مجله تخصصی اپیدمیولوژی ایران؛ ۱۳۹۵، دوره ۱۲، شماره ۳: صفحات ۶۳-۵۶.

مقاله پژوهشی

روش‌های بازنمونه‌گیری بوت استرپ و جک نایف در تحلیل بقای بیماران مبتلا به تالاسمی ماژور

رضا علی‌اکبری خویی[۱]، عنایت‌اله بخشی[۲]، آزیتا آذرکیوان[۳]، اکبر بیگلریان[۲]

[۱] فوق لیسانس آمار زیستی، گروه آمار زیستی، دانشگاه علوم بهزیستی و توانبخشی، تهران، ایران

[۲] دانشیار، گروه آمار زیستی، دانشگاه علوم بهزیستی و توانبخشی، تهران، ایران

[۳] دانشیار، فوق تخصص هماتولوژی انکولوژی کودکان، مؤسسه عالی آموزشی و پژوهشی طب انتقال خون، مرکز تحقیقات، درمانگاه تالاسمی، تهران، ایران

نویسنده رابط: اکبر بیگلریان، نشانی: تهران، اوین، بلوار دانشجو، بن‌بست کودکیار، دانشگاه علوم بهزیستی و توانبخشی، گروه آمار زیستی، تلفن: ۲۲۱۸۰۱۴۶

پست الکترونیک: abiglarian@uswr.ac.ir

تاریخ دریافت: ۹۴/۰۹/۰۹؛ پذیرش: ۹۵/۰۲/۰۴

مقدمه و اهداف: حجم نمونه کم می‌تواند تجزیه و تحلیل داده‌های آماری از جمله داده‌های بقا را تحت تأثیر قرار دهد. کاهش حجم نمونه می‌تواند به دلایل مختلف، از جمله از دست رفتن اطلاعات رخ دهد؛ یعنی داده‌های گم شده در برخی از متغیرها وجود داشته باشد. هدف این مطالعه، براورد پارامترها با استفاده از روش‌های بازنمونه‌گیری بوت استرپ و جک نایف در تحلیل بقای بیماران مبتلا به تالاسمی ماژور است.

روش کار: در این مطالعه هم‌گروه تاریخی از اطلاعات ۲۹۶ بیمار مبتلا به تالاسمی ماژور که در طی سال‌های ۱۳۷۳ تا بهار سال۱۳۹۲ به کلینیک ظفر در شهر تهران مراجعه کرده بودند؛ استفاده شد. برای تحلیل مدل‌های بقای مختلف پارامتری به داده‌ها برازش داده شد و پس از انتخاب بهترین مدل، از الگوریتم‌های بازنمونه‌گیری بوت استرپ و جک نایف استفاده شد. تحلیل داده‌ها با نرم‌افزار STATA نسخه ۱۲ انجام شد.

یافته‌ها: نتایج حاصل از روش‌های بازنمونه‌گیری بیانگر کاهش خطاهای استاندارد و کوتاه‌تر شدن فواصل اطمینان بودند. همچنین به‌کارگیری روش‌های باز نمونه‌گیری بوت استرپ و جک نایف منجر به معنی‌داری گروه‌های سنی و نسبت والدین ($P< ۰/۰۰۱$) در مقایسه با مدل لوگ-نرمال ($P>۰/۹۰۰$) گردید.

نتیجه‌گیری: با مقایسه طول فواصل اطمینان، روش بازنمونه‌گیری جک نایف عملکرد بهتری نسبت به روش بازنمونه‌گیری بوت استرپ داشت و بنابراین استفاده از این روش، در نمونه‌های با اندازه کم پیشنهاد می‌شود.

واژگان کلیدی: تحلیل بقا، روش‌های بازنمونه‌گیری، تالاسمی ماژور

مقدمه

حجم نمونه مناسب یکی از مهم‌ترین مسائل مربوط به تحلیل‌های آماری است. حجم نمونه کم می‌تواند باعث عدم دست‌یابی به نتایج متقن و زیر سؤال رفتن اعتبار نتایج شود. تعیین حجم نمونه مناسب برای مطالعه‌های مختلف متفاوت است که با فرمول‌های مناسب قابل دست‌یابی است (۱). پس از انجام مطالعه‌های میدانی، به طور معمول پژوهش‌گران با اطلاعات ازدست‌رفته مواجه می‌شوند، که همین امر باعث کاهش حجم نمونه می‌شود. برای مثال، ممکن است برخی از اطلاعات مورد نیاز در پرونده‌های پزشکی بیماران ثبت نشده باشد و همین باعث کاهش ناخواسته حجم نمونه شود. بدیهی است که برآوردهای نااریب با حجم نمونه مکفی در اختیار قرار می‌گیرد. نبود نمونه به اندازه کافی منجر به مخدوش شدن اعتبار و صحت برآوردهای به‌دست‌آمده برای مدل خواهد شد. برای مقابله با کاهش حجم نمونه که به واسطه ریزش مشاهدات، در برخی یا تمام متغیرها، رخ می‌دهد؛ راه‌های مختلفی وجود دارد. یکی از این راه‌ها که بر محدودیت گفته شده فایق می‌آید، به‌کارگیری روش‌های بازنمونه‌گیری[۱] است (۱،۲). بوت استرپ[۲] روشی است که فارغ از بسیاری از فرضیه‌ها، با تولید نمونه‌های فراوان، شرایط نمونه را به شرایط جامعه نزدیک می‌کند، و با در نظرگیری تمام حالت‌های تشکیل نمونه، می‌توان برآوردهای صحیح و نااریبی از ضرایب و نیز فاصله اطمینان گزارش نمود (۲). به‌عبارتی، با استفاده از این روش می‌توان برآوردهای نسبتاً دقیقی از پارامترهای جامعه و همین‌طور

[۱] Bootstrap
[۲] Jackknife

[Downloaded from irje.tums.ac.ir on 2024-09-08]

ضریب‌های رگرسیونی به دست آورد (۳). این روش در سال ۱۹۷۹ میلادی توسط افرون[۱] ارایه شد که به‌وسیله آن می‌توان مشکل نبود نمونه بزرگ را از بین برد (۴). جک نایف[۲] نیز یکی دیگر از روش‌های بازنمونه‌گیری می‌باشد، که با به‌کارگیری روشی خاص اعتبار برآوردهای پژوهش را بیشتر کرده و آثار سوء ناشی از کاهش حجم نمونه را به حداقل می‌رساند. این روش نیز توسط کوینویل[۳] در سال ۵۶-۱۹۴۹ میلادی ارایه شد و بسط آن توسط توکی[۴] در سال ۱۹۵۸ میلادی صورت گرفت. در این روش، مشاهدات یک‌به‌یک از مطالعه خارج شده و در ادامه برآوردهای مورد نظر به‌دست می‌آیند. درپایان، تمام براوردهای مراحل مختلف با هم تلفیق شده و برآوردهای موسوم به جک نایف به دست می‌آیند (۵،۶).

از فواید روش باز نمونه‌گیری بوت‌استرپ می‌توان به سادگی آن اشاره نمود. این روش برای برآورد خطای استاندارد و بازه‌ی اطمینان برای برآورد زننده‌های پیچیده پارامترهای توزیع، مثل نقطه‌های صدکی، نسبت‌ها، نسبت بخت‌ها و ضرایب همبستگی و هم‌چنین بررسی و کنترل پایداری نتایج مناسب است. از طرفی، در برخی مواقع، روش‌های بازنمونه‌گیری تضمینی برای نمونه متناهی نیست. اگر نتایج مهم باشند؛ باید تا جایی‌که توان محاسباتی رایانه‌ها و زمان اجرا اجازه می‌دهد؛ تعداد نمونه‌ها را زیاد کرد. البته باید توجه داشت که افزایش تعداد نمونه‌ها تغییری در افزایش اطلاعات، ناشی از داده‌های اولیه، نمی‌دهد؛ بلکه فقط اثر خطای نمونه‌گیری تصادفی را کاهش می‌دهد. لازم به ذکر است، روش جک نایف که برای برآورد مقدار اریبی و واریانس مورد استفاده قرار می‌گیرد، در مقایسه با روش بوت استرپ پیچیدگی کم‌تری دارد (۶،۷). از این روش‌ها برای ارزیابی ثبات مدل‌ها در برخی از مطالعه‌های بقای استفاده شده است. از جمله چن[۵] و همکاران که برای بررسی ثبات مدل کاکس از روش بازنمونه‌گیری بوت‌استرپ استفاده کردند. آن‌ها در مطالعه خود از نمونه‌های تولید شده تصادفی با جایگذاری و با حجم یکسان از نمونه اصلی استفاده نمودند، و برای نمونه جدید مدل رگرسیون کاکس را برازش دادند (۸). دن[۶] و همکاران در مطالعه‌ای دیگر با استفاده از روش بازنمونه‌گیری بوت‌استرپ به ارزیابی مدل رگرسیون کاکس پرداختند (۹). زاناتا[۷] و همکاران برای برآورد خطای معیار نرخ‌های بقای ۱۰ ساله، از روش جک‌نایف استفاده کردند. آن‌ها هم‌چنین برای مقایسه نرخ‌های بقا، خطاهای معیار حاصل را ملاک عمل قرار دادند (۱۰). مولینا[۸] و همکاران نیز از روش جک‌نایف برای برآورد خطای معیار نرخ‌های بقای یک‌ساله استفاده کردند (۱۱).

هدف این مطالعه، برآورد پارامترها با استفاده از روش‌های بازنمونه‌گیری بوت استرپ و جک نایف در تحلیل بقای بیماران مبتلا به تالاسمی ماژور به‌منظور شناسایی عوامل خطر این بیماران است.

روش کار

در این مطالعه که به‌صورت هم‌گروهی تاریخی انجام شد، اطلاعات ۲۹۶ نفر بیمار مبتلا به تالاسمی ماژور که در طی سال‌های ۱۳۷۳ تا بهار سال ۱۳۹۲ به کلینیک تالاسمی ظفر شهر تهران مراجعه کرده بودند؛ بررسی شد. داده‌ها و اطلاعات مورد نیاز از طریق پرونده خوانی در بخش بایگانی کلینیک ثبت گردید. متغیرهای زمان ابتلا و تشخیص تا زمان مرگ به‌عنوان متغیر وابسته (زمان بقا) در نظر گرفته شد. وضعیت تزریق خون (بدون تزریق، نامنظم، منظم)؛ سن، سن زمان ابتلا و تشخیص؛ سن در زمان نخستین تزریق دسفرال؛ سن در زمان اولین تزریق خون؛ وضعیت طحال برداری (برداشته شده، برداشته نشده)؛ نوع دسفرال تزریقی (زیرجلدی، پمپی، هیچ‌کدام، هر دو)؛ نسبت پدر و مادر (غریبه، فامیل دور، فامیل درجه‌یک)؛ میزان تحصیلات پدر و مادر (بی‌سواد، زیر دیپلم، دیپلم، دانشگاهی)؛ زادگاه (هم‌جوار دریا، غیر هم‌جوار دریا) و گروه سنی بیماران (متولدین پیش از سال۱۳۶۰ و متولدین پس از سال ۱۳۶۰) به‌عنوان متغیرهای مستقل در نظر گرفته شدند. لازم به ذکر است که این متغیرها با رایزنی با متخصصان بالینی ذی‌ربط و وضعیت ثبت آن‌ها در پرونده‌ها انتخاب شدند. تعداد مشاهدات نامشخص برای متغیرها در جدول شماره ۱ آمده است.

همان‌گونه که در جدول شماره ۱ آمده است، متغیرهای مطالعه دارای مقادیر نامشخص بسیاری هستند که این خود باعث کاهش حجم نمونه در هنگام تحلیل آماری داده‌ها خواهد شد.

---

[۱] Efron
[۲] Resampling
[۳] Quenouille
[۴] Tukey
[۵] Chen
[۶] Chen
[۷] Zanata
[۸] Molina

[Downloaded from irje.tums.ac.ir on 2024-09-08]

برای تحلیل از مدل‌های پارامتری بقا استفاده شد. پس از برازش مدل‌های مختلف به داده‌های بیماران تالاسمی ماژور و مقایسه آن‌ها با یکدیگر مدل پارامتری لوگ‌نرمال به‌عنوان بهترین مدل برای شناسایی عوامل خطر این بیماری انتخاب شد. برای این مدل ملاک اطلاع آکاییکه برابر ۱۸/۷۳ به دست آمد (۱۲). در ادامه، برای تعدیل و از بین بردن ناپایداری ناشی از کاهش حجم نمونه، روش‌های بازنمونه‌گیری بوت استرپ و جک نایف برای مدل لوگ-نرمال با استفاده از نرم‌افزار stata نسخه ۱۲ به کار گرفته شد.

۱) روش بازنمونه‌گیری بوت استرپ:

ایده‌ی اصلی در مورد روش بازنمونه‌گیری بوت استرپ از آن‌جا ناشی می‌شود که استنباط در مورد یک جامعه از روی یک نمونه می‌تواند به‌صورت استنباط در مورد نمونه از روی واحدهای بازنمونه‌گیری شده درآید. برای مثال، از حجم نمونه به‌اندازه n به دفعات مکرر و بسیار زیاد به روش نمونه‌گیری تصادفی ساده با جایگذاری نمونه گرفته می‌شود. پس از اتمام این گام، آماره‌های برآورد شده در هر مرحله با یکدیگر تلفیق شده و آماره‌های موسوم به بوت استرپ حاصل می‌شوند (۶،۷). برای این منظور، فرض کنید که نمونه‌های مستقل $\mathbf{y}_n = (y_1, y_2, ..., y_n)$ در اختیار باشند. پس از تولید بازنمونه‌های $\mathbf{y}_n^* = (y_1^*, y_2^*, ..., y_n^*)$ به تعداد زیاد (برای مثال به تعداد B)، تقریب مونت کارلوی براورد بوت استرپ برای میانگین و واریانس آماره Tn به صورت زیر به دست می‌آید:

$$\hat{E}^*(T_n) = \frac{1}{B}\sum_{b=1}^{B}\left(T_n(y^*_{n,b})\right)$$

$$\hat{\mathrm{var}}^*\left(T_n(\mathbf{y}^*_n)\mid \mathbf{y}_n\right) = \frac{1}{B-1}\sum_{b=1}^{B}\left(T_n(\mathbf{y}^*_{n,b}) - \frac{1}{B}\sum_{j=1}^{B}T_n(\mathbf{y}^*_{n,j})\right)^2 = S^{*2}$$

پس از براورد میانگین و واریانس پارامتر مورد نظر، فاصله اطمینان ۹۵ درصد برای آن به‌صورت زیر قابل محاسبه است.

$$\left(\hat{E}^*(T_n) - t_{0.975}(n-1)S^*\,;\, \hat{E}^*(T_n) + t_{0.975}(n-1)S^*\right).$$

لازم به ذکر است انجام این مراحل مستلزم به‌کارگیری یک نرم‌افزار مناسب آماری است.

۲) روش بازنمونه‌گیری جک نایف:

روش بازنمونه‌گیری جک نایف یک روش نمونه‌گیری بدون جایگذاری است، که برای برآورد مقدار اریبی و واریانس، روشی مناسب و مفید است. برای مثال، در محاسبه‌ی فاصله اطمینان آماره مورد نظر (مانند میانگین) روال کار در این نوع روش باز نمونه‌گیری بدین صورت است که مشاهدات تک به تک از نمونه اصلی خارج می‌شود و نمونه جک نایف ایجاد می‌شود. در ادامه، برای هر نمونه‌ی جک نایف آماره مورد نظر محاسبه می‌شود:

$$y_j^* = nY_n - (n-1)Y_j \qquad j = 1, 2, ..., n$$

که در آن Y_n بیان‌گر برآوردگر نمونه کامل، Y_j بیان‌گر برآوردگر نمونه فاقد مشاهده j-ام و y_j^* بیان‌گر j- امین شبه مقدار تولید شده است. حال میانگین برآورد شده y^* و واریانس y^*، به‌ترتیب به‌صورت

$$\bar{y}^* = \frac{\sum_{j=1}^{n} y_j^*}{n}$$

و

$$S^{*2} = \frac{1}{n}\left(\frac{\sum_{j=1}^{n} y_j^{*2} - n\bar{y}^{*2}}{n-1}\right)$$

محاسبه می‌شوند (۶).

در ادامه، فاصله اطمینان ۹۵ درصد برای میانگین y^* بر اساس توزیع تی استودنت به‌صورت زیر قابل دست‌یابی است:

$$\left(\bar{y}^* - t_{0.975}(n-1)S^*\,;\, \bar{y}^* + t_{0.975}(n-1)S^*\right)$$

لازم به ذکر است، به دلیل پیچیدگی کم‌تری که روش جک نایف در مقایسه با روش بوت استرپ دارد، اجرای این روش با نرم‌افزار در مدت زمان کم‌تری انجام می‌شود.

یافته‌ها

میانگین± انحراف معیار سنی بیماران ۰/۴۷±۲۹/۱۱ سال بود. زمان بقای این بیماران ۴۹-۵ سال بود. میانگین± خطای معیار زمان بقای این بیماران، با استفاده از روش کاپلان- مایر، ۱/۱۲±۴۱/۶۶ به دست آمد. نکته مهم کاهش شدید حجم نمونه هنگام برازش بهترین مدل روی داده‌های مورد نظر بود. در واقع، با برازش مدل لوگ-نرمال حجم نمونه از ۲۹۶ به ۵۰ کاهش یافت. در جدول شماره ۲، نتیجه‌ی مربوط به برازش مدل لوگ-نرمال روی داده‌ها گزارش شده است. در ادامه، از روش‌های بازنمونه‌گیری استفاده شد. برای این منظور، مشاهده‌های گمشده از تحلیل خارج شده و نمونه‌های بوت‌استرپ و جک‌نایف بر اساس داده‌های کامل تولید شدند، و نمونه‌های با اندازه ۵۰ بازتولید شدند. برای روش جک‌نایف تعداد ۵۰ بار نمونه‌های با حجم ۵۰ تولید شد و به دلیل رخداد خطای اجرای برنامه در روش بازنمونه‌گیری بوت استرپ، تعداد تکرار به ۱۵ بار محدود شد. در پایان از نمونه‌های ایجاد شده برای برازش مدل بقای لوگ-نرمال استفاده شد. (جدول‌های شماره ۴و۳).

[Downloaded from irje.tums.ac.ir on 2024-09-08]

با انجام روش‌های بازنمونه‌گیری بوت استرپ و جک نایف، خطاهای معیار، مقادیر احتمال و فواصل اطمینان مربوط به متغیرها به شدت تحت تأثیر قرار گرفته و تغییر کردند (جدول‌های شماره ۴و۳). به‌عبارتی، این روش‌ها منجر به کاهش قابل ملاحظه در خطاهای استاندارد شدند و همین کاهش باعث معنی‌داری برخی از متغیرها شد. در عین حال، فواصل اطمینان نیز به‌طور قابل ملاحظه‌ای تنگ‌تر شدند. طول فواصل اطمینان برای هر سه روش در جدول شماره ۵ گزارش شده است. با مقایسه‌ی طول فواصل اطمینان، مشاهده می‌شود که طول فواصل اطمینان در روش جک نایف، در بیش‌تر متغیرها، تنگ‌تر از طول فواصل اطمینان در روش بوت استرپ شده است.

جدول شماره ۱- وضعیت مشاهده‌های نامشخص در متغیرهای بررسی شده بیماران مبتلا به تالاسمی ماژور

متغیر	تعداد	درصد
جنس بیمار		
مرد/زن	۲۹۱	۹۸/۳
نامشخص	۵	۱/۷
زادگاه بیماران		
هم‌جوار/ غیر هم‌جوار دریا	۲۵۲	۸۵/۱
نامشخص	۴۴	۱۴/۹
تحصیلات مادر		
بی‌سواد/ زیر دیپلم/ دیپلم/ دانشگاهی	۲۰۵	۶۹/۳
نامشخص	۹۱	۳۰/۷
تحصیلات پدر		
بی‌سواد/ زیر دیپلم/ دیپلم/ دانشگاهی	۲۰۵	۶۹/۳
نامشخص	۹۱	۳۰/۷
وضعیت پیشامد		
رخداد/عدم رخداد مرگ	۲۰۶	۶۹/۶
نامشخص	۹۰	۳۰/۴
گروه سنی*		
متولد پیش/پس از سال ۱۳۶۰	۲۸۷	۹۶/۹
نامشخص	۹	۳/۱
وضعیت طحال		
برداشته شده/نشده	۲۰۵	۶۹/۳
نامشخص	۹۱	۳۰/۷
نوع دسفرال		
زیر جلدی/ پمپی/ هر دو/ هیچ‌کدام	۱۸۶	۶۲/۸
نامشخص	۱۱۰	۳۷/۲
نسبت والدین		
غریبه/ فامیل دور/ فامیل درجه‌یک	۱۷۵	۵۹/۱
نامشخص	۱۲۱	۴۰/۹

* با توجه به شروع اقدامات جدی درمانی و بهداشتی در دهه‌ی ۶۰، این دسته‌بندی صورت گرفته است.

[Downloaded from irje.tums.ac.ir on 2024-09-08]

جدول شماره ۲- نتایج حاصل از برازش مدل بقای لوگ-نرمال روی داده‌های تالاسمی

متغیرها	ضریب رگرسیونی	خطای معیار	مقدار احتمال	فاصله اطمینان ۹۵ درصد
ضریب ثابت	۱/۹۷	۵۷/۱۱	۰/۹۷۲	(-۱۰۹/۹۶و۱۱۳/۹۰)
زادگاه	۰/۳۷	۰/۱۲	۰/۰۰۲	(۰/۱۶و۰/۶۱)
سن ابتلا	۰/۳۳	۰/۱۲	۰/۰۰۶	(۰/۱۰و۰/۵۶)
سن دسفرال	-۰/۰۷	۰/۰۲	<۰/۰۰۱	(-۰/۰۴ و-۰/۱۱)
طحال برداری	۱/۴۶	۱۴/۴۴	۰/۹۱۹	(-۲۶/۸۴و۲۹/۷۶)
تحصیلات مادر	۰/۳۰	۰/۰۹	<۰/۰۰۱	(۰/۱۲و۰/۴۹)
نسبت والدین	۰/۵۷	۱۲/۴۱	۰/۹۶۳	(-۲۳/۷۵و۲۴/۸۹)
سن شروع	-۰/۱۲	۰/۰۵	۰/۰۱۶	(-۰/۰۲ و-۰/۲۲)
گروه سنی	-۰/۷۷	۲۶/۹۲	۰/۹۷۷	(-۵۳/۵۳ و ۵۱/۹۹)

جدول شماره ۳- نتایج حاصل از برازش مدل بقای لوگ-نرمال روی داده‌های تالاسمی پس از انجام روش بازنمونه‌گیری بوت استرپ

متغیرها	ضریب رگرسیونی	خطای معیار	مقدار احتمال	فاصله اطمینان ۹۵ درصد
ضریب ثابت	۱/۹۷	۰/۰۶۰	<۰/۰۰۱	(۱/۸۵ و ۲/۰۹)
زادگاه	۰/۳۷	۰/۰۲۰	<۰/۰۰۱	(۰/۲۶ و ۰/۴۹)
سن ابتلا	۰/۳۳	۰/۰۱۰	<۰/۰۰۱	(۰/۲۹ و ۰/۳۷)
سن دسفرال	-۰/۰۷	۰/۰۰۴	<۰/۰۰۱	(-۰/۰۸ و-۰/۰۶)
طحال برداری	۱/۴۶	۰/۷۵۰	۰/۰۵۲	(-۰/۰۱ و ۲/۹۳)
تحصیلات مادر	۰/۳۰	۰/۰۰۳	<۰/۰۰۱	(۰/۲۹ و ۰/۳۱)
نسبت والدین	۰/۵۷	۰/۰۱۰	<۰/۰۰۱	(۰/۵۵ و ۰/۵۸)
سن شروع	-۰/۱۲	۰/۰۰۳	<۰/۰۰۱	(-۰/۱۲ و-۰/۱۱)
گروه سنی	-۰/۷۷	۰/۲۹۰	۰/۰۰۸	(-۱/۳۵ و-۰/۲)

جدول شماره ٤- نتایج حاصل از برازش مدل بقای لوگ-نرمال بر داده‌های تالاسمی پس از انجام روش بازنمونه‌گیری جک نایف

متغیرها	ضریب رگرسیونی	خطای معیار	مقدار احتمال	فاصله اطمینان ۹۵ درصد
ضریب ثابت	۱/۹۷	۰/۹۹۰	۰/۰۵۳	(-۰/۰۳ و ۳/۹۷)
زادگاه	۰/۳۷	۰/۰۰۰۲	P<۰/۰۰۱	(۰/۳۷ و ۰/۳۸)
سن ابتلا	۰/۳۳	۰/۰۰۱	P<۰/۰۰۱	(۰/۳۲ و ۰/۳۳)
سن دسفرال	-۰/۰۷	۰/۰۰۰۲	P<۰/۰۰۱	(-۰/۰۷ و-۰/۰۶)
طحال برداری	۱/۴۶	۰/۹۸۰	۰/۱۴۵	(-۰/۵۳ و ۳/۴۵)
تحصیلات مادر	۰/۳۰	۰/۰۰۱	P<۰/۰۰۱	(۰/۲۹ و ۰/۳۰)
نسبت والدین	۰/۵۷	۰/۱۲۰	P<۰/۰۰۱	(۰/۳۲ و ۰/۸۲)
سن شروع	-۰/۱۲	۰/۰۰۱	P<۰/۰۰۱	(-۰/۱۲ و-۰/۱۱)
گروه سنی	-۰/۷۷	۰/۰۳	P<۰/۰۰۱	(-۰/۸۴ و-۰/۷۰)

[Downloaded from irje.tums.ac.ir on 2024-09-08]

جدول شماره ۵- طول فواصل اطمینان برای مدل بقای لوگ - نرمال پیش و پس از انجام روش‌های بازنمونه‌گیری بوت استرپ و جک نایف

متغیرها	روش معمولی	روش بوت استرپ	روش جک نایف
ضریب ثابت	۲۲۳/۸۷	۰/۲۴	۴/۰۰
زادگاه	۰/۴۴	۰/۲۳	۰/۰۱
سن ابتلا	۰/۴۶	۰/۱۸	۰/۰۱
سن دسفرال	۰/۰۷	۰/۰۲	۰/۰۱
طحال برداری	۵۶/۶۰	۲/۹۴	۳/۹۸
تحصیلات مادر	۰/۳۷	۰/۰۲	۰/۰۱
نسبت والدین	۴۸/۶۵	۰/۰۳	۰/۵
سن شروع	۰/۲۰	۰/۰۱	۰/۰۱
گروه سنی	۱۰۵/۵۲	۱/۱۵	۰/۱۴

بحث

در برخی از مطالعه‌ها، حجم نمونه کم به دلایل مختلف، مانند وجود داده‌های گمشده، رخ می‌دهد و باعث مخدوش شدن اعتبار نتایج حاصل از تحلیل‌های آماری و به‌طور خاص تحلیل بقا می‌شود (۱۵). به‌کارگیری روش‌ها و الگوریتم‌هایی که بتواند آثار سوء ناشی از کم شدن حجم نمونه را تقلیل دهد، ضروری به نظر می‌رسد. روش‌های بازنمونه‌گیری بوت استرپ و جک نایف از جمله روش‌هایی هستند که می‌توانند برای این منظور به کار گرفته شوند. کورتال دِبوس[۱] و همکارانش روش بوت استرپ را برای به‌کارگیری در موارد مختلفی از جمله وقتی که اندازه نمونه برای انجام یک استنباط آماری ناکافی است، پیشنهاد کرده‌اند (۱۶). بنی‌اسدی[۲] و همکاران روش بازنمونه‌گیری بوت استرپ را در رگرسیون لجستیک و کاربرد آن در تحلیل داده‌های مربوط به سرطان سینه به‌کار گرفتند. آن‌ها نشان دادند که در صورت کمبود حجم نمونه در رگرسیون لجستیک، به‌کارگیری روش بازنمونه‌گیری بوت استرپ و نیز افزایش تکرارها، باعث نزدیکی نتایج حاصل به نتایج حاصل از داده‌های کامل می‌شود (۱۴). در واقع، روش باز نمونه‌گیری بوت‌استرپ یک روش ساده است که می‌تواند برای برآورد خطای استاندارد و بازه اطمینان مورد استفاده قرار گیرد. استفاده از این نوع نمونه‌گیری می‌تواند منجر به کاهش اثر خطای نمونه‌گیری تصادفی شود (۶). روش بازنمونه‌گیری جک نایف آثار سوء ناشی از کاهش حجم نمونه را به حداقل می‌رساند و برای براورد واریانس و میزان اریبی هم قابل استفاده است (۵،۷).

ژو[۳] و همکاران با استفاده از روش ترکیبی بازنمونه‌گیری جک نایف با میانگین حسابی نشان دادند که این روش برای بازسازی ژن از عملکرد بالاتری برخوردار است (۱۷). آن[۴] و همکاران از روش‌های باز نمونه‌گیری برای ارزیابی بالینی پیش‌بینی بقای کلی بیماران، نتایج فاز سوم آزمایه بالینی، با استفاده از آزمایش‌های شبیه‌سازی شده فاز دوم استفاده کردند و نتایج مشابهی با اندازه‌های مربوط به ارزیابی‌های بالینی نتیجه گرفتند (۱۸). پوی‌ریر[۵] و همکاران یک روش بازنمونه‌گیری بوت‌استرپ را در تجزیه و تحلیل زمان تا نفروپاتی با یک مدل مخاطرات متناسب کاکس با داده‌های پلی‌مورفیسم تک نوکلئوتیدی استفاده نمودند. آن‌ها نتیجه گرفتند که به طور کلی این روشِ کاهش - اریبی، برای مطالعه‌های پیچیده شامل متغیرهای کمی، کیفی، زمان تا رخداد (بقا) قابل استفاده است (۱۹). تانگ[۶] و همکاران هشت فاصله اطمینان برای اختلاف بین دو نسبت همبسته مبتنی بر روش‌های مختلف، از جمله روش بازنمونه‌گیری بوت استرپ، ایجاد کردند و عرض فاصله اطمینان کوتاه‌تری نتیجه گرفتند (۲۰). سوریانو[۷] و همکاران در مطالعه‌ای به ارزیابی روش جک‌نایف و بوت‌استرپ برای اندازه‌های زوجی پرداختند. یافته‌های شبیه‌سازی این مطالعه نشان داد که دقت برآورد فاصله اطمینان در روش جک‌نایف بهتر از بوت‌استرپ به‌دست آمد. در عین حال گزارش کردند که پوشش فاصله اطمینان روش جک‌نایف نسبت به تغییرات تعداد خوشه و اندازه خوشه استوار است (۲۱). در این مطالعه، پس از برازش مدل‌های

[۱] Korthals-de Bos
[۲] Baniasadi M
[۳] Zhou C
[۴] Ann MW
[۵] Poirier JG
[۶] Tang NS
[۷] Severiano A

[Downloaded from irje.tums.ac.ir on 2024-09-08]

بقای مختلف و انتخاب مدل لوگ-نرمال به‌عنوان بهترین مدل، به دلیل وجود داده‌های گم شده در متغیرها، نمونه از ۲۹۶ به ۵۰ کاهش یافت. با به‌کارگیری الگوریتم‌های بازنمونه‌گیری بوت استرپ و جک نایف خطاهای استاندارد کوچک‌تر و طول فواصل اطمینان تنگ‌تری برای ضرایب متغیرهای مدل لوگ-نرمال به دست آمد. با مقایسه طول فواصل اطمینان، روش جک نایف در مقایسه با روش بوت استرپ به‌طور نسبی عملکرد بهتری داشت. یکی از محدودیت‌های اجرایی در تحلیل داده‌ها، در روش بوت استرپ رخ داد. در این روش نرم‌افزار قادر به انجام تکرارهای بیش‌تر نبود و براوردهای این روش با ۱۵ تکرار انجام شد.

نتیجه‌گیری

یافته‌های این مطالعه نشان داد که به‌کارگیری روش‌های بازنمونه‌گیری بوت استرپ و جک نایف باعث کاهش خطای استاندارد و نیز تنگ‌تر شدن فواصل اطمینان برای ضرایب متغیرهای موجود در مدل‌های بقای دارای حجم نمونه کم می‌شوند. در این بین، روش بازنمونه‌گیری جک نایف در مقایسه با روش بازنمونه‌گیری بوت استرپ به نسبت عملکرد بهتری داشت.

تشکر و قدردانی

نویسندگان این مقاله بر خود لازم می‌دانند مراتب تشکر و قدردانی خود را از سرکار خانم فروغی مسؤول بخش بایگانی کلینیک ظفر و تمامی کارکنانی که در جمع‌آوری داده‌ها همکاری داشته‌اند، اعلام نمایند.

منابع

1. Dupont WD, Plummer WD Jr. Power and sample size calculations for studies involving linear regression. Control Clin Trials. 1998; 19: 589-601.
2. Efron B. The bootstrap and Modern statistics. JASA. 2000; 95: 1293-96.
3. Efron B, Tibshirani RJ. An introduction to the bootstrap. NewYork, Chapman and Hall/CRC. 1993, 15.
4. Efron B. Better bootstrap confidence intervals (with discussions) J Amer Statist Assoc. 1987; 82: 171- 200.
5. Crowley PH. Resampling methods for computation-intensive data analysis in ecology and evolution. Ann Rev Ecol Sys. 1992; 23: 405-47.
6. Shao J, Tu D. The Jackknife and Bootstrap. Springer-Verlag, Inc. 1995, pp:17-19,60,104,127.
7. Sahinler S, Topuz D. Bootstrap and jackknife resampling algorithms for estimation of regression parametters. J App Quan Meth. 2007; 2: 188-99.
8. Chen C, Sun P, Dai QS, Weng HW, Li HP, Ye S. The Glasgow Prognostic Score predicts poor survival in cisplatin-based treated patients with metastatic nasopharyngeal carcinoma. PLoS One. 2014; 9: e112581.
9. Denne C, Maag S, Heussen N, Häusler M. A new method to analyse the pace of child development: Cox regression validated by a bootstrap resampling procedure. BMC Pediatr. 2010; 10:12. doi: 10.1186/1471-2431-10-12.
10. Zanata RL, Fagundes TC, Freitas MC, Lauris JR, Navarro MF. Ten-year survival of ART restorations in permanent posterior teeth. Clin Oral Investig. 2011; 15: 265-71. doi: 10.1007/s00784-009-0378-x.
11. Molina GF, Faulks D, Mazzola I, Mulder J, Frencken JE. One year survival of ART and conventional restorations in patients with disability. BMC Oral Health. 2014; 14:49. doi: 10.1186/1472-6831-14-49.
12. Ali Akbari Khoei R, Bakhshi E, Azarkeivan A, Biglarian A. Survival analysis of the thalassemia major patients using parametric and semiparametric survival models. J Health Admin (JHA). 2015; 18: 82-91.
13. Hanna SR. Confidence limits for air quality model evaluations, as estimated by bootstrap and jackknife resampling methods. Atmos Environ. 1989; 23: 1385-98.
14. Baniasadi M, BaBaie G, Zeraati H, Memari F. Application of bootstrap sample-resample method in logistic regression in analysis of breast cancer data. SJSPH. 2006; 4: 9-18.
15. Mir Mohammadkhani M, Holakouei Naieni K. Application of multiple imputations in medical and epidemiological research. SJSPH. 2012; 9: 77-88.
16. Korthals-de Bos IB, Hoving JL, van Tulder MW, Rutten-van Mölken MP, Adèr HJ, et al. Cost effectiveness of physiotherapy manual therapy, and general practitioner care for neck pain: economic evaluation alongside a randomized controlled trial. BMJ. 2003; 326: 911.
17. Zhou C, Zhang SW, Liu F. An ensemble method for reconstructing gene regulatory network with jackknife resampling and arithmetic mean fusion. Int J Data Min Bioinform. 2015;12: 328-42.
18. An MW, Han Y, Meyers JP, Bogaerts J, Sargent DJ, Mandrekar SJ. Clinical Utility of Metrics Based on Tumor Measurements in Phase II Trials to Predict Overall Survival Outcomes in Phase III Trials by Using Resampling Methods. J Clin Oncol. 2015; 33: 4048-57.
19. Poirier JG, Faye LL, Dimitromanolakis A, Paterson AD, Sun L, Bull SB. Resampling to Address the Winner's Curse in Genetic Association Analysis of Time to Event. Genet Epidemiol. 2015; 39: 518-28.
20. Tang NS, Li HQ, Tang ML, Li J. Confidence interval construction for the difference between two correlated proportions with missing observations. J Biopharm Stat. 2015 Jan 29. [Epub ahead of print]
21. Severiano A, Carriço JA, Robinson DA, Ramirez M, Pinto FR. Evaluation of jackknife and bootstrap for defining confidence intervals for pairwise agreement measures. PLoS One. 2011; 6: e19539.

[Downloaded from irje.tums.ac.ir on 2024-09-08]

Iranian Journal of Epidemiology 2016; 12(3): 56-63.
Original Article

Bootstrap and Jackknife Resampling Methods in Survival Analysis of Patients with Thalassemia Major

Ali Akbari Khoei R[1], Bakhshi E[2], Azarkeivan A[3], Biglarian A[2]

1- MSc in Biostatistics, Department of Biostatistics, University of Social Welfare and Rehabilitation Sciences, Tehran, Iran
2- Associate Professor, Department of Biostatistics, University of Social Welfare and Rehabilitation Sciences, Tehran, Iran
3- Associate Professor, Pediatric Hematology Oncology, Iranian Blood Transfusion Organization (IBTO) – High Institute for Research and Education in Transfusion Medicine, Thalassemia Clinic, Tehran, Iran
Corresponding author: Biglarian A, abiglarian@uswr.ac.ir

Background and Objectives: A small sample size can influence the results of statistical analysis. A reduction in the sample size may happen due to different reasons, such as loss of information, i.e. existing missing value in some variables. This study aimed to apply bootstrap and jackknife resampling methods in survival analysis of thalassemia major patients.

Methods: In this historical cohort study, the data of 296 patients with thalassemia major who were visited at Zafar Clinic, Tehran, from 1994 to 2013 were used. Parametric survival models were used to analyze the data. The log – normal survival model was selected as the best model and then the bootstrap and jackknife resampling algorithms were used for this model. Data analysis was carried out with the STATA 12.0 software.

Results: The results of the resampling methods showed that standard errors decreased and confidence intervals were shortened. In addition, the result of the bootstrap and jackknife resampling methods showed that age group and the relationship of the parents ($P<0.001$) were significant compared with the log-normal model ($P>0.900$).

Conclusion: Comparison of the confidence intervals suggests that the jackknife resampling method can be used when the sample size is small.

Keywords: Survival analysis, Resampling methods, Thalassemia major

[Downloaded from irje.tums.ac.ir on 2024-09-08]