

مدل‌های شکنندگی و خطرهای متناسب برای تحلیل داده‌های بقای فضایی

کیومرث مترجم** محسن محمدزاده^{۱*} و آمنه آبیاری*

*گروه آمار، دانشگاه تربیت مدرس، تهران

**موسسه آموزش عالی صدرالمتهین، تهران

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

چکیده: مدل خطرهای متناسب کاکس یکی از پرکاربردترین مدل‌ها برای برآوردن 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 برآوردن داده‌های بقا با وجود عوامل خطر ناشناخته بررسی و کارایی این دو مدل هنگامی که 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 مورد توجه 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

ناشناخته در مدل بندی داده‌های بقا می‌شود. در نظر نگرفتن خطر ناشناخته سبب می‌شود قسمت عمده‌ای از تغییرات تابع خطر که می‌تواند با وجود 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. از این‌رو از مدل‌های شکنندگی برای لحاظ کردن اثر عوامل مؤثر ناشناخته استفاده می‌شود.

وایل و همکاران [۱] برای اولین بار عبارت شکنندگی را برای مدل‌های بقای تک متغیره مورد استفاده قرار دادند. سپس کلایتون و کازیک [۲] مدل‌های شکنندگی را به کار بردند. لنکستر [۳] زمان بیکاری را با استفاده از اثر تصادفی مورد بررسی قرار داد و مدل خطرات متناسب آمیخته را در مطالعات اقتصادی متداول نمود. مدل‌های شکنندگی در مطالعات طول عمر توسط زلترمن [۴] مورد استفاده قرار گرفت. ترنیو و گرامبش [۵] یک رهیافت درست‌نمایی جزئی توانیده^۱ را برای این مدل‌ها ارائه کردند. گوتیرز [۶] مدل‌های شکنندگی^۲ و شکنندگی مشترک^۳ را مورد مطالعه 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] مدل‌های شکنندگی همبسته را با استفاده از توابع مفصل شکنندگی معرفی و مورد استفاده قرار داد. در واقع مدل‌های شکنندگی نوع خاصی از مدل‌های اثرات تصادفی برای مدل‌بندی داده‌های بقا هستند. اما هنگامی که بین مشاهدات همبستگی فضایی وجود داشته باشد یعنی داده 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 به صورت یک میدان تصادفی وارد مدل شکنندگی کرده و شناسایی پذیری مدل بقای فضایی ارائه شده 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 مورد مطالعه قرار دادند. جیانگ

1- Penalized partial likelihood

2- Frailty model

3- Shared frailty model

و همکاران [۱۲] روشی برای برازش مدل بقای همبسته 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 مطالعه 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 به-عنوان 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 به‌واسطه 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

$$f(t) = h(t)S(t) \quad (1)$$

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

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

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 $X = (X_1, \dots, X_K)$ باشد. در این صورت مدل خطرهای متناسب کاکس به صورت

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

تعریف می‌شود، که در آن $X = (X_1, \dots, X_K)$ بردار متغیرهای تبیینی، $\beta' = (\beta_1, \dots, \beta_k)$ 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 به کار 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. از این رو از مدل‌های شکنندگی برای لحاظ کردن اثر عوامل ناشناخته استفاده می‌شود. 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 به صورت ضربی به تابع خطر افزود.

اگر X بردار متغیرهای تبیینی و U بردار متغیرهای ناشناخته باشد، که هر دو آن‌ها در تابع خطر مؤثرند، با فرض استقلال X و U مدل خطرهای متناسب کاکس در زمان t برای آزمودنی i به صورت $h_0(t)\exp(\beta'x_i + \psi'u_i)$ 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 به عنوان اثر شکنندگی z_i در مدل خطر کاکس ضرب می‌شود. بنابراین مدل خطر در زمان t و به شرط معلوم بودن z_i و بردار متغیرهای تبیینی برای آزمودنی i به صورت

$$h(t|z_i, h_0, X) = z_i h_0(t) \exp(\beta'x_i), \quad i = 1, \dots, n \quad (2)$$

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 به صورت $h_0(t)e^{\beta'X}$ متداول است.

در مدل شکنندگی با فرض استقلال اثر شکنندگی از متغیرهای تبیینی، تغییرات کل تابع خطر و در نتیجه 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 مورد استفاده قرار می‌گیرند.

معمولاً برای ساده‌سازی محاسبات میانگین متغیر شکنندگی برابر یک در نظر گرفته می‌شود. در مقاله حاضر توزیع متغیر شکنندگی لگ‌نرمال در نظر گرفته شده 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:

$$S(t) = \exp(-H(t)). \quad (3)$$

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

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

حاصل می‌شود. بنابراین تابع درست‌نمایی مدل خطرهای متناسب به صورت

$$L(\beta) = \prod_{i=1}^n \left(h_0(t_i) e^{\beta'X_i} \right)^{\delta_i} \exp(-H_0(t_i) e^{\beta'X_i}) \quad (4)$$

است، که در آن

$$\delta_i = \begin{cases} 1 & T_i^* \leq C_i \\ 0 & T_i^* > C_i \end{cases}$$

متغیر تصادفی سانسور و C_i زمان سانسور واحد i ام است. در صورت معلوم بودن شکل تابع خطر پایه با ماکسیمم کردن تابع درستنمایی (۴) نسبت به پارامترهای رگرسیونی، برآورد ضرایب β به راحتی حاصل می شود.

در حالت نیمه پارامتری فرض توزیعی در ارتباط با $h_0(\cdot)$ در نظر گرفته نمی شود و تابع خطر پایه به روش ناپارامتری برآورد می شود. در این حالت چون تابع درستنمایی شامل یک تابع خطر ناشناخته است، برای برآورد ضرایب رگرسیونی β استفاده از تابع درستنمایی معمول امکان پذیر نیست. بنابراین از تابع درستنمایی اصلاح شده که حاوی اطلاعات کافی در خصوص پارامترها است و همزمان به تابع خطر پایه وابسته نیست استفاده می شود. کاکس [۱۳] با معرفی تابع درستنمایی جزئی^۱، پارامترهای مدل را برآورد کرد. روش او مشکلات فنی برآورد پارامترها با در نظر گرفتن تابع خطر پایه به عنوان یک پارامتر مزاحم با بعد متناهی را حل کرد. در این روش تابع خطر پایه، یک تابع گسسته در نظر گرفته می شود که به جز زمان های t_i که اتفاق مورد نظر رخ داده باشد صفر است و تابع خطر پایه تجمعی به صورت

$$H_0(t) = \sum_{t \leq t_j} h_0(t_j) \quad (5)$$

حاصل می شود و با جایگذاری (۵) در (۴) و انجام محاسبات، تابع درستنمایی جزئی به صورت

$$\prod_{i=1}^n \left[h_0(t_i) e^{\beta X_i} \exp \left(-h_0(t_i) \sum_{j \in R(t_i)} e^{\beta X_j} \right) \right]^{\delta_i} \quad (6)$$

به دست می آید. از مشتق جزئی (۶) نسبت به تابع خطر پایه و برابر صفر قرار دادن آن داریم

$$\hat{h}_0(t) = \frac{\delta_i}{\sum_{j \in R(t_i)} e^{\beta X_j}} \quad (7)$$

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

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

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

نکته قابل توجه در استفاده از مدل خطرهای متناسب کاکس آن است که روش ماکسیمم درست‌نمایی جزئی کاکس به جای 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 به صورت یک تابع پله‌ای است. این موضوع اهمیت زمان دقیق اندازه‌گیری را مشخص می‌کند.

در مدل شکنندگی (۲) به طور کلی اثرات تصادفی z_i ها غیرقابل مشاهده‌اند و فرض می‌شود این اثرات از یک توزیع پیروی می‌کنند. بنابراین به شرط مشاهده z_i ها تابع درست‌نمایی به صورت

$$L(\beta | z) = \prod_{i=1}^n \left(z_i h_0(t_i) e^{\beta X_i} \right)^{\delta_i} \exp(-z_i H_0(t) e^{\beta X_i})$$

و تابع بقا به صورت

$$S(t|z) = e^{\int_0^t h(s|z) ds} = e^{-z \int_0^t h_0(s) ds} = e^{-z H_0(t)}$$

به دست خواهند آمد. بنابراین تابع بقا به صورت

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

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

$$f(t) = -h_0(t) LA'(H_0(t))$$

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

$$E(z) = -LA'(\circ), \quad Var(z) = LA''(\circ) - (LA'(\circ))^2$$

که در آن LA' و LA'' به ترتیب 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 به علت تعدد متغیرهای تبیینی یا هزینه زیاد در نظر گرفتن تمامی متغیرهای تبیینی در مطالعه، متغیرهای تبیینی کم‌اهمیت‌تر در مطالعه در نظر 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 به صورت 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 به دنبال آن تابع بقا شود.

در این بخش به منظور بررسی تأثیر عوامل خطر ناشناخته بر مدل‌های خطرهای متناسب کاکس و شکنندگی، در یک مطالعه شبیه‌سازی مجموعه‌ای از داده‌های بقا و بقای فضایی با متغیرهای تبیینی متفاوت تولید شده و با برآزش 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، انجام شده است. در این دو مطالعه زمان بقا یک‌بار متأثر از موقعیت مکانی و بار دیگر بدون تأثیرگذاری موقعیت واحد آزمایشی در نظر گرفته شده است. بر این اساس چهار مجموعه داده به صورت

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 به ترتیب با میانگین‌های $۷/۳۳, ۲۵$ و ۸۰ و واریانس‌های $۲/۴۴, ۶/۲۵$ و ۲۰ در نظر گرفته شده است که به ترتیب نشان‌دهنده میزان اسید اوریک، سن و وزن بیماران است. در مجموعه‌های دوم و چهارم به منظور 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 $[۰, ۱]$ و X_3 که معرف میزان فشارخون افراد است از توزیع نرمال با میانگین ۱۳ و واریانس ۲ تولید شده‌اند. برای تمامی مجموعه داده‌ها بر اساس سه متغیر تبیینی تابع خطر به صورت

$$h(t) = h_0(t) \exp(\cdot / \delta X_1 + \cdot / \epsilon X_2 + \cdot / \gamma X_3)$$

در نظر گرفته شد. داده‌های بقا بر اساس متغیرهای تبیینی و از توزیع وایبول با پارامترهای $\gamma = 1$ و $\nu = 2$ تولید شد. در این مقاله داده‌های بقا از طریق تابع خطر پایه تولید شده‌اند. 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

$$S(t|X) = \exp(-H_0(t) \exp(\beta'X)) \quad (۸)$$

که در آن $H_0(\cdot)$ تابع خطر پایه تجمعی است که به صورت

$$H_0(t) = \int_0^t h_0(u) du$$

محاسبه می‌شود. با توجه به آنکه $S(t|X) = 1 - F(t|X)$ و $F(\cdot)$ دارای توزیع یکنواخت پیوسته در بازه $[0, 1]$ است، لذا $S(\cdot)$ نیز از این توزیع پیروی می‌کند و بر اساس (۸) داریم

$$U = \exp(-H_0(t) \exp(\beta'X))$$

که در آن U یک متغیر تصادفی از توزیع یکنواخت پیوسته در بازه $[0, 1]$ است و داریم

$$T = H_0^{-1}(-\log(U) \exp(-\beta'X)). \quad (۹)$$

لذا برای تولید داده‌های بقا از رابطه (۹) استفاده شده است. برای تولید داده‌های بقا از توزیع وایبول، تابع خطر پایه تجمعی $H_0(t) = \gamma t^\nu$ در نظر گرفته شد. بنابراین می‌توان متغیر تصادفی زمان بقا را با استفاده از یک متغیر تصادفی یکنواخت پیوسته U به صورت

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

تولید کرد. در این شبیه‌سازی مقادیر $\nu = 2$ و $\gamma = 1$ برابر یک در نظر گرفته شده‌اند. در این مقاله با در نظر گرفتن یک میدان تصادفی گاوسی^۱ روش بندر و همکاران [۱۵]، برای تولید داده‌های بقای فضایی تعمیم داده شد. معمولاً برای مدل بندی داده‌های فضایی از میدان تصادفی به صورت $\{Z(s); s \in D \subseteq R^d; d \geq 1\}$ استفاده می‌شود (محمدزاده [۱۶]). این میدان تصادفی گاوسی نامیده می‌شود، هرگاه برای هر $m \geq 1$ توزیع توأم

یک متغیر تصادفی پیوسته U با توزیع یکنواخت $u(0,1)$ داریم:

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

که در آن $Z(\cdot)$ میدان تصادفی گاوسی موردنظر و s موقعیت مکانی برحسب طول و عرض موقعیت واحد آماری است. با توجه به آنکه داده‌های بقای فضایی می‌توانند به صورت مشبکه-ای^۱ یا زمین آماری^۲ مشاهده شوند، در این مقاله داده‌های بقای فضایی زمین آماری با استفاده از تابع تغییرنگار در فواصل منظم تولید شده‌اند. برای تولید داده‌ها، n اثر فضایی روی یک مشبکه $\sqrt{n} \times \sqrt{n}$ در نظر گرفته شده و بر اساس ماتریس فاصله گره‌های این مشبکه، n داده

از میدان تصادفی گاوسی با تابع تغییرنگار^۳ گاوسی به صورت $C(d) = \sigma^2 \exp\left(-\frac{\|d\|^2}{a^2}\right)$

تولید شده است، که در آن $\|d\|$ فاصله اقلیدسی بین نقاط است و مقادیر $a=2$ و $\sigma^2=5$ به ترتیب برای واریانس و دامنه میدان تصادفی در نظر گرفته شده‌اند. به همین ترتیب داده بقا و بقای فضایی با ۲۰ و ۸۰ درصد سانسور و ۵۰۰ و ۱۰۰۰ بار تکرار تولید شده‌اند. برای مقایسه عملکرد دو مدل از معیارهای میانگین توان دوم خطا^۴ (MSE) و درصد قدرمطلق اربیبی^۵ (MAPB) برآوردها استفاده شده است. در این مطالعه داده‌های بقا و بقای فضایی بر اساس سه متغیر تبیینی در نظر گرفته شده تولید شده‌اند اما یک‌بار با هر سه متغیر تبیینی و بار دیگر دو متغیر تبیینی کم‌اهمیت‌تر در نظر گرفته شده است. در این مطالعه در هر مجموعه داده متغیر تبیینی تأثیرگذارتر بر داده‌ها (وزن و فشارخون) حذف شده 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 ارائه شده‌اند. 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 به مراتب دقیق‌تر از مدل خطرهای متناسب است.

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

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

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

شکنندگی			کاکس			پارامتر	سانسور	تکرار
MSE	MAPB	برآورد	MSE	MAPB	برآورد			
۰/۰۰۱	۵/۴۱۴	۰/۵۲۷	<۰/۰۰۱	۱/۲۵۱	۰/۱۵۰۶	β_1		
۰/۰۰۱	۴/۱۵۰۲	۰/۶۲۷	<۰/۰۰۱	۰/۳۴۳	۰/۱۶۰۲	β_2		
۰/۰۰۱	۵/۰۷۰	۰/۷۳۵	<۰/۰۰۱	۰/۱۸۸۹	۰/۱۷۰۶	β_3	۲۰	
۰/۰۱۰	۱۹/۷۷۴	۰/۴۰۱	۰/۰۲۸	۳۳/۲۳۶	۰/۳۳۴	β_1		
۰/۰۴۵	۳۵/۵۴۸	۰/۳۸۷	۰/۰۷۶	۴۶/۰۹۷	۰/۳۲۳	β_2		۵۰۰
۰/۰۰۱	۵/۱۵۶	۰/۵۲۶	<۰/۰۰۱	۱/۲۷۹	۰/۱۵۰۶	β_1		
<۰/۰۰۱	۳/۱۵۳	۰/۶۱۹	<۰/۰۰۱	۰/۶۴۴	۰/۱۵۹۶	β_2		
۰/۰۰۱	۵/۲۳۶	۰/۷۳۷	<۰/۰۰۱	۱/۳۵۴	۰/۱۷۰۹	β_3	۸۰	
۰/۰۰۸	۱۷/۱۸۱۵	۰/۴۱۱	۰/۰۱۲	۲۱/۶۳۷	۰/۳۹۲	β_1		
۰/۰۳۲	۲۹/۶۵۰	۰/۴۲۲	۰/۰۳۸	۳۲/۳۰۲	۰/۴۰۶	β_2		
<۰/۰۰۱	۴/۴۵۹	۰/۵۲۲	<۰/۰۰۱	۰/۶۰۴	۰/۱۵۰۳	β_1		
۰/۰۰۱	۴/۶۷۲	۰/۶۲۸	<۰/۰۰۱	۰/۱۸۶۰	۰/۱۶۰۵	β_2		
۰/۰۰۱	۴/۵۳۰	۰/۷۳۲	<۰/۰۰۱	۰/۶۶۲	۰/۱۷۰۵	β_3	۲۰	
۰/۰۰۵	۱۴/۶۷۱	۰/۴۲۷	۰/۰۲۸	۳۳/۷۵۱	۰/۳۳۱	β_1		
۰/۰۰۴	۹/۹۴۰	۰/۵۴۰	۰/۰۳۰	۲۸/۱۸۴۲	۰/۴۲۷	β_2		۱۰۰۰
۰/۰۰۱	۴/۹۶۳	۰/۵۲۵	<۰/۰۰۱	۰/۱۸۳۱	۰/۱۵۰۴	β_1		
۰/۰۰۱	۶/۰۲۷	۰/۶۳۶	<۰/۰۰۱	۲/۰۵۰	۰/۱۶۱۲	β_2		
۰/۰۰۱	۴/۵۶۲	۰/۷۳۲	<۰/۰۰۱	۰/۱۸۹۸	۰/۱۷۰۶	β_3	۸۰	
۰/۰۰۴	۱۲/۲۰۰	۰/۴۳۹	۰/۰۰۸	۱۷/۹۷۸	۰/۴۱۰	β_1		
۰/۰۰۱	۴/۴۲۴	۰/۶۲۷	۰/۰۰۱	۱/۴۰۸	۰/۱۵۹۲	β_2		

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 گرفته شده‌اند 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 افزایش یافته است اما این افزایش برای مجموعه اول فراوان تر است و علت آن پراکندگی بیشتر متغیر تبیینی وزن در مجموعه اول و در نتیجه 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

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

به‌طور کلی 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

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

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 مورد بررسی 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. از این‌رو 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 مورد توجه 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] Gutierrez, 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, Trabiati 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.