رویکرد آنتروپی به تنظیم جدول عمر، مطالعهٔ موردی: جدول عمر ایران

رضوان رضائی، غلامحسین یاری*
دانشگاه علم و صنعت ایران، دانشکده ریاضی
دریافت ۹۸/۰۲/۲۵ پذیرش ۹۸/۰۸/۱۳

چکیده

تحلیل بقا و به‌ویژه برآورد توزیع بقا از موضوعات مهم در علوم آماری است. روش‌های پارامتری و ناپارامتری مختلفی برای برآورد توزیع بقا مطرح شده‌اند. در این ارتباط توزیع‌های بقای تئوریکی مشخص شده‌اند و پارامترهایشان به‌کمک روش‌هایی مانند برآوردگر حداکثر درست‌نمایی و برآوردگر بیزی به‌دست می‌آیند. از جمله روش‌های ناپارامتری نیز می‌توان به‌روش کاپلان‌مایر، رگرسیونی کاکس و جدول عمر اشاره کرد. علاوه بر این، یکی دیگر از مباحثی که در تحلیل بقا اهمیت زیادی دارد است طبقه‌بندی داده‌ها است که همواری و نیکویی برازش دو نیاز اساسی برای آن محسوب می‌شوند. از سوی دیگر در نظریه احتمال بر اساس عبارت تعریف‌شده‌ی پایه‌ای آنتروپی، دو مدل بهینه‌سازی یکی بر اساس اصل بیشینه آنتروپی (ME) و دیگری بر پایهٔ اصل کمینه معیار کولبک-لیبلر (MKL) به منظور برآورد توزیع احتمال ارائه شده است. در این مقاله، رویکرد دو مدل بهینه‌سازی فوق را به برآورد توزیع بقا و توزیع احتمال به‌ویژه برای داده‌های طبقه‌بندی شده بررسی می‌کنیم. در این پژوهش‌ها علاوه بر بررسی مدل‌های پارامتریک، روش ناپارامتری جدیدی که یک تابع هدف ترکیب شده از دو اصل ME و MKL و یک ضریب برای اطمینان از درجهٔ نیکویی برازش و هموارسازی برآوردها که نشان‌دهندهٔ اولویت این دو شاخص در طبقه‌بندی داده‌ها است را به‌کار می‌بریم. ما از این روش برای برآورد توزیع احتمال مرگ و میر سن مشخص (ستون q_x) در جدول عمر، استفاده می‌کنیم. در پایان به‌کمک این روش جدول عمر زنان و مردان ایران در سال ۱۳۹۰ (ه.ش) را تنظیم می‌کنیم.

واژه‌های کلیدی: تحلیل بقا، نظریهٔ اطلاع، اصل بیشینه آنتروپی، اصل کمینه کولبک- لیبلر، جدول عمر.

مفاهیم و مقدمات

تحلیل بقا در آمار زیستی و عمل‌گر اهمیت زیادی دارد. در این مقاله قصد داریم توزیع طول عمر که از موضوعات مهم تحلیل بقا است را به‌کمک نظریه اطلاع برآورد کنیم و سپس جدول عمر زنان و مردان ایران در سال ۱۳۹۰ (ه.ش) را به این وسیله تنظیم کنیم. برای این منظور ابتدا برخی از مفاهیم مقدماتی مربوط به نظریه اطلاع و جدول عمر را ارائه می‌کنیم.

تعریف ۱. فرض کنید X یک متغیر تصادفی پیوسته با تابع چگالی $f_X(x)$ و با تکیه‌گاه S_X باشد آن‌گاه *آنتروپی شانون* X به‌صورت $h(X)=-\int_{S_X} f(x)\log f(x)dx$ تعریف می‌شود [۱۱].

معیار فوق که میزان عدم قطعیت متغیر تصادفی X را نشان می‌دهد اولین بار به‌وسیلهٔ شانون[1] معرفی شد. پس از آن برخی تعمیم‌های این تعریف و همین‌طور شاخص‌های مرتبط دیگر از جمله شاخص آنتروپی نسبی ارائه شد. این شاخص به‌وسیلهٔ کولبک و لیبلر[2] با نام آنتروپی نسبی[3] معروف و بدین‌صورت تعریف می‌شود:

*نویسنده مسئول yari@iust.ac.ir

1. Shannon
2. Kullback & Leibler
۳. این معیار به‌عنوان یک شاخص برای بررسی نیکویی برازش یک توزیع احتمال $f(x)$ در مقابل توزیع احتمال مرجع $g(x)$ استفاده می‌شود. به بیان دقیق‌تر این نام‌گذاری به‌دلیل کارایی این معیار در تعیین میزان افتراق (تفاوت) بین دو تابع چگالی احتمال است.

[DOI: 10.52547/mmr.7.2.279]

تعریف ۲. برای هر دو متغیر تصادفی پیوسته X و Y با توابع چگالی به‌ترتیب f و g، آنتروپی نسبی که معیار کولبک-لیبلر نیز نامیده می‌شود از رابطهٔ (۱) به‌دست می‌آید [۹].

$$KL(f(x)\|g(x)) = \int_R f(x)\ln(f(x)/g(x))dx, \qquad (۱)$$

تعمیم‌هایی از شاخص KL نیز معرفی شده است، از جمله می‌توان به تعریفی که به‌وسیلهٔ لیو[1] ارائه شده است اشاره کرد که در ادامه آن را مطرح می‌کنیم.

تعریف ۳. (آنتروپی نسبی بر اساس تابع بقا (KLS)) $X_1, X_2, \ldots$ یک دنبالهٔ مثبت، مستقل و هم توزیع از متغیرهای تصادفی با تابع بقای غیرصعودی $\bar{F}(x;\theta) = P_\theta(X > x)$ که θ بردار پارامترهای نامعلوم و $\bar{G}(x) = \sum_{i=0}^{n-1}(1-i/n)I_{[X_{(i)}, X_{(i+1)})}(x)$ تابع بقای تجربی بر اساس یک نمونهٔ تصادفی nتایی از $\bar{F}(x;\theta)$ که I تابع مشخصه و $(0 = X_{(0)} \le) X_{(1)} \le X_{(2)} \le \ldots \le X_{(n)}$ نمونه مرتب شده است، را در نظر بگیرید. در این صورت معیار آنتروپی نسبی یا کولبک-لیبلر بر اساس توابع بقای $\bar{G}(x)$ و $\bar{F}(x;\theta)$ به‌صورت رابطهٔ (۲) محاسبه می‌شود [۱۰].

$$KLS(\bar{G}_n(x)\|\bar{F}(x)) = \int_0^\infty [\bar{G}_n(x)\ln(\bar{G}_n(x)/\bar{F}(x))] - [\bar{G}_n(x) - \bar{F}(x)]dx, \qquad (۲)$$

در رابطهٔ (۲) $\bar{F}(x;\theta)$ توزیع مرجع است و KLS فاصلهٔ $\bar{G}(x)$ تا $\bar{F}(x;\theta)$ را می‌سنجد. رابطهٔ (۲) را می‌توان بدین‌صورت بیان کرد [۱۴]:

$$KLS(\bar{G}_n(x)\|\bar{F}(x)) = \int_0^\infty [\bar{G}_n(x)\ln(\bar{G}_n(x)/\bar{F}(x))] - [\bar{X} - E(X)]dx, \qquad (۳)$$

(تعاریف بالا برای متغیر تصادفی گسسته X با تابع جرم احتمال $P(X = x)$ نیز با جابه‌جایی $\int, \Sigma$ برقرار است. هم‌چنین برای مطالعهٔ جزئیات بیش‌تر به [۱۴] مراجعه کنید.)

از سوی دیگر در جمعیت‌شناسی پیش‌بینی احتمال مرگ افراد جامعه برای بررسی ساختار بقای جامعه اهمیت خاصی دارد. این پیش‌بینی تنها بر اساس محاسبه احتمال مرگ افراد پیشین در شرایط نسبتاً همگن و با استفاده از اطلاعات جمعیتی جمع‌آوری‌شده، امکان‌پذیر است. علاوه بر این در اکچوئری تعهدات یک شرکت بیمه عمر به طول عمر بیمه‌شدگان وابسته است. جدول عمر[2] یکی از مهم‌ترین کشفیات جمعیت‌شناسی و از نخستین ابزارهای آماری است که برای بررسی حوزه‌های اصلی جمعیت چون مرگ‌ومیر، امید زندگی[3]، مهاجرت[4]، باروری[5] و هم‌چنین ساختار رشد جمعیت به‌کار گرفته شده است. ادموندهالی و جان‌گرانت[6] اولین دانشمندانی بودند که (به‌صورت جداگانه) جدول عمرهایی را برای لهستان و انگلیس ابداع کردند [۱]. جدول عمر، جدول مرگ‌ومیر[7] نیز نامیده می‌شود و دارای انواع مختلفی است: جدول عمر کامل، جدول عمر خلاصه، جدول عمر مقطعی و جدول عمر نسلی [۱]. در ایران تاکنون جدول مرگ‌ومیری که به‌صورت رسمی منتشر شود و مبنای محاسبات جمعیتی و بیمه‌ای باشد، انتشار نیافته است. مشکل اساسی، در دست نبودن آمار کل و دقیق فوت‌شدگان برحسب سن در زمان فوت است. با این حال در سال‌های اخیر کوشش‌هایی از طرف بعضی سازمان‌ها و پژوهش‌گران انجامشده است. شاید از اولین کارهایی که می‌توان در زمینه ساخت جدول عمر در ایران نام برد، تلاشی است که نهاپتیان و خزانه در سال ۱۳۵۲ در طرحی با عنوان «بررسی و

1. Lio
2. Life table
3. Life Expectancy
4. Migration
5. Fertility
6. J.Grant and Edmund Halley
7. Mortality table

[DOI: 10.52547/mmr.7.2.279]

برآورد میزان‌های اختصاصی سنی مرگ و میر و باروری در ایران»، داشتند. آنها جدول عمر کشور را در بازه‌های ۵ ساله و به‌تفکیک شهر و روستا و جنس تنظیم کردند. بعد از آن افرادی از جمله ک.ال کهلی، شمس، دکتر حسین ملک افضلی و محمودی، دکتر حسین ملک افضلی و همکاران، نوراللهی، زنجانی و کوششی، میرزایی، کوششی و ناصری، نوراللهی و نجاتیان جداول عمر با روش‌های متفاوت و با اهداف بعضاً متفاوتی، طراحی و تنظیم کرده‌اند [۱]. برخی از ستون‌های مهم این جدول بدین‌قرار هستند:

x و $x+d$ (x سن شروع، $x+d$ سن پایان دوره و d فاصلهٔ آن دو سن به سال تمام است برای مثال در جدول عمر کامل $d=1$ است.)، l_x (تعداد بازماندگان تا سن درست x که در این جدول‌ها با ۱۰۵ نفر شروع شده است.)، d_x (تعداد مرگ‌ومیرها از سن x تا $x+d$)، e_x (امید زندگی در سن x است که $e_x=\int_0^\infty l_a da$)، q_x (احتمال مرگ‌ومیر در هر بازه x تا $x+d$ است که در تشکیل یک جدول نقش کلیدی ایفا می‌کند.) و M_x (میزان مرگ‌ومیر در وسط گروه سن x و $x+d$ را نشان می‌دهد و معمولاً با m_x مشخص می‌شود. با فرض ثبات نرخ لحظه‌ای مرگ و میر در بازه سنی x تا $x+d$، رابطهٔ $q_x=1-\exp(-m_x)$ برقرار است [۱۲].). برای جزئیات بیش‌تر در مورد سایر ستون‌های جدول عمر و روابط بین آن‌ها به [۱] مراجعه کنید.

معیار KLS و برآورد پارامترهای توزیع گمپرتز

در بررسی‌های بقای سیستم‌ها به‌منظور مدل‌سازی داده‌های مرگ و میر و طول عمر آن‌ها، برخی توزیع‌های آماری استفاده می‌شود. از جملهٔ این توزیع‌ها، توزیع گمپرتز، گمپرتز-مکهام، وایبل، هلیگمن پولارد نوع ۱، ۲ و ۳ و سیلر و هم‌چنین تعمیم‌های از آنها مانند توزیع‌های BGG[1]، OLGG[2] و EGWG[3] هستند.

در این‌جا از معیار KLS برای بررسی توزیع‌های پارامتریک مرگ و میر استفاده می‌کنیم. برای این منظور کافی است تابع بقا و امید ریاضی توزیع موردنظر را در رابطهٔ (۳) جای‌گذاری کرده و سپس نسبت به هرکدام از پارامترها با در نظر گرفتن ثابت بودن سایر پارامترها از عبارت KLS به‌دست آمده مشتق گرفته و آنها را برابر صفر قرار داده و سپس دستگاه به‌دست‌آمده را حل کرد [۱۴]. در این‌جا برای جلوگیری از دور شدن هدف این بخش از ارائهٔ جزئیات بیش‌تر این روش صرف‌نظر می‌کنیم و خواننده را به بررسی [۱۴] دعوت می‌کنیم و فقط نتایج مربوط به استفاده از این روش را برای توزیع گمپرتز با تابع بقا $\bar{F}(X)=\exp(-(\lambda/c)\exp(cx-1))$ و امید ریاضی $E(X)=(1/c)\exp(\lambda/c)[(\lambda/c)-\ln(\lambda/c)-\gamma]$ که c و λ پارامترهای این توزیع و γ ثابت اویلر است، ارائه می‌کنیم. بر اساس آن‌چه بالا گفته شد با جای‌گذاری $\bar{F}(X)$ و $E(X)$ توزیع گمپرتز در رابطهٔ (۳) داریم:

$$KLS(\bar{G}_n(x)\|\bar{F}(x))=\sum_{k=1}^{n}(1-\frac{k-1}{n})\ln(1-\frac{k-1}{n})\Delta x_k+-(\bar{x}-(1/c)\exp(\lambda/c)[(\lambda/c)-\ln(\lambda/c)-\gamma])$$
$$+\sum_{k=1}^{n}(1-\frac{k-1}{n})\frac{\lambda}{c}(\frac{1}{c}e^{cx_{(k)}}-x_{(k)}-\frac{1}{c}e^{cx_{(k-1)}}+x_{(k-1)}).$$

سپس با مشتق‌گیری از مقدار KLS فوق، به‌ترتیب نسبت به پارامتر λ (با فرض ثابت بودن پارامتر c) و c (با فرض ثابت بودن پارامتر λ) و دستگاه معادلهٔ زیر به‌دست می‌آید.

1. The beta generalized Gompertz distribution
2. The Odd Log-Logistic Generalized Gompertz Distribution
3. The Exponentaited Generalized Weibull-Gompertz Distribution

[DOI: 10.52547/mmr.7.2.279]

$$\begin{cases} \frac{1}{c}\sum_{k=1}^{n}(1-\frac{k-1}{n})(\frac{1}{c}(e^{cx_{(k)}}-e^{cx_{(k-1)}})-\Delta x_{(k)})+\frac{e^{\frac{\lambda}{c}}}{c^2}(\frac{\lambda}{c}-\ln(\frac{\lambda}{c})-\gamma)+\frac{e^{\frac{\lambda}{c}}}{c^2}\frac{\lambda-c}{c\lambda}=0, \\ (1+\frac{\lambda}{c})\frac{-e^{\frac{\lambda}{c}}}{c^2}(\frac{\lambda}{c}-\ln(\frac{\lambda}{c})-\gamma)+\frac{e^{\frac{\lambda}{c}}}{c}(\frac{1}{c}-\frac{\lambda}{c})+\sum_{k=1}^{n}[(1-\frac{k-1}{n})(\frac{-\lambda}{c})(\frac{1}{c}(e^{cx_{(k)}}-e^{cx_{(k-1)}})-\Delta x_{(k)})]+ \\ \sum_{k=1}^{n}[(1-\frac{k-1}{n})(\frac{\lambda}{c})(\frac{-1}{c^2}(e^{cx_{(k)}}-e^{cx_{(k-1)}})+\frac{1}{c}(x_{(k)}e^{cx_{(k)}}-x_{(k-1)}e^{cx_{(k-1)}}))]=0. \end{cases}$$

چنان‌که ملاحظه می‌شود جواب‌های این دستگاه به‌طور سر راست قابل محاسبه نیست و باید از روش‌های عددی مانند نیوتن، شبه نیوتن، روش‌های تکرار و از نرم‌افزارهای مربوطه مانند MATLAB کمک گرفت. به‌طور مثال در بخش‌های بعد از این روش برای مدل‌سازی مرگ و میر ۲۰۸ موش استفاده می‌کنیم و نتایج را در جدول ۱ ارائه می‌کنیم. در ادامه، رویکرد آنتروپی به داده‌های طبقه‌بندی شده به‌ویژه برای تنظیم یک جدول عمر را مورد مطالعه و بررسی قرار می‌دهیم. قبل از آن به‌طور خلاصه اصول بهینه‌سازی آنتروپی را شرح می‌دهیم.

بررسی اجمالی اصول بهینه‌سازی آنتروپی

آنتروپی نظریهٔ اطلاع، به‌وسیلهٔ شانون و اصل بهینه‌سازی آنتروپی تحت عنوان اصل بیشینه آنتروپی به‌وسیلهٔ جینز ارائه شده است [۱۱]. کولبک و همکاران نیز با معرفی اصل کمینه معیار کولبک-لیبلر شرایط استفاده از آنتروپی را وسعت دادند و آن را از یک شاخص نظریهٔ اطلاع به یک ابزار آمار استنباطی تبدیل کردند [۹]. به بیان دقیق‌تر، بهینه‌سازی آنتروپی شامل دو اصل بیشینه آنتروپی (ME[1]) و اصل کمینه معیار کولبک-لیبلر (MKL[2]) است و این دو اصل برای برآورد توزیع احتمال با استفاده از مفهوم آنتروپی به‌کار می‌روند. اولی یک توزیع احتمال را تنها بر پایهٔ اطلاعات معلوم، بدون اضافه کردن هرگونه اطلاعات ذهنی دیگری، برآورد می‌کند و دومی برای برآورد کردن یک توزیع احتمال که نزدیک‌ترین توزیع به توزیع پیشین است، استفاده می‌شود.

جینز اعتقاد داشت که با داشتن میانگین، معمولاً بی‌شمار توزیع سازگار وجود دارد. ME ما را تشویق به انتخاب توزیعی که آنتروپی شانون را به حداکثر رسانده و به‌طور هم‌زمان سازگار با محدودیت میانگین است، می‌کند. برای این منظور فرض کنید Θ یک متغیر تصادفی گسسته در فضای احتمال (Ω, F, P) باشد که در آن $\Omega=\{\theta_1,\theta_1,..,\theta_n\}$ و مقدار نامعلوم $P(\Omega=\theta_i)=p_i, i=1,2,...,n$ است و به‌کمک برخی اطلاعات معلوم $g_i(\theta)(j=1,2,...,m)$ مانند میانگین، واریانس، j امین گشتاور و غیره برآورد می‌شود. ME بدین‌صورت توصیف می‌شود:

$$\max \quad H(p)=-\sum_{i=1}^{n}p_i\ln p_i \qquad s.t\begin{cases}\sum_{i=1}^{n}p_i=1 \\ \sum_{i=1}^{n}p_i g_j(\theta)=E_j,\ j=1,2,...,m \\ p_i\geq 0, i=1,2,...,n\end{cases}$$

که در آن $H(p)$ آنتروپی توزیع احتمال روی Ω است.

1. Maximum Entropy
2. Minimum KL

[DOI: 10.52547/mmr.7.2.279]

هم‌چنین اگر یک توزیع پیشین برای Θ یعنی $Q(\Theta=\theta_i)=q_i$ وجود داشته باشد، آن‌گاه MKL می‌تواند برای به‌دست آوردن توزیع برآورد شدهٔ دیگری که از نظر آماری نزدیک‌ترین به توزیع پیشین تحت همان قیود ME است، استفاده شود. MKL بدین‌صورت مدل‌بندی می‌شود:

$$\min \quad K(P,Q)=-\sum_{i=1}^{n} p_i \ln(p_i/q_i) \qquad s.t \begin{cases} \sum_{i=1}^{n} p_i = 1 \\ \sum_{i=1}^{n} p_i g_j(\theta) = E_j,\ j=1,2,...,m \\ p_i \geq 0, i=1,2,...,n \end{cases}$$

که $K(P,Q)$ معیار کولبک لیبلر است.

دو مدل مذکور به‌کمک روش لاگرانژ [۸] می‌توانند حل شوند که جواب‌ها بدین‌صورت است:

$$p_i^{(\max)} = \exp[-\alpha_0 - \sum_{i=1}^{m} \alpha_i g_j(\theta)], \qquad (۴)$$

$$p_i^{(\min)} = q_i \exp[-\beta_0 - \sum_{i=0}^{m} \beta_i g_j(\theta)] \qquad (۵)$$

و $\alpha_1,\alpha_2,...,\alpha_n$ و $\beta_1,\beta_2,...,\beta_n$ ضرایب لاگرانژ مدل‌های مذکور هستند.

«طبقه‌بندی داده‌ها» با استفاده از اصول بهینه‌سازی آنتروپی

یک گروه زنده که جمعیت اولیه آن را با l_0 نشان می‌دهیم، در نظر بگیرید. در زمان $x\geq 0$ جامعه‌ی زنده را با $l_x(x=0,1,...,T)$ مشخص و فرض می‌کنیم که $l_T=0$ باشد. هم‌چنین $d_x(x=0,1,...,T)$ را تعداد مرگ و میر جامعه از x تا $x+1$ در نظر بگیرید آن‌گاه

$$d_x = l_x - l_{x+1}. \qquad (۶)$$

مقدار d_x معمولاً از مشاهدات به‌دست می‌آید. در تحلیل بقا q_x برای نشان دادن احتمال مرگ و میر که به معنای احتمال مرگ یک فرد در سال بعد است، استفاده می‌شود. احتمال مرگ و میر یک پارامتر اساسی برای ساختن یک جدول عمر است و اهمیت بسیار زیادی در علم بیمهٔ عمر دارد. در شرایط واقعی این احتمال از داده‌های نمونه می‌تواند به دست آید و به‌صورت $\hat{q}_x = d_x/l_x$ مشخص می‌شود. بررسی و چرایی این که $\sum_{x=0}^{T}\hat{q}_x \neq 1$ آسان است. به‌هرحال، $\hat{q}_x$ برآوردی از q_x است که در شرایط واقعی نزدیک‌ترین مقدار ممکن را برای احتمال مرگ و میر ارائه می‌کند. این فرایند در بیمهٔ عمر «طبقه‌بندی داده» نامیده می‌شود. علاوه بر این، برآورد q_x بر اساس اطلاعات نمونه است؛ بنابراین ما باید از اطلاعات موجود در نمونه به‌طور کامل استفاده کنیم و سعی کنیم بهترین اطلاعات اضافی ممکن را وارد مسئله کنیم. این امر یک پایه و اساس منطقی برای استفادهٔ ME و MKL را می‌تواند فراهم کند. حال به‌منظور استفاده از ME و MKL، احتمال مرگ و میر را به‌صورت (۷) تعریف می‌کنیم.

$$\tilde{q}_x = d_x/l_0,\ x=0,1,...,T, \qquad (۷)$$

(نسبتی از احتمال مرگ جامعه در گروه‌های سنی مختلف به جامعهٔ اولیه). $\tilde{q}_x$ نیاز توزیع احتمال را می‌تواند فراهم کند (زیرا در اصول احتمال کلموگروف صدق می‌کند: $\forall x: 0\leq \tilde{q}_x \leq 1$ و $\sum_{x=0}^{T}\tilde{q}_x = 1$) و روابط

[DOI: 10.52547/mmr.7.2.279]

$q_0 = \tilde{q}_0, \tilde{d}_x = \tilde{q}_x - \tilde{q}_{x+1}, \hat{q}_x = \tilde{d}_x / \tilde{l}_x$ نیز برقرار است، بنابراین برآورد q_x، یعنی $\hat{q}_x$ را می‌توان این بار با $\tilde{q}_x$ مشخص و برآورد آن به‌روش ME از حل رابطهٔ (۸) حاصل می‌شود.

$$\max \quad H = -\sum_{i=1}^{n} \hat{q}_x \ln \hat{q}_x \qquad s.t \begin{cases} \sum_{i=1}^{n} \hat{q}_x = 1 \\ \sum_{i=1}^{n} x\hat{q}_x = E_1 \\ \sum_{i=1}^{n} (x - E_1)^j \hat{q}_x = E_j,\ j = 2,...,k \\ \hat{q}_x \geq 0,\ x = 1,2,...,T \end{cases} \tag{۸}$$

که در آن $\hat{q}_x$ برآوردی از $\tilde{q}_x$، E_1 میانگین داده‌های نمونه و E_j گشتاور j ام است، از سوی دیگر $\tilde{q}_x$ مانند یک توزیع پیشین برای احتمال مرگ می‌تواند در نظر گرفته شود، در نتیجه یک مدل MKL برای برآورد کردن $\tilde{q}_x$ به‌صورت (۹) است.

$$\min \quad K = -\sum_{i=1}^{n} \hat{q}_x \ln(\hat{q}_x / \tilde{q}_x) \qquad s.t \begin{cases} \sum_{i=1}^{n} \hat{q}_x = 1 \\ \sum_{i=1}^{n} x\hat{q}_x = E_1 \\ \sum_{i=1}^{n} (x - E_1)^j \hat{q}_x = E_j,\ j = 2,...,k \\ \hat{q}_x \geq 0,\ x = 1,2,...,T \end{cases} \tag{۹}$$

حل معادلات (۸) و (۹) به‌کمک روش لاگرانژ با ضرایب α_j, β_j به‌صورت روابط (۱۰) و (۱۱) است.

$$\hat{q}_x^{(\max)} = \exp[-\sum_{i=0}^{m} \alpha_j (x - E_1)^j], \tag{۱۰}$$

$$\hat{q}_x^{(\min)} = \tilde{q}_x \exp[-\sum_{i=0}^{m} \beta_j (x - E_1)^j] \tag{۱۱}$$

داده‌های طبقه‌بندی شده به‌وسیلهٔ ترکیب دو روش ME و MKL (MEMKL)

چنان‌که در بخش‌های قبل گفته شد همواری و نیکویی برازش همیشه مهم‌ترین موضوعات موردتوجه در طبقه‌بندی داده‌ها بوده‌اند هرچند تکنیک‌های بسیار دیگری نیز توسعه یافته‌اند. به‌عنوان مثال، گسترده‌ترین روش مورد استفاده تابه‌حال، به‌خصوص به‌وسیلهٔ بیمه‌دانان آمریکای شمالی برای ساخت جداول عمر، روش ویتاکر-هندرسون[1] از طبقه‌بندی است [۲]. این روش در کار بولمن[2] و ویتکار[3] معرفی شد و به نظریه هندرسون کمک کرد [۴]، [۵]، [۶]، [۱۳]. روش ویتاکر-هندرسون مقادیر طبقه‌بندی شده را با کمینه کردن مقدار

$$M = F + hS = \sum_{x=1}^{n} \omega_x (v_x - u_x)^2 + h \sum_{x=1}^{n-z} [\Delta^z v_x]^2 \tag{۱۲}$$

1. Whittaker-Henderson
2. Bohlmann
3. Whittaker

[DOI: 10.52547/mmr.7.2.279]

به‌دست می‌آورد که F و S به‌ترتیب مقادیر وزندارشدهٔ نیکویی برازش برای داده‌های اولیه و ضریب همواری هستند. u_x توزیع پیشین مرگ و میر و v_x مرگ و میر برآورد شده است، یعنی نتیجهٔ طبقه‌بندی داده و $\Delta^z v_x$ از z^{th} اختلاف v_x (معمولاً $z=2,3$ یا بیش‌تر) است. ω_x ضریب وزنی و h عامل تعدیل مثبت بین نیکویی برازش و همواری است. این روش به‌طور گسترده‌ای استفاده شده و تبدیل به یک منطق اساسی در طبقه‌بندی داده و دیگر روش‌های جاری در این زمینه شده است.

به‌طورکلی، مشخصه‌سازی روش‌های مختلف طبقه‌بندی داده ممکن است از دو جهت باشد، یکی به‌کار بردن نیکویی برازش و همواری و دیگری چگونگی اندازه‌گیری همواری و نیکویی برازش؛ بنابراین، طبقه‌بندی داده را به‌عنوان یک موضوع دو هدفه می‌توان نگاه کرد. ازیک طرف، نتایج طبقه‌بندی داده‌ها باید هموار باشد و از سوی دیگر باید به داده‌های اصلی نزدیک باشند. در ادامه یک رویکرد جدید از طبقه‌بندی داده که هر دو روش ME و MKL را در مدل زیر می‌تواند ترکیب کند را مطرح می‌کنیم [۷]:

$$\min \quad G=\mu\sum_{x=1}^{T}\hat{q}_x\ln\hat{q}_x+(1-\mu)\sum_{x=1}^{T}\hat{q}_x\ln(\hat{q}_x/\tilde{q}_x) \qquad s.t\begin{cases}\sum_{x=1}^{T}\hat{q}_x=1\\ \sum_{x=1}^{T}x\hat{q}_x=E_1\\ \sum_{x=1}^{T}(x-E_1)^j\hat{q}_x=E_j,\ j=2,...,k\\ \hat{q}_x\geq 0, x=1,2,...,T\end{cases} \tag{۱۳}$$

که در آن $0\leq\mu\leq 1$ عامل تعدیل داده شده بین همواری و نیکویی برازش، E_1 امید ریاضی و E_jها گشتاورهای مرکزی مرتبه j ام متغیر تصادفی X هستند. (در معادلهٔ (۱۳) ملاحظه می‌شود زمانی که μ را ۰ قرار می‌دهیم تنها روش MKL و زمانی که آن را برابر ۱ قرار می‌دهیم تنها روش ME، یعنی به‌ترتیب معادلات (۹) و (۸)، برقرار می‌ماند.) از آن‌جاکه معادلات (۸) و (۹) قابل‌حل هستند، به‌آسانی می‌توان نتیجه گرفت که معادلهٔ (۱۳) نیز قابل‌حل است.

در مدل بالا، ME به‌عنوان شاخص همواری و MKL به‌عنوان شاخص نیکویی برازش استفاده می‌شوند. از این‌رو، این دو شاخص با یک ضریب خطی μ برای انعکاس دادن وزن‌های مختلف همواری و نیکویی برازش، یک‌پارچه می‌شوند. دلیل اتخاذ یک ترکیب محدب از همواری و نیکویی برازش، مطمئن شدن از تحدب تابع هدف و قابل حل بودن مدل پیشنهاد شده است. در واقع، چگونگی تصمیم گرفتن وزن مناسب بین همواری و نیکویی برازش موضوعی بسیار بحث برانگیز است و معمولاً با روش تجربی تعیین می‌شود. در ادامه این بخش از این روش ترکیبی برای طبقه‌بندی داده‌های مرگ ۲۰۸ موش استفاده می‌کنیم.

مثال: داده‌های جدول (۱) مربوط به بررسی طول عمر ۲۰۸ موش است که تا ستون ۵ برگرفته از [۳] است. در این جدول مقدار x معرف طول عمر موش‌ها است و مقادیرش از ۱ تا ۱۴ سال تغییر می‌کند. در ستون چهارم، پنجم و ششم جدول (۱) مقادیر برآورد شده هر طبقه به‌ترتیب به‌کمک روش بیشینه درست‌نمایی (MLE)، روش بیزی (BE) و روش کولبک-لیبلر بر اساس تابع بقا (KLS) محاسبه و ارائه شده است. مقادیر پارامترهای برآورد شده c,λ به‌روش KLS با نقطه شروع ۰/۱۵ و ۰/۱۵ به‌کمک دستور fsolve در نرم‌افزار MATLAB به‌ترتیب ۰/۰۷۵۷ و ۰/۲۸۰۷ هستند. در پایان این مدل‌ها به‌کمک شاخص‌های χ^2 و $[\Delta^4\tilde{q}_x]^2$ که به‌ترتیب مربوط به نیکویی برازش و همواری مدل هستند با هم مقایسه شدند. چنان‌که ملاحظه می‌شود روش KLS برای برازش مدل احتمالاتی در مقابل دو روش

[DOI: 10.52547/mmr.7.2.279]

MLE و BE مناسب نیست چرا که شاید از نظر نیکویی برازش نسبت به دو روش دیگر بهتر عمل کرده اما از نظر هموواری بسیار ضعیف عمل کرده است. (برای جزئیات بیش‌تر به [۷] مراجعه کنید.)

در ادامه، در جدول (۲) مقادیر $\tilde{q}_x$ به‌کمک روش MEMKL برای مقادیر مختلفی از μ تحت قیود مطرح شده در رابطهٔ (۱۳) با k برابر ۵ محاسبه و مدل‌های به‌دست‌آمده به‌کمک شاخص‌های χ^2 و $[\Delta^4\tilde{q}_x]^2$ با هم مقایسه شده‌اند. همهٔ محاسبات این جدول مربوط به حل رابطهٔ (۱۳) هستند که به‌کمک دستور fmincon در نرم‌افزار MATLAB به‌دست‌آمده‌اند و مقدار Exit Flag همه‌ی مدل‌ها ۱ است. نتایج این جدول نشان می‌دهد مدل‌های مربوط به μ برابر ۱ و ۹/۰ بیش‌ترین مقدار χ^2 و کم‌ترین مقدار $[\Delta^4\tilde{q}_x]^2$ و مدل‌های مربوط به μ برابر ۰ و ۰/۱ کم‌ترین مقدار χ^2 و بیش‌ترین مقدار $[\Delta^4\tilde{q}_x]^2$ را در بین مدل‌ها دارند.

جدول ۱. داده‌های مربوط به طول عمر ۲۰۸ موش

x	d_x	$\tilde{q}_x$	MLE	BE	KLS
۱	۳	۰/۰۱۴۴	۰/۰۳۱۱	۰/۰۳۲۱	۰/۰۹۱۹
۲	۳	۰/۰۱۴۴	۰/۰۱۶۸	۰/۰۱۷۳	۰/۱۰۸۴
۳	۶	۰/۰۲۸۸	۰/۰۲۴۴	۰/۰۲۴۹	۰/۱۲۳۱
۴	۶	۰/۰۲۸۸	۰/۰۳۴۹	۰/۰۳۵۴	۰/۱۳۳۰
۵	۱۶	۰/۰۷۶۹	۰/۰۴۹۱	۰/۰۴۹۶	۰/۱۳۴۶
۶	۱۴	۰/۰۶۷۳	۰/۰۶۷۶	۰/۰۶۸	۰/۱۲۴۹
۷	۲۵	۰/۱۲۰۲	۰/۰۹۰۱	۰/۰۹۰۲	۰/۱۰۳۲
۸	۲۰	۰/۰۹۶۲	۰/۱۱۴۴	۰/۱۱۴	۰/۰۷۳۳
۹	۳۲	۰/۱۵۳	۰/۱۳۵۱	۰/۱۳۴۲	۰/۰۴۲۵
۱۰	۲۵	۰/۱۲۰۲	۰/۱۴۳۷	۰/۱۴۲۳	۰/۰۱۸۸
۱۱	۲۷	۰/۱۲۹۸	۰/۱۳۱	۰/۱۲۹۷	۰/۰۱۵۹
۱۲	۱۳	۰/۰۶۲۵	۰/۰۹۵۳	۰/۰۹۴۸	۰/۰۱۱۱
۱۳	۱۱	۰/۰۵۲۹	۰/۰۵۰۱	۰/۰۵۰۵	۰/۰۱۰۱
۱۴	۷	۰/۰۳۳۷	۰/۰۱۶۵	۰/۰۱۷	۰/۰۱۰۰
χ^2			۰/۰۷۵۶	۰/۰۷۳۷	۰/۰۰۲۳
$[\Delta^4\tilde{q}_x]^2$			۰/۰۰۰۸	۰/۰۰۰۷	۲/۴۸۷۷

توجه داریم که برای برازش بر داده‌ها، مدلی مناسب‌تر است که مقدار $G=\mu[\Delta^4\tilde{q}_x]^2+(1-\mu)\chi^2$ کم‌تری داشته باشد. برای انتخاب مدل مناسب، دو مدل اول و آخر را با وجود این‌که هر دو از کم‌ترین مقدار G برخوردارند از مقایسات حذف می‌کنیم زیرا که صرفاً به‌ترتیب فقط به نیکویی برازش و هموواری مدل توجه دارند. سپس مدل‌های مربوط به μ برابر ۰/۸ و ۰/۹ را برای برازش بر داده‌ها مناسب معرفی می‌کنیم زیرا کم‌ترین مقدار G را در بین سایر مدل‌ها به خود اختصاص داده‌اند؛ اما با توجه به این‌که در مدل مربوط به μ برابر ۰/۹ نیز بیش‌ترین توجه به نیکویی برازش است مدل مربوط به μ برابر ۰/۸ را که توازن مناسب‌تری بین دو شاخص مذکور برقرار کرده است را مناسب‌تر معرفی می‌کنیم. به‌طورکلی از نتایج بالا، می‌توان نتیجه گرفت که روش پیشنهادی کاراست و ضریب تعدیل نقش مهمی در مبادله کردن هموواری و نیکویی برازش ایفا می‌کند که باعث ایجاد انعطاف‌پذیری در طبقه‌بندی داده‌ها است.

تنظیم جدول عمر خلاصه ایران در سال ۱۳۹۰ به‌کمک روش MEMKL

در این بخش قصد داریم جدول عمر زنان و مردان ایران را در سال ۱۳۹۰ به‌کمک روش MEMKL تنظیم کنیم. برای این منظور از داده‌های [۱] استفاده می‌کنیم که در جدول ۳ ارائه شده‌اند. نتایج مربوط به استفاده از روش مذکور برای این داده‌ها در جدول ۴ تنظیم شده است. این جدول گویای این موضوع است که در بیش‌تر مواقع به‌ازای تعداد قیود ثابت، مقدار G مدل به‌ازای μ از ۰ تا ۰/۵ افزایش و سپس به‌ازای μ از ۰/۵ تا ۱ کاهش می‌یابد. هم‌چنین واضح

[DOI: 10.52547/mmr.7.2.279]

است مدل‌ها بسیار به مقدار μ وابسته هستند از این‌رو،ا بهتر است به‌طور تجربی انتخاب شود. علاوه بر این، زمانی که μ برابر ۰ است یعنی زمانی که شاخص نیکویی برازش اهمیت کاملی در مقابل شاخص همواری مدل دارد مقادیر G بسیار کوچک‌تر از مواقع دیگر است.

جدول ۲. $\tilde{q}_x$ مربوط به جدول ۱ به‌کمک روش MEMKL برای مقادیر مختلفی از μ

x \ μ	۰	۰.۱	۰.۲	۰.۳	۰.۴	۰.۵	۰.۶	۰.۷	۰.۸	۰.۹	۱
۱	۰/۰۱۴۶	۰/۰۱۴۴	۰/۰۱۴۲	۰/۰۱۴	۰/۰۱۳۹	۰/۰۱۳۷	۰/۰۱۳۵	۰/۰۱۳۳	۰/۰۱۳۲	۰/۰۱۳	۰/۰۱۲۸
۲	۰/۰۱۴۲	۰/۰۱۴۵	۰/۰۱۴۹	۰/۱۵۳	۰/۰۱۵۶	۰/۰۱۶	۰/۰۱۶۳	۰/۰۱۶۷	۰/۰۱۷۱	۰/۰۱۷۴	۰/۰۱۷۸
۳	۰/۰۲۴۸	۰/۰۲۸۲	۰/۰۲۸۱	۰/۰۲۷۸	۰/۰۲۷۶	۰/۰۲۷۴	۰/۰۲۷۲	۰/۰۲۶۹	۰/۰۲۶۶	۰/۰۲۶۳	۰/۰۲۶۱
۴	۰۲۸۸/۰	۰/۰۲۹۷	۰/۰۳۰۷	۰/۰۳۱۷	۰/۰۳۲۶	۰/۰۳۷۷	۰/۰۳۴۷	۰/۰۳۵۷	۰/۰۳۶۷	۰/۰۳۷۸	۰/۰۳۸۸
۵	۰/۰۷۷۴	۰/۰۷۵۵	۰/۰۷۳۷	۰/۰۷۱۲	۰/۰۶۹۱	۰/۰۶۷	۰/۰۶۵	۰/۰۶۳	۰/۰۶۱۱	۰/۰۵۹۱	۰/۰۵۷۲
۶	۰/۰۶۸۲	۰/۰۶۹۵	۰/۰۷۰۸	۰/۰۷۲۱	۰/۰۷۳۶	۰/۰۷۴۹	۰/۰۷۵۹	۰/۰۷۲۲	۰/۰۷۸۳	۰/۰۷۹۵	۰/۰۸۰۷
۷	۰/۱۲۱۶	۰/۱۲	۰/۱۱۸۷	۰/۱۱۷۲	۰/۱۱۵۵	۰/۱۱۴	۰/۱۱۲۶	۰/۱۱۰۸	۰/۱۰۹۳	۰/۱۰۷۷	۰/۱۰۶۱
۸	۰/۰۹۶۷	۰/۰۹۹۶	۰/۱۰۲۴	۰/۱۰۵۲	۰/۱۰۸۲	۰/۱۱۱۲	۰/۱۱۴۲	۰/۱۱۷۴	۰/۱۲۰۵	۰/۱۲۳۷	۰/۱۲۶۷
۹	۰/۱۵۲۸	۰/۱۵۱۲	۰/۱۴۹۵	۰/۱۴۷۸	۰/۱۴۶۱	۰/۱۴۴۴	۰/۱۴۲۷	۰/۱۴۱	۰/۱۳۹۲	۰/۱۳۷۵	۰/۱۳۵۸
۱۰	۰/۱۱۹۵	۰/۱۲۰۳	۰/۱۲۱۳	۰/۱۲۲۴	۰/۱۲۳۳	۰/۱۲۴۱	۰/۱۲۵	۰/۱۲۵۷	۰/۱۲۶۷	۰/۱۲۷۵	۰/۱۲۸۳
۱۱	۰/۱۲۸۸	۰/۱۲۶۷	۰/۱۲۴۴	۰/۱۲۲	۰/۱۱۹۹	۰/۱۱۷۷	۰/۱۱۵۴	۰/۱۱۳۲	۰/۱۱۱۱	۰/۱۰۸۹	۰/۱۰۶۷
۱۲	۰/۰۶۲۳	۰/۰۶۳۸	۰/۰۶۵۵	۰/۰۶۷	۰/۰۶۸۷	۰/۰۷۰۳	۰/۰۷۲	۰/۰۷۳۷	۰/۰۷۵۳	۰/۰۷۷	۰/۰۷۸۷
۱۳	۰/۰۵۳	۰/۰۵۳۱	۰/۰۵۳	۰/۰۵۳	۰/۰۵۲	۰/۰۵۲۸	۰/۰۵۲۸	۰/۰۵۲۶	۰/۰۵۲۵	۰/۰۵۲۴	۰/۰۵۲۲
۱۴	۰/۰۳۳	۰/۰۳۳۶	۰/۰۳۳۴	۰/۰۳۳	۰/۰۳۳	۰/۰۳۲۹	۰/۰۳۲۷	۰/۰۳۲۶	۰/۰۳۲۴	۰/۰۳۲۳	۰/۰۳۲۲
χ^2	۶e-۰۵	۰/۰۰۰۴	۰/۰۰۱۴	۰/۰۰۳	۰/۰۰۵۴	۰/۰۰۸۳	۰/۰۱۱۹	۰/۰۱۶۳	۰/۰۲۱۱	۰/۰۲۶۸	۰۳۳۱/۰
$[\Delta^4\tilde{q}_x]^2$	۰/۶۱۳۱	۰/۵۰۱	۰/۳۹۶۵	۰/۳۰۴۳	۰/۲۲۲۴	۰/۱۵۴۹	۰/۱۰۰۷	۰/۰۵۵۹	۰/۰۲۵۵	۰/۰۰۶۴	۰/۰۰۰۱
G	۶e-۰۵	۰/۰۵۰۴	۰/۰۸۰۴	۰/۰۹۳۴	۰/۰۹۲۲	۰/۰۸۱۶	۰/۰۶۵۲	۰/۰۴۴	۰/۰۲۶۴	۰/۰۰۸۵	۰/۰۰۰۱

جدول ۳. داده‌های مربوط به مرگ‌ومیر زنان و مردان ایران سال ۱۳۹۰

جنسیت	مردان		زنان	
X	d_x	$\tilde{q}_x$	d_x	$\tilde{q}_x$
۰	۲۳۸۹	۰/۰۲۳۸۹	۲۰۱۹	۰/۰۲۰۱۹
۱	۳۹۸	۰/۰۰۳۹۸	۳۷۷	۰/۰۰۳۷۷
۵	۲۴۷	۰/۰۰۲۴۷	۱۸۴	۰/۰۰۱۸۴
۱۰	۲۰۷	۰/۰۰۲۰۷	۱۵۰	۰/۰۰۱۵
۱۵	۳۹۸	۰/۰۰۳۹۸	۲۴۷	۰/۰۰۲۴۷
۲۰	۵۵۰	۰/۰۰۵۵	۳۴۸	۰/۰۰۳۴۸
۲۵	۵۴۲	۰/۰۰۵۴۲	۴۳۳	۰/۰۰۴۳۳
۳۰	۶۰۸	۰/۰۰۶۰۸	۵۴۴	۰/۰۰۵۴۴
۳۵	۷۹۰	۰/۰۰۷۹	۷۳۴	۰/۰۰۷۳۴
۴۰	۱۱۹۵	۰/۰۱۱۹۵	۱۰۶۱	۰/۰۱۰۶۱
۴۵	۱۹۸۶	۰/۰۱۹۸۶	۱۶۴۴	۰/۰۱۶۴۴
۵۰	۳۲۱۲	۰/۰۳۲۱۲	۲۴۶۸	۰/۰۲۴۶۸
۵۵	۵۱۶۹	۰/۰۵۱۶۹	۳۷۰۲	۰/۰۳۷۰۲
۶۰	۷۶۹۵	۰/۰۷۶۹۵	۵۶۶۴	۰/۰۵۶۶۴
۶۵	۱۰۹۷۱	۰/۱۰۹۷۱	۸۹۰۴	۰/۰۸۹۰۴
۷۰	۱۴۵۶۵	۰/۱۴۵۶۵	۱۳۲۸۸	۰/۱۳۲۸۸
۷۵	۱۶۹۶۱	۰/۱۶۹۶۱	۱۷۵۵۴	۰/۱۷۵۵۴
۸۰	۳۲۱۱۶	۰/۳۲۱۱۶	۴۰۶۷۹	۰/۴۰۶۷۹

[DOI: 10.52547/mmr.7.2.279]

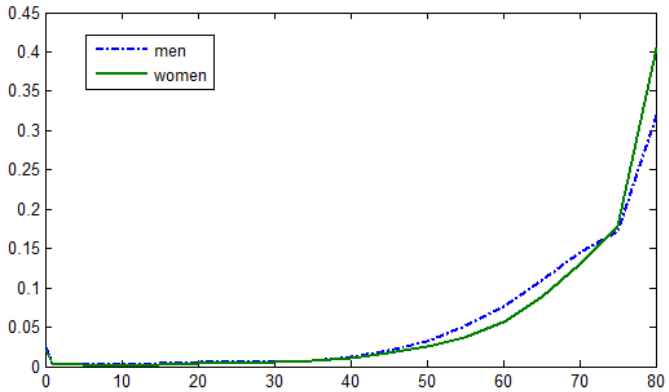
در این‌جا با حذف دو مدل ME و MKL (به‌تنهایی) و مقایسه مقادیر G باقی‌مانده، مقدار k و μ را به‌ترتیب ۴ و ۰/۱ انتخاب می‌کنیم. در پایان بر اساس مدل منتخب، جدول عمر سال ۱۳۹۰ زنان و مردان ایران را در جدول ۵ تنظیم و روند مرگ و میر سنین مشخص شده آنها را در نمودار شکل ۱ رسم کرده‌ایم. چنان‌که در این شکل ملاحظه می‌شود مقادیر برآورد شده احتمالات مرگ و میر مردان و زنان ایرانی بسیار نزدیک به‌هم است و هر دو روند صعودی دارند و در سنین ۷۵ تا ۸۰ سالگی احتمال مرگ و میر زنان بیش‌تر از مردان تخمین زده شده است.

جدول ۴. $G\times10^{-6}$ مدل‌های مربوط به داده‌های جدول ۳ به‌کمک روش MEMKL برای مقادیر مختلفی از μ و k

جنسیت	مردان					زنان				
μ	k					k				
	۱	۲	۳	۴	۵	۱	۲	۳	۴	۵
۰	۱	۰	۱	۳۵	۱۱۳۷۵۴	۰	۰	۱	۱۳	۱۸۶۶۶۶
۰/۰	۳۵۰۲۹	۲۶۱۶۷۶	۲۳۵۷	۲۳۴۸	۱۰۳۵۷۰	۵۴۵۴	۳۸۳۱	۳۷۵۸	۳۴۵۷	۲۵۹۶۴۹
۲/۰	۸۶۱۶۷	۵۵۹۴۹۶	۴۴۳۱	۴۲۸۴	۸۶۵۲۱	۱۲۷۹۵	۷۲۳۴	۶۸۵۳	۶۴۲۸	۲۱۷۷۸۲
۳/۰	۱۴۶۸۰۴	۸۴۹۲۸۴	۶۲۱۷	۴۹۸۵	۹۲۴۶۰	۲۱۵۱۱	۹۹۳۵	۹۰۹۶	۸۰۰۸	۲۱۴۲۸۵
۴/۰	۲۱۱۰۵۶	۱۰۹۹۶۹	۷۳۴۶	۶۳۶۷	۷۵۴۶۳	۳۰۸۸۲	۱۱۷۱۶	۱۰۵۲۱	۱۰۲۰۳	۱۵۴۳۹۱
۵/۰	۲۷۴۴۹۲	۱۲۹۸۱۲	۸۰۳۴	۶۷۲۶	۵۹۲۵۸	۴۰۴۹۸	۱۲۸۱۱	۱۰۹۲۴	۱۰۰۷۷	۱۳۱۲۰۵
۶/۰	۳۲۸۱۳۲	۱۳۹۷۹۶	۸۰۲۷	۶۶۵۶	۴۷۱۳۷	۴۹۱۸۱	۱۲۸۹۹	۱۰۶۰۱	۱۰۳۴۸	۱۰۱۸۱
۷/۰	۳۵۶۸۱۳	۱۳۶۴۷۷	۷۳۷۵	۵۸۸۷	۶۳۱۲	۵۴۵۷۹	۱۱۹۳۴	۹۲۵۷	۹۵۲۴	۹۲۴۵
۸/۰	۳۳۶۷۸۶	۱۱۷۴۲۵	۵۸۸۰	۴۷۷۰	۵۱۹۲	۵۴۳۹۹	۹۵۷۱	۷۱۴۹	۸۱۲۴	۷۸۶۶
۹/۰	۲۳۷۰۱۲	۷۴۹۱۳۹	۳۴۸۹	۳۰۱۹	۳۳۶۰	۴۰۳۹۰	۶۰۱۲	۴۱۰۴	۶۰۹۴	۵۸۲۸
۱	۵	۱۱۳	۵۱	۶۱۹	۹۱۸	۲۰	۵۹۲	۱۰۵	۳۴۹۳	۲۲۹۱۱

نتیجه‌گیری

در این مقاله از معیارهای نظریه اطلاع برای بررسی توزیع بقا و توزیع احتمال مرگ و میر کمک گرفتیم. از جملهٔ این معیارها اصل بیشینه آنتروپی شانون و اصل کمینه آنتروپی کولبک-لیبلر هستند. در بررسی مرگ و میر یکی از ابزارهای موجود طبقه‌بندی اطلاعات، جدول عمر است و در طبقه‌بندی داده‌ها هموواری و نیکویی برازش مدل برازش داده شده اهمیت ویژه‌ای دارد. این دو معیار به‌ترتیب معادل با اصل بیشینه آنتروپی شانون و اصل کمینه آنتروپی کولبک لیبلر هستند. ما با کمینه کردن مدل ترکیبی دو اصل فوق، توزیع احتمالاتی مرگ و میر را برآورد کردیم و از این روش برای تنظیم جدول عمر استفاده کردیم. بر این اساس جدول عمر زنان و مردان ایران در سال ۱۳۹۱ تنظیم شده است. به‌طورکلی، می‌توان نتیجه گرفت که روش پیشنهادی کاراست و ضریب تعدیل نقش مهمی در مبادله کردن هموواری و نیکویی برازش ایفا می‌کند که باعث ایجاد انعطاف‌پذیری در طبقه‌بندی داده‌ها است.



شکل ۱. نمودار روند احتمال مرگ و میر زنان و مردان ایران در سال ۱۳۹۰، برآورد شده به‌کمک روش MEMKL

[DOI: 10.52547/mmr.7.2.279]

جدول ۵. جدول عمر زنان و مردان ایران ۱۳۹۰ به‌روش MEMKL

جنسیت	مردان					زنان				
x	l_x	d_x	$\hat{q}_x$	${}_dm_x$	e_x	l_x	d_x	$\hat{q}_x$	${}_dm_x$	e_x
۰	۱۰۰۰۰۰	۲۳۸۹	۰۲۳۲۸۴/۰	۰۲۳۵۵۹/۰	۶۵/۷۰	۱۰۰۰۰۰	۲۰۱۹	۰۱۹۷۳/۰	۰۱۹۹۲۷/۰	۴۱/۷۳
۱	۹۷۶۱۱	۳۹۸	۰۰۴۵۳/۰	۰۰۴۵۴/۰	۳۸/۷۱	۹۷۹۸۱	۳۷۷	۰۰۴۱۷۷/۰	۰۰۴۱۸۶/۰	۹۲/۷۳
۵	۹۷۲۱۳	۲۴۷	۰۰۲۷۰۷/۰	۰۰۲۷۱۱/۰	۶۶/۶۷	۹۷۶۰۴	۱۸۴	۰۰۱۹۸۶/۰	۰۰۱۹۸۸/۰	۲۰/۷۰
۱۰	۹۶۹۶۶	۲۰۷	۰۰۲۱۵۹/۰	۰۰۲۱۶۱/۰	۸۳/۶۲	۹۷۴۲۰	۱۵۰	۰۰۱۵۹۶/۰	۰۰۱۵۹۷/۰	۳۳/۶۵
۱۵	۹۶۷۵۲	۳۹۸	۰۰۳۸۴۱/۰	۰۰۳۸۴۸/۰	۹۶/۵۷	۹۷۲۷۰	۲۴۷	۰۰۲۴۱۸/۰	۰۰۲۴۲۱/۰	۴۳/۶۰
۲۰	۹۶۳۶۱	۵۵۰	۰۰۵۱۹/۰	۰۰۵۲۰۴/۰	۱۹/۵۳	۹۷۰۲۲	۳۴۸	۰۰۳۳۵۶/۰	۰۰۳۳۶۲/۰	۵۷/۵۵
۲۵	۹۵۸۱۰	۵۴۲	۰۰۵۲۸/۰	۰۰۵۲۹۴/۰	۴۸/۴۸	۹۶۶۷۵	۴۳۳	۰۰۴۱۷۵/۰	۰۰۴۱۸۴/۰	۷۶/۵۰
۳۰	۹۵۲۶۸	۶۰۸	۰۰۶۰۶۳/۰	۰۰۶۰۸۱/۰	۷۴/۴۳	۹۶۲۴۱	۵۴۴	۰۰۵۳۱۴/۰	۰۰۵۳۲۸/۰	۹۸/۴۵
۳۵	۹۴۶۶۰	۷۹۰	۰۰۸۰۲۴/۰	۰۰۸۰۵۶/۰	۰۰/۳۹	۹۵۶۹۸	۷۳۴	۰۰۷۲۹/۰	۰۰۷۳۱۷/۰	۲۳/۴۱
۴۰	۹۳۸۷۰	۱۱۹۵	۰۱۲۲۱۳/۰	۰۱۲۲۸۸/۰	۳۱/۳۴	۹۴۹۶۳	۱۰۶۱	۰۱۰۶۹۹/۰	۰۱۰۷۵۷/۰	۵۲/۳۶
۴۵	۹۲۶۷۵	۱۹۸۶	۰۲۰۰۷۸/۰	۰۲۰۲۸۲/۰	۷۲/۲۹	۹۳۹۰۲	۱۶۴۴	۰۱۶۷۴۱/۰	۰۱۶۸۸۳/۰	۹۱/۳۱
۵۰	۹۰۶۸۹	۳۲۱۲	۰۳۲۴۸۲/۰	۰۳۳۰۲۱/۰	۳۱/۲۵	۹۲۲۵۸	۲۴۶۸	۰۲۵۰۳۳/۰	۰۲۵۳۵۲/۰	۴۳/۲۷
۵۵	۸۷۴۷۷	۵۱۶۹	۰۵۱۴۹۳/۰	۰۵۲۸۶۶/۰	۱۴/۲۱	۸۹۷۹۱	۳۷۰۲	۰۳۷۴۸۸/۰	۰۳۸۲۰۹/۰	۱۱/۲۳
۶۰	۸۲۳۰۹	۷۶۹۵	۰۷۶۸۳۴/۰	۰۷۹۹۴۶/۰	۳۰/۱۷	۸۶۰۸۹	۵۶۶۴	۰۵۶۰۹۴/۰	۰۵۷۷۲۹/۰	۹۹/۱۸
۶۵	۷۴۶۱۳	۱۰۹۷۱	۱۰۸۶۵/۰	۱۱۵۰۱۸/۰	۸۱/۱۳	۸۰۴۲۵	۸۹۰۴	۰۸۸۳۴۶/۰	۰۹۲۴۹۵/۰	۱۴/۱۵
۷۰	۶۳۶۴۲	۱۴۵۶۵	۱۴۵۱۶۸/۰	۱۵۶۸۵/۰	۷۳/۱۰	۷۱۵۲۱	۱۳۲۸۸	۱۳۱۲۳۱/۰	۱۴۰۶۷۸/۰	۶۹/۱۱
۷۵	۴۹۰۷۷	۱۶۹۶۱	۱۷۱۸۹۱/۰	۱۸۸۶۱/۰	۱۵/۸	۵۸۲۳۳	۱۷۵۵۴	۱۷۸۸/۰	۱۹۶۹۸۹/۰	۷۵/۸
۸۰	۳۲۱۱۶	۳۲۱۱۶	۳۲۰۱۱۳/۰	۳۸۵۸۲۹/۰	۱۱/۶	۴۰۶۷۲	۴۰۶۷۹	۴۰۵۵۲۶/۰	۵۲۰۰۷۸/۰	۴۱/۶

منابع

۱. طه حسینی، "ساخت جدول عمر سالانه برای ایران"، مرکز آمار ایران، پژوهشکدهٔ آمار (۱۳۹۲).

2. Alicja S. N., William F. S., "An extension of the Whittaker-Henderson method of graduation", Scand. Actuar J, 1 (2012) 70-79.
3. Ananda M. M, Dalpatadu R. J., Singh A. K., "estimating parameters of the force of mortality in actuarial studies", Actuar. Res. Clear. House, 1 (1993) 129-141.
4. Bohlmann G., "Ein Ausgleichungs ProblME", In Nachrichten von der Königl Gesellschaft der Wissenschaften zu Göttingen, MathMEatisch-physikalische Klasse; Horstmann, L., Ed., Commissionsverlag der Dieterich'schen Universitätsbuchhandlung: Göttingen, Germany, (1899) 260-271.
5. Henderson R., "A new method of graduation", Trans. Actuar. Soc. Am., 25 (1924) 29-53.
6. Henderson R., "Further rMEarks on graduation", Trans. Actuar. Soc. Am., 26 (1925) 52-74.
7. He D., Huang Q. Gao J., "A New Entropy Optimization Model for Graduation of Data in Survival Analysis", Entropy, 14 (2012) 1306-1316.
8. Kapur J. N., Kesavan H. K., "Entropy optimization Principles with Applications", AcadMEic Press Inc.: San Diego, CA, USA, (1992).
9. Kullback S., Leibler R. A., "on Information and Sufficiency", Annals of MathMEatical Statistics, 22 (1) (1951) 79-86.
10. Liu J., "Information theoretic content and probability", Ph.D.Thesis, University of Florida, USA, (2007) 25-26.

[DOI: 10.52547/mmr.7.2.279]

11. M. Cover T., A. Thomas J., "Elements of information theory", Second Edition, Published by John Wiley & Sons, Inc., Hoboken, New Jersey (2006).
12. Schoen R., "Modeling multigroup populations", plenum press, New York, (1899).
13. Whittaker E. T., "On a new method of graduation", Proc. Edinb. Math. Soc, 41 (1923) 63-75.
14. Yari G. H., Mirhabibi A., Saghafi A., "Estimation of the Weibull parameters by Kullback Leibler divergence of Survival functions", Appl. Math. Inf. Sci, 7, No. 1 (2013) 187-192.

[DOI: 10.52547/mmr.7.2.279]