

Comparison of Marginal Logistic Model with Repeated Measures and Conditional Logistic Model in Risk Factors Affecting Hypertension

Neda Gilani¹,
Anoshirvan Kazemnejad¹,
Farid Zayeri²,
Jamshid Yazdani³

¹ Department of Biostatistics, Faculty of Medical Sciences, Tarbiat Modares University, Tehran, Iran

² Department of Biostatistics, Faculty of Paramedical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran

³ Psychiatry and Behavioral Sciences Research Center, Department of Epidemiology and Biostatistics, Faculty of Health, Mazandaran University of Medical Sciences, Sari, Iran

(Received May 9, 2010 ; Accepted March 19, 2011)

Abstract

Background and purpose: To analyze the data in which the correlation between observations are to be considered, a general method is using marginal model with repeated measures, yet there is another method called conditional model with random clusters. According to the binary responses, the aim of the present study is to compare the efficiency of these two models in studying the risk factors affecting hypertension in Mazandaran province.

Materials and methods: This cross-sectional study is part of research studying the risk factors of the non-infectious diseases in Mazandaran province in 2007 with a sample size of 1000 cases. First, stratified sampling was employed and then random cluster sampling was used in each stratum. In all cases, blood pressure was measured three times through standard methods. The obtained data were analyzed using χ^2 test and fitting marginal and conditional logistic models.

Results: Of the total of 1000 (500 male and 500 female) cases, 38 percent had high blood pressure. In both models four variables of age, physical activity, body-mass-index, and the consumption of fruits and vegetables were identified as factors affecting hypertension. In addition, gender and fasting blood sugar were statistically significant only in marginal logistic model. Using goodness-of-fit criteria, values of all the statistics for conditional logistic model with random effects of individuals were lower in comparison with the marginal logistic model with repeated measures.

Conclusion: It can be concluded that although one cannot decisively select a model as the most appropriate one, conditional logistic model fits were somewhat better than the marginal logistic model with repeated measures, so it can be used to analyze such data.

Key words: Hypertension, marginal logistic model with repeated measures, conditional logistic model with random clusters, goodness-of-fit criteria

مقایسه مدل‌های لجستیک حاشیه‌ای با اندازه‌گیری مکرر و لجستیک شرطی در بررسی عوامل موثر بر پرفشاری خون

ندا گیلانی^۱انوشیروان کاظم نژاد^۲فرید زابری^۳جمشید یزدانی^۴

چکیده

سابقه و هدف: دو تا از مهمترین روش‌های تحلیل برای داده‌های همبسته عبارتند از استفاده از مدل حاشیه‌ای با اندازه‌گیری مکرر و دیگری بکارگیری مدل‌های شرطی با خوشه‌های تصادفی می‌باشد. با توجه به دو حالتی بودن متغیر پاسخ، هدف از مطالعه حاضر مقایسه کارایی دو مدل لجستیک حاشیه‌ای با اندازه‌گیری مکرر و لجستیک شرطی با خوشه‌های تصادفی برای بررسی عوامل خطر موثر بر بیماری پرفشاری خون در استان مازندران می‌باشد.

مواد و روش‌ها: این مطالعه مقطعی قسمتی از طرح بررسی عوامل خطر بیماری‌های غیر واگیردار استان مازندران در سال ۱۳۸۶ با حجم نمونه ۱۰۰۰ نفر بود. روش نمونه‌گیری بصورت طبقه‌ای و در داخل طبقات بصورت تصادفی خوشه‌ای یک مرحله‌ای بود. در تمامی افراد فشار خون سه بار به روش استاندارد اندازه‌گیری شد. داده‌های حاصل با استفاده از آزمون Chi-square و برازش مدل‌های لجستیک حاشیه‌ای و شرطی مورد تجزیه و تحلیل قرار گرفت.

یافته‌ها: از مجموع ۱۰۰۰ فرد (۵۰۰ مرد و ۵۰۰ زن) مورد مطالعه، ۳۸ درصد دارای بیماری پرفشاری خون بوده‌اند. در هر دو مدل، متغیرهای گروه سنی، فعالیت بدنی، نمایه توده بدنی و مصرف میوه و سبزی از عوامل تاثیرگذار بر پرفشاری خون شناخته شدند. اضافه بر این متغیرهای جنسیت و قند خون ناشتا فقط در مدل لجستیک حاشیه‌ای معنی‌دار آماری گردید. با استفاده از معیارهای نیکویی برازش، مقادیر تمامی آماره‌ها برای مدل لجستیک شرطی با اثر تصادفی افراد نسبت به مدل لجستیک حاشیه‌ای با اندازه‌گیری مکرر، کمتر بدست آمد.

استنتاج: اگرچه نمی‌توان با قاطعیت مدل مشخصی را به عنوان مناسب‌ترین مدل معرفی کرد ولی از میان دو مدل مذکور، مدل لجستیک شرطی تا حدودی برازش بهتری داشته و می‌توان برای تحلیل چنین داده‌هایی از آن بهره برد.

واژه‌های کلیدی: بیماری پرفشاری خون، مدل لجستیک حاشیه‌ای با اندازه‌گیری مکرر، مدل لجستیک شرطی با خوشه‌های تصادفی، معیارهای برازش

مقدمه

در تحلیل کلاسیک آماری معمولاً فرض می‌شود که مشاهدات از هم مستقل هستند، در حالی که این فرض در بسیاری از مطالعات به خصوص در حوزه پزشکی برقرار نیست و در مطالعات زیادی مشاهدات به هم وابسته

E-mail: kazem_an@modares.ac.ir

مؤلف مسئول: انوشیروان کاظم نژاد - تهران: دانشگاه تربیت مدرس، دانشکده علوم پزشکی، گروه آمار زیستی

۱- دانشجو کارشناسی ارشد، گروه آمار زیستی، دانشکده علوم پزشکی، دانشگاه تربیت مدرس

۲. گروه آمار زیستی، دانشکده علوم پزشکی، دانشگاه تربیت مدرس

۳. گروه آمار زیستی، دانشکده پیراپزشکی، دانشکده علوم پزشکی شهید بهشتی

۴. مرکز تحقیقات روان‌پزشکی و علوم رفتاری، گروه آمار و اپیدمیولوژی، دانشکده بهداشت، دانشکده علوم پزشکی مازندران

تاریخ دریافت: ۸۹/۲/۱۹ تاریخ ارجاع جهت اصلاحات: ۸۹/۴/۱۲ تاریخ تصویب: ۸۹/۱۲/۲۸

هستند. برای مثال داده‌هایی که در طول زمان جمع‌آوری می‌شوند- مانند اندازه‌گیری فشار خون هر فرد در دوره‌های زمانی- یا داده‌هایی که در خوشه‌هایی دسته‌بندی می‌شوند. مشاهدات حاصل از یک آزمودنی در دوره‌های زمانی یا دسته‌بندی شده در یک خوشه شباهت بیشتری نسبت به مشاهدات آزمودنی‌ها یا خوشه‌های دیگر دارند. در نتیجه معمولاً دارای همبستگی مثبتی با هم هستند. برای تحلیل درست داده‌هایی که همبسته هستند، به یک مدل آماری نیازمندیم که در آن همبستگی میان مشاهدات لحاظ شود^(۱). در مدل‌بندی حاشیه‌ای از پاسخ‌های خوشه‌بندی شده، ساختار وابستگی توأم را به عنوان یک پارامتر مزاحم در نظر می‌گیرند. اثر تصادفی در این مدل از نوع R-side است^(۲،۳). درحالی‌که در مدل‌بندی شرطی این وابستگی را به صورت عبارتی از سطح خوشه‌ای فرض کرده که دارای مقادیر یکسان برای مشاهدات داخل یک خوشه هستند که این مقادیر از مقادیر مشاهدات واقع در خوشه های دیگر متفاوت هستند. اثر تصادفی در این مدل از نوع G-side خواهد بود. بر اساس این دو مدل کلی، هرگاه متغیر پاسخ دوحالتی و مشاهدات همبسته باشند می‌توان از مدل لجستیک حاشیه‌ای با اندازه‌گیری مکرر و مدل لجستیک شرطی با خوشه‌های تصادفی استفاده کرد^(۴-۶). مدل لجستیک حاشیه‌ای با اندازه‌گیری مکرر که توسط Zeger و Liang در سال ۱۹۸۶ معرفی شد؛ تعمیمی بر مدل‌های خطی تعمیم یافته^۱ با لحاظ همبستگی بین داده‌ای است. در این مدل، همبستگی موجود بین پاسخ‌ها برای هر آزمودنی به عنوان پارامتر در نظر گرفته می‌شود. برآورد پارامترها و همچنین همبستگی میان مشاهدات دریافتی از هر فرد در آن با روش معادلات برآورد تعمیم یافته^۲ به دست می‌آید. مدل لجستیک شرطی با خوشه‌های تصادفی در رده مدل‌های آمیخته خطی تعمیم یافته^۳ قرار می‌گیرد. از این

مدل‌ها زمانی استفاده می‌شود که بخواهیم اثرهای تصادفی را به عنوان بخشی از پیش بین خطی در مدل‌های خطی تعمیم یافته قرار دهیم. استفاده از تقریب لاپلاس که بر پایه درست‌نمایی تقریبی است، یکی از روش‌های استنباطی مدل‌های آمیخته خطی تعمیم یافته محسوب می‌شود. اگرچه مدل‌های با اندازه‌گیری مکرر، کاربردی‌ترین مدل در همبستگی‌های با دوره‌های زمانی است، مدل‌های شرطی با خوشه‌های تصادفی در برخی شرایط می‌توانند مناسب‌تر باشند. چون محاسبه تابع درست‌نمایی مدل‌های شرطی با خوشه‌های تصادفی به نوعی فرم بسته‌ای ندارد. در گذشته استفاده محدودی از این مدل‌ها می‌شد اما امروزه با پیشرفت‌هایی که در زمینه نرم‌افزارهای آماری روی داده است، استفاده کاربردی از این مدل‌ها در حال افزایش است^(۷،۸).

فشارخون بالا عبارت است از افزایش فشار وارده از جریان خون به دیواره رگ‌های خونی. فشار خون بالا یکی از مهمترین عوامل خطر بروز بیماری‌های عروق کرونری و شایع‌ترین عامل بروز نارسایی قلبی و سکته مغزی در بسیاری از کشورها و نیز مهمترین علت ایجاد نارسایی کلیه می‌باشد. در حقیقت بعد از دیابت، مهم‌ترین دلیل شایع نارسایی مزمن کلیوی پرفشاری خون است. با تغییر الگوی اپیدمیولوژیک بیماری‌ها در کشورهای پیشرفته و در حال توسعه و افزایش امید به زندگی، افزایش استرس‌های ناشی از زندگی شهرنشینی و تغییر الگوهای غذایی، پرفشاری خون به عنوان مهمترین عامل خطر برای زمینه‌ساز بیماری‌های قلبی-عروقی، تهدیدی جدی تلقی می‌شود^(۹-۱۷).

هدف این مطالعه استفاده از مدل‌های لجستیک حاشیه‌ای با اندازه‌گیری مکرر و لجستیک شرطی با خوشه‌های تصادفی برای بررسی عوامل خطر موثر بر بیماری پرفشاری خون در استان مازندران است.

1. Generalized linear models (GLM)

2. GEE

3. Generalized linear mixed models (GLMM)

مواد و روش ها

این مطالعه از نوع کاربردی می باشد و داده های مورد استفاده در آن که به شیوه مقطعی گردآوری شده است، قسمتی از طرح بررسی عوامل خطر بیماری های غیر واگیردار در استان مازندران در سال ۱۳۸۶ بود. روش نمونه گیری به صورت طبقه ای (طبقات شامل شهرها، روستاها و در شهرهای بزرگ، مناطق شهری) و در داخل طبقات بصورت تصادفی خوشه ای یک مرحله ای بود. برای بالا بردن کارایی نمونه گیری ترتیب قرار دادن طبقات در شهرها بر اساس موقعیت اقتصادی- اجتماعی بود. نمونه گیری در کلیه مراحل بصورت تخصیص متناسب^۱ صورت گرفت. حجم نمونه به گونه ای بود که حداقل تعداد نمونه در هر خانه^۲ سنی- جنسی ۱۰۰ نفر باشد. بنابراین در هر گروه سنی تعداد ۲۰۰ نفر (۱۰۰ زن و ۱۰۰ مرد) انتخاب شدند. در نهایت باتوجه به بررسی در پنج گروه سنی، حجم نمونه در استان برابر ۱۰۰۰ نفر و در هر خوشه ۲۰ نفر بود که بدین ترتیب در استان ۵۰ خوشه بررسی شد (تصویر ۱).

جمع آوری اطلاعات در این مطالعه از سه طریق زیر انجام گرفته است: پرسشنامه همراه با مصاحبه حضوری؛ مطالعات بالینی با اندازه گیری قد، وزن، دور کمر و فشار خون؛ اندازه گیری قندخون ناشتا، کلسترول تام و HDL پلاسما.

متغیرهای مورد سنجش

ابتلا به بیماری پر فشاری خون: متغیر پاسخ عبارت است از بیماری پر فشاری خون (که دو حالتی می باشد: پر فشاری خون ندارد = ۰ و پر فشاری خون دارد = ۱) که این متغیر برای هر فرد در سه دوره زمانی با فواصل زمانی یک دقیقه اندازه گیری شد. این متغیر دو حالتی با توجه به تعریف JNC-7^۳ بدست آمد؛ یعنی فشارخون

سیستولیک بیشتر یا مساوی ۱۴۰ میلی متر جیوه و یا فشار دیاستولیکی بیشتر یا مساوی ۹۰ میلی متر جیوه به عنوان بیماری پر فشاری خون در نظر گرفته شد (۹). همچنین کلیه افراد تحت درمان دارویی ضد فشار نیز در گروه هیپرتانسیو قرار گرفتند.

متغیرهای قند خون ناشتا، کلسترول و HDL با نقاط برش به ترتیب ۱۲۶، ۲۴۰ و ۳۵ به صورت متغیرهای کیفی در نظر گرفته شدند. متغیرهای متوسط فعالیت بدنی و مصرف میوه و سبزی به ترتیب از مجموع متوسط فعالیت و ورزش یک فرد در طول هفته و متوسط تعداد مصرفی واحدهای میوه و سبزی در طول یک هفته بدست آمد. همچنین متغیرهای سن، جنس، محل سکونت (شهر، روستا) و نمایه توده بدنی (BMI^۴) نیز به عنوان متغیرهای کمکی در نظر گرفته شد.

تجزیه و تحلیل اطلاعات

مدل لجستیک حاشیه ای با اندازه گیری مکرر

مدل لجستیک با اندازه گیری مکرر که از نوع مدل های رگرسیونی شبه پارامتری است؛ برای تحلیل مشاهدات دوحالتی همبسته در مطالعات طولی از نوع اندازه گیری مکرر و مطالعات چند سطحی کاربرد فراوان دارد. ۳ متغیر دوحالتی فشارخون را با توجه به سه دوره زمانی با (Y_1, Y_2, Y_3) نشان می دهیم. قرار می دهیم:

$$\pi_i = P(Y_1 = i_1, Y_2 = i_2, Y_3 = i_3)$$

ملاحظه می شود $2^3 = 8$ جدول متقاطع از توزیع توأم (Y_1, Y_2, Y_3) بدست می آید. برآورد پارامترها و همچنین همبستگی میان مشاهدات دریافتی از هر فرد برای این مدل با روش معادلات برآورد تعمیم یافته بدست می آید. در روش GEE^۵ نیازی به تعیین کامل توزیع

4. Body mass index

5. Generalized estimating equation

1. Proportional

2. Cell

3. Joint national committee

داده ای نیست. به عبارتی، سازگاری برآورد کننده GEE تنها به تعیین درست توابع میانگین یعنی $1 \leq i \leq n$ در

$$\sum_{i=1}^N \dot{\mu}_i' V_i^{-1} (y_i - \mu_i) = 0 \quad \text{معادله}$$

بستگی دارد که در آن $y_i = (y_{ij})_{1 \leq j \leq n}$ ، $\mu_i = E(y_i)$ و $V_i = \text{var}(y_i) = (A_i^{1/2} R A_i^{1/2}) \phi$ بوده و ϕ پارامتر بیش پراکنش است. $\dot{\mu} = \frac{\partial \mu}{\partial \beta}$

مدل لجستیک شرطی با خوشه های تصادفی

مدل لجستیک شرطی با خوشه های تصادفی در رده مدل های آمیخته خطی تعمیم یافته قرار می گیرد. یک مدل $GLMM^1$ را در حالت کلی می توان به صورت زیر در نظر گرفت:

$$y = \mu + e$$

μ برداری از مقدار مورد انتظار شرطی مشاهدات

است که به پارامترهای مدل با یک تابع پیوند g مربوط می شود. به طوری که:

$$\mu = E(y | \alpha)$$

$$g(\mu) = X\beta + Z\alpha$$

X و Z ماتریس های طرح معلوم از اثرهای ثابت و تصادفی هستند، β و α به ترتیب برداری از پارامترهای اثر ثابت و تصادفی همانند مدل آمیخته نرمال هستند. بر اساس مدل لجستیک شرطی با خوشه های تصادفی، هر فرد را می توان به عنوان یک خوشه در نظر گرفت که این خوشه ها در واقع سطوحی تصادفی محسوب می گردند. بر اساس این مدل، مشاهدات داخل هر خوشه با هم همبسته بوده (اندازه گیری فشارخون مربوط به یک فرد) ولی با مشاهدات سایر خوشه ها همبستگی لحاظ نمی شود. در نتیجه مدل به صورت مدلی آمیخته لجستیک با اثر تصادفی افراد (اثر خوشه ها) و چندین اثر ثابت خواهد بود. اگر Y_{i3}, Y_{i2}, Y_{i1} تعداد موارد دارای بیماری فشارخون خوشه i ام با توجه به سه

زمان باشند؛ احتمال ابتلا به بیماری برای هر فرد با فرض ارتباط خطی لجیت احتمال ها و خوشه ها، به صورت مدل لجستیک با اثر تصادفی فرد خواهد بود. استنباط آماری این مدل با استفاده از تقریب لاپلاس بدست آمد.

معیارهای مقایسه

در این مطالعه برای مقایسه کارایی مدل ها از معیارهای جدول شماره ۱ استفاده شده است. در جدول شماره ۱؛ n نشان دهنده بعد مدل و n^* اندازه داده ای و d - تعداد پارامترهای کوواریانس + رتبه X - است. عموماً هدف این معیارهای برازش، اندازه گیری برازش مدل برآورد است. هر قدر مقادیر آن ها، کم تر باشد نشان دهنده مناسب بودن مدل است (۱۹، ۱۸).

جدول شماره ۱: معیارهای اطلاع

نام معیار	فرمول معیار اطلاع	مرجع
AIC ²	$-2l + 2d$	Akaike (1974)
AICC ³	$-2 + 2dn^* / (n^* - d - 1)$	Hurvich and Tsai (1989)
BIC ⁴	$-2l + d \log n$	Schwarz (1978)
CAIC ⁵	$-2l + d (\log n + 1)$	Bozdogan (1987)
HQIC ⁶	$-2l + 2 \log \log n$	Hannan and Quinn (1979)

برای بررسی اثر عوامل مختلف بر فشار خون بالا به گزارش نتایج آزمون Chi-square و مقایسه حاصل از برازش دو مدل لجستیک حاشیه ای و شرطی پرداخته شد. نتایج با استفاده از پردازش های GLIMMIX نرم افزار SAS نسخه ۹/۲ بدست آمده اند که در این پردازش امکان تحلیل همزمان متغیرهای مستقل کمی و کیفی برای اولین بار میسر شده است. همچنین سطح معنی داری آماری ۰/۰۵ در نظر گرفته شد (۲۲-۲۰).

یافته ها

هدف اولیه یافتن متغیرهای کیفی مرتبط با بیماری پرفشاری خون بود تا از آن ها بتوان در برازش مدل ها

2. Akaike information criterion
3. Small sample bias corrected version of AIC
4. Bayesian inference criterion
5. Consistent Akaike's information criterion
6. Hannan Quinn information criteria

1. Generalized linear mixed model

بهره گرفت. با استفاده از آزمون Chi-square و بر اساس نتایج جدول شماره ۲ ملاحظه می شود که؛ از مجموع ۹۹۹ فرد مورد مطالعه، ۳۸ درصد دارای بیماری پرفشاری خون بوده اند. نتایج جدول شماره ۲ نشان می دهد که: نوع شغل و رده سنی با این بیماری مرتبط بوده در حالی که چنین چیزی در رابطه با متغیر محل سکونت دیده نشد. افرادی که میزان قند خون ناشتای آن ها بیشتر از ۱۲۶ mg/dL بوده است (۷۴/۵ درصد) بیشتر از افرادی که قند خون ناشتای آن ها پایین تر یا مساوی ۱۲۶ mg/dL (۴۱ درصد) دچار پرفشاری خون می شوند و این که افراد با کلسترول خون کمتر از ۲۴۰ (۴۱ درصد) در مقایسه با افرادی که میزان کلسترول خون آن ها بالاتر از ۲۴۰ بوده است (۵۵/۹ درصد) کمتر مبتلا به فشار خون بالا می شوند. در حالت کلی بدون در نظر گرفتن گروه های سنی تفاوتی بین دو جنس از لحاظ ابتلا به بیماری پرفشاری خون وجود نداشت (p=۰/۱۲۲).

جدول شماره ۳ ابتلا به بیماری پرفشاری خون را به تفکیک گروه های سنی در دو جنس نشان می دهد. ابتلا به پرفشاری خون با افزایش سن، افزایش یافته و در مردان در گروه های سنی ۱۵ تا ۲۴ و ۴۵ تا ۵۴ سال این درصد بیشتر از زنان است اما در بقیه گروه های سنی در زنان بیشتر مشاهده شد.

جدول شماره ۴ بیانگر تحلیل مدل لجستیک حاشیه ای با اندازه گیری مکرر و مدل لجستیک شرطی با خوشه های تصادفی است. در این جدول برآورد ضرایب و نسبت شانس (OR)، خطای استاندارد و فاصله اطمینان ۹۵ درصد ضرایب به همراه آزمون مربوطه به آن - که بی تاثیر بودن متغیر مربوطه بر وضعیت فشارخون افراد را آزمون می کند - برای هر یک از متغیرها به تفکیک دو مدل آورده شده است. بر اساس این جدول در هر دو مدل، متغیرهای گروه سنی، فعالیت بدنی، نمایه توده بدنی و مصرف میوه و سبزی از عوامل تاثیر گذار بر بیماری پرفشاری خون شناخته شدند. اضافه

بر این متغیرهای جنسیت و قند خون ناشتا فقط در مدل لجستیک حاشیه ای از لحاظ آماری معنی دار گردید و دیگر متغیرهای زمینه ای مورد بررسی از لحاظ آماری معنی دار نبودند. در جدول شماره ۵ با توجه به معیارهای اطلاع، آماره های نیکویی برازش و مقادیر لگ درست نمایی که از هر دو مدل حاشیه ای و شرطی بدست آمدند، در دو مدل با هم مقایسه شدند. تمامی این آماره ها برای مدل لجستیک شرطی با اثر تصادفی افراد، دارای مقادیر کمتری نسبت به مدل لجستیک حاشیه ای با اندازه گیری مکرر هستند. مقادیر کمتر این آماره ها دلالت بر بهتر بودن مدل برازشی دارد.

جدول شماره ۲: توزیع فراوانی افراد بر اساس ابتلا به بیماری فشارخون و متغیرهای پایه رسته ای

متغیر	دارای فشارخون تعداد(درصد)	سالم تعداد(درصد)	سطح معنی داری
کل افراد	۳۸۰ (۳۸)	۶۱۹ (۶۲)	
جنسیت			۰/۱۲۲
مرد	۱۸۴ (۳۶/۹)	۳۱۵ (۶۳/۱)	
زن	۱۹۶ (۳۹/۲)	۳۰۴ (۶۰/۸)	
محل سکونت			
روستایی	۲۵۹ (۳۹/۳)	۴۰۰ (۶۰/۷)	۰/۱۲۶
شهری	۱۲۱ (۳۵/۶)	۲۱۹ (۶۴/۴)	
نوع شغل			
کارمند	۲۴ (۲۳/۳)	۷۹ (۷۶/۷)	
آزاد	۹۸ (۳۶/۷)	۱۶۹ (۶۳/۳)	
در حال تحصیل-سرباز	۱۷ (۲۰/۴)	۶۶ (۷۹/۵)	۰/۰۰۰
خانه دار	۱۷۵ (۴۳/۶)	۲۲۶ (۵۶/۴)	
بیکار	۱۴ (۳۷/۸)	۲۳ (۶۲/۲)	
سایر مشاغل	۵۲ (۴۸/۱)	۵۶ (۵۱/۹)	
سن			
۱۵ - ۲۴	۳۵ (۱۷/۵)	۱۶۵ (۸۲/۵)	
۲۳ - ۳۴	۴۳ (۲۱/۷)	۱۵۵ (۷۸/۳)	
۳۵ - ۴۴	۶۷ (۳۲/۵)	۱۳۹ (۶۷/۵)	۰/۰۰۰
۴۵ - ۵۴	۱۰۳ (۵۳/۱)	۹۱ (۴۶/۹)	
۵۵ - ۶۴	۱۳۲ (۶۵/۷)	۶۹ (۴۳/۳)	
قندخون ناشتا			
<= ۱۲۶	۲۷۳ (۴۰/۹)	۳۹۵ (۵۹/۱)	۰/۰۰۰
> ۱۲۶	۳۸ (۷۴/۵)	۱۳ (۲۵/۵)	
میزان کلسترول خون			
<= ۲۴۰	۲۴۹ (۴۱)	۳۵۹ (۵۹)	۰/۰۰۰
> ۲۴۰	۶۲ (۵۵/۹)	۴۹ (۴۴/۱)	

جدول شماره ۳: ابتلا به بیماری به تفکیک جنس و گروه های سنی

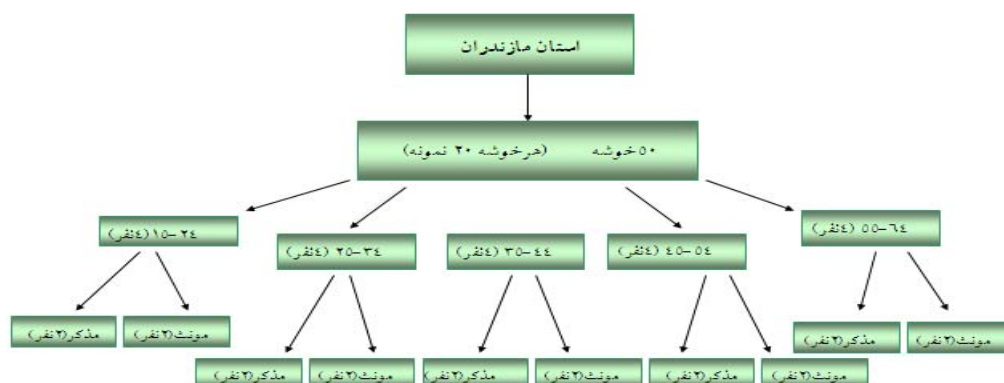
گروه سنی	ابتلا به پرفشاری خون	
	مرد	زن
	تعداد (درصد)	تعداد (درصد)
۱۵-۲۴	۲۸ (۱۵/۲)	۷ (۳/۶)
۲۵-۳۴	۱۹ (۱۰/۳)	۲۴ (۱۲/۲)
۳۵-۴۴	۳۲ (۱۷/۴)	۳۵ (۱۷/۹)
۴۵-۵۴	۵۱ (۲۷/۷)	۵۲ (۲۶/۵)
۵۵-۶۴	۵۴ (۲۹/۳)	۷۸ (۳۹/۸)
کل	۱۸۴ (۱۰۰)	۱۹۶ (۱۰۰)

جدول شماره ۵: مقایسه مدل ها با استفاده از آماره های نیکویی

معیارهای برازش	لجستیک حاشیه ای	لجستیک شرطی
-2 Res Log Pseudo-Likelihood	۶۵۳۱/۸۱	۶۳۴۸/۲۶
Pseudo-AIC	۶۵۷۳/۸۱	۶۳۹۰/۲۶
Pseudo-AICC	۶۵۷۴/۵۱	۶۳۹۰/۹۵
Pseudo-CAIC	۶۶۸۲/۹۱	۶۵۰۷/۴۸
Pseudo-HQIC	۶۶۱۴/۶۹	۶۴۲۷/۴

جدول شماره ۴: تعیین عوامل موثر بر پرفشاری خون با استفاده از مدل های لجستیک حاشیه ای با اندازه گیری مکرر و لجستیک شرطی با خوشه های تصادفی

متغیرها	لجستیک حاشیه ای با اندازه گیری مکرر				لجستیک شرطی با خوشه های تصادفی			
	$\hat{\beta}$ (SE)	نسبت شانس (OR)	فاصله اطمینان ۹۵ درصد	آماره آزمون	سطح معنی داری	$\hat{\beta}$ (SE)	نسبت شانس (OR)	فاصله اطمینان ۹۵ درصد
جنسیت								
مذکر	۰/۶۶ (۰/۳۰)	۱/۹۳	۰/۰۶ و ۱/۲۵	۲/۱۸	۰/۰۲۹	۰/۶۷ (۰/۴۸)	۱/۹۵	۰/۲۸ و ۱/۶۱
مونث								
گروه سنی								
۱۵-۲۴	-۱/۸۶ (۰/۳)	۰/۱۶	-۲/۴۴ و -۱/۲۷	-۶/۲۱	۰/۰۰۰	-۱/۸۹ (۰/۴۶)	۰/۱۵	-۲/۸۰ و -۰/۹۹
۲۵-۳۴	-۱/۵۳ (۰/۲۱)	۰/۲۲	-۱/۹۴ و -۱/۱۲	-۷/۱۷	۰/۰۰۰	-۱/۵۷ (۰/۳۵)	۰/۲۱	-۲/۲۷ و -۰/۸۸
۳۵-۴۴	-۱/۱۳ (۰/۱۹)	۰/۳۲	-۱/۵۱ و -۰/۷۵	-۵/۸۰	۰/۰۰۰	-۱/۱۶ (۰/۳۳)	۰/۳۱	-۱/۸۲ و -۰/۵۱
۴۵-۵۴	-۱/۴۲ (۰/۱۹)	۰/۶۶	-۰/۷۸ و ۰/۰۵	-۲/۲۶	۰/۰۲۴	-۰/۴۰ (۰/۳۳)	۰/۶۷	-۱/۰۵ و ۰/۲۴
۵۵-۶۴								
نوع شغل								
کارمند	-۰/۶۹ (۰/۴۹)	۰/۵۰	-۱/۶۶ و ۰/۲۷	-۱/۴۱	۰/۱۶	-۰/۱۶ (۰/۶۲)	۰/۸۵	-۱/۰۵ و -۱/۳۶
آزاد	-۰/۱۴ (۰/۴۳)	۰/۸۷	-۰/۹۸ و ۰/۷	-۰/۳۳	۰/۷۴	-۰/۱۱ (۰/۸۷)	۱/۱۲	-۱/۸۲ و ۱/۵۹
بیکار	-۰/۱۰ (۰/۶۲)	۰/۹۰	-۱/۳۲ و ۱/۱۲	-۰/۱۶	۰/۸۷	-۰/۷۲ (۰/۶۹)	۰/۴۹	-۲/۰۸ و ۰/۶۵
خانه دار	۰/۷۱ (۰/۴۶)	۲/۰۳	-۰/۱۴ و ۱/۵۵	۱/۶۴	۰/۱۰	۰/۷۲ (۰/۶۷)	۲/۰۵	-۰/۶۰ و ۲/۰۵
سایر	۰/۱۴ (۰/۴۳)	۱/۱۵	-۰/۷ و ۰/۹۹	۰/۳۳	۰/۷۴	۰/۱۶ (۰/۶۵)	۱/۱۷	-۱/۱۳ و ۱/۴۴
در حال تحصیل								
قدخون ناشتا								
کمتر از ۱۲۶	-۰/۶۳ (۰/۲۸)	۰/۵۳	-۱/۱۷ و ۰/۰۹	-۲/۲۸	۰/۰۲	-۰/۶۳ (۰/۵۹)	۰/۵۳	-۱/۷۸ و ۰/۵۲
بیشتر از ۱۲۶								
کلسترول								
کمتر از ۲۴۰	-۰/۱۹ (۰/۲۱)	۰/۸۳	-۰/۶ و ۰/۲۲	-۰/۹۰	۰/۳۷	-۰/۱۹ (۰/۴۰)	۰/۸۳	-۰/۹۸ و ۰/۶۱
بیشتر از ۲۴۰								
HDL								
کمتر از ۳۵	۰/۰۸ (۰/۲۵)	۱/۰۸	-۰/۴۰ و ۰/۵۶	۰/۳۲	۰/۷۵	۰/۰۸ (۰/۴۵)	۱/۰۸	-۰/۸۱ و ۰/۹۷
بیشتر از ۳۵								
نمایه توده بدنی BMI	۰/۰۸ (۰/۰۱)	-	۰/۰۵ و ۱/۰۱	۵/۹۵	۰/۰۰۱	۰/۰۸ (۰/۲۴)	-	۰/۰۳ و ۱/۱۳
مصرف میوه و سبزی	-۰/۰۹ (۰/۰۴)	-	-۰/۱۷ و ۰/۰۷	-۲/۱۴	۰/۰۳۲	-۰/۲۸ (۰/۰۸)	-	-۰/۴۵ و ۰/۱۱
فعالیت بدنی	-۰/۰۰۰۲ (۰/۰۰۰۱)	-	-۰/۰۰۰۴ و ۰/۰۰۰۱	-۱/۷۷	۰/۰۴۸	-۰/۰۰۰۳ (۰/۰۰۰۱)	-	-۰/۰۰۰۵ و ۰/۰۰۰۱



تصویر شماره ۱: ساختار نمونه گیری

بحث

به عنوان مناسب‌ترین مدل معرفی کرد ولی از میان دو مدل مذکور، مدل لجستیک شرطی تا حدودی برازش بهتری داشته و می‌توان برای تحلیل چنین داده‌هایی از آن بهره برد.

مطالعات دیگر نیز در چنین مواقعی مدل شرطی را پیشنهاد کرده‌اند و نشان داده‌اند که مدل حاشیه‌ای مناسب چنین داده‌هایی نیست. بعنوان نمونه در مطالعه Wang و Devereux در سال ۲۰۰۶ بر داده‌های فشارخون، مدل‌های خطی تعمیم یافته و مدل‌های آمیخته نرمال مقایسه شد که در هر دو مدل متغیرهای سن، دیابتی بودن فرد، اضافه وزن و چاقی و الکلی بودن بر پرفشاری خون موثر بودند ولی ضرایب دو مدل با هم متفاوت بود (۲۳). در مطالعات انجام شده انواع دیگری از مدل‌های شرطی و حاشیه‌ای بر روی داده‌های مشابه، برازش شده است ولی پیشنهاد شده در صورتی که داده‌ها از نوع چند سطحی و خوشه‌ای باشند از مدل‌های شرطی استفاده گردد (۲۹-۲۴).

هدف این مطالعه اهمیت توجه به ساختار داده‌ها (داده‌های گرد آوری شده در طول زمان مانند داده‌های فشارخون) و برازش مدل مناسب جهت چنین داده‌هایی بود تا همبستگی میان مشاهدات در آن لحاظ شود. با توجه به ساختار داده‌ای، دو مدل لجستیک حاشیه‌ای با اندازه‌گیری مکرر و لجستیک شرطی با خوشه‌های تصادفی افراد برازش داده شد و برای مقایسه مدل‌ها از معیارهای نیکویی برازش استفاده شد. نتایج نشان دادند که مقادیر برآورد ضرایب در مدل لجستیک حاشیه‌ای با GEE و مدل لجستیک شرطی با روش تقریب لاپلاس به هم نزدیک هستند. خطای استاندارد برآورد پارامترها در مدل لجستیک شرطی نسبت به مدل لجستیک حاشیه‌ای با GEE بیشتر برآورد شده است. عبارتی در مدل GEE ضرایب دچار کم برآوردی است. همچنین مقادیر آماره‌های نیکویی برازش برای مدل لجستیک شرطی کمتر از مدل لجستیک حاشیه‌ای بود. نتیجه آن که اگرچه در پایان نمی‌توان با قاطعیت مدل مشخصی را

References

1. Diggle P, Liang K, Zeger S. Analysis of Longitudinal Data. 2nd ed. New York: Oxford University Press, 2002.
2. Dobson A. An Introduction to generalized linear Models. 2nd ed. London: Chapman & Hall/CRC, 2001.
3. Kevin H, Kim N. General Linear Models: Theory and applications using SAS software. 2nd ed. London: Chapman & Hall/CRC, 2007.
4. Charles E, McCulloch C, Seare S. Generalized Linear and mixed models. 1st ed. London: Wiley Interscience Publication; 2001. P 220-246.
5. Brady W, Kathleen B, Andrzej T. Linear mixed models: A practical guide using statistical software. 1st ed, London: Chapman & Hall/CRC; 2007. P 9-13.
6. Muller K. Introduction to mixed modeling Univariate, Multivariate and mixed models. 1st ed. London: Wiley Interscience Publications, 2006.
7. Jiang J. Linear and Generalized Linear Mixed Models and Their Applications. 1st ed. New York: Springer Science; 2007. P 119-122.
8. Breslow NE, Clayton DG. Approximate Inference in generalized linear mixed models. J Am Stat Assoc 1993; 88(421): 9-25.
9. Chobanian AV, Bakris GL, Black HR, Cushman WC, Green LA, Izzo JL, et al. The

- Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation and Treatment of High Blood Pressure: The JNC7 report. JAMA 2003; 289(19): 2560-2572.
10. Kriby M. NICE and Hypertension. Sage 2006; 6(4): 177-179.
 11. Cohen JD. Hypertension Epidemiology and Economic Burden: Refining Risk Assessment To Lower Costs. Manag Care 2009; 18(10): 51-58.
 12. Karimloo M. Application of General Linear Models in Blood pressure. Tehran Univ Med Sci 1991; 12(3): 19-27 (Persian).
 13. Amirkhizi F, Siyasi F, Minaei S, Jalali M, Dorosti A, Chamri M. Blood pressure status of rural women in Kerman and its relationship with anthropometric indexes. Lorestan Univ Med Sci 2008; 10(2): 31-38 (Persian).
 14. Sadeghi M, Roohafza H, Sadry Gh, Bahonar A, Saeidi M, Asgari S, et al. Prevalence of high blood pressure and its relation with cardiovascular risk factors. Qazvin Univ Med Sci 2003; 26(3): 46-52 (Persian).
 15. Azizi A, Abasi M, Abdoli Gh. The prevalence of Hypertension and its Association with age, sex and BMI in a population being educated using community based medicine in Kermanshah. IJEM 2008; 10(4): 323-329 (Persian).
 16. Ahmadi A, Hasanzade J, Rajaei fard A. Risk Factors of Hypertension in Koohrang city of Charmahal o bakhtiyari in 2007. IREA 2008; 4(2): 19-25 (Persian).
 17. Goodarzi M, Badakhsh M, Masinaei N, Abbas zade M. Hypertension prevalence in over 18 years old. JIUMS 2004; 11(43): 821-828 (Persian).
 18. Jiang J. Goodness-of-fit tests for mixed model diagnostics. Ann Statist 2001; 29(4): 1137-1164.
 19. Brown H, Prescott R. Applied Mixed Models in Medicine. 2nd ed, London: John Wiley and Sons; 2006. P 33-59.
 20. Pinheiro JC, Chao EC. Efficient Laplacian and Adaptive Gaussian Quadrature Algorithms for Multilevel Generalized Linear Mixed Models. J Comput Graph Stat 2006; 15(1): 58-81.
 21. Brady T, West Kathleen B, Welch Andrzej T. Linear mixed models A practical guide using statistical software. Linear mixed models A Practical Guide Using Statistical Software. London: Chapman & Hall/CRC: Taylor & Francis Group, 2007.
 22. Little R, Milliken G. SAS/STAT 9.2 User's Guide, Chapter 38: The GLIMMIX Procedure. 1st ed. Vancouver: SAS Publishing, 2006.
 23. Wang W, Devereux R. A Longitudinal Study of Hypertension Risk Factors and Their Relation to Cardiovascular Disease. Hypertension 2006; 47(3): 403-409.
 24. Fouriaud C, Jacquint-Salord MC, Degoulet P, Aime F, Lang T, Laprugne J, et al. Influence of socio professional condition on blood pressure levels and hypertension control. Am J Epidemiol 1984; 120(1): 7286-7292.
 25. Suarez LM, Criqui H, Barrett-Connor E. Spouse Concordance for systolic and diastolic Blood pressure. Am J Epidemiol 1983; 118(3): 345-351.
 26. Lane D. Ethnic differences in Hypertension and blood pressure control in the UK. QJM 2001; 94(7): 391-396.
 27. Kurt R, Maurer C. Blood pressure levels for person ages 6 months-74 years, 1976-80.

- National Center for Health Statistics. DHHS 1983; 11(233): 10-69.
28. Yadav S, Boddula R, Genitta G, Bhatia V, Bansal B, Kongara S, et al. Prevalence and risk factors of pre-hypertension and hypertension in an affluent north Indian population. *Indian J Med Res* 2008; 128(6): 712-720.
29. Roca C, Suarez A, Ceballos J. Control of hypertension in patients at high risk of cardiovascular disease. *QJM* 2005; 98(8) 581-588.