



مقایسه مدل‌های آمیخته خطی تعمیم‌یافته و مدل‌های خطی تعمیم‌یافته در تعیین عوامل

مرتبط با بیماری دیابت نوع ۲ در استان یزد

نویسندگان: حسین فلاح‌زاده^۱، فریبا اسدی^۲، مسعود رحمانیان^۳، مهدی عمادی^۴

۱. استاد آمار زیستی، گروه آمار و اپیدمیولوژی، دانشگاه علوم پزشکی و خدمات بهداشتی درمانی شهید صدوقی یزد
۲. نویسنده مسئول: دانشجوی کارشناسی ارشد آمار زیستی، دانشگاه علوم پزشکی و خدمات بهداشتی درمانی شهید صدوقی یزد
- تلفن تماس: ۰۹۳۵۵۴۳۵۶۰۳ Email: fariba.asadi3@gmail.com
۳. فوق تخصص بیماریهای غدد داخلی و متابولیسم، مرکز تحقیقات دیابت یزد
۴. دانشیار گروه آمار، دانشگاه فردوسی مشهد

چکیده

مقدمه: بیماری دیابت از جمله بیماری‌های مزمنی است که شیوع آن بسیار زیاد و روز به روز در حال افزایش است. در این مطالعه ضمن تعیین عوامل موثر بر بیماری دیابت به مقایسه دو مدل خطی تعمیم‌یافته و خطی آمیخته تعمیم‌یافته می‌پردازیم.

روش بررسی: داده‌های این مطالعه مربوط به طرح تحقیقاتی بررسی شاخص‌های اپیدمیولوژیک بیماری دیابت بزرگسالان در گروه سنی ۳۰ سال و بالاتر شهری استان یزد می‌باشد. در این مطالعه جمعا ۲۷۹۵ نفر با انجام آزمایش قند خون از لحاظ ابتلا به دیابت بررسی شدند. برای تحلیل داده‌ها با استفاده از مدل رگرسیون لجستیک آمیخته و رگرسیون لجستیک معمولی از نرم افزار R استفاده شد.

یافته‌ها: در این مطالعه متغیرهای سابقه خانوادگی ابتلا به دیابت، سن، شاخص توده بدنی و دور کمر به دور باسن در هر دو مدل معنی‌دار شد ($p < 0.001$)، با این تفاوت متغیر که شغل در مدل رگرسیون لجستیک معمولی در سطح ۰/۱ معنی‌دار بود ولی در مدل رگرسیون لجستیک آمیخته معنی‌دار نبود. همچنین، متغیرهای مساحت خانه، سطح تحصیلات و جنسیت در هیچ یک معنی‌دار نشد. با توجه به مقادیر نسبت شانس نیز در برخی از آنها شاهد تفاوت قابل توجهی بین دو مدل هستیم. با توجه به خطای استاندارد ضرایب و مقایسه مقادیر آن در دو مدل، شاهد کم برآوردی در مدل رگرسیون لجستیک معمولی بودیم.

نتیجه‌گیری: بکارگیری مدل‌های آمیخته تعمیم‌یافته منجر به نتایج دقیق‌تری می‌شود و از کم برآوردی خطای استاندارد ضرایب جلوگیری می‌کند.

واژه‌های کلیدی: مدل‌های خطی آمیخته تعمیم‌یافته، رگرسیون لجستیک، دیابت، تقریب لاپلاس، مدل‌های

خطی تعمیم‌یافته

طلوع بهداشت

دو ماهنامه علمی پژوهشی

دانشکده بهداشت یزد

سال پانزدهم

شماره: دوم

خرداد و تیر ۱۳۹۵

شماره مسلسل: ۵۶

تاریخ وصول: ۱۳۹۳/۸/۱۷

تاریخ پذیرش: ۱۳۹۳/۱۰/۲۱



مقدمه

(Models) (۱۰-۱۳) کلاسی از مدل‌های رگرسیونی است که شامل مدل‌های رگرسیون خطی می‌شود و علاوه بر آن بسیاری از مدل‌های غیرخطی مهم بکار رفته در تحقیقات پزشکی زیستی را نیز دربر می‌گیرد. این مدل‌ها رگرسیون معمولی را با اضافه کردن پاسخ‌های غیرنرمال و تابع اتصالی از توزیع‌های خانواده نمایی (مثل دوجمله‌ای و پواسن) گسترش داده اند و مدل‌های خطی آمیخته تعمیم یافته با اضافه کردن اثرات تصادفی (خوشه) در پیشگوه‌های خطی باعث گسترش بیشتر آنها شده اند (۹)، به همین دلیل مدل‌های خطی آمیخته تعمیم یافته بهترین ابزار برای آنالیز داده‌های غیرنرمال که شامل اثرات تصادفی هستند به شمار می‌روند (۳). مدل رگرسیون لجستیک حالتی از مدل‌های خطی تعمیم یافته است که بیشترین کاربرد را در مطالعات پزشکی دارد. شکل کلی این مدل به صورت زیر است:

$$\ln \frac{\pi_i}{1 - \pi_i} = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k$$

که β_0 عرض از مبدا جامعه است و به صورت لگاریتم بخت‌های موفقیت وقتی تمامی کوریت‌ها مقدار صفر می‌گیرند تفسیر می‌شود و β_1 که شیب جامعه گفته می‌شود به صورت تغییر لگاریتم بخت‌های (log odds) موفقیت به ازای یک واحد تغییر در x_1 وقتی سایر متغیرها ثابت نگه داشته شده‌اند تفسیر می‌شود (۱۴). همان‌طور که گفته شد مدل‌های آمیخته خطی تعمیم یافته اثرات تصادفی را به پیشگوی خطی اضافه می‌کنند. پس مدل رگرسیون لجستیک آمیخته را می‌توان به صورت زیر بیان کرد:

$$\ln \frac{\pi_i}{1 - \pi_i} = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + b_1 z_1 + \dots + b_m z_m$$

در بسیاری از مطالعات با داده‌هایی روبرو هستیم که داری ساختار سلسله‌مراتبی و یا طولی هستند، بعنوان مثال اندازه‌گیری کلسترول خون افراد در چند خانواده یا در چندین شهر، استان و یا کشور و یا اندازه‌گیری فشار خون افراد در طول زمان. در این‌گونه داده‌ها بدلیل وجود همبستگی درون خوشه‌ها (مثلاً خانواده، شهر، استان و یا کشور در مثال اول و افراد در مثال دوم) فرض استقلال برقرار نمی‌باشد و به همین دلیل مدل‌های خطی (۱،۲) برای تحلیل این داده‌ها مناسب نیستند. گاهی اوقات محققان به اشتباه این همبستگی را نادیده می‌گیرند و یا آن را بعنوان اثرات ثابت در نظر می‌گیرند (۳). عدم در نظر گرفتن این همبستگی میان داده‌ها گرچه ممکن است در ضرایب تاثیر چندانی نداشته باشد ولی بر انحراف معیار و فاصله اطمینان و همچنین آزمون‌ها تاثیر زیادی خواهد داشت (۴). پس باید مدلی را استفاده کرد که این همبستگی‌ها در آن لحاظ شود که این مدل‌ها به مدل‌های خطی آمیخته (۵،۶) معروف هستند مدل‌های خطی با اثرات آمیخته یک روش مهم برای تحلیل داده‌های طولی و سلسله‌مراتبی است که در علوم مختلف از جمله کشاورزی، ژئوفیزیک، زیست‌شناسی و پزشکی استفاده می‌شود. فرض معمول در برازش این مدل‌ها آن است که داده‌ها دارای توزیع نرمال هستند (۹-۳،۷). از طرفی در بسیاری از داده‌های پزشکی متغیر پاسخ توزیع نرمال ندارد و معمولاً به صورت شمارشی و یا دو حالتی است. در این‌گونه موارد بسیاری از محققان سعی می‌کنند با تبدیل داده‌ها آنها را نرمال کنند ولی همیشه اینکار امکان‌پذیر نیست و باید از مدل‌های تعمیم یافته استفاده کرد. مدل‌های خطی تعمیم یافته Generalized Linear



پانکراس منجر به نقص تولید انسولین می‌شود و در نوع دو مقاومت پیشرونده بدن به انسولین وجود دارد که در نهایت ممکن است به تخریب سلول‌های بتای پانکراس و نقص کامل تولید انسولین منجر شود. در دیابت نوع دو مشخص است که عوامل ژنتیکی، چاقی و کم‌تحرکی نقش مهمی در ابتلای فرد دارند (۲۳-۲۰). از آنجایی که داده‌های این مطالعه به صورت خوشه‌ای گردآوری شده است پس با توجه به توضیحات داده شده استفاده از مدل‌های آمیخته خطی ضروری می‌باشد. در این مطالعه ضمن تعیین عوامل موثر بر بیماری دیابت به مقایسه دو مدل خطی تعمیم‌یافته و خطی آمیخته تعمیم‌یافته می‌پردازیم.

روش بررسی

این مطالعه از نوع توصیفی-تحلیلی و به صورت مقطعی بوده و داده‌های آن مربوط به طرح تحقیقاتی بررسی شاخص‌های اپیدمیولوژیک بیماری دیابت بزرگسالان در گروه سنی ۳۰ سال و بالاتر شهری استان یزد است که در سال ۱۳۷۷ جمع‌آوری شده است. انتخاب نمونه‌ها در این مطالعه به صورت خوشه‌ای بوده که شامل ۹۴ خوشه ۲۰ خانواری بوده و جمعا ۲۷۹۵ نفر با انجام آزمایش قند خون از لحاظ ابتلا به دیابت بررسی شدند. در این مطالعه تقسیم بندی دیابت براساس معیارهای سازمان جهانی بهداشت (WHO) صورت گرفته است. پس از ورود داده‌ها به نرم افزار R3.0.1 متغیرهای سن، سابقه خانوادگی ابتلا به دیابت، شاخص توده بدنی، تحصیلات، دور کمر، شغل، دور باسن، مساحت خانه و جنسیت به عنوان متغیرهای مستقل و متغیرهای خانوار و محل سکونت به عنوان اثر تصادفی در نظر گرفته شدند. از آنجایی که متغیر پاسخ یعنی ابتلا به دیابت دو حالتی بود (دارد، ندارد) از تابع اتصال لجیت استفاده شد. سپس با

که در آن b_i ها ضرایب اثرات تصادفی هستند و فرض بر این است که این ضرایب دارای توزیع نرمال با میانگین صفر و ماتریس واریانس کواریانس می‌باشند. علت اینکه خوشه را به عنوان اثر ثابت در نظر نمی‌گیرند اینست که مطالعات معمولا تعداد زیادی خوشه دارند که اگر به عنوان اثر ثابت با آنها رفتار شود مدل شامل تعداد زیادی پارامتر می‌شود و موجب پیچیدگی مدل و نتایج نادرست می‌گردد ولی اگر خوشه را به عنوان اثر تصادفی در نظر بگیریم آنگاه تنها یک پارامتر شرطی در مدل پراکندگی را توصیف خواهد کرد (۹). مدل‌های آمیخته خطی تعمیم‌یافته (Generalized Linear Mixed Models) (۸ و ۹ و ۱۴ و ۱۵) ترکیبی از مدل‌های خطی تعمیم‌یافته و مدل‌های آمیخته خطی می‌باشند که برای تحلیل داده‌های همبسته با پاسخ غیر نرمال بکار می‌روند (۱۱). گرچه از زمان معرفی این مدل‌ها سالها می‌گذرد ولی در سال‌های اخیر استفاده از این مدل‌های و تحقیق در مورد روش‌های برآورد آنها بدلیل پیشرفت نرم افزارهای مختلف بسیار مورد توجه محققان قرار گرفته است و موضوع اصلی بسیاری از تحقیقات است (۱۸-۱۶، ۱۱).

دیابت از جمله بیماری‌های مزمن در چند دهه اخیر است که شیوع آن در تمامی کشورها از جمله ایران روز به روز افزایش پیدا می‌کند. این بیماری که به آن بیماری قند نیز گفته می‌شود (۱۹)، یک اختلال متابولیک (سوخت و سازی) در بدن است. در این بیماری توانایی تولید انسولین در بدن از بین می‌رود و یا بدن در برابر انسولین مقاوم شده و بنابراین انسولین تولیدی نمی‌تواند عملکرد طبیعی خود را انجام دهد. نقش اصلی انسولین پایین آوردن قند خون توسط مکانیزم‌های مختلفی است. دیابت دو نوع اصلی دارد. در دیابت نوع یک تخریب سلول‌های بتا در



مدل حذف شدند. معیار مقایسه مدلها معیار آکائیک و روش برازش مدل لجستیک آمیخته روش لاپلاس بود.

یافته‌ها

در جدول شماره یک ضرایب و مقادیر معنی داری هر یک از متغیرهای موجود در برازش مدل نهایی و در جدول دو مقادیری نسبت شانس و فاصله اطمینان در هر دو مدل آورده شده است. براساس نتایج حاصل از برازش مدل رگرسیون لجستیک معمولی متغیرهای شغل، تحصیلات و مساحت خانه معنی دار نشده است. همچنین شانس ابتلا به دیابت در افرادی که سابقه فامیلی ابتلا به دیابت در آنها وجود نداشته ۵۲ درصد کمتر از آنهاست که در بستگان آنها سابقه ابتلا به دیابت وجود داشته است و با بالا رفتن سن نیز شانس ابتلا به دیابت نیز به شدت بالا رفته است. از طرفی با توجه به این دو جدول خطر مذکور در افراد لاغر و نرمال تفاوت چندانی نداشته در صورتی که در افراد چاق و دارای اضافه وزن شانس ابتلا به دیابت به ترتیب ۳/۷۲ و ۵/۷۷ برابر افراد لاغر است.

استفاده از دستور glmer مدل رگرسیون لجستیک آمیخته و توسط دستور glm مدل رگرسیون لجستیک به داده ها برازش داده شد. دستور glmer اثرات ثابت و اثرات تصادفی را به صورت مجزا به ما می‌دهد ولی نتایج اثرات ثابت را برای کل جامعه‌ای که نمونه از آن گرفته شده است می‌توان تعمیم داد، در صورتی که اثرات تصادفی تنها مربوط به نمونه است.

در هر دو مدل پس از رفع هم خطی به بررسی معنی داری هر یک از عوامل به روش پس رو پرداخته شد. به این صورت که بعد از برازش مدل اولین عاملی که بیشترین مقدار معنی داری را داشت (متغیر جنسیت) از مدل خارج کرده و برازش مدل را روی سایر متغیرهای باقی مانده انجام دادیم. پس از تعیین مدل نهایی فرضیات مدل بررسی و نتایج دو مدل نهایی لجستیک آمیخته و کلاسیک با یکدیگر مقایسه شد.

لازم به ذکر است که در مدل نهایی متغیر جنسیت و اثر متقابل تحصیلات و شغل به دلیل عدم معنی داری و بهتر شدن برازش از

جدول ۱: مقادیر ضرایب متغیرها و معنی داری آنها براساس مدل رگرسیون لجستیک آمیخته و معمولی

مدل رگرسیون لجستیک معمولی		مدل رگرسیون لجستیک آمیخته		متغیر
مقدار معنی داری	ضریب در مدل (SE)	مقدار معنی داری	ضریب در مدل (SE)	
<۰/۰۰۱	(۰/۱۰)-۰/۷۳	<۰/۰۰۱	(۰/۱۰)-۰/۷۰	سابقه خانوادگی
<۰/۰۰۱	(۰/۱۳)۰/۵۴	<۰/۰۰۱	(۰/۱۴)۰/۵۵	گروه سنی ^۱
<۰/۰۰۱	(۰/۱۴)۱/۳۱	<۰/۰۰۱	(۰/۱۵)۱/۴۱	۴۰-۴۹
<۰/۰۰۱	(۰/۱۶)۱/۷۵	<۰/۰۰۱	(۰/۱۸)۱/۸۷	۵۰-۶۴
۰/۱۰	(۰/۱۹)۰/۳۲	۰/۱۲	(۰/۲۱)۰/۳۲	+۶۵
<۰/۰۰۱	(۰/۱۹)۰/۷۸	<۰/۰۰۱	(۰/۲۱)۰/۸۳	شاخص توده بدنی ^۲
<۰/۰۰۱	(۰/۲۱)۰/۸۴	<۰/۰۰۱	(۰/۲۳)۰/۹۲	نرمال
۰/۱۲	(۰/۱۴)-۰/۱۹	۰/۳۱	(۰/۱۵)-۰/۱۵	چاق
<۰/۰۰۱	(۰/۵۱)۱/۶۷	<۰/۰۰۱	(۰/۵۳)۱/۸۸	اضافه وزن
۰/۳۰	(۰/۰۰)۰/۰۰	۰/۱۸	(۰/۰۰)۰/۰۰	تحصیلات ^۳
۰/۶۳	(۰/۲۲)-۰/۱۰	۰/۳۳	(۰/۲۴)-۰/۲۳	دور کمر به دور باسن
۰/۵۷	(۰/۱۸)-۰/۱۰	۰/۶۱	(۰/۱۹)-۰/۰۹	مساحت خانه
۰/۰۹	(۰/۳۳)-۰/۵۵	۰/۱۸	(۰/۳۵)-۰/۴۷	کارگر
۰/۳۰	(۰/۲۳)-۰/۲۴	۰/۲۴	(۰/۲۵)-۰/۲۸	خانه دار
۰/۶۵	(۰/۱۹)-۰/۰۸	۰/۷۳	(۰/۲۰)-۰/۰۶	کشاورز
-	-	۰/۴۲	۰/۱۲	بازنشسته
-	-	<۰/۰۰۱	۰/۳۸	سایر
-	-	۰/۴۲	۰/۱۲	واریانس اثر تصادفی اول (خانواده)
-	-	<۰/۰۰۱	۰/۳۸	واریانس اثر تصادفی دوم (محل سکونت)

گروه‌های مرجع به ترتیب شامل: ۱. گروه سنی ۳۹-۳۰ * ۲. لاغر * ۳. زیر دیپلم * ۴. مرد * ۵. کارمند می‌باشند.



جدول ۲: مقادیر نسبت شانس و فواصل اطمینان متغیرها براساس برازش مدل‌های رگرسیون لجستیک آمیخته و معمولی

فاصله اطمینان (۰.۹۵٪)				نسبت شانس (OR)		متغیر
لجستیک آمیخته		لجستیک معمولی		لجستیک آمیخته	لجستیک معمولی	
حد پایین	حد بالا	حد پایین	حد بالا			
۰/۵۸	۰/۳۹	۰/۶۱	۰/۳۹	۰/۴۸	۰/۴۹	سابقه خانوادگی (ندارد)
۲/۲۴	۱/۳۲	۲/۳۰	۱/۳۱	۱/۷۲	۱/۷۳	گروه سنی ^۱
۴/۹۳	۲/۸	۵/۶۰	۳/۰۳	۳/۷۲	۴/۱۲	۵۰-۶۴
۸/۰۴	۴/۱۵	۹/۳۷	۴/۵۴	۵/۷۷	۶/۵۳	۶۵+
۲/۰۳	۰/۹۳	۲/۰۹	۰/۹۱	۱/۳۷	۱/۳۸	شاخص توده بدنی ^۲
۳/۲۳	۱/۴۸	۳/۴۹	۱/۵۲	۲/۱۹	۲/۳۱	نرمال
۳/۵۴	۱/۵۱	۳/۹۶	۱/۶۰	۲/۳۲	۲/۵۲	چاق
۱/۰۸	۰/۶۲	۱/۱۵	۰/۶۳	۰/۸۲	۰/۸۵	اضافه وزن
۱۴/۵	۱/۹۵	۱۸/۹۶	۲/۲۸	۵/۳۲	۶/۵۸	تحصیلات ^۳
۱/۰۰	۰/۹۹	۱/۰۰	۰/۹۹	۱/۰۰	۱/۰۰	دور کمر به دور باسن
۱/۴۰	۰/۵۷	۱/۲۷	۰/۴۸	۰/۸۹	۰/۷۹	مساحت خانه
۱/۲۹	۰/۶۳	۱/۳۲	۰/۶۲	۰/۹۰	۰/۹۰	کارگر ^۴
۱/۱۰	۰/۲۹	۱/۲۵	۰/۳۰	۰/۵۷	۰/۶۲	خانه دار
۱/۲۴	۰/۴۹	۱/۲۲	۰/۴۵	۰/۷۸	۰/۷۴	کشاورز
۱/۳۴	۰/۶۲	۱/۴۰	۰/۶۱	۰/۹۱	۰/۹۳	بازنشسته
						سایر

گروه‌های مرجع به ترتیب شامل: ۱. گروه سنی ۳۹-۳۰ * ۲. لاغر * ۳. زیر دیپلم * ۴. کارمند می باشند.

دیابت تشخیص داده نشد، ولی همه آنها شانس کمتری در ابتلا به دیابت نسبت به کارمندان داشته و کمترین آن مربوط به کشاورزان بود.

همچنین یافته‌ها حاکی از آن است که با یک واحد افزایش در شاخص دور کمر به دور باسن نسبت شانس ابتلا به دیابت ۶/۵۸ برابر افزایش می‌یابد. متغیر سطح تحصیلات و مساحت خانه در این مدل معنی‌دار نشد.

مقدار واریانس در سطح خانوار برابر ۰/۱۲ برآورد شد که نشان می‌دهد همبستگی میان افراد داخل هر خانواده وجود دارد و قابل توجه است. همچنین برآورد مقدار واریانس اثر تصادفی دوم یعنی ناحیه برابر ۰/۳۸ با خطای معیار ۰/۶۲ بدست آمد ($P < 0.001$) که نشان‌دهنده همبستگی معنی‌داری بین افراد ساکن در هر ناحیه از استان یزد با ابتلا به دیابت است.

همچنین این خطر در افراد کشاورز ۴۳ درصد کمتر از کارمندان بود. مقدار نسبت شانس شاخص دور کمر به دور باسن در این مدل ۵/۳۲ با فاصله اطمینان (۱/۹۵، ۱۴/۵) بدست آمد که نشان می‌دهد با یک واحد افزایش این شاخص شانس ابتلا به دیابت ۵/۳۲ برابر افزایش می‌یابد.

نتایج رگرسیون لجستیک آمیخته نشان می‌دهد که شانس ابتلا به دیابت در افرادی با تحصیلات دیپلم و بالاتر ۱۵ درصد کمتر از افراد زیر دیپلم و نسبت شانس در افراد بدون سابقه خانوادگی ابتلا به دیابت ۰/۴۹ برابر افرادی است که در خانواده آنها سابقه ابتلا به دیابت وجود دارد. با بالا رفتن سن نیز نسبت شانس ابتلا به دیابت بالا رفته طوری که افراد بالای ۶۵ سال ۶/۵۳ برابر نسبت به افراد در گروه سنی ۳۹-۳۰ سال بیشتر در معرض خطر ابتلا به دیابت هستند. گرچه شغل از عوامل موثر بر



بحث و نتیجه‌گیری

به طور کلی عوامل خطر متعددی نظیر سن، فشار خون، فشارهای روحی و روانی، شغل، تغذیه، سابقه خانوادگی ابتلا به دیابت و... در ابتلا به دیابت نقش دارند که بسیاری از آنان را در این مطالعه توسط دوروش رگرسیونی مورد بررسی قرار دادیم. از آنجایی که مدل‌های خطی برای داده‌های پیوسته استفاده می‌شوند ولی معمولاً در عمل با مشاهدات گسسته زیادی روبرو هستیم. مک و نلدر (۱۹۸۹) (۱۰) گسترش یافته مدل‌های خطی با نام مدل‌های خطی تعمیم یافته را پیشنهاد کردند که مدل رگرسیون لجستیک یکی از انواع این مدل‌هاست. آنها اشاره کردند که عناصر کلیدی یک مدل خطی کلاسیک یعنی مدل رگرسیون خطی عبارتند از: مشاهدات مستقلند (۸)، میانگین مشاهدات تابعی خطی از کوریت هاست (۱۴)، و واریانس مشاهدات ثابت است. از طرفی در بسیاری از داده‌ها بخصوص داده‌های که به صورت طولی و یا خوشه‌ای جمع‌آوری شده‌اند شاهد نوعی همبستگی در داده‌ها هستیم. گسترش یافته مدل‌های خطی تعمیم یافته یعنی مدل‌های خطی آمیخته تعمیم یافته با اصلاح موارد دو سه موجب در نظر گرفتن این همبستگی‌ها می‌شود به این صورت که میانگین مشاهدات وابسته به تابعی خطی از برخی کوریت‌ها از طریق یک تابع اتصال است (۹)، و واریانس مشاهدات تابعی از میانگین است. داده‌های این مطالعه به صورت نمونه‌گیری خوشه‌ای جمع‌آوری شده بودند پس نوعی همبستگی درون آنها وجود دارد. در این مطالعه ما به مقایسه دو مدل رگرسیونی لجستیک آمیخته و معمولی پرداختیم. در یک مطالعه مقطعی که به منظور بررسی عوامل خطر دیابت در سال ۱۹۹۶ روی افراد ۱۵ ساله و بالاتر در ممری مرکز شهر کیپ تاون در آفریقای جنوبی انجام

شد نتایج آنالیز رگرسیون نشان داد که سن، دور کمر، مصرف انرژی کم و سابقه خانوادگی ابتلا به دیابت معنی‌دار بود در حالی که جنس و چاقی معنی‌دار نبودند (۲۴). در این مطالعه نیز متغیرهای سابقه خانوادگی، سن، شاخص توده بدنی و دور کمر به دور باسن در هر دو مدل معنی‌دار شده است، ولی متغیر جنس معنی‌دار نبود، در حالی که در مطالعه انجام شده در اصفهان صورت گرفت متغیرهای سن، جنس، فشارخون، شاخص توده بدنی و سابقه فامیلی ابتلا به دیابت در ابتلا به دیابت معنی‌دار شد (۲۶). در این مطالعه متغیر شغل در مدل رگرسیون لجستیک معمولی در سطح ۰/۱ معنی‌دار بود ولی در مدل رگرسیون لجستیک آمیخته معنی‌دار نبود. همچنین متغیرهای جنس، مساحت خانه و تحصیلات در هیچ یک معنی‌دار نشد. با توجه به مقادیر نسبت شانس نیز در برخی از آنها شاهد تفاوت قابل توجهی بین دو مدل هستیم. با توجه به خطای استاندارد ضرایب و مقایسه مقادیر آن در دو مدل شاهد کم برآوردی در مدل رگرسیون لجستیک معمولی بودیم. در مطالعه‌ای که اسکرنال و همکارانش در سال ۲۰۰۲ انجام دادند نیز به این نتیجه رسیدند که در نظر نگرفتن همبستگی موجود در داده‌ها باعث کم برآوردی خطای استاندارد ضرایب رگرسیونی می‌شود (۲۵). با توجه به نتایج بدست آمده از این دو مدل می‌توان گفت که مدل رگرسیون لجستیک آمیخته که حالتی از مدل‌های خطی آمیخته تعمیم یافته است به دلیل در نظر گرفتن همبستگی در داده‌های سلسله‌مراتبی نتایج دقیق‌تری نسبت به مدل رگرسیون لجستیک ارائه می‌دهد.

نتایج این مطالعه نشان داد که سابقه خانوادگی ابتلا به دیابت، سن و شاخص توده بدنی و دور کمر به دور باسن از عوامل مرتبط



تقدیر و تشکر

این مقاله حاصل نتایج پایان‌نامه کارشناسی ارشد در دانشگاه علوم پزشکی شهید صدوقی یزد می‌باشد. بدین وسیله از آقایان مهندس محمد حسین احمدیه و دکتر محمد افخمی اردکانی که داده‌ها را در اختیار نویسندگان این مقاله قرار دادند تشکر و قدر دانی می‌شود. همچنین از جناب آقای دکتر مهدی یاسری استادیار گروه اپیدمیولوژی و آمارزیستی دانشگاه تهران به دلیل کمک‌های بی‌دریغشان بی‌نهایت سپاسگزاری می‌شود.

با ابتلا به دیابت است؛ گرچه عوامل سن و سابقه خانوادگی ابتلا به دیابت قابل تغییر نیستند ولی می‌توان با شناسایی افراد در معرض خطر و آموزش آنها از شیوع دیابت جلوگیری نمود. همچنین از آنجایی که بکارگیری مدل‌های آمیخته تعمیم یافته منجر به نتایج دقیق‌تر می‌شود و از کم برآوردی خطای استاندارد ضرایب جلوگیری می‌کند، این مدل‌ها به عنوان مدل برتر برای تحلیل داده‌های همبسته در مطالعات طولی و چند سطحی توصیه می‌شود.

References

- 1- Radhakrishna Rao C, Toutenburg H. Linear models (Least Squares and Alternatives): Springer New York; 1995.
- 2- Searle SR. Linear models. Wiley Classics Library: John Wiley & Sons; 2012.
- 3- Bolker BM, Brooks ME, Clark CJ, Geange SW, Poulsen JR, Stevens MHH, et al. Generalized linear mixed models: a practical guide for ecology and evolution. Trends Ecology Evolution 2009; 24(3): 127-35.
- 4- nori Jlyany K, Muhammad K, Azam K, Eshraghian M, Zeraati H, Akaberi A, et al. Application of logistic regression mixed model on the combined factors of goiter disease observed data on health and disease.
- 5- Bapat R. Linear Mixed Models. Linear Algebra and Linear Models: Springer; 2012. p. 99-114.
- 6- McCulloch CE, Neuhaus JM. Generalized linear mixed models. Encyclopedia of Environmetrics 2013.
- 7- McCulloch CE. Generalized linear mixed models: Wiley Online Library; 2006.
- 8- Isik F. Generalized Linear Mixed Models. Fourth International Workshop on the Genetics of Host-Parasite Interactions in Forestry; 31 July; Eugene, Oregon, USA. North Carolina State University North Carolina State University 2011. p. 16-24.
- 9- Agresti A. Categorical Data Analysis. simultaneously in Canada. 492-3 p.
- 10- McCullagh P, Nelder J. Generalized Linear Models. Edition S, Editor 1989.
- 11- Venables WN, Dichmont CM. GLMs, GAMs and GLMMs: an overview of theory for applications in fisheries research. Fisheries Res 2004; 70(2): 319-37.



- 12- Abramovich F, Grinshtein V. Model selection and minimax estimation in generalized linear models. arXiv preprint arXiv:14098491.2014.
- 13- Myers RH, Montgomery DC, Vining GG, Robinson TJ. Generalized linear models: with applications in engineering and the sciences: John Wiley & Sons; 2012.
- 14- Fitzmaurice GM, Laird NM, Ware JH. Generalized Linear Mixed Models. applied longitudinal analysis: John Wiley & Sons; 2012.
- 15- Wiley J, Sons. Generalized Linear Mixed Models. Encyclopedia of statistics in behavioral science. p. 1-4.
- 16- Hadfield JD. MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R package. J Statistical Software 2010; 33(2): 1-22.
- 17- Groll A, Tutz G. Variable selection for generalized linear mixed models by L₁-penalized estimation. Statistics and Computing 2014; 24(2): 137-54.
- 18- Nakagawa S, Schielzeth H. A general and simple method for obtaining R² from generalized linear mixed-effects models. Methods in Ecology and Evolution. 2013; 4(2): 133-42.
- 19- Diabetes Blue Circle Symbol. International Diabetes Federation 2006.
- 20- Larejani B, Zahedi F. Epidemiology of diabetes mellitus in Iran. Iran J Diabetes Metabolism 2001; 1(1): 1-8.
- 21- M. H. Definition and classification of diabetes mellitus and the new criteria for diagnosis. 2000.
- 22- Association AD. Diagnosis and classification of diabetes mellitus. Diabetes Care. 2008 .(Supplement 1): S55-S60.
- 23- Consultation W. Definition, diagnosis and classification of diabetes mellitus and its complications: Part; 1999.
- 24- Levitt NS, Katzenellenbogen JM, Bradshaw D, Hoffman MN, Bonnici F. The prevalence and identification of risk factors for NIDDM in urban Africans in Cape Town, South Africa. Diabetes Care 1993; 16(4): 601-7.
- 25- Skronal D HS. Multilevel Logistic Regression. Statistics in Medicine. 2002; 3: 411-20



26- Merati M, Feizei A, Bager Nejad M. Prevalence of high blood pressure and diabetes and risk factors associated with them, based on a large study of the general population-an application of multivariate logistic regression models. *Health Sys Res* 2012; 8(2): 193-203.

Archive of SID



Comparison of Generalized Linear Mixed and Generalized Linear Models in Determining Type II Diabetes Related Factors in Yazd

Fallahzadeh H(Ph.D)¹, Asadi F(M.Sc)², Rahmanian M(Ph.D)³, Emadi M(Ph.D)⁴

1. Professor of Biostatistics, Department of Biostatistics, Shahid Sadoughi University of Medical Sciences, Yazd, Iran.
2. Corresponding Author: Graduate student of Biostatistics, Shahid Sadoughi University of Medical Sciences, Yazd, Iran
3. Professor of Endocrine Diseases And Metabolism, Diabetes Research Center, Yazd, Iran.
4. Associate professor, Department of Statistics, Ferdowsi University of Mashhad, Iran.

Abstract

Introduction: Diabetes mellitus is a chronic disease, its prevalence is very high and is increasing recently. In this study, in addition to determining type II diabetes related factors, we compared the generalized linear and generalized linear mixed models.

Methods: Data is related to research project to investigate the epidemiological characteristics of diabetes in adults aged 30 years and older in the province of Yazd. In this study, 2,795 people were screened with a blood glucose test for diabetes. We for data analysis by the mixed logistic and ordinary logistic regression used the R software.

Results: In this study, four variables of family history of diabetes, age, body mass index and waist circumference to hip circumference were significant in both models (p-value <.001). Job was a significant variable in the ordinary logistic regression model in level significant .1 but not significant in the mixed logistic regression model. The education, area of housing and gender not significant in neither logistic mixed model nor ordinary logistic model. According to the values of the odds ratio also, we saw quite differences between the two models. Judging from standard error of the coefficients and comparison of the their values in both models seen underestimate in ordinary logistic regression model

Conclusion: The use of generalized linear mixed models lead to more accurate results and prevents underestimated standard error of the coefficients.

Keywords: Diabetes; Generalized linear mixed models; Logistic regression; Laplace approximation; Generalized linear models