

## سرشماری یا نمونه‌گیری یا شاید هر دو<sup>۱</sup> محمد حسین نجاتیان<sup>۲</sup>

### چکیده

متنی که در ذیل می‌آید، مطالبی است که توسط خانم لین بیلارد<sup>۳</sup> رئیس سابق انجمن آمار ایالات متحده در دفاع از کاربرد روش‌های نمونه‌گیری در جریان سرشماری سال ۲۰۰۰ آن کشور ارائه شده است.

هر چند استفاده از روش‌های نمونه‌گیری در زمینه‌های مختلف مربوط به سرشماری مثل عملیات میدانی، ارزیابی و ارتقای کیفیت و... ممکن است مستلزم پذیرش مقدار معینی از ریسک باشد، اما در عین حال، اگر بر حفظ و تداوم روش‌های موجود پافشاری شود، بی‌گمان هیچگونه بهبودی در روش‌ها به وجود نخواهد آمد. به همین دلیل باید با پذیرش حد معقولی از ریسک به استفاده از فنون نمونه‌گیری در زمینه‌های مختلف سرشماری روی آورد.

به ویژه اگر محاسبات اولیه نشان دهد که پذیرش حجم اندکی از خطاهای نمونه‌گیری می‌تواند موجب کاهش حجم بزرگی از خطاهای غیر نمونه‌گیری شود، هرگز نباید در استفاده از این روش‌ها تردیدی نشان داد.

### مقدمه

<sup>۱</sup> این مقاله در بولتن انجمن آمار ایالات متحده آمریکا، شماره‌ی ۲۵، بهار ۱۹۹۹ منتشر شده است.

<sup>۲</sup> آمارشناس و کارشناس آمارهای جمعیت.

<sup>۳</sup> Lynne Billard

گزیبده مطالب آماری - ۶۰  
سرشماری یا نمونه‌گیری یا شاید هر دو

یکی از چالش‌های عمده که در مورد سرشماری سال ۲۰۰۰ وجود دارد، استفاده از نمونه‌گیری در این سرشماری است. از جمله مطالب قابل بحث در این مورد، باوری است که پاره‌ای از نمایندگان مجلس دارند مبنی بر اینکه در شمارش کامل (اصلاح شده بر پایه خالص کم‌شماری) بیشتر احتمال دارد که معرف بودن نتایج از دست برود تا عدم انجام این اصلاح. پاره‌ای دیگر از نمایندگان بر این باورند که اصلاح نتایج در فرم کم‌شماری‌ها موجب اختصاص دلارهای بیشتری به منطقه انتخابی آنان می‌شود. به همین دلیل بازار شایعات و انتشار مطالب اغواگرانه بسیار گرم شده بود.

در این شرایط جای تعجب نیست که ضعف اطلاعات یا حتی تولید اطلاعات نادرست (حداقل از دید علمی) شیوع یابد، بنابراین، نگارنده، در سال ۱۹۹۶ به عنوان رئیس وقت انجمن آمار آمریکا گروهی را برای بررسی ادعای کنگره تشکیل داد. ادعایی که طی آن اظهار می‌شد آمارشناسان در پاسخ به این سوال که آیا نمونه‌گیری یک روش علمی متعبر است یا نه، به دو دسته (یا همانگونه که ادعا شده بود، به دو دسته مساوی) تقسیم شده‌اند. گروه مزبور، همانگونه که انتظار می‌رفت، نمونه‌گیری را به عنوان یک روش علمی معتبر، مورد تایید قرار داد. در نتیجه‌ی چالش مزبور، هر روز بیشتر رنگ و بوی سیاسی پیدا کرد و بحث‌ها به جای حمله‌ی مستقیم به نمونه‌گیری، به حوزه‌ی قانون اساسی سوق پیدا کرد.<sup>۱</sup> با وجود این، در گزارش رسانه‌ها و نیز در آرای صادره از کنگره اعلام شد که ظاهراً در مورد نمونه‌گیری و برنامه‌های دفتر سرشماری در زمینه‌ی چگونگی استفاده از نمونه‌گیری در سرشماری سال ۲۰۰۰، سوء برداشت‌هایی اساسی وجود دارد.

در نتیجه در جولای سال ۱۹۹۸، از نگارنده دعوت شد که همراه با اوژن اریکسون<sup>۱</sup> و رابرت هیل<sup>۲</sup> در یک مصاحبه مطبوعاتی در باشگاه ملی مطبوعات به تشریح و توجیه

<sup>۱</sup> علت این پدیده الزام دولت به انجام سرشماری (Census) به موجب قانون اساسی آمریکا است. (مترجم)

<sup>۱</sup> Demograher - Statistician

<sup>۲</sup> Geogsapher - Statistician

سرشماری سال ۲۰۰۰ پردازد. وظیفه‌ی مشخص نگارنده در آن مصاحبه مطبوعاتی، تشریح نمونه‌گیری به عنوان یک ابزار علمی و کاربرد اخص آن در سرشماری سال ۲۰۰۰ بود. مقاله حاضر، متن رسمی مطالبی است که در آن موقعیت ارائه شد. این توجیه، بعداً برای اعضا و کارکنان کنگره نیز تکرار شد.

باید تأکید شود که نمونه‌گیری مولفه اصلی علم آمار است، همچنانکه ستاره‌ها مؤلفه اصلی ستاره‌شناسی هستند. علاوه بر این، انواع برنامه‌های نمونه‌گیری که از طرف دفتر سرشماری ایالات متحده برای سرشماری سال ۲۰۰۰ در نظر گرفته شده است، مدت‌ها مورد بحث و بررسی قرار گرفته و اگر به درستی به کار گرفته شود، از نظر علمی انتخاب مناسبی خواهد بود. این طرح‌ها قبلاً نیز در حوزه‌های بزرگی مورد استفاده قرار گرفته است. اما به هر حال استفاده از روش (Capture-recapture) در سرشماری، امر تازه‌ای است. در زمینه‌های عملیاتی، عده معدودی هستند که در استفاده مفید و موثر از این طرح‌ها در زمینه‌های پیشنهادی، تردید دارند. دل مشغولی‌ها بر محور این سوال دور می‌زند که آیا می‌توان سرانجام، خطاهای اندازه‌گیری را به سطوحی پائین‌تر از خطاهای (همیشه موجود) نمونه‌گیری، کاهش داد یا نه؟ به هر حال، اعضای آکادمی ملی علوم، طرح‌ها و برنامه‌های خاصی را که دفتر سرشماری برای سرشماری سال ۲۰۰۰ تدارک دیده بود، مورد تأیید قرار داده است. اما بدون شک چالش‌ها، همچنان ادامه خواهد داشت.

### سرشماری یا نمونه‌گیری

همیشه، پذیرش شرایط موجود - در مقابل تحولات نوین - با آسودگی بیشتری، همراه است. بیشتر ما، در مقابل هرگونه تغییر، حتی تغییراتی که سرانجام منجر به شرایط بهتر خواهد شد، سرسختانه مقاومت می‌کنیم. همین مقاومت در مقابل مسایلی که به درستی قابل فهم نیستند نیز وجود دارد، بدون توجه به این نکته که آنچه به ظاهر ساده و آسان می‌نماید مثل شمارش یک به یک افراد یک جامعه - می‌تواند به مراتب پیچیده‌تر از آن باشد که به طور غریزی احساس می‌شود.

گزییده مطالعه ب آماری - ۶۰  
سرشماری یا نمونه‌گیری یا شاید هر دو

عکس‌العمل‌های عمومی در مقابل اخبار مربوط-انجام اولین جراحی‌های قلب توسط کریستین بارنارد، حدود ۳۰ سال پیش احتمالاً هنوز در خاطره‌ها باقی است. جدا از شگفتی‌های غیر قابل انکاری که در رفتار مردم مشاهده می‌شد، پاسخ عمومی به آن منفی بود و انجام جراحی‌های مزبور را تجویز نمی‌کرد. بسیاری از این مخالفت‌ها، از یک ذهنیت قدیمی سرچشمه می‌گرفت. هرچند که به هر حال مخالفت‌های سنجیده‌ای نیز از جانب گروه‌های کوچک علمی مشاهده می‌شد (اما سرانجام معلوم شد که این مخالفت‌ها نیز بی‌اساس بوده است). هنوز هم، معدود افرادی هستند که تداوم این پیشرفت پزشکی را انکار می‌کنند. در مورد مبارزه علیه سرطان نیز سرنوشت مشابهی وجود دارد که می‌تواند به عنوان قیاس پزشکی دیگری مورد توجه قرار گیرد. اما در عین حال امروزه افراد مبتلا به سرطان، آخرین مداوای پزشکی موجود را می‌طلبند یا در واقع تمنا می‌کنند و هرگز راضی به مداوا شدن با روش‌های مرسوم در ۲۰ یا ۳۰ سال قبل نیستند چه رسد به مسامحه کاری‌های بی‌خطر صدها سال پیش.

### ارتباط این مفاهیم با سرشماری‌های نوین

ایالات متحده آمریکا، از زمانی که اولین سرشماری را انجام داد و طی آن حدود ۴ میلیون (۳/۹ میلیون) نفر را شمارش کرد، تغییرات بسیاری کرده است. امروز تعداد افراد کشور به پیش از ۲۵۰ میلیون نفر رسیده و مساحت آن چهار برابر شده است. در سال ۱۷۹۰ کشور عمدتاً از یک جامعه روستایی تشکیل می‌شد، ولی امروز غالباً شهری است. عمده‌ترین تحول شاید در زمینه حمل و نقل رخ داده باشد. امروز مردم کشور طی چند ساعت هزاران مایل جابجا می‌شوند. بنابراین شمارش ۴ میلیون نفر که در زمان سرشماری سال ۱۷۹۰ دشواری‌های خود را داشت، امروزه کار سهل و ساده‌ای است. در اولین سرشماری‌ها، ماموران سرشماری برای رسیدن به واحدهای آماری و انجام کار خود، از اسب استفاده می‌کردند. طی دهه‌های بعدی با تغییر شرایط، تحولات بسیاری در جهت دستیابی به صحیح‌ترین اطلاعات رخ داده است. بنابر این، از اسب سواری که در اولین سرشماری‌ها به مردم مراجعه می‌کردند، مسیری بسیار طولانی طی

شده تا ارتشی بزرگ از زنان خانه‌دار جای آنان را بگیرند. سپس با تبدیل شدن زنان خانه‌دار به نیروی کار مزد و حقوق‌بگیر و کاهش تعداد آنان، مسلماً تعداد اندک زنان خانه‌دار تکافوی نیاز این ارتش بزرگ را نمی‌داد و بنابراین فرم‌های سرشماری به آدرس خانوارها پست شد. در آغاز، این پرسشنامه‌ها به موقع برگشت داده می‌شدند. اما در سرشماری سال ۲۰۰۰ برآوردهای مربوط به میزان بازگشت (R.R) پرسشنامه‌های پست شده، حول ۵۵ درصد است. بنابر این، بازهم نیاز به روش‌های جدید، احساس می‌شود. برای سرشماری سال ۲۰۰۰ هم ارسال پستی پرسشنامه و هم برقراری ارتباط تلقی با پاسخگویان، همراه با یک بسیج تبلیغاتی و جلب مشارکت گروه‌های اجتماعی، طراحی شده است. تا خانوارها را به عودت پرسشنامه‌ها تشویق و ترغیب کنند.

طی این سال‌ها، گونه‌های دیگری از تحول نیز پشت‌سر گذاشته شده است. به عنوان مثال در سال ۱۸۹۰ دستگاه شمارش‌گر Hollerith (نوعی رایانه قدیمی) برای اولین بار به خدمت گرفته شد و جایگزین روش نامطمئن شمارش دستی پرسشنامه‌ها گردید. اطلاعات مربوط به سن، درآمد، و نژاد افراد در گروه‌های معین همراه با سایر ویژگی‌ها استخراج شد. در سال ۲۰۰۰ برای اولین بار مقدمات لازم برای شمارش افراد دارای نژاد مختلط نه به صورت فردی بلکه در گروه‌بندی‌هایی، تدارک دیده شده است. پرسش‌های دیگری نیز اضافه شده است که تماماً در جهت پاسخگویی به نیازهای اطلاعاتی در سطح منطقه‌ای و ملی است. روش‌های نوینی برای تهیه فهرست آدرس‌ها و تضمین حداکثر صحت و به‌هنگام بودن آنها، پیش‌بینی شده است. کاربرد فن‌آوری‌های نوین نظیر اسکنرهای چشمی (Optical scanner) برنامه‌ریزی شده است. با مشاهده‌ی وسعت و دامنه‌ی این تحولات، هیچ‌یک از ما، از انجام سرشماری به همان روش‌های منسوخ شده قدیمی دفاع نمی‌کنیم.

هر یک از این تغییرات بر پایه این شناخت حادث شده است که روش یا روش‌های موجود به دلایلی غیر قابل اجرا شده‌اند. این دلایل غالباً با صحت و/یا توانایی گردآوری فرم‌های تکمیل شده در ارتباط است. هر چند روش‌های جدید نیز می‌تواند با منابع جدیدی از خطا همراه باشد، اما باور بر این بوده است که کسب دانش بیشتر در مورد منابع قبلی خطا و نیز با افزایش سودمندی حاصل، قابل جبران خواهد بود. فکر کنید

گزیبده مطالاب آماری - ۶۰  
سرشماری یا نمونه‌گیری یا شاید هر دو

اگر قرار بود امروزه پرسشنامه‌های دریافتی بدون استفاده از ماشین‌های شمارش‌گر Hollerith و اخلاف آن و به صورت دستی شمارش می‌شد، چه مدت وقت و چه منابعی از نیروی انسانی مورد نیاز می‌بود و چه حجم از عدم صحت و دقت به طور اجتناب‌ناپذیری ایجاد می‌شد. البته ماشین‌های Hollerith نیز خطاهای خاص خود را به وجود می‌آورد- خطای پردازش. در نتیجه، برگزارکنندگان سرشماری‌های مزبور، روشی برای تعیین و مقابله با این‌گونه خطاها اتخاذ کردند. نتیجه‌ی این کنش‌ها، پیشرفت بسیار زیادی بود که در آن زمان در جهان به وجود آمد.

در عین حال که پاره‌ای از پیشرفت‌ها معلول اتفاقات تصادفی بوده‌اند، اما بیشتر آن‌ها در نتیجه‌ی پژوهش‌های زیاد و جامع به ثمر نشسته‌اند. مانند پژوهش‌هایی که در زمینه جامعه‌شناسی، جغرافیا، جمعیت‌شناسی، علوم رایانه‌ای، آمار، علوم سیاسی و اقتصاد و در واقع در هریک از بسیاری رشته‌های علوم که در اجرای سرشماری مؤثرند، به عمل آمده است. شناخت کم‌شماری‌ها - وجود آن در هر مورد و نیز ویژگی‌های افتراقی آن (برحسب منطقه و نژاد) - یکی از آن رخداد‌های تصادفی است. این پدیده زمانی کشف شد که تعداد مردان ثبت نام‌کننده برای اعزام به جبهه‌های جنگ جهانی دوم از تعداد مردان شمارش شده در سرشماری در همان زمان، فزونی یافت. بررسی‌های انجام‌شده روشن ساخت که در واقع این موضوع یک پدیده واقعی است و به‌دنبال آن پژوهش‌های بیشتری که در سال‌های اخیر در همین زمینه به عمل آمد، به ارائه روش‌هایی برای تصحیح ارقام حاصل از سرشماری‌ها و رفع اثر کم‌شماری‌ها و نیز مهم‌تر از همه، اجرای بهتر سرشماری منجر شد.

اکنون که سرشماری سال ۲۰۰۰ را پیش رو داریم، در کجای این مسیر تکاملی قرار گرفته‌ایم؟ اجازه دهید فقط یک جنبه از سرشماری را در نظر بگیریم. چگونه می‌توان یک شمارش کامل را صحیح (حتی‌الامکان) توصیف کرد؟

در این سوال دوجزء متمایز وجود دارد. بی‌پاسخ‌ها و موضوعات کم‌شماری شده. ابتدا، یادآوری می‌شود که ۱۰۰ درصد خانوارها، پرسشنامه‌های سرشماری را دریافت می‌دارند. سپس در نقطه‌ای از زمان، تعدادی از خانوارها پرسشنامه‌های تکمیل شده را

سرشماری یا نمونه‌گیری یا شاید هر دو  
گزیده مطالب آماری- ۶۱

عودت خواهند داد- که اصطلاحاً آنرا " شمارش آغازین " می‌خوانیم. سپس، در مرحله پی‌گیری، تعداد دیگری از خانوارها، پرسشنامه‌های تکمیل شده را ارسال می‌دارند که البته شامل خانوارهایی که پرسشنامه خود را با تأخیر عودت داده‌اند نیز می‌شود. تعدادی از افراد پرسشنامه‌های خود را به صورت تکراری ارسال می‌دارند. بنابر این به صورت تصویری خواهیم داشت:

شمارش آغازین	شمارش به دنبال پی‌گیری بی‌پاسخ‌ها	دوباره شماری‌ها	از قلم افتاده‌ها
--------------	--------------------------------------	--------------------	------------------

این پدیده را می‌توان به صورت ذیل نیز تصور کرد:

از قلم افتاده‌ها	موارد شناخته شده
------------------	------------------

کل جامعه

در اینجا، تعداد کل موارد شناخته شده، مرکب از تعداد افرادی است که پرسشنامه‌های خود را عودت داده‌اند به اضافه‌ی افرادی از بدون پاسخ‌ها که پس از پی‌گیری، پرسشنامه خود را ارسال کرده‌اند، البته پس از تعدیل به خاطر دوباره‌شماری‌ها.

در این مورد، اولین مطلب، دستیابی به اندازه درست، شمارش به دنبال پی‌گیری بی‌پاسخ‌ها با استفاده از یک روش با بهره‌وری هزینه‌ای بالا، است. بی‌پاسخ‌ها، برخلاف از قلم افتاده‌ها، معرف افرادی است که پرسشنامه خود را عودت نداده‌اند، اما با صرف وقت و کوشش قابل حصول هستند و بنابر این، قابل سرشماری خواهند بود. در سرشماری سال ۱۹۹۰، حدود یک‌سوم جامعه بی‌پاسخ بودند. متأسفانه، فرایند پی‌گیری بسیار گران است، چرا که برای انجام آن، لازم است پرسشگرانی استخدام شوند و آموزش ببینند.

گزییده مطالعه مطلب آماری - ۶۰  
سرشماری یا نمونه‌گیری یا شاید هر دو

علاوه بر این انجام فرایند مزبور چندماه به طول می‌انجامد، و نتایج حاصل از این فرایند نیز دارای این خاصیت است که صحت در ارقام به دست آمده، به طور فزاینده‌ای کاهش می‌یابد. به عنوان مثال در سرشماری ۱۹۹۰، خطای اندازه‌گیری در مورد پرسشنامه‌های دریافت شده در اثر پی‌گیری‌های انجام شده در ماه مه، ۶/۶ درصد و برای ماه اوت - دسامبر برابر ۲۸/۴ درصد برآورد شده بود. به همین دلیل حجم خطای اندازه‌گیری یکی از عوامل مورد توجه است. در نتیجه، این سوال مطرح می‌شود که آیا می‌توان تعداد بی‌پاسخ‌ها را با صحت بیشتری اندازه‌گیری کرد؟

متأسفانه، به هر حال پاره‌ای از افراد سرشماری نشده باقی خواهند ماند. تعداد حقیقی افراد جامعه برابر حاصل جمع تعداد افراد شناسایی شده و تعداد افراد شناخته نشده و در نتیجه، شمارش نشده است. مسلماً اگر تمامی افراد پاسخ‌دهند به طوری که کل افراد جامعه مورد سرشماری قرار گیرند، کم‌شماری وجود نخواهد داشت. [موقتاً فرض می‌شود که تمامی پاسخ‌های دریافتی صحیح است، یعنی خطای اندازه‌گیری وجود ندارد]. اما واقعیت آنست که حتی اگر منابع نامحدود در اختیار باشد و حداکثر کوشش نیز به عمل آید، در جهان امروز، شمارش تمام افراد، عملاً غیر ممکن است. حتی کشورهایایی که در طی عملیات میدانی سرشماری قانون منع عبور و مرور را به جامعه تحمیل می‌کنند، آن‌را "ناکارا" یافته‌اند. حتی این روش نیز منجر به پوشش صحیح نخواهد شد، هرچند که از نظر مفهومی باید نتیجه بخش باشد. (اگر به گذشته بازگردیم فقدان آزادی‌ها زمینه‌ساز تحمیل منع عبور و مرور می‌شد). بنا بر این سوال دوم این است که؛ چگونه می‌توان تعداد افراد از قلم افتاده را تا آنجا که ممکن است به طور صحیح تعیین کرد؟

همان‌گونه که در گذشته، روش‌هایی در زمینه‌ی مطالب بحرانی مربوط به سرشماری‌ها مورد پژوهش قرار گرفته‌است، در حال حاضر نیز روش‌های علمی مناسبی عرضه شده است که امکان می‌دهد اندازه دقیقتری از تعداد بی‌پاسخ‌ها را به دست آورد و نیز به بهترین اندازه در مورد از قلم افتاده‌ها، دست یافت. در این مورد، بهترین روش علمی برای دستیابی به این مقادیر، علم آمار به ویژه نمونه‌گیری است.

قبل از پرداختن به مباحث نمونه‌گیری مطروحه در سرشماری سال ۲۰۰۰، پاره‌ای اصول پایه‌ای به صورت گذرا مورد توجه قرار خواهد گرفت. ابتدا به این سوال می‌پردازیم که چرا نمونه؟ و این که نمونه‌گیری چیست؟

### چرا نمونه؟

الف- پاره‌ای از داده‌ها غیر قابل دسترس هستند.

ب - گاهی اوقات جامعه به گونه‌ای است که شمارش کامل افراد جامعه عملی است، اما در عین حال:

۱ - فرایند مشاهده خود به خود موجب از بین بردن افراد جامعه خواهد شد. به عنوان مثال در اندازه‌گیری طول عمر یا عملکرد اشیایی نظیر فیوزها، کبریت‌ها، حتی اگر امکان آزمایش تمامی اعضای جامعه وجود داشته باشد، پس از انجام آزمایش، چیزی برای استفاده یا فروش باقی نخواهد ماند.

۲ - نمونه‌گیری دارای مزایای اقتصادی است مثل مواردی که در آن‌ها زمان و منابع محدود هستند. در این حالت لازم است بین اطلاعاتی که به دست می‌آید و هزینه‌ی مربوط به آن، تعادلی ایجاد شود.

۳ - اندازه‌گیری‌های فردی، در یک تمام شماری ممکن است با دقت کمتری همراه باشد تا در یک نمونه‌گیری، به عنوان مثال، گردآوری اطلاعاتی که پاسخگویان نسبت به آن اطلاعات حساسیت دارند دارای خطاهای اندازه‌گیری بزرگی است که خارج از حد انتظار خواهد بود. برای نشان دادن این پدیده در یک مطالعه بزرگ کلاسیک در مورد اطلاع مربوط سن افراد، (در فیلادلفیا)، وقتی آمارگیران پس از گذشت ۸ تا ۱۰ روز به پاسخگویانی که خود با آنها مصاحبه کرده بودند مراجعت کردند و مجدداً سن آنان را مورد پرسش قرار دادند، بیش از ۱۰ درصد پاسخگویان سن خود را یک سال یا بیشتر از آنچه که ۸ تا ۱۰ روز قبل اظهار داشته بودند، اعلام می‌کردند. (و وقتی در نوبت دوم پرسشگرانی بجز پرسشگران اولیه مورد استفاده قرار گرفتند

گزیده مطالعه ب آماري - ۶۰  
سرشماری یا نمونه‌گیری یا شاید هر دو

تفاوت مزبور به بیش از ۱۷ درصد رسید<sup>۱</sup>). انجام مصاحبه و کسب اطلاع در مورد ویژگی‌های اشتغال یا درآمد نیز نوعاً با مقادیر بزرگی از خطاهای اندازه‌گیری همراه است. در عین حال انتظار می‌رود که پاره‌ای از این خطاها از طریق آموزش مناسبتر پرسشگران تحت کنترل درآید. ناگفته پیداست که هر قدر تعداد پرسشگران بیشتر باشد خطاهای اندازه‌گیری نیز افزایش خواهد یافت. بنابراین صحیح‌ترین اطلاعات در واقع از طریق انتخاب علمی یک نمونه و استفاده از تعداد کمتری پرسشگر با کیفیت مناسب به دست خواهد آمد.

ازین رو در محتوای یک سرشماری افراد از قلم افتاده، در طبقه‌بندی فوق در گروه الف قرار می‌گیرند و بی‌پاسخ‌ها اگر محدودیتی در منابع نباشد در بند (۳) از طبقه ب و در صورت محدود بودن منابع در بندهای (۲) و (۳) از طبقه ب قرار می‌گیرند.

### نمونه‌گیری چیست؟

نمونه‌گیری فرایند انتخاب افراد از جامعه است به گونه‌ای که اطلاعات کسب شده از نمونه، انعکاس درستی از اطلاعات متناظر در کل جامعه باشد.

فرایند نمونه‌گیری شامل تهیه طرح نمونه‌گیری نیز هست. طرح نمونه‌گیری بسته به ویژگیهای اصلی جامعه مورد مطالعه و نیز نوع اطلاعات مورد نظر برای آمارگیری تغییر می‌کند. در این مورد امکانات بسیاری وجود دارد. اما در عین حال یک طرح نمونه‌گیری خوب، واجد پاره‌ای ویژگی‌های عمومی خواهد بود. یک جنبه پایه‌ای آن است که هر واحد جامعه یک شانس معین برای انتخاب شدن دارد. تعیین حجم نمونه، مسلماً از مؤلفه‌های اساسی و مهم است. در بین سایر عوامل، میزان صحت مورد نظر، حدود تغییرات یا میزان انحراف از مقدار حقیقی (ولی نامعلوم) صفت مورد مطالعه در کل

<sup>۱</sup> نگاه کنید به پالمرجی. ال ۱۹۴۳ عوامل مؤثر در تغییر پذیر پاسخگوها در آمارگیری‌ها ژورنال انجمن آمار آمریکا، (۱۴۳، ۳۸-۱۵۲)

سرشماری یا نمونه‌گیری یا شاید هر دو  
گزیده مطالب آماری - ۶۱

جامعه و نیز معیاری از واریانس آن در جامعه باید مورد توجه قرار گیرد. بنابراین اگر در انتخاب نمونه از یک جامعه، معیارهای مربوط، به یک طرح نمونه‌گیری مناسب رعایت شده‌باشد، آماره‌های محاسبه شده بر پایه نتایج حاصل از نمونه (مثل میانگین نمونه‌ای) نیز برآورد مناسبی از آماره متناظر در کل جامعه خواهد بود.

برای روشن شدن مطلب، فرض کنید در یک شرایط معین (که در آن، جامعه مورد مطالعه از نوعی یک‌نواختی منطقی برخوردار است)، نمونه‌گیری تصادفی ساده، مناسب‌ترین طرح نمونه‌گیری تشخیص داده شده است. در پیوست شماره ۱، فرمول مربوط به محاسبه‌ی حجم نمونه لازم برای برآورد میانگین جامعه ارائه شده است. بر مبنای محاسبات انجام شده که نتایج آن در جداول ضمیمه آمده است، ملاحظه می‌شود. برای آن که مقدار نمونه‌ای در فاصله ۵ واحد ( $\beta=5$ ) از مقدار حقیقی جامعه در سطح ۹۵٪ دقت باشد، حجم نمونه  $n$ ، برحسب حجم جامعه  $N$  به شرح ذیل خواهد بود.

حجم جامعه $N$	۱۰۰۰۰	۲۰۰۰۰	۵۰۰۰۰	۱۰۰۰۰۰
حجم نمونه $n$	۱۵/۳	۱۵/۴	۱۵/۴	۱۵/۴ (۱۶)

به عبارت دیگر حجم نمونه مورد نیاز برابر ۱۶ فرد است. لازم به ذکر است که در مقایسه با جوامع واقعی، جامعه مفروض بسیار کوچک است، باید توجه داشت که علی‌رغم افزایش شدید و سریع حجم جامعه، حجم نمونه چندان افزایشی ندارد. در عین حال اگر در جامعه‌ای با حجم  $N=100000$ ، سطح دقت افزایش داده شود، ارقام ذیل برای حجم نمونه  $n$  به دست خواهد آمد.

دقت	٪۹۵	٪۹۹	٪۹۹/۹
-----	-----	-----	-------

گزیده مطالعه مطالعه آماری - ۶۰  
 سرشماری یا نمونه‌گیری یا شاید هر دو

۳۷/۲(۳۸)	۲۶/۵(۲۷)	۱۵/۴(۱۶)	حجم نمونه n
----------	----------	----------	-------------

بدین ترتیب، برای نیل به دقت ۹۹/۹٪ حجم نمونه مورد نیاز برابر ۳۸ واحد است. [ مفهوم دقت ۹۹/۹٪ آن است که اگر ۱۰۰۰ نمونه انتخاب شود، در ۹۹۹ نمونه آماره محاسبه شده از نمونه در فاصله  $\beta$  واحد از مقدار حقیقی آن در جامعه است و تنها در یک نمونه، خارج از این فاصله خواهد بود. البته، عملاً فقط یک نمونه گرفته می‌شود. مطلوبیت و تمایز فوق‌العاده‌ای که این نمونه دارد آن است که یکی از ۹۹۹ نمونه باشد. ] این که حجم نمونه به مراتب کوچک‌تر از آن است که قبل از انجام محاسبات، انتظار می‌رفت، نه فقط شگفت‌آور و تعجب‌برانگیز است، بلکه شاید غیرقابل باور هم باشد. بهتر است به دو مثال ذیل توجه شود: اول، یک آشپز که در حال تهیه یک ظرف سوپ است، یک قاشق از آن را می‌چشد تا بداند آیا نیازی به نمک بیشتر دارد یا نه. اگر سوپ داخل ظرف را کاملاً هم زده باشند، فارغ از بزرگی ظرف سوپ، آشپز مطمئن خواهد بود که همان یک قاشق معیار دقیقی از میزان شوری یا بی‌نمکی سوپ به دست خواهد داد. دوم، یک مورد کاملاً غیر عادی را نظر بگیرید. فرض کنید در جامعه‌ای که تمامی افراد آن دارای درآمد یکسان هستند، در نظر است مطالعه‌ای در مورد درآمد افراد انجام شود. اگر درآمد فقط یک نفر از جامعه معین شود، بله نمونه‌ای با حجم یک نفر، تمامی اطلاعات مورد نظر در مورد درآمد افراد جامعه مزبور روشن خواهد شد. اما، غالباً بین افراد جامعه، تفاوت‌هایی وجود خواهد داشت، همین تفاوت‌ها ضرورت توجه به واریانس جامعه را در تهیه طرح نمونه‌گیری الزام می‌بخشد.

واریانس، به اشکال مختلف ظهور می‌یابد. یک دریاچه را در نظر بگیرید که در آن فقط دو نوع ماهی، کوچک و بزرگ وجود دارد. یعنی پاره‌ای از آنها بسیار کوچک و پاره‌ای دیگر بسیار بزرگ. فرض کنید یک نمونه با حجم ۳ ماهی از این دریاچه گرفته شود، اگر اتفاقاً هر ۳ ماهی از نوع بزرگ یا از نوع کوچک باشد، متوسط وزن ماهی‌های نمونه به هیچ‌وجه منعکس کننده متوسط وزن ماهی‌های دریاچه نخواهد بود. در این

حالت، یک طرح نمونه‌گیری بهتر، طرحی است که در آن ماهی‌ها به دو طبقه، تقسیم و از هر طبقه تعدادی ماهی نمونه انتخاب شود. یا آن که، به جای ماهی‌های کوچک و بزرگ، ممکن است در دریاچه، ماهی‌های سفید و راه‌راه وجود داشته باشد، در این حالت می‌توان ماهی‌ها را بر حسب رنگ طبقه‌بندی کرد. به طور کلی، طبقات متعدد و متفاوتی وجود دارد و بخشی از طرح نمونه‌گیری انتخاب طبقه‌بندی مناسب است.

یک طرح نمونه‌گیری تصادفی ساده، در واقع ساده‌ترین طرح نمونه‌گیری است. اما طرح‌هایی که برای سرشماری طراحی شده یا در سرشماری مورد استفاده قرار گرفته است، بسیار پیچیده‌تر است. با وجود این، اصول و مبانی پایه‌ای همان است که عنوان شد. به طور کلی، اینگونه طرح‌های پیچیده‌تر، دارای سودمندی بیشتری است تا طرح‌های تصادفی ساده.

در فرایند سرشماری، طرح‌های طبقه‌بندی می‌تواند برای برآورد تعداد بی‌پاسخ‌ها، مورد استفاده قرار گیرد تا نیازی به جستجو برای یافتن هریک از این واحدها نباشد و به جای این کار پرسشگران به تعداد اندکی از خانوارهای بی‌پاسخ مراجعه کنند. این فرایند، چون به تعداد کمتری پرسش‌گر نیاز دارد، بهره‌بروری بیشتری دارد و از آنجا که در مدت زمانی کمتر و سریعتر به انجام می‌رسد، موجب کاهش خطای اندازه‌گیری نیز خواهد شد. در طرح‌های سرشماری، طبقه‌بندی‌ها معطوف به تغییرات منطقه‌ای یا محلی، یا سن و نژاد خواهد بود.

### کم‌شماری

اندازه‌گیری کم‌شماری را می‌توان از طریق طرح‌های نمونه‌گیری به روش سنتی (Capture-recapture) یا (C.R) به موثرترین صورت، برآورد کرد. ساده‌ترین شکل این گونه طرح‌ها در پیوست شماره ۲ عرضه شده است. به این منظور، قبل از اجرای سرشماری یک نمونه انتخاب می‌شود (برای تضمین استقلال آماری آن). پس از اجرای سرشماری وقتی رقم  $C_1$  مربوط به اطلاع حاصل از سرشماری بدست آمد، می‌توان از طریق مقابله ی یافته‌های سرشماری با داده‌های حاصل از نمونه‌گیری به رقم  $\alpha$  رسید

گزیده مطالعات آماری - ۶۰  
سرشماری یا نمونه‌گیری یا شاید هر دو

که معرف افرادی است که هم در سرشماری و هم در نمونه‌گیری ثبت شده‌اند. براساس این اطلاع، تعداد کل واحدها، شامل واحدهای شمارش شده و واحدهای از قلم افتاده قابل محاسبه خواهد بود. بدین ترتیب برآوردی از واحدهای قلم افتاده بدست خواهد آمد. توجه به این نکته هم خالی از لطف نیست که در این روش، اقلام دیگری نیز قابل برآورد است. یکی از این اقلام، تعداد واحدهایی است که در سرشماری از قلم افتاده ولی در نمونه‌گیری ثبت شده‌اند، و اطلاع دیگر مربوط به تعداد واحدهایی است که هم در سرشماری و هم در نمونه‌گیری از قلم افتاده‌اند.

البته، در عمل، برنامه‌های سرشماری، ضرورتاً، پیچیده‌تر خواهد بود. اما در عین حال اصول و مبانی نظری آن، تغییری نخواهد داشت. بنابراین، به عنوان مثال، ترکیب روش (C.R) با طرح‌های مبتنی بر طبقه‌بندی، امکان می‌دهد که اینگونه اقلام برای هر طبقه و در نتیجه برای کل کشور، محاسبه شود.

اشاره به نکته دیگری نیز ضروری است. فرض کنید که یافته‌های سرشماری و نیز نتایج حاصل از طرح‌های نمونه‌گیری (C.R) در دست باشد. به طوری که گذشت اطلاعات مربوط به پاره‌ای از افراد ( $\alpha$ ) هم در سرشماری و هم در نمونه‌گیری ثبت شده است. در نتیجه، این امکان وجود دارد که دو مجموعه از اطلاعات با یکدیگر مقایسه شده و معلوم شود که آیا ویژگی‌های ثبت شده در دو منبع یکسان است یا نه. مسلماً پاره‌ای از آن‌ها یکسان و پاره‌ای دیگر متفاوت خواهند بود. در مواردی که اطلاعات ثبت شده در دو منبع برای یک واحد معین یکسان نیست، پرسش‌گران می‌توانند با مراجعه مجدد، نسبت به رفع تناقض اقدام کنند. به این ترتیب، اطلاعات نمونه‌ای مزبور، علاوه بر استفاده‌ای که در برآورد کم شماری مربوط به سرشماری دارند، می‌توانند در مورد کنترل کیفیت اطلاعات سرشماری نیز مورد استفاده قرار گیرند. توجه به این نکته نیز لازم است که هرگز فرض نمی‌شود که یافته‌های نمونه‌گیری بهتر از نتایج سرشماری است. تنها فرضی که لازم است وجود استقلال دو منبع است.

**نظر نهایی**

مردان و زنانی که در تاریخ ماندگار می‌شوند، افرادی هستند که تغییرات و تحولات را با آغوش باز می‌پذیرند و در واقع از آن حمایت و پشتیبانی می‌کنند. تحولاتی که منجر به پیشرفت به سوی آینده‌ای بهتر خواهد شد. و بر عکس پاره‌ای از شخصیت‌های تاریخی نیز به خاطر مخالفت با این‌گونه پیشرفت‌ها چهره‌ای نامطلوب و ناپسند دارند، مثل افرادی که دانش جدید عصر خود مبنی بر گرد بودن زمین را بر نمی‌تابیدند و بر مسطح بودن زمین که عقیده رایج عصر آنان بود، پای می‌فشردند.

ما می‌توانیم و باید به این واقعیت ببالیم که دفتر سرشماری و کارکنان آن، با شناخت نمونه‌گیری به عنوان یک روش علمی معتبر، جزو گروهی به شمار می‌آیند که گرد بودن زمین را باور داشتند و با بینه علمی مسطح بودن آنرا رد کردند. اگر به مثال مربوط به بیماری سرطان باز گردیم، بیماران در خواست دارند که با بهترین مداوای پزشکی که علم می‌تواند پیشنهاد کند، مورد معالجه قرار گیرند. با همین قیاس، دفتر سرشماری نیز در جستجوی بهترین‌هایی است که به نظر می‌رسد روش‌های نوین علمی می‌تواند ارائه کند. همان‌گونه که ما آگاهی داریم سرطان هنوز به طور کامل مهار نشده است، هیچ فرد معتبری نیز ادعا نمی‌کند که روش‌های مورد استفاده در سرشماری‌ها به کمال خود رسیده‌اند. پژوهش‌ها، همچنان ادامه دارند تا بهترین روش‌ها را برای اجرای هر چه دقیقتر و هر چه موثرتر سرشماری‌ها بیابند. در عین حال، دفتر سرشماری به اهمیت استفاده از روش‌های علمی روز آمد، بجای روش‌های مربوط به دهه‌های قبل، به خوبی واقف است و ما باید از این بابت سپاسگزار باشیم.

### پیوست شماره ۱:

#### نمونه‌گیری تصادفی ساده

فرض کنید برآورد یک میانگین مجهول در یک جامعه نظیر متوسط مصرف روزانه آب یک شهر، مورد نظر باشد. تعداد کل افراد جامعه معین و برابر  $N$  است (مثلاً ۱۰۱۰۰ نفر). همچنین فرض کنید اطلاعات موجود از گذشته حاکی است که ۹۵ درصد

گزیده مطالعه مطالعه آماری - ۶۰  
 سرشماری یا نمونه‌گیری یا شاید هر دو

مصرف آب افراد این جامعه در فاصله ۲۰ واحدی از میانگین کلی جامعه است (یعنی سنجه پراکندگی آن  $\sigma=10$  یا  $\sigma^2=100$  است).

علاوه بر این، فرض کنید که در نظر است برآورد مزبور دارای دامنه تغییرات  $\beta$  واحد در حول میانگین حقیقی جامعه و میزان دقت  $\alpha\%$  باشد (که در ادبیات آمار با علامت  $Z_\alpha$  نشان داده می‌شود) در این شرایط، تعداد نمونه  $n$  برای حصول به شرایط فوق عبارت خواهد بود از:

$$n = \frac{N\sigma^2}{(N-1)\beta^2 / Z_\alpha^2 + \sigma^2}$$

جدول ۱ و ۲، نتیجه محاسبات مربوط تعداد نمونه  $n$  را برحسب مقادیر مختلف  $N$  با فرض اینکه برآورد میانگین مورد نظر در فاصله  $\beta$  واحد از مقدار حقیقی میانگین جامعه قرار داشته باشد (مثلاً  $\beta=5$ ) و دقت  $95\%$  نشان می‌دهد. (برای دقت  $95\%$  درصد،  $Z_\alpha=1/96$  و برای دقت  $99\%$  درصد،  $Z_\alpha=2/576$ ). جدول ۳ و ۴، نتایج محاسبات مشابهی را برای  $\beta=2$  به دست می‌دهد.

**جدول ۱- تعداد نمونه  $n$  با  $\beta=5$  با دقت  $95\%$  درصد**

حجم جامعه $N$	۱۰۰۰۰	۱۵۰۰۰	۲۰۰۰۰	۵۰۰۰۰	۱۰۰۰۰۰
حجم نمونه $n$	۱۵/۳	۱۵/۴	۱۵/۴	۱۵/۴	۱۵/۴

**جدول ۲- برآورد تعداد نمونه  $n$  با  $\beta=5$  با دقت  $99\%$  درصد**

حجم جامعه $N$	۱۰۰۰۰	۱۵۰۰۰	۲۰۰۰۰	۵۰۰۰۰	۱۰۰۰۰۰
حجم نمونه $n$	۲۶/۵	۲۶/۵	۲۶/۵	۲۶/۵	۲۶/۵

جدول ۳- برآورد تعداد نمونه  $n$  با  $\beta=2$  با دقت ۹۵ درصد

حجم جامعه N	۱۰۰۰۰	۱۵۰۰۰	۲۰۰۰۰	۵۰۰۰۰	۱۰۰۰۰۰
حجم نمونه n	۹۱/۱	۹۵/۴	۹۵/۶	۹۵/۹	۹۵/۹

جدول ۴- برآورد تعداد نمونه  $n$  با  $\beta=2$  با دقت ۹۹ درصد

حجم جامعه N	۱۰۰۰۰	۱۵۰۰۰	۲۰۰۰۰	۵۰۰۰۰	۱۰۰۰۰۰
حجم نمونه n	۱۶۳/۲	۱۶۴/۱	۱۶۴/۵	۱۶۵/۳	۱۶۵/۶

## پیوست ۲

### محاسبه کم‌شماری‌ها

فرض کنید در نظر است تعداد پرندگان در یک منطقه معین (N) اندازه‌گیری شود. بررسی‌های انجام شده، منجر به انتخاب طرح مرسوم به (C.R) گردیده است. در اجرای این روش، فرض کنید که ابتدا ۲۰۰ پرنده (نمونه) مورد بررسی قرار گرفته، علامت‌گذاری و رها سازی شده‌اند. سپس یک نمونه مستقل ۱۰۰ عضوی از پرندگان مورد بررسی قرار گرفته‌است. مقایسه دو نمونه نشان داده‌است که از بین ۱۰۰ پرنده نمونه دوم، ۸۰ پرنده دارای علامت و ۲۰ پرنده فاقد علامت بوده‌اند، یعنی ۸۰ درصد از پرندگان گرفته شده در مرحله دوم در بررسی اولی نیز وجود داشته‌اند.

جمع	نمونه‌گیری دوم		
	مشاهده نشده در نمونه اول	مشاهده شده در نمونه اول	
۲۰۰	۱۲۰	۸۰	علامت‌گذاری شده نمونه‌گیری اول
۵۰	۳۰	۲۰	علامت‌گذاری نشده

گزیبده مطالعه ب آماری - ۶۰  
 سرشماری یا نمونه‌گیری یا شاید هر دو

۱۵۰	۱۰۰	جمع
-----	-----	-----

براساس همین اطلاعات- اگر به درستی گردآوری شده باشد. می‌توان برآوردی از کل تعداد پرندگان به دست داد. کافی است فرض شود که میزان پوشش در آمارگیری اول نیز ۸۰ درصد است و به عبارت دیگر نمونه اول برابر ۸۰ درصد کل پرندگان است.

$$N \times \frac{80}{100} = 200 \quad \text{یعنی}$$

$$N = \frac{100}{80} \times 200 = 250 \quad \text{و بنابر این}$$

از ۲۵۰ پرنده موجود در منطقه، در واقع ۸۰ پرنده در هر دو نمونه‌گیری مشاهده شده‌اند، ۱۲۰ پرنده فقط در نوبت اول مشاهده و در نوبت دوم مشاهده نشده‌اند ۲۰ پرنده فقط در نوبت دوم مشاهده شده و در نوبت اول به نمونه اصابت نکرده‌اند. بنابر این جمعاً ۲۲۰ پرنده (۲۰+۲۰۰=۲۲۰) مورد بررسی قرار گرفته است و ۳۰ پرنده (۲۵۰-۲۲۰=۳۰) در هیچ‌یک از دو نوبت، مورد مطالعه قرار نگرفته است. توجه این نکته نیز لازم است که

$$N=250=(200 \times 100)/80 \quad \text{و نیز} \quad 30=20 \times \frac{120}{80}$$

در مورد سرشماری، دور اول کسب اطلاع در واقع در حین عملیات میدانی سرشماری و مراجعه به خانوارها، انجام خواهد شد. و مثلاً  $C_1$  مشاهده متناظر با پرندگان علامت‌گذاری شده در نوبت اول است. صرفاً به خاطر ادامه بررسی، فرض کنید این رقم  $C_1$  برابر باشد با تعداد افرادی که پرسشنامه خود را تکمیل و عودت داده‌اند، به اضافه‌ی تمامی افرادی که در آغاز جزو بی‌پاسخ‌ها بوده‌اند ولی سپس پرسشنامه خود را تکمیل و ارسال داشته‌اند (با استفاده یا بدون استفاده از روش‌های نمونه‌گیری) در مرحله بعد، یک نمونه

سرشماری یا نمونه‌گیری یا شاید هر دو  
 گزیده مطالب آماری - ۶۱

با حجم  $S_1$  انتخاب می‌شود؛ که متناظر با دور دوم علامت‌گذاری در مثال پرندگان فوق خواهد بود. تعداد مواردی از این دو منبع که مقابله و بریکدیگر منطبق شده‌اند، برابر  $\alpha$  فرض می‌شود. یعنی در واقع  $\alpha$  تعداد افرادی از جامعه آماری است که هم در سرشماری و هم در نمونه‌گیری بعدی ثبت شده‌اند. همچنین  $C = S_1 - \alpha$  تعداد افرادی است که در نمونه شمرده شده ولی در سرشماری پوشش داده نشده‌اند. و همین طور  $d$  تعداد افرادی است که در هیچ‌یک از دو منبع مشاهده نشده‌اند. بنابراین، متناظر با آنچه در مثال پرندگان مشاهده شد، جدول ذیل به دست خواهد آمد:

جمع	نمونه‌گیری		
	به حساب نیامده	به حساب آمده	
$C_1$	$b$	$\alpha$	سرشماری به حساب آمده
$C_2$	$d$	$C$	به حساب نیامده
$N=?$	$S_2$	$S_1$	جمع

تعداد کل افراد یا حجم جامعه ( $N$ ) برابر خواهد بود با:  $N = \alpha + b + c + d$  در این رابطه اجزا  $(\alpha + b + c)$  معلوم است. چرا که  $\alpha$  و  $b$  از اطلاعات مربوط به سرشماری و  $C$  از اطلاعات مربوط به نمونه‌گیری به دست آمده است. ولی می‌توان  $d$  را نیز به صورت  $d = \frac{b \times c}{\alpha}$  برآورد کرد. پس از انجام عملیات جبری لازم، نتیجه ذیل به دست خواهد آمد:

$$N = \frac{(\alpha + b) \times (\alpha + c)}{\alpha} = \frac{c_1 \times s_1}{\alpha}$$

مفهوم این رابطه به شرح ذیل است:

(رقم حاصل از نمونه‌گیری) (رقم حاصل از سرشماری)  
 تعداد حالات مقابله شده

گزیده مطالعه ب آماری - ۶۰  
سرشماری یا نمونه‌گیری یا شاید هر دو

$N=$

مبانی نظری نمونه‌گیری به روش (C.R)، ابتدا در سال‌های ۱۸۰۰ و به منظور برآورد حجم جوامع متعلق به حیات وحش ابداع شد. یکی از فرضیات اصلی این روش آمارگیری آن است که هر فرد آماری (پرنده، حیوان و ...) دارای احتمال مساوی برای انتخاب شدن است. مطالعات نشان داده است که کم‌شماری در سرشماریها بسته به منطقه، نژاد و سطح درآمد، به طور برجسته و قابل توجهی تغییر می‌کند، بنابراین، اگر فقط یک طرح (CR) ساده به اجرا درآید، ضرورتاً فرض مزبور محقق نخواهد شد. همین امر، استفاده از یک طرح (CR) ولی به شکل طبقه‌بندی شده را ضرورت می‌بخشد تا یکسانی داخل هر طبقه حاصل شود. البته تلطیف‌های دیگری نیز وجود دارد، ولی هیچ‌یک از آنها با مبانی و اصول مطرح شده در اینجا، ناسازگاری ندارند.