به‌کارگیری روش پساطبقه‌بندی قضاوتی در نمونه‌گیری خوشه‌ای دومرحله‌ای

علی نجفی مجیدآبادی، طاهره رحمانی، نادر نعمت‌الهی*
دانشگاه علامه طباطبائی، گروه آمار

دریافت ۹۸/۰۳/۰۷ پذیرش ۹۸/۰۸/۱۸

چکیده

روش پساطبقه‌بندی قضاوتی روشی برای طبقه‌بندی مشاهدات با استفاده از یک متغیر کلیدی است، به‌طوری‌که طبقه‌بندی بعد از انتخاب نمونه انجام می‌شود. در این مقاله این روش را در نمونه‌گیری خوشه‌ای دومرحله‌ای به‌کار می‌بریم. به‌عبارت دیگر در نمونه‌گیری خوشه‌ای دومرحله‌ای در مرحله‌ی دوم به‌جای نمونه‌گیری تصادفی ساده بدون جای‌گذاری از روش پساطبقه‌بندی قضاوتی و روش پساطبقه‌بندی قضاوتی تعمیم‌یافته استفاده کرده و براوردگرهای جدیدی را برای میانگین جامعه ارائه می‌کنیم. در نهایت با استفاده از شبیه‌سازی مونت‌کارلو با مجموعه داده‌های واقعی و داده‌های تولیدی از توزیع‌های متقارن و نامتقارن، براوردگرهای پیشنهادی را با براوردگر میانگین نمونه‌گیری خوشه‌ای دومرحله‌ای ساده را مقایسه می‌کنیم. نتایج شبیه‌سازی نشان می‌دهد که، در بیش‌تر حالت‌ها، براوردگرهای پیشنهادی عمل‌کرد بهتری نسبت به براوردگر میانگین نمونه‌گیری دومرحله‌ای ساده دارند.

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

مقدمه

در بررسی‌های آماری، جمع‌اوری داده‌های مورد نظر برای استنباط در مورد جامعه، اغلب از طریق نمونه‌گیری انجام می‌گیرد. در براورد پارامترهای یک جامعه با استفاده از نمونهٔ انتخابی از جامعه، انتخاب روش نمونه‌گیری که با صرف زمان و هزینه کم‌تر، منجر به ارائهٔ براوردگری کارا شود، بسیار مهم است. به‌ویژه در مواقعی که اندازه‌گیری مشخصهٔ بررسی شده در جامعه نیازمند زمان و هزینهٔ بسیار زیادی است. ساده‌ترین و در عین حال پرکاربردترین روش نمونه‌گیری، نمونه‌گیری تصادفی ساده[1] (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 جامعهٔ 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 هزینهٔ زیادی باشد. در بعضی بررسی‌ها، چارچوبی از تک تک اعضای جامعه در دسترس نیست بلکه چارچوب به‌صورت 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 مرحلهٔ نمونه‌گیری خوشه‌ای دومرحله‌ای اغلب از نمونه‌گیری تصادفی ساده استفاده می‌شود. به‌این ترتیب که اگر جامعه به N خوشه تقسیم شده باشد، در مرحلهٔ اول نمونه‌گیری n خوشه را به‌صورت تصادفی ساده انتخاب می‌کنیم و سپس در مرحلهٔ دوم از هر خوشهٔ منتخب تعدادی از واحدها را به‌صورت تصادفی ساده به‌عنوان واحدهای

---

*نویسنده مسئول nematollahi@atu.ac.ir

1. Simple Random Sampling

نمونه اصلی برمی‌گزینیم که به آن نمونه‌گیری خوشه‌ای دومرحله‌ای با انتخاب نمونهٔ تصادفی ساده در مرحلهٔ دوم[1] ($TSCSRS$) گوییم.

در نمونه‌گیری تصادفی ساده به‌علت این که هیچ کنترلی روی اعضای نمونهٔ انتخابی وجود ندارد، ممکن است نمونه‌های انتخابی نمایان‌گر خصوصیات جامعهٔ مورد نظر نباشند و براوردگرهای حاصل دقت کافی نداشته نباشند. در چنین حالتی به‌طور معمول برای افزایش دقت براورگرهای حاصل، اندازهٔ 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 مجموعهٔ رتبه‌ای[2] (RSS) را برای براورد پارامترهای جامعه پیشنهاد کرد که بسیار کاراتر از روش نمونه‌گیری تصادفی ساده با اندازهٔ نمونهٔ برابر است. این افزایش کارایی در نتیجهٔ رتبه‌بندی واحدها حاصل می‌شود که با هزینهٔ کمی صورت می‌گیرد. هالز و دل [۲] با استفاده از این روش نمونه‌گیری میانگین محصول علوفه و میانگین ارتفاع درختان یک جنگل را براورد کردند. تاکاهاسی و واکی‌موتو [۳] میانگین حسابی مشاهده‌های حاصل از نمونهٔ مجموعهٔ رتبه‌ای را به‌عنوان براوردگر ناااریبی برای میانگین جامعه ارائه کردند و نشان دادند که این براوردگر کاراتر از براوردگر میانگین جامعه در حالت نمونه‌گیری تصادفی ساده با اندازهٔ نمونه برابر است.

در دو دههٔ 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[3] ($ERSS$) به‌وسیلهٔ سماوی و همکاران [۴]، روش RSS جفت‌شده[4] ($PaRSS$) به‌وسیلهٔ مطلق [۵]، روش RSS میانه‌ای[5] ($MRSS$) به‌وسیلهٔ مطلق [۶]، روش RSS دوگانه[6] ($DRSS$) به‌وسیلهٔ ال‌صالح و ال‌قدیری [۷]، روش RSS جزئی[7] ($PRSS$) به‌وسیلهٔ حق و همکاران [۸] و روش RSS پیوندی[8] ($HRSS$) به‌وسیلهٔ حق و همکارن [۹] برای براورد میانگین جامعه پیشنهاد شده است. برخی از پژوهش‌گران نیز روش‌های مختلف RSS را در نمونه‌گیری‌های پیچیده‌تر به‌کار 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 مرحلهٔ دوم روش نمونه‌گیری دومرحله‌ای، نعمت‌الهی و همکاران [۱۴] از روش RSS و حق [۱۵] از روش $HRSS$ استفاده کرده‌اند که آنها را به‌ترتیب $TSCRSS$ و $TSCHRSS$ می‌نامیم. آنها نشان داده‌اند که براوردگر میانگین جامعه براساس $TSCRSS$ و $TSCHRSS$ کاراتر از براوردگر میانگین جامعه براساس $TSCSRS$ است. اوزتورک [۱۶] براوردگرهایی برای میانگین جامعهٔ متناهی و واریانس آن در نمونه‌گیری خوشه‌ای دومرحله‌ای با طرح نمونه‌گیری RSS ارائه کرد و نشان داد که کارایی براوردگرهای 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 مجموعهٔ رتبه‌ای نامتعادل[9] ($URSS$) 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 پساطبقه‌بندی قضاوتی[1] (JPS) پیشنهاد شده است

روش پساطبقه‌بندی قضاوتی به‌وسیلهٔ 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 براوردگرهای میانگین در روش پساطبقه‌بندی قضاوتی و روش نمونه‌گیری از مجموعهٔ رتبه‌ای تحت تابع زیان توان دوم خطا ناپذیرفتنی هستند. فری و فیمن [۲۲] براوردگری بهبودیافته برای میانگین روش پساطبقه‌بندی قضاوتی ارائه کردند و نشان دادند که براوردگر جدید ارائه‌شده کارایی بیش‌تری نسبت به براوردگرهای موجود دارد. هم‌چنین دست‌براورده و همکاران [۲۳] و زمان‌زاده و وانگ [۲۴] پژوهش‌هایی در زمینهٔ براورد مشخصه‌های جامعه در نمونه‌گیری پساطبقه‌بندی قضاوتی انجام دادند. وانگ و همکاران [۲۵] و محمدقاسمی و همکارن [۲۶] براوردگرهای 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 مرحلهٔ دوم از نمونه‌گیری تصادفی ساده بدون جای‌گذاری، نمونه‌گیری از مجموعهٔ رتبه‌ای ساده و پیوندی استفاده شده است. از آن‌جاکه روش پساطبقه‌بندی قضاوتی نسبت به نمونه‌گیری تصادفی ساده اغلب منجر به براوردگری کاراتر از میانگین جامعه می‌شود، بنابراین انتظار می‌رود که به‌کارگیری روش پساطبقه‌بندی قضاوتی در مرحلهٔ دوم نمونه‌گیری دومرحله‌ای موجب افزایش کارایی براوردگر میانگین جامعه شود. از این‌رو، در این مقاله برای براورد میانگین جامعه از روش نمونه‌گیری دومرحله‌ای استفاده می‌کنیم که در آن انتخاب نمونه‌ها در مرحلهٔ اول به روش تصادفی ساده بدون جای‌گذاری انجام می‌گیرد. در مرحلهٔ دوم نمونه‌گیری، برای انتخاب نمونه‌ها از روش پساطبقه‌بندی قضاوتی و تعمیم‌یافته‌های آن استفاده می‌کنیم و براوردگرهای جدیدی را برای میانگین جامعه ارائه می‌دهیم. این روش‌ها را با $TSCJPS$ نشان می‌دهیم. در این روش براوردگر نااریب 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 داه‌های تولیدی از دو توزیع متقارن و نامتقارن، نشان می‌دهیم که براوردگر میانگین در نمونه‌گیری دومرحله‌ای با به‌کارگیری روش پساطبقه‌بندی قضاوتی در مرحلهٔ دوم، کارایی بیش‌تری نسبت به نمونه‌گیری دومرحله‌ای معمول بر اساس نمونه‌گیری تصادفی واحدها دارد.

مواد و روش‌ها

روش پساطبقه‌بندی قضاوتی

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

1. Judgment Post Stratification

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

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

که در آن Y_{jr} نمونهٔ کمکی rام، برای رتبه‌بندی Y_j است. برای به‌دست آوردن نمونهٔ پساطبقه‌بندی قضاوتی، هر ستون که دارای H واحد است را به‌صورت افزایشی مرتب می‌کنیم و بدون اندازه‌گیری خصوصیت Y در آن‌ها رتبهٔ Y_j را بین $Y_{j1},...,Y_{j,H-1},Y_j$ برای $j=1,...,m$ به‌صورت قضاوتی تعیین کرده و با R_j نشان می‌دهیم، که R_jها دارای توزیع یکنواخت گسسته در مجموعه‌ی $\{1,...,H\}$ هستند. بنابر این نمونهٔ حاصل از روش *JPS* به‌صورت (Y_j,R_j)، $j=1,...,m$ 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 رتبهٔ یکسان باشند. بنابراین روش *JPS* یک نوع نمونه‌گیری پساطبقه‌بندی است که واحدهای دارای رتبهٔ 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 $U_1,\ldots,U_m$ نمونه‌ای شامل m درخت در یک جنگل و $Y_1,...,Y_m$ ارتفاع این درختان باشند. برای طبقه‌بندی قضاوتی این نمونه به‌ازای هر درخت، $H-1$ درخت را به‌طور تصادفی از جنگل انتخاب کرده و رتبهٔ آن درخت را در این مجموعه‌ی Hتایی 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 رتبهٔ یکسان هستند را در یک طبقه قرار می‌دهیم بنابر این H طبقه خواهیم داشت، که البته همهٔ 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 نمونهٔ پساطبقه‌بندی قضاوتی برای $j=1,...,m$ و $r=1,...,H$ به‌صورت زوج‌های $(Y_{[r]j},R_j=r)$ خواهد بود، که در آن $Y_{[r]j}$ واحد j ام نمونه‌ی mتایی اولیه است که در پساطبقهٔ rام قرار گیرد، یعنی دارای رتبهٔ قضاوتی r باشد.

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

$$I_{rj}=\begin{cases}1 & R_j=r\\ 0 & o.w.\end{cases},\quad n_r=\sum_{j=1}^{m}I_{rj},\quad I_r=\begin{cases}1 & n_r>0\\ 0 & n_r=0\end{cases},\quad h=\sum_{r=1}^{H}I_r,\quad J_r=\begin{cases}\dfrac{1}{n_r} & n_r>0\\ 0 & o.w.\end{cases}$$

در این صورت براوردگر میانگین جامعه بر اساس روش پساطبقه‌بندی قضاوتی به‌صورت (۱) است:

$$\hat{\mu}_{JPS}=\frac{1}{h}\sum_{r=1}^{H}\overline{Y}_{[r]}I_r, \tag{۱}$$

که در آن

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

برای به‌دست آوردن ویژگی‌های براوردگر $\hat{\mu}_{JPS}$ به لم زیر که به‌وسیلهٔ دست‌براورده و همکاران [۲۳] ارائه‌شده است نیاز داریم.

لم ۱. در روش پساطبقه‌بندی قضاوتی داریم:

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

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

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

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

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

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

برای برهان به‌دست‌براورده و همکاران [۲۳] مراجعه شود.

قضیهٔ ۱. اگر $\mu_{[r]}$ و $\sigma^2_{[r]}$ به ترتیب میانگین و واریانس rامین آمارهٔ مرتب قضاوتی در یک نمونه تصادفی ساده به اندازه‌ی H باشند، آن‌گاه امید و واریانس براوردگر میانگین پساطبقه‌بندی قضاوتی $\hat{\mu}_{JPS}$ داده‌شده در (۱) بدین‌صورت است:

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

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

برای برهان به‌دست‌براورده و همکاران [۲۳] مراجعه شود.

با توجه به قضیهٔ ۱، $\hat{\mu}_{JPS}$ براوردگری نااریب برای میانگین جامعه است.

یافته‌ها

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

نمونه‌گیری خوشه‌ای دومرحله‌ای یکی از روش‌های نمونه‌گیری است که در عمل بسیار پرکاربرد است. این روش نمونه‌گیری دارای دومرحله است. در مرحلهٔ اول، نمونه‌ها از واحدهای نمونه‌گیری اولیه[1] (PSU) انتخاب می‌شوند. در مرحلهٔ دوم، نمونه‌ها از واحدهای نمونه‌گیری ثانویه[2] (SSU) در هر یک از PSUهای انتخاب‌شده، انتخاب می‌شوند. در نمونه‌گیری خوشه‌ای دومرحله‌ای اغلب روشی برای انتخاب PSU و SSUها از قبل مشخص می‌شود. روش نمونه‌گیری تصادفی ساده اغلب برای انتخاب PSU و SSUها در هر دو مرحله استفاده می‌شود.

فرض کنید جامعهٔ مورد نظر دارای N خوشه (N واحد نمونه‌گیری مرحلهٔ اول، PSU) بوده به‌طوری‌که iامین خوشه دارای M_i واحد (M_i واحد نمونه‌گیری مرحلهٔ دوم، SSU) است. هم‌چنین فرض کنید خصیصهٔ مورد نظر Y دارای میانگین μ و واریانس σ^2 بوده و میانگین و واریانس آن در خوشهٔ iام به‌ترتیب 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 مرحلهٔ اول، m_i تعداد واحدهای انتخاب‌شده از نمونهٔ iام در مرحلهٔ دوم و Y_{ij}، مقدار خصیصهٔ مورد نظر برای jامین نمونهٔ تصادفی ساده انتخابی ($j=1,...,m_i$) در iامین خوشه انتخابی ($i=1,...,n$) باشد، آن‌گاه نمونهٔ به‌دست آمده بدین‌صورت است:

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

براوردگر نااریب میانگین جامعه براساس این روش به‌صورت (۳) است:

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

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

قضیهٔ ۲. در روش $TSCSRS$ براوردگر $\hat{\mu}_{TSCSRS}$ براوردگری نااریب برای میانگین جامعه است و دارای واریانس زیر است

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

1. Primary Sampling Unit
2. Secondary Sampling Unit

که در آن $S_b^2=\frac{1}{N-1}\sum_{i=1}^{N}(M_i\mu_i-\overline{M}\mu)^2$.

برای برهان به ککران [۲۹] مراجعه شود.

در عمل، اغلب این روش نمونه‌گیری در جامعه‌های بزرگ که به گروه‌های جمعیتی خوشه‌بندی شده‌اند، مورد استفاده قرار می‌گیرد. اگر پارامتر مورد نظر ما میانگین جامعه باشد، با به‌کارگیری روش پساطبقه‌بندی قضاوتی در مرحله‌ی دوم می‌توانیم کارایی براوردگر میانگین را افزایش دهیم. این روش را نمونه‌گیری دومرحله‌ای با روش پساطبقه‌بندی قضاوتی در مرحلهٔ دوم[1] $(TSCJPS)$ می‌نامیم که انتظار می‌رود کارایی آن در براورد میانگین جامعه بیش از نمونه‌گیری تصادفی خوشه‌ای دومرحله‌ای باشد. روش $TSCJPS$ بدین‌شرح است.

ابتدا همانند روش نمونه‌گیری خوشه‌ای دومرحله‌ای متداول (انتخاب نمونهٔ تصادفی ساده بدون جای‌گذاری در هر دو مرحلهٔ نمونه‌گیری)، n خوشهٔ تصادفی در مرحلهٔ اول انتخاب و در خوشهٔ iام، m_i واحد انتخاب می‌شود تا نمونهٔ داده‌شده در (۲) به‌دست آید. اکنون با استفاده از روش پساطبقه‌بندی قضاوتی نمونه تصادفی سادهٔ به‌دست آمده در مرحلهٔ دوم را بهبود می‌بخشیم. برای این منظور برای هر یک از واحدهای n نمونهٔ m_iتایی یک نمونهٔ کمکی به اندازه‌ی $H-1$ به‌صورت زیر انتخاب می‌کنیم و رتبهٔ هر یک از Y_{ij}ها در میان نمونه‌های Hتایی را تعیین می‌کنیم.

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

نمونهٔ پساطبقه‌بندی قضاوتی به اندازهٔ m_i به‌صورت زوج‌های 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 پساطبقه rام خوشهٔ iام انتخابی قرار می‌گیرد، یعنی دارای رتبهٔ قضاوتی r است. اکنون برای $i=1,\dots,n$ قرار می‌دهیم.

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

اگر $\hat{\mu}_{i,JPS}$ میانگین نمونهٔ پساطبقه‌بندی قضاوتی در خوشهٔ iام باشد، آن‌گاه

$$\hat{\mu}_{i,JPS}=\frac{1}{h_i}\sum_{r=1}^{H}\overline{Y}_{i[r]}I_{ir}, \qquad (۴)$$

که در آن

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

مشابه با براوردگر میانگین $\hat{\mu}_{TSCSRS}$ در نمونه‌گیری خوشه‌ای دومرحله‌ای متداول داده‌شده در (۳)، براوردگر میانگین جامعه $\mu=\frac{1}{N\overline{M}}\sum_{i=1}^{N}M_i\mu_i$ در روش $TSCJPS$ را بدین‌صورت معرفی می‌کنیم

1. Two- Stage Cluster Judgement Post Stratification

$$\hat{\mu}_{TSCJPS}=\frac{1}{n\bar{\bar{M}}}\sum_{i=1}^{n}M_i\,\hat{\mu}_{i,JPS}=\frac{1}{n\bar{\bar{M}}}\sum_{i=1}^{n}\sum_{r=1}^{H}\frac{M_i}{h_i}\bar{Y}_{i[r]}I_{ir}.$$

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

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

برهان: با توجه به قسمت الف قضیهٔ ۱، در مرحلهٔ دوم نمونه‌گیری خوشه‌ای دومرحله‌ای برای هر $i=1,\dots,n$، $\hat{\mu}_{i,JPS}$ داده‌شده در (۴) براوردگری نااریب برای میانگین خوشهٔ iام، 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

$$\begin{aligned}E(\hat{\mu}_{TSCJPS})&=E\left(\frac{1}{n\bar{M}}\sum_{i=1}^{n}M_i\,\hat{\mu}_{i,JPS}\right)\\&=E_1\left\{E_2\left(\frac{1}{n\bar{\bar{M}}}\sum_{i=1}^{n}M_i\,\hat{\mu}_{i,JPS}\right)\right\}\\&=E_1\left\{\frac{1}{n\bar{M}}\sum_{i=1}^{n}M_i\,E_2\left(\hat{\mu}_{i,JPS}\right)\right\}\\&=E_1\left\{\frac{1}{n\bar{M}}\sum_{i=1}^{n}M_i\,\mu_i\right\}\\&=\frac{1}{N\bar{\bar{M}}}\sum_{i=1}^{N}M_i\,\mu_i=\mu\end{aligned}\tag{۵}$$

برابری آخر رابطهٔ (۵) به‌دلیل آن است که در مرحلهٔ 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.

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

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

که در آن

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

و $\mu_{i[r]}$ و $\sigma_{i[r]}^2$ به‌ترتیب میانگین و واریانس rامین آمارهٔ مرتب قضاوتی در نمونهٔ تصادفی به اندازهٔ H در خوشهٔ iام هستند.

برهان: با استفاده از رابطهٔ واریانس مکرر داریم

$$Var(\hat{\mu}_{TSCJPS})=Var_1\left[E_2(\hat{\mu}_{TSCJPS})\right]+E_1\left[Var_2\left(\hat{\mu}_{TSCJPS}\right)\right]=A+B\tag{۶}$$

از رابطهٔ (۵) داریم که $E_2\left(\hat{\mu}_{TSCJPS}\right)=\frac{1}{n\bar{\bar{M}}}\sum_{i=1}^{n}M_i\,\mu_i$. بنابراین از ویژگی‌های نمونه‌گیری تصادفی ساده بدون جای‌گذاری در مرحلهٔ اول نمونه‌گیری داریم.

$$A=Var_1\left[E_2\left(\hat{\mu}_{TSCJPS}\right)\right]=Var_1\left[\frac{1}{n\bar{M}}\sum_{i=1}^{n}M_i\,\mu_i\right]=\left(1-\frac{n}{N}\right)\frac{S_b^2}{n\bar{M}^2}\tag{۷}$$

از طرفی با استفاده از قسمت دوم قضیهٔ ۱، در مرحلهٔ دوم نمونه‌گیری در خوشهٔ iام، $i=1,\dots,n$ 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 رابطهٔ (۶) داریم که

$$\begin{aligned}B&=E_1\left[Var_2(\hat{\mu}_{TSCJPS})\right]=E_1\left[Var_2\left(\frac{1}{n\bar{M}}\sum_{i=1}^{n}M_i\,\hat{\mu}_{i,JPS}\right)\right]\\&=E_1\left[\frac{1}{n^2\bar{M}^2}\sum_{i=1}^{n}M_i^2Var_2(\hat{\mu}_{i,JPS})\right]\\&=E_1\left[\frac{1}{n^2\bar{M}^2}\sum_{i=1}^{n}M_i^2E_2\left(\frac{J_{i1}I_{i1}}{h_i^2}\right)\sum_{r=1}^{H}\sigma_{i[r]}^2+\frac{1}{n^2\bar{M}^2}\sum_{i=1}^{n}M_i^2Var_2\left(\frac{I_{i1}}{h_i}\right)\frac{H}{H-1}\sum_{r=1}^{H}(\mu_{i[r]}-\mu_i)^2\right]\\&=\frac{1}{n\bar{M}^2}E_1\left\{\frac{1}{n}\sum_{i=1}^{n}M_i^2E_2\left(\frac{J_{i1}I_{i1}}{h_i^2}\right)\sum_{r=1}^{H}\sigma_{i[r]}^2\right\}+D^2\end{aligned}\tag{۸}$$

با جای‌گذاری روابط (۷) و (۸) در رابطهٔ (۶) نتیجه حاصل می‌شود.

نمونه‌گیری دومرحله‌ای با به کارگیری روش پساطبقه‌بندی قضاوتی تعمیم‌یافته در مرحلهٔ دوم

در این بخش می‌خواهیم با استفاده از براودگرهای تعمیم‌یافته که به‌وسیلهٔ وانگ و همکاران [۲۵] و محمدقاسمی و همکاران [۲۶] برای نمونه‌گیری پساطبقه‌بندی قضاوتی ارائه شده‌اند، دو براوردگر تعمیم‌یافته برای براوردگر میانگین جامعه با استفاده از نمونه‌گیری دومرحله‌ای با به کارگیری روش پساطبقه‌بندی قضاوتی تعمیم‌یافته در مرحلهٔ دوم ارائه دهیم. ابتدا براوردگر وانگ و همکاران [۲۵] را بیان می‌کنیم. برای این منظور میانگین مشاهده‌های دارای رتبهٔ r را با $\mu_{[r]}$ نمایش می‌دهیم، وانگ و همکاران [۲۵] با استفاده از لم زیر نشان دادند $\mu_{[1]}\le\mu_{[2]}\le\ldots\le\mu_{[H]}$ است.

لم ۲. فرض کنید $Y_{[1]},\ldots,Y_{[H]}$ به‌طور تصادفی مرتب‌شده باشد. در این صورت برای هر y داریم

$$F_{[1]}(y)\ge\ldots\ge F_{[H]}(y)$$

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

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

توجه کنید که اگر دو توزیع به‌طور تصادفی مرتب‌شده باشند (برای مثال $F(x)\ge G(x)$)، آن‌گاه برای هر تابع φ غیر نزولی داریم

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

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

$$\mu_{[1]}\le\mu_{[2]}\le\ldots\le\mu_{[H]}$$

سپس با جای‌گزین کردن شکل هم‌نوا شده میانگین‌های نمونه $\bar{Y}_{[r]}$ در رابطهٔ $\hat{\mu}_{JPS}=\frac{1}{h}\sum_{r=1}^{H}\bar{Y}_{[r]}I_r$ براوردگر میانگین هم‌توان را در روش نمونه‌گیری پساطبقه‌بندی قضاوتی بدین‌صورت به‌دست آوردند.

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

به‌طوری که به‌ازای $n_r>0$، $\bar{Y}_{[r]}^{iso}=\max_{t\le r}\min_{s\ge r}\sum_{g=t}^{s}\frac{n_g\bar{Y}_{[g]}}{n_{rs}}$ و $n_{rs}=\sum_{g=t}^{s}n_g$ است. از این‌رو، ما از این روش در نمونه‌گیری دومرحله‌ای استفاده کرده و براوردگری برای میانگین جامعه ارائه می‌دهیم. برای این منظور در نمونه‌گیری

خوشه‌ای دومرحله‌ای، فرض کنید با انتخاب n خوشه از N خوشه در مرحلهٔ اول نمونه‌گیری و سپس m_i واحد از M_i واحد خوشهٔ iام در مرحلهٔ دوم نمونه‌گیری، $\hat{\mu}_{w_i}$ براوردگر پیشنهادی وانگ و همکاران [۲۵] در خوشهٔ iام انتخابی باشد. در این صورت با توجه به این‌که $\mu = \frac{1}{N\bar{M}}\sum_{i=1}^{N} M_i \mu_i$، براوردگری جدید برای میانگین جامعه در روش نمونه‌گیری دومرحله‌ای بدین‌صورت ارائه می‌دهیم:

$$\hat{\mu}_{new_1} = \frac{1}{n}\sum_{i=1}^{n}\frac{M_i}{\bar{M}}\hat{\mu}_{w_i}$$

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

محمدقاسمی و همکاران [۲۶] با توجه به ایده‌ای که بیان شد به‌جای $\left\{(Y_j, R_j), j=1,...,m\right\}$ از نمونهٔ مرتب‌شدهٔ آن استفاده کردند. به این ترتیب که $\left\{\left(Y_{(j)}, R_{[j]}\right), j=1,\ldots,m\right\}$ را نمونهٔ مرتب‌شدهٔ طرح *JPS* در نظر گرفتند که در آن $Y_{(1)} \le Y_{(2)} \le \ldots \le Y_{(m)}$ آماره‌های مرتب‌شدهٔ $Y_1, Y_2, \ldots, Y_m$ هستند و $R_{[j]}$ رتبهٔ متناظر با $Y_{(j)}$ است. در این حالت ممکن است رابطهٔ $R_{[1]} \le R_{[2]} \le \ldots \le R_{[m]}$ برقرار نباشد. ایدهٔ معرفی براوردگر جدید به‌وسیلهٔ محمدقاسمی و همکاران [۲۶] این است که از تصادفی قرار گرفتن هر یک از مشاهدات درون طبقات چشم‌پوشی کرده و مشاهدات را به‌صورت $\left\{\left(Y_{(j)}, R_{(j)}\right), j=1,\ldots,m\right\}$ در نظر بگیریم که در آن $Y_{(1)} \le Y_{(2)} \le \ldots \le Y_{(m)}$ و $R_{(1)} \le R_{(2)} \le \ldots \le R_{(m)}$ باشد. در نتیجه براوردگر پیشنهادی محمدقاسمی و همکاران [۲۶] بدین صورت است:

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

که در آن

$$\overline{Y}^{*}_{(r)} = \frac{1}{n_r}\sum_{j=1}^{m} Y_{(j)} I_{rj}, \quad I_{rj} = \begin{cases} 1 & R_{(j)} = r \\ 0 & \text{درغیر این‌صورت} . \end{cases}, \quad n_r = \sum_{j=1}^{m} I_{rj}$$

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

$$\hat{\mu}_{new_2} = \frac{1}{n}\sum_{i=1}^{n}\frac{M_i}{\bar{M}}\hat{\mu}_{g_i} .$$

شبیه‌سازی

در این بخش با استفاده از دو مطالعهٔ شبیه‌سازی شده، کارایی روش $TSCJPS$ نسبت به نمونه‌گیری دومرحله‌ای تصادفی با به‌کارگیری نمونه‌گیری تصادفی سادهٔ بدون جای‌گذاری در هر دو مرحله بررسی می‌شود. برای مقایسهٔ کارایی براوردگرها به‌دلیل دشوار بودن محاسبه میانگین توان دوم خطا و اریبی براوردگرهای جدید از شبیه‌سازی مونت‌کارلو استفاده می‌کنیم. در پژوهش اول از داه‌های 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 به‌وسیلهٔ مرکز آمار ایران [۲۸] جمع‌اوری شده‌اند، استفاده می‌شود. در این بررسی واحدهای نمونه‌گیری مرحلهٔ اول یک استان یا در مواردی چند استان کوچک است که با هم ادغام شده‌اند و واحدهای نمونه‌گیری مرحلهٔ دوم گل‌خانه‌ها هستند.

فرض کنید بخواهیم میانگین ارزش محصولات گل‌خانه‌های کشور را براورد کنیم. برای این منظور براوردگرهای پیشنهادی و براوردگر میانگین در نمونه‌گیری خوشه‌ای دومرحله‌ای ساده را استفاده کرده و کارایی براوردگرهای حاصل یعنی، $\hat{\mu}_{TSCJPS}$، $\hat{\mu}_{new_2}$، $\hat{\mu}_{new_1}$، و $\hat{\mu}_{TSCSRS}$ را با یک‌دیگر مقایسه می‌کنیم. در محاسبهٔ براوردها 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 رابطهٔ زیر به‌دست می‌آید استفاده می‌شود (دل و کلاتر [۳۰]).

$$X=\rho\left(\frac{Y-\mu}{\sigma}\right)+\sqrt{1-\rho^2}Z \qquad (۹)$$

که در آن μ و σ میانگین و انحراف معیار متغیر Y (در خوشهٔ 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 به‌دست می‌آید و رتبه‌بندی براساس مقادیر X انجام می‌شود.

در داده‌های گل‌خانه‌ای $N=25$ واحد نمونه‌گیری مرحلهٔ اول PSU وجود دارد. تعداد واحدهای نمونه‌گیری مرحلهٔ دوم در هر یک از PSUها در جدول ۱ نشان داده‌شده است. در مرحله اول نمونه‌گیری n واحد (استان) به روش نمونه‌گیری تصادفی سادهٔ بدون جای‌گذاری انتخاب می‌شود. سپس در هر یک از PSUهای انتخاب‌شده m_i واحد نمونه‌گیری مرحلهٔ دوم (گل‌خانه‌ها) با چهار روش پساطبقه‌بندی قضاوتی، پساطبقه‌بندی قضاوتی تعمیم‌یافته و نمونه‌گیری تصادفی سادهٔ 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 براورد میانگین جامعه، در هر تکرار نمونه‌گیری میانگین نمونه‌های حاصل از سه روش محاسبه می‌شود. انتخاب نمونه‌ها و محاسبهٔ براودگرها با استفاده از نرم‌افزار R انجام می‌گیرد. مقدار دقیق میانگین جامعه که از چارچوب قابل محاسبه است

بنابراین می‌توان براوردی از میانگین توان دوم خطای هر یک از براوردگرهای $\hat{\mu}_k$ برابر با $\mu=40/5$ است. یعنی $M\hat{S}E\left(\hat{\mu}\right)$ را بدین‌صورت به‌دست آورد:

$$M\hat{S}E\left(\hat{\mu}_k\right)=\frac{1}{50000}\sum_{d=1}^{50000}\left(\hat{\mu}_{d_k}-\mu\right)^2, \quad k=1,2,3,4$$

که در آن $\hat{\mu}_{d_1}$، $\hat{\mu}_{d_2}$، $\hat{\mu}_{d_3}$ و $\hat{\mu}_{d_4}$ به‌ترتیب براوردگرهای $\hat{\mu}_{TSCJPS}$، $\hat{\mu}_{new_1}$، $\hat{\mu}_{new_2}$ و $\hat{\mu}_{TSCSRS}$ در dامین تکرار نمونه‌گیری هستند. برای مقایسهٔ کارایی هر دو براوردگر از کارایی نسبی براوردگرهای حاصل استفاده می‌کنیم، کارایی نسبی $\hat{\mu}_1$ نسبت به $\hat{\mu}_2$ بدین‌صورت تعریف می‌شود:

$$RE\left(\hat{\mu}_1,\hat{\mu}_2\right)=\frac{M\hat{S}E\left(\hat{\mu}_2\right)}{M\hat{S}E\left(\hat{\mu}_1\right)}$$

جدول ۱. تعداد واحدهای نمونه‌گیری مرحلهٔ 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 براوردگرهای $\hat{\mu}_{TSCJPS}$، $\hat{\mu}_{new_1}$، $\hat{\mu}_{new_2}$ نسبت به براوردگر $\hat{\mu}_{TSCSRS}$ را به‌ترتیب با RE_{TSCJPS}، RE_1 و RE_2 نمایش می‌دهیم که برای مقادیر $H=2,4,6,8,10$، $n=3,6$ و $m_i=3,10$ در جدول‌های ۲ و ۳ ارائه دادیم.

جدول‌های ۲ و ۳ نشان می‌دهند که براورد حاصل از روش $TSCJPS$ تقریباً کارایی برابری با براوردگر حاصل از نمونه‌گیری دومرحله‌ای تصادفی سادهٔ بدون جای‌گذاری در هر مرحلهٔ نمونه‌گیری دارد. وقتی که از روش پساطبقه‌بندی قضاوتی تعمیم‌یافته (روش وانگ و همکاران [۲۵] و روش محمدقاسمی و همکاران [۲۶]) در مرحلهٔ دوم نمونه‌گیری دو مرحله‌ای استفاده می‌شود، براوردگرهای حاصل از این روش در بیش‌تر موارد کارایی بیش‌تری نسبت به براورد حاصل از نمونه‌گیری دومرحله‌ای تصادفی سادهٔ بدون جای‌گذاری در هر مرحلهٔ نمونه‌گیری دارند، هم‌چنین نسبت به براوردگر 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 براوردگر $\hat{\mu}_{TSCJPS}$ نسبت به براوردگر $\hat{\mu}_{TSCSRS}$ تقریباً ثابت، اما دقت براوردگرهای $\hat{\mu}_{new_1}$ و $\hat{\mu}_{new_2}$ نسبت به $\hat{\mu}_{TSCSRS}$ کاهش می‌یابد، هم‌چنین وقتی که رتبه‌بندی قوی است با افزایش H برای N، n و m_i ثابت، کارایی براوردگر $\hat{\mu}_{TSCJPS}$ نسبت به براوردگر $\hat{\mu}_{TSCSRS}$ تقریباً ثابت است، ولی بهبود کارایی براوردگرهای $\hat{\mu}_{new_1}$ و $\hat{\mu}_{new_2}$ نسبت به $\hat{\mu}_{TSCSRS}$ همواره افزایشی است. براوردگرهای $\hat{\mu}_{new_1}$ و $\hat{\mu}_{new_2}$ وابسته به رتبه‌بندی هستند، هر چه رتبه‌بندی دقیق‌تر باشد

براوردگرهای $\hat{\mu}_{new_1}$ و $\hat{\mu}_{new_2}$ براوردگرهای کارایی هستند. زمانی که m_i کوچک $(m_i = 3)$ و رتبه‌بندی ضعیف باشد با افزایش H برای N و n ثابت براوردگرهای $\hat{\mu}_{new_1}$ و $\hat{\mu}_{new_2}$ براوردگرهای بهتری در مقایسه با براوردگرهای $\hat{\mu}_{TSCJPS}$ و $\hat{\mu}_{TSCSRS}$ 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 براوردگرها تقریباً کارایی برابری با براوردگر $\hat{\mu}_{TSCSRS}$ دارند.

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

		$m_i=3$			$m_i=10$		
H	ρ	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 براوردگرهای $\hat{\mu}_{TSCJPS}$، $\hat{\mu}_{new_1}$ و $\hat{\mu}_{new_2}$ نسبت به براوردگر $\hat{\mu}_{TSCSRS}$ برای $n=6$ و $N=25$

		$m_i=3$			$m_i=10$		
H	ρ	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 مقایسهٔ کارایی براوردگرها می‌پردازیم. در جدول‌های ۴ و ۵ مقایسه کارایی نسبی براساس این توزیع‌ها آورده شده است که خوشه‌بندی بدین‌صورت انجام می‌شود: برای توزیع نرمال در خوشهٔ اول ۴۹ داده از توزیع $N(25,100)$، برای خوشهٔ دوم ۴۸ داده از توزیع $N(24,100)$، برای خوشهٔ سوم ۴۷ داده از توزیع $N(23,100)$ و به‌همین ترتیب برای خوشهٔ بیست و پنجم، ۲۵ داده از توزیع $N(1,100)$ تولید می‌کنیم. هم‌چنین برای توزیع نمایی در خوشهٔ اول ۴۹ داده از توزیع $E(25)$، برای خوشهٔ دوم ۴۸ داده از توزیع $E(24)$، برای خوشهٔ سوم ۴۷ داده از توزیع $E(23)$ و به‌همین ترتیب برای خوشهٔ بیست و پنجم، ۲۵ داده از توزیع $E(1)$ تولید می‌کنیم. با توجه به این خوشه‌بندی و داده‌های تولید شده، همانند قبل

کارایی نسبی براوردگرهای $\hat{\mu}_{TSCJPS}$، $\hat{\mu}_{new_1}$ و $\hat{\mu}_{new_2}$ نسبت به براوردگر $\hat{\mu}_{TSCSRS}$ که به‌ترتیب با RE_{TSCJPS}، RE_1 و RE_2 نمایش داده شده برای مقادیر $H=2,4,6,8,10$،

جدول ۴. مقایسهٔ کارایی براوردگرهای $\hat{\mu}_{TSCJPS}$، $\hat{\mu}_{new_1}$ و $\hat{\mu}_{new_2}$ نسبت به براوردگر $\hat{\mu}_{TSCSRS}$ برای $\rho=0/99$، $n=3$ و $N=25$

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 براوردگر $\hat{\mu}_{TSCJPS}$ تقریباً برابر براوردگر $\hat{\mu}_{TSCSRS}$ است. در هر دو توزیع $\hat{\mu}_{new_1}$ و $\hat{\mu}_{new_2}$ نسبت به $\hat{\mu}_{TSCSRS}$ در بیش‌تر موارد براوردگرهای بهتری هستند، هم‌چنین در توزیع نرمال با مقایسهٔ مقادیر کارایی برای $m_i=3$ ($m_i=10$) 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. مقایسهٔ کارایی براوردگرهای $\hat{\mu}_{TSCJPS}$، $\hat{\mu}_{new_1}$ و $\hat{\mu}_{new_2}$ نسبت به براوردگر $\hat{\mu}_{TSCSRS}$ برای $\rho=0/99$، $n=6$ و $N=25$

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 براوردگرهای جدیدی برای میانگین جامعه به‌همراه واریانس این براوردگر در طرح نمونه‌گیری دومرحله‌ای با به‌کارگیری روش پساطبقه‌بندی قضاوتی در مرحلهٔ دوم و تعمیمی از این براوردگر ارائه شد. هم‌چنین نشان دادیم که براوردگرهای پیشنهادی، براوردگرهایی نااریب برای میانگین جامعه هستند. نتایج شبیه‌سازی نشان می‌دهد که این براوردگرها در بسیاری از موارد نسبت به براوردگر 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، "براوردگر جدید میانگین در طرح نمونه‌گیری طبقه‌بندی قضاوتی با مرتب کردن مشاهدات درون طبقات، مجله علوم آماری، جلد ۱۰، شمارهٔ 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, NewYork (1977).
30. Dell T. R., Clutter J. L., "Ranked Set Sampling Theory with Order Statistics Background", Biometrics, (1972) 545-555.