

اندیشه آماری، بهار و تابستان ۱۳۹۱، شماره پیاپی ۳۳

سال هفدهم شماره اول، ص ۲۹-۴۳

کاربردی از مدل‌های رگرسیون لجستیک ترتیبی دو سطحی در تعیین عوامل مؤثر بر بار اقتصادی دیابت نوع دو در ایران

مریم هادی پور^۱، راضیه جعفرآقایی، قاسم یادگارفر، آوات فیضی، فرید ابوالحسنی

چکیده:

در سال‌های اخیر، مدل‌های رگرسیونی چند سطحی به طور چشمگیری در علوم مختلف از جمله پزشکی، روانشناسی، اقتصاد و سایر علوم توسعه یافته‌اند. این مدل‌ها برای داده‌هایی با ساختار سلسله مراتبی که هر سطح پایینی در سطوح بالاتر لانه گزیده است، کاربرد دارند. برای مدل کردن این نوع داده‌ها با متغیرهای پاسخ گسسته (مانند دو یا چند حالتی، شمارشی، ترتیبی و...) از مدل‌های رگرسیونی تعمیم یافته استفاده می‌شود. در این مقاله ابتدا مدل رگرسیونی لجستیک ترتیبی دو سطحی معرفی شده و روش‌های مختلفی برای برآورد پارامترهای مدل شرح داده می‌شود. سپس کاربرد این مدل با استفاده از داده‌های مطالعه برآورد هزینه دیابت در ایران که توسط مرکز غدد درون ریز و متابولیسم و دانشگاه علوم پزشکی تهران در سال ۱۳۸۵ گردآوری شده‌اند، در تعیین عوامل مؤثر فردی و محیطی بر بار اقتصادی دیابت نوع دو بر بیماران دیابتی پرداخته می‌شود.

واژه‌های کلیدی: بار اقتصادی بیماری دیابت نوع دو، رگرسیون لجستیک ترتیبی دو سطحی، روش‌های شبه درستمایی حاشیه‌ای و پیشگویانه، روش مربع‌بندی گوسی، روش مربع‌بندی گوسی سازوار.

^۱maryam.hadipoor@yahoo.com

۱ مقدمه

(از جمله گسسته، دو حالتی، شمارشی و ترتیبی)

می‌باشند (مک کولا و سیرل، ۲۰۰۱).

در این مقاله از مدل رگرسیونی لجستیک ترتیبی دو سطحی که حالت خاصی از مدل‌های رگرسیونی چند سطحی تعمیم یافته است، برای برازش داده‌های دیابت استفاده می‌کنیم. بخش ۲ به معرفی این مدل‌ها پرداخته است. بخش ۳ روش‌های برآورد پارامترها در مدل رگرسیونی چند سطحی تعمیم یافته را شرح می‌دهد. در بخش ۴ مدل رگرسیون لجستیک ترتیبی دو سطحی را به داده‌های دیابت برازش داده و در بخش ۵ به بحث و نتیجه‌گیری خواهیم پرداخت.

مدل‌های رگرسیونی چند سطحی نقش مهمی را در پژوهش‌های مربوط به علوم پزشکی، روانشناسی، اقتصاد و سایر علوم ایفا می‌کنند. در این مدل‌ها، اثرات متغیرهای پیش‌بین سطح یک، در سطوح بالاتر تصادفی در نظر گرفته می‌شوند. به عنوان مثال برای بررسی اثر هوش بر موفقیت تحصیلی دانش‌آموزان در یک مدرسه، به دلیل قرار گرفتن دانش‌آموزان در کلاس‌های مختلف با داده‌های سلسله مراتبی روبرو هستیم. در این حالت دانش‌آموزان واحدهای سطح یک و کلاس‌ها واحدهای سطح دو را تشکیل می‌دهند. هنگامی که هدف تحقیق بررسی اثر متغیر پیش‌بین در سطح فردی بر روی متغیر پاسخ همان سطح (مانند بررسی اثر هوش دانش‌آموز بر موفقیت تحصیلی) باشد، تحلیل داده‌ها در سطح فردی کاراتر از تحلیل در سطح خوشه‌ای است (هدکر و گینز، ۱۹۹۴). اما هنگامی که بررسی اثرات متغیرهای سطح فردی (مانند جنس و هوش) در حضور متغیرهای سطوح بالاتر (مانند اندازه کلاس) و همچنین اثر متقابل آن‌ها بر متغیر وابسته مد نظر باشد، لزوم استفاده از مدل‌های چند سطحی آشکار می‌شود (هدکر و گینز، ۱۹۹۴). مدل‌های رگرسیونی چند سطحی تعمیم یافته^۲، رده‌ای از مدل‌های رگرسیونی چند سطحی برای انواع مختلفی از متغیرهای وابسته

۲ مدل رگرسیونی دو سطحی تعمیم یافته

به طور کلی مدل‌های رگرسیونی خطی تعمیم یافته شامل سه مؤلفه می‌باشند (اگرستی، ۲۰۰۷) :

(۱) مؤلفه تصادفی که نشان‌دهنده متغیر پاسخ (Y) با یک توزیع احتمال است.

(۲) مؤلفه خطی یا سیستماتیک که متغیرهای پیش‌بین در مدل را نشان می‌دهد.

(۳) تابع پیوند، تابعی از مقادیر مورد انتظار متغیر پاسخ $E(Y)$ را مشخص می‌کند که توسط مدل خطی به مقادیر متغیرهای پیش‌بین مدل مربوط می‌شود.

شکل کلی مدل‌های خطی تعمیم یافته دو سطحی

^۲Generalized Multilevel Regression Models

عبارتست از:

به عبارتی احتمال تعلق فرد شماره i -ام به رده‌های s و بالاتر متغیر پاسخ را نشان می‌دهد و بنابراین مدل

رگرسیون ترتیبی با تابع پیوند لجیت برابر است با

$$\mu_{ij} = E(Y_{ij} | \nu_j, x_{ij}) \quad (1)$$

که در آن i ($i = 1, \dots, n_j$) به واحدهای سطحیک و j ($j = 1, \dots, N$) به واحدهای سطح

دو دلالت دارند و مقدار مورد انتظار متغیر پاسخ از

طریق تابع پیوند η_{ij} با متغیرهای پیش‌بین به‌صورت

زیر مرتبط شده است.

زیر است.

$$\logit(\gamma_i^{(s)}) = \beta_2 x_i - \alpha^{(s)} + u_{0j} \quad (4) \quad \eta_{ij} = g(\mu_{ij}) = x_{ij}\beta + \nu_j \quad (2)$$

که در آن u_{0j} اثر تصادفی سطح دو در مدل می‌باشد

(مک کولا و نلدر، ۱۹۸۳).

 ν_j اثر تصادفی خوشه j -ام می‌باشد که به جمله $X_{ij}\beta$ اضافه می‌شود. این اثر تصادفی، تأثیر خوشه j -ام را روی واریانس مشاهدات درونی‌اش که توسط

متغیرهای پیش‌بین توضیح داده نشده‌اند، نشان می‌دهد.

فرض می‌شود که ν_j دارای توزیع $N(0, \sigma_\nu^2)$ باشد.

مدل رگرسیونی لجستیک ترتیبی:

۳ روش‌های برآورد پارامترها در

مدل‌های رگرسیونی چند سطحی

تعمیم یافته

چندین روش متفاوت برای مدل‌بندی پاسخ ترتیبی

وجود دارد مانند (۱) شانس متناسب، (۲) طبقه‌های

همجوار^۴ و (۳) نسبت پیوستگی^۵. در این مقاله

با توجه به اینکه مدل‌های رگرسیونی چند سطحی

تعمیم یافته بسیار متداول شده‌اند، روش‌های برآوردیابی

پارامترهای آن‌ها از اهمیت ویژه‌ای برخوردار است.

روش‌ها و برنامه‌های کامپیوتری متنوعی برای برازش

این مدل‌ها پیشنهاد شده‌اند. با این وجود نمی‌توان

روش یکتایی را برای برآورد پارامترها که در همه

حالات کارا باشد، در نظر گرفت. در این مدل‌ها

تابع درستمایی داده‌های مشاهده شده که برای برآورد

 $(1, \dots, S)$ رده باشد به‌طوری‌که

$$\gamma_i^{(s)} = P(y_i > s) \quad , \quad s = 1, \dots, S-1$$

^۳ Cumulative(proportional)odds (PO)^۴ Adjacent Categories^۵ Continuation Ratio

پارامترها به کار می‌رود، یک تابع حاشیه‌ای می‌باشد این معنا که در برآورد پارامترهای مدل بیش برآوردی که نسبت به اثرات تصادفی انتگرال‌گیری شده است می‌کنند. به منظور رفع این مشکل از روش‌های PQL که در حالت کلی شکل بسته‌ای ندارد. بنابراین و MQL که توسط مک کولا و سیرل (۲۰۰۱) از روش‌های تقریبی برای برآورد پارامترها استفاده بررسی شده‌اند، استفاده می‌شود. این روش‌ها توسط می‌شود (رب هسکت و همکاران، ۲۰۰۲ و هدرک، خطی‌سازی مدل از طریق بسط سری تیلور انجام می‌شوند و از روش‌های حداقل مربعات تعمیم‌یافته (۲۰۰۳). روش‌های تقریبی رایجی که در این زمینه به کار می‌روند عبارتند از تقریب لاپلاس، تکراری (IGLS^{۱۱}) و یا حداقل مربعات تعمیم‌یافته روش‌های بیزی، روش‌های شبه درستنمایی حاشیه‌ای و پیشگویانه (MQL^۶ و PQL^۷)، روش مربع‌بندی پارامترهای مدل استفاده می‌کنند.

گوسی (GQ^۸) و روش مربع‌بندی گوسی سازوار مرتبه PQL و MQL با در نظر گرفتن تعداد جملاتی (AGQ^۹). تقریب لاپلاس بر اساس بسط سری که در بسط تیلور تحت خطی‌سازی قرار می‌گیرند، تیلور می‌باشد و هنگامی که توزیع پسین تقریباً نرمال حاصل می‌شود. به عنوان مثال اگر سری تیلور تا و یا حجم نمونه درون خوشه‌ها بزرگ باشد، خوب مرتبه دوم بسط داده شود، برآوردهای PQL_2 و عمل می‌کند (رادنباش و همکاران، ۲۰۰۰). در MQL_2 به دست می‌آیند. این روش‌ها برای حالتی رهیافت بیزی پارامترها و اثرات تصادفی مدل، به عنوان که توزیع پاسخ به شرط اثرات تصادفی تقریباً نرمال متغیر تصادفی در نظر گرفته می‌شوند و برای نمونه‌گیری است، خوب عمل می‌کنند. همچنین برای داده‌هایی از توزیع پسین و برآورد پارامترها از روش زنجیره با پاسخ‌های دو حالتی و حجم نمونه‌ای درون خوشه‌ای مارکف مونت کارلو (McMC^{۱۰}) استفاده می‌شود بزرگ، کارا می‌باشند (مک کولا و سیرل، ۲۰۰۱).

(بران و دراپر، ۲۰۰۵). در مدل‌بندی چند سطحی اما در حالتی که حجم نمونه درون خوشه کوچک با پاسخ‌های غیریوسته، روش‌های برآورد ماکزیمم باشد، به طور ضعیف عمل می‌کنند (مک کولا و درستنمایی به طور محاسباتی افراطی عمل می‌کنند، به سیرل، ۲۰۰۱ و رد ریگز و گلدمن، ۲۰۰۱). روش‌های PQL و MQL در نرم‌افزارهای MLwiN (راسباش و همکاران، ۲۰۰۴) و HLM (رادنباش و همکاران، ۲۰۰۵) قابل دسترس هستند.

^۶Marginal Quasi likelihood

^۷predictive Quasi likelihood

^۸Gaussian Quadrature

^۹Adaptive Gaussian Quadrature

^{۱۰}Monte Carlo Markov Chain

^{۱۱}Iterative Generalized Least Square

^{۱۲}Restricted Iterative Generalized Least Square

روش GQ توسط گینز و هدر در سال ۱۹۹۷ برای و روش های GQ و AGQ در نرم افزار STATA مدل های خطی تعمیم یافته به کار برده شد. این روش به تعداد زیادی نقاط مربع بندی برای تقریب تابع درست نمایی نیاز دارد و با مشکلاتی همراه است که برای پاسخ های دو حالتی توسط لیسافر و اسپینز (۲۰۰۱) بررسی شده اند. راندنباش و بریک (۲۰۰۲) با استفاده از مطالعات شبیه سازی نشان دادند که GQ بهتر از PQL تقریب می زند. روش GQ در مواردی بویژه برای داده هایی با پاسخ برنولی که

حجم نمونه درون خوشه کم و همبستگی درون خوشه ای زیاد است، به طور ضعیف عمل می کند (لیسافر و اسپینز، ۲۰۰۱). برای غلبه بر مشکلات GQ، از روش AGQ که به نقاط مربع بندی کمتری نیاز دارد، استفاده می شود (رب هسکت و همکاران، ۲۰۰۲). رب هسکت و همکاران (۲۰۰۵) برای مقایسه عملکرد دو روش GQ و AGQ، از مدل های دو سطحی با پاسخ های دو حالتی و حجم های نمونه ای و همبستگی های درون خوشه ای متفاوتی استفاده کردند و نشان دادند در حالت هایی که حجم نمونه درون خوشه بزرگ و همبستگی درون خوشه ای زیاد باشد، AGQ بهتر از GQ عمل می کند. روش های AGQ و GQ توسط دستور gllamm در STATA قابل اجرا هستند (رب هسکت و اسکروندال، ۲۰۰۸ و اسکروندال و رب هسکت، ۲۰۰۳). در این مقاله از روش های PQL و MQL در نرم افزار MLwiN

داده های مورد استفاده در طرح حاضر از مطالعه کشوری « برآورد هزینه دیابت در ایران در سال ۸۵ » می باشد که به صورت مقطعی و توسط مرکز تحقیقات غدد درون ریز و متابولیسم دانشگاه علوم پزشکی تهران و معاونت سلامت وزارت بهداشت در ۲۷ استان بر روی بیماران دیابتی تهیه شده است. این مطالعه برای اولین بار و با روش نمونه گیری تصادفی خوشه ای چند مرحله ای در ایران انجام شده است. هدف از برآزش مدل توضیحی تعیین عوامل مؤثر فردی و محیطی بر بار اقتصادی دیابت نوع دو بر بیماران دیابتی می باشد. متغیر پاسخ (بار اقتصادی تحمیل شده بر بیمار) از تلفیق دو متغیر دو حالتی کاهش و جذب درآمد ماهیانه بیمار (کمتر از ۳۰ درصد، ۳۰ درصد و بیشتر) به دلیل مشکلات مرتبط با دیابت تعیین گردید. لازم به ذکر است که جذب درآمد ماهیانه بیمار به دلیل هزینه های صرف شده مربوط به بیماری دیابت تعریف می شود که

۴ برازش مدل رگرسیون

لجستیک ترتیبی دو سطحی

به داده های دیابت

این متغیر به صورت درصد اندازه‌گیری شده است. جهانی است و توسط چند پرسش از فرد بیمار به دست بر این اساس، متغیر پاسخ ترتیبی (سه رسته‌ای) می‌آید.

به دست آمد. رسته‌های متغیر پاسخ ۱ - دو متغیر

کاهش و جذب درآمد با پاسخ مقادیر کمتر از 30%

به عنوان بار اقتصادی نرمال ۲ - مقادیر بالای 30%

به عنوان بار اقتصادی کمرشکن و ۳ - در صورتی که

یکی از متغیرها بالای 30% و دیگری پایین 30%

باشد به عنوان بار اقتصادی متوسط در نظر گرفته شد

۱۳. تعداد افراد مورد مطالعه ۳۹۱۶ بود که مورد

مصاحبه رو در رو قرار گرفتند. از این تعداد ۳۲۳۴

پرسشنامه به طور کامل تکمیل گردید و مورد تحلیل قرار گرفتند. پایایی پرسشنامه توسط یک نمونه

مقدماتی ۷۰ نفری به تأیید رسیده است. متغیرهای

مورد مطالعه در سطح اول، متغیرهای وضعیت سلامتی،

وضعیت اقتصادی، محل سکونت، جنسیت، وضعیت

بیمه بودن یا نبودن بیمار، سن و تعداد سال‌های

آگاهی از ابتلا به بیماری در نظر گرفته شدند. متغیرهای

تعداد پزشک متخصص در استان (یا مرکز استان)،

تعداد کلینیک درمانی، نسبت شهرنشینی، نرخ باسوادی

و نرخ بیکاری به عنوان متغیرهای سطح دوم مورد

مطالعه قرار گرفتند. لازم به ذکر است که متغیر

وضعیت سلامتی بیمار، از طریق شاخص HUI^{۱۴}

اندازه‌گیری می‌شود. این شاخص، شاخص استاندارد

^{۱۳}Health System Performance Assessment,

Christopher Murray, David Evans, WHO 2003

^{۱۴}Health Utility Index

۵ یافته‌ها

در این بخش، نتایج مطالعه حاضر، در دو قسمت

توصیفی و تحلیلی ارائه می‌شوند.

۱۰۵ توصیف داده‌ها

نتایج این قسمت در قالب جداولی در انتهای مقاله

آورده شده است.

۲۰۵ تحلیل داده‌ها

به منظور برآزش مدل به داده‌های دیابت، ابتدا مدل

رگرسیونی لجستیک ترتیبی دو سطحی با عرض از

مبدأ تصادفی و بدون در نظر گرفتن متغیرهای پیش‌بین

در سطح دو و سپس با در نظر گرفتن متغیرهای

پیش‌بین در سطح دو مدل از طریق روش AGQ

(در نرم افزار STATA) به برآورد پارامترهای مدل

پرداخته شده است. نتایج در جداول ۳ و ۴ نشان

داده شده‌اند. در این قسمت برای اثبات لزوم استفاده

از مدل دو سطحی می‌توان از معیار ضریب تفکیک

واریانس (VPC^{۱۵}) که نشان‌دهنده میزان وابستگی

داده‌های درون خوشه‌ای است استفاده کرد (گلدشتین

^{۱۵}Variance Partition Coefficient

و همکاران، ۲۰۰۲). فرمول این ضریب برای مدل‌های رگرسیونی لجستیک، به صورت زیر می‌باشد

$$VPC = \frac{\sigma_{0j}^2}{\sigma_{0j}^2 + \frac{\pi^2}{3}}$$

با توجه به فرمول بالا، مقدار ضریب تفکیک واریانس برابر ۱۹/۰ به دست می‌آید که بیان کننده میزان بالای همبستگی افراد درون یک استان است. لازم به ذکر است که متغیرهای پیش‌بین وضعیت سلامتی و تعداد سال‌های آگاهی از بیماری در سطح یک همبستگی معنی‌دار داشته ($P\text{-value} < 0/01, r = 0/17$) و بنابراین برای بوجود نیامدن مشکل هم‌خطی، متغیر پیش‌بین تعداد سال‌های آگاهی از بیماری از مدل حذف شده است. همان گونه که از جدول ۳ مشخص است در مدل‌بندی متغیر پاسخ دو معادله حاصل می‌شود، اولین معادله مربوط به رده اول متغیر پاسخ یعنی رده بار نرمال اقتصادی و دومین معادله مربوط به دورده بار نرمال و متوسط اقتصادی می‌باشد. همچنین قابل ذکر است که برای مدل‌بندی متغیر پاسخ ترتیبی، از تابع پیوند لجیت با فرض شانس متناسب استفاده کردیم. در این فرض اثر (یا شانس) تک تک متغیرهای پیش‌بین بر متغیر پاسخ، در همه رده‌ها یکسان در نظر گرفته می‌شود.

تأثیرگذار باشد. در مثال بیان شده، اگر چه هر دو مدل آماری اثرات معنی‌دار مشابهی را از متغیرهای پیش‌بین نشان دادند، ولی این مدل‌ها منجر به نتایج متفاوتی در رابطه با اندازه اثرات و اندازه واریانس اثر تصادفی شدند. در حالت رگرسیون لجستیک ترتیبی دو سطحی بدون متغیرهای پیش‌بین سطح دو، واریانس اثر تصادفی یعنی واریانس u_j از این مقدار در حالت رگرسیون لجستیک ترتیبی دو سطحی با اضافه شدن متغیرهای پیش‌بین سطح دو در مدل بیشتر شده که نشان می‌دهد اضافه کردن متغیرهای پیش‌بین در سطح دو واریانس بخشی از این واریانس تبیین نشده در سطح دو را توضیح می‌دهد که منجر به کاهش واریانس سطح دو مدل می‌شود. در جداول ۴ و ۵ برای مدل‌های برازش داده شده دو سطحی، مشاهده می‌شود که $\sigma_{\nu_0}^2, \sigma_{\nu_1}^2 \neq 0$ و بنابراین مقدار وابستگی بین مشاهدات به صورت آماری معنادار است. این مقدار وابستگی در مشاهدات را می‌توان با معیار ضریب تفکیک واریانس (VPC^{16}) نشان داد به‌طوریکه هر چه این مقدار بیشتر باشد، شدت وابستگی مشاهدات درون یک خوشه بیشتر است و به عبارتی لزوم استفاده از مدل‌های رگرسیونی چند سطحی را بیان می‌دارد (گلدشتین و همکاران، ۲۰۰۲).

مقدار VPC با توجه به فرمول بخش ۵.۲ برای مدل

۶ بحث و نتیجه‌گیری

در این مقاله نشان دادیم که انتخاب مدل‌های آماری می‌تواند بر نتایج به دست آمده از یک مجموعه داده

¹⁶ Variance Partition Coefficient

(۱) برابر ۲۱/۰ و برای مدل (۲) برابر ۱۹/۰ به زنان شانس بیشتری برای تحمل کردن بار کمرشکن می‌باشد. مقدار به‌دست آمده از مدل (۱) نشان و متوسط اقتصادی بیماری دیابت نسبت به بار نرمال دهنده این است که خوشه‌بندی بیماران درون استان‌ها دارند (۱۹/۳ = نسبت شانس). همچنین هر چقدر حدود ۲۰ درصد از پراکندگی موجود در مشاهدات وضعیت اقتصادی بیمار بدتر باشد، شانس بیشتری را که توسط متغیرهای پیش‌بین موجود در مدل برای تحمیل بار کمرشکن اقتصادی بیماری دیابت توضیح داده نشده است، تعیین می‌کند. مقدار VPC را داراست. به‌عنوان مثال شانس تعلق بیماران به به‌دست آمده از مدل (۲) نشان می‌دهد در حالتیکه رده بار کمرشکن نسبت به نرمال و متوسط برای متغیرهای پیش‌بین به مدل اضافه شده است، خوشه شدن بیماران درون استان‌ها ۱۹ درصد از تغییرپذیری یافته‌ها از جدول ۴ نشان می‌دهد با افزایش شاخص داده‌ها (یعنی واریانس تبیین نشده) را توضیح می‌دهد. شانس تعلق بیمار به رده بار کمرشکن سلامتی، بیمار کمتر می‌شود (به عبارتی نسبت شانس ۲۰ درصد کاهش می‌یابد) و بنابراین بیماران در وضع (ژاکوینز و همکاران، ۱۹۸۹). لازم به ذکر است که VPC به عوامل گوناگونی مانند حجم نمونه در خوشه‌های مختلف، افزایش و یا کاهش متغیرهای پیش‌بین و نوع متغیر پاسخ بستگی دارد (هدکر و گینز، ۱۹۹۴). در مدل (۲) با آوردن اثر سطح دو (یعنی استان‌های مختلف) در مدل رگرسیونی، برای تحمیل بار کمرشکن اقتصادی بر بیمار می‌باشد. ملاحظه می‌شود که اثر متغیرهای پیش‌بین سطح یک مانند جنسیت، محل سکونت، وضعیت سلامتی، وضعیت بیمه بودن و سن بر بار اقتصادی بیماری دیابت نوع دو تحمیل بر بیمار معنی‌دار است ($P -$ value $< 0/05$). همان گونه که از جدول ۴ مشاهده می‌شود ضریب متغیرهای جنسیت و وضعیت اقتصادی مثبت می‌باشد که نشان می‌دهد مردان نسبت

تعداد پزشک متخصص در سطح استان، شانس تعلق صفر می‌باشد و آماره آزمون دارای توزیع خی دو با بیمار به رده بار کمرشکن بیماری افزایش می‌یابد. درجه آزادی P است. مقدار (۷۵/۹) به دست آمده این مقدار حاکی از آن است که به دلیل کمبود از آزمون نسبت درستیابی نشان می‌دهد که مدل تخصیص بودجه به هزینه‌های درمانی بیماران توسط حاصل از جدول ۵ با دو پارامتر اضافی در مقایسه با سیستم ارائه خدمات بهداشتی کشور، مراجعه بیماران مدل جدول ۴، معنی‌دار است و بنابراین پارامترهای به پزشکان متخصص بیشتر می‌شود و باعث تحمیل اضافه شده به مدل غیر صفر می‌باشند ($P\text{-value} < 0/05$). همچنین واریانس اثر تصادفی در سطح هر چه بیشتر بار کمرشکن اقتصادی بیماری بر بیمار دوی مدل، غیر صفر و معنی‌دار است ($P\text{-value} < 0/05$) که لزوم استفاده از مدل دو سطحی را نشان می‌دهد. مدیریت بهتر بر این موضوع را نشان می‌دهد. نکته قابل ذکر این که هنگام برآورد پارامترها در مدل‌های

رگرسیون چند سطحی معمولاً فرض می‌شود توزیع محدودیت‌های تحقیق: یکی از نقاط ضعف اثر تصادفی (u_j) نرمال است در صورتیکه در مقاله حاضر پس از برآورد اثر تصادفی با استفاده از روش بیز تجربی ملاحظه شد که توزیع اثر تصادفی غیر نرمال و چوله به راست می‌باشد. برای حل این مشکل روش‌های نوین برآوردیابی بیضوی^{۱۷} ابداع شده‌اند که تا کنون به صورت جامع عمومیت نیافته‌اند. برای مقایسه دو مدل مفروض از آزمون نسبت درستیابی (LRT^{۱۸}) استفاده می‌کنیم که به صورت زیر تعریف می‌شود

$$LRT = -2 \log \frac{L(\hat{\theta}_0)}{L(\theta)}$$

که در آن P نشان‌دهنده تعداد پارامترها، $\hat{\theta}$ برآورد پارامترهای مدل و $\hat{\theta}_0$ برآورد پارامترها تحت فرض بالای نمونه، تعداد زیاد متغیرهای پیش‌بین در سطح ۱ و ۲ و به کارگیری روش‌های دقیق مانند استفاده از پرسشنامه استاندارد HUI برای تعیین میزان سلامتی بیمار اشاره کرد.

^{۱۷}Elliptical Estimation Methods^{۱۸}Likelihood Ratio Test

جدول ۱: میانگین، انحراف معیار، دامنه تغییرات و میانه مشخصات افراد مورد مطالعه.

متغیرهای سطح یک	میانگین	انحراف استاندارد	$Max\ Min$	میانه
سن (سال)	59/27	11/79	100 13	60
مدت آگاهی از ابتلا به دیابت (سال)	8/11	6/82	50 1	6
شاخص سلامتی (HUI)	0/62	0/39	1 - 0.495	0/72

جدول ۲: توزیع فراوانی بیماران بر حسب سطوح مختلف متغیرهای پیش‌بین سطح یک.

متغیرهای سطح یک	محل سکونت: روستایی شهری	جنسیت: زن مرد	وضعیت بیمه: بله خیر	وضعیت اقتصادی: بد متوسط خوب
فراوانی	2508 726	1249 1985	440 2794	734 1870 630
(%)	(%78) (%22)	(%38/6) (%61/4)	(%13/6) (%86/4)	(%19/5) (%57/8) (%22/7)

جدول ۳: میانگین، انحراف معیار، دامنه تغییرات و میانه متغیرهای پیش‌بین در سطح استان.

متغیرهای سطح دو	میانگین	انحراف استاندارد	$Max\ Min$	میانه
نسبت شهرنشینی (درصد)	67/61	14/92	93.91 47.11	61/17
نرخ باسوادی (درصد)	84/86	3/79	91.27 68.01	85/01
نرخ بیکاری (درصد)	9/94	2/73	18.90 6.70	9/5
تعداد پزشک متخصص	63/37	66/23	209 4	36
تعداد مراکز درمانی	43/85	43/57	145 6	23

جدول ۴: نتایج برازش مدل رگرسیون لجستیک ترتیبی دو سطحی بدون وارد کردن متغیرهای سطح دو (مدل ۱).

متغیرهای پیش‌بین	$\text{logit}(\gamma_{ij})$ (فاصله اطمینان ۹۵٪ برای OR)	p-مقدار
اثرات ثابت در سطح یک جنسیت (مرد به زن)	1/18 (2/70, 3/90) (0/09)	> 0/05
محل سکونت (شهر به روستا)	- 0/2 (0/67, 0/99) (0/10)	> 0/05
سن	- 0/01 (0/92, 1/07) * (0/04)	> 0/05
وضعیت سلامتی	- 1/7 (0/15, 0/23) (0/11)	> 0/05
وضعیت اقتصادی (بد به خوب)	0/83 (1/74, 3/00) (0/14)	> 0/05
وضعیت اقتصادی (متوسط به خوب)	0/3 (1/06, 1/70) (0/12)	> 0/05
عرض از مبدا	$\text{logit}(\gamma_{ij}^{(1)})$ 0/36 (0/80, 2/60) * (0/30) $\text{logit}(\gamma_{ij}^{(2)})$ 0/36 (4/30, 13/50) * (0/30)	> 0/05
مولفه واریانس σ_{0j}^2	0/87 (0/11)	> 0/05
آماره لگاریتم درستمایی	-1991/5992	

اعداد داخل پرانتز خطای استاندارد (SE) را نشان می‌دهند.

p مقدار مربوط به نسبت شانسیها می‌باشد.

* در سطح ۵/۰ غیر معنی‌دار

جدول ۵: نتایج برازش مدل رگرسیون لجستیک ترتیبی دو سطحی با وارد کردن متغیرهای سطح دو (مدل ۲).

متغیرهای پیش‌بین	$logit(\gamma_{ij})$ (فاصله اطمینان ۹۵٪ برای OR)	p-مقدار
اثرات ثابت در سطح یک جنسیت (مرد به زن)	1/16 (2/67, 3/78) (0/09)	> 0/05
محل سکونت (شهر به روستا)	- 0/2 (0/67, 0/99) (0/10)	> 0/05
وضعیت سلامتی	- 1/65 (0/15, 0/23) (0/11)	> 0/05
وضعیت اقتصادی (بد به خوب)	0/9 (1/87, 3/23) (0/14)	> 0/05
وضعیت اقتصادی (متوسط به خوب)	0/3 (1/06, 1/70) (0/11)	> 0/05
عرض از مبدا	$logit(\gamma_{ij}^{(1)})$ - 1/3 (0/13, 0/60) * (0/38) $logit(\gamma_{ij}^{(2)})$ 0/36 (0/69, 2/90) * (0/37)	
اثرات ثابت در سطح دو نسبت شهر نشینی	- 0/04 (0/95, 0/97) * (0/005)	
تعداد پزشک متخصص	0/009 (1/007, 1/01) (0/001)	> 0/05
اثرات تصادفی σ_{1j}^2	0/78 (0/12)	> 0/05
آماره لگاریتم درستنمایی آزمون نسبت درستنمایی	-1986/72 9/75	

اعداد داخل پرانتز خطای استاندارد (SE) را نشان می‌دهند.

p مقدار مربوط به نسبت شانسیها می‌باشد.

* در سطح ۵/۰ غیر معنی‌دار

مراجع

- [1] Agresti, A. (2007). *An introduction to categorical data analysis*. 3rd ed. Wiley series.
- [2] Browne, W.J. and Draper, D. (2005). A comparison of Bayesian and likelihood methods for fitting multilevel models. *Submitted*. Downloadable from <http://multilevel.ioe.ac.uk/team/materials/wbrssa>.
- [3] Burnham, K.P. and Anderson, D.R. (1998). *Model selection and multimodel inference*. 2nd ed. Spring, New York.
- [4] Gibbons, R.D. and Hedeker, D., (1997). Random Effects Probit and Logistic Regression Models for Three Level Data. *Biometrics*, **53**, 1527-1537.
- [5] Goldstein, H., Rasbash, J. and Browne, W.J. (2002). Partitioning Variation in Multilevel Models. *Understanding Statistics*.
- [6] Hedeker, D., Gibbons, R.D. and Flay, B.R. (1994). Random Effects Regression Models for clustered data with an example from smoking prevention research. *Journal of Consulting and Clinical Psychology*, **62**, 757-765.
- [7] Hedeker, D. (2003). A mixed effects multinomial logistic regression model. *Statistics in Medicine*, **22**, 1433-1446.
- [8] Hedeker, D. (2005). *Generalized Linear Mixed Models*. Encyclopedia of Statistics in Behavioral Science. Wiley, New York.
- [9] Jacobs, D.R., Jeffery, R.W. and Hannan, P.J. (1989). Methodological issues in worksite health intervention research: II, computation of variance in worksite data. In Johnson, K., LaRosa, J.H., Scheirer, and Wolle, J.M. (Eds). *Proceedings of the*

1988 methodological issues in worksite research conference (pp. 77-88). Airlie, VA: US. Department of Health and Human Services.

- [10] Litte, R.C., Milliken, G.A., Stroup, W.A., Wolfinger, R.D. and Schabenberger, O. (2006). *SAS for Mixed Models*, Second Edition. Cary, NC: SAS Institute.
- [11] Lesaffre, E. and Spiessens, B. (2001). On the effect of the number of quadrature points in a logistic random effects model: an example. *Applied Statistics*, **50**, 325-335.
- [12] McCullach, P. and Nelder, J.A. (1983). *Generalized Linear Models*. London: Chapman and Hall.
- [13] McCulloch, C.E. and Searle, S.R. (2001). *Generalized Linear and Mixed Models*. Wiley, New York.
- [14] Rodrigues, G. and Goldman, N. (2001). Improved estimation procedures for multilevel models with binary response: A case study. *Journal of the Royal Statistical Society*, **164**, 339-355.
- [15] Rabe-Hesketh, S., Skrondal, A. and Pickles, A. (2002). Reliable estimation of generalized linear mixed models using adaptive quadrature. *The Stata Journal*, **2**, 1-21.
- [16] Rabe-Hesketh, S., Skrondal, A. (2008). *Multilevel and Longitudinal Modeling Using Stata*. 2nd ed. College Station, Tx: Stata Press.
- [17] Rasbash, J. Steele, F. Browne, W. and Prosser, B. (2004). A User Guide to MLwiN Version 2.0. London: Institute of Education. Downloadable from <http://multilevel.ioe.ac.uk/download/manuals.html>.

- [18] Raudenbush, S.W., Yang, M.L. and Yosef, M. (2000). Maximum likelihood for generalized linear models with nested random effects via high order multivariate Laplace approximation. *Jornal of Computational and Graphical Statistics*, **9**, 141-157.
- [19] Raudenbush, S.W. and Bryk, A.S. (2002). *Hierarchical Linear Models: Application and Data Analysis Methods*, Second Edition. Newbury Park, CA: Sage.
- [20] Raudenbush, S., Bryk, A., Cheog, Y.F. and Congdon, R. (2005). HLM 6: Hierarchical Linear and Nonlinear Modeling Lincolnwood, *Scientific Software International*.
- [21] Spiegelhalter, D.J., Best, N.G., Carlin, B.P. and Van der Linde, A. (2002). Bayesian measures of model complexity and fit (whit discution). *J. Roy. Statist. Soc. B*, **64**, 583-640.
- [22] Skrondal, A. and Rabe- Hesketh, S. (2003). Multilevel logistic Regression for polytomous data and rankings *Psychometrika*, **68**, 267-287.