

# تحلیل بقای بیماران مبتلا به سرطان پستان با استفاده از مدل کاکس تعمیم یافته

مرتضی حاجی حسینی، جواد فردمال، رقیه حسن زاده

مرکز پژوهش دانشجویان، گروه آمار زیستی، دانشکده بهداشت، دانشگاه علوم پزشکی همدان

مرکز تحقیقات مدلسازی بیماری‌های غیرواگیر، گروه آمار زیستی، دانشکده بهداشت، دانشگاه علوم پزشکی همدان

گروه آمار زیستی، دانشکده علوم پزشکی، دانشگاه تربیت مدرس

## چکیده

یکی از پرکاربردترین مدل‌ها در تحلیل فرآیندهای بقا مدل کاکس و حالت‌های توسعه یافته آن می‌باشد. پذیره اساسی مدل کاکس، متناسب بودن خطرها در طول زمان است که در صورت برقرار نبودن آن می‌توان از حالت‌های توسعه یافته مدل کاکس استفاده کرد. در این مقاله قصد داریم کاربرد یکی از حالت‌های توسعه یافته‌ی مدل کاکس در تحلیل بقای بیماران مبتلا به سرطان پستان را مطرح کنیم.

واژه‌های کلیدی: تحلیل بقا، سرطان پستان، مدل کاکس تعمیم یافته.

## ۱ مقدمه

اگر مشاهدات یک فرآیند بقا در بازه‌های زمانی مشخص و بصورت گسسته مانند ویزیت‌های دوره‌ای بیماران دیده شوند، فرآیند را زمان گسسته می‌نامند، در غیر اینصورت فرآیند زمان پیوسته می‌باشد، اگرچه نمی‌توان بصورت پیوسته فرآیند را پیگیری کرد اما این تمایز در انتخاب مدل مناسب مفید است [۳]. در مطالعات پزشکی زمان بقا می‌تواند به صورت پیشرفت بیماری، پاسخ به درمان، عود و یا مرگ تعریف شود. علاوه بر آن، در مطالعه برخی بیماری‌ها مانند سرطان‌ها، زمان بقا می‌تواند بیانگر زمان از شروع درمان تا دریافت پاسخ یا مدت زمان

بطور کلی فرآیند تصادفی حادثه‌ای است که در طول زمان رخ می‌دهد و دارای یک جزء تصادفی می‌باشد [۱۱]. فرآیندی مانند ثبت مقادیر متغیر پاسخ در لحظات متفاوتی از زمان نمونه‌ای تجربی از فرآیندهای تصادفی است. بنابراین فرآیند تصادفی شامل متغیر پاسخی مانند  $Y_t$  است که در طول زمان بصورت تصادفی مقادیر مختلفی را شامل می‌شود [۹]. فرآیندی که در آن  $Y_t$  زمان تا رخ دادن حادثه‌ای مورد نظر باشد را فرآیند بقا می‌نامیم [۵].

## ۲ مدل کاکس

مدل کاکس برای مشخص کردن رابطه بین مخاطره حادثه‌ی مورد نظر و متغیرهای کمکی می‌باشد، که به مخاطره پایه حادثه‌ی مورد نظر تحت بررسی بستگی ندارد [۴]. از آنجا که این مدل شامل یک بخش ناپارامتری به نام مخاطره پایه و یک بخش پارامتری می‌باشد، مدل نیمه پارامتری نامیده می‌شود [۷]. در مدل کاکس مخاطره رخداد مورد نظر برای فرد  $i$  ام بصورت زیر محاسبه می‌شود [۷]:

$$\lambda(t, X_i) = \lambda_0(t) e^{(\sum_{k=1}^p \beta_k X_{ik})},$$

که در آن  $\lambda_0$  تابع نامنفی نامشخصی (بخش ناپارامتریک تابع مخاطره) از زمان است که به عنوان مخاطره پایه مشخص می‌شود،  $X_i$  برداری با  $P$  مولفه شامل  $X_i$  ها که نشان‌دهنده‌ی مشخصات مربوط به فرد  $i$  ام و  $\beta$  برداری  $1 \times P$  از ضرایب رگرسیونی ( $\beta_k$ ) می‌باشد. پذیره مهم این مدل، متناسب بودن خطرها<sup>۱</sup> در طول زمان است [۶]. به این معنی که نسبت مخاطره برای دو شخص  $i$  و  $j$  با بردار مشخصات  $X_i$  و  $X_j$  ثابت می‌باشد.

$$\begin{aligned} HR = \frac{\lambda(t, X_i)}{\lambda(t, X_j)} &= \frac{\lambda_0(t) e^{(\sum \beta_k X_{ik})}}{\lambda_0(t) e^{(\sum \beta_k X_{jk})}} \\ &= \frac{e^{(\sum_{k=1}^p \beta_k X_{ik})}}{e^{(\sum_{k=1}^p \beta_k X_{jk})}} = \theta, \end{aligned}$$

به عبارتی دیگر:

$$\lambda(t, X_i) = \theta \lambda(t, X_j).$$

بهبود باشد [۱]. فرض کنید متغیر تصادفی  $T$ ، زمان بقا و تابع‌های  $f(t)$ ،  $F(t)$  و  $S(t)$  به ترتیب تابع چگالی، تابع توزیع و تابع بقا مربوط به آن باشند. در این صورت تابع نرخ مخاطره  $h(t)$  بصورت زیر تعریف می‌شود:

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{Pr(t \leq T < t + \Delta t | T > t)}{\Delta t} = \frac{f(t)}{s(t)}.$$

تابع نرخ مخاطره بیانگر تابع چگالی احتمال رخ دادن حادثه در زمان  $t$  می‌باشد، با این شرط که تا لحظه  $t$  حادثه‌ای رخ نداده باشد. در تحلیل بقا سه تابع  $S(t)$ ،  $f(t)$  و  $h(t)$  هم ارزش هستند؛ به این معنا که با داشتن یکی از آن‌ها می‌توان دو تابع دیگر را بدست آورد. رایج‌ترین مدل برای مشخص کردن اثر متغیرهای مختلف بر رخداد مورد نظر، مدل کاکس است. مدل کاکس اولین بار در سال ۱۹۷۲ توسط کاکس مطرح شد [۱]. در شرایطی که پذیره اصلی این مدل یعنی پذیره متناسب بودن خطرها در طول زمان برقرار نباشد، توان آزمون‌های آماری کاهش می‌یابد [۱]. یکی از روش‌های متداول جایگزین، استفاده از مدل کاکس تعمیم یافته می‌باشد [۱۲]. برقرار نبودن فرض اساسی مدل کاکس به معنای وجود متغیر و/یا متغیرهایی است که اثر یا مقدار آن (ها) در طول زمان تغییر می‌کند [۶]. بنابراین اثر متقابلی از زمان وارد مدل می‌شود و اثر متغیرهای کمکی مدل بر بقای بیماران بررسی می‌شود. در مطالعات پزشکی، متغیرهایی مانند وضعیت عود وجود دارند که به زمان وابسته هستند، در نتیجه پذیره اساسی مدل کاکس برای چنین متغیرهایی برقرار نخواهد بود. لذا در این مقاله به دنبال استفاده از روش‌های جایگزین مانند مدل کاکس تعمیم یافته هستیم.

<sup>۱</sup> Proportional Hazard

پذیره متناسب بودن خطرها را می‌توان با استفاده از

$$X_i(t) = (X_{i1}, X_{i2}, \dots, X_{ip_1}, X_{i1}(t), X_{i2}(t), \dots, X_{ip_2}(t)).$$

چند روش، از جمله روش‌های گرافیکی، نیکویی برازش، بررسی اثر متقابل با زمان و بررسی خطاهای شونفلد بررسی کرد [۷، ۱۳]. همچنین برآورد ضرایب مدل کاکس با استفاده از روش حداکثر درستنمایی جزئی می‌باشد که به دلیل نبودن روش حل صریح و محاسبات پیچیده برای برآورد اثر متغیرهای کمکی از روش‌های تکرار عددی استفاده می‌شود. تابع درستنمایی جزئی مدل کاکس بصورت زیر تعریف می‌شود [۵]:

$$\begin{aligned} HR(t) &= \frac{\lambda(t, X_i(t))}{\lambda(t, X_j(t))} \\ &= e^{[\sum_{k=1}^{p_1} \beta_k [X_{ik} - X_{jk}] + \sum_{k=1}^{p_2} \delta_k [X_{ik}(t) - X_{jk}(t)]]}. \end{aligned}$$

از این روش می‌توان به عنوان روشی برای ارزیابی پذیره متناسب بودن خطرها نیز استفاده کرد، به این صورت که چنانچه ضریب متغیر کمکی وابسته به زمان در مدل معنادار نشد، به معنای مستقل از زمان بودن متغیر مورد بررسی است. به منظور برآورد ضرایب رگرسیونی در مدل کاکس تعمیم یافته از روش حداکثر درستنمایی جزئی تعمیم یافته و برای حل معادلات حاصل از روش‌های عددی مانند نیوتون-رافسون استفاده می‌شود [۸، ۱۰].

در این عبارت منظور از  $R(t_i)$  مجموعه تمام افرادی است که قبل از  $t_i$  سانسور یا مرگ برای آن‌ها اتفاق نیفتاده است.

$$L(\beta) = \prod_{m=1}^n \frac{\exp(\sum_{k=1}^p \beta_k X_{ik})}{\sum_{j \in R(t_i)} \exp(\sum_{k=1}^p \beta_k X_{jk})}$$

### ۳ مدل کاکس تعمیم یافته

برقرار نبودن فرض اساسی مدل کاکس به معنای وجود متغیر و/یا متغیرهایی است که اثر یا مقدار آن‌ها در طول زمان تغییر می‌کند [۶]. زمانی که در مدل مورد بررسی متغیرهای کمکی مستقل از زمان و/یا وابسته به زمان وجود دارند، یکی از گزینه‌های رایج برای تحلیل اطلاعات، از مدل کاکس تعمیم یافته می‌باشد [۷]. فرض کنید  $P_1 + P_2$  متغیر کمکی داریم که  $P_1$  متغیر آن مستقل از زمان و با نمادهای  $X_{P_1}, \dots, X_2, X_1$  و  $P_2$  متغیر دیگر آن وابسته به زمان با نمادهای  $X_{P_2}(t), \dots, X_2(t), X_1(t)$  باشند، آنگاه تابع مخاطره برای فرد  $i$  به صورت زیر نوشته می‌شود:

$$\lambda(t, X_i(t)) = \lambda_0(t) e^{[\sum_{k=1}^{p_1} X_{ik} \beta_k + \sum_{k=1}^{p_2} \delta_k X_{ik}(t)]},$$

### ۴ کاربرد

در این مطالعه اطلاعات مربوط به ۵۲۷ بیمار زن مراجعه کننده به مرکز تشخیصی درمانی دارالایتام مهدیه همدان طی سالهای ۱۳۹۰-۱۳۸۳ با تشخیص اولیه سرطان پستان استفاده شده است. اطلاعاتی شامل سن در زمان تشخیص (کمتر از ۵۰ سال و بزرگتر از ۵۰ سال)، اندازه تومور (کمتر از ۲، بزرگتر از ۵ و بین ۵-۲ سانتی‌متر)، نوع تومور (داکتال، لوبولار و مدولار) و وضعیت عود (دارد، ندارد) می‌باشد که از پرونده پزشکی بیماران و به وسیله چک لیست جمع آوری شده است. اطلاعات کامل مربوط

به داده‌های مورد استفاده در این مطالعه در جای دیگر منتشر شده است [۲]. جدول ۱ خلاصه اطلاعات مربوط به افراد شرکت‌کننده در این مطالعه را نشان می‌دهد (جدول ۱ را در پیوست ملاحظه بفرمایید). میانه بقای گزارش شده در جدول ۱، کوچکترین زمان بقایی است که در آن مقدار تابع بقا برابر یا کمتر از ۰/۵ باشد [۷]. با توجه به اینکه تمام بیماران در ابتدای مطالعه عود نداشته‌اند و در طول پیگیری ۱۲۷ بیمار عود را تجربه کرده‌اند (جدول ۱) و با در نظر گرفتن این نکته که تغییر وضعیت عود برای هر بیمار بر مخاطره مرگ وی اثر دارد، این متغیر در این مطالعه به عنوان متغیر وابسته به زمان در نظر گرفته شده است. همچنین برای بررسی وابسته به زمان بودن متغیرهای دیگر از رابطه همبستگی بین تابعی از زمان بقا و تابعی از خطاهای شونفلد استفاده شده است [۱۳] که بر اساس نتایج این روش متغیرهای سن در زمان تشخیص و تعداد گره لنفاوی فرض متناسب بودن نسبت خطرها در طول زمان را حفظ نکرده بودند. در نهایت مشخص شد که متغیرهای سن در زمان تشخیص و اندازه تومور در سطح خطای ۰/۰۵ بر بقای بیماران مبتلا به سرطان پستان اثر معناداری دارند (جدول ۲ را در پیوست ملاحظه بفرمایید). همچنین مشخص شد که در این مطالعه متغیر وابسته به زمان عود اثر معناداری بر بقای بیماران مبتلا به سرطان پستان ندارد (جدول ۲).

قرار دارند ( $P=0/009$ ). همچنین بیمارانی با اندازه تومور بیش از ۵ سانتی‌متر ۲ برابر بیماران با اندازه تومور کمتر مساوی ۲ سانتی‌متر در معرض خطر مرگ به علت این بیماری هستند. ( $P=0/003$ ) با توجه به این‌که متناسب بودن خطرها، پذیره اساسی مدل کاکس می‌باشد، در شرایط برقرار نبودن این پذیره گزینه‌های مختلفی از جمله مدل کاکس تعمیم یافته وجود دارد که در این مقاله کاربرد این مدل در تحلیل اطلاعات بقا بیماران مبتلا به سرطان پستان که در آن متغیرهای کمکی وابسته به زمان وجود دارد، نشان داده شد.

## مراجع

- [1] Cox, P.R. (1972). Life tables. Wiley Online Library.
- [2] Faradmal, J., Mafi, M., Sadighi Pashaki, A., Karami, M. and Roshanaei, G. (2014). Factors affecting survival in breast cancer patients referred to the darol aitam-e mahdieh center. Zanzan University of Medical Sciences Journal, 22 (93), 105-115.
- [3] Jazwinski, A.H. (2007). Stochastic processes and filtering theory. Courier Corporation.
- [4] Jenkins, S.P. (2005). Survival analysis. Unpublished manuscript, Institute for Social and Economic Research, University of Essex, Colchester, UK.

## ۵ بحث و نتیجه‌گیری

با توجه به نتایج مدل برازش داده شده، بیمارانی که بیش از ۵۰ سال داشتند ۱/۴ (۱/۰۸-۱/۹) برابر بیشتر از بیماران دیگر در معرض خطر مرگ به علت سرطان پستان

- [13] Schonfeld, D. (1982). Partial residuals for the proportional hazards model. *Biometrika*, 69, 51-55.
- [5] Klein, J.P. and Moeschberger, M.L. (2003). *Survival analysis: techniques for censored and truncated data*. Springer Science and Business Media.

### پیوست

- جداول ۱ و ۲ در ذیل ارائه شده اند.
- [6] Klein, M. (2005). *Survival analysis: a self-learning text*. Springer.
- [7] Kleinbaum, D.G. and Klein, M. (1996). *Survival analysis*. Springer.
- [8] Lai, D.C.F., Tung, H.K.K., Wong, M.C.S. and Ng, S. (2010). *Professional Financial Computing Using Excel and VBA*, Wiley.
- [9] Lindsey, J.K. (2004). *Statistical analysis of stochastic processes in time*. Cambridge University Press.
- [10] McLachlan, G. and Krishnan, T. (2007). *The EM algorithm and extensions*. John Wiley and Sons.
- [11] Ross, S.M. (1996). *Stochastic processes*. John Wiley and Sons, New York.
- [12] Saegusa, T. D. and Chen, Y.Q. (2014). Hypothesis testing for an extended cox model with time-varying coefficients. *Biometrics*, 70 (3), 619-628.

جدول ۱: اطلاعات توصیفی و مقایسه میانه بقا براساس متغیرهای کمکی

متغیر	گروه	تعداد	میانه بقا (سال)	آماره آزمون والد	مقدار-P
سن در زمان تشخیص (سال)	بزرگتر مساوی ۵۰	۱۷۵	۷/۳	۶/۵	۰/۰۱
	کوچکتر از ۵۰	۳۵۲	۹/۰۵		
اندازه تومور (سانتی متر)	کمتر مساوی ۲	۳۱۹	۸/۸	۹/۳	۰/۰۰۹
	بین ۲-۵	۱۷۷	۹/۱		
	بزرگتر از ۵	۳۱	۵/۷		
نوع تومور	داکتال	۴۶۸	۸/۸	۰/۴	۰/۸
	لبولار	۳۲	۱۱/۲		
	مدولار	۲۷	NA		
عود	ندارد	۴۰۰	NA	۱۱/۹	۰/۰۰۹
	دور/نزدیک	۱۲۷	۶/۵		

NA: نرم افزار مقدار خاصی را گزارش نداده است.

جدول ۲: بررسی اثر متغیرهای کمکی بر خطر مرگ بیماران

متغیر	گروه	نسبت مخاطره	آماره آزمون والد	مقدار-P
سن در زمان تشخیص (سال)	کوچکتر از ۵۰	مرجع	۲/۵	۰/۰۰۹
	بزرگتر مساوی ۵۰	۱/۴		
اندازه تومور (سانتی متر)	کمتر مساوی ۲	مرجع	۲/۹	۰/۰۰۳
	بین ۲-۵	۱/۲		
	بزرگتر از ۵	۲/۰۱		
نوع تومور	داکتال	مرجع	-۰/۲	۰/۸
	لبولار	۰/۹		
	مدولار	۰/۹		
عود	ندارد	مرجع	۰/۰۷	۰/۹
	دور/نزدیک	۱/۰۲		

طبقه مرجع: طبقه‌ای است که نسبت خطر بدست آمده در طبقات دیگر با آن طبقه مقایسه می‌شوند.