بسط رگرسیون منطقی برای تحلیل داده‌های طولی: رگرسیون منطقی انتقال

پروین سربخش[۱]، یدالله محرابی[۲]، فرید زایری[۳]، مریم‌السادات دانشپور[۴]

چکیده:

برای شناسایی و لحاظ کردن اثرات متقابل بین متغیرهای پیش‌بین در مدلهای رگرسیونی، می‌توان از روش رگرسیون منطقی استفاده کرد که در آن متغیرهای پیش‌بین جدید به صورت ترکیبات منطقی از متغیرهای دو حالتی اولیه ساخته و وارد مدل می‌شوند تا تقابل بین متغیرهای مستقل در قالب این ترکیب منطقی لحاظ شود. تاکنون رگرسیون منطقی برای تحلیل داده‌هایی با پاسخهای مستقل از هم معرفی و استفاده شده است ولی با وجود این‌که در مطالعات علمی مشاهدات همبسته به دلایل مختلف رخ میدهد، تابه حال رگرسیون منطقی برای تحلیل مشاهدات همبسته طولی انجام نشده است. به دلیل اهمیت بررسی اثرات متقابل در مطالعات طولی، در این مقاله مبانی نظری بسط رگرسیون منطقی برای تحلیل داده‌های طولی ارائه و مدل رگرسیون منطقی انتقال برای شناسایی اثرات متقابل تاثیرگذار بر متغیر پاسخ طولی دوحالتی پیشنهاد شده است. برنامه رایانه‌ای مربوط به مدل رگرسیون منطقی انتقال با به کارگیری معیار اطلاع آکائیکه (AIC) به عنوان تابع امتیاز مدل نوشته و پارامترهای مدل برآورد شده‌اند. برای ارزیابی عملکرد مدل پیشنهادی مطالعه شبیه‌سازی در سناریوهای مختلف اجرا شده که نتایج آن نشانگر عملکرد قابل قبول مدل پیشنهادی در یافتن اثرات متقابل موثر بر پاسخ طولی دوحالتی است. به عنوان مثال کاربردی نیز تحلیل ارتباط بین پلی‌مورفیسم‌ها و سایر عوامل خطر با سطح پایین HDL خون در طول زمان در مطالعه قند و لیپید تهران با مدل پیشنهادی انجام شده است.

واژه‌های کلیدی: پلی‌مورفیسم، رگرسیون منطقی، مدل انتقال با ساختار زنجیر مارکف، مطالعات طولی، HDL.

۱ مقدمه

یکی از روشهای مرسوم آماری برای آنالیز داده ها و بررسی ارتباط بین متغیرهای پیش بین و متغیر پاسخ، مدلهای رگرسیونی است. در اغلب کاربردهای عملی مدلهای رگرسیونی، یک مدل رگرسیونی می‌تواند ارتباط اثرات اصلی متغیرهای پیش‌بین را روی پاسخ نشان دهد. اثرات متقابل بین متغیرها همواره قابل شناسایی و لحاظ شدن در مدل نیستند و در صورت لحاظ شدن در مدل نیز به دلیل پیچیده شدن آن، از دوطرفه و نهایتا سه طرفه تجاوز نمی کنند.

زمانی که تعداد متغیرهای پیش بین زیاد باشد و به ویژه این متغیرها دو حالتی باشند اثرات متقابل مراتب بالاتر بین این متغیرها می‌تواند روی متغیر پاسخ تاثیر بگذارد. برای لحاظ کردن چنین تقابلهایی در مدلهای رگرسیونی، می‌توان به جای استفاده از تک تک متغیرها در برازش مدل، یک عبارت ترکیبی از آنها ساخت و به عنوان متغیر مستقل جدید وارد مدل کرد. رگرسیون منطقی یک روش رگرسیونی و کلاس‌بندی تعمیم یافته است که در آن متغیرهای جدید پیش‌بین به صورت ترکیبات بولی از متغیرهای دو حالتی اولیه ساخته و وارد مدل می‌شوند تا ارتباط و تقابل بین متغیرهای مستقل در قالب این ترکیبهای بولی ظاهر شود [۱۸].

تاکنون، رگرسیون منطقی برای مطالعات مورد-شاهدی یا هم‌گروهی با مشاهدات مستقل معرفی و استفاده شده است. به عنوان نمونه، اثرات متقابل موثر روی سرطان پروستات توسط اتزیونی و همکارانش با استفاده از رگرسیون منطقی گزارش شده است [۶]. همچنین نیز شوند[۵] در مقاله خود رگرسیون منطقی را برای تعیین ترکیبات موثر SNP-ها در ایجاد بیماری سرطان سینه در یک مطالعه مورد ـ شاهدی به کار بست [۱۷]. در یک مطالعه

[۱]استادیار گروه آمار و اپیدمیولوژی، دانشکده بهداشت، دانشگاه علوم پزشکی تبریز
[۲]استاد گروه اپیدمیولوژی، دانشکده بهداشت، دانشگاه علوم پزشکی شهید بهشتی
[۳]دانشیار گروه آمار زیستی، دانشکده پیراپزشکی، علوم پزشکی شهید بهشتی
[۴]دکترای آمار زیستی، مرکز تحقیقات پروتئومیکس، دانشگاه علوم پزشکی شهید بهشتی

[۵]Holger Schwender

نیز، رگرسیون منطقی لجستیک شرطی برای تحلیل داده‌های مورد-شاهدی جور شده برای تحلیل داده‌های اسکیزوفرنی بکار رفته است [۱۰].

با وجود این‌که مطالعات طولی با پاسخ‌های همبسته جزء مطالعات مرسوم در تحقیقات علمی هستند، تاکنون رگرسیون منطقی برای تحلیل این نوع از داده‌های هم‌بسته بسط داده نشده است. از اهدافی که با طراحی مطالعات طولی دنبال می شود شناسایی تاثیر متغیرهای پیش‌بین بر روی متغیر پاسخ در طول زمان است همچنین علاقه‌مند به بررسی این پرسش هستیم که آیا میزان ارتباط متغیر پاسخ با متغیرهای پیش‌بین در طول زمان تغییر می کند یا نه. به عبارت دیگر، آیا بین متغیرهای پیش‌بین و زمان اثر متقابل وجود دارد یا نه. طبیعی است که اگر بین متغیرهای پیش‌بین، اثرات متقابل مهم و تاثیرگذار بر متغیر پاسخ طولی وجود داشته باشد، و یا اگر اثر متقابل با زمان وجود داشته باشد، شناسایی و وارد کردن این اثرات متقابل باعث بهبود برازش مدل طولی خواهد شد. زمانی که تعداد متغیرهای پیش‌بین زیاد باشد و اثرات متقابل مراتب بالا نیز بالقوه وجود داشته باشد جستجوی دستی و شانسی این برهم‌کنش‌ها عملا بی‌فایده خواهد بود. لذا به نظر می‌توان برای شناسایی اثرات متقابل بین متغیرهای پیش‌بین دوحالتی در مدلهای طولی از تئوری رگرسیون منطقی استفاده شود. لذا با توجه به اهمیت بحث اثرات متقابل در تحلیل داده‌های طولی، هدف از این پژوهش تعمیم رگرسیون منطقی برای تحلیل داده‌های همبسته طولی است. جهت ارائه نحوه انجام تعمیم، ابتدا مدل رگرسیون منطقی و مدلهای طولی به اختصار مرور می‌شود و در ادامه نحوه بسط رگرسیون منطقی به داده‌های طولی و مدل پیشنهادی محققین توضیح داده خواهد شد.

۲ روش انجام تحقیق

۱.۲ ترکیبات بولی (منطقی)

با توجه به منطق جبر بولی، معمول‌ترین و آسان ترین راه نمایش ترکیبات بولی استفاده از عملگرهای $\vee(or),\quad \wedge(and), {}^{c}(not)$ و استفاده از کروشه ها می باشد. به عنوان مثال، اگر $X_1, X_2, \ldots, X_{10}$ متغیرهای پیش‌بین دو حالتی باشد، یک ترکیب بولی از این متغیرهای دوحالتی می‌تواند به صورت زیر ساخته شود که متشکل از ٦ متغیر دوحالتی اولیه است و با استفاده از عملگرها نشان داده می شود:

$$L = \left\{(X_1 \wedge X_2) \wedge \left[(X_3 \wedge X_{10}) \vee \left(X_5 \wedge (X_3^c \vee X_6)\right)\right]\right\}$$

با توجه به اینکه این ترکیب، یک ترکیب بولی از متغیرهای دوحالتی است، ارزش و مقدار آن نیز دوحالتی وبه صورت دو مقدار صفر یا یک خواهد بود [١٦].

۲.۲ رگرسیون منطقی

رگرسیون منطقی یک روش رگرسیونی تعمیم یافته و جدید است که در آن متغیرهای پیش‌بین به صورت ترکیبهای بولی از متغیرهای دو حالتی ساخته می‌شود این ترکیبهای ساخته شده نشان‌گر تقابل و تعامل بین متغیرهای پیش‌بین دوحالتی مورد بررسی هستند به طوری که متغیرهای پیش‌بینی که با هم در یک ترکیب ظاهر می‌شوند اثرات برهمکنشی با یکدیگر دارند که بر روی متغیر پاسخ تاثیرگذار است. به طور مشخص، در رگرسیون منطقی، به دنبال یک متغیر دو حالتی جدید هستیم که حاصل یک ترکیب منطقی بولی مطلوب از متغیرهای دو حالتی اولیه است طوری که استفاده از این متغیر جدید به عنوان متغیر پیش‌بین، در مقایسه با سایر ترکیبات بولی ممکن، بهترین برازش را برای متغیر پاسخ داشته باشد. این رگرسیون در زمینه داده‌های SNP، توالی ژنی و غربالگری بیماریهای چند عاملی کاربرد دارد و به دلیل استفاده از ترکیبات بولی منطقی رگرسیون منطقی (Logic Regression) نامیده شده است [١٤].

فرض کنید $X_1, X_2, \ldots, X_k$ متغیرهای پیش‌بین دو حالتی $Z_1, \ldots, Z_p$، متغیرهای پیش‌بین کمی و Y متغیر پاسخ است. مدل رگرسیون منطقی به صورت زیر تعریف می‌شود:

$$g\big(E(Y)\big) = \beta_0 + \sum_{i=1}^{p} \gamma_i Z_i + \sum_{j=1}^{t} \beta_j L_j$$

که در آن L_j یک عبارت بولی از متغیرهای پیش‌بین دوحالتی X_i و $g(E(Y))$ یک تابع پیوند است. Z_i نیز بردار مربوط به متغیرهای پیش‌بین کمی است که قابلیت حضور در مدل را دارند. قالب ارائه شده فوق می‌تواند شامل رگرسیون خطی $g\big(E(Y)\big) = E(Y)$ و

رگرسیون لجستیک $g(E(Y)) = \log \frac{E(Y)}{1-E(Y)}$ مدل مخاطرات متناسب کاکس یا سایر مدلهای خطی تعمیم یافته باشد. به عبارت دیگر متغیر پاسخ (Y) می‌تواند ازهر نوعی (به عنوان مثال کمی، کیفی، دو حالتی، زمان بقا) باشد و متغیرهای پیش‌بین جدید مدل (Lj) که به صورت ترکیبات بولی از متغیرهای پیش‌بین دوحالتی (Xj) ساخته می‌شوند، از طریق تابع پیوند $g(E(Y))$ با متغیر پاسخ در ارتباط هستند. به طور کلی، برای هر مدلی یک تابع امتیازتعریف می‌شود که نشان دهنده کیفیت مدل مفروض است. برای مثال در رگرسیون خطی تابع امتیاز، مجموع مربعات خطا و در رگرسیون لجستیک آماره انحراف[۶] است. به شرط تعریف تابع امتیاز مناسب، مدل رگرسیون منطقی قابلیت تعمیم به انواع مدلهای خطی تعمیم یافته را دارد [۱۸].

در رگرسیون منطقی هدف یافتن عبارت بولی است به طوری که تابع امتیازتعیین شده مدل را مینیمم کند. در عمل می‌توان با یک سری متغیر داده شده تعداد بسیار زیادی ترکیب منطقی ساخت و روش مستقیم نیز برای فهرست کردن همه ترکیبات منطقی وجود ندارد که بتوان برای گزینش بهترین مدل، همه پیش‌بینی های متفاوت رادر اختیار داشت پس امکان ارزیابی کامل همه ترکیبات منطقی ممکن وجود ندارد. لذا برای یافتن چنین ترکیبات منطقی از الگوریتم‌های جستجویی مثل الگوریتم Simulated Annealing استفاده می‌شود. الگوریتم Simulated Annealing یک الگوریتم جستجوی تصادفی است که در فضای حالتهای ممکن ترکیبات منطقی، بر مبنای تابع امتیاز تعیین شده به دنبال یافتن بهترین ترکیبات منطقی است به طوری که با حضور این ترکیبات منطقی یافت شده در مدل به عنوان متغیرهای پیش‌بین جدید تابع امتیاز مدل مینیمم شود [۱۴].

۳.۲ مطالعات طولی

مشخصه مطالعات طولی این است که افراد به طور مکرر در طول زمان اندازه گیری می‌شوند. از این نظر در تقابل با مطالعات مقطعی هستند که در آنها برای هر فرد تنها یک پیامد اندازه گیری می‌شود. مشاهدات افراد در طول زمان باعث ایجاد همبستگی در بین پاسخ ها برای هر فرد می‌شود که در نظر نگرفتن این جمله باعث بیش برآورد شدن واریانس خواهد شد. یکی از انواع مدلهایی که برای تحلیل داده های طولی استفاده می‌شود مدلهای انتقال یا مارکف هستند که در این مدلها مساله همبستگی موجود بین پاسخهای هر واحد را با مدلبندی پاسخ در زمان حال به شرط پاسخ در زمانهای گذشته بررسی می‌کنند. در این روش، پاسخ های تکراری به صورت یک فرایند تصادفی در نظر گرفته شده و مدلبندی بر این اساس انجام می‌گیرد [۵].

در این مقاله، "مدل حاشیه ای انتقال زنجیر مارکف با پاسخ دو حالتی" که توسط گونکالوس[۷] و همکاران معرفی شده است [۸] جهت توسعه رگرسیون منطقی برای تحلیل داده‌های طولی انتخاب شد. در این مدل برای بیان رفتار همبستگی سریالی بین مشاهدات پی در پی از نسبت شانس استفاده می‌شود و برای ساختن مدل لجستیک برای احتمال حاشیه‌ای موفقیت با مقادیر پارامترهای وابستگی داده شده تاخیر -۱ و تاخیر -۲ زنجیر مارکف مرتبه اول و دوم پارامتری می‌شود. به دلیل این‌که در این مدل، متغیر پاسخ به صورت حاشیه‌ای در مقابل متغیرهای پیش‌بینی مدل‌بندی می‌شود و در نتیجه می‌توان شدت اثر هر متغیر را بر پاسخ به صورت حاشیه‌ای بررسی کرد و از طرف دیگر برای میزان وابستگی مشاهدات به یکدیگر، یک مقدار عددی برآورد می‌شود به نظر برای داده‌های طولی که همبستگی بین مشاهدات در طول زمان معمولاً از خاصیت خودبازگشتی[۸] برخوردار است مدل مناسبی است لذا برای توسعه رگرسیون منطقی برای تحلیل داده‌های طولی، این مدل انتخاب شد. از این رو در ادامه توضیحاتی راجع به این مدل داده می‌شود.

۴.۲ مدل‌های حاشیه‌ای از طریق زنجیره‌های مارکف دوحالتی مرتبه اول و دوم

فرض کنید $y_{it} \in \{0,1\}$ که پاسخ فرد i-ام $(i = 1, \ldots, n)$ در زمان t $(t = 1, \ldots, T_i)$ است، مقادیر تصادفی تولید شده با میانگین $P(y_{it} = 1) = \theta_{it}$ باشد. تمام مشاهدات مربوط به فرد i-ام به

[۶] Deviance

[۷] Goncalves

[۸] Atou Regressive

این صورت است: $(y_{it},\dots,y_{iT_i})$. متناظر با فرد i–ام در زمان t–ام، تعداد p متغیر پیش‌بین موجود است که با بردار X_{it} نشان داده می‌شود. هدف برازش یک مدل رگرسیون لجستیک است که متغیرهای کمکی و توزیع احتمال پاسخ را به این صورت به هم ربط دهد:

$$\log itP(Y_{it}=1)=\log it\theta_{it}=x_{it}^T\beta.$$

در مدل زنجیر مارکف مرتبه اول با پاسخ دوحالتی، توزیع مشترک $(Y_1,\dots,Y_T)$ با توزیع Y_1 مجموعه‌ای از ماتریس انتقال یک مرحله‌ای، یا به طور معادل مجموعه‌ای از احتمالات شرطی مشخص می‌شود:

$$p_j=P\{Y_t=1\mid Y_{t-1}=j\},\quad (t=2,\dots,T),(j=0,1).$$

برای یک جفت از مشاهدات پی در پی (Y_t,Y_{t-1}) به طوری که توزیع حاشیه‌ای Y_{t-1} معلوم باشد، p_j طوری انتخاب می‌شود که $\theta_t=E(4)$ مقدار از قبل تخصیص داده شده‌ای باشد. از آن‌جایی که دو p_j تحت کنترل وجود دارد، این اجازه را می‌دهد تا جزء دیگری از توزیع مشترک برای تنظیم وابستگی انتخاب شود. در تحلیل داده‌های دو حالتی، کمیت ارجح برای اندازه‌گیری وابستگی بین متغیرها نسبت شانس است:

$$OR(Y_t,Y_{t-1})=\frac{P\{Y_{t-1}=Y_t=1\}P\{Y_{t-1}=Y_t=0\}}{P\{Y_{t-1}=0,Y_t=1\}P\{Y_{t-1}=1,Y_t=0\}}=\frac{p_1/(1-p_1)}{p_0/(1-p_0)}.$$

برای این تنظیمات لازم است p_1 و p_0 طوری به دست آیند که در معادلات زیر صدق کنند:

$$\psi_1=OR(Y_t,Y_{t-1})=\frac{p_1/(1-p_1)}{p_0/(1-p_0)},$$

$$\theta_1=\theta_{t-1}p_1+(1-\theta_{t-1})p_0.$$

لذا احتمالات انتقال یک گام به این صورت محاسبه می‌شوند:

$$p_j=\frac{(2j-1)[1-\delta_0+(\psi_1-1)\theta_{t-1}]+(\psi_1-1)\theta_t}{2(\psi_1-1)[1-j+(2j-1)\theta_{t-1}]},\quad j=(0,1)$$

که در آن:

$$\delta_0^2=1+(\psi_1-1)\{\psi_1(\theta_1-\theta_{t-1})^2-(\theta_t+\theta_{t-1})^2+2(\theta_t+\theta_{t-1})\}.$$

در حالت $\psi_1=1$ متغیرها مستقل از هم هستند و $p_j=\theta_t$ است[۸].

در مدل فوق، هیچ الزامی به ثابت نگه داشتن ψ_1 در طول زمان وجود ندارد. بنابراین، رویکرد فوق می‌تواند به زنجیره مارکف مرتبه دوم گسترش یابد. یعنی مشخصاً، هدف تولید نباله $(Y_1,\dots,Y_T)$ است به طوری که $E(Y_T)=\theta_t$ و θ_t–ها مقادیر از قبل تخصیص داده شده دارند $(t=1,\dots,T)$ و ساختار وابستگی فرایند نیز از نوع زنجیر مارکف مرتبه دوم است. این تنظیمات و ویژگی‌ها منجر به در نظر گرفتن توزیع مشترک سه جزء این فرایند در زمان‌های (Y_{t-2},Y_{t-1},Y_t) می‌شود. جدول احتمالات ۲×۲×۲ مربوطه با استفاده از هفت مولفه مشخص می‌شود با توجه به این محدودیت که مجموع باید یک باشد و سه محدودیت القاء شده توسط میانگین‌های حاشیه‌ای $(\theta_{t-2},\theta_{t-1},\theta_t)$، چهار مؤلفه برای تعیین وابستگی ساختار باقی می‌ماند. محدودیت‌ها و شرایط زیر جهت تعیین وابستگی ساختار باقی می‌ماند:

$$OR(Y_{t-1},Y_{t-2})=\psi_1=OR(Y_{t-1},Y_t)\quad (۱)$$

$$OR(Y_{t-2},Y_t|Y_{t-1}{=}0){=}\psi_2=OR(Y_{t-2},Y_t|Y_{t-1}{=}1),\quad (۲)$$

که در آن ψ_1 و ψ_2 دو مقدار مثبت از قبل تخصیص داده شده است و به ترتیب میزان وابستگی مرتبه اول و میزان وابستگی مرتبه دوم نامیده می‌شوند.

احتمالات انتقال نیز به صورت زیر تعریف می‌شوند:

$$p_{hj}=P(Y_t=1\mid Y_{t-2}=h,Y_{t-1}=j),\qquad h,j=0,1,\qquad (t=3,\dots,T).$$

بنابراین مساله تبدیل به یافتن احتمالات انتقال دو گام p_{hj} می‌شود طوری که واجد شرایط مذکور باشد. برای محاسبه این احتمالات انتقال با استفاده از نمادگذاری‌های زیر:

$$p'_j=p\{Y_t=1\mid Y_{t-2}=j\},\qquad p''_j=p\{Y_t=1\mid Y_{t-2}=j\},$$

شرایط (۱) و (۲) به:

$$\frac{p'_1(1-p'_0)}{p'_0(1-p'_1)}=\psi_1=\frac{p''_1(1-p''_0)}{p''_0(1-p''_1)}$$

و

$$\frac{p'_{10}(1-p'_{00})}{p'_{00}(1-p'_{10})}=\psi_2=\frac{p''_{11}(1-p''_{01})}{p''_{01}(1-p''_{11})}$$

تبدیل می‌شود. با در نظر گرفتن شرایط مذکور، باید میانگین حاشیه‌ای Y_t برابر:

$$\theta_t = p_{11}p_1''\theta_{t-2} + p_{10}(1-p_1'')\theta_{t-2} + p_{01}p_0''(1-\theta_{t-2}) + p_{00}(1-p_0'')(1-\theta_{t-2}),$$

باشد، همچنین در نظر گرفتن توزیع توام بین $(Y_{t'}Y_{t-1})$ به:

$$p_1'\theta_{t-2} = p_{11}p_1''\theta_{t-2} + p_{01}p_0''(1-\theta_{t-2}),$$

منجر می‌شود که با انجام عملیات جبری این احتمالات انتقال محاسبه خواهد شد [۸].

در مدل انتقال زنجیر مارکف با پاسخ دو حالتی استنباط درستنمایی بر پایه اطلاعات مربوط به نمونه n-تایی از افراد که مستقل از هم فرض می‌شوند انجام می‌شود. اگر y_{it} مشاهده t-ام مربوط به فرد i-ام، $(t = 1,\ldots,T)$، $(i = 1,\ldots,n)$ و y_{i0} تمام مشاهدات مربوط به فرد i-ام باشد، سهم فرد i-ام در لگاریتم درستنمایی برای پارامترهای (β,λ) که در آن $\lambda = (\log\psi_1, \log\psi_2)$، به این صورت می‌باشد [۸]:

$$\begin{aligned} loglikelihood(\beta,\lambda \mid y_{i0}) &= \left[y_{i1}\log it(\theta_{i1}) + \log(1-\theta_{i1})\right] \\ &+ \left[y_{i2}\log it(p'_{y_{i1}})\right] + \log(1-p'_{y_{i1}}) \\ &+ \sum_{t=3}^{T}\left[y_{it}\log it(p_{y_{it-2},y_{it-1}}) \right. \\ &\left. + \log(1-p_{y_{it-2},y_{it-1}})\right], \end{aligned}$$

که در آن کروشه اول مربوط به سهم مشاهده اول، کروشه دوم مربوط به سهم مشاهده دوم و کروشه سوم مربوط به سهم مشاهدات $(Y_3,\ldots,Y_T)$ در تابع درستنمایی مدل است. این تابع درستنمایی از طریق θ-ها و احتمالات انتقال یک گام و دوگام (p,p') با متغیرهای پیش‌بین ارتباط پیدا می‌کند. واضح است، تابع لگاریتم درستنمایی برای کل نمونه، از مجموع لگاریتم درستنمایی تک‌تک افراد نمونه به‌دست می‌آید:

$$loglikelihood = \sum_{i=1}^{n} ll(\beta,\lambda \mid y_i).$$

آماره AIC برای مدل فوق به شکل

$$AIC = -2loglikelihood + 2q,$$

محاسبه می‌شود که در آن q تعداد پارامترهای موجود در مدل است.

۵.۲ مدل پیشنهادی: مدل رگرسیون منطقی انتقال

با توجه به این‌که رگرسیون منطقی تاکنون برای تحلیل داده‌ها با پاسخ‌های طولی بسط داده نشده است، لذا در تحقیق حاضر بر مبنای مدل انتقال ذکر شده در بخش قبلی، "مدل رگرسیون منطقی انتقال" در قالب مدل رگرسیون منطقی با تابع پیوند

$$\log itP(Y_{it}=1) = \log it\theta_{it} = Z_{it}^T Y\gamma_{it} + L_{it}^T\beta_{it}$$

و استفاده از آماره AIC به عنوان تابع امتیاز مدل، معرفی شده به طوری که در آن $y_{it} \in \{0,1\}$ پاسخ فرد i-ام، $(i = 1,\ldots,n)$، در زمان t، $(t = 1,\ldots,T_i)$، بوده و هر مشاهده دارای میانگین $P(Y_{it}=1) = \theta_{it}$ است. L_{it} بردار ترکیب‌های منطقی حاصل از متغیرهای پیش‌بین دو حالتی X_{it} بوده و Z_{it} نیز بردار مربوط به متغیرهای پیش‌بین کمی است که قابلیت حضور در مدل را دارند. ساختار وابستگی بین مشاهدات متغیر پاسخ به طور مشابه با مدل انتقال به صورت زنجیر مارکف مرتبه اول (ψ_1) و مرتبه دوم (ψ_2) تعریف می‌شود. همچنین تابع درستنمایی و میزان AIC مدل نیز به طور مشابه با مدل انتقال با در نظر گرفتن ترکیبات منطقی یافت شده به عنوان متغیرهای پیش‌بین جدید مدل محاسبه می‌شوند. بنابراین الگوریتم جستجوی Annealing با جستجو در فضای ترکیبات منطقی ممکن، ترکیب‌های منطقی را یافته و شناسایی می‌کند که طبق آمار AIC حاصل از مدل رگرسیون منطقی انتقال، کمترین امتیاز و در نتیجه بهترین برازش را در مدل رگرسیون منطقی انتقال دارند.

برنامه مربوط به مدل رگرسیون منطقی انتقال در محیط FORTRAN 77 نوشته شد و پارامترهای این مدل با استفاده از الگوریتم DALL که یک برنامه ماکزیمم‌سازی توابع درستنمایی است [۹] برآورد شد. برنامه نوشته شده با بسته نرم‌افزاری مربوط به رگرسیون منطقی LogicReg [۱۴] ادغام شد. بسته نرم‌افزاری ادغام شده در محیط سیستم عامل لینوکس مجدداً کامپایل شد تا قابل اجرا در محیط R باشد و در نهایت داده‌ها که با استفاده از این بسته نرم‌افزاری تصحیح و تکمیل شده که رگرسیون منطقی انتقال در آن گنجانده شده بود، در محیط R(2.14.1) تجزیه و تحلیل شدند. عملکرد مدل نیز در شرایط و سناریوهای مختلفی شامل میزان‌های همبستگی متفاوت، شدت اثر تقابل بین متغیرها،

تقابل‌های مختلف و حجم‌نمونه‌های متفاوت با مطالعه شبیه‌سازی بررسی شد.

۳ مطالعه شبیه‌سازی

برای ارزیابی مدل رگرسیون منطقی انتقال پیشنهاد شده در این تحقیق و مقایسه تئوریک آن با مدل حاشیه‌ای انتقال صرفاً با اثرات اصلی به عنوان مدل مرسوم، براساس توزیع‌های احتمالی دوجمله‌ای و با در نظر گرفتن اثر متقابل بین متغیرهای مستقل دوحالتی و نیز میزان وابستگی بین اندازه‌های تکراری متغیرهای پاسخ با تعداد ۳ تکرار برای پاسخ، داده‌های مورد نیاز از مدل انتقالی زنجیر مارکف مرتبه اول شبیه‌سازی شد. به این ترتیب که برای هر متغیر پاسخ، ۱۰ متغیر مستقل دو حالتی X_1 تا X_{10} از توزیع برنولی با احتمال موفقیت p برابر ۰/۵ تولید شدند. در مطالعه شبیه‌سازی برای هر متغیر پاسخ یک اثر متقابل بین دو متغیر مستقل برنولی به صورت $X_1 \vee X_2$ به صورت ترکیب منطقی L تعریف شد که برای هر تکرار به ترتیب L_1 و L_2 و L_3 نامگذاری شدند.

برای متغیر پاسخ، سه اندازه تکراری $(t = 1, 2, 3)$ هر کدام با احتمال موفقیت از پیش تعیین شده θ_t که مرتبط با اثر متقابل ایجاد شده L است، به صورت

$$\theta_t = 1/(1 + \exp(-\beta * L_t)$$

در نظر گرفته شد. مقدار همبستگی یا نسبت شانس در زنجیر مارکف مرتبه اول بین پاسخ‌های متوالی پاسخ برابر ψ_1 در نظر گرفته شد. سپس احتمالات انتقال یک گام طبق روابط زیر محاسبه شد

$$\theta_t = \theta_{t-1}p_1 + (1 - \theta_{t-1})p_0 (26.3),$$

$$\psi_1 = \frac{p_1/(1-p_1)}{p_0/(1-p_0)},$$

$$p_j = \frac{(2j-1)[1-\delta_0+(\psi_1-1)\theta_{t-1}]+(\psi_1-1)\theta_t}{2(\psi_1-1)[1-j+(2j-1)\theta_{t-1}]}, \quad j=(0,1)$$

$$\delta_0^2 = 1+(\psi_1-1)\{\psi_1(\theta_t-\theta_{t-1})2-(\theta_t+\theta_{t-1})2+2(\theta_t+\theta_{t-1})\}.$$

با شروع از اولین پاسخ، y_1 از توزیع برنولی با میانگین θ_1 تولید شد.

$$y_1 \sim bernulli(\theta_1).$$

سپس شبیه‌سازی y_2 به این صورت انجام گرفت که برای y_1 های برابر یک، y_2 از توزیع برنولی با احتمال p_1 و برای y_1 های برابر صفر، y_2 از توزیع برنولی با میانگین p_0 تولید شد به صورتی که در نهایت احتمال موفقیت y_2 برای تمام y_1 ها برابر θ_2 شد:

$$p_1 = p(y_2 = 1 \mid y_1 = 1),$$

$$p_0 = p(y_2 = 1 \mid y_1 = 0),$$

$$p_j = \frac{(2j-1)[1-\delta_0+(\psi_1-1)\theta_1]+(\psi_1-1)\theta_2}{2(\psi_1-1)[1-j+(2j-1)\theta_1]}, \quad j=(0,1)$$

$$\left\{ \begin{array}{l} y_1 = 1 \Rightarrow y_2 \sim bernulli(p_1) \\ y_1 = 01 \Rightarrow y_2 \sim bernulli(p_0) \end{array} \right\}.$$

سپس شبیه‌سازی y_3 به این صورت انجام گرفت که برای y_2 های برابر یک، y_3 از توزیع برنولی با احتمال p'_1 و برای y_2 های برابر صفر، y_3 از توزیع برنولی با میانگین p'_0 تولید شد به صورتی که در نهایت احتمال موفقیت y_3 برای تمام y_2 ها برابر θ_3 شد.

$$p'_1 = p(y_3 = 1 \mid y_2 = 1),$$

$$p'_0 = p(y_3 = 1 \mid y_2 = 0),$$

$$p'_j = \frac{(2j-1)[1-\delta_0+(\psi_1-1)\theta_2]+(\psi_1-1)\theta_3}{2(\psi_1-1)[1-j+(2j-1)\theta_2]}, \quad j=(0,1)$$

$$\left\{ \begin{array}{l} y_1 = 1 \Rightarrow y_2 \sim bernulli(p_1) \\ y_1 = 01 \Rightarrow y_2 \sim bernulli(p_0) \end{array} \right\}.$$

به این ترتیب با وجود اثر متقابل L، پاسخ‌های همبسته‌ای تولید شد که دارای میزان همبستگی مشخص و احتمالات انتقال متناسب با میانگینشان هستند. برای تمام میزان‌های وابستگی مراحل فوق با انجام محاسبات مربوط به آن، میزان وابستگی اجرا شد.

تعداد داده‌های شبیه‌سازی شده برای هر زمان برابر ۵۰، ۲۰۰، ۵۰۰ و ۱۰۰۰ بود که برای هر سری از داده‌ها نیز میزان وابستگی بین مشاهدات متوالی (نسبت شانس) در ۵ حالت یعنی میزان وابستگی برابر: $\psi_1 = 0.2$، $\psi_1 = 0.5$، $\psi_1 = 1$، $\psi_1 = 2$ و $\psi_1 = 5$ در نظر گرفته شد. شدت تاثیر اثر متقابل با پاسخ نیز در چهار حالت $\beta = 0$، $\beta = 0.5$، $\beta = 1.5$ و $\beta = 3$ در نظر گرفته شد.

داده‌های تولید شده با استفاده از بسته نرم‌افزاری ساخته شده برای مدل پیشنهادی در محیط R تحلیل و برای ارزیابی مدل پیشنهادی و مقایسه تئوریک آن با مدل انتقالی مرسوم با اثرات اصلی، از این تحلیل‌ها استفاده شد. معیار ارزیابی عملکرد مدل

پیشنهاد شده طبق داده‌های شبیه سازی، درصد تشخیص صحیح اثر متقابل واقعی در نظرگرفته شد یعنی مدل رگرسیون منطقی انتقالی از 500 سری داده شبیه‌سازی شده در چند درصد مواقع قادر به تشخیص صحیح اثر متقابل در نظر گرفته شده می‌باشد. در این مدل‌ها برای بررسی میزان صحت و دقت برآوردها از شاخص‌های درصد اریبی نسبی[۹] [۱۳] و MSE استفاده شد. جهت مقایسه نیکویی برازش مدل پیشنهادی و مدل انتقال مرسوم نیز میانگین مقادیر AIC حاصل از مدل‌های رگرسیون منطقی انتقال که قادر به تشخیص صحیح اثر متقابل بودند با میانگین مقادیر AIC مدل‌های انتقال حاصل از اثرات اصلی به عنوان مدل مرسوم و معمول مقایسه شد.

۱.۳ نتایج حاصل از مطالعه شبیه‌سازی

نتایج مربوط به نتایج مطالعه شبیه‌سازی که در جداول ۱ تا ٤ گزارش شده است، نشان می‌دهد که در سری اول شبیه‌سازی با اثر متقابل ساده $L = X_1 \vee X_2$ در همه سناریوها، مقدار AIC مدل رگرسیون منطقی انتقال از مدل انتقال با اثرات اصلی بهتر است و این به دلیل شناسایی و لحاظ شدن اثر متقابل $L = X_1 \vee X_2$ در مدل می‌باشد زیرا این اثر متقابل شامل اثرات اصلی X_1 و X_2 و همچنین اثر تعاملی بین X_1 و X_2 است که در مدل انتقال با اثرات اصلی این تعامل لحاظ نمی‌شود ولی در رگرسیون منطقی انتقال کل این عبارت در مدل به عنوان ترکیب منطقی شناسایی و وارد مدل می‌شود. تفاوت بین AIC مدل‌ها نیز ناشی از این لحاظ شدن این تقابل تاثیرگذار است. با افزایش حجم نمونه، عملکرد رگرسیون منطقی در تشخیص صحیح اثر متقابل واقعی یا به عبارت بهتر، توان آن افزایش می‌یابد به طوری که در حجم نمونه ۱۰۰۰، در ضعیف‌ترین شدت اثر تقابل نیز بیش از ۹۵ درصد مواقع اثر متقابل به درستی تشخیص داده شده است و در حجم نمونه ۲۰۰ و شدت اثر برابر ۳، صد درصد مواقع جمله اثر متقابل درست تشخیص داده شده است. همچنین طبق جداول مشاهده می‌شود که شدت ارتباط اثر متقابل با پاسخ نیز در عملکرد مدل در تشخیص صحیح اثرات متقابل تاثیرگذار است به طوری که هرچه شدت ارتباط اثر متقابل با پاسخ بیشتر باشد، به عبارت دیگر، اثر متقابل قوی‌تر و تاثیرگذارتر باشد، درصد تشخیص صحیح این اثرات متقابل افزایش می‌یابد که کاملا مورد انتظار و معقول است. از طرف دیگر، میزان وابستگی بین مشاهدات متوالی یا همان نسبت شانس بین دو مشاهده متوالی تاثیر چندانی بر عملکرد مدل در تشخیص صحیح اثر متقابل ندارد و در حجم نمونه و ضریب β مشخص، مدل رفتار مشابهی را در یافتن اثرات متقابل صحیح در میزان‌های متفاوت وابستگی از خود نشان می‌دهد.

٤ مثال کاربردی

جامعه مورد بررسی برای مثال کاربردی، داده‌های جمعیت مطالعه قند و لیپید تهران[۱۰] (TLGS) است. مطالعه قند و لیپید تهران یک مطالعه آینده‌نگر است که روی یک نمونه از جمعیت منطقه ۱۳ تهران انجام شده و هدف آن تعیین شیوع بیماری‌های غیر واگیر و ترویج سبک زندگی سالم در این جمعیت است. در این تحقیق داده‌های ۳ مرحله از مطالعه مورد بررسی قرار می‌گیرد: مرحله اول که مطالعه مقطعی بود از اسفند ۱۳۷۷ تا شهریور ۱۳۸۰ به طول انجامید؛ سپس افراد وارد مرحله ۲ مطالعه شدند که از مهر ۱۳۸۰ شروع شد و در شهریور ٨٤ به پایان رسید و پس از آن مرحله سوم اجرا شد که از سال ٨٤ تا اسفند ٨٦ ادامه پیدا کرد اطلاعات کامل مربوط به این مطالعه در منبع مربوطه آمده است [۲]. از میان افراد شرکت کنندگان مطالعه قند و لیپید تهران تعداد ۳۲۹ نفر با سن ۲۰ سال یا بیشتر بودند که در هر ۳ فاز مطالعه حضور داشتند و تمام اطلاعات مورد نیاز برای تحلیل را دارا بودند، که در تحقیق حاضر، داده‌های طولی مربوط به این ۳۲۹ نفر جهت بررسی ارتباط بین پلی‌مورفیسم‌ها و سایر عوامل خطر با سطح پایین HDL خون به عنوان متغیر پاسخ در طول زمان مورد بررسی قرار گرفت.

مقادیر HDL کمتر از ٤٠ میلی‌گرم بر دسی‌لیتر برای مردان و کمتر از ۵۰ میلی‌گرم بر دسی‌لیتر برای زنان به عنوان سطح پایین HDL تعریف شد [١٦]. برخی از عوامل خطر موثر بر داشتن سطح پایین HDL که در این تحقیق تاثیر آن‌ها به همراه پلی‌مورفیسم‌ها بر روی متغیر پاسخ بررسی شده است به صورت: چاقی شکمی

[۹] Relative bias

[۱۰] Tehran Lipid and Glucose Study

(اندازه دور کمر ≤ ۹۵ سانتی‌متر)، تری‌گلیسرید ≤ ۱۵۰ میلی‌گرم بر دسی‌لیتر، فشار خون ≤۱۳۰.۸۵ mm/Hg یا درمان برای پرفشاری خون، گلوکز ناشتای پلاسما ≤ ۱۱۰ میلی‌گرم بر دسی‌لیتر یا مصرف داروی پایین آورنده قند خون مطابق تعریف مولفه‌های سندروم متابولیک بر اساس معیار ATP III [۱] تعریف شوند. همچنین افرادی که در زمان جمع‌آوری داده‌ها به صورت روزانه یا گه‌گاه سیگار مصرف می‌کردند، به عنوان سیگاری در نظر گرفته شدند. متغیر زمان که در این داده‌ها ابتدا به صورت متغیر کیفی سه حالته که فاز زمانی انجام مطالعه را نشان می‌داد، بود به سه متغیر ظاهری[۱۱] تبدیل شد و با در نظر گرفتن مرحله اول مطالعه به عنوان سطح رفرنس، دو متغیر ظاهری ساخته شده برای مراحل ۲ و ۳ مطالعه جهت برازش مدل وارد فضای جستجوی الگوریتم Annealing شدند.

تاثیرات متقابل بین پلی مورفیسم ژن‌های ApoB، ApoCIII، ApoA1M1، ABCA1، SRB1، ApoB، ApoA1M2 و پلی‌مورفیسم ژن ApoAIV که احتمالاً با اختلال در سطح HDL مرتبط هستند [۱۱، ۷، ۳] و سایر عوامل خطر با استفاده از مدل پیشنهادی بررسی شدند. به منظور تجزیه و تحلیل داده‌ها، هر پلی‌مورفیسم به صورت متغیر تصادفی X با مقادیر ۰، ۱ و ۲ در نظر گرفته شد (به عنوان نمونه، نوکلئوتید AA، GA/AG و GG، به ترتیب با ۰ و ۱ و ۲ کدگذاری شدند). سپس این متغیر به دو متغیر دو حالتی با عنوان‌های ژن غالب[۱۲] (X_R) و ژن مغلوب[۱۳] تبدیل شد. متغیر ژن غالب و ژن مغلوب به این صورت تعریف می‌شوند [۱٤]:

$$\begin{cases} X_D = 1, & X \geq 1, \\ X_D = 0, & X = 0, \end{cases}$$

$$\begin{cases} X_R = 1, & X = 2, \\ X_R = 0, & X \leq 1, \end{cases}$$

به این ترتیب، تعداد $2p$ پیش‌بینی کننده دو حالتی از P تا SNP به دست می‌آید. در تحقیق حاضر نیز از شکل دو حالتی غالب و مغلوب این پلی‌مورفیسم‌ها استفاده شده است.

جهت انجام این بررسی، از دو مدل پیشنهادی یعنی رگرسیون منطقی انتقال مارکف مرتبه اول و رگرسیون منطقی انتقال مارکف مرتبه دوم با تابع امتیاز AIC برای تحلیل داده‌های طولی مورد بررسی استفاده شد. در تنظیمات مربوط به الگوریتم Annealing برای یافتن ترکیبات منطقی مناسب، حداکثر تعداد ۳ ترکیب منطقی به عنوان متغیرهای پیش‌بین جدید مدل در نظر گرفته شد که این ترکیبات نهایتاً می‌توانند متشکل از ۸ متغیر دو حالتی اولیه باشند که در فضای جستجوی الگوریتم Annealing تمام پلی‌مورفیسم‌های مورد بررسی، متغیرهای سن، جنسیت، مصرف سیگار و مرحله زمانی انجام مطالعه، تری‌گلیسرید، فشار خون بالا، دور کمر بالا، قند خون ناشتا بالا حضور داشتند و بهترین ترکیبات منطقی توسط الگوریتم شناسایی شده و اثر حاشیه‌ای ترکیبات یافت شده در طول زمان بر پاسخ دو حالتی یعنی داشتن سطح پایین HDL برآورد شد. لذا به طور خلاصه دو مدل رگرسیونی به داده‌ها برازش یافت: اولین مدل، مدل رگرسیون منطقی انتقال با ساختار وابستگی زنجیر مارکف مرتبه اول بود که در آن الگوریتم Annealing از بین تمام متغیرهای پیش‌بین دو حالتی مورد بررسی، تعداد ۳ ترکیب منطقی متشکل از نهایتاً ۸ متغیر دو حالتی اولیه را طوری می‌یابد که در پایان مدل رگرسیون منطقی انتقال برازش یافته با این ترکیبات منطقی به عنوان متغیرهای پیش‌بین جدید مدل، کمترین AIC را نسبت به مدل‌هایی با ترکیبات منطقی دیگر داشته باشد. به طور مشابه، مدل دوم نیز مدل رگرسیون منطقی انتقال با ساختار وابستگی زنجیر مارکف مرتبه دوم بود که در آن الگوریتم Annealing از بین تمام متغیرهای پیش‌بین دو حالتی مورد بررسی، تعداد ۳ ترکیب منطقی متشکل از ۸ متغیر را طوری می‌یابد مدل برازش یافته کمترین AIC را داشته باشد. ضرایب مربوط به هر کدام از این ترکیبات منطقی، میزان همبستگی مرتبه اول و دوم و میزان AIC هر مدل نیز برآورد شدند. در نهایت با استفاده از مدل پیشنهادی توانستیم مدل رگرسیون منطقی انتقال را به داده‌های طولی طوری برازش دهیم که اولاً اثرات متقابل بین متغیرهای پیش‌بین در قالب ترکیبات منطقی شناسایی شده و با برازش مدل با استفاده از این ترکیبات منطقی به عنوان متغیرهای پیش‌بین جدید، این اثرات در مدل لحاظ شود و

[۱۱] dummy

[۱۲] dominant

[۱۳] recessive

ثانیاً وابستگی بین مشاهدات متغیرها پاسخ نیز در تحلیل‌ها در نظر گرفته شده و دقت و کارایی برآوردگرها با در نظر گرفته شدن این وابستگی در مدل‌بندی افزایش یابد.

۱.٤ یافته‌های حاصل از تحلیل داده‌های مثال کاربردی با مدل رگرسیون منطقی انتقال

نتایج مدل رگرسیون منطقی انتقال با ساختار همبستگی زنجیر مارکف مرتبه اول با حضور تمام متغیرهای مورد بررسی موثر بر داشتن سطح پایین HDL در فضای جستجوی الگوریتم Annealing نشان می‌دهد که ترکیب منطقی "داشتن دور کمر بالا و تریگلیسرید بالا" با نسبت شانس ۲/۲۹ دارای اثر متقابل تاثیرگذار بر پاسخ است، همچنین تقابلی بین زمان انجام مطالعه، پلیمورفیسم ApoA1M1 و Apo CIII دیده می‌شود که این تقابل تاثیر افزایشی بر داشتن سطح پایین HDL دارد. آخرین ترکیب منطقی یافت شده در این مدل، ترکیب منطقی شامل تقابلی از متغیرهای جنسیت، فشار خون بالا و SRB1 است. مردان با فشار خون بالا یا افرادی با SRB1=AA دارای شانس کمتری برای داشتن HDL پایین هستند یعنی این ترکیب تاثیر کاهشی بر پاسخ دارد. نسبت شانس برای این ترکیب برابر ۰/۳۸ است. شاخص میزان وابستگی مشاهدات متوالی برای این مدل حدود ۲/۵۰ برآورد شد که نشانگر وابستگی شدید مشاهدات متوالی متغیر پاسخ به هم است. معیار اطلاع آکائیک این مدل (AIC) برابر ۱۰۰۰/۷۲ بدست آمد. در این مدل متغیر سن هم به که به عنوان متغیر تعدیلی کمی در مدل حضور داشت، تاثیر معنی داری بر داشتن سطح پایین HDL نداشت. نتایج مربوط به این مدل به تفصیل در جدول ۵ گزارش شده است.

با توجه به ماهیت داده‌ها که از نوع مشاهدات طولی هستند و می‌توانند وابستگی مارکفی مرتبه دوم نیز داشته باشند، در مرحله بعدی مدل منطقی انتقال مارکف مرتبه دوم نیز به داده‌ها برازش یافت تا همبستگی مرتبه دوم بین داده‌ها نیز بررسی و در صورت وجود در مدل لحاظ شود. مدل رگرسیون منطقی انتقال مارکف مرتبه دوم با ۳ ترکیب منطقی و ۸ متغیر برای بررسی تاثیر پلیمورفیسم‌ها بر داشتن سطح پایین HDL در طول زمان با تعدیل برای تمام متغیرهای مداخله‌گر برازش یافت نتایج این مدل که در جدول ٦ گزارش شده است. طبق این مدل، تقابل بین فشار خون، جنسیت و SRB1 مغلوب اثر کاهشی بر شانس داشتن HDL پایین دارد به این صورت که "مردان با فشار خون بالا یا افراد با SRB1 مغلوب" نسبت شانسی برابر ۰/۳۷ برای داشتن هیپولیپیدمی نسبت به سایر افراد فاقد این ترکیب داشته‌اند ($p < ۰/۰۰۱$). همچنین اثر متقابل بین زمان و پلیمورفیسم ApoCIII وجود دارد و فاز ۲ مطالعه یا ApoCIII مغلوب شانس داشتن HDL پایین را حدوداً ۲ برابر می‌کند ($p < ۰/۰۰۱$). آخرین ترکیب منطقی برای این مدل نشان میدهد که داشتن تریگلیسرید بالا به همراه یک یا هر دو عامل خطر دور کمر و فشار خون بالا شانس داشتن HDL پایین را ۲/۵ برابر افزایش می‌دهد ($p < ۰/۰۰۱$). لگاریتم میزان وابستگی مرتبه اول $\log(\psi_1)$ در این مدل برابر ۲/۳۲ با فاصله اطمینان ٪۹۵ برابر (۲/۸٦ و ۱/۷۸) و میزان وابستگی مرتبه دوم $\log(\psi_2)$ نیز برابر ۱/٦۵ با فاصله اطمینان ٪۹۵ برابر (۲/۲۵ و ۱/۰٦) برآورد شد که هر دو معنی‌دار هستند و نشانگر وجود وابستگی مرتبه اول و دوم بین مشاهدات پاسخ هستند ($p < ۰/۰۰۱$). مقدار AIC این مدل نسبت به مدل رگرسیون منطقی انتقال زنجیر مارکف مرتبه اول کمتر و برابر ۹۷٤/۹۷ به دست آمد.

۵ بحث و بررسی

هدف این پژوهش، بسط رگرسیون منطقی برای تحلیل مشاهدات وابسته به هم و در حالت خاص آن، تحلیل داده‌های طولی بود. ضرورت لحاظ کردن وابستگی بین مشاهدات در تحلیل داده‌های طولی و از طرف دیگر ضرورت بررسی، شناسایی و لحاظ کردن اثرات متقابل در مدل‌بندی داده‌های طولی و محدود بودن رگرسیون منطقی به مطالعات مقطعی، انگیزه پیشنهاد مدل طولی منطقی را ایجاد کرد. با توجه به مبانی نظری رگرسیون منطقی و نیز با انتخاب مدلهای طولی انتقال با ساختار وابستگی زنجیر مارکف مرتبه اول و زنجیر مارکف مرتبه دوم و با در نظر گرفتن AIC این مدل‌ها به عنوان تابع امتیاز لازم برای جستجوی الگوریتم Annealing، مدل رگرسیون منطقی انتقال معرفی شد و مطالعه شبیه‌سازی در سناریوهای مختلف اجرا وداده‌های تولید شده با مدل پیشنهادی و مدل انتقال با اثرات اصلی تحلیل شدند که در همه سناریوها و حالت‌ها، نتایج نشانگر عملکرد خوب مدل رگرسیون منطقی انتقال و برتری آن نسبت به مدل انتقال صرفا با اثرات اصلی بود. در تمام

حجم نمونه‌ها و میزان‌های مختلف وابستگی بین مشاهدات متوالی، شدت اثر برابر صفر ($\beta = 0$) نیز جهت برآورد میزان خطای نوع اول (α) در نظر گرفته شد که نتایج نشان داد در بدترین حالت، این خطا برابر ۰/۰۰۸ بوده است به این معنی که حداکثر ٤ بار در ۵۰۰ سری داده شبیه‌سازی شده با شدت اثر متقابل مد نظر برابر صفر، مدل به اشتباه این اثر متقابل را شناسایی کرده است. همچنین شاخص "میزان تشخیص صحیح اثرات متقابل توسط مدل" به عنوان توان مدل در یافتن این اثرات تعبیر و تفاضل توان از عدد یک به عنوان خطای نوع دوم (β) در نظر گرفته شد. به عنوان مثال زمانی که مدل قادر به یافتن صحیح اثر متقابل در ۷۷ درصد مواقع است توان مدل ۰/۷۷ و خطای نوع دوم مدل برابر ۰/۲۳ است. توان مدل پیشنهادی در یافتن یک اثر متقابل زمانی‌که واقعا چنین تقابلی وجود دارد، به حجم نمونه و شدت اثر تقابل بستگی داشت به‌طوری‌که با اثر متقابل $L = X_1 \vee X_2$، حجم نمونه ۵۰ و شدت اثر ۰/۵ حداکثر توان مدل ۰/۰٦ است در حالی که در همین اثر متقابل و همین حجم نمونه ولی با شدت اثر تقابل برابر ۳ توان مدل حداقل ۰/۹٦ است. از طرفی با تقابل $L = X_1 \vee X_2$ و حجم نمونه ۱۰۰۰، حتی در ضعیف‌ترین شدت اثر نیز، حداقل توان مدل برابر ۰/۹۷ می‌باشد.

به عنوان مثال کاربردی نیز مدل پیشنهاد شده بر روی داده‌های TLGS استفاده شد و اثرات متقابل بین برخی از پلیمورفیسم‌ها و متغیرهای تعدیلگر بر روی سطح پایین HDL شناسایی شد که در مدل رگرسیون منطقی انتقال برازش یافته که تمام متغیرهای موثر بر HDL در فضای جستجوی الگوریتم حضور داشتند بود، ترکیب منطقی "داشتن تریگلیسرید بالا و دور کمر بالا" با سطح پایین HDL ارتباط معنی‌داری به صورت اثرات متقابل نشان داد. در تحقیقات و مطالعات متعدد پزشکی، ارتباط معکوس بین این دو متغیر با سطح HDL مشاهده شده است به طوری که افراد با سطح پایین HDL دارای دور کمر و TG بالاتری هستند (۱۲، ۱۹، ۲۰). همچنین پلیمورفیسمهای Apo A1M1، Apo CIII و SRB1 نیز در این مدل ظاهر شده‌اند که در تقابل با یکدیگر و با سایر متغیرهای تعدیل شده هستند. طبق مطالعات قبلی، تاثیر گذاری این پلی‌مورفیسم‌ها به صورت اثر اصلی بر سطح HDL بررسی شده (۱۱، ۱۵) ولی در این پژوهش، اثرات اصلی و متقابل آنها در طول زمان برای اولین بار با استفاده از مدل رگرسیون منطقی انتقال پیشنهاد شده بررسی شد. همچنین طبق آخرین ترکیب منطقی یافت شده در این مدل، دیده می‌شود که مردان با فشار خون بالا یا افراد با ژنوتیپ AA در SRB1 دارای نسبت شانس کمتری برای داشتن سطح پایین HDL هستند. تقابل بین جنسیت و فشار خون یافته‌ای است که توسط داده‌ها نیز تایید شد.

مقایسه مدل‌های مارکف مرتبه اول و دوم با هم نشان می‌دهد که متغیرهای ظاهر شده در مدلهای متناظر باهم تقریبا یکسان هستند از این مقایسه نظیر به نظیر می‌توان نتیجه گرفت که کاهش AIC مدل‌های مارکف مرتبه دوم نسبت به اول به دلیل لحاظ کردن وابستگی مرتبه دوم بین داده‌هاست و لذا در نظر گرفتن این وابستگی باعث بهبود مدل شده و لزوم برازش مدل مارکف مرتبه دوم با توجه به ماهیت داده‌ها را تایید می‌کند.

٦ نتیجه‌گیری نهایی

هدف رگرسیون منطقی به عنوان یکی از روش‌های آنالیز اثرات متقابل، یافتن ترکیب‌های بولی از متغیرهای دوحالتی اولیه است به طوری که این ترکیب‌های بتوانند پیامد مدنظر را به نحو بهتری پیش‌بینی کنند. مزیت رگرسیون منطقی نسبت به سایر روش‌های تجزیه و تحلیل متغیرهای دوحالتی مانند روش شبکه‌های عصبی مصنوعی و درخت تصمیم‌گیری، این است که یافته‌های رگرسیون منطقی به طور کامل به شکل یک مدل رگرسیون نوشته می‌شود و در نتیجه قادر به تعیین شدت اثر این تقابل‌ها بوده و امکان تفسیر ضرایب، انجام آزمون فرضیه در مورد ضرایب و همچنین ارزیابی کفایت مدل با استفاده از تابع امتیاز آن مدل وجود دارد. این روش رگرسیونی در بررسی‌های متعددی برای تحلیل داده‌های حاصل از مطالعات هم‌گروهی، مورد–شاهدی و مورد–شاهدی جور شده استفاده شده است و به دلیل اهمیت بررسی اثرات متقابل در داده‌های طولی و اهمیت بررسی برهم‌کنش بین متغیرها با زمان، در مطالعه حاضر نیز رگرسیون منطقی برای تحلیل داده‌های طولی معرفی و استفاده شد.

جدول ۱: مقدار آماره AIC برای مدل‌های رگرسیون منطقی انتقال و انتقال با اثرات اصلی و شاخص‌های ارزیابی برای برآوردهای حاصل از مدل رگرسیون منطقی انتقال در مطالعه شبیه‌سازی با حجم نمونه 50 و تعداد 500 تکرار با اثر متقابل ایجاد شده: $L = X_1 \vee X_2$

مدل	$\beta = 0$					$\beta = 0.5$				
	میزان وابستگی (ψ_1) در زنجیر مارکف مرتبه اول					میزان وابستگی (ψ_1) در زنجیر مارکف مرتبه اول				
	0.2	0.5	1	2	5	0.2	0.5	1	2	5
AIC رگرسیون منطقی انتقال	−*	201.8	−	204.72	−	186.49	194.48	196.83	195.14	181.39
AIC انتقال با اثرات اصلی	204.15	216.46	218.96	216.49	204.66	200.03	210.23	212.91	210.36	199.65
درصد تشخیص صحیح اثر متقابل توسط مدل پیشنهادی	صفر	0.8	صفر	0.6	صفر	5	4.4	3.2	5.6	6
درصد اریبی نسبی برآورد وابستگی	−	18	−	−3.5	−	−30	−16	−9	−2.5	33.8
ESE برآورد میزان وابستگی	−	0.28	−	0.11	−	0.01	0.06	0.29	1.36	24.24
درصد اریبی نسبی برآورد β	−	−	−	−	−	114	128	140	136	130
MSE برآورد β	−	2.54	−	1.35	−	0.32	0.66	0.56	0.6	0.21

* برای مواقعی که هیچ مدلی اثر متقابل مد نظر را شناسایی نکرده است، شاخصی هم قابل محاسبه نبود بنابراین این موارد با (−) نشان داده شده است.

مدل	$\beta = 1.5$					$\beta = 3$				
	میزان وابستگی (ψ_1) در زنجیر مارکف مرتبه اول					میزان وابستگی (ψ_1) در زنجیر مارکف مرتبه اول				
	0.2	0.5	1	2	5	0.2	0.5	1	2	5
AIC رگرسیون منطقی انتقال	153.11	158.35	159.32	157.16	151.31	94.88	96.7	97.15	96.55	94.47
AIC انتقال با اثرات اصلی	167.54	172.58	173.81	172.21	165.8	107.29	109.34	109.75	109.51	107.24
درصد تشخیص صحیح اثر متقابل توسط مدل پیشنهادی	64	60.6	60.2	61.4	67	98.4	97.6	96.6	98	
درصد اریبی نسبی برآورد وابستگی	−50	−20	−8	−2.5	5.2	−85	−60	−42	−23	−9.4
ESE برآورد میزان وابستگی	0.03	0.11	1.13	3.95	23.3	0.72	12.16	13.79	65.3	114.82
درصد اریبی نسبی برآورد β	12	12.67	13.33	18	14.67	4	4	4.33	6.33	7.67
MSE برآورد β	0.05	0.07	0.14	0.22	0.18	0.28	0.3	0.3	0.4	0.39

جدول ۲: مقدار آماره AIC برای مدل‌های رگرسیون منطقی انتقال و انتقال با اثرات اصلی و شاخص‌های ارزیابی برای برآوردهای حاصل از مدل رگرسیون منطقی انتقال در مطالعه شبیه‌سازی با حجم نمونه 200 و تعداد 500 تکرار با اثر متقابل ایجاد شده: $L = X_1 \vee X_2$

مدل	$\beta = 0$					$\beta = 0.5$				
	میزان وابستگی (ψ_1) در زنجیر مارکف مرتبه اول					میزان وابستگی (ψ_1) در زنجیر مارکف مرتبه اول				
	0.2	0.5	1	2	5	0.2	0.5	1	2	5
AIC رگرسیون منطقی انتقال	−*	201.8	−	204.72	−	186.49	194.48	196.83	195.14	181.39
AIC انتقال با اثرات اصلی	204.15	216.46	218.96	216.49	204.66	200.03	210.23	212.91	210.36	199.65
درصد تشخیص صحیح اثر متقابل توسط مدل پیشنهادی	صفر	0.8	صفر	0.6	صفر	5	4.4	3.2	5.6	6
درصد اریبی نسبی برآورد وابستگی	−	18	−	−3.5	−	−30	−16	−9	−2.5	33.8
ESE برآورد میزان وابستگی	−	0.28	−	0.11	−	0.01	0.06	0.29	1.36	24.24
درصد اریبی نسبی برآورد β	−	−	−	−	−	114	128	140	136	130
MSE برآورد β	−	2.54	−	1.35	−	0.32	0.66	0.56	0.6	0.21

* برای مواقعی که هیچ مدلی اثر متقابل مد نظر را شناسایی نکرده است، شاخصی هم قابل محاسبه نبود بنابراین این موارد با (−) نشان داده شده است.

مدل	$\beta = 1.5$					$\beta = 3$				
	میزان وابستگی (ψ_1) در زنجیر مارکف مرتبه اول					میزان وابستگی (ψ_1) در زنجیر مارکف مرتبه اول				
	0.2	0.5	1	2	5	0.2	0.5	1	2	5
AIC رگرسیون منطقی انتقال	153.11	158.35	159.32	157.16	151.31	94.88	96.7	97.15	96.55	94.47
AIC انتقال با اثرات اصلی	167.54	172.58	173.81	172.21	165.8	107.29	109.34	109.75	109.51	107.24
درصد تشخیص صحیح اثر متقابل توسط مدل پیشنهادی	64	60.6	60.2	61.4	67	98.4	97.6	96.6	98	
درصد اریبی نسبی برآورد وابستگی	−50	−20	−8	−2.5	5.2	−85	−60	−42	−23	−9.4
ESE برآورد میزان وابستگی	0.03	0.11	1.13	3.95	23.3	0.72	12.16	13.79	65.3	114.82
درصد اریبی نسبی برآورد β	12	12.67	13.33	18	14.67	4	4	4.33	6.33	7.67
MSE برآورد β	0.05	0.07	0.14	0.22	0.18	0.28	0.3	0.3	0.4	0.39

جدول ۳: مقدار آماره AIC برای مدل‌های رگرسیون منطقی انتقال و انتقال با اثرات اصلی و شاخص‌های ارزیابی برای برآوردهای حاصل از مدل رگرسیون منطقی انتقال در مطالعه شبیه‌سازی با حجم نمونه 500 و تعداد 500 تکرار با اثر متقابل ایجاد شده: $L = X_1 \vee X_2$

مدل	$\beta = 0$					$\beta = 0.5$				
	میزان وابستگی (ψ_1) در زنجیر مارکف مرتبه اول					میزان وابستگی (ψ_1) در زنجیر مارکف مرتبه اول				
	0.2	0.5	1	2	5	0.2	0.5	1	2	5
AIC رگرسیون منطقی انتقال	−*	−	−	−	−	1883.88	1985.07	2012.32	1985.25	1874.76
AIC انتقال با اثرات اصلی	1942.69	2062.24	2091.87	2061.96	1942.52	1900.5	2002.91	2029.85	2002.25	1891.22
درصد تشخیص صحیح اثر متقابل توسط مدل پیشنهادی	صفر	صفر	صفر	صفر	صفر	صفر	83.8	78.4	76.8	77
درصد اریبی نسبی برآورد وابستگی	−	−	−	−	−	صفر	−2	−1	−0.5	صفر
ESE برآورد میزان وابستگی	−	−	−	−	−	صفر	صفر	0.02	0.08	0.61
درصد اریبی نسبی برآورد β	−	−	−	−	−	6	8	8	8	4
MSE برآورد β	−	−	−	−	−	0.01	0.01	0.01	0.01	0.01

* برای مواقعی که هیچ مدلی اثر متقابل مد نظر را شناسایی نکرده است، شاخصی هم قابل محاسبه نبود بنابراین این موارد با (−) نشان داده شده است.

مدل	$\beta = 1.5$					$\beta = 3$				
	میزان وابستگی (ψ_1) در زنجیر مارکف مرتبه اول					میزان وابستگی (ψ_1) در زنجیر مارکف مرتبه اول				
	0.2	0.5	1	2	5	0.2	0.5	1	2	5
AIC رگرسیون منطقی انتقال	1534.82	1579.64	1592.63	1576.88	1510.79	932.04	946.67	951.44	947.25	927.56
AIC انتقال با اثرات اصلی	1580.92	1624.46	1637.12	1622.19	1559.65	988.56	1002.72	1007.42	1003.04	984.77
درصد تشخیص صحیح اثر متقابل توسط مدل پیشنهادی	85	100	100	100	100	100	100	100	100	
درصد اریبی نسبی برآورد وابستگی	−5	−2	−1	−0.5	صفر	−5	صفر	1	−0.5	صفر
ESE برآورد میزان وابستگی	صفر	0.01	0.04	0.12	0.86	0.01	0.03	0.14	0.43	3.61
درصد اریبی نسبی برآورد β	صفر	صفر	0.67	0.67	صفر	0.33	0.33	0.33	0.33	0.33
MSE برآورد β	0.01	0.01	0.01	0.02	0.02	0.03	0.02	0.02	0.03	0.03

جدول ٤: مقدار آماره AIC برای مدل‌های رگرسیون منطقی انتقال و انتقال با اثرات اصلی و شاخص‌های ارزیابی برای برآوردهای حاصل از مدل رگرسیون منطقی انتقال در مطالعه شبیه‌سازی با حجم نمونه 1000 و تعداد 500 تکرار با اثر متقابل ایجاد شده: $L = X_1 \vee X_2$

مدل	$\beta = 0$					$\beta = 0.5$				
	میزان وابستگی (ψ_1) در زنجیر مارکف مرتبه اول					میزان وابستگی (ψ_1) در زنجیر مارکف مرتبه اول				
	0.2	0.5	1	2	5	0.2	0.5	1	2	5
AIC رگرسیون منطقی انتقال	−*	4099.36	−	4104.1	−	3767.07	3972.64	4024.39	3969.04	3744.18
AIC انتقال با اثرات اصلی	3871.1	4111.09	4170.36	4111.61	3871.45	3789.43	3993.51	4045.02	3989.87	3767.57
درصد تشخیص صحیح اثر متقابل توسط مدل پیشنهادی	صفر	صفر	0.2	صفر	0.8	صفر	99.2	97.6	97.2	96.6
درصد اریبی نسبی برآورد وابستگی	−	2	−	−3.5	−	صفر	صفر	صفر	−0.5	1
ESE برآورد میزان وابستگی	−	صفر	−	−0.04	−	صفر	صفر	0.01	0.04	0.27
درصد اریبی نسبی برآورد β	−	−	−	−	−	صفر	صفر	2	2	صفر
MSE برآورد β	−	3.52	−	2.65	−	صفر	0.01	0.01	0.01	0.01

* برای مواقعی که هیچ مدلی اثر متقابل مد نظر را شناسایی نکرده است، شاخصی هم قابل محاسبه نبود بنابراین این موارد با (−) نشان داده شده است.

مدل	$\beta = 1.5$					$\beta = 3$				
	میزان وابستگی (ψ_1) در زنجیر مارکف مرتبه اول					میزان وابستگی (ψ_1) در زنجیر مارکف مرتبه اول				
	0.2	0.5	1	2	5	0.2	0.5	1	2	5
AIC رگرسیون منطقی انتقال	3061.13	3150.47	3176.65	3148.06	3014.96	1862.22	1892.44	1900.11	1891.52	1853.8
AIC انتقال با اثرات اصلی	3145.94	3232.18	3257.62	3229.06	3103.42	1966.27	1995.88	2003.73	1994.57	1958.63
درصد تشخیص صحیح اثر متقابل توسط مدل پیشنهادی	98.4	100	100	100	100	100	100	100	100	
درصد اریبی نسبی برآورد وابستگی	صفر	صفر	صفر	−5	صفر	−5	−2	−2	−0.5	−1
ESE برآورد میزان وابستگی	صفر	صفر	0.02	0.06	0.45	صفر	0.02	0.06	0.24	1.6
درصد اریبی نسبی برآورد β	0.67	0.67	0.67	صفر	صفر	صفر	صفر	صفر	صفر	صفر
MSE برآورد β	0.01	0.01	0.01	0.01	0.01	0.03	0.03	0.03	0.02	0.02

جدول ۵: یافته‌های حاصل از برازش مدل رگرسیون منطقی انتقال مارکف مرتبه اول با 3 ترکیب منطقی و 8 متغیر برای بررسی تاثیر پلی‌مورفیسم‌ها بر داشتن سطح پایین HDL در طول زمان

مدل رگرسیون منطقی انتقال مارکف مرتبه اول	ترکیب منطقی یافت شده	نسبت شانس با فاصله اطمینان 95%	معیار اطلاع آکائیکه مدل	$\log(\psi_1)$
3 ترکیب منطقی متشکل از 8 متغیر	ترکیب منطقی اول: (تری گلیسرید بالا **و** دور کمر بالا)	$2.29(1.51, 3.48)$	1000.72	2.5
	$(Apo \neq +/+)$ **یا** $(ApoA1M1 = CC)$	$2.30(1.77, 2.99)$		
	ترکیب منطقی سوم: (مردان با فشار خون بالا) **یا** $(SRBI = AA$	$0.38(0.25, 0.59)$		

جدول ۶: یافته‌های حاصل از برازش مدل رگرسیون منطقی انتقال مارکف مرتبه دوم با 3 ترکیب منطقی و 8 متغیر برای بررسی تاثیر پلی‌مورفیسم‌ها بر داشتن سطح پایین HDL در طول زمان

مدل رگرسیون منطقی انتقال مارکف مرتبه اول	ترکیب منطقی یافت شده	نسبت شانس با فاصله اطمینان 95%	معیار اطلاع آکائیکه مدل	$\log(\psi_1)$	$\log(\psi_2)$
3 ترکیب منطقی متشکل از 8 متغیر	ترکیب منطقی اول: (تری گلیسرید بالا **و** دور کمر بالا) **یا** فشار خون بالا))	$(1.66, 3.76)$ 2.50	974.97	2.32	1.65
	ترکیب منطقی دوم: (مردان با فشار خون بالا) **یا** $(SRB1 = AA$	$(0.23, 0.59)$ 0.37			
	ترکیب منطقی سوم: (مرحله ۲ مطالعه **یا** $(ApoCII = GG$	$(1.57, 2.57)$ 2.01			

مراجع

[1] Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III) final report. Circulation 2002 Dec 17; 106 (25): :3143-421.2.

[2] Azizi F, Rahmani M, et al. Cardiovascular risk factors in an Iranian urban population: Tehran Lipid and Glucose Study (Phase 1). Sozial- und Präventivmedizin/Social and Preventive Medicine. 2002;47(6): 408-26.

[3] Brown CM, Rea TJ, et al. The contribution of individual and pairwise combinations of SNPs in the APOA1 and APOC3 genes to interindividual HDL-C variability. J Mol Med (Berl) 2006 Jul; 84(7): 561-72 Epub 2006 May 17.

[4] Carty CL, Heagerty P, et al. Interaction between Fibrinogen and IL-6 Genetic Variants and Associations with Cardiovascular Disease Risk in the Cardiovascular Health Study. Annals of Human Genetics. 74(1): 1-10.

[5] Diggle PJ, Liang KY, et al.; 1994 Analysis of Longitudinal data. oxford university press. New York.

[6] Etzioni R, Falcon S, et al. Prostate-specific antigen and free prostate-specific antigen in the early detection of prostate cancer: do combination tests improve detection? Cancer Epidemiol Biomarkers Prev 2004 Oct; 13(10): 1640-5.

[7] Frikke-Schmidt R. Context-dependent and invariant associations between APOE genotype and levels of lipoproteins and risk of ischemic heart disease: a review. Scand J Clin Lab Invest Suppl 2000; 233: 3-25.

[8] Goncalves MH, Azzalini A. Using Markov chains for marginal modelling of binary longitudinal data in anexact likelihood approach. Metron - International Journal of Statistics. 2008; 0(2): 157-81.

[9] Ishiguro M, Akaike H. 1989 DALL, Davidon's algorithm for log likelihood maximization: a Fortran subroutine for statistical model buildersInstitute of Statistical Mathematics.

[10] Li Q, Fallin MD, et al. Detection of SNP-SNP interactions in trios of parents with schizophrenic children. Genet Epidemiol 2010 Jul; 34(5):396-406.

[11] McCarthy JJ, Lehner T, et al. Association of genetic variants in the HDL receptor, SR-B1, with abnormal lipids in women with coronary artery disease. Journal of Medical Genetics. 2003 June 1, 2003; 40(6):453-8.

[12] Miller M, Langenberg P, et al. Impact of lowering triglycerides on raising HDL-C in hypertriglyceridemic and non-hypertriglyceridemic subjects. Int J Cardiol 2007 Jul 10; 119(2): 192-5 Epub 2006 Oct 18.

[13] Moineddin R, Matheson F, et al. A simulation study of sample size for multilevel logistic regression models. BMC Medical Research Methodology. 2007; 7:34.

[14] Ruczinski I, Kooperberg C, et al. Logic Regression. Journal of Computational and Graphical Statistics. 2003; 12(3):475-511.

[15] Sarbakhsh P, Mehrabi Y, et al. Logic regression analysis of association of gene polymorphisms with low HDL: Tehran Lipid and Glucose Study. Gene 2013 Jan 25; 513(2): 278-81 doi: 101016/jgene201210084 Epub 2012 Nov 10.

[16] Sarbakhsh P, Mehrabi y, et al. Logic regression and its application in predicting diseases. Andish-ye-amari. [Research]. 16(1):34-46.

[17] Schwender H, Ickstadt K. Identification of SNP interactions using logic regression. Biostatistics. 2008 January 1, 2008; 9(1):187-98.

[18] Schwender H, Ruczinski I. Logic regression and its extensions. Adv Genet 2010; 72: 25-45.

[19] Seidell JC, Perusse L, et al. Waist and hip circumferences have independent and opposite effects on cardiovascular disease risk factors: the Quebec Family Study. Am J Clin Nutr 2001 Sep; 74(3): 315-21.

[20] Williams PT, Krauss RM, et al. Associations of lipoproteins and apolipoproteins with gradient gel electrophoresis estimates of high density lipoprotein subfractions in men and women. Arterioscler Thromb 1992 Mar; 12(3): 332-40.