مقایسه برآورد پارامترها در مدلهای کاکس و شکنندگی با وجود عوامل خطر ناشناخته

سلیمان خیری[1]، دکتر سقراط فقیه‌زاده، دکتر محمدرضا مشکانی، دکتر محمود محمودی، دکتر غلامرضا بابائی

چکیده مقاله

یکی از متداولترین مدلهای مورد استفاده در تحلیل داده‌های بقا مدل خطر متناسب کاکس می‌باشد. در این مدل، ممکن است تابع خطر به یک مجموعه از عوامل خطر ناشناخته یا غیرقابل اندازه‌گیری بستگی داشته باشد. به دلیل اینکه امکان قرار دادن این عوامل در مدل کاکس وجود ندارد، قسمتی از تغییرات پاسخ که می‌توانست با وجود آنها تبیین گردد به جمله خطا اضافه می‌شود. این مسئله باعث می‌گردد که برآورد پارامترها در این مدل اریب و گمراه کننده گردند. برای برطرف کردن این مشکل، یک اثر تصادفی به نام اثر شکنندگی به عنوان نماینده عوامل خطر ناشناخته در مدل کاکس قرار داده و مدل حاصل را مدل شکنندگی می‌نامند.

در این بررسی با انجام یک مطالعه شبیه‌سازی، برآورد پارامترها در مدلهای کاکس و شکنندگی با وجود عوامل خطر ناشناخته مقایسه شده است. نتایج حاصل از شبیه‌سازی در حالت داده‌های بقای یک متغیره نشان می‌دهد که هرچه تعداد عوامل خطر ناشناخته بیشتر باشد، میزان اریبی و میانگین توان دوم خطای برآوردها در مدل کاکس نسبت به مدل شکنندگی بیشتر می‌گردند. همچنین در حالت داده‌های بقای دو متغیره، هرچه میزان همبستگی زمانهای بقا بیشتر باشد، میزان اریبی و خطای برآوردها در مدل کاکس بیشتر می‌شوند. در این حالت، مدل شکنندگی با لحاظ نمودن همبستگی زمانهای بقا، نقص مدل کاکس را به اندازهٔ قابل ملاحظه‌ای برطرف می‌کند.

● *واژه‌های کلیدی: تحلیل بقا، عوامل خطر ناشناخته، داده‌های بقای چندمتغیره، مدل خطر متناسب کاکس، مدل شکنندگی.*

مقدمه

در بسیاری از مطالعات پزشکی، گروهی از آزمودنی‌ها در طی دورهٔ معینی وارد مطالعه گشته و مورد پیگیری قرار می‌گیرند تا حادثه‌ای معین برای آنها به وقوع بپیوندد. از این رو یکی از مسائل مورد علاقه محققات مطالعه فاصله زمانی از ورود به مطالعه تا وقوع حادثه موردنظر و تعیین رابطه آن با متغیرهای کمکی برای بررسی عوامل خطر می‌باشد. مشاهدات زمانی حاصل از این مطالعات را که با داده‌های سانسور شده همراه هستند "داده‌های بقا" و مجموعه روشهای آماری برای تجزیه و تحلیل آنها را "تحلیل بقا" می‌نامند.

داده‌های بقا ممکن است یک متغیره یا چند متغیره باشند. اکثر داده‌های بقا از نوع داده‌های بقای یک متغیره هستند، در این نوع داده‌ها هر آزمودنی تحت مطالعه، حادثه نهایی را حداکثر یک بار تجربه خواهد کرد و فرض می‌شود زمانهای بقا برای آزمودنی‌های مختلف مستقل از یکدیگر هستند. داده‌های بقای چند متغیره متشکل از چندین گروه مستقل از زمانهای بقا هستند. با این فرض که زمانهای بقا در هر گروه بواسطه تأثیر متغیرهای ناشناخته ژنتیکی و محیطی همبسته بوده، اما در گروههای متفاوت مستقل از هم هستند. (۱)

شناسایی تابع توزیع بقا و تابع خطر و برآورد پارامترهای آنها از مهمترین اهداف تحلیل بقا می‌باشد. تابع بقا رفتار احتمالاتی زمان بقا را نشان می‌دهد و تابع خطر نمایانگر قدرت آنی خطر در زمان مشخص می‌باشد. این دو تابع اطلاعات هم ارزی در مورد زمانهای بقا در اختیار محقق قرار می‌دهند. با این حال در بسیاری از کاربردهای بقا، هدف مورد توجه بررسی رابطه متغیرهای کمکی با زمان بقا می‌باشد.

در بعضی از مطالعات، بخصوص مطالعاتی که روی انسان انجام می‌گیرد ممکن است عواملی ناشناخته غیر از متغیرهای کمکی وجود داشته باشند که توزیع زمان بقا و به دنبال آن تابع خطر را شدیداً تحت تأثیر قرار دهند. اما به دلیل ناشناخته بودن و یا ناتوانی در اندازه‌گیری آنها قادر به گنجاندن انها در مدل رگرسیونی نباشیم (مانند عوامل ناشناخته ژنتیکی یا محیطی). در حالت کلی چنانچه عاملی روی متغیر پاسخ اثرگذار باشد، لازم است مقدار آنرا برای هر آزمودنی به دست آورده و به عنوان متغیر کمکی در مدل وارد نمائیم، اما اگر چنین کاری امکان‌پذیر نباشد تغییراتی از تابع خطر که می‌توانست با وجود عامل مربوطه تبیین شده و کاهش یابد، با جمله خطا جمع شده و در نتیجه باعث افزایش تغییرات تابع خطر نسبت به حالتی که عامل موردنظر در مدل وجود داشته باشد می‌گردد. تغییرات افزایش یافته منجر به تغییر شکل تابع خطر گردیده و باعث می‌شود که ضرایب رگرسیونی تغییر نموده و از نظر عددی به صفر نزدیکتر شده و در حالت کلی موجب به دست آوردن برآوردگرهای اریب و گمراه‌کننده‌ای از پارامترهای رگرسیونی گردند. عوامل خطر ناشناخته در حالت داده‌های بقای یک متغیره باعث افزایش ناهمگنی (Heterogeneity) زمانهای بقای آزمودنیها گردیده و در حالت داده‌های بقای چندمتغیره باعث ایجاد همبستگی بین زمانهای بقا در درون گروهها می‌گردد. در هر دو حالت فوق، برآورد پارامترها با استفاده از مدل معمول خطر متناسب کاکس (Cox Propotrional Hazard Model)، اریب و

[1] - دانشجوی دکترای آمار زیستی دانشگاه تربیت مدرس

گمراه‌کننده می‌گردد. اما با استفاده از مدلهای شکنندگی (Frailty Models) می‌توان این نقص را برطرف نمود (۱)، (۲) و (۳).

در این بررسی به کمک یک مطالعه شبیه‌سازی نشان می‌دهیم که با وجود عوامل خطر ناشناخته، مدل خطر کاکس برآورد اریب و گمراه‌کننده‌ای از پارامترها بدست داده اما مدل شکنندگی برآورد دقیقتری از پارامترها، بخصوص برای داده‌های بقای چند متغیره بدست می‌دهد.

روشها

الف- مدل خطر متناسب کاکس

فرض کنیم یک نمونه تصادفی از زمانهای بقا وجود داشته و T_i نشان دهنده زمان بقا و $\lambda_i(t)$ تابع خطر در زمان t برای آزمودنی i باشد. در مدل خطر متناسب، تابع خطر در زمان t برای آزمودنی i با وجود بردار متغیرهای کمکی $Z_i = (Z_{i1}, \ldots, Z_{ip})$ به صورت زیر می‌باشد:

$$\lambda_i(t'\lambda_o, z_i) = \lambda_o(t)\exp(\beta' z_i)$$

که در آن β بردار پارامترهای رگرسیون و $\exp(\beta' z_i)$ تابع خطر نسبی می‌باشد. تحت این مدل متغیرهای کمکی دارای اثرات ضربی روی تابع خطر بوده و پارامترهای رگرسیونی همچون لگاریتم خطرات نسبی تفسیر می‌گردند. تابع $\lambda_o(t)$ تابع خطر مبنا است که سطح عمومی خطر را برای کلیهٔ آزمودنیها مشخص می‌سازد.

ب- مدل شکنندگی برای داده‌های بقای یک متغیره

شرط استفاده از مدل کاکس در حالت داده‌های بقای یک متغیره، این است که صرفنظر از اثر متغیرهای کمکی، بین زمانهای بقا ناهمگنی وجود نداشته باشد. ناهمگنی زمانهای بقا را می‌توان ناشی از تأثیر یکسری عوامل خطر ناشناخته در نظر گرفت و برای مقابله با آن یک اثر تصادفی غیرقابل اندازه‌گیری در تابع خطر قرار می‌دهیم. اگر Z بردار متغیرهای معلوم و W بردار متغیرهای ناشناخته باشند که هر دو در تابع خطر تأثیر داشته باشند، در این صورت مدل حقیقی خطرات متناسب کاکس در زمان t به شکل $\lambda_o(t)\exp(\beta' z_i+\psi' w)$ خواهد شد. چون W ناشناخته یا غیرقابل اندازه‌گیری است $\exp(\psi w)$ را تصادفی فرض نموده و آن را به عنوان اثر تصادفی یا اثر شکنندگی Y در نظر گرفته و به شکل زیر در مدل قرار می‌دهیم. این مدل نشان می‌دهد که خطر در زمان t به شرط معلوم بودن متغیر Y و وجود بردار متغیرهای کمکی برای آزمودنی i به صورت زیر می‌باشد:

$$\lambda_i(t \setminus y_i, \lambda_o, \beta) = y_i \lambda_o(t)\exp(\beta' z_i)$$

مدل فوق را مدل کاکس اصلاح شده یا مدل شکنندگی بقا می‌نامند (۴). در نتیجه بکارگیری این مدل، با فرض اینکه اثر شکنندگی مستقل از متغیرهای کمکی موجود در مدل باشد آنگاه تغییرات کل تابع خطر و به دنبال آن تغییرات تابع بقا به اثر متغیرهای نامعلوم، اثر عوامل خطر ناشناخته و اثر تصادفی مدل تجزیه می‌گردد. تفسیر اثر شکنندگی به این صورت می‌باشد که با فرض ثابت بودن مقادیر متغیرهای تبیینی معلوم، چنانچه برآورد y_i بزرگتر از برآورد y_j باشد، گوییم آزمودنی i شکننده‌تر از آزمودنی j است. یعنی خطر شکست برای آزمودنی i بیشتر از آزمودنی j است. در حالت کلی چنانچه برآورد y_i بزرگتر از یک باشد، گوییم خطر نسبی عوامل ناشناخته برای آزمودنی i بیشتر از حد متوسط است و برعکس. (۱) و (۵).

مدلهای شکنندگی برای داده‌های بقای چندمتغیره

مدلهای شکنندگی بطور گسترده برای مدل بندی همبستگی داده‌های بقای چندمتغیره بکار گرفته شده است. فرض کنیم داده‌های بقای چندمتغیره متشکل از n گروه مستقل بوده و در هر گروه k زمان بقای همبسته وجود داشته باشد. برای سادگی اندیس i را برای گروه و اندیس j را برای اعضای درونی گروه قرار می‌دهیم $(i = 1, \ldots, n \;,\; j = 1,\ldots, k)$. در حالت کلی تعداد زمانها در هر گروه می‌تواند متفاوت باشد که در آن صورت k به k_i تبدیل می‌شود. در داده‌های بقای چندمتغیره، در هر گروه عوامل خطر غیرقابل مشاهده و اندازه‌گیری (همچون عوامل ژنتیکی و محیطی) وجود دارد که باعث ایجاد همبستگی زمانهای بقا در هر گروه می‌گردند.

به خاطر همبستگی زمانهای بقا در هر گروه، مدل رگرسیون کاکس قابل استفاده نبوده و لازم است این مدل را به نحوی اصلاح نمود که بتواند همبستگی زمانهای بقا را تبیین نماید. فرض کنیم T_{ij} زمان بقای عضو j از گروه i باشد. به منظور مدل بندی زمانهای بقای همبسته، یک مقدار تصادفی و مشترک برای اعضای هر گروه به نام اثر شکنندگی درنظر گرفته و تابع خطر هر عضو گروه را در این مقدار مشترک ضرب می‌کنیم، دراین صورت تابع خطر برای آزمودنی j از گروه i در زمان t برابر با حاصل ضرب $y_i\lambda(t)$ می‌باشد که در آن $\lambda(t)$ مدل کاکس و Y مقدار اثر شکنندگی گروه i است. بنابراین تابع خطر شرطی در زمان t به شرط متغیر تصادفی Y و با وجود متغیرهای کمکی Z به صورت زیر می‌باشد:

$$\lambda_{ij}(t/y_i,z_{ij}) = y_i\, \lambda_o(t) \exp(\beta\, z_{ij}) \quad i = 1, \ldots, n\, ; j = 1, \ldots, k$$

که در آن β بردار پارامترهای رگرسیونی مجهول و $\lambda_0(t)$ تابع خطر مبنا است. در حالت کلی فرض می‌شود که متغیرهای تصادفی Y مستقل از هم و هم توزیع باشند. در اینجا نیز همانند مدل بندی داده‌های بقای یک متغیره توزیع Y را توزیع مثبتی فرض نموده و عموماً میانگین آن را یک در نظر می‌گیرند. واریانس Y تغییرات زمان بقای بین گروهها را نشان می‌دهد و هرچه مقدار آن بیشتر باشد همبستگی زمانهای بقای درون گروهها و تغییرات زمانهای بقا بین گروهها بیشتر است (۱).

باتوجه به اینکه تابع خطر مثبت است در هر دو حالت داده‌های یک متغیره و چندمتغیره بقا لازم است برای اثر شکنندگی توزیع مثبتی در نظر گرفته شود. توزیعهای مثبتی چون گاما، پایای مثبت (Positive Stable)، لگ نرمال، گاوسی وارون (Inverse Gaussian) و توزیع واریانس توانی (Power Variance) با میانگین متناهی برای متغیر Y در نظر گرفته می‌شود. اما در عمل بیشتر از توزیع گاما و توزیع پایای مثبت استفاده می‌شود. همچنین به خاطر ساده‌سازی محاسبات و از آنجا که اثر شکنندگی نماینده خطر نسبی عوامل ناشناخته می‌باشد در بیشتر موارد میانگین توزیع اثر شکنندگی را برابر با یک فرض می‌نمایند (۶)، (۷) و (۸).

برآورد پارامترها در مدلهای شکنندگی با استفاده از روش درستنمایی جزئی کاکس امکان‌پذیر نبوده و لازم است که اثر شکنندگی و پارامترهای مدل به صورت همزمان برآورد گردند. برای انجام این کار روشهای متفاوتی وجود دارد اما ساده‌ترین آنها الگوریتم EM می‌باشد پس از برازش مدل، مقادیر Yها برآورد می‌گردند (۹).

د-شبیه‌سازی

با کمک شبیه‌سازی به مقایسه برآورد پارامترهای مدل کاکس و مدل شکنندگی با وجود عوامل خطر ناشناخته پرداخته‌ایم. در این بررسی زمانهای بقای ساختگی از یک مدل رگرسیون معلوم تولید نموده، سپس دو مدل فوق را به این زمانها برازانده‌ایم. باتوجه به اینکه انگیزه به کارگیری مدل شکنندگی برای داده‌های یک متغیره و چندمتغیره بقا متفاوت است، لازم است دو بررسی جداگانه برای هرکدام انجام دهیم. برای تولید زمانهای بقای تصادفی از مدل رگرسیون معلوم وایبول استفاده کرده‌ایم. توزیع وایبول دارای دو پارامتر به نامهای پارامتر شکل α و پارامتر مقیاس η بوده و کاربرد زیادی در تحلیل بقا دارد تابع چگالی توزیع وایبول با پارامترهای (α,η) به صورت زیر است:

$$f_T(t) = \alpha\eta\,(\alpha\eta)^{a-1}\exp(-(\eta t)^{a}) \quad ; \quad t>O,\ \alpha>O,\ \eta>O$$

در بسیاری از کاربردها پارامتر شکل این توزیع بستگی زیادی با متغیرهای کمکی ندارد. بنابراین در مدل رگرسیون وایبول پارامتر شکل را ثابت در نظر گرفته، پارامتر مقیاس را با متغیرهای کمکی مرتبط می‌کنند (۱۰).

الف- تولید داده‌های بقای یک متغیره

برای تولید زمانهای بقای یک متغیره فرض نموده‌ایم که پارامتر مقیاس توزیع وایبول متأثر از ۵ متغیر کمکی با ضرایب معلوم به صورت زیر باشد:

$$\eta(x) = \exp(0.5x_1+0.07x_2-x_3+0.3x_4-0.3x_5)$$

به منظور تولید اعداد تصادفی برای متغیرهای کمکی از توزیعهای زیر استفاده کرده‌ایم:

مقدار x_1 را از توزیع برنولی که معرف جنس است، مقدار x_2 را از توزیع گاما با میانگین ۲۰ و واریانس ۴ که معرف سن است، مقدار x_3 را از توزیع برنولی که معرف گروه تیمار است، مقدار x_4 را از توزیع پواسن با میانگین ۰/۵ که معرف تعداد بیماری همراه است و مقدار x_5 را از توزیع نرمال با میانگین ۱۲ و واریانس ۲ که معرف مقدار فشار خون می‌باشد، تولید نموده‌ایم.

پس از تولید مقادیر متغیرهای کمکی برای هر آزمودنی، مقدار η را از رابطه قبل محاسبه نموده و متناظر با آن یک عدد تصادفی از توزیع وایبول با $a = 2$ و η به عنوان زمان بقای آن آزمودنی تولید می‌نماییم. این کار را تکرار می‌کنیم تا یک نمونه n مشاهده‌ای از زمانهای بقا تولید گردد. چنانچه $\eta_i<\eta_j$ باشد انتظار می‌رود زمان بقای تولید شده برای آزمودنی i بزرگتر از زمان بقای تولید شده برای آزمودنی j باشد. مقادیر زمانهای بقای تولید شده دارای چولگی مثبت بوده و در دامنه صفر تا ۷۰ قرار دارند. در مرحله بعد با به کارگیری سانسور راست و نوع اول، زمانهای بقای بزرگتر از ۲۰ را به عنوان زمان بقای سانسور شده در نظر می‌گیریم در نتیجه این کار حدود ۱۵ تا ۲۵ درصد از زمانها سانسور می‌گردند. با این کار مجموعه داده‌های بقا برای برازش مدلهای کاکس و شکنندگی آماده شده‌اند.

با انجام رگرسیون کاکس برحسب کلیه متغیرهای موجود مشاهده می‌شود که با افزایش حجم نمونه و افزایش تعداد تکرار نمونه‌گیری، متوسط برآورد پارامترها به دو برابر مقادیر اصلی پارامترها میل می‌کند. این مسئله از آنجا ناشی می‌شود که با ملاحظه توابع خطر کاکس و وایبول، در حالت کلی ضرایب پارامترها در مدل کاکس، a برابر ضرایب پارامترهای متناظر در مدل وایبول می‌باشد.

باتوجه به‌اینکه مقیاس اندازه‌گیری متغیرها در مطالعه حاضر مختلف است برای بدست آوردن تفسیر ساده و روشن از اریبی برآوردها از محک اریبی نسبی که برحسب درصد گزارش می‌شود استفاده می‌کنیم. اگر θ مقدار واقعی پارامتر و $\hat{\theta}$ برآورد آن باشد، اریبی منفی نسبی برابر با $\frac{E(\theta-\hat{\theta})}{\theta}$ می‌باشد. همچنین به خاطر آنکه با وجود عوامل خطر ناشناخته در برآوردها کم‌برآوردی بوجود می‌آید، میزان اریبی منفی نسبی مثبت خواهدشد. برای مقایسه برآورد پارامترها از دو معیار اریبی منفی نسبی و میانگین توان دوم خطا (Mean Square Error) استفاده کرده‌ایم. بر اساس این معیارها مدلی بهتر است که میزان اریبی و میانگین توان دوم خطای (MSE) برآوردهای آن کمتر باشد.

ب-تولید داده‌های بقای دومتغیره

با فرض اینکه تنها متغیر اثرگذار بر زمان بقا متغیر تیمار باشد. زمان بقای مورد j از گروه i را از توزیع وایبول با پارامتر شکل $a = 2$ و پارامتر مقیاس η_{ij} تولید می‌نماییم که در آن

$$\eta_{ij} = \exp(x_{1ij}) \ ; \ j = 1,2 \quad i = 1,\ldots,n$$

مقدار x_{ij} که معرف گروه تیمار می‌باشد از توزیع برنولی تولید می‌گردد. چنانچه در رابطه η_{ij} متغیر مشترک x_{2i} را اضافه نماییم آنگاه مقدار η برای دو آزمودنی هر گروه در یک جهت تغییر نموده که در نتیجه آن زمان بقای همبسته تولید خواهد گشت. هر اندازه که تعداد متغیرهای مشترک را زیاد کنیم مقادیر η_{i1} و η_{i2} همبستگی بیشتری خواهند داشت و در نتیجه آن زمانهای بقای زوجی با همبستگی بالاتری تولید خواهد شد. در اینجا متغیرهای مشترک، نقش عوامل خطر ناشناخته را ایفا می‌نمایند که هرچه تعداد آنها بیشتر گردد همبستگی زمانهای بقا بیشتر خواهد شد.

نتایج

الف-داده‌های بقای یک متغیره

همانگونه که در بخش قبل تشریح شد زمانهای بقای یک متغیره تولید شده متأثر از پنج متغیر کمکی (پنج عامل خطر) هستند. برای نشان دادن اثر متغیرهای خطر ناشناخته، ابتدا مدلهای کاکس و شکنندگی را با وجود پنج متغیر برازانده سپس در هر مرحله یک متغیر از متغیرهای موجود در مدل را کم می‌کنیم (به عنوان متغیر خطر ناشناخته) و تأثیر آن را در صد اریبی منفی

نسبی و میانگین توان دوم خطای (MSE) دیگر برآوردهای موجود در مدل ملاحظه می‌نماییم. به عنوان نمونه در مدل ۳ جدول (۱) متغیرهای X_1, X_2, X_3 نقش متغیرهای معلوم و متغیرهای X_4, X_5 نقش متغیرهای ناشناخته را ایفا می‌کنند. برای اینکه اثر حجم نمونه را در این مسئله روشن سازیم، مسئله را برای سه حجم نمونه ۳۰، ۱۰۰ و ۳۰۰ مشاهده‌ای انجام داده‌ایم. بنابراین باتوجه به تعداد متغیرها و حجم‌های نمونه و دو مدل خطر کاکس و شکنندگی، در حالت کلی ۳۰ مدل مختلف برازانده شده و برازش هر کدام از این مدلها ۲۰۰۰ بار تکرار شده است. مقادیر درصد اریبی منفی نسبی و MSE برآورد پارامترهای موجود در مدل در جدول (۱) آمده است.

باتوجه به مقادیر جدول (۱) نتایج زیر بدست می‌آید:

۱ـ باتوجه به نتایج مدل ۱ چنانچه هر پنج متغیر تأثیرگذار بر زمانهای بقا در مدل وجود داشته باشند، تفاوت قابل ملاحظه‌ای میان درصد اریبی منفی نسبی و مقدار MSE برآوردهای مدل کاکس و شکنندگی وجود ندارد. همچنین اریبی برآوردها با افزایش حجم نمونه به مقدار صفر میل می‌کند.

۲ـ با کاهش تعداد متغیرهای موجود در مدل (افزایش متغیرهای ناشناخته)، میزان اریبی منفی نسبی برآوردهای مدل کاکس افزوده شده است تا آنجا که در مدل آخر که تنها متغیر X_1 در مدل وجود دارد، بیش از ۵۰ درصد اریبی منفی نسبی دیده می‌شود.

۳ـ مقادیر اریبی برآوردها در مدل شکنندگی در مقایسه با مدل کاکس بسیار کمتر است. اما اگر تعداد متغیرهای ناشناخته در برابر متغیرهای موجود در مدل زیاد گردد، مدل شکنندگی نیز قادر به ارائه برآوردهای نااریب نخواهد بود.

۴ـ با به کارگیری مدل شکنندگی در مقایسه با مدل کاکس در حضور حجم نمونه بالا، MSE برآوردها خیلی کمتر شده است. یعنی مدل شکنندگی برآورد دقیق‌تری برای پارامترها ارائه می‌دهد.

۵ـ باتوجه به نتایج مدلهای ۲ تا ۵ با وجود عوامل خطر ناشناخته، افزایش حجم نمونه تأثیری در اصلاح اریبی برآوردهای مدل کاکس نداشته اما میزان اریبی برآوردهای مدل شکنندگی را اندکی کاهش داده است. از طرفی باتوجه به کاهش MSE برآوردها در هر دو مدل، نتیجه می‌شود که با افزایش حجم نمونه دقت برآوردها افزایش می‌یابد.

جدول ۱: محاسبه اریبی منفی نسبی و MSE با استفاده از مدلهای کاکس و شکنندگی براساس ۲۰۰۰ بار تکرار در داده‌های بقای یک‌متغیره با فرض اینکه متغیرهای اثرگذار بر زمانهای بقا، متغیرهای $X_۱, X_۲, X_۳, X_۴, X_۵$ می‌باشند. (×)

مدل		حجم نمونه											
		۳۰				۱۰۰				۳۰۰			
	متغیر	خطر کاکس		خطر شکنندگی		خطر کاکس		خطر شکنندگی		خطر کاکس		خطر شکنندگی	
	موجود	اریبی	MSE	اریبی	MSE	اریبی	MSE	اریبی	MSE	اریبی	MSE	اریبی	MSE
۱	$X_۱$	۱۵/۸	۰/۴۴۴	۱۶/۲	۰/۴۵۶	۳/۴	۰/۰۶۹	۶/۹	۰/۰۷۹	۲/۱	۰/۲۲	۲/۶	۰/۰۲۲
	$X_۲$	۱۸/۴	۰/۰۲۸	۱۶/۵	۰/۰۲۶	۴/۳	۰/۰۰۴	۶/۱	۰/۰۰۵	۰/۸	۰/۰۰۱	۲/۱	۰/۰۰۱
	$X_۳$	۱۷/۱	۰/۷۲۸	۱۷/۳	۰/۷۶۳	۳/۹	۰/۰۳	۵/۸	۰/۱۱۹	۱/۲	۰/۰۲۷	۲/۱	۰/۰۳۴
	$X_۴$	۲۲/۵	۰/۲۴۲	۲۱/۸	۰/۲۸۷	۵/۸	۰/۰۳۳	۷/۱	۰/۰۳۳	۲/۲	۰/۰۰۹	۳/۲	۰/۰۱
	$X_۵$	۱۷/۲	۰/۵۴	۱۸/۶	۰/۰۶۴	۴/۳	۰/۰۰۸	۶	۰/۰۱۰	۱/۳	۰/۰۲	۲/۴	۰/۰۰۳
۲	$X_۱$	۲۶/۸	۰/۳۹۸	۱/۲۱	۰/۴۷۰	۳۴/۵	۰/۱۸۳	۵/۴	۰/۱۴۲	۲۶/۸	۰/۱۵۵	۱۲/۶	۰/۰۵۷
	$X_۲$	۲۴/۳	۰/۰۰۷	۸/۱۲	۰/۰۲۸	۳۴	۰/۰۰۳	۴/۳	۰/۰۰۸	۳۶/۱	۰/۰۰۶	۱۴/۴	۰/۰۰۳
	$X_۳$	۲۶/۹	۰/۶۹۷	۱۹	۰/۷۲	۳۵/۴	۰/۵۷۳	۴/۹	۰/۲۰۳	۳۷	۰/۵۷۲	۱۲/۹	۰/۱۳۸
	$X_۴$	۲۰/۱	۰/۲۰۸	۱۱	۰/۲۴۷	۳۲/۱	۰/۰۷۲	۳/۳	۰/۰۶۴	۲۵/۶	۰/۰۵۶	۱۲/۸	۰/۰۲۳
۳	$X_۱$	۳۲/۲	۰/۲۸۴	۲۳/۲	۰/۴۳	۳۷/۴	۰/۲۰۱	۱۳/۲	۰/۱۳۷	۳۹/۳	۰/۱۷۴	۱۸/۲	۰/۰۷۰
	$X_۲$	۳۰/۴	۰/۰۰۷	۲۲/۶	۰/۰۲۶	۳۶/۲	۰/۰۰۴	۱۲/۳	۰/۰۰۸	۳۹	۰/۰۰۴	۱۸/۲	۰/۰۰۳
	$X_۳$	۳۰/۸	۰/۷۲۱	۲۱/۱	۱/۰۰۶	۳۸/۴	۰/۶۵۸	۱۳/۲	۰/۲۲۳	۴۰/۱	۰/۶۶۵	۱۸/۹	۰/۱۹۵
۴	$X_۱$	۴۷/۲	۰/۴۵۴	۳۸/۹	۰/۴۸۸	۴۸/۷	۰/۲۹۴	۳۲/۵	۰/۲۰۱	۴۹/۵	۰/۲۶۲	۳۱/۵	۰/۱۶۳
	$X_۲$	۴۲/۱	۰/۰۰۸	۳۲/۸	۰/۰۲۲	۴۷/۲	۰/۰۰۵	۳۳/۶	۰/۰۰۸	۴۹/۱	۰/۰۰۶	۳۲	۰/۰۰۵
۵	$X_۱$	۵۰/۲	۰/۴۶۴	۴۲/۳	۰/۴۵۷	۵۰/۹	۰/۳۱	۳۷/۷	۰/۲۲۶	۵۰/۱	۰/۲۶۸	۳۴	۰/۱۷۶

(×) تذکر: مقادیر اریبی موجود در جدول مقادیر اریبی منفی نسبی هستند که برحسب درصد گزارش شده‌اند.

ب- داده‌های بقای دومتغیره

ایده استفاده از مدل شکنندگی در داده‌های بقای چندمتغیره، توصیف همبستگی زمانهای بقا می‌باشد. برای نشان دادن این موضوع چند مدل با درجات مختلف همبستگی زمانهای بقا در نظر می‌گیریم. فرض می‌کنیم که در همه این مدلها تنها متغیر مورد علاقه شاخص، گروه تیمار است. هدف مقایسه برآورد ضریب این متغیر با استفاده از مدلهای کاکس و شکنندگی با وجود همبستگی زمانهای بقا می‌باشد. در مدل ۱ جدول (۲)، از زمانهای بقای مستقل استفاده شده است. در نتیجه همبستگی زمانهای بقا صفر شده است. در مدلهای بعد، با روشی که در بخش قبل ارائه شده از زمانهای بقای همبسته استفاده شده است. با اضافه نمودن متغیرهای ناشناخته، میزان همبستگی زمانها به ترتیب افزایش یافته تا آنجا که در مدل آخر این میزان به ۳۵ درصد رسیده است. نتایج این بررسی برای حجم نمونه ۶۰ مشاهده‌ای (۳۰ گروه زوجی)، ۲۰۰ مشاهده‌ای (۱۰۰ گروه زوجی) و ۶۰۰ مشاهده‌ای (۳۰۰ گروه زوجی) با ۱۰۰۰ بار برازش برای هر مدل در جدول ۲ آمده است.

جدول ۲: اریبی منفی نسبی و MSE مدل کاکس و شکنندگی بر اساس ۱۰۰۰ بار تکرار در داده‌های بقای دومتغیره

حجم نمونه	مدل		۱	۲	۳	۴	۵
۶۰	ضریب همبستگی زمانها		۰/۰۱	۰/۰۶	۰/۱۷	۰/۲۲	۰/۳۴
	کاکس	اریبی	۳/۷	۶/۸	۲۰/۰	۲۷/۳	۳۳/۲
		MSE	۰/۱۵۶	۰/۱۷۱	۰/۲۹۴	۰/۴۱۱	۰/۵۷۲
	شکنندگی	اریبی	۴/۷	۰/۷	۰/۷	۱	۴/۷
		MSE	۰/۱۷۶	۰/۱۷۷	۰/۲۲۲	۰/۲۳۸	۰/۲۸
۲۰۰	ضریب همبستگی زمانها		۰/۰	۰/۰۶	۰/۱۷	۰/۲۲	۰/۳۵
	کاکس	اریبی	۰/۴	۹	۲۲/۴	۲۸/۴	۳۵
		MSE	۰/۰۳۸	۰/۰۷۲	۰/۲۳۲	۰/۳۵۷	۰/۵۲۵
	شکنندگی	اریبی	۲	۱	۰/۱	۰/۳	۳/۴
		MSE	۰/۰۴۶	۰/۰۵۵	۰/۰۶۸	۰/۰۷۴	۰/۰۸۸
۶۰۰	ضریب همبستگی زمانها		۰	۰/۶	۰/۱۷	۰/۲۳	۰/۳۴
	کاکس	اریبی	۰/۱	۹/۲	۲۲/۹	۲۹/۶	۳۵/۴
		MSE	۰/۰۱۳	۰/۰۴۶	۰/۲۲	۰/۳۶۲	۰/۵۱۳
	شکنندگی	اریبی	۰/۴	۰/۱	۰/۵	۰/۴	۴
		MSE	۰/۰۱۴	۰/۰۲	۰/۰۲۳	۰/۰۲۳	۰/۰۳۴

تذکر: مقادیر اریبی موجود در جداول مقادیر اریبی منفی نسبی هستند که برحسب درصد گزارش شده‌اند.

باتوجه به مقادیر جدول (۲) نتایج زیر بدست می‌آید:

۱- باتوجه به مدل ۱ چنانچه زمانهای بقا مستقل باشند تفاوت قابل ملاحظه‌ای میان درصد اریبی منفی نسبی و میانگین توان دوم خطای برآورد اثر تیمار در مدلهای کاکس و شکنندگی وجود ندارد.

۲- با افزایش میزان همبستگی زمانهای بقا، درصد اریبی منفی نسبی برآورد اثر تیمار در مدل کاکس افزایش یافته و این میزان تا ۳۵ درصد نیز رسیده، اما میزان اریبی منفی برآورد اثر تیمار توسط مدل شکنندگی بسیار ناچیز و قابل صرف نظر کردن است. یعنی مدل شکنندگی با وجود همبستگی زمانهای بقا برآورد تقریباً نااریب از پارامترها ارائه می‌دهد.

۳- با افزایش میزان همبستگی، میانگین توان دوم خطا (MSE) در مدل کاکس افزایش یافته اما این افزایش برای مدل شکنندگی بسیار ناچیز است. یعنی مدل شکنندگی با وجود همبستگی زمانهای بقا برآورد دقیقتری نسبت به مدل کاکس ارائه می‌دهد.

۴- گرچه افزایش حجم نمونه باعث کاهش میانگین خطای برآورد در هر دو مدل شده، اما افزایش حجم نمونه کمکی به اصلاح اریبی منفی برآورد در مدل کاکس نکرده است.

بحث

کلمه شکنندگی در ابتدا توسط ووپل به منظور تشریح پیامدهای ناشی از وجود چندین منبع تغییرات، برای داده‌های یک متغیره طول عمر پیشنهاد شده است (۱۱). اما براساس مطالعات موجود (۸) و نتایج شبیه‌سازی حاضر مباستر است که مدل شکنندگی برای تبیین همبستگی موجود در داده‌های بقای چندمتغیره بکار رود. این مدل اولین بار در سال ۱۹۸۷ بوسیلهٔ کلیتون برای داده‌های بقای دومتغیره به منظور بررسی همبستگی زمان بقای حمله قلبی پدر و پسر به کار گرفته شده است (۱۲).

مدلهای شکنندگی در بقا به مدلهای تصادفی بقا نیز مشهور شده‌اند. این

مدلها، مدلهای نسبتاً جدیدی در بقا بوده و بصورت گسترده‌ای در دههٔ ۱۹۹۰ مورد مطالعه قرار گرفته و هم اکنون نیز موضوع بسیاری از تحقیقات را به خود اختصاص داده است. مشکلات تکنیکی در برآورد پارامترهای این مدل باعث شده که در مقایسه با مدل کاکس کمتر استفاده شوند. اخیراً تجزیه و تحلیل حالات خاصی از این مدلها در نرم‌افزارهای آماری SAS و Splus آمده است (۱).

هندرسون و عمان (۱۳) به روش نظری مسئله عدم استفاده از مدل شکنندگی هنگامی که اثر شکنندگی وجود دارد را بررسی نموده و نتیجه گرفته‌اند که در این حالات اریبی منفی در برآوردهای رگرسیون ایجاد می‌گردد و بزرگی اریبی به واریانس اثر شکنندگی و توزیع آن بستگی دارد. شوماخر و همکاران (۱۴) نشان داده‌اند که چگونه حذف یک عامل مهم، باعث کاهش برآورد خطر نسبی می‌گردد. کیدینگ و همکاران (۳) به صورت نظری نشان داده‌اند که چگونه حذف یکی از دو متغیر کمکی باعث افزایش واریانس تابع خطر و اریبی برآورد دیگر متغیر موجود در مدل می‌گردد. ایشان همچنین تأثیر به کارگیری اثر شکنندگی را در اصلاح واریانس تابع خطر و اریبی برآورد نشان داده‌اند. در عین حال ایشان نتیجه گرفته‌اند که برای مقابله با اثر متغیرهای ناشناخته در داده‌های یک متغیره بقا بهتر است از مدل زمان شکست شتابیده (Accelerated Failure Time Model) استفاده شود.

در بحث شبیه‌سازی این مطالعه، زمانهای بقا از مدل رگرسیون وایبول تولید شده و داده سانسور شده از نوع سانسور راست و دورهٔ مطالعه بر اساس روش سانسور نوع اول تعیین شده است. همچنین برای اثر شکنندگی از توزیعهای گاما و گاوسی وارون استفاده شده است. به عنوان مطالعات دیگر می‌توان این بررسی را با تولید مشاهدات از دیگر مدلهای رگرسیون بقا، بکارگیری روش سانسور نوع دوم و دیگر توزیعهای شکنندگی انجام داد. کلیه محاسبات شبیه‌سازی با نرم‌افزار Splus انجام شده است. علاقمندان جهت دریافت برنامه تهیه شده می‌توانند با نویسنده تماس بگیرند.

مراجع

1- Hougaard, P., *Analysis of multivariate survival data, New York: Springer-Verlag, 2000.*

2- Hougaard P., Myglegaard P. and Borch-Johansen K., *Heterogeneity models of disease susceptibility, with application to diabetic nephropathy, Biometrics, 1994, 50, 1178-1188.*

3- Keiding N., Andersen P. K. and Klein P. J., *The role of frailty models and accelerated failure time models in describing heterogeneity due to omitted covariates, Statistics in Medicine, 1997, Vol. 16, 215-224.*

4- Clayton, D., *A monte carlo method for Bayesian inference in frailty models, Biometrics, 1991, 47, 467-485.*

5- Hosmer D. W. and Lemeshow S., *Applied survival analysis, 1999, wiley.*

6- Bolstad W. and manda O. S., *Investigating child mortality in Malawi using family and community random effects: A Bayesian analysis, Journal of the American Statistical Association, 2001, 96, 12-19.*

7. Oakes, D., *Bivariate survival models induced by frailty. Journal of the American statistical Association, 1989, 84, 487-493.*

8- Hougaard P., *Frailty: In Encyclopedia of Biostatistics, Armitage P. and colton T., New York: Wiley, 1997.*

9- Klein, J. P., *Semiparametric estimation of random effects using the Cox model based on the EM algorithm, Biometrics. 1992, 48, 795-806.*

10- Lawless J. F., *Parametric models in survival analysis, In Encyclopedia of Biostatistics, Armitage P. and Colton T., New York: Wiley, 1997.*

11- Vaupel J. W., Manton K. G., and Stallavd E., *The impact of heterogeneity in individual frailty on the dynamics of mortality, Demography, 1997, 16, 439-454.*

12- Clayton D. G., *A. model for association in bivariate life-table and its application in epidemiological studies of chronic disease incidence, Biometrica. 1978, 65, 141-151.*

13- Henderson R. and Oman P., *Effect of frailty on marginal regression, 1999, 61, 367-379.*

14- Schumacher M., Olschewski M. and Schmoor C., *The impact of heterogeneity on the comparison of survival times, Statistics in Medicinge, 1987, 6, 773-784.*