

روش‌های نمونه‌گیری از مجموعه‌ی رتبه‌دار استوار و کاربرد آن‌ها در برآورد هزینه‌ی مسکن خانوارهای روستایی استان مازندران در سال ۱۳۹۲

میرحامد هاشمی‌مجد^{†*} و روشنگ علی‌اکبری صبا[‡]

[†] سازمان مدیریت و برنامه‌ریزی استان مازندران
[‡] پژوهشکده‌ی آمار

چکیده: نمونه‌گیری از مجموعه‌ی رتبه‌دار (RSS^۱) از روش‌های شناخته‌شده‌ای است که برتری آن نسبت به روش نمونه‌گیری تصادفی ساده توسط آمارشناسان نشان داده شده است. با ایجاد تغییراتی در ساختار RSS می‌توان دقت و کارایی آن را در برآورد پارامترهای مرکزی جامعه، افزایش داد. از این‌رو در سال‌های اخیر آمارشناسان انواع مختلفی از این روش، مانند روش‌های نمونه‌گیری از مجموعه‌ی رتبه‌دار استوار را معرفی کرده‌اند. استفاده از این روش‌ها سبب افزایش هرچه بیش‌تر کارایی برآوردگرهای مرکزی جامعه می‌شود. در این مقاله، با هدف شناسایی روشی کاراتر در برآورد هزینه‌ی مسکن خانوارهای روستایی استان مازندران، کارایی روش‌های نمونه‌گیری از مجموعه‌ی رتبه‌دار کرانگین استوار و نمونه‌گیری از مجموعه‌ی رتبه‌دار کرانگین استوار دوگانه نسبت به نمونه‌گیری تصادفی ساده مورد بررسی قرار می‌گیرد.

واژگان کلیدی: نمونه‌گیری از مجموعه‌ی رتبه‌دار؛ نمونه‌گیری از مجموعه‌ی رتبه‌دار کرانگین استوار؛ نمونه‌گیری از مجموعه‌ی رتبه‌دار کرانگین استوار دوگانه؛ هزینه‌ی مسکن؛ کارایی.

۱- مقدمه

یکی از متداول‌ترین روش‌ها برای انتخاب نمونه از جامعه، روش نمونه‌گیری تصادفی ساده است. این روش دارای مزیت‌های فراوانی از جمله اجرای آسان و قابل درک و دقت

* نویسنده‌ی عهده‌دار مکاتبات

مطلوب نتایج حاصل از آن است. از سویی دیگر یکی از عیب‌های مهم این روش، نبود کنترلی توسط پژوهش‌گر در انتخاب واحدهای نمونه است و گاهی این امر می‌تواند منجر به انتخاب بدنمونه‌هایی شود که نمایانگر جامعه نیستند. تلاش‌های اخیر آمارشناسان برای کاهش احتمال رخداد بدنمونه بر استفاده از اطلاعات گذشته‌ی واحدهای جامعه متمرکز بود تا زمانی که مک‌این‌تایر [۹] روش نمونه‌گیری جدیدی با نام نمونه‌گیری از مجموعه‌ی رتبه‌دار (RSS) را معرفی کرد. هدف نهایی روش معرفی‌شده توسط مک‌این‌تایر، انتخاب مشاهده‌هایی از جامعه بود که با احتمال بیش‌تری نسبت به نمونه‌گیری تصادفی ساده و با تعداد مشاهده‌های برابر، تمامی سطوح جامعه را در بر گیرد. پس از وی تاکاهاسی و واکیمتو [۱۲] در مقاله‌ای بنیادی، به بیان و اثبات نظریه‌های اساسی در شیوه‌ی نمونه‌گیری RSS پرداختند. دل و کلاتر [۶] نشان دادند که در صورت وجود یا نبود خطای رتبه‌بندی، میانگین نمونه‌ی RSS یک برآوردگر نااریب برای میانگین جامعه است. مارتین و همکاران [۸] روش نمونه‌گیری از مجموعه‌ی رتبه‌دار را برای ارزیابی گونه‌های گیاهی در جنگل‌های بلوط یک منطقه به کار گرفتند. سماوی و همکاران [۱۱] روش نمونه‌گیری از مجموعه‌ی رتبه‌دار کرانگین (ERSS^۱) را برای موقعیت‌هایی که تشخیص واحدهای رتبه‌بندی پرهزینه و دشوار است، پیش‌نهاد کردند. نعمت‌الهی و همکاران [۱۰] روش نمونه‌گیری مجموعه‌ی رتبه‌دار را در نمونه‌گیری خوشه‌ای دومرحله‌ای به‌کار بردند و ویژگی‌های روش نمونه‌گیری جدید را معرفی کردند. جعفری جوزانی و جانسون [۷] روش نمونه‌گیری از مجموعه‌ی رتبه‌دار را در جامعه‌های متناهی و بر پایه‌ی نمونه‌گیری‌های با جای‌گذاری مورد استفاده قرار دادند. پس از آن‌ها الناصر و بنی‌مصطفی [۲] روش نمونه‌گیری از مجموعه‌ی رتبه‌دار کرانگین استوار (RERSS^۳) را مطرح کردند و به بررسی کارایی برآوردگر میانگین آن تحت توزیع‌های متقارن و نامتقارن پرداختند. العمری [۳ و ۴] روش نمونه‌گیری از مجموعه‌ی رتبه‌دار کرانگین استوار دوگانه (DRERSS^۴) را معرفی و از آن در برآورد پارامترهای میانگین و میانه‌ی جامعه استفاده کرد. العمری و الناصر [۵] برآوردی برای میانه‌ی جامعه با استفاده از روش RERSS معرفی کردند. نمونه‌گیری از مجموعه‌ی رتبه‌دار یکی از روش‌های نمونه‌گیری از جامعه‌هایی است که بتوان به‌سادگی و با صرف کم‌ترین هزینه واحدهای آن را با توجه به متغیر مورد علاقه رتبه‌بندی کرد. در این روش نمونه‌گیری، رتبه‌بندی واحدهای نمونه باعث افزایش نظم و انتخاب نمونه‌ای می‌شود که نماینده‌ی مناسبی از جامعه‌ی مورد پژوهش است. این روش با کاهش

تعداد مشاهده‌های لازم برای اندازه‌گیری و افزایش درستی برآوردها با استفاده از رتبه‌بندی واحدها، دستیابی به برآوردهایی کاراتر را برای استنباط درباره‌ی پارامتر مورد نظر جامعه فراهم می‌کند.

در این مقاله با هدف برآورد میانگین و میانه‌ی جامعه ابتدا در بخش ۲ به معرفی روش‌های نمونه‌گیری از مجموعه‌ی رتبه‌دار می‌پردازیم. در بخش ۳ به بررسی کارایی روش‌های نمونه‌گیری نام‌برده‌شده در این مقاله در برآورد میانگین و میانه‌ی هزینه‌ی مسکن خانوارهای روستایی ساکن استان مازندران [۱] می‌پردازیم. در نهایت در بخش آخر جمع‌بندی و نتیجه‌گیری از مطالب گفته‌شده ارائه می‌کنیم.

۲- مروری بر روش‌های نمونه‌گیری از مجموعه‌ی رتبه‌دار

در این بخش روش‌های نمونه‌گیری از مجموعه‌ی رتبه‌دار مورد بحث در مقاله به اختصار توضیح داده می‌شود.

نمونه‌گیری از مجموعه‌ی رتبه‌دار (ساده)

شیوه‌ی انتخاب یک نمونه‌ی n تایی ($n = mk$) از جامعه به روش نمونه‌گیری از مجموعه‌ی رتبه‌دار به صورت زیر است:

گام نخست: ابتدا k نمونه‌ی k تایی به روش تصادفی ساده از جامعه‌ی مورد بررسی انتخاب می‌کنیم؛

گام دوم: در هر نمونه‌ی تصادفی، نمونه‌ها را بر اساس قضاوت شخصی یا کارشناسی رتبه‌بندی می‌کنیم؛

گام سوم: در اولین نمونه‌ی تصادفی k تایی، واحد دارای کوچک‌ترین رتبه را انتخاب می‌کنیم. در دومین نمونه‌ی تصادفی k تایی، واحد دارای دومین رتبه را انتخاب می‌کنیم و به همین صورت تا در k امین نمونه‌ی تصادفی، واحد دارای بزرگ‌ترین رتبه را انتخاب می‌کنیم؛ و

گام چهارم: گام‌های یک تا سه را m بار تکرار می‌کنیم تا اندازه‌ی نمونه‌ی مورد نیاز برای تحلیل آماری به دست آید.

قضیه ۱- یک برآوردگر نارایب برای میانگین جامعه بر اساس روش نمونه‌گیری از مجموعه‌ی رتبه‌دار برابر است به:

$$\hat{\mu}_{RSS} = \bar{X}_{RSS} = \frac{1}{mk} \sum_{j=1}^m \sum_{i=1}^k X_{i(i:k)j}$$

با واریانس

$$Var(\bar{X}_{RSS}) = \frac{1}{mk} \sigma^2 - \frac{1}{mk} \sum_{i=1}^k (\hat{\mu}_{(i)} - \mu)^2$$

که در آن $X_{i(i:k)j}$ ، i آماره‌ی ترتیبی در نمونه‌ی تصادفی ساده‌ی k تایی در Z آماره‌ی تکرار و $\hat{\mu}_{(i)}$ میانگین آن است که به صورت زیر به دست می‌آید.

$$\hat{\mu}_{(i)} = \int_{-\infty}^{+\infty} x f_{(i)}(x) dx$$

که در آن $f_{(i)}(x)$ تابع چگالی آماره‌ی ترتیبی i ام در نمونه‌ی تصادفی k تایی است.

نمونه‌گیری از مجموعه‌ی رتبه‌دار کرانگین

یک نمونه‌ی n تایی ($n = mk$) به روش نمونه‌گیری از مجموعه‌ی رتبه‌دار کرانگین از طریق پی‌موندن گام‌های زیر حاصل می‌شود.

گام نخست: ابتدا k نمونه‌ی تصادفی k تایی به روش تصادفی ساده از جامعه‌ی مورد بررسی انتخاب می‌کنیم.

گام دوم: نمونه‌های به دست آمده از گام نخست را بر اساس قضاوت شخصی یا کارشناسی رتبه‌بندی و مرتب می‌کنیم.

گام سوم: در صورت زوج بودن k ، از مجموعه نمونه‌های با شماره‌ی کوچک‌تر یا برابر $\frac{k}{2}$ ، کوچک‌ترین واحد رتبه‌بندی شده و از مجموعه نمونه‌های با شماره‌ی بزرگ‌تر از $\frac{k}{2}$ ، بزرگ‌ترین واحد رتبه‌بندی شده را انتخاب می‌کنیم.

در صورت فرد بودن k ، از مجموعه نمونه‌های با شماره‌ی کوچک‌تر یا برابر $\frac{k-1}{2}$ ،

کوچک‌ترین واحد رتبه‌بندی شده و از مجموعه نمونه‌های با شماره‌ی بزرگ‌تر یا برابر $\frac{k+1}{2}$ ،

بزرگ‌ترین واحد رتبه‌بندی شده و از مجموعه نمونه‌ی با شماره‌ی $\frac{k+1}{p}$ ، واحد دارای رتبه‌ی $\frac{k+1}{p}$ را انتخاب می‌کنیم.

گام چهارم: گام‌های یک تا سه را m بار تکرار می‌کنیم تا اندازه‌ی نمونه‌ی مورد نیاز برای تحلیل آماری به دست آید.

قضیه ۲- یک برآوردگر نارایب برای میانگین جامعه‌های متقارن بر اساس روش نمونه‌گیری از مجموعه‌ی رتبه‌دار کرانگین برابر

$$\bar{X}_{ERSS} = \frac{1}{mk} \sum_{j=1}^m \left\{ \sum_{i=1}^k X_{i(1:k)j} + \sum_{i=\frac{k+1}{p}}^k X_{i(k:k)j} + I \times X_{\frac{k+1}{p}(\frac{k+1}{p}:k)j} \right\}$$

با واریانس

$$Var(\bar{X}_{ERSS}) = \frac{1}{pmk} \left(\sigma_{(1:k)}^2 + \sigma_{(k:k)}^2 \right)$$

است. $X_{i(l:k)j}$ برابر با l امین آماره‌ی ترتیبی در مجموعه‌ی تصادفی k تایی i ام از چرخه‌ی j ام است و $\sigma_{(l:k)}^2$ نیز واریانس l امین آماره‌ی ترتیبی در مجموعه‌های تصادفی k تایی است و I برابر یک تابع نشان‌گر است که به صورت زیر تعریف می‌شود.

$$I = \begin{cases} 1 & \text{اگر } k \text{ فرد باشد} \\ 0 & \text{اگر } k \text{ زوج باشد} \end{cases}$$

سماوی و همکاران [۱۱] روش نمونه‌گیری از مجموعه‌ی رتبه‌دار کرانگین را با فرض قابل تشخیص بودن کوچک‌ترین و بزرگ‌ترین واحد در مجموعه‌ی نمونه‌ها، معرفی کردند. اما در صورت وجود داده‌های دورافتاده، استفاده از روش نمونه‌گیری ERSS ممکن است منجر به ارایی‌ی برآوردهای نادرستی از پارامترهای جامعه شود. برای از بین بردن یا کاهش اثر مشاهده‌های دورافتاده‌ی موجود در مجموعه‌ی داده‌ها استفاده از روش‌های نمونه‌گیری RERSS توسط الناصر و بنی‌مصطفی [۲] و DRERSS توسط العمری [۳] پیش‌نهاد شد.

نمونه‌گیری از مجموعه‌ی رتبه‌دار کرانگین استوار

برای به‌دست آوردن یک نمونه‌ی $n = mk$ تایی از جامعه‌ی مورد بررسی با استفاده از روش نمونه‌گیری از مجموعه‌ی رتبه‌دار کرانگین استوار، گام‌های زیر را انجام می‌دهیم.

گام نخست: ابتدا k نمونه‌ی k تایی به روش تصادفی ساده از جامعه‌ی مورد نظر انتخاب می‌کنیم.

گام دوم: نمونه‌های به‌دست آمده از گام نخست را رتبه‌بندی می‌کنیم.
گام سوم: ضریب r را برابر با جزء صحیح αk قرار می‌دهیم که در آن α عددی ثابت و در بازه‌ی $(0, 0.5)$ است.

گام چهارم: در صورت زوج بودن k ، از مجموعه نمونه‌های با شماره‌ی کوچک‌تر یا برابر $\frac{k}{2}$ ، واحد دارای رتبه‌ی $r + 1$ و از مجموعه نمونه‌های با شماره‌ی بزرگ‌تر از $\frac{k}{2}$ ، واحد دارای رتبه‌ی $r - k$ را انتخاب می‌کنیم. در صورت فرد بودن k ، از مجموعه نمونه‌های با شماره‌ی کوچک‌تر یا برابر $\frac{k-1}{2}$ ، واحد دارای رتبه‌ی $r + 1$ و از مجموعه نمونه‌های با شماره‌ی بزرگ‌تر یا برابر $\frac{k+1}{2}$ ، واحد دارای رتبه‌ی $r - k$ و از مجموعه نمونه‌ی با شماره‌ی $\frac{k+1}{2}$ ، واحد دارای رتبه‌ی $\frac{k+1}{2}$ را انتخاب می‌کنیم.

گام پنجم: گام‌های یک تا چهار را m بار تکرار می‌کنیم تا اندازه‌ی نمونه‌ی مورد نیاز برای تحلیل آماری به‌دست آید.

قضیه‌ی ۳- یک برآوردگر نارایب برای میانگین جامعه‌های متقارن بر اساس روش نمونه‌گیری از مجموعه‌ی رتبه‌دار کرانگین استوار برابر

$$\bar{X}_{RERSS} = \frac{1}{mk} \sum_{j=1}^m \left\{ \sum_{i=1}^{\lfloor \frac{k}{2} \rfloor} X_{i(r+1:k)j} + \sum_{i=k-\lfloor \frac{k}{2} \rfloor+1}^k X_{i(k-r:k)j} + I \times \right. \\ \left. X_{\frac{k+1}{2} \left(\frac{k+1}{2} : k \right) j} \right\}$$

با واریانس

$$\text{Var}(\bar{X}_{\text{RERSS}}) = \frac{1}{m^2 k^2} \left(\left[\frac{k}{2} \right] \sigma_{(r+1:k)}^2 + \left[\frac{k}{2} \right] \sigma_{(k-r:k)}^2 + I \times \sigma_{\left(\frac{k+1}{2}; k\right)}^2 \right)$$

است.

نمونه‌گیری از مجموعه‌ی رتبه‌دار کرانگین استوار دوگانه

برای به‌دست آوردن یک نمونه‌ی $n = mk$ تایی از جامعه‌ی مورد بررسی با استفاده از روش نمونه‌گیری از مجموعه‌ی رتبه‌دار کرانگین استوار دوگانه گام‌های زیر را در نظر می‌گیریم:

گام نخست: ابتدا k^2 نمونه‌ی k تایی به روش تصادفی ساده از جامعه‌ی مورد نظر انتخاب می‌کنیم.

گام دوم: نمونه‌های به‌دست آمده از گام نخست را رتبه‌بندی می‌کنیم.
گام سوم: ضریب r را برابر با جزء صحیح ak قرار می‌دهیم که در آن α عددی ثابت و در بازه‌ی $(0, 0.50)$ است.

گام چهارم: در صورت زوج بودن k ، از مجموعه نمونه‌های با شماره‌ی کوچک‌تر یا برابر $\frac{k^2}{4}$ ، واحدهای دارای رتبه‌ی $r + 1$ و از مجموعه نمونه‌های با شماره‌ی بزرگ‌تر از $\frac{k^2}{4}$ ، واحدهای دارای رتبه‌ی $k - r$ را انتخاب می‌کنیم. در صورت فرد بودن k ، از مجموعه نمونه‌های با شماره‌ی کوچک‌تر یا برابر $\frac{k(k-1)}{4}$ ، واحدهای دارای رتبه‌ی $r + 1$ ، از k مجموعه نمونه‌ی بعدی، واحدهای دارای رتبه‌ی $\frac{k+1}{4}$ و از $\frac{k(k-1)}{4}$ مجموعه نمونه‌ی باقی‌مانده، واحدهای دارای رتبه‌ی $k - r$ را انتخاب می‌کنیم. اجرای صحیح این گام منجر به انتخاب k مجموعه نمونه‌ی k تایی می‌شود.

گام پنجم: در k مجموعه نمونه‌ی k تایی به‌دست آمده از گام قبل همانند گام چهارم در روش RERSS به انتخاب نمونه می‌پردازیم.

گام ششم: گام‌های یک تا پنج را m بار تکرار می‌کنیم تا اندازه‌ی نمونه‌ی مورد نیاز به‌دست آید.

قضیه ۴- یک برآوردگر نارایب برای میانگین جامعه‌های متقارن بر اساس روش نمونه‌گیری از مجموعی رتبه‌دار کرانگین استوار دوگانه برابر

$$\bar{X}_{DRERSS}^* = \frac{1}{mk} \sum_{j=1}^m \left\{ \sum_{i=1}^{\lfloor \frac{k}{2} \rfloor} X_{i(r+1:k)j}^* + \sum_{i=k-\lfloor \frac{k}{2} \rfloor+1}^k X_{i(k-r:k)j}^* + I \times X_{\frac{k+1}{2}(\frac{k+1}{2}:k)j}^* \right\}$$

با واریانس

$$Var(\bar{X}_{DRERSS}) = \frac{1}{m^2 k^2} \left(\left[\frac{k}{2} \right] \sigma_{(r+1:k)}^{2*} + \left[\frac{k}{2} \right] \sigma_{(k-r:k)}^{2*} + I \times \sigma_{(\frac{k+1}{2}:k)}^{2*} \right)$$

است.

در قضیه ۴ علامت * به معنای انجام دوباره‌ی رتبه‌بندی است و تأکید می‌کند که $X_{i(r+1:k)j}^*$ ممکن است برابر با $X_{i(r+1:k)j}$ نباشد. برای برآورد میانه‌ی جامعه با استفاده از روش‌های نمونه‌گیری SRS، RSS، ERSS، RERSS و DRERSS از تابع میانه، به صورت زیر استفاده می‌کنیم.

$$\widehat{M}_l = \text{median}\{S_l\}; \quad l = SRS, RSS, ERSS, RERSS, DRERSS$$

که در آن S_l به معنای نمونه‌ی انتخاب‌شده از جامعه با استفاده از l امین روش‌های نمونه‌گیری نام‌برده شده در این مقاله است.

در عمل گاهی برای انتخاب نمونه با استفاده از روش‌های نمونه‌گیری از مجموعی رتبه‌دار، مقدار k را عددی کوچک در نظر می‌گیریم تا از بروز خطای رتبه‌بندی جلوگیری شود. در این مقاله فرض می‌کنیم رتبه‌بندی‌ها کامل و بدون خطا انجام شده‌اند.

۳- مطالعه‌ی کاربردی

در این بخش با استفاده از نرم‌افزار SAS-9.4 به بررسی کارایی برآوردهای میانگین و میانه‌ی حاصل از روش‌های نمونه‌گیری SRS، RSS، ERSS، RERSS و DRERSS می‌پردازیم. برای این منظور از داده‌های آمارگیری هزینه و درآمد خانوارهای روستایی سال ۱۳۹۲ در استان مازندران [۱]، استفاده شده است. آمارگیری هزینه و درآمد خانوار هر ساله توسط مرکز آمار ایران انجام می‌شود. در این مطالعه‌ی کاربردی متغیر هزینه‌ی سالانه‌ی خانوار در بخش مسکن به‌عنوان متغیر مورد علاقه انتخاب شده است. در جدول ۱ خلاصه‌ای از آماره‌های توصیفی متغیرهای هزینه‌ی مسکن خانوارهای روستایی استان مازندران ۱۳۹۲ آورده شده است.

جدول ۱- آماره‌های توصیفی خانوارهای روستایی ساکن استان مازندران در سال ۱۳۹۲

هزینه‌ی مسکن خانوار (ده میلیون ریال)	
میانگین حسابی	۲/۴۶۱
میانه	۲/۱۹۶
واریانس	۰/۰۲۱
ضریب تغییرات	۵۸/۱۹۷
ضریب چولگی	۴۳/۴۲

۱-۳- روش‌های انتخاب نمونه و برآورد

طرح آمارگیری هزینه و درآمد خانوار روستایی توسط مرکز آمار ایران در سال ۱۳۹۲ براساس آمارگیری از ۱۹۴۳۷ خانوار نمونه در نقاط روستایی کشور اجرا شده است که از این تعداد، ۶۱۷ خانوار در نقاط روستایی استان مازندران ساکن بودند. خانوارهای روستایی استان مازندران که در این آمارگیری به‌عنوان نمونه انتخاب شده‌اند، به‌عنوان جامعه‌ی آماری این مطالعه در نظر گرفته شده‌اند و از بین آن‌ها نمونه‌هایی با استفاده از روش‌های نام‌برده‌شده انتخاب شده است. هدف از این مطالعه‌ی کاربردی، شناسایی روشی کارا تر و معرفی آن برای برآورد میانگین و میانه‌ی هزینه‌ی مسکن خانوارهای روستایی ساکن استان مازندران طی یک سال است.

برای این منظور ۵۰۰۰ نمونه‌ی مستقل به‌صورت زیر از جامعه‌ی مورد نظر انتخاب می‌کنیم.

ابتدا با استفاده از روش نمونه‌گیری تصادفی ساده، نمونه‌هایی k تایی (۱۲، ۱۰، ۶، ۵) k از جامعه‌ی مورد نظر انتخاب می‌کنیم. حال با انجام رتبه‌بندی کامل و بدون خطا بر اساس متغیر مورد علاقه، واحدهای نمونه‌ای انتخاب‌شده را مرتب می‌کنیم. در هر مجموعه نمونه‌ی رتبه‌بندی‌شده بر اساس روش‌های نمونه‌گیری RSS، ERSS، RERSS و DRERSS به انتخاب نمونه می‌پردازیم. در هر روش، مقدار میانگین و میانه‌ی جامعه را محاسبه می‌کنیم. این عمل را ۵۰۰۰ بار به‌طور مستقل تکرار می‌کنیم. در نهایت متوسط میانگین و میانه‌های به‌دست آمده را به‌عنوان برآوردی از پارامترهای جامعه‌ی مورد نظر، محاسبه می‌کنیم. به‌دلیل بزرگ بودن مقدار میانگین توان دوم خطا از معیار میانگین نسبی توان دوم خطا برای مقایسه‌ی این روش‌ها استفاده شده است.

اندازه‌ی نمونه‌ی مورد نظر برای بررسی کارایی روش‌های نمونه‌گیری از رابطه‌ی (۱) به‌دست آمده است.

$$(1) \quad n = \frac{Z_{1-\frac{\alpha}{2}}^2 \sigma_X^2}{e^2 \bar{X}^2 + \frac{Z_{1-\frac{\alpha}{2}}^2 \sigma_X^2}{N}}$$

که در آن با قرار دادن α برابر ۰/۰۵، $Z_{0.975}$ (صدک توزیع نرمال استاندارد) برابر ۱/۹۶،

σ_X^2 (واریانس داده‌ها) برابر ۰/۲۱ ده میلیون ریال، e (خطای نسبی) برابر ۰/۱، \bar{X} (میانگین حسابی داده‌ها) برابر ۲/۴۶۱ ده میلیون ریال و N (اندازه‌ی جامعه) برابر ۶۱۷ خانوار، اندازه‌ی نمونه برابر با ۶۰ خانوار به‌دست می‌آید.

۲-۳- کارایی برآورد میانگین جامعه

میانگین نسبی توان دوم خطا (RMSE)^۵ برای محاسبه‌ی کارایی برآوردگر میانگین، به‌صورت زیر محاسبه شده است.

$$\hat{\mu}_l = \frac{1}{5000} \sum_{i=1}^{5000} \hat{\mu}_{l,i}$$

$$\widehat{bias}(\hat{\mu}_l) = \frac{1}{5000} \sum_{i=1}^{5000} (\hat{\mu}_{l,i} - \hat{\mu}_l)$$

$$\widehat{RMSE}(\hat{\mu}_l) = \frac{MSE(\bar{X}_{SRS})}{\mu} = \frac{1}{5000\mu} \sum_{i=1}^{5000} (\hat{\mu}_{l,i} - \hat{\mu}_l)^2$$

$$(2) \quad \widehat{RE}_l = \frac{RMSE(\bar{X}_{SRS})}{RMSE(\bar{X}_l)} = \frac{\frac{MSE(\bar{X}_{SRS})}{\mu}}{\frac{MSE(\bar{X}_l)}{\mu}} = \frac{MSE(\bar{X}_{SRS})}{MSE(\bar{X}_l)};$$

که در آن‌ها

$$l = RSS, ERSS, RERSS, DRERSS$$

$\hat{\mu}_{l,i}$ برابر میانگین i اُمین تکرار مستقل روش نمونه‌گیری l ام و $\bar{X}_l = \hat{\mu}_l$ برابر متوسط میانگین‌های به‌دست آمده از ۵۰۰۰ تکرار روش نمونه‌گیری l ام است. برآوردهای کارایی برآوردگر میانگین حاصل از نمونه‌گیری‌های SRS، ERSS، RSS و RERSS و DRERSS برای $k = 5, 6, 10, 12$ در جدول ۲ خلاصه شده است.

جدول ۱- برآوردهای کارایی برآوردگر میانگین هزینه‌ی مسکن حاصل از روش‌های SRS، ERSS و RERSS و DRERSS با اندازه‌ی مجموعه‌ی ۵، ۶، ۱۰ و ۱۲ و اندازه‌ی نمونه‌ی نهایی ۶۰ برای خانوارهای روستایی ساکن استان مازندران در سال ۱۳۹۲ بر حسب ده میلیون ریال

روش نمونه‌گیری						
DRERSS	RERSS	ERSS	RSS	SRS	معیار	k
۰/۰۰۱۲	۰/۰۰۳۱	۰/۰۱۰۷	۰/۰۰۰۷	۰/۰۱۲	RMSE	۵
۹/۸۹۲	۳/۸۳۴	۱/۱۳	۱/۷۳	۱	RE	
۰/۰۰۱	۰/۰۰۲۹	۰/۰۱۳	۰/۰۰۶۹	۰/۰۱۲	RMSE	۶
۱۲/۱۱۹	۴/۰۹۷	۰/۸۳۳	۱/۷۶۶	۱	RE	
۰/۰۰۰۷	۰/۰۰۲۵	۰/۰۱۶۷	۰/۰۰۵۱۳	۰/۰۱۲	RMSE	۱۰
۱۶/۵۲	۴/۸۹	۰/۷۲	۲/۳۶	۱	RE	
۰/۰۰۰۷	۰/۰۰۲۳	۰/۰۱۸۳	۰/۰۰۵۲۷	۰/۰۱۲	RMSE	۱۲
۱۷/۵۵	۵/۲۱۵	۰/۶۶	۲/۲۹۵	۱	RE	

۳-۳- کارایی برآورد میانه‌ی جامعه

کارایی برآوردگر میانه‌ی حاصل از روش‌های نمونه‌گیری نام‌برده‌شده در این مقاله، مشابه با روابط در قسمت میانگین محاسبه شده است.

برآورد کارایی برآوردگر میانه‌ی حاصل از نمونه‌گیری‌های SRS، RSS، ERSS، RERSS و DRERSS برای $k = 5, 6, 10, 12$ در جدول ۳ خلاصه شده است.

جدول ۳- برآوردهای کارایی برآوردگر میانه هزینه‌ی مسکن حاصل از روش‌های SRS، RSS، ERSS، RERSS و DRERSS با اندازه‌ی مجموعه‌ی ۵، ۶، ۱۰ و ۱۲ و اندازه‌ی نمونه‌ی نهایی ۶۰ برای خانوارهای روستایی ساکن استان مازندران در سال ۱۳۹۲ بر حسب ده میلیون ریال

روش نمونه‌گیری						
DRERSS	RERSS	ERSS	RSS	SRS	معیار	k
۰/۰۰۳۳	۰/۰۰۴۵	۰/۰۰۹۸	۰/۰۰۰۶	۰/۰۱۲	\widehat{RMSE}	۵
۳/۶۱۸	۲/۶۲۵	۱/۲۰۵	۱/۹۵۵	۱	\widehat{RE}	
۰/۰۰۰۳	۰/۰۰۵۴	۰/۰۰۱۷	۰/۰۰۵۷	۰/۰۱۲	\widehat{RMSE}	۶
۳/۷۶۶	۲/۲۰۶	۰/۷۰۵	۲/۰۵۹	۱	\widehat{RE}	
۰/۰۰۲۶	۰/۰۰۱۰۳	۰/۰۰۱۴	۰/۰۰۴۱	۰/۰۱۲	\widehat{RMSE}	۱۰
۴/۴۷	۱/۱۴۷	۰/۸۶۹	۲/۸۶	۱	\widehat{RE}	
۰/۰۰۲۱	۰/۰۰۱۸	۰/۰۰۱۳	۰/۰۰۳۸	۰/۰۱۲	\widehat{RMSE}	۱۲
۵/۵۶۹	۱/۱	۰/۸۹۶	۳/۱۰۴	۱	\widehat{RE}	

۴- بحث و نتیجه‌گیری

با توجه به مقدارهای به‌دست آمده از رابطه‌های \widehat{RMSE} و \widehat{RE} که در جدول‌های ۲ و ۳ خلاصه شده است مشخص می‌شود که:

- در داده‌های واقعی با احتمال حضور مشاهده‌های دورافتاده، روش نمونه‌گیری از مجموعه‌ی رتبه‌دار کرانگین کارایی کم‌تر و عمل‌کرد نامطلوبی نسبت به روش نمونه‌گیری تصادفی ساده در برآورد پارامترهای مرکزی (میانگین و میانه) جامعه دارد.

- روش نمونه‌گیری از مجموعه‌ی رتبه‌دار حتی در داده‌هایی با احتمال حضور مشاهده‌های دورافتاده نیز روشی کاراتر از نمونه‌گیری تصادفی ساده در برآورد میانگین و میانه‌ی جامعه است اما دارای بیش‌ترین کارایی نیست.
- روش‌های نمونه‌گیری از مجموعه‌ی رتبه‌دار کرانگین استوار و روش نمونه‌گیری از مجموعه‌ی رتبه‌دار کرانگین استوار دوگانه، کاراتر از روش نمونه‌گیری تصادفی ساده در برآورد پارامترهای مرکزی جامعه هستند.
- در بین تمامی روش‌های نمونه‌گیری بررسی‌شده در این مقاله، روش نمونه‌گیری از مجموعه‌ی رتبه‌دار کرانگین استوار دوگانه دارای بیش‌ترین کارایی نسبت به نمونه‌گیری تصادفی ساده در برآورد پارامترهای مرکزی جامعه است.
- افزایش اندازه‌ی مجموعه نمونه‌های گردآوری‌شده سبب افزایش کارایی روش‌های نمونه‌گیری RSS، ERSS و DRERSS می‌شود. افزایش مقدار k تنها در نمونه‌گیری از مجموعه‌ی رتبه‌دار کرانگین استوار و برای برآورد میانه‌ی جامعه، سبب کاهش کارایی آن شده است.
- به نظر می‌رسد استفاده از روش‌های نمونه‌گیری از مجموعه‌ی رتبه‌دار کرانگین استوار و نمونه‌گیری از مجموعه‌ی رتبه‌دار کرانگین استوار دوگانه برای برآورد هزینه‌ی مسکن خانوارها، که متغیری با توزیع چوله است، روشی با کارایی بسیار بالا باشد. از این رو استفاده از آن در آمارگیری‌ها توصیه می‌شود.

توضیحات

1. Ranked Set Sampling
2. Extreme Ranked Set Sampling
3. Robust Extreme Ranked Set Sampling
4. Double Robust Extreme Ranked Set Sampling
5. Relative Mean Squared Error

مرجع‌ها

[۱] مرکز آمار ایران، نتایج آمارگیری هزینه و درآمد خانوار سال ۱۳۹۲.

- [2] Al-Nasser, D.A., and Bani-Mustafa, A. (2009). Robust Extreme Ranked Set Sampling, *Journal of Statistical Computation and Simulation*, **79**, 859–867.
- [3] Al-Omari, A.I. (2010a). Estimation of Mean Based on Modified Robust Extreme Ranked Set Sampling, *Journal of Statistical Computation and Simulation*, **81**, 1055–1066.
- [4] Al-Omari, A.I. (2010b). Estimation of the Population Median of Symmetric an Asymmetric Distributions using Double Robust Extreme Ranked Set Sampling, *Journal of Revista Investigation Operational*, **31**, 199–207.
- [5] Al-Omari, A.I. and Al-Nasser, A.D. (2012). On the Population Median Estimation Using Robust Extreme Ranked Set Sampling, *Monte Carlo Methods and Applications*, **18**, 109–118.
- [6] Dell, Tr. and Clutter, J.I. (1972). Ranked Set Sampling Theory with Order Statistics Background, *Biometrics*, **28**, 545–555.
- [7] Jafari Jozani, M. and Johnson, B.C. (2011). Design based Estimation for Ranked Set Sampling in Finitepopulations, *Environmental and Ecological Statistics*, **18**, 663–685.
- [8] Martin, W.L, Sharik, T.L, Oderwald, R.G. and Smith, D.W. (1980). *Evaluation of Ranked Set Sampling for Estimating Shrub Phytomass in Appalachian Oak Forests*, Publication Number Fws-4-80, School Of Forestry And Wildlife Resources, Virginia Polytechnic Institute and State University, Blacksburg.
- [9] McIntyre, G.A. (1952, Reprinted 2005). A Method for Unbiased Selective Sampling, Using Ranked Sets, *Aust J Agric Res* 1952, **3**, 385–390. (Reprinted In *The American Statistician* 2005, **59**, 230–232.
- [10] Nematollahi, N., Salehi, M.M. and Saba, R.A. (2008). Two-Stage Cluster Sampling with Ranked Set Sampling In The Secondary Sampling Frame, *Communications In Statistics – Theory And Methods*, **37**, 2404–2415.

- [11] Samawi, H.M., Ahmed, M.S. and Abu-Dayyeh, W. (1996). Estimating the Population Mean Using Extreme Ranked Set Sampling, *Biometrical Journal*, **38**, 577–586.
- [12] Takahasi, K. and Wakimoto, K. (1968). On Unbiased Estimates of the Population Mean Based on the Sample Stratified by Means of Ordering, *Ann Inst Stat Math*, **20**, 1–31.

میرحامد هاشمی‌مجد

فوق لیسانس آمار اجتماعی

ساری، خیابان معلم، سازمان مدیریت و برنامه‌ریزی استان مازندران، معاونت آمار و اطلاعات، گروه اجرایی طرح‌های آماری و سرشماری‌ها.

رایانشانی: hamedhashemi772@yahoo.com

روشنک علی‌اکبری صبا

دکتری آمار

تهران، خیابان فاطمی، خیابان باباطاهر، خیابان سرتیپ فکوری، شماره‌ی ۱۴۵، پژوهشکده‌ی آمار.

رایانشانی: r_saba@src.ac.ir

