

## معرفی و کاربرد مدل کارت برای طبقه‌بندی تعداد فرزندان ایده‌آل

### زنان ۱۵-۴۹ ساله استان سمنان<sup>۱</sup>

آرزو باقری<sup>۲</sup>

مهسا سعادت<sup>۳</sup>

حجیه بی بی رازقی نصرآباد<sup>۴</sup>

هدف اصلی این مقاله معرفی و کاربرد مدل درختی کارت برای طبقه‌بندی وقایع جمعیتی است که از آن برای طبقه‌بندی تعداد فرزندان ایده‌آل طرح "بررسی رفتارهای ازدواج و باروری زنان حداقل یکبار ازدواج کرده، ۱۵-۴۹ ساله در استان سمنان-۱۳۹۱" استفاده شده است. یافته‌های مطالعه نشان داد که مدل درختی کارت با استفاده از ضریب جینی با احتمالات پیشین برآوردی از دقت کافی و مناسب برای پیش‌بینی طبقه تعداد فرزندان ایده‌آل برخوردار بود. نتایج برازش این مدل برای طبقه‌بندی تعداد فرزندان ایده‌آل ۴۰۵ زن حداقل یکبار ازدواج کرده با متغیرهای پیش‌بین سن در اولین ازدواج، نوع ازدواج، سطح تحصیلات، وضعیت شغلی، محل تولد و کوهورت موالید، نشان داد که وضعیت شغلی، نقش متفاوتی در میان کوهورت‌های موالید ایفا می‌کند. نزدیکی نظرات و نیات باروری کوهورت موالید اول و دوم به یکدیگر، از دیگر نتایج

۱. این مقاله مستخرج از طرح کاوش داده‌های جمعیتی با استفاده از درخت تصمیم (ابلاغ شماره ۲۰/۱۵۲۸۳ مورخ ۹۳/۱۱/۵) است که با حمایت مالی مؤسسه مطالعات و مدیریت جامع و تخصصی جمعیت کشور در سال ۱۳۹۳ انجام شده است. نویسندگان مقاله بر خود لازم می‌دانند تا از جناب آقای دکتر محمد جلال عباسی شوازی که با ارائه نظرات ارزشمند خود بر غنای این مقاله افزودند کمال تشکر را داشته باشند.

۲. استادیار آمار کاربردی مؤسسه مطالعات و مدیریت جامع و تخصصی جمعیت کشور، -abagheri\_000@ya-hoo.com

۳. استادیار آمار زیستی مؤسسه مطالعات و مدیریت جامع و تخصصی جمعیت کشور، نویسنده مسئول، -mahsa-saadati@gmail.com

۴. استادیار جمعیت‌شناسی مؤسسه مطالعات و مدیریت جامع و تخصصی جمعیت کشور، -hajiieh.raze-ghi@gmail.com

بدست آمده در این مقاله می‌باشد. زنان شاغل نسبت به زنان غیرشاغل در کوهورت موالید اول تعداد فرزندان ایده‌آل بیشتری دارند. همچنین، سن در اولین ازدواج برای کوهورت‌های موالید دوم و سوم که محل تولد آنان روستا بوده است، نیز نقش مهمی بر روی تعداد فرزندان ایده‌آل دارد.

واژگان کلیدی: داده‌کاوی، طبقه‌بندی، درخت تصمیم کارت، تعداد فرزندان ایده‌آل، زنان، سمنان

### مقدمه

داده‌کاوی<sup>۱</sup> فرآیندی است که در آغاز دهه ۹۰ میلادی پا به عرصه ظهور گذاشته و با نگرشی نو، به مسئله استخراج اطلاعات از پایگاه داده‌ها می‌پردازد. از سال ۱۹۹۵ داده‌کاوی به صورت جدی وارد مباحث آمار شد و در سال ۱۹۹۶، اولین شماره مجله «کشف دانش از پایگاه داده‌ها» منتشر گردید. در حال حاضر، داده‌کاوی مهمترین فناوری، جهت بهره‌برداری مؤثر از داده‌های حجیم می‌باشد و اهمیت کاربرد آن رو به فزونی است. هدف مدل حاصل از داده‌کاوی، اغلب طبقه‌بندی<sup>۲</sup> یا پیش‌بینی دقیق داده‌های جدید بر مبنای شناخت مدل و استفاده از آن برای پیش‌بینی می‌باشد. طبقه‌بندی روشی برای تحلیل داده‌هاست که مدلی برای توصیف دسته‌ها و طبقات مهم موجود در داده‌ها تولید می‌کند و یک فرایند دو مرحله‌ای متشکل از مرحله آموزشی (ساخت مدل طبقه‌بندی) و مرحله پیش‌بینی طبقه داده جدید (استفاده از مدل ساخته شده برای طبقه‌بندی داده‌های جدید) است. این مسئله در علوم اجتماعی به شکل گسترده‌ای کاربرد دارد (لریس، لومینی و مانا<sup>۳</sup> ۲۰۱۱).

درخت تصمیم<sup>۴</sup> یکی از مفیدترین روش‌ها در مسائل طبقه‌بندی است که فضای داده‌ها را به نواحی مستطیلی تقسیم می‌کند و یک مشاهده براساس ناحیه‌ای که در آن قرار می‌گیرد، طبقه‌بندی می‌شود. درخت تصمیم که معمولاً با نام درخت طبقه‌بندی<sup>۵</sup> شناخته می‌شود به عنوان یکی از مدل‌های طبقه‌بندی برای مواجهه با مجموعه داده‌ها

1. Data Mining
2. Knowledge Discovery in Database
3. Classification
4. Loris, Lumini and Manna
5. Decision Tree
6. Classification Tree

شامل تعداد زیاد متغیر بسیار مناسب می‌باشد (جیاوی و جی<sup>۱</sup> ۲۰۰۶). درخت تصمیم با استفاده از متغیرهای پیش‌بین یک گراف غیرچرخشی شبیه درخت را برای انجام طبقه‌بندی معرفی می‌نماید. این درخت از روش‌های ناپارامتری استفاده کرده و الگوی درختی آن می‌تواند به‌طور مستقیم برای تشخیص طبقه یک مشاهده جدید مورد استفاده قرار گیرد که همین نکته این روش را به یک روش کارا و کاربردی برای طبقه‌بندی تبدیل کرده است (پادگورلک<sup>۲</sup> و همکاران ۲۰۰۲؛ اولسن و دلن<sup>۳</sup> ۲۰۰۸).

دلیل استفاده از مدل درختی این است که روش‌های معمول آماری به سه دلیل عمده زیر نمی‌توانند در تشخیص و پیش‌بینی‌های علوم اجتماعی و جمعیت‌شناسی مورد استفاده قرار گیرند:

- در مطالعات اجتماعی معمولاً مقادیر بسیاری از متغیرهای پیش‌بین وجود دارند که حجم زیاد متغیرها استفاده از روش‌های معمول را با مشکلات و پیچیدگی‌هایی روبرو می‌سازد.
- روش‌های معمول آماری نیازمند وجود برخی فرضیات نظیر نرمال بودن توزیع و واریانس‌های برابر می‌باشند که در اکثر داده‌های واقعی معمولاً چنین فرضیاتی برقرار نخواهد بود.
- نتایج به‌دست آمده از روش‌های معمول آماری به سادگی قابل تفسیر نمی‌باشند. در حالی‌که درخت تصمیم یکی از روش‌های مناسب طبقه‌بندی می‌باشد که با وجود مشکلات ذکر شده می‌تواند به خوبی عمل نماید.
- روش‌های آماری موجود و درخت تصمیم که برای طبقه‌بندی و پیش‌بینی داده‌ها به‌کار می‌روند را می‌توان با یکدیگر به صورت زیرمقایسه نمود (لانگ<sup>۴</sup> و همکاران ۱۹۹۳؛ شوارتز<sup>۵</sup> و همکاران ۲۰۰۳؛ ملو<sup>۶</sup> و همکاران ۲۰۰۶). و مطالعات فرسبی
- تحلیل‌های تک متغیره: روش ساده آماری با کاربرد آسان است اما با توجه به در نظر نگرفتن روابط میان متغیرها، این روش از دقت کمتری نسبت به سایر روش‌های آماری برخوردار می‌باشد.

1. Jiawi and Jay
2. Podgorelec
3. Olson and Delen
4. Long
5. Schwarzer
6. Mello

- تحلیل‌های چند متغیره: این روش‌ها دقت آماری را افزایش می‌دهند. با این وجود، کارایی این روش‌ها به برقراری پیش‌فرض‌های بسیاری که بررسی آن‌ها پیچیده است، بستگی دارد.
- درخت تصمیم: این روش که روشی ناپارامتری است، دقت آماری را افزایش می‌دهد، کاربرد آن آسان است، تفسیر نتایج آن به سهولت انجام می‌گیرد و زمانی که داده‌ها با یک مدل خطی ساده به دقت برازش نیابند، دقیق‌تر از رگرسیون خطی عمل می‌نماید. با این‌که الگوهای درختی برای طبقه‌بندی قدمت زیادی دارند ولی معرفی کامل و جامع درخت طبقه‌بندی با تقسیمات دوتایی تحت عنوان مدل طبقه‌بندی و رگرسیون درختی (کارت)<sup>۱</sup> اولین بار توسط بریمن<sup>۲</sup> و همکارانش (۱۹۸۴) صورت گرفت. کوئینلان<sup>۳</sup> (۱۹۸۶) مدل‌های ID۳ و C۴,۵ را برای معرفی یک ساختار درختی با تقسیمات بیشتر از دو معرفی نمود. درخت طبقه‌بندی کارت با این‌که بر مبنای متغیرهای پیش‌بین کمی طراحی شده است، قابلیت تعمیم به متغیرهای رشته‌ای را نیز دارد. این روش در زمینه تشخیص و پیش‌بینی به‌طور گسترده در مقایسه با انواع دیگر درخت‌های طبقه‌بندی مورد استفاده قرار می‌گیرد (ویلکینسون<sup>۴</sup> ۱۹۹۲؛ مارشال<sup>۵</sup> ۲۰۰۱).
- استفاده از درخت تصمیم به دلیل دقت و سرعت الگوریتم‌های آن، همچنین سادگی در تفسیر نتایج به‌طور گسترده‌ای در علوم جمعیتی، کامپیوتری، پزشکی، روانشناسی، اقتصادی و غیره کاربرد دارد (بهویار<sup>۶</sup> ۲۰۱۴؛ رمضان‌خوانی<sup>۷</sup> و همکاران ۲۰۱۴؛ لریس، لومینی و مانا ۲۰۱۱). در این مقاله به منظور آشنایی پژوهشگران جمعیتی با مدل درختی کارت و بکارگیری این مدل برای طبقه‌بندی وقایع جمعیتی، از اطلاعات طرح «بررسی رفتارهای ازدواج و باروری زنان حداقل یکبار ازدواج کرده، ۱۵-۴۹ ساله در استان سمنان-۱۳۹۱» استفاده شد (رازقی نصرآباد ۱۳۹۳). از آن‌جا که به‌کارگیری مدل مناسب به‌منظور تبیین عوامل مؤثر بر روی تحولات باروری که تحولات جمعیتی را در چند دهه اخیر به دنبال داشته است، ضروری می‌باشد، هدف اصلی این مقاله استفاده از مدل مذکور در

1. Classification & Regression Tree (CART)

2. Breiman

3. Quinlan

4. Wilkinson

5. Marshall

6. Bhuyar

7. Ramezankhani

طبقه‌بندی تعداد فرزندان ایده‌آل زنان حداقل یکبار ازدواج کرده، ۱۵-۴۹ ساله استان سمنان بود.

روند تحولات باروری در ایران حاکی از آن است که در سه دهه اخیر باروری کاهش چشمگیری داشته و میزان باروری کل از هفت تولد برای هر زن در سال ۱۳۵۸ به ۲/۱ تولد در سال ۱۳۷۹، ۱/۹ تولد در سال ۱۳۸۵ و ۱/۸ تولد در سال ۱۳۹۰ کاهش یافته است (عباسی شوازی و حسینی چاووشی ۱۳۹۲). استان سمنان (جامعه مطالعه)، به تبعیت از تغییرات جمعیتی و کاهش باروری کشور، جزء استان‌هایی است که باروری زیر سطح جایگزینی را تجربه می‌کند. تحولات باروری استان سمنان نشان می‌دهد که میزان باروری کل این استان از ۴/۹ تولد در سال ۱۳۵۵ به ۳/۷ تولد در سال ۱۳۷۰ کاهش یافته، سپس با سرعت بیشتر در سال ۱۳۷۹ به ۲ فرزند رسیده، از سال ۱۳۸۵ کاهش ادامه داشته تا در سال ۱۳۹۰ به زیر سطح جایگزینی ۱/۵۹ فرزند رسیده است (همان).

با کاهش باروری و رسیدن به زیر سطح جایگزینی و یا پایین‌تر، نیات و ایده‌آل‌های باروری مورد توجه محققان زیادی قرار گرفته است. ایده‌آل باروری بخشی از فرایند تصمیم‌گیری باروری می‌باشد که تعداد فرزندان دلخواه افراد را نشان می‌دهد (میلر<sup>۱</sup> ۱۹۹۴) و مبنای انگیزشی برای نیات رفتاری آنان فراهم می‌کند (پروگینی و باگزی<sup>۲</sup> ۲۰۰۱). ایده‌آل‌های باروری می‌توانند به دو مفهوم مختلف ارجاع داده شوند، بنابراین به شکل سؤالی که از افراد پرسیده می‌شود، بستگی دارد. اولین مفهوم به ایده‌آل‌های باروری که به لحاظ اجتماعی مطلوب و ایده‌آل شناخته می‌شود، مرتبط است و دومین مفهوم به نیاتی اشاره دارد که به درک فردی اشخاص در مورد تعداد ایده‌آل خانواده مرتبط است. پاسخگویان در پیمایش‌های باروری و خانواده، اغلب ایده‌آل اجتماعی باروری<sup>۳</sup> را به عنوان درک خود از ایده‌آل جامعه تفسیر می‌کنند؛ به عنوان مثال اهداف باروری که به وسیله جامعه مثبت ارزشگذاری می‌شود و می‌توان این اهداف را با سئوالات: «شما فکر می‌کنید تعداد ایده‌آل و مناسب فرزند برای یک خانواده چه تعدادی است؟» (تستا<sup>۴</sup> ۲۰۰۶) یا «به نظر شما یک زوج ازدواج کرده چه تعداد فرزند باید داشته باشند؟» (لایت برنی و مکدونالد<sup>۵</sup> ۱۹۸۲؛ عباسی شوازی و همکاران ۱۳۸۳) مشخص نمود. در این مطالعه تمرکز بر درک فردی اشخاص

1. Miller
2. Perugini and Bagozzi
3. Fertility Social Ideal
4. Testa
5. Lightbourne and McDonald

از تعداد ایده‌آل می‌باشد. که از آن به عنوان ایده‌آل‌های فردی باروری<sup>۱</sup> نام می‌برند. این مفهوم به لحاظ تجربی ایده‌آل شخصی پاسخگویان در مورد تعداد فرزندان را اندازه‌گیری می‌کند. ایده‌آل فردی باروری انگیزه‌ها، نگرش‌ها و اعتقادات فردی را منعکس می‌کند و پاسخگویان ارزیابی شخصی خود را از اهداف بیان می‌کنند (میلر ۱۹۹۴). این مفهوم با سؤال «شما فکر می‌کنید چند فرزند برای خانواده شما ایده‌آل است؟» پرسیده می‌شود.

در بسیاری از جوامع بعد ایده‌آل و یا هنجار باروری، دو فرزند می‌باشد که به سطح جانشینی باروری<sup>۲</sup> نیز نزدیک است. مطالعات روی ایده‌آل‌های باروری نشان می‌دهد که اگرچه در کشورهای زیادی باروری واقعی به زیر سطح جایگزینی رسیده، اما باروری ایده‌آل دو فرزند است (ون دکا<sup>۳</sup> ۲۰۰۱).

وستاف<sup>۴</sup> (۲۰۱۰) با استفاده از داده‌های طرح سلامت و جمعیت، مطالعه‌ای با هدف بررسی تعداد فرزندان ایده‌آل در کشورهای در حال توسعه انجام داد. نتایج مطالعه وستاف نشان داد که در ۱۸ کشور از آسیا و آفریقای شمالی، دامنه فرزندان از دو در هند تا چهار در اردن و پاکستان نوسان دارد. در آمریکای لاتین و کارائیب دامنه از ۲/۲ در برزیل تا ۳/۷ در گواتمالا است. بالاترین تعداد ایده‌آل در آفریقای مرکزی و غربی بوده که دامنه‌اش از ۴/۸ در غنا تا ۹/۱ در نیجرو ۹/۲ در چاد می‌باشد.

در ایران اکثر مطالعات در حوزه باروری معطوف به تعداد فرزندان زنده به دنیا آورده زنان است و در مطالعه‌ای که در سال ۱۳۹۳ توسط سعادت‌ی و باقری انجام گرفت به مدلسازی درختی تعداد فرزندان زنده به دنیا آمده زنان سمنانی پرداخته شد. با توجه به نقش تعداد فرزندان ایده‌آل در باروری، این مطالعه بر تعداد ایده‌آل فرزندان تمرکز دارد و برای طبقه‌بندی و تبیین آن در استان سمنان از مدل درختی کارت استفاده شده است.

#### 1. Fertility Individual Ideals

۲. سطح جانشینی باروری معادل ۲/۱ می‌باشد که چنانچه میزان باروری کل نیز در این سطح و یا بالاتر باشد، می‌توان چنین استنباط کرد که زنان در نسل بعدی به اندازه کافی دختر برای جایگزین کردن خود در جمعیت دارند. زمانی که از هر زن یک دختر برای تجدید نسل باقی بماند میزان خالص تجدید نسل یک و باروری برابر با سطح جایگزینی می‌گردد.

3. Van de Kaa

4. Westoff

### مبانی نظری و تجربی

تمایلات و رفتارهای باروری، با دورویکرد ساختاری<sup>۱</sup> و ایده‌ای<sup>۲</sup> مطالعه می‌شوند. یکی از تبیین‌های مهم ساختاری نظریه مدرنیزاسیون است که با تأکید بر تغییرات ساختی و گذار از اقتصاد کشاورزی به اقتصاد صنعتی، تغییرات باروری را به تغییرات ساختاری در زندگی اجتماعی به جهت ظهور نیروهای جدید اجتماعی و سه جریان صنعتی شدن، شهرنشینی و آموزش همگانی مرتبط می‌کند. در فرایند مدرنیزاسیون، افراد به دنبال حداکثر سود و منفعت فردی خود هستند. براساس این تئوری، به دلیل فروپاشی ترتیبات سنتی حاکم بر زندگی روزمره افراد، همراه با استقرار سبک‌های نوین زندگی شهری و صنعتی، نگرش‌ها و تمایلات فرزندآوری تغییر می‌کند (بکر<sup>۳</sup> ۱۹۸۱؛ لیبنشتاین<sup>۴</sup> ۱۹۵۷).

کالدول<sup>۵</sup> (۱۹۷۶) تغییرات در رفتارها و تمایلات باروری را در جریان ثروت بین نسل‌ها می‌داند. علت این امر ظهور و جایگزینی خانواده هسته‌ای به جای خانواده گسترده است. در واقع، کالدول به رابطه مستقیم میان ساختار خانواده و باروری تأکید می‌کند. از نظر وی، در جوامع سنتی که ساختار خانواده گسترده عمومیت دارد، فرزندان منافع اقتصادی دارند و جریان ثروت رو به بالا و از فرزندان به والدین است. بنابراین، تصمیم‌گیری درباره داشتن فرزند زیاد، تصمیمی عقلانی محسوب می‌شود. در جوامع مدرن، هزینه فرزندان بالاست، والدین بیشتر به کیفیت فرزندان توجه دارند و جریان ثروت از بالا به پایین است. از این رو، پاسخ عقلانی افراد در کاهش تعداد فرزندان است.

مک‌دونالد (۲۰۰۲) با طرح نظریه اجتناب از خطر، بعد دیگری از نظریه‌های مربوط به هزینه و فایده فرزندان را ارائه می‌دهد. طبق نظر مک‌دونالد، هزینه‌ها و منافع در واقع هزینه‌ها و منافع آینده‌اند و براین اساس افراد درباره هزینه‌ها و منافع اطمینان ندارند. از این رو تصمیم‌گیری درباره باروری وابسته به جهت‌گیری‌های آینده خواهد بود. اگر افراد به این نتیجه برسند که آینده به لحاظ اقتصادی، اجتماعی و زندگی شخصی نامطمئن خواهد بود، ممکن است به لحاظ امنیتی دچار تردید شوند، زیرا در پی دفع مخاطره‌ها هستند. بر مبنای این نظریه فرض می‌شود که عدم اطمینان از شرایط اقتصادی اجتماعی

1. Structural
2. Ideational
3. Becker
4. Leibenstein
5. Caldwell
6. Risk Aversion



باعث می‌شود که افراد فرزند کمتری داشته باشند، متمایل به داشتن فرزند کمتری باشند و یا تشکیل خانواده و داشتن فرزند (بیشتر) را به آینده دورتر (دوران اطمینان اقتصادی) محول سازند، زیرا «تشکیل خانواده یا والد شدن جز تعهدات بلند مدت و به شدت نیازمند منابع است و زمانی که شرایط کاری و درآمدی افراد ناامن باشد، این ترتیبات به آینده موکول می‌شوند یا مورد چشم‌پوشی قرار می‌گیرند» (برناردی، کلارنر و اچ واندرا<sup>۱</sup> ۲۰۰۸). دسته‌ای دیگر از تبیین‌ها بر نقش تغییرات ایده‌ای و نگرشی به‌منزله یکی از مؤلفه‌های مهم تغییرات جمعیتی تأکید بیشتری دارند. یکی از تبیین‌های ایده‌ای مهم، تئوری انتقال دوم جمعیتی<sup>۲</sup> است. کاهش باروری و رسیدن به زیر سطح جانشینی با تغییرات در ارزش‌ها، هنجارها و رفتارها همراه شده است. این پدیده‌ای است که به‌منزله تغییرات تشکیل خانواده و فرزندآوری به آن تأکید می‌شود. انتقال دوم جمعیتی بحث می‌کند که دلایل تغییر در شاخص‌های جمعیتی نتیجه تغییر ارزش‌هاست (وندکا ۱۹۹۷). تصمیمات درباره مسیر زندگی و رفتارها، با ایده‌آل‌ها درباره روش‌های مناسب زندگی در خانواده پیوند دارد (لسته‌هاق و مورس<sup>۳</sup> ۲۰۰۰). از جمله آن‌که ارزش‌ها در زمینه فرزندآوری در خانواده مدرن، به سمت کاهش تعداد فرزندان است. با این حال، در مواردی نیز ایده‌آل‌ها و ارزش‌ها از فرزندآوری و داشتن فرزند حمایت می‌کند. از جمله هنگامی که بر فرزندآوری به‌منزله هدف ثانویه بعد از خودشکوفایی و افزایش سرمایه اجتماعی و رسیدن به زندگی با مرکزیت خانواده تمرکز می‌شود. در واقع، کاهش تعداد فرزندان برای اهمیت دادن زیاد به خانواده و فرزندان اتفاق می‌افتد (وندکا ۱۹۹۷).

براساس نظریه رفتار برنامه‌ریزی شده، هر شخص برای انجام یک رفتار، نیت انجام آن را از قبل در خود ایجاد می‌کند. این نیت همواره به‌صورت یک آمادگی ذهنی رفتاری باقی می‌ماند، تا این‌که در فرصت و زمان مناسب، فرد اقدام به تبدیل آن به عمل کند (آجزن<sup>۴</sup> ۲۰۰۵). این تئوری معتقد است که انسان‌ها معمولاً بین تصمیم‌گیری‌های خود و درگیر شدن در یک رفتار، بطور منطقی عمل کرده و بر طبق ترجیحاتشان رفتار می‌کنند. برای مثال، وقتی زن و شوهر هر دو قصد داشتن فرزند در آینده نزدیک را داشته باشند، به احتمال زیاد این فرزندآوری رخ خواهد داد. اما چنان‌چه فقط یکی از زوجین چنین قصد و نیتی را داشته باشد، احتمال وقوع فرزندآوری پایین خواهد بود.

1. Bernardi, Klärner and H Von der
2. Second Demographic Transition (SDT)
3. Lesthaeghe and Moors
4. Ajzen



طبق نظریه رفتار برنامه‌ریزی شده، نیت یک تعیین‌کننده رفتار می‌باشد و چنان‌چه نیت رفتاری به درستی اندازه‌گیری شوند، میزان قابل ملاحظه‌ای از واریانس در رفتار واقعی را تبیین می‌کنند. نظریه رفتار برنامه‌ریزی شده مبتنی بر این فرض است که مردم معمولاً به شیوه عقلانی رفتار می‌کنند، یعنی آن‌ها پس از دریافت اطلاعات موجود و بررسی پیامدهای اعمال خود، اقدام به انجام یک رفتار می‌کنند. از این‌رو، نیت فرد برای انجام یا عدم انجام یک رفتار مهمترین تعیین‌کننده بلافصل آن عمل است.

در چارچوب نظریه‌های فوق محققان مختلف به بررسی عوامل مؤثر بر تعداد ایده‌آل فرزندان پرداخته‌اند. گلدستاین، لوتز و تستا<sup>۱</sup> (۲۰۰۳) اشاره می‌کنند که ایده‌آل باروری افراد جوان در آلمان و اتریش که در بستری از باروری زیر سطح جایگزینی زندگی می‌کنند، به کمتر از دو فرزند کاهش یافته است. از نظر آن‌ها شرایط نسلی باروری باعث چرخشی رو به پایین برای باروری ایده‌آل می‌شود، به گونه‌ای که تعداد ایده‌آل فرزندان هر نسل تحت تأثیر رژیم باروری که در آن رشد یافته‌اند، می‌باشد. از نظر این محققان هر دو باروری واقعی و ایده‌آل، هر چند در سطوح غیریکنواخت، روند کاهشی مشابهی را طی می‌کنند.

از نظر بونگارت<sup>۲</sup> (۲۰۰۱) رتبه نهایی باروری برابر با تعداد ایده‌آل، ضرب در مجموعه‌ای از عواملی که تعداد ایده‌آل را مشخص می‌کنند، می‌باشد. بر این اساس می‌توان گفت رتبه دلخواه تحت تأثیر مجموعه‌ای از عوامل تسهیل‌کننده و بازدارنده، از جمله اختلالات باروری، موالید ناخواسته، داشتن شریک جنسی، ترکیب جنسیتی قابل قبول و عامل نهایی هزینه‌های فرصت از دست رفته می‌باشد (کاسنل وال و مورگان<sup>۳</sup> ۲۰۰۳).

ترکیب جنسی فرزندان یکی دیگر از عواملی است که نیت باروری افراد را تحت تأثیر قرار می‌دهد. پرتی و ساینگ<sup>۴</sup> (۲۰۱۰) نشان دادند که در هند خانواده‌هایی که تعداد فرزند دختر بیشتر و یا این‌که فقط فرزند دختر دارند، نسبت به کسانی که تعداد فرزندان دختر و پسر آن‌ها با یکدیگر برابر و یا تعداد فرزندان پسر بیش از تعداد فرزند دختر است، تعداد فرزند بیشتری را ایده‌آل می‌دانند و این افراد ممکن است در آینده تصمیم برای فرزندآوری داشته باشند.

مطالعه عباسی شوازی و همکاران (۱۳۸۳) با عنوان تحولات باروری در ایران که در چهار استان منتخب کشور شامل آذربایجان غربی، سیستان و بلوچستان، گیلان و

1. Goldstein, Lutz and Testa

2. Bongaarts

3. Quesnel-Vallée and Morgan

4. Pretti and Singh

یزد انجام شد در این زمینه از جمله مطالعات جامعی است که علاوه بر مطالعه عوامل مختلف اقتصادی، اجتماعی و فرهنگی مؤثر بر باروری، به چگونگی تحولات سریع باروری پرداخته و رفتار باروری و نگرش زنان نسبت به باروری ایده‌آل را مورد بررسی قرار داده است. براساس نتایج این مطالعه بجز استان سیستان و بلوچستان که ۴۵ درصد از زنان چهار فرزند را ایده‌آل می‌دانند در بقیه استان‌ها دو فرزند مطلوب است و داشتن بیش از سه یا چهار فرزند به عنوان خانواده بزرگ، رفتاری غیرهنجاری محسوب می‌شود.

حسینی و عباسی شوازی (۱۳۸۸) در مطالعه‌ای با عنوان تغییرات اندیشه‌ای و تأثیر آن بر رفتار و ایده‌آل‌های باروری زنان ترک و کرد، نشان دادند که هر دو گروه قومی دو فرزند را به عنوان شمار فرزندان ایده‌آل برای یک زوج می‌دانند هر چند ۲۰ تا ۲۹ درصد از زنان کرد داشتن سه فرزند و یا بیشتر را ایده‌آل دانسته‌اند. از طرف دیگر درصد نسبتاً بالایی از زنان کرد یک فرزند را بعد ایده‌آل فرزند می‌دانند.

عباسی شوازی و عسکری ندوشن (۱۳۸۴) در مطالعه‌ای با عنوان تغییرات خانواده و کاهش باروری در ایران: مطالعه موردی در یزد، تأثیر ابعاد و تغییرات خانواده بر نگرش‌ها و رفتارهای باروری را مورد بررسی قرار دادند. نتایج این مطالعه نشان داده است که زوج‌هایی که در محیط شهر پرورش یافته‌اند، زنانی که از سطوح بالاتر تحصیلی برخوردارند و زنانی که در سنین بالاتر ازدواج می‌کنند، در مقایسه با سایرین فرزندان کمتری را برای هر زوج کافی می‌دانند.

رازقی نصرآباد، عباسی شوازی و حسینی چاووشی (۱۳۹۳) در یک مطالعه کیفی با عنوان پدیدارشناسی زمان تولد اولین فرزند در بین زنان تهرانی نشان دادند که بسیاری از زنان نیاز عاطفی به فرزند، تکمیل بعد خانواده و هنجار جامعه را از دلایل اصلی تمایل به فرزندآوری می‌دانستند و از نظر آن‌ها داشتن دو فرزند برای تأمین این هدف کافی است. مرور مطالعات پیشین نشان می‌دهد که بررسی تعداد ایده‌آل فرزندان و عوامل مؤثر بر آن در سال‌های اخیر مورد توجه بسیاری از جمعیت‌شناسان قرار گرفته است. براساس این مطالعات، مجموعه‌ای از عوامل اقتصادی، اجتماعی، نگرش‌ها و ایده‌آل‌های افراد تصمیمات باروری را تحت تأثیر قرار می‌دهد. با توجه به این‌که در این مقاله، هدف اصلی معرفی روش مدل درختی کارت به منظور طبقه‌بندی تعداد ایده‌آل فرزندان می‌باشد، برخی از مشخصه‌های جمعیتی و اجتماعی مؤثر بر تعداد ایده‌آل باروری در این مدل وارد شده است (برای آشنایی با تأثیر سایر عوامل بر تعداد ایده‌آل فرزندان می‌توان به طرح رازقی نصرآباد (۱۳۹۳) مراجعه نمود).

## روش

برای طبقه‌بندی تعداد فرزندان ایده‌آل از اطلاعات طرح «بررسی رفتارهای ازدواج و باروری زنان حداقل یکبار ازدواج کرده، ۱۵-۴۹ ساله در استان سمنان-۱۳۹۱» استفاده شد (رازقی نصرآباد ۱۳۹۳). داده‌های این طرح که هدف آن، مطالعه تغییرات ایده‌آل‌ها و رفتارهای باروری و شناخت عوامل مؤثر بر آنها بود، در یک مطالعه پیمایشی-مقطعی با استفاده از پرسشنامه ساختاریافته در پاییز سال ۱۳۹۱ جمع‌آوری گردید. ۴۰۵ زن ۱۵-۴۹ ساله متعلق به خانوارهای معمولی ساکن استان سمنان که حداقل یکبار ازدواج کرده‌اند، نمونه طرح را تشکیل داد که با توجه به فرمول کوکران و با در نظر گرفتن سطح خطای ۵ درصد به صورت زیر محاسبه شد.

$$n = \frac{(Z_{1-\alpha/2})^2 p(1-p)}{d^2} \quad (1)$$

در این معادله:

$Z_{1-\alpha/2}$ : صدک  $1 - \alpha/2$  بالایی توزیع نرمال استاندارد در سطح معنی‌داری ۹۵ درصد (تقریباً برابر با ۲).

$p$ : نسبت زنان ۱۵-۴۹ ساله حداقل یکبار ازدواج کرده (این زنان شامل زنان دارای همسر، بی‌همسر بر اثر فوت همسر و بی‌همسر بر اثر طلاق می‌باشند) از کل زنان ۱۰ ساله و بیشتر که متعلق به یکی از خانوارهای معمولی ساکن در استان سمنان طبق اطلاعات سرشماری سال ۱۳۸۵ هستند (۰/۶۷).  
 $d$ : خطای نسبی نمونه‌گیری (۰/۰۵).

در نظر گرفته می‌شود. در نتیجه تعداد ۴۰۵ نمونه با در نظر گرفتن نرخ بی‌پاسخی ۱/۱۴۴ بدست آمد.

در این طرح با توجه به اهداف آن، متغیرهای گوناگونی اندازه‌گیری شد که تعدادی از آن‌ها با در نظر گرفتن تأثیرشان بر روی طبقه‌بندی تعداد فرزندان ایده‌آل، به عنوان متغیرهای پیش‌بین انتخاب شدند.

## تعاریف و مفاهیم

متغیرهای این مطالعه به صورت زیر تعریف می‌شوند که تعداد فرزندان ایده‌آل در زمان تحقیق متغیر پاسخ و سایر متغیرها به عنوان متغیر پیش‌بین در نظر گرفته شده‌اند:  
تعداد فرزندان ایده‌آل در زمان تحقیق: بیانگر تعداد فرزندان می‌باشد که از نظر

پاسخگویان ایده آل است و با سؤال «در حال حاضر چند فرزند برای شما ایده آل است؟» مشخص شد. این متغیر، از نوع نسبتی و گسسته است که در این مطالعه با چهار سطح ۰، ۱، ۲ و ۳ فرزند و بیشتر اندازه گیری شد.

**سن در اولین ازدواج:** این متغیر بیانگر سن پاسخگو در زمان اولین ازدواج بوده و برحسب سال اندازه گیری شده است.

**نوع ازدواج:** این متغیر که نسبت فامیلی بین زن و شوهر را نشان می دهد با دو گزینه خویشاوند و غیرخویشاوند مشخص شده است.

**سطح تحصیلات:** این متغیر در دو سطح دیپلم و بالاتر و زیر دیپلم تعیین شده است. **وضعیت شغلی:** اشتغال انجام هر گونه فعالیت است که پاسخگو به عنوان شغل خود اعلام و بابت آن مزد دریافت می کند و یا به نوعی در تولید درآمد خود سهیم است (اعم از اشتغال در منزل و یا خارج از منزل) (عباسی شوازی و همکاران ۱۳۸۳). این متغیر به صورت دو سطحی با دو گزینه شاغل و غیرشاغل سنجیده شده است. غیرشاغلان کلیه زنان خانه دار، محصل و مستمری بگیر را در بردارند.

**محل تولد:** این متغیر به صورت دو سطحی با دو گزینه شهر و روستا مشخص شده است.

**کوهورت مولید:** کوهورت مولید عبارت است از مولید یک گروه از افراد که در دوره یکسانی متولد شده اند و زمان فردی مشابهی را در بستر تاریخی یکسانی تجربه کرده اند (رایدر<sup>۱</sup> ۱۹۶۵). نمونه بررسی شده شامل زنان ۱۵-۴۹ ساله هستند که متولدین سالهای ۱۳۴۰ تا ۱۳۷۵ را شامل می شوند و در این مطالعه در سه کوهورت مولید قرار گرفته اند. متولدین سالهای ۱۳۴۰ تا ۱۳۴۹ به عنوان کوهورت مولید دهه ۱۳۴۰، متولدین سالهای ۱۳۵۰ تا ۱۳۵۹ به عنوان کوهورت مولید دهه ۱۳۵۰ و متولدین سالهای ۱۳۶۰ تا ۱۳۷۵ به عنوان کوهورت مولید بعد از سال ۱۳۶۰ در نظر گرفته شده اند.

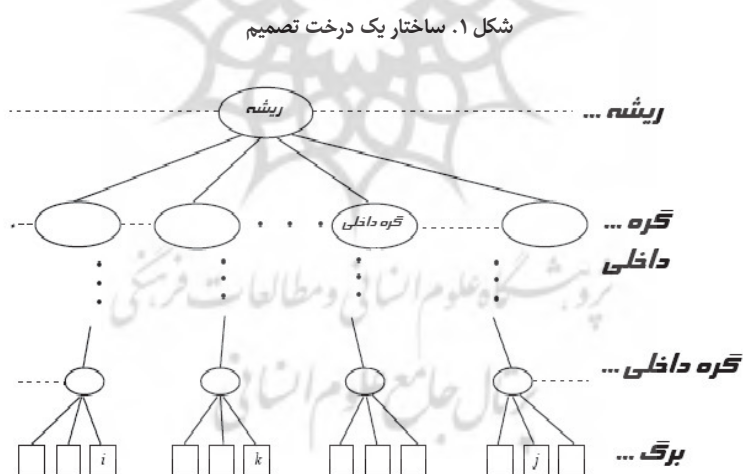
#### روش آماری (روش تجزیه و تحلیل داده ها)

یکی از رایج ترین کاربردهای داده کاوی، استفاده از آن برای ساختن مدل هایی برای پیش بینی طبقه یک مشاهده براساس ویژگی های آن می باشد که طبقه بندی نام دارد. طبقه بندی یکی از روش های داده کاوی است که در ابتدا با استفاده از داده های آموزشی، الگویی را تعیین می کند و سپس از آن برای طبقه بندی مشاهدات جدید استفاده می نماید.

1. Ryder

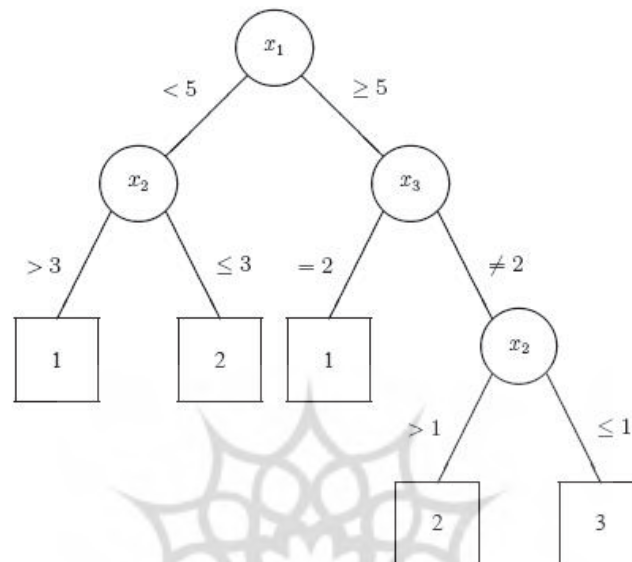
یک درخت تصمیم از سه جزء اصلی ریشه<sup>۱</sup>، گره داخلی<sup>۲</sup> و گره خارجی (برگ)<sup>۳</sup> تشکیل شده است. در روند ساخت درخت تصمیم ابتدا یک متغیر پیش‌بین به عنوان ریشه انتخاب می‌گردد و با توجه به یک سری از سؤالات و ویژگی‌ها به چندین گره داخلی تقسیم می‌شود؛ به تقسیمات ایجاد شده یک زیر درخت می‌گویند. به عنوان مثال، مشاهدات براساس گروه‌های سنی مختلف به دو یا چند گروه و یا براساس جنسیت به دو گروه تقسیم می‌شوند. هر گره داخلی نیز مانند ریشه به گره‌های دیگری تقسیم می‌شود تا در نهایت به هر گره یک طبقه از متغیر پاسخ منتسب می‌شود که این گره‌ها برگ نامیده می‌شوند، در واقع برگ‌ها، آخرین لایه گراف درخت تصمیم هستند. ساختار یک درخت تصمیم در شکل (۱) نشان داده شده است.

مثالی از یک درخت تصمیم در شکل (۲) نشان داده شده است. به عنوان مثال، برای یک فرد، اگر  $X_1$  کوچکتر از پنج و  $X_2$  بزرگتر از سه باشد، آنگاه این فرد به طبقه اول تعلق دارد. اگر  $X_1$  بزرگتر یا مساوی پنج و فرد به طبقه دو متغیر  $X_3$  متعلق باشد آنگاه فرد به گروه اول تعلق دارد. جدول (۱) قوانین مستخرج از این درخت را نشان می‌دهد.



شکل ۲. مثالی از درخت طبقه‌بندی با سه متغیر پیش‌بین (بیضی‌ها) و متغیر پاسخ سه طبقه‌ای (مستطیل‌ها)

1. Root
2. Internal node
3. External node (leaf)



جدول ۱. قوانین مستخرج از درخت تصمیم شکل (۲)

آن گاه فرد متعلق به طبقه اول	۱- اگر $x_1$ کوچکتر از ۵ و $x_2$ بزرگتر از ۳
آن گاه فرد متعلق به طبقه دوم	۲- اگر $x_1$ کوچکتر از ۵ و $x_2$ کوچکتر یا مساوی ۳
آن گاه فرد متعلق به طبقه اول	۳- اگر $x_1$ بزرگتر یا مساوی ۵ و $x_3$ مساوی ۲
آن گاه فرد متعلق به طبقه دوم	۴- اگر $x_1$ بزرگتر یا مساوی ۵ و $x_3$ مخالف ۲ و $x_2$ بزرگتر از ۱
آن گاه فرد متعلق به طبقه سوم	۵- اگر $x_1$ بزرگتر یا مساوی ۵ و $x_3$ مخالف ۲ و $x_2$ کوچکتر یا مساوی ۱

درخت تصمیم در بسیاری از زمینه‌ها به کار برده می‌شود اما مهمترین جایگاه و کاربرد آن زمانی است که پژوهشگر با مشاهدات و متغیرهای گوناگون سروکار دارد و نمی‌تواند به سادگی از روش‌های کلاسیک موجود استفاده نماید (ساهو و کومار، ۲۰۱۴). این روش یکی از روش‌های آماری ناپارامتری می‌باشد که با به‌کارگیری روش‌های نسبتاً ساده، یک الگوی طبقه‌بندی را برای مشاهدات موجود معرفی می‌نماید. با این‌که این روش از راهکارهای

نسبتاً ساده‌ای استفاده می‌نماید ولی در زمینه تشخیص و پیش‌بینی می‌تواند به خوبی روش‌های پیچیده عمل نماید (ویتن و فرانک<sup>۱</sup> ۲۰۰۵).

معمولاً برای سنجش کیفیت یک مدل پیش‌بینی، از دقت آن مدل استفاده می‌شود. روش اساسی در برآورد و سنجش مدل داده‌کاوی، ابتدا ساختن مدلی با استفاده از بخشی از مجموعه داده‌ها و سپس به کارگیری سایر داده‌ها برای برآورد خطای پیش‌بینی در مدل مورد نظر است. داده‌هایی که در بخش اول مورد استفاده قرار می‌گیرند، نمونه آموزشی<sup>۲</sup> و داده‌هایی که در بخش دوم به منظور اعتبارسنجی حاصل از نمونه آموزشی مورد استفاده قرار می‌گیرند، نمونه آزمون<sup>۳</sup> نامیده می‌شوند (تیموفیو<sup>۴</sup> ۲۰۰۴). این راهکار ساده که تنها برای مجموعه داده‌های بزرگ کاربرد دارد، بر این فرض استوار است که نمونه آموزشی و نمونه آزمون، توزیع یکسان ولی ناشناخته دارند.

به طور کلی برای شناخت و معرفی یک درخت تصمیم مناسب و کارا باید سه مرحله زیر طی شود:

- در مرحله اول باید شرایط و معیارهایی برای ساخت و رشد یک درخت تصمیم معرفی شود.
  - در مرحله دوم سعی می‌شود درختی با اندازه کوچکتر ولی با همان دقت یا دقت بیشتر ایجاد شود.
  - در مرحله سوم باید مناسب بودن درخت معرفی شده بررسی شود.
- شاخص جینی<sup>۵</sup> توسط بریمن و همکارانش در سال ۱۹۸۴ برای انتخاب متغیر مناسب در درخت طبقه‌بندی معرفی گردید. این شاخص معمولاً در مدل‌های درختی با تقسیمات دوتایی در هر گره، مورد استفاده قرار می‌گیرد. شاخص جینی با استفاده از معادله (۲) تعریف می‌شود.

$$i(t) = Gini(t) = 1 - \sum_{j=1}^k p_j^2 [c = c_j | T = t] \quad (2)$$

رابطه فوق زمانی که مشاهدات فقط متعلق به یک طبقه باشند، برابر صفر و وقتی

1. Witten and Frank
2. Learning sample
3. Test sample
4. Thimofeew
5. Gini Index



احتمال تعلق به هر طبقه با هم برابر باشد بیشترین مقدار ممکن را اختیار می‌نماید. در این معادله مقدار  $p$  با استفاده از دو پیش‌فرض احتمالات برآوردی و برابر متعلق به هر طبقه از روی داده‌ها قابل محاسبه است.

الگوریتم کارت برای هر دو نوع متغیر پاسخ کیفی و کمی کاربرد دارد. این روش برای متغیرهای مستقل کمی طراحی شده است، اما قابل تعمیم و استفاده برای متغیرهای کیفی نیز می‌باشد. درخت طبقه‌بندی کارت از شاخص جینی برای تعیین متغیرهای مناسب استفاده می‌نماید، با این حال می‌توان از شاخص‌های دیگری نظیر آنتروپی و آماره کای-دو نیز استفاده نمود (تیموفیو ۲۰۰۴). مزیت شاخص جینی نسبت به آنتروپی و شاخص‌های دیگر سرعت بالاتر آن در انجام محاسبات می‌باشد. بنابراین در این مطالعه از شاخص جینی استفاده شده است.

بدیهی است که در تمام روش‌های طبقه‌بندی دستیابی به الگویی که بدون خطا باشد، غیرممکن است، ولی هدف معرفی الگویی با کمترین خطا می‌باشد. از این رو یکی از مهمترین نکات مورد توجه در معرفی یک الگوی درختی، بررسی دقت و کارایی آن می‌باشد. هدف از معرفی یک درخت طبقه‌بندی دستیابی به یک الگوی طبقه‌بندی براساس مشاهدات موجود می‌باشد به طوری که این الگوی معرفی شده از دقت مناسبی برای مشاهدات جدید برخوردار باشد. یعنی این مدل بتواند در مورد یک مشاهده جدید به خوبی و با کمترین خطای ممکن، تصمیم‌گیری نماید. با توجه به این که در نظر گرفتن میزان طبقه‌بندی اشتباه براساس نمونه‌ای که درخت طبقه‌بندی با توجه به آن ساخته شده است، به عنوان میزان خطا مناسب نمی‌باشد، روش‌هایی برای برآورد میزان خطای یک درخت طبقه‌بندی مبتنی بر روش‌های روایی متقاطع<sup>۲</sup> مورد استفاده قرار می‌گیرند. روایی متقاطع  $k$  مرتبه‌ای برای زمانی که اندازه مشاهدات موجود با توجه به متغیرهای پیش‌بین کافی نباشد، مناسب است. در این روش مشاهدات موجود به‌طور تصادفی به  $k$  قسمت با اندازه‌های مساوی تقسیم می‌شوند و در هر مرحله  $1-k$  قسمت به عنوان نمونه آموزشی و قسمت دیگر به عنوان نمونه آزمون در نظر گرفته می‌شود. با توجه به نمونه آموزشی معرفی شده درخت مناسب ایجاد می‌شود و خطای مربوط به آن براساس نمونه آزمون برآورد می‌گردد. از آن جا که مشاهدات به  $k$  قسمت تقسیم شده‌اند این اقدام نیز  $k$  بار صورت می‌پذیرد و  $k$  اندازه خطا به دست می‌آید که میانگین این خطاها به عنوان برآوردی از

1. Chi-square Statistic

2. Cross validation

### معرفی و کاربرد مدل کارت برای طبقه‌بندی تعداد فرزندان ایده‌آل...

خطای مدل درختی با توجه به کل مشاهدات به دست می‌آید. این روش از دقت مناسبی در برآورد خطا برخوردار است و برای نمونه‌های کوچک مناسب می‌باشد (شافر<sup>۱</sup> ۱۹۹۳). در محاسبه خطا معمولاً از ماتریس اغتشاش<sup>۲</sup> استفاده می‌شود. این ماتریس که برای یک متغیر دو سطحی الگویی مشابه شکل (۳) دارد مقادیر واقعی طبقه متغیر مورد نظر را با مقادیر پیش‌بینی شده طبقه آن متغیر توسط مدل مقایسه می‌کند.

شکل ۳. ماتریس اغتشاش برای طبقه‌بندی یک متغیر دو سطحی

۱	۰	مقدار پیش‌بینی طبقه	
b	a		مقدار واقعی طبقه
d	c		

معادله (۳) نحوه محاسبه دقت را برای ماتریس شکل (۳) نشان می‌دهد.

$$دقت = \frac{a+d}{a+b+c+d} \quad (۳)$$

هر چه طبقات پیش‌بینی شده و واقعی به یکدیگر نزدیکتر باشند، مدل از دقت بالاتری برخوردار خواهد بود. به عبارتی دیگر در این ماتریس هر چه مقادیر روی قطر اصلی قرار گیرند، دقت مدل بیشتر است. در حالت ایده‌آل کلیه عناصر سازنده ماتریس بر روی قطر اصلی قرار می‌گیرند که در این حالت دقت مدل ۱۰۰ درصد خواهد شد و در بدترین حالت هیچ یک از مؤلفه‌های ماتریس روی قطر اصلی قرار نداشته و دقت مدل برابر صفر می‌شود. در عمل دقت مدل، عددی مابین این دو مقدار فرین است؛ بنابراین هر چه دقت مدل بزرگتر باشد، اعتبار آن مدل در پیش‌بینی متغیر مورد نظر بیشتر است.

با وجود چندین الگوریتم مختلف برای معرفی و ساخت درخت تصمیم، الگوریتم کارت امتیازاتی دارد که استفاده از آن را در عمل تا حد زیادی افزایش داده است. کارت، قواعد ساده‌ای برای تعیین گروه‌های با ریسک بالا یا پایین نسبت به متغیر مورد نظر فراهم می‌کند (هلف پی آر دی<sup>۳</sup> ۲۰۰۲). در واقع، کارت زیرگروه‌های همگن را با استفاده از روش‌های ناپارامتری استخراج می‌کند. در نتیجه این الگوریتم برای آنالیز اکتشافی داده‌ها

1. Schaffer
2. Confusion Matrix
3. Holfprd

و ایجاد طبقه‌بندی ساده و قابل تفسیر، همچنین تعریف ضوابط پیش‌بینی با کمترین پیش‌فرض‌ها مفید است (سیسیالانو و مولا<sup>۱</sup> ۲۰۰۰). سادگی در تفسیر نتایج، آزاد توزیع بودن شاخص‌های به کار رفته در ساخت درخت و نحوه برخورد آن با داده‌های گمشده و دورافتاده از مزایای مهم مدل کارت است. در این مقاله نیز با توجه به این موارد از مدل کارت برای طبقه‌بندی تعداد فرزندان ایده‌آل در زمان تحقیق استفاده شد. در ادامه به مدلسازی این متغیر با استفاده از نرم افزار STATISTICA نسخه ۷ پرداخته شده است.

#### یافته‌ها

جدول (۲) توزیع و درصد فراوانی متغیرهای تعداد فرزندان ایده‌آل در زمان تحقیق، کهورت مولید، وضعیت شغلی، سطح تحصیلات، نوع ازدواج و محل تولد طرح بررسی رفتارهای ازدواج و باروری زنان حداقل یکبار ازدواج کرده، ۱۵-۴۹ ساله در استان سمنان-۱۳۹۱ را نشان می‌دهد. با توجه به این جدول تعداد ایده‌آل  $70/4$  درصد زنان، ۲ فرزند، ۱۱ درصد زنان یک فرزند و  $4/4$  درصد آنان بی‌فرزندی می‌باشد.

میانگین سن در اولین ازدواج  $3/15 \pm 20/57$  سال و فراوانی ازدواج در سن ۱۸ سالگی در میان زنان ۱۵-۴۹ ساله بیشتر از سایر سنین بوده است.

پژوهشگاه علوم انسانی و مطالعات فرهنگی  
پرتال جامع علوم انسانی

1. Siciliano and Mola

معرفی و کاربرد مدل کارت برای طبقه بندی تعداد فرزندان ایده آل...

جدول ۲. توزیع و درصد فراوانی متغیرهای مطالعه

درصد فراوانی	فراوانی	متغیر	
		مقدار	نام
۴/۴	۱۸	۰	تعداد فرزندان ایده آل در زمان تحقیق (متغیر پاسخ)
۱۱/۱	۴۵	۱	
۷۰/۴	۲۸۵	۲	
۱۴/۱	۵۷	۳ و بیشتر	
۳۲/۳	۱۳۱	دهه ۱۳۴۰	کوهورت مولید
۳۵/۸	۱۴۵	دهه ۱۳۵۰	
۳۱/۹	۱۲۹	دهه ۱۳۶۰ به بعد	
۸۰/۲	۳۲۵	غیرشاغل	وضعیت شغلی
۱۹/۸	۸۰	شاغل	
۶۷/۱	۲۷۱	زیردیپلم	سطح تحصیلات
۳۳/۰	۱۳۴	دیپلم و بالاتر	
۵۹/۳	۲۴۰	غیرخویشاوندی	نوع ازدواج
۴۰/۷	۱۶۵	خویشاوندی	
۷۷/۳	۳۱۳	شهر	محل تولد
۲۲/۷	۹۲	روستا	
	۴۰۵		کل نمونه

جدول توافقی متغیرهای پاسخ و متغیرهای پیش بین مورد مطالعه در جدول (۳) ارائه شده است. در کلیه کوهورت های مولید دو فرزندی بیشترین تعداد ایده آل فرزندان بوده

است. در کوهوت موالید اول (دهه ۱۳۴۰)، بی فرزندی ایده‌آل هیچ زنی نبوده است. این کوهوت بیشترین درصد (۲۶/۷ درصد) سه فرزند و بیشتر را ایده‌آل دانسته‌اند. بیشترین درصد ایده‌آل بی فرزندی با ۸/۵ درصد متعلق به کوهوت موالید سوم بوده است. دو گروه شاغل و غیر شاغل بیشترین ایده‌آل‌شان دو فرزندی بوده است و ایده‌آل سه فرزندی و بیشتر در میان غیر شاغلان (۱۴/۲ درصد) و شاغلین (۱۳/۸ درصد) تقریباً با یکدیگر برابر است. در میان سطوح تحصیلی ایده‌آل دو فرزندی در هر دو سطح دیپلم و بالاتر و زیر دیپلم بیشترین مقدار را به خود اختصاص داده‌اند. زنان با تحصیلات زیر دیپلم بیشتر از زنان با تحصیلات دیپلم و بالاتر تعداد سه فرزند و بیشتر را ایده‌آل خود دانسته‌اند. زنان با تحصیلات دیپلم و بالاتر بیش از دو برابر زنان با تحصیلات زیر دیپلم بی فرزندی را ایده‌آل دانسته‌اند. در میان زنان با ازدواج خویشاوندی ۱۰/۹ درصد تک فرزندی را ایده‌آل خود دانسته‌اند این در حالی است که ۱۱/۳ درصد زنان با ازدواج غیرخویشاوندی تک فرزندی را ایده‌آل دانسته‌اند. ۲۱/۷ درصد از زنان متولد روستا سه فرزند و بیشتر را در مقابل ۱۱/۸ درصد زنان متولد شهر ایده‌آل دانسته‌اند. ایده‌آل بی فرزندی در میان زنان متولد شهر و روستا به ترتیب برابر ۴/۸ و ۳/۳ درصد است. بیشترین تعداد ایده‌آل ۳ فرزند و بیشتر با ۱۹/۴ درصد در میان زنانی است که در سنین ۱۹-۲۱ سال ازدواج کرده‌اند. بیشترین تعداد ایده‌آل بی فرزندی را با ۸/۶ درصد زنانی که در بازه ۲۲-۲۴ سالگی ازدواج کرده‌اند، به خود اختصاص داده‌اند. ایده‌آل ۲ فرزندی در میان کلیه گروه‌های سنی ازدواج بیشترین میزان را به خود اختصاص داده است.

پژوهشگاه علوم انسانی و مطالعات فرهنگی  
پرتال جامع علوم انسانی

معرفی و کاربرد مدل کارت برای طبقه‌بندی تعداد فرزندان ایده‌آل...

جدول ۳. جدول توافقی متغیر پاسخ و متغیرهای پیش‌بین مورد مطالعه

درصد فراوانی	تعداد فرزندان ایده‌آل در زمان تحقیق (متغیر پاسخ)				متغیر	
	۳ <	۲	۱	۰	مقدار	نام
۱۰۰	۲۶/۷	۶۷/۲	۶/۱	۰/۰	دهه ۱۳۴۰	کوهورت مولید
۱۰۰	۱۱/۷	۷۲/۴	۱۱/۰	۴/۸	دهه ۱۳۵۰	
۱۰۰	۳/۹	۷۱/۳	۱۶/۳	۸/۵	دهه ۱۳۶۰ به بعد	
۱۰۰	۱۴/۲	۷۰/۵	۱۰/۲	۵/۲	غیرشاغل	وضعیت شغلی
۱۰۰	۱۳/۸	۷۰/۰	۱۵/۰	۱/۴	شاغل	
۱۰۰	۱۵/۹	۶۹/۰	۱۲/۲	۳/۰	زیردیپلم	سطح تحصیلات
۱۰۰	۱۰/۴	۷۳/۱	۹/۰	۷/۵	دیپلم و بالاتر	
۱۰۰	۱۱/۳	۷۳/۳	۱۱/۳	۴/۲	غیرخویشاوندی	نوع ازدواج
۱۰۰	۱۸/۲	۶۶/۱	۱۰/۹	۴/۸	خویشاوندی	
۱۰۰	۱۱/۸	۷۲/۲	۱۱/۲	۴/۸	شهر	محل تولد
۱۰۰	۲۱/۷	۶۶/۱	۱۰/۹	۳/۳	روستا	
۱۰۰	۱۸/۵	۶۶/۹	۱۰/۸	۳/۸	<۱۸	سن در اولین ازدواج
۱۰۰	۱۹/۴	۷۳/۸	۵/۸	۱/۰	۱۹-۲۱	
۱۰۰	۷/۰	۶۸/۸	۱۵/۶	۸/۶	۲۲-۲۴	
۱۰۰	۹/۱	۷۷/۳	۱۱/۴	۲/۳	>۲۵	

مدل درختی کارت برای طبقه‌بندی تعداد فرزندان ایده‌آل با معیار جینی و احتمالات پیشین برآوردی (مدل ۱) و احتمالات پیشین برابر (مدل ۲) با وارد نمودن متغیرهای پیش‌بین شامل سن در اولین ازدواج، نوع ازدواج، سطح تحصیلات، وضعیت شغلی، محل تولد و کوهورت مولید با فرض استفاده از معیار انتخاب ضریب جینی و احتمالات پیشین برآوردی و برابر، به ترتیب درخت طبقه‌بندی



نامه انجمن جمعیت شناسی ایران

شکل (۴) که شامل ۵ گره و ۶ برگ و درخت طبقه‌بندی شکل (۵) که شامل ۸ گره و ۹ برگ است، ساخته شد. همان‌گونه که ملاحظه می‌شود کلیه متغیرهای پیش‌بین به جز نوع ازدواج به عنوان گره در مدل درختی کارت (۱) و کلیه متغیرهای پیش‌بین به جز محل تولد به عنوان گره در مدل درختی کارت (۲)، وارد شده‌اند. در واقع مدل برازش شده در میان متغیرهای پیش‌بین وارد شده با توجه به معیارهای برازش مدل، این دو متغیر را از مدل حذف نموده است.

جدول (۴) ماتریس اغتشاش را که براساس آن می‌توان دقت مدل (۱ و ۲) را محاسبه نمود، نشان می‌دهد. این ماتریس مقادیر طبقه‌بندی درست و نادرست براساس مدل برازش یافته را نشان می‌دهد. مقادیر مشخص شده در جدول (۴)، طبقه‌بندی درست یا به عبارتی دقت مدل درختی (۱ و ۲) را بیان می‌کند. در نتیجه دقت مدل (۱) عبارتست از:

جدول ۴. ماتریس اغتشاش مدل (۱ و ۲)

طبقه مشاهده شده تعداد فرزندان ایده‌آل		مدل (۱)				مدل (۲)					
		طبقات پیش‌بینی شده				طبقات پیش‌بینی شده					
تعداد	درصد کل	جمع	۳	۲	۱	۰	جمع	۳	۲	۱	۰
		۰	۰	۱۸	۱	۱۷	۰	۰	۱۸	۰	۱
۱	۰	۴/۴۴	۰	۰/۲۵	۱/۷۳	۲/۴۷	۴/۴۴	۰/۲۵	۴/۲۰	۰	۰
۱	۱	۲۵	۷	۶	۲۶	۶	۴۵	۴	۴۱	۰	۰
۱	۱	۱۱/۱۱	۱/۷۳	۱/۴۸	۶/۴۲	۱/۴۸	۱۱/۱۱	۰/۹۹	۱۰/۱۲	۰	۰
۲	۰	۲۸۵	۶۳	۵۷	۱۴۰	۲۵	۲۸۵	۱۳	۲۷۲	۰	۰
۲	۱	۷۰/۳۷	۱۵/۵۶	۱۴/۰۷	۳۴/۵۷	۶/۱۷	۷۰/۳۷	۳/۲۱	۶۷/۱۶	۰	۰
۳	۰	۵۷	۳۲	۵	۱۸	۲	۵۷	۱۶	۴۱	۰	۰
۳	۱	۱۴/۰۷	۷/۹۰	۱/۲۳	۴/۴۴	۰/۴۹	۱۴/۰۷	۳/۹۵	۱۰/۱۲	۰	۰
۳	۱	۴۰۵	۱۱۲	۶۹	۱۹۱	۴۳	۴۰۵	۳۴	۳۷۱	۰	۰
۳	۱	۱۰۰	۲۵/۱۸	۱۷/۰۴	۴۷/۱۶	۱۰/۶۲	۱۰۰	۸/۴۰	۹/۶۰	۰	۰

$$دقت = \frac{0+0+272+16}{405} = \frac{288}{405} = 0.71 \quad (۴)$$

یعنی ۷۱ درصد موارد به درستی طبقه‌بندی شده‌اند و ۲۹ درصد خطا وجود دارد. همچنین دقت مدل درختی (۲) عبارتست از:



معرفی و کاربرد مدل کارت برای طبقه‌بندی تعداد فرزندان ایده‌آل...

$$(5) \quad \text{دقت} = \frac{10+26+57+32}{405} = \frac{125}{405} = 0.31$$

یعنی تنها ۳۱ درصد موارد به درستی طبقه‌بندی شده‌اند و ۶۹ درصد خطا وجود دارد. با توجه به خطای بالای مدل (۲)، نمی‌توان از آن برای پیش‌بینی تعداد فرزندان ایده‌آل استفاده کرد.

جدول (۵) مخاطره و خطای استاندارد مدل (۲و۱) را برای دو مجموعه داده آموزشی و آزمون نشان می‌دهد. در برازش مدل درختی داده‌ها به دو دسته داده آموزشی و آزمون تقسیم می‌شوند و مدل بر روی هر یک از این داده‌ها برازش می‌یابد. در واقع داده آموزشی برای برازش مدل و داده آزمون برای تأیید اعتبار مدل مورد استفاده قرار می‌گیرند. هر چه مقادیر مخاطره محاسبه شده برای این دو مجموعه داده به هم نزدیک باشد، نشان دهنده اعتبار بالای مدل برازش یافته می‌باشد. با توجه به این‌که در مدل (۱) این مقادیر برای هر دو مجموعه داده تقریباً با هم برابر است، می‌توان گفت که مدل از اعتبار مناسبی برخوردار می‌باشد.

جدول ۵. مخاطره و خطای استاندارد برای دو مجموعه داده آموزشی و آزمون مدل (۲و۱)

مدل (۲)		مدل (۱)		مجموعه آموزشی
خطای استاندارد	مخاطره	خطای استاندارد	مخاطره	
۰/۰۲۳	۰/۴۹۱	۰/۰۲۳	۰/۲۸۹	
۰/۰۲۰	۰/۸۲۶	۰/۰۲۲	۰/۲۸۹	مجموعه آزمون (براساس اعتبارسنجی متقابل k-مرتبه‌ای)

جدول (۶)، احتمالات پیشین برآورد شده که در مدلسازی به‌کار رفته‌اند را نشان می‌دهد. در مدل (۱) احتمالات پیشین براساس مقادیر موجود هر طبقه تعداد فرزندان زنده به دنیا آمده محاسبه شده است. در حالی‌که در مدل (۲) از احتمالات پیشین برابر استفاده شده؛ از آن‌جا که برای تعداد فرزندان زنده به دنیا آمده ۴ طبقه در نظر گرفته شده است مقدار این احتمال ۰/۲۵۰ محاسبه گردیده است.

جدول ۶. احتمالات پیشین برآورد شده مدل (۲و۱)

مدل (۲)	مدل (۱)	تعداد	طبقه
احتمال پیشین	احتمال پیشین		
۰/۲۵۰	۰/۰۴۴	۱۸	۰
۰/۲۵۰	۰/۱۱۱	۴۵	۱
۰/۲۵۰	۰/۷۰۴	۲۸۵	۲
۰/۲۵۰	۰/۱۴۱	۵۷	۳

- با مقایسه دقت مدل‌های (۱) و (۲) نتیجه زیر به دست می‌آید:
- دقت مدل (۱) بالا و دقت مدل (۲) بسیار پایین است. این دقت پایین ممکن است به دلیل فرض برابری احتمالات پیشین باشد که در محاسبات از آن استفاده شده است. همان‌گونه که در جدول (۲) نشان داده شده است، درصد تعداد فرزندان ایده‌آل ۰، ۱، ۲ و ۳ و بیشتر به ترتیب برابر ۴/۴، ۱۱/۱، ۷۰/۴ و ۱۴/۱ می‌باشد. در نتیجه فرض برابری احتمالات پیشین، می‌تواند دقت مدل را کاهش دهد.

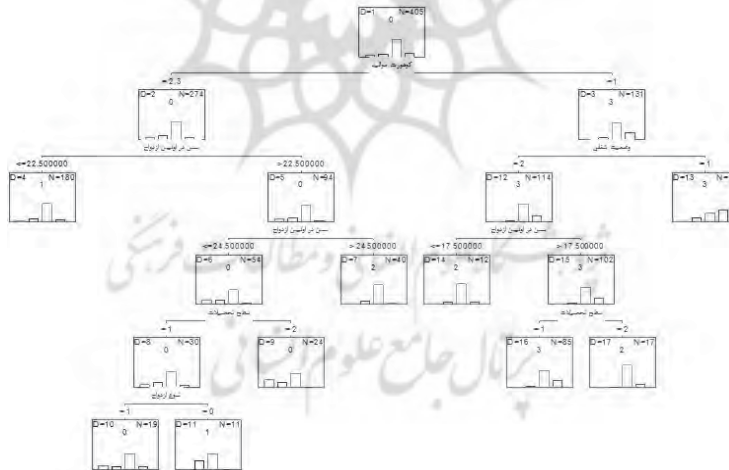
پژوهشگاه علوم انسانی و مطالعات فرهنگی  
پرتال جامع علوم انسانی

معرفی و کاربرد مدل کارت برای طبقه‌بندی تعداد فرزندان ایده‌آل...

شکل ۴. مدل درختی کارت برای طبقه‌بندی تعداد فرزندان ایده‌آل با معیار جینی و احتمالات پیشین برآوردی (مدل ۱)



شکل ۵. مدل درختی کارت برای طبقه‌بندی تعداد فرزندان ایده‌آل با معیار جینی و احتمالات پیشین برابر (مدل ۲)



کوهورت مولید: دهه ۱=۱۳۴۰، دهه ۲=۱۳۵۰، دهه ۳=۱۳۶۰

وضعیت شغلی: شاغل=۱، غیرشاغل=۲

سطح تحصیلات: زیر دیپلم=۱، دیپلم و بالاتر=۲

نوع ازدواج: غیرخویشاوندی=۰، خویشاوندی=۱، محل تولد: شهر=۱، روستا=۲

با توجه به دقت مناسب مدل (۱)، نتایج زیر از درخت تصمیم شکل (۳) قابل استخراج است. لازم به ذکر است که متغیر کوهورت موالید در ریشه این درخت قرار گرفته و مؤثرترین متغیر برای تقسیم‌بندی در نظر گرفته شده است. این درخت براساس این‌که زنان در کوهورت موالید (۱) یا کوهورت موالید (۲ یا ۳) قرار گرفته باشند به دو بخش تقسیم شده است. سمت چپ درخت مربوط به زنان در کوهورت موالید اول است. برای زنان در کوهورت موالید (۱)، درخت بر حسب وضعیت شغلی به دو شاخه شاغل (کد=۱) و غیرشاغل (کد=۲) تقسیم شده است. برای زنان در کوهورت موالید (۱) که شاغلند، درخت بر حسب سطح تحصیلات به دو زیرشاخه زیر دیپلم (کد=۱) و دیپلم و بالاتر (کد=۲) تقسیم شده است. سمت راست این درخت مربوط به زنان در کوهورت موالید دوم و سوم است. برای زنان در کوهورت موالید (۲ و ۳)، درخت بر حسب محل تولد به دو شاخه شهر (کد=۱) و روستا (کد=۲) تقسیم شده است. برای زنان کوهورت موالید دوم و سوم که محل تولدشان روستا بوده است بر حسب این‌که در سن ۲۱/۵ سالگی و کمتر از آن یا بالاتر از ۲۱/۵ سالگی ازدواج کرده باشند، درخت به دو شاخه تقسیم شده است. برگ‌های درخت نشان‌دهنده تعداد فرزندان ایده‌آل است که توسط درخت به صورت زیر طبقه‌بندی شده‌اند:

- ایده‌آل زنان در کوهورت موالید اول (دهه ۱۳۴۰) که شاغلند، ۳ فرزند و بیشتر بوده است. ایده‌آل فرزندان در این کوهورت که غیرشاغل بوده‌اند، بدون تأثیر سطح تحصیلات ۲ فرزند است.
- ایده‌آل زنان در کوهورت موالید دوم (دهه ۱۳۵۰) و سوم (دهه ۱۳۶۰ به بعد) که محل تولدشان شهر است، ۲ فرزند می‌باشد.
- ایده‌آل زنان در کوهورت موالید دوم (دهه ۱۳۵۰) و سوم (دهه ۱۳۶۰ به بعد) که محل تولدشان روستا است، بسته به این‌که سن ازدواجشان ۲۱/۵ سال یا کمتر یا بیشتر از ۲۱/۵ باشد به ترتیب ۳ فرزند و بیشتر یا ۲ فرزند است.

### نتیجه‌گیری

بررسی ادبیات تحقیق در این مقاله نشان داده است که تمایلات و ایده‌آل‌های باروری به عنوان یکی از عوامل تأثیرگذار بر رفتارهای باروری می‌باشند و روند تغییرات باروری بستگی به چگونگی و نوع تغییر این فرایند دارد. ایده‌آل‌های باروری اگر نتوانند سطوح باروری را بطور دقیق پیش‌بینی کنند، می‌توانند در مورد چگونگی تغییرات باروری در آینده اطلاعاتی

ارائه دهند (رازقی نصرآباد و سرایی ۱۳۹۳). با کاهش باروری، آگاهی از ایده‌آل‌های افراد در زمینه تعداد فرزندان و تعیین‌کننده‌های آن، به منظور اتخاذ سیاست‌های جمعیتی متناسب با بستر اقتصادی، اجتماعی و فرهنگی جامعه ایران اخیراً مورد توجه محققان جمعیتی قرار گرفته است. با توجه به اهمیت طبقه‌بندی داده‌ها در علوم اجتماعی و جمعیت‌شناسی با استفاده از روش‌هایی که علاوه بر دقت مطلوب در طبقه‌بندی، دارای تفسیری عینی و ساده نیز باشد، هدف مقاله حاضر استفاده از درخت تصمیم برای مدل‌سازی داده‌های جمعیتی بوده است. بدین منظور با استفاده از تحلیل ثانویه داده‌های طرح «بررسی رفتارهای ازدواج و باروری زنان حداقل یکبار ازدواج کرده، ۱۵-۴۹ ساله در استان سمنان-۱۳۹۱» از مدل کارت در طبقه‌بندی تعداد فرزندان ایده‌آل زنان ۱۵-۴۹ ساله استان سمنان استفاده گردید. هدف اصلی این مقاله معرفی روش درختی کارت، با استفاده از متغیرهای جمعیتی و زمینه‌ای مؤثر بر طبقه‌بندی تعداد ایده‌آل فرزند بوده است. یکی از مسائل مورد علاقه پژوهشگران در مطالعات جمعیتی، پیش‌بینی طبقه تعداد فرزندان ایده‌آل با استفاده از تحلیل ممیزی<sup>۱</sup>،  $K$  نزدیکترین همسایگی<sup>۲</sup>، رگرسیون لجستیک<sup>۳</sup> و رگرسیون پروبیت<sup>۴</sup> است که از جمله روش‌های متداول در این زمینه می‌باشند (انواکره ۲۰۰۷؛ رحمان، احمد و هوکو<sup>۵</sup> ۲۰۰۸). به منظور انجام تحلیل ممیزی خطی پیش‌فرض‌های دارا بودن توزیع نرمال چندمتغیره کلیه متغیرهای پیش‌بین و برابری ماتریس‌های واریانس-کوواریانس در هر طبقه باید برقرار باشند که پیش‌فرض نرمال بودن ضروری است، با این حال این روش بدون توجه به برقراری این پیش‌فرض برای کلیه متغیرها توسط پژوهشگران به‌کار برده می‌شود. از سوی دیگر، این روش تنها برای متغیرهای پیش‌بین پیوسته طراحی شده و متغیرهای پیش‌بین رسته‌ای باید به متغیرهای ظاهری<sup>۶</sup> تبدیل شوند که این عمل، منجر به افزایش تعداد متغیرها می‌گردد. به علاوه کلیه متغیرهایی که در ترکیب خطی وارد می‌شوند باید کامل باشند، به عبارت دیگر مشاهده‌ای که یک مقدار گمشده داشته باشد از تحلیل حذف می‌گردد که منجر به اریبی حاصل از کاهش تعداد نمونه می‌گردد. همچنین اگر

1. Discriminant analysis
2. Kth nearest neighbor
3. Logistic
4. Probit
5. Nwakeze
6. Rahman, Ahmad and Hoque
7. Dummy

متغیرهای پیش‌بین شامل هر دو نوع متغیر پیوسته و دو حالتی باشند، این روش نتایج غیر قابل اعتمادی به دنبال خواهد داشت.

روش‌های رگرسیون لجستیک و پروبیت نیز از دیگر روش‌های پارامتری در مطالعات طبقه‌بندی می‌باشند. برآمد نهایی این روش‌ها، برآورد نسبت مواردی است که در طبقات مختلف متغیر وابسته قرار گرفته‌اند. این روش‌ها همانند تحلیل ممیزی خطی، آزاد-توزیع<sup>۲</sup> نیستند، روشی برای تحلیل نمونه‌ها با مقادیر گمشده در یک متغیر ندارند، تنها برای متغیرهای وابسته رسته‌ای قابل استفاده هستند و همانند کلیه مدل‌های پارامتری، همه متغیرهای به کار رفته در تحلیل توسط پژوهشگر تعیین می‌گردند.

روش‌های که در دهه‌های اخیر با پیشرفت نرم‌افزارهای رایانه‌ای برای طبقه‌بندی داده‌ها به کار می‌رود، درخت تصمیم است که به دلیل انعطاف‌پذیری و ویژگی‌های خاص، به خصوص خروجی آن که یک گراف می‌باشد و تفسیر آن را ساده‌تر می‌نماید، مقبولیت عام یافته است (ابراهیم<sup>۳</sup> و همکاران ۲۰۰۸).

نتایج این مطالعه نشان داد که دو سوم زنان در مطالعه حاضر، دو فرزند را بعنوان بعد ایده‌آل ذکر کردند. به منظور طبقه‌بندی داده‌ها دو مدل درختی به این داده‌ها برازش یافت؛ در مدل اول کلیه متغیرهای پیش‌بین شامل سن در اولین ازدواج، نوع ازدواج، سطح تحصیلی، وضعیت شغلی، محل تولد و کوهورت مولید در فرایند مدلسازی وارد شدند. در مدل (۱) از معیار انتخاب جینی و احتمالات پیشین برآوردی استفاده شد. در مدل استخراج شده کلیه متغیرهای پیش‌بین به جز نوع ازدواج در مدل درختی ظاهر شدند و در طبقه‌بندی تعداد فرزندان ایده‌آل نقش داشتند. در مدل (۲) نیز کلیه متغیرهای پیش‌بین مدل (۱) در فرایند مدلسازی وارد شدند و از معیار انتخاب جینی و احتمالات پیشین برابر استفاده شد. در مدل (۲)، کلیه متغیرهای پیش‌بین به جز محل تولد در مدل درختی ظاهر شدند و در طبقه‌بندی تعداد فرزندان ایده‌آل نقش داشتند. از میان دو مدل اول و دوم، از آنجا که دقت مدل دوم بسیار پایین (حدود ۳۱ درصد) بود، مدل اول به عنوان مدل نهایی انتخاب و تفسیر و استنباط بر پایه آن انجام گرفت. براساس مدل (۱)، نتایج زیر بدست می‌آید:

وضعیت شغلی بر روی تعداد فرزندان ایده‌آل کوهورت مولید اول مؤثر است؛ ایده‌آل شاغلین، ۳ فرزند و بیشتر و غیرشاغلین ۲ فرزند است. تعداد فرزندان ایده‌آل بیشتر زنان

1. Outcome
2. Distribution-free
3. Ibrahim

شاغل نسبت به زنان غیرشاغل ممکن است به دلیل داشتن امنیت اقتصادی آن‌ها باشد که امکان داشتن فرزندان بیشتر را برای آن‌ها نسبت به سایر زنان در آن کوهورت فراهم می‌کند. اما در میان کوهورت‌های دیگر وضعیت شغلی نقشی در تعیین تعداد ایده‌آل زنان نداشت و برای هر دو گروه زنان شاغل و غیر شاغل، تعداد دو فرزند را ایده‌آل بدست آمد. محل تولد و سن در اولین ازدواج بر روی تعداد فرزندان ایده‌آل کوهورت مولید دوم و سوم مؤثر است؛ متولدین شهر بدون تأثیر هیچ متغیر دیگری تعداد ۲ فرزند را ایده‌آل می‌دانند. در حالی که متولدین روستا بر حسب این که در سن کمتر از ۲۱/۵ سال یا بیشتر از ۲۱/۵ سال ازدواج کرده‌اند به ترتیب تعداد ۳ فرزند و بیشتر و ۲ فرزند را ایده‌آل می‌دانند. در تبیین این نتایج می‌توان گفت که عمده زنان که در سنین پایین‌تری ازدواج کرده‌اند به کوهورت اول که متولدین دهه ۱۳۴۰ می‌باشند تعلق دارند. برعکس زنانی که در سنین بالاتر ازدواج کرده‌اند عمدتاً متعلق به کوهورت‌های جوانتر هستند و تعداد فرزند کمتری را ترجیح می‌دهند. این یافته با نتایج سایر مطالعات مطابقت می‌کند؛ براساس مطالعه عباسی شوازی و عسکری ندوشن (۱۳۸۴) زنانی که در سنین بالاتر ازدواج می‌کنند، در مقایسه با سایرین تعداد فرزندان کمتری دارند، به هنگام ازدواج باروری ایده‌آل پایین‌تری داشته‌اند، با احتمال بیشتری قبل از اولین حاملگی خود از روش‌های پیشگیری استفاده کرده‌اند و فرزندان کمتری را برای هر زوج کافی می‌دانند. این نکته که زنان کوهورت مولید دوم و سوم با سن ازدواج بیشتر از ۲۱/۵ سال، همچنان تمایل به داشتن ۲ فرزند دارند قابل تأمل است. این مسئله نشان می‌دهد که کلیه افراد در این دو کوهورت فارغ از هر متغیر دیگری ۲ فرزند را ایده‌آل می‌دانند. با در نظر گرفتن این مسئله برای زنان کوهورت مولید سوم، چنانچه شرایط به گونه‌ای منطبق بر نیازهای این نسل تغییر یابد، آن‌ها نیز مانند کوهورت‌های پیشین ۲ فرزند به دنیا خواهند آورد. این یافته در سایر مطالعات (رازقی نصرآباد و میرزایی، ۱۳۹۱؛ عباسی شوازی و عسکری ندوشن، ۱۳۸۴؛ حسینی و بگی، ۱۳۹۱) نیز نشان داده شده است و بر طبق این مطالعات نیز اکثر زنان دو فرزند را به عنوان تعداد ایده‌آل ذکر می‌کنند. همگرایی در ایده‌آل دو فرزند در سطوح مختلف نسلی می‌تواند ناشی از تغییرات اجتماعی باشد که در سالیان اخیر در جهت تمایل به فرزند کمتر بوده است و از طریق برخی مکانیزم‌ها نظیر افزایش هزینه‌های زندگی، افزایش هزینه‌های تربیت فرزندان، تغییرات نگرشی در جامعه و ...، ایده‌آل‌های فعلی افراد را تحت الشعاع قرار داده باشد. در مطالعه رازقی نصرآباد و سرایی (۱۳۹۳) که بر ارزش فرزندان تمرکز داشت، زنان در هر سه کوهورت به هزینه‌های مالی که فرزندان برای خانواده به وجود می‌آورند، اذعان داشته و



درک نسبتاً یکسانی از هزینه‌های فرزندان داشتند. بر طبق این مطالعه در هر سه کوهورت عنصر تجربه زندگی در این زمینه اهمیت دارد. تجربه نسل جوان و نسل‌های قدیمی‌تر نشان داده که تربیت فرزندان نیاز به صرف هزینه‌های مادی فراوان دارد.

یافته‌های مطالعه نشان می‌دهد که با پیشروی در سطوح کوهورت، نقش متغیرهای جمعیتی و زمینه‌ای به عنوان عوامل تأثیرگذار بر نگرش زنان در خصوص ایده‌آل باروری کاهش می‌یابد. در جامعه مورد نظر به نظر می‌رسد نوعی همگرایی در ایده‌آل باروری به سمت دو فرزند وجود دارد. یافته‌های این مطالعه همسو با سایر مطالعات حکایت از این دارد که در فرهنگ ایرانی تشکیل خانواده و فرزندآوری به عنوان یک هنجار و ارزش محسوب می‌شود و بسیاری از افراد تمایل زیادی برای پدر و مادر شدن دارند و ایده‌آل دو فرزند در بین همه زنان وجود دارد از این رو اتخاذ تدابیری به منظور تسهیل ازدواج و فرزندآوری و فراهم نمودن شرایطی که زنان بتوانند تعداد فرزندان را که می‌خواهند به دنیا بیاورند، یک امر ضروری است.

از دیدگاه آماری، نتایج این مقاله حاکی از آن است که تفسیر یافته‌ها بدون در نظر گرفتن اثر همزمان متغیرهای مورد مطالعه در طرح، ممکن است منجر به گمراهی گردد. به عنوان نمونه بدون در نظر گرفتن مدل درختی و تنها با استفاده از جدول توافقی (۳)، وضعیت شغلی اثر معنی‌داری بر روی تعداد فرزندان ایده‌آل زنان ۴۹-۱۵ ساله ندارد ( $p=0/301$ )، این در حالی است که در تحلیل داده‌ها با استفاده از درخت تصمیم، وضعیت شغلی بر روی طبقه‌بندی تعداد فرزندان ایده‌آل کوهورت مولید اول مؤثر است. بنابراین پیشنهاد می‌شود در تحلیل چندمتغیره داده‌های جمعیتی، به منظور دستیابی به نتایج معتبر، از روش درخت تصمیم که دقت آماری را در طبقه بندی افزایش می‌دهد، استفاده گردد.

#### منابع

- حسینی، حاتم و محمد جلال عباسی شوازی (۱۳۸۸). «تغییرات اندیشه‌ای و تأثیر آن بر رفتارهای ایده‌آل باروری زنان ترک و کرد». پژوهش زنان، دوره ۷، شماره ۲: ۵۵-۸۴.
- حسینی، حاتم و بلال بگی (۱۳۹۱). «فرهنگ و باروری، بررسی تمایلات باروری زنان کرد در شهرستان مهاباد». فصلنامه مطالعات راهبردی، شماره ۵۸: ۱۶۱-۱۲۱.
- رازقی نصرآباد، حجه بی بی و محمد میرزایی (۱۳۹۱). «شکاف موجود بین شمار فرزندان موجود و ایده‌آل در استان‌های منتخب سمنان، کهگیلویه و بویراحمد و هرمزگان». نامه

- انجمن جمعیت‌شناسی ایران، سال هفتم، شماره ۱۳: ۱۷۶-۱۴۹.
- رازقی نصرآباد، حجیه بی بی (۱۳۹۳). «طرح بررسی رفتارهای ازدواج و باروری زنان حداقل یکبار ازدواج کرده، ۱۵-۴۹ ساله در استان سمنان-۱۳۹۱». تهران: مؤسسه مطالعات و مدیریت جامع و تخصصی جمعیت کشور.
- رازقی نصرآباد، حجیه بی بی و حسن سرایی (۱۳۹۳). «تحلیل کوهورتی ارزش فرزند در استان سمنان». زن در توسعه و سیاست، دوره ۱۲ شماره ۱: ۲۲۹-۲۵۰.
- رازقی نصرآباد، حجیه بی بی، محمد جلال عباسی شوازی و میمنت حسینی چاووشی (۱۳۹۳). «پدیدارشناسی زمان تولد اولین فرزند در بین زنان تهرانی». مطالعات راهبردی زنان، سال شانزدهم، شماره ۶۳: ۹۷-۵۷.
- سعادت، مهسا و آرزو باقری (۱۳۹۳). «طرح کاوش داده‌های جمعیتی با استفاده از درخت تصمیم». تهران: مؤسسه مطالعات و مدیریت جامع و تخصصی جمعیت کشور.
- عباسی شوازی، محمد جلال، میمنت حسینی چاووشی، پیترو مک‌دونالد و بهرام دلاور (۱۳۸۳). «تحولات باروری در ایران شواهدی از چهار استان منتخب». تهران: وزارت بهداشت درمان و آموزش پزشکی.
- عباسی شوازی محمد جلال و عباس عسکری ندوشن (۱۳۸۴). «تغییرات خانواده و کاهش باروری در ایران مطالعه موردی استان یزد». نامه علوم اجتماعی، شماره ۲۵: ۳۵-۷۵.
- عباسی شوازی، محمد جلال و میمنت حسینی چاووشی (۱۳۹۲). «طرح تحولات باروری در ایران در چهار دهه اخیر کاربرد و ارزیابی روش فرزندان خود در برآورد باروری با استفاده از داده‌های سرشماری ۱۳۶۵ تا ۱۳۹۰». تهران: پژوهشکده آمار ایران.

Ajzen, I. (2005). *Attitudes, personality, and behavior*. McGraw-Hill Education (UK).

Becker, G. S. (1981). *A treatise on the family*. Enlarged. Harvard University Press, Cambridge Massachusetts, 1(5): 16.

Bernardi, L., A. Klärner, and H. Von der Lippe (2008). "Job insecurity and the timing of parenthood: A comparison between Eastern and Western Germany". *European Journal of Population/Revue Européenne de Démographie*, 24(3): 287-313.

Bhuyar, V. (2014). "Comparative Analysis of Classification Techniques on

- Soil Data to Predict Fertility Rate for Aurangabad District". *International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)*, 3(2): 200-203.
- Bongaarts, J. (2001). "Fertility and reproductive preferences in post-transition societies". *Population and Development Review*, 27: 260-281.
- Breiman, L., J. Friedman, C. J. Stone and R. A. Olshen. (1984). *Classification and regression trees*. CRC press.
- Caldwell, J. C. (1976). «Toward a restatement of demographic transition theory». *Population and Development Review*, 2 (3/4): 321-366.
- Goldstein, J., W. Lutz, and M. R. Testa (2003). "The emergence of sub-replacement family size ideals in Europe." *Population Research and Policy Review*, 22(5-6): 479-496.
- Holfprd T. R. (2002). *Multivarite Methods in Epidemiology*. New York, Oxford University Press.
- Ibrahim N. A., A. Kudus, I. Daud and M.R Abubakr (2008) "Decision tree for competing risks survival probability in breast cancer study". *PWASET* 2008; 28:15-19.
- Jiawei H, and B.Jay (2006). *Data Mining: Concepts and Techniques*, Second Edition, University of Illinois at Urbana-Champaign.
- Leibenstein, H. (1957). *Economic Backwardness and Economic Growth*. John Wiley and Sons, New York.
- Lesthaeghe, R. and G. Moors (2000).» Recent trends in fertility and household formation in the industrialized world». *Review of Population and Social Policy*, 9(1): 121-170.
- Lightbourne, R.E. and A.L. MacDonald. (1982). Family size preferences. World Fertility Survey Comparative Studies Cross National Surveys 14. United Kingdom: Charlesworth Ltd.
- Long, W. J., J. L. Griffith, H. P.Selker, and R. B. D'agostino, (1993). "A comparison of logistic regression to decision-tree induction in a medical

- domain". *Computers and Biomedical Research*, 26(1): 74-97.
- Loris, N., A. Lumini, and C. Manna. (2011). "A Data Mining Approach for Predicting the Pregnancy Rate in Human Assisted Reproduction". *Advanced Computational Intelligence Paradigms in Healthcare*, 5: 97-111.
- Marshall, R. J. (2001). "The use of classification and regression trees in clinical epidemiology". *Journal of Clinical Epidemiology*, 54(6): 603-609.
- McDonald, P. (2002). "Sustaining fertility through public policy: The range of options". *Population*. 57(3):417-445.
- Mello, F. C., L. G. Bastos, S. L. Soares, V. M. Rezende, M. B. Conde, R. E. Chaisson, and G. L. Werneck, (2006). "Predicting smear negative pulmonary tuberculosis with classification trees and logistic regression: a cross-sectional study". *BMC Public Health*, 6(1), 43.
- Miller, W. B. (1994). «Childbearing motivations, desires, and intentions: a theoretical framework». *Genetic, Social, and General Psychology Monographs*, 120: 223-258
- Nwakeze NM. (2007). "The demand for children in Anambra State of Nigeria: A logit analysis". *African Population Studies*, 22(2):175-201.
- Olson, D. L., and D. Delen. (2008). *Advanced data mining techniques*. Springer Science & Business Media.
- Perugini, M., and R. P. Bagozzi. (2001). «The role of desires and anticipated emotions in goal-directed behaviours: Broadening and deepening the theory of planned behavior». *The British Journal of Social Psychology* 40(1): 79-98.
- Podgorelec, V., P. Kokol, B. Stiglic, and I. Rozman (2002). "Decision trees: an overview and their use in medicine". *Journal of Medical Systems*, 26(5): 445-463.
- Preeti, D and L.S. Singh. (2010). "Time varying and unvarying factors affecting ideal and actual family size in North India". Annual Meeting Program Population Association of America. Princeton University, Dallas

- texas, April 15-17.
- Quinlan, J. R. (1986). «Induction of decision trees». *Machine Learning*, 1(1): 81-106.
- Quesnel-Vallée, A., and S. P. Morgan (2003). «Missing the target? Correspondence of fertility intentions and behavior in the US». *Population Research and Policy Review*, 22(5-6): 497-525.
- Rahman M, T. Ahmad, and A. Hoque (2008). «Factors affecting children ever born in slum areas of Rajshahi city corporation, Bangladesh». *Middle East Journal of Nursing* 2(4): 5-10.
- Ramezankhani, A., O. Pournik, J. Shahrabi, D. Khalili, F. Azizi, and F. Hadaegh. (2014). «Applying decision tree for identification of a low risk population for type 2 diabetes. Tehran Lipid and Glucose Study». *Diabetes Research and Clinical Practice*, 105(3): 391-398.
- Ryder, N. B. (1965). «The cohort as a concept in the study of social change». *American Sociological Review* 30(6): 843-861.
- Sahoo, A. J., and Y.Kumar. (2014). «Seminal quality prediction using data mining methods», *Technology and Health Care*, 22(4): 531-545.
- Schaffer, C. (1993). «Selecting a classification method by cross-validation». *Machine Learning*, 13(1): 135-143.
- Schwarzer G., T. Nagata, D. Mattern, R. Schmelzeisen, and M. Schumacher (2003). "Comparison of Fuzzy inference, logistic regression and classification trees (cart): prediction of cervical lymph node metastasis in carcinoma of tongue", *Methods in Medicine*, 42(5): 572-577.
- Siciliano R, and F. Mola (2000). "Multivariate data analysis and modeling through classification and regression tree". *Computational Statistics & Data Analysis*. 32: 285-301.
- Testa, M. R. (2007). Childbearing preferences and family issues in Europe: evidence from the Eurobarometer 2006 survey. Vienna Yearbook of Population Research.

- Thimofeew R. (2004). "Classification and Regression Trees (CART) theory and application". *Master of Science Dissertation, Berlin, Humboldt University*, 2004.
- Van de Kaa, D. J. (2001). "Postmodern fertility preferences: from changing value orientation to new behavior". *Population and Development Review* 27: 290-331.
- Van de Kaa, D. (1997). "Options and Sequences: Europe's demographic patterns". *Journal of the Australian Population Association* 14 (1): 1-30.
- Westoff, C.F. (2010). "Desired Number of Children: 2000-2008", DHS Comparative Reports No. 25, Office of Population Research, Princeton University.
- Wilkinson, L. (1992). "Tree structured data analysis: AID, CHAID and CART", SYSTAT Joint Software Conference.
- Witten, I. H., and E. Frank. (2005). *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann.

