مجله علوم آماری، پاییز و زمستان ۱۴۰۰
جلد ۱۵، شماره ۲، ص ۵۳۳ – ۵۴۸
DOI: 10.29252/jss.15.2.533
مقاله پژوهشی

پیش‌نهادی برای به‌کارگیری نمونه‌گیری مجموعه‌رتبه‌ای در آمارگیری هزینه و درآمد خانوار

روشنک علی‌اکبری صبا[۱]، نسرین ابراهیمی[۲]، لیدا کلهری ندرآبادی[۱]، آسیه عباسی[۲]

[۱] پژوهشکده آمار
[۲] مرکز آمار ایران

تاریخ دریافت: ۱۳۹۹/۱۲/۰۳ تاریخ پذیرش و انتشار: ۱۴۰۰/۰۶/۳۰

چکیده: روش نمونه‌گیری مجموعه‌رتبه‌ای با بهره‌گیری از اطلاعات رتبه‌بندی واحدها، نمونه معرف‌تری از جامعه در اختیار طراحان آمارگیری قرار می‌دهد که توزیع نمونه‌گیری آن به توزیع واقعی جامعه نزدیک‌تر است. در این مقاله برای اطمینان از اثربخشی نمونه‌گیری مجموعه‌رتبه‌ای در آمارگیری‌های بزرگی که به منظور تهیه آمارهای رسمی کشور انجام می‌گیرند، درصدد استفاده از این روش نمونه‌گیری برای بهبود کارایی براوردهای کلیدی آمارگیری هزینه و درآمد خانوار مرکز آمار ایران هستیم. نتایج بررسی نشان می‌دهد استفاده از نمونه‌گیری مجموعه‌رتبه‌ای در طراحی آمارگیری هزینه و درآمد خانوار می‌تواند منجر به بهبود کارایی براوردهای کلیدی طرح گردد به شرطی که متغیر رتبه‌بندی مورد استفاده همبستگی بالایی با متغیرهای اصلی طرح داشته باشد. بدیهی است در صورت نبود متغیری مناسب و در دسترس برای رتبه‌بندی واحدها می‌توان از اطلاعات چارچوب نمونه‌گیری برای ساخت متغیری همبسته با متغیرهای کلیدی طرح استفاده کرد.

واژه‌های کلیدی: آمارگیری هزینه و درآمد خانوار، نمونه‌گیری مجموعه‌رتبه‌ای، درآمد مورد انتظار.

آدرس الکترونیکی نویسنده مسئول مقاله: روشنک علی‌اکبری صبا، roshanaksaba22@gmail.com
کد موضوع‌بندی ریاضی (۲۰۱۰): 62D05 62J07.

۱ مقدمه

آمارگیری هزینه و درآمد خانوار غنی‌ترین منبع اطلاعاتی رسمی خانواری است که برای محاسبه شاخص‌های آماری در حوزه هزینه‌های زندگی، الگوی مصرف و درآمد خانوارها و نیز مباحث اقتصادی مانند سبد مصرفی خانوارها، ضریب جینی، توزیع عادلانه درآمد، سهم خانوارهای کم‌درآمد و بسیاری موضوعات مشابه دیگر مورد استفاده آمارشناسان، اقتصاددانان و محققین کشور قرار می‌گیرد. این آمارگیری با بیش از ۵۰ سال سابقه اجرا، در حال حاضر اطلاعات مفیدی در سطح استان و کل کشور برای برنامه‌ریزی‌های کلان ارائه می‌کند. آمارگیری هزینه و درآمد خانوار یک آمارگیری نمونه‌ای است که برای فراهم آمدن امکان محاسبه براوردهای درست طراحی نمونه‌گیری دقیق و پیچیده‌ای دارد.

طراحی نمونه‌گیری مقوله‌ای تخصصی در آمارگیری‌های نمونه‌ای است که بستری برای انجام استنباطهای صحیح در مورد مشخصه‌های جامعه فراهم می‌آورد. هر طرح نمونه‌گیری سه محور عمده دارد: روش نمونه‌گیری، اندازهٔ نمونه و در نهایت روش براوردی که با استفاده از آن، اطلاعات واحدهای نمونه برای استنباط در مورد کل جامعه به کار می‌رود. روش نمونه‌گیری و اطلاعات پیشین در مورد ویژگی‌های جامعه اغلب در تعیین اندازهٔ نمونه نقش دارند. روشی که برای براورد پارامترهای مورد نظر در جامعه به کار می‌رود نیز براساس روش نمونه‌گیری تعیین می‌شود. از این رو انتخاب روش نمونه‌گیری مناسب در آمارگیری‌های نمونه‌ای از اهمیت بالایی برخوردار است.

در آمارگیری هزینه و درآمد خانوار مانند اغلب آمارگیری‌های خانواری که برای محاسبه و ارائه آمارهای رسمی انجام می‌گیرند، از روش‌های نمونه‌گیری چندمرحله‌ای طبقه‌بندی‌شده بر پایه انتخاب تصادفی واحدهای جامعه استفاده می‌شود. در این روش‌ها که از روش‌های نمونه‌گیری احتمالی محسوب می‌شوند هر یک از واحدهای جامعه شانسی برای انتخاب در نمونه دارند و احتمال انتخاب آن‌ها مقدار مشخصی است. از آن جایی که انتخاب واحدها در روش‌های نمونه‌گیری احتمالی به‌صورت تصادفی بوده و احتمال انتخاب آن‌ها مشخص است، استفاده از آن‌ها منجر به براوردهای قابل اعتماد از پارامترهای موردنظر در جامعه می‌شود و امکان ارائه براوردی از خطای استاندارد هر براوردگر را نیز به‌عنوان معیاری برای ارزیابی کیفیت براوردها فراهم می‌سازد (رضایی و همکاران، ۱۳۹۷).

با توجه به اهمیت درستی و اعتبار نتایج حاصل از آمارگیری‌های نمونه‌ای، استفاده از روش‌های نمونه‌گیری که با صرف هزینه‌ای ناچیز منجر به افزایش کارایی براوردها می‌شوند می‌تواند در طراحی آمارگیری‌ها سودمند باشد، به‌ویژه زمانی که اندازه‌گیری صفت موردنظر در جامعه زمان‌بر بوده و نیازمند

صرف هزینهٔ زیادی باشد. نمونه‌گیری مجموعه‌رتبه‌ای[1] (RSS) یکی از روش‌هایی است که استفاده از آن اغلب منجر به براوردهایی کارا از پارامترهای مورد نظر در جامعه می‌شود.

هنگامی که هزینهٔ انتخاب واحدهای نمونه و رتبه‌بندی آن‌ها براساس اطلاعات موجود در مقایسه با هزینهٔ اندازه‌گیری مقدار واقعی صفت مورد نظر قابل چشم‌پوشی باشد نمونه‌گیری مجموعه رتبه‌دار با ارائهٔ براوردهای کاراتر به نمونه‌گیری تصادفی ساده ترجیح داده می‌شود. این روش نمونه‌گیری با استفاده از اندازهٔ یک صفت مناسب از واحدهای نمونه و رتبه‌بندی آن‌ها، انعطاف‌پذیری نمونه‌گیری قصدی در انتخاب نمونه‌ای معرف از جامعه را با امکان محاسبهٔ خطای نمونه‌گیری و ارزیابی براوردها در نمونه‌گیری تصادفی ساده تلفیق می‌کند. نمونه‌گیری مجموعه‌رتبه‌ای که برای اولین بار توسط مک اینتایر (۱۹۵۲) معرفی شد در دو دههٔ اخیر مورد توجه آمارشناسان و محققان بسیاری قرار گرفت و روش‌های نمونه‌گیری مشابهی با استفاده از ایدهٔ اصلی مک اینتایر معرفی شد. برای مروری بر این روش‌ها به‌عنوان مثال به کاور و همکاران (۱۹۹۵) یا چن و همکاران (۲۰۰۴) مراجعه نمایید.

بسیاری از تحقیقات در خصوص نمونه‌گیری مجموعه‌رتبه‌ای بر براورد پارامترهای جوامع نامتناهی تمرکز دارند و اغلب از روش‌های مدل‌یار یا مدل‌مبنا برای استنباط در مورد پارامترهای جامعه بهره جسته‌اند، این در حالیست که در عمل، اغلبِ جوامع مورد بررسی، جوامعی متناهی هستند. عملکرد نمونه‌گیری مجموعه‌رتبه‌ای در براورد میانگین یک جامعه متناهی، نخستین بار توسط تاکاهاسی و فوت‌سویا (۱۹۸۸) بررسی شد. آن‌ها نشان دادند تحت شرایطی خاص، براورد معرفی‌شده براوردی نااریب برای میانگین جامعه است و کارایی این براورد را برای اندازه نمونه $n = ۲$ به دست آوردند. آنها همچنین نشان دادند دقت نسبی براورد میانگین جامعهٔ متناهی در نمونه‌گیری مجموعه‌رتبه‌ای بر اساس نمونه‌گیری بدون جایگذاری بیش از براورد میانگین جامعه در نمونه‌گیری تصادفی ساده است. پاتیل و همکاران (۱۹۹۵) نتیجهٔ بررسی آن‌ها را به جوامع متناهی بزرگتر در حالت کلی تعمیم دادند. پس از آن مطالعات دیگری نیز در خصوص نمونه‌گیری مجموعه‌رتبه‌ای و کاربردهای آن در جوامع متناهی صورت گرفت که از آن جمله می‌توان به دیشپند و همکاران (۲۰۰۶) و بوزا (۲۰۰۹) اشاره کرد. الصالح و سماوی (۲۰۰۷) و ازدمیر و گوک‌پینار (۲۰۰۷) احتمال شمول واحدهای جامعهٔ متناهی را در برخی روش‌های نمونه‌گیری مجموعه‌رتبه‌ای معرفی و با نمونه‌گیری تصادفی ساده با اندازه نمونه مشابه مقایسه کردند. جعفری جوزانی و جانسون (۲۰۱۰) به بررسی روش‌های مختلف نمونه‌گیری مجموعه‌رتبه‌ای در جوامع متناهی که توسط دیشپند و همکاران (۲۰۰۶) توصیف شده بودند، پرداختند و برای براورد میانگین یا مجموع صفت مورد نظر در جوامع متناهی از یک رویکرد طرح‌مبنا

[1] Ranked set sampling

استفاده کردند.

در این مقاله پس از اشاره‌ای کوتاه به روش نمونه‌گیری مجموعه‌رتبه‌ای اولیه و کارایی آن، از نمونه‌گیری مجموعه‌رتبه‌ای معرفی‌شده توسط دیشپند و همکاران (۲۰۰۶) برای نمونه‌گیری از خوشه‌های طرح آمارگیری هزینه و درآمد خانوار مرکز آمار ایران در سال ۱۳۹۷ استفاده می‌شود تا افزایش کارایی حاصل از این روش نمونه‌گیری نسبت به روش مورد استفاده در طرح فعلی بررسی گردد. برای این منظور ابتدا در بخش ۲، روش نمونه‌گیری مجموعه‌رتبه‌ای که توسط مک اینتایر معرفی شده است ارائه می‌شود. در بخش ۳ به‌صورت خاص، راهکار استفاده از نمونه‌گیری مجموعه‌رتبه‌ای در جوامع متناهی تشریح می‌شود. بخش ۴ به توضیحاتی در خصوص آمارگیری هزینه و دارمد خانوار مرکز آمار ایران اختصاص یافته است. به‌کارگیری روش نمونه‌گیری مجموعه‌رتبه‌ای نیازمند اطلاعاتی برای رتبه‌بندی واحدهای نمونه است. روش پیشنهادی برای رتبه‌بندی واحدهای نمونه در بخش ۵ ارائه می‌شود. در نهایت در بخش ۶ با انجام یک مطالعه شبیه‌سازی بر اساس داده‌های واقعی آمارگیری هزینه و درآمد خانوار سال ۱۳۹۷ به براورد کارایی روش نمونه‌گیری پیشنهادی پرداخته می‌شود.

۲ نمونه‌گیری مجموعه‌رتبه‌ای

برای به دست آوردن یك نمونهٔ مجموعه‌رتبه‌ای به اندازهٔ $n = mn'$ فرض کنید $X_{j1}, \ldots, X_{jn'}$، نمونهٔ تصادفی سادهٔ باجایگذاری به اندازهٔ n'^2 از X (جامعهٔ موردنظر) با تابع چگالی $f(\cdot)$، میانگین μ و واریانس σ^2 باشد. فرایند انتخاب نمونه به‌صورت زیر است

- انتخاب n'^2 نمونهٔ تصادفی بدون جایگذاری از جامعه و تخصیص نمونه‌ها به n' زیرمجموعهٔ هم‌اندازه به‌صورت تصادفی
- رتبه‌بندی واحدهای هر زیرمجموعه براساس صفت کمکی مورد نظر
- انتخاب و اندازه‌گیری کوچک‌ترین واحد از اولین زیرمجموعهٔ n'تایی، انتخاب و اندازه‌گیری دومین واحد کوچک از دومین زیرمجموعهٔ n'تایی، ادامهٔ این فرایند تا انتخاب و اندازه‌گیری بزرگ‌ترین واحد از n'امین زیر مجموعهٔ n'تایی
- در صورت لزوم، تکرار فرایند نمونه‌گیری در m چرخه برای رسیدن به نمونه‌ای به‌اندازهٔ $n = n'm$. به‌عنوان مثال نمونه‌گیری مجموعهٔ رتبه‌دار با اندازه $n = 3m$ با m چرخه به صورت زیر است.

$$
\text{چرخه (۱)} \begin{cases} X_{11}^{(1)} \le X_{11}^{(2)} \le X_{11}^{(3)} & \Rightarrow X_{11}^{(1)} \\ X_{21}^{(1)} \le X_{21}^{(2)} \le X_{21}^{(3)} & \Rightarrow X_{21}^{(2)} \\ X_{31}^{(1)} \le X_{31}^{(2)} \le X_{31}^{(3)} & \Rightarrow X_{31}^{(3)} \end{cases}
$$

$$
\text{چرخه (۲)} \begin{cases} X_{12}^{(1)} \le X_{12}^{(2)} \le X_{12}^{(3)} & \Rightarrow X_{12}^{(1)} \\ X_{22}^{(1)} \le X_{22}^{(2)} \le X_{22}^{(3)} & \Rightarrow X_{22}^{(2)} \\ X_{32}^{(1)} \le X_{32}^{(2)} \le X_{32}^{(3)} & \Rightarrow X_{32}^{(3)} \end{cases}
$$

$$
\dots \quad \dots \quad \dots \quad \dots \quad \dots \quad \dots \quad \dots
$$

$$
\text{چرخه (m)} \begin{cases} X_{1m}^{(1)} \le X_{1m}^{(2)} \le X_{1m}^{(3)} & \Rightarrow X_{1m}^{(1)} \\ X_{2m}^{(1)} \le X_{2m}^{(2)} \le X_{2m}^{(3)} & \Rightarrow X_{2m}^{(2)} \\ X_{3m}^{(1)} \le X_{3m}^{(2)} \le X_{3m}^{(3)} & \Rightarrow X_{3m}^{(3)} \end{cases}
$$

اگر $X_{ij}^{(i)}$ مقدار مرتب‌شده iام $(i = 1, \dots, n')$ از iامین نمونهٔ تصادفی به اندازهٔ n' در jامین چرخه $(j = 1, 2, \dots, m)$ باشد، آنگاه نمونهٔ

$$
X_{11}^{(1)}, \dots, X_{n'1}^{(n')}; X_{12}^{(1)}, \dots, X_{n'2}^{(n')}; X_{1m}^{(1)}, \dots, X_{n'm}^{(n')}
$$

یک نمونهٔ مجموعه‌رتبه‌ای به اندازهٔ $n = mn'$ با m چرخه (هر یک به اندازهٔ n') است. براورد میانگین جامعه با استفاده از نمونه‌گیری مجموعه‌رتبه‌ای به‌صورت (تاکاهاسی و واکی‌موتو، ۱۹۹۸)

$$
\hat{\mu}_{RSS} = \frac{1}{mn'} \sum_{j=1}^{m} \sum_{i=1}^{n'} X_{ij}^{(i)}
$$

معرفی شده است. تاکاهاسی و واکی موتو (۱۹۶۸) نشان دادند $\hat{\mu}_{RSS}$ براوردگری ناااریب برای میانگین جامعه μ است و واریانس آن عبارت است از:

$$
\mathrm{Var}(\hat{\mu}_{RSS}) = \frac{1}{mn'^2} \sum_{i=1}^{n'} \sigma_{(i)}^2 = \frac{\sigma^2}{mn'} - \frac{1}{mn'^2} \sum_{i=1}^{n'} (\mu_i - \mu)^2
$$

که در آن $\mu_{(i)}$ و $\sigma^{۲}_{(i)}$ به ترتیب میانگین و واریانس i امین آماره مرتب در نمونه‌ای تصادفی به اندازهٔ n' هستند. آن‌ها همچنین نشان دادند اگر $\hat{\mu}_{RSS}$ و $\hat{\mu}^{w}_{SRS}$ به ترتیب براورد میانگین جامعه بر اساس نمونه‌گیری مجموعه‌رتبه‌ای و نمونه‌گیری تصادفی سادهٔ با جایگذاری باشند، آنگاه

$$۱ \leq \frac{\mathrm{Var}(\hat{\mu}^{w}_{SRS})}{\mathrm{Var}(\hat{\mu}_{RSS})} \leq \frac{n'+۱}{۲}$$

که در آن n' اندازهٔ هر یك از چرخه‌ها در نمونه‌گیری مجموعه‌رتبه‌ای است.

۳ نمونه‌گیری مجموعه‌رتبه‌ای در جوامع متناهی

جامعهٔ متناهی $U = \{۱, \ldots, N\}$ با مقادیر ثابت (اما نامعلوم) صفت مورد نظر Y و صفت کمکی X را در نظر بگیرید (نظریه طرح‌مبنا). انتخاب نمونه به روش نمونه‌گیری مجموعه‌رتبه‌ای از جامعه متناهی ثابت برای نخستین بار توسط پاتیل و همکاران (۱۹۹۵) برای براورد میانگین جامعه معرفی شد. آن‌ها انتخاب نمونه‌ها از جامعه را بدون جایگذاری در نظر گرفتند تا همبستگی منفی بین واحدهای انتخاب‌شده موجب کاهش تغییرپذیری براورد و افزیش کارایی آن شود. همین موضوع منجر به محدودیت‌هایی در تعیین اندازهٔ نمونه و کارایی ناشی از آن شد (صبا، ۱۳۹۷). پس از آن روش‌های مختلفی برای انتخاب نمونه و محاسبه براوردها توسط آمارشناسان مختلف ارائه شد. روش مورد استفاده در این مقاله، نمونه‌گیری ارائه‌شده توسط دیشپند و همکاران (۲۰۰۶) در سطح صفر آن است. دیشپند و همکارانش پیشنهاد دادند در هر گام از نمونه‌گیری، نمونه‌ای به روش تصادفی سادهٔ بدون جایگذاری از جامعهٔ متناهی انتخاب شود، سپس بعد از رتبه‌بندی و انتخاب واحد مورد نظر برای اندازه‌گیری، همه واحدهای نمونهٔ انتخاب‌شده به جامعه بازگردانده شوند (چه واحد انتخاب‌شده برای اندازه‌گیری و چه واحدهای انتخاب‌شده برای رتبه‌بندی). جعفری جوزانی و جانسون (۲۰۱۰) نشان دادند استفاده از این روش نمونه‌گیری بر پایه نظریه طرح‌مبنا در جوامع متناهی منجر به براوردگرهای دقیق‌تری برای میانگین یا مجموع جامعه نسبت به نمونه‌گیری تصادفی ساده حتی با اندازه نمونه کم می‌شود. از این رو در این بخش از مقاله، روش نمونه‌گیری مجموعه‌رتبه‌ای دیشپند و همکاران (۲۰۰۶) در سطح صفر برای بهبود کارایی براوردهای اصلی آمارگیری هزینه و درآمد خانوار سال ۱۳۹۷مورد بررسی قرار می‌گیرد. بدین منظور در بخش بعد، روش نمونه‌گیری مورد استفاده در آمارگیری هزینه و درآمد خانوار سال ۱۳۹۷ به‌طور خلاصه مورد بررسی قرار می‌گیرد. سپس در بخش پنجم پیشنهاد انتخاب نمونه‌ها به روش نمونه‌گیری مجموعه‌رتبه‌ای ارائه می‌شود. بر اساس آنچه گفته شد،

به‌کارگیری روش نمونه‌گیری مجموعه‌رتبه‌ای نیازمند اطلاعاتی برای رتبه‌بندی واحدهای نمونه است. روش پیشنهادی برای رتبه‌بندی واحدهای نمونه در بخش ۵ ارائه می‌شود.

۴ روش نمونه‌گیری در آمارگیری هزینه و درآمد خانوار

آمارگیری از هزینه و درآمد خانوارها یکی از قدیمی‌ترین آمارگیری‌های مرکز آمار ایران است که سابقه اجرای آن به بیش از ۵۰ سال پیش بازمی‌گردد. در طرح نمونه‌گیری آمارگیری از هزینه و درآمد خانوارهای شهری و روستایی سال ۱۳۹۷، برای به دست آوردن براورد سطوح و تغییرات پارامترها با دقت مطلوب، از نمونه‌گیری چرخشی استفاده شده است که برای انجام بهینه آن نمونه‌ها از نمونه پایه انتخاب شده‌اند. نمونه پایه، نمونه بزرگی است که می‌توان از آن برای تأمین نیازهای چند آمارگیری یا چند دوره از یك آمارگیری، زیرنمونه‌هایی انتخاب کرد. نمونه‌های آمارگیری از هزینه و درآمد خانوارهای شهری و روستایی سال ۱۳۹۷ طی سه مرحله و با استفاده از روش‌های نمونه‌گیری احتمالی انتخاب شده‌اند. در مرحله اول، خوشه‌های نمونه در هر یك از طبقات نمونه پایه، به روش سیستماتیك انتخاب شده‌اند. در مرحله دوم در هر خوشه نمونه، گروه‌های چرخش شش خانواری ساخته شده و سه گروه چرخش بر اساس الگوی چرخش، برای آمارگیری در هرسال تعیین شده‌اند. در مرحله سوم، داخل هر یك از گروه‌های چرخش، دو خانوار به روش نمونه‌گیری تصادفی ساده به‌عنوان خانوار نمونه اصلی انتخاب شده‌اند و چهار خانوار باقیمانده در گروه چرخش به‌عنوان خانوارهای جایگزین در نظر گرفته شده‌اند. بر اساس راهنمای آمارگیری، در برخی شرایط خاص در هر گروه چرخش امکان استفاده از خانوارهای جایگزین به جای خانوارهای نمونه فراهم بوده است. هر خوشهٔ نمونه در آمارگیری سال ۱۳۹۷ شامل یك بلوك/آبادی، بخشی از یك بلوك/آبادی بزرگ یا در مواردی مجموعه‌ای از چند بلوك/آبادی کوچك است که بر اساس اطلاعات سرشماری عمومی نفوس و مسکن سال ۱۳۹۵ به‌گونه‌ای ساخته شده است که در مناطق شهری حداقل شامل ۳۵۰ خانوار باشد. در نقاط روستایی نیز اندازهٔ خوشه‌ها به‌گونه‌ای در نظر گرفته شده است که خانوارهای نمونه در طرح‌های آمارگیری نیروی کار و هزینه و درآمد خانوار را در طول نمونه پایه، تأمین کند (دفتر روش‌شناسی آماری و طرح‌های نمونه‌گیری مرکز آمار ایران، ۱۳۹۷).

با توجه به طرح نمونه‌گیری استفاده‌شده در آمارگیری هزینه و درآمد خانوار سال ۱۳۹۷، خانوارهای نمونه در داخل هر یک از خوشه‌های منتخب برای چند سال پس از طراحی نمونهٔ پایه قبلاً توسط دفتر روش‌شناسی آماری و طرح‌های نمونه‌گیری انتخاب شده است. با این وجود در این پژوهش یک طرح نمونه‌گیری کاراتر برای اجرای آمارگیری در سال‌های آتی پیشنهاد می‌شود. در روش پیشنهادی، انتخاب

خوشه‌های نمونه (واحدهای نمونه‌گیری مرحله اول) از نمونه پایه به روش نمونه‌گیری مجموعه‌رتبه‌ای صورت می‌گیرد. از آنجایی که بر اساس طرح نمونه‌گیری کنونی مرکز آمار ایران، داخل هر یک از طبقات نمونه پایه، پیش از انتخاب سیستماتیک خوشه‌ها، چارچوب خوشه‌ها بر اساس متغیرهای کلیدی آمارگیری مرتب نمی‌شوند، فرض می‌شود بین انتخاب سیستماتیک خوشه‌ها و انتخاب خوشه‌ها به روش تصادفی ساده در داخل طبقات اختلاف چندانی وجود ندارد. با تکیه بر این فرض می‌توان نتیجه انتخاب خوشه‌ها به روش نمونه‌گیری مجموعه‌رتبه‌ای را داخل هر یک از طبقات نمونه پایه با انتخاب تصادفی سادهٔ خوشه‌ها در طبقات مقایسه کرد. برای نزدیک بودن طرح نمونه‌گیری پیشنهادی به طرح کنونی مرکز آمار ایران، از براورد هورویتز تامپسون برای براورد پارامترها استفاده خواهد شد. با توجه به نااریب بودن این براورد در هر دو روش نمونه‌گیری، از کارایی نسبی براوردها برای مقایسه روش‌ها استفاده خواهد شد.

۵ انتخاب خوشه‌ها به روش نمونه‌گیری مجموعه‌رتبه‌ای

شیرکوند در سال ۱۳۹۲ ضمن بررسی روش‌های براورد مختلف در نمونه‌گیری مجموعه‌رتبه‌ای از جوامع متناهی، موضوع انتخاب خوشه‌های نمونه در آمارگیری از هزینه و درآمد خانوار سال ۱۳۹۰ به روش نمونه‌گیری مجموعه‌رتبه‌ای را بررسی کرد. مطالعه ایشان که بر اساس یک روش نمونه‌گیری خودگردان بر پایه نمونه‌های آمارگیری از هزینه و درآمد خانوارهای شهری سال ۱۳۹۰ در استان تهران بود، برتری استفاده از نمونه‌گیری مجموعه‌رتبه‌ای را نسبت به نمونه‌گیری تصادفی ساده در انتخاب واحدهای نمونهٔ مرحلهٔ اول نشان داد. همان‌گونه که در بخش‌های قبل اشاره شد برای استفاده از نمونه‌گیری مجموعه‌رتبه‌ای وجود اطلاعاتی مناسب مانند متغیری کمکی که همبستگی بالایی با متغیرهای اصلی دارد برای رتبه‌بندی واحدهای آماری مورد نیاز است. شیرکوند (۱۳۹۲) از اطلاع درآمد مورد انتظار خانوارها به‌عنوان متغیر رتبه‌بندی استفاده کرده است. این متغیر که با استفاده از اطلاعات خانوارها و امکانات و تسهیلات و ویژگی‌های آن‌ها در چارچوب نمونه‌گیری برای هریک از خانوارهای جامعه ساخته می‌شود توسط زاهدیان و همکاران (۱۳۹۲) پیشنهاد شد. آن‌ها نشان دادند متغیر درآمد مورد انتظار خانوارها همبستگی بالایی با متغیر درآمد خانوارها دارد، که اطلاع آن در آمارگیری از هزینه و درآمد خانوارها جمع‌آوری می‌شود.

۵.۱ رتبه‌بندی خوشه‌ها

هزینه و درآمد یک خانوار از کلیدی‌ترین متغیرهایی است که سطح اقتصادی آن خانوار را نشان می‌دهد. برای ساخت متغیری همبسته با متغیرهای هزینه و درآمد خانوارها که امکان دسترسی به آن برای واحدهای نمونه مراحل مختلف (خانوار، خوشه، ...) در چارچوب نمونه‌گیری آمارگیری هزینه و درآمد خانوار نیز فراهم باشد از داده‌های سرشماری عمومی نفوس و مسکن ۱۳۹۵ استفاده شده است. به دلیل نبود امکان دسترسی به نمونه پایه خانواری مرکز آمار ایران و نیز چارچوب کامل سرشماری ۱۳۹۵، از اطلاعات خانوارها در فایل دو درصد سرشماری ۱۳۹۵ استفاده شده است که نمونه‌ای بیش از ۴۸۲ هزار خانوار معمولی ساکن و گروهی از کل خانوارهای کشور را در بر دارد (مرکز آمار ایران، ۱۳۹۵، الف). روش نمونه‌گیری در فایل دو درصد سرشماری، در سطح هر منطقه‌ی شهری / روستایی شهرستان‌های کشور به‌صورت تصادفی ساده طبقه‌بندی شده در نظر گرفته شده است. طبقه‌بندی در سطح هر شهرستان بر اساس نقاط شهری / روستایی شهرستان انجام شده است (مرکز آمار ایران، ۱۳۹۵، ب).

ساخت متغیر رتبه‌بندی بر پایه داده‌های فایل دو درصد سرشماری، با استفاده از آزمون تقریب وسع[1] (PMT) انجام شده است. این آزمون یکی از پرکاربردترین روش‌های گروه‌بندی خانوارها از نظر سطح اقتصادی آن‌ها محسوب می‌شود. با استفاده از آزمون تقریب وسع در گام نخست ساخت متغیر رتبه‌بندی، درآمد مورد انتظار خانوارها با برازش یک مدل رگرسیونی به دست آمده است. برای این منظور ابتدا با استفاده از داده‌های آمارگیری از هزینه و درآمد خانوار ۱۳۹۵ متغیرهایی که با درآمد خانوار همبستگی بیشتری دارند و اطلاعات آن‌ها از داده‌های آمارگیری در دسترس است، انتخاب شدند. با توجه به اهداف بررسی، متغیرهایی برای براورد درآمد و رتبه‌بندی خانوارها انتخاب شدند که اطلاعات آن‌ها در سرشماری ۱۳۹۵ نیز گردآوری شده است. متغیرهای شهری یا روستایی بودن محل سکونت، سن سرپرست خانوار، مدرک تحصیلی سرپرست خانوار، وضع زناشویی سرپرست خانوار، تعداد افراد باسواد در خانوار، تعداد محصلین خانوار، تعداد شاغلین خانوار، سرانه سطح زیربنای محل سکونت خانوار، و نحوه تصرف محل سکونت برای براورد درآمد خانوارها انتخاب شدند. با برازش یک مدل رگرسیونی خطی با عرض از مبدأ بر روی داده‌های آمارگیری از هزینه و درآمد خانوار ۱۳۹۵ در هر استان درآمد سرانه مورد انتظار خانوارها در این سال براورد شد. سپس با استفاده از اطلاع تعداد اعضای خانوار، درآمد مورد انتظار خانوارهای نمونه نیز براورد شد. درآمد مورد انتظار براوردشده ضریب همبستگی قابل توجهی با متغیرهای کلیدی آمارگیری هزینه و درآمد خانوار دارد که مقادیر آن در جدول ۱ ارائه شده است.

[1] Proxy Means Test

جدول ۰۱. همبستگی درآمد مورد انتظار براوردشده و متغیرهای کلیدی آمارگیری هزینه و درآمد خانوار

مقدار	ضریب همبستگی درآمد مورد انتظار براوردشده				
	درآمد خانوار	هزینه کل خالص خانوار	هزینه کل ناخالص خانوار	هزینه خوراک خانوار	هزینه غیرخوراک خانوار
ضریب همبستگی	۰٫۶۵	۰٫۵۸	۰٫۵۷	۰٫۳۶	۰٫۵۵

پس از اطمینان از عملکرد مطلوب مدل رگرسیونی، این مدل بر داده‌های فایل دو درصد سرشماری عمومی نفوس و مسکن ۱۳۹۵ برازش داده شد و تقریبی از درآمد خانوارهای این فایل به دست آمد. از آنجایی که بر اساس طرح نمونه‌گیری آمارگیری هزینه و درآمد خانوار ۱۳۹۷ و با توجه به چارچوب‌سازی از طریق فهرست‌برداری در این آمارگیری امکان برقراری ارتباط دقیق بین خانوارهای نمونه و خانوارهای سرشماری ۱۳۹۵ در نقاط شهری کشور وجود نداشت مقرر شد میانگین درآمد مورد انتظار خانواهای فایل دو درصد سرشماری به خوشه‌های نمونه در آمارگیری سال ۱۳۹۷ منتسب گردند.

فایل‌های خلاصه آمارگیری هزینه و درآمد خانوار در نقاط شهری اطلاعات ۲۰۳۵۰ خانوار و در نقاط روستایی اطلاعات ۱۸۶۱۰ خانوار را در برداشت که با تجمیع این دو فایل اطلاعات ۳۸۹۶۰ خانوار نمونه در سال ۱۳۹۷ در دسترس بود. همانگونه که اشاره شد متغیر درآمد مورد انتظار تنها برای خانوارهایی قابل براورد بود که اطلاعات آن‌ها در فایل دو درصد سرشماری ۱۳۹۵ وجود داشت. در نهایت، اشتراک مجموعهٔ این خانوارها با نمونه‌های آمارگیری هزینه و درآمد خانوار ۱۳۹۷ شامل ۲۶۵۶۳ خانوار بود (۲۰۳۵۰ خانوار شهری و ۶۲۱۳ خانوار روستایی). با میانگین‌گیری از درآمد مورد انتظار خانوارها در سطح خوشه‌های آمارگیری از هزینه و درآمد ۱۳۹۷ مقدار میانگین درآمد مورد انتظار خانوارها برای حدود ۸۸ درصد خوشه‌ها به دست آمد. اطلاعات سایر خوشه‌ها با جانهی اطلاع میانگین در طبقه یا میانگین در سطح مناطق شهری و روستایی کشور تکمیل شد. به عبارت دیگر با اجرای این فرایند، اطلاع میانگین درآمد مورد انتظار خانوارها برای تمام خوشه‌های نمونه در آمارگیری سال ۱۳۹۷ به دست آمد. با توجه به داده‌های در دسترس، مقرر شد از اطلاع میانگین درآمد مورد انتظار خوشه‌ها برای رتبه‌بندی هر خوشه استفاده شود و نمونه‌گیری مجموعه‌رتبه‌ای برای انتخاب خوشه‌های نمونه در آمارگیری هزینه و درآمد خانوار مورد بررسی قرار گیرد.

۵.۲ انتخاب خوشه‌ها

در داخل هر یک از طبقات، نمونه‌ای با اندازه ۱۵ خوشه بر اساس روش توضیح داده شده در بخش ۳ با استفاده از متغیر میانگین درآمد مورد انتظار خانوارها در خوشه‌ها انتخاب شد. به عبارت دیگر از خوشه‌های چارچوب مورد بررسی در هر طبقه ۱۵ نمونه تصادفی ساده بدون جایگذاری انتخاب شد سپس در هر نمونه، ۱۵ خوشه منتخب بر اساس متغیر رتبه‌بندی (میانگین درآمد مورد انتظار خانوارهای خوشه) مرتب شدند و با الگوی ارائه‌شده در بخش ۲ نمونه‌ای به روش RSS با یک چرخه از آن‌ها انتخاب شد. مقایسه نتایج حاصل با نتایج روش تصادفی ساده در بخش بعد توضیح داده شده است.

شایان ذکر است بر اساس طرح نمونه‌گیری هزینه و درآمد خانوار سال ۱۳۹۷، انتخاب خوشه‌ها در این طرح به روش سیستماتیک انجام شده است. مانند بسیاری از تحقیقات مشابه با فرض عدم مرتب بودن چارچوب خوشه‌ها، در این مطالعه، نمونه‌گیری تصادفی ساده و نمونه‌گیری سیستماتیک روش‌هایی با کارایی مشابه فرض شده‌اند. از این رو نتایج نمونه‌گیری مجموعه‌رتبه‌ای با نمونه‌گیری تصادفی ساده در انتخاب خوشه‌ها مقایسه شده است نه نمونه‌گیری سیستماتیک.

۶ کارایی نمونه‌گیری مجموعه‌رتبه‌ای در آمارگیری هزینه و درآمد خانوار

در این بخش، کارایی نمونه‌گیری مجموعه‌رتبه‌ای در انتخاب واحدهای نمونه‌گیری مرحله اول (خوشه‌ها) در آمارگیری هزینه و درآمد خانوار مورد بررسی قرار می‌گیرد. برای این منظور از روش نمونه‌گیری خودگردان استفاده می‌شود و زیرنمونه‌هایی به روش نمونه‌گیری مجموعه‌رتبه‌ای از نمونهٔ اصلی انتخاب می‌شوند. فرایند نمونه‌گیری ۱۰۰۰ مرتبه تکرار می‌شود و براوردی از کارایی براوردهای اصلی طرح با استفاده از مقدار نسبی میانگین مربعات خطای براوردها به دست می‌آید. از آنجایی که عملکرد نمونه‌گیری مجموعه‌رتبه‌ای در مقایسه با نمونه‌گیری تصادفی ساده (بدون جایگذاری) هم به اندازهٔ نمونه و هم به کسر نمونه‌گیری بستگی دارد و از آنجایی که در این پژوهش دسترسی به چارچوب خوشه‌ها امکان‌پذیر نیست و چارچوبِ در اختیار در برخی طبقه‌ها تنها دو خوشه نمونه را شامل می‌شود، خوشه‌های نمونه به جای انتخاب از چارچوب خوشه‌های هر طبقه، از چارچوب خوشه‌های نقاط شهری هر استان انتخاب شده‌اند تا محدودیت‌های دسترسی به داده‌ها در این پژوهش منجر به نتیجه‌گیری اشتباه نگردد. بر این اساس برای پیشگیری از ایجاد خطا در مقایسه دو روش نمونه‌گیری، زیرنمونه‌هایی با همان تعداد نمونه به روش نمونه‌گیری تصادفی ساده از خوشه‌ها نیز انتخاب شدند. سپس کارایی نمونه‌گیری مجموعه‌رتبه‌ای با نمونه‌گیری تصادفی ساده از خوشه‌های نمونه در

نقاط شهری هر استان مقایسه شد. فرایند انجام این شبیه‌سازی به شرح زیر است.

الف- پس از ساخت متغیر رتبه‌بندی، که توضیحات آن در بخش قبل ارائه شد، در نقاط شهری هر استان، زیرنمونه‌ای به اندازهٔ ۱۵ خوشه به روش نمونه‌گیری مجموعه‌رتبه‌ای از خوشه‌های نمونه آن استان انتخاب شد. برای فراهم شدن امکان مقایسه، با ثابت نگه داشتن تعداد نمونه، زیرنمونه‌ای به اندازهٔ ۱۵ خوشه به روش تصادفی ساده از خوشه‌های نمونه در نقاط شهری هر استان نیز انتخاب شد. بر این اساس تعداد کل خوشه‌های نمونه در نقاط شهری ۴۶۵ خوشه در نظر گرفته شد.

ب- با توجه به ماهیت جامعه (متناهی) و استفاده از روش‌های طرح‌پایه در این آمارگیری، وزن‌های نمونه‌گیری خوشه‌ها برای هر دو شیوه نمونه‌گیری محاسبه شدند. برای محاسبه وزن‌ها در نمونه‌گیری مجموعه‌رتبه‌ای از روابط ارائه‌شده توسط جعفری جوزانی و جانسون (۲۰۱۰) و برای محاسبه وزن‌های نمونه‌گیری در روش تصادفی ساده، از طرح نمونه‌گیری آمارگیری هزینه و درآمد خانوار ۱۳۹۷ استفاده شده است. شایان ذکر است از آنجایی که در آمارگیری هزینه و درآمد خانوار وزن‌های نمونه‌گیری برای همه خانوارهای نمونه در داخل یک خوشه یکسان است، براوردهای نهایی را می‌توان تابعی از مشخصه‌های خوشه‌های نمونه (میانگین خوشه و وزن یکسان خانوارها در خوشه) در نظر گرفت.

ج- پس از تعیین وزن‌های نمونه‌گیری، براوردهای هورویتز تامپسون میانگین متغیرهای کلیدی طرح شامل درآمد، هزینه کل خالص، هزینه کل ناخالص، هزینه خوراک و هزینه غیر خوراک خانوار به دست آمد.

د- این فرایند ۱۰۰۰ مرتبه تکرار شد تا امکان براورد کارایی روش‌های نمونه‌گیری فراهم گردد.

ه- کارایی نسبی براوردهای حاصل از نمونه‌گیری مجموعه‌رتبه‌ای خوشه‌ها نسبت به براوردهای نمونه‌گیری تصادفی ساده با استفاده از رابطه

$$RE(\hat{\mu}_k) = \frac{MSE_{SRS}(\hat{\mu}_k)}{MSE_{RSS}(\hat{\mu}_k)}$$

به دست آمد، که در آن $\hat{\mu}_k$ برای $k = ۱, \ldots, ۵$ میانگین متغیرهای کلیدی طرح است.

نتایج شبیه‌سازی در جدول ۲ نشان می‌دهد نمونه‌گیری مجموعه‌رتبه‌ای در براورد میانگین اغلب متغیرهای کلیدی طرح منجر به براوردهای کاراتری نسبت به نمونه‌گیری تصادفی ساده می‌شود. از آنجایی که در این مطالعه، متغیر رتبه‌بندی (درآمد مورد انتظار خانوار) همبستگی ضعیفی با هزینه خوراک خانوارها دارد (جدول ۱) استفاده از نمونه‌گیری مجموعه‌رتبه‌ای افزایشی در کارایی براورد میانگین این متغیر ایجاد نمی‌کند. بدیهی است در صورت وجود امکان براورد متغیری برای رتبه‌بندی دقیق‌تر خانوارها مقدار کارایی نسبی براوردهای نمونه‌گیری مجموعه‌رتبه‌ای افزایش خواهد داشت.

جدول ۲. کارایی نسبی نمونه‌گیری مجموعه‌رتبه‌ای نسبت به نمونه‌گیری تصادفی ساده در براورد میانگین متغیرهای کلیدی آمارگیری هزینه و درآمد خانوارهای شهری

مقدار کارایی نسبی	پارامتر مورد نظر				
	درآمد خانوار	هزینه کل خالص خانوار	هزینه کل ناخالص خانوار	هزینه خوراک خانوار	هزینه غیرخوراک خانوار
	۱٫۲	۱٫۳	۱٫۳	۱٫۰	۱٫۳

۷ بحث و نتیجه‌گیری

استفاده از روش نمونه‌گیری مجموعه‌رتبه‌ای در طراحی آمارگیری هزینه و درآمد خانوار می‌تواند منجر به بهبود قابل توجهی در کارایی براوردهای کلیدی طرح گردد به شرطی که متغیر رتبه‌بندی مورد استفاده همبستگی بالایی با متغیرهای اصلی طرح داشته باشد. در صورت نبود متغیری مناسب و در دسترس برای رتبه‌بندی واحدها می‌توان ابتدا از اطلاعات چارچوب برای ساخت متغیری همبسته با متغیرهای کلیدی مورد بررسی استفاده کرد. میزان بهبود کارایی حاصل از به‌کارگیری نمونه‌گیری مجموعه‌رتبه‌ای به میزان همبستگی متغیر رتبه‌بندی با متغیر اصلی مورد بررسی بستگی دارد. بر اساس نتایج شبیه‌سازی شده در این مقاله و با توجه به همه محدودیت‌های موجود در انجام شبیه‌سازی، به نظر می‌رسد در شرایطی مشابه آنچه در مقاله تشریح شده است کارایی نمونه‌گیری مجموعه‌رتبه‌ای در براورد اغلب پارامترهای کلیدی طرح (همه موارد به‌غیر از میانگین هزینه خوراک) حدود ۲۰ یا ۳۰ درصد بیش از نمونه‌گیری تصادفی ساده باشد. شایان ذکر است راهکار پیشنهادی برای رتبه‌بندی واحدها و استفاده از نمونه‌گیری مجموعه‌رتبه‌ای در این مقاله منجر به افزایش کارایی براوردهای کلیدی در طرح آمارگیری هزینه و درآمد خانوار مرکز آمار ایران شده است، در حالی که این افزایش کارایی باعث افزایش هزینه‌های آمارگیری یا پیچیدگی اجرای آمارگیری نخواهد شد و این موضوع نقطه قوتی برای راهکار ارائه شده در مقاله محسوب می‌گردد. به‌عبارت دیگر درصورت استفاده از نمونه‌گیری مجموعه‌رتبه‌ای هیچ هزینه‌ای به هزینه‌های مرحله اجرای آمارگیری اضافه نخواهد شد و تنها هزینه استفاده از این روش، هزینه کارشناسی تعریف متغیر رتبه‌بندی و انتخاب نمونه‌ها با روش RSS است. انتخاب نمونه‌ها با به‌کارگیری نرم‌افزارهای آماری زمانبر نبوده و به سرعت انجام می‌گیرد.

استفاده از نمونه‌گیری مجموعه‌رتبه‌ای، خطای غیر نمونه‌گیری خاصی به براوردها تحمیل نمی‌کند زیرا آمارگیری از واحدها طبق روال همیشه انجام می‌گیرد و فقط با به‌کارگیری این روش نمونه‌گیری، واحدهای نمونه منتخب، معرف بهتری از جامعه مورد بررسی خواهند بود. بدیهی است اگر استفاده از نمونه‌گیری

مجموعه‌رتبه‌ای با هدف کاهش تعداد نمونه‌ها برای رسیدن به کارایی روش نمونه‌گیری تصادفی ساده باشد حتی ممکن است خطاهای غیرنمونه‌گیری کاهش یابند. اگر هدف از به‌کارگیری این روش نمونه‌گیری افزایش کارایی با همان تعداد نمونه مورد نظر در آمارگیری باشد نیز به دلیل اضافه نشدن منبعی جدید برای بروز خطاها انتظار می‌رود خطای غیر نمونه‌گیری افزایش نیابد.

با توجه به اصول و ویژگی‌های نمونه‌گیری مجموعه‌رتبه‌ای انتظار می‌رود با دسترسی به چارچوب خانوارهای داخل خوشه‌ها در آمارگیری هزینه و درآمد خانوار و استفاده از نمونه‌گیری مجموعه‌رتبه‌ای برای انتخاب خانوارها در مرحله آخر نمونه‌گیری، افزایش بیشتری در کارایی براوردها ایجاد گردد. بدین منظور لازم است متغیرهای رتبه‌بندی مناسب در چارچوب آتی خانواری در نمونه‌گیری هزینه و درآمد خانوار ساخته شوند تا حداکثر بهره ممکن از به‌کارگیری نمونه‌گیری مجموعه‌رتبه‌ای در طراحی این آمارگیری محقق گردد.

درصورت دسترسی به شیوه محاسبه وزن‌های نمونه‌گیری، استفاده از روش‌های پیچیده‌تر نمونه‌گیری مجموعه‌رتبه‌ای برای افزایش بیشتر کارایی براودهای آمارگیری هزینه و درآمد خانوار قابل بررسی است. همچنین شایان ذکر است نتایج ارائه‌شده در این مقاله به شرایط شبیه‌سازی انجام‌شده بستگی دارد و آمارشناسان محترم آمارگیری هزینه و درآمد خانوار در صورت صلاحدید با دسترسی به داده‌های بیشتر می‌توانند شبیه‌سازی‌های دقیق‌تری نیز انجام دهند.

تقدیر و تشکر

نویسندگان مقاله از داوارن، سردبیر و ویراستار محترم مجله که نظرات و پیشنهادهای ارزشمندشان باعث ارتقای سطح مقاله شد صمیمانه تشکر و قدردانی می‌نماید.

مراجع

دفتر روش‌شناسی آماری و طرح‌های نمونه‌گیری مرکز آمار ایران (۱۳۹۷)، دستورالعمل طرح نمونه‌گیری آمارگیری از هزینه و درآمد خانوارهای شهری/روستایی- سال ۱۳۹۷.

رضایی قهرودی، ز.، علی‌اکبری صبا، ر.، زاهدیان، ع. (۱۳۹۷)، روش‌های آمارگیری. پژوهشکده‌ی آمار.

زاهدیان، ع.، علی‌اکبری صبا، ر.، دانش‌پرور، ن.، باقری، ف. (۱۳۹۲)، تعیین توزیع دو دهک پایین جامعه در مناطق جغرافیایی. پژوهشکده‌ی آمار.

شیرکوند، ز. (۱۳۹۲)، براورد طرح‌پایه نمونه‌گیری از مجموعهٔ رتبه‌دار در جامعه‌های متناهی. پایان‌نامه‌ی کارشناسی ارشد. دانشگاه علامه طباطبایی.

علی‌اکبری صبا، ر. (۱۳۹۷)، روش نمونه‌گیری مجموعه رتبه‌دار و برخی کاربردهای آن. رساله دکتری. دانشگاه آزاد اسلامی، واحد علوم و تحقیقات.

مرکز آمار ایران (۱۳۹۷)، دادگان آمارگیری از هزینه و درآمد خانوارهای شهری/روستایی- سال ۱۳۹۷.

مرکز آمار ایران (۱۳۹۵، الف)، داده‌های دو درصد سرشماری عمومی نفوس و مسکن سال ۱۳۹۵.

مرکز آمار ایران (۱۳۹۵، ب)، گزارش نمونه‌گیری از فایل سرشماری عمومی نفوس و مسکن سال ۱۳۹۵.

Al-Saleh, M. F. and Samawi, H. M. (2007), A Note on Inclusion Probability in Ranked Set Sampling and Some of Its Variations, *Test* , **16**, 198–209.

Bouza, C. N. (2009), Ranked Set Sampling and Randomized Response Procedures for Estimating the Mean of a Sensitive Quantitative Character, *Metrika* **70**, 267–277.

Chen, Z., Bai Z. and Sinha, B. (2004), *Ranked Set Sampling: Theory and Applications*, Lecture Notes in Statistics, Springer, New York.

Deshpande, J. V., Frey, J. and Ozturk, O. (2006), Nonparametric Ranked-Set Sampling Confidence intervals for Quantiles of a Fnite Population. *Environmental and Ecological Statistics*, **13**, 25–40.

Jafari Jozani, M. and Johnson, B. C. (2010), Design Based Estimation for Ranked Set Sampling in Finite Population, *Environmental and Ecological Statistics*, **18**, 663-685.

Kaur, A., Patil, G., Sinha A. and Taillie, C. (1995), Ranked Set Sampling: An Annotated Bibliography, *Environmental and Ecological Statistics*, **2**, 25–54.

McIntyre, G. A. (1952), A Method for Unbiased Selective Sampling Using Ranked Sets, *Australian Journal of Agricultural Research*,**3**, 385– 390.

Ozdemir, Y. A. and Gokpinar, F. (2007), A Generalized Formula for Inclusion Probabilities in Ranked Set Sampling, *Hacettepe Journal of Mathematics and Statistics*, **36**, 89–99.

Patil, G. P., Sinha, A. K. and Taillie, C. (1995), Finite Population Corrections for Ranked Set Sampling, *Annals of the Institute of Statistical Mathematics*, **47**, 621–636.

Takahasi, K. and Futatsuya, M. (1988), Ranked Set Sampling from a Fnite Population, *Proceedings of the Institute of Statistical Mathematics*, **36**, 55–68.

Takahasi, K. and Futatsuya, M. (1998), Dependence Between Order Statistics in Samples from Fnite Population and Its Application to Ranked Set Sampling, *Annals of the Institute of Statistical Mathematics*, **50**, 49–70.

Takahasi, K. and Wakimoto, K. (1968), On Unbiased Estimates of the Population Mean Based on the Sample Stratified by Means of Ordering, *Annals of the Institute of Statistical Mathematics*, **21**, 249-255.

Journal of Statistical Sciences, Autumn and Winter, 2021
Vol. 15, No. 2, pp 533-548
DOI: 10.29252/jss.15.2.533

A Suggestion for Using Ranked Set Sampling in Household Expenditure and Income Survey

Aliakbari، Saba R.[1]، Ebrahimi، N.[2]، Kalhori Nadrabadi، L.[1] & Abbasi، A.[2].
[1] Statistical Research and Training Center, Tehran, Iran.
[2] Statistical Center of Iran, Tehran, Iran.

Abstract: The ranked set sampling method uses the ranking information of the units to provide a more representative sample of the population to the survey designers. The sampling distribution is closer to the actual distribution of the population. In this article, to ensure the effectiveness of ranked set sampling in extensive surveys conducted to prepare official statistics, we intend to use this sampling method to improve the efficiency of key estimates of household expenditure and income survey of the Statistics Center of Iran. The results show that using ranked set sampling to design household expenditure and income surveys can improve the efficiency of key estimates of the study, provided that the ranking variable used has a high correlation with the main variables of the study. Obviously, in the absence of a suitable and available variable for ranking the units, the information of the sampling frame can be used to construct a ranking variable correlated with the key variables of the survey.

Keywords: Household expenditure and income survey, ranked set sampling, expected income.

Mathematics Subject Classification (2010): 62J07, 62D05.