

## مدل توأم داده‌های شمارشی طولی و زمان تا رخداد پیشامد صفر آماسیده با استفاده از مدل AFT: مطالعه موردی مجموعه داده‌های HIV/AIDS

مجتبی زینلی نجف‌آبادی، احسان بهرامی سامانی

گروه آمار، دانشکده علوم ریاضی، دانشگاه شهید بهشتی

تاریخ دریافت: ۱۳۹۸/۱۲/۱۳ تاریخ پذیرش و انتشار: ۱۳۹۹/۰۶/۳۰

**چکیده:** این مقاله با هدف مدل‌بندی توأم داده‌های شمارشی طولی تعداد سلول‌های CD4 و زمان تا رخداد مرگ در افراد مبتلا به HIV، مبتنی بر مدل زمان شکست شتابیده پایه ریزی شده است. برای مدل‌بندی پاسخ شمارشی طولی، از مدل رگرسیون آمیخته خطی تعمیم‌یافته تحت خانواده توزیع‌های سری توانی استفاده شده است و برای داده‌های زمان تا رخداد پیشامد، مدل پارامتری زمان شکست شتابیده تحت توزیع وایبل در نظر گرفته شده است. این دو پاسخ از طریق اثرات تصادفی همبسته با توزیع نرمال به یکدیگر مرتبط می‌شوند. فرض می‌شود پاسخ‌های طولی و زمان تا رخداد به شرط اثرات تصادفی و متغیرهای کمکی مستقل هستند. در نظر گرفتن تعداد صفر زیاد برای دو پاسخ و مکانیسم سانسور از راست، مدل توأمی را تولید کرده است که تاکنون توسط محققان دیگر مورد بررسی قرار نگرفته است. پارامترهای مدل توأم نیز با روش‌های مونت کارلوی زنجیر مارکوفی برآورد شدند.

**واژه‌های کلیدی:** مدل زمان شکست شتابیده، زمان تا رخداد پیشامد، سانسور از راست، خانواده توزیع‌های سری توانی، مدل رگرسیون آمیخته خطی تعمیم‌یافته.

## ۱ مقدمه

در پایان سال ۲۰۱۸، در جهان بیش از ۳۷/۸ میلیون نفر با ویروس اچ.آی.وی (HIV) زندگی می‌کردند و از ابتدای همه‌گیری این ویروس تاکنون تقریباً ۳۲ میلیون نفر در اثر ابتلا به آن جان باخته‌اند (سازمان بهداشت جهانی، ۲۰۱۸). اگرچه، همه‌گیری این ویروس در بین مناطق و کشورهای جهان، متفاوت است. با این حال منطقه صحرای آفریقا نسبت به سایر نقاط جهان، به شدت آسیب دیده باقی مانده است. در مطالعات بالینی و پزشکی، داده‌های طولی و زمان تا رخداد<sup>۱</sup> (بقا یا زمان شکست<sup>۲</sup>) نقش مهمی را ایفا می‌کنند و در عمل، عمدتاً با هم مشاهده می‌شوند (هندرسون و همکاران، ۲۰۰۰؛ گئو و کارلین، ۲۰۰۴؛ وو و همکاران، ۲۰۱۲). داده‌های طولی مانند تعداد سلول‌های CD4، از اندازه‌گیری‌های روی افراد مبتلا به ویروس HIV در یک سری از مقاطع زمانی، حاصل می‌شوند. ارتباط این داده‌ها با داده‌های زمان تا رخداد مرگ یک فرد مبتلا، مدت زمان وقوع این رویداد در هر فرد را طی دوره مطالعه معین، توصیف می‌کند. از نقطه نظر پزشکی، معمولاً زمان رخداد مرگ به تعداد سلول‌های CD4، بستگی دارد و لذا تحلیل توأم داده‌های همبسته زمان تا رخداد و شمارشی طولی، مثر ثمر است (سیاتیس و داویدین، ۲۰۰۴).

روش‌های زیادی برای مدل‌بندی حاشیه‌ای این داده‌ها از قبیل مدل‌های آمیخته خطی یا مدل‌های آمیخته خطی تعمیم‌یافته برای پاسخ طولی و مدل‌های پارامتری وایبل یا نیمه پارامتری خطرات متناسب کاکس<sup>۳</sup> (کاکس، ۱۹۷۲) برای پاسخ زمان تا رخداد پیشامد وجود دارد. مدل‌بندی حاشیه‌ای پاسخ‌ها، علی‌رغم سادگی در محاسبات، به دلیل اینکه نمی‌توانند وابستگی بین پاسخ‌ها را در نظر بگیرند، ممکن است منجر به نتایج و برآوردهای اریب و ناکارآمد شوند. به‌ویژه این تحلیل‌ها زمانی‌که پاسخ طولی با سلامتی افراد در ارتباط است، نامناسب هستند (گئو و کارلین، ۲۰۰۴). در این حالت، مدل‌سازی همزمان<sup>۴</sup> ضمن در نظر گرفتن وابستگی بین پاسخ‌ها، می‌تواند نتایج را بهتر و استنباط‌ها در رابطه با میزان تاثیر متغیرهای کمکی بر فرآیندهای طولی و بقا را معتبرتر سازد (هتفیلد و همکاران، ۲۰۱۲). گاهی اوقات در تحلیل داده‌های بقا و استفاده از مدل کاکس، فرض خطرات متناسب، برقرار نیست که در این اوقات عموماً، مدل‌های پارامتری زمان شکست شتابیده<sup>۵</sup> (وی، ۱۹۹۲)، به عنوان جایگزینی مناسب مورد استفاده قرار می‌گیرند (کلاین و موسبرگر، ۲۰۰۳؛ سنگ و همکاران، ۲۰۰۵). در مقابل مدل کاکس، مدل‌های زمان

<sup>۱</sup>Time to event<sup>۲</sup>Survival or Failure time<sup>۳</sup>Cox proportional hazards<sup>۴</sup>Simultaneous<sup>۵</sup>Accelerated Failure Time (AFT)

شکست شتابیده می‌توانند بهترین مشخص‌سازی و تفسیر عوامل ریسک (خطر) روی زمان شکست را داشته باشند. مدل‌های پارامتری زمان شکست شتابیده مثل وایبل و لگ نرمال توسط محققان زیادی مورد مطالعه قرار گرفته‌اند. یکی از راه‌های انتخاب این توزیع‌ها، ممکن است بر اساس تابع (نرخ) خطرشان باشد. بسیاری از محققان بر این باورند که داده‌های بقا در ارتباط با افراد آلوده به ویروس HIV، تحت برخی شرایط، ممکن است در زمان شروع بیماری، خطر افزایشی و در ادامه خطر کاهش‌ی داشته باشند (لون و همکاران، ۲۰۰۵). از طرفی نرخ خطر توزیع وایبل، بسته به مقدار پارامتر شکل، یکنوا است (کاهش‌ی است اگر مقدار پارامتر شکل کمتر از ۱ و افزایش‌ی است اگر این مقدار بیشتر از ۱، باشد). به این خاطر در این مقاله از توزیع وایبل برای مدل‌بندی پاسخ زمان تا رخداد پیشامد استفاده شد.

فرض کنید  $T_i$  برای  $i = 1, \dots, n$ ، متغیر زمان تا رخداد پیشامد و  $x_i = (1, x_{i1}, \dots, x_{ip})'$  برداری  $(p+1)$ -بعدی از متغیرهای کمکی باشد. در مدل زمان شکست شتابیده فرض می‌شود که رابطه خطی بین لگاریتم  $T_i$  و  $x_i$  به صورت  $\log T_i = x_i' \beta + \sigma \varepsilon_i$  وجود دارد، که در آن  $\beta = (\beta_0, \dots, \beta_p)'$  بردار پارامترهای رگرسیونی نامعلوم،  $\sigma$  ( $\sigma > 0$ ) پارامتر مقیاس نامعلوم و  $\varepsilon_i$  جملات خطای مستقل و بیان‌گر انحراف مقادیر  $\log T_i$  از بخش خطی مدل است.  $\varepsilon_i$  نمونه‌ای تصادفی از متغیر  $\varepsilon$  با تابع چگالی معلوم  $f_\varepsilon(\cdot, d)$ ، تابع بقای  $1 - F_\varepsilon(\cdot, d) = S_\varepsilon(\cdot, d)$  و تابع خطر  $h_\varepsilon = f_\varepsilon / S_\varepsilon$  است (گنجعلی و باغفلکی، ۱۳۹۶). عموماً توزیع  $\varepsilon_i$ ، چنان اختیار می‌گردد تا  $T_i$  از توزیع وایبل، لگ نرمال یا لگ لوژیستیک پیروی کند. اگر توزیع خطا، نرمال در نظر گرفته شود آنگاه  $T$  توزیع لگ نرمال و اگر توزیع خطا، لوژیستیک باشد آنگاه  $T$  توزیع لگ لوژیستیک خواهد داشت. توزیع وایبل هم تابعی خطی از توزیع مقدار حدی (مینیمم) استاندارد است (کلاین و موسبرگر، ۲۰۰۳). در این مقاله سعی بر آن است که با کمک مدل‌سازی توأم دو پاسخ شمارشی طولی و زمان تا رخداد پیشامد، که هر دو در نقطه صفر آماسیده شده‌اند، پارامترها به‌طور همزمان برآورد و تحلیل گردند. پاسخ‌ها از طریق اثرهای تصادفی همبسته به یکدیگر مرتبط می‌شوند.

در بخش ۲ چارچوب نظری مدل‌بندی شامل مدل رگرسیون آمیخته خطی تعمیم‌یافته برای پاسخ شمارشی طولی صفرآماسیده و مدل زمان شکست شتابیده برای پاسخ زمان تا رخداد پیشامد آماسیده در صفر، به همراه مدل توأم پیشنهادی ارائه شده است. در بخش ۳ به منظور نشان دادن کارایی مدل توأم پیشنهادی، آزمایش شبیه‌سازی با پارامترها و حجم نمونه مختلف اجرا شده است. در بخش ۴ داده‌های HIV/AIDS توصیف و مدل توأم به این داده‌ها برازش داده شده است. علی‌رغم پیچیدگی محاسبات مدل توأم، پارامترها برآورد و خطاهای استاندارد محاسبه شده‌اند. در نهایت بحث و نتیجه‌گیری ارائه شده است.

## ۲ چارچوب نظری مدل‌بندی توأم

یکی از رویکردهای مدل‌سازی همزمان، مدل‌سازی توأم است که به علت استفاده از تابع درستنمایی توأم، نه تنها منجر به افزایش دقت برآوردها می‌شود بلکه از محبوبیت زیادی نیز برخوردار است (سویتینگ و تامپسون، ۲۰۱۱). مدل‌سازی توأم پاسخ‌های طولی و زمان تا رخداد پیشامد به‌طور گسترده مورد توجه محققان قرار گرفته است. مرور خوبی از آن را می‌توان در سیاتیس و داویدین (۲۰۰۴) دنبال کرد. الگوی مدل‌سازی توأم شامل رویکردهای مختلفی از جمله الگوی آمیخته<sup>۱</sup> یا مدل‌های انتخابی است. این مدل‌ها را می‌توان با افزودن اثرهای تصادفی مشترک، توسعه داد که به آنها الگوی آمیخته تصادفی یا انتخاب تصادفی نیز می‌گویند و به‌طور کلی در کلاس مدل‌های اثرات تصادفی یا پارامتر مشترک<sup>۲</sup> فردنگر (مختص آزمودنی)<sup>۳</sup> قرار دارند (سوزا، ۲۰۱۱). در این مقاله برای مدل‌سازی دو مجموعه داده، از مدل اثرات تصادفی با فرض توزیع نرمال، استفاده شده است. این داده‌ها به شرط اثرات تصادفی و متغیرهای کمکی، مستقل هستند (لی و همکاران، ۲۰۱۷). در ادامه پاسخ شمارشی طولی، تحت مدل رگرسیون آمیخته خطی تعمیم‌یافته و توابع ربط لگاریتم و لوچیت و پاسخ زمان تا رخداد پیشامد تحت یک مدل زمان شکست شتابیده که از توزیع وایبل پیروی می‌کند مدل‌بندی خواهند شد. همچنین ویژگی آماسیدگی در صفر برای هر دو پاسخ تحت مکانیسم سانسور از راست، در نظر گرفته می‌شود. مسئله آماسیدگی در داده‌ها، زمانی رخ می‌دهد که متغیر پاسخ، در یک نقطه یا چند نقطه از تکیه‌گاهش، فراوانی بیش از اندازه‌ای داشته باشد که معروف‌ترین آن، آماسیدگی در نقطه صفر است. ویژگی آماسیدگی ممکن است در هر نقطه غیر صفر یا حتی به‌طور همزمان، در چندین نقطه مختلف از تکیه‌گاه نیز رخ دهد (بهرامی سامانی و خجسته‌بخت، ۱۳۹۹). همچنین، تابع درستنمایی توأم مربوط به دو پاسخ ارائه می‌شود و برآورد پارامترها نیز با روش‌های شبیه‌سازی مونت کارلوی زنجیر مارکوفی (MCMC)، مورد بحث قرار خواهند گرفت.

نمونه‌ای شامل  $n$  فرد را طی بازه مطالعه  $(0, T]$ ، در نظر بگیرید. فرد  $i$ ام مجموعه‌ای از داده‌های طولی  $y_i(t)$ ، را از پاسخ شمارشی در زمان‌های پیگیری  $t_{ij}, j = 1, \dots, n_i$  تولید می‌کند که  $n_i$  تعداد پیگیری برای فرد  $i$ ام ( $i = 1, \dots, n$ ) است. در حقیقت داده‌های طولی برای فرد  $i$ ام عبارت از  $y_i(t) = \{y_i(t_{ij}), j = 1, \dots, n_i\}$  است که در زمان‌های  $t_{ij}$  ثبت شده‌اند. برای فرآیند بقا نیز فرض کنید  $T_i = \min(T_i^*, C_i)$ ، زمان رخداد مشاهده شده برای فرد  $i$ ام است که در آن  $T_i^*$  زمان رخداد

<sup>1</sup>Pattern-Mixture

<sup>2</sup>Shared Parameter

<sup>3</sup>Subject-Specific

(واقعی) مرگ و  $C_i$  زمان سانسور است و فرض می‌شود که مستقل از  $T_i^*$  باشد (سانسور بی‌اطلاع<sup>۱</sup>). همچنین متغیر نشانگر وضعیت سانسور  $\delta_i = I(T_i^* \leq C_i)$  با مقادیر ۱ (در حالت مشاهده کامل) و ۰ (در حالت سانسور از راست) تعریف می‌گردد. فرض کنید ماتریس متغیرهای کمکی فرآیندهای طولی و بقا نیز به ترتیب با  $X$  و  $X(t)$  برای بخش‌های ثابت و  $W$  و  $W(t)$  برای بخش‌های تصادفی مشخص شوند. برخی از این متغیرهای کمکی مستقل از زمان هستند و ماتریس‌های بخش تصادفی، زیرماتریس‌هایی از ماتریس‌های بخش ثابت در نظر گرفته می‌شوند.

## ۲.۱ مدل رگرسیون آمیخته خطی تعمیم‌یافته برای پاسخ شمارشی طولی صفر آماسیده

مدل‌های خطی تعمیم‌یافته در مواقعی که پاسخ مورد نظر توزیع نرمال ندارد مورد استفاده قرار می‌گیرند. فرض کنید پاسخ شمارشی طولی، دارای توزیع سری‌های توانی<sup>۲</sup> با تعداد صفر زیاد و پارامتر  $\lambda_i(t)$  است که با نماد (ZIPSD) نشان داده می‌شود. در این صورت تابع جرم احتمال  $Y_i(t)$  به صورت

$$P(Y_i(t) = y_i(t); \lambda_i(t)) = \begin{cases} \pi_{1i}(t) + (1 - \pi_{1i}(t)) \frac{a_0}{b(\lambda_i(t))} & y_i(t) = 0 \\ (1 - \pi_{1i}(t)) \frac{a_{y_i(t)}(\lambda_i(t))^{y_i(t)}}{b(\lambda_i(t))} & y_i(t) > 0 \end{cases}$$

است، که در آن  $0 \leq \pi_{1i}(t) \leq 1$  و  $b(\lambda) = \sum_{y=0}^{\infty} a_y \lambda^y$ ،  $\lambda > 0$ ،  $a_y > 0$ ، پارامتر صفرآماسیده است که احتمال آمیختگی بین توزیع سری توانی و توزیع تباهیده در صفر را نشان می‌دهد (بهرامی و همکاران، ۲۰۱۲). این پارامتر به همراه  $\lambda_i(t)$ ، از طریق توابع ربط لجیت و لگاریتمی به متغیرهای کمکی مرتبط هستند. یعنی

$$\begin{aligned} \text{logit}(\pi_{1i}(t)) &= X'_{1i}(t)\beta_1 + W'_{1i}(t)\eta_{1i} \\ \log(\lambda_i(t)) &= X'_{2i}(t)\beta_2 + W'_{2i}(t)\eta_{2i} \end{aligned}$$

که در آن  $X_{\ell i}(t)$  و  $\beta_{\ell} = (\beta_{\ell 0}, \dots, \beta_{\ell p_1})'$  به ازای  $\ell = 1, 2$  به ترتیب بردارهای  $p_1$ -بعدی متغیرهای کمکی زمان-وابسته<sup>۳</sup> و ضرائب رگرسیونی بخش ثابت مدل طولی هستند. بردارهای  $X_{\ell i}(t)$  ممکن است همپوشانی داشته باشند و هر دو دارای عضو اول ۱ (برای عرض از مبدأ) هستند. همچنین

<sup>1</sup>Non-Informative

<sup>2</sup>Power Series Distribution (PSD)

<sup>3</sup>Time-Dependence (Time-Varying)

۱۵۲ ..... مدل توأم داده‌های شمارشی طولی و زمان

$\eta_{\ell i} = (\eta_{\ell i 1}, \dots, \eta_{\ell i q_1})'$ ، اثرات تصادفی متناظر با بردارهای  $q_1$ -بعدی متغیرهای کمکی  $W_{\ell i}(t)$  هستند ( $q_1 < p_1$ ). بردارهای اثرات تصادفی  $\eta_1$  و  $\eta_2$  مستقل و  $\eta_{\ell} \sim MVN_{q_1}(\circ, \Sigma_{\eta_{\ell}})$  که در آن  $\Sigma_{\eta_{\ell}}$  ماتریس کواریانس  $\eta_{\ell}$  است. اگر قرار داده شود  $\beta = (\beta_1', \beta_2')$  و  $\eta_i = (\eta_{1i}', \eta_{2i}')'$  آنگاه

$$Y_i(t) | \eta_i; \beta \sim ZIPSD(\lambda_i(t), \pi_{1i}(t)). \quad i = 1, \dots, n \quad (۱)$$

همچنین، تابع درستنمایی شرطی داده‌های مشاهده شده،  $y_i(t)$ ، برای فرد  $i$ ام،  $Ly$ ، به صورت

$$Ly(y_i(t) | \eta_i) = P_{LC}(Y_i(t) = y_i(t) | \eta_i; \beta)$$

است، که در آن  $P_{LC}(\cdot | \cdot)$  تابع جرم احتمال مربوط به بخش شمارشی طولی است. فرض می‌شود که به شرط  $\eta_i$ ، اندازه‌های طولی از یکدیگر مستقل هستند. بنابراین

$$P_{LC}(Y_i(t) = y_i(t) | \eta_i; \beta) = \prod_j P(Y_i(t_{ij}) = y_i(t_{ij}) | \eta_i; \beta)$$

که در آن  $P(Y_i(t_{ij}) = y_i(t_{ij}) | \eta_i; \beta)$ ، عبارت از تابع جرم احتمال پاسخ شمارشی است که از توزیع ZIPS، پیروی کرده و برابر است با

$$\begin{aligned} & (P(Y_i(t_{ij}) = \circ | \eta_i; \beta))^{I(Y_i(t_{ij}) = \circ)} (P(Y_i(t_{ij}) = y_i(t_{ij}) | \eta_i; \beta))^{I(Y_i(t_{ij}) > \circ)} \\ &= \left\{ \pi_{1i}(t_{ij}) + (1 - \pi_{1i}(t_{ij})) \frac{a_{\circ}}{b(\lambda_i(t_{ij}))} \right\}^{I(Y_i(t_{ij}) = \circ)} \\ &\times \left\{ (1 - \pi_{1i}(t_{ij})) \frac{a_{y_i(t_{ij})}(\lambda_i(t_{ij}))^{y_i(t_{ij})}}{b(\lambda_i(t_{ij}))} \right\}^{I(Y_i(t_{ij}) > \circ)} \end{aligned}$$

## ۲.۲ مدل زمان شکست شتابیده برای پاسخ زمان تا رخداد پیشامد صفر آماسیده

برای فرآیند بقا یک مدل رگرسیون زمان شکست شتابیده با اثرات آمیخته به صورت

$$\log(T_i^* | \vartheta_{1i}; \gamma_1, \sigma) = X_{1i}' \gamma_1 + W_{1i}' \vartheta_{1i} + \sigma \varepsilon_i$$

را در نظر بگیرید، که در آن  $X_{\tau i}$  و  $\gamma_1 = (\gamma_0, \dots, \gamma_{p_1})'$  به ترتیب بردارهای  $p_1$ -بعدی از متغیرهای کمکی مستقل از زمان و ضرائب رگرسیونی متناظر آنها مربوط به بخش ثابت مدل بقا هستند.  $X_{\tau i}$  دارای عضو اول ۱ است و  $\vartheta_{1i} = (\vartheta_{1i1}, \dots, \vartheta_{1iq_1})'$  اثرات تصادفی متناظر با بردار  $q_1$ -بعدی متغیرهای کمکی  $W_{\tau i}$  است ( $q_1 < p_1$ ). به علاوه، فرض کنید توزیع شرطی  $T_i^*$  به صورت

$$F(t_i|\vartheta_{1i}, \vartheta_{\tau i}; \gamma_1, \gamma_{\tau}, \alpha) = \pi_{\tau i} + (1 - \pi_{\tau i})F_{T^* > \cdot}(t_i|\vartheta_{1i}; \gamma_1, \alpha)$$

است، که در آن  $F_{T^* > \cdot}(t|\vartheta_1; \gamma_1, \alpha) = P(T^* \leq t | T^* > \cdot, \vartheta_1; \gamma_1, \alpha)$  یک توزیع شرطی پیوسته برای بخش غیر صفر  $T_i^*$  و پارامتر  $\vartheta_{\tau}; \gamma_{\tau}$  است،  $\pi_{\tau} = P(T^* = \cdot | \vartheta_{\tau}; \gamma_{\tau})$  احتمال شرطی بخش صفر است (بریکرز و گراولز، ۲۰۱۶). توزیع شرطی  $F_{T^* > \cdot}$ ، از طریق مدل زمان شکست شتابیده با توزیع وایبل تحت رگرسیون آمیخته خطی تعمیم یافته مدل بندی می شود. تابع چگالی شرطی و تابع بقا به صورت

$$f_{T^* > \cdot}(t|\vartheta_1; \gamma_1, \alpha) = \lambda^* \alpha t^{\alpha-1} \exp(-\lambda^* t^{\alpha})$$

$$S_{T^* > \cdot}(t|\vartheta_1; \gamma_1, \alpha) = e^{-\lambda^* t^{\alpha}}$$

خواهند بود، که در آن  $\log(\lambda^*) = -\alpha\{X'_{\tau}\gamma_1 + W'_{\tau}\vartheta_1\}$  و  $\alpha = \sigma^{-1}$ . برای مدل بندی  $\pi_{\tau}$ ، مدل رگرسیونی لوژیستیک  $\text{logit}(\pi_{\tau}) = X'_{\tau}\gamma_{\tau} + W'_{\tau}\vartheta_{\tau}$ ، در نظر گرفته می شود. بردارهای اثرات تصادفی  $\vartheta_1$  و  $\vartheta_{\tau}$  مستقل و  $\vartheta_l \sim MVN_{q_l}(\cdot, \Sigma_{\vartheta_l})$  که در آن  $\Sigma_{\vartheta_l}$  ماتریس کواریانس  $\vartheta_l$  است و اگر قرار داده شود  $\gamma = (\gamma'_1, \gamma'_{\tau})'$  و  $\vartheta_i = (\vartheta'_{1i}, \vartheta'_{\tau i})'$  داریم

$$T_i|\vartheta_i; \gamma, \alpha \sim ZIWeiAFT(\alpha, \lambda_i^*, \pi_{\tau i}) \quad (2)$$

همچنین، تابع درستنمایی شرطی  $t_i$  و  $\delta_i$ ، مربوط به  $i$ امین فرد،  $L_s$ ، در حالت سانسور از راست به صورت

$$L_s(t_i, \delta_i|\vartheta_i) = [f(t_i|\vartheta_i; \gamma, \alpha)]^{\delta_i} [S(t_i|\vartheta_i; \gamma, \alpha)]^{1-\delta_i}$$

$$= [(1-\pi_{\tau i})f_{T^* > \cdot}(t_i|\vartheta_i; \gamma, \alpha)]^{\delta_i} [(1-\pi_{\tau i})(1-F_{T^* > \cdot}(t_i|\vartheta_i; \gamma, \alpha))]^{1-\delta_i}$$

$$= (1-\pi_{\tau i}) [f_{T^* > \cdot}(t_i|\vartheta_i; \gamma, \alpha)]^{\delta_i} [S(t_i|\vartheta_i; \gamma, \alpha)]^{1-\delta_i}$$

خواهد بود، زیرا

$$\frac{dF(t_i|\boldsymbol{\vartheta}_i; \boldsymbol{\gamma}, \alpha)}{dt_i} = (1 - \pi_{\tau i}) \frac{dF_{T^* > \circ}(t_i|\boldsymbol{\vartheta}_{\backslash i}; \boldsymbol{\gamma}_{\backslash}, \alpha)}{dt_i}$$

و همچنین  $f_{T^* > \circ}(t|\cdot) = d/dt F_{T^* > \circ}(t|\cdot)$  (کلاین و موسبرگر، ۲۰۰۳).

## ۲.۳ مدل توأم

مدل رگرسیونی آمیخته خطی تعمیم‌یافته برای فرآیند شمارشی طولی آماسیده در صفر (۱) و مدل زمان شکست شتابیده برای فرآیند بقای صفر آماسیده (۲)، از طریق اثرات تصادفی همبسته  $(\boldsymbol{\eta}, \boldsymbol{\vartheta})'$ ، که به ازای هر فرد مستقل هستند، به یکدیگر مرتبط می‌شوند (هندرسون و همکاران، ۲۰۰۰؛ سنگ و همکاران، ۲۰۰۵). در اینجا، فرض می‌شود  $(\boldsymbol{\eta}, \boldsymbol{\vartheta})' \sim MVN(\circ, D_{\boldsymbol{\eta}, \boldsymbol{\vartheta}})$ ، که در آن  $D_{\boldsymbol{\eta}, \boldsymbol{\vartheta}}$ ، ماتریس کواریانس اثرات تصادفی دو مولفه طولی و بقا است و به صورت  $D_{\boldsymbol{\eta}, \boldsymbol{\vartheta}} = \begin{bmatrix} \Sigma_{\boldsymbol{\eta}} & \rho \Sigma_{\boldsymbol{\eta}} \Sigma_{\boldsymbol{\vartheta}} \\ \rho \Sigma_{\boldsymbol{\eta}} \Sigma_{\boldsymbol{\vartheta}} & \Sigma_{\boldsymbol{\vartheta}} \end{bmatrix}$  تعریف می‌شود که  $\Sigma_{\boldsymbol{\eta}}$ ،  $\Sigma_{\boldsymbol{\vartheta}}$  و  $\rho$  به ترتیب ماتریس‌های کواریانس اثرات تصادفی مولفه‌های طولی، بقا و همبستگی بین دو مولفه (زیرمدل) هستند. اگر  $\rho = \circ$ ، آنگاه مدل توأم به مدل‌های مستقل تبدیل می‌شوند. به این ترتیب، مدل توأم به صورت است. در اینجا داده‌های مشاهده شده  $\{(\mathbf{y}_i(t), T_i = t_i, \delta_i)\}_{i=1}^n$  و پارامترهای مدل توأم

$$\mathbf{Y}_i(t) | \boldsymbol{\eta}_i; \boldsymbol{\beta} \sim ZIPSD(\lambda_i(t), \pi_{\backslash i}(t)) \quad T_i | \boldsymbol{\vartheta}_i; \boldsymbol{\gamma}, \alpha \sim ZIW e i A F T(\alpha, \lambda_i^*, \pi_{\tau i})$$

$$\log(\lambda_i(t)) = \mathbf{X}'_{\tau i}(t) \boldsymbol{\beta}_{\tau} + \mathbf{W}'_{\tau i}(t) \boldsymbol{\eta}_{\tau i} \quad \log(\lambda_i^*) = -\alpha \{ \mathbf{X}'_{\tau i} \boldsymbol{\gamma}_{\backslash} + \mathbf{W}'_{\tau i} \boldsymbol{\vartheta}_{\backslash i} \}$$

$$\text{logit}(\pi_{\backslash i}(t)) = \mathbf{X}'_{\backslash i}(t) \boldsymbol{\beta}_{\backslash} + \mathbf{W}'_{\backslash i}(t) \boldsymbol{\eta}_{\backslash i} \quad \text{logit}(\pi_{\tau i}) = \mathbf{X}'_{\tau i} \boldsymbol{\gamma}_{\tau} + \mathbf{W}'_{\tau i} \boldsymbol{\vartheta}_{\tau i}$$

$\boldsymbol{\Psi} = (\boldsymbol{\beta}, \boldsymbol{\gamma}, D_{\boldsymbol{\eta}, \boldsymbol{\vartheta}}, \sigma)$  هستند. در این صورت، تابع درستنمایی مدل توأم به صورت

$$L(\boldsymbol{\Psi}; \mathbf{y}(t), t, \delta) = \prod_{i=1}^n \int L_y(\mathbf{y}_i(t) | \boldsymbol{\eta}_i) L_s(t_i, \delta_i | \boldsymbol{\vartheta}_i) f(\boldsymbol{\eta}_i, \boldsymbol{\vartheta}_i) d\boldsymbol{\eta}_i d\boldsymbol{\vartheta}_i$$

خواهد بود، که در آن  $f(\cdot, \cdot)$  تابع چگالی احتمال بردار اثرات تصادفی  $(\boldsymbol{\eta}_i, \boldsymbol{\vartheta}_i)$  است.

## ۲.۴ شناساپذیری پارامترهای مدل

شناساپذیری مدل‌های آماری، یکی از مهم‌ترین ویژگی‌های پارامترهای یک مدل آماری مناسب است. اگر پارامترهای یک مدل آماری شناسا نباشند، آنگاه برآوردهای ارباب با انحراف معیار بالا برای پارامترهای مدل موردنظر به دست می‌آید و این موضوع سبب می‌شود که نتایج و استنباط‌های اشتباه و غلطی در پژوهش موردنظر حاصل گردد. از سویی دیگر، با این موضوع زمانی بیشتر برخورد می‌کنیم که تعداد پارامترها در یک مدل آماری افزایش پیدا کند، به‌خصوص در تحلیل داده‌های آمیخته همبسته، با مدل‌هایی سرکار خواهیم داشت که تعداد پارامترهای آن از تحلیل جداگانه این داده‌ها بیشتر خواهد شد. بنابراین در این مدل‌ها، شناساپذیری پارامترهای مدل، نقش بسزایی را ایفا می‌کند. همچنین در مواقعی که پارامترهای مدل شناساپذیر نباشند رفتار مدل غیرمنطقی و برآوردها نامعقول است. در شبیه‌سازی اینگونه مدل‌ها، مواردی مانند عدم همگرایی الگوریتم‌ها، مشاهده مقادیر خطای خیلی کوچک یا خیلی بزرگ، برآوردهای خارج از دامنه فضای پارامتر، فواصل اطمینان با طول زیاد مشاهده می‌شود و بعضاً با تغییر نقطه اولیه، برآوردها به طرز چشمگیری تغییر می‌کنند. از آن جا که مدل ارائه شده در این مقاله مدل همزمانی از مدل‌های خطی تعمیم یافته است، لذا پارامترهای مدل شناساپذیر هستند (بهرامی سامانی و گنجعلی، ۱۳۹۸).

## ۳ مطالعه شبیه‌سازی

در این بخش برای نشان دادن عملکرد و کارایی مدل توأم پیشنهادی که در بخش قبل ارائه شده است، مطالعه شبیه‌سازی با حجم  $n = 50$ ،  $100$  و  $500$  با تعداد  $500$  تکرار، انجام می‌شود. ابتدا بردار تصادفی  $U$  به طول  $n$  از توزیع یکنواخت  $(0, 1)$ ، تولید و سپس برای تولید داده‌های شمارشی طولی صفرآماسیده،  $y_{ij}$ ،  $i = 1, \dots, n$ ،  $j = 1, 2$  به صورت

$$\begin{cases} y_{ij} = 0 & U_{ij} < \pi_{ij} = \frac{\exp(\beta_{10} + \beta_{1j}x_{1ij})}{1 + \exp(\beta_{10} + \beta_{1j}x_{1ij})} \\ y_{ij} \sim \text{Poisson}(\lambda_{ij}) & \text{جای دیگر} \end{cases}$$

عمل می‌شود. پارامتر توزیع پواسن از طریق تابع ربط  $\lambda_{ij} = \exp(\beta_{20} + \beta_{2j}X_{2ij} + \sigma_{b_i}b_i)$  به متغیرهای کمکی مرتبط می‌شود، که در آن  $b_i$ ، اثر تصادفی مشترک بین دو مولفه طولی و بقا است به‌طوری‌که  $(0, \sigma_{b_i}^2) \sim N(0, \sigma_{b_i}^2)$  و  $\sigma_{b_i} = 0.5$ . برای تولید داده‌های بقای صفرآماسیده،  $t_i$ ، از مدل زمان

شکست شتابیده،  $t_i^* = \exp(\gamma_{10} + \gamma_{11}X_{1i} + \sigma_{b_i}b_i + 0.5 \times \varepsilon_i)$ ، زمان‌های رخداد تولید و به صورت

$$\begin{cases} t_i = 0 & U_i < \pi_{1i} = \frac{\exp(\gamma_{20} + \gamma_{21}x_{1i})}{1 + \exp(\gamma_{20} + \gamma_{21}x_{1i})} \\ t_i = \min(t_i^*, C_i) & \text{جای دیگر} \end{cases}$$

عمل می‌شود، که در آن  $\varepsilon_i$  از توزیع مقدار حدی شبیه‌سازی می‌شود. در این فرآیند، مقادیر اولیه پارامترها عبارت‌اند از:  $\beta_1 = (\beta_{10}, \beta_{11}, \beta_{12})' = (1, 1, -1)'$ ،  $\beta_2 = (\beta_{20}, \beta_{21}, \beta_{22})' = (3, -1, 1)'$ ،  $\gamma_1 = (\gamma_{10}, \gamma_{11})' = (2, -1)'$  و  $\gamma_2 = (\gamma_{20}, \gamma_{21})' = (1, 1)'$ . کلیه ماتریس‌های طرح از توزیع برنولی با احتمال موفقیت ۰/۴، تولید می‌شوند. همچنین، زمان سانسور از توزیع یکنواخت  $(0, 5)$ ، تولید شده‌است که به‌طور متوسط منجر به ۲۷٪ سانسور در داده‌ها می‌گردد. برازش مدل توأم در نرم‌افزار “R” صورت گرفت و با استفاده از تابع “nlminb”، ماکسیمم تابع درستنمایی توأم و برآورد پارامترها محاسبه گردید. همچنین ماتریس اطلاع فیشر مربوط به برآورد پارامترهای مدل و سپس عکس آن محاسبه شده، درایه‌های روی قطر اصلی، واریانس‌های مربوط به برآورد پارامترهای مدل هستند. در نهایت خطای استاندارد برآورد پارامترهای مدل از جذر مثبت واریانس‌های موجود در عکس ماتریس اطلاع فیشر به دست آمد. برای به دست آوردن ماتریس اطلاع فیشر از دستور “fdHess” استفاده می‌شود. در ادامه و در جدول ۱، برآوردها، خطای استاندارد، اریبی و میانگین مربع خطا (MSE)، ارائه شده‌اند. همچنین، در شکل ۱، MSE برآوردها با اندازه نمونه‌های مختلف رسم شده‌اند. نتایج نشان می‌دهد که مقادیر برآورد به مقادیر واقعی نزدیک‌اند و با افزایش حجم نمونه، این فاصله کمتر می‌شود. همچنین مقادیر خطای برآورد نیز کوچک است. از سوی دیگر، میزان اریبی نزدیک به صفر است که با افزایش حجم نمونه، گرایش به صفر دارد. قابل ذکر است که MSE نیز مقادیر کوچکی اختیار کرده است و با افزایش حجم نمونه، این مقادیر نیز به صفر میل می‌کنند. لذا برآوردها سازگار هستند و مدل توانمندی بالایی در برآورد پارامترها دارد.

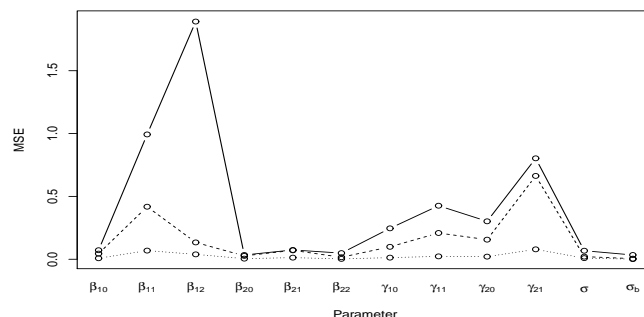
## ۴ تحلیل داده‌های HIV/AIDS

در این بخش مدل توأم ارائه شده را روی مجموعه داده‌هایی که توسط **گلدمن و همکاران (۱۹۹۶)** جمع‌آوری شده‌است، اعمال می‌شود. این داده‌ها یک مطالعه طولی بر روی ۴۶۷ فرد با سندروم نقص ایمنی اکتسابی (AIDS) است، که تاثیر درمان AZT<sup>۱</sup> برای آنها بررسی شده است. هدف اصلی تحلیل وابستگی و ارتباط

<sup>۱</sup>Zidovudine

جدول ۱. نتایج شبیه سازی مدل توأم به تکنیک حجم نمونه با در نظر گرفتن مکانیسم سانسور از راست.

MSE	n = 500			n = 100			n = 50			پارامتر			
	اریبی	خطای برآورد	برآورد	MSE	اریبی	خطای برآورد	برآورد	MSE	اریبی		خطای برآورد	برآورد	
0.008	-0.009	0.009	1.009	0.045	0.028	0.011	0.072	0.072	-0.025	0.067	1.025	1.0	$\beta_{10}$
0.069	0.155	0.212	0.825	0.418	0.266	0.448	0.524	0.493	0.587	0.805	0.413	1.0	$\beta_{11}$
0.039	0.095	0.172	-1.095	0.134	-0.070	0.36	-0.093	1.891	-1.179	0.708	0.179	-1.0	$\beta_{12}$
0.005	-0.056	0.039	2.056	0.027	-0.145	0.078	2.125	0.334	-0.155	0.908	2.155	2.0	$\beta_{20}$
0.013	0.077	0.085	-1.077	0.071	-0.176	0.199	-0.824	0.074	0.179	0.205	-1.179	-1.0	$\beta_{21}$
0.003	0.006	0.051	0.994	0.016	0.061	0.113	0.039	0.049	0.072	0.011	0.978	1.0	$\beta_{22}$
0.013	-0.065	0.092	2.065	0.099	-0.180	0.058	2.180	0.246	-0.351	0.206	2.051	2.0	$\gamma_{10}$
0.023	-0.069	0.125	-0.031	0.009	0.225	0.012	-1.225	0.426	0.583	0.293	-1.583	-1.0	$\gamma_{11}$
0.021	-0.059	0.122	1.059	0.156	-0.253	0.303	1.252	0.302	-0.299	0.36	1.299	1.0	$\gamma_{20}$
0.079	-0.009	0.066	1.000	0.663	0.61	0.528	0.39	0.803	0.653	0.773	0.547	1.0	$\gamma_{21}$
0.009	0.076	0.055	0.424	0.021	0.098	0.106	0.402	0.068	0.222	0.119	0.468	0.5	$\sigma$
0.001	-0.021	0.027	0.521	0.005	0.052	0.051	0.428	0.034	0.165	0.802	0.325	0.5	$\sigma_b$

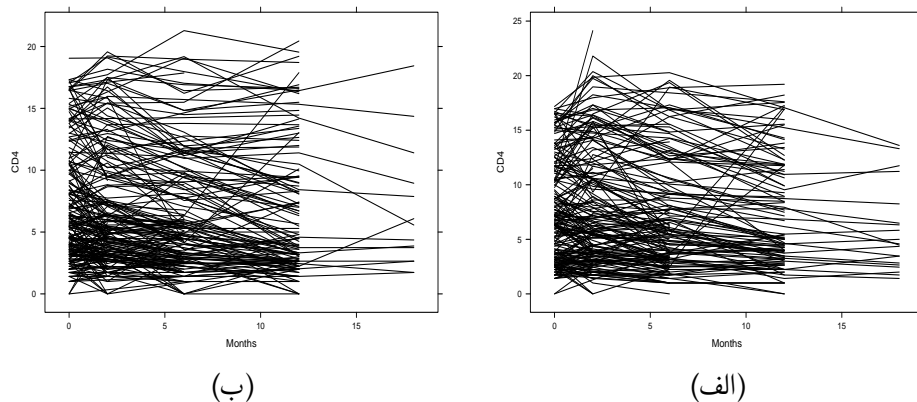


شکل ۱. مقادیر MSE پارامترها برای حجم نمونه ۵۰ (خط ممتد)، ۱۰۰ (خط چین) و ۵۰۰ (نقطه چین).

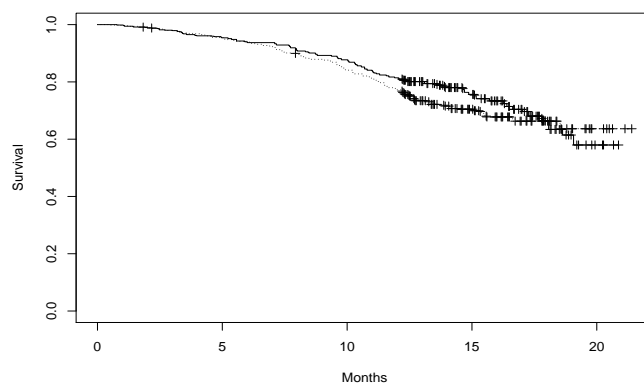
بین تعداد سلول‌های CD4 و زمان مرگ افراد و مقایسه اثربخشی و اطمینان به دو داروی ضدویروسی ddI<sup>۱</sup> و ddC<sup>۲</sup> است. افراد به طور تصادفی این دو دارو را دریافت می‌کنند و تعداد سلول‌های CD4 آنها در زمان ورود به دوره مطالعه، و همچنین ۲، ۶، ۱۲ و ۱۸ ماه پس از آن ثبت شده‌اند ( $n_i \leq 5$ ). بسیاری از افراد که در نقطه زمانی اول ( $t = 0$ )، پیشامد مرگ را تجربه کردند تعداد سلول‌های ایمنی بدنشان صفر گزارش شده است. یعنی بسیاری از افرادی که در مجموعه خطر قرار داشتند در زمان صفر، پیشامد مرگ را تجربه کردند. لذا فرآیند بقا در نقطه صفر آماسیده است. همچنین، تعداد سلول‌های CD4 این افراد نیز صفر اندازه‌گیری و گزارش شده است. به‌طور کلی تعداد سلول‌های CD4، میزان سلامت سیستم ایمنی بدن را نشان می‌دهد. تعداد طبیعی سلول‌های CD4، در یک فرد، بیش از ۵۰۰ سلول در هر میلی‌متر مکعب خون است. اگر تعداد CD4، کمتر از ۳۰۰ باشد، فرد مبتلا به AIDS تشخیص داده می‌شود. بنابراین، اندازه‌گیری و سنجش تعداد CD4، یک شاخص مهم است و می‌تواند به شناخت و آگاهی میزان پیشرفت ویروس HIV و نهایتاً ابتلا به ایدز، کمک کند. بنابراین، تعداد سلول‌های CD4، در طول زمان، به عنوان پاسخ طولی در نظر گرفته شد. در پایان دوره مطالعه ۱۸۸ فرد جان خود را از دست داده بودند که منجر به ۵۹/۷٪ سانسور در داده‌ها گردید. حجم نمونه در پنج نقطه زمانی (۲۲، ۱۰۲، ۱۵۳، ۱۸۲، ۲۳۱) برای گروه دارویی ddI و (۱۴، ۱۲۳، ۱۵۷، ۱۸۶، ۲۳۶) برای گروه دارویی ddC است. داده‌های مشاهده نشده با گذشت زمان به علت مرگ و میر، انصراف و ترک مطالعه و عدم مراجعه به درمانگاه، به شدت در حال افزایش بوده است. به عنوان یک تحلیل توصیفی، حالت‌های طولی تعداد سلول‌های CD4 افراد و برآورد کاپلان مایر منحنی بقا برای هر دو دارو، به ترتیب در شکل‌های ۲ و ۳، نمایش داده شده‌است. شکل ۲،

<sup>۱</sup>Didanosine

<sup>۲</sup>Zalcitabine



شکل ۲. تغییرات تعداد سلول‌های CD4 افراد، برای دو داروی الف: ddI و ب: ddC.



شکل ۳. برآوردهای کاپلان مایر احتمال بقا برای ddI (نقطه‌چین) و ddC (خط ممتد).

نرخ بالایی از گمشدگی را در طول زمان به تفکیک دو داروی ضد ویروسی نشان می‌دهد. همچنین باتوجه به شکل ۳، احتمال زنده ماندن افراد دریافت کننده تیمار ddC، اندکی بیشتر از احتمال زنده ماندن افراد دریافت کننده تیمار ddI، بعد از ماه ششم از دوره مطالعه است. فرآیند شمارشی طولی عبارت است از

$$y_{ij}|b_i; \beta \sim ZIPoisson(\lambda_{ij}, \pi_{ij})$$

و پارامترها با توابع ربط

$$\text{logit}(\pi_{ij}) = \beta_{10} + \beta_{11}t_{ij} + \beta_{12}t_{ij} Drug_i + \beta_{13} Gender_i + \beta_{14} PrevOI_i$$

۱۶۰ ..... مدل توأم داده‌های شمارشی طولی و زمان

$$\begin{aligned}
 & + \beta_{15} AZT_i \\
 \log(\lambda_{ij}) & = \beta_{20} + \beta_{21j}t_{ij} + \beta_{22j}t_{ij} Drug_i + \beta_{23} Gender_i + \beta_{24} PrevOI_i \\
 & + \beta_{25} AZT_i + \sigma_{b_i}b_i
 \end{aligned}$$

به متغیرهای کمکی و اثرات تصادفی مرتبط می‌شوند. در اینجا  $y_{ij}$ ، تعداد سلول‌های CD4 فرد  $i$  ام ( $i = 1, \dots, 467$ ) در زمان  $j$ ام ( $j = 1, 2$ ) و زمان رخداد مرگ نیز به عنوان پاسخ بقاء، در نظر گرفته شد. برای این فرآیند، مدل پارامتری زمان شکست شتابیده با توزیع وایبل با تعداد صفر زیاد، به صورت

$$\begin{aligned}
 T_i|b_i; \gamma, \alpha & \sim ZIWeiAFT(\alpha, \lambda_i^*, \pi_{2i}) \\
 \log(\lambda_i^*) & = -\alpha\{\gamma_{10} + \gamma_{11} Drug_i + \gamma_{12} Gender_i + \gamma_{13} PrevOI_i \\
 & + \gamma_{14}AZT_i + \sigma_{b_i}b_i\} \\
 \text{logit}(\pi_{2i}) & = \gamma_{20} + \gamma_{21} Drug_i + \gamma_{22} Gender_i + \gamma_{23} PrevOI_i + \gamma_{24}AZT_i
 \end{aligned}$$

است. در مجموعه داده‌ها  $Gender_i$  معرف جنسیت (زن=۰، مرد=۱)،  $Drug_i$  نوع دارو (ddC=۰، ddI=۱)،  $PrevOI_i$  معرف عفونت‌های فرصت‌طلب (موثر) قبلی (AIDS=۰، no AIDS=۱) و تاثیر درمان  $AZT_i$  (failure=۰، intolerance=۱) است. در جدول ۲، نتایج برازش مدل توأم روی داده‌های واقعی، شامل برآورد پارامترها (Est.)، خطای برآورد (S.E.) و فواصل اطمینان ۹۵٪، ارائه شده است. همانطور که ملاحظه می‌شود اختلاف متغیر کمکی  $PrevOI$ ، از مقدار صفر، در هر دو مولفه طولی و بقاء معنی‌دار شده است. مقدار برآورد شده ضریب رگرسیونی این متغیر در مولفه طولی، ۲/۹۹- و بازه اطمینان ۹۵٪ آن، (۰/۱۸-، ۵/۹۶۲-) است. از این‌رو، یک فرد که در ابتدای ورود به مطالعه مبتلا به ایدز تشخیص داده شده بود دارای تعداد سلول‌های CD4 به مراتب کمتر از یک فرد بدون تشخیص ایدز، بوده است. همچنین، اختلاف معنی‌داری بین دو نوع دارو (ddI و ddC) وجود ندارد.

## بحث و نتیجه‌گیری

برای تحلیل داده‌های حاصل از دو پاسخ همبسته طولی و زمان تا رخداد پیشامد، که هر دو در نقطه صفر آماسیده شده‌اند، از مدل‌سازی توأم استفاده شد. پارامترهای پاسخ طولی، از طریق مدل رگرسیونی آمیخته

جدول ۲. نتایج برازش مدل توأم برای تحلیل داده‌های AIDS تحت مکانیسم سانسور از راست.

پاسخ زمانی تا رخداد			پاسخ شارشی طولی		
۹۵٪	پایه	<i>S.E.</i>	<i>Est.</i>	۹۵٪	پایه
اطمینان	۹۵٪			اطمینان	
(۰/۴۱۵, ۰/۵۷۶)	۰/۴	۰/۴۹۶	$\sigma$	(-۱۲۸۲۷, -۰/۱۹۳)	۳/۸۳
(۲۰۰۱۵, ۴۰۱)	۰/۴۹	۳۰۱۲	$Constant(\gamma_{10})$	-۶۵۶	$Constant(\beta_{10})$
(-۰/۳۱۵, ۰/۰۷۷)	۰/۰۹۸	-۰/۱۱۹	$Drug(\gamma_{11})$	-۶۶۴	$Time(\beta_{11})$
(-۰/۰۶۱۵, ۰/۳۱۳)	۰/۲۳۲	-۰/۱۵۱	$Gender(\gamma_{12})$	-۰/۲۴	$Drug \times Time(\beta_{12})$
(۰/۳۶۴, ۰/۸۹۸)	۰/۱۳۴	۰/۶۳۱	$PrevOI(\gamma_{13})$	-۲/۸	$Gender(\beta_{13})$
(-۰/۰۶۳, ۰/۰۶۷)	۰/۰۳۳	۰/۰۰۲	$AZT(\gamma_{14})$	-۲/۸۹	$PrevOI(\beta_{14})$
(-۲۱۷۵۶, ۶۴۵)	۷/۵۱	-۷/۶۵۳	$Constant(\gamma_{20})$	-۰/۶۹	$AZT(\beta_{15})$
(-۲۵۱۰۷, ۱۶۸۱۴)	۱۰/۴۸	-۴۱۴۷	$Drug(\gamma_{21})$	-۰/۲۳	$Time(\beta_{112})$
(-۱۰/۳۳۸, ۵۵۳۴)	۴/۶۸	-۲۶۰۲	$Gender(\gamma_{22})$	-۲/۴۹	$Drug \times Time(\beta_{22})$
(-۵۱۴, -۰/۸۲۲)	۱/۰۵۴	-۳/۳۱	$PrevOI(\gamma_{23})$	۲/۶۱۸	$Constant(\beta_{20})$
(-۱۰/۳۷۵, ۹/۲۷۴)	۴/۹۱۲	-۰/۵۵۱	$AZT(\gamma_{24})$	-۱/۵۶۱	$Time(\beta_{21})$
				۰/۳۶	$Drug \times Time(\beta_{22})$
				۰/۱۱	$Gender(\beta_{23})$
				۰/۴۰۷۶	$PrevOI(\beta_{24})$
				۰/۴۰۷۰۶	$AZT(\beta_{25})$
				۰/۰۷۴	$Time(\beta_{212})$
				۲/۲۹۹	$Drug \times Time(\beta_{222})$
				۰/۲۲	
				۰/۲۶	$\sigma_b$
				۰/۵۳۱	اثر تصادفی

پارامتر  
آماسیگی ( $\pi_1$ )

پارامتر توزیع  
پواسن ( $\lambda$ )

خطی تعمیم یافته، به متغیرهای کمکی مرتبط شدند. چون اکثر مطالعات بقاء، بیانگر آن است که برآورد پلرامترها توسط مدل‌های پارامتری، کارایی بالاتری نسبت به مدل خطرات متناسب کاکس دارند، برای مدل‌بندی پاسخ زمان تا رخداد پیشامد، از مدل پارامتری زمان شکست شتابیده با توزیع وایبل و مکانیسم سانسور از راست استفاده شد. این دو پاسخ، از طریق اثرات تصادفی همبسته به یکدیگر مرتبط گردیدند. برای بررسی عملکرد مدل توأم پیشنهادی، آزمایش شبیه‌سازی صورت گرفت و از جمله نتایج شبیه‌سازی می‌توان به کاهش خطای استاندارد برآوردها، با افزایش اندازه نمونه، اشاره کرد. در بخش کاربرد، مجموعه داده‌های مربوط به افراد مبتلا به ویروس HIV، مورد تحلیل قرار گرفتند. از چهار متغیر کمکی موجود در مدل، متغیر کمکی *PrevOI* در سطح ۵٪ معنی‌دار شد. لذا افرادی که در شروع فرآیند مطالعه، مبتلا به بیماری ایدز، بوده‌اند نسبت به افرادی که بدون علائم بیماری تشخیص داده شده‌اند دارای تعداد سلول‌های CD4 کمتری هستند. همچنین نتایج دلالت بر وجود اختلاف معنی‌دار بین دو گروه دارویی ندارند. از سوی دیگر، برآورد انحراف استاندارد پارامتر پیوند دهنده دو پاسخ،  $\sigma_b$ ، معادل ۵۳۱٪ است که اختلاف معنی‌داری از صفر دارد. لذا به‌طور شهودی پیوند بین دو پاسخ تأیید می‌شود.

## تقدیر و تشکر

از داوران و ویراستار محترم مجله که با توصیه‌های ارزشمند خود سبب ارتقای این مقاله و ارائه بهتر آن شده‌اند کمال تشکر را داریم.

## مراجع

- بهرامی سامانی، ا. و خجسته‌بخت، ن. (۱۳۹۹)، تحلیل نیم‌پارامتری مدل‌های رگرسیونی برای پاسخ‌های سری توانی آماسیده صفر با متغیرهای تبیینی گمشده، مجله علوم آماری، ۱۴، ۹۵-۱۱۲.
- بهرامی سامانی، ا. و گنجعلی، م. (۱۳۹۸)، تحلیل داده‌های آمیخته گسسته-پیوسته، انتشارات دانشگاه شهید بهشتی تهران، تهران.
- گنجعلی، م. و باغفلکی، ت. (۱۳۹۶)، مبانی و مدل‌بندی بیزی داده‌ها: با استفاده از برنامه نویسی BUGS و نرم‌فزار R، انتشارات دانشگاه شهید بهشتی، تهران.
- Bahrami, S. E., Ganjali, M. and Amirian, Y. (2012), Likelihood Estimation

for Longitudinal Zero-Inflated Power Series Regression Models, *Journal of Applied Statistic*, **39**, 1965–1974.

Braekers, R. and Grouwels, Y., (2016), A Semi-Parametric Cox’s Regression Model for Zero Inflated Left-Censored Time To Event Data, *Communications in Statistics – Theory and Methods*, **45**, 1969-1988.

Cox, D. R. (1972), Regression Models and Life-Tables, *Journal of the Royal Statistical Society, Series B*, **34**: 187–220.

Goldman, A. I., Carlin, B. P., Crane, L. R., Launer, C., Korvick, J. A., Deyton, L. and Abrams, D. I. (1996), Response of CD4+ and Clinical Consequences to Treatment Using ddI or ddC in Patients with Advanced HIV Infection, *Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology*, **11**: 161–169.

Guo, X. and Carlin, B. P. (2004), Separate and Joint Modeling of Longitudinal and Event Time Data Using Standard Computer Packages, *Journal American Statistical Association*, **58**: 16-24.

Hatfield, A., Hodges, S., and Carlin, P. (2012), Combining Longitudinal and Survival Information in Bayesian Joint Models: When Are Treatment Estimates Improved? *Biostatistics*, **0**: 1-31.

Henderson, R., Diggle, P. and Dobson, A. (2000), Joint Modeling of Longitudinal Measurements and Event Time Data, *Biostatistics*, **4**:465-480.

Klein, J. P. and Moeschberger, M. L. (2003), *SURVIVAL ANALYSIS- Techniques for Censored and Truncated Data*, Second Edition, Statistics for Biology and Health, Springer LLC.

- Lawn, S. D., Myer, L., Orrell, C., Bekker, L. G. and Wood, R. (2005), Early Mortality Among Adult Saccessing a Community-Based Antiretroviral Service in South Africa: Implications for Programme Design, *AIDS*, **19**, 2141-2148.
- Li, S., Zheng, M. and Gao, S. (2017), Joint Modeling of Longitudinal Cholesterol Measurements and Time To Onset of Dementia in An Elderly African American Cohort, *Biostatistics & Epidemiology*, **1**: 148-160.
- Sousa, I. (2011), A Reviw on Joint Modelling of Longitudinal Measurements And Time-To-Event, *REVSTAT Statistical Journal*, **9**: 57-81.
- Sweeting, M. J. and Thompson, S. G. (2011), Joint Modelling of Longitudinal And Time-To-Event Data with Application to Predicting Abdominal Aortic Aneurysm Growth and Rupture, *Biometrical Journal*, **53**: 750–763.
- Tseng, Y., Hsieh, F. and Wang, J. L. (2005), Joint Modelling of Accelerated Failure Time and Longitudinal Data, *Biometrika*, **92**: 587-603.
- Tsiatis, A. A. and Davidian, M. (2004), Joint Modeling of Longitudinal and Time-To-Event Data: An Overview, *Statistica Sinica*, **14**: 809-834.
- Wei, L. J. (1992), The Accelerated Failure Time Model: A Useful Alternative to the Cox Regression Model in Survival Analysis, *Stat Med*, **11**: 1871-9.
- WHO, (2018), Global Health Observatory Data; Available at: [www.who.int/gho/hiv/epidemic\\_status/cases\\_all/en](http://www.who.int/gho/hiv/epidemic_status/cases_all/en).
- Wu, L., Liu, W., Grace, Y. and Hung, Y. (2012), Analysis of Longitudinal and Survival Data: Joint Modeling, Inference Methods, and Issues, *Journal of Probability and Statistics*. **1**: 1-17.

Journal of Statistical Sciences, Spring and Summer, 2021  
Vol. 15, No. 1, pp 147-164  
DOI: 10.29252/jss.15.1.23

## **Joint Model of Longitudinal Count and Time to Event Data with Excess Zeros Using the AFT Model: A Case Study of the HIV / AIDS Dataset**

Zeinali Najafabadi, M. and Bahrami Samani, E.  
Department of Statistics, Shahid Beheshti University, Tehran, Iran.

**Abstract:** This article aims to joint modeling of longitudinal CD $\kappa$  cells count and time to death in HIV patients based on the AFT model. The modeling of the longitudinal count response, a GLME model under the family of PSD, was used. In contrast, for the TTE data, the parametric AFT model under the Weibull distribution was investigated. These two responses are linked through random effects correlated with the normal distribution. The longitudinal and survival data are then assumed independent, given the latent linking process and any available covariates. Considering excess zeros for two responses and right censoring, presented a joint model that has not yet been investigated by other researchers. The parameters were also estimated using MCMC methods.

**Keywords:** Accelerated failure time model, Time to event, Right censoring, Family of power series distribution, Generalized linear mixed effect (GLME) model.

**Mathematics Subject Classification (2010):** 62J05, 62J12.