تشریح مدل‌بندی دوسطحی در داده‌های طولی فشار خون

محمد غلامی، قاسم یادگارفر
گروه آمار زیستی، دانشگاه تربیت مدرس تهران
گروه آمار و اپیدمیولوژی، دانشگاه علوم پزشکی اصفهان

روش تحلیل چندسطحی روشی کارا در تحلیل داده‌های همبسته و طولی می‌باشد. در این مقاله سعی داریم تا علاوه بر تشریح جنبه‌های نظری این تحلیل، به تجزیه و تحلیل عوامل موثر بر فشار خون در یک مطالعه‌ی طولی، بصورت گذشته‌نگر و با استفاده از پرونده های پزشکی پرسنل شاغل در کارخانه پلی اکریل اصفهان در طی سال‌های ۷۰ تا ۸۰ بپردازیم.

واژه‌های کلیدی: تحلیل چندسطحی، مطالعه همگروهی گذشته نگر، کارخانه پلی اکریل، فشار خون دیاستولیک.

۱ مقدمه

در بسیاری از مطالعات انسانی و علوم بیولوژیکی ساختار خوشه ای یا سلسله مراتبی قابل مشاهده می‌باشد. بطور مثال، فرزندان یک خانواده از لحاظ مشخصات فیزیکی و روحی، دارای شباهت بیشتری نسبت به دیگر افراد جامعه می‌باشند.(در این مثال فرزندان به عنوان سطح یک و خانواده‌ها به عنوان سطح ۲ در نظر گرفته می‌شوند) از جنبه آماری شباهت افراد به یکدیگر موید مستقل نبودن داده‌ها می‌باشد. دو نفری که متعلق به یک پدر و مادر هستند از لحاظ بسیاری از شرایط مانند یکدیگرند[۱]. برای مثال اگر فرزند یک خانواده دارای بیماری خاصی مانند هپاتیت باشد امکان داشتن هپاتیت در برادر و خواهر وی افزایش می‌یابد. و یا در مطالعات اندازه های تکرار شده بر روی متغیرهای اثر گذار بر فشار خون، فشار خون نمونه‌ها در زمانهای مختلف اندازه گرفته می‌شود. مسلما فشار خون هر فرد بخاطر وجود شرایط ذاتیش دارای تغییرات مخصوص به خود در طول زمان می‌باشد. در این مثال مجموعه اندازه‌های گرفته شده در زمان‌های مختلف برای هر یک از اعضای نمونه واحدهای سطح اول می‌باشند که برای هر واحدی از سطح دوم (انسان‌ها) تعریف می‌شوند. اما در توجیه اینکه چرا اندازه‌های تکراری در مطالعات طولی به عنوان سطح یک در نظر گرفته می‌شوند، می‌توان به همبستگی

Archive of SID
www.SID.ir

مشاهدات تکراری به خاطر تعلق خاص به یک فرد را عنوان نمود. همبستگی‌هایی از این نوع در طرح‌های مقطعی در ساختارهای آشیانه ای دیده می‌شوند. آنالیز جند سطحی معمولا برای تجزیه و تحلیل داده‌های اجتماعی و زیستی که در بیش از یک سطح قرار گرفته‌اند، به کار می‌رود و در واقع حالت بسط داده شده از مدل‌های خطی تعمیم‌یافته می‌باشد که در آن علاوه بر مدل بندی متغیر پاسخ ضرایب رگرسیونی نیز مدل‌بندی می‌شود. این روش، روشی کارا در مدل‌سازی داده‌هایی با ساختار آشیانه‌ای می‌باشد و هدف آن مدل‌بندی متغیر وابسته بر اساس تابعی از متغیرهای پیشگو(مستقل) در بیشتر از یک سطح است[۲]. آنالیز چندسطحی با اسامی دیگری نظیر مدل‌های خطی آشیانه‌ای، مدل‌های با ضرایب تصادفی و مدل‌های با اثرات تصادفی نیز شناخته می‌شود. از مزایای روش تحلیل چندسطحی به طرح اندازه مکرر در تحلیل داده‌های طولی می‌توان به قدرت تحمل داده‌های گمشده در برازش مدل بدون نیاز به برآورد و یا از دست دادن داده ها، کنترل اثرات محیطی و ارائه مدل جمعی و فردی برای افراد مورد مطالعه اشاره نمود. اما یکی از اصلاحات مهم در تحلیل چندسطحی، تعریف سطح می‌باشد. سطح جزئی از داده‌های آشیانه‌ای می‌باشد. سطح یک متغیر توسط نحوه جمع آوری داده‌ها مشخص می‌شود. و در واقع سطح، مقطع جمع آوری داده‌ها می‌باشد.[۱]

۲ مدل دو سطحی

همانگونه که در تعریف تحلیل چند سطحی بیان نمودیم، این روش حالت بسط داده شده‌ای از مدل‌های خطی تعمیم‌یافته است که در آن علاوه بر مدل‌بندی متغیر پاسخ، ضرایب رگرسیونی نیز مدل‌بندی می‌شوند. حالت نمادین این تعریف را می‌توانید در قالب معادله ۱ برای متغیر پاسخ Y با K متغیر کمکی $X_1, \cdots, X_k$ و همچنین n متغیر کمکی $W_1, \cdots, W_n$ جهت توضیح ضرایب سطح یک مشاهده نمایید. معادله ۲ نیز فرضیات این مدل را در حالت کلی بیان می‌کند.

level 1:

$$Y_{ij} = \beta_{0j} + \beta_{1j}X_{1ij} + \beta_{2j}X_{2ij} + \cdots + \beta_{kj}X_{kij} + e_{ij} \tag{1}$$

Level 2:

$$\begin{aligned}\beta_{0j} &= \gamma_{00} + \gamma_{01}W_{1j} + \gamma_{02}W_{2j} + \cdots + \gamma_{0n}W_{nj} + u_{0j}\\ \beta_{1j} &= \gamma_{10} + \gamma_{11}W_{1j} + \gamma_{12}W_{2j} + \cdots + \gamma_{0n}W_{1j} + u_{1j}\end{aligned}$$

Archive of SID

www.SID.ir

$$\vdots$$

$$\beta_{kj} = \gamma_{k0} + \gamma_{k1}W_{1j} + \gamma_{k2}W_{2j} + \cdots + \gamma_{kn}W_{nj} + u_{0j}$$

$$U = \begin{bmatrix} u_{0j} \\ u_{1j} \\ \vdots \\ u_{kj} \end{bmatrix} \approx N(0,G)\ ,\ G = \begin{bmatrix} \sigma^2_{u0} & \sigma_{01} & \cdots & \sigma_{0k} \\ \sigma_{01} & \sigma^2_{u1} & \cdots & \sigma_{1k} \\ \vdots & \vdots & \ldots & \vdots \\ \sigma_{0k} & \sigma_{1k} & \cdots & \sigma^2_{uk} \end{bmatrix} \tag{2}$$

$$e_{0ij} \approx N(0,R)\ ,\ Cov(\underline{U},\underline{e}) = 0$$

همانگونه که مشخص است، سطح یک مدل (level 1) شبیه نوعی از مدل رگرسیون چندگانه، و قسمت دوم آن (Level 2) نشان دهنده نحوه ارتباط ضرایب رگرسیونی با متغیرهای سطح دو می‌باشد. و این یعنی اینکه شیب‌ها و عرض از مبدها خود تابعی از متغیر سطح دوم یعنی W می‌باشد. همانگونه که در معادله۱ مشخص است در این روش، علاوه بر منبع خطا e، منبع $u_{\circ}, u_{1}, \cdots, u_k$در مدل موجود می‌باشد. بیان به فرم معادلات ۱ و۲ علاوه بر نمایش همه متغیرهای مستقل و وابسته ماهیت چندسطحی مدل را نیز نمایش می دهد. به جای استفاده از چنین سیستم معادلاتی، می‌توانیم بخش‌های سطح دوی مدل را در داخل سطح یک جایگذاری کرده و پس از مرتب‌سازی به معادله‌ای به فرم۳ دست یابیم.

$$Y_{ij} = \underbrace{\sum_{x=1}^{k}\sum_{y=\circ}^{n}\gamma_{xy}X_{xij}}_{\text{قسمت ثابت}} + \underbrace{\sum_{x=1}^{k}u_{xj}X_{xij} + u_{\circ j} + e_{ij}}_{\text{قسمت تصادفی}} \tag{۳}$$

ارائه مدل آماری به فرم 3 دارای مزایایی نسبت به فرم 1 می‌باشد. که از آن جمله می‌توان به فشرده‌تر بودن نحوه ارائه مدل، تعیین ساده‌تر اثرات ساده و تصادفی و نیز نزدیک بودن این مدل به خروجی نرم‌افزارهای متداول آماری اشاره نمود. اما علی رقم مزایای بیان شده تشخیص و تفسیر سطوح مدل در بیان به این فرم بسیار مشکل می‌باشد. در حالت کلی می‌توان معادله 3 را بصورت ماتریسی و به فرم معادله 4 نوشت.

$$y = X\alpha + Z\beta + e \tag{۴}$$

که در آن ماتریس X مشاهدات مستقل برای اثرات ثابت، Z ماتریس مشاهدات مستقل برای اثرات تصادفی، α بردار ضرایب ثابت مدل، β بردار ضرایب تصادفی مدل و در نهایت e خطای مدل می‌باشد.

Archive of SID
www.SID.ir

۳ روش‌های برازش مدل

در مدل‌های چند سطحی ما علاقه مند به برآورد سه نوع پارامتر: ضرایب ثابت، مولفه‌های تصادفی و همچنین پارامترهای پراکندگی (واریانس و کواریانس) مدل هستیم. باید توجه نمود که ضرایب تصادفی ضرایبی پنهان می‌باشند. یعنی مولفه‌هایی غیر قابل مشاهده ولی اثر گذار بر سیستم. خوشبختانه علی رغم غیرقابل مشاهده بودن، این ضرایب قابل برآورد می‌باشند. از روش های برآورد می‌توانیم به روش درستنمایی ماکزیمم (ML)، روش باقیمانده درستنمایی ماکزیمم[1] ($REML$) روش برآورد تکراری حداقل مربعات تعمیم یافته[2] ($IGLS$)، روش برآورد تکراری حداقل مربعات تعمیم یافته محدود شده[3] ($RIGLS$)، روش‌های بیزی، روش معادلات برآورد یابی تعمیم یافته[4] (GEE) و روش‌های مونت کارلویی ($MCMC$) اشاره نماییم. روش باقیمانده درست‌نمایی ماکزیمم که در برخی از متون به آن روش درستنمایی مقید[5] نیز گفته می‌شود در واقع همان روش ML می‌باشد که در آن تابع درستنمایی اصلاح شده است. از مزایای روش REML به روش MLE می‌توان به دو نکته عدم وابستگی مولفه‌های واریانس مدل به برآورد اثرات ثابت مدل و دخیل شدن درجه آزادی برآورد اثرات ثابت مدل در برآورد مولفه‌های واریانس اشاره نمود([۴]و[۵]). در روش IGLS که در بخش ۱.۳ به آن خواهیم پرداخت. برآورد اثرات ثابت و تصادفی مدل بر اساس یک روش تکراری محاسبه می‌گردند. شایان ذکر است که نتایج به دست آمده از این روش مشابه روش ML می‌باشد. نحوه محاسبه روش RIGLS نیز مانند روش IGLS می‌باشد. با این تفاوت که در محاسبه مولفه ماتریس واریانس-کواریانس مدل به جای استفاده از معادله ۷ از معادله $(y - X\hat{\alpha})(y - X\hat{\alpha})' = V - X(X'V^{-1}X)^{-1}X'$ بهره برده می‌شود. در این روش مولفه $X(X'V^{-1}X)^{-1}X'$ در واقع واریانس ضرایب ثابت مدل می‌باشد. که از مجموع ماتریس واریانس کل کم شده و باعث نااریبی برآوردهای واریانس مدل می‌شود. از این رو برآوردهای محاسبه شده در روش RIGLS معادل با روش REML می‌باشد.[۱]
از دیگر روش های برآورد می توانیم به روش های مدل خطی بیزی اشاره نماییم . در این روش برای هر یک از ضرایب β_j یک توزیع پیشین با واریانس σ_u^2 در نظر گرفته شده و برآوردها بر اساس آن محاسبه می‌شود. در روش برآوردهای بیزی کامل برای تک تک

1Residual Maximum Likelihood Estimation

2Iterative generalized least squares

3Restricted iterative generalized least squares

4Generalized Estimating Equations

5Restricted Maximum Likelihood Estimation

Archive of SID
www.SID.ir

ضرایب ثابت و تصادفی مانند β_j و σ_u^2 یک توزیع پیشین در نظر گرفته شده و برآوردیابی بر اساس آن انجام می‌شود. با قرار دادن فرض نرمال برای توزیع ضرایب ثابت و تصادفی، برآوردهای روش بیزی کاملا مشابه برآوردهای روش IGLS و RIGLS خواهد شد. برایکی و رادن‌باش [۶] نحوه استفاده از الگوریتم EM را برای محاسبه چنین برآوردهایی توضیح داده‌اند. همچنین از روش‌های برآورد می‌توانیم به روش معادلات برآوردی تعمیم یافته GEE اشاره نمود[۷]. تفاوت اصلی GEE ها با دیگر روش‌ها در برآورد ماتریس واریانس V با استفاده از رگرسیون عادی یا دستورالعمل گشتاوری بر پایه باقیمانده‌های خام می‌باشد. این روش برآورد بیشتر در فکر برآورد درست اثرات ثابت مدل است تا بررسی اثرات تصادفی مدل. بنابراین برآوردهای ارائه شده توسط این روش سازگار و نه کاملا کارآمد می‌باشند. از دیگر خصوصیات GEE، می‌توان به برآورد سریع و با مفروضات کمتر در مورد ماتریس واریانس V اشاره نمود. و در پایان از روش‌های مونت کارلویی خصوصا نمونه‌گیری گیبز برای حالاتی که حجم نمونه کم باشد، می‌توان در جهت دستیابی به برآوردهای دقیق استفاده نمود[۸]. نرم افزارهای HLM ،MLwiN (به عنوان نرم‌افزارهای تخصصی) و R، SPSS،SAS،SPLUS، LISREL، Stata و Mplus(به عنوان نرم افزارهای عمومی)قابلیت برازش مدل‌های چندسطحی با پاسخ نرمال را دارا می‌باشند، در ادامه به توضیح روش IGLS به عنوان روش برازش مدل نرم افزار MLwiN خواهیم پرداخت.

۳-۱ روش برآورد تکراری حداقل مربعات تعمیم یافته (IGLS)

همانگونه که گفته شد در روش IGLS برآورد اثرات ثابت و مولفه‌های واریانس مدل بر اساس یک روش تکراری محاسبه می‌شوند. طبق بخش ضمیمه، واریانس متغیر پاسخ برابر است با V که این واریانس خود به دو قسمت خطا (R) و تغییرات بین خوشه ای (ZGZ') تقسیم می‌شود. در صورتی که ضرایب ثابت مدل (α) برآورد گردیده و از متغیر پاسخ کم شده و سپس حاصلضرب خارجی آنها محاسبه شود. ماتریسی به دست می آید که امید ریاضی این ماتریس برابر با ماتریس V می‌باشد[۹]. باید به این نکته توجه نمود که $y - X\hat{\alpha}$ بردار خطای مدل نیست. زیرا بردار خطای مدل $y - X\hat{\alpha} - Z\hat{\beta}$ می‌باشد. در روش IGLS با مساوی قرار دادن ماتریس حاصلضرب خارجی $(y - \hat{\alpha})(y - X\hat{\alpha})'$ با ماتریس V و حل معادلات خطی توام، با استفاده از روش تکراری به محاسبه مولفه‌های ماتریس واریانس مدل پرداخته می‌شود [۹]. همانگونه که در بخش ۲ ضمیمه ثابت شده است برآورد درستنمایی ماکزیمم α برابر

Archive of SID
www.SID.ir

است با

$$\hat{\alpha} = (X'V^{-1}X)^{-1}X'V^{-1}y \qquad (5)$$

همانگونه که در معادله ۵ مشخص است برآورد ضرایب ثابت مدل خود تابعی از برآورد ماتریس واریانس-کواریانس V می‌باشد. در روش IGLS برای برآورد اولیه $\hat{\alpha}$ با فرض نبود هیچگونه مولفه واریانس-کواریانس $G = 0$ و تنها با وجود مولفه خطا، $\hat{\alpha}$ به صورت معادله ۶ برآورد می‌شود.

$$\hat{\alpha} = (X'X)^{-1}X'y \qquad (6)$$

سپس با مساوی قرار دادن ماتریس حاصلضرب خارجی $y - X\hat{\alpha}$ (معادله ۷) و حل معادله توام آن برآورد مولفه‌های ماتریس واریانس-کواریانس مدل V به دست می آید.

$$(y - X\hat{\alpha}(y - X\hat{\alpha})' = V \qquad (7)$$

پس از برآورد مولفه‌های ماتریس واریانس-کواریانس V با استفاده از دستگاه معادلات ۷ از این مرحله به بعد $\hat{\alpha}$ از معادله ۵ محاسبه می‌شود این بدین معنی است که پس از برآورد $\hat{\alpha}$ در معادله ۷ مولفه‌های ماتریس واریانس-کواریانس مدل از نو محاسبه شده و برآورد آن به جای ماتریس V جایگذاری می‌گردد.عمل محاسبه $\hat{\alpha}$ باز دوباره با استفاده از برآورد V جدید به دست می آید. عمل تکراری برآورد $\hat{\alpha}$ و $\hat{V}$ را تا جایی ادامه می‌دهیم که به برازش‌هایی با دقت مورد نظر خود دست یابیم.

۳-۲ برآورد اثرات تصادفی مدل

با توجه به اینکه برای برآورد اثرات تصادفی فرض نرمالیتی اعمال می‌شود و این برآوردها از طریق IGLS به طور مستقیم قابل برآورد نیستند با به دست آوردن پارامترهای واریانس و اثرات ثابت از طریق روش IGLS و با استفاده از فرمول‌هایی که برای برآورد اثرات تصادفی از روش MLE بخش سوم ضمیمه ثابت شده‌است برآورد اثرات تصادفی مدل عبارتست از

$$\hat{\beta} = GZ'V^{-1}(y - Xa)$$

و واریانس این برآورد گر برابر است با

$$var(\hat{\beta}) = GZV^{-1}ZG - GZ'V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}ZG$$

Archive of SID
www.SID.ir

۳-۳ روش های آزمون پارامترها و شاخص‌های کفایت مدل:

به منظور تست آماری ضرایب مدل از آزمون نسبت درستنمایی،آماره انحراف و آماره والد و به منظور انتخاب مدل از معیار اطلاع آکائیک(AIC)[6] که به صورت $AIC = -Ln\ L + K$ یا $AIC = -۲Ln\ L + ۲K$ و یا معیار اطلاع بیزی[7](BIC) که به صورت $BIC = -۲Ln\ L + p\ Ln(n)$ تعریف می‌شود استفاده می‌کنیم.(L تابع بزرگنمایی می‌باشد) هر مدلی که دارای AIC یا BIC کمتری باشد مدل مناسب‌تری برای برازش مدل بر روی داده‌ها می‌باشد [۱۰]. شاخص دیگری که در تحلیل داده‌های چندسطحی بسیار مناسب می‌باشد ICC[8] می‌باشد. این معیار شاخصی از همبسته بودن داده‌های درون خوشه‌ای می‌باشد و عبارتست از نسبت واریانس بین خوشه‌ای به کل واریانس موجود

$$ICC = \frac{\sigma_{u_۰}^۲}{\sigma_{u_۰}^۲ + \sigma_e^۲}$$

این شاخص درجه همبستگی داده‌ها را گزارش می‌نماید. هرچه مقدار عددی این شاخص به یک نزدیک‌تر باشد به معنای این است که مدل‌بندی چندسطحی مدل مناسبی برای تحلیل می‌باشد.

۴ به کارگیری مدل بندی دوسطحی برای تعیین عوامل موثر بر فشار خون :

پرفشاری خون یکی از بزرگترین مسائل سلامت هم در کشورهای توسعه یافته و هم در حال توسعه است. این مشکل شایع، بی علامت و در عین حال به سادگی قابل شناسایی و درمان پذیر است. که علی رغم پیشرفت روز افزون علم مکانیزم خطر زایی آن هنوز دقیقا روشن نشده‌است و در صورت عدم درمان سبب ایجاد عوارض کشنده می‌شود. از عوامل موثر بر فشار خون می‌توان به عواملی مانند سن، نژاد، ژنتیک ، جنسیت ، وزن تولد پایین و ... به عنوان عوامل غیر قابل اصلاح و همچنین چاقی، مصرف الکل، شرایط تغذیه‌ای،

[6] Akaike Information Criterion

[7] Bayesian Information Criterion

[8] Intra Class Correlation

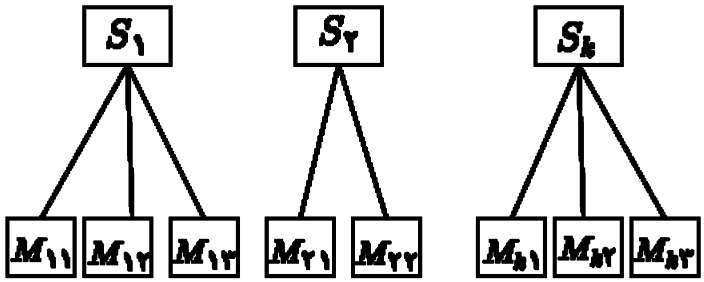
Archive of SID

www.SID.ir

فعالیت بدنی، شرایط محیطی نظیر درجه حرارت و ارتفاع، استرس‌های محیطی و عوامل روانی-اجتماعی، کار در نوبت کاری، مواجهه‌های مختلف شغلی نظیر مواجهه با سر و صدای زیاد، مواجهه با سرب و ... به عنوان عوامل قابل اصلاح اشاره نمود[۱۱]. داده‌های استفاده شده در این مطالعه با مراجعه به پرونده‌های پزشکی کارکنان شاغل در پلی اکریل اصفهان در طی سال‌های ۷۰ تا ۸۷ جمع آوری شده‌است. تعداد همگروه مورد نظر در این مطالعه ۱۰۱۷ نفر بود. معیار ورود یک فرد به مطالعه استخدام رسمی و یا غیر رسمی فرد در طی سالهای ۷۰ تا ۸۷ به شرط عدم ثبت بیماری خاص در پرونده پزشکی و معیار خروج آن استفاده از داروهای کاهش فشار خون در هر زمانی از مطالعه می‌باشد. متغیر پاسخ در این مطالعه فشار خون دیاستولیک و همچنین متغیرهای BMI، سن، جنسیت، نوبتکاری، وضعیت تأهل، میزان تحصیلات، نوع فعالیت، نوع استخدام به عنوان متغیرهای توضیح دهنده در نظر گرفته شده‌اند. تجزیه و تحلیل داده‌های این طرح به وسیله بسته‌های نرم افزاری MLwiN و SPSS انجام گشت. از ۱۰۱۷ نفر شرکت کننده در این مطالعه تعداد ۲۲ نفر(۲/۲%) زن و تعداد ۹۹۵ نفر (۹۷/۸%) مرد بودند. در این جمعیت تعداد ۵۷۳ نفر(۵۶/۳%) مجرد و تعداد ۴۴۳ نفر (۴۳/۶%) را افراد متاهل تشکیل می‌دادند. تعداد ۶۸ نفر (۶/۷%) کار اداری و تعداد ۹۴۹ نفر (۹۳/۳%) دارای شغل غیر اداری بودند. از افراد مورد مطالعه ۸۴۲ نفر (۸۲/۸%) بصورت قراردادی و ۱۷۵ (۱۷/۲%) بصورت رسمی استخدام شده بودند. میانگین تعداد فرزندان (۰/۰۰۹) ۰/۰۴۳۲ با فاصله اطمینان (۰/۰۶۰۳, ۰/۰۲۴۳) بود.توزیع سطح تحصیلات شرکت کنندگان در مطالعه عبارت بود از ۴۱ نفر (۴%) زیر دیپلم، ۴۷۶ نفر(۴۶/۸%) دیپلم و ۴۹۰ نفر (۴۸/۲%) تحصیلات فوق دیپلم به بالا افراد مورد مطالعه در سه گروه روزکار (۵۴%) ۵۰۸ نفر، دونوبتکار (۶%) ۵۸ نفر و سه نوبتکار (۴۰%) ۳۷۶ نفر دسته بندی شده بودند. میانگین تعداد تکرار مشاهده برای هر نفر (۰/۰۵%) ۴/۳۶ با فاصله اطمینان (۴/۴۶, ۴/۲۶) با میانه ۴ بود. همچنین پس از ساخت متغیر وضعیت ازدواج نتایج به این قرار که ۵۷۳ نفر (۶۴/۸%) مجرد، ۲۸۱ نفر (۳۱/۸%) متاهل با یک فرزند و نهایتا ۲۹ نفر (۳/۳%) متاهل با بیش از یک فرزند بودند. برای برازش مدل ، ابتدا مدل با عرض از مبدا تصادفی را برازش دادیم. سطح یک این برازش ، مشاهدات و سطح دو کارگران در نظر گرفته شد. نمودار گرافیکی این مدل را می‌توانید در شکل ۱ مشاهده نمایید .

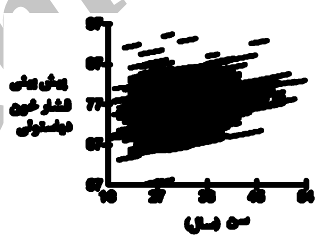
Archive of SID
www.SID.ir

شکل ۱: نمایش سطوح مطالعه در برازش مدل دو سطحی
S نماینده کارگران و M نمایش دهنده مشاهدات در طول زمان می‌باشد.

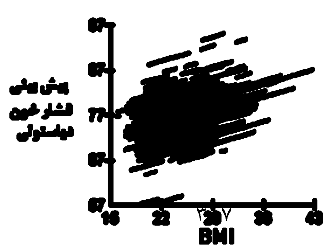


مقدار عددی ICC بدست آمده در این مطالعه برابر ۳۰٪ بود. همچنین فرض صفر بودن واریانس بین خوشه‌ای برای فشار خون دیاستولیک رد گردید ($p < ۰/۰۰۱$). این دو شاخص نشان دهنده این نکته است که همبستگی بین داده‌های فردی بسیار زیاد می‌باشد و این یعنی اینکه مدل بندی دوسطحی برازش خوبی را برای داده‌های موجود ارائه خواهد داد.

شکل ۲: نتایج حاصل ازبرازش مدل دوسطحی به داده های فشار خون



شکل ۳: رابطه BMI با فشار خون دیاستولیک برای هر نفر



Archive of SID
www.SID.ir

جدول ۱: ضرایب متغیرهای مستقل با فشار خون دیاستولیک

ضریب	برآورد	انحراف معیار	فاصله اطمینان بالا	فاصله اطمینان پایین	p-value
عرض از مبدا	۵۸/۶۸	۲/۱۸	۶۳/۱۵	۵۴/۲۱	<۰/۰۰۱
BMI	۰/۴۱	۰/۰۵	۰/۵۱	۰/۳۱	<۰/۰۰۱
سن-۲۵	۰/۲	۰/۰۴	۰/۲۸	۰/۱۲	<۰/۰۰۱
کار اداری(کارگری)*	-۱/۲۸	۰/۸۶	۰/۴۱	-۲/۹۷	۰/۱۴
رسمی(قراردادی)	-۰/۱۲	۰/۵۸	۰/۹۷	-۱/۲۶	۰/۷۷
زن(مرد)	۱۰/۲۵	۱/۲۴	۱۲/۶۸	۷/۸۲	<۰/۰۰۱
نوبتکاری					۰/۸۶
دو نوبتکار(روزکار)	۰/۰۳	۰/۷۸	۱/۵۶	-۱/۵	۰/۹۷
سه نوبتکار(روزکار)	۰/۲۴	۰/۴۴	۱/۱	-۰/۶۲	۰/۵۹
تحصیلات					۰/۰۶
دیپلم(زیر دیپلم)	-۱/۴۴	۰/۶	۰/۲۶	-۲/۶۲	۰/۱۱
فوق دیپلم(زیر دیپلم)	۰/۵۷	۰/۴۳	۱/۳۹	-۰/۲۵	۰/۱۷
وضعیت ازدواج					۰/۰۴
متاهل بدون فرزند(مجرد)	-۰/۶	۰/۴۱	-۰/۲	-۱/۴	۰/۰۳
متاهل با فرزند(مجرد)	-۱/۶۰	۱/۱۲	۰/۶۱	-۲/۱۱	۰/۰۹
$Var(u_{0j})$	۱۸/۲۶	۱/۳۹	۲۱/۲۲	۱۵/۳۹	<۰/۰۰۱
$Var(e_{ij})$	۳۹/۶۶	۱/۲۱	۵۶/۲۳	۳۴/۰۹	<۰/۰۰۱

*: سطوح مرجع انتخاب شده می‌باشند

۵ بحث و نتیجه گیری :

قبل از بحث در مورد نتایج باید به این نکته اشاره کرد که منبع کلیه مطالب گفته شده در این قسمت را می‌توانید در منبع [۱۱] این مقاله جستجو کنید. به علت زیاد بودن تعداد منابع از ذکر تک تک این منابع خودداری نموده و شما را به منبع ۱۱ این مقاله رجوع می‌دهیم. جدول متغیرهای موثر بر فشار خون دیاستولیک در جدول ۱ قابل مشاهده می‌باشد. همانگونه که این جدول نمایش می‌دهد متغیر BMI با فشار خون دیاستولیک رابطه نشان داد ($p < ۰/۰۰۱$). جهت این ارتباط مانند مطالعات گذشته در جهت مثبت بوده و با افزایش یک واحد بر شاخص BMI فشار خون دیاستولیک را ۰/۴۱ میلیمتر جیوه افزایش می‌داد. متغیر سن نیز مانند مطالعات گذشته با فشار خون دیاستولیک رابطه مثبت نشان داد ($p < ۰/۰۰۱$). با افزایش

www.SID.ir

یک سال بر سن افراد فشار خون دیاستولی ۰/۲ میلیمتر جیوه افزایش می‌یافت. نوبتکاری با فشار خون دیاستولی ($p = ۰/۸۷۶$) رابطه نشان نداد. در مطالعات پیشین روابط متناقضی در مورد رابطه با فشار خون و نوبتکاری گزارش شده‌است. که اکثر آنها در جهت تایید فرضیه افزایش فشار خون نوبتکارها به نسبت روزکارها بوده است. رابطه بین متغیر وضعیت ازدواج با فشار خون دیاستولیک ($p = ۰/۰۴$) پذیرفته شد. با متاهل شدن و سپس بچه دار شدن افراد فشار خون سیستولی مرتبا کاهش می یافت. البته مطالعات مختلف نیز نتایج متناقضی در مورد اثر وضعیت تاهل با فشار خون نشان داده اند. البته نکته مهمی که ممکن است رابطه مشاهده شده در مورد تغییر وضعیت تاهل را در این مطالعه مخدوش کند عامل سن است زیرا با افزایش سن، افراد متاهل و سپس فرزند دار می‌شوند. بنابراین امکان دارد که رابطه مشاهده شده ناشی از سن باشد و نه وضعیت تاهل. البته از آنجایی که سن در این مطالعه تغییرات زیادی را دارا نبود احتمال مخدوش شدگی رابطه مشاهده شده با سن بسیار ناچیز می‌باشد.

همچنین نتایج نشان دهنده عدم ارتباط نوع شغل با فشار خون دیاستولیک ($p = ۰/۱۴$) بود. نتایج همچنین نشان دهنده کاهش فشار خون کارگران نسبت به کارمندان اداری بود. در مطاعات مختلف فشار شغلی از عوامل موثر بر فشار خون بوده است. البته در مشاهده رابطه نوع شغل و فشار خون باید به این نکته نیز توجه داشت که کارگران بخاطر انجام کارهای فیزیکی عموما افرادی لاغرتر هستند. ضمن اینکه افرادی کارگر می‌شوند که از لحاظ پزشکی ا فرد سالم‌تری باشند.

نوع استخدام نیز با فشار خون و دیاستولیک ($p = ۰/۷۷$) رابطه نشان نداد. ذکر این نکته خالی از لطف نیست که افراد شاغل رسمی دارای سن بیشتر نسبت به افراد با استخدام قراردادی بودند.

متغییر تحصیلات نیز با فشار خون دیاستولیک ($p = ۰/۰۶$) رابطه نشان نداد. در اکثر مطالعاتی که بین سطح تحصیلات و فشار خون رابطه مشاهده شده بود، این رابطه یک رابطه کاهشی بود یعنی اینکه با افزایش تحصیلات ا فراد فشار خون کاهش می‌یافت. همچنین نتایج نشان داد که مردها فشار خون دیاستولیک ($p < ۰/۰۰۱$) بیشتری نسبت به زنان دارا می‌باشند. این ارتباط می‌تواند ناشی از آرامش بیشتر مردان نسبت به زنان باشد. این افزایش فشار خون برای فشار خون دیاستویک ۱۰/۲۵ میلیمتر جیوه بود. این ارتباط با مطالعات پیشین نیز هماهنگی داشت [۱۱].

Archive of SID

www.SID.ir

۶ ضمیمه:

بخش اول: محاسبه امید و واریانس متغیر تصادفی Y

$$E(y) = E(Xa + Z\beta + e) = X\alpha \tag{۸}$$

$$\begin{aligned} V(y) = var(X\alpha + Z\beta + e) &= var(X\alpha) + var(Z\beta) + var(e) \\ &= Z\ var(\beta)Z' + var(e) \Longrightarrow V(y) = ZGZ' + R \end{aligned}$$

بخش دو: محاسبه ضرایب ثابت مدل (α)

همانگونه که می‌دانید تابع درستنمایی نرمال چند متغیره عبارتست از

$$L = \frac{\exp[-\frac{۱}{۲}(Y - X\alpha)'V^{-۱}(Y - X\alpha)}{(۲\pi)^{(\frac{۱}{۲})^n}|V|^{\frac{۱}{۲}}} \tag{۹}$$

با لگاریتم گرفتن از معادله بالا داریم

$$\log(L) = K - \frac{۱}{۲}[\log|V| + (Y - X\alpha)'V^{-۱}(Y - X\alpha)]$$

با گرفتن مشتق ماتریسی از معادله‌بالا داریم

$$\begin{aligned} \frac{\partial}{\partial\alpha}\log(L) &= \frac{\partial}{\partial\alpha}(K - \frac{۱}{۲}[\log|V| + (Y - X\alpha)'V^{-۱}(Y - X\alpha)]) \\ &= \frac{\partial(Y - X\alpha)'V^{-۱}(Y - X\alpha)}{\partial\alpha} = -۲X'V^{-۱}(Y - Xa) \end{aligned}$$

حال با مساوی قرار دادن معادله بالا با صفر داریم

$$\hat{\alpha} = (X'V^{-۱}X)^{-۱}X'V^{-۱}Y$$

Archive of SID
www.SID.ir

بخش سوم: محاسبه مولفه‌های تصادفی مدل

در ابتدا اندیس گذاری زیر را در نظر بگیرید.

γ_R : پارامترهای واریانس ماتریس R
γ_G : پارامترهای واریانس ماتریس G
γ : پارامترهای واریانس ماتریس V

بر اساس اندیس گذاری بالا می‌توانیم تابع درستنمایی را به صورت زیر تعریف کنیم:

$$L(\alpha,\beta,\gamma,y)=L(\alpha,\gamma_R,\gamma|\beta)L(\gamma_G,\beta) \tag{۱۰}$$

با داشتن توزیع نرمال شرطی می‌توانیم معادله بالا را به صورت زیر بازنویسی کنیم

$$\begin{aligned} L(\alpha,\beta,\gamma,y) \quad &\propto \quad |R|^{-۱/۲}\exp[-\frac{۱}{۲}(y-X\alpha-Z\beta)'R^{-۱}(y-Xa-Z\beta)] \\ &\times|G|^{-۱/۲}\exp(-۱/۲\beta'G^{-۱}\beta) \end{aligned}$$

با لگاریتم گیری از معادله قبل داریم

$$\log(L)=-\frac{۱}{۲}[\log|R|+(y-Xa-Z\beta)'R^{-۱}(y-Xa-Z\beta)]+\log|G|+\beta'G^{-۱}\beta]+K$$

که در آن K همه مقادیر ثابتی است که در ماکزیمم نمودن تابع درستنمایی نقشی ایفا نمی‌کند با مشتق گیری از β در معادله بالا داریم.

$$\partial\log(L)/\partial\beta=Z'R^{-۱}(y-Xa-Z\beta)-G^{-۱}\beta=-(Z'R^{-۱}Z+G^{-۱})\beta+Z'R^{-۱}(y-Xa)$$

با مساوی قرار دادن این معادله با صفر داریم

$$(Z'R^{-۱}Z+G^{-۱})\hat{\beta}=Z'R^{-۱}(y-Xa)$$

$$\hat{\beta}=(Z'R^{-۱}Z+G^{-۱})^{-۱}Z'R^{-۱}(y-Xa)$$

با دانستن این نکته که $V=ZGZ'+R$می‌توانیم معادله قبل را به فرم دیگر، و به شکل زیر بنویسیم.

$$\hat{\beta}=GZ'V^{-۱}(y-Xa)$$

Archive of SID
www.SID.ir

تشکر و قدردانی

بدین وسیله از بنیاد ملی نخبگان به خاطر حمایت‌های بی دریغ این نهاد خدمتگذار و نخبه پرور از اینجانب و همچنین آقای دکتر صنعتی بخاطر زحمات بی اندازه شان کمال تشکر و قدردانی را دارم.

مراجع

[1] Goldstein, H. Multilevel Models in Educational and Social Research ,London 1987, Griffin.

[2] Hox, J. Multilevel Analysis: Techniques and Applications. Mahwah 2002., N.J.: Lawrence Erlbaum Associates

[3] Douglas A.Luke,Multilevel modeling.A sage university paper series 2004 .Quanititive applications in social sciences;143

[4] McCulloh,C.Searel,S.Generalized,Liner,and Mixed Models,Wiley 2001,New York.

[5] Brown,H . Prescott,R(2006). Applied Mixed Model in Medicine , Second Edition , John Wiley & Sons,Ltd ,UK

[6] Bryk, A.S. and Raudenbush,S.W. :Hierarchical linear models. Newbury Park: 1992 Sage.

[7] Liang, K. and Zeger, S.L: Longitudinal data analysis using generalized linear models. Biometrica 1986; 73:45-51

[8] Zeger, S.L. and Karim, M.R: Generalised linar models with random effects ; a Gibbs Sampling approach. Journal of the American Statistical Society 1991; 86:79-102

[9] Brown,H . Prescott,R. Applied Mixed Model in Medicine , Second Edition , John Wiley & Sons,Ltd , 2006;UK

[10] Myers R.H., Montgomery D.C., Vining G.G., Generalized Linear Models with Application in Engineering and Sciences, John Wiley &

Archive of SID

www.SID.ir

Sons, Inc. 2002, New York.

[۱۱] غلامی، م. کاربرد روش تحلیل چند سطحی در بررسی برآورد کننده‌های تغییرات طولی فشار خون در یک کوهورت گذشته‌نگر شاغل در پلی اکریل اصفهان از سال ۷۰ تا ۸۷، ۱۳۸۸ رساله کارشناسی ارشد. دانشکده علوم پزشکی اصفهان، دانشکده بهداشت.

Archive of SID

www.SID.ir