IJABS 2017: 4:3

Original Article

A Cross Sectional Study of Iranian Women and Sex Preference for Children

Mahsa Saadati[1*]

1. *Assistant Professor of National Population Studies & Comprehensive Management Institute, Tehran, Iran.*
** (Corresponding author:, Mahsa Saadati , Email: mahsa.saadati@psri.ac.ir)*

(Received: 15 Mar 2017; Revised:30 April 2017 ; Accepted:31 July 2017)

Abstract

Introduction: Children sex preference may have significant effects on fertility behavior, which is an influential component of population dynamics and could control the population size, structure, and composition. The main objective of this study is to investigate affecting factors on Iranian women's child sex preference through applying Classification and Regression Trees algorithm, which is an effective and easy to interpret non-parametric classification method.

Methods: A cross-sectional study was conducted to collect demographical data of 1250 Iranian women aged 15-49. To classify child sex preference for children, age, educational level, place of residence, and number of siblings for women, were nominated as predictors using the SPSS-22 statistical software.

Results: Women's age, educational level and number of siblings were remained in extracted decision tree. The validity of the resulted tree was confirmed by 0.71 accuracy, which means 71% of women's sex preference, has been classified correctly.

Conclusions: The most important determinant of women's child sex preference was age. It could be concluded that educated Iranian women in different age cohorts are in favor of having girls.

Declaration of Interest: None

Key words: Sex differences, Fertility preferences, Child, Women, Decision trees

Introduction

Globally, children sex preference have been a noticeable subject in demographic studies because of its potential negative social and demographic implications (1). Male preference is prevalent in many parts of the world particularly in South and East Asia, parts of the Middle East, and North Africa. In Bangladesh and Nepal, more than 95% and in Burkina Faso and Senegal, more than 30% of women had a preference for having a male child (2). Edlund and Lee resulted in a very insignificant tendency for South Korean women in a good condition to have more boys (2). Daughter preference has also been observed in some West African countries including Ghana (21.3%), Malawi (21.2%) and Liberia (22.2%) (1).

In Asia, the boy preference of many parents has led to perhaps abortion or directly killing of a large number of girls which cause to an unbalanced sex ratios (3). From the early 1980s, families could determine their children sex and also had age old son preference which resulted in unusually high male sex ratios over female in a number of Asian countries, such as China and India, two of the most populated countries (4-6).

Kugler and Kumar studied the number of Indian children for the families by using data from nationally representative household surveys (7). In spite of a strong preference for boys in India, families have children desire in the situation that their first child is a girl. It

means that Indian parents will continue to their children bearing until they reach to their desired sex. Thus, in this society, a convenient covariate to predict the probability of having a second child or the families total number of children could be the first child's sex (7).

American parents, specifically fathers, are also in favor of having boys over girls. Comparing to Asia, these son preference are less severe, but it has some outcomes. Dahl and Moretti (2008) showed that it seems fathers mainly motivate son preference of American parents. Though, women had only a slight preference for daughters (8).

Development of sex preference for children has also been documented in some studies on advanced western societies (9-12). The first reason of this progress might be due to the medical improvements which assist parents to choose deliberately their child's sex (13). The next reason is that in modern low-fertility societies, the sex composition of previous children might influence on the couples' childbearing behavior (14).

Infanticide, sex-selective abortions, or sex selection technologies joint with sever gender preference may lead to a serious bias of sex ratio. This imbalance between the two sexes could cause a postponement in the marriage, or a growth in the number of never married people. Moreover, gender preference may have extensive consequences for fertility behaviors of families. Parents who have the tendency of one or more definite sex children may make larger families comparing to the others (15).

Sex preference can also lead to gender bias in the allocation of food and health care (16). Birth intervals have been observed to be largest for women having equal number of boys and girls, intermediate for those having more boys than girls and shortest for those having more girls than boys. This may imply that sex preference tends to increase fertility levels (16).

Shahbaziyan et al. (2014) investigated that whether sex preference of parents in Iran and more specifically women can change the family fertility behavior in Kangavar city (17). They studied fertility according to sex preference, educational level, job status and women's age. They found that there are a high relationship between educational level and gender preference for children. They showed that more sons led to less fertility. In addition, Mansurian and Khushnevis (2006) studied the sex preference influences of ever married women on their fertility behavior in Tehran. They found that fertility behavior in families having more boys is lower than families with more girls. The women in this study had preference to have more boys (18). Hejazi (2013) studied attitudes of employed Iranian women in Isfehan province to have second child (19). They have considered women's age, educational level, sex preference, and job status as predictors. They failed to find any sex preference for their children in their study.

One of the important determinants of fertility behaviors is sex preference for children and there are not so many studies examined Iranian women's sex preference for children as a response variable and determined its direct influential factors. Therefore, the main aim of this article is to investigate factors which affect Iranian women's sex preference for children by applying Classification & Regression Trees algorithm (CART) as an applicable method.

Methods

Most of researchers modeled sex preference by logistic regression. There are a number of reasons for difficulties of the traditional statistical methods like logistic regression to investigate sex preference. Logistic regression is poorly suitable for multiple evaluations. When there are many possible influential predictors, the task of variable selection is very problematic as well. Complicated interactions or patterns may occur in the model which is another vital issue. In addition, a common, but incorrect, method of handling missing data in most of traditional statistical methods is to exclude cases with missing values; this is both inefficient and runs the risk of introducing bias in the analysis (20-22).

To model sex preference, we applied the CART algorithm which is obtained by recurrently dividing the data, fitting a simple prediction model within each division. The

resulted algorithm graphically is a decision tree (23).

CART Algorithm

Several statistical algorithms for building decision trees are available, including CART (24), C4.5 (25), CHAID (Chi-Squared Automatic Interaction Detection) (26) and QUEST (Quick, Unbiased, Efficient, Statistical Tree) (27). Table (1) provides a brief comparison of the four most widely used decision tree methods (28, 29).

Table 1. Comparison of Different Decision Tree Algorithms

Methods	CART	C4.5	CHAID	QUEST
Measure used to select input variable	Gini index; Towing criteria	Entropy info-gain	Chi-square	Chi-square for categorical variables; j-way ANOVA for continuous/ordinal variables
Pruning	Pre-pruning using a single pass algorithm	Pre-pruning using a single pass algorithm	Pre-pruning using chi-square test for independence	Post-pruning
Dependent variable	Categorical/ Continuous	Categorical/ Continuous	Categorical	Categorical
Input variables	Categorical/ Continuous	Categorical/ Continuous	Categorical/ Continuous	Categorical/ Continuous
Split at each node	Binary; Split on linear combinations	Multiple	Multiple	Binary; Split on linear combinations

CART analysis consists of four steps; the first one contains of tree building by nodes splitting recursively. According to the distribution of node classes in the learning dataset and the decision cost matrix, each resulting node is allocated to a predicted class. The allocation of a predicted class to each node occurs regardless that node is successively split into child nodes. The next step involves in stopping the tree building process. In this situation, a greatest tree has been created, which possibly greatly over-fits the information contained within the learning dataset. The third step contains tree pruning, which results in a procedure for making simpler and simpler trees by cutting off unimportant nodes. The last step consists of optimal tree selection from pruned trees, in the way that it fits but not over-fit the information in the learning dataset.

Although the CART algorithm manual commends to investigate with different splitting measures, these measures will give similar results if response is a binary categorical variable. The Gini, followed by Twoing are the two most common splitting function (23). Gini index is a contamination-based criterion that measures the differences between the probability distributions of the target variable's values. Some previous works have applied the Gini index (24, 30). He indexed is defined as:

$$Gini(y,S) = 1 - \sum_{c_j \epsilon dom(y)} \left(\frac{\left|\sigma_{y=c_j} S\right|}{|S|}\right)^2 \qquad (1)$$

Where S is a training set and y is the probability vector of the target variable. Therefore, the assessment criterion for selecting the attribute a_i is defined as:

$$GiniGain(a_i,S) = Gini(y,S) - \sum_{v_{i,j} \epsilon dom(a_i)} \left(\frac{\left|\sigma_{a_i=v_{i,j}} S\right|}{|S|}\right) Gini\left(y, \sigma_{a_i=v_{i,j}} S\right) \qquad (2)$$

Data Description

The data from 1250 ever married women aged 15-49 years having a child (children) and intended to have more child (children) in "Childbearing Attitudes and Its Social, Economic and Cultural Factors" survey analyzed in this study (31). In the present study, although some behavioral questions were asked, we did not carry out any intervention.

Therefore, there was no requirement to obtain ethical code. Sex preferences of these women were assessed by two different questions, which measured the number of their Children Ever Born (CEB) and desired number of children. Multistage stratified sampling was used to select the women who referred in public health and treatment centers to vaccinate their children in 31 provinces in Iran, 2014. Different factors may affect women's sex preference. The list of dependent (response) and independent variables (factors) used in this study are as follows:

Sex preference for children (response variable): From 6231 women aged 15-49 in the survey (31); those did not want more children and were childless deleted from interested population. Then, the women's number of ever born and desired boys summed up and the same index calculated for girls as well in the resulted population in the previous step. The difference between the sum of girls and boys computed as preference values. Sex preference was categorized as *boy preference, girl preference,* and *no sex preference*. All respondents with zero preference values were classified as having no preference for a child's sex and deleted from interested population. All respondents with negative preference values were classified as having preference for boys and all those women with positive preference values were classified as having preference for girls. Thus respondents with boy or girl preferences were made the final sample in this study were 1250 women.

Place of residence: This is a place where women were living in the study time that could be even urban or rural areas.

Women's age: A four categorical variable with levels of 10-19, 20-29, 30-39, and 40-49 that is considered to measure the age of women in the survey time.

Women's educational level: It was considered as a categorical variable with five categories of under secondary, high school and diploma, associate and bachelor, master and above, and religious degree.

Women with different number of siblings: It has three categories as equal, more sisters, and more brothers, which represented women with equal number of sisters or brothers, more sisters compared to brothers or more brothers compared to sisters, respectively.

Results

The main aim of this study was to classify gender preference of 15-49 years old Iranian women. 51.04% and 48.96% of women preferred to have boys and girls, respectively. 68.6% of women are living in the urban area. 54.9% of them are 20-29 years old. Women's educational level 74.6% is diploma and less. 43.3% of women have more sisters than brothers while 38.7% of women are vice versa. Table (2) shows women's sex preference crossed by predictors in this study. According to the results of this table, women who lived in urban areas (50.3%), had more than 30 years old (52.7% in age 30-39 and 53.8% in age 40-49), with educational level of associate and bachelor and higher (51.9% of associate and bachelor levels and 58.6% of master and above levels), had equal siblings or more sisters (51.1% of equal number of siblings and 50.8% of more sisters) preferred girls. While other women in each level of predictors preferred boys. None of the predictors had significant influential on women's sex preference (Table, 2).

Figure (1) presents decision tree of women's sex preference according to the selected predictors in this study. The rules of the extracted tree are as follows:

- Women whose age were 10 to 29 years old and their educational levels were associate and above or diploma and lower including religious degree preferred to have girls or boys, respectively.
- Women whose age were 30 to 49 years old and they had equal number of siblings and more sisters preferred to have girls.
- Women whose age were 30 to 49 years old and had more brothers according to their educational levels that were under secondary or high school and above preferred to have boys or girls, respectively.

Table (3) presents misclassification matrix for classification model. It specifies the precision of the classification model. In this table, the shaded cells indicate accurate classification of the tree on Figure (1). Equation (3) contains

the calculation method of the classification precision for this tree. It shows the correct proportion of total number, which predicted by tree. The results state that the accuracy of the model is 71%, which indicates 71% ages of women's sex preference have been classified correctly.

Table (4) shows the classification tree risks and standard errors for training and learning data. As mentioned before, to fit CART algorithm, data divided to two different groups

Table 2. Women's Gender Preference by Predicted Variables

Variables		Children preference (Response Variable)			Chi-Square Test	Asymp. Sig. (2-sided)
Name	Value	Girl	Boy	Total		
Place of Residence	Urban	50.3	49.7	100	1.935	<0.164
	Rural	46.1	53.9	100		
Women's Age	10-19	42.6	57.4	100	5.289	<0.152
	20-29	46.6	53.4	100		
	30-39	52.7	47.3	100		
	40-49	53.8	46.2	100		
Women's Educational Level	Under secondary	45.9	54.1	100	4.876	<0.300
	High school & diploma	49.7	50.3	100		
	Associate and Bachelor	51.9	48.1	100		
	Master & above	58.6	41.4	100		
	Religious Degree	0.0	100.0	100		
Women's Difference Number of Siblings	Equal	51.1	48.9	100	3.027	<0.220
	More sisters	50.8	49.2	100		
	More brothers	45.9	54.1	100		

Table 3. Misclassification Matrix for Classification Tree

Observed Category	**Predicted Category**		**Total**
Girl preference	Girl preference	Boy preference	
	426	186	612
Boy preference	172	466	638
Total	598	652	1250

Table 4. Risks and Standard Errors of the Classification Tree for Training and Learning Data

	Risk	**Standard error**
Learning set	0.446	0.014
***k*-fold cross validity of training set**	0.483	0.014

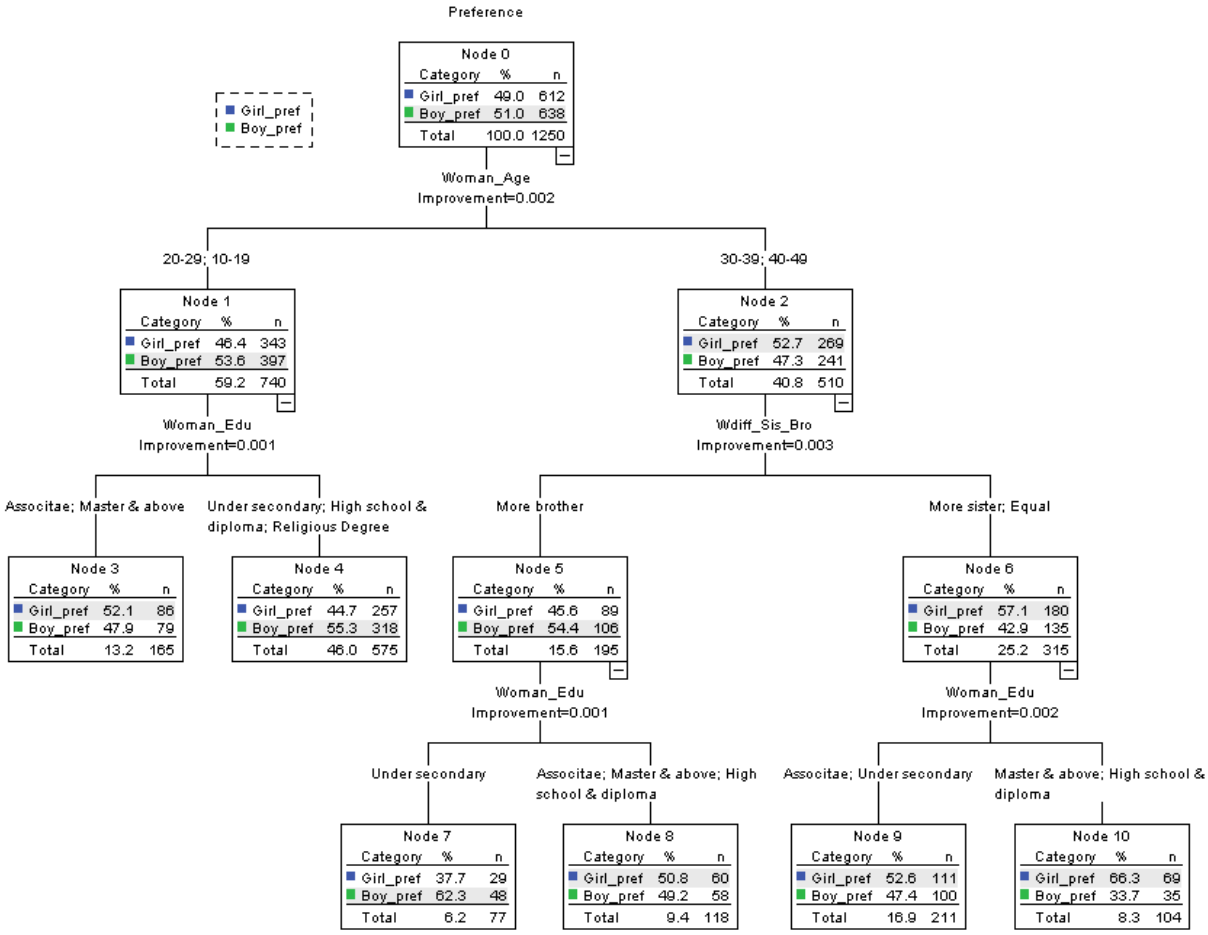

Figure 1. Decision Tree of Women's Sex Preference by Independent Variables

$$Accuracy = \frac{426+466}{1250} = 0.71 \qquad (3)$$

of training and learning data and the model fits to these two groups. Indeed, training and learning data are used for fitting and confirming the validity of the model, respectively. When the risks of these two data sets are close to each other, it confirms the validity of the fitted model. According to the results of Table (4), the equality of these values confirms the validity of classification model, which is proposed by the classification tree in Figure (1).

Discussion

What are the main reasons for parents to prefer children of one sex over another one? It can be due to the fact that children of a particular sex may provide certain services, such as financial, social, or emotional benefits for the family. In developing countries, for example, sons are preferred comparing to daughters because they may provide assistance in agriculture and fishing (32). Moreover, sons are also prized for continuing the family name. On the other hand, daughters could be reliable in providing parents' old-age assistance, helping with household tasks or caring for their younger siblings. Thus, many families with in big favor for sons consider having at least one daughter (10).

Morgan et al. (1988) in his study stated that boys might diminish parents' separation risk, since fathers' responsibilities and marital consistencies are greater when they have sons (33). Mothers may also prefer girls because of raising them easier or more satisfying companions (34).

Brockmann (2001) also debated that the value of having daughters are growing due to the fact that they are participating more in labor-force and burden of ageing is increasing (35).

Moreover, the improvement of girl preference might be foster because of changing prospects regarding the division of work and family duties in the family and a more positive assessment of women's role in society (36).

Due to the consequences of sex preference for children on couple's fertility behavior such as sex-selective abortions and abnormally changes of sex ratios, it is important to investigate determinants of sex preference. Most of researchers modeled sex preference by logistic regression, which was developed by statistician. Hank and Kohler (2002) studied sex preference by multinomial logistic regression according to predictors such as age, educational level, cultural variables and sex distribution of previous children in Germany (37). Rai et al. (2014) examined sex preference by multinomial and Binary logistic regression in Nepal by selected predictors such as age, number of children, sex of the last child, educational level, and job and economic status (38). Frempong and Codjoe (2013) considered age, place of residence, region, religion, job status, lineage, and education, as predictors to study sex preference in Ghanaians family by binary logistic regression (39). Mansurian and Khushnevis (2006) have considered age, education, the total number of children, marriage duration, place of residence, sex composition of children as nominated predictors and found their influence on fertility by applying logistic regression in Iran (18).

There are a number of reasons for difficulties of logistic regression to investigate sex preference such as poorly suitable for modeling many possible predictor variables, generally difficult to model interactions and no efficient procedure to handle missing data. Due to the advantages of the CART algorithm compared to logistic regression, in this article, the CART algorithm was applied to model sex preference of 1250 women aged 15-49. However, there is not any reference of applying CART algorithm to sex preference, Saadati and Bagheri (2015) and Bagheri and Saadati (2014, 2015) employed this algorithm to ideal number of children and CEB data (20-22).

Following results has been drawn from the extracted decision tree:

- Without considering any independent variables, women in this study preferred to have a son. This result is in favor of the results in countries such as India, China, and Korea. Pande and Astone (2007) discussed that this desire is extremely ingrained in social, economic and cultural elements (40).
- Educated women in young (10-29) and old (30-49) age groups preferred to have girls. The same result has been reported by Shahbaziyan et al. (2014) (17). Some authors such as Wongboonsin and Ruffalo, (1995) mentioned that variations in sex preference among countries and regions could be linked with factors involve the individual characteristics of parents, especially their level of education (41). The conclusion of Frempong and Codjoe (2013) was against girl preference. They resulted that Ghanaian women who had higher education had higher likelihood to prefer son (39).
- Women's age categories have also an important role on the resulted decision tree. Sex preference of women in young (10-29) and old (30-49) age groups were different. This is contrary to the findings of Westley and Choe's (2007) study in Pakistan (42). They suggested that young and old adults were less likely to prefer a daughter. Because sons are regarded as economic assets and security particularly during old age.
- Another important factor on women's sex preference in this study was the number of women's siblings. Some of the researchers such as Lyngstad and Prskawetz (2010) considered the number of siblings influenced on fertility (43). They measured social interaction through the cross-sibling influences on fertility of Norwegian families. The authors studied the data, which included the siblings' fertility, education, income, and marital histories. Kazemipour (2014) also highlighted to the influence of the number of women's siblings on their fertility. However, less attention has been devoted to the influence of women's siblings on their sex preference for children (31).

This article is extracted from a survey under the title of "Mining Demographic Data by Decision Tree" which is supported by National Population Studies and Comprehensive Management Institute in 2014 by the registered number of 20/15283. Also, the author would like to acknowledge Associate Professor Sahla Kazemipour as executer of Childbearing Attitudes and Its Social, Economic and Cultural Factors survey in 2014.

References

1. Fuse K. Variations in attitudinal gender preferences for children across 50 less-developed countries. Demogr Res. 2010; 23: 1031-48.
2. Edlund L, Lee C. Son preference, sex selection and economic development: theory and evidence from South Korea. Discussion papers. 2009;910(04).
3. Francis D.r. Do Fathers Prefer Sons? .2015 . http://www.nber.org/digest/oct04/ w10281.html.
4.Scharping T. Birth Control in China 1949-2000: Population policy and demographic development: Routledge; 2013.
5. Zorina ZA, Obozova TA. New data on the brain and cognitive abilities of birds. Zool Zhurnal. 2011;90(7):784-802.
6. Schloegl C, Schmidt J, Boeckle M, Weiss BM, Kotrschal K. Grey parrots use inferential reasoning based on acoustic cues alone. Proc R Soc Lond B Biol Sci. 2012;279(1745):4135-42.
7. Kugler AD, Kumar S. Preference for Boys, Family Size, and Educational Attainment in India. Demography. 2017;54(3):835-59.
8. Dahl GB, Moretti E. The demand for sons. Rev Econ Stu. 2008;75(4):1085-120.
9. Seidl C. The desire for a son is the father of many daughters. J Popul Econ. 1995;8(2):185-203.
10. Diekmann A, Schmidheiny K. Do Parents of Girls Have a Higher Risk of Divorce? An Eighteen- Country Study. J Marriage Fam. 2004;66(3):651-60.
11. Lundberg S. Sons, daughters, and parental behaviour. Oxford Rev Econ Pol. 2005;21(3):340-56.
12. Raley S, Bianchi S. Sons, daughters, and family processes: Does gender of children matter? Annu Rev Sociol. 2006;32:401-21.
13. Dahl E, Gupta RS, Beutel M, Stoebel-Richter Y, Brosig B, Tinneberg H-R, et al. Preconception sex selection demand and preferences in the United States. Fertil Steril. 2006;85(2):468-73.
14. Gray E, Evans A. Parity progression in Australia: What role does sex of existing children play? Aust J Soc Issues. 2005;40(4):505.
15. Gray E, Morrison NM. Influence of combinations of sexes of children on family size. J Hered. 1974;65:169-74.
16. Mishra V, Roy TK, Retherford RD. Sex differentials in childhood feeding, health care, and nutritional status in India. Populat develop Rev. 2004;30(2):269-95.
17. Shahbaziyan S. Golami, A., Shahbaziyan S. The Rule of Sex Preference in Fertility Behavior of Women in Kangavar, J Ilam Med Sci. 2014; 22(6):133-42.
18. Mansurian MK. Khushnevis, A. sex preference of ever married women on fertility behaiviour; a case study in Tehran. Soc Human Sci Shiraz J. 2006;47:perso129-46.
19. Hejazi. N. Employed Women's attitude to second child and its inflectional factors. Health System Res J. 2013;9(7):771-81.
20. Saadati M. Bagheri, A. Mining Children Ever Born Data; Classification Tree Approach. Indian J Sci Technol. 2015;8(30):1-8.
21. Bagheri A, Saadati M. Applying CART Decision Tree for Mining Fertility Behavior. Hakim Seyed Esmail Jorjani J. 2015;3(2):63-88.
22. Bagheri A, Saadati M. Introduction and Application of CART Model to Classify Ideal Number of Children for 15-49 Year-Old Women, Semnan providence. J Populat Associat Iran. 2014;17:77-111.
23. Timofeev R. Classification and regression trees (CART) theory and applications. Humboldt University, Berlin. 2004.
24. Breiman L, Friedman J, Olshen R. Classification and Regression Trees. Belmont, California: Wadsworth. 1984. Inc Ref Type: Generic.
25. Quinlan JR. C4. 5: programs for machine learning: Elsevier; 2014.
26. Kass GV. An exploratory technique for investigating large quantities of categorical data. Appl Statist. 1980:119-27.
27. Loh W-Y, Shih Y-S. Split selection methods for classification trees. Stat sinica. 1997:815-40.

28. Bhukya DP, Ramachandram S. Decision tree induction: an approach for data classification using AVL-tree. Int J Comput Elect Eng. 2010;2(4):660.
29. Lin N, Noe D, He X. Tree-based methods and their applications. Springer Handbook of Engineering Statistics: Springer; 2006. p. 551-70.
30. Gelfand SB, Ravishankar C, Delp EJ, editors. An iterative growing and pruning algorithm for classification tree design. Systems, Man and Cybernetics, 1989 Conference Proceedings, IEEE International Conference on; 1989: IEEE.
31. Kazemipour. Sh. Childbearing Attitudes and its social, economical and cultural factors, Statistical Research Center, Tehran, Iran.2014
32. Bhatia JC. Sex preference of children in rural Ghana (West Africa). Health Popul: Perspectives Issues. 1984;7(1):32-47.
33. Morgan SP, Lye DN, Condran GA. Sons, daughters, and the risk of marital disruption. Am J Sociol. 1988;94(1):110-29.
34. Marleau JD, Saucier J-F. Preference for a first-born boy in Western societies. J Biosoc Sci. 2002;34(1):13-27.
35. Brockmann H. Girls preferred? Changing patterns of sex preferences in the two German states. Eur Sociol Rev. 2001;17(2):189-202.
36. Hank K, Kohler H-P. Gender preferences for children in Europe: Empirical results from 17 FFS countries. Demogr Res. 2000;2.
37. Hank K, Kohler H-P. Gender preferences for children revisited: new evidence from Germany. at: ideas repec org/p/dem/wpaper/wp-2002-017 html (accessed 6 October. 2002.
38. Rai P, Paudel IS, Ghimire A, Pokharel PK, Rijal R, Niraula SR. Effect of gender preference on fertility: cross-sectional study among women of Tharu community from rural area of eastern region of Nepal. Reprod Health. 2014;11(1):15.
39. Frempong GA, Codjoe SNA.. Education and sex preference of children in Ghana, . Regional Institute for Population Studies. 2013; University of Ghana. P.O. Box LG 96. Legon, in XXVII IUSSP International Population Conference, Session 171: Poster Session on Fertility. Busan, Korea, 26-31 August 2013.
40. Pande RP, Astone NM. Explaining son preference in rural India: the independent role of structural versus individual factors. Popul Res Policy Rev. 2007;26(1):1-29.
41. Wongboonsin K, Ruffolo VP. Sex preference for children in Thailand and some other South-East Asian countries. Asia-Pac Popul J. 1995;10(3):43-62.
42. Westly S, Choe Kim M. How does son preference affect population in Asia. Asia Pac. 2007(84):1-12.
43. Lyngstad TH, Prskawetz A. Do siblings' fertility decisions influence each other? Demography. 2010;47(4):923-34.