



پژوهشکده‌ی آمار

جمهوری اسلامی ایران
مرکز آمار ایران
پژوهشکده‌ی آمار

نمونه‌گیری متعادل شده و امکان‌سنجی استفاده از آن در آمارگیری‌های ملی

نمونه گیری متعادل شده و امکان سنجی استفاده از آن در آمار گیری های ملی

فاطمه هرنندی (مجری)

محمد رضا فریدروhani

زهرة فلاح محسن خانی

فرهاد مهران (مشاور)



پژوهشکده های آمار

گروه پژوهشی طرح های فنی و روش های آماری

پیش‌گفتار

امروزه آمارگیری‌های نمونه‌ای به‌عنوان مکمل روش سرشماری در تهیه‌ی آمارهای رسمی کشور به صورت گسترده‌ای مورد استفاده قرار می‌گیرد. به منظور دستیابی به برآوردهایی مطلوب از ویژگی‌های جامعه مورد نظر، انتخاب نمونه به روش‌های مناسب به نحوی که منجر به نمونه‌ای تصادفی و در عین حال معرف جامعه مورد بررسی گردد، از اهمیت به‌سزایی برخوردار است. نمونه‌گیری متعادل شده به‌عنوان یک روش نمونه‌گیری احتمالی که می‌تواند به نمونه‌ای معرف از جامعه مورد بررسی منجر شود به این مفهوم که برآورد مجموع متغیرهای کمکی بر مبنای نمونه مزبور با مقدار معلوم آن‌ها برابر گردد؛ از دیرباز مورد توجه آمارشناسان بوده است، اما به این دلیل که روش‌های کارا برای انتخاب نمونه متعادل شده که در عین حال نمونه‌ای تصادفی هم باشد، وجود نداشت و روش‌های موجود نیز تنها در شرایط خاص قابل استفاده بود، این شیوه نمونه‌گیری در آمارگیری‌های ملی وارد نشده بود. به تازگی (۲۰۰۴)، روشی به نام روش مکعب برای انتخاب نمونه متعادل شده پیشنهاد گردیده است که امکان انتخاب نمونه تصادفی با احتمال‌های انتخاب برابر یا نابرابر که در عین حال بر اساس متغیرهای متعددی، متعادل نیز باشد را فراهم می‌سازد. این روش بنابر ادعای استفاده‌کنندگان آن، منجر به افزایش ۲۰ تا ۹۰ درصدی کارایی آمارگیری‌ها می‌شود.

پژوهشکده‌ی آمار با توجه به رسالت خود در زمینه ارتقای کیفیت فرآیند تولید اطلاعات در نظام آماری کشور، معرفی روش‌های نوین و مناسب تولید آمار را در دستور کار دارد. مجموعه‌ی حاضر با هدف معرفی روش نمونه‌گیری متعادل شده، جایگاه آن در بین روش‌های مختلف نمونه‌گیری از جوامع

محدود و سنجش کارایی آن در آمارگیری‌های ملی، گامی است در این راه که امید می‌رود مورد توجه و استفاده طراحان آمارگیری‌های نمونه‌ای قرار گیرد.

اجرای این طرح با مدیریت سرکار خانم فاطمه هرنندی و همکاری جناب آقای دکتر محمدرضا فرید روحانی و سرکار خانم زهره فلاح محسن‌خانی انجام شده است که در این‌جا از یکایک این عزیزان تشکر و قدردانی می‌شود.

مشاور محترم طرح، جناب آقای دکتر فرهاد مهران نیز علاوه بر ارائه راهنمایی‌های بسیار ارزشمند، با انجام بخشی از مراحل اصلی کار، گروه مطالعاتی را بیش از پیش یاری نمودند که بدین وسیله از همکاری صمیمانه ایشان تشکر و قدردانی می‌شود.

از سرکار خانم ناظریان که زحمت تایپ این گزارش را به عهده داشته و با حوصله فراوان آن را به انجام رساندند نیز بسیار تشکر می‌شود.

از خوانندگان محترم تقاضا می‌شود، نظریات اصلاحی خود در ارتباط با محتوای مجموعه حاضر را به گروه پژوهشی طرح‌های فنی و روش‌های آماری پژوهشکده‌ی آمار منعکس نمایند.

گروه پژوهشی طرح‌های فنی و روش‌های آماری
پژوهشکده‌ی آمار

فهرست مطالب

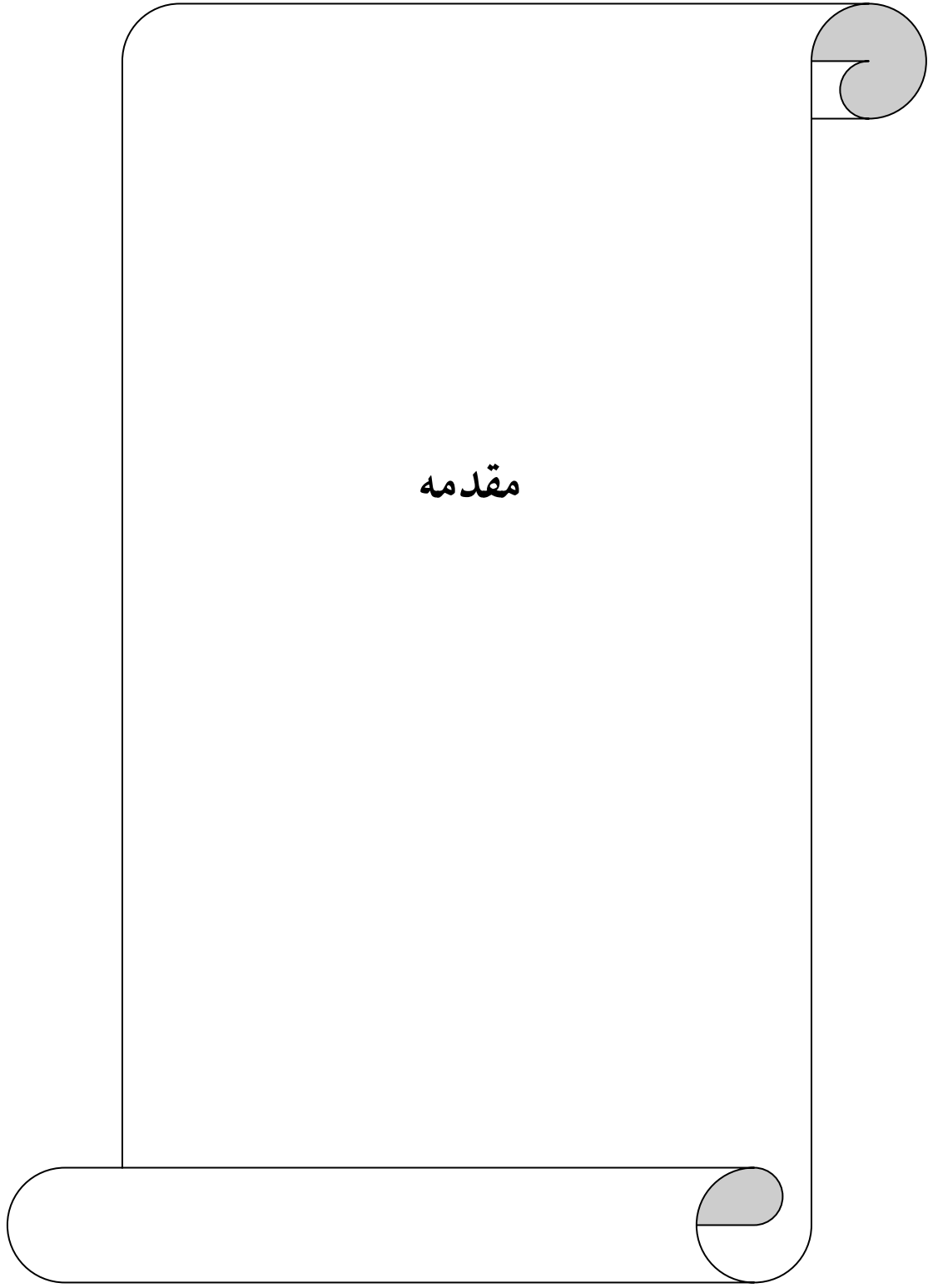
صفحه

عنوان

۱	مقدمه
۷	فصل ۱- کلیات
۹	۱-۱- تاریخچه آمارگیری نمونه‌ای
۱۳	۱-۲- ویژگی‌های آمارگیری نمونه‌ای
۱۴	۱-۳- نمونه‌گیری احتمالی در مقایسه با نمونه‌گیری غیر احتمالی
۱۹	۱-۴- نمونه معرف
۲۷	فصل ۲- روش‌های دستیابی به راهبرد معرف
۳۱	۲-۱- رهیافت مدل مبنا
۳۴	۲-۲- رهیافت طرح مبنا
۳۸	۲-۳- مقایسه دو رهیافت
۳۸	۲-۳-۱- اشکالات وارد شده به رهیافت طرح مبنا
۴۰	۲-۳-۲- اشکالات وارد شده به رهیافت مدل مبنا
۴۲	۲-۴- جمع‌بندی
۴۵	۲-۵- روش‌های دستیابی به راهبرد معرف در رهیافت طرح مبنا
۴۶	۲-۵-۱- طبقه‌بندی
۴۷	۲-۵-۲- خوشه‌بندی
۴۷	۲-۵-۳- نمونه‌گیری سیستماتیک از لیست مرتب شده
۴۸	۲-۵-۴- احتمال‌های انتخاب نابرابر

۴۸ ۲-۵-۵- برآوردگرهای کالیبره
۵۷ فصل ۳- نمونه‌گیری متعادل شده
۵۹ ۳-۱- مفاهیم پایه در نمونه‌گیری احتمالی
۵۹ ۳-۱-۱- جامعه آماری و اطلاعات مورد نظر
۶۱ ۳-۱-۲- چارچوب نمونه‌گیری
۶۱ ۳-۱-۳- احتمال انتخاب نمونه
۶۶ ۳-۱-۴- نمایش هندسی انتخاب نمونه
۶۷ ۳-۱-۵- برآورد
۶۷ ۳-۲- نمونه‌گیری متعادل شده
۶۸ ۳-۲-۱- حالت‌هایی خاص از طرح‌های نمونه‌گیری متعادل شده
۶۹ ۳-۲-۲- تاریخچه نمونه‌گیری متعادل شده
۷۱ ۳-۲-۳- روش کل شماری
۷۴ ۳-۲-۳-۱- معایب روش کل شماری
۷۵ ۳-۲-۴- روش مکعب
۷۵ ۳-۲-۴-۱- نمایش هندسی نمونه‌گیری متعادل شده در روش مکعب
۷۸ ۳-۲-۴-۲- شرایط تحقق طرح نمونه‌گیری متعادل شده
۸۲ ۳-۲-۴-۳- روش‌های انتخاب تصادفی رأسی از K (مرحله پرواز)
۸۴ ۳-۲-۴-۴- روش‌های انتخاب یکی از کل نمونه‌های مجاور رأس انتخاب شده K (مرحله فرود)
۸۹ ۳-۲-۴-۵- کیفیت تعادل

۱۸۹	نرم افزار ۳-۲-۴-۶
۹۰	استفاده از نمونه گیری متعادل شده در آمارگیری های ملی ۳-۳
۹۳	فصل ۴- شبیه سازی
۹۷	۱-۴- داده های شبیه سازی
۱۰۲	۲-۴- هدف شبیه سازی
۱۰۳	۳-۴- تعداد نمونه شبیه سازی و احتمال های انتخاب
۱۰۳	۴-۴- روش شبیه سازی
۱۰۵	۵-۴- نتایج شبیه سازی
۱۱۶	۶-۴- نتیجه گیری و پیشنهادات
۱۱۹	مراجع اصلی
۱۲۳	سایر مراجع
۱۳۱	پیوست ها



هدف نمونه‌گیری استفاده از اطلاعات جزئی از یک جامعه برای نتیجه‌گیری درباره کل جامعه می‌باشد. غالباً نمونه‌گیری را در مقابل سرشماری قرار می‌دهند و مقایسه می‌کنند. به علت عملکرد محدودتر، هزینه نمونه‌گیری اساساً کم‌تر از سرشماری است و به همین دلیل نیز روش نمونه‌گیری، جمع‌آوری اطلاعات بیش‌تر و دقیق‌تر اعضای جامعه را ممکن می‌سازد. در عوض نمونه‌گیری همراه اشتباهات ناشی از بررسی جزء به جای کل می‌باشد که باید حتی‌الامکان آن را با استفاده از روش‌های کارا به حداقل رساند.

امروزه در تهیه آمارهای رسمی کشورها از روش نمونه‌گیری به عنوان مکمل روش سرشماری و در بعضی کشورها حتی به عنوان جایگزین آن استفاده می‌شود.

نمونه حاصل از نمونه‌گیری را از دو دیدگاه می‌توان ارزیابی کرد:

- معرف بودن: تا چه حد نمونه ساختار اساسی جامعه را محفوظ نگاه داشته است؟
- تصادفی بودن: آیا واحدهای انتخاب شده در نمونه و همچنین واحدهای انتخاب نشده همگی احتمال انتخاب مثبت و مشخص داشته‌اند؟

دیدگاه اول به روش نمونه‌گیری توجه ندارد بلکه ارزش نمونه در مشخصات نمونه حاصل می‌باشد. در نقطه مقابل، دیدگاه دوم به مشخصات نمونه انتخاب شده توجه ندارد و ارزش نمونه بر اساس نحوه نمونه‌گیری و رعایت اصول آن تعیین می‌شود.

بررسی تاریخیچه‌ی آمارگیری‌های نمونه‌ای نشان می‌دهد که در سال‌های اولیه، دیدگاه اول، دیدگاه غالب بوده، اما با گذشت زمان و آشکار شدن ضعف‌های نمونه‌گیری‌های غیر تصادفی، دیدگاه دوم از مقبولیت گسترده‌ای در بین آمارشناسان برخوردار گردید به نحوی که امروزه، تصادفی بودن انتخاب نمونه، اصل لازم هر آمارگیری از جامعه محدود تلقی می‌شود هر چند به تنهایی برای اعتبار نتایج آمارگیری کافی نیست. برای معتبر نمودن نتایج آمارگیری مبتنی بر نمونه‌های تصادفی لازم است از تمامی اطلاعات موجود در مورد جامعه مورد بررسی قبل و بعد از نمونه‌گیری استفاده شود

تا از معرف بودن نمونه انتخابی اطمینان حاصل شود. در واقع تنها با تلفیق این دو دیدگاه در یک نمونه‌گیری، می‌توان به برآوردهایی با دقت بالا از ویژگی‌های مورد نظر جامعه دست یافت. نمونه‌گیری متعادل شده را می‌توان به صورت ترکیبی از این دو دیدگاه بیان کرد که در آن، هدف به‌دست آوردن نمونه تصادفی می‌باشد به نحوی که نمونه حاصل از جهاتی معرف جامعه مورد بررسی باشد.

نمونه‌گیری متعادل شده، بحثی است که از دیرباز و حتی قبل از طرح نظریه نمونه‌گیری احتمالی توسط نیمن (۱۹۳۴) مطرح بوده است اما به دلیل آن‌که روش‌های کارا برای انتخاب نمونه متعادل شده که در عین حال نمونه‌ای تصادفی هم باشد، وجود نداشت و روش‌های موجود نیز یا محدود به شرایط خاص بوده و یا امکان عملیاتی شدن نداشته‌اند، استفاده از این شیوه نمونه‌گیری علی‌رغم جذابیت مفهوم آن، در آمارگیری‌های ملی وارد نشده بود. در طی سالیان متمادی نیز بسیاری از آمارشناسان زبده، تنها راه حل‌های موردی برای نمونه‌گیری متعادل شده ارائه می‌نمودند.

در سال ۲۰۰۴، دوپل و تیه روشی کلی برای انتخاب نمونه متعادل شده به نام روش مکعب که امکان انتخاب نمونه تصادفی با احتمال‌های انتخاب برابر یا نابرابر که در عین حال بر اساس متغیرهای متعددی معرف نیز باشد را پیشنهاد نمودند. این روش هم اکنون توسط مرکز آمار فرانسه برای اکثر آمارگیری‌های ملی مورد استفاده قرار می‌گیرد و این ادعا وجود دارد که استفاده از نمونه‌گیری متعادل شده منجر به افزایش ۲۰ تا ۹۰ درصدی کارایی این آمارگیری‌ها شده است (دوپل و تیه، ۲۰۰۴).

هر چند هدف اصلی گزارش حاضر، معرفی روش نمونه‌گیری متعادل شده و امکان‌سنجی استفاده از آن در آمارگیری‌های ملی می‌باشد اما به منظور نشان دادن جایگاه این روش در بین روش‌های مختلف نمونه‌گیری از جوامع محدود و تشریح این مسئله که چرا بحث متعادل بودن نمونه انتخابی از دیرباز مورد توجه و علاقه آمارشناسان بوده است، هر یک از مباحث مطرح شده در فوق به تفصیل در فصول مختلف این گزارش مورد بحث و بررسی قرار گرفته‌اند.

در فصل اول گزارش، کلیاتی در مورد آمارگیری‌های نمونه‌ای ارائه شده است که شامل تاریخچه، ویژگی‌های این آمارگیری‌ها و مقایسه بین نمونه‌گیری‌های احتمالی و غیر احتمالی می‌باشد. در آخرین قسمت این فصل، مفهوم نمونه معرف و راهبرد معرف بیان شده است.

در فصل دوم گزارش، روش‌های دستیابی به راهبرد معرف شامل رهیافت مدل مبنا و طرح مبنا مورد بحث و بررسی قرار گرفته است و پس از مقایسه این دو رهیافت، اهم روش‌هایی که در رهیافت طرح مبنا برای بهبود طرح نمونه‌گیری و روش برآورد با استفاده از اطلاعات کمکی وجود دارند که یکی از آن‌ها نیز نمونه‌گیری متعادل شده می‌باشد، بیان شده است.

فصل سوم گزارش به تشریح نمونه‌گیری متعادل شده اختصاص یافته است. در این فصل ابتدا مروری بر مفاهیم پایه در نمونه‌گیری احتمالی می‌شود و پس از بیان تاریخچه نمونه‌گیری متعادل شده، جزئیات روش مکعب که در انواع مختلف نمونه‌گیری و برای هر تعداد متغیر کمکی برای متعادل نمودن نمونه، قابل استفاده می‌باشد، ارائه شده است.

در فصل چهارم گزارش با انجام یک شبیه‌سازی با استفاده از یک مجموعه داده واقعی، عملکرد نمونه‌گیری متعادل شده مورد ارزیابی قرار گرفته است. از آن‌جا که برآوردهای کالیبره نیز از طریق تولید برآوردهای صحیح از متغیرهای کمکی همبسته با متغیرهای مورد بررسی سعی در بهبود برآوردهای مورد نظر دارند، در ارزیابی مذکور، عملکرد این نوع برآوردها نیز در کنار نمونه‌گیری متعادل شده بررسی شده است. در پایان این فصل نتیجه‌گیری و پیشنهادات این طرح پژوهشی ارائه شده است.

فصل اول

کلیات

۱-۱- تاریخچه آمارگیری نمونه‌ای

امروزه آمارگیری نمونه‌ای در سطحی گسترده توسط سازمان‌های دولتی، محققان بازار، طراحان نظرسنجی‌ها، محققان اجتماعی و سایر افراد استفاده می‌شود. ویژگی اصلی که موجب شده آمارگیری نمونه‌ای علی‌رغم سابقه نسبتاً کوتاه ۱۰۰ ساله خود که عمدتاً نیز از سال ۱۹۳۰ به بعد بوده، به یک چنین جایگاهی برسد، به وجود آمدن روش‌های نمونه‌گیری کارا برای جمع‌آوری مقرون به صرفه داده‌ها بوده است.

در نشست‌های متعدد مؤسسه بین‌المللی آمار^۱ (ISI) در قرن گذشته، آمارشناسان در مورد اعتبار علمی انواع نمونه‌گیری‌ها از جامعه محدود، بحث و گفتگو داشته‌اند. اولین بار نیکلای کیائر رئیس دفتر آمار نروژ در همایش مؤسسه بین‌المللی آمار در سال ۱۸۹۵ (کیائر، ۱۸۹۶-۱۸۹۵)، استفاده از روش نمونه‌گیری از جوامع محدود را که وی آن را «روش معرف» و دیگران آن را «بررسی جزئی» یا «روش‌های غیر مستقیم» می‌نامیدند، مطرح نمود. پیشنهاد او ابتدا با مقاومت زیادی از جانب آمارشناسان دولتی که معتقد بودند تنها پوشش کامل جامعه (تمام شماری) مطلوب است، مواجه شد اما پس از پی‌گیری و پافشاری مصرانه وی بر نظرات خود در کنفرانس‌های متعدد، سرانجام ISI در سال ۱۹۰۳ قطعنامه‌ای را مبنی بر استفاده از روش معرف تصویب نمود (کیائر، ۱۹۰۵). پس از آن هم‌چون موارد مشابه در تاریخ علم، که آنچه در ابتدا جدید محسوب می‌شود پس از گذشت مدتی، تثبیت شده و توسط عده زیادی مورد استفاده قرار می‌گیرد، نظریه‌ی جدید بسط و توسعه یافت. لوسین مارچ در سال ۱۹۰۳ برای اولین بار نظریه نمونه‌گیری احتمالی را مطرح کرد (کیائر، ۱۹۰۵). باولی و همکاران وی در ۱۹۰۶ یک آمارگیری نمونه‌ای واقعی را انجام دادند (باولی و دیگران، ۱۹۱۵). ادین در سال ۱۹۱۰، اولین استفاده مستند از تصادفی‌سازی فیزیکی را انجام داد (ادین، ۱۹۱۲). وی معتقد بود که تنها نمونه‌گیری

^۱ International Statistical Institute

تصادفی به معنی دقیق کلمه می‌تواند روش معرف باشد در صورتی که بسیاری دیگر از کاربران اولیه، نمونه‌گیری قصدی^۱ از مناطق جغرافیایی بزرگ را روش معرف می‌دانستند.

در ماه می سال ۱۹۲۴ دفتر مؤسسه بین‌المللی آمار، کمیسیونی را برای مطالعه کاربرد روش معرف در آمار تعیین نمود. اعضای این کمیسیون عبارت بودند از پروفیسور باولی، پروفیسور جینی، ینسن، مارچ، پروفیسور استیوارت و پروفیسور زیزک. کمیسیون نتایج مطالعات خود را در قالب یک گزارش در سال ۱۹۲۵ (ینسن، ۱۹۲۶) ارائه نمود. به این گزارش مقاله‌ای از ینسن تحت عنوان «روش معرف در عمل» و مقاله‌ای از باولی تحت عنوان «اندازه‌گیری دقت در نمونه‌گیری» به همراه دو مقاله کوتاه‌تر از استیوارت و مارچ ضمیمه شده بود. این گزارش و ضمائم آن تصویر مناسبی از روش‌های نمونه‌گیری به‌کار گرفته شده پس از آن در آمارگیری‌های مختلف و همچنین نظریه آماری‌ای که این روش‌ها بر اساس آن استوار بودند را ارائه می‌نماید. نتایج این گزارش در مجموعه‌ای از مصوبات مؤسسه بین‌المللی آمار به شرح زیر گنجانده شد.

با توجه به این‌که در بسیاری از موارد به علت عدم امکان حصول داده‌های آماری کامل، لازم است نتایج کلی را بر اساس بررسی‌های جزئی به‌دست آورد؛

و با توجه به این‌که حتی در مواردی که داده‌های کامل در دسترس هستند، این امکان وجود دارد که کار با بخشی از داده‌های مزبور کفایت نماید مشروط بر آن‌که انتخاب این بخش بر اساس یک روش معقول و منطقی انجام پذیرد؛ و نظر به این‌که صرفه‌جویی در نیروی انسانی، زمان و هزینه که با محدود کردن بررسی به بخشی از داده‌ها حاصل می‌شود، اغلب این امکان را فراهم خواهد ساخت که استفاده‌ی جامع‌تری از اطلاعات در دسترس انجام گیرد و همچنین نسبت به‌کار با کل داده‌ها، درک عمیق‌تری از موضوع به‌دست آید؛

I. پیرو مصوبه نشست برلین در سال ۱۹۰۳، مجدداً بر مزایای قابل ملاحظه‌ای که می‌توان از کاربرد روش معرف تحت شرایط زیر به‌دست آورد تأکید می‌شود:

¹ Purposive Sampling

نتایج یک بررسی جزئی تنها در مواردی قابل تعمیم است که نمونه‌ی به‌کار رفته ماهیتاً به اندازه‌ی کافی معرف کل (جامعه) باشد. با چنین ملاحظاتی، نمونه به طرق مختلفی قابل انتخاب است که در ذیل به دو حالت اصلی آن اشاره می‌شود:

الف) انتخاب تصادفی

تعدادی از واحدها با شانس دقیقاً یکسان شمول در نمونه، انتخاب می‌گردند. بنابراین دقت نتایج به تعداد نمونه وابسته است که باید به اندازه‌ی کافی بزرگ باشد تا از انحرافات اتفاقی ناچیز تأثیر نپذیرد.

ب) انتخاب قصدی

گروه‌هایی از واحدها به طریقی انتخاب می‌گردند که بر اساس نتایج حاصل از تمامی آن‌ها، ویژگی‌های تقریباً یکسان با کل جامعه به‌دست آید. در واقع گروه‌های انتخاب شده باید همان ویژگی را که کل جامعه دارد به نمونه بدهند. به‌منظور اطلاع از دقت برآورد، لازم است به تعداد کافی گروه در نمونه منظور شود به قسمی که بتوان تغییر ویژگی‌های مورد نظر بین گروه‌ها را اندازه‌گیری نمود. اما از آنجایی که دقت نتایج، اغلب به میزان قابل توجهی از اصول به‌کار رفته در انتخاب نمونه تأثیر می‌پذیرد، کنترل‌های زیر توصیه می‌شود:

- انتخاب نمونه باید حداقل دو بار بر اساس قاعده یکسان انجام گیرد؛ پس از مقایسه‌ی نمونه‌ها، می‌توان آن‌ها را ترکیب نمود (این توصیه در مورد انتخاب تصادفی نیز قابل استفاده است).
- برای تمامی تکرارها، ارتباط بین جزء و کل باید به‌طور مرتب کنترل شود.

II توصیه می‌شود حتی‌الامکان بررسی به صورتی ترتیب داده شود که دقت نتایج را بتوان تحت عبارات ریاضی بیان نمود و به علاوه همراه نتایج، اطلاعی از میزان خطای آن‌ها نیز ارائه شود.

III همان‌گونه که در مصوبه سال ۱۹۰۳ درخواست شد مجدداً تکرار می‌کنیم که لازم است در گزارش نتایج هر بررسی معرف، شرح کامل و روشنی از روش انتخاب نمونه ارائه شود.

در گزارش ISI همچنین ذکر شده بود که هنگامی که ترکیب جامعه مورد نظر، متفاوت از جوامعی باشد که مبنای بازی‌های شانسی^۱ هستند، انتخاب قصدی نمونه ارجحیت دارد. همان‌گونه که ملاحظه می‌شود در این مصوبات، اهمیت قابل توجهی به روش نمونه‌گیری قصدی داده شده است اما در همان سال

¹ Games of Chance

(۱۹۲۶) در دفتر سرشماری ایتالیا، خطای فاحشی در نتایج حاصل از نمونه‌ای قصدی از بخش‌های سرشماری به وقوع پیوست که نه تنها ضعف این روش را آشکار ساخت بلکه تهدیدی برای تخریب اعتبار کلی نمونه‌گیری‌ها بود (جینی و گالوانی، ۱۹۲۹). جریان از این قرار بود که آمارشناسان ایتالیایی، جینی و گالوانی برای نگهداری درصدی از فرم‌های تکمیل شده سرشماری جمعیت ۱۹۲۱ ایتالیا به‌منظور استفاده‌های بعدی از روش نمونه‌گیری قصدی استفاده کردند با این هدف که نمونه حاصل معرف کل کشور از جنبه‌های دموگرافی، اجتماعی، اقتصادی و ویژگی‌های جغرافیایی باشد. به این منظور با آزمون و خطا، ۲۹ بخش از ۲۱۴ بخش کشور را انتخاب کردند به نحوی که برای ۷ متغیر مهم (متغیرهای کنترل)، مقادیر متوسط به‌دست آمده بر اساس نمونه، بسیار نزدیک به همین مقادیر در کل کشور باشد. آن‌ها پس از استخراج نتایج با این مشکل مواجه شدند که هر چند نتایج حاصل از نمونه برای میانگین‌های متغیرهای کنترل بسیار خوب بود اما مقادیر به‌دست آمده برای میانگین‌های سایر متغیرها بر اساس نمونه قصدی انتخاب شده در اغلب موارد به‌طور معنی‌داری با مقادیر واقعی کشور تفاوت داشتند. نیمن در مقاله مهم خود در سال ۱۹۳۴ با تحلیل کار آمارشناسان ایتالیایی، ضعف روش نمونه‌گیری قصدی را که ناشی از اتکای آن به مفروضاتی غیر قابل اثبات در مورد احتمال‌های پیشین بود^۱، نشان داد و ثابت نمود تنها روش معرفی که می‌توان برای کاربرد عمومی توصیه نمود، نمونه‌گیری تصادفی طبقه‌بندی شده است. در واقع مقاله نیمن با پی‌ریزی اساس نظریه نمونه‌گیری و آشکار نمودن قدرت نمونه‌گیری احتمالی در حل مسائل برآورد آماری، مستقل از هر گونه اطلاعی در مورد قوانین احتمال پیشین، نمونه‌گیری قصدی را که به‌عنوان مطرح‌ترین روش نمونه‌گیری تا آن زمان از مقبولیت خاصی برخوردار بود، به حاشیه راند.

در این مقاله وی همچنین تئوری استنباط بر اساس فاصله اطمینان را ارائه، نمونه‌گیری احتمالی از خوشه‌ها را مطرح و انتساب بهینه برای نمونه‌گیری طبقه‌بندی شده را بیان نمود.

^۱ فرض پایه در روش انتخاب قصدی استفاده شده این بود که رگرسیون متغیرهای کنترل روی متغیرهای مورد نظر خطی است.

این مقاله در دهه ۱۹۴۰ با تعدادی مقاله دیگر دنبال شد که در مورد روش‌های اصلی نمونه‌گیری احتمالی بودند. این کار توسط آمارشناسانی از قبیل کوکران، دمینگ، هنسن، هورویتز، مدوآ، ماهالانوبیس و بیتس انجام شد. در پایان آن دهه موضوع آمارگیری نمونه‌ای به حد کافی مستحکم و منسجم شده بود که بتوان اولین مجموعه از کتب مرجع تخصصی را ارائه نمود. اولین نسل از کتاب‌های درسی در زمینه نمونه‌گیری حوالی ۱۹۵۰ نوشته شد. در سال‌های بعد از آن نیز پیشرفت تئوری و عملی قابل توجهی در روش‌های نمونه‌گیری و استنباط آماری و همچنین روش‌های محاسبه خطای نمونه‌گیری در آمارگیری‌ها حاصل شد به نحوی که اکثر ادارات آمار برای تولید آمار رسمی^۱، اقدام به طراحی آمارگیری‌های مبتنی بر نمونه‌گیری احتمالی نمودند.

۱-۲- ویژگی‌های آمارگیری نمونه‌ای

موضوع آمارگیری نمونه‌ای، فرآیند انتخاب اعضای جامعه برای نمونه و فرآیند برآورد است. طرح نمونه باید با توجه به اهداف آمارگیری تهیه شود و بخش اساسی طرح کلی آمارگیری باشد. نقش ملاحظات هزینه در تعیین یک طرح نمونه کارا بسیار مهم است. بدین منظور برای طراحی یک نمونه کارا، مدل هزینه‌های مربوط به جنبه‌های مختلف فرآیند آمارگیری باید تهیه شود. در این صورت، هدف طرح نمونه را می‌توان به شکل حداقل نمودن هزینه کل آمارگیری برای رسیدن به سطح مورد نظر واریانس برآوردهای آمارگیری (خطای نمونه‌گیری) یا برعکس به صورت مینیمم کردن واریانس برآوردها، برای سطح مشخص هزینه کل، تعریف کرد.

یک ویژگی متمایز کننده آمارگیری نمونه‌ای از نمونه‌گیری معمول در ادبیات آماری این است که جامعه‌ای که نمونه از آن انتخاب می‌شود، دارای اندازه محدود است و هر چند اغلب خیلی بزرگ است اما جامعه نامحدودی که در تئوری استاندارد آماری فرض می‌شود نیست. یک ویژگی دیگر آن نیز این

^۱ Formal Statistics

است که اغلب میزان قابل توجهی اطلاعات در مورد اعضای جامعه وجود دارد. روش‌های مختلف استفاده از اطلاعات مکمل، منجر به تهیه روش‌های متنوع آمارگیری شده است. اطلاعات مکمل را می‌توان در مرحله انتخاب نمونه، در مرحله برآورد یا در هر دو مرحله استفاده نمود.

۱-۳- نمونه‌گیری احتمالی در مقایسه با نمونه‌گیری غیر احتمالی

یک تمایز اصلی در روش‌های نمونه‌گیری، بین نمونه‌گیری‌های احتمالی و غیر احتمالی است. در نمونه‌گیری احتمالی هر عضو جامعه، احتمال انتخاب مشخص و مثبت دارد. این نوع نمونه‌گیری مستلزم وجود چارچوبی است که واحدهای آن دارای ارتباطی یک به یک، یک به چند، چند به یک یا چند به چند با واحدهای جامعه مورد بررسی هستند. در نمونه‌گیری احتمالی، پس از تعیین جامعه مورد نظر و تهیه چارچوب مرتبط با آن، اعضای نمونه با استفاده از یک الگوریتم احتمالاتی از چارچوب مذکور انتخاب می‌شوند به نحوی که هر عنصر چارچوب شانس انتخابی معین و مثبت داشته باشد و بتوان احتمال انتخاب نمونه حاصل را محاسبه نمود.

نمونه‌گیری غیر احتمالی، به هر نوع نمونه‌گیری که قادر به محقق نمودن شرایط نمونه‌گیری احتمالی نباشد، اطلاق می‌شود. در این نوع نمونه‌گیری لزومی به تهیه چارچوب نیز وجود ندارد.

در نمونه‌گیری احتمالی، برای به‌دست آوردن برآورد، اطلاعات حاصل از نمونه را با لحاظ نمودن احتمال‌های انتخاب آن‌ها به‌نحوی در قالب یک برآوردگر ترکیب می‌کنیم که در صورت تکرار نمونه‌گیری به دفعات زیاد (تمام حالات ممکن بر اساس طرح نمونه‌گیری)، میانگین برآوردهای حاصل از نمونه‌های مختلف با نتایج حاصل از آمارگیری کل جامعه یکسان شود (برآوردگر ناریب).

مزیت عمده نمونه‌گیری احتمالی این است که تئوری آماری را می‌توان برای به‌دست آوردن خواص برآوردگرها به‌کار برد و از آن‌جا که احتمال انتخاب هر نمونه مشخص است، امکان محاسبه خطای نمونه‌گیری و در نتیجه بازه اطمینان برای پارامترهای جامعه بر اساس اطلاعات آمارگیری وجود دارد.

خطای نمونه‌گیری، میزان تغییرپذیری نتایج، ناشی از بررسی نمونه‌ای از عناصر به جای کل آن‌ها را نشان می‌دهد و سنجش از کیفیت طرح نمونه‌گیری و نتایج آمارگیری را ارائه می‌کند. در نمونه‌گیری غیر احتمالی چنین امکانی وجود ندارد و تنها می‌توان به ارزیابی ذهنی و شخصی در مورد نتایج اتکا نمود. علاوه بر آن چون هیچ نظریه خاصی برای روش‌های نمونه‌گیری غیر احتمالی وجود ندارد، نمی‌توان هیچ‌گونه قضاوت کلی در مورد این روش‌ها داشت و هر کاربرد این روش‌ها باید جداگانه مورد ارزیابی قرار گیرد. علی‌رغم ضعف تئوریک نمونه‌گیری غیر احتمالی، انواع مختلف آن به دلیل کم هزینه بودن و سهولت اجرا مورد استفاده قرار می‌گیرد. سه نوع متداول نمونه‌گیری غیر احتمالی عبارتند از:

- نمونه‌گیری اتفاقی، بی‌دردسر یا شانسی^۱

در این روش نمونه‌گیری، عناصر نمونه به صورت اتفاقی با هدف استنباط در مورد کل جامعه انتخاب می‌شوند. به‌عنوان مثال‌هایی از این نمونه‌گیری می‌توان از نمونه داوطلبان، مصاحبه‌های کنار خیابان و نمونه خوانندگان تکمیل‌کننده پرسشنامه‌های مندرج در نشریات نام برد.

- نمونه‌گیری قضاوتی یا قصدی یا انتخاب کارشناسی^۲

در این شیوه نمونه‌گیری، عناصر نمونه با دقت فراوان انتخاب می‌شوند تا نمونه‌ای «معرف»^۳ حاصل شود. بدین منظور انتخاب نمونه‌ها به نحوی انجام می‌شود که برای متغیرهای کنترل، برآوردهای حاصل از نمونه با مقدار واقعی آن در جامعه یکسان شود. فرض اساسی در روش انتخاب قصدی این است که متغیرهای مورد نظر در آمارگیری با متغیرهای کنترل هم‌بسته هستند. در واقع اعتبار نتایج حاصل از انتخاب قصدی بر مبنای فرضیاتی در مورد جامعه مورد بررسی است که مشکل بتوان صحت آن‌ها را آزمون نمود. در مورد انتخاب کارشناسی نیز، مطالعات انجام شده مشخص

¹ Haphazard, Convenience or Accidental Sampling

² Judgment or Purposive Sampling or Expert Choice

³ Representative