مقایسه مدل‌های آمیخته خطی تعمیم‌یافته و مدل‌های خطی تعمیم‌یافته در تعیین عوامل مرتبط با بیماری دیابت نوع۲ در استان یزد

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

۱. استاد آمارزیستی،گروه آمار واپیدمیولوژی،دانشگاه علوم پزشکی و خدمات بهداشتی درمانی شهید صدوقی یزد

۲. نویسنده مسئول: دانشجوی کارشناسی‌ارشد 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 ومتابولیسم، مرکز تحقیقات دیابت یزد

۴. دانشیار گروه آمار، دانشگاه فردوسی مشهد

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

چکیده

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

روش بررسی: داده‌های این مطالعه مربوط به طرح تحقیقاتی بررسی شاخص‌های اپیدمیولوژیک بیماری دیابت بزرگسالان در گروه سنی ۳۰سال وبالاتر 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 ورگرسیون لجستیک معمولی از نرم افزار R استفاده شد.

یافته‌ها: در این مطالعه متغیرهای سابقه خانوادگی ابتلا به دیابت، سن، شاخص توده بدنی و دورکمر به دور باسن در هردو مدل معنی‌دار شد ($p<.۰۰۱$) ، با این تفاوت متغیر که شغل در مدل رگرسیون لجستیک معمولی در سطح۰/۱ معنی‌دار بود ولی در مدل رگرسیون لجستیک آمیخته معنی دار نبود. همچنین، متغیرهای مساحت خانه، سطح تحصیلات و جنسیت در هیچ یک معنی دار نشد. با توجه به مقادیر نسبت شانس نیز در برخی از آنها شاهد تفاوت قابل‌توجهی بین دو مدل هستیم. با توجه به خطای استاندارد ضرایب ومقایسه مقادیر آن در دو مدل، شاهد کم برآوردی در مدل رگرسیون لجستیک معمولی بودیم.

نتیجه‌گیری: بکارگیری مدل‌های آمیخته تعمیم‌یافته منجر به نتایج دقیق‌تر می‌شود و ازکم برآوردی خطای استاندارد ضرایب جلوگیری می‌کند.

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

مقدمه

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 ویا کشور ویا اندازه‌گیری فشار خون افراد در طول زمان. در این‌گونه داده‌ها بدلیل وجود همبستگی درون خوشه ها(مثلا خانواده، شهر، استان ویا کشور در مثال اول وافراد در مثال دوم) فرض استقلال برقرار نمی‌باشد وبه همین دلیل مدل‌های خطی (۱،۲) برای تحلیل این داده‌ها مناسب نیستند. گاهی اوقات محققان به اشتباه این همبستگی را نادیده می‌گیرند ویا 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 وفاصله اطمینان وهمچنین آزمونها تاثیر زیادی خواهد داشت(۴). پس باید مدلی را استفاده کرد که این همبستگی ها 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 وسلسه مراتبی است که در علوم مختلف از جمله کشاورزی، ژئوفیزیک، زیست شناسی وپزشکی 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 ومعمولا به صورت شمارشی ویا دوحالتی است. در اینگونه موارد بسیاری از محققان سعی می‌کنند با تبدیل داده ها آنها را نرمال کنند ولی همیشه اینکار امکان‌پذیر نیست و باید از مدلهای تعمیم یافته استفاده کرد. مدلهای خطی تعمیم یافته Generalized Linear (Models)(۱۳-۱۰) کلاسی از مدلهای رگرسیونی است که شامل مدلهای رگرسیون خطی می‌شود وعلاوه 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 وتابع اتصالی از توزیع‌های خانواده نمایی(مثل دوجمله‌ای وپواسن)گسترش داده اندومدلهای خطی آمیخته تعمیم یافته با اضافه کردن اثرات تصادفی (خوشه) در پیشگوهای 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 تعمیم‌یافته است که بیشترین کاربرد را در مطالعات پزشکی دارد. شکل کلی این مدل به صورت زیر است:

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

که β_0 عرض از مبدا جامعه است وبه 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 x1 وقتی سایر متغیرها ثابت نگه‌داشته شده‌اند تفسیر می‌شود(۱۴). همان‌طور که گفته شد مدلهای آمیخته خطی تعمیم‌یافته اثرات تصادفی را به پیشگوی خطی اضافه می‌کنند. پس مدل رگرسیون لجستیک آمیخته را می‌توان به صورت زیر بیان کرد:

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

که در آن b_i ها ضرایب اثرات تصادفی هستند وفرض 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 وموجب پیچیدگی مدل ونتایج نادرست می‌گردد ولی اگر خوشه را به عنوان اثر تصادفی درنظر بگیریم آنگاه تنها یک پارامتر شرطی در مدل پراکندگی را توصیف خواهد کرد(۹). مدلهای آمیخته خطی تعمیم‌یافته (Generalized Linear Mixed Models) (۱۵و۱۴و۹و۸) ترکیبی از مدلهای خطی تعمیم‌یافته ومدلهای آمیخته خطی می‌باشند که برای تحلیل داد ه‌های همبسته با پاسخ غیر نرمال بکار میروند(۱۱). گرچه از زمان معرفی این مدلها سالها میگذرد ولی در سال‌های اخیر استفاده از این مدلهای وتحقیق در مورد روش‌های برآورد آنها بدلیل پیشرفت نرم افزارهای مختلف بسیار مورد توجه محققان قرارگرفته است وموضوع اصلی بسیاری از تحقیقات است(۱۸-۱۶،۱۱).

دیابت از جمله بیماری‌های مزمن در چند دهه اخیر است که شیوع آن درتمامی کشورها ازجمله 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 پانکراس منجر به نقص تولید انسولین می‌شود و در نوع دو مقاوت 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 خوشه ای 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 دومدل خطی تعمیم‌یافته وخطی آمیخته تعمیم‌یافته می‌پردازیم.

روش بررسی

این مطالعه از نوع توصیفی- تحلیلی وبه صورت مقطعی بوده وداده‌های آن مربوط به طرح تحقیقاتی بررسی شاخص‌های اپیدمیولوژیک بیماری دیابت بزرگسالان در گروه سنی ۳۰سال وبالاتر 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 داده ها به نرم افزارR3.0.1 متغیرهای سن، سابقه خانوادگی ابتلا به دیابت، شاخص توده بدنی، تحصیلات، دور کمر، شغل، دورباسن، مساحت خانه وجنسیت به عنوان متغیرهای مستقل ومتغیرهای خانوار ومحل سکونت به عنوان اثر تصادفی در نظر گرفته شدند. از آنجایی که متغیر پاسخ یعنی ابتلا به دیابت دوحالتی بود(دارد،ندارد) از تابع اتصال لوجیت استفاده شد. سپس با

استفاده از دستور glmer مدل رگرسیون لجستیک آمیخته وتوسط دستور glm مدل رگرسیون لجستیک به داده ها برازش داده شد. دستور glmer اثرات ثابت واثرات 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، درصورتی‌که اثرات تصادفی تنها مربوط به نمونه است.

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

لازم به ذکر است که در مدل نهایی متغیر جنسیت واثر متقابل تحصیلات وشغل به دلیل عدم معنی‌داری وبهتر شدن برازش از مدل حذف شدند. معیار مقایسه مدلها معیار آکائیک و روش برازش مدل لجستیک آمیخته روش لاپلاس بود.

یافته‌ها

در جدول شماره یک ضرایب ومقادیر معنی داری هر یک از متغیرهای موجود در برازش مدل نهایی ودر جدول دو مقادیر نسبت شانس و فاصله اطمینان در هردو مدل آورده شده است. براساس نتایج حاصل از برازش مدل رگرسیون لجستیک معمولی متغیرهای شغل، تحصیلات ومساحت خانه معنی‌دار نشده است. همچنین شانس ابتلا به دیابت در افرادی که سابقه فامیلی ابتلا به دیابت در آنها وجود نداشته ۵۲ درصد کمتر از آنهایی است که در بستگان آنها سابقه ابتلا به دیابت وجود داشته است و با بالارفتن سن نیز شانس ابتلا به دیابت نیز به شدت بالا رفته است. از طرفی با توجه به این دوجدول خطر مذکور در افراد لاغر ونرمال تفاوت چندانی نداشته در صورتی که در افراد چاق و دارای اضافه وزن شانس ابتلا به دیابت به ترتیب ۳/۷۲ و۵/۷۷ برابر افراد لاغر است.

جدول۱: مقادیر ضرایب متغیرها ومعنی‌داری آنها براساس مدل رگرسیون لجستیک آمیخته ومعمولی

متغیر		مدل رگرسیون لجستیک آمیخته		مدل رگرسیون لجستیک معمولی	
		ضریب در مدل(SE)	مقدار معنی‌داری	ضریب در مدل(SE)	مقدارمعنی‌داری
سابقه خانوادگی		(۰/۱۰)-۰/۷۰	<۰/۰۰۱	(۰/۱۰)-۰/۷۳	<۰/۰۰۱
گروه سنی[1]	۴۰-۴۹	(۰/۱۴)۰/۵۵	<۰/۰۰۱	(۰/۱۳)۰/۵۴	<۰/۰۰۱
	۵۰-۶۴	(۰/۱۵)۱/۴۱	<۰/۰۰۱	(۰/۱۴)۱/۳۱	<۰/۰۰۱
	۶۵+	(۰/۱۸)۱/۸۷	<۰/۰۰۱	(۰/۱۶)۱/۷۵	<۰/۰۰۱
شاخص توده بدنی[2]	نرمال	(۰/۲۱)۰/۳۲	۰/۱۲	(۰/۱۹)۰/۳۲	۰/۱۰
	چاق	(۰/۲۱)۰/۸۳	<۰/۰۰۱	(۰/۱۹)۰/۷۸	<۰/۰۰۱
	اضافه وزن	(۰/۲۳)۰/۹۲	<۰/۰۰۱	(۰/۲۱)۰/۸۴	<۰/۰۰۱
تحصیلات[3]		(۰/۱۵)-۰/۱۵	۰/۳۱	(۰/۱۴)-۰/۱۹	۰/۱۲
دورکمر به دورباسن		(۰/۵۳)۱/۸۸	<۰/۰۰۱	(۰/۵۱)۱/۶۷	<۰/۰۰۱
مساحت خانه		(۰/۰۰)۰/۰۰	۰/۱۸	(۰/۰۰)۰/۰۰	۰/۳۰
شغل[5]	کارگر	(۰/۲۴)-۰/۲۳	۰/۳۳	(۰/۲۲)-۰/۱۰	۰/۶۳
	خانه دار	(۰/۱۹)-۰/۰۹	۰/۶۱	(۰/۱۸)-۰/۱۰	۰/۵۷
	کشاورز	(۰/۳۵)-۰/۴۷	۰/۱۸	(۰/۳۳)-۰/۵۵	۰/۰۹
	بازنشسته	(۰/۲۵)-۰/۲۸	۰/۲۴	(۰/۲۳)-۰/۲۴	۰/۳۰
	سایر	(۰/۲۰)-۰/۰۶	۰/۷۳	(۰/۱۹)-۰/۰۸	۰/۶۵
واریانس اثر تصادفی اول(خانواده)		۰/۱۲	۰/۴۲	-	-
واریانس اثر تصادفی دوم(محل سکونت)		۰/۳۸	<۰/۰۰۱	-	-

گروه‌های مرجع به ترتیب شامل: ۱.گروه سنی۳۹-۳۰ * ۲.لاغر * ۳.زیر دیپلم * ۴.مرد * ۵.کارمند می‌باشند.

جدول۲: مقادیر نسبت شانس وفواصل اطمینان متغیرها براساس برازش مدلهای رگرسیون لجستیک آمیخته و معمولی

متغیر		نسبت شانس (OR)		فاصله اطمینان(۹۵٪)			
		لجستیک آمیخته	لجستیک معمولی	لجستیک آمیخته		لجستیک معمولی	
				حدپایین	حدبالا	حد پایین	حدبالا
سابقه خانوادگی(ندارد)		۰/۴۹	۰/۴۸	۰/۳۹	۰/۶۱	۰/۳۹	۰/۵۸
گروه سنی [1]	۴۰-۴۹	۱/۷۳	۱/۷۲	۱/۳۱	۲/۳۰	۱/۳۲	۲/۲۴
	۵۰-۶۴	۴/۱۲	۳/۷۲	۳/۰۳	۵/۶۰	۲/۸	۴/۹۳
	۶۵+	۶/۵۳	۵/۷۷	۴/۵۴	۹/۳۷	۴/۱۵	۸/۰۴
شاخص توده بدنی [2]	نرمال	۱/۳۸	۱/۳۷	۰/۹۱	۲/۰۹	۰/۹۳	۲/۰۳
	چاق	۲/۳۱	۲/۱۹	۱/۵۲	۳/۴۹	۱/۴۸	۳/۲۳
	اضافه وزن	۲/۵۲	۲/۳۲	۱/۶۰	۳/۹۶	۱/۵۱	۳/۵۴
تحصیلات [3]		۰/۸۵	۰/۸۲	۰/۶۳	۱/۱۵	۰/۶۲	۱/۰۸
دورکمر به دورباسن		۶/۵۸	۵/۳۲	۲/۲۸	۱۸/۹۶	۱/۹۵	۱۴/۵
مساحت خانه		۱/۰۰	۱/۰۰	۰/۹۹	۱/۰۰	۰/۹۹	۱/۰۰
شغل [4]	کارگر	۰/۷۹	۰/۸۹	۰/۴۸	۱/۲۷	۰/۵۷	۱/۴۰
	خانه دار	۰/۹۰	۰/۹۰	۰/۶۲	۱/۳۲	۰/۶۳	۱/۲۹
	کشاورز	۰/۶۲	۰/۵۷	۰/۳۰	۱/۲۵	۰/۲۹	۱/۱۰
	بازنشسته	۰/۷۴	۰/۷۸	۰/۴۵	۱/۲۲	۰/۴۹	۱/۲۴
	سایر	۰/۹۳	۰/۹۱	۰/۶۱	۱/۴۰	۰/۶۲	۱/۳۴

گروه‌های مرجع به ترتیب شامل: ۱.گروه سنی۳۹-۳۰ * ۲.لاغر *۳.زیر دیپلم*۴.کارمند می باشند.

همچنین این خطر در افراد کشاورز۴۳ درصد کمتر از کارمندان بود. مقدار نسبت شانس شاخص دور کمر به دور باسن در این مدل ۵/۳۲ با فاصله اطمینان (۱۴/۵، ۱/۹۵) 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 کمترازافراد زیر دیپلم ونسبت 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 بالارفتن سن نیز نسبت شانس ابتلا به دیابت بالا رفته طوری که افراد بالای ۶۵ سال ۶/۵۳ برابر نسبت به افراد درگروه سنی ۳۹-۳۰ سال بیشتر در معرض خطر ابتلا به دیابت هستند. گرچه شغل از عوامل موثر بر دیابت تشخیص داده نشد، ولی همه آنها شانس کمتری در ابتلا به دیابت نسبت به کارمندان داشته وکمترین آن مربوط به کشاورزان بود.

همچنین یافته ها حاکی از آن است که با یک واحد افزایش در شاخص دورکمر 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 وقابل توجه است. همچنین برآورد مقدار واریانس اثر تصادفی دوم یعنی ناحیه برابر ۰/۳۸ با خطای معیار۰/۶۲ بدست آمد($P<۰/۰۰۱$) که نشان‌دهنده همبستگی معنی‌داری بین افراد ساکن در هرناحیه از استان یزد با ابتلا به دیابت است.

بحث و نتیجه‌گیری

به طور کلی عوامل خطر متعددی نظیرسن، فشار خون، فشارهای روحی وروانی، 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 معمولا در عمل با مشاهدات گسسته زیادی روبرو هستیم. مک ونلدر(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)، وواریانس مشاهدات ثابت است. از طرفی در بسیاری از داده ها بخصوص داده‌های که به صورت طولی ویا خوشه ای جمع آوری شده اند شاهد نوعی همبستگی در داده ها هستیم.گسترش یافته مدلهای خطی تعمیم یافته یعنی مدلهای خطی آمیخته تعمیم یافته با اصلاح موارد دو وسه موجب در نظر گرفتن این همبستگی ها می‌شود به این صورت که میانگین مشاهدات وابسته به تابعی خطی از برخی کوریت‌ها ازطریق 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.دراین مطالعه ما به مقایسه دو مدل رگرسیونی لجستیک آمیخته ومعمولی پرداختیم.در یک مطالعه مقطعی که به منظور بررسی عوامل خطر دیابت در سال ۱۹۹۶ روی افراد ۱۵ساله وبالاتر در ممری مرکز شهر کیپ تاون در آفریقای جنوبی انجام شد نتایج آنالیز رگرسیون نشان داد که سن،دور کمر،مصرف انرژی کم وسابقه خانوادگی ابتلا به دیابت معنی‌دار بود در حالی که جنس و چاقی معنی دار نبودند(۲۴). در این مطالعه نیز متغیرهای سابقه خانوادگی، سن، شاخص توده بدنی و دورکمر به دور باسن در هردو مدل معنی دار شده است، ولی متغیر جنس معنی دار نبود، در حالی که در مطالعه انجام شده در اصفهان صورت گرفت متغیرهای سن، جنس، فشارخون، شاخص توده بدنی وسابقه فامیلی ابتلا به دیابت در ابتلا به دیابت معنی‌دار شد(۲۶). در این مطالعه متغیر شغل در مدل رگرسیون لجستیک معمولی در سطح۰/۱ معنی دار بود ولی در مدل رگرسیون لجستیک آمیخته معنی دار نبود. همچنین متغیرهای جنس، مساحت خانه وتحصیلات در هیچ یک معنی دار 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 ومقایسه مقادیر آن در دو مدل شاهد کم برآوردی در مدل رگرسیون لجستیک معمولی بودیم.در مطالعه ای که اسکرنال وهمکارانش در سال ۲۰۰۲ انجام دادند نیز به این نتیجه رسیدند که در نظر نگرفتن همبستگی موجود در داده ها 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 درنظر گرفتن همبستگی در داده‌های سلسله مراتبی نتایج دقیق تری نسبت به مدل رگرسیون لجستیک ارائه می‌دهد.

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

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

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 داده ها را در اختیار نویسندگان این مقاله قرار دادند تشکرو قدر دانی می‌شود. همچنین از جناب آقای دکتر مهدی یاسری استادیار گروه اپیدمیولوژی و آمارزیستی دانشگاه تهران به دلیل کمک‌های بی دریغشان بی‌نهایت سپاسگزاری 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 L 1-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 Federation2006.

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.

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, it's 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