

بکارگیری روش باز نمونه گیری بوت استرپ در رگرسیون لجستیک و کاربرد آن در تحلیل داده های مربوط به بیماران مبتلا به سرطان سینه

محمد بنی اسدی: کارشناس ارشد گروه آمار زیستی، دانشگاه تربیت مدرس

دکتر غلامرضا بابایی: دانشیار گروه آمار زیستی، دانشگاه تربیت مدرس

دکتر حجت زراعتی: استادیار گروه اپیدمیولوژی و آمار زیستی، دانشکده بهداشت و انستیتو تحقیقات بهداشتی، دانشگاه علوم پزشکی

تهران - نویسنده رابط zeraati@modares.ac.ir

دکتر فریدون معماری: استادیار گروه جراحی عمومی دانشکده پزشکی، دانشگاه علوم پزشکی تهران

دریافت: ۸۳/۷/۲۷ پذیرش: ۸۴/۲/۴

چکیده:

زمینه و هدف: ارائه فواصل اطمینان برای پارامترهای مجهول در مطالعات مختلف معمولاً براساس پذیرش پیش فرض هایی است که از آن جمله می توان به ضرورت بزرگ بودن حجم نمونه (به اندازه کافی) اشاره داشت. هدف از انجام این تحقیق، بررسی چگونگی به کارگیری روش بوت استرپ (Bootstrap) در مدل رگرسیون لجستیک (Logistic Regression) و نشان دادن ضرورت بکارگیری روش بوت استرپ در مدل های رگرسیونی لجستیک در مواردی است که اطلاعات کافی در اختیار محقق نباشد.

روش کار: برای این منظور اطلاعات مربوط به ۱۵۰ بیمار را که در فاصله سال های ۱۳۷۸ تا ۱۳۸۰ در انستیتو کانسر مجتمع بیمارستانی امام خمینی تهران به علت ابتلا به سرطان سینه مورد عمل جراحی قرار گرفته اند استخراج کردیم، سپس نمونه تصادفی ۵۰ تایی از بین ۱۵۰ بیمار استخراج گردید.

نتایج: با استفاده از روش معمول رگرسیون لجستیک، مدل مناسب و معنی داری بر اطلاعات اولیه برازش شد و برای هر یک از ضرایب مدل، فاصله اطمینان و خطای معیار محاسبه گردید. نتایج حاصل از کاربرد روش معمول و استفاده از روش بوت استرپ در نمونه های متفاوت نشان می دهد که کاربرد روش بوت استرپ در مواردی که با حجم نمونه ناکافی مواجه هستیم و شرایط مناسب برای استفاده از روش معمول رگرسیون لجستیک فراهم نیست، ضروری بوده و نتایج به دست آمده با افزایش تکرارهای بوت استرپ، به نتایج حاصل از داده های کامل نزدیک می شود.

نتیجه گیری: این امر، هم در مورد برآورد ضرائب مدل و هم در مورد فاصله اطمینان و خطای مربوط به ضرائب، صادق است.

واژگان کلیدی: بوت استرپ، رگرسیون لجستیک، سرطان پستان، باز نمونه گیری

مقدمه:

سالانه بیش از ۸ میلیون نفر به انواع سرطان مبتلا می شوند (متوسط بروز سالانه ۱۴۳ در ۱۰۰۰۰۰ نفر می باشد) سرطان های ریه (با ۱۲۲۱۰۰۰ مورد)، معده (با ۱۰۱۳۰۰۰ مورد)، و پستان (حدود یک میلیون مورد) فراوانترین سرطان ها را در سطح دنیا تشکیل می دهند. طبق نظر کتب مرجع در کشور آمریکا از هر ۹ زن یک

سرطان پستان شایعترین سرطان در بین زنان بوده و ۱۸٪ کل سرطان های این گروه را تشکیل می دهد به گونه ای که سالانه در جهان یک میلیون مورد جدید از این بیماری بروز می کند. در کشورهای توسعه یافته حدود یک سوم کل سرطان های زنان را تشکیل می دهد. در دنیا

نفر و طبق بعضی مقالات جدیدتر از هر ۸ زن یک نفر مبتلا به سرطان پستان می‌گردد (برونر ۱۳۷۱). نسبت بروز سرطان پستان در کم شیوع ترین نقاط جهان به پر شیوع ترین نقاط آن یک به پنج است (محققی ۱۳۸۱). به عبارت دیگر حتی در کم شیوع ترین نقاط جهان، این سرطان درصد قابل توجهی از زنان را مبتلا می‌سازد. از دهه چهل میلادی میزان بروز سرطان پستان به طور ثابت در حال افزایش بوده است (محققی ۱۳۸۱). طبق نظر سازمان بهداشت جهانی هم اکنون سالانه حدود دو درصد به میزان بروز سرطان پستان در سراسر جهان افزوده می‌شود (محققی ۱۳۸۱).

مطالعات انجام شده در این زمینه نشان دهنده اهمیت بسیار زیاد این بیماری می‌باشد. آمار شیوع سرطان پستان در ایران حاکی از آن است که ایران جزو مناطق کم شیوع نمی‌باشد. تخمین زده می‌شود که کشور ما در حد وسط مناطق با حداکثر شیوع و حداقل شیوع بوده و میزان بروز در بین زنان ما نیز در حال افزایش می‌باشد. در مطالعه حاضر سعی شده است که به وسیله تکنیک های آماری عوامل مؤثر بر درمان این بیماری شناسایی شود.

روش رگرسیون لجستیک یکی از روشهایی است که توسط آن می‌توان اثرهمزمان متغیرهای مختلف را بر نتیجه درمان مورد بررسی قرار داد (Agresti A. ۱۹۹۶). روش های مرسوم ارائه برآورد برای پارامترهای رگرسیون لجستیک بر مبنای پذیرش تئوری نمونه های بزرگ است، و استفاده از این روش ها برای نمونه های کوچک خیلی قابل اعتماد نیست. باید توجه داشت که برای برازش مدل رگرسیونی لجستیک بر یک سری اطلاعات، مجبور به پذیرش فرضیاتی هستیم که عدم برقراری هر یک از فرضیات باعث اریبی زیادی در تحلیل ما می‌شود. از آن جمله می‌توان داشتن حجم نمونه معین

و مشخص راعنوان کرد که برای تعیین حجم نمونه باتوجه به فرمول ارائه شده، نیز باید فرضیاتی راپبذیریم؛ داشتن حجم نمونه معین، شرایط برقراری قضیه حد مرکزی واستفاده از تئوری نمونه های بزرگ را فراهم می‌کند وتنها به شرط در اختیار داشتن حجم نمونه لازم مجاز به استفاده از روش های رگرسیونی برای تحلیل اطلاعات تحقیق خواهیم بود (Agresti A. ۱۹۹۶). برای استفاده از روش های روش های مرسوم جهت برآورد پارامترها در رگرسیون لجستیک پذیرش پیش فرض هایی مبنی بر نرمال بودن توزیع و بزرگ بودن نمونه (به اندازه کافی) ضروری است، و عدم برقراری این پیش فرض ها اعتبار استنتاج های انجام شده را تحت الشعاع قرار خواهد داد. این که تا چه اندازه برآورد ضرایب، با توجه به فرضیات پذیرفته شده، صحیح و دقیق هستند یا به عبارت دیگر، نمونه ای که طبق آن، این ضرائب به دست آمده است، تا چه اندازه می‌تواند بیانگر جامعه واقعی باشد، نیز جای بحث دارد (Efron B. ۲۰۰۰). بوت استرپ روشی است که فارغ از بسیاری از فرضیات، با ایجاد نمونه های فراوان، شرایط نمونه را به شرایط جامعه نزدیک می‌کند و بادر نظرگرفتن تمامی حالات تشکیل نمونه، می‌توان از صحت برآورد ضرائب و برآورد فاصله اطمینان برای ضرائب، مطمئن تر شد (Efron B. and Tibshirani ۱۹۹۳). در استفاده از این روش نیازمند پذیرش پیش فرض های مربوط به نرمال بودن توزیع و نمونه بزرگ نخواهیم بود، و نتایج از اعتبار مناسبی برخوردار خواهند شد، هرچند بهایی متناظر با طویل تر شدن فواصل اطمینان خواهیم پرداخت. چون در مطالعات پزشکی_آماری دستیابی به نمونه های بزرگ برای رسیدن به نتیجه های مطلوب به عنوان یک مشکل اساسی محسوب می‌شود (به گونه ای که در صورت فقدان نمونه کافی شرایط برای استفاده از روش رگرسیون لجستیک

تومور شامل طول، عرض و ضخامت (میلی متر)، تعداد غدد لنفاوی درگیر، وجود متاستاز (داشته _ نداشته)، محل متاستاز، این که بیمار در چه مرحله ای از بیماری به پزشک مراجعه کرده است (در سیستم TNM با اعداد ۱ تا ۴)، گروه خونی بیمار، تعداد واحد خون دریافتی حین عمل، تکنیک جراحی به کار رفته برای بیمار، نوع پاتولوژی، مدت زمان بیهوشی و مدت زمان بستری بعد از عمل جراحی.

عوامل ذکر شده در بالا در قالب متغیرهای کیفی و کمی دسته بندی و کدگذاری شده و با استفاده از روش رگرسیون لجستیک عوامل مؤثر بر متغیر پاسخ تعیین شده است (برای نمونه ۱۵۰ تایی). طبق محاسبات انجام شده، حجم نمونه مورد نیاز برای انجام این تحقیق برابر ۱۵۰ به دست آمده است. در ادامه مطالعه کاربرد بوت استرپ را در رگرسیون لجستیک با فرض اینکه نمونه کافی در اختیار نباشد، با کمک ماشین یک نمونه ۵۰ تایی به تصادف از بین نمونه های موجود (۱۵۰ تایی) انتخاب کرده ایم، بر اساس این نمونه کاهش یافته (۵۰ تایی) و با استفاده از نرم افزار یکبار نمونه ۵۰ تایی (برابر با نمونه پایه) و یکبار نمونه ۱۵۰ تایی (برابر با حجم نمونه لازم برای انجام رگرسیون لجستیک در این مطالعه) به عنوان نقطه شروع عملیات باز نمونه گیری بوت استرپ تولید کردیم و تکرارهای بوت استرپ را ۲۵۰، ۵۰۰، ۷۵۰ و ۱۰۰۰ در نظر گرفتیم. نتایج حاصل از این روش تلفیقی را با روش معمول رگرسیون لجستیک مقایسه کرده ایم. برای هر یک از نمونه های بوت استرپ، ضرایب مدل رگرسیون لجستیک برآورد شده و همچنین برای هر ضریب فاصله اطمینان بوت استرپی و خطای معیار با روش بوت استرپ برآورد شده است. برای بررسی تاثیر بوت استرپ، برآوردهای بوت استرپی با برآوردهای

مناسب نخواهد بود (چگونگی استفاده از روش باز نمونه گیری بوت استرپ در این مطالعه مورد بررسی قرار می گیرد؛ این روش در سال ۱۹۷۹ توسط افرون (Efron) ارائه شد که به وسیله آن می توان مشکل نبود نمونه بزرگ را از بین برد (Efron B. ۱۹۸۱). با استفاده از این روش می توان برآوردهای نسبتاً دقیقی از پارامترهای جامعه و همینطور ضرایب رگرسیونی به دست آورد (Efron B. ۱۹۸۷).

در این تحقیق ضمن بیان چگونگی کاربرد روش بوت استرپ در رگرسیون لجستیک، کارایی استفاده از این روش را در شناسایی عوامل مؤثر بر بهبود بیماران مبتلا به سرطان سینه که مورد عمل جراحی قرار گرفته اند، بررسی کرده ایم.

روش کار:

برای انجام این مطالعه اطلاعات مربوط به ۱۵۰ بیمار از پرونده های موجود در انستیتو کانسر مجتمع بیمارستانی امام خمینی تهران استخراج شد. برای انتخاب این پرونده ها از روش نمونه گیری تصادفی ساده استفاده شده است. در این مطالعه متغیر پاسخ مدل در دو حالت بهبودی کامل و عدم بهبودی کامل در این مرحله از درمان در نظر گرفته شده است؛ با این تعریف که تجویز درمانهای مجدد یا تکمیلی در فاصله شش ماه از انجام عمل جراحی به عنوان عدم بهبودی کامل و غیر از آن به عنوان بهبودی کامل در این مرحله از درمان لحاظ شده است. از جمله عواملی که به عنوان متغیرهای مدل از پرونده های مذکور استخراج شده و تاثیر آن ها بر متغیر پاسخ مورد بررسی قرار گرفته است به شرح زیر است:

سن بیمار (سال)، وضعیت تأهل بیمار (متاهل - غیرمتاهل)، تعداد فرزندان بیمار، طرف درگیر (وجود تومور در پستان چپ یا راست و یا هر دو طرف)، اندازه

حاصل از روش معمول رگرسیون لجستیک برای هر متغیر مقایسه شده است.

برای آشنایی بیشتر با روش بوت استرپ، الگوریتم ساده ای از چگونگی انجام این روش ارائه شده است.

مرحله ۱- نمونه بوت استرپ $x_1^*, x_2^*, \dots, x_n^*$ را با روش نمونه گیری تصادفی ساده با جایگذاری از نمونه مشاهده شده x_1, x_2, \dots, x_n به دست می آوریم.

مرحله ۲- آماره بوت استرپ $T^* = T(x_1^*, K, x_n^*)$ را محاسبه می کنیم.

مرحله ۳- مراحل ۱ و ۲ را B بار تکرار و B آماره بوت استرپ $\{T_b^*; b=1, K, B\}$ را محاسبه می کنیم (Hall P. ۱۹۹۲).

نتایج :

از ۱۵۰ بیمار مورد بررسی، ۸۰ بیمار (۵۲/۳٪) دارای تومور در سینه راست بوده اند. ۱۳۸ نفر (۹۲/۸٪) متأهل بوده و از این بین، ۱۳۱ نفر آنها دارای فرزند بودند. توزیع سنی بیماران نشان می دهد که بیشتر بیماران بین ۳۵ تا ۵۰ سن دارند.

۷۱ نفر (۴۶/۳٪) دچار متاستاز بوده و ۱۲ نفر (۷/۸٪) از بیماران در مرحله یک (I) بیماری، ۵۸ نفر (۳۷/۹٪) در مرحله دو (II)، ۶۱ نفر (۳۹/۹٪) در مرحله سه (III) بیماری و بقیه در مرحله چهار (IV) بیماری به بیمارستان مراجعه نموده اند. ۱۰۹ نفر (۷۱/۲٪) ماستوکتومی کامل (MRM) شده اند و ۱۲ نفر (۷/۸٪) نیز رزکسیون و اکسیژن و ۱۴ نفر (۹/۲٪) کوادرانوکتومی و بقیه تحت تکنیک های دیگر جراحی قرار گرفته اند و پاتولوژی ۱۲۲ نفر آنها (۷۹/۷٪) داکتال کارسینوما بوده است.

آزمون آماری خی دو و t در یک تحلیل ساده برای سنجش ارتباط متغیرهای کیفی و کمی با نتیجه درمان نشان داد که در بین متغیرهای کیفی هیچ یک از متغیرها اثر معنی داری بر نتیجه درمان نداشته لیکن متغیرهای سن، طول تومور، عرض تومور، قطر تومور، تعداد غدد لنفاوی درگیر، طول مدت بیهوشی و طول مدت بستری بر نتیجه درمان اثر معنی داری داشته اند ($p < 0/01$ ، جداول ۱ و ۲). برای بررسی اثر همزمان متغیرهای مورد بررسی بر نتیجه درمان از مدل رگرسیون لجستیک به روش پسرو (Back ward) استفاده شد که در نتیجه ملاحظه شد که متغیرهای سن، قطر تومور، تعداد تومور، طول مدت بیهوشی و وجود متاستاز بر نتیجه درمان مؤثر بوده اند (جدول ۳).

در ادامه برای بررسی حاصل کاربرد روش بوت استرپ در شرایط عدم برقراری پیش فرض های مربوط به کاربرد مدل های رگرسیونی و یا کافی نبودن تعداد نمونه ها، روش بوت استرپ یک بار برای همه داده ها و بار دیگر برای نمونه کم (۵۰) با تکرارهای مختلف بکار برده شد و در هر روش ضرایب مدل های رگرسیونی، خطای معیار و فاصله اطمینان هر ضریب تعیین و با روش اصلی مقایسه شد که در نتیجه ملاحظه گردید:

- اگر شرایط مربوط به برآورد ضرایب مدل های رگرسیونی برقرار نباشد و از بوت استرپ برای رفع این مشکل با استفاده از تمامی نمونه ها استفاده شود، ملاحظه می شود که (جدول ۴):

- با افزایش تکرارهای بوت استرپ، برآورد ضرایب متغیرها، به برآورد آن ها به روش معمول نزدیک می شود.
- با افزایش تکرارهای بوت استرپ، طول فاصله اطمینان های برآورد شده کوچکتر می شود.
- با افزایش تکرارهای بوت استرپ، خطای برآورد شده کاهش می یابد.

بوت استرپ خطای برآورد به میزان قابل توجهی کاهش می یابد و همین طور ضرایب به دست آمده برای متغیرهای مدل با استفاده از این روش به ضرایب حاصل شده از روش معمول رگرسیون لجستیک در داده های کامل نزدیک می شود با توجه به نتایج حاصله این امر در مورد طول بازه فاصله اطمینان مربوط به ضرایب هر یک از متغیرهای مدل نیز صادق است. بنابراین با توجه به نتایج این تحقیق، استفاده از الگوریتم بوت استرپ در زمانی که حجم نمونه کافی در اختیار نباشد، راه مناسبی جهت برقرار کردن شرایطی است که بتوان از رگرسیون لجستیک استفاده کرد. استفاده از نمونه های بوت استرپ با حجمی برابر با حجم نمونه محاسبه شده در مدل رگرسیون لجستیک با تعداد تکرار بیش از ۷۵۰، باعث حصول نتایجی نزدیک به نتایج به دست آمده از کاربرد روش رگرسیون لجستیک بر داده های کامل می شود.

تشکر و قدردانی :

نویسندگان مراتب تشکر و قدردانی خویش را از معاونت محترم پژوهشی دانشگاه علوم پزشکی و خدمات بهداشتی درمانی تهران که بودجه لازم جهت اجرای این پژوهش را فراهم ساختند ابراز می دارند. همچنین لازم می دانیم که مراتب سپاسگزاری خود را از جناب آقای اکبر نعمت پناه ناظم فنی واحد آمار و مدارک پزشکی انستیتو کانسر و سرکار خانم میرجمهری ناظمه فنی واحد مدارک پزشکی بخش های شیمی درمانی انستیتو کانسر که در استخراج اطلاعات ما را یاری کردند اعلام نماییم.

۲- اگر حجم نمونه کافی در اختیار نباشد و از بوت استرپ برای رفع این مشکل استفاده شود (جدول ۵) :

- با افزایش حجم نمونه های بوت استرپ ، برآورد ضرایب متغیرها به برآورد آن ها در نمونه کامل نزدیک می شود.

- با افزایش حجم نمونه های بوت استرپ ، طول فاصله اطمینان های برآورد شده کوچکتر می شود.

- با افزایش حجم نمونه های بوت استرپ، خطای برآورد شده کاهش می یابد.

بحث:

با توجه به این که برای استفاده از روش رگرسیون لجستیک در تحلیل داده ها وجود شرایطی همچون استقلال متغیرهای توضیحی و همچنین در اختیار بودن حجم نمونه کافی الزامی است در صورت برقرار نبودن هر یک از این شرایط امکان استفاده از روش رگرسیون لجستیک به روش معمول و سنتی وجود ندارد. در تحقیق حاضر نبود حجم نمونه کافی مورد بررسی قرار گرفته است. نتایج به دست آمده در این تحقیق نشان می دهد که استفاده از بوت استرپ باعث نزدیک شدن نتایج حاصل از این روش در حالت کمبود داده به نتایج به دست آمده از روش معمول رگرسیون لجستیک می شود. همچنین دیده می شود که حجم نمونه بوت استرپ و تعداد تکرارهای بوت استرپ در نزدیک شدن نتایج به نتایج حاصل از روش معمول بی تاثیر نیست. نمودارهای موجود در این مقاله مؤید این موضوع است.

نتیجه گیری :

بررسی های ما نشان می دهد زمانی که حدود یک سوم حجم نمونه مورد نیاز در دسترس است با افزایش تکرارهای بوت استرپ و همین طور افزایش حجم نمونه

جدول ۱- ارتباط متغیرهای کیفی مستقل و متغیر پاسخ

| آزمون خی دو (χ^2) | | متغیر |
|--------------------------|-------------|--------------|
| P – Value | مقدار آماره | |
| ۰/۸۸۲ | ۰/۶۶۳ | طرف درگیر |
| ۰/۵۱۴ | ۲/۲۹۰ | وضعیت تاهل |
| ۰/۰۷۵ | ۵/۱۸۶ | محل متاستاز |
| ۰/۸۱۰ | ۱/۵۹۲ | درجه پیشرفت |
| ۰/۴۵۶ | ۴/۶۷۹ | تکنیک جراحی |
| ۰/۹۳۳ | ۰/۴۳۴ | نوع پاتولوژی |

جدول ۲- ارتباط متغیرهای کمی مستقل و متغیر پاسخ

| آزمون (t) | | متغیر |
|-----------|-------------|------------------------|
| P – Value | مقدار آماره | |
| ۰/۰۰ | ۹/۱۴۷ | سن بیمار |
| ۰/۷۴۳ | ۰/۳۲۹ | تعداد فرزندان |
| ۰/۰۰ | ۱۴/۲۱۵ | اندازه طول تومور |
| ۰/۰۰ | ۱۰/۳۳۱ | اندازه عرض تومور |
| ۰/۰۰۳ | ۳/۰۳۴ | اندازه قطر تومور |
| ۰/۰۰ | ۶/۸۳۱ | تعداد غدد لنفاوی درگیر |
| ۰/۳۶۶ | ۰/۹۰۹ | تعداد واحد خونی |
| ۰/۰۰۰ | ۶/۶۱۶ | مدت بیهوشی |
| ۰/۰۰۰ | ۷/۰۹۴ | مدت بستری |

جدول ۳- برآورد ضرائب، خطا و طول فاصله اطمینان برای داده‌های کامل به روش معمول (رگرسیون لجستیک)

| متغیر | β | S.E | Wald | Sig. | $\exp(\beta)$ | طول بازه |
|------------------|---------|-------|--------|-------|---------------|----------|
| مقدار ثابت | -۱۵/۲۶۳ | ۳/۱۳۵ | ۲۳/۷۰۱ | ۰/۰۰۰ | ۰ | |
| سن | ۰/۲۳۷ | ۰/۰۶۲ | ۱۴/۶۸۴ | ۰/۰۰۰ | ۱/۲۶۷ | ۰/۳۰۷ |
| قطر تومور | -۰/۵۳۶ | ۰/۱۸۰ | ۸/۸۳۰ | ۰/۰۰۳ | ۰/۵۸۵ | ۰/۴۲۲ |
| تعداد غدد لنفاوی | ۰/۴۷۷ | ۰/۱۲۵ | ۱۴/۵۳۳ | ۰/۰۰۰ | ۵/۳۰۳ | ۱۲/۲۹۸ |
| مدت بیهوشی | ۱/۶۶۸ | ۰/۵۰۵ | ۱۰/۹۱۴ | ۰/۰۰۱ | ۵/۳۰۳ | ۱۲/۲۹۸ |
| متاستاز | ۲/۱۶۰ | ۰/۷۰۶ | ۹/۳۵۲ | ۰/۰۰۲ | ۸/۶۷۲ | ۳۲/۴۴۵ |

جدول ۴- برآورد ضرایب، خطا و طول بازه برای واحدهای کامل به روش بوت استرپ

| متغیر | تکرار بوت استرپ | برآورد ضریب | برآورد خطا | طول بازه |
|------------------|-----------------|-------------|------------|----------|
| مقدار ثابت | B = ۲۵۰ | -۱۹/۲۶۳۰ | ۱۷/۲۵۸۲ | ۱۳/۱۱۵۹ |
| | B = ۵۰۰ | -۱۹/۲۳۲۰ | ۱۵/۳۷۹۷ | ۱۲/۴۰۱۹ |
| | B = ۷۵۰ | -۱۸/۷۹۶۳ | ۱۲/۹۵۲۵ | ۱۱/۹۸۶۳ |
| | B = ۱۰۰۰ | -۱۸/۷۰۹۴ | ۸/۶۸۰۶ | ۱۱/۸۸۹۴ |
| سن | B = ۲۵۰ | ۰/۳۰۳۹ | ۰/۳۱۸۹ | ۰/۲۸۴۸ |
| | B = ۵۰۰ | ۰/۳۰۴۱ | ۰/۳۲۰۸ | ۰/۲۴۷۲ |
| | B = ۷۵۰ | ۰/۲۹۲۸ | ۰/۲۳۲۲ | ۰/۲۳۸۶ |
| | B = ۱۰۰۰ | ۰/۲۸۹۶ | ۰/۱۶۴۰ | ۰/۲۲۷۴ |
| قطر تو مور | B = ۲۵۰ | ۰/۶۶۷۶ | ۰/۵۴۶۴ | ۱/۷۷۳۸ |
| | B = ۵۰۰ | ۰/۶۷۷۹ | ۰/۷۴۰۲ | ۱/۱۱۶۱ |
| | B = ۷۵۰ | ۰/۶۵۴۸ | ۰/۵۷۱۳ | ۰/۹۹ |
| | B = ۱۰۰۰ | ۰/۶۴۸۲ | ۰/۴۳۵۶ | ۰/۸۷ |
| تعداد غدد لنفاوی | B = ۲۵۰ | ۰/۵۹۴۹ | ۰/۵۵۶۶ | ۰/۵۴۶۹ |
| | B = ۵۰۰ | ۰/۵۷۹۴ | ۰/۴۴۱۳ | ۰/۵۳۹۱ |
| | B = ۷۵۰ | ۰/۵۷۵۷ | ۰/۳۷۰۲ | ۰/۵۰۳۷ |
| | B = ۱۰۰۰ | ۰/۵۷۴۵ | ۰/۲۷۲۴ | ۰/۲۷۲۱ |
| مدت بیهوشی | B = ۲۵۰ | ۲/۰۵۶۲ | ۱/۵۴۸۲ | ۱/۹۸۷۷ |
| | B = ۵۰۰ | ۲/۰۷۷۲ | ۱/۴۰۴۱ | ۱/۶۱۸۸ |
| | B = ۷۵۰ | ۲/۰۳۴۶ | ۰/۹۸۹۳ | ۱/۵۱۹۸ |
| | B = ۱۰۰۰ | ۲/۰۳۰۹ | ۰/۷۸۴۶ | ۱/۳۸۱۴ |
| متاستاز | B = ۲۵۰ | ۲/۷۲۴۷ | ۲/۱۵۰۶ | ۴/۱۷۴۱ |
| | B = ۵۰۰ | ۲/۷۶۲۰ | ۳/۱۸۷۵ | ۴/۰۳۲۵ |
| | B = ۷۵۰ | ۲/۶۷۹۹ | ۱/۹۷۷۶ | ۳/۸۸۸ |
| | B = ۱۰۰۰ | ۲/۶۱۵۵ | ۱/۷۸۷۸ | ۳/۸۷۳۷ |

جدول ۵- برآورد ضرایب، خطا و طول بازه برای واحدهای نمونه گیری شده به روش بوت استرپ

| متغیر (حجم نمونه بوت استرپ) | تکرار بوت استرپ | برآورد ضریب | برآورد خطا | طول بازه |
|--------------------------------|-----------------|-------------|------------|----------|
| مقدار ثابت (n = ۵۰) | B = ۲۵۰ | -۹۷/۵۲۰۹ | ۱۵۵/۳۹۱ | ۱۸۶/۹۴۶۲ |
| | B = ۵۰۰ | -۹۱/۱۳۰۰ | ۱۴۵/۳۹۲ | ۱۱۶/۱۵۶۱ |
| | B = ۷۵۰ | -۸۹/۱۴۷۷ | ۱۴۹/۱۴۳ | ۹۷/۱۷۱۹ |
| مقدار ثابت (n = ۱۵۰) | B = ۱۰۰۰ | -۸۸/۱۴۱۲ | ۱۴۵/۱۷۲ | ۷۴/۴۵۷۸ |
| | B = ۲۵۰ | -۳۴/۵۷۹ | ۷۰/۴۳۶۲ | ۲۸/۶۴۱۷ |
| | B = ۵۰۰ | -۳۱/۱۴۱۴ | ۵۹/۹۰۰۰ | ۲۷/۸۷۲ |
| سن (n = ۵۰) | B = ۷۵۰ | -۳۱/۱۴۴۱ | ۵۶/۵۵۷۱ | ۲۷/۵۹۳۴ |
| | B = ۱۰۰۰ | -۲۹/۲۵۰۳ | ۵۰/۱۹۲۶ | ۲۱/۲۷۴۶ |
| | B = ۲۵۰ | ۱/۳۵۷۸ | ۲/۳۴۳ | ۲/۵۴۵ |
| سن (n = ۱۵۰) | B = ۵۰۰ | ۱/۲۷۵۷ | ۲/۲۱۶ | ۱/۲۲۵۴ |
| | B = ۷۵۰ | ۱/۲۳۷۰ | ۲/۱۹۰ | ۰/۹۹۱۵ |
| | B = ۱۰۰۰ | ۱/۱۸۰۴ | ۲/۱۱۸ | ۰/۸۹۲۵ |
| قطر تومور (n = ۵۰) | B = ۲۵۰ | ۰/۴۶۵۱ | ۱/۰۷۳۳ | ۰/۴۰۶۷ |
| | B = ۵۰۰ | ۰/۴۱۰۹ | ۰/۸۵۹۲ | ۰/۳۹۳۸ |
| | B = ۷۵۰ | ۰/۴۰۴۶ | ۰/۸۸۴۲ | ۰/۳۸۷ |
| قطر تومور (n = ۱۵۰) | B = ۱۰۰۰ | ۰/۳۷۹۲ | ۰/۷۵۷۰ | ۰/۳۵۶۵ |
| | B = ۲۵۰ | -۰/۸۹۶۳ | ۸/۱۲۳ | ۲۸/۱۵۳۲ |
| | B = ۵۰۰ | -۰/۸۰۹۹ | ۷/۹۷۹ | ۲۳/۵۷۲۷ |
| تعداد غدد درگیر (n = ۵۰) | B = ۷۵۰ | -۰/۴۷۴۳ | ۶/۴۲۰ | ۲۱/۵۵۸۱ |
| | B = ۱۰۰۰ | -۰/۲۳۲۳ | ۵/۵۹۶ | ۲۰/۱۲۲ |
| | B = ۲۵۰ | -۰/۱۱۰۵ | ۰/۶۰۷۸ | ۲/۵۲۴۳ |
| تعداد غدد درگیر (n = ۱۵۰) | B = ۵۰۰ | -۰/۰۹۶۲ | ۰/۴۳۶۷ | ۲/۳۸۹۳ |
| | B = ۷۵۰ | -۰/۰۸۷۶ | ۰/۴۲۰۷ | ۲/۳۵۴۹ |
| | B = ۱۰۰۰ | -۰/۰۷۲۱ | ۰/۳۷۹۶ | ۲/۲۲۰۴ |
| تعداد غدد درگیر (n = ۵۰) | B = ۲۵۰ | ۲/۷۱۲۰ | ۵/۱۲۲۷ | ۱۶/۰۵۱۵ |
| | B = ۵۰۰ | ۲/۶۱۵۰ | ۴/۹۰۱۵ | ۶/۸۰۲۰ |
| | B = ۷۵۰ | ۲/۵۳۱۵ | ۴/۸۳۱۵ | ۵/۷۲۱۷ |
| | B = ۱۰۰۰ | ۲/۴۸۱۵ | ۴/۸۴۲۷ | ۴/۵۶۱۵ |

جدول ۵ (ادامه) - برآورد ضرایب، خطا و طول بازه برای واحدهای نمونه گیری شده به روش بوت استرپ

| | | | | |
|---------|--------|--------|----------|-----------------|
| ۱/۱۸۸۹ | ۱/۹۵۸۵ | ۰/۸۱۵۰ | B = ۲۵۰ | |
| ۰/۹۲۲۲ | ۱/۶۹۸۶ | ۰/۷۳۹۱ | B = ۵۰۰ | تعداد غدد درگیر |
| ۰/۸۴۴۸ | ۱/۵۴۱۴ | ۰/۷۱۱۵ | B = ۷۵۰ | (n = ۱۵۰) |
| ۰/۷۵۹۲ | ۱/۳۸۷۴ | ۰/۶۳۳۰ | B = ۱۰۰۰ | |
| ۴۳/۳۸۶ | ۱۵/۵۰۵ | ۸/۳۶۸۳ | B = ۲۵۰ | |
| ۱۶/۴۴۷۹ | ۱۵/۱۴۸ | ۸/۱۹۹۳ | B = ۵۰۰ | مدت بیهوشی |
| ۱۴/۰۴۵۱ | ۱۴/۴۸۸ | ۸/۱۸۹۷ | B = ۷۵۰ | (n = ۵۰) |
| ۸/۳۱۶۲ | ۱۱/۳۹۰ | ۷/۹۰۸۷ | B = ۱۰۰۰ | |
| ۲/۷۴۲۹ | ۴/۰۸۶۰ | ۳/۲۳۵۱ | B = ۲۵۰ | |
| ۲/۶۲۲۵ | ۳/۴۲۳۲ | ۳/۰۶۰۸ | B = ۵۰۰ | مدت بیهوشی |
| ۲/۲۴۲۲ | ۳/۷۶۴۱ | ۳/۰۹۳۴ | B = ۷۵۰ | (n = ۱۵۰) |
| ۲/۱۰۸۱ | ۳/۱۱۳۲ | ۲/۹۷۱۶ | B = ۱۰۰۰ | |
| ۹۲/۹۲۱۹ | ۲۴/۶۲۸ | ۴/۱۸۹۷ | B = ۲۵۰ | متاستاز |
| ۸۳/۵۵۸۴ | ۲۴/۰۴۱ | ۴/۱۲۶۸ | B = ۵۰۰ | (n = ۵۰) |
| ۸۰/۸۹۹۱ | ۲۲/۵۸۳ | ۴/۱۲۰۱ | B = ۷۵۰ | |
| ۶۸/۵۶۶۶ | ۲۰/۸۵۷ | ۳/۱۰۳۳ | B = ۱۰۰۰ | |
| ۵/۵۳۳۱ | ۳/۷۴۹۸ | ۱/۲۷۰۵ | B = ۲۵۰ | |
| ۵/۳۲۷۳ | ۱/۸۲۴۳ | ۱/۲۲۹۸ | B = ۵۰۰ | متاستاز |
| ۵/۱۵۲۲ | ۱/۵۵۵۰ | ۱/۰۹۸۹ | B = ۷۵۰ | (n = ۱۵۰) |
| ۴/۷۴۳۵ | ۱/۳۴۱۴ | ۱/۰۸۷۹ | B = ۱۰۰۰ | |

Efron B. (۲۰۰۰) The Bootstrap and Modern Statistics. *JASA*. ۹۵(۴۵۲):۱۲۹۳-۱۲۹۶.

Hall P. (۱۹۹۲) Bootstrap confidence intervals in nonparametric regression. *Ann. Statist.* ۲۰:۶۹۵-۷۱۱.

Efron B. (۱۹۸۱) Nonparametric standard error and confidence intervals (with discussions). *Canadian J.Statist.* ۹:۱۳۹-۱۷۲.

Efron B. (۱۹۸۷) Better Bootstrap confidence intervals (with discussions) *J.Amer. Statist. Assoc.* ۸۲:۱۷۱-۲۰۰.

منابع :

برونر، سودارت (۱۳۷۱). پرستاری بیماری های زنان و پستان، دکتر مرتضی دلاورخان، پروانه بیشه بان. محقق، محمد علی (۱۳۸۱). مقدمه و مروری بر اپیدمیولوژی سرطان پستان در ایران، تهران، مجموعه مقالات چهاردهمین همایش سالیانه انستیتو کانسر.

Agresti A. (۱۹۹۶) An Introduction to Categorical Data Analysis. John Wiley.

Efron B and Tibshirani R.J. (۱۹۹۳) An introduction to the Bootstrap. NewYork. Chapman and Hall.