براورد کم‌شماری مرگ و میر با روش‌های مختلف گیر و بازگیر در شهر گرگان-۱۳۸۱

فاطمه اسلمی[۱]

چکیده

اغلب برای براورد مرگ و میر در یک جامعه‌ی بسته دو روش پیشنهاد می‌شود: روش اول استفاده از مدل‌های مختلف لگ خطی و برازش آن‌ها به داده‌های مرگ می باشد که در آن با استفاده از مدل‌های شبه استقلال می‌توان فرم بسته‌ای برای براورد جمعیت مزبور به دست آورد.

روش دوم استفاده از روش پوشش نمونه‌ای برای براورد میزان مرگ و میر می‌باشد. در این روش همپوشانی اطلاعات نقش مهمی در براورد اندازه‌ی جمعیت و در نهایت براورد کم‌شماری دارد. مفهوم پوشش نمونه‌ای با براورد اندازه‌ی کل جمعیت ارتباط مستقیم دارد یعنی هر چه میزان همپوشانی اطلاعات بیشتر شود، میزان کم شماری کاهش می‌یابد و بالعکس.

در این مطالعه با استفاده از روش‌های مختلف گیر و بازگیر و سیستم‌های ثبت چندگانه میزان کم‌شماری در داده‌های مرگ و میر شهر گرگان در سال ۱۳۸۱ را براورد خواهیم کرد.

برای داده‌های مرگ و میر شهر گرگان در روش لگ خطی میزان کم‌شماری ۲۹ درصد (۳۱۲ مورد کم‌شماری فوت) و در روش پوشش نمونه‌ای میزان کم شماری ۲۲/۴۱ درصد

---

[۱] دانشجوی کارشناسی ارشد آمار زیستی دانشگاه علوم پزشکی شهید بهشتی

(۲۴۰ مورد کم‌شماری فوت) براورد شد. با تعمیم نتایج حاصل به کل کشور، میزان مرگ و میر خام با روش لگ خطی۵/۲۴ در هزار و با روش پوشش نمونه‌ای۵/۰ در هزار براورد می‌شود.

واژگان کلیدی: کم شماری مرگ و میر – روش گیر بازگیر – روش لگ خطی- روش پوشش نمونه‌ای – شهر گرگان

مقدمه

تبیین وضعیت مرگ و میر در جامعه و علل آن در جامعه به منظور شناسایی عوامل خطر و مقابله با عوامل سبب‌ساز آن راهکار اصلی برای افزایش طول عمر و بالا بردن سلامت سلامت انسان‌ها است. اطلاعات مربوط به تعداد موارد فوت کشور از ابعاد متعددی مورد نیاز بوده و کاربرد دارد، از آن جمله: تعیین حق بیمه‌ی زندگی، پیش‌بینی‌های جمعیتی، تعیین امید زندگی، برنامه‌ریزی‌های اقتصادی و اجتماعی و.... وجود آمارهای متعدد و گاه متناقض از سوی سازمان‌های متولی ثبت اطلاعات مرگ و میر به دلیل کم‌شماری[۱] که در این خصوص وجود دارد، موجب عدم اطمینان برنامه‌ریزان و تصمیم‌گیران به این آمارها شده است. به منظور تعیین میزان کم‌شماری و براورد صحیح تعداد متوفیان، روش‌های متعددی وجود دارد که در این مقاله از روش‌گیر و بازگیر[۲] برای براورد آن استفاده می‌کنیم.

روش‌های گیر و بازگیر ابتدا برای براورد تعداد و کم‌شماری در جوامع حیوانی به کار می‌رفته و در طول یک دهه‌ی اخیر استفاده از این روش‌ها در جوامع انسانی به طور چشم‌گیری افزایش پیدا کرده است. امروزه این روش به طور گسترده در بخش سلامت برای براورد اندازه‌ی یک جامعه‌ی پنهان و میزان‌های بروز و شیوع بیماری‌ها و یا پیشامدها به کار می‌رود.

---

۱ *Undercount*

۲ *Capture-Recapture*

برای نخستین بار لاپلاس[۱] در سال ۱۷۸۶ از روش گیر و بازگیر برای براورد تعداد جمعیت جامعه‌ی فرانسه استفاده کرد. به‌رغم این‌که برخی از نویسندگان، قرن ۱۸ را زمان آغاز به کارگیری این روش ذکر کرده‌اند ولی بیشتر نویسندگان، زمان آغاز به کارگیری این روش را با کار پترسن[۲] به رسمیت می‌شناسند.

این روش برای اولین بار در مورد جامعه‌ی غیرانسانی توسط پترسن در سال ۱۸۹۶ برای براورد تعداد کل ماهی‌ها به کار رفته است. هر چند روش او یکی از ساده‌ترین روش‌های گیر و بازگیر است، ولی‌هنوز هم از براوردگر معرفی شده توسط او به عنوان یک براورد اولیه استفاده می‌شود. در سال ۱۹۳۸ خانم اشنی بل[۳]، روش پترسن -- لینکلن را تعمیم داده و پایه‌ی روش گیر و بازگیر چندگانه[۴] را بنا نهاد [1, 6].

در سال‌های اخیر به ویژه در دهه‌ی ۹۰ میلادی، همگام با پیشرفت و ابداع روش‌های مختلف آماری به کارگیری این روش در جوامع انسانی به سرعت گسترش یافت، به طوری که برای مسائل همه‌گیری‌شناسی[۵] در جوامع انسانی‌از این روش زیاد استفاده می‌شود[2, 4]. در این خصوص سکار و دمینگ[۶] (۱۹۴۹) از دو نمونه برای براورد میزان‌های ولادت و مرگ در هند استفاده کردند و جزو پیشگامان به کارگیری روش در جوامع انسانی هستند [1]. وایتس و سیدل[۷] در سال ۱۹۶۸ برای براورد تعداد بیماران بیمارستانی از سیستم ثبت سه نمونه‌ای استفاده کردند [7]. در سال ۱۹۹۴ مقایسه‌ی میزان‌های شیوع تعداد سندروم داون در مکان‌های مختلف توسط راسل و

---

۱ *Laplace*

۲ *Peterson*

۳ *Schnabel*

۴ *Multiple capture-recapture*

۵ *Epidemiology*

۶ *Sekar & deming*

۷ *Wittes & Sidel*

ارتن[۱] انجام شده است که در آن از روش گیر و بازگیر برای براورد تعداد گم شده‌ها در طول فعالیت‌های نظارتی استفاده شده است [11].

هدف از این تحقیق براورد کم‌شماری و تعدیل میزان‌های مرگ شهر گرگان با استفاده از مدل‌های مختلف گیر و بازگیر از جمله مدل لگاریتم خطی[۲] و روش پوشش نمونه‌ای[۳] است.

به منظور اجرای این روش در براورد تعداد موارد فوت شهر گرگان اطلاعات از سه منبع اداره‌ی آمار و اطلاع رسانی وزارت بهداشت، درمان و آموزش پزشکی، معاونت سلامت وزارت بهداشت، درمان و آموزش پزشکی و بالاخره سازمان ثبت احوال کشور، جمع‌آوری و با یافتن موارد مشترک منابع، تحت فرض بسته بودن جامعه به تجزیه و تحلیل داده‌ها پرداخته شده است.

مواد و روش‌ها

بیشترین تحقیق در زمینه‌ی گیر و بازگیر در نیمه‌ی دوم قرن بیستم انجام شده است. در ساده‌ترین روش نمونه‌گیری، n_1 حیوان از جامعه‌ی هدف با اندازه‌ی ناشناخته‌ی N صید می‌شود و سپس این حیوانات را علامت‌گذاری کرده و رها می‌کنند. پس از مدت زمان لازم برای اختلاط حیوانات علامت دار با سایر اعضای جامعه - که نباید این دوره طولانی شود - یک نمونه‌ی ثانویه به حجم n_2 از این جامعه استخراج می‌شود. این نمونه شامل اعضای علامت‌دار و بدون علامت می‌باشد.

بدیهی است که اگر در نمونه‌ی دوم اعضای علامت‌دار و غیر علامت دار را دوباره علامت زده و به جامعه برگردانیم و این کار را K بار انجام دهیم، براوردگر بهتر و

---

[۱] *Russel & orton*

[۲] *Log-Linear model*

[۳] *Coverage method*

دقیق‌تری برای N به دست می‌آید. این روش را روش گیر و بازگیر چند مرحله‌ای (چندگانه) یا آزمایش اشنی بل گویند.

مدل‌های گیر و بازگیر در جوامع انسانی تحت عنوان" سیستم ثبت چندگانه"[۱] به کار می‌رود. در این روش هر فهرست را به عنوان یک نمونه (دام) در نظر گرفته و اسامی ثبت شده در فهرست‌ها را به عنوان مشخصه و برچسب منحصر به فرد در نظر می‌گیریم، این وضعیت شبیه وضعیت گیر و بازگیر برای براورد فراوانی حیوانات خواهد بود. بنابراین در این روش برای براورد تعداد افراد دارای ویژگی موردنظر، نیازی به صید، علامت‌گذاری و بازصید نمی‌باشد، بلکه فهرست‌ها و افرادی که در هر دو فهرست حضور دارند، به عنوان گیر و بازگیر در نظر گرفته می‌شوند. بنابر این "گیرش در یک نمونه" در یک مطالعه‌ی گیر و بازگیر چندگانه با "افراد ثبت شده در یک فهرست"مطابقت دارد و "احتمال گیرش" با "احتمال تحقیق"یا "احتمال تعیین هویت کردن" سازگار است. همچنین در این مطالعه " نمونه" به فهرست یا منبع اطلاق می‌شود.

- مدل پترسن - لینکلن (دو نمونه‌ای)

براوردگر پترسن به روش‌های مختلف به دست می‌آید. یکی از این روش‌ها، روش " استدلال تناسبی"[۲] است. همان‌طور که قبلاً گفته شد در روش گیر و بازگیر دو نمونه‌ای پس از رها کردن n_1 حیوانات علامت‌گذاری شده به جامعه، یک نمونه‌ی ثانویه به حجم n_2 از جامعه استخراج می‌شود. فرض کنید تعداد اعضای علامت‌دار در این نمونه m2 باشد به طور شهودی انتظار می‌رود نسبت اعضای دارای علامت در نمونه‌ی دوم یعنی $\frac{m_2}{n_2}$ برابر با نسبت اعضای دارای علامت در کل جامعه یعنی $\frac{n_1}{N}$ باشد. حال با مساوی قرار دادن این دو نسبت و حل آن بر حسب N خواهیم داشت:

---

[۱] *Multiple-record system*

[۲] *Proportionality arguments*

(۲-۱)

$$\frac{m_2}{n_2} = \frac{n_1}{N} \quad \rightarrow \quad N = \frac{n_1 \times n_2}{m_2}$$

به آزمایش ساده‌ی یاد شده آزمایش پترسن و به براوردگر آن "براوردگر پترسن"یا "شاخص لینکلن"و به این روش، "گیر و بازگیر دو مرحله‌ای"یا "بازگیرش منفرد" می‌گویند.

اعتبار براورد آزمایش پترسن – لینکلن وابسته به مفروضات اساسی زیر می‌باشد:

- *جامعه تحت مطالعه بسته*[۱] *باشد*: در طول فرایند نمونه‌گیری تغییراتی به واسطه‌ی تولد، مرگ و مهاجرت وجود ندارد و به عبارت دیگر شانس انتخاب شدن یک واحد در هر بار نمونه‌گیری تقریباً ثابت فرض شده است.

- *استقلال*[۲] : نمونه‌ها، مستقل از یکدیگر می‌باشند. یعنی احتمال انتخاب شدن یک نمونه در یک مرحله مستقل از احتمال انتخاب یا عدم انتخاب در مراحل دیگر است.

- همگنی[۳] : همه‌ی اعضا در جامعه‌ی تحت مطالعه، احتمال‌های یکسان برای مشاهده شدن در هر نمونه را داشته باشند.

- *تطبیق کامل*[۴] : اعضای مشاهده شده در یک مرحله می‌توانند کاملاً با مرحله‌ی دیگر تطبیق داده شوند.

از مزایای روش پترسن – لینکلن می‌توان به سادگی محاسبات آن اشاره نمود. از محدودیت‌های اصلی این روش که جزو نقاط ضعف این روش نیز به حساب می‌آید،

---

[۱] *Closed population*

[۲] *Independence*

[۳] *Homogeneity*

[۴] *Complete matching*

شرط استقلال نمونه‌ها از یکدیگر است، ولی با وجود این استفاده از این براورد به عنوان یک براورد اولیه و مقدماتی بسیار مفید می‌باشد.

همان طور که گفته شد براوردگر پترسن - لینکلن به صورت زیر می‌باشد:

$$N_p = \frac{n_1 \times n_2}{m_2}$$

که Np براورد N با روش پترسن می‌باشد. این براوردگر یک براوردگر اریب است (اگر $m_2 = 0$ باشد، آن‌گاه $Np = \infty$ خواهد شد).

چپمن[1] در سال ۱۹۵۱ نشان داد که این براوردگر برای جامعه "بهترین براوردگر مجانباً نرمال" است، ولی برای نمونه‌های کوچک دارای اریبی[2] بالایی می‌باشد و به بیش‌براورد[3] پارامتر N خواهد انجامید. به همین خاطر وی براوردگر اصلاح شده‌ی زیر را که دارای اریبی کمتری است به جای براوردگر (۱-۲) پیشنهاد می‌کند.

(۲-۲)

$$N_{ch} = \frac{(n_1 + 1)(n_2 + 1)}{(m_2 + 1)} - 1$$

با استفاده از روش براورد "حداکثر درستنمایی" نیز همین براوردگر به دست می‌آید.

پژوهشگاه علوم انسانی و مطالعات فرهنگی

- مدل‌های لگ خطی

در این روش ابتدا یک جدول t بعدی ناقص (t تعداد فهرست‌ها می باشد)؛ شامل $2^t - 1$ خانه مشاهده شده و یک خانه‌ی مشاهده نشده تشکیل می‌شود. در یک گیر و بازگیر سه نمونه‌ای می‌توانیم داده‌های به دست آمده از سه نمونه (فهرست) را در یک جدول سه بعدی ناقص شامل هشت خانه مطابق با جدول ۱-۲ مرتب کنیم.

---

[1] *Chapman*

[2] *Bias*

[3] *overestimate*

جدول ۱ -۲: فراوانی‌های قابل مشاهده در وضعیت گیر و بازگیر سه نمونه ای

		نمونه‌ی اول			
		حضور		عدم حضور	
		نمونه‌ی دوم		نمونه‌ی دوم	
		حضور	عدم حضور	حضور	عدم حضور
نمونه‌ی	حضور	X_{111}	X_{121}	X_{211}	X_{221}
سوم	عدم حضور	X_{112}	X_{122}	X_{212}	X_{222}

که در هر x_{ijk} اندیس ۲ نشان دهنده‌ی عدم حضور در نمونه و اندیس ۱ نشان دهنده‌ی حضور در نمونه‌ی متناظر می باشد. به عنوان مثال X_{121} بیان کننده فراوانی افرادی است که در نمونه‌ی (فهرست) اول و سوم حضور دارند ولی در نمونه‌ی (فهرست) دوم حضور ندارند.

سپس مدل‌های لگ خطی مختلفی به این جدول ناقص برازش داده و آن‌گاه مناسب‌ترین مدل را انتخاب کرده و با این فرض که اثر متقابل سه طرفه وجود ندارد، خانه‌ی مشاهده نشده‌ی X_{222} را براورد و از آن‌جا براورد، اندازه‌ی کل جمعیت به دست آورده می‌شود.

در این روش معیار انتخاب بهترین مدل، معیار AIC[1] می‌باشد.

$$AIC = G^2 - 2df$$

$$G^2 = 2\sum x_{ijk}\ Log\left(\frac{x_{ijk}}{m_{ijk}}\right)$$

[1] *Akaik Information Criterion*

که در آن G^2 آماره نسبت درستنمایی[1] و df درجه‌ی آزادی آن می‌باشد.

- روش پوشش نمونه‌ای

تعریف کلی پوشش نمونه‌ای برای یک نمونه‌ی مفروض عبارت است از نسبتی (درصدی) از جمعیت که در آن نمونه قرار گرفته است. در این روش با استفاده از همپوشانی اطلاعات به براورد کم‌شماری و در نهایت براورد اندازه‌ی کل جمعیت می‌پردازیم. به طور کلی مفهوم پوشش نمونه‌ای با براورد اندازه‌ی کل جمعیت در ارتباط می‌باشد. یعنی هر چه میزان همپوشانی اطلاعات بیشتر شود، میزان کم‌شماری کاهش می‌یابد و بالعکس. علاوه بر این، روش پوشش نمونه‌ای را می‌توان در هر وضعیتی حتی در وضعیت عدم استقلال نمونه‌ها از یکدیگر نیز به خوبی استفاده کرد. پس در این روش اندازه‌ای کمّی برای همپوشانی اطلاعات و پارامترهایی برای منبع وابستگی براورد می‌شود و در نتیجه می‌توان اندازه‌ی جمعیت را از طریق رابطه‌ی آن با پوشش نمونه‌ای براورد شده و پارامترهای وابستگی به دست آورد.

روش پوشش نمونه‌ای به عنوان یک روش جایگزین روش‌های یاد شده مطرح می‌باشد، که برای وارد نمودن منبع وابستگی در براورد از همپوشانی اطلاعات استفاده می‌کند. مزایای مهم این روش به شرح ذیل می‌باشند:

الف) هیچ گونه انتخاب مدل و یا مقایسه‌ی مدل نیاز نمی‌باشد.

ب) وقتی که تعداد فهرست‌ها افزایش می‌یابد، هیچ‌گونه مشکلی به وجود نمی‌آید.

یکی از محدودیت‌های روش پوشش نمونه‌ای این است که برای به دست آوردن براورد دقیق اندازه‌ی جمعیت و برای مدل‌بندی وابستگی، میزان همپوشانی اطلاعات در فهرست‌ها حداقل باید ۵۵ درصد باشد.

---

[1] *Likelihood-ratio statistic*

نتایج

به کارگیری مدل‌های گیر و بازگیر در مطالعات مربوط به جوامع انسانی با چندین فرض اساسی همراه است که عبارتند از:

در به کارگیری این روش‌ها و مدل‌ها، فرض کرده‌ایم که جمعیت مورد مطالعه بسته است. یعنی اندازه‌ی جمعیت در طول دوره‌ی مطالعه بدون تغییر می‌ماند. همچنین باید تمام مشخصات (علائم شناسایی) افراد درون فهرست‌ها به طور صحیح ثبت و تطبیق شده باشند.

فرض دیگر این است که احتمال توأم گیرش برای هر عضو در تمام فهرست‌ها باید بزرگتر از صفر باشد (یعنی هر فرد باید شانس ثبت شدن در تمام سه نمونه را به طور توأم داشته باشد) تا این‌که بتوانیم همپوشانی اطلاعات را به دست آوریم. بنابراین هر فرد باید شانس ثبت شدن در هر فهرست را داشته باشد و عدم ثبت صرفاً از نوع یک صفر تصادفی (گم شدن به دلیل شانس کم) باشد نه یک صفر ساختاری (گم شدن به دلیل غیرممکن بودن).

برای براورد تعداد مرگ‌های مربوط به شهر گرگان در سال ۱۳۸۱سه روش مختلف را به کار برده‌ایم. فهرست متوفیان ثبت شده به وسیله‌ی اداره‌ی آمار و اطلاع رسانی وزارت بهداشت، درمان و آموزش پزشکی را فهرست اول و فهرست متوفیان ثبت شده به وسیله‌ی معاونت سلامت وزارت بهداشت، درمان و آموزش پزشکی را فهرست دوم و فهرست متوفیان ثبت شده به وسیله‌ی سازمان ثبت احوال کشور را فهرست سوم در نظر گرفته‌ایم.

بدون احتساب مرگ و میر نوزادان در کل ۱۰۷۱ مورد مرگ از سه منبع گزارش شده است که در آن اداره‌ی آمار و اطلاع‌رسانی تعداد ۷۷۵ مورد، معاونت سلامت ۵۵۷ مورد و سازمان ثبت احوال ۶۶۹ مورد گزارش کرده‌اند. پس از مقایسه‌ی فهرست‌های مرگ و میر این سه بخش با یکدیگر و به دست آوردن تعداد موارد مشترک آن‌ها تعداد کل مرگ‌ها براورد می‌شود.که برای انجام محاسبات از نرم افزار SAS استفاده شده است.

- نتایج مقایسه‌ی فهرست‌ها با یکدیگر

همان طور که قبلاً نیز اشاره شد، یکی از مفروضات اصلی برای معتبر بودن این براوردگرها استقلال نمونه‌ها از یکدیگر است. در اینجا بدون در نظر گرفتن این فرضیه نتایج حاصل در جدول ۱-۳ آمده است.

جدول ۱-۳: نتایج به دست آمده از مقایسه‌های دو به دوی نمونه‌ها با یکدیگر

نمونه‌ها	براوردگر پترسن	براوردگر چپمن	خطای معیار چپمن	فاصله‌ی اطمینان ۹۵درصد براوردگر چپمن	
				کران بالا	کران پایین
فهرست‌های ۱و۲	۱۲۳۳	۱۲۳۳	۳۰	۱۳۰۲	۱۱۸۴
فهرست‌های ۱و۳	۹۷۱	۹۷۱	۱۱	۹۸۹	۹۴۸
فهرست‌های ۲و۳	۱۲۴۲	۱۲۴۱	۳۶	۱۳۲۷	۱۱۸۴
فهرست‌های ۱و۳ ادغام شده و۲	۱۲۸۰	۱۲۷۹	۲۶	۱۳۳۷	۱۲۳۵

بر اساس نتایج جدول ۱-۳ براوردهای قابل توجیهی از روی فهرست‌های (۱و۲) و (۲و۳) برای تعداد کل فوت‌شدگان به دست می‌آید.

ولی با توجه به این‌که در کل ۱۰۷۱ مورد مرگ متمایز توسط این سه منبع گزارش شده است، لذا براورد فهرست‌های ۱و۳ یعنی ۹۷۱ مورد، قابل پذیرش نمی‌باشد (یعنی مقدار براورد شده از تعداد فوت ثبت شده نیز کمتر است) و این

می‌تواند حاکی از وابستگی بین این دو منبع باشد. از این رو می‌توان از اطلاعات این تجزیه و تحلیل مقدماتی برای مدل‌بندی مدل‌های پیچیده‌تر استفاده نمود. در ادامه دو فهرست اول و سوم (بخش آمار و سازمان ثبت احوال) را همان طور که وایتس و همکارانش پیشنهاد دادند [7] با هم ادغام کرده و یک نمونه‌ی جدید را تشکیل می‌دهیم و سپس نمونه‌ی جدید را با نمونه‌ی دوم (معاونت سلامت) مقایسه می‌کنیم. با انجام این کار تعداد مرگ‌ها ۱۲۸۰ به دست می‌آید که براورد قابل توجیهی است. براوردهای مرگ و میر از فهرست‌های ۱ و ۲ و فهرست‌های ۱ و ۲ که به ترتیب برابر با ۱۲۳۲ و ۱۲۴۲ می‌باشد تقریباً نزدیک به‌هم است. در حالت ادغام فهرست‌های ۱ و ۳ اختلاف محسوسی در تعداد مرگ‌ها مشاهده می‌شود.

- نتایج اجرای مدل‌های لگاریتم خطی

با استفاده از نتایج به دست آمده از انطباق فهرست‌ها با یکدیگر یک جدول سه طرفه‌ی ناقص (شامل هفت خانه‌ی قابل مشاهده و یک خانه‌ی غیرقابل مشاهده، جدول ۱) تشکیل می‌شود. سپس مدل‌های لگ خطی مختلف به خانه‌های مشاهده شده‌ی جدول برازش داده می‌شود، آن‌گاه با توجه به معیار AIC مناسب‌ترین مدل انتخاب می‌شود. سپس با فرض نبودن اثر متقابل سه تایی، مدل انتخاب شده برای براورد خانه‌ی مشاهده نشده استفاده می‌گردد و از آن‌جا براورد اندازه‌ی کل مرگ‌ها به دست می‌آید.

نتیجه‌ی حاصل در جدول۲-۳ نشان داده شده است.

جدول ۲-۳: برازش مدل‌های لگ خطی بر روی داده‌های مرگ شهرستان گرگان

مدل‌های لگاریتم خطی	معیار AIC	درجه‌ی آزادی	براورد کل مرگ‌ها	خطا معیار مجانبی براورد	فاصله‌ی اطمینان ۹۵درصد	
					کران بالا	کران پایین
اشباع شده	۰	۰	۱۴۹۸	۱۱۰	۱۷۷۲	۱۳۳۱
(۱۲-۱۳)	۲/۳	۱	۱۳۸۳	۶۴	۱۵۳۵	۱۲۸۰
(۱۲-۲۳)	۱۳۶/۳۱	۱	۱۱۱۷	۱۰	۱۱۴۰	۱۱۰۲
(۱۳-۲۳)	۶/۰۷	۱	۱۳۱۴	۴۰	۱۴۰۷	۱۲۴۷
(۱۲-۳)	۴۸/۸۳	۲	۱۱۳۳	۱۱	۱۱۵۸	۱۱۱۵
(۱۳-۲)	۵/۸۹	۲	۱۲۸۰	۲۶	۱۳۳۷	۱۲۳۵
(۲۳-۱)	۱۵۲/۰۴	۲	۱۱۳۹	۱۱	۱۱۶۵	۱۱۲۱
مستقل	۱۵۹/۱۲	۳	۱۱۰۵	۷	۱۱۲۱	۱۰۹۲

در این جدول منظور از مدل اشباع شده این است که آثار اصلی و آثار متقابل دوتایی در مدل حضور دارند. همچنین به عنوان نمونه منظور از مدل ۱۳-۱۲ مدلی است که فاقد اثر متقابل دو تایی (۳ و۲) می باشد و بقیه‌ی آثار (شامل آثار اصلی و آثار متقابل دیگر به جز مورد یاد شده) را دارد و یا مدل ۳-۱۲ مدلی است که فقط شامل آثار اصلی و اثر متقابل دو تایی (۲ و ۱) می‌باشد. مدل مستقل نیز مدلی است که فقط شامل آثار اصلی باشد. سایر مدل‌ها نیز به طور مشابه قابل تفسیر است.

پس از برازش مدل‌های لگاریتم خطی مختلف به خانه‌های مشاهده شده که در جدول۲-۳ نشان داده شده است در نهایت مدل ۱۳-۱۲ به عنوان مناسب‌ترین مدل قابل توجیه انتخاب می‌شود (یعنی مدلی که در آن نمونه‌ی اول و سوم مستقل هستند). این مدل دارای دو ویژگی مهم می‌باشد، اول اینکه مقدار معیار آکائیک این مدل کوچک بوده و (AIC = ۲/۰۳) دوم، خطای براورد آن نیز کم است (Se=۶۴) در نتیجه تحت این

مدل تعداد کل متوفیان ۱۳۸۳ نفر با خطای معیار ۶۴ نفر برآورد شده و فاصله‌ی اطمینان ۹۵ درصد این برآورد (۱۵۲۵ و ۱۲۸۰) به دست می‌آید.

در نتیجه میزان کم‌شماری (تعداد متوفیان ثبت نشده) به اندازه‌ی ۳۱۲ مورد، معادل با ۲۹درصد به دست آمد.

- نتایج اجرای روش پوشش نمونه‌ای

در این روش با توجه به این‌که می‌توان پوشش نمونه‌ای را در هر وضعیتی حتی در وضعیت عدم استقلال نمونه‌ها از یکدیگر نیز به خوبی برآورد نمود، ابتدا میزان پوشش نمونه‌ای را برآورد می‌نماییم، تا این‌که مشخص شود از کدام روش باید برای برآورد اندازه‌ی جمعیت استفاده نمود. در جدول ۳-۳ مقدار $\hat{N}_0$ در وضعیتی که فرض استقلال بین نمونه‌ها برقرار باشد، به کار می‌رود و مقدار $\hat{N}$ در شرایطی به کار می‌رود که حداقل بین یک جفت نمونه وابستگی وجود داشته و میزان همپوشانی اطلاعات نیز نسبتاً زیاد باشد (بیش از ۵۵درصد). مقدار $\hat{N}_1$ نیز در شرایطی به کار می‌رود که دست‌کم بین یک جفت نمونه وابستگی وجود داشته و میزان همپوشانی اطلاعات کم باشد (کمتر از ۵۵درصد) و یا خطای برآورد از یک سوم برآورد اندازه‌ی جمعیت تجاوز نماید.

همان طور که در جدول ۳-۳ ملاحظه می‌شود، میزان پوشش نمونه‌ای در این مطالعه ۷۹/۷ درصد برآورد شده است. لذا با توجه به این‌که میزان همپوشانی اطلاعات نسبتاً زیاد می‌باشد، می‌توان انتظار داشت که داده‌ها حاوی اطلاعات کافی برای برآورد کم‌شماری خواهند بود. در ضمن متوسط تعداد افرادی که بیش از یک بار توسط منابع ثبت شده اند ۹۳۹ نفر برآورد شده است.

با توجه به اینکه دست کم بین یک جفت نمونه وابستگی وجود دارد و میزان همپوشانی اطلاعات نیز نسبتاً زیاد می‌باشد، لذا از مقدار $\hat{N}$ جدول ۳-۳ به عنوان برآورد اندازه‌ی جمعیت استفاده می‌کنیم. در نتیجه تعداد کل فوت‌شدگان به اندازه‌ی ۱۳۱۱ نفر و با خطای معیار ۱۵ نفر برآورد شد و فاصله‌ی اطمینان ۹۵ درصدی برای این

براورد (۱۳۲۱، ۱۲۹۸) به دست آمد. در ضمن، میزان کم‌شماری به اندازه‌ی ۲۴۰ مورد، معادل ۱۸/۳۱ درصد براورد شد.

جدول ۳-۳ : نتایج به دست آمده از روش پوشش نمونه‌ای

براوردهای N	M	D	C	براورد کل مرگ‌ها	خطای معیار	فاصله اطمینان ۹۵درصد	
						کران بالا	کران پایین
$\hat{N}_0$	۱۰۷۱	۹۳۹	۰/۷۹۷	۱۱۷۸	۹	۱۱۸۲	۱۱۷۰
$\hat{N}$	۱۰۷۱	۹۳۹	۰/۷۹۷	۱۳۱۱	۱۵	۱۳۲۱	۱۲۹۸
$\hat{N}_1$	۱۰۷۱	۹۳۹	۰/۷۹۷	۱۲۷۳	۱۶	۱۲۸۴	۱۲۶۲

M در جدول۳-۳ تعداد افراد متمایز ثبت شده به وسیله‌ی سازمان‌های مختلف و **D** متوسط تعداد افراد ثبت شده به وسیله‌ی سازمان‌های مختلف و **C** براورد پوشش نمونه‌ای می باشد.

در این جدول خطای معیار براورد اندازه‌ی جمعیت بر اساس هزار تکرار بوت استرپ[1] به دست آمده و ممکن است برای این خطا در آزمایشهای مختلف مقادیر متفاوتی بدست آید. $\hat{N}_0$ براورد اندازه‌ی جمعیت در وضعیتی است که منابع از یکدیگر مستقل می‌باشند، $\hat{N}$ براورد اندازه‌ی جمعیت در وضعیتی است که بین منابع وابستگی وجود دارد و میزان همپوشانی بین منابع به اندازه‌ی کافی بزرگ می‌باشد (دست‌کم ۵۵درصد) و $\hat{N}_1$ براورد یک مرحله‌ای اندازه‌ی جمعیت برای پوشش نمونه‌ای کم می‌باشد. استفاده از این براورد هنگامی توصیه می‌شود که خطای معیار براورد شده نسبتاً بزرگ باشد.

---

[1] *Bootstrap*

بحث و نتیجه‌گیری

همان‌طور که قبلاً گفته شد در به کارگیری این مدل‌ها فرض بر بسته بودن جمعیت مورد مطالعه می‌باشد، یعنی اندازه‌ی جمعیت در طول دوره‌ی مطالعه بدون تغییر باقی می‌ماند. همچنین فرض شده است که تمامی مشخصات افراد به طور کامل ثبت و منطبق شده است.

با روش لگاریتم خطی میزان کم‌شماری مرگ و میر در شهر گرگان ۲۹ درصد به دست آمد که با تعمیم آن به کل کشور، با فرض عدم تفاوت میان میزان مرگ و میر در شهر و روستا، میزان مرگ در کشور ۵/۲۴ در هزار براورد می‌شود.

با روش پوشش نمونه‌ای میزان کم‌شماری فوت در شهر مذکور ۲۲ درصد براورد شده که با تعمیم آن به کل کشور، میزان مرگ و میر ۵ در هزار برای کشور براورد می‌شود.

طی سال‌های اخیر سازمان‌های مرتبط با موضوع جمعیت، هر یک به طور جداگانه براورد شاخص مرگ و میر در کشور را ارائه نموده‌اند و ناسازگاری‌های بسیار بین میزان‌های مرگ و میر اعلام شده توسط این سازمان‌ها مشکلاتی در ارائه‌ی اطلاعات پایه برای کشور ایجاد کرده است.

با طرح این مسئله در شورای عالی آمار، مسئولیت ساماندهی آمارهای مرگ و میر کشور به مرکز آمار ایران سپرده شد که در راستای آن مرکز آمار ایران با اجرای راهکار کوتاه مدت و با روش‌های غیرمستقیم میزان مرگ و میر در کشور را برابر با ۵/۵ براورد کرد که به عنوان مبنا برای کل کشور در نظر گرفته شده است.

همچنین میزان مرگ و میر توسط بخش جمعیت سازمان ملل ۵/۵ در هزار تخمین زده شده است. [۳]

همانگونه که ملاحظه می‌شود میزان مرگ و میر براورد شده با استفاده از روش‌های ارائه شده در این مقاله به میزان مرگ و میر ارائه شده توسط سازمان ملل نزدیک می‌باشد. شایان ذکر است که در این بررسی از اطلاعات شهر گرگان استفاده شده و

گزارش سازمان ملل مربوط به کل کشور است و به نظر می رسد بالاتر بودن آن براورد به دلیل در نظرگرفتن مناطق روستایی در براورد باشد.

بنابر نتایج حاصل می‌توان پیشنهاد نمود که روش به کار رفته در این مقاله برای سایر استان‌ها و کل کشور قابل اجرا و همچنین نتایج آن مقرون به واقعیت می‌باشد.

امید است که نتایج به دست آمده در این مقاله بتواند پاسخگوی کمبود اطلاعات در زمینه شاخص‌های پایه‌ای مرگ و میر کشور باشد.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

فهرست منابع:

۱) زنجانی حبیب ا... ، نوراللهی طه، جداول مرگ و میر ایران برای‌سال۱۳۷۵، مؤسسه‌ی پژوهش تأمین اجتماعی، تهران ۱۳۷۹، صفحه ۳۷.

۲) ترجمه آیت‌اللهی، سنفی، عباسی، رضویان، روش‌های تحلیل جمعیتی، انتشارات دانشگاه شیراز.

۳) سالنامه‌ی آماری کشور، سال ۱۳۸۱، مرکز آمار ایران.

۴) Yvonne M.M.Bishop, Stephen E. Fienberg and paul. W Holland, Discrete Multivariate Analysis : Theory and practice, Chapter 6, 229, The MIT press Cambridge, Massachusetts and London.

۵) Ceorge A. F. Seber, Capture-Recapture: Before and after Euring, Department of statistics and Actuarial science simon fraser university Burnaby,2000, BC, Canada.

۶) International Working Group for Disease Monitoring and Forecasting, Capture-recapture and multiple-record systems estimation II: applications in human diseases, Am J Epidemiol, 1995, 142, 1059-68.

۷) Seber GF, The Estimation of Animal Abundance and Related Parameters, 2nd edition, 1982, Griffin, London.

۸) Wittes J, Colton T, Sidel VW, A generalisation of the simple capture-recapture model with applications to epidemiological research, J Chronic Disease, 1968, 21, 287-301.

۹) Fisher N, Turner SW, Pugh R, Taylor C, Estimating numbers of homeless and homeless mentally ill people in north east Westminster by using capture-recapture analysis, BMJ, 1994, 308, 27-30.

۱۰) LaPorte RE, Assessing the human condition: capture-recapture techniques, Br Med , 1994, 308, 5-6.

۱۱) Giovanni Corrao, Vincenzo Bagnardi, Giovanni Vittadini, Sergio Favilli, Capture-recapture methods to size alcohol related problems in a population, J Epidemiol Community Health, 2000, 54, 603-610 .

۱۲) Evens, M.A. Bonett, D.G & Mcdonal L.L, A general Theory for Modeling Capture -recapture data from a closed population, Biometrics , 1997, 50 , 396 - 405.

۱۳) Cormak.R.M, log linear models for capture-recapture, Biometrics, 1989, 45, 395 - 413.

۱۴) S.T.Buckland, I. B. J. Goudie, and D.L Borchers, wildlife population Assessment: past Development and Future Directions, Biometrics, 2000, 56, 1-12.

۱۵) Chao, A . Tsay, P K. Lin, S-H . Shau, W-Y. Chao, D-Y, Tutorial in biostatistics - The applications of capture-recapture models to epidemiological data. Statistics in Medicine, 2001, 20, 3123 – 3157.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی