

تحلیل بیزی داده‌های ریسک‌های رقابتی مدل‌سازی شده با استفاده از تابع مفصل مانای مثبت

محمد رضا آخوند، انوشیروان کاظم‌نژاد*، ابراهیم حاجی‌زاده

گروه آمار زیستی، دانشگاه تربیت مدرس، تهران، ایران

تاریخ پذیرش: ۸۹/۲/۱۵

تاریخ دریافت: ۸۸/۹/۲

چکیده

مقدمه: در سال‌های اخیر استفاده از توابع مفصل جهت مدل‌سازی داده‌های بقاء چند متغیره گسترش فزاینده‌ای یافته است که از جمله داده‌های بقاء چند متغیره داده‌های ریسک‌های رقابتی می‌باشند.

هدف: در این مطالعه قصد داریم تحلیل بیزی داده‌های ریسک‌های رقابتی مدل‌سازی شده با استفاده از تابع مفصل مانای مثبت را ارائه کنیم. در پایان نیز مدل پیشنهادی را بر روی داده‌های کارآزمایی بالینی داروی دی اتیل استیل بسترول اجرا خواهیم کرد. در این کارآزمایی ۵۰۶ بیمار مبتلا به سرطان پروستات تحت درمان با دزهای مختلفی از داروی دی اتیل استیل بسترول قرار گرفته‌اند.

روش بررسی: پس تشکیل تابع درستنمایی به کمک توابع مفصل، با انتخاب توزیع‌های پیشین مناسب برای پارامترها توزیع پسین پارامترها را با استفاده از الگوریتم متروپلیس- هستینگ و نمونه‌گیری برشی بدست خواهیم آورد.

نتایج: برازش مدل بیزی به داده‌ها نشان داد که تاثیر نوع درمان بر روی زمان مرگ در اثر سرطان پروستات وابسته به سن و وزن بیمار می‌باشد، که نتایج حاصل از روش‌های کلاسیک نیز نتیجه بدست آمده را تأیید می‌کند.

نتیجه‌گیری: برآورد ضرائب همبستگی تاوکنندال بدست آمده در مدل بیزی تغییرات کمتری نسبت به مدل کلاسیک از خود نشان می‌دهد.

واژه‌های کلیدی: تحلیل بیزی، توابع مفصل، ریسک‌های رقابتی

مقدمه

در داده‌های بقاء حادثه مورد علاقه ممکن است به دلایل مختلفی روی دهد. برای مثال، اگر علاقه‌مند به تحلیل داده‌های زمان تا وقوع مرگ در اثر بیماری قلبی باشیم، برای برخی افراد تحت مطالعه ممکن است مرگ به دلایل دیگری غیر از بیماری قلبی، رخ دهد. این دلایل متفاوت شکست، بنام ریسک‌های رقابتی^۱ شناخته می‌شوند. گلی و همکاران^(۱) ریسک رقابتی را به صورت حادثه‌ای که مانع مشاهده حادثه مورد علاقه می‌شود و یا احتمال وقوع آنرا تغییر می‌دهد، تعریف کرده‌اند.

مدل‌های مختلفی برای تحلیل داده‌های ریسک‌های رقابتی پیشنهاد شده است. از جمله مدل‌های مورد استفاده در تحلیل داده‌های ریسک‌های رقابتی، مدل مخاطره متناسب ویژه^۲- علت کاکس^۳ است. در این روش برای تابع خطر ویژه^۲- علت بعدی نامتناهی در نظر گرفته می‌شود، در نتیجه اگر علاقه‌مند به برآورد احتمالات بقاء ویژه^۲- علت^۳ باشیم، خطرات بدست آمده توابع پله‌ای بسیار غیر معقولی می‌باشند که دارای نوار اطمینان بسیار عریضی‌اند. از دیگر روش‌های مورد استفاده در تحلیل داده‌های ریسک‌های رقابتی مدل‌های کاملاً پارامتری می‌باشند که بیشتر این مدل‌ها ساختاری مستقل برای ریسک‌های رقابتی در نظر می‌گیرند. به عبارت دیگر فرض می‌کنند که زمان‌های شکست برای ریسک‌های مختلف دو به دو از هم مستقل می‌باشند. مهمترین دلیل برای فرض چنین ساختار مستقلی برای ریسک‌های رقابتی وابسته، عدم شناسایی پذیری^۴ مدل در هنگام در نظر گرفتن ساختاری وابسته است. بنابراین انتخاب مدل مستقل هم باعث سادگی مدل و هم موجب رفع مشکل شناسایی پذیری مدل می‌گردد. اما در بسیاری از موارد در نظر گرفتن چنین ساختاری می‌تواند بسیار گمراه کننده باشد. اسکالر و کریر^(۲) مدلی را برای ریسک‌های رقابتی وابسته بر مبنای توابع مفصل ارائه کردند که علاوه بر در نظر گرفتن همبستگی میان ریسک‌های مختلف، مشکل شناسایی پذیری مدل‌های کاملاً پارامتری را نداشت و در عین حال دامنه‌ای گسترده از همبستگی‌ها را از طریق پارامتر وابستگی میان زمان‌های بقاء فراهم می‌کرد. در این مدل توزیع‌های حاشیه‌ای به گونه‌ای کاملاً پارامتری انتخاب می‌شوند و می‌توان تاثیر متغیرهای کمکی را بر روی ریسک‌های مختلف بدست آورد.

در سال‌های اخیر استفاده از توابع مفصل در زمینه‌های مختلف علمی گسترش فزاینده‌ای یافته است و کتاب‌هایی نیز در این زمینه نوشته شده است که از جمله آنها می‌توان به نلسن،^(۳) جو^(۴) و چروبینی و همکاران^(۵) اشاره کرد. محققین مختلفی نیز توابع مفصل را برای مدل‌سازی داده‌های بقاء چند متغیره بکار گرفته‌اند که از میان آنها می‌توان به موارد زیر اشاره کرد. آکرز^(۶) روشی را جهت مدل‌سازی همبستگی در مدل‌های شکنندگی دو متغیره پیشنهاد کرد. شی و لوئیس^(۷) روش‌های برآورد پارامتری و نیمه-پارامتری همبستگی را در تحلیل داده‌های بقا دو متغیره مورد بررسی قرار دادند. ژانگ و کلین^(۸)، کریر^(۹) و کیاشف و همکاران^(۱۰) از توابع مفصل جهت تحلیل داده‌های ریسک‌های رقابتی استفاده کردند. هی و لاولس^(۱۱) از مدل‌های تکه‌ای ثابت^۵ و توابع اسپلاین^۶ برای توابع خطر پایه در مدل‌های بقاء چند متغیره استفاده کردند. آنها تابع مفصل کلایتون را جهت مدل‌سازی چنین داده‌های به کار گرفتند.

¹-Competing Risks

²-Cause Specific Cox Proportional Hazard Mode

³-Cause Specific Survival Probabilities

⁴-Identifiability

⁵-Picewise Constant Model

⁶-Spline Function

اما با وجود برخی مزایای روش های بیزی نسبت به روش های کلاسیک همچون نیرومندی در مقابل کم بودن اندازه نمونه و عدم نیاز به استفاده از توزیع های مجانبی جهت برآوردها تاکنون روش های بیزی کمتر در تحلیل داده های مدل سازی شده با استفاده از توابع مفصل به کار گرفته شده اند و مقالات بسیار کمی در زمینه تحلیل بیزی داده های بقاء چند متغیره مدل سازی شده با استفاده از توابع مفصل می توان یافت. تنها مقالات منتشره در این زمینه مربوط می شود به گوستافون و همکاران^(۱۲) که از نمونه گیری نقاط مهم^۱ جهت برازش توابع مفصل به داده های بقاء دومتغیره استفاده کردند. آنها ساختار وابستگی میان زمان های بقاء را از طریق تابع مفصل نرمال ایجاد کردند. رومئو و همکاران^(۱۳) نیز روش های بیزی را در تحلیل داده های بقا دومتغیره مدل سازی شده با استفاده از توابع مفصل به کار گرفتند.

در این مطالعه هدف آن است که پس از مدل سازی داده های ریسک های رقابتی با استفاده از تابع مفصل مانای مثبت، از روش های بیزی برای برآورد پارامترها استفاده شود. با توجه به اینکه محاسبه توزیع پسین پارامترها با استفاده از روش های تحلیلی امکان پذیر نمی باشد، برای انجام تحلیل بیزی لازم است از روش های شبیه سازی مونت کارلوی زنجیر مارکوفی^۳ استفاده شود. استنباط بیزی با استفاده از روش های شبیه سازی مونت کارلوی زنجیر مارکوفی در سال های اخیر با گسترش توانمندی های محاسباتی بسیار مورد توجه قرار گرفته و قادر است برای هر حجم نمونه دلخواه، استنباط نسبتا دقیقی از پارامترها ارائه دهد.

در این مقاله پس از معرفی توابع مفصل، به چگونگی مدل سازی داده های بقاء ریسک های رقابتی با استفاده از توابع مفصل خواهیم پرداخت. سپس با انتخاب توزیع های پیشین مناسب جهت پارامترها، روش بیزی را در تحلیل چنین داده هایی ارائه خواهیم کرد و در پایان نیز مدل بدست آمده را بر روی داده های ریسک های رقابتی بیماران مبتلا به سرطان پروستات اجرا خواهیم کرد.

مواد و روش ها

توابع مفصل به صورت توزیع های چندمتغیره ای که دارای توزیع های حاشیه ای یکنواخت می باشند تعریف می - گردند. به بیان دقیق تر تابع مفصل، به صورت تابعی دومتغیره همانند $[0,1] \rightarrow [0,1]^2 : C$ تعریف می شود که در شرایط زیر صدق کند :

$$1. \text{ برای هر } u \in [0,1] \quad C(u,0) = C(0,u) = 0$$

$$2. \text{ برای هر } u \in [0,1] \quad C(u,1) = C(1,u) = u$$

$$3. \text{ برای تمام } 0 \leq u_1 \leq u_2 \leq 1 \text{ و } 0 \leq v_1 \leq v_2 \leq 1$$

$$C([u_1, u_2] \times [v_1, v_2]) = C(u_2, v_2) - C(u_1, v_2) - C(u_2, v_1) + C(u_1, v_1) \geq 0$$

خانواده های مختلفی از توابع مفصل وجود دارند که از جمله مهمترین آنها می توان به خانواده توابع مفصل ارشمیدسی اشاره کرد^(۱۴). در این خانواده از توابع مفصل تابع بقاء دومتغیره به صورت زیر تعریف می گردد:

¹ Important Sampling

² Markov Chain Monte Carlo

$$S(t_1, t_2) = \varphi_\alpha \left(\varphi_\alpha^{-1} (S_1(t_1)) + \varphi_\alpha^{-1} (S_2(t_2)) \right) \quad (1)$$

که φ_α تابع مولد نامیده می‌شود، $\varphi_\alpha : [0, +\infty] \rightarrow [0, 1]$ ، $\varphi(0) = 1$ ، $\varphi' < 0$ ، $\varphi'' > 0$ و φ_α^{-1} معکوس تابع مولد می‌باشد، α نیز نشان دهنده میزان همبستگی میان زمان‌های بقاء می‌باشد.

مهمترین توابع مفصلی که در این کلاس قرار می‌گیرند عبارتند از مفصل مانای مثبت،^(۱۵و۱۶) مفصل فرانک^(۱۷) و مفصل کلایتون^(۱۸). با استفاده از ضریب همبستگی تاوکنندال بهتر می‌توان میزان همبستگی میان متغیرها را درک کرد که در توابع مفصل ارشمیدسی به صورت زیر قابل محاسبه است:

$$\tau_\alpha(T_1, T_2) = 4 \int_0^1 \frac{\varphi^{-1}(t)}{\varphi^{-1'}(t)} dt + 1 \quad (2)$$

جهت انجام تحلیل داده‌ها نیاز به انتخاب یک تابع مفصل می‌باشد، که بدلیل برخی خواص مناسب از جمله متناسب بودن خطر در حاشیه‌ها از تابع مفصل ارائه شده توسط هوگارد^(۱۵و۱۶) جهت تشکیل تابع بقاء توام استفاده می‌کنیم.

$$\tau_\alpha(T_1, T_2) = 4 \int_0^1 \frac{\varphi^{-1}(t)}{\varphi^{-1'}(t)} dt + 1 \quad (3)$$

با استفاده از تابع مفصل بالا تابع بقاء توام به صورت زیر بدست می‌آید:

$$S(t_1, t_2) = \exp \left\{ - \left[\left(-\ln(S_1(t_1)) \right)^\alpha + \left(-\ln(S_2(t_2)) \right)^\alpha \right]^\alpha \right\} \quad (4)$$

که در آن $S_1(t_1)$ ، $S_2(t_2)$ توابع حاشیه‌ای و α پارامتر همبستگی است. هنگامیکه $\alpha \rightarrow 1$ استقلال میان حاشیه‌ها را خواهیم داشت یعنی $S(t_1, t_2) = S(t_1) \times S(t_2)$. با مشتق‌گیری از تابع بقاء توام بالا نسبت به t_1, t_2 می‌توان تابع چگالی توام را بدست آورد.

از مهمترین توزیع‌های مورد استفاده در تحلیل داده‌های بقاء توزیع وایبل می‌باشد. این توزیع دارای نرخ خطری است که می‌تواند صعودی، نزولی و یا ثابت باشد. با در نظر گرفتن توزیع وایبل برای حاشیه‌ها $T_j \sim \text{weibull}(r_j, \lambda_j)$ با تابع چگالی

$$f_j(t | r_j, \lambda_j) = r_j \lambda_j t^{r_j-1} e^{-\lambda_j t^{r_j}}, \quad r_j > 0, \quad \lambda_j > 0, \quad j = 1, 2 \quad (5)$$

و تابع بقاء به صورت زیر

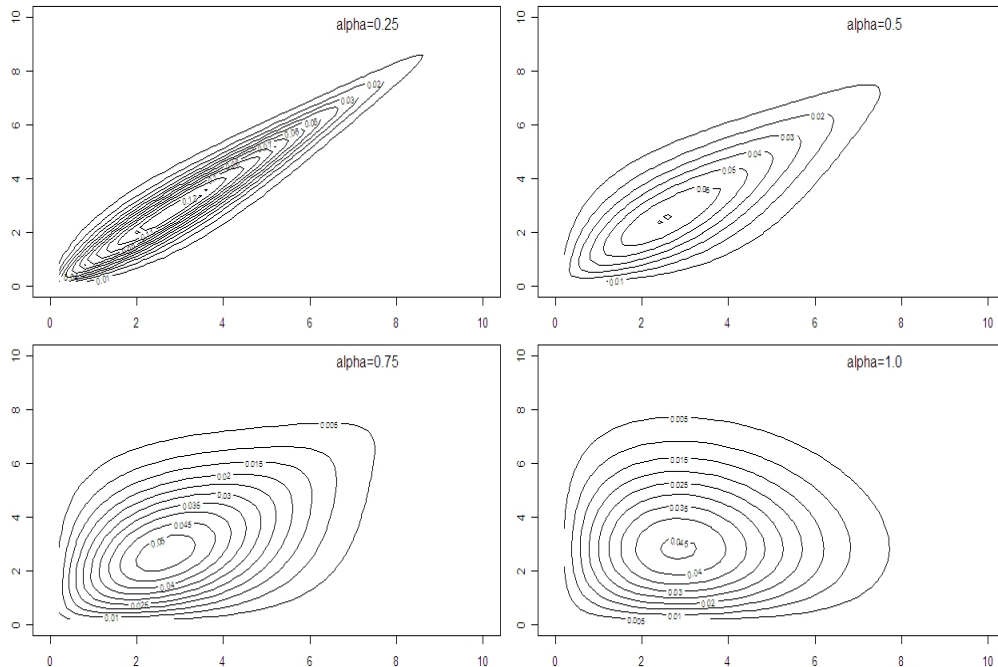
$$S_j(t | r_j, \lambda_j) = e^{-\lambda_j t^{r_j}}, \quad r_j > 0, \quad \lambda_j > 0, \quad j = 1, 2 \quad (6)$$

تابع بقاء دومتغیره توزیع وایبل با استفاده از مفصل مانای مثبت به صورت زیر بدست می‌آید:

$$S(t_1, t_2) = \exp \left\{ - \left[\left(\lambda_1 t_1^{r_1} \right)^\alpha + \left(\lambda_2 t_2^{r_2} \right)^\alpha \right]^\alpha \right\} \quad (7)$$

که به سادگی و با دوبار مشتق‌گیری از تابع بقاء دومتغیره بالا می‌توان تابع چگالی توام را بدست آورد. شکل (۱) تابع چگالی توام وایبل بدست آمده بر مبنای مقادیر مختلف پارامتر همبستگی مفصل مانای مثبت را نشان می‌دهد. اشکال

زیر به ازاء مقادیر برابر ۰/۲۵، ۰/۵، ۰/۷۵، ۱/۰۰ بدست آمده‌اند. همانطوریکه مشاهده می‌گردد با نزدیک شدن مقدار پارامتر وابستگی به یک همبستگی کاهش می‌یابد. برای این مفصل با استفاده از معادله (۲) به سادگی می‌توان میزان ضریب همبستگی تاوکنندال را به صورت $\tau_\alpha = 1 - \alpha$ بدست آورد.



شکل ۱- تابع چگالی توام و ایبل بدست آمده با استفاده از تابع مفصل مانای مثبت و با در نظر گرفتن $\tau = 2$ ، $\lambda = 4$ و مقادیر مختلف پارامتر همبستگی

تابع درستمایی داده‌های ریسک‌های رقابتی

داده‌های ریسک‌های رقابتی با علت شکست را در نظر بگیرید. همچنین فرض کنید T_{ij} ، زمان شکست به دلیل i -امین ریسک برای j -امین فرد در نمونه باشد که $i = 1, 2, \dots, n$ ، $j = 1, 2, \dots, m$ است. در این حالت یک مجموعه از داده‌ها به صورت $\{z_i, c_{ij}\}$ خواهیم داشت که در آن $T_i = \min(T_{ij}, j = 1, 2, \dots, m)$ زمان‌های بقاء واقعی، $z_i = \min(T_i, C_i)$ و زمان‌های سانسور شدن است. در اینجا ماتریس نشانگر وضعیت به صورت $c_{ij} = I(T_i = T_{ij})$ و بردار نشانگر وضعیت به صورت $c_i = \sum_{j=1}^m c_{ij}$ تعریف می‌گردد. همچنین فرض می‌شود که برای $T_{ij} \neq T_{ik}$ ، $k \neq j$

در داده‌های ریسک‌های رقابتی با دوگروه از افراد مواجه هستیم گروهی که حادثه برای آنها در اثر یکی از ریسک‌ها بوقوع می‌پیوندد و گروه دیگر افرادی که تا پایان مطالعه حادثه‌ای برای آنها رخ نمی‌دهد. برای افرادی که حادثه برای آنها براساس وقوع یکی از ریسک‌ها مثلا ریسک ام رخ می‌دهد، سهم آنها در تابع درستمایی به صورت زیر قابل محاسبه است:

$$f^{(j)}(t) = \lim_{\Delta t \rightarrow 0} \frac{\Pr(T_1 > t, \dots, t \leq T_j \leq t + \Delta t, \dots, T_m > t | T \geq t)}{\Delta t} = -\frac{\partial}{\partial t_j} S(t_1, \dots, t_m) |_{t_k = t, \forall k} \quad (8)$$

و برای گروه دیگری که تا پایان مدت زمان مطالعه حادثه مورد نظر برای آنها به وقوع نمی‌پیوندد، مشارکت آنها در تشکیل تابع درستنمایی به صورت $S_T(t) = \Pr(T_1 > t, \dots, T_m > t)$ خواهد بود. با فرض اینکه متغیر سانسور شدن برای فرد $I=1$ ، C_1 مستقل از زمان بقاء مربوط به $-$ امین ریسک T_{1j} است، می‌توان تابع درستنمایی را به صورت زیر نوشت: (۱۹)

$$L = \prod_{i=1}^n \left(\prod_{j=1}^m [f^{(j)}(z_i)]^{c_{ij}} \right) [S_T(z_i)]^{1-c_i} \tag{۹}$$

در حالتی که فقط دو دلیل جهت وقوع شکست داشته باشیم ($j = 2$) باشد، در اینصورت با استفاده تابع بقاء دومتغیره بدست آمده از طریق تابع مفصل، می‌توان تابع درستنمایی را به صورت زیر نوشت :

$$L = \prod_{i=1}^n \left[\frac{\partial C_\alpha(S_{1\theta_1}(z_i), S_{2\theta_2}(z_i))}{\partial S_{1\theta_1}(z_i)} f_{1\theta_1}(z_i) \right]^{c_{i1}} \left[\frac{\partial C_\alpha(S_{1\theta_1}(z_i), S_{2\theta_2}(z_i))}{\partial S_{2\theta_2}(z_i)} f_{2\theta_2}(z_i) \right]^{c_{i2}} \tag{۱۰}$$

که در آن توابع چگالی حاشیه‌ای، $S_{1\theta_1}$ ، $S_{2\theta_2}$ ، توابع بقاء حاشیه‌ای و θ_{11} ، θ_{21} ، پارامترهای موجود در حاشیه‌ها و C_α و C_α متغیرهای وضعیت سانسور شدن در اثر ریسک اول و دوم می‌باشند. با استفاده از تابع درستنمایی ارائه شده در بالا و در نظر گرفتن توزیع وایبل برای حاشیه‌ها خواهیم داشت :

$$L(\alpha, r_1, r_2, \lambda_1, \lambda_2 | z, c_1, c_2) = \prod_{i=1}^n r_1^{c_{i1}} r_2^{c_{i2}} \lambda_1^{\frac{1}{\alpha} c_{i1}} \lambda_2^{\frac{1}{\alpha} c_{i2}} z_i^{(r_1/\alpha - 1)c_{i1} + (r_2/\alpha - 1)c_{i2}} \tag{۱۱}$$

$$\left[(\lambda_1 z_i^{r_1})^{\frac{1}{\alpha}} + (\lambda_2 z_i^{r_2})^{\frac{1}{\alpha}} \right]^{-(\alpha-1)(c_{i1}+c_{i2})} \times \exp \left\{ - \left[(\lambda_1 z_i^{r_1})^{\frac{1}{\alpha}} + (\lambda_2 z_i^{r_2})^{\frac{1}{\alpha}} \right]^\alpha \right\}$$

همچنین با وارد کردن متغیرهای کمکی از طریق $\lambda_j = \exp(\beta_{0j} + \beta_{1j} X)$ در مدل، تابع درستنمایی به صورت زیر بدست خواهد آمد :

$$L(\alpha, r_1, r_2, \beta_1, \beta_2 | z, c_1, c_2, X) = \prod_{i=1}^n r_1^{c_{i1}} r_2^{c_{i2}} z_i^{(r_1/\alpha - 1)c_{i1} + (r_2/\alpha - 1)c_{i2}} e^{(\beta_{01} + \beta_{11} X_i) c_{i1} / \alpha} e^{(\beta_{02} + \beta_{12} X_i) c_{i2} / \alpha} s_{i\alpha}^{(\alpha-1)(c_{i1}+c_{i2})} e^{-s_{i\alpha}^\alpha} \tag{۱۲}$$

که در آن $s_{i\alpha} = (z_i^{r_1} e^{(\beta_{01} + \beta_{11} X)} + z_i^{r_2} e^{(\beta_{02} + \beta_{12} X)})$ و $\beta_j = (\beta_{0j}, \beta_{1j})$ می‌باشد.

تحلیل بیزی داده‌های ریسک‌های رقابتی

جهت تحلیل بیزی لازم است که توزیع پیشین برای پارامترها انتخاب شود، که با در نظر گرفتن توزیع‌های پیشین به صورت $\beta_{1j} \sim \text{Normal}(c_{1j}, d_{1j}^2)$ و $\beta_{0j} \sim \text{Normal}(c_{0j}, d_{0j}^2)$ ، $r_j \sim \exp(r_{0j})$ ، $\alpha \sim \text{Beta}(a, b)$ پسین توام برای پارامترها $(\alpha, r_1, r_2, \beta_1, \beta_2)$ به صورت زیر بدست می‌آید :

$$\pi(\alpha, r_1, r_2, \beta_1, \beta_2 | z, c_1, c_2, X) \propto \alpha^{a-1} (1-\alpha)^{b-1} e^{-r_{01} r_1 - r_{02} r_2} e^{\sum_i (\beta_{01} + \beta_{11} X_i) c_{i1} / \alpha - (\beta_{01}^2 - 2\beta_{01} c_{01}) / 2d_{01}^2 - (\beta_{11}^2 - 2\beta_{11} c_{11}) / 2d_{11}^2} \tag{۱۳}$$

$$e^{\sum_i (\beta_{02} + \beta_{12} X_i) c_{i2} / \alpha - (\beta_{02}^2 - 2\beta_{02} c_{02}) / 2d_{02}^2 - (\beta_{12}^2 - 2\beta_{12} c_{12}) / 2d_{12}^2} \prod_{i=1}^n r_1^{c_{i1}} r_2^{c_{i2}} z_i^{(r_1/\alpha - 1)c_{i1} + (r_2/\alpha - 1)c_{i2}} s_{i\alpha}^{(\alpha-1)(c_{i1}+c_{i2})} e^{-s_{i\alpha}^\alpha}$$

در حالتی که $K \geq 1$ متغیر کمکی در مدل وجود داشته باشد، با در نظر گرفتن توزیع پیشین ضرائب رگرسیونی به صورت $\beta_{kj} \sim \text{Normal}(c_{kj}, d_{kj}^2)$ ($j = 1, 2, k = 1, 2, \dots, K$) و با شرط استقلال میان آنها، توزیع پسین توام برای پارامترها به صورت زیر بدست می‌آید :

$$\pi(\alpha, r_1, r_2, \beta_1, \beta_2 | z, c_1, c_2, X) \propto e^{\sum_i (\beta_{01} + \sum_{k=1}^K \beta_{k1} X_{ik}) c_{i1} / \alpha - (\beta_{01}^2 - 2\beta_{01}c_{01}) / 2d_{01}^2 - \sum_{k=1}^K (\beta_{k1}^2 - 2\beta_{k1}c_{k1}) / 2d_{k1}^2} e^{\sum_i \left(\beta_{02} + \sum_{k=1}^K \beta_{k2} X_{ik} \right) c_{i2} / \alpha - (\beta_{02}^2 - 2\beta_{02}c_{02}) / 2d_{02}^2 - \sum_{k=1}^K (\beta_{k2}^2 - 2\beta_{k2}c_{k2}) / 2d_{k2}^2} \alpha^{a-1} (1-\alpha)^{b-1} e^{-r_{01}r_1 - r_{02}r_2} \prod_{i=1}^n r_1^{\delta_{i1}} r_2^{\delta_{i2}} \quad (14)$$

$$z_i^{(r_1/\alpha - 1)c_{i1} + (r_2/\alpha - 1)c_{i2}} s_{\alpha i}^{(\alpha-1)(c_{i1} + c_{i2})} e^{-s_{\alpha i}^a}$$

که در آن $s_{\alpha i} = (z_i^{r_1} e^{\left(\beta_{01} + \sum_{k=1}^K \beta_{k1} X_{ik} \right)} + z_i^{r_2} e^{\left(\beta_{02} + \sum_{k=1}^K \beta_{k2} X_{ik} \right)})$ می باشد. به خاطر پیچیدگی و ابعاد زیاد توزیع توام پسین بدست آمده، امکان محاسبه توزیع پسین پارامترهای مدل به روش تحلیلی وجود ندارد. لذا برای تقریب توزیع پسین پارامترها از روش های مونت کارلوی زنجیره مارکوفی استفاده می گردد. در این روش با نمونه گیری متوالی از توزیع های شرطی کامل پارامترها، زنجیره های مارکوفی تشکیل خواهد شد که در صورت همگرایی می توان نمونه های تولید شده را نمونه هایی از توزیع های حاشیه ای پسین دانست و بر اساس آن استنباط در مورد پارامترهای مدل را انجام داد. دوره قبل از همگرایی زنجیر مارکوف، دوره تطبیق^۱ نامیده می شود و از نتایج این دوره در برآورد پارامترها استفاده نمی گردد. ساده ترین روش مونت کارلوی زنجیره مارکوفی، نمونه گیری گیبز^۲ است. از این روش، موقعی استفاده می گردد که توزیع های شرطی کامل از نوع شناخته شده ای باشند. در غیر این صورت، با استفاده از الگوریتم های دیگری چون الگوریتم متروپلیس - هستینگ^۳، نمونه گیری ردی^۴ و نمونه گیری ردی سازوار^۵ و نمونه گیری برشی^۶ نمونه گیری از توزیع های شرطی کامل امکان پذیر است. (۲۴-۲۰)

کاربرد: شناسایی عوامل موثر بر مرگ در اثر سرطان پروستات

به منظور به کار گیری مدل بیزی بر روی داده های واقعی از داده های معرفی شده توسط اندرو و هرزبرگ^(۲۵) استفاده می گردد. داده ها از کارآزمایی بالینی داروی دی اتیل استیل بسترول^(DES) بر روی بیماران مراحل ۳ و ۴ بیماری سرطان پروستات بدست آمده است. ۵۰۶ بیمار در طول سال های ۱۹۶۷-۱۹۶۹ وارد این کارآزمایی شده اند و دزهای مختلفی از دارو برای آنها تجویز شده است. آزمایش به صورت دوسوکور بوده و درمان به صورت قرص (دارونما)، ۰/۲ میلی گرم، ۱ میلی گرم و مصرف روزانه ۵ میلی گرم از داروی دی اتیل استیل بسترول بوده است. بیماران بر اساس پروتکل در فواصل شش ماهه مورد پیگیری قرار گرفته اند.

¹ -Burn-in Period

² -Gibbs Sampling

³ - Metropolis Hastings

⁴ - Rejection Sampling

⁵ - Adaptive Rejection Sampling

⁶ - Slice Sampling

⁷ - Diethylstilbestrol

از ۴۸۳ بیمار با داده‌های کامل، ۱۳۵ مرگ از سرطان پروستات (۰.۲۰)، ۲۱۹ مرگ از سایر بیماری‌ها (۰.۴۵) و ۱۳۹ زمان سانسور شده (۰.۲۹) مشاهده گردید. هدف بررسی تاثیر داروی دی اتیل استیل بسترول بر بقاء بیماران می‌باشد. با توجه به اثرات جانبی ممکن این دارو بر بیماری‌های قلبی و یا سایر سرطان‌ها، جهت ارزیابی واقعی مزایای درمان با این دارو نه تنها مرگ در اثر سرطان پروستات بلکه مرگ در اثر سایر دلایل نیز باید مورد توجه قرار گیرد.

روش‌های مختلفی برای طبقه بندی‌های ریسک‌های مرگ در داده‌های سرطان پروستات مورد استفاده قرار گرفته است. کی،^(۲۶) لان و مک نیل^(۲۷) که بر روی مدل‌های مخاطره متناسب ویژه- علت کاکس مطالعه کرده‌اند، دلایل مرگ را به صورت مرگ در اثر سرطان، بیماری‌های قلبی و سایر بیماری‌ها طبقه بندی کردند. کی ضرائب رگرسیونی برای هر ریسک را به صورت جداگانه و با در نظر گرفتن مرگ از سایر دلایل به صورت سانسور شده برآورد کرد. بیار و گرین^(۲۸) همچنین چن و همکاران^(۲۹) مرگ‌های غیر از سرطان پروستات را در گروهی دیگر طبقه بندی کردند. در این تحقیق ما مرگ‌ها را به صورت مرگ در اثر سرطان پروستات و مرگ در اثر سایر علل تقسیم بندی کرده‌ایم. همچنین اثر متغیرهای کمکی زیر در مدل مورد بررسی قرار گرفت. نوع درمان (دارونما یا ۰/۲ میلی گرم DES به عنوان " گروه درمانی با دز پایین" و با کد صفر، ۱/۰۰ میلی گرم تا ۵/۰۰ میلی گرم DES به عنوان " گروه درمانی با دز بالا" با کد ۱)، Rx، سن بیمار در زمان تشخیص، Age، وزن استاندارد شده (وزن بر حسب کیلوگرم- قد بر حسب سانتی متر +۲۰۰)، Wt، وضعیت عملکرد (فعالیت به صورت طبیعی با کد صفر و حداقل ۵۰٪ اوقات روز در بستر با کد یک)، Pf، سابقه بیماری قلبی (نداشتن سابقه با کد صفر و داشتن سابقه با کد ۱)، Hx، میزان هموگلوبین در ۱۰۰ میلی گرم، Hg، اندازه اولیه تومور بر حسب سانتی متر مربع، Sz، شاخصی ترکیبی از مرحله تومور و درجه هیستولوژی، Sg.

روش اجرای شبیه سازی MCMC

با به کارگیری توزیع پسین بدست آمده^(۱۲) در نرم افزار WinBugs^(۳۰) و در نظر گرفتن توزیع پیشین برای پارامترها به صورت $\alpha \sim Beta(1,1)$ ، $r_j \sim exp(0.001)$ ، $\beta_{kj} \sim Normal(0,100^2)$ ، $k = 0,1$ ، $j = 1,2,\dots,K$ ، و استفاده از نمونه گیری برشی و الگوریتم متروپلیس-هستینگ می توان توزیع پسین جهت پارامترها را بدست آورد. برای اطمینان از همگرایی زنجیره‌های مارکف از معیارهمگرایی گوک^(۳۱) استفاده شده است. پس از اجرای شبیه سازی در نرم افزار WinBugs نتایج حاصله جهت بررسی همگرایی و انجام استنباط توسط بسته BOA^(۳۲) در نرم افزار R^(۳۳) مورد فراخوانی قرار گرفت. تمام شبیه سازی‌ها با استفاده از زنجیره‌هایی به طول ۱۰۰۰۰ و با در نظر گرفتن ۵۰۰۰ مشاهده اول به عنوان دوره تطبیق انجام گرفت.

نتایج و بحث

به منظور یافتن مدل مناسب برای داده‌ها، با استفاده حذف به شیوه پسرو^۱ و کنار گذاشتن متغیرهای غیر معنی‌دار، مدل‌های مختلفی به داده‌ها برازش گردید. برای هر کدام از مدل‌ها مقدار آماره $DIC^{(۳۴)}$ مورد محاسبه قرار گرفت. در کلیه مدل‌ها مقدار DIC با در نظر گرفتن مقادیر اولیه برای پارامترها به صورت $\beta_{kj} = 0$ ، $\alpha = 0.5$ ، $r_1 = r_2 = 1$ و تعداد تکرار برابر ده هزار و در نظر گرفتن پنج هزار مشاهده اول به عنوان دوره تطبیق بدست آمد. مدل‌ها و مقادیر DIC بدست آمده برای هر یک از آنها در جدول (۱) آمده است. براساس معیار DIC مدل با کمترین مقدار DIC بعنوان مدل مناسب در میان مجموعه‌ای از مدل‌ها در نظر گرفته می‌شود. برای مدل‌های جدول (۱)، مقدار ضریب همبستگی تاوکنندال و بازه‌های اطمینان بدست آمده برای آنها در جدول (۲) ارائه شده است.

همانطوریکه در جدول (۱) ملاحظه می‌گردد مدل (۵) دارای کمترین مقدار DIC می‌باشد در نتیجه به عنوان مدل مناسب در نظر گرفته می‌شود. خلاصه‌های آماری پسین پارامترها، شامل برآوردهای نقطه‌ای (میانگین، میانه، انحراف معیار) و فواصل اطمینان (چندک ۲/۵ و ۹۷/۵ درصد) پارامترها برای مدل (۵) در جدول (۳) آورده شده است.

جدول ۱- مدل‌های برازش شده به داده‌ها همراه با مقدار DIC بدست آمده برای آنها

مدل	سرطان پروستات	سایر عوامل	معیار اطلاع DIC	تعداد پارامترها
۱	$Rx \times (Age+Wt+Hx+Hg+Sz+Sg+Pf)$	$Rx \times (Age+Wt+Hx+Hg+Sz+Sg+Pf)$	۳۶۲۳/۷۴	۳۵
۲	$Rx \times (Age+Wt+Hg)+Sz+Sg$	$Rx \times (Age+Wt)+Hx+Sg$	۳۶۰۳/۷۸	۲۱
۳	$Rx \times (Age+Wt)+Hg+Sz+Sg$	$Rx \times (Age+Wt)+Hx+Sg$	۳۶۰۱/۸۹	۲۰
۴	$Rx \times (Age+Wt)+Hg+Sz+Sg$	$Wt+Hx+Sg+Rx \times Age$	۳۵۹۹/۳۷	۱۹
۵	$Rx \times (Age+Wt)+Hg+Sz+Sg$	$Rx \times Age + Wt+Hx$	۳۵۹۸/۸۲	۱۸

جدول ۲- مقادیر ضریب همبستگی تاوکنندال بدست آمده از برازش مدل‌های مختلف به داده‌ها

شماره مدل	مقدار همبستگی تاو کنندال	کران پایین	کران بالا
۱	۰/۰۵۷۷	۰/۰۰۶۴	۰/۱۲۸۲
۲	۰/۰۸۳۸	۰/۰۰۲۸	۰/۲۵۶۴
۳	۰/۰۷۵	۰/۰۰۹۴	۰/۱۶۶۷
۴	۰/۰۹۷۴	۰/۰۰۳	۰/۳۰۰۱
۵	۰/۱۱۴	۰/۰۰۳۹	۰/۳۲۳۳

همانطوریکه بر اساس فواصل اطمینان ارائه شده در جدول (۳) ملاحظه می‌گردد بر روی زمان مرگ در اثر سرطان پروستات عواملی همچون میزان هموگلوبین، اندازه اولیه تومور، شاخص ترکیبی مرحله بیماری و درجه هیستولوژی، اثر متقابل نوع درمان و سن همچنین نوع درمان و وزن موثر می‌باشند. افزایش اندازه اولیه تومور و شاخص ترکیبی مرحله بیماری و درجه هیستولوژی باعث کاهش زمان بقاء و افزایش خطر مرگ بر اثر سرطان پروستات می‌گردد. در حالیکه

جدول ۳- خلاصه‌های آماری پسین بدست آمده بر اساس تحلیل بیزی داده‌های ریسک‌های رقابتی

	پارامترها	میانگین	انحراف معیار	چندک ۲/۵ درصد	چندک ۹۷/۵ درصد
مرگ در اثر سرطان پروستات	مقدار ثابت	-۳/۲۸۸	۰/۸۶۲۸	-۵/۱۲۹	-۱/۸۸۴
	نوع درمان	-۰/۹۴۷۳	۱/۱۳۵	-۳/۰۵۱	۱/۶۷۸
	سن	-۰/۰۴۲۳۸	۰/۰۱۰۱۶	-۰/۶۱۶۴	-۰/۰۲۰۲
	وزن	۰/۰۰۱۰۵۸	۰/۰۰۵۵۳۴	-۰/۱۱۰۷	۰/۰۰۹۳۴۳
	میزان هموگلوبین	-۰/۱۳۶۳	۰/۰۳۹۱۲	-۰/۱۹۹۸	-۰/۰۵۹۰۷
	اندازه اولیه تومور	۰/۰۳۸۴۳	۰/۰۰۵۹۲۸	۰/۰۲۶۷۴	۰/۰۵۰۲۸
	شاخص ترکیبی	۰/۲۶۷۵	۰/۰۴۰۸۴	۰/۱۹۰۹	۰/۳۴۴۶
	نوع درمان × سن	۰/۰۳۸۲۳	۰/۰۱۵۳۸	۰/۰۱۰۴۳	۰/۰۶۵۹۹
	نوع درمان × وزن	-۰/۰۲۴۴۷	۰/۰۰۹۹۵۸	-۰/۰۴۲۴۸	-۰/۰۰۶۳۴۲
	پارامتر شکل	۱/۰۲۴	۰/۸۲۶۹	۰/۸۷۳	۱/۱۹۵
مرگ در اثر سایر علل	مقدار ثابت	-۲/۵۱۹	۱/۲۲۹	-۴/۶۷۸	-۰/۶۳۷
	نوع درمان	-۴/۳۴۳	۱/۰۳۶	-۶/۰۵۴	-۲/۱۶۶
	سن	۰/۰۰۶۳۰۴	۰/۰۱۲۰۴	-۰/۰۱۱۹۸	۰/۰۲۵۴
	وزن	-۰/۰۱۴۵۵	۰/۰۰۵۵	-۰/۰۲۴۴۵	-۰/۰۰۴۳۰۲
	سابقه بیماری قلبی	۰/۶۸۰۱	۰/۱۳۳۷	۰/۴۱۱۱	۰/۹۳۲۷
	نوع درمان × سن	۰/۰۶۰۰۱	۰/۰۱۳۹۱	۰/۰۳۱۵۳	۰/۰۸۲۳۷
	پارامتر شکل	۰/۶۹۰۱	۰/۰۴۲۵۲	۰/۶۰۷۱	۰/۷۷۴۷
	پارامتر همبستگی	۰/۸۸۶	۰/۰۸۷۶۹	۰/۶۷۶۷	۰/۹۹۶۱

افزایش میزان هموگلوبین باعث کاهش خطر مرگ در اثر سرطان پروستات و افزایش زمان بقاء می‌شود. تاثیر نوع درمان نیز بر روی خطر مرگ در اثر سرطان پروستات وابسته به سن و وزن می‌باشد. در مرگ بر اثر سایر علل نیز نوع درمان، وزن، سابقه بیماری قلبی و اثر متقابل سن و نوع درمان معنی‌دار و متغیرهای میزان هموگلوبین، اندازه اولیه تومور و شاخص ترکیبی مرحله بیماری و درجه هیستولوژی اثر معنی‌داری بر روی مرگ در اثر سایر علل نداشتند. شاخص وضعیت عملکرد نیز بر روی هیچکدام از دلایل مرگ اثر معنی‌داری از خود نشان نداد.

در این تحقیق ابتدا چگونگی مدل‌سازی داده‌های ریسک‌های رقابتی با استفاده از توابع مفصل ارائه گردید. استفاده از توابع مفصل برای مدل‌سازی داده‌های ریسک‌هایی رقابتی مزایایی همچون عدم نیاز به در نظر گرفتن ساختاری مستقل برای ریسک‌های رقابتی وابسته، رفع مشکل شناسایی پذیری موجود در مدل‌های کاملاً پارامتری و امکان در نظر گرفتن توزیع‌های حاشیه‌ای متفاوت برای هر کدام ریسک‌ها را فراهم می‌کند. سپس روش بیزی تحلیل

داده‌های ریسک‌های رقابتی پیشنهاد شد که دارای مزایایی همچون نیرومندی در مقابل کم بودن اندازه نمونه و عدم نیاز به استفاده از توزیع‌های مجانبی جهت برآورد پارامترها می‌باشد. به منظور به کار گیری مدل بیزی بر روی داده‌های واقعی از داده‌های معرفی شده توسط اندرو و هرزبرگ^(۲۵) استفاده گردید که بر اساس نتایج بدست آمده شاخص‌های میزان هموگلوبین، اندازه اولیه تومور، شاخص ترکیبی مرحله بیماری و درجه هیستولوژی، اثر متقابل نوع درمان و سن همچنین نوع درمان و وزن بر مرگ بر اثر سرطان پروستات دارای اثری معنی‌دار بودند.

بیار و گرین^(۲۸) نیز در مطالعه خود نشان دادند که تاثیر نوع درمان بر روی زمان بقاء وابسته به سن می‌باشد. در مدلی که آنها برازش کردند اثر متقابل نوع درمان و شاخص ترکیبی مرحله بیماری و درجه هیستولوژی نیز معنی‌دار بود که در مدل بیزی برازش شده به داده‌ها این اثر معنی‌دار نبود. در مطالعه چن و همکاران^(۲۹) نیز نوع درمان، اندازه اولیه تومور، شاخص ترکیبی مرحله بیماری و درجه هیستولوژی اثر معنی‌داری بر روی زمان بقاء سرطان پروستات نشان دادند، اما در این مطالعه هیچگونه اثر متقابلی معنی‌دار نبود.

در مطالعه کی^(۲۶) نیز که مرگ‌ها به صورت مرگ در اثر سرطان و سایر علل تقسیم بندی شده بود، میزان هموگلوبین، اندازه اولیه تومور و شاخص ترکیبی مرحله بیماری و درجه هیستولوژی اثر معنی‌داری بر روی زمان بقاء سرطان نشان دادند. در مطالعه لان و مک نیل^(۲۷) هم نوع درمان، میزان هموگلوبین، اندازه اولیه تومور و شاخص ترکیبی مرحله بیماری و درجه هیستولوژی بر روی خطر مرگ در اثر سرطان معنی‌دار بود.

مقادیر بدست آمده برای پارامترها در این تحقیق بسیار مشابه برآوردهای بدست آمده توسط اسکالر و کریر بود.^(۲) با وجود اینکه آنها از روش‌های کلاسیک و تابع مفصل فرانک جهت برازش مدل به داده‌ها استفاده کردند اما نتایج بدست آمده بسیار مشابه نتیجه حاصل از برازش مدل بیزی و استفاده از تابع مفصل مانای مثبت می‌باشد. تنها تفاوت در نتایج برآوردهای بدست آمده برای پارامتر همبستگی می‌باشد. آنها نشان دادند که با استفاده از تابع مفصل فرانک و مدل‌سازی داده‌ها به روش کلاسیک مقدار همبستگی حاصل از برازش مدل‌های مختلف به داده‌ها با میزان پیچیدگی مدل تغییر می‌کند. به گونه‌ای که دامنه تغییرات ضرائب همبستگی تاوکنندال بدست آمده برای مدل‌های مختلف از $-0/47$ با فاصله اطمینان 95% ($0/15$ ، $-0/70$) تا $0/55$ با فاصله اطمینان 95% ($0/69$ ، $0/34$) بود. در حالیکه در این تحقیق میزان همبستگی حاصل از برازش مدل‌های مختلف به داده‌ها گرچه با میزان میزان پیچیدگی مدل تغییر می‌کند اما تغییرات بسیار کمی دارد.

نتیجه گیری

این کار را می‌توان اولین گام در تحلیل بیزی داده‌های ریسک‌های رقابتی مدل‌سازی شده با استفاده از توابع مفصل در نظر گرفت، چون انعطاف‌پذیری توابع مفصل این امکان را فراهم می‌کند که از سایر توزیع‌های قابل استفاده در تحلیل داده‌های بقاء جهت توزیع‌های حاشیه‌ای استفاده نمود. همچنین حتی جهت انعطاف‌پذیری بیشتر مدل می‌توان حاشیه‌ها را بصورت نیمه‌پارامتری و یا کاملاً ناپارامتری در نظر گرفت. با وجود خانواده‌های مختلفی از توابع مفصل جهت انواع گوناگونی از همبستگی‌ها، می‌توان از سایر خانواده‌های توابع مفصل نیز جهت مدل‌سازی داده‌های ریسک‌های رقابتی استفاده کرد.

References:

1. Gooley, T.A., Leisenring, W., Crowley, J., Storer, B.E., and Leisenring, W., *Stat in Med.*, **18**, 695 (1999).
2. Escarela, G., and Carrier, J.F., *Statist. Med.*, **12**, 333 (2003).
3. Nelsen, R., *An Introduction to Copulas.*, Springer, NewYork (2006).
4. Joe, H. , *Multivariate Models and Dependence Concepts.*, Chapman & Hall, London (1997).
5. Cherubini, U., Luciano, E., and Vecchiato, W., *Copula Methods in Finance*, John Wiley and Sons, NewYork (2004).
6. Oakes, D., *J. A. S. A.*, **84**, 487(1989).
7. Shih, J.H. and Louis, T.A., *Biometrics*, **51**, 1384 (1995).
8. Zheng, M., and Klein, J.P., *Biometrika*, **82**, 127(1995).
9. Carrier, J. , *Transact. Soci. Actuaries*, **XLVI**, 45(1994).
10. Kiashev, V.K., Dimitrina, D.S., and Hamberman, S., *Ins. Math. Econ.*, **41**, 339 (2007).
11. He, W., and Lawless, J.F., *Biometrics*, **59**, 837 (2003).
12. Gustafson, P., Aeschliman, D., Levy, and AR., *Lifetime Data Anal.*, **9**, 5 (2003).
13. Romeo, J.S., Tanaka, N.I., and Pedroso-de-Lina, A.C., *Lifetime Data Anal.*, **12**, 205 (2006).
14. Genest, C.M.J., *Amer. Statistician*, **40**, 280 (1986).
15. Hougaard, P., *Biometrika*, **73**, 387 (1986).
16. Hougaard, P., *Biometrika*, **73**, 671 (1986).
17. Frank, M.J., *Aequa. Math.*, **19**, 194 (1979).
18. Clayton, D.G., *Biometrika*, **65**, 141 (1978).
19. Eldant-Johnson, R.C., *J. N. Survival Models and Data Analysis*, Wiley, NewYork (1980).
20. Carlin, B.P., Louis, T., *Bayes and Emprical Bayes Methods for Data Analysis*, chapman and Hall, NewYork (2000).
21. Gilks, W.R., *Bayesian Statistics 4*, eds J.M. Bernardo, Berger, J.O., Dawid, A.P., and Smith, A.F.M., Oxford University Press, NewYork (1992).
22. Gilks, W.R., *A. W. P. Appl. Statist.*, **41**, 337(1992).
23. Gilks, W.R., R.S., Spiegelhalter, D.J., *Markov Chain Monte Carlo in Practice*, Chapman and Hall, London (1996).
24. Neal, R., *Ann. Stat.*, **31**, 705 (2003)
25. Andrews, D.F., and Herzberg, A.M., *Data: A Collection of Problems from Many Fields for the Student and Research Worker*, Springer, NewYork (1985).
26. Kay, R., *Biometrics*, **42**, 203 (1986).
27. Lunn, M., and McNeil, D., *Biometrics*, **51**, 524 (1995).
28. Byar, D.P., and Green, S., *Bulletin du Cancer*, **67**, 477 (1980).
29. Chen, S.C., Fine, J.P., and Wei, L.J., *Biometrics*, **54**, 219 (1998).
30. Spiegelhalter, D.J., Thomas, A., and Best, N.G., *WinBugs 1.4. Computer Program. Imperical College AND MRC Biostatistics unit*, IPH. Cambridge, UK (2003).
31. Geweke, J., In: Bernarod, J.M., Berger, J.O., Dawid, A.P., and Smith, A.F.M., *Bayesian Statistics, vol 4.*, Oxford University Press, Oxford (1992).
32. Smith, BG., <http://www.public-health.Uiowa.edu/boa/> (2005).
33. Venables, W.N., .Smith, D.M., <http://www.r-project.org/> (2009).
34. Spiegelhalter, D.J., Best, N.G., Carlin, B.P., and van der Linde, A., *J. Roy. Statist. Soc. B*, **64**, 583 (2002).

Surf and download all data from SID.ir: www.SID.ir

Translate via STRS.ir: www.STRS.ir

Follow our scientific posts via our Blog: www.sid.ir/blog

Use our educational service (Courses, Workshops, Videos and etc.) via Workshop: www.sid.ir/workshop