

◆ *Original Research Article*

**Provincial Analysis of Women's Universality of Marriage in Iran:
Application of Machine Learning in Modeling Survival Function**

Mehdi Khalili¹, Fatemeh Torabi^{2*}

Abstract

Various studies have examined the transformations of marriage at the national and regional levels, but there are still unknowns at the provincial level within the country. The purpose of this article is to investigate the universality of women's marriage in the provinces of the country using the survival random forest method, one of the common algorithms in machine learning. Using the household expenditure and income survey data from 2008 to 2019, the general status of women's marriage was investigated by defining a threshold level (marriage rate less than 95% for women aged 45-49). The findings of this research indicate that, among the research variables, high school and university education levels had the greatest impact on reducing the universality of women's marriages. The proportion of urbanization in each province is another influential variable in reducing the universality of women's marriages. Out of the total of 31 provinces of the country, 11 provinces still had the universality of marriage until the end of the study period, but the marriage rate of women aged 45-49 in other provinces has decreased to below 95%. The high marriage rates in most of the provinces that do not enjoy the universality of marriage now indicate the existence of a suitable platform for facilitating marriage and expanding its universality in the country. Otherwise, it is expected that this trend will continue to decrease in the future.

Keywords: Universality of marriage, Random forest, Machine learning, Survival function, Education, Urbanization

Received: 2023-09-19

Accepted: 2023-11-03

1 PhD Candidate in Demography, University of Tehran, Tehran, Iran: m.m.khalilii13@gmail.com

2* Professor, Department of Demography, University of Tehran, Tehran, Iran (Corresponding Author): fatemeh_torabi@ut.ac.ir

E-ISSN: 2008-3742 / © Population Association of Iran. This is an open access article under the CC BY 4.0 License (<https://creativecommons.org/licenses/by/4.0/>).

DOI: <https://doi.org/10.22034/jpai.2024.2011980.1308>

Extended Abstract

Introduction

Over the past four decades, demographic, socio-economic, and cultural changes have significantly impacted the age at first marriage, particularly for women. The increasing delay in the age of first marriage is well-documented in many studies, with education frequently cited as a key factor. Additionally, marriage squeeze and shifting nuptiality patterns have contributed to changes in marriage trends among Iranian women. In this article, we analyze the evolution of women's marriage patterns at the provincial level, highlighting cultural differences across Iran's provinces. While overall trends are well understood, many provincial-level nuances remain unexplored. The unequal employment opportunities and educational attainment level among different provinces doubles the importance of provincial surveys on marriage. Many other studies in Iran have addressed the importance of the role of education in delaying women's marriage; however, there are still many unknowns regarding provincial differences. This article attempts to address the effect of various provincial variables and measure the delay in women's marriage according to different economic and social conditions.

Method and data

This study utilizes the Household Expenditure and Income Survey (HEIS) to analyze changes in the universality of marriage among women across different provinces in Iran. For modeling the survival function, we employed the random survival forest, a robust machine-learning algorithm for analyzing the data. Three independent variables—education, urbanization, and total labor force participation—were used to explore provincial differences in the universality of marriage. All data and algorithms were implemented in Python. This method, with high accuracy and examining the possible states of a variable, allows for a detailed examination of the changes in the dependent variable. The survival function will also be estimated more accurately. To achieve this goal, the research data was first trained with the desired algorithm and then the model was applied using test data.

Findings

Our findings indicate that the universality of marriage has significantly changed over the past decades, with substantial provincial-level variations. The proportion of married women in the 45-49 age group in all provinces of the country in 2009 was

above 95%, indicating the high prevalence of marriage among all provinces of the country. The highest proportion of married women aged 45-49 among the provinces of the country is in the provinces of Sistan and Baluchestan, Kohgiluyeh and Boyer Ahmad, Yazd, Khorasan Razavi, South Khorasan, and Hormozgan. In contrast, the lowest proportion is observed in the provinces of Gilan, Markazi, and Lorestan. Therefore, until the beginning of 2009, all provinces still enjoyed the universality of marriage. Between 2009 and 2020, the universality of marriage among women declined in many provinces, and by the end of 2020, approximately 20 provinces had experienced a universality rate below 95%. The results of the multivariate analysis further show that, among other factors, secondary education had the most significant impact on the decline in women's universality of marriage. Examining the effect of independent variables on the dependent variable shows that the variables of high school education, university education, urbanization, and labor force participation will have the greatest impact on the dependent variable, respectively. As a result, across different provinces, the role of education and the level of economic participation continues to be important in women's marital status.

Conclusion and discussion

The provincial analysis of the universality of women's marriage in Iran reveals that its decline began in 2008. Before this, there was no significant change, particularly in less developed provinces. Differences in women's access to education and varying levels of urbanization and economic participation have contributed to the uneven decline in the marriage trend. The survival function calculated in this study shows that by the end of 2019, 11 provinces—Gilan, Mazandaran, West Azerbaijan, Kerman, Razavi Khorasan, Sistan and Baluchistan, Kurdistan, Yazd, Qom, Golestan, and South Khorasan—maintained universality in marriage. In contrast, the remaining provinces experienced declines at various points during the study period, with the proportion of married women aged 45-49 falling below 95%. The analysis of variables influencing the survival function indicates that high school and university education significantly reduced the universality of marriage in the provinces. Following education, urbanization emerged as the second most important factor explaining the decline. Despite these shifts, in 22 provinces, the proportion of married women aged 45-49 remains relatively high. Although this ratio has decreased, many provinces have a marriage rate above 90%.

Citation:

Khalili, M. & Torabi, F. (2024), Provincial Analysis of Women's Universality of Marriage in Iran: Application of Machine Learning in Modeling Survival Function, *Journal of Population Association of Iran*, 19(38), 143-177.

<https://doi.org/10.22034/jpai.2024.2011980.1308>

ارجاع:

خلیلی، مهدی؛ ترابی، فاطمه (۱۴۰۳). تحلیل استانی عمومیت ازدواج زنان در ایران: کاربرد یادگیری ماشین در مدل‌سازی تابع بقا، نامه انجمن جمعیت‌شناسی ایران، ۱۹(۳۸)، ۱۷۷-۱۴۳.

<https://doi.org/10.22034/jpai.2024.2011980.1308>

تحلیل استانی عمومیت ازدواج زنان در ایران: کاربرد یادگیری ماشین در مدل‌سازی تابع بقا

مهدی خلیلی^۱، فاطمه ترابی^{۲*}

چکیده

مطالعات مختلفی دگرگونی‌های ازدواج را در سطوح ملی و منطقه‌ای بررسی کرده‌اند، اما در سطح استان‌های کشور هنوز ناشناخته‌هایی وجود دارد. هدف این مقاله بررسی وضعیت عمومیت ازدواج زنان در استان‌های کشور با استفاده از روش جنگل تصادفی بقا به‌عنوان یکی از الگوریتم‌های رایج در یادگیری ماشین است. با استفاده از داده‌های طرح هزینه و درآمد خانوار از سال ۱۳۸۸ تا ۱۳۹۹ وضعیت عمومیت ازدواج زنان با تعریف یک سطح آستانه‌ای (نسبت ازدواج کمتر از ۹۵٪ برای زنان ۴۹-۴۵ ساله) مورد بررسی قرار گرفت. یافته‌های این تحقیق حاکی از آن است که در بین متغیرهای تحقیق، سطوح تحصیلی دبیرستان و دانشگاه بیشترین تأثیر را بر کاهش عمومیت ازدواج زنان داشته است. نسبت شهرنشینی در هر استان نیز متغیر تأثیرگذار دیگری بر کاهش عمومیت ازدواج زنان است. از مجموع ۳۱ استان کشور، ۱۱ استان تا پایان دوره مطالعه همچنان از عمومیت ازدواج برخوردار بودند اما نسبت ازدواج زنان ۴۹-۴۵ ساله در سایر استان‌ها به پایین‌تر از ۹۵٪ کاهش یافته است. بالا بودن نسبت‌های ازدواج در اکثر استان‌هایی که اکنون از عمومیت ازدواج برخوردار نیستند، نشان‌دهنده وجود بستر مناسب برای تسهیل ازدواج و گسترش عمومیت آن در کشور است. در غیر این صورت انتظار می‌رود این روند در آینده نیز همچنان کاهش باشد.

واژگان کلیدی: عمومیت ازدواج، جنگل تصادفی، یادگیری ماشین، تابع بقا، تحصیلات، شهرنشینی، ایران.

تاریخ پذیرش: ۱۴۰۲/۰۸/۱۳

تاریخ دریافت: ۱۴۰۲/۰۶/۲۹

۱. دانشجوی دکتری جمعیت‌شناسی، دانشگاه تهران، تهران، ایران؛ m.m.khalilii13@gmail.com

۲. استاد گروه جمعیت‌شناسی، دانشگاه تهران، تهران، ایران (نویسنده مسئول)؛ fatemeh_torabi@ut.ac.ir

DOI: <https://doi.org/10.22034/jpai.2024.2011980.1308>

مقدمه و بیان مسئله

گمانه‌زنی‌ها درباره علل تأثیرگذار بر تاخیر ازدواج زنان در ادبیات جمعیت‌شناسی و سایر رشته‌ها بسیار است. در جمعیت‌شناسی اما به واسطه نقش مستقیم ازدواج بر تحولات باروری، به ازدواج زنان و تغییرات آن توجه ویژه‌ای شده است. افزایش قابل توجه در دستیابی به فرصت‌های تحصیلی زنان در سال‌های بعد از انقلاب همگام با ورود نسل‌های متولدین دوره بیش‌زایی^۱ به سن ازدواج و باروری، مطالعه نقش تحصیلات در اثرگذاری بر تغییرات ازدواج و باروری را به مقصدی مهم برای محققان تبدیل کرده است. رشد جمعیت تحصیل‌کرده‌ها و تغییر ترکیب جمعیت بی‌سواد به نفع جمعیتی که نسبت به گذشته باسوادتر است، یکی از دلایل مهم در توجه ویژه به نقش تحصیلات است. هر چند تحصیلات به خودی خود در کاهش یا افزایش باروری فرضیه تأیید شده‌ای به شمار می‌رود، اما بیشتر از طریق افزایش سن ازدواج و زمان‌بندی^۲ فرزندآوری به خصوص در کشورهای دارای میزان بالای باروری نکاحی عمل می‌کند (Jones and Gubhaju, 2009). در همین راستا، در دهه‌های اخیر فرضیات متعددی در خصوص نحوه تأثیر تحصیلات بر خانواده و ازدواج و رابطه آن با تحولات جمعیتی شکل گرفته است.

در دهه‌های اخیر، تحولات اقتصادی و اجتماعی در کنار گذار جمعیتی، همانند سایر نهادها، نهاد خانواده و ازدواج را نیز تحت تأثیر قرار داده است. از دیدگاه نظریه‌پردازان نوسازی، تحولات در خانواده به تغییرات در زندگی اجتماعی وابسته است که متأثر از سه جریان صنعتی شدن، شهرنشینی و آموزش همگانی است. بنابراین، تغییرات ساختار خانواده و ازدواج در انطباق با تغییرات شرایط اجتماعی و اقتصادی صورت گرفته است (عباسی شوازی و صادقی، ۱۳۸۴). با توجه به این تغییرات، در بسیاری از نقاط دنیا سن ازدواج تغییر کرده است و همگام با شرایط

1. Baby boom
2. Timing

جامعه هر روز در حال تغییر است (Bloom, 2020). در نتیجه در بسیاری از نقاط دنیا در چند سال اخیر، بخش قابل توجهی از تحقیقات معطوف به دگرگونی‌های ازدواج و خانواده بوده است (Torabi and Baschieri, 2010; Rallu, 2006; Malhotra, 1997; Muhsam, 1974). در ایران نیز، میانگین سن ازدواج بار اول زنان و مردان از سال ۱۳۴۵ تا ۱۳۹۵ تغییرات قابل توجهی داشته است به طوری که این مقدار برای مردان از ۲۵ سال در سال ۱۳۴۵ به ۲۷ سال در سال ۱۳۹۵ رسیده و برای زنان نیز از ۱۸ سال در سال ۱۳۴۵ به ۲۳ سال در سال ۱۳۹۵ افزایش یافته است (فتحی و همکاران، ۱۳۹۶). همان‌طور که مشاهده می‌شود تغییرات در میانگین سن ازدواج برای هر دو جنس رخ داده است، اما برای زنان محسوس‌تر بوده چرا که میانگین سن ازدواج در مردان، ۲ سال و در زنان ۵ سال افزایش یافته است.

بررسی‌ها در تبیین عوامل افزایش سن ازدواج به خصوص برای زنان در دهه‌های قبل نقش مؤثر تحصیلات و عوامل اقتصادی را نشان می‌دهد (ترابی و شمس قهفرخی، ۱۴۰۰). آموزش و تحصیلات می‌تواند زنان را در معرض بسیاری از ایده‌ها و نگرش‌های جدیدی قرار دهد که به نوعی آن‌ها را از ازدواج دور می‌کند (Ikamari, 2005; Zhao, 2017). بنابراین بخش قابل توجهی از تأخیر در ازدواج زنان مربوط به تغییرات نگرشی حاصل از تحصیلات است. به طوری که زنان با تحصیلات بالا کمتر در سنین پایین‌تر ازدواج می‌کنند و اوج ازدواج آنان در سنین بالاتر است (Goldestein and Kenny, 2001). این تغییر الگوی سنی ازدواج از سنین پایین به سنین بالاتر شاید مهم‌ترین خروجی اثر تحصیلات بر ازدواج زنان باشد. به عبارت دیگر، با افزایش نسبت زنان تحصیل کرده در مقایسه با گذشته بخش قابل توجهی از بازار ازدواج منعکس کننده الگوی ازدواج زنانی است که هم تحصیل کرده هستند و هم الگوی ازدواج متأخری را تجربه می‌کنند. نسبت‌های متفاوت تحصیلی و شغلی در میان مردان و زنان، علاوه بر تأثیرگذاری بر بازار ازدواج می‌تواند به الگوی ازدواج متفاوت در یک جامعه تبدیل شود. برای مثال وجود مضیقه ازدواج در افراد

تحصیل کرده باعث می‌شود برخی از این افراد با گروه‌های سنی دیگر و حتی نژادهای دیگر ازدواج کنند (Hirchman and Matras, 1971). این الگوی متأخر ازدواج زنان در سال‌های اخیر دغدغه علمی و سیاست‌گذاری مهمی را مبنی بر تأخیر در ازدواج زنان و در نهایت کاهش ازدواج در ایران ایجاد کرده است. یافته‌های رازقی نصرآباد و رحیمی (۱۳۹۳) افزایش میانگین سال‌های تجرد زنان و مردان ایران در طول دوره‌های ۱۳۸۵ و ۱۳۹۰ و نقش تأثیرگذار تحصیلات در این زمینه را نشان می‌دهد. همچنین بسیاری از تحقیقاتی که در ایران در تبیین ازدواج زنان انجام شده است به نقش تحصیلات تأکید کرده‌اند (کاظمی‌پور، ۱۳۸۳؛ رضادوست و ممبئی، ۱۳۹۰؛ رحیمی و همکاران، ۱۳۹۴؛ محمودیان، ۱۳۸۳؛ Torabi et al., 2013).

در کنار تحصیلات متغیرهای دیگری را نیز می‌توان برشمرد که بر روند تغییرات ازدواج زنان ایران مؤثر بوده‌اند. ویژگی متفاوت شهرهای صنعتی، فضای شغلی و دستیابی متفاوت به امکانات مختلف می‌تواند منجر به شکل‌گیری انواع جدید خانواده و نوع متفاوت ازدواج به خصوص برای زنان باشد (Puschmann and Solli, 2014). با گذشت چندین دهه از تحولات اقتصادی اجتماعی ایران، تحقیقات مختلف نشان‌دهنده اثر این تغییرات بر الگوی ازدواج است. با گسترش شهرنشینی در ایران، میانگین سن در ازدواج اول برای افراد شهرنشین به مراتب بالاتر از این مقدار در مناطق روستایی است (بگی، ۱۴۰۱). شهرنشینی در بسیاری از نقاط ایران به معنای ایجاد فرصت‌های تحصیلی و شغلی بیشتر به خصوص برای زنان است. بنابراین، افزایش سال‌های حضور در مدرسه برای زنان مناطق شهری به طور قابل توجهی بیشتر از زنان روستایی است. در نتیجه، الگوی ازدواج زنان در مناطق روستایی و شهری نشان می‌دهد که زنان روستایی در ایران بسیار زودتر از زنان شهری ازدواج می‌کنند (Saadati and Bagheri, 2017). همچنین استقلال زنان، میزان بیکاری مردان و وضعیت اشتغال زنان نیز از عوامل مهم تأخیر در ازدواج دختران در بین شهرستان‌های کشور به شمار می‌رود (حسینی و همکاران، ۱۳۹۷). تفاوت سن ازدواج زنان بر حسب فعالیت آنان ممکن است ناشی از آن باشد که زنان شاغل احتمال ازدواج در سنین بالاتر را نیز برای خود

متصور هستند لذا در انتخاب همسر و سواس بیشتری به خرج می‌دهند و همین امر به دیر ازدواج کردن آنان خواهد انجامید (عسکری‌ندوشن و همکاران، ۱۳۹۷).

علاوه بر ویژگی‌های اقتصادی-اجتماعی که در بالا بدان اشاره شد، ویژگی‌های فرهنگی و تفاوت‌های نگرشی مهمی وجود دارد که می‌تواند تبیین‌کننده تغییرات ازدواج زنان در ایران باشد. در بافت فرهنگی و قومی

بنابر آنچه که در بالا گفته شد، فضای کلی تحقیقات داخلی و خارجی در خصوص عوامل مؤثر بر تغییرات ازدواج زنان به خوبی روشن است. تحصیلات زنان یکی از مهم‌ترین عواملی است که در کاهش و تأخیر ازدواج زنان بدان تأکید شده است. در کنار تحصیلات، شهرنشینی و عوامل اقتصادی و فرهنگی نیز توانسته بخشی از دگرگونی‌های ازدواج زنان را تبیین کند. با این وجود، بخش مبهم ماجرا به تفاوت‌های استانی و نقش این عوامل بر تغییرات ازدواج زنان در سطح استان‌های کشور برمی‌گردد. در سطح کلی برآوردها از کاهش عمومیت ازدواج و در نتیجه افزایش تجرد قطعی زنان در سال‌های آینده خبر می‌دهد (کوششی و خلیلی، ۱۳۹۹). در سطح شهرستان استان‌های کشور نیز برخی از مطالعات بر تفاوت در نسبت‌های تجرد صحنه گذاشته‌اند. نتایج تحلیل‌ها در سطح شهرستان‌های کشور نشان می‌دهد که شهرستان‌های غرب و جنوب غرب کشور بیشترین میزان تجرد و شهرستان‌های مرکز و شمال شرق کشور کمترین میزان تجرد در دو جنس را تجربه می‌کنند (امامی و همکاران، ۱۴۰۰). اما در سطح استانی اطلاعات دقیقی درباره این موضوع در دسترس نیست. مقاله حاضر با تکیه بر چنین خلاء پژوهشی، سعی در بررسی عوامل تأثیرگذار بر تغییرات ازدواج زنان در سطح استان‌های کشور دارد. سوال اصلی این پژوهش آن است که عواملی که در مطالعات پیشین در تبیین تغییرات ازدواج بدان اشاره شده است آیا در سطح استانی هم صادق است؟ همچنین استان‌های مختلف چه تفاوتی در تجربه چنین تغییراتی دارند. فرض اصلی این مطالعه آن است که، استان‌های مختلف به واسطه

سطح توسعه‌یافتگی و شرایط فرهنگی-اجتماعی مختلف ممکن است تجربه متفاوتی از تغییرات ازدواج زنان و تأثیرپذیری از متغیرهای مهم در این زمینه داشته باشند.

روش و داده‌های تحقیق

روش پژوهش حاضر تحلیل ثانویه داده‌های موجود است. داده‌ها و مدل در زبان برنامه‌نویسی پایتون^۱ نوشته و اجرا شده است. برای بررسی تغییرات عمومیت ازدواج در بین استان‌های کشور از داده‌های طرح هزینه درآمد خانوار در بین سال‌های ۱۳۸۸ تا ۱۳۹۹ استفاده شده است. تعداد مشاهدات کل تحقیق ۴۰۲ عدد بوده است که برابر است با تعداد ۳۱ استان از سال ۱۳۸۸ تا ۱۳۹۹. نسبت ازدواج کرده‌های زنان ۴۹-۴۵ ساله در هر استان در بین سال‌های مذکور محاسبه شد و سپس کاهش نسبت زنان ازدواج کرده‌ی هر استان به زیر ۹۵٪ به‌عنوان وقوع رخداد عدم عمومیت ازدواج در نظر گرفته شد. انتخاب این سطح آستانه‌ای برای مبنا قرار دادن کاهش عمومیت ازدواج براساس مطالعه جونز (۲۰۱۵) است که این سطح را ملاک کاهش عمومیت ازدواج می‌داند. ملاک شروع تحلیل با سال ۱۳۸۸ نیز از طریق بررسی نسبت زنان ازدواج کرده ۴۹-۴۵ ساله در سال‌های قبل از آن انتخاب شده است. ابتدا از سال ۱۳۸۰ نسبت زنان ازدواج کرده در گروه سنی ۴۹-۴۵ سال ارزیابی شد که براساس آن تغییر قابل توجهی در بین استان‌های کشور تا قبل از سال ۱۳۸۸ در این نسبت مشاهده نشد و در نتیجه این سال به‌عنوان سال مبدا انتخاب شد. در بین استان‌های کشور، تنها استان البرز از سال ۱۳۸۹ بر مبنای تقسیمات جدید کشوری به استان‌های کشور اضافه شده است. برای سال قبل از آن نسبت ازدواج کرده‌های زنان تهران برای استان البرز اعمال شده است.

متغیرهای مستقل تحقیق شامل تحصیلات، شهرنشینی و نرخ مشارکت اقتصادی کل است. سطح تحصیلات زنان به سه گروه بی‌سواد و ابتدایی راهنمایی، دبیرستان و دانشگاهی تقسیم شده

1. python

است. به دلیل عدم دسترسی به داده‌های سری زمانی در خصوص نرخ مشارکت اقتصادی زنان در بازه ۱۳ ساله مطالعه، از نرخ مشارکت اقتصادی کل استفاده شده است. در ادامه روش جنگل تصادفی بقاء و مدل‌سازی تابع بقاء که در این مقاله مورد استفاده بوده‌اند معرفی خواهد شد.

۱) **روش جنگل تصادفی بقاء:** روش جنگل تصادفی بقاء^۱ یکی از روش‌های توسعه داده شده بر اساس الگوریتم‌های یادگیری ماشین است که پایه و اساس آن الگوریتم درخت تصمیم‌گیری و جنگل تصادفی است. روش جنگل تصادفی جزء الگوریتم‌های یادگیری نظارت شده^۲ است که برای هر دو کاربرد طبقه‌بندی^۳ و رگرسیون مورد استفاده قرار می‌گیرد. جنگل تصادفی در واقع الگوریتم توسعه داده درخت تصمیم‌گیری است که بر طبق آن چندین درخت تصمیم از طریق نمونه‌گیری تصادفی ساخته می‌شود و مدل نهایی از طریق درخت‌های تصمیم ایجاد شده ارزیابی می‌شود (Geron, 2019). یکی دیگر از برتری‌های جنگل تصادفی نسبت به درخت تصمیم‌گیری این است که با ایجاد هزاران درخت تصمیم می‌توان تا حد زیادی از بیش‌برازشی^۴ مدل که ضعف اصلی الگوریتم درخت تصمیم است، جلوگیری کرد. به طور کلی تمامی الگوریتم‌های یادگیری ماشین در معرض دو خطر جدی قرار دارند: بیش‌برازشی و کم‌برازشی^۵. بیش‌برازشی زمانی رخ می‌دهد که مدل بیش از حد پیچیده و اصطلاحاً داده‌های آموزش^۶ را حفظ کرده است و قابل تعمیم به داده‌های تست نیست (در یادگیری ماشین، منظور از داده‌های آموزش داده‌هایی است که برای به‌کارگیری الگوریتم، آموزش داده می‌شود و داده‌های تست نیز داده‌های است برای سنجش دقت الگوریتم مدل بر این داده‌ها اعمال می‌شود). در این شرایط

1 Random survival forest

2 Supervised learning

3 Classification

۴ برای اطلاعات بیشتر در خصوص الگوریتم‌های یادگیری نظارت شده رجوع کنید به واردهان (۲۰۲۰).

5 Overfitting

6 Underfitting

7 Training data

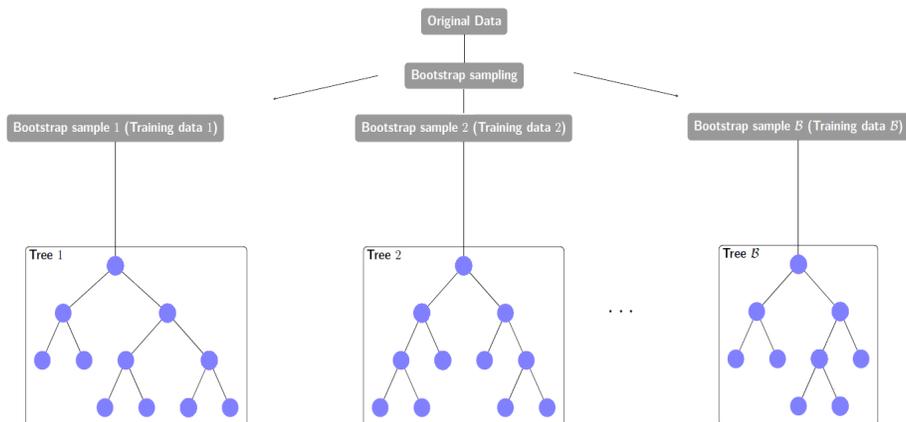
مدل قابل اعتبار نیست چرا که ممکن است تمام داده‌های پرت را نیز مورد بررسی قرار دهد. کم برازشی زمانی رخ می‌دهد که الگوریتم نتوانسته است به اندازه کافی داده‌های آموزش را مورد بررسی قرار دهد. راه‌های مختلفی برای حل دو مشکل بیش برازشی و کم برازشی پیشنهاد شده است. برای مثال می‌توان به روش اعتبار سنجی متقابل^۱ و یا تغییر ابر پارامترها^۲ اشاره کرد (Muller and Guido, 2016).

در الگوریتم جنگل تصادفی با ایجاد درخت‌های تصمیم متعدد از طریق نمونه‌گیری تصادفی تا حد زیادی می‌توان از بیش‌برازشی و کم‌برازشی مدل جلوگیری کرد. در این روش، همچون تمام الگوریتم‌های یادگیری ماشین ابتدا داده‌ها به دو مجموعه داده‌های آموزش و داده‌های تست تقسیم می‌شوند. ابتدا مدل از طریق داده‌های آموزش، آموزش داده می‌شود و سپس برای ارزیابی این یادگیری، مدل بر مجموعه داده‌های تست اعمال می‌شود. ارزیابی اولیه و کلی مدل از طریق اعمال آن بر مجموعه داده‌های تست ایجاد می‌شود (Jo, 2021). به طور کلی، اگر مدل توانسته باشد به خوبی آموزش دیده باشد باید بر روی داده‌های تست هم عملکرد خوبی داشته باشد در غیر این صورت مدل مناسب داده‌های مورد استفاده نیست. در صورت آموزش مناسب مدل بر داده‌های آموزش، بعد از این مرحله برای ساختن درخت‌های تصمیم متعدد از نمونه بوت‌استرپ^۳ استفاده می‌شود. این نمونه به معنای نمونه‌گیری مکرر از یک مجموعه داده به همراه جانشینی است که خروجی آن یک مجموعه داده است که حجم آن به اندازه داده اولیه است اما امکان دارد برخی از داده‌ها حذف و یا برخی تکرار شوند. شکل ۱ نحوه ایجاد نمونه بوت‌استرپ از داده اصلی برای هر مقدار درخت تصمیم را نشان می‌دهد.

1 Cross-validation

2 Hyperparameters

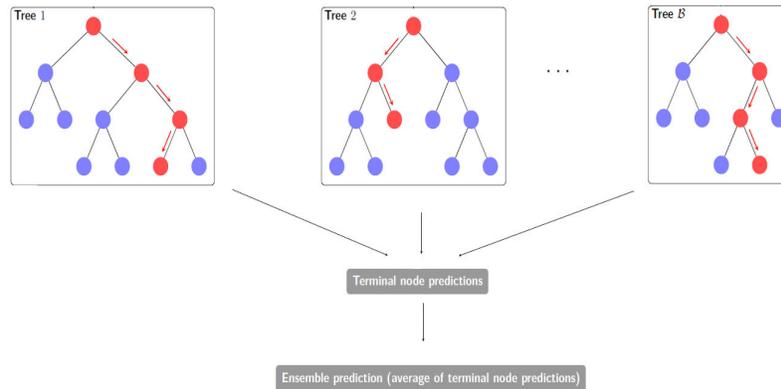
3 Bootstrap sample



شکل ۱- نمونه بوت استرپ از داده اصلی برای B مقدار درخت تصمیم

همان‌طور که در شکل ۱ مشاهده می‌شود، براساس نمونه بوت‌استرپ مقدار مشخصی از درخت تصمیم برای داده‌های تحقیق ایجاد می‌شود که مدل نهایی از طریق رأی‌گیری از این تعداد مشخص از درخت‌ها ایجاد و به کار گرفته می‌شود. بنابراین می‌توان مجموعه متناهی از درخت‌های تصمیم $\mathcal{T} = \{N_1, N_2, \dots, N_m\}$ را متصور شد که به صورت سلسله مراتبی با هم مرتبط هستند. در هر درخت تصمیم اولین گره 1 که در بالای هر درخت قرار دارد به‌عنوان گره ریشه 2 شناخته می‌شود. در ادامه در هر مرحله گره اصلی به چند زیرگروه تقسیم می‌شود که هر کدام از شاخه‌ها دربردارنده اطلاعات مربوط به داده است و در نهایت باید از بین آن‌ها یک گره نهایی انتخاب شود. شکل ۲ فرآیند رأی‌گیری در درخت تصمیم را نشان می‌دهد.

1 node
2 Root node



شکل ۲- فرآیند رأی‌گیری در بین درخت‌های تصمیم مختلف

همان‌طور که در شکل ۲ مشخص است، فرآیند رأی‌گیری در مدل جنگل تصادفی برای هر درخت تصمیم اعمال می‌شود که در نهایت تنها یک گره به‌عنوان پاسخ اصلی و گره پایانی^۱ انتخاب می‌شود. در شکل ۲ مسیر انتخاب گره پایانی در بین ۳ درخت تصمیم با رنگ قرمز مشخص شده است. مزیت اصلی الگوریتم جنگل تصادفی ایجاد درختان تصمیم مختلف است که با رأی‌گیری در هر درخت امکان بررسی تمامی حالت‌های ممکن پارامترهای مدل را به محقق خواهد داد و موجب افزایش قدرت پیش‌بینی مدل می‌شود.

۲) مدل‌سازی تابع بقا: برای مدل‌سازی تابع بقا از طریق الگوریتم جنگل تصادفی ابتدا تعداد N درخت تصمیم از طریق فرآیند بوت‌استرپینگ ایجاد می‌شود سپس برای هر درخت تصمیم تابع بقای مجزا ایجاد می‌شود و متغیرهای تأثیرگذار در مدل برای هر درخت رأی‌گیری می‌شود. در مرحله بعد براساس هر گره پایانی تابع بقای هر درخت تصمیم مشخص می‌شود. در نهایت نیز با استفاده از داده‌های OOB مقدار خطای پیش‌بینی مدل مشخص خواهد شد (Ishwaran et al., 2008). به‌طور میانگین، نمونه بوت‌استرپ حدوداً ۳۶ درصد از داده‌ها را شامل نمی‌شود و بنابراین

1. Terminal node

این داده‌ها وارد مدل نخواهند شد. به‌منظور ارزیابی مدل، از این داده‌ها برای اعتبارسنجی متقابل استفاده می‌شود که به خطای خارج از کیسه^۱ یا OOB معروف است. در نتیجه داده‌های نمونه‌گیری شده که در مدل محاسبه می‌شوند داخل کیسه^۲ نامیده می‌شوند (Fantazzini and Figini, 2009).

روش جنگل تصادفی بقا، تابع بقا، $S(t|H) = \mathbb{P}\{T \geq t|H\}$ را از طریق معادله زیر برآورد می‌کند:

$$H(t|X) = \int_{(0,t)}^x \frac{F(du|X)}{S(u|X)}, \quad F(u|X) = \mathbb{P}\{T \geq t|H\}$$

با در دست داشتن هر تعداد مشخص از درخت تصمیم تابع بقا از طریق میانگین‌گیری گره‌های پایانی بدست می‌آید.^۳ اگر $H_b(t|X)$ و $S_b(t|X)$ به ترتیب تابع تجمعی مخاطره و تابع بقا برای b مقدار درخت تصمیم باشند در نتیجه توابع برآورد شده برابر است با:

$$\bar{H}(t|X) = \frac{1}{ntree} \sum_{b=1}^{ntree} H_b(t|X), \quad \bar{S}(t|X) = \frac{1}{ntree} \sum_{b=1}^{ntree} S_b(t|X).$$

همچنین برای برآورد خطای پیش‌بینی مدل، OOB، از روش مشابهی یعنی میانگین‌گیری از گره‌های پایانی در هر درخت تصمیم استفاده می‌شود. اگر $O_i = O(X_i)$ برابر باشد با درخت‌های تصمیمی که در آن مورد i خارج از کیسه باشد بنابراین:

$$\bar{S}(t|X_i) = \frac{1}{|O_i|} \sum_{b \in O_i} S_b(t|X_i)$$

در ادامه برای بررسی اثر متغیرهای مستقل^۱ بر تابع بقا و همچنین ارزیابی اهمیت دو مورد i و j در دو متغیر مستقل X_i و X_j می‌گوییم مقدار i تاثیر کمتری دارد اگر:

1. Out of bag error
2. In-bag

۳. برای اطلاعات بیشتر در زمینه الگوریتم جنگل تصادفی رجوع کنید به لویپه (۲۰۱۴)

$$\sum_{i=1}^m \bar{H}(t_1|X_i) > \sum_{j=1}^m \bar{H}(t_1|X_j)$$

که در آن $t_1 < t_2 < \dots < t_m$ برابر است زمان وقوع رخداد در مدل. در نهایت شاخص بریر نیز برای ارزیابی مدل از طریق معادله زیر محاسبه می‌شود. اگر $\bar{S}(t|X)$ برآوردی از تابع بقا باشد مقدار شاخص بریر^۲ برای این تابع برابر است با:

$$\bar{BS}(t) = \frac{1}{n} \sum_{i=1}^n \left\{ \frac{\bar{S}2(t|X)I\{T_i \leq t\}\delta_i}{\bar{G}(T_i - |X_i)} + \frac{(1 - \bar{S}(t|X_i))^2 I\{T_i > t\}}{\bar{G}(t|X_i)} \right\}$$

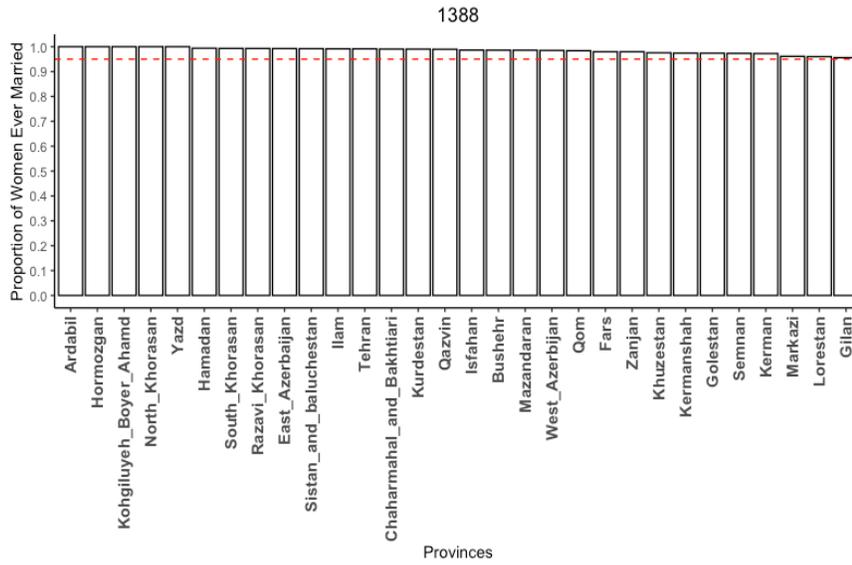
که در آن $\bar{G}(t|X_i)$ برآوردی تابع بقای سانسور شده است (Ishwaran et al. 2011).

یافته‌های تحقیق

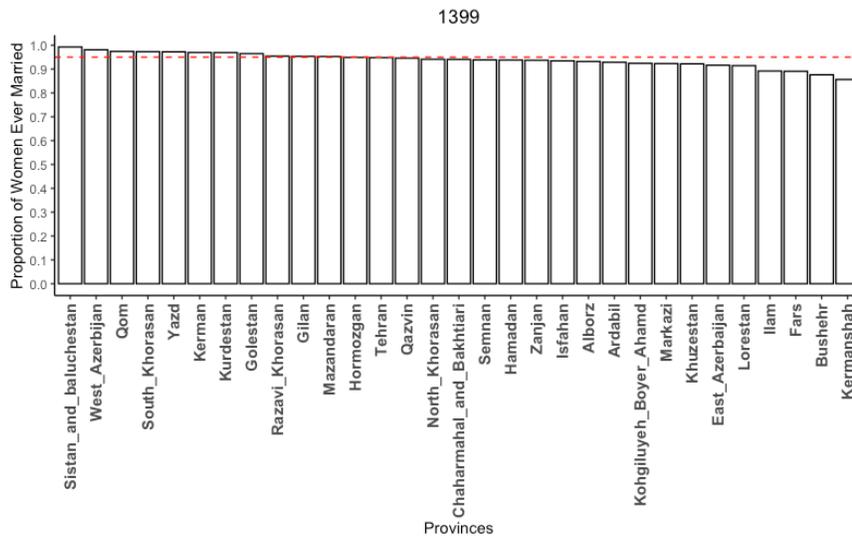
در این مطالعه، نسبت زنان ازدواج کرده در گروه سنی ۴۹-۴۵ ساله در بین استان‌های کشور وارد مدل تحقیق شده است. بررسی ابتدایی این نسبت‌ها برای استان‌های مختلف نشان‌دهنده تغییرات محسوس در نسبت زنان ازدواج کرده تقریباً در تمامی استان‌های کشور است. شکل‌های ۳ و ۴ این نسبت‌ها را در بین استان‌های مختلف برای دو سال ۱۳۸۸ و ۱۳۹۹ (زمان شروع و پایان مطالعه) نشان می‌دهد.

1 features

2 Brier index



شکل ۳- نسبت ازدواج کرده‌های زنان ۴۹-۴۵ ساله در بین استان‌های کشور. ۱۳۸۸



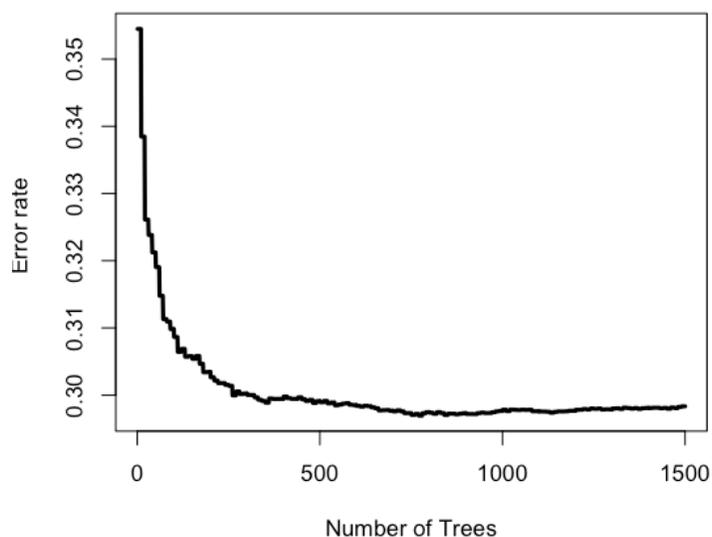
شکل ۴- نسبت زنان ازدواج کرده ۴۹-۴۵ ساله در بین استان‌های کشور. ۱۳۹۹

همان‌طور که در شکل ۳ مشخص است، نسبت زنان ازدواج کرده در گروه سنی ۴۹-۴۵ سال در تمامی استان‌های کشور در سال ۱۳۸۸ بالاتر از ۹۵٪ بوده است (خط چین قرمز نسبت ازدواج کرده‌های زنان در سطح ۹۵٪ را نشان می‌دهد) که بیان‌گر عمومیت بالای ازدواج در بین تمامی استان‌های کشور است. همچنین بیشترین نسبت زنان ازدواج کرده ۴۵-۴۹ ساله در بین استان‌های کشور مربوطه به استان‌های سیستان و بلوچستان، کهگیلویه و بویر احمد، یزد، خراسان رضوی، خراسان جنوبی و هرمزگان است. در حالی که کمترین مقدار این نسبت در استان‌های گیلان، مرکزی و لرستان مشاهده می‌شود. بنابراین، همان‌گونه که قبلاً اشاره شد، تا شروع سال ۱۳۸۸ هنوز تمامی استان‌ها از عمومیت ازدواج برخوردار بودند.

۱۲ سال بعد و در سال ۱۳۹۹ این نسبت‌ها به کلی متفاوت است. همان‌طور که در شکل ۴ مشخص است، به جز ۱۱ استان تمامی استان‌های کشور نسبت ازدواج پایین‌تر از ۹۵٪ را برای زنان ۴۹-۴۵ تجربه کرده‌اند. این استان‌ها عبارتند از استان‌های گیلان، مازندران، آذربایجان غربی، کرمان، خراسان رضوی، سیستان و بلوچستان، کردستان، یزد، قم، گلستان و خراسان جنوبی. به عبارت دیگر، این استان‌ها همچنان از عمومیت ازدواج برخوردارند. پایین‌ترین مقدار این نسبت در بین تمامی استان‌های کشور در سال ۱۳۹۹ مربوط به استان کرمانشاه با نسبت ۸۵٪ است. اگرچه در ۲۰ استان کشور نسبت ازدواج کرده‌های زنان ۴۹-۴۵ ساله پایین‌تر از عمومیت ازدواج است، اما همچنان نسبت زنان ازدواج کرده ۴۹-۴۵ ساله در اکثر استان‌های کشور در سطحی بالای ۹۰ درصد قرار دارد که نشان‌دهنده رواج بالای ازدواج در کشور است. در بخش بعد به مدل‌سازی این نسبت‌ها و در نهایت توابع بقای استانی اشاره می‌شود.

به‌کارگیری الگوریتم‌های یادگیری ماشین در مدل‌سازی تابع بقا نیازمند ملاحظات گوناگونی است که مهم‌ترین آن بررسی صحت و دقت مدل در فرآیند تحلیل داده است. شاخص‌های مختلفی برای ارزیابی این الگوریتم‌ها وجود دارد که بسته به نوع الگوریتم مورد استفاده می‌تواند متفاوت باشد. یکی از مهم‌ترین شاخص‌های ارزیابی جنگل تصادفی در ابتدا تعداد درخت

تصمیم‌گیری بهینه برای مدل است. شکل ۵ نمودار تعداد درختان تصمیم‌گیری مورد استفاده برای این تحقیق و میزان خطای متناظر با آن را براساس هر تعداد درخت تصمیم نشان می‌دهد.

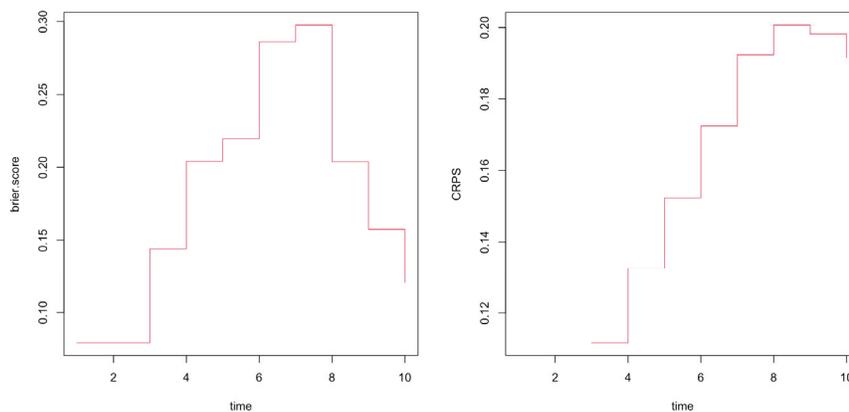


شکل ۵- تعداد درخت تصمیم و میزان خطای متناظر با آن

انتظار می‌رود با افزایش تعداد درختان تصمیم مقدار خطای مدل کاهش یابد. بنابر شکل ۵، تعداد درخت تصمیم در مقدار کمتر از ۵۰۰ درخت مقدار خطای نسبتاً بالایی را نشان می‌دهد اما با افزایش درختان مقدار خطا کاهش می‌یابد و رفته‌رفته به یک سطح پایدار می‌رسد. اینکه با چه تعداد درخت تصمیم می‌توان بهترین نتیجه را از مدل گرفت دقیقاً تابع شکل فوق و مقدار خطا در هر مقدار درخت تصمیم است. با این حال، افزایش تعداد آن لزوماً به معنای دقت بیشتر مدل نیست. تعداد بهینه درخت تصمیم همبستگی بالایی با تعداد مشاهدات در داده‌های مورد استفاده و متغیرهای تحقیق دارد. در داده‌های بسیار بزرگ انتظار می‌رود با تعداد درخت تصمیم کمتری درصد خطای مدل کاهش یابد. در این تحقیق همان‌طور که شکل ۵ نیز نشان می‌دهد، تعداد

۱۵۰۰ درخت تصمیم برای به‌کارگیری مدل استفاده شده است که درصد خطای آن به زیر ۰/۳۰ رسیده است و تعداد بهینه درختان تصمیم را برای این مدل نشان می‌دهد.

از جمله شاخص‌های مهم دیگر در ارزیابی الگوریتم جنگل تصادفی بقا شاخص نمره بریر^۱ و شاخص CPRS است. این شاخص نشان‌دهنده دقت مدل در تعداد واحد زمانی است که واقعه ممکن است در آن رخ بدهد. این شاخص بین صفر و یک متغیر است و هر چقدر عدد آن به صفر نزدیک‌تر باشد بیانگر دقت بیشتر مدل است. شکل ۶ هر دو شاخص CPRS و نمره بریر را نشان می‌دهد.



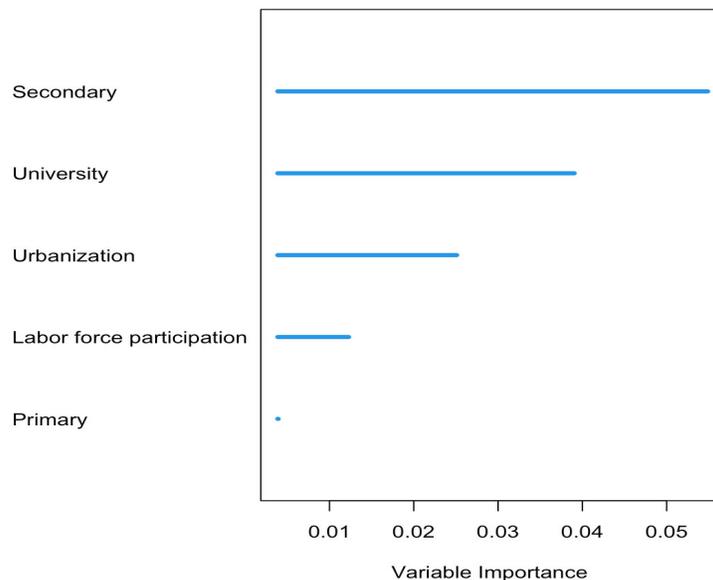
شکل ۶- شاخص بریر و CPRS به تفکیک زمان مورد مطالعه

براساس شکل ۶، شاخص بریر و CPRS هرگز به بالاتر ۰/۳۰ درصد نزدیک نمی‌شوند که بیانگر دقت قابل قبول مدل است. در شاخص بریر افزایشی و سپس کاهش‌ی بودن مدل در سال‌های وقوع رویداد نشان‌دهنده دقت مدل در سال‌های ابتدایی و پایانی داده مورد مطالعه است. به بیان دیگر، نوع پراکندگی و همبستگی متغیرهای تحقیق با زمان رخداد واقعه در

¹ Brier index

سال‌های پایانی و ابتدایی تحقیق نسبت به سال‌های میانی به نحو دقیق‌تری در مدل منعکس شده است. در شاخص CPRS نیز روند افزایشی نمودار نشان‌دهنده دقت بالای مدل در سال‌های ابتدایی در بین مشاهدات و کاهش آن در سال‌های پایانی است با این حال مقدار کمتر از ۰/۵۰ نشان‌دهنده دقت قابل قبول مدل در هر دو شاخص است.

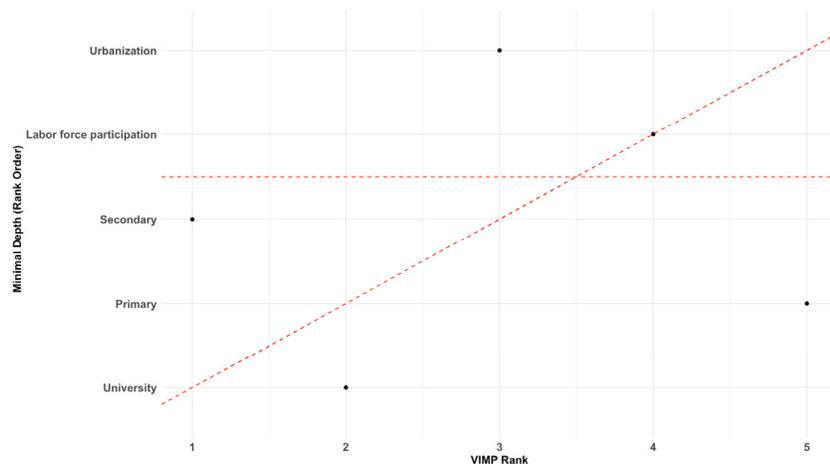
شاخص بربر شاخص بسیار مهمی در ارزیابی نهایی مدل و همچنین شاخص خوبی برای سنجش این موضوع است که متغیرهای مستقل مدل آیا توانسته‌اند تبیین‌کننده متغیر وابسته باشد یا خیر. شکل ۷ تأثیر متغیرهای مستقل مدل برای ۱۵۰۰ درخت تصمیم بر تابع بقا عمومیت ازدواج در بین استان‌های کشور را نشان می‌دهد. همان‌طور که پیش‌تر نیز گفته شد، تأثیر نهایی هر یک از متغیرهای تحقیق بر عمومیت ازدواج زنان از طریق رأی‌گیری در هر درخت تصمیم و در نهایت میانگین‌گیری از تمام گره‌های پایانی انجام می‌شود.



شکل ۷- متغیرهای تأثیرگذار بر تابع بقای عمومیت ازدواج

بنابر شکل ۷، بیشترین متغیری که توانسته است بر کاهش عمومیت ازدواج زنان و وقوع رخداد در استان‌های کشور تأثیرگذار باشد تحصیلات دبیرستانی و بعد از آن تحصیلات دانشگاهی است. شهرنشینی و میزان مشارکت اقتصادی کل نیز در رتبه‌های بعدی تأثیرگذاری بر متغیر وابسته و وقوع رویداد قرار دارند. میزان شهرنشینی در هر استان علاوه بر اثر مستقیمی که می‌تواند داشته باشد، از طریق افزایش دستیابی به فرصت‌های تحصیلی می‌تواند بر کاهش عمومیت ازدواج زنان تأثیرگذار باشد. تحصیلات ابتدایی به همراه میزان مشارکت اقتصادی کل کمترین تأثیر را بر وقوع رویداد و کاهش عمومیت ازدواج زنان داشته است با این حال تأثیر این دو متغیر صفر نیست. اهمیت نسبی هر متغیر نسبت به سایر متغیرها از طریق مقایسه بین‌استانی در مدل صورت می‌گیرد. بر این اساس، تغییراتی که استان‌های مختلف به لحاظ نسبت‌های تحصیلی و میزان استغال و شهرنشینی تجربه می‌کنند در تصمیم‌گیری نهایی مدل موثر است و خروجی نهایی که نقش متغیرهای مهم را به ترتیب نشان می‌دهد براساس تفاوت‌های استانی اعمال می‌شود.

همان‌طور که در شکل ۷ مشخص است، مقیاس عددی برای سنجش تأثیرگذاری متغیرها به صورت نسبت آن با سایر متغیرها و نه به صورت یک عدد مطلق و استاندارد شده بیان شده است و در نتیجه تنها می‌توان آن را نسبت به سایر مقادیر تفسیر کرد. در نتیجه، متغیر تحصیلات دبیرستانی بیشتر از سایر متغیرها تغییرات متغیر وابسته را توضیح می‌دهد که این متغیر به اندازه ۰/۰۶ و تحصیلات دانشگاهی به اندازه ۰/۰۴ از تغییرات متغیر وابسته را تبیین می‌کنند. همچنین، با حذف متغیر تحصیلات دبیرستانی انتظار می‌رود به مقدار ۰/۰۶ مدل دچار کم‌برازشی شود. وجه دیگر اهمیت متغیرهای تحقیق در شکل ۸ نشان داده شده است.



شکل ۸- ترتیب اهمیت متغیرهای مستقل براساس شاخص عمق حداقلی در برآورد تابع بقا

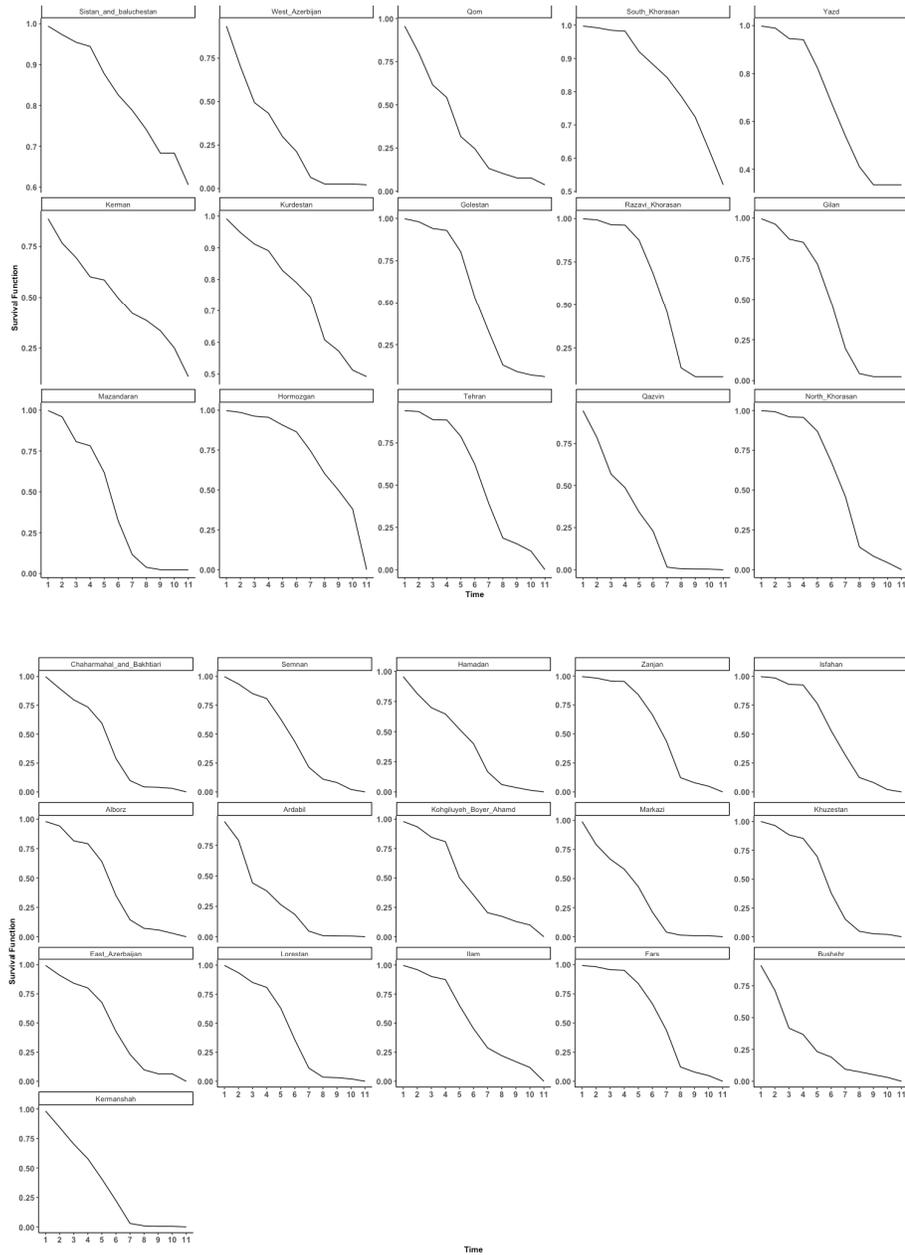
شکل ۸ ترتیب اهمیت متغیرهای تحقیق در برآورد تابع بقا در استان‌های مختلف را نشان می‌دهد. همان‌طور که انتظار می‌رود، ترتیب اهمیت متغیرها برابر است با تحصیلات دبیرستان، تحصیلات دانشگاهی، شهرنشینی، میزان مشارکت اقتصادی کل و تحصیلات ابتدایی. نکته مهم در شکل ۸ ارزیابی ترتیب اهمیت متغیرهای مستقل تحقیق براساس شاخص عمق حداقلی^۱ است. براساس این شاخص متغیرهای تحقیق براساس تعداد دفعات جدا شدن آن‌ها در فرآیند رأی‌گیری در درخت‌های تصمیم‌گیری طبقه‌بندی می‌شوند. به بیان دیگر، شاخص عمق حداقلی بیان می‌کند که در فرآیند تصمیم‌گیری در بین درخت‌های تصمیم چند مرتبه از یک متغیر مستقل برای رأی‌گیری نهایی استفاده شده است و چه تعداد از رأی‌گیری‌ها براساس آن متغیر بوده است. در شکل ۸ متغیرهای سطح تحصیلات دبیرستان، شهرنشینی و میزان مشارکت اقتصادی کل به‌عنوان بیشترین معیار رأی‌گیری در بین درخت‌های تصمیم انتخاب شده‌اند.

1 - Minimal Depth

با ارزیابی کلی مدل و بررسی اثرات متغیرهای تحقیق بر تابع بقای مدل، می‌توان تابع بقای کلی را در بین استان‌های مختلف براساس هر تعداد درخت تصمیم مختلف برآورد کرد. بر این اساس، با وارد کردن تأثیر هر یک از متغیرهای مدل بر استان‌های مختلف، تابع بقا با تأثیر تمامی متغیرها برای هر استان بدست می‌آید. شکل ۹ تابع بقای هر استان را در طول دوره مطالعه نشان می‌دهد. ترتیب استان‌ها براساس شکل ۴ است، استان‌هایی که در سال ۱۳۹۹ به ترتیب از نسبت ازدواج کرده‌های بالاتری برای زنان ۴۹-۵۰ ساله برخوردار بوده‌اند.

بنابر شکل ۹، عمده استان‌های کشور در طول دوره مطالعه به سطح آستانه‌ای کاهش نسبت زنان ازدواج کرده ۴۵-۴۹ ساله به پایین‌تر از ۰/۹۵ نزدیک شده‌اند. همان‌طور که پیش‌تر نیز گفته شد، تنها ۱۱ استان واقعه مورد نظر را تا پایان سال مورد بررسی، ۱۳۹۹، تجربه نکرده‌اند. در حالی که نسبت زنان ازدواج کرده ۴۵-۴۹ ساله سایر ۲۰ استان با فاصله‌های زمانی مختلف تا پایان سال ۱۳۹۹ به مقداری پایین‌تر از ۰/۹۵ رسیده است. الگوی تابع بقا در ۳ استان سیستان و بلوچستان، آذربایجان غربی و قم حاکی از آن است که نه تنها تا پایان سال مطالعه واقعه مورد نظر را تجربه نکرده‌اند بلکه در سطح بسیار بالایی از احتمالات تابع بقا قرار دارند. برای این سه استان احتمالات بقا تا پایان سال مطالعه به ترتیب برابر است با ۰/۹۹، ۰/۹۸ و ۰/۹۷۳ است. این ۳ استان در تحلیل اولیه نیز در شمار استان‌هایی قرار داشتند که در هر دو سال ۱۳۹۹ و ۱۳۸۸ از عمومیت ازدواج برخوردار بودند. همچنین استان‌هایی مانند بوشهر، اردبیل، قزوین، کهگیلویه و بویر احمد و چهارمحال و بختیاری اگرچه واقعه مورد نظر را تجربه کرده‌اند، با این حال، احتمال بقای بالاتری را برای واقعه مورد نظر نشان می‌دهند. برای این استان‌ها روند تابع بقا با شدت بیشتری به سمت تجربه واقعه مورد نظر پیش می‌رود. برخلاف این استان‌ها، استان‌های هرمزگان، زنجان و فارس قرار دارند که با یک شیب ملایم و کندی به سمت تجربه واقعه مورد نظر پیش می‌روند.

تحلیل استانی عمومیت ازدواج زنان در ایران: کاربرد یادگیری ماشین در مدل‌سازی تابع بقا ۱۶۷



شکل ۹- تابع بقا براساس مدل جنگل تصادفی بقا برای استان‌های کشور

الگوی متفاوت استانی در تابع بقا، نشان‌دهنده سطح متفاوت از تأثیرگذاری متغیرهای مستقل بر متغیر مورد مطالعه و وقوع رخداد است. دستیابی متفاوت به سطوح تحصیلی و نسبت‌های مختلفی از شهرنشینی و میزان مشارکت اقتصادی اینگونه توانسته است بر احتمال بقا و تجربه واقعه مورد نظر تأثیرگذار باشد. از سوی دیگر الگوی متأخر ازدواج در بین زنان می‌تواند عامل دیگری برای تفاوت‌ها در احتمالات بقای استانی باشد. در برخی از استان‌ها مانند هرمزگان الگوی احتمال بقا نشان‌دهنده سرعت بیشتر تجربه واقعه در این استان است در حالی که در برخی دیگر از استان‌ها مانند البرز این روند آرام و بطئی است. زمان‌بندی متفاوت استان‌های مختلف در رسیدن به موقعیت عدم عمومیت ازدواج و مقادیر متفاوت احتمال بقا در آن‌ها به خوبی نشان‌دهنده تأثیر متغیرهای مورد مطالعه بر تابع بقا است. از طرف دیگر، با توجه به محدودیت داده‌های مورد مطالعه در این تحقیق، ممکن است چنین تفاوت‌های استانی مربوط به سایر متغیرهای اجتماعی و اقتصادی و همچنین متغیرهای فرهنگی، قومی و مذهبی باشد که در این مطالعه سنجیده نشده است. این موضوع به خصوص برای استان‌هایی بیشتر صادق است که در طول دوره مطالعه واقعه مورد نظر را تجربه نکرده باشند. در این استان‌ها این امکان وجود دارد که متغیرهای دیگری بر عمومیت ازدواج تأثیرگذار باشد.

نتیجه‌گیری

برآوردهایی که از تغییرات ازدواج انجام می‌شود عموماً طیف وسیعی از متغیرهای تأثیرگذار اجتماعی-اقتصادی و جمعیتی را شامل می‌شود. این موضوع تا حد زیادی به ماهیت بین رشته‌ای مطالعات ازدواج و خانواده و همچنین تأثیرپذیری آن از عوامل مختلف در یک جامعه برمی‌گردد. با این حال، مطالعات متعددی که در این زمینه انجام شده نقش متغیرهای مهمی را آشکار کرده است که می‌تواند تا حد زیادی تبیین‌کننده بخش عمده تغییرات ازدواج زنان باشد. تحصیلات به‌عنوان یکی از مهم‌ترین عواملی است که در بسیاری از تحقیقات خارجی و داخلی در تأثیرگذاری تأخیر در ازدواج و کاهش عمومیت ازدواج زنان بدان اشاره شده است. افزایش سال‌های حضور در دوره‌های تحصیلی و داشتن چشم‌انداز شغلی بعد از تحصیلات دلیل عمده

تأخیر در ازدواج است که فراوان بدان پرداخته شده است. سطوح مختلف تحصیلی زنان می‌تواند تأثیرات متفاوتی بر الگوی سنی ازدواج آنان داشته باشد. علاوه بر این، سطح متفاوت دستیابی زنان به تحصیلات در نقاط شهری و روستایی باعث شده است تا شهرنشینی فارغ از اثر خاص خود، از طریق ایجاد فرصت‌های متفاوت زنان به تحصیلات در تأخیر در ازدواج زنان تأثیرگذار باشد.

در این مقاله، با تکیه بر ادبیات پیشین اثر برخی از متغیرهای مهم همچون تحصیلات، شهرنشینی و میزان مشارکت اقتصادی بر روند عمومیت ازدواج زنان در استان‌های مختلف مورد بررسی قرار گرفت. با سطح متفاوتی از شهرنشینی و دستیابی به فرصت‌های آموزشی زنان، انتظار می‌رود روند تأخیر در ازدواج آنان نیز به تفکیک استان‌ها مختلف متفاوت باشد. از طرفی دیگر، براساس مطالعه عباسی شوازی (۱۳۸۰) همگرایی‌ای که در کاهش باروری در استان‌های مختلف کشور اتفاق افتاده است، فرض این همگرایی برای کاهش عمومیت ازدواج در استان‌های مختلف را قوی‌تر می‌کند. اگرچه ازدواج در بافت سنتی-دینی ایران همچنان از ارزش بالایی برخوردار است، اما تأثیرگذاری برخی از عوامل و شرایط متفاوت استان‌های مختلف را در این زمینه نباید نادیده گرفت.

برای بررسی چنین تفاوت‌هایی در سطح استان‌های کشور، در این مقاله از یک روش جدید در مدل‌سازی توابع بقا استفاده شده است. الگوریتم‌های یادگیری ماشین به‌واسطه دقت بالا و قابلیت مدل‌سازی با داده‌ها و متغیرهای فراوان به سرعت در حال گسترش هستند. این روش‌ها قابلیت‌های مهمی از جمله آموزش داده‌ها را برای محقق فراهم می‌کنند که از طریق آن به ماشین این اجازه را می‌دهد که با کمترین دخالت محقق مدلی که در ذیل داده پنهان شده است را کشف و به کار بگیرد. با این حال استفاده از این الگوریتم‌ها نیازمند ملاحظات بسیاری است که باید از سوی محقق به دقت بررسی شود. ماهیت احتمالی این روش‌ها باعث می‌شود خطای حاصل از مدل‌سازی به دقت قابل بررسی و نمایش باشد که این امکان را فراهم می‌کند تا محقق بتواند

مدل متناسب با داده خود را به نحو دقیق‌تری انتخاب کند. ضعف عمده چنین الگوریتم‌هایی پیچیدگی آن است که می‌تواند به کارگیری این مدل‌ها را با مشکل مواجه کند. نکته مهمی که در استفاده از این الگوریتم‌ها باید به یاد داشت آن است که همیشه در معرض بیش‌برازشی و یا کم‌برازشی قرار دارند. در این مقاله، تحلیل استانی تابع بقا در استان‌های مختلف کشور از طریق روش جنگل تصادفی بقا و تعریف یک سطح آستانه‌ای، نسبت ازدواج کمتر از ۹۵ درصد برای زنان ۴۹-۵۰ سال، انجام شده است.

تحلیل استانی تغییرات عمومیت ازدواج زنان در ایران نشان می‌دهد که روند کاهشی آن از سال ۱۳۸۸ آغاز شده است. تا قبل از این سال، تغییر محسوسی در عمومیت ازدواج زنان به خصوص در استان‌های کمتر توسعه‌یافته مشاهده نشده است. دستیابی متفاوت زنان به تحصیلات و همچنین شهرنشینی و میزان مشارکت اقتصادی متفاوت توانسته است منجر به تفاوت قابل توجه در روند کاهشی عمومیت ازدواج زنان شود. تابع بقای محاسبه شده در این مطالعه نشان می‌دهد که از بین کل استان‌های کشور ۱۱ استان گیلان، مازندران، آذربایجان غربی، کرمان، خراسان رضوی، سیستان و بلوچستان، کردستان، یزد، قم، گلستان و خراسان جنوبی تا پایان سال ۱۳۹۹ همچنان از عمومیت ازدواج برخوردار بوده‌اند. این در حالی است که بقیه استان‌ها در سال‌های مختلف دوره مورد مطالعه واقعه را تجربه کرده‌اند و نسبت ازدواج کرده‌های زنان ۴۹-۵۰ ساله به زیر ۰/۹۵ رسیده است. بررسی متغیرهای تأثیرگذار در تابع بقای استان‌ها نیز نشان داد که سطح تحصیلات دبیرستان و دانشگاه بیشترین تأثیر را در کاهش عمومیت ازدواج زنان در استان‌های کشور دارد. بعد از تحصیلات، شهرنشینی به‌عنوان مهم‌ترین متغیر تبیین‌کننده کاهش عمومیت ازدواج به شمار می‌رود.

علیرغم کاهش عمومیت ازدواج و تغییرات قابل توجه در تاخیر ازدواج زنان در اکثر استان‌های کشور، ۲۲ استان، بررسی‌های استانی حاکی از آن است که نسبت ازدواج کرده‌های زنان ۴۹-۵۰ ساله همچنان در سطح بالایی قرار دارد. با وجود کاهش این نسبت، در بسیاری از

استان‌های کشور نسبت ازدواج کرده‌های زنان در سطحی بالاتر از ۰/۹۰ قرار دارد. تحلیل توابع بقای استانی نیز نشان می‌دهد، به جز ۱۱ استان مذکور بسیاری از استان‌ها تا پایان سال مطالعه واقعه مورد نظر را تجربه نکرده‌اند. در نتیجه، اگرچه عمومیت ازدواج در ایران در سال ۱۳۹۹ در بسیاری از استان‌های کشور نسبت به سال ۱۳۸۸ از بین رفته است، نسبت زنان ازدواج کرده ۴۵-۴۹ ساله به ندرت به سطحی پایین‌تر از ۰/۸۷ رسیده است که نشان‌دهنده رواج بالای ازدواج اما با یک الگوی متأخر در کشور است.

تفاوت‌های فرهنگی و اجتماعی در هر یک استان‌های کشور می‌تواند تأثیر گوناگونی بر روند تغییرات ازدواج داشته باشد که فراتر از بحث این مقاله است. در شرایط مساوی و در صورت وجود سطوح یکسان متغیرهای مستقل در استان‌های کشور، انتظار می‌رود شرایط فرهنگی خاص هر استان تبیین‌کننده تغییرات عمومیت ازدواج زنان باشد. عواملی همچون فردگرایی، خودتحقق‌بخشی، سرمایه فرهنگی و ارزش‌های دینی می‌تواند در کنار سایر عوامل اجتماعی و اقتصادی توضیح‌دهنده تغییرات ازدواج در ایران باشد (عسکری‌ندوشن و همکاران، ۱۳۹۵). اهمیت این موضوع با توجه به سطح توسعه‌یافتگی مختلف استان‌ها و همچنین تفاوت‌های قومی و فرهنگی ملازم با آن بیشتر نیز می‌شود. برای مثال، عمومیت ازدواج در دو استان قم و خراسان رضوی حتی تا پایان سال ۱۳۹۹ در سطح بالایی قرار دارد که می‌تواند تا حدودی منعکس‌کننده نقش فضای مذهبی در این دو استان باشد.

در این تحقیق تحصیلات به‌عنوان تأثیرگذارترین متغیر بر کاهش عمومیت ازدواج زنان همچون سایر مطالعات، نشان‌دهنده اهمیت بالای این متغیر در تغییرات ازدواج است. در برخی از مطالعات به الگوی U شکل تأثیرگذاری تحصیلات بر سن ازدواج زنان اشاره شده است (خلیلی، ۱۴۰۰؛ عسکری‌ندوشن و همکاران، ۱۳۹۷). به‌نحوی که، زنان با تحصیلات کم و زنان با تحصیلات بالا بیشترین تأخیر در ازدواج را تجربه می‌کنند و از شانس ازدواج کمتری برخوردار هستند. با این حال، بررسی‌های مقاله حاضر نشان می‌دهد که الگوی تأثیرگذاری تحصیلات بر

ازدواج زنان وقتی در سطح استان‌های کشور نگاه شود، می‌تواند متفاوت باشد و لزوماً رابطه U شکل بین تحصیلات و تاخیر در ازدواج زنان را تایید نکند. از طرفی دیگر، این امکان نیز وجود دارد که با افزایش سهم زنان تحصیل کرده نسبت به زنان بی‌سواد در کل جمعیت، الگوی ازدواج این زنان در درازمدت غالب شود و حتی با افزایش سهم این زنان ممکن است شانس ازدواج آنان نیز نسبت به زنان بی‌سواد افزایش یابد. آیسن و استوینسون در تایید این نکته (۲۰۱۰) معتقدند، هرچه از نسل‌های گذشته به نسل‌های امروزی می‌رسیم، زنان تحصیل کرده نسبت به گذشته از شانس ازدواج بیشتری برخوردار هستند که دلیل آن افزایش سهم این زنان نسبت به کل جمعیت زنان است.

یکی از محدودیت‌های اصلی تحلیل بقا یافتن متغیرهای مستقلی است که مقدار آن‌ها در طول زمان تغییر کند. در این مقاله، با توجه به محدودیت در دسترسی به داده‌های سری زمانی در بازه ۱۲ ساله تحقیق، سعی شد تا متغیرهایی برای بررسی تغییرات ازدواج انتخاب شوند که اطلاعات آن‌ها از سال ۱۳۸۸ تا ۱۳۹۹ موجود باشد. با این حال، می‌توان متغیرهای دیگری را نیز وارد مدل کرد و اثر آن را بر روند تغییرات عمومیت ازدواج زنان مورد مطالعه قرار داد.

منابع

- امامی، عادل؛ عسکری ندوشن، عباس؛ ترکاشوند مرادآبادی، محمد؛ و ترابی، فاطمه. (۱۴۰۰). تأخیر ازدواج در ایران: تحلیل تفاوت‌های شهرستانی در الگوهای مجرد مردان و زنان ایرانی. مطالعات جمعیتی، (۱)۷، ۳۵-۶۵.
https://jips.nipr.ac.ir/article_144524.html
- بگی، میلاد. (۱۴۰۱). دلایل و تعیین‌کننده‌های تاخیر در ازدواج در ایران. جامعه‌شناسی کاربردی، (۱)۳۴، ۳۱-۵۰.
<https://doi.org/10.22108/jas.2022.134561.2312>
- ترابی، فاطمه؛ و شمس قهفرخی، زهرا. (۱۴۰۰). عوامل موثر بر تاخیر در ازدواج در ایران و جهان. نامه انجمن جمعیت‌شناسی ایران، (۳۲)۱۶، ۱۴۱-۱۷۵.
<https://doi.org/10.22034/jpai.2022.547779.1219>
- حسینی، حاتم؛ صادقی، رسول؛ و رضایی، اردشیر. (۱۳۹۸). عوامل موثر بر تاخیر ازدواج دختران در شهرستان‌های ایران. توسعه محلی (روستایی-شهری)، (۱)۱۱، ۱۴۷-۱۷۲.
https://jrd.ut.ac.ir/article_74459.html

تحلیل استانی عمومیت ازدواج زنان در ایران: کاربرد یادگیری ماشین در مدل‌سازی تابع بقا ۱۷۳

خلیلی، مهدی. (۱۴۰۰). تحلیل نسلی-مقطعی ازدواج زنان ایران با استفاده از جداول عمر زناشویی. نامه انجمن جمعیت‌شناسی ایران، ۱۶(۳۱)، ۳۱-۷. <https://doi.org/10.22034/jpai.2022.542847.1205>

رازقی نصرآباد، حجیه بی‌بی؛ و رحیمی، علی. (۱۳۹۳). بررسی تحولات امید زندگی مجردی در ایران با استفاده از جداول خالص زناشویی: دوره ۱۳۸۵ تا ۱۳۹۰. نامه انجمن جمعیت‌شناسی ایران، ۹(۱۸)، ۹۵-۱۱۸. https://www.jpaiassoc.ir/article_23422.html

رحیمی، علی؛ کاظمی‌پور، شهلا؛ و رازقی نصرآباد، حجیه بی‌بی. (۱۳۹۴). تحلیل گذار مجرد زنان و مردان در ایران طی سال‌های ۱۳۴۵ تا ۱۳۹۰. مطالعات اجتماعی ایران، ۹(۳)، ۸۰-۱۰۹. http://www.jss-isa.ir/article_26754.html

رضادوست، کریم؛ و ممینی، ایمان. (۱۳۹۰). بررسی رابطه بین تأخیر سن ازدواج و متغیرهای میزان درآمد، میزان تحصیلات و تعدادی دیگر از متغیرها در زنان شاغل. مشاوره کاربردی، ۱۱(۱)، ۱۰۳-۱۲۰. https://jac.scu.ac.ir/article_10266.html

عباسی شوازی، محمدجلال. (۱۳۸۰). همگرایی رفتارهای باروری در ایران میزان روند و الگوی سنی باروری در استان‌های کشور در سال‌های ۱۳۵۱ و ۱۳۷۵. نامه علوم اجتماعی، ۱۸(۱۸)، ۲۰۱-۲۳۱. https://jnoe.ut.ac.ir/article_15210.html

عباسی شوازی، محمدجلال؛ و صادقی، رسول. (۱۳۸۴). قومیت و الگوهای ازدواج در ایران. پژوهش زنان، ۳(۱)، ۲۵-۴۷. https://jwdp.ut.ac.ir/article_13238.html

عسکری ندوشن، عباس؛ فتوحی تفتی، فاطمه؛ و علی مندگاری، ملیحه. (۱۳۹۷). تفاوت‌های اقتصادی-اجتماعی مرتبط با سن ازدواج زنان در ایران: مطالعه تطبیقی دختران جوان در آستانه‌ی ازدواج و زنان همسر دار ۱۵ تا ۴۹ ساله. مطالعات راهبردی زنان، ۲۰(۷۹)، ۲۱۰-۱۷۱. <https://doi.org/10.22095/JWSS.2018.81515>

عسکری ندوشن، عباس؛ عباسی شوازی، محمدجلال؛ و پیری محمدی، مریم. (۱۳۹۵). سن ایده‌آل ازدواج و عوامل تعیین‌کننده آن در شهر یزد. مطالعات راهبردی زنان، ۱۹(۷۳)، ۳۵-۶۳. <https://doi.org/10.22095/jwss.2016.44383>

فتحی، الهام؛ جاوید، نورمحمد؛ سرخیل، بهناز؛ و زبرجد، سودابه. (۱۳۹۶). تغییرات میانگین سن ازدواج در ایران. آمار، ۲۴، ۱۲-۸. <https://amar.srtc.ac.ir/article-1-281-fa.html>

کاظمی‌پور، شهلا. (۱۳۸۳). تحول سن ازدواج زنان در ایران و عوامل جمعیتی موثر بر آن. پژوهش زنان، ۲(۳)، ۱۰۳-۱۲۴. https://jwdp.ut.ac.ir/article_13256.html

کوششی، مجید؛ و خلیلی، مهدی. (۱۳۹۹). پیش‌بینی عمومیت ازدواج زنان ایران (بر پایه تحلیل الگوهای همسرگزینی سنی). بررسی مسائل اجتماعی ایران، ۱۱(۲)، ۴۰-۷. <https://doi.org/10.22059/ijsp.2021.82670>

محمودیان، حسین. (۱۳۸۳). سن ازدواج در حال افزایش: بررسی عوامل پشتیبان. نامه علوم اجتماعی / مطالعات جامعه‌شناختی، ۲۴، ۲۷-۵۳. https://jnoe.ut.ac.ir/article_10552.html

Abbasi-Shavazi, M. J. (2001). Convergence of fertility behaviors in Iran: Levels, trends, and age patterns of fertility in the provinces of the country during 1972 and 1996. *Journal of Social Sciences Letter / Sociological Review*, 18(18), 201-231. [In Persian]. https://jnoe.ut.ac.ir/article_15210.html

Abbasi-Shavazi, M. J., & Sadeghi, R. (2005). Ethnicity and marriage patterns in Iran. *Woman in Development & Politics / Women's Research*, 3(1), 25-47. [In Persian]. https://jwdp.ut.ac.ir/article_13238.html

Askari-Nodoushan, A., Abbasi Shavazi, M. J., & Piri Mohammadi, M. (2016). Ideal age at marriage and its correlates in Yazd City. *Women's Strategic Studies*, 19(73), 35-63. [In Persian]. <https://doi.org/10.22095/jwss.2016.44383>

Askari-Nodoushan, A., Fotouhi-Tafti, F., & Alimondegari, M. (2018). Socioeconomic differential of female age at marriage in Iran: Comparative study of female youth at marriageable age and female married 15-49 years old. *Women's Strategic Studies*, 20(79), 171-210. [In Persian]. <https://doi.org/10.22095/JWSS.2018.81515>

Bagi, M. (2023). Reasons for delayed marriage and its determinants in Iran. *Journal of Applied Sociology*, 34(1), 31-50. [In Persian]. <https://doi.org/10.22108/jas.2022.134561.2312>

Bhrolcháin, M. N. (2001). Flexibility in the marriage market. *Population: An English Selection*, 13(2), 9-48. <https://doi.org/10.3917/popu.p2001.13n2.0047>

Bloome, D., & Ang, S. (2020). Marriage and union formation in the United States: Recent trends across racial groups and economic backgrounds. *Demography*, 57(5), 1753-1786. <https://doi.org/10.1007/s13524-020-00910-7>

- Emami, A., Askari-Nodoushan, A., Torkashvand Moradabadi, M., & Torabi, F. (2022). Marriage postponement in Iran: District-level analysis of youth singlehood patterns. *Iranian Population Studies*, 7(1), 35-65. [In Persian].
https://jips.nipr.ac.ir/article_144524.html
- Fantazzini, D., & Figini, S. (2009). Random survival forests models for SME credit risk measurement. *Methodology and Computing in Applied Probability*, 11(1), 29-45.
<https://doi.org/10.1007/s11009-008-9078-2>
- Fathi, E., Javid, N., Sarkheil, B., & Zebarjad, S. (2017). Changes in the average age of marriage in Iran. *Amar / Statistics*, 24, 8-12. [In Persian]. <https://amar.srtc.ac.ir/article-1-281-fa.html>
- Géron, A. (2019). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow* (2nd ed.). O'Reilly Media, Inc.
- Goldstein, J. R., & Kenney, C. T. (2001). Marriage delayed or marriage forgone? New cohort forecasts of first marriage for U.S. women. *American Sociological Review*, 66(4), 506-519.
<https://doi.org/10.1177/000312240106600402>
- Hirschman, C., & Matras, J. (1971). A new look at the marriage market and nuptiality rates, 1915-1958. *Demography*, 8(4), 549-569. <https://doi.org/10.2307/2060691>
- Hosseini, H., Sadeghi, R., & Rezaei, A. (2019). Determinants of delay marriage among girls in Iran. *The Journal of Community Development (Rural-Urban)*, 11(1), 147-172. [In Persian].
https://jrd.ut.ac.ir/article_74459.html
- Ikamari, L. (2005). The effect of education on the timing of marriage in Kenya. *Demographic Research*, 12, 1-28. <https://doi.org/10.4054/demres.2005.12.1>
- Isen, A., & Stevenson, B. (2010). Women's education and family behavior: Trends in marriage, divorce and fertility. *National Bureau of Economic Research Working Paper, No. 15725*.
<https://doi.org/10.3386/w15725>
- Ishwaran, H., Kogalur, U. B., Blackstone, E. H., & Lauer, M. S. (2008). Random survival forests. *The Annals of Applied Statistics*, 2(3), 841-860. <https://doi.org/10.1214/08-aos169>
- Ishwaran, H., Kogalur, U. B., Chen, X., & Minn, A. J. (2011). Random survival forests for high-dimensional data. *Statistical Analysis and Data Mining*, 4(1), 115-132.
<https://doi.org/10.1002/sam.10103>
- Jo, T. (2021). *Machine learning foundations: Supervised, unsupervised, and advanced learning*. Springer International Publishing.

- Jones, G. (2010). Changing marriage patterns in Asia. *SSRN Electronic Journal*.
<https://doi.org/10.2139/ssrn.1716533>
- Jones, G. W., & Gubhaju, B. (2009). Factors influencing changes in mean age at first marriage and proportions never marrying in the low-fertility countries of East and Southeast Asia. *Asian Population Studies*, 5(3), 237–265.
<https://doi.org/10.1080/17441730903351487>
- Kazemipour, Sh. (2004). The evolution of women's age at marriage in Iran and the demographic factors affecting it. *Woman in Development & Politics / Women's Research*, 2(3), 103-124. [In Persian]. https://jwdp.ut.ac.ir/article_13256.html
- Khalili, M. (2021). Cohort and period analysis of marriage change of Iranian women using nuptiality life tables. *Journal of Population Association of Iran*, 16(31), 7-31. [In Persian].
<https://doi.org/10.22034/jpai.2022.542847.1205>
- Koosheshi, M., & Khalili, M. (2021). Forecasting the universality of marriage among Iranian women (Based on analysis of nuptiality patterns and first marriage demand form men). *Journal of Social Problems of Iran*, 11(2), 7-40. [In Persian].
<https://doi.org/10.22059/ijsp.2021.82670>
- Lampard, R. (2013). Age at marriage and the risk of divorce in England and Wales. *Demographic Research*, 29, 167–202. <https://doi.org/10.4054/demres.2013.29.7>
- Loupe, G. (2014). *Understanding random forests: From theory to practice*. Ph.D. Dissertation, University of Liège, Faculty of Applied Sciences.
- Mahmoudian, H. (2004). Rising age at marriage: An examination of supporting factors. *Journal of Social Sciences Letter / Sociological Review*, 24, 27-53. [In Persian].
https://jnoe.ut.ac.ir/article_10552.html
- Malhotra, A. (1997). Gender and the timing of marriage: Rural-urban differences in Java. *Journal of Marriage and the Family*, 59(2), 434. <https://doi.org/10.2307/353481>
- Muhsam, H. V. (1974). The marriage squeeze. *Demography*, 11(2), 291–299.
<https://doi.org/10.2307/2060564>
- Müller, A. C., & Guido, S. (2016). *Introduction to machine learning with Python: A guide for data scientists*. O'Reilly Media, Inc.
- Puschmann, P., & Solli, A. (2014). Household and family during urbanization and industrialization: Efforts to shed new light on an old debate. *The History of the Family: An International Quarterly*, 19(1), 1–12. <https://doi.org/10.1080/1081602x.2013.871570>

- Rahimi, A., Kazempour, Sh., & Razaghi Nasrabad, H. B. (2015). Analysis of the transition to celibacy among women and men in Iran during the years 1966 to 2011. *Iranian Social Studies*, 9(3), 80-109. [In Persian]. http://www.jss-isa.ir/article_26754.html
- Rallu, J. L. (2006). Female deficit and the marriage market in Korea. *Demographic Research*, 15, 51-60. <https://doi.org/10.4054/demres.2006.15.3>
- Razeghi-Nasrabad, H. B., & Rahimi, A. (2015). Changes in singlehood life expectancy using the net nuptiality table in Iran, 2006-2011. *Journal of Population Association of Iran*, 9(18), 96-119. [In Persian]. https://www.jpaiassoc.ir/article_23422.html
- Rezadoust, K., & Mombeni, I. (2011). The relationship of income, graduation and other variables with delay in marriage among single employed women. *Biannual Journal of Applied Counseling*, 1(1), 103-120. [In Persian]. https://jac.scu.ac.ir/article_10266.html
- Saadati, M., & Bagheri, A. (2017). Determinants of Iranian youths' marriage age: A parametric survival analysis approach. *Social Determinants of Health*, 3(4), 177-185. <https://doi.org/10.22037/sdh.v3i4.18529>
- Sabbah-Karkaby, M., & Stier, H. (2017). Links between education and age at marriage among Palestinian women in Israel: Changes over time. *Studies in Family Planning*, 48(1), 23-38. <https://doi.org/10.1111/sifp.12015>
- Torabi, F., & Baschieri, A. (2010). Ethnic differences in transition to first marriage in Iran: The role of marriage market, women's socio-economic status, and process of development. *Demographic Research*, 22, 29-62. <https://doi.org/10.4054/demres.2010.22.2>
- Torabi, F., & Shams Ghahfarokhi, Z. (2021). Factors affecting marriage delay in Iran and around the world. *Journal of Population Association of Iran*, 16(32), 141-175. [In Persian]. <https://doi.org/10.22034/jpai.2022.547779.1219>
- Torabi, F., Baschieri, A., Clarke, L., & Abbasi-Shavazi, M. J. (2013). Marriage postponement in Iran: Accounting for socio-economic and cultural change in time and space. *Population, Space and Place*, 19(3), 258-274. <https://doi.org/10.1002/psp.1710>
- Verdhan, V. (2020). *Supervised learning with Python: Concepts and practical implementation using Python*. Apress. <https://doi.org/10.1007/978-1-4842-6156-9>
- Zhao, C., Chen, B., & Li, X. (2023). Rising housing prices and marriage delays in China: Evidence from the urban land transaction policy. *Cities*, 135(104214), 104214. <https://doi.org/10.1016/j.cities.2023.104214>