مجله علمی دانشگاه علوم پزشکی سمنان- جلد ۵، شماره ۱ و ۲، پائیز و زمستان ۱۳۸۲

روش پوشش نمونه‌ای برای برآورد اندازه جمعیت و کاربرد آن در کم‌شماری مرگ‌ومیر

ایمان تقدسی[۱](M.Sc)، انوشیروان کاظم‌نژاد*[۱](Ph.D)، ناهید سنجری[۲](Ph.D)، سقراط فقیه‌زاده[۱](Ph.D)، علی‌رضا ابدی[۳](Ph.D)

۱- دانشگاه تربیت مدرس، دانشکده علوم پزشکی، گروه آمار زیستی

۲- دانشگاه شیراز، دانشکده علوم پایه، گروه آمار

۳- دانشگاه علوم پزشکی شهید بهشتی، گروه آمار زیستی

چکیده

سابقه و هدف: یکی از معضلات و مشکلاتی که در سر راه برنامه‌ریزان، تصمیم‌گیران و مسئولین اجرایی کشور قرار دارد و آنها را در اتخاذ تصمیمات صحیح دچار مشکل می‌نماید، وجود آمارهای متعدد و گاهاً متناقض از پدیده‌های واحد می‌باشد. تولید آمارهای متعدد و گاهاً متناقض از جانب سازمان‌های مربوط به ثبت اطلاعات مرگ‌ومیر، موجب عدم اطمینان برنامه‌ریزان و تصمیم‌گیران به این آمارها گردیده است. این اختلاف در آمارها بیان‌گر این است که در سیستم جمع‌آوری اطلاعات این سازمان‌ها نوعی کم‌شماری (Undercount) وجود دارد. یکی از روش‌هایی که می‌تواند این نقیصه را برطرف نماید، روش پوشش نمونه‌ای (Sample coverage approach) می‌باشد. این مطالعه بر روی اطلاعات مرگ‌ومیر سال ۱۳۷۹ شهرستان بوشهر و به منظور برآورد تعداد متوفیان این شهرستان و تصحیح کم‌شماری انجام گردید.

مواد و روش‌ها: نخست، اطلاعات مرگ‌ومیر شهرستان بوشهر از سه منبع: بخش توسعه نظام آماری و بخش توسعه شبکه و ارتقاء سلامت وزارت بهداشت، درمان و آموزش پزشکی و سازمان ثبت احوال گردآوری شد. سپس اطلاعات هر منبع که براساس نام، نام خانوادگی، جنسیت و محل سکونت افراد متوفی بود، وارد رایانه گردید. آن‌گاه به منظور یافتن موارد مشترک منابع از نرم افزارهایی نظیر بانک اطلاعاتی Access استفاده گردید و پس از تجزیه و تحلیل داده‌ها با استفاده از نرم‌افزار Splus، نتایج زیر به دست آمد.

یافته‌ها: بدون احتساب مرگ‌ومیر نوزادان، تعداد کل متوفیان شهرستان بوشهر در سال ۱۳۷۹، ۶۶۰ نفر با خطای معیار ۵ نفر برآورد گردید؛ در حالی‌که تنها ۶۴۱ مورد مرگ توسط سازمان‌های مربوطه گزارش شده بود. در نتیجه تعداد کم‌شماری توسط این سازمان‌ها به طور متوسط ۱۹ نفر معادلِ ۲/۸۸ درصد برآورد گردید.

نتیجه‌گیری: یافته‌های این تحقیق نشان می‌دهند که در سیستم جمع‌آوری اطلاعات مرگ‌ومیر، کم‌شماری وجود دارد و با استفاده از روش پوشش نمونه‌ای می‌توان برآورد قابل قبولی برای آن به دست آورد.

واژه‌های کلیدی: روش پوشش نمونه‌ای، هم‌پوشانی اطلاعات، کم‌شماری، بوت‌استرپ

مقدمه

یکی از دغدغه‌های مدیران، تصمیم‌گیران و مسئولین اجرایی کشور، وجود آمارهای متعدد و گاهاً متناقض از پدیده‌های واحد می‌باشد. به عبارت دیگر آمار و ارقام متفاوتی از جانب سازمان‌های مختلف در ارتباط با یک پدیده واحد تولید می‌شود. در حال حاضر در کشور ما، سازمان ثبت احوال و وزارت بهداشت، درمان و آموزش پزشکی، هر یک با شیوه‌های متفاوتی به ثبت اطلاعات مربوط به مرگ‌ومیر

*نویسنده مسئول. تلفن: ۸۰۱۳۰۳۰-۰۲۱، نمابر: ۸۰۱۳۰۳۰-۰۲۱، E-mail: kezem_an@modares.ac.ir

[Downloaded from koomeshjournal.semums.ac.ir on 2024-07-31]

می‌پردازند، اما آمار و ارقامی که تولید می‌کنند اغلب با هم تفاوت دارد و این اختلاف در آمارها این فرضیه را که ممکن است کم‌شماری در این سیستم‌ها وجود داشته باشد، تقویت می‌کند. به عنوان نمونه تعداد متوفیان استان بوشهر در سال ۱۳۷۹ توسط بخش توسعه نظام آماری وزارت بهداشت، درمان و آموزش پزشکی ۳۲۸۲ نفر [۱] و توسط بخش توسعه شبکه و ارتقاء سلامت وزارت بهداشت، درمان و آموزش پزشکی ۳۰۱۰ نفر [۲] و توسط سازمان ثبت احوال کشور ۲۹۱۴ نفر گزارش گردیده است؛ که اختلاف در این آمارها بیان‌گر نوعی کم‌شماری در سیستم‌های جمع‌آوری اطلاعات این سازمان‌ها می‌باشد. لذا برای به‌دست آوردن میزان کم‌شماری از روش پوشش نمونه‌ای استفاده می‌کنیم. به طور کلی در هر زمینه‌ای که حداقل دو فهرست (لیست) ناکامل داشته باشیم، این روش قابل اجرا می‌باشد. از آنجا که در اپیدمیولوژی، لیست‌ها می‌توانند از منابع مختلفی از قبیل ثبت‌های بیمارستانی، پرونده‌های پزشکی، نسخه‌های پزشکی، ویزیت دکترها و... تهیه شوند، بنابراین اجرای روش پوشش نمونه‌ای، عملی خواهد بود. به عبارت دیگر، هرگاه یک موضوع خاص (مانند دیابت) را بتوان حداقل از دو طریق (منبع، فهرست، دام و یا بازبینی) مشاهده کرد؛ اجرای روش پوشش نمونه‌ای عملی خواهد بود.

ایده پوشش نمونه‌ای در گونه‌های گیاهی و برای برآورد فراوانی وحوش به سال‌های دور برمی‌گردد [۱۰]، همچنین کاربرد این روش در برآورد فراوانی وحوش روی مجموعه داده‌های موش صحرایی و مجموعه داده‌های خرچنگ نشان داده شده است [۱۱]. اما کاربرد این روش در جوامع انسانی و بر روی داده‌های اپیدمیولوژیکی در سال‌های اخیر، توسعه یافته است و مطالعه حاضر یکی از نادر مطالعاتی است که بر روی داده‌های اپیدمیولوژیکی و برای برآورد کم‌شماری مرگ‌ومیر در کشور انجام می‌شود. در این راستا، بر اساس سه مجموعه داده اپیدمیولوژیکی با میزان هم‌پوشانی کم، متوسط و زیاد و برای برآورد اندازه جمعیت مورد مطالعه، مورد استفاده قرار گرفته است [۷]. همچنین از این روش برای برآورد کم‌شماری در زمینه سرشماری نیز استفاده شده است [۵]. روش مشابهی برای داده‌های اپیدمیولوژیکی و برای برآورد تعداد افراد مبتلا به ویروس هپاتیت A در شمال تایوان و نیز برای برآورد تعداد مبتلایان به دیابت نوع دوم در ایتالیا استفاده شده است. اهداف اصلی این روش ارائه اندازه‌ای برای به کمیت درآوردن هم‌پوشانی اطلاعات و ارائه پارامترهایی برای به کمیت درآوردن منبع وابستگی می‌باشد. سپس می‌توان اندازه جمعیت را از طریق رابطه بین اندازه جمعیت و پوشش نمونه‌ای برآورد شده و پارامترهای وابستگی به‌دست آورد [۱۲]. همچنین کاربرد این روش بر روی مجموعه‌های متنوعی از داده‌های اپیدمیولوژیکی با تعداد منابع مختلف نشان داده شده است [۶]. در این مقاله سعی داریم وضعیت کلی چند نمونه‌ای (t تا) را مورد بحث و بررسی قرار داده و کاربرد آن را روی داده‌های مرگ‌ومیر سال ۱۳۷۹ شهرستان بوشهر نشان دهیم و در پایان، درباره مزایا و محدودیت‌های این روش پشنهادهایی ارائه نماییم.

مواد و روش‌ها

در نخستین قدم با توجه به اهداف مطالعه، گردآوری اطلاعات از طریق اخذ لیست‌های موجود مرگ‌ومیر از سازمان‌های مرتبط با آن صورت گرفت. سپس با توجه به این‌که اطلاعات مرگ‌ومیر مورد استفاده در این مقاله به‌صورت خام بود، لذا قبل از به‌کارگیری این روش روی مجموعه داده‌ها و نتیجه‌گیری، ابتدا اطلاعات مربوط به هر منبع را که بر اساس نام، نام خانوادگی، جنسیت و محل سکونت افراد متوفی بود، رکورد به رکورد وارد رایانه گردید. سپس به منظور یافتن موارد مشترک دو به دو منابع یا هر سه منبع، از بانک اطلاعاتی Access استفاده نمودیم. از آنجا که شیوه‌های متفاوتی در نحوه ثبت اسامی نوزادان فوت شده (اعم از سقط، مرده‌زایی، مرگ نوزادی و...) در هر یک از سه سیستم وجود دارد و با توجه به اینکه نوزادانی که در بدو تولد فوت می‌کنند فاقد شناسنامه می‌باشند، در نتیجه معمولاً اطلاعات مرگ آنها در سیستم ثبت احوال ثبت نمی‌گردد. در ضمن مرگ این

[Downloaded from koomeshjournal.semums.ac.ir on 2024-07-31]

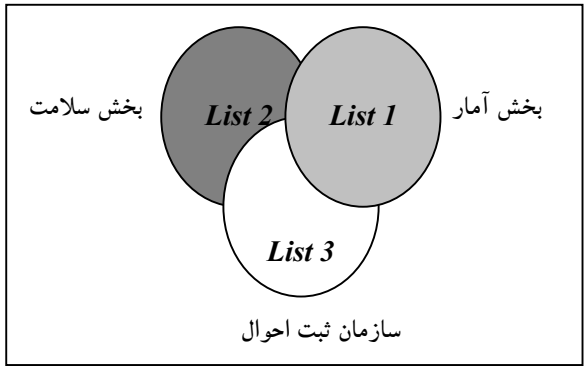
نوزادان در بخش‌های فعال در نظام بهداشت، درمان و آموزش پزشکی نیز با شیوه‌های متفاوتی ثبت می‌گردد. بنابراین تطبیق اسامی نوزادان نتیجه‌ای به همراه نخواهد داشت و از آنجا که تعداد این نوزادان در لیست‌های بخش‌های فعال در نظام بهداشت و درمان تأثیر می‌گذارند؛ ممکن است نتایج را مخدوش سازند و به بیش–برآورد (Over-estimate) بیانجامند.

بنابراین برای برآورد کم‌شماری تعداد مرگ‌ومیر ناشی از این سه سیستم، اسامی نوزادان را از لیست‌های بهداشت، درمان و آموزش پزشکی حذف نموده و سپس به نتیجه‌گیری پرداختیم.

لازم به ذکر است که اطلاعات مرگ‌ومیر بخش سلامت وزارت بهداشت، درمان و آموزش پزشکی عموماً از بیمارستان‌ها و زایشگاه‌ها، پزشکی قانونی، گورستان‌ها، خانه‌های بهداشت، رابطین بهداشت، مراکز بهداشتی درمانی روستایی و مراکز بهداشتی درمانی شهری جمع‌آوری می‌شود و اطلاعات مرگ‌ومیر بخش آمار وزارت بهداشت، درمان و آموزش پزشکی عموماً از طریق گورستان‌های تحت پوشش دانشگاه‌های علوم پزشکی جمع‌آوری می‌گردد و در هر دو این سیستم‌ها نیازی نیست که فرد متوفی حتماً دارای شناسنامه باشد. در صورتی که اطلاعات مرگ‌ومیر سازمان ثبت احوال بر اساس مراجعه بازماندگان فرد متوفی به همراه شناسنامه او به مراکز این سازمان جمع‌آوری می‌گردد. به منظور راحتی کار همان‌طور که در شکل ۱ نشان داده شده است، لیست متوفیات ثبت شده به وسیله بخش توسعه نظام آماری وزارت بهداشت، درمان و آموزش پزشکی را لیست اول و لیست متوفیات ثبت شده به وسیله بخش توسعه شبکه و ارتقاء سلامت وزارت بهداشت، درمان و آموزش پزشکی را لیست دوم و لیست متوفیات ثبت شده به وسیله سازمان ثبت احوال را لیست سوم در نظر گرفته‌ایم. در ضمن در این مقاله عباراتی از قبیل لیست، منبع یا سیستم به نمونه اشاره دارند و ممکن است بر حسب مورد از هر یک استفاده شده باشد.

با توجه به اینکه سه لیست در اختیار داریم، در کل هشت مقدار (هفت مقدار قابل مشاهده $Z_{۱۰۰}$، $Z_{۰۱۰}$، $Z_{۰۰۱}$، $Z_{۱۱۰}$، $Z_{۱۰۱}$، $Z_{۰۱۱}$ و $Z_{۱۱۱}$ و یک مقدار غیرقابل مشاهده $Z_{۰۰۰}$) وجود دارد، که در هر Z_{ijk}، اندیس صفر نشان دهنده عدم حضور فرد متوفی در لیست مربوطه و اندیس یک نشان دهنده حضور فرد متوفی در لیست مربوطه می‌باشد. به عنوان مثال، $Z_{۱۰۰}$ معرف تعداد متوفیانی است که فقط در لیست اول حضور دارند و $Z_{۰۱۱}$ معرف تعداد متوفیانی است که در لیست‌های دوم و سوم حضور دارند ولی در لیست اول حضور ندارند. به همین ترتیب، سایر مقادیر مشاهده شده نیز دارای تفسیرهای مشابهی می‌باشند. مقدار غیرقابل مشاهده $Z_{۰۰۰}$ معرف تعداد متوفیانی است که در هیچ یک از لیست‌ها حضور ندارند.

شکل ۱. تصویری کلی از هم‌پوشانی اطلاعات منابع



زمانی که بر روی یک نمونه جمع بسته می‌شود، اندیس مربوط به آن نمونه به وسیله علامت «+» جایگزین می‌شود. به عنوان مثال؛

$$Z_{+۱۱} = Z_{۰۱۱} + Z_{۱۱۱} \quad \text{یا}$$

$$Z_{++۱} = Z_{۰۰۱} + Z_{۰۱۱} + Z_{۱۰۱} + Z_{۱۱۱}$$

و n_j $\left(j = ۱,۲,۳,\ldots,t\right)$، معرف تعداد افرادی است که در نمونه j اُم حضور دارند. بنابراین به ازای $t = ۳$، داریم:

$n_۱ = Z_{۱++}$، $n_۲ = Z_{+۱+}$ و $n_۳ = Z_{++۱}$.

در این مقاله وضعیتی را در نظر گرفته‌ایم که ناهمگنی میان افراد وجود ندارد و تنها وابستگی بین منابع وجود دارد، یعنی احتمال ثبت یک فرد توسط یک سیستم به سوابق ثبت او توسط سیستم‌های دیگر وابسته است ولی این احتمال در میان

[Downloaded from koomeshjournal.semums.ac.ir on 2024-07-31]

افراد مختلف تفاوتی نمی‌کند. برای وضعیت کلی که چند نمونه (t تا) وجود دارد، برآوردگر پوشش نمونه‌ای، که به وسیله چائو و همکارانش [۱۲،۶] معرفی گردید، عبارت است از:

$$\hat{C} = ۱ - \frac{۱}{t}\sum_{k=۱}^{t}\frac{S_k}{n_k} \qquad (۱)$$

که متوسط نسبت افرادی است که بیش از یک بار توسط سیستم‌های مختلف ثبت شده‌اند. توجه کنید که در رابطه فوق، t عبارت از تعداد لیست‌های مورد استفاده در مطالعه می‌باشد و n_k عبارت از تعداد افرادی است که در لیست k اُم حضور دارند و $S_k = Z_{s_۱ s_۲ \cdots s_t} I\left[s_k = ۱, s_j = ۰, j \neq k\right]$ معرف تعداد افرادی است که فقط در یک نمونه (لیست) حضور دارند. در نتیجه، این برآوردگر، متمم کسر مجموعه‌های تک‌عضوی می‌باشد و بدیهی است که مجموعه‌های تک‌عضوی اطلاعات هم‌پوشانی را در برنمی‌گیرند. اینک D را به‌صورت زیر تعریف می‌کنیم:

$$D = M - \frac{۱}{t}\sum_{k=۱}^{t} S_k \qquad (۲)$$

که در آن $\frac{۱}{t}\sum_{k=۱}^{t} S_k$ متوسط تعداد افرادی است که فقط یک بار توسط منابع ثبت شده‌اند و M عبارت از تعداد کل افراد متمایزی است که توسط منابع ثبت شده‌اند ($M = \sum Z_{ijk} I[i + j + k \succ ۰]$). بنابراین می‌توان D را به عنوان متوسط تعداد افرادی که بیش از یک بار بر روی t لیست حضور یافته‌اند تفسیر نمود.

برآوردگر اندازه جمعیت که به وسیله چائو و همکارانش معرفی گردید، برای وضعیتی کلی که t نمونه وجود دارد از رابطه زیر به‌دست می‌آید:

$$\hat{N} = \left[\frac{D}{\hat{C}} - \frac{۱}{t\hat{C}}\sum\sum_{i<j} A(i,j)\right]\left\{۱ - \frac{۱}{t\hat{C}}\sum_{i<j}\sum\frac{A(i,j)B(i,j)}{n_i n_j}\right\}^{-۱} \qquad (۳)$$

که در رابطه فوق، t عبارت از تعداد لیست‌های مورد استفاده در مطالعه می‌باشد و n_k عبارت از تعداد افرادی است که در لیست k–اُم حضور دارند. مقادیر $\hat{C}$ و D نیز در روابط (۱) و (۲) معرفی گردیدند. عناصر $A(i,j)$ و $B(i,j)$ نیز توابعی از مقادیر مختلف Z_{ijk} می‌باشند و به ازای مقادیر مختلف i و j به‌صورت زیر تعریف می‌شوند:

$$H(i,j) = Z_{k_۱ k_۲ \cdots k_t} I\left[k_i = ۱, k_j = +, k_n = ۰, n \neq i, j\right]،$$
$$A(i,j) = H(i,j) + H(j,i)$$
و $B(i,j) = Z_{k_۱ k_۲ \cdots k_t} I\left[k_i = k_j = ۱, k_n = +, n \neq i, j\right]$

که $I[A]$ نمایانگر تابع نشانگر برای پیشامد A می‌باشد.

با توجه به روابط فوق، $H(i,j)$ عبارت است از تعداد افرادی که در نمونه i اُم حضور دارند ولی در نمونه‌های n اُم $(n \neq i,j)$ حضور ندارند. همچنین $B(i,j)$ عبارت است از تعداد افرادی که در نمونه‌های i اُم و j اُم حضور دارند. $A(i,j)$ نیز عبارت است از مجموع تعداد افرادی که در نمونه i اُم حضور دارند ولی در نمونه‌های n اُم $(n \neq i,j)$ حضور ندارند و تعداد افرادی که در نمونه j اُم حضور دارند ولی در نمونه‌های n اُم $(n \neq i,j)$ حضور ندارند. به عنوان مثال مقادیر $A(i,j)$ و $B(i,j)$ به ازای $(i < j)$ در مورد داده‌های این تحقیق عبارتند از:

$$A(۱,۲) = H(۱,۲) + H(۲,۱) = Z_{۱+۰} + Z_{+۱۰},$$
$$B(۱,۲) = Z_{۱۱+}$$
$$A(۱,۳) = H(۱,۳) + H(۳,۱) = Z_{۱۰+} + Z_{+۰۱},$$
$$B(۱,۳) = Z_{۱+۱}$$
$$A(۲,۳) = H(۲,۳) + H(۳,۲) = Z_{۰۱+} + Z_{۰+۱}$$
$$B(۲,۳) = Z_{+۱۱},$$

اینک با جای گذاری مقادیر $A(i,j)$، $B(i,j)$، $\hat{C}$ و D در رابطه (۳) و به ازای $t = ۳$ ، برآورد اندازه جمعیت در وضعیتی که نمونه‌ها به یکدیگر وابسته‌اند و میزان هم‌پوشانی اطلاعات نسبتاً زیاد است، به صورت زیر به‌دست آورده می‌شود:

[Downloaded from koomeshjournal.semums.ac.ir on 2024-07-31]

$$\hat{N} = \frac{\frac{\left(Z_{+11} + Z_{1+1} + Z_{11+}\right)}{3\hat{C}}}{1 - \frac{1}{3\hat{C}}\left[\frac{\left(Z_{1+0} + Z_{+10}\right)Z_{11+}}{n_1 n_2} + \frac{\left(Z_{10+} + Z_{+01}\right)Z_{1+1}}{n_1 n_3} + \frac{\left(Z_{0+1} + Z_{01+}\right)Z_{+11}}{n_2 n_3}\right]} \quad (4)$$

سرانجام با توجه به رابطه زیر می‌توان تعداد افرادی را که در هیچ یک از لیست‌ها ثبت نشده‌اند، به‌دست آورد:

$$\hat{Z}_{000} = \hat{N} - M \quad (5)$$

به طور کلی برای به‌دست آوردن واریانس برآوردگر فوق و به دنبال آن تشکیل فواصل اطمینان مربوطه، می‌توان از روش ناپارامتری بوت‌استرپ (Bootstrap) استفاده نمود [۹]. همان‌طور که داروش و همکارانش (۱۹۹۳) نشان داده‌اند، فراوانی‌های گیرش در اکثر مواقع دارای توزیع چندجمله‌ای می‌باشند [۸]. یک نمونه بوت‌استرپ $\{Z^*_{000}, Z^*_{100}, \cdots, Z^*_{111}\}$ از یک توزیع چندجمله‌ای با مجموع خانه‌های $\hat{N}$ (حجم نمونه) و احتمالات خانه‌ای $\{\hat{Z}_{000}/\hat{N}, \hat{Z}_{100}/\hat{N}, \cdots, \hat{Z}_{111}/\hat{N},\}$ برای هر خانه قابل مشاهده و احتمال $1 - \sum Z^*_{ijk} I[i+j+k>0]/\hat{N}$ برای خانه مشاهده نشده تولید می‌شود. که در آن؛ $\hat{Z}_{000} = \hat{N} - \sum Z_{ijk} I[i+j+k>0]$ تعداد اعضای مشاهده نشده در هر سه نمونه می‌باشد. در هر تکرار، برآورد بوت‌استرپ $\hat{N}^* - \sum Z^*_{ijk} I[i+j+k>0]$ و $\hat{N}^*$ براساس هفت خانه قابل مشاهده $\{Z^*_{100}, \cdots, Z^*_{111}\}$ محاسبه می‌شود. پس از B مرتبه تکرار، واریانس بوت‌استرپ $\hat{N}$ عبارت است از واریانس نمونه آن B مرتبه برآوردهای اندازه جمعیت. پس از آن که واریانس برآوردگر $\hat{N}$ به‌دست آمد، می‌توان از تبدیل لگاریتمی چائو (۱۹۸۹) استفاده نموده و یک فاصله اطمینان تقریبی برای اندازه جمعیت ارائه نمود.

برنهام در سال ۱۹۸۷ و پس از او، چائو در سال ۱۹۸۹ تبدیل زیر را پیشنهاد نمودند [۴]. آنها فرض کردند تعداد اعضایی که از جامعه گرفته نشده‌اند، یعنی $(\hat{N} - M)$ دارای توزیع لگ-نرمال خواهد بود. به عبارت دیگر، $\log(\hat{N} - M)$ را به عنوان یک متغیر تصادفی با توزیع تقریبی نرمال درنظر گرفتند. با این فرض کران‌های بازه اطمینان به صورت زیر خواهند بود:

$$\left[M + (\hat{N} - M)/c \quad , \quad M + c.(\hat{N} - M)\right] \quad (6)$$

که در آن $(\hat{N} - M)$ بر تعداد اعضای گرفته نشده دلالت دارد و مقدار c از رابطه زیر به‌دست می‌آید.

$$c = \exp\left\{1/96\left[\log\left(1 + \frac{Var(\hat{N})}{(\hat{N} - M)^2}\right)\right]^{\frac{1}{2}}\right\}$$

کسر هم‌پوشانی نسبتاً کوچک به این مفهوم است که مجموعه‌های تک عضوی نسبتاً زیادی وجود دارد، در این حالت به دلیل هم‌پوشانی ناکافی نمی‌توان کم‌شماری را به طور دقیق اندازه‌گیری نمود. در نتیجه، معمولاً برآوردگر موجود در رابطه (۴) دارای خطای معیار بزرگ می‌باشد. سؤال این است که هم‌پوشانی اطلاعات چقدر باید بزرگ باشد؟ برای پاسخ به این سؤال، مطالعات شبیه سازی پیشین به ازای $N = 200$ نشان داده‌اند که میزان هم‌پوشانی اطلاعات باید حداقل ۵۵ درصد باشد [۷،۵]. یک راه‌کار عملی‌تر برای داده‌های وابسته این است که می‌توان از خطای معیار بوت‌استرپ برآوردگر استفاده نمود. به این ترتیب که اگر خطای معیار بوت‌استرپ برآوردگری از یک سوم برآورد اندازه جمعیت بیشتر گردد، در نتیجه این برآورد فاقد ارزش و اعتبار خواهد بود.

نتایج

بدون احتساب مرگ‌ومیر نوزادان در کل ۶۴۱ مورد مرگ از سه منبع: بخش توسعه نظام آماری (۵۸۳ مورد)، بخش

[Downloaded from koomeshjournal.semums.ac.ir on 2024-07-31]

توسعه شبکه و ارتقاء سلامت (۵۹۶ مورد) و سازمان ثبت احوال (۴۴۶ مورد) گزارش گردید. که پس از مقایسه لیست‌های مرگ‌ومیر این سه بخش با یکدیگر، تعداد موارد مشترک دو به دو لیست‌ها و یا هر سه لیست در جدول ۱ نشان داده شده است.

جدول ۱. مقایسه منابع بدون احتساب مرگ‌ومیر نوزادان در سال ۱۳۷۹

منابع	لیست‌های ۱و۲	لیست‌های ۱و۳	لیست‌های ۲و۳	لیست‌های ۱،۲و۳
تعداد موارد مشترک	$Z_{11+}=۵۷۷$	$Z_{1+1}=۴۰۱$	$Z_{+11}=۴۰۴$	$Z_{111}=۳۹۸$

سپس با استفاده از جدول ۱ به محاسبه مقادیر Z_{ijk} (فراوانی‌های گیرش) پرداخته شد، که نتایج آن در جدول ۲ آمده است.

با توجه به این که در هر Z_{ijk}، اندیس صفر نشان‌دهنده عدم حضور فرد متوفی در لیست مربوطه و اندیس یک نشان‌دهنده حضور فرد متوفی در لیست مربوطه می‌باشد؛ بنابراین مقادیر Z_{ijk} را می‌توان با توجه به تعبیری که برای هر Z_{ijk} به ازای (i,j,k)های مختلف وجود دارد، به‌دست آورد و برای محاسبه آنها فرمول بسته‌ای وجود ندارد. به عنوان مثال؛ $Z_{11.}$ معرف تعداد متوفیانی است که در لیست‌های اول و دوم حضور دارند ولی در لیست سوم حضور ندارند. همچنین $Z_{1..}$ معرف تعداد متوفیانی است که فقط در لیست اول حضور دارند. بنابراین طریقه محاسبه مقادیر $Z_{11.}$ و $Z_{1..}$ عبارتند از:

$$Z_{1..}=۵۸۳-(۵۷۷+۴۰۱-۳۹۸)=۳$$ و

$Z_{11.}=۵۷۷-۳۹۸=۱۷۹$ به همین ترتیب، سایر مقادیر مشاهده شده نیز دارای تفسیرها و محاسبات مشابهی می‌باشند.

حال برای برآورد $Z_{...}$ یعنی تعداد متوفیانی که توسط هیچ یک از سازمان‌ها ثبت نگردیده‌اند، ابتدا باید پوشش نمونه‌ای را بر اساس رابطه (۱) برآورد نمود، که در این مطالعه ۹۶/۲ درصد برآورد گردید. با توجه به میزان نسبتاً زیاد پوشش نمونه‌ای برآورد شده، عقیده ما این است که داده‌ها حاوی اطلاعات کافی برای برآورد کم‌شماری می‌باشند. در ضمن متوسط تعداد افرادی که بیش از یک بار بر روی سه لیست ثبت شده‌اند براساس رابطه (۲)، ۶۲۲ نفر برآورد گردید.

جدول ۲. داده‌های مربوط به مرگ‌ومیر شهرستان بوشهر در سال ۱۳۷۹

لیست‌های مرگ‌ومیر			داده‌ها
بخش آمار	بخش سلامت	ثبت احوال	
۱	۱	۱	$Z_{111}=۳۹۸$
۱	۱	۰	$Z_{11.}=۱۷۹$
۱	۰	۱	$Z_{1.1}=۳$
۱	۰	۰	$Z_{1..}=۳$
۰	۱	۱	$Z_{.11}=۶$
۰	۱	۰	$Z_{.1.}=۱۳$
۰	۰	۱	$Z_{..1}=۳۹$
۰	۰	۰	$Z_{...}=?$

اکنون با توجه به این‌که می‌دانیم بین دو منبع (بخش توسعه نظام آماری و بخش توسعه شبکه و ارتقاء سلامت وزارت بهداشت و درمان) وابستگی وجود دارد، بنابراین از برآوردگر ارائه شده در رابطه (۴) استفاده نموده و تعداد متوفیان به اندازه ۶۶۰ نفر و با خطای معیار ۵ نفر برآورد گردید و یک فاصله اطمینان ۹۵ درصدی برای این برآورد با استفاده از رابطه (۶) برابر با (۶۷۴،۶۵۲) به‌دست آمد. در ضمن خطای معیار برآورد شده با استفاده از روش بوت‌استرپ و براساس ۱۰۰۰ تکرار محاسبه شده است. در نهایت با استفاده از رابطه (۵)، میزان کم‌شماری به اندازه ۱۹ نفر، معادلِ ۲/۸۸ درصد برآورد گردید.

بحث

در به‌کارگیری این روش فرض کرده‌ایم که جمعیت مورد مطالعه بسته می‌باشد. یعنی اندازه جمعیت در طول دوره مطالعه بدون تغییر می‌ماند. فرض دیگر این است که تمام مشخصات افراد درون لیست‌ها باید به‌طور صحیح ثبت و تطبیق گردیده باشند. همان‌طورکه توسط بِلک و همکارش نشان داده شده، با وجود این‌که این فرض در اکثر مطالعات اپیدمیولوژیکی بدیهی

[Downloaded from koomeshjournal.semums.ac.ir on 2024-07-31]

می‌باشد، اما رعایت این فرض در اکثر کشورهای در حال توسعه یکی از موانع و مشکلات اساسی استفاده از این روش می‌باشد [۳]. زیرا ممکن است نام یک فرد توسط منابع مختلف به صورت‌های مختلفی ثبت شده باشد. به عنوان مثال، ممکن است نام یک فرد توسط یک سیستم به‌صورت «رحمان علیپور» و در سیستم دیگر به‌صورت «رحمن علی‌پور» ثبت شده باشد. که این دو نام اگرچه هر دو نام یک شخص می‌باشند، لیکن از آنجا که مقایسات توسط رایانه صورت می‌گیرد، رایانه این دو اسم را به عنوان دو شخص مختلف در نظر می‌گیرد. ولی در کشورهای توسعه یافته که افراد دارای کد ملی می‌باشند این مشکل وجود ندارد.

در کشور ما علاوه بر آن‌که، نگارش‌های متفاوتی برای یک اسم وجود دارد، اغلب اسامی دارای پیشوند، پسوند و یا هر دو می‌باشند. که هر یک از این موارد منجر به عدم انطباق فهرست‌ها گردیده و موجب بیش‌شماری در برآوردها می‌گردد. به عنوان نمونه می‌توان به اسامی بتول (بطول)، بیگم (بگم)، رحمان (رحمن)، حمزه (همزه)، آیشه (عایشه)، زنگویی (زنگوئی)، اسماعیل (اسمائیل،اسمعیل)، علینژاد (علی‌نژاد)، حسنپور (حسن‌پور)، ترکزاده (ترک‌زاده) و... که تلفظ یکسان ولی نگارش متفاوت دارند اشاره نمود. لازم به ذکر است، در تحقیق حاضر تناقض‌های نگارشی بسیاری وجود داشت که به منظور جلوگیری از اطاله کلام از ذکر بیشتر این اسامی پرهیز می‌نمائیم. لذا برای برطرف نمودن این مشکل در تحقیق حاضر با صرف وقت بیشتر و چندین‌بار نظارت و بازبینی مجدد لیست‌ها به طور هم‌زمان به نتیجه مطلوب دست یافتیم.

فرض پیچیده‌تر دیگر این است که هر فرد باید شانس ثبت شدن در تمام سه نمونه را به‌طور توأم داشته باشد. یکی دیگر از محدودیت‌های روش پوشش نمونه‌ای این است که برای به‌دست آوردن برآورد دقیق اندازه جمعیت و برای مدل‌بندی وابستگی، میزان هم‌پوشانی اطلاعات باید حداقل ۵۵ درصد باشد؛ در صورتی که میزان هم‌پوشانی اطلاعات کم باشد، یک حد پایین برای نمونه‌هایی با وابستگی مثبت توصیه می‌شود و روشن است که یک حد پایین با دقت بیشتر از یک برآوردگر نقطه‌ای بی‌دقت، قابل استفاده و استناد می‌باشد. مزایای این روش نسبت به روش‌های دیگری که در این زمینه کاربرد دارند از جمله روش مدل‌های لگاریتم خطی این است که در این روش هیچ‌گونه انتخاب مدل و یا مقایسه مدل نیاز نمی‌باشد و نیز با افزایش تعداد لیست‌ها مشکلی به وجود نمی‌آید.

منابع

[۱] حسینی سیدعلی، نشریه آمار مرگ‌ومیر و علل مربوطه در سال ۱۳۷۹، دفتر توسعه و هماهنگی نظام آماری و دانشگاه‌های علوم پزشکی و خدمات بهداشتی درمانی کشور، مرداد ۱۳۸۰، صفحات: ۵۶-۲۵.

[۲] نقوی محسن، سیمای مرگ‌ومیر در ده استان کشور در سال ۱۳۷۹، مرکز توسعه شبکه و ارتقاء سلامت، دی ۱۳۷۹، صفحات: ۷۰-۹۰.

[3] Black JFP, McLarty DG, Capture-recapture technique: difficult to use in developing countries, Br Med J, 1994; 55: 294-301.

[4] Chao A, Estimating population size for sparse data in capture-recapture experiments, Biometrics, 1989; 45: 427-38.

[5] Chao A, Tsay PK, A sample coverage approach to multiple-system estimation with application to census undercount, J Am Stat Assoc, 1998; 93: 283-93.

[6] Chao A, Tsay PK, Shau WY, Chao DY, The application of capture-recapture models to epidemiological data, Stat Med, 2001; 20: 3123-57.

[7] Chao A, Tsay PK, Shau WY, Chao DY, Population size estimation for capture-recapture models with applications to epidemiological data proceeding of Biometrics section, Am Stat Assoc, 1996; 91: 108-17.

[8] Darroch JN, Fienberg SE, Glonek GFV, Junker B, A three-sample multiple-recapture approach to census population estimation with heterogeneous catchability. J Am Stat Assoc, 1993; 88: 1137-48.

[9] Efron B, Tibshirani RJ, An introduction to bootstrap, Chapman and Hall, 1993; New York, p. 101-50.

[10] Good IJ, On the population frequencies of species and the estimation of population parameters, Biometrika, 1953; 40: 237-64.

[Downloaded from koomeshjournal.semums.ac.ir on 2024-07-31]

[11] Lee SM, Chao A, Estimating population size via sample coverage for closed capture-recapture models, Biometrics, 1994; 50: 88-97.

[12] Tsay PK, Chao A, Population size estimation for capture-recapture models with applications to epidemiological data, J Appl Stat, 2001; 28: 25-36.

[Downloaded from koomeshjournal.semums.ac.ir on 2024-07-31]