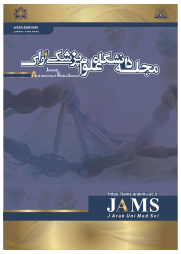
JAMS

مجله دانشگاه علوم پزشکی اراک

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

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



مقاله پژوهشی

تجزیه و تحلیل داده‌های سانسورشده فاصله ای با استفاده از تکنیک جانهی تصادفی

دانیال حبیبی[۱]، محمد رفیعی[۱*]

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

اطلاعات مقاله

تاریخ دریافت: ۹۷/۰۱/۲۷

تاریخ پذیرش: ۹۷/۰۴/۲۰

تاریخ انتشار: ۹۷/۱۰/۰۱

واژگان کلیدی

جانهی 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

چکیده

زمینه و هدف: سانسور فاصله‌ای در مطالعات پزشکی در داده‌هایی که به صورت دوره‌ای تکرار می‌شوند رخ می‌دهند. روش‌های مرسومی به منظور بررسی این نوع داده وجود دارد. هدف از انجام این مطالعه، بررسی تکنیک جانهی تصادفی در تجزیه و تحلیل داده‌های سانسورشده 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 جانهی تصادفی خوب و قابل قبول است. مقدار p حاصل در داده‌های واقعی نشان می‌دهد که با استفاده از تکنیک جانهی تصادفی تفاوتی وجود ندارد.

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

تمامی حقوق انتشار این اثر، متعلق به دانشگاه علوم پزشکی اراک می‌باشد.

www.SID.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 (۶-۴).

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 علایم ویروس در بدن فرد مشاهده نگردد و در آزمایش بعدی علایم 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 منتقل 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 دایمی در نظر گرفت؛ معمولا فردی که برای بررسی دندان‌های خود به‌طور مرتب به دندان‌پزشک مراجعه می‌کند، اگر در اولین ویزیت خود پوسیدگی دندان و یا ظهور اولین دندان دایمی مشاهده نشود و در ویزیت بعدی مشاهده شود در نتیجه می‌توان اطلاعات حاصل را به‌عنوان داده‌های سانسورشده فاصله‌ای در نظر گرفت (۹).

در دو دهه اخیر، تعداد مقالات زیادی در زمینه داده‌های سانسورشده 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 رگرسیونی تحت مدل خطرات متناسب استفاده نمود و هم‌چنین آزمون پارامتریک نمره را هنگامی‌که 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، به‌طوری‌که هر دو نوع مشاهده دقیق و سانسورشده 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 سانسورشده اطلاعات مرتبط افراد مشخص است، اما زمان دقیق بقاء مشخص نیست و این در حالیست که در بیشتر موارد مربوط به داده‌های بریده‌شده، 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 سانسورشدن 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 (∞+و۱) است. تابع توزیع بریده شده از سمت چپ در a به صورت زیر است:

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

Archive of SID

www.SID.ir

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

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

تابع بریده شده فاصله ای 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 سانسورشده فاصله ای این است که با استفاده از روش جانهی آن‌ها را به داده کامل تبدیل کنیم و برای آزمودن آزمون فرضیه از روش‌های موجود در داده‌های کامل استفاده شود. یک روش معمول انتخاب $T_i = \frac{L_i + R_i}{2}$ به‌عنوان روش جانهی نقطه میانی است. هم‌چنین اگر $T_i = L_i$ و $T_i = R_i$ را در نظر بگیریم، به ترتیب روش جانهی از سمت چپ و راست را خواهیم داشت. داده‌ها به‌صورت سانسورشده فاصله‌ای $(L_i, R_i], i = 1,\ldots,n$ تولید شده‌اند و هدف، مقایسه تابع بقای دو جامعه می‌باشد. بنابراین، فرضیه صفر به‌صورت زیر خواهد بود:

در این مطالعه روشی تحت عنوان روش جانهی تصادفی برای آنالیز داده‌های سانسورشده فاصله‌ای به کار می‌گیریم. در روش جانهی تصادفی، عددی تصادفی مانند w را از توزیع بریده شده از چپ در a و از راست در b تولید می‌کنیم. در مورد داده‌های سانسور شده فاصله ای، جانهی 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 داده های سانسورشده فاصله‌ای $[L, R], i=1,2,\ldots,n$ به مجموعه اعداد کامل $w_1, w_2, \ldots, w_n$ تبدیل می‌شود. بدین صورت داده‌های سانسورشده فاصله ای، به داده‌های کامل تبدیل می‌شود و از روش های موجود در داده‌های کامل می‌توان برای تجزیه و تحلیل داده‌های سانسورشده فاصله‌ای استفاده نمود.

هدف از انجام این مطالعه، بررسی روش جانهی تصادفی به عنوان روشی در تجزیه و تحلیل داده‌های سانسورشده فاصله‌ای با استفاده از تکنیک شبیه‌سازی است که در نهایت با داده واقعی مورد ارزیابی قرار می‌گیرد.

۲. ملاحظات اخلاقی

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

۳. مواد و روش‌ها

برای شبیه سازی داده سانسور 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 سانسور فاصله ای است که مقدار دقیق آن مشخص نباشد و تنها درون یک فاصله (L,R) مشاهده گردد که $T \in (L,R)$ و $L \le R$ با احتمال یک است.

برای شبیه سازی داده سانسورشده فاصله‌ای نیاز به تولید سه تایی مرتب (t_i,l_i,r_i) ،$i=1,2,\ldots,n$ داریم. t_i توسط شبیه‌سازی از تابع توزیع S(t) به‌دست می‌آید. ولی روش تولید l_i و r_i با استفاده از شبیه سازی مشابه تولید t_i نیست. در طول عمر واقعی l_i و r_i برای هر فرد به‌طور پیوسته غیرممکن است، خصوصا 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 زمان ها بررسی شده‌اند و این مجموعه زمان بررسی را با P نشان داده‌اند که $P=\{p_1,\ldots,p_k\}$ و هم‌چنین احتمال حضور هر فرد در هر یک از k، $j=1,2,\ldots,k$ ،pj را با q نشان داده‌اند، به طوری‌که $0 \le q \le 1$ است. اگر q=۱ باشد، انتظار می‌رود که افراد در همه زمان ها حضور داشته باشند. اگر q=۰ باشد، انتظار می‌رود که افراد در همه زمان ها حضور نداشته باشند. اگر $0 < q < 1$ باشد، انتظار می‌رود که افراد در برخی از زمان ها احتمال حضور داشته باشند و برخی دیگر از زمان ها حضور نداشته باشند.

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

Archive of SID
www.SID.ir

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

فرضیات شبیه سازی عبارت اند از: K زمان بالقوه بررسی وجود داشته باشد که باید در طرح مشخص شده باشند. همه‌ی افراد در اولین زمان بررسی (p_1) مشاهده شوند. افراد در طول پیگیری با احتمال q حضور یابند. زمان‌های بقاء از یک تابع توزیع معلوم S(t) تولید شوند. برای تولید اولین داده سانسورشده فاصله ای برای اولین فرد یا (l_1,r_1) مراحل زیر را انجام می‌دهیم. در مرحله اول u_2 را از توزیع U(۰,۱) تولید می‌کنیم. درمرحله دوم تابع نشان‌گر زیر را تعریف می‌کنیم:

$$I=\begin{cases}1, & \text{اگرفرددرزمان } p_2 \text{حضورداشته‌باشد} \quad (u_2\le q)\\ 0, & \text{اگرفرددرزمان } p_2 \text{حضورنداشته‌باشد} \quad (u_2>q)\end{cases}$$

در مرحله سوم، مراحل ۱ و ۲ را برای همه‌ی Pj زمان بررسی ممکن j=۳،۴،...k تکرار می‌کنیم. در مرحله چهارم برداری از همه‌ی حضورهای k عضو از P تهیه می‌کنیم. این بردار مجموعه ای از زمان‌های حضور واقعی یا همان A_i را نشان می‌دهد. برای مثال، بردار (۱و۰و۰و۱و۰و۰و۰و۱) برای فرد اول نشان می‌دهد:

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

در مرحله پنجم، بزرگترین عضو A_1 که کوچک‌تر از t_1 است را به‌عنوان l_1 و کوچک‌ترین عضو A_1 که بزرگ‌تر از t_1 است را به‌عنوان r_1 در نظر می‌گیریم. روابط زیر به‌سادگی نتیجه می‌شود:

$$\begin{cases}t_1<a_{11}, & \text{سانسورشده‌چپ} \Rightarrow (l_1,r_1)=(0,a_{11})\\ t_1<a_{1m_1}, & \text{سانسورشده‌راست} \Rightarrow (l_1,r_1)=(a_{1m_1},+\infty)\\ t_1=a_{1j}, & \text{داده‌کامل} \Rightarrow l_1=r_1=a_{1j}\end{cases}$$

در مرحله ششم، مراحل ۱ تا ۵ را برای به‌دست آوردن فواصل بقیه مشاهدات تکرار کنید. بدین صورت n داده سانسورشده فاصله‌ای به فرم (l_i,r_i) برای i=۱،۲،...n به دست می‌آیند.

در روش جانهی تصادفی ابتدا اعداد تصادفی w_i ،i=۱،۲،...n برای هر داده سانسورشده فاصله ای (l_i,r_i) تولید می‌شود. این اعداد تصادفی، از توزیع F_0 که از سمت چپ و راست از بازه سانسورشده فاصله ای بریده شده‌اند، تولید شده‌اند. در اینجا F_0 تابع توزیع تحت فرض صفر است.

در مرحله اول، اعداد w_i ،i=۱،۲،...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 شده‌اند، تولید می‌شوند. درمرحله دوم، آزمون رتبه لگاریتمی برای داده‌های کامل (Wi) مورد استفاده قرار می‌گیرد و سپس p مقدار حاصل از آزمون رتبه لگاریتمی محاسبه می‌شود.

در مرحله سوم، مراحل اول و دوم را B دفعه تکرار می‌شود و سپس میانگین p مقدار را به‌دست می‌آوریم.

در مرحله چهارم، براساس میانگین p مقدار که از مرحله سوم به‌دست آمده نسبت به رد یا پذیرش فرض صفر اقدام می‌شود. دو نمونه n=۱۰۰ و n=۵۰ در هر گروه برای مقایسه در نظر گرفته شده است. برای زمان های بقای سانسور فاصله ای (L,R],i=1,...,n از توزیع نمایی با میانگین $\exp(\alpha\square+\square\beta)$, $\exp(\alpha)$ که به ترتیب مرتبط با جامعه اول و دوم بوده‌اند استفاده شده‌اند (۹, ۱۳). پارامترهای توزیع نمایی α=۲ و β = -۰/۸، -۰/۴، ۰، ۰/۴، ۰/۸ است.

ما از الگوریتم ارائه شده توسط کیانی و آراسان در سال ۲۰۰۲ برای تولید داده های سانسورشده فاصله ای استفاده کرده ایم (۱۷). در شبیه‌سازی مونت کارلو سطح معنی‌داری تمام آزمون‌ها در سطح ۰/۰۵ و با ۵۰۰۰ تکرار در نظر گرفته شده است. برای احتمالات حضور افراد (q) به ترتیب هنگامی که n=۱۰۰ و n=۵۰ است، ما q را برابر با ۰/۹، ۰/۷۵ و ۰/۵ در نظر گرفته‌ایم. برای مثال، q=۰/۵ بدین معناست که یک فرد برای حضور در ویزیت بعدی ۵۰ درصد شانس دارد، در حالی‌که q=۰/۹ نشان دهنده احتمال بیشتری برای حضور در ویزیت بعدی است. در نهایت نتایج را با یک مجموعه داده واقعی به بررسی می‌نماییم.

۴. یافته‌ها

جداول ۱ و ۲ نتایج حاصل از شبیه سازی را با استفاده از آزمون جانهی تصادفی به ترتیب با حجم نمونه ۵۰ و ۱۰۰

Archive of SID
www.SID.ir

نشان می‌دهد. برآورد توان بر اساس احتمال حضور ۰/۵، ۰/۷۵ و ۰/۹ به‌دست 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 بتای۰/۸- 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 دندان‌پزشکی می‌باشد. این داده شامل ۲۰۷ کودک بوده است که اطلاعات آن‌ها در شهر ارومیه جمع آوری شده است. پیشامد موردنظر 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 جانهی تصادفی تفاوتی در بقای اولین دندان مولر در افراد دختر و پسر وجود ندارد (p=۰/۴۷۳۶).

جدول ۱. برآورد توان در n=۵۰ و α=۰۵/۰ 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=۱۰۰ و α=۰۵/۰ 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 جانهی تصادفی به عنوان تکنیکی برای تجزیه و تحلیل اطلاعات هنگامی که با داده‌های سانسورشده فاصله ای مواجهه هستیم، پرداخته است.

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 سانسورشده 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=۵۰ و n=۱۰۰ است،q برابر با ۰/۵، ۰/۷۵ و ۰/۹در نظر گرفته شد.

پاکیاری و همکاران مطالعه ای با عنوان مقایسه منحنی بقا در داده‌های سانسورشده انجام داده‌اند (۹). در این مطالعه، بررسی و مقایسه روش‌های رایج در تجزیه و تحلیل اطلاعات در زمینه داده‌های سانسورشده 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 جانهی است. نتایج حاصل بیان‌گر آن است که آزمون پارامتریک نمره و روش جانهی عملکرد بهتری نسبت به سه روش ناپارامتریک دارند.

به عنوان نمونه در مطالعه پاکیاری 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 و ۰/۸ به ترتیب توان ۰/۹۷۳۸، ۰/۵۱۶۶، ۰/۰۶۲۸، ۰/۵۰۴۴ و ۰/۹۷۸۴ به‌دست آمده است. در روش جانهی 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 جانهی 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 جانهی تصادفی در مقایسه با سایر روش‌های رایج که در مطالعه پاکیاری و همکاران بیان شده است، عملکرد خوب و قابل قبولی دارد و حتی در بعضی از موارد بهترین عملکرد را دارا می‌باشد.

Archive of SID
www.SID.ir

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

۶. نتیجه‌گیری

روش‌های مرسومی در تجزیه و تحلیل داده‌های مربوط به سانسور فاصله‌ای وجود دارد. در این مطالعه نیز تکنیک روش جانهی 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.

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 بین المللی ناشران مجلات پزشکی را دارا بودند.

۹. تضاد منافع

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

Archive of SID

www.SID.ir

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 ʿulūm-i pizishkī-i Arāk. 2016; 19(6):11-8.

Archive of SID

www.SID.ir

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 Habibi[1], 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.

This open-access article distributed under the terms of the Creative Commons Attribution NonCommercial 4.0 License (CC BY-NC 4.0).

www.SID.ir