مقایسه نتایج مدل رگرسیون لجستیک معمولی و لجستیک نیرومند در مدل‌بندی عوامل مرتبط با بیماری پره‌دیابت

فاطمه عزیزی مایوان[1]
مهدی جباری نوقابی[2]، علی تقی‌پور[3]
محمدتقی شاکری[4]*، مهسا مکرم[5]

۱- گروه بهداشت عمومی، دانشگاه علوم پزشکی نیشابور، نیشابور، ایران.
۲- گروه آموزشی آمار، دانشکده علوم ریاضی، دانشگاه فردوسی، مشهد، ایران.
۳- گروه اپیدمیولوژی، مرکز تحقیقات عوامل اجتماعی موثر بر سلامت، دانشگاه علوم پزشکی مشهد، مشهد، ایران.
۴- گروه آمار زیستی، مرکز تحقیقات عوامل اجتماعی موثر بر سلامت، دانشگاه علوم پزشکی مشهد، مشهد، ایران.
۵- گروه جمعیت‌شناسی، کمیته تحقیقات دانشجویی، دانشگاه آزاد اسلامی، واحد تهران مرکزی، تهران، ایران.

*نویسنده مسئول: مشهد، خیابان دانشگاه، دانشکده بهداشت عمومی. تلفن: ۳۸۵۱۵۱۱۶-۰۵۱
E-mail: shakerimt@mums.ac.ir

چکیده

دریافت: ۱۳۹۷/۰۱/۲۰ ویرایش: ۱۳۹۷/۰۱/۲۷ پذیرش: ۱۳۹۷/۰۷/۱۵ آنلاین: ۱۳۹۷/۰۷/۳۰

زمینه و هدف: با توجه به اینکه خطر ابتلا به دیابت در افراد پره‌دیابتیک بسیار بالا است، تعیین عوامل موثر بر پره‌دیابت دارای اهمیت می‌باشد. این مطالعه با هدف مقایسه نتایج مدل رگرسیون لجستیک معمولی و لجستیک نیرومند در مدل‌بندی عوامل مرتبط با بیماری پره‌دیابت انجام شد.

روش بررسی: این مطالعه که از نوع مقطعی-تحلیلی است روی ٦٤٦٠ نفر از افراد بالای ۳۰ سال، شرکت‌کننده در طرح غربالگری دیابت دانشگاه علوم پزشکی مشهد، از مهر تا آذر ۱۳۸۹ انجام شد. با توجه به میزان قندخون ناشتای افراد، ٥٤١٤ نفر سالم و ١٠٤٦ نفر به‌عنوان پره‌دیابتیک شناسایی شدند. سن، جنس، نمایه توده بدن، فشارخون سیستولیک، فشارخون دیاستولیک و نسبت کمر به باسن در مورد هر فرد اندازه‌گیری شد. مدل رگرسیون لجستیک معمولی روی داده‌ها برازش شد. سپس داده‌های پرت مشخص و سه مدل نیرومند Mallow، WBY و BY برازش شد. آنگاه مدل‌ها با هم مقایسه گردیدند.

یافته‌ها: متغیرهای سن، نمایه توده بدن و فشارخون سیستولیک در همه مدل‌ها از لحاظ آماری معنادار شدند (P<۰/۰۱) و متغیر نسبت کمر به باسن معنادار نشد (P>۰/۱). تعداد ۵۵۲ داده‌ی پرت با خطای بدرده‌بندی در مدل معمولی وجود داشت. مقادیر کای دو پیرسون و سطح زیرمنحنی راک در مدل Mallow به‌طور تقریبی فرقی با مدل معمولی نداشت. اما در مدل‌های WBY و BY بهنسبت بیشتر بود.

نتیجه‌گیری: با توجه به نتایج این پژوهش سن بالا، نمایه توده بدنی و فشارخون بالا در ابتلا به بیماری پره‌دیابت موثر می‌باشند. همچنین مدل‌های رگرسیون نیرومند WBY و BY برازش بهتر و توان پیشگویی بالاتری نسبت به رگرسیون لجستیک معمولی در مدل‌بندی عوامل گفته‌شده در ارتباط با پره‌دیابت دارند.

کلمات کلیدی: نمایه توده بدنی، دیابت شیرین، مدل‌های لجستیک، مرحله پره‌دیابت.

مقدمه

دیابت یکی از شایع‌ترین بیماری‌های مزمن در تقریبا تمامی کشورها است و به‌خاطر تغییر الگوی زندگی مردم به سمت تحرک کمتر و چاقی بیشتر همچنان روند افزایشی دارد. دیابت نوع دو شایع‌ترین نوع دیابت در دنیاست و به‌طور تقریبی ۹۰٪ بیماران دیابتی را شامل می‌شود.[1] شیوع جهانی دیابت در سال ٢٠١٤ در میان افراد بزرگسال بالای ۱۸ سال، ۹٪ برآورد شد.[2] در سال ۲۰۱۲ تعداد مرگ‌هایی که در دنیا به‌طور مستقیم بر اثر دیابت رخ داده‌اند، ۱/۵ میلیون مورد برآورد شد. بیشتر از ۸۰٪ موارد مرگ ناشی از دیابت در کشورهای با درآمد کم یا متوسط رخ داده است.[3] سازمان جهانی بهداشت پیش‌بینی کرده است که دیابت هفتمین علت مرگ در سال ۲۰۳۰ خواهد بود.[5] شیوع دیابت در ایران در سال ۲۰۱۱، ۱۱/۱٪ برآورد شد که این میزان تا سال ۲۰۳۰ به ۱۲/۸٪ خواهد رسید.[1] این بیماری خطر ابتلا به بیماری

قلبی و سکته را افزایش می‌دهد. همچنین باعث صدمه‌ی کلیه، ایجاد زخم در پا و نیاز احتمالی به قطع عضو و آسیب در شبکیه چشم می‌شود.[۵و۶] خطر مرگ در افراد دیابتی حداقل دو برابر افراد غیردیابتی است.[۷]

پره‌دیابت یا پیش از دیابت مرحله‌ای است که در آن اگرچه قند خون بالاتر از حد نرمال است اما آنقدر بالا نیست که بتوان آن را در دسته‌ی دیابت نوع ۲ به‌حساب آورد. در این مرحله، قند خون ناشتا بین ۱۲۵-۱۰۰ mg/dl و یا دو ساعت پس از آزمایش تحمل گلوکز با ۷۵ g گلوکز، قند خون بین ۱۹۹-۱۴۰ mg/dl می‌باشد. براساس گزارش انجمن دیابت آمریکا برخی از عوامل خطرزای دیابت نوع ۲ عبارتند از داشتن اضافه وزن، سن بیشتر از ۴۵ سال، عدم فعالیت بدنی، پیشینه‌ی خانوادگی دیابت، پرفشاری خون (فشارخون بالاتر یا مساوی ۱۴۰/۹۰ mmHg یا دریافت درمان پرفشاری خون)، کلسترول HDL سرم برابر یا کمتر از ۳۵ mg/dl و یا تری‌گلیسرید سرم برابر یا بالاتر از ۲۵۰ mg/dl.[۸] احتمال اینکه افراد مبتلا به پره‌دیابت بدون رعایت پیشگیری طی ده سال یا کمتر مبتلا به دیابت شوند، بسیار بالا است مگر اینکه شیوه‌ی زندگی خود را با تلاش برای کاهش وزن، رعایت رژیم غذایی مناسب و تحرک بیشتر، تغییر دهند.[۹] بنابراین شناسایی افراد مبتلا به پره‌دیابت و تعیین الگوی ابتلا به این بیماری جهت درمان و پیشگیری از ابتلا به آن بسیار با اهمیت می‌باشد. انجام این کار نیازمند به روش‌های مناسب آماری است. رگرسیون لجستیک یک روش استاندارد برای مدل‌سازی و بیان ارتباط بین یک متغیر پاسخ کیفی با یک یا چند متغیر توضیحی کیفی یا کمی است. در واقع تفاوت اصلی رگرسیون لجستیک با رگرسیون خطی شرط کیفی بودن متغیر پاسخ است.[۱۰]

روش معمول برآورد پارامترها در مدل رگرسیون لجستیک، روش ماکزیمم درست‌نمایی (MLE) است. اما برآورد پارامترها در این روش بسیار تحت تاثیر وجود داده‌های پرت قرار می‌گیرد و ممکن است در حضور این داده‌ها بسیار غیر واقعی ارایه شود.[۱۱] این مطالعه با هدف مدل‌بندی عوامل مرتبط با بیماری پره‌دیابت با استفاده از مدل‌های رگرسیون لجستیک معمولی و نیرومند و مقایسه نتایج آن‌ها انجام گرفت.

روش بررسی

این مطالعه به‌صورت مقطعی- تحلیلی بر روی داده‌های مطالعه‌ای انجام گرفت که تحت عنوان غربالگری دیابت در سطح کلان شهر مشهد از نیمه‌ی مهر تا اواخر آذر ۱۳۸۹ در دانشگاه علوم پزشکی مشهد انجام گرفت. تمامی افراد بالای ۳۰ سال و داوطلب ساکن در بافت مرکزی شهر مشهد وارد مطالعه شده و افراد مبتلا به دیابت از مطالعه خارج گردیدند. از مجموع داده‌های حاصل پس از مدیریت داده‌ها و حذف داده‌های ناقص حجم جمعیتی بالغ بر ۶۴۶۰ نفر در مطالعه کنونی وارد شدند. این طرح پژوهشی با کد اخلاق IR.MUMS.REC.1395.363 در دانشگاه علوم پزشکی مشهد به تصویب رسیده است.

ازآنجایی‌که نتایج حاصل از مدل رگرسیون لجستیک در حجم نمونه‌های بالا دقیق‌تر می‌باشد، تمامی داده‌های مناسب و کامل این مطالعه مورد تحلیل آماری قرار گرفت. تمامی اندازه‌گیری‌ها شامل سن و قد، وزن، جنس، فشارخون سیستولیک، فشارخون دیاستولیک، قندخون ناشتا، دور کمر و دور باسن در مورد هر فرد اندازه‌گیری و ثبت شد. منظور از دور کمر، محیط دور کمر واقع در قسمت بالای ناف و دور باسن، محیط دور پهن‌ترین و برجسته‌ترین قسمت بدن در پایین خط کمر می‌باشد. به‌منظور استفاده از شاخص‌های چاقی مناسب به‌جای قد و وزن، از نمایه توده بدن (BMI) که ترکیبی از هر دو است، استفاده شد و به‌جای دور کمر و دور باسن از شاخص نسبت دور کمر به باسن استفاده گردید. میزان قندخون ناشتا به‌عنوان معیار تشخیص پره‌دیابت در نظر گرفته شد. به این صورت که افراد با قندخون ناشتای ۱۲۵-۱۰۰ mg/dl به‌عنوان پره‌دیابتیک و کمتر از ۱۰۰ به‌عنوان سالم شناسایی شدند. برای انتخاب مدل، ارتباط آماری تک‌تک متغیرهای موردنظر با ابتلا به پره‌دیابت را با استفاده از Chi-square test و Independent samples t-test در سطح معناداری P<۰/۲ بررسی کردیم. سپس مدل رگرسیون لجستیک معمولی را با استفاده از متغیرهای معنادار شده، روی داده‌ها برازش دادیم. آنگاه داده‌های پرت با توجه به شاخص باقیمانده‌ی استیودنتیزه‌ی پیرسون مشخص گردیدند. سه مدل نیرومند Mallow، WBY و BY که بر مبنای نیرومندتر کردن روش MLE است،[۱۲] روی داده‌ها برازش شد. سطح معناداری P<۰/۰۵ در نظر گرفته شد. سپس این مدل‌ها با استفاده از معیار Chi-square پیرسون از لحاظ نیکویی برازش مورد مقایسه قرار گرفتند. همچنین سطح زیرمنحنی راک برای مقایسه‌ی توان پیش‌بینی مدل‌ها محاسبه گردید. نرم‌افزار مورد استفاده R Project for Statistical Computing, Version R 3.1.2 (www.r-project.org) بود.

یافته‌ها

در میان ٦٤٦٠ نفر شرکت‌کننده در این مطالعه ٤٥٠٧ نفر (٦٩/٨٪) مرد و ١٩٥٣ نفر (٣٠/٢٪) زن بودند که ١٠٤٦ نفر (١٦/٢٪) از مجموع آنان پره‌دیابتیک و ٥٤١٤ نفر (٨٣/٨٪) سالم بودند. میانگین سن و ویژگی‌های بالینی افراد سالم و پره‌دیابتیک در جدول ١ آمده است. Chi-square test حاکی از همگنی دو گروه دیابتیک و پره‌دیابتیک به لحاظ جنس بود (P=٠/٤٣١). اختلاف میانگین‌های هر یک از متغیرهای توضیحی کمی در افراد سالم و پره‌دیابتیک با توجه به Independent samples t-test معنادار شد (P=٠/٠٠١).

در مدل رگرسیون لجستیک چندگانه متغیر فشارخون دیاستولیک را به‌خاطر معنادار نشدن (P=٠/٢٨١) از مدل حذف شد. برای بررسی هم‌خطی میان متغیرهای توضیحی سن، نمایه توده بدن، فشارخون سیستولیک و نسبت دور کمر به دور باسن، مقادیر شاخص تورم واریانس محاسبه شد. این مقادیر همگی نزدیک به یک بودند که بیانگر عدم وجود هم‌خطی قابل توجه بین متغیرها است. برای بررسی خطی بودن رابطه‌ی متغیرهای توضیحی معنادار شده در مدل که همگی به‌صورت کمی هستند، Lowess smoothed logit در برابر هر یک از این متغیرها رسم و خطی بودن رابطه‌ی آن‌ها تایید شد. آزمون نیکویی برازش Chi-square نیز برای این مدل معنادار شد (P=٠/٠٠١).

در مدل نهایی رگرسیون لجستیک معمولی متغیر عرض از مبدا و متغیرهای توضیحی سن، نمایه توده بدن و فشارخون سیستولیک معنادار شدند (P=٠/٠٠١)، اما متغیر کمر به باسن معنادار نشد (P=٠/١٥). اندازه‌ی ضرایب رگرسیونی برآوردشده از این مدل در جدول ٢ آمده است.

جدول ١: مقایسه توزیع میانگین متغیرهای فردی و بالینی بیماران پره‌دیابتیک با افراد سالم

متغیرها	**سالم** (انحراف‌معیار± میانگین)	**پره‌دیابتیک** (انحراف‌معیار± میانگین)	P*
سن (سال)	٤٣/٩٠±١٠/٣٥	٤٦/٤٠±١٠/٤٦	٠/٠٠١
نمایه توده بدن (kg/m^2)	٢٦/٥٢±٤/١٣	٢٩/١٧±٤/٣١	٠/٠٠١
فشارخون سیستولیک (mmHg)	١١٥/٨٩±١٤/١٨	١٢٠/٠٩±١٥/٩٥	٠/٠١٣
فشارخون دیاستولیک (mmHg)	٧٤/٢٦±٩/٩٠	٧٦/٥٣±١٠/٤٣	٠/٨٠٥
نسبت کمر به باسن	٠/٩٢±٠/٠٩	٠/٩٣±٠/٠٨	٠/١٥٦

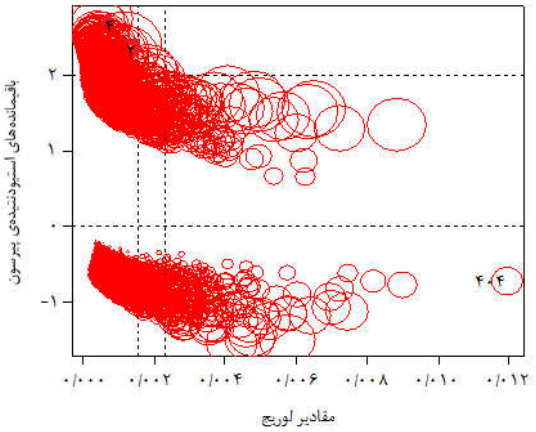
* گزارش شده از مدل رگرسیون لجستیک چندگانه، P<٠/٠٥ معنادار در نظر گرفته شد.

جدول ٢: برآورد ضرایب رگرسیون در مدل‌های مختلف مورد مطالعه

متغیرها	**مدل رگسیون لجستیک معمولی** (انحراف‌معیار± ضریب رگرسیونی)	**مدل Mallow** (انحراف‌معیار± ضریب رگرسیونی)	**مدل WBY** (انحراف‌معیار± ضریب رگرسیونی)	**مدل BY** (انحراف‌معیار± ضریب رگرسیونی)
سن (سال)	٠/٠١٩١±٠/٠٠٣٤	٠/٠١٩١±٠/٠٠٣٤	٠/٠٢٠٤±٠/٠٠٣٣	٠/٠١٩١±٠/٠٠٣٣
نمایه توده بدن (kg/m^2)	٠/١٣٩٢±٠/٠٠٨٤	٠/١٤٠١±٠/٠٠٨٤	٠/١٤٦٣±٠/٠٠٨٦	٠/١٣٨٧±٠/٠٠٨٤
فشارخون سیستولیک (mmHg)	٠/٠٠٨٢±٠/٠٠٢٥	٠/٠٠٨٢±٠/٠٠٢٥	٠/٠٠٦٨±٠/٠٠٢٦	٠/٠٠٨١±٠/٠٠٢٥
نسبت کمر به باسن	−٠/٦١١٣±٠/٤٢٤٨	−٠/٥٧٦٩±٠/٤٢٥٧	−٠/٤٧٧٥±٠/٤٢٢٨	−٠/٦١١٦±٠/٤٢٣٤

جدول ۳: مقادیر Chi-square پیرسون و سطح زیرمنحنی راک برای مدل رگرسیون لجستیک معمولی و سه مدل نیرومند Mallow، WBY و BY

مدل رگرسیون لجستیک	معمولی	Mallow	WBY	BY
Chi-square پیرسون	۶۴۸۳	۶۴۷۶	۶۰۸۶	۶۰۴۵
سطح زیرمنحنی راک (درصد)	۶۸/۸۱	۶۸/۸۲	۷۲/۲۶	۷۲/۳۳



نمودار ۱: نمودار تاثیر

نمودارهای هموارشده‌ی Lowess مربوط به مقادیر احتمال برآوردشده از مدل در برابر باقیمانده‌های استیودنتیده‌ی پیرسون و استیودنتیده‌ی انحراف رسم شد، که در آن‌ها شیب خط به‌طور معناداری غیرصفر شد. این نکته بیانگر عدم کفایت برازش مدل و اثر معنادار داده‌های پرت در مدل بود.

در نمودار تاثیر که در نمودار ۱ رسم شده است، محور افقی آن مربوط به مقادیر لوریج (Leverage) و محور عمودی آن مربوط به مقادیر باقیمانده‌های استیودنتیده‌ی پیرسون و اندازه‌ی دایره‌ها متناسب با مقادیر فاصله‌ی کوک (Cook' distance) مشاهده‌ی مربوطه می‌باشد.

مراحل برازش مدل‌های رگرسیون لجستیک نیرومند مانند مدل رگرسیون لجستیک معمولی انجام شد. خطی بودن رابطه‌ی متغیرهای توضیحی معنادار شده با لوجیت‌ها در هر سه مدل بررسی و تایید گردید. نتایج حاصل از برازش سه مدل نیرومند Mallow، WBY و BY به همراه نتایج مدل رگرسیون لجستیک معمولی در جدول ۲ آورده شده است.

در جدول ۳ معیار Chi-square پیرسون برای مقایسه‌ی نیکویی برازش و سطح زیرمنحنی راک برای مقایسه‌ی توان پیشگویی مدل‌ها به‌کار رفته است. با توجه به جدول، مقدار Chi-square پیرسون برای هر سه مدل نیرومند کمتر از مدل معمولی است. در مدل Mallow این مقدار هفت واحد کمتر از مدل معمولی است. در مدل WBY، ۳۹۷ واحد و در مدل BY، ۴۳۵ واحد کمتر از مدل معمولی است. مقدار سطح زیرمنحنی راک برای مدل Mallow کمابیش برابر با مدل معمولی است. اما برای مدل‌های WBY و BY ۴-۳٪ بیشتر است.

بحث

در همه مدل‌های برازش‌شده در این مطالعه متغیرهای سن، نمایه توده بدن و فشارخون سیستولیک معنادار شدند که این نتیجه منطبق با گزارش سالیانه‌ی انجمن دیابت آمریکا در سال ۲۰۱۵ در رابطه با عوامل خطرزای دیابت و پره‌دیابت می‌باشد.[۸]

در مطالعه Díaz-Redondo و همکاران که در میان جامعه اسپانیایی انجام شده است، ارتباط فشارخون با پره‌دیابت تایید شد.[۱۳] در پژوهش کنونی متغیر نسبت کمر به باسن معنادار نشد درحالی‌که در مطالعات دیگر اثر این متغیر در ابتلا به دیابت تایید شده است.[۱۴و۱۵]

در بسیاری از مطالعاتی که تاکنون در زمینه‌ی دیابت و پره‌دیابت انجام شده است، از رگرسیون لجستیک استفاده شده است.[۱۶-۱۸] اما مطالعه‌ای که در این زمینه از رگرسیون لجستیک نیرومند استفاده کرده باشد، یافت نشد.

تعداد ۵۵۲ داده‌ی پرت از نوع بدرده‌بندی در مدل رگرسیون لجستیک معمولی وجود داشت. داده‌های بدرده‌بندی‌شده دارای لوریج کوچک و باقیمانده استیودنتیده‌ی بزرگ می‌باشند.[۱۲] تمام داده‌های پرت مربوط به افراد پره‌دیابتیک بود. علت آن می‌تواند وجود برخی عوامل دیگر مرتبط با پره‌دیابت باشد که در مطالعه کنونی وارد نشده

است. Hosmer و همکاران نیز در مطالعه خود یکی از عوامل وجود خطای بدرده‌بندی در رگرسیون لجستیک را حذف حداقل یک متغیر توضیحی مهم در مدل بیان کرده‌اند.[۱۹]

در مطالعات متعددی روش‌های مختلف رگرسیون لجستیک نیرومند معرفی و ارایه شده است[۱۱و۲۲-۲۰] که در پژوهش کنونی سه تا از روش‌های ارایه‌شده برای نیرومندتر کردن روش MLE، یعنی روش‌های Mallow، WBY و BY برای برازش روی داده‌ها و مقایسه با روش رگرسیون لجستیک معمولی، انتخاب شد. اندازه‌ی ضرایب در این سه مدل تفاوت چندانی با مدل رگرسیون لجستیک معمولی نداشت. با توجه به مقادیر Chi-square پیرسون، مدل Mallow از لحاظ نیکویی برازش تفاوت چندانی با مدل معمولی نداشت، درحالی‌که دو مدل WBY و BY مدل‌های به‌نسبت بهتری بودند. همچنین قدرت پیشگویی مدل Mallow با توجه به سطح زیرمنحنی راک به‌طور تقریبی تفاوتی با مدل معمولی نداشت. اما دو مدل WBY و BY توان پیشگویی بهتری داشتند (٤-۳٪ بیشتر). با توجه به این نتایج، استفاده از مدل Mallow برای این داده‌ها توصیه نمی‌شود. اما دو مدل WBY و BY را می‌توان به مدل رگرسیون لجستیک معمولی ترجیح داد. در مطالعه Ahmad و همکاران مدل رگرسیون لجستیک معمولی و چهار مدل نیرومند Mallow، شوئپه، WBY و BY، با استفاده از داده‌های شبیه‌سازی‌شده و داده‌های واقعی مورد مقایسه قرار گرفتند و عملکرد آن‌ها در حضور انواع داده‌های پرت ارزیابی شد که در آن مطالعه، مدل WBY و سپس مدل Mallow عملکرد بهتری نسبت به بقیه مدل‌ها داشتند.[۱۲]

در رگرسیون لجستیک، نقاط پرت خطرناک از نوع نقاط نافذ بد می‌باشد.[۲۳] این داده‌ها را می‌توان به‌این‌صورت که مقدار لوریج و باقیمانده‌ی آن‌ها هر دو بزرگ باشد، تشخیص داد. با توجه به نمودار تاثیر رسم‌شده در شکل ۲، چنین داده‌ای در مطالعه کنونی وجود نداشت و همه داده‌های پرت از نوع تنها بدرده‌بندی بودند. بنابراین این انتظار وجود داشت که مدل نیرومند تفاوت خیلی زیادی با مدل معمولی نداشته باشد. همچنین با توجه به مطالعه Sarkar و همکاران با افزایش حجم نمونه، کنترل داده‌های پرت اثر کمتری در برآورد پارامترها دارد.[۲۴] بنابراین حجم نمونه‌ی زیاد عامل دیگری برای چندان بهتر نشدن مدل نیرومند نسبت به مدل معمولی بوده است. با توجه به نتایج این پژوهش، سن بالا، نمایه توده بدنی و فشارخون بالا در ابتلا به بیماری پره‌دیابت موثر می‌باشند. همچنین مدل‌های رگرسیون نیرومند WBY و BY برازش بهتر و توان پیشگویی بالاتری نسبت به مدل رگرسیون لجستیک معمولی در مدل‌بندی عوامل گفته‌شده در ارتباط با پره‌دیابت دارند.

سپاسگزاری: این مقاله حاصل پایان‌نامه تحت عنوان "مقایسه نتایج مدل رگرسیون لجستیک معمولی و مدل رگرسیون لجستیک نیرومند در بیماران پره دیابتیک" در مقطع کارشناسی ارشد در سال ۱۳۹۵ و کد IR.MUMS.REC.1395.363 می‌باشد که با حمایت دانشگاه علوم پزشکی و خدمات بهداشتی درمانی مشهد اجرا شده است.

References

1. Whiting DR, Guariguata L, Weil C, Shaw J. IDF diabetes atlas: global estimates of the prevalence of diabetes for 2011 and 2030. *Diabetes Res Clin Pract* 2011;94(3):311-21.
2. Mendis S, Davis S, Norrving B. Organizational update: the world health organization global status report on noncommunicable diseases 2014; one more landmark step in the combat against stroke and vascular disease. *Stroke* 2015;46(5):e121-2.
3. World Health Organization (WHO). Global Health Estimates: Deaths by Cause, Age, Sex and Country, 2000-2012. Geneva: WHO; 2014.
4. Mathers CD, Loncar D. Projections of global mortality and burden of disease from 2002 to 2030. *PLoS Med* 2006;3(11):e442.
5. World Health Organization (WHO). Global status report on noncommunicable diseases 2010. Description of the global burden of NCDs, their risk factors and determinants. [Internet] Geneva: WHO; 2011 [cited 2018 Aug 15]. Available from: http://www.who.int/nmh/publications/ncd_report2010/en/
6. Morrish NJ, Wang SL, Stevens LK, Fuller JH, Keen H. Mortality and causes of death in the WHO Multinational Study of Vascular Disease in Diabetes. *Diabetologia* 2001;44 Suppl 2:S14-21.
7. Roglic G, Unwin N, Bennett PH, Mathers C, Tuomilehto J, Nag S, et al. The burden of mortality attributable to diabetes: realistic estimates for the year 2000. *Diabetes Care* 2005;28(9):2130-5.
8. American Diabetes Association. Classification and diagnosis of diabetes. *Diabetes Care* 2015;38(Suppl 1):S8-16.
9. Alberti KG, Zimmet PZ. Definition, diagnosis and classification of diabetes mellitus and its complications. Part 1: diagnosis and classification of diabetes mellitus provisional report of a WHO consultation. *Diabet Med* 1998;15(7):539-53.
10. Cabrera AF. Logistic regression analysis in higher education: An applied perspective. In: Smart JC, editor. Higher Education: Handbook of Theory and Research. Vol. 10. New York, NY: Agathon Press; 1994. P. 225-56.
11. Kalina J. Highly robust methods in data mining. *Serbian J Manag* 2013;8(1):9-24.

12. Ahmad S, Ramli NM, Midi H. Robust estimators in logistic regression: a comparative simulation study. *J Modern Appl Stat Methods* 2010;9(2):18.
13. Díaz-Redondo A, Giráldez-García C, Carrillo L, Serrano R, García-Soidán FJ, Artola S, et al. Modifiable risk factors associated with prediabetes in men and women: a cross-sectional analysis of the cohort study in primary health care on the evolution of patients with prediabetes (PREDAPS-Study). *BMC Fam Pract* 2015;16:5.
14. Hadaegh F, Bozorgmanesh M, Safarkhani M, Khalili D, Azizi F. Predictability of body mass index for diabetes: Affected by the presence of metabolic syndrome? *BMC Public Health* 2011;11(1):1.
15. Wei W, Xin X, Shao B, Zeng FF, Love EJ, Wang BY. The relationship between anthropometric indices and type 2 diabetes mellitus among adults in north-east China. *Public Health Nutr* 2015;18(9):1675-83.
16. Aram-Ahmaddi M, Bahrampour A, Najafipour H. Predicting type two diabetes and determination of effectiveness of risk factors applying logistic regression model. *J Kerman Univ Med Sci* 2014;21(2):103-13.
17. Heikes KE, Eddy DM, Arondekar B, Schlessinger L. Diabetes risk calculator. *Diabetes Care* 2008;31(5):1040-5.
18. Tabaei BP, Herman WH. A multivariate logistic regression equation to screen for diabetes: development and validation. *Diabetes Care* 2002;25(11):1999-2003.
19. Hosmer Jr DW, Lemeshow S. Applied Logistic Regression. New York, NY: John Wiley & Sons, Inc.; 2004.
20. Feng J, Xu H, Mannor S, Yan S, editors. Robust Logistic Regression and Classification. *Adv Neural Inf Process Syst*; 2014.
21. Croux C, Haesbroeck G. Implementing the Bianco and Yohai estimator for logistic regression. *Comput Stat Data Anal* 2003;44(1):273-95.
22. Carroll RJ, Pederson S. On robustness in the logistic regression model. *J Roy Stat Soc B (Methodological)* 1993:693-706.
23. Copas J. Binary regression models for contaminated data. *J Roy Stat Soc B (Methodological)* 1988;50:225-65.
24. Sarkar SK, Midi H, Rana M. Detection of outliers and influential observations in binary logistic regression: An empirical study. *J Appl Sci* 2011;11(1):26-35.

TUMJ Tehran University Medical Journal, October 2018; Vol. 76, No. 7: 452-458 Original Article

Comparison of ordinary logistic regression and robust logistic regression models in modeling of pre-diabetes risk factors

Fateme Azizi Mayvan M.Sc.[1]
Mehdi Jabbari Nooghabi Ph.D.[2]
Ali Taghipour Ph.D.[3]
Mohammad Taghi Shakeri Ph.D.[4*]
Mahsa Mokarram Ph.D. Student[5]

1- Department of Public Health, Neyshabur University of Medical Sciences, Neyshabur, Iran.
2- Department of Statistics, School of Mathematics, Ferdowsi University, Mashhad, Iran.
3- Department of Epidemiology, Social Determinants of Health Research Center, Mashhad University of Medical Sciences, Mashhad, Iran.
4- Department of Biostatistics, Social Determinants of Health Research Center, Mashhad University of Medical Sciences, Mashhad, Iran.
5- Department of Demographics, Student Research Committee, Islamic Azad University, Central Tehran Branch, Tehran, Iran.

[*] Corresponding author: School of Public Health, Daneshgah St., Mashhad, Iran.
Tel: +98 51 38515116
E-mail: shakerimt@mums.ac.ir

Abstract

Received: 09 Apr. 2018 Revised: 16 Apr. 2018 Accepted: 07 Oct. 2018 Available online: 22 Oct. 2018

Background: Regarding the increased risk of developing type 2 diabetes in pre-diabetic people, identifying pre-diabetes and determining of its risk factors seems so necessary. In this study, it is aimed to compare ordinary logistic regression and robust logistic regression models in modeling pre-diabetes risk factors.

Methods: This is a cross-sectional study and conducted on 6460 people, over 30 years old, who have participated in the screening of diabetes plan in Mashhad city that it was done by Mashhad University of Medical Sciences from October to December 2010. According to the fasting blood sugar criteria, 5414 individuals were identified as healthy and 1046 individuals were identified as pre-diabetic. Age, gender, body mass index, systolic blood pressure, diastolic blood pressure and waist-to-hip ratio were measured for every participant. The data was entered into the Microsoft Excel 2013 (Microsoft Corp., Redmond, WA, USA) and then analysis of the data was done in R Project for Statistical Computing, Version R 3.1.2 (www.r-project.org). Ordinary logistic regression model was fitted on the data. The outliers were identified. Then Mallow, WBY and BY robust logistic regression models were fitted on the data. And then, the robust models were compared with each other and with ordinary logistic regression model according to goodness of fit and prediction ability using Pearson's chi-square and area under the receiver operating characteristic (ROC) curve respectively.

Results: Among the variables that were included in the ordinary logistic regression model and three robust logistic models, age, body mass index and systolic blood pressure were statistically significant ($P< 0.01$) but waist-to-hip ratio was not statistically significant ($P> 0.1$). There were 552 outliers with misclassification error in the ordinary logistic regression model. Pearson's chi-square value and area under the ROC curve value in the Mallow model were almost the same as for ordinary logistic regression model. But it was relatively higher in BY and WBY models.

Conclusion: Based on results of this study age, overweight and hypertension are risk factors of prediabetes. Also, WBY and BY models were better than ordinary logistic regression model, according to goodness of fit criteria and prediction ability.

Keywords: body mass index, diabetes mellitus, logistic models, prediabetic state.