طرح نمونه‌گیری خوشه‌ای دومرحله‌ای در پژوهش‌های میدانی
در پژوهش‌های میدانی

کامران میرزایی [۱]، مریم پارساییان [۲]

تاریخ دریافت: ۱۴۰۱/۱۱/۱۰
تاریخ پذیرش: ۱۴۰۲/۰۲/۲۲

چکیده:
در مقاله حاضر سعی شده است راهنماییِ کاملاً عملیاتی و مفیدی در خصوص اجرای طرح نمونه‌گیری خوشه‌ای دومرحله‌ای در پژوهش‌های میدانی ارائه شود که با توجه به شرایط و امکانات فعلی موجود در کشور، از پرکاربردترین طرح‌های نمونه‌گیری است. به‌علاوه برای استفاده بیشتر محققان در این حوزه، در این مقاله کدهایی برای اجرای نمونه‌گیری خوشه‌ای دومرحله‌ای با حجم غیریکسان بر اساس احتمال متناسب با اندازه به همراه مثالی در مورد نمونه‌گیری خوشه‌ای دومرحله‌ای با احتمال متناسب با اندازه (PPS) با داده‌های فرضی با استفاده از نرم‌افزار R آورده شده است.
واژه‌های کلیدی: چارچوب نمونه‌گیری، احتمال متناسب با حجم نمونه، نمونه‌گیری تصادفی، نمونه‌گیری خوشه‌ای دومرحله‌ای، پژوهش میدانی

۱ مقدمه

طرح‌های پژوهش‌های آزمایشی در دو حالت کنترل‌شده (آزمایشگاه) و محیط طبیعی (میدانی) انجام می‌شوند. یکی از محدودیت‌های اصلی طرح‌های آزمایشگاهی این است که این احتمال وجود دارد که شرکت‌کنندگان با آگاهی از اینکه موردتحقیق قرار دارند رفتار متفاوتی را از خود نشان دهند که این مشکل را می‌توان با جمع‌آوری داده‌ها از محیط طبیعی علیرغم دشواری بیشتر کنترل متغیرها برطرف نمود.

مشاهده در پژوهش‌های میدانی نقش بسیار مهمی دارد و می‌تواند به دو صورت ساختارمند مانند پرسشنامه و بدون ساختار مانند مشاهده رفتارهای مختلف در محیط انجام شود. در مشاهده ساختارمند بر مبنای طرحی تعریف شده، رفتارها طبقه‌بندی و ثبت می‌شوند؛ بنابراین لازم است از سؤال‌های مشخص، بسته‌پاسخ و قابل کمی شدن استفاده شود. این در حالی است که در مشاهده بدون ساختار، رفتارها بدون چارچوب قبلی ثبت می‌شوند بدین معنی که مشاهدات با برداشت آزاد پژوهشگر جمع‌آوری و ثبت می‌شوند [۱].

امروزه در دنیا با ایجاد بانک‌های اطلاعاتی عظیم، امکان دستیابی به داده‌های بسیار مفیدی برای محققین فراهم شده است. در بسیاری از ارگان‌های کشور از سال‌ها قبل به‌منظور مکانیزه نمودن روند ثبت، ذخیره، تجزیه‌وتحلیل و نمایش داده‌ها و همچنین ارائه خدمات الکترونیک به ذینفعان و ... سامانه‌هایی را در بخش‌ها و زمینه‌های مختلف در بستر اینترنت ایجاد نموده‌اند که با استفاده از این سامانه‌ها، دستیابی به کلیه واحدهای اپیدمیولوژیک در جمعیت‌های مختلف به‌راحتی امکان‌پذیر شده است و این امر زمینه بسیار مناسبی را برای تعیین چارچوب‌های نمونه‌گیری برای محققین فراهم می‌نماید.

این مقاله باهدف کاربرد برای همه متخصصین حوزه‌های مختلف نگاشته شده است و شامل راهنمایی‌های عملی برای محققانی است که از این تکنیک نمونه‌گیری در مطالعات خود استفاده می‌کنند. به عبارت دقیق‌تر بعد از توضیح مختصری از نمونه و شرح روش‌های نمونه‌گیری بر اساس چگونگی انتخاب نمونه، طرح نمونه‌گیری خوشه‌ای دومرحله‌ای شرح داده می‌شود. در انتها مزایای این روش نمونه‌گیری و مسائل و چالش‌های رایجی که در نمونه‌گیری خوشه‌ای دومرحله‌ای روی می‌دهند، مانند عدم پاسخ، خطای اندازه‌گیری و اثرات خوشه‌بندی بحث و بررسی می‌شوند.

۱.۱ نمونه‌گیری و روش‌های نمونه‌گیری

ادبیات نمونه‌گیری پیمایشی چندین روش استفاده از اطلاعات کمکی خاص را برای دستیابی به روش‌های نمونه‌گیری و برآورد کارآمدتر

[۱] اپیدمیولوژیست، سازمان دامپزشکی کشور
[۲] دانشجوی دکتری سنجش و اندازه‌گیری، دانشکده روانشناسی و علوم تربیتی، دانشگاه تهران، تهران، ایران. (نویسنده مسئول: maryam.parsaeian@ut.ac.ir
)

توصیف می‌کند. اگر اطلاعات موردنیاز از قبل برای کل جمعیت در دسترس نباشد، ممکن است جمع‌آوری اقلام خاص برای یک نمونه بزرگ‌تر و اولیه و سپس استفاده از آن برای بهبود تخمین از یک نمونه کوچک‌تر و نهایی سودمند باشد. این روش عموماً نمونه‌گیری دومرحله‌ای نامیده می‌شود [۱۶].

به‌طورکلی استفاده از نمونه‌ای که نماینده واقعی از جامعه باشد به دلیل اینکه مستلزم هزینه کمتری است بسیار رایج است. چراکه در موقعیت‌های زندگی واقعی مواردی وجود دارد که فهرستی از عناصر جامعه در دسترس نیست و یا اگر موجود باشد به دلیل ماهیت توزیع جامعه مقرون‌به‌صرفه نیست لذا انتخاب نمونه مناسبی که نماینده‌ای از ویژگی‌های جامعه باشد حائز اهمیت است [۳]. در واقع هدف از استخراج نمونه مناسبی از جامعه دستیابی به استنباطی درباره جامعه است. دقت روش استنباط متأثر از دو عامل حجم نمونه و مقدار تغییرات در داده‌هاست [۴]؛ بنابراین لازم است فرآیند انتخاب برای تصمیم‌گیری در این خصوص که کدام اعضای جامعهٔ هدف و به چه صورت وارد گروه مطالعه شوند و همچنین فرآیند برآورد محاسبه آماره‌های نمونه به‌طور دقیق مشخص شوند. در واقع این دو موضوع ارتباط بسیار نزدیکی باهم دارند؛ به‌طوری‌که طریقه انتخاب افراد مشمول نمونه‌گیری بر روش برآوردها تأثیرگذار خواهد بود. از سوی دیگر برآورد یک پارامتر از نمونه همیشه متأثر از خطای تصادفی می‌باشد که به‌طورکلی قابل‌حذف نبوده و بهترین شیوه برخورد با آن، کنترل خطاهای اجتناب‌ناپذیر با به‌کارگیری «روش نمونه‌گیری درست» و «حجم نمونه مناسب» است [۵]. در ادامه به شرح مختصری از هر یک از این دو مورد پرداخته می‌شود.

روش‌های نمونه‌گیری بر اساس چگونگی انتخاب نمونه به دو دسته روش‌های احتمالاتی و روش‌های غیراحتمالاتی تقسیم می‌شوند. در نمونهٔ احتمالاتی، هر یک از عناصر جامعه دارای احتمالی معلوم و غیرصفر برای گنجانیده شدن در نمونه هستند بنابراین می‌توان برآوردهای نااریبی از پارامترهای جامعه را بر اساس داده‌های نمونه به دست آورد ولی نمونهٔ غیراحتمالاتی مبتنی بر برنامهٔ نمونه‌گیری است که این خصیصه را ندارد بدین معنی که بسیار محتمل است که چنین روش انتخاب نمونه‌ای به برآوردهایی با اریبی بسیار زیاد منجر شود. در واقع دلیل استفاده از نمونه‌های غیراحتمالاتی این است که نمونه‌گیری احتمالاتی غالباً شیوه‌ای وقت‌گیر و پرهزینه است که ممکن است در بسیاری از وضعیت‌ها امکان‌پذیر نباشد. از سوی دیگر به‌منظور اجرای نمونه‌گیری احتمالاتی لازم است فهرستی از اعضای جامعه موجود باشد تا بتوان نمونه‌ای را استخراج کرد درحالی‌که چنین الزامی برای روش‌های غیراحتمالاتی وجود ندارد [۶]. لذا در تحقیقات کاربردی که محقق نیازمند برآوردهای نزدیک به واقعیت از نمونه‌هاست تا بتواند در جهت تعیین ویژگی‌های واقعی جامعه قدم‌های مؤثری بردارد، می‌بایست نمونه‌گیری‌ها از نوع احتمالاتی باشند و به‌صورت تصادفی انجام شوند تا نتیجه حاصل‌شده هر چه بیشتر به واقعیت نزدیک شود. یکی از روش‌های احتمالاتی بسیار رایج، نمونه‌گیری خوشه‌ای است. انگیزه رایج نمونه‌گیری خوشه‌ای کاهش هزینه از طریق افزایش کارایی نمونه‌گیری است [۳]؛ بنابراین در روش‌های احتمالاتی نمونه‌گیری به چارچوب نمونه‌گیری نیاز است که شامل فهرست نهایی افراد موردمطالعه برای تحلیل است [۷].

از سوی دیگر انتخاب نمونه در روش‌های احتمالاتی به صورت‌های مختلف از جمله نمونه‌گیری تصادفی ساده و نمونه‌گیری متناسب با حجم نمونه انجام می‌شود. در نمونه‌گیری تصادفی ساده شانس انتخاب هر واحد برابر است؛ اما اغلب واحدها ممکن است از نظر اندازه متفاوت باشند. اگر در این مورد از نمونه‌گیری تصادفی ساده استفاده شود، ممکن است نتیجه مورد انتظار یا اطلاعات موردنظر به دست نیاید. در این شرایط، چنین اطلاعات فرعی را می‌توان در انتخاب نمونه به‌منظور برآورد دقیق‌تر پارامترهای جامعه مورداستفاده قرار داد. احتمال انتخاب نمونه‌ها برای واحدهای مختلف به‌اندازه آن‌ها بستگی دارد که به آن‌ها نمونه‌برداری با احتمال متناسب با اندازه گفته می‌شود که با علامت اختصاری PPS نشان داده می‌شوند [۸].

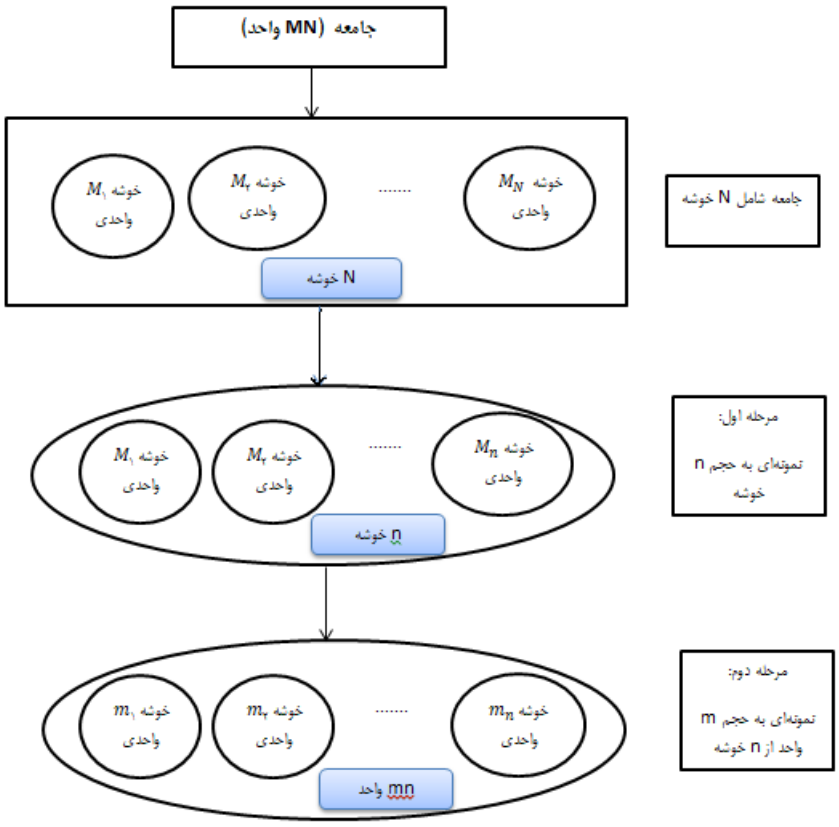
نمونه‌برداری PPS به عناصر «بزرگ‌تر» در لیست شانس بیشتری برای انتخاب نسبت به عناصر «کوچک‌تر» می‌دهد. به‌طور خاص، احتمال انتخاب یک عنصر با اندازه آن رابطه مستقیم دارد. اگر یک عنصر دو برابر بزرگ‌تر از عنصر دیگر باشد، شانس نمونه‌برداری دو برابر خواهد داشت [۹]. به‌منظور اطمینان از اینکه همه افراد موردمطالعه در جامعه صرف‌نظر از اندازه خوشه خود، احتمال انتخاب یکسانی دارند، هر یک از سطوح سلسله مراتبی قبل از سطح نهایی باید بر اساس اندازه واحدهای نهایی که شامل تعداد واحدهای یکسانی از هر خوشه در آخرین سطح سلسله مراتبی هستند، نمونه‌برداری شوند جمله کمی گنگ است. همچنین این روش، برنامه‌ریزی برای کار میدانی را تسهیل می‌کند زیرا در هر واحد انتخاب‌شده تعداد از پیش تعیین‌شده افراد در نمونه حضور دارند و زمانی که واحدهای نمونه‌برداری به‌طور قابل‌توجهی از نظر اندازه متفاوت هستند این روش بسیار مفید است، زیرا این اطمینان را می‌دهد که واحدهایی که در خوشه‌های بزرگ‌تر هستند همان احتمال ورود به نمونه را دارند که واحدهای نمونه در خوشه‌های کوچک‌تر هستند و بالعکس [۷].

۲ طرح نمونه‌گیری خوشه‌ای دومرحله‌ای

نمونه‌گیری خوشه‌ای در یک مرحله و یا تحت شرایط خاصی در بیش از یک مرحله انجام می‌شود. در نمونه‌گیری خوشه‌ای یک مرحله‌ای، نمونه‌ای از خوشه‌ها به‌طور تصادفی ساده انتخاب می‌شوند به‌طوری‌که همه عناصری که در داخل هر خوشه نمونه قرار دارند موردبررسی قرار می‌گیرند. ولی تحت دو شرط خاص به‌منظور افزایش کارایی لازم است نمونه‌گیری خوشه‌ای در دو مرحله انجام شود. اول اینکه خوشه‌های انتخاب‌شده بزرگ‌تر از آن هستند که همه عناصر داخل خوشه‌ها را بتوان در نمونه قرار داد و دوم اینکه عناصری که در داخل خوشه‌ها قرارگرفته‌اند نسبت به متغیرهای مورداندازه‌گیری بسیار همگن باشند که در این حالت در نمونه قرار دادن همه عناصر در خوشه نمونه باعث فزونگی بسیار زیادی می‌شود. در چنین شرایطی غالباً نمونه‌گیری از عناصر انتخاب‌شده درون خوشه نمونه بهتر از انتخاب همه عناصر است [۶]. به‌عبارت‌دیگر نمونه‌گیری خوشه‌ای دومرحله‌ای زمانی استفاده می‌شود که اندازه خوشه‌ها بزرگ باشد و مشاهده تمام واحدهای داخل آن‌ها دشوار یا از نظر هزینه گزاف باشد.

خوشه‌ها در نمونه‌گیری خوشه‌ای دومرحله‌ای ممکن است دارای حجم‌های یکسان یا غیریکسان باشند. بهترین روش برای انتخاب خوشه‌های با حجم برابر، نمونه‌گیری تصادفی ساده است و درصورتی‌که خوشه‌ها دارای حجم‌های غیریکسانی باشند بهترین راه، نمونه‌ای است که متناسب با حجم هر خوشه انتخاب شود [۲].

در نمونه‌گیری خوشه‌ای دومرحله‌ای، خوشه‌ها معمولاً به‌عنوان واحدهای نمونه‌گیری اولیه PSU و واحدهایی که در مرحله دوم به‌عنوان واحدهای نمونه‌گیری ثانویه انتخاب می‌شوند، SSU نامیده می‌شوند. در نمونه‌گیری تصادفی خوشه‌ای دومرحله‌ای، اصولاً می‌توان از هر نوع طرح نمونه‌گیری در دو مرحله استفاده کرد که منجر به ترکیب‌های متعددی می‌شود. به‌عنوان‌مثال یکی از ترکیب‌ها در دو مرحله (SI, SI) است که در آن هر دو واحد PSU و SSU با نمونه‌گیری تصادفی ساده انتخاب می‌شوند و ترکیب دیگر، (PPS, PPS) است که در آن هر دو واحد PSU و SSU با نمونه‌گیری با احتمال متناسب با اندازه خوشه‌ها[3] انتخاب‌شده‌اند. طرح کلی از نمونه‌برداری خوشه‌ای دومرحله‌ای با حجم غیر یکسان در شکل ۱ آورده شده است [۵].



شکل ۱. طرح شماتیکی از نمونه‌گیری خوشه‌ای دومرحله‌ای به شیوه (PPS, PPS)

[3] Probability Proportional to Size (PPS)

در نمونه‌گیری خوشه‌ای دومرحله‌ای، ابتدا همه واحدهای جامعه به N خوشه گروه‌بندی می‌شوند به‌طوری‌که هر خوشه دارای $\mathrm{M_i}$ عنصر است که $\mathrm{i}=1,2,\ldots,\mathrm{N}$ است (که می‌توان حجم خوشه‌ها را یکسان نیز در نظر گرفت یعنی $\mathrm{M_i}=\mathrm{M}$). سپس یک نمونه تصادفی n تایی از خوشه‌ها بیرون کشیده می‌شود. سپس از هر یک از n خوشه انتخاب‌شده، نمونه‌های تصادفی از عناصر خوشه به حجم $\mathrm{m_i}$ گرفته می‌شود [۳]. لذا در هر یک از مراحل نمونه‌گیری خوشه‌ای دومرحله‌ای، منبع تغییراتی وجود دارد که باید موردتوجه قرار گیرند. به‌عبارت‌دیگر در مرحله اول، تفاوت بین واحدهای نمونه‌گیری اولیه یا خوشه‌ها و در مرحله دوم تفاوت بین افراد موجود در داخل خوشه‌ها بایستی مدنظر قرار داده شوند. هر یک از این منبع تغییرات را می‌توان با استفاده از دو شاخص «اثر طرح[۴]» و «ضریب همبستگی درون خوشه‌ای[۵]» بررسی کرد [۶]. اثر طرح بر تعداد نمونه‌های لازم در یک طرح نمونه‌گیری پیچیده مانند نمونه‌گیری خوشه‌ای دلالت دارد تا دقت این طرح برابر با طرح نمونه‌گیری تصادفی ساده شود. به‌عنوان‌مثال اگر اثر طرح برابر با ۲ باشد، این بدین معنی است که در یک طرح پیچیده دو برابر حجم نمونه نسبت به طرح نمونه‌گیری تصادفی ساده نیاز است تا به دقتی برابر با طرح نمونه‌گیری تصادفی ساده به دست آید. به‌عبارت‌دیگر با فرض یک حجم نمونه ثابت اثر طرح عبارت از نسبت بین تغییرات مشاهده‌شده در یک طرح پیچیده و تغییراتی که در صورت به‌کارگیری یک طرح نمونه‌گیری تصادفی ساده[۶] مورد انتظار می‌باشد [۷]؛ بنابراین درصورتی‌که نمونه‌گیری خوشه‌ای به‌عنوان طرح نمونه‌گیری پیچیده در نظر گرفته شود اثر طرح از رابطه (۱) محاسبه می‌شود:

$$\mathrm{D}=\frac{\mathrm{S^2_{cluster}}}{\mathrm{S^2_{str}}} \qquad (۱)$$

که در آن منظور از $\mathrm{S^2_{cluster}}$، واریانس مربوط به نمونه‌گیری خوشه‌ای است و $\mathrm{S^2_{str}}$ واریانس مربوط به نمونه‌گیری تصادفی ساده است.

ضریب همبستگی درون خوشه‌ای نشان‌دهنده میزان شباهت بین افراد در یک خوشه است و برای تنظیم محاسبه اندازه نمونه برای در نظرگرفتن اثر خوشه‌بندی استفاده می‌شود. این شاخص ارتباط داده‌های خوشه‌ای را با مقایسه واریانس بین خوشه‌ها $(\mathrm{S^2_b})$ و واریانس داخل خوشه‌ها $(\mathrm{S^2_w})$ با استفاده از فرمول (۲) محاسبه می‌نماید.

$$\rho=\frac{\mathrm{S^2_b}}{\mathrm{S^2_b}+\mathrm{S^2_w}} \qquad (۲)$$

با توجه به فرمول (۲) مشخص است که هرقدر واریانس داخل خوشه‌ای بیشتر باشد (تفاوت‌های افراد در داخل خوشه‌ها بیشتر باشد)، ضریب همبستگی داخل خوشه‌ای کمتر خواهد بود. مقدار ضریب همبستگی درون خوشه‌ای همواره در بازه $[0,1]$ قرار می‌گیرد. وقتی این ضریب برابر با ۱ باشد بدین معنی است که تمامی افراد داخل خوشه کاملاً مشابه هم و یکسان هستند. ولی اگر متغیر موردمطالعه ما به‌صورت کاملاً تصادفی بین خوشه‌ها توزیع‌شده باشد، ضریب همبستگی درون خوشه‌ای برابر با صفر خواهد بود. به‌طورکلی خیلی به‌ندرت پیش می‌آید که مقدار ضریب همبستگی داخل خوشه‌ای نزدیک ۱ باشد و معمولاً مقادیر کمتر یا مساوی ۰٫۲، بین ۰٫۲ تا ۰٫۴ و بیشتر یا مساوی ۰٫۵ به ترتیب به‌عنوان درجه یکنواختی کم، متوسط و زیاد مطرح می‌باشند [۵].

بین ضریب همبستگی درون خوشه‌ای و اثر طرح رابطه‌ای وجود دارد. به‌عبارت‌دیگر با داشتن هریک از این دو می‌توان دیگری را به دست آورد که در رابطه (۳) آورده شده است:

$$D=1+(m-1)\,\rho \qquad (۳)$$

که در آن m بیانگر میانگین تعداد نمونه جمع‌آوری‌شده در هر واحد است. طبق رابطه (۳) نتیجه گرفته می‌شود درصورتی‌که $\mathrm{D}=1$ باشد طرح خوشه‌ای با طرح نمونه‌گیری تصادفی ساده مشابه خواهد بود و اعمال هیچ تعدیلی هم برای حجم نمونه نیاز نیست.

شاخص ضریب همبستگی درون خوشه‌ای بر حجم نمونه داخل خوشه‌ای تأثیر دارد و به‌طورکلی، هرچه مقدار این ضریب به یک نزدیک‌تر باشد، حجم نمونه داخل خوشه‌ای کاهش می‌یابد (چراکه قاعدتاً در این حالت آزمایش تعداد کمی از افراد برای به دست آوردن اطلاعات درست کافی است)، اما تعداد خوشه‌ها یا همان واحدهای نمونه‌گیری اولیه (PSU ها) به دلیل افزایش تفاوت‌های کلی بین خوشه‌ای افزایش پیدا خواهد کرد [۵].

۳ تعیین حجم نمونه

هر مشاهده‌ای که از جامعه گرفته می‌شود شامل مقدار معینی اطلاعات در مورد پارامترهای جامعه موردنظر است. مسلماً کسب اطلاعات در مورد هر مشاهده مستلزم هزینه‌ای است بنابراین محقق باید تعیین کند که چه مقدار اطلاعات جمع‌آوری کند. بدیهی است در صورت جمع‌آوری اطلاعات ناقص، برآورد دقیقی از پارامترهای جامعه به دست نمی‌آید و از طرف دیگر جمع‌آوری بیش‌ازحد اطلاعات منجر به‌صرف هزینه بیشتر می‌شود. مقدار اطلاعات به‌دست‌آمده از نمونه به حجم نمونه و واریانس

[4] Design Effect
[5] Intra-cluster correlation Coefficient
[6] Simple Random Sampling (SRS)

داده‌ها بستگی دارد [۴].

در طراحی شیوه نمونه‌برداری و محاسبه حجم نمونه مؤثر بایستی نسبت مورد انتظار، محدوده خطای قابل‌قبول یا همان دقت و سطح اطمینان معلوم باشند. با توجه به اینکه در کتب مختلف در خصوص این اجزاء به‌تفصیل بحث شده است، از توضیح در خصوص این موارد در این مقاله خودداری می‌گردد.

حجم نمونه به عواملی از جمله دقت مدنظر محقق، نسبت مورد انتظار پدیده موردمطالعه و سطح اطمینان موردنیاز بستگی دارد. برای تعیین حجم نمونه بایستی یک تعادلی بین دقت مطالعه و هزینه‌های قابل انجام در طرح برقرار گردد [۱۱].

اندازه جامعه اغلب تأثیر کم و یا ناچیزی بر حجم نمونه دارد و در بیشتر موارد می‌توان آن را نادیده گرفت. ولي به‌طورکلی زمانی که حجم جامعه نسبت به حجم نمونه کوچک باشد، حجم نمونه کاهش می‌یابد. افزایش جمعیت بزرگ (مثلاً از ۱۰۰۰۰ تا ۱۰۰۰۰۰۰۰۰) عموماً تأثیری بر حجم نمونه ندارد. مي‌توان گفت واریانس در تعیین حجم نمونه بسیار مهم‌تر است [۱۰].

در نمونه‌گیری خوشه‌ای دومرحله‌ای، حجم نمونه با در نظر گرفتن عوامل متعددی از جمله سطح دقت مطلوب، واریانس خوشه‌ها، تعداد خوشه‌ها و ضریب همبستگی درون خوشه‌ای تعیین می‌شود. حجم نمونه یکسان در یک طرح پیچیده مانند نمونه‌گیری خوشه‌ای دومرحله‌ای در یك فرآیند پنج مرحله‌اي تعیین می‌شود. به‌منظور درک بهتر از این مبحث، فرض کنید می‌خواهیم طی یک نمونه‌گیری خوشه‌ای دومرحله‌ای، اندازه نمونه حیوانات را تعیین کنیم. بدین منظور لازم است ابتدا اندازه نمونه حیوانات در یک خوشه محاسبه شود (n) و ثانیاً تعداد خوشه‌هایی که باید نمونه‌برداری شوند (m) محاسبه شود [۱۴]. تعداد حیواناتی که باید در یک خوشه (گله) به‌صورت تصادفی نمونه‌برداری شوند را می‌توان به طبق رابطه (۴) تعیین کرد:

$$n = \frac{z_{\alpha/2}\, p\,(1-p)}{e^2} \qquad (۴)$$

که در آن $z_{\alpha/2}$، p و e به ترتیب بیانگر سطح دقت موردنظر، احتمال متناسب با حجم خوشه‌ها و سطح تغییرپذیری در خوشه‌ها هستند.

مرحله دوم نمونه‌گیری خوشه‌ای دومرحله‌ای شامل انتخاب تصادفی تعداد خوشه‌ها از جامعه است. در این مرحله تخمین تعداد خوشه‌های موردنیاز بر اساس سطح اطمینان (α) (که معمولاً ۹۵% در نظر گرفته می‌شود)، احتمال متناسب با حجم خوشه‌ها (p)، دقت مطالعه (e) و حساسیت[۷] (Se) و ویژگی[۸] (Sp) آزمون موردمطالعه بر اساس رابطه (۵) تعیین می‌شود [۱۰].

$$m = \frac{log\,(1-\alpha)}{log\,[Sp\,(1-p) + (1-Se)\,p]} \qquad (۵)$$

در مرحله سوم، اثر طرح (D) با استفاده از رابطه (۳) برآورد می‌شود.

در مرحله چهارم، حجم نمونه محاسبه‌شده در مرحله اول تعدیل می‌شود تا تعداد واحدهای نمونه‌برداری ثانویه (افراد) موردنیاز در یک طرح نمونه‌گیری خوشه‌ای دومرحله‌ای حاصل شود بدین منظور لازم است مقدار n در D ضرب شود که حاصل با $n_{adjusted}$ نمایش داده می‌شود.

در نهایت برای محاسبه تعداد خوشه‌هایی که بایستی مورد نمونه‌برداری قرار بگیرد کافی است عدد حاصل‌شده در مرحله چهارم بر مقدار m تقسیم شود. عدد به‌دست‌آمده بیانگر تعداد واحدهای نمونه‌برداری اولیه (خوشه‌های) موردنیاز برای برآورد احتمال پدیده موردمطالعه است که با استفاده از یک طرح نمونه‌گیری خوشه‌ای دومرحله‌ای به‌دست‌آمده است [۵].

۴ اجرای خوشه‌بندی دومرحله‌ای با حجم غیریکسان در R

به‌منظور طراحی نمونه‌گیری خوشه‌ای دومرحله‌ای لازم است متغیرهایی برای شناسایی واحدهای نمونه‌گیری اولیه و ثانویه و همچنین تعداد کل خوشه‌ها در جامعه (N) و تعداد کل عناصر در هر خوشه انتخابی (M_i) تعریف شوند.

به‌منظور استخراج نمونه خوشه‌ای دومرحله‌ای از داده $Voorst$ استفاده‌شده است. $Voorst$ در واقع جامعه‌ای شامل بی‌نهایت نقاط است که منطقه موردمطالعه توسط سلول‌های یک شبکه خوب گسسته شده است و متغیر موردمطالعه، غلظت ماده آلی خاک (SOM) برای تمام مراکز سلول‌های شبکه شبیه‌سازی‌شده است. بیست‌وچهار مربع ۰٫۵ کیلومتری ساخته‌شده است که به‌عنوان PSU عمل می‌کنند. به دلیل مناطق ساخته‌شده، جاده‌ها و غیره، PSU ها در $Voorst$ اندازه نابرابر دارند، به‌عنوان‌مثال، تعداد SSU ها (نقاط) در PSU ها متفاوت است.

با استفاده از کد زیر می‌توان تعیین کرد که همه نقاط گسسته در کدام ناحیه‌ها قرار می‌گیرند. در اولین مرحله در کد R، خوشه متغیر به $grdVoorst$ اضافه می‌شود که نشان می‌دهد یک واحد به کدام خوشه تعلق دارد.

```
library(dplyr)
```

[7] sensitivity
[8] specificity

```
library(base)
cell\_size = 25
w = 500 \#width of zones
grdVoorst = grdVoorst \%$>$\%
mutate(zone\_s1 = s1 \%$>$\%
findInterval(min(s1)
+ 1:11 * w + 5.0 * cell\_size),
zone\_s2 = s2 \%$>$\%
findInterval(min(s2) + w + 5.0 *
cell\_size), psu = str\_c(zone\_s1,
zone\_s2, sep = "\_"))
```

از تابع زیر برای انتخاب یک نمونه تصادفی خوشه‌ای دومرحله‌ای از یک جامعه نامتناهی که توسط تعداد محدودی از نقاط گسسته شده استفاده می‌شود:

```
twostage = function(sframe, psu, n, m) \{
units = sample(nrow(sframe), size = n, replace =
TRUE)
mypsusample = sframe\$psu[units]
ssunits = NULL
for (i in seq\_len(length(mypsusample))) \{
ssunit = sample(
x = which(sframe\$psu == mypsusample[i]),
size = m, replace = TRUE)
ssunits = c(ssunits, ssunit)
\}
psudraw = rep(c(1:n), each = m)
mysample = data.frame(ssunits, sframe[ssunits, ],
psudraw)
mysample
\}
```

با توجه به کد فوق بدیهی است که PSU و SSU هر دو به‌صورت با جایگزینی انتخاب شدند. اگر یک مرکز سلول شبکه‌ای انتخاب شود، یک نقطه به‌طور کاملاً تصادفی از آن سلول شبکه انتخاب می‌شود. این کار با جابجایی مرکز سلول شبکه به یک نقطه تصادفی در سلول شبکه انتخاب‌شده با تابع $jitter$ انجام می‌شود. درصورتی‌که یک PSU بیش از یک بار انتخاب‌شده باشد، برآوردهای متعددی از میانگین آن در PSU استفاده می‌شود. در ادامه از تابع $twostage$ برای انتخاب چهار بار یک PSU با احتمال‌های متناسب با حجم نمونه استفاده می‌شود. مرحله دوم حجم نمونه برای همه PSU ها برابر ۱۰ باشد. این SSU ها با نمونه‌گیری تصادفی ساده انتخاب می‌شوند [۱۳].

```
n = 4
m = 10
cell\_size = 25
set.seed\eqref{GrindEQ__314_}
mysample = grdVoorst \%$>$\%
twostage(psu = "psu", n = n, m = m)
\%$>$\% mutate(s1 = s1 \%$>$\%
jitter(amount =cell\_size / 2),
s2 = s2 \%$>$\%
jitter(amount = cell\_size / 2))
```

۵ مثال

برای درک بهتر مطالب بیان‌شده در این مقاله، در اینجا مثالی با اعداد و ارقام فرضی (جدول شماره ۱) بیان می‌گردد. فرض کنید می‌خواهیم از بین ۳۰ خوشه موجود در جدول، تعداد ۱۰ خوشه را به روش نمونه‌برداری با احتمال متناسب با اندازه (PPS) انتخاب نماییم. ابتدا کلیه واحدهای اولیه موجود در جمعیت موردمطالعه (ستون الف) و ظرفیت‌های مربوطه یا همان تعداد افراد موجود در هر واحد (ستون ب) را وارد فهرست می‌نماییم و فراوانی تجمعی این ظرفیت‌ها را نیز محاسبه می‌نماییم (ستون ج). حال با تقسیم تعداد کل افراد جمعیت موردمطالعه، بر تعداد خوشه‌ای که می‌بایست نمونه‌گیری شود، عدد مربوط به‌تناوب نمونه‌برداری را به دست می‌آوریم (۲۰۰۰ = ۱۰/۲۰۰۰۰). در مرحله بعد بایستی یک عدد را به‌صورت تصادفی بین یک تا عدد به‌دست‌آمده انتخاب نماییم (برای مثال: ۹۰۵) و به تعداد خوشه‌ای که باید مورد انتخاب قرار گیرند، عدد مربوط به‌تناوب نمونه‌برداری را به آن اضافه نماییم (۹۰۵ + ۲۰۰۰ = ۲۹۰۵ − ۲۹۰۵ + ۲۰۰۰ = ۴۹۰۵ و ۴۹۰۵ + ۲۰۰۰ = ۶۹۰۵ و ...). سپس تک‌تک اعداد به‌دست‌آمده به ترتیب با اعداد موجود در ستون فراوانی تجمعی مقایسه می‌شوند و هرکدام در اولین ردیفی قرار می‌گیرند که فراوانی تجمعی مربوطه، مساوی و یا بزرگ‌تر از آن‌ها باشد. بدین شکل این ردیف‌ها به‌عنوان

خوشه‌هایی که باید نمونه‌گیری شوند انتخاب و مشخص خواهند شد (ستون د).

به‌منظور محاسبه احتمال انتخاب هر خوشه (ستون ه) لازم است تعداد خوشه مورد نمونه‌گیری در حاصل تقسیم ظرفیت هر خوشه بر حجم جامعه موردمطالعه ضرب شود.

به‌علاوه برای محاسبه احتمال انتخاب هر فرد داخل هر خوشه (ستون ز) کافی است تعداد افراد مورد نمونه‌گیری در هر خوشه بر حجم خوشه مزبور تقسیم شود که مقادیر حاصل تا دو رقم اعشار در جدول ۱ گرد شده‌اند.

وزن پایه کلی[9] انتخاب هر فرد در جمعیت موردمطالعه (ستون ح) نیز عکس ضرب دو احتمال فوق‌الذکر خواهد بود؛ به عبارت ساده‌تر وزن پایه کلی، عکس احتمال انتخاب هر فرد در جمعیت است که همان‌طور که مشاهده می‌گردد برای هریک از افراد واردشده در مطالعه یکسان و مشابه هم می‌باشند که در جدول ۱ تا دو رقم اعشار گرد شده‌اند.

بر اساس نتایج نشان داده‌شده در جدول ۱ مشاهده می‌شود که تنها تعداد ۱۰ خوشه از بین ۳۰ خوشه مورد انتخاب قرارگرفته است که محاسبات آن به‌صورت دستی کار راحتی است؛ اما تصور کنید برای مثال قرار باشد از بین ۲۵۰۰۰ واحد پرورش مرغ گوشتی موجود در کشور، تعداد ۶۰۰ واحد را به شکل بالا انتخاب و وارد مطالعه نمایید. بدیهی است انجام محاسبات فوق و مقایسه تک‌تک فراوانی‌های تجمعی با اعداد به‌دست‌آمده به‌عنوان تناوب نمونه‌برداری، کاری بسیار مشکل و یا غیرقابل انجام خواهد بود. برای حل این مشکل نیز نویسندگان این مقاله با استفاده از نرم‌افزار R کد ذیل را تهیه نموده‌اند که جهت استفاده علاقه‌مندان در اینجا آورده شده است. در واقع این برنامه در محیط R تمامی اقدامات فوق‌الذکر برای انتخاب خوشه‌های مورد نمونه‌گیری را به‌صورت خودکار انجام می‌دهد و لیست خوشه‌های انتخاب‌شده را در قالب یک فایل اکسل ارائه می‌نماید.

```
rm(list=ls())
RandomUnits $<$- read.csv2(file = "12- Random Units
- R Imput.csv" ,sep=";")
dimRandomUnits $<$- dim(RandomUnits)
rowRandomUnits $<$- dimRandomUnits [1]
rowRandomUnits
columnRandomUnits $<$- dimRandomUnits[2]
columnRandomUnits
AllPoultryUnits $<$- read.csv2(file = "2
2- All Poultry Units - R Imput.csv" ,sep=";")
dimAllPoultryUnits $<$- dim(AllPoultryUnits)
rowAllPoultryUnits $<$- dimAllPoultryUnits [1]
rowAllPoultryUnits
columnAllPoultryUnits $<$- dimAllPoultryUnits[2]
columnAllPoultryUnits
i $<$- 1
j $<$- 1
k $<$- 1
class(RandomUnits\$Sample.No)
AllPoultryUnits\$Selected $<$- 0
for (i in seq(1, rowRandomUnits, by=1)) \{
for (j in seq(k, rowAllPoultryUnits, by=1)) \{
if (RandomUnits\$Sample.No[i] $<$
AllPoultryUnits\$Cumulative[k])\{
AllPoultryUnits\$Selected[k] $<$-
RandomUnits\$Sample.No[i]
k $<$- k+1
print(k)
break
\}
k $<$- k+1
print(k)
\}
\}
write.csv(AllPoultryUnits,"MozhiBariMyLoves.csv",
row.names=FALSE,fileEncoding = "UTF- 8" )
```

[9] Overall Basic Weight

جدول ۱. مراحل انجام طرح نمونه‌گیری با احتمال متناسب با اندازه خوشه‌ها

الف- شماره خوشه	ب- حجم خوشه	ج- فراوانی تجمعی	د- تعداد خوشه‌های موردمطالعه	ه- احتمال انتخاب خوشه	و- تعداد نمونه‌گیری در هر خوشه	ز- احتمال انتخاب هر فرد در هر خوشه	ح- وزن پایه کلی نمونه‌گیری
۱	۱۰۲۸	۱۰۲۸	۹۰۵	۰/۵۱	۳۰۰	۰/۲۹	۶/۶۷
۲	۵۵۵	۱۵۸۳		۰/۲۸	۳۰۰	۰/۵۴	۶/۶۷
۳	۳۹۰	۱۹۷۳		۰/۲۰	۳۰۰	۰/۷۷	۶/۶۷
۴	۱۳۰۹	۳۲۸۲	۲۹۰۵	۰/۶۵	۳۰۰	۰/۲۳	۶/۶۷
۵	۶۹۸	۳۹۸۰		۰/۳۵	۳۰۰	۰/۴۳	۶/۶۷
۶	۹۰۷	۴۸۸۷		۰/۴۵	۳۰۰	۰/۳۳	۶/۶۷
۷	۴۳۲	۵۳۱۹	۴۹۰۵	۰/۲۲	۳۰۰	۰/۶۹	۶/۶۷
۸	۸۹۷	۶۲۱۶		۰/۴۵	۳۰۰	۰/۳۳	۶/۶۷
۹	۶۷۷	۶۸۹۳		۰/۳۴	۳۰۰	۰/۴۴	۶/۶۷
۱۰	۵۰۱	۷۳۹۴	۶۹۰۵	۰/۲۵	۳۰۰	۰/۶۰	۶/۶۷
۱۱	۸۶۷	۸۲۶۱		۰/۴۳	۳۰۰	۰/۳۵	۶/۶۷
۱۲	۸۶۷	۹۱۲۸	۸۹۰۵	۰/۴۳	۳۰۰	۰/۳۵	۶/۶۷
۱۳	۱۰۰۲	۱۰۱۳۰		۰/۵۰	۳۰۰	۰/۳	۶/۶۷
۱۴	۱۰۹۴	۱۱۲۲۴	۱۰۹۰۵	۰/۵۵	۳۰۰	۰/۲۷	۶/۶۷
۱۵	۶۶۸	۱۱۸۹۲		۰/۳۳	۳۰۰	۰/۴۵	۶/۶۷
۱۶	۵۰۰	۱۲۳۹۲		۰/۲۵	۳۰۰	۰/۶	۶/۶۷
۱۷	۸۳۵	۱۳۲۲۷	۱۲۹۰۵	۰/۴۲	۳۰۰	۰/۳۶	۶/۶۷
۱۸	۳۹۶	۱۳۶۲۳		۰/۲۰	۳۰۰	۰/۷۶	۶/۶۷
۱۹	۶۳۰	۱۴۲۵۳		۰/۳۲	۳۰۰	۰/۴۸	۶/۶۷
۲۰	۴۸۳	۱۴۷۳۶		۰/۲۴	۳۰۰	۰/۶۲	۶/۶۷
۲۱	۳۱۹	۱۵۰۵۵	۱۴۹۰۵	۰/۱۶	۳۰۰	۰/۹۴	۶/۶۷
۲۲	۵۶۹	۱۵۶۲۴		۰/۲۸	۳۰۰	۰/۵۳	۶/۶۷
۲۳	۹۸۷	۱۶۶۱۱		۰/۴۹	۳۰۰	۰/۳۰	۶/۶۷
۲۴	۵۹۸	۱۷۲۰۹	۱۶۹۰۵	۰/۳۰	۳۰۰	۰/۵۰	۶/۶۷
۲۵	۳۷۵	۱۷۵۸۴		۰/۱۹	۳۰۰	۰/۸۰	۶/۶۷
۲۶	۳۸۷	۱۷۹۷۱		۰/۱۹	۳۰۰	۰/۷۸	۶/۶۷
۲۷	۴۶۵	۱۸۴۳۶		۰/۲۳	۳۰۰	۰/۶۵	۶/۶۷
۲۸	۷۵۱	۱۹۱۸۷	۱۸۹۰۵	۰/۳۸	۳۰۰	۰/۴۰	۶/۶۷
۲۹	۳۶۵	۱۹۵۵۲		۰/۱۸	۳۰۰	۰/۸۲	۶/۶۷
۳۰	۴۴۸	۲۰۰۰۰		۰/۲۲	۳۰۰	۰/۶۷	۶/۶۷

۶ بحث و نتیجه‌گیری

نمونه‌گیری خوشه‌ای دومرحله‌ای یک تکنیک پرکاربرد در تحقیقات پیمایشی است که انتخاب با PPS در نمونه‌های خوشه‌ای دومرحله‌ای (چندمرحله‌ای) رایج است [۹] و شامل گروه‌بندی افراد یا واحدهای مناسب در خوشه‌ها و سپس انتخاب تصادفی نمونه‌ای از خوشه‌ها برای جمع‌آوری داده‌ها است. در واقع نمونه‌گیری خوشه‌ای دومرحله‌ای روشی رایج برای انجام نمونه‌برداری است، به‌ویژه زمانی که جمعیت موردمطالعه زیاد و از نظر جغرافیایی پراکنده باشد. در این روش نمونه‌گیری، از مجموعه‌ای از واحدها مانند خانوار، مدرسه، بیمارستان یا گله در خصوص دام‌ها، نمونه‌گیری می‌شود و سپس مرحله دوم نمونه‌گیری در هر خوشه انتخابی انجام می‌شود تا واحدهای جداگانه برای مطالعه انتخاب شوند. بااینکه نمونه‌گیری خوشه‌ای دومرحله‌ای می‌تواند یک روش کارآمد و مقرون‌به‌صرفه برای جمع‌آوری داده‌ها باشد، مسائل و چالش‌های متعددی می‌تواند در طول فرآیند نمونه‌گیری ایجاد شود.

درصورتی‌که نمونه‌گیری به‌صورت پیمایشی (زمینه‌یابی) انجام شود بدین معنی که داده‌های جمع‌آوری‌شده حاصل از داده‌های پرسشنامه باشد که توسط آزمودنی‌ها تکمیل می‌شوند دو مسئله اساسی در تمام روش‌های نمونه‌گیری از جمله نمونه‌گیری خوشه‌ای دومرحله‌ای رخ می‌دهد که عبارت‌اند از ۱) عدم پاسخگویی که زمانی رخ می‌دهد که واحدها یا افراد منتخب از شرکت در مطالعه خودداری می‌کنند یا امکان تماس با آن‌ها وجود ندارد. عدم پاسخگویی می‌تواند سوگیری را به نمونه وارد کند، زیرا ویژگی‌های کسانی که پاسخ نمی‌دهند ممکن است با کسانی که پاسخ نمی‌دهند متفاوت باشد. برای کاهش این سوگیری، محققان می‌توانند از تکنیک‌های مختلفی مانند ارائه مشوق‌ها، استفاده از تماس‌های پیگیری یا نمونه‌برداری بیش‌ازحد در مرحله اول برای پاسخگویی به پاسخ‌های داده نشده استفاده کنند. ۲) خطای اندازه‌گیری که می‌تواند از عدم دقت در سؤالات نظرسنجی یا جمع‌آوری داده‌ها ناشی شود که منجر به نتایج مغرضانه یا نادرست شود. برای به حداقل رساندن خطای اندازه‌گیری، محققان می‌توانند از پرسش‌های نظرسنجی استاندارد و آموزش برای گردآورندگان داده استفاده کنند تا از سازگاری در جمع‌آوری داده‌ها در میان خوشه‌ها اطمینان حاصل کنند [۱۵].

از دیگر چالش‌های نمونه‌گیری خوشه‌ای دومرحله‌ای می‌توان به مسائل مربوط به انتخاب خوشه‌ها و نمونه‌گیری از افراد درون خوشه‌ها اشاره کرد. برای مثال، انتخاب خوشه‌ها می‌تواند بر نمایندگی نمونه تأثیر بگذارد و اگر تکنیک نمونه‌گیری تصادفی نباشد یا گروه‌های خاصی به‌طور سیستماتیک حذف شوند، نمونه‌گیری از افراد درون خوشه‌ها می‌تواند منجر به سوگیری شود [۱۶]. به‌عبارت‌دیگر یک مسئله بالقوه خطر سوگیری انتخاب است، زیرا خوشه‌ها و افراد درون خوشه‌ها ممکن است نماینده کل جامعه نباشند.

مزایای استفاده از نمونه‌گیری خوشه‌ای دومرحله‌ای را می‌توان به بدین‌صورت بیان کرد: در شرایطی که جمعیت از نظر جغرافیایی پراکنده است و انجام یک نمونه‌گیری تصادفی ساده را دشوار می‌کند می‌تواند بسیار مفید است. چراکه قادر است جمعیت‌های بزرگ و متنوع را مدیریت کند که باعث بهبود کارایی و کاهش خطاهای نمونه‌گیری می‌شود. در واقع با تقسیم جمعیت به خوشه‌ها، تعداد افراد موردبررسی کاهش می‌یابد که می‌تواند در زمان و منابع صرفه‌جویی کند. علاوه بر این، با انتخاب تصادفی خوشه‌ها و افراد در هر خوشه، نمونه نماینده جامعه است و خطر خطاهای نمونه‌گیری را کاهش می‌دهد. در نهایت، روش نمونه‌گیری خوشه‌ای دومرحله‌ای امکان شناسایی خوشه‌های درون جامعه را فراهم می‌کند که ممکن است بینش مفیدی را در مورد موضوع مطالعه ارائه دهد. به‌عنوان‌مثال، در یک مطالعه بهداشت عمومی، شناسایی خوشه‌هایی با نرخ بالای یک بیماری خاص می‌تواند به هدف قرار دادن مداخلات و منابع در مناطقی که بیشتر نیاز دارند کمک کند.

در نتیجه روش نمونه‌گیری خوشه‌ای دومرحله‌ای ابزار ارزشمندی در انجام پیمایش‌ها و مطالعات پژوهشی به‌ویژه در شرایطی است که جمعیت بزرگ، پراکنده و ناهمگن است. در واقع این روش به ارائه نتایج دقیق‌تر با هزینه و زمان کمتر در مقایسه با سایر تکنیک‌های نمونه‌گیری شناخته‌شده است. بدین ترتیب محققان با بررسی دقیق مسائل و چالش‌های بالقوه‌ای که می‌تواند در طول فرآیند نمونه‌گیری به وجود بیاید از روش نمونه‌گیری خوشه‌ای دومرحله‌ای استفاده می‌کنند تا از اعتبار و قابلیت اطمینان نتایج اطمینان حاصل کنند و به نتایج مطلوبی دست یابند.

مراجع

[۱] قاسمی، ح.، مرجع پژوهش. ۱۳۹۹، تهران: اندیشه آرا.

[۲] عمیدی، ع.، نظریه نمونه‌گیری و کاربردهای آن. ۱۳۹۴، تهران: مرکز نشر دانشگاهی.

[3] Ugwu, M.C. and M.S. Madukaife, M. S. (2022). Two-stage cluster sampling with unequal probability sampling in the first stage and ranked set sampling in the second stage. *Statistics in Transition new series (SiTns)*. **23(3)**, 199-214.

[4] Scheaffer, R. L. et al. (2011). *Elementary survey sampling*. Cengage Learning.

[5] Ferrari, G. et al. (2016). *Foot and mouth disease vaccination and post-vaccination monitoring*. FAO/OIE.

[6] Levy, P. S. and Lemeshow, S. (2013). *Sampling of populations: methods and applications*. John Wiley & Sons.

[7] SI, S. I. *Steps in applying Probability Proportional to Size (PPS) and calculating Basic Probability Weights*.

[8] Alam, M., Sumy, S. A. and Parh, Y. A. (2015). Selection of the samples with probability proportional to size. *Science Journal of Applied Mathematics and Statistics*. **3(5)**. 230-233.

[9] Piazza, T. (2010). Fundamentals of applied sampling. *Handbook of survey research* . 139-168.

[10] Salman, M. D. (2008).*Animal disease surveillance and survey. systems: methods and applications*. John Wiley & Sons.

[11] Magnani, R. (1999). *Sampling guide*, Food Security and Nutrition Monitoring (IMPACT) Project.

[12] Salman, M. (2003). *Chapter 1. Animal Disease Surveillance and Survey Systems*.

[13] Brus, D. J. (2022). *Spatial Sampling with R.* CRC Press.

[14] Levy, P. and Lemeshow, S. (1980). *Sampling for Health Professionals Lifetime Learning Publications*. Belmont, California.

[15] Gall, B. (2005). *Educational research: An introduction.* Educational design research.

[16] Berg, S. (1972). Double Sampling for Cluster-or Two-Stage Sampling. *International Statistical Review/Revue Internationale de Statistique*. **40**, 13-22.

Two-stage Cluster Sampling Design in Field Researches continuing title in this line (if required)

Kamran Mirzaie[1] and Maryam Parsaeian[2]

Abstract:

In the present article, an attempt has been made to provide a fully operational and useful guide regarding the implementation of two-stage cluster sampling plan in field research, which is one of the most widely used sampling plans according to the current conditions and facilities in the country. . In addition, for further use by researchers in this field, this article provides codes for implementing two-stage cluster sampling with unequal size based on probability proportional to size along with an example on two-stage cluster sampling with probability proportional to size (PPS) with hypothetical data using R software is included.

Keywords: sampling frame, probability proportional to sample size, random sampling, two-stage cluster sampling, field research.

[1] **Iran Veterinary Organization, Epidemiologist**

[2] **Ph.D. Candidate, Faculty of Psychology and Education, University of Tehran, Tehran, Iran,** $maryam.parsaeian@ut.ac.ir$