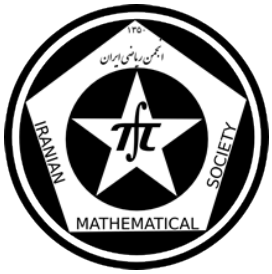

بسم الله الرحمن الرحیم

به نام خدا



اولین همایش ملی ریاضیات زیستی

کتابچه مقالات

۲۱ و ۲۲ اسفند ماه ۱۳۹۷

دانشگاه نیشابور

پیش‌گفتار

با یاری خداوند متعال توفیق اجرای اولین همایش ملی ریاضیات زیستی را با همت ستودنی اساتید و دانشجویان و حمایت گروه ریاضی و آمار و همچنین مسئولین محترم دانشگاه نیشابور و سازمان‌ها و دستگاه‌های مختلف کشور در روزهای ۲۱ و ۲۲ اسفند ۱۳۹۷ در دانشگاه نیشابور یافتیم. تا با ارایه آخرین دست‌آوردهای علمی، بحث و تبادل نظر بین دانشگاهیان، پژوهش‌گران و کارشناسان در این زمینه، زمینه‌ساز ارتباط قوی‌تر بین دانشگاهیان وزارت های علوم، تحقیقات و فناوری و بهداشت، درمان و آموزش پزشکی شده تا باعث کیفیت مطلوب تر زندگی مردم گردیم.

در این همایش، ۳ سخنرانی کلیدی با عناوین نحوه بهینه تجویز داروی تاموکسیفن در مهار متاستاز استخوان، مدل سازی پیامد بیماران آلوده به ویروس HIV به کمک سیستم های معادلات دیفرانسیل و آنالیز تک سلولی ها توسط اساتید مدعو در نظر گرفته شد. متقاضیان شرکت در کنفرانس تعداد ۶۲ مقاله به دبیرخانه ارسال نمودند که کمیته علمی کنفرانس به همراه داوران محترم تعداد ۲۴ مقاله را به‌عنوان سخنرانی، ۲۷ مقاله را برای ارایه به صورت پوستر و ۱ کارگاه آموزشی را پذیرفت.

همکاری صمیمانه اساتید دانشگاه‌های سراسر کشور بویژه اعضای کمیته های اجرایی و علمی را قدر می نهیم. همکاری این عزیزان در ارایه پیشنهادهای سازنده، کمک شایانی به هر چه بهتر اجرا شدن همایش و متون علمی این کتابچه بوده است. در پایان، کمیته علمی کنفرانس از تلاش‌های بی‌دریغ و زحمات صادقانه همه مسوولین، همکاران هیات علمی و دانشجویان گروه ریاضی و آمار دانشکده علوم پایه، و کارکنان بخش‌های مختلف دانشگاه نیشابور، تشکر و قدردانی می‌کند.

کمیته برگزاری اولین همایش ملی ریاضیات زیستی
دانشگاه نیشابور
۲۱ و ۲۲ اسفند ماه ۱۳۹۷

اعضای کمیته اجرایی

۱. دکتر رضا محمدی (دبیر کمیته اجرایی)	دانشگاه نیشابور
۲. دکتر مژگان افخمی گلی	دانشگاه نیشابور
۳. دکتر احسان انجیدنی	دانشگاه نیشابور
۴. دکتر بهاره خطیب آستانه	دانشگاه نیشابور
۵. دکتر محمد حسین رحمانی دوست	دانشگاه نیشابور
۶. دکتر محمد شیرازیان	دانشگاه نیشابور
۷. دکتر سید محسن صالح	دانشگاه نیشابور
۸. دکتر سید مهدی صالحی	دانشگاه نیشابور
۹. دکتر داوود عمارتی مقدم	دانشگاه نیشابور

اعضای کمیته علمی

فهرست مقالات

عنوان ______ صفحه

اولین همایش ملی ریاضیات زیستی
۲۱ و ۲۲ اسفندماه ۱۳۹۷
همایش ملی ریاضیات زیستی

مدلسازی ریاضی زیست‌حسگر گلوکز و حل عددی با استفاده از روش هم‌مکانی توابع پایه شعاعی

آمنه طالعی [۱]، **دانشکده ریاضی، دانشگاه صنعتی شیراز، شیراز، ایران**

چکیده: زیست‌حسگر گلوکز از اولین و پرکاربردترین زیست‌حسگرهای موجود در بازار است، که مدل ریاضی آن را می‌توان به صورت معادلات دیفرانسیل پاره‌ای بیان کرد. در این مقاله قصد داریم به بررسی حل عددی معادله نفوذ- واکنش زیست‌حسگر بپردازیم. برای حل عددی، از روش هم‌مکانی با استفاده از توابع پایه شعاعی برای گسسته‌سازی متغیر متناظر با غلظت گلوکز مایعاتی مانند خون و غلظت محصولات حاصل از واکنش استفاده می کنیم. همچنین متغیر زمان را به روش تفاضل‌متناهی پسرو گسسته‌سازی می‌کنیم.

واژه‌های کلیدی: زیست‌حسگر، روش توابع پایه شعاعی، معادله نفوذ- واکنش.

۱ مقدمه

زیست‌حسگر نخستین‌بار توسط پروفسور لیلاند. سی. کلارک در سال ۱۹۵۶ معرفی شد و پس از آن در سال ۱۹۷۵ اولین زیست‌حسگر تجاری که برای اندازه‌گیری گلوکز به روش آمپرومتری بود، معرفی شد (۴). گلوکز با آنزیم گلوکزاکسیداز در فرایند اکسیداسیون کاتالیستی به اکسولاکتون و آب‌اکسیژنه تبدیل می‌شود. از طریق فرایند تجزیه‌ی آب‌اکسیژنه حاصل از این واکنش برروی سطح الکترود و انتقال الکترونی حاصل از آن، سیگنالی در مدار ایجاد می‌شود. زیست‌حسگر آمپرومتری وسیله‌ای است که این سیگنالها را اندازه‌گیری می‌کند و در نتیجه میزان گلوکز مایعاتی مانند خون مشخص می‌شود. این عملکرد را می‌توان به صورت واکنش زیر نشان داد:

$$E + S \leftrightarrows ES \rightarrow E + P$$

که E، S، ES وPبه ترتیب متناظر با آنزیم گلوکزاکسیداز، گلوکز خون(سوبسترا)، کمپلکس آنزیم-سوبسترا و محصول حاصل از واکنش است. با افزایش جمعیت و درگیرشدن افراد با بیماری‌های مختلف، اهمیت استفاده از زیست‌حسگرها در تشخیص بیماری به دلیل تشخیص بیولوژیکی بسیار خاص و پایدار و قابل‌اعتماد آنها، مشخص می‌شود. شبیه‌سازی عددی فرصت مطالعه‌ی ویژگی‌های زیست‌حسگر گلوکز بدون نیاز به انجام آزمایش‌های پرهزینه و زمان‌بر را فراهم می‌کند.

مدل ریاضی زیست‌حسگر اغلب به صورت معادلات دیفرانسیل غیرخطی است که به ندرت جواب دقیق برای آن وجود دارد. بنابراین روش‌های عددی برای حل اینگونه معادلات بسیار موثر خواهد بود. در این مقاله به حل عددی مدل ریاضی یک بعدی برای زیست‌حسگر آمپرومتری گلوکز بر پایه‌ی آنزیم گلوکزاکسیداز می پردازیم. روش هم‌مکانی با استفاده از توابع پایه‌شعاعی برای گسسته‌سازی غلظت گلوکز مایعاتی مانند خون و غلظت محصولات حاصل از واکنش و روش تفاضل‌متناهی پسرو برای گسسته‌سازی متغیر زمان پیشنهاد می‌شود.

[۱] آمنه طالعی: a.taleei@sutech.ac.ir

۲ مدل ریاضی زیست‌حسگر

مدل ریاضی زیست‌حسگر شامل معادلات نفوذ-واکنش با جمله‌ی غیرخطی طرح میکائیلیس-منتن است (۳). مطابق قانون فیک، اتصال واکنش کاتالیزشده با آنزیم در لایه‌ی آنزیم گلوکزاکسیداز با نفوذ در فضای یک بعدی تشریح شده و به معادلات ذیل منجر می‌شود:

$$\frac{\partial \hat{S}}{\partial \hat{t}} = D_{\hat{S}} \frac{\partial^۲ \hat{S}}{\partial \hat{x}^۲} - \frac{V_{max} \hat{S}}{K_M + \hat{S}}, \tag{۱.۲}$$

$$\frac{\partial \hat{P}}{\partial \hat{t}} = D_{\hat{P}} \frac{\partial^۲ \hat{P}}{\partial \hat{x}^۲} - \frac{V_{max} \hat{S}}{K_M + \hat{S}}, \tag{۲.۲}$$

که $\hat{x}$ و $\hat{t}$ به ترتیب فاصله و زمان را نشان می‌دهند. $D_{\hat{S}}$ و $D_{\hat{P}}$ ضرایب انتشار هستند. V_{max}، حداکثر سرعت آنزیم و K_M، ثابت میکائیلیس است. ثابت میکائیلیس غلظت سوبسترایی است که در آن نصف حداکثر چگالی واکنش کاتالیز شده با آنزیم به دست آمده است. شرایط اولیه و مرزی را می‌توان به فرم زیر در نظر گرفت:

$$\hat{S}(\hat{x}, ۰) = ۰, \quad \hat{x} \in [۰, d), \quad \hat{S}(d, ۰) = \hat{S}_۰, \tag{۳.۲}$$

$$\hat{P}(\hat{x}, ۰) = ۰, \quad \hat{x} \in [۰, d), \quad \hat{P}(d, ۰) = \hat{P}_۰, \tag{۴.۲}$$

$$\frac{\partial \hat{S}}{\partial \hat{x}}(۰, \hat{t}) = ۰, \quad \hat{P}(۰, \hat{t}) = ۰, \quad \hat{t} > ۰, \tag{۵.۲}$$

$$\hat{S}(d, \hat{t}) = \hat{S}_۰, \quad \hat{P}(d, \hat{t}) = \hat{P}_۰, \quad \hat{t} > ۰. \tag{۶.۲}$$

جریان سنجش شده به عنوان پاسخ زیست‌حسگر که به شارش ماده‌ی الکترواکتیو در سطح الکترود وابسته است به طور صریح از قوانین فیک و فارادی برحسب اکسایش و یاکاهش $\hat{P}$ به دست می‌آید (۱):

$$\hat{i} = n_e F D_{\hat{P}} \frac{\partial \hat{P}}{\partial \hat{x}}(۰, \hat{t}), \tag{۷.۲}$$

که n_e تعداد الکترون‌های وارده در انتقال شار و F ثابت فارادی است.

برای تعیین پارامترهای اصلی حاکم بر مدل ریاضی (۱)-(۷)، متغیرهای بعد دار $\hat{x}$ و $\hat{t}$ و غلظت‌های مجهول $\hat{S}$ و $\hat{P}$ را با پارامترهای بدون بعد زیر جایگزین می‌کنیم:

$$S = \frac{\hat{S}}{\hat{S}_۰}, \; P = \frac{\hat{P}}{\hat{S}_۰}, \; x = \frac{\hat{x}}{d}, \; t = \frac{D_{\hat{S}} \hat{t}}{d^۲}, \; r = \frac{D_{\hat{p}}}{D_{\hat{S}}}, \; \mu = \frac{v_m d^۲}{D_{\hat{S}} \hat{S}_۰}, \; k = \frac{k_m}{\hat{S}_۰}. \tag{۸.۲}$$

معادلات (۱) و (۲) به صورت زیر نوشته می‌شوند:

$$\frac{\partial S}{\partial t} = \frac{\partial^۲ S}{\partial x^۲} - \frac{\mu S}{k + S}, \tag{۹.۲}$$

$$\frac{\partial P}{\partial t} = r \frac{\partial^۲ P}{\partial x^۲} + \frac{\mu S}{k + S}. \tag{۱۰.۲}$$

شرایط اولیه و مرزی مدل بی بعد به صورت زیر است:

$$S(x,۰)=۰,\quad P(x,۰)=۰,\quad x\in[۰,۱), \tag{۱۱.۲}$$

$$\frac{\partial S}{\partial x}(۰,t)=۰,\quad P(۰,t)=۰,\quad t>۰, \tag{۱۲.۲}$$

$$S(۱,t)=۱,\quad P(۱,t)=۰,\quad t\geq ۰, \tag{۱۳.۲}$$

همچنین چگالی جریان بی‌بعد به صورت زیر می‌باشد:

$$\frac{\hat{i}d}{n_eFD_{\hat{P}}\hat{S}_۰}=\frac{\partial P}{\partial x}(۰,t). \tag{۱۴.۲}$$

درصورتی که غلظت گلوکز در لایه‌ی آنزیمی خیلی کمتر از ثابت میکائیلیس شود ($S\ll k$)، تابع غیرخطی میکائیلیس –منتن به صورت مرتبه اول ساده می‌شود و معادلات (۹) و (۱۰) به صورت زیر بدست می‌آید:

$$\frac{\partial S}{\partial t}=\frac{\partial^۲ S}{\partial x^۲}-\frac{\mu S}{k}, \tag{۱۵.۲}$$

$$\frac{\partial P}{\partial t}=r\frac{\partial^۲ P}{\partial x^۲}+\frac{\mu S}{k}. \tag{۱۶.۲}$$

۳ اعمال روش پیشنهادی

در این مقاله، گسسته‌سازی متغیر زمان به روش تفاضلات متناهی پسرو و گسسته‌سازی متغیر مکان به روش هم‌مکانی بر پایه‌ی توابع پایه شعاعی در نظر گرفته می‌شود. با انتخاب N گره در بازه $[۰,۱]$، توابع مجهول Pو S و مشتقات آن‌ها را در هر گام زمانی درونیابی می‌کنیم. انواع متعددی از توابع شعاعی وجود دارد. در این مقاله تابع چندمربعی تحت پارامتر شکل c، را به عنوان تابع شعاعی انتخاب می‌کنیم(۲). با قرار دادن تقریب تابع و مشتقات آن در معادلات دیفرانسیل مفروض و با استفاده از روش هم مکانی، به یک دستگاه معادلات جبری دست می‌یابیم.

با فرض Δt به عنوان گام زمانی، مشتق متغیر زمان را در معادلات (۱۵) و (۱۶) به صورت $\frac{S^{n+۱}-S^n}{\Delta t}$ تقریب می‌زنیم و با به کاربردن تقریب $S^{n+۱}(x)=\sum_{j=۱}^{N}{\lambda_j}^{n+۱}\phi(r_j)$ و $P^{n+۱}(x)=\sum_{j=۱}^{N}{\alpha_j}^{n+۱}\phi(r_j)$ داریم:

$$\sum_{j=۱}^{N}{\lambda_j}^{n+۱}[-\Delta t\phi_{xx}(r_{ij})+(۱+\Delta t\frac{\mu}{k})\phi(r_{ij})]=\sum_{j=۱}^{N}{\lambda_j}^{n}\phi(r_{ij}), \tag{۱.۳}$$

$$\sum_{j=۱}^{N}{\alpha_j}^{n+۱}[-r\Delta t\phi_{xx}(r_{ij})+\phi(r_{ij})]-\Delta t\frac{\mu}{k}\sum_{j=۱}^{N}{\lambda_j}^{n+۱}\phi(r_{ij})=\sum_{j=۱}^{N}{\alpha_j}^{n}\phi(r_{ij}). \tag{۲.۳}$$

با اعمال شرایط مرزی به دستگاه $L_۱Y^{n+۱}=L_۲Y^n+C$ می‌رسیم به‌طوریکه:

$$L_۱=\begin{bmatrix}A & O\\ M & D\end{bmatrix};\ L_۲=\begin{bmatrix}B & O\\ O & B\end{bmatrix};\ Y^k=\begin{bmatrix}\lambda;\ \alpha\end{bmatrix}^T,\ k=n,n+۱,$$

$$C(۱:N-۱)=۰;\ C(N)=۱;\ C(N+۱:۲N)=۰,$$

$$A=\begin{bmatrix} \phi_x(r_{11}) & \dots & \phi_x(r_{1N}) \\ -\Delta t\phi_{xx}(r_{21})+(1+a\Delta t)\phi(r_{21}) & \dots & -\Delta t\phi_{xx}(r_{2N})+(1+a\Delta t)\phi(r_{2N}) \\ \vdots & \vdots & \vdots \\ -\Delta t\phi_{xx}(r_{(N-1)1})+(1+a\Delta t)\phi(r_{(N-1)1}) & \dots & -\Delta t\phi_{xx}(r_{(N-1)N})+(1+a\Delta t)\phi(r_{(N-1)N}) \\ \phi(r_{N1}) & \dots & \phi(r_{NN}) \end{bmatrix},$$

$$D=\begin{bmatrix} \phi_x(r_{11}) & \dots & \phi_x(r_{1N}) \\ -r\Delta t\phi_{xx}(r_{21})+\phi(r_{21}) & \dots & -r\Delta t\phi_{xx}(r_{2N})+\phi(r_{2N}) \\ \vdots & \vdots & \vdots \\ -r\Delta t\phi_{xx}(r_{(N-1)1})+\phi(r_{(N-1)1}) & \dots & -r\Delta t\phi_{xx}(r_{(N-1)N})+\phi(r_{(N-1)N}) \\ \phi_x(r_{N1}) & \dots & \phi_x(r_{NN}) \end{bmatrix},$$

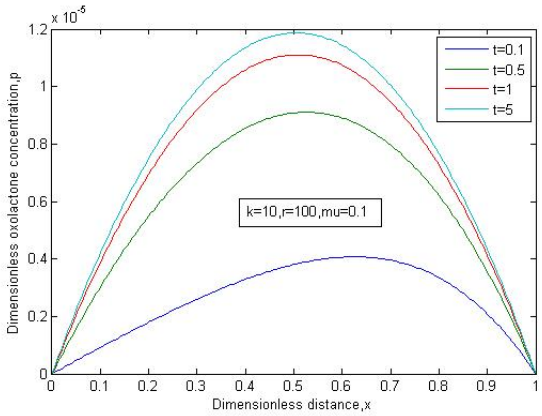
$$M=\begin{bmatrix} 0 & \dots & 0 \\ -a\Delta t\phi_{xx}(r_{21})+\phi(r_{21}) & \dots & -a\Delta t\phi_{xx}(r_{2N})+\phi(r_{2N}) \\ \vdots & \vdots & \vdots \\ -a\Delta t\phi_{xx}(r_{(N-1)1})+\phi(r_{(N-1)1}) & \dots & -a\Delta t\phi_{xx}(r_{(N-1)N})+\phi(r_{(N-1)N}) \\ 0 & \dots & 0 \end{bmatrix},$$

$$B=\begin{bmatrix} 0 & \dots & 0 \\ \phi(r_{21}) & \dots & \phi(r_{2N}) \\ \vdots & \vdots & \vdots \\ \phi(r_{(N-1)1}) & \dots & \phi(r_{(N-1)N}) \\ 0 & \dots & 0 \end{bmatrix}.$$

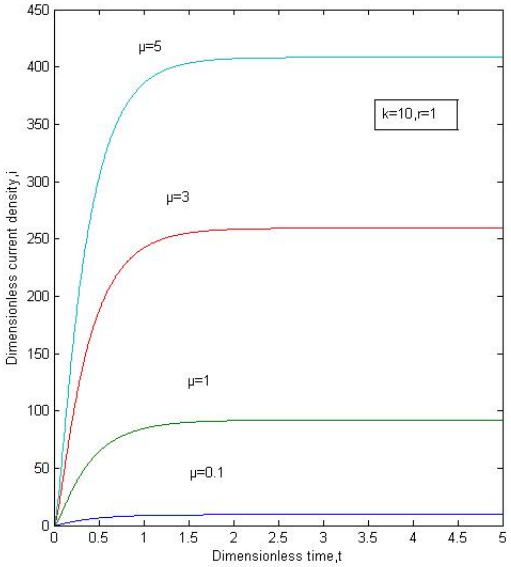
۴ نتایج عددی

غلظت گلوکز، محصول اکسولاکتون و آب‌اکسیژنه حاصل از واکنش و چگالی جریان زیست‌حسگر به پارامتر بی‌بعد $a=\mu/k$، بستگی دارند. μ، سرعت واکنش آنزیم را با سرعت انتقال توده در سراسر لایه‌ی آنزیمی می‌سنجد و پارامتر مهمی در تعیین پارامترهای بهینه مانند ضخامت لایه‌ی آنزیمی است.

برای محاسبه پاسخ زیست‌حسگر، تغییر غلظت اکسولاکتون و آب‌اکسیژنه در سطح مبدل ارزیابی می‌شود. همانطور که در شکل (۱) مشاهده می‌شود غلظت محصول در $t=5$، به حالت پایا می‌رسد. شکل (۲) نشان می‌دهد که با افزایش مدول μ، چگالی جریان زیست‌حسگر افزایش می‌یابد. با توجه به تاثیر مستقیم ضخامت غشای آنزیمی بر مدول μ، می‌توان نتیجه گرفت که واکنش حالت پایا همیشه در غشای ضخیم‌تر به نسبت غشای نازک‌تر، بالاتر است.



شکل ۱: غلظت اکسولاکتون در مقادیر زمانی مختلف



شکل ۲: چگالی جریان با مقادیر مختلف مدول μ

۵ نتیجه‌گیری

با استفاده از روش هم‌مکانی توابع پایه شعاعی به بررسی مدل ریاضی زیست‌حسگر آمپرومتری پرداختیم.چگالی جریان زیست‌حسگر به واکنش گلوکز با آنزیم گلوکزاکسیداز بررسی شد و نشان دادیم که پاسخ به مدول μ وضخامت غشای آنزیمی وابسته است.

مراجع

[1] R. Baronas, , F. Ivanauskas, , J. Kulys, *Mathematical Modeling of Biosensors: An Introduction for Chemists and Mathematicians*, Springer series on chemical sensors and biosensors, 2010.

[2] G. E. Fasshauer, *Meshfree Approximation Methods with Matlab*, Interdisciplinary Mathematical Sciences - Vol. 6, World Scientific Publishers, Singapore, 2007.

[3] O.M. Kirthiga, L. Rajendran, Approximate analytical solution for non-linear reaction diffusion equations in a mono-enzymatic biosensor involving Michaelis- Menten kinetics, *Journal of Electroanalytical Chemistry*, 751, (2015), 119-127.

[4] J. Wang, Glucose biosensors: 40 years of advances and challenges, *Electroanalysis*, 13 (2001), 983-988.

[5] M. Ylilammi, L. Lehtinen, Numerical analysis of a theoretical one-dimensional amperometric enzyme sensor, *Medical & Biological Engineering & Computing*, 26 (1988), 81-87.

اولین همایش ملی ریاضیات زیستی
۲۱ و ۲۲ اسفندماه ۱۳۹۷
همایش ملی ریاضیات زیستی
دانشگاه نیشابور

کاربرد نمونه‌گیری مجموعه رتبه‌دار در مسائل پزشکی و زیستی

حسن اسفندیاری‌فر، گروه ریاضی، دانشگاه امام علی (ع)، تهران، ایران و سید مهدی صالحی [1]، گروه ریاضی، دانشگاه نیشابور، نیشابور، ایران

چکیده: دونه‌گیری مجموعه‌ی رتبه‌دار برای اولین بار توسط مک‌اینتایر (۱۹۵۲) و با انگیزه‌ی یافتن روشی جدید در کنار نمونه‌گیری تصادفی ساده، معمول‌ترین روش در علم نمونه‌گیری، معرفی شد. یکی از ضعف‌های روش سنتی نمونه‌گیری تصادفی ساده این است که در صورت کم بودن اندازه‌ی نمونه، به‌سختی می‌توان یک نمونه‌ی مناسب معرف جامعه بر حسب آن یافت. درحالی‌که نمونه‌گیری مجموعه‌ی رتبه‌دار، که بر پایه‌ی رتبه‌بندی قضاوتی یا چشمی می‌باشد، بر این مشکل غلبه کرده است. بنابراین در صورت برقراری شرایط اعمال این روش رتبه‌دار در مواردی که اندازه‌گیری دقیق واحدهای جامعه هزینه‌بر یا زمان‌بر است، بهتر است جایگزین روش تصادفی ساده شود. در این مقاله به برخی از کاربردهای طرح نمونه‌گیری مجموعه‌ی رتبه‌دار در مسائل پزشکی و زیستی اشاره می‌کنیم.

واژه‌های کلیدی: نمونه‌گیری مجموعه‌ی رتبه‌دار، نمونه‌گیری تصادفی ساده، آزمایش‌های کلینیکی، آلودگی‌های زیستی

۱ مقدمه

در مبحث نمونه‌گیری، پژوهش‌گر همواره با مسأله‌ی انتخاب روش مناسب برای جمع‌آوری داده‌ها مواجه است، به‌طوری‌که وی به‌دنبال روش‌هایی است که دقت را برای برآورد ویژگی‌های یک جامعه‌ی آماری افزایش داده و هزینه را کمینه کند. شرایط زیر را در نظر بگیرید : اندازه‌گیری دقیق واحدهای جامعه‌ی آماری سخت، پرهزینه و یا زمان‌بر است، اما می‌توان آن‌ها را به‌سادگی و با کمترین هزینه رتبه‌بندی کرد. این رتبه‌بندی را می‌توان به‌صورت چشمی یا استفاده از یک نظر کارشناسانه روی متغیر مورد نظر یا حتی یک متغیر دیگر مرتبط با آن انجام داد. در چنین شرایطی، روش نمونه‌گیری مجموعه‌ی رتبه‌دار می‌تواند جایگزین مناسبی برای روش سنتی تصادفی ساده باشد. دلیل این امر این است که در اغلب موارد، برآوردگرهای حاصل از روش نمونه‌گیری مجموعه‌ی رتبه‌دار عملکرد بهتری نسبت به برآوردگرهای حاصل از روش نمونه‌گیری تصادفی دارند، بنابراین با تعداد نمونه‌ی کمتر در روش اول، می‌توان به دقت یکسان در روش دوم رسید. کاربرد عمده‌ی این روش نمونه‌گیری در مطالعات زیست محیطی، کشاورزی و آزمایشگاهی است. به‌طور مثال فرض کنید مایل به تخمین میانگین طول عمر درختان یک جنگل هستیم. این آزمایش به‌دلیل قطع کردن درختان هزینه‌بر خواهد بود. اما درختان را بر اساس ارتفاع آن‌ها (که با طول عمر آن‌ها وابستگی دارد) می‌توان به‌صورت چشمی مرتب کرد و در نتیجه با به‌کارگیری روش نمونه‌گیری

[1] سید مهدی صالحی: salehi2sms@gmail.com

مجموعه‌ی رتبه‌دار متحمل هزینه‌ی کمتر شد. ارزیابی میزان خطرپذیری برخی اماکن همانند مراکز تبدیل پسماند، کارخانه‌های پتروشیمی و سایت‌های هسته‌ای معمولا بسیار هزینه‌بر است. اما در اکثر موارد، اطلاعاتی از قبیل رکوردهای ثبت شده، عکس‌ها و برخی شناسه‌های فیزیکی وجود دارد که می‌توان بر اساس آن‌ها به هدف اصلی دست یافت. در برخی موارد خاص، سطح آلودگی این قبیل اماکن را می‌توان بر اساس سرنخ‌های چشمی یا حتی کاغذهای شیمیایی–واکنشی و مطالعات الکترومغناطیسی به‌دست آورد. در این راستا یک مثال ملموس در تاریخ وجود دارد: چند سال پس از ۱۹۷۰، پژوهشی برای برآورد میزان پلوتونیم (یک عنصر رادیواکتیو که در بسیاری از فرآیندهای هسته‌ای مورد استفاده قرار می‌گیرد) در سطح خاک ناحیه‌ای مشخص و محصور شده در مجاورت سایت نوادا (واقع در آمریکا) انجام گرفت. نمونه‌های خاک از نقاط متفاوتی از آن ناحیه‌ی محصور جمع‌آوری شدند تا با به‌کار بردن روش‌های رادیوشیمیایی میزان تجمع عنصر پلوتونیم را در هر متر مربع سطح خاک معین کنند. اما این آزمایش بسیار گران است. در حالی‌که یک متغیر کمکی به نام $FIDLER$ وجود دارد که با استفاده از آن می‌توان نمونه های خاک را به آسانی رتبه‌بندی کرد که از انجام آزمایش رادیو شیمیایی بسیار ارزانتر می‌باشد. محققانی روش نمونه‌گیری مجموعه‌ی رتبه‌دار با متغیر کمکی ذکر شده را روی داده‌های تحقیق بالا اعمال کردند و نتیجه گرفتند که روش نمونه‌گیری مجموعه‌ی رتبه‌دار کارایی بسیار بالاتری نسبت به نمونه‌گیری تصادفی دارد (برای مشاهده‌ی مثال‌های بیشتر به چن و همکاران، ۲۰۰۴ مراجعه شود).

در این مقاله ابتدا در بخش دوم به مقایسه اندازه‌ی نمونه‌ی مورد نیاز برای برآورد میانگین جامعه‌ی نرمال با استفاده از روش‌های نمونه‌گیری مجموعه‌ی رتبه‌دار (که ازین به بعد با RSS میشناسیم) و نمونه‌گیری تصادفی ساده (SRS) می‌پردازیم. سپس در بخش سوم چند مثال و پیشنهاد برای کاربرد روش RSS در مسائل پزشکی از قبیل آزمایش‌های کلینیکی و همچنین مسائل مربوط به آلودگی‌های زیست محیطی ارائه می‌دهیم.

۲ مقایسه‌ی اندازه‌ی نمونه‌ی مورد نیاز RSS و SRS در برآورد میانگین جامعه‌ی نرمال

تعریف: فرض کنید $(X_1, X_2, \cdots, X_m)$ یک نمونه‌ی تصادفی باشد. این نمونه را به‌صورت چشمی یا قضاوتی مرتب کرده، سپس کوچک‌ترین آن‌ها یعنی $X_{1:m}$ را انتخاب می‌کنیم و با $X_{(1)1}$ نشان می‌دهیم و بقیه را کنار می‌گذاریم. نمونه‌ی تصادفی دیگری به اندازه‌ی m گرفته و پس از مرتب کردن آن به‌صورت چشمی یا قضاوتی، دومین کوچک‌ترین واحد آن را انتخاب می‌کنیم و با $X_{(2)2}$ نشان می‌دهیم و سپس بقیه را کنار می‌گذاریم. این کار را m بار انجام می‌دهیم و $X_{(m)m}$ را از نمونه‌ی mام انتخاب می‌کنیم و با $X_{m,m}$ نشان می‌دهیم. در نهایت به $(X_{(1)1}, X_{(2)2}, \cdots, X_{(m)m})$ یک نمونه‌ی مجموعه‌ی رتبه‌دار (RSS) به اندازه‌ی m گفته می‌شود. در واقع این تعریف روش نمونه‌گیری مجموعه‌ی رتبه‌دار یک‌حلقه‌ای را نشان می‌دهد. اگر این رویه را r بار تکرار کنیم، یک نمونه‌ی مجموعه‌ی رتبه‌دار r–حلقه‌ای به اندازه‌ی نمونه‌ی rm به‌صورت

$$\mathbf{X}_{RSS} = \left\{X_{(i)i,j},\ i = 1, \cdots, m,\ j = 1, \cdots, r\right\} \tag{۱.۲}$$

حاصل خواهد شد. لازم به‌ذکر است که اندیس j در این‌جا بیانگر شمارنده‌ی حلقه می‌باشد.

حال یک جامعه‌ی آماری با میانگین مجهول μ و واریانس σ^2 را در نظر بگیرید. فرض کنید مایلیم μ را برآورد کنیم. در این‌صورت بدیهی است که

$$\bar{X}_{SRS} = \frac{1}{n}\sum_{i=1}^{n} X_i,$$

میانگین یک نمونه‌ی تصادفی با اندازه‌ی $n = rm$، برای μ ناریب است. میانگین یک نمونه‌ی مجموعه‌ی رتبه‌دار r–حلقه‌ای با اندازه‌ی زیرنمونه‌ی m، یعنی

$$\bar{X}_{RSS} = \frac{1}{rm}\sum_{j=1}^{r}\sum_{i=1}^{m} X_{(i)i,j}, \tag{۲.۲}$$

نیز یک برآوردگر ناااریب برای μ می‌باشد. در این‌صورت می‌توان نشان داد (تاکاهاشی و واکیموتو، ۱۹۶۸)

$$Var(\bar{X}_{RSS}) = \frac{\sigma^2}{n} - \frac{1}{nm}\sum_{i=1}^{m}(\mu_{i:m} - \mu)^2, \qquad (۳.۲)$$

به‌طوری‌که $\mu_{i:m}$ بیانگر امید ریاضی iامین آماره‌ی ترتیبی از یک نمونه‌ی تصادفی ساده به اندازه‌ی m می‌باشد. بنابراین بدون در نظر گرفتن توزیع جامعه همواره داریم

$$Var(\bar{X}_{RSS}) < Var(\bar{X}_{SRS}). \qquad (۴.۲)$$

پس $\bar{X}_{RSS}$ کارایی بیشتری نسبت به $\bar{X}_{SRS}$ دارد. در ادامه می‌خواهیم شدت برتری برآوردگر اول را نسبت به دومی بیابیم زمانی که توزیع جامعه‌ی آماری نرمال استاندارد باشد. برای این منظور، اندازه‌ی SRS را N در نظر بگیرید، در نتیجه $Var(\bar{X}_{SRS}) = \frac{1}{N}$. همچنین اندازه‌ی RSS را n در نظر بگیرید (یک‌حلقه‌ای)، در این‌صورت خواهیم داشت

$$Var(\bar{X}_{RSS}) = \frac{1}{n} - \frac{1}{n^2}\sum_{i=1}^{n}\mu_{i:n}^2. \qquad (۵.۲)$$

آرنولد و همکاران (۲۰۰۸) مقادیر عددی $\mu_{i:n}$ را برای توزیع نرمال ارائه داده‌اند. بر این اساس، می‌توان واریانس برآوردگرهای دو طرح مد نظر را در شکل ۱ مقایسه نمود. همان‌طور که مشاهده می‌شود استفاده از RSS بسیار به‌صرفه‌تر از SRS می‌باشد، چرا که به عنوان مثال یک RSS به اندازه‌ی ۱۵ کارایی مشابهی با یک SRS با اندازه‌ی ۱۰۰ دارد.

۳ کاربرد RSS در مسائل پزشکی و زیست‌محیطی

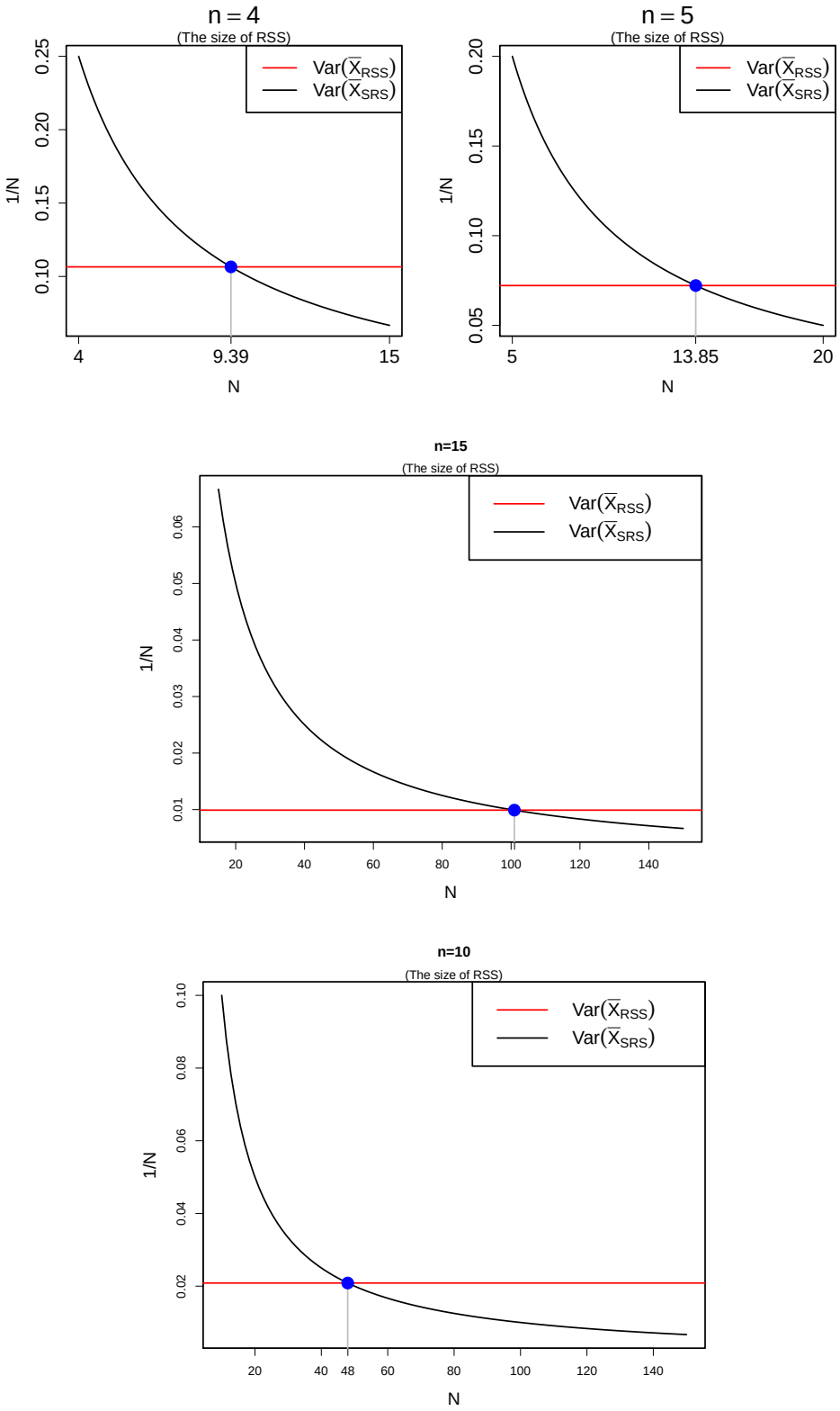
در این بخش چند مثال در رابطه با مسائل بالینی و همچنین زیست محیطی ارائه می‌شود که در همگی آن‌ها امکان اجرای طرح RSS وجود دارد.

تعیین سطح بیلیروبین خون در نوزادان: روش استاندارد برای تعیین شدت زردی، اندازه گیری بیلی روبین سرم میباشد که مستلزم نمونه گیری از ورید محیطی است که این کار باعث تحمیل استرس و درد به نوزاد و هم باعث صرف هزینه و وقت میگردد (گلستان و همکاران، ۱۳۸۶). این در حالی است که می‌توان به‌سادگی سطح بیلیروبین را بر اساس میزان زردرنگی پوست بدن نوزاد رتبه‌بندی کرد. در واقع بر اساس قانون کرامر هرچه میزان زردی پوست از سر به سمت نواحی پایین بدن سرایت کند، بیلیروبین خون بیشتر است.

آزمایش‌های کلینیکی: آزمایش‌های کلینیکی به‌طور معمول از هزینه‌ی بالایی برخوردار هستند. بنابراین برای این‌که نتیجه‌ی تحقیق در این زمینه دقیق‌تر و ارزان‌تر باشد، می‌توان از طرح‌های رتبه‌دار استفاده کرد. برای این منظور، شکل ظاهری (میزان چاقی)، سابقه‌ی فامیلی و یا سن افراد می‌تواند به‌عنوان یک متغیر کمکی برای رتبه‌بندی افراد و در نتیجه اعمال نمونه‌گیری مجموعه‌ی رتبه‌دار مورد استفاده قرار گیرد.

برآورد میزان آلودگی هسته‌ای در یک منطقه: وسیله‌ای که اندازه‌گیری دقیق را انجام می‌دهد، دزمتر نام دارد. این آزمایش هزینه‌بر است. اما می‌توان نمونه‌های مشکوک را ابتدا بر اساس روش‌های آشکارساز از قبیل $X50B$ ، $IM-147$ (اشعه گاما)، $ANPDR27J$ (بتا، گاما) و $ND-RAD-302$ رتبه‌بندی کرد، سپس با استفاده از دزمتر میزان دقیق اثر را محاسبه نمود.

برآورد میزان آلودگی شیمیایی در یک منطقه: برای آگاهی از میزان آلودگی شیمیایی روش‌های فیزیکی گوناگونی وجود دارد که غالبا پرهزینه هستند. این روش‌ها عبارتند از روش‌های طیف‌سنجی تداخل امواج نور، تشخیص تغییرات pH، قانون نفوذ گراهام و فعالیت‌های کاتالیکی. در حالی‌که لوله‌های مخصوصی وجود دارند که دارای نوارهای رنگی برای مشخص کردن نوع و شدت عوامل می‌باشند. بنابراین می‌توان نمونه‌ها را ابتدا با کمک آن‌ها رتبه‌بندی کرد سپس اندازه‌گیری دقیق فیزیکی را انجام داد.



شکل ۱: مقایسه‌ی اندازه‌ی نمونه‌ی مورد نیاز RSS و SRS در برآورد میانگین جامعه‌ی نرمال

برآورد میزان آلودگی میکروبی و بیولوژیکی: تشخیص آلودگی‌های میکروبی نیز نیازمند صرف هزینه‌های بسیار می‌باشد. یکی از روش‌های تشخیص میزان این نوع آلودگی‌ها، استفاده از لوله‌های GK می‌باشد که محیط کشت را برای آن میکروب فراهم می‌آورد. این در حالی است که می‌توان با استفاده از ابزار $HS-3$ نمونه‌ها را رتبه‌بندی کرد.

۴ نتیجه‌گیری

در بخش دوم این مقاله مشاهده شد که در صورت نرمال بودن جامعه، استفاده از RSS بسیار به‌صرفه تر از SRS در برآورد میانگین این جامعه‌ی آماری می‌باشد. زیرا با یک میزان دقت یکسان داده‌شده، نیاز به واحدهای بیشتری از SRS نسبت به RSS داریم. به‌طور دقیق‌تر اگر (N,n) به‌ترتیب بیانگر اندازه‌های نمونه‌ی SRS و RSS باشد، آن‌گاه میانگین‌های حاصل از این دو نمونه با اندازه‌های (۴و۱۰)، (۵و۱۴)، (۱۰و۴۸) و (۱۵و۱۰۰) و ... کارایی یکسانی خواهند داشت. در نتیجه استفاده از RSS در آزمایش‌های پزشکی و زیست‌محیطی مطرح شده در بخش سوم، که همگی هزینه‌بر هستند، به‌صرفه خواهد بود.

مراجع

[6] Chen, Z., Bai, Z. and Sinha, B. K. (2004), Ranked Set Sampling: Theory and Applications. *Springer-Vergla*, New York.

[7] McIntyre, G.A. (1952), A method for unbiased selective sampling, using ranked sets. *Australian Journal of Agricultural Research,* **3**, 385–390.

[8] Takahasi, K. and Wakimoto, K. (1968), On unbiased estimates of the population mean based on the sample stratified by means of ordering. *Annals of the Institute of Statistical Mathematics*, **20**, 1–31.

[9] Wolfe DA. *Ranked set sampling: its relevance and impact on statistical inference*. ISRN Probab. Stat. 2012.

کاربرد ضریب همبستگی درون گروهی (ICC) در تعیین حجم نمونه برای داده‌های خوشه‌ای و تاثیر آن بر توان مطالعه در پژوهش‌های علوم پزشکی

فرزانه برومند[۳،۱]، حبیب الله اسماعیلی[۲]، محمد تقی شاکری[۲]، نوشین اکبری شارک [۳]، [۲]دکترای تخصصی آمار زیستی، استاد مرکز تحقیقات عوامل اجتماعی موثر بر سلامت، دانشگاه علوم پزشکی مشهد، مشهد، ایران، [۳]دانشجوی دکترای آمار زیستی، کمیته تحقیقات دانشجویی، مرکز تحقیقات عوامل اجتماعی موثر بر سلامت، دانشکده بهداشت، دانشگاه علوم پزشکی مشهد، مشهد، ایران.

چکیده: ددر پژوهش های بالینی، برای دستیابی به برآورد دقیق و قابل اعتماد حجم نمونه، لازم است در چارچوب طراحی اجرای تحقیق، آزمون آماری مناسبی برای فرضیات مورد نظر به کار گرفته شود. در عمل به موارد زیادی برمیخوریم که میان فرضیات مورد آزمون، طراحی مطالعه، روش تحلیل آماری و محاسبه تعداد نمونه هماهنگی لازم وجود ندارد. قطعا این ناهماهنگی ها می تواند موجب مخدوش شدن اعتبار و جامعیت مورد نظر در پژوهش های بالینی شود. در بسیاری از مطالعات پزشکی، طراحی مطالعه منجر به وجود آمدن ساختار خاصی در داده‌ها می‌شود. این ساختار با نام ساختار خوشه‌ای شناخته می‌شود. در این شرایط اگر پژوهشگر بدون توجه به آنچه گفته شد از روش های معمول برای تعیین حجم نمونه استفاده کند، حجم نمونه را بیشتر از حد نیاز برآورد خواهد کرد. ضریب همبستگی درون گروهی یا ICC یک شاخص آماری است که به واسطه آن می‌توان درجه توافق و همبستگی اعضای درون یک گروه (خوشه) را تعیین کرد. مقدار ICC بر تعیین حجم نمونه برای داده هایی با ساختار خوشه ای تاثیرگذار است. حجم نمونه موثر اصطلاحی است که برای توصیف رابطه حجم نمونه درون هر خوشه در مقایسه با حجم نمونه کل استفاده می‌شود. با توجه به اندازه ICC و اثر طرح (DE) ممکن است حجم نمونه مورد نیاز برای بدست آمدن معنی داری آماری کاهش یابد. حتی مقدار کوچک ICC می‌تواند با افزایش حجم درون هر خوشه، منجر به بزرگ شدن DE و به تبع آن کاهش ESS شود. بنابراین در تعیین حجم نمونه در داده های خوشه‌ای از کنار ICC کوچک هم نباید به راحتی گذشت. افزایش تعداد خوشه‌ها، بیشتر از افزایش حجم درون خوشه‌ها می‌تواند به بالا بردن توان مطالعه منجر شود. بنابراین یکی از راه‌های افزایش توان مطالعه در داده‌هایی با ساختار خوشه‌ای افزایش تعداد خوشه‌هاست. توجه داریم فرمول‌های استاندارد برای تعیین حجم نمونه باید بکار گرفته شود تنها درصورت وجود ساختار خوشه‌ای در داده‌ها برای رسیدن به حجم نمونه موثر می‌توان تعدیل‌سازی لازم را انجام داد.

واژه‌های کلیدی: ضریب همبستگی درون گروهی، حجم نمونه موثر، داده‌های خوشه‌ای، توان مطالعه، کارآزمایی بالینی.

[۱] فرزانه برومند: Boroumandf951@mums.ac.ir

۱ مقدمه

در پژوهش‌های بالینی، برای دستیابی به برآورد دقیق و قابل اعتماد حجم نمونه، لازم است در چارچوب طراحی اجرای تحقیق، آزمون آماری مناسبی برای فرضیات مورد نظر به کار گرفته شود. فرضیه مورد آزمون باید بیانگر اهداف مورد مطالعه باشد. در عمل به موارد زیادی برمیخوریم که میان فرضیات مورد آزمون، طراحی مطالعه، روش تحلیل آماری و محاسبه تعداد نمونه هماهنگی لازم وجود ندارد. قطعا این ناهماهنگی‌ها می‌تواند موجب مخدوش شدن اعتبار و جامعیت مورد نظر در پژوهش‌های بالینی شود. (چاو و همکاران (۲۰۰۷)) در بسیاری از مطالعات پزشکی، طراحی مطالعه منجر به وجود آمدن ساختار خاصی در داده‌ها می‌شود. در این ساختار که با نام ساختار خوشه‌ای شناخته می‌شود، مجموعه‌ی داده‌ها در گروه‌ها یا در اصطلاح خوشه‌هایی قرار می‌گیرد به طوری که مشاهدات درون هر خوشه عمدتا دارای همبستگی مثبت هستند. تحلیل‌هایی که این همبستگی را به حساب نمی‌آورند غالبا منجر به استنباط‌هایی گمراه‌کننده می‌شوند. (دیگل و همکاران (۱۹۹۴) و فیتزموریس و همکاران (۲۰۰۸)) در این شرایط اگر پژوهشگر بدون توجه به آنچه گفته شد از روش های معمول برای تعیین حجم نمونه استفاده کند، حجم نمونه را بیشتر از حد نیاز برآورد خواهد کرد. زیرا شباهت ایجاد شده به دلیل ساختار طراحی مطالعه بین اعضای درون یک خوشه، می تواند تغییر پذیری متغیر پاسخ بدست آمده از نمونه های مربوط به یک خوشه را کاهش می دهد. این تغییرپذیری باعث کاهش حجم نمونه مورد نیاز برای رسیدن به یک توان معین است. این موضوع برای پژوهش هایی با طرح کارآزمایی بالینی و پژوهش های حوزه سلامت بسیار اهمیت دارد، زیرا طراحی مطالعه در بسیاری از این پژوهش ها، باعث ایجاد خوشه در داده ها می شود. (دانر و کلار (۲۰۰۰)) در آمار ضریب همبستگی درون گروهی[1] یا ICC یک شاخص آماری است که به واسطه آن می‌توان درجه توافق و همبستگی اعضای درون یک گروه (خوشه) را تعیین کرد. (دانر و کلار (۲۰۰۰)) این شاخص اولین بار توسط فیشر[2] در سال ۱۹۵۴ به عنوان یک اصلاح برای ضریب همبستگی پیرسن معرفی شد. در حالی که ICC مدرن با استفاده از میانگین مربعات آنالیز واریانس محاسبه می‌شود. (کو و لی (۲۰۱۶)) ICC را میتوان با استفاده از نرم‌افزارهای مختلف آماری به آسانی محاسبه کرد. هدف این مقاله اهمیت استفاده از ICC در تعیین حجم نمونه برای داده‌های خوشه‌ای و تاثیر آن بر توان آماری مطالعه است.

۲ روش

همانطور که پیش‌تر در مقدمه بیان شد ICC، اندازه‌ای از همبستگی درون گروهی یا درون خوشه‌ای را نشان می‌دهد. محاسبه این شاخص با بررسی واریانس بین خوشه و درون خوشه محاسبه می‌شود. با تقسیم واریانس بین خوشه ها به مجموع واریانس بین و درون خوشه‌ها مقدار ICC بدست می‌آید. (کلیپ و همکاران (۲۰۰۴))

$$ICC = \frac{S_b^2}{S_b^2 + S_w^2} \tag{۱.۲}$$

که در آن S_b^2 واریانس بین خوشه ای و S_w^2 واریانس درون خوشه ای است. مقدار ICC براساس این فرمول مقداری بین صفر و یک است. کوچک بودن S_w^2 باعث می‌شود ICC به یک نزدیک شود که نشان دهنده ی بزرگ بودن همبستگی درون خوشه ای است. همینطور بزرگ بودن S_w^2 باعث می شود ICC به صفر نزدیک شود که نشان دهنده کوچک بودن همبستگی درون خوشه ای است. (دانر و کلار (۲۰۰۰)، کلیپ و همکاران (۲۰۰۴))

اینکه مقدار ICC چگونه بر حجم نمونه در داده های خوشه‌ای تاثیرگذار است را با بیان مفهومی به نام حجم نمونه موثر[3] توضیح خواهیم داد. حجم نمونه موثر اصطلاحی است که برای توصیف رابطه حجم نمونه درون هر خوشه در مقایسه با حجم نمونه کل استفاده

[1] Interclass correlation coefficient
[2] Fisher
[3] Effective Sample Size
[4] Design Effect

می‌شود. با توجه به اندازه ICC و اثر طرح[۴] (DE) ممکن است حجم نمونه مورد نیاز برای بدست آمدن معنی داری آماری کاهش یابد. حجم نمونه موثر (ESS) از رابطه زیر بدست می آید.

$$ESS = \frac{m.k}{DE} \qquad (۲.۲)$$

که در آن m حجم خوشه ها، k تعداد خوشه ها، $m.k$ حجم نمونه کل و DE اثر طرح است. اثر طرح از رابطه زیر بدست می آید:

$$DE = ۱ + ICC(m - ۱) \qquad (۳.۲)$$

اگر $ICC = ۰$ در این صورت $DE = ۱$ و حجم نمونه موثر با حجم نمونه کل برابر خواهد بود. $(ESS = m.k)$

اگر $ICC > ۰$ (حتی اگر ICC خیلی کوچک باشد) در این صورت مقدار DE با بزرگ بودن حجم درون خوشه (m) بزرگ می شود در این حالت ESS یا حجم نمونه موثر کاهش می یابد بنابراین به حجم نمونه کمتری برای معنی داری آماری نیاز است. (کلیپ و همکاران (۲۰۰۴)، مورای و همکاران (۱۹۹۴)، مورای و شورت (۱۹۹۶).)

اگر $ICC = ۱$ درنتیجه $DE = m$ و به تبع آن $ESS = K$ به عبارت دیگر در این شرایط تنها کافی است از هر خوشه یک نمونه اختیار شود. پیش تر هم ذکر شد در صورتی که $ICC = ۱$ به این معناست که تمامی اعضای درون خوشه باهم برابرند. استفاده از این روابط در مرحله قبل آزمایش برای تعیین حجم نمونه مناسب در پژوهش های بالینی کار برد فراوان دارد. (کاچران (۲۰۰۷))

۳ مثال کاربردی

برای بررسی تاثیر ماساژ بر سطح بیلی روبین نوزادان نارس یک مطالعه کارآزمایی طراحی شد. در این مطالعه نوزادان نارسی که سن حاملگی مادرانشان بین ۳۰ الی ۳۶ هفته بود و در بخش $NICU$، بستری بودند به طور تصادفی به دو گروه ماساژ درمانی با روغن کنجد و ماساژ درمانی با روش متداول (کنترل) تخصیص داده شدند. چهار پرستار، ماساژ درمانی را بر عهده داشتند. تخصیص این چهار پرستار به گروه ماساژ درمانی با روغن کنجد و کنترل به صورت تصادفی انجام شد.(۲ نفر گروه کنترل ۲ نفر گروه ماساژ درمانی با روغن کنجد) تعیین حجم نمونه برای این مطالعه به طور معمول بر اساس فرمول مقایسه میانگین ها در دو گروه مستقل، در سطح معنی داری ۰/۰۵ و توان ۸۰ درصد و اطلاعات بدست آمده از مقالات مشابه $S_۱ = ۱/۹$ و $S_۲ = ۲/۳۴$ و $\overline{X}_۱ = ۱۱/۰۱$ و $\overline{X}_۲ = ۹/۵$ برای هر گروه ۳۲ بدست آمد. ۳۲ نوزاد گروه ماساژ درمانی ۳۲ نوزاد گروه کنترل که در مجموع حجم نمونه مورد نیاز ۶۴ می‌باشد.

$$n = \frac{\left(Z_{۱-\frac{\alpha}{۲}} + Z_{۱-\beta}\right)^۲ (S_۱^۲ + S_۲^۲)}{\left(\overline{X}_۲ - \overline{X}_۱\right)^۲} \qquad (۱.۳)$$

تعیین حجم نمونه بدون در نظر گرفتن ساختار خوشه ای داده‌ها برای رسیدن به توان ۸۰ درصد بدست آمده است. حال آنکه بین نوزادانی که توسط یه پرستار خاص ماساژ می‌بینند همبستگی وجود دارد و این همبستگی باعث بوجود آمدن خوشه می‌شود. طراحی این مطالعه باعث بوجود آمدن چهار خوشه (به تعداد پرستارانی که نوزادان را ماساژ می دهند) شده است. پس از محاسبه ICC برای این مطالعه (۰/۰۲۵) و برآورد حجم نمونه موثر بر اساس فرمول دو، مشخص شد توان آزمون در این شرایط ۰/۶۹ درصد است که این یعنی با درنظر نگرفتن ساختار خوشه‌ای داده‌ها در تعیین حجم نمونه، مقدار قابل توجهی از توان آزمون از دست رفته است.

۴ تاثیر ICC و DE بر محاسبه توان و حجم نمونه

برای واضح شدن تاثیر ICC و طراحی مطالعه بر محاسبه توان و حجم نمونه، سه استراتژی متفاوت را در مورد K، m و mk بکار می‌بریم تا تاثیر تغییرات هریک را بر توان آزمون بررسی کنیم.

۱- ثابت بودن حجم نمونه کل، mk

در این قسمت توان آزمون برای مثال کاربردی را به ازای مقادیر مختلف m و k به طوری که mk ثابت باشد، شبیه‌سازی می‌کنیم.

همانطور که در جدول ۱ مشاهده می‌کنید مقدار کوچک ICC هم می‌تواند تاثیر بزرگی بر توان آزمون داشته باشد. مقدار mk برابر با عدد ۶۴، بدست آمده براساس فرمول چهار اختیار شده است. همانطور که در سطر اول جدول شماره یک مشاهده می‌کنید در این شرایط توان آزمون برابر ۶۹ درصد است. اما با افزایش تعداد خوشه‌ها و کاهش حجم درون هر خوشه می‌توان توان آزمون را به مقدار مطلوب ۸۰ درصد رساند. پاسخ این سوال که چرا با افزایش تعداد خوشه‌ها و کاهش حجم درون خوشه‌ها توان افزایش می‌یابد، پیش تر در مقدمه گفته شد، تغییر پذیری درون خوشه‌ها کم است بنابراین با افزایش تعداد خوشه می‌توان تغییرپذیری را افزایش داد از طرفی چون در این استراتژی mk مقدار ثابتی است. با افزایش تعداد خوشه‌ها حجم درون خوشه کاهش می‌یابد. در این مثال اگر بخواهیم با ۶۴ نمونه به توان مطلوب ۸۰ درصد دست یابیم در طراحی مطالعه باید هر نوزاد توسط یک پرستار ماساژ داده شود که یعنی در این مطالعه به ۶۴ پرستار نیاز است. ممکن است این روش برای افزایش توان مطالعه شدنی نباشد.

جدول ۱: حجم نمونه موثر و توان در صورت ثابت بودن mk

تعداد ماساژ دهندگان k	تعداد نوزادانm	تعداد کل mk	$ICC = ۰/۰۲۵$		$tTest^*\ POWER$
			DE	ESS	
۴	۱۶	۶۴	۱/۳۸	۴۷	۰/۶۹
۸	۸	۶۴	۱/۱۸	۵۵	۰/۷۵
۱۶	۴	۶۴	۱/۰۸	۶۰	۰/۷۸
۳۲	۲	۶۴	۱/۰۳	۶۳	۰/۷۹
۶۴	۱	۶۴	۱/۰۰	۶۴	۰/۸۰

در جدول فوق مقادیرمحاسبه شده برای ESS به بالا گرد شده است و به دلیل وجود دو گروه، به ESSهای فرد یک عدد اضافه شده است.

*توان محاسبه شده برای آزمون تی مستقل براساس حجم نمونه مورد نیاز (ESS) محاسبه شده است.

۲- ثابت بودن تعداد خوشه ها، k

در این قسمت توان آزمون برای مثال کاربردی را به ازای مقادیر مختلف m و mk به طوری که k ثابت باشد، شبیه سازی می‌کنیم.

با توجه به جدول ۲ در صورت ثابت بودن تعداد خوشه ها با افزایش حجم درون هر خوشه توان مطالعه افزایش می یابد. قابل ذکر است حتی با وجود ICC کوچک در این مطالعه تاثیر قابل توجه آن بر توان آزمون قابل مشاهده است.

در جدول فوق مقادیرمحاسبه شده برای ESS به بالا گرد شده است و به دلیل وجود دو گروه، به ESSهای فرد یک عدد اضافه شده است.

*توان محاسبه شده برای آزمون تی مستقل براساس حجم نمونه مورد نیاز (ESS) محاسبه شده است.

ثابت بودن حجم درون خوشه ها، m

در این قسمت توان آزمون برای مثال کاربردی را به ازای مقادیر مختلف k و mk به طوری که m ثابت باشد، شبیه‌سازی می‌کنیم.

جدول ۲: حجم نمونه موثر و توان در صورت ثابت بودن k

تعداد ماساژ دهندگان k	تعداد نوزادانm	تعداد کلmk	$ICC = ۰/۰۲۵$		$Test^*\ tPOWER$
			DE	ESS	
۴	۱۰	۴۰	۱/۲۳	۳۳	۰/۵۴
۴	۲۰	۸۰	۱/۴۸	۵۵	۰/۷۵
۴	۳۰	۱۲۰	۱/۷۳	۷۰	۰/۸۴
۴	۴۰	۱۶۰	۱/۹۸	۸۲	۰/۸۹

همانطور که در جدول ۳ مشاهده می‌کنید افزایش تعداد خوشه منجر به افزایش توان مطالعه خواهد شد.

جدول ۳: حجم نمونه موثر و توان در صورت ثابت بودن m

تعداد ماساژ دهندگان k	تعداد نوزادانm	تعداد کلmk	$ICC = ۰/۰۲۵$		$Test^*\ t\ POWER$
			DE	ESS	
۴	۱۰	۴۰	۱/۲۳	۳۳	۰/۵۴
۶	۱۰	۶۰	۱/۲۳	۴۹	۰/۷۰
۸	۱۰	۸۰	۱/۲۳	۶۶	۰/۸۲
۱۰	۱۰	۱۰۰	۱/۲۳	۸۲	۰/۸۹
۱۲	۱۰	۱۲۰	۱/۲۳	۹۸	۰/۹۳

در جدول فوق مقادیرمحاسبه شده برای ESS به بالا گرد شده است و به دلیل وجود دو گروه، به ESSهای فرد یک عدد اضافه شده است.

*توان محاسبه شده برای آزمون تی مستقل براساس حجم نمونه مورد نیاز (ESS) محاسبه شده است.

بحث و نتیجه گیری

به طور خلاصه شاخص ICC در آمار نشان دهنده درجه همبستگی و توافق است. ICC اندازه‌ای از همبستگی مقادیر درون یک خوشه را نشان می‌دهد. حجم نمونه در پژوهش‌های بالینی معمولا کم است با این حال اثر طرح (DE) با افزایش حجم درون هر خوشه، به مقدار قابل توجهی افزایش می یابد. افزایش DE، حجم نمونه موثر برای رسیدن به معنی داری آماری را کاهش می دهد. همچنین مقدار ICC در محاسبه ESS در مطالعاتی که داده های آن ساختار خوشه ای دارند، موثر است. حتی مقدار کوچک ICC می‌تواند با افزایش m، منجر به بزرگ شدن DE و به تبع آن کاهش ESS شود. بنابراین در تعیین حجم نمونه در داده های خوشه‌ای از کنار ICC کوچک هم نباید به راحتی گذشت. زیاد بودن تعداد خوشه ها و کم بودن اعضای درون خوشه می‌تواند به کوچک شدن DE بیانجامد. افزایش تعداد خوشه‌ها، بیشتر از افزایش حجم درون خوشه‌ها می‌تواند به بالا بردن توان مطالعه منجر شود. بنابراین یکی از راه‌های افزایش توان مطالعه در داده هایی با ساختار خوشه ای افزایش تعداد خوشه هاست. توجه داریم فرمول‌های استاندارد برای تعیین حجم نمونه باید بکار گرفته شود تنها درصورت وجود ساختار خوشه ای در داده ها برای رسیدن به حجم نمونه موثر می‌توان این تعدیل سازی را انجام داد.

مراجع

[10] CHOW, S.-C., WANG, H. & SHAO, J. 2007. *Sample size calculations in clinical research*, CRC press.

[11] COCHRAN, W. G. 2007. *Sampling techniques*, John Wiley & Sons.

[12] DIGGLE, P., LIANG, K.-Y. & ZEGER, S. L. 1994. Longitudinal data analysis. *New York: Oxford University Press,* 5**,** 13.

[13] DONNER, A. & KLAR, N. 2000. Design and analysis of cluster randomization trials in health research.

[14] FITZMAURICE, G., DAVIDIAN, M., VERBEKE, G. & MOLENBERGHS, G. 2008. *Longitudinal data analysis*, CRC Press.

[15] KILLIP, S., MAHFOUD, Z. & PEARCE, K. 2004. What is an intracluster correlation coefficient? Crucial concepts for primary care researchers. *The Annals of Family Medicine,* 2**,** 204-208.

[16] KOO, T. K. & LI, M. Y. 2016. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of chiropractic medicine,* 15**,** 155-163.

[17] MURRAY, D. M., ROONEY, B. L., HANNAN, P. J., PETERSON, A. V., ARY, D. V., BIGLAN, A., BOTVIN, G. J., EVANS, R. I., FLAY, B. R. & FUTTERMAN, R. 1994. Intraclass correlation among common measures of adolescent smoking: estimates, correlates, and applications in smoking prevention studies. *American Journal of Epidemiology,* 140**,** 1038-1050.

[18] MURRAY, D. M. & SHORT, B. 1996. Intraclass correlation among measures related to alcohol use by school aged adolescents: estimates, correlates and applications in intervention studies. *Journal of Drug Education,* 26**,** 207-230.

معادلات دیفرانسیل کسری

سوسن امیری [۱]، محمد حسین رحمانی دوست گروه ریاضی، دانشگاه نیشابور، نیشابور

چکیده: در سال ۱۹۶۵ میلادی هوپیتال نامه‌ای به لایپ‌نیتز نوشت. و این سؤال را مطرح کرد که « در نمادگذاری $\frac{d^n y}{dx^n}$، هنگامی که $n = \frac{۱}{۲}$ باشد چه رخ می‌دهد؟» لایپ‌نیتز پاسخ داد: «این امر در حال حاضر پارادوکسی آشکار است که روزی به عنوان یکی از مفیدترین مباحث علمی مطرح خواهد گشت.» محاسبات دیفرانسیل و انتگرال کسری به طور گسترده در حوزه‌هایی چون ریاضیات زیستی، اکولوژی، فیزیک، شیمی و غیره استفاده می‌شود. مدل‌هایی با مشتقات مرتبه کسری، بر حسب هدف و نوع سیستم با یکدیگر متفاوت هستند. در این میان، معادلات دیفرانسیل کسری [۱] به طور گسترده در مدل‌های بیولوژی استفاده می‌شود. این دو معادله از جنبه‌های مختلفی همچون ارائه راه‌حل‌های معادلات دیفرانسیل کسری و نیز ویژگی‌های رفتار مختلف راه‌حل‌ها، بررسی شده‌اند. یکی از مهم‌ترین کاربردهای این معادلات در فیزیک در شبیه‌سازی حرکت و شبیه‌سازی سیستم‌های بیولوژیکی می‌باشد.

واژه‌های کلیدی: مشتق کسری گرانولد–لتنیکوف، مشتق کسری ریمان–لیوویل، مشتق کسری کاپتوی

۱ مقدمه

محاسبات دیفرانسیل و انتگرال کسری، شاخه‌ای از محاسبات کلاسیک بوده که مرتبه مشتقات و انتگرال را به یک مرتبه غیر عددی صحیح تعمیم می‌دهد. این نوع محاسبات در ابتدا بیش از ۳۰۰ سال پیش از سوی ال هوپیتال و لایپ‌نیتز به صورت مشتقاتی با مرتبه $\alpha = \frac{۱}{۲}$ ارائه شدند. درسال‌های اخیر معادلات دیفرانسیل کسری به دلیل پیدایش مکرر در مکانیسم‌های جریان، بیولوژی، الکتروشیمی و مسائل فنی و فیزیکی دیگر، محققین بسیاری را بر آن داشته است تا در این زمینه کارهای زیادی انجام دهند. بسیاری از پژوهشگران از معادلات دیفرانسیل کسری برای توسعه مدل‌های خود در بیولوژی از جمله تعاملات بین شکار و شکارچی، فارماکرکنیتیک (بخشی از داروسازی که با جذب و پخش دارو در بدن سر و کار دارد) ، عفونت HIV و غیره استفاده کرده‌اند. امروزه روش‌های زیادی برای حل این نوع مسائل ابداع شده‌اند و افراد زیادی این روش‌ها را به کار گرفته‌اند تا بتوانند جواب واقعی این مسائل را بیابند.

[1] Fractional differential equation (FDE)

[۱] سوسن امیری: Sousan.amirikhan@gmail.com

۱.۱ مشتق کسری گرانولد-لتنیکوف

تابع پیوسته $f(t) \in C[a,b]$ را به صورت زیر در نظر می‌گیریم. مشتق مرتبه اول تابع $f(t)$ چنین است:

$$f'(t) = \frac{df}{dt} = \lim_{h\to ۰} \frac{f(t) - f(t-h)}{h}, \tag{۱.۱}$$

$$f''(t) = \frac{d^۲f}{dt^۲} = \lim_{h\to ۰} \frac{f(t) - ۲f(t-h) + f(t-۲h)}{h^۲}, \tag{۲.۱}$$

$$f'''(t) = \frac{d^۳f}{dt^۳} = \lim_{h\to ۰} \frac{f(t) - ۳f(t-h) + ۳f(t-۲h) - f(t-۳h)}{h^۳}, \tag{۳.۱}$$

$$f^n(t) = \frac{d^nf}{dt^n} = \lim_{h\to ۰} \frac{۱}{h^n} \sum_{r=۰}^{n} (-۱)^{n-r} \binom{n}{r} f\big(t-(n-r)h\big) \tag{۴.۱}$$

رابطه (۴.۱) را می‌توان به صورت زیر بازنویسی کرد:

$$f^n(t) = \lim_{h\to ۰} \frac{۱}{h^n} \sum_{r=۰}^{n} (-۱)^{r} \binom{n}{r} f(t-rh). \tag{۵.۱}$$

اکنون تعمیم روابط (۱.۱) و (۵.۱) به صورت زیر خواهد بود:

$$f_h^p(t) = \frac{۱}{h^p} \sum_{r=۰}^{n} (-۱)^{r} \binom{p}{r} f(t-rh),$$

$${}_a^C D_t^p f(t) = \sum_{k=۰}^{m} \frac{f^{(k)}(a)(t-a)^{-p+k}}{\Gamma(-p+k+۱)} + \frac{۱}{\Gamma(-p+m+۱)} \int_۰^x (t-\tau)^{m-p} \times f^{(m+۱)}(\tau)d\tau.$$

۲.۱ مشتق کسری ریمان-لیوویل

مشتق کسری ریمان-لیوویل به صورت زیر تعریف می‌شود:

$${}_aD_t^p f(t) = (\frac{d}{dt})^{m+۱} \frac{۱}{\Gamma(m-p+۱)} \int_a^t (t-\tau)^{m-p} f(\tau)d\tau, \quad m \le p < m+۱.$$

۳.۱ مشتق کسری کاپتو

فرض کنید $f \in A^k[t_۰,T]$ و $k \in \mathbb{N}$ باشد که $A^k[t_۰,T]$ مجموعه‌ای از تمام توابع $f : [t_۰,T] \to \mathbb{R}$ بوده که $f^{k-۱}$ پیوسته است و $t \in [t_۰,T]$. در این صورت برای هر $k-۱ < \alpha < k$، مشتق مرتبه کسری کاپتوی به صورت زیر تعریف می‌شود:

$${}_{t_۰}^C D_t^\alpha f(t) = \frac{۱}{\Gamma(k-\alpha)} \int_{t_۰}^t (t-\tau)^{k-\alpha-۱} f^{(k)}(\tau)d\tau,$$

که $\Gamma(.)$ اشاره به تابع گاما دارد.

۲ نتایج اصلی

لم ۱.۲. اگر $f = (f_۱, \dots, f_n) \in C^۱$، آنگاه مساله مقدار اولیه

$$D^{\alpha_i} y_i(t) = f_i(t, y_۱, \dots, y_n),\ y_i^{(k)} = C_k^i,\ ۱ \le i \le n,\ ۱ \le k \le m_i$$

معادل معادله انتگرال ولترای زیر است:

$$y_i(t) = \sum_{k=۰}^{m_i} C_k^i \frac{l^k}{k!} + I^{\alpha_i} f_i(t, y_۱, \dots, y_n),\ ۱ \le i \le n. \tag{۱.۲}$$

اثبات. به (۱۹) مراجعه شود. □

قضیه ۲.۲. فرض کنید

$$f = (f_۱, \dots, f_n) : w \to \mathbb{R}^n \in C^۱$$

که در آن

$$w = [۰, \chi^*] \times \prod_{j=۱}^{n} [y_j(۰) - l_j, y_j(۰) + l_j],\ \chi^* > ۰,\ l_j > ۰,\ \forall j$$

در این صورت دستگاه معادلات

$$D^{\alpha_i} y_i(t) = f_i(t, y_۱, \dots, y_n),\ y_i^{(k)} = C_k^i,\ ۱ \le i \le n,\ ۱ \le k \le m_i \tag{۲.۲}$$

که $m_i < \alpha_i < m_i + ۱$ و D^{α_i} یک جواب یکتا به صورت $\bar{y} : [۰, \chi] \to \mathbb{R}$ دارد که

$$\chi = \min \left\{ \chi^*, \left(\frac{l\Gamma(\alpha_i + ۱)}{[۱ + \alpha] \parallel f \parallel} \right)^{\frac{۱}{\alpha_i}}, \left(\frac{lk!}{[۱ + \alpha]|C_k^i|} \right)^{\frac{۱}{k}} \right\},\ i = ۱, ۲, \dots, n$$

$$k = ۱, ۲, \dots, m_i,\ \ l = \min\{l_۱, l_۲, \dots, l_n\},\ \ \alpha = \max\{\alpha_۱, \alpha_۲, \dots, \alpha_n\}.$$

لازم به ذکر است که $[۱ + \alpha]$ نشان دهنده جزء صحیح $\alpha + ۱$ می‌باشد. **اثبات.** با توجه به لم (۱.۲) معادله (۲.۲) معادل است با

$$y - i(t) = \sum_{k=۰}^{m_i} C_k^i \frac{t^k}{k!} + I^{\alpha_i} f_i(t, \bar{y}),\ ۱ \le i \le n.$$

قرار می‌دهیم $\bar{A} = \big(A_۱(\bar{y}), \dots, A_n(\bar{y})\big)$ که

$$A_i[\bar{y}(t)] = \sum_{k=۰}^{m-۱} C_k^i \frac{t^k}{k!} + \frac{۱}{\Gamma(\alpha_i)} \int_۰^t (t - s)^{\alpha_i - ۱} f(s, \bar{y}(s)) ds. \tag{۳.۲}$$

ناگفته نماند که یک نقطه ثابت از عملگر A یک جواب معادله (۲.۲) است. چون $f \in C^۱$، به طور موضعی با یک همسایگی به شعاع l و با ثابت لیپ‌شیتز است.

فرض کنید

$$U = \{\bar{y} \in B\ :\ |y_i - y_i(۰)| \le l,\ ۱ \le i \le n,\ \chi < l\}.$$

به وضوح U یک زیر مجموعه بسته و محدب و غیر تهی از $C([۰,\chi]^n)$ که f لیپ‌شیتز است.
چون f_i $(i = ۱,۲,\dots)$ روی مجموعه فشرده w پیوسته بوده، بنابراین روی w پیوسته یکنواخت خواهد بود. تعریف پیوستگی یکنواخت را می‌نویسیم:

$$\forall \varepsilon > ۰,\ \exists \delta > ۰,\ \ |\bar{y} - \bar{z}| < \delta,\ \ |f_i(t,\bar{y}) - f_i(t,\bar{z})| < \varepsilon^*$$
$$\varepsilon^* = \min\{\frac{\varepsilon\Gamma(\alpha_۱ + ۱)}{\chi^{\alpha_۱}}, \dots, \frac{\varepsilon\Gamma(\alpha_n + ۱)}{\chi^{\alpha_n}}.$$

اکنون فرض می‌کنیم $\bar{y}, \bar{z} \in U$ به طوری که $|\bar{y} - \bar{z}| < \delta$. بنابراین با توجه به نامساوی

$$|f_i(t,\bar{y}) - f_i(t,\bar{z})| < \varepsilon^* \qquad (۴.۲)$$

داریم

$$|A_i\bar{y}(t) - z_i\bar{y}(t)| = \left|\frac{۱}{\Gamma(\alpha_i)}\int_۰^t (t-s)^{\alpha_i - ۱}[f_i(s,\bar{y}(s)) - f_i(s,\bar{z}(s))]ds\right|$$
$$< \varepsilon^* \left|\frac{۱}{\Gamma(\alpha_i)}\int_۰^t (t-s)^{\alpha_i - ۱}\right| \le \varepsilon^* \frac{\chi^{\alpha_i}}{\Gamma(\alpha_i + ۱)} \le \varepsilon,\ \forall i$$

با توجه به پیوستگی A_iها، A نیز پیوسته خواهد بود.
به علاوه برای $\bar{y} \in U$ و $t \in [۰,\chi]$ داریم

$$|A_i\bar{y}(t) - C_۰^i| = \left|\sum_{k=۱}^{m_i} C_k^i \frac{t^k}{k!} + \frac{۱}{\Gamma(\alpha_i)}\int_۰^t (t-s)^{\alpha_i - ۱} f(s,\bar{y}(s))ds\right|$$
$$\le \left|\sum_{k=۱}^{m_i} C_k^i \frac{t^k}{k!}\right| + \frac{\parallel f \parallel \chi^{\alpha_i}}{\Gamma(\alpha_i + ۱)} \le l,\ \forall i$$

بنابراین نشان داده شد که A نگاشتی از مجموعه U به خودش است. اکنون با استقرا ثابت می‌کنیم که برای هر $\bar{y}, \bar{z} \in U$:

$$|A_i^n\bar{y}(s) - A_i^n\bar{z}(s)| \le \frac{(Ls^{\alpha_i})^n}{\Gamma(۱ + n\alpha_i)}|\bar{y} - \bar{z}|,\ \ n = ۰,۱,۲,\dots.$$

اگر $n = ۰$، بدیهی است که تساوی بالا برقرار است. فرض کنیم تساوی در گام $n - ۱$ برقرار باشد. ثابت می‌شود که در گام nام نیز برقرار است.

$$|A_i^n\bar{y}(t) - A_i^n\bar{z}(t)| = |A_i(A_i^{n-۱}\bar{y}(t)) - A_i(A_i^{n-۱}\bar{z}(t))|$$
$$= \frac{۱}{\Gamma(\alpha_i)}\int_۰^t (t-s)^{\alpha_i - ۱}[f_i(s,A_i^{n-۱}\bar{y}(s)) - f_i(s,A_i^{n-۱}\bar{z}(s))]ds.$$

با توجه به این که f روی U لیپ‌شیتز است، از استقرا به دست می‌آید:

$$|A_i^n\bar{y}(t) - A_i^n\bar{z}(t)| \le \frac{L}{\Gamma(\alpha_i)}\int_۰^t (t-s)^{\alpha_i - ۱)} \parallel A_i^{n-۱}\bar{y}(s) - A_i^{n-۱}\bar{z}(s) \parallel ds$$
$$\le \frac{L^n}{\Gamma(\alpha_i)\Gamma(۱ + (n-۱)\alpha_i)}\int_۰^t (t-s)^{\alpha_i - ۱} s^{\alpha_i(n-۱)}|\bar{y} - \bar{z}|ds$$
$$\le \frac{L^n|\bar{y} - \bar{z}|}{\Gamma(\alpha_i)\Gamma(۱ + (n-۱)\alpha_i)}\int_۰^t (t-s)^{\alpha_i - ۱} s^{\alpha_i(n-۱)}ds$$
$$= \frac{L^n|\bar{y} - \bar{z}|I^{\alpha_i}t^{\alpha_i(n-۱)}}{\Gamma(۱ + (n-۱)\alpha_i)} = \frac{(Lt^{\alpha_i})^n|\bar{y} - \bar{z}|}{\Gamma(۱ + n\alpha_i)} \le \frac{L^n|\bar{y} - \bar{z}|}{\Gamma(۱ + n\alpha_i)}\chi^{n\alpha_i}.$$

در نتیجه

$$|A_i^n \bar{y}(t) - A_i^n \bar{z}(t)| \leq \frac{(L\chi^{\alpha_i})^n}{\Gamma(۱ + n\alpha_i)} |\bar{y} - \bar{z}|.$$

با در نظر گرفتن

$$\beta_n = \frac{(L\chi^{\alpha_i})^n}{\Gamma(۱ + n\alpha_i)}$$

و به منظور به کار بردن قضیه نقطه ثابت باناخ تعمیم یافته نیاز به همگرایی $\sum_{n=۰}^{\infty} \beta_n$ داریم:

$$\sum_{n=۰}^{\infty} \frac{(L\chi^{\alpha_i})^n}{\Gamma(۱ + n\alpha_i)} = E_{\alpha_i}(L\chi^{\alpha_i})$$

که یک تابع میتاگ-لفلر از مرتبه α_i روی $L\chi^{\alpha_i}$ است. بنابراین با توجه به قضیه نقطه ثابت باناخ تعمیم یافته، A_i به یک نقطه ثابت همگرا بوده و یکتاست. □

۳ نتیجه‌گیری

در این مقاله، وجود و یکتایی جواب‌های دستگاه معادلات دیفرانسیل کسری غیر خطی

$$D^{\alpha_i} y_i(t) = f_i(t, y_۱, \ldots, y_n),\ y_i^{(k)} = C_k^i,\ ۱ \leq i \leq n,\ ۱ \leq k \leq m_i$$

که D^{α_i}، $m_i < \alpha_i < m_i + ۱$ بیان کننده مشتق کسری کاپتو می‌باشد، بررسی شده است.

مراجع

[۱۹] ح. جعفری، مقدمه‌ای بر معادلات دیفرانسیل کسری، چاپ اول، دانشگاه مازندران، بابلسر، ۱۳۹۲.

[20] A. D. Obembe, H. Y. Al-Yousef, M. E. Hossain, S. A. Abu-Khamsin, Fractional derivatives and their applications in reservoir engineering problems, *a review. J. Pet. Sci. Eng.*, **157**, 312-327 (2017).

[21] F. Mainardi, *Fractional Calculus and Waves in Linear Viscoelasticity: An Introduction to Mathematical Models*, Imperial College Press, London (2010).

[22] R. L. Magin, *Fractional Calculus in Bioengineering*, Begell House, West Redding (2006).

اولین همایش ملی ریاضیات زیستی
۲۱ و ۲۲ اسفندماه ۱۳۹۷
همایش ملی ریاضیات زیستی

مقایسه‌های چندگانه در کارآزمایی‌های بالینی ناپست‌تری

سعیده حاجبی خانیکی [۱]، دانشجوی دکتری آمار زیستی، مرکز تحقیقات عوامل اجتماعی موثر بر سلامت، دانشگاه علوم پزشکی مشهد، مشهد، ایران، دکتر محمدتقی شاکری استاد آمارزیستی، مرکز تحقیقات عوامل اجتماعی موثر بر سلامت، دانشگاه علوم پزشکی مشهد، مشهد، ایران، احسان برادران سیرجانی کارشناس ارشد آمار ریاضی، دانشگاه آزاد اسلامی واحد مشهد، مشهد، ایران، نگار سنگ سفیدی دانشجوی کارشناسی ارشد آمار زیستی، مرکز تحقیقات عوامل اجتماعی موثر بر سلامت، دانشگاه علوم پزشکی مشهد، مشهد، ایران

چکیده: در سال‌های اخیر به علت پیشرفت های پزشکی و دارویی نیاز به استفاده از کارآزمایی‌های بالینی ناپست‌تری برای مقایسه ی درمان‌های جدید کم هزینه تر با درمان استاندارد قبلی بیشتر احساس می‌شود. در این مقاله روش‌های مقایسه چندگانه درمان استاندارد با درمان های جدید برای ارزیابی اثربخشی آنها مورد بررسی قرار گرفته است. اگر ارزیابی درمان ها براساس یک یا چند متغیر پاسخ باشد، روش‌های پیشنهادی متفاوت است. در این مقاله آزمون‌های اجتماع-اشتراک، اشتراک-اجتماع، رویه هلم، رویه هوچبرگ، روش آزمون بسته و اصلاحات آنها به عنوان روش هایی که خطای نوع اول کلی خانواده ی مقایسه‌ها را در کارآزمایی بالینی ناپست تری کنترل میکنند، پیشنهاد شده‌اند.

واژه‌های کلیدی: کارآزمایی بالینی ناپست‌تری، مقایسه‌های چندگانه، رویه هلم، رویه هوچبرگ، روش آزمون بسته.

۱ مقدمه

کارآزمایی‌های بالینی با گروه کنترل فعال (درمان استاندارد) که در آن هدف این است تا نشان دهیم درمان جدید به طور غیرقابل قبولی بدتر از درمان استاندارد نیست را کارآزمایی‌های "ناپست‌تری [۱]" می‌نامند(۲۳). امروزه نیاز به استفاده از کارآزمایی‌های بالینی ناپست‌تری در تحقیقات پزشکی و داروسازی به دلیل ماهیت طرح و امکان جایگزینی درمان های کم هزینه تر با اثرات جانبی کمتر بیشتر از گذشته احساس می‌شود.

اساس کار کارآزمایی‌های بالینی ناپست‌تری به این صورت است که اگر حدپایین فاصله اطمینان برای تفاوت متغیر پاسخ بین

[1] Non-inferiority

[۱] سعیده حاجبی خانیکی: hajebis971@mums.ac.ir

گروه‌های مورد مقایسه به طور معنی داری کمتر از یک مقدار از پیش تعیین شده (δ) نباشد، آنگاه با خطای نوع اول α می‌توان اظهار داشت که درمان جدید به طور غیر قابل قبولی بدتر از درمان استاندارد نیست یا بعبارت دیگر دلیلی بر رد فرض ناپست‌تری نخواهیم داشت. مقدار δ یا میزان تفاوت قابل قبول بین درمان استاندارد و درمان جدید نیز از مطالعات مشابه یا متا آنالیزهای انجام شده بدست می‌آید(۲۴).

مقایسه‌های چندگانه در هر کارآزمایی بالینی با مشکلاتی مواجه است. این مشکلات جنبه‌های زیادی دارد، از جمله وجود چندین درمان، چندین متغیر پاسخ برای اندازه گیری اثربخشی درمان و چندین بازه زمانی. اما همه منجر به یک مشکل مشابه می‌شوند: عدم به کارگیری روش درست برای مقایسه چندگانه، احتمال اینکه به اشتباه کارآیی یک دارو یا درمان را بپذیریم، را افزایش می دهد.

در کارآزمایی‌های ناپست‌تری نقش فرضیه‌های صفر و جایگزین به نوعی معکوس می‌شود. در آزمون‌های برتری،[۲] که هدف تایید برتری یک درمان جدید است، خطای نوع اول احتمال نتیجه برتری یک درمان زمانیکه اثرات درمان یکسان است، می‌باشد. در این مطالعات، خطای نوع اول احتمال نتیجه گیری ناپست‌تری است وقتی که کنترل فعال به طور قابل توجهی بهتر از درمان جدید است.

کنترل میزان خطای نوع I می‌تواند به روش‌های مختلفی تعریف شود. معروف ترین روش برای کارآزمایی‌های بالینی، کنترل خطای خانوادگی[۳] (FWE) است که احتمال رد حداقل یک فرضیه درست صفراست. در مورد آزمون چندین متغیر پاسخ در کارآزمایی‌های ناپست‌تری، FWE به معنای احتمال نتیجه گیری حداقل یک ناپست‌تری در یک متغیر پاسخ است زمانی که ناپست‌تری واقعا درست نباشد. در بخش ۲ به بررسی روش های مختلف برای مقایسه درمان استاندارد با چندین درمان در حالتی که فقط یک متغیر پاسخ داریم و در بخش ۳ در حالتی که چندین متغیر پاسخ داریم، پرداخته می شود.

۲ مقایسه چندین درمان با درمان استاندارد

تیمارهای آزمایشی چندگانه را می‌توان با یک درمان استاندارد با کارایی شناخته شده مقایسه کرد. مثلا هنگامی که دوزهای متعدد از یک دارو در نظر گرفته می‌شود یا زمانی که درمان‌های متعدد تجربی با یک درمان استاندارد برای مقایسه کارآیی آن ها طرح ریزی می شود.(۲۵) در هر مورد، خانواده ی فرضیاتی که برای آن نرخ خطای نوع I باید کنترل شود، بایستی در نظر گرفته شده و برای کنترل FWE تعدیل مناسب انجام شود. دو ساختار کلی موجود به صورت زیر است:

۱. پیدا کردن زیرمجموعه‌ای از تیمارها که ناپست‌تر از کنترل نیستند. در این حالت فرضیه‌ها به این صورت نوشته می‌شوند:

$$H_0 : \mu_C - \mu_{E,i} > \delta_i \qquad \text{برای تمام } i\text{ها}$$
$$Ha: \mu_C - \mu_{E,i} \leq \delta_i \qquad \text{برای حداقل یک } i$$

که در آن μ_C میانگین متغیر پاسخ در گروه کنترل و $\mu_{E,i}$میانگین متغیر پاسخ در گروه درمانی i ام و δ_i تفاوت قابل قبول بین آن هاست.

۲. نشان دادن اینکه تمام تیمارها ناپست‌تر از گروه کنترل نیستند: در این حالت فرضیه‌ها به این صورت نوشته می‌شوند:

$$H_0 : \mu_C - \mu_{E,i} > \delta_i \qquad \text{برای حداقل یک } i$$
$$Ha: \mu_C - \mu_{E,i} \leq \delta_i \qquad \text{برای تمام } i\text{ها}$$

در مورد اول آزمون اجتماع–اشتراک (intersection–union) و رویه‌های مقایسه‌های چندگانه استاندارد مورد نظر قرار میگیرند.

[2] Superiority
[3] Familywise error

در مورد دوم آزمون اشتراک–اجتماع (union–intersection) نتیجه شده و غالبا هیچ تعدیلی برای خطای نوع I اسمی نیاز نیست.((۲۶))

۱.۲ تیمارهای غیرترتیبی : انتخاب زیرمجموعه

زمانی که هدف مقایسه‌ی چندگانه‌ی تیمارهای آزمایشی(درمان های مختلف) باشد که ترتیب در تیمارها در آزمون اجتماع–اشتراک اهمیت نداشته باشد، راههای زیادی برای کنترل FWE وجود دارد.

۱. یک راه این است که مقایسه‌های مختلف تیمارها با درمان استاندارد را مانند مقایسه‌های غیرمرتبط در نظر گرفته و هیچ تصحیحی انجام ندهیم. اما وقتی که یک گروه کنترل برای مقایسه با چندین تیمار آزمایشی در کارآزمایی تاییدی[۴] استفاده می‌شود این روش نامناسب است زیرا فقط CWE [۵] را کنترل کرده و FWE کنترل نخواهد شد. یعنی تصحیحی برای مقایسه‌ها به صورت چندگانه اعمال نمیکند.

۲. روش دیگر استفاده از تصحیح بن–فرونی می‌باشد که در آن هر یک از I تیمار آزمایشی با درمان استاندارد در سطح $\frac{\alpha}{I}$ مقایسه می‌شود. در اینجا فرضیه ناپست‌تری زمانی نتیجه می‌شود که حد بالایی فاصله اطمینان در سطح یاد شده کمتر از δ_i باشد. این روش خطای نوع اول را کنترل کرده و فواصل اطمینان همزمان تولید می‌کند.

۳. رویه‌های قوی تری مانند رویه هلم[۶] هم وجود دارد. این رویه نیز FWE را کنترل کرده و بعلاوه بطور یکنواختی قدرتمند تر از رویه بن–فرونی است. کارکرد این رویه به این صورت است که پی–مقدارهای مقایسه‌های مختلف را بدست آورده و آنها را مرتب می‌کنیم. سپس کوچکترین پی–مقدار $p(۱)$ با $\frac{\alpha}{I}$ مقایسه می‌شود. اگر $p(۱) < \frac{\alpha}{I}$ فرضیه مربوط به تیمار و مقایسه‌ی نظیر آن رد شده و کوچکترین پی–مقدار بعدی با $\frac{\alpha}{I-۱}$ مقایسه شده و این روند ادامه پیدا می‌کند. اگر در یکی از مراحل رابطه $(p(i) < \frac{\alpha}{I-i+۱})$ برقرار نباشد، آنگاه رویه متوقف شده و نتیجه‌ای در مورد سایر مقایسه‌های آزمون نشده نمی‌توان گرفت. از آنجایی که در کارآزمایی‌های ناپست‌تری نتیجه گیری بر اساس فواصل اطمینان صورت می‌گیرد، باید با روشی مشابه کار کرد. به این صورت که برپایه ی حفظ خطای نوع اول یکطرفه‌ی $\frac{\alpha}{۲}$ اگر ناپست‌تری بر اساس حداقل یکی از فواصل اطمینان دو طرفه‌ی $۱۰۰ \times (۱ - \alpha/I)$□ نتیجه گرفته شود، رویه به مرحله دوم می‌رود. در مرحله دوم اگر ناپست‌تری بر اساس حداقل دوتا از فواصل اطمینان دو طرفه‌ی $۱۰۰ \times (۱ - \alpha/(I-۱))$□ نتیجه گرفته شود، رویه به مرحله‌ی بعدی می‌رود و به همین ترتیب مراحل طی می‌شود. هر زمان این رویه متوقف شود (بعبارت بهتر ناپست‌تری برای حداقل j تا از I مقایسه برقرار نباشد) آنگاه فقط برای $j - ۱$ مقایسه از مراحل قبل نتیجه‌ی نهایی گرفته می‌شود(۲۷).

۴. یکی از روش‌های معمول دیگر هم برای مواقعی که هدف مقایسه‌ی بیش از یک گروه آزمایشی با کنترل فعال باشد، استفاده از آزمون دانت[۷] است. این آزمون فاصله اطمینانی را به فرم $\overline{x}_i - \overline{x}_c \pm d_\alpha(I-۱, f) \times se$ تولید می‌کند که در آن f درجه آزادی و برابر با تعداد کل مشاهدات منهای تعداد گروههای آزمایشی و I تعداد گروههای آزمایشی و کنترل است. (۲۵)

۲.۲ تیمارهای ترتیبی : انتخاب زیرمجموعه

زمانیکه ترتیب خاصی بین تیمارها وجود داشته باشد، باید جهت کارایی بیشتر، ساختارهای بیشتری به مقایسه‌ها اضافه شود. مثلا دوزهای چندگانه‌ی یک تیمار آزمایشی کارآیی متفاوتی دارند به این صورت که دوزهای بیشتر کارآیی بالاتری دارد. اگر هدف مقایسه‌ی دوزهای

[4] Confirmatory trial
[5] Comparison-wise error
[6] Holm procedure
[7] Dunnett's test

مختلف با درمان استاندارد جهت تعیین ناپست‌تر بودن تیمارها نسبت به آن باشد، روش دنباله‌ای ثابت [8] را می‌توان استفاده کرد. به این صورت که بالاترین دوز (بالاترین کارآیی) با درمان استاندارد مقایسه میشود، اگر ناپست‌تری نتیجه گرفته شود، دوز بالایی بعدی مقایسه می‌شود و به همین ترتیب ادامه پیدا می‌کند تا زمانیکه یک دوز ناپست‌تر از کنترل نباشد. در این روش تا زمانی که مراحل ادامه پیدا میکند میتوان مقایسه‌ها را در همان سطح α انجام داد. یکی از اصلاحاتی که بر روش دنباله ای ثابت صورت گرفته ، رویه ی فال–بک [9] می‌باشد. در این روش زمانیکه مقایسه‌های اولیه منجر به رد فرض صفر نشود، مقداری از نرخ خطای نوع I برای مقایسه‌های بعدی نگه داشته می‌شود. (۲۸)

۳ ناپست‌تری روی چندین متغیر پاسخ

مقایسه تیمار آزمایشی و درمان استاندارد برپایه ی چندین متغیر پاسخ معمولا در مطالعات کارآزمایی بالینی بسیار مورد توجه است. در واقع برای افزایش کارآیی، بسیاری از اوقات چندین متغیر جهت تصمیم‌گیری نهایی جمع‌آوری شده و در استنباط‌ها مورد استفاده قرار می‌گیرد. اما مساله‌ی مهمی که باید در نظر گرفته شود حفظ سطح کل FWE می‌باشد. گام اول در تحلیل، ایجاد اولویت بین مقایسه‌ها است. مهمترین متغیرها که باید تعدادشان نیز بسیار کم باشد را مقایسه‌های اولیه، گروه بعدی مقایسه‌های ثانویه و به همین ترتیب تمام گروههای ممکن از متغیرها برای مقایسه را تشکیل می‌دهیم.

۱.۳ چندین برآمد در یک خانواده [10]

ابتدا با حالتی شروع می‌کنیم که فقط یک خانواده از متغیرهای پاسخ مهم یعنی متغیرهای پاسخ گروه اولیه را مبنای تصمیم گیری قرار دهیم.

۱. اگر بین پاسخ ها یک ترتیب معنادار از لحاظ اهمیت متغیرها وجود داشته باشد، آنگاه همان روش رویه دنباله ثابت را می‌توان مورد استفاده قرار داد. با این رویه، ابتدا ناپست‌تری را برای مهمترین متغیر بررسی کرده و متغیرهای دیگر لحاظ نمیشوند مگر اینکه ناپست‌تری برای متغیر قبلی ثابت شود.

۲. اصلاح روش دنباله ثابت همانطور که اشاره شد روش فال–بک است. با این روش حتی اگر یک یا چند متغیر پاسخ در نتایج استفاده نشود، تمام مقایسه‌ها را می‌توان لحاظ و محاسبه کرد.(۲۹)

۳. روش هلم

۴. هوچبرگ [11] نیز روشی را برپایه روش هلم ارائه کرده که در آن بجای رویکرد رو به عقب از روشی رو به جلو استفاده می‌شود. به عبارت ساده تر ناپست‌تری برای متغیر پاسخ نظیر بزرگترین پی–مقدار که در رابطه‌ی $p(j) \leq \frac{\alpha}{(I-j+1)}$ صدق کند و تمام متغیرهای پاسخ با پی–مقدارهای کوچکتر از آن پذیرفته می‌شود. این تعریف شامل تمام متغیرهایی خواهد بود که در روش هلم ناپست‌تری را نتیجه می‌دهند.(۳۰)

۵. روش دیگری که FWE را کنترل می‌کند روش آزمون بسته [12] نام دارد. روش آزمون بسته تمام زیر مجموعه‌های ممکن از

[8] Fixed sequence approach
[9] Fall back procedure
[10] Single family
[11] Hochberg
[12] Closed testing procedure
[13] global null hypothesis

فرضیه‌ها را در نظر میگیرد. در اینجا با J متغیر پاسخ، $\sum_{k=۱}^{j}\binom{j}{k} = ۲^j - ۱$ زیرمجموعه خواهیم داشت که درون هر زیر مجموعه، یک فرضیه صفر فراموضعی[۱۳] به این صورت در نظر گرفته می‌شود:

فرضیه صفر: ناپست‌تری در هیچ کدام از متغیرهای پاسخ زیر مجموعه ی مورد بررسی وجود ندارد.

فرضیه جایگزین: ناپست‌تری در حداقل یکی از متغیرهای پاسخ زیر مجموعه ی مورد بررسی وجود دارد.

و نهایتاً ناپست‌تری در مورد یک برآمد خاص در صورتی نتیجه گرفته می‌شود که آن متغیر در تمام زیرمجموعه هایی که در بردارنده ی آن هستند، نتیجه گرفته شود(۳۱)

توجه ۱.۳. در استفاده از روش آزمون بسته مطلوب است که آزمون‌ها α–فرسا[۱۴] باشد. یعنی در آزمون فرضیه صفر در هر زیرمجموعه، خطای نوع اول دقیقا مقدار α بوده و کمتر از آن نباشد. بعنوان مثال در روش بن–فرونی این مساله لحاظ نمی شود. زیرا خطای نوع اول در یک زیرمجموعه ی مورد نظر دارای کران بالای $\sum_{k=۱}^{I}\alpha_k I(k)$ می‌باشد که $I(k)$ فقط زمانی مقدار ۱ را می‌پذیرد که متغیر پاسخ Iام در زیر مجموعه مورد بررسی باشد. در این حالت فقط در مورد زیرمجموعه ای که شامل تمام متغیرهای پاسخ است این مجموع برابر α خواهد بود.

۲.۳ چندین متغیر پاسخ در چندین خانواده

اگر چند خانواده از متغیرهای پاسخ مهم بوده (برآمدهای گروه اولیه، ثانویه و ...). و چند خانواده از برآمدها مبنای تصمیم‌گیری قرار گیرد، برای این موضوع ۲ استراتژی را در نظر میگیریم که عبارتند از استراتژی کنترل پیاپی[۱۵] و کنترل موازی[۱۶].

در کنترل پیاپی باید ابتدا تمام برآمدها در یک خانواده معنی دار شوند و سپس خانواده بعدی مورد بررسی قرار گیرد (مراحل کار این روش مانند روش دنباله ثابت است). آزمون‌های درون هر خانواده میتواند از هر یک از روش‌های بن–فرونی، هلم، فال–بک و .. که FWE را در سطح α حفظ میکند، انجام شود.

در روش کنترل موازی، اگر حداقل یک متعیر پاسخ در یک خانواده معنی دار شود میتوان خانواده ی بعدی را نیز مورد بررسی قرار داد. با این وجود اگر ناپست‌تری برای بعضی از متغیرهای پاسخ در یک خانواده فقط نتیجه گرفته شود، متغیرها در خانواده ی بعدی در سطح خطای نوع اول کوچکتری آزمون می‌شوند. همچنین آزمون درون یک خانواده از متغیرها در این روش توان کمتری نسبت به روش کنترل پیاپی دارد. (۳۲)

۴ نتیجه‌گیری

در این مقاله با توجه به اهمیت روز افزون کارآزمایی‌های ناپست‌تری، به بررسی روش‌های مقایسه‌های چندگانه در این نوع کارآزمایی ها پرداخته شد. در این نوع کارآزمایی‌ها که مبنای تصمیم‌گیری بر اساس دو یا چند متغیر پاسخ مهم می‌باشد روش‌های مختلفی ارائه شد که تمام آنها بر پایه‌ی حفظ سطح خطای نوع اول می‌باشند.

مراجع

[23] FOOD, U. & ADMINISTRATION, D. (2016). Non-inferiority clinical trials to establish effectiveness. Guidance for industry.

[14] α-exhaustive
[15] serial gatekeeping
[16] parallel gatekeeping

[24] SCHUMI, J. & WITTES, J. T.(2011). Through the looking glass: understanding non-inferiority. *Trials,* 12**,** 106.

[25] DUNNETT, C. W. (1964). New tables for multiple comparisons with a control. *Biometrics,* 20**,** 482-491.

[26] BERGER, R. L. & HSU, J. C. (1996) Bioequivalence trials, intersection-union tests and equivalence confidence sets. *Statistical Science,* 11**,** 283-319.

[27] HOLM, S. 1979. A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics***,** 65-70.

[28] WIENS, B. L. (2003). A fixed sequence Bonferroni procedure for testing multiple endpoints. *Pharmaceutical Statistics,* 2**,** 211-215.

[29] WIENS, B. L. & DMITRIENKO, A. (2005). The fallback procedure for evaluating a single family of hypotheses. *Journal of Biopharmaceutical Statistics,* 15**,** 929-942.

[30] HOCHBERG, Y. (1988). A sharper Bonferroni procedure for multiple tests of significance. *Biometrika,* 75**,** 800-802.

[31] MARCUS, R., ERIC, P. & GABRIEL, K. R. (1976). On closed testing procedures with special reference to ordered analysis of variance. *Biometrika,* 63**,** 655-660.

[32] DMITRIENKO, A., WIENS, B. L., TAMHANE, A. C. & WANG, X. (2007). Tree□structured gatekeeping tests in clinical trials with hierarchically ordered multiple objectives. *Statistics in medicine,* 26**,** 2465-2478.

معادلات ساختاری بیزی

نجمه شکیبایی [۱]، کمیته تحقیقات دانشجویی، گروه آمار زیستی، دانشکده علوم پزشکی دانشگاه اصفهان، اصفهان، عقیل علایی
گروه آمار، دانشکده علوم ریاضی، دانشگاه صنعتی اصفهان

چکیده: مدل‌سازی معادلات ساختاری بیزی [۱] نسل دوم تجزیه و تحلیل آماری چند متغیره بسیار نیرومند از خانواده رگرسیون چند متغیری است که با اعتبار و قابلیت اعتماد به نمرات مشاهده شده مورد استفاده قرار می‌گیرد. به عبارتی الگوهای فرضی از ارتباطات مستقیم وغیرمستقیم در میان یک مجموعه از متغیرهای مشاهده شده و پنهان بررسی می‌شود که این الگوهای فرضی در مطالعات بالینی بسیار به چشم می‌خورد. این مقاله، روش و کاربرد مدل سازی معادلات ساختاری بیزی همراه با یک مثال در پژوهش های بالینی را معرفی نموده است.

واژه‌های کلیدی: معادلات ساختاری بیزی، تجزیه و تحلیل آماری، بیز

۱ مقدمه

مدل‌سازی معادله ساختاری دیدگاهی است که در آن الگوهای فرضی از ارتباطات مستقیم و غیرمستقیم در میان یک مجموعه از متغیرهای مشاهده شده و پنهان بررسی می‌شود. هم‌چنین یک متغییر می‌تواند همزمان هم نقش پیش‌بینی‌کننده و هم نتیجه را ایفا کند. کاربرد اصلی آن در موضوعات چند متغیره‌ای است که نمی‌توان آنها را به شیوه دومتغیری با در نظر گرفتن هربار یک متغیر مستقل با یک متغیر وابسته انجام داد. از دیگر کاربردهای مدل سازی معادلات ساختاری می‌توان به موارد تجزیه وتحلیل مدل با سازه‌های پنهان، اجرای تحلیل عاملی تاییدی، تجزیه رگرسیون با مشکلات چندهم‌خطی، تجزیه و تحلیل مسیر با چند متغیر وابسته، آزمون همبستگی و کوواریانس دریک مدل، مدل سازی روابط داخلی بین متغیرها به طور هم‌زمان دریک مدل، مدل‌سازی و آزمون متغیرهای میانجی در یک مدل و تجزیه و تحلیل متغیرهای تعدیلگر در یک مدل زیر اشاره نمود.(۳۳)

اهمیت این تکنیک در پژوهش های بالینی از آنجاست که غالباً در این حوزه از مطالعات، پژوهشگران به بررسی روابط بین متغیرهای مختلف در قالب مدل یا شبکه‌ای از روابط می‌پردازند؛ بنابراین آنان، مبتنی بر فرضیه‌های خود در مورد روابط بین متغیرها، شمای کلی از این روابط را در قالب مدلی از پیش ساخته طراحی می‌نمایند(۳۴). در چنین وضعیتی پژوهشگر با کمک مدل‌سازی معادلات ساختاری

[1] BSEM

[۱] نجمه شکیبایی: najmeh.shakibaei@gmail.com

بیزی و با کمک نرم افزارهای موجود سوالات پیش رو را پاسخ می‌دهد.

از اولین مطالعات بیزی معادلات ساختاری می‌توان به شاینز و همکاران(۳۵)؛ انصاری و جدی(۳۶)، دانسون(۳۷)، لی(۳۸)، اشاره کرد. همچنین با گسترش نرم افزارها از مطالعات کاربرد مدلسازی معادلات ساختاری در حوزه بالینی نیز می‌توان به مدل‌سازی معادلات ساختاری بیزی: کاربرد سلامت و شاخص سلامت اشاره نمود(۳۹؛ ۴۰). با توجه به مفید بودن این روش آماری در حوزه‌های بالینی بر آن شدیم تا مرور مختصری با استفاده از منابع مطالعه شده در این زمینه به معرفی روش معادلات ساختاری با استفاده از نرم افزار وین باگ[۲] بپردازیم.

۲ متن اصلی

اطلاعات این مطالعه مروری با استفاده از مرور و خلاصه سازی مقالات و کتب مرتبط و از طریق جستجوی هدف‌مند کتابخانه‌ای جهت دستیابی به کتب و منابع الکترونیکی شامل موتور جستجوگر گوگل و بانک‌های اطلاعاتی جهت دستیابی به کتب و مقالات به دست آمد. نتایج پژوهشگران عرصه علوم انسانی و آموزش علوم سلامت قادرند تا با استفاده از مدل‌سازی معادلات ساختاری به مناسب بودن مدل‌های مفهومی یا کاربرد آن در جامعه مورد مطالعه پی‌ببرند. یافته‌های این مطالعه مروری در قالب معرفی مفهوم و گامهای کلی که پژوهشگران لازم است مدنظر قرار دهند ارائه شده است.

۱.۲ تعریف مدل‌سازی معادلات ساختاری

مدل معادله ساختاری اساساً ترکیب مدل‌های مسیر و مدل‌های تحلیل عاملی تاییدی است. تکنیک‌های آماری مدل سازی معادلات ساختاری از طریق محاسبه واریانس مشترک یا عمومی بین متغیرهای آشکار این کار را انجام می شود. مدل سازی معادلات ساختاری همانند تحلیل عاملی نقش تفکیک واریانس مربوط به هر شاخص را به عهده دارد(۳۴). اطلاعات مورد نیاز برای تحلیل در قالب پارامترهای مدل در اختیار نرم افزار قرار داده می شوند متغیرها و پارامترهای موجود در تحقیق شامل:

تعریف ۱.۲ (متغیر پنهان یا سازه ها یا عامل ها). متغیرهایی که نمی‌توان آنها را مستقیماً مشاهده یا مورد سنجش قرار داد. همچون عشق، نفرت، رضایت، کیفیت و . . .

تعریف ۲.۲ (متغیر آشکار یا مشاهده شده). متغیرهایی که مستقیماً اندازه‌گیری می‌شوند. مانند قد، وزن، فشارخون و . . . این متغیرها رابه منظور تعریف یا استنباط در مورد متغیر پنهان به کار می‌بریم.

تعریف ۳.۲. متغیرهای وابسته و مستقل متغیرها چه آشکار و چه پنهان، همچنین می‌توانند به عنوان متغیرهای مستقل و وابسته تعریف شوند.

تعریف ۴.۲ (متغیر مستقل یا برون زا). متغیرهایی هستند که تحت تاثیر متغیرهای موجود درمدل نیستند. این متغیرها حداقل یک مسیر به متغیردیگروارد می کنند.

تعریف ۵.۲ (متغیر وابسته یا درون زا). متغیرهایی هستند که مقادیرآنها توسط مدل برآورد می شود. این متغیرها حداقل یک مسیررا ازمتغیردیگردریافت می کنند.

ترکیب اصلی معادلات ساختاری کلی را می‌توان به دو خرده مدل تقسیم بندی کرد:

۱. مدل اندازه گیری

۲. مدل ساختاری.

[2] Win Bugs

مدل اندازه‌گیری ارتباط بین متغیرهای مشاهده شده و مشاهده نشده را تعریف می‌کند. به عبارت دیگر، بر مبنای اندازه‌گیری، پژوهش‌گر تعریف می‌کند که کدام متغیرهای مشاهده‌پذیر می‌توانند جهت اندازه‌گیری متغیرهای پنهان مورد استفاده قرار بگیرند. بنابراین مدل اندازه‌گیری، مدل تحلیل عاملی تاییدی را مشخص می‌کند که به وسیله آن هر اندازه روی عامل خاصی بارگذاری می‌شود. درمقایسه، مدل ساختاری، ارتباط بین متغیرهای مشاهده نشده را تعریف می‌کند (۳۳) .یکی از رایج‌ترین روش های برآورد، روش بیشینه درستنمایی[۳] است. در اکثر برنامه های کامپیوتری تحلیل مدل های معادلات ساختاری است و نیز اکثر مطالعات منتشر شده استفاده از این روش را گزارش نموده‌اند. از آنجایی که این روش برآورد، بر مبنای پیش فرض‌هایی همچون مطالعه روی حجم نمونه‌های بزرگ، پیوسته بودن متغیرها استوار است پژوهشگران لازم است از برقرار بودن این شرایط اطمینان حاصل نمایند(۳۴). اما در مقابل روش بیزی وابستگی کمتر به این فرضیات را دارد و جوابگو برای نمونه‌های کوچک می‌باشد.(۴۲)

۳ تشریح مراحل انجام معادلات ساختاری بیزی

۱.۳ تدوین مدل

فرض کرده $Y = (Y_1, Y_2, \ldots, Y_n)$ ماتریس داده‌ها باشند. معادله اندازه گیری به صورت $Y_i = \lambda\omega_i + \varepsilon_i$ در نظر گرفته می‌شود. که Y_i معرف بردار $p \times 1$ از متغیرهای مشاهده شده برای توصیف بردار تصادفی $q \times 1$ متغیرهای پنهان ω_i ، λ ماتریس $p \times q$ از ضرایب رگرسیون Y_i روی ω_i و ε_i بردار تصادفی $p \times 1$ ازخطای اندازه گیری است که متغیر پنهان ω_i به صورت زیر در نظر گرفته می‌شود.

$$\omega_i = \begin{pmatrix} \eta_i \\ \xi_i \end{pmatrix} \tag{۱.۳}$$

η_i معرف بردار تصادفی $q_1 \times 1$ از متغیرهای پنهان وابسته و ξ_i معرف بردار تصادفی $q_2 \times 1 (q_2 = q - q_1)$ از متغیرهای پنهان مستقل است. بنابراین معادله ساختاری به صورت زیر در نظر گرفته می‌شود. که β ماتریس رابطه متغیرهای پنهان درونی، Γ ماتریس $q_1 \times q_2$ رابطه میان متغیرهای پنهان درونی و بیرونی و σ بردار تصادفی $q_1 \times 1$ از باقیمانده‌ها است.

$$\eta_i = \beta\eta_i + \Gamma\xi_i + \sigma_i \tag{۲.۳}$$

۲.۳ تشخیص مدل

در مساله تشخیص این سوال مطرح می شود که آیا بر اساس داده‌های نمونه‌ای(شامل شده درماتریس کواریانس نمونه‌ای s) و مدل نظری (تعریف شده بوسیله ماتریس کواریانس جامعه)، می‌توان مجموعه منحصر به فردی از برآورد پارامترها یافت؟

۳.۳ برآورد بیزین پارامترها

فرض کرده M یک SEM دلخواه از پارامترهای ناشناخته θ و داده‌های مشاهده شده با حجم نمونه n باشد. در رویکرد بیزی θ تصادفی با توزیع پیشین $p(\theta)$ است. برآورد بیزین بر اساس داده‌های مشاهده شده Y و توزیع پیشین θ صورت می‌گیرد. یک مسئله انتخاب تابع پیشین مناسب است. مراجعه به (۴۲)

برآوردهای بیزی از θ را معمولاً به عنوان میانگین یا مد تابع پسین $\theta|y$ در نظر می‌گیریم. اگر بتوانیم به اندازه کافی تعدادی از مشاهدات $\theta|y$ را شبیه سازی کنیم می‌توانیم میانگین و آماره‌های مفید دیگر را از طریق مشاهدات شبیه سازی به دست آوریم. برای حل کردن مشکل، کافی است روش‌های قابل اعتماد و کارآمد برای استخراج مشاهدات از توزیع پسین به دست آوریم. به جای کار

[3] Maximum likelihood

کردن با تابع چگالی $p(\theta|y)$ بر تابع چگالی $p(\theta,\eta|y)$ متمرکز می‌شویم (η مجموعه متغیرهای پنهان). تابع چگالی $\theta,\eta|y$ فرم بسته ای ندارد و دسترسی به آن سخت است، براساس مجموعه داده کامل (y,η) توزیع شرطی $(\theta|\eta,y)$ و توزیع $(\eta|\theta,y)$ به آسانی استخراج می‌شود. . بنابراین بعضی روش‌های $MCMC$ اجرا می‌کنیم تا مشاهدات را از طریق مشاهدات تکراری چگالی شرطی آن‌ها $(\theta|\eta,y)$ و $(\eta|\theta,y)$ به دست آوریم. در نهایت توزیع توام از (θ^t,η^t) بعد از تعداد تکرار کافی به توزیع پسین $(\theta|\eta,y)$ همگرا می‌شود. استنتاج‌های آماری از مدل بر اساس نمونه شبیه سازی شده از مشاهدات $p(\theta|\eta,y)$ صورت می‌گیرد. . به عبارت دیگر $\{(\theta^t,\eta^t);t=۱,۲,\ldots,T\}$ برآورد θ وانحراف استاندار آن

$$\hat{\theta}=T^{-۱}\sum_{i=۱}^{T}\theta^i \tag{۳.۳}$$

$$Var(\theta|Y)=(T-۱)^{-۱}\sum_{i=۱}^{T}(\theta^t-\hat{\theta})(\theta^t-\hat{\theta})^{-۱} \tag{۴.۳}$$

زمانی که T به سمت ∞ می‌رود $\hat{\theta}$ به $E(\theta|y)$ میل می‌کند به عبارت دیگر: برای هر ، Y_i و W_i نشان دهنده بردار از متغیرهای پنهان باشد، $E(W_i|Y_i)$ میانگین پسین است. . برآوردگر بیز W_i از طریق: $E(W_i|Y_i)=T^{-۱}\sum_{i=۱}^{T}W_i$ که W_i امین ستون از η^t است. این برآورد W_i می‌تواند برای آنالیز داده‌های پرت و باقی مانده‌ها و ارزیابی نیکویی برازش مدل مورد استفاده قرار گیرد(۴۲).

۴.۳ آزمون مدل

معیارهای تشخیص همگرایی

معیارهای تشخیصی اولیه شامل خطای مونت کارلو، نمودار تریس، نمودار تکرارها در مقابل مقادیر تولید شده را بدست می‌دهد. نوساناتی حول یک محور بدون روند یا تناوب، گویای وضعیت خوب زنجیره است.همچنین می‌توان از $densitybutton$ که یک برآورد تقریبی کرنل از تابع چگالی تابع احتمال پسین پارامترها فراهم می‌کند و نمودار اتوکرولیشن[۴] استفاده نمود. برای دستیابی به معیار های دقیق‌تر همگرایی می‌توان از آزمون گِلمن رابین و آزمون رفتری و لویس استفاده نمود.(۴۳)

شاخص‌های نیکویی برازش

انواع شاخص‌های نیکویی برازش مورد استفاده عبارتند از معیار اطلاع دویانس و ارزیابی پیش بینی کننده پسین. معیار اطلاع دویانس توسط اشپیگل هالتر به عنوان اندازه ای برای مقایسه مدلها و بررسی کفایت مدل معرفی شد. مقادیر کمتر DIC اشاره به برازش بهتر دارد.

ارزیابی پیش بینی کننده پسین

نمونه‌های تولید شده از توزیع پیش بینی کننده پسین با داده های واقعی مقایسه می‌گردد. دراین تکنیک اگر مدل برازش خوبی به داده‌ها داشته باشد داده‌های شبیه‌سازی شده تحت مدل باید نزدیک به داده‌های مشاهده شده باشد. مقادیر نزدیک ۵.۰ ارزیابی کننده پسین نشان دهنده نزدیک بودن توزیع داده‌های واقعی و داده‌های شبیه سازی شده به یکدیگر است. در حالی که مقادیر نزدیک به صفر یا یک این احتمال بیانگر وجود اختلاف بین آنهاست.

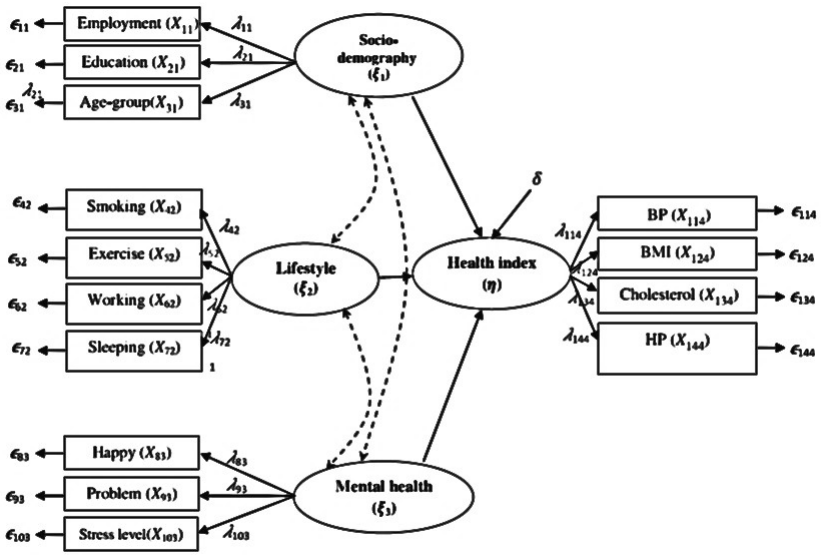
[4] Autocorrelation

۵.۳ اصلاح مدل

اگر برازش یک مدل نظری به قوتی نبود که انتظار داشتیم، آنگاه گام بعدی،اصلاح مدل وارزیابی مدل جایگزین و اصلاح شده است.و در نهایت به تفسیر نتایج می‌پردازیم.

۴ بررسی یک مطالعه

هدف این مطالعه مدل سازی شاخص سلامت بر اساس معادله ساختاری بیزی است. مطالعات مختلف نشان می‌دهند که بسیاری از عوامل بر سلامت افراد از جمله ویژگی‌های جمعیت شناسی، سبک زندگی سلامت روان و... تاثیر می‌گذارند. داده‌های مطالعه از ۵۰۳۵ نفر افراد بزرگتر از ۱۸ سال در$huluLangat$(منطقه‌ای در مالزی) جمع آوری شده است. متغیرهای پنهان این مطالعه ویژگی‌های جمعیت شناختی، سبک زندگی و سلامت ذهن است. مدل سازی معادلات این مطالعه برای $i = ۱, ۲, \ldots, n$ به صورت زیر است.



شکل ۱:

که در آن λ ماتریس ضرایب عاملی به دست آمده از رگرسیون Y_i بر ω_i است.

$$y_i = \lambda\omega_i + \varepsilon_i$$

بنابراین متغیر های پنهان به صورت $\omega_i = (\eta, \xi_{۱i}, \xi_{۲i}, \xi_{۳i})^{-۱}$ و معادله ساختاری برای توضیح روابط میان متغیرهای پنهان به صورت $\eta_i = \Gamma\xi_i + \sigma_i$ است.

نتایج این مطالعه نشان می‌دهد که میان شاخص سلامت و ویژگی‌های جمعیت شناختی رابطه معنادار وجود دارد هم‌چنین میان شاخص سلامت و سبک زندگی اما میان شاخص سلامت و سلامت روان رابطه معناداری یافت نشد که ممکن است تحت تاثیرمتغیرهای مهم دیگر چون رنج و مشقت، جنون جوانی، افسردگی، اضطراب و اختلال اجتماعی و... باشد و باید آن‌ها نیز در مدل گنجانده شوند(۴۴).

۵ نتیجه گیری

در مجموع، مدل یابی معادلات ساختاری به دلیل دارا بودن قابلیت های متعدد و نیز غلبه بر محدودیت های روش های کلاسیک، کاربردهای بسیاری در مطالعات بالینی دارد. از آنجا که پدیده های انسانی غالباً دارای علت های چندگانه بوده و مطالعه آنها مستلزم لحاظ کردن متغیرها و سازه های متعدد و در برخی کوریت‌ها به دلیل نحوه جواب دادن افراد به سوالات ممکن است با تعداد زیادی داده گمشده مواجه شویم بنابراین در این موارد رویکرد بیزی پاسخ بهتری می‌دهد. همچنین با توجه به استفاده از داده ها برای تازه کردن اطلاعات پیشین موجود از پارامترو قرار گرفتن اطلاعات تازه شده در قالب توزیع پسین پارامتر، وابستگی کمتر به فرضیات، و هم چنین زمانی که تعداد عوامل زیاد می‌شود، کشف ارتباط بین متغیرها پیچیده تر می‌شود. در این شرایط معمولا مدل بندی بیزی در تفکیک واریانس‌ها موفق عمل می‌کند.(۴۵؛ ۳۴) امید است پژوهشگران فعال در حوزه بالینی بر اصول علمی مربوطه احاطه داشته و بتوانند از قابلیت های این تکنیک ها به طور مناسبی در مطالعات خود بهره مند گردند.

مراجع

[33] Lomax RG, Schumacker RE. A beginner’s guide to structural equation modeling. psychology press; 2004 Jun 24.

[34] Alavi,M, Structural Equation Modeling (SEM) in Health Sciences Education Researches: An Overview of the Method and Its Application , Iranian Journal of Medical Education 2013: 13(6) / 530

[35] R. Scheines, H. Hoijtink, and A. Boomsma, Bayesian estimation and testing of structural equation models, Psychometrika 64 (1999), pp. 37–52.

[36] Ansari, K. Jedidi, and S. Jagpal, A hierarchical Bayesian methodology for treating heterogeneity in structural equation models, Market. Sci. 19 (2000), pp. 328–347.

[37] S.Y. Lee and J.Q. Shi, Bayesian analysis of structural equation model with fixed covariates, Struct. Equ. Model. 7 (2000), pp. 411–430

[38] Dunson DB, Palomo J, Bollen K. Bayesian structural equation modeling. SAMSI TR2005-5. 2005 Jul 27

[39] S.Y. Lee, X.Y. Song, S. Skeviton, andY.T. Hao, Application of structural equation models to quality of life, Struct. Equ. Model. 12 (2005), pp. 435–453

[40] Stojanovski, E. and K. Mengersen, Bayesian Structural Equation Models: A Health Application,

[41] Cook, J., 2012. Conjugate prior relationships.

[42] Xin-Yuan Song and Sik-Yum Lee,Basic and Advanced Bayesian 10. Structural Equation Modeling With Applications in the Medical and Behavioral Sciences

[43] Ntzoufras I. Bayesian modeling using WinBUGS. John Wiley, Sons; 2011 Sep 20.

[44] Ferra Yanuar a , Kamarulzaman Ibrahim b, Abdul Aziz Jemain, Bayesian structural equation modeling for the health index, , Journal of Applied Statistics, 2014

[45] Fon Sim Ong, Kok Wei Khong, Ken Kyid Yeoh, Osman Syuhaily, Othman Mohd. Nor, A comparison between Structural Equation Modelling (SEM) and Bayesian SEM approaches on in-store behaviour, ,

تجزیه و تحلیل پایداری مدل شکار- شکارچی با واکسیناسیون و مهاجرت در شکار

محمد شیرازیان، محمد حسین رحمانی دوست و مرضیه شمس آبادی [۱]**، دانشگاه نیشابور ،نیشابور ،ایران**

چکیده: ددر این مقاله به بررسی یک مدل سه بعدی که از شکار مستعد، شکار واکسینه شده و شکارچی تشکیل شده و فرمول بندی می شود می پردازیم. پاسخ عملکردی از نوع لوتکا- ولترا می باشد. همچنین پویایی دستگاه از قبیل جواب های کراندار، شرایط وجود تعادل و پایداری نقاط تعادلی به صورت ریاضی وار تجزیه و تحلیل خواهد شد.

واژه‌های کلیدی: لوتکا- ولترا، مدل شکار-شکارچی، پایداری.

۱ مقدمه

برای مطالعه ی رفتار پویایی یک مدل ، مدل سازی ریاضی به عنوان یک ابزار مؤثر برای توصیف و تحلیل، مورد استفاده قرار می گیرد. در سال ۱۷۹۸، اقتصاددان بریتانیایی، مالتوس یک مدل نوع گونه ای راشکل داد و سپس توسط ورهاست اصلاح شد. در ابتدا لوتکا و ولترا مدل شکار –شکارچی را پیشنهاد دادند. بعد از آن مدل شکار-شکارچی به یک بخش تحقیقاتی مهم در ریاضیات کاربردی تبدیل شد. بخاطرمدل کرماک –مکندیک روی دستگاه های $SIRS$ ریاضی اپیدمیولوژیکی به یک موضوع جالب تحقیقاتی تبدیل شد. نویسندگان بسیاری در مورد مدل شکار شکارچی مطالعه کرده اند ومقاله هایی در مطبوعات منتشر نموده اند. از جمله جهری که یک مدل شکار شکارچی از نوع لوتکا- ولترا را با وجود بیماری در گونه ی شکار مطرح و پایداری موضعی و سراسری آن را بررسی نمودند (۴۸). در این مقاله، یک مدل سه بعدی با وجود واکسیناسیون و مهاجرت در شکار بررسی می شود.

۲ بررسی مدل ریاضی

مدل شامل دو جمعیّت است. شکار که تراکم جمعیّت آن با $N(t)$ مشخص می شود و شکارچی که تراکم جمعیّت آن با $P(t)$ نشان داده می شود. اینجا t متغییر زمان است. فرضیّات زیر برای ساختن مدل در نظر گرفته می شود.

۱) جمعیّت شکار به طور منطقی با نرخ رشد $r(r > ۰)$ رشد می یابد و ظرفیّت حمل K $(K > ۰)$ در غیاب بیماری،

[۱] مرضیه شمس آبادی: m.shamsabadi.96@gmail.com

واکسیناسیون و شکارگری است. لذا داریم:

$$\frac{dS}{dt} = rS(۱ - \frac{S}{K})$$

۲) شکار واکسینه شده ی $V(t)$ گروه جداگانه ای است و ϕ میزان واکسیناسیون شکار مستعد است θ میزان بی اثر شدن واکسن می باشد.

۳) جمعیّت شکار با وجود واکسیناسیون به دو دسته تقسیم می شود. که شامل شکار مستعد و شکار واکسینه شده است. بنابراین کلّ جمعیّت شکار در لحظه ی t عبارتند از: $N(t) = S(t) + V(t)$.

۴) شکارچی با شکار کردن شکار مستعد و واکسینه شده آشکارا به نتیجه ی مشابهی می رسد با تفاوت بازده ی جستجو که به ترتیب با $P_۲, P_۱$ نشان داده می شود.

۵) پاسخ عملکردی شکارچی از نوع لوتکا –ولترا است.

۶) ضرایب تبدیل شکار مستعد و واکسینه شده به شکارچی به ترتیب با $q_۲, q_۱$ نشان داده می شوند که متفاوت هستند.

۷) جمعیّت شکار دارای نرخ مهاجرت $m_۲, m_۱$ می باشند که به ترتیب مربوط به نرخ مهاجرت شکار مستعد و واکسینه شده است.

۸) نرخ مرگ و میر طبیعی سه گونه ی شکار مستعد، شکار واکسینه و شکارچی متفاوت است و به ترتیب با $d_۳, d_۲, d_۱$ نشان داده می شوند.

از مدل ریاضی با فرضیّات فوق معادلات دیفرانسیل زیر نتیجه می شود:

$$\begin{aligned}
\frac{dS}{dt} &= rS(۱ - \frac{S}{k}) - \phi S + \theta V - p_۱ PS - m_۱ S - d_۱ S,\\
\frac{dV}{dt} &= \phi S - \theta V - p_۲ PV - m_۲ V - d_۲ V,\\
\frac{dP}{dt} &= q_۱ p_۱ PS + q_۲ p_۲ PV - d_۳ P.
\end{aligned} \tag{۱.۲}$$

با شرایط اولیّه:

$$P(۰) = P_۰ \geqslant ۰, V(۰) = V_۰ \geqslant ۰, S(۰) = S_۰ \geqslant ۰،$$

توجه ۱.۲. اگر $\phi = ۰$ آنگاه واکسیناسیون وجود ندارد. بنابراین داریم: $\lim_{t\to\infty} V(t) = ۰$.

توجه ۲.۲. در این مدل بین مرگ ومیر با مهاجرت تفاوت قائل می شویم ولی جالب است توجه داشته باشیم که شرایط مهاجرت در مدل فوق مشابه شرایط مرگ و میر است.

کرانداری جواب ها

قضیه ۳.۲. همه ی جواب های دستگاه (۱.۲) همواره در ناحیه ی مثبت $R^۳_+$ هستند.

اثبات.

چون تمام پارامترها نامنفی هستند. سمت راست (۱.۲) یک تابع هموار از متغییر های (S, V, P) در $\frac{۱}{۸}$ مثبت فضای $\Omega = \{(S, V, P) : S \geqslant ۰, V \geqslant ۰, P \geqslant ۰\}$ است. چون دستگاه فوق همگن است لذا $(S = ۰, V = ۰, P = ۰)$ یک جواب دستگاه است. طبق قضیه وجود و یکتایی جواب؛ منحنی جواب مورد نظراز اوّلین ربع شروع می شود و در آن باقی می ماند. واین یعنی هیچ منحنی فضای مختصاتی را قطع نمی کند. توجه داریم که Ω یک مجموعه ی ثابت است. □

قضیه ۴.۲. همه ی جوابهای دستگاه (۱.۲) به طور یکنواخت کراندار هستند.

اثبات. فرض می کنیم که $\chi = S + V + P$ بنابراین مشتق χ نسبت به زمان عبارتند از

$$\frac{d\chi}{dt} = \frac{dS}{dt} + \frac{dV}{dt} + \frac{dP}{dt}$$

حال برای هر $\mu > 0$ داریم:

$$\frac{d\chi}{dt} + \mu\chi = rS(1 - \frac{S}{K}) - p_1PS - m_1S - d_1S - p_2PV - m_2V - d_2V$$
$$- q_1p_1PS + q_2p_2PV - d_3P + \mu\chi$$

بنابراین

$$\frac{d\chi}{dt} + \mu\chi \leqslant k\frac{(r+\mu)^2}{4r} - (m_2 + d_2 - \mu)V - (d_2 - \mu)P$$

اگر $\mu < \min(m_2 + d_2, d_3)$ آنگاه داریم:

$$\frac{d\chi}{dt} + \mu\chi \leqslant k\frac{(r+\mu)^2}{4r} = \eta$$

بنابراین با استفاده از قضیه نابرابری دیفرانسیل داریم:

$$0 < \chi(S,V,P) < (\frac{\eta}{\mu})(1 - \exp(-\mu t)) + \chi(S_0, V_0, P_0)\exp(-\mu t)$$

بنابراین زمانی که $t \to \infty$ نتیجه می شود $0 < \chi(S,V,P) < \frac{\eta}{\mu}$ لذا همه ی جواب های دستگاه (۱.۲) به فضای زیر محدود می شوند.

$$\Lambda = \{(S,V,P) \in R_+^3 : \chi = \frac{\eta}{\mu} + \varepsilon, \forall\varepsilon > 0\}$$

□

۱.۲ نقاط تعادلی مدل

برای دستگاه (۱.۲) نقاط تعادلی زیر وجود دارد.

۱) تعادل بدیهی $E_0 = (0,0,0)$ همواره وجود دارد.

۲)تعادل بدون شکارچی $E_1 = (S_1, V_1, 0)$ وجود دارد اگر شرط زیر برقرار باشد.

$$(r - \phi - d_1 - m_1 + \frac{\theta\phi}{\theta + d_2 + m_2}) > 0. \qquad (۲.۲)$$

که اینجا

$$S_۱ = \frac{K}{r}(r - \phi - m_۱ - d_۱ + \frac{\theta\phi}{\theta + m_۲ + d_۲})$$
$$V_۱ = \frac{\phi K}{r(\theta + m_۲ + d_۲)}(r - \phi - m_۱ - d_۱ + \frac{\theta\phi}{\theta + m_۲ + d_۲}$$

۳)تعادل درونی $E_۲ = (S_۲, V_۲, P_۲)$ وجود دارد اگر شرایط زیر برقرار باشد.

$$(d_۳ - p_۲q_۲V_۲) > ۰$$

و $P_۲$ ریشه ی مثبت معادله ی زیرباشد.

$$\left(-\frac{rd_۳(\theta + d_۲ + m_۲ + p_۲P_۲)^۲}{K}\right) + [r\theta + (r - \phi)(d_۲ + m_۲) - (\theta + d_۲ + m_۲)(d_۱ + m_۱ + p_۱P_۲)$$
$$- P_۲(-r + \phi + d_۱ + m_۱ + p_۱P_۲)p_۲][p_۱(\theta + d_۲ + m_۲ + p_۲P_۲)q_۱ + \phi p_۲q_۲] = ۰. \qquad (۳.۲)$$

در این نقطه داریم:

$$S_۲ = \frac{d_۳ - p_۲q_۲V_۲}{p_۱q_۱},$$
$$V_۲ = \frac{\phi d_۳}{(\theta p_۱q_۱ + d_۲p_۱q_۱ + m_۲p_۱q_۱ + \phi p_۲q_۲ + p_۱p_۲q_۱P_۲)}.$$

۲.۲ بررسی پایداری نقاط تعادل

۱) نقطه تعادل بدیهی $E_۰ = (۰, ۰, ۰)$ همواره وجود دارد.

می دانیم برای تجزیه و تحلیل پایداری مدل از ماتریس ژاکوبین استفاده می کنیم لذا داریم:

$$\acute{J} = \begin{bmatrix} r - \frac{۲rS}{K} - \phi - p_۱P - m_۱ - d_۱ & \theta & -p_۱S \\ \phi & -\theta - p_۲P - m_۲ - d_۲ & -p_۳V \\ q_۱p_۱P & q_۲p_۲P & q_۱p_۱S + q_۲p_۲V - d_۳ \end{bmatrix}_{۳\times۳}$$

بنابراین ماتریس ژاکوبین ارزیابی شده در $E_۰$ به صورت زیر می باشد.

$$\acute{J}(E_۰) = \begin{bmatrix} r - \phi - m_۱ - d_۱ & \theta & ۰ \\ \phi & -\theta - m_۲ - d_۲ & ۰ \\ ۰ & ۰ & -d_۳ \end{bmatrix}$$

از اینرو چند جمله ای مشخصه ماتریس فوق عبارتند از:

$$(-\lambda - d_۳)(-\theta\phi - (r - \phi - d_۱ - m_۱ - \lambda)(\lambda + \theta + d_۲ + m_۲)) = ۰ \qquad (۴.۲)$$

لذا یکی از مقادیر ویژه ی $\acute{J}(E_۰)$ برابر $-d_۳$ می باشد و دو ریشه ی باقیمانده ی (۴.۲) از معادله درجه دوّم زیر به دست می آید.

$$(\lambda - r + \phi + d_۱ + m_۱)(\lambda + \theta + d_۲ + m_۲) - \theta\phi = ۰$$

اکنون با استفاده از محک روت -هرویتز اگر شرایط زیر برقرار باشد $E_۰$ پایدار موضعی است .

$$\begin{cases} (-r + \theta + \phi + d_۱ + d_۲ + m_۱ + m_۲) > ۰, \\ (-r + \phi + d_۱ + m_۱)(\theta + d_۲ + m_۲) - \theta\phi > ۰. \end{cases} \qquad (۵.۲)$$

۲) در نقطه تعادلی بدون شکارچی $E_۱ = (S_۱, V_۱, ۰)$ ماتریس ژاکوبین به صورت زیر می باشد.

$$\acute{J}(E_۱) = \begin{bmatrix} r - \frac{۲rS_۱}{K} - \phi - m_۱ - d_۱ & \theta & -p_۱S_۱ \\ \phi & -\theta - m_۲ - d_۲ & -p_۲V_۱ \\ ۰ & ۰ & q_۱p_۱S_۱ + q_۲p_۲V_۱ - d_۳ \end{bmatrix}$$

لذا معادله مشخصه ی $J\acute{(E_۱)}$ به صورت زیر می باشد.

$$\frac{۱}{K}(-\lambda - d_۳ + p_۱q_۱S_۱ + p_۲q_۲V_۱)[(۲rS_۱ + K(-r + \lambda + \phi + d_۱ + m_۱))(\lambda + \theta + d_۲ + m_۲) - K\theta\phi] = ۰. \qquad (۶.۲)$$

بنابراین یکی از مقادیر ویژه برابر است با $\lambda_۱ = -d_۳ + p_۱q_۱S_۱ + p_۲q_۲V_۱$

و دو ریشه ی باقیمانده ی چندجمله ای مشخصه ی (۵.۲) ازمعادله ی درجه دوّم زیر به دست می آید.

$$(\lambda + \frac{۲rS_۱}{K} - r + \phi + d_۱ + m_۱)(\lambda + \theta + d_۲ + m_۲) - \theta\phi = ۰. \qquad (۷.۲)$$

بنا به محک روت - هرویتز $E_۱$ پایدار است اگر

$$\begin{cases} (-r + \frac{۲rS_۱}{K} + \theta + \phi + d_۱ + d_۲ + m_۱ + m_۲) > ۰, \\ (\frac{۲rS_۱}{K} - r + \phi + d_۱ + m_۱)(\theta + d_۲ + m_۲) - \theta\phi > ۰, \\ \lambda_۱ = (-d_۳ + p_۱q_۱S_۱ + p_۲q_۲V_۱) < ۰ \end{cases} \qquad (۸.۲)$$

که اینجا $S_۱ > ۰$ ، $V_۱ > ۰$.

۳) در نقطه تعادل درونی $E_۲ = (S_۲, V_۲, P_۲)$ ماتریس ژاکوبین برابر است با

$$\acute{J}(E_۲) = \begin{bmatrix} \alpha_{۱۱} & \alpha_{۱۲} & \alpha_{۱۳} \\ \alpha_{۲۱} & \alpha_{۲۲} & \alpha_{۲۳} \\ \alpha_{۳۱} & \alpha_{۳۲} & \alpha_{۳۳} \end{bmatrix}$$

که اینجا

$$\alpha_{۱۱} = r - \frac{۲rS_۲}{K} - \phi - p_۱P_۲ - m_۱ - d_۱, \alpha_{۱۲} = \theta, \alpha_{۱۳} = -p_۱S_۲,$$
$$\alpha_{۲۱} = \phi, \alpha_{۲۲} = -\theta - p_۲P_۲ - m_۲ - d_۲, \alpha_{۲۳} = -p_۲V_۲,$$
$$\alpha_{۳۱} = q_۱p_۱P_۲, \alpha_{۳۲} = q_۳p_۳P_۲, \alpha_{۳۳} = q_۱p_۱S_۲ + q_۲p_۲V_۲ - d_۳$$

لذا معادله مشخصه ماتریس فوق به صورت زیر می باشد.

$$\lambda^۳ + B_۱\lambda^۲ + B_۲\lambda + B_۳ = ۰$$

$$B_۱ = \text{-}\square\square A = -(\alpha_{۱۱} + \alpha_{۲۲} + \alpha_{۳۳}),$$

مجموع مینور های مرتبه ی دوّم برابر است با $B_۲$

$$B_۲ = (\alpha_{۱۱}\alpha_{۲۲} - \alpha_{۱۲}\alpha_{۲۱}) + (\alpha_{۱۱}\alpha_{۳۳} - \alpha_{۱۳}\alpha_{۳۱}) + (\alpha_{۲۲}\alpha_{۳۳} - \alpha_{۲۳}\alpha_{۳۲})$$

$$B_۳ = -\square\square\square A = -[(\alpha_{۱۱}(\alpha_{۲۲}\alpha_{۳۳} - \alpha_{۳۲}\alpha_{۲۳}) + \alpha_{۱۲}(\alpha_{۳۱}\alpha_{۲۳} - \alpha_{۳۳}\alpha_{۲۱})) + \alpha_{۱۳}(\alpha_{۲۱}\alpha_{۳۲} - \alpha_{۳۱}\alpha_{۲۲})]$$

با توجه به محک روت -هرویتز $E_۲$ پایدار موضعی است اگر داشته باشیم:

$$\begin{cases} B_۱B_۲ > B_۳, \\ B_i > ۰ \end{cases}$$

برای $i = ۱, ۲, ۳$.

۳ نتیجه‌گیری

در مدل ارائه شده نقطه تعادلی بدیهی E_0 همواره وجود دارد. و اگر در شرط (۵.۲) صدق کند آنگاه این نقطه پایدار است. از لحاظ ریاضی این حالت ناپایدار است زیرا اگر این حالت پایدار باشد به این معنی است که گونه منقرض شده است. نقطه تعادلی E_1 اگر پایدار باشد به این معنی است که جمعیّت شکار بدون شکارچی مدت طولانی زنده می ماند. و مهمترین نقطه تعادلی؛ تعادل E_2 است زیرا این نقطه همزیستی همه ی گونه ها در اکوسیستم را نشان می دهد. این نقطه برای تعادل زیستی بسیار ضروری است.

تشکر و قدردانی

بدینوسیله نویسنده تشکّر وسپاسگزاری می نماید از مسئولین دانشگاه نیشابور، بویژه اساتید محترم گروه ریاضی که از محضرشان بهرمند شدیم.

مراجع

[46] H. W. Hethcote, W. D. Wang, L. T. Han and M. Zhien, *A predator- prey model with infected prey*, Theor. Popul. Biol. 66 (2004), 259- 268.

[47] S. Jana , T. K. Kar, *Modeling and analysis of a prey- predator system with disease in the prey*, Chaos Solitons Fractals. 47 (2013), 42- 53.

[48] A. Johri, et al. *Study of a prey- predator model with diseased prey*, Int.J.Contemp. Math. Sci. 7 (2012), no. 9- 12, 489- 498.

[49] S. Kumar, H. Kharbanda, *Stability analysis of prey- predator model with infection, migration and vaccination in prey*, arXiv preprint arXiv:1709.10319 (2017).

مطالعه عددی یک مدل ریاضی بیماری ناشی از آلودگی آب

یونس طالعی دانشگاه تبریز، تبریز، ایران، حجت اله عبادیزاده [1]،
دانشگاه امام علی (ع)، ایران

چکیده: یماری تب تیفوئید یکی از شایع ترین بیماری ناشی از آلودگی مواد غذایی و منابع آب و یکی از مشکلات زیست محیطی در حال حاضر جهان صنعتی می‌باشد. این بیماری عفونی مانند سایر بیماری‌ها از جمله ایدز، هپاتیت و ... در قالب یک دستگاه معادلات دیفرانسیل غیرخطی مدل‌بندی می‌شود. در این مقاله، یک روش عددی طیفی ماتریسی برای مطالعه عددی مدل ریاضی تب تیفوئید معرفی می شود. در اینجا، هدف پیدا کردن جواب تقریبی با دقت مناسب برای تابع میزان افراد محافظت شده، تحت درمان و احیا شده از طریق آموزش برای جلوگیری و درمان بیماری تیفوئید می‌باشد. نتایج حاصل از اجرای روش، دقت و کارایی روش را نشان می‌دهد.

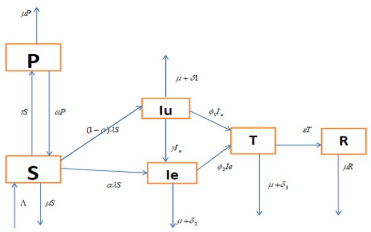
واژه‌های کلیدی: بیماری تب تیفوئید، عوامل آلودگی آب، مدل ریاضی، روش گالرکین ماتریسی

۱ مقدمه

بیماری تب تیفوئید یک بیماری عفونی است، که در اثر باکتری سالمونلا تیفی ایجاد می‌شود، این بیماری واگیردار بوده و از طریق آب و غذای آلوده گسترش می‌یابد و معمولاً با تب بالا، اسهال، بی‌اشتهایی و سردرد بروز می‌کند. باکتری تیفوئید در آب‌های گل آلود تا یک ماه و در یخ تا سه ماه زنده می‌ماند، در اثر گرمای ۶۰ تا ۱۰۰ درجه از میان می‌رود و بخصوص نور آفتاب به سرعت سبب انهدام باکتری می‌شود. در برابر خشکی هم تا دو ماه مقاومت دارد. این بیماری عفونی ممکن است به شکل تک‌گیر یا همه‌گیر درآید. در فصل پائیز و تابستان بیشتر به صورت همه گیر در می‌آید، این همه‌گیری بخصوص در اجتماعات مانند سربازخانه‌ها و مدارس دیده می‌شود. این بیماری در ایران در تمام فصول فراوان است ولی در تابستان و پائیز بیشتر است. در زمان جنگ که تمام این شرایط موجود است شیوع این بیماری چشمگیر است. افراد تازه وارد که به مناطق آلوده وارد می‌شوند بیشتر از افراد بومی مبتلا می‌گردند، کم شدن ترشح اسید معده هم در آلودگی مؤثر است زیرا اسید معده خاصیت از بین بردن این باکتری را دارد. مدل‌های ریاضی در تحلیلی رفتار دینامیکی این بیماری مانند بیماری‌های ایدز، مالاریا، هپاتیت و ... مورد استفاده قرار می‌گیرد. این مقاله، به بررسی جواب‌های عددی یک مدل ریاضی بیماری تیفوئید که یک دستگاه معادلات دیفرانسیل غیر خطی می‌باشد، می‌پردازیم. روش گالرکین ماتریسی، یک روش طیفی با دقت بالا برای تقریب مسائل فیزیکی و مدل‌های ریاضی می‌باشد که در آن، جواب های مساله با پایه های استاندارد از بعد متناهی تقریب زده

[1] حجت اله عبادیزاده: ebadizadeh.h@gmail.com

می‌شود.



شکل ۱: مدل اپیدمی تب تیفوئید

۲ مدل ریاضی

در این بخش به معرفی مدل ریاضی رفتار دینامیکی بیماری تب تیفوئید (شکل ۱) می‌پردازیم. این مدل ریاضی، توسط پیرلسن [۱] در سال ۱۹۸۹ بصورت یک دستگاه معادلات دیفرانسیل غیر خطی بصورت زیر معرفی شد:

$$\begin{cases} \frac{dP}{dt} = \tau S - (\mu+\omega)P - \mu S \\ \frac{dS}{dt} = \Lambda - S(\beta_1 I_u + \beta_2 I_e + \beta_3 T) + \omega P - (\tau+\mu)S \\ \frac{dI_u}{dt} = (1-\alpha)S(\beta_1 I_u + \beta_2 I_e + \beta_3 T) - (\mu+\delta_1+\phi_1+\gamma)I_u \\ \frac{dI_e}{dt} = \alpha S(\beta_1 I_u + \beta_2 I_e + \beta_3 T) - (\mu+\delta_2+\phi_2)I_e + \gamma I_u \\ \frac{dT}{dt} = \phi_1 I_u + \phi_2 I_e - (\mu+\delta_3+\varepsilon)T \\ \frac{dR}{dt} = \varepsilon T - \mu R \end{cases} \quad (۱.۲)$$

در اینجا $P(t)$ میزان افراد محافظت شده از عفونت در لحظه t ، $S(t)$ میزان افراد حساس درلحظه t ، $I_u(t)$ میزان افراد بیمار آموزش ندیده درلحظه t ، $I_e(t)$ میزان افراد بیمار آموزش دیده درلحظه t ، $T(t)$ میزان افراد تحت درمان درلحظه t و R میزان افراد احیا شده درلحظه t می‌باشند. در این مقاله، هدف اصلی ارایه یک روش عددی ماتریسی بر پایه استاندارد $1, t, t^2, ..., t^N$ (روش گالرکین ماتریسی) برای مدل ریاضی عفونت بیان شده در (۱.۲) می‌باشد. قرار می‌دهیم

$$\begin{cases} A_1 = \tau S - (\mu+\omega)P - \mu S \\ A_2 = \Lambda - S(\beta_1 I_u + \beta_2 I_e + \beta_3 T) + \omega P - (\tau+\mu)S \\ A_3 = (1-\alpha)S(\beta_1 I_u + \beta_2 I_e + \beta_3 T) - (\mu+\delta_1+\phi_1+\gamma)I_u \\ A_4 = \alpha S(\beta_1 I_u + \beta_2 I_e + \beta_3 T) - (\mu+\delta_2+\phi_2)I_e + \gamma I_u \\ A_5 = \phi_1 I_u + \phi_2 I_e - (\mu+\delta_3+\varepsilon)T \\ A_6 = \varepsilon T - \mu R \end{cases} \quad (۲.۲)$$

در اینصورت خواهیم داشت:

$$|\frac{\partial A_1}{\partial P}| = \mu+\omega < \infty, \quad |\frac{\partial A_1}{\partial S}| = \tau, |\frac{\partial A_1}{\partial Ie}| = 0, |\frac{\partial A_1}{\partial Iu}| = 0, |\frac{\partial A_1}{\partial T}| = 0, |\frac{\partial A_1}{\partial R}| = 0,$$

بطور مشابه برای $i = 2, ..., 6$ ، A_i حاصل مشتق‌های نسبی کراندار خواهد بود، لذا مدل دینامیکی ۱.۲ بنا به قضیه زیر دارای جواب یکتا و پیوسته می‌باشد.

---

[۱] *Perelson*

قضیه ۱.۲. (وجود و یکتایی جواب) برای معادله دیفرانسیل مرتبه اول غیر خطی

$$x'(t) = f(t, x(t)), \quad x(۰) = x. \tag{۳.۲}$$

اگر تابع $f(., x(.))$ در شرط لیپ‌شیتز صدق کند، بعبارتی دیگر $|f(., x(.)) - f(., y(.))| < L|x - y|$ یا اینکه تغییرات f نسبت به مولفه دوم کراندار باشد، یعنی داشته باشیم $|\frac{\partial f(., x(.))}{\partial x}| < \infty$. آنگاه مساله مقدار اولیه (۳.۲) دارای جواب پیوسته یکتا خواهد بود. (ر. ک. معادلات دیفرانسیل عادی، محمود حصارکی ۱۳۹۴).

۳ نتایج اصلی

در این قسمت به پیاده سازی روش گالرگین ماتریسی برای حل عددی مساله پرداخته و سپس روش پیشنهادی در قالب یک مثال عددی بررسی می‌شود.

۱.۳ روش گالرکین ماتریسی

اگر فرض می‌کنیم $P_N(t) = \sum_{k=۰}^{N} p_k t^k$ ، $S_N(t) = \sum_{k=۰}^{N} s_k t^k$ ، $I_{u,N}(t) = \sum_{k=۰}^{N} u_k t^k$ ، $I_{e,N}(t) = \sum_{k=۰}^{N} e_k t^k$ ، $T_N(t) = \sum_{k=۰}^{N} T_k t^k$ ، $R_N(t) = \sum_{k=۰}^{N} r_k t^k$ تقریب‌هایی از جواب‌های مساله در فضای توابع با بعد متناهی باشند، آنگاه دارای نمایش ماتریسی بصورت زیر خواهند بود:

$$P_N(t) = \mathbf{EX}_N(t)\mathbf{P}, \;\; S_N(t) = \mathbf{EX}_N(t)\mathbf{S}, \;\; I_{u,N}(t) = \mathbf{EX}_N(t)\mathbf{Iu}, \;\; I_{e,N}(t) = \mathbf{EX}_N(t)\mathbf{Ie},$$

$$T_N(t) = \mathbf{EX}_N(t)\mathbf{T}, \;\; R_N(t) = \mathbf{EX}_N(t)\mathbf{R}.$$

بطوریکه

$$\mathbf{P} = [p_., p_۱, ..., p_N]^T, \;\; \mathbf{S} = [s_., s_۱, ..., s_N]^T, \; \mathbf{Iu} = [u_., u_۱, ..., u_N]^T,$$
$$\mathbf{Ie} = [۱, e_۱, ..., e_N], \; \mathbf{T} = [T_., T_۱, ..., T_N]^T, \; \mathbf{R} = [r_., r_۱, ..., r_N]^T, \;\; \mathbf{EX}_N(t) = [۱, t, ..., t^N].$$

قضیه ۱.۳. با فرض اینکه ماتریس M و $\widehat{\mathbf{S}}$ بصورت

$$\mathbf{M} = \begin{pmatrix} ۰ & ۰ & ۰ & \dots & ۰ \\ ۱ & ۰ & ۰ & \dots & ۰ \\ ۰ & ۲ & ۰ & \dots & ۰ \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ ۰ & ۰ & ۰ & N & ۰ \end{pmatrix}, \;\; \widehat{\mathbf{S}} = \begin{pmatrix} s_۰ & ۰ & ۰ & \dots & ۰ \\ s_۱ & s_۰ & ۰ & \dots & ۰ \\ s_۲ & s_۱ & s_۰ & \dots & ۰ \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ s_N & s_{N-۱} & s_{N-۲} & \dots & s_۰ \\ ۰ & s_N & s_{N-۱} & \dots & s_۱ \\ ۰ & ۰ & s_N & \dots & s_۲ \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ ۰ & ۰ & ۰ & \dots & s_N \end{pmatrix}.$$

باشند، آنگاه خواهیم داشت:

۱.

$$\frac{dP_N}{dt} = \mathbf{PMEX}_N(t), \;\; \frac{dS_N}{dt} = \mathbf{SMEX}_N(t), \;\; \frac{dI_{u,N}}{dt} = \mathbf{IuMEX}_N(t),$$
$$\frac{dI_{e,N}}{dt} = \mathbf{IeMEX}_N(t), \;\; \frac{dT_N}{dt} = \mathbf{TMEX}_N(t), \;\; \frac{dR_N}{dt} = \mathbf{RMEX}_N(t).$$

۲.

$$S_N I_{u,N} = \mathbf{EX}_{۲N}(t)\widehat{\mathbf{S}}\mathbf{Iu}, \;\; S_N I_{e,N} = \mathbf{EX}_{۲N}(t)\widehat{\mathbf{S}}\mathbf{Ie}, \;\; S_N T_N = \mathbf{EX}_{۲N}(t)\widehat{\mathbf{S}}\mathbf{T}.$$

۲.۳ بیان روش

با جایگذاری نمایش های ماتریسی بیان شده در قضیه ۱.۳ در جملات سیستم غیر خطی ۱.۲ خواهیم داشت:

$$\begin{cases} G_1 := \mathbf{PMEX}_N(t) - \tau\mathbf{EX}_N(t)\mathbf{S} + (\mu+\omega)\mathbf{EX}_N(t)\mathbf{P} + \mu\mathbf{EX}_N(t)\mathbf{S} \\ G_2 := \mathbf{SMEX}_N(t) - \Lambda + \beta_1\mathbf{EX}_{2N}(t)\widehat{\mathbf{S}}\mathbf{Iu} + \beta_2\mathbf{EX}_{2N}(t)\widehat{\mathbf{S}}\mathbf{Ie} + \beta_3\mathbf{EX}_{2N}(t)\widehat{\mathbf{S}}\mathbf{T} - \omega\mathbf{EX}_N(t)\mathbf{P} - (\tau+\mu)\mathbf{EX}_N(t)\mathbf{S} \\ G_3 := \mathbf{IuMEX}_N(t) - (1-\alpha)(\beta_1\mathbf{EX}_{2N}(t)\widehat{\mathbf{S}}\mathbf{Iu} + \beta_2\mathbf{EX}_{2N}(t)\widehat{\mathbf{S}}\mathbf{Ie} + \beta_3\mathbf{EX}_{2N}(t)\widehat{\mathbf{S}}\mathbf{T}) + (\mu+\delta_1+\phi_1+\gamma)\mathbf{EX}_N(t)\mathbf{Iu} \\ G_4 := \mathbf{IeMEX}_N(t) - \alpha(\beta_1\mathbf{EX}_{2N}(t)\widehat{\mathbf{S}}\mathbf{Iu} + \beta_2\mathbf{EX}_{2N}(t)\widehat{\mathbf{S}}\mathbf{Ie} + \beta_3\mathbf{EX}_{2N}(t)\widehat{\mathbf{S}}\mathbf{T}) + (\mu+\delta_2+\phi_2)\mathbf{EX}_N(t)\mathbf{Ie} - \gamma\mathbf{EX}_N(t)\mathbf{Iu} \\ G_5 := \mathbf{TMEX}_N(t) - \phi_1\mathbf{EX}_N(t)\mathbf{Iu} - \phi_2\mathbf{EX}_N(t)\mathbf{Ie} + (\mu+\delta_3+\varepsilon)\mathbf{EX}_N(t)\mathbf{T} \\ G_6 := \mathbf{RMEX}_N(t) - \varepsilon\mathbf{EX}_N(t)\mathbf{T} + \mu\mathbf{EX}_N(t)\mathbf{R} \end{cases} \tag{۱.۳}$$

بنا به روش طیفی گالرکین به ازای $k = 0, 1, ..., N-1$ خواهیم داشت:

$$< G_1, t^k > = < G_2, t^k > = < G_3, t^k > = < G_4, t^k > = < G_5, t^k > = < G_6, t^k > = 0. \tag{۲.۳}$$

و از طرفی برای شرط های اولیه خواهیم داشت :

$$P_N(0) = p_0,\ \ S_N(0) = s_0,\ \ I_{u,N}(0) = u_0,\ \ I_{e,N}(0) = e_0,\ \ T_N(0) = T_0,\ \ R_N(0) = r_0.$$

در نتیجه با محاسبه ضرب های داخلی (۲.۳) و استفاده از شرط های اولیه یک دستگاه معادلات جبری غیرخطی حاصل می شود، در نتیجه با حل این دستگاه مقادیر بردارهای مجهول $\mathbf{P}, \mathbf{S}, \mathbf{Iu}, \mathbf{Ie}, \mathbf{T}, \mathbf{R}$ حاصل می‌شوند.

۴ کران خطا

در این بخش، یک کران بالا برای خطای مطلق برای جواب تقریبی بدست آمده از روش پیشنهادی بر اساس درجه همواری جواب مساله ارائه می‌دهیم. ابتدا یک کران بالا برای خطای مطلق $P_N(t)$ تعیین می‌کنیم. تابع خطا را بصورت

$$e_{N,P}(t) = P(t) - P_N(t)$$

تعریف می‌کنیم. از آنجاییکه جواب تقریبی $P_N(t)$ بعنوان یک بهترین تقریب برای $P(t)$ از یک زیر فضای با بعد متناهی $\square$ می باشد، لذا طبق قضیه تیلور با باقیمانده لاگرانژ می‌توان نوشت:

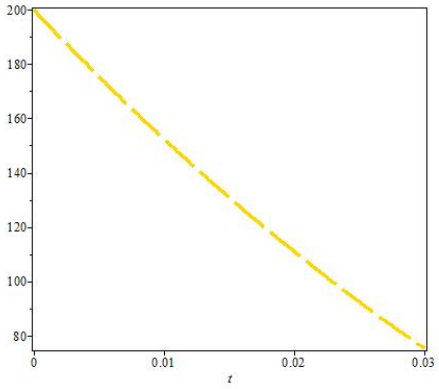
$$|e_{N,P}(t)| = |P(t) - P_N(t)| \le \frac{t^{N+1}}{(N+1)!}P^{(N+1)}(\zeta) \le \frac{M_p}{(N+1)!}$$

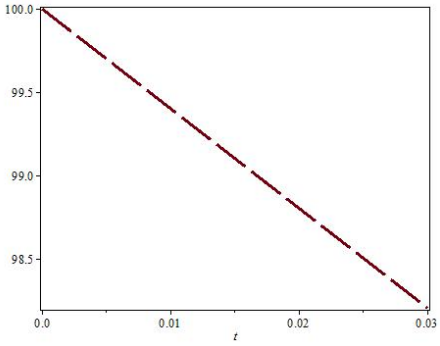
که در آن $M_p = \max_{t\in[0,1]} |P^{(N+1)}(t)|$ می‌باشد. بنابراین، مشابه با روند بالا، خواهیم داشت:

$$|e_{N,S}(t)| \le \frac{M_S}{(N+1)!},\ \ |e_{N,Iu}(t)| \le \frac{M_{Iu}}{(N+1)!},\ \ |e_{N,Ie}(t)| \le \frac{M_{Ie}}{(N+1)!},$$
$$|e_{N,T}(t)| \le \frac{M_T}{(N+1)!},\ \ |e_{N,R}(t)| \le \frac{M_R}{(N+1)!}.$$

۵ نتایج عددی

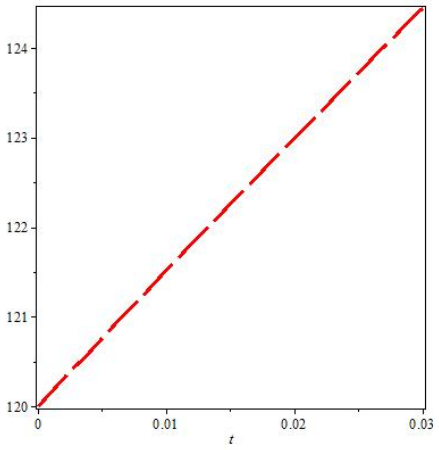
مثال ۱.۵. به عنوان مثال شرایط $P(0) = 100,\ S(0) = 200,\ I_u(0) = 140,\ \ I_e(0) = 120,\ \ T(0) = 80,\ \ R(0) = 60$ را در نظر می‌گیریم که در آن مقادیر پارامترها در جدول ۱ داده شده است. نتایج عددی بدست آمده در شکل‌ها، دقت و کارایی روش پیشنهادی را به ازای $N = 5$ نشان می‌دهد.

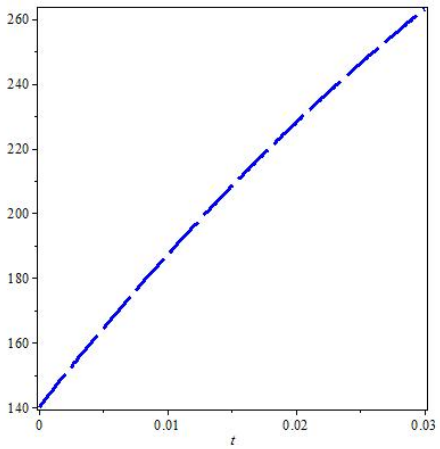




شکل ۲: رفتار تقریبی تابع $P(t)$

شکل ۳: رفتار تقریبی تابع $S(t)$



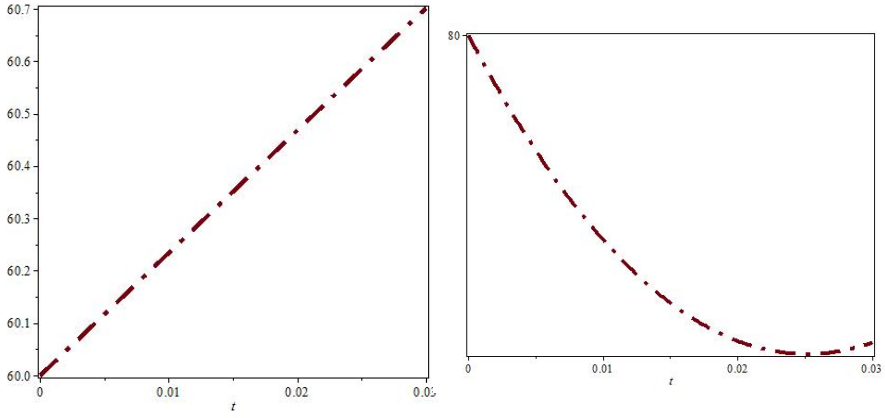


شکل ۴: رفتار تقریبی تابع $Iu(t)$

شکل ۵: رفتار تقریبی تابع $Ie(t)$

جدول ۱: مقادیر پارامترها

$\delta_1, \delta_2, \delta_3$	α	τ	ω	β_3	β_2	β_1	μ	γ	ε	ϕ_2	ϕ_1	Λ
۰۰۷۵.۰	۰۰۷۲.۰	۱.۰	۵.۰	۰۵.۰	۰۵.۰	۵.۱	۱۴۲.۰	۶.۰	۴.۰	۳.۰	۰۴.۰	۲۰۰



شکل ۶: رفتار تقریبی تابع $T(t)$

شکل ۷: رفتار تقریبی تابع $R(t)$

۶ نتیجه‌گیری

مدل‌های ریاضی مسائل و بررسی رفتار آن‌ها در قالب الگوریتم های عددی یکی از ابزارهای مناسب و قابل اعتماد پژوهش‌های علمی- ریاضی و بین رشته ای می‌باشد. قابلیت تفسیر رفتارهای دینامیکی سیستم های خطی و غیر خطی با زبان الگوریتم ریاضی می تواند شامل روش‌های عددی مختلف باشد. در این مقاله، یک روش عددی به زبان ساده ماتریسی برای بررسی مدل بیماری تب تیفوئید معرفی شد که می‌تواند برای سیستم های مشابه نیز قابل تعمیم باشد.

مراجع

[۵۰] غ. قربانی، ع. توانا و ح. پرهیزگار، بررسی اپیدمی تیفوئید در یک واحد نظامی، مجله طب نظامی، ۵ (۱۳۸۲)، ۳۹-۴۲.

[۵۱] م. ملکوتیان و همکاران، بررسی میزان شیوع و مرگ و میر ناشی از بیماری‌های منتقله از آب آشامیدنی و غذا مجله ارتقای ایمنی و پیشگیری از مصدومیت ها، ۳ (۱۳۹۴)، ۱۵-۲۴.

[۵۲] م. حصارکی، و. رومی ،معادلات دیفرانسیل عادی، .(۱۳۹۴)

اولین همایش ملی ریاضیات زیستی
۲۱ و ۲۲ اسفندماه ۱۳۹۷
همایش ملی ریاضیات زیستی
دانشگاه نیشابور

بررسی مدل ریاضی شیوع یک عامل بیماری

حجت‌الله عبادی‌زاده [۱]، دانشکده علوم‌پایه، گروه ریاضی، دانشگاه افسری امام علی(ع)، علی عزیزی مایوان، دانشکده علوم‌پایه، گروه ریاضی، دانشگاه افسری امام علی (ع)، سید مهدی کاظمی تربقان، دانشکده علوم‌پایه، گروه ریاضی، دانشگاه بجنورد

چکیده: در این مقاله به ارائه و بررسی یک مدل ریاضی برای شیوع یک عامل بیماری در جامعه می‌پردازیم. در این مدل افراد جامعه را به چهار دسته‌ی: افراد مستعد دریافت عامل بیماری، افراد آلوده به عامل بیماری، افراد بیماری که تحت درمان قرار گرفته‌اند و افرادی که بهبود یافته‌اند، تقسیم می‌کنیم. سپس دستگاه معادلات دیفرانسیل مدل را ارائه داده و با استفاده از ماتریس نسل دوم عدد شیوع را محاسبه می‌نماییم. در پایان، در کنار تحلیل دینامیکی مدل، به بررسی شرایط پایداری حول نقطه تعادل دستگاه معادلات می‌پردازیم.

واژه‌های کلیدی: عدد شیوع، ماتریس نسل دوم، مدل ریاضی بیماری. **رده‌بندی ریاضی** (۲۰۱۰): ۹۲D۳۰، ۳۴D۲۳.

۱ مقدمه

ویروس یک عامل عفونی ذره‌بینی است که تنها قادر است درون سلول‌های زنده یک جاندار رشد کرده و تکثیر شود. ویروس‌ها تقریبا در تمام اکوسیستم‌های روی کره زمین یافت می‌شوند و به‌عنوان یکی از عوامل مهم ایجاد بیماری در یک جاندار و انتقال آن به سایر جانداران شناخته می‌شود. از جمله بیماری‌هایی که دارای منشأ ویروسی هستند می‌توان به سرماخوردگی و آنفولانزا اشاره کرد که بیماری‌هایی شایع و تقریبا بی‌خطر هستند، اما بیماری‌های خطرناکی چون ایدز و انواع هپاتیت نیز دارای منشأ ویروسی هستند. شیوع ویروس یک بیماری در جامعه، در صورت فراهم بودن شرایط، می‌تواند به یک همه‌گیری (اپیدمی) منجر شود. اما برای کنترل و جلوگیری از گسترش یک عامل ویروسی در جامعه، فهمیدن مکانیسم و فرایند دینامیکی انتقال ویروس یک امر بسیار مهم است.

در سال ۱۷۶۰، دانیل برنولی اولین مدل ریاضی در زمینه اپیدمیولوژی را ارائه داد که نشان می‌داد واکسیناسیون در برابر بیماری آبله سودمند است [۷]. مدل برنولی زمینه‌ساز ورود مدل‌های ریاضی به مطالعات زیستی و بیولوژیک شد به‌طوری‌که در اوایل قرن بیستم مطالعات گسترده و جامعی بر روی شیوع بیماری‌های عفونی، با استفاده از مدل‌های ریاضی، صورت پذیرفت [۳, ۴]. به‌عنوان مثال، در سال ۱۹۰۶ هامر یک مدل برای بررسی گسترش بیماری سرخک ارائه داد، در سال ۱۹۱۱ یک فیزیکدان به نام راس، با استفاده از معادلات دیفرانسیل، به طراحی یک مدل برای شیوع بیماری مالاریا از طریق پشه پرداخت و در سال ۱۹۲۶ این کرماک و مک کندریک بودند که مدل معروف SIR را برای اپیدمی بیماری طاعون در لندن در سال ۱۶۶۶ ارائه نمودند [۱, ۶, ۷].

در اپیدمیولوژی، مدل‌های ریاضی فهم باارزشی از روند گسترش یک بیماری در میان انسان‌ها ارائه دادند و فرضیه‌هایی برای بهبود

[۱] حجت‌الله عبادی‌زاده: ebadizadeh.h@gmail.com

استراتژی کنترل بیماری‌ها مطرح نمودند. امروزه مدل‌های ریاضی در تحلیل دینامیک بیماری‌های زیادی مانند ایدز، مالاریا، هپاتیت و غیره مورد استفاده قرار می‌گیرند. این مدل‌ها نقش مهمی در درک بهتر از الگوهای اپیدمی جهت کنترل بیماری دارند.

در این مقاله یک مدل ریاضی برای شیوع یک عامل بیماری (ویروس) در جامعه ارائه می‌دهیم که در این مدل جمعیت را به چهار دسته تقسیم کرده‌ایم: افرادی که مستعد دریافت عامل بیماری هستند، افرادی که به عامل بیماری آلوده شده‌اند، افراد بیماری که تحت درمان قرار می‌گیرند و کسانی که بهبود یافته‌اند. با توجه به اینکه با گذشت زمان تعداد جمعیت مورد نظر تغییر می‌کند، مدل ما یک مدل وابسته به زمان خواهد بود.

مباحث ارائه شده در این مقاله به این صورت است که: در بخش دوم به مباحث مربوط به مدل‌سازی ریاضی که شامل تعیین متغیرها و پارامترها، فرمول‌بندی معادلات و تشکیل دستگاه معادلات متناظر با آن که در اینجا یک دستگاه معادلات معمولی غیرخطی است پرداخته می‌شود. در بخش سوم نیز نقطه تعادل عاری از بیماری و نقطه تعادل شیوع بیماری ارائه شده و با استفاده از ماتریس نسل دوم مدل ریاضی، فرمول عدد شیوع محاسبه می‌گردد. علاوه‌براین، پایداری دستگاه معادلات مدل موردنظر نیز مورد بحث قرار داده می‌شود.

۲ مدل‌سازی ریاضی

با توجه به مطالب گفته شده در بخش قبل، کل جمعیت جامعه مورد مطالعه را در زمان t با $N(t)$ نمایش می‌دهیم. حال اگر $S(t)$ مجموعه افرادی از جامعه باشد که مستعد دریافت عامل بیماری هستند، $I(t)$ مجموعه افرادی که آلوده به عامل بیماری باشند، $T(t)$ مجموعه همه بیمارانی باشد که تحت درمان قرار می‌گیرند و $R(t)$ مجموعه همه کسانی باشد که بهبود یافته‌اند، آنگاه داریم:

$$N(t) = S(t) + I(t) + T(t) + R(t).$$

این مدل را به اختصار $SITR$ می‌نامیم. در این مدل، نرخ انتقال بیماری را β و نرخ مرگ‌ومیر طبیعی و زادوولد را به‌ترتیب μ و λ در نظر می‌گیریم. همچنین نرخ انتخاب افراد بیمار برای درمان را γ، نرخ بهبود بیماران تحت درمان را η و نرخ بهبود بیماران بدون استفاده از خدمات درمانی را α در نظر می‌گیریم. به این ترتیب اگر کسر δ بیانگر تعداد افراد بیمار تحت درمانی باشد که همچنان می‌توانند بیماری را انتقال دهند، آنگاه تعداد افراد جدیدی که در زمان t آلوده می‌شوند برابر خواهد بود با $\beta S(t)(I(t) + \delta T(t))$.

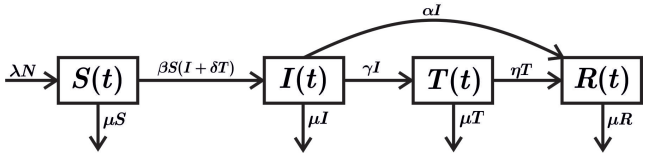
حال با توجه به متغیرها و پارامترهای تعریف شده در بالا، به فرمول‌بندی مدل ریاضی $SITR$ می‌پردازیم که منجر به دستگاه معادلات دیفرانسیل غیرخطی زیر می‌شود:

$$\begin{aligned}
\dot{S} &= -\beta S(I + \delta T) - \mu S + \lambda N,\\
\dot{I} &= \beta S(I + \delta T) - (\alpha + \gamma + \mu)I,\\
\dot{T} &= \gamma I - (\eta + \mu)T,\\
\dot{R} &= \eta T + \alpha I - \mu R.
\end{aligned} \tag{۱.۲}$$

توجه کنید که تمام پارامترها در معادلات بالا نامنفی هستند. در شکل ۱ نمودار گردشی این مدل ارائه شده است.

اگر N را برابر با یک در نظر بگیریم، در این‌صورت مجموعه‌های S، I، T و R کسری از N بوده و داریم:

$$S + I + T + R = ۱,$$



شکل ۱: نمودار گردشی مدل $SITR$

بنابراین $R = ۱ - S - I - T$. به‌این‌ترتیب می‌توان دستگاه معادلات (۱) را به‌صورت زیر در نظر گرفت:

$$\begin{aligned} \dot{S} &= -\beta S(I + \delta T) - \mu S + \lambda, \\ \dot{I} &= \beta S(I + \delta T) - (\alpha + \gamma + \mu)I, \\ \dot{T} &= \gamma I - (\eta + \mu)T. \end{aligned} \tag{۲.۲}$$

۳ نقاط تعادل و عدد شیوع

جواب‌های دستگاه معادلات (۲) بصورت $Q = (S, I, T)$ هستند که قضیه زیر مثبت بودن این جواب‌ها را نشان می‌دهد [۵].

قضیه ۱: اگر مقادیر $S(۰)$، $I(۰)$و $T(۰)$ نامنفی باشند، آنگاه برای $t > ۰$ مقادیر $S(t)$، $I(t)$ و $T(t)$ نامنفی هستند.

اگر فرض کنیم که در جمعیت N هیچ بیماری وجود نداشته باشد، در این صورت نقطه تعادل عاری از بیماری دستگاه (۲) به‌صورت $Q_۰ = (S_۰, ۰, ۰)$ خواهد بود که $S_۰ = \frac{\lambda}{\mu}$. سمت راست معادلات دستگاه (۲) را برابر با صفر قرار داده و نقطه تعادل دیگر دستگاه (۲) که به‌صورت (S^*, I^*, T^*) است را محاسبه می‌کنیم. پس داریم

$$\begin{aligned} S^* &= \frac{-(\alpha + \gamma + \mu)I^* - \lambda}{\mu}, \\ T^* &= \frac{\gamma I^*}{\eta + \mu}. \end{aligned}$$

عدد شیوع بیماری، تعداد افرادی است که یک فرد بیمار، در طول دوره بیماری خود، آنها را مبتلا می‌کند. با توجه به [۲]، عدد شیوع $R_۰$ درواقع همان شعاع طیفی ماتریس نسل دوم $FV^{-۱}$ است. ماتریس‌های F و V را از دستگاه (۲) به‌صورت زیر بدست می‌آوریم:

$$F = \begin{bmatrix} \frac{\beta\lambda}{\mu} & \frac{\beta\lambda\delta}{\mu} \\ ۰ & ۰ \end{bmatrix},$$

و

$$V = \begin{bmatrix} \alpha + \gamma + \mu & ۰ \\ -\gamma & \eta + \mu \end{bmatrix}.$$

پس داریم:

$$V^{-۱} = \frac{۱}{(\alpha + \gamma + \mu)(\eta + \mu)} \begin{bmatrix} \eta + \mu & \gamma \\ ۰ & \alpha + \gamma + \mu \end{bmatrix},$$

که در نهایت

$$R_0 = \frac{\beta\lambda}{\mu(\alpha+\gamma+\mu)}.$$

در ادامه، با استفاده از ژاکوبین دستگاه معادلات (۲)، پایداری سیستم را حول نقطه تعادل عاری از بیماری، یعنی $(\frac{\lambda}{\mu}, 0, 0)$، بررسی می‌کنیم.

قضیه ۲: نقطه تعادل عاری از بیماری بطور مجانبی پایدار است اگر $R_0 < 1$ و ناپایدار است اگر $R_0 > 1$. **اثبات.** ماتریس ژاکوبی دستگاه (۲) را به‌صورت زیر محاسبه می‌کنیم:

$$J_0 = \begin{bmatrix} -\mu & \frac{-\beta\lambda}{\mu} & \frac{-\beta\lambda\delta}{\mu} \\ 0 & \alpha & \frac{\beta\lambda\delta}{\mu} \\ 0 & \gamma & -(\eta+\mu) \end{bmatrix}.$$

پس چندجمله‌ای مشخصه J_0 به‌صورت $(x+\mu)(x^2+bx-c)$ خواهد بود که

$$b = \eta+\alpha+\gamma+2\mu-\frac{\beta\lambda}{\mu}$$
$$c = (\frac{\beta\lambda}{\mu}-(\alpha+\gamma+\mu))(\eta+\mu)-\frac{\beta\lambda\gamma\delta}{\mu}.$$

با توجه به [۲]، در صورتی که جواب‌های چندجمله‌ای مشخصه دارای بخش حقیقی منفی باشند، آنگاه دستگاه پایدار مجانبی خواهد بود و این در صورتی اتفاق می‌افتد که $R_0 < 1$. همچنین اگر $R_0 > 1$ آنگاه دستگاه ناپایدار خواهد بود. □

مراجع

[53] F. Brauer, P. Van den, J. Wu, *Mathematical Epidemiology*, Springer-Verlag Berlin, 2008.

[54] P.V.D. Driessche and J. Watmough, *Reproduction numbers and subthreshold endemic equilibria for compartmental models of disease transmission*, Math Biosci. $180 : 29-48$, 2002

[55] M. Farkas, *Dynamical Models in Biology*, Academic Press, 2001.

[56] Z. Ma, J. Li, *DynamicalModeling and Analysis of Epidemics*, World Scientific Publishing Co. Pte. Ltd., 2009.

[57] I. Shaban, E. S. Massaw, O. Daniel Makinde, *Modelling the effect of screening on the spread of HIV infection in a population with variable inflow of infective immigrants*, Scientific Research and Essays, $6 : 4397-4405$, 2011.

[58] E. Vynnycky, R. G. White, *An Introduction to Infectious Disease Modelling*, Oxford: Oxford University Press, 2010.

[59] X. Q. Zhao, *Dynamical Systems in Population Biology*, Springer-Verlag New York, 2003.

یک مدل ریاضی برای هپاتیت Bو بررسی اثر واکسیناسیون در رشد بیماری

حجت‌الله عبادی‌زاده [۱]، دانشکده علوم‌پایه، گروه ریاضی، دانشگاه افسری امام علی(ع)، علی عزیزی مایوان، دانشکده علوم‌پایه، گروه ریاضی، دانشگاه افسری امام علی (ع)، سید مهدی کاظمی تربقان، دانشکده علوم‌پایه، گروه ریاضی، دانشگاه بجنورد

چکیده: این مقاله یک مدل ریاضی برای بیماری هپاتیت B ارائه می‌شود. در این مدل جامعه مورد مطالعه بر اساس نوع فعالیت افراد که ممکن است منجر به ابتلای آنها به هپاتیت B گردد، به دو دسته تقسیم می‌شود که نرخ واکسیناسیون در این دو دسته می‌تواند متفاوت باشد. در پایان، فرمول عدد شیوع بر اساس پارامترهای مدل محاسبه می‌شود.

واژه‌های کلیدی: بیماری هپاتیت B ، عدد شیوع، ماتریس نسل دوم.

۱ مقدمه

در سال ۱۷۶۰، دانیل برنولی اولین مدل ریاضی در زمینه اپیدمیولوژی را ارائه داد که نشان می‌داد واکسیناسیون در برابر بیماری آبله سودمند است [۱۲]. مدل برنولی زمینه‌ساز ورود مدل‌های ریاضی به مطالعات زیستی و بیولوژیک شد. در اپیدمیولوژی، مدل‌های ریاضی فهم باارزشی از روند گسترش یک بیماری در میان انسان‌ها ارائه دادند و فرضیه‌هایی برای بهبود استراتژی کنترل بیماری‌ها مطرح نمودند. امروزه مدل‌های ریاضی در تحلیل دینامیک بیماری‌ها زیادی مانند ایدز، مالاریا، هپاتیت و غیره مورد استفاده قرار می‌گیرند. این مدل‌ها نقش مهمی در درک بهتر از الگوهای اپیدمی جهت کنترل بیماری دارند.

هپاتیت Bیک عفونت کبدی خطرناک است که ناشی از ویروس HBV از خانواده هپادنا ویریده می‌باشد. عفونت هپاتیت B ممکن است کوتاه مدت باشد (هپاتیت B حاد) یا مدت زمان زیادی طول بکشد (هپاتیت B مزمن). در برخی بیماران که عفونت هپاتیت Bمزمن شده، ممکن است منجر به نارسایی کبدی، سرطان کبد یا سیروز شود (سیروز کبدی بیماری است که باعث زخم و سخت شدن پایدار بافت کبدی می‌شود). راه‌های انتقال این بیماری همانند بیماری ایدز است با این تفاوت که شیوع هپاتیت B در دنیا در حدود ۵۰ تا ۶۰ برابر بیشتر از شیوع بیماری ایدز است. آمارهای جهانی نشان می‌دهد که نزدیک به ۲ میلیارد انسان در سرتاسر دنیا آلوده به ویروس هپاتیت B هستند و سالانه ۵۰ میلیون نفر به این تعداد اضافه می‌شوند. طبق تخمین‌هایی که زده می‌شود، سالانه ۶۰۰۰۰۰ نفر نیز از عوارض ناشی از این بیماری جان خود را از دست می‌دهند [۱، ۹].

اصلی‌ترین راه‌های انتقال HBVروابط جنسی، انجام تزریقات با سرنگ و سوزن آلوده و انتقال از مادر به جنین است. به این ترتیب، معتادان تزریقی که از سرنگ و سوزن مشترک استفاده می‌کنند، کارکنان بخش‌های مراقبت‌های بهداشتی، کادر درمان و کسانی که

[۱] حجت‌الله عبادی‌زاده: ebadizadeh.h@gmail.com

با خون انسان در تماس هستند و افرادی که روابط جنسی محافظت نشده دارند، در معرض خطر ابتلا به هپاتیت B قرار دارند. البته فرد می‌تواند با یک واکسن از بروز بیماری پیشگیری کند. اگر فردی مبتلا به هپاتیت B باشد، باید اقدامات لازم را به‌کار گیرد تا از انتقال این بیماری به دیگران پیشگیری کند. آمارهای گردآوری شده در چند سال اخیر نشان می‌دهد که در ایران شیوع هپاتیت B کمتر از ۳ درصد است [۴]. برنامه‌هایی که در ایران برای کنترل و کاهش شیوع بیماری هپاتیت B اجرا شدند، از جمله برنامه آگاهی‌بخشی عمومی نسبت به هپاتیت B و راه‌های انتقال آن و همچنین طرح واکسیناسیون که از سال ۱۹۹۳ برای نوزادان، کارکنان مراکز بهداشتی و سایر افراد در معرض خطر صورت پذیرفت، باعث شدند که درصد شیوع هپاتیت B در ایران از ۵/۳ درصد در سال ۱۹۹۰ به ۱۴/۲ درصد در سال ۲۰۰۰ برسد [۲].

تحقیقات و مقالات بسیاری در زمینه اپیدمیولوژی بیماری هپاتیت B وجود دارد که در آنها از مدل‌های ریاضی استفاده شده است. بعنوان نمونه می‌توان به [۳] اشاره کرد که در آن از یک مدل ساده ریاضی برای بررسی اثر حامل‌های ویروس در انتقال ویروس HBVاستفاده شده است. در [۵] از مدل‌های ریاضی برای بهبود استراتژی ریشه‌کنی هپاتیت B در نیوزلند کمک گرفته شده است. استفاده از مدل‌های ریاضی در بررسی دینامیک انتقال و شیوع هپاتیت B در چین [۱۱] و در بررسی اثر واکسیناسیون در کنترل هپاتیت B [۶] نیز از جمله کاربردهای مدل‌های ریاضی در تحقیقات مربوط به اپیدمیولوژی هپاتیت B است. چند محقق ایرانی نیز اخیراً در [۱۰] به ارائه یک مدل ریاضی برای بررسی دینامیک شیوع هپاتیت B در بخشی از استان کرمان پرداخته‌اند.

در این مقاله یک مدل ریاضی برای بیماری هپاتیت B ارائه می‌دهیم که در این مدل جمعیت را به شش دسته تقسیم کرده‌ایم: افرادی که مستعد بیماری هستند، افرادی که در دوره نهفتگی بیماری قرار دارند، افرادی که که تحت واکسیناسیون قرار می‌گیرند، افرادی که دچار هپاتیت B حاد هستند، افرادی که دچار هپاتیت B مزمن هستند و کسانی که بهبود یافته‌اند. همچنین کل جمعیت را به دو دسته‌ی کسانی که معتاد تزریقی هستند و یا به طریقی با تزریقات و خون انسان سروکار دارند و کسانی که این‌گونه نیستند، تقسیم می‌کنیم. با توجه به اینکه با گذشت زمان تعداد جمعیت مورد نظر تغییر می‌کند، مدل ما یک مدل وابسته به زمان خواهد بود.

مباحث ارائه شده در این مقاله به این صورت است که: مطالب مربوط به مدل‌سازی ریاضی شامل تعیین متغیرها و پارامترها، فرمول‌بندی معادلات و تشکیل دستگاه معادلات متناظر با آن که در اینجا یک دستگاه معادلات معمولی غیرخطی است، در بخش دوم ارائه می‌شود. در بخش سوم ، ماتریس نسل دوم مدل ریاضی در نقطه تعادل عاری از بیماری ارائه شده و فرمول عدد شیوع محاسبه می‌گردد.

۲ مدل‌سازی ریاضی

در این بخش به ارائه مدل ریاضی هپاتیت B برای دو گروه متمایز از افراد می‌پردازیم. گروه اول معتادان تزریقی، کارکنان بخش بهداشت و کسانی هستند که به طریقی با تزریقات و خون انسان سروکار دارند که این گروه را N_bمی‌نامیم. گروه دوم نیز سایر افراد جامعه هستند، یعنی کسانی که متعلق به N_b نباشند، که گروه دوم را N_n می‌نامیم. کل جمعیت مورد نظر را با فرض مهاجرت ثابت، در زمان t ، $N(t)$ در نظر می‌گیریم و داریم $N(t) = N_b(t) + N_n(t)$.

گروه N_b را به شش دسته تقسیم می‌کنیم: $S_b(t)$: افرادی که مستعد بیماری هستند، $E_b(t)$: افراد آلوده به ویروس HBV که بیماری آنها در مرحله نهفتگی بیماری قرار دارد، $I_b(t)$: افرادی که دچار بیماری هپاتیت B حاد گشته‌اند، $C_b(t)$: افرادی که بیماری هپاتیت B آنها مزمن شده است، $V_b(t)$: افرادی که تحت واکسیناسیون قرار گرفته‌اند، $R_b(t)$: افرادی که بهبود یافته‌اند. به همین ترتیب گروه N_n را نیز به شش دسته $S_n(t)$، $E_n(t)$، $I_n(t)$، $C_n(t)$، $V_n(t)$ و $R_n(t)$ تقسیم می‌کنیم. به‌عبارت دیگر:

$$N_b(t) = S_b(t) + E_b(t) + I_b(t) + C_b(t) + V_b(t) + R_b(t),$$

$$N_n(t) = S_n(t) + E_n(t) + I_n(t) + C_n(t) + V_n(t) + R_n(t).$$

این دسته‌ها متغیرهای مدل هستند. این نکته قابل ذکر است که طرح واکسیناسیون هپاتیت B برای نوزادان بطور کامل انجام می‌شود و این‌که واکسیناسیون به معنی مصونیت کامل از هپاتیت B نیست. در ادامه پارامترهای مؤثر در مدل را برای گروه N_b بصورت زیر ارائه می‌دهیم:

β: نرخ انتقال بیماری،

p_b: نرخ واکسیناسیون افراد مستعد بیماری،

α_b: نرخ واکسیناسیون افرادی که در مرحله نهفتگی بیماری قرار دارند،

t: نرخ جابجایی افراد از مرحله E_b به مرحله I_b،

γ: نرخ خارج شدن از مر حله I_b،

r: نرخ جابجایی از مرحله C_b به R_b،

δ: احتمال بیمار شدن افرادی که تحت واکسیناسیون قرار گرفته‌اند،

η: نرخ جابجایی از V_b به R_b،

μ: نرخ مرگ‌ومیر طبیعی،

λ: نرخ زاد و ولد.

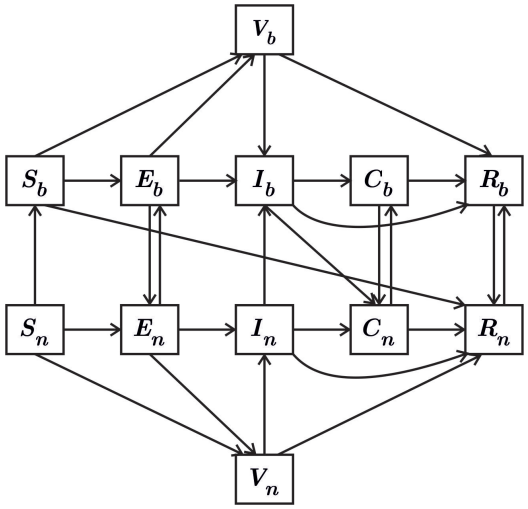
تمام پارامترهای بالا برای گروه N_n نیز برقرارند بجز نرخ‌های واکسیناسیون که برای گروه N_n این نرخ‌ها p_n و α_n خواهند بود. اگر $I = I_b + I_n$ و $C = C_b + C_n$ باشند، θ را احتمال جابجایی افراد از I به C در نظر می‌گیریم و همچنین فرض می‌کنیم که u نرخ جابجایی افراد از گروه N_n به گروه N_b باشد.

حال با توجه به متغیرها و پارامترهای تعریف شده در بالا، به فرمول‌بندی مدل ریاضی بیماری هپاتیت B می‌پردازیم که منجر به دستگاه معادلات دیفرانسیل غیرخطی زیر می‌شود:

$$
\begin{aligned}
\dot{S_b} &= -\beta(۱-p_b)S_b(I+C) - (\mu+p_b)S_b + uS_n \\
\dot{E_b} &= \beta(۱-p_b)S_b(I+C) - (\alpha_b + (۱-\alpha_b)t + \mu)E_b + uE_n \\
\dot{I_b} &= (۱-\alpha_b)tE_b - (\mu + \gamma(۱-\theta) + \theta)I_b + \beta\delta(I+C)V_b + uI_n \\
\dot{C_b} &= \theta I_b - (\mu + r)C_b + uC_n \\
\dot{V_b} &= \alpha_b E_b + p_b S_b - (\beta\delta(I+C) + \eta)V_b + uV_n \\
\dot{R_b} &= \gamma(۱-\theta)I_b + \eta V_b + rC_b - \mu R_b + uR_n \\
\dot{S_n} &= -\beta(۱-p_n)S_n(I+C) - (u+\mu+p_n)S_n \\
\dot{E_n} &= \beta(۱-p_n)S_n(I+C) - (\alpha_n + (۱-\alpha_n)t + \mu + u)E_n + \lambda(E_b+E_n) \\
\dot{I_n} &= (۱-\alpha_n)tE_n - (\mu + \gamma(۱-\theta) + \theta + u)I_n + \beta\delta(I+C)V_n \\
\dot{C_n} &= \theta I_n - (\mu + r + u)C_n + \lambda(I+C) \\
\dot{V_n} &= \alpha_n E_n + p_n S_n - (\beta\delta(I+C) + \eta + u)V_n \\
\dot{R_n} &= \gamma(۱-\theta)I_n + \eta V_n + rC_n - (\mu+u)R_n + \lambda(R_b + R_n + S_b + S_n)
\end{aligned}
\tag{۱.۲}
$$

در شکل ۱ نمودار گردشی این مدل ارائه شده است.

توجه داشته باشید، در این مدل فرض بر این است که هر نوزاد متولد شده از مادر مبتلا به هپاتیت B حاد و یا مزمن، دچار هپاتیت B مزمن می‌شود. همچنین نرخ جابجایی از N_n به N_b را می‌توان همان نرخ گرایش افراد گروه N_n به اعتیاد تزریقی در نظر گرفت.



شکل ۱: نمودار گردشی مدل ریاضی بیماری هپاتیت B

۳ نقطه تعادل رهایی از بیماری و عدد شیوع

جواب‌های دستگاه معادلات (۱) بصورت $Q = (S_b, E_b, I_b, C_b, V_b, S_n, E_n, I_n, C_n, V_n)$ هستند که قضیه زیر مثبت بودن این جواب‌ها را نشان می‌دهد [۷].

قضیه ۱: اگر مقادیر $S_b(۰)$، $E_b(۰)$، $I_b(۰)$، $C_b(۰)$، $V_b(۰)$، $S_n(۰)$، $E_n(۰)$، $I_n(۰)$، $C_n(۰)$ و $V_n(۰)$ نامنفی باشند، آنگاه برای $t > ۰$ مقادیر $S_b(t)$، $E_b(t)$، $I_b(t)$، $C_b(t)$، $V_b(t)$، $S_n(t)$، $E_n(t)$، $I_n(t)$، $C_n(t)$ و $V_n(t)$ نامنفی هستند.

اگر فرض کنیم که در جمعیت N هیچ بیماری وجود نداشته باشد، در این صورت نقطه تعادل عاری از بیماری دستگاه (۱) به‌صورت $Q_۰ = (S_{b۰}, ۰, ۰, ۰, V_{b۰}, S_{n۰}, ۰, ۰, ۰, V_{n۰})$ خواهد بود.

عدد شیوع بیماری، تعداد افرادی است که یک فرد بیمار، در طول دوره بیماری خود، آنها را مبتلا می‌کند. با توجه به [۷]، عدد شیوع $R_۰$ درواقع همان شعاع طیفی ماتریس نسل دوم $FV^{-۱}$ است. ماتریس‌های F و V را از دستگاه (۱) به‌صورت زیر بدست می‌آوریم:

$$F = \begin{bmatrix} ۰ & \beta(۱-p_b)S_{b۰} & \beta(۱-p_b)S_{b۰} & ۰ & \beta(۱-p_b)S_{b۰} & \beta(۱-p_b)S_{b۰} \\ ۰ & ۰ & ۰ & ۰ & ۰ & ۰ \\ ۰ & ۰ & ۰ & ۰ & ۰ & ۰ \\ ۰ & \beta(۱-p_n)S_{n۰} & \beta(۱-p_n)S_{n۰} & ۰ & \beta(۱-p_n)S_{n۰} & \beta(۱-p_n)S_{n۰} \\ ۰ & ۰ & ۰ & ۰ & ۰ & ۰ \\ ۰ & ۰ & ۰ & ۰ & ۰ & ۰ \end{bmatrix},$$

و

$$V=\begin{bmatrix} \alpha_b+(۱-\alpha_b)t+\mu & ۰ & ۰ & -u & ۰ & ۰ \\ -(۱-\alpha_b)t & m & -\beta\delta V_{b۰} & ۰ & -\beta\delta V_{b۰}-u & -\beta\delta V_{b۰} \\ ۰ & -\theta & \mu+r & ۰ & ۰ & -u \\ -\lambda & ۰ & ۰ & \alpha_n+(۱-\alpha_n)t+\mu-\lambda & ۰ & ۰ \\ ۰ & -\beta\delta V_{n۰} & -\beta\delta V_{n۰} & -(۱-\alpha_n)t & n & -\beta\delta V_{n۰} \\ ۰ & -\lambda & -\lambda & ۰ & -\theta-\lambda & \mu+r+u-\lambda \end{bmatrix},$$

که $m=\mu+\gamma(۱-\theta)+\theta-\beta\delta V_{b۰}$ و $n=\mu+\gamma(۱-\theta)+\theta+u-\beta\delta V_{n۰}$ است. شعاع طیفی ماتریس $FV^{-۱}$ به‌صورت زیر خواهد بود:

$$\begin{aligned} R_{۰}=&\beta(۱-p_n)S_{n۰}\big[\big(\frac{۱-\alpha_n t}{bx}\big)\big(۱-z+\frac{\lambda+\theta}{f}\big)+\big(\frac{۱-n_۲-n_۳-n_۴}{bn_۱}\big)\big(\frac{-y(۱-\alpha_n t)}{x}+u(۱-\alpha_b t)\big)\big]\\ &+\beta(۱-p_b)S_{b۰}\big[\big(\frac{۱-\alpha_n t}{bx}\big)\big(\frac{\lambda}{(۱-t)\alpha_b+t+\mu}\big)\big(۱-z+\frac{\lambda+\theta}{f}\big)\\ &+\big(\frac{۱-n_۲-n_۳-n_۴}{n_۱}\big)\big(\frac{\lambda u(۱-\alpha_b t)}{b((۱-t)\alpha_b+t+\mu)}-\frac{\lambda y(۱-\alpha_n t)}{b((۱-t)\alpha_b+t+\mu)}+g\big)\big]. \end{aligned}$$

که $a=\gamma(۱-\gamma)\theta+\mu-\beta\delta V_{b۰}$، $b=(۱-t)\alpha_n+t+\mu-\lambda(۱+u)$، $c=\beta\delta V_{n۰}\big(\frac{\mu+r-\theta}{\theta}\big)$، $d=\gamma(۱-\gamma)\theta+u-\beta\delta V_{n۰}$، $e=\big(\frac{\mu+r-\theta}{\theta}\big)(\gamma(۱-\gamma)\theta+\mu-\beta\delta V_{b۰}$، $f=\frac{-\lambda(\mu+r+\theta)}{\theta}$، $g=\frac{u\lambda}{\theta}+\mu+r+u-\lambda$، $h=\beta\delta V_{n۰}\big(\frac{u-\theta}{\theta}\big)$، $x=d+\frac{c(\theta+\lambda)}{f}$، $y=\frac{e(\theta+\lambda)}{f}\beta\delta V_{b۰}-u$، $z=(-\frac{(\mu+r)}{\theta})\frac{(\theta+\lambda)}{f}$، $n_۱=\frac{h}{x}-\frac{cg}{fx}$، $n_۲=-\beta\delta V_{b۰}-\frac{eg}{f}+yn_۱$، $n_۳=\frac{g}{f}+\big(\frac{\lambda+\theta}{f}\big)n_۱$ و $n_۴=\frac{g}{f}\big(\frac{\mu+r}{\theta}\big)+\frac{u}{\theta}+zn_۱$.

مراجع

[60] R. Akbari, A. Vahidian, A. A. Heydari, A. Heydari, *The Analysis of a Disease-free Equilibrium of Hepatitis B model*, Sahand communiation in mathematical analysis, $3(2):1-11, 2016$.

[61] S. Alavian, B. Hajarizadeh, M. Ahmadzad-Asl, A. Kabir, K. Bagheri-Lankarani, *Hepatitis B Virus Infection in Iran: A Systematic Review*, Hepatitis Monthly, $8:281-294, 2008$.

[62] R.M. Anderson and R.M. May, *Infectious Disease of Humans: Dynamics and Control*, Oxford University Press, Oxford, 1991.

[63] H. Keyvani, M. Sohrabi, F. Zamani, H. Poustchi, H. Ashrafi, F. Saeedian, M.Mooadi, N. Motamed, H. Ajdarkosh, M. Khonsari, G. Hemmasi, M. Ameli, A. Kabir, M. Khodadost, *A Population Based Study on Hepatitis B Virus in Northern Iran*, *Amol*, Hepat Mon, $14:1-8, 2014$.

[64] G.F. Medley and N.A. Lindop, *Hepatitis-B virus endemicity: heterogeneity*, *cat-astrophic dynamics and control*, Nature Medicine, $7(5):619-624, 2001$.

[65] J. Pang, J.A. Cui, and X. Zhou, *Dynamical behavior of a hepatitis B virus transmission model with vaccination*, Journal of Theoretical Biology, $265:572-578, 2010$.

[66]]. Shaban, S. S. Massaw, O. Daniel Makinde, *Modelling the effect of screening on the spread of HIV infection in a population with variable inflow of infective immigrants*, Scientific Research and Essays, $6 : 4397 - 4405, 2011$.

[67] P. Van den Driessche and J. Watmough, *Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission*, Math. Biosci., $180(1) : 29 - 48, 2002$.

[68] K. Wang, A. Fan, A. Torres, *Global Properties of an Improved Hepatitis B Virus Model*, Nonlinear analysis: Real World applications, $11(4) : 3131 - 3138, 2010$.

[69] T. Waezizadeh, M. Mohammad Rezaei, *Dynamical Behavior of HBV in a Population*, Cankaya University Journal of Science and Engineering, $14 : 29 - 41, 2017$.

[70] S.J. Zhao, Z.Y. Xu, and Y. Lu, *A mathematical model of hepatitis B virus transmission and its application for vaccination strategy in China*, Int.J.Epidemiol., $29 : 744 - 752, 2000$.

[71] X. Q. Zhao, *Dynamical Systems in Population Biology*, Springer-Verlag New York, 2003.

اولین همایش ملی ریاضیات زیستی
۲۱ و ۲۲ اسفندماه ۱۳۹۷
همایش ملی ریاضیات زیستی
دانشگاه نیشابور

کاربرد رگرسیون لجستیک جریمه شده با لاسوی تطبیقی در تحلیل داده‌های با بعد بالا

نجمه عسگری فرسنگی [۱]، کارشناس ارشد آمار زیستی، گروه اپیدمیولوژی و آمار زیستی، دانشگاه علوم پزشکی مشهد، مشهد، ایران
راضیه یوسفی کارشناس ارشد آمار زیستی، گروه اپیدمیولوژی و آمار زیستی، دانشگاه علوم پزشکی مشهد، مشهد، ایران
محمد تقی شاکری استاد آمار زیستی، گروه اپیدمیولوژی و آمار زیستی، دانشگاه علوم پزشکی مشهد، مشهد، ایران

چکیده: با تغییر سبک زندگی و افزایش شیوع سرطان، تشخیص اولیه بیماری از اهمیت زیادی برخوردار می‌باشد. در راستای شناسایی ژن‌های افتراق دهنده‌ی بیماران از افراد سالم در مطالعات سرطان‌شناسی، الگوهای بیان ژن در دو بافت سالم و سرطانی با یکدیگر مقایسه می‌شوندکه به‌دلیل بالا بودن بعد داده‌های بیان ژنی استفاده از روش‌های معمول آماری از جمله رگرسیون لجستیک مناسب نیست. به این منظور از روش های رگرسیونی اصلاح شده استفاده می‌شود. رگرسیون لجستیک جریمه شده با لاسو یکی از محبوب ترین روش ها است که انتخاب متغیر و برآورد احتمالات را به‌طور همزمان انجام می‌دهد. لاسو تمامی متغیرها را به یک اندازه جریمه می‌کند در صورتی که دارای درجه اهمیت یکسانی نیستند. در این راستا رگرسیون لجستیک جریمه شده با لاسوی تطبیقی در جهت اصلاح جریمه‌ی لاسو معرفی گردید که وزن‌های متفاوتی برای ضرایب در عبارت جریمه در نظر می‌گیرد و در نتیجه نتایج لاسو را بهبود می‌بخشد.

واژه‌های کلیدی: رگرسیون لجستیک جریمه شده، لاسوی تطبیقی، داده‌های با بعد بالا

۱ مقدمه

با پیشرفت تکنیک‌هایی با فناوری بالا و تولید حجم عظیم داده در بیولوژی، به‌کار بردن روش‌هایی برای تحلیل این داده‌ها در جهت پی بردن به سیستم پیچیده پویای حیات مورد توجه است. در بین فناوری‌های بسیار پیشرفته‌ی اخیر، ریزآرایه‌ی[۱] دی‌ان‌ای (DNA)[۲] از مشهورترین‌ها است که بیان هزاران ژن را به‌طور همزمان در بافت موردنظر اندازه‌گیری می‌کند. در مطالعات سرطان‌شناسی، الگوهای بیان ژن در دو بافت سالم و سرطانی با یکدیگر مقایسه می‌شوند. غالباً به‌دلیل هزینه نسبتاً بالا، آزمایش‌های ریزآرایه‌ای معمولاً در تعداد اندکی نمونه انجام می‌شوند.

انتخاب متغیر در شرایطی که تعداد متغیرها (p) بسیار بیشتر از تعداد نمونه‌ها (n) است(داده‌های با بعد بالا) در روش‌های آماری جدید

[1] Microarray
[2] Deoxyribonucleic acid

[۱] نجمه عسگری فرسنگی: asgarifn941@gmail.com

و یادگیری ماشین از اهمیت ویژه‌ای برخوردار بوده و استفاده از روش‌های آماری در طبقه بندی داده‌های با بعد بالا (شرایط $p > n$) کاری چالش برانگیز است(۷۲).

یکی از محبوب‌ترین مدل‌های خطی تعمیم یافته، برای تحلیل رابطه‌ی یک یا چند متغیر توضیحی بر متغیر پاسخ دو یا چندتایی رگرسیون لجستیک است که برای دستیابی به نتایج پایا[۳] و قابل قبول به حجم داده‌هایی بیش از آنچه که در رگرسیون خطی معمولی مواجه هستیم نیاز می‌باشد(۷۳). از دیدگاه آماری نیز وجود تعداد زیادی متغیر، منجر به بیش‌برازشی شده و عملکرد طبقه‌بندی را کاهش می‌دهد((۷۴)، (۷۵)).

مدل‌های رگرسیونی جریمه شده در این شرایط کارا هستند که عبارات جریمه مختلفی در مقالات مورد بررسی قرار گرفته‌اند((۷۶)–(۷۸)). یکی از محبوب‌ترین عبارات جریمه، عملگر انتخاب و انقباض کمترین قدرمطلق (LASSO) [۴] است که توسط تیبشیرانی در سال ۱۹۹۶ برای رگرسیون خطی معرفی شد(۷۷). لاسو با انتخاب متغیرهای کمتر در داده‌های با بعد بالا مدل‌های ساده، تفسیرپذیر و دارای صحت پیش‌بینی بالا ارائه می‌دهد.

جریمه کردن تمام متغیرها به یک اندازه در صورتی که دارای درجه‌ی اهمیت یکسانی نیستند یکی از ضعف‌های لاسو است. به منظور بهبود نتایج لاسو زو، لاسوی تطبیقی [۵] را پیشنهاد کرد که وزن‌های متفاوتی برای ضرایب در نظر می‌گیرد و نشان داد اگر وزن‌ها وابسته به داده‌ها بوده، لاسوی وزن‌دار شده دارای خاصیت الهام بخشی شده، اریبی در انتخاب [۶] کاهش یافته و مدل پایدارتری حاصل می‌شود(۷۹). در این مقاله رگرسیون لجستیک جریمه شده با لاسوی تطبیقی معرفی شده و کارایی روش از نظر تعداد متغیرهای انتخابی و صحت پیش بینی مدل در مجموعه داده‌ی ریزآرایه بررسی گردید.

۲ رگرسیون لجستیک جریمه شده با لاسوی تطبیقی

رگرسیون لجستیک یکی از روش‌های محبوب آماری در پزشکی است که در تحلیل‌ها به‌طور گسترده استفاده می‌شود ولی شواهدی وجود دارد که در داده‌های باحجم نمونه کم و تعداد متغیر زیاد نتایج پایایی نداشته و کند عمل می‌کند. به منظور استفاده از رگرسیون لجستیک در داده‌های با بعد بالا، با اضافه کردن عبارتی نامنفی تحت عنوان جریمه به منفی لگاریتم تابع درست‌نمایی نمونه، مدل قابل استفاده می‌شود.

عبارات جریمه متفاوتی در متون مورد بررسی قرار گرفته‌اند که یکی از محبوب‌ترین عبارات جریمه، جریمه l_1 معروف به لاسو است که این عبارت جریمه قابل تعمیم به مدل‌های خطی تعمیم یافته، از جمله رگرسیون لجستیک نیز است. علی‌رغم مزیت لاسو برای استفاده در داده‌های با بعد بالا، یکی از ضعف‌های آن در نظر گرفتن جریمه‌ی یکسان برای تمام متغیرها می‌باشد(۸۰). به منظور بهبود نتایج لاسو زو، لاسوی تطبیقی را پیشنهاد کرد(۷۹). لاسوی تطبیقی وزن‌های متفاوتی برای ضرایب مختلف در نظر گرفته و برآورد ضرایب مدل از رابطه (۱.۲) به‌دست می‌آید.

$$\widehat{\beta}_{APLR} = \underset{\beta}{argmin} \left[-\sum_{i=1}^{n} \{ y_i \ln(\pi_i) + (1 - y_i) \ln(1 - \pi_i) \} + \lambda \sum_{j=1}^{p} w_j |\beta_j| \right] \qquad (۱.۲)$$

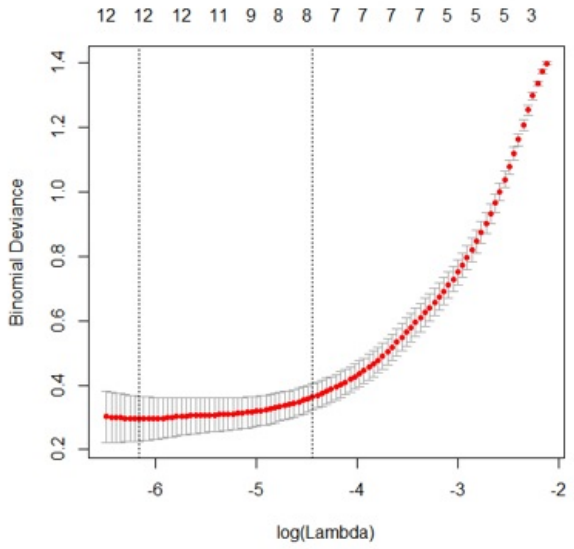
در رابطه (۱.۲) $w_j = (|\beta_j|)^{-\gamma}$ بردار وزن‌های معلوم بوده و γ یک مقدار مثبت ثابت فرض می‌شود. زو نشان داد، اگر وزن‌ها وابسته به داده‌ها باشند لاسوی وزن‌دار شده دارای خاصیت الهام بخشی می‌شود یعنی احتمال انتخاب مجموعه‌ای از متغیرها به‌درستی، همگرا به یک است که در نتیجه اریبی در انتخاب کاهش یافته و مدل پایدارتری به‌دست می‌آید(۸۱). زو از معکوس برآوردهای حداکثر

[3] Stable
[4] Least Absolute Shrinkage and Selection Operator
[5] Adaptive LASSO
[6] Selection bias

درست‌نمایی به‌عنوان وزن اولیه استفاده کرد، که در داده‌های با بعد بالا معتبر نیستند(۷۹). بوهلمن و وندگیر به‌منظور به‌کارگیری لاسوی تطبیقی در مجموعه داده‌های با بعد بالا از برآوردگرهای لاسو برای وزن اولیه لاسوی تطبیقی استفاده کردند(۸۲).

۳ مثال کاربردی در داده های واقعی

به منظور بررسی کارکرد لاسوی تطبیقی در داده های با بعد بالا ، از داده‌های ریزآرایه سرطان روده بزرگ شامل ۲۲۲۷۷ متغیر(ژن) و حجم نمونه ۱۱۱ با کد GSE44861 در پایگاه عمومی GEO [۷] استفاده شد. داده‌ها در محیط نرم‌افزاری R3.5.1 فراخوانی و با استفاده از بسته‌ی آماری glmnet، مدل لاسوی تطبیقی به داده‌ها برازش داده‌شد. به این منظور ابتدا مدل رگرسیون لجستیک جریمه شده با لاسو به داده‌ها برازش داده شد سپس از معکوس ضرایب مدل لاسو به عنوان وزن اولیه در لاسوی تطبیقی استفاده گردید. پارامتر لاندای مناسب برای برازش مدل، مقداری است که به‌ازای آن مدل برازش شده دارای کمترین خطا باشد. با استفاده از اعتبارسنجی متقاطع پارامتر لاندای بهینه برای برازش مدل تعیین و مدل نهایی برازش داده شد و تنها ۱۲ متغیر توسط مدل انتخاب گردید (شکل ۱).



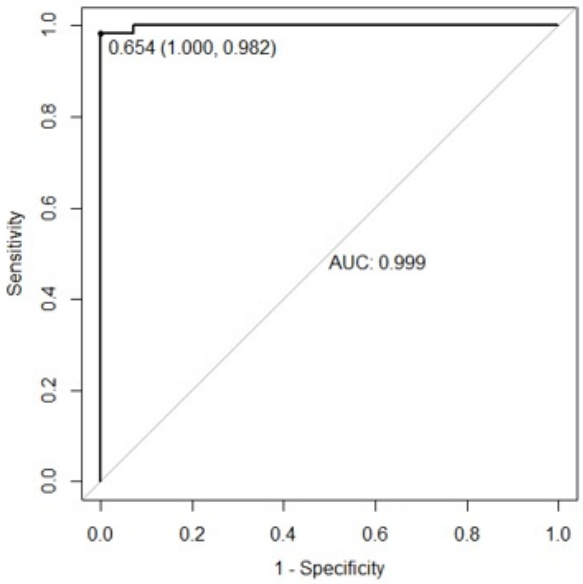
شکل ۱: اعتبارسنجی متقاطع مدل لاسوی تطبیقی

صحت مدل لاسوی تطبیقی با استفاده از سطح زیر منحنی مشخصه عملکرد (ROC) (۰/۹۹۶, ۱)۰/۹۹۹ ، حساسیت و ویژگی مدل نیز به ترتیب ۰/۹۸ و یک بود (شکل ۲).

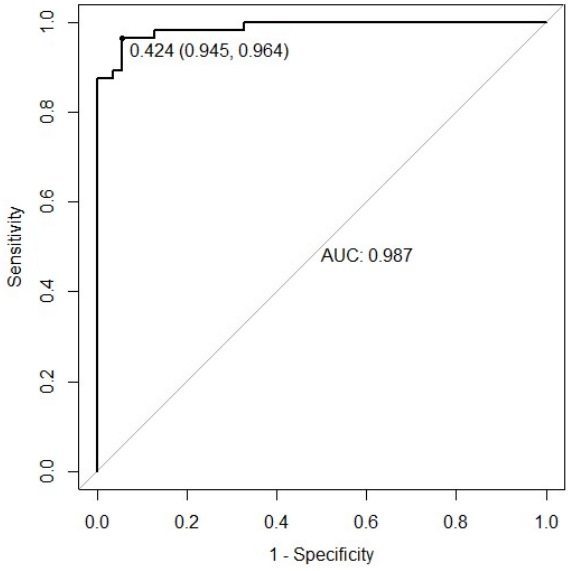
۴ بحث

کاربرد مهم داده‌های ریزآرایه شناسایی ژن‌های مرتبط با سرطان است که به‌دلیل بالا بودن بعد این داده‌ها، روش‌های رگرسیونی معمول کارایی لازم را ندارند. به منظور استفاده از رگرسیون لجستیک برای طبقه بندی افراد به دو دسته بیمار و سالم در داده‌های با بعد بالا از روش‌های رگرسیونی جریمه شده استفاده می‌شود. عبارت جریمه‌ی لاسو به دلیل برآورد ضرایب و انتخاب متغیر به طور همزمان از محبوب‌ترین‌ها است. لاسو دارای معایبی می‌باشد که زو در راستای بهبود بخشیدن به نتایج آن لاسوی تطبیقی را معرفی کرد.

[7] Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo)



شکل ۲: منحنی مشخصه عملکرد مدل لاسوی تطبیقی



شکل ۳: منحنی مشخصه عملکرد مدل لاسو

زو نشان داد که لاسوی تطبیقی در شرایط $p < n$ ضرایب غیر صفر را با احتمال یک به درستی انتخاب می‌کند(۷۹). هوآنگ و همکاران در مطالعه‌ای عملکرد لاسوی تطبیقی در مقایسه با لاسو را در شرایط $n > p$ با استفاده از شبیه‌سازی و در داده‌های واقعی بررسی کردند و نشان دادند در داده‌های با بعد بالا، لاسو و لاسوی تطبیقی از نظر عملکرد یکسان بوده هرچند لاسوی تطبیقی از لاسو از نظر تعداد متغیرهای انتخابی و صحت پیش‌بینی کمی بهتر عمل می‌کند(۸۳). در مطالعه‌ای الجمال و لی رگرسیون لجستیک جریمه شده با لاسوی تطبیقی با وزن اولیه منطبق بر همبستگی را معرفی کرده و نشان دادند از لاسوی تطبیقی با وزن اولیه‌ی برآوردگر لاسو از نظر تعداد متغیرهای انتخابی و صحت پیش‌بینی، مدل بهتر عمل می‌کند، که البته در صورت وجود تنها یک همبستگی کامل بین متغیرها نمی‌توان از این روش استفاده کرد(۸۴). در پژوهشی رئیسی و همکاران به منظور شناسایی ژن‌هایی که نقش مهمی در سرطان مثانه دارند

از دو روش جریمه شده‌ی لاسو و لاسوی تطبیقی استفاده کردند. در این مطالعه‌ی مورد شاهدی ۲۵ بیمار و ۲۳ فرد سالم شرکت داشتند که میزان بیان ۲۲ ژن در آن‌ها اندازه‌گیری شد. ۱۲ ژن با صحت پیش‌بینی ۰/۷۱ در روش لاسو و ۴ ژن با صحت پیش‌بینی ۰/۸۹ با استفاده از روش لاسوی تطبیقی شناسایی شد که صحت پیش‌بینی دو مدل اختلاف معنی داری داشتند(p-value= ۰/۰۰۹)(۸۵).

نتایج مطالعه حاضر نیز نشان داد که لاسوی تطبیقی در داده‌های با بعد بالا قابلیت انتخاب تعداد کمی متغیر با صحت پیش بینی بالا را داشته و مدلی ساده و تفسیرپذیر نتیجه می‌دهد.

مراجع

[72] Piao Y, Piao M, Park K, Ryu K-H. *An ensemble correlation-based gene selection algorithm for cancer classification with gene expression data*. Bioinformatics (Oxford, England). 2012;28:3306–15.

[73] Agresti A. *An introduction to categorical data analysis. 2nd ed.* New Jersey: John Wiley & Sons 2007.

[74] Chen K-H, Wang, K.-J., Wang, K.-M., & Angelia, M.-A. *Applying particle swarm optimization-based decision tree classifier for cancer classification on gene expression dataAppl Soft Comput.* 2014;24:773-80.

[75] Peng H, Fu Y, Liu J, Fang X, Jiang C. *Optimal gene subset selection using the modified SFFS algorithm for tumor classification. Neural Comput Appl.* 2013;23:1531–8.

[76] Liang Y, Liu C, Luan X-Z, Leung K-S, Chan T-M, Xu Z-B, et al. *Sparse logistic regression with a L1/2 penalty for gene selection in cancer classification*. BMC Bioinformatics. 2013;14:198-210.

[77] Tibshirani R. *Regression shrinkage and selection via the lasso. J Royal Statist Soc B*. 1996;58(1) :267-88.

[78] Zhenqiu L, Feng J, Guoliang T, Suna W, Fumiaki S, Ming T. *Sparse logistic regression with Lp penalty for biomarker identification.* Stat Appl Genet Mol Biol. 2007;6:1-22.

[79] Zou H. *The adaptive lasso and its oracle properties.* J Am Stat Assoc. 2006;101(476):1418-29.

[80] Zou H, Hastie T. *Regularization and variable selection via the elastic net.* J R Stat Soc Series B Stat Methodol. 2005;67(2):301-20.

[81] Fan J, Li R. *Variable Selection via Nonconcave Penalized Likelihood and its Oracle Properties.* J Am Stat Assoc. 2001;96(456):1348-60.

[82] Bühlmann P, VanDeGeer S. *Statistics for high-dimensional data: Methods,theory and applications [Internet]*. Switzerland: Springer, Berlin, Heidelberg; 2011. Available from: https://link.springer.com/book/10.1007/978-3-642-20192-9.

[83] Huang J, Ma S, Zhang C-H. *Adaptive lasso for sparse high-dimensional regression models.* Stat Sin. 2008;18:1603-18.

[84] Algamal Z, Lee M. *Regularized logistic regression with adjusted adaptive elastic net for gene selection in high dimensional cancer classification.* Comput Biol Med. 2015;(6 (136-45.

[85] Raeisi H, Jaberipoor M, Zare N, Hosseini A. *The Role of 22 Genes Expression in Bladder Cancer by Adaptive LASSO*. Iran J Cancer Prev. 2016;9(6).

تحلیلی بر مدل بیماری سرطان مبتنی بر نظریه آشوب

میلاد فهیمی [۱]، **کارشناسی ارشد ریاضیات کاربردی دانشگاه سمنان کاظم نوری عضو هیات علمی دانشکده ریاضی و علوم کامپیوتر دانشگاه سمنان لیلا ترک زاده عضو هیات علمی دانشکده ریاضی و علوم کامپیوتر دانشگاه سمنان**

چکیده: ددر این مقاله، یک مدل ریاضیاتی سه بعدی مبتنی بر نظریه آشوب–بی نظمی برای بیماری سرطان مورد بررسی و تحلیل قرارگرفته است؛ بدین منظور در راستای واقع گرایانه شدن نتایج حاصل از تحلیل مقدار پارامترها ، در تشکیل فرم دیفرانسیلی مدل ، از مشتقات مرتبه کسری کاپوتو–لیوویل استفاده شده است و سپس با استفاده از خواص موجود در حسابان کسری ، شرایط لازم برای وجود جوابی برای مدل مطرح شده مورد بررسی قرار گرفته است.

واژه‌های کلیدی: مدل سرطان ، نظریه آشوب ، حسابان کسری ، مشتق کسری

۱ مقدمه

مدل های ریاضیاتی در ساختار های مختلف به مطالعه و تحلیل رفتار بیماری ها از قبیل آنفولانزا ، ایدز و سرطان پرداخته‌اند و هر یک جنبه های خاصی از آن بیماری از جمله نحوه کنترل ، انتشار و انقراض آن را مورد بررسی قرار داده اند. در همه این مدل ها هدف اصلی و اساسی مطالعات، رسیدن به فهم واقع بینانه از مکانیسم رشد یا کنترل بیماری و پیش بینی رفتار آن است.امروزه با توسعه علوم ریاضی و ظهور مفاهیم جدید در این علم، همواره سعی شده مدل هایی ارائه شود که تطابق بیشتری با واقعیت دارند[۴،۳]. برای نمونه با استفاده از مفاهیم معادلات دیفرانسیل کسری می توان مدل های واقع گرایانه تری ارائه داد. مبداء حساب و دیفرانسیل کسری به اواخر قرن هفدهم بازمی گردد که نیوتن و لایب نیتز از پیشگامان این حوزه بودند. استفاده از مشتقات کسری در مسائل اپیدمی ، رویکردی سودمند و واقع گرایانه برای مطالعه رفتار بیماری ها است که در همین راستا در سال های اخیر پژوهش هایی مبتنی بر نظریه آشوب–بی نظمی برای مطالعه رفتار بیماری ها انجام شده است[۳،۲،۱] .

[۱] میلاد فهیمی: fahimimilad@gmail.com

۱.۱ ساختار ریاضیاتی مدل بیماری سرطان:

در این مقاله مدل مطالعاتی رفتار بیماری سرطان به شرح زیر نظر گرفته شده است :

$$\begin{aligned}\dot{x}_۱ &= x_۱(t)(۱ - x_۱(t)) - Ax_۱(t)x_۲(t) - B(x_۱)(t)x_۳(t),\\ \dot{x}_۲(t) &= Cx_۲(t)(۱ - x_۲(t)) - Dx_۱(t)x_۲(t),\\ \dot{x}_۳(t) &= E\frac{x_۱(t)x_۳(t)}{x_۱(t) + F} - Gx_۱(t)x_۳(t) - Hx_۳(t),\end{aligned}$$

که در آن

$$A, B, C, D, E, F, G, H$$

پارامترهای مورد بررسی در معادله هستند[۲]. در این مدل سه دسته سلول های درگیر بیماری مورد مطالعه قرار می‌گیرد که در آن $x_۱(t)$ تعداد سلولهای سرطانی در لحظه t و $x_۲(t)$ تعداد سلولهای سالم در لحظه t و $x_۳(t)$ تعداد سلول‌های دارای مصونیت نسبت به بیماری در لحظه t را مشخص می کند . معادله اول میزان تغییر در تعداد سلول های سرطانی در لحظه را نشان می دهد و معادله دوم میزان تغییر در تعداد سلول های سالم که دارای روند رشد طبیعی هستند را نشان می دهد و معادله سوم میزان آمادگی سیستم ایمنی سلول ها در برابر بیماری را نشان می دهد. در این مقاله به منظور واقع گرایانه شدن بررسی ها و نتایج از عملگرهای مشتق کسری استفاده شده است ؛ بدین منظور به مقدماتی از حسابان کسری می پردازیم.

تعریف ۱.۱. عملگر مشتق کسری کاپوتو - لیوویل به صورت زیر قابل تعریف است [۴] :

$$\overset{\cdot}{CFC}\mathfrak{D}_t^{\gamma}\{f(t)\} = \frac{M(\gamma)}{n-\gamma}\int_{۰}^{t} f^n(\theta)\exp[-\frac{\gamma}{n-\gamma}(t-\theta)]\mathrm{d}\theta, \quad n - ۱ < \gamma \leq n,$$

که در آن $M(\gamma)$ تابع نرمال است به گونه ای که : $M(۰) = M(۱) = ۰$

۲ نتایج اصلی

حال در این بخش مدل مورد مطالعه برای بیماری سرطان را با استفاده از مشتقات کسری از نوع کاپوتو بررسی می کنیم و لذا به اعمال شرایط آن می پردازیم.

۱.۲ ساختار تغییر یافته مدل بیماری سرطان با اعمال عملگرها در حیطه حسابان کسری :

بنابراین شکل تغییر یافته مدل با در نظر گرفتن کرنل های نمایی به صورت زیر قابل ارائه خواهد بود: مدل(۲)

$$\begin{aligned}
C\dot{F}C\mathcal{D}_t^{\gamma}x_1(t) &= x_1(t)(1-x_1(t)) - Ax_1(t)x_2(t) - Bx_1(t)x_3(t),\\
C\dot{F}C\mathcal{D}_t^{\gamma}x_2(t) &= Cx_2(t)(1-x_2(t)) - Dx_1(t)x_2(t),\\
C\dot{F}C\mathfrak{D}_t^{\gamma} &= E\frac{x_1(t)x_3(t)}{x_1(t)+F} - Gx_1(t)x_3(t) - Hx_3(t),
\end{aligned}$$

با شرایط اولیه به صورت :

$$x_{1(0)}(t) = x_1(0); \quad x_{2(0)}(t) = x_2(0); \quad x_{3(0)}(t) = x_3(0).$$

که در آن $C\dot{F}C\mathfrak{D}_t$ نشان دهنده مشتق کسری از نوع کاپوتو می باشد. حال معادلات دیفرانسیل در مدل(۲) را به فرم معادلات انتگرال تغییر می دهیم و لذا داریم : مدل (۳)

$$\begin{aligned}
x_1(t) - x_1(0) &= C\dot{F}I_t^{\gamma}[x_1(t)(1-x_1(t)) - Ax_1(t)x_2(t) - Bx_1(t)x_3(t)],\\
x_2(t) - x_2(0) &= C\dot{F}I_t^{\gamma}[Cx_2(t)(1-x_2(t)) - Dx_1(t)x_2(t)],\\
x_3(t) - x_3(0) &= C\dot{F}I_t^{\gamma}[E\frac{x_1(t)x_3(t)}{x_1(t)+F} - Gx_1(t)x_3(t) - Hx_3(t)],
\end{aligned}$$

پس با اعمال مشتق کسری طبق تعریف (۱) خواهیم داشت:

$$\begin{aligned}
\Longrightarrow x_1(t) = x_1(0) &+ \frac{2(1-\gamma)}{(2-\gamma)M(\gamma)}[x_1(t)(1-x_1(t)) - Ax_1(t)x_2(t) - Bx_1(t)x_3(t)]\\
t &+ \frac{2\gamma}{(2-\gamma)M(\gamma)}\int_0^t [x_1(s)(1-x_1(s)) - Ax_1(s)x_2(s) - Bx_1(s)x_3(s)]\mathrm{d}s,\\
\Longrightarrow x_2(t) = x_2(0) &+ \frac{2(1-\gamma)}{(2-\gamma)M(\gamma)}[Cx_2(t)(1-x_2(t)) - Dx_1(t)x_2(t)]\\
&+\frac{2\gamma}{(2-\gamma)M(\gamma)}\int_0^t [Cx_2(s)(1-x_2(s)) - Dx_1(s)x_2(s)]\mathrm{d}s,\\
\Longrightarrow x_3(t) = x_3(0) &+ \frac{2(1-\gamma)}{(2-\gamma)M(\gamma)}[E\frac{x_1(t)x_3(t)}{x_1(t)+F} - Gx_1(t)x_3(t)Hx_3(t)]\\
&+\frac{2\gamma}{(2-\gamma)M(\gamma)}\int_0^t [E\frac{x_1(t)x_3(t)}{x_1(t)+F} - Gx_1(s)x_3(s) - Hx_3(s)]\mathrm{d}s.
\end{aligned}$$

بنابراین کرنل ها به صورت زیر قابل تعریف خواهند بود :

$$\begin{aligned}
\tau(t,x_1(t)) &= x_1(1-x_1(t)) - Ax_1(t)x_2(t) - Bx_1(t)x_3(t),\\
\upsilon(t,x_2(t)) &= Cx_2(t)(10x_2(t)(1-x_2(t)) - Dx_1(t)x_2(t),\\
\phi(t,x_3(t)) &= E\frac{x_1(t)x_3(t)}{x_1(t)+F} - Gx_1(t)x_3(t) - Hx_3(t).
\end{aligned}$$

قضیه ۱.۲. شرایط لیپ شیتز برای هر یک کرنل های سه گانه برقرار است .

اثبات. برای کرنل های سه گانه مورد نظر داریم :

$$\begin{aligned}\|\tau(t,x_1(t))-\tau(t,X_1(t))\| &= \|(x_1(t)-X_1(t))[1-(x_1(t)-X_1(t))]\\ &\quad -A(x_1(t)-X_1(t))x_2(t)-B(x_1(t)-X_1(t))x_3(t)\|,\\ \|\upsilon(t,x_2(t))-\upsilon(t,X_2(t))\| &= C\|(x_2(t)-X_2(t))[1-(x_2(t)-X_2(t))]\\ &\quad -Dx_1)t)(x_2(t)-X_2(t))\|,\end{aligned}$$

حال با استفاده از نامساوی کوشی داریم :

$$\begin{aligned}\|\phi(t,x_3(t))-\phi(t,X_3(t))\| = E\Bigg\| &\frac{x_1(t)(x_3(t)-X_3(t)))}{x_1(t)+F}\\ &-Gx_1(t)(x_3(t)-X_3(t)))-H(x_3(t)-X_3(t))\Bigg\|.\end{aligned}$$

$$\begin{aligned}\|\tau(t,x_1(t))-\tau(t,X_1(t))\| &\le \|(x_1(t)-X_1(t))[1-(x_1(t)-X_1(t))]\\ &\quad -A(x_1(t)-X_1(t))x_2(t)-B(x_1(t)-X_1(t))x_3(t)\|,\\ \|\upsilon(t,x_2(t))-\upsilon(t,X_2(t))\| &= C\|(x_2(t)-X_2(t))[1-(x_2(t)-X_2(t))]\\ &\quad -Dx_1)t)(x_2(t)-X_2(t))\|,\end{aligned}$$

$$\begin{aligned}\|\phi(t,x_3(t))-\phi(t,X_3(t)))\| \le E\Bigg\| &\frac{x_1(t)(x_3(t)-X_3(t))}{x_1(t)+F}\\ &-Gx_1(t)(x_3(t)-X_3(t))-H(x_3(t)-X_3(t))\Bigg\|;\end{aligned}$$

همچنین داریم :

$$\begin{aligned}x_1(t) &= \frac{2(1-\gamma)}{(2-\gamma)M(\gamma)}\tau(t,x_{1(n-1)})+\frac{2\gamma}{(2-\gamma)M(\gamma)}\int_0^t \tau(s,x_{1(n-1)})\mathrm{d}s,\\ x_2(t) &= \frac{2(1-\gamma)}{(2-\gamma)M(\gamma)}\tau(t,x_{2(n-1)})+\frac{2\gamma}{(2-\gamma)M(\gamma)}\int_0^t \tau(s,x_{2(n-1)})\mathrm{d}s,\\ x_3(t) &= \frac{2(1-\gamma)}{(2-\gamma)M(\gamma)}\phi(t,x_{3(n-1)})+\frac{2\gamma}{(2-\gamma)M(\gamma)}\int_0^t \phi(s,x_{3(n-1)})\mathrm{d}s.\end{aligned}$$

حال با در نظر گرفتن نرم هر یک از پارامترها و با استفاده از نامساوی مثلثی داریم:

$$\|Y_n(t)\| = \|x_{۱(n)}(t) - X_{۱(n-۱)}(t)\| \leq \frac{۲(۱-\gamma)}{(۲-\gamma)M(\gamma)}\|\tau(t, x_{۱(n-۱)}(t)) - \tau(t, X_{۱(n-۲)}(t))\| + \frac{۲\gamma}{(۲-\gamma)M(\gamma)}\left\|\int_۰^t \left[\tau(s, x_{۱(n-۱)}(s)) - \tau(s, X_{۱(n-۲)}(s))\right]\right\| \mathrm{d}s,$$

$$\|\Phi_n(t)\| = \|x_{۲(n)}(t) - X_{۲(n-۱)}(t)\| \leq \frac{۲(۱-\gamma)}{(۲-\gamma)M(\gamma)}\|\upsilon(t, x_{۲(n-۱)}(t)) - \upsilon(t, X_{۲(n-۲)}(t))\| + \frac{۲\gamma}{(۲-\gamma)M(\gamma)}\left\|\int_۰^t \left[\upsilon(s, x_{۲(n-۱)}(s)) - \upsilon(s, X_{۲(n-۲)}(s))\right]\right\| \mathrm{d}s,$$

$$\|\Psi_n(t)\| = \|x_{۳(n)}(t) - X_{۳(n-۱)}(t)\| \leq \frac{۲(۱-\gamma)}{(۲-\gamma)M(\gamma)}\|\phi(t, x_{۳(n-۱)}(t)) - \phi(t, X_{۳(n-۲)}(t))\| + \frac{۲\gamma}{(۲-\gamma)M(\gamma)}\left\|\int_۰^t \left[\phi(s, x_{۳(n-۱)}(s)) - \phi(s, X_{۳(n-۲)}(s))\right]\right\| \mathrm{d}s,$$

که در آن :

$$x_{۱(n)}(t) = \sum_{m=۰}^{\infty} Y_m(t); \quad x_{۲(n)}(t) = \sum_{m=۰}^{\infty} \Phi_m(t); \quad x_{۳(n)}(t) = \sum_{m=۰}^{\infty} \Psi_m(t).$$

از طرفی چون کرنل ها دارای خاصیت لیپ شیتز هستند لذا داریم :

$$\|Y_n(t)\| = \|x_{۱(n)}(t) - X_{۱(n-۱)}(t)\| \leq \frac{۲(۱-\gamma)}{(۲-\gamma)M(\gamma)}\| \triangle \|(x_{۱(n-۱)}(t)) - X_{۱(n-۲)}(t))\| + \frac{۲\gamma}{(۲-\gamma)M(\gamma)} \triangle_۲ \int_۰^t \left\|x_{۱(n-۱)}(s)) - X_{۱(n-۲)}(s))\right\| \mathrm{d}s,$$

$$\|\Phi_n(t)\| = \|x_{۲(n)}(t) - X_{۲(n-۱)}(t)\| \leq \frac{۲(۱-\gamma)}{(۲-\gamma)M(\gamma)} \triangle_۳ \|x_{۲(n-۱)}(t)) - X_{۲(n-۲)}(t)\| + \frac{۲\gamma}{(۲-\gamma)M(\gamma)} \triangle_۴ \int_۰^t \left\|x_{۲(n-۱)}(s) - X_{۲(n-۲)}(s)\right\| \mathrm{d}s,$$

$$\|\Psi_n(t)\| = \|x_{۳(n)}(t) - X_{۳(n-۱)}(t)\| \leq \frac{۲(۱-\gamma)}{(۲-\gamma)M(\gamma)} \triangle_۵ \|x_{۳(n-۱)}(t)) - X_{۳(n-۲)}(t)\| + \frac{۲\gamma}{(۲-\gamma)M(\gamma)} \triangle_۶ \int_۰^t \left\|x_{۳(n-۱)}(s) - X_{۳(n-۲)}(s)\right\| \mathrm{d}s,$$

$$\|Y_n(t)\| = \|x_{۱(n)}(t) - X_{۱(n-۱)}(t)\| \leq \frac{۲(۱-\gamma)}{(۲-\gamma)M(\gamma)}\| \triangle \|(x_{۱(n-۱)}(t)) - X_{۱(n-۲)}(t))\| + \frac{۲\gamma}{(۲-\gamma)M(\gamma)} \triangle_۲ \int_۰^t \left\|x_{۱(n-۱)}(s)) - X_{۱(n-۲)}(s))\right\| \mathrm{d}s,$$

$$\|\Phi_n(t)\| = \|x_{۲(n)}(t) - X_{۲(n-۱)}(t)\| \leq \frac{۲(۱-\gamma)}{(۲-\gamma)M(\gamma)} \triangle_۳ \|x_{۲(n-۱)}(t)) - X_{۲(n-۲)}(t)\| + \frac{۲\gamma}{(۲-\gamma)M(\gamma)} \triangle_۴ \int_۰^t \left\|x_{۲(n-۱)}(s) - X_{۲(n-۲)}(s)\right\| \mathrm{d}s,$$

$$\|\Psi_n(t)\| = \|x_{۳(n)}(t) - X_{۳(n-۱)}(t)\| \leq \frac{۲(۱-\gamma)}{(۲-\gamma)M(\gamma)} \triangle_۵ \|x_{۳(n-۱)}(t)) - X_{۳(n-۲)}(t)\| + \frac{۲\gamma}{(۲-\gamma)M(\gamma)} \triangle_۶ \int_۰^t \left\|x_{۳(n-۱)}(s) - X_{۳(n-۲)}(s)\right\| \mathrm{d}s,$$

و لذا اثبات کامل است . □

قضیه ۲.۲. برررسی وجود جواب های مدل (۲): مدل شماره (۲) دارای جواب است .

اثبات. حال با فرض این که :

$$x_۱(t) = x_{۱(n)}(t) - \Theta_{۱(n)}(t); \quad x_۲(t) = x_{۱(n)}(t) - \Theta_{۲(n)}(t); \quad x_۱(t) = x_{۳(n)}(t) - \Theta_{۳(n)}(t)$$

$$\begin{aligned}
&\|x_۱(t) - \frac{۲(۱-\gamma)}{(۲-\gamma)M(\gamma)}\tau(t,x_۱) - x_۱(۰) - \frac{۲\gamma}{(۲-\gamma)M(\gamma)}\int_۰^t \tau(s,x_۱(s))ds\| \\
&\quad \leq \|\Theta_{۱(n)}(t)) + \{\frac{۲(۱-\gamma)}{(۲-\gamma)M(\gamma)}\Delta_۱ + \frac{۲\gamma}{(۲-\gamma)M(\gamma)}\Delta_۲ t\}\|\Theta_{۱(n)}(t)\|, \\
&\|x_۲(t) - \frac{۲(۱-\gamma)}{(۲-\gamma)M(\gamma)}v(t,x_۲) - x_۲(۰) - \frac{۲\gamma}{(۲-\gamma)M(\gamma)}\int_۰^t v(s,x_۲(s))ds\| \\
&\quad \leq \|\Theta_{۲(n)}(t)) + \{\frac{۲(۱-\gamma)}{(۲-\gamma)M(\gamma)}\Delta_۳ + \frac{۲\gamma}{(۲-\gamma)M(\gamma)}\Delta_۴ t\}\|\Theta_{۲(n)}(t)\|, \\
&\|x_۳(t) - \frac{۲(۱-\gamma)}{(۲-\gamma)M(\gamma)}\phi(t,x_۳) - x_۳(۰) - \frac{۲\gamma}{(۲-\gamma)M(\gamma)}\int_۰^t \phi(s,x_۳(s))ds\| \\
&\quad \leq \|\Theta_{۳(n)}(t)) + \{\frac{۲(۱-\gamma)}{(۲-\gamma)M(\gamma)}\Delta_۵ + \frac{۲\gamma}{(۲-\gamma)M(\gamma)}\Delta_۶ t\}\|\Theta_{۳(n)}(t)\|,
\end{aligned}$$

اگر از دو طرف معادلات فوق حد بگیریم به گونه ای که : $n \to \infty$

$$\begin{aligned}
x_۱(t) &= \frac{۲(۱-\gamma)}{(۲-\gamma)M(\gamma)}\tau(t,x_۱(t)) + x_۱(۰) + \frac{۲\gamma}{(۲-\gamma)M(\gamma)}\int_۰^t \tau(s,x_۱(s))ds, \\
x_۲(t) &= \frac{۲(۱-\gamma)}{(۲-\gamma)M(\gamma)}v(t,x_۲(t)) + x_۲(۰) + \frac{۲\gamma}{(۲-\gamma)M(\gamma)}\int_۰^t v(s,x_۲(s))ds, \\
x_۳(t) &= \frac{۲(۱-\gamma)}{(۲-\gamma)M(\gamma)}\phi(t,x_۳(t)) + x_۳(۰) + \frac{۲\gamma}{(۲-\gamma)M(\gamma)}\int_۰^t \phi(s,x_۳(s))ds,
\end{aligned}$$

پس ، مدل (۲) دارای جواب است و لذا اثبات کامل است . □

۳ نتیجه‌گیری

در این مقاله با توجه به اعمال عملگرهای مشتقات کسری بر مدل سه بعدی مطالعه بیماری سرطان و بهره گیری از تعاریف موجود در حیطه حسابان کسری ، می توان این گونه استنتاج کرد که مدل سه بعدی بیماری سرطان ، با توجه به شرایط تعریف شده برای کرنل های آن ، قطعا دارای جواب است. همچنین بهره گیری از مشتقات مرتبه کسری در این مدل ، امکان بررسی دقیق تر تغییر هر یک از پارامترهای اساسی را فراهم می آورد.

مراجع

[86] S. P.Banks, Analysis and structure of a new chaotic system, *Int. J. Bifurc . Chaos 9*,(1999) :1571–1583.

[87] K.Itik ; S.P. Banks, Chaos in a three-dimensional cancer model , *Int. J. Bifurc. Chaos 20*, (2010): 71–79 .

[88] J. Francisco Gómez-Aguilar, M.Guadalupe López , Chaos in a Cancer Model via Fractional Derivatives with Exponential Decay and Mittag-Leffler Law , *Entropy. J.*,(2017) :3-19

[89] N.Sheikh; Ali. Saqib, M.Khan, A comparative study of Atangana-Baleanu and Caputo-Fabrizio fractional derivatives to the convective flow of a generalized Casson fluid *Eur. Phys. J.132* (2017) 10-39

تحلیلی بر مدل تصادفی اپیدمی با نرخ شیوع غیرخطی بیماری

میلاد فهیمی [۱]، **کارشناسی ارشد ریاضیات کاربردی دانشگاه سمنان کاظم نوری عضو هیات علمی دانشکده ریاضی و علوم کامپیوتر دانشگاه سمنان لیلا ترک زاده عضو هیات علمی دانشکده ریاضی و علوم کامپیوتر دانشگاه سمنان ایمیل**

چکیده: در این مقاله معادله دیفرانسیل معمولی برای مدل اپیدمی بیماری به صورت (SIQ) مورد بررسی قرار می گیرد و با اعمال نرخ یکسان سازی شیوع و ایجاد شرایط مناسب لم ایتو دوبلین ، یک معادله دیفرانسیل تصادفی تحت عنوان مدل تصادفی $(SIQS)$ اپیدمی بیماری تشکیل می شود و سپس وجود جواب و یکتایی آن برای معادله دیفرانسیل تصادفی ارائه شده مورد بررسی قرار می گیرد. در پایان با استفاده از روش شبیه سازی عددی اویلر ماریاما و برنامه نویسی متلب نمودار نوسانات پارامترها مورد بررسی قرار می گیرد .

واژه‌های کلیدی: مدل اپیدمی، نرخ شیوع غیرخطی ، مدل تصادفی ، فرمول ایتو ، معادله دیفرانسیل تصادفی

۱ مقدمه

مدل سازی ریاضی در علوم مختلف از جمله مهندسی و شیمی و زیست به دفعات استفاده شده است. در سال‌های اخیر استفاده از مدل های ریاضی برای مطالعه مسائل گوناگون علوم زیستی مورد توجه پژوهشگران قرار گرفته است. در مسائل اپیدمی، مدل های ریاضی برای تحلیل رفتار بیماری هایی از قبیل آنفولانزا و ایدز و سرطان استفاده شده‌اند تا جنبه هایی از نحوه انتشار و انقراض بیماری ها را ارئه دهند. یکی از راه های رایج برای مدل سازی مسائل اپیدمی، استفاده از معادلات دیفرانسیل معمولی می باشد. این دسته از مدل ها نخستین بار توسط $Kermak$ و $MCKendrick$ در سال ۱۹۷۲ میلادی معرفی شد که در آن افراد مورد مطالعه در ۳ گروه مستعد پذیرش بیماری و بهبود یافته تقسیم می شوند. [۱] در همین راستا مدل هایی از قبیل $SIRV$ و SIR و $SEIR$ و $SEIRS$ به عنوان مدل هایی در حیطه مدلسازی قطعی مطرح و توسیع یافتند[۷-۵] . از آنجا که در مدل سازی قطعی همواره با محدودیت هایی روبرو می شویم که امکان تطابق مدل ها با مسائل واقعی در آن وجود ندارد؛ لذا باید از مدل هایی مبتنی بر معادلات دیفرانسیل تصادفی استفاده شود تا مدل سازی به صورت واقع گرایانه تری انجام شود.

در این مقاله ابتدا با تعریف نویز سفید یا اختلالات تصادفی و ایجاد زمینه ظهور حرکت براونی ، مدلی تصادفی برای مدل سازی رفتار بیماری ها تعریف می شود.

[۱] میلاد فهیمی: fahimimilad@gmail.com

۱.۱ مبانی نظری پژوهش

اگر فرض کنیم که $\mathbb{R}^d_+ = \{x \in \mathbb{R}^d : x_i > ۰,\ ۱ \le i \le d\}$ ، در این صورت معادله دیفرانسیل تصادفی d-بعدی به صورت زیر قابل تعریف خواهد بود:

$$dX(t) = f(t, X(t))d_t + g(t, \alpha(t))dB_t,\ t \ge t.$$

با مقدار اولیه $X(t.) = x.$ و حرکت براونی m بعدی B_t .

بنابراین عملگر دیفرانسیلی L را به صورت زیر تعریف می کنیم:

$$L = \frac{\partial}{\partial t} + \sum_{i=۱}^{d} f_i(x,t)\frac{\partial}{\partial x_i} + \frac{۱}{۲}\sum_{i,j=۱}^{d}[g^T(x,t)g(x,t)]_{i,j}\frac{\partial^۲}{\partial x_i \partial x_j}$$

حال اگر L بر روی تابع $V \in C$ اعمال شود به گونه ای که: $C^{۲\times۱}(\mathbb{R}^d \times [t.,\infty); \mathbb{R}_+)$

$$LV(x,t) = V_t(x,t) + V_x(x,t)f(x,t) + \frac{۱}{۲}trace[g^T(x,t)V_{xx}(x,t)g(xt)]$$

که در آن:

$$V_t = \frac{\partial V}{\partial t},\quad V_x = (\frac{\partial V}{\partial x_۱}, ..., \frac{\partial V}{\partial x_d})^T,\quad V_{xx} = (\frac{\partial^۲ V}{\partial x_i \partial x_j})$$

بنابراین لم ایتو – دوبلین به صورت زیر قابل ارائه خواهد بود :

$$dV(x,t) = LV(x,t)d_t + V_x(x,t)g(x,t)dB_t$$

۲.۱ ساختار ریاضیاتی مدل تصادفی اپیدمی

حال معادله دیفرانسیل تصادفی $(SIQS)$ با اعمال نرخ واقع گرایانه شیوع اپیدمی (incidenc rate)

$$g(I)S = \frac{SI}{۱ + \alpha I}$$

به فرم زیر قابل ارائه خواهد بود: (مدل شماره (۱))

$$\begin{cases} dS(t) = \left[A - \dfrac{SI}{۱+\alpha I} - dS + \gamma I + PQ\right]d_t + \sigma_۱(s - s^*)dB_۱(t) \\ dI(t) = \left[\dfrac{SI}{۱+\alpha I} - (\gamma + \delta + d)I\right]d_t + \sigma_۲(I - I^*)dB_۲(t) \\ dQ(t) = \left[\delta I - (P+d)Q\right]d_t + \sigma_۳(Q - Q^*)dB_۳(t) \end{cases}$$

که در آن: A نرخ همبستگی شاخص S به نرخ تولد و مهاجرت و d نرخ مرگ طبیعی و P و γ نشان دهنده متوسط تعداد افراد بهبود یافته هستند و σ نشان دهنده متوسط (نرخ) اشخاص مبتلا به بیماری عفونی است و همچنین B_i به ازای $(i = ۱,۲,۳)$ حرکت های براونی استاندارد مستقل هستند که در آن σ_i ها به ازای $(i = ۱,۲,۳)$ نویزها (white noise) هستند حال اگر در مدل بیماری فرض کنیم که $\sigma_i = ۰\ \ (i = ۱,۲,۳)$ لذا مدل فوق به صورت زیر خواهد بود:

$$\begin{cases} S(t) = A - \dfrac{SI}{۱+\alpha I} - dS + \gamma I + PQ \\ I(t) = \dfrac{SI}{۱+\alpha I} - (\gamma + \delta + d)I \\ Q(t) = \delta I - (P+d)Q \end{cases}$$

حال با توجه به روش استخراج عدد مولد عفونت [۳،۲] ، در این مدل تصادفی عدد مولد عفونت (Repraduction number) به صورت $R_{\cdot} = \frac{A}{d(\gamma+\delta+d)}$ خواهد بود.

تعریف ۱.۱. اگر $X(t) = (S(t), I(t), Q(t))$ جواب معادلات مدل (۱) باشد، گوییم این جواب برای مدل به طور تصادفی کراندار است هرگاه به ازای هر $\varepsilon \in (۰, ۱)$ ، مقدار ثابت $\delta = \delta(\varepsilon) > ۰$ وجود داشته باشند به گونه ای که به ازای هر مقدار اولیه $(S(۰), I(۰), Q(۰))$ ، جواب $X(t)$ دارای خاصیت زیر باشد: $\limsup P\{|X(t)| > \delta\} < \varepsilon$.

۲ نتایج اصلی

در این بخش ، وجود ، یکتایی و کرانداری جواب معادله دیفرانسیل ذکر شده در مدل و نیز شبیه سازی عددی مدل مورد بررسی قرار می گیرد .

۱.۲ بررسی وجود ، یکتایی و کرانداری جواب

قضیه ۱.۲. معادله (۱) دارای جواب منحصر به فرد $(S(t), I(t), Q(t))$ می باشد و این جواب با احتمال یک در $\mathbb{R}^{۳}_{+}$ بسته خواهد بود.

اثبات. تابع V را به صورت $V : \mathbb{R}^{۳}_{+} \to \mathbb{R}_{+}$ تعریف می کنیم به گونه ای که:

$$V(S,I,Q) = \ln(S_{(t)}, I_{(t)}, Q_{(t)}) = \ln(S_{(t)}) + \ln(I_{(t)}) + \ln(Q_{(t)})$$

حال با اعمال لم ایتو–دوبلین به صورت زیر داریم:

$$\begin{aligned}\Longrightarrow dV(S,I,Q) &= V_t d_t + V_s d_s + V_I d_I + V_Q d_Q + \\ &+ \tfrac{۱}{۲} V_{SS} d_S d_S + \tfrac{۱}{۲} V_{II} d_I d_I + \tfrac{۱}{۲} V_{QQ} d_Q d_Q + \\ &+ \tfrac{۱}{۲} V_{SQ} d_S d_Q + \tfrac{۱}{۲} V_{SI} d_S d_I + \tfrac{۱}{۲} V_{IQ} d_I d_Q\end{aligned}$$

لذا با اعمال آن داریم:

$$\begin{aligned}&\Longrightarrow V_S d_S + V_I d_I + V_Q d_Q = \frac{S'(t)}{S(t)} + \frac{I'(t)}{I(t)} + \frac{Q'(t)}{Q(t)} = \\ &= \left[\frac{A}{S} - \frac{I}{۱+\alpha I} - \frac{dS}{S} + \frac{\gamma I}{S} + \frac{PQ}{S}\right] d_t - \frac{\sigma_۱(S-S^*)}{S} dB_۱(t) \\ &+ \left[\frac{S}{۱+\alpha I} - (\gamma+\delta+d)\right] d_t + \frac{\sigma_۲(I-I^*)}{I} dB_۲(t) \\ &+ \left[\frac{\delta I}{Q} - (P+d)\right] d_t + \frac{\sigma_۳(Q-Q^*)}{Q} dB_۳(t) \\ &= \left[\frac{A}{S} - \frac{I}{۱+\alpha I} - \frac{dS}{S} + \frac{\gamma I}{S} + \frac{PQ}{S}\frac{S}{۱+\alpha I} - (\gamma+\delta+d) + \frac{\delta I}{Q}\right. \\ &\left. - (P+d)\right] d_t - \sigma_۱(۱ - \frac{S^*}{S}) dB_۱(t) + \sigma_۲(۱ - \frac{Q^*}{Q}) dB_۲(t) \\ &+ \sigma_۳(۱ - \frac{Q^*}{S}) dB_۳(t)\end{aligned}$$

همچنین داریم :

$$\begin{aligned}
&\Longrightarrow \frac{۱}{۲}V_{SS}d_Sd_S + \frac{۱}{۲}V_{II}d_Id_I + \frac{۱}{۲}V_{QQ}d_Qd_Q \\
&= \frac{۱}{۲}\Big(\sigma_۱(S-S^*) + \sigma_۲(I-I^*) + \sigma_۳(Q-Q^*)\Big)d_t \\
&\Longrightarrow dV(S,I,Q) \leq C
\end{aligned}$$

که در آن C یک عدد مثبت ثابت می باشد. حال اگر از دو طرف معادله فوق انتگرال نسبت به t بگیریم، لذا با توجه به [۷] نتیجه زیر حاصل می شود:

$$+\infty > V(S(۰),I(۰),Q(۰)) + KT \geq +\infty$$

و لذا با وجود تناقض فوق، حکم ثابت است. □

قضیه ۲.۲ (بررسی کرانداری). به ازای هر مقدار اولیه $(S(۰),I(۰),Q(۰)) \in \mathbb{R}^۳_+$ ، جواب معادله های مدل (۱) به طور تصادفی کراندار است.

اثبات. با فرض این که $V(S,I,Q) = e^t(S^\theta, I^\theta, Q^\theta)$ باشد و $\theta \in (۰,۱)$ ، در این صورت با اعمال شرط ایتو-دوبلین، فرم دیفرانسیلی زیر حاصل می گردد:

$$\begin{aligned}
dV(S,I,Q) &= e^t\Big\{(S^\theta + I^\theta + Q^\theta) + HS^{\theta-۱}(A - \frac{\beta SI}{۱+\alpha I} - dS \\
&+ \gamma I + PQ) + HI^{\theta-۱}[\frac{\beta SI}{۱+\alpha I} - (\gamma+\delta+d)I)] \\
&+ \theta Q^{\theta-۱}[\delta I - (P+d)Q] + \frac{\theta(\theta-۱)}{۲}[\sigma_۱^۲S^\theta(۱-\frac{S^*}{S})^۲ \\
&+ \sigma_۲^۲I^\theta(۱-\frac{I^*}{I})^۲ + \sigma_۳^۲Q^\theta(۱-\frac{Q^*}{Q})^۲]\Big\}dt \\
&+ e^t\theta[\sigma_۱S^\theta(۱-\frac{S^*}{S})dB_۱(t) + \sigma_۲I^\theta(۱-\frac{I^*}{I})dB_۲(t) + \sigma_۳Q^\theta(۱-\frac{Q^*}{Q})dB_۳(t)] \\
&\leq e^t\Big\{۳ + A - d(S+I+Q) + \frac{\theta(\theta-۱)}{۲}[\sigma_۱^۲S^\theta(۱-\frac{S^*}{S})^۲ \\
&+ \sigma_۲^۲I^\theta(۱-\frac{I^*}{I})^۲ + \sigma_۳^۲Q^\theta(۱-\frac{Q^*}{Q})^۲]\Big\}dt \\
&+ e^t\theta[\sigma_۱S^\theta(۱-\frac{S^*}{S})dB_۱(t) + \sigma_۲I^\theta(۱-\frac{I^*}{I})dB_۲(t) + \sigma_۳Q^\theta(۱-\frac{Q^*}{Q})dB_۳(t)] \\
&\leq (۳+A)e^tdt + e^t\theta[\sigma_۱S^\theta(۱-\frac{S^*}{S})dB_۱(t) + \sigma_۲I^\theta(۱-\frac{I^*}{I})dB_۲(t) \\
&+ \sigma_۳Q^\theta(۱-\frac{Q^*}{Q})dB_۳(t)].
\end{aligned}$$

و لذا با انتگرال گرفتن از دو طرف عبارت فوق به نامساوی زیر دست می یابیم:

$$\begin{aligned}
&E(V(S,I,Q)) \leq V(S(۰),I(۰),Q(۰)) + (۳+A)(e^t - ۱) \\
&\Longrightarrow e^{-t}E(V(S,I,Q)) \leq e^{-t}V(S(۰),I(۰),Q(۰)) + ۳ + A \qquad (۱.۲)
\end{aligned}$$

لذا با تعریف نرم به صورت $|X|$ داریم:

$$|X(t)| = (S^{۲}(t) + I^{۲}(t) + Q^{۲}(t))^{\frac{۱}{۲}}$$

و لذا داریم:

$$|X(t)|^{\theta} = (S^{۲}(t) + I^{۲}(t) + Q^{۲}(t))^{\frac{\theta}{۲}} \leq ۳^{\frac{\theta}{۲}} \max\{S^{\theta}, I^{\theta}, Q^{\theta}\}$$
$$\leq ۳^{\frac{\theta}{۲}}(S^{\theta}, I^{\theta}, Q^{\theta})$$

بنابراین با توجه به رابطه (۲):

$$\limsup E|X(t)|^{\theta} \leq ۳^{\frac{\theta}{۲}}(۳ + A) < +\infty$$

حال با در نظر گرفتن $\theta = \frac{۱}{۲}$ ، ثابت $\delta_{۱} > ۰$ وجود دارد به گونه ای که:

$$\limsup E|X(t)|^{\frac{۱}{۲}} \leq \delta_{۱}$$

بنابراین به ازای هر $\varepsilon > ۰$ و قرار دادن $\delta = \frac{\delta_{۱}^{۲}}{\varepsilon^{۲}}$ در اعمال نامساوی چبیشف، نتیجه می گیریم که:

$$P\{|X(t)| > \delta\} \leq \frac{E|X(t)|^{\frac{۱}{۲}}}{\delta^{\frac{۱}{۲}}} = \frac{\delta_{۱}}{\delta^{\frac{۱}{۲}}} = \varepsilon$$

و لذا با توجه به تعریف (۱): جواب $X(t)$ برای مدل (۱) به طور تصادفی کراندار خواهد بود. □

۲.۲ شبیه سازی عددی

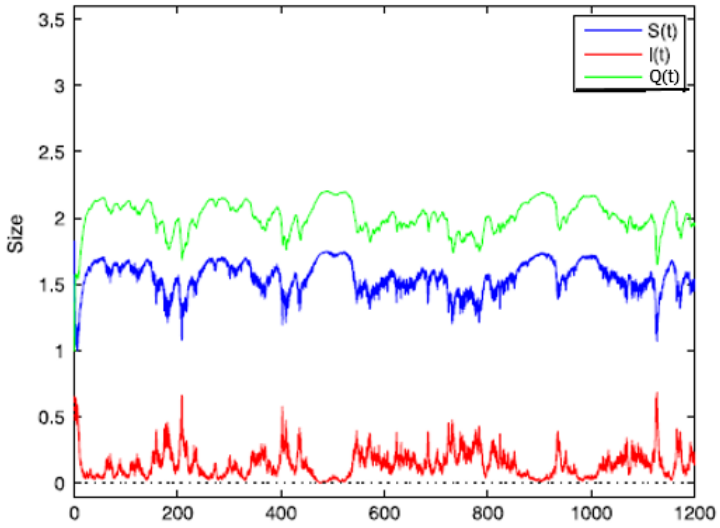
با استفاده از طرح گسسته سازی اویلر ماریاما [۴] داریم:

$$S_{k+۱} = S_k + [A - \frac{S_k I_k}{۱ + \alpha I_k} - dS_k + \gamma I_k + PQ_k]\Delta t + \sigma_{۱}(S_k - S^*)$$
$$I_{k+۱} = I_k + [\frac{S_k I_k}{۱ + \alpha I_k} - (\gamma + \delta + d)I_k]\Delta t + \sigma_{۲}(I_k - I^*)\Delta wt$$
$$Q_{k+۱} = Q_k + [\delta I_k - (P + d)Q_k]\Delta t + \sigma_{۳}(Q_k - Q^*)\Delta wt$$

که در آن $\Delta w_t \sim \sqrt{\Delta t}Z$ و $Z \sim N(۰, ۱)$ به ازای $t = ۱, ۲, ..., n$ به عبارت دیگر Z متغیر تصادفی گاوسی می باشد.
حال با قراردادن مقادیر اولیه به صورت $(S(۰), I(۰), Q(۰)) = (۱, ۱\square۵, ۱)$ و در نظر گرفتن:

$$A = ۱, \quad \alpha = ۰\square۷, \quad d = ۰\square۵, \quad \delta = ۰\square۵, \quad P = ۰\square۳, \quad \gamma = ۰\square۰۱$$

لذا با توجه به شکل (۱) می توان نتیجه گرفت که به ازای $\sigma_{۱} = \sigma_{۲} = \sigma_{۳} = ۰$ ، مدل (۱) به طور تصادفی پایدار بوده و این امر با تغییر مقادیر پارامترها تغییری نخواهد کرد.
شکل(۱)- نمودار نوسانات پارامترهای اساسی بر حسب زمان



۳ نتیجه‌گیری

در این مقاله با توجه به مدل تعیینی–قطعی اپیدمی ، مدل تصادفی اپیدمی تشکیل شده است. در فرآیند تشکیل این معادله از راهبرد اعمال نرخ شیوع غیرخطی اپیدمی برای واقعی تر شدن پارامترها و نیز با در نظر گرفتن نویزها(نوسانات) و حرکت براونی برای دستیابی به معادلات دیفرانسیل تصادفی استفاده شده است. با توجه به قضایای مطرح شده ، وجود ، یکتایی و کرانداری تصادفی جواب معادلات موجود در مدل تصادفی استنتاج می شود . همچنین با شبیه سازی عددی مبتنی بر طرح گسسته سازی اویلر – ماریاما در نرم افزار متلب و تحلیلی نتایج عددی حاصل از ترسیم نمودار نوسانات پارامتر های اساسی مدل ، می توان پایداری مدل به ازای تغییرات پارامترها را نتیجه گرفت.

مراجع

[90] V. Capasso, A generalisation of the Kermack–McKendrick deterministic epidemic model , *Math. Biosci.42*, (1978), 43-61 .

[91] O. Diekmann , J. A. P. Heesterbeek , J. A. J. Metz , On the definition and the computation of the basic reproduction ratio R0 in models for infectious diseases in heterogeneous populations , *Journal of Mathematical Biology* , Volume 28 ,(1990) : 365–382 .

[92] J. Stoer and R. Bulirsch, *Introduction to Numerical Analysis*, 2nd ed., Springer-Verlag, Berlin: 1993.

[93] D.J. Higham , An algorithmic introduction to numerical simulation of stochastic differential equations, *SIAM Rev. 43* , (2001) : 525–546 .

[94] Q.liu , Q.chen , Analysis of the deterministic and stochastic SIRS epidemic models with nonlinear incidence , *physica A* (2015) : 1-24

[95] Q.liu , Q.chen , Dynamics of a stochastic SIR epidemic model with saturated incidence , *App.math.comput 282*, (2016) : 155-166

[96] F.Wei , F.Fangxiang chen , Stochastic permanence of an SIQS epidemic model with saturated incidence and independent random perturbations , *physica , A453*, (2016):99 - 107 .

اولین همایش ملی ریاضیات زیستی
۲۱ و ۲۲ اسفندماه ۱۳۹۷
همایش ملی ریاضیات زیستی
دانشگاه نیشابور

نتایجی از C^3 -کدهای خودمکمل ماکسیمال

فریبا فیاضی [۱]، **دانشکده علوم ریاضی، دانشگاه قم، قم، ایران، سید علیرضا اشرفی دانشکده علوم ریاضی، دانشگاه کاشان، کاشان، ایران، احمد غلامی دانشکده علوم ریاضی، دانشگاه قم، قم، ایران**

چکیده: در سال ۲۰۱۴، النا فیمل و همکارانش، ۲۱۶ C^3 – کد خود مکمل ماکسیمال را به ۲۷ کلاس هم ارزی شامل ۸ کد افراز کردند و زیرگروهی از گروه متقارن روی پایه های نوکلئوتید (آدنین A، تیمین T، سیتوزین C و گوانین G) ساختند. در این مقاله، ما نمایش جایگشتی و جدول سرشت را برای گروه ساخته شده متناظر با C^3 – کدهای خود مکمل ماکسیمال بدست می آوریم و سپس با استفاده از آن نوع ارتعاشات IRو رامان را بین کدون های مورد نظر مشخص می کنیم.

واژه‌های کلیدی: C^3 – کد خود مکمل ماکسیمال ، نمایش جایگشتی، جدول سرشت، ارتعاشات، کدهای ژنتیکی

۱ مقدمه

اسیدهای نوکلئوتید بسپارهایی (پلیمرهایی) با زنجیر طولانی و وزن مولکولی بالا متشکل از نوکلئوتید ها هستند. هرنوکلئوتید از یک مولکول قند ۵ کربنی و یک مولکول باز نیتروژن دار، تشکیل شده است. در هسته هر سلول مولکول هایی قرار دارند که اساسی ترین اطلاعات حیات را در خود ذخیره کرده اند. این مولکول ها، دئوکسی ریبونوکلئوتید اسید (DNA) نام دارند. DNAاز چهار نوع مولکول ساخته شده است، که به آنها " نوکلئوتید " می گوییم. این چهار نوکلئوتید عبارتند از: آدنین (A) ،گوانین (G) ، سیتوزین (C) و تیمین (T) . این مولکول ها خود به دو زیر گروه تقسیم می شوند: آدنین و گوانین در گروه پورین قرار می گیرند و سیتوزین و تیمین در گروه پیرامیدین جای داده می شوند. نوکلئوتیدها از طریق پیوند هیدروژنی به هم متصل می شوند. سه کدهای ژنتیکی با سه حرف از پایه های اسید نوکلئوتید ساخته می شوند. فرض کنیم $\beta = \{T(U), C, A, G\}$ گروه متقارن S_β روی β به صورت زیر تعریف می شود:

$$S_\beta = \{\pi : \beta \to \beta | \ \pi \text{دو سویی است}\}$$

(S_β, o) یک گروه است که در آن o عمل ترکیب روی توابع است. حال به معرفی نگاشت π می پردازیم. منظور از π_{AGCT} نگاشت زیر است:

$$\pi : (A, T, C, G) \to (G, A, T, C)$$

[۱] فریبا فیاضی: fayazifariba@yahoo.com

$$\pi(A) = G, \pi(T) = A, \pi(C) = T, \pi(G) = C.$$

نگاشت دو سویی $\pi : \beta \to \beta$ برای مجموعه کدون ها (یک دنباله از سه نوکلئوتید در دنباله DNA) نگاشت زیر را القا می کند که آن را نیز با π نشان می دهیم:

$$\pi : \beta^۳ \to \beta^۳$$

در (۱۰۱) چهار نگاشت از انتقال های اسید نوکلئوتیدها به صورت زیر تعیین و نام گذاری شده اند:

$$i = I(or\ id) : (A, T, C, G) \to (A, T, C, G)$$

$$c = SW : (A, T, C, G) \to (T, A, G, C)$$

$$p = YR : (A, T, C, G) \to (G, C, T, A)$$

$$r = KM : (A, T, C, G) \to (C, G, A, T)$$

که در آن ها اولی نگاشت همانی، دومی نگاشت مکمل، سومی نگاشت پیرامیدین- پورین و چهارمی نگاشت آمینو- کیتو است. برای هر زیرگروه X از $\beta^۳$، ۶۴ کدون در $\beta^۳$ وجود دارد.

حال می خواهیم گروه متقارن $(S_۳, o)$ را بررسی کنیم که نمایش آن به صورت زیر است:

$$S_۳ = \{\alpha : \{۱, ۲, ۳\} \to \{۱, ۲, ۳\} | \ \alpha \text{ است سویی دو}\}$$

به وضوح هر α یک نگاشت روی مجموعه کدون ها $(\beta^۳)$ القا می کند. به عنوان مثال جایگشت (۱۳۲) کدون $(b_۱, b_۲, b_۳)$ را به $(b_۳, b_۱, b_۲)$ می برد. بنابراین برای هر مجموعه دلخواه $X \subseteq \beta^۳$ ، $\pi(X)$ یک تبدیل روی نوکلئوتیدها است در حالی که $\alpha(X)$ یک جایگشت از موقعیت کدون است. همانطور که در زیر نشان خواهیم داد، این تفاوت از دیدگاه بیولوژیکی نقش مهمی دارد. در ادامه روی گروه $(A_۳, o)$ به عنوان زیرگروهی از $(S_۳, o)$ تمرکز می کنیم که یکی از نمایش های آن به صورت زیر است:

$$A_۳ = \{\alpha_. = (۱)(۲)(۳), \alpha_۱ = (۱۲۳), \alpha_۲ = (۱۳۲)\}.$$

همانطور که در بالا ذکر کردیم گروه $S_۳$ روی مجموعه کدون ها عمل می کند و لذا می توان گفت دو کدون $x_۱, x_۲ \in \beta^۳$ به طور دوری هم ارزند هرگاه نگاشت $\alpha \in A_۳$ وجود داشته باشد به طوری که:

$$\alpha(x_۱) = x_۲$$

و در این حالت می نویسیم $x_۱ \sim x_۲$. برای مثال، برای کدون دلخواه ATG داریم:

$\alpha_۱(ATG) = TGA$ و $\alpha_۲(ATG) = GAT$. لذا: $ATG \sim GAT \sim TGA$. در این مقاله از هم ارزی کلاس های تزویجی استفاده خواهیم کرد. واضح است که کدون های AAA, TTT, CCC, GGG به یک کلاس هم ارزی تعلق دارند و بقیه ۲۰ کلاس های هم ارزی نیز هر کدام شامل سه عضو هستند. جایگشت (۲)(۱۳) از $S_۳$ را در نظر بگیرید که عضوی از $A_۳$ نیست. وارون این جایگشت خودش است. این وارون را با $\leftarrow$ نشان می دهیم و جایگشت وارون ساز می نامیم. برای کدون دلخواه $x = (b_۱, b_۲, b_۳) \in \beta^۳$ داریم: $x^{\leftarrow} = (b_۳, b_۲, b_۱)$. جایگشت های وارون ساز نقش مهمی در بیولوژی دارند زیرا برخی از پروتئین ها (که معمولا در میتوکندری هستند)، می توانند در جهت وارون کد گذاری شوند.

در زیست شناسی مولکولی دنباله ها را در سه قالب متفاوت در جهت ۳ $\to$ ۵ می خوانند ، هر کدام از یک نوکلئوتید متفاوت در یک سه گانه آغاز می شود. دنباله $ATGACGGTACGATTG...$ را در نظر بگیرید:

قالب ۰: خواندن دنباله از اولین کدون موجود$\{ATG, ACG, ...\}$.

قالب ۱: در قالب ۰ یک نوکلئوتید را رد می کنیم و بقیه کدون ها را می خوانیم $\{TGA, CCG, ...\}$.

قالب ۲: در قالب ۰ دو نوکلئوتید را رد می کنیم و بقیه کدون ها را می خوانیم $\{GAC, GGT, ...\}$.

توزیع ۶۴ نوکلئوتید در سه قالب از ژن ها یکسان نیست و لذا می توان سه مجموعه از نوکلئوتیدها را به صورت $X_۲, X_۱, X_۰$ مشخص نمود (۹۷) .

تعریف ۱.۱. نگاشت مکمل $\pi : \beta^۳ \to \beta^۳$ با ضابطه ی $\pi(uv) = \pi(u)\pi(v)$ تعریف می شود که در آن $u \in \beta^n$ و $v \in \beta^{n-۱}$ تعریف می شود.

تعریف ۲.۱. یک کلمه از مجموعه $S \subseteq S_\beta^n$ یک کد نامیده می شود هرگاه برای هر $x_۱, x_۲, ..., x_n$ و $y_۱, y_۲, ..., y_m$ در S و به ازای $n, m \geq ۱$ داشته باشیم:

$$x_i = y_j, n = m, \forall\, i = ۱, ..., n,\ j = ۱, ..., m.$$

توجه کنید که همواره کدها به صورت خطی خوانده می شوند.

جایگشت تری نوکلئوتید $W_۰ = l_۰l_۱l_۲$ که در آن $l_i \in \beta$ به صورت زیر است:

$$P(W_۰) = W_۱ = l_۲l_۱l_۰, \quad P(P(w_۰)) = P(w_۱) = W_۲ = l_۱l_۰l_۲.$$

مجموعه های $X_۰$ ، $X_۱ = P(X_۰)$ و $X_۲ = P(X_۱)$ شامل ۲۰ کد دایره ای تری نوکلئوتید ماکسیمال هستند. همچنین:

$$C(X_۰) = X_۰, C(X_۱) = X_۲, C(X_۲) = X_۱.$$

تعریف ۳.۱. فرض کنیم $X \subseteq \beta^۳$. مجموعه کدون های X را یک کد مستدیر تری نوکلئوتید گوییم اگر هر کلمه از حروف β که به صورت مستدیر نوشته شده است حداکثر یک تجزیه از کلمات Xداشته باشد. کد مستدیر تری نوکلئوتید ماکسیمال است هرگاه برای هر $x \in S_\beta$ و $x \notin X$ ، $X \cup \{x\}$ کد مستدیر تری نوکلئوتید نباشد. همچنین برای کلمات ۳ یا ۴ حرفی می توان گفت کد مستدیر تری نوکلئوتید ماکسیمال است هرگاه دقیقا شامل ۲۰ کدون باشد یعنی $|X| = ۲۰$.

لازم به ذکر است که، اگر $X_۰$ کد مستدیر باشد آنگاه لزوما $X_۱ = P(X_۰)$و $X_۲ = P(X_۱)$ کدهای مستدیر نیستند. مثال بعدی این موضوع را ثابت می کنند.

فرض کنیم $Y_۰ = \{GGC, CCG\}$ کد مستدیر باشد.

$Y_۱ = P(Y_۰) = \{GCG, CGC\}$ کد مستدیر نیست زیرا دارای دو تجزیه زیر است:

$$w = GCG, CGC,\ w' = CGC, GCG.$$

تعریف ۴.۱. $X \subseteq \beta^۳$ یک $C^۳$-کد از X است، هرگاه $X_۱, X_۲$ کدهای مستدیر باشند به طوری که داشته باشیم:

$$\alpha_۱(X) = X_۱,\ \alpha_۲(X) = X_۲ \quad \forall \alpha \in S_۳.$$

بنابر تعریف هر $C^۳$- کد یک کد مستدیر است.

اگر $X_۰$ $C^۳$- کد باشد، آنگاه $X_۱ = P(X_۰)$و $X_۲ = P(X_۱)$ کدهای مستدیر هستند. مثال بعدی این موضوع را ثابت می کند.

فرض کنیم $Y_. = \{CGG, GCC\}$ کد مستدیر باشد اما C^3- کد نیست. $P(Y_.) = Y_1 = \{GGC, CCG\}$ کد مستدیر است اما $P(P(Y_.)) = Y_2 = \{GCG, CGC\}$ کد مستدیر نیست زیرا دارای دو تجزیه زیر است:

$$w = GCG, CGC,\ w' = CGC, GCG.$$

برای هر کدون $x = N_1N_2N_3$، آنتی کدون x را با $\overleftarrow{c(x)} = N_1'N_2'N_3'$ نشان می دهیم که در آن N_i' ها، مکمل نوکلئوتیدهای N_i هستند یعنی $N_i' = c(N_i)$.

تعریف ۵.۱. $X \subseteq \beta^3$ یک خود مکمل است هرگاه برای هر $x \in X$ آنتی کدون $\overleftarrow{c(x)}$ وجود داشته باشد به طور ی که:

$$x \in X \iff \overleftarrow{c(x)} \in X.$$

قضیه ۶.۱. (۱) همانی و جایگشت های وارون ساز در موقعیت پایه های یک کدون تنها جایگشت هایی هستند که مستدیر بودن را در یک کدون مستدیر $X \subseteq \beta^3$حفظ می کنند.

(۲)فرض کنید $X \subseteq \beta^3$ یک کد دایره ای تری نوکلئوتید باشد. برای هر $\pi \in S_\beta$، $\pi(X)$ نیز یک کد دایره ای تری نوکلئوتید است. همچنین اگر X یک C^3- کد باشد $\pi(X)$ نیز C^3 -کد است.

قضیه ۷.۱. فرض کنید $\pi \in S_\beta$. آنگاه $co\pi = \pi oc$ اگر و فقط اگر $\pi(X)$ یک کد خود مکمل مستدیر تری نوکلئوتید باشد که در آن $X \subseteq \beta^3$ هر کد خود مکمل مستدیر تری نوکلئوتید دلخواه است.

بنابر قضایای فوق می توان گفت ۲۱۶ C^3 - کد خودمکمل ماکسیمال به ۲۷ کلاس هم ارزی که در هر کلاس ۸ کد وجود دارد، تقسیم می شود. لذا می توان زیرگروه L از S_β را بصورت زیر تشکیل داد.

$$L = \{id, c, p, r, \pi_{AT}, \pi_{CG}, \pi_{ACTG}, \pi_{AGTC}\}$$

گروهL زیرگروه نرمال S_β نیست اما زیرگروهی یکریخت با گروه دو وجهی از مرتبه ۸ است. این زیرگروه را با D_8 نشان می دهیم. در ادامه می خواهیم جدول سرشت متناظر گروه Lمتناظر با C^3 -کد خودمکمل ماکسیمال را بدست آوریم و با کمک آن نوع ارتعاش بین کدون ها را مشخص کنیم.

۲ نتایج اصلی

در ابتدا مفاهیمی از نظریه سرشت را بیان می کنیم (۱۰۲) . یک همریختی از یک گروه به توی گروهی از ماتریس ها و جایگشت ها را به ترتیب یک نمایش ماتریسی یا جایگشتی آن گروه می نامند. در حالت اول تابعی از گروه به توی میدان را که هر عضو را به اثر ماتریس متناظر در نمایش تصویر می کند را سرشت متناظر با ان نمایش می نامند. سرشت χ از یک نمایش جایگشتی ρ تعداد اعضایی است که تحت عمل روی گروه ثابت می مانند. $\chi(1)$ را درجه سرشت گوییم و سرشت ها از درجه یک را سرشت خطی نامیم. بنابر نتیجه ای معروف سرشت ها روی کلاس های تزویج ثابت هستند و اگر g و h مزدوج باشند آنگاه $\chi_\rho(h) = \chi_\rho(g)$. در حالت کلی تابع $\phi : G \to \mathbb{C}$ را یک تابع کلاسی گویند هرگاه ϕ روی کلاس های تزویج G ثابت باشد. سرشت متناظر با یک نمایش تحویل ناپذیر را یک سرشت تحویل ناپذیر می نامند. بنابر نتیجه ای معروف در نظریه نمایش سرشت χ تحویل ناپذیر است اگر و تنها اگر $\frac{1}{|G|}\sum_{g\in G}\chi(g)\chi(g^{-1}) = 1$. مجموعه همه سرشت های تحویل ناپذیر گروه را با $Irr(G)$ نشان می دهند. بنابر نتیجه ای معروف، هر سرشت یک گروه را می توان به صورت ترکیبی خطی از سرشت های تحویل ناپذیر نوشت که ضرائب ان اعدادی صحیح و نامنفی هستند. به علاوه $|G| = \sum_{i=1}^{k}\chi_i(1)^2$. جدول سرشت یک گروه، یک جدول r در r است که در آن r تعداد کلاس های تزویج گروه مورد نظر و سطرهای جدول در حقیقت مقادیر سرشت های تحویل ناپذیرند و در بالای هر کدام از ستون های آن نماینده یکی از کلاس های تزویج گروه قرار دارد. قضایای زیر را نیز از مرجع (۱۰۲) داریم:

قضیه ۱.۲. (روابط تعامد کلی [قضیه ۲.۱۳، ص. ۱۹]): برای هر عضو $h \in G$ داریم:

$$\frac{1}{|G|}\sum_{g\in G}\chi_i(gh)\chi_j(g^{-1}) = \delta_{ij}\frac{\chi_i(h)}{\chi_i(1)}.$$

لم ۲.۲. ([لم ۲.۱۵، ص. ۲۰]):فرض کنید که G یک گروه متناهی باشد و χ_i سرشت از G و ρ نمایشی که χ_iاز آن حاصل شده است. اگر $g \in G$و $|G| = n$ آنگاه:

۱. $\rho(g)$با یک ماتریس قطری به شکل $diag(v_1, ..., v_f)$ متشابه است.

۲. $v_i^n = 1$.

۳. $\chi(g) = \sum v_i, |\chi(g) \leq 1$.

۴. $\chi(g^{-1}) = \bar{\chi(g)}$.

قضیه ۳.۲. (رابطه تعامد دوم [قضیه ۲.۱۸، ص. ۲۱]): فرض کنید g و h دو عضو دلخواه از گروه G باشند. در این صورت داریم:

حال فرض کنیم گروه L متناظر با C^3 – کد خود مکمل ماکسیمال باشد. یک نمایش جایگشتی از گروه L روی مجموعه متقارن S_β^3 ، همریختی $\rho : L \to S_\beta^3$ است. گروه L دارای پنج کلاس تزویجی زیر است:

$$\{e\}, \{\pi_{AGTC}, \pi_{ACTG}\}, \{\pi_{AT}, \pi_{CG}\}, \{c\}, \{r, p\}$$

لذا دارای پنج سرشت تحویل ناپذیر خواهد بود و می توان جدول سرشت گروه مورد نظر را با استفاده از مفاهیم بیان شده، بدست آورد.

۳ نتیجه‌گیری

در این مقاله جدول سرشت گروه L متناظر با C^3 – کد خود مکمل ماکسیمال را بدست آوردیم و سپس با کمک آن نوع ارتعاشات بین کدون های مورد نظر را مشخص نمودیم.

مراجع

[97] Arquès DG, Michel CJ, A complementary circular code in the protein coding genes, *J Theor Biol*.(1996), no. 182, 45–58.

[98] Benard E, Michel CJ, Transition and transversion on the common trinucleotide circular code, *Comput Biol J ID*. (2013), no. 795418:10.

[99] Christian J. Michel, The maximal C^3 self-complementary trinucleotide circular code X in genes of bacteria, eukaryotes, plasmids and viruses, *Journal of Theoretical Biology*. (2015), no. 380: 156-177.

[100] Fimmel E, et al, Circular codes, symmetries and transformations, *J. Math. Biol.DOI 10.1007/s00285-014-0806-7* (2014).

[101] Gonzalez D. L., Giannerini S., Rosa R. Strong short-range correlations and dichotomic codon classes in coding DNA sequences, *Phys. Rev. E* (2008) 78(5, ID 051918).

[102] Isaacs, Irving Martin, Character theory of finite groups, *Academic press, INC* (1976).

مدل‌های ریاضی ایمن درمانی سرطان

فائزه قربانی [۱]، **سید محسن صالح گروه ریاضی، دانشگاه نیشابور، نیشابور**

چکیده: طبق تحقیقات اخیر، امیدوارکننده‌ترین مسیرها در عرصه تحقیقات سرطانی پیش‌گیری از ابتلا به سرطان و ایمن درمانی است. ایمن درمانی در واقع نوع خاصی از شیمی درمانی است. با این تفاوت که در روش شیمی درمانی سیتوتاکسیک، سلول های تومور مستقیماً مورد هدف قرار می‌گیرند؛ در حالی که ایمن درمانی سلول‌های تومور را به طور غیرمستقیم و با افزایش کارایی سلول‌های ایمنی بدن از بین می‌برد. هدف از این مقاله، طراحی الگوهای ایمن درمانی و شیمی درمانی سرطان است.

واژه‌های کلیدی: مدل‌سازی ریاضی، سرطان، شیمی درمانی، ایمن درمانی

۱ مقدمه

در این مقاله، برخی از انواع تلاش‌های پژوهشگران در طراحی مدل‌های ریاضی و شبیه‌سازی شیمی درمانی، ایمن درمانی و ترکیبی از آن دو برای سرطان تشریح شده است.

سرطان مهم‌ترین عامل تلفات در جهان است. ارتقای وضعیت واکنش ایمنی با استفاده از تزریق و ارائه مؤثر درمان‌های ایمنی، مسیری امیدوار کننده برای افزایش شانس بقا و همچنین بهبود کیفیت معیشتی بیماران سرطانی است. با توجه به واکنش‌های متفاوت بیماران نسبت به درمان‌های مختلف ، ایجاد ابزارهای پیش‌بینی کننده برای تعیین بهترین روش درمانی برای هرکدام از بیماران حائز اهمیت است. الگوهای ریاضی می‌توانند به کشف تعاملات بین درمان‌ها و درک واکنش‌ها به دوزدهی کمک کند.

هدف نهایی از این مطالعات، ارائه چارچوبی برای عصر جدید در اندازه‌گیری دوزها و همچنین بهینه کردن تعداد دفعات و میزان استفاده از روش‌های درمانی از جمله شیمی درمانی یا پرتو درمانی است.

۲ طراحی الگوهای شیمی درمانی سیتوتاکسیک

در اوایل دهه نود میلادی، مدل ریاضی واکنش لَمفوسیت T سیتوتاکسیک نسبت به رشد تحریک سلول‌های ایمنی در برابر تومور به عنوان مدلی دو جمعیتی مطرح شد. از نتایج بسیار مهم این مدل، ثبت پدیده‌های رکود و فرار تومورها است. پیرو همین الگو، تحقیقاتی

[۱] **فائزه قربانی:** nastaranvafayi74@gmail.com

صورت گرفت که نشان داد روش‌های شیمی درمانی بسیار شدید در کنترل تومور ضعیف عمل می‌کنند و شیمی درمانی بسیار تهاجمی به سیستم ایمنی بدن آسیب وارد می‌کند (۱۰۳).

تحقیقات دیگری نشان می دهد که برخی از بیماران واکنش های ناهماهنگ و غیرمنتظره ای نسبت به درمان‌های سرطانی در مقایسه با روش های شیمی درمانی نشان می دهند.

برای درک بهتر این اتفاق، الگوی ریاضی تعاملات تومور با سیستم ایمنی و ردیابی سه جمعیت سلولی، سلول‌های تومور، سلول‌های ایمنی و سلول‌های عادی مورد بررسی قرار گرفت. از اکتشافات جالب آنالیز این الگو، آن بود که تنها در هنگام وجود سیستم ایمنی و قدرت کافی آن، تعادل کاملاً عاری از تومور مشاهد می‌شد (۱۰۴؛ ۱۰۵). سپس با افزودن جمعیت چهارم که میزان غلظت دارو در سیستم را نشان می‌دهد، معادله واکنش دارو در بدن معرفی شد که در آن هر معادله جمعیت سلولی با عبارتی برای تشریح آسیب‌های وارده به سلول‌ها توسط شیمی درمانی را داراست.

$$\frac{dE}{dt} = s - dE\frac{rET}{\sigma + T} - c_1 ET - k_1(1 - e^{-u}), \tag{۱.۲}$$

$$\frac{dT}{dt} = a_1 T(1 - b_1 T) - c_2 ET - c_3 NT - k_2(1 - e^{-u}), \tag{۲.۲}$$

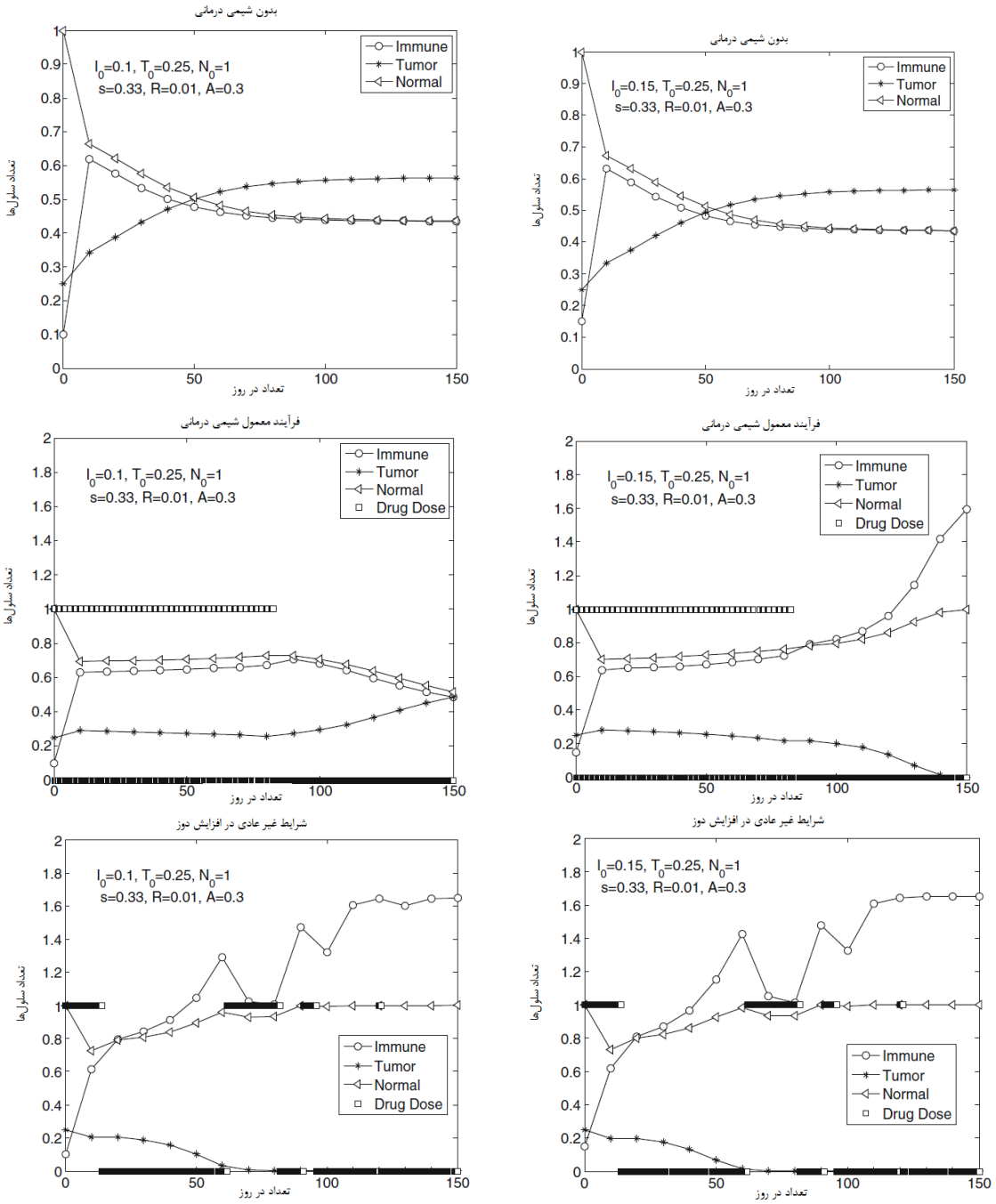
$$\frac{dN}{dt} = a_2 N(1 - b_2 N) - c_4 NT - k_3(1 - e^{-u}), \tag{۳.۲}$$

$$\frac{du}{dt} = v(t) - d_2 u. \tag{۴.۲}$$

در این معادلات E بیانگر جمعیت سلول‌های ایمنی تاثیر گذار، T جمعیت سلول‌های تومور، N نشان دهنده جمعیت سلول‌های نرمال و u میزان غلظت داروی شیمی درمانی می‌باشد. نرخ آسیب رسانی به هر جمعیت سلولی با حروف k_1، k_2 و k_3 در معادلات (۳.۲)-(۱.۲) نشان داده شده است که با کمترین نرخ آسیب رسانی به سلول‌های نرمال و بیشترین نرخ آسیب رسانی به سلول‌های تومور معرفی شده‌اند. با در نظر داشتن شرایط درمانی، شبیه‌سازی با برنامه‌های دوز دهی فرضی امکان‌پذیر شد. سادگی نسبی محاسبات این الگو کارایی تئوری کنترل بهینه را میسر می‌نماید. به شکل ۱ رجوع شود.

در این مدل، هدف به حداقل رساندن اندازه کل تومور در بازه زمانی از پیش تعیین شده است در حالی که به جمعیت سلول‌های نرمال آسیب چندانی نرسد. در این مورد، هدف دست یافتنی است. اما شبیه سازی‌ها نوساناتی را در اندازه‌های جمعیت سلولی به وجود آورد.

خوشبختانه هدف تابع، منحصر به فرد نیست و گزینه‌های ناشی از عوارض جانبی دیگر می‌تواند به نتایج متفاوتی منجر شود. برای مثال ممکن است اندازه تومور را به حداقل رساند در حالی که نواسانات در جمعیت سلول نیز به حداقل رسیده باشد. در سایر موارد، هدف به حداقل رساندن همزمان اندازه تومور و کل میزان دوزدهی شیمی درمانی است. علاوه بر این، می‌خواهیم بدانیم درصورتی که نقطه پایانی زمان درمان تومور نیز بهینه گردد نتایج درمانی چگونه تغییر می‌کنند؟ (۱۰۴؛ ۱۰۵؛ ۱۰۶).



شکل ۱: مقایسه پروتکل‌های شیمی درمانی و ایمن درمانی

۳ طراحی الگوهای ایمن درمانی

هدف از ایمن درمانی، تقویت توان دفاعی خود سیستم در برابر تومور است. از مزایای این رویکرد نسبت به شیمی درمانی سیتوتاکسیک، مسمومیت نسبتاً کمتر و در برخی موارد توانایی مورد هدف قراردادن سلول‌های تومور است که داروهای سیتوتاکسیک قادر به نفوذ در آن‌ها نیستند.

تحقیقات اصلی در این زمینه الگویی از معادلات دیفرانسیل عادی سه جمعیتی درمان $IL2$ است. این الگو سلول‌های تومور،

جمعیتی از سلول‌های تاثیرگذار ایمنی فعال شده شامل سیتوتاکسیک T، ماکروفاژ و سلول‌های کشنده که نقش تهاجمی دارند و تراکمی از سیتوکین $IL۲$ که نقش افزایش قدرت سیستم ایمنی را دارد، ردیابی می‌کند.

$IL۲$ یکی از سیتوکین‌های ضد التهابی است که در یک سلول میزبان به عنوان فعالیت ایمنی طبیعی ایجاد می‌گردد. تزریق تراکم‌های $IL۲$ به عنوان بخشی از روش‌های خارجی درمان سرطان به کار گرفته می‌شود که می‌تواند بخشی از سلول‌های تومور را نابود کند (۱۰۸).

$IL۲$ داخلی نیز از طریق تعاملات میان جمعیت تأثیرگذار و سلول‌های سرطانی تحریک می‌شود. در این الگو مشاهده می‌شود که میزان بسیار پایین $IL۲$ فعالیت سیستم ایمنی را به میزان کافی بالا نمی‌برد در حالی که میزان بی رویه آن نیز می‌تواند تاثیرات مخربی به وجود آورد. پس می‌توان نتیجه گرفت که $IL۲$ به تنهایی در هر دوز نمی‌تواند رشد تومور را کنترل کند، اما میزان صحیح ترکیب با نوعی از داروهای شیمی درمانی می‌تواند منجر به از بین رفتن موفقیت آمیز تومور شود.

۴ طراحی الگوهای ترکیبی شیمی درمانی و ایمن درمانی

ترکیب شیمی درمانی با ایمن درمانی می‌تواند حداقل از دو جنبه حائز اهمیت باشد:

۱–محافظت بیمار از عفونت

۲–مبارزه با خود سرطان.

درمان‌های سیتوتاکسیک سیستم ایمنی بدن را تضعیف می‌کنند و همین امر، بیمار را مستعد عفونت‌های خطرناکی می‌کند. حفظ سیستم ایمنی قوی به ویژه در زمان شیمی درمانی برای کنترل موفقیت آمیز رشد تومور امری ضروری می‌باشد. بنابراین مدلی برپایه ترکیب شیمی درمانی سیتوتاکسیک با ایمن درمانی ایجاد شد. این مدل شش جمعیت را ردیابی می‌کند:

۱–سلول‌های تومور،

۲–سلول‌های NK (لنفوسیت‌هایی از سیستم ایمنی بدن هستند که کشنده طبیعی هستند)،

۳–لنفوسیت $CD۸ + T$ (نوعی لنفوسیت T (گلبول سفید خون) که سلول‌های سرطانی را از بین می‌برد و نقش کشنده دارد)،

۴–جمعیت لنفوسیت وارد شده به گردش خون،

۵–غلظت شیمی درمانی سیتوتاکسیک

۶–غلظت $IL۲$ خارجی (۱۰۷).

ایمن درمانی متشکل از تقویت سلول‌های $CD۸ + T$ و تزریق $IL۲$ هستند و شیمی درمانی با یک سیستم خاص سیتوتاکسیک و چرخه‌های غیرسلولی معرفی شده است. این مدل ثابت کرد که تزریق داروهای ایمن درمانی همزمان با شیمی درمانی موثرترین روش نابودی سرطان است. سوال مهمی که در بررسی روش‌های درمانی ترکیبی مطرح می‌شود، وجود راهی برای بهینه‌سازی برنامه‌های درمانی است.

۵ نتیجه‌گیری

در این مقاله، مدل‌های شیمی درمانی و همچنین ایمن درمانی سرطان مورد بحث قرار گرفت و بهترین و کاملترین نوع درمان ارائه شد. امید داریم مطالعه این مقاله تلاشی الهام‌بخش در زمینه الگوهای ریاضی درمان سرطان برای خوانندگان باشد.

مراجع

[103] K. Roesch, D. Hasenclever, M. Scholz, Modelling Lymphoma Therapy and Outcome, *Bull Math Biol*, **76**(2) (2014) 401-430.

[104] L. de Pillis, A. Radunskaya, *A mathematical tumor model with immune resistance and drug therapy*, an optimal control approach. J Theor Med, **3** (2001) 79-100.

[105] L. de Pillis, A. Radunskaya, *The dynamics of an optimally controlled tumor model*, a case study. Math Comput Model (Special Issue), **37** (2003) 1221-1244.

[106] L. de Pillis, W. Gu, K. Fister, T. Head, K. Maples, A. Murugan et al, *Chemotherapy for tumors*, an analysis of the dynamics and a study of quadratic and linear optimal controls. Math Biosci, **209** (2007) 292-315.

[107] L. dePillis, A. Radunskaya, C. Wiseman, A validated mathematical model of cell-mediated immune response to tumor growth, *Cancer Res*, **65**(1) (2005) 7950-7958.

[108] D. Kirschner, J. Panetta, Modeling immunotherapy of the tumor-immune interaction, *J Math Biol*, **37**(3) (1998) 235-252.

اولین همایش ملی ریاضیات زیستی
۲۱ و ۲۲ اسفندماه ۱۳۹۷
همایش ملی ریاضیات زیستی
دانشگاه نیشابور

بررسی نقاط تعادل وپایداری یک مدل شکار – شکارچی به همراه یک منبع زیستی

فرزانه لطفی [۱]، محمد شیرازیان گروه ریاضی – دانشگاه نیشابور– نیشابور– ایران

چکیده: این کار به مدل شکار – شکارچی در محیطی که در آن ظرفیت حمل به صورت یک متغیری از زمان فرض شده است می‌پردازد ویکی از گونه‌ها از دیگری تغذیه می‌کند. یک منبع مشترک توسط دو گونه‌ی رقابتی مصرف می‌شود وهمزمان شکارچی شکار را مصرف می‌کند. هدف اصلی ما شامل دوحالت می‌باشد :مدلی با ظرفیت حمل متغیر ودیگری رقابت درون گونه‌ای . حالت‌های پایدار را مشخص می‌کنیم ورفتارهای دینامیکی را مورد بررسی قرار می‌دهیم. همچنین پایداری سراسری ومحلی نقطه تعادل داخلی را توسط معیار روت –هورویتز بررسی می‌کنیم وسپس یک تابع لیاپانوف مناسب متناظر با آن را ارائه می‌دهیم.

واژه‌های کلیدی: مدل شکار – شکارچی، ظرفیت حمل، نقطه تعادل، تابع لیاپانوف

۱ مقدمه

در اطراف ما بسیاری از مسائل جالب زیستی وجود دارد که نیاز به مدل‌سازی ریاضی برای درک رفتار بحرانی و ماهیت ذاتی سیستم دارد. هدف اصلی توسعه مدل، محاسبه تغییرات در جمعیت خاصی نیست بلکه مطالعه میزان پیچیدگی موجود در محیط‌زیست است. در جمعیت زیستی عموما ظرفیت حمل میزان واقعی جمعیت را بر اساس محیط آن پیش‌بینی می‌کند. ظرفیت حمل یک عامل بسیار مهم است زیرا نشان‌دهنده‌ی سرعت و بالاترین سطح رشد جمعیت می‌باشد. در تحقیقات اخیر، ظرفیت حمل مقدار ثابت فرض شده است اما در محیط متغیر پویا ظرفیت حمل به عنوان یک متغیر حالت در نظر گرفته شده است در مدلی که توسط هایل [۱] وبورز [۲] معرفی شده است جمعیت شکار – شکارچی را به طور لجستیک با ظرفیت حمل ثابت به صورت زیر رشد می‌کند:

$$\frac{dx}{dt} = r_۱ x\left(۱ - \frac{x}{k_۱}\right) - aF(x)y,$$
$$\frac{dy}{dt} = r_۲ y\left(۱ - \frac{y}{k_۲}\right) + bF(x)y,$$

[1] Hoyle
[2] Bowers

[۱] فرزانه لطفی: farzanelotfi96@gmail.com

که در آن X و Y توده‌ی زیستی جمعیت شکار و شکارچی است که به صورت لجستیک با نرخ‌های رشد $r_۱$ و $r_۲$ ظرفیت‌های حمل $k_۱$ و $k_۲$ به ترتیب رشد می‌کنند. $F(x)$ تابع پاسخ را نشان می‌دهد (۱۱۰). در مدل شکار– شکارچی معرفی شده توسط لسلی [۳] و گاور [۴]، ظرفیت حمل جمعیت شکار مقدار ثابت k فرض شده است و ظرفیت حمل شکارچی با ظرفیت حمل شکار متناسب است (۱۱۱؛ ۱۱۲). این کار می‌تواند به بیشتر از دو گونه در مدل زنجیره‌غذایی گسترش یابد.

گاهی اوقات مشاهده می‌شود که غنی‌سازی منبع زیستی از چند طریق بر محیط زیست تاثیر می‌گذارد. غنی‌سازی منبع که برای رشد جمعیتی سودمند است ممکن است باعث کاهش تنوع گونه‌ها گردد. درک درستی از برهم‌کنش گونه‌های مختلف در یک محیط زیست برای مدیریت منابع زیستی قابل برداشت خیلی مهم است. هر گونه ، جزئی از سیستم‌های زیستی پیچیده است که با دیگر گونه‌ها برهم‌کنش دارد این برهم‌کنش‌ها ممکن است شکل‌های مختلفی از قبیل رقابتی ، شکار کردن ، همزیستی (انگلی)، همزیستی دو موجود را داشته باشند. شکار درون گروهی ترکیبی از اولی و دومی است یعنی رقابتی وشکار کردن که به معنی کشتن ومصرف کردن گونه‌هایی که سهمی در منبع زیستی محدود و مشترک دارند است. شکارچی درون گروهی غالب، دو سود به‌دست‌می‌آورد تغذیه کردن و حذف رقیب‌بالقوه . شکار درون گروهی از هر دو شکار کلاسیک و شکار رقابتی متمایز است زیرا رقبای بالقوه را کاهش می‌دهد. بنابراین تاثیر قویتری روی دینامیکهای جمعیت به نسبت شکار ورقابت به همراه دارد. برداشت از گونه‌های مختلف روی سیستم محیط زیست تاثیر زیادی می‌گذارد. در اینجا شکار درون گروهی به همراه دینامیکهای شکار– شکارچی در سه معادله دیفرانسیل برای شکار، شکارچی و منابع رخ می‌دهد. نتایج غنی‌سازی منبع زیستی رادر سطح کم، متوسط و بالا بررسی می‌کنیم همچنین نقاط تعادل زیستی احتمالی را بررسی می‌کنیم.

۲ فرمول‌بندی مدل

برای مطالعه سیستم شکار – شکارچی مدلی را ارائه می‌دهیم که توسعه‌ی مدل پیشنهاد شده توسط لسلی (۱۱۱) و بیسنر و راس (۱۰۹) وسافوان وهمکاران (۱۱۳) می‌باشد. استراتژی برداشت مستقل را به عنوان تعمیم ممکن برای یافته‌های جدید در نظر می‌گیریم :

$$\begin{aligned}
\frac{dX}{dt} &= r_۱X\left(۱-\frac{X}{mZ}\right)-aXY-E_۱X,\\
\frac{dY}{dt} &= r_۲Y\left(۱-\frac{Y}{nZ}\right)+bXY-E_۲Y,\\
\frac{dZ}{dt} &= Z\left(c-uX-vY\right)
\end{aligned} \tag{۱.۲}$$

X و Y توده‌ی زیستی جمعیت شکار وشکارچی است که به صورت لجستیکی به ترتیب با نرخ رشد $r_۱$ و $r_۲$ رشد می‌کند. Z منبع زیستی را نشان می‌دهد. سومین معادله منبع زیستی که به صورت پیوسته با زمان تغییر می‌کند جمله‌های mZ و nZ به‌ترتیب ظرفیت‌های حمل زیست‌محیطی برای جمعیت شکار وشکارچی هستند در اینجا، $m+n=۱$. اگر $m<n$ باشد بدین معناست که گونه‌های شکار دسترسی به منبع کمتری در مقایسه با شکارچی دارند.منبع زیستی با نرخ توسعه c رشد می‌کند. uZX و vZY به ترتیب نرخ‌های مصرف منبع توسط گونه‌های شکار وشکارچی هستند که در آن v و u ثابت هستند. $E_۱$ و $E_۲$ به ترتیب برداشت‌های مستقل از گونه‌های شکار و شکارچی هستند.

[3] Leslie
[4] Gower

۳ تجزیه و تحلیل مدل

با استفاده از تبدیلات $x = \frac{bX}{r_1}$ ، $y = \frac{aY}{r_1}$ ، $z = \frac{bpZ}{r_1}$ ، $\tau = r_1 t$ ، $\alpha = \frac{r_2}{r_1}$ ، $\beta = \frac{bp}{aq}$ ، $\gamma = \frac{c}{r_1}$ ، $\delta = \frac{u}{b}$ ، $\epsilon = \frac{v}{a}$ ، $\mu = \frac{E_1}{r_1}$ ، $\nu = \frac{E_2}{r_1}$ در معادلات (۱.۲) به شکل بی بعد زیر نشان داده می‌شود:

$$\begin{aligned}
\frac{dx}{d\tau} &= x\left(1 - \frac{x}{z}\right) - xy - \mu x, \\
\frac{dy}{d\tau} &= \alpha y\left(1 - \frac{\beta y}{z}\right) + xy - \nu y \\
\frac{dz}{d\tau} &= z(\gamma - \delta x - \epsilon y)
\end{aligned} \tag{۱.۳}$$

برای به دست آوردن نقاط تعادل، قرار می دهیم $\frac{dx}{d\tau} = 0$، $\frac{dy}{d\tau} = 0$ و $\frac{dz}{d\tau} = 0$. بنابراین

$$x\left(1 - \frac{x}{z}\right) - xy - \mu x = 0 \Rightarrow \begin{cases} x = 0 \\ \left(1 - \frac{x}{z}\right) - y - \mu = 0 \end{cases}$$

$$\alpha y\left(1 - \frac{\beta y}{z}\right) + xy - \nu y = 0 \Rightarrow \begin{cases} y = 0 \\ \alpha\left(1 - \frac{\beta y}{z}\right) + x - \nu = 0 \end{cases}$$

$$z(\gamma - \delta x - \epsilon y) = 0 \Rightarrow \begin{cases} z = 0 \\ \gamma - \delta x - \epsilon y = 0 \end{cases}$$

دو نقطه تعادل به صورت $\left(\frac{\gamma}{\delta}, 0, \frac{\gamma}{\delta(1-\mu)}\right)$ ، $\left(0, \frac{\gamma}{\epsilon}, \frac{\alpha\beta\gamma}{\epsilon(\alpha-\nu)}\right)$ و نقطه تعادل داخلی توسط $(\hat{x}, \hat{y}, \hat{z})$ داده می‌شود، که در آن

$$\begin{aligned}
\hat{y} &= \frac{\gamma - \delta\hat{x}}{\epsilon} \\
\hat{z} &= \frac{\epsilon\hat{x}}{\phi + \delta\hat{x}} \\
\phi &= \epsilon(1 - \mu) - \gamma
\end{aligned} \tag{۲.۳}$$

و $\hat{x}$ در معادله‌ی زیر صدق می‌کند.

$$(\alpha\beta\delta^2 + \epsilon^2)\hat{x}^2 + ((\alpha - \nu)\epsilon^2 + \alpha\beta\delta(\phi - \gamma))\hat{x} - \alpha\beta\gamma\phi = 0.$$

۱.۳ بررسی پایداری نقاط تعادل

لم ۱.۳. فرض کنید $\delta_1 = r_1 - E_1$ و $\delta_2 = r_2 - E_2$ آنگاه :

- اگر $\delta_2 > 0$ سیستم مدل (۱.۳) یک نقطه تعادل بدون شکار $\left(0, \frac{\gamma}{\epsilon}, \frac{\alpha\beta\gamma}{\epsilon(\alpha-\nu)}\right)$ دارد که پایدار است هرگاه $\gamma > (1 - \mu)\epsilon$ ، ناپایدار است هرگاه $\gamma < (1 - \mu)\epsilon$. ، در این حالت نقطه تعادل غیربدیهی، پایدار است بر حسب میزان برداشت این نقطه پایدار است هرگاه $E_1 > r_1 - \frac{ac}{\nu}$ ، ناپایدار است هرگاه $E_1 < r_1 - \frac{ac}{\nu}$.
- اگر $\delta_1 > 0$ سیستم (۱.۳) یک نقطه تعادل بدون شکارچی دارد $\left(\frac{\gamma}{\delta}, 0, \frac{\gamma}{\delta(1-\mu)}\right)$ پایدار است هرگاه $\nu > \alpha + \frac{\gamma}{\delta}$ ، ناپایدار است هرگاه $\nu < \alpha + \frac{\gamma}{\delta}$ ، در این حالت نقطه تعادل غیربدیهی، پایدار است. بر حسب میزان برداشت این نقطه پایدار است هرگاه $E_2 > r_2 + \frac{bc}{u}$ ، ناپایدار است هرگاه $E_2 < r_1 + \frac{bc}{u}$.

□□□. اگر $\hat{x} < \frac{\gamma}{\delta}$ و $\epsilon(۱-\mu) > \gamma$ آنگاه سیستم (۱.۳) دارای یک نقطه تعادل داخلی است.

اثبات. ژاکوبین در حالت کلی برای معادله (۱.۳) به شکل زیر است:

$$J(x,y,z) = \begin{pmatrix} ۱ - \frac{۲x}{z} - y - \mu & -x & \frac{x^۲}{z^۲} \\ y & \alpha - \frac{۲\alpha\beta y}{z} + x - \nu & \frac{\alpha\beta y^۲}{z^۲} \\ -\delta z & -\epsilon z & \gamma - \delta x - \epsilon y \end{pmatrix} \tag{۳.۳}$$

□. با جایگذاری نقطه تعادل بدون شکار داریم:

$$J\left(۰, \frac{\gamma}{\epsilon}, \frac{\alpha\beta\gamma}{\epsilon(\alpha-\nu)}\right) = \begin{pmatrix} ۱ - \frac{\gamma}{\epsilon} - \mu & ۰ & ۰ \\ \frac{\gamma}{\epsilon} & \nu - \alpha & \frac{(\nu-\alpha)^۲}{\alpha\beta} \\ -\frac{\delta\alpha\beta\gamma}{\epsilon(\nu-\alpha)} & \frac{\alpha\beta\gamma}{(\nu-\alpha)} & ۰ \end{pmatrix} \tag{۴.۳}$$

که مقادیر ویژه آن عبارتند از:

$$۱ - \frac{\gamma}{\epsilon} - \mu, \quad \frac{۱}{۲}(-\alpha + \nu \pm \sqrt{\alpha^۲ - ۲\alpha\nu + \nu^۲ + ۴\gamma\nu - ۴\gamma\alpha}).$$

سیستم پایدار است هر گاه تمام مقادیر ویژه فوق منفی باشند، یعنی

- $۱ - \frac{\gamma}{\epsilon} - \mu < ۰$ در نتیجه $\gamma < (۱-\mu)\epsilon$
- $\frac{۱}{۲}(-\alpha + \nu \pm \sqrt{\alpha^۲ - ۲\alpha\nu + \nu^۲ + ۴\gamma\nu - ۴\gamma\alpha}) < ۰$ بنابراین $۴\gamma\nu - ۴\gamma\alpha < ۰$ یا $\gamma(\nu-\alpha) < ۰$ و در نتیجه با فرض $\delta_۲ > ۰$ داریم $\nu - \alpha < ۰$ یا $\nu < \alpha$.

در غیر این صورت سیستم ناپایدار خواهد بود.

□□. با جایگذاری نقطه تعادل بدون شکارچی در ماتریس ژاکوبین (۳.۳) داریم:

$$J\left(\frac{\gamma}{\delta}, ۰, \frac{\gamma}{\delta(۱-\mu)}\right) = \begin{pmatrix} \mu - ۱ & -\frac{\gamma}{\delta} & (\mu-۱)^۲ \\ ۰ & \alpha + \frac{\gamma}{\delta} - \nu & ۰ \\ \frac{\gamma}{\mu-۱} & \frac{\gamma\epsilon}{\delta(\mu-۱)} & ۰ \end{pmatrix} \tag{۵.۳}$$

که مقادیر ویژه آن عبارتند از:

$$\alpha - \nu + \frac{\gamma}{\delta}, \quad \frac{۱}{۲}(-۱ + \mu \pm \sqrt{۱ - ۲\mu + \mu^۲ + ۴\gamma\mu - ۴\gamma})$$

سیستم پایدار است هر گاه تمام مقادیر ویژه فوق منفی باشند یعنی

- $\alpha - \nu + \frac{\gamma}{\delta} < ۰$ در نتیجه $\alpha + \frac{\gamma}{\delta} < \nu$
- $\frac{۱}{۲}(-۱ + \mu \pm \sqrt{۱ - ۲\mu + \mu^۲ + ۴\gamma\mu - ۴\gamma}) < ۰$ بنابراین $۴\gamma\mu - ۴\gamma < ۰$ یا $\gamma(\mu-۱) < ۰$ و در نتیجه با فرض $\delta_۱ > ۰$ داریم $\mu - ۱ < ۰$ یا $\mu < ۱$.

در غیر این صورت سیستم ناپایدار است.

□

قضیه ۲.۳. اگر $\hat{z} \leq \min\{\frac{\epsilon}{\delta}, \frac{۱}{\epsilon}, ۱\}$ آنگاه سیستم (۱.۳) پایدار مجانبی محلی است.

اثبات. معادله مشخصه ماتریس ژاکوبین سیستم مدل (۱.۳) در نقطه تعادل داخلی، یعنی $\det(J(\hat{x}, \hat{y}, \hat{z}) - \lambda I) = ۰$ داریم:

$$\lambda^۳ + a_۱\lambda^۲ + a_۲\lambda + a_۳ = ۰$$

که در آن

$$\begin{aligned}
a_۱ &= \frac{\hat{x} + \hat{y}\alpha\beta}{\hat{z}^۲} \\
a_۲ &= \frac{\hat{x}\hat{y}\hat{z}^۲ + \hat{x}\hat{y}\alpha\beta + \hat{x}^۲\hat{z}\delta + \hat{y}^۲\hat{z}\alpha\beta\epsilon}{\hat{z}^۲} \\
a_۳ &= \frac{\hat{x}^۲\hat{y}\alpha\beta\delta - \hat{x}\hat{y}^۲\hat{z}\alpha\beta\delta + \hat{x}^۲\hat{y}\hat{z}\epsilon + \hat{x}\hat{y}^۲\alpha\beta\epsilon}{\hat{z}^۲}
\end{aligned}$$

به وضوح $a_۱ > ۰$ $a_۲ > ۰$ اگر $\hat{z} < \frac{\epsilon}{\delta}$ آنگاه $a_۳ > ۰$ اکنون اگر $۱ - \hat{z}\epsilon > ۰$ و $۱ - \hat{z} > ۰$ آنگاه

$$\begin{aligned}
a_۱a_۲ - a_۳ = &\frac{\hat{x} + \hat{y}\alpha\beta}{\hat{z}^۲}\frac{\hat{x}\hat{y}\hat{z}^۲ + \hat{x}\hat{y}\alpha\beta + \hat{x}^۲\hat{z}\delta + \hat{y}^۲\hat{z}\alpha\beta\epsilon}{\hat{z}^۲} \\
&- \frac{\hat{x}^۲\hat{y}\alpha\beta\delta - \hat{x}\hat{y}^۲\hat{z}\alpha\beta\delta + \hat{x}^۲\hat{y}\hat{z}\epsilon + \hat{x}\hat{y}^۲\alpha\beta\epsilon}{\hat{z}^۲} \\
&= [\hat{x}^۲\hat{y}\hat{z}^۲(۱ - \hat{z}\epsilon) + \hat{x}^۲\hat{y}\alpha\beta + \hat{x}^۳\hat{z}\delta + \hat{x}\hat{y}^۲\hat{z}\alpha\beta\epsilon(۱ - \hat{z}) + \alpha\beta\hat{x}\hat{y}^۲\hat{z}^۲ \\
&+ \hat{x}\hat{y}^۲\alpha^۲\beta^۲ + \alpha\beta\delta\hat{x}^۲\hat{y}\hat{z}(۱ - \hat{z}) + \alpha^۲\beta^۲\epsilon\hat{y}^۳\hat{z} + \hat{x}\hat{y}^۲\hat{z}^۳\alpha\beta\delta]/\hat{z}^۴ \\
&> ۰,
\end{aligned}$$

بنابراین $a_۱ > ۰$ ، $a_۳ > ۰$ و $a_۱a_۲ - a_۳ > ۰$ برقرار است هرگاه $\hat{z} \leq \{\min \frac{\epsilon}{\delta}, \frac{۱}{\epsilon}, ۱\}$ با استفاده از معیار روت- هورویتز سیستم، پایدار مجانبی محلی است. □

قضیه ۳.۳. سیستم (۱.۳) در ناحیه‌ای که در آن $\alpha\beta y > \hat{z}(\epsilon + \delta)z$ باشد پایدار مجانبی سراسری است که z چگالی تعادل منابع زیستی را نشان می‌دهد.

اثبات. برای اثبات پایداری سراسری سیستم یک تابع لیاپانوف می‌سازیم.

$$V(x, y, z) = \left[(x - \hat{x}) - \hat{x}\log(\frac{x}{\hat{x}})\right] + H_۱\left[(y - \hat{y}) - \hat{y}\log(\frac{y}{\hat{y}})\right] + H_۲\left[(z - \hat{z}) - \hat{z}\log(\frac{z}{\hat{z}})\right]$$

اکنون داریم :

$$\begin{aligned}
\frac{dv}{dt} &= \frac{dv}{dx}\frac{dx}{dt} + \frac{dv}{dy}\frac{dy}{dt} + \frac{dv}{dz}\frac{dz}{dt}\frac{dv}{dt} = \frac{x-\hat{x}}{x}\frac{dx}{dt} + H_1\frac{y-\hat{y}}{y}\frac{dy}{dt} + H_2\frac{z-\hat{z}}{z}\frac{dz}{dt} \\
&= (x-\hat{x})\left[1 - \frac{x}{z} - y - \mu - 1 + \frac{\hat{x}}{\hat{z}} + \hat{y} + \mu\right] + H_1(y-\hat{y})\left[\alpha - \alpha\beta\frac{y}{z}\right. \\
&\left. + x - \nu - \alpha + \alpha\beta\frac{\hat{y}}{\hat{z}} - \hat{x} + \nu\right] + H_2(z-\hat{z})\left[\gamma - \delta x - \epsilon y - \gamma + \delta\hat{x} + \epsilon\hat{y}\right] \\
&= -(x-\hat{x})\left[\frac{z(x-\hat{x}) - x(z-\hat{z})}{z\hat{z}} + (y-\hat{y})\right] + H_1(y-\hat{y})\left[\alpha\beta\frac{y(z-\hat{z}) - z(y-\hat{y})}{z\hat{z}} + x - \hat{x}\right] \\
&- H_2\delta(z-\hat{z})(x-\hat{x}) - H_2\epsilon(z-\hat{z})(y-\hat{y}) \\
&= -\frac{(x-\hat{x})^2}{\hat{z}} + \frac{x(x-\hat{x})(z-\hat{z})}{z\hat{z}} - (x-\hat{x})(y-\hat{y}) - H_1\frac{\alpha\beta(y-\hat{y})^2}{\hat{z}} + H_1\frac{\alpha\beta(y-\hat{y})(z-\hat{z})y}{z\hat{z}} \\
&+ H_1(y-\hat{y})(x-\hat{x}) - H_2\delta(x-\hat{x})(z-\hat{z}) - H_2\epsilon(y-\hat{y})(z-\hat{z}) \\
&\leq -\left[\frac{(x-\hat{x})^2}{\hat{z}} + H_1\alpha\beta\frac{(y-\hat{y})^2}{\hat{z}}\right] + \frac{1}{2}\left[H_2\epsilon + \frac{H_1\alpha\beta y}{z\hat{z}}\right]\left[(y-\hat{y})^2 + (z-\hat{z})^2\right] \\
&+ \frac{1}{2}\left[-H_2\delta + \frac{x}{z\hat{z}}\right]\left[(x-\hat{x})^2 + (z-\hat{z})^2\right] \\
&= \left[\frac{1}{2}\left(-H_2\delta + \frac{x}{z\hat{z}}\right) - \frac{1}{\hat{z}}\right](x-\hat{x})^2 + \left[\frac{1}{2}\left(-H_2\epsilon + \frac{H_1\alpha\beta y}{z\hat{z}}\right) - \frac{H_1\alpha\beta}{\hat{z}}\right](y-\hat{y})^2 \\
&+ \frac{1}{2}\left[-H_2\epsilon + \frac{H_1\alpha\beta y}{z\hat{z}} - H_2\delta + \frac{x}{z\hat{z}}\right](z-\hat{z})^2 < 0
\end{aligned}$$

هرگاه $\alpha\beta y + x > \hat{z}(\epsilon+\delta)z$ در ناحیه‌ای که در $\frac{H_1\alpha\beta y}{z\hat{z}} + \frac{x}{z\hat{z}} > H_2\epsilon + H_2\delta$ سیستم مدل (۱.۳) پایدار مجانبی سراسری است.

□

مراجع

[109] B. Basener and D. S. Ross, Booming and crashing populations and Easter Island, SIAM J. Appl. Math. 65 (2005) 684

[110] A. Hoyle and R. Bowers, When is evolutionary branching in predator–prey systems possible with an explicit carrying capacity? Math. Biosci. 210 (2007) 1.

[111] P. H. Leslie, A stochastic model for studying the properties of certain biological systems by numerical methods, Biometrika 45 (1958) 16.

[112] P. H. Leslie and J. C. Gower, The properties of a stochastic model for the predator– prey type of interaction between two species, Biometrika 47 (1960) 219.

[113] H. Safuan, I. Towers, Z. Jovanoski and H. Sidhu, ANZIAMJ 53 (2012) C172.

مقایسه فواصل اطمینان برای نسبت شانس

اعظم مسلمی، استادیار، هیئت علمی گروه آمار زیستی،دانشکده پزشکی، دانشگاه علوم پزشکی اراک، اراک، ایران محمد رفیعی، دانشیار، هیئت علمی گروه آمار زیستی، دانشکده پزشکی، دانشگاه علوم پزشکی اراک، اراک، ایران راشد پورحمیدی، دانشجوی ارشد آمار زیستی، دانشکده پزشکی، دانشگاه علوم پزشکی اراک، اراک، ایران سردار جهانی، دانشجوی ارشد آمار زیستی، دانشکده پزشکی، دانشگاه علوم پزشکی اراک، اراک، ایران کورش رضایی، مربی، هیئت علمی گروه پرستاری، دانشکده پرستاری، دانشگاه علوم پزشکی اراک، اراک، ایران مهشید عسکری [1]، دانشجوی ارشد آمار زیستی، دانشکده پزشکی، دانشگاه علوم پزشکی اراک، اراک، ایران

چکیده: در مسائل تجزیه و تحلیل داده ها بخش کوچکی از جامعه که در قالب نمونه استخراج شده است کاربرد دارد. در استنباط آماری از برآوردهای نقطه ای و فاصله ای (فاصله اطمینان) برای برآورد پارامتر مورد نظر استفاده می کنند. برآورد فاصله ای دقت بالاتری نسبت به برآورد نقطه ای دارد. در مطالعات بالینی که از نسبت شانس استفاده می شود، این شاخص شدت ارتباط بین هر یک از متغیرهای مستقل و وجود یا فقدان بیماری را فراهم می کند. در این مقاله هشت فاصله اطمینان نسبت شانس را که شامل فواصل اطمینان دقیق و مجانبی هستند، بر اساس طول فاصله اطمینان مورد بررسی قرار می‌دهیم.

پژوهش مذکور به بررسی انواع فاصله اطمینان نسبت شانس می‌پردازد و مقادیر آنها در مثالی تحت یک کار آزمایی بالینی یک سوکور محاسبه می‌شوند. مشارکت‌کنندگان این کار آزمایی بالینی ۸۰ نفر از بیماران مبتلابه سکته قلبی در بیمارستان‌های اراک در سال ۱۳۹۷ بودند. نمونه‌ها با استفاده از روش تخصیص تصادفی به دو گروه مداخله و کنترل تقسیم شدند. به‌منظور جمع‌آوری داده‌ها از چک‌لیست بررسی اختلال پس از سانحه $(PTSD)$ استفاده شد. برای بیماران در گروه مداخله مراقبت معنوی به مدت (سه روز) انجام شد و یک ماه پس از ترخیص بیمار از وی خواسته شد تا چک‌لیست را تکمیل کند. تجزیه‌وتحلیل داده‌ها از طریق نرم‌افزار R نسخه ۳.۵.۲ باشد. و به منظور تحلیل داده ها از آزمون‌های آماری کای اسکور، نسبت شانس، فاصله اطمینان ولف لوجیت، گارت، هموار مستقل، کارن فلید دقیق، کارن فیلد مید-پی، اگرستی-مین و باپتیستا-پیک دقیق وباپتیستا-پیک مید-پی استفاده شد.

در این مطالعه مقدار نسبت شانس ۰/۱۱ است. ازسه فاصله اطمینان مجانبی بهترین فاصله اطمینان مجانبی گارت با طول فاصله ۲/۰۳۶ می باشد؛ و از بین پنج فاصله اطمینان دقیق بهترین فاصله اطمینان باپتیستا- پیک مید-پی با طول فاصله ۱/۹۷۳ است.

در مطالعه حاضر انواع فاصله اطمینان برای نسبت شانس بررسی شدند. این فواصل شامل فواصل اطمینان دقیق و فواصل اطمینان مجانبی بودند. از سه فاصله اطمینان مجانبی ذکرشده بهترین فاصله اطمینان گارت است، که طول کمتر و دقت بیشتری دارد. اما در بین پنج فاصله اطمینان دقیق، فاصله اطمینان نسبت شانس باپتیستا- پیک مید-پی طول کمتر و دقت بهتری دارد. لذا در این مطالعه بهترین فاصله اطمینان باپتیستا- پیک مید-پی است.

---

[1] مهشید عسکری: mahshidasgri70@gmail.com

واژه‌های کلیدی: نسبت شانس، فاصله اطمینان، سکته قلبی،$PTSD$ ،طول فاصله اطمینان

۱ مقدمه

نسبت شانس [۱] مقیاس مهمی برای دو نسبت مستقل دوجمله‌ای است. اغلب در مطالعات مورد شاهدی و مقطعی و کوهورت مورداستفاده قرار می‌گیرد. همچنین در کار آزمایی بالینی یک معیار در متاآنالیزها محسوب می‌شود[۱]. نسبت شانس مقیاسی از ارتباط بین پیامد و مواجهه است و تقریب مناسبی برای بیماری‌های نادر می‌باشد. اگر $OR = ۱$ مواجهه بر پیامد مؤثر نیست و اگر $OR > ۱$ مواجهه بر پیامد تأثیر مثبت و افزایشی دارد و در صورتی $OR < ۱$ مواجهه بر پیامد تأثیر منفی و کاهشی دارد[۲]. آمار در پزشکی به‌کرات به مقایسه دو نسبت دوحالتی می‌پردازد. متغیر دوحالتی متغیری است که دارای دو نتیجه موفقیت و شکست است و موفقیت لزوماً نشان‌دهنده یک نتیجه مطلوب نیست. نتایج دو گروه مستقل را می‌توان در جدول ۲×۲ خلاصه کرد. جدول توافقی که اولین بار توسط لاندرسون [۲] ارائه شد، به‌صورت زیر است[۳].

جدول ۱: جدول توافقی ۲×۲

	موفقیت	شکست	جمع
کنترل	$n_{۱۱}$	$n_{۱۲}$	$n_{۱+}$
مداخله	$n_{۲۱}$	$n_{۲۲}$	$n_{۲+}$
جمع	$n_{+۱}$	$n_{+۲}$	N

تعداد افراد را در گروه‌ها با ($n_{۱+}, n_{۲+}$) نشان می‌دهیم. در این جدول، حاشیه ثابت فرض شده است. جدول توافقی با حاشیه ثابت در مطالعات کار آزمایی بالینی و مورد شاهدی و کوهورت بسیار کاربرد دارد. $n_{۱۱}$ موفقیت در گروه اول دارای توزیع دوجمله‌ای با پارامتر $n_{۱+}$ و $p_۱$ است و $n_{۲۱}$ موفقیت در گروه دوم دارای توزیع دوجمله‌ای با پارامتر $n_{۲+}$ و $p_۲$ است. لذا برآورد موفقیت‌ها را از نسبت درستمایی ماکزیمم به‌صورت مقابل به دست می‌آوریم.

$$p_۱ = \frac{n_{۱۲}}{n_{۱+}} \quad , \quad p_۲ = \frac{n_{۲۱}}{n_{۲+}}$$

و نسبت شانس را به‌صورت مقابل برآورد می‌کنیم

$$\widehat{\theta} = \frac{n_{۱۱}/n_{۱۲}}{n_{۲۱}/n_{۲۲}} = \frac{n_{۱۱}n_{۱۲}}{n_{۲۱}n_{۲۲}} \qquad (۱.۱)$$

فاصله اطمینان برای نسبت شانس شامل فواصل اطمینان مجانبی و دقیق است. فاصله اطمینان مجانبی بر اساس تقریبی از توزیع نرمال ساخته می‌شود. فاصله اطمینان دقیق بر اساس توزیع دقیق مشاهدات ساخته می‌شود فاصله اطمینان دقیق به ویژه در حالت حجم نمونه کوچک و نسبت نزدیک به ۰ و ۱ کاربرد دارد.

۲ مواد و روشها

داده های این مطالعه از یک کار آزمایی بالینی یک سوکور بدست آمد. که بر بیماران سکته قلبی حاد بستری در بخش‌هایccu بیمارستان‌های امیرکبیر و امیرالمؤمنین اراک در دو گروه مداخله و کنترل اجرا شد. در گروه مداخله مراقبت معنوی درمدت بستری (سه

[۱] $ratioOdds$
[۲] $Lydersen$

روز) انجام شد، که شامل امید افزایی معنوی یک مداخله محتوا محور در سه جلسه است و به ۴ محور اصلی می‌پردازد تا از این طریق امید بیماران را افزایش دهد. این چهار محور عبارت‌اند از: ارتباط باخدا، ارتباط با خود، ارتباط با دیگران و ارتباط با طبیعت است. یک ماه پس از ترخیص بیمار به‌منظور جمع‌آوری داده‌ها از چک‌لیست بررسی اختلال استرس پس از سانحه PCL [۳] استفاده شد.

در این مطالعه با متغیر دوحالتی روبرو هستیم که نتایج آن در جداول توافقی تنظیم می‌شود. یکی از مهم‌ترین شاخص‌هایی که در این نوع جداول برآورد می‌شود نسبت شانس است. برای یک جدول توافقی ۲×۲ با فراوانی مورد انتظار نسبت شانس به‌صورت مقابل بدست می‌آید.

$$\theta = \frac{p_1/(1-p_1)}{p_2/(1-p_2)} \tag{۱.۲}$$

و θ به‌صورت معادله (۱) برآورد می‌شود. برای برآورد دقیقتر نسبت شانس از فواصل اطمینان استفاده می کنیم. معیار بررسی عملکرد فواصل اطمینان، طول فاصله اطمینان است. هرچه طول کمتر باشد، فاصله اطمینان از دقت بالاتری برخوردار است. فواصل اطمینان OR در زیر اشاره شده است.

۱.۲ $logitWoolf$

فاصله اطمینان مجانبی ولف لوجیت ابتدا توسط ولف [۴] [۴] ارائه شد. این فاصله اطمینان بر اساس تقریب توزیع نرمال از $\log(\widehat{\theta})$ محاسبه شده و به صورت زیر تعریف می شود.

$$woolflogit = \log(\widehat{\theta}) \pm z_{(\alpha/2)}\sqrt{\frac{1}{n_{11}} + \frac{1}{n12} + \frac{1}{n21} + \frac{1}{n22}} \tag{۲.۲}$$

اگر یکی از سلول‌های جدول توافقی ۲×۲ صفر شود، انحراف معیار بی‌نهایت و فاصله اطمینان ناآگاهی بخش می‌شود و اطلاعات درستی نمی‌دهد.

۲.۲ $logitadjustedGart$

گارت [۵] [۵] با اضافه کردن ۰/۵ به هر سلول جدول توافقی ۲×۲، فاصله اطمینان زیر را ارائه کرد.

$$\tilde{n11} = n11 + ۰/۵,\ \tilde{n22} = n22 + ۰/۵,\ \tilde{n12} = n12 + ۰/۵,\ \tilde{n21} = n21 + ۰/۵$$

$$gartadjustedlogit = \log(\widehat{\theta}) \pm z_{(\alpha/2)}\sqrt{\frac{1}{\tilde{n_{11}}} + \frac{1}{\tilde{n_{12}}} + \frac{1}{\tilde{n_{21}}} + \frac{1}{\tilde{n_{22}}}} \qquad \tilde{\theta} = \frac{\tilde{n_{11}}\tilde{n_{22}}}{\tilde{n_{12}}\tilde{n_{21}}} \tag{۳.۲}$$

این فاصله اطمینان درزمانی که بیشتر سلول‌های جدول توافقی صفر هستند، بسیار مفید است. اگر $n_{11} = ۰$ و $n_{21} \neq ۰$ آنگاه $\widehat{\theta} = ۰$ و کران پایین فاصله اطمینان($L > ۰$) می شود و اگر $n_{11} = n_{1+}$ ، $n_{21} = n_{2+}$ یا اگر $n_{21} = ۰$ و $n_{11} \neq ۰$ آنگاه $\widehat{\theta} = \infty$ و کران بالای فاصله اطمینان تعریف نشده ($U = \infty$) است.

---

[۳] $ListCheckdiseaseheartIschemic$

[۴] $Woolf$

[۵] $Gart$

۳.۲ $logitIndependence - smooth$

اگرستی [۶] [۶] با اضافه کردن مقدار نامنفی $c_{ij} = \frac{۲n_{i+n+j}}{N^۲} \quad i,j = ۱,۲$ به هر سلول جدول توافقی ۲×۲، فاصله اطمینانی مشابه فاصله اطمینان گارت ارایه کرد. بنابراین در این فاصله اطمینان اگر $c = ۰$ و $c = ۰/۵$ باشد به ترتیب فاصله اطمینان گارت و ولف به دست می آید.

۴.۲ $conditionalexactCornfield$

اگر تعداد موفقیت‌ها ($n_{+۱}$)و تعداد شکست‌ها ($n_{+۲}$) همچنین تمام مجموع توأم حاشیه‌ای در جدول (۱) ثابت باشد احتمال مشاهدات در جدول توافقی با $x_{۱۱}$ موفقیت دارای توزیع فوق هندسی غیر مرکزی است. که تابع احتمال آن به صورت زیر است:

$$f(x_{۱۱}|\theta) = \frac{\binom{n_{۱+}}{x_{۱۱}}\binom{n_{۲+}}{n_{+۱}-x_{۱۱}}\theta^{x_{۱۱}}}{\sum_{i=n_{max}}^{n_{min}}\binom{n_{۱+}}{i}\binom{n_{۲+}}{n_{+۱}-i}\theta^{i}} \tag{۴.۲}$$

و زمانی که $n_{min} = min(n_{۱+}, n_{+۱}), \quad n_{max} = max(۰, n_{۱+} - n_{+۱})$ می باشد. کارن فیلد[۷] [۷] با معکوس کردن دو آزمون یک‌طرفه دقیق شرطی فاصله اطمینان شرطی دقیق (L,U) را معرفی کرد که (L,U) با حل روابط زیر می‌توان به دست می آیند.

$$\sum_{x_{۱۱}=n_{۱۱}}^{n_{min}} f(x_{۱۱}|L) = \alpha/۲ \qquad \sum_{x_{۱۱}=n_{max}}^{n_{۱۱}} f(x_{۱۱}|U) = \alpha/۲ \tag{۵.۲}$$

فاصله اطمینان دقیق کارن فیلد همیشه با آزمون دقیق $Fisher - Irwin$ مطابقت دارد [۶]. همچنین ، کارن فیلد فواصل تقریبی مجانبی را به فواصل شرطی دقیق ترجیح می‌دهد.

۵.۲ $intervalmid - pCornfield$

این فاصله اطمینان شبه دقیق [۸] (براساس $pointmid$ ساخته می شود) هستند [۸]؛ و با حل روابط زیر به دست می آید.

$$\sum_{x_{۱۱}=n_{۱۱}}^{n_{min}} f(x_{۱۱}|L_{cm}) - \frac{۱}{۲}f(n_{۱۱}|L_{cm}) = \alpha/۲. \tag{۶.۲}$$

$$\sum_{x_{۱۱}=n_{max}}^{n_{۱۱}} f(x_{۱۱}|U_{cm}) - \frac{۱}{۲}f(n_{۱۱}|U_{cm}) = \alpha/۲. \tag{۷.۲}$$

---

[۶] $Agresti$

[۷] $Cornfield$

[۸] $quasi - exact$

$conditionalexactBatista - Pike$ ۶.۲

باتیستا و پیک [۹] [۹] و [۱۰] بر اساس معکوس آزمون دوطرفه، فاصله اطمینان را پیشنهاد دادند که با حل روابط زیر به دست می آید:

$$\sum_{x_{11}=n_{max}}^{n_{min}} f(x_{11}|L).I(f(x_{11}|L) \leq f(n_{11}|L)) = \alpha \qquad (۸.۲)$$

$$\sum_{x_{11}=n_{max}}^{n_{min}} f(x_{11}|U).I(f(x_{11}|L) \leq f(n_{11}|U)) = \alpha \qquad (۹.۲)$$

$mid - pBatista - Pike$ ۷.۲

این فاصله اطمینان از نوع فاصله اطمینان شبه دقیق است[۱۱]، به‌صورت زیر محاسبه می‌شود.

$$\sum_{x_{11}=n_{max}}^{n_{min}} f(x_{11}|L).I(f(x_{11}|L) \leq f(n_{11}|L)) - \frac{1}{2}f(n_{11}|L) = \alpha \qquad (۱۰.۲)$$

$$\sum_{x_{11}=n_{max}}^{n_{min}} f(x_{11}|U).I(f(x_{11}|L) \leq f(n_{11}|U)) - \frac{1}{2}f(n_{11}|U) = \alpha \qquad (۱۱.۲)$$

$unconditionalexactAgresti - Min$ ۸.۲

اگرستی و مین [۱۲] بر اساس معکوس یک آزمون $score$ دوطرفه دقیق غیرشرطی فاصله اطمینان دقیق غیرشرطی را برای نسبت شانس پیشنهاد دادند:

$$T(n|\theta_0) = [n_{1+}(\widehat{p_1} - \tilde{p_1})]^2[\frac{1}{n_{1+}\tilde{p_1}(1-\tilde{p_1})} + \frac{1}{n_{2+}\tilde{p_2}(1-\tilde{p_2})}] \qquad (۱۲.۲)$$

که در آن،

$$\theta_0 = \frac{p_1/(1-p_1)}{p_2/(1-p_2)} \qquad (۱۳.۲)$$

می باشد سپس روابط زیر را ارائه می‌شود :

$$R(T(n)|p_1, \theta_0) = \sum_{|T(x)| \geqslant |T(n)|} f(x|p_1, \theta_0) \qquad (۱۴.۲)$$

$$R(T(n)|\theta_0) = sup_{p_1} R(T(n)|p_1, \theta_0) \qquad (۱۵.۲)$$

بنابراین با حل روابط زیر فاصله اطمینان اگرستی مین به دست می آید.

$$R(T(n)|L) = \alpha \qquad R(T(n)|U) = \alpha \qquad (۱۶.۲)$$

اگر $\theta_0 < L$ در نتیجه $R(T(n)|\theta_0) < \alpha$ و در صورتی که $\theta_0 > U$ می توان نتیجه گرفت $R(T(n)|\theta_0 > U$

[۹] $Batista - Pike$

۳ نتایج

دراین پژوهش درمجموع ۸۰ نفر شرکت داشتند و افراد به طور تصادفی به دو گروه کنترل و مداخله تخصیص داده شدند. با توجه به جدول (۲) از ۴۰ نفر تحت گروه مداخله ۷ نفر دارای $PTSD > ۵۰$ و از ۴۰ نفردر گروه کنترل ۲۶ نفردارای $PTSD > ۵۰$ بودند. در نتیجه در گروه مداخله ۶۵درصد و در گروه کنترل ۱۷/۵ درصد افراد دارای اختلال استرس پس از سانحه بودند($PTSD > ۵۰$).

جدول ۲: نتایج اختلال استرس در بیماران سکته قلبی حاد

	$PTSD$		
مراقبت معنوی	بیشتر از ۵۰	کمتر از ۵۰	جمع
	درصد	درصد	درصد
مداخله	۷	۳۳	۴۰
	(۱۷/۵)	(۸۲/۵)	(۵۸/۸)
کنترل	۲۶	۱۴	۴۰
	(۶۵)	(۳۵)	(۴۱/۳)
			۸۰
$p - value = ۰/۰۰۰$		$\chi^۲ = ۱۸/۶۲$	

با توجه به اینکه مقدار $OR = ۰/۱۱$ می باشد می‌توان نتیجه گرفت شانس استرس پس از سانحه در بیماران سکته قلبی حاد در گروه کنترل ۹/۱ برابر گروه مراقبت معنوی است ($p < ۰/۰۵$).

با توجه به نتایج جدول(۳) می‌توان نتیجه گرفت در بین سه فواصل اطمینان مجانبی، گارت با مقدار ۲/۰۳۶ طول کمتر از ولف لوجیت و مستقل هموار است پس دقت بالاتری دارد. از میان پنج فاصله اطمینان دقیق طول فاصله اطمینان باپتیستا-پیک دقیق برابر با مقدار ۱/۹۷۳ است چون طول کمتر، دارای دقت بیشتری می باشد. و بیشترین طول متعلق به فاصله اطمینان کارن فیلد دقیق برابر با مقدار ۲/۳۴۱ است. بنابراین این فاصله اطمینان دقت کمتری نسبت به سایر فواصل اطمینان دارد.

جدول ۳: فواصل اطمینان برای OR در بیماران سکته قلبی حاد.

	فاصله اطمینان		
	کران پایین	کران بالا	طول
ولف لوجیت	۳/۰۹	۲۴/۸۴	۲/۰۸۴
گارت	۲/۹۵	۲۲/۶۰	۲/۰۳۶
مستقل هموار	۲/۹۶	۲۲/۷۵	۲/۰۳۹
کارن فیلد دقیق	۲/۷۹	۲۸/۹۸	۲/۳۴۱
کارن فیلد مید-پی	۳/۰۶	۲۵/۶۳	۲/۱۲۵
باپیستا-پیک دقیق	۲/۸۵	۲۵/۳۹	۲/۱۸۷
باپیستا-پیک مید-پی	۳/۱۷	۲۲/۸۰	۱/۹۷۳
اگرستی-مین	۰/۰۳	۰/۳۱	۲/۳۳۵

۴ بحث

در این پژوهش با توجه به اینکه حجم نمونه متوسط است. با مقایسه فواصل اطمینان مجانبی، طول فاصله اطمینان گارت برابر ۲/۰۳۶ نسبت به فواصل اطمینان ولف لوجیت و هموار مستقل کمتر است. لذا در بین فواصل اطمینان مجانبی نسبت شانس، فاصله اطمینان گارت از دقت بالاتری برخوردار است. در بین فواصل اطمینان دقیق طول فاصله اطمینان باپتیستا–پیک مید–پی برابر ۱/۹۷۳ است. بنابراین دقت این فاصله اطمینان بیشتر از تمام فواصل اطمینان نسبت شانس است.

در مطالعاتی که اگرستی و مین در سال ۲۰۰۲ [۱۲] و فیگرلند [۱۰] و همکاران در سال ۲۰۱۱ [۱۳] انجام دادند. فواصل اطمینان نسبت شانس ازنظر پوشش احتمالی موردبررسی قرار گرفتند. این مطالعات برای بررسی حجم نمونه‌های کوچک و بزرگ انجام شد. نتایج مطالعات نشان داد، که سه فاصله اطمینان مجانبی نسبت شانس عبارت‌اند از: ولف لوجیت، گارت و مستقل هموار عملکرد مشابهی در حجم نمونه کوچک و متوسط دارند. برای نمونه کوچک فاصله اطمینان گارت پوشش احتمالی محافظه‌کاری نسبت به فاصله اطمینان ولف لوجیت و هموار مستقل دارد و هر سه فاصله اطمینان برای نسبت نزدیک به صفر و یک محافظه‌کار عمل می‌کند. هنگامی‌که حجم نمونه کاهش می‌یابد هر سه فاصله اطمینان پوشش احتمالی مشابهی دارند. با افزایش نسبت شانس هر سه فاصله اطمینان واگرا می‌شوند و فاصله اطمینان گارت بین این سه فاصله اطمینان بهتر است. همچنین این فاصله اطمینان از دو فاصله اطمینان ولف لوجیت و هموار مستقل کوتاه‌تر است.

در این مطالعات بررسی پوشش احتمالی برای فواصل اطمینان دقیق، شامل کارن فیلد دقیق و کارن فیلد مید–پی و باپتیستا–پیک دقیق و باپتیستا–پیک مید–پی و اگرستی–مین نیز انجام شده، برای نمونه‌های کوچک هر دو فاصله اطمینان کارن فلید بسیار محافظه‌کار عمل می‌کنند و با افزایش حجم نمونه تغییر نمی‌کنند. فاصله باپتیستا–پیک دقیق محافظه‌کاری ویژه برای حجم نمونه‌های کم در بین فواصل اطمینان دقیق دارد و فواصل اطمینان اگرستی–مین کمتر محافظه‌کار است. لذا بهترین فاصله اطمینان دقیق همان باپتیستا–پیک مید–پی است.

در مطالعه‌ای دیگر که در سال ۲۰۱۴ [۱۴] تحت عنوان مقایسه فواصل اطمینان دو نسبت مستقل دوحالتی با استفاده از روش گرافیکی و میانگین متحرک توسط لود و دانه [۱۱] انجام شد. از روش گرافیکی و حجم نمونه‌های کوچک و بزرگ استفاده شد. به‌منظور، سادگی محاسبات بدون نرم‌افزار، روش برون لی جفری [۱۲] [۱۵] پیشنهادشده است. در اکثر موارد پوشش احتمال یک‌طرفه پیشنهاد می‌شود. روش گارت–نم [۱۳] [۱۶] تنها استثنا در روش تصحیح چولگی است. با توجه به تصحیح منظم که شامل یک خطای پیوستگی برای بهبود عملکرد نمونه‌های کوچک و انتخاب یک تصحیح پیوستگی برای بررسی پوشش احتمالی محافظه‌کار است. با استفاده از میانگین وزن‌دار شده یک روش ترکیبی جدید به دست می‌آید که تقریباً با گارت–نم عملکرد یکسانی دارد.

۵ نتیجه گیری

به‌طورکلی فاصله اطمینان باپتیستا–پیک مید–پی برای نسبت شانس عملکرد بهتری نسبت به سایر فواصل اطمینان دقیق و مجانبی دارد و این نتیجه با یافته‌های فیگرلند و همکارانش مطابقت دارد.

---

[۱۰] $Fagerland$
[۱۱] $DaneandLaud$
[۱۲] $JeffreysBrown - Li$
[۱۳] $Gart - Nam$

مراجع

[114] JJ. Deeks and DG. Altman, *Effect measures for metaanalysis of trials with binary outcomes.*, In: M. Egger , G. Davey Smith and DG. Altman (eds) *Systematic reviews in health care: meta-analysis in context* 2nd edn., BMJ Publishing Group, London, 2001, pp. 313–335.

[115] M. Szumilas, Explaining odds ratios, *Journal of the Canadian academy of child and adolescent psychiatry.*, 19 (2010), 227.

[116] S. Lydersen ,MW. Fagerland and P. Laake , Tutorial in biostatistics: recommended tests for association in 2×2 tables, *Stat Med.*, 28 (2009), 1159–1175.

[117] B. Woolf , On estimating the relation between blood group and disease, *Ann Human Gene.*, 19 (1955), 251–253.

[118] JJ. Gart , Alternative analyses of contingency tables, *J R Stat Soc B Stat Meth.*, 28 (1966), 164–179.

[119] A. Agresti , On logit confidence intervals for the odds ratio with small samples, *Biometrics.*, 55 (1999), 597–602.

[120] J. Cornfield , A statistical problem arising from retrospective studies, *Pro Third Berkeley Sym Math Stat Probab.*, 4 (1956), 135–148.

[121] KF. Hirji , S-J. Tan and RM. Elashoff , A quasi-exact test for comparing two binomial proportions, *Stat Med.*, 10 (1991), 1137–1153.

[122] J. Baptista and MC. Pike , Exact two-sided confidence limits for the odds ratio in a 2×2 table, *J R Stat Soc C Appl Stat.*, 26 (1977), 214–220.

[123] TE. Sterne , Some remarks on confidence or fiducial limits, *Biometrika.*, 41 (1954), 275–278.

[124] MW. Fagerland , Exacten and mid-p confidence intervals for the odds ratio, *Stata Journal.*, 12 (2012), 505.

[125] A. Agresti and Y. Min , Unconditional small-sample confidence intervals for the odds ratio, *Biostatistics.*, 3 (2002), 379–386.

[126] MW. Fagerland, S. Lydersen, and P. Laake, Recommended confidence intervals for two independent binomial proportions,*Statistical methods in medical research.*, 24 (2015), 224–254.

[127] PJ. Laud, and A. Dane, Confidence intervals for the difference between independent binomial proportions: comparison using a graphical approach and moving averages, *Pharmaceutical statistics.*, 13 (2014), 294–308.

[128] L. Brown, and Li. Xuefeng, Confidence intervals for two sample binomial distribution, *Journal of Statistical Planning and infrence.*, 130 (2005), 359–375.

[129] JJ. Gart, and JM. Nam, Approximate interval estimation of the ratio of binomial parameters: a review and corrections for skewness, *Biometrics.*, 44 (1988), 323–338.

برآورد طول دوره کمون در بیماران مبتلا به HIV در استان خراسان رضوی

نجمه نخعی راد [۱]، **گروه ریاضی وآمار، دانشگاه آزاد اسلامی واحد مشهد وحید فکور گروه آمار، دانشگاه فردوسی مشهد**

چکیده: ددر این مقاله به برآورد طول دوره کمون در بیماران مبتلا به HIV در استان خراسان رضوی پرداخته شده است . دوره کمون یا دوره نهفتگی ویروس HIV به مدت زمان از ورود ویروس به بدن تا آشکار شدن ایدز اطلاق می‌شود. در ابتدا طول دوره کمون در حالت ناپارامتری با استفاده از جدول طول عمر و برآوردگر کاپلان مه- یر برآورد شده است. سپس با استفاده از یک مدل پارامتری پیشنهادی برآورد پارامتری طول دوره کمون به دست آورده شده است. و در انتها عوامل مؤثر بر طول دوره کمون توسط روش رگرسیونی کاکس شناسایی شده اند.

واژه‌های کلیدی: سانسور راست، جدول طول عمر، برآوردگر کاپلان مه-یر، رگرسیون کاکس.

۱ مقدمه

در سال ۱۹۸۱، در نیویورک، با مشاهده چندین بیماری وخیم شبیه سرطان و دیده شدن انواع متفاوت دیگری از عفونت‌های ناشناخته در ایالت متحده امریکا، توجه دانشمندان به ویروس مشترکی جلب شد که با تغییر شکل مداوم، سیستم ایمنی بدن را از کار می‌اندازد و در نتیجه امکان ابتلا به انواع بیماری های خطرناک دیگر را فراهم می‌کند. اولین مواردی که مشاهده شد در بین معتادان تزریقی و همچنین همجنس‌گرایان مرد قرار داشت که به دلیل نامعلومی سیستم دفاعی بدن آن‌ها ضعیف شده بود و علایم بیماری التهاب ریه پنوموسیستیک کارینی (Pneumocystis Carinii) PC در آن‌ها مشاهده می‌شد. سپس نوعی سرطان پوست نادر با نام کاپوسی سارکوما (Kaposi's Sarcoma) KS در میان مردان همجنس‌گرا گزارش شد. گزارش موارد بیشتری از بیماری PC و KS زنگ خطری برای مرکز پیشگیری و کنترل بیماری بود. در آن اوایل مرکز کنترل بیماری هنوز نامی رسمی برای این بیماری انتخاب نکرده بود و معمولاً نام این بیماری را با بیماری که مریض با آن در ارتباط بود بیان می‌کردند. واژه ایدز در همایشی در ژوئیه ۱۹۸۲ معرفی شد و از سپتامبر ۱۹۸۲، مرکز کنترل بیماری از واژه ایدز برای نسبت دادن این بیماری استفاده کرد. همزمان در آفریقا نیز موارد متعددی از شیوع این علائم دیده شد که منجر به مرگ و میر فراوانی در این قاره گردید. دانشمندان بر این عقیده اند که شکل ابتدایی این ویروس در شامپانزه ها وجود داشته است که به دلیل تغذیه ساکنان گینه بیسائو و قصابی این حیوانات از طریق زخم های انسانی به انسان منتقل

[۱] نجمه نخعی راد: najme.rad@gmail.com

شده و در طول یک قرن تکامل یافته و به دیگران انتقال یافته است. راه های عمده انتقال این ویروس عبارتند از: استفاده از سوزن و سرنگ مشترک (تزریق مواد مخدر با سرنگ مشترک، خالکوبی، وسایل آلوده کار به خصوص در بیمارستان ها و دندانپزشکی ها و ...)، رفتارهای جنسی پرخطر، خون و فرآورده های خونی و از مادر مبتلا به جنین. (بنت و همکاران، ۲۰۱۴)

HIV (Human Immune-deficiency Virus) یا ویروس نقص ایمنی انسانی، ویروسی است که سبب صدمه به دستگاه ایمنی بدن می‌شود و هنگامی که فرد مبتلا به HIV شده یا اصطلاحاً HIV مثبت می‌شود، ویروس به تدریج شروع به از بین بردن سیستم ایمنی بدن وی کرده به طوریکه بدن نمی‌تواند حتی با بیماری های جزئی مقابله کند و مستعد ابتلا به انواع بیماری ها می‌شود. ویروس HIV در دو نوع HIV۱ و HIV۲ است که HIV۲ نادرتر و کم خطرتر می‌باشد. با پیشرفت ویروس و درگیری سلول های ایمنی موجود در خون (لنفوسیت ها) و همچنین سلول های ایمنی موجود در بافت ها مانند مغز استخوان، طحال، کبد و گره های لنفاوی این سلول ها دیگر توانائی تولید پادتن برای مقابله با بیماری ها را نداشته و فرد مبتلا به بیماری هایی نظیر انواع سرطان ها، توکسوپلاسمای مغزی و ... می‌گردد.

یک چنین وضعیتی ایدز AIDS (Acquired Immune Deficiency Syndrome) نامیده می‌شود که همراه با پیدایش علائم و نشانه های گوناگون در بدن می‌باشد. برخی ازنشانه ها، همانطور که پیش تر گفته شد در سال ۱۹۸۱ در افراد بالغ جوانی دیده شد که دچار نقص ایمنی مادرزادی نبودند و این تعجب پزشکان را برانگیخت و به همین دلیل مجموعه علائم را نشانه های نقص ایمنی اکتسابی یا ایدز نام نهادند چون نمی‌توانستند آن را به بیماری مشخصی نسبت دهند و هنوز عامل آن ناشناخته بود. دوره کمون یا انکوباسیون در لغت به معنای تکوین بیماری عفونی از هنگام ورود پاتوژن تا پیدایش نشانه های بالینی می‌باشد. طبق همین تعریف، دوره کمون در بیماری ایدز از هنگام ورود ویروس HIV به بدن فرد تا پیدایش علائم ایدز ادامه دارد. این مدت زمان معمولاً سال ها به طول می‌انجامد و در طول این مدت فرد ناقل بیماری بوده بدون اینکه علائمی از بیماری را نشان دهد. شروع بیماری ایدز از دو طریق قابل تشخیص است:

۱. شمارش سلول های Tcd۴ موجود در هر میلی متر خون: هر گاه تعداد این سلول ها به کمتر از ۲۰۰ عدد در هر میلی متر برسد اصطلاحاً بیمار وارد مرحله ایدز شده است.

۲. از طریق علائم و نشانگان: گاه ممکن است فرد تعداد سلول های Tcd۴ بیش از ۲۰۰ داشته باشد ولی علائم و بیماری های از قبیل: قارچ دهانی، توکسوپلاسمای مغزی، سرطان پوست، PC ، و یا پلاکت های مقاوم به درمان در وی مشاهده گردد، در این حالت نیز بیمار وارد مرحله ایدز شده است. (بنت و همکاران، ۲۰۱۴)

۲ اطلاعاتی درباره تحقیق:

اهداف تحقیق:

همانطور که پیش تر گفته شد فرد مبتلا به HIV در طول دوره کمون ناقل این ویروس بوده در صورتی که به هیچ طریقی به جز آزمایش خون قابل تشخیص نمی‌باشد. این افراد که معمولاً دارای رفتارهای پرخطر نیز می‌باشند با ظاهری سالم باعث انتقال ویروس به سایرین می‌گردند. لذا برآورد طول دوره کمون این بیماری دارای اهمیت می‌باشد.

هدف کلی:

برآورد طول دوره کمون در بیماران مبتلا به HIV در خراسان رضوی.

اهداف فرعی:

۱. برآورد طول دوره کمون به روش ناپارامتری (جدول طول عمر و برآوردگر کاپلان مه–یر)

۲. برآورد طول دوره کمون براساس یک مدل پارامتری پیشنهادی

۳. یافتن عوامل مؤثر بر طول دوره کمون توسط روش رگرسیونی کاکس

نوع مطالعه:

از آنجایی که در حین اجرای تحقیق دخالتی در وضعیت موجود افراد صورت نپذیرفته، مطالعه به طور کلی یک مطالعه غیر تجربی یا مشاهده ای می‌باشد و طبق تعریف مطالعه مقطعی، مطالعه ای از این نوع نیز محسوب می‌گردد. در مطالعه مقطعی، تمام جامعه یا قسمتی از آن انتخاب شده و آزمایشگر داده های مورد نظرش را گردآوری می‌کند. چون اطلاعات جمع آوری شده، فقط آنچه را که در یک نقطه زمانی درباره ی پارامترهای جامعه وجود دارد نشان می دهد، به این نوع مطالعه مقطعی می‌گویند. همچنین این مطالعه با توجه به برخی سؤال ها که از سوابق افراد شده است، مشمول مطالعه گذشته نگر نیز می‌شود.

جامعه آماری:

جامعه آماری، کلیه بیماران مبتلا به HIV در استان خراسان رضوی را شامل می‌شود.

واحد آماری:

واحد آماری، فرد بیمار مبتلا به HIV می‌باشد.

روش نمونه گیری و حجم نمونه:

با توجه به مشخص نبودن چارچوب دقیق جامعه آماری از جمله حجم جامعه مورد بررسی یعنی تعداد افراد مبتلا به HIV در استان خراسان رضوی و محدودیت شناسایی و معرفی این افراد، نمونه مورد بررسی شامل کلیه افرادی است که به مرکز مشاوره بیماری های عفونی رفتاری وابسته به اداره بهداشت خراسان رضوی مراجعه کرده و یا توسط این مرکز و مراکزی از قبیل مراکز درمانی، انتقال خون، بهزیستی و زندان های استان شناسایی شده اند و در حال حاضر تحت پوشش این مرکز هستند.

روش و ابزار گردآوری اطلاعات :

برای جمع آوری اطلاعات از پرسش نامه استفاده شد که طبق اطلاعات موجود در پرونده های بیماران در مرکز بیماری های عفونی رفتاری تکمیل گردید و در مواردی که اطلاعات درج شده در پرونده ناقص بود از پزشک و افراد مشغول به کار در مرکز و همچنین بیماران نیز سؤالاتی پرسیده شد.

محدودیت های در تحقیق:

۱. حجم کم نمونه به دلیل شناسایی مشکل بیماران مبتلا به HIV و گاه در عین شناسایی عدم تمایل فرد به ادامه مراجعه به مرکز

۲. تعداد مشاهدات سانسور شده ی زیاد در مطالعه به دلیل نوپا بودن مرکز و فعالیت شناسایی بیماران مبتلا از طریق پایگاه های دیده وری

۳. اطلاعات ثبت نشده در پرونده بیماران

پیش فرض ها در انجام تحقیق:

۱. از آنجائیکه زمان دقیق آلوده شدن فرد به ویروس HIV به طور دقیق قابل تشخیص نیست و ویروس HIV تنها از راه های گفته شده در قبل قابل انتقال است لذا از هر فرد با توجه به نوع رفتار پرخطر یا زمان اولین تزریق مشترک پرسیده شده و تاریخ آلوده شدن فرد به ویروس HIV نقطه وسط بازه‌ی زمان اولین رفتار پرخطر و زمان تشخیص ابتلا به HIV در نظر گرفته شده است.

۲. اطلاعات موجود در پرونده بیماران قابل اطمینان بوده و به طور دقیق ثبت شده اند.

۳. در طول دوره کمون هیچ دارویی که باعث تأثیر بر طول دوره گردد، مصرف نمی‌شود.

۴. بیمار آلوده به ویروس HIV حتماً روزی وارد مرحله ایدز می‌شود.

مدل تعریف شده برای داده ها:

مطالعه از سال ۱۳۷۵ آغاز شده و تا ۱۳۸۷ ادامه یافته است. در طول این مدت افراد در زمان های مختلف با مراجعه به مرکز و یا شناسایی، وارد مطالعه شده اما زمان پایان مطالعه برای تمام آن ها یکسان است. بیماران از زمان ورود به مطالعه تا شروع ایدز یا سانسور پیگیری شده اند.

متغیرهای تحقیق:

متغیر وابسته: طول دوره کمون به ماه متغیر وابسته است که متغیری کمی و مقیاسی است.

متغیرهای مستقل: متغیرهای مستقل بر اساس اطلاعات موجود در پرونده های بیماران در نظر گرفته شده اند چون به دلیل سختی دسترسی به بیماران (مثلاً به دلیل عدم آگاهی برخی خانواده ها امکان تماس تلفنی وجود نداشت، گذشته از اینکه اسم و آدرس این افراد محرمانه نیز می‌باشد و همچنین چون نوبت آزمایشات برای بیماران HIV مثبت، سالیانه و برای بیماران ایدزی هر سه ماه یکبار است ملاقات حضوری با تعداد زیادی از بیماران امکان پذیر نبود) امکان تعریف متغیرهای دیگر وجود نداشت. متغیرهای تعریف شده عبارتند از: جنسیت، سن، وضعیت تأهل، اعتیاد به مواد مخدر، اعتیاد به سیگار، میزان تحصیلات، شغل، درآمد، سابقه حضور در زندان، عامل ایجاد کننده بیماری، نوع ویروس، نحوه پیگیری، ابتلا به بیماری های هپاتیت $C(HCV)$ ، هپاتیت $B(HBS)$ ، سل (TB)، $VDRL$ و ویروس های CMV، $HTLV$۱ و $HTLV$۲.

روش های تجزیه و تحلیل:

۱. جدول طول عمر

۲. برآوردگر کاپلان مه-یر

۳. رگرسیون کاکس

۴. استفاده از یک مدل پارامتری پیشنهادی برای برآورد زمان تا ایدز در داده های سانسور شده از راست

تجزیه و تحلیل با نرم افزارهای R و $SPSS$ انجام شده است.

۳ نتایج تحقیق:

۱.۳ اطلاعات توصیفی:

از مجموع ۶۴ فرد حاضر در مطالعه، ۵۳ نفر مرد و ۱۱ نفر زن هستند. میانگین سن این افراد ۳۶/۱۶ و انحراف معیار سن ۸/۴۷ سال می‌باشد. ۲۱ نفر متاهل و مابقی مجرد، مطلقه و بیوه هستند. ۸۷/۵ درصد افراد تحصیلات زیر دیپلم داشته و ۸۴/۴ درصد سابقه حضور در زندان را داشته اند. ۷۰ درصد افراد اعتیاد به سیگار و ۱۷/۲ درصد اعتیاد به مواد مخدر داشته و ۷۰/۴ درصد متادون مصرف کرده یا در ترک کامل هستند. ۸۰ درصد مبتلایان سابقه استفاده از سوزن و سرنگ مشترک، ۱۱ درصد رابطه جنسی پر خطر و ۹ درصد تجربه هر دو مورد را داشته اند. ۶۲ نفر از افراد آلوده به ویروس HIV۱ و ۲ نفر آلوده به ویروس HIV۲ هستند. تعداد مبتلایان به هر یک از بیماری‌های هپاتیت $C(HCV)$ ، هپاتیت $B(HBS)$ ، سل (TB)، $VDRL$ و ویروس‌های CMV، $HTLV$۱ و $HTLV$۲. در جدول زیر آمده است:

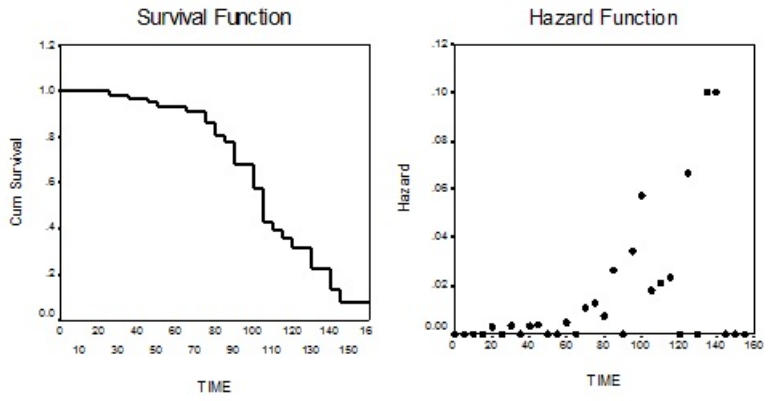
بیماری	HCV	HBS	TB	$VDRL$	CMV	$HTLV۱$	$HTLV۲$
دارد	۳۹	۹	۱۲	۰	۶	۵	۲
ندارد	۲۵	۵۵	۵۲	۶۴	۵۸	۵۹	۶۲

۲.۳ برآورد ناپارامتری طول دوره کمون:

همانطور که گفته شد مطالعه از سال ۱۳۷۵ آغاز شده و تا ۱۳۸۷ ادامه یافته است. در طول این مدت افراد در زمان های مختلف با مراجعه به مرکز و یا شناسایی وارد مطالعه شده اما زمان پایان مطالعه برای تمام آن ها یکسان است. بیماران از زمان ورود به مطالعه تا شروع ایدز یا سانسور پیگیری شده اند. به این ترتیب از مجموع ۶۴ فرد آلوده به ویروس HIV از آنجا که ۳۶ نفر تا پایان مطالعه پیشامد ایدز را تجربه نکرده اند سانسور می‌شوند. برای برآورد ناپارامتری طول دوره کمون از جدول طول عمر و برآوردگر کاپلان مه–یر استفاده کرده ایم:

جدول طول عمر:

زمان‌های تشخیص ایدز را ۵ ماهه در نظر گرفته و احتمالات ۵ ماهه را محاسبه کرده ایم. نتایج نشان می دهد که میانه طول دوره کمون برای مشاهدات ۱۰۲/۵۸ ماه است. نمودار تابع بقا تجمعی و تابع مخاطره به دست آمده توسط این روش در شکل ۱ آمده است.



شکل ۱: نمودار تابع بقا و تابع مخاطره به دست آمده توسط جدول طول عمر

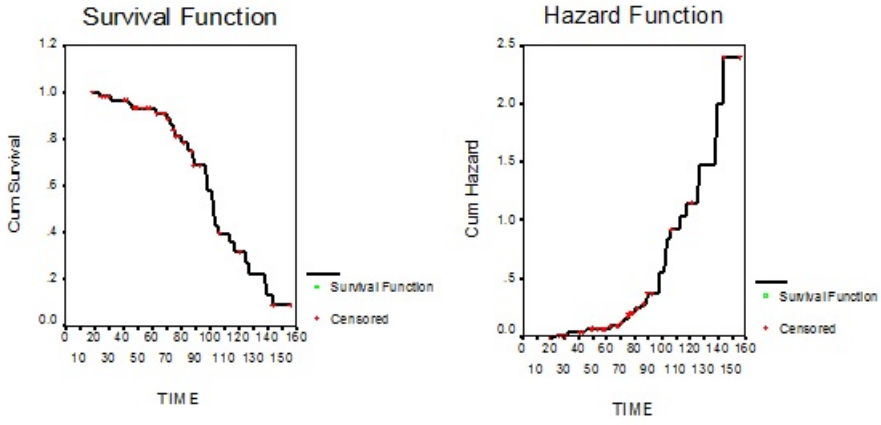
برآوردگر کاپلان مه–یر: در این روش میانگین و میانه طول دوره کمون به ترتیب ۱۰۵ و ۱۰۲ ماه به دست آمده است (جدول ۱). نمودار تابع بقا تجمعی و تابع مخاطره به دست آمده توسط این روش در شکل ۲ آمده است.

جدول ۱: میانگین و میانه طول دوره کمون در روش کاپلان مه–یر

	Survival Time	Confidence Interval ٪۹۵	Standard Error
Mean	۱۰۵	(۹۵, ۱۱۵)	۵
Median	۱۰۲	(۹۵, ۱۰۹)	۴

۳.۳ برآورد پارامتری طول دوره کمون با استفاده از یک مدل پیشنهادی:

y را یک متغیر برنولی تعریف می‌کنیم. به طوریکه $y = ۱$ نشان دهنده‌ی وقوع ایدز و $y = ۰$ بیانگر عدم رسیدن به مرحله ایدز باشد. فرض کنید T زمان تشخیص ایدز و X بردار متغیرهای توضیحی باشد که می‌تواند برای هر فرد تعریف شود. در این مدل احتمال وقوع



شکل ۲: نمودار تابع بقا و تابع مخاطره به دست آمده توسط روش کاپلان مه–یر

ایدز پس از آلوده شدن با ویروس HIV را برابر نسبت افراد آلوده در نمونه که به مرحله ایدز رسیده اند به کل افراد در نمونه در نظر گرفته و توزیع زمان تا پدیدار شدن ایدز را

$$f(t|y = ۱, X) = \lambda\delta t^{\delta-۱} exp(-\lambda t^{\delta}),$$

فرض می‌کنیم که $\lambda, \delta > ۰$.

اگر ایدز در زمان t_i برای نفرiام مشاهده شود در این صورت سهم درستنمایی برای این شخص برابر است با

$$p(y_i = ۱) f(t_i|y = ۱, X)$$

و اگر نفر i ام تا زمان t_i بدون مشاهده ایدز پیگیری شود سهم درستنمایی وی

$$p(y_i = ۰|X) + pr(y_i = ۱|X) \int_{t_۱}^{\infty} f(u|y = ۱, X)du$$

می‌باشد. این مدل در سال ۱۹۸۸ توسط لویی و همکاران هنگامی‌که t معلوم یا سانسور شده باشد بحث شده است. برآورد درستنمایی ماکسیمم پارامترهای مدل در جدول ۲ آمده است ($-۲logL = ۳۵۹\square۸۳۳$) . میانگین طول دوره کمون با استفاده از این مدل ۰۶/۱۰۰ ماه بدست آمده است که با میانگین بدست آمده در روش های ناپارامتری تفاوت چندانی ندارد.

جدول ۲: برآورد درستنمایی ماکسیمم پارامترهای مدل

Coefficients	Estimate	std. error
delta	۱/۲۵۹۷۴۲۸۴۶	۰/۰۷۵۲۸۰۹۰
lambda	۰/۰۰۲۷۵۶۳۳۹	۰/۰۰۰۸۰۶۵۳

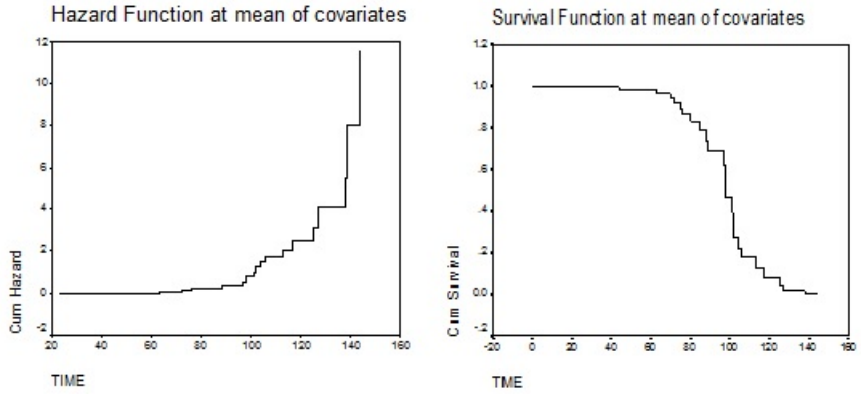
۴.۳ یافتن عوامل مؤثر بر طول دوره کمون توسط روش رگرسیونی کاکس :

رگرسیون کاکس در ۱۲ مرحله تکرار شده است به این ترتیب که ابتدا تمام متغیرهای توضیحی وارد مدل شده، سپس در طی ۱۲ مرحله از مدل خارج می‌شوند و در مرحله آخر با انتخاب خطای ۰/۰۵ متغیرهای جنسیت، اعتیاد به سیگار، اعتیاد به مواد مخدر (اعتیاد و مصرف متادون)، تحصیلات، راه انتقال (استفاده از سوزن و سرنگ مشترک)، ابتلا به هپاتیت C و ویروس $HTLV$۱ با ضرایب آمده در جدول زیر در مدل حضور دارند. نمودار تابع بقا تجمعی و تابع مخاطره به دست آمده توسط این روش در شکل ۳ آمده است.

Variables in the Equation

		B	SE	Wald	df	Sig.	Exp(B)
					a		
Step 12	SEX	3.587	1.411	6.466	1	.011	36.126
	AGE	-.075	.044	2.910	1	.088	.928
	SMOKING	-1.297	.507	6.556	1	.010	.273
	ADICTION			8.110	3	.044	
	ADICTION(1)	-1.323	1.281	1.067	1	.302	.266
	ADICTION(2)	-6.546	2.570	6.488	1	.011	.001
	ADICTION(3)	-2.137	1.181	3.274	1	.070	.118
	EDUCAT	.586	.267	4.815	1	.028	1.797
	FACTOR			14.634	2	.001	
	FACTOR(1)	1.649	2.049	.648	1	.421	5.204
	FACTOR(2)	5.431	1.641	10.958	1	.001	228.319
	HCV	-2.103	.774	7.381	1	.007	.122
	HTLV1	3.009	.772	15.179	1	.000	20.268

جدول ۳: متغیرهای موجود در مدل رگرسیون کاکس در مرحله آخر



شکل ۳: نمودار تابع بقا و تابع مخاطره به دست آمده توسط مدل رگرسیون کاکس

مراجع

[130] Bennett, J. E., Dolin, R., & Blaser, M. J. (2014). *Principles and practice of infectious diseases (Vol. 1)*. Elsevier Health Sciences.

[131] Lui, K.- J., Darraw, W. W., Rutherford, G. W., (1988). *A model – based estimate of the mean incubation period for AIDS in homosexual men.* Science 240, 1333-1335

خواص مجانبی برآوردگرهای کاپلان مه-یر و نلسون-آلن برای داده های سانسور شده ی وابسته

نجمه نخعی راد [1]، گروه ریاضی و آمار، دانشگاه آزاد اسلامی واحد مشهد وحید فکور گروه آمار، دانشگاه فردوسی مشهد

چکیده: در مطالعات بقا، برآورد تابع بقا که در توصیف مدت زمان تا وقوع پیشامدی خاص به کار می‌رود و همچنین تابع تجمعی خطر از اهمیت بسیاری برخوردار می‌باشد. در حالتی که زمان های بقا مستقل از یکدیگرند، پیشتر برآوردگرهای این توابع معرفی و خواص مجانبی آن ها بررسی شده است. در این مقاله برقراری خواص مجانبی برآوردگرهای کاپلان-مه یر و نلسون-آلن تابع تجمعی خطر برای همان برآوردگرهای پیشین، در صورت وابستگی زمان های بقا، در مدل سانسور راست نشان داده شده است که شامل سازگاری قوی و همگرایی ضعیف می‌باشد. پیش از بیان همگرایی ضعیف، ابتدا نشان می دهیم که تحت شرایط نظم می توان این برآوردگرها را به صورت میانگین متغیرهای تصادفی به اضافه یک جمله باقیمانده نمایش داد و در ادامه از آن در اثبات همگرایی ضعیف استفاده می‌کنیم.

واژه‌های کلیدی: α-آمیزنده، داده های سانسور شده ی وابسته، برآوردگر کاپلان مه-یر، برآوردگر نلسون-آلن، سازگاری قوی، همگرایی ضعیف

۱ مقدمه

تحلیل بقا اصطلاحی است که برای بیان تحلیل داده های مربوط به زمان، از یک شروع خوش تعریف تا وقوع یک یا چند پیشامد مشخص یا نقطه پایان به کار می رود و در علوم کاربردی نظیر پزشکی ، زیست شناسی، اپیدمی‌شناسی، مهندسی، اقتصاد و جمعیت شناسی مطرح است. در تحلیل داده های بقا به دو دلیل عمده نمی توان از روش های استاندارد تحلیل آماری استفاده نمود. دلیل اول اینکه داده های بقا به طور نرمال توزیع نشده اند و اگثر توزیع ها چوله به راست هستند. در نتیجه پذیرفتن فرض نرمال بودن این قبیل داده ها معقول نمی‌باشد که البته این مشکل را می توان با تبدیل داده ها برای رسیدن به توزیع متقارن تر برطرف نمود. هر چند که رویکرد بهتر، قبول یک مدل توزیعی جایگزین برای داده ها می‌باشد. دلیل دوم اینکه این داده ها به صورت هایی ظاهر می شوند که مشکلاتی را در تحلیل آن ها ایجاد می‌کند. از آن جمله وجود مشاهدات سانسور شده در بین آن ها است . مشاهدات سانسور شده زمانی رخ می دهند که وضعیت بقای واحدهای مورد مطالعه در رابطه با وقوع یا عدم وقوع پیشامدی خاص در دوره زمانی مشخصی، معلوم باشد. سانسور را می توان به انواع سانسور راست، چپ و فاصله ای تقسیم بندی کرد. سانسور از راست که در این مقاله از آن صحبت شده است در سه

[1] نجمه نخعی راد: najme.rad@gmail.com

نوع کلی سانسور نوع I، سانسور نوع II و سانسور نوع III یا سانسور تصادفی می‌باشد:

سانسور نوع I : در سانسور نوع I پیشامد مطلوب تنها وقتی قابل مشاهده است که قبل از یک زمان از پیش تعیین شده رخ دهد. مطالعه ای را در نظر بگیرید که با تعداد ثابتی از اعضا (بیماران، حیوانات...) در زمانی مشخص آغاز می‌شود. به لحاظ صرفه جویی در زمان و هزینه مطالعه باید پیش از اینکه تمام اعضا به پیشامد مطلوب رسیده باشند خاتمه یابد و نتایج گزارش شوند. در این حالت اگر پیشامد مورد نظر برای عضوی رخ دهد، زمان تا پیشامد این عضو معلوم می‌باشد ولی اگر عضو تا پایان مطالعه بقا داشت و هیچ حادثه ای غیر از پیشامد مورد نظر باعث خارج شدن عضو از مطالعه نگردد در این صورت مشاهده، سانسور شده از راست با زمان بقای برابر با طول دوره مطالعه است.

حال مطالعه ای را در نظر بگیرید که در آن اعضا، زمان های سانسور مشخص ولی متفاوت از یکدیگر دارند. این شکل از سانسور نوع I سانسور فزاینده نوع I نامیده می‌شود.

نوع دیگری از سانسور که تعمیمی از سانسور نوع I می‌باشد به این صورت است که اعضا در زمان های متفاوتی وارد مطالعه می‌شوند اما زمان پایان مطالعه برای تمام اعضا یکسان است و از قبل توسط آزمایشگر تعیین می‌شود. زمان های سانسور اعضاء هنگامی‌که وارد مطالعه می‌شوند مشخص است لذا در این نوع مطالعه هر عضو زمان سانسور ثابت و خاص خود را دارد. این شکل از سانسور، سانسور تعمیم یافته نوع I نامیده می‌شود.

سانسور نوع II: نوع دیگر سانسور راست، سانسور نوع II است که در آن اعضا در یک زمان مشابه وارد مطالعه می‌شوند و مطالعه تا زمانی که اولین r عضو، پیشامد مطلوب را تجربه کند ادامه می یابد. r عددی از پیش تعیین شده می‌باشد($r \leq n$). یک چنین مطالعه ای صرفه جوئی در زمان و هزینه را در پی دارد زیرا ممکن است زمان بسیار زیادی طول بکشد تا تمام اعضا پیشامد مطلوب را تجربه کنند. این نوع از سانسور معمولاً در تست عمر کالاها و تجهیزات استفاده می‌شود. نوعی سانسور که از سانسور نوع II نتیجه می‌شود سانسور فزاینده نوع II می‌باشد. این شکل از سانسور همانند سانسور فزاینده نوع I می‌باشد یعنی زمان در مطالعه ی (یا زمان سانسور) اعضا مشخص ولی متفاوت از یکدیگر است با این تفاوت که زمان های سانسور، در سانسور فزاینده نوع II برخلاف سانسور فزاینده نوع I تصادفی می‌باشد. در سانسور فزاینده نوع I زمان های سانسور از قبل توسط آزمایشگر مشخص می‌شوند و مقادیری ثابت می‌باشند. اما در سانسور فزاینده نوع II بدین گونه است که اعضای در مطالعه به □ گروه به حجم های n_i تقسیم می شوند. خاتمه آزمایش در هر گروه وقتی است که تعداد r_i عضو آن گروه پیشامد مطلوب را تجربه کنند لذا اعضای باقیمانده که پیشامد مطلوب را تجربه نکرده اند سانسور می‌شوند. بنابراین زمان های سانسور تصادفی می‌باشند.

سانسور نوع III : گاهی اوقات ممکن است اعضای وارد شده به یک مطالعه پیشامدهایی غیر از پیشامد مطلوب در مطالعه را تجربه کنند و این باعث حذف آنها از ادامه مطالعه گردد. پیشامدهای نظیر مرگ های تصادفی یعنی مرگ در اثر هر عاملی غیر از پیشامد مطلوب در مطالعه، مهاجرت، عدم میل به شرکت در ادامه مطالعه و .... در چنین مواقعی پیشامد مورد نظر قابل مشاهده نیست. این نوع سانسور، سانسور تصادفی نامیده می‌شود.

در حالتی که مشاهدات سانسور شده ی تصادفی از راست و دو به دو مستقل از یکدیگرند، برآوردگر تابع بقا به برآوردگر کاپلان مه- یر (۱۹۵۸) مشهور است و خواص مجانبی آن در طول سال های اخیر به طور وسیع توسط محققین بسیاری از جمله برسلو و کرولی (۱۹۷۴)، پترسون(۱۹۷۷)، گیل(۱۹۸۳، ۱۹۸۱، ۱۹۸۰) ، لو و سینگ (۱۹۸۶) ، ونگ (۱۹۸۷)، و اشتوت و ونگ (۱۹۹۳) مورد مطالعه قرار گرفته است. این در حالیست که نتایج اندکی برای مشاهدات بقا در حالتی که مستقل نیستند در دسترس است: ولکل و کرولی (۱۹۸۴) و یینگ و وی (۱۹۹۴) سازگاری و نرمال بودن مجانبی برآوردگر کاپلان مه- یر را تحت وابستگی های نیمه مارکفی و آمیختگی بیان کردند.

در ادامه ابتدا در بخش دوم متغیرها، تعاریف و پیش فرض های مورد نیاز در دنباله بحث را معرفی می‌کنیم. سپس در بخش سوم سازگاری قوی دو برآوردگر کاپلان مه- یر تابع بقا و نلسون آلن تابع تجمعی خطر را نشان خواهیم داد. در بخش چهارم همگرایی ضعیف این دو برآوردگر را بررسی می کنیم.

۲ تعاریف، نمادها و پیش فرض ها

فرض کنید $X_1, \cdots, X_n$ دنباله ای از زمان های بقا برای n فرد باشند. فرض کنید این دنباله دو به دو مستقل فرض نشده اند اما دارای یک تابع توزیع حاشیه ای پیوسته، مشترک و نامعلوم $F(t) = P(X_i \leq t)$ می‌باشند به طوریکه $F(0) = 0$.
و همچنین فرض کنید که متغیرهای تصادفی X_i سانسور شده از راست با متغیرهای تصادفی سانسور Y_i باشند در این صورت تنها (Z_i, δ_i) قابل مشاهده است که در آن

$$Z_i = X_i \wedge Y_i \qquad \text{و} \qquad \delta_i = I(X_i \leq Y_i) \qquad i = 1, \cdots, n$$

$I(A)$ متغیر تصادفی نشانگر پیشامدA و$\wedge$ نشانه مینیمم است.
در این مدل سانسور تصادفی، زمان های سانسور Y_i، $i = 1, \cdots, n$ متغیرهای تصادفی مستقل و هم توزیع با تابع توزیع $G(y) = P(Y_i < y)$ فرض شده اند به طوریکه $G(0) = 0$، که همچنین مستقل از متغیرهای تصادفی X_i می‌باشند.
هدف ارائه استنباط ناپارامتری درباره F بر پایه مشاهدات سانسور شده $(Z_i, \delta_i); i = 1, \cdots, n$ می‌باشد برای این منظور دو فرآیند تصادفی $N_n(t)$ و $Y_n(t)$ را بر $[0, \tau]$ به صورت زیر تعریف می‌کنیم:

$$N_n(t) = \sum_{i=1}^{n} I(Z_i \leq t, \delta_i = 1) = \sum_{i=1}^{n} I(X_i \leq t \wedge y_i)$$
$$Y_n(t) = \sum_{i=1}^{n} I(Z_i \geq t)$$

که $N_n(t)$ تعداد مشاهدات سانسور نشده کوچکتر یا مساوی t و $Y_n(t)$ تعداد مشاهدات سانسور شده و سانسور نشده ی بزرگتر یا مساوی t می‌باشد.

برآوردگر کاپلان مه -یر $(K-M)$ تابع بقا به صورت

$$1 - F_n(t) = \prod_{s \leq r} \left(1 - \frac{dN_n(s)}{Y_n(s)}\right)$$

است که در آن $dN_n(s) = N_n(s) - N_n(s^-)$. برای تابع توزیع F بر $[0, \tau]$ ، تابع تجمعی خطر Δ درحالتی که F پیوسته است (گیل، ۱۹۸۰) به صورت:

$$\Delta(t) = \int_0^t \frac{dF(s)}{1 - F(s^-)} = -\log(1 - F(t))$$

تعریف می‌شود که تابع تجربی آن ، $\Delta_n(t)$، که به برآوردگر نلسون- آلن $\Delta(t)$ معروف است به صورت زیر می‌باشد.

$$\Delta_n(t) = \int_0^t \frac{dN_n(s)}{Y_n(s)}$$

پیشتر فرض کردیم که زمان های بقا دو به دو مستقل نیستند و اکنون فرض می‌کنیم دارای وابستگی α-آمیزنده با تعریف زیر می‌باشند:

تعریف ۱.۲. دنباله $\{X_i, i = 0, \pm 1, \pm 2, \cdots\}$ یک دنباله ی α-آمیزنده از متغیرهای تصادفی است هرگاه، برای دو مجموعه $A \in F_1^k(X), B \in F_{k+n}(X)$ و عدد صحیح مثبت n:

$$\alpha(n) = \sup_{A \in F_1^k(X), B \in F_{k+n}(X)} |P(A \cap B) - P(A)P(B)| \to 0; \quad n \to \infty$$

لم ۲.۲ (کای ۱۹۹۸). اگر $\{X_n; n \geq ۱\}$ دنباله ای از متغیرهای تصادفی α-آمیزنده با ضریب آمیختگی $\alpha(n)$ و $\{Y_n; n \geq ۱\}$ دنباله ای از متغیرهای تصادفی iid و مستقل از $\{X_n; n \geq ۱\}$ باشد آن گاه $\{(X_n, Y_n); n \geq ۱\}$ و همچنین $\{X_n \wedge Y_n; n \geq ۱\}$ دنباله ای از متغیرهای تصادفی α-آمیزنده با ضریب آمیختگی $۴\alpha(n)$می‌باشند.

فرض کنید H تابع توزیع Z_iها به صورت زیر داده شده باشد:

$$\bar{H} = ۱ - H = \bar{F}\bar{G} = (۱ - F)(۱ - G)$$

زمان های (احتمالاً نامتناهی) τ_H, τ_G, τ_F را به صورت

$$\tau_F = \inf\{y : F(y) = ۱\}$$

تعریف می‌کنیم. به سادگی دیده می شود که $\tau_H = \tau_F \wedge \tau_G$ (ونگ و اشتوت ۱۹۹۳)
با تعریف

$$F_*(t) = P(Z \leq t, \delta = ۱)$$

داریم:

$$F_*(t) = \int_۰^\infty F(t \wedge z)dG(z) = \int_۰^t (۱ - G(z))dF(z)$$

قرار می دهیم:

$$\alpha = \alpha(F, G) = P(X \leq Y) = \int_۰^\infty F(z)dG(z) = \int_۰^\infty [۱ - G(z)]dF(z)$$

فرض می‌کنیم $\alpha > ۰$. واضح است که $F(t)/\alpha$ تابع توزیع شرطی Z به شرط $\delta = ۱$ می‌باشد.
اگر $\tau_{F_*} = \inf\{t : F(t) = \alpha\}$ به آسانی مشاهده می‌شود که $\tau_H = \tau_{F_*}$ پس $\tau_{F_*} = \tau_F \wedge \tau_G$
به این ترتیب برای تابع پیوسته F، تابع تجمعی خطر را می توان به صورت

$$\Delta(t) = \int_۰^t \frac{F_*(s)}{\bar{H}(s)}$$

نوشت. قرار می دهیم:

$$\bar{N}_n(t) = N_n(t)/n \qquad , \qquad \bar{Y}_n(t) = Y_n(t)/n$$

پیش از اثبات خواص مجانبی برآوردگرهای کاپلان مه- یر و نلسون-آلن فرض های زیر را می پذیریم.
$K_۱$: $\{X_j; j \geq ۱\}$ دنباله ای از متغیرهای تصادفی α- آمیزنده با تابع توزیع پیوسته است.
$K_۲$: متغیرهای زمان سانسور $\{Y_j; j \geq ۱\}$، iid با تابع توزیع پیوسته G و مستقل از $\{X_j; j \geq ۱\}$ می‌باشند.
$K_۳$: $\alpha(n) = O(n^{-v})$ برای یک $v > ۳$

۳ بررسی سازگاری قوی برآوردگرهای کاپلان مه- یر و نلسون- آلن

در این بخش به بررسی سازگاری قوی دو برآوردگر کاپلان مه -یر و نلسون -آلن می پردازیم. قضیه زیر سازگاری قوی را برای برآوردگر نلسون آلن نشان می دهد:

قضیه ۱.۳. تحت فرض های $(k_1)-(k_3)$

$$\sup_{r\geq 0}|\bar{Y}_n(t)-\bar{H}_n(t)|=O(a_n)\quad a.s. \tag{۱.۳}$$

$$\sup_{r\geq 0}|\bar{N}_n(t)-F_*(t)|=O(a_n)\quad a.s. \tag{۲.۳}$$

که $a_n=(\frac{\log\log n}{n})^{\frac{1}{2}}$ و در نتیجه برای هر $0<\tau<\tau_H$

$$\sup_{0\leq r\leq\tau}|\Delta_n(t)-\Delta(t)|=O(a_n)\quad a.s. \tag{۳.۳}$$

پیش از اثبات قضیه ۱.۳ لم زیر را بیان می‌کنیم:

لم ۲.۳ (کای و روساس، ۱۹۹۲). اگر دنباله $\{X_n, n\geq 1\}$ دنباله ای مانا و α- آمیزنده از متغیرهای تصادفی با تابع توزیع F و ضریب آمیختگی $\alpha(n)=O(n^{-v})$، $v>3$ و F_n تابع توزیع تجربی بر پایه $X_1,\cdots,X_n$ باشد. آنگاه

$$\sup_{x\in R}|F_n(x)-F(x)|=O(a_n)\quad a.s.$$

که در آن $a_n=(\frac{\log\log n}{n})^{\frac{1}{2}}$.

اثبات. قضیه ۱.۳

از لم ۲.۲ به آسانی مشاهده می‌شود که $\{(Z_i,\delta_i); i\geq 1\}, \{Z_i; i\geq 1\}$ دو دنباله از متغیرهای تصادفی α- آمیزنده مانا می‌باشند . پس قسمت اول و دوم قضیه با استفاده از لم ۲.۳ و این حقیقت که $\bar{N}_n$ و $\bar{Y}_n$ و $1-$ تابع های تجربی هستند، نتیجه می شود. برای اثبات قسمت سوم، از لم ۲ در گیل (۱۹۸۱) و دو قسمت اول قضیه، نتیجه می‌شود که:

$$\sup_{0\leq r\leq\tau}|\Delta_n(t)-\Delta(t)|\leq\frac{2\rho(\bar{N}_n,F_*)}{\bar{Y}_n(\tau)}+\frac{\rho(\bar{Y}_n,\bar{H})[\bar{N}_n(\tau)+\rho(\bar{N}_n,F_*)]}{\bar{Y}_n(\tau)[\bar{Y}_n(\tau)-\rho(\bar{Y}_n(\tau),\bar{H}]}=O(a_n)\quad a.s$$

که ρ سوپریمم بر $[0,\tau]$ است و به این ترتیب حکم ثابت می‌شود. □

پس از اثبات سازگاری قوی برآوردگر نلسون- آلن حال ، در قضیه بعد سازگاری قوی را برای برای برآوردگر کاپلان مه-یر نشان می دهیم.

قضیه ۳.۳. تحت فرض های $(k_1)-(k_3)$ برای هر $0<t<\tau_H$

$$\sup_{0\leq r\leq\tau}|F_n(t)-F(t)|=O(a_n)\qquad a.s$$

اثبات. با نوشتن بسط تیلور برای تابع $e^{-\Delta(t)}$ در نقطه $\Delta_n(t)$ و تابع $e^{\log(1-F_n(t))}$ در نقطه $\Delta_n(t)$ داریم:

$$\begin{aligned}
F_n(t)-F(t)&=1-F(t)-(1-F_n(t))\\
&=e^{-\Delta(t)}-e^{\log(1-F_n(t))}\\
&=e^{-\Delta_n^*(t)}[\Delta_n(t)-\Delta(t)]+e^{-\Delta_n^{**}(t)}[-\log(1-F_n(t))-\Delta_n(t)]\\
&=e^{-\Delta_n^*(t)+\Delta(t)}e^{-\Delta(t)}[\Delta_n(t)-\Delta(t)]\\
&+e^{-\Delta_n^{**}(t)+\Delta_n(t)}e^{-\Delta_n(t)+\Delta(t)}e^{-\Delta(t)}[-\log(1-F_n(t))-\Delta_n(t)]\\
&=I+II
\end{aligned} \tag{۴.۳}$$

که Δ^* و Δ^{**} در روابط زیر صدق می‌کنند:

$$\rho(\Delta_n^*, \Delta) \leq \rho(\Delta_n, \Delta) \tag{۵.۳}$$

$$\rho(\Delta_n^{**}, \Delta_n) \leq \rho(-\log(۱ - F_n), \Delta_n) \leq C(\tau)/n \tag{۶.۳}$$

از طرفی می دانیم

$$|e^{-x} - e^{-y}| \leq |x - y|; \qquad x, y \geq ۰$$

پس:

$$\begin{aligned} I &= (e^{-\Delta_n^*(t)+\Delta(t)} - ۱)e^{-\Delta(t)}[\Delta_n(t) - \Delta(t)] + e^{-\Delta(t)}[\Delta_n(t) - \Delta(t)] \\ &\leq |e^{-\Delta_n^*(t)} - e^{-\Delta(t)}||\Delta_n(t) - \Delta(t)| + e^{-\Delta(t)}[\Delta_n(t) - \Delta(t)] \\ &= O(\rho^۲(\Delta_n, \Delta)) + e^{-\Delta(t)}[\Delta_n(t) - \Delta(t)] \end{aligned} \tag{۷.۳}$$

و

$$\begin{aligned} II &= [۱ + (-\Delta_n^{**}(t) + \Delta_n(t)) + (-\Delta_n^{**}(t) + \Delta_n(t))^۲/۲ + \cdots] \\ &[۱ + (-\Delta_n(t) + \Delta_n(t)) + (-\Delta_n(t) + \Delta_n(t))^۲/۲ + \cdots]e^{-\Delta(t)}(-\log(۱ - F_n(t) - \Delta_n(t)) \\ &= O(n^{-۱}) \qquad a.s \end{aligned} \tag{۸.۳}$$

و به این ترتیب از (۵.۳)-(۸.۳) نتیجه می‌شود:

$$F_n(t) - F(t) = \bar{F}(t)[\Delta_n(t) - \Delta(t)] + O(a_n^۲) \tag{۹.۳}$$

در نتیجه با استفاده از قضیه قبل حکم ثابت می‌شود.

□

۴ بررسی همگرایی در توزیع برآوردگرهای کاپلان مه -یر و نلسون- آلن

پیش از بیان قضایای مربوط به همگرایی ضعیف ابتدا نشان می دهیم که برآوردگر کاپلان مه-یر و برآوردگر نلسون-آلن را می توان به صورت میانگین متغیرهای تصادفی به اضافه باقیمانده ای از مرتبه $O(n^{-\frac{۱}{۲}}(\log n)^{-\lambda})$، $\lambda > ۰$ نوشت و سپس از آن در اثبات همگرایی ضعیف این دو برآوردگر استفاده می‌کنیم. این مطلب در قضیه ۱.۴ بیان شده است. برای این منظور قرار می دهیم:

$$g(x) = \int_۰^x (\bar{H}(s))^{-۲} dF_*(s)$$

و برای x, z حقیقی مثبت و $\delta = ۰$ یا ۱

$$\zeta(z, \delta, x) = g(z \wedge x) - I(z \leq x, \delta = ۱)/\bar{H}(z)$$

مشاهده می‌شود که

$$\begin{aligned} &E(\zeta(Z_i, \delta_i, x)) = ۰ \\ &Cov(\zeta(Z_i, \delta_i, x), \zeta(Z_i, \delta_i, t)) = g(s \wedge t) \end{aligned}$$

قضیه ۱.۴ (کای ۱۹۹۸). تحت فرض های $(k_۱)-(k_۳)$

$$i)\Delta_n(t)-\Delta(t)=-\frac{۱}{n}\sum_{i=۱}^{n}\zeta(Z_i,\delta_i,x)+r_n(t)$$
$$ii)F_n(t)-F(t)=-\frac{\bar{F}(t)}{n}\sum_{i=۱}^{n}\zeta(Z_i,\delta_i,x)+r_n(t)$$

که برای هر $\tau<\tau_H$

$$\sup_{۰\le r\le\tau}|r_n(t)|=O(b_n)\qquad a.s$$

و

$$b_n=n^{-\frac{۱}{۲}}(\log n)^{-\lambda}$$

حال با استفاده از نمایش برآوردگرهای کپلان مه-یر و نلسون-آلن به صورت میانگین متغیرهای تصادفی، همگرایی ضعیف آن ها را در قضیه های زیر اثبات می‌کنیم.

قضیه ۲.۴. تحت فرض های $(k_۱)-(k_۳)$ برای $t\in[۰,\tau]$ و هر $\tau<\tau_H$

$$\sqrt{n}[\Delta_n(t)-\Delta(t)]\xrightarrow{L}N(۰,\sigma^۲(t))$$

که

$$\sigma^۲(t)=Var(\zeta(Z_i,\delta_i,t))+۲\sum_{j=۲}^{\infty}Cov(\zeta(Z_۱,\delta_۱,t),\zeta(Z_j,\delta_j,t))$$

اثبات. طبق لم ۲.۲ $\{\zeta(Z_i,\delta_i,t),i=۱,۲,\cdots\}$ دنباله ای پایا، α- آمیزنده و کراندار از متغیرهای تصادفی است. به این ترتیب با استفاده از قضیه ۱.۴ و قضیه ۴۰۵۰۱۸ در ایبراگیمف و لی نیک (۱۹۷۱) حکم به سادگی ثابت می شود. □

قضیه ۳.۴. تحت فرض های $(k_۱)-(k_۲)$ برای $t\in[۰,\tau]$ و هر $\tau<\tau_H$

$$\sqrt{n}[F_n(t)-F(t)]\xrightarrow{L}N(۰,\sigma^۲(t)\bar{F}^۲(t))$$

اثبات. با استفاده از قضیه ۲.۴ و رابطه (۹.۳) حکم به سادگی ثابت می‌شود. □

مراجع

[132] Breslow, N., Crowley, J. (1974). *A large sample study of the life table and product limit estimator under random censorship.* Ann. Statist. 2, 437-453.

[133] Cai, Z. (1998). *Asymptotic properties of Kaplan- Meier estimator for censored dependent data.* Statistics & Probability Lett. 37, 381-389.

[134] Cai, Z., Roussas, G. G. (1992). *Uniform strong estimation under - mixing , with rate*. Statist. probab. Lett. 15, 47-55.

[135] Csorgo, M., Revesz, P. (1981). *Strong Approximatoin in probability and Statistics* . Academic Press, New York,

[136] Gill, R. D. (1980). *Censoring and Stochastic Intergrals.* Mathematical Center Tracts No. 124, Mathematisch Centrum, Amsterdam.

[137] Gill, R. D. (1981). *Testing with replacement and the product limite estimator.* Ann. Statist. 9, 853-860.

[138] Gill, R.D. (1983). *Large sample behavior of the product limit estimator on the whole line.* Ann. Statist. 11, 49-56.

[139] Ibragimov, I. A., Linnik Yu, V. (1971). *Independent and Stationary Sequences of Random Varibles*. Walters Noordhoff, Groningen, the Nehterlands.

[140] Kaplan, E. L., Meier, P. (1958). *Non parametric estimation from incomplete observation.* J. Amer. Statist. Assoc. 53, 457-481

[141] Lo, S. H., Singh, K. (1986). *The product limit estimator and the bootstrap: some asymptotic representations.* Probab. Theory Rel. Fields 71, 455-465.

[142] Peterson, A. V. (1977). *Expressing the Kaplan-Meier estimator sa a function of empirical subsurvival functions.* . J. Amer. Statist. Assoc. 72, 854-858.

[143] Stute, W., Wang, J.L. (1993). *A strong law under random censorship.* Ann. Statist. 21, 1591-1607.

[144] Voelkel, J., Crowley, J. (1984). *Nonparametric inference for a class of seim-Markov process with censored observation.* Ann. Statist. 12, 142-160.

[145] Wang, J. G. (1987). *A note on the uniform consistency of Kaplan – Meier estimator.* Ann. Statist. 15, 1313-1316.

[146] Ying, Z., Wei, L. J. (1994). *The Kaplan – Meier estimate for dependent failure time observations.* J. Multivariate Anal. 50, 17- 29.

کاربرد تکنیک مدلسازی معادلات ساختاری در شناسایی عوامل خطر بیماری فشار خون

راضیه یوسفی [1]، کارشناسی ارشد آمارزیستی، گروه اپیدمیولوژی و آمارزیستی، دانشگاه علوم پزشکی مشهد، آزاده ساکی، استادیار آمارزیستی، گروه اپیدمیولوژی و آمارزیستی، دانشگاه علوم پزشکی مشهد

چکیده: در ساختاربندی مدل‌های چندسطحی تک معادله‌ای، معمولا یک متغیر پاسخ بر اساس یک یا چندین متغیر تبیینی مدل‌بندی می‌شود. با این حال، مثال‌های بیشماری مرتبط با برخی از پدیده‌های اجتماعی و اقتصادی وجود دارد که مدل‌های تک معادله‌ای به تنهایی نمی‌توانند تشریح کاملی از ارتباط بین متغیرهای پاسخ و تبیینی را فراهم کنند. مدل‌های چندسطحی به ظاهر نامربوط می‌توانند چنین کمبودهایی را برطرف کنند. همانند سایر مدل‌های خطی، مدل‌های چندسطحی به ظاهر نامربوط نیز دربرگیرنده پیش فرض‌هایی مانند نرمال بودن توزیع خطاها است که ممکن است فرض مذکور در مسائل واقعی برقرار نباشند. در این صورت، استفاده از توزیع‌های نامتقارن مانند توزیع چوله-نرمال می‌تواند مثمرثمر باشد. در مقاله حاضر، با در نظر گرفتن فرض چوله-نرمال برای خطای سطح دوم مدل‌های چندسطحی به ظاهر نامربوط، نحوه بکارگیری این دست مدل‌ها مورد مطالعه قرار می‌گیرد. سپس، با استفاده از رویکرد بیزی برآوردهای مدل و استنباط‌های آماری مرتبط انجام می‌شود. علاوه بر این، با برازش مدل‌های پیشنهادی در مورد داده‌های هزینه- درآمد سال ۱۳۹۰ خانوارهای روستایی ایرانی، نحوه عملکرد آن‌ها مورد بررسی قرار می‌گیرد.

واژه‌های کلیدی: مدل‌های چندسطحی به ظاهر نامربوط، توزیع چوله-نرمال، استنباط بیزی، هزینه و درآمد.

مقدمه

با توجه به رشد و گسترش جمعیت و به طبع آن افزایش شیوع بیماری‌های مختلف، امروزه مطالعات زیادی در خصوص بیماری ها و شناسایی ریسک فاکتورهای موثر بر ان ها انجام می‌شود. در سال‌های اخیر، مزایای مدل‌های معادلات ساختاری به کمک تکنیک‌های معمول آماری مورد ملاحظه قرار گرفته و موجب افزایش بکارگیری تکنیک مدلسازی معادلات ساختاری در تحلیل داده‌ها و آزمون اینگونه مدلها گردیده است. در این مطالعه سعی شد در مدلسازی فشارخون و ریسک فاکتورهای مرتبط با آن، برآورد پارامترها براساس روش درستنمایی ماکزیمم محاسبه گردد.

مواد و روش ها

در این مطالعه مقطعی مبتنی بر جمعیت، مدلسازی معادلات ساختاری به منظور مدلسازی ارتباطات بین عوامل مستعدکننده بیماری فشارخون بکارگرفته شد. اطلاعات مربوط به نمونه تصادفی شامل ۹۷۶۵ نفر از مطالعه مشهد مورد استفاده قرار گرفت. داده‌ها به کمک نرم افزار Amos22 مورد تحلیل قرار گرفتند.

نتایج

[1] راضیه یوسفی : o.akhgari@gmail.com

مدل مفهومی براساس شاخص های نیکویی برازش

$$(CFI = ۰/۹۳۹, TLI = ۰/۹۰۸, NFI = ۰/۹۳۵, RMSEA = ۰/۰۴, SRMR = ۰/۰۳۷)$$

مورد تایید قرار گرفت. متغیر سن و پس از آن چاقی و کم تحرکی بیشترین تاثیر را روی فشار خون نشان دادند.

نتیجه گیری

یافته‌ها شواهدی را برای تایید مدل مفهومی در نظر گرفته شده در مورد عوامل مستعدکننده بیماری فشارخون نشان دادند که می‌توانند در بکارگیری سیاست‌های پیشگیرانه در جلوگیری از ابتلا به این بیماری مفید باشند.

۱ مقدمه

با توجه به رشد و گسترش جمعیت و به طبع آن افزایش شیوع بیماری‌های مختلف، امروزه مطالعات زیادی در خصوص بیماری ها و شناسایی ریسک فاکتورهای موثر بر ان ها انجام می‌شود. از جمله روش‌هایی که به منظور انجام اینگونه تحلیل ها کاربرد دارد مدل های رگرسیونی هستند. اما از آنجاییکه در اغلب این بیماری ها بیشتر عوامل مرتبط همبستگی دارند بکارگیری مدل های رگرسیون معمول که با فرض عدم هم خطی و عدم همبستگی بین پیشگوها برازش داده می‌شوند مناسب نیست. بنابراین نیاز به مدل های کامل تری وجود دارد که بتواند ساختارهای همبستگی را تشخیص دهد و عوامل اصلی و مستقل را در غالب متغیرهای پنهان به همراه سایر متغیرهای آشکار مورد بررسی قرار دهد. بدین منظور مدلسازی معادلات ساختاری می‌تواند مورد توجه قرار گیرد. مدل یابی معادله ساختاری یک تکنیک تحلیل چند متغیری کلی و بسیار نیرومند از خانواده رگرسیون چند متغیری و به بیان دقیق تر بسط مدل خطی کلی است که به پژوهشگر امکان می‌دهد مجموعه ای از معادلات رگرسیون را بصورت همزمان مورد آزمون قرار دهد. مدلسازی معادلات ساختاری می‌تواند به راحتی برای تحلیل های پیچیده که شامل روابط مفروض بین یک یا چند متغیر پنهان مستقل و یک یا چند متغیر پنهان وابسته باشد، بکار گرفته شود ((۱۴۷)-(۱۴۹)). از مزایای اصلی مدل‌های معادلات ساختاری میتوان به انعطافپذیری، مدلسازی متغیرهای پنهان، در نظر گرفتن خطا و آزمون مدل‌های مفروض اشاره نمود(۱۴۸). مباحث عملی مملو از متغیرهای پنهانی است که نمی‌توانند برای هر فرد اندازه گیری شوند. در سال‌های اخیر مدل‌های معادلات ساختاری به طور فزاینده ای در مسایل مهمی همچون علوم رفتاری، پزشکی و علوم اجتماعی بکار گرفته شدند و بنابر این امروزه تحلیل مدل‌های معادلات ساختاری بخش مهمی از تحقیقات آماری را به خود اختصاص می‌دهد. نرم افزارهای گوناگون تحلیل این معادلات را بر مبنای مدل‌های خطی که در آنها رابطه بین متغیرهای پنهان خطی است انجام می‌دهند(۱۵۰).

با توجه به گسترش کاربرد مدل‌های معادلات ساختاری، در این مقاله سعی شده است به معرفی این روش و بیان کاربرد آن در شناسایی عوامل خطر بیماری ها پرداخته شود. بنابراین پس از شرح این تکنیک، شناسایی عوامل خطر بیماری فشار خون در یک نمونه واقعی با استفاده از مدلسازی معادلات ساختاری انجام گرفت. امید است با تعیین مدل مناسب گامی در جهت پیشگیری اولیه از بروز این بیماری و گسترش آن در جامعه برداشته شود.

۲ روش کار

در این مطالعه به بررسی عوامل خطر بیماری فشار خون بر روی داده‌های مطالعه مشهد پرداخته شده است. مطالعه مشهد که از سال ۱۳۸۷ در شهر مشهد با هدف تعیین شیوع عوامل خطر ساز بیماری‌های غیر واگیردار و بهبود شیوه ی زندگی برای کاهش این عوامل خطر ساز آغاز شد، شامل ۹۷۶۵فرد بین سن۶۵-۳۵ سال می باشد. روش نمونه گیری طبقه ای- خوشه ای بود. به طوری که کل شهرستان مشهد به سه مرکز بهداشت ۳،۲،۱ تقسیم بندی شده بود و خوشه‌ها مناطق تحت پوشش این مرکز بودند. توضیحات بیشتردر مورد نحوه

استخراج نمونه و فرایند جمع آوری اطلاعات در مطالعه غیور مبرهن و همکاران آمده است. (۱۵۱) متغیرهای مورد استفاده در این مطالعه عبارت بودند از: متغیر پنهان موقعیت اقتصادی و اجتماعی با نشانگرهای میزان درامد خانوار در ماه، میزان تحصیلات سرپرست خانوار (برحسب تعداد سال‌های تحصیل) و تعداد اعضای خانوار، متغیر پنهان فشارخون براساس دو متغیر فشارخون سیستولیک و دیاستولیک افراد، متغیر پنهان تغذیه ناسالم شامل میزان اسید چرب اشباع، کلسترول، ساکاروز و سدیم مصرفی افراد در رژیم روزانه آنها، متغیر پنهان چاقی براساس دو شاخص مربوط به چاقی BMI و WHR ، مغیرهای مشاهده شده دیگری که در این تحلیل وارد شدند عبارت بودند از سن، میزان فعالیت فیزیکی افراد، میزان قند خون و تری گلیسیرید خون.

یک مدل معادلات ساختاری برای براورد اثرات فاکتورهای خطر موثر بر بالا رفتن فشارخون تعیین گردید.پس از تعریف مدل مفهومی‌شناسایی پذیر بودن و مناسب بودن مدل مورد توجه قرار گرفت. همانطور که اشاره شد هدف مدل‌های معادلات ساختاری آزمون این فرضیه است که آیا ماتریس واریانس مشاهده شده برای یک مجموعه از متغیرهای اندازه‌گیری شده، برابر با ماتریس کوواریانس باز تولید شده از طریق مدل مفروض است یا خیر و این فرضیه به صورت زیر بیان میشود:

$$\Sigma = \Sigma(\theta)$$

که در آنΣ: ماتریس کوواریانس تولید شده ازجامعه توسط متغیرهای مشاهده شده و $\Sigma(\theta)$: ماتریس کوواریانس باز تولید شده از مدل مفروض براساس θ است. مدل‌های معادلات ساختاری پایه، شامل مدل اندازه گیری و مدل ساختاری هستند. مدل اندازه گیری جزئی از مدل ساختاری است که نحوه سنجش یک متغیر پنهان را با استفاده از دو یا تعداد بیشتری متغیر مشاهده شده تعریف می‌کند. مدل اندازه گیری عبارت است از:

$$x_i = \Lambda\omega_i + \varepsilon_i \quad i = ۱, ۲, \ldots, n$$

که در آن x_i یک بردار $p \times ۱$ از متغیرهای نشانگر توضیحی برای $q \times ۱$ بردار تصادفی متغیرهای پنهان ω_i و Λ یک ماتریس $p \times q$ از بارهای ضرایب بدست آمده از رگرسیون x_i بر ω_i است و ε_i یک بردار تصادفی $p \times ۱$ از خطاهای اندازه گیری است که دارای توزیع $\mathrm{N}(۰, \psi_\varepsilon)$ هستند. فرض می‌شود که برای $\mathrm{i} = 1, 2, \cdots, \mathrm{n}$ ،ω_i ها مستقلند و از توزیع $\mathrm{N}(0,\phi)$ پیروی می‌کنند، و با بردار تصادفی ε_i ناهمبسته هستند. مدل ساختاری جزئی از مدل‌های معادلات ساختاری است که نشان می‌دهد متغیرهای پنهان (و گاهی اوقات آشکار) چگونه بر یکدیگر اثر می‌گذارند. این مدل به ترتیب زیر است:

$$\eta_i = \beta\eta_i + \Gamma\xi_i + \delta_i \quad i = ۱, \cdots, n$$

وقتی β یک ماتریس $m \times m$ از پارامترهای ساختاری است که بیان کننده رابطه بین متغیرهای پنهان درون زا است و قطر آن صفر است. Γ ماتریس پارامتر رگرسیون $m \times n$ برای ارتباط دادن متغیرهای درون زا و برون زا است. δ_i بردار $m \times ۱$ که فرض می‌شود از توزیع $\mathrm{N}(۰, \psi_\delta)$ پیروی می‌کند و ψ_δ یک ماتریس کوواریانس قطری است. همچنین فرض می‌شود که δ_i با ξ_i ناهمبسته اند ((۱۴۷)،(۱۵۲)). هر مدل اندازه گیری متشکل از سه نوع متغیر است که شامل متغیر پنهان، متغیرمشاهده شده و متغیر خطا می‌شوند. محقق همواره به معیارهایی نیاز دارد تا با کمک گرفتن از آنها و تفسیرشان، به ارزیابی مدل تدوین شده بر مبنای چارچوب نظری و پیشینه تجربی دست زند(۱۵۳). برخی از این شاخص ها عبارتند از:

★ شاخص کای اسکوئر

شاخص کای اسکوئر را می‌توان به عنوان عمومی ترین و پرکاربردترین شاخص برازش در مدل سازی معادله ساختاری تلقی کرد. اولین نکته ای که درباره تفسیر مقدار این شاخص می‌توان اظهار کرد این است که هرچه مقدار آن کوچکتر باشد برازش داده‌ها به مدل بهتر است. فرمول محاسبه این شاخص عبارت است از:

$$\chi^{۲} = (n - ۱) \times F_{ML}$$

که در آن:

$$F_{ML} = ln\,|\Sigma| - ln\,|S| + tr\left[(S)\left(\Sigma^{-۱}\right)\right] - K$$

$ln\,|S|$ لگاریتم طبیعی دترمینال ماتریس کوواریانس مشاهده شده و $ln\,|\Sigma|$ لگاریتم طبیعی دترمینال ماتریس کوواریانس باز تولید شده است.

به دلیل حساس بودن این شاخص به حجم نمونه از شاخص اصلاح شده آن نسبت به درجه آزادی استفاده می‌شود. هرچند در صورتی که حجم نمونه بسیار بالا باشد شاخص کای اسکوئر اصلاح شده نیز مقادیر بزرگی خواهد داشت و قابل اتکا نیست (۱۵۴).

★ شاخص NFI

این شاخص به افزودن پارامتر به مدل حساس است به نحوی که هرچه پارامتر به مدل افزوده شود مقدار این شاخص نیز افزایش می‌یابد.

$$NFI = \frac{\chi_n^{۲} - \chi_m^{۲}}{\chi_n^{۲}}$$

مقدار قابل قبول برای این شاخص حداقل ۰/۹۰ و مقداری که نشان دهنده ی یک برازش خوب است حداقل ۰/۹۵ می باشد.

★ شاخص TLI

این شاخص تلاش می‌کند تا نقطه ضعف شاخص NFI را در به حساب نیاوردن جریمه شاخص برای افزودن پارامتر مرتفع کند. این شاخص بر مبنای متوسط ضرایب همبستگی بین متغیرها در مدل قرار دارد. هرچه این ضرایب کوچکتر باشند شاخص توکر لویس نیز مقدار کوچکتری را نشان خواهد داد. این شاخص بین صفر تا یک تغییر می‌کند و مقدار۰/۹۵ یا بیشتر منعکس کننده ی یک مدل خوب است.

$$TLI = \frac{\left(\chi_n^{۲}/df_n\right) - \left(\chi_m^{۲}/df_m\right)}{\left(\chi_n^{۲}/df_n\right) - ۱}$$

★ شاخص CFI

این شاخص نیز بر مبنای همبستگی بین متغیرهای حاضر در مدل قرار دارد به نحوی که ضرایب بالای همبستگی بین آن ها به مقادیر بالای این شاخص می‌انجامد. مقدار آن از فرمول بدست می‌آید. مقدار بیشتر از ۰/۹۰ برای این شاخص بیانگر برازش مناسب مدل است.

$$CFI = \frac{\left(\chi_n^{۲} - df_n\right) - \left(\chi_m^{۲} - df_m\right)}{\left(\chi_n^{۲} - df_n\right)}$$

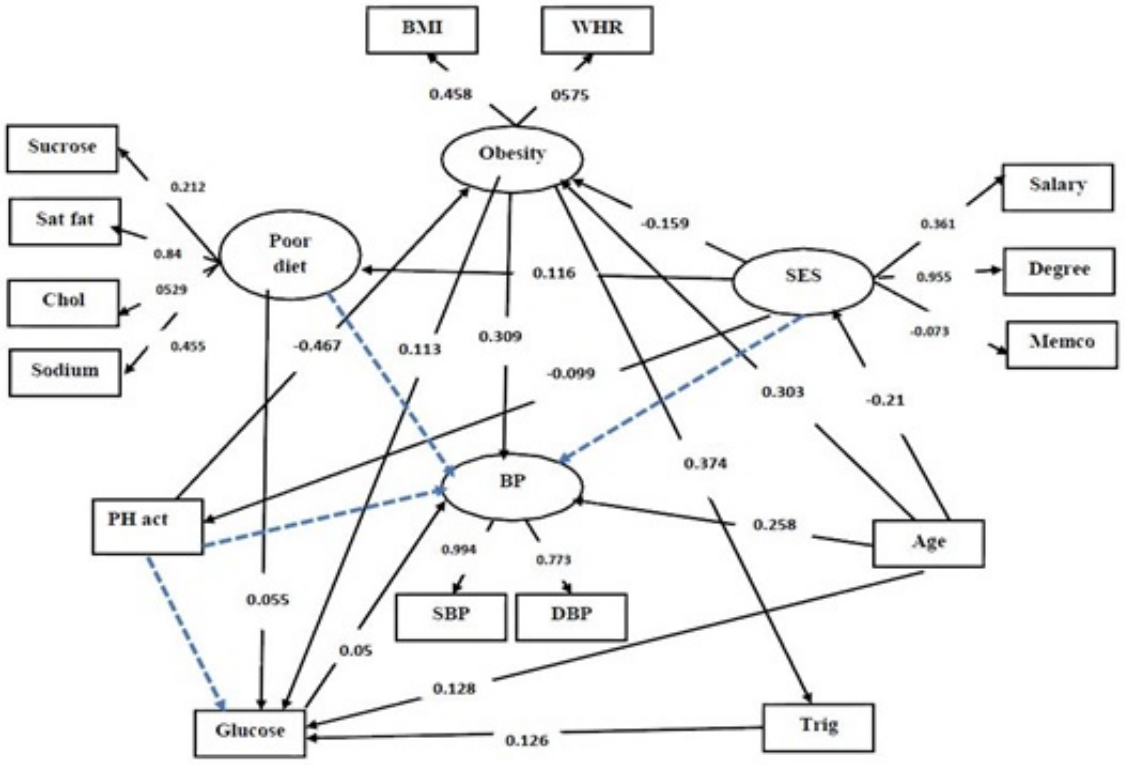
★ شاخص RMSEA

این شاخص بر مبنای تحلیل ماتریس باقیمانده قرار دارد. مدل‌های قابل قبول دارای مقدار ۰/۵۰ یا کوچکتر برای این شاخص هستند. برازش مدل‌هایی که دارای مقادیر بالاتر از ۰/۱ هستند ضعیف برآورد می‌شود.

$$RMSEA = \sqrt{\frac{\chi^2_m - df_m}{(n-1) \times df_m}}$$

۳ نتایج

از ۹۷۶۵ فرد شرکت داده شده در تحلیل، حدود ۴۰درصد مرد و۶۰درصد زن بودند. میانگین سنی افراد ۱/۴۸ بود. ۲۳درصد از کل جمعیت به بیماری پر فشاری خون مبتلا بودند. (فشارخون سیستولیک بالای ۱۴۰ و فشارخون دیاستولیک بالای ۹۰). شیوع این بیماری در بین مردان بیشتر از زنان بود. (۲۳/۳ درصد در مقابل ۲۴/۴ درصد). برای مشخصات توصیفی نمونه مورد مطالعه به مطالعه غیور و همکاران مراجعه شود(۱۵۱)



شکل ۱: ضرایب برآورد شده از برازش مدل مفهومی

مدل مفهومی براساس شاخص های نیکویی برازش، برازش مناسبی به داده‌ها نشان داد

$$(CFI = ۰/۹۳۹, TLI = ۰/۹۰۸, NFI = ۰/۹۳۵, SRMR = ۰/۰۳۷, RMSEA = ۰/۰۴)$$

. نمودار ۱ ضرایب مسیر و بارهای عاملی بدست آمده را نشان می‌دهد. اثرات مستقیم سن، چاقی و قندخون بر فشارخون معنی دار بودند. همانطور که ملاحظه می‌شود، متغیر موقعیت اقتصادی و اجتماعی با اینکه اثر مستقیم بر فشارخون ندارد اما اثر غیر مستقیم بر آن دارد. بررسی اثرات کل متغیرها نشان می‌دهد متغیرسن بیشترین تاثیر بر روی فشارخون را دارد. متغیر فعالیت بدنی اثر مستقیم بر روی

فشارخون ندارد اما اثر کل نسبتا بالایی دارد. منفی بودن این مقدار نشان می‌دهد که با افزایش فعالت بدنی، فشارخون کاهش می‌یابد. اثرات کل متغیرها در جدول ۱ آمده اند .فعالیت بدنی به شدت موجب کاهش احتمال بروز چاقی می‌گردد و همچنین تری گلیسرید خون را نیزکاهش می‌دهد. چاقی بر فشارخون سیستولیک نسبت به فشارخون دیاستولیک تاثیرگذارتر است. در مجموع کلیه متغیرها تاثیر بیشتری روی فشارخون سیستولیک نشان داده اند.

۴ بحث

در چند دهه گذشته محققان زیادی بر روی عوامل موثر بر بیماری فشارخون مطالعاتی را انجام داده اند. در این مطالعات مدل‌های گوناگونی نیز برای شناسایی این عوامل بکارگرفته شد. از جمله زنگ و همکاران در مطالعه خود عوامل خطر مربوط به فشارخون را به کمک یک مدل رگرسیون لجستیک بررسی کردند(۱۵۵) در مطالعه ون نیز این مدل برای شناسایی ارتباط بین فشارخون و عوامل خطر آن بکارگرفته شد((۱۵۵)، (۱۵۶)). اما در کلیه این مدلها تنها اثرات عوامل در مقایسه بین افراد مبتلا به این بیماری و افراد سالم مورد توجه قرار گرفته است، در صورتیکه تاثیر این عوامل در بین افراد در معرض خطر و حتی در میان سالمین به منظور پیشگیری از ابتلا نیز اهمیت ویژه‌ای دارد. بنابراین در این مطالعه سعی شده است با در نظر گرفتن یک مدل آماری به بررسی عواملی بپردازیم که در همه‌ی افراد جامعه از سالم گرفته تا افراد مبتلا به پرفشاری خون در مرحله‌ی حاد، موجب افزایش خطر بیماری می‌گردد. در این شرایط نیاز به مدلی بود که بتواند عوامل مختلف را در یک تحلیل چند متغیره وارد کند و از طرفی همبستگی این عوامل را در نظر بگیرد. بنابراین این مطالعه براساس یک مدل پیشبینی با استفاده از یک مطالعهی مقطعی برای شناسایی عوامل مستعدکننده‌ی بیماری فشارخون در افراد بالای ۳۵ سال صورت گرفت. متغیر فشارخون به عنوان متغیر پاسخ بصورت پنهان و متغیرهای چاقی، موقعیت اقتصادی و اجتماعی و تغذیه‌ی ناسالم به شکل متغیرهای پنهان تعریف شدند. بعلاوه متغیر سن متغیر برونزای این مدل بود. متغیرهای میانجی قندخون و تریگلیسرید نیز به مدل اضافه شدند. انتخاب این متغیرها با مطالعات زیادی هماهنگ بود(۱۵۷) .مطالعات زیادی از مدلسازی معادلات ساختاری جهت شناسایی عوامل موثر بر بیماری ها و از جمله فشار خون بهره گرفته اند. ((۱۶۰)-(۱۵۸)).

پس از تعیین متغیرها و مدل مفهومی، مدل کلاسیک معادلات ساختاری با استفاده از روش برآورد MLR برازش یافت و سپس شاخص های نیکویی برازش بررسی گردید. مقدار شاخص CFI و TLI و NFI بیشتر از ۰٫۹، شاخص RMSEA و SRMR کمتر از ۰٫۵۰ همگی گواهی بر مناسب بودن مدل کلاسیک در این مطالعه بودند. مدل مفهومی با یافته‌های پژوهش برازش مناسبی داشت. سن، چاقی و کمبود فعالیت بدنی عواملی هستند که موجب افزایش ریسک ابتلا به بیماری پرفشاری خون میگردند. این یافته‌ها نیز با پژوهش های دیگر همسو است.درمطالعه برملاگ و همکاران رابطه قوی میان افزایش فشارخون سیستولیک و دیاستولیک در هر دو گروه دارای فشارخون بالا و تحت درمان و گروه افراد با سطح پایین فشارخون مشاهده شد. (۱۶۱). ککینوس و همکاران در مطالعه ای اذعان داشتند که می‌توان پذیرفت که افزایش فعالیت بدنی با شدت و دوره مناسب با عث کاهش فشارخون می‌گردد. بنابراین برای پیشگیری اولیه در یک برنامه سلامت ملی باید برنامه‌های ورزشی مورد توجه ویژه قرار گیرد(۱۶۲).موسویان و همکاران تاثیرات فعالیت بدنی مشخصی را بر فشارخون زنان یائسه دارای اضافه وزن بررسی کرده اند. در این مطالعه این فعالیت موجب کاهش میانگین فشارخون سیستولیک و دیاستولی شده است (۱۶۳).

۵ نتیجه گیری

تکنیک مدلسازی معادلات ساختاری این امکان را فراهم نمود تا به طور همزمان به ارزیابی کیفیت سنجش متغیرها و مقبولیت اثرات مستقیم و غیرمستقیم و همچنین تعامل‌های تعریف شده میان متغیرها بپردازیم. براساس نتایج بدست آمده به نظر می‌رسد انتخاب شیوهی زندگی سالم شامل تحرک بدنی کافی و درنتیجه جلوگیری از بروز چاقی برای پیشگیری و مدیریت افزایش فشارخون امری ضروری به نظر می‌رسد.

جدول ۱: ضرایب بدست آمده از مدل معادلات ساختاری

	سن	موقعیت اقتصادی اجتماعی	فعالیت بدنی	چاقی	تری گلیسرید	تغذیه ناسالم	قندخون	فشارخون
موقعیت اقتصادی اجتماعی	−۰/۲۱	–	–	–	–	–	–	–
فعالیت بدنی	۰/۰۰۲۱	−۰/۰۹۹	–	–	–	–	–	–
چاقی	۰/۳۲۷	−۰/۱۱۳	−۰/۴۶۷	–	–	–	–	–
تری گلیسرید	۰/۱۲۲	−۰/۰۴۲	−۰/۱۷۵	۰/۳۷۴	–	–	–	–
تغذیه ناسالم	−۰/۰۲۴	۰/۱۱۶	–	–	–	–	–	–
قندخون	۰/۱۷۹	−۰/۰۱۲	−۰/۰۷۵	۰/۱۶۱	۰/۱۲۶	۰/۰۵۵	–	–
فشارخون	۰/۳۶۸	−۰/۰۳۵	−۰/۱۴۸	۰/۳۱۷	۰/۰۰۶	۰/۰۰۳	۰/۰۵	–
سدیم	−۰/۰۱۱	۰/۰۵۳	–	–	–	۰/۴۵۵	–	–
ساکاروز	−۰/۰۰۵	۰/۰۲۵	–	–	–	۰/۲۱۲	–	–
جمعیت خانوار	۰/۰۱۵	−۰/۰۷۳	–	–	–	–	–	–
کلسترول	−۰/۰۱۳	۰/۰۶۱	–	–	–	۰/۵۲۹	–	–
اسیدچرب اشباع	−۰/۰۲	۰/۰۹۷	–	–	–	۰/۸۴۱	–	–
تحصیلات	−۰/۲۰۱	۰/۹۵۵	–	–	–	–	–	–
درآمد	−۰/۰۷۶	۰/۳۶۱	–	–	–	–	–	–
شاخص دورکمربه دور هیپ	۰/۱۸۸	−۰/۰۶۵	−۰/۲۶۹	۰/۵۷۵	–	–	–	–
شاخص توده بدنی	۰/۱۵	−۰/۰۵۲	−۰/۲۱۴	۰/۴۵۸	–	–	–	–
فشارخون سیستول	۰/۳۶۶	−۰/۰۳۵	−۰/۱۴۷	۰/۳۱۵	۰/۰۰۶	۰/۰۰۳	۰/۰۴۹	۰/۹۹۴
فشارخون دیاستول	۰/۲۸۵	۰/۰۲۷	۰/۱۱۴	۰/۲۴۵	۰/۰۰۵	۰/۰۰۲	۰/۰۳۸	۰/۷۷۳

مراجع

[147] Ullman JB. *Structural equation modeling: Reviewing the basics and moving forward.* Journal of personality assessment. 2006;87 (1):35-50.

[148] Altindag I, Genc A. *Bayesian Nonlinear Structural Equation Modeling.* Journal of Selcuk University Natural and Applied Science. 2015;4(1):174-92.

[149] Song XY, Xia YM, Lee SY. *Bayesian semiparametric analysis of structural equation models with mixed continuous and unordered categorical variables. Statistics in medicine.* 2009;28(17):2253-76.

[150] Lee S-Y, Song X-Y. *Model comparison of nonlinear structural equation models with fixed covariates.* Psychometrika. 2003;68(1):27-47.

[151] Ghayour-Mobarhan M, Moohebati M, Esmaily H, Ebrahimi M, Parizadeh SMR, Heidari-Bakavoli AR, et al. *Mashhad stroke and heart atherosclerotic disorder (MASHAD) study: design, baseline characteristics and 10-year cardiovascular risk estimation.* International journal of public health. 2015;60(5):561-72.

[152] Kaplan D. *Structural equation modeling: Foundations and extensions:* Sage Publications; 2008.

[153] Sivo SA, Fan X, Witta EL, Willse JT. *The search for" optimal" cutoff properties: Fit index criteria in structural equation modeling.* The Journal of Experimental Education. 2006;74(3):267-88.

[154] Kline RB. *Principles and practice of structural equation modeling*: Guilford publications; 2015.

[155] Ahmad WMAW, Nawi MABA, Aleng NA, Halim NA, Mamat M, Hamzah MP, et al. *Association of Hypertension with Risk Factors Using Logistic Regression*. Applied Mathematical Sciences. 2014;8(52):2563-72.

[156] Van Onna M. *Bayesian estimation and model selection in ordered latent class models for polytomous items.* Psychometrika. 2002;67(4):519-38.

[157] Bardenheier BH, Bullard KM, Caspersen CJ, Cheng YJ, Gregg EW, Geiss LS. *A Novel Use of Structural Equation Models to Examine Factors Associated With Prediabetes Among Adults Aged 50 Years and Older.* Diabetes Care. 2013;36(9):2655-62.

[158] Cois A, Ehrlich R. *Analysing the socioeconomic determinants of hypertension in South Africa: a structural equation modelling approach.* BMC public health. 2014;14(1):1.

[159] Shen B-J, Todaro JF, Niaura R, McCaffery JM, Zhang J, Spiro III A, et al. *Are metabolic risk factors one unified syndrome? Modeling the structure of the metabolic syndrome X.* American Journal of Epidemiology. 2003;157(8):701-11.

[160] Nock NL, Wang X, Thompson CL, Song Y, Baechle D, Raska P, et al., editors. *Defining genetic determinants of the Metabolic Syndrome in the Framingham Heart Study using association and structural equation modeling methods.* BMC proceedings; 2009: BioMed Central.

[161] Bramlage P, Pittrow D, Wittchen H-U, Kirch W, Boehler S, Lehnert H, et al. *Hypertension in overweight and obese primary care patients is highly prevalent and poorly controlled.* American journal of hypertension. 2004;17(10):904-10.

[162] Kokkinos PF, Giannelou A, Manolis A, Pittaras A. *Physical activity in the prevention and management of high blood pressure.* Hellenic J Cardiol. 2009;50(1):52-9.

[163] Mousavian As, SHaherian S, Namvar f, GHanbarzadeh m. *Comparing effect of walking and selected aerobic exercise on blood pressure in overweight postmenopausal women.* Journal of Nursing and Midwifery Urmia University of Medical Sciences. 2014;12(16):427-35. eng @ 2008-6326, 2014.

مدلسازی معادلات ساختاری بیزی و کاربرد آن در مدلسازی عوامل خطر بیماری ها

راضیه یوسفی [۱]، کارشناسی ارشد آمارزیستی، گروه اپیدمیولوژی و آمارزیستی، دانشگاه علوم پزشکی مشهد آزاده ساکی استادیار آمارزیستی، گروه اپیدمیولوژی و آمارزیستی، دانشگاه علوم پزشکی مشهد

چکیده: دلسازی معادله ساختاری، ابزاری قدرتمند در دست پژوهشگر است که وی را در چگونگی تدوین مبانی و چارچوب نظری پژوهش در قالب مدل های اندازه گیری و ساختاری یاری می‌رساند. اما این تکنیک نیاز به برقراری پیش فرض‌هایی دارد که مشکلاتی ایجاد می‌کند. به منظور حل این مشکلات رویکرد بیزی درمورد تحلیل بسیاری از مدلهای معادله‌ی ساختاری پیشرفته پیشنهاد شد. در این مطالعه سعی شده است ضمن معرفی رویکرد بیزی مدلسازی معادلات ساختاری به بیان نتایج این روش در داده‌های واقعی نیز پرداخته شود.

مواد و روش ها

در این مطالعه مقطعی مبتنی بر جمعیت، مدلسازی معادلات ساختاری بیزی با استفاده از روش نمونه گیری گیبس معرفی و در داده‌های واقعی مطالعه مشهد بکار گرفته شد. مدل‌های بیزی با پیشین های فاقد اطلاع نرمال، یکنواخت و پیشین حاوی اطلاع به داده ها برازش یافتند.

نتایج

در این مطالعه با توجه به نزدیک بودن برآوردهای دو مدل فاقد اطلاع این نتیجه حاصل شد که انتخاب توزیع فاقد اطلاع تاثیر چندانی در برآورد نداشت و در اینحالت برآوردها براساس اطلاعات موجود در دادهها استخراج شدند. مدل بیزی حاوی اطلاع زمان بیشتری برای رسیدن به همگرایی نیاز داشت. اما بررسی آماره DIC مدل ها نشان داد که مدل های بیزی فاقد اطلاع با پیشینهای یکنواخت و نرمال تعریف شده یکسان بودند اما با بکارگیری اطلاعات پیشین شاخص DIC به اندازهی ۱۳۹ واحد کاهش یافت.

نتیجه گیری

نتایج این مطالعه نشان داد که بکارگیری رویکرد بیزی معادلات ساختاری نتایجی حداقل در سطح رویکرد کلاسیک ارائه می‌کند و در صورتی که اطلاعات پیشین مناسبی وجود داشته باشد منجر به نتایج مناسب تر و دستیابی به برآوردها در شرایط دشوار می‌گردد.

واژه‌های کلیدی: مدلسازی معادلات ساختاری، تکنیک بیزی، نمونه گیر گیبس

[۱] راضیه یوسفی:

۱ مقدمه

مدلسازی معادله ساختاری، ابزاری قدرتمند در تدوین مبانی و چارچوب نظری پژوهش در قالب مدل‌های اندازه‌گیری و ساختاری با فرضیات معین است.((۱۶۷)–(۱۶۴)). اما اغلب ملاحظه می‌شود که در داده‌های جمع آوری شده در یک مطالعه این فرضیات برقرار نیستند و بنابراین دراین شرایط نمیتوان روش درستنمایی ماکزیمم را بکار برد زیرا شاخصهای برازش مدل، برآورد پارامترها و خطاهای استاندارد با افزایش عدم نرمالیتی، هرچه بیشتر اریب خواهند شد (۱۶۸). اثرات تخلف از فرضیات نرمالیتی باعث بوجود آمدن مقادیر بزرگ کیدو و در نتیجه رد بسیاری از مدلها خواهد شد. (۱۶۹).

به منظور حل این مشکلات رویکرد بیزی با چند مزیت در مورد تحلیل بسیاری از مدلهای معادله‌ی ساختاری پیشرفته پیشنهاد شد. اول اینکه روش‌های بیزی دامنه فرضیات مورد آزمون را گسترش می‌دهد و نتایج می‌توانند به طرق شهودی و بدون اتکا به آزمون‌های معناداری فرض صفر تفسیر شوند. دوم، برآورد بیزی میتواند یافته‌های پیشین موجود در تحقیقات مشابه قبلی یا بدست آمده از متاآنالیز را با داده‌های جدید ترکیب کند. سوم، چون برآورد بیزی میتواند یافته‌های قبلی را بکار گیرد، داده‌های بدست آمده از مطالعات با حجم نمونه‌ی کم کمتر دچار مشکل خواهند شد. چهارم، روشهای بیزی امکان برآورد مدلهارا وقتی روشهای سنتی برآورد به علت پیچیدگی مدل ناکارا هستند، فراهم می‌کند ((۱۶۶)، (۱۷۰)، (۱۷۱)). توسعه و کاربرد رویکرد بیزی مدلسازی معادلات ساختاری، نسبتا کند بوده است اما با تکنولوژی جدید و روش‌هایی چون نمونه گیر گیبس، که امروزه به کمک نرم افزارها به آسانی قابل دستیابی است، امکان پذیر گشته است. ((۱۷۶)–(۱۷۲)). دراین مطالعه سعی شد پس از شرح مختصری از مفاهیم مربوط به رویکرد بیزی معادلات ساختاری به بیان نتایج بدست آمده از این روش در یک داده واقعی پرداخته شود.

۲ تعیین مدل معادلات ساختاری بیزی

فرض کنید مدل اندازه گیری به صورت زیر باشد:

$$y = \alpha + \Lambda\eta + Kx + \varepsilon$$

وقتی y بردازی از متغیرهای آشکار باشد.α برداری از عرض از مبدا مدل اندازه گیری، Λ ماتریس بارهای عاملی ،η برداری از متغیرهای پنهان، ماتریسی از ضرایب رگرسیون مربوط به متغیرهای آشکار به مشاهده شده x، ε بردارخطا و Σ ماتریس کوواریانس قطری می‌باشد. مدل ساختاری عامل های مشترک را به یکدیگر و به متغیرهای آشکار x مربوط می‌کند و بصورت زیر بیان می‌شود:

$$\eta = \nu + \mathrm{B}\eta + \Gamma x + \zeta$$

وقتی ν برداری از عرضازمبدا مدلساختاری،BوΓماتریسهای ضرایب ساختاری، و ζ برداری از خطاهای ساختاری با ماتریس کوواریانس Ψ می‌باشد.

۳ نمونه گیری زنجیره مارکوف مونت کارلو برای مدلهای معادلات ساختاری بیزی

روش های زنجیره مارکوف مونت کارلو عمومی ترین روش برای دستیابی به براورد بیزی در مسائل است (۱۷۷). نمونه گیر گیبس یکی از این روشهاست که ایده‌ی اولیه آن توسط متروپلیس و نلر در سال ۱۹۵۳ ارائه و سپس با مقاله ای که گیمان و گیمان در مورد مدل های پردازش تصویر ارائه دادند وارد مرحله جدیدی شد(۱۷۸). رویکرد بیزی با در نظر گرفتن η بعنوان داده گمشده شروع می‌شود. فرض کنیم $\eta = (\eta_1, \eta_2, \cdots, \eta_n)$ سپس، داده های مشاهده شده $Y = (y_1, y_2, \cdots, y_n)$ با η در تحلیل پسین ترکیب می‌شوند.

نمونه گیر گیبس توزیع پسین $[\theta, \eta|y]$ را به کمک الگوریتم زیر فراهم می‌کند. در تکرار $(s+1)$ مقادیر جاری $\left(\theta^{(s)}, \eta^{(s)}, Y^{(s)}\right)$ استفاده شده و نمونه گیر گیبس به ترتیب زیر اجرا می‌شود:

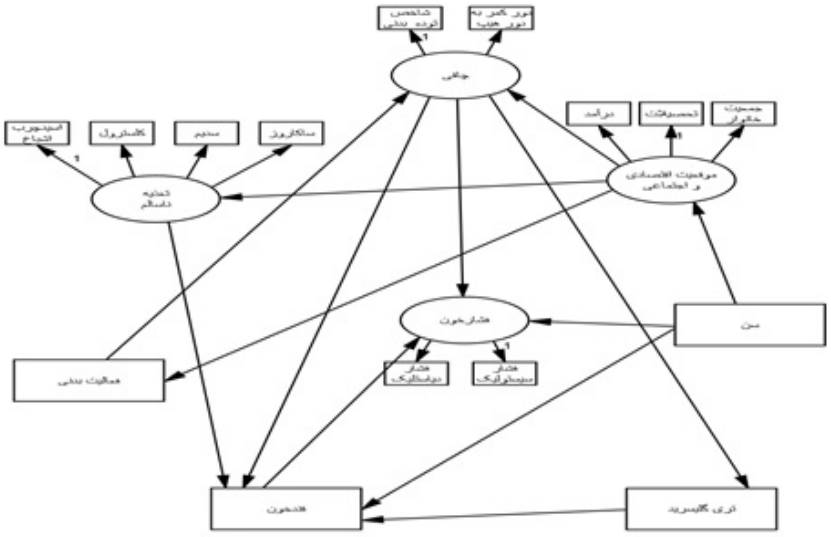
۱. تولید متغیر تصادفی $\eta^{(s+1)}$از توزیع شرطی $\left(\eta \mid \theta^{(s)}, Y^{(s)}\right)$

۲. تولید متغیر تصادفی $\theta^{(s+1)}$از توزیع شرطی $\left(\eta \mid \theta^{(s+1)}, Y^{(s)}\right)$

۳. تولید $\left(Y^{s+1}\right)$ از $P\left(Y \mid \eta^{(s+1)}, \theta^{(s+1)}\right)$ ((۱۷۹)،(۱۸۰)).

۴ مدلسازی معادلات ساختاری بیزی برای شناسایی عوامل خطر بیماری فشار خون

در این بخش اطلاعات مربوط به مطالعه مشهد مورد استفاده قرار گرفت(۱۸۱). یک مدل معادلات ساختاری برای براورد اثرات فاکتورهای خطر موثر بر فشارخون تعیین گردید. (نمودار ۱)



شکل ۱: نمودار مدل مفهومی تدوین شده

برازش مدل در هرسه حالت بیزی با پیشین فاقد اطلاع نرمال، یکنواخت و پیشین حاوی اطلاع انجام و مدل ها مقایسه گردیدند. (جدول۱ و ۲)

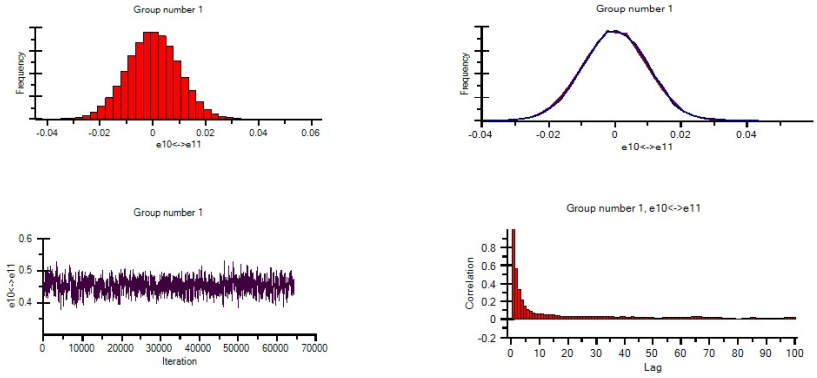
جدول ۱: مقایسه مدل های بیزی

مدل بیزی	تعداد نمونه تولید شده	تعداد زنجیره	PSR	$PPP-value$	DIC	سرعت همگرایی
پیشین فاقد اطلاع یکنواخت	۵۵۱۹۲	۲	۱/۰۰۱۸	۰/۵	۶۵۶۲۵/۳۶	۱
پیشین فاقد اطلاع نرمال	۵۶۵۰۱	۲	۱/۰۰۱۷	۰/۵۲	۶۵۶۲۶/۱۷	۲
پیشین حاوی اطلاع	۵۶۷۳۱	۶۴	۱/۰۰۱۸	۰/۵۵	۶۵۴۸۶/۴۲	۳

جدول ۲: میانگین‌های پسین با پیشین‌های متفاوت

مسیر			پیشین ۱	میانگین پسین	پیشین۲	میانگین پسین
فشارخون	←	فشارخون دیاستول	$N(۰, ۱۰^{۱۰})$	۰/۴۷۳	$N(۰, ۱۰^{۱۰})$	۰/۴۷۰
چاقی	←	شاخص دورکمربه دور هیپ	$N(۰, ۱۰^{۱۰})$	۰/۰۴۰	$N(۰, ۱۰^{۱۰})$	۰/۰۳۸
موقعیت اقتصادی اجتماعی	←	درآمد	$N(۰, ۱۰^{۱۰})$	۰/۵۳۶	$N(۰, ۱۰^{۱۰})$	۰/۵۴۱
سن	←	چاقی	$N(۰, ۱۰^{۱۰})$	۰/۴۰۳	$N(۰, ۱۰^{۱۰})$	۰/۴۱۰
سن	←	فشارخون	$N(۰, ۱۰^{۱۰})$	۰/۵۷۵	$N(۰, ۱۰^{۱۰})$	۰/۶۰۷
چاقی	←	فشارخون	$N(۰, ۱۰^{۱۰})$	۰/۵۱۷	$N(۰, ۱۰^{۱۰})$	۰/۴۳۱
فعالیت بدنی	←	چاقی	$N(۰, ۱۰^{۱۰})$	−۱/۸۰۰	$N(۰, ۱۰^{۱۰})$	−۱/۹۴۳
قندخون	←	فشارخون	$N(۰, ۱۰^{۱۰})$	۰/۲۳۰	$N(۰, ۱۰^{۱۰})$	۰/۲۳۹
موقعیت اقتصادی اجتماعی	←	فعالیت بدنی	$N(۰, ۱۰^{۱۰})$	−۰/۰۳۴	$N(۰, ۱۰^{۱۰})$	−۰/۰۳۵
چاقی	←	قندخون	$N(۰, ۱۰^{۱۰})$	۰/۰۴۱	$N(۰, ۱۰^{۱۰})$	۰/۰۳۵
تغذیه ناسالم	←	کلسترول	$N(۰, ۱۰^{۱۰})$	۱/۰۹۷	$N(۰, ۱۰^{۱۰})$	۱/۰۹۹
تغذیه ناسالم	←	قندخون	$N(۰, ۱۰^{۱۰})$	۰/۰۲۶	$N(۰, ۱۰^{۱۰})$	۰/۱۲۰
موقعیت اقتصادی اجتماعی	←	تغذیه ناسالم	$N(۰, ۱۰^{۱۰})$	۰/۱۱۶	$N(۰, ۱۰^{۱۰})$	−۰/۰۴۰
موقعیت اقتصادی اجتماعی	←	جمعیت خانوار	$N(۰, ۱۰^{۱۰})$	−۰/۰۶۸	$N(۰, ۱۰^{۱۰})$	−۰/۲۳۳
موقعیت اقتصادی اجتماعی	←	چاقی	$N(۰, ۱۰^{۱۰})$	−۰/۲۰۷	$N(۰, ۱۰^{۱۰})$	۰/۰۶۴
تری گلیسرید	←	قندخون	$N(۰, ۱۰^{۱۰})$	۰/۰۶۳	$N(۰, ۱۰^{۱۰})$	−۰/۲۲۸
سن	←	موقعیت اقتصادی اجتماعی	$N(۰, ۱۰^{۱۰})$	−۰/۲۲۵	$N(۰, ۱۰^{۱۰})$	۰/۰۶۳
سن	←	قندخون	$N(۰, ۱۰^{۱۰})$	۰/۰۶۲	$N(۰, ۱۰^{۱۰})$	۰/۲۲۸
چاقی	←	تری گلیسرید	$N(۰, ۱۰^{۱۰})$	۰/۲۷۰	$N(۰, ۱۰^{۱۰})$	۰/۳۵۱
تغذیه ناسالم	←	ساکاروز	$N(۰, ۱۰^{۱۰})$	۰/۳۵۷	$N(۰, ۱۰^{۱۰})$	۰/۹۲۵
تغذیه ناسالم	←	سدیم	$N(۰, ۱۰^{۱۰})$	۰/۹۲۳	$N(۰, ۱۰^{۱۰})$	۰/۰۲۷
کوواریانس خطاها						
فشارخون سیستولیک	و	دیاستولیک	-	-	$N(۰, ۰/۰۱)$	۰/۰۰۱
درآمد	و	تحصیلات	-	-	$N(۰, ۰/۰۱)$	۰/۰۱۶
درآمد	و	جمعیت خانوار	-	-	$N(۰, ۰/۰۱)$	۰/۰۰۲
تحصیلات	و	جمعیت خانوار	-	-	$N(۰, ۰/۰۱)$	۰/۰۰۱
اسیدچرباشباع	و	کلسترول	-	-	$N(۰, ۰/۰۱)$	-
اسیدچرباشباع	و	سدیم	-	-	$N(۰, ۰/۰۱)$	۰/۰۰۴
اسیدچرباشباع	و	ساکاروز	-	-	$N(۰, ۰/۰۱)$	۰/۰۰۱
کلسترول	و	سدیم	-	-	$N(۰, ۰/۰۱)$	۰/۰۰۵
کلسترول	و	ساکاروز	-	-	$N(۰, ۰/۰۱)$	۰/۰۰۱
سدیم	و	ساکاروز	-	-	$N(۰, ۰/۰۱)$	۰/۰۰۵

نمودار ۲ نمودارهای همگرایی مربوط به یکی از مسیرهای موجود را نشان می‌دهد. در این نمودارها همگرایی به وضوح دیده می‌شود. از آنجاییکه مقادیر نمونه‌ها هم مثبت و هم منفی هستند و در اطراف صفر نیز بصورت متقارن قرار گرفته‌اند بنابراین می‌توان نتیجه گرفت به لحاظ آماری فرضیه برابری این پارامترها با صفر می‌تواند پذیرفته شود. بنابراین پیش فرض مدلسازی معادلات ساختاری مبنی بر مستقل بودن خطاهای متغیرهای نشانگر به این ترتیب تامین شده است.



شکل ۲: نمودار کوواریانس خطای بین متغیرهای فشارخون سیستولیک و دیاستولیک

۵ بحث

در این مطالعه با توجه به اینکه توزیع متغیرها غیرنرمال شناخته شد و قبل از بکارگیری برآوردهای درستنمایی ماکزیمم نیاز به برقراری پیش فرض‌های زیادی وجود داشت تا برآوردهای نااریب بدست آیند، بنابراین رویکرد بیزی معادلات ساختاری مورد توجه قرار گرفت. مودن (۱۸۲) و مودنوآسپارو (۱۶۵) در مطالعات خود بیان داشتند که پارامترهایی چون کوواریانس خطاها که در رویکرد کلاسیک شناسایی‌پذیر نیستند و امکان برآورد آنها وجود ندارد به کمک تکنیک بیزی قابل برآورد هستند، به کمک تکنیک بیزی امکان آزمون مستقیم احتمال اینکه یک پارامتر در یک محدوده قرار گیرد، فراهم می‌گردد و از طرفی بکارگیری توزیع‌های پیشین به دستیابی به برآوردهای دقیق‌تری منجر می‌شود. در این مطالعه با توجه به نزدیک بودن برآوردهای دو مدل فاقد اطلاع این نتیجه حاصل شد که انتخاب توزیع فاقد اطلاع تاثیر چندانی در برآورد نداشت و در این حالت برآوردها براساس اطلاعات موجود در داده‌ها استخراج شدند. مطالعات مودن (۱۸۲)، جلمن(۱۸۳) و اشبی (۱۸۴) نیز بر این نکته اذعان داشتهاند که بکارگیری توزیع پیشین فاقد اطلاع موجب می‌شود برآوردهای پسین براساس تابع درستنمایی استخراج شود.

اما در مرحله بعد برازش مدل بیزی حاوی اطلاع به این ترتیب صورت گرفت که پیش فرض تکنیک معادلات ساختاری مبنی بر صفر بودن کوواریانس خطاهای مربوط به نشانگرهای یک عامل به عنوان پیشین بکار رفت. در رویکرد کلاسیک مدلسازی معادلات ساختاری امکان در نظر گرفتن فرض مستقل بودن خطای نشانگرها به علت شناسایی ناپذیر شدن مدل وجود ندارد، در حالیکه در مسائل واقعی ممکن است بین نشانگرهای یک عامل با توجه به ارتباط‌شان همبستگی وجود داشته باشد. به این منظور برای کوواریانس بین نشانگرهای یک عامل توزیع نرمال با میانگین صفرو واریانس ۰/۰۱ درنظرگرفته شدند. زیفور و اسوالد نیز در مطالعه خود که در زمینهی علوم رفتاری انجام شده است این پیشین را بکار گرفته و به مدل مناسبتری دست یافته اند (۱۷۱). مدل بیزی حاوی اطلاع زمان بیشتری برای رسیدن به همگرایی نیاز داشت. اما بررسی آماره DIC مدل ها نشان داد که مدل های بیزی فاقد اطلاع با پیشینهای یکنواخت و نرمال تعریف شده یکسان بودند اما با بکارگیری اطلاعات پیشین شاخص DIC به اندازهی ۱۳۹ واحد کاهش یافت. این کاهش نشان داد که وجود این همبستگی ها باعث مقدار هرچند جزئی کمبود برازش در مدل بوده است.

۶ نتیجه گیری

نتایج این مطالعه نشان داد که بکارگیری رویکرد بیزی معادلات ساختاری نتایجی حداقل در سطح رویکرد کلاسیک ارائه می‌کند و در صورتی که اطلاعات پیشین مناسبی وجود داشته باشد منجر به نتایج مناسبتر و دستیابی به برآوردها در شرایط دشوار می‌گردد. بنابراین با توجه به ماهیت رخدادهای طبیعی که منجر به بروز شرایط دشوار در تحلیل میگردد به نظر میرسد رویکرد بیزی میتواند به عنوان راه حلی در این شرایط مورد توجه قرار گیرد.

مراجع

[164] Hoyle RH. *Handbook of structural equation modeling*: Guilford Press; 2012.

[165] Muthen LK, Muthén BO. 2010 *Mplus user's guide. Muthén and Muthén,* Los Angeles. 1998.

[166] Altindag I, Genc A. *Bayesian Nonlinear Structural Equation Modeling*. Journal of Selcuk University Natural and Applied Science. 2015;4(1):174-92.

[167] Browne MW. *Generalized least squares estimators in the analysis of covariance structures*. ETS Research Bulletin Series. 1973;1973(1):i-36.

[168] Yanuar F, *editor The Estimation Process in Bayesian Structural Equation Modeling Approach*. Journal of Physics: Conference Series; 2014: IOP Publishing.

[169] Hancock G, Mueller R. *Structural Equation Modeling: A Second Course (Quantitative Methods in Education and the Behavioral Sciences)*. IAP-Information Age Publishing Inc, Charlotte. 2006.

[170] Dunson DB. *Bayesian latent variable models for clustered mixed outcomes*. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2000;62(2):355-66.

[171] Zyphur MJ, Oswald FL. *Bayesian Estimation and Inference A User's Guide*. Journal of Management. 2013:0149206313501200. English.

[172] Arminger G, Muthén BO. *A Bayesian approach to nonlinear latent variable models using the Gibbs sampler and the Metropolis-Hastings algorithm*. Psychometrika. 1998;63(3):271-300.

[173] Lee S-Y, Song X-Y. *Model comparison of nonlinear structural equation models with fixed covariates*. Psychometrika. 2003;68(1):27-47.

[174] Ansari A, Jedidi K. *Bayesian factor analysis for multilevel binary observations*. Psychometrika. 2000;65(4):475-96.

[175] Van Onna M. *Bayesian estimation and model selection in ordered latent class models for polytomous items*. Psychometrika. 2002;67(4):519-38.

[176] Song X-Y, Lee S-Y. *Analysis of structural equation model with ignorable missing continuous and polytomous data*. Psychometrika. 2002;67(2):261-88.

[177] Efron B. *The bootstrap and markov-chain monte carlo*. Journal of biopharmaceutical statistics. 2011;21(6):1052-62.

[178] Geman S, Geman D. *Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images*. IEEE Transactions on pattern analysis and machine intelligence. 1984 (6):721-41.

[179] Lee S-Y. *Structural equation modeling: A Bayesian approach*: John Wiley & Sons; 2007.

[180] Lee S-Y, Song X-Y. *Basic and advanced Bayesian structural equation modeling: With applications in the medical and behavioral sciences*: John Wiley & Sons; 2012.

[181] Ghayour-Mobarhan M, Moohebati M, Esmaily H, Ebrahimi M, Parizadeh SMR, Heidari-Bakavoli AR, et al. *Mashhad stroke and heart atherosclerotic disorder (MASHAD) study: design, baseline characteristics and 10-year cardiovascular risk estimation*. International journal of public health. 2015;60(5):561-72.

[182] Muthén B. *Bayesian analysis in Mplus: A brief introduction*. Unpublished manuscript www statmodel com/download/IntroBayesVersion. 2010;203.

[183] Gelman A, Carlin JB, Stern HS, Rubin DB. *Bayesian data analysis*: Chapman & Hall/CRC Boca Raton, FL, USA; 2014.

[184] Ashby D. *Bayesian statistics in medicine: a 25 year review*. Statistics in medicine. 2006;25(21):3589-631.