



JAMS

مجله دانشگاه علوم پزشکی اراک

دوره بیست و یکم، شماره شش، آذر و دی ۱۳۹۷

journal homepage: <http://jams.arakmu.ac.ir>



مجله دانشگاه علوم پزشکی اراک

مقاله پژوهشی

تجزیه و تحلیل داده‌های سانسور شده فاصله ای با استفاده از تکنیک جانهای تصادفی

دانیال حبیبی^۱، محمد رفیعی^{*۱}

۱. گروه آمار زیستی، دانشگاه علوم پزشکی اراک، اراک، ایران.

چکیده

زمینه و هدف: سانسور فاصله‌ای در مطالعات پزشکی در داده‌هایی که به صورت دوره‌ای تکرار می‌شوند رخ می‌دهند. روش‌های مرسوم به منظور بررسی این نوع داده وجود دارد. هدف از انجام این مطالعه، بررسی تکنیک جانهای تصادفی در تجزیه و تحلیل داده‌های سانسور شده فاصله‌ای است.

مواد و روش‌ها: با استفاده از تکنیک شبیه‌سازی مونت کارلو به ارزیابی توان این روش می‌پردازیم و هم‌چنین در نهایت با استفاده از مجموعه داده واقعی عملکرد آن را ارزیابی می‌کنیم. این مجموعه داده مربوط به اطلاعات دندان پزشکی شهر ارومیه است که شامل ۲۰۷ کودک بوده است. تمام محاسبات با استفاده از نرم‌افزار R 3.2.3 انجام گرفته است.

یافته‌ها: نتایج حاصل از شبیه‌سازی نشان می‌دهد که توان آزمون محاسبه شده در تکنیک جانهای تصادفی خوب و قابل قبول است. مقدار p حاصل در داده‌های واقعی نشان می‌دهد که با استفاده از تکنیک جانهای تصادفی تفاوتی وجود ندارد.

نتیجه‌گیری: روش جانهای تصادفی در مقایسه با سایر روش‌های مرسوم به عنوان یک روش جایگزین می‌تواند مورد استفاده قرار بگیرد.

اطلاعات مقاله

تاریخ دریافت: ۹۷/۰۱/۲۷

تاریخ پذیرش: ۹۷/۰۴/۲۰

تاریخ انتشار: ۹۷/۱۰/۰۱

واژگان کلیدی

جانهای تصادفی

سانسور فاصله‌ای

شبیه‌سازی مونت کارلو

*نویسنده مسئول:

محمد رفیعی

آدرس پستی: ایران، اراک، دانشگاه علوم پزشکی اراک، گروه آمار زیستی.

نمابر: +98 86 3417 3529

Email: rafeie@arakmu.ac.ir

۱. مقدمه

تحلیل بقاء از منظر علم آمار بدین صورت است که با استفاده از روش‌های مختلف به تجزیه و تحلیل متغیر تصادفی نامنفی پرداخته می‌شود (۱). به علت چولگی به راست و سانسور داده‌ها، روش‌های موجود آماری قادر به پاسخگویی برای آنالیز این‌گونه داده‌ها نیستند (۲، ۳). سانسور داده‌ها در واقع بدین مفهوم است که ما زمان دقیق رخداد را نمی‌دانیم، البته اطلاعات مربوط به قبل از زمان مورد بررسی را در اختیار داریم (۴-۶).

سانسور فاصله ای در مطالعات پزشکی در داده‌هایی که به صورت دوره ای اندازه‌گیری می‌شوند، مشاهده می‌گردند. اگر پیامد مورد نظر در زمان اولین ویزیت (زمان L) مشاهده نگردد اما در زمان ویزیت بعدی (زمان R) مشاهده گردد، زمان واقعی بروز بیماری در بازه زمانی بین L تا R خواهد بود. در این حالت گفته می‌شود که زمان واقعی بروز بیماری به طریق فاصله ای سانسور شده است (۷).

به عنوان مثال، زمان ابتلا به ویروس HIV را در نظر بگیرید. فردی ممکن است برای ارزیابی سلامتی خود دو بار در فاصله زمانی مختلف آزمایش HIV دهد؛ بدین منظور اگر در اولین آزمایش علائم ویروس در بدن فرد مشاهده نگردد و در آزمایش بعدی علائم ویروس در بدن مشاهده گردد، از آنجایی که زمان ورود ویروس به بدن مشخص نیست و تنها می‌دانیم بعد از زمان اولین آزمایش و قبل از دومین آزمایش ویروس به بدن منقل شده است، بنابراین زمان ورود ویروس به بدن را سانسور فاصله ای در نظر می‌گیریم (۸).

مثال دیگری را می‌توان در ارتباط با زمان اولین پوسیدگی دندان و یا زمان اولین ظهور دندان دائمی در نظر گرفت؛ معمولاً فردی که برای بررسی دندان‌های خود به‌طور مرتب به دندان‌پزشک مراجعه می‌کند، اگر در اولین ویزیت خود پوسیدگی دندان و یا ظهور اولین دندان دائمی مشاهده نشود و در ویزیت بعدی مشاهده شود در نتیجه می‌توان اطلاعات حاصل را به‌عنوان داده‌های سانسور شده فاصله‌ای در نظر گرفت (۹).

در دو دهه اخیر، تعداد مقالات زیادی در زمینه داده‌های سانسور شده فاصله‌ای و هم‌چنین نحوه تجزیه و تحلیل آن‌ها با عنوان آزمون لاک رتبه ای تعمیم یافته ارائه شده است (۱۰). پتو و پتو یک روش برای مقایسه گروه‌ها تحت جایگزینی نوع لمان ارائه داده‌اند. در این روش هم مشاهدات سانسور فاصله‌ای و هم مشاهدات دقیق در آنالیز می‌توانند حضور داشته باشند (۱۱). فینگلشتاین از روش رگرسیون تحت مدل خطرات متناسب استفاده نمود و هم‌چنین آزمون پارامتریک نمره را هنگامی که شاخص درمانی به عنوان متغیر است، به کار برد. در این روش می‌توان به مقایسه درمانی K نمونه ای پرداخت (۱۲). سان آزمون ناپارامتریک بدون فرض مدل خطر متناسب را توسعه داد. پس از آن ژائو و سان این روش را بهبود داده‌اند، به‌طوری که هر دو نوع مشاهده دقیق و سانسور شده می‌توانند وجود داشته باشند (۱۳). سان و همکاران، کلاس جدیدی از آزمون‌های لاک رتبه ای تعمیم یافته ارائه کرده‌اند که بر این اساس مشاهده دقیق نمی‌تواند حضور داشته باشد (۱۴). ژائو و همکاران به اصلاح آزمون لاک رتبه ای پرداخته‌اند تا امکان حضور مشاهدات دقیق فراهم گردد (۱۵).

در داده‌های سانسور شده اطلاعات مرتبط افراد مشخص است، اما زمان دقیق بقاء مشخص نیست و این در حالیست که در بیشتر موارد مربوط به داده‌های بریده شده، به نقص کامل اطلاعات در مورد پیشامد مدنظر اشاره می‌شود. به عبارت دیگر، به دلیل این‌که فرد مشاهده نمی‌شود، داده‌های بریده شده در داده وجود ندارد، ولی در سانسور شدن برای فرد پیشامد مدنظر رخ می‌دهد اما زمان دقیق آن مشخص نیست (۱۶).

می‌توان گفت داده‌های بریده شده از توزیعی می‌آیند که تکیه گاه آن محدود شده است. به عنوان نمونه، توزیع بریده شده از سمت چپ، یک توزیع نمایی است که تکیه گاه آن در بازه $(0, \infty)$ است. تابع توزیع بریده شده از سمت چپ در a به صورت زیر است:

$$F(x) = \frac{F(x) - F(a)}{1 - F(a)}$$

فاصله‌ای با استفاده از تکنیک شبیه‌سازی است که در نهایت با داده واقعی مورد ارزیابی قرار می‌گیرد.

۲. ملاحظات اخلاقی

این مطالعه با کد ۹۲-۱۵۹-۱۲ به تصویب کمیته اخلاق پژوهشی دانشگاه علوم پزشکی اراک رسیده است.

۳. مواد و روش‌ها

برای شبیه‌سازی داده‌های سنسور شده فاصله‌ای فرض کنید T نشان دهنده یک متغیر تصادفی نامنفی و پیوسته و همچنین زمان بقای یک فرد با تابع بقاء $S(t)$ باشد. T هنگامی یک سنسور فاصله‌ای است که مقدار دقیق آن مشخص نباشد و تنها درون یک فاصله (L, R) مشاهده گردد که $T \in (L, R)$ و $L \leq R$ با احتمال یک است.

برای شبیه‌سازی داده‌های سنسور شده فاصله‌ای نیاز به تولید سه تایی مرتب (t_i, l_i, r_i) داریم. t_i توسط شبیه‌سازی از تابع توزیع $S(t)$ به دست می‌آید. ولی روش تولید l_i و r_i با استفاده از شبیه‌سازی مشابه تولید t_i نیست. در طول عمر واقعی l_i و r_i برای هر فرد به‌طور پیوسته غیرممکن است، خصوصاً در مطالعاتی که به صورت پیگیری هستند. کیانی و آراسان (۱۷) یک مجموعه زمان بررسی را برای شبیه‌سازی در نظر گرفتند و فرض کرده‌اند که افراد در این زمان‌ها بررسی شده‌اند و این مجموعه زمان بررسی را با P نشان داده‌اند که $P = \{p_1, \dots, p_k\}$ و همچنین احتمال حضور هر فرد در هر یک از $j=1, 2, \dots, k$ را p_j با q نشان داده‌اند، به طوری که $0 \leq q \leq 1$ است. اگر $q=1$ باشد، انتظار می‌رود که افراد در همه زمان‌ها حضور داشته باشند. اگر $q=0$ باشد، انتظار می‌رود که افراد در همه زمان‌ها حضور نداشته باشند. اگر $0 < q < 1$ باشد، انتظار می‌رود که افراد در برخی از زمان‌ها احتمال حضور داشته باشند و برخی دیگر از زمان‌ها حضور نداشته باشند.

براساس مقدار q ، هر فرد دارای یک مجموعه زمان‌های بررسی واقعی است که این مجموعه را با A نشان داده‌اند. مثلاً برای

تابع توزیع بریده شده از سمت راست در b به صورت زیر است:

$$F(x) = \frac{F(x)}{F(b)}$$

تابع بریده شده فاصله‌ای از سمت چپ در a و از سمت راست در b به صورت زیر است:

$$F(x) = \frac{F(x) - F(a)}{F(b) - F(a)}$$

یک روش ساده برای تحلیل داده‌های سنسور شده فاصله‌ای این است که با استفاده از روش جانهای آن‌ها را به داده کامل تبدیل کنیم و برای آزمودن آزمون فرضیه از روش‌های موجود در داده‌های کامل استفاده شود. یک روش معمول انتخاب به‌عنوان روش جانهای نقطه میانی است.

همچنین اگر $T_i = L_i$ و $T_i = R_i$ را در نظر بگیریم، به ترتیب روش جانهای از سمت چپ و راست را خواهیم داشت. داده‌ها به‌صورت سنسور شده فاصله‌ای $(L_i, R_i], i=1, \dots, n$ تولید شده‌اند و هدف، مقایسه تابع بقای دو جامعه می‌باشد. بنابراین، فرضیه صفر به‌صورت زیر خواهد بود:

در این مطالعه روشی تحت عنوان روش جانهای تصادفی برای آنالیز داده‌های سنسور شده فاصله‌ای به کار می‌گیریم. در روش جانهای تصادفی، عددی تصادفی مانند w را از توزیع بریده شده از چپ در a و از راست در b تولید می‌کنیم. در مورد داده‌های سنسور شده فاصله‌ای، جانهای تصادفی به معنی تولید تصادفی عدد w در بازه $[L, R]$ است. عدد w از توزیع بریده شده از چپ در L و از راست در R به تصادف تولید می‌شود. به کمک این روش داده‌های سنسور شده فاصله‌ای w_1, w_2, \dots, w_n به مجموعه اعداد کامل $[L, R], i=1, 2, \dots, n$ تبدیل می‌شود. بدین صورت داده‌های سنسور شده فاصله‌ای، به داده‌های کامل تبدیل می‌شود و از روش‌های موجود در داده‌های کامل می‌توان برای تجزیه و تحلیل داده‌های سنسور شده فاصله‌ای استفاده نمود.

هدف از انجام این مطالعه، بررسی روش جانهای تصادفی به عنوان روشی در تجزیه و تحلیل داده‌های سنسور شده

سانسور شده فاصله ای بریده شده اند، تولید شده اند. در اینجا F_0 تابع توزیع تحت فرض صفر است.

در مرحله اول، اعداد $w_i, i=1,2,\dots,n$ به طور تصادفی از تابع توزیع F_0 که از سمت چپ در نقطه L_i و از سمت راست در نقطه R_i بریده شده اند، تولید می شوند. در مرحله دوم، آزمون رتبه لگاریتمی برای داده های کامل (W_i) مورد استفاده قرار می گیرد و سپس p مقدار حاصل از آزمون رتبه لگاریتمی محاسبه می شود.

در مرحله سوم، مراحل اول و دوم را B دفعه تکرار می شود و سپس میانگین p مقدار را به دست می آوریم.

در مرحله چهارم، براساس میانگین p مقدار که از مرحله سوم به دست آمده نسبت به رد یا پذیرش فرض صفر اقدام می شود. دو نمونه $n=100$ و $n=50$ در هر گروه برای مقایسه در نظر گرفته شده است. برای زمان های بقای سانسور فاصله ای $(L,R], i=1,\dots,n$ از توزیع نمایی با میانگین $\exp(\alpha\beta)$ استفاده شده اند (۹، ۱۳). پارامترهای توزیع نمایی $\alpha=2$ و $\beta = -0.8, -0.4, 0, 0.4, 0.8$ است.

ما از الگوریتم ارائه شده توسط کیانی و آراسان در سال ۲۰۰۲ برای تولید داده های سانسور شده فاصله ای استفاده کرده ایم (۱۷). در شبیه سازی مونت کارلو سطح معنی داری تمام آزمون ها در سطح 0.05 و با 5000 تکرار در نظر گرفته شده است. برای احتمالات حضور افراد (q) به ترتیب هنگامی که $n=100$ و $n=50$ است، ما q را برابر با $0.09, 0.075$ و 0.05 در نظر گرفته ایم. برای مثال، $q=0.05$ بدین معناست که یک فرد برای حضور در ویزیت بعدی 50 درصد شانس دارد، در حالی که $q=0.09$ نشان دهنده احتمال بیشتری برای حضور در ویزیت بعدی است. در نهایت نتایج را با یک مجموعه داده واقعی به بررسی می نماییم.

۴. یافته ها

جداول ۱ و ۲ نتایج حاصل از شبیه سازی را با استفاده از آزمون جانهای تصادفی به ترتیب با حجم نمونه 50 و 100

نام فرد داریم $A_i = \{a_{i1}, a_{i2}, \dots, a_{imi}\}$ ، توجه کنید که $A_i \subset P$.

فرضیات شبیه سازی عبارت اند از: K زمان بالقوه بررسی وجود داشته باشد که باید در طرح مشخص شده باشند. همه ی افراد در اولین زمان بررسی (p_1) مشاهده شوند. افراد در طول پیگیری با احتمال q حضور یابند. زمان های بقاء از یک تابع توزیع معلوم $S(t)$ تولید شوند. برای تولید اولین داده سانسور شده فاصله ای برای اولین فرد یا (I_1, r_1) مراحل زیر را انجام می دهیم. در مرحله اول u_2 را از توزیع $U(0,1)$ تولید می کنیم. در مرحله دوم تابع نشان گر زیر را تعریف می کنیم:

$$I = \begin{cases} 1, & \text{اگر فرد در زمان } p_2 \text{ حضور داشته باشد} \\ 0, & \text{اگر فرد در زمان } p_2 \text{ حضور نداشته باشد} \end{cases} \quad (u_2 \leq q)$$

در مرحله سوم، مراحل ۱ و ۲ را برای همه ی P_j زمان بررسی ممکن $j=3,4,\dots,k$ تکرار می کنیم. در مرحله چهارم برداری از همه ی حضورهای k عضو از P تهیه می کنیم. این بردار مجموعه ای از زمان های حضور واقعی یا همان A_i را نشان می دهد. برای مثال، بردار $(0.10, 0.10, 0.10, 0.10)$ برای فرد اول نشان می دهد:

$$A_1 - (a_{11}, a_{12}, a_{13}, a_{14}) - (p_1, p_4, p_6, p_9)$$

در مرحله پنجم، بزرگترین عضو A_1 که کوچکتر از t_1 است را به عنوان I_1 و کوچکترین عضو A_1 که بزرگتر از t_1 است را به عنوان r_1 در نظر می گیریم. روابط زیر به سادگی نتیجه می شود:

$$\begin{cases} t_1 < a_{11}, & \text{سانسور شده چپ} \Rightarrow (l_1, r_1) = (0, a_{11}) \\ t_1 < a_{1m_1}, & \text{سانسور شده راست} \Rightarrow (l_1, r_1) = (a_{1m_1}, +\infty) \\ t_1 = a_{1j}, & \text{داده کامل} \Rightarrow l_1 = r_1 = a_{1j} \end{cases}$$

در مرحله ششم، مراحل ۱ تا ۵ را برای به دست آوردن فواصل بقیه مشاهدات تکرار کنید. بدین صورت n داده سانسور شده فاصله ای به فرم (I_i, r_i) برای $i=1,2,\dots,n$ به دست می آیند.

در روش جانهای تصادفی ابتدا اعداد تصادفی $w_i, i=1,2,\dots,n$ برای هر داده سانسور شده فاصله ای (I_i, r_i) تولید می شود. این اعداد تصادفی، از توزیع F_0 که از سمت چپ و راست از بازه

با استفاده از روش مونت کارلو به شبیه‌سازی این تکنیک و در نهایت با استفاده از یک داده واقعی به بررسی عملکرد آن پرداخته شد.

کل فرآیند شبیه‌سازی توسط نرم‌افزار R انجام شده است. ما از الگوریتم ارائه شده توسط کیانی و آراسان در سال ۲۰۰۲ برای تولید داده‌های سانسور شده فاصله‌ای استفاده کرده‌ایم. در شبیه‌سازی مونت کارلو سطح معنی‌داری تمام آزمون در سطح ۰/۰۵ و با ۵۰۰۰ تکرار در نظر گرفته شده است. برای احتمالات حضور افراد (q) به ترتیب هنگامی که $n=50$ و $n=100$ است، q برابر با ۰/۵، ۰/۷۵، ۰/۹ در نظر گرفته شد.

پاکبازی و همکاران مطالعه ای با عنوان مقایسه منحنی بقا در داده‌های سانسور شده انجام داده‌اند (۹). در این مطالعه، بررسی و مقایسه روش‌های رایج در تجزیه و تحلیل اطلاعات در زمینه داده‌های سانسور شده فاصله‌ای با استفاده از روش شبیه‌سازی مونت کارلو صورت گرفته و در نهایت با داده واقعی مورد ارزیابی قرار گرفته است. روش‌های رایج استفاده شده شامل سه آزمون ناپارامتریک لاک رتبه‌ای تعمیم یافته، آزمون پارامتریک نمره و آزمون جانهای است. نتایج حاصل بیان‌گر آن است که آزمون پارامتریک نمره و روش جانهای عملکرد بهتری نسبت به سه روش ناپارامتریک دارند.

به عنوان نمونه در مطالعه پاکبازی و همکاران، در شبیه‌سازی آزمون نمره در حجم نمونه ۵۰ با احتمال حضور ۰/۵ و هم‌چنین با مقادیر بتای ۰/۸، ۰/۴، ۰، ۰/۴، ۰/۸ به ترتیب توان ۰/۹۷۳۸، ۰/۵۱۶۶، ۰/۰۶۲۸، ۰/۵۰۴۴ و ۰/۹۷۸۴ به دست آمده است. در روش جانهای نیز توان به ترتیب برابر با ۰/۹۷۰۰، ۰/۴۸۰۴، ۰/۰۵۰۶، ۰/۵۰۷۰ و ۰/۹۷۱۸ برآورد شده است. در این مطالعه، با استفاده از روش جانهای تصادفی به ترتیب توان ۰/۹۷۲۴، ۰/۵۱۵۶، ۰/۰۵۱۴، ۰/۵۰۶۲ و ۰/۹۷۹۰ حاصل شده است.

نتایج این مطالعه نشان می‌دهد که روش جانهای تصادفی در مقایسه با سایر روش‌های رایج که در مطالعه پاکبازی و همکاران بیان شده است، عملکرد خوب و قابل قبولی دارد و حتی در بعضی از موارد بهترین عملکرد را دارا می‌باشد.

نشان می‌دهد. برآورد توان بر اساس احتمال حضور ۰/۵، ۰/۷۵ و ۰/۹ به دست آمده است. به عنوان نمونه، اگر احتمال حضور فردی ۰/۵ باشد (بدین معنا که یک فرد برای حضور در ویزیت بعدی ۵۰ درصد شانس دارد)، در حجم نمونه ۵۰ با مقدار بتای ۰/۸، برآورد توان برابر با ۰/۹۷۲۴ است.

در نهایت آزمون را با استفاده از یک مجموعه داده واقعی مورد ارزیابی قرار دادیم (۱۸). این مجموعه داده مربوط به اطلاعات دندان پزشکی می‌باشد. این داده شامل ۲۰۷ کودک بوده است که اطلاعات آن‌ها در شهر ارومیه جمع آوری شده است. پیشامد مورد نظر اولین خرابی دندان دائمی است. هدف از انجام این مثال مقایسه بقای اولین دندان مولر در افراد دختر و پسر است. مقدار p حاصل نشان می‌دهد که با استفاده از تکنیک جانهای تصادفی تفاوتی در بقای اولین دندان مولر در افراد دختر و پسر وجود ندارد ($p=0/4736$).

جدول ۱. برآورد توان در $n=50$ و $\alpha=0/5$ با احتمال حضور ۰/۵، ۰/۷۵ و ۰/۹

β	احتمال حضور		
	۰/۵	۰/۷۵	۰/۹
۰/۸	۰/۹۷۲۴	۰/۹۷۳۰	۰/۹۷۵۶
۰/۴	۰/۵۱۵۶	۰/۴۹۵۶	۰/۵۲۹۱
۰	۰/۰۵۱۴	۰/۰۵۱۶	۰/۰۵۰۹
۰/۴	۰/۵۰۶۲	۰/۵۰۹۱	۰/۵۱۰۴
۰/۸	۰/۹۷۹۰	۰/۹۷۸۲	۰/۹۷۶۱

جدول ۲. برآورد توان در $n=100$ و $\alpha=0/5$ با احتمال حضور ۰/۵، ۰/۷۵ و ۰/۹

β	احتمال حضور		
	۰/۵	۰/۷۵	۰/۹
۰/۸	۱	۱	۱
۰/۴	۰/۸۰۰۱	۰/۸۱۰۷	۰/۸۰۹۸
۰	۰/۰۵۱۰	۰/۰۵۰۸	۰/۰۵۱۴
۰/۴	۰/۸۰۹۷	۰/۸۱۵۴	۰/۸۱۱۵
۰/۸	۱	۱	۱

۵. بحث

این مطالعه به بررسی روش جانهای تصادفی به عنوان تکنیکی برای تجزیه و تحلیل اطلاعات هنگامی که با داده‌های سانسور شده فاصله ای مواجه هستیم، پرداخته است.

۷. تقدیر و تشکر

این مطالعه برگرفته از پایان‌نامه دانشجویی است و هیچ‌گونه حامی مالی نداشته است. نویسندگان بر خود لازم می‌دانند از تمامی کسانی که در انجام این پژوهش همکاری و مساعدت نمودند تشکر و سپاس‌گزاری نمایند.

۸. سهم نویسندگان

تمامی نویسندگان معیارهای استاندارد نویسندگی بر اساس پیشنهادات کمیته بین‌المللی ناشران مجلات پزشکی را دارا بودند.

۹. تضاد منافع

بدین‌وسیله نویسندگان تصریح می‌نمایند که هیچ‌گونه تضاد منافی در خصوص پژوهش حاضر وجود ندارد.

در مطالعه حاضر با به‌کار بردن روش جانهای تصادفی مقدار p در مجموعه داده واقعی برابر با $۰/۴۷۳۶$ به‌دست آمده است که با مطالعه پاکسازی و همکاران مطابقت دارد و مقدار آن به سایر روش‌ها بسیار نزدیک است (در آزمون نمره برابر با $۰/۴۶۷۶$ و در روش جانهای برابر با $۰/۳۴۲۲$ است).

۶. نتیجه‌گیری

روش‌های مرسوم در تجزیه و تحلیل داده‌های مربوط به سانسور فاصله‌ای وجود دارد. در این مطالعه نیز تکنیک روش جانهای تصادفی به عنوان روشی برای تجزیه و تحلیل ارائه شده است که نتایج حاصل بیان‌گر آن است که می‌تواند به عنوان یک روش جایگزین مورد استفاده قرار گیرد.

References

1. Miller Jr RG. Survival analysis: John Wiley & Sons; 2011.
2. Collett D. Modelling Survival Data in Medical Research. Boca Raton, Fla: Chapman & Hall/CRC. 2003.
3. Habibi D, Rafiei M, Chehrei A, Shayan Z, Tafaqodi S. Comparison of Survival Models for Analyzing Prognostic Factors in Gastric Cancer Patients. Asian Pacific journal of cancer prevention: APJCP. 2018; 19(3):749.
4. Cox DR. Analysis of survival data: Routledge; 2018.
5. Kalbfleisch J, Prentice RL. statistical analysis of failure time data. 1980.
6. Klein JP, Van Houwelingen HC, Ibrahim JG, Scheike TH. Handbook of survival analysis: CRC Press; 2013.
7. Klein J, Moeschberger M. Survival analysis: statistical methods for censored and truncated data. New York, NY: Springer-Verlag; 2003.
8. Kleinbaum DG, Klein M. Survival analysis: Springer; 2010.
9. Pakyari RE, Habibi D. On Comparison of Survival Curves with Interval Censored Data. Jordan Journal of Mathematics and Statistics. 2016; 9(3):203-15.
10. Peace KE, Sun J, Chen D-GD. Interval-censored time-to-event data: methods and applications: Chapman and Hall/CRC; 2012.
11. Peto R, Peto J. Asymptotically efficient rank invariant test procedures. Journal of the Royal Statistical Society Series A (General). 1972: 185-207.
12. Finkelstein DM. A proportional hazards model for interval-censored failure time data. Biometrics. 1986: 845-54.
13. Zhao Q, Sun J. Generalized log-rank test for mixed interval-censored failure time data. Statistics in Medicine. 2004; 23(10):1621-9.
14. Sun J. A non-parametric test for interval-censored failure time data with application to AIDS studies. Statistics in medicine. 1996; 15(13):1387-95.
15. Zhao X, Zhao Q, Sun J, Kim JS. Generalized Log-Rank Tests for Partly Interval-Censored Failure Time Data. Biometrical Journal. 2008; 50(3):375-85.
16. Mills M. Introducing survival and event history analysis: Sage; 2010.
17. Kiani K, Arasan J, editors. Simulation of interval censored data in medical and biological studies. International Journal of Modern Physics: Conference Series; 2012: World Scientific.
18. Pakyari R, Rafiei M, Abolhasani S. Statistical Analysis of Failure Time of the Molar Tooth Using Goodness-of-Fit Technique for Interval Censored Data. Majallah-i dānishgāh-i 'ulūm-i pizishkī-i Arāk. 2016; 19(6):11-8.



JAMS

Journal of Arak University of Medical Sciences
2018; 21(6)

Journal Homepage: <http://jams.arakmu.ac.ir>



ORIGINAL RESEARCH

Analysis of Interval Censored Data Using Random Imputation Technique

Danial Habibi¹, Mohammad Rafiei^{1*}

1. Department of Biostatistics, Arak University of Medical Sciences, Arak, Iran.

ARTICLE INFORMATION

Article history:

Received: 16 April 2018

Accepted: 11 July 2018

Published online: 22 December 2018

Keywords:

Interval censored data

Monte Carlo study

Random imputation

* Corresponding Author:

Mohammad Rafiei; Department of Biostatistics, Arak University of Medical Sciences, Arak, Iran.

Fax: +98 86 3417 3529

Email: rafeie@arakmu.ac.ir

ABSTRACT

Background and Aim: Interval censored data occur in repeated data in medical studies. There are common methods to analysis this type of data. The purpose of this study is to examine the random imputation technique in the analysis of interval censored data.

Materials and Methods: Using the Monte Carlo simulation technique, we evaluate the power of Random Imputation method, and finally we assess its performance using the actual data set. Actual dataset is related to dental information in Urmia, which contains 207 children. All calculations are done using R 3.2.3 software.

Findings: The simulation results show that the power of random imputation technique is good and acceptable. The p-value in real data shows that there is no difference using the random imputation technique.

Conclusion: Random imputation technique can be used as an alternative method in comparison with other conventional methods.

© Copyright (2018) Arak University of Medical Sciences

Cite this article as: Habibi D., Rafiei M. Analysis of Interval Censored Data Using Random Imputation Technique. J Arak Uni Med Sci. 2018; 21(6): 26-33.