

برآورد واریانس تابع بقاء شرطی مبتنی بر مدل مخاطره متناسب کاکس در حالتی که پراکندگی مشاهدات بسیار زیاد باشد

هادی جباری نوقابی<sup>۱</sup>، مهدی جباری نوقابی<sup>۱</sup>، باقر مقدس زاده براز<sup>۲</sup>

<sup>۱</sup> دانشگاه فردوسی مشهد

<sup>۲</sup> دانشگاه پیام نور مشهد

چکیده: در تحلیل بقاء اغلب علاقه مندی زمان بقاء را که وابسته به برخی متغیرهای توضیحی و زمینه‌ای است را پیش‌بینی نمائیم. راه متداول پیش‌بینی بقاء از روی متغیرهای تبیینی استفاده از مدل ناپارامتری مخاطره متناسب کاکس است. اما در شرایط خاصی از جمله اینکه داده‌ها پراکنده باشند یا در بدهی زیاد رخ دهند؛ به طوری که مشاهدات کسی در مجموعه متغیرهای پیش‌بینی کننده قرار بگیرند، استفاده از مدل مخاطره متناسب کاکس جایز نیست. زیرا این مدل از دستیافت کیدینگ<sup>۱</sup> (۱۹۹۵) که برای میانگین گرفتن روی منحنی‌های بقاء فردی است؛ استفاده می‌نماید. در چنین مواردی باید از برآوردگر دیگری مبتنی بر مدل مخاطره متناسب پیشنهادی کاکس استفاده نمود که ضمن برخورداری از مزایای روش استاندارد (برآورد تابع بقاء شرطی توسط مدل کاکس)، مشکل اخیر را برطرف نماید. نکته قابل توجه برآورد واریانس این برآوردگر جدید است که در این تحقیق به آن می‌پردازیم. مقایسه واریانس این برآوردگر جدید و واریانس برآوردگر استاندارد می‌تواند روش مناسبی برای انتخاب برآوردگر بهینه تابع بقاء شرطی مبتنی بر مدل پیشنهادی کاکس باشد. از این رو در این تحقیق برآوردگر جدید معرفی و واریانس آن محاسبه می‌شود. در پایان واریانس دو برآوردگر برای یک مطالعه کاربردی روی بقاء بیماران مبتلا به سرطان معده مقایسه می‌گردد.

واژه‌های کلیدی: تابع بقاء شرطی، مدل مخاطره متناسب کاکس، برآورد کاپلان مایر، درستیابی جزئی، مخاطره‌های متناسب، سرطان معده

## ۱ مقدمه

یکی از اهداف اصلی آنالیز بقاء به دست آوردن یک برآورد برای تابع بقاء به شرط الگوهای هم‌عامل مشخص است. گرچه استنباط در مورد مدل مخاطره‌های متناسب، خصوصیات اساسی نرخ مخاطره را نادیده می‌گیرد؛ اما از دست دادن تابع بقاء درست همانند توابع بقاء شرطی نامشخص برای انجام برآورد بعدی از کمیت‌ها معمول است. تهیه چنین اطلاعاتی ممکن است به راهنمایی

1) Keiding, N

در مورد تصمیم گرفتن برای یک مفهوم کاربردی کمک کند. در این تحقیق با توجه به شرایط خاصی که ممکن است در عمل بوجود آید که داده‌ها در تعدادهای زیاد رخ دهند یا اینکه پراکندگی مشاهدات به نحوی باشد که داده‌های کمی در مجموعه متغیرهای توضیحی قرار گیرند؛ از برآوردگر جدید مخاطره متناسب تابع بقاء شرطی استفاده می‌شود که در واقع مکملی برای روش استاندارد برآورد تحت مدل پیشنهادی کاکس است. در این تحقیق واریانس برآوردگر جدید برآورد می‌گردد به طوری که این برآورد واریانس برای محاسبه آسان است و تمام خصوصیات معمول مدل کاکس را داراست.

در یک مطالعه بقاء با  $n$  آزمودنی، فرض می‌کنیم  $T_1, T_2, \dots, T_n$  زمان‌های شکست بالقوه باشد و  $C_1, C_2, \dots, C_n$  زمان‌های سانسور برای آزمودنی باشد. فرض می‌کنیم  $X_i = \min(T_i, C_i)$  و  $\delta_i = I(T_i \leq C_i)$  که در آن  $I$  تابع نشانگر می‌باشد. همچنین تعریف می‌کنیم:

$$y_i(t) = I(X_i \geq t), \quad i = 1, 2, \dots, n \quad (1)$$

فرض می‌شود زمان شکست هر آزمودنی مرتبط با هم‌عامل‌های  $Z_i$  یا متغیرهای تبیینی  $Z_i$ ،  $i = 1, 2, \dots, n$  باشد. مدل مخاطره‌های متناسب کاکس<sup>۲</sup> یک شکل ساده برای ارتباط بین تابع مخاطره  $\lambda(\cdot)$  برای یک آزمودنی در زمان  $t$  و مقدار مشاهده شده<sup>۱</sup> یک بردار  $1 \times p$  هم‌عامل  $Z$  همانند

$$\lambda(t|z) = \lambda_0(t) \cdot \exp(\beta'z) \quad (2)$$

در نظر می‌گیرد که  $\lambda_0(t)$  یک تابع مخاطره مبنای ثابت است و  $\beta$  یک بردار از پارامترهای لگاریتم خطر نسبی است. در روش استاندارد، استنباط آماری روی  $\beta$  معمولاً روی ماکزیم‌سازی درستنمایی جزئی کاکس<sup>۳</sup> صورت می‌گیرد.

$$L(\beta) = \prod_{i=1}^n \left[ \frac{\exp(\beta'z_i)}{\sum_{j=1}^n y_j(X_i) \cdot \exp(\beta'z_j)} \right]^{\delta_i} \quad (3)$$

که در آن  $\hat{\beta}$  برآورد درستنمایی جزئی ماکزیم  $\beta$  است. تحت مدل (۲)، احتمال بقاء شرطی

$$S(t|z) = P(T > t | Z = z)$$

اغلب توسط

$$\hat{S}(t|z) = \exp \left\{ -\hat{\Lambda}_0(t) \cdot \exp(\hat{\beta}'z) \right\} \quad (4)$$

(برسلاو<sup>۴</sup>، ۱۹۷۲ و ۱۹۷۴) برآورد می‌شود. که

$$\hat{\Lambda}_0(t) = \sum_{X_i \leq t} \frac{\delta_i}{\sum_{j=1}^n y_j(X_i) \cdot \exp(\beta' z_j)} \quad (5)$$

این برآورد به یک نقطه تنهای Z مربوط می‌شود. یک سؤال طبیعی مطرح است که احتمال بقاء به شرط آن‌که هم‌عامل‌ها به بعضی زیر مجموعه‌های H متعلق باشند؛ چیست؟ مجموعه H ممکن است برای مثال یک گروه سنی، یا یک بازه مشخص از اندازه‌های پیوسته یا ترکیبی از این‌ها باشد. بطور کلی فرض می‌کنیم H یک زیر مجموعه از فضای p-بعدی اقلیدسی باشد. یک امکان برای بکارگیری معادلات (۴) و (۵) در برآورد S(t|z) برای همه zها در H و بنابراین برای میانگین روی یک برآورد توزیع Z مناسب است. اما برای هر زیرمجموعه H، این دست‌یافت می‌تواند سخت باشد. بخصوص وقتی که داده‌ها پراکنده باشند یا در بعدهای زیاد رخ دهند که مشاهدات بسیار کمی در H قرار می‌گیرند. بنابراین احتمالات بقاء به شرط آن‌که  $Z \in H$  مستقیماً برآورد می‌گردند.

استنباط در مورد مدل مخاطره‌های متناسب اساساً روی توزیع هم‌عامل‌های Z به شرط زمان‌های شکست تمرکز می‌کند. که در اینجا به عنوان احتمال آن‌که زمان بقاء بزرگتر از t باشد به شرط آن‌که Z به یک زیرمجموعه H متعلق باشد؛ یعنی

$$P(T > t | Z \in H)$$

بیان می‌شود. توسط قانون بیز این احتمال آشکارا برحسب توزیع شرطی Z در  $T=t$  با توزیع کناری T بیان می‌شود.  
با فرض

$$\pi_i(\beta, t) = y_i(t) \cdot \exp(\beta z_i) / \sum_{j=1}^N y_j(t) \cdot \exp(\beta z_j) \quad (6)$$

در درستی‌هایی جزئی (۳) هر عبارت  $\pi_i(\beta, X_i)$  به احتمال شرطی در زمان  $X_i$  از یک مشاهده شکست معروف است.

تحت مدل (۲)، تابع توزیع شرطی Z به شرط  $T=t$  به طور سازگار توسط

$$\hat{P}(Z \leq z | T = t) = \sum_{(j: Z_j \leq z)} \pi_j(\hat{\beta}, t) \quad (7)$$

برآورد می‌شود.

توسط فرمول بیز و با فرض  $\hat{P}(\cdot) = 1 - \hat{S}(\cdot)$  برآوردگر از چپ پیوسته کاپلان مایر  $F(\cdot)$  باشد.

4) Breslow, N

و  $t_1 < t_2 < \dots < t_k$  زمان‌های شکست متمایز باشند و  $W(t_i)$  اندازه پله  $\hat{F}$  در  $t_i$  باشد؛ چنانچه مشاهده پایانی یک شکست داریم

$$\hat{S}(t|Z \in H) = \frac{\int_t^{\infty} \hat{P}(Z \in H|u) d\hat{F}(u)}{\int_0^{\infty} \hat{P}(Z \in H|u) d\hat{F}(u)} = \frac{\sum_{t_i > t} \hat{P}(Z \in H|t_i) \cdot W(t_i)}{\sum_{i=1}^k \hat{P}(Z \in H|t_i) \cdot W(t_i)} \quad (8)$$

وقتی مشاهده پایانی یک شکست نباشد و  $\sum_{i=1}^k W(t_i) < 1$  داریم

$$\hat{S}(t|Z \in H) = \frac{\sum_{t_i > t} \hat{P}(Z \in H|t_i) \cdot W(t_i) + \hat{P}(Z \in H|T > t_k) \cdot \{1 - \sum_{i=1}^k W(t_i)\}}{\sum_{i=1}^k \hat{P}(Z \in H|t_i) \cdot W(t_i) + \hat{P}(Z \in H|T > t_k) \cdot \{1 - \sum_{i=1}^k W(t_i)\}} \quad (9)$$

که در آن

$$\hat{P}(Z \in H|T > t_k) = \frac{\sum_H y_j(t_k^+)}{\sum_{j=1}^n y_j(t_k^+)} \quad (10)$$

و  $t_k^+$  زمان بلافاصله بعد از زمان  $t_k$  است.

## ۲ برآورد واریانس $\hat{S}(t|Z \in H)$

فرض می‌کنیم مدل مخاطره‌های متناسب (۲) با  $\beta = \beta_0$  را داشته باشیم. برای به دست آوردن واریانس تقریبی عبارت (۹) در هر  $t$ ، این دستیافت لینک<sup>۵</sup> را بکار می‌گیریم که قابل انعطاف است و می‌تواند توسط بسته‌های نرم افزاری استاندارد برای برازش مدل رگرسیون کاکس محاسبه شود.

در منبع پراکندگی در  $\hat{S}(t|Z \in H)$  وجود دارد، یکی توسط برآورد احتمال شرطی بقاء (۹) به شرط  $\beta$  و دیگری بعلت تردید در  $\beta$  ایجاد می‌شود. توسط مرتبه اول بسط سری تیلور داریم:

$$\hat{S}(t|Z \in H) = \hat{S}(t|Z \in H)|_{\beta=\beta_0} + (\hat{\beta} - \beta_0) \frac{\partial \hat{S}(t|Z \in H)}{\partial \beta} \Big|_{\beta=\beta_0} \quad (11)$$

که  $\hat{\beta}$  روی قطعه خط بین  $\beta_0$  و  $\hat{\beta}$  واقع است. می‌توان نشان داد که  $\frac{\partial \hat{S}(t|Z \in H)}{\partial \beta} \Big|_{\beta=\beta_0}$  بطور تقریبی ثابت است و  $\hat{\beta} - \beta_0$  بطور تقریبی مستقل از  $\hat{S}(t|Z \in H)|_{\beta=\beta_0}$  است. بنابراین

5) Link, C. L. (1984)

واریانس  $\hat{S}(t|Z \in H)$  توسط

$$\text{Var} \{ \hat{S}(t|Z \in H) \} \approx \text{Var} \{ \hat{S}(t|Z \in H) | \beta \} + \left\{ \frac{\partial \hat{S}(t|Z \in H)}{\partial \beta} \right\}_{\beta = \bar{\beta}}^T \cdot \text{Var}(\hat{\beta}) \quad (12)$$

تقریب زده می‌شود.

اولین عبارت در تخمین (۱۲)، پراکندگی ناشی از برآورد بقاء شرطی و دومین عبارت، پراکندگی به علت  $\hat{\beta}$  را می‌دهد. برآورد واریانس تقریبی  $\hat{\beta}$  تحت مدل (۲) آسان است. بنابراین همه مسائل باقی مانده برای عبارت دوم سمت راست عبارت (۱۲) محاسبه مشتق جزئی  $\hat{S}(t|Z \in H)$  نسبت به  $\beta$  است. طبق (۹) داریم:

$$\hat{S}(t|Z \in H) = \frac{\sum_{t_i > t} C_i + B_1}{\sum_{i=1}^k C_i + B_1}$$

که در آن  $C_i = \hat{P}(Z \in H | t_i) \cdot W(t_i)$  و  $B_1$  توسط رابطه (۱۳) تعریف می‌شود.

$$\hat{P}(Z \in H | T > t_k) = \left\{ 1 - \sum_{i=1}^k W(t_i) \right\} \quad (13)$$

در نتیجه

$$\frac{\partial \hat{S}(t|Z \in H)}{\partial \beta} = \frac{\left( \sum_{t_i > t} \frac{\partial C_i}{\partial \beta} + \frac{\partial B_1}{\partial \beta} \right) \left( \sum_{i=1}^k C_i + B_1 \right) - \left( \sum_{i=1}^k \frac{\partial C_i}{\partial \beta} + \frac{\partial B_1}{\partial \beta} \right) \left( \sum_{t_i > t} C_i + B_1 \right)}{\left( \sum_{i=1}^k C_i + B_1 \right)^2}$$

اما با فرض  $D_i = \frac{\partial C_i}{\partial \beta}$  برای هر  $i$

$$\frac{\partial \hat{S}(t|Z \in H)}{\partial \beta} = \frac{\left( \sum_{t_i > t} D_i \right) \left( \sum_{t_i \leq t} C_i \right) - \left( \sum_{t_i \leq t} D_i \right) \left( \sum_{t_i > t} C_i + B_1 \right)}{\left( \sum_{i=1}^k C_i + B_1 \right)^2} \quad (14)$$

همچنین طبق تعریف

$$D_i = \frac{\partial C_i}{\partial \beta} = \frac{\partial}{\partial \beta} \hat{P}(Z \in H | t_i) \cdot W(t_i)$$

که طبق (۷)

$$\frac{\partial}{\partial \beta} \hat{P}(Z \in H | t_i) = \frac{\partial \sum_H y_j(t) \cdot \exp(\hat{\beta} z_j)}{\sum_{j=1}^n y_j(t) \cdot \exp(\hat{\beta} z_j)}$$

و با فرض

$$L = \sum_H y_j(t) \cdot \exp(\hat{\beta} z_j)$$

$$M = \sum_{j=1}^n y_j(t) \cdot \exp(\hat{\beta} z_j)$$

داریم:

$$\frac{\partial}{\partial \beta} \hat{P}(Z \in H | t_i) = \frac{\partial L}{\partial \beta M}$$

$$\frac{\partial}{\partial \beta} \hat{P}(Z \in H | t_i) = \frac{\sum_H y_j(t) z_j \exp(\hat{\beta} z_j) \cdot M - \sum_{j=1}^n y_j(t) z_j \exp(\hat{\beta} z_j) \cdot L}{M^2}$$

$$= \frac{\sum_H y_j(t) z_j \exp(\hat{\beta} z_j)}{M} - \frac{\sum_{j=1}^n y_j(t) z_j \exp(\hat{\beta} z_j) \cdot L}{M^2}$$

$$= \frac{L}{M} \left[ \frac{\sum_H y_j(t) z_j \exp(\hat{\beta} z_j)}{L} - \frac{\sum_{j=1}^n y_j(t) z_j \exp(\hat{\beta} z_j)}{M} \right]$$

در نتیجه

$$\left[ \frac{y_j(t) z_j \exp(\hat{\beta} z_j)}{\sum_{j=1}^n y_j(t) \cdot \exp(\hat{\beta} z_j)} \right]$$

که در آن

$$\exp(\hat{\beta} z_j)$$

$$\exp(\hat{\beta} z_j)$$

چون  $1 - \hat{S}(\cdot)$

در سمت راست

بود به صورت زیر

فرض می کنیم

(۱)

که در آن

بنابر این

در نتیجه

$$\begin{aligned} \frac{\partial}{\partial \beta} \hat{P}(Z \in H|t_i) &= \frac{\sum_{j=1}^n y_j(t) \cdot \exp(\hat{\beta} z_j)}{\sum_{j=1}^n y_j(t) \cdot \exp(\hat{\beta} z_j)} \\ &\times \left[ \frac{\sum_H y_j(t) z_j \exp(\hat{\beta} z_j)}{\sum_H y_j(t) \exp(\hat{\beta} z_j)} - \frac{\sum_{j=1}^n y_j(t) z_j \exp(\hat{\beta} z_j)}{\sum_{j=1}^n y_j(t) \cdot \exp(\hat{\beta} z_j)} \right] \\ &= \hat{P}(Z \in H|t) \cdot [E_p(Z|t, \pi^H) - E_p(Z|t, \pi)] \end{aligned}$$

که در آن

$$E_p(Z|t, \pi) = \frac{\sum_{j=1}^n y_j(t) z_j \exp(\hat{\beta} z_j)}{\sum_{j=1}^n y_j(t) \cdot \exp(\hat{\beta} z_j)}$$

$$E_p(Z|t, \pi^H) = \frac{\sum_H y_j(t) z_j \exp(\hat{\beta} z_j)}{\sum_H y_j(t) \cdot \exp(\hat{\beta} z_j)}$$

چون  $\hat{F}(\cdot) = 1 - \hat{S}(\cdot)$  برآوردگر از چپ پیوسته کاپلان مایر  $F(\cdot)$  است بنابراین عبارت نخست در سمت راست معادله (۱۲) یعنی  $\hat{V}ar \{ \hat{S}(t|Z \in H) | \beta \}$  می تواند توسط فرمول گرین وود<sup>۶</sup> به صورت زیر برآورد شود.  
فرض می کنیم  $S(t_k)$  تابع بقاء در زمان  $t_k$  باشد؛ لذا

$$\begin{aligned} S(t_k) &= P_r(T > t_1) \cdot P_r(T > t_2) \dots P_r(T > t_k | T > t_{k-1}) \\ &= P_1 \cdot P_2 \dots P_k \end{aligned}$$

که در آن

$$P_j = P_r(T > t_j | T > t_{j-1}) \quad \forall j$$

بنابراین

$$\hat{S}(t_k) = \prod_{j: t_j \leq t} P_j = \prod_{j: t_j \leq t} \left( 1 - \frac{d_j}{n_j} \right)$$

6) Green-Wood

کاپلان-مایر (۱۹۵۸) فرمول گرین وود (۱۹۲۶) زیر را برای برآورد واریانس  $\prod \hat{p}_i$  استفاده نمودند:

$$\hat{Var} \left( \prod_{t_i \leq t} \hat{p}_i \right) \approx \left( \prod_{t_i \leq t} \hat{p}_i \right)^2 \cdot \left\{ \prod_{t_i \leq t} \left( 1 + \frac{\hat{q}_i}{n_i \hat{p}_i} \right) - 1 \right\} \quad (15)$$

که در آن احتمال دوجمله‌ای از یک نمونه تصادفی به حجم  $n_i$  است و  $\hat{q}_i = 1 - \hat{p}_i$

$$\hat{Var} \left\{ \hat{S}(t|Z \in H)|_{\beta} \right\} \approx \hat{S}(t|Z \in H)|_{\beta}^2 \cdot \left\{ \prod_{t_i \leq t} \left( 1 + \frac{\hat{q}_i}{n_i \hat{p}_i} \right) - 1 \right\} \quad (16)$$

$$\begin{aligned} \hat{P}_i &= \hat{P}_r(T > t_i | T > t_{i-1}, Z \in H) = \frac{\hat{P}_r(T > t_i, T > t_{i-1}, Z \in H)}{\hat{P}_r(T > t_{i-1}, Z \in H)} \\ &= \frac{\hat{P}_r(T > t_i, Z \in H)}{\hat{P}_r(T > t_{i-1}, Z \in H)} \\ &= \frac{\hat{P}_r(T > t_i | Z \in H) \cdot \hat{P}_r(Z \in H)}{\hat{P}_r(T > t_{i-1} | Z \in H) \cdot \hat{P}_r(Z \in H)} \\ &= \frac{\hat{P}_r(T > t_i | Z \in H)}{\hat{P}_r(T > t_{i-1} | Z \in H)} \end{aligned}$$

طبق رابطه (۹) داریم:

$$\hat{P}_r(T > t_i | Z \in H) = \frac{\sum_{j=i+1}^k C_j + B_1}{\sum_{j=1}^k C_j + B_1}$$

$$\hat{P}_r(T > t_{i-1} | Z \in H) = \frac{\sum_{j=i}^k C_j + B_1}{\sum_{j=1}^k C_j + B_1}$$

پس

$$\frac{C_j + B_1}{\sum_{j=1}^k C_j + B_1}$$

$$\frac{C_j + B_1}{\sum_{j=1}^k C_j + B_1}$$

که  $n_i = \sum_{j=1}^k y_j(t_i)$  تعداد برای  $\hat{Var}(\hat{\beta})$  نیز از دستیافت

که در آن لگاریتم تابع درست

$$\left. \begin{aligned} & \\ & \end{aligned} \right\}$$

لذا

$$\left[ \frac{\sum_{j \in S} n_j \exp(\beta' x_j)}{\exp(\beta' x_j)} \right]$$

۳ مثال کاربرد

در یک مطالعه کاربرد درمان، گروه قومی، گ تومور، مدت دیسنزای



پس

$$\hat{P}_i = \hat{P}_r(T > t_i | T > t_{i-1}, Z \in H) = \frac{\sum_{j=i+1}^k C_j + B_1}{\sum_{j=i}^k C_j + B_1}$$

که  $n_i = \sum_{j=1}^k y_j(t_i)$  تعداد افراد زنده مانده تا زمان  $t_i$  است. برای  $\hat{Var}(\hat{\beta})$  نیز از دستیافت کاکس (۱۹۷۲ و ۱۹۷۵) به صورت زیر استفاده می‌کنیم:

$$\hat{Var}(\hat{\beta}) = - \left\{ \frac{\partial^2 Lik(\beta)}{\partial \beta_p \partial \beta_q} \right\}^{-1}$$

که در آن لگاریتم تابع درستنمایی جزئی  $\beta$  به صورت زیر است.

$$Lik(\beta) = \sum_{i=1}^k \left[ \beta' z_i - Ln \left\{ \sum_{j \in R_i} \exp(\beta' z_j) \right\} \right]$$

لذا

$$\frac{\partial Lik(\beta)}{\partial \beta_p} = \sum_{i=1}^k z_{ip} - \frac{\sum_{j \in R_i} z_{jp} \exp(\beta' z_j)}{\sum_{j \in R_i} \exp(\beta' z_j)}$$

$$\frac{\partial^2 Lik(\beta)}{\partial \beta_p \partial \beta_q} = - \sum_{i=1}^k \frac{\left\{ \frac{\sum_{j \in R_i} \exp(\beta' z_j) \left\{ \sum_{j \in R_i} z_{jp} z_{jq} \exp(\beta' z_j) \right\}}{\left\{ \sum_{j \in R_i} \exp(\beta' z_j) \right\}^2} \right\}}{\left\{ \sum_{j \in R_i} \exp(\beta' z_j) \right\}} - \frac{\left\{ \sum_{j \in R_i} z_{jp} \exp(\beta' z_j) \right\} \left\{ \sum_{j \in R_i} z_{jq} \exp(\beta' z_j) \right\}}{\left\{ \sum_{j \in R_i} \exp(\beta' z_j) \right\}^2}$$

### ۳ مثال کاربردی

در یک مطالعه کاربردی بقاء بیماران مبتلا به سرطان معده را از روی متغیرهای مستقل کیفیت درمان، گروه قومی، گروه خونی، جنس، سن، اعتیاد به مواد مخدر، اعتیاد به دخانیات، نوع نومور، مدت دیسفاژی به ماه، درجه دیسفاژی و سابقه سرطان در خانواده توسط روش استاندارد

و برآوردگر جدید پیش‌بینی می‌نماییم. توسط رگرسیون کاکس به روش Backward برای تعداد ۱۴۹ بیمار با ۱۸ مورد سانسور، ۲۱ مرحله متغیرهای توضیحی وارد معادله رگرسیونی می‌شوند. در مرحله اول تأثیر تمام متغیرهای مستقل روی زمان بقاء بررسی می‌شود. سپس با توجه به معنی داری ضرایب رگرسیونی به ترتیب متغیرها از معادله رگرسیونی حذف می‌گردند و در مرحله ۲۱ مدل نهایی به دست می‌آید. با توجه به جدول ۱ که متغیرهای مستقل موجود در مدل را در مرحله نهایی نشان می‌دهد، با ۹۵٪ اطمینان تنها دو متغیر کیفیت درمان (TRE) و نوع تومور آدنوکارسینوم (Type) به ترتیب با P- مقادیرهای ۰٫۰۰۴ و ۰٫۰۰۸ در زمان بقاء مؤثرند.

Variables in the Equation

Step	TRE	B	SE	Wald	df	Sig.	Exp(B)
21	Type	-0.759	0.264	8.236	1	0.004	0.468
		0.474	0.177	7.132	1	0.008	1.606

جدول ۱. متغیرهای موجود در مدل رگرسیون در مرحله نهایی

پس از برآورد تابع بقاء توسط مدل رگرسیون کاکس، تابع بقاء توسط برآوردگر جدید برای هر یک از گروه‌های خطر که بر اساس دو متغیر توضیحی مهم کیفیت درمان (TRE) و نوع تومور آدنوکارسینوم (Type) به دست می‌آید؛ برآورد می‌شود. بدین ترتیب واریانس‌های به دست آمده از روش استاندارد و برآوردگر جدید توسط روش آماری آزمون استودنت مقایسه می‌گردند. نتایج به شرح جدول ۲ است.

### T-Test

Group Statistics

COD	N	Mean	Std Deviation	Std. Error Mean
برآوردگر جدید واریانس	164	30.66628	392.6508696	30.66088
برآوردگر زیرگروه کاپلان مایر	167	1.01E-02	0.1161363	8.99E-03

95% Confidence Interval of the Difference	
Lower	Upper
-29.114	90.4264
-29.888	91.1999

جدول ۲. نتایج

جدول فوق نشان روش جدید و برآورد برای میانگین تفاوت

### مراجع

- [۱] میلی، آر. ج. فردوسی،
- [۲] چپ تی تربیت مدل
- [۳] مقدس ز سرطان دکتري،
- the uncer- regression
- ision model statist., 10,
- s and life- 7.
- a. Biomet-

Independent Samples Test

	Levene's Test for Equality of Variances	t-test for Equality of Means								
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
وارانس Equal variances assumed	4.121	.043	1.009	329	.314	30.656199	30.383393	-29.114	90.4264	
Equal variances not assumed			1.000	163.0	.319	30.656199	30.660883	-29.888	91.1999	

جدول ۲. نتایج آزمون استیودنت برای مقایسه مقادیر برآورد شده واریانس تابع بقا توسط دو روش

جدول فوق نشان می دهد که با ۵ درصد خطا تفاوت معنی داری بین واریانس برآورد شده توسط روش جدید و برآوردگر حالت استاندارد برای تابع بقا، وجود ندارد. فاصله اطمینان ۹۵ درصد برای میانگین تفاوت واریانس برآورد شده توسط دو روش نیز، مؤید مطلب فوق است.

### مراجع

- [۱] میلر، آر. جی، ترجمه بزرگنیا، ابوالقاسم و رضایی یزند، حجت. تحلیل بقا، مشهد، دانشگاه فردوسی، ۱۳۸۰.
- [۲] چپ تی لی، مترجم دکتر انوشیروان کاظم نژاد، تحلیل بقای کاربردی، تهران، دانشگاه تربیت مدرس، ۱۳۸۰.
- [۳] مقدس زاده، پافر، مدل چند حالتی برای تجزیه و تحلیل بقا و کاربرد آن در مبتلایان به سرطان مری و معده همراه با یک بیماری دیگر در استان خراسان، پایان نامه برای اخذ درجه دکتری، دانشگاه تربیت مدرس، دانشکده پزشکی، ۱۳۸۰.
- [4] Altman, D.G. and Anderson, P.K. (1986), A note on the uncertainty of a survival probability estimated from Cox's regression model, *Biometrika*, 73, 722-724.
- [5] Anderson, P. K. and Gill, R. D. (1982), Cox's regression model for counting processes: a large sample study, *Ann. Statist.*, 10, 1100-1120.
- [6] Breslow, N. (1972), Discussion on 'Regression models and life-tables' (by D. R. Cox), *J. R. Statist. Soc. B*, 34, 216-217.
- [7] Breslow, N. (1974), Covariance analysis of survival data. *Biometrics*, 30, 89-99.

Backward برای تعداد رگرسیون می شوند. سپس با توجه به می گردند و در مرحله موجود در مدل را در (TRE) و نوع تومور مان بقا مؤثرند.

Step	1
21	T

برآوردگر جدید برای هر (TRE) و نوع تومور

بد توسط روش آماری

### T-Test

COD
جدید واریانس
برگروه کابلان مایر

- [8] Cox, D.R. (1972), Regression models and life-tables. (with discussion). *J. R. Statist. Soc. B*, 34, 187-220.
- [9] Cox, D.R. (1975), Partial likelihood. *Biometrika*, 62, 269-276.
- [10] Kalbfleisch, J. D. and Prentice, R. L. (1980), *The statistical Analysis of Failure Time Data*. New York: Wiley.
- [11] Kaplan, E. L. and Meier, P. (1958), Non-parametric estimation from incomplete observations, *J. Am. Statist. Ass.*, 53, 457-481.
- [12] Keiding, N. (1995), Historical controls and modern survival analysis. *Lifetime Data Anal.*, 1, 19-25.
- [13] Keiding, N. and Knudman, M. W. (1990), Survival analysis in natural history studies of disease, *Statist. Med.*, 9, 1221-1222.
- [14] Lin, D. Y., Fleming, T. R. and Wei, L. J. (1994), Confidence bands for survival curves under the proportional hazards mode, *Biometrika*, 81, 73-81.
- [15] O'Quigley, J. and Xu, R. (1998), Godness-of-fit in survival analysis, In *Encyclopedia of Biostatistics* (des P. Armitage and T. Colton), 1731-1745. chichester: Wiley.
- [16] Lachin, J. M. (2000), *Biostatistical Methods, The Assessment of Relative Risks*, New York: Wiley.

مجموعه مفاه

تعمین اند

فرشید جم

پژوهشکده

چکیده:

موضوعی

حاضر اهمیت

تابعی آمار

اشاره شد

نیاستند

اندازه نم

باشد، قا

با فرضیه

برای تم

برآوردگی

(۱)

اما به

(۲)

که

الف

ب

ج

د

ه

و

ز

ح

ط

ی

ک

خ

گ

چ

پ

ت

ث

ج

د

ه

و

ز

ح

ط

ی

ک

خ

گ

چ

پ

ت

ث

ج

د

ه

و

ز

ح

ط

ی

ک

خ

گ

چ

پ

ت

ث

ج

د

ه

و

ز

ح

ط

ی

ک

خ

گ

چ

پ

ت

ث

ج

د

ه

و

ز

ح

ط

ی

ک

خ

گ

چ

پ

ت

ث

ج

د

ه

و

ز

ح

ط

ی

ک

خ

گ

چ

پ

ت

ث

ج

د

ه

و

ز

ح

ط

ی

ک

خ

گ

چ

پ

ت

ث

ج

د

ه

و

ز

ح

ط

ی

ک

خ

گ

چ

پ

ت

ث

ج

د

ه

و

ز

ح

ط

ی

ک

خ

گ

چ

پ

ت

ث

ج

د

ه

و

ز

ح

ط

ی

ک

خ

گ

چ

پ

ت

ث

ج

د

ه

و

ز

ح

ط

ی

ک

خ

گ

چ

پ

ت

ث

ج

د

ه

و

ز

ح

ط

ی

ک

خ

گ

چ

پ

ت

ث

ج

د

ه

و

ز

ح

ط

ی

ک

خ

گ

چ

پ

ت

ث

ج

د

ه

و

ز

ح

ط

ی

ک

خ

گ

چ

پ

ت

ث

ج

د

ه

و

ز

ح

ط

ی

ک

خ

گ

چ

پ