ± 4.7	0.001*	18.0

$|SB_k|$ is the absolute average standardized bias of *k*-th covariate. Paranthes indicates percentage of related baseline characteristic. *shows the significant covariate comparing with p-value=0.05.

As can be seen in Table1, 12167 (71%) individuals are not smokers including 64% female, 49% older than 45 years, 30.1% working, 50.7% having chronic disease, 43.8% being exposed tobacco, and 16.5 % using alcohol. The percent $|SB_k|$ values are also higher than 10% for gender, age, working status, chronic disease, tobacco exposure, alcohol usage, and BMI. According to the binary logistic regression analysis results, only income level is not significant before matching (p-value =0.536).

The baseline characteristics between the control and treatment groups after matching are given in Table 2. According to the results, the percent $|SB_k|$ decreases after matching for all methods. However, nearest (1:1), Mahalanobis, and Mah-Ps matchings cause more than 10% $|SB_k|$ for some of the covariates. On the other hand, caliper and stratified matchings provide $|SB_k|$ less than 10%. Caliper matching method provides the least $|SB_k|$ for gender, marital status, chronic disease, tobacco exposure, and BMI whereas stratified matching generally method reveals the least $|SB_k|$ for age group, working status, alcohol usage. Meantime, Mah-Ps has the least $|SB_k|$ for income level. Stratified and Mahalanobis matching are the lowest for education level with the same $|SB_k|$.

In addition, the percent $|SB|$ results with the average percent *BR* is shown Table 3. The average standardized bias before matching is 28.9%. The average standardized bias of all matching methods less than 10% whereas caliper (3.2%) and stratified (4.0%) matching methods have the least values. According to the percent *BR*, caliper and stratified matching methods have the highest value (81.0% and 76.8%) among the other methods. On the other hand, Nearest matching have the lowest average percent BR (56.5%) with 7.8% average percent *SB.*

The data set was splitted as 70% training and 30% test data to evaluate classification performances of the matching methods. Confusion matrix of binary classification is a two by two table formed by counting of the number of the true positive, false positive, true negative, and false negative of a binary classification method. The most common validation measurements are accuracy, sensitivity, and specificity deriving from the confusion matrix. Accuracy indicates the correctly classification rate of the methods, while sensitivity and specificity assess the accuracy rate. The method can be very specific without being sensitive, or it can be very sensitive without being specific. The accuracy, sensitivity, and specificity rates of the methods are shown in Table 4.

As can be seen, accuracy rate of before matching is almost similar with the rate of stratified matching. Caliper matching has the least accuracy rate among matching methods. Meantime, sensitivity and specificity rates of before matching and stratified matching are high among the other method.

Table 2: Baseline characteristics between control and treatment groups after matching

Methods	Nearest Matching (1:1)			Caliper Matching			Stratified Matching			Mahalanobis Matching			Mah-Ps Matching		
Variables	Cigarette using		$\lvert SB_k \rvert$ (%)	Cigarette using		$\lvert SB_k \rvert$ (%)	Cigarette using		$\lvert SB_k \rvert$ (%)	Cigarette using		$\lvert SB_k \rvert$ (%)	Cigarette using		$\lvert SB_k \rvert$ (%)
	No (N=5075)	Yes (N=5075)		No (N=4436)	Yes (N=4436)		No (N=12164)	Yes (N=5075)		No (N=5075)	Yes (N=5075)		No (N=5075)	Yes (N=5075)	
Female	1984 (39.1)	1777 (35.0)	8.5	1808 (40.8)	1733 (39.1)	3.5	7794 (64.0)	1777 (35.0)	3.7	2207 (43.5)	1777 (35.0)	17.4	2201 (43.4)	1777 (35.0)	17.2
≥45 year	2078 (40.9)	1923 (37.9)	6.3	1810 (40.8)	1741 (39.2)	3.2	5961 (49.0)	1923 (37.9)	1.0	2044 (40.3)	1923 (37.9)	4.9	2008 (39.6)	1923 (37.9)	3.4
Higher Education	938 (18.5)	815 (16.1)	6.4	789 (17.8)	742 (16.7)	2.8	1749 (14.4)	815 (16.1)	0.1	817 (16.1)	815 (16.1)	0.1	822 (16.2)	815 (16.1)	0.4
Married	3704 (73.3)	3657 (72.1)	2.1	3159 (71.2)	3178 (71.6)	1.0	8254 (67.9)	3657 (72.1)	1.9	3628 (71.5)	3657 (72.1)	1.3	3621 (71.3)	3657 (72.1)	1.6
Working	2541 (50.1)	2790 (55.0)	9.8	2173 (49.0)	2250 (50.7)	3.5	3667 (30.2)	2790 (55.0)	1.5	2517 (49.6)	2790 (55.0)	10.8	2427 (47.8)	2790 (55.0)	14.4
≥2541TL	1852 (36.5)	1759 (34.7)	3.8	1590 (35.8)	1549 (34.9)	1.9	3952 (32.5)	1759 (34.7)	3.3	1792 (35.3)	1759 (34.7)	1.4	1791 (35.3)	1759 (34.7)	1.3
Chronic Disease	2346 (46.2)	2294 (45.2)	2.1	2028 (45.7)	2031 (45.8)	0.1	6162 (50.7)	2294 (45.2)	1.4	2304 (45.4)	2294 (45.2)	0.4	2269 (44.7)	2294 (45.2)	1.0
Tobacco Exposure	746 (14.7)	1095 (21.6)	17.9	608 (13.7)	725 (16.3)	7.4	812 (6.7)	1095 (21.6)	8.6	811 (16.0)	1095 (21.6)	14.4	785 (15.5)	1095 (21.6)	15.8
Alcohol	1857 (36.6)	2209 (43.7)	14.6	1592 (35.9)	1679 (37.8)	4.1	2008 (16.5)	2219 (43.7)	0.3	1725 (34.0)	2219 (43.7)	20.1	1694 (33.4)	2219 (43.7)	21.4
BMI	26.1±4.7	25.8±4.7	6.4	25.8 ± 4.8	26.0± 4.7	4.0	26.6 ± 5.3	25.8 ± 4.8	18.0	26.2±4.8	25.8±4.7	8.4	26.0±4.7	25.8±4.7	4.3

$\lvert SB_k \rvert$ presents absolute average standardized bias of k-th covariate. Parenthesis indicates percentage of related baseline characteristic.

Table 3: Average percent standardized bias and bias reduction results

Methods	$\|SB\|$	BR
Before Matching	28.9	-
Nearest Matching (1:1)	7.8	56.5
Caliper Matching	3.2	81.0
Stratified Matching	4.0	76.8
Mahalanobis Matching	7.9	76.0
Mah-Ps Matching	8.1	76.3

Table 4: Classification performances of methods

Methods	Accuracy (%)	Sensitivity (%)	Specificity (%)
Before Matching	75.3	77.9	65.7
Nearest Matching (1:1)	58.3	55.8	64.8
Caliper Matching	52.3	51.6	54.7
Stratified Matching	75.5	77.6	64.8
Mahalanobis Matching	58.3	55.8	64.9
Mah-Ps Matching	58.3	55.8	64.8

4. Discussion

Observational studies have been widely using to estimate treatment effects on outcomes to reduce bias. Greedy Matching methods have generally preferred in many fields because of the calculation simplicity. In this study, nearest, caliper, stratified, and combined propensity score and Mahalanobis metric matching methods are investigated in terms of bias reduction and classification performances on cigarette usage of individuals in Turkey. Results indicated that gender, age group, education level, marital status, working status, chronic disease, body mass index, tobacco exposure, and alcohol usage of individuals are significant on cigarette usage in Turkey. For this reason, sociological and medical effects of these variables should be discussed for future studies in terms of cigarette usage. On the other hand, income levels of individuals do not significantly effect on cigarette usage. It means that individuals in Turkey tend to smoke regardless of their income.

When it comes to the comparison of greedy matching methods, Caliper matching method has the lowest average standardized bias. It also provides the highest average bias reduction. Besides, stratified matching method reveals the second lowest average standardized bias and provides the second highest average bias reduction. It can be mentioned that Caliper method is slightly superior to stratified matching according to the average standardized bias and bias reduction. On the other hand, stratified matching gives the highest accuracy rate whereas caliper matching has low accuracy rate.
Even though before matching provides high accuracy rate, it still produces high average standardized bias. Thus, the necessity and importance of the greedy matching methods is clear in terms of the average standardized bias. Briefly, caliper and stratified matching methods should be preferred if the researchers consider bias reduction whereas stratified matching method should be used in classification studies.

Conflicts of interest

There is no conflict of interest.

References

[1] Rosenbaum P.R., Rubin, D.B. Assessing Sensitivity to an Unobserved Binary Covariate in an Observational Study with Binary Outcome. *J. R. Stat. Soc. Series B*, 45 (1983) 212–218.
[2] Morgan C.J. Reducing Bias Using Propensity Score Matching. *J. Nucl. Cardiol.*, 25 (2018) 404-406.

[3] Rubin D.B. Estimating Causal Effects From Large Data Sets Using Propensity Scores. *Ann. Int. Med.*, 15 (1997) 757-763.

[4] Larsen M.D. An Analysis of Survey Data on Smoking Using Propensity Scores. *Sankhyā: Ind. J. Stat., Series B*, (1960-2002), (1999) 61.

[5] Austin P.C. A Tutorial and Case Study in Propensity Score Analysis: An Application to Estimating the Effect of In-hospital Smoking Cessation Counselling on Mortality. *Multivar. Behav. Res.*, 46 (2011) 119-151.

[6] Taylor G., Girling A., McNeill A., and Aveyard P. Does Smoking Cessation Result in Improved Mental Health? A Comparison of Regression Modelling and Propensity Score Matching. *BMJ Open*, (2015) 5:e008774. doi: 10.1136/bmjopen-2015-008774.

[7] Bidwell L.C., Palmer R.H.C., Brick L., Madden P.A.F., Heath A.C., and Knopik V.S. A Propensity Scoring Approach to Characterizing the Effects of Maternal Smoking During Pregnancy on Offspring's Initial Responses to Cigarettes and Alcohol. *Behav. Gen.*, 46 (2016) 416–430.

[8] Lin K.F., Wu H.F., Huang W.C., Tang P.L., Wu M.T., and Wu F.Z. Propensity Score Analysis of Lung Cancer Risk in a Population with High Prevalence of Non-Smoking Related Lung Cancer. *BMC Pulm. Med.*, 17 (2017) 120.

[9] Sahota S., Lovecchio F., Harold R.E., Beal M.D., and Manning D.W. The Effect of Smoking on Thirty-Day Postoperative Complications after Total Joint Arthroplasty: A Propensity Score-Matched Analysis. *J. Arthroplasty*, 33 (2017) 30-35.

[10] Song Y., Lee D., Suh D.C., Kim J.G., Kim J.K., Han M., Liu H., Zhao L., Kim E.H., Jung S.C., Lee D.G., Koo H.J., Kim M.J., Baek S., Hwang S.M., Kim B.J., Kim Y.J., Cho H.J., Kim S.J., Jeon S.B., and Kim J.S. Cigarette Smoking Preferentially Affects Intracranial Vessels in Young Males: A Propensity-Score Matching Analysis. *Neurointervention*, 14 (2019) 43-52.

[11] Rubin D.B., Thomas N. Matching Using Estimated Propensity Scores: Relating Theory to Practice. *Biometrics*, 52 (1996) 249-264.

[12] Caliendo M., Kopeinig S. Some Practical Guidance for the Implementation of Propensity Score Matching. *J. Econ. Surv.*, 22 (2005) 31-72.

[13] Ho D.E., Imai K., King G., and Stuart E.A. MatchIt: Nonparametric Preprocessing for Parametric Causal Inference. *J. Stat. Soft.*, 42 (2011) 1–28.

[14] Stuart E.A., Rubin D.B. Matching with Multiple Control Groups and Adjusting for Group Differences. *J. Ed. Behav. Stat.*, 33 (2008) 279-306.

[15] Cochran W.G. The Effectiveness of Adjustment by Subclassification in Removing Bias in Observational Studies. *Biometrics*, 24, (1965) 295–313.

[16] Guo X.S., Fraser M.W. Propensity score analysis: statistical methods and applications. 2nd ed. London, Thousand Oaks: Sage Publications 2015; pp. 100-150.

[17] Olmos A., Govindasamy P. Propensity Scores: A Practical Introduction Using R. *J. MultiDiscip. Eva.*, 11 (2015) 1-88.

[18] Normand S.L., Landrum M.B., Guadagnoli E., Ayanian J.Z., Ryan T.J., Cleary P.D., and McNeil B.J. Validating Recommendations for Coronary Angiography Following an Acute Myocardial Infarction in the Elderly: A Matched Analysis Using Propensity Scores. *J. Clin. Epide.*, 54 (2001) 387-398.

[19] Cochran W.G., Rubin, D.B. Controlling Bias in Observational Studies: A Review. *Ind. J. Stat.*, 35 (1973) 417–446.

[20] Turkish Statistical Institute. Available at: http://www.tuik.gov.tr. Retrieved Oct 23, 2019.