

برآورد نارایب نسبت جمعیت‌های پنهان در معرض بیماری‌های پرخطر

مهسا سعادت^۱، آرزو باقری^۱

مقاله پژوهشی

چکیده

مقدمه: از آن‌جا که سلامت جامعه در اثر بروز بیماری‌های پرخطر با مخاطره همراه می‌باشد، همواره جمعیت‌های در معرض این بیماری‌ها به خصوص جمعیت‌های پنهان، مورد توجه پژوهشگران و سیاست‌گذاران بهداشت عمومی قرار دارند. روش‌های متداولی که برای نمونه‌گیری و محاسبه برآورد نسبت این جمعیت‌ها مورد استفاده محققان قرار می‌گیرد، اغلب منجر به کم یا بیش‌برآوردی این نسبت‌ها در جمعیت‌های مورد مطالعه می‌گردد. با وجود معرفی روش‌های نمونه‌گیری کارایی همچون روش نمونه‌گیری پاسخگو محور از بیش از دو دهه پیش تاکنون، اما به دلیل عدم آشنایی پژوهشگران این حوزه با نحوه محاسبه برآوردها در نمونه‌های به دست آمده از این روش، کمتر از آن در برآورد نسبت جمعیت‌های پنهان استفاده می‌گردد. هدف از انجام مطالعه حاضر، معرفی برآوردهای نسبت‌های جمعیتی برای متغیرهای کیفی مانند ابتلا به بیماری، با استفاده از برآوردهای احتمال روابط درون و برون‌گروهی و اندازه شبکه اجتماعی پاسخگویان بود.

روش‌ها: با فرض وجود روابط دوطرفه در جمعیت هدف و انجام نمونه‌گیری با جای‌گذاری، در این مطالعه برآورد نسبت‌های جمعیتی با استفاده از برآوردهای احتمال روابط درون و برون‌گروهی و اندازه شبکه اجتماعی پاسخگویان محاسبه گردید.

یافته‌ها: تئوری‌های موجود و شبیه‌سازی‌های رایانه‌ای نشان داد که برآوردهای معرفی شده برای نسبت جمعیت‌های پنهان، به طور مجانبی نارایب بودند و نرخ همگرایی بالایی داشتند.

نتیجه‌گیری: عدم انتخاب صحیح روش نمونه‌گیری و همچنین، محاسبه برآورد نسبت جمعیت‌های پنهان که در معرض بیماری‌های پرخطر هستند و در سیاست‌گذاری‌های بهداشتی تأثیرگذار می‌باشند، نتایج قابل‌قبولی در رسیدن به اهداف این سیاست‌گذاری‌ها ارایه نخواهد داد.

واژه‌های کلیدی: جمعیت‌های پنهان در معرض بیماری‌های پرخطر، روش نمونه‌گیری پاسخگو محور، برآوردهای نارایب مجانبی، متغیرهای کیفی

ارجاع: سعادت مهسا، باقری آرزو. برآورد نارایب نسبت جمعیت‌های پنهان در معرض بیماری‌های پرخطر. مجله تحقیقات نظام سلامت ۱۳۹۵؛ ۱۲ (۴): ۵۲۶-۵۲۰

پذیرش مقاله: ۱۳۹۵/۹/۱۵

دریافت مقاله: ۱۳۹۵/۳/۱۸

اصلاح شده نمونه‌گیری ارجاع زنجیره‌ای و نمونه‌گیری پاسخگو محور (Respondent driven sampling) که به گروهی از روش‌های نمونه‌گیری احتمالی با عنوان سازوار (Adaptive) تعلق دارند، توسط Heckathorn معرفی شد (۲). روش نمونه‌گیری پاسخگو محور با انتخاب نمونه‌های اولیه (Initial samples) و یا هسته‌ها (Seeds) آغاز می‌شود. انتخاب هسته‌ها به طور عمدتاً غیر تصادفی است و مرحله اول عضوگیری این نمونه‌گیری از آنان آغاز می‌گردد. به هسته‌ها تعداد مشخصی کوپن (Coupon) برای نمونه‌گیری از آشنایان و همتایان (Peers) داده می‌شود. کوپن در واقع سهمیه هر فرد برای به عضویت درآوردن همتایان خود است که در این روش اغلب سه کوپن به هر عضوگیر تخصیص داده می‌شود (۳). اطلاعاتی در مورد طرح، زمان کاری و مکان ستاد آمارگیری و شماره شناسایی منحصر به فرد جهت مرتبط نمودن عضوگیر و عضو شونده بر روی کوپن‌ها درج شده است. عضو شونده نیز پس از اتمام مصاحبه خود، کوپن‌هایی را برای توزیع میان همتایان دریافت می‌نماید و در نتیجه، زنجیره‌های عضوگیری (Recruitment chains) تشکیل می‌شوند. علاوه بر آن، این روش نیاز به جمع‌آوری اطلاعات مربوط به درجه (Degree)

مقدمه

جمعیت‌های پنهان (Hidden) جوامعی هستند که نسبت کوچکی از کل جمعیت، اما اغلب مؤثر در سلامت جامعه را تشکیل می‌دهند. نمونه‌گیری از این جمعیت‌ها به دلیل عدم دسترسی به چارچوب نمونه‌گیری که نمونه‌ها از آن استخراج می‌شوند، با مشکلات ویژه‌ای در استفاده از روش‌های استاندارد آماری روبه‌رو می‌باشد. اعضای چنین جمعیت‌هایی، موضوع پژوهش‌های بسیاری در حوزه‌های مختلف علمی اعم از سلامت و بهداشت عمومی (مانند معتادان، مبتلایان به بیماری‌های پرخطری همچون ایدز و کارگران جنسی)، برنامه‌ریزی و سیاست‌گذاری (مانند مهاجران غیر قانونی و بی‌خانمان‌ها) و هنر و فرهنگ (مانند موزیسین‌های زیرزمینی و هنرمندان سالمند) هستند. از آن‌جا که روش‌های نمونه‌گیری متداول در نمونه‌گیری از این جوامع ناکارآمد می‌باشد، روش‌های ارجاع زنجیره‌ای (Chain referral) به دلیل توانایی در رسیدن به اعضای جمعیت پنهان، بسیار مورد توجه پژوهشگران قرار گرفته است. با این حال، تعمیم استنباط‌های حاصل از نمونه‌های به دست آمده با استفاده از این روش‌ها معتبر نیست (۱).

۱- استادیار، مؤسسه مطالعات و مدیریت جامع و تخصصی جمعیت کشور، تهران، ایران

نویسنده مسؤول: آرزو باقری

Email: arezoo.bagheri@psri.ac.ir

روشها

شکل ۱ (قسمت الف) الگوی جمعیت فرضی و شکل ۱ (قسمت ب) اختلاف میان روابط درون گروهی (خطوط نقطه چین) و برون گروهی (خطوط متصل) را برای یک جمعیت فرضی که از دو گروه دایره (A) و لوزی (B) تشکیل شده است را نشان می‌دهد (۲۰).

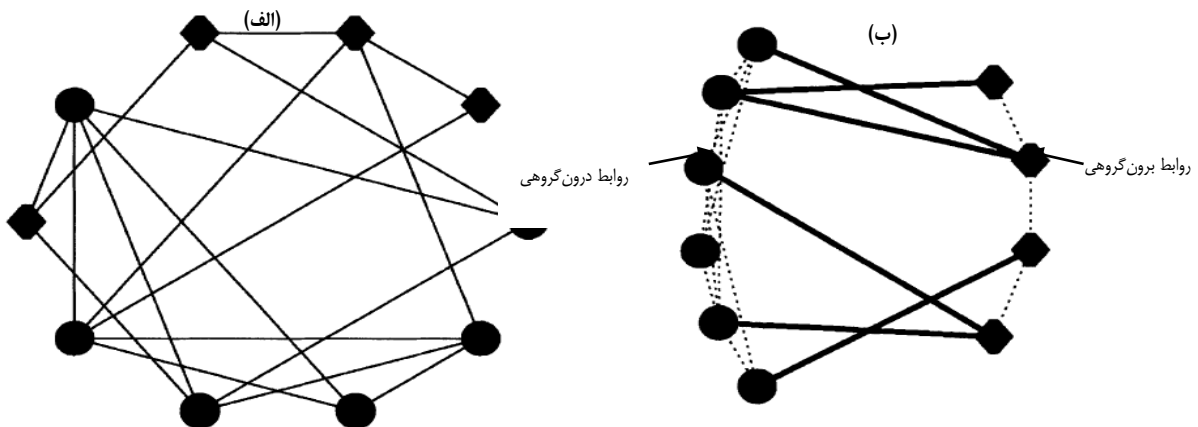
به عنوان مثال، جمعیت فرضی از دو گروه A و B با عفونت HIV مثبت و منفی تشکیل شده است. هدف، برآورد نسبت جمعیت هر یک از این دو گروه در جمعیت می‌باشد. در شکل ۱ (قسمت ب)، ۶ گره (Nodes) که عضو گروه A و ۴ گره که عضو B هستند، قرار دارد. ۶ ارتباط (Edge) (تعداد خطوط متصل بین این دو گروه)، از گروه A به گروه B مشاهده می‌شود. از دو طریق می‌توان تعداد ارتباطات بین گروهی را محاسبه نمود. اگر همه اطلاعات مربوط به شبکه اجتماعی افراد در ماتریسی مانند X قرار گیرد، در این ماتریس در صورتی که رابطه مستقیمی بین فرد i و j وجود داشته باشد و $X_{ij} = 1$ ، در غیر این صورت $X_{ij} = 0$ است. روابط دوطرفه (Reciprocal) هستند؛ یعنی اگر $X_{ij} = 1$ آن‌گاه $X_{ji} = 1$ (اگر فرد A فرد B را می‌شناسد، فرد B نیز فرد A را خواهد شناخت) خواهد بود. به منظور سادگی، این روابط رابطه‌های دوستی در نظر گرفته می‌شوند (هرچند می‌توانند هرگونه رابطه دیگری نیز باشند). درجه فرد i، D^i تعریف می‌گردد که در واقع تعداد دوستانی است که این فرد دارد ($D^i = \sum_j X_{ij}$). T_a تعداد کل دوستانی است که از افراد گروه A نشأت می‌گیرد. اگر N_a تعداد اعضای گروه A و D_a مجموع درجه‌های اعضای گروه A باشد، بدین ترتیب T_a به صورت رابطه ۱ تعریف می‌شود.

$$T_a = \sum_{i \in A} D^i = N_a \cdot D_a \quad \text{رابطه ۱}$$

$$T_b = \sum_{i \in B} D^i = N_b \cdot D_b$$

اگر ماتریس X مفروض باشد، احتمال یک رابطه دوستی تصادفی که عضوی از گروه A را به عضوی از گروه B مرتبط می‌سازد، به صورت رابطه ۲ تعریف می‌شود.

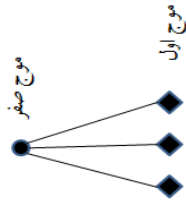
$$S_{ab} = \frac{T_{ab}}{T_a} \quad \text{رابطه ۲}$$



شکل ۱. الگوی روابط در یک جمعیت فرضی (الف) و اختلاف میان روابط درون و برون گروهی (ب)

هر پاسخگو دارد. تعداد افرادی که واجد شرایط مشارکت در طرح هستند و در دوره زمانی مشخص شده در طرح، عضو آن‌ها را ملاقات کرده است، درجه عضوگیری را تشکیل می‌دهند. در واقع، احتمال انتخاب برای هر شرکت کننده، متناسب با درجه آن فرد است. اندازه شبکه‌های اجتماعی افراد به منظور وزن دهی برآوردها و تولید برآوردهای نارایب در این روش نمونه‌گیری جمع‌آوری می‌شود. شبکه‌های اجتماعی (Social networks) ساختار اجتماعی افرادی است که از طریق یک یا دو عامل وابستگی به هم مرتبط شده‌اند. یکی دیگر از مشخصه‌های مهم در این روش نمونه‌گیری، هوموفیلی (Homophily) است که در واقع ابزاری برای ارزیابی سطح آریبی و همچنین، معرف بودن نمونه‌های پاسخگو محور فراهم می‌کند. هوموفیلی نشان می‌دهد که کدام یک از انواع روابط اجتماعی درون گروهی و یا برون گروهی، عضوگیری را کنترل می‌نماید.

در سال‌های اخیر، از این روش نمونه‌گیری در بررسی میزان عفونت HIV (Human immunodeficiency virus) میان معناتان تزریقی (۴-۶)، کارگران جنسی (۷)، مردان همجنس‌گرا (۸، ۹) و مهاجران (۱۰) در کشورهای مختلف و در ایرن میان معناتان تزریقی (۱۱، ۱۲) و کارگران جنسی (۱۳، ۱۴) استفاده می‌شود. همچنین، مطالعات بسیاری به معرفی و بررسی روش‌های استنباط آماری انواع متغیرها در نمونه‌گیری‌های موفق از جمعیت‌های پنهان پرداخته‌اند که هر یک از آن‌ها بسته به نوع نمونه‌گیری، پیچیدگی و ویژگی‌های خاص خود را دارد (۱۵-۱۹). از آن‌جا که بسیاری از متغیرهای مورد بررسی در مطالعات مختلف بهداشتی (مانند ابتلای افراد به بیماری)، متغیرهای کیفی هستند و روش‌های استنباط آماری برای برآورد نسبت جمعیت این متغیرها با متغیرهای کمی تفاوت دارد؛ هدف اصلی از انجام این تحقیق، بررسی نحوه محاسبه برآوردهای نسبت جمعیت برای این نوع متغیرها در نمونه‌گیری پاسخگو محور بود که از تحلیل‌های ساختار شبکه نشأت می‌گیرد. بدین منظور به بررسی روش برآورد نسبت جمعیت برای این متغیرها پرداخته شد. بخش یافته‌ها نیز به طور مجانبی نارایب بودن برآوردهای معرفی شده با استفاده از شبیه‌سازی‌های رایانه‌ای را نشان می‌دهد.



شکل ۲. موج و دوره عضوگیری

$$EI(e_{j \rightarrow k})_{r=x} = \begin{cases} \text{رابطه ۷} & \text{اگر رابطه } e_{j \rightarrow k} \text{ در دوره عضوگیری } \lambda \text{ام انتخاب شده باشد} \\ \cdot & \text{در غیر این صورت} \end{cases}$$

همچنین، به منظور محاسبه برآوردهای نسبت جمعیت در روش نمونه‌گیری پاسخگو محور، فرضیات زیر در نظر گرفته می‌شود.

✓ نمونه‌گیری همراه با جایگذاری است. این فرض محاسبات را آسان‌تر می‌نماید و با افزایش تعداد جمعیت، تغییرات در جمعیت حذف می‌شود (۱۲).

✓ بین هر دو نفر در شبکه جمعیت‌های پنهان رابطه‌ای وجود دارد. بسیاری از جمعیت‌های پنهان مانند معنادان تزیقی، شبکه دوستی متراکم دارند در نتیجه، این فرض برای آنان منطقی به نظر می‌رسد.

✓ تمام پاسخگویان تنها یک کوپن دریافت می‌کنند. نتایج به دست آمده در این حالت به راحتی قابل تعمیم به حالتی است که تعداد کوپن‌های عضوگیری بیشتر از یک کوپن باشد.

✓ وقتی پاسخگویان عضوگیری می‌نمایند، عضوگیری به صورت تصادفی از تمام روابط موجود میان آنان انجام می‌گیرد. این فرض با در نظر گرفتن نمونه‌گیری با جایگذاری می‌تواند به صورت رابطه ۸ مطرح گردد.

$$\Pr[EI(e_{j \rightarrow k})_{r=x+1} = \mathbb{1} | NI(j)_{w=x} = \mathbb{1}] = \frac{1}{D_j} \quad \text{رابطه ۸}$$

✓ هسته‌ها با احتمال متناسب با روابط خود انتخاب می‌شوند؛ بدین معنی که احتمال انتخاب فردی با $2K$ رابطه دوستی، دو برابر احتمال انتخاب فردی با K رابطه دوستی است. این فرض می‌تواند به صورت رابطه ۹ مطرح شود.

$$\Pr[NI(j)_{w=0} = \mathbb{1}] = \frac{D_j}{\sum_{i \in N} D_i} \quad \text{رابطه ۹}$$

از آن‌جا که هسته‌ها در مطالعاتی که بر روی جمعیت‌های پنهان انجام می‌شود، با احتمال متناسب با درجه انتخاب نمی‌شوند (به طور غیر تصادفی انتخاب می‌شوند)، فرایند نمونه‌گیری به نقطه‌ای همگرا تبدیل می‌شود که در آن برآوردها نارایب می‌باشند (۲۱).

برای محاسبه برآورد نارایب نسبت جمعیت، ابتدا باید نشان داد که چگونه می‌توان از نحوه عضوگیری برای برآورد نمودن احتمال روابط بین گروهی S_{ab} و S_{ba} استفاده نمود. دوم این که، برآورد اندازه درجه افراد در گره‌های D_a و D_b با استفاده از اندازه شبکه اجتماعی که خود آنان اظهار می‌کنند، قابل برآورد است. سوم این که، این برآوردها می‌توانند ترکیب شوند تا برآورد نسبتی از جمعیت که در هر یک از این گروه‌ها قرار می‌گیرد، P_a و P_b به دست آید.

وقتی که T_{ab} تعداد روابطی که نشان دهنده ارتباط میان عضوی از گروه A با عضوی از گروه B است، باشد؛ از آن‌جا که روابط دو طرفه هستند، تعداد روابط از گروه A به B برابر با تعداد روابط از گروه B به A است. این تعداد روابط را می‌توان از دو طریق محاسبه نمود.

- حاصل ضرب تعداد روابطی که از گروه A نشأت گرفته است. T_a در احتمال این که یکی از این روابط با فردی از گروه B ، S_{ab} باشد.
- حاصل ضرب تعداد روابطی که از گروه B نشأت گرفته است. T_b در احتمال این که یکی از این روابط با فردی از گروه A ، S_{ba} باشد.

$$T_a \cdot S_{ab} = T_{ab} \quad \text{رابطه ۳}$$

$$T_b \cdot S_{ba} = T_{ba}$$

با مساوی قرار دادن طرفین روابط ۳ با یکدیگر و استفاده از تعاریف T_a و T_b در رابطه ۱، معادله ۴ حاصل می‌شود.

$$N_a D_a S_{ab} = N_b D_b S_{ba} \quad \text{رابطه ۴}$$

رابطه ۴ حاوی اطلاعاتی از روابط و مشخصه‌های شبکه اجتماعی افراد است. بنابراین، خصوصیات روابط از طریق اطلاعات مربوط به اندازه شبکه اجتماعی استنباط می‌شود. حتی اگر اطلاعات کاملی در مورد شبکه اجتماعی در اختیار باشد؛ یعنی D_a, D_b, S_{ab} و S_{ba} معلوم باشد، هنوز دو مجهول اندازه‌های جمعیت گروه A و گروه B ، N_a و N_b ، وجود دارد. اگر هر دو طرف رابطه ۴ بر N تعداد کل جمعیت تقسیم شود، آنگاه P_a و P_b در معادله $N_a D_a S_{ab} = N_b D_b S_{ba}$ قابل محاسبه خواهد بود که با در نظر گرفتن $P_a + P_b = 1$ و حل سیستمی با دو معادله و دو مجهول، مقدار نسبت‌های جمعیت را می‌توان از روابط ۵ با استفاده از اطلاعات مربوط به ساختار شبکه‌هایی که گروه‌ها را در جمعیت به هم متصل می‌کند، محاسبه نمود.

$$P_a = \frac{S_{ba} D_b}{S_{ba} D_b + S_{ab} D_a} \quad \text{رابطه ۵}$$

$$P_b = \frac{S_{ab} D_a}{S_{ab} D_a + S_{ba} D_b}$$

روابط فوق برای هر ساختار شبکه‌ای با روابط دوطرفه صحیح می‌باشد (۲۲، ۲۱). به منظور ایجاد برآوردهای نسبت جمعیت در نمونه‌گیری پاسخگو محور، باید فرضیاتی در مورد جامعه مورد مطالعه و همچنین، نحوه عضوگیری در نظر گرفته شود. فرایند انتخاب نمونه‌ها در این روش نمونه‌گیری به صورت روشی برای انتخاب گره‌ها و روابط در نظر گرفته می‌شود؛ بدین معنی که ابتدا در موج صفر گره‌ها و سپس گره‌ها رابطه‌هایی که دوره عضوگیری (Recruitment period) اول را تعیین می‌نمایند، انتخاب می‌شوند. رابطه‌ها در دوره عضوگیری اول، گره‌ها در موج اول را مشخص می‌کنند. این فرایند که گره‌ها روابط و روابط گره‌ها را تعیین می‌کنند تا رسیدن به اندازه نمونه مطلوب، ادامه خواهد یافت. شکل ۲ نشان دهنده مفاهیم موج و دوره عضوگیری برای عضوی از گروه A که سه عضو از گروه B را عضوگیری نموده است، می‌باشد.

می‌توان $NI(j)_{w=x}$ و $EI(e_{j \rightarrow k})_{r=x}$ را به صورت رابطه ۶ و ۷ تعریف نمود (۲۰).

$$NI(j)_{w=x} = \begin{cases} \text{اگر گره } \lambda \text{ام در موج } \lambda \text{ام انتخاب شده باشد} & \mathbb{1} \\ \cdot & \text{در غیر این صورت} \end{cases} \quad \text{رابطه ۶}$$

در این حالت نیز نسبتی با برآوردهای ناریب به دست می‌آید که به طور مجانبی ناریب است. برآورد P_B نیز به صورت مشابه قابل محاسبه می‌باشد. برای تعمیم برآورد جمعیت‌های پنهان برای متغیرهای کیفی با بیش از دو گروه، مدل دوطرفه در یک سیستم با N گروه می‌تواند با مجموعه‌ای از معادلات نمایش داده شود که یکی از این معادلات نشان دهنده برابری مجموع نسبت جمعیت‌ها در زیرگروه‌ها با یک است ($P_A + P_B + P_C + \dots + P_N = 1$) و سایر معادلات بیانگر مؤلفه‌های دوطرفه‌ای برای هر زوج از گروه‌ها می‌باشد (تعداد این زوج‌ها برابر با $\frac{N(N-1)}{2}$ می‌باشد). در این حالت، سیستم معادلات بیش تبیین شده (Over-determined) است. روش‌های مناسب برای حل سیستم‌های معادلات در این حالت، روش کمترین مربعات خطی با منطقی مشابه با رگرسیون خطی و هموارسازی داده‌ها (Data smoothing) می‌باشد (۲، ۲۱، ۲۲).

یافته‌ها

در بخش قبل برآورد نسبت جمعیت معرفی گردید و در این بخش ناریبی برآوردهای معرفی شده با استفاده از شبیه‌سازی‌های رایانه‌ای بررسی خواهد شد. Heckathorn و Salganik در این شبیه‌سازی‌ها جمعیتی تولید کردند و سپس به طور مکرر نمونه‌هایی را با جایگذاری از آن انتخاب نمودند (۲۰). در شبیه‌سازی‌ها، شاخص‌های جمعیتی مختلفی همچون تعداد افراد جمعیت N ، نسبت جمعیت در گروه A ، P_A ، انتخاب توزیع درجه جمعیت، $\Gamma(D)$ که از آن درجه گروه‌ها (رابطه ۲) به دست می‌آید، قابل تغییر است. در واقع، برای درجه معلوم D ، $\Gamma_a(D)$ احتمال آن است که گرهی از گروه A دارای درجه D باشد. در نهایت، روابط درون‌گروهی بین دو گروه می‌تواند تغییر کند. در برخی موقعیت‌ها، ساختار دوستی ممکن است به شدت هموفیلی باشد (۲۸) و بدان معنی است که افراد در هر گروه شانس بیشتری را برای برقراری ارتباط با همان گروه دارند.

در شکل ۳ نتایج این شبیه‌سازی‌ها برای دامنه بزرگی از مقادیر نسبت جمعیت رسم شده است. مقدار متوسط برآوردها که با دایره مشخص شده‌اند، مشابه با مقادیر واقعی هستند که با خطوط نقطه‌چین نشان داده شده است. حاصل میانگین این برآوردها در تکرارهای زیاد، مقادیر واقعی نسبت جمعیت در گروه A می‌باشد. با اندازه نمونه ۵۰۰، مقدار اربیی ناچیز به دست می‌آید. خطوط عمودی نشان دهنده فواصل اطمینان ۹۰ درصدی برای برآوردها می‌باشد. بنابراین، به طور مجانبی ناریب بودن برآوردهای پیشنهادی ارایه می‌شود. شبیه‌سازی‌های عددی برای جمعیتی با شاخص‌هایی که پیش از این مطرح شد، در شکل ۴ نمایش داده شده است. دایره‌ها بیانگر متوسط برآورد در ۱۰۰ هزار نمونه است. نرخ همگرایی سریع و اندازه اربیی‌ها کوچک هستند. با اندازه نمونه ۵۰۰، اربیی کمتر از ۰/۰۰۱ و قابل چشم‌پوشی می‌باشد.

بحث

با این که بیش از دو دهه از معرفی روش نمونه‌گیری پاسخگو محور به عنوان روشی موفق و کارا برای نمونه‌گیری از جمعیت‌های پنهان می‌گذرد، اما همچنان در بسیاری از مطالعات که با جمعیت‌های پنهان در معرض خطر بیماری‌های

برای محاسبه برآوردهای S_{ab} و S_{ba} در رابطه ۲، باید چگونگی انتخاب عضوگیری‌های مشاهده شده از بین عضوگیری‌های ممکن مشخص گردد. این اطلاع از طریق ثبت شماره منحصر به فرد کوپن‌های عضوگیری قابل بررسی است. برای هر متغیر، ماتریس عضوگیری R در نمونه به دست آمده به صورت رابطه ۱۰ محاسبه می‌شود.

$$\hat{T} = \begin{bmatrix} \hat{T}_{aa} & \hat{T}_{ab} \\ \hat{T}_{ba} & \hat{T}_{bb} \end{bmatrix} \quad \text{رابطه ۱۰}$$

که در آن، عضوگیری فردی از گروه A توسط فرد دیگری از گروه A ، \hat{T}_{aa} ؛ عضوگیری فردی از گروه A توسط فردی از گروه B ، \hat{T}_{ab} ؛ عضوگیری فردی از گروه B توسط فرد دیگری از گروه B ، \hat{T}_{bb} و عضوگیری فردی از گروه B توسط فردی از گروه A ، \hat{T}_{ba} در نظر گرفته شده است. از آن‌جا که عضوگیری‌های مشاهده شده نمونه تصادفی از تمام روابط هستند، برآوردهای ناریب برای S_{ab} و S_{ba} به صورت رابطه ۱۱ تعریف می‌شود.

$$\frac{\hat{T}_{ab}}{\hat{T}_a} \widehat{S_{ab}} = \frac{\hat{T}_{ab}}{\hat{T}_{aa} + \hat{T}_{ab}} \quad \text{رابطه ۱۱}$$

$$\frac{\hat{T}_{ba}}{\hat{T}_b} \widehat{S_{ba}} = \frac{\hat{T}_{ba}}{\hat{T}_{bb} + \hat{T}_{ba}}$$

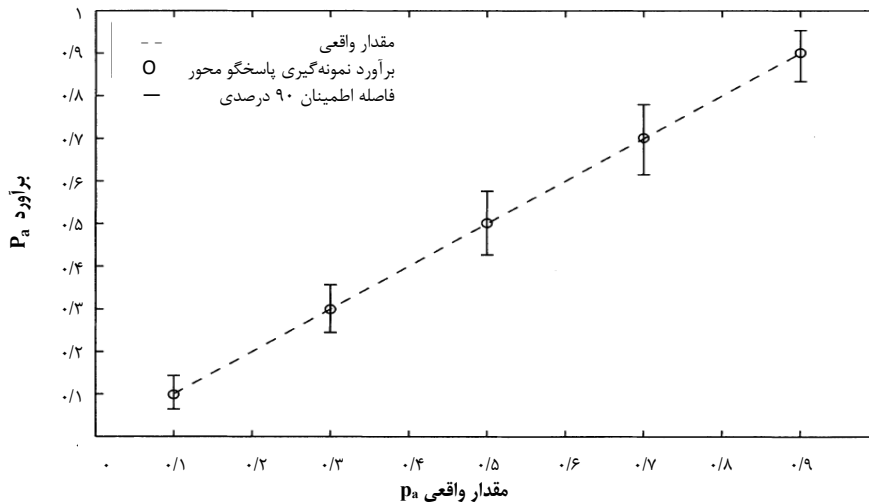
از آن‌جا که میانگین نمونه، برآوردر خوبی برای اندازه درجه نیست، باید اطلاعات نمونه به گونه‌ای تعدیل شود تا اطلاعات دقیقی را در مورد جامعه تولید نماید (۲۳). وقتی هدف، برآورد متوسط اندازه درجه افراد در گروه A ، D_a است؛ پژوهشگران نباید درجه هسته‌ها را نیز در نظر بگیرند؛ چرا که هسته‌ها به صورت غیر تصادفی انتخاب شده‌اند و افرادی با درجه‌های بزرگ هستند (۲۰). از آن‌جا که تعداد هسته‌ها نسبت به تعداد کل نمونه خیلی کوچک است؛ در نتیجه، اطلاعات کمی با در نظر نگرفتن آنان استفاده نخواهد شد. برای ایجاد برآورد متوسط اندازه درجه جمعیت در گروه A ، Heckathorn و Salganik میزان \widehat{D}_a را به صورت رابطه ۱۲ محاسبه نمودند (۲۰).

$$\widehat{D}_a = \frac{n_a}{\sum_{i=1}^{n_a} \frac{1}{\hat{T}_i}} \quad \text{رابطه ۱۲}$$

از آن‌جا که در رابطه فوق (۲۴) اندازه نمونه n_a به جای مخرج در صورت کسر قرار گرفته است، به نظر می‌رسد که برآوردگری برای متوسط درجه جمعیت نباشد؛ در صورتی که این برآوردر معادل با نسبت دو برآوردر Hansen و Hurwitz (۲۵) است که یکی برآورد کننده تعداد کل روابط \hat{T}_a (روابطی که از افراد گروه A نشأت می‌گیرد) و دیگری برآوردکننده N_a می‌باشد (۲۰). صورت و مخرج \widehat{D}_a هر دو برآوردرهای Hansen و Hurwitz هستند (۲۵) که ناریب می‌باشند (۲۶). نسبت دو برآوردر ناریب وقتی اندازه نمونه n باشد، به طور مجانبی درای اربیی n^{-1} است (۲۷). بنابراین، با بزرگ شدن n_a ، اندازه نمونه گروه A ، $D_a \rightarrow E[\widehat{D}_a]$ است. این مقدار اربیی در مورد اندازه نمونه‌های بزرگ قابل چشم‌پوشی می‌باشد (۲۷).

پس از برآورد اطلاعات مورد نیاز در مورد شبکه اجتماعی که افراد را به هم مرتبط می‌سازد، از این اطلاعات می‌توان برای برآورد P_a و P_b ، نسبت جمعیت‌های پنهان برای متغیرهای کیفی استفاده نمود. P_a با استفاده از روابط ۱۱ و ۱۲ به صورت رابطه ۱۳ قابل برآورد است.

$$\hat{P}_a = \frac{\hat{D}_b \cdot \widehat{S_{ba}}}{\hat{D}_a \cdot \widehat{S_{ab}} + \hat{D}_b \cdot \widehat{S_{ba}}} \quad \text{رابطه ۱۳}$$

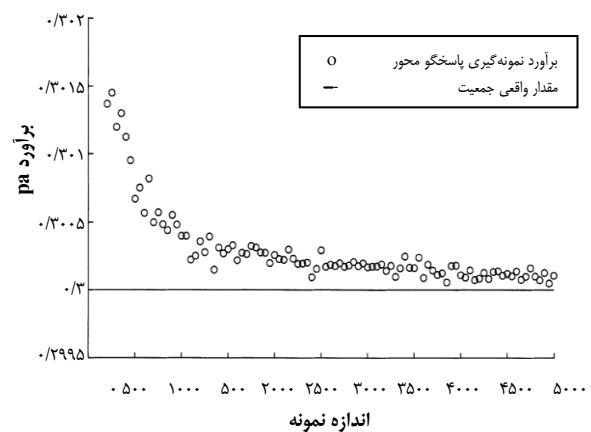


شکل ۳. شبیه‌سازی عددی برای اثبات به طور مجانبی نارایب بودن برآورد نسبت جمعیت

مؤسسه‌ای و نمونه‌گیری ارجاع زنجیره‌ای (دو روش نمونه‌گیری متفاوت) و مقایسه نمونه‌گیری ارجاع زنجیره‌ای و پاسخگو محور (دو روش برآورد متفاوت) پرداختند (۳۱). از نمونه‌های حاصل از روش نمونه‌گیری ارجاع زنجیره‌ای (۳۲) با استفاده از روش برآورد بحث شده در مطالعه حاضر، برآوردهای نمونه‌گیری پاسخگو محور به دست آمد. نتایج مطالعه Heckathorn نشان داد که عدم در نظر گرفتن طرح نمونه‌گیری موفق و روش برآورد مناسب، منجر به نتیجه‌گیری‌های گمراه‌کننده می‌شود. علت اصلی عدم به کارگیری نمونه‌گیری‌های مناسب در نفوذ به جمعیت‌های پنهان توسط پژوهشگران، عدم آگاهی آنان از نحوه محاسبه برآوردها و استنباط‌ها در این روش‌های نمونه‌گیری است. هدف اصلی از انجام پژوهش حاضر، بررسی برآوردهای نسبت جمعیت‌های پنهان برای متغیرهای کیفی برآمده از روش نمونه‌گیری پاسخگو محور بود. برآوردهای معرفی شده در این مطالعه برای محاسبه نسبت‌های جمعیت متغیرهای کیفی قابل استفاده می‌باشند، هرچند با اندکی اغماض می‌توان از برآوردهای معرفی شده برای متغیرهای کمی نیز که بر حسب چندک‌ها طبقه‌بندی شده‌اند، استفاده نمود (۲۲).

تشکر و قدردانی

تحقیق حاضر برگرفته از طرح «استنباط آماری نمونه‌گیری پاسخگو محور برای برآورد پارامترهای جمعیتی» می‌باشد که با حمایت مؤسسه مطالعات و مدیریت جامع و تخصصی جمعیت کشور در سال ۱۳۹۴ با شماره ابلاغ طرح ۲۰/۱۸۶۱۶ انجام گرفت.



شکل ۴. شبیه‌سازی عددی برای اثبات ادعای سریع بودن نرخ همگرایی به برآورد واقعی

خاص در حوزه‌های سلامت و بهداشت مواجه هستند، اغلب از روش‌های نمونه‌گیری متداولی همچون نمونه‌گیری مؤسسه‌ای جهت بررسی این جمعیت‌ها استفاده می‌شود که نتایج حاصل از این پژوهش‌ها در اغلب موارد گمراه‌کننده است (۳۰، ۲۹). از آن‌جا که از این اطلاعات اغلب برای سیاست‌گذاری‌ها استفاده می‌شود؛ نتایج نادرست منجر به بیراهه رفتن اهداف سیاست‌گذاری‌ها می‌گردد. Heckathorn و Jeffri در مطالعه خود به مقایسه نتایج حاصل از نمونه‌گیری

References

1. Kalton G. Introduction to survey sampling. New York, NY: SAGE; 1983.
2. Heckathorn DD. Respondent-driven sampling: a new approach to the study of hidden populations restricted access. Social Problems 1997; 44(2): 174-99.
3. Abdul-Quader AS, Heckathorn DD, Sabin K, Saidel T. Implementation and analysis of respondent driven sampling: lessons learned from the field. J Urban Health 2006; 83(6 Suppl): i1-i5.
4. Mumtaz GR, Weiss HA, Thomas SL, Riome S, Setayesh H, Riedner G, et al. HIV among people who inject drugs in the Middle East and North Africa: systematic review and data synthesis. PLoS Med 2014; 11(6): e1001663.

5. Young AM, DiClemente RJ, Halgin DS, Sterk CE, Havens JR. Drug users' willingness to encourage social, sexual, and drug network members to receive an HIV vaccine: a social network analysis. *AIDS Behav* 2014; 18(9): 1753-63.
6. Stormer A, Tun W, Guli L, Harxhi A, Bodanovskaia Z, Yakovleva A, et al. An analysis of respondent driven sampling with Injection Drug Users (IDU) in Albania and the Russian Federation. *J Urban Health* 2006; 83(6 Suppl): i73-i82.
7. Johnston LG, Huong PT, Hien MT. HIV prevalence and risk factors among injecting drug users in Ho Chi Minh City, Vietnam. *Proceedings of the 16th International Conference on the Reduction of Drug Related Harm*; 2005 Mar 20-24; Belfast, Northern Ireland.
8. Liu H, Liu H, Cai Y, Rhodes AG, Hong F. Money boys, HIV risks, and the associations between norms and safer sex: a respondent-driven sampling study in Shenzhen, China. *AIDS Behav* 2009; 13(4): 652-62.
9. Johnston LG, Khanam R, Reza M, Khan SI, Banu S, Alam MS, et al. The effectiveness of respondent driven sampling for recruiting males who have sex with males in Dhaka, Bangladesh. *AIDS Behav* 2008; 12(2): 294-304.
10. Montealegre JR, Risser JM, Selwyn BJ, McCurdy SA, Sabin K. Effectiveness of respondent driven sampling to recruit undocumented Central American immigrant women in Houston, Texas for an HIV behavioral survey. *AIDS Behav* 2013; 17(2): 719-27.
11. Malekinejad M, Mohraz M, Razani N, Akbari G, McFarland W, Khairandish P, et al. High HIV prevalence in a respondent-driven sampling survey of injection drug users in Tehran, Iran. *AIDS Behav* 2015; 19(3): 440-9.
12. Rahnama R, Mohraz M, Mirzazadeh A, Rutherford G, McFarland W, Akbari G, et al. Access to harm reduction programs among persons who inject drugs: findings from a respondent-driven sampling survey in Tehran, Iran. *Int J Drug Policy* 2014; 25(4): 717-23.
13. Kazerooni PA, Motazedian N, Motamedifar M, Sayadi M, Sabet M, Lari MA, et al. The prevalence of human immunodeficiency virus and sexually transmitted infections among female sex workers in Shiraz, South of Iran: by respondent-driven sampling. *Int J STD AIDS* 2014; 25(2): 155-61.
14. Navadeh S, Mirzazadeh A, Mousavi L, Haghdoost A, Fahimfar N, Sedaghat A. HIV, HSV2 and syphilis prevalence in female sex workers in Kerman, south-east Iran; using respondent-driven sampling. *Iran J Public Health* 2012; 41(12): 60-5.
15. Bagheri A, Saadati M. A Survey on link-tracing sampling methods (social networks) and its application on demography [Research]. Tehran, Iran: National Population Studies and Comprehensive Management Institute; 2014. [In Persian].
16. Bagheri A, Saadati M. Respondent driven sampling statistical inferences to estimate demographical parameters [Research]. Tehran, Iran: National Population Studies and Comprehensive Management Institute; 2015. [In Persian].
17. Bagheri A. Respondent driven sampling: a new approach to sampling rare and hidden diseases. *J Health Syst Res* 2016; 11(4): 753-61.
18. Saadati M, Bagheri A. Respondent driven sampling method compared with other sampling methods of hidden populations. *Iran J Epidemiol* 2016; 12(2): 63-74.
19. Bagheri A, Saadati M. Exploring the effectiveness of chain referral methods in sampling hidden populations. *Indian J Sci Technol* 2015; 8(30): 1-8.
20. Salganik MJ, Heckathorn DD. Sampling and estimation in hidden populations using respondent-driven sampling. *Sociol Methodol* 2004; 34(1): 193-240.
21. Heckathorn DD. Respondent-driven sampling ii: deriving valid population estimates from chain-referral samples of hidden populations. *Social Problems* 2002; 49(1): 11-34.
22. Heckathorn DD. Extensions of respondent-driven sampling: analyzing continuous variables and controlling for differential recruitment. *Sociol Methodol* 2007; 37(1): 151-208.
23. Eland-Goossens MA, Van De Goor LA, Vollemans EC, Hendriks VM, Garretsen HF. Snowball Sampling Applied to Opiate Addicts Outside the Treatment System. *Addiction Research* 1997; 5(4): 317-30.
24. Verma V. Sampling elusive populations: Applications to studies of child labour. Geneva, Switzerland: International Labour Organization; 2013.
25. Hansen MH, Hurwitz WN. On the theory of sampling from finite populations. *Ann Appl Stat* 1943; 14(4): 333-62.
26. Brewer KR, Hanif M. Sampling with Unequal Probability. In: Cochran WG, Editor. *Sampling Techniques*. 3rd ed. New York, NY: Wiley Eastern Limited; 1977.
27. Erdos P, Renyi A. On random graphs. In: Cochran WG, Editor. *Sampling techniques*. New York, NY: Wiley Eastern Limited; 1977. p. 290-7.
28. McPherson M, Smith-Lovin L, Cook JM. Birds of a Feather: Homophily in Social Networks. *Annu Rev Sociol* 2001; 27(1): 415-44.
29. Thompson SK, Frank O. Model-based estimation with link-tracing sampling designs. *Survey methodology* 2000; 26(1): 87-98.
30. Thompson SK, Collins LM. Adaptive sampling in research on risk-related behaviors. *Drug Alcohol Depend* 2002; 68(Suppl 1): S57-S67.
31. Heckathorn DD, Jeffri J. Finding the beat: Using respondent-driven sampling to study jazz musicians. *Poetics* 2001; 28(4): 307-29.
32. Goodman L.A. Snowball sampling. *Ann Math Stat* 1961; 20: 572-9.

Unbiased Estimator of Population Proportion for Hidden Populations Exposed to High-Risk Diseases

Mahsa Saadati¹, Arezoo Bagheri¹

Abstract

Background: Since society health is threatened by high risk diseases, populations exposed to these diseases, especially hidden populations, always attract the attention of researchers and policy makers in the field of public health. Conventional methods that are used by researchers for sampling and calculating population proportion estimation often lead to underestimation or overestimation of these proportions in the studied populations. Efficient sampling methods such as respondent-driven sampling (RDS) method have been introduced more than two decades ago. However, due to the unfamiliarity of researchers in this field with the technique of calculating estimations for samples in this method, this sampling method is less applied in estimating proportions of hidden populations. The main objective of the current study was to introduce estimators of population proportions for qualitative variables such as disease occurrence through estimates of the probability of intergroup and intragroup relations and respondents' social network size.

Methods: In the present study, by assuming the existence of reciprocal relationships in the population and sampling with replacement, the population proportions were computed through estimating the probability of intergroup and intragroup relations and social network size of respondents.

Findings: Existing theories and computer simulations showed that estimators introduced for proportions of hidden populations were asymptotically unbiased and had a high rate of convergence.

Conclusion: The lack of selection of a suitable sampling method and computing method for estimating proportions of hidden populations, which are exposed to high-risk diseases and are effective in health policies, will not provide acceptable results in achieving the objectives of this policy.

Keywords: Hidden populations exposed to high-risk diseases, Respondent-driven sampling method, Asymptotically unbiased estimators, Qualitative variables

Citation: Saadati M, Bagheri A. Unbiased Estimator of Population Proportion for Hidden Populations Exposed to High-Risk Diseases. J Health Syst Res 2017; 12(4): 520-6.

1- Assistant Professor, National Population Studies and Comprehensive Management Institute, Tehran, Iran

Corresponding Author: Arezoo Bagheri, Email: arezoo.bagheri@psri.ac.ir