مجله‌ی پژوهش‌های آماری ایران
پژوهشکده‌ی آمار

دوره‌ی ۶، شماره‌ی ۱، بهار و تابستان ۱۳۸۸، صص ۱۲۱–۱۳۹

J. Statist. Res. Iran **6** (2009): 121–139

رهیافت درستنمایی تجربی وکاربرد آن در تحلیل بقا

محدثه صفاکیش* و حمیدرضا نواب‌پور

دانشگاه علامه‌طباطبایی

چکیده. در بیش‌تر مطالعه‌هایی که صورت می‌گیرد، علاقه‌مند به استنباط درباره‌ی توزیع جامعه و پارامترهای آن هستیم. از جمله روش‌های معمول برای براورد پارامترهای جامعه –زمانی‌که توزیع معلوم است– روش ماکسیمم درستنمایی است. براوردهای حاصل از این روش در حالت حدی ویژگی‌های مطلوب بسیاری دارند. نااریبی، واریانس مینیمم و توزیع نرمال به همراه روش دلتا استنباط پیرامون براوردگرها و توابع آن‌ها را میسر می‌سازد. در صورت ناشناخته بودن توزیع جامعه، روش‌های ناپارامتری بسیاری وجود دارند. برخی از این روش‌ها مانند روش بازنمونه‌گیری خودگردان[۱]، بر پایه‌ی تکرار نمونه‌گیری از یک نمونه‌ی اولیه پایه‌ریزی شده‌اند. در این مقاله رهیافت ناپارامتری درستنمایی تجربی به‌منظور استفاده‌ی بهتر از اطلاعات کمکی برای استنباط درباره‌ی پارامترهای جامعه معرفی و چگونگی ساختن ناحیه‌ی اطمینان با استفاده از آن بیان می‌شود. همچنین استفاده از این روش در براورد مدل رگرسیون میانه‌ی زمان شکست در حضور داده‌های سانسور شده از راست نشان داده می‌شود و ناحیه‌ی اطمینان بردار پارامترهای مدل و زیرمجموعه‌ی دلخواه از آن‌ها به دست می‌آید. سرانجام مدل رگرسیون میانه‌ی طول عمر برای مجموعه‌ی داده‌های مربوط به بیماران مبتلا به سرطان مغز استخوان براورد می‌شود. همچنین ناحیه‌ی اطمینان برای بردار پارامترها و بازه‌های اطمینان برای تک تک ضریب‌های رگرسیونی به دست می‌آید.

واژگان کلیدی. درستنمایی تجربی؛ معادله‌ی براورد؛ براوردگر کاپلان–مه‌یر؛ رگرسیون میانه؛ سانسور از راست؛ درستنمایی تجربی نیمرخ؛ روش خودگردان درستنمایی تجربی.

*نویسنده‌ی عهده‌دار مکاتبات

۱ مقدمه

استنباط درباره‌ی توزیع جامعه و پارامترهای آن در دو حالت پارامتری (توزیع جامعه معلوم) و ناپارامتری (توزیع نامعلوم) صورت می‌گیرد. روش ماکسیمم درستنمایی از جمله روش‌های پارامتری معمول برای براورد پارامترهای توزیع یک جامعه است. ویژگی‌هایی چون نااریبی، واریانس مینیمم و نرمال بودن به همراه روش دلتا استنباط پیرامون براوردگرها و توابع آن‌ها را با استفاده از این روش میسر می‌سازد. در این روش تابع درستنمایی از روی تابع توزیع جامعه به دست می‌آید.

در حالتی‌که توزیع جامعه مشخص نیست و امکان نرمال بودن آن نیز توسط آزمون‌های نیکویی برازش تأیید نمی‌شود و حتی تبدیل‌های نرمال کننده‌ی داده‌ها هم مؤثر واقع نشوند، استفاده از تقریب نرمال برای ساختن بازه‌ی اطمینان و آزمون فرض، درستی استنباط‌ها را مورد تردید قرار می‌دهد. در چنین شرایطی بهتر است از روش‌های مشابه ناپارامتری برای استنباط پیرامون پارامترهای جامعه استفاده شود. برخی از این روش‌ها مانند روش بازنمونه‌گیری خودگردان، بر پایه‌ی تکرار نمونه‌گیری از یک نمونه‌ی اولیه پایه‌ریزی شده‌اند. در مقابل همه‌ی روش‌های باز نمونه‌گیری، درستنمایی تجربی به عنوان یك روش ناپارامتری زمانی که توزیع جامعه نامشخص است برای ساختن ناحیه‌های اطمینان و آزمون فرض روی پارامترهای مورد نظر جامعه ابداع شده است. استفاده از نسبت درستنمایی تجربی برای نخستین بار توسط توماس و گرانکه‌مه‌یر (۱۹۵۷) به کار رفته است. گسترش این نظریه توسط اُئن (۱۹۸۸ و ۱۹۹۰) انجام شد. وی ناحیه‌ی اطمینان درستنمایی تجربی را برای میانگین یك متغیر تصادفی بر اساس مشاهده‌های مستقل و هم‌توزیع ارایه داد. اُئن (۱۹۸۸) تابع نسبت درستنمایی تجربی را در حالت یك متغیره تعریف کرد و از آن برای ساختن بازه‌های اطمینان برای میانگین، چندک‌ها و تابع‌های مشتق‌پذیر آن‌ها استفاده کرد. اُئن در سال ۱۹۹۰، تعمیمی از این مسئله را در حالت چندمتغیره برای تابع‌های بردار مقدار ارایه داد. پس از آن افراد بسیاری در زمینه‌های مختلف از این روش استفاده کردند از جمله‌ی آن‌ها تچین و تسائو (۲۰۰۳) هستند که توزیع حدی آماره‌ی نسبت درستنمایی تجربی بردار پارامترهای مدل رگرسیون میانه را به‌صورت مجموع موزون توزیع‌های خی‌دوی استاندارد به دست آوردند، سپس با مد نظر قرار دادن این ویژگی به استنباط پیرامون بردار پارامترها تحت داده‌های بقای سانسور شده از راست[۲] پرداختند. سوبرامانیان (۲۰۰۷) به‌منظور محاسبه‌ی ناحیه‌ی اطمینان برای زیرمجموعه‌ی دلخواه از پارامترهای مدل رگرسیون میانه، آماره‌ی نسبت درستنمایی تجربی نیم‌رخ را تعریف کرد و توزیع حدی آن را به دست آورد.

در بخش دوم، رهیافت درستنمایی تجربی برای استنباط پیرامون پارامترهایی که به‌صورت تابع براورد نااریب بیان می‌شوند؛ ارائه شده سپس آماره‌ی نسبت درستنمایی تجربی برای پارامتر میانگین، در حالت

یک متغیره و چند متغیره به‌عنوان حالت خاصی از آن، معرفی می‌شود. در بخش سوم مدل رگرسیون میانه‌ی زمان شکست در حضور داده‌های سانسور شده از راست معرفی شده و روش درستنمایی تجربی برای ساختن ناحیه‌ی اطمینان بردار پارامترهای مدل به کار برده می‌شود. در بخش آخر نتایج مربوط به برازش مدل رگرسیون میانه بر روی مجموعه داده‌های پیوند مغز استخوان و ناحیه‌ی اطمینان بردار پارامترهای این مدل ارایه می‌شود.

۲ درستنمایی تجربی[۳]

درستنمایی تجربی یک رهیافت ناپارامتری است که از اطلاعات کمکی موجود درباره‌ی جامعه استفاده کرده و توزیع را به‌گونه‌ای براورد می‌کند که بیش‌ترین مقدار درستنمایی را داشته باشیم. در این روش تابع درستنمایی بدون نیاز به دانستن توزیع جامعه، بر اساس احتمال‌های نسبت داده‌شده به هر یک از مشاهده‌ها ساخته می‌شود. هال و اسکالا (۱۹۹۰)، برخی از مهم‌ترین ویژگی‌های روش درستنمایی تجربی را به‌صورت زیر معرفی می‌کنند.

۱. ناحیه‌های درستنمایی تجربی نیاز به براورد شاخص‌های خاص نظیر چولگی و . . . ندارند، در حالی‌که در روش خودگردان تصحیح اریبی شتابیده[۴] نیاز به براوردی از چولگی است و روش خودگردان صدک-t بدون داشتن براورد کارا از خطای استاندارد براوردگر پارامتر مورد نظر به نتیجه نمی‌رسد.

۲. ناحیه‌ی اطمینان در این روش بدون نیاز به داشتن کمیت محوری ساخته می‌شود. این خصوصیت در مسائلی که دست‌یابی به واریانس براوردگر مشکل است، -مثل حالتی‌که می‌خواهیم برای ضریب همبستگی بازه‌ی اطمینان بسازیم- مزیت بزرگی به نظر می‌رسد.

۳. ناحیه‌های اطمینان درستنمایی تجربی حافظ دامنه‌ی تغییرات پارامتر بوده و قابل تبدیل هستند. حافظ دامنه بودن[۵] به این معنا است که ناحیه‌ی به دست آمده از روش درستنمایی تجربی زیر مجموعه‌ی ناسره[۶] از دامنه‌ی تغییرات پارامتر است. برای مثال ناحیه‌ای که برای ضریب همبستگی به دست می‌آید حتماً در فاصله‌ی بین ۱- تا ۱+ قرار می‌گیرد. قابل تبدیل بودن یعنی برای ساختن ناحیه‌ی اطمینان برای پارامتر $g(\theta)$ کافی است تابع $g(\cdot)$ روی ناحیه‌ی به دست آمده برای پارامتر θ اثر داده شود. دارا بودن این دو ویژگی برای ناحیه‌های درستنمایی تجربی در حالی است که ناحیه‌های خودگردان صدک-t هیچ یك از این ویژگی‌ها را ندارند، و روش تصحیح اریبی شتابیده نیز تنها تحت تبدیل‌های پایای پارامتر بدون تغییر است.

۴. شکل ناحیه‌ی درستنمایی تجربی از روی داده‌ها به دست می‌آید. بنا بر این نیازی به تعیین شکل ناحیه قبل از ساختن ناحیه‌ی اطمینان مثل روش‌های خودگردان و تقریب نرمال نیست.

اُئن (۱۹۸۸، ۱۹۹۰ و ۱۹۹۱) با استفاده از ایده‌ی مطرح‌شده توسط توماس و گرانکه‌مه‌یر (۱۹۵۷) آماره‌ی نسبت درستنمایی تجربی را برای مسائل ناپارامتری بیان کرده است. وی همچنین نشان داده که این آماره دارای توزیع خی‌دو است. علاوه بر این چگونگی یافتن ناحیه‌ی اطمینان و آزمون برای پارامترهایی که به‌صورت تابعی چون $\theta(F)$ از تابع توزیع F قابل بیان هستند را مطرح کرده است. ویژگی‌های مجانبی و تصحیح‌هایی که برای آماره‌ی نسبت درستنمایی تجربی وجود دارند توسط دیسیچیو و همکاران (۱۹۸۹)، هال و اسکالا (۱۹۹۰) بیان و بررسی شده‌اند.

۲٫۱ درستنمایی تجربی و معادله‌های براورد ناااریب

فرض کنید $x_1, \ldots, x_n$ مشاهده‌های مستقل و هم‌توزیع (i.i.d.) از توزیع نامعلوم F باشند و θ بردار پارامتری pبعدی مورد نظر جامعه باشند. در این حالت تابع درستنمایی تجربی به‌صورت زیر تعریف می‌شود (اُئن، ۱۹۹۰):

$$L(F) = \prod_{i=1}^{n} p_i = \prod_{i=1}^{n} dF(x_i).$$

همان‌طور که می‌دانید هر یک از واحدها با احتمال مثبتی در نمونه قرار می‌گیرند. چنان‌چه p_i را به‌عنوان احتمال مشاهده‌ی x_i در نظر بگیریم، خواهیم داشت: $\sum_{i=1}^{n} p_i = 1$ و $p_i \geqslant 0$. با در نظر گرفتن این دو قید براورد ماکسیمم درستنمایی تابع توزیع جامعه با استفاده از فن ضرایب لاگرانژ به ازای $p_i = \frac{1}{n}$، برابر با تابع توزیع تجربی $F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I(x_i \leqslant x)$ به دست می‌آید. با توجه به این مطلب تابع نسبت درستنمایی تجربی به‌صورت زیر تعریف می‌شود:

$$R(F) = \frac{L(F)}{L(F_n)} = \frac{\prod_{i=1}^{n} p_i}{\prod_{i=1}^{n} \frac{1}{n}} = \prod_{i=1}^{n} np_i. \qquad (1)$$

در بسیاری از مسائل، اطلاعات کمکی درباره‌ی پارامترها را می‌توان با r تابع براورد ناااریب که $r \geqslant p$ است، به‌صورت زیر بیان کرد.

$$g(x,\theta) = \{g_1(x,\theta), \ldots, g_r(x,\theta)\}, \qquad E_F\{g(x,\theta)\} = 0$$

در این حالت تعداد تابع‌های براورد (r) بیش‌تر از تعداد پارامترها (p) است. به‌منظور استفاده‌ی بهتر از اطلاعات موجود پیرامون جامعه رابطه‌ی $\sum_{i=1}^{n} p_i g(x_i, \theta) = 0$ در نظر گرفته می‌شود. رهیافت درستنمایی تجربی با مد نظر قرار دادن این قید علاوه بر دو شرط قبلی، براورد مناسب‌تری از p_iها به دست می‌دهد. بنا بر این مسئله‌ی یافتن p_iها به گونه‌ای است که تابع درستنمایی تجربی ماکسیمم شود. به عبارت دیگر:

$$\max L(F) = \max_{\{p_i \geqslant 0, \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i g(x_i,\theta) = 0\}} \prod_{i=1}^{n} p_i$$

با به‌کار بردن فن ضرایب لاگرانژ احتمال‌های p_i ، $i = 1, \ldots, n$ به‌صورت زیر حاصل می‌شوند:

$$p_i = \frac{1}{n} \left\{ \frac{1}{1 + t'(\theta) g(x_i, \theta)} \right\} \qquad i = 1, \ldots, n, \tag{2}$$

که در آن $t'(\cdot)$ بردار ضرایب لاگرانژ است و از حل معادله‌ی زیر با استفاده از روش عددی الگوریتم نیوتن–رافسون به دست می‌آید.

$$\sum_{i=1}^{n} p_i g(x_i, \theta) = \frac{1}{n} \sum_{i=1}^{n} \frac{g(x_i, \theta)}{1 + t'(\theta) g(x_i, \theta)} = 0.$$

با جایگذاری مقدار p_i در رابطه‌ی (۱) لگاریتم تابع نسبت درستنمایی تجربی نیم‌رخ به‌صورت زیر تعریف می‌شود:

$$\begin{aligned} \log R(F) &= \log \prod_{i=1}^{n} \left\{ 1 + t'(\theta) g(x_i, \theta) \right\}^{-1} \\ &= -\sum_{i=1}^{n} \log \left\{ 1 + t'(\theta) g(x_i, \theta) \right\} = -\ell_E(\theta). \end{aligned}$$

برای به دست آوردن $\tilde{\theta}$، براورد ماکسیمم درستنمایی تجربی پارامتر θ، تابع $\ell_E(\theta)$ نسبت به θ مینیمم می‌شود. سپس این مقدار در رابطه‌ی (۲) جایگذاری شده و براورد ماکسیمم درستنمایی تجربی برای p_i به‌صورت زیر به دست می‌آید:

$$\tilde{p}_i = \frac{1}{n} \left\{ \frac{1}{1 + t'(\tilde{\theta}) g(x_i, \tilde{\theta})} \right\} \qquad i = 1, \ldots, n.$$

سرانجام این مقدار را به‌عنوان براورد جرم احتمال در نقطه‌ی x_i در نظر گرفته و براورد ماکسیمم درستنمایی تجربی تابع توزیع به‌صورت زیر حاصل می‌شود:

$$\tilde{F}_n(x) = \sum_{i=1}^{n} \tilde{p}_i I(x_i \leqslant x) = \sum_{i=1}^{n} \frac{I(x_i \leqslant x)}{1 + t'(\tilde{\theta}) g(x_i, \tilde{\theta})}.$$

براوردگرهای درستنمایی تجربی پارامتر θ و تابع توزیع F دارای توزیع مجانبی نرمال هستند (تچین و لاولس، ۱۹۹۴). یکی دیگر از ویژگی‌های مشترک میان نسبت درستنمایی تجربی و پارامتری، توزیع مجانبی خی‌دو برای این آماره‌ها است. آماره‌ی نسبت درستنمایی تجربی برای آزمون فرض صفر $H_۰ : \theta = \theta_۰$، به‌صورت رابطه‌ی زیر تعریف می‌شود. این آماره با فرض برقراری شرط‌های خاصی دارای توزیع مجانبی خی‌دو با p (تعداد پارامترها) درجه‌ی آزادی است (تچین و لاولس، ۱۹۹۴).

$$W_E(\theta_۰) = ۲\{\ell_E(\theta_۰) - \ell_E(\tilde{\theta})\}$$

برای مثال فرض کنید علاقه‌مند به استنباط پیرامون میانگین جامعه، μ، باشیم. در این صورت تابع براورد نااریب به‌صورت $g(x, \mu) = \bar{x} - \mu$ تعریف شده و قید مورد استفاده در فرایند ماکسیمم‌سازی با تابع $\sum_{i=۱}^{n} p_i x_i = \mu$ جایگزین می‌شود. همچنین برای تعمیم این نظریه به مشاهده‌های مستقل و ناهم‌واریانس[۷] در تحلیل رگرسیونی، درستنمایی تجربی موزون ابداع شده است (وو، ۲۰۰۴).

۳ مدل رگرسیون میانه[۸]

در تحلیل بقا برای استفاده از اطلاعات متغیرهای کمکی معمولاً با در نظر گرفتن مدلی میان متغیر زمان شکست و متغیرهای کمکی و برازش مدل به داده‌های موجود با در نظر گرفتن شرط‌های خاصی به پیش‌بینی زمان شکست پرداخته می‌شود. در میان تمام روش‌هایی که در این زمینه وجود دارد مدل رگرسیون میانه زمانی که توزیع زمان شکست چوله است، جایگزین مناسبی برای رگرسیون میانگین است. در این مدل‌ها میانه‌ی زمان شکست (به‌جای میانگین آن‌ها) را به‌عنوان متغیر پاسخ در نظر گرفته و آن را طبق رابطه‌ی خطی زیر

$$m_i = \mathrm{med}(T_i|Z_i) = \beta' Z_i + \varepsilon_i$$

در ارتباط با بردار متغیرهای کمکی Z_i قرار می‌دهند. در این رابطه $Z_i = (۱, X_i)$ بردار $p + ۱$-بعدی شامل متغیرهای کمکی و β بردار پارامترهای نامعلوم رگرسیون است. خطاهای ε_i متغیرهای تصادفی مستقل هستند که با فرض معلوم بودن $Z_i = z_i$ دارای چگالی نامعلوم $f_\varepsilon(\varepsilon_i|z_i)$ و میانه‌ی صفر می‌باشند. از آن‌جا که برخی از واحدها سانسور می‌شوند؛ بنا بر این آن‌چه را که می‌توان در عمل مشاهده کرد، به‌صورت زیر نمایش داده می‌شود:

$$Y_i = \min(C_i, T_i), \quad \delta_i = I\{T_i \leqslant C_i\}, \qquad i = ۱, \ldots, n$$

که در آن C_i، زمان سانسور واحد iام با توزیع نامعلوم $G(\cdot)$ و تابع بقای $1-G(\cdot)$ است و تابع نشان‌گر δ_i وضعیت سانسور یا شکست واحد iام را مشخص می‌کند. همچنین فرض می‌شود زمان شکست و زمان سانسور شده از یکدیگر مستقل هستند ($C_i \perp T_i$). به عبارت دیگر سانسور داده‌ها در این مدل گمشدگی کاملاً تصادفی[9] در نظر گرفته می‌شود. برای استفاده از روش درستنمایی تجربی تابع براورد ناااریب زیر را در نظر می‌گیریم (تچین و تسائو، ۲۰۰۳):

$$S_n(\beta)=\sum_{i=1}^{n} Z_i\left\{\frac{I(Y_i\geq \beta' Z_i)}{1-\hat{G}(\beta' Z_i)}-\frac{1}{2}\right\}\approx 0 \tag{3}$$

که در آن $1-\hat{G}(\cdot)$ براورد کاپلان-مه‌یر برای تابع بقای متغیر سانسور است و به‌صورت زیر به دست می‌آید:

$$1-\hat{G}(t)=\prod_{u\leq t}\left\{1-\frac{\Delta N^c(u)}{Y(u)}\right\},$$

که در آن

$$\Delta N^c(u)=\sum_{i=1}^{n} I(Y_i\leqslant u,\delta_i=0),$$
$$Y(u)=\sum_{i=1}^{n} I(Y_i\geqslant u).$$

به‌دلیل ناپیوستگی تابع $S_n(\beta)$، براورد نقطه‌ای $\hat{\beta}$ (ریشه‌ی تابع) با به‌کارگیری یک روش جستجوی شبکه‌ای[10] به دست می‌آید. یافتن ریشه‌ی این تابع معادل به دست آوردن مقداری از بردار پارامترهاست که به ازای آن نرم تابع برداری $S_n(\beta)$ مینیمم می‌شود (یینگ و همکاران، ۱۹۹۵). در این روش مقدارهای ممکن برای پارامتر β به‌صورت یک شبکه‌ی $[a,b]$ در نظر گرفته می‌شود. سپس به ازای مقدارهای مختلف درون این شبکه، مقدار تابع مورد نظر را محاسبه می‌کنیم. از میان مقدارهای مختلف β، آن‌که تابع را بهینه می‌کند (در این‌جا کم‌ترین مقدار نرم تابع $S_n(\beta)$ را می‌دهد)، پاسخ مورد نظر است. تابع درستنمایی تجربی با در نظر گرفتن معادله‌ی براورد (۳) به ازای β_0 (مقدار واقعی پارامتر β) به‌صورت زیر تعریف می‌شود:

$$L(\beta_0)=\left\{\sup\prod_{i=1}^{n} p_i : p_i\geqslant 0,\sum_{i=1}^{n} p_i=1,\sum_{i=1}^{n} p_i W_{ni}=0\right\},$$

$$W_i(\beta_0,\hat{G})=W_{ni}=Z_i\left\{\frac{I(Y_i\geqslant \beta_0' Z_i)}{1-\hat{G}(\beta_0' Z_i)}-\frac{1}{2}\right\}.$$

مشابه روش به کار رفته در بخش قبل با استفاده از فن ضریب‌های لاگرانژ، p_i از رابطه‌ی زیر به دست می‌آید:

$$p_i = \frac{۱}{n}(۱ + \lambda' W_{ni})^{-۱}, \qquad i = ۱, \ldots, n,$$

و بردار ضریب‌های لاگرانژ λ از حل معادله‌ی زیر با استفاده از الگوریتم نیوتن-رافسون به دست می‌آید.

$$\sum_{i=۱}^{n} p_i W_{ni} = \sum_{i=۱}^{n} \frac{۱}{n} \left(۱ + \lambda' W_{ni}\right)^{-۱} = ۰.$$

مشابه بخش قبل ضریب (۲) برابر منفی لگاریتم تابع درستنمایی تجربی به صورت زیر به دست می‌آید:

$$\begin{aligned}\ell(\beta_۰) &= -۲ \log \prod_{i=۱}^{n} (۱ + \lambda' W_{ni})^{-۱} \\ &= ۲ \sum_{i=۱}^{n} \log(۱ + \lambda' W_{ni}).\end{aligned}$$

این تابع در حالت حدی با مجموع موزون از متغیرهای تصادفی خی‌دو به‌صورت زیر هم‌توزیع است (تچین و تسائو، ۲۰۰۳):

$$\ell(\beta_۰) \xrightarrow{L} \sum_{i=۱}^{p+۱} \hat{k}_i \chi^۲_{i,۱}$$

که در آن $\hat{k}_i$ها ویژه‌مقدارهای ماتریس کوواریانس حدی $\hat{\Gamma}_۱^{-۱}\hat{\Gamma}$ هستند و $\chi^۲_{i,۱}$ متغیرهای تصادفی مستقل خی‌دو هر کدام با یک درجه‌ی آزادی می‌باشند. ماتریس‌های $\hat{\Gamma}_۱^{-۱}$ و $\hat{\Gamma}$ به‌صورت زیر تعریف می‌شوند:

$$\hat{\Gamma}_۱ = \frac{۱}{n} \sum_{i=۱}^{n} \left\{ \frac{I(Y_i \geqslant \hat{\beta}' Z_i)}{۱ - \hat{G}(\hat{\beta}' Z_i)} - \frac{۱}{۲} \right\}^۲ Z_i Z_i', \tag{۴}$$

$$\hat{\Gamma}_۲ = \frac{۱}{۴n} \sum_{i=۱}^{n} (۱ - \delta_i) \left\{ \frac{\sum_{j=۱}^{n} Z_j I(\hat{\beta}' Z_j \geqslant Y_i)}{\sum_{j=۱}^{n} I(Y_j \geqslant Y_i)} \right\}^{\otimes ۲}$$

$$\hat{\Gamma} = \hat{\Gamma}_۱ - \hat{\Gamma}_۲ \tag{۵}$$

که برای هر بردار دلخواه a داریم: $a^{\otimes ۲} = aa'$. بنا بر این ناحیه‌ی اطمینان برای بردار پارامترهای مدل رگرسیونی به‌صورت رابطه‌ی زیر به دست می‌آید (تچین و تسائو، ۲۰۰۳):

$$R_\alpha(\beta) = \left\{ \beta : \ell(\beta) \leqslant c_\alpha \right\}, \tag{۶}$$

که c_α چندک $\alpha-$ ۱ام توزیع $\sum_{i=1}^{p+1}\hat{k}_i\chi^2_{i,1}$ است. همچنین به‌منظور مطالعه‌ی زیرمجموعه‌ی دلخواه از پارامترهای مدل مانند $\beta^{(1)}_{q\times 1}$، آماره‌ی نسبت درستنمایی تجربی نیم‌رخ به‌صورت زیر تعریف می‌شود:

$$\ell_s(\beta^{(1)}) = \min_{\beta^{(2)}\in N} \ell(\beta^{(1)},\beta^{(2)}),$$

$$N = \left\{\beta^{(2)} : \left\|\beta^{(2)} - \hat{\beta}^{(2)}\right\| = O\left(\frac{1}{\sqrt[3]{n}}\right)\right\}$$

که در آن $\|\cdot\|$ نرم اقلیدسی و $\hat{\beta}^{(2)}$ براورد نقطه‌ای برای زیر بردار $\beta^{(2)}_{(p-q)\times 1}$ است که با به کار بردن روش جستجوی شبکه‌ای از حل معادله‌ی $S_n(\beta^{(1)},\beta^{(2)}) = 0$ به دست می‌آید. این آماره دارای توزیع حدی مجموع موزون از متغیرهای تصادفی خی‌دوی مستقل است ولی به‌دلیل نبود براورد مناسب برای ماتریس واریانس-کوواریانس حدی، سوبرامانیان (۲۰۰۷) توزیع نمونه‌گیری این آماره را با استفاده از روش خودگردان تولید کرده و از روی آن مقدار بحرانی مورد نیاز برای ساختن ناحیه‌ی اطمینان را محاسبه می‌کند. چگونگی ساختن ناحیه‌ی اطمینان در بخش بعد می‌آید.

۳٫۱ درستنمایی تجربی نیم‌رخ خودگرداننده[11]

برای انجام آزمون فرض $H_0 : \beta^{(1)} = \beta^{(1)}_0$ در برابر فرض کلی $H_1 : \beta^{(1)} \neq \beta^{(1)}_0$ از نمونه‌ی اصلی (Y_i,δ_i,Z_i)، $i = 1,\dots,n$، به‌روش تصادفی ساده‌ی باجایگذاری به تعداد b بار نمونه به اندازه‌ی n گرفته و نمونه‌های تکراری را با (Y_i^*,δ_i^*,Z_i^*) نمایش می‌دهیم. پس از محاسبه‌ی مقدارهای $W_i^*((\beta^{(1)}_0,\beta^{(2)}),\hat{G})$، که در آن $\hat{G}$ براوردگر کاپلان-مه‌یر بر مبنای نمونه‌ی خودگردان است؛ دنباله‌ی مقدارهای $\ell^*(\beta^{(1)}_0,\beta^{(2)})$ را به ازای هر نمونه‌ی خودگردان محاسبه می‌کنیم. در گام بعد با مینیمم کردن $\ell^*(\cdot)$ نسبت به $\beta^{(2)}\in N$، نسبت درستنمایی تجربی نیم‌رخ خودگردان $\ell_s(\beta^{(1)})$، برای هر بازنمونه[12] به دست می‌آید. ازکنار هم قرار دادن این مقدارها توزیع نمونه‌گیری آماره‌ی نسبت درستنمایی تجربی نیم‌رخ حاصل می‌شود. پس از مرتب کردن مقدارهای به دست آمده در یک ترتیب افزایشی، مقدار $b(1-\alpha)$ام را به دست آورده و با نماد $b_{(1-\alpha)}$ نشان می‌دهیم. در واقع مقدار بحرانی مورد نظر همان چندک $(1-\alpha)$ام توزیع نمونه‌گیری نسبت درستنمایی تجربی نیم‌رخ است. بنا بر این ناحیه‌ی اطمینان برای زیر بردار $\beta^{(1)}_{q\times 1}$ به‌صورت زیر حاصل می‌شود (سوبرامانیان، ۲۰۰۷):

$$R_\alpha(\beta^{(1)}_0) = \left\{\beta^{(1)} : \ell_s(\beta^{(1)}) \leqslant b_{(1-\alpha)}\right\}$$

۴ کاربرد

در این بخش مدل رگرسیون میانه‌ی زمان شکست روی مجموعه‌ی داده‌های پیوند مغز استخوان[۱۳] تولید شده توسط آوالاس و همکاران (۱۹۹۳) مورد بررسی قرار می‌گیرد. داده‌ها شامل اطلاعات مربوط به ۴۳ بیمار مبتلا به نوعی سرطان مغز استخوان به نام لیمفوما[۱۴] است که در واحد پیوند مغز استخوان دانشگاه ایالتی اُهایو مورد مطالعه و بررسی قرار گرفته‌اند.

در علم تومورشناسی[۱۵] روش معمول برای بازسازی سلول‌های مغز استخوان از بین رفته در اثر شیمی‌درمانی، انجام عمل پیوند مغز استخوان است. در این بررسی مدت زمان سپری‌شده پس از دریافت سلول‌های پیوندی تا زمان مرگ بیمار یا رد پیوند توسط بدن بیمار، به‌عنوان زمان شکست در نظر گرفته می‌شود. همان‌طور که در بخش سوم بیان شد، در هر مطالعه‌ی تحلیل بقا علاوه بر متغیر پاسخ (زمان شکست) متغیر نشان‌گری که بیان‌گر وضعیت بیمار است، اندازه‌گیری می‌شود. در این مطالعه نشان‌گر δ_i به‌صورت زیر تعریف می‌شود.

$$\delta_i : \begin{cases} ۱ & \text{بیمار } i\text{ام مرده یا پیوند را پس زده است} \\ ۰ & \text{بیمار } i\text{ام از مطالعه خارج شده است} \end{cases}$$

به همراه دو متغیر بیان‌شده در بالا، متغیرهای کمکی زمان انتظار برای دریافت پیوند مغز استخوان پس از تشخیص بیماری ($X_۲$ بر حسب ماه) و شاخص کارنوفسکی[۱۶] بیمار قبل از عمل پیوند ($X_۱$)، برای هر یک از ۴۳ بیمار تحت مطالعه اندازه‌گیری و ثبت شدند.

در علم تومور شناسی مطالعه‌ی چگونگی اثر گذاری این دو متغیر کمکی روی میانه‌ی طول عمر بیمار پس از پیوند مغز استخوان اهمیت بسزایی دارد. بنا بر این با هدف مطالعه‌ی اثر این دو متغیر مدل رگرسیون میانه بر اساس مطالب بیان‌شده در بخش سوم به‌صورت زیر درنظر گرفته می‌شود:

$$T_i = \beta_۰ + \beta_۱ X_{۱i} + \beta_۲ X_{۲i} + \epsilon_i, \qquad i = ۱, \ldots, ۴۳$$

که در آن متغیر T_i لگاریتم زمان شکست، β_j، $(j = ۰, ۱, ۲)$ پارامترهای نامعلوم مدل و ϵ_iها خطاهای مدل و دارای ویژگی‌های معرفی‌شده در قبل هستند.

در این بخش براورد نقطه‌ای ضریب‌های رگرسیونی مدل بالا با استفاده از روش معرفی‌شده در بخش سوم به دست می‌آیند. سپس ناحیه‌ی اطمینان ۹۵ درصد بردار پارامترها (β) بر اساس رابطه‌ی (۶) ساخته می‌شود. در انتها بازه‌ی اطمینان ۹۵ درصد برای تک تک پارامترهای مدل به‌روش درستنمایی تجربی نیم‌رخ (معرفی شده در بخش قبل) محاسبه می‌شوند.

۴٫۱ براورد نقطه‌ای پارامترها

براورد نقطه‌ای پارامترهای مدل رگرسیون میانه از حل معادله‌ی (۳) به دست می‌آیند. دامنه‌ی تغییرات پارامترها و مقدار اولیه‌ی مورد نیاز برای اجرای این روش با استفاده از تابع crq که یکی از تابع‌های بسته‌ی quantreg در نرم‌افزار R است، به‌صورت زیر به دست می‌آید. این تابع مدل رگرسیون میانه را با استفاده از روش معرفی‌شده توسط پورت‌نوی (۲۰۰۳) براورد می‌کند. در زیر براورد ضریب‌های رگرسیونی همراه با بازه‌ی ۹۵ درصد آن‌ها آمده است.

جدول ۱. مقدار اولیه‌ی پارامترهای مدل و بازه‌های اطمینان متناظر

$\hat{\beta}_i = b_i$	حد بالای بازه‌ی اطمینان	حد پایین بازه‌ی اطمینان
$-۰٫۲۴۲۷۹$	$۱٫۳۸۰۲۱$	$-۱٫۸۶۵۸۰$
$۰٫۰۳۸۹۴۰$	$۰٫۰۶۳۲۱$	$۰٫۰۱۴۶۷۰$
$-۰٫۰۰۲۰۳$	$۰٫۰۰۷۱۹$	$-۰٫۰۱۱۲۵$

هر یک از بازه‌های به دست آمده را به ۱۰ قسمت تقسیم کرده و نقطه‌های حاصل را با خط‌های عمودی و افقی به یکدیگر متصل می‌کنیم، به‌طوری که یک شبکه از نقطه‌ها به دست آید. به این ترتیب مجموعه‌ای از مقدارهای ممکن برای بردار پارامترهای مدل به‌صورت زیر حاصل می‌شود.

$$\Big\{ (b_{۰i}, b_{۱i}, b_{۲i}) \Big| b_{۰i} = -۱٫۸۶۵۸ + \frac{i}{۱۱}\Big(۱٫۳۸۰۲ + ۱٫۸۶۵۸\Big),$$
$$b_{۱i} = ۰٫۰۱۴۶۷ + \frac{i}{۱۱}\Big(۰٫۰۶۳۲۱ - ۰٫۰۱۴۶۷\Big),$$
$$b_{۲i} = -۰٫۰۱۱۲۵ + \frac{i}{۱۱}\Big(۰٫۰۰۷۱۹ + ۰٫۰۱۱۲۵\Big), \quad i = ۱, \ldots, ۱۰ \Big\}.$$

در میان این مقدارها آن بردار که کوچک‌ترین نرم تابع $S_n(\beta)$ را نتیجه دهد و در شرط‌های

$$\sum_{i=۱}^{n} p_i = ۱ \quad \text{و} \quad \sum_{i=۱}^{n} p_i W_{ni} = ۰$$

صدق کند را به‌عنوان براورد نقطه‌ای پارامترهای مدل رگرسیون میانه در نظر می‌گیریم. پس از اجرای این مرحله‌ها، مدل رگرسیون برازش یافته به‌صورت زیر به دست می‌آید:

$$m_i = \mathrm{med}(T_i | X_{۱i}, X_{۲i}) = ۰٫۹۳۰۲۶ + ۰٫۰۱۹۲۷ X_{۱i} + ۰٫۰۰۱۲۲ X_{۲i}$$

که در آن m_i میانه‌ی زمان شکست به شرط مقداراهای مشخصی از متغیرهای کمکی است.

در علم تومورشناسی شاخص کارنوفسکی معیاری است که توانایی بیمار را برای انجام اعمال حیاتی ضروری که برای زنده ماندن لازم هستند، اندازه‌گیری می‌کند. هرچه این معیار بزرگ‌تر و نزدیک به ۱۰۰ باشد، توانایی بیمار بیش‌تر است و با نزدیک شدن به ۱۰ از توان بیمار برای ادامه‌ی زندگی کاسته می‌شود. به عبارت دیگر زمان مرگ بیمار نزدیک‌تر می‌شود. مثبت بودن ضریب این متغیر ($b_۱$) در مدل براورد شده بیان‌گر این حقیقت است که میانه‌ی طول عمر بیمار با افزایش معیار کارنوفسکی اندازه‌گیری‌شده قبل از عمل پیوند بیش‌تر می‌شود. این مطلب همان چیزی است که از تعریف شاخص انتظار می‌رفت. زمان انتظار برای دریافت سلول‌های پیوندی از دیگر عوامل مؤثر بر رد یا پذیرش پیوند است. بر خلاف انتظار ضریب مثبت متغیر مدت زمان انتظار سپری‌شده برای دریافت سلول‌های پیوندی ($b_۲$) بیان می‌کند که با افزایش زمان انتظار، میانه‌ی طول عمر بیمار پس از عمل پیوند افزایش می‌یابد.

۴٫۲ ناحیه‌ی اطمینان توأم بردار پارامترها (β)

براساس مطالب بیان‌شده در بخش سوم، ناحیه‌ی اطمینان ۹۵ درصد بردار پارامترهای β مجموعه‌ای به‌صورت رابطه‌ی (۶) است که در آن $c_{۰/۰۵}$ چندک ۹۵ درصد بالای توزیع آماره‌ی نسبت درستنمایی تجربی $\ell(\beta)$ می‌باشد. ماتریس کوواریانس حدی این آماره با در نظر گرفتن رابطه‌های (۴) و (۵) به‌صورت زیر به دست می‌آید:

$$\hat{\Gamma}_۱^{-۱}\hat{\Gamma} = \begin{bmatrix} ۰/۹۸۹۹۴۶۹ & -۰/۸۱۴۹۹۸۰۹۵ & -۰/۳۷۰۶۴۵۱۲۴ \\ ۰/۰۰۰۰۶۴۲۵۱ & ۱/۰۰۵۱۸۴۵۳۲۰ & ۰/۰۰۲۳۴۴۰۷۴۰ \\ ۰/۰۰۰۰۱۷۴۹۸ & ۰/۰۰۱۴۲۲۷۶۸۰ & ۱/۰۰۰۶۴۲۰۶۳۰ \end{bmatrix}.$$

برای یافتن مقدار بحرانی ناحیه‌ی اطمینان به توزیع آماره‌ی $\ell(\beta)$ نیاز است. وقتی n بزرگ می‌شود، آماره‌ی نسبت درستنمایی تجربی با ترکیب خطی از توزیع‌های خی‌دو با یک درجه‌ی آزادی هم‌توزیع است (تچین و تسائو، ۲۰۰۳). بنا بر این برای شبیه‌سازی توزیع نمونه‌گیری آماره‌ی نسبت درستنمایی تجربی از رابطه‌ی زیر که با آن هم‌توزیع است استفاده کرده و توزیع حدی این آماره را تولید می‌کنیم.

$$(۰/۹۹۹۹۹۲۴)\chi^۲_{۱,۱} + (۰/۹۹۹۹۴۳۲)\chi^۲_{۲,۱} + (۰/۹۹۵۸۳۷۸)\chi^۲_{۳,۱}$$

که در آن ضریب‌ها ویژه مقدارهای ماتریس $\hat{\Gamma}_۱^{-۱}\hat{\Gamma}$ هستند. روش کار به این صورت است که در هر بار تکرار نمونه‌گیری سه عدد تصادفی از توزیع خی‌دو با یک درجه‌ی آزادی تولید کرده و رابطه‌ی بالا را محاسبه می‌کنیم. این روند را به تعداد $b = ۱۰۰۰$ بار تکرار کرده تا توزیع نمونه‌گیری آماره‌ی نسبت درستنمایی تجربی براورد شود. پس از مرتب کردن ۱۰۰۰ مقدار به دست آمده، صدک ۹۵ام را محاسبه کرده و آن را

به‌عنوان $c_{۰/۰۵} = ۷/۹۱۹۱۹۹۲$ در نظر می‌گیریم. با قرار دادن این مقدار در رابطه‌ی (۶) ناحیه‌ی اطمینان ۹۵ درصد برای بردار پارامترهای مدل به‌صورت زیر به دست می‌آید:

$$\{\beta : \ell(\beta) \leqslant ۷/۹۱۹۱۹۹۲\}$$

۴٫۳ بازه‌ی اطمینان برای هر یک از پارامترهای مدل (β_j)

با توجه با این‌که تفسیر ضریب‌های مدل رگرسیونی تنها در صورت معنی‌دار بودن آن‌ها دارای اهمیت است، بنا بر این در این قسمت بازه‌های اطمینان ۹۵ درصد را برای تک تک پارامترهای مدل محاسبه کرده و در مورد معنی‌دار بودن آن‌ها نتیجه‌گیری می‌کنیم. همان‌طور که در بخش سوم بیان شد، استنباط پیرامون هر زیر بردار دلخواه از پارامترهای مدل رگرسیون میانه با محاسبه‌ی آماره‌ی نسبت درستنمایی تجربی نیم‌رخ امکان‌پذیر است. با بزرگ شدن n، آماره‌ی نسبت درستنمایی تجربی نیم‌رخ با ترکیب موزون از توزیع‌های خی‌دو با یک درجه‌ی آزادی هم‌توزیع است. در عمل ماتریس کوواریانس حدی این آماره به ازای مقدارهای مختلفی از متغیرهای کمکی به‌سادگی براورد نمی‌شود. بنا بر این استفاده از این توزیع حدی برای یافتن مقدار بحرانی مورد نظر مناسب نیست. برای حل این مشکل توزیع نمونه‌گیری آماره‌ی نسبت درستنمایی تجربی نیم‌رخ با به‌کارگیری روش خودگردان[۱۷] شبیه‌سازی می‌شود (سوبرامانیان، ۲۰۰۷). بر خلاف روش معمول که بازنمونه‌گیری[۱۸] به روش تصادفی ساده‌ی باجایگذاری است؛ در این مقاله روش خودگردان با اندکی تغییر و با استفاده از روش نمونه‌گیری با احتمال‌های نابرابر باجایگذاری از نمونه‌ی اصلی اجرا می‌شود. نمادهای p_i همان احتمال‌های درستنمایی تجربی هستند که از رابطه‌ی زیر به دست می‌آیند:

$$p_i = \frac{۱}{۴۳}\left(\frac{۱}{۱+\lambda' W_{۴۳i}}\right), \qquad i = ۱, \ldots, ۴۳, \tag{۷}$$

جدول ۲. احتمال‌های درستنمایی تجربی

i	p_i	i	p_i	i	p_i	i	p_i
۱	۰/۰۳۴۹۰۸۱۳	۲	۰/۰۳۳۸۸۳۶۵	۳	۰/۰۱۲۶۸۹۴۲	۴	۰/۰۳۱۹۹۹۰۵
۵	۰/۰۱۶۴۸۹۳۳	۶	۰/۰۱۶۴۸۹۳۳	۷	۰/۰۲۹۵۳۰۸۰	۸	۰/۰۲۸۲۲۹۸۵
۹	۰/۰۲۸۱۳۷۶۹	۱۰	۰/۰۱۸۴۹۶۸۱	۱۱	۰/۰۲۷۹۸۴۱۰	۱۲	۰/۰۲۵۴۰۳۷۵
۱۳	۰/۰۲۵۶۳۸۷۳	۱۴	۰/۰۲۱۱۱۸۶۵	۱۵	۰/۰۲۴۶۸۶۴۶	۱۶	۰/۰۲۴۵۰۲۱۴
۱۷	۰/۰۲۴۲۴۵۶۲	۱۸	۰/۰۲۲۲۲۰۰۰	۱۹	۰/۰۲۲۳۸۷۷۴	۲۰	۰/۰۲۳۰۱۶۸۹
۲۱	۰/۰۲۲۹۵۵۴۵	۲۲	۰/۰۲۵۰۸۰۵۱	۲۳	۰/۰۲۳۸۵۶۲۰	۲۴	۰/۰۲۴۰۹۶۰۳
۲۵	۰/۰۲۴۶۴۸۲۹	۲۶	۰/۰۲۲۴۸۳۹۰	۲۷	۰/۰۲۲۴۵۳۱۸	۲۸	۰/۰۲۵۰۳۷۲۱
۲۹	۰/۰۲۲۳۳۰۳۰	۳۰	۰/۰۲۲۲۳۸۱۵	۳۱	۰/۰۲۲۲۳۸۱۵	۳۲	۰/۰۲۶۵۳۹۳۱
۳۳	۰/۰۲۱۷۴۶۶۵	۳۴	۰/۰۲۰۷۳۲۹۵	۳۵	۰/۰۳۲۲۴۲۰۰	۳۶	۰/۰۲۰۳۹۵۰۴
۳۷	۰/۰۲۰۷۸۳۶۱	۳۸	۰/۰۴۵۲۱۹۶۶	۳۹	۰/۰۳۶۹۴۰۸۶	۴۰	۰/۰۲۰۳۸۴۲۷
۴۱	۰/۰۱۹۶۴۷۰۳	۴۲	۰/۰۱۹۵۲۴۱۵	۴۳	۰/۰۱۹۲۴۷۶۹		

$$W_{۴۳i} = \begin{bmatrix} ۱ \\ X_{۱i} \\ X_{۲i} \end{bmatrix} \left\{ \frac{I(Y_i \geqslant ۰/۹۳۰۲۶ + ۰/۰۱۹۲۷X_{۱i} + ۰/۰۰۱۲۲X_{۲i})}{۱ - \hat{G}(۰/۹۳۰۲۶ + ۰/۰۱۹۲۷X_{۱i} + ۰/۰۰۱۲۲X_{۲i})} - \frac{۱}{۲} \right\}.$$

بردار ضریب‌های لاگرانژ λ از حل معادله‌ی زیر به‌روش عددی به دست می‌آید.

$$g(\lambda) = \sum_{i=۱}^{۴۳} \left(\frac{W_{۴۳i}}{۱ + \lambda' W_{۴۳i}} \right) = ۰.$$

با درنظر گرفتن این مطلب که p_iها مقدار احتمال هستند، رابطه‌های زیر را داریم:

$$۰ < \frac{۱}{۴۳} \left(\frac{۱}{۱ + \lambda' W_{۴۳i}} \right) < ۱ \Rightarrow ۱ + \lambda' W_{۴۳i} \geqslant \frac{۱}{۴۳}.$$

نکته‌ی قابل ذکر این است که در این‌جا الگوریتم نیوتن–رافسون با چک کردن این شرط در هر مرحله از تکرار الگوریتم، تعدیل می‌شود. بنا بر این مقداری از λ که به ازای آن p_i منفی می‌شود، به دست نمی‌آید. با قرار دادن λی به دست آمده از این روش در رابطه‌ی (۷)، p_iها به‌صورت جدول ۲ حاصل می‌شوند.

پس از محاسبه‌ی p_iها به ترتیبی که در بالا آمده است، نمونه‌گیری با احتمال‌های نابرابر با جایگذاری به اندازه‌ی n=۱۰۰۰ بار تکرار می‌شود. سپس از هر نمونه مقدار آماره‌ی $\ell_s(\beta_j)$ از رابطه‌ی زیر محاسبه می‌شود:

$$\ell_s(\beta_j) = \min_{(\beta_k,\beta_i)\in N} \ell\left(\beta_j, (\beta_k, \beta_i)\right)$$

$$N=\left\{\begin{bmatrix}\beta_k\\ \beta_i\end{bmatrix}:\left\|\begin{matrix}\beta_k-b_k\\ \beta_i-b_i\end{matrix}\right\|\leqslant\frac{۱}{\sqrt[۳]{۴۳}},\quad i,j,k=۰,۱,۲\quad i,k\neq j\right\}$$

که در آن b_i و b_k همان براوردهای نقطه‌ای به دست آمده در قبل هستند. به این ترتیب توزیع نمونه‌گیری آماره‌ی نسبت درستنمایی تجربی نیم‌رخ $\ell_s(\beta_j)$ به‌روش خودگردان درستنمایی تجربی شبیه‌سازی کرده و چندک‌های ۹۵ درصد و ۵ درصد از این توزیع را به دست می‌آوریم. بنا بر این بازه‌ی اطمینان برای پارامتر $\beta_j, j=۰,۱,۲$ مجموعه‌ی مقدارهایی از پارامتر مورد نظر است که آماره‌ی نسبت درستنمایی تجربی نیم‌رخ متناظر با آن‌ها درون بازه‌ی میان دو چندک به دست آمده از توزیع نمونه‌گیری $\ell_s(\beta_j)$ قرار بگیرند. با اجرای این روش بازه‌های اطمینان ۹۵ درصد برای هر یک از پارامترها به‌صورت زیر به دست می‌آیند:

جدول ۳. بازه‌های اطمینان درستنمایی تجربی برای پارامترها

براورد نقطه‌ای ضریب‌ها $\hat{\beta}_i$	حد بالای بازه‌ی اطمینان	حد پایین بازه‌ی اطمینان
۰٫۹۳۰۲۶	۱٫۶۴۰۲۶۳	۰٫۲۵۷۴۲۵
۰٫۰۱۹۲۷	۰٫۰۳۱۷۸۰	۰٫۰۰۵۲۷۸
۰٫۰۰۱۲۲	۰٫۰۱۰۰۷۶	−۰٫۰۰۸۴۰۶

همان‌طور که از ویژگی‌های روش درستنمایی تجربی انتظار می‌رفت، بازه‌های به دست آمده متقارن نیستند. این عدم تقارن نشان‌دهنده‌ی چولگی در توزیع تک تک پارامترها است. با توجه به بازه‌های به دست آمده برای دو پارامتر $\beta_۰$ و $\beta_۱$ مثبت بودن ضریب شاخص کارنوفسکی و عرض از مبدأ مشخص است. همچنین با مشاهده‌ی بازه‌ی به دست آمده برای پارامتر $\beta_۲$ می‌توان نتیجه گرفت که متغیر متناظر $(X_۲)$ اثر معنی‌داری روی میانه‌ی طول عمر بیمار ندارد. بنا بر این مدل جدید با یک متغیر کمکی $(X_۱)$ به‌صورت زیر نوشته می‌شود:

$$m_i=\mathrm{med}(T_i|X_{۱i})=۰٫۹۳۰۲۶+۰٫۰۱۹۲۷X_{۱i}\qquad(۸)$$

نمادها همان است که در قبل معرفی شده است.

بنا بر این، با قرار دادن مقدار به ازای شاخص کارنوفسکی در مدل به دست آمده به‌راحتی می‌توان میانه‌ی طول عمر بیمار پس از عمل پیوند را پیش‌گویی کرد. برای مثال، براورد نقطه‌ای لگاریتم مبنای ۱۰ میانه‌ی طول عمر برای بیماری با شاخص کارنوفسکی ۳۰، مقدار $۰٫۹۳۰۲۶+۰٫۰۱۹۲۷(۳۰)=۱٫۵۰۸۳۶$ است. در نتیجه میانه‌ی طول عمر بیمار پس از عمل پیوند برابر با ۳۲ ($۱٫۵۰۸۳۶^{۱۰}$) روز می‌باشد. به‌همین

ترتیب میانه‌ی طول عمر بیمار پس از عمل پیوند به ازای سایر مقدارهای شاخص کارنوفسکی، محاسبه شده و در جدول ۴ آمده است.

جدول ۴. براورد میانه‌ی طول عمر به ازای مقدارهای مختلف کارنوفسکی

کارنوفسکی (X_1)	میانه‌ی طول عمر (روز)
۱۰	۱۴
۲۰	۲۱
۳۰	۳۲
۴۰	۵۰
۵۰	۷۸
۶۰	۱۲۲
۷۰	۱۹۰
۸۰	۲۹۷
۹۰	۴۶۲
۱۰۰	۷۲۰

۴٫۴ بحث و نتیجه‌گیری

در این بخش مدل رگرسیون میانه برای مطالعه‌ی اثر دو عامل معیار کارنوفسکی و مدت زمان انتظار برای دریافت پیوند روی میانه‌ی طول عمر بیماران مبتلا به سرطان مغز استخوان که پس از یک دوره‌ی شیمی‌درمانی تحت عمل پیوند مغز استخوان قرار گرفته‌اند، براورد شد. پس از محاسبه‌ی براورد نقطه‌ای پارامترها مشخص شد میانه‌ی طول عمر بیماران با افزایش شاخص کارنوفسکی اندازه‌گیری‌شده قبل از عمل پیوند بیش‌تر می‌شود، که با یادآوری تعریف این معیار نتیجه‌ای کاملاً منطقی است. همچنین بازه‌ی اطمینان ۹۵ درصدی به دست آمده برای این پارامتر بیان می‌کند که متغیر متناظر (شاخص کارنوفسکی) روی میانه‌ی طول عمر مؤثر است و نباید از مدل خارج شود.

نتیجه به دست آمده برای اثرگذاری متغیر مدت زمان انتظار بیان می‌کند که با افزایش زمان انتظار سپری‌شده برای دریافت سلول‌های پیوندی، میانه‌ی طول عمر افزایش می‌یابد. البته با توجه به بازه‌ی به دست آمده برای این پارامتر دیده می‌شود که پارامتر مورد نظر تفاوت معنی‌داری با صفر ندارد. بنا بر این نتیجه‌ی به دست آمده از تفسیر این ضریب مورد استفاده قرار نمی‌گیرد. در نهایت با توجه به مطالب بیان شده در بالا مدل رگرسیون میانه‌ی مناسب برای این مطالعه به‌صورت رابطه‌ی (۸) است.

توضیحات

۱. Bootstrap Resampling Method

۲. Right Censored

۳. Empirical Likelihood

۴. Accelerated Bias Correction Bootstrap Method

۵. Range Preserving

۶. Improper Subset

۷. Heteroscedastic

۸. Median Regression Model

۹. Missing Completely at Random

۱۰. Grid Search Method

۱۱. Bootstrapping Profile Empirical Likelihood

۱۲. Resample

۱۳. Bone Marrow Transplantation

۱۴. Lymphoma

۱۵. Oncology

۱۶. Karnofsky

۱۷. Bootstrap

۱۸. Resampling

مرجع‌ها

Avalos, B.R., Klein, J.L., Kapoor, N., Tutschka, P.J., and Coprlan, E.A. (1993). Preparation for marrow transplantation in Hodgkin's and Non Hodgkin's lymphoma using Bu/CY, *Bone Marrow Transplantation*, **12**, 133-138.

DiCiccio, T.J., Hall, P. and Romano, J.P. (1989). Comparison of parametric and empirical likelihood functions, *Biometrika*, **76**, 465-476.

Hall, P. and Scala, B.L. (1990). Methodology and algorithm of empirical likelihood, *International Statistical Review*, **58**, 102-127.

Kaplan, E.L. and Meier, P. (1958). Nonparametric estimation from incomplete observation, *journal of the American Statistical Association*, **53**, 457-481.

Owen, A. (1988). Empirical likelihood ratio confidence intervals for single functional, *Biometrika*, **75**, 237-249.

Owen, A. (1990). Empirical likelihood ratio confidence regions, *The Annals of Statistics*, **18**, 90-120.

Owen, A. (1991). Empirical likelihood for linear models, *The Annals of Statistics*, **19**, 1725-1747.

Portnoy, S. (2003). Censored regression quantiles, *journal of the American Statistical Association*, **98**, 1001-1012.

Qin, G. and Lawless, J. (1994). Empirical likelihood and general estimating equations, *The Annals of Statistics*, **22**, 300-325.

Qin, G. and Tsao, M. (2003). Empirical likelihood inference for median regression models for censored survival data, *Journal of Multivariate Analysis*, **85**, 416-430.

Subramanian, S. (2007). Censored median regression and profile empirical likelihood, *Statistical Methodology*, **4**, 493-503.

Thomas, D.R. and Grunkemeier, G.R. (1957). Confidence interval estimation of survival probabilities for censored data, *journal of the American Statistical Association*, **70**, 865-871.

Wu, C. (2004). Weighted empirical likelihood inference, *Statistics and Probability Letters*, **66**, 67-79.

Ying, Z., Tung, S.H., and Wei, L.J. (1995). Survival analysis with median regression models, *journal of the American Statistical Association*, **90**, 178-184.

محدثه صفاکیش
گروه آمار، دانشکده‌ی اقتصاد،
دانشگاه علامه‌طباطبایی،
تهران، ایران.
رایانشانی: $m.safakish\_atu@yahoo.com$

حمیدرضا نواب‌پور
گروه آمار، دانشکده‌ی اقتصاد،
دانشگاه علامه‌طباطبایی،
تهران، ایران.
رایانشانی: $h.navvabpour@srtc.ac.ir$