

# امکان‌سنجی تولید جمعیت شبیه‌سازی شده برای طرح نیروی کار مرکز آمار ایران به روش همگذاشتی

مجری

اشکان شباک

همکاران

حامد لروند

کاوه کیانی

عباس مرادی

نادر حکیمی‌پور



پژوهشکده‌ی آمار

گروه پژوهشی پردازش داده‌ها و اطلاع‌رسانی

بهار ۱۳۹۸



## پیش‌گفتار

دشواری روزافزون گردآوری اطلاعات به روش‌های سنتی به دلیل پیچیدگی‌های امروزی جوامع آماری، نیاز به مطالعه برای دگرگونی یا بهنگام سازی روش‌های آمارگیری را ضروری کرده است. استفاده از دیگر منابع داده‌ها و مدل‌سازی، از روش‌هایی هستند که می‌توانند به عنوان جایگزین روش‌های آمارگیری به کار رفته یا به افزایش دقت برآوردها و استنباط‌های ناشی از آمارگیری‌های سنتی یاری رسانند. مدل‌های ریزشیه‌سازی توجه بسیاری از پژوهش‌گران حوزه‌ی جمعیت‌شناختی را به خود جلب کرده است. معمولاً یک مجموعه داده قادر به تأمین همه‌ی نیازهای محققان برای تصمیم‌گیری‌ها و برنامه‌ریزی‌ها نیست و این مسئله پژوهشگر را وادار به استفاده از انواع مختلفی از داده‌های موجود مانند ریزداده‌ها و داده‌های گردآوری شده در سرشماری‌ها می‌کند. جداول سرشماری شامل اطلاعاتی در مورد توزیع جغرافیایی جمعیت است اما معمولاً اطلاعات جزئی‌تر در مورد ویژگی‌های افراد یا خانوار، از ریزداده‌های حاصل از آمارگیری‌های نمونه‌ای به دست می‌آید. در اغلب موارد اطلاعات حاصل از این آمارگیری‌ها در سطوح بزرگتر موجود است.

به عنوان مثال در طرح «آمارگیری نیروی کار» که توسط مرکز آمار ایران به صورت فصلی انجام می‌شود، اطلاعات کافی برای ارائه برآوردها در سطح نواحی بزرگ مانند استان در اختیار محقق قرار می‌گیرد، اما مشکل اصلی زمانی ظاهر می‌شود که علاقمند به برآورد شاخص‌ها یا پارامترهای موردنظر در سطح کوچک‌ناحیه‌ها باشیم. کوچک‌ناحیه به جمعیتی اطلاق می‌شود که امکان ارائه‌ی برآورد قابل اعتماد برای آن به دلیل محدودیت‌های داده‌های موجود وجود ندارد. کوچک‌ناحیه می‌تواند یک ناحیه‌ی جغرافیایی و یا یک گروه جمعیتی باشد. در اغلب موارد ریزداده‌ها در کوچک‌ناحیه‌ها موجود نیستند یا تعداد نمونه به قدری ناچیز است که قادر به ارائه‌ی برآوردهای دقیق در آن سطح نیستیم. یکی از روش‌های برآورد در سطح کوچک‌ناحیه‌ها که در سال‌های اخیر مورد توجه زیادی قرار گرفته است، روش ساخت جامعه همگذاشتی است که به عنوان یک ابزار آماری مهم، می‌تواند در چنین مواردی مورد استفاده قرار گیرد. در واقع روش ساخت داده‌ی همگذاشتی، علاوه بر تأمین ریزداده‌ها در سطوح جغرافیایی کوچک، مجموعه داده‌ای تولید می‌کند که اصول محرمانگی را رعایت می‌کند و قابلیت انتشار عمومی برای استفاده‌ی محققان و موارد آموزشی دارد.

بر این اساس با هدف پاسخگویی به نیاز موجود در خصوص اطلاع از وضعیت اشتغال و بیکاری جمعیت در کوچک‌ناحیه‌ها، طرح پژوهشی «امکان‌سنجی تولید جمعیت شبیه‌سازی شده برای طرح نیروی کار مرکز آمار

ایران به روش همگذاشتی» در دستور کار پژوهشکده‌ی آمار قرار گرفت. این پژوهش سعی در معرفی روشی برای شبیه‌سازی جمعیت نیروی کار ایران در سطح استان‌ها و شهرستان‌ها دارد. از بین شهرستان‌های کشور چهار شهرستان اسلام‌آباد غرب، گچساران، دهلران و ابهر به عنوان پایلوت در نظر گرفته شده‌اند. به منظور مقایسه دقت برآورد نرخ بیکاری به این روش با مقدار واقعی آن، این کار را برای ۳۱ استان کشور نیز انجام داده‌ایم که نتایج مطلوبی به دست آمد.

این پژوهش در گروه پردازش داده‌ها و اطلاع‌رسانی پژوهشکده‌ی آمار توسط آقای دکتر اشکان شباک به عنوان مجری طرح و آقایان حامد لرونند، دکتر کاوه کیانی، دکتر عباس مرادی و دکتر نادر حکیمی‌پور به عنوان همکاران اصلی طرح پژوهشی انجام شده است که بدین وسیله از همه این عزیزان صمیمانه تشکر و قدردانی می‌شود.

گروه پژوهشی پردازش داده‌ها و اطلاع‌رسانی

پژوهشکده‌ی آمار

# فهرست مطالب

۱	۱- کلیات تحقیق
۱	۱-۱- مقدمه
۲	۲-۱- بیان مسئله
۵	۳-۱- اهمیت و ضرورت پژوهش
۶	۴-۱- اهداف تفصیلی پژوهش
۶	۵-۱- تعاریف و مفاهیم کلی مورد نیاز
۶	۱-۵-۱- داده‌های همگذاشتی
۷	۱-۵-۲- کوچک‌ناحیه
۸	۱-۵-۳- برآوردهای کوچک‌ناحیه‌ای
۹	۱-۵-۴- متغیر رسته‌ای
۹	۱-۵-۵- متغیر پیوسته
۹	۱-۵-۶- کار
۹	۱-۵-۷- بیکار
۱۰	۱-۵-۸- شاغل
۱۰	۱-۵-۹- جامعه هدف
۱۱	۱-۵-۱۰- جامعه آمارگیری
۱۱	۱-۵-۱۱- واحد آماری
۱۱	۱-۵-۱۲- جمعیت فعال اقتصادی
۱۱	۱-۵-۱۳- جمعیت غیرفعال اقتصادی
۱۱	۱-۵-۱۴- جمعیت همگذاشتی

۱۳	۲- مروری بر ادبیات پژوهش
۱۳	۱-۲- مقدمه
۱۳	۲-۲- طرح آمارگیری نیروی کار
۱۶	۳-۲- برآورد کوچک‌ناحیه‌ای نرخ بیکاری
۱۶	۲-۳-۱- پژوهش‌های داخلی برای برآورد کوچک‌ناحیه‌ای نرخ بیکاری
۱۷	۲-۳-۲- پژوهش‌های خارجی برای برآورد کوچک‌ناحیه‌ای نرخ بیکاری
۱۸	۴-۲- ساخت جمعیت همگذاشتی برای برآورد کوچک‌ناحیه‌ها
۲۱	۳- ریزش‌بیه‌سازی جمعیت همگذاشتی
۲۱	۱-۳- مقدمه
۲۲	۲-۳- ریزش‌بیه‌سازی
۲۳	۳-۳- بازسازی همگذاشتی
۲۵	۴-۳- بهینه‌سازی ترکیباتی
۲۹	۵-۳- تولید مدل-مینا
۳۰	۶-۳- استفاده از بسته‌ی نرم‌افزاری simPop در نرم‌افزار R برای تولید جمعیت همگذاشتی
۳۴	۷-۳- روش تولید جمعیت همگذاشتی با بسته‌ی simPop
۳۷	۴- شبیه‌سازی متغیرها در جمعیت همگذاشتی
۳۷	۱-۴- مقدمه
۳۷	۲-۴- مدل رگرسیون لجستیک
۴۰	۳-۴- شبیه‌سازی متغیرهای رسته‌ای
۴۱	۴-۴- شبیه‌سازی متغیرهای پیوسته
۴۳	۵- روش‌شناسی برآورد نرخ بیکاری سال‌های ۱۳۹۵ و ۱۳۹۶ با تولید جمعیت همگذاشتی
۴۳	۱-۵- مقدمه
۴۳	۲-۵- ساختارخانوار
۴۴	۳-۵- شبیه‌سازی متغیرهای رسته‌ای
۴۵	۴-۵- پسا کالیبره کردن جمعیت همگذاشتی

۴۶ ..... ۵-۵- تحلیل نتایج

۴۸ ..... ۶-۵- مقایسه تابع توزیع و پراکندگی تحت برآورد هوریتز- تامپسون و همگذاشتی

۵۰ ..... ۷-۵- نتایج برآورد نرخ بیکاری برای استان‌های کشور

۶۱ ..... ۶- نتیجه‌گیری

۶۱ ..... ۱-۶- مقدمه

۶۲ ..... ۲-۶- تحلیل نتایج طرح

۶۲ ..... ۳-۶- پیشنهادات

۶۵ ..... پیوست

۶۷ ..... مرجع‌ها





# فهرست جدول‌ها

- جدول ۳-۱- ریز داده‌های آمارگیری نمونه‌ای ..... ۲۷
- جدول ۳-۲- توزیع خانوارها بر حسب بعد خانوار ..... ۲۷
- جدول ۳-۳- توزیع اعضای خانوار بر حسب سن ..... ۲۷
- جدول ۳-۴- جدول بندی خانوارهای منتخب بر اساس قید کوچک ناحیه‌ای بعد خانوار ..... ۲۸
- جدول ۳-۵- جدول بندی خانوارهای منتخب بر اساس قید کوچک ناحیه‌ای سن عضو خانوار ..... ۲۸
- جدول ۳-۶- جدول بندی خانوارهای منتخب بر اساس قید کوچک ناحیه‌ای بعد خانوار پس از جابه‌جایی خانوار a با خانوار d ..... ۲۸
- جدول ۳-۷- جدول بندی خانوارهای منتخب بر اساس قید کوچک ناحیه‌ای سن عضو خانوار پس از جابه‌جایی خانوار a با خانوار d ..... ۲۹
- جدول ۳-۸- جدول بندی خانوارهای c و d منتخب بر اساس قید کوچک ناحیه‌ای بعد خانوار ..... ۲۹
- جدول ۳-۹- جدول بندی خانوارهای c و d منتخب بر اساس قید کوچک ناحیه‌ای سن عضو خانوار ..... ۲۹
- جدول ۳-۱۰- توابع مهم مورد استفاده در بسته‌ی نرم‌افزاری simPop ..... ۳۱
- جدول ۳-۱۱- جمعیت افراد به تفکیک جنس و وضع سواد ..... ۳۳
- جدول ۳-۱۲- جمعیت افراد به تفکیک سن و جنس ..... ۳۳
- جدول ۳-۱۳- برآورد هوریتز-تامپسون تعداد افراد بر حسب سن و توانایی صحبت کردن به زبان فارسی به تفکیک جنس ..... ۳۳
- جدول ۳-۱۴- برآورد هوریتز-تامپسون تعداد افراد بر حسب وضع سواد و توانایی صحبت کردن به زبان فارسی به تفکیک جنس ..... ۳۴
- جدول ۳-۱۵- برآورد تعدیل شده تعداد افراد بر حسب سن و توانایی صحبت کردن به زبان فارسی به تفکیک جنس ..... ۳۴
- جدول ۳-۱۶- برآورد تعدیل شده تعداد افراد بر حسب وضع سواد و توانایی صحبت کردن به زبان فارسی به تفکیک جنس ..... ۳۴

- جدول ۱-۵- متغیرهای مورد مطالعه برای تولید جامعه همگذاشتی ..... ۴۶
- جدول ۲-۵- مقایسه نرخ بیکاری به روش همگذاشتی، روش هورتز-تامپسون و نرخ واقعی بیکاری سال ۱۳۹۵ ..... ۴۷
- جدول ۳-۵- مقایسه نرخ بیکاری به روش همگذاشتی، روش هوریتز-تامپسون و نرخ واقعی بیکاری براساس ۲ درصد خام از داده‌های سرشماری سال ۱۳۹۵ ..... ۴۸
- جدول ۴-۵- مقایسه نرخ بیکاری سال ۱۳۹۵ در استان‌ها به روش‌های مختلف ..... ۵۱
- جدول ۵-۵- مقایسه نرخ بیکاری در فصل بهار سال ۱۳۹۵ در استان‌ها به روش‌های مختلف ..... ۵۲
- جدول ۶-۵- مقایسه نرخ بیکاری در فصل تابستان سال ۱۳۹۵ در استان‌ها به روش‌های مختلف ..... ۵۳
- جدول ۷-۵- مقایسه نرخ بیکاری در فصل پاییز سال ۱۳۹۵ در استان‌ها به روش‌های مختلف ..... ۵۴
- جدول ۸-۵- مقایسه نرخ بیکاری در فصل زمستان سال ۱۳۹۵ در استان‌ها به روش‌های مختلف ..... ۵۵
- جدول ۹-۵- مقایسه نرخ بیکاری سال ۱۳۹۶ در استان‌ها به روش‌های مختلف ..... ۵۶
- جدول ۱۰-۵- مقایسه نرخ بیکاری در فصل بهار سال ۱۳۹۶ در استان‌ها به روش‌های مختلف ..... ۵۷
- جدول ۱۱-۵- مقایسه نرخ بیکاری در فصل تابستان سال ۱۳۹۶ در استان‌ها به روش‌های مختلف ..... ۵۸
- جدول ۱۲-۵- مقایسه نرخ بیکاری در فصل پاییز سال ۱۳۹۶ در استان‌ها به روش‌های مختلف ..... ۵۹
- جدول ۱۳-۵- مقایسه نرخ بیکاری در فصل زمستان سال ۱۳۹۶ در استان‌ها به روش‌های مختلف ..... ۶۰

# فهرست شکل‌ها

شکل ۱-۲- دقت جمعیت همگذاشتی و برآورد به دست آمده از طرح آمارگیری برای پیشرفت تحصیلی ..... ۲۰

شکل ۱-۳- چگونگی تولید جامعه همگذاشتی در simPop ..... ۳۵

شکل ۱-۵- نحوه توزیع جمعیت به روش هوریتز-تامپسون و جمعیت همگذاشتی ..... ۴۹

شکل ۲-۵- تابع توزیع وضع فعالیت با دو روش هوریتز-تامپسون و جمعیت همگذاشتی ..... ۵۰



# کلیات تحقیق

## ۱-۱- مقدمه

بدون شک داشتن آمار و اطلاعات دقیق پایه و اساس برنامه‌ریزی‌های کلان کشورها است. داشتن اطلاعات دقیق در هر کشوری به عنوان سرمایه‌ای ملی برای آن کشور محسوب می‌شوند. سوالی که به ذهن می‌رسد، این است که این اطلاعات دقیق چگونه جمع‌آوری می‌شوند؟ برای بدست آوردن اطلاعات دقیق، اغلب از جامعه مورد نظر سرشماری و آمارگیری نمونه‌ای انجام می‌شود. سرشماری را به دلیل اینکه به زمان و هزینه زیادی نیازمند است نمی‌توان ماهانه و یا حتی سالانه انجام داد، بلکه بسیاری از کشورها از جمله ایران سرشماری را هر ده سال یکبار انجام می‌دهند. بنابراین اطلاعاتی که از طریق سرشماری بدست می‌آیند به روز نیستند و برای برنامه‌ریزی مناسب نخواهند بود. در مقابل آمارگیری نمونه‌ای به دلیل اینکه به زمان و هزینه کمتری نیازمند است را می‌توان به صورت ماهانه، فصلی و سالانه انجام داد. در واقع براساس آمارگیری نمونه‌ای، به جای جمع‌آوری اطلاعات از کل جامعه مدنظر، از نمونه‌ای که معرف واقعی جامعه است اطلاعات را جمع‌آوری می‌کنند. در طرح‌های آماری مبتنی بر آمارگیری نمونه‌ای باید به اندازه نمونه انتخاب شده دقت داشت. اگر اندازه نمونه ناکافی باشد برآوردهایی که برای پارامترهای جامعه معرفی می‌شوند، از دقت بالایی برخوردار نخواهند بود. در آمارگیری برآوردهایی که تنها مبتنی بر داده‌های نمونه‌ای باشند را برآوردهای مستقیم<sup>۱</sup> می‌نامند. هرچه اندازه نمونه بزرگ‌تر باشد برآوردهای مستقیم کاراتر و دقیق‌تر هستند. اگر اندازه نمونه یک ناحیه بزرگ باشد به طوری که برآورد مستقیم با دقت بالا را نتیجه دهد به آن ناحیه «ناحیه بزرگ» می‌گویند. یک ناحیه را «کوچک ناحیه»<sup>۲</sup> می‌نامیم اگر برآوردهای مستقیم دقت و کارایی لازم را در برآورد پارامترهای ناحیه نداشته باشند. به عبارت بهتر، کوچک بودن ناحیه بیانگر ناکافی بودن اندازه نمونه‌ای برای محاسبه برآوردهای دقیق است. در

<sup>1</sup> Direct Estimators

<sup>2</sup> Small Area

نمونه‌گیری از یک جامعه هنگامی که اندازه‌ی نمونه‌ای در سطح ملی بهینه می‌شود، ممکن است برای یک کوچک ناحیه بسیار کم یا حتی صفر باشد؛ در نتیجه برآوردهای مستقیم دقت کافی برای برآورد پارامترهای کوچک ناحیه‌ی مورد نظر را ندارند. وجود یک نمونه جامع شامل تمام کوچک ناحیه‌ها بطوری که اندازه‌ی نمونه‌ای مربوط به هر کوچک ناحیه، برای پشتیبانی برآوردهای مستقیم به اندازه‌ی کافی بزرگ باشد، تقریباً غیرممکن است. ساخت جامعه همگذاشتی یک روش مناسب برای برآورد کوچک ناحیه‌ای است که در این طرح به مطالعه آن می‌پردازیم. در این فصل ضمن بیان مسأله و ضرورت اجرای طرح به معرفی کلیات دیگر تحقیق از قبیل تعاریف و مفاهیم کلی می‌پردازیم.

## ۱-۲- بیان مسأله

همانگونه که در بالا گفته شد، یکی از مهمترین مسائلی که پایه و اساس برنامه‌ریزی‌ها می‌باشد، جمع‌آوری اطلاعات است. یکی از مهمترین طرح‌های آماری که نقش بالقوه‌ای در برنامه‌ریزی‌های اقتصادی کشورها دارد طرح آمارگیری نیروی کار است. این آمارها اغلب براساس تعریف‌ها و استانداردهای بین‌المللی و با محاسبه‌ی شاخص‌های شناخته شده به دست می‌آیند. رایج‌ترین منبع اطلاعاتی که برای برآورد شاخص‌های نیروی کار مورد استفاده قرار می‌گیرد اطلاعات حاصل از آمارگیری‌های نمونه‌ای است، که به دلیل اهمیت موضوع، در اغلب کشورها به صورت یک آمارگیری مستمر در دوره‌های زمانی متوالی اجرا می‌شود. طراحی آمارگیری‌های نیروی کار معمولاً به گونه‌ای صورت می‌گیرد که برآوردهای حاصل از اجرای آن در سطح کل کشور و استان‌های مختلف بهینه باشند و خطاهای نسبی برآوردها از مقدار تعیین شده‌ای بیشتر نباشند.

آمارگیری نیروی کار ایران که از سال ۱۳۸۴ با طراحی جدید به صورت فصلی اجرا می‌شود برآوردهای قابل اعتمادی در سطح کل کشور و استان‌ها ارائه می‌کند اما برای برنامه‌ریزی‌های بهتر و اجرای سیاست‌های مناسب‌تر توسط برنامه‌ریزان و سیاست‌گذاران، برآورد شاخص‌های نیروی کار در سطح مناطق جغرافیایی کوچکتر از جمله شهرستان‌ها نیز مورد نیاز است. از آنجایی که آمارگیری نیروی کار ایران برای محاسبه‌ی برآوردهای مورد نظر در سطح کل کشور و استان‌ها بهینه شده است، بکارگیری روش‌های برآورد معمول برای محاسبه‌ی برآوردهای نیروی کار در سطح شهرستان به دلیل نبود تعداد نمونه‌ی کافی در شهرستان‌ها منطقی نیست. در چنین شرایطی استفاده از روش‌های برآورد کوچک‌ناحیه‌ای یکی از راهکارهایی است که امکان محاسبه‌ی برآوردهای مورد نظر با سطح درستی مطلوب را برای مناطق جغرافیایی یا زیرجامعه‌های کوچک فراهم می‌آورد. در این روش‌ها معمولاً با استفاده از اطلاعات کمکی برآوردهایی قابل قبول و با خطایی کمتر از برآوردهای مستقیم به دست می‌آید.

همانگونه که گفته شد، کیفیت داده‌ها در آمارگیری‌ها، اندازه‌ی دقت برآوردها و نیز اعتبار استنباط‌های پس از آمارگیری از مهمترین و اصلی‌ترین موضوعاتی هستند که دغدغه‌ی دست‌اندرکاران آمارگیری و دیگر کارشناسان و آمارشناسان را تشکیل می‌دهند. در این میان هزینه یک آمارگیری همواره متغیری در برابر دقت و کیفیت آن بوده است. از سوی دیگر مشکل بی‌پاسخی، به طور روزافزونی در آمارگیری‌ها افزایش می‌یابد که هم منجر به افزایش هزینه و نیز خطای آمارگیری می‌شود. خطای پوشش نیز از دیگر عواملی است که به دلایل گوناگون و گاه ناگزیر (مانند شرایط جغرافیایی) رخ می‌دهد که همراه با افزایش هزینه، دقت آمارگیری را به شدت کاهش می‌دهد.

محدود بودن پاسخگویان و یا جامعه‌ی آمارگیری نیز مشکل دیگری است که آمارشناسان و آمارگیران اغلب با آن روبرو هستند. از همین رو است که به ویژه در سال‌های اخیر لزوم تغییر در روش‌های آمارگیری و یا جایگزینی آن مورد توجه آمارشناسان بوده است. به عنوان مثال توجه به آمارهای ثبتي و به کارگیری داده‌های ثبتي در آمارگیری‌ها از نخستین و البته مهمترین روش‌هایی است که مورد توجه آمارشناسان بوده و به ویژه در سال‌های اخیر کاربرد و اهمیت بیشتری پیدا کرده است. تمام مشکلاتی که در بالا گفته شد بر اساس روش‌های طرح-مبنا می‌باشد که با استفاده از آمارگیری انجام می‌گیرند. البته امروزه روش‌های سنتی طرح-مبنا، به ویژه به دلیل افزایش بار خطاهای غیر نمونه‌گیری (به ویژه خطای بی‌پاسخی، خطای اندازه‌گیری و خطای پوشش)، به کمک روش‌های پیچیده‌تری تلاش در افزایش دقت و به خصوص کاهش واریانس برآوردهای مورد نظر دارند. از این روی با بهره‌گیری از اطلاعات و متغیرهای کمکی و با مدل‌سازی و یا وزن‌دهی متناسب با شرایط نمونه‌گیری، ابتدا برآوردهای کالبدنی مناسبی را محاسبه کرده و سپس با توجه به ویژگی‌های آماری طرح نمونه‌گیری (مانند امید ریاضی و واریانس طرح نمونه‌گیری) آن را تعدیل می‌کنند. این روش را نمونه‌گیری مدل-یار-مبنا گویند. گاه در عمل شرایطی اتفاق می‌افتد که به دلیل حجم نمونه کوچک، محدودیت نمونه‌گیری ناشی از شرایط جغرافیایی و یا افزایش بسیار زیاد خطای بی‌پاسخی روش‌های طرح-مبنا و مدل-یار مبنا از نظر اصول آماری معنی دار نبوده و کارایی ندارند. در چنین مواردی آمارشناسان روش مدل-مبنا را توصیه کرده‌اند که البته به دلایل فنی و محاسبات پیچیده و نیز دقت پایین، سازمان‌های آماری تا کنون رغبت چندانی برای بکارگیری این روش در فرآیند تولید آمار خود، به جز کاربرد در برآورد ناحیه‌های کوچک، نداشته‌اند.

روش مدل-مبنا به جز کاربرد در موارد معدود مانند آمارگیری‌های کوچک‌ناحیه‌ها، هنوز در دیگر موارد گسترش نیافته و در حال نظریه‌پردازی است. (در حال حاضر بیشتر روش‌های آمارگیری نمونه‌ای طرح-مبنا<sup>۳</sup>، مدل-یار-مبنا<sup>۴</sup> هستند.) برای توضیح بیشتر باید گفت که به طور کلی آمارگیری نمونه‌ای برای بدست آوردن اطلاعات آماری از پارامترهای یک جامعه با مطالعه‌ی موردی نمونه‌ای کوچکتر از آن جامعه طراحی و انجام می‌شود. استنباط آماری برای برآورد کردن و یا پیش‌بینی پارامترهای مورد نظر در جامعه نیز معمولاً به یکی از رویکردهای «طرح-مبنا، مدل-یار-مبنا و مدل-مبنا» صورت می‌گیرد. روش طرح-مبنا، همان روش سنتی و معمول است که بر اساس نظریه نمونه‌گیری و بر پایه‌ی توزیع نمونه‌ای برآوردها، ناشی از طرح نمونه‌گیری، پایه‌ریزی شده است. روش مدل-مبنا با تاکید کمتر بر توزیع نمونه‌ای برآوردها، تلاش دارد با استفاده از گمانه‌زنی یک مدل آماری به برآورد پارامترهای مورد نظر و یا انجام پیش‌بینی‌های لازم بپردازد. در روش مدل-یار-مبنا نیز که بیشتر به روش طرح-مبنا نزدیک است، استنباط‌های آماری با استفاده از اطلاعات کمکی (صفات یا متغیرهای کمکی) دقیق‌تر می‌شوند. معمولاً در این روش برآوردها و پیش‌گویی‌ها با بکارگیری یک مدل رگرسیونی و یا سری زمانی که بر پایه ارتباط میان پارامترهای مورد نظر و متغیرهای کمکی شکل گرفته است به دست می‌آیند. این برآوردها سپس با توجه به طرح نمونه‌گیری و ویژگی‌های توزیع نمونه‌ای تعدیل شده و برای استنباط نهایی درباره پارامترها به کار می‌روند.

<sup>3</sup> Designed-based

<sup>4</sup> Model-assisted based

به دلایل ویژگی‌های آماری خوبی که این‌گونه استنباط‌ها (مدل-مبنا) دارند و افزایش دقت برآوردها با بهره‌گیری از صفات کمکی، کاربرد گسترده‌ای در انجام آمارگیری‌های نمونه‌ای دارند. به ویژه در مواردی که به دلایل هزینه و یا دسترسی به نمونه‌ها امکان انجام نمونه‌گیری به آسانی مهیا نباشد.

از جمله کاربردهای روش‌های مدل-مبنا، ساخت جمعیت همگذاشتی است (تمپل و همکاران (۲۰۱۷)). در واقع برای ساخت جمعیت همگذاشتی از مدل‌های آماری مختلفی مانند رگرسیون لجستیک استفاده می‌شود. به منظور دسترسی به جمعیت آماری از ناحیه‌ی مورد نظر همراه با ویژگی‌های جمعیت‌شناختی آن‌ها (مانند توزیع سنی-جمعیتی و ...) ایده‌ی ایجاد جمعیت‌های همگذاشتی مطرح شده است. جمعیت‌هایی از این نوع در حالت کلی از ترکیب ریز داده‌های حاصل از سرشماری که در دسترس است، ساخته می‌شوند. هنگام ایجاد این جمعیت‌ها سعی بر این است که توزیع و همبستگی میان ویژگی‌های عامل‌ها در جمعیت همگذاشتی مشابه با ریز داده‌های سرشماری بوده و همچنین تعداد عامل‌های عضو هر گروه با داده‌های تجمیع شده‌ی موجود هماهنگ باشد. در بیشتر مواقع دسترسی به ریز داده‌ها و اطلاعات برای بسیاری از محققان به دلایل محرمانگی امکان‌پذیر نیست، مزیت جامعه همگذاشتی به دلیل مصنوعی بودن داده‌ها این است که اصول محرمانگی را به طور کامل رعایت می‌کند و توزیع آن با جامعه واقعی مشابه است. محرمانگی در آمار رسمی از مهمترین و بنیادی‌ترین مفاهیم به شمار می‌رود. اهمیت این کار تا آنجا است که اصل ششم از اصول ده‌گانه بنیادین آمار که با خرد جمعی سازمان‌های مرجع آماری و متخصصان آمار رسمی در سطح جهانی تهیه شده و توسط بخش آمار سازمان ملل در سال ۱۹۹۴ تصویب و ابلاغ شده است، به این امر اختصاص دارد. شایان‌گفتن است که این اصول برای سازمان‌های آماری منشور کاری و حرفه‌ای بوده و به منزله دستورالعمل و راهنمای اصلی تدوین فعالیت‌ها است. در این اصل با تاکید بر لزوم محرمانه نگاه داشتن اطلاعات شخصی واحدهای آماری (اعم از اشخاص حقیقی و حقوقی) بیان می‌شود: داده‌های شامل اطلاعات شخصی پاسخگویان که به منظور تولید اطلاعات آماری گردآوری می‌شوند، محرمانه محسوب شده و باید تنها برای اهداف آماری مورد استفاده قرار گیرند. این موضوع به عنوان یکی از عناوین کلیدی در همه قوانین آماری کشورها اشاره شده است. از جمله در ماده ۷ قانون مرکز آمار ایران مصوب سال ۱۳۵۳ اشاره شده که آمار و اطلاعاتی که ضمن آمارگیری‌های مختلف از افراد و موسسات جمع‌آوری می‌شود محرمانه خواهد بود و جز در تهیه آمارهای کلی و عمومی نباید مورد استفاده قرار گیرد. استفاده و مطالبه و استناد به اطلاعات جمع‌آوری شده از افراد و موسسات به هیچ وجه در مراجع قضایی و اداری و مالیاتی و نظایر آن مجاز نخواهد بود. از سوی دیگر در بسیاری موارد به ویژه در کاربردهای پژوهشی، کاربران و پژوهشگران به شدت نیازمند دسترسی به ریز داده‌ها هستند. با توجه به این موضوع در صورتی که بتوان روشی پیدا کرد که در عین اینکه ریز داده‌ها را در اختیار کاربران قرار داد محرمانگی آن را حفظ کرد گام مهمی در جهت استفاده درست و بهینه از آمار برداشته شده است. بنابراین محققان می‌توانند بدون داشتن دغدغه از افشای اطلاعات، از جامعه همگذاشتی به جای جامعه واقعی برای انجام پژوهش‌های خود استفاده کنند. در این مطالعه قصد داریم که جامعه همگذاشتی را برای چهار شهرستان دهلران، ابهر، اسلام‌آباد غرب و گچساران و همچنین کل استان‌های کشور را بسازیم و نرخ بیکاری سال‌های ۱۳۹۵ و ۱۳۹۶ را براساس این جامعه ساخته شده برآورد کنیم.



### ۱-۳- اهمیت و ضرورت پژوهش

یک مسأله متداول در همه جوامع بشری، پایش و کنترل شاخص‌های مهم اقتصادی-اجتماعی است. اشتغال و بیکاری، از جمله موضوعات اساسی اقتصاد هر کشور است، به گونه‌ای که افزایش و کاهش بیکاری، به عنوان یکی از شاخص‌های توسعه یافتگی جوامع تلقی می‌شود. نرخ بیکاری یکی از شاخص‌هایی است که برای ارزیابی شرایط اقتصادی کشور مورد استفاده قرار می‌گیرد. پایش و اعمال سیاست در خصوص کنترل یا ایجاد تغییر دلخواه در نرخ بیکاری از دغدغه‌های دولتمردان محسوب می‌شود. طرح آمارگیری نیروی کار با هدف برآورد شاخص‌های نیروی کار از جمله نرخ بیکاری به صورت فصلی و سالانه و همچنین تغییرات آن در کل کشور و استان‌ها با روش نمونه‌گیری چرخشی<sup>۵</sup> اجرا می‌شود.

نیروی انسانی یکی از مهمترین عوامل پیشرفت استان‌ها، شهرستان‌ها و مناطق کشور است و کشورها در تلاشند که از این مزیت خود به هر نحو که امکان‌پذیر است استفاده کنند. کشورها نیز با اتکا به توانایی‌های نیروی کار بالقوه و موجود در تلاشند وضعیت اقتصادی خود را بهبود دهند. برای اطلاع از وضعیت نیروی انسانی نیاز به گردآوری اطلاعات نیروی انسانی به روش‌های مختلف است.

همان‌گونه که اشاره شد دشواری روزافزون گردآوری اطلاعات به روش‌های سنتی به دلیل پیچیدگی‌های امروزی جوامع آماری، نیاز به مطالعه برای دگرگونی یا بهنگام‌سازی روش‌های آمارگیری را ضروری کرده است. استفاده از دیگر منابع داده‌ها و مدل‌سازی، از روش‌هایی هستند که می‌توانند به عنوان جایگزین روش‌های آمارگیری به کار رفته و یا به افزایش دقت برآوردها و استنباط‌های ناشی از آمارگیری‌های سنتی یاری رسانند. یکی از کاربردهای برآوردهای کوچک ناحیه‌ای درک بهتر مقدار و تغییرات شاخص‌های جامعه میان مناطق جغرافیایی است، به همین دلیل یکی از شایع‌ترین تقاضاهای سیاست‌گذاران، ساختن برآوردهایی با دقت بالا در کوچک‌ناحیه‌های جغرافیایی است. یکی از طرح‌های آمارگیری که در مرکز آمار ایران انجام می‌شود، طرح آمارگیری نیروی کار است. این طرح به طور فصلی با هدف برآورد شاخص‌های نیروی کار مانند نرخ بیکاری در سطح کل کشور، و در هر استان اجرا می‌شود. اما نیاز تصمیم‌گیران به داشتن برآورد در سطح کوچک‌تر از استان مانند شهرستان، مرکز آمار ایران را بر آن داشته تا با استفاده از روش‌هایی مانند برآورد کوچک ناحیه‌ای، به این تقاضاها پاسخ دهد. محدودیت‌های اعتباری و نیروی انسانی مانع از آن می‌شود که بتوان آمارگیری‌ها را با اندازه‌ی نمونه بزرگتر اجرا کرد به طوری که برآوردها با دقت بالا در سطح شهرستان‌ها قابل ارائه باشد. بنابراین استفاده از روش‌های برآورد کوچک‌ناحیه‌ای می‌تواند راهگشای این مسئله باشد. یکی از روش‌های برآورد کوچک‌ناحیه‌ای، شبیه‌سازی جمعیت کوچک‌نواحی به وسیله ساختن جمعیت همگذاشتی است که در سال‌های اخیر مورد توجه زیادی قرار گرفته است. داده‌های همگذاشتی هم داده‌های مورد نیاز در سطح کوچک‌ناحیه‌ها را فراهم می‌کنند.

همچنین از دیگر ویژگی‌های همگذاشتی این است که داده‌هایی را تولید می‌کند که اصول محرمانگی را رعایت می‌کنند و بنابراین قابلیت انتشار عمومی برای محققان را دارند. همچنین داده‌های همگذاشتی به گونه‌ای ساخته می‌شود که هیچ‌گونه نیازی به ادیت و جانپی و ... ندارد.

<sup>5</sup> Rotation Sampling

براساس مطالعات انجام شده تاکنون در کشور، پژوهشی در زمینه برآورد شاخص‌های طرح نیروی کار در سطح شهرستان به وسیله جامعه همگذاشتی انجام نشده است که این پژوهش بنا دارد تا این کار را انجام داده و به عنوان اولین مطالعه عملی در این زمینه گشایش‌گر این راه باشد.

## ۱-۴- اهداف تفصیلی پژوهش

همان‌گونه که اشاره شد و از عنوان این طرح نیز برمی‌آید مهمترین هدف این پژوهش این است که به طور کلی به این پرسش پاسخ دهد که «آیا می‌توان براساس اطلاعات آمارگیری طرح نیروی کار و اطلاعات کمکی دیگر مانند سرشماری جامعه‌ای مصنوعی ساخت که اصول محرمانگی در آن رعایت شود و برآوردهای دقیقی برای شاخص‌های طرح نیروی کار در کوچک‌ناحیه‌ها بدست آورد؟» با توجه به این هدف کلی و با محوریت آن، انتظار می‌رود در پایان این پژوهش، بتوان به پرسش‌های تفصیلی‌تر زیر نیز پاسخ داد:

- مفاهیم، اصول کلی و روش‌های ساخت جامعه همگذاشتی چه هستند؟
- چه اقدامات و مطالعات دیگری توسط آمارشناسان و یا سازمان‌های آماری جهان در زمینه استفاده از جامعه همگذاشتی انجام شده است؟
- آیا برآوردهای بدست آمده در روش همگذاشتی به برآوردهای مستقیم بدست آمده از طرح آمارگیری نیروی کار در سطح استان‌ها نزدیک هستند؟

## ۱-۵- تعاریف و مفاهیم کلی مورد نیاز

در این بخش برخی از مفاهیم کلی که در این طرح مطرح می‌شوند را معرفی می‌کنیم. مفاهیم تخصصی و فنی مورد نیاز نیز در بخش‌های بعدی معرفی خواهند شد.

### ۱-۵-۱- داده‌های همگذاشتی<sup>۶</sup>

داده‌های همگذاشتی مجموعه‌ای از داده‌های جایگزین شده با داده‌های واقعی هستند که به منظور حفظ محرمانگی و کاهش ریسک افشای اطلاعات شخصی و خصوصی افراد، خانوارها، نهادها و ... به کار می‌روند. این داده‌ها از لحاظ آماری بسیار شبیه به داده‌های واقعی هستند به گونه‌ای که می‌توان از آن‌ها به جای داده‌های واقعی برای انجام تحقیقات و پژوهش‌هایی که با داده‌های محرمانه سر و کار دارند، استفاده کرد. تولید داده‌های همگذاشتی یک روش جایگزینی داده‌های واقعی با داده‌های همگذاشتی است که به منظور حفظ حریم خصوصی به کار می‌روند. داده‌های همگذاشتی به صورت تصادفی و با اعمال محدودیت‌هایی با داده‌های واقعی جایگزین می‌شوند. داده‌های همگذاشتی به طور عمده به سه دسته تقسیم می‌شوند:

**الف. داده‌های کاملاً همگذاشتی<sup>۷</sup>:** این داده‌ها به طور کامل همگذاشتی هستند و شامل هیچ داده‌ی واقعی نیست. مولدهای داده‌های کاملاً همگذاشتی توابع چگالی و ویژگی (متغیر)های واقعی را شناسایی می‌کنند و

<sup>۶</sup> Synthetic Data

<sup>۷</sup> Fully Synthetic Data

پارامترهای این توابع چگالی را برآورد می‌کنند. بنابراین برای هر ویژگی (متغیر)، داده‌های محرمانه به صورت تصادفی از مقادیر برآورد شده توابع چگالی، برآورد می‌شوند. روش جانهی چندگانه و روش بوت استرپ، روش‌های کلاسیکی هستند که برای تولید داده‌های کاملاً همگذاشتی استفاده می‌شوند. از آنجایی که داده‌های منتشر شده به این روش، کاملاً همگذاشتی هستند و شامل اطلاعات اصلی نیستند، این روش در حفظ حریم خصوصی و محرمانگی قوی است.

ب. داده‌های تقریباً همگذاشتی<sup>۸</sup>: این داده‌ها به طور کامل همگذاشتی نیستند بلکه ویژگی‌هایی که حساس‌تر هستند را با داده‌های همگذاشتی جایگزین می‌کند. در واقع در اینجا داده‌هایی همگذاشتی می‌شوند که ریسک افشای آن‌ها بالا است. پوشش مقادیر واقعی با مقادیر همگذاشتی مانع شناسایی مجدد می‌شود، بنابراین محرمانگی در داده‌های منتشر شده حفظ می‌شود. معمولاً از روش‌های جانهی چندگانه و مدل-مبنا برای تولید داده‌های تقریباً همگذاشتی استفاده می‌شود. این روش‌ها برای مقادیر گمشده در داده‌های واقعی نیز مفید هستند. ریسک افشای اطلاعات در این داده‌ها در مقایسه با داده‌های کاملاً همگذاشتی بالاتر است زیرا این داده‌ها، داده‌های واقعی را شامل می‌شوند.

ج. داده‌های همگذاشتی ترکیبی<sup>۹</sup>: این داده‌ها با استفاده از داده‌های اصلی و داده‌های همگذاشتی تولید می‌شوند. برای هر رکوردی از داده‌های واقعی، نزدیکترین رکورد در داده‌های همگذاشتی انتخاب می‌شود و ترکیب این دو، داده‌های همگذاشتی ترکیبی را تشکیل می‌دهند. داده‌های همگذاشتی مزایای داده‌های تقریباً همگذاشتی و کاملاً همگذاشتی را دارد. بنابراین ریسک افشای اطلاعات در این دسته از داده‌ها نسبت به دو دسته قبل کمتر است اما به هزینه و زمان زیادی برای پردازش داده‌ها نیاز است.

در این طرح از روش داده‌های کاملاً همگذاشتی برای تولید جمعیت همگذاشتی استفاده می‌شود.

## ۱-۵-۲- کوچک ناحیه<sup>۱۰</sup>

آمارگیری‌های نمونه‌ای زمان زیادی است که به عنوان وسیله‌ای کم هزینه برای به دست آوردن اطلاعاتی درباره‌ی طیف گسترده‌ای از موضوع‌های مورد علاقه در بازه‌های زمانی متداول، شناخته شده‌اند. این آمارگیری‌ها در عمل برای فراهم آوردن برآوردهایی نه تنها از جامعه‌ی کل مورد نظر بلکه برای انواع گوناگون زیر جامعه‌ها (حوزه‌ها) نیز به طور گسترده به کار می‌روند. حوزه‌ها می‌توانند به وسیله‌ی ناحیه‌های جغرافیایی یا گروه‌های جمعیت شناختی- اجتماعی یا دیگر زیرجامعه‌ها تعریف شوند. مثال‌های حوزه‌ی جغرافیایی (ناحیه) شامل ایالت یا استان، شهرستان، حوزه‌ی شهرداری، منطقه‌ی مدرسه‌ها، منطقه‌ی بیمه‌ی بیکاری، ناحیه کلان شهرها و ناحیه‌ی خدمات تندرستی است. از سوی دیگر، حوزه‌ی جمعیت شناختی- اجتماعی می‌تواند به یک گروه خاص سنی- جنسی- نژادی درون یک ناحیه‌ی جغرافیایی بزرگ، اشاره داشته باشد. بنابراین به طور کلی، به ناحیه‌های جغرافیایی کوچک (استان، شهر، شهرستان، ...) که دامنه‌ی کوچکی از ناحیه‌های جغرافیایی بزرگتر را شامل

<sup>8</sup> Partially Synthetic Data

<sup>9</sup> Hybrid Synthetic Data

<sup>10</sup> Small Area

می‌شوند، کوچک ناحیه می‌گوییم. البته کوچک ناحیه‌ها لزوماً ناحیه‌های جغرافیایی نیستند. کوچک ناحیه می‌تواند ناظر بر خصیصه‌ی خاصی که شامل تعداد نمونه کم است شامل سن، جنسیت و ... باشد. در این پژوهش شهرستان به عنوان یک کوچک ناحیه در نظر گرفته شده است.

### ۱-۵-۳- برآوردهای کوچک ناحیه‌ای

برآوردهای کوچک ناحیه‌ای به دو گروه عمده‌ی زیر تقسیم می‌شوند:

- برآوردهای مستقیم
- برآوردهای نامستقیم<sup>۱۱</sup>

#### ✓ برآوردهای مستقیم

این برآوردها تنها از داده‌های نمونه‌ای مربوط به نواحی موردنظر استفاده می‌کنند و معمولاً در ناحیه‌هایی که اندازه‌ی نمونه‌ای آن‌ها به قدر کافی بزرگ است، به کار می‌روند. برآوردهای مستقیم اصولاً طرح مبنا هستند و از وزن نمونه‌گیری استفاده می‌کنند. این برآوردها در کوچک ناحیه‌ها تنها از نمونه‌های مربوط به همان نواحی استفاده می‌کنند که به دلیل بهینه نبودن تعداد نمونه‌ها دقت کمی دارند. برای افزایش دقت و کاهش انحراف استاندارد در برآورد پارامترهای کوچک ناحیه‌ها از برآوردهای نامستقیم استفاده می‌شود:

#### ✓ برآوردهای نامستقیم

گاهی برای برآورد معتبر پارامتر مورد نظر در یک کوچک ناحیه، به دلیل کوچک بودن اندازه‌ی نمونه‌ای و کافی نبودن مشاهده‌ها یا اطلاعات در آن کوچک ناحیه، باید چند کوچک ناحیه را با هم ادغام کنیم و یا اطلاعات کمکی و اضافی از کوچک ناحیه‌های مشابه، یا از اطلاعات همان کوچک ناحیه در مطالعه‌های گذشته، یا ترکیبی از این دو، گردآوری کنیم و سپس برآورد پارامترها را انجام دهیم. چنین برآوردهای نامستقیم می‌نامند. انواع برآوردهای نامستقیم به شرح زیر هستند:

- برآوردهای نامستقیم حوزه‌ای<sup>۱۲</sup>
- برآوردهای نامستقیم زمانی<sup>۱۳</sup>
- برآوردهای نامستقیم حوزه‌ای و زمانی<sup>۱۴</sup>.

برآوردهای نامستقیم حوزه‌ای از مقدارهای متغیر مورد نظر مربوط به حوزه‌ی دیگر، اما نه از دوره‌ی زمانی دیگر، بهره می‌گیرد. برآوردهای نامستقیم زمانی از مقدارهای مورد نظر مربوط به دوره‌ی زمانی دیگر برای حوزه‌ی مورد نظر، اما نه از حوزه‌ی دیگر، استفاده می‌کند. از سوی دیگر، برآوردهای نامستقیم حوزه‌ای و زمانی از مقدارهای متغیر مورد نظر مربوط به حوزه‌ی دیگر و نیز دوره‌ی زمانی دیگر بهره می‌گیرد.

<sup>11</sup> Indirect Estimators

<sup>12</sup> Domain Indirect Estimators

<sup>13</sup> Time Indirect Estimators

<sup>14</sup> Domain and Time Indirect Estimators