مجله علوم آماری، پاییز و زمستان ۱۳۸۶
جلد ۱، شماره ۲، ص ۱۵۷-۱۷۰

تحلیل پیشامدهای بازگشتی با استفاده از مدل شکنندگی وابسته به زمان

محمودرضا گوهری [۱]، محمود محمودی [۲]، کاظم محمد [۲]، عین‌اله پاشا [۳]
[۱] گروه آمار و ریاضی، دانشگاه علوم پزشکی ایران، [۲] گروه آمارحیاتی و اپیدمیولوژی، دانشگاه علوم پزشکی تهران، [۳] گروه آمار، دانشگاه تربیت معلم

تاریخ دریافت: ۱۳۸۶/۵/۲۵ تاریخ آخرین بازنگری: ۱۳۸۶/۱۲/۱۹

چکیده : داده‌های بازگشتی یکی از انواع مهم داده‌های بقا هستند که ویژگی عمدهٔ آنها همبستگی بین مشاهدات است. این ویژگی استفاده از مدل‌های بقا مانند رگرسیون کاکس که در آنها مستقل بودن مشاهدات یکی از فرضیات اصلی مدل است را ناممکن می‌سازد. مدل‌های شکنندگی یکی از رویکردهای عمده برای تحلیل داده‌های بازگشتی هستند. در این مدل‌ها یک متغیر شکنندگی ثابت به عنوان اثر خصوصیات فردی یا نماینده همه عواملی که سبب وابستگی بین مشاهدات می‌شوند به صورت ضربی وارد تابع خطر می‌شود. در مقاله حاضر یک مدل شکنندگی با اثر تصادفی وابسته به زمان بر مبنای مدل‌های خطر نیمه پارامتری کاکس و با مولفه‌های شکنندگی که دارای فرایند گامای تکه‌ای هستند معرفی شده و کارایی آن در یک مطالعه شبیه‌سازی با یک مدل شکنندگی گاما با اثرات ثابت مقایسه گردیده است.

واژه‌های کلیدی : پیشامد بازگشتی، مدل شکنندگی، فرایند پواسن آمیخته.

آدرس الکترونیک مسئول مقاله: محمودرضا گوهری، mgohari@iums.ac.ir

۱ مقدمه

در بسیاری از مطالعات پزشکی با بیماری‌ها یا عوارضی مواجه هستیم که بیمار پس از درمان و بهبودی، مجددا امکان ابتلا به بیماری را دارا می‌باشد. به این گونه از بیماری‌ها عودپذیر یا بازگشتی [۱] می‌گویند. عفونت‌های مکرر ناشی از دستگاه دیالیز، تعداد حملات صرع، اپیزودهای مکرر بیماری‌های روانی یا بستری شدن متوالی در بیمارستان، مثال‌هایی ازپیشامدهای بازگشتی هستند، که به دو نوع همسان و غیر همسان تقسیم می‌شوند. در نوع همسان رخداد مجدد بیماری از نوع قبلی است، اما در نوع غیر همسان بیماری در ظهور مجدد شکل یا شدت متفاوتی نسبت به قبل دارد. پیشامدهای غیر همسان معمولا توسط مدل‌های چندحالته [۲] تحلیل می‌شوند. در این مقاله ما با داده‌های بازگشتی نوع همسان کار می‌کنیم. مدل‌هایی که برای تحلیل داده‌های بازگشتی همسان به کار می‌روند یا تعداد پیشامدها یا فاصله زمانی بین آنها را مورد بررسی قرار می‌دهند. هنگامی که تعداد پیشامدها در یک محدوده زمانی هدف مطالعه باشد، داده‌ها بر اساس فرایند پواسن مدل‌بندی می‌شوند و در مطالعاتی که زمان بین پیشامدها مورد توجه باشد، مدل‌هایی بر اساس فرایند تجدید [۳] مورد استفاده قرار می‌گیرند. صرف نظر از متغیر مورد مطالعه، داده‌های بازگشتی دارای این ویژگی هستند که بین زمان رخداد پیشامدهای متوالی مربوط به یک فرد همبستگی وجود دارد. این همبستگی ناشی از دو منبع عمدهٔ ویژگیهای فردی و اثر پیشامدهای قبلی است. ویژگی‌های وراثتی و ژنتیکی، محیط رشد و زندگی از عواملی هستند که سبب تفاوت افراد می‌شوند. از سوی دیگر پیشامدهای قبلی نیز ممکن است در زمان وقوع پیشامدهای جدید تاثیر گذار باشند. اگر بدون توجه به همبستگی درونی بین زمان پیشامدهای فرد از مدل‌های معمول مانند رگرسیون خطر متناسب [۴]، برای تحلیل داده‌ها استفاده شود برآوردها اریب و واریانس برآورد پارامترها کمتر از مقدار واقعی برآورد می‌شوند (کاکس، ۱۹۸۳). در

---
[۱] Recurrent
[۲] Multistate
[۳] Regeneration
[۴] Proportional Hazard

مدل‌های رگرسیونی نیمه پارامتری به منظور وارد کردن همبستگی درونی داده‌ها دو رویکرد عمده وجود دارد. در رویکرد اول از یک سو واریانس استوار جایگزین واریانس پارامترها شده و از سوی دیگر با قرار دادن محدودیت‌هایی روی مجموعه خطر و مقیاس زمانی مورد استفاده، وابستگی بین پیشامدها به صورت غیر مستقیم وارد مدل می‌گردد (ترنو و گرمش، ۲۰۰۰). در این مدل‌ها تابع خطر برحسب رتبه پیشامدها طبقه‌بندی و شکل تابع خطر مطابق با تعداد پیشامدهای قبلی تغییر می‌کند. ازمهمترین این مدل‌ها می توان به مدل شرطی (پرنتیس و همکاران، ۱۹۸۱) و مدل حاشیه‌ای (وی و همکاران، ۱۹۸۹) اشاره نمود. مدل‌های غیر طبقه‌ای نیز وجود دارند که پیشامدها بدون توجه به تعداد یا زمان وقوع رخدادهای قبلی و بر اساس مدل پواسن مدل بندی می‌شوند (آندرسن و گیل، ۱۹۸۲). کوک (۲۰۰۲) بحث کاملی از انواع مدل‌های تحلیل پیشامدهای بازگشتی ارائه نموده است.

در رویکرد دوم همبستگی بین پیشامدها از طریق یک اثر تصادفی وارد مدل می‌شود. این مدل‌ها به نام مدل‌های اثرات تصادفی یا مدل‌های شکنندگی[۵] شناخته می‌شوند. در مدل‌های شکنندگی برای هر فرد یک اثر تصادفی اختصاصی در نظر گرفته می‌شود که بیان کننده خصوصیات فردی و اختصاصی هر فرد است. این مدل‌ها در بخش ۲ به اختصار معرفی می‌شوند. مدل‌های شکنندگی درمقاله لاولس (۲۰۰۲)، آلن (۱۹۹۸) و آکز (۱۹۹۲) آورده شده است. در مدل‌های شکنندگی اثر اختصاصی فرد در طول زمان ثابت فرض می‌شود. اما در برخی مواقع این فرض چندان واقعی به نظر نمی‌رسد، زیرا خصوصیات افراد در طول زمان ثابت و یکنواخت نیست و دستخوش تغییر می‌شود.

تعداد مدل‌هایی که دارای متغیر شکنندگی غیر ثابت در طول زمان هستند بسیار اندک است. پایک (۱۹۹۴) یک مدل شکنندگی برای داده‌های موازی پیشنهاد نمود که متغیر شکنندگی توسط خطاهای وابسته به زمان مقادیر متفاوتی در طول زمان اختیار می‌کنند. یائو و مک گقیلکریست (۱۹۹۸) یک مدل شکنندگی را در قالب مدل‌های خطی تعمیم یافته پیشنهاد کردند. ماندا (۲۰۰۵) یک مدل بیزی را برای داده‌های بازگشتی با متغیر شکنندگی وابسته به زمان پیشنهاد نمود. هوگارد (۲۰۰۰)

[۵] Frailty

نیز برخی از مدل‌های شکنندگی وابسته به زمان را معرفی نموده است.

یکی از دلایل عمده توسعه نیافتن مدل‌های شکنندگی وابسته به زمان، پیچیدگی محاسبات آنها به ویژه محاسبه تابع درستنمایی حاشیه‌ای است که باعث شده برخی از این مدل‌ها در قالب مدل‌های بیزی مطرح شود، زیرا در این مدل‌ها توسط روش‌های MCMC[۶] انتگرال گیری روی تابع درستنمایی به وسیله روش‌های عددی قابل انجام است (کلایتون، ۱۹۹۱، سینها و دِی، ۱۹۹۷).

در مقاله حاضر یک مدل شکنندگی وابسته به زمان پیشنهاد می‌گردد. این مدل که مدل شکنندگی جمعی تکه‌ای[۷] نامیده شده در بخش ۳ معرفی می‌شود. در این مدل متغیر شکنندگی دارای یک فرایند گاما است که در طول بازه‌های زمانی دارای مقادیر ثابت و در نقاط برش دارای نموهایی می‌باشد. انتگرال تابع درستنمایی مدل پیشنهادی دارای شکل بسته می‌باشد. تابع درستنمایی مدل در بخش۴ مورد بررسی قرار می‌گیرد. میزان کارایی مدل بر اساس میزان اریبی و واریانس برآوردگرها توسط یک مطالعه شبیه‌سازی در بخش ۵ نمایش داده می‌شود.

۲ مدل‌های شکنندگی

در مدل‌های آماری همیشه عوامل ناشناخته یا اندازه‌گیری نشده‌ای وجود دارند که بر رفتار متغیر پاسخ تاثیر دارند اما در مدل وارد نشده‌اند. در داده‌های بازگشتی این عوامل را می‌توان به عنوان یک متغیر پنهان[۸] در نظر گرفت که ویژگیهای فردی و اختصاصی افراد را در بر دارند و سبب وابستگی بین زمان و تعداد پیشامدهای متوالی فرد می‌شوند. اگر این همبستگی درونی نادیده گرفته شود تعداد پیشامدهای هر فرد متناظر با یک فرایند پواسن غیر همگن است.

اگر فرد iام در طول زمان$[\circ, \tau_i)$ مشاهده شود و تابع نشانگر$Y_i(t) = I(t \le t_i)$ نشان دهنده وضعیت فرد در مجموعه خطر باشد، آنگاه فرایند تعداد پیشامدهای فرد iام در طول زمان مشاهده، $\{N_i(u), u \ge \circ\}$ ، یک فرایند پواسن با تابع شدت

[۶] Monte Carlo Markov Chain

[۷] Piecewise additive frailty

[۸] Latent

$Y_i(t)\rho_i(t)$ است که در آن $\rho_i(t)$ شدت فرایند است که می‌توان آنرا تابعی از متغیرهای مستقل به صورت $\rho_i(t) = \rho_\circ(t)g(x,\beta)$ در نظر گرفت. تابع $g(x,\beta)$ تابعی p بعدی از متغیرهای کمکی است که معمولا به صورت $\exp(x_i',\beta)$ فرض می‌شود تا تناظری با تابع خطر رگرسیون کاکس ایجاد شود.

تحت فرایند پواسن انتظار می‌رود که میانگین و واریانس داده‌ها با یکدیگر برابر باشند، اما در عمل معمولا این شرایط برقرار نیست و داده‌ها دارای واریانس بیشتری هستند. این مقدار واریانس اضافی به عامل یا عواملی نسبت داده می‌شود که سبب وابستگی داده‌ها با یکدیگر شده و فرض مستقل بودن آنها را نقض می‌کنند. برای تبیین این واریانس اضافی یک مدل پواسن آمیخته[۹] (یا بیش پراکنده[۱۰]) به داده‌ها برازش می‌شود (لاولس، ۱۹۸۷). در این مدل یک متغیر غیر قابل مشاهده شده w_i به عنوان نماینده همه عوامل ناشناخته، اندازه‌گیری نشده یا غیر قابل اندازه‌گیری که مخصوص فرد بوده و باعث همبستگی داده‌های یک فرد می‌شوند وارد مدل می‌گردد. فرض می‌شود این متغیر تصادفی(یا شکنندگی) دارای اثر ضربی روی شدت فرایند است. در این حالت فرایند $N_i(t)$ به شرط معلوم بودن x_i و w_i یک فرایند پواسن با شدت زیر است.

$$w_i\rho_\circ(t)\exp(x_i'\beta) \qquad (1)$$

متغیرهای شکنندگی w_i مستقل، هم توزیع، دارای توزیع $G(w_i)$ و مستقل از فرایند سانسور شدن فرض می‌شوند. تابع درستنمایی کامل فرایند پواسن با شدت رابطه (۱) به صورت

$$L = \Pi_{i=1}^{\infty} \int_\circ^\infty \Pi_{j=1}^{n_i} \frac{\rho_\circ(t_{ij})}{\mu_\circ(t_{ij})}(w_i\mu_i(\tau_i))^{n_i}\exp(-w_i\mu_i(\tau_i))dG(w_i)$$

است که در آن $\mu_i(\tau_i) = \int_\circ^{\tau_i}\rho_i(t)dt$. با دقت در این رابطه دیده می‌شود که برای برآورد پارامترها فقط اطلاع از تعداد پیشامدها کافی است و نیازی به دانستن زمان دقیق پیشامدها نیست. در این مشاهدات افراد به شرط معلوم بودن متغیر شکنندگی، مستقل از یکدیگر هستند.

---
[۹] Mixed Poisson

[۱۰] Over Dispersed

برای توزیع متغیرهای شکنندگی توزیع‌های مثبت مختلفی مانند گاما، لگ نرمال، پایدار مثبت[۱۱] (هوگارد، ۱۹۸۶) در نظر گرفته می‌شود. از میان توزیع‌های مختلف، توزیع گاما به دلیل خواص ریاضی مناسبی که دارد بیشتر از سایر توزیع‌ها به کار برده می‌شود.

۳ مدل شکنندگی جمعی تکه‌ای

در مدل‌های شکنندگی اثر تصادفی در طول زمان ثابت فرض می‌شود. این فرض در برخی از مطالعات واقعی به نظر نمی‌رسد. زیرا در حالی که برخی از ویژگیهای افراد مربوط به وراثت هستند، سایر خصوصیات تابعی از محیط و رفتارهای فردی است که ممکن است در طول زمان دچار تغییر و تحول شوند.

عوامل وراثتی سبب ایجاد همبستگی طولانی مدت یا پایدار می‌شوند، اما سایر عوامل ناشناخته و اندازه گیری نشده می‌توانند باعث وابستگی کوتاه‌مدت بین زمان پیشامدها باشند. بنابر این مدل‌هایی که متغیر شکنندگی تابعی از زمان بوده و به جای یک مقدار ثابت دارای یک فرایند تصادفی باشد بهتر می‌توانند رفتار تابع خطر را در طول زمان توصیف نمایند. مدل پیشنهادی ما یک مدل شکنندگی با وابستگی کوتاه‌مدت است که متغیر شکنندگی دارای یک فرایند گامای تکه‌ای است. فرض کنیم فرد iام در طول زمان $[\circ, \tau_i]$ مشاهده و در زمان های $t_{i\backslash} < \ldots < t_{in_i}$ پیشامد مورد مطالعه را تجربه کرده است. اگر a_K طولانی‌ترین زمان مشاهده در مطالعه باشد، فاصله$[\circ, a_K]$ را می‌توان توسط نقاط برش $\circ < a_1 < \ldots < a_K$ به K بازه زمانی $I_k = (a_{k-1}, a_k]$، $k = 1, \ldots, K$ تقسیم نمود. با تعریف تابع نشانگر $w_k = I_k(a_{k-1} < t < a_k)$، تعداد پیشامدهای فرد iام در بازه زمانی kام برابر است با

$$n_{ik} = \sum_{j=1}^{n_i} w_k(t_{ij}) \qquad i = 1, \ldots, m \qquad k = 1, \ldots, K$$

اگر z_k اثر تصادفی فرد iام در زمان kام باشد، آنگاه شدت فرایند برابر

$$\rho_{ik}(t \mid z_{ik}, x_i) = z_{ik} \exp(\Phi_k + x_i'\beta) \qquad (۲)$$

[۱۱] Positive Stable

است، که در آن x_i بردار p بعدی متغیرهای کمکی، β بردار p بعدی از پارامترها و Φ_k لگاریتم شدت خطر پایه در بازه kام است. اثر تصادفی z_{ik} در رابطه (۲) را می‌توان به دو اثر اختصاصی فرد و اثر اختصاصی فرد-زمان به صورت

$$z_{ik} = w_i + u_{ik} \tag{۳}$$

تقسیم نمود، که در آن w_i اثر اختصاصی فرد iام است که در طول زمان بین مشاهدات فرد به اشتراک گذاشته می‌شود و سبب وابستگی طولانی مدت بین آنها می شود و u_{ik} اثر اختصاصی فرد-زمان است. این اثر در عین حال که به فرد iام اختصاص دارد در بازه‌های مختلف مقادیر متفاوتی را اختیار می‌کند. با فرض آنکه متغیرهای w_i و u_{ik} متقابلاً مستقل و دارای توزیع گاما با گشتاورهای $E(w_i) = ۱$، $Var(w_i) = \gamma_۰$، $E(u_{ik}) = ۱$ و $Var(u_{ik}) = \gamma_k$ باشند، تعداد پیشامدها دارای توزیع شرطی

$$n_{ik} \mid z_{ik} \simeq Poisson\big(z_{ik} exp(x_i'\beta + \Phi_k)\big)$$

خواهد بود. در رابطه (۲) اثر فردی را می توان به صورت آمیخته با اثر بازه اول نیز در نظر گرفت. یعنی در بازه اول اثر همه عوامل ثابت فرد مانند عوامل ژنتیکی و وراثتی حضور دارند و در بازه‌های بعدی اثر اختصاصی فرد-زمان آشکار می‌شود. در این حالت شدت فرایند در رابطه (۳) برابر است با

$$\rho_{ik}(t \mid z_{ik}, x_i) = u_{ik} \exp(\Phi_k + x_i'\beta) \qquad k = ۱, \ldots, K$$

۴ برآورد پارامترها

مدل‌های اثرات تصادفی جزو مدل‌های خطی تعمیم‌یافته محسوب شده و یک مدل تمام پارامتری هستند. در نتیجه روش ماکسیمم درستنمایی برای برآورد پارامترهای این مدل‌ها استفاده می‌شود. در حالت کلی اگر $g(u_i, \theta)$ تابع چگالی احتمال بردار متغیرهای تصادفی u_i و $f(y_i \mid u_i, \beta)$ تابع چگالی شرطی پاسخ به شرط متغیر

تصادفی باشد، تابع درستنمایی کامل مدل برابر است با

$$L(\beta,\theta) = \int_{dim(u_i)} f(y_i \mid u_i) g(u_i \mid \theta) du_i$$

برای برآورد پارامترها بر اساس روش ماکسیمم درستنمایی دو رویکرد عمده وجود دارد. در رویکرد اول متغیر تصادفی (شکنندگی) یک پارامتر اخلال محسوب شده و توسط انتگرال‌گیری حذف می‌شود. در رویکرد دوم اثرات تصادفی u_i نیز مورد توجه بوده و مانند اثرات ثابت (پارامترهای رگرسیونی) برآورد می‌شوند. پارامترهای مدل شکنندگی جمعی تکه‌ای از تابع درستنمایی حاشیه‌ای برآورد می‌شوند. تابع درستنمایی حاشیه‌ای مدل را می‌توان به صورت حاصلضرب درستنمایی k بازه به صورت:

$$\begin{aligned} L(t_{ij} \mid x_i, w_i, u_{ik}) &= \int \Pi_{k=1}^{K} \Pi_{i=1}^{m} (z_{ik} e^{\Phi_k + x_i'\beta})^{n_{ik}} \exp(z_{ik} e_{ik} e^{\Phi_k + x_i'\beta}) \\ &\times \quad g(w_i) \Pi_{k=1}^{K} g(u_{u_{ik}}) dw_i du_i \end{aligned}$$

نوشت که در آن e_{ik} طول مدت مشاهده فرد iام در بازه kام به صورت زیر است.

$$e_{ik} = \begin{cases} \tau_i - a_{k-1} & a_{k-1} < \tau_i < a_k \\ a_k - a_{k-1} & \tau_i \geq a_k \\ 0 & \text{جاهای دیگر} \end{cases}$$

لگاریتم تابع درستنمایی به صورت

$$\begin{aligned} L(\theta) = \sum_i \sum_k (\Phi_k + x_i'\beta) n_{ik} - \frac{1}{\gamma_0} \log(1 + \gamma B_{i0}) - \frac{1}{\gamma_k} \log(1 + \gamma_k B_{ik}) \\ + \log \sum_{\ell=0}^{n_{ik}} (e^{n_{ik}}) \frac{\Gamma(\frac{1}{\gamma_0} + \ell)}{\Gamma(\frac{1}{\gamma_0})} \frac{\Gamma(\frac{1}{\gamma_k} + n_{ik} - \ell)}{\Gamma(\frac{1}{\gamma_k})} \frac{1}{(\gamma_0 + B_{i0})^{\ell}} \frac{1}{(\frac{1}{\gamma_k} + B_{ik})^{n_{ik} - \ell}} \end{aligned}$$

نوشته می‌شود، که در آن $B_{ik} = e_{ik} e^{\Phi_k + x_i'\beta}$ و $B_{i0} = \sum_k e_{ik} e^{\Phi_k + x_i'\beta}$. برآورد پارامترها از حل معادلات درستنمایی با استفاده از روش‌های تکراری مانند نیوتن-رافسون بدست می‌آید (کلین، ۱۹۹۲).

۵ شبیه‌سازی

به منظور ارزیابی کارایی مدل شکنندگی جمعی تکه‌ای در برآورد پارامترها و مقایسه آنها با مدل شکنندگی با اثرات ثابت گاما یک مطالعه شبیه‌سازی اجرا گردید. هدف از این مطالعه ارزیابی عملکرد دو مدل بر اساس مقدار اریبی و مقایسه انحراف معیار تجربی با انحراف معیار مجانبی است. جهت تعیین اثر تعداد بازه‌ها روی عملکرد مدل پیشنهادی، داده‌ها با تعداد بازه‌های مختلف تولید و مدل به آنها برازش گردید. برای تولید داده‌ها نمونه‌هایی ۲۰۰ تایی که هر عضو دارای سه فرایند پواسن مستقل است در سه بازه زمانی مختلف در نظر گرفته شد. در هر بازه تعداد پیشامدهای هر عضو از یک فرایند پواسن با نرخ $(w_i + u_{ik})e^{\Phi_k + x_1\beta_1 + x_2\beta_2}$ استخراج گردید. متغیر کمکی x_1 از توزیع گسسته برنولی و متغیر کمکی x_2 نیز از توزیع پیوسته یکنواخت $U(0, 1)$ تولید گردیدند. پارامترهای رگرسیونی نیزبه صورت $\beta_1 = 2$ و $\beta_2 = -1$ در نظر گرفته شدند. متغیرهای شکنندگی u_{i1}, u_{i2}, w_i و u_{i3} نیز از توزیع های گاما با پارامتر مقیاس به ترتیب $\gamma_0 = 0/5, \gamma_1 = 0/8, \gamma_2 = 0/6, \gamma_3 = 0/4$ تولید شدند. لگاریتم شدت پایه (Φ_k) نیز به صورت بردار سه عضوی $(-1, 0/5, 1)$ در نظر گرفته شد. استخراج نمونه های ۲۰۰ تایی ۵۰۰ مرتبه تکرار گردید. فرایند تولید داده‌ها به روش فوق به ازای یک مجموعه دیگر از پارامترهای مقیاس γ_k $(k = 0, 1, 2, 3)$ نیز اجرا گردید.

تولید داده‌های فوق و برازش مدل توسط نرم‌افزار R صورت گرفته است. نتایج بدست آمده از برازش دو مدل شکنندگی جمعی تکه‌ای و مدل شکنندگی گاما به داده‌های شبیه‌سازی شده به روش فوق در جدول ۱ آورده شده است. مدل شکنندگی گاما که به داده‌ها برازش داده شده است یک مدل شکنندگی گاما با تابع خطر پایه متفاوت در بازه‌ها است. این مدل حالت ساده مدل جمعی تکه‌ای با اثر شکنندگی ثابت است و با مدل شکنندگی گامای معمولی که در آن تابع خطر پایه ثابت است کمی متفاوت است. در جدول ۱ خطای معیار تجربی[۱۲](ESE) و خطای معیار مجانبی[۱۳](ASE) پارامترها هستند. نتایج بدست آمده از برازش دو مدل نشان

[۱۲] Emperical Standard Deviation

[۱۳] Asymptotic Standard Deviation

می‌دهد که مقدار اریبی برآوردها در مدل شکنندگی تکه‌ای نسبت به مدل شکنندگی گاما کمتر است. بنابر این برازش یک مدل شکنندگی به داده‌هایی که از یک توزیع ناهمگن در زمان تولید شده‌اند برآوردهای اریب را نتیجه می‌دهد. همچنین مقدار انحراف تجربی برآوردهای دو مدل تفاوت چندانی با یکدیگر ندارد. در برخی از موارد انحراف معیار در مدل جمعی تکه‌ای کمتر باشد. برای تعیین اثر تعداد بازه‌های زمانی در کارایی مدل، به داده‌های تولید شده در بازه‌های زمانی ۵ و ۱۰ تایی نیز دو مدل برازش شدند. نتایج حاصل نشان می دهد که تعداد بازه‌های زمانی تاثیری در برآوردها نداشته است. با افزایش تعداد بازه‌های زمانی مقدار پارامترهای مشترک با مدل‌های کوچکتر تفاوتی را نشان نمی‌دهد. البته باید توجه داشت که در صورتی که میانگین تعداد پیشامدها برای هر فرد در هر بازه زمانی کم باشد برآوردها اریبی بیشتری خواهند داشت.

جدول ۱: برآوردهای دو مدل شکنندگی گاما و جمعی تکه‌ای در داده‌های سه‌فرایندی

پارامتر	مقدار حقیقی	شکنندگی جمعی تکه‌ای			شکنندگی گاما		
		اریبی	ESE	ASE	اریبی	ESE	ASE
$\beta_۱$	۲	۰/۰۲۹	۰/۱	۰/۰۹۵	۰/۱۵	۰/۰۹۶	۰/۰۹۸
$\beta_۲$	−۱	۰/۱۳	۰/۰۵	۰/۰۵۳	۰/۰۱۴	۰/۰۵۲	۰/۰۵۲
$\Phi_۱$	−۱	−۰/۱	۰/۱۱	۰/۱	۰/۰۵۹	۰/۰۱	۰/۰۱۲
$\Phi_۲$	۰/۵	۰/۷	۰/۰۸	۰/۰۸	۰/۰۶۳	۰/۰۵۶	۰/۰۵۵
$\Phi_۳$	۱	۰/۶	۰/۰۷	۰/۰۷۱	۰/۰۲۲	۰/۰۷	۰/۰۷۸
$\gamma_۰$	۰/۵	۰/۱۴	۰/۰۹۲	۰/۰۹۵	۰/۰۹۱	۰/۰۹	۰/۰۵۹
$\gamma_۱$	۰/۸	۰/۰۹	۰/۰۸۹	۰/۰۸۸	--	--	--
$\gamma_۲$	۰/۶	−/۱	۰/۰۱	۰/۰۱۱	--	--	--
$\gamma_۳$	۰/۴	۰/۷	۰/۰۲۱	۰/۰۲۲	--	--	--
$\beta_۱$	۲	۰/۰۱۸	۰/۱	۰/۰۹۶	۰/۲۵	۰/۰۹	۰/۰۹۸
$\beta_۲$	−۱	−۰/۰۱۲	۰/۰۸۵	۰/۰۸۵	۰/۰۱۴	۰/۰۵	۰/۰۵۲
$\Phi_۱$	−۱	−۰/۰۱۷	۰/۱۱	۰/۱۰۱	۰/۲۰۵	۰/۰۱	۰/۰۱۲
$\Phi_۲$	۰/۵	۰/۰۶۲	۰/۰۸	۰/۰۸	۰/۰۶۳	۰/۰۶۵	۰/۰۵۵
$\Phi_۳$	۱	−۰/۰۲	۰/۰۱۴	۰/۰۱۲	۰/۰۲۹	۰/۰۷	۰/۰۷۸
$\gamma_۰$	۰/۵	۰/۰۶۸	۰/۰۱۲	۰/۰۹۵	۰/۰۹۱	۰/۰۹	۰/۰۵۹
$\gamma_۱$	۰/۲	۰/۰۱	۰/۰۸۸	۰/۰۸۷	--	--	--
$\gamma_۲$	۰/۵	−۰/۰۶۸	۰/۰۱	۰/۰۱۱	--	--	--
$\gamma_۳$	۰/۸	۰/۰۱۳	۰/۰۲۱	۰/۰۲۲	--	--	--

بحث و نتیجه‌گیری

امروزه مدل‌های شکنندگی یکی از عمده‌ترین رویکردها برای تحلیل داده‌های بازگشتی به شمار می‌روند. با وجود این که در این مدل‌ها امکان تفسیر حاشیه‌ای

ضرایب وجود ندارد، اما به دلیل دارا بودن منطق شهودی ساده و در عین حال قوی برای در نظر گرفتن همبستگی بین پیشامدها کاربرد فراوانی پیدا کرده‌اند. در این مقاله یک مدل شکنندگی وابسته به زمان برای تحلیل داده‌های بازگشتی ارائه گردید. بر خلاف مدل‌های شکنندگی معمول که در آنها اثر تصادفی در طول زمان ثابت فرض می‌شود، اثر تصادفی در مدل شکنندگی جمعی تکه‌ای دارای یک فرایند گاما است. متغیر شکنندگی در طول بازه‌های زمانی ثابت ولی در انتهای هر بازه دارای نمو است. وابستگی متغیر شکنندگی به زمان، انعطاف مدل جهت در نظر گرفتن تغییر در ویژگیهای فردی یا محیطی که در طول زمان رخ می‌دهد را فراهم می‌کند. برخی از تغییرات فردی درونی و برخی ناشی از تغییر در عوامل بیرونی هستند. در برخی از مطالعات متغیر شکنندگی وابسته به متغیرهایی که خود به زمان وابسته هستند در نظر گرفته شده است (ابولیبده و همکاران، ۱۹۹۰، فونگ و همکاران، ۲۰۰۱). با وجود تغییر در ویژگیها در طول زمان، انتظار نمی‌رود این تغییرات در فاصله‌های زمانی خیلی کوتاه انجام شود. به عبارت دیگر اثر تصادفی به صورت لحظه‌ای تغییر نمی‌کند و متغیر شکنندگی تابعی از t به صورت $w(t)$ نیست. ولی منطقی به نظر می‌رسد که تابعی از بازه‌های زمانی I_k باشد. نتایج شبیه‌سازی در این مقاله نشان داد که تعداد بازه‌ها تاثیری در دقت و صحت برآوردهای مدل ندارد. با این وجود در کاربردهای عملی معمولا تعداد ۳ تا ۱۰ بازه برای مدل‌های تکه‌ای کافی به نظر می‌رسد. تعداد بیشتر بازه‌ها سبب افزایش پارامترها و پیچیده شدن مدل می‌گردد.

تقدیر و تشکر

نویسندگان از داوران محترم که پیشنهادات ارزنده ایشان موجب بهبود این مقاله گردید، کمال تشکر و سپاسگزاری را دارند.

مراجع

Aalen, O. O. (1998), *Hetrogenity in Survival Analysis*, Statistics in Medicine **7**, 1121-37.

Abu-libdeh, H., Turnbull B. W. and Clark L. (1990), *Analysis of Multivariate Recurrent Events in Longitudinal Studies: Application to Skin Cancer Prevention Trial*, Biometrics **46**, 1017-34.

Andersen, P. K. and Gill, R. D. (1982), *Cox's Regression Models for Counting Processes: a Large Sample Study*, Annals of Statistics, **10**, 1100.

Clayton, D. G. (1991), *A Monte Carlo Method for Bayesian Inference in Frailty Models.*, Biometrics **47**, 467-85.

Cook, R. and Lawless, J. F.(2002), *Analysis of Repeated Events*, Statistical Methods in Medical Research, **11**, 141-166.

Cox, D. R. (1983),*Some Remarks on Overdispersion*, Biometrika, **70**, 269-74.

Fong, D., Lam, K. F., Lawless, J. F. and Lee, Y. W. (2001), *Dynamic Random Effects for Times Between Repeated Events*, Lifetime Data Analysis, **7**, 345-362.

Hougaard, P. (2000), *Analysis of Multivariate Survival Data*, New York, Springer.

Hougaard, P. (1986), *Survival Models for Heterogenous Populations Derived from Stable Distributions*, Biometrika, **73**, 387-96.

Klein, J. P. (1992), *Semiparametric Estimation of Random Effects Using the Cox Model Based on the EM Algorithm*, Biometrics, **48**, 795-806.

Lawless, J. F., (2002), *Statistical Models and Methods for Lifetime Data*, New York, John Wiley & Sons.

Lawless, J. F. (1987), *Negative Binomial Regression Models*, The Canadian Journal of Statistics,**15(3)**, 209-26.

Manda, S. O. and Meyer, R. (2005), *Bayesian Inference for Recurrent Events Data Using Time-Dependent Frailty*, Statistics in Medicine, 24, 1263-74

Oakes, D., (1992), *Frailty Models for Multiple Eevent Times*, Survival Analysis: State of the Art, Netherland, Kluwer

Paik, M. C., Wei-Yann, T. and Ottman, R. (1994), *Multivariate Survival Analysis Using Piece Wise Gamma Frailty*, Biometrics **50**, 975-988.

Prentice, R.l., Williams, B. J. and Peterson, A. V. (1981), *On the Regression Analysis of Multivariate Failure Time Data*, Biometrika **68**, 373-79.

Sinha, D., Dey, D. K. (1997), *Semiparametric Bayesian Analysis of Survival Data*, Journal of the American Statistical Association, **92**, 1195-1212.

Therneau, T. M. and Grambsch, P. M. (2000), *Modeling Survival Data: Extending the Cox Model*, New York, Springer.

Wei, L. G., Lin, D. Y. and Weissfeld, L., (1989), *Regression Analysis of Multivariate Incomplete Failure Time Data by Modeling Marginal Distributions*, Journal of American Statistical Association, **84**, 1065-73.

Yau, K. K. and McGilchrist, C. A. (1998), *ML and REML Estimation in Survival Analysis with Time Dependent Correlated Frailty*, Statisitcs in Medicine, **17**, 1201-121.