Factors Influencing Drug Injection History among Prisoners: A Comparison between Classification and Regression Trees and Logistic Regression Analysis

Azam Rastegari[1], Ali Akbar Haghdoost PhD[2], Mohammad Reza Baneshi PhD[3]

Original Article

Abstract

Background: Due to the importance of medical studies, researchers of this field should be familiar with various types of statistical analyses to select the most appropriate method based on the characteristics of their data sets. Classification and regression trees (CARTs) can be as complementary to regression models. We compared the performance of a logistic regression model and a CART in predicting drug injection among prisoners.

Methods: Data of 2720 Iranian prisoners was studied to determine the factors influencing drug injection. The collected data was divided into two groups of training and testing. A logistic regression model and a CART were applied on training data. The performance of the two models was then evaluated on testing data.

Findings: The regression model and the CART had 8 and 4 significant variables, respectively. Overall, heroin use, history of imprisonment, age at first drug use, and marital status were important factors in determining the history of drug injection. Subjects without the history of heroin use or heroin users with short-term imprisonment were at lower risk of drug injection. Among heroin addicts with long-term imprisonment, individuals with higher age at first drug use and married subjects were at lower risk of drug injection. Although the logistic regression model was more sensitive than the CART, the two models had the same levels of specificity and classification accuracy.

Conclusion: In this study, both sensitivity and specificity were important. While the logistic regression model had better performance, the graphical presentation of the CART simplifies the interpretation of the results. In general, a combination of different analytical methods is recommended to explore the effects of variables.

Keywords: Classification and regression trees, Logistic regression model, History of drug injection, Drug abuse

Citation: Rastegari A, Haghdoost AA, Baneshi MR. **Factors Influencing Drug Injection History among Prisoners: A Comparison between Classification and Regression Trees and Logistic Regression Analysis.** Addict Health 2013; 5(1-2): 7-15.

Received: 08.07.2012 **Accepted:** 23.10.2012

1- MSc Student, Research Center for Health Services Management, Institute for Futures Studies in Health, Kerman University of Medical Sciences, Kerman, Iran

2- Professor , Department of Epidemiology, Regional Knowledge Hub for HIV/AIDS Surveillance, Kerman University of Medical Sciences, Kerman, Iran

3- Assistant Professor, Research Center for Modeling in Health, Institute for Futures Studies in Health, Kerman University of Medical Sciences, Kerman, Iran

Correspondence to: Mohammad Reza Baneshi PhD, Email: m_baneshi@kmu.ac.ir

Introduction

Nowadays, researchers choose relevant statistical methods based on the assumptions and circumstances of their study. One of the major problems in medical studies is determining independent variables with the greatest impact on the outcomes. Although the classic method for these studies is using regression models, the prerequisites of each method should be evaluated before its implication. As regression models require a linear relationship between dependent and independent variables, their use in the absence of such a relation may be misleading. On the other hand, possible complex interactions and patterns between variables cannot be identified unless interaction terms are added to the regression model which in turn increases the complexity of the model and makes its interpretation difficult.[1]

The reliability of predictive models depends on the sample size and the number of variables. A model with high number of variables and small sample size will not have reliable results. Generally, obtaining appropriate regression coefficients entails a minimum of 10 outcome events per variable.[2,3]

Since regression models are usually designed to predict the status of future patients, a mathematical formula needs to be developed based on the calculated regression coefficients. Moreover, besides interpretation of results, some researchers may seek for suitable charts and graphs to present the results in an understandable way for individuals without advanced statistical information. Therefore, researchers' familiarity with alternative methods of regression analysis will enable them to deal with various situations through the most appropriate model.

While some researchers prefer classic statistical methods, more accurate evaluations of a particular study's data and specifications may suggest better models. Classification and regression trees (CARTs) are alternative methods to categorize and predict important medical events such as survival in patients with breast cancer, risk of cardiovascular diseases, and the incidence of death in renal patients. They provide the chance of graphical interpretation without the limitations of regression models.[1,4,5] CARTs comprise a root node, branches, and leaves (terminal nodes). The root node is placed at the top of the tree and includes all observations. It is then split by an independent variable. Afterward, the goodness of split criterion is applied to select the best split on the variable, i.e. the split that maximizes the reduction in the degree of heterogeneity at the corresponding node. This procedure is repeated for all variables. Nodes without any branches are called terminal nodes or leaves.[6-8]

Despite the many benefits and wide use of CARTs, they have some disadvantages. Most importantly, the parent node is split into child nodes using only one variable. Adjustments for other variables are not considered and it is not possible to estimate odds ratios. In addition, since all variables and their levels are tested as possible cutoff points to select branches, the model may be sensitive even to small changes in data. Nevertheless, the results of all statistical techniques depend on data sets.[1,4,5]

Addiction is a crucial issue with destructive impacts. Among various forms of addiction, injection drug abuse undoubtedly imposes the greatest health effects on the society. As a high risk behavior, injection drug abuse not only has legal, psychological, and social aspects but also increases the risk of hepatitis B and C. Most cases (65%) of human immunodeficiency virus (HIV) infection in Iran are caused by injection drug abuse.[9] The global prevalence of hepatitis C among intravenous drug users has been estimated at 50-90%.[10]

Therefore, factors leading to injection drug abuse should be identified and prevented. The high prevalence of drug abuse in prisons and difficult access to drugs have increased the tendency for drug injection. On the other hand, broad use of shared syringes elevates the risk of infection with HIV and hepatitis viruses. As the prisoners can communicate diseases to other individuals after release, prisons should receive extensive attention as places with high potential for spreading risky behaviors and infectious diseases. This study hence used the CART to determine factors influencing injection drug abuse in prisons of Iran.

Methods

Random sampling was used to select 13 small and 14 big prisons (with ≤ 300 and > 300 prisoners, respectively). Afterward, 200 prisoners were systematically selected from each prison and their information was collected by a questionnaire. Finally, 2720 subjects were recruited according to the aim of this research.

History of injection (yes/no answers) was the dependent variable. Age, years of imprisonment, age at first drug use, education (illiterate, able to read and write, primary school, secondary school, university), occupation before arrest (truck driver, seasonal worker, unemployed, and other businesses), and marital status (single, married, and other) were considered as independent variables. Other independent variables including reason for arrest (drug trafficking, murder, acts incompatible with chastity, fights, robbery, financial crimes, and smuggling), kind of drug used one month before arrest (marijuana, weed, ecstasy, opium, heroin, crack, crystal, methadone, and alcohol), and knowledge about HIV were answered as yes or no.

CARTs use the Gini coefficient in order to assess inhomogeneity in binary splits of parent node to child nodes. Gini coefficient is the most widely used criterion for measuring the inhomogeneity. Gini coefficient takes a value of zero when all observations at a node belong to one level of a dependent variable. It takes its maximum amount (0.5) when observations are equally distributed in various levels of a dependent variable. Gini coefficient is calculated for all levels of all variables. Therefore, the best split on a variable will be the one that minimizes the Gini coefficient. This process continues until one of the termination criteria is met. Termination criteria include user-defined limits (the minimum number of parent and child nodes) or pure terminal nodes (i.e. when all observations belong to the same level of a dependent variable).[11]

Performing several trials to obtain the final model substantially increases type I error. Therefore, level of significance of each split variable will be adjusted by Bonferroni correction.[12] In this method, the tree is first fitted to the data with the greatest number of nodes. Since this tree has numerous nodes, it is highly complex while containing the least wrong categorizations. Prediction at each node is made based on prior probability (the weight of observations in each category of the dependent variable) and misclassification cost (number of falsely categorized observations).[13]

Pruning should hence result in a final tree with the lowest complexity and misclassification cost. Pruning starts from terminal nodes. A terminal node will be deleted only if its elimination causes a misclassification cost which is significantly lower than the reduction in complexity. The complexity parameter of the new tree is then calculated. This process continues to reach the root node. Finally, complexity parameters are plotted against tree size (number of terminal nodes) and the optimal tree is selected.[1,14,15]

In the present study, the CART and logistic regression models were fitted to the data set. The minimum numbers of observations in parent and child nodes were considered as 100 and 50, respectively. The models were compared in terms of sensitivity, specificity, and accuracy.

Fitting a model makes it useful in prediction and analysis of new data sets. We first randomly allocated data two a training group (75%) and a testing group (25%). After fitting the models on the training group, they were applied to the testing group. The fitted model was then used to predict the samples belonging to different classes of the response variable. Considering the low number of outcomes (having the history of injection) in the overall data set of this study, the largest proportion of data was allocated to the training group. Otherwise, the low incidence of outcomes in the training group could affect the quality of the model. Finally, SPSS FOR Windows 16.0 (SPSS Inc., Chicago, IL, USA) was used to evaluate the fitness of the models.

Results

The study sample consisted of 2720 prisoners with mean age of 32.82 ± 8.56 years old and mean history of imprisonment was 2.18 ± 2.35 years. While most participants (65.8%) had primary or secondary school education, a few (13.5%) were illiterate or could only read and write and 20.7% had a university degree. Seasonal workers, truck drivers, and those with other jobs comprised 12.8%, 5.4%, and 78.1% of the whole population, respectively. Overall, 3.7% were unemployed. The majority of subjects (52.1%) were married, 34.8% were single and 13.1% had other marital status.

The most common reasons for arrest were drug trafficking (52.6%) and robbery (25.8%). In total, 22.7% had the history of drug injection

(Table 1). Heroin and crack (51.5%), opium (47.5%), and crystal (13.3%) were the most widely used drugs. The mean age at first drug use was 20.79 ± 6.36 years old.

The CART suggested four variables of heroin, imprisonment history, age at first drug use, and marital status (Figure 1). In fact, heroin users with imprisonment of less than three years were at lower risk of injecting drugs and about 70% of them did not have the history of drug injection. On the other hand, heroin addicts with more than three years of imprisonment and less than 16 years of age at first

Table 1. Frequency and percentage of prisoners with history of drug injection

Data	History of drug injection	
	Yes	No
	n (%)	n (%)
Training data group	458 (22.4)	1591 (77.6)
Testing data group	160 (23.8)	511 (76.2)
Total data	618 (22.7)	2101 (77.3)

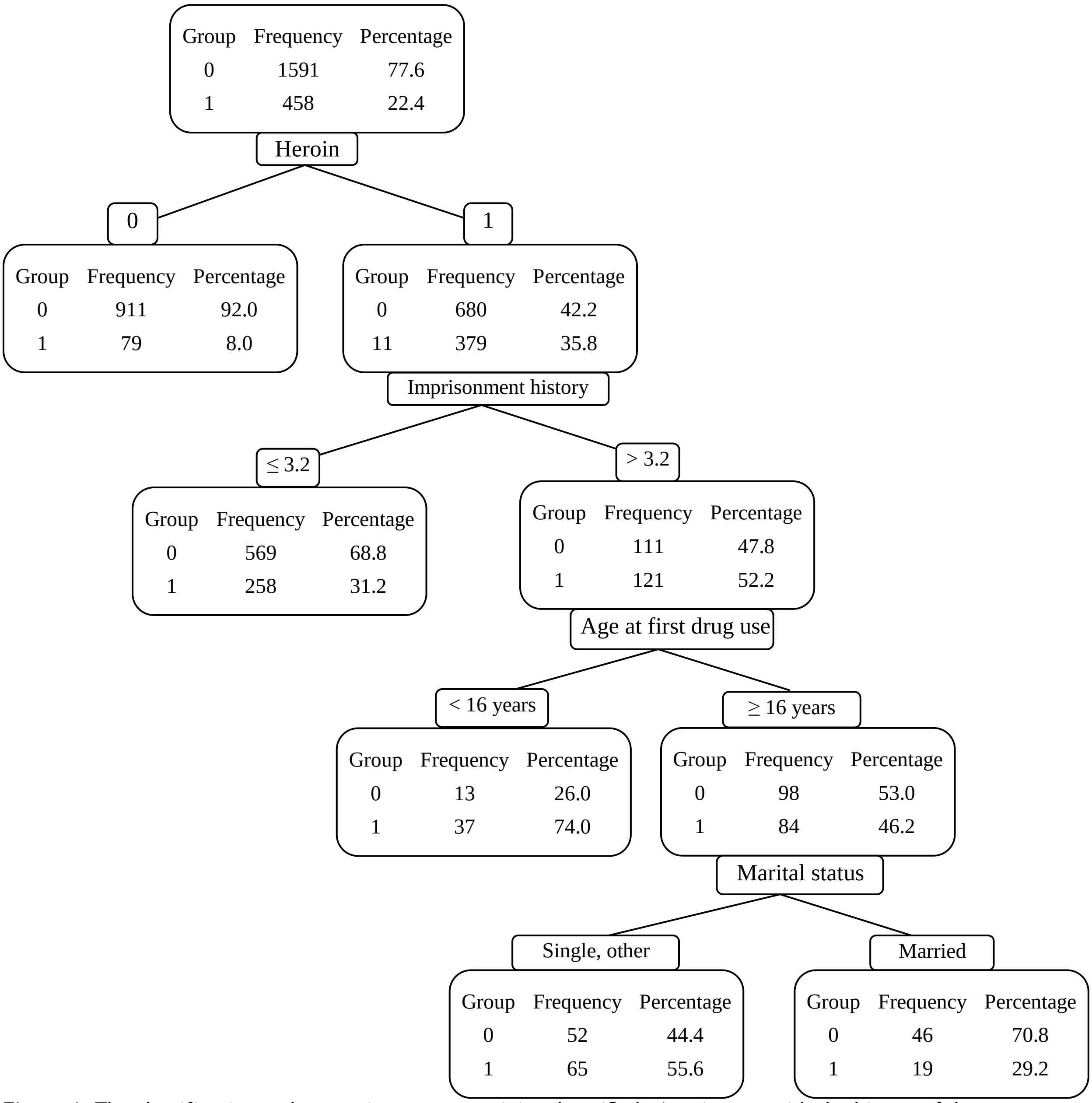

Figure 1. The classification and regression tree on training data (Code 1: prisoners with the history of drug injection. At each node, the group with the highest percentage was considered as the predicting group)

drug use were at higher risk of drug injection (about 74% of these individuals had the history of drug injection). Marital status was the predictor of drug injection among subjects who had started using drugs after 16 years of age. More precisely speaking, while being single was associated with higher risk of drug injection, the contrary was true about being married (70% of married participants did not have the history of drug injection).

Eight variables, i.e. significant variables in the CART plus age and history of opium, methadone, and ecstasy use one month before arrest, remained in the final logistic regression model. According to the logistic regression model, every one year increase in imprisonment increased the risk of drug injection by 15%. In contrast, every one year increase in age at first drug use decreased the risk of drug injection by 10%. Moreover, single prisoners or those with other marital status were at higher risk of drug injection compared to married subjects (53% and 116%, respectively). Apparently, the results of the logistic regression model and the CART were similar.

Table 2 summarizes the sensitivity, specificity, and accuracy of each part of the CART and the regression model. As it is seen, adding a variable to the logistic regression model increased its sensitivity but decreased its specificity. However, the same was not true in the CART. Therefore, although the sensitivity of the logistic regression model was higher than the CART, they had the same level of accuracy and specificity. The sensitivity of the final regression model and the CART were 27% and

Table 2. The results of the logistic regression model and the classification and regression tree in training data group

	Variable	Odds ratio	95% Confidence interval	Sensitivity (%)	Specificity (%)	Accuracy (%)
Logistic regression model						
First step	Heroin use	6.42	4.95-8.32	0.0	100	77.6
Second step	Heroin use	7.00	5.39-9.20	8.3	97.6	77.6
	History of arrest	1.17	1.12-1.23			
Third step	Heroin	6.18	4.7-8.12	15.3	96.8	78.6
	History of arrest	1.17	1.14-1.25			
	Age at first drug use	0.92	0.9-0.94			
Fourth step	Heroin use	5.84	4.4-7.70	20.1	96.2	79.2
	History of arrest	1.18	1.13 -1.24			
	Age at first drug use	0.92	0.89-0.94			
	Marital status	1.16	0.90-1.50			
	Other	2.18	1.57-3.028			
Fifth step	Heroin	4.30	3.2-5.90	27.1	95.2	79.0
	History of arrest	1.15	1.10-1.22			
	Age at first drug use	0.90	0.88-0.92			
	Marital status	1.53	1.14-2.05			
	Other	2.16	1.55-3.02			
	Age	1.05	1.03-1.07			
	Opium use	0.60	0.38-0.70			
	Methadone use	2.21	1.37-3.50			
	Ecstasy use	9.50	1.47-62.54			
Tree model						
First step	Heroin use	-	-	0.0	100	77.0
Second step	Heroin use	-	-	26.4	93.0	78.0
	History of arrest	-	-	-	-	-
Third step	Heroin use	-	-	-	-	-
	History of arrest	-	-	8.1	99.2	79.0
	Age at first drug use	-	-	-	-	-
Forth step	Heroin use	-	-	-	-	-
	History of arrest	-	-	-	-	-
	Age at first drug use	-	-	14.2	96.7	78.0
	Marital status	-	-	-	-	-

Table 3. Comparing the results of the logistic regression model and the classification and regression tree (CART) on total data and training and testing data

	Model	Sensitivity (%)	Specificity (%)	Accuracy (%)
Training data	Logistic regression	27.1	95.2	79.0
	CART*	14.2	96.7	78.0
Testing data	Logistic regression	30.0	94.0	78.0
	CART	25.0	96.7	80.0
Total data	Logistic regression	31.9	94.0	79.9
	CART	24.9	95.5	79.5

* CART: Classification and regression tree

14%, respectively (Table 3). Assessing the functionality of the two models in predicting the dependent model using the testing group revealed higher sensitivity of the regression model (30% vs. 25%). Nevertheless, the specificity and accuracy of the CART were about 2% higher than the regression model. Considering the entire data set, the logistic regression model and CART had the area under the receiver operating characteristic (ROC) curve of 0.798 and 0.765, respectively.

Discussion

According to the CART, history of heroin use one month before arrest, history of imprisonment, age at first drug use, and marital status can predict the risk of drug injection. Using regression model required four additional independent variables. Comparison between the findings of the two models suggested heroin use, low age at first drug use, longer imprisonment, and being single to increase the risk of drug injection. Most studies have indicated a direct relation between heroin use and drug injection. In fact, as addicts will need more heroin over time, its high cost will force them to inject drugs. In a study on 7743 drug addicts, Yarmohammadi Vasl and Ghanadi found that early onset of drug use was significantly associated with drug injection.[9] Similarly, the Iranian Center for Prison Education reported a relation between drug injection and longer prison sentences.[16]

As the logistic regression model had greater number of significant variables, it had higher sensitivity in identifying people with history of drug injection. However, the specificity and accuracy of the CART and the regression model were the same. Therefore, the logistic regression model was more practical than the CART. Nevertheless, interpretation and future use of the CART model are simpler.

The present study sought to compare CART and logistic regression models in predicting medical implications. Sensitivity, specificity, goodness of fit, coefficient of determination, sum of squared errors (difference between real and predicted values), and the area under the ROC curve are among the indices used to compare the performance of various models. However, the goodness of fit index and coefficient of determination of regression trees cannot be calculated. Hence, sensitivity and specificity are mainly used in comparisons between logistic regression models and CARTs. Models to recognize both healthy and ill patients need to be evaluated in terms of not only sensitivity but also specificity. Since we tried to accurately identify subjects with and without the history of drug injection, we calculated both the sensitivity and specificity of the models. Low sensitivity of the two models in this study suggests that the selected independent variables were not the best predictors of the dependent variable. Therefore, other effective independent variables should also be considered. As high classification accuracy does not guarantee high sensitivity and specificity, interpretation of the results requires the simultaneous assessment of all criteria.

In medical studies, models have to be fitted according to their performance on independent data and future application. If lower levels of sensitivity and specificity are calculated using the training data than the testing data, the model will have a poor performance. As we mentioned, the two models in this study had similar performance on training data.

Furthermore, number of outcome events per variable should not be disregarded when comparing models. Based on previous research, less than five outcome events per variable will increase error. Therefore, a minimum of 10 outcome events per variable will be required for valid regression coefficient and confidence intervals.[2,3] In our study, number of outcome events per variable was 25 in the whole data set and 18.3 in the testing group.

CARTs facilitate the evaluation of interactive effects of variables. In the present study for instance, age at first drug use could predict drug injection in individuals with history of heroin use and longer imprisonment. Assessment of interactive effects in regression models necessitates additional terms which will complicate the model. Another advantage of CARTs is the use of surrogate splitters to handle missing data. In fact, a CART simply substitutes missing data at a node by another variable that splits the node with the highest homogeneity. In contrast, complex methods to replace missing data in regression models substantially increase the volume of analysis.[17,18]

Colombet et al. compared the efficacy of regression models and CARTs in estimating the risk of heart disease in 15444 patients. They found the CART to be more accurate (69% vs. 65%).[19] In a study by Tsien et al. to predict heart failure in 1252 patients with chest pain, the areas under ROC curves of the two mentioned models were very close.[20] Delan et al. predicted five-year survival of patients with breast cancer by regression models and CARTs. They reported higher sensitivity, specificity, and the area under ROC curve for the regression model.[21] Keshtkar et al. suggested the CART to have higher sensitivity, accuracy, and specificity in determination of factors affecting the intensity of preeclampsia.[22] Ma et al. published similar findings.[23]

Although the number of event per variable is an important factor in performance of regression models and CARTs, it has not been mentioned by most previous studies. All aspects of studies cannot hence be compared. Regression models have better performance than CARTs in studies with greater number of events per variable.[24] Likewise, the regression model showed better performance than CART in our study with 18 outcome events per variable in the training data set. Different numbers of outcome events per variable in the present study and previous research can justify their inconsistencies. Therefore, further studies with different numbers of outcome events per variable and different sample sizes are recommended.

Conclusion

CARTs are represented graphically, simple to interpret, and able to identify interactive effects of variables. Moreover, they easily deal with the problems of missing data. They are suggested as complementary to regression models for better explanation of how independent variables affect dependent variables.

Conflict of Interest

The Authors have no conflict of interest.

Acknowledgements

Appreciation goes to Dr. Abbas Sedaghat, Dr. Ehsan Mostafavi, Dr. Mehdi Osoli, Dr. Soudabeh Navadeh, Dr. Leila Sajadi, and the HIV Care Center of Kerman University of Medical Sciences, Epidemic Diseases Management Center, and Prisons and Security and Corrective Measures Organization.

This article was the result of an MSc thesis (project number: 90/37). The research design was approved by the Deputy of Research and Technology at Kerman University of Medical Sciences.

References

1. Lewis RJ. An introduction to classification and regression tree (CART) analysis. Proceedings of the Annual Meeting of the Society for Academic Emergency Medicine; 2000 May; San Francisco, USA.
2. Bagley SC, White H, Golomb BA. Logistic regression in the medical literature: standards for use and reporting, with particular attention to one medical domain. J Clin Epidemiol 2001; 54(10): 979-85.
3. Van Voorhis CRW, Morgan BL. Understanding power and rules of thumb for determining sample sizes. Tutorials Quantitative Methods for Psycholog 2007; 3(2): 43-50.
4. Indurkhya N, Weiss SM. Estimating performance gains for voted decision trees. Intelligent Data Analysis 1998; 2(1-4): 303-10.
5. Podgorelec V, Kokol P, Stiglic B, Rozman L. Decision trees: An overview and their use in medicine. Journal of Medical Systems 2002; 26(5): 445-63.
6. Bensic M, Sarlija N, Zekic-Susac M. Modelling small-business credit scoring by using logistic regression, neural networks and decision trees. Intelligent Systems in Accounting, Finance and

Management 2006; 13(3): 133-50.
7. Wilkinson, L. Tree structured data analysis: AID, CHAID and CART. Proceedings of the Sawtooth/SYSTAT Joint Software Conference; 1992; Sun Valley, USA.
8. Heping Z. Recursive partitioning and tree-based methods. In: Gentle JE, Hardle W, Mori Y, editors. Handbook of computational statistics: concepts and methods.New York City, NY: Springer; 2004. p. 813-40.
9. Yarmohammadi Vasl M, Ghanadi F. Study of the transition from non-injection to injection in heroin users. Journal of Hamadan University of Medical Sciences 2010; 17(2): 50-6.
10. Zakizadeh M, Sadeghian AA, Bagheri Nesami M. Seroprevalence of hepatitis C infection and associated factors in addicts imprisoned at Khezerabad prison, Sari. J Shaheed Sadoughi Univ Med Sci 2006; 14(2): 29-37. [In Persian].
11. Bradford JP, Kunz C, Kohavi R, Brunk C, Brodley CE. Pruning decision trees with misclassification costs. ECE Technical Reports 1998 p. 131-6.
12. Jensen D, Schmill MD. Adjusting for multiple comparisons in decision tree pruning. 3th International Conference on Knowledge Discovery and Data Mining (KDD-97); 1997 Aug 14-17; California, United States of America. 1997 p. 195-8.
13. Quinlan JR. Induction of decision trees. Machine Learning 1986; 1(1): 81-106.
14. Chang LY, Chen WC. Data mining of tree-based models to analyze freeway accident frequency. J Safety Res 2005; 36(4): 365-75.
15. Rastogi R, Shim K. A decision tree classifier that integrates building and pruning. Data Mining and Knowledge Discovery 2000; 4: 315-44.
16. Bolhari J, Mirzamani SM. Evaluation of drug abuse in prisons in Iran. Tehran, Iran: Iran Drug Control Headquarters: 2001.
17. Weiss SM, Indurkhya N. Decision-rule solutions for data mining with missing values. Advances in Artificial Intelligence 2000 p. 1-10.
18. Lobo O, Numao M. Ordered estimation of missing values. Methodologies for Knowledge Discovery and Data Mining 1999; 499-503.
19. Colombet I, Ruelland A, Chatellier G, Gueyffier F, Degoulet P, Jaulent MC. Models to predict cardiovascular risk: comparison of CART, multilayer perceptron and logistic regression. Proc AMIA Symp 2000; 156-60.
20. Tsien CL, Fraser HS, Long WJ, Kennedy RL. Using classification tree and logistic regression methods to diagnose myocardial infarction. Stud Health Technol Inform 1998; 52 Pt 1: 493-7.
21. Delan D, Walker G, Kadam A. Predicting breast cancer survivability: a comparison of three data mining methods. Artificial Intelligence in Medicine 2005; 34(2): 113-27.
22. Keshtkar AA, Majdzade SR, Mohammad K, Ramezanzade F, Borna S, Azemikhah A, et al. Determination of effectvie factors on preeclampsia severity the application of classification and regression trees. J Gorgan Univ Med Sci 2006; 8(2): 47-54. [In Persian].
23. Ma CM, Chao CM, Wu VC, Cheng BW. Predicting patients at risk of acute renal failure in intensive care units using artificial intelligence tools. 2008.
24. Baneshi MR. Statistical models in prognostic modelling with many skewed variables and missing data: a case study in breast cancer. [PhD Thesis]. Edinburgh: Edinburgh University. 2009.

تعیین عوامل مؤثر بر سابقه تزریق مواد در زندانیان: مقایسه مدل رگرسیون لجستیک و مدل درختی

اعظم رستگاری[1]، دکتر علی اکبر حق‌دوست[2]، دکتر محمد رضا بانشی[3]

مقاله پژوهشی

چکیده

مقدمه: به دلیل اهمیت مطالعات انجام شده در زمینه پزشکی، پژوهشگران باید با انواع روش‌های آماری آشنا باشند تا به این وسیله بتوانند مناسب‌ترین روش آماری را انتخاب نمایند. روش درختی به عنوان یک روش مکمل برای مدل‌های رگرسیونی قابل استفاده است. در این مطالعه به بررسی و مقایسه ویژگی‌های دو مدل رگرسیون لجستیک و مدل درختی پرداخته شد.

روش‌ها: داده‌ها مربوط به ۲۷۲۰ زندانی می‌باشد که به بررسی عوامل مؤثر بر داشتن سابقه تزریق مواد در آن‌ها پرداخته شد. داده‌ها به دو قسمت آموزشی و آزمایشی تقسیم گردید و دو مدل روی مجموعه آموزشی برازش داده و عملکرد آن روی مجموعه آزمایشی بررسی شد.

یافته‌ها: تعداد متغیر معنی‌دار در مدل رگرسیونی و درختی به ترتیب ۸ و ۴ بود. متغیرهای هروئین، سابقه حبس، سن شروع مصرف مواد و وضعیت تأهل از عوامل مهم برای تعیین داشتن سابقه تزریق بود. افرادی که سابقه مصرف هروئین نداشتند و یا با وجود مصرف هروئین سابقه حبس کوتاه مدت داشتند، کمتر در معرض خطر تزریق مواد بودند. افراد هروئینی با سابقه حبس طولانی، افرادی که سن شروع مصرف مواد آن‌ها بالا بود و متأهلان نیز افراد کم‌خطری بودند. حساسیت مدل لجستیک از مدل درختی بالاتر بود، ولی ویژگی و صحت دسته‌بندی دو روش مشابه بود.

نتیجه‌گیری: در مطالعه حاضر، هر دو معیار حساسیت و ویژگی مهم است. با این وجود، عملکرد مدل لجستیک بهتر بود اما، نمودار گرافیکی مدل درختی امکان تفسیر نتایج را برای عموم مخاطبان امکان‌پذیر می‌کند. استفاده از روش‌های آنالیز مختلف برای کشف تأثیر متغیرها توصیه می‌شود.

واژگان کلیدی: مدل درختی، مدل رگرسیون لجستیک، سابقه تزریق، مصرف مواد

ارجاع: رستگاری اعظم، حق‌دوست علی اکبر، بانشی محمد رضا. **تعیین عوامل مؤثر بر سابقه تزریق مواد در زندانیان: مقایسه مدل رگرسیون لجستیک و مدل درختی.** مجله اعتیاد و سلامت ۱۳۹۲؛ ۵ (۲-۱): ۱۵-۷.

تاریخ دریافت: ۹۱/۴/۱۸ **تاریخ پذیرش:** ۹۱/۸/۲

۱- دانشجوی کارشناسی ارشد، مرکز تحقیقات مدیریت خدمات بهداشتی، مرکز مطالعات آینده در سلامت، دانشگاه علوم پزشکی کرمان، کرمان، ایران
۲- استاد، گروه اپیدمیولوژی، مرکز منطقه‌ای آموزش نظام مراقبت HIV/ ایدز، دانشگاه علوم پزشکی کرمان، کرمان، ایران
۳- استادیار، مرکز تحقیقات مدل‌سازی در سلامت، مرکز مطالعات آینده در سلامت، دانشگاه علوم پزشکی کرمان، کرمان، ایران
نویسنده مسؤول: دکتر محمد رضا بانشی Email: m_baneshi@kmu.ac.ir