باسمه تعالی

مجموعه مقالات

دومین کارگاه آموزشی

قابلیت اعتماد و کاربردهای آن

قطب علمی داده های ترتیبی و فضایی دانشگاه فردوسی مشهد

با همکاری

گروه های آمار دانشگاه اصفهان و دانشگاه فردوسی مشهد

و انجمن آمار ایران

مردادماه ۱۳۹۰

دبیر کارگاه:

آقای دکتر مجید اسدی عضو هیأت علمی گروه آمار دانشگاه اصفهان و عضو هسته قطب علمی داده های ترتیبی و فضایی

اعضای کمیته علمی:

- دکتر جعفر احمدی، دانشگاه فردوسی مشهد
- دکتر مجید اسدی، دانشگاه اصفهان
- دکتر فیروزه حقیقی، دانشگاه تهران
- دکتر بهاءالدین خالدی، دانشگاه رازی کرمانشاه
- دکتر احمد خدادادی، دانشگاه شهید بهشتی
- دکتر غلامرضا محتشمی برزادران، دانشگاه فردوسی مشهد

اعضای کمیته اجرایی (به ترتیب حروف الفبا):

- دکتر مجید اسدی، دانشگاه اصفهان
- دکتر محمد بهرامی، دانشگاه اصفهان
- دکتر هادی جباری نوقابی، دانشگاه فردوسی مشهد
- دکتر مهدی توانگر، دانشگاه اصفهان

محورها:

الف) عمومی (ویژه دانشجویان تحصیلات تکمیلی و کاربران قابلیت اعتماد)

روش های آماری در استنباط داده های قابلیت اعتماد، روش های ناپارامتری در استنباط داده های قابلیت اعتماد، آزمون های طول عمر تسریع یافته، الگوهای تعمیر و نگهداری سیستم ها

ب) تخصصی

قابلیت اعتماد سیستم های منسجم، داده های ترتیبی در قابلیت اعتماد، ترتیب های تصادفی در قابلیت اعتماد، استنباط آماری در قابلیت اعتماد، مدل های تنش- مقاومت، قابلیت اعتماد و نظریه اطلاع، مدل های تعمیر و نگهداری سیستم ها، آزمون های تسریع کننده، مفاهیم سالخوردگی، وابستگی در قابلیت اعتماد

پیشگفتار

در راستای ارتقای سطح علمی اعضای هیأت علمی و دانشجویان تحصیلات تکمیلی دانشگاه ها و مراکز آموزش عالی کشور و کاربران قابلیت اعتماد در مراکز صنعتی، دومین کارگاه آموزشی قابلیت اعتماد و کاربردهای آن روزهای ۱۸ و ۱۹ خردادماه ۱۳۹۰ توسط قطب علمی داده های ترتیبی و فضایی دانشگاه فردوسی مشهد و با همکاری گروه های آمار دانشگاه اصفهان و دانشگاه فردوسی مشهد و انجمن آمار ایران در محل دانشکده علوم ریاضی دانشگاه اصفهان برگزار گردید. در این کارگاه دو روزه بیش از ۸۵ نفر از اعضای هیأت علمی، دانشجویان تحصیلات تکمیلی و کاربران قابلیت اعتماد از دانشگاه ها و مراکز علمی و صنعتی شرکت نمودند. پس از فراخوان کارگاه، کمیته علمی اقدام به بررسی مقالات دریافتی نموده و تعداد بیست مقاله را به صورت سخنرانی و هشت مقاله را به صورت پوستر برای نمایش مورد موافقت قرار داد. همچنین تعداد شش سخنرانی عمومی که شامل بحث های کاربردی و کلاسیك قابلیت اعتماد بود در دو روز برگزاری کارگاه ارایه گردید. سخنرانان عمومی این کارگاه به شرح زیر بودند:

- آقای دکتر جواد بهبودیان، استاد گروه آمار دانشگاه شیراز
- آقای دکتر ناصررضا ارقامی، استاد گروه آمار دانشگاه فردوسی مشهد
- آقای دکتر مجید اسدی، استاد گروه آمار دانشگاه اصفهان
- آقای دکتر غلامرضا محتشمی برزادران، دانشیار گروه آمار دانشگاه فردوسی مشهد

در این کارگاه آقای دکتر مجید رضایی استادیار گروه آمار دانشگاه بیرجند در خصوص استفاده از نرم افزار R در مباحث قابلیت اعتماد سخنرانی نمودند. متن سخنرانی های عمومی این کارگاه در پایگاه اطلاعاتی قطب علمی داده های ترتیبی و فضایی دانشگاه فردوسی مشهد به نشانی http://osdce.um.ac.ir قابل دسترسی است. متن کامل تعداد سه مقاله ی فارسی و دوازده مقاله ی انگلیسی پذیرفته شده پس از تأیید کمیته علمی در این مجموعه تدوین شده است که بخش اول شامل مقالات انگلیسی و بخش دوم شامل مقالات فارسی است.

برگزاری این کارگاه علاوه بر تلاش خستگی ناپذیر کمیته های علمی و اجرایی، مرهون پشتیبانی گروه آمار دانشگاه اصفهان و مسئولین محترم آن دانشگاه بوده است. بدین وسیله مراتب سپاس و قدرانی قطب علمی داده های ترتیبی و فضایی گروه آمار دانشگاه فردوسی مشهد را ابراز می نماییم.

با آرزوی توفیق الهی

قطب علمی داده های ترتیبی و فضایی

فهرست

برآورد ضریب همبستگی کندال برای داده‌های بقای وابسته از طریق مدل‌های مفصل ارشمیدسی دو متغیره تحت سانسور تصادفی

جباری نوقابی، ه.

گروه آمار دانشگاه فردوسی مشهد

email: Jabbarinh@um.ac.ir

مختاری، ا. و آذرنوش، ح.ع.

گروه آمار دانشگاه آزاد اسلامی واحد مشهد

چکیده

بردار زمان بقای (T_1, T_2) را که به وسیله مدل مفصل ارشمیدسی مدل‌بندی می‌شود، در نظر می‌گیریم. در این مقاله با برازش مفصل کلایتون به داده‌ها و تعمیم ضریب همبستگی کندال برای داده‌های سانسور شده از راست، یک برآورد گشتاوری ساده برای پارامتر وابستگی مدل‌های مفصل ارشمیدسی به دست می‌آوریم.

واژه های کلیدی: مفصل‌های ارشمیدسی، مفصل کلایتون، ضریب همبستگی τ کندال، داده‌های بقای وابسته

کد نویسی ریاضی: ۶۲N۰۲.

۱ مقدمه

تاکنون مدل‌های زیادی برای بررسی داده‌های بقای چند متغیره‌ی وابسته پیشنهاد شده‌اند. در میان آن‌ها، مدل‌های ارشمیدسی بسیار متداول هستند. در این زمینه باندین روچ و لیانگ (۱۹۹۶)، ونگ و ول (a و b ۲۰۰۰) و هوگارد (۲۰۰۰) کار کرده‌اند. براین اساس تابع بقای دو متغیره $S(t_1, t_2)$ با توابع حاشیه‌ای $S(t_2) = S(0, t_2)$ و $S(t_1) = S(t_1, 0)$ را می‌توانیم به وسیله مفصل ارشمیدسی با درنظر گرفتن وابستگی بین دو زمان بقا مدل‌بندی کنیم. در این صورت داریم:

$$S(t_1, t_2) = p[q(S_1(t_1)) + q(S_2(t_2))] \tag{1}$$

که در آن، تابع $q(.)$ یک تابع محدب و نزولی بر بازه $[0, 1]$ است و $q(1) = 0$. همچنین $q(.)$ معکوس تابع $p(.)$ است. جنست و ریوست (۱۹۹۳) و اکس (۱۹۸۹) در این زمینه تحقیق کرده‌اند. در این مقاله برای مدل‌بندی وابستگی بین داده‌های بقا از مفصل کلایتون استفاده می‌کنیم که در آن $p(S) = (1 + S)^{\frac{-1}{\alpha}}$. با توجه به مفصل کلایتون تابع

بقای دو متغیره به صورت زیر خواهد بود:

$$S(t_۱,t_۲)=[\frac{۱}{S_۱(t_۱)^{-\alpha}+S_۲(t_۲)^{-\alpha}-۱}]^{\frac{۱}{\alpha}}.$$

متغیرهای تصادفی و مستقل V و U به صورت زیر تعریف می‌کنیم:

$$U=\frac{q(S_۱(T_۱))}{q(S_۱(T_۱))+q(S_۲(T_۲))} \quad (۲)$$

$$V=S(T_۱,T_۲)=p[q(S_۱(T_۱))+q(S_۲(T_۲))] \quad (۳)$$

جنست و ریوست (۱۹۹۳) ثابت کردند که اگر $(T_۱,T_۲)$ از مفصل ارشمیدسی با توابع بقای حاشیه‌ای $S_۱(t_۱)$ و $S_۲(t_۲)$ پیروی کند، آن‌گاه U دارای توزیع یکنواخت بر فاصله‌ی $[۰,۱]$ و V دارای توزیع کندال با تابع چگالی $k(v)=\frac{q(v)q(v)}{q(v)^۲}$ است که بر فاصله‌ی $[۰,۱]$ تعریف می‌شود و در آن $q(.)$ و$q(.)$ به ترتیب مشتقات مرتبه اول و دوم تابع $q(.)$ است.

فرض می‌کنیم بردار $(C_۱,C_۲)$ نشان‌دهنده‌ی سانسور تصادفی راست با تابع توزیع دلخواه در ازای بردار $(T_۱,T_۲)$ باشد. در بخش دوم، تابع توزیع شرطی $V=S(T_۱,T_۲)$ هنگامی که $T_۱$ و $T_۲$ هر دو یا یکی از آن‌ها سانسور شده‌اند، بررسی می‌شود. در بخش سوم برآوردگر مناسبی برای ضریب همبستگی τی کندال در معرض داده‌های سانسور شده معرفی می‌شود. در نهایت در بخش چهارم نتایج حاصل از برآوردگر ضریب همبستگی τی کندال که با استفاده از شبیه‌سازی بدست آمده است، ارایه می‌گردد.

۲ مواد و روش‌ها

در این بخش دو قضیه و یک نتیجه در مورد خواص متغیرهای تصادفی زمان بقا در معرض مشاهدات سانسور شده از راست بیان می‌شود.

فرض کنید $(T_۱,T_۲)$ یک جفت متغیر تصادفی باشد طوری که تابع توزیع آن‌ها توسط یک مفصل ارشمیدسی تولید شده است. همچنین فرض کنید $(T_۱,T_۲)$ بوسیله بردار $(C_۱,C_۲)$ از راست سانسور شده باشد که $(C_۱,C_۲)$ دارای توزیع توأم پیوسته دلخواهی است. قضیه زیر توزیع شرطی $V=S(T_۱,T_۲)$ را با فرض زمان‌های بقای $T_۱$ و $T_۲$ ارایه می‌کند.

قضیه ۱. (ونگ و اکس، ۲۰۰۸)

۱. تابع توزیع شرطی $(V \mid T_۱>c_۱,T_۲>c_۲)$ برابر است با

$$F_۱(v,c_۱,c_۲)=\frac{۱}{S(c_۱,c_۲)}[v-\frac{q(v)-q(S(c_۱,c_۲))}{q(v)}]; \quad ۰\le v\le S(c_۱,c_۲)$$

۲. تابع توزیع شرطی $(V \mid T_1 > c_1, T_2 = t_2)$ برابر است با

$$F_2(v, c_1, t_2) = \frac{q(S(c_1, t_2))}{q(v)}; \quad 0 \leq v \leq S(c_1, t_2)$$

۳. تابع توزیع شرطی $(V \mid T_1 = t_1, T_2 > c_2)$ برابر است با

$$F_2(v, t_1, c_2) = \frac{q(S(t_1, c_2))}{q(v)}; \quad 0 \leq v \leq S(t_1, c_2)$$

نتیجه ۱. تحت شرایط قضیه ۱ نتایج زیر بدست می‌آید:

$$\begin{aligned} i) E(V \mid T_1 > c_1, T_2 > c_2) &= \frac{S(c_1, c_2)}{2} - q[S(c_1, c_2)] \int_0^1 \frac{du}{q[uS(c_1, c_2]} \\ &+ \int_0^1 \frac{q[uS(1, c_2]}{q[uS(c_1, c_2)]} du. \end{aligned}$$

$$ii) E(V \mid T_1 > c_1, T_2 = t_2) = S(c_1, t_2) - S(c_1, t_2) q[S(c_1, t_2)] \int_0^1 \frac{du}{q[uS(c_1, t_2)]}.$$

$$iii) E(V \mid T_1 = t_1, T_2 > c_2) = S(t_1, c_2) - S(t_1, c_2) q[S(t_1, c_2)] \int_0^1 \frac{du}{q[uS(t_1, c_2)]}.$$

برهان. تحت شرایط قضیه۱، برای هر $0 < v < S(c_1, c_2)$ در مورد قسمت (i) داریم:

$$\begin{aligned} E(V \mid T_1 > c_1, T_2 > c_2) &= \int_v [1 - F(v, c_1, c_2)] dv \\ &= \int_0^{S(c_1, c_2)} dv - \int_0^{S(c_1, c_2)} \frac{v}{S(c_1, c_2)} dv \\ &+ \frac{\int_0^{S(c_1, c_2)} \frac{\phi(v)}{\phi(v)}}{S(c_1, c_2)} dv \\ &- \frac{\int_0^{S(c_1, c_2)} \frac{\phi(S(c_1, c_2)}{\phi(v)}}{S(c_1, c_2)} dv \end{aligned} \tag{4}$$

باتغییر متغیر $u = \frac{v}{S(c_1, c_2)}$ داریم:

$$\begin{aligned} E(V \mid T_1 > c_1, T_2 > c_2) &= v \mid_0^{S(c_1, c_2)} - \frac{v^2}{2S(c_1, c_2)} \mid_0^{S(c_1, c_2)} + \int_0^1 \frac{\phi(uS(c_1, c_2))}{\phi(uS(c_1, c_2))} du \\ & -\phi(S(c_1, c_2)) \int_0^1 \frac{du}{\phi(uS(c_1, c_2))} \\ &= S(c_1, c_2) - \frac{S(c_1, c_2)}{2} + \int_0^1 \frac{\phi(uS(c_1, c_2))}{\phi(uS(c_1, c_2))} du \\ & -\phi(S(c_1, c_2)) \int_0^1 \frac{du}{\phi(uS(c_1, c_2))}. \end{aligned}$$

برای اثبات (ii) داریم:

$$\begin{aligned} E(V \mid T_1 > c_1, T_2 = t_2) &= \int_v [1 - F(v, c_1, t_2)] dv \\ &= \int_0^{S(c_1, c_2)} dv - [\int_0^1 \frac{du}{\phi(uS(c_1, t_2))}] \phi(S(c_1, t_2) S(c_1, t_2). \end{aligned}$$

برای اثبات (iii) داریم:

$$\begin{aligned} E(V \mid T_1 = t_1, T_2 > c_2) &= \int_v [1 - F(V, t_1, c_2)] dv \\ &= \int_0^{S(c_1,c_2)} dv - [\int_0^1 \frac{du}{\phi(uS(t_1,c_2))}]\phi(S(t_1,c_2)S(t_1,c_2). \end{aligned}$$

قضیه ۲. (ونگ و اکس، ۲۰۰۸) فرض کنید (T_1, T_2) یک بردار تصادفی باشد که توسط مفصل کلایتون[1] مدل‌بندی می‌شود. آن‌گاه بنا به قضیه ۱ و فرع ۱ روابط زیر را حاصل می‌شوند:

$$\begin{aligned} &i) \quad E(V \mid T_1 > c_1, T_2 > c_2) = \frac{S(c_1,c_2)}{2}(\frac{\alpha+1}{\alpha+2}) \\ &ii) \quad E(V \mid T_1 > c_1, T_2 = t_2) = S(c_1,t_2)(\frac{\alpha+1}{\alpha+2}) \\ &iii) \quad E(V \mid T_1 = t_1, T_2 > c_2) = S(t_1,c_2)(\frac{\alpha+1}{\alpha+2}) \end{aligned} \tag{5}$$

پس در مدل کلایتون داریم:

$$E(V \mid T_1 > x_1, T_2 > x_2) = \frac{E(V \mid T_1 > x_1, T_2 = x_2)}{2}$$

و

$$E(V \mid T_1 > x_1, T_2 > x_2) = \frac{E(V \mid T_1 = x_1, T_2 > x_2)}{2}$$

واضح است که وقتی $\alpha = 0$ و T_1 و T_2 مستقل باشند، روابط (۵) به صورت زیر نوشته می‌شوند:

$$\begin{aligned} &i) \quad E(V \mid T_1 > c_1, T_2 > c_2) = \frac{S(c_1,c_2)}{4} = \frac{S_1(c_1)S_2(c_2)}{4} \\ &ii) \quad E(V \mid T_1 > c_1, T_2 = t_2) = \frac{S_1(c_1)S_2(t_2)}{2} \\ &iii) \quad E(V \mid T_1 = t_1, T_2 > c_2) = \frac{S_1(t_1)S_2(c_2)}{2} \end{aligned}$$

بنابراین هنگامی که α افزایش می‌یابد، امید ریاضی‌های شرطی نیز افزایش می‌یابند. دلیل این خاصیت این است که تابع بقا $S(t_1, t_2)$ در مدل مفصل کلایتون یک تابع صعودی از α است. همچنین هنگامی که $\alpha \to \infty$

$$E(V \mid T_1 > c_1, T_2 > c_2) \to \frac{S(c_1,c_2)}{2}$$

و

$$E(V \mid T_1 > c_1, T_2 = t_2) \to S(c_1, t_2)$$

و

$$E(V \mid T_1 = t_1, T_2 > c_2) \to S(t_1, c_2).$$

[1] Clayton

۳ کاربرد

یک راه ساده برای برآورد پارامترهای مدل‌های مفصل ارشمیدسی، استفاده از ضریب همبستگی τی کندال است. این ضریب همبستگی به صورت زیر تعریف می‌شود:

$$\tau = E[sign\{T_{1i} - T_{1j}\}sign\{T_{2i} - T_{2j}\}], 0 < \tau < 1$$

که در آن (T_{1i}, T_{2i}) و (T_{1j}, T_{2j}) مشاهدات مستقل از (T_1, T_2) بوده و $sign$ تابع علامت بوده و به صورت زیر تعریف می‌شود:

$$sign(T_{1i} - T_{1j}) = \begin{cases} 1 & (T_{1i} - T_{1j}) > 0 \\ 0 & (T_{1i} - T_{1j}) = 0 \\ -1 & (T_{1i} - T_{1j}) < 0 \end{cases}$$

اگر δ_1 و δ_2 به ترتیب وضعیت سانسور T_1 و T_2 را نشان دهد، رابطه‌ی بین τ و $V = S(T_1, T_2)$ در حالت کلی با استفاده از فرمول زیر بیان می‌شود:

$$\begin{aligned} \tau &= g(\theta) = 4E[V] - 1 \\ &= 4E[V(1-\delta_1)(1-\delta_2)] + E[V\delta_1(1-\delta_2)] + E[V(1-\delta_1)\delta_2] \\ &+ E[V\delta_1\delta_2]\} - 1 \end{aligned}$$

که در آن $\delta_1 = I(T_1 \le C_1)$ و $\delta_2 = I(T_2 \le C_2)$.

با استفاده از روابط قضیه ی ۱ داریم:

$$\begin{aligned} g(\hat{\theta}) &= \frac{4}{n}\sum_i [\frac{S(C_{1i}, C_{2i})}{2} - q(S(C_{1i}, C_{2i})\int_0^1 \frac{du}{q(uS(C_{1i}, C_{2i}))} \\ &+ \int_0^1 \frac{q(uS(C_{1i}, C_{2i}))}{q(uS(C_{1i}, C_{2i}))} du](1-\delta_{1i})(1-\delta_{2i}) \\ &+ \frac{4}{n}\sum_i [S(C_{1i}, T_{2i}) - S(C_{1i}, T_{2i})q(S(C_{1i}, T_{2i})) \\ &\times \int_0^1 \frac{du}{q(uS(C_{1i}, T_{2i}))}]\delta_{1i}(1-\delta_{2i}) \\ &+ \frac{4}{n}\sum_i [S(T_{1i}, C_{2i}) - S(T_{1i}, C_{2i})q(S(T_{1i}, C_{2i})) \\ &\times \int_0^1 \frac{du}{q(uS(T_{1i}, C_{2i}))}](1-\delta_{1i})\delta_{2i} \\ &+ \frac{4}{n}\sum_i S(T_{1i}, T_{2i})\delta_{1i}\delta_{2i} - 1 \end{aligned} \quad (6)$$

بنا به نتایج جنست و ریوست (۱۹۹۳) در مدل مفصل کلایتون داریم: $\tau = \frac{\alpha}{\alpha+2}$. پس با تعریف $X_{1i} = min(T_{1i}, C_{1i})$ و $X_{2i} = min(T_{2i}, C_{2i})$ و جایگذاری آن‌ها در رابطه (۶) داریم:

$$\begin{aligned}(\frac{\alpha}{\alpha+2})+1 &= \frac{4}{n}\sum_i S(X_{1i}, X_{2i})(1-\delta_{1i})(1-\delta_{2i})[\\ &\quad (\frac{1}{2}) + \frac{\delta_{1i}}{(1-\delta_{1i})} + \frac{\delta_{2i}}{(1-\delta_{2i})} + \frac{\delta_{1i}\delta_{2i}}{(1-\delta_{1i})(1-\delta_{2i})}(\frac{\alpha+2}{\alpha+1})]\\ &= \frac{4}{n}(\frac{\alpha+1}{\alpha+2})\sum_i S(X_{1i}, X_{2i})(1-\delta_{1i})(1-\delta_{2i})[\\ &\quad \frac{(\alpha+1)+(\alpha+1)\delta_{1i}+(\alpha+1)\delta_{2i}+(1-\alpha)\delta_{1i}\delta_{2i}}{2(1-\delta_{1i})(1-\delta_{2i})(\alpha+1)}]\end{aligned}$$

در نتیجه با کمی محاسبات ساده داریم:

$$\sum_i S(X_{1i}, X_{2i})[1+\delta_{1i}+\delta_{2i}+\frac{1-\alpha}{1+\alpha}\delta_{1i}\delta_{2i}] - n = 0 \qquad (7)$$

حال اگر در رابطه‌ی (۷) به جای α از برآورد آن به صورت $\alpha_n = \frac{2\tau_n}{1-\tau_n}$ استفاده کنیم، می توانیم برآوردی برای ضریب همبستگی τ ی کندال به دست آوریم زمانی که داده‌ی سانسور شده داریم. با استفاده از برآورد تابع بقای دو متغیره نیز که توسط دابروسکا(۱۹۸۸) بیان شده است، داریم:

$$\tau_n = \frac{\sum_i (X_{1i}, X_{2i})(1+\delta_{1i}+\delta_{2i}+\delta_{1i}\delta_{2i}) - n}{\sum_i (X_{1i}, X_{2i})(-1-\delta_{1i}-\delta_{2i}+3\delta_{1i}\delta_{2i}) + n}$$

بدیهی است هنگامی که هیچ داده‌ی سانسور شده‌ای نداشته باشیم(یعنی به ازای هرi، $\delta_{1i} = \delta_{2i} = 1$)، خواهیم داشت:

$$\tau_n = \frac{4}{n}\sum_{i=1}^{n}(T_{1i}, T_{2i}) - 1 = \frac{4}{n}\sum_{i=1}^{n} V_i - 1$$

۴ بحث و نتیجه گیری

در این مقاله با استفاده از قضیه ۱ و نتیجه‌ی حاصل از آن و مدل مفصل کلایتون یک برآورد گشتاوری برای τی کندال زمانی که داده‌ی سانسور شده داریم، ارایه شد. همچنین با استفاده از قضیه ۱ می‌توان پارامتر مجهول مدل‌های مفصل ارشمیدسی را برآورد کرد.

در یک مطالعه شبیه‌سازی شده، ابتدا یک نمونه‌ی تصادفی ۱۰۰ تایی از مدل های مفصل کلایتون، گامبل، نرمال و فرانک با $\tau = 0/1, 0/2, 0/3, 0/4, 0/5, 0/6, 0/7, 0/8, 0/9$ ، بردارهای زمان را تولید می‌کنیم. بردارهای سانسور را نیز از توزیع یکنواخت در بازه $(0, 1)$ تولید می‌کنیم. پس از تولید داده‌ها، مشاهده می‌کنیم که ۲۰ درصد آن‌ها سانسور شده‌اند. برای ارزیابی دقت برآوردگر ضریب همبستگی τی کندال در حالت داده‌های سانسور شده، اریبی و انحراف معیار برآوردها را به صورت تجربی با تکرار مستقل فرآیند شبیه‌سازی به اندازه‌ی ۱۰۰۰ بار محاسبه می‌کنیم. جدول ۱ تا ۴ خلاصه‌ی نتایج شبیه‌سازی با نرم افزار R را نشان می‌دهد.

جدول ۱: برآورد اریبی و انحراف معیار برآوردهای ضریب همبستگی τی کندال برای مفصل کلایتون

α	۰/۲۲	۰/۵	۰/۸۵	۱/۲۳	۲	۳	۴/۶۶	۸	۱۸
τ	۰/۱	۰/۲	۰/۳	۰/۴	۰/۵	۰/۶	۰/۷	۰/۸	۰/۹
اریبی برآورد شده	-۱/۰۹	-۱/۱۹	-۱/۲۸	-۱/۳۹	-۱/۴۸	-۱/۵۸	-۱/۶۹	-۱/۷۸	-۱/۸۹
انحراف معیار برآورد شده	۰/۰۰۴۳	۰/۰۰۴۶	۰/۰۰۵۰۹	۰/۰۰۴۷	۰/۰۰۴۶	۰/۰۰۵۲	۰/۰۰۵۱	۰/۰۰۵۷	۰/۰۰۵۶

جدول ۲: برآورد اریبی و انحراف معیار برآوردهای ضریب همبستگی τی کندال برای مفصل نرمال

α	۰/۱۵	۰/۳۰	۰/۴۵	۰/۵۸	۰/۷۵	۰/۸۰	۰/۸۹	۰/۹۵	۰/۹۸
τ	۰/۱	۰/۲	۰/۳	۰/۴	۰/۵	۰/۶	۰/۷	۰/۸	۰/۹
اریبی برآورد شده	-۱/۰۹	-۱/۱۸	-۱/۲۸	-۱/۳۸	-۱/۴۸	-۱/۵۸	-۱/۶۹	-۱/۷۹	-۱/۸۹
انحراف معیار برآورد شده	۰/۰۰۴۳	۰/۰۰۴۸	۰/۰۰۴۷	۰/۰۰۵۳	۰/۰۰۴۹	۰/۰۰۵۲	۰/۰۰۵۸	۰/۰۰۵۴	۰/۰۰۶۰

جدول ۳: برآورد اریبی و انحراف معیار برآوردهای ضریب همبستگی τی کندال برای مفصل گامبل

α	۱/۱۱	۱/۲۵	۱/۴۲	۱/۶۶	۲/۰۰	۲/۵۰	۳/۳۰	۵/۰۰	۱۰/۰۰
τ	۰/۱	۰/۲	۰/۳	۰/۴	۰/۵	۰/۶	۰/۷	۰/۸	۰/۹
اریبی برآورد شده	-۱/۰۸	-۱/۱۹	-۱/۲۹	-۱/۳۸	-۱/۴۹	-۱/۵۸	-۱/۶۸	-۱/۸۹	-۱/۸۸
انحراف معیار برآورد شده	۰/۰۰۳۸	۰/۰۰۴۶	۰/۰۰۴۵	۰/۰۰۵۱	۰/۰۰۴۸	۰/۰۰۴۹	۰/۰۰۵۷	۰/۰۰۵۴	۰/۰۰۵۹

جدول ۴: برآورد اریبی و انحراف معیار برآوردهای ضریب همبستگی τی کندال برای مفصل فرانک

α	۱/۰	۲/۰	۳/۰	۵/۰	۶/۰	۸/۰	۱۲/۰	۲۰/۰	۴۰/۰
τ	۰/۱	۰/۲	۰/۳	۰/۴	۰/۵	۰/۶	۰/۷	۰/۸	۰/۹
اریبی برآورد شده	-۱/۰۸	-۱/۱۸	-۱/۲۹	-۱/۳۹	-۱/۴۹	-۱/۵۸	-۱/۶۹	-۱/۷۸	-۱/۸۸
انحراف معیار برآورد شده	۰/۰۰۴۴	۰/۰۰۴۵	۰/۰۰۴۷	۰/۰۰۵۳	۰/۰۰۵۱	۰/۰۰۵۱	۰/۰۰۵۲	۰/۰۰۵۴	۰/۰۰۵۸

مراجع

[۱] Bandeen-Roche، K. J.، Liang، K.Y. .(۱۹۹۶) Modelling failure-timeassociations in data with multiple levels of clustering. *Biometrika*.، ۸۳، .۳۹–۲۹

[۲] Dabrowska، D.M.، .(۱۹۸۸) Kaplan - Meier estimate on the plane. *Ann. Statist* ، ۱۶، .۱۴۸۹–۱۴۷۵

[۳] Genest، C.، Rivest، L.P.، .(۱۹۹۳) Statistical inference procedures for bivariate Archimedean copulas. *J.Amer. Statist. Assoc.*، ۸۸، .۱۰۴۳–۱۰۳۴

[۴] Hougaard، P.، .(۲۰۰۰) Analysis of Multivariate Survival Data. Springer-Verlag، New York، Inc.

[۵] Wang، A.، Oakes، D.، .(۲۰۰۸) Some properties of the Kendall distribution in bivariate Archimedean copula models under censoring. *Statistics. Probability.*، ۷۸، .۲۵۸۳–۲۵۷۸

[۶] Wang، J.، Zafra، P.، .(۲۰۰۹) Estimating Bivariate Survival Function by Volterra Estimator Using Dynamic Programming Techniques. *Journal of Data Science.*، ۷، .۳۸۰–۳۶۵

[۷] Wang، W.، Wells، M.T.، (۲۰۰۰a). Model selection and semiparametric inference for bivariate failure-time data. *J. Amer. Statist. Assoc.*، ۴۴۹، .۷۲–۶۲

[۸] Wang، W.، Wells، M.T.، (۲۰۰۰b). Estimation of kendalls tau under censoring. *Statist. Sinica.*، ۱۰، .۱۲۱۸–۱۱۹۹

آزمونهای طول عمر تسریع یافته بر اساس میانگین باقیمانده عمر

محسن بهلولی زفره

دانشگاه اصفهان، گروه آمار

email: mohsen.bohlooli@yahoo.com

چکیده

به دست آوردن نمونه یا داده در مبحث قابلیت اعتماد یا به عبارتی مطالعات طول عمر یک تفاوت عمده با سایر مباحث آماری دارد و آن اینکه در مطالعات طول عمر حصول یک داده مستلزم از کار افتادن یک قطعه یا محصول است که زمان طولانی را می‌طلبد. آزمون‌های طول عمر تسریع یافته به منظور به دست آوردن سریع‌تر طول عمر محصولات با قابلیت اعتماد بالا به کار گرفته می‌شود. در این مقاله با استفاده از این آزمون‌ها و استفاده از میانگین باقیمانده طول عمر متناسب به برآورد تابع قابلیت اعتماد می‌پردازیم.

واژه های کلیدی: آزمون‌های طول عمر تسریع یافته، میانگین باقیمانده طول عمر متناسب، تابع قابلیت اعتماد، فشار.

کد نویسی ریاضی: ۶۲N۰۵.

۱ مقدمه

امروزه ارائه محصولاتی با خصوصیاتی بهتر و قابلیت اعتماد بالاتر برای تولیدکنندگان، اصلی اساسی به شمارمی‌آید. مصرف کنندگان انتظار دارند که محصولات مورد نیاز، مدت زمان طولانی‌تری کار کنند. از اینرو ضمانت در مورد قابلیت اعتماد، یکی از ویژگی‌های مهم محصولات شده است و در نتیجه برآورد هرچه دقیق‌تر قابلیت اعتماد از طریق انجام آزمایش سیستم‌ها در مراحل مختلف تولید محصول یک رکن اساسی است. اما انجام آزمون در شرایط طبیعی مستلزم صرف زمان بسیار طولانی و معمولاً از بین بردن تعداد زیادی از واحدهاست. از اینرو انجام آزمون‌های قابلیت اعتماد در شرایط طبیعی پرهزینه و غیر عملی است و این باعث بوجود آمدن آزمون‌های طول عمر تسریع یافته شده است. به عبارت دیگر آزمون‌های طول عمر تسریع یافته برای بدست آوردن سریع‌تر اطلاعات توزیع طول عمر یک محصول به کار می‌رود.

در آزمون‌های تسریع یافته واحدهای آزمون در معرض شرایطی سخت‌تر و شدیدتر از حالت طبیعی قرار می‌گیرند و در

نتیجه طول عمر واحدها نسبت به حالت طبیعی کوتاهتر می‌شود. سپس بعد از انجام آزمایش طول عمرهای تسریع یافته بدست آمده برای بدست آوردن برآوردها در شرایط طبیعی برون‌یابی می‌شوند. معمول‌ترین و پرکاربردترین روش برای تسریع آزمایش، اعمال فشار به واحد آزمون است. فشارهایی که معمولاً برای تسریع آزمون به کار می‌رود دما، ولتاژ، لرزش، رطوبت و ... است. آزمون از طریق اعمال فشار به چند نوع تقسیم می‌شود که مهمترین آنها فشار- ثابت ، فشار- مرحله‌ای .

این نوع آزمون توسط پژوهشگرانی همچون کلیپینسکی و نلسون (۱۹۷۵)، میکر و نلسون (۱۹۷۵)، نلسون و کیلپینسکی (۱۹۷۶)، نلسون و میکر (۱۹۷۸)، میکر و هان (۱۹۸۵) و یانگ (۱۹۹۴) مطالعه شده است.

۲ میانگین باقیمانده عمر

فرض کنید T نشان دهنده‌ی طول عمر یک سیستم باشد اگر چنین سیستمی در زمان t هنوز در حال کار باشد آنگاه باقیمانده‌ی عمر شرطی آن عبارتست از

$$(T-t\,|\,T>t)\,,$$

که تابع قابلیت اعتماد آن برابر است با

$$R\,(x\,|\,t)=\frac{R(x+t)}{R(t)},\qquad x>\cdot,\quad t>\cdot.$$

میانگین باقیمانده‌ی عمر T که یکی از معیارهای مهم در مطالعات طول عمر است و آن را با $m(t)$ نشان می‌دهیم به شکل زیر تعریف می شود.

$$m(t)=\mathrm{E}\,(T-t\,|\,T>t)\,.$$

$m(t)$ عبارت است از متوسط زمان باقیمانده از طول عمر سیستم، با فرض این که تا زمان t زنده بوده است. به راحتی می‌توان نشان داد به شرط $R(t)>\cdot$ داریم

$$\begin{aligned} m(t) &= \int_{\cdot}^{\infty} R\,(x\,|\,t)\,dx \\ &= \frac{\int_{t}^{\infty} R(x)dx}{R(t)}. \end{aligned}$$

قضیه ۱. اگر T دارای تابع قابلیت اعتماد R و میانگین باقیمانده عمر m باشد آنگاه

$$R(t)=\frac{m(\cdot)}{m(t)}e^{-\int_{\cdot}^{t}\frac{1}{m(x)}dx} \tag{۱}$$

برهان. هال و ولنر (۱۹۸۱) را ببینید.

۳ میانگین باقیمانده عمر طول عمر متناسب

پس از آنکه باتاچارجی (۱۹۸۲) و همچنین هال و ولنر (۱۹۸۱) خواصی از میانگین باقیمانده عمر را مورد بررسی قرار دادند و مَگولوری و ژانگ (۱۹۹۴) مدل رگرسیونی برای برآورد میانگین باقیمانده عمر ارائه دادند. اُکس و داسو (۱۹۹۰) مفهوم میانگین باقیمانده طول عمر متناسب ($PMRL$) را با توجه به مدل ارائه شده توسط کاکس (۱۹۷۲) معرفی نمودند. گوپتا و کرمانی (۱۹۹۸) کاربردها و خواص $PMRL$ را بررسی کردند.

تعریف ۱. دو توزیع با تابع‌های قابلیت اعتماد R_0 و R را دارای میانگین باقیمانده عمر متناسب می‌نامیم و با $PMRL$ نشان می‌دهیم هر گاه برای $\theta > 0$ (θ مستقل از زمان) داشته باشیم

$$m(t) = \theta m_0(t), \; \forall t, \tag{۲}$$

که در آن $m(t)$ تابع میانگین باقیمانده عمر متناظر با تابع قابلیت اعتماد R و $m_0(t)$ تابع میانگین باقیمانده عمر متناظر با تابع قابلیت اعتماد R_0 است. $m_0(t)$ را تابع میانگین باقیمانده عمر مبنا می‌نامند.

در بحث ما $m(t)$ را تابع میانگین باقیمانده طول عمر در شرایط تسریع یافته و $m_0(t)$ را تابع میانگین باقیمانده عمر در شرایط طبیعی در نظر می‌گیریم.

۴ مدل و استنباط آماری

فرض می‌کنیم n واحد در آزمایش قرار گرفته‌اند و ممکن است برخی از واحدها سانسور شوند. داده‌ها را به صورت $\{(T_i, z_i); i = 1, ..., n\}$ نمایش می‌دهیم که در آن T_i زمان شکست واحد i-ام و z_i فشار وارده بر واحد i-ام را نشان می‌دهد. فرض می‌کنیم که نمونه بدست آمده مستقل و همتوزیع هستند. مدل زیر را در نظر می‌گیریم

$$m(t; \mathbf{z}, \boldsymbol{\beta}) = \exp(\boldsymbol{\beta}' z) m_0(t), \quad \forall t, \tag{۳}$$

که در آن $\mathbf{z}' = (z_1, z_2, ..., z_p)$ بردار متغیر کمکی است و $\boldsymbol{\beta}$ بردار ضرایب مربوط به متغیرهای کمکی است. با این فرض که تابع میانگین باقیمانده مبنا خود به پارامترهایی مانند $\boldsymbol{\theta}$ بستگی دارد یعنی $m_0(t) = m_0(t, \boldsymbol{\theta})$. بنابراین تابع چگالی احتمال بر اساس رابطه (۲) به صورت زیر بدست می‌آید

$$\begin{aligned} f(t; \mathbf{z}, \boldsymbol{\beta}, \boldsymbol{\theta}) = {} & \tfrac{m_0(0)}{m_0^2(t)} \exp\left(-\boldsymbol{\beta}'\mathbf{z}\right) \\ & \times \exp\left(-\tfrac{1}{\exp(\boldsymbol{\beta}'\mathbf{z})} \int_0^t \tfrac{du}{m_0(u)}\right) \left(\exp\left(\boldsymbol{\beta}'\mathbf{z}\right) m'_0(t) + 1\right). \end{aligned} \tag{۴}$$

حال اگر u نشان دهنده واحدهای سانسور نشده و c نشان دهنده واحدهای سانسور شده باشد تابع درستنمایی به صورت زیر خواهد شد

$$L(\beta, \theta) = \prod_u f(t_u; \mathbf{z_u}, \beta, \theta) \prod_c R(t_c; \mathbf{z_c}, \beta, \theta).$$

و تابع لگاریتم درستنمایی عبارتست از

$$\log L(\beta,\theta) = \sum_u \ln f(t_u; \mathbf{z_u}, \beta, \theta) + \sum_u \ln R(t_c; \mathbf{z_c}, \beta, \theta).$$

با جایگذاری تابع چگالی (۴) در تابع لگاریتم درستنمایی بدست می‌آوریم

$$\begin{aligned}&\log L(\boldsymbol{\beta},\boldsymbol{\theta}) = \\ &\sum_u \ln\left(\frac{m_\cdot(\cdot)}{m_\cdot^{۲}(t_u)}\exp\left(-\boldsymbol{\beta}'\mathbf{z}_u\right)\exp\left(-\frac{۱}{\exp(\boldsymbol{\beta}'\mathbf{z}_u)}\int_\cdot^{t_u}\frac{du}{m_\cdot(u)}\right)\left(\exp\left(\boldsymbol{\beta}'\mathbf{z}_u\right)m'_\cdot(t_u)+۱\right)\right) \\ &+\sum_c \ln\left(\frac{m_\cdot(\cdot)}{m_\cdot^{۲}(t_c)}\exp\left(-\boldsymbol{\beta}'\mathbf{z}_c\right)\exp\left(-\frac{۱}{\exp(\boldsymbol{\beta}'\mathbf{z}_c)}\int_\cdot^{t_c}\frac{du}{m_\cdot(u)}\right)\left(\exp\left(\boldsymbol{\beta}'\mathbf{z}_c\right)m'_\cdot(t_c)+۱\right)\right).\end{aligned}$$

و با کمی ساده کردن خواهیم داشت

$$\begin{aligned}&\log L(\boldsymbol{\beta},\boldsymbol{\theta}) = \\ &\sum_u \left[\ln\left(m_\cdot(\cdot)\right) - ۲\ln\left(m_\cdot(t_u)\right) - \exp\left(-\boldsymbol{\beta}'\mathbf{z}_u\right)\int_\cdot^{t_u}\frac{du}{m_\cdot(u)} + \ln\left(m'_\cdot(t_u)+\exp\left(-\boldsymbol{\beta}'\mathbf{z}_u\right)\right)\right] \\ &+\sum_c \left[\ln\left(m_\cdot(\cdot)\right) - ۲\ln\left(m_\cdot(t_c)\right) - \exp\left(-\boldsymbol{\beta}'\mathbf{z}_c\right)\int_\cdot^{t_c}\frac{du}{m_\cdot(u)} + \ln\left(m'_\cdot(t_c)+\exp\left(-\boldsymbol{\beta}'\mathbf{z}_c\right)\right)\right].\end{aligned}$$

برای بدست آوردن برآورد پارامترهای β و $\boldsymbol{\theta}$ قرار می‌دهیم

$$I_\cdot(t) = \int_\cdot^t \frac{du}{m_\cdot(u)},$$

$$D_\cdot(t,\boldsymbol{\theta}) = \frac{\partial m_\cdot(t,\boldsymbol{\theta})}{\partial t},$$

$$F_\cdot(t,\boldsymbol{\theta}) = \ln\left[\frac{m_\cdot(t,\boldsymbol{\theta})}{m_\cdot(\cdot,\boldsymbol{\theta})}\right],$$

و

$$\xi(t,\mathbf{z},\boldsymbol{\beta},\boldsymbol{\theta}) = \ln\left(\frac{D_\cdot(t,\boldsymbol{\theta})+\exp\left(-\boldsymbol{\beta}'\mathbf{z}\right)}{m_\cdot(t,\boldsymbol{\theta})}\right).$$

اگر زمان‌های سانسور واحدها $c_۱, c_۲, \ldots, c_n$ در نظر گرفته شده باشند و قرار دهیم $t_i = \min(T_i, c_i)$ و $\delta_i = I(T_i \le c_i)$ آنگاه

$$\log L(\boldsymbol{\beta},\boldsymbol{\theta}) = \sum_{i=۱}^n [\delta_i\xi(t_i,\mathbf{z},\beta,\theta) - \exp\left(-\beta'\mathbf{z}_u\right)I_\cdot(t_i) - F_\cdot(t_i)].$$

و برآورد ماکزیمم درستنمایی پارامترهای β و $\boldsymbol{\theta}$ از حل معادلات زیر بدست می‌آیند

$$\sum_{j=۱}^n\left[I_\cdot(t_j,\theta) - \frac{j}{e_j(\boldsymbol{\beta},\boldsymbol{\theta})}\right]\exp\left(-\boldsymbol{\beta}'\mathbf{z}\right)z_i = \cdot, \quad i = ۱, ..., p,$$

$$\sum_{i=۱}^n\left[{}_i\xi'(t_i,\mathbf{z},\boldsymbol{\beta},\boldsymbol{\theta}) - \exp\left(-\boldsymbol{\beta}'\mathbf{z}\right)I'_\cdot(t_i) - F'_\cdot(t_i)\right] = \cdot,$$

که در آن

$$e_i(\boldsymbol{\beta},\boldsymbol{\theta}) = D_0(t_i,\boldsymbol{\theta}) + \exp\left(-\boldsymbol{\beta}'\mathbf{z}\right),$$

و $\xi'(t_i,\mathbf{z},\boldsymbol{\theta},\boldsymbol{\theta})$ و $F'_0(t_i)$ نشان دهنده بردار گرادیان $\xi(t,\mathbf{z},\boldsymbol{\beta},\boldsymbol{\theta})$ و $F_0(t,\boldsymbol{\theta})$ نسبت به بردار پارامتر $\boldsymbol{\theta}$ است. این معادلات از روش‌های عددی حل می‌شوند و سپس بعد از بدست آوردن برآورد پارامترها می‌توان تابع قابلیت اعتماد در شرایط طبیعی را با استفاده از داده‌های تسریع شده برآورد نمود.

مثال ۱ . یک آزمایش را برای مطالعه اثر فشار ولتاژ بر روی یک نوع دیود نوری (LED) انجام شده است. دو مجموعه ۳۰ تایی از این LEDها در معرض فشار ثابت ولتاژ ۷ و ۸ ولت قرار گرفته‌اند. در فشار ۷ ولت آزمایش تا شکست ۲۷ ام ادامه می‌یابد و در فشار دوم تا شکست ۲۵ ام. داده‌های شکست حاصل از این آزمایش در جدول ۱ خلاصه شده‌اند. تابع میانگین باقیمانده عمر مبنا به صورت زیر در نظر گرفته شده است.

$$m_0(u) = \gamma_0 + \gamma_1 u + \gamma_2 u^2.$$

با جایگذاری در مدل $PMRL$ بدست می‌آوریم

$$m(t) = (\gamma_0 + \gamma_1 t + \gamma_2 t^2)\exp(\beta z).$$

لذا تابع قابلیت اعتماد به صورت زیر بدست می‌آید

$$\begin{aligned} R(t) &= \frac{m_0(0)}{m_0(t)} \exp\left(-\exp(-\beta z)\int_0^t \frac{du}{m_0(u)}\right) \\ &= \frac{\gamma_0}{\gamma_0+\gamma_1 t+\gamma_2 t^2} \exp\left(-\exp(-\beta z)\int_0^t \frac{du}{\gamma_0+\gamma_1 u+\gamma_2 u^2}\right). \end{aligned}$$

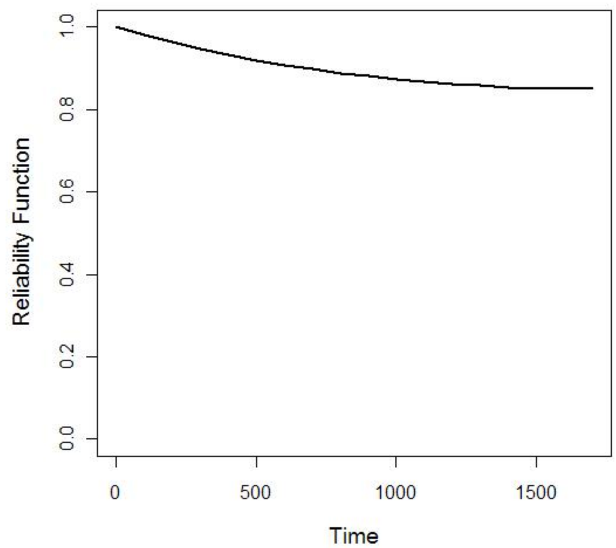
با استفاده از روش‌های عددی برآورد پارامترهای مدل $PMRL$ در جدول۲ آورده شده است. بر اساس این برآوردها تابع قابلیت اعتماد در سطح فشار طبیعی یعنی ۴ ولت در شکل ۱ رسم شده است.

جدول ۵: داده‌های به کار رفته در مثال ۱

فشار	زمانهای شکست									
۷ ولت	۲	۲	۳	۳	۵	۷	۸	۲۴	۲۶	۲۷
	۳۷	۹۷	۱۳۰	۱۵۵	۱۷۱	۲۱۵	۳۴۶	۴۰۸	۴۹۹	۵۴۴
	۶۰۰	۶۸۶	۱۰۱۰	۱۱۳۵	۲۳۴۵	۲۳۶۵	۳۵۱۰			
۸ ولت	۲	۳	۴	۴	۵	۵	۷	۸	۱۲	۱۵
	۲۴	۲۶	۴۵	۴۹	۵۲	۵۸	۶۰	۱۴۳	۱۷۰	۱۸۹
	۱۹۰	۳۰۷	۳۵۹	۵۶۳	۷۳۵					

جدول ۶: برآورد پارامترهای مدل PMRL در مثال ۱

پارامتر	برآورد
$\eta_۰$	۶۳۴۷/۷۲
$\eta_۱$	۱/۲۹۱
$\eta_۲$	-۰/۰۰۰۳۷
β	-۰/۰۰۰۹



شکل ۱: برآورد تابع قابلیت اعتماد در فشار طبیعی طبق مدل PMRL در مثال ۱

مراجع

[۱] Cox، D.R. .(۱۹۷۲) Regression Models and Life Tables (With Discussion). *J. Roy. Stat. Soc. B*، ۳۴، .۲۲۰–۱۸۷

[۲] Gupta، R.C. and Kirmani، S. .(۱۹۹۸) On the Proportional Mean Residual Life Model and Its Implications. *Statistics*، ۳۲، .۱۸۷–۱۷۵

[۳] Hall، W.J. and Wellner، J.A. .(۱۹۸۱) Mean Residual Life، Proceedings of the International Symposium on Statistics and Related Topics، Amsterdam: North-Holland .۲۸۳–۱۶۹

[۴] Kielpinski،T.J. and Nelson، W.B. .(۱۹۷۵) Optimum Censored Accelerated Life Tests for Normal and Lognormal Life Distributions. *IEEE Transactions on Reliability*، **R-۲۴**، .۳۲۰–۳۱۰

[۵] Maguluri، G. and Zhang، C. .(۱۹۹۴) Estimation in the Mean Residual Life Regression Model. *Journal of the Royal Statistical Society B*، ۳، .۴۸۹–۴۷۷

[۶] Meeker، W.Q. and Hahn، G.J. .(۱۹۸۵) How to Plan an Accelerated Life Test، Some Practical Guidelines. The ASQC References in Quality Control:Statistical Techniques، Vol. .۱۰

[۷] Meeker، W.Q. and Nelson، W.B. .(۱۹۷۵) Optimum Accelerated Life-Tests for Weibull and Extreme Value Distributions. *IEEE Transactions on Reliability*، **R-۲۴**، .۳۳۲–۳۲۱

[۸] Nelson، W.B. .(۱۹۹۰) Accelerated Testing: Statistical Models، Test Plans، and Data Analyses. John Wiley Sons، Inc.، New York.

[۹] Nelson، W.B. and Kielpinski، T.J. .(۱۹۷۶) Theory for Optimum Censored Accelerated Life Tests for Normal and Lognormal Life Distributions. *Technometrics*، ۱۸، .۱۱۴–۱۰۵

[۱۰] Nelson، W.B. and Meeker، W.Q. .(۱۹۷۸) Theory for Optimum Accelerated Censored Life Tests for Weibull and Extreme Value Distributions. *Technometrics*، ۲۰، .۱۷۷–۱۷۱

[۱۱] Oakes، D. and Dasu، T. .(۱۹۹۰) A Note on Residual Life. *Biometrika*، ۲۰، .۴۱۰–۴۰۹

[۱۲] Yang، G. .(۲۰۰۷) Life Cycle Reliability Engineering، John Wiley Sons، Inc.، Hoboken، New Jersey.

تحلیل بقا با نرم افزار R

یاسر مهرعلی

گروه آمار دانشگاه اصفهان

email: yasermehrali@gmail.com

مجید اسدی

گروه آمار دانشگاه اصفهان

email: m.asadi@sci.ui.ac.ir

چکیده

در این مقاله به تحلیل داده های بقا در نرم افزار R می پردازیم. ابتدا داده های سانسور شده از راست و سپس داده های سانسور فاصله ای را تحلیل می نماییم. در هر قسمت به تحلیل مثالهایی از داده های واقعی می پردازیم.

واژه های کلیدی: برآورد، تابع بقا، تابع بقا تجربی، تابع مخاطره، توزیع نمایی، توزیع وایبل، رگرسیون.

کد نویسی ریاضی: ۶۲N۰۱، ۶۲N۰۲، ۶۲N۰۳.

۱ تحلیل داده های سانسور شده از راست

تئوری این بخش مبتنی بر فصل ۱۰ کتاب دابسون[۴] می باشد.

۱.۱ توابع بقا و توابع مخاطره

توزیع های که در تحلیل بقا مورد استفاده قرار می گیرند:

```
> names(survreg.distributions)
```

۱.۱.۱ توزیع نمایی

```
> x <- seq(.1,6, length=100)
> tetha=2; mean=1/tetha
> fexp=dsurvreg(x,mean, scale=1,dist="exponential")
```

```
> Fexp=psurvreg(x,mean, scale=1,dist="exponential")
> Sexp=1-Fexp
> hexp=fexp/ Sexp
> Hexp=-log(Sexp)
> par(mfrow=c(2,2))
> plot(x,fexp,typ="l",ylab="f")
> plot(x,Fexp,typ="l",ylab="F")
> plot(x,hexp,typ="l",ylab="h")
> plot(x,Hexp,typ="l",ylab="H")
> par(mfrow=c(1,1))

> par(mfrow=c(1,2))
> plot(x,Hexp,typ="l",xlab="y",ylab="H(y)")
> plot(x,Hexp,log="xy",typ="l",xlab="log(y)",ylab="log H(y)")
> par(mfrow=c(1,1))
```

۲.۱.۱ توزیع وایبل

```
> x <- seq(.1,10, length=100)
> lambda=5; tetha=2;scale=1/lambda; mean=gamma(1+scale)*tetha
> fwei=dsurvreg(x,mean, scale,dist="weibull")
> Fwei=psurvreg(x,mean, scale,dist="weibull")
> Swei=1-Fwei
> hwei=fwei/ Swei
> Hwei=-log(Swei)
> par(mfrow=c(2,2))
> plot(x,fwei,typ="l",ylab="f")
> plot(x,Fwei,typ="l",ylab="F")
> plot(x,hwei,typ="l",ylab="h")
>plot(x,Hwei,typ="l",ylab="H")
```

```
> par(mfrow=c(1,2))
> plot(x,Hwei,typ="l",ylab="H(y)")
> plot(x,Hwei,log="xy",typ="l",xlab="log(y)",ylab="log H(y)")
> par(mfrow=c(1,1))

> #-----10.3 Empirical survivor function----NULL-
```

۲.۱ تابع بقا تجربی

۱.۲.۱ مثال: زمانهای بهبود بیماری

محاسبه تابع توزیع تجربی، با استفاده از داده های زمان های بهبود بیماران سرطان خون (گهان، ۱۹۶۵) توضیح داده می شود. در دو گروه کنترل و تیمار هر کدام ۲۱ بیمار مورد مطالعه قرار گرفته است؛ و زمانهای بهبود بیماری اندازه گیری شده است. در گروه تیمار بیش از نیمی از مشاهدات سانسور شده است. (سانسور از راست). داده ها عبارتند از:

```
> Remission=matrix(scan(),ncol=3,byrow=T,dimnames =
 list(NULL,c('group','censor','time')))
1  0  10      0  1  11      0  1  1
1  0  11      0  1  12      0  1  1
1  1  13      0  1  12      0  1  2
1  1  16      0  1  15      0  1  2
1  0  17      0  1  17      0  1  3
1  0  19      0  1  22      0  1  4
1  0  20      0  1  23      0  1  4
1  1  22      1  1  6       0  1  5
1  1  23      1  1  6       0  1  5
1  0  25      1  1  6       0  1  8
1  0  32      1  0  6       0  1  8
1  0  32      1  1  7       0  1  8
```

```
1  0  34     1  0  9      0  1  8
1  0  35     1  1  10     0  1  11

> group=factor(Remission[,1],labels=c("Control","Treatment"))
; censor=Remission[,2]; time=Remission[,3]

> Surv(time, censor)

> sfit=survfit( Surv(time, censor)~group);sfit

> par(mfrow=c(1,2))
> plot(sfit,lty=1:2,ylab="survival"
,main="Figure 10.3")
> legend(x=27,y=1,levels(group),lty =1:2)
> plot(sfit,lty=1:2,fun="event",ylab="cumulative events")
> legend(27,.4,levels(group),lty =1:2)

> par(mfrow=c(1,2))
> plot(sfit,lty=1:2,fun="cumhaz",xlab="y",ylab="cumulative hazard")
> plot(sfit,lty=1:2,log='xy',fun="cumhaz",xlab="log(y)",ylab="log H(y)"
,main="Figure 10.4")
> legend(1,2.8,levels(group),lty =1:2)
> par(mfrow=c(1,1))
```

۳.۱ برآورد

۱.۳.۱ مثال: مدل نمایی ساده

ابتدا در حالت بدون سانسور

```
> # not censored
```

```
> fitexp<-fitdist(time,"exp",method="mme");fitexp
```

آزمون های نیکویی برازش

```
> summary(fitexp)
```

نمودار توزیع های تجربی و نظری

```
> plot(fitexp)
```

اکنون حالت سانسور شده را در نظر می گیریم

```
> # censored
> time
> rtime=ifelse(time*censor!=0,time*censor,NA); rtime
> d1<-data.frame(left=time,right=rtime)

> fitexpcens<-fitdistcens(d1,"exp")
> summary(fitexpcens)
```

نمودار توزیع تجربی و توزیع نظری

```
> plot(fitexpcens)
```

۲.۳.۱ مثال: مدل وایبل

ابتدا در حالت بدون سانسور

```
> # not censored
> fitwei<-fitdist(time,"weibull",method="mme");fitwei
> fitwei<-fitdist(time,"weibull");fitwei
```

آزمون های نیکویی برازش

```
> summary(fitwei)
```

نمودار توزیع های تجربی و نظری

```
> plot(fitwei)
```

اکنون حالت سانسور شده را در نظر می گیریم

```
> # censored
> fitweicens<-fitdistcens(d1,"weibull")
> summary(fitweicens)
```

نمودار توزیع تجربی و توزیع نظری

```
> plot(fitweicens)
```

۴.۱ رگرسیون برای یک مدل بقا پارامتری

۱.۴.۱ مثال: زمانهای بهبود

حالت اول: پاسخ ها نمایی

```
> expfit=survreg(Surv(time, censor)~group, dist="exponential")
```

آنالیز واریانس مدل

```
> summary(expfit)
```

ماتریس واریانس- کواریانس برآورد ضرایب

```
> expfit$var
```

مقادیر برازش شده و مانده های مدل

```
> predict(expfit)
> r=residuals(expfit, type="response");r
```

نمودارهای تشخیصی

```
> qqnorm(r)
> qqline(r)
```

```
> residuals(expfit, type="deviance")
> residuals(expfit, type="working")
```

حالت دوم: پاسخ ها وایبل

```
> weifit=survreg(Surv(time, censor)~group, dist="weibull")
```

آنالیز واریانس مدل

```
> summary(weifit)
```

ماتریس واریانس-کواریانس برآورد ضرایب

```
> weifit$var
```

۲ تحلیل داده های سانسور فاصله ای

۱.۲ تابع بقا تجربی

مثال: زمانهای بهبود بیماران سرطان سینه بیماران در دو گروه طبقه بندی شده اند. گروه اول فقط پرتو درمانی شده اند(Rad)، گروه دوم هم پرتو درمانی و هم شیمی درمانی شده اند(RadChem). زمان بهبود آنها بصورت بازه ثبت شده است. بنابراین به جای داده های معمولی، با داده های سانسور فاصله ای سروکار داریم.

```
> data(bcos)
> attach(bcos)
> bcos
```

داده های سانسور شده را بصورت زیر می سازیم

```
> Surv(left,right,type="interval2")
```

تابع بقا تجربی بصورت زیر بدست می آید

```
> library(interval)
> icout<-icfit(Surv(left,right,type="interval2")~treatment, data=bcos)
> summary(icout)
```

```
> par(mfrow=c(1,2))
> plot(icout)
> plot(icout,dtype="cdf")
> par(mfrow=c(1,1))
```

```
> par(mfrow=c(1,2))
> plot(icout[1])
> plot(icout[2])
> par(mfrow=c(1,1))
```

```
> sfit=survfit( Surv(left,right,type="interval2")~treatment,data=bcos)
> summary(sfit)
```

```
> par(mfrow=c(1,2))
> plot(sfit,lty=1:2,fun="cumhaz",xlab="y",ylab="cumulative hazard")
> plot(sfit,lty=1:2,log='xy',fun="cumhaz",xlab="log(y)",ylab="log H(y)")
> legend(5,2.8,title="treatment",levels(treatment),lty =1:2)
> par(mfrow=c(1,1))
```

۲.۲ برآورد

```
> d2<-bcos[,1:2]
> d2[d2==Inf]=NA ; d2
```

```
> fitexpcens<-fitdistcens(d2,"exp")
```

```
> summary(fitexpcens)
> plot(fitexpcens)

> fitweicens<-fitdistcens(d2, distr="weibull")
> summary(fitweicens)
> plot(fitweicens)
```

۳.۲ مدل سازی

```
> testresult<-ictest(Surv(left,right,type="interval2")~treatment,data=bcos)
> testresult

> testresult<-ictest(Surv(left,right,type="interval2")
~treatment,data=bcos,alternative="greater")
> testresult

> testresult<-ictest(Surv(left,right,type="interval2")
~treatment,data=bcos,alternative="less")
> testresult
```

پیوست

برخی کتب مرجع عمومی R در زمینه آمار:

سطح ۱:

۱ A Beginner's Guide to R- Alain F. Zuur - Elena N. Ieno -Erik H.W.G. Meesters-Springer

۲ R for Beginners-Emmanuel Paradis

۳ The R Guide Version ۲.۵ W. J. Owen

۴ simpleR، Using R for Introductory Statistics-John Verzani

سطح ۲:

۵ Using R for Introductory Statistics John Verzani-CHAPMAN and HALL/CRC

۶ Introductory Statistics with R Peter Dalgaard Springer

۷ Using R for Data Analysis and Graphics Introduction، Code and Commentary-J H Maindonald

۸ Statistics-An Introduction using R-Michael J. Crawley and Imperial College London، UK

۹ icebreakeR Andrew Robinson

سطح ۳:

۱۰ An Introduction to R: Software for StatisticalModelling and Computing Course-Materials and Exercises-Petra Kuhnert and Bill Venables

۱۱ Data Analysis and Graphics Using R-an Example-Based Approach -John Maindonald

۱۲ Modern Applied Statistics with S-Fourth edition-W. N. Venables and B. D. Ripley-Springer

۱۳ Statistical Analyses using s-Plus-Second Edition- Brian S. Eueritt-CHAPMAN and HALUCRC

۱۴ Introduction to the R Project for Statistical Computing for use at ITC D G Rossiter

برای آشنایی با نرم افزار R، منبع ۳ (مهرعلی و صالحی)را ملاحظه نمایید. منبع ۶ زیر درباره تحلیل بقا با نرم افزار R می باشد.

لینک زیر به معرفی پکیج ها و توابع نرم افزار R در زمینه تحلیل بقا می پردازد

http://cran.r-project.org/web/views/Survival.html

پکیج زیر شامل توابعی برای برآورد پارامترها در مدلهای قابلیت اعتماد نرم افزار می باشد

http://cran.r-project.org/web/views/Survival.html

مراجع

[۱] آشفته افشین، بزرگ نیا ابوالقاسم. تحلیل بقا. *اندیشه آماری سال هفتم، شماره دوم.*

[۲] مهرعلی یاسر (۱۳۸۸). آشنایی با نرم افزار R. *امید ۱۴ و ۱۵ (نشریه انجمن آمار دانشگاه اصفهان).*

[۳] مهرعلی یاسر، صالحی مهدی (۱۳۸۹). کارگاه آشنایی با نرم افزار R. *دهمین کنفرانس آمار ایران.*
لینک دریافت فایل

http://daneshamari.blogfa.com/post-۷۷.aspx

[۴] A. J. Dobson، .(۲۰۰۲) An Introduction to Generalized Linear Models.

[۵] M. Laure، D. Muller .(۲۰۰۹) ‘fitdistrplus’ R Package- Help to fit of a parametric distribution to non-censored or censored data.

[۶] M. Tableman، J. S. Kim .(۲۰۰۳) Survival Analysis Using S Analysis of Time-to-Event Data.

[۷] T. Therneau .(۲۰۰۹) ‘Survival’ R Package- Survival analysis، including penalised likelihood.