Published online 2019 April 9.

Research Article



Application of Classification and Regression Trees in Iranian Women's Childbearing Desire

Arezoo Bagheri¹ and Mahsa Saadati^{1,*}

¹National Population Studies and Comprehensive Management Institute, Tehran, Iran

Received 2018 October 31; Revised 2019 February 17; Accepted 2019 February 20.

Abstract

Socio-economic factors, in addition to value and attitude changes, have had a major impact on fertility behavior of women toward childbearing. The main purpose of this study is to investigate childbearing desire of 6183 women in provinces with total fertility rate (TFR) lower (TFR \leq 2) and upper (TFR > 2) than replacement level by classification and regression trees (CART) algorithm, as one of the most applicable classification trees. In this study women's job, place of residence, age, opinion, marriage age, educational level, and children ever born were investigated as predictors. The accuracy of fitted classification trees for CD in provinces with TFR \leq 2 and TFR > 2 were 0.68 and 0.72, respectively. As a result, women's children ever born and age had significant roles on their desire to have a child in all provinces while their opinion toward childbearing had just effect on provinces with TFR \leq 2. In provinces with TFR \leq 2, 10 -39 years old women with 2 children and less CEB were willing to have another child. Women with 2 or less children at the age of 10 -29 as well as women with 3 or more than 3 children at the age of 20 -29 had a desire for more children in provinces with TFR \geq 2. If the women have a positive attitude toward childbearing and their age is not high, they will desire to give birth to a child until they reach the ideal number of their children. Thus, policies in changing their attitudes toward childbearing and creating the necessary facilities to prevent delays in marriage could be an important step in changing fertility rates.

Keywords: Total Fertility Rate, Childbearing Desire, Classification and Regression Trees (CART)

1. Background

Nowadays, most countries in the world are somehow in a state of population imbalance and are facing consequences of this problem. Some countries are struggling with problems of over-growth of the population, while the other communities are negatively affected by the decline in population growth. Thus demographic policies in each country can play an important role at reforming such a situation. Over a period of three decades, fertility in Iran has been astonishingly reduced. Since 2006, total fertility rate (TFR) or the number of children born alive, which is expected by each woman during her fertility course, was reached below the replacement level (1, 2).

There are several reasons and perspectives to explain fertility transitions. In many demographic studies, improving the status of women and women's autonomy were explained as influential factors in reducing fertility (3-5). Many studies have examined the relationship between women's status and their fertility behaviors, education and employment in the labor market as important indi-

cators of women's independence (6). Promoting women's education causes an increase in their employment opportunities and, as a result, increase their ability to decision making in their family which could be a reason for lower TFR (3).

To control this fertility decline and prevent more decreasing, conducting new studies with the aim of awareness and identification of women's desires toward child-bearing are necessary. In this regard, ideal number of children (INC), which is the desired number of children families like to bear, has recently considered to be an interesting issue for many researchers who studied determinants influenced fertility behavior of Iranian families (7-11). In most of these studies, researchers used logistic regression to investigate the influential factors on fertility rates.

However, the method that is used in recent decades due to the development of computer software for the classification of data is the decision tree, which is generally accepted due to its flexibility and specific features, especially its output, which is a graph and makes its interpretation simpler. Recently, some studies applied classification trees

^{*}Corresponding author: National Population Studies and Comprehensive Management Institute, Tehran, Iran. Email: mahsa.saadati@psri.ac.ir

for demographical data (10, 12).

Since the developmental level of different provinces of Iran in different studies has been effective in their fertility, ignoring the differences between provinces in fertility analysis leads to inaccurate results (13, 14). Therefore, in this study, provinces were divided into two categories based on their TFR in order to analyze data more precisely. These categories were constructed based on provinces' TFRs in Iran, which were calculated in the study of McDonald et al. (2). According to the values of TFR and replacement level, provinces were divided into two categories of TFR < 2, and TFR > 2. Childbearing desire (CD), which is resulted from this question "considering the number of children you have, do you desire to have more children?", was analyzed by applying one of the most applicable classification trees, CART, in this study. Many factors affect women's CD, which result in changing ideal family size such as the sex of siblings, marital satisfaction, and socioeconomic status of the family (15). According to the target population and measurement tool of this study, women's job status, age, opinion toward childbearing, marriage age, educational level, and children ever born (CEB) were investigated as predictors.

2. Methods

In this cross-sectional study, childbearing attitudes and its social, economic, and cultural factors were investigated by statistical research center under the supervision of Dr. Kazemipour in 2014. Here, a structured questionnaire was reviewed by 6231 married women aged from 15 to 49 years old in 31 provinces in Iran during autumn 2014 (15). In all provinces, a multi-stage stratified random sampling was used to select women who referred to public health and treatment centers to vaccinate their children based on the size and distribution of the population. A questionnaire was used to collect demographic, socio-economic, and fertility behavior characteristics of women, and only 16 questions (items) about value of children, including cultural questions (Opinin1: families with a child have a greater sense of happiness than childless families, Opinin3: childbearing strengthens the power of commitment of the parents, Opinin8: life without a child is cold and soulless, Opinin9: good children are a blessing and God will help for their expenditure, Opinin11: the existence of a child strengthens the family), social questions (Opinin4: these days if you have more than two children, people will blame you, Opinin13: these days if you don't have a child, people will blame you, Opinin10: childbearing spoil mother's physical fitness, Opinin16: spending money for what you may like is better to spend it for childbearing, Opinin5: childbearing is difficult and draws comfort from an individual, Opinin12: childbearing can spoil women's educational and career progress, which are more important than childbearing for them, Opinin2: these days many people prefer to have children later owing to lack of trust to their husbands), and economic questions (Opinin6: children will be parental support in oldage, Opinin14: concerns and uncertainties about the future make people unwilling for childbearing, Opinin15: if the government protects families, especially employed women, they will have more children, Opinin7: the cost of raising a child prevents parents to have a child).

These questions were scored from completely disagree (1) to completely agree (5) scores in the Likert scale in five points. In this study, the opinion toward childbearing was constructed from the value of children's questions as a categorical variable with three categories of negative, neutral, and positive. The reliability and content validity of questionnaire were confirmed by Cronbach's alpha, which at least 0.82 for each sub-item, and 10 demographers, respectively. In this study, women's CD were analyzed by CART algorithm based on selected predictors, including job status, age, opinion toward childbearing, marriage age, educational level, and children ever born (CEB) as the most influential factors, which postponed childbearing and could decrease demand for more children. Women's CD were assessed by a question that "considering the number of children already you have, do you desire to have another child (15)?"

The CART algorithm is one of the most applicable classification trees which extract binary splits. CART as a nonparametric statistical method established for classification problems (16). Based on type of dependent variable, categorical or continuous, classification or regression tree were created by CART algorithm, respectively. Extracting accurate set of data classifiers is the main purpose of CART algorithm. It is robust against the outliers and collinearities, can use both categorical and continuous variables, take into account missing data, detect interactions, and can be considered an exploratory analysis (16). The CART methodology is done in three phases of construction or building of maximum tree, selection of right tree size, and classification of new data (17). In this study, Gini index was used as a binary splitting rule; for node t and target variable by k categories, it is defined as Equation 1:

$$Gini(t) = 1 - \sum_{j=1}^{k} p^{2} [c = c_{j} | T = t]$$
 (1)

Where p is the probability that a node t belongs to C_j class and is estimated by $|C_{j,D}|/|D|$ (|D| is the size of D subset). The sum is computed over k categories (16). Classification tree can be extracted by a number of statistical software such as SPSS, SAS, R, and STATISTICA. The SPSS version22 was used to apply the CART algorithm to women's CD in this study.

3. Results

Out of 6183 married women aged from 15 - 49 years, 4898 (79.22%) and 1285 (20.78%) of them lived in provinces with TFR \leq 2 and TFR > 2, respectively. Also, 63.7% and 71.8% of women in provinces with TFR \leq 2 and TFR > 2, respectively expressed a desire for childbearing. Most of the women had a tendency of having children considering their CEB in this study (overall percentage is 65.4). Almost 14% of women were employed. More than 70% and 80% of women also lived in the urban area and were at the age of 20 - 39 years old. Women in the age groups of 10 - 19, 20 - 29, 30 - 39, and 40 - 49 had average age of 18.13 \pm 0.965, 25.51 \pm 2.586, 33.40 \pm 2.657, and 42.85 \pm 2.767 years old, respectively. Most of the women with negative (36.0%) and positive (39.8%) opinions toward childbearing were in provinces with TFR \leq 2 and TFR > 2, respectively. Most of the women (95%) married in their 10 - 29 years old. The percentage of high school and above-educated women in provinces with TFR ≤ 2 was 66.1% in comparison to the same group in provinces with TFR > 2 (60.3%). Childlessness is very low in the data (0.2%). The CEB of more than 85% of women was 1 - 2 children. Almost 23 against 13 percent of women in provinces with TFR > 2 comparing to TFR \le 2 had 3 and more children. Comparison of these variables between two province groups showed that there were significant differences between women's age, opinion toward childbearing, educational level, and CEB at the significance level of 0.01. Thus two different CART models were fitted.

The CD and selected predictors are shown in Table 1. The results of this table for TFR \leq 2 showed that variables of place of residence (P value = 0.01), age (P value < 0.01), opinion (P value < 0.01), and CEB (P value < 0.01), and for TFR > 2 variables of place of residence (P value = 0.035), age (P value < 0.01), opinion (P value = 0.005), and CEB (P value < 0.01) had significant association with CD. The women's educational level had different roles on CD according to province divisions. This variable had a significant association with CD in provinces with TFR \leq 2 (P value < 0.01) while no significant association with CD was observed in provinces with TFR > 2 (P value = 0.711).

Figures 1 and 2 indicate classification trees of CD in provinces with TFR \leq 2 and TFR > 2. Table 2 represents the misclassification matrix, which indicates the accuracy of classification models of CD in provinces with TFR \leq 2 and TFR > 2. Based on the shaded cells in Table 2, correct classification or accuracy of the classification trees in Figures 1 and 2 were obtained. The accuracy of the classification trees for these models can be calculated by Equations 2 and 3.

$$Accuracy = \frac{2856 + 489}{4898} = 0.68 \tag{2}$$

$$Accuracy = \frac{805 + 123}{1285} = 0.72 \tag{3}$$

Classification accuracy equals to 0.68 and 0.72 means that CD of 68% and 72% of women have been classified correctly (this value indicates that misclassifications of these models are equal to 32% and 28%).

In classification tree of CD in provinces with TFR \leq 2, only job status did not enter the tree while in another group, job status, residence, and marriage age were not displayed in the tree. As a result, women's children ever born and the age had a significant role in their desire to have a child in all provinces while their opinion toward childbearing had just impact on provinces with TFR \leq 2. The CEB was the most influential factor on CD as it was presented in the root of CART trees in Figures 1 and 2.

We can extract the rules 1 to 3 from the classification tree of CD in Figure 1:

- -Women at 10 39 years old with 2 children and less CEB desired to have more children.
- Women at 40 49 years old with 2 children, less CEB, and negative and neutral opinions in comparison to positive opinion toward childbearing were not willing to have more children.
- Women with 3 children and more CEB were not willing to have more children.

Rules 4 to 6 were also extracted from the classification tree of CD in Figure 2:

- -Women with 2 children and less desired for childbearing.
- -Women with more than 3 children at the age of 20 29 were willing to have another child.
- -Women with more than 3 children at the age of 30 49 did not express a desire for childbearing.

4. Discussion

Persistent low fertility is not desirable for any population. Given the consequences of continuous fertility below the replacement level, adopting programs to prevent

Table 1. Childbearing Desire Crossed by Selected Predictors Variables Childbearing Desire (Response Variable) Childbearing Desire (Response Variable) Test Statistic Test Statistic P Value P Value Total Sample — Total Sample Yes Iob status 0.462^a 0.497 0.553 62.6 37.4 Employed 676 69.9 30.1 173 Unemployed 36.1 4222 72.1 27.9 1112 63.9 Place of residence 6.56^a 0.010 4.46^a 0.035 62.6 37.4 3517 70.2 29.8 918 66.5 Rural 33.5 1381 76.0 24.0 367 301.27^b 70.82b Age < 0.01 < 0.01 20 - 29 25.9 2285 81.2 18.8 62.5 30 - 39 44.4 2120 37.5 563 62 40 - 49 62.2 365 50.0 37.8 50.0 10.52^a 87.41 < 0.01 0.005 Opinion toward childbearing 44.0 65.7 367 1763 34.3 Neutral 64.7 72.7 35.3 1557 27.3 407 Positive 71.5 28.5 1578 75.5 24.5 511 Marriage age 1.98b 0.29^b 0.159 62.0 38.0 1790 71.5 28.5 474 35.0 28.5 754 30 - 39 61.3 38.7 271 80.4 19.6 56 40 - 49 85.7 14.3 0.0 100.0 28.12^b 0.14^b **Educational level** < 0.01 0.711 Illiterate 47.4 52.6 116 80.6 19.4 62 Primary and middle school 40.1 1542 70.3 29.7 448 59.9 34.8 1897 70.1 29.9 High school/diploma University 67.5 32.5 1343 74.8 25.2 317 Children ever born (CEB) 250.43^b 87.5 12.5 100.0 0.0 1-2 67.9 32.1 4274 77.6 22.4 987 3 - 4 34.8 65.2 584 52.2 47.8 245 32 49 31.3 68.8 51.0 49.0

b Linear by linear association.

persistent fertility reduction and to increase it at least to the replacement level are necessary. Any policy to increase fertility or prevent its further reduction requires knowing and understanding the factors and conditions affecting fertility (3). The main purpose of this study was to investigate the influential factors on CD. Some studies such as Saadati and Bagheri concluded about dividing provinces according to TFR and get different results in different divisions. In this study, the provinces were divided into two groups with TFR \leq 2 and TFR > 2 (12, 13). Provinces with TFR \leq 2 in comparison to TFR > 2 had less desire to have the third child.

Almost 70% of women with 1 or 2 children in all provinces desired for childbearing while most of the

women with more than 2 children did not have any willing to have another child (Table 1). In this study, CEB was the most influential factor on CD as it was presented in the first layer of CART trees in Figures 1 and 2. In agreement with other studies, it can be concluded that by increasing the number of CEB, the tendency toward childbearing is decreased (10,18,19). It is important to mention that women's age also played a significant role in CD. In provinces with TFR \leq 2, most of the women with lower CEB (2 and fewer children) in older ages (40 - 49 years old) did not have a tendency toward childbearing. Moreover, most women with higher CEB (3 and more children) in older ages (30 - 49 years old) did not desire to have another child in provinces with TFR > 2. These results are important evidence of the

^a Pearson chi-square test.

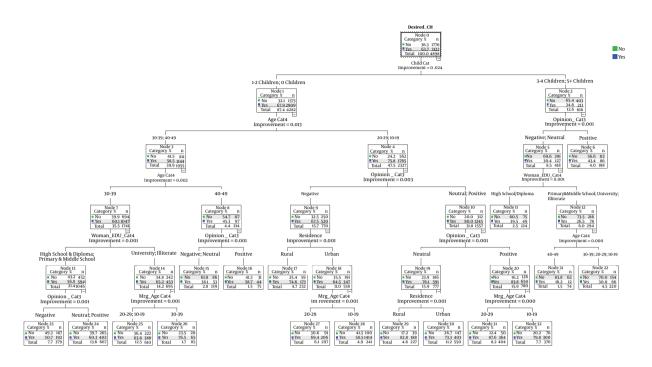


Figure 1. Classification tree of CD in provinces with TFR \leq 2

Observed Category	Predicted Category		T-4-1
	Yes	No	Total
ΓFR ≤ 2 (Figure 1)			
	2856 (58.32)	266 (5.42)	3122 (63.74)
	1287 (26.28)	489 (9.98)	1776 (36.26)
	4143 (84.59)	755 (15.41)	4898 (100)
TFR > 2 (Figure 2)			
	805 (62.64)	118 (9.18)	923 (71.82)
	239 (18.60)	123 (9.57)	362 (28.17)
	1044 (81.24)	241 (18.76)	1285 (100)

^a Values are expressed as No. (%).

influential role of women's age besides CEB. Abbasi-Shavazi and Khajesalehi (18), Hosseini and Baigi (20), and Hejazi (21) obtained similar results. They reported that by increasing women's age the desire for more children declines.

On the basis of this study, opinions which consider cultural, economic and social components somewhat affect the reduction of women's desires in provinces with TFR \leq 2. Hosseini and Beigi (20) also presented similar results. Women's negative opinions, including reducing the sense

of happiness with the presence of the child, blaming by others when having more children, the cost of childbearing, and children are obstacles to work and educational progress could diminish the desire to have more children. In the studies of Abbasi-Shavazi and khajesalehi (18), and Hejazi (21), the educational level did not have a significant impact on willingness to childbearing similar to the results of this study. It seems that other issues are involved in the context of the relationship between education and childbearing. In other words, education alone cannot have a significant impact on the need for motherhood, fatherhood, and childbearing. The other factors such as CEB and age, in addition to the direct expected effects, may play the role of moderating in relationship between education and childbearing.

In this study, women's residence place did not have a significant influence on CD in all provinces.

This result is consistent with another study (20). Also, the marriage age is supposed to be one of the factors affecting fertility in the results of this study. Increasing the average marriage age is always reported as one of the major reasons for low fertility (15). In Kazemipour's study, in 2014 (15), fertility attitude and marriage age of youth in marriage age and 15 - 49 married women had inverse re-

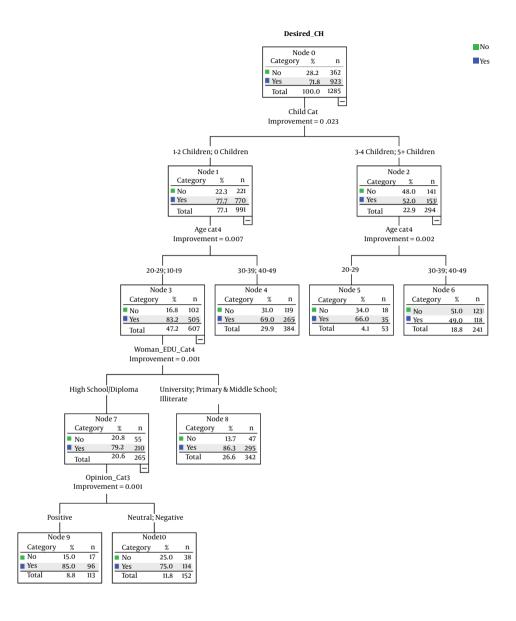


Figure 2. Classification tree of CD in provinces with TFR > 2

lations. This result is similar to the results of Kariman et al. study (22). They mentioned that women who married in the higher ages compared with the others had less CD. However, in this study similar to the results of educational level, this variable alone cannot have a significant impact on CD, which was mentioned by Hosseini et al. (19) as well. A lot of studies emphasized the probability of unwillingness of employed women in comparison to unemployed one to have a child. While in this study, consistent with

Hosseini and Baigi study (20), this variable was not significant in CD. Also, the marriage age is one of the factors affecting fertility according to the results of this study in provinces with TFR \leq 2 while it is in contrast to the univariate results of Table 1. Increasing the average marriage age is always reported as one of the major reasons for low fertility (15).

In conclusion, since the most important variables affecting the desire of women to have a child were the num-

ber of currently born children (CEB), their age, and their opinion toward childbearing, policymakers should provide new insights to strengthen women's positive opinion toward childbearing, and they marry in younger ages to have suitable time for reaching their ideal number of children.

Footnotes

Conflict of Interests: The authors declare that they have no conflict of interest.

Funding/Support: This study was supported by National Population Studies and Comprehensive Management Institute in 2014 with the registered number of 20/15283.

References

- Salehi-Isfahani D, Abbasi-Shavazi MJ, Hosseini-Chavoshi M. Family planning and fertility decline in rural Iran: The impact of rural health clinics. *Health Econ*. 2010;19 Suppl:159–80. doi: 10.1002/hec.1613. [PubMed: 20552711].
- McDonald P, Hosseini-Chavoshi M, Abbasi-Shavazi MJ, Rashidian A. An assessment of recent Iranian fertility trends using parity progression ratios. *Demograp Res.* 2015;32(58):1581-602. doi: 10.4054/Dem-Res.2015.32.58.
- Abbasi-Shavazi MJ, McDonald P, Hosseini-Chavoshi M. The fertility transition in Iran: Revolution and reproduction. Netherlands: Springer; 2009. doi: 10.1007/978-90-481-3198-3.
- 4. Hosseini H, Torabi F, Bagi B. Demand for long-acting and permanent contraceptive methods among Kurdish women in Mahabad, Iran. *J Biosoc Sci.* 2014;**46**(6):772–85. doi: 10.1017/S0021932013000710. [PubMed: 24406051].
- Kariman N, Simbar M, Ahmadi F, Vedadhir AA. Concerns about one's own future or securing child's future: Paradox of childbearing decision making. *Health*. 2014;6(10):1019–29. doi: 10.4236/health.2014.610128.
- 6. Erfani A, McQuillan K. The changing timing of births in Iran: An explanation of the rise and fall in fertility after the 1979 Islamic revolution. *Biodemography Soc Biol.* 2014;**60**(1):67–86. doi: 10.1080/19485565.2014.899428. [PubMed: 24784988].
- Dallaire DH, Pineda AQ, Cole DA, Ciesla JA, Jacquez F, Lagrange B, et al. Relation of positive and negative parenting to children's depressive symptoms. *J Clin Child Adolesc Psychol*. 2006;35(2):313–22. doi: 10.1207/s15374424jccp3502_15. [PubMed: 16597227]. [PubMed Central: PMC3152307].

- Erfani A, McQuillan K. Rapid fertility decline in Iran: Analysis of intermediate variables. *J Biosoc Sci.* 2008;40(3):459-78. doi: 10.1017/S002193200700243X. [PubMed: 17850688].
- Abbasi-Shavazi MJ, Torabi F. Women's education and fertility in Islamic countries. In: Groth H, Sousa-Poza A, editors. *Population dynamics in Muslim countries*. Berlin, Heidelberg: Springer; 2012. p. 43–62.
- Bagheri A, Saadati M, Razeghi Nasrabad HBB. [Introduction and application of CART model for classifying 15 49 year old women ideal number of children in Semnan province]. J Populat Associat Iran. 2014;9(17):77-111. Persian.
- Kariman N, Simbar M, Ahmadi F, Vedadhir AA. Socioeconomic and emotional predictors of decision making for timing motherhood among Iranian women in 2013. *Iran Red Crescent Med J.* 2014;16(2). e13629. doi: 10.5812/ircmj.13629. [PubMed: 24719733]. [PubMed Central: PMC3965866].
- Saadati M, Bagheri A. Educated Iranian women in favor of having girls: CART classification approach. European Population Conference. Mainz, Germany. 2016.
- Saadati M, Bagheri A. [Study of ideal marriage interval to childbearing in terms of youth at the threshold of marriage]. *Payesh*. 2016;17(2):239–50. Persian.
- Bagheri A, Saadati M. [Analysis of the childlessness ideal survival time of young's at the threshold of marriage: The parametric log normal model]. Pajoohande. 2016;21(4):199-209. Persian.
- Kazemipour S. Childbearing attitudes and its social, economic and cultural factors. Tehran: Statistical Research Center; 2014.
- Breiman L. Classification and regression trees. New York: Routledge: Tailor & Francis Group; 2017.
- Han J, Pei J, Kamber M. Data mining: Concepts and techniques. Elsevier Science; 2011.
- Abbasi-Shavazi MJ, Khajesalehi Z. [An assessment on the impact of women's autonomy education and social participation on childbearing intention in Sirjan city]. Women Dev Politic. 2013;11(1):45–65. Percian
- Hosseini H, Asgari-Nadushan A, Moradi N. [Comparative study of childbearing desires Shiite and Sunni Kurdish women in rural areas of Kamyaran city]. J Woman Fam Stud. 2016;4(1):63–84. Persian.
- Hosseini H, Baigi B. [Economic, social, cultural and demographic determinants of childbearing tendencies of married women referring to Hamedan health centers in 2012]. J Kermanshah Univ Med Sci. 2014;18(1):35–43. Persian.
- 21. Hejazi N. [The attitude of employed women to bring the second child and the factors affecting it]. [Health Sys Res. 2013;9(7):771-81. Persian.
- Kariman N, Amerian M, Jannati P, Salmani F. Factors influencing first childbearing timing decisions among men: Path analysis. *Int J Reprod Biomed (Yazd)*. 2016;14(9):589–96. [PubMed: 27738661]. [PubMed Central: PMC5054296].