

مدل‌بندی فضایی داده‌های بقای سانسور شده

کیومرث مترجم، محسن محمدزاده*، آمنه آبیاری؛ گروه آمار، دانشگاه تربیت مدرس

پذیرش ۹۳/۷/۲۲

دریافت ۹۳/۱/۲۵

چکیده

در تحلیل بقا تشخیص منابع خطر برای مجموعه داده‌های بقا اهمیت ویژه‌ای دارد. برخی از این عوامل قابل شناسایی هستند و در قالب متغیرهای تبیینی در مدل‌هایی مانند مدل مخاطرات متناسب کاکس قابل بیانند. اما برخی دیگر از این عوامل خطر، ناشناخته و قابل شناسایی یا اندازه‌گیری نیستند. یکی از این منابع تأثیرگذار، همبستگی فضایی میان داده‌های بقای سانسور شده است که در پژوهش‌هایی که تا کنون صورت گرفته کمتر به آن توجه شده است. این مقاله مدل بقای فضایی برای تحلیل داده‌های بقای فضایی زمین آماری راست سانسور معرفی می‌شود و با پژوهشی شبیه‌سازی کارایی مدل‌های کاکس، شکنندگی و بقای فضایی در برازش به این‌گونه داده‌های بقا بررسی می‌شود. سپس نحوه کاربست مدل بقای فضایی برای تحلیل داده‌های مربوط به زمان ابتلا به بیماری سرکوسپوریوز در باغات زیتون نشان داده شده است.

واژه‌های کلیدی: مدل مخاطرات متناسب کاکس، عوامل خطرناشناخته، اثر تصادفی فضایی، مدل بقای فضایی.

مقدمه

در بسیاری از پژوهش‌های آماری رفتار متغیرهایی تحت عنوان مدت زمان، مانند دوره نقاهت بیماری، مدت زمان ابتلا به بیماری تا مرگ بیمار، مدت زمان تصویب لایحه‌ای در مجلس، مدت زمان تدوین آئین‌نامه قوانین، طول عمر یک قطعه و مدت زمان جذب یک دانش‌آموخته دانشگاه به بازار کار بررسی می‌شود. در این‌گونه پژوهش‌ها به‌ویژه در بررسی‌های پزشکی، هدف یافتن و مدل‌بندی منابع تأثیرگذار بر مدت زمان بررسی شده است تا با تغییر منابع خطر، مدت زمان مذکور کنترل شود. این‌گونه داده‌های برحسب زمان، داده‌های بقا و روش‌های تحلیل آن‌ها تحلیل بقا نامیده می‌شوند. در روش‌های تحلیل بقا مرسوم‌ترین مدل در برازش به داده‌های بقا مدل کاکس است، اما استفاده از این مدل مستلزم برقراری چند فرض است. یکی از مفروضات مدل کاکس استقلال داده‌های بقا است که در صورت تشخیص نادرست یا بی‌توجهی به این فرض، استفاده از مدل کاکس برای برازش به مجموعه‌ای از داده‌های بقای همبسته می‌تواند موجب تحلیل نادرست و گمراهی در اخذ تصمیم صحیح شود. وقتی مشاهدات مستقل نباشند و وابستگی آن‌ها تابعی از فاصله بین موقعیت‌های مشاهدات باشد، داده‌های فضایی نامیده می‌شوند [۱] اگر داده‌های بقا شامل چنین ویژگی‌ای باشند نحوه تحلیل آن‌ها متفاوت است.

معمولاً برای وارد کردن عوامل خطر ناشناخته در مدل بقا از مدل‌های شکنندگی استفاده می‌شود که در آن یک متغیر تصادفی با تکیه‌گاه مثبت و میانگین یک در مدل کاکس وارد می‌شود به همین دلیل مدل‌های شکنندگی را کاکس اصلاح شده نیز می‌نامند. گوتیرز مدل‌های شکنندگی و شکنندگی مشترک پارامتری بقا را بررسی کرد [۱۱].

*نویسنده مسئول mohsen_m@modares.ac.ir

هندرسون و همکاران زمان بقای بیماران مبتلا به سرطان خون را با توجه به همبستگی فضایی میان زمان‌های بقا بررسی کردند [۱۲]. ژائو و همکاران با استفاده از درخت‌های پولیا^۱ یک رهیافت ناپارامتری منعطف برای مدل‌بندی داده‌های بقای فضایی ارائه کردند [۱۸]. ژائو هموارسازی فضایی خطر پایه بقا^۲ را برای مدل‌بندی بقای درخت‌های پولیا به‌کار برد [۱۷]. هوانگ و همکاران آماره کاوشی فضایی^۳ را براساس مدل نمایی برای داده‌های بقای سانسور شده و سانسور نشده پیوسته ارائه کردند [۱۳]. دارموفال مدل‌بندی فرایندهای سیاسی^۴ وابسته فضایی را مطرح کرد [۹]. لی و لین مدل‌های انتقالی نرمال نیمه‌پارامتری برای داده‌های بقای همبسته فضایی را مطرح کردند [۱۴]. کوک و لی آزمون‌های جای‌گشتی برای شناسایی خوشه‌های فضایی با داده‌های سانسور شده را بررسی کردند [۷]. کوک و همکاران شناسایی خوشه‌بندی فضایی برای برآمدهای طولی را بررسی کردند [۶]. در این مقاله سه مدل مخاطرات متناسب، شکنندگی و بقای فضایی معرفی و عمل‌کرد آن‌ها در برازش به داده‌های بقای همبسته فضایی مقایسه می‌شود.

در بخش ۲ مدل بقای فضایی به تفصیل معرفی می‌شود. در بخش ۳ طی بررسی شبیه‌سازی مجموعه‌ای از داده‌های بقای فضایی تولید و کارایی سه مدل کاکس، شکنندگی و بقای فضایی در برازش مدل به داده‌ها مقایسه می‌شود. در بخش ۴ داده‌های مربوط به بیمار سرکوسپوریوز باغات زیتون با استفاده از مدل بقای فضایی معرفی شده و مدل‌بندی می‌شود. سپس به بحث و نتیجه‌گیری پرداخته خواهد شد.

مدل‌های بقای فضایی

به‌طور کلی برای بررسی و تحلیل داده‌های بقا از تابع خطر استفاده می‌شود. تابع خطر $h(t)$ پتانسیل لحظه‌ای وقوع اتفاق مورد نظر را به شرط رخ ندادن اتفاق مورد نظر تا زمان t برای هر مشاهده آماری نشان می‌دهد. مرسوم‌ترین تابع خطر را کاکس [۸] بدین‌صورت تعریف کرد:

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

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. در برخی پژوهش‌ها عوامل تأثیرگذار بر داده‌های بقا قابل مشاهده یا قابل اندازه‌گیری نیستند. در این صورت از مدل‌های شکنندگی برای برازش به داده‌های بقا استفاده می‌شود. ایده مؤلفه شکنندگی را اولین بار واپل و همکاران [۱۶] برای مدل‌های بقا تک متغیره مطرح کردند، سپس کلایتون و کازیک [۵] به مدل‌های شکنندگی چندمتغیره تعمیم دادند. در واقع مدل‌های شکنندگی نوعی مدل با اثرات تصادفی برای تحلیل داده‌های بقا هستند. تابع خطر مدل شکنندگی بدین‌صورت است:

$$h(t|Z, h_0, X) = Zh_0(t) \exp(\beta'X)$$

که در آن t زمان بقا، $h_0(\cdot)$ تابع خطر پایه، X بردار متغیرهای تبیینی، β بردار اثر متغیرهای تبیینی و Z اثر شکنندگی است. مدل شکنندگی تعمیمی از مدل کاکس با هدف مدل‌بندی همبستگی درون خوشه‌ای است.

برای در نظر گرفتن عوامل خطر ناشناخته از مدل‌های شکنندگی استفاده می‌شود. اما در صورتی‌که اطلاعات

1. Poly trees
2. Spatial smoothing of baseline survival
3. Scan Statistic
4. Political Event Processes

بیش‌تری در مورد منبع اثر در اختیار باشد لازم است با استفاده از اطلاعات اضافی به نتایج دقیق‌تری دست یافت. یکی از این عوامل ناشناخته تأثیرگذار بر داده‌های بقا، موقعیت مکانی مشاهده مورد نظر است به‌طوری که هر داده بقا برحسب مکان استقرار خود با داده‌های دیگر وابستگی مکانی داشته باشد. به عبارت دیگر مشاهدات نزدیک به هم وابسته‌تر و مشاهدات دورتر از هم وابستگی کم‌تری دارند. اما این اثر فضایی ممکن است به‌طور مستقیم قابل شناسایی یا قابل مشاهده نباشد. بنابراین با در نظر گرفتن همبستگی فضایی در مدل‌های شکنندگی^۵ رده جدیدی از مدل‌های بقا با عنوان مدل‌های بقای فضایی ارائه می‌شود.

برای مدل‌بندی داده‌های فضایی به‌طور معمول از میدان تصادفی^۶ استفاده می‌شود، که مجموعه‌ای از متغیرهای تصادفی مانند $\{Z(s); s \in D\}$ است، که در آن مجموعه اندیس‌گذار D زیرمجموعه‌ای از فضای اقلیدسی $d \geq 1$ بعدی R^d است. ساختار همبستگی این میدان تصادفی با تغییرنگار

$$C(h) = \text{Cov}(Z(s), Z(s+h)), \quad s, h \in D$$

مشخص می‌شود. بنابراین برای منظور کردن همبستگی فضایی داده‌ها با لحاظ کردن یک میدان تصادفی $Z(\cdot)$ با تغییرنگار معلوم، تابع خطر فضایی به‌صورت

$$h(s, t|X, Z(s)) = h_0(t) \exp(\beta'X + Z(s))$$

در نظر گرفته می‌شود، که در آن زمان بقا، $h_0(\cdot)$ تابع خطر پایه، X بردار متغیرهای تبیینی، β بردار ضرائب و s موقعیت جغرافیایی مشاهده میدان تصادفی $Z(s)$ است.

چنانچه میدان تصادفی $Z(\cdot)$ گاوسی^۷ باشد استنباط مدل بقای فضایی براساس تابع درست‌نمایی رتبه‌ای انجام می‌شود. لی و ریان مدل بقای فضایی را در حالتی که میان زمان‌های بقای چند ناحیه خودهمبستگی فضایی وجود داشته باشد، معرفی کردند [۱۵]. در مدل ارائه شده لی و ریان خودهمبستگی فضایی صرفاً بین نواحی در نظر گرفته می‌شود و برای مشاهدات درون هر ناحیه فرض استقلال برقرار است [۱۵]. در بسیاری از مواقع این همبستگی فضایی ممکن است بین مشاهدات درون نواحی نیز وجود داشته باشد یعنی به‌عبارتی داده‌ها از نوع زمین‌آماری باشند، در این حالت مدل ارائه شده قابل استفاده نخواهد بود.

در این مقاله مدل بقای فضایی برای حالتی که زمان‌های بقا به‌صورت زمین‌آماری باشند معرفی می‌شود. با تعمیم مدل بقای فضایی تابع درست‌نمایی رتبه‌ای زمان‌های بقای فضایی از طریق مدل جدید بدین‌صورت است:

$$L(\beta, \Theta) = \int \cdots \int \prod_{i=1}^n \left\{ \frac{\exp\{\beta'X_{ij} + Z_i\}}{\sum_{i'} Y_{i'}(t_i) \exp\{\beta'X_{i'} + Z_{i'}\}} \right\}^{\delta_i} dF(Z_1, \dots, Z_n) \quad (1)$$

که در آن $Z_i = Z(s_i)$ و $Y_{i'}(t_i)$ تابع نشان‌گر بدین‌صورت است:

$$Y_{i'}(t_i) = \begin{cases} 1, & t_{i'} \geq t_i \\ 0, & t_{i'} < t_i \end{cases}$$

n تعداد کل افراد بررسی شده و

$$\delta_i = \delta(t_i) = \begin{cases} 1, & t_i \leq t_c \\ 0, & t_i > t_c \end{cases}$$

5. Frailty models
6. Random field
7. Gaussian random field

تابع نشان‌گر مربوط به سانسور داده بقای فرد i ام است به طوری که t_c زمان سانسور است.

برای دستیابی به این مدل ابتدا فرض کنید $T_{(1,0)} < \dots < T_{(L,0)}$ نشان‌دهنده L زمان شکست مجزا باشد، $T_{(\ell,k)}$ برای $k = 1, \dots, c_\ell$ نشان‌دهنده c_ℓ زمان بقا سانسور شده بلافاصله پس از زمان $T_{(\ell,0)}$ و پیش از زمان $T_{(\ell+1,0)}$ است. باید توجه داشت که با فرض پیوسته بودن توزیع زمان بقا، احتمال وجود گره در زمان‌های بقا

صفر است. فلمینگ و هرینگتون [۱۰] نشان دادند، احتمال تمام رتبه‌بندی‌های سازگار برابر

$$P(T_{(1,0)} < \dots < T_{(L,0)}, T_{(\ell,0)} < T_{(\ell,k)}, k = 1, \dots, c_\ell, \ell = 1, \dots, L) \quad (۲)$$

است، که در آن رتبه‌بندی نسبی بین زمان‌های بقا سانسور شده در نظر گرفته نمی‌شود. سپس نگاهشستی یک‌به‌یک از (ℓ, k) برای $k = 0, \dots, c_\ell$ و $\ell = 1, \dots, L$ به شمارنده اصلی i برای $i = 1, \dots, n$ به صورت $i_{(\ell,k)} \leftrightarrow (\ell, k)$ در نظر گرفته می‌شود. احتمال این که مقدار $T_{(\ell,k)}$ برای $k = 0, \dots, c_\ell$ به شرط متغیرهای تیبینی مشاهده شده، اثرهای تصادفی مشاهده نشده و $T_{(\ell,0)} = t$ ، بیش‌تر از $T_{(\ell,0)}$ باشد عبارتست از:

$$g_\ell(t) = P(T_{(\ell,k)} > T_{(\ell,0)}, k = 1, \dots, c_\ell | T_{(\ell,0)} = t, X_{(\ell,k)}, Z_{i_{(\ell,k)}}, k = 1, \dots, c_\ell) \\ = e^{-H_0(t) \sum_{k=1}^{c_\ell} \exp\{X'_{(\ell,k)}\beta + Z_{i_{(\ell,k)}}\}}$$

تابع چگالی شرطی $T_{(\ell,0)}$ نیز بدین صورت است:

$$f(t | X_{(\ell,0)}, Z_{i_{(\ell,0)}}) = h_0(t) \exp\{\beta' X_{(\ell,0)} + Z_{i_{(\ell,0)}}\} e^{H_0(t) \exp\{\beta' X_{(\ell,0)} + Z_{i_{(\ell,0)}}\}}$$

بنابراین می‌توان احتمال رتبه (۲) را بدین صورت نوشت:

$$\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \int_0^{\infty} \int_{t_1}^{\infty} \dots \int_{t_{L-1}}^{\infty} \prod_{\ell=1}^L f(t_\ell | X_{(\ell,0)}, Z_{i_{(\ell,0)}}) g_\ell(t_\ell) dt_L \dots dt_1 dF(Z_1, \dots, Z_n) \\ = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \prod_{\ell=1}^L \frac{\exp\{\beta' X_{(\ell,0)} + Z_{i_{(\ell,0)}}\}}{\sum_{(\ell',k') \in R_\ell} \exp\{\beta' X_{(\ell',k')} + Z_{i_{(\ell',k')}}\}} dF(r_1, \dots, r_M)$$

که در آن $dF(Z_1, \dots, Z_n)$ توزیع توام اثرهای تصادفی و R_ℓ مجموعه‌های مخاطره روی $T_{(\ell,0)}$ است، یعنی $R_\ell = \{(j, k): j = \ell, \dots, L, k = 0, \dots, c_j\}$ در نهایت در فرایندهای مخاطره می‌توان عبارت درست‌نمایی را به صورت رابطه (۱) به دست آورد.

با ماکسیم کردن تابع درست‌نمایی رتبه‌ای می‌توان برآورد پارامترهای رگرسیونی و پارامترهای فضایی را به دست آورد. چون تابع درست‌نمایی فرم بسته‌ای ندارد و از روش‌های عددی مونت کارلو برای محاسبه برآوردهای ماکسیم درست‌نمایی پارامترها بهره گرفته شده است.

بررسی شبیه‌سازی

برای مقایسه عمل‌کرد مدل‌های کاکس، شکنندگی و بقای فضایی جدید در صورت وجود همبستگی فضایی در داده‌ها، مجموعه داده بقای فضایی تولید می‌شود. از آنجاکه روش‌های موجود تولید داده‌های بقای فضایی عمدتاً زمان‌بر هستند روشی برای تولید داده‌های بقای فضایی معرفی می‌شود.

بندر و همکاران با استفاده از تابع بقا روشی برای شبیه‌سازی داده‌های بقای کلاسیک ارائه کردند. در اینجا با تعمیم این روش و وارد کردن یک اثر تصادفی فضایی، روشی ساده برای تولید داده‌های بقای فضایی ارائه می‌شود. بدین منظور با توجه به تابع بقا داریم [۴].

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

که در آن

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

بنابراین براساس مدل کاکس تابع توزیع تجمعی مدل کاکس بدین صورت است:

$$F(t|X) = 1 - \exp[-H_0(t) \exp(\beta'X)]$$

فرض کنید Y یک متغیر تصادفی با تابع توزیع F باشد. در این صورت $U = F(Y)$ دارای توزیع یکنواخت $U(0,1)$ است. از این رو، $S(Y) = 1 - U$ نیز دارای توزیع $U(0,1)$ است. بنابراین برای تولید داده بقا براساس مدل کاکس داریم

با در نظر گرفتن یک میدان تصادفی گاوسی با یک ساختار کواریانس معتبر و اضافه کردن آن به عنوان اثر فضایی در مدل کاکس داده بقای فضایی بدین صورت تولید می‌شود:

$$T = H_0^{-1} \left[-\log(U) \times \exp(-\beta'x + Z(s)) \right]$$

که در آن $Z(\cdot)$ میدان تصادفی گاوسی مورد نظر و s موقعیت مکانی برحسب طول و عرض موقعیت هر واحد آماری است. حال با در نظر گرفتن یک تابع خطر پایه دل‌خواه می‌توان به تولید داده بقای فضایی پرداخت. بنابراین می‌توان متغیر تصادفی زمان بقا فضایی را با استفاده از یک متغیر تصادفی یکنواخت پیوسته U و تابع خطر پایه نمایی بدین صورت تولید کرد:

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

در این روش می‌توان تابع خطر پایه را براساس توزیع‌هایی با تکیه‌گاه مثبت مانند وایبول، گاما، گامپرتز در نظر گرفت که با توجه به فرم معرفی شده در حالت کلی به صورت سراسر قابل محاسبه است. در این بررسی برای تولید n اثر فضایی یک شبکه $\sqrt{n} \times \sqrt{n}$ در نظر گرفته شده و بر اساس ماتریس فاصله گره‌های این شبکه، n داده از میدان تصادفی گاوسی با تابع تغییرنگار نمایی به صورت $C(h) = \sigma^2 \exp(-||h||/a)$ تولید شده است، که در آن $||h||$ فاصله اقلیدسی بین دو موقعیت فضایی است و $\sigma^2 = 1$ و $a = 1$ و $\gamma = 1$ در نظر گرفته شده‌اند.

بر این اساس سه مجموعه داده شامل ۱۰۰ داده بقای فضایی با تابع خطر

$$h(s, t|X, Z(s)) = h_0(t) \exp(\beta X + Z(s))$$

تولید شده است. شبیه‌سازی این مجموعه داده با منظورکردن $\beta = 1$ ، تابع خطر پایه نمایی با میانگین یک و در نظر گرفتن یک متغیر تبیینی از توزیع نرمال استاندارد با سانسورهای ۲۰، ۴۰، ۶۰ و ۸۰ درصد و تکرارهای ۱۰۰، ۲۰۰ و ۵۰۰ صورت گرفته است [۲].

در مرحله اول به داده‌های شبیه‌سازی شده، مدل کاکس و شکنندگی با متغیر شکنندگی گاوسی برازش داده شد. سپس با روش مونت کارلو تابع درست‌نمایی رتبه‌ای ماکسیمم و پارامترهای مدل بقای فضایی برآورد شدند. با دو ملاک میانگین توان دوم مانده‌ها^۸ (MSE) و میانگین درصد قدرمطلق اریبی^۹ (MAPB) برآوردها نتایج حاصل از برازش سه مدل به داده‌ها ارزیابی و مقایسه شده‌اند. چنان‌که در جدول ۱ ملاحظه می‌شود، مقدار ملاک MSE بیانگر آن است که مدل شکنندگی نسبت به مدل مخاطرات متناسب، نتایج دقیق‌تری در رابطه با برآورد پارامترهای رگرسیونی ارائه کرده است اما با افزایش درصد سانسور داده‌ها دقت برآورد پارامترها در هر دو مدل به‌طور چشمگیری کاهش یافته است. برآوردهای حاصل از مدل کاکس کم‌برآورد و برآورد پارامترهای رگرسیونی مدل بقای فضایی کمی بیش‌برآورد هستند. با افزایش سانسور داده‌ها مقدار ملاک MSE برای برآورد پارامترهای مدل بقای فضایی نیز افزایش یافته است اما این افزایش نسبت به مدل کاکس و شکنندگی بسیار کمتر است. هرچند براساس ملاک MSE مدل شکنندگی نسبت به مدل مخاطرات متناسب نتایج دقیق‌تری در رابطه با برآورد پارامترهای رگرسیونی ارائه کرده است اما با افزایش درصد سانسور داده‌ها دقت برآورد پارامترها در هر دو مدل به‌طور چشمگیری کاهش یافته است. چنان‌که در جدول ۲ ملاحظه می‌شود براساس ملاک MSE استفاده از مدل بقای فضایی به‌طور چشمگیری موجب افزایش دقت برآورد پارامترهای رگرسیونی شده و براساس ملاک MAPB در تکرارهای ۲۰۰ و ۵۰۰ میزان اریبی برآوردهای مدل بقای فضایی بسیار کمتر از اریبی برآوردهای مدل کاکس و شکنندگی است.

جدول ۱. برآورد پارامتر β مدل‌های مخاطرات متناسب کاکس و شکنندگی

مدل	درصد سانسور	تکرار								
		۵۰۰		۲۰۰		۱۰۰				
		MSE	MAPB	برآورد	MSE	MAPB	برآورد	MSE	MAPB	برآورد
کاکس	۲۰	۰/۰۳۰۵	۱۹/۴۵۱۹	۰/۸۰۵۵	۰/۰۴۸۰	۲۱/۹۰۳۰	۰/۷۸۱۰	۰/۰۵۳۹	۲۳/۲۰۷۰	۰/۷۶۷۹
	۴۰	۰/۰۴۰۰	۲۰/۰۰۹۵	۰/۷۹۹۹	۰/۰۴۹۱	۲۱/۱۶۶۶	۰/۷۷۸۳	۰/۰۵۸۰	۲۴/۰۹۳۱	۰/۷۵۹۱
	۶۰	۰/۰۴۵۸	۲۱/۳۹۳۱	۰/۷۸۶۱	۰/۰۵۷۴	۲۳/۹۵۰۵	۰/۷۶۰۵	۰/۰۶۲۹	۲۵/۰۷۳۵	۰/۷۴۹۲
	۸۰	۰/۰۴۶۰	۲۱/۴۴۴۳	۰/۷۸۵۶	۰/۰۶۴۵	۲۵/۳۹۹۷	۰/۷۴۶۰	۰/۰۷۸۲	۲۷/۹۶۱۲	۰/۷۲۰۴
شکنندگی	۲۰	۰/۰۳۰۵	۱۷/۴۷۴۵	۰/۸۰۵۵	۰/۰۳۳۸	۱۸/۳۷۳۲	۰/۸۱۶۳	۰/۰۳۵۴	۱۸/۸۰۳۱	۰/۸۱۲۰
	۴۰	۰/۰۳۲۹	۱۸/۱۴۴۲	۰/۸۱۸۶	۰/۰۳۶۱	۱۹/۰۰۸۵	۰/۸۰۹۹	۰/۰۳۹۲	۱۹/۸۰۷۹	۰/۸۰۱۹
	۶۰	۰/۰۳۳۳	۱۸/۲۴۳۶	۰/۸۱۷۶	۰/۰۴۲۰	۲۰/۵۰۰۰	۰/۷۹۵۰	۰/۰۴۸۱	۲۱/۹۲۵۸	۰/۷۸۰۷
	۸۰	۰/۰۳۸۸	۱۹/۶۹۷۶	۰/۸۰۳۰	۰/۰۴۵۳	۲۱/۲۸۷۵	۰/۷۸۷۱	۰/۰۵۲۹	۲۳/۰۰۱۳	۰/۷۷۰۰

مثال کاربردی

در این بخش با استفاده از مدل بقای فضایی زمین آماری داده‌های شیوع بیماری سرکوسپوریوز ۱۰ در باغ زیتون واقع در غرب کشور بررسی می‌شود. بیماری یاد شده یکی از بیماری‌های شایع در باغات زیتون است که هر سال سبب وارد آمدن خسارات بسیار زیادی به باغداران می‌شود. سرکوسپوریوز یک بیماری قارچی است که در آن قارچ دارای اسکلروت‌هایی^{۱۱} به‌رنگ قهوه‌ای مایل به سبز و دایره‌ای یا تا حدودی سیلندری شکل است. این قارچ باعث به‌وجود آمدن لکه در برگ‌های زیتون می‌شود. علائم بیماری به شکل لکه‌های خاکستری تا سربی رنگی است که در سطح زیرین برگ به‌وجود می‌آید و سپس کاملاً سیاه می‌شود. سطوح بالایی برگ‌ها نیز زرد

8. Mean Square Error

9. Mean Absolute Percentage Bias

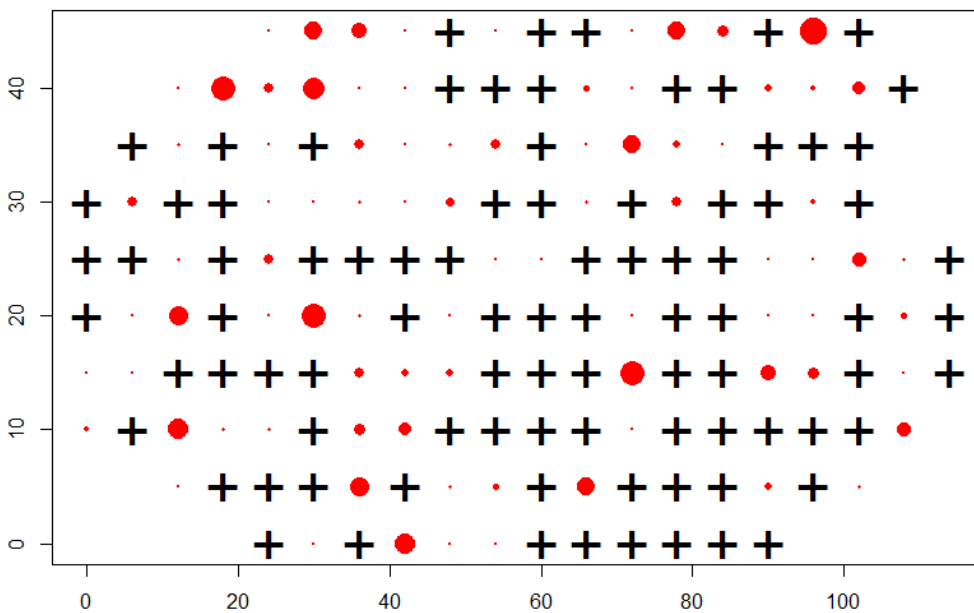
10. Cercosporiose

11. Sclerotia

شده و در نهایت سبب ریزش برگ‌ها از شاخه می‌شود. اطلاعات اندکی درباره نحوه شیوع این بیماری وجود دارد. عامل بارش نقش مهمی در انتقال قارچ از برگ‌های آلوده به برگ‌های سالم دارد. در ضمن باد هم می‌تواند قارچ را از برگ‌های افتاده در کف باغ به سایر برگ‌ها منتقل کند. از این رو، به‌نظر می‌رسد موقعیت قرارگیری و فاصله درختان از یکدیگر نیز می‌تواند یکی از عوامل تأثیرگذار در شیوع این بیماری باشد.

جدول ۲. برآورد پارامترهای مدل بقای فضایی

درصد سانسور	پارامتر	تکرار								
		۵۰۰			۱۰۰					
		MSE	MAPB	برآورد	MSE	MAPB	برآورد			
۲۰	σ^2	۰/۰۰۰۸	۲/۸۰۱۸	۱/۰۲۸۰	۰/۰۲۱۵	۱۴/۶۷۹۰	۰/۸۵۳۲	۰/۰۵۴۲	۲۳/۲۸۵۱	۱/۲۳۲۹
	a	۰/۰۰۰۶	۲/۵۱۷۶	۰/۹۷۴۸	۰/۰۱۱۳	۱۰/۶۳۰۵	۰/۸۹۳۷	۰/۰۱۸۹	۱۳/۷۶۱۸	۰/۸۶۲۴
		۰/۰۰۰۵	۲/۳۰۱۸	۰/۹۷۷۰	۰/۰۰۰۹	۲/۹۷۰۷	۰/۹۱۰۳	۰/۰۰۱۰	۳/۲۲۶۰	۰/۹۶۷۷
۴۰	σ^2	۰/۰۰۴۶	۶/۸۰۶۹	۱/۰۶۸۱	۰/۰۲۷۴	۱۶/۵۶۶۳	۱/۱۶۵۷	۰/۰۵۸۸	۲۴/۲۴۱۲	۱/۲۴۲۴
	a	۰/۰۲۶۴	۱۶/۲۳۳۱	۰/۸۳۷۷	۰/۰۲۸۳	۱۶/۸۰۸۴	۰/۸۳۱۹	۰/۰۶۳۹	۲۵/۲۶۹۵	۰/۷۴۷۳
		۰/۰۰۶۹	۸/۳۲۹۴	۱/۹۱۶۷	۰/۰۲۴۶	۱۵/۶۷۴۴	۰/۸۴۳۳	۰/۰۳۱۶	۱۷/۷۶۸۹	۰/۸۲۲۳
۶۰	σ^2	۰/۰۰۶۵	۸/۰۸۷۶	۱/۰۸۰۹	۰/۰۲۸۵	۱۶/۸۷۳۹	۰/۸۳۱۳	۰/۰۶۸۵	۲۶/۱۷۶۸	۱/۲۶۱۸
	a	۰/۰۳۳۷	۱۸/۳۴۷۸	۰/۸۱۶۵	۰/۰۵۲۰	۲۲/۷۹۹۹	۰/۷۷۲۰	۰/۰۸۵۰	۲۹/۱۵۳۶	۰/۷۰۸۵
		۰/۰۲۵۹	۱۶/۱۰۱۹	۰/۸۳۹۰	۰/۰۲۹۹	۱۷/۲۸۸۳	۰/۸۲۷۱	۰/۰۵۴۴	۲۳/۳۳۲۳	۰/۷۶۶۷
۸۰	σ^2	۰/۰۰۷۰	۸/۳۵۵۲	۱/۰۸۳۶	۰/۰۳۳۷	۱۸/۳۴۷۱	۱/۱۸۳۵	۰/۰۷۷۴	۲۷/۸۱۸۷	۱/۲۷۸۲
	a	۰/۰۳۵۳	۵/۷۹۳۸	۰/۸۱۲۱	۰/۰۶۲۲	۲۴/۹۳۰۷	۰/۷۵۰۵	۰/۰۹۲۳	۳۰/۳۷۶۹	۰/۶۹۶۲
		۰/۰۴۰۰	۱۹/۹۸۹۲	۰/۸۰۰۱	۰/۰۵۰۸	۲۲/۵۴۴۵	۰/۷۷۴۶	۰/۰۹۲۰	۳۰/۳۳۱۵	۰/۶۹۶۷



شکل ۱. موقعیت درختان در باغ و زمان ابتلا به بیماری (مشاهدات و داده‌های سانسور شده)

در این پژوهش باغی به مساحت تقریبی ۵۰۰۰ متر مربع که در آن ۱۷۳ اصله نهال وجود دارد که به‌مدت دو ماه و به‌صورت روزانه در خصوص آلودگی درختان به بیماری سرکوسپوریوز، پیگیرانه بررسی شده است. سه متغیر تبیینی شامل سن درخت، گونه درختان زیتون (شامل سه گونه) و ارتفاع در مدل منظور شده است. در صورت آلوده شدن هر درخت به بیماری با توجه به علائم یاد شده، زمان ابتلا یادداشت شده است. در پایان

بررسی‌ها تعداد ۸۵ اصله از درختان باغ مبتلا به بیماری شده و مابقی سالم مانده‌اند که در حقیقت سانسور راست صورت گرفته است. از این رو، در این بررسی با سانسور راست تقریباً ۵۱ درصدی روبه‌رو هستیم. موقعیت قرارگیری درختان در شکل ۱ نمایش داده شده است که در آن درختانی که مبتلا به بیماری شده‌اند با دایره و درختانی که تا پایان بررسی مبتلا نشده‌اند با علامت به‌علاوه نشان داده شده‌اند. مساحت دایره‌های قرمز رنگ نشان‌دهنده زمان ابتلا به بیماری است به‌طوری که دایره‌های با مساحت کمتر نشان‌دهنده ابتلا در روزهای اول بررسی است.

جدول ۳. برآورد پارامتر مدل‌های برازنده شده به داده‌های بیماری درختان زیتون

پارامتر	کاکس		شکندگی		بقای فضایی	
	برآورد	انحراف استاندارد	برآورد	انحراف استاندارد	برآورد	انحراف استاندارد
	-۰/۳۸۱	۰/۱۲۹	-۰/۶۰۰	۰/۱۹۲	-۰/۵۸۱	۰/۰۱۵
	-۰/۶۶۲	۰/۱۵۶	-۱/۰۲۶	۰/۲۲۱	-۱/۳۹۳	۰/۰۶۲
	۰/۳۵۱	۰/۱۱۹	۰/۴۷۷	۰/۱۸۱	۰/۵۹۹	۰/۰۸۸
	-	-	-	-	۱/۰۰۱	۰/۱۰۳
	-	-	-	-	۰/۳۰۴	۰/۰۴۲
<i>L</i>	-۳۹۶/۶۶۷		-۳۲۰/۰۴۶		-۲۹۹/۶۶۲	
AIC	۷۹۹,۳۳۳۶		۶۴۶,۰۹۱۸		۶۰۹,۳۲۵۳	

در این مرحله برای تعیین میزان تأثیر هر یک از متغیرهای تبیینی و همچنین موقعیت مکانی درختان بر ابتلا به بیماری سه مدل کاکس، شکندگی و بقای فضایی به داده‌ها برازش داده شد. چنان‌که در جدول ۳ ملاحظه می‌شود برآورد ضرایب متغیرهای تبیینی سن و گونه درختان در هر سه مدل منفی است که نشان‌دهنده اثر عکس بر خطر (زمان) ابتلا به بیماری است. به‌طور مثال در مدل بقای فضایی ضریب متغیر تبیینی سن درختان برابر ۰/۵۸۱- است بنابراین اگر دو درخت با سن متفاوت اما با شرایط یکسان تحت عوامل دیگر مد نظر باشند خطر ابتلا به بیماری برای درختی که سن بیشتری دارد به‌ازای هر سال $0/559 = e^{-0/581}$ برابر خطر ابتلا به بیماری در درخت با سن کمتر است. پس هر چقدر سن درخت بیشتر باشد خطر ابتلا به بیماری برای آن کمتر خواهد بود اما خطر ابتلا به بیماری در درختان بلندتر بیشتر است. چنان‌که ملاحظه می‌شود انحراف معیار برآوردگر پارامترهای رگرسیونی در مدل فضایی بسیار کمتر از انحراف معیار در مدل‌های کاکس و شکندگی است. برای مقایسه مدل‌ها از ملاک اطلاع آکائیک^{۱۲} $AIC = 2k - 2 \ln(L(\beta, \Theta))$ استفاده شده است [۳]، که در آن k تعداد پارامترهای مدل است. براساس این ملاک، مدلی که مقدار AIC آن کمتر باشد برازش بهتری نسبت به مدل‌های دیگر به داده‌ها ارائه می‌کند. بنابراین نتایج حاکی از برتری مدل بقای فضایی بر دو مدل دیگر است.

بحث و نتیجه‌گیری

در این مقاله مدل بقای فضایی برای داده‌های زمین آماری معرفی شد و در بررسی شبیه‌سازی عمل‌کرد این مدل با مدل‌های کاکس و شکندگی مقایسه عددی شد. نتایج نشان می‌دهد که هنگام وجود همبستگی فضایی در داده‌ها، مدل‌های کاکس و شکندگی قادر به ارائه تحلیلی مناسب از داده‌ها نیستند و عملاً در صورت وجود همبستگی فضایی نتایج این مدل‌ها قابل اعتماد نیست. به‌ویژه با افزایش اثر فضایی داده‌ها ضعف مدل‌های کاکس و شکندگی در یک برازش مناسب به داده‌ها پررنگ‌تر می‌گردد. نتایج بررسی شبیه‌سازی نشان می‌دهد که

12. Akaike Information Criterion

افزایش درصد سانسور سبب کاهش دقت برآورد پارامترهای مدل کاکس و شکنندگی می‌شود. در مدل بقای فضایی نیز با افزایش سانسور داده‌ها مقدار ملاک MSE برای برآورد پارامترهای مدل نیز افزایش یافته است یعنی در این حالت نیز با افزایش درصد سانسور دقت برآوردها کاهش می‌یابد. اما این کاهش دقت برای مدل‌های بقای فضایی نسبت به مدل کاکس و شکنندگی بسیار کمتر است. در هر صورت با افزایش درصد سانسور میزان اریبی برآورد پارامترها در هر سه مدل افزایش می‌یابد که باید به آن توجه شود. در ادامه با برآزش مدل‌های مورد بحث به داده‌های مربوط به زمان ابتلا به بیماری سرکوسپوریوز درختان زیتون، مدل بقای فضایی برآزش مناسب‌تری به داده‌ها را نشان داد. با برآزش مدل بقای فضایی تابع خطر نسبی دقیق‌تری به داده‌ها برآزش داده شد. طبق نتایج حاصل هرچه فاصله درختان از درخت مبتلا به بیماری بیشتر باشد خطر ابتلا به بیماری کمتر خواهد بود.

تقدیر و تشکر

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

منابع

۱. محمدزاده م.، *آمار فضایی و کاربردهای آن*، مرکز نشر آثار علمی دانشگاه تربیت مدرس، تهران، ایران (۱۳۹۱).
۲. مترجم ک.، محمدزاده م.، آبیاری آ.، *عملکرد مدل‌های کاکس و شکنندگی برای داده‌های بقای فضایی*، مجله مدل‌سازی پیشرفته ریاضی، دوره ۴، شماره ۲، پاییز و زمستان ۱۳۹۳، (۱۳۹۴) ۱۰۱-۱۱۷.
3. Akaike H., "Information theory and an extension of the maximum likelihood principle", in Petrov, B.N.; Csáki, F., 2nd International Symposium on Information Theory, Tsahkadsor, Armenia, USSR, September 2-8, 1971, Budapest (1973) 267-281.
4. Bender R., Augustin T., Blettner M., "Generating survival times to simulate Cox proportional hazards models", *Statistics in Medicine*, 24 (2005) 1713-1723.
5. Clayton D.G., Cuzik J., "Multivariate generalization of the proportional hazards model", *Journal of the Royal Statistical Society, Series A*, 148 (1985) 82-117.
6. Cook A., Gold D., Li Y., "Spatial Cluster Detection for Longitudinal Outcomes using Administrative Regions" *Communications in Statistics Theory and Methods*, 42 (2013) 2105-2117.
7. Cook A.J., Li Y., "Notes on Permutation Tests for Spatial Cluster Detection with Censored Outcome Data", *Biometrics*, 64 (2008) 1289-1292.
8. Cox D.R., "Regression models and life-tables", *Journal of the Royal Statistical Society, Series B*, 34 (1972) 187-220.

9. Darmofal D., "Bayesian Spatial Survival Models for Political Event Processes", *American Journal of Political Science*, 53 (2009) 241-257.
10. Fleming T.R., Harington D.P., "Counting Processes and Survival Analysis", Wiley, New York (1991).
11. Gutierrez R.G., "Parametric frailty and shared frailty survival models", *Stata Journal*, 2 (2002) 22-44.
12. Henderson R., Shimakura S., Gorst D., "Modeling Spatial Variation in Leukemia Survival Data", *Journal of the American Statistical Association (Taylor and Francis)*, 97 (2002) 965-972.
13. Huang L., Kulldorff M., Gregorio D., "A Spatial Scan Statistic for Survival Data", *Biometrics*, 63 (2007) 109-118.
14. Li Y., Lin X., "Semiparametric Normal Transformation Models for Spatially Correlated Survival Data", 101 (2006) 591-603.
15. Li Y., Ryan L., "Modeling Spatial Survival Data Using Semiparametric Frailty Models", *Biometrics*, 58 (2002) 287-297.
16. Vaupel J.W., Manton K.G., Stallard E., "The impact of heterogeneity in individual frailty on the dynamics of mortality", *Demography*, 16 (1979) 439-454.
17. Zhao L., "Mixtures of Polya Trees for Flexible Spatial Survival Modeling", ProQuest, UMI Dissertation Publishing (2011).
18. Zhao L., Hanson T.E., Carlin B.P., "Mixtures of Polya Trees for Flexible Spatial Frailty Survival Modeling", *Biometrika*, 96 (2009) 263-276.