

به کارگیری روش پساتبقه‌بندی قضاوتی در نمونه‌گیری خوشه‌ای دومرحله‌ای

علی نجفی مجیدآبادی، طاهره رحمانی، نادر نعمت‌الهی*

دانشگاه علامه طباطبائی، گروه آمار

پذیرش ۹۸/۰۸/۱۸

دریافت ۹۸/۰۳/۰۷

چکیده

روش پساتبقه‌بندی قضاوتی روشی برای طبقه‌بندی مشاهدات با استفاده از یک متغیر کلیدی است. به طوری که طبقه‌بندی بعد از انتخاب نمونه انجام می‌شود. در این مقاله این روش را در نمونه‌گیری خوشه‌ای دومرحله‌ای به کار می‌بریم. به عبارت دیگر در نمونه‌گیری خوشه‌ای دومرحله‌ای در مرحله دوم به جای نمونه‌گیری تصادفی ساده بدون جای‌گذاری از روش پساتبقه‌بندی قضاوتی و روش پساتبقه‌بندی قضاوتی تعمیم‌یافته استفاده کرده و برآوردهای جدیدی را برای میانگین جامعه ارائه می‌کنیم. در نهایت با استفاده از شبیه‌سازی مونت کارلو با مجموعه داده‌های واقعی و داده‌های تولیدی از توزیع‌های متقارن و نامتقارن، برآوردهای پیشنهادی را با برآوردهای میانگین نمونه‌گیری خوشه‌ای دومرحله‌ای ساده را مقایسه می‌کنیم. نتایج شبیه‌سازی نشان می‌دهد که، در بیش‌تر حالت‌ها، برآوردهای پیشنهادی عمل‌کرد بهتری نسبت به برآوردهای میانگین نمونه‌گیری دومرحله‌ای ساده دارند.

واژه‌های کلیدی: روش پساتبقه‌بندی قضاوتی - نمونه‌گیری از مجموعه‌ی رتبه‌ای - نمونه‌گیری خوشه‌ای دومرحله‌ای

مقدمه

در بررسی‌های آماری، جمع‌آوری داده‌های مورد نظر برای استنباط در مورد جامعه، اغلب از طریق نمونه‌گیری انجام می‌گیرد. در برآورد پارامترهای یک جامعه با استفاده از نمونه‌انتخابی از جامعه، انتخاب روش نمونه‌گیری که با صرف زمان و هزینه کم‌تر، منجر به ارائه برآوردهای کارا شود، بسیار مهم است. به‌ویژه در مواقعی که اندازه‌گیری مشخصه بررسی شده در جامعه نیازمند زمان و هزینه بسیار زیادی است. ساده‌ترین و در عین حال پرکاربردترین روش نمونه‌گیری، نمونه‌گیری تصادفی ساده^۱ (SRS) است. در این روش، تمام واحدهای جامعه، شانس برابری برای انتخاب در نمونه را دارند. برای انتخاب نمونه در این روش، باید یک چارچوب از تمامی اعضای جامعه بررسی شده در دسترس باشد که در این چارچوب مشخصات تمام واحدهای جامعه ثبت شده باشد. در این حالت ممکن است اعضای نمونه تصادفی در کل جامعه پخش شده باشند و اندازه‌گیری آنها زمان‌بر بوده و نیازمند هزینه زیادی باشد. در بعضی بررسی‌ها، چارچوبی از تک تک اعضای جامعه در دسترس نیست بلکه چارچوب به صورت گروه‌ها یا خوشه‌هایی است و نمونه‌گیری با انتخاب از میان این خوشه‌ها انجام می‌شود. در این حالت با انتخاب برخی از خوشه‌ها و اندازه‌گیری برخی از اعضای خوشه‌های انتخابی، می‌توان به‌طور اساسی در میزان هزینه صرفه‌جویی کرد.

در هر دو مرحله نمونه‌گیری خوشه‌ای دومرحله‌ای اغلب از نمونه‌گیری تصادفی ساده استفاده می‌شود. به این ترتیب که اگر جامعه به N خوشه تقسیم شده باشد، در مرحله اول نمونه‌گیری n خوشه را به صورت تصادفی ساده انتخاب می‌کنیم و سپس در مرحله دوم از هر خوشه منتخب تعدادی از واحدها را به صورت تصادفی ساده به عنوان واحدهای

*نویسنده مسئول nematollahi@atu.ac.ir

1. Simple Random Sampling

نمونه اصلی برمی‌گزینیم که به آن نمونه‌گیری خوشه‌ای دومرحله‌ای با انتخاب نمونه تصادفی ساده در مرحله دوم^۱ (*TSCSRS*) گوییم.

در نمونه‌گیری تصادفی ساده به علت این که هیچ کنترلی روی اعضای نمونه انتخابی وجود ندارد، ممکن است نمونه‌های انتخابی نمایانگر خصوصیات جامعه مورد نظر نباشند و برآوردهای حاصل دقت کافی نداشته باشند. در چنین حالتی به طور معمول برای افزایش دقت برآوردهای حاصل، اندازه نمونه را به اندازه کافی بزرگ انتخاب می‌کنند تا دقت مطلوب حاصل شود که این خود موجب افزایش هزینه‌ها می‌شود. در چنین شرایطی، اگر بتوان واحدهای جامعه را با کمترین هزینه رتبه‌بندی کرد، مک‌این‌تایر [۱] روش نمونه‌گیری از مجموعه رتبه‌ای^۲ (*RSS*) را برای برآورد پارامترهای جامعه پیشنهاد کرد که بسیار کارتر از روش نمونه‌گیری تصادفی ساده با اندازه نمونه برابر است. این افزایش کارایی در نتیجه رتبه‌بندی واحدها حاصل می‌شود که با هزینه کمی صورت می‌گیرد. هالز و دل [۲] با استفاده از این روش نمونه‌گیری میانگین محصول علفه و میانگین ارتفاع درختان یک جنگل را برآورد کردند. تاکاهاسی و واکی‌موتو [۳] میانگین حسابی مشاهده‌های حاصل از نمونه مجموعه رتبه‌ای را به عنوان برآوردگر نارایی برای میانگین جامعه ارائه کردند و نشان دادند که این برآوردگر کارتر از برآوردگر میانگین جامعه در حالت نمونه‌گیری تصادفی ساده با اندازه نمونه برابر است.

در دو دهه گذشته، بسیاری از پژوهش‌گران روش *RSS* را در مسئله‌های پارامتری و ناپارامتری استفاده کرده و انواع روش‌های *RSS* تعمیم‌یافته را معرفی کرده‌اند. برای مثال روش *RSS* کرانگین^۳ (*ERSS*) به وسیله سماوی و همکاران [۴]، روش *RSS* جفت‌شده^۴ (*PaRSS*) به وسیله مطلق [۵]، روش *RSS* میانه‌ای^۵ (*MRSS*) به وسیله مطلق [۶]، روش *RSS* دوگانه^۶ (*DRSS*) به وسیله ال‌صالح و ال‌قدیری [۷]، روش *RSS* جزئی^۷ (*PRSS*) به وسیله حق و همکاران [۸] و روش *RSS* پیوندی^۸ (*HRSS*) به وسیله حق و همکاران [۹] برای برآورد میانگین جامعه پیشنهاد شده است. برخی از پژوهش‌گران نیز روش‌های مختلف *RSS* را در نمونه‌گیری‌های پیچیده‌تر به کار برده‌اند. برای مثال سماوی [۱۰] از *RSS*، سماوی و سعید [۱۱] از *ERSS*، ابراهیم و همکاران [۱۲] از *MRSS* و مهدی‌زاده و زمان‌زاده [۱۳] از *PaRSS* در نمونه‌گیری طبقه‌ای استفاده کرده‌اند. هم‌چنین در انتخاب نمونه در مرحله دوم روش نمونه‌گیری دومرحله‌ای، نعمت‌الهی و همکاران [۱۴] از روش *RSS* و حق [۱۵] از روش *HRSS* استفاده کرده‌اند که آنها را به ترتیب *TSCRSS* و *TSCRSS* می‌نامیم. آنها نشان داده‌اند که برآوردگر میانگین جامعه براساس *TSCRSS* و *TSCRSS* کارتر از برآوردگر میانگین جامعه براساس *TSCSRS* است. اوزتورک [۱۶] برآوردهایی برای میانگین جامعه متناهی و واریانس آن در نمونه‌گیری خوشه‌ای دومرحله‌ای با طرح نمونه‌گیری *RSS* ارائه کرد و نشان داد که کارایی برآوردهای پیشنهادی به ضریب هم‌بستگی درون خوشه‌ها بستگی ندارد. وانگ و همکاران [۱۷] و آهن و همکاران [۱۸] از *RSS* در طرح‌های تصادفی خوشه‌ای استفاده کردند هم‌چنین وانگ و همکاران [۱۹] به بررسی نمونه‌گیری از مجموعه رتبه‌ای نامتعادل^۹ (*URSS*) در بررسی‌های تصادفی شده خوشه‌ای پرداختند.

با وجود کاربردهای بسیار موفق *RSS*، پژوهش‌گران هنوز هم تمایلی به استفاده از روش *RSS* در مسائل عملی ندارند زیرا آنها اغلب ترجیح می‌دهند از روش‌های نمونه‌گیری استفاده کنند که بتوانند از داده‌های حاصل از آنها برای

1. Two- Stage Cluster Simple Random Sampling
2. Ranked Set Sampling
3. Extreme Ranked Set Sampling
4. Pair Ranked Set Sampling
5. Median Ranked Set Sampling
6. Double Ranked Set Sampling
7. Partial Ranked Set Sampling
8. Hybrid Ranked Set Sampling
9. Unbalanced Ranked Set Sampling

اهداف مختلف استفاده نمایند. در روش RSS ، مشاهدات دارای توزیع یکسان نیستند و اطلاعات رتبه‌بندی قابل تفکیک از مشاهدات نیست. به‌عبارت دیگر، رتبه‌ها بخش جدایی‌ناپذیر ساختار داده‌ها هستند و نمی‌توان آنها را نادیده گرفت. در مواردی که محقق علاقه‌ای به استفاده از روش‌های نمونه‌گیری پیچیده از قبیل RSS نداشته باشد و بخواهد تحلیل‌های خود را با استفاده از روش نمونه‌گیری تصادفی ساده شروع کرده، و در صورتی که تمایل داشته باشد، اطلاعات رتبه‌بندی را در مرحله بعدی به مدل اضافه کند، روشی با عنوان روش پساتبقة‌بندی قضاوتی^۱ (JPS) پیشنهاد شده است.

روش پساتبقة‌بندی قضاوتی به‌وسیله مک‌ایچرن و همکاران [۲۰] پیشنهاد شد. استفاده از این طرح نمونه‌گیری هنگامی نسبت به نمونه‌گیری تصادفی ساده اولویت و کارایی بیشتری دارد که پیدا کردن رتبه مشاهدات ارزان‌تر و ساده‌تر از اندازه‌گیری دقیق آنها باشد. فری و اوزتورک [۲۱] نشان دادند زمانی که رتبه‌بندی‌ها با استفاده از متغیرهای کمکی انجام می‌شوند برآوردگرهای میانگین در روش پساتبقة‌بندی قضاوتی و روش نمونه‌گیری از مجموعه رتبه‌ای تحت تابع زیان توان دوم خطا ناپذیرفتنی هستند. فری و فیمن [۲۲] برآوردگری بهبودیافته برای میانگین روش پساتبقة‌بندی قضاوتی ارائه کردند و نشان دادند که برآوردگر جدید ارائه‌شده کارایی بیشتری نسبت به برآوردگرهای موجود دارد. هم‌چنین دست‌برآورده و همکاران [۲۳] و زمان‌زاده و وانگ [۲۴] پژوهش‌هایی در زمینه برآورد مشخصه‌های جامعه در نمونه‌گیری پساتبقة‌بندی قضاوتی انجام دادند. وانگ و همکاران [۲۵] و محمدقاسمی و همکاران [۲۶] برآوردگرهای تعمیم‌یافته‌ای برای روش JPS ارائه کردند. امیدوار و همکاران [۲۷] بررسی‌هایی در مورد مدل‌های آمیزه‌ای متناهی با استفاده از روش JPS انجام دادند.

در پژوهش‌های انجام شده تاکنون، در نمونه‌گیری خوشه‌ای دومرحله‌ای در مرحله دوم از نمونه‌گیری تصادفی ساده بدون جای‌گذاری، نمونه‌گیری از مجموعه رتبه‌ای ساده و پیوندی استفاده شده است. از آن‌جاکه روش پساتبقة‌بندی قضاوتی نسبت به نمونه‌گیری تصادفی ساده اغلب منجر به برآوردگری کاراتر از میانگین جامعه می‌شود، بنابراین انتظار می‌رود که به‌کارگیری روش پساتبقة‌بندی قضاوتی در مرحله دوم نمونه‌گیری دومرحله‌ای موجب افزایش کارایی برآوردگر میانگین جامعه شود. از این‌رو، در این مقاله برای برآورد میانگین جامعه از روش نمونه‌گیری دومرحله‌ای استفاده می‌کنیم که در آن انتخاب نمونه‌ها در مرحله اول به روش تصادفی ساده بدون جای‌گذاری انجام می‌گیرد. در مرحله دوم نمونه‌گیری، برای انتخاب نمونه‌ها از روش پساتبقة‌بندی قضاوتی و تعمیم‌یافته‌های آن استفاده می‌کنیم و برآوردگرهای جدیدی را برای میانگین جامعه ارائه می‌دهیم. این روش‌ها را با $TSCJPS$ نشان می‌دهیم. در این روش برآوردگر ناریب جامعه و واریانس آن را به‌دست آورده و با انجام دو مطالعه شبیه‌سازی روی داده‌های مربوط به سرشماری عمومی کشاورزی سال ۱۳۸۲ مرکز آمار ایران [۲۸] و داده‌های تولیدی از دو توزیع متقارن و نامتقارن، نشان می‌دهیم که برآوردگر میانگین در نمونه‌گیری دومرحله‌ای با به‌کارگیری روش پساتبقة‌بندی قضاوتی در مرحله دوم، کارایی بیشتری نسبت به نمونه‌گیری دومرحله‌ای معمول بر اساس نمونه‌گیری تصادفی واحدها دارد.

مواد و روش‌ها

روش پساتبقة‌بندی قضاوتی

در برخی بررسی‌ها یک نمونه تصادفی ساده‌ی m تایی از جامعه جمع‌آوری و تلاش می‌شود با افزودن اطلاعات اضافی به این نمونه، دقت آن افزایش یابد. برای این منظور در دهه اخیر روشی به‌نام روش پساتبقة‌بندی قضاوتی پیشنهاد شده است. شیوه انتخاب نمونه‌ی JPS بدین‌شرح است.

فرض کنید نمونه‌ای تصادفی ساده شامل m عضو را از جامعه‌ای با میانگین μ و واریانس σ^2 انتخاب و خصوصیت مورد نظر یعنی Y_1, \dots, Y_m را ثبت کرده‌ایم. حال به‌ازای هر کدام از Y_j ها، یک مجموعه تصادفی بدون جای‌گذاری شامل $H-1$ عضو جامعه یعنی $Y_{j1}, \dots, Y_{j,H-1}$ که $H = 2, \dots, N$ را در نظر می‌گیریم. بنابر این، کل نمونه انتخابی به‌صورت زیر خواهد بود.

$$\begin{array}{cccc} Y_1 & Y_2 & \dots & Y_m \\ Y_{11} & Y_{21} & \dots & Y_{m1} \\ Y_{12} & Y_{22} & \dots & Y_{m2} \\ \vdots & \vdots & \vdots & \vdots \\ Y_{1,H-1} & Y_{2,H-1} & \dots & Y_{m,H-1} \end{array}$$

که در آن Y_{jr} نمونه کمکی r ام، برای رتبه‌بندی Y_j است. برای به‌دست آوردن نمونه پساتبقه‌بندی قضاوتی، هر ستون که دارای H واحد است را به‌صورت افزایشی مرتب می‌کنیم و بدون اندازه‌گیری خصوصیت Y در آن‌ها رتبه Y_j را بین $Y_{j1}, \dots, Y_{j,H-1}, Y_j$ برای $j = 1, \dots, m$ به‌صورت قضاوتی تعیین کرده و با R_j نشان می‌دهیم، که R_j ها دارای توزیع یکنواخت گسسته در مجموعه‌ی $\{1, \dots, H\}$ هستند. بنابر این نمونه حاصل از روش *JPS* به‌صورت (Y_j, R_j) ، $j = 1, \dots, m$ است. توجه شود که هر چه مقدار H کوچک‌تر باشد، رتبه‌بندی واحدها دقیق‌تر است.

در این روش ممکن است برخی از مشاهدات Y_j دارای رتبه یکسان باشند. بنابر این روش *JPS* یک نوع نمونه‌گیری پساتبقه‌بندی است که واحدهای دارای رتبه یکسان در یک طبقه قرار می‌گیرند و تعیین طبقات آنها بعد از نمونه‌گیری انجام می‌شود. رتبه‌بندی واحدهای نمونه‌ای می‌تواند به‌صورت چشمی یا با استفاده از یک متغیر کمکی تعیین شود. برای مثال فرض کنید U_1, \dots, U_m نمونه‌ای شامل m درخت در یک جنگل و Y_1, \dots, Y_m ارتفاع این درختان باشند. برای طبقه‌بندی قضاوتی این نمونه به‌ازای هر درخت، $H-1$ درخت را به‌طور تصادفی از جنگل انتخاب کرده و رتبه آن درخت را در این مجموعه‌ی H تایی از درخت‌ها به‌صورت قضاوتی تعیین می‌کنیم. این رتبه‌بندی قضاوتی می‌تواند به‌صورت چشمی و با نگاه به درختان صورت گیرد یا برای مثال با استفاده از متغیر کمکی قطر درختان (با این فرض که بین قطر و ارتفاع درختان هم‌بستگی زیادی وجود دارد) انجام شود.

توجه داشته باشید که واحدهایی که دارای رتبه یکسان هستند را در یک طبقه قرار می‌دهیم بنابر این H طبقه خواهیم داشت، که البته همه آنها لزوماً ناتهی نیستند، به‌عبارت دیگر ممکن است در یک یا چند طبقه مشاهده‌ای وجود نداشته باشد. از آن‌جاکه رتبه‌بندی‌ها در عمل به‌صورت قضاوت شخصی یا به‌کمک یک متغیر کمکی که با متغیر اصلی هم‌بستگی دارد انجام می‌شود، بنابر این رتبه‌بندی احتمالاً با خطا همراه است. در این صورت نمونه پساتبقه‌بندی قضاوتی برای $j = 1, \dots, m$ و $r = 1, \dots, H$ به‌صورت زوج‌های $(Y_{[r]j}, R_j = r)$ خواهد بود، که در آن $Y_{[r]j}$ واحد j ام نمونه‌ی m تایی اولیه است که در پساتبقه r ام قرار گیرد، یعنی دارای رتبه قضاوتی r باشد.

برای تعریف برآوردگر میانگین *JPS* استاندارد، فرض کنید n_r تعداد $\{j; R_j = r\}$ یعنی تعداد مشاهده‌ها با رتبه قضاوتی r باشد. در این صورت بردار (n_1, \dots, n_H) دارای توزیع چندجمله‌ای با تعداد آزمایش m و احتمال‌های برابر $\frac{1}{H}$ است. بنابر این، ممکن است به‌ازای برخی مقادیر r داشته باشیم $n_r = 0$ هم‌چنین I_r, I_{rj} و J_r را بدین‌صورت تعریف می‌کنیم:

$$I_{rj} = \begin{cases} 1 & R_j = r \\ 0 & \text{ow} \end{cases}, \quad n_r = \sum_{j=1}^m I_{rj}, \quad I_r = \begin{cases} 1 & n_r > 0 \\ 0 & n_r = 0 \end{cases}, \quad h = \sum_{r=1}^H I_r, \quad J_r = \begin{cases} \frac{1}{n_r} & n_r > 0 \\ 0 & \text{ow} \end{cases}$$

در این صورت برآوردگر میانگین جامعه بر اساس روش پس‌اطبقه‌بندی قضاوتی به صورت (۱) است:

$$\hat{\mu}_{JPS} = \frac{1}{h} \sum_{r=1}^H \bar{Y}_{[r]} I_r, \quad (1)$$

که در آن

$$\bar{Y}_{[r]} = \begin{cases} 0 & n_r = 0 \\ \frac{1}{n_r} \sum_{j=1}^m Y_{[r]j} I_{rj} & n_r > 0 \end{cases}$$

برای به‌دست آوردن ویژگی‌های برآوردگر $\hat{\mu}_{JPS}$ به لم زیر که به وسیله دست‌برآورده و همکاران [۲۳] ارائه شده است نیاز داریم.

لم ۱. در روش پس‌اطبقه‌بندی قضاوتی داریم:

$$E\left(\frac{I_r}{h}\right) = E\left(\frac{I_1}{h}\right) = \frac{1}{H} \quad \text{الف) برای } r = 1, \dots, H$$

$$\text{Cov}\left(\frac{I_r}{h}, \frac{I_s}{h}\right) = \text{Cov}\left(\frac{I_1}{h}, \frac{I_2}{h}\right) = -\frac{1}{H-1} \text{Var}\left(\frac{I_1}{h}\right), \quad r \neq s \quad \text{ب) برای}$$

ج) تابع احتمال $\frac{I_r}{h}$ ، $r = 1, \dots, H$ ، به صورت زیر است

$$p\left(\frac{I_r}{h} = u\right) = \begin{cases} \left(\frac{H-1}{H}\right)^m, & u = 0 \\ \frac{1}{H^m} \binom{H-1}{k-1} \sum_{j=1}^k (-1)^{j-1} (k-j+1)^m, & u = \frac{1}{k}, k = 1, \dots, H \end{cases}$$

$$E\left(\frac{J_r^2}{h^2}\right) = \frac{1}{H^m} \left\{ \frac{1}{m} + \sum_{k=2}^H \sum_{j=1}^{k-1} \sum_{n_r=1}^{m-k+1} \frac{(-1)^{j-1} (H-1) \binom{m}{n_r} \binom{k-1}{j-1} (k-j)^{m-n_r} \right\} \quad (2)$$

$$\text{Var}\left(\frac{I_r}{h}\right) = \frac{1}{m^2} \sum_{k=1}^{H-1} \left(\frac{k}{H}\right)^{m-1} \quad (3)$$

برای برهان به دست‌برآورده و همکاران [۲۳] مراجعه شود.

قضیه ۱. اگر $\mu_{[r]}$ و $\sigma_{[r]}^2$ به ترتیب میانگین و واریانس r امین آماره مرتب قضاوتی در یک نمونه تصادفی ساده به اندازه H باشند، آن‌گاه امید و واریانس برآوردگر میانگین پس‌اطبقه‌بندی قضاوتی $\hat{\mu}_{JPS}$ داده شده در (۱) بدین صورت است:

$$E(\hat{\mu}_{JPS}) = \frac{1}{H} \sum_{r=1}^H \mu_{[r]} = \mu \quad \text{الف)}$$

$$Var(\hat{\mu}_{JPS}) = E\left(\frac{J_1 I_1^2}{h^2}\right) \sum_{r=1}^H \sigma_{[r]}^2 + Var\left(\frac{I_1}{h}\right) \frac{H}{H-1} \sum_{r=1}^H (\mu_{[r]} - \mu)^2 \quad (ب)$$

برای برهان به دست برآورده و همکاران [۲۳] مراجعه شود.

با توجه به قضیه ۱، برآوردگری ناریب برای میانگین جامعه است.

یافته‌ها

نمونه‌گیری خوشه‌ای دومرحله‌ای با به کارگیری روش پس‌اطبقه‌بندی قضاوتی در مرحله دوم

نمونه‌گیری خوشه‌ای دومرحله‌ای یکی از روش‌های نمونه‌گیری است که در عمل بسیار پرکاربرد است. این روش نمونه‌گیری دارای دومرحله است. در مرحله اول، نمونه‌ها از واحدهای نمونه‌گیری اولیه^۱ (PSU) انتخاب می‌شوند. در مرحله دوم، نمونه‌ها از واحدهای نمونه‌گیری ثانویه^۲ (SSU) در هر یک از PSU ‌های انتخاب‌شده، انتخاب می‌شوند. در نمونه‌گیری خوشه‌ای دومرحله‌ای اغلب روشی برای انتخاب PSU و SSU ‌ها از قبل مشخص می‌شود. روش نمونه‌گیری تصادفی ساده اغلب برای انتخاب PSU و SSU ‌ها در هر دو مرحله استفاده می‌شود.

فرض کنید جامعه مورد نظر دارای N خوشه (N واحد نمونه‌گیری مرحله اول، PSU) بوده به طوری که i امین خوشه دارای M_i واحد (M_i واحد نمونه‌گیری مرحله دوم، SSU) است. هم‌چنین فرض کنید خصیصه مورد نظر Y دارای میانگین μ و واریانس σ^2 بوده و میانگین و واریانس آن در خوشه i ام به ترتیب μ_i و σ_i^2 باشد. در نمونه‌گیری خوشه‌ای دومرحله‌ای با استفاده از روش تصادفی ساده بدون جای‌گذاری در هر دو مرحله، اگر n تعداد خوشه‌های انتخاب‌شده در مرحله اول، m_i تعداد واحدهای انتخاب‌شده از نمونه i ام در مرحله دوم و Y_{ij} مقدار خصیصه مورد نظر برای j امین نمونه تصادفی ساده انتخابی ($j=1, \dots, m_i$) در i امین خوشه انتخابی ($i=1, \dots, n$) باشد، آن‌گاه نمونه به دست آمده بدین صورت است:

$$\begin{matrix} Y_{11} & Y_{21} & \dots & Y_{n1} \\ Y_{12} & Y_{22} & \dots & Y_{n2} \\ \vdots & \vdots & \vdots & \vdots \\ Y_{1m_1} & Y_{2m_2} & \dots & Y_{nm_n} \end{matrix} \quad (۲)$$

برآوردگر ناریب میانگین جامعه براساس این روش به صورت (۳) است:

$$\hat{\mu}_{TSCSRS} = \frac{1}{nM} \sum_{i=1}^n M_i \bar{Y}_i = \frac{1}{nM} \sum_{i=1}^n \sum_{j=1}^{m_i} \frac{M_i}{m_i} Y_{ij}, \quad (۳)$$

که در آن $\bar{M} = \frac{1}{N} \sum_{i=1}^N M_i$ و $\bar{Y}_i = \frac{1}{m_i} \sum_{j=1}^{m_i} Y_{ij}$ است.

قضیه ۲. در روش $TSCSRS$ برآوردگر $\hat{\mu}_{TSCSRS}$ برآوردگری ناریب برای میانگین جامعه است و دارای واریانس زیر است

$$Var(\hat{\mu}_{TSCSRS}) = \left(1 - \frac{n}{N}\right) \frac{S_b^2}{nM} + \frac{1}{nN} \sum_{i=1}^n \left(\frac{M_i}{\bar{M}}\right)^2 \left(\frac{M_i - m_i}{M_i - 1}\right) \frac{\sigma_i^2}{m_i},$$

1. Primary Sampling Unit
2. Secondary Sampling Unit

$$S_b^2 = \frac{1}{N-1} \sum_{i=1}^N (M_i \mu_i - \bar{M} \mu)^2$$

که در آن μ میانگین جامعه باشد. برای برهان به ککران [۲۹] مراجعه شود.

در عمل، اغلب این روش نمونه گیری در جامعه های بزرگ که به گروه های جمعیتی خوشه بندی شده اند، مورد استفاده قرار می گیرد. اگر پارامتر مورد نظر ما میانگین جامعه باشد، با به کارگیری روش پساتبقة بندی قضاوتی در مرحله دوم می توانیم کارایی برآوردگر میانگین را افزایش دهیم. این روش را نمونه گیری دومرحله ای با روش پساتبقة بندی قضاوتی در مرحله دوم ($TSCJPS$) می نامیم که انتظار می رود کارایی آن در برآورد میانگین جامعه بیش از نمونه گیری تصادفی خوشه ای دومرحله ای باشد. روش $TSCJPS$ بدین شرح است.

ابتدا همانند روش نمونه گیری خوشه ای دومرحله ای متداول (انتخاب نمونه تصادفی ساده بدون جای گذاری در هر دو مرحله نمونه گیری)، n خوشه تصادفی در مرحله اول انتخاب و در خوشه i ام، m_i واحد انتخاب می شود تا نمونه داده شده در (۲) به دست آید. اکنون با استفاده از روش پساتبقة بندی قضاوتی نمونه تصادفی ساده به دست آمده در مرحله دوم را بهبود می بخشیم. برای این منظور برای هر یک از واحدهای n نمونه m_i تایی یک نمونه کمکی به اندازه $H-1$ به صورت زیر انتخاب می کنیم و رتبه هر یک از Y_{ij} ها در میان نمونه های H تایی را تعیین می کنیم.

$$\begin{matrix} Y_{i1} & Y_{i2} & \dots & Y_{im_i} \\ Y_{i1,1} & Y_{i2,1} & \dots & Y_{im_i,1} \\ \vdots & \vdots & \vdots & \vdots \\ Y_{i1,H-1} & Y_{i2,H-1} & \dots & Y_{im_i,H-1} \end{matrix}$$

نمونه پساتبقة بندی قضاوتی به اندازه m_i به صورت زوج های مرتب $(Y_{i[r]j}, R_{ij} = r)$ ، $i=1, \dots, n$ ، $j=1, \dots, m_i$ ، $r=1, \dots, H$ است که در آن $Y_{i[r]j}$ واحد j ام نمونه است که در پساتبقة r ام خوشه i ام انتخابی قرار می گیرد، یعنی دارای رتبه قضاوتی r است. اکنون برای $i=1, \dots, n$ قرار می دهیم.

$$I_{ij} = \begin{cases} 1 & R_{ij} = r \\ 0 & o.w. \end{cases}, \quad n_{ir} = \sum_{j=1}^{m_i} I_{ij}, \quad I_{ir} = \begin{cases} 1 & n_{ir} > 0 \\ 0 & n_{ir} = 0 \end{cases}, \quad h_i = \sum_{r=1}^H I_{ir}, \quad J_{ir} = \begin{cases} \frac{1}{n_{ir}} & n_{ir} > 0 \\ 0 & o.w. \end{cases}$$

اگر $\hat{\mu}_{i,JPS}$ میانگین نمونه پساتبقة بندی قضاوتی در خوشه i ام باشد، آن گاه

$$\hat{\mu}_{i,JPS} = \frac{1}{h_i} \sum_{r=1}^H \bar{Y}_{i[r]} I_{ir}, \quad (4)$$

که در آن

$$\bar{Y}_{i[r]} = \begin{cases} 0 & n_{ir} = 0 \\ \frac{1}{n_{ir}} \sum_{j=1}^{m_i} Y_{i[r]j} & n_{ir} > 0 \end{cases}$$

مشابه با برآوردگر میانگین $\hat{\mu}_{TSCSRS}$ در نمونه گیری خوشه ای دومرحله ای متداول داده شده در (۳)، برآوردگر میانگین

$$\mu = \frac{1}{NM} \sum_{i=1}^N M_i \mu_i$$

جامعه $TSCJPS$ را بدین صورت معرفی می کنیم

$$\hat{\mu}_{TSCJPS} = \frac{1}{nM} \sum_{i=1}^n M_i \hat{\mu}_{i,JPS} = \frac{1}{nM} \sum_{i=1}^n \sum_{r=1}^H \frac{M_i Y_{i[r]}^-}{h_i} I_{ir}$$

ویژگی‌های $\hat{\mu}_{TSCJPS}$ در قضیه‌های زیر آورده شده است.

قضیه ۳. براوردگر $\hat{\mu}_{TSCJPS}$ براوردگری نارایب برای میانگین جامعه، μ است.

برهان: با توجه به قسمت الف قضیه ۱، در مرحله دوم نمونه‌گیری خوشه‌ای دومرحله‌ای برای هر $i = 1, \dots, n$ ، $\hat{\mu}_{i,JPS}$ داده‌شده در (۴) براوردگری نارایب برای میانگین خوشه i ام، μ_i است. بنابر این با تخصیص دادن اندیس ۱ و ۲ به ترتیب برای مراحل اول و دوم نمونه‌گیری داریم $E_2(\hat{\mu}_{i,JPS}) = \mu_i$ ، $i = 1, \dots, n$ ، و در نتیجه

$$\begin{aligned} E(\hat{\mu}_{TSCJPS}) &= E\left(\frac{1}{nM} \sum_{i=1}^n M_i \hat{\mu}_{i,JPS}\right) \\ &= E_1\left\{E_2\left(\frac{1}{nM} \sum_{i=1}^n M_i \hat{\mu}_{i,JPS}\right)\right\} \\ &= E_1\left\{\frac{1}{nM} \sum_{i=1}^n M_i E_2(\hat{\mu}_{i,JPS})\right\} \quad (5) \\ &= E_1\left\{\frac{1}{nM} \sum_{i=1}^n M_i \mu_i\right\} \\ &= \frac{1}{NM} \sum_{i=1}^N M_i \mu_i = \mu \end{aligned}$$

برابری آخر رابطه (۵) به دلیل آن است که در مرحله اول نمونه‌گیری $M_1 \mu_1, \dots, M_n \mu_n$ نمونه‌ای تصادفی ساده از $M_1, \mu_1, \dots, M_n, \mu_n$ است.

قضیه ۴. در روش $TSCJPS$ ، واریانس $\hat{\mu}_{TSCJPS}$ عبارت است از

$$Var(\hat{\mu}_{TSCJPS}) = \left(1 - \frac{n}{N}\right) \frac{S_b^2}{nM^2} + \frac{1}{nM^2} E_1 \left\{ \frac{1}{n} \sum_{i=1}^n M_i^2 E_2 \left(\frac{J_{i1} I_{i1}}{h_i^2} \right) \sum_{r=1}^H \sigma_{i[r]}^2 \right\} + D^2$$

که در آن

$$D^2 = \frac{1}{nM^2} E_1 \left\{ \frac{1}{n} \sum_{i=1}^n M_i^2 Var_2 \left(\frac{I_{i1}}{h_i} \right) \frac{H}{H-1} \sum_{r=1}^H (\mu_{i[r]} - \mu_i)^2 \right\}$$

و $\mu_{i[r]}$ و $\sigma_{i[r]}^2$ به ترتیب میانگین و واریانس r امین آماره مرتب قضاوتی در نمونه تصادفی به اندازه H در خوشه i ام هستند.

برهان: با استفاده از رابطه واریانس مکرر داریم

$$Var(\hat{\mu}_{TSCJPS}) = Var_1[E_2(\hat{\mu}_{TSCJPS})] + E_1[Var_2(\hat{\mu}_{TSCJPS})] = A + B \quad (6)$$

از رابطه (۵) داریم که $E_2(\hat{\mu}_{TSCJPS}) = \frac{1}{nM} \sum_{i=1}^n M_i \mu_i$ ، بنابراین از ویژگی‌های نمونه‌گیری تصادفی ساده بدون جای‌گذاری در مرحله اول نمونه‌گیری داریم.

$$A = Var_1\left[E_2(\hat{\mu}_{TSCJPS})\right] = Var_1\left[\frac{1}{nM} \sum_{i=1}^n M_i \mu_i\right] = \left(1 - \frac{n}{N}\right) \frac{S_b^2}{nM^2} \quad (7)$$

از طرفی با استفاده از قسمت دوم قضیه ۱، در مرحله دوم نمونه‌گیری در خوشه i ام، $i = 1, \dots, n$ داریم

$$Var_2(\hat{\mu}_{i,JPS}) = E_2 \left(\frac{J_{i1} I_{i1}}{h_i^2} \right) \sum_{r=1}^H \sigma_{i[r]}^2 + Var_2 \left(\frac{I_{i1}}{h_i} \right) \frac{H}{H-1} \sum_{r=1}^H (\mu_{i[r]} - \mu_i)^2$$

بنابر این با توجه به رابطه (۶) داریم که

$$\begin{aligned} B &= E_1 [Var_2(\hat{\mu}_{TSCJPS})] = E_1 \left[Var_2 \left(\frac{1}{nM} \sum_{i=1}^n M_i \hat{\mu}_{i,JPS} \right) \right] \\ &= E_1 \left[\frac{1}{n^2 \bar{M}^2} \sum_{i=1}^n M_i^2 Var_2(\hat{\mu}_{i,JPS}) \right] \\ &= E_1 \left[\frac{1}{n^2 \bar{M}^2} \sum_{i=1}^n M_i^2 E_2 \left(\frac{J_{i1} I_{i1}}{h_i^2} \right) \sum_{r=1}^H \sigma_{i[r]}^2 + \frac{1}{n^2 \bar{M}^2} \sum_{i=1}^n M_i^2 Var_2 \left(\frac{I_{i1}}{h_i} \right) \frac{H}{H-1} \sum_{r=1}^H (\mu_{i[r]} - \mu_i)^2 \right] \quad (\lambda) \\ &= \frac{1}{n \bar{M}^2} E_1 \left\{ \frac{1}{n} \sum_{i=1}^n M_i^2 E_2 \left(\frac{J_{i1} I_{i1}}{h_i^2} \right) \sum_{r=1}^H \sigma_{i[r]}^2 \right\} + D^2 \end{aligned}$$

با جای گذاری روابط (۷) و (۸) در رابطه (۶) نتیجه حاصل می شود.

نمونه گیری دومرحله ای با به کارگیری روش پساطبقه بندی قضاوتی تعمیم یافته در مرحله دوم

در این بخش می خواهیم با استفاده از برآوردهای تعمیم یافته که به وسیله وانگ و همکاران [۲۵] و محمدقاسمی و همکاران [۲۶] برای نمونه گیری پساطبقه بندی قضاوتی ارائه شده اند، دو برآوردهای تعمیم یافته برای برآوردهای میانگین جامعه با استفاده از نمونه گیری دومرحله ای با به کارگیری روش پساطبقه بندی قضاوتی تعمیم یافته در مرحله دوم ارائه دهیم. ابتدا برآوردهای وانگ و همکاران [۲۵] را بیان می کنیم. برای این منظور میانگین مشاهده های دارای رتبه r را با $\mu_{[r]}$ نمایش می دهیم، وانگ و همکاران [۲۵] با استفاده از لم زیر نشان دادند $\mu_{[1]} \leq \mu_{[2]} \leq \dots \leq \mu_{[H]}$ است.

لم ۲. فرض کنید $Y_{[1]}, \dots, Y_{[H]}$ به طور تصادفی مرتب شده باشد. در این صورت برای هر y داریم

$$F_{[1]}(y) \geq \dots \geq F_{[H]}(y)$$

که در آن $F_{[r]}(y)$ تابع توزیع تجمعی درون پساطبقه r ام است.

برهان: برای اثبات به وانگ و همکاران [۲۵] مراجعه شود.

توجه کنید که اگر دو توزیع به طور تصادفی مرتب شده باشند (برای مثال $F(x) \geq G(x)$)، آن گاه برای هر تابع φ غیر نزولی داریم

$$E_G(\varphi(X)) \geq E_F(\varphi(X))$$

وانگ و همکاران [۲۵] با قرار دادن $\varphi(X) = X$ به این نتیجه رسیدند:

$$\mu_{[1]} \leq \mu_{[2]} \leq \dots \leq \mu_{[H]}$$

سپس با جایگزین کردن شکل هم نوا شده میانگین های نمونه $\bar{Y}_{[r]}$ در رابطه $\hat{\mu}_{JPS} = \frac{1}{h} \sum_{r=1}^H \bar{Y}_{[r]} I_r$ برآوردهای میانگین

هم توان را در روش نمونه گیری پساطبقه بندی قضاوتی بدین صورت به دست آوردند.

$$\hat{\mu}_w = \frac{1}{h} \sum_{r=1}^H \bar{Y}_{[r]}^{iso} I_r$$

به طوری که به ازای $n_r > 0$ ، $\bar{Y}_{[r]}^{iso} = \max_{t \leq r} \min_{s \geq r} \sum_{g=t}^s \frac{n_g \bar{Y}_{[g]}}{n_{rs}}$ و $n_{rs} = \sum_{g=t}^s n_g$ است. از این رو، ما از این روش در

نمونه گیری دومرحله ای استفاده کرده و برآوردهای میانگین جامعه ارائه می دهیم. برای این منظور در نمونه گیری

خوشه‌ای دومرحله‌ای، فرض کنید با انتخاب n خوشه از N خوشه در مرحله اول نمونه‌گیری و سپس m_i واحد از M_i واحد خوشه i ام در مرحله دوم نمونه‌گیری، $\hat{\mu}_{wi}$ بروردگر پیشنهادی وانگ و همکاران [۲۵] در خوشه i ام انتخابی باشد. در این صورت با توجه به این که $\mu = \frac{1}{NM} \sum_{i=1}^N M_i \mu_i$ ، بروردگری جدید برای میانگین جامعه در روش نمونه‌گیری دومرحله‌ای بدین صورت ارائه می‌دهیم:

$$\hat{\mu}_{new1} = \frac{1}{n} \sum_{i=1}^n \frac{M_i}{M} \hat{\mu}_{wi}$$

حال در ادامه بروردگر تعمیم‌یافته محمدقاسمی و همکاران [۲۶] را بیان می‌کنیم تا با استفاده از آن بروردگر دیگری برای برورد میانگین در نمونه‌گیری دومرحله‌ای ارائه دهیم.

محمدقاسمی و همکاران [۲۶] با توجه به ایده‌ای که بیان شد به جای $\{(Y_j, R_j), j=1, \dots, m\}$ از نمونه مرتب‌شده آن استفاده کردند. به این ترتیب که $\{(Y_{(j)}, R_{[j]}), j=1, \dots, m\}$ را نمونه مرتب‌شده طرح JPS در نظر گرفتند که در آن $Y_{(1)} \leq Y_{(2)} \leq \dots \leq Y_{(m)}$ آماره‌های مرتب‌شده Y_1, Y_2, \dots, Y_m هستند و $R_{[j]}$ رتبه متناظر با $Y_{(j)}$ است. در این حالت ممکن است رابطه $R_{[1]} \leq R_{[2]} \leq \dots \leq R_{[m]}$ برقرار نباشد. ایده معرفی بروردگر جدید به وسیله محمدقاسمی و همکاران [۲۶] این است که از تصادفی قرار گرفتن هر یک از مشاهدات درون طبقات چشم‌پوشی کرده و مشاهدات را به صورت $\{(Y_{(j)}, R_{(j)}), j=1, \dots, m\}$ در نظر بگیریم که در آن $Y_{(1)} \leq Y_{(2)} \leq \dots \leq Y_{(m)}$ و $R_{(1)} \leq R_{(2)} \leq \dots \leq R_{(m)}$ باشد. در نتیجه بروردگر پیشنهادی محمدقاسمی و همکاران [۲۶] بدین صورت است:

$$\hat{\mu}_g = \frac{1}{H} \sum_{r=1}^H \bar{Y}_{(r)}^*$$

که در آن

$$\bar{Y}_{(r)}^* = \frac{1}{n_r} \sum_{j=1}^m Y_{(j)} I_{rj}, \quad I_{rj} = \begin{cases} 1 & R_{(j)} = r \\ 0 & \text{در غیر این صورت} \end{cases}, \quad n_r = \sum_{j=1}^m I_{rj}$$

اگر $n_r > 0$ باشد. چنانچه $n_r = 0$ باشد، در این صورت $\bar{Y}_{(r)}^*$ به صورت میانگین آمیخته رده یا رده‌های مجاور آن تعریف می‌شود. از آنجاکه محاسبه میانگین توان دوم خطا و اریبی بروردگر $\hat{\mu}_g$ به آسانی امکان‌پذیر نیست محمدقاسمی و همکاران [۲۶] با استفاده از یک مطالعه شبیه‌سازی نشان دادند که این بروردگر نسبت به بروردگر میانگین استاندارد پس‌اطبقه‌بندی قضاوتی بهتر عمل می‌کند. از این رو، ما از این روش در نمونه‌گیری دومرحله‌ای به جای نمونه‌گیری تصادفی ساده استفاده می‌کنیم و بروردگری برای میانگین جامعه ارائه می‌دهیم. در نمونه‌گیری خوشه‌ای دومرحله‌ای، فرض کنید با انتخاب n خوشه از N خوشه در مرحله اول نمونه‌گیری و سپس انتخاب m_i واحد از M_i واحد خوشه i ام در مرحله دوم نمونه‌گیری، $\hat{\mu}_{gi}$ بروردگر پیشنهادی محمدقاسمی و همکاران [۲۶] در خوشه i ام انتخابی باشد. در این صورت با توجه به این که $\mu = \frac{1}{nM} \sum_{i=1}^N M_i \mu_i$ است بروردگر جدیدی برای میانگین جامعه در روش نمونه‌گیری دومرحله‌ای بدین صورت ارائه می‌دهیم:

$$\hat{\mu}_{new2} = \frac{1}{n} \sum_{i=1}^n \frac{M_i}{M} \hat{\mu}_{gi}$$

شبیه‌سازی

در این بخش با استفاده از دو مطالعه شبیه‌سازی شده، کارایی روش $TSCJPS$ نسبت به نمونه‌گیری دومرحله‌ای تصادفی با به کارگیری نمونه‌گیری تصادفی ساده بدون جای‌گذاری در هر دو مرحله بررسی می‌شود. برای مقایسه کارایی برآوردگرها به دلیل دشوار بودن محاسبه میانگین توان دوم خطا و اریبی برآوردگرهای جدید از شبیه‌سازی مونت‌کارلو استفاده می‌کنیم. در پژوهش اول از داده‌های واقعی و در پژوهش دوم از داده‌های تولید شده از دو توزیع متقارن و نامتقارن استفاده می‌کنیم.

در پژوهش اول، از داده‌های گل‌خانه‌های ایران که در سرشماری عمومی کشاورزی ۱۳۸۲ به وسیله مرکز آمار ایران [۲۸] جمع‌آوری شده‌اند، استفاده می‌شود. در این بررسی واحدهای نمونه‌گیری مرحله اول یک استان یا در مواردی چند استان کوچک است که با هم ادغام شده‌اند و واحدهای نمونه‌گیری مرحله دوم گل‌خانه‌ها هستند.

فرض کنید بخواهیم میانگین ارزش محصولات گل‌خانه‌های کشور را برآورد کنیم. برای این منظور برآوردگرهای پیشنهادی و برآوردگر میانگین در نمونه‌گیری خوشه‌ای دومرحله‌ای ساده را استفاده کرده و کارایی برآوردگرهای حاصل یعنی، $\hat{\mu}_{TSCJPS}$ ، $\hat{\mu}_{new_1}$ ، $\hat{\mu}_{new_2}$ و $\hat{\mu}_{TSCSRS}$ را با یکدیگر مقایسه می‌کنیم. در محاسبه برآوردها نیاز به رتبه‌بندی واحدهای نمونه داریم. برای رتبه‌بندی ناقص از مدل دل و کلاتر [۳۰] استفاده می‌کنیم. در این مدل برای رتبه‌بندی از یک متغیر کمکی X که از روی متغیر اصلی Y مطابق رابطه زیر به دست می‌آید استفاده می‌شود (دل و کلاتر [۳۰]).

$$X = \rho \left(\frac{Y - \mu}{\sigma} \right) + \sqrt{1 - \rho^2} Z \quad (9)$$

که در آن μ و σ میانگین و انحراف معیار متغیر Y (در خوشه مربوط)، ρ ضریب هم‌بستگی X و Y ، و Z متغیر تصادفی نرمال استاندارد است. هر چه مقدار ρ به یک نزدیک‌تر باشد رتبه‌بندی دقیق‌تر است. از این‌رو، در شبیه‌سازی دو مقدار $\rho = 0/99$ برای رتبه‌بندی قوی و $\rho = 0/50$ برای رتبه‌بندی ضعیف در نظر گرفته می‌شود. با تولید داده Z از توزیع نرمال استاندارد و قرار دادن در (۹) مقادیر متغیر کمکی X به دست می‌آید و رتبه‌بندی براساس مقادیر X انجام می‌شود.

در داده‌های گل‌خانه‌ای $N = 25$ واحد نمونه‌گیری مرحله اول PSU وجود دارد. تعداد واحدهای نمونه‌گیری مرحله دوم در هر یک از PSU ها در جدول ۱ نشان داده شده است. در مرحله اول نمونه‌گیری n واحد (استان) به روش نمونه‌گیری تصادفی ساده بدون جای‌گذاری انتخاب می‌شود. سپس در هر یک از PSU های انتخاب شده m_i واحد نمونه‌گیری مرحله دوم (گل‌خانه‌ها) با چهار روش پساتبقه‌بندی قضاوتی، پساتبقه‌بندی قضاوتی تعمیم‌یافته و نمونه‌گیری تصادفی ساده بدون جای‌گذاری انتخاب می‌شود. برای مقایسه کارایی چهار روش، انجام نمونه‌گیری ۵۰۰۰۰ بار تکرار می‌شود. نتایج برای مقادیر مختلف n ، m_i و H ارائه می‌شود.

برای برآورد میانگین جامعه، در هر تکرار نمونه‌گیری میانگین نمونه‌های حاصل از سه روش محاسبه می‌شود. انتخاب نمونه‌ها و محاسبه برآوردگرها با استفاده از نرم‌افزار R انجام می‌گیرد. مقدار دقیق میانگین جامعه که از چارچوب قابل محاسبه است

بنابراین می‌توان برآوردی از میانگین توان دوم خطای هر یک از برآوردگرهای $\hat{\mu}_k$ برابر با $\mu = 40/5$ است. یعنی $MSE(\hat{\mu})$ را بدین صورت به دست آورد:

$$MSE(\hat{\mu}_k) = \frac{1}{50000} \sum_{d=1}^{50000} (\hat{\mu}_{dk} - \mu)^2, \quad k = 1, 2, 3, 4$$

که در آن $\hat{\mu}_{d_1}, \hat{\mu}_{d_2}, \hat{\mu}_{d_3}, \hat{\mu}_{d_4}$ به ترتیب برآوردهای $\hat{\mu}_{TSCIPS}, \hat{\mu}_{new_1}, \hat{\mu}_{new_2}$ و $\hat{\mu}_{TSCRS}$ در d آمین تکرار نمونه‌گیری هستند. برای مقایسه کارایی هر دو برآوردها از کارایی نسبی برآوردهای حاصل استفاده می‌کنیم، کارایی نسبی $\hat{\mu}_1$ نسبت به $\hat{\mu}_2$ بدین صورت تعریف می‌شود:

$$RE(\hat{\mu}_1, \hat{\mu}_2) = \frac{MSE(\hat{\mu}_2)}{MSE(\hat{\mu}_1)}$$

جدول ۱. تعداد واحدهای نمونه‌گیری مرحله دوم در هر PSU

PSU_i	M_i	PSU_i	M_i	PSU_i	M_i
۱	۴۲	۲	۱۶۲	۳	۵۳۸
۴	۳۸	۵	۳۳	۶	۲۰
۷	۱۶۷	۸	۹۳۶	۹	۶۸۰
۱۰	۶۱	۱۱	۹۳	۱۲	۴۰
۱۳	۱۴	۱۴	۲۶	۱۵	۲۷
۱۶	۲۰	۱۷	۱۴	۱۸	۲۷۵
۱۹	۳۲	۲۰	۷۵۰	۲۱	۸۴
۲۲	۱۴	۲۳	۱۸	۲۴	۲۶
۲۵	۳۰				

کارایی نسبی برآوردهای $\hat{\mu}_{TSCIPS}, \hat{\mu}_{new_1}, \hat{\mu}_{new_2}$ نسبت به برآوردها $\hat{\mu}_{TSCRS}$ را به ترتیب با RE_{TSCIPS} ، RE_1 و RE_2 نمایش می‌دهیم که برای مقادیر $n = 3, 6, 10, H = 2, 4, 6, 8, 10$ و $m_i = 3, 10$ در جدول‌های ۲ و ۳ ارائه دادیم.

جدول‌های ۲ و ۳ نشان می‌دهند که برآورد حاصل از روش $TSCRS$ تقریباً کارایی برابری با برآوردها حاصل از نمونه‌گیری دومرحله‌ای تصادفی ساده بدون جای‌گذاری در هر مرحله نمونه‌گیری دارد. وقتی که از روش پس‌اطبقه‌بندی قضاوتی تعمیم‌یافته (روش وانگ و همکاران [۲۵] و روش محمدقاسمی و همکاران [۲۶]) در مرحله دوم نمونه‌گیری دو مرحله‌ای استفاده می‌شود، برآوردهای حاصل از این روش در بیش‌تر موارد کارایی بیش‌تری نسبت به برآورد حاصل از نمونه‌گیری دومرحله‌ای تصادفی ساده بدون جای‌گذاری در هر مرحله نمونه‌گیری دارند، هم‌چنین نسبت به برآوردها پس‌اطبقه‌بندی قضاوتی ساده نیز عمل‌کرد بهتری دارند. مقایسه مقادیر کارایی نسبی، نشان می‌دهد که با افزایش m_i برای N, n, ρ و H ثابت، کارایی برآوردها $\hat{\mu}_{TSCIPS}$ نسبت به برآوردها $\hat{\mu}_{TSCRS}$ تقریباً ثابت، اما دقت برآوردهای $\hat{\mu}_{new_1}$ و $\hat{\mu}_{new_2}$ نسبت به $\hat{\mu}_{TSCRS}$ کاهش می‌یابد، هم‌چنین وقتی که رتبه‌بندی قوی است با افزایش H برای N, n و m_i ثابت، کارایی برآوردها $\hat{\mu}_{TSCIPS}$ نسبت به برآوردها $\hat{\mu}_{TSCRS}$ تقریباً ثابت است، ولی بهبود کارایی برآوردهای $\hat{\mu}_{new_1}$ و $\hat{\mu}_{new_2}$ نسبت به $\hat{\mu}_{TSCRS}$ همواره افزایشی است. برآوردهای $\hat{\mu}_{new_1}$ و $\hat{\mu}_{new_2}$ وابسته به رتبه‌بندی هستند، هر چه رتبه‌بندی دقیق‌تر باشد

برآوردگرهای $\hat{\mu}_{new_1}$ و $\hat{\mu}_{new_2}$ برآوردگرهای کارایی هستند. زمانی که m_i کوچک ($m_i = 3$) و رتبه‌بندی ضعیف باشد با افزایش H برای N و n ثابت برآوردگرهای $\hat{\mu}_{new_1}$ و $\hat{\mu}_{new_2}$ برآوردگرهای بهتری در مقایسه با برآوردگرهای $\hat{\mu}_{TSCSRS}$ و $\hat{\mu}_{TSCJPS}$ هستند هم‌چنین بهبود کارایی آن‌ها افزایشی است. زمانی که رتبه‌بندی ضعیف و $m_i = 10$ است، با افزایش H برای N و n ثابت این برآوردگرها تقریباً کارایی برابری با برآوردگر $\hat{\mu}_{TSCSRS}$ دارند.

جدول ۲. مقایسه کارایی برآوردگرهای $\hat{\mu}_{new_1}$ ، $\hat{\mu}_{TSCJPS}$ و $\hat{\mu}_{new_2}$ نسبت به برآوردگر $\hat{\mu}_{TSCSRS}$ برای $n = 3$ و $N = 25$

H	ρ	$m_i = 3$			$m_i = 10$		
		RE_1	RE_2	RE_{TSCJPS}	RE_1	RE_2	RE_{TSCJPS}
۲	۰/۹۹	۱/۰۳	۱/۰۳	۱/۰۳	۱/۰۰	۱/۰۰	۱/۰۰
۴		۱/۱۴	۱/۱۷	۰/۹۸	۱/۰۶	۱/۰۶	۱/۰۴
۶		۱/۳۱	۱/۳۴	۰/۹۵	۱/۰۷	۱/۰۸	۱/۰۰
۸		۱/۵۵	۱/۵۹	۰/۹۹	۱/۱۰	۱/۱۰	۱/۰۰
۱۰		۱/۶۹	۱/۷۳	۰/۹۸	۱/۱۶	۱/۱۶	۱/۰۲
۲	۰/۵۰	۱/۰۰	۰/۹۹	۱/۰۰	۰/۹۸	۰/۹۵	۰/۹۷
۴		۱/۰۸	۱/۱۱	۱/۰۱	۱/۰۲	۰/۹۹	۱/۰۰
۶		۱/۰۹	۱/۱۵	۰/۹۵	۰/۹۹	۰/۹۷	۰/۹۶
۸		۱/۲۵	۱/۳۲	۱/۰۱	۱/۰۱	۰/۹۹	۰/۹۸
۱۰		۱/۲۶	۱/۳۷	۰/۹۶	۱/۰۳	۱/۰۱	۰/۹۸

جدول ۳. مقایسه کارایی برآوردگرهای $\hat{\mu}_{TSCJPS}$ ، $\hat{\mu}_{new_1}$ و $\hat{\mu}_{new_2}$ نسبت به برآوردگر $\hat{\mu}_{TSCSRS}$ برای $n = 6$ و $N = 25$

H	ρ	$m_i = 3$			$m_i = 10$		
		RE_1	RE_2	RE_{TSCJPS}	RE_1	RE_2	RE_{TSCJPS}
۲	۰/۹۹	۱/۰۰	۱/۰۰	۱/۰۰	۰/۹۹	۰/۹۹	۰/۹۹
۴		۱/۲۰	۱/۲۳	۱/۰۳	۱/۰۶	۱/۰۶	۱/۰۳
۶		۱/۳۷	۱/۴۱	۰/۹۹	۱/۰۹	۱/۱۰	۱/۰۳
۸		۱/۵۶	۱/۶۰	۱/۰۱	۱/۱۲	۱/۱۳	۱/۰۱
۱۰		۱/۷۱	۱/۷۷	۱/۰۱	۱/۱۶	۱/۱۶	۱/۰۱
۲	۰/۵۰	۰/۹۶	۰/۹۵	۰/۹۵	۰/۹۸	۰/۹۴	۰/۹۷
۴		۱/۰۵	۱/۰۹	۰/۹۷	۰/۹۹	۰/۹۶	۰/۹۷
۶		۱/۱۴	۱/۲۱	۰/۹۸	۱/۰۰	۰/۹۷	۰/۹۷
۸		۱/۲۵	۱/۳۴	۱/۰۱	۱/۰۳	۱/۰۰	۰/۹۹
۱۰		۱/۲۶	۱/۳۷	۰/۹۶	۱/۰۳	۱/۰۱	۰/۹۸

در پژوهش دوم، با تولید داده از توزیع متقارن نرمال و توزیع نامتقارن نمایی، به مقایسه کارایی برآوردگرها می‌پردازیم. در جدول‌های ۴ و ۵ مقایسه کارایی نسبی براساس این توزیع‌ها آورده شده است که خوشه‌بندی بدین‌صورت انجام می‌شود: برای توزیع نرمال در خوشه اول ۴۹ داده از توزیع $N(25,100)$ ، برای خوشه دوم ۴۸ داده از توزیع $N(24,100)$ ، برای خوشه سوم ۴۷ داده از توزیع $N(23,100)$ و به‌همین ترتیب برای خوشه بیست و پنجم، ۲۵ داده از توزیع $N(1,100)$ تولید می‌کنیم. هم‌چنین برای توزیع نمایی در خوشه اول ۴۹ داده از توزیع $E(25)$ ، برای خوشه دوم ۴۸ داده از توزیع $E(24)$ ، برای خوشه سوم ۴۷ داده از توزیع $E(23)$ و به‌همین ترتیب برای خوشه بیست و پنجم، ۲۵ داده از توزیع $E(1)$ تولید می‌کنیم. با توجه به این خوشه‌بندی و داده‌های تولید شده، همانند قبل

کارایی نسبی برآوردگرهای $\hat{\mu}_{TSCJPS}$ ، $\hat{\mu}_{new1}$ و $\hat{\mu}_{new2}$ نسبت به برآوردگر $\hat{\mu}_{TSCSRS}$ که به ترتیب با RE_1 ، RE_2 و RE_{TSCJPS} نمایش داده شده برای مقادیر $H = 2, 4, 6, 8, 10$ ، جدول ۴. مقایسه کارایی برآوردگرهای $\hat{\mu}_{TSCJPS}$ ، $\hat{\mu}_{new1}$ و $\hat{\mu}_{new2}$ نسبت به برآوردگر $\hat{\mu}_{TSCSRS}$ برای $\rho = 0/99$ ، $N = 25$ و $n = 3$

H	توزیع	$m_i = 3$			$m_i = 10$		
		RE_1	RE_2	RE_{TSCJPS}	RE_1	RE_2	RE_{TSCJPS}
۲	نمایی	۰/۹۱	۰/۹۳	۰/۹۱	۰/۹۵	۰/۹۵	۰/۹۵
۴		۱/۱۳	۱/۱۵	۱/۰۰	۱/۰۰	۱/۰۱	۰/۹۸
۶		۱/۲۲	۱/۲۴	۰/۹۷	۱/۰۲	۱/۰۳	۰/۹۸
۸		۱/۳۵	۱/۳۷	۰/۹۹	۱/۱۰	۱/۱۱	۱/۰۱
۱۰		۱/۴۴	۱/۴۶	۱/۰۱	۱/۱۳	۱/۱۴	۱/۰۰
۲	نرمال	۰/۹۸	۰/۹۹	۰/۹۷	۰/۹۶	۰/۹۸	۰/۹۶
۴		۱/۲۶	۱/۳۳	۱/۰۰	۱/۲۵	۱/۳۱	۱/۱۰
۶		۱/۳۹	۱/۴۶	۰/۹۸	۱/۳۹	۱/۴۶	۱/۰۵
۸		۱/۳۰	۱/۳۶	۰/۹۴	۱/۶۲	۱/۷۰	۱/۰۸
۱۰		۱/۳۰	۱/۳۵	۰/۹۷	۱/۶۷	۱/۷۴	۱/۰۱

$n = 3, 6$ ، $\rho = 0/99$ و $m_i = 3, 10$ را در جدول‌های ۴ و ۵ ارائه دادیم. از این جدول‌ها مشاهده می‌شود که در توزیع‌های نرمال و نمایی کارایی برآوردگر $\hat{\mu}_{TSCJPS}$ تقریباً برابر برآوردگر $\hat{\mu}_{TSCSRS}$ است. در هر دو توزیع $\hat{\mu}_{new1}$ و $\hat{\mu}_{new2}$ نسبت به $\hat{\mu}_{TSCSRS}$ در بیش‌تر موارد برآوردگرهای بهتری هستند، هم‌چنین در توزیع نرمال با مقایسه مقادیر کارایی برای $m_i = 3$ (با افزایش H بهبود کارایی نسبی همواره افزایشی نیست) (است)، اما در توزیع نمایی برای m_i ثابت با افزایش H ، بهبود کارایی افزایشی است.

جدول ۵. مقایسه کارایی برآوردگرهای $\hat{\mu}_{TSCJPS}$ ، $\hat{\mu}_{new1}$ و $\hat{\mu}_{new2}$ نسبت به برآوردگر $\hat{\mu}_{TSCSRS}$ برای $\rho = 0/99$ ، $N = 25$ و $n = 6$

H	توزیع	$m_i = 3$			$m_i = 10$		
		RE_1	RE_2	RE_{TSCJPS}	RE_1	RE_2	RE_{TSCJPS}
۲	نمایی	۰/۹۰	۰/۹۱	۰/۹۰	۰/۹۷	۰/۹۷	۰/۹۷
۴		۱/۰۸	۱/۱۱	۰/۹۳	۱/۰۱	۱/۰۲	۰/۹۸
۶		۱/۳۲	۱/۳۵	۱/۰۰	۱/۰۸	۱/۰۸	۱/۰۰
۸		۱/۴۲	۱/۴۴	۰/۹۷	۱/۱۰	۱/۱۰	۰/۹۹
۱۰		۱/۵۳	۱/۵۱	۱/۰۰	۱/۱۲	۱/۱۲	۰/۹۶
۲	نرمال	۰/۹۵	۰/۹۶	۰/۹۴	۰/۹۵	۰/۹۶	۰/۹۴
۴		۱/۲۴	۱/۳۴	۰/۹۸	۱/۲۳	۱/۳۰	۱/۰۷
۶		۱/۳۶	۱/۴۳	۰/۹۸	۱/۴۶	۱/۵۴	۱/۰۸
۸		۱/۲۹	۱/۳۵	۰/۹۵	۱/۶۱	۱/۶۹	۱/۰۲
۱۰		۱/۲۵	۱/۲۸	۱/۰۳	۱/۷۲	۱/۸۰	۱/۰۰

بحث و نتیجه‌گیری

در این مقاله برآوردهای جدیدی برای میانگین جامعه به‌همراه واریانس این برآوردها در طرح نمونه‌گیری دومرحله‌ای با به‌کارگیری روش پساتبقه‌بندی قضاوتی در مرحله دوم و تعمیمی از این برآوردها ارائه شد. همچنین نشان دادیم که برآوردهای پیشنهادی، برآوردهایی نارایب برای میانگین جامعه هستند. نتایج شبیه‌سازی نشان می‌دهد که این برآوردها در بسیاری از موارد نسبت به برآوردها متداول در نمونه‌گیری دومرحله‌ای کارایی بیشتری دارند.

منابع

1. McIntyre G. A., "A Method for Unbiased Selective Sampling Using Ranked Sets", Australian Journal of Agricultural Research, 3 (1952) 385-390.
2. Halls L. K., Dell T. R., "Trial of Ranked Set Sampling for Forage Yields", Forest Science, 12 (1966) 22-26.
3. Takahasi K., Wakimoto K., "On Unbiased Estimates of the Population Mean Based on the Sample Stratified by Means of Ordering", Annals of the Institute of Statistical Mathematics, 20 (1968) 1-31.
4. Samawi H. M., Ahmed M. S., Abo-Dayyeh W., "Estimating of the Population Mean Using Extreme Ranked Set Samples", Biometrical Journal, 38 (1996) 577-586.
5. Muttlak H. A., "Pair Rank Set Sampling", Biometrical Journal, 38 (1996) 879-885.
6. Muttlak H. A., "Median Ranked Set Sampling", Journal of Applied Statistical Science, 6 (1997) 245-255.
7. Al- Saleh M. F., Al-Kadiri M. A., "Double Ranked Set Sampling", Statistics and Probability Letters, 48 (2000) 205-212.
8. Haq A., Brown J., Moltehanova E., Al-Omari A. I., "Partial Ranked Set Sampling Design", Environmetrics, 24 (2013) 201-207.
9. Haq A., Brown J., Moltehanova E., "Hybrid Ranked Set Sampling Scheme", Journal of Statistical Computation and Simulation, 86 (2016) 1-28.
10. Samawi H. M., "Stratified Ranked Set Sampled", Pakistan Journal of Statistics, 12 (1) (1996) 9-16.
11. Samawi H. M., Saeid L. J., "Stratified Extreme Ranked Set Sample with Application to Ratio Estimators", Journal of Modern Applied Statistical Methods, 3 (2004) 117-133.
12. Ibrahim A. A., "Investigating the use of Stratified Percentile Ranked Set Sampling Method for Estimating the Population Mean", Journal of Mathematics, 30 (2011) 351-368.
13. Mahdizadeh M., Zamanzade E., "Stratified pair ranked set sampling", Communications in Statistics-Theory and Methods, 47 (2018) 5904-5915.
14. Nematollahi N., Salehi M. M., Saba R. A., "Two- Stage Cluster Sampling with Ranked Set Sampling in the Secondary Sampling Frame", Communications in Statistics Theory and Methods, 37 (2008) 2404-2415.

15. Haq A., "Two- Stage Cluster Sampling with Hybrid Ranked Set Sampling in the Secondary Sampling Frame", *Communications in Statistics Theory and Methods*, 46 (17) (2017) 8450-8467.
16. Ozturk O., "Two- Stage Cluster Sampling with Ranked Set Sampling Designs", *Annals of the Institute of Statistical Mathematics*, 71 (1) (2017) 63-91.
17. Wang X., Lim J., Stokes L., "Using Ranked Set Sampling With Cluster Randomized Designs for Improved Inference on Treatment Effects", *Journal of the American Statistical Association*, 111 (516) (2017) 1576-1590.
18. Ahn S., Wang X., Lim J., "On Unbalanced Group Sizes in Cluster Randomized Designs Using Balanced Ranked Set Sampling", *Statistics and Probability Letters*, 123 (2017) 210-217.
19. Wang X., Ahn S., Lim J., "Unbalanced Ranked Set Sampling in Cluster Randomized Studies" *Journal of Statistical Planning and Inference*, 187 (2017) 1-16.
20. MacEachern S. N., Stasny E. A., Wolfe D. A., " Judgment Post-Stratification with Imprecise Rankings", *Biometrics*, 60 (2004) 207-215.
21. Frey J., Ozturk O. , "Constrained Estimation Using Judgment Post-Stratification", *Annals of the Institute of Statistical Mathematics*, 63 (2011) 769-789.
22. Frey J., Feeman T.G., "An Improved Mean Estimator for Judgment Post-Stratification", *Computational Statistics and Data Analysis*, 56 (2012) 418-426.
23. Dastbaravarde A., Arghami N. R., Sarmad M., "Some Theoretical Results Concerning Non-parametric Estimation by Using a Judgment Post-Stratification Sample", *Communication in Statistics Theory and Methods*, 45 (2016) 2181-2203.
24. Zamanzade E., Wang X., "Estimation of Population Proportion for Judgment Post-Stratification", *Computational Statistics and Data Analysis*, 112 (2017) 257-269.
25. Wang X., Lim J., Stokes L., "A Nonparametric Mean Estimator for Judgment Post-stratified Data", *Biometrics*, 64 (2008) 355-363.
۲۶. محمد قاسمی، حامد زمان‌زاده، احسان محمدی محمد، "برآوردگر جدید میانگین در طرح نمونه‌گیری طبقه‌بندی قضاوتی با مرتب کردن مشاهدات درون طبقات، مجله علوم آماری، جلد ۱۰، شماره ۱ (۱۳۹۵)، ۱۲۹-۱۳۷.
27. Omidvar S., Jafari Jozani M., Namatollahi N., "Judgment Post-Stratification in Finite Mixture Modeling: An Example in Estimating the Prevalence of Osteoporosis", *Statistical in Medicine*, 37(30) (2018) 4823-4836.
۲۸. مرکز آمار ایران، نتایج سرشماری عمومی کشاورزی (۱۳۸۲).
29. Cochran W. G., "Sampling Techniques", 3rd Edition. Wiley, New York (1977).
30. Dell T. R., Clutter J. L., "Ranked Set Sampling Theory with Order Statistics Background", *Biometrics*, (1972) 545-555.