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 تعمیم‌یافته 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 درستنمایی حاشیه‌ای و پیشگویانه، روش مربع‌بندی گوسی، روش مربع‌بندی گوسی سازوار.

[۱] maryam_hadipoor@yahoo.com

۱ مقدمه

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. هنگامی‌که 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 گیبنز، 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 گیبنز، ۱۹۹۴). مدل‌های رگرسیونی چند سطحی تعمیم‌یافته [۲] ، رده‌ای از مدل‌های رگرسیونی چند سطحی برای انواع مختلفی از متغیرهای وابسته (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 تعمیم‌یافته 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 تعمیم‌یافته 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.

۲ مدل رگرسیونی دو سطحی تعمیم‌یافته

به طور کلی مدل‌های رگرسیونی خطی تعمیم‌یافته 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 تعمیم‌یافته دو سطحی

[۲] Generalized Multilevel Regression Models

عبارتست از:

$$\mu_{ij} = E(Y_{ij} \mid \nu_j, x_{ij}) \qquad (۱)$$

که در آن i $(i = 1, \ldots, n_j)$ به واحدهای سطح یک و j $(j = 1, \ldots, N)$ 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.

$$\eta_{ij} = g(\mu_{ij}) = x_{ij}\beta + \nu_j \qquad (۲)$$

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:

چندین روش متفاوت برای مدل‌بندی پاسخ ترتیبی وجود دارد مانند ۱) شانس متناسب، [۳] ۲) طبقه‌های همجوار [۴] و ۳) نسبت پیوستگی [۵] . در این مقاله از روش شانس متناسب استفاده می‌کنیم. فرض کنید متغیر پاسخ یک متغیر ترتیبی با S $(s = 1, \ldots, S)$ رده باشد به‌طوری‌که

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

به عبارتی احتمال تعلق فرد شماره i - ام به رده‌های s و بالاتر متغیر پاسخ رانشان می‌دهد و بنابراین مدل رگرسیون ترتیبی با تابع پیوند لجیت برابر است با

$$\begin{aligned} logit(\gamma_i^{(s)}) &= ln\frac{\gamma_i^{(s)}}{1-\gamma_i^{(s)}} \\ &= \beta_2 x_i - \alpha^{(s)} \end{aligned} \qquad (۳)$$

مدل رگرسیون لجستیک ترتیبی دو سطحی به صورت زیر است.

$$logit(\gamma_i^{(s)}) = \beta_2 x_i - \alpha^{(s)} + u_{0j} \qquad (۴)$$

که در آن u_{oj} اثر تصادفی سطح دو در مدل می‌باشد (مک کولا و نلدر، ۱۹۸۳).

۳ روش‌های برآورد پارامترها در مدل‌های رگرسیونی چند سطحی تعمیم‌یافته

با توجه به اینکه مدل‌های رگرسیونی چند سطحی تعمیم‌یافته 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 درستنمایی داده‌های مشاهده شده که برای برآورد

[۳] Camulative(proportional)odds (PO)

[۴] Adjacent Categories

[۵] Continuation Ratio

پارامترها به کار می‌رود، یک تابع حاشیه‌ای می‌باشد که نسبت به اثرات تصادفی انتگرال‌گیری شده است که در حالت کلی شکل بسته‌ای ندارد. بنابراین از روش‌های تقریبی برای برآورد پارامترها استفاده می‌شود (رب هسکت و همکاران، ۲۰۰۲ و هدکر، ۲۰۰۳). روش‌های تقریبی رایجی که در این زمینه به کار می‌روند عبارتند از تقریب لاپلاس، روش‌های بیزی، روش‌های شبه درستنمایی حاشیه‌ای و پیشگویانه (MQL [۶] و PQL [۷])، روش مربع‌بندی گوسی (GQ [۸]) و روش مربع‌بندی گوسی سازوار (AGQ [۹]). تقریب لاپلاس بر اساس بسط سری تیلور می‌باشد و هنگامی‌که توزیع پسین تقریباً نرمال و یا حجم نمونه درون خوشه‌ها بزرگ باشد، خوب عمل می‌کند (رادنباش و همکاران، ۲۰۰۰). در رهیافت بیزی پارامترها و اثرات تصادفی مدل، به‌عنوان متغیر تصادفی در نظر گرفته می‌شوند و برای نمونه‌گیری از توزیع پسین و برآورد پارامترها از روش زنجیره مارکف مونت کارلو (McMC [۱۰]) استفاده می‌شود (بران و دراپر، ۲۰۰۵). در مدل‌بندی چند سطحی با پاسخ‌های غیرپیوسته، روش‌های برآورد ماکزیمم درستنمایی به‌طور محاسباتی افراطی عمل می‌کنند، بسیرل، ۲۰۰۱ و ردریگز و گلدمن، ۲۰۰۱). روش‌های PQL و MQL در نرم‌افزارهای MLwiN (راسباش و همکاران، ۲۰۰۴) و HLM (رادنباش و همکاران، ۲۰۰۵) قابل دسترس هستند.

این معنا که در برآورد پارامترهای مدل بیش برآوردی می‌کنند. به‌منظور رفع این مشکل از روش‌های PQL و MQL که توسط مک کولا و سیرل (۲۰۰۱) بررسی شده‌اند، استفاده می‌شود. این روش‌ها توسط خطی‌سازی مدل از طریق بسط سری تیلور انجام می‌شوند و از روش‌های حداقل مربعات تعمیم‌یافته تکراری (IGLS [۱۱]) و یا حداقل مربعات تعمیم‌یافته وزن‌دار شده تکراری (RIGLS [۱۲]) برای برآورد پارامترهای مدل استفاده می‌کنند.

مرتبه PQL و MQL با در نظر گرفتن تعداد جملاتی که در بسط تیلور تحت خطی‌سازی قرار می‌گیرند، حاصل می‌شود. به‌عنوان مثال اگر سری تیلور تا مرتبه دوم بسط داده شود، برآوردهای PQL_2 و MQL_2 به‌دست می‌آیند. این روش‌ها برای حالتی که توزیع پاسخ به شرط اثرات تصادفی تقریباً نرمال است، خوب عمل می‌کنند. همچنین برای داده‌هایی با پاسخ‌های دو حالتی و حجم نمونه‌ای درون خوشه‌ای بزرگ، کارا می‌باشند (مک کولا و سیرل، ۲۰۰۱). اما در حالتی‌که حجم نمونه درون خوشه کوچک باشد، به‌طور ضعیف عمل می‌کنند (مک کولا و

---

[۶] Marginal Quasi likelihood

[۷] predictive Quasi likelihood

[۸] Guassian Quadrature

[۹] Adaptive Guassian Quadrature

[۱۰] Monte Carlo Markov Chain

[۱۱] Iterative Generalized Least Square

[۱۲] Restricted Iterative Generalized Least Square

روش GQ توسط گیبنز و هدکر در سال ۱۹۹۷ برای مدل‌های خطی تعمیم‌یافته به‌کار برده شد. این روش به تعداد زیادی نقاط مربع‌بندی برای تقریب تابع درستنمایی نیاز دارد و با مشکلاتی همراه است که برای پاسخ‌های دو حالتی توسط لیسافر و اسپیسنز (۲۰۰۱) بررسی شده‌اند. رادنباش و بریک (۲۰۰۲) با استفاده از مطالعات شبیه‌سازی نشان دادند که GQ بهتر از PQL تقریب می‌زند. روش GQ در مواردی بویژه برای داده‌هایی با پاسخ برنولی که حجم نمونه درون خوشه کم و همبستگی درون خوشه‌ای زیاد است، به طور ضعیف عمل می‌کند (لیسافر و اسپیسنز، ۲۰۰۱). برای غلبه بر مشکلات GQ، از روش AGQ که به نقاط مربع‌بندی کمتری نیاز دارد، استفاده می‌شود (رب هسکت و همکاران، ۲۰۰۲). رب هسکت و همکاران (۲۰۰۵) برای مقایسه عملکرد دو روش GQ و AGQ، از مدل‌های دو سطحی باپاسخ‌های دو حالتی و حجم‌های نمونه‌ای و همبستگی‌های درون خوشه‌ای متفاوتی استفاده کردند و نشان دادند در حالت‌هایی که حجم نمونه درون خوشه بزرگ و همبستگی درون خوشه‌ای زیاد باشد، AGQ بهتر از GQ عمل می‌کند. روش‌های 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 روش‌های PQL و MQL در نرم افزار MLwiN و روش‌های GQ و AGQ در نرم افزار STATA برای برآورد پارامترها استفاده شد و سپس نتایج مذکور مقایسه گردید.

۴ برازش مدل رگرسیون لجستیک ترتیبی دو سطحی به داده‌های دیابت

داده‌های مورد استفاده در طرح حاضر از مطالعه کشوری « برآورد هزینه دیابت در ایران در سال ۸۵ » می‌باشد که به‌صورت مقطعی و توسط مرکز تحقیقات غدد درون‌ریز 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 نمونه‌گیری تصادفی خوشه‌ای چند مرحله‌ای در ایران انجام شده است. هدف از برازش مدل توضیحی تعیین عوامل مؤثر فردی و محیطی بر بار اقتصادی دیابت نوع دو بر بیماران دیابتی می‌باشد. 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 هزینه‌های صرف شده مربوط به بیماری دیابت تعریف می‌شود که

این متغیر به صورت درصد اندازه‌گیری شده است. بر این اساس، متغیر پاسخ ترتیبی (سه رسته‌ای) به‌دست آمد. رسته‌های متغیر پاسخ ۱- دو متغیر کاهش و جذب درآمد با پاسخ مقادیر کمتر از 30% به‌عنوان بار اقتصادی نرمال ۲- مقادیر بالای 30% به‌عنوان بار اقتصادی کمرشکن و ۳- در صورتی که یکی از متغیرها بالای 30% و دیگری پایین 30% باشد به‌عنوان 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 به‌عنوان 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 جهانی است و توسط چند پرسش از فرد بیمار به‌دست می‌آید.

۵ یافته‌ها

در این بخش، نتایج مطالعه حاضر، در دو قسمت توصیفی و تحلیلی ارائه می‌شوند.

۱.۵ توصیف داده‌ها

نتایج این قسمت در قالب جداولی در انتهای مقاله آورده شده‌است.

۲.۵ تحلیل داده‌ها

به منظور برازش 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

[۱۳] Health System Performance Assessmant, Christopher Murray , David Evans, WHO 2003

[۱۴] Health Utility Index

[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 به‌دست 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 ($P-value < 0/01, r = 0/17$) 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 دو رده 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.

۶ بحث و نتیجه‌گیری

در این مقاله نشان دادیم که انتخاب مدل‌های آماری می‌تواند بر نتایج به‌دست آمده از یک مجموعه داده 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[۱۶]) 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

[۱۶] Variance Partition Coefficient

(۱) برابر ۲۱/۰ و برای مدل (۲) برابر ۱۹/۰ می‌باشد. مقدار به‌دست آمده از مدل (۱) نشان دهنده این است که خوشه‌بندی بیماران درون استان‌ها حدود ۲۰ درصد از پراکندگی موجود در مشاهدات را که توسط متغیرهای پیش‌بین موجود در مدل توضیح داده نشده است، تعیین می‌کند. مقدار VPC به‌دست آمده از مدل (۲) نشان می‌دهد در حالتیکه متغیرهای پیش‌بین به مدل اضافه شده است، خوشه شدن بیماران درون استان‌ها ۱۹ درصد از تغییرپذیری داده‌ها(یعنی واریانس تبیین نشده) را توضیح می‌دهد. مجموعه داده حاضر، مقدار نسبتاً بالا و منطبق با مقادیر مقبول گزارش شده از مقالات قبلی می‌باشد (ژاکوبز 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 گیبنز، ۱۹۹۴). در مدل (۲) با آوردن اثر سطح دو (یعنی استان‌های مختلف) در مدل رگرسیونی، 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 به زنان شانس بیشتری برای تحمل کردن بار کمرشکن و متوسط اقتصادی بیماری دیابت نسبت به بار نرمال دارند(۱۹/۳ = نسبت شانس). همچنین هر چقدر وضعیت اقتصادی بیمار بدتر باشد، شانس بیشتری برای تحمیل بار کمرشکن اقتصادی بیماری دیابت را داراست. به‌عنوان مثال شانس تعلق بیماران به رده بار کمرشکن نسبت به نرمال و متوسط برای بیماران فقیر نسبت به غنی برابر ۴۶/۲ می‌باشد. یافته‌ها از جدول ۴ نشان می‌دهد با افزایش شاخص سلامتی، شانس تعلق بیمار به رده بار کمرشکن بیمار کمتر می‌شود (به عبارتی نسبت شانس ۲۰ درصد کاهش می‌یابد) و بنابراین بیماران در وضع سلامتی بهتر، شانس بیشتری برای تعلق داشتن به رده نرمال و متوسط اقتصادی بیماری دیابت دارند. همچنین شانس تحمیل بار کمرشکن بیماری برای بیماران ساکن شهر نسبت به روستا برابر ۸/۰ می‌باشد که نشان می‌دهد سکونت در روستا عامل خطری برای تحمیل بار کمرشکن اقتصادی بر بیمار می‌باشد. برآورد منفی ضریب متغیر نسبت شهرنشینی در استان نشان می‌دهد با افزایش شهرنشینی، شانس تعلق بیمار به رده بار کمرشکن بیماری کاهش یافته و بیان‌گر این است که سطح پایین محیط زندگی بیمار عامل خطری برای تحمیل بار کمرشکن اقتصادی بیماری دیابت به حساب می‌آید. برآورد مثبت ضریب متغیر تعداد پزشک متخصص نشان می‌دهد با افزایش

تعداد پزشک متخصص در سطح استان، شانس تعلق بیمار به رده بار کمرشکن بیماری افزایش می‌یابد. این مقدار حاکی از آن است که به دلیل کمبود تخصیص بودجه به هزینه‌های درمانی بیماران توسط سیستم ارائه خدمات بهداشتی کشور، مراجعه بیماران به پزشکان متخصص بیشتر می‌شود و باعث تحمیل هر چه بیشتر بار کمرشکن اقتصادی بیماری بر بیمار می‌شود که لزوم سیاست‌گذاری هر چه سریعتر و مدیریت بهتر بر این موضوع را نشان می‌دهد. نکته قابل ذکر این که هنگام برآورد پارامترها در مدل‌های رگرسیونی چند سطحی معمولاً فرض می‌شود توزیع 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 به‌صورت جامع عمومیت نیافته‌اند.

برای مقایسه دو مدل مفروض از آزمون نسبت درستنمایی(LRT [۱۸]) استفاده می‌کنیم که به‌صورت زیر تعریف می‌شود

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

که در آن P نشان‌دهنده تعداد پارامترها، $\hat{\theta}$ برآورد پارامترهای مدل و $\hat{\theta}_0$ برآورد پارامترها تحت فرض صفر می‌باشد و آماره آزمون دارای توزیع خی‌دو با درجه آزادی P است. مقدار(۷۵/۹) به‌دست آمده از آزمون نسبت درستنمایی نشان می‌دهد که مدل حاصل از جدول ۵ با دو پارامتر اضافی در مقایسه با مدل جدول ۴، معنی‌دار است و بنابراین پارامترهای اضافه شده به مدل غیر صفر می‌باشند ($P-value < 0/05$). همچنین واریانس اثر تصادفی در سطح دوی مدل، غیر صفر و معنی‌دار است ($P-value < 0/05$) که لزوم استفاده از مدل دو سطحی را نشان می‌دهد.

محدودیت‌های تحقیق: یکی از نقاط ضعف مطالعه، نادقیق بودن متغیر وابسته از نقطه نظر اندازه گیری است. در حقیقت بهتر می‌بود همراه دو سؤال میزان کاهش و جذب درآمد ماهیانه در پرسشنامه، میزان درآمد ماهیانه بیمار نیز پرسیده می‌شد تا متغیر پاسخ یعنی بار اقتصادی تحمیل شده بر بیمار به نحوی تعدیل یابد. همچنین به‌دلیل پرسشنامه‌ای بودن ابزار اندازه‌گیری، ممکن است مطالعه در معرض خطای غیر تصادفی یادآوری (recall bias) قرار گرفته باشد. از نقاط قوت مطالعه می‌توان به حجم 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	$1 \quad -0.495$	0/72

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

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

جدول ۳ : میانگین، انحراف معیار، دامنه تغییرات و میانه متغیرهای پیش بین 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	$logit(\gamma_{ij})$ (فاصله اطمینان 95% برای OR)	مقدار $-p$
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	$-0/01$ $(0/92, 1/07)$ $*(0/04)$	
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	$logit(\gamma_{ij}^{(1)})$ $0/36$ $(0/80, 2/60)$ $*(0/30)$	
	$logit(\gamma_{ij}^{(2)})$ $0/36$ $(4/30, 13/50)$ $*(0/30)$	
مولفه واریانس		
σ_{0j}^2	$0/87$ $(0/11)$	$> 0/05$
آماره لگاریتم درستنمایی	$-1991/5992$	

اعداد داخل پرانتز خطای استاندارد(SE) را نشان می دهند.

p مقدار مربوط به نسبت شانسها می‌باشد.

$*$ در سطح ۰/۰۵ 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 95% 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	$-1/3$ $(0/13, 0/60)$ $*(0/38)$ $logit(\gamma_{ij}^{(1)})$	
	$0/36$ $(0/69, 2/90)$ $*(0/37)$ $logit(\gamma_{ij}^{(2)})$	
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	$-1986/72$	
آزمون نسبت درستنمایی	$9/75$	

اعداد داخل پرانتز خطای استاندارد(SE) را نشان می دهند.

p مقدار مربوط به نسبت شانسها می‌باشد.

$*$ در سطح ۰/۰۵ غیر معنی‌دار

مراجع

[1] Agresti, A. (2007).*An introduction to categorical data analysis*. 3nd 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. *Jornal 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] Littele, R.C., Milliken,G.A., Stroup, W.A., Wolfinger, R.D. and Schabenberger, O. (2006).*SAS for Mixed Models*, Secend Edition. Cary, NC: SAS Institute.

[11] Lesaffre, E. and Spiessense, 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. *Jornal 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 Jornal*, **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.