دانشور
پزشکی

یک خط مشی جدید در مدل‌بندی توافق داده‌های ترتیبی

نویسندگان: علیرضا اکبرزاده‌باغبان[1]، دکتر غلامرضا بابایی[2] و دکتر انوشیروان کاظم نژاد[2]

۱- دانشجوی دکتری آمار زیستی، دانشکده علوم پزشکی، دانشگاه تربیت مدرس
۲- دانشیارگروه آمار زیستی، دانشکده علوم پزشکی، دانشگاه تربیت مدرس

چکیده

سابقه و اهداف: در آماربرای تعیین تطابق بین اندازه گیری‌های حاصل از مقیاس‌های ترتیبی، روش‌هایی از قبیل محاسبه شاخص‌های توافق و همچنین مدل‌بندی توافق موجود است. در این مقاله ضمن بیان مشکلات استفاده از شاخص‌های توافق، موضوع مدل‌بندی ساختار توافق مورد بررسی قرار گرفته است.

روش بررسی: این کار به کمک مدل جدید «پیوند مربع امتیازات» انجام گرفته و نتیجه این مدل با نتایج حاصل از مدل‌های موجود «پیوند یکنواخت»، «پارامتر قطری» و «توافق به‌علاوه پیوند یکنواخت» از طریق بیان دقیق جزئیات و برازش آن‌ها به داده‌های مربوط به شاخص ترتیبی نیاز به درمان ارتدنسی مقایسه شده است.

یافته‌ها: نتایج نشان داد که مدل پیوند یکنواخت (با p=۰/۰۴۹) به داده‌ها برازش ندارد، در حالی‌که مدل‌های پارامتر قطری (با p=۰/۱۳۳)، پیوند مربع امتیازات (با p=۰/۱۸۳) و توافق به‌علاوه پیوند یکنواخت (با p=۰/۶۸۴) برای داده‌ها مناسبند.

بحث: در بین مدل‌هایی که تنها یک پارامتر بیش از مدل استقلال داشتند، مدل پیوند مربع امتیازات، بهترین برازش را نشان داد، در حالی که مدل توافق به‌علاوه پیوند یکنواخت، به دلیل دارا بودن پارامترهای بیش‌تر، بهتر ازهمه مدل‌ها بود. به‌علاوه نتایج هر دو مدل پیوند مربع امتیازات، و توافق به‌علاوه پیوند یکنواخت نشان داد از بین دو دندان‌پزشکی که نمونه‌ها را نرخ‌گذاری کردند، دندان‌پزشک ۱ نسبت به دندان‌پزشک ۲ تمایل به نرخ‌گذاری (طبقه‌بندی) بالاتر دارد.

واژه‌های کلیدی: مدل‌بندی ساختار توافق، مدل پیوند، شاخص نیاز به درمان ارتدنسی (IOTN)

دوماهنامه علمی - پژوهشی
دانشگاه شاهد
سال دوازدهم - شماره ۵۴
دی ۱۳۸۳

مقدمه

فرض کنیم دو آزماینده به‌طور جداگانه، هر کدام از افراد یک نمونه را طبق یک مقیاس ترتیبی طبقه‌بندی می‌کنند. در چنین گروه‌بندی‌هایی به‌ندرت توافق کاملی بین آزماینده‌ها وجود دارد، یعنی داده‌ها صد در صد قابل اعتماد نیستند. ارزیابی قابلیت اعتماد (reliability) از منظر توافق بین آزماینده‌ها قابل بررسی است [1].

فرض کنیم نرخ‌گذاری توأم این دو آزماینده در یک جدول توافقی آمده باشد. در چنین جدولی، دو موضوع را می‌توان بررسی کرد: ۱) همگنی و ناهمگنی توزیع‌های حاشیه‌ای این دو نرخ‌گذاری. معمولاً برای گروه‌های پاسخ ترتیبی علاقه‌مندیم بدانیم آیا گروه‌بندی یک آزماینده نسبت به آزماینده دیگر، میل به بالاتر (پایین‌تر) بودن دارد یا دقیقاً بر هم منطبقند. ۲) توافق

بین آزماینده‌ها. این موضوع، فراوانی روی قطر اصلی در توزیع توأم نرخ‌گذاری‌ها را پوشش می‌دهد. این مقاله عمدتاً به موضوع دوم می‌پردازد. برای بررسی کامل موضوع اول مرجع [۲] را ببینید.

ساده‌ترین و متداول‌ترین راه برای بررسی توافق بین آزماینده‌ها، محاسبه شاخص‌های توافق بین نرخ‌گذاری‌های آن‌ها است. در این میان ضرایب کاپا و کاپا وزن‌دار، بیش‌تر از سایر شاخص‌ها مورد استفاده قرار می‌گیرند [۳و۴]. مقالات متعددی در خصوص نقاط ضعف این شاخص‌ها و مشکلات استفاده از آن‌ها منتشر شده است. پاره‌ای از این مشکلات عبارتند از: ۱) این شاخص‌ها برای مقایسه فراوانی‌های مشاهده شده و مورد انتظار، مدل استقلال بین نرخ‌گذاری‌ها را به‌عنوان رابطه مبنا مورد استفاده قرار می‌دهند و لذا در صورتی که این فرض برقرار نباشد، استفاده از آن‌ها می‌تواند گمراه‌کننده باشد و اتفاقاً در داده‌های با مقیاس ترتیبی به ندرت فرض استقلال برقرار است [۱]. ۲) این ضرایب فقط «سطح توافق روی هم مشاهده شده» را با «سطح توافق روی هم مورد انتظار» مقایسه می‌کنند و به الگوی توافق (نحوه پراکندگی داده‌ها در جدول) توجهی ندارند [۵]. ۳) این شاخص‌ها فقط برای جداول مربعی که در آن‌ها هر دو آزماینده سطوح یکسانی از نرخ‌گذاری را تشخیص دهند قابل محاسبه است؛ یعنی در صورتی‌که یکی از سطوح مقیاس ترتیبی، توسط یک آزماینده تشخیص داده شود ولی توسط آزماینده دیگر تشخیص داده نشود، این ضرایب قابل محاسبه نیستند. ۴) کاپا وزن‌دار که از آن فقط برای بررسی توافق بین نرخ‌گذاری‌های ترتیبی استفاده می‌شود، بیش‌تر از نسبت دقیق توافق نسبت به انتخاب وزن‌ها حساس است [۱]؛ به این معنا که با دادن مجموعه‌های متفاوتی از وزن‌ها به سلول‌های جدول توافقی، مقادیر مختلفی برای این ضریب به دست می‌آیند و در نتیجه، تفسیر این ضریب دشوار می‌شود. ۵) این ضرایب تنها اندازه‌هایی از توافقند و هیچ اطلاعی در مورد ساختار یا الگوی توافق در اختیار نمی‌گذارند [۱]. بنابراین بهتر است به جای توصیف ساختار توافق داده‌های چنین جدولی توسط یک عدد، این ساختار را مدل‌بندی کنیم. اولین بار این کار در سال ۱۹۸۵ توسط تانر و یانگ [۶] انجام پذیرفت. این بحث با برآورد معادله رگرسیون خطی برای دو متغیر کمی، به‌جای محاسبه ضریب همبستگی خطی پیرسن برای آن دو متغیر، قابل مقایسه است.

در این مقاله برای مدل‌بندی ساختار مذکور، مدل پیوند مربع امتیازات برای نخستین بار معرفی گردیده و از مدل‌های موجود پارامتر قطری، پیوند یکنواخت و توافق به‌علاوه پیوند یکنواخت نیز استفاده شده است. در نهایت با برازاندن این مدل‌ها به داده‌های مربوط به شاخص ترتیبی، نیاز به درمان ارتدنسی مدل‌ها را با هم مقایسه کرده، بهترین مدل را ارائه می‌کنیم و به تفسیر پارامترهای آن می‌پردازیم.

در این مقاله، نرخ‌گذاری دو نفر آزماینده و مدل‌بندی ساختار توافق بین داده‌های آن‌ها مورد نظر است، در حالی‌که این بحث‌ها تماماً برای موقعیت‌هایی که در آن‌ها دو روش تشخیصی و یا دو ابزار تشخیصی داریم، یا یک آزماینده و یک استاندارد طلایی داریم و یا یک آزماینده در دو زمان هر کدام از افراد یک نمونه را مورد ارزیابی قرار می‌دهد، قابل استفاده‌اند.

مدل‌ها

۱- مدل‌پارامتر قطری (diagonal parameter model)

فرض کنیم هر کدام از n فرد یک نمونه، توسط دو آزماینده A و B به‌طور مستقل به یکی از r گروه جدا از هم در یک مقیاس ترتیبی نسبت داده می‌شوند. اگر نتیجه این دو نرخ‌گذاری را در یک جدول مربع متقاطع بیاوریم و $(i, j = 1,2,\ldots,r)$ $m_{ij} = n\pi_{ij}$ را معرف فراوانی مورد انتظار نرخ‌گذاری i توسط آزماینده اول و نرخ‌گذاری j توسط آزماینده دوم در نظر بگیریم، ساده‌ترین مدل برای بیان ساختارتوافق داده‌های موجود در چنین جدولی عبارت است از:

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

که در آن λ_i^A و λ_j^B به ترتیب اثر نرخ‌گذاری آزماینده‌های A و B است و پارامتر $\delta(i,j)$ که فقط برای سلول‌های روی قطر اصلی در نظر گرفته می‌شود، معرف توافقی است که ورای شانس انتظار داریم. در حالت خیلی ساده $\delta(i,j)$ را می‌توان به‌صورت زیر در نظر گرفت که طبق آن برای همه سلول‌های روی قطر اصلیِ جدول فقط یک پارامتر در نظر گرفته می‌شود [۶]:

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

همان‌گونه که ملاحظه می‌شود این مدل نسبت به مدل استقلال، تنها عبارت $\delta(i,j)$ را بیش‌تر دارد تا بتوان به کمک آن توافق موجود در داده‌های حاصل از دوبار اندازه‌گیری را منظور کرد. معمولاً مدل استقلال به این نوع از داده‌ها برازش ندارد؛ زیرا در نرخ‌گذاری‌های متقاطع مقیاس ترتیبی رابطه مستقیم و غالباً محکمی بین سطوح نرخ‌گذاری آزماینده‌ها وجود دارد [۷-۹]. در صورتی که مدل استقلال به داده‌ها برازش مناسبی داشته باشد، یعنی $\delta(i,j)$ برابر صفر باشد، نتیجه می‌گیریم که هیچ توافقی ورای آنچه به تصادف انتظار داریم بین دو آزماینده وجود ندارد [۷]. از آن‌جا که تعداد درجات آزادی برای بررسی مناسبت مدل استقلال $(r-1)^2$ است، لذا تعداد درجات آزادی برای بررسی مناسبت این مدل می‌شود $df=(r-1)^2-1$ [۹].

برای مقایسه وضعیت نرخ‌گذاری آزماینده‌ها از پارامتر نسبت بخت موضعی (local odds ratio) استفاده شده است که به‌صورت زیر تعریف می‌شود:

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

با توجه به این‌که همه مدل‌های این مقاله به‌صورت لگ خطی هستند، برای همه آن‌ها ابتدا لگاریتم نسبت بخت موضعی محاسبه شده است. نهایتاً هنگام تفسیر نتایج، با انجام یک تبدیل معکوس، مقدار θ_{ij} به‌دست آمده است. با این توضیحات، لگاریتم نسبت بخت موضعی برای این مدل عبارت است از:

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

در حالت $|i-j|>1$ هیچ کدام از چهار سلولی که درمحاسبه نسبت بخت موضعی دخالت دارند، یعنی سلول‌های (i,j)، (i+1 , j+1)، (i+1,j)، (i,j+1) روی قطر اصلی نیستند. برای حالت $|i-j|=1$ تنها یکی از این چهار سلول روی قطر اصلی است و در حالت $i=j$ دو تا از این سلول‌ها روی قطر اصلی هستند.

تعمیم حالت قبل برای $\delta(i,j)$ را می‌توان به‌صورت زیر در نظر گرفت:

$$.\delta(i,j) = \begin{cases} \delta_i & i=j, i=1,2,\dots,r \\ 0 & i \neq j \end{cases}$$

بر عکس حالت قبل، در این حالت مدل روی قطر اصلی جدول اشباع است، زیرا برای هر کدام از سلول‌های روی قطر اصلی، یک اثر در نظر گرفته شده است. با توجه به قیود لازم روی پارامترها، تعداد درجات آزادی برای بررسی مناسبت مدل در این حالت $df=(r-1)(r-2)$ می‌شود [۱۰]. در اغلب موارد در نظر گرفتن تنها یک پارامتر برای همه سلول‌های روی قطر اصلی کافی است و نیازی به در نظر گرفتن حالت اشباع شده نیست؛ ضمن این‌که برازش مدل اشباع شده و برآورد پارامترهای آن در بسیاری از موارد مقدور نیست [۶]. لذا در این مقاله نیز مدل به‌صورت اشباع نشده در نظر گرفته شده است.

اگر فرض کنیم آزماینده‌ها توافق ندارند، مدل فوق در هر دو حالت این نتیجه را دربر دارد که نرخ‌گذاری آن‌ها مستقلند. این رفتار ممکن است در مقیاس نرخ‌گذاری ترتیبی دیده نشود، زیرا بین این نرخ‌گذاری‌ها رابطه مثبت متوسط تا قوی دیده می‌شود [۹]. لذا باید برای مقیاس‌های ترتیبی مدل‌های دیگری را در نظر گرفت.

۲- مدل پیوند خط به خط

(Linear by linear association model)

برای طبقه‌بندی متقاطع داده‌های ترتیبی حاصل از نرخ‌گذاری دو آزماینده، مدل پیوند خط به خط، الگوی ساده و غالباً مناسبی را ارائه می‌کند [۹]. این مدل به‌صورت زیر است:

$$\log m_{ij} = \mu + \lambda_i^A + \lambda_j^B + \beta u_i u_j \quad (۲)$$

در این مدل $u_1 < u_2 < ... < u_r$ امتیازات ثابتی هستند که به گروه‌های ترتیبیِ 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:

$df=(r-1)^2-1$. برای این مدل داریم

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

۲- الف) مدل پیوند یکنواخت

(uniform association model)

برازش مدل ۲ منوط به اتخاذ تصمیمی در مورد امتیازات گروه‌ها است. یک راه برای انجام این کار- که در این مقاله مورد نظر نیست- این است که امتیازات گروه‌ها را به منزله پارامترهایی در نظر گرفته، با منظور کردن قیود لازم، آن‌ها را از روی داده‌ها برآورد کنیم [۱و۹]. راه دوم برای انجام این کار- که معمولاً اجرا می‌شود- در نظر گرفتن شماره هر گروه به عنوان امتیاز آن گروه است، یعنی $u_i = i$. با این فرض، مدل ۲ به مدل پیوند یکنواخت که توسط گودمن [۷] معرفی گردید، تبدیل می‌شود و بدیهی است که در این صورت داریم: $\log \theta_{ij} = \beta$.

۲- ب) مدل پیوند مربع امتیازات

(square scores association model)

حالت خاص $u_i = i$ در مدل قبل، ساده‌ترین حالت از حالات کلی است که در آن‌ها فواصل بین امتیازات گروه‌ها مساوی (equal-interval scores) هستند. فواصل مساوی بین امتیازات گروه‌ها بدین‌معنا است که امتیازات طبق یک تصاعد حسابی به گروه‌ها نسبت داده می‌شوند. کلیه این حالات مدل پیوند خط به خط (۲) را به مدل پیوند یکنواخت تبدیل می‌کنند. به‌عنوان مثال می‌توان برای سادگی محاسبه در یک مقیاس ترتیبی ۵ حالته، امتیازات گروه‌ها را به ترتیب ۲-، ۱-، ۰، ۱و۲ در نظر گرفت [۹].

در بسیاری از موارد، مساوی بودن فاصله بین گروه‌های ترتیبی منطقی نیست؛ زیرا تفاوت خصوصیت مورد نظر بین گروه‌های مجاور می‌تواند بسیار زیاد باشد، ولی محدودیت سطوح موجب می‌گردد تا این گروه‌ها در کنار هم قرار گیرند. یکی از موقعیت‌هایی که در آن می‌توان این نقص را تا حدی برطرف کرد، مدل‌بندی ساختار داده‌های ترتیبی است؛ زیرا می‌توان با انتساب امتیازات واقعی تر به این گروه‌ها، خلأ به‌وجود آمده را تاحدی پر کرد. به عبارت دیگر، هنگامی که خصوصیت مورد نظر - مانند میزان درد - با قاعده‌ای غیر از تصاعد حسابی افزایش می‌یابد، اختصاص امتیازات به گروه‌ها را در قالب این تصاعد می‌توان چشم‌پوشی از طبیعت آن خصوصیت تلقی کرد.

حالت خاص دیگری که می‌خواهیم برای تحقق هدف فوق برای اولین بار در این مقاله در نظر بگیریم و محاسن آن را نسبت به روش‌های مرسوم بررسی کنیم عبارت است از $u_i = i^2$ ؛ بدین‌معنا که امتیاز هر گروه را برابر توان دوم شماره آن گروه در نظر می‌گیریم. در این صورت مدل ۲ را «مدل پیوند مربع امتیازات» می‌نامیم. برای این مدل، لگاریتم نسبت بخت موضعی عبارت است از:

$$\log \theta_{ij} = \beta i^2 j^2 + \beta(i+1)^2 (j+1)^2 - \beta i^2 (j+1)^2 - \beta(i+1)^2 j^2$$
$$= \beta(4ij + 2i + 2j + 1)$$

که بعد از ساده کردن به‌صورت زیر تبد یل می‌شود:

$$\log \theta_{ij} = \begin{cases} \beta(2i+1)^2 & i = j \\ \beta(i+j)(2(\max(i,j)+1) & |i-j| = 1 \end{cases}$$

با توجه به این‌که به ازای کلیه حالات |i-j|>1 نمی‌توان فرمول بسته‌ای برای لگاریتم نسبت بخت موضعی به‌دست آورد، از ذکر آن خودداری می‌کنیم؛ ضمن این‌که همان‌طور که خواهیم دید در تفسیر نتایج، عمده توجه فقط به حالت اول، یعنی i=j است.

۳- مدل توافق به‌علاوه پیوند یکنواخت

(agreement plus uniform association model)

اگرچه برای توصیف رابطه بین گروه‌بندی‌های ترتیبی، مدل‌های پیوند یکنواخت و پیوند مربع امتیازات می‌توانند مناسب باشند، ولی چون این مدل‌ها به رفتار ویژه سلول‌های روی قطر اصلی توجهی نمی‌کنند، گزینه خوبی برای مدل‌بندی توافق نیستند. مدل زیر با در نظر گرفتن اثر سلول‌های روی قطر اصلی، ورای آنچه مدل‌های پیوند یکنواخت و پیوند مربع امتیازات ارائه می‌کنند، 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 به‌دلیل رابطه مبنایی که بین نرخ‌گذاری‌های ترتیبی وجود دارد. ۳) توافق مازاد روی قطر اصلی که ورای شانس و رابطه مبنا است [۹].

این مدل که قادر به تفکیک سه مؤلفه مذکور است و به آن مدل توافق به‌علاوه پیوند یکنواخت اطلاق می‌گردد به‌صورت زیر است [۹]:

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

که‌در آن $\delta(i,j)$ و $\beta u_i u_j$ جملاتی‌هستند که‌به‌ترتیب در مدل‌های پارامتر قطری‌و پیوند خطی‌به‌خط معرفی‌شده‌اند. همان‌طور که ملاحظه می‌شود این مدل نسبت به مدل استقلال دو پارامتر بیش‌تر دارد و لذا تعداد درجات آزادی برای بررسی مناسبت آن می‌شود $df=(r-1)^2-2$.

ــ حالات خاص: ۱) $\delta=0$ که در این صورت، این مدل به ازای امتیازات ثابت $u_i=i$ به مدل پیوند یکنواخت و به ازای $u_i=i^2$ به مدل پیوند مربع امتیازات تبدیل می‌شود. ۲) $\beta=0$ که در این صورت، این مدل به مدل پارامتر قطری تبدیل می‌شود. ۳) $\beta=\delta=0$ که در این صورت این مدل به مدل استقلال تبدیل می‌شود. برای سادگی محاسبات و جلوگیری از پیچیدگی موضوع، در ادامه بحث، امتیازات گروه‌ها را برای مدل ۳ به‌صورت $u_i=i$ در نظر می‌گیریم. در این شرایط، لگاریتم نسبت بخت موضعی برای این مدل می‌شود:

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

نتایج

برای بررسی و مقایسه کارایی مدل‌های مذکور، از داده‌های مربوط به طرح ارزیابی روایی و پایایی شاخص نیاز به درمان ارتدنسی، (Index of Orthodontic Treatment Need: IOTN)، که در تابستان سال ۱۳۸۲ در مرکز تحقیقات علوم دندان‌پزشکی واقع در دانشگاه علوم پزشکی شهید بهشتی انجام شد، استفاده می‌کنیم. در این طرح، هر کدام از ۲۰ دندان‌پزشک به‌طور مستقل این شاخص را روی ۶۹ نمونه اندازه‌گیری کردند. شاخص مذکور که به کمک ۳۱ مشخصه و به‌صورت بیش‌ترین امتیاز (maximum score) اندازه‌گیری می‌شود، ترتیبی 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 ارتدنسی‌مطرح‌است مورد استفاده قرارمی‌گیرد [۱۱-۱۷].

با توجه به این‌که هدف ما در این مقاله ارائه راهکاری جدید در مدل‌بندی ساختار توافق زوجی داده‌ها و مقایسه آن با راهکارهای موجود است، لذا تنها از داده‌های مربوط به اندازه‌گیری دو نفر از این دندان‌پزشکان استفاده شده است. از آن‌جا که سطح ۱ شاخص مذکور خیلی کم دیده شد، سطوح ۱ و ۲ این شاخص را با هم در نظر گرفته، به‌عنوان یک سطح

جدول ۱: طبقه‌بندی متقاطع نرخ‌گذاری‌های دندانپزشکان همراه با فراوانی‌های مورد انتظاربرای مدل‌های توافق به‌علاوه پیوند یکنواخت و پیوند مربع امتیازات

دندانپزشک۲ / دندانپزشک۱	نیاز جزئی یا عدم‌نیاز به درمان	نیاز متوسط به درمان	نیاز شدید به درمان	نیاز بسیار شدید به درمان
نیاز جزئی یا عدم نیاز به درمان	۱# ۰/۹۶$ ۰/۵۱&	۰ ۰/۰۴ ۰/۴۱	۰ ۰ ۰/۰۳	۰ ۰ ۰
نیاز متوسط به درمان	۰ ۰/۸۲ ۱/۶۳	۶ ۴/۹۴ ۳/۱۵	۰ ۰/۲۳ ۱/۲۲	۰ ۰ ۰
نیاز شدید به درمان	۲ ۱/۱۸ ۰/۸۲	۵ ۵/۳۰ ۷/۴۰	۳۸ ۳۸/۸۱ ۳۷/۱۶	۱ ۰/۷۰ ۰/۶۱
نیاز بسیار شدید به درمان	۰ ۰/۰۳ ۰	۰ ۰/۷۲ ۰/۰۲	۵ ۳/۹۶ ۴/۵۸	۱۱ ۱۱/۲۳ ۱۱/۳۹

فراوانی مشاهده شده

$ فراوانی مورد انتظاربرای مدل توافق به‌علاوه پیوند یکنواخت

& فراوانی مورد انتظاربرای مدل پیوند مربع امتیازات

جدول ۲: خلاصه‌ای از مدل‌های برازش داده شده به داده‌های جدول ۱

مدل	آماره‌نیکویی برازش	درجه‌آزادی	سطح‌معناداری
۱. پیوند یکنواخت	۱۵/۵۵	۸	۰/۰۴۳۹
۲. پارامتر قطری	۱۲/۴۲	۸	۰/۱۳۳۵
۳.پیوندمربع‌امتیازات	۱۱/۳۵	۸	۰/۱۸۲۵
۴.توافق‌به‌علاوه‌پیوند یکنواخت	۴/۸۰	۷	۰/۶۸۴۱

منظور کردیم. با این شرایط، نرخ‌گذاری این دو دندان‌پزشک، در جدول توافقی ۴×۴ شماره ۱ آمده است. برای برازاندن مدل‌ها به داده‌های این جدول از نرم‌افزار SAS استفاده شد و پارامترهای این مدل‌ها با روش حداکثر درستنمایی (ML) برآورد گردیدند.

همان‌طور که دیده می‌شود ۸ سلول از ۱۶ سلول جدول دارای فراوانی مشاهده شده صفر هستند. این وضعیت عموماً برای چنین مجموعه داده‌هایی اتفاق می‌افتد، زیرا بررسی اعتبار داده‌ها و اندازه‌گیری توافق آن‌ها در یک پیش مطالعه، قبل از مطالعه اصلی و با حجم نمونه کم انجام می‌شود [۱۸] . زمانی که آزمون نیکویی برازش برای چنین داده‌هایی به کاربرده می‌شود، توزیع آماره نسبت درستنمایی G^2 توسط «توزیع مجذور خی» به خوبی تقریب زده نمی‌شود، اما این آماره برای مقایسه مدل‌ها خوب عمل می‌کند [۹].

جدول ۲ به ترتیب وضعیت و میزان برازندگی، آماره نیکویی برازش مدل‌های مذکور را همراه با درجه آزادی و سطح معناداری آن‌ها در بر دارد. ملاک پذیرش یک مدل به‌عنوان مدل مناسب آن‌است‌که مقدار احتمال آماره نیکویی برازش آن حداقل ۰/۰۵ باشد.

بنابراین مدل استقلال برای این داده‌ها مناسب نیست؛ زیرا برای این مدل، مقدارآماره نیکویی برازش و سطح معناداری به‌ترتیب برابرند با $G^2 = ۶۵/۰۷$ و $p= ۰/۰۰۰۰$.

ملاحظه می‌شود که با وارد کردن پارامتر β به مدل استقلال و رسیدن به مدل پیوند یکنواخت، اصلاح بسیار بزرگی اتفاق افتاده به‌طوری که مقدار آماره نیکویی برازش به ۱۵/۵۵تقلیل یافته، اما با $p=۰/۰۴۹۳$ هنوز این مدل به داده‌ها برازش پیدا نکرده است.

برای مدل پارامتر قطری مقدار آماره نیکویی برازش باز هم کاهش یافته است. این مدل با $p=۰/۱۳۳۵$ ضعیف‌ترین برازش را به داده‌ها نشان می‌دهد. مدل مناسب بعد، مدل پیوند مربع امتیازات است. این مدل با $G^2 =۱۱/۳۵$ و $p=۰/۱۸۲۵$ از مدل پارامتر قطری بهتر است. بنابراین چنانچه بخواهیم مدل‌هایی را که فقط یک پارامتر نسبت به مدل استقلال بیش‌تر دارند با هم مقایسه و مدلی را انتخاب کنیم، مدل پیوند مربع امتیازات بهترین انتخاب است. برای این مدل، مقدار ضریب β برابر ۰/۱۰ با خطای معیار تقریبی ۰/۰۲ برآورد گردیده و با $p<۰/۰۰۰۱$ معنادار شد. ملاک معنادار بودن پارامترهای مدل‌ها آن است که مقدار احتمال آماره آزمون آن‌ها حداکثر ۰/۰۵ باشد.

به‌منظور دستیابی به نتیجه بهتر، مدل توافق به‌علاوه پیوند یکنواخت را بررسی می‌کنیم. این حدس از آن‌جا ناشی می‌شود که این مدل دارای پارامترهای بیش‌تری نسبت به مدل‌های قبل است. از جدول ۲ می‌بینیم که برازش این مدل به داده‌ها کاملاً بهتر شده است؛ زیرا در این حالت $G^2 = ۴/۸۰$ و $p= ۰/۶۸۴۱$ است. برای این مدل، پارامتر پیوند یکنواخت برابر ۱/۴۹ با خطای معیار تقریبی ۰/۶۹ و پارامتر قطر اصلی برابر ۱/۷۸ با خطای معیار تقریبی ۰/۵۵ برآورد گردیدند و به‌ترتیب با $p<۰/۰۵$ و $p<۰/۰۱$ از نظر آماری معنادار شدند.

بحث و نتیجه‌گیری

جدول ۱ فراوانی مشاهده شده در نتیجه نرخ‌گذاری دو دندان‌پزشک را همراه با فراوانی مورد انتظار برای مدل‌های پیوند مربع امتیازات و توافق به‌علاوه پیوند یکنواخت در بر دارد. می‌بینیم که تحت این مدل‌ها، فراوانی‌های مورد انتظار و مشاهده شده به هم نزدیک هستند. مجموع مربعات باقی‌مانده‌های این مدل‌ها به ترتیب برابرند با: ۲۴/۳۲ و ۵/۰۴. یکی از دلایل بهتر بودن مدل توافق به‌علاوه پیوند یکنواخت، نسبت به مدل پیوند مربع امتیازات، بیش‌تر بودن تعداد پارامترها در این مدل است. بنابراین در صورتی که قادر به برازاندن مدل توافق به‌علاوه پیوند یکنواخت به داده‌ها باشیم، بهترین انتخاب را انجام داده‌ایم؛ زیرا هم رابطه بین سطوح مقیاس، و هم اثر مربوط به قطر اصلی جدول را منظور کرده‌ایم. اما تجربه نویسندگان این مقاله حاکی است برازاندن این مدل و برآورد پارامترهای آن – به‌دلیل کم بودن حجم نمونه در بررسی‌های این چنینی– در اغلب موارد مقدور نیست، در حالی که مدل‌های ساده تر را می‌توان به راحتی برازش داد و پارامترهای آن‌ها را برآورد کرد.

از طرفی با وجود ترتیبی بودن مقیاس داده‌ها، دیدیم که مدل پیوند یکنواخت به داده‌ها برازش ندارد، در حالی‌که مدل پیوند مربع امتیازات به داده‌ها برازش دارد. لذا در هر مورد که مقیاس داده‌ها ترتیبی باشد، در نظر گرفتن فواصل مساوی برای امتیازات گروه‌ها، می‌تواند انتخاب مناسبی نباشد.

از آن‌جا که در مدل پارامتر قطری هیچ فرضی برای امتیازات گروه‌ها قائل نمی‌شویم و در این مدل با گروه‌های ترتیبی، مانند سطوح مقیاس اسمی برخورد می‌شود، می‌توان از آن برای مدل‌بندی ساختار توافق داده‌های اسمی نیز استفاده کرد. البته این خاصیت در مدل‌بندی ساختار توافق داده‌های ترتیبی، ضعف این مدل محسوب می‌گردد، زیرا در این صورت، اطلاع موجود در رتبه‌های گروه‌ها نادیده گرفته می‌شود.

با توجه به اهمیت امتیازات گروه‌های ترتیبی در مدل‌بندی ساختار توافق داده‌ها، بعضی از محققین این امتیازات را پارامتردر نظر گرفته، آن‌ها را از روی داده‌ها برآورد کرده‌اند [۱و۹] . اگرچه این کار به ظاهر بسیار ارزشمند به نظر می‌رسد، ولی انجام آن بر یک فرض غیرمعقول استوار شده است؛ بدین‌معنا که در این کار،امتیاز گروه اول برابر ۱ و امتیاز گروه آخر برابر r فرض شده و سایر امتیازات برآورد گردیده‌اند. بدیهی است که در این صورت، سایر امتیازات مقیدند تا بین ۱ و r برآورد گردند. قطعاً بهترین امتیازات برآورد شده آن‌هایی هستند که بدون هیچ قیدی برآورد گردند، ولی چنانچه لازمه برآورد امتیازات، مشخص کردن دو تا از آن‌ها باشد، شاید بهتر باشد که u_1 را برابر ۱ و u_2 را برابر ۲ در نظر بگیریم و اجازه دهیم تا سطوح بالایی مقیاس بدون هیچ قید اضافه دیگر برآورد گردند.

یکی دیگر از مزایای مدل جدید پیوند مربع امتیازات این است که در این مدل، علاوه بر منظور کردن امتیاز گروه‌ها در زمان مدل‌بندی، هنگام تفسیر نتایج نیز به جایگاه گروه‌ها توجه می‌شود؛ در حالی که سایر مدل‌های مورد بحث در این مقاله، امتیازات گروه‌ها را حداکثر در زمان مدل‌بندی در نظرمی‌گیرند. مثلاً نتیجه مدل توافق به‌علاوه پیوند یکنواخت را در حالت i=j می‌توان به این صورت تفسیر کرد که به ازای ۳و۲و۱ =i بخت نرخ‌گذاری ۱ +i نسبت به i توسط دندان‌پزشک ۱ تقریباً ۱۵۶ = $\exp(\hat{\beta}+2\hat{\delta})$ بار نسبت به همین بخت در دندان‌پزشک ۲ بزرگ‌تر است؛ یعنی دندان‌پزشک ۱ نسبت به دندان‌پزشک ۲ تمایل به نرخ‌گذاری بالاتر دارد و این تمایل در همه سطوح یکسان است. اما طبق مدل پیوند مربع امتیازات، این تمایل با تغییر سطوح مقیاس عوض می‌شود. نتایج مدل پیوند مربع امتیازات را درحالت i=j، به ازای ۳ و۲و۱ =i با توجه‌به رابطه $\hat{\theta}_{ii}=\exp(\hat{\beta}(2i+1)^2)=\exp(0.1(2i+1)^2)$ در جدول۳ خلاصه کرده‌ایم. همان‌طور که از این جدول مشخص است تمایل به نرخ‌گذاری بالاتر در دندان‌پزشک ۱ نسبت به دندان‌پزشک ۲ در سطوح بالای مقیاس بیش‌تر است.

جدول ۳: برآورد نسبت بخت موضعی درحالت i=j برای مدل پیوند مربع امتیازات

i	۱	۲	۳
$\hat{\theta}_{ii}$	۲/۵	۱۲/۲	۱۳۴/۳

نهایتاً باید بگوییم که نتایج حاصل از مدل‌های این مقاله هنگامی ارزشمندتر و آگاهی بخش‌تر خواهند بود که نرخ‌گذاری آزماینده‌ها را با یک استاندارد طلایی (gold standard) مقایسه کرده، در مورد چگونگی ارزیابی آن‌ها نسبت به استاندارد طلایی اظهار نظر کنیم.

پیوست

نحوه اثبات لگاریتم نسبت بخت موضعی در مدل پیوند مربع امتیازات

فرض کنیم هر کدام از n فرد یک نمونه، توسط دو آزماینده A و B به طورمستقل به یکی از r گروه جدا از هم در یک مقیاس ترتیبی نسبت داده می‌شوند. اگر نتیجه این دو نرخ‌گذاری را در یک جدول مربع متقاطع بیاوریم و $(i,j=1,2,...,r)$ $m_{ij}=n\pi_{ij}$ رامعرف فراوانی مورد انتظار نرخ‌گذاری i توسط آزماینده اول و نرخ‌گذاری j توسط آزماینده دوم در نظر بگیریم، مدل پیوند خط به خط الگوی ساده وغالباً مناسبی، برای مدل‌بندی ساختار توافق داده‌های این جدول ارائه می‌کند. این مدل به‌صورت زیر است:

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

که در آن λ_i^A و λ_j^B به‌ترتیب اثر نرخ‌گذاری آزماینده‌های A و B بوده و $u_1<u_2<...<u_r$ امتیازات ثابتی هستند که به گروه‌های ترتیبی پاسخ نسبت داده می‌شوند و $\beta u_i u_j$ مبین رابطه موجود بین سطوح نرخ‌گذاری است. برای مقایسه وضعیت نرخ‌گذاری آزماینده‌ها از طریق مدل‌بندی، از نسبت بخت موضعی

(local odds ratio) استفاده می‌شود که به‌صورت زیر تعریف می‌شود:

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

لگاریتم نسبت بخت موضعی برای مدل فوق عبارت است از $\log\theta_{ij} = \beta(u_{i+1} - u_i)(u_{j+1} - u_j)$. با فرض $u_i = i$ مدل فوق به مدل پیوند یکنواخت با $\log\theta_{ij} = \beta$ تبدیل می‌شود. فرض اخیر نمی‌تواند برای همه موقعیت‌ها مناسب باشد، زیرا ممکن است تفاوت بین سطوح مقیاس ترتیبی یک عدد ثابت نباشد. بنابراین مدل پیوند خط به خط را در حالت $u_i = i^2$ در نظر می‌گیریم و نام آن‌را مدل پیوند مربع امتیازات می‌نهیم. با توجه به خواص تابع لگاریتم، می‌توانیم لگاریتم نسبت بخت موضعی را به‌صورت زیر بنویسیم:

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

حال اگر فرض $u_i = i^2$ را در نظر بگیریم، عبارت فوق می‌شود:

$$\log\theta_{ij} = \mu + \lambda_i^A + \lambda_j^B + \beta i^2 j^2 + \mu + \lambda_{i+1}^A + \lambda_{j+1}^B + \beta(i+1)^2(j+1)^2$$
$$- \mu - \lambda_i^A - \lambda_{j+1}^B - \beta i^2 (j+1)^2 - \mu - \lambda_{i+1}^A - \lambda_{j+1}^B - \beta(i+1)^2 j^2$$

عبارت فوق بعد از ساده کردن به رابطه زیر تبدیل می‌شود:

$$\log\theta_{ij} = \beta i^2 j^2 + \beta(i+1)^2(j+1)^2 - \beta i^2 (j+1)^2 - \beta(i+1)^2 j^2$$

$$= \beta(4ij + 2i + 2j + 1)$$

حال اگر در دو حالت $i = j$ و $|i-j| = 1$ رابطه فوق را ساده کنیم به همان رابطه موجود در مقاله خواهیم رسید. یعنی:

$$\log\theta_{ij} = \begin{cases} \beta(2i+1)^2 & i = j \\ \beta(i+j)(2(\max(i,j)+1) & |i-j| = 1 \end{cases}$$

تشکر و قدردانی

نویسندگان این مقاله مراتب تشکر و قدردانی خود را نسبت به جناب آقای دکتر محمدرضا صفوی استادیار دانشکده دندان‌پزشکی دانشگاه علوم پزشکی شهید بهشتی و رئیس مرکز تحقیقات علوم دندان‌پزشکی، سرکارخانم دکتر مهتاب نوری استادیار دانشکده دندان‌پزشکی و معاون پژوهشی آن مرکز و همچنین آقای دکتر علی‌اکبر سفیدرودی فارغ‌التحصیل دکتری دندان‌پزشکی آن دانشگاه اعلام می‌دارند.

منابع

1. Perkins SM, Becker MP. Assessing rater agreement using marginal association models. Stat Med 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. May SM. Modelling observer agreement – an alternative to kappa. J Clin Epidemiol 1994; 44:1315-1324.
4. Cohen J. A coefficient of agreement for nominal scales. Education and Psychological Measures 1960; 20:37-46.
5. Light RJ. Measures of response agreement for qualitative data: some generalizations and alternatives. Psychological Bulletin 1971; 5:365-377.
6. Tanner MA, Young MA. Modelling agreement among raters. Journal of the American Statistical Association 1985; 80:175-180.
7. Goodman LA. Simple models for the analysis of association in cross-classifications having ordered categories. Journal of the American Statistical Association 1979; 74:537-552.
8. Agresti A. Categorical Data Analysis. Wiley: New York, 2002.
9. Agresti A. A model for agreement between ratings on an ordinal scale. Biometrics 1988; 44:539-548.
10. Becker MP, Agresti A. Log-linear modelling of pairwise interobserver agreement on a categorical scale. Stat Med 1992; 11:101-114.
11. Brook PH, Shaw WC. The development of an index for orthodontic treatment priority. Eur J Orthod 1989; 11:309-320.
12. Tarvit DJ, Freer TJ. Assessing malocclusion- the time factor. J Orthod 1998; 25:31-34
13. Richmond F, O'brien KD, Roberts CT, et al. Dentists variation in determination of Orthodontic Treatment Need. J Orthod 1994; 21:65-68.
14. Cooper S, Mandll NA, Dibiase D, et al. The reliability of the Index of Orthodontic Treatment Need over time. J Orthod 2000; 27:47-53.

15. Richmond F, Robetrs CT, Andrews M. Use of the Index of Orthodontic Treatment Need (IOTN) in assessing the need for Orthodontic Treatment pre- and- post-applience therapy. J Orthod 1994; 21:175-184.

16. Bentel MJ, Vig Wl, Shanker S, et al. Efficacy of training dental students in the Index of Orthodontic Treatment Need. American Journal of Orthodontic and Dentofacial Orthopedics 1996; 110(2):197-205.

17. Waring D, Jones JW. Does the GDP need to know about IOTN?. Dental Update 2003; 30:123-30.

18. Fleiss JL. The Design and Analysis of Clinical Experiments. Wiley: New York, 1999.