

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

فاطمه اسلمی^۱

چکیده

اغلب برای برآورد 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 می‌باشد که در آن با استفاده از مدل‌های شبه استقلال می‌توان فرم بسته‌ای برای برآورد جمعیت مزبور به دست آورد.

روش دوم استفاده از روش پوشش نمونه‌ای برای برآورد میزان مرگ و میر می‌باشد. در این روش همپوشانی اطلاعات نقش مهمی در برآورد اندازه‌ی جمعیت و در نهایت برآورد کم‌شماری دارد. مفهوم پوشش نمونه‌ای با برآورد اندازه‌ی کل جمعیت ارتباط مستقیم دارد یعنی هر چه میزان همپوشانی اطلاعات بیشتر شود، میزان کم‌شماری 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 برآورد 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 کم‌شماری ۲۲/۴۱ درصد

^۱ دانشجوی کارشناسی ارشد آمار زیستی دانشگاه علوم پزشکی شهید بهشتی

برآورد کم‌شماری مرگ و میر با ... ————— گزیده مطالب آماری - ۶۳

(۲۴۰ مورد کم‌شماری فوت) برآورد شد. با تعمیم نتایج حاصل به کل کشور، میزان مرگ و میر خام با روش لگ خطی ۵/۲۴ در هزار و با روش پوشش نمونه‌ای ۵/۰ در هزار برآورد می‌شود.

واژگان کلیدی: کم‌شماری 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 برآورد صحیح تعداد متوفیان، روش‌های متعددی وجود دارد که در این مقاله از روش گیر و بازگیر^۲ برای برآورد آن استفاده می‌کنیم.

روش‌های گیر و بازگیر ابتدا برای برآورد 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 برآورد 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 برآورد 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 برآورد تعداد کل ماهی‌ها به کار رفته است. هر چند روش او یکی از ساده‌ترین روش‌های گیر و بازگیر است، ولی هنوز هم از برآوردگر معرفی شده توسط او به عنوان یک برآورد 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 انسانی از این روش زیاد استفاده می‌شود [2, 4]. در این خصوص سکار و دمینگ^۶ (۱۹۴۹) از دو نمونه برای برآورد 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 برآورد 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

برآورد کم‌شماری مرگ و میر با ... ————— گزیده مطالب آماری - ۶۳

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

هدف از این تحقیق برآورد 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 برآورد تعداد موارد فوت شهر گرگان اطلاعات از سه منبع اداره‌ی آمار و اطلاع‌رسانی 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 علامت‌دار 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 علامت‌دار را دوباره علامت زده و به جامعه برگردانیم و این کار را K بار انجام دهیم، برآوردگر بهتر و

^۱ Russel & orton

^۲ Log-Linear model

^۳ Coverage method

گزیده مطالب آماری - ۶۳ ————— برآورد کم‌شماری مرگ و میر با ...

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 برآورد فراوانی حیوانات خواهد بود. بنابراین در این روش برای برآورد 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)

برآوردگر 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 m_2 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)

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

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

اعتبار برآورد 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

گزیده مطالب آماری - ۶۳ ————— برآورد کم‌شماری مرگ و میر با ...

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

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

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

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

چپمن^۱ در سال ۱۹۵۱ نشان داد که این برآوردگر برای جامعه "بهترین برآوردگر مجانباً نرمال" است، ولی برای نمونه‌های کوچک دارای اریبی^۲ بالایی می‌باشد و به بیش‌برآورد^۳ پارامتر N خواهد انجامید. به همین خاطر وی برآوردگر اصلاح شده‌ی زیر را که دارای اریبی کمتری است به جای برآوردگر (۲-۱) پیشنهاد می‌کند.

(۲-۲)

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

با استفاده از روش برآورد "حداکثر درست‌نمایی" نیز همین برآوردگر به دست می‌آید.

- مدل‌های لگ خطی

در این روش ابتدا یک جدول t بعدی ناقص (t تعداد فهرست‌ها می‌باشد)؛ 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 ۲-۱ مرتب کنیم.

^۱ Chapman

^۲ Bias

^۳ overestimate

برآورد کم‌شماری مرگ و میر با ... ————— گزیده مطالب آماری - ۶۳

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

		نمونه‌ی اول			
		حضور		عدم حضور	
نمونه‌ی سوم	حضور	نمونه‌ی دوم		نمونه‌ی دوم	
		حضور	عدم حضور	حضور	عدم حضور
نمونه‌ی اول	حضور	X_{111}	X_{121}	X_{211}	X_{221}
نمونه‌ی اول	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 برآزش 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 برآورد و از آن‌جا برآورد، اندازه‌ی کل جمعیت به دست آورده می‌شود.

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

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

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

¹Akaiik Information Criterion

گزیده مطالب آماری - ۶۳ ————— برآورد کم‌شماری مرگ و میر با ...

که در آن G^2 آماره نسبت درست‌نمایی^۱ 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 برآورد کم‌شماری و در نهایت برآورد اندازه‌ی کل جمعیت می‌پردازیم. به طور کلی مفهوم پوشش نمونه‌ای با برآورد 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 اندازه‌ی کمی برای همپوشانی اطلاعات و پارامترهایی برای منبع وابستگی برآورد می‌شود و در نتیجه می‌توان اندازه‌ی جمعیت را از طریق رابطه‌ی آن با پوشش نمونه‌ای برآورد 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 برآورد 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 برآورد دقیق اندازه‌ی جمعیت و برای مدل‌بندی وابستگی، میزان همپوشانی اطلاعات در فهرست‌ها حداقل باید ۵۵ درصد باشد.

^۱ 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 این که 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 برآورد تعداد مرگ‌های مربوط به شهر گرگان در سال ۱۳۸۱ سه روش مختلف را به کار برده‌ایم. فهرست متوفیان ثبت شده به وسیله‌ی اداره‌ی آمار و اطلاع‌رسانی 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 برآورد می‌شود. که برای انجام محاسبات از نرم افزار SAS استفاده شده است.

گزیده مطالب آماری -۶۳ ————— برآورد کم‌شماری مرگ و میر با ...

- نتایج مقایسه‌ی فهرست‌ها با یکدیگر

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

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

نمونه‌ها	برآوردگر پترسن	برآوردگر چپمن	خطای معیار چپمن	فاصله‌ی اطمینان ۹۵ درصد برآوردگر چپمن	
				کران بالا	کران پایین
فهرست‌های ۲و۱	۱۲۳۳	۱۲۳۳	۳۰	۱۳۰۲	۱۱۸۴
فهرست‌های ۳و۱	۹۷۱	۹۷۱	۱۱	۹۸۹	۹۴۸
فهرست‌های ۳و۲	۱۲۴۲	۱۲۴۱	۳۶	۱۳۲۷	۱۱۸۴
فهرست‌های ۳و۱ ادغام شده و ۲	۱۲۸۰	۱۲۷۹	۲۶	۱۳۳۷	۱۲۳۵

بر اساس نتایج جدول ۳-۱ برآوردهای قابل توجیهی از روی فهرست‌های (۲و۱) و (۳و۲) 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 برآورد فهرست‌های ۳و۱ یعنی ۹۷۱ مورد، قابل پذیرش نمی‌باشد (یعنی مقدار برآورد شده از تعداد فوت ثبت شده نیز کمتر است) و این

برآورد کم‌شماری مرگ و میر با ... ————— گزیده مطالب آماری - ۶۳

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 همان‌طور 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 برآورد قابل توجیهی است. برآوردهای مرگ و میر از فهرست‌های ۱ و ۲ و فهرست‌های ۱ و ۲ که به ترتیب برابر با ۱۲۳۲ و ۱۲۴۲ می‌باشد تقریباً نزدیک به هم 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 برآزش داده می‌شود، آن‌گاه با توجه به معیار AIC مناسب‌ترین مدل انتخاب می‌شود. سپس با فرض نبودن اثر متقابل سه تایی، مدل انتخاب شده برای برآورد خانه‌ی مشاهده نشده استفاده می‌گردد و از آن‌جا برآورد اندازه‌ی کل مرگ‌ها به دست می‌آید.

نتیجه‌ی حاصل در جدول ۲-۳ نشان داده شده است.

گزیده مطالب آماری - ۶۳ ————— برآورد کم‌شماری مرگ و میر با ...

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

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

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 ۱۲-۱۳ مدلی است که فاقد اثر متقابل دو تایی (۲ و ۳) می باشد و بقیه‌ی آثار (شامل آثار اصلی و آثار متقابل دیگر به جز مورد یاد شده) را دارد و یا مدل ۱۲-۳ 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 جدول ۲-۳ نشان داده شده است در نهایت مدل ۱۲-۱۳ 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 ($AIC = 2/03$) دوم، خطای برآورد آن نیز کم است ($Se=64$) در نتیجه تحت این

برآورد کم‌شماری مرگ و میر با ... ————— گزیده مطالب آماری - ۶۳

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 ۲۹ درصد به دست آمد.

- نتایج اجرای روش پوشش نمونه‌ای

در این روش با توجه به این که 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 این که 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 ۵۵ درصد). مقدار \hat{N}_1 نیز در شرایطی به کار می‌رود که دست کم بین یک جفت نمونه وابستگی وجود داشته و میزان همپوشانی اطلاعات کم باشد (کمتر از ۵۵ درصد) 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 این که 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

گزیده مطالب آماری - ۶۳ ————— برآورد کم‌شماری مرگ و میر با ...

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

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

فاصله اطمینان ۹۵ درصد		خطای معیار	برآورد کل مرگ‌ها	C	D	M	برآوردهای N
کران پایین	کران بالا						
۱۱۷۰	۱۱۸۲	۹	۱۱۷۸	۰/۷۹۷	۹۳۹	۱۰/۷۱	\hat{N}_0
۱۲۹۸	۱۳۲۱	۱۵	۱۳۱۱	۰/۷۹۷	۹۳۹	۱۰/۷۱	\hat{N}
۱۲۶۲	۱۲۸۴	۱۶	۱۲۷۳	۰/۷۹۷	۹۳۹	۱۰/۷۱	\hat{N}_1

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

در این جدول خطای معیار برآورد اندازه‌ی جمعیت بر اساس هزار تکرار بوت استرپ^۱ به دست آمده و ممکن است برای این خطا در آزمایشهای مختلف مقادیر متفاوتی بدست آید. \hat{N}_0 برآورد اندازه‌ی جمعیت در وضعیتی است که منابع از یکدیگر مستقل می‌باشند، \hat{N} برآورد اندازه‌ی جمعیت در وضعیتی است که بین منابع وابستگی وجود دارد و میزان همپوشانی بین منابع به اندازه‌ی کافی بزرگ می‌باشد (دست کم ۵۵ درصد) و \hat{N}_1 برآورد یک مرحله‌ای اندازه‌ی جمعیت برای پوشش نمونه‌ای کم می‌باشد. استفاده از این برآورد هنگامی توصیه می‌شود که خطای معیار برآورد شده نسبتاً بزرگ باشد.

^۱ Bootstrap

بحث و نتیجه‌گیری

همان‌طور که قبلاً گفته شد در به‌کارگیری 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 به‌طور 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 برآورد می‌شود.

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

طی سال‌های اخیر سازمان‌های مرتبط با موضوع جمعیت، هر یک به‌طور جداگانه برآورد 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 برآورد کرد که به‌عنوان 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 برآورد 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

گزیده مطالب آماری -۶۳ ————— برآورد کم‌شماری مرگ و میر با ...

گزارش سازمان ملل مربوط به کل کشور است و به نظر می‌رسد بالاتر بودن آن برآورد به دلیل در نظر گرفتن مناطق روستایی در برآورد باشد. بنابراین 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.

فهرست منابع:

۱) زنجانی حبیب ا...، نوراللهی طه، جداول مرگ و میر ایران برای سال ۱۳۷۵، مؤسسه‌ی پژوهش تأمین اجتماعی، تهران ۱۳۷۹، صفحه ۳۷.

برآورد کم‌شماری مرگ و میر با ... ————— گزیده مطالب آماری - ۶۳

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) George 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.

گزیده مطالب آماری -۶۳ ————— برآورد کم‌شماری مرگ و میر با ...

- 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.