

شبیه‌سازی جمعیتی کوچک‌ناحیه‌ها در سال پایه

محمدتقی معطی^۱

حمیدرضا نواب‌پور^۲

چکیده

در ایران پیش‌بینی‌های جمعیتی به دلیل نبود اطلاعات کمی، اغلب در سطح ناحیه‌های بزرگ همانند استان و کشور انجام می‌شود. اما تولید اطلاعات جمعیتی و ویژگی‌های آن در سطح کوچک‌ناحیه^۳ها (بخش، شهر، ناحیه‌ی شهرداری و ...)، توان پاسخ‌گویی به نیازهای آماری پژوهش‌گران، برنامه‌ریزان و تصمیم‌سازان را بالا می‌برد. با توجه به این‌که به‌طور عموم تمامی اطلاعات جمعیتی در کوچک‌ناحیه‌ها در دسترس نیستند و داشتن اطلاع از توزیع شمارش^۴ و ویژگی‌های افراد و خانوارها (به‌عنوان مثال سن - جنسیتی، نوع خانوار و ...) برای یک کوچک‌ناحیه ضروری است، لذا شبیه‌سازی جمعیت کوچک‌ناحیه‌ها به‌صورت هم‌گذاشتی^۵ با ویژگی‌هایی که مدنظرند، از اهمیت ویژه‌ای برخوردار است. از آنجاکه دسترسی به بسیاری از اطلاعات جمعیتی و ویژگی‌های آن در کوچک‌ناحیه‌ها از طریق سرشماری‌ها امکان‌پذیر نیست، بسیاری از محققان را بر آن داشته است تا برای دستیابی به اطلاعات مورد نیاز خود از کوچک‌ناحیه‌ها به دنبال راه‌کارهایی باشند. در سال‌های اخیر روش‌های مختلفی برای شبیه‌سازی‌های جمعیتی ابداع شده است، از جمله‌ی این روش‌ها می‌توان به روش نمونه‌بنا

۱ دانشجوی دکتری آمار دانشگاه علامه طباطبایی، mt_moeti@yahoo.com

۲ دانشیار آمار دانشگاه علامه طباطبایی، (نویسنده مسئول)، h.navvabpour@src.ac.ir

3 Small Area
4 Frequency Distribution
5 Synthetic

اشاره کرد. این مقاله با ارایه راه‌کارهایی برای اصلاح نقاط ضعف این روش، از آن برای شبیه‌سازی جمعیت کوچک‌ناحیه‌های نمونه با ویژگی‌های متفاوت به‌صورت هم‌گذاشتی برای سال پایه، ۱۳۸۵، استفاده می‌کند. به‌علاوه با توجه به چند معیار عمل‌کرد این روش با اطلاعات سرشماری مقایسه می‌شود.

واژگان کلیدی: نمونه‌مبنا، شبیه‌سازی جمعیتی، کوچک‌ناحیه، جمعیت هم‌گذاشتی.

مقدمه و بیان مسأله

اطلاع از تعداد کل جمعیت و ویژگی‌های اجتماعی و اقتصادی اعضای آن که برخی اطلاعات آن‌ها در سرشماری گردآوری نشده است و چگونگی تحولات آن‌ها در فاصله‌ی سال‌های بین دو سرشماری عمومی نفوس و مسکن جمعیت‌شناسان را به‌سوی ابداع و استفاده از روش‌هایی برای پیش‌بینی جمعیت در ناحیه‌های جغرافیایی به‌ویژه کوچک‌ناحیه‌ها که یافته‌های سرشماری حتی در سال اجرای سرشماری نیز در آن سطوح وجود ندارند و یا منتشر نمی‌شوند، ترغیب کرده است. در این مقاله با توجه به این‌که ارائه‌ی جدول‌های سرشماری در قالب خلاصه‌ی نتیجه‌های تفصیلی سرشماری در سطح استان منتشر می‌شوند و استخراج نتیجه‌ها فقط درباره‌ی برخی ویژگی‌های خاص در سطح شهرستان‌ها انجام می‌شوند و با توجه به نیاز روز افزون به استفاده از داده‌ها در سطح‌های کوچک‌تر مثل بخش و شهر و ...، لازم است روش‌هایی برای تحقق این هدف معرفی شوند. منظور از پیش‌بینی جمعیت در کوچک‌ناحیه‌ها دستیابی به تعداد کل جمعیت به‌تنهایی نیست بلکه هدف رسیدن به تعداد جمعیت همراه با ویژگی‌های اجتماعی و اقتصادی خانوارها و اعضای درون آن‌ها است. به‌عنوان مثال آگاهی از وضع اشتغال و بیکاری گروه‌های سنی جمعیت فعال در یک کوچک‌ناحیه کمک بزرگی به برنامه‌ریزان برای تأمین اعتبارات مورد نیاز برای فراهم آوردن زیر ساخت‌های لازم برای ایجاد موقعیت‌های شغلی تازه در آن ناحیه است. هدف از شبیه‌سازی جمعیت هم‌گذاشتی در سال اجرای سرشماری (سال پایه)، دستیابی به توزیع‌های چندبعدی از ویژگی‌های توأم خانواری و فردی است که اغلب در مجموعه‌ی نتیجه‌های سرشماری‌ها وجود ندارند. بحثی که ممکن است مطرح شود این است، زمانی که اطلاعی مثل توزیع سنی-جنسیتی یا توزیع بعد و نوع خانوارها در سرشماری

اندازه‌گیری شده است چرا در ساخت جمعیت پایه، دوباره این توزیع‌ها به‌شیوه‌ی هم‌گذاشتی شبیه‌سازی می‌شوند. علت اصلی این است که این توزیع‌ها معمولاً در سطح کشور و استان تهیه می‌شوند و ویژگی‌های دیگری از این دست که مورد توجه پژوهشگران و برنامه‌ریزان هستند در موارد محدود با شرایط خاص تنها در صورت درخواست کتبی و رسمی، آن هم با حفظ محرمانگی اطلاعات فردی و سپری کردن سلسله مراتب موجود ممکن است در اختیار درخواست‌کننده قرار بگیرند. چرا که دستیابی به اطلاعاتی فراتر از آنچه که در چارچوب نتیجه‌های سرشماری‌ها به‌صورت استاندارد و از پیش تعیین شده استخراج می‌شوند، فرایندی هزینه‌بر و زمان‌بر است. به‌علاوه ممکن است برخی از کاربران نیاز به اطلاعات ترکیبی (سرشماری و آمارگیری نمونه‌ای) در سطح کوچک‌ناحیه‌ها داشته باشند که مراکز آماری ملی قادر به تهیه و اطلاع‌رسانی آن‌ها نیستند.

از جمله روش‌های شبیه‌سازی جمعیتی می‌توان به روش‌های بازسازی هم‌گذاشتی^۱ (SR) (ویلسون و پونال ۱۹۷۶) اشاره کرد. این روش‌ها عموماً از فن برازش متناسب تکراری^۲ (IPF) (دمینگ و استفان ۱۹۴۰) و نمونه‌ای از جامعه‌ی هدف برای به‌دست آوردن توزیع‌های توام استفاده می‌کنند (بکمن و همکاران ۱۹۹۶، هانگ و ویلیامسون ۲۰۰۲، گوآ و بهات ۲۰۰۷، آرنتره و همکاران ۲۰۰۷، یه و همکاران ۲۰۰۹). بسیاری از روش‌های SR توزیع توام خانوارهای مشاهده‌شده و خانوارهای شبیه‌سازی شده و یا توزیع توام افراد مشاهده‌شده و افراد شبیه‌سازی شده را تطابق می‌دهند ولی این تطبیق به‌صورت هم‌زمان بین خانوارها و افراد انجام نمی‌شود. برای رفع این محدودیت روش‌های مختلفی برای مطابقت ویژگی‌های خانواری و ویژگی‌های فردی ارائه شده است. از جمله‌ی آن‌ها، روش روزآمدسازی متناسب تکراری^۳ (IPU) که توسط یه و همکاران (۲۰۰۹) توسعه یافته، مورد توجه بیش‌تری است.

روش بهینه‌سازی ترکیبی^۴ (CO) از دیگر روش‌های شبیه‌سازی جمعیتی است. این روش یک جمعیت هم‌گذاشتی را به‌وسیله‌ی محدودیت‌های مورد استفاده‌ی کوچک‌ناحیه از جمله ویژگی‌های مورد علاقه و نمونه‌ای از جامعه‌ی هدف برای هر کوچک‌ناحیه شبیه‌سازی می‌کند

1 Synthetic Reconstruction

2 Iterative Proportional Fitting

3 Iterative Proportional Updating

4 Combinatorial Optimization

(ووآس و ویلیامسون ۲۰۰۱ و هانگ و ویلیامسون ۲۰۰۲). روش بازسازی هم‌گذاشتی نمونه‌آزاد (گارجیلو و همکاران ۲۰۱۰، بارتله‌می و توانیت ۲۰۱۲، لنورماند و دی‌فیوانت ۲۰۱۳) از جمله روش‌هایی است که برای شبیه‌سازی جمعیتی با ویژگی‌های خاص به‌کار می‌رود. در این مقاله روش روزآمدسازی متناسب تکراری نمونه‌مبنا (یه و همکاران ۲۰۰۹) که از این پس آن را به اختصار روش نمونه‌مبنا می‌نامیم، برای شبیه‌سازی جمعیتی با ویژگی‌های خاص معرفی می‌کنیم. در ادامه به بررسی بیش‌تر مشکلاتی که در روش شبیه‌سازی ممکن است ایجاد شود می‌پردازیم. در پایان کیفیت اطلاعات حاصل از به‌کارگیری شبیه‌سازی را با انتخاب چهار کوچک‌ناحیه‌ی نمونه‌ای با موقعیت‌های جغرافیایی مختلف، به‌صورت کابردی مورد ارزیابی قرار می‌دهیم.

مفهوم شناسی

فرض کنید در یک کوچک‌ناحیه‌ی مورد مطالعه، n فرد در مجموعه‌ی X و m خانوار در مجموعه‌ی Y وجود دارند. می‌خواهیم با استفاده از برخی ویژگی‌های منتخب از این افراد و خانوارها در دو مجموعه‌ی X و Y ، مجموعه‌ی جدیدی از خانوارهای تکمیل‌شده به‌دست آوریم، این مجموعه را با P نشان می‌دهیم. تعریف می‌کنیم،

$$X = \{x_1, x_2, \dots, x_n\}$$

$$Y = \{y_1, y_2, \dots, y_m\}$$

اکنون همه‌ی ویژگی‌های منتخب افراد را در مجموعه‌ی q عضوی T و همه‌ی ویژگی‌های منتخب خانوارها را در مجموعه‌ی p عضوی U در نظر می‌گیریم. ویژگی هر فرد را با t_x ،

و ویژگی هر خانوار را با $u_y = 1, 2, \dots, p$ نشان می‌دهیم، لذا

$$T = \{t_1, t_2, \dots, t_q\}$$

$$U = \{u_1, u_2, \dots, u_p\}$$

اکنون تعداد افراد مربوط به هر ویژگی فردی را با مجموعه‌ی n_T و تعداد خانوارهای مربوط

به هر ویژگی خانواری را با مجموعه‌ی n_U به‌صورت زیر تعریف می‌کنیم:

$$n_T = \{n_{t_1}, n_{t_2}, \dots, n_{t_q}\}$$

$$n_U = \{n_{u_1}, n_{u_2}, \dots, n_{u_p}\}.$$

۱. ویژگی‌های فردی

ویژگی‌های زیر از جمله ویژگی‌های مهمی هستند که از آن‌ها در این مقاله برای افراد استفاده می‌شود:

(۱) گروه سنی: [۰-۵]، [۵-۱۰]، ...، [۸۵>]، و

(۲) نسبت خانوادگی: سرپرست خانوار تک‌نفره، سرپرست خانوار تک‌والدی، سرپرست خانوار زوجی بدون فرزند، سرپرست خانوار هسته‌ای (خانوارهای زوجی با فرزند)، سرپرست سایر خانوارها (به خانوارهایی به جز خانوارهای مذکور اطلاق می‌شود مانند خانوارهای فرزند سرپرست)، فرزند خانوار، شریک سرپرست خانوار، افراد سایر خانوارها.

۲. ویژگی‌های خانواری

ویژگی‌های زیر از جمله ویژگی‌های مهمی هستند که از آن‌ها در این مقاله برای خانوار استفاده می‌شود:

(۱) انواع خانوار: خانوار تک‌نفره، خانوار زوجی بدون فرزند، خانوار تک‌والد، خانوار هسته‌ای، سایر خانوارها، و

(۲) بعد خانوار: ۱، ۲، ۳، ۴، ۵، ۶، ۷ و بیش‌تر.

۳. جامعه‌های مرجع

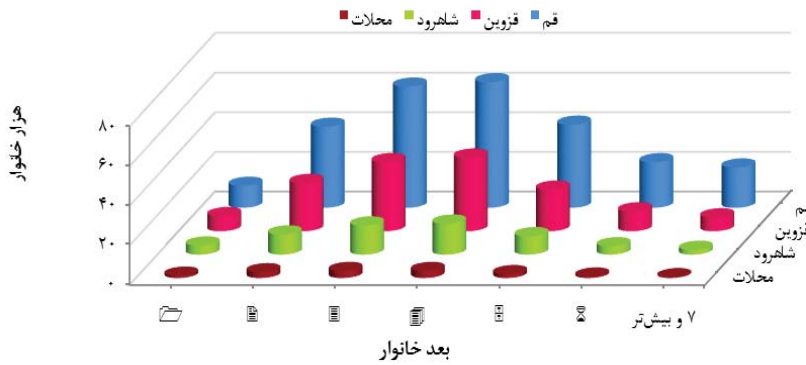
چهار شهرستان قم، قزوین، شاهرود و محلات را به‌عنوان چهار جامعه‌ی مرجع در نظر گرفته‌ایم. این شهرستان‌ها به دلیل داشتن جمعیت‌های متفاوت (جمعیت شهرستان‌های منتخب در سرشماری سال ۱۳۸۵ از حدود ۵۰۰۰۰ نفر تا ۱۰۰۰۰۰۰ نفر بوده است)، نوع خانوارهای مختلف (تک‌نفره، زوجی بدون فرزند، تک‌والد، هسته‌ای)، بعد خانوار متفاوت، موقعیت‌های جغرافیایی مختلف (از نظر مکان شهرستان و تعداد دهستان‌ها و روستاهای مختلف)، از همه مهم‌تر

شهرستان‌هایی که در طول سال‌های اخیر کمتر دچار تغییرات جغرافیایی شده‌اند، ملاک انتخاب بوده و در سرشماری عمومی نفوس و مسکن سال ۱۳۸۵ به ترتیب دارای ۲۶۱۵۸۶، ۱۴۱۶۷۹ و ۶۰۷۹۲ و ۱۳۹۳۳ خانوار و به ترتیب دارای جمعیتی بالغ بر ۱،۰۳۴۰۷۹، ۵۲۷۰۵۰ و ۲۲۰۷۵۹ و ۴۷۷۵۵ نفر بوده‌اند. برای شبیه‌سازی جمعیت این شهرستان‌ها با ویژگی‌های توام فردی و خانواری به توزیع‌های شمارشی یا توزیع‌های احتمال کوچک‌ناحیه‌ی مورد مطالعه نیاز داریم. توزیع‌های مورد نیاز به شرح زیر می‌باشند.

جدول ۱: توزیع‌های شمارشی گروه سنی افراد، سرپرست خانوارها، سرپرست انواع خانوارها، و فرزندان در سرشماری ۱۳۸۵ در چهار شهرستان منتخب

| گروه سنی | ت | ۵-۹ | ۱۰-۱۴ | ۱۵-۱۹ | ۲۰-۲۴ | ۲۵-۲۹ | ۳۰-۳۴ | ۳۵-۳۹ | ۴۰-۴۴ | ۴۵-۴۹ | ۵۰-۵۴ | ۵۵-۵۹ | ۶۰-۶۴ | ۶۵-۶۹ | ۷۰-۷۴ | ۷۵-۷۹ | ۸۰-۸۴ | ۸۵+ | مجموع | | |
|--------------------------|----------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|------|------|
| | | | | | | | | | | | | | | | | | | | مرد | زنان | |
| افراد | مجموع | ۳۳۷۷ | ۳۵۱۷ | ۳۹۷۹ | ۵۰۴۰ | ۵۳۷۸ | ۴۳۵۰ | ۳۳۶۹ | ۳۴۸۶ | ۳۳۰۰ | ۲۷۶۸ | ۲۲۰۱ | ۱۳۴۸ | ۱۲۰۴ | ۱۲۷۱ | ۷۰۳ | ۵۲۰ | ۳۲۰ | | | |
| | شماره‌رو | ۱۵۹۸۶ | ۱۵۹۸۲ | ۱۵۹۳۸ | ۱۶۱۲۱ | ۱۶۱۱۱ | ۱۶۱۴۰ | ۱۶۱۹۱ | ۱۶۲۰۳ | ۱۶۲۱۰ | ۱۶۲۰۳ | ۱۶۲۱۰ | ۱۶۲۱۰ | ۱۶۲۱۰ | ۱۶۲۱۰ | ۱۶۲۱۰ | ۱۶۲۱۰ | ۱۶۲۱۰ | | | |
| | فروین | ۳۳۸۷۳ | ۳۳۶۵۵ | ۳۶۷۵۲ | ۴۶۹۰۷ | ۴۶۵۱۶ | ۳۵۵۱۶ | ۴۵۱۰۶ | ۳۹۱۶۸ | ۳۳۱۹۶ | ۲۶۹۷۲ | ۲۱۲۹۱ | ۱۵۵۲۳ | ۱۳۶۷۲ | ۱۳۶۷۲ | ۴۱۳۷ | ۵۰۴۴ | ۳۳۲۹ | ۱۶۶۹ | ۱۶۶۹ | ۱۶۶۹ |
| سرپرست خانوار | مجموع | ۰ | ۰ | ۳ | ۳۵ | ۴۱۷ | ۱۱۹۱ | ۱۶۳۱ | ۱۶۴۳ | ۱۷۵۹ | ۱۵۷۷ | ۱۱۶۱ | ۷۷۵ | ۷۶۵ | ۸۶۰ | ۵۰۹ | ۴۰۷ | ۲۶۰ | | | |
| | شماره‌رو | ۰ | ۰ | ۰ | ۷ | ۲۴۱ | ۲۳۵۲ | ۲۶۸۶ | ۲۶۰۳ | ۲۶۳۷ | ۲۶۰۳ | ۲۶۰۳ | ۲۶۰۳ | ۲۶۰۳ | ۲۶۰۳ | ۲۶۰۳ | ۲۶۰۳ | ۲۶۰۳ | | | |
| | فروین | ۰ | ۰ | ۱۱ | ۳۳۷ | ۵۰۳۲ | ۱۶۶۹۱ | ۲۰۶۲۷ | ۱۹۴۶۷ | ۱۷۸۱۵ | ۱۴۵۲۴ | ۱۱۶۷۸ | ۶۷۹۹ | ۶۷۹۹ | ۶۷۹۹ | ۶۷۹۹ | ۶۷۹۹ | ۶۷۹۹ | ۶۷۹۹ | ۶۷۹۹ | ۶۷۹۹ |
| سرپرست خانوارهای تک نفره | مجموع | ۰ | ۰ | ۱۲ | ۵۳ | ۲۰۲ | ۲۰۲ | ۱۱۷ | ۱۰۸ | ۷۹ | ۳۴ | ۲۷ | ۱۳ | ۱۱ | ۱۱ | ۱۱ | ۱۱ | ۱۱ | | | |
| | شماره‌رو | ۰ | ۰ | ۰ | ۲ | ۱۲۲ | ۴۵۸ | ۲۰۲ | ۱۱۷ | ۱۰۸ | ۷۹ | ۳۴ | ۲۷ | ۱۳ | ۱۱ | ۱۱ | ۱۱ | ۱۱ | | | |
| | فروین | ۰ | ۰ | ۷ | ۲۰۲ | ۱۰۲ | ۴۳۶ | ۲۰۲ | ۱۱۷ | ۱۰۸ | ۷۹ | ۳۴ | ۲۷ | ۱۳ | ۱۱ | ۱۱ | ۱۱ | ۱۱ | | | |
| سرپرست خانوارهای تک والد | مجموع | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | | | |
| | شماره‌رو | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | | | |
| | فروین | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | | | |
| سرپرست خانوارهای زوجی | مجموع | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | | | |
| | شماره‌رو | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | | | |
| | فروین | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | | | |
| فرزندان | مجموع | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | | | |
| | شماره‌رو | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | | | |
| | فروین | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | | | |

توزیع‌های جدول ۱ و شکل ۱ را می‌توان به‌طور دقیق و تکمیل شده از مرکز آمار ایران (با استفاده از اطلاعات سرشماری) استخراج کرد. در ادامه به توضیح بیشتر در مورد شبیه‌سازی جمعیت هم‌گذاشتی از اطلاعات افراد درون خانوار با استفاده از روش نمونه‌مبنا می‌پردازیم.



شکل ۱: نمودار توزیع شمارش بعد خانوار در سرشماری ۱۳۸۵ در چهار شهرستان منتخب

روش‌های شبیه‌سازی جمعیت

روش‌های مختلفی برای شبیه‌سازی‌های جمعیتی وجود دارد که از جمله‌ی آن‌ها می‌توان به روش نمونه‌مبنا اشاره کرد.

روش نمونه‌مبنا

این روش ابتدا با درصدی (نمونه‌ای) از کل خانوارهای شمارش شده در یک سرشماری، P ، که آن را با P_j نشان می‌دهیم، شروع می‌شود. برای بازسازی مجدد P از وزن‌های وابسته به خانوارهای نمونه‌ای یا افراد درون این خانوارها (W_j) به منظور تطبیق تعداد کل افراد X و خانوارهای Y استفاده می‌شود. در این روش برای هر ویژگی خانوار یا فرد درون خانوار j ، هدف، تطبیق دادن مجموع موزون (WS_j) با محدودیت‌های برآورد شده e_j به وسیله‌ی تنظیم کردن وزن‌ها است. وزن W_j ، وزن خانوار یا فرد درون خانوار j ام از نمونه‌ی موزون و e_j برآوردی از تعداد کل ویژگی j ام خانوار یا فرد درون خانوار در P می‌باشد

$$\hat{e}_j = \begin{cases} n_{u_j} & j = 1, 2, \dots, p \\ n_{t_j} & j = p + 1, p + 2, \dots, p + q \end{cases}$$

که در آن $\{n_{tk}\}_{1 \leq k \leq q}$ تعداد افراد مربوط به هر ویژگی فردی و $\{n_{ui}\}_{1 \leq i \leq p}$ تعداد خانوارهای مربوط به هر ویژگی خانواری را نشان می‌دهند. این برآورد به‌طور جداگانه برای هر نوع فرد و خانوار می‌تواند با استفاده از نتیجه‌های سرشماری انجام شود.

در این روش ابتدا ویژگی‌های افراد و خانوارهای مجموعه‌ی P_g در یک ماتریس که آن را ماتریس D نامیده، و در جدول ۲ نمایش داده شده است، خلاصه می‌کنیم. فرض کنیم تعداد خانوارهایی که به‌عنوان نمونه از کل جامعه‌ی P در نظر می‌گیریم برابر عدد معلوم N باشد. پس تعداد اعضای مجموعه‌ی P_g برابر N است، یعنی $|P_g| = N$ ، که نماد I نشانگر کاردینال (۰) است. ویژگی‌های خانوار (u_p) ، از دو دسته ویژگی نوع خانوار در ۵ حالت و بعد خانوار در ۷ حالت تشکیل شده‌اند. به‌نظر می‌رسد که تمام ویژگی‌های خانوار برابر $7 \times 5 = 35$ حالت است، که این طور نیست زیرا مشخصه‌هایی مانند خانوارهای تک‌نفره با بعد ۲ وجود ندارد. نمادهای زیر ۱۹ ویژگی را که می‌توان برای خانوار در نظر گرفت مشخص می‌کند، لذا $p = 19$.

- (u_1) خانوارهای تک‌نفره،
- $(u_7 - u_6)$ خانوارهای تک‌والد دو نفره تا خانوارهای تک‌والد بیش از ۶ نفر،
- (u_8) خانوارهای زوجی بدون فرزند،
- $(u_9 - u_{13})$ خانوارهای هسته‌ای سه نفره تا خانوارهای هسته‌ای بیش از ۶ نفر، و
- $(u_{14} - u_{19})$ سایر خانوارهای دو نفره تا سایر خانوارهای بیش از ۶ نفر.

ویژگی‌های افراد (t_q) ، نیز از دو دسته ویژگی گروه سنی در ۱۸ حالت و نسبت خانوادگی در ۸ حالت تشکیل شده‌اند، نمادهای زیر ۱۲۵ ویژگی را که می‌توان برای افراد در نظر گرفت مشخص می‌کند، پس داریم $q = 125$.

- $(t_1 - t_{15})$ رده سنی ۴ تا رده سنی ۱۸ سرپرست خانوار تک‌نفره،
- $(t_{16} - t_{21})$ رده سنی ۳ تا رده سنی ۱۸ سرپرست خانوار تک‌والد،

- $(t_{۲۲} - t_{۲۶})$ رده سنی ۴ تا رده سنی ۱۸ سرپرست خانوار زوجی بدون فرزند،
- $(t_{۲۷} - t_{۳۱})$ رده سنی ۴ تا رده سنی ۱۸ سرپرست خانوار هسته‌ای،
- $(t_{۳۲} - t_{۳۷})$ رده سنی ۳ تا رده سنی ۱۸ سرپرست سایر خانوارها،
- $(t_{۳۸} - t_{۴۱})$ رده سنی ۱ تا رده سنی ۱۴ فرزندان خانوار،
- $(t_{۴۲} - t_{۴۷})$ رده سنی ۳ تا رده سنی ۱۸ شریک سرپرست خانوار،
- $(t_{۴۸} - t_{۴۹})$ رده سنی ۱ تا رده سنی ۱۸ دیگر افراد سایر خانوارها، و

همچنین در این ماتریس، بردار E برآورد محدودیت‌های e_j و δ_b درصد قدر مطلق انحراف برآورد محدودیت‌های e_j از مجموع موزون اولیه، رابطه (۱) و δ_a درصد انحراف برآورد محدودیت‌های e_j از مجموع موزون نهایی را نشان می‌دهند.

$$(۱) \quad \delta_b = \frac{|\bar{e}_j - ws_j|}{\bar{e}_j}$$

یکی از مشکلات روش نمونه‌مبنا وجود خانه‌ی صفر است. یعنی ممکن است برآورد محدودیت‌های جامعه که در بردار E ماتریس D قرار دارند، یک عدد غیر صفر باشد در صورتی‌که برخی از اعضای بردار ws که مجموع موزون محدودیت‌های نمونه است، صفر باشند. به‌عنوان مثال در کاربرد این مقاله در چند مورد سلول صفر وجود دارد که از جمله‌ی آنها می‌توان به محدودیت‌های $t_1, t_۳, t_۶, t_۷, t_۱۰, t_۱۱, t_۱۲, t_۱۳, t_۱۴, t_۱۵, t_۱۸$ اشاره کرد. این محدودیت‌ها ۹ محدودیت از ۱۴۴ محدودیت در شهرستان محلات هستند که در نمونه صفرند ولی در جامعه صفر نیستند.

جدول ۲: جدول طراحی شده برای گردآوری داده‌ها با استفاده از روش نمونه‌مبنا

| شماره خانوار | W | ${}_1u$ | ${}_2u$ | ... | u_p | ${}_1t$ | ${}_2t$ | ... | t_q | ${}_1W$ | ${}_2W$ | ... | W_{p+q} |
|--------------|---|---------|---------|-----|-------|--------------|--------------|-----|-----------|---------|---------|-----|-----------|
| ۱ | | | | | | | | | | | | | |
| ۲ | | | | | | | | | | | | | |
| ... | | | | | | | | | | | | | |
| N | | | | | | | | | | | | | |
| WS | | | | | | | | | | | | | |
| E | - | ${}_1e$ | ${}_2e$ | ... | e_p | ${}_1e_{p+}$ | ${}_2e_{p+}$ | ... | e_{p+q} | | | | |
| δ_b | | | | | | | | | | | | | |
| ${}_1WS$ | | | | | | | | | | | | | |
| ${}_2WS$ | | | | | | | | | | | | | |
| ... | | | | | | | | | | | | | |
| WS_{p+q} | | | | | | | | | | | | | |
| δ_a | | | | | | | | | | | | | |

یکی دیگر از مشکلات روش نمونه‌مبنا زمانی است که در خروجی، برخی از وزن‌های نهایی (ستون W_{p+q} در ماتریس D در آخرین تکرار) کوچک‌تر از یک می‌شوند. این بدین معنی است که فراوانی این نوع خانوارها در جامعه کم‌تر از یک است که در عمل غیر ممکن است. در مثال این مقاله این مشکل فقط در شهرستان محلات دیده شد که تعدادی از خانوارهای نمونه‌ای وزنی کوچک‌تر از یک داشته‌اند.

معیارهای ارزیابی عمل کرد

برای آزمون عمل کرد روش نمونه‌مبنا از روی جامعه‌های مرجع (توزیع‌های معرفی شده در جدول ۱ و شکل ۱) با استفاده از الگوریتم معرفی شده در این روش، افراد و خانوارهای این جامعه‌ها را بازسازی کرده و برای هر یک از شهرستان‌های منتخب، این الگوریتم را با استفاده از نرم افزار R اجرا کردیم. در روش نمونه‌مبنا این الگوریتم برای هر یک از چهار شهرستان منتخب ۱۰۰ بار تکرار شده و نتیجه‌ی آخرین تکرار ثبت شده است. در این روش با توجه به اینکه نمونه انتخابی ثابت است در هر تکرار وزن‌های نهایی هر خانوار آنقدر تصحیح می‌شوند تا

بهترین نتیجه حاصل شود (ملاک انتخاب بهترین نتیجه درصد انحراف برآورد محدودیت‌های e_j از مجموع موزون نهایی است که در رابطه (۱) بیان شده است). نتیجه‌های روش نمونه‌مبنا به تفکیک شهرستان‌های منتخب در جدول‌های ۳ و ۴ آمده‌اند.

جدول ۳: توزیع شمارش شبیه‌سازی شده گروه سنی افراد، سرپرست خانوار، سرپرست انواع خانوار، و

فرزندان در سال ۱۳۸۵ در چهار شهرستان منتخب به‌وسیله‌ی روش نمونه‌مبنا

| بیشتر از ۸۵ | ۸۰-۸۴ | ۷۵-۷۹ | ۷۰-۷۴ | ۶۵-۶۹ | ۶۰-۶۴ | ۵۵-۵۹ | ۵۰-۵۴ | ۴۵-۴۹ | ۴۰-۴۴ | ۳۵-۳۹ | ۳۰-۳۴ | ۲۵-۲۹ | ۲۰-۲۴ | ۱۵-۱۹ | ۱۰-۱۴ | ۵-۹ | ۰-۴ | گروه سنی | |
|-------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|----------|------------------|
| ۱۰۶ | ۱۵۲ | ۱۳۲ | ۲۱۷ | ۱۱۵ | ۸۱ | ۵۵ | ۵۳ | ۳۵ | ۲۷ | ۲۳ | ۲۸ | ۳۹ | ۵۳ | ۱۴ | ۰ | ۰ | ۰ | محلات | سرپرست |
| ۲۳۸ | ۴۷۷ | ۵۸۳ | ۷۶۴ | ۴۸۰ | ۴۱۹ | ۲۱۷ | ۱۶۲ | ۱۳۱ | ۷۹ | ۱۰۸ | ۱۱۷ | ۲۰۲ | ۴۵۸ | ۱۲۲ | ۲ | ۰ | ۰ | شاهرود | خانوارهای تک‌فرد |
| ۳۸۱ | ۷۳۴ | ۹۶۶ | ۱۳۲۲ | ۹۵۹ | ۶۱۶ | ۳۶۶ | ۲۸۴ | ۲۳۱ | ۱۸۵ | ۲۳۶ | ۲۵۱ | ۳۸۹ | ۴۳۶ | ۱۰۲ | ۷ | ۰ | ۰ | قزوین | تک‌فرد |
| ۶۵۸ | ۱۲۶۶ | ۱۳۴۷ | ۲۱۲۰ | ۱۲۷۵ | ۱۰۶۷ | ۵۶۸ | ۵۵۳ | ۳۶۶ | ۲۸۳ | ۳۶۳ | ۳۶۱ | ۴۷۷ | ۴۱۸ | ۹۲ | ۵ | ۰ | ۰ | قم | تک‌فرد |
| ۱۵ | ۱۹ | ۳۱ | ۵۱ | ۶۵ | ۶۹ | ۷۸ | ۱۰۴ | ۹۸ | ۱۲۱ | ۷۲ | ۴۸ | ۳۵ | ۸ | ۱ | ۰ | ۰ | ۰ | محلات | سرپرست |
| ۲۹ | ۷۱ | ۱۳۳ | ۱۹۰ | ۲۵۶ | ۴۴۰ | ۴۳۳ | ۵۰۴ | ۴۲۴ | ۳۷۸ | ۲۸۲ | ۱۷۲ | ۱۰۹ | ۴۳ | ۵ | ۰ | ۰ | ۰ | شاهرود | خانوارهای تک‌فرد |
| ۶۹ | ۱۱۲ | ۲۱۳ | ۴۱۰ | ۴۹۹ | ۶۹۲ | ۸۷۵ | ۸۹۶ | ۸۲۷ | ۶۸۹ | ۴۸۰ | ۲۶۶ | ۹۹ | ۳۰ | ۰ | ۰ | ۰ | ۰ | قزوین | ولد |
| ۱۱۶ | ۲۰۰ | ۳۲۹ | ۶۴۸ | ۸۲۴ | ۱۱۹۴ | ۱۴۲۶ | ۱۸۹۹ | ۱۸۴۰ | ۱۴۷۸ | ۱۱۷۳ | ۶۵۳ | ۲۴۷ | ۷۵ | ۴ | ۰ | ۰ | ۰ | قم | تک‌فرد |
| ۱۳۶ | ۲۲۶ | ۳۲۱ | ۵۴۱ | ۵۶۶ | ۵۷۳ | ۵۹۶ | ۶۹۹ | ۱۳۲۷ | ۱۵۰۳ | ۱۴۹۸ | ۱۴۸۴ | ۱۰۴۷ | ۲۹۶ | ۸ | ۰ | ۰ | ۰ | محلات | سرپرست |
| ۲۶۰ | ۷۲۱ | ۱۰۶۴ | ۱۶۷۷ | ۱۹۲۲ | ۲۴۳۲ | ۲۹۰۷ | ۳۱۱۱ | ۳۴۹۹ | ۳۵۴۱ | ۳۶۴۴ | ۳۸۶۱ | ۴۲۵۰ | ۴۵۷۳ | ۴۵ | ۰ | ۰ | ۰ | شاهرود | خانوارهای تک‌فرد |
| ۴۵۹ | ۱۱۹۰ | ۲۰۴۷ | ۴۰۰۳ | ۴۱۶۱ | ۴۵۷۰ | ۴۶۷۸ | ۴۹۸۲ | ۵۲۰۱ | ۵۵۲۱ | ۵۷۵۱ | ۶۰۷۸ | ۶۴۱۶ | ۶۸۱۳ | ۹۴ | ۱ | ۰ | ۰ | قزوین | زوجین |
| ۸۹۹ | ۲۱۵۲ | ۳۱۸۱ | ۶۰۵۷ | ۶۸۰۷ | ۷۹۰۶ | ۸۹۱۹ | ۱۰۰۵۹ | ۱۰۳۰۶ | ۱۰۷۹۰ | ۱۱۲۰۳ | ۱۱۶۹۳ | ۱۲۲۰۷ | ۱۲۷۰۰ | ۱۳۲۰۰ | ۱۳۷۰۰ | ۱۴۲۰۰ | ۱۴۷۰۰ | قم | زوجین |
| ۱۰ | ۱۶ | ۳۹ | ۵۱ | ۵۰ | ۵۲ | ۶۰ | ۹۲ | ۱۳۱ | ۱۱۸ | ۱۱۴ | ۵۷ | ۶۵ | ۵۱ | ۲۵ | ۶ | ۰ | ۰ | محلات | فرزندان |
| ۴۵ | ۹۱ | ۱۵۴ | ۲۷۱ | ۳۳۹ | ۴۰۴ | ۳۹۹ | ۵۴۸ | ۶۵۴ | ۵۴۲ | ۵۰۴ | ۴۵۲ | ۴۴۷ | ۱۷۰ | ۶۹ | ۵ | ۰ | ۰ | شاهرود | خانوارهای تک‌فرد |
| ۶۹ | ۱۳۳ | ۲۱۹ | ۵۱۳ | ۵۶۶ | ۸۲۰ | ۹۰۹ | ۱۱۶۲ | ۱۳۴۵ | ۱۳۳۷ | ۱۲۱۳ | ۱۱۳۵ | ۱۰۷۰ | ۶۸۱ | ۱۴۰ | ۳ | ۰ | ۰ | قزوین | زوجین |
| ۱۳۱ | ۲۳۴ | ۵۱۴ | ۱۰۰۳۳ | ۱۱۶۷ | ۱۳۲۱ | ۱۶۷۱ | ۲۵۵۸ | ۲۸۲۴ | ۲۷۲۷ | ۲۶۳۲ | ۲۶۰۷ | ۲۴۷۷ | ۱۴۴۳ | ۲۹۲ | ۲۱ | ۰ | ۰ | قم | ولد و زوجین |

جدول ۴: توزیع شمارش شبیه‌سازی شده نوع خانوارها در سال ۱۳۸۵ به‌وسیله‌ی روش نمونه‌مبنا

| شهرستان | نوع خانوار | روش | تک‌فرد | تک‌والد | زوجی | سایر خانوارها |
|---------|------------|-------|--------|---------|-------|---------------|
| محلات | سرشماری ۸۵ | ۱۱۲۹ | ۷۹۴ | ۱۱۱۲۲ | ۸۸۸ | |
| | نمونه مبنا | ۱۱۲۹ | ۸۱۴ | ۱۱۰۹۱ | ۹۳۷ | |
| شاهرود | سرشماری ۸۵ | ۴۵۴۹ | ۳۴۶۷ | ۴۷۵۷۷ | ۵۱۹۹ | |
| | نمونه مبنا | ۴۵۴۹ | ۳۴۷۰ | ۴۷۵۸۰ | ۵۰۹۴ | |
| قزوین | سرشماری ۸۵ | ۷۵۸۵ | ۶۱۸۲ | ۱۱۶۳۶۱ | ۱۱۵۵۱ | |
| | نمونه مبنا | ۷۵۸۵ | ۶۱۶۶ | ۱۱۵۸۹۴ | ۱۱۵۱۶ | |
| قم | سرشماری ۸۵ | ۱۱۰۷۹ | ۱۲۰۸۹ | ۲۱۴۹۶۶ | ۲۳۴۵۲ | |
| | نمونه مبنا | ۱۱۰۷۹ | ۱۲۲۰۶ | ۲۱۶۷۰۷ | ۲۳۷۳۰ | |

برای ارزیابی عملکرد این روش از معیاری که شمارش‌های مشاهده‌شده (شمارش سرشماری) و مورد انتظار (شمارش شبیه‌سازی شده) خانوارها و افراد مقایسه می‌کند و توسط لنورماند و دی‌فیوانت (۲۰۱۳) معرفی شده است، استفاده می‌شود.

$$(۲) \quad PGP = 1 - \frac{1}{\gamma} \times \frac{\sum_{k=1}^p |o_k - E_k|}{\sum_{k=1}^p o_k}$$

معیار PGP همواره در بازه ۰/۵ تا ۱ تغییر می‌کند. نزدیک شدن این معیار به عدد یک نشانگر این است که شمارش‌های مشاهده‌شده (شمارش سرشماری) و مورد انتظار (شمارش شبیه‌سازی شده) هم‌توزیع‌اند.

که در آن O_k و E_k به ترتیب شمارش‌های مشاهده‌شده و مورد انتظار خانوار یا افراد درون خانوار k را نشان می‌دهند.

در آماره‌ی آزمونی دیگر، از فاصله‌ی خی‌دو استفاده می‌شود (لنورماند و دی‌فیوانت ۲۰۱۳). این فاصله از یک توزیع خی‌دو با $p - 1$ درجه‌ی آزادی پیروی می‌کند.

$$(۳) \quad \chi^2 = \frac{\sum_{k=1}^p (o_k - E_k)^2}{\sum_{k=1}^p o_k}$$

معیار دیگری برای ارزیابی عملکرد دو روش نمونه‌آزاد و نمونه‌مبنا، محاسبه درصد انحراف نتیجه‌های واقعی از نتیجه‌های شبیه‌سازی شده است که به صورت زیر محاسبه می‌شود:

$$(۴) \quad \delta_k = \frac{|E_k - o_k|}{E_k} \times 100 \quad \forall k = 1, 2, \dots, p$$

که δ_k درصد انحراف نتیجه‌های واقعی از نتیجه‌های شبیه‌سازی شده در هر رده سنی k را نشان می‌دهد.

مقایسه‌ی روش نمونه‌مبنا با اطلاعات سرشماری ۸۵

برای مقایسه‌ی روش نمونه‌مبنا با اطلاعات سرشماری ۸۵، از سه معیار معرفی شده‌ی (۲) تا (۴) بالا به صورت زیر استفاده می‌شود:

- ۱- مقایسه‌ی شمارش‌های مشاهده‌شده (شمارش سرشماری) و مورد انتظار (شمارش شبیه‌سازی شده) انواع خانوار برای هر شهرستان از روی مقدار آماری آزمون یا p مقدار آن‌ها، رابطه (۳)،
- ۲- مقایسه‌ی میانگین درصد انحراف شمارش‌های مشاهده‌شده (شمارش سرشماری) و مورد انتظار (شمارش شبیه‌سازی شده) انواع خانوار برای هر شهرستان با استفاده از معیار معرفی شده در رابطه‌ی (۴)،
- ۳- مقایسه‌ی PGP به‌دست آمده از شمارش‌های مشاهده‌شده (شمارش سرشماری) و مورد انتظار (شمارش شبیه‌سازی شده) انواع خانوار برای هر شهرستان با استفاده از معیار معرفی شده در رابطه‌ی (۲)، و
- ۴- زمان اجرای برنامه در شهرستان‌های مختلف.

۱. مقایسه‌ی نتیجه‌های جدول‌های گروه‌های سنی مختلف

با محاسبه‌ی مقدار آماری (۲) و به‌دست آوردن p مقدار متناظر برای شبیه‌سازی جمعیت هم‌گذاشتی نوع خانوار در شهرستان‌های منتخب، جدول ۵ ساخته شد.

جدول ۵ نشان می‌دهد p مقدارهای در بیش‌تر موارد با افزایش اندازه‌ی جامعه بزرگ‌تر می‌شوند.

جدول ۵: p مقدارهای سرپرست انواع خانوارهای شبیه‌سازی شده چهار شهرستان منتخب به‌وسیله‌ی

روش نمونه‌مبنا

| شهرستان | تک نفره | تک والد | زوجی | سایر خانوارها |
|---------|---------|---------|--------|---------------|
| محلات | ۰/۹۹۹۸ | ۰/۹۹۹۳ | ۰/۹۹۹۹ | ۰/۹۹۹۴ |
| شاهرود | ۱ | ۰/۹۹۹۹ | ۱ | ۰/۹۹۸۸ |
| قزوین | ۱ | ۰/۹۹۹۹ | ۰/۹۹۹۹ | ۰/۹۹۹۹ |
| قم | ۱ | ۰/۹۹۹۹ | ۰/۹۹۹۶ | ۰/۹۹۹۹ |

۲. مقایسه‌ی درصد انحراف نتیجه‌های جدول‌های نوع خانوار (گارجیولو و همکاران ۲۰۱۰)

با محاسبه میانگین درصد انحراف‌ها از رابطه‌ی (۴) برای انواع خانوار در شهرستان‌های منتخب، جدول ۶ ساخته شد.

جدول ۶ نشان می‌دهد میانگین درصد انحراف تعداد انواع خانوارهای شبیه‌سازی شده در بیش‌تر موارد با افزایش اندازه‌ی جامعه کم‌تر می‌شوند.

جدول ۶: درصد انحراف شمارش سرشماری از شمارش شبیه‌سازی شده‌ی انواع خانوارهای چهار شهرستان منتخب به‌وسیله‌ی روش نمونه‌مبنا

| شهرستان | تک نفره | تک والد | زوجی | سایر خانوارها | میانگین انحراف |
|---------|---------|---------|------|---------------|----------------|
| محلات | ۰ | ۲/۵۳ | ۰/۲۸ | ۵/۴۸ | ۲/۰۷ |
| شاهرود | ۰ | ۰/۱۰ | ۰/۰۱ | ۲/۰۲ | ۰/۵۳ |
| قزوین | ۰ | ۰/۱۶ | ۰/۰۴ | ۰/۱۰ | ۰/۰۸ |
| قم | ۰ | ۰/۹۷ | ۰/۸۱ | ۱/۱۹ | ۰/۷۴ |

۳. مقایسه‌ی روش نمونه‌مبنا با اطلاعات سرشماری ۸۵، با استفاده از معیار **PGP** (نورماند و دی‌فیوانت ۲۰۱۳)

با نتیجه‌های محاسبه‌ی معیار **PGP** از رابطه‌ی (۲) برای انواع خانوار در شهرستان‌های منتخب، جدول ۷ ساخته شد.

نتیجه‌ها در جدول ۷ نشان می‌دهند که معیار **PGP** در شبیه‌سازی سرپرست خانوارها در بیش‌تر موارد با افزایش اندازه‌ی جامعه بزرگ‌تر می‌شوند. به عبارت دیگر شمارش‌های شبیه‌سازی شده به شمارش‌های سرشماری سال ۸۵ نزدیک شده‌اند.

جدول ۷: مقدار معیار **PGP** برای سرپرست انواع خانوارهای شبیه‌سازی شده چهار شهرستان منتخب

به‌وسیله‌ی روش نمونه‌مبنا

| شهرستان | تک نفره | تک والد | زوجی | سایر خانوارها |
|---------|---------|---------|--------|---------------|
| محلات | ۱ | ۰/۹۸۵۰ | ۰/۹۹۸۶ | ۰/۹۳۷۵ |
| شاهرود | ۱ | ۰/۹۹۹۵ | ۰/۹۹۹۸ | ۰/۹۸۹۵ |
| قزوین | ۱ | ۰/۹۹۹۱ | ۰/۹۹۹۷ | ۰/۹۹۹۵ |
| قم | ۱ | ۰/۹۹۵۲ | ۰/۹۹۹۶ | ۰/۹۹۴۱ |

۴. زمان اجرای برنامه

یکی از مزیت‌های روش نمونه‌مبنا تاثیر ناپذیری قابل توجه اجرای کدهای R به‌منظور شبیه‌سازی جمعیت‌ها از اندازه‌ی جمعیت جامعه‌ی مورد نظر می‌باشد. جدول ۸ مدت زمان اجرای کدهای R را برای شبیه‌سازی جمعیت هم‌گذاشتی سال پایه (۱۳۸۵) نشان می‌دهد.

جدول ۸: مدت زمان اجرای برنامه‌ی رایانه‌ای R به‌وسیله‌ی روش نمونه‌مبنا در چهار شهرستان منتخب

| شهرستان | نمونه مبنا |
|---------|------------|
| محلات | ۷ ثانیه |
| شاهرود | ۸ ثانیه |
| قزوین | ۸ ثانیه |
| قم | ۱۸ ثانیه |

به‌طور کلی مزایا و معایب روش شبیه‌سازی نمونه‌مبنا را می‌توان به‌صورت زیر خلاصه نمود.

الف) مزایا

- ۱) سرعت بالای اجرای کدهای نرم افزار R، به‌ویژه برای کوچک‌ناحیه‌ها با جمعیت زیاد،
- ۲) استفاده تنها از توزیع‌های حاشیه‌ای بدون نیاز به توزیع‌های کمکی جامعه یا ناحیه‌های بزرگ‌تر،
- ۳) آسان بودن نوشتن کدهای الگوریتم، و
- ۴) با اضافه کردن ستونی به ماتریس D (محدودیت جدید) به‌راحتی می‌توان ویژگی‌های جدید را مورد بررسی قرار داد.

ب) معایب

- ۱) ممکن است ویژگی یا محدودیتی در جامعه باشد که در نمونه‌ی منتخب موجود نباشد (به-عنوان مثال فرض کنید در جامعه‌ای تعداد ۳ سرپرست خانوار تک‌والد گروه سنی چهارم وجود داشته باشد ولی در نمونه‌ی منتخب این خانوارها موجود نباشند) در این‌صورت در برآورد محدودیت‌های جامعه نیز این برآورد صفر می‌شود یعنی خانه‌ی صفر ایجاد می‌شود، و
- ۲) ممکن است شمارش شبیه‌سازی شده یک نوع خانوار زیر یک باشد.

راه‌کارهایی برای مقابله با ضعف‌های روش نمونه‌مبنا در شبیه‌سازی جمعیت هم‌گذاشتی

به‌طور کلی می‌توان مشکلات شبیه‌سازی با روش نمونه‌مبنا را به‌وجود دو مشکل خانه‌ی صفر و وزن‌های زیر یک تقسیم‌بندی کرد. اکنون راه‌کارهای پیش‌نهادی برای مقابله با این مشکلات ارائه می‌شوند. برای رفع مشکل خانه‌ی صفر پژوهشگران راه‌کارهایی ارائه نموده‌اند که در زیر به برخی از آن‌ها اشاره می‌کنیم.

۱. قرار دادن یک وزن کوچک نزدیک به صفر در یکی از خانه‌های ستون محدودیت حاوی خانه‌ی صفر (بکمن و همکاران ۱۹۹۶، یه و همکاران ۲۰۰۹)

در این حالت خانه‌ی صفر محدودیت (ستون) متناظر را با عددی نزدیک به صفر پر می‌کنیم (این عدد را کم‌تر از یک انتخاب می‌کنیم تا تاثیر چندانی در نمونه نداشته باشد) در این صورت محدودیت، دیگر شامل خانه‌ی صفر نیست ولی باید توجه داشته باشیم که این عدد تاثیر چندانی بر وزن‌های نهایی نداشته باشد یعنی بعضی از وزن‌ها را زیر یک یا خیلی بزرگ نکند در صورت ایجاد این مشکل، عدد نزدیک به صفر را آن‌قدر در ستون محدودیت مورد نظر تغییر می‌دهیم تا این مشکل رفع شود.

۲. افزایش اندازه‌ی نمونه‌ای

با افزایش اندازه‌ی نمونه‌ای معمولاً اطلاعاتی که در جامعه وجود دارد در نمونه نیز یافت می‌شود و با پر شدن خانه‌های نمونه‌ای مشکل خانه‌ی صفر نیز برطرف می‌شود. البته این حالت پر هزینه است و استفاده از آن توصیه نمی‌شود.

لازم به‌ذکر است در این مقاله برای برطرف کردن مشکل خانه‌ی صفر از یک راه‌کار جدید به شرح زیر استفاده شده است.

۳. ادغام محدودیتی که شامل خانه‌ی صفر است، با محدودیت قبل یا بعد مرتبط با آن

به‌عنوان مثال در این حالت، محدودیت t_1 که معرف سرپرست خانوارهای تک‌نفره گروه سنی چهارم است را با گروه سنی بعدی یعنی محدودیت t_2 ، سرپرست خانوارهای تک‌نفره گروه

سنی پنجم ادغام کرده و برآورد متناظر با این خانه‌ی جدید را در جامعه که در بردار E خلاصه شده‌اند، به دست می‌آوریم. سپس برآورد خانه‌ی ادغام شده را در نسبت جامعه‌ای محدودیت‌های t_1 و t_2 ضرب می‌کنیم تا برآورد جداگانه برای این دو خانه به دست آید. در این مقاله برای برطرف کردن مشکل وزن‌های زیر یک سه راه‌کار جدید به شرح زیر ارائه شده است.

۱. کاستن محدودیت‌ها

در این حالت می‌توان برای هر توزیع چند بعدی که به عنوان خروجی در نظر گرفته می‌شود، یک ماتریس D ی جدید متناظر با محدودیت‌هایی که این توزیع چندبعدی را محاسبه می‌کنند بسازیم. البته مشروط بر این که این محدودیت‌ها از محدودیت‌های حذف شده مستقل باشند. به عنوان مثال زمانی که می‌خواهیم توزیع چند بعدی که فقط به خانوارهای تک‌والد و هسته‌ای مرتبط هستند، بسازیم، اگر محدودیت‌های مربوط به خانوارهای تک‌نفره و سایر خانوارها را حذف کنیم در این صورت خروجی وزن‌های نهایی اصلاح می‌شود.

۲. مدیریت تکرار الگوریتم

به طور کلی می‌دانیم که تعداد تکرارهای الگوریتم روش نمونه‌مبنا معمولاً تا زمانی ادامه می‌یابد که δ_a (درصد انحراف برآورد محدودیت‌های e_j از مجموع موزون نهایی، رابطه‌ی (۱)) مینیمم شود. در این روش می‌توان تعداد تکرارها را تا زمانی مدیریت کرد که در ستون وزن‌های نهایی وزن زیر یک موجود نباشد، مشروط بر اینکه δ_a معنادار باشد.

۳. تغییر در الگوریتم در هنگام مواجهه با وزن زیر یک

در این حالت می‌توان الگوریتم را طوری تغییر داد که هر گاه با وزن زیر یک مواجهه شود آن را یک در نظر گرفته و الگوریتم را ادامه دهد به این ترتیب الگوریتم را آن قدر ادامه می‌دهیم تا به δ_a ی معنادار برسیم.

لازم به ذکر است در این مقاله از روش مدیریت تکرار الگوریتم برای رفع مشکل وزن‌های زیر یک که تنها در شهرستان محلات وجود داشته، استفاده شده است.

بحث و نتیجه گیری

لازم به ذکر است که چهار شهرستان نمونه را طوری انتخاب کرده‌ایم که از لحاظ فرهنگی، اجتماعی، جغرافیایی و جمعیتی با یکدیگر متفاوت باشند. همچنین نمونه‌ی استفاده شده در روش نمونه مینا، برگرفته از اطلاعات ۲٪ سرشماری سال ۱۳۸۵ سایت مرکز آمار ایران است. با توجه به نتیجه‌های مقایسه‌ی روش نمونه مینا با اطلاعات سرشماری سال ۱۳۸۵ می‌توان نتیجه گرفت هر چه تعداد افراد و خانوار کوچک‌ناحیه‌ی مورد بررسی بیش‌تر شود نتیجه‌های روش نمونه مینا عمل‌کرد بهتری دارد. همچنین زمان اجرای کدهای برنامه روش نمونه مینا به‌طور قابل توجهی کم می‌باشد. به‌طور کلی می‌توان نتیجه گرفت که شبیه‌سازی جمعیت هم‌گذاشتی با روش نمونه مینا، در نوشتن کدهای برنامه ساده‌تر، سرعت اجرای آن بالاتر و در مجموع می‌توان گفت که نتیجه‌های عمل‌کرد روش نمونه مینا با افزایش اندازه‌ی جمعیت دقیق‌تر می‌شود.

منابع

- Arentze, T., Timmermans, H. J. P., and Hofman, F. (2007), Creating Synthetic Household Populations: Problems and Approach, *Transportation Research Record: Journal of the Transportation Research Board*, 2014: 85-91.
- Barthélemy, J., and Toint, P. L. (2012), Synthetic Populations: Review of the Different Approaches, Working paper. FUNDP - University of Namur.
- Beckman, R. J., Baggerly, K. A., and McKay, M. D. (1996), Creating Synthetic Baseline Populations, *Transportation Research Part A*, 30: 415-429.
- Deming, W. E., and Stephan, F. F. (1940), On a Least Squares Adjustment of a Sampled Frequency Table When the Expected Marginal Totals are Known, *Annals of Mathematical Statistics*, 11(4): 427-444.
- Gargiulo, F., Ternes, S., Huet, S., and Deffuant, G. (2010), An Iterative Approach for Generating Statistically Realistic Populations of Households, *PLoS ONE*, 5(1), e8828.
- Guo, J. Y., and Bhat, C. R. (2007), Population Synthesis for the Microsimulating Travel Behavior, *Transportation Research Record: Journal of the Transportation Research Board*, 2014: 92-101.
- Huang, Z., and Williamson, P. (2002), A Comparison of Synthetic Reconstruction and Combinatorial Optimisation Approaches to the Creation of Small-Area Microdata, Working paper. Population microdata unit, Liverpool, United Kingdom: University of Liverpool, Department of Geography.

- Lenormand, M., and Deffuant, G. (2013), Generating a Synthetic Population of Individuals in Households: Sample-Free Vs Sample-Based Methods, *Journal of Artificial Societies and Social Simulation*, 16(4): 12.
- Wilson, A. G., and Pownall, C. E. (1976), A New Representation of the Urban System for Modelling and for the Study of Micro-level Interdependence, *Area*, 8(4): 246-254.
- Voas, D., and Williamson, P. (2001), Evaluating Goodness-of-Fit Measures for Synthetic Microdata, *Geographical and Environmental Modeling*, 5: 177-200.
- Ye, X., Konduri, K., Pendyala, R. M., Sana, B., and Waddel, P. (2009), A Methodology to Match Distributions of Both Household and Person Attributes in the Generation of Synthetic Populations, Washington, U.S.A., Transportation Research Board - 88th Annual Meeting.