Archive of SID

Jundishapur Education Development Journal
Quarterly Journal of Medical Sciences Education and Development Center
Year 13, No. 1, 2022 Spring Season

Teaching Multivariate Linear mixed model in Multiple Longitudinal Data to the Medical Students: A Case Study of Child Development Data

Maryam Moradi: Department of Statistics, Science and Research Branch, Islamic Azad University, Tehran, Iran.
Kambiz Ahmadi Angali*: Department of Biostatistics, Faculty of Health, Ahvaz Jundishapur University of Medical Sciences, Ahvaz, Iran.
Mohammad Hassan Behzadi: Department of Statistics, Science and Research Branch, Islamic Azad University, Tehran, Iran.
Rahman Farnoosh: Department of Mathematics, Iran University of Science and Technology, Tehran, Iran.

Abstract: The aim of this study was to teach medical students the multivariate linear mixed statistical model in multiple longitudinal data: A case study of child development data. The research method was descriptive-analytical and based on longitudinal data. The statistical population of the study consisted of 100 students who were randomly selected and participated in two training workshops based on statistical modeling of t and normal mixed multivariate linear distributions based on children's growth data (height, weight and head circumference). After the training course, they participated in practical test. Mean and standard deviation and statistical modeling of multivariate mixed linear t-distribution were used to analyze the data. The results showed that the amount of height variable parameters according to MtLMM and MnLMM models for breastfed infants was significantly higher than formula ($P < 0.05$). Also, the estimation of weight variable parameters for infants who used formula was significantly higher ($P < 0.05$) compared to infants who consumed only formula. The trainings provided to the two groups led to a significant increase in students' learning ($P < 0.05$). The students stated that teaching a multivariate statistical model in longitudinal data revealed the important of using this model in medical research and life sciences.

Keywords: education, breastfeeding, mixed model, longitudinal data, medical sciences.

***Corresponding author:** Department of Biostatistics, Faculty of Health, Ahvaz Jundishapur University of Medical Sciences, Ahvaz, Iran.

Email: ahmadikambiz00@gmail.com

www.SID.ir

Archive of SID

مقدمه

به منظور بررسی اثر درمان‌ها در مطالعات پزشکی، متغیر پاسخ به طور مکرر در طول زمان اندازه‌گیری می‌شود که به آنها مطالعات طولی گفته می‌شود. در اغلب تحقیقات کاربردی، بخصوص در مباحث مربوط به علوم زیستی، پزشکی و مطالعات اقتصادی، عموماً به دلیل تفاوت‌های خاصی که بین واحدهای آزمایشی وجود دارد، تغییرپذیری پاسخ‌های مربوط به واحدها نیاز به مدلسازی ویژه‌ای دارد. درصد عمده‌ای از این تغییرپذیری غالباً به سبب وجود تفاوت‌هایی است که قبل و یا در حین آزمایش ایجاد شده است. اگر این منبع تغییرپذیری قابل تفکیک از خطا و اثرهای آزمایشی باشد آنگاه می‌توان دقت آزمایش‌ها را با برازش مدل‌های تحلیل واریانس افزایش داد و نیز توزیع‌هایی که بهتر برازنده داده‌ها هستند را انتخاب نمود (دیویس ۲۰۰۲). جمع آوری اطلاعات در مورد پیامدهای طولی متعدد به طور فزاینده ای در بسیاری از موارد بالینی متداول است. در بسیاری از موارد، مطلوب است که این نتایج را به طور مشترک مدل سازی کنیم. با این حال، در مجموعه داده های بزرگ، با پیامدهای زیاد، بار محاسباتی اغلب مانع از مدل سازی همزمان چندین نتیجه در یک مدل واحد می شود (هیوز و همکاران ۲۰۲۱). روش تحلیل این نوع از داده‌ها، تحلیل اندازه‌گیری تکراری است که تنها از یک نوع ساختار همبستگی استفاده می‌کند و در صورتی که این ساختار دارای همبستگی مناسب نباشد نتایج معتبر نیست. برای رفع این مشکل مدل‌های آمیخته جایگزین مناسبی است. همچنین ممکن است افراد شرکت‌کننده در مطالعه از یک جامعه ناهمگن آمده باشند. اگر داده های جمع‌آوری شده از یک مطالعه پزشکی موردنظر هر سه ویژگی فوق را همزمان داشته باشند، لازم است این موارد در مدل آماری لحاظ شود، مدل اثرات آمیخته خطی در داده‌های طولی چندمتغیره، مدلی است که این هدف را دنبال می‌کند (دیویس ۲۰۰۲). هدف از معرفی این مدل‌ها، ارائه تحلیل های پرکاربرد و مفید در مورد داده‌های مذکور می باشد که علاوه بر لحاظ نمودن وابستگی‌های موجود بین داده‌ها، نیازهایی از جمله برازش مدل را نیز ارائه می‌دهد.

مدل‌های آمیخته تاریخچه‌ای بسیار قدیمی داز قرن ۱۸ میلادی دارند. توزیع‌های آمیخته اولین بار صورت یک توزیع نرمال آمیخته متناهی دومؤلفه‌ای معرفی شد (پیرسون ۱۸۹۴). محققینی از جمله (وولف ۱۹۶۵) برآورد ماکسیمم درست نمایی توزیع‌های آمیخته را مورد مطالعه اولیه قرار دادند. سپس (دمپستر و دیگران ۱۹۷۷) الگوریتم EM را برای برآورد ماکسیمم درستنمایی پارامترها در مدل های آمیخته ارائه کردند (اورت و هاند۱۹۸۱) نظریه و کاربردهای توزیع‌های آمیخته و ویژگی‌های آنها را مورد بررسی قرار داده‌اند. با ظهور کامپیوترهای با سرعت بالا، توجه به سمت برآورد ML پارامترها در یک توزیع آمیخته بیشتر شد. پرکاربردترین مدل خطی با اثرات آمیخته برای پاسخ‌های پیوسته با اثرات تصادفی و خطاهای درون گروهی، توزیع‌های نرمال هستند که نسبت به داده‌های پرت حساس هستند. این داده‌های پرت برای مدل‌های اثرات آمیخته نسبت به مدل هایی با اثرات ثابت بیشترین مشکل را ایجاد می‌کنند، زیرا آنها در اثرات تصادفی، در خطاهای درون گروهی یا در هر دو رخ می‌دهند و برآورد مدل در عمل سخت خواهد شد (کرودر ۱۹۹۰). مدل‌های اثرات آمیخته غیرخطی چندمتغیره (MNLMM) (Multivariate linear mixed model) به دلیل انعطاف‌پذیری آنها برای تجزیه و تحلیل داده‌های طولی به طور افزایشی مورد استفاده قرار می‌گیرند. این مدل‌ها به دلیل انعطاف‌پذیری نسبت به انتخاب ساختارهای مختلف واریانس-کواریانس و ویژگی های مهم دیگر و عدم نیاز به مفروضات محدودکننده برای اندازه‌گیری تکراری مناسبند (وانگ و فان ۲۰۱۱). از آنجا که مدل های خطی آمیخته دامنه بسیار گسترده ای از داده ها را شامل می شوند ولی بسیاری از محققین و دانشجویان پزشکی به دلیل عدم آشنایی با این مدل ها، اثرات تصادفی را نادیده می گیرند که این امر موجب می شود برخی از پارامترها به اشتباه معنی دار شوند. استفاده

www.SID.ir

Archive of SID

درست از این مدل ها موجب می شود از بسیاری از این نتایج اشتباه جلوگیری شود. لذا با توجه به پرکاربرد بودن داده های طولی در علوم زیستی و پزشکی، مطالعه حاضر به معرفی و آموزش روش تجزیه و تحلیل آماری به دانشجویان پزشکی با تکیه بر داده های طولی رشد کودکان انجام شد. به کارگیری مدل آمیخته توزیع t - چندمتغیره برای داده‌های رشد کودکان نادر می‌باشد و هدف این است که نشان دهیم روش چندمتغیره t با یک الگوریتم تدوین شده به راحتی قابل استفاده در داده‌های علوم پزشکی می‌باشد.

روش کار

با توجه به اینکه هدف پژوهش حاضر آموزش مدل آماری آمیخته خطی چندمتغیره در داده های طولی چندگانه به دانشجویان پزشکی: مطالعه موردی داده های رشد کودکان بود، روش پژوهش بر حسب هدف توصیفی-تحلیلی و بر حسب نوع داده از نوع داده های طولی است. مدل آماری به کار گرفته شده در پژوهش مدل آمیخته خطی چندمتغیره توزیع تی و توزیع نرمال بود.

جهت رعایت اخلاق پژوهشی کد اخلاق IR.SRBIAU.REC.1398.05 در کمیته اخلاق دانشگاه ثبت شد. جامعه آماری پژوهش شامل ۱۰۰ دانشجوی پزشکی دانشگاه علوم پزشکی اهواز بود که براساس حجم نمونه $n = \frac{(Z^2_{1-\alpha/2} - Z^2_{1-\beta})S^2}{d^2}$ و سطح خطای پنج درصد ($Z_{1-\alpha/2} = 1.96$) و توان آزمون ۲۰درصد ($Z_{1-\beta} = 0.84$)، S انحراف معیار نمره یادگیری و d اختلاف میانگین نمره یادگیری در دو گروه و به صورت تصادفی انتخاب تا بر روی داده های رشد کودکان مراجعه کننده به مراکز بهداشتی شهر اهواز کار کنند. داده‌های رشد کودکان با اندازه‌گیری سری آنتروپومتریک وزن، قد و دور سر (۷۵۰ نفر) از بدو تولد تا ۱۲ماهگی، همچنین نوع شیردهی تغذیه با شیرمادر، شیرخشک یا ترکیبی از هر دو در نهایت ثبت شد. شرایط ورود به مطالعه شامل موارد زیر بود: وزن هنگام تولد بیش از ۲۵۰۰ گرم، سن حاملگی هنگام تولد بین ۳۷ تا ۴۲ هفته. کودکان در طی بازه زمانی شش مرتبه (هنگام تولد، دو، چهار، شش، نه و ۱۲ماه) مورد پایش و اندازه‌گیری شاخص‌های رشد قرار گرفتند. نوزادان براساس نوع تغذیه در سه دسته تقسیم‌بندی شدند. گروه اول تغذیه انحصاری با شیرمادر است که طبق گفته سازمان بهداشت جهانی، در شش ماه اول زندگی فقط از شیرمادر و مکمل ویتامین استفاده می‌کند (پی ایکس 2000) . دو گروه دیگر شامل تغذیه با شیرخشک و گروهی که هم با شیرمادر و هم با شیرخشک تغذیه می کنند. یک مدل طولی سه متغیره شامل قد، وزن و اندازه دورسر به عنوان یک متغیر وابسته با استفاده از مدل‌های آمیخته چندمتغیره به داده‌ها برازش داده شد. در مطالعات طولی معمولاً وجود داده-هایی با بیش از یک متغیر پاسخ و به طور مکرر در طول زمان است که منجر به داده‌های طولی چندمتغیره در یک موضوع خاص و تحت زمان‌های متوالی خاص می‌شود. مدل آمیخته خطی چندمتغیره (MLMM) محبوب‌ترین روش برای تجزیه‌وتحلیل مشترک بیش از یک سری داده-های طولی است. در داده‌های طولی غالباً همبستگی وجود دارد و این همبستگی را باید در مدل‌های آماری در نظر گرفت. بنابراین، داده‌های هر فرد در هر زمان که تحت تأثیر داده‌های زمان قبلی قرار گرفته و میزان متغیرهای پاسخ هر فرد در هر زمان نیز ممکن است با هم ارتباط داشته باشند. مدل اثرات آمیخته خطی در داده‌های طولی چندمتغیره مدلی است که این وابستگی را در نظر گرفته و برآورد پارامترها در این شرایط هدف این مدل می‌باشد. مدل چندمتغیره آمیخته خطی در این مطالعه براساس قد تولد (سانتی‌متر)، وزن هنگام تولد (کیلوگرم)، اندازه دور سر هنگام تولد (سانتی‌متر) و نوع تغذیه نوزاد تنظیم شد و تأثیر متغیرهای مستقل (داده‌های طولی) بر متغیر وابسته به طور همزمان بررسی شد. میزان یادگیری و تأثیر آموزش این روش آماری بر روی دانشجویان با استفاده از

www.SID.ir

Archive of SID

پرسشنامه انجام شد. شرایط ورود به مطالعه دانشجویان شامل تمایل به شرکت در مطالعه و تحصیل در ترم آخر پزشکی بود. در صورت عدم شرکت دانشجو در کارگاه آموزشی یا آزمون های دوره، از مطالعه حذف می شد. بعد از مشخص شدن افراد واجد شرایط، آنان در دو گروه جهت آموزش و کار بر روی داده های رشد کودکان تقسیم شدند. سپس هر گروه به صورت تصادفی ساده به ۲ گروه آموزش مدل آمیخته خطی توزیع تی چندمتغیره و مدل آمیخته خطی توزیع نرمال چندمتغیره تخصیص یافتند. در ابتدای مطالعه دانشجویان در پیش آزمون شرکت نمودند. طبق برنامه کارگاه آموزشی از روز بعد از آزمون شروع شد. طبق برنامه، کارگاه آموزشی از روز بعد از آزمون شروع شد. آموزش عملی این مطالعه یک هفته طول کشید و روز بعد از اتمام آموزش عملی، مجدداً یک پس آزمون عملی شبیه سازی شده اجرا شد و دانشجویان شرکت کننده در کارگاه در آن شرکت کردند. برای تجزیه و تحلیل پژوهش از روش های آمار توصیفی و استنباطی استفاده شد. در بخش توصیفی برای شاخص های میزان یادگیری از میانگین و انحراف معیار و در بخش استنباطی به برآورد پارامترهای رگرسیونی و تست های آماری t-test Paired ، تحلیل کوواریانس و در صورت توزیع غیر نرمال از آزمون ویلکاکسون استفاده شد و کار بر داده های رشد کودکان با مدل آمیخته براساس توزیع نرمال و توزیع تی و به تفکیک شیرمادر، شیرخشک و یا شیر ترکیبی انجام شد. تجزیه‌وتحلیل داده‌ها با استفاده از مدل آمیخته خطی چندمتغیره انجام شد. از روش همبستگی پیرسون و ازمون فرضیات آماری در سطح معناداری ۰/۰۵ بهره بردیم. تجزیه‌و تحلیل آماری با استفاده از نرم افزار Spss نسخه ۲۳ و نرم افزار R نسخه ۴،۰،۲ صورت پذیرفت.

روش آماری

در حالت کلی یک مدل خطی با اثرهای آمیخته به صورت زیر معرفی می‌شود:

$$\mathbf{Y}_{N\times 1} = \mathbf{X}_{N\times p}\boldsymbol{\beta}_{p\times 1} + \mathbf{Z}_{N\times q}\mathbf{u}_{q\times 1} + \mathbf{e}_{N\times 1},$$

که در آن Y بردار مقادیر مشاهده شده متغیر پاسخ، ماتریس پررتبه ستونی X شامل متغیرهای توضیحی (کمکی)، β بردار پارامترهای مجهول اثرهای ثابت، بردار u نمایانگر اثرهای تصادفی و ماتریس Z شامل برخی متغیرهای توضیحی و نشانگر است که با توجه به فرض های موجود درباره اثرهای تصادفی u در مدل ساخته می شوند. e نیز برداری از مؤلفه‌های خطای تصادفی مدل است که مستقل از u توزیع می‌گردد. استنباط آماری پارامترهای مدل براساس تابع درستنمایی حاشیه‌ای

$$L(\boldsymbol{\beta},\boldsymbol{\theta}\mid \mathbf{x},\mathbf{y}) = \int_{R^q} f(\mathbf{y}\mid \mathbf{x},\mathbf{u},\boldsymbol{\beta})g(\mathbf{u}\mid\boldsymbol{\theta})d\mathbf{u}$$

انجام می‌شود که معمولاً در آن تابع چگالی اثرهای تصادفی $g(u|\theta)$ و چگالی شرطی-پاسخ $f((y)|x,u,\beta)$ نرمال فرض می‌شوند.

برای داده های طولی مطالعه شامل وزن، قد و اندازه دور سر نوزادان هنگام تولد به ترتیب متغیرهای Y_{ij1}, Y_{ij2}, Y_{ij3} را تعریف می‌کنیم، وقتی که i=1,2,...,25 تعداد نمونه و j=1,...,6 تعداد ماه‌های نظارت بر رشد نوزاد است (هنگام تولد، دو، چهار، شش، نه و ۱۲ ماه).

تابع چگالی توزیع آمیخته چندمتغیره خطی (MLMM) به فرم تابع زیر تعریف می‌شود (وانگ ۲۰۱۱)

وقتی که $Y_i = [y_{i1} : \cdots : y_{ir}]$ یک ماتریس پاسخ $n_{i\times r}$ شامل r بردار قابل مشاهده $y_{ij} = (y_{ij1}, \ldots, y_{ijni})$ است.

برای توزیع تی آمیخته خطی (MtLMM) برای بردار پاسخ Y_i داریم (Wang W, 2011) :

$$y_i = X_i\beta + Z_i b_i + e_i \quad with \quad \begin{bmatrix} b_i \\ e_i \end{bmatrix} \sim t_{(q_2+n_i)r}\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix} D & 0 \\ 0 & R_i \end{bmatrix}, v_i\right), \qquad \textbf{(1-1)}$$

www.SID.ir

Archive of SID

وقتی کـه $(\beta_1^T, \ldots, \beta_r^T)^T \beta$ = پـارامتر رگرسـیونی است. q_1 -بردار که اثرات ثابت از j-امین برآمد را توصیف می‌کند و $(b_{i1}^T, \ldots, b_{i1}^T)^T b_i$ = که یک q_{2r}-بردار از اثرات تصادفی است.

یافته ها

این بخش نتایج تجزیه و تحلیل داده‌های واقعی که در اختیار دو گروه از دانشجویان جهت آمـوزش مـدل آمـاری توزیع های تی و نرمال آمیخته چنـدمتغیره براسـاس داده های طولی رشد کودکان شامل قد، وزن و انـدازه دور سـر در نتیجه نوع تغذیه مصرفی قـرار گرفتـه بـود را در مـدل نشان می‌دهد. ایـن نتـایج در ایـن بخـش جهـت مقایسـه کارایی و عملکرد مدل هـای آمیختـه و بـرازش بهتـر ایـن مدل ها در داده های طولی مورد استفاده در علوم پزشکی آورده می شود. و در نهایت در قسـمت بحـث بـه اهمیـت کاربرد این مدل ها پرداخته می شود.

جدول ۱ میانگین و انحراف معیار وزن، قد و اندازه دور سر نوزادان از بدو تولد تا ۱۲ ماهگی را همـراه بـا افـزایش مقدار نشان می دهد. هم چنین این جدول اطلاعات رشد کودک را با توجه به سه نوع تغذیه دریافتی دسـته بنـدی می‌کند. نتایج جدول ۱ نشان داد کـه میـانگین و انحـراف معیار قد و اندازه دور سر نوزادانی که از شـیر مـادر تغذیـه می کنند در مقایسه با دو گروه دیگر بیشـتر و میـانگین و انحراف معیار نوزادانی کـه از شـیر ترکیبـی اسـتفاده مـی کنند از گروه شیر مادر و شیرخشک نیز بیشتر است.

جدول ۱. میانگین و انحراف معیار وزن، قد و اندازه دور سر نوزادان

Table 1. Mean and standard deviation of weight, height and head circumference of infants

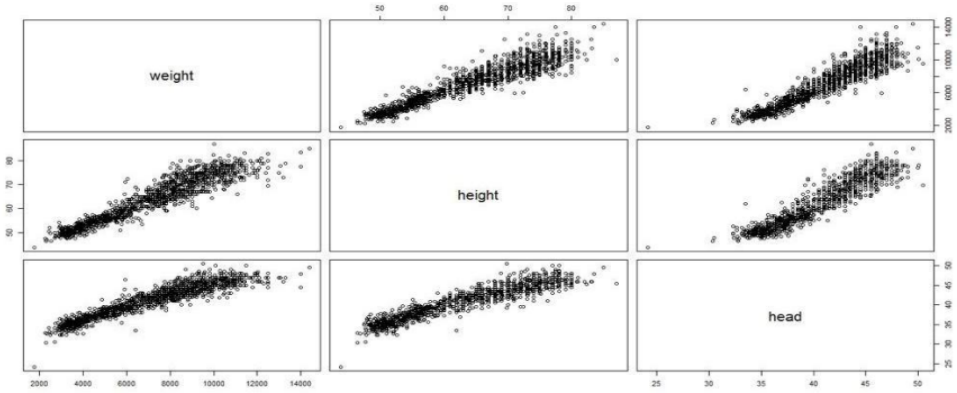
	بدو تولد		۲ماهگی		۴ماهگی		۶ماهگی		۸ماهگی		۱۲ماهگی	
	میانگین	انحراف معیار	میانگین	انحراف معیار	میانگین	انحراف معیار	میانگین	انحراف معیار	میانگین	انحراف معیار	میانگین	انحراف معیار
وزن (گرم)	3520.82	481.15	5156.24	675.05	7147.62	828.71	8236.01	972.04	9381.00	1187.80	10434.69	6014.67
قد (سانتیمتر)	50.85	1.85	56.27	2.50	63.21	2.71	67.21	2.74	72.45	2.81	76.00	2.68
دور سر (سانتی متر)	35.10	1.43	38.10	1.31	41.29	1.39	42.99	1.29	44.82	1.50	45.98	1.46

	شیرمادر	شیرخشک	ترکیبی
میانگین و انحراف معیار وزن (گرم)	3400 ± 0.4	3300 ± 0.5	3510 ± 0.5
میانگین و انحراف معیار قد(سانتیمتر)	51.2 ± 1.2	49.8 ± 2.1	50.1 ± 1.7
میانگین و انحراف معیار اندازه دور سر(سانتیمتر)	42.15 ± 1.8	41.29 ± 2.1	41.87 ± 1.1

www.SID.ir

Archive of SID

شکل ۱ برآورد پارامترهای قد، وزن و اندازه دور سر نوزادان از بدو تولد تا ۱۲ ماهگی را در سه گروه تغذیه¬ای نشان می¬دهد. طبق نمودارها، متوسط سرعت رشد تا ۱۲ ماه در تمام اندازه¬گیری¬ها: قد ، وزن و اندازه دور سر در گروه شیرمادر بهتر از گروه-های تغذیه شده با شیرخشک و ترکیبی بود. همچنین، تغذیه ترکیبی رشد بهتری نسبت به تغذیه با شیرخشک نشان داد.



شکل ۱. نمودار پراکنش ماتریس متغیرهای وزن، قد و اندازه دور سر نوزادان

Figure 1. Matrix distribution diagram of variables of weight, height and head circumference of infants

جدول ۲ برآورد پارامترهای مدل تحت MtLMM و MnLMM را برای متغیرهای وزن، قد و اندازه دور سر نوزادان نشان می‌دهد. مقدار پارامترهای متغیر قد برای نوزادانی که از شیرمادر تغذیه می‌کنند، به طور قابل توجهی بالاتر از شیرخشک بود (P = ۰/۰۰۰۲). همچنین برآورد پارامترهای متغیر وزن بر حسب مدل‌های MtLMM و MnLMM برای نوزادانی که از شیر ترکیبی استفاده می‌کنند در مقایسه با نوزادانی که فقط شیر خشک مصرف می‌کنند به طور قابل توجهی بالاتر (P = ۰۰۰۰) بود.

جدول ۲. برآورد پارامترهای مدل رگرسیون چندمتغیره براساس MtLMM و MNLMM

Table 2. Estimation of multivariate regression model parameters based on MtLMM and MNLMM

		پارامترها	MtLMM	p-value	MNLMM	p-value
قد	ثابت	β_{01}	63.40	0.0000	64.29	0.0000
	شیرمادر	β_{11}	-7.16	0.0002	0.107	0.9865
	شیرترکیبی	β_{21}	5.59	0.0947	-0.27	0.9680
	شیرخشک	Refrence Level				
وزن	ثابت	β_{02}	6915.89	0.0000	7262.73	0.0000
	شیرمادر	β_{12}	-1843.03	0.0000	115.60	0.0000
	شیرترکیبی	β_{22}	1745.85	0.0000	-223.97	0.0000
	شیرخشک	Refrence Level				
اندازه دور سر	ثابت	β_{03}	41.41	0.0000	41.45	0.0000
	شیرمادر	β_{13}	-1.49	0.5143	-0.01	0.9979
	شیرترکیبی	β_{23}	1.99	0.4343	-0.43	0.9506
	شیرخشک	Refrence Level				

www.SID.ir

Archive of SID

مقایسه میانگین نمره دانشجویان در انجام کار عملی در پیش آزمون و پس آزمون براساس آزمون t-student نشان داد که میانگین نمره پیش آزمون در بین دو گروه با (P = ۰/۵۳) اختلاف معناداری نداشت اما در پس آزمون بین دو گروه مطالعه با (P < 0.05) اختلاف آماری معناداری وجود داشت. با بکارگیری آزمون لون، با توجه به سطح معنی داری (۰,۰۹۳) برابری واریانس های دو گروه تایید می‌گردد. نتایج تحلیل کوواریانس بیان می کند که دوره های آموزشی بر متغیر یادگیری با اندازه اثر (۰,۹۱۷) تأثیرگذار است (جدول ۳).

جدول ۳. شاخص های توصیفی دانشجویان در حیطه میزان یادگیری

Table 3. Descriptive indicators of students in the field of learning

آزمون / گروه	پیش آزمون		پس آزمون		p-value	Leven's Test		اثرات بین گروهی
	میانگین و انحراف معیار	تعداد	میانگین و انحراف معیار	تعداد		F	p-value	اندازه اثر
گروه اول	12.05 ± 3.17	50	24.83 ± 3.21	50	0.003	1.823	0.093	0.917
گروه دوم	12.17 ± 2.94	50	23.51 ± 3.97	50	0.016			

بحث و نتیجه گیری

مدل های آمیخته خطی دامنه گسترده ای از داده ها را شامل می شوند و در سالهای اخیر در بسیاری از تحقیقات به ویژه تحقیقات زیستی و پزشکی به کار رفته است، ولی با وجود پیشرفت نرم افزارهای گوناگون در این زمینه بسیاری از محققین و دانشجویان به دلیل عدم آشنایی با این مدل ها، معمولا یا سعی در نرمال کردن داده ها دارند که در بیشتر مواقع این کار امکان پذیر نیست (مانند داده های باینری یا شمارشی) یا با وجود اثرات تصادفی آنها را نادیده گرفته یا آنها را به عنوان اثرات ثابت در نظر می گیرند که این کار موجب می شود برخی از پارامترها به اشتباه معنی دار شوند (فلاح زاده و اسدی ۱۳۹۸). اهمیت این موضوع، مسئله اصلی پژوهش حاضر است که با هدف آموزش مدل آماری آمیخته خطی چندمتغیره در داده های طولی چندگانه به دانشجویان پزشکی: مطالعه موردی داده های رشد کودکان انجام شد. در این مقاله از داده‌های طولی رشد نوزادان شامل قد، وزن و اندازه دور سر استفاده کردیم. با توجه به کاربرد زیاد داده های طولی در حوزه علوم پزشکی و زیستی و لزوم تولید علم از طریق پژوهش، افزایش مهارت های آماری دانشجویان ضروری می باشد. اغلب دانشجویان در تحصیلات دانشگاهی خود فقط یک دوره مقدماتی آمار را می گذرانند، این دوره جایی است که مدرسان آمار، دانشجویان را به استفاده از نرم افزارهای آماری جهت سهولت پژوهش و یا به کار بردن آمار در زندگی و حرفه شان ترغیب نمی کنند. روش های سنتی و تنها آموزش تئوریک آمار در حال حاضر از کارائی لازم برخوردار نیست، زیرا برای ایجاد پیوند محکم بین آمار و کاربرد آن در دنیای واقعی با عدم موفقیت روبرو بوده است (ییلماز ۱۹۹۶). در سال های اخیر، آموزش دهندگان آمار توجه خود را در بازاندیشی فرایند آموزش آمار در هر دو سطح دانشگاهی و پیش دانشگاهی متمرکز کرده اند. مدرسان آمار بر روی افزایش بهبود شناختی آموزش، مهارتها و دانشی که از دانشجویان انتظار می رود تمرکز کرده اند (گال و همکاران ۱۹۹۴). یکی دیگر از تغییرات مهم که بیش از چند دهه تأثیر عمده ای بر آموزش و یادگیری آمار گذاشته است، ادغام کامپیوتر به خصوص در کلاس های درس آمار و برنامه نویسی آماری می باشد. پیشرفت در فناوری های رایانه ای از یک سو منجر به افزایش دسترسی دانشجویان و از سوی دیگر رشد روزافزون در

www.SID.ir

Archive of SID

توسعه بسته های آماری کاربرپسند جهت انجام پژوهش های علمی شده است. استفاده دانشجویان از کامپیوتر در کلاس درس آمار به انجام سریع تر و کارامدتر وظایف محاسباتی و تحقیقاتی منجر شده، در نتیجه آنها به جای محاسبات می توانند تمرکز بیشتری بر مفاهیم آمار و پژوهش داشته باشند (گارفیلد ۱۹۹۵). بسیاری از محققان راهبردهای نرم افزاری ابتکاری به کار رفته برای یادگیری آمار در دوره های کارشناسی و تحصیلات تکمیلی را بیان کرده اند. هم چنین در استفاده از تکنولوژی در مورد ترکیباتی که شامل برنامه های نرم افزاری با روش های جدید برنامه های درسی و منابع اینترنتی هستند بحث می کنند. (بارتس ۲۰۰۱) در مطالعه ای که انجام داد دریافت علیرغم توصیه مدرسین آمار به استفاده از تحلیل رایانه ای داده ها، اما نقش این روش در دانشگاه ها و مؤسسات علمی کمرنگ بوده و از نقطه نظر وی اساتید فکر می کنند که درک مفاهیم آماری همراه با نرم افزارهای آماری می تواند بسیار مشکل باشد و در واقع ممکن است استفاده از نرم افزارها و برنامه های رایانه ای موجب تأخیر در درک و آموزش مفاهیم آماری گردد. در این مطالعه دانشجویان کار عملی بر داده های رشد کودکان درنتیجه استفاده از تغذیه را با به کارگیری مدل های آمیخته تجربه کردند. مدل آماری نشان داد که نوزادانی که از شیر مادر تغذیه می‌کنند از نظر اندازه دورسر و قد رشد بهتری نسبت به نوزادانی که با شیرخشک و شیرترکیبی تغذیه می شوند، داشتند. مدل های آمیخته برازش بهتری به داده ها می دهند و مانع از معنادار شدن نادرست برخی پارامترهای مدل می شوند. در گذشته به دلیل روش های محاسباتی پیچیده این مدل ها کمتر مورد استقبال قرار گرفته اند ولی امروزه به دلیل پیشرفت نرم افزارهای گوناگون روش های به دست آوردن اثرات ثابت و تصادفی در مدل های آمیخته خطی در حال افزایش است و موضوع اصلی بسیاری از تحقیقات شده است (فلاح زاده و اسدی ۱۳۹۸). لی سون در سال ۲۰۱۱ در مقاله ای به مقایسه تفاوت روش های برآورد مدل های خطی آمیخته و مدل های خطی آمیخته تعمیم یافته پرداخت و نتیجه گرفت این روش ها انعطاف پذیرتر هستند. (گیلانی و همکاران ۱۳۹۰) به مقایسه مدل لجستیک حاشیه ای و مدل لجستیک آمیخته برای بررسی عوامل مؤثر بر بیماری پرفشاری خون پرداختند و نتیجه گرفتند که مدل لجستیک آمیخته برازش بهتری به داده ها می دهد.

در پژوهش حاضر نیز کاربرد مدل های آمیخته در داده های طولی رشد کودک بررسی شد و نشان داد که این مدل برازش بهتری به داده ها دارد. زیرا در بیشتر مطالعات در مورد اثرات تغذیه بر رشد نوزاد، تجزیه و تحلیل مقطعی در فواصل مختلف، معمولاً این تأثیرات تا ۱۲ ماهگی در نظر گرفته شد و شاخص‌های رشد در هر زمان، بدون در نظر گرفتن همبستگی حاصل از اندازه-گیری های طولی مورد مطالعه قرار گرفتند. همچنین شاخص‌های رشد به صورت تک متغیره در نظر گرفته شدند و هیچ ارتباط معنی‌داری بین آنها دیده نمی‌شود. در مطالعه حاضر، از مدلی استفاده شد که می‌تواند وابستگی حاصل از اندازه‌گیری‌های طولی و شاخص‌های اندازه‌گیری تن‌سنجی قد، وزن و اندازه دور سر را در نظر بگیرد. روش-های پیشنهادی به پزشک اجازه می‌دهد تا داده های طولی را در شرایط متنوعی به کار گیرد. تجزیه و تحلیل و مدل‌سازی چندمتغیره با اثرات تصادفی کمتر ارزیابی شده است. این مطالعه نشان می‌دهد که محققان علوم پزشکی می‌توانند این نوع مدل‌ها را بیشتر ارزیابی و تجزیه و تحلیل کنند.

References

Bartz, A. (2001). Computer and software use in teaching the beginning statistics course. Teaching of Psychology, 28(2), 147-150.

Crowder, M. (1990). Analysis of Repeated Measures, Chapman & Hall/CRC Monographs on Statistics and Applied Probability Book 41.

www.SID.ir

Archive of SID

Dempster, A. P., Laird, N. M., and Rubin. D. B., (1977). Maximum Likelihood from Incomplete Data Via the EM Algorithm. Journal of the Royal Statistical Society, Series B, 39(1), 1–38.

Davis, C.S., (2002). Statistical Methods for the Analysis of Repeated Measurements. Springer Texts in Statistics, New York.

Everitt, B. S., Hand, D. J., (1981). Finite Mixture Distributions, Springer, Netherlands.

Fallahzadeh, H., Asadi, F., (2019). Generalized Linear Mixed Models: Introduction, Estimation Methods and Their Application in Medical Studies. Paramedical Sciences and Military Health. 14(1), 33-38.

Gal, I., Ginsburg, L., (1994). The role of beliefs and attitudes in learning statistics: Towards an assessment framework. Journal of Statistics Education, 2, 1-15.

Garfiel. J, (1995). How students learn statistics. International Statistical Review, 1, 25-34.

Gilani, N., Kazemnejad. A., Zayeri. F., Yazdani. J., (2011). Comparison of Marginal Logistic Model with Repeated Measures and Conditional Logistic Model in Risk Factors Affecting Hypertension. JOURNAL OF MAZANDARAN UNIVERSITY OF MEDICAL SCIENCES, 21(82), 27-35.

Hughes, DM., García-Fiñana, M., Wand, MP., (2021). Fast approximate inference for multivariate longitudinal data. Biostatistics: doi: 10.1093/biostatistics/kxab021. Epub ahead of print. PMID: 33991420.

Pearson, K. (1894). Contributions to the Theory of Mathematical Evolution. Philosiphcal Transitions of the Royal Society of London, 185, 71–110.

Wolfe, J. H. (1965). A Computer Program for the Computation of Maximum Likelihood Analysis of Types, Research Memo. SRM 65-12. San Diego: U.S. Naval Personal Research Activity.

Wang, W.L., and Fan, T.H., (2011). Estimation in multivariate t linear mixed models for multiple longitudinal data. Statistica sinica 21, 1857-1880.

Yilmaz, M. R. (1996). The challenge of teaching statistics to nonspecialist. Journal of Statistics Education, 4(1), 1-9.

www.SID.ir

Archive of SID

مجله‌ی توسعه‌ی آموزش جندی‌شاپور
فصلنامه‌ی مرکز مطالعات و توسعه‌ی آموزش علوم پزشکی
سال سیزدهم، شماره ۱، بهار ۱۴۰۱

آموزش مدل آماری آمیخته خطی چندمتغیره در داده های طولی چندگانه به دانشجویان پزشکی: مطالعه موردی داده های رشد کودکان

مریم مرادی: دانشجوی دکتری آمار، گروه آمار، دانشکده علوم پایه، واحد علوم و تحقیقات، دانشگاه آزاد اسلامی، تهران، ایران.
کامبیز احمدی انگالی*: عضو هیئت علمی، گروه آمار زیستی، دانشکده بهداشت، دانشگاه علوم پزشکی جندی شاپور،اهواز، ایران.
محمدحسن بهزادی: عضو هیئت علمی، گروه آمار، دانشکده علوم پایه، واحد علوم و تحقیقات، دانشگاه آزاد اسلامی، تهران، ایران.
رحمان فرنوش: عضو هیئت علمی، گروه آمار، دانشگاه علم و صنعت ایران، تهران، ایران.

چکیده: پژوهش حاضر با هدف آموزش مدل آماری آمیخته خطی چندمتغیره در داده های طولی چندگانه به دانشجویان پزشکی: مطالعه موردی داده های رشد کودکان انجام شده است. روش پژوهش توصیفی-تحلیلی و بر پایه داده های طولی بود. جامعه آماری پژوهش شامل ۱۰۰ دانشجو بود که به طور تصادفی انتخاب و در دو کارگاه آموزش مبتنی بر مدلسازی آماری توزیع های تی و نرمال آمیخته خطی چندمتغیره بر پایه داده های رشد کودکان (قد، وزن و اندازه دور سر) شرکت کردند. شرکت کنندگان قبل و بعد از دوره آموزشی در یک آزمون عملی شرکت نمودند. از میانگین و انحراف معیار و مدل بندی آماری آمیخته خطی چندمتغیره توزیع تی و نرمال برای تحلیل داده ها استفاده شد. نتایج حاصل از پژوهش نشان داد که مقدار پارامترهای متغیر قد بر حسب مدل‌های MtLMM و MnLMM برای نوزادانی که از شیر مادر تغذیه می کنند به طور قابل توجهی بالاتر از شیر خشک بوده ($P < 0.05$). همچنین برآورد پارامترهای متغیر وزن برای نوزادانی که از شیر ترکیبی استفاده می‌کنند در مقایسه با نوزادانی که فقط شیر خشک مصرف می‌کنند به طور قابل توجهی بالاتر ($P < 0.05$) بود. آموزشهای ارائه شده به دو گروه منجر به افزایش معنی داری در یادگیری دانشجویان شد ($P < 0.05$). دانشجویان بیان کردند که آموزش مدل آماری آمیخته چندمتغیره در داده های طولی باعث آشکارشدن اهمیت کاربرد این مدل در تحقیقات پزشکی و علوم زیستی گردید.
واژگان کلیدی: آموزش، شیرمادر، مدل آمیخته، داده طولی، علوم پزشکی.

****نویسنده مسؤول:*** عضو هیئت علمی، گروه آمار زیستی،دانشکده بهداشت، دانشگاه علوم پزشکی جندی شاپور،اهواز، ایران.

Email: ahmadikambiz00@gmail.com

www.SID.ir