مقایسه مدل های حاشیه ای و انتقال برای تحلیل پاسخ های دو حالتی: یک مطالعه شبیه سازی

فرید زایری[1]، سوده شهسواری[2]، احمدرضا باغستانی[2]، سارا جام برسنگ[2*]،وحید لهرابیان[3]

1) مرکز تحقیقات پروتئومیکس، دانشکده پیراپزشکی، دانشگاه علوم پزشکی شهید بهشتی
2) گروه آمار زیستی، دانشکده پیراپزشکی، دانشگاه علوم پزشکی شهید بهشتی
3) گروه فیزیک پزشکی، دانشکده پزشکی، دانشگاه علوم پزشکی ایلام

تاریخ دریافت: **91/8/9** تاریخ پذیرش: **91/11/15**

چکیده

مقدمه: مطالعات طولی در بسیاری از شاخه های علوم، خصوصاً در علوم پزشکی بسیار رایج است. در دو دهه اخیر مدل های بسیار زیادی برای بررسی این گونه داده ها پیشنهاد شده است. این مدل ها هم چنین برای بررسی متغیرهای پاسخ مختلف از جمله ترتیبی و دوحالتی که در مطالعات پزشکی کاربرد بسیاری دارند، بسط داده شده‌اند. با این که تاکنون مقالات بسیاری به مقایسه مدل های حاشیه ای و اثرات تصادفی پرداخته اند، مطالعات کمتری در مورد مقایسه مدل های حاشیه ای و انتقال انجام گرفته است.

مواد و روش ها: در این مطالعه با استفاده از شبیه سازی با مقدارهای مختلف از همبستگی داده ها، دقت این دو مدل به وسیله محاسبه سطح زیر منحنی راکمورد مقایسه قرار گرفت و برتری هر یک از دو مدل در شرایط مختلف بیان گردید.

یافته های پژوهش: اگر افراد سالم با سرعت زیادی بیمار شوند(شدت انتقال از وضعیت سالم به بیمار) و مدت زمان بیشتری را در وضعیت بیماری بگذرانند، استفاده از مدل انتقال برای پیش بینی مناسب تر است. هم چنین در شرایطی که شدت انتقال بین دو وضعیت مطالعه یکسان باشد و یا روندی برعکس حالت قبل داشته باشد، به این معنی که افراد سالم زمان بیشتری را در وضعیت سلامتی بگذرانند و در صورت بیمار شدن با سرعت بیشتری بهبود یابند، استفاده از مدل حاشیه ای برای پیش بینی مناسب تر است.

بحث و نتیجه گیری: نتایج به دست آمده از مطالعه حاضر به پژوهشگر این امکان را می‌دهد که اگر هدف از مطالعه پیش بینی پاسخ باشد، علاوه بر در نظر گرفتن مطالب ذکر شده، میزان شدت انتقال بین وضعیت ها را نیز جهت انتخاب مدل، برای رسیدن به نتایج دقیق تر در نظر بگیرد.

واژه‌های کلیدی: داده های دوحالتی، همبستگی مارکوفی، مدل حاشیه ای، مدل انتقال، سطح زیر منحنی راک

* **نویسنده مسئول:** گروه آمار زیستی، دانشکده پیراپزشکی، دانشگاه علوم پزشکی شهید بهشتی

Email: s.jambarsang@sbmu.ac.ir

[Downloaded from sjimu.medilam.ac.ir on 2024-04-20]

مقدمه

امروزه مطالعات طولی در بسیاری از شاخه های علوم، خصوصاً در علـــوم پزشـــکی بسیار رایج است. داده ها در چنین مطالعاتی حاصل از اندازه گیری مکرر افراد یا واحدهای آزمایشی هستند. این نوع اندازه گیری واحدها سبب می شود داده های حاصل دارای همبستگی درونی باشند. به دلیل وجود همبستگی بین مقادیر به دست آمده، مدل های معمول در تحـــلیل داده ها کاربردی نیست و به مدل هایی نیاز است که این همبستگی درونی را در تحلیل داده ها و محاسبه پارامترها به حـــساب آورنـــد،(1). در دو دهه اخیر مدل های بسیار زیادی برای بررسی این گونه داده ها پیشنهاد شده است،(2). سه دسته اصلی این مدل ها، مدل های حاشیه ای، اثرات تصادفی و انتقال(مارکوف) هستند،(۳،۴). این مدل ها هم چنین برای بررسی متغیرهای پاسخ مختلف از جمله ترتیبی و دوحالتی که در مطالعات پزشکی کاربرد بسیاری دارند، بسط داده شده اند.(7-5)

مدل های ارایه شده برای تحلیل داده های طولی هر کدام به شکلی همبستگی داده ها را در برآورد پارامتر اعمال می‌کنند. امـــا ماهیت همبستگی درون داده ها نیز اشکـــال مختـــلفی مانـــند تـــبادل پذیر (exchangeable)، خودبازگشتی(مارکـــوف) یا بـدون ساختـــار(unstructured)دارد،(8). در ایـن میـــان همبستگی مارکوفی به دلیل درک عمومی آسان تر و تسهیل در امر برازش مدل و کاهش پارامترهای آن در اکثر مواقع، از مقبولیت زیادی برخوردار است. همبستگی مارکوفی به این معنی است که برآمدهای آزمایش در هر زمان به برآمدهای پیشین و در حالت مارکوف مرتبه اول فقط به برآمد پیشین آن وابسته است. یکی از مثال های رایج برای درک مفهوم مارکوف، برآورد احتمال ابری یا آفتابی بودن چند روز متوالی به شرط دانستن وضعیت هوا در روز اول است،(9). در این مقاله منظور از مدل مارکوف، مارکوف مرتبه اول است. از میان سه دسته کلی مدل های طولی که بیان شد، دو مدل حاشیه ای و انتقال قابلیت احتساب ویژگی مارکوف را در برآورد پارامترهای خود دارند.(3)

با این که تاکنـــون مقالات بسیاری به مقایسه مدل های حاشیه ای و اثرات تصادفی پرداخته اند، مطالعات کمتری در مورد مقایسه مدل های حاشیه ای و انتقال انجام گرفته است. لیانگ و زیگر در مطالعه ای ضمن بیان تفاوت های بنیادی این دو مدل از نظر مفهوم نشان دادند در اکثر مواقع قدر مطلق پارامترهای برآورد شده از مدل انتقال مقداری کوچک تر از پارامترهای حاصل از مدل حاشیه ای دارند،(10). ورنر و همکاران نیز این رابطه را برای مقادیر مختلف همبستگی بین متغیر مستقل در هر زمان و پاسخ در یک زمان قبل بررسی کردند.(11)

داده های دوحالتی سهم عمده ای از داده های حاصل از مطالعات پزشکی را تشکیل می‌دهند،(12). منحنی راک پرکاربـــردترین ابزار برای بررسی توان پیش بینی مدل های مرتبط با پاسخ های دوحالتی است. سطح زیر منحنی راک، شاخص مناسبی برای بیان این تـــوانایی است،(13). در این مقاله برآورد احتمال ها برای یک سری داده دوحالتی شبیه سازی شده، با استفاده از دو مدل حاشیه ای و انتقال به دست آمده و سپس از شاخص سطح زیر منحنی راک برای مقایسه دقت این دو مدل استفاده شده اســت. در بخش های بعدی مختصری از ماهیت مدل های حاشیه ای و انتـــقال مورد استفاده در این مطالعه و هم چنین داده های شبیه سازی شده به منظور مقایسه دو مدل بیان شده است. در انتها نیز نتایج شبیه سازی و بحث ارایه می شود.

مواد و روش ها

مدل حاشیه‌ای

در مدل حاشیه ای اثر متغیرهای تبیینی بر روی پاسخ و همبستگی درونی داده ها به صورت جداگانه مدل بندی می‌شوند،(14). در مدل حاشیه ای میانگین جامعه به عنوان تابعی از متغیرهای تبیینی به صورت، $h(\mu_{ij}) = X_{ij}^{'}\beta$ بیان می‌شود، که h در آن یک تابع پیوند نامیده شده و مثلاً وضعیتی که پاسخ ها دوحالتی هستند، می تواند تابع لوجیت باشد. در این مدل، برآورد پارامترها همبستگی درون موردی با استفاده از روش معادلات برآوردی تعمیم یافته(GEE) با در نظر گرفتن یک ماتریس همبستـــگی اعمـــال

[Downloaded from sjimu.medilam.ac.ir on 2024-04-20]

می شود،(3). در این مطالعه ماتریس همبستگی به شکلی در نظر گرفته می شود که دارای خاصیت مارکوف باشد.

در این حالت مدل به شکل زیر بیان می‌شود:

$$\log it \Pr(Y_{ij}=1)=\beta_0+\beta_1 t_{ij}$$

که در مطالعه حاضر t_{ij} زمان های مشاهده پاسخ در نظر گرفته شده است. ساختار همبستگی به صورت $\varepsilon_j=\alpha\,\varepsilon_{j-1}+z_j$ می‌باشد که zj متغیرهای تصادفی دو به دو مستقل و دارای توزیع N(0,σ2) هستند. هم چنین $\varepsilon_j \sim N(0,\Sigma_j)$ می‌باشد،(8)، که در آن

$$\Sigma_j=\sigma^2\begin{bmatrix}1 & \rho & \rho^2 & \cdots & \rho^{t-1}\\ \rho & 1 & \rho & \cdots & \rho^{t-2}\\ \rho^2 & \rho & 1 & \cdots & \rho^{t-3}\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ \rho^{t-1} & \rho^{t-2} & \rho^{t-3} & \cdots & 1\end{bmatrix}$$

مدل انتقال(مارکوف)

یک زنجیر مارکوف مرتبه اول به وسیله ماتریس احتمال انتقال آن مشخص می شود. یک ماتریس 2×2 که درایه های آن، π_{ab} نشان‌دهنده احتمال قرار گرفتن در وضعیت b در زمان t به شرط قرار داشتن در وضعیت a در زمان $t-1$ است. رابطه بین احتمال های انتقال و متغیرهای تبیینی با استفاده از یک مدل لوجستیک که حاصل ترکیب دو مدل لوجستیک برای هر یک از حالت های پاسخ در زمان $t-1$، به راحتی به شکل زیر قابل بیان است،(14):

$$\log it \Pr(Y_{ij}=1|Y_{ij-1}=y_{ij})=X_{ij}^{'}\beta_0+y_{ij-1}X_{ij}^{'}\alpha$$

در این مطالعه مدل با یک متغیر مستقل که زمان است به شکل زیر در نظر گرفته شده است:

$$\log it \Pr(Y_{ij}=1|Y_{ij-1}=y_{ij})=t_{ij}^{'}\beta+\alpha y_{ij-1}$$

منحنی مشخصه عملکرد (Receiving Operating Characteristics Curve)

منحنی مشخصه عملکرد(منحنی راک) یک ابزار تشخیصی بسیار مفید خصوصاً در علوم پزشکی است و رابطه بین حساسیت و ویژگی را در یک آزمون بررسی می کند. برآوردهای بر مبنای مدل منحنی راک به این دلیل که فرصت ارزیابی اثر متغیرهای پیش بینی را روی دقت مدل های آماری ایجاد می کند بسیار مفید هستند،(13). این منحنی برای ارزیابی دقت و توان پیش بینی آزمون های تشخیصی و مدل های آماری ملاک های متعددی از جمله مساحت زیر منحنی (AUC)، طول تصویر منحنی(PLC) و ناحیه پیچ خورده منحنی(ASC) دارد،(15). در بین این ملاک ها مساحت زیر منحنی راک رایج تر است. این اندازه مقداری بین صفر و یک را می‌گیرد و مقادیر نزدیک یک دقت بالاتر را نشان می‌دهند.(13)

شبیه سازی

برای بررسی این مطلب که در صورت برقراری خاصیت مارکوف در داده ها، کدام یک از دو مدل حاشیه ای یا انتقال که قابلیت احتساب این ویژگی را دارند، برازش بهتری دارند، یک مطالعه شبیه سازی به این شرح طراحی شد. با استفاده از بسته نرم افزاری msm،(16)، یک سری داده دودویی با همبستگی مارکوفی با 3 نوع ماتریس شدت انتقال شبیه سازی شد. این داده ها با تکرارهای زمانی 3، 5،10 و 15 تایی و در حجم نمونه های 20 و 50 تایی تولید شدند و برای هر ترکیب از این تکرار، حجم نمونه و ماتریس شدت انتقال، شبیه سازی 1000 بار تکرار شد.

برای مقایسه قابلیت رده بندی صحیح دو مدل در هر ترکیب شبیه سازی شده به وسیله دو مدل ذکر شده احتمال های برآورد و سطح زیر منحنی راک محاسبه گردید،(17). در انتها میانگین دو سطح زیر منحنی به عنوان معیار مقایسه دو مدل در نظر گرفته شد.

یافته های پژوهش

هدف از انجام این مطالعه ارزیابی دو مدل مارکف و حاشیه ای با همبستگی مارکف مرتبه اول وقتی خاصیت مارکفی در داده ها برقرار است بود. بدین منظور داده هایی با تکرارهای متفاوت و ماتریس های شدت انتقال مختلف تولید شد که پس از مدل بندی، با استفاده از سطح زیر منحنی راک دقت دو مدل مورد ارزیابی قرار گرفت.

همان‌گونه که از جدول شماره 1 مشاهده می‌شود m تعداد موارد شامل شده در مطالعه را نشان می‌دهد که برای آن با دو مقدار 20 و 50 شبیه سازی اجرا شد. به علاوه تعداد اندازه‌گیری های تکرار شده

[Downloaded from sjimu.medilam.ac.ir on 2024-04-20]

متغیر پاسخ برای هر مورد با n نمایش داده شده است که در این مطالعه تعداد تکرارها ۳،۵،۱۵،۱۰ در نظر گرفته شد. هر کدام از اعداد جدول نشان دهنده مساحت زیـــر منحنی راک است که مقادیر بالاتر نشان دهنده دقت بالاتر در پیش بینی مدل است. برای مثال مقدار 0/8473 نشان می دهد با سه اندازه گیری تکرار شده برای پاسخ و برازش مدل انتقال با ماتریس شدت انتقال اول دقت پیش بینی خوبی وجود دارد در حالی که با همین مدل در حالی که ماتریس انتقال دوم استفاده شده و تعداد تکرارها زیاد است دقت پیش بینی 0/5750 با توجه به این که نزدیک 0/5 است مناسب نیست.

زمانی که ماتریس شدت انتقال به صورت 1 تعریف شده است که در آن افراد با شدت کمتری در وضعیت سلامت باقی می‌مانند و در عوض تغییر وضعیت فرد سالم به بیمار با شدت بیشتری رخ می‌دهد، وقتی تکرارها زیاد است(10 و 15) مدل انتقال دقت بیشتری را به خود اختصاص داده است و در تکرارهای کم اختلاف دقت در دو مدل ناچیز است. به علاوه وقتی افراد با احتمال های یکسانی در وضعیت خود باقی می مانند یا تغییر وضعیت می‌دهند معمولاً مدل حاشیه ای بهـــتر عمل می کـــند، خصـــوصاً زمـــانی که تعداد تکرارها کم باشد. در حالتی برعکـــس 1 که فرد با احتمال کمتری در وضعیت خـــود بـــاقی می ماند و با شدت بیـــشتری تمایـــل به تغییر وضعـــیت دارد نـــیز مـــدل حاشـــیه ای بر مـــدل انتقال برتـــری دارد. هـــم چنین نـــتایج مشاهده شـــده بـــرای حالت هـــای مخـــتلف تعـــداد تکـــرارهـــا و ماتریـــس های شـــدت انتقال در هر دو نمونه 20 و 50 تایی مشابه بود.

جدول شماره 1. برآورد سطح زیر منحنی راک برای مدل های حاشیه ای و انتقال

ماتریس انتقال		مدل	تعداد بیماران=20				تعداد بیماران=50			
			تعداد تکرار ها برای هر فرد				تعداد تکرار ها برای هر فرد			
			3	5	10	15	3	5	10	15
0/85	0/15	انتقال	0/8473	0/7989	0/7197	0/6795	0/8502	0/7953	0/7140	0/6740
750	0/25	حاشیه ای	/8465	0/7790	0/6833	0/6333	0/8471	0/7788	0/6820	0/6337
0/50	0/50	انتقال	0/6251	0/6209	0/5901	0/5750	0/6582	0/6114	0/5813	0/5682
0/50	0/50	حاشیه ای	0/7666	0/6903	0/6053	0/5729	0/7660	0/6900	0/6085	0/5737
0/80	0/20	انتقال	0/6165	0/6122	0/5674	0/5536	0/6425	0/6068	0/5654	0/5451
0/20	0/80	حاشیه ای	0/7563	0/6749	0/5923	0/5645	0/7578	0/6754	0/5951	0/5651

بحث و نتیجه گیری

اصطلاح حاشیه ای به این معنی است که میانگین پاسخ در هر موقعیت تنها به مقدار متغیر همراه مورد نظر وابسته است و نه به اثرات تصادفی یا پاسخ های پیشین، در صورتی که در مدل انتقال و به طور کلی مدل هـــای شرطـــی(مارکـــوف) مـــیانگین پاسخ به پاسخ هـــای قبـــلی نیز وابسته است. استفاده از مدل های حاشیـــه ای راه ساده تـــری برای بسط مدل های خطی تعمیم یافته به مدل طولی است. این روش میانگین پاسخ در هر موقعیت، $E(Y_{ij} \mid X_{ij})$ را به راحتی با یک تابع ربط مدل بندی می کند. به دلیل این‌که تاکید اصلی بر روی میانگین پاسخ و ارتباط آن با متغیر همراه است، مدل حاشیه ای نیازی به مشخص بودن توزیع کامل بردار پاسخ های تکرار شده ندارد،(1). این ویژگی خصوصاً در پاسخ های دو و چندحالتی بسیار راه‌گشا است. با این حال یک فرضیه مبهم در این مدل اغلب نادیده گرفته می‌شود. در مدل های حاشیه ای فرض می‌شود میانگین شرطی j امین پاسخ به شرط $x_{i1},...,x_{in}$ تنها به x_{ij} به شکل زیر وابسته است:

$$E(Y_{ij} \mid X_i) = E(Y_{ij} \mid X_{i1},..., X_{in}) = E(Y_{ij} \mid X_{ij})$$

در متغیرهای همـــراهی که در طول زمـــان تغییر نمی کنند، این فرضیه ضرورتاً حفظ می‌شود. هم چنین

[Downloaded from sjimu.medilam.ac.ir on 2024-04-20]

برای متغیرهایی که در طول زمان تغییر می‌کنند اما بر اساس طراحی مطالعه ثابت در نظر گرفته شده‌اند این وضعیت برقرار است. با این حال برای متغیرهای همراه غیرثابت در زمان که به صورت تصادفی در طول مطالعه تغییر کنند، این فرضیه نقض می شود. به این ترتیب در شرایطی که متغیرهای همراه غیرثابت در طول زمان وجود دارد، با ملاحظه بیشتری باید از یک مدل حاشیه ای استفاده شود.(18)

همان طور که پیش از این نیز ذکر شد در مدل حاشیه ای چون مدل بندی میانگین پاسخ به صورت جداگانه از همبستگی درونی داده ها انجام می گیرد، این شکل از مدل بندی نقش مهمی در تعبیر پارامترهای مدل دارد. پارامترها در این مدل چگونگی تغییرات میانگین پاسخ در طول زمان و نوع ارتباط این تغییرات با متغیرهای همراه را توصیف می‌کند. این نکته قابل توجه است که چون برآوردها در این مدل نسبت به ساختار کواریانس استوار(Robust) هستند تعبیر پارامترهای مدل به فرضیات در مورد ماهیت و مقدار همبستگی درونی داده ها بستگی ندارد.(14)

مدل های انتقال مدل سازی را به طور هم زمان روی میانگین و همبستگی درونی انجام می‌دهند. این کار به وسیله شرطی کردن هر برآمد روی بقیه برآمدها و یا مجموعه ای از بقیه برآمدها، صورت می‌گیرد. به دلیل ماهیت دنباله ای داده های طولی استفاده از مدل های انتقال مناسب به نظر می رسد. در مدل های انتقال توزیع شرطی هر پاسخ به صورت تابعی معلوم از پاسخ های پیشین و متغیرهای همراه بیان می شود.(18)

اگرچه مدل های مارکوف و خودبازگشتی پیشینه ای طولانی و گسترده از نظر استفاده در داده های سری زمانی دارند، کاربرد آن ها در داده های طولی با محدودیت هایی همراه است. برخی جنبه های مدل های انتقال مناسب بودن آن ها را در مدل سازی داده های طولی تحت تاثیر قرار می‌دهد. از جمله محدودیت ها می‌توان به این موارد اشاره کرد، به طور کلی مدل های انتقال برای داده های تکراری که در بازه های زمانی یکسان اندازه گیری شده اند بسط داده شده است. هم چنین کاربرد آن ها در موقعیت هایی که داده های گم شده، زمان های گم شده و اندازه گیری های نامنظم و فاصله های غیریکسان بین اندازه گیری ها وجود داشته باشد، بسیار مشکل است،(3). علاوه بر آن تعبیر ضرایب رگرسیونی با تغییر ترتیب همبستگی سریالی تغییر می کند. در نهایت ممکن است در بسیاری از مطالعات این ضرایب هدف معمول را دنبال نکنند، چون وابستگی به داده های قبلی می‌تواند سبب شود اثر متغیرهای مورد علاقه کم برآورد شود، در واقع دیدن پاسخ ها را به شرط پاسخ های قبلی در شرایطی که انتظار داریم متغیر همراه بر مقدار میانگین پاسخ در تمام موقعیت ها اثر داشته باشد، سبب شود این اثر نادیده گرفته شود.(18)

در این مطالعه که هدف مقایسه مدل حاشیه ای و انتقال بود نتایج زیر به دست آمد:

وضعیت اول حالتی است که ماتریس شدت انتقال به گونه ای است که فرد با شدت کمتری تمایل دارد در وضعیت سالم قرار بگیرد یعنی شدت انتقال از وضعیت بیمار به سالم 0/25 و ماندن در وضعیت سالم 0/15 است و در عوض تمایل به بودن در وضعیت بیماری با شدت بیشتری رخ می دهد که فرد بیمار با شدت انتقال 0/75 تمایل دارد در همان وضعیت باقی بماند و فرد سالم با شدت انتقال 0/85 به وضعیت بیماری می رود. در این حالت تقریباً در کل شرایط های مطالعه برازش مدل انتقال مناسب تر است البته زمانی که تعداد تکرارها زیاد است اختلاف دقت به نفع مدل انتقال بیشتر می‌شود.

وضعیت دوم حالتی است که ماتریس شدت انتقال نشان می‌دهد تمایل فرد برای تغییر وضعیت به وضعیت دیگر یا ماندن در وضعیت کنونی برابر 0/5 است و شدت انتقال در کل وضعیت ها یکسان است. در این حالت مدل حاشیه ای همواره بهتر عمل می کند و اختلاف دقت با تکرارهای اندازه گیری کمتر به نفع مدل حاشیه‌ای بیشتر است.

وضعیت سوم ماتریس شدت انتقال حالتی را نشان می‌دهد که فرد با شدت بیشتری 0/8 تمایل به تغییر وضعیت دارد در حالی که تمایل به ماندن در وضعیت کنونی با شدت کمتر و برابر 0/2 است. در این حالت اختلاف دقت وقتی اندازه گیری های تکرار شده برای

[Downloaded from sjimu.medilam.ac.ir on 2024-04-20]

پاسخ کم است به نفع مدل حاشیه ای است در حالی که با تکرارهای بیشتر دقت مدل انتقال بیشتر است.

مزایا و معایب ذکر شده در بالا و هم چنین هدف محقق از برازش مدل باعث انتخاب یکی از دو مدل انتقال یا حاشیه ای در داده های طولی می‌شود. نتایج به دست آمده از مطالعه حاضر به پژوهشگر این امکان را می دهد که اگر هدف از مطالعه پیش بینی پاسخ باشد، علاوه بر در نظر گرفتن مطالب ذکر شده، میزان شدت انتقال بین وضعیت ها را نیز جهت انتخاب مدل، برای رسیدن به نتایج دقیق تر در نظر بگیرد. به این ترتیب در مواردی که شدت انتقال افراد از وضعیت اول(صفر، سالم بودن،...) به وضعیت دوم(یک، بیمار شدن، ...) زیاد و شدت انتقال به عکس این وضعیت کم باشد و یا به عبارت بهتر افراد سالم با سرعت زیادی بیمار شوند(شدت انتقال از وضعیت سالم به بیمار) و مدت زمان بیشتری را در وضعیت بیماری بگذرانند، استفاده از مدل انتقال برای پیش بینی مناسب تر است. هم چنین در شرایطی که شدت انتقال بین دو وضعیت مطالعه یکسان باشد و یا روندی برعکس حالت قبل داشته باشد، به این معنی که افراد سالم زمان بیشتری را در وضعیت سلامتی بگذرانند و در صورت بیمار شدن با سرعت بیشتری بهبود یابند، استفاده از مدل حاشیه ای برای پیش بینی مناسب تر است.

References

1-Parzen M, Ghosh S, Lipsitz S. A generalized linear mixed model for longitudinal binary data with a marginal logit link function. Ann Appl Stat 2011;5(1):449-67.

2-Liu I, Agresti A. The Analysis of ordered categorical data. An overview and a survey of recent developments. TEST 2005;14:1-73.

3-Diggle PJ, Heagerty P, Liang KY, Zeger SL, editors. Analysis of longitudinal data. 2th ed. Oxford University Press; 2004.

4-Agresti A, editors. Analysis of ordinal categorical data. 2th ed. New York: Wiley; 2010.

5-Heagerty P. Marginalized transition models and likelihood inference for longitudinal categorical data. Biometrics 2002;58:342-51.

6-Heagerty P. Marginally specified logisic-normal models for longitudinal binary data. Biometrics 1999;55:247-57.

7-Agresti A, Lang J. A proportional odds model with subject-specific effect for repeated ordered categorical responses. Biometrika 1993;80:527-34.

8-Fitzmaurice GM, Laird NM, Ware JH, editors. Applied Longitudinal Analysis. 2th ed. John Wiley & sons; 2011.

9-Rabiner LR. A Tutorial on hidden markov models and selected applications in speech recognition. Proc IEEE 1989;77:257-86.

10-Zeger SL, Liang KY. An overview of methods for the analysis of longitudinal data. Stat Med 1992;11:1825-39.

11-Sullivan PM, Andersona GL. A cautionary note on inference for marginal regression models with longitudinal data and general correlated response data. Commun Stat Simul Comput 1994;23:939-51.

12-Hu FB, Goldberg J, Hedeker D, Flay BR, Pentz MA. Comparison of population-averaged and subject-specific approaches for analyzing repeated binary outcomes. Am J Epidemiol 1998;147:694-703.

13-Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 1982;143:29-36.

14-Rezaei Ghahroodi Z, Ganjali M. Models for Longitudinal analysis of binary response data for identifying the effects of different treatments on insomnia. Appl Math Sci 2010;4:3067-82.

15-Lee WC, Hsiao CK. Alternative Summary indices for the receiver operating characteristic curve. Epidemiology 1996;7: 605-11.

16-Titman AC, Sharples LD. Model diagnostics for multi-state models. Stat Methods Med Res 2010;19:621-51.

17-Honghu Liu GL, Cumberland WG, Wu T. Testing Statistical Significance of the Area under a Receiving Operating Characteristics Curve for Repeated Measures Design with Bootstrapping. J Data Sci 2005;3:257-78.

18-Fitzmaurice G, Molenberghs G. Advances in longitudinal data analysis. An historical perspective. In: Longitudinal Data Analysis. Chapman & Hall/CRC; 2008.

[Downloaded from sjimu.medilam.ac.ir on 2024-04-20]

Comparing Marginal and Transition Models In The Analysis of Binary Longitudinal Data: a Simulation Study

Zayeri F[1], Shahsavari S[2], Baghestani A.R[2], Jambarsang S[2], lohrabian V[3]*

(Received: 30 Oct. 2012 Accepted: 3 Feb. 2013)

Abstract

Introduction: Longitudinal studies widely used in many branches of science, especially medical science. In the past two decades, many models have been proposed for the evaluation of such data. These models are also developed to investigate various response variables such that categorical variables that have many applications in medical research. While many papers have compared the marginal and random effects models but fewer studies have compared the marginal and transition models.

Materials & Methods: This paper used simulation with different scenarios for correlation, sample size and repeated measures and calculated ROC Curve for two models. Then estimated AUC for them and comparison was performed.

Findings: If intensity transition for healthy to disease was large and patient remained disease for a long time that is better used transition model to prediction. In addition, in conditions that intensity transition in any status was equal or trends of transition intensity was inverse of position1 that healthy people remained health for a long time and patients recovered quickly, that is better used marginal model.

Discussion & Conclusion: In analysis of longitudinal data for achieving more accurate results when purpose of study was prediction, that is better intensity transition between status was considered and then decided which model is choosen.

Keywords: binary data, markov correlation, marginal model, transition model, roc area under curve

1. Proteomics Research Center, Faculty of Paramedical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran
2. Dept of Biostatistics, Faculty of Paramedical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran
3. Dept of Medical Physics, Faculty of Medicine, Ilam University of Medical Sciences, Ilam, Iran
**(corresponding author)*

[Downloaded from sjimu.medilam.ac.ir on 2024-04-20]