

اندیشه آماری، بهار و تابستان ۱۳۹۱، شماره پیاپی ۳۳

سال هفدهم شماره اول، ص ۴۳-۲۹

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 پژوهشکی، 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 گستته (مانند دو یا چند حالتی، شمارشی، ترتیبی و...) از مدل‌های رگرسیونی تعیین 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 درستنمایی حاشیه‌ای و پیشگویانه، روش مریع‌بندی گوسی، روش مریع‌بندی گوسی سازوار.

^۱maryam_hadipoor@yahoo.com

۱ مقدمه

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

می‌باشد (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 برآورد 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 برآورد 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 مراتی روی رو 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 کاراتراز تحلیل در سطح خوش‌های است (هدکر و گیبز، 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 گیبز، ۱۹۹۴). مدل‌های رگرسیونی چند سطحی تعمیم‌یافته^۲، ردیه از مدل‌های رگرسیونی چند سطحی برای انواع مختلفی از متغیرهای وابسته

۲ مدل رگرسیونی دو سطحی

تعمیم‌یافته

به طور کلی مدل‌های رگرسیونی خطی تعمیم‌یافته شامل سه مؤلفه می‌باشد (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.

شكل کلی مدل‌های خطی تعمیم‌یافته 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 $i = 1, \dots, n_j$ به واحدهای سطح

یک و $j = 1, \dots, N$ به واحدهای سطح

دو دلالت دارند و مقدار مورد انتظار متغیر پاسخ از

طريق تابع پیوند η_{ij} با متغیرهای پیش‌بین به صورت مدل رگرسیون لجستیک ترتیبی دو سطحی به صورت

زیر است.

$$\begin{aligned} logit(\gamma_i^{(s)}) &= \ln \frac{\gamma_i^{(s)}}{1 - \gamma_i^{(s)}} \\ &= \beta_2 x_i - \alpha^{(s)} \end{aligned} \quad (3)$$

$$\eta_{ij} = g(\mu_{ij}) = x_{ij}\beta + \nu_j \quad (2)$$

ν_j اثر تصادفی خوشه j -ام می‌باشد که به جمله که در آن ν_{0j} اثر تصادفی سطح دو در مدل می‌باشد

اضافه می‌شود. این اثر تصادفی، تأثیر خوشه $X_{ij}\beta$ (مک‌کولا و نلدر، ۱۹۸۳).

ز-ام را روی واریانس مشاهدات درونی اش که توسط

متغیرهای پیش‌بین توضیح داده نشده‌اند، نشان می‌دهد.

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

روش‌های برآورد پارامترها در مدل‌های رگرسیونی چند سطحی

مدل رگرسیونی لجستیک ترتیبی:

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

از روش شانس متناسب استفاده می‌کنیم. فرض تعییم‌یافته بسیار متداول شده‌اند، روش‌های برآوردهایی

کنید متغیر پاسخ یک متغیر ترتیبی با S (= s) پارامترهای آن‌ها از اهمیت ویژه‌ای برخوردار است.

روش‌ها و برنامه‌های کامپیوتری متنوعی برای برازش R رده باشد به طوری که

این مدل‌ها پیشنهاد شده‌اند. با این وجود نمی‌توان $\gamma_i^{(s)} = P(y_i > s), s = 1, \dots, S - 1$

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

^۳Cumulative(proportional)odds (PO)

^۴Adjacent Categories

^۵Continuation Ratio

پارامترها به کار می‌رود، یکتابع حاشیه‌ای می‌باشد این معنا که در برآورد پارامترهای مدل بیش برآورده می‌کنند. به منظور رفع این مشکل از روش‌های PQL و MQL که توسط مک کولا و سیرل (۲۰۰۱) که در حالت کلی شکل بسته‌ای ندارد. بنابراین بررسی شده‌اند، استفاده می‌شود. این روش‌ها توسط از روش‌های تقریبی برای برآورد پارامترها استفاده می‌شود (رب هسکت و همکاران، ۲۰۰۲ و هدکر، خطی‌سازی مدل از طریق بسط سری تیلور انجام می‌شوند و از روش‌های حداقل مربعات تعیین یافته زمینه به کار می‌روند عبارتند از تقریب لایاس، تکراری (IGLS^{۱۱}) و یا حداقل مربعات تعیین یافته روش‌های بیزی، روش‌های شبه درستمایی حاشیه‌ای وزن دار شده تکراری (RIGLS^{۱۲}) برای برآورد و پیشگویانه (MQL^۶ و PQL^۷)، روش مریع‌بندی پارامترهای مدل استفاده می‌کنند.

گوسی (GQ^۸) و روش مریع‌بندی گوسی سازوار مرتبه PQL و MQL با درنظر گرفتن تعداد جملاتی AGQ^۹). تقریب لایاس 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 خوش‌ها 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 داده‌های 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).

اما در حالتی که حجم نمونه درون خوش‌ها کوچک با پاسخ‌های غیرپیوسته، روش‌های برآورد ماکریزم باشد، به طور ضعیف عمل می‌کنند (مک کولا و درستمایی به طور محاسباتی افراطی عمل می‌کنند، سیرل، ۲۰۰۱ و دریگز و گلدمان، ۲۰۰۱). روش‌های MQL و PQL 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 هدکر در سال ۱۹۹۷ برای و روش‌های GQ و AGQ در نرم افزار STATA مدل‌های خطی تعیین‌بافته به کار برده شد. این روش برای برآورد پارامترها استفاده شد و سپس نتایج مذکور مقایسه گردید. به تعداد زیادی نقاط مربع‌بندی برای تقریب تابع درستنمای نیاز دارد و با مشکلاتی همراه است که برای پاسخ‌های دو حالتی توسط لیسافر و اسپیسائز (۲۰۰۱) بررسی شده‌اند. رادنباش و بربیک (۲۰۰۲) با استفاده از مطالعات شبیه‌سازی نشان دادند که برای پاسخ‌های دو حالتی توسط لیسافر و اسپیسائز (۲۰۰۱) بررسی شده‌اند. رادنباش و بربیک (۲۰۰۲)

۴ برازش مدل رگرسیون لجستیک ترتیبی دو سطحی به داده‌های دیابت

داده‌های مورد استفاده در طرح حاضر از مطالعه زیاد است، به طور ضعیف عمل می‌کند (لیسافر و کشوری «برآورد هزینه دیابت در ایران در سال ۸۵» می‌باشد که به صورت مقطعی و توسط مرکز اسپیسائز، ۲۰۰۱). برای غلبه بر مشکلات GQ، از روش AGQ که به نقاط مربع‌بندی کمتری نیاز تحقیقات غدد درون‌ریز و متابولیسم دانشگاه علوم پزشکی تهران و معاونت سلامت وزارت بهداشت در ۲۷ استان بر روی بیماران دیابتی تهیه شده است. روب هسکت و همکاران (۲۰۰۵) برای این مطالعه برای اولین بار و با روش نمونه‌گیری مقایسه عملکرد دو روش GQ و AGQ، از مدل‌های AGQ بهتر از GQ عمل می‌کند. روش‌های اقتصادی تحمیل شده بر بیمار (از تلفیق دو متغیر دو سطحی با پاسخ‌های دو حالتی و حجم‌های نمونه‌ای تصادفی خوش‌های چند مرحله‌ای در ایران انجام شده و همبستگی‌های درون خوش‌های متفاوتی استفاده است. هدف از برازش مدل توضیحی تعیین عوامل کردن و نشان دادن در حالت‌هایی که حجم نمونه مؤثر فردی و محیطی بر بار اقتصادی دیابت نوع درون خوش‌بزرگ و همبستگی درون خوش‌های زیاد دو بر بیماران دیابتی می‌باشد. متغیر پاسخ (بار باشد، AGQ بهتر از GQ عمل می‌کند. روش‌های اقتصادی تحمیل شده بر بیمار) از تلفیق دو متغیر دو حالتی کاهش و جذب درآمد ماهیانه بیمار (کمتر از AGQ و GQ توسط دستور gllamm در STATA قابل اجرا هستند (رب هسکت و اسکروندا، ۲۰۰۸، ۳۰ درصد، ۳۰ درصد و بیشتر) به دلیل مشکلات و اسکروندا و رب هسکت، ۲۰۰۳). در این مقاله مرتبط با دیابت تعیین گردید. لازم به ذکر است که از روش‌های MLwiN و PQL در نرم افزار MQL و PQL به بیماری دیابت تعریف می‌شود که شده مربوط به بیماری دیابت تعریف می‌شود که

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 ۳۰%
به عنوان بار اقتصادی نرمال ۲ - مقادیر بالای ۳۰%

به عنوان بار اقتصادی کمرشکن و ۳ - در صورتی که در این بخش، نتایج مطالعه حاضر، در دو قسمت یکی از متغیرها بالای ۳۰% و دیگری پایین ۳۰% توصیفی و تحلیلی ارائه می‌شوند.

باشد به عنوان بار اقتصادی متوسط در نظر گرفته شد

۱۰.۵ توصیف داده‌ها

نتایج این قسمت در قالب جداولی در انتهای مقاله آورده شده است. مصاحبه رو در رو قرار گرفتند. از این تعداد ۳۲۳۴ پرسشنامه به طور کامل تکمیل گردید و مورد تحلیل قرار گرفتند. پایانی پرسشنامه توسط یک نمونه

۲۰.۵ تحلیل داده‌ها

به منظور برآشش مدل به داده‌های دیابت، ابتدا مدل رگرسیونی لجستیک ترتیبی دو سطحی با عرض از آگاهی از ابتلا به بیماری در نظر گرفتن متغیرهای پیش‌بین مبدأ تصادفی و بدون در نظر گرفتن متغیرهای پیش‌بین در سطح دو و سپس با در نظر گرفتن متغیرهای تعداد پژوهش 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 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 (گلدوست).

^{۱۴}Health System Performance Assessment, Christopher Murray , David Evans, WHO 2003

^{۱۵}Variance Partition Coefficient

^{۱۶}Health Utility Index

و همکاران، ۲۰۰۲). فرمول این ضریب برای مدل‌های تأثیرگذار باشد. در مثال بیان شده، اگر چه هر دو رگرسیونی لجستیک، به صورت زیر می‌باشد

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

پیش‌بین نشان دادند، ولی این مدل‌ها منجر به نتایج متفاوتی در رابطه با اندازه اثرات و اندازه واریانس با توجه به فرمول بالا، مقدار ضریب تفکیک واریانس برابر ۱۹٪ 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 ($P-value < 0/01, r = 0/17$) و بنابراین برای بوجود نیامدن مشکل همخطی، متغیر پیش‌بین تعداد سال‌های آگاهی از بیماری از مدل حذف شده است. همان‌گونه 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 شناس متناسب استفاده کردیم. در این فرض اثر (یا شناس) 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 به دست آمده از یک مجموعه داده

مقدار VPC با توجه به فرمول بخش ۵.۲ برای مدل

^{۱۶} Variance Partition Coefficient

(۱) برابر $21/0$ و برای مدل (۲) برابر $19/0$ به زنان شانس بیشتری برای تحمل کردن بار کمرشکن می‌باشد. مقدار به دست آمده از مدل (۱) نشان و متوسط اقتصادی بیماری دیابت نسبت به بار نرمال دهنده این است که خوشبتدی بیماران درون استانها $19/3 =$ نسبت شانس). همچنین هر چقدر دارند(۳) درصد از پراکندگی موجود در مشاهدات وضعیت اقتصادی بیمار بدتر باشد، شانس بیشتری را که توسط متغیرهای پیش‌بین موجود در مدل برای تحمل بار کمرشکن اقتصادی بیماری دیابت توضیح داده نشده است، تعیین می‌کند. مقدار VPC را داراست. به عنوان مثال شانس تعلق بیماران به به دست آمده از مدل (۲) نشان می‌دهد در حالتیکه رده بار کمرشکن نسبت به نرمال و متوسط برای متغیرهای پیش‌بین به مدل اضافه شده است، خوش بیماران فقیر نسبت به غنی برابر $46/2$ می‌باشد.

شدن بیماران درون استانها 19 درصد از تغییرپذیری یافته‌ها از جدول 4 نشان می‌دهد با افزایش شاخص داده‌ها (یعنی واریانس تبیین نشده) را توضیح می‌دهد. سلامتی، شانس تعلق بیمار به رده بار کمرشکن مجموعه داده حاضر، مقدار نسبتاً بالا و منطبق با بیمار کمتر می‌شود (به عبارتی نسبت شانس 20 درصد کاهش می‌یابد) و بنابراین بیماران در وضع (ژاکوبز و همکاران، 1989). لازم به ذکر است سلامتی بهتر، شانس بیشتری برای تعلق داشتن به VPC به عوامل گوناگونی مانند حجم نمونه در که نشان می‌دهد سکونت در روستا عامل خطری خوش‌های مختلف، افزایش و یا کاهش متغیرهای پیش‌بین و نوع متغیر پاسخ بستگی دارد (هدکر و بیماران ساکن شهر نسبت به روستا برابر $8/0$ می‌باشد 1994). در مدل (۲) با آوردن اثر سطح که نشان می‌دهد سکونت در روستا عامل خطری دو (یعنی استان‌های مختلف) در مدل رگرسیونی، برای تحمل بار کمرشکن اقتصادی بیماری دیابت دارد. ملاحظه می‌شود که اثر متغیرهای پیش‌بین سطح برآورد منفی ضریب متغیر نسبت شهرنشینی در استان یک مانند جنسیت، محل سکونت، وضعیت سلامتی، نشان می‌دهد با افزایش شهرنشینی، شانس تعلق بیمار وضعیت بیمه بودن و سن بر بار اقتصادی بیماری به رده بار کمرشکن بیماری کاهش یافته و بیان گر دیابت نوع دو تحملی بر بیمار معنی دارد ($P -$ این است که سطح پایین محیط زندگی بیمار عامل خطری برای تحمل بار کمرشکن اقتصادی بیماری $value < 0/05$). همان گونه که از جدول 4 مشاهده می‌شود ضریب متغیرهای جنسیت و وضعیت دیابت به حساب می‌آید. برآورد مثبت ضریب متغیر اقتصادی مثبت می‌باشد که نشان می‌دهد مردان نسبت تعداد پژوهشک متخصص نشان می‌دهد با افزایش

تعداد پزشک متخصص در سطح استان، شناس تعلق صفر می‌باشد و آماره آزمون دارای توزيع 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 درستنمایی 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 می‌باشد ($< P-value$) هر چه بیشتر بار کمرشکن اقتصادی بیماری بر بیمار (۰/۰۵). همچنین واریانس اثر تصادفی در سطح می‌شود که لزوم سیاست‌گذاری هر چه سریعتر و دوی مدل، غیر صفر و معنی‌دار است ($< P-value$) مدیریت بهتر بر این موضوع را نشان می‌دهد. نکته (۰/۰۵) که لزوم استفاده از مدل دو سطحی را نشان قابل ذکر این که هنگام برآورد پارامترها در مدل‌های می‌دهد.

رگرسیونی چند سطحی معمولاً فرض می‌شود توزيع **محدودیت‌های تحقیق**: یکی از نقاط ضعف اثر تصادفی (μ_u) نرمال است در صورتیکه در مقاله مطالعه، نادقيق بودن متغير وابسته از نقطه نظر اندازه حاضر پس از برآورد اثر تصادفی با استفاده از روش گیری است. در حقیقت بهتر می‌بود همراه دو سؤال بیز تجربی ملاحظه شد که توزيع اثر تصادفی غیر میزان کاهش و جذب درآمد ماهیانه در پرسشنامه، نرمال و چوله به راست می‌باشد. برای حل این میزان درآمد ماهیانه بیمار نیز پرسیده می‌شد تا متغير مشکل روش‌های نوین برآوردهای بیضوی^{۱۷} ابداع پاسخ یعنی بار اقتصادی تحمیل شده بر بیمار به شده‌اند که تا کنون به صورت جامع عمومیت نیافتها ند. نحوی تعديل یابد. همچنین به دلیل پرسشنامه‌ای برای مقایسه دو مدل مفروض از آزمون نسبت بودن ابزار اندازه‌گیری، ممکن است مطالعه در معرض درستنمایی (LRT^{۱۸}) استفاده می‌کنیم که به صورت خطای غیر تصادفی یادآوری (recall bias) قرار گرفته باشد. از نقاط قوت مطالعه می‌توان به حجم بالای نمونه، تعداد زیاد متغيرهای پیش‌بین در سطح ۱ و ۲ و به کارگیری روش‌های دقیق مانند استفاده از پرسشنامه استاندارد HUI برای تعیین میزان سلامتی بیمار اشاره کرد.

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

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

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

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

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 (%22/7) (%57/8) (%19/5)

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

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

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

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

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

p مقدار مربوط به نسبت شانسها می‌باشد.

* در سطح ۵٪ غیر معنی‌دار

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

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

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

 p مقدار مربوط به نسبت شانسها می‌باشد.

* در سطح ۵٪ 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] Littlele, 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. *Jurnal 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.