نامه انجمن جمعیت‌شناسی ایران / سال دهم، شمارهٔ ۱۹، بهار و تابستان ۹۴، ۸۹-۱۰۷

شبیه‌سازی جمعیتی کوچک‌ناحیه‌ها در سال پایه

محمدتقی معطی[۱]

حمیدرضا نواب‌پور[۲]

چکیده

در ایران پیش‌بینی‌های جمعیتی به‌دلیل نبود اطلاعات کمکی، اغلب در سطح ناحیه‌های بزرگ همانند استان و کشور انجام می‌شود. اما تولید اطلاعات جمعیتی و ویژگی‌های آن در سطح کوچک‌ناحیه[۳]ها (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] ویژگی‌های افراد و خانوارها (به‌عنوان مثال سنی– 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. ازآنجاکه دسترسی به بسیاری از اطلاعات جمعیتی و ویژگی‌های آن در کوچک‌ناحیه‌ها از طریق سرشماری‌ها امکان پذیر نیست، بسیاری از محققان را بر آن داشته است تا برای دست‌یابی به اطلاعات مورد نیاز خود از کوچک ناحیه‌ها به دنبال راه‌کارهایی باشند. در سال‌های اخیر روش‌های مختلفی برای شبیه‌سازی‌های جمعیتی ابداع شده است، از جمله‌ی این روش‌ها می‌توان به روش نمونه‌مبنا

---

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

۲دانشیار آمار دانشگاه علامه طباطبایی، (نویسنده مسئول)، h.navvabpour@srtc.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 شبیه-سازی 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[1]، 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

1 Based Sample

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 اطلاعی فراتر از آنچه که در چارچوب نتیجه-های 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[1] (SR) (ویلسون و پونال ۱۹۷٦) اشاره کرد. این روش‌ها عموما از فن برازش متناسب تکراری[2] (IPF) (دمینگ و استفان ۱۹٤۰) و نمونه‌ای از جامعه‌ی هدف برای به‌دست آوردن توزیع‌های توام استفاده می‌کنند (بکمن و همکاران ۱۹۹٦، هانگ و ویلیامسون ۲۰۰۲، گوآ و بهات ۲۰۰۷، آرنتزه 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 ارایه شده است. از جمله‌ی آن‌ها، روش روزامدسازی متناسب تکراری[3] (IPU) که توسط یه و همکاران (۲۰۰۹) توسعه یافته، مورد توجه بیش‌تری است.

روش بهینه‌سازی ترکیبی[4] (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، بارته‌لمی 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، $x = 1,2,\dots,q$ و ویژگی هر خانوار را با u_y، $y = 1,2,\dots,p$ 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:

$$n_T = \left\{ n_{t_1}, n_{t_2}, \dots, n_{t_q} \right\}$$
$$n_U = \left\{ n_{u_1}, n_{u_2}, \dots, n_{u_p} \right\}.$$

۱. ویژگی‌های فردی

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

۱) گروه سنی: [۵-۰]، [۱۰-۵]، ...، [<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، ٤، ٥، ٦، ۷ و بیش‌تر.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

۳. جامعه‌های مرجع

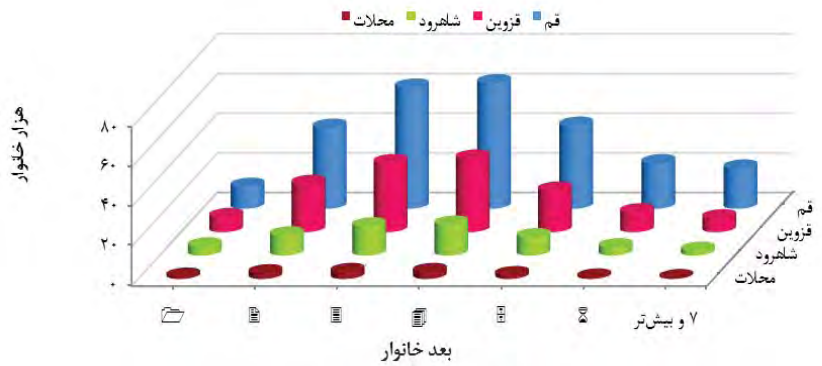
چهار شهرستان قم، قزوین، شاهرود و محلات را به‌عنوان چهار جامعه‌ی مرجع در نظر گرفته‌ایم. این شهرستان‌ها به دلیل داشتن جمعیت‌های متفاوت (جمعیت شهرستان‌های منتخب در سرشماری سال ۱۳۸۵از حدود ۵۰۰۰۰ نفر تا ۱۰۰۰۰۰۰ نفر بوده است)، نوع خانوارهای مختلف (تک‌نفره، زوجی بدون فرزند، تک‌والد، هسته‌ای)، بعد خانوار متفاوت، 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 به‌ترتیب دارای ۲۶۱۵۸۶، ۱۴۱۶۷۹، ۶۰۷۹۲ و ۱۳۹۳۳ خانوار و به‌ترتیب دارای جمعیتی بالغ بر ۱۰۳۴۰۷۹، ۵۲۷۰۵۰، 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 به‌شرح 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		۰-۴	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	بیشتر از ۸۵
افراد	تحلات	۳۲۶۷	۳۵۱۷	۳۹۷۹	۵۰۹۰	۵۳۷۸	۴۳۵۰	۳۶۲۹	۳۸۹۶	۳۲۹۰	۲۷۶۸	۲۲۰۱	۱۳۳۹	۱۲۰۴	۱۲۷۱	۷۰۳	۵۲۰	۳۴۰	
	شاهرود	۱۵۹۸۶	۱۵۶۸۲	۱۸۵۳۸	۲۶۱۳۱	۲۷۷۸۰	۲۲۱۹۰	۱۹۸۶۱	۱۵۳۲۰	۱۳۵۰۳	۱۲۴۲۰	۹۵۳۲	۵۹۲۵	۴۸۸۹	۴۵۲۲	۲۰۹۰	۱۶۶۴	۱۸۶۸	۸۵۸
	قزوین	۱۳۸۷۳	۱۳۶۴۵۵	۴۷۷۵۲	۶۲۹۰۷	۶۵۵۱۶	۵۵۱۵۷	۴۵۱۰۶	۳۹۱۶۸	۳۳۸۹۶	۲۶۹۷۲	۲۱۲۴۱	۱۲۶۷۲	۱۵۵۴۳	۹۱۳۷	۵۰۹۴	۳۳۲۹	۱۹۶۹	
	قم	۸۸۳۹۸	۸۷۷۲۰	۹۹۶۰۳	۱۲۴۸۳۳	۱۲۸۶۸۵	۱۱۰۲۴۴	۸۵۵۷۹	۷۲۲۱۹	۵۹۷۱۹	۴۴۴۲۱	۳۸۰۵۶	۲۳۵۰۱	۱۸۸۱۲	۱۵۴۱۶	۸۱۱۲	۶۰۷۷	۳۱۰۸	
سرپرست خانوار	تحلات	۰	۰	۳	۳۵	۴۱۷	۱۱۹۱	۱۶۳۱	۱۶۹۳	۱۷۵۹	۱۵۷۷	۱۲۶۱	۷۹۰	۷۷۵	۷۶۵	۸۶۰	۵۰۹	۴۰۷	۲۶۰	
	شاهرود	۰	۰	۷	۲۴۱	۲۳۵۲	۶۶۰۶	۷۶۰۳	۷۵۳۷	۶۹۳۶	۶۷۰۳	۵۳۲۴	۳۹۵۵	۳۶۹۵	۲۹۹۷	۲۹۰۲	۱۹۳۴	۱۳۶۰	۵۶۲	
	قزوین	۰	۰	۱۱	۳۳۷	۵۰۳۲	۱۶۴۶۱	۲۰۶۲۷	۱۹۴۶۷	۱۷۸۱۵	۱۴۵۲۴	۱۱۶۷۸	۸۶۶۶	۷۶۹۹	۶۱۷۳	۶۳۱۴	۳۵۱۲	۲۱۹۲	۴۷۱	
	قم	۰	۰	۳۰	۶۴۸	۱۱۰۸۹	۳۵۴۸۵	۴۰۲۳۵	۳۶۶۵۵	۳۲۱۴۹	۲۷۷۶۹	۲۱۸۴۳	۱۳۵۰۳	۱۱۳۹۳	۹۹۷۸	۹۸۰۷	۵۳۴۸	۳۹۴۲	۱۷۶۲	
سرپرست خانوارهای تک نفره	تحلات	۰	۰	۰	۱۴	۵۳	۳۹	۲۸	۲۳	۲۷	۳۴	۵۳	۵۵	۸۱	۱۱۵	۲۱۷	۱۵۲	۱۳۲	۱۰۶	
	شاهرود	۰	۰	۲	۱۲۲	۴۵۸	۲۰۲	۱۱۷	۱۰۸	۷۹	۱۳۱	۱۶۲	۲۱۷	۴۱۹	۴۸۰	۷۶۴	۵۸۳	۴۷۷	۲۲۸	
	قزوین	۰	۰	۷	۱۰۲	۴۳۶	۳۸۹	۲۵۱	۲۳۶	۱۸۵	۲۳۱	۲۸۴	۳۶۶	۶۸۶	۹۵۹	۱۳۷۲	۹۶۶	۷۳۴	۳۸۱	
	قم	۰	۰	۵	۹۲	۴۱۸	۴۷۷	۳۱۱	۲۶۳	۲۸۳	۳۶۶	۵۵۳	۵۶۸	۱۰۶۷	۱۲۷۵	۲۱۲۰	۱۳۴۷	۱۲۷۶	۶۵۸	
سرپرست خانوارهای تک والد	تحلات	۰	۰	۰	۱	۸	۳۵	۴۸	۷۲	۱۲۱	۹۸	۱۰۴	۷۸	۶۹	۶۵	۵۱	۲۱	۱۳	۱۰	
	شاهرود	۰	۰	۰	۵	۴۳	۱۰۹	۱۷۲	۲۸۲	۳۷۸	۴۲۳	۵۰۳	۴۳۳	۴۴۰	۲۵۶	۱۹۰	۱۳۳	۷۱	۲۹	
	قزوین	۰	۰	۰	۳۰	۹۹	۲۷۷	۴۸۱	۶۹۱	۸۲۹	۸۹۹	۸۷۷	۶۴۴	۵۰۰	۴۱۱	۲۱۳	۱۱۲	۶۹	
	قم	۰	۰	۰	۴	۷۵	۲۴۶	۶۴۹	۱۱۶۵	۱۵۶۴	۱۸۲۵	۱۸۸۵	۱۱۸۹	۸۲۱	۶۴۵	۳۲۷	۱۹۹	۷۹	
سرپرست خانوارهای زوجی	تحلات	۰	۰	۰	۸	۲۹۶	۱۰۴۹	۱۴۸۸	۱۵۰۳	۱۵۰۸	۱۳۳۲	۱۰۰۳	۵۹۸	۵۷۵	۵۳۷	۵۴۲	۳۲۱	۲۲۶	۱۳۶	
	شاهرود	۰	۰	۰	۴۵	۱۵۷۴	۵۹۲۸	۶۸۶۲	۶۶۴۳	۵۹۳۸	۵۴۹۶	۴۱۰۹	۲۹۰۶	۲۴۳۲	۱۹۲۲	۱۶۷۷	۱۰۶۴	۷۲۱	۲۶۰	
	قزوین	۰	۰	۱	۹۴	۳۸۸۲	۱۴۸۹۹	۱۸۸۵۷	۱۷۵۲۹	۱۵۵۹۶	۱۲۱۱۰	۹۳۲۵	۶۷۰۸	۵۴۹۴	۴۱۴۵	۲۰۵۳	۱۱۹۳	۴۶۰	
	قم	۰	۰	۴	۱۶۱	۴۸۶۱	۳۲۳۱۲	۳۶۶۴۵	۳۲۶۲۲	۲۷۶۱۲	۲۲۷۴۶	۱۶۸۸۱	۹۸۲۳	۷۸۲۹	۶۷۴۸	۶۰۲۰	۳۱۶۶	۲۱۴۰	۸۹۶	
فرزندان خانوارهای تک والد و زوجی	تحلات	۳۱۵۸	۳۴۲۴	۳۸۶۴	۴۶۶۷	۳۵۵۵	۱۳۹۸	۴۳۷	۱۹۵	۹۰	۵۵	۴	۱۶	۱	۰	۱	۰	۰	۰	
	شاهرود	۱۵۲۴۳	۱۵۲۷۲	۱۸۰۵۷	۲۳۶۸۸	۱۸۳۶۹	۷۰۵۳	۱۹۵۳	۸۲۹	۳۵۳	۱۵۰	۲۸	۶۷	۱۴	۵	۰	۰	۰	۰	
	قزوین	۳۵۴۰۲	۳۵۵۴۹	۴۵۵۳۴	۵۷۲۹۸	۴۳۵۴	۱۷۷۰۷	۵۰۰۸	۱۸۰۷	۷۳۳	۳۰۲	۱۶۲	۵۲	۲۰	۱۱	۰	۰	۰	۰	
	قم	۸۵۴۶۳	۸۵۵۱۲	۹۲۷۶۳	۱۱۲۶۹۶	۷۹۳۳۳	۳۰۴۳۰	۸۲۰۲	۲۷۵۵	۱۱۳۱	۵۲۰	۲۵۱	۷۴	۳۱	۱۱	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_s نشان می‌دهیم، شروع می‌شود. برای بازسازی مجدد P از وزن‌های وابسته به خانوارهای نمونه‌ای یا افراد درون این خانوارها (W_jها) به‌منظور 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[1] (WS_j) با محدودیت‌های براورد شده‌ی e_j به‌وسیله‌ی تنظیم کردن وزن‌ها است. وزن W_j، وزن خانوار یا فرد درون خانوار j ام از نمونه‌ی موزون و e_j براوردی از تعداد کل ویژگی j ام خانوار یا فرد درون خانوار در P می‌باشد

$$\widehat{e_j} = \begin{cases} n_{u_j} & j = ۱,۲,\dots,p \\ n_{t_j} & j = p+۱, p+۲, \dots, p+q \end{cases}$$

1 Weighted Sum

که در آن $\{n_{t_k}\}_{1\le k\le q}$ تعداد افراد مربوط به هر ویژگی فردی و $\{n_{u_l}\}_{1\le l\le p}$، تعداد خانوارهای مربوط به هر ویژگی خانواری را نشان می‌دهند. این براورد 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_s 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 درنظر می‌گیریم برابر عدد معلوم N باشد. پس تعداد اعضای مجموعه‌ی P_s برابر N است، یعنی $|P_s| = N$، که نماد $|.|$ 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 $۷ \times ۵ = ۳۵$ 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 $p = ۱۹$.

- ($u_۱$) خانوارهای تک‌نفره،
- ($u_۲ - u_۷$) خانوارهای تک‌والد دو نفره تا خانوارهای تک‌والد بیش از ٦ نفر،
- ($u_۸$) خانوارهای زوجی بدون فرزند،
- ($u_۹ - u_{۱۳}$) خانوارهای هسته‌ای سه نفره تا خانوارهای هسته‌ای بیش از ٦ نفر، و
- ($u_{۱٤} - u_{۱۹}$) سایر خانوارهای دو نفره تا سایر خانوارهای بیش از ٦ 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 $q = ۱۲۵$.

- ($t_۱ - t_{۱۵}$) رده سنی ٤ تا رده سنی ۱۸ی سرپرست خانوار تک‌نفره،
- ($t_{۱٦} - t_{۳۱}$) رده سنی ۳ تا رده سنی ۱۸ی سرپرست خانوار تک‌والد،

- ($t_{۳۲} - t_{٤٦}$) رده سنی ٤ تا رده سنی ۱۸ی سرپرست خانوار زوجی بدون فرزند،
- ($t_{٤٧} - t_{٦١}$) رده سنی ٤ تا رده سنی ۱۸ی سرپرست خانوار هسته‌ای،
- ($t_{٦٢} - t_{٧٧}$) رده سنی ۳ تا رده سنی ۱۸ی سرپرست سایر خانوارها،
- ($t_{٧٨} - t_{٩١}$) رده سنی ۱ تا رده سنی ١٤ی فرزندان خانوار،
- ($t_{٩٢} - t_{١٠٧}$) رده سنی ۳ تا رده سنی ۱۸ی شریک سرپرست خانوار،
- ($t_{١٠٨} - t_{١٢٥}$) رده سنی ۱ تا رده سنی ۱۸ی دیگر افراد سایر خانوارها، و

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

$$\delta_b = \frac{|\widehat{e_j} - ws_j|}{\widehat{e_j}} \quad (۱)$$

یکی از مشکلات روش نمونه‌مبنا وجود خانه‌ی صفر است. یعنی ممکن است براورد محدودیت‌های جامعه که در بردار Eی ماتریس D قرار دارند، یک عدد غیر صفر باشد در صورتی‌که برخی از اعضای بردارws 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 $t_{١}$، $t_{١٣}$، $t_{١٦}$، $t_{١٧}$، $t_{٣٠}$، $t_{٣١}$، $t_{٣٢}$، $t_{٦١}$ و $t_{١٠٨}$ اشاره کرد. این محدودیت‌ها ۹ محدودیت از ١٤٤ محدودیت در شهرستان محلات هستند که در نمونه صفرند ولی در جامعه صفر نیستند.

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

شماره خانوار	W	۱u	۲u	…	u_p	۱t	۱t	…	t_q	۱W	۲W	…	W_{p+q}
۱													
۲													
…													
N													
WS													
E	-	۱e	۲e	…	e_p	۱e_{p+}	۲e_{p+}	…	e_{p+q}				
δ_b													
۱WS													
۲WS													
…													
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 عمل‌کرد

برای آزمون عمل‌کرد روش نمونه‌مبنا از روی جامعه‌های مرجع (توزیع‌های معرفی شده در جدول ۱ و شکل ۱) با استفاده از الگوریتم معرفی‌شده 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

بهتـــرین نتیجـــه حاصـــل شـــود (مـــلاک انتخـــاب بهتـــرین نتیجـــه درصـــد انحـــراف بـــراورد محدودیت‌های e_j از مجموع موزون نهایی است که در رابطه (۱) بیان شده اســت). نتیجـه‌هـای روش نمونه‌مبنا به‌تفکیک شهرستان‌های منتخب در جدول‌های ۳ و ٤ 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

گروه سنی		۴ـ۰	۹ـ۵	۱۴ـ۱۰	۱۹ـ۱۵	۲۴ـ۲۰	۲۹ـ۲۵	۳۴ـ۳۰	۳۹ـ۳۵	۴۴ـ۴۰	۴۹ـ۴۵	۵۴ـ۵۰	۵۹ـ۵۵	۶۴ـ۶۰	۶۹ـ۶۵	۷۴ـ۷۰	۷۹ـ۷۵	۸۴ـ۸۰	بیشتر از ۸۵
سرپرست خانوارهای تک نفره	محلات	۰	۰	۰	۱٤	۵۳	۳۹	۲۸	۲۳	۳۷	۳٤	۵۳	۵۵	۸۱	۱۱۵	۲۱۷	۱۳۲	۱۵۲	۱۰۶
	شاهرود	۰	۰	۲	۱۲۲	٤۵۸	۲۰۲	۱۱۷	۱۰۸	۷۹	۱۳۱	۲۱۷	۱۶۲	٤۱۹	٤۸۰	۷۶٤	۵۸۳	٤۷۷	۲۲۸
	قزوین	۰	۰	۷	۱۰۲	٤۳۶	۳۸۹	۲۵۱	۲۳۶	۱۸۵	۲۳۱	۳۶۶	۲۸٤	۶۸۶	۹۵۹	۱۳۷۲	۹۶۶	۷۳٤	۳۸۱
	قم	۰	۰	۵	۹۲	٤۱۸	٤۷۷	۳۱۱	۲۶۳	۲۸۳	۳۶۶	۵۶۳	۵۶۸	۱۰۶۷	۱۲۷۵	۲۱۲۰	۱۳٤۷	۱۲۷۶	۶۵۸
سرپرست خانوارهای تک والد	محلات	۰	۰	۰	۱	۸	۳۵	٤۸	۷۲	۱۲۱	۹۸	۱۰٤	۷۸	۶۹	۶۵	۵۱	۳۱	۱۹	۱۵
	شاهرود	۰	۰	۰	۵	۴۳	۱۰۹	۱۷۲	۲۸۲	۳۷۸	٤۲٤	۵۰٤	٤۳۳	٤٤۰	۲۵۶	۱۹۰	۱۳۳	۷۱	۲۹
	قزوین	۰	۰	۰	۰	۳۰	۹۹	۲۶۶	٤۸۰	۶۸۹	۸۲۷	۸۹۶	۸۷۵	۶۹۲	٤۹۹	٤۱۰	۲۱۳	۱۱۲	۶۹
	قم	۰	۰	۰	٤	۷۵	۲٤۷	۶۵۳	۱۱۷۳	۱۵۷۸	۱۸٤۰	۱۸۹۹	۱٤۲۶	۱۱۹٤	۸۲٤	۶٤۸	۳۲۹	۲۰۰	۱۱۶
سرپرست خانوارهای زوجی	محلات	۰	۰	۰	۸	۲۹۶	۱۰٤۷	۱٤۸٤	۱٤۹۸	۱۵۰۳	۱۳۲۷	۹۹۹	۵۹۶	۵۷۳	۵۳۶	۵٤۱	۳۲۱	۲۶۶	۱۳۶
	شاهرود	۰	۰	۰	٤۵	۱۵۷۳	۵۹۲۵	۶۸۶۱	۶۶٤٤	۵۹٤۱	۵٤۹۹	٤۱۱۱	۲۹۰۷	۲٤۳۲	۱۹۲۲	۱۶۷۷	۱۰۶٤	۷۲۱	۲۶۰
	قزوین	۰	۰	۱	۹٤	۳۸۷۳	۱٤۸۵۶	۱۸۷۸۸	۱۷٤۵۱	۱۵۵۲۱	۱۲۰۵۱	۹۲۸۲	٤۱۳۱	۵٤۷۰	۶۶۷۸	٤۰۰۳	۲۰٤۷	۱۱۹۰	٤۵۹
	قم	۰	۰	٤	۲۶۲	۹۱۹۶	۳۲٤۷۰	۳۶۹۳۸	۳۲۹۰۳	۲۷۹۰۷	۲۳۰٤۶	۱۷۰۵۹	۹۹۱۹	۷۹۰۶	۶۸۰۷	۶۰۰۳	۳۱۸۱	۲۱۵۲	۸۹۹
فرزندان خانوارهای تک والد و زوجی	محلات	۰	۰	۶	۲۵	۵۱	۶۵	۵۷	۱۱٤	۱۱۸	۱۳۱	۹۲	۶۰	۵۲	۵۰	۵۱	۳۹	۱۶	۱۰
	شاهرود	۰	۰	۵	۶۹	۱۷۰	٤٤۷	٤۵۲	۵۰٤	۵٤۲	۶۵٤	۵٤۸	۳۹۹	٤۰٤	۳۳۹	۲۷۱	۱۵٤	۹۱	٤۵
	قزوین	۰	۰	۳	۱٤۰	۶۸۱	۱۰۷۰	۱۲۳۵	۱۲۱۳	۱۳۳۷	۱۳٤۵	۱۱۶۲	۹۰۹	۸۲۰	۵۶۶	۵۱۳	۲۷۹	۱۷۳	۶۹
	قم	۰	۰	۲۱	۲۹۲	۱٤٤۳	۲٤۷۷	۲۶۰۷	۲۶۳۲	۲۷۲۷	۲۸۲٤	۲۵۵۸	۱۶۷۱	۱۳۲۱	۱۱٤۷	۱۰۳۳	۵۱٤	۳۳٤	۱۳۱

جدول ٤: توزیع شمارش شبیه‌سازی شده نوع خانوارها در سال ۱۳۸۵ به‌وسیله‌ی روش نمونه‌مبنا

شهرستان	نوع خانوار روش	تک نفره	تک والد	زوجی	سایر خانوارها
محلات	سرشماری ۸۵	۱۱۲۹	۷۹٤	۱۱۱۲۲	۸۸۸
	نمونه مبنا	۱۱۲۹	۸۱٤	۱۱۰۹۱	۹۳۷
شاهرود	سرشماری ۸۵	٤۵٤۹	۳٤۶۷	٤۷۵۷۷	۵۱۹۹
	نمونه مبنا	٤۵٤۹	۳٤۷۰	٤۷۵۸۰	۵۰۹٤
قزوین	سرشماری ۸۵	۷۵۸۵	۶۱۸۲	۱۱۶۳۶۱	۱۱۵۵۱
	نمونه مبنا	۷۵۸۵	۶۱۶۶	۱۱۵۸۹٤	۱۱۵۱۶
قم	سرشماری ۸۵	۱۱۰۷۹	۱۲۰۸۹	۲۱٤۹۶۶	۲۳٤۵۲
	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 می کند و توسط لنورماند و دی‌فیوانت (۲۰۱۳) معرفی شده است، استفاده می‌شود.

$$PGP = ۱ - \frac{۱}{۲} \times \frac{\sum_{k=۱}^{p} |O_k - E_k|}{\sum_{k=۱}^{p} O_k} \qquad (۲)$$

معیار PGP [1]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 به‌ترتیب شمارش‌های مشاهده‌شده و مورد انتظار خانوار یا افراد درون خانوار kام را نشان می‌دهند.

در آماره‌ی آزمونی دیگر، از فاصله‌ی خی‌دو استفاده می‌شود (لنورماند و دی‌فیوانت ۲۰۱۳). این فاصله از یک توزیع خی‌دو با $p - ۱$ درجه‌ی آزادی پیروی می‌کند.

$$\chi^{۲} = \frac{\sum_{k=۱}^{p} (O_k - E_k)^{۲}}{\sum_{k=۱}^{p} O_k} \qquad (۳)$$

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

$$\delta_k = \frac{|E_k - O_k|}{E_k} \times ۱۰۰ \qquad \forall k = ۱,۲,\dots,p \qquad (٤)$$

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) تـا (٤) بالا به‌صورت زیر استفاده می‌شود:

---

1 Proportion of Good Prediction

۱- مقایسه‌ی شمارش‌های مشاهده‌شده (شمارش سرشماری) و مورد انتظار (شمارش شبیه‌سازی شده) انواع خانوار برای هر شهرستان از روی مقدار آماره‌ی 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 (٤)،

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

٤- زمان اجرای برنامه در شهرستان‌های مختلف.

۱. مقایسه‌ی نتیجه‌های جدول‌های گروه‌های سنی مختلف

با محاسبه‌ی مقدار آماره‌ی (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	۰/۹۹۹٤
شاهرود	۱	۰/۹۹۹۹	۱	۰/۹۹۸۸
قزوین	۱	۰/۹۹۹۹	۰/۹۹۹۹	۰/۹۹۹۹
قم	۱	۰/۹۹۹۹	۰/۹۹۹٦	۰/۹۹۹۹

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

با محاسبه میانگین درصد انحراف‌ها از رابطه‌ی (٤) برای انواع خانوار در شهرستان‌های منتخب، جدول ٦ ساخته شد.

جدول ٦ نشان می‌دهد میانگین درصد انحراف تعداد انواع خانوارهای شبیه‌سازی شده در بیش‌تر موارد با افزایش اندازه‌ی جامعه کم‌تر می‌شوند.

جدول ٦: 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	۵/٤۸	۲/۰۷
شاهرود	۰	۰/۱۰	۰/۰۱	۲/۰۲	۰/۵۳
قزوین	۰	۰/۱٦	۰/۰٤	۰/۱۰	۰/۰۸
قم	۰	۰/۹۷	۰/۸۱	۱/۱۹	۰/۷٤

۳. مقایسه‌ی روش نمونه‌مبنا با اطلاعات سرشماری ۸۵، با استفاده از معیار PGP (لنورماند 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	سایرخانوارها
محلات	۱	۰/۹۸۵۰	۰/۹۹۸٦	۰/۹۳۷۵
شاهرود	۱	۰/۹۹۹۵	۰/۹۹۹۸	۰/۹۸۹۵
قزوین	۱	۰/۹۹۹۱	۰/۹۹۹۷	۰/۹۹۹۵
قم	۱	۰/۹۹۵۲	۰/۹۹۹٦	۰/۹۹٤۱

٤. 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 اجـرای کـدهای 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
٤) 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 (به‌عنوان 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 براورد محدودیت‌های جامعه نیز این براورد 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 ارایه می‌شوند. برای رفع مشکل خانه‌ی صفر پژوهشگران راه‌کارهایی ارایه 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 این‌صورت 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 معمولا 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 براورد متناظر با این خانه‌ی جدید را در جامعه که در بـردار E خلاصـه شده‌اند، به‌دست می‌آوریم. سپس براورد خانه‌ی ادغام شده را در نسبت جامعه‌ای محدودیت‌های t_1 و t_2 ضرب می‌کنیم تا براورد جداگانه برای این دو خانه به‌دست آید.

در این مقاله برای برطرف کردن مشکل وزن‌های زیر یک سه راه‌کار جدید به شرح زیر ارایه شده است.

۱. کاستن محدودیت‌ها

در این حالت می‌توان برای هر توزیع چند بعدی که به‌عنوان خروجی در نظـر گرفتـه مـی‌شـود، یک ماتریس Dی جدید متناظر با محدویت‌هایی که این توزیع چندبعـدی را محاسـبه مـی‌کننـد بسازیم. البته مشروط بر این‌که این محدودیت‌ها از محدودیت‌های حـذف شـده مسـتقل باشـند. به‌عنوان مثال زمانی که می‌خواهیم توزیع چند بعدی که فقط به خانوارهای تک‌والـد و هسـته‌ای مرتبط هستند، بسازیم، اگر محدودیت‌های مربوط به خانوارهای تـک‌نفـره و سـایر خانوارهـا را حذف کنیم در این صورت خروجی وزن‌های نهایی اصلاح می‌شود.

۲. مدیریت تکرار الگوریتم

به‌طور کلی می‌دانیم که تعداد تکرارهای الگوریتم روش نمونه‌مبنا معمولا تا زمانی ادامه مـی‌یابـد که δ_a (درصد انحراف براورد 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 نهـایی 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 مواجـه شـود آن را یک در نظر گرفته و الگوریتم را ادامه دهد به این ترتیب الگوریتم را آن‌قدر ادامه می‌دهیم تـا بـه δ_aی معنادار برسیم.

لازم به‌ذکر 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 به‌ذکر 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 مبنا، برگرفته از اطلاعات ۲٪ سرشماری سال ۱۳۸۵ سایت مرکز آمار ایران است. با توجه به نتیجه‌های مقایسه‌ی روش نمونه‌مبنا با اطلاعات سرشماری سال ۱۳۸۵ می‌توان نتیجه گرفت هر چه تعداد افراد و خانوار کوچک‌ناحیه‌ی مورد بررسی بیش‌تر شود نتیجه‌های روش نمونه‌مبنا عمل‌کرد بهتری دارد. همچنین زمان اجرای کدهای برنامه روش نمونه‌مبنا به‌طور قابل توجهی کم می‌باشد. به‌طور کلی می‌توان نتیجه گرفت که شبیه‌سازی جمعیت هم‌گذاشتی با روش نمونه‌مبنا، در نوشتن کدهای برنامه ساده‌تر، سرعت اجرای آن بالاتر و در مجموع می‌توان گفت که نتیجه‌های عمل‌کرد روش نمونه‌مبنا باافزایش 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 Representationof 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.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی