Archive of SID

دانشور
پزشکی

ارزیابی قابلیت تشخیص دندان‌پزشکان در اندازه‌گیری شاخص نیاز به درمان ارتدنسی (IOTN) با استفاده از مدل‌بندی توافق

نویسندگان: علیرضا اکبرزاده‌باغبان[1]، دکتر غلامرضا بابایی[*2]، دکتر انوشیروان کاظم نژاد[3] و دکتر مهتاب نوری[4]

۱. دانشجوی دکتری آمار زیستی دانشکده علوم پزشکی دانشگاه تربیت مدرس
۲. دانشیار گروه آمار زیستی دانشکده علوم‌پزشکی‌دانشگاه‌تربیت‌مدرس
۳. استاد گروه آمار زیستی دانشکده علوم‌پزشکی دانشگاه تربیت مدرس
۴. استادیار دانشکده دندان‌پزشکی دانشگاه علوم پزشکی شهید بهشتی

* نویسنده مسئول: Email: ziabar1@yahoo.com

دوماهنامه علمی – پژوهشی
دانشگاه شاهد
سال سیزدهم – شماره ۶۰
دی ۱۳۸۴

تاریخ وصول: ۸۳/۲/۲۶
تاریخ پذیرش: ۸۴/۱/۲۴

چکیده

مقدمه: بهترین روش برای تعیین تطابق بین اندازه‌گیری‌های حاصل از مقیاس‌های رتبه‌ای در آمار، مدل‌بندی ساختار توافق بین آن اندازه‌گیری‌ها است. یکی از موضوعات مهم مرتبط با بحث توافق اندازه‌گیری‌ها از طریق مدل‌بندی، قابلیت تشخیص رده‌های آن مقیاس رتبه‌ای است.

روش بررسی: در این مقاله به کمک سه مدل آماری، قابلیت تشخیص دندان‌پزشکان در اندازه‌گیری شاخص نیاز به درمان ارتدنسی (IOTN) ارزیابی و بررسی می‌شود که دندان‌پزشکان در اختصاص افراد به سطوح مختلف این شاخص چگونه عمل می‌کنند. شاخص مذکور در دو زمان مجزا و به فاصله یک روز، توسط ۱۰ دندان‌پزشک و به‌طور مستقل روی ۲۰ کست استاندارد مطالعه ارتدنسی، با استفاده از خط کش استاندارد *DHC* اندازه‌گیری شد. مدل‌های آماری مورد استفاده در این مقاله، جهت محاسبه قابلیت تشخیص رده‌ها توسط دندان‌پزشکان عبارتند از مدل پارامتر قطری، مدل پیوند یکنواخت و مدل توافق به‌علاوه پیوند یکنواخت.

یافته‌ها: در مورد داده‌های مربوط به ۷ دندان‌پزشک، مدل پارامتر قطری (با حداقل سطح معناداری ۰/۰۹) برازش داشت در حالی‌که برای ۳ مجموعه داده باقیمانده، مدل پیوند یکنواخت (با حداقل سطح معناداری ۰/۴۰) مناسب بود. مدل توافق به‌علاوه پیوند یکنواخت در هیچ مورد برازش مناسب نداشت.

بحث: نتایج این مدل‌ها نشان داد که این دندان‌پزشکان با میانگین ۰/۹۸۱۲ در تشخیص و تفکیک سطوح مختلف شاخص دارای قابلیت بالا هستند. کم‌ترین قابلیت تشخیص برابر ۰/۹۴۱۵ بود که ضریب توافق کاپا برای دوبار اندازه‌گیری این دندان‌پزشک ۰/۴۲ محاسبه شد.

واژه‌های کلیدی: قابلیت تشخیص رده، داده رتبه‌ای، مدل پیوند، شاخص نیاز به درمان ارتدنسی (IOTN)

www.SID.ir

Archive of SID

مقدمه

فرض کنیم آزماینده‌ای در دو زمان متفاوت، هر کدام از افراد یک نمونه را طبق یک مقیاس رتبه‌ای طبقه‌بندی کند، به‌طوری که نتیجه طبقه‌بندی در زمان اول روی نتیجه طبقه‌بندی در زمان دوم هیچ تأثیری نداشته باشد. می‌توان نتیجه توأم این دو بار تشخیص را در یک جدول متقاطع مربع آورد. در چنین جدولی، سه موضوع را می‌توان بررسی کرد:

۱. قابلیت تشخیص رده‌ها: زمانی که رده‌های نرخ‌گذاری به‌صورت رتبه‌ای هستند، یکی از موضوعات مهم مرتبط با مبحث تطابق اندازه‌گیری‌ها، قابلیت تشخیص رده‌ها است که این موضوع عمدتاً نسبت به رده‌های مجاور مطرح می‌شود. در این مورد باید معین کنیم که آیا رده‌های مقیاس رتبه‌ای برای آزماینده قابل تفکیک هستند یا خیر [۱].

۲. همگنی یا ناهمگنی توزیع‌های حاشیه‌ای نرخ‌گذاری‌های آزماینده در دو زمان: معمولاً برای متغیرهای رتبه‌ای می‌خواهیم بدانیم آیا گروه‌بندی یک آزماینده در زمان دوم نسبت به زمان اول بالاتر (پایین‌تر) است یا این‌که این دو مجموعه از تشخیص‌ها دقیقاً بر هم منطبقند [۲].

۳. توافق نرخ گذاری‌های آزماینده در دو زمان اندازه‌گیری: این موضوع در واقع فراوانی روی قطر اصلی در توزیع توأم نرخ گذاری‌ها است [۳و۴]. در این مقاله ما فقط به موضوع اول می‌پردازیم.

اکثر مقیاس‌های رتبه‌ای، ذهنی هستند. به‌عنوان مثال، تشخیص این‌که وضعیت بیماری یا نیاز به درمان یک فرد در چه سطحی از یک مقیاس رتبه‌ای قرار دارد، غالباً مشکل است و حتی آزماینده‌های خیلی مجرب، در دو بار اندازه‌گیری روی افراد مشترک، نظرهای متفاوتی ارائه می‌کنند. یکی از دلایل این اتفاق می‌تواند تفاوت معنای رده‌ها برای آزماینده‌ها باشد [۳].

معمولاً برای بررسی توافق بین نرخ‌گذاری‌های یک مقیاس رتبه‌ای در دو زمان یا ارزیابی اعتبار این داده‌ها از ضریب توافق کاپا استفاده می‌شود؛ اما استفاده از این ضرایب از بسیاری جهات خالی از اشکال نیست و توصیه می‌شود به‌جای این کار، ساختار توافق داده‌های حاصل را مدل‌بندی کنیم [۱،۵و۶]. لذا در این مقاله قصد داریم با استفاده از مدل‌های آماری مندرج در پیوست ۱، ساختار مذکور را مدل‌بندی کرده، به کمک آن به بررسی قابلیت تشخیص دندان‌پزشکان در تشخیص و تفکیک سطوح مختلف شاخص رتبه‌ای نیاز به درمان ارتدنسی بپردازیم [۷و۸].

داده‌های این مقاله قسمتی از داده‌های مربوط به «طرح ارزیابی روایی و پایایی شاخص نیاز به درمان ارتدنسی» (Index of Orthodontic Treatment Need: IOTN) است که در تابستان سال ۱۳۸۲ در مرکز تحقیقات علوم دندان‌پزشکی انجام شد. در یک مرحله از این طرح، هر کدام از ۱۰ دندان‌پزشک که به‌عنوان آزماینده حضور داشتند، این شاخص را در دو زمان متفاوت و به فاصله یک روز، به‌طور مستقل روی ۲۰ کست استاندارد مطالعه ارتدنسی با استفاده از خط کش استاندارد *DHC* اندازه‌گیری کردند. این نمونه‌ها به‌صورت تصادفی از قسمت بایگانی بخش ارتدنسی دانشکده دندان‌پزشکی دانشگاه علوم پزشکی شهید بهشتی انتخاب شدند. همه تمهیدات لازم اندیشیده شد تا ارزیابی اول این دندان‌پزشکان روی ارزیابی دوم آن‌ها هیچ تأثیری نداشته باشد. به‌علاوه چون مطالعه روی کست گچی انجام شد، فاصله زمانی یک روزه نمی‌تواند اثری روی نمونه‌ها داشته باشد. هدف طرح این بود که اگر اندازه‌گیری‌های دندان‌پزشکان معتبر باشد، این افراد شاخص مذکور را برای نوجوانان دبیرستانی شهر تهران اندازه‌گیری کنند. این شاخص به کمک ۳۰ مشخصه و به‌صورت بیش‌ترین امتیاز (maximum score) اندازه‌گیری می‌شود؛ بدین‌صورت که برای هر فرد، بیش‌ترین رتبه از بین رتبه‌هایی که به هر کدام از این ۳۰ مشخصه داده می‌شود، به‌عنوان شاخص نیاز به درمان ارتدنسی

www.SID.ir

Archive of SID

محسوب می‌گردد. این شاخص رتبه‌ای بوده، دارای ۵ سطح عدم نیاز به درمان، نیاز جزئی به درمان، نیاز متوسط به درمان، نیاز شدید به درمان و نیاز بسیار شدید به درمان است و در بسیاری از کشورهای اروپایی و آسیایی – که در آن‌ها بیمه درمان ارتدنسی مطرح است– مورد استفاده قرار می‌گیرد. لازم به ذکر است که این شاخص عمدتاً برای اندازه‌گیری نیاز به درمان ارتدنسی در سطح جامعه کاربرد دارد [۷]. از نظر نحوه اندازه‌گیری می‌توان این شاخص را با شاخص وضعیت پریودونتال (CPI) مقایسه کرد؛ زیرا در ارزیابی وضعیت پریودونتال نیز بیش‌ترین رتبه‌ای که در شش نقطه از لثه هر فرد حاصل می‌شود، به‌عنوان شاخص CPI وی محسوب می‌شود. فرم ثبت مشخصه‌ها و اندازه‌گیری نهایی «شاخص رتبه‌ای نیاز به درمان ارتدنسی» در پیوست ۲ آمده است.

مواد و روش‌ها

برای ارزیابی قابلیت تشخیص دندان‌پزشکان در تفکیک سطوح مختلف شاخص رتبه‌ای نیاز به درمان ارتدنسی (IOTN) ازمدل‌های آماری توافق به‌علاوه پیوند یکنواخت [۹] T پارامتر قطری [۶] و پیوند یکنواخت [۱۰و۱۱]، که به‌طور کامل در پیوست ۱ شرح داده شده، استفاده شد. هر سه مدل مذکور حالت خاصی از مدل‌های لگ خطی بوده، مدل‌های پارامتر قطری و پیوند یکنواخت نیزحالات خاصی از مدل توافق به‌علاوه پیوند یکنواختند [۳]. جهت تکمیل شدن بحث و مقایسه کامل اندازه‌گیری‌های انجام شده شاخص در دو زمان مختلف توسط دندان‌پزشکان، ضرایب توافق کاپا نیز محاسبه شده اند [۴و ۱۲].

تعداد ۲۰ کست استاندارد مطالعه ارتدنسی که به‌صورت تصادفی از قسمت بایگانی بخش ارتدنسی دانشکده دندان‌پزشکی دانشگاه علوم پزشکی شهید بهشتی انتخاب شده بودند، دو بار و با فاصله یک روز توسط هر کدام از دندان‌پزشکان به‌وسیله خط کش استاندارد *DHC* مورد ارزیابی قرار گرفتند. این تعداد به دلیل محدودیت‌های عملی اتخاذ گردید؛ ولی با توجه به این‌که حجم نمونه در مطالعات مربوط به ارزیابی روایی و پایایی کم است، می‌تواند نمونه قابل قبولی باشد [۱۳]. از آن‌جا که سطوح ۱،۲و۳ توسط دندان‌پزشکان به‌ندرت تشخیص داده شده بود، این سه سطح با هم ترکیب شد و نهایتاً روی داده‌های مربوط به ۱۰ دندان‌پزشک در یک مقیاس رتبه‌ای سه حالته (با سطوح نیاز متوسط و پایین‌تر به درمان، نیاز شدید به درمان، نیاز خیلی شدید به درمان) کار شد. نهایتاً قابلیت تشخیص دندان‌پزشکان برآورد و با هم مقایسه گردید.

در این راستا، بعد از محاسبه ضریب کاپا برای هر کدام از۱۰ جدول تشکیل شده از جدول‌بندی متقاطع این ۱۰ دندان‌پزشک در دو زمان، با توجه به طبیعت رتبه‌ای بودن مقیاس شاخص جدید، ابتدا مدل توافق به‌علاوه پیوند یکنواخت، سپس مدل پیوند یکنواخت و سرانجام مدل پارامتر قطری مورد نظر قرار گرفتند تا بهترین مدل آماری ممکن انتخاب شود. ملاک مقایسه مدل‌ها و انتخاب آن‌ها، مقادیر آماره نیکویی برازش (Goodness of fit statistics) و سطح معناداری این آماره‌ها بود؛ بدین‌معنا که بعد از برازش مدل‌ها، آن‌هایی مورد نظر قرار می‌گرفتند که سطح معناداری آماره نیکویی برازش آن‌ها حداقل ۰/۰۵ بود. به‌علاوه در صورتی که برای بعضی از این جداول بیش از یک مدل مناسب بود، برای انتخاب بهترین مدل، آن مدلی مورد قبول قرار می‌گرفت که مقدار آماره نیکویی برازش کم‌تر و در نتیجه سطح معناداری بالاتری داشت [۳].

برای استخراج جداول دو بعدی و محاسبه ضرایب کاپا از نرم‌افزار SPSS ۹ و برای برازش مدل‌های مورد نظر روی داده‌های جداول مذکور و برآورد پارامترهای این مدل‌ها – با روش حداکثر درستنمایی – از نرم‌افزار SAS ۸ استفاده کردیم. قابلیت تشخیص را نیز به کمک فرمول نویسی در نرم‌افزار اکس ۲۰۰۰ به‌دست آوردیم.

www.SID.ir

Archive of SID

قابلیت تشخیص رده

قابلیت تشخیص در حالت کلی برای رده‌های i و j و به کمک نسبت بخت (odds ratio) زیر تعریف می‌شود [۱،۳و۱۴] که در آن m_{ij} فراوانی مورد انتظار سلول موجود در سطر i و ستون j جدول‌بندی متقاطع حاصل از نرخ‌گذاری در دو زمان است:

$$\tau_{ij} = (m_{ii}m_{jj})/(m_{ij}m_{ji})$$

اما قابلیت تشخیص رده‌های مجاور اهمیت بیش‌تری دارد. لذا ما نیز در این مقاله این موضوع را برای رده‌های مجاور هم i و $j=i+1$ در نظر گرفته، به کمک رابطه $\gamma_{i,i+1} = 1-(\tau_{i,i+1})^{-1}$ برآورد می‌کنیم [۱۴و۱۵].

برای مدل توافق به‌علاوه پیوند یکنواخت، با فرض این‌که امتیاز هر رده از مقیاس رتبه‌ای برابر رتبه (شماره) آن رده باشد ($u_i = i$)، قابلیت تشخیص رده‌های i و $i+1$ عبارت از $\gamma_{i,i+1} = 1-\exp(-\beta-2\delta)$ خواهد بود که در آن β معرف ضریب جمله پیوند یکنواخت و δ معرف ضریب جمله پارامتر قطری در مدل است. مدل پارامتر قطری، حالت خاصی از مدل توافق به‌علاوه پیوند یکنواخت به ازای $\beta = 0$ است. بنابراین قابلیت تشخیص برای این مدل برابر است با $1-\exp(-2\delta)$. مدل پیوند یکنواخت نیز حالت خاصی از مدل توافق به‌علاوه پیوند یکنواخت به ازای $\delta = 0$ است. پس قابلیت تشخیص برای این مدل برابراست با $1-\exp(-\beta)$.

با این تعریف، قابلیت تشخیص رده حداکثر برابر ۱ به‌دست خواهد آمد؛ ولی چون پارامترهای β و δ غالباً درعمل اعداد مثبتی برآورد می‌گردند، لذا عموماً خواهیم داشت $0 \le \gamma_{i,i+1} \le 1$. در این صورت هرقدر این پارامتر به عدد ۱ نزدیک‌تر باشد، مطلوب‌تر بوده، قابلیت تشخیص آزماینده‌ها بیش‌تر خواهد بود و هرچه این مقدار به صفر نزدیک‌تر باشد، قابلیت تشخیص آزماینده‌ها کم‌تر است. از این دیدگاه می‌توان قابلیت تشخیص را با پارامتر ضریب تعیین (R^2) در مدل‌های رگرسیون مقایسه کرد. برای ۱۰ جدولی که ما در این مقاله ارائه کرده‌ایم نیز چنین وضعیتی دیده شد.

طبق تعاریف موجود، قابلیت تشخیص رده برای آن دسته ازجداول دو بعدی به‌دست می‌آید که یک بعد آن را نرخ‌گذاری آزماینده اول و بعد دیگر آن را نرخ‌گذاری آزماینده دوم تشکیل می‌دهد. در چنین جداولی می‌توان قابلیت تشخیص رده را برآورد کرد؛ ولی نمی‌توان آن را به آزماینده خاصی نسبت داد، مگر این‌که هر دو آزماینده استاندارد طلایی (gold standard) باشند که در این صورت، بدون نگرانی از این‌که قابلیت تشخیص رده را به کدام یک نسبت دهیم، آن را برآورد می‌کنیم. در این مقاله تأکید ما بر قابلیت تشخیص دندان‌پزشکان در تفکیک رده‌های مجاور شاخص نیاز به درمان ارتدنسی است؛ زیرا ما این مقدار را برای جداولی برآورد کرده‌ایم که در آن‌ها هر دو بعد جدول توسط یک دندان‌پزشک حاصل گردیده است.

همچنین تحت شرایطی، بعضی از رده‌ها غیر قابل تشخیص خواهند بود [۱،۳و۱۴]. به‌عنوان مثال در صورتی که نتیجه نرخ‌گذاری‌های دو زمان یک آزماینده از نظر آماری مستقل باشد، هیچ‌کدام از رده‌ها قابل تشخیص نخواهند بود. البته چنین وضعیتی هرگز در بررسی داده‌های توافق آزماینده‌ها اتفاق نمی‌افتد؛ زیرا قطعاً آزماینده‌هایی که برای اندازه‌گیری انتخاب می‌شوند، دارای تخصص و تجربه لازم هستند و لذا نرخ‌گذاری مجزای آن‌ها در دو زمان هرگز از نظر آماری مستقل نخواهد بود. لازم به ذکر است که چنانچه در مدل توافق به‌علاوه پیوند یکنواخت هر دو پارامتر β و δ برابر صفر باشند – که در این صورت این مدل به مدل استقلال تبدیل می‌شود- هیچکدام از رده‌ها قابل تشخیص نخواهند بود.

یافته‌ها

قبلاً نیز گفته شد که بنا برپاره‌ای ملاحظات تئوری و عملی، در این مقاله سطوح ۱، ۲و۳ شاخص رتبه‌ای ۵ حالته «شاخص رتبه‌ای نیاز به درمان ارتدنسی» با هم ترکیب شدند و سرانجام با داده‌های مربوط به ۱۰ دندان‌پزشک برای یک مقیاس رتبه‌ای ۳ حالته در دو

www.SID.ir

Archive of SID

جدول ۱. ضرایب کاپا، بهترین مدل‌های برازش شده با جزئیات مربوطه و قابلیت تشخیص برای ۱۰ جدول ۳×۳ حاصل از نرخ‌گذاری هرکدام از دندان‌پزشکان در دو زمان مختلف

شماره ردیف (ستون ۱)	مقدار کاپا (ستون۲)	بهترین مدل برازش شده (ستون ۳)	مقدار آماره نیکویی برازش مدل (ستون٤)	مقدار احتمال مربوط به برازش مدل (ستون۵)	مقدار برآورد شده پارامتر (ستون۶)	مقدار آماره مرتبط با پارامتر مدل (ستون ۷)	مقدار احتمال مرتبط با پارامتر مدل (ستون۸)	قابلیت تشخیص دندان‌پزشکان (ستون ۹)
۱	۰/۴۸	پیوندیکنواخت	۰/۳۰	۰/۹۶	۳/۶۱	۶/۳۴	۰/۰۱۱۸	۰/۹۷۲۹
۲	۰/۶۱	″	۲/۹۶	۰/۴۰	۳/۶۰	۸/۰۵	۰/۰۰۴۶	۰/۹۷۲۶
۳	۰/۴۲	″	۱/۴۶	۰/۶۹	۲/۸۴	۵/۶۰	۰/۰۱۸۰	۰/۹۴۱۵
۴	۰/۵۹	پارامتر قطری	۲/۷۵	۰/۴۳	۲/۴۵	۵/۷۶	۰/۰۱۶۴	۰/۹۹۲۵
۵	۰/۶۷	″	۴/۵۹	۰/۲۰	۲/۶۹	۷/۹۵	۰/۰۰۴۸	۰/۹۹۵۳
۶	۰/۴۰	″	۳/۳۵	۰/۳۴	۱/۷۳	۳/۸۹	۰/۰۴۸۷	۰/۹۶۸۵
۷	۰/۶۰	″	۶/۴۳	۰/۰۹	۲/۱۲	۸/۹۴	۰/۰۰۲۸	۰/۹۸۵۶
۸	۰/۶۱	″	۴/۱۱	۰/۲۵	۲/۵۴	۶/۸۲	۰/۰۰۹۰	۰/۹۹۳۸
۹	۰/۷۲	″	۵/۸۵	۰/۱۲	۲/۳۶	۱۱/۲۸	۰/۰۰۰۸	۰/۹۹۱۱
۱۰	۰/۷۸	″	۲/۱۲	۰/۵۵	۳/۰۴	۷/۹۸	۰/۰۰۴۷	۰/۹۹۷۷

زمان و در قالب ۱۰ جدول ۳×۳ کار شد. مدل توافق به‌علاوه پیوند یکنواخت با داده‌های هیچ‌کدام ازاین ۱۰ جدول مطابقت نداشت. در ۳ مورد مدل پیوند یکنواخت و در ۷ مورد دیگر مدل پارامتر قطری بهترین برازش را نشان دادند. خلاصه‌ای از کارهای انجام شده روی این مجموعه داده در جدول ۱ نشان داده شده است.

لازم به ذکر است که تعدادی از سلول‌های مربوط به هرکدام از ۱۰ جدول مذکور خالی (دارای فراوانی مشاهده شده صفر) بود. این وضعیت عموماً برای چنین مجموعه داده‌هایی اتفاق می‌افتد. زمانی که آزمون‌های نیکویی برازش برای چنین داده‌هایی به کار برده می‌شود، توزیع آماره نسبت درستنمایی (G^2) توسط توزیع کای- دو به خوبی تقریب زده نمی‌شود؛ اما این آماره برای مقایسه مدل‌ها خوب عمل می‌کند [۳،۱۵،۱۶و۱۷].

ستون۲ جدول فوق به کمک نرم‌افزار SPSS، ستون‌های۳ تا ۸ توسط نرم‌افزار SAS و ستون ۹ به کمک فرمول‌نویسی در نرم‌افزار اکسل محاسبه گردیده‌اند. تعداد درجات آزادی در بررسی مناسبت مدل براساس آماره نیکویی برازش برای هر دو مدل پیوند یکنواخت و پارامتر قطری در حالت کلی برابر است با $df = (r-1)^2 - 1$ که در آن r تعداد سطوح مقیاس رتبه‌ای است. در این تحقیق، مقیاس رتبه‌ای تبدیل شده نهایی دارای ۳ سطح است و لذا درجه آزادی برای هر ۱۰ مدل موجود در جدول نیز برابر ۳ است [۳].

طبق بررسی‌های انجام شده در ۱۰ جدول مذکور هیچ‌کدام از سطوح «شاخص نیاز به درمان ارتدنسی» برای هیچ‌کدام از دندان‌پزشکان غیر قابل تشخیص نبود، زیرا هیچ‌کدام از نسبت‌های بخت برابر ۱ نبود که یکی از شروط مربوط به غیرقابل تشخیص بودن رده‌ها است [۱و۱۴] و لذا مبادرت به محاسبه قابلیت تشخیص آن‌ها کردیم. برآورد قابلیت تشخیص در مدل پیوند یکنواخت از طریق $1-\exp(-\hat{\beta})$ و در مدل پارامتر قطری از طریق $1-\exp(-2\hat{\delta})$ به‌دست می‌آید که در آن $\hat{\beta}$ و $\hat{\delta}$ به ترتیب مقادیر برآورد شده پارامترهای β و δ به روش حد اکثر درستنمایی هستند.

www.SID.ir

Archive of SID

بحث و نتیجه‌گیری

همان‌گونه که از ستون ۲ جدول ۱ ملاحظه می‌شود، برای هیچ‌کدام از دندان‌پزشکان توافق کامل و یا تقریباً کامل (به‌عنوان مثال با ضریب توافق حداقل ۰/۸) بین نرخ‌گذاری‌ها در دو زمان دیده نمی‌شود. با توجه به این‌که رده‌ها توسط دندان‌پزشکان قابل تشخیص بودند، قابلیت تشخیص آن‌ها را محاسبه کردیم. طبق ستون ۹ جدول، همه دندان‌پزشکان دارای قابلیت بالایی در تفکیک و تشخیص سطوح مختلف شاخص نیاز به درمان ارتدنسی هستند، به‌طوری که دندان‌پزشک شماره ۱۰ با مقدار برآورد شده تقریباً برابر یک دارای بیش‌ترین قابلیت تشخیص و دندان‌پزشک شماره ۳ با مقدار برآورد شده ۰/۹۴۱۵ دارای کم‌ترین قابلیت تشخیص است. برای این دو دندان‌پزشک، مقدار ضریب توافق برآورد شده کاپا به ترتیب برابرند با ۰/۷۸ و ۰/۴۲. اگرچه دندان‌پزشک شماره ۳ در مقایسه با سایرین دارای قابلیت پایین‌تری در تشخیص سطوح مجاور شاخص است، ولی از آن‌جا که این مقدار به‌طور قابل ملاحظه به عدد یک نزدیک است، می‌توان گفت که در مجموع، تفکیک سطوح این شاخص رتبه‌ای برای دندان‌پزشکان دشوار نیست. استنباط اخیر از روی میانگینِ رویِ هم قابلیت تشخیص این ۱۰ دندان‌پزشک که برابر ۰/۹۸۱۲ است نیز تایید می‌شود.

دندان‌پزشکانی که برای آن‌ها مدل پارامتر قطری مناسب بود، با میانگین ۰/۹۸۹۲ دارای قابلیت تشخیص بالاتری نسبت به سایر دندان‌پزشکان بودند. این دندان‌پزشکان – که مدل پیوند یکنواخت برای مدل‌بندی ارزیابی آن‌ها مناسب بود – دارای میانگین قابلیت تشخیص ۰/۹۶۲۳ بودند.

اگر ستون ۲ و ۹ جدول ۱ را با هم مقایسه کنیم می‌بینیم که بجز در بعضی موارد، با افزایش ضریب توافق، قابلیت تشخیص نیز اضافه شده است. در توضیح مواردی که در آن‌ها این روند دیده نمی‌شود نیز باید گفت ضریب توافق، سطح کلی توافق مشاهده شده را با سطح کلی توافق مورد انتظار مقایسه می‌کند و به تفاوت بین الگوهای توافق مشاهده شده و مورد انتظار غیرحساس است [۶]، در حالی که چون قابلیت تشخیص رده‌ها از طریق مدل‌های آماری برآورد می‌گردد، الگوی مذکور را در نظر گرفته، قابلیت دندان‌پزشکان را در تشخیص و تفکیک سطوح مختلف مقیاس بیان می‌کند و لذا دارای اعتبار بیش‌تر است.

پیشنهاد

در بررسی قابلیت تشخیص رده‌ها توسط آزماینده‌ها در یک مقیاس رتبه‌ای، ابتدا ضرایب توافق کاپا را برای جدول متقاطع حاصل از نرخ‌گذاری آزماینده‌ها در دو زمان محاسبه می‌کنیم. در صورت وجود توافق کامل (۱= کاپا)، بدون انجام مدل‌بندی اظهار می‌داریم که رده‌ها به خوبی توسط آزماینده‌ها قابل تشخیص هستند. در غیر این صورت، برای رسیدن به قابلیت تشخیص رده‌ها توسط آزماینده‌ها، می‌توان مدل‌های پارامتر قطری، پیوند یکنواخت و توافق به‌علاوه پیوند یکنواخت را مورد نظر قرار داد. بعد از رسیدن به بهترین مدل مناسب و برآورد پارامترهای آن، قابلیت تشخیص رده‌ها به آسانی قابل برآورد خواهد بود. توصیه می‌شود در انجام این کار، اعم از محاسبه ضریب توافق و همچنین مدل‌بندی ساختار توافق برای محاسبه قابلیت تشخیص رده‌ها، مخصوصاً اگر تعداد سطوح متغیر زیاد باشد، تعداد نمونه در حد قابل قبول باشد.

به‌علاوه با توجه به این‌که تنها یک پارامتر(δ) برای سلول‌های روی قطر اصلی در مدل پارامتر قطری و تنها یک پارامتر (β) برای رابطه بین سطوح متغیر رتبه‌ای در مدل پیوند یکنواخت در نظر گرفته می‌شود، قابلیت تشخیص برای همه رده‌های مجاور یکسان برآورد می‌گردد؛ در حالی که اگر می‌توانستیم پارامترهای مختلفی در این مدل‌ها در نظر بگیریم، علاوه بر این‌که ممکن بود به مدل‌های قوی تری برسیم که به داده‌ها برازش مناسب تری دارند، می‌توانستیم برآوردهای مختلف نزدیک‌تر به واقعیتی از قابلیت تشخیص

www.SID.ir

Archive of SID

رده‌های مجاور متفاوت داشته باشیم. به‌عنوان مثال، قابلیت تشخیص رده‌های ۱و۲ می‌توانست متفاوت از قابلیت تشخیص رده‌های ۲و۳ برآورد گردد. نویسندگان این مقاله در نظر دارند تا در صورت امکان در گام‌های بعد به این مقوله بپردازند.

منابع

1. Perkins SM, Becker MP. Assessing rater agreement using marginal association models. *Statistics In Medicine* 2002; 21:1743-1760.
2. Koch GG, Landis JR, Freeman JL, Freeman DH, Lehnen RG. A general methodology for the analysis of experiments with repeated measurement of categorical data. *Biometrics* 1977; 33:133-158.
3. Agresti A. A model for agreement between ratings on an ordinal scale. *Biometrics* 1988; 44:539-548.
4. Kraemer HC, Periakoil VS, Noda A. Tutorial in biostatistics, kappa coefficients in medical research. *Stat Med* 2002; 21:2109-2129.
5. Kraemer HC, Periyakoil VS, Noda A. Tutorial in biostatistics, kappa coefficient in medical research. *Stat Med* 2002; 21:2109-2129.
6. Tanner MA, Young MA. Modelling agreement among raters. *JASA* 1985; 80:175-180.
7. Brook PH, Shaw WC. The development of an index for orthodontic treatment priority. *European Journal of Orthodontic* 1989; 11:309-320.
8. Richmond S, Shaw WC, O'Brien KD et al. The relationship between the Index of Orthodontic Treatment Need and the consensus opinion of a panel of 74 dentists. *BDJ* 1995; 178:370-374.
9. Agresti A. Categorical Data Analysis. Wiley: New York, 2002.
10. Goodman LA. Simple models for the analysis of association in cross-classifications having ordered categories. *JASA* 1979; 74:537-552.
11. May SM. Modelling observer agreement – an alternative to kappa. *J Clin Epidemiol* 1994; 44:1315-1324.
12. Cohen J. A coefficient of agreement for nominal scales. *Education and Psychological Measures* 1960; 20:37-46.
13. Fleiss JL. The Design and Analysis of Clinical Experiments. Wiley: New York, 1999.
14. Darroch JN, McCloud PI. Category distinguishability and observer agreement. *Australian Journal of Statistics* 1986; 28:371-388.
15. Becker MP, Agresti A. Log-linear modelling of pairwise interobserver agreement on a categorical scale. *Stat Med* 1992; 11:101-114.
16. Haberman SJ. Loglinear models and frequency tables with small expected counts. *Annals of Statistics* 1977; 5:1148-1169.
17. Agresti A, Yang M. An empirical investigation of some effects of sparseness in contingency tables. *Computational Statistics and Data Analysis* 1986; 5:9-21.

www.SID.ir

Archive of SID

پیوست۱ - مدل‌ها

برای آشنایی با مدل‌های مورد استفاده در این مقاله، شرح مختصری در خصوص آن‌ها ارائه می‌شود. از آن‌جا که همه مدل‌های مذکور از مدل استقلال نتیجه می‌شوند، بحث را از این مدل آغاز می‌کنیم.

۱. مدل استقلال (مدل پیوند پوچ)

فرض کنیم هر کدام از n فرد نمونه، توسط دو آزماینده A و B به‌طورمستقل به یکی از r گروه جدا از هم در یک مقیاس ترتیبی نسبت داده می‌شود. اگر نتیجه این دو نرخ‌گذاری را در یک جدول مربع متقاطع بیاوریم و $m_{ij} = n\pi_{ij}$ را معرف فراوانی مورد انتظار نرخ‌گذاری i توسط آزماینده اول و نرخ‌گذاری j توسط آزماینده دوم در نظر بگیریم، ساده‌ترین مدل برای بیان ساختار داده‌های موجود در چنین جدولی، مدل استقلال است که فرمول ریاضی آن عبارت است از:

$$\log m_{ij} = \mu + \lambda_i^A + \lambda_j^B \quad i=1,2,\ldots,r \; j=1,2,\ldots,r \qquad (۱$$

که در آن λ_i^A و λ_j^B به ترتیب اثر نرخ‌گذاری آزماینده‌های A و B هستند. این مدل معمولاً به جداول توافقی از نوع مذکور برازش ندارد؛ زیرا با توجه به تجربه آزماینده‌ها، انتظار می‌رود نرخ‌گذاری‌های آن‌ها روی تعداد مشترکی از افراد به هم وابسته باشند. در صورتی که این مدل به داده‌ها برازش مناسبی داشته باشد نتیجه می‌گیریم که هیچ توافقی ورای آنچه به تصادف انتظار داریم بین دو آزماینده وجود ندارد که چنین چیزی عملاً محال است [۸] . برای مقایسه وضعیت نرخ‌گذاری آزماینده‌ها براساس مدل‌های مذکور، از نسبت بخت موضعی (local odds ratio) استفاده می‌شود که به‌صورت زیر تعریف می‌شود:

$$\theta_{ij} = \frac{m_{ij}m_{i+1,j+1}}{m_{i,j+1}m_{i+1,j}} .$$

این پارامتر برای مدل فوق برابر یک است و لذا لگاریتم آن همواره صفر است. تعداد درجات آزادی برای بررسی مناسبت این مدل که گودمن آن را مدل پیوند پوچ (null association model) نامیده $df=(r-1)^2$ است [۳].

۲. مدل پارامتر قطری

این مدل ابتدا در سال ۱۹۸۵ توسط تانر و یانگ و برای مدل‌بندی داده‌های با مقیاس اسمی ارائه گردید [۶] . شکل کلی این مدل عبارت است از:

$$\log m_{ij} = \mu + \lambda_i^A + \lambda_j^B + \delta(i,j) \qquad (۲$$

که در آن

$$\delta(i,j) = \begin{cases} \delta & i=j \\ 0 & i \neq j \end{cases} .$$

پارامتر δ در این مدل که برای سلول‌های روی قطر اصلی در نظر گرفته شده معرف توافقی است که ورای شانس انتظار داریم. لگاریتم نسبت بخت موضعی برای این مدل چنین است:

$$\log\theta_{ij} = \begin{cases} 2\delta & i=j \\ -\delta & |i-j|=1 \\ 0 & |i-j|>1 \end{cases}$$

طبیعی است که در حالت $|i-j|>1$ هیچ‌کدام از چهار سلولی که در محاسبه نسبت بخت موضعی ظاهر می‌شوند [یعنی سلول‌های (i,j)، $(i+1,j+1)$، $(i+1,j)$، $(i,j+1)$] روی قطر اصلی نیست و لذا برای این حالت فراوانی مورد انتظار از طریق رابطه ۱ به‌دست می‌آید. برای حالت $|i-j|=1$ تنها یکی از چهار سلول روی قطر اصلی و برای حالت $i=j$ دو تا از سلول‌ها روی قطر اصلی هستند.

www.SID.ir

Archive of SID

این مدل تنها یک پارامتر بیش‌تر از مدل استقلال دارد و لذا تعداد درجات آزادی برای بررسی مناسبت آن می‌شود $df=(r-1)^2-1$. اگر فرض کنیم آزماینده‌ها توافق ندارند، مدل پارامتر قطری می‌گوید که نرخ‌گذاری آن‌ها مستقلند. این رفتارممکن است در مقیاس نرخ‌گذاری ترتیبی دیده نشود، زیرا بین این نرخ‌گذاری‌ها رابطه مثبت متوسط تا قوی دیده می‌شود [۳]. لذا برای مقیاس‌های ترتیبی مدل دیگری را در نظر می‌گیریم.

۳. مدل پیوند یکنواخت

برای طبقه‌بندی متقاطع داده‌های ترتیبی حاصل از نرخ‌گذاری دو آزماینده، مدل پیوند خط به خط (linear by linear association) الگوی ساده وغالباً مناسبی ارائه می‌کند [۳].این مدل به‌صورت:

$$\log m_{ij} = \mu + \lambda_i^A + \lambda_j^B + \beta u_i u_j \qquad (۳$$

است که در آن $u_1 < u_2 < \ldots < u_r$ امتیازات ثابتی هستند که به گروه‌های ترتیبی پاسخ نسبت داده می‌شوند و $\beta u_i u_j$ مبین رابطه موجود بین سطوح نرخ‌گذاری است. این مدل نیز نسبت به مدل استقلال تنها یک پارامتر بیش‌تر دارد و لذا تعداد درجات آزادی برای بررسی مناسبت آن عبارت است از: $df=(r-1)^2-1$. لگاریتم نسبت بخت موضعی برای این مدل چنین است:

$$\log\theta_{ij} = \beta(u_{i+1} - u_i)(u_{j+1} - u_j)$$

با فرض $u_i=i$، مدل ۳ به مدل پیوند یکنواخت (model uniform association) که توسط گودمن [۱۰] معرفی گردید، تبدیل می‌شود. واضح است که در این صورت داریم:

$$\log\theta_{ij} = \beta$$

٤. مدل توافق به‌علاوه پیوند یکنواخت

اگرچه برای توصیف رابطه بین گروه‌بندی‌های ترتیبی، مدل پیوند یکنواخت مناسب است، ولی چون این مدل به رفتار ویژه سلول‌های روی قطر اصلی توجهی نمی‌کند، انتخاب خوبی برای مدل‌بندی توافق نیست. مدل زیر با در نظر گرفتن اثر سلول‌های روی قطر اصلی، ورای آنچه مدل پیوند ارائه می‌کند این موضوع را پوشش می‌دهد. به عبارت دیگر، کل توافق موجود به سه مؤلفه زیر افراز می‌گردد: ۱) توافق شانسی (آن میزان از توافقی که در صورت استقلال آزماینده‌ها نیز انتظار داریم)، ۲) توافق به دلیل رابطه مبنایی که بین نرخ‌گذاری‌های ترتیبی وجود دارد، ۳) توافق مازاد روی قطر اصلی که ورای شانس و رابطه مبنا است.

این مدل که قادر به تفکیک سه مؤلفه مذکور است و به آن مدل توافق به‌علاوه پیوند یکنواخت (agreement model plus uniform association) اطلاق می‌گردد به‌صورت زیر است [۳]:

$$\log m_{ij} = \mu + \lambda_i^A + \lambda_j^B + \beta u_i u_j + \delta(i,j) \qquad (۴$$

که در آن $\delta(i,j)$ و $\beta u_i u_j$ جملاتی هستند که به ترتیب در مدل‌های پارامتر قطری و پیوند یکنواخت معرفی شده‌اند. همان‌طور که ملاحظه می‌شود این مدل دو پارامتر بیش‌تر از مدل استقلال دارد و لذا تعداد درجات آزادی برای بررسی مناسبت آن چنین است: $df=(r-1)^2-2$

حالات خاص: ۱) $\delta = 0$ که در این صورت این مدل به ازای امتیازات ثابت $u_i = i$ به مدل پیوند یکنواخت ۳ تبدیل می‌شود. ۲) $\beta = 0$ که در این صورت این مدل به مدل پارامتر قطری ۲ تبدیل می‌شود. ۳) $\beta = \delta = 0$ که در این صورت این مدل به مدل استقلال ۱ تبدیل می‌شود. لگاریتم نسبت بخت موضعی برای این مدل چنین می‌شود:

$$\log\theta_{ij} = \begin{cases} \beta + 2\delta & i = j \\ \beta - \delta & |i-j| = 1 \\ \beta & |i-j| > 1 \end{cases}$$

www.SID.ir

Archive of SID

پیوست۲ – فرم اطلاعاتی بررسی نیاز به درمان‌های ارتودنسی

شماره مسلسل: شماره پرونده/کست:

جزء سلامت دندانی (dental health component)

Character	Grade 2	Grade 3	Grade 4		Grade 5
Hypodontia (missing)			**محدود (حداکثر یک دندان در هر کواردانت)**		**شدید (بیش از یک دندان در یک کواردانت)**
Overjet	3.5-6mm With Competent Lips	3.5-6mm With Incomp. Lips	6-9mm		> 9mm
Reverse overjet	0-1mm	1-3.5mm	S- **مشکلات** +1-3.5 M	> 3.5mm	> 3.5mm S-M **مشکلات** +
Crossbite	**با** ≤ 1mm	**با** 1-2mm **فاصله بین** ICR **و** RCP	Scissors bite	**با بیش از** 2mm **فاصله بین** RCP **و** ICP	
Contact Displacement	1-2mm	2-4mm	> 4mm		
Openbite	1-2mm	2-4mm	> 4mm		
Overbite	> 3.5mm **بدون تماس با لثه**	**در تماس با لثه بدون تروما**	**در تماس با لثه همراه تروما**		

اکلوژن Prenormal **یا** Postnormal (2g)

دندان اضافی (4x)

دندان نیمه رویش یافته که با فشار دندان‌های مجاور منحرف شده باشد (4t)

تأخیر یا توقف رویش به علت شلوغی جابه‌جایی و حضور دندان‌های اضافی، شیری یا پاتولوژی (5i)

آنومالی‌های کرانیوفاسیال و شکاف لب و کام (5p)

دندان آنکیلوتیک (submerged) **شیری** (5s)

نام معاینه‌کننده:

تاریخ معاینه:

با توجه به وضعیت دندانی بیمار در رتبه ..... **قرار می‌گیرد.**

www.SID.ir