

تحلیل داده‌های بقای دارای سانسور وابسته با استفاده از مدل خطرهای متناسب کاکس بر اساس تابع مفصل

وحید قوامی قنبرآبادی^۱ - محمدتقی شاکری^۱ - دکتر حبیب‌الله اسماعیلی^۲
دکتر هادی جباری نوقابی^۳

^۱ دانشگاه علوم پزشکی مشهد

^۲ دانشگاه علوم پزشکی مشهد، دانشکده بهداشت

^۳ دانشگاه فردوسی مشهد، دانشکده علوم ریاضی

چکیده:

در مدل خطرهای متناسب کاکس، یکی از فرض‌های اساسی برقراری فرض استقلال میان زمان سانسور و زمان پیشامد می‌باشد. در مطالعات بالینی زمانی که سانسور به دلایلی همچون خروج فرد از مطالعه یا مخاطره رقابتی رخ می‌دهد، همواره نگرانی در خصوص اعتبار نتایج مبتنی بر فرض استقلال سانسورها وجود دارد؛ جهت بررسی فرض استقلال به معرفی تعمیمی از مدل خطرهای متناسب کاکس با استفاده از تابع مفصل می‌پردازیم. این مدل امکان انجام آنالیز حساسیت، جهت بررسی فرض استقلال را فراهم می‌نماید. با استفاده از شبیه‌سازی ابتدا مناسب بودن مدل ارائه شده اثبات گردید سپس تاثیرات عدم برقراری فرض استقلال بین زمان سانسور و پیشامد بر روی برآورد پارامترها بررسی گردید. کلیه محاسبات مربوط به تعمیم تابع درست‌نمایی و برآورد پارامترها و شبیه‌سازی با استفاده از نرم‌افزار R انجام شد. با استفاده از شبیه‌سازی مشخص گردید که عدم برقراری فرض استقلال می‌تواند باعث نتایج اریب در مدل خطرهای متناسب کاکس گردد و میزان اریبی با میزان وابستگی میان سانسور و پیشامد ارتباط مستقیم دارد. در ادامه با استفاده از این مدل، برقراری فرض استقلال زمان سانسور و پیشامد در مجموعه داده‌های واقعی مربوط به بیماران دارای سرطان پستان مورد بررسی قرار گرفت.

کلمات کلیدی مدل خطرهای متناسب کاکس، آنالیز حساسیت، تابع مفصل، خطرات رقابتی، سرطان پستان

۱ مقدمه

در مطالعات مربوط به تحلیل بقا، یکی از روش‌های متداول برای پیش‌بینی احتمال بقا، استفاده از مدل خطرهای متناسب کاکس می‌باشد؛ در این روش یکی از فرض‌های اساسی، استقلال میان زمان سانسور و زمان پیشامد مورد نظراست (۱). در مطالعات بالینی زمانی که یک سانسور به دلایلی همچون کناره‌گیری بیمار^۱ از مطالعه یا یک مخاطره رقابتی^۲ رخ می‌دهد، همواره نگرانی در خصوص اعتبار نتایج مبتنی بر فرض استقلال سانسورها وجود دارد (۲). یکی از راهکارهای بررسی وجود فرض استقلال انجام آنالیز حساسیت است (۳). آنالیز حساسیت، یک مطالعه در جهت بررسی نتایج یک روش آماری، در صورت تغییر مفروضات اساسی آن روش آماری است. حال اگر بتوان مدل خطرهای متناسب کاکس را به نحوی تعمیم داد که میزان وابستگی میان سانسور و پیشامد در مدل لحاظ گردد، آن گاه می‌توان با فرض کردن درجات خاصی از وابستگی نتایج حاصل بر روی برآورد پارامترها را بررسی نمود و به عبارتی یک آنالیز حساسیت انجام داد. یکی از روش‌های انجام آنالیز حساسیت استفاده از تابع مفصل می‌باشد (۲). زینگ و کلین^۳ (۱۹۹۵) یک تابع مفصل را برای توزیع توام مخاطرات رقابتی فرض نموده و توزیع‌های حاشیه‌ای آن را به روش ناپارامتری و تحت تابع مفصل مذکور برآورد نمودند (۴). هانگ و چنگ^۴ (۲۰۰۸) تابع مفصل ارایه شده توسط زینگ و کلین (۱۹۹۵) را جهت انجام یک آنالیز حساسیت برای مدل خطرهای متناسب کاکس تعمیم دادند (۲). در این تحقیق ما با استفاده از نتایج مطالعات فوق به معرفی روشی جهت بررسی فرض استقلال زمان سانسور و زمان پیشامد در داده‌های بقا می‌پردازیم.

۲ روش کار

با توجه به استفاده از تابع مفصل در ابتدا به معرفی تابع مفصل می‌پردازیم. تابع مفصل، تابعی است که تابع توزیع چند متغیره چند بعدی را به توابع حاشیه‌ای یک بعدی آن مرتبط می‌سازد (۵). ما در این تحقیق از توابع مفصل دو بعدی استفاده می‌نماییم؛ از آنجاییکه انتخاب نادرست تابع مفصل سبب آریبی ناچیز در برآورد پارامترها می‌گردد لذا ما در این تحقیق از تابع مفصل فرانک^۵ به دلیل خواص آن از جمله امکان تغییر تاو کندال در بین مقادیر ۱- تا ۱+ استفاده نمودیم (۲).

^۱ patient withdrawal
^۲ Competing risk
^۳ Zheng and Klein
^۴ Huang and Zhang,
^۵ Frank copula

تابع مفصل فرانک: فرض کنید که u و v در بازه $[0, 1]$ باشند و α پارامتر تابع مفصل باشد که درجه وابستگی را نشان می دهد. حال تابع مفصل فرانک به صورت زیر تعریف می شود:

$$C(u, v; \alpha) = \log_{\alpha} \left\{ 1 + \frac{(\alpha^u - 1)(\alpha^v - 1)}{\alpha - 1} \right\}, \alpha > 0, \alpha \neq 1. \quad (1)$$

تبدیلات میان τ کندال و پارامتر وابستگی α در تابع مفصل قابل محاسبه می باشد. ارتباط میان α و τ در تابع مفصل فرانک به صورت زیر است (۵):

$$\tau = 1 + 4(-\log \alpha)^{-1} \{D_1(-\log \alpha) - 1\}, \alpha > 0, \alpha \neq 1. \quad (2)$$

$D_k(X)$ (تابع دیبای^۶) به صورت زیر تعریف می شود که k یک عدد صحیح مثبت است:

$$D_k(x) = \frac{k}{x^k} \int_0^x \frac{t^k}{e^t - 1} dt. \quad (3)$$

۳ مدل های با زمان پیشامد و سانسور آگاهی بخش

فرض می کنیم T زمان پیشامد، C زمان سانسور آگاهی بخش و S زمان سانسور بدون اطلاع باشند. زمان مشاهده شده یا به عبارتی زمان بقا را به صورت $X = \min(T, C, S)$ نشان می دهیم. همچنین دو تابع نشانگر $\delta_1 = I(X = T)$ و $\delta_2 = I(X = C)$ را تعریف می نماییم. دو متغیر مستقل Z و W با ابعاد به ترتیب $p \times n$ و $q \times n$ را که با زمان های سانسور آگاهی بخش و زمان پیشامد مرتبط هستند، در نظر می گیریم. حال توابع خطر برای زمان پیشامد T_1 و زمان سانسور وابسته C_1 به ترتیب به صورت زیر تعریف می گردند:

$$\lambda_i(t|Z_i, W_i) = \lambda_0(t) \exp(Z_i' \beta); \quad \Psi_i(t|Z_i, W_i) = \Psi_0(t) \exp(W_i' \beta_c) \quad (4)$$

که در آن ها، $\lambda_0(t)$ و $\Psi_0(t)$ توابع خطر پایه نامشخص، β و β_c پارامترهای مجهول به ترتیب با ابعاد $1 \times p$ و $1 \times q$ می باشند.

توابع توزیع تجمعی حاشیه ای برای T_i و C_i به صورت زیر تعریف می گردند:

$$F_i(t|Z_i, W_i) = 1 - S_i(t|Z_i, W_i) = 1 - \exp\{-\Lambda_0(t) \exp(Z_i' \beta)\}; \quad (5)$$

$$G_i(t|Z_i, W_i) = 1 - R_i(t|Z_i, W_i) = 1 - \exp\{-\Psi_0(t) \exp(W_i' \beta_c)\} \quad (6)$$

^۶ Debye function

که $S_i(t)$ و $R_i(t)$ توابع بقا به ترتیب برای T و C و $A_0(t)$ و $\Psi_0(t)$ توابعی خطر تجمعی پایه به ترتیب برای T و C می‌باشند. حال برای یک تابع مفصل مفروض مانند C با پارامتر α ، تابع توزیع توأم تجمعی T_i و C_i به صورت زیر قابل مدل‌بندی است:

$$H_i(t, c) = Pr(T_i \leq t, C_i \leq c) = C\{F_i(t), G_i(c); \alpha\} \quad (V)$$

به‌طور کلی تا زمانی که فرد پیشامد مورد نظر را تجربه نکرده باشد دارای جرم واحد می‌باشد، اما به محض مشاهده پیشامد، تمام جرم خود را از دست می‌دهد. بدیهی است در حالت برقراری فرض استقلال میان سانسور و پیشامد، در صورتی که فرد سانسور گردد جرمش به صورت یکنواخت بر روی تمام نقاط پیشامد بعد از خود توزیع می‌گردد و این در حالی است که در حالت وجود سانسور وابسته این توزیع جرم به صورت یکنواخت انجام نمی‌شود. سوالی که مطرح می‌باشد اینست که در حالت سانسور وابسته جرم این افراد به چه صورت بر روی نقاط پیشامد بعد از خود توزیع می‌گردد؟

۴ توابع احتمال شرطی بقا

فرض کنید، x_i که $i = 1, \dots, n$ نقاط زمانی مرتب شده به صورت صعودی و بدون وجود گره باشند. اگر فرد i ام در زمان x_i ، $i = 1, \dots, n$ سانسور شده باشد (سانسور آگاهی بخش) سپس برای زمان پیشامد $x_j > x_i$ می‌توانیم احتمال این که فرد i ام در زمان x_j ازدست رفته باشد را محاسبه نماییم. زینگ و کلین (۱۹۹۴) با انجام محاسباتی تحت فرض تابع توزیع توأم مشخص شده توسط تابع مفصل C برای $t > x_i$ نشان دادند که:

$$P_i(t) = Pr(T_i \geq t | T_i \geq x_i, C_i = x_i) = \frac{Pr(T_i \geq t, C_i = x_i)}{Pr(T_i \geq x_i, C_i = x_i)} = \frac{1 - C_\nu\{F_i(t), G_i(x_i); \alpha\}}{1 - C_\nu\{F_i(x_i), G_i(x_i); \alpha\}} \quad (۸)$$

که در آن

$$C_\nu(a, b, \alpha) = \frac{\partial C(u, \nu; \alpha)}{\partial \nu} \Big|_{(u, \nu) = (a, b)} \quad (۹)$$

حال تابع احتمال بقای شرطی (۹) را با $P_i(t)$ نشان می‌دهیم (۶). جرمی که شخص سانسور شده i ام در زمان x_j ازدست می‌دهد را با $D_i(x_j)$ نمایش می‌دهیم که x_j زمان پیشامد فرد i ام می‌باشد و $x_j > x_i$. حال داریم:

$$D_i(x_j) = P_i(x_{j-1}) - P_i(x_j) \quad (۱۰)$$

برای سانسور آگاهی بخش نیز چون β_c در $P_i(t)$ دخیل است، لذا بایستی این احتمال مجدداً برآورد گردد. جهت درک بهتر، پیشامدها و سانسورها را در چارچوب مخاطرات رقابتی تعریف می‌نماییم؛ لذا می‌توانیم سانسور وابسته را به عنوان پیشامد در نظر گرفته و توابع مشابه را برآورد کنیم. به طور خاص برای فرد i ام که در زمان x_i پیشامد را مشاهده نموده است، می‌توانیم احتمال این که فرد در زمان $x_i > c$ به صورت وابسته سانسور گردد را محاسبه نماییم. برای $x_i > c$ داریم:

$$Q_i(c) = Pr(C_i \geq c | C_i \geq x_i, T_i = x_i) = \frac{Pr(C_i \geq c, T_i = x_i)}{Pr(C_i \geq x_i, T_i = x_i)} = \frac{1 - C_u\{F_i(x_i), G_i(c); \alpha\}}{1 - C_u\{F_i(x_i), G_i(x_i); \alpha\}} \quad (11)$$

که در آن

$$C_u(a, b, \alpha) = \left. \frac{\partial C(u, v; \alpha)}{\partial u} \right|_{(u,v)=(a,b)} \quad (12)$$

فرض کنیم که $E_i(x_j)$ بیان‌کننده جرمی است که شخص i ام که در زمان x_i پیشامد را مشاهده نموده است در زمان سانسور وابسته x_j از دست بدهد. داریم:

$$E_i(x_j) = Q_i(x_{j-1}) - Q_i(x_j) \quad (13)$$

۵ توابع درست‌نمایی جزئی

برای هر فرد با سانسور آگاهی بخش، فرض می‌کنیم که سهم آن در تابع درست‌نمایی به صورت تدریجی در هر نقطه‌ای که پیشامد مشاهده می‌شود، کاهش یافته و به وسیله $P_i(x_j)$ و $D_i(x_j)$ نشان داده می‌شود. حال تابع درست‌نمایی جزئی تعمیم‌یافته کاکس را برای پیشامد به صورت زیر تعریف می‌کنیم:

$$L^{(T)}(\beta) = \prod_{j=1}^n \prod_{i=1}^j \left\{ \frac{P_i(x_j) \exp(Z_i' \beta)}{\sum_{k=1}^n P_k(x_j) \exp(Z_k' \beta)} \right\}^{D_i(x_j)} \quad (14)$$

حال جهت خوش‌تعریف شدن تساوی (۵) برخی تعدیلات را اعمال می‌نماییم. اولاً برای $k \leq j$ مقدار $P_k(x_j)$ را برابر یک در نظر می‌گیریم؛ ثانیاً برای $j > i$ داریم $P_i(x_j) = 0$. ثالثاً با توجه به تعریف $D_i(x_j)$ برای فرد i ام که پیشامد مورد نظر را تجربه کرده است برای $j > i$ فرض می‌کنیم: $D_i(x_j) = 0$ و $D_i(x_i) = 1$.

حال تابع درستنمایی جزئی تعمیم یافته کاکس را برای حالتی که سانسور وابسته را به عنوان پیشامد در نظر می گیریم، تعریف می کنیم:

$$L^{(C)}(\beta_c) = \prod_{j=1}^n \prod_{i=1}^j \left\{ \frac{Q_i(x_j) \exp(W'_i \beta_c)}{\sum_{k=1}^n Q_k(x_j) \exp(W'_k \beta_c)} \right\}^{E_i(x_j)} \quad (15)$$

جهت خوش تعریف شدن تساوی (۶) بصورت مشابه اولاً برای $k \leq j$ مقدار $Q_k(x_j)$ را برابر یک در نظر می گیریم؛ ثانیاً این که برای یک فرد i ام که به صورت وابسته سانسور شده است برای $i > j$ ما $Q_i(x_j) = 0$ فرض می نماییم. ثالثاً با توجه به تعریف $E_i(x_j)$ برای فرد i ام که به صورت وابسته سانسور شده است، فرض می کنیم: $E_i(x_i) = 1$ و $E_i(x_j) = 0$. همچنین برای فرد i ام که به صورت مستقل سانسور شده است برای $j \leq i$ داریم $P_i(x_j) = Q_i(x_j) = 1$ و برای $j > i$ نیز داریم $P_i(x_j) = Q_i(x_j) = 0$. همچنین در سانسورهای مستقل برای تمام j ها فرض می کنیم $E_i(x_j) = D_i(x_j) = 0$. حال بایستی β و β_c از طریق ماکسیمم ساختن تابع درستنمایی جزئی توام تعمیم یافته زیر برآورد گردند:

$$L(\beta, \beta_c) = L^{(T)}(\beta) L^{(C)}(\beta_c) \quad (16)$$

از آنجایی که توابع درستنمایی جزئی توام تعمیم یافته (۵) و (۶) شامل مقادیر مجهولی همچون $E_i(x_j), D_i(x_j), Q_i(x_j), P_i(x_j)$ هستند؛ لذا جهت برآورد پارامترها از روش تکرارهای عددی استفاده می شود.

قابل ذکر است که کلیه محاسبات مربوط به برآورد پارامترها و نیز شبیه سازی که در ادامه خواهد آمد از نرم افزار R.2.12.1 انجام شده است.

۶ یافته ها

در این بخش ابتدا با استفاده از شبیه سازی به بررسی دقت مدل پرداخته و در ادامه مدل تعمیم یافته ارائه شده را جهت بررسی فرض استقلال زمان سانسور و زمان پیشامد بر روی داده های واقعی مربوط به بیماران دارای سرطان پستان استفاده می کنیم.

۱.۶ شبیه سازی

در ابتدا تعداد ۳۰۰ مجموعه داده بقا را که شرط استقلال زمان پیشامد و زمان سانسور در آن ها برقرار نبوده، به حجم ۲۰۰ بر اساس معادلات مشخص بقا با پارامترهای معلوم تولید می نماییم. جهت این امر از تابع مفصل فرانک با مقدار پارامتر معلوم

۰.۰۰۰۰۰۰۰۱۲۵۸ که معادل τ کندال ۰.۸ می‌باشد، استفاده شده است. این مجموعه داده‌ها را با استفاده از مدل خطرهای متناسب کاکس، مدل معرفی شده در این تحقیق با تابع مفصل فرانک و میزان وابستگی های معادل T کندال ۰.۸، ۰.۵ و ۰.۲ مدل بندی می‌نماییم.

۲.۶ نحوه تولید متغیرهای وابسته زمان پیشامد و سانسور آگاهی بخش

جهت تولید داده‌ها دو متغیر مستقل یکی از توزیع برنولی با احتمال ۰.۵ و دیگری از توزیع یکنواخت در بازه $(-۵, ۵)$ ایجاد می‌نماییم. سپس با استفاده از توزیع وایل، توزیع‌های حاشیه‌ای مربوط به زمان‌های پیشامد و سانسور آگاهی بخش را به صورت زیر به دست می‌آوریم؛ برای زمان t داریم:

$$S(t) = \exp\{-0.5t^2 \exp(\beta_1 B + \beta_2 A)\} \quad (17)$$

$$R(t) = \exp\{-0.5t \exp(\beta_{1c} B + \beta_{2c} A)\} \quad (18)$$

که در آن $\beta_1 = 0.5$, $\beta_2 = 0.1$, $\beta_{1c} = 0.6$, $\beta_{2c} = 0.3$

جهت تولید مشاهدات وابسته از تابع مفصل فرانک با $\tau = 0.8$ استفاده نمودیم که برای این امر می‌توان از دستورات کتابخانه Copula در نرم‌افزار R2.12.1 استفاده نمود.

علاوه بر این متغیر تصادفی S_i از توزیع یکنواخت $(0, 5)$ را جهت زمان سانسور مستقل تولید می‌نماییم. بنابراین با توجه به تعریف سانسور راست مشاهدات به صورت $X_i = \min(T_i, C_i, S_i)$ تعریف می‌گردند. همان‌طور که ملاحظه می‌شود عدم برقراری فرض استقلال در مدل خطرهای متناسب کاکس سبب ارزیابی شدید نتایج از مقادیر حقیقی می‌شود. در جدول ۱ نتایج حاصل از این مرحله آورده شده است.

ملاحظه می‌گردد که در صورت استفاده از تابع مفصل فرانک با $\tau = 0.8$ پارامترهای برآورد شده به مقدار واقعی آن‌ها بسیار نزدیک است. ضمناً میزان مقادیر MSE نسبت به استفاده از مدل خطرهای متناسب کاکس کوچکتر می‌باشد. لذا این نتایج نشان‌دهنده دقت روش پیشنهادی در برآورد پارامترها است. همچنین ملاحظه می‌گردد در صورتی که درجه وابستگی (τ کندال) به اشتباه اختصاص یافته ($\tau = 0.5$ و $\tau = 0.2$) و مورد استفاده قرار گیرد شاهد ارزیابی قابل ملاحظه‌ای در نتایج مربوط به برآورد پارامترها خواهیم بود به طوری که هرچه میزان انحراف از وابستگی حقیقی بیشتر باشد، میزان ارزیابی از مقادیر حقیقی پارامترها نیز بیشتر خواهد بود.

جدول ۱: مقایسه نتایج حاصل از استفاده از مدل‌های مختلف

پارامتر	مقدار حقیقی	مقدار پارامتر در مدل کاکس	\sqrt{MSE} مدل کاکس	برآورد پارامتر در مدل کاکس	\sqrt{MSE} مدل کاکس	مقدار پارامتر در مدل فرانک	\sqrt{MSE} مدل فرانک	برآورد پارامتر در مدل فرانک	\sqrt{MSE} مدل فرانک	مقدار پارامتر در مدل وول	\sqrt{MSE} مدل وول	برآورد پارامتر در مدل وول	\sqrt{MSE} مدل وول
β_1	۰.۵	۰.۲۲۶	۰.۲۸	۰.۴۷۱	۰.۳۱	۰.۳۹۸	۰.۱۵۲	۰.۳۳۷	۰.۱۵۵	۰.۳۳۷	۰.۱۵۵	۰.۳۳۷	۰.۱۵۵
β_2	۰.۱	۰.۰۵۴	۰.۰۵۶	۰.۱۰۵	۰.۰۰۹	۰.۰۷۶	۰.۰۲۷	۰.۰۶۳	۰.۰۴۲	۰.۰۶۳	۰.۰۴۲	۰.۰۶۳	۰.۰۴۲
β_{c1}	۰.۶	۰.۸۴۱	۰.۲۵	۰.۶۱۸	۰.۰۱۵	۰.۶۷۴	۰.۰۸۷	۰.۷۷۳	۰.۱۸۴	۰.۷۷۳	۰.۱۸۴	۰.۷۷۳	۰.۱۸۴
β_{c2}	۰.۳	۰.۳۴۱	۰.۰۵	۰.۲۹۱	۰.۰۰۹	۰.۳۱۷	۰.۰۱۸	۰.۳۲۴	۰.۰۲۵	۰.۳۲۴	۰.۰۲۵	۰.۳۲۴	۰.۰۲۵

۷ تحلیل داده‌های بیماران سرطان پستان

پرونده مربوط به ۹۹۶ بیمار مبتلا به سرطان پستان که طی سال‌های ۱۳۸۰ لغایت ۱۳۸۸ جهت درمان به بخش آنکولوژی بیمارستان‌های قائم (عج) و امید مشهد مراجعه نموده‌اند مورد بررسی قرار گرفت که ۵۵۹ بیمار شرایط ورود به مطالعه را پیدا کردند. قابل ذکر است که شرایط خروج از مطالعه عبارت بودند از عدم جراحی بیمار، عدم اتمام دوره درمان به طور کامل و قرار داشتن در مرحله IV بیماری. شروع مطالعه برای افراد زمان جراحی ایشان است. میانگین مدت زمان پیگیری افراد ۳۲.۹ ماه می‌باشد. هدف اصلی این مطالعه بررسی تاثیر BMI بر بقای بدون بیماری بیماران دارای سرطان پستان پس از درمان ایشان است. از آنجایی که برخی عوامل همچون مرحله بیماری و سن بیمار از عوامل موثر شناخته شده بر روی بقای بیماران می‌باشد لذا این متغیرها نیز به مدل اضافه گردیده است. از تعداد ۵۵۹ بیمار وارد شده در مطالعه تعداد ۱۵۱ نفر عود داشته‌اند که به عنوان پیشامد مورد نظر در نظر گرفته می‌شود. از تعداد ۴۰۸ نفر باقیمانده تعداد ۵۵ نفر که آخرین وضعیت بالینی ایشان از نظر پزشک مساعد نبوده است از ادامه همکاری با پزشک خودداری نموده‌اند که آنها را به عنوان سانسور آگاهی بخش در نظر می‌گیریم. تعداد ۳۵۳ نفر باقیمانده نیز در آخرین وضعیت مشاهده شده دارای شرایط مساعدی بوده‌اند که مشاهدات مربوط به این افراد به عنوان سانسور بدون اطلاع یا سانسور مستقل در نظر گرفته می‌شود. از لحاظ سنی (age groups) بیماران همانند مطالعه مورد اشاره در منبع (۸) به دو گروه زیر ۳۵ سال، و بزرگتر یا مساوی ۳۵ سال تقسیم شدند. از لحاظ شاخص توده بدنی (BMI) نیز بیماران بر اساس معیار معرفی شده توسط سازمان بهداشت جهانی (WHO) که افراد با BMI بالاتر از ۳۰ را به عنوان افراد چاق معرفی می‌کند به دو گروه با BMI کمتر از ۳۰ و بزرگتر یا مساوی ۳۰ تقسیم شدند (۹).

با توجه به نظر پزشک متخصص، بیماران دارای مراحل I و IIA و IIB جز مرحله اولیه و بیماران دارای مراحل IIIB، IIIA و IIIC جز مرحله پیشرفته بیماری قرار گرفتند که این متغیر را با stage نشان داده‌ایم.

جهت بررسی فرض استقلال میان زمان‌های پیشامد و زمان‌های مربوط به سانسور آگاهی بخش از مدل تعمیم‌یافته و با استفاده از تابع مفصل فرانک و با τ کندال با مقادیر صفر و ۰.۲ و ۰.۵ و ۰.۸ استفاده شد.

برای برآورد انحراف معیار پارامترها از خودگردان سازی^۷ با ۱۰۰ بار تکرار استفاده شده است. نتایج در جدول ۲ قابل مشاهده است.

جدول ۲: نتایج مربوط به برآورد پارامترها

$\tau = 0.8$	$\tau = 0.5$	$\tau = 0.2$	$\tau = 0$		متغیر
-۰.۳۳۸ ۰.۱۸۹ -۰.۷۰۸ ۰.۰۳۲ ۰.۰۷۳	-۰.۳۶۱ ۰.۱۹۲ -۰.۷۳۷ ۰.۰۱۵ ۰.۰۶۰	-۰.۳۵۲ ۰.۲۰۹ -۰.۷۶۲ ۰.۰۵۸ ۰.۰۹۲۱	-۰.۳۷۴ ۰.۲۴۹ -۰.۸۶۲ ۰.۱۱۴ ۰.۱۳۳	برآورد پارامتر انحراف پایین بازه اطمینان کران پایین بازه اطمینان کران بالای بازه اطمینان مقدار P	گروه سنی
۰.۲۵۷ ۰.۱۶۸ -۰.۰۷۲ ۰.۷۸۰ ۰.۱۲۶	۰.۲۷۹ ۰.۱۶۱ -۰.۰۳۶ ۰.۵۹۴ ۰.۰۸۳۱	۰.۲۹۲ ۰.۱۷۲ -۰.۰۴۵ ۰.۶۲۹ ۰.۰۸۹	۰.۲۸۵ ۰.۱۸۱ -۰.۰۶۹ ۰.۶۴۰ ۰.۱۱۵	برآورد پارامتر انحراف معیار پارامتر کران پایین بازه اطمینان کران بالای بازه اطمینان مقدار P	BMI
۰.۴۶۹ ۰.۱۲۷ ۰.۲۲۰ ۰.۷۱۸ ۰.۰۰۰۲	۰.۴۹۲ ۰.۱۳۶ ۰.۲۲۵ ۰.۷۵۸ ۰.۰۰۰۲	۰.۵۲۷ ۰.۱۴۹ ۰.۲۳۵ ۰.۸۱۹ ۰.۰۰۰۴	۰.۵۵۶ ۰.۱۶۴ ۰.۲۳۴ ۰.۷۲ ۰.۰۰۱	برآورد پارامتر انحراف معیار پارامتر کران پایین بازه اطمینان کران بالای بازه اطمینان مقدار P	مرحله بیماری

با توجه به نتایج به دست آمده می‌توان برآوردهای حاصل از مدل خطرهای متناسب کاکس را بر مبنای فرض استقلال معتبر دانست. در ادامه آزمون متناسب بودن خطرها را بر اساس باقیمانده‌های وزنی انجام می‌دهیم (۱۰). جهت انجام این آزمون از دستور cox.zph در نرم‌افزار R استفاده نمودیم که فرض متناسب بودن خطرها^۸ در مورد هیچکدام از متغیرها در سطح ۰.۰۵ رد نشد. در ادامه با استفاده از روش گام به گام پیشرونده مبتنی بر آزمون والد به انتخاب متغیرهای مستقل می‌پردازیم. در ابتدا متغیرهای گروه سنی (agegroup)، سطح شاخص توده بدن (BMI) مرحله بیماری

^۷ Bootstrapping

^۸ Proportional hazards assumption

(stage) مدل را به همراه اثرات متقابل آن‌ها به مدل وارد می‌نماییم. بر اساس نتایج به دست آمده، تنها متغیر مرحله بیماری در مدل باقی می‌ماند.

متغیر	برآورد (B)	خطای استاندارد	آماره والد	درجه آزادی	P - Value	Exp(B)
مرحله بیماری	۰.۵۳۹	۰.۱۶۳	۱۰.۸۸۱	۱	۰.۰۰۱	۱.۷۱۴

جدول ۳: متغیرهایی که بر اساس روش گام به گام پیشرونده، در مدل باقی می‌مانند

با توجه به نتایج فوق بر اساس مدل خطرهای متناسب کاکس تنها اثر مرحله بیماری در بقای بیماران معنی‌دار است به طوری که خطر عود بیماری در مرحله پیشرفته ۱.۷ برابر خطر عود در بیماران است که در مرحله اولیه بیماری قرار داشته و مورد درمان قرار گرفته‌اند.

۸ بحث و نتیجه‌گیری

با استفاده از توابع مفصل، مدل خطرهای متناسب کاکس را به نحوی تعمیم دادیم که امکان لحاظ نمودن ارتباط میان سانسور و پیشامد در مدل وجود داشته باشد. با استفاده از شبیه‌سازی علاوه بر اطمینان از دقت مدل، نشان دادیم که وابستگی میان زمان پیشامد و سانسور سبب آریبی در برآورد پارامترها گشته و میزان این آریبی با میزان وابستگی ارتباط مستقیم دارد. در مجموعه داده‌های مربوط به بیماران دارای سرطان پستان با روش ارایه شده فرض استقلال مورد تایید قرار گرفت. لذا با توجه به برقراری فرض استقلال با استفاده از مدل خطرهای متناسب کاکس به تحلیل مجدد اطلاعات پرداختیم. نتایج حاصل از استفاده از مدل خطرهای متناسب کاکس نشان‌دهنده تاثیر مرحله بیماری در بقای بیماران بود به نحوی که خطر عود بیماری در افرادی که در زمان تشخیص بیماری در مرحله پیشرفته بیماری قرار داشته‌اند ۱.۷ برابر بیماران است که در زمان تشخیص اولیه بیماری در مرحله اولیه بیماری قرار داشته‌اند. بنابراین تشخیص بیماری در مراحل اولیه نقش مهمی در موفقیت درمان و عدم عود بیماری دارد.

منابع

- 1 Kleinbaum D. G, Klein M. Introduction to Survival Analysis. In: Survival Analysis A Self-Learning Text. Second edition. New York. Springer 2005; 1-42.

- 2 Huang X, Zhang N. Regression Survival Analysis with an Assumed Copula for Dependent Censoring: A Sensitivity Analysis Approach. *Biometrics* 2008; 64: 1090-1099.
- 3 Kleinbaum D. G, Klein M. Competing Risks Survival Analysis. In: *Survival Analysis A Self-Learning Text*. Second edition. New York. Springer 2005; 403-407.
- 4 Zheng M. and Klein J. P. Estimates of marginal survival for dependent competing risks based on an assumed copula. *Biometrika* 1995; 82: 127-138
- 5 Nelsen R. B. *An Introduction to Copulas*. New York: Springer, 2006.
- 6 Zheng M and Klein J. P. A self-consistent estimator of marginal survival functions based on dependent competing risk data and an assumed copula. *Communications in Statistics, Part A-Theory and Methods* 1994; 23: 2299-2311.
- 7 Nelsen R. B. Properties of a one-parameter family of bivariate distributions with specified marginals. *Communications in Statistics, Part A: Theory and Methods* 1986; 15: 2377-3285.
- ۸ فضل علیزاده ع، حاتمی الف، بررسی تاثیر کاهش دوره رادیوتراپی بعد از ماستکتومی بر میزان عود موضعی در بیماران مبتلا به سرطان پستان، فصلنامه پژوهشی پژوهنده، ۱۳۸۳؛ دوره ۹، مهر و آبان، شماره ۴۰: صفحات ۲۱۴-۲۰۹.
- ۹ سایت سازمان بهداشت جهانی به آدرس الکترونیکی: <http://apps.who.int/bmi>
- 10 Grambsch P, and Therneau T. Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika* 1994; 81: 515-26.