مجله مدل‌سازی پیشرفته ریاضی، دوره ۴، شماره ۲، پاییز و زمستان ۱۳۹۳

مدل‌های شکنندگی و خطرهای متناسب برای تحلیل داده‌های بقای فضایی

کیومرث مترجم** محسن محمدزاده[1]* و آمنه آبیار*

*گروه آمار، دانشگاه تربیت مدرس، تهران

**موسسه آموزش عالی صدرالمتالهین، تهران

تاریخ دریافت: ۱۳۹۳/۷/۲۸ تاریخ پذیرش: ۱۳۹۴/۵/۱۵

چکیده: مدل خطرهای متناسب کاکس یکی از پرکاربردترین مدل‌ها برای برازاندن به داده‌های بقا است که بر اساس فرض‌های همگنی جامعه، استقلال و هم‌توزیع بودن داده‌های بقا بنا شده است. اما در بسیاری از مواقع خطرهای واحدهای آماری متفاوت بوده و فرض همگنی جامعه برقرار نیست. یکی از دلایل این تفاوت وجود عوامل خطر ناشناخته یا مشاهده نشده است که لحاظ نکردن آن‌ها و استفاده از مدل‌هایی همچون مدل خطرهای متناسب کاکس می‌تواند نتایج گمراه‌کننده‌ای را به همراه داشته باشد. در این‌گونه موارد از مدل شکنندگی که اثر عوامل ناشناخته را در نظر می‌گیرد استفاده می‌شود. در این مقاله عملکرد مدل‌های شکنندگی و خطرهای متناسب کاکس در برازش داده‌های بقا باوجود عوامل خطر ناشناخته بررسی و کارآیی این دو مدل هنگامی‌که منبع اثر عوامل خطر ناشناخته همبستگی فضایی داده‌های بقا باشد، مورد بررسی قرار می‌گیرد.

واژه‌های کلیدی: داده‌های بقا، مدل خطرهای متناسب کاکس، مدل شکنندگی، مدل بقای فضایی.

رده‌بندی ریاضی (۲۰۱۰): ۶۲N۰۲، ۶۲H۱۱.

۱- مقدمه

یکی از مسائل موردتوجه محققان در تحلیل داده‌های بقا وجود عوامل خطر ناشناخته است. گاهی به دلیل شرایط خاص مطالعه یا وجود برخی محدودیت‌های اقتصادی امکان اندازه‌گیری برخی از عوامل خطر در قالب متغیرهای تبیینی وجود ندارد. همچنین در نظر نگرفتن برخی عوامل تأثیرگذار بر زمان بقا یا غیرقابل مشاهده بودن این عوامل موجب وجود عوامل خطر

۱- آدرس الکترونیکی نویسنده مسئول مقاله: mohsen_m@modares.ac.ir

ناشناخته در مدل بندی داده‌های بقا می‌شود. در نظر نگرفتن خطر ناشناخته سبب می‌شود قسمت عمده‌ای از تغییرات تابع خطر که می‌تواند باوجود عوامل ناشناخته در مدل تبیین شود، با جمله خطای مدل جمع شده، باعث افزایش تغییرات تابع خطر، تشخیص نادرست تابع بقا و برآورد نامناسب ضرایب رگرسیونی مدل شود. مدل خطرهای متناسب کاکس یکی از پرکاربردترین مدل‌ها در برازش به داده‌های بقا است اما در نظر گرفتن عوامل ناشناخته مؤثر بر داده‌ها در مدل کاکس میسر نیست. ازاین‌رو از مدل‌های شکنندگی برای لحاظ کردن اثر عوامل مؤثر ناشناخته استفاده می‌شود.

واپل و همکاران [۱] برای اولین بار عبارت شکنندگی را برای مدل‌های بقای تک متغیره مورداستفاده قراردادند. سپس کلایتون و کازیک [۲] مدل‌های شکنندگی را به‌کار بردند. لنکستر [۳] زمان بیکاری را با استفاده از اثر تصادفی موردبررسی قرار داد و مدل خطرات متناسب آمیخته را در مطالعات اقتصادی متداول نمود. مدل‌های شکنندگی در مطالعات طول عمر توسط زلترمن [۴] مورداستفاده قرار گرفت. ترنیو و گرامبش [۵] یک رهیافت درستنمایی جزئی تاوانیده[1] را برای این مدل‌ها ارائه کردند. گوتیرز [۶] مدل‌های شکنندگی[2] و شکنندگی مشترک[3] را موردمطالعه و ویژگی‌های آن‌ها را مقایسه کرد. خیری و همکاران [۷] از مدل شکنندگی همبسته برای تحلیل عوامل خطر در پیوند قرنیه استفاده کردند. وینک [۸] مدل-های شکنندگی همبسته را با استفاده از توابع مفصل شکنندگی معرفی و مورداستفاده قرار داد. در واقع مدل‌های شکنندگی نوع خاصی از مدل‌های اثرات تصادفی برای مدل‌بندی داده‌های بقا هستند. اما هنگامی‌که بین مشاهدات همبستگی فضایی وجود داشته باشد یعنی دادهٔ بقا تابعی از موقعیت فضایی مشاهده آن‌ها باشد مدل‌های شکنندگی کلاسیک برای تحلیل داده‌های بقا مناسب نیست. در این حالت می‌توان اثر تصادفی غیرقابل مشاهده را وارد مدل شکنندگی کرد. آسم کودکان در شهر تهران و ابتلا به بیماری سرطان پوست در ارتفاعات مختلف، که در آن‌ها تابش اشعه خورشید متفاوت است، مثال‌هایی از داده‌های بقای فضایی هستند. لی و ریان [۹] اثر فضایی غیرقابل مشاهده را به‌صورت یک میدان تصادفی وارد مدل شکنندگی کرده و شناسایی پذیری مدل بقای فضایی ارائه‌شده را نشان دادند. یکی از ویژگی‌های مدل شکنندگی فضایی امکان پیشگویی زمان‌های بقا در موقعیت‌هایی است که در آن‌ها مشاهده‌ای نداریم. دیوا و همکاران [۱۰] همبستگی فضایی میان داده‌های بقا را در مدل‌های بقای بیز سلسله مراتبی لحاظ کردند. هانسون و همکاران [۱۱] یک مدل بیز سلسله مراتبی را برای لحاظ کردن اثر همبستگی فضایی-زمانی در داده‌های سرطان سینه در ایالت آیوا موردمطالعه قراردادند. جیانگ

1- Penalized partial likelihood
2- Frailty model
3- Shared frailty model

و همکاران [۱۲] روشی برای برازش مدل بقای همبستهٔ فضایی به مجموعه داده‌های هم‌گروه را ارائه کردند.

هدف این مقاله مقایسه عملکرد مدل‌های کاکس و شکنندگی در برازش به داده‌های بقا با عوامل ناشناخته ناشی از نادیده گرفتن متغیر تبیینی و اثر تصادفی فضایی با متغیرهای تبیینی مختلف است. در بخش ۲ مدل‌های خطرهای متناسب کاکس و شکنندگی تشریح می‌شوند. در بخش ۳ به چگونگی برآورد پارامترهای مدل‌ها پرداخته می‌شود. در بخش ۴ با در نظر گرفتن دو مطالعهٔ فرضی، داده‌های بقا و بقای فضایی شبیه‌سازی شده و نتایج برازش مدل‌های خطرهای متناسب و شکنندگی به داده‌ها ارائه می‌گردد. نهایتاً بحث و نتیجه‌گیری در بخش ۵ ارائه می‌شود.

۲- مدل‌های خطرهای متناسب کاکس و شکنندگی

مدل خطرهای متناسب کاکس یک مدل رگرسیونی است، که در آن زمان رخداد حادثه به‌عنوان متغیر وابسته در نظر گرفته می‌شود و متغیرهای کمکی که بر زمان رخداد حادثه مؤثرند به‌عنوان متغیرهای تبیینی در مدل منظور می‌شوند. یکی از پیش‌فرض‌های اساسی در این مدل متناسب بودن خطر برای تمامی واحدهای آماری است. البته این شرط ممکن است در پاره‌ای از مواقع تأمین نشود، اما به‌واسطهٔ سادگی یکی از مدل‌های پرکاربرد نزد کاربران حوزه تحقیقات پزشکی است. وقتی داده‌های بقا از توزیعی معلوم با تکیه‌گاه مثبت مانند وایبول پیروی کنند، تحلیل آن‌ها توسط مدل‌های پارامتری به‌صورت

$$f(t)=h(t)S(t) \qquad (۱)$$

انجام می‌شود، که در آن $f(\cdot)$ تابع چگالی زمان‌های بقا، $S(t)=P(T^*\geq t)$ تابع بقا و

$$h(t)=\lim_{\Delta t\to 0}\frac{P\left(t\leq T^*<t+\Delta t|T^*\geq t\right)}{\Delta t}$$

تابع خطر است، به‌طوری‌که T^* متغیر زمان بقا است. تابع خطر سرعت خطر رخداد اتفاق موردنظر در زمان مشخص t را نشان می‌دهد و تابع بقا احتمال آنکه زمان بقای واحد آزمایشی موردنظر بیشتر از t واحد زمانی باشد را محاسبه می‌کند. همچنین تابع خطر تجمعی به‌صورت

$$H(t)=\int_0^t h(u)\,du$$

تعریف می‌شود، که میزان خطر آنی وقوع رخداد موردنظر را از ابتدای مطالعه تا زمان t

محاسبه می‌کند. یکی از مدل‌های متداول برای تحلیل داده‌های بقا، مدل خطرهای متناسب معروف به مدل کاکس (کاکس، [۱۳]) است. فرض کنید $h(t|X)$ تابع خطر در زمان t با بردار متغیرهای تبیینی $X=(X_1,\ldots,X_K)$ باشد. در این صورت مدل خطرهای متناسب کاکس به‌صورت

$$h(t|X)=h_0(t)e^{\beta' X}$$

تعریف می‌شود، که در آن $X=(X_1,\ldots,X_K)$ بردار متغیرهای تبیینی، $\beta'=(\beta_1,\ldots,\beta_k)$ بردار اثرهای ثابت و $h_0(\cdot)$ تابع خطر پایه است [۱۳]. در این مدل فرض می‌شود تمامی افراد جامعه دارای تابع خطر پایه یکسان هستند. یکی از ویژگی‌های مدل کاکس این است که بدون در نظر گرفتن هیچ‌گونه فرض توزیعی در ارتباط با تابع خطر پایه می‌توان آن را به داده‌های بقا برازش داد، هرچند در مدل پارامتری کاکس می‌توان برای تابع خطر پایه فرض توزیعی نیز در نظر گرفت. بنابراین در استفاده از مدل کاکس لازم است تابع خطر پای $h_0(\cdot)$ با روش نا-پارامتری یا پارامتری برآورد شود. بر همین اساس مدل خطرهای متناسب کاکس به دو صورت پارامتری و نیمه‌پارامتری استفاده می‌شود. در حالت پارامتری می‌توان توزیع‌های بقا مانند نمایی، وایبول و گومپرتز را به‌کار برد (یاشین، [۱۴]). در حالت پارامتری، مناسب بودن برازش تابع بقا در مدل کاکس به داده‌ها از اهمیت بسزایی برخوردار است.

علاوه بر این، لحاظ کردن اثر عوامل ناشناخته مؤثر بر داده‌های بقا در مدل نیز می‌تواند موجب بهبود کارایی مدل شود. اما منظور کردن عوامل ناشناخته مؤثر بر داده‌های بقا در مدل کاکس میسر نیست. ازاین‌رو از مدل‌های شکنندگی برای لحاظ کردن اثر عوامل ناشناخته استفاده می-شود. همگن بودن زمان‌های بقا یکی از شرط‌های مهم استفاده از مدل خطرهای متناسب کاکس است. اما وجود عوامل خطر ناشناخته موجب ناهمگنی زمان‌های بقا می‌شود که برای کنترل آن می‌توان یک اثر تصادفی غیرقابل اندازه‌گیری را به‌صورت ضربی به تابع خطر افزود.

اگر X بردار متغیرهای تبیینی و U بردار متغیرهای ناشناخته باشد، که هردوی آن‌ها در تابع خطر مؤثرند، با فرض استقلال X و U مدل خطرهای متناسب کاکس در زمان t برای آزمودنی i به‌صورت $h_0(t)\exp(\beta' x_i+\psi' u_i)$ برای $i=1,\ldots,n$ خواهد بود، که در آن ψ بردار اثرات متغیرهای شکنندگی است. چون U ناشناخته و غیرقابل اندازه‌گیری است عبارت $\exp(\psi' u_i)$ تصادفی در نظر گرفته می‌شود و به‌عنوان اثر شکنندگی z_i در مدل خطر کاکس ضرب می‌شود. بنابراین مدل خطر در زمان t و به‌شرط معلوم بودن z_i و بردار متغیرهای تبیینی برای آزمودنی i به‌صورت

$$h(t|z_i,h_0,X)=z_i h_0(t)\exp(\beta' x_i),\qquad i=1,\ldots,n \qquad (۲)$$

خواهد بود، که مدل کاکس اصلاح‌شده یا مدل شکنندگی بقا نامیده می‌شود. اگرچه در تعریف کلی مدل شکنندگی تابع خطر غیرشرطی به هر شکل قابل مدل‌بندی است، اما به‌دلیل استفاده گسترده از مدل خطرهای متناسب کاکس، مدل شکنندگی بر مبنای مدل خطر کاکس به‌صورت $h_0(t)e^{\beta' X}$ متداول است.

در مدل شکنندگی با فرض استقلال اثر شکنندگی از متغیرهای تبیینی، تغییرات کل تابع خطر و درنتیجه تغییرات تابع بقا به اثر متغیرهای معلوم، اثر عوامل ناشناخته و اثر تصادفی مدل تجزیه می‌شود. مدل شکنندگی با تخصیص اثر تصادفی، وجود تفاوت در آزمودنی‌ها را بر روی برآورد اثر متغیرهای درون مدل لحاظ می‌کند. با توجه به مثبت بودن تابع خطر، متغیر شکنندگی دارای توزیعی با تکیه‌گاه مثبت در نظر گرفته می‌شود. انتخاب توزیع مناسب برای متغیر شکنندگی از اهمیت زیادی برخوردار است به‌ویژه به‌واسطه ارتباط تابع بقا با تبدیل لاپلاس تابع چگالی متغیر شکنندگی توزیع‌هایی که فرم لاپلاس بسته‌ای دارند مانند توزیع گاما، لگ نرمال، گاوسی وارون، پواسون مرکب و توزیع‌های خانواده واریانس توانی با میانگین متناهی مورداستفاده قرار می‌گیرند.

معمولاً برای ساده‌سازی محاسبات میانگین متغیر شکنندگی برابر یک در نظر گرفته می‌شود. در مقاله حاضر توزیع متغیر شکنندگی لگ‌نرمال در نظر گرفته‌شده است. در مدل شکنندگی اگر برآورد z_i بزرگ‌تر از z_j باشد بدان معنی است که آزمودنی i شکننده‌تر از آزمودنی j بوده و خطر شکست برای آزمودنی i بیشتر از آزمودنی j است.

۳-برآورد پارامترهای مدل‌ها

برای برآورد پارامترهای مدل خطرهای متناسب کاکس با روش ماکسیمم درستنمایی باید تابع درستنمایی این مدل محاسبه شود. بر اساس مشاهدات (t_i, δ_i, X_i)، $i=1,\dots,n$ و تابع خطر پایه $h_0(\cdot)$ در مدل خطرهای متناسب کاکس داریم:

$$S(t)=\exp\left(-H(t)\right). \qquad (۳)$$

بنابراین با استفاده از روابط (۱) و (۳) تابع چگالی زمان‌های بقا به‌صورت

$$f(t)=h(t)\exp\left(-H(t)\right)=h_0(t)e^{\beta' X}\exp\left(-H_0(t)e^{\beta' X}\right)$$

حاصل می‌شود. بنابراین تابع درستنمایی مدل خطرهای متناسب به‌صورت

$$L(\beta)=\Pi_{i=1}^{n}\left(h_0(t_i)e^{\beta' X_i}\right)^{\delta_i}\exp\left(-H_0(t)e^{\beta' X_i}\right) \qquad (۴)$$

است، که در آن

$$\delta_i = \begin{cases} ۱ & T_i^* \le C_i \\ ۰ & T_i^* > C_i \end{cases}$$

متغیر تصادفی سانسور و C_i زمان سانسور واحد iام است. درصورت معلوم بودن شکل تابع خطر پایه با ماکسیمم کردن تابع درستنمایی (۴) نسبت به پارامترهای رگرسیونی، برآورد ضرایب β به‌راحتی حاصل می‌شود.

در حالت نیمه‌پارامتری فرض توزیعی در ارتباط با $h_0(\cdot)$ در نظر گرفته نمی‌شود و تابع خطر پایه به روش ناپارامتری برآورد می‌شود. در این حالت چون تابع درستنمایی شامل یک تابع خطر ناشناخته است، برای برآورد ضرایب رگرسیونی β استفاده از تابع درستنمایی معمول امکان‌پذیر نیست. بنابراین از تابع درستنمایی اصلاح‌شده که حاوی اطلاعات کافی در خصوص پارامترها است و هم‌زمان به تابع خطر پایه وابسته نیست استفاده می‌شود. کاکس [۱۳] با معرفی تابع درستنمایی جزئی[1]، پارامترهای مدل را برآورد کرد. روش او مشکلات فنی برآورد پارامترها با در نظر گرفتن تابع خطر پایه به‌عنوان یک پارامتر مزاحم با بعد متناهی را حل کرد. در این روش تابع خطر پایه، یک تابع گسسته در نظر گرفته می‌شود که به‌جز زمان‌های t_i که اتفاق موردنظر رخ‌داده باشد صفر است و تابع خطر پایه تجمعی به‌صورت

$$H_0(t) = \sum_{t \le t_j} h_0(t_j) \tag{۵}$$

حاصل می‌شود و با جایگذاری (۵) در (۴) و انجام محاسبات، تابع درستنمایی جزئی به‌صورت

$$\Pi_{i=1}^{n} \left[h_0(t_i) e^{\beta' X_i} \exp\left(-h_0(t_i) \sum_{j \in R(t_i)} e^{\beta' X_j} \right) \right]^{\delta_i} \tag{۶}$$

به‌دست می‌آید. از مشتق جزئی (۶) نسبت به تابع خطر پایه و برابر صفر قرار دادن آن داریم

$$\hat{h}_0(t) = \frac{\delta_i}{\sum_{j \in R(t_i)} e^{\beta' X_j}} \tag{۷}$$

با جایگذاری (۷) در (۶)، تابع درستنمایی جزئی خطرهای متناسب کاکس به‌صورت

1- Partial likelihood function

$$L\left(\beta\right)=\Pi_{i=1}^{n}\left(\frac{e^{\beta' X_i}}{\sum_{j\in R\left(t_i\right)} e^{\beta' X_j}}\right)^{\delta_i}$$

حاصل می‌شود، که از مشتق اول آن برای برآورد بردار پارامتر β و از مشتق دوم برای برآورد واریانس برآوردگرها استفاده می‌شود.

نکته قابل‌توجه در استفاده از مدل خطرهای متناسب کاکس آن است که روش ماکسیمم درستنمایی جزئی کاکس به‌جای مقدار مشاهدات از رتبه داده‌ها استفاده می‌کند بنابراین بدون تغییر رتبه مشاهدات حتی با تغییر مقدار مشاهدات نیز تغییری در برآورد پارامترها صورت نمی‌گیرد. بنابراین برآورد پارامترها به‌صورت یک تابع پله‌ای است. این موضوع اهمیت زمان دقیق اندازه‌گیری را مشخص می‌کند.

در مدل شکنندگی (۲) به‌طورکلی اثرات تصادفی z_i ها غیرقابل مشاهده‌اند و فرض می‌شود این اثرات از یک توزیع پیروی می‌کنند. بنابراین به‌شرط مشاهدهٔ z_i ها تابع درستنمایی به‌صورت

$$L\left(\beta \mid z\right)=\Pi_{i=1}^{n}\left(z_i h_0\left(t_i\right)e^{\beta' X_i}\right)^{\delta_i}\exp\left(-z_i H_0\left(t\right)e^{\beta' X_i}\right)$$

و تابع بقا به‌صورت

$$S\left(t|z\right)=e^{\int_0^t h\left(s|z\right)ds}=e^{-z\int_0^t h_0\left(s\right)ds}=e^{-zH_0\left(t\right)}$$

به‌دست خواهند آمد. بنابراین تابع بقا به‌صورت

$$S\left(t\right)=E\left(S\left(t|z\right)\right)=E\left(e^{zH_0\left(t\right)}\right)=LA\left(H_0\left(t\right)\right)$$

حاصل می‌شود، که در آن $LA\left(\cdot\right)$ تبدیل لاپلاس است. بنابراین داریم

$$f\left(t\right)=-h_0\left(t\right)LA'\left(H_0\left(t\right)\right)$$

$$h\left(t\right)=h_0\left(t\right)\frac{LA'\left(H_0\left(t\right)\right)}{LA\left(H_0\left(t\right)\right)}$$

$$E\left(z\right)=-LA'\left(0\right),\quad Var\left(z\right)=LA''\left(0\right)-\left(LA'\left(0\right)\right)^2$$

که در آن LA' و LA'' به‌ترتیب مشتق اول و دوم تبدیل لاپلاس هستند.

۴- مطالعه شبیه‌سازی

هدف این بخش بررسی تأثیر عوامل خطر ناشناخته بر مدل‌های خطرهای متناسب کاکس و شکنندگی است. این عوامل خطر ناشناخته می‌تواند حاصل از نادیده گرفتن یک یا چند متغیر تبیینی یا غیرقابل اندازه‌گیری بودن یک عامل مؤثر بر داده بقا باشد. در مطالعات بقا گاهی به‌علت تعدد متغیرهای تبیینی یا هزینه زیاد در نظر گرفتن تمامی متغیرهای تبیینی در مطالعه، متغیرهای تبیینی کم‌اهمیت‌تر در مطالعه درنظر گرفته نمی‌شوند. اما باید توجه داشت که تشخیص متغیرهای تبیینی و تعیین درجه تأثیرگذاری آن‌ها در مدل از اهمیت به‌سزایی برخوردار است. همچنین یکی از عواملی که در بسیاری از تحقیقات پزشکی به‌صورت مستقیم غیرقابل اندازه‌گیری یا غیرقابل مشاهده است اثر همبستگی فضایی میان داده‌های بقا است. نادیده گرفتن این همبستگی فضایی می‌تواند موجب استنباط نادرست در برآورد تابع خطر و به‌دنبال آن تابع بقا شود.

در این بخش به‌منظور بررسی تأثیر عوامل خطر ناشناخته بر مدل‌های خطرهای متناسب کاکس و شکنندگی، در یک مطالعه شبیه‌سازی مجموعه‌ای از داده‌های بقا و بقا فضایی با متغیرهای تبیینی متفاوت تولیدشده و با برازش مدل‌های کاکس و شکنندگی به داده‌ها، برآوردهای مدل رگرسیون کاکس و مدل شکنندگی مورد مقایسه قرار گرفته‌اند. در این شبیه‌سازی داده‌ها بر اساس دو مطالعه فرضی برای مدت‌زمان بهبود بیماران به‌عنوان زمان بقا، انجام‌شده است. در این دو مطالعه زمان بقا یک‌بار متأثر از موقعیت مکانی و بار دیگر بدون تأثیرگذاری موقعیت واحد آزمایشی در نظر گرفته‌شده است. بر این اساس چهار مجموعه داده به‌صورت

۱. داده‌های بقا کلاسیک با متغیرهای تبیینی گاما
۲. داده‌های بقا کلاسیک با متغیرهای تبیینی مختلف
۳. داده‌های بقای فضایی با متغیرهای تبیینی گاما
۴. داده‌های بقای فضایی با متغیرهای تبیینی مختلف

هر یک با ۲۰ و ۸۰ درصد سانسور تولید شد. در مجموعه‌های اول و سوم سه متغیر تبیینی X_1, X_2, X_3 از خانواده توزیع گاما به‌ترتیب با میانگین‌های ۷/۳۳، ۲۵ و ۸۰ و واریانس‌های ۲/۴۴، ۶/۲۵ و ۲۰ در نظر گرفته‌شده است که به‌ترتیب نشان‌دهنده میزان اسید اوریک، سن و وزن بیماران است. در مجموعه‌های دوم و چهارم به‌منظور تولید مقادیر تصادفی متغیرهای تبیینی، مقدار X_1 را که معرف سن بیمار است از توزیع گاما با میانگین ۵۰ و واریانس ۲۵، X_2 که معرف نرخ بهبود بیماری است از توزیع یکنواخت پیوسته در بازه $[0,1]$ و X_3 که معرف میزان فشارخون افراد است از توزیع نرمال با میانگین ۱۳ و واریانس ۲ تولیدشده‌اند. برای تمامی مجموعه داده‌ها بر اساس سه متغیر تبیینی تابع خطر به‌صورت

$$h(t)=h_{\circ}(t)\exp(\cdot/\Delta X_{۱}+\cdot/۶X_{۲}+\cdot/۷X_{۳})$$

در نظر گرفته شد. داده‌های بقا بر اساس متغیرهای تبیینی و از توزیع وایبول با پارامترهای $\gamma=۱$ و $\nu=۲$ تولید شد. در این مقاله داده‌های بقا از طریق تابع خطر پایه تولیدشده‌اند. بندر و همکاران [۱۵] روشی برای تولید داده‌های بقا از طریق تابع خطر پایه ارائه کردند. در این روش با معلوم بودن تابع خطر پایه در مدل خطرهای متناسب داریم

$$S(t|X)=\exp\left(-H_{\circ}(t)\exp(\beta'X)\right) \tag{۸}$$

که در آن $H_{\circ}(\cdot)$ تابع خطر پایه تجمعی است که به‌صورت

$$H_{\circ}(t)=\int_{\circ}^{t}h_{\circ}(u)\,du$$

محاسبه می‌شود. با توجه به آنکه $S(t|X)=۱-F(t\,|\,X)$ و $F(\cdot)$ دارای توزیع یکنواخت پیوسته در بازه $[\circ,۱]$ است، لذا $S(\cdot)$ نیز از این توزیع پیروی می‌کند و بر اساس (۸) داریم

$$U=\exp\left(-H_{\circ}(t)\exp(\beta'X)\right)$$

که در آن U یک متغیر تصادفی از توزیع یکنواخت پیوسته در بازه $[\circ,۱]$ است و داریم

$$T=H_{\circ}^{-۱}\left(-\log(U)\exp(-\beta'X)\right). \tag{۹}$$

لذا برای تولید داده‌های بقا از رابطه (۹) استفاده‌شده است. برای تولید داده‌های بقا از توزیع وایبول، تابع خطر پایه تجمعی $\mathrm{H}_{\circ}(\mathrm{t})=\gamma \mathrm{t}^{\nu}$ در نظر گرفته شد. بنابراین می‌توان متغیر تصادفی زمان بقا را با استفاده از یک متغیر تصادفی یکنواخت پیوسته U به‌صورت

$$\mathrm{T}=-\left[\frac{\log(\mathrm{U})}{\gamma\exp(\beta'\,\mathrm{X})}\right]^{\frac{۱}{\nu}}$$

تولید کرد. در این شبیه‌سازی مقادیر $\nu=۲$ و $\gamma=۱$ برابر یک در نظر گرفته‌شده‌اند. در این مقاله با در نظر گرفتن یک میدان تصادفی گاوسی[1] روش بندر و همکاران [۱۵]، برای تولید داده‌های بقای فضایی تعمیم داده شد. معمولاً برای مدل بندی داده‌های فضایی از میدان تصادفی به‌صورت $\left\{Z(s);s\in D\subseteq R^{d};d\geq ۱\right\}$ استفاده می‌شود (محمدزاده [۱۶]). این میدان تصادفی گاوسی نامیده می‌شود، هرگاه برای هر $m\geq ۱$ توزیع توأم

1- Gaussian random field

$\left(Z\left(s_1\right),\ldots,Z\left(s_m\right)\right)$، نرمال چندمتغیره باشد. برای تولید داده‌های بقای فضایی با استفاده از یک متغیر تصادفی پیوسته U با توزیع یکنواخت $u\left(0,1\right)$ داریم:

$$\mathrm{T}=-\left[\frac{\log\left(\mathrm{U}\right)}{\gamma\exp\left(\beta'\,\mathrm{X}+Z\left(s\right)\right)}\right]^{\frac{1}{\nu}}$$

که در آن $Z\left(\cdot\right)$ میدان تصادفی گاوسی موردنظر و s موقعیت مکانی برحسب طول و عرض موقعیت واحد آماری است. با توجه به آنکه داده‌های بقای فضایی می‌توانند به‌صورت مشبکه-ای[1] یا زمین آماری[2] مشاهده شوند، در این مقاله داده‌های بقای فضایی زمین آماری با استفاده از تابع تغییرنگار در فواصل منظم تولیدشده‌اند. برای تولید داده‌ها، n اثر فضایی روی یک مشبکه $\sqrt{n}\times\sqrt{n}$ در نظر گرفته‌شده و بر اساس ماتریس فاصله گره‌های این مشبکه، n داده از میدان تصادفی گاوسی با تابع تغییرنگار[3] گاوسی به‌صورت $\mathrm{C}\left(d\right)=\sigma^2\exp\left(-\frac{\left\|\mathrm{d}\right\|^2}{\mathrm{a}^2}\right)$ تولیدشده است، که در آن $\left\|\mathrm{d}\right\|$ فاصله اقلیدسی بین نقاط است و مقادیر $\sigma^2=5$ و $\mathrm{a}=2$ به ترتیب برای واریانس و دامنه میدان تصادفی در نظر گرفته‌شده‌اند. به همین ترتیب ۴۰۰ داده بقا و بقای فضایی با ۲۰ و ۸۰ درصد سانسور و ۵۰۰ و ۱۰۰۰ بار تکرار تولیدشده‌اند. برای مقایسه عملکرد دو مدل از معیارهای میانگین توان دوم خطا[4] (MSE) و درصد قدرمطلق اریبی[5] (MAPB) برآوردها استفاده‌شده است. در این مطالعه داده‌های بقا و بقای فضایی بر اساس سه متغیر تبیینی در نظر گرفته‌شده تولیدشده‌اند اما یک‌بار با هر سه متغیر تبیینی و بار دیگر دو متغیر تبیینی کم‌اهمیت‌تر در نظر گرفته‌شده است. در این مطالعه در هر مجموعه داده متغیر تبیینی تأثیرگذارتر بر داده‌ها (وزن و فشارخون) حذف‌شده است. نتایج برازش مدل‌های کاکس و شکنندگی به چهار مجموعه داده شبیه‌سازی شده با در نظر گرفتن سه و دو متغیر تبیینی در جداول ۱ تا ۴ ارائه‌شده‌اند. همان‌طور که در جدول ۱ ملاحظه می‌شود با در نظر گرفتن سه متغیر تبیینی، مدل خطرهای متناسب و شکنندگی تقریباً عملکرد یکسانی داشته‌اند اما با حذف یک متغیر تبیینی مدل شکنندگی عملکرد بهتری نسبت به مدل خطرهای متناسب داشته است. بر اساس نتایج با حذف یک متغیر تبیینی، برآورد ضرایب متغیرهای تبیینی باقی-

1- Lattice
2- Geostatistical
3- Covariogram
4- Mean Square Error
5- Mean Absolute Percentage Bias

مانده در مدل (کاکس و شکنندگی) نادقیق‌تر از حالتی است که تمامی متغیرهای تبیینی در نظر گرفته شود. حذف یک متغیر به معنای وجود عوامل خطر ناشناخته در مدل است. جدول ۱ نشان می‌دهد برآورد ضرایب در این حالت در مدل شکنندگی به خاطر وجود متغیر شکنندگی که اثر عوامل خطر ناشناخته را در مدل تبیین می‌کند به‌مراتب دقیق‌تر از مدل خطرهای متناسب است.

جدول ۱: برآورد پارامترها با مدل خطرهای متناسب و شکنندگی برای داده‌های مجموعه ۱

تکرار	سانسور	پارامتر	کاکس			شکنندگی		
			برآورد	MAPB	MSE	برآورد	MAPB	MSE
۵۰۰	۲۰	β_1	۰/۵۰۲	۰/۳۰۸	<۰/۰۰۱	۰/۵۱۶	۳/۲۷۰	<۰/۰۰۱
		β_2	۰/۶۰۳	۰/۵۳۷	<۰/۰۰۱	۰/۶۲۱	۳/۵۰۳	<۰/۰۰۱
		β_3	۰/۷۰۴	۰/۵۳۷	<۰/۰۰۱	۰/۷۲۵	۳/۵۰۳	۰/۰۰۱
		β_1	۰/۱۶۸	۶۶/۳۷۸	۰/۱۱۰	۰/۲۰۴	۵۹/۱۰۴	۰/۰۸۷
		β_2	۰/۱۷۲	۷۱/۲۶۳	۰/۱۸۳	۰/۲۰۴	۶۶/۰۵۱	۰/۱۵۷
	۸۰	β_1	۰/۵۱۴	۲/۸۹۳	<۰/۰۰۱	۰/۵۳۹	۷/۷۳۸	۰/۰۰۱
		β_2	۰/۶۱۱	۱/۸۹۰	<۰/۰۰۱	۰/۶۴۰	۶/۵۸۴	۰/۰۰۲
		β_3	۰/۷۱۳	۱/۹۴۶	<۰/۰۰۱	۰/۷۴۶	۶/۶۵۰	۰/۰۰۲
		β_1	۰/۲۹۲	۴۱/۵۱۳	۰/۰۴۳	۰/۳۰۰	۳۹/۹۶۹	۰/۳۹۹
		β_2	۰/۲۱۰	۶۴/۹۸۴	۰/۱۵۲	۰/۲۱۶	۶۴/۰۵۶	۰/۱۴۷
۱۰۰۰	۲۰	β_1	۰/۵۰۱	۰/۲۳۸	<۰/۰۰۱	۰/۵۱۵	۳/۰۶۴	<۰/۰۰۱
		β_2	۰/۶۰۳	۰/۴۳۱	<۰/۰۰۱	۰/۶۲۰	۳/۳۰۵	<۰/۰۰۱
		β_3	۰/۷۰۴	۰/۵۴۰	<۰/۰۰۱	۰/۷۲۴	۳/۳۸۴	۰/۰۰۱
		β_1	۰/۱۸۸	۶۲/۳۶۶	۰/۰۹۷	۰/۲۵۹	۴۸/۱۱۷	۰/۰۵۸
		β_2	۰/۱۸۶	۶۹/۰۴۵	۰/۱۷۲	۰/۲۶۳	۵۶/۱۷۶	۰/۱۱۴
	۸۰	β_1	۰/۵۱۰	۱/۹۸۰	<۰/۰۰۱	۰/۵۳۳	۶/۶۶۱	۰/۰۰۱
		β_2	۰/۶۰۹	۱/۴۶۶	<۰/۰۰۱	۰/۶۳۵	۵/۸۷۰	۰/۰۰۱
		β_3	۰/۷۱۱	۱/۵۰۴	<۰/۰۰۱	۰/۷۴۳	۶/۰۹۳	۰/۰۰۲
		β_1	۰/۳۰۶	۳۸/۸۸۵	۰/۰۳۸	۰/۳۳۱	۳۳/۸۹۴	۰/۰۲۹
		β_2	۰/۲۳۹	۶۰/۲۱۶	۰/۱۳۱	۰/۲۶۷	۵۵/۵۱۹	۰/۱۱۱

جدول ۲: برآورد پارامترها با مدل خطرهای متناسب و شکنندگی برای داده‌های مجموعه ۲

تکرار	سانسور	پارامتر	کاکس			شکنندگی		
			برآورد	MAPB	MSE	برآورد	MAPB	MSE
۵۰۰	۲۰	β_1	۰/۵۰۶	۱/۲۵۱	<۰/۰۰۱	۰/۵۲۷	۵/۴۱۴	۰/۰۰۱
		β_2	۰/۶۰۲	۰/۳۴۳	<۰/۰۰۱	۰/۶۲۷	۴/۵۰۲	۰/۰۰۱
		β_3	۰/۷۰۶	۰/۸۸۹	<۰/۰۰۱	۰/۷۳۵	۵/۰۷۰	۰/۰۰۱
		β_1	۰/۳۳۴	۳۳/۲۳۶	۰/۰۲۸	۰/۴۰۱	۱۹/۷۷۴	۰/۰۱۰
		β_2	۰/۳۲۳	۴۶/۰۹۷	۰/۰۷۶	۰/۳۸۷	۳۵/۵۴۸	۰/۰۴۵
	۸۰	β_1	۰/۵۰۶	۱/۲۷۹	<۰/۰۰۱	۰/۵۲۶	۵/۱۵۶	۰/۰۰۱
		β_2	۰/۵۹۶	۰/۶۴۴	<۰/۰۰۱	۰/۶۱۹	۳/۱۵۳	<۰/۰۰۱
		β_3	۰/۷۰۹	۱/۳۵۴	<۰/۰۰۱	۰/۷۳۷	۵/۲۳۶	۰/۰۰۱
		β_1	۰/۳۹۲	۲۱/۶۳۷	۰/۰۱۲	۰/۴۱۱	۱۷/۸۱۵	۰/۰۰۸
		β_2	۰/۴۰۶	۳۲/۳۰۲	۰/۰۳۸	۰/۴۲۲	۲۹/۶۵۰	۰/۰۳۲
۱۰۰۰	۲۰	β_1	۰/۵۰۳	۰/۶۰۴	<۰/۰۰۱	۰/۵۲۲	۴/۴۵۹	<۰/۰۰۱
		β_2	۰/۶۰۵	۰/۸۶۰	<۰/۰۰۱	۰/۶۲۸	۴/۶۷۲	۰/۰۰۱
		β_3	۰/۷۰۵	۰/۶۶۲	<۰/۰۰۱	۰/۷۳۲	۴/۵۳۰	۰/۰۰۱
		β_1	۰/۳۳۱	۳۳/۷۵۱	۰/۰۲۸	۰/۴۲۷	۱۴/۶۷۱	۰/۰۰۵
		β_2	۰/۴۲۷	۲۸/۸۴۲	۰/۰۳۰	۰/۵۴۰	۹/۹۴۰	۰/۰۰۴
	۸۰	β_1	۰/۵۰۴	۰/۸۳۱	<۰/۰۰۱	۰/۵۲۵	۴/۹۶۳	۰/۰۰۱
		β_2	۰/۶۱۲	۲/۰۵۰	<۰/۰۰۱	۰/۶۳۶	۶/۰۲۷	۰/۰۰۱
		β_3	۰/۷۰۶	۰/۸۹۸	<۰/۰۰۱	۰/۷۳۲	۴/۵۶۲	۰/۰۰۱
		β_1	۰/۴۱۰	۱۷/۹۷۸	۰/۰۰۸	۰/۴۳۹	۱۲/۲۰۰	۰/۰۰۴
		β_2	۰/۵۹۲	۱/۴۰۸	۰/۰۰۱	۰/۶۲۷	۴/۴۲۴	۰/۰۰۱

در این مطالعه متغیرهای تبیینی دو مجموعه اول و سوم از یک خانواده توزیع و در مجموعه‌های دوم و چهارم از سه توزیع مختلف در نظر گرفته‌شده‌اند که در مجموعه‌های اول و سوم متغیر تبیینی وزن و در مجموعه‌های دیگر متغیر تبیینی فشارخون با داشتن بیشترین واریانس تأثیرگذارترین متغیر تبیینی بر داده‌های بقا و بقا فضایی هستند. همان‌طور که در جداول ۱ و ۲ ملاحظه می‌کنید با در نظر نگرفتن مؤثرترین متغیر تبیینی در هر مجموعه میانگین توان دوم

خطا و درصد قدرمطلق اریبی افزایش‌یافته است اما این افزایش برای مجموعه اول فراوان‌تر است و علت آن پراکندگی بیشتر متغیر تبیینی وزن در مجموعه اول و درنتیجه تأثیرگذاری بیشتر این متغیر تبیینی بر داده‌های بقای مجموعه اول است.

جدول ۳: برآورد پارامترها با مدل خطرهای متناسب و شکنندگی برای داده‌های مجموعه ۳

			کاکس			شکنندگی		
تکرار	سانسور	پارامتر	برآورد	MAPB	MSE	برآورد	MAPB	MSE
۵۰۰	۲۰	β_1	۰/۲۱۸	۵۶/۴۸۸	۰/۰۸۰	۰/۳۳۳	۳۳/۳۴۷	۰/۰۲۸
		β_2	۰/۲۵۷	۵۷/۱۵۳	۰/۱۱۸	۰/۳۹۵	۳۴/۱۵۷	۰/۰۴۲
		β_3	۰/۲۹۸	۵۷/۳۵۸	۰/۱۶۱	۰/۴۶۰	۳۴/۲۷۶	۰/۰۵۸
		β_1	۰/۱۴۶	۷۰/۷۹۲	۰/۱۲۵	۰/۱۸۱	۶۳/۸۸۵	۰/۱۰۲
		β_2	۰/۱۴۱	۷۶/۵۸۳	۰/۲۱۱	۰/۱۷۱	۷۱/۵۷۰	۰/۱۸۴
	۸۰	β_1	۰/۲۶۵	۴۷/۰۱۱	۰/۰۵۵	۰/۳۲۳	۳۵/۴۴۲	۰/۰۳۱
		β_2	۰/۳۲۲	۴۶/۳۵۵	۰/۰۷۷	۰/۳۹۳	۳۴/۴۴۹	۰/۰۴۳
		β_3	۰/۳۷۰	۴۷/۱۲۲	۰/۱۰۹	۰/۴۵۵	۳۵/۰۳۳	۰/۰۶۰
		β_1	۰/۲۲۹	۵۴/۱۵۲	۰/۰۷۳	۰/۲۳۸	۵۲/۴۸۲	۰/۰۶۹
		β_2	۰/۱۸۳	۶۹/۵۴۴	۰/۱۷۴	۰/۱۹۰	۶۸/۲۸۰	۰/۱۶۸
۱۰۰۰	۲۰	β_1	۰/۲۱۳	۵۷/۴۱۱	۰/۰۸۲	۰/۳۲۹	۳۴/۲۲۰	۰/۰۲۹
		β_2	۰/۲۵۱	۵۸/۱۲۲	۰/۱۲۲	۰/۳۹۰	۳۴/۹۴۴	۰/۰۴۴
		β_3	۰/۲۹۷	۵۷/۵۶۴	۰/۱۶۲	۰/۴۵۹	۳۴/۳۷۱	۰/۰۵۸
		β_1	۰/۱۵۶	۶۸/۸۳۸	۰/۱۱۸	۰/۲۰۸	۵۸/۳۸۸	۰/۰۸۵
		β_2	۰/۱۵۴	۷۴/۲۸۷	۰/۱۹۹	۰/۲۱۰	۶۴/۹۸۱	۰/۱۵۲
	۸۰	β_1	۰/۲۶۳	۴۷/۳۲۴	۰/۰۵۶	۰/۳۱۹	۳۶/۲۵۰	۰/۰۳۳
		β_2	۰/۳۲۵	۴۵/۷۶۷	۰/۰۷۵	۰/۳۸۸	۳۵/۳۷۷	۰/۰۴۵
		β_3	۰/۳۷۳	۴۶/۷۱۸	۰/۱۰۷	۰/۴۵۳	۳۵/۳۳۳	۰/۰۶۱
		β_1	۰/۲۵۱	۴۹/۸۴۳	۰/۰۶۲	۰/۲۶۷	۴۶/۶۸۰	۰/۰۵۴
		β_2	۰/۲۰۷	۶۵/۵۶۹	۰/۱۵۵	۰/۲۲۴	۶۲/۶۱۵	۰/۱۴۱

به‌طورکلی با افزایش درصد سانسور، اریبی و میانگین توان دوم خطا بیشتر می‌شود. همان‌طور که در جداول ۳ و ۴ ملاحظه می‌شود چون در مدل‌های کاکس و شکنندگی تنها اثر فضایی

داده‌ها تبیین نشده است، میانگین توان دوم خطا و درصد قدرمطلق اریبی برآوردها کاهش قابل‌ملاحظه‌ای نسبت به جداول ۱ و ۲ نشان می‌دهد. به‌علاوه صرف‌نظر از اینکه متغیرهای تبیینی دارای توزیع یکسان یا متفاوتی باشند باوجود تمامی متغیرهای تبیینی در مدل نیز مقادیر برآوردها و تغییرات آن‌ها به‌شدت تحت تأثیر اثر فضایی داده‌ها قرار گرفته و عملاً نتایج به‌دست‌آمده توسط مدل‌های کاکس و شکنندگی بسیار دور از واقعیت است. لازم به ذکر است که حذف یک متغیر تبیینی در این حالت هم تأثیر نامطلوبی بر برآورد پارامترها داشته است.

جدول ۴: برآورد پارامترها با مدل خطرهای متناسب و شکنندگی برای داده‌های مجموعه ۴

			کاکس			شکنندگی		
تکرار	سانسور	پارامتر	برآورد	MAPB	MSE	برآورد	MAPB	MSE
۵۰۰	۲۰	β_1	۰/۲۰۹	۵۸/۲۰۰	۰/۰۸۵	۰/۲۸۷	۴۲/۶۲۳	۰/۰۴۵
		β_2	۰/۲۴۴	۵۹/۲۷۲	۰/۱۲۶	۰/۳۳۸	۴۳/۶۷۰	۰/۰۶۹
		β_3	۰/۲۸۹	۵۸/۷۷۰	۰/۱۶۹	۰/۳۹۷	۴۳/۳۲۵	۰/۰۹۲
		β_1	۰/۱۸۲	۶۳/۵۳۶	۰/۱۰۱	۰/۲۲۵	۵۵/۰۰۵	۰/۰۷۶
		β_2	۰/۱۷۳	۷۱/۲۲۱	۰/۱۸۳	۰/۲۱۲	۶۴/۵۹۲	۰/۱۵۰
	۸۰	β_1	۰/۲۹۷	۴۰/۵۶۸	۰/۰۴۱	۰/۳۲۰	۳۶/۰۷۶	۰/۰۳۳
		β_2	۰/۳۶۶	۳۹/۰۲۶	۰/۰۵۵	۰/۳۹۱	۳۴/۸۷۶	۰/۰۴۴
		β_3	۰/۴۱۴	۴۰/۸۹۲	۰/۰۸۲	۰/۴۴۵	۳۶/۴۵۴	۰/۰۶۵
		β_1	۰/۲۶۱	۴۷/۷۹۸	۰/۰۵۷	۰/۲۷۳	۴۵/۳۶۰	۰/۰۵۱
		β_2	۰/۲۷۲	۵۴/۷۲۷	۰/۱۰۸	۰/۲۸۳	۵۲/۸۸۴	۰/۱۰۱
۱۰۰۰	۲۰	β_1	۰/۲۰۸	۵۸/۴۲۵	۰/۰۸۵	۰/۲۸۵	۴۲/۹۱۹	۰/۰۴۶
		β_2	۰/۲۴۵	۵۹/۱۰۷	۰/۱۲۶	۰/۳۳۹	۴۳/۵۶۷	۰/۰۶۸
		β_3	۰/۲۹۰	۵۸/۵۷۷	۰/۱۶۸	۰/۳۹۷	۴۳/۳۰۷	۰/۰۹۲
		β_1	۰/۱۸۶	۶۲/۸۴۸	۰/۰۹۹	۰/۲۳۶	۵۲/۸۷۸	۰/۰۷۰
		β_2	۰/۲۳۳	۶۱/۱۴۴	۰/۱۳۵	۰/۲۹۴	۵۰/۹۴۲	۰/۰۹۳
	۸۰	β_1	۰/۲۸۷	۴۲/۵۲۲	۰/۰۴۵	۰/۳۱۱	۳۷/۷۵۴	۰/۰۳۶
		β_2	۰/۳۴۱	۴۳/۱۲۸	۰/۰۶۷	۰/۳۷۱	۳۸/۰۸۵	۰/۰۵۲
		β_3	۰/۴۱۵	۴۰/۷۱۷	۰/۰۸۱	۰/۴۴۶	۳۶/۲۷۱	۰/۰۶۴
		β_1	۰/۲۶۳	۴۷/۴۵۸	۰/۰۵۶	۰/۲۷۸	۴۴/۳۰۶	۰/۰۴۹

β_2	۰/۳۶۳	۳۹/۵۵۸	۰/۰۵۶	۰/۳۸۱	۳۶/۴۸۸	۰/۰۴۸

۵- بحث و نتیجه‌گیری

در این مقاله عملکرد مدل‌های کاکس و شکنندگی برای داده‌های بقا و بقای فضایی با درصد سانسورهای متفاوت و با در نظر گرفتن متغیرهای تبیینی از یک و چند خانواده موردبررسی قرار گرفت. نتایج بیانگر آن است که در نظر نگرفتن یک یا چند متغیر تبیینی تأثیر بسزایی در عملکرد مدل برازش داده‌شده به داده‌ها دارد. ازاین‌رو هنگام حذف یک متغیر تبیینی از مدل یا نادیده گرفتن آن باید در نظر داشت که علاوه بر ضریب ثابت متغیر تبیینی، میزان تأثیرگذاری این متغیر تبیینی بر مجموعه داده‌ها نیز باید در نظر گرفته شود و دقت بیشتری در استخراج متغیرهای تبیینی مؤثر بر داده بقا لازم است. یکی از این عوامل که ممکن است توسط محققان موردتوجه قرار نگیرد و حتی اندازه‌گیری نشود اثر فضایی در برخی مطالعات زمان بقا است. نتایج شبیه‌سازی این واقعیت را نشان می‌دهند که در نظر نگرفتن اثر فضایی در مدل می‌تواند کارایی مدل خطرهای متناسب و شکنندگی را تا میزان قابل‌ملاحظه‌ای کم کند.

تقدیر و تشکر

نویسندگان از داوران محترم مجله که نظرات ارزنده آن‌ها موجب بهبود مقاله شد و از حمایت قطب علمی داده‌های ترتیبی و فضایی دانشگاه فردوسی مشهد قدردانی می‌نمایند.

مراجع

[1] Vaupel, J.W., Manton, K.G. and Stallard, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality, *Demography*, **16**, 439-454.

[2] Clayton, D.G. and Cuzik, J. (1985). Multivariate generalization of the proportional hazards model, *Journal of the Royal Statistical Society*, Series A, **148**, 82-117.

[3] Lancaster, T. (1979). Econometric methods for the duration of unemployment, *Econometrics*, **47**, 939-956.

[4] Zelterman, D. (1992). A statistical distribution with an unbounded hazard function and its application to a theory from demography, *Biometrics*, **48**, 807-818.

[5] Therneau, T.M. and Grabsch, P.M. (2000). *Modeling survival data: extending the Cox model*, New York: Springer.

[6] Gutierresz, R.G. (2002). Parametric frailty and shared frailty survival models, *Stata Journal*, **2**, 22-44.

[7] Kheiri, S., Meshkani, M. R. and Faghihzadeh, S. (2005). A correlated frailty model for analysing risk factors in bilateral corneal graft rejection for Keratoconus: A Bayesian approach, *Statistics in Medicine*, **24**, 2681–2693.

[8] Wienke, A. (2011). *Frailty models in survival analysis*, United States of America: Chapman and Hall/CRC Biostatistics Series.

[9] Li, Y. and Ryan, L. (2002). Modeling spatial survival data using semiparametric frailty models, *Biometrics*, **58**, 287-297.

[10] Diva, U., Banerjee, S. and Dey, D.K. (2007). Modeling spatially correlated survival data for individuals with multiple cancers, *Statistical Modeling*, **7**, 191-213.

[11] Hanson, T., Jara, A. and Zhao, L. (2012), A Bayesian semi-parametric temporally stratified proportional hazards model with spatial frailties, *Bayesian Analysis*, **7**, 147-188.

[12] Jiang, H., Patrick E., Brown, H.R. and Shimakura, S. (2014). Geostatistical survival models for environmental risk assessment with large retrospective cohorts, *Journal of the Royal Statistical Society*: Series A (Statistics in Society), **177**, 679–695.

[13] Cox, D.R. (1972), Regression models and life-tables, *Journal of the Royal Statistical Society*, Series B, **34**, 187-220.

[14] Yashin, A.I., Vaupel, J.W. and Iachine I. (1995). Correlated individual frailty: an advantageous approach to survival analysis of bivariate data, *Mathematical Population Studies*, **5**, 1-15.

[15] Bender, R., Augustin, T. and Blettner, M. (2005). Generating survival times to simulate Cox proportional hazards models, *Statistics in Medicine*, **24**, 1713-1723.

[۱۶] محمدزاده، م. (۱۳۹۱). آمار فضایی و کاربردهای آن، مرکز نشر آثار علمی دانشگاه تربیت مدرس، تهران، ایران.

Frailty and Proportional Hazards Models for Analysis of Spatial Survival Data

Kiumars Motarjem[**], Mohsen Mohammadzadeh[*], Ameneh Abyar[*]

[*]Department of Statistics, Trabiat Modares University, Tehran, Iran

[**]Sadrolmotalehin Institute, Tehran, Iran

Abstract

One of the most widely used models for fitting survival data is Cox proportional hazards model that is based on homogeneity, independence and equi-distributed of survival data. However, in many cases, hazards of statistical units are different and the assumption of population homogeneity is not established. One of the reasons for such deference is the unknown or unobserved risk factors, which may lead to some misleading models if there is no concern for them or some models such as Cox proportional hazard models have to be implemented. In such cases, regarding the unknown risk factors, frailty models are used. In this paper, the performances of the Cox and frailty models for survival and spatial survival data with unknown risk factors are considered. The efficiency of these models whilst the source of unknown risk factors is the spatial correlation of survival data is also examined.

***Keywords*:** Survival data, Cox proportional hazards model, Frailty model, Spatial survival model.

Mathematics Subject Classification (2010): 62N02, 62H11.