مجله مدل‌سازی پیشرفته ریاضی
سال ۱۴۰۲، دوره ۱۳، شماره ۳، ص ۴۴۸-۴۶۶

شاپا چاپی: ۲۲۵۱-۸۰۸۸
DOI: 10.22055/JAMM.2023.42846.2132

مدل‌سازی داده‌های چندمتغیره طولی با استفاده از توابع مفصل جفتی واین

محمدصادق لؤلؤ،(۱) محمدرضا آخوند،(۲) [۱] کامبیز احمدی انگالی،(۱) فاطمه برازجانی(۳)

(۱) گروه آماروایپیدمیولوژی، دانشکده بهداشت، دانشگاه علوم پزشکی جندی شاپور اهواز، اهواز، ایران
(۲) گروه آمار، دانشکده علوم ریاضی و کامپیوتر، دانشگاه شهید چمران اهواز، اهواز، ایران
(۳) مرکز تحقیقات تغذیه و بیماری‌های متابولیک، دانشگاه علوم پزشکی جندی شاپور اهواز، اهواز، ایران

دبیر مسئول: رحیم چینی‌پرداز

تاریخ دریافت: ۱۴۰۱/۱۰/۳۰ تاریخ پذیرش: ۱۴۰۲/۱۰/۲

چکیده: در برخی مطالعات پزشکی ممکن است چندین اندازه‌گیری بر روی هر بیمار داشته باشیم. در چنین شرایطی یک روش، به کارگیری اثرات تصادفی در مدل‌سازی داده‌ها است. گاهی این داده‌های طولی ممکن است برای چندین متغیر پاسخ اندازه‌گیری شود، در این حالت اگر چه می‌توان پاسخ‌ها را به صورت مجزا مدل‌بندی کرد اما چنین رویکردی موجب کاهش توان و کارایی در برآورد اثرات متغیرهای کمکی روی متغیر پاسخ می‌گردد. در چنین مدل‌هایی علاوه بر تحلیل وابستگی بین اندازه‌های مکرر مربوط به هریک از متغیرهای پاسخ، وابستگی بین پاسخ‌ها نیز باید مدل شود. از جمله روش‌هایی که در سال‌های اخیر توجه بسیاری از محققان را برای مدل‌سازی داده‌های چند متغیره به خود جلب کرده است، مدل‌سازی داده‌ها با استفاده از تابع مفصل است. از مهمترین مزیت‌های بکارگیری تابع مفصل نسبت به مدل‌سازی چند متغیره طولی داده‌ها به روش کلاسیک این است می‌توان علاوه بر توزیع نرمال هر توزیع دیگری غیر از نرمال را به عنوان توزیع‌های حاشیه‌ای در نظر گرفت. همچنین توزیع‌های حاشیه‌ای حتی می‌توانند توزیع‌های متفاوتی داشته باشند. در شرایطی که داده‌ها ساختاری چند متغیره داشته باشند یکی از راه‌های تشکیل توزیع‌های چندمتغیره استفاده از مفصل‌های جفتی واین است. ما در این مطالعه با استفاده از تابع مفصل‌های مختلف به کمک مفصل‌های جفتی واین ساختار طولی چندمتغیره‌ای را تشکیل می‌دهیم و این مدل‌ها را با مدل حاصل از برازش تابع مفصل نرمال چند متغیره مقایسه می‌کنیم. سپس بهترین مدل را با استفاده از معیار اطلاع آکائیک معرفی کرده و در پایان مدل ارائه شده را بر روی داده‌های برآورد اثر تغذیه بر رشد نوزادان به کار خواهیم گرفت.

واژه‌های کلیدی: اندازه‌گیری طولی، تابع مفصل نرمال چندمتغیره، مفصل‌های جفتی واین، رشد نوزاد، تغذیه نوزاد.

رده‌بندی ریاضی: 62H05; 62N02; 62J05; 62J12

[۱] نویسنده مسئول مقاله

(M.R. Akhoond) mr.akhoond@scu.ac.ir

۱ مقدمه

امروزه بخش قابل توجهی از طرح‌های مطالعاتی در علوم مختلف، از جمله آزمایش‌های کلینیکی و اطلاعات وابسته به موجودات زنده، ایجاب می‌کند که واحدهای آزمایش در زمان‌ها و یا موقعیت‌های مشخصی، به منظور کنترل تفاوت‌ها بین آزمودنی‌ها اندازه‌گیری شوند، دراین نوع آزمایش‌ها که هر واحد آزمایشی به طور متناوب اندازه‌گیری می‌شود معمولا از اصطلاح اندازه‌گیری مکرر و در حالتی که هر واحد نمونه‌گیری در زمان‌های متفاوت بررسی شود، تحلیل داده‌های طولی استفاده می‌گردد (دیگل [۲] و همکاران، ۲۰۰۲).
هدف اولیه مطالعات طولی، بررسی تغییرات متغیر پاسخ در طول زمان و نیز استخراج عامل‌های تاثیرگذار بر این تغییرات می‌باشد. به دلیل وجود اندازه‌گیری‌های تکراری، امکان ارزشیابی تغییرات درون هر واحد نیز فراهم می‌شود که این امکان در مطالعات مقطعی که پیامد تنها در یک زمان مشخص اندازه‌گیری می‌شود، وجود ندارد. در این حالت برخلاف سایر آزمایش‌ها بجای مشاهدات مستقل با مجموعه‌ای از مشاهدات وابسته سروکار خواهیم داشت. به دلیل آنکه اندازه‌های مکرر از دو بعد زمان و واحدهای آزمایشی تشکیل می شوند استفاده از مدل‌های مناسب که اثر مؤلفه‌های مذکور را در توصیف این داده‌ها به درستی لحاظ کند ضرورری است. مسئله مهم و کابردی دیگر که به ویژه در تحلیل داده‌های بالینی طولی مورد توجه قرار می‌گیرد، تحلیل همزمان یا چند متغیره‌ی طولی است. در این نوع مطالعات گاهی موقعیت‌هایی وجود دارد که در آن ها، هدف محقق آن است تا چندین متغیر را به دلیل ارتباط و همبستگی بین آنها، بطور همزمان مورد بررسی قرار دهد.
اگر در تحلیل داده‌ها با مواردی مواجه شویم که لازم است چند متغیر پاسخ به صورت همزمان از روی تعدادی متغیر کمکی پیش‌بینی گردند، این گونه مدل‌ها را مدل‌های چند متغیره می‌نامند. اگر چه می توان پاسخ‌ها را به صورت مجزا مدل‌بندی کرد اما چنین رویکردی موجب کاهش توان و کارایی در پیش‌بینی پاسخ‌ها می‌گردد. مزایای مدل‌های توام که همبستگی بین پاسخ‌ها را در نظر می گیرد عبارست از کنترل بهتر خطای نوع اول و افزایش کارایی در برآورد اثرات متغیرهای کمکی روی متغیر پاسخ است (دی لیون [۳] و همکاران، ۲۰۰۸)
. تمامی روش‌های تحلیل چندمتغیره داده‌های طولی از یک سو به بررسی سیستماتیک بین برآمدهای چندگانه و از سوی دیگر به تبیین ارتباط متغیرها با اندازه‌گیری‌های مکرر با عوامل موجود در مطالعه می پردازد. مدل‌های چند متغیره‌ی طولی با در نظر گرفتن همبستگی‌های متفاوت در تحلیل، به برآوردهایی با اریبی کمتر، سازگاری بالاتر و بنابراین دقت بیشتر منجر می‌شوند (فتیزموریس [۴] و همکاران، ۲۰۰۹).
عمومی‌ترین ابزار آماری برای تحلیل این مشاهدات، استفاده از رگرسیون چندمتغیره طولی است که ساختار همبستگی بین متغیرها را نیز لحاظ می‌نماید. تحلیل رگرسیونی یک روش آماری برای پیش‌بینی مقادیر یک یا چند متغیر پاسخ از مجموعه‌ای از مقادیر متغیرهای پیش‌بینی کننده است (جانسون [۵] و همکاران، ۲۰۰۲). مهمترین پیش فرض استفاده از روش رگرسیون چندمتغیره طولی فرض نرمال بودن توزیع‌های حاشیه‌ای است.
از جمله روش‌هایی که در سال‌های اخیر توجه بسیاری از محققان را برای مدل سازی داده‌های چندمتغیره به خود جلب کرده است، مدل‌سازی داده‌ها با استفاده از تابع مفصل است. ساختار توزیع های توام و بررسی ویژگی‌های آن با توزیع‌های حاشیه‌ای معلوم در صورت عدم برقراری فرض استقلال بین متغیرها کار پیچیده‌ای است. تابع مفصل برای ساخت یا بررسی توزیع‌های چند متغیره وقتی که توزیع‌های حاشیه‌ای یک متغیره و ساختار وابستگی آن‌ها معلوم است، بسیار مفید است.
از جمله مهمترین مزیت به‌کارگیری تابع مفصل نسبت به رگرسیون چندمتغیره طولی این است که نیازی به فرض نرمال بودن توزیع‌های حاشیه‌ای ندارد و هر توزیع دیگری غیر از نرمال را می‌توان به عنوان توزیع حاشیه‌ای متغیرهای پاسخ به‌کار گرفت. خصوصا این که الزامی در یکسان بودن توزیع‌های حاشیه‌ای هم وجود ندارد و در عمل می‌توان چندین متغیر پاسخ با توزیع‌های حاشیه‌ای متفاوت را همزمان مدل‌سازی کرد. از دیگر مزیت‌های به کارگیری تابع مفصل در مدل‌سازی مشاهدات این است که توزیع‌های حاشیه‌ای می‌توانند پارامتری و یا حتی ناپارامتری در نظر گرفته شوند، همچنین توابع مفصل از ضرایب همبستگی ناپارامتری بهره می‌گیرند و توزیع‌های شرطی به راحتی از روی تابع مفصل قابل محاسبه‌اند.
توابع مفصل به عنوان یک ابزار مناسب برای فرمول‌بندی و ساخت خانواده توزیع‌های چند متغیره و نیز روشی برای مطالعه اندازه‌های آزاد از مقیاس وابستگی، مورد توجه محققان قرار گرفته‌اند. در سال‌های اخیر با افزایش توجه به استفاده از تابع مفصل کتاب‌های بسیاری در این زمینه نوشته شده است که مهمترین آنها جو [۶] (۱۹۹۷) و نلسن [۷] (۲۰۰۶) می‌باشند. از آنجایی که بر اساس هر یک از توابع مفصل می‌توان توزیع‌های چند متغیره مختلفی را ساخت، جی نست و مکای [۸] (۱۹۸۶) تابع چگالی احتمال چند خانواده از توابع مفصل را در حالت کلی بدست آورده‌اند. نلسن [۹] (۱۹۹۱) روی مفصل‌ها و اندازه همبستگی تمرکز کرد. سانگ [۱۰] (۲۰۰۰) یک کلاس از مدل‌های پراکندگی چند متغیره تولید شده از مفصل نرمال را ارائه کرد. معرفی توابع مفصل مختلف، شرح خواص و ویژگی های آن به طور مبسوط در کتاب نلسن (۲۰۰۶)

[2] Diggle
[3] DeLeon
[4] Fitzmaurice
[5] Johnson
[6] Joe
[7] Nelsen
[8] Genest, MacKay
[9] Nelsen
[10] Song

آمده است. زیمر وتریودی [۱۱] (۲۰۰۶) جهت تحلیل یک مدل انتخابی برای پاسخ‌های شمارشی از یک چارچوب مفصل سه متغیره استفاده کردند. تریودی و زیمر (۲۰۰۷) روش‌هایی را به منظور یافتن مفصلی که به بهترین برازش را به مشاهدات داشته باشد، معرفی کرده‌اند. جی نست [۱۲] و همکاران (۲۰۰۷) در مورد مزایا و معایب استفاده از تابع مفصل برای مدل‌سازی داده‌های گسسته بحث کردند. سانگ و همکاران (۲۰۰۹) با استفاده از مفصل نرمال یک ابزار انعطاف‌پذیر و یکپارچه ارائه کردند که مدل‌های خطی تعمیم یافته یک بعدی مجزا را یکی می‌کند تا یک آنالیز رگرسیونی توأم از پاسخ‌های همبسته گسسته- پیوسته و آمیخته بدست آید. کلو و پایوا [۱۳] (۲۰۰۹) به مرور مطالعات انجام شده در زمینه مدل‌های رگرسیونی با استفاده از تابع مفصل پرداختند. دی لئون و وو [۱۴] (۲۰۱۱) با استفاده از تعریف متغیر پنهان برای پاسخ گسسته، مفصل نرمال را برای ساخت مدل توأم به‌کار بردند. رادیک و همکاران [۱۵] (۲۰۱۳) با استفاده از تابع مفصل‌های مختلف پاسخ‌های دوحالتی وضعیت بیمه و وضعیت بستری ۱۳۱۳۲ بیمار مراجعه کننده به یک بیمارستان را با رگرسیون اسپلاین مدل بندی کردند و نتایج مطالعه خود را با مفصل جو به عنوان مفصل کاراتر ارایه دادند. ثابتی [۱۶] و همکاران (۲۰۱۴) استنباط بیزی را برای مدل‌های مفصل شرطی سازگار با موقعیت‌های رگرسیونی که در آن پاسخ‌های دومتغیره پیوسته یا آمیخته هستند، انجام دادند. نیکلولوپولس و جو [۱۷] (۲۰۱۵) با استفاده از تابع مفصل‌های نرمال، گامبل و تی استیودنت طرح‌های عاملی برای سه پرسشنامه نگرش زندگی، نگرش در مورد محیط زیست و نگرش در مورد علم را مدل‌سازی کردند و نتایج هر کدام را برای مفصل کاراتر ارایه نمودند. جیریایی [۱۸] و همکاران (۲۰۱۶) با استفاده از تابع مفصل نرمال یک کلاس کلی از مدل‌های چند متغیره گسسته–پیوسته (آمیخته) تحت عنوان تابع توزیع‌های مفصل نرمال ارایه دادند. تا کنون روش‌های متعددی برای مدل‌سازی دو یا چند پاسخ با اندازه‌گیری طولی ارائه شده است. یکی از پرکاربردترین مدل‌هایی که در تحلیل این نوع داده‌ها به کار می رود، مدل‌های خطی با اثرات آمیخته (LMM) است (لرد و ویر[۱۹]،۱۹۸۲) . جورجیوا [۲۰] (۲۰۰۱) یک مدل خطی تعمیم یافته با اثرات تصادفی (GLMM) جداگانه برای دو پاسخ با اندازه‌گیری خوشه‌ای در نظر گرفت سپس با به‌کارگیری یک توزیع دو متغیره نرمال برای اثرات تصادفی یک مدل توام را ارائه داد. وثنیج و همکاران (۲۰۱۳) از تابع مفصل دومتغیره نرمال برای مدل‌سازی داده‌های پیوسته طولی با اثر تصادفی نرمال استفاده نمود. وو و دی لیون [۲۱] (۲۰۱۴) یک مدل دومتغیره گسسته–پیوسته با به‌کارگیری تابع مفصل نرمال را جهت برآورد اثر چند دز یک دارو بر پاسخ‌های رشد وزنی و نبود نقص عضو چندین جنین موش ارایه نمود. روی [۲۲] (۲۰۱۶) نیز با تعمیم مطالعه وثنیج [۲۳](۲۰۱۳) انواع حالت‌های ممکن اثر تصادفی در مدل طولی با تابع مفصل دومتغیره نرمال و همچنین با توزیع‌های حاشیه‌ای متفاوت ارائه و به کمک شبیه سازی داده‌ها مدل‌ها را مقایسه نمود. مطالعات ذکر شده محدود به مدل‌سازی داده‌های طولی در حالت دو متغیر با استفاده توابع مفصل است و این مدل‌ها در حالت طولی چندمتغیره مورد بحث قرار نگرفته‌اند، دایونگ کیم[۲۴] و همکاران (۲۰۱۳) با تاکید بر این موضوع که شناسایی یک تابع مفصل چند متغیره مناسب برای مدل‌سازی با ساختار وابستگی در داده‌های چند متغیره ساده نیست، توسعه تابع مفصل به کمک واین را برای غلبه بر این مشکل پیشنهاد دادند. به منظور آشکار کردن و درک کامل الگوهای وابستگی پیچیده و پنهان در داده‌های چند متغیره، مخلوطی از مفصل‌های واین پیشنهاد شده است. از آنجایی که یک مفصل واین دارای پارامترهای متعددی است که وابستگی را از طریق ساخت تکراری مفصل های جفتی در نظر می‌گیرد، این روش می تواند پیچیده‌گی الگوهای چندمتغیره را تسهیل کند. به دنبال کار دایونگ و همکارانش، استفاده از مفصل‌های جفتی واین، برای مدل‌سازی چندمتغیره و به‌خصوص داده‌های طولی، توسط محققین شدت گرفت. روسکن[۲۵] و همکاران (۲۰۱۷) روش انعطاف‌پذیر خود را برای مدل‌سازی همبستگی بین اجزای داده‌های طولی چند متغیره غیرنرمال با استفاده از رویکرد مفصل‌های جفتی واین معرفی کردند. شی [۲۶] و همکاران (۲۰۱۸) برای پیش‌بینی خسارت‌های بیمه با داده‌های طولی چند متغیره از مفصل‌های جفتی واین بهره گرفتند. لین [۲۷] و همکاران (۲۰۲۰) از مفصل‌های جفتی واین برای مدل‌سازی داده‌های طولی باینری استفاده کردند. سفیدی و همکاران (۲۰۲۲) از مفصل‌های جفتی واین برای مدل‌سازی داده‌های طولی رتبه‌ای استفاده کردند. یانگ [۲۸] و همکاران (۲۰۲۲) مفصل‌های جفتی واین دو بخشی [۲۹] را برای مدل‌سازی داده‌های بیمه با اندازه‌گیری طولی به کار

---

[11] Zimmer, Trivedi
[12] Genest
[13] Kolev, Paiva
[14] DeLeon, Wu
[15] Radice
[16] Sabeti
[17] Nikoloulopoulos, Joe
[18] Jiryaie
[19] Laird, Ware
[20] Gueorguieva
[21] Wu, de Leon
[22] Roy
[23] Withanage
[24] Daeyoung Kim
[25] Ruscone
[26] Shi
[27] Lin
[28] Yang
[29] Two part Dvine copula

بردند. مشابه مطالعات مذکور ما در مطالعه حاضر، با کمک مفصل‌های جفتی واین یک مدل چندمتغیره طولی برای داده‌های رشد نوزاد با استفاده از تابع مفصل‌های فرانک، کلایتون و FGM ارائه کرده‌ایم و با مدل طولی بدست آمده با استفاده از مفصل نرمال چندمتغیره مورد مقایسه قرار داده‌ایم. هدف اصلی ما معرفی مفصل‌های جفتی واین و کاربرد فراگیر آن در ساخت مدل‌های چند متغیره در تمامی حوضه‌ها به خصوص داده‌های طولی در علوم پزشکی است. با تاکید بر این موضوع که مدل‌سازی با این روش پیچدگی‌های روش‌های کلاسیک در ساخت مدل‌های چندمتغیره طولی را به خوبی تسهیل می‌کند و می‌تواند برای داده‌های طولی با ابعاد بالای متغیرهای پاسخ کاربردی باشد. در مطالعه حاضر علاوه بر بکارگیری تابع مفصل نرمال برای پاسخ‌های پیوسته، از مفصل‌های فرانک، کلایتون و FGM نیز برای مدل‌سازی این پاسخ‌ها با اندازه‌گیری طولی استفاده شده است. بدین منظور در بخش دوم مروری بر روش تحلیل مطالعات طولی و معرفی تابع مفصل و ویژگی‌های آن پرداخته می‌شود. در بخش سوم مدل‌سازی داده‌های سه متغیره طولی با استفاده از تابع مفصل مورد بررسی قرار می‌گیرد. در بخش چهارم مثال کاربردی رشد کودک با استفاده از این مدل بررسی می‌شود و در بخش پنجم به بحث و نتیجه گیری در مورد نتایج بدست آمده خواهیم پرداخت.

۲ مروری بر روش تحلیل داده‌های طولی و تابع مفصل

تحلیل ساختار همبستگی داده‌های طولی

همبستگی درون داده‌های طولی موجب می‌گردد که مشاهدات ثبت شده برای یک فرد در زمان یا حالات مختلف و یا بین افراد یک خوشه، شبیه به هم باشد. به دلیل همبستگی بین پاسخ‌های هر آزمودنی، روش‌های برآوردیابی ویژه‌ای برای تحلیل آماری مورد نیاز است که منجر به استنباط‌های معتبری گردد. با توجه به اهداف نهایی جمع آوری داده‌های طولی و نیز طبق نظر تحلیل‌گر در برخورد با پارمترهای همبستگی، مدل‌های متفاوتی مانند حاشیه‌ای ، اثرهای تصادفی و انتقال توسط محققان مختلف توسعه یافته و در مباحث کاربردی مورد استفاده قرار گرفته‌اند. در این مطالعه روش استفاده از اثر تصادفی برای تحلیل ساختار همبستگی درون فردی بکار گرفته شده است. در مدل‌های خطی با اثرهای تصادفی، عموماً فرض بر آن است که برخی ضرایب رگرسیونی برای آزمودنی‌ها به دلایل متفاوت، به خود آزمودنی وابسته بوده و دارای توزیع جداگانه می‌باشد. به عبارت دیگر، همبستگی بین مشاهدات در داده‌های تکراری یا داده‌های موجود در یک خوشه را می‌توان ناشی از داشتن تأثیر تصادفی یکسان در فرد یا یک خوشه دانست که این اثر در خوشه‌ی دیگر متفاوت است. در حقیقت همین ویژگی است که موجب ایجاد همبستگی در اندازه‌های ثبت شده از یک فرد یا یک خوشه می گردد (فتیزموریس و همکاران، ۲۰۰۹). در حالت کلی یک مدل خطی با اثر تصادفی به صورت زیر معرفی می شود:

$$\boldsymbol{Y}_{N\times ۱} = X^T_{N\times P}\boldsymbol{\beta}_{P\times ۱} + Z^T_{N\times q}\boldsymbol{b}_{q\times ۱} + \varepsilon_{N\times ۱}$$

که در آن $\boldsymbol{Y}$ بردار مقادیر مشاهده شده متغیر پاسخ، X ماتریس پررتبه ستونی شامل متغیرهای مستقل یا کمکی، $\boldsymbol{\beta}$ بردار پارمترهای مجهول اثرهای ثابت، $\boldsymbol{b}$ بردار نمایانگر اثرهای تصادفی و ماتریس Z شامل برخی متغیرهای توضیحی و نشانگر که با توجه به فرض های موجود درباره اثرهای تصادفی $\boldsymbol{b}$ در مدل ساخته می‌شود. $\boldsymbol{\epsilon}$ نیز برداری از مؤلفه‌های خطای تصادفی مدل است که مستقل از $\boldsymbol{b}$ توزیع می‌گردد. استنباط آماری پارمترهای مدل بر اساس تابع درستنمایی حاشیه‌ای

$$L(\boldsymbol{\beta},\Theta|X,Y) = \int_{R^q} f(y|x,\boldsymbol{b},\boldsymbol{\beta})g(\boldsymbol{b}|\Theta)d\boldsymbol{b}$$

انجام می‌شود که معمولاً در آن تابع چگالی اثرهای تصادفی $g(\boldsymbol{b}|\Theta)$ و چگالی $f(y|x,\boldsymbol{b},\boldsymbol{\beta})$ نرمال فرض می‌شود. همانطور که بیان شد در مدل‌سازی داده‌های چندمتغیره طولی با تابع مفصل می‌توان توزیع $f(y|x,\boldsymbol{b},\boldsymbol{\beta})$ را توزیعی غیرنرمال فرض کرد. یکی از متداول‌ترین روش‌های برآوردیابی پارمترهای مجهول با اثرهای آمیخته، روش حداکثر درستنمایی لی و نلدر [۳۰] (۱۹۹۶) است. استفاده از این روش منجر به حل انتگرال‌های پیچیده در تابع درستنمایی حاشیه‌ای، بخصوص در حالت غیر نرمال بودن توزیع اثرهای تصادفی شده که به روش‌های تکراری عددی پیشرفته نیاز دارد. در مطالعات اخیر روش مربع بندی گوس–نیوتن به عنوان یک روش موثر برای تقریب عددی انتگرال‌های پیچیده به منظور یافتن برآوردحداکثر درستنمایی پارمترها در مدل‌های با اثر تصادفی نرمال ارائه شده است (پین هیرو و چائو [۳۱]، ۲۰۰۶).

مروری بر تابع مفصل

تابع مفصل ابزاری است که ارتباط بین یک تابع توزیع چند متغیره با توزیع های حاشیه‌ای یک بعدی و ساختار همبستگی آن را نشان می‌دهد. تابع مفصل، تابعی است که توزیع چند متغیره را به حاشیه‌های یک متغیره آنها پیوند می‌دهد. C یک تابع مفصل است اگر $\mathrm{C} : [۰,۱]^۲ \to [۰,۱]$ در شرایط زیر صدق کند :

[30] Lee, Nelder

[31] Pinheiro, Chao

۱. $C(u,\circ)=C(\circ,u)=\circ$

۲. $C(u,۱)=u, C(۱,v)=v$

۳. اگر $v_۱, v_۲, u_۱, u_۲ \in [\circ, ۱]; u_۲ \geq u_۱, v_۲ \geq v_۱$ آنگاه

$$C\left(u_۲, v_۲\right)-C\left(u_۱, v_۲\right)-C\left(u_۲, v_۱\right)+C\left(u_۱, v_۱\right) \geq \circ$$

.

خانواده‌هایی از تابع مفصل

خانواده‌های مختلفی از توابع مفصل در تجزیه و تحلیل‌ها مورد استفاده قرار می‌گیرند که از جمله مهمترین آن‌ها می‌توان خانواده توابع مفصل بیضوی که خود شامل برخی از توابع مفصل مانند، مفصل گوسین (نرمال)، مفصل تی، مفصل نرمال چوله، مفصل تی چوله و خانواده توابع مفصل‌های ارشمیدسی که با استفاده از یک تابع مولد و براساس تبدیل لاپلاس بدست می‌آیند را نام برد. خانواده‌های توابع مفصل بسیاری در این دسته قرار می‌گیرند که مهمترین آن‌ها، مفصل فرانک، مفصل کلایتون، مفصل گامبل است. ساده‌ترین تابع مفصل، مفصل حاصلضرب است. تابع توزیع این مفصل در حالت دو متغیره به صورت زیر تعریف می‌شود:

$$C\left(u_۱, u_۲\right)=u_۱ u_۲$$

مفصل حاصلضرب یک معیار مهم است، زیرا متناظر با استقلال میان دو متغیر می باشد.
براساس قضیه اسکالر برای هر توزیع توام $F_{Y_۱,Y_۲}(y_۱, y_۲)$ از یک جفت متغیرتصادفی $(Y_۱, Y_۲)$ با توابع توزیع حاشیه‌ای $F_{Y_۱}(y_۱)$ و $F_{Y_۲}(y_۲)$ تابع دو متغیره وجود دارد، به طوری که برای هردو عدد حقیقی $y_۱$ و $y_۲$ رابطه

$$F_{Y_۱,Y_۲}\left(y_۱, y_۲\right)=C\left(F_{Y_۱}\left(y_۱\right), F_{Y_۱}\left(y_۱\right); \theta\right)$$

برقرار است. علاوه بر این اگر $Y_۱$ و $Y_۲$ متغیرهای تصادفی پیوسته باشد، آن گاه تابع مفصل C یکتاست و برعکس آن نیز برقرار است، یعنی برای هر دو متغیر $Y_۱$ و $Y_۲$ و هر تابع مفصل C، تابع $F_{Y_۱,Y_۲}(y_۱, y_۲)$ یک تابع توزیع دومتغیره با توزیع‌های حاشیه‌ای $F_{Y_۱}(y_۱)$ و $F_{Y_۲}(y_۲)$ است. در این روش تابع درستنمایی را در حالت خاص برای یک مدل دو متغیره $(y_۱, y_۲)$ به دست می آورند. فرض کنید یک نمونه تصادفی N تایی داشته باشیم که $y_k \sim F_k\left(y_k; \boldsymbol{x}_k, \boldsymbol{\beta}_k\right)$ و $\boldsymbol{x}_k$ به عنوان بردار متغیرهای کمکی و $\boldsymbol{\beta}_k$ به عنوان بردار پارمترها باشند. در این حالت برای $k=۱,۲$ تابع چگالی‌های حاشیه‌ای به صورت زیربدست می آید:

$$f_k\left(y_k; \boldsymbol{x}_k, \boldsymbol{\beta}_k\right)=\partial F_k\left(y_k; \boldsymbol{x}_k, \boldsymbol{\beta}_k\right) / \partial y_k$$

تابع چگالی توام $Y_۱$ و $Y_۲$ به صورت زیر محاسبه می‌شود:

$$f\left(y_۱, y_۲; \theta\right)=\partial^۲ C\left(F_۱\left(y_۱\right), F_۲\left(y_۲\right); \theta\right) / \partial y_۱ \partial y_۲=c_{۱۲}\left(F_۱\left(y_۱\right), F_۲\left(y_۲\right); \theta\right) f_۱\left(y_۱\right) f_۲\left(y_۲\right),$$

که در آن $c_{۱۲}\left(F_۱\left(y_۱\right), F_۲\left(y_۲\right); \theta\right)$ تابع چگالی مفصل است و از طریق مشتق‌گیری از تابع توزیع مفصل به فرم زیر بدست می آید:

$$c_{۱۲}\left(F_۱\left(y_۱; \boldsymbol{x}_۱, \boldsymbol{\beta}_۱\right), \mathrm{F}_۲\left(y_۲; \boldsymbol{x}_۲, \boldsymbol{\beta}_۲\right); \theta\right)=\partial^۲ C\left(F_۱\left(y_۱; \boldsymbol{x}_۱, \boldsymbol{\beta}_۱\right), \mathrm{F}_۲\left(y_۲; \boldsymbol{x}_۲, \boldsymbol{\beta}_۲\right); \theta\right) / \partial F_۱ \partial F_۲.$$

برآورد پارمترها

برای نوشتن تابع درستنمایی ابتدا تابع درستنمایی به شرط اثرات تصادفی نوشته می‌شود و سپس با انتگرال‌گیری بر روی اثرات تصادفی تابع درستنمایی برای پارامترهای مدل بدست می‌آید:

$$L(\boldsymbol{\Theta})=\prod_{i=۱}^n \int \prod_{t=۱}^T f_{Y_{it۱}, Y_{it۲} \mid b_i}(.|.) f\left(b_i\right) d b_i.$$

در نتیجه لگاریتم تابع درستنمایی به صورت زیر بدست می آید:

$$\begin{aligned} l(\boldsymbol{\Theta}) &= \log L(\theta) \\ &= \sum_{i=۱}^N\left[\log f\left(y_{i۱}; \boldsymbol{x}_{i۱}, \boldsymbol{\beta}_۱\right)+\log f\left(y_{i۲}; \boldsymbol{x}_{i۲}, \boldsymbol{\beta}_۲\right)+\log \mathrm{c}_{۱۲}\left(F_۱\left(y_{i۱}; \boldsymbol{x}_{i۱}, \boldsymbol{\beta}_۱\right), F_۲\left(y_{i۲}; \boldsymbol{x}_{i۲}, \beta_۲\right); \theta\right)\right] \end{aligned} \tag{۱.۲}$$

که در آن $\boldsymbol{\Theta}$ بردار کلیه پارامترهای مدل می‌باشد. با قرار دادن $l'(\boldsymbol{\Theta}) = \partial l(\boldsymbol{\Theta})/\partial\boldsymbol{\Theta}$ به عنوان معادلات امتیاز و $l''(\boldsymbol{\Theta}) = \partial^{۲} l(\boldsymbol{\Theta})/\partial\boldsymbol{\Theta}\partial\boldsymbol{\Theta}^T$ به عنوان ماتریس هسین، برآورد ماکسیمم درستنمایی پارامترها به وسیله حل معادلات درستنمایی $l'(\boldsymbol{\theta}) = ۰$ به دست می‌آید. این معادلات در حالت کلی غیرخطی هستند اما با استفاده از روش گوس-نیوتن به راحتی محاسبه می‌شوند. تحت برقراری شرایط نظم $\hat{\boldsymbol{\Theta}}$ سازگار است و به طور مجانبی دارای توزیع نرمال به صورت زیر می‌باشد:

$$\sqrt{N}(\hat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}) \to N\left[۰, -\left(\frac{۱}{N}\frac{\partial^{۲} l(\hat{\boldsymbol{\Theta}})}{\partial\hat{\boldsymbol{\Theta}}\partial\hat{\boldsymbol{\Theta}}^T}\right)^{-۱}\right].$$

مدل خطی با اثر تصادفی برای داده های طولی

فرض کنید Y_{itk} توزیع حاشیه‌ای kام $(k = ۱, ۲)$ برای فرد iام $(i = ۱, \ldots, n)$ در زمان tام $(t = ۱, \ldots, T)$ باشد. مدل خطی با اثر تصادفی به صورت زیر ارایه می‌شود:

$$Y_{itk} = \boldsymbol{X}_{itk}^T \boldsymbol{\beta}_k + \boldsymbol{Z}_{itk}^T \boldsymbol{b}_{ik} + \epsilon_{itk} \qquad (۲.۲)$$

که $\boldsymbol{X}_{itk}^T$ بردار متغیرهای مستقل (کمکی) و $\boldsymbol{\beta}_k$ بردار ضرایب رگرسیونی متناظر با آن‌ها، $\boldsymbol{Z}_{itk}$ بردار متغیرهای کمکی برای اثرات تصادفی و $\boldsymbol{b}_{ik}$ بردار اثرات تصافی است و ϵ_{itk} خطای مدل و دارای توزیع $N(۰, \sigma^{۲})$ است. ساده‌ترین نوع مدل‌های خطی چند متغیره با اندازه‌گیری طولی با بکارگیری پارامتر یکسان برای هر مدل حاشیه‌ای و فرض $Z_{itk} = ۱$ است که به مدل با اثرات تصادفی عرض از مبدا یکسان[32] معروف است. در این حالت مدل (۲.۲) به شکل زیر ساده می‌شود:

$$Y_{itk} = X_{itk}^T \boldsymbol{\beta}_k + b_i + \epsilon_{itk} \qquad (۳.۲)$$

دراین مطالعه کلیه مدل‌ها مبتنی بر مدل (۳.۲) ارایه می‌شود و فرض شده است که $b_i \sim N(۰, \sigma^{۲})$ می‌باشد. جهت مشاهده سایر حالت‌های $\boldsymbol{Z}_{itk}^T \boldsymbol{b}_{ik}$ در مدل کلی (۲.۲) به کتاب لرد و ویر (۱۹۸۲) یا فتیزموریس (۲۰۰۹) می‌توان مراجعه کرد. در مدل‌های خطی آمیخته برای داده های طولی[33] فرض می‌شود توزیع متغیر وابسته به شرط اثر تصادفی دارای توزیع نرمال است. طبق مدل (۳.۲)، $Y_{itk}|b_i \sim N(\mu_{itk}(b_i), \sigma^{۲})$، که در آن $\mu_{itk} = X_{itk}^T \boldsymbol{\beta}_{ik} + b_i$ است. همانطور که بیان شد از جمله مهمترین مزیت‌های استفاده از تابع مفصل برای مدل‌سازی این است که توزیعی حاشیه‌ای برای متغیر پاسخ را می‌توان هر توزیع دلخواه در نظر گرفت یعنی توزیع $Y_{itk}|b_i$ می‌تواند هر توزیع دیگری غیر از نرمال باشد. این در حالی است که یکی از پیش فرض‌های اصلی مدل‌های خطی برقراری فرض نرمال برای توزیع متغیر وابسته است. در این مطالعه توزیع لجستیک و نرمال به عنوان توزیع حاشیه‌های تابع مفصل استفاده شده است. در مدل دو متغیره فرض می‌شود:

$$Y_{it۱}|b_i \sim N(\mu_{it۱}(b_i), \sigma_۱^{۲}) \qquad (۴.۲)$$

$$Y_{it۲}|b_i \sim logistic(\mu_{it۲}(b_i), \sigma_۲) \qquad (۵.۲)$$

که در (۴.۲)، $\mu_{it۱} = X_{it۱}^T \boldsymbol{\beta}_۱ + b_i$ و به عنوان میانگین توزیع نرمال شرطی و $\sigma_۱$ انحراف معیار آن و در (۵.۲)، $\mu_{it۲} = X_{it۲}^T \boldsymbol{\beta}_۲ + b_i$ به عنوان پارامتر مکان و $\sigma_۲$ پارامتر مقیاس توزیع لجستیک شرطی است.

تابع درستنمایی با استفاده ازتوابع مفصل نرمال، فرانک، کلایتون و حاصلضرب برای مدل‌های دومتغیره

تابع توزیع شرطی مفصل نرمال از رابطه زیر به دست می‌آید:

$$F_{Y_{it۱}, Y_{it۲}|b_i}(\cdot|\cdot) = \Phi_۲[\Phi^{-۱}(u_{it۱}(b_i)), \Phi^{-۱}(u_{it۲}(b_i)); \theta]$$

[32] Shared random intercept

[33] Linear mixed models for longitudinal data

که در آن برای

$$u_{itk}(b_i) = F_{Y_{itk}|b_i}(y_{itk}|b_i)$$

$u_{itk}(b_i)$ها توزیع شرطی یکنواخت در بازه‌ی $(0, 1)$ هستند و $F_{Y_{itk}|b_i}(y_{itk}|b_i)$ تابع توزیع تجمعی شرطی با استفاده از حاشیه‌های (۴.۲) و (۵.۲) به دست می‌آید. θ پارامتر وابستگی بین Y_{it1}, Y_{it2} برای تابع مفصل نرمال است (که در این تابع مفصل همان ضریب همبستگی پیرسون است) که به صورت زیر تعریف می‌شود:

$$\theta = \rho = corr(U_{it1}(b_i), U_{it2}(b_i)|b_i)$$

و $\rho = 0$ اگر و تنها اگر Y_{it1}, Y_{it2} از هم مستقل باشند. b_i اثر تصادفی مربوط به فرد iام است که فرض می‌شود $b_i \sim N(0, \sigma^2)$ باشد. تابع چگالی شرطی دومتغیره Y_{it1}, Y_{it2} با استفاده از مفصل نرمال به شکل زیر به دست می‌آید (وثیج و همکاران، ۲۰۱۳).

$$f_{Y_{it1},Y_{it2}|b_i}(\cdot|\cdot) = \frac{\phi_2(\Phi^{-1}(u_{it1}(b_i)), \Phi^{-1}(u_{it2}(b_i)); \theta)}{\phi(\Phi^{-1}\{u_{it1}(b_i)\})\phi(\Phi^{-1}\{u_{it2}(b_i)\})} f_{Y_{it1}|b_i}(y_{it1}|b_i) \times f_{Y_{it2}|b_i}(y_{it2}|b_i) \tag{۶.۲}$$

که در آن برای $k = 1, 2$، تابع چگالی شرطی با استفاده از حاشیه‌های (۴.۲) و (۵.۲) است و ϕ_2, ϕ به ترتیب چگالی نرمال استاندارد و چگالی نرمال استاندارد دو متغیره است. حال به کمک رابطه (۱.۲) می‌توان تابع درستنمایی دومتغیره شرطی با استفاده از تابع مفصل نرمال را از طریق تابع چگالی (۶.۲) محاسبه و پارامترهای مدل را به روش گوس‌نیوتن برآورد نمود.

مشابه (۶.۲) برای سایر توابع مفصل چگالی شرطی آن‌ها از روی تابع چگالی‌های مفصل به دست می‌آید. عبارات (۷.۲)، (۸.۲)، (۹.۲) و (۱۰.۲) به ترتیب چگالی شرطی دو متغیره برای توابع مفصل فرانک، کلایتون، FGM و حاصلضرب را نشان می‌دهد.

$$f_{Y_{it1},Y_{it2}|B_i}(\cdot|\cdot) = \frac{\theta(1-e^{-\theta})e^{-\theta[(u_{it1}(b_i)+(u_{it2}(b_i)))]}}{[(1-e^{-\theta})-(1-e^{-\theta(u_{it1}(b_i))})(e^{-\theta(u_{it2}(b_i))}-1)]^2} f_{Y_{it1|B_i}}(y_{it1}|b_i) \times f_{Y_{it2|B_i}}(y_{it2}|b_i). \tag{۷.۲}$$

$$f_{Y_{it1},Y_{it2}|b_i}(\cdot|\cdot) = \left[(u_{it1}(b_i))^{-\theta} + (u_{it2}(b_i))^{-\theta} - 1\right]^{-2-\frac{1}{\theta}} \times [(u_{it1}(b_i))(u_{it1}(b_i))]^{-\theta-1} \times [\theta+1] \times f_{Y_{it1}|b_i}(y_{it1}|b_i) \times f_{Y_{it2}|b_i}(y_{it2}|b_i) \tag{۸.۲}$$

$$f_{Y_{it1},Y_{it2}|b_i}(\cdot|\cdot) = [1 + \theta(2u_{it1}(b_i) - 1)(2u_{it2}(b_i) - 1)]\, f_{Y_{it1}|b_i}(y_{it1}|b_i) \times f_{Y_{it2}|b_i}(y_{it2}|b_i). \tag{۹.۲}$$

$$f_{Y_{it1},Y_{it2}|b_i}(\cdot|\cdot) = f_{Y_{it1}|b_i}(y_{it1}|b_i) \times f_{Y_{it2}|b_i}(y_{it2}|b_i) \tag{۱۰.۲}$$

در کلیه عبارات بالا θ پارامتر وابستگی تابع مفصل، $u_{itk}(b_i) = F_{Y_{itk|b_i}}(y_{itk}|b_i)$ تابع توزیع شرطی حاشیه‌ها و $f_{Y_{itk|b_i}}(y_{itk}|b_i)$ تابع چگالی شرطی حاشیه‌ها برای $k = 1, 2$ است. تابع درستنمایی و برآورد پارامترهای مدل مشابه تابع مفصل نرمال به کمک رابطه (۱.۲) و چگالی‌های شرطی (۷.۲)، (۸.۲)، (۹.۲) و (۱۰.۲) به دست می‌آید.

۳ مدل‌بندی سه متغیره طولی با استفاده از تابع مفصل

در این بخش دو روش واین [34] و تابع مفصل نرمال سه متغیره برای مدل سازی سه متغیره معرفی و روش برآورد پارامتر به‌وسیله آن‌ها توضیح داده می‌شود.

روش واین

روش واین با بکارگیری مفصل‌های جفتی [35] از متغیرها، یک کلاس بزرگ‌تر از توزیع‌های چند بعدی را می‌سازد. از پیشگامان معرفی کننده این روش مدل سازی چندمتغیره از مفصل‌های جفتی، می‌توان از جو (۱۹۹۷) ، بدفورد و کوک(۲۰۰۲) و آس و همکاران(۲۰۰۹) نام برد. ابتدا به شرح این روش می پردازیم.

[34] Vine
[35] Pair-copula

همانطور که بیان شد برای دو متغیر $Y_۱, Y_۲$ تابع توزیع مفصل به شکل زیر معرفی می‌شود:

$$F_{Y_۱,Y_۲}(y_۱, y_۲; \theta) = C(F_{Y_۱}(y_۱), F_{Y_۲}(y_۲))$$

همچنین با توجه به تعریف تابع چگالی شرطی، $Y_۲$ به شرط $Y_۱$ که به صورت زیر نوشته می‌شود:

$$f_{۲|۱}(y_۲|y_۱) = \frac{f_{۱,۲}(y_۱, y_۲)}{f_۱(y_۱)}$$

بنابراین

$$f_{۲|۱}(y_۲|y_۱) = \frac{f_{۱,۲}(y_۱, y_۲)}{f_۱(y_۱)} = c_{۱۲}(F_۱(y_۱), F_۲(y_۲))f_۲(y_۲) \tag{۱.۳}$$

که در آن $c_{۱۲}(F_۱(y_۱), F_۲(y_۲))$ همان چگالی مفصل دو متغیره است. رابطه بالا را برای هر i, j می‌توان به فرم کلی زیر نوشت:

$$f_{j|i}(y_j|y_i) = c_{ij}(F_i(y_i), F_j(y_j))f_j(y_j) \tag{۲.۳}$$

از طرفی تابع چگالی یک توزیع سه متغیره براساس قانون‌های احتمال شرطی به صورت زیر نوشته می‌شود:

$$f_{۱,۲,۳}(y_۱, y_۲, y_۳) = f_۱(y_۱)f_{۲|۱}(y_۲|y_۱)f_{۳|۱,۲}(y_۳|y_۱, y_۲) \tag{۳.۳}$$

و عبارت $f_{۳|۱,۲}(y_۳|y_۱, y_۲)$ را می‌توان از طریق فرم کلی (۹.۲) به فرم زیر بسط داد:f

$$\begin{aligned} f_{۳|۱,۲}(y_۳|y_۱, y_۲) &= c_{۱۳|۲}\left(F_{۱|۲}(y_۱|y_۲), F_{۳|۲}(y_۳|y_۲)\right) f_{۳|۲}(y_۳|y_۲) \\ &= c_{۱۳|۲}\left(F_{۱|۲}(y_۱|y_۲), F_{۳|۲}(y_۳|y_۲)\right) c_{۲۳}(F_۲(Y_۲), F_۳(Y_۳))f_۳(y_۳) \end{aligned} \tag{۴.۳}$$

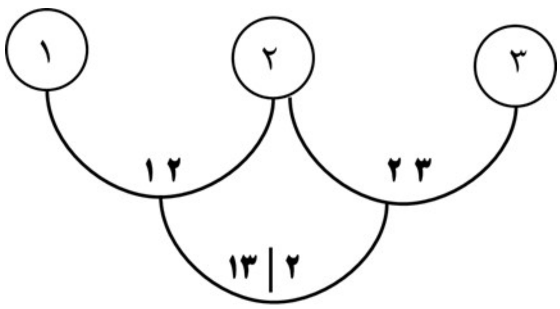
حال با جایگذاری (۴.۳) و (۱.۳) در عبارت (۳.۳) چگالی سه متغیره $Y_۱, Y_۲, Y_۳$ به شکل زیر به دست می‌آید:

$$\begin{aligned} f_{۱,۲,۳}(y_۱, y_۲, y_۳) =& f_۱(y_۱)c_{۱۲}(F_۱(y_۱), F_۲(y_۲))f_۲(y_۲) \\ & c_{۱۳|۲}(F_{۱|۲}(y_۱|y_۲), F_{۳|۲}(y_۳|y_۲))c_{۲۳}(F_۲(y_۲), F_۳(y_۳))f_۳(y_۳). \end{aligned}$$

اگر $y_۱|y_۲$ را از $y_۳|y_۲$ مستقل فرض کنیم یعنی فرض کنیم $c_{۱۳|۲}(F_{۱|۲}(y_۱|y_۲), F_{۳|۲}(y_۳|y_۲))$، عبارت بالا به شکل زیر ساده می‌شود

$$f_{۱,۲,۳}(y_۱, y_۲, y_۳) = f_۱(y_۱)f_۲(y_۲)f_۳(y_۳)c_{۱۲}(F_۱(y_۱), F_۲(y_۲))c_{۲۳}(F_۲(y_۲), F_۳(y_۳)). \tag{۵.۳}$$

شکل ۱ یک مدل واین سه متغیره را نشان می‌دهد، از آن جهت که ترسیم این نوع نمودار از مدل‌های چند متغیره واین شبیه به خوشه‌های



شکل ۱: شکل شماتیک یک مدل واین سه متغیره

انگور است، این مدل vine به معنی خوشه‌ی انگور نام گرفته است.

در عبارت (۵.۳)، c_{12} و c_{23} هر تابع مفصلی می‌توانند باشند و الزامی به یکسان در نظر گرفتن توابع مفصل هر جفت متغیر در این روش وجود ندارد. این خاصیت روش واین را به یک روش بسیار توانمند و ساده در برازش مدل‌های چندبعدی به کمک تابع مفصل تبدیل کرده است. چگالی شرطی سه متغیره به روش واین به کمک عبارت (۵.۳) به شکل زیر نوشته می‌شود.

$$\begin{aligned} f(y_{it1}, y_{it2}, y_{it3}|b_i) =& f_{Y_{it1}|b_i}(y_{it1}|b_i) f_{Y_{it2}|b_i}(y_{it2}|b_i) f_{y_{it3}|b_i}(y_{it3}|b_i) \\ & c_{12}\left(F_{Y_{it1}|b_i}(y_{it1}|b_i), F_{Y_{it2}|b_i}(y_{it2}|b_i); \theta_{12}\right) \\ & c_{23}\left(F_{Y_{it2}|b_i}(y_{it2}|b_i), F_{Y_{it3}|b_i}(y_{it3}|b_i); \theta_{23}\right) \end{aligned} \qquad (۶.۳)$$

در این عبارت θ_{12} و θ_{23} به ترتیب پارامتر وابستگی تابع مفصل برای c_{12} و c_{23} است. حال با محاسبه لگاریتم تابع درستنمایی با استفاده از تابع چگالی توام سه متغیره (۶.۳) مشابه حالت دومتغیره می‌توان برآورد پارامترها را به روش حداکثر درستنمایی و با کمک روش گوس نیوتن محاسبه کرد.

تابع مفصل سه متغیره نرمال

تابع توزیع مفصل k متغیره نرمال به شکل زیر تعریف می‌شود.

$$C(u_1, \ldots, u_k; R) = \Phi_k(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_k))$$

که در آن R ماتریس همبستگی تابع مفصل است و یک ماتریس مربعی $k \times k$ است.

فرم ماتریسی تابع چگالی مفصل نرمال چند متغیره نیز به صورت زیر تعریف می‌شود

$$c(\boldsymbol{q}; R) = \frac{1}{|R|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}\boldsymbol{q}^T(R^{-1} - I)\boldsymbol{q}\right\} \qquad (۷.۳)$$

که در آن q بردار k بعدی و $\boldsymbol{q} = (q_1 = \Phi^{-1}(F_{Y_1}(y_1)), \ldots, q_k = \Phi^{-1}(F_{Y_k}(y_k)))$، R^{-1} معکوس ماتریس همبستگی تابع مفصل R و $|R|$ دترمینان آن است. I ماتریس همانی $k \times k$ است. برای مفصل سه متغیره با $k = 3$، ماتریس همبستگی تابع مفصل به شکل زیر تعریف می‌شود.

$$R = \begin{pmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 & \rho_{13} \\ \rho_{12} & \rho_{13} & 1 \end{pmatrix}$$

که عناصر ρ_{ij} همبستگی دو به دو مفصل برای هر سه متغیر است وقتی $i, j = 1, 2, 3$ و دترمینان R به فرم زیر به دست می‌آید:

$$|R| = 1 - \rho_{12} - \rho_{13} - \rho_{23} + 2\rho_{12}\rho_{13}\rho_{23}$$

ضمناً فرم باز شده عبارت $\left\{-\frac{1}{2}\boldsymbol{q}^T(R^{-1} - I)\boldsymbol{q}\right\}$ به صورت زیر است

$$\begin{aligned} &\left\{-\frac{1}{2}\boldsymbol{q}^T(R^{-1} - I)\boldsymbol{q}\right\} \\ &= -\frac{1}{2|R|}\begin{bmatrix} q_1^2(1-\rho_{23}) + q_2^2(1-\rho_{13}) + q_3^2(1-\rho_{12}) - \\ 2q_1q_2(\rho_{12} - \rho_{23}\rho_{13}) - 2q_1q_3(\rho_{13} - \rho_{12}\rho_{23}) - 2q_2q_3(\rho_{23} - \rho_{12}\rho_{13}) \end{bmatrix} \end{aligned}$$

، که با جایگذاری در عبارت (۷.۳) فرم غیرماتریسی تابع چگالی سه متغیره نرمال به دست می‌آید.

اگر فرض کنیم $\rho_{ij} = \rho$ برای هر $i, j = 1, 2, 3$ یعنی $R = \begin{pmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{pmatrix}$ باشد عبارات بالا به فرم زیر ساده‌تر می‌شوند:

$$|R| = 1 - 3\rho + 2\rho^3$$

و

$$\left\{-\frac{۱}{۲}\boldsymbol{q}^T(R^{-۱}-I)\boldsymbol{q}\right\} = -\frac{۱}{۲|R|}\left[(۱-\rho)(q_۱^۲+q_۲^۲+q_۳^۲) - ۲(\rho-\rho^۲)(q_۱q_۲+q_۱q_۳+q_۲q_۳)\right]$$

بنابراین تابع چگالی شرطی سه متغیره به کمک تابع مفصل نرمال به شکل زیر نوشته می‌شود:

$$f_{Y_{ij۱},Y_{ij۲},Y_{ij۳}|b_i}(\cdot|\cdot) = c(\boldsymbol{q}; R, b_i) f_{Y_{ij۱}|b_i}(y_{ij۱}|b_i) f_{Y_{ij۲}|b_i}(y_{ij۲}|b_i) f_{Y_{ij۳}|b_i}(y_{ij۳}|b_i) \qquad (۸.۳)$$

که در آن $c(\boldsymbol{q}; R, b_i)$ تابع چگالی شرطی سه متغیره تابع مفصل نرمال طبق (۷.۳) با $k = ۳$ است و

$$\boldsymbol{q} = \left(q_۱ = \Phi^{-۱}(F_{Y_{it۱}|b_i}(y_{it۱}|b_i)), q_۲ = \Phi^{-۱}(F_{Y_{it۲}|b_i}(y_{it۲}|b_i)), q_۳ = \Phi^{-۱}(F_{Y_{it۳}|b_i}(y_{it۳}|b_i))\right)$$

حال با محاسبه لگاریتم تابع درستنمایی از (۵.۳) مشابه حالت دومتغیره می‌توان برآورد پارامترها را با کمک روش گوس‌نیوتن بدست آورد.
برای مقایسه مدل‌ها می‌توان از معیارهای اطلاع نظیر معیار اطلاع آکائیک استفاده کرد. معیار اطلاع آکائیک با فرمول زیر برای مقایسه مدل‌ها مورد استفاده قرار گرفت و کمتر بودن مقدار آن حکایت از بهتر بودن مدل برازش شده دارد.

$$AIC = -۲\log L + ۲P$$

که در آن L تابع درستنمایی و P تعداد پارامترهای مدل می‌باشند. کلیه محاسبات با استفاده از کتابخانه Proc nlmixe در نرم افزار SAS نسخه ۹.۴ انجام شده است.

۴ مثال کاربردی

رشد عبارت است از بزرگ شدن اجزای بدن که نقش بسیار مهمی را در سلامت ایفا می‌کند و از ارکان مهم سلامت جامعه محسوب می‌شود (توتونچی و همکاران، ۲۰۰۹) استفاده از شاخص‌های رشد و وضع تغذیه‌ی کودکان به عنوان معیاری برای سنجش سطح بهداشت یک منطقه بر پایه این واقعیت است که رشد ضعیف برای بیشتر کودکان به معنی انحراف از شرایط محیطی مطلوبی است که باید زمینه رشد و نمو را در کودکان به بهترین شکل فراهم کند (ریحانی و همکاران، ۲۰۰۰). در اکثر مطالعات برای پایش رشد فیزیکی و سلامت کودکان از شاخص‌های تن‌سنجی نظیر وزن، قد و دور سر و بویژه وزن کردن مرتب استفاده می‌شود (امیرحکیمی و همکاران، ۲۰۱۵) پایش منظم قد، وزن و دور سر کودکان موجب کشف به موقع اختلالات رشد و آغاز زود هنگام اقدام‌های پیشگیری و درمانی می‌شود (گرین [۳۶] و همکاران، ۲۰۰۱). مهم‌ترین عامل موثر بر روند رشد جسمانی کودک، تغذیه است (اسکولادوتیر [۳۷] و همکاران، ۲۰۰۵). در مراحل اول زندگی شیر مادر بدون شک ایده آل ترین غذا است که علاوه بر فراهم کردن کلیه مواد مغذی کودک، دارای مزایایی از نظر ایمنی و سلامت روانی نیز می‌باشد (بهرمن و همکاران، ۲۰۰۰). در اکثر مطالعات که اثر تغذیه نوزادان را بر رشد آنان بررسی کرده‌اند جهت تعدیل مدل پیشنهادی خود و بررسی دقیق تر، فاکتورهایی همانند جنس، رتبه تولد، بعد خانوار، شاخص‌های سلامت مادر، وضعیت اقتصادی و فرهنگی خانواده را نیز به کار گرفته‌اند (ویکتورا [۳۸] و همکاران، ۲۰۱۶). در این مطالعه کل ۱۹۵ نوزاد متولد شده روستاهای شهرستان بوشهر که طی سال ۹۴ به مرکز بهداشت مراجعه کرده بودند با معیار ورود به مطالعه وزن تولد بیش از ۲۵۰۰ گرم، سن جنینی بین ۳۷ تا ۴۲ هفته و بدون عارضه مادرزادی و مراجعه کامل در طول مطالعه بررسی شدند، در این مدت ۶ مرتبه (بدو تولد، دو ماهگی، چهار ماهگی، شش ماهگی، هفت ماهگی، نه ماهگی) مراجعه داشتند که در مجموع ۱۲۰ کودک مورد مطالعه قرار گرفتند. در هر بار مراجعه قد (به سانتی متر)، وزن (به کیلوگرم) و دور سر (به سانتی متر) کودک با استفاده از متر و ترازوهای مخصوص نوزادان در مرکز بهداشتی اندازه گیری شد. سن کودک در زمان مراجعه به مرکز با دقت روز ثبت گردید. شیرخوران بر اساس نوع تغذیه به سه گروه تقسیم می‌شوند. گروه اول شیر مادر خواران انحصاری که بر اساس تعریف سازمان بهداشت جهانی (WHO) کودکی است که تنها از شیر مادر و قطره ویتامین در اولین شش ماه زندگی استفاده کرده باشد. دو گروه دیگر شامل شیر خشک خوار و گروهی که از هر دو شیر مادر و شیر خشک تغذیه می‌کردند. بنابراین متغیرهای وابسته در این مطالعه شامل وزن نوزاد به کیلوگرم، قد نوزاد به سانتی متر، دور سر نوزاد به سانتی متر می‌باشد. متغیرهای مستقل شامل سن مادر به سال، سن پدر به سال، شاخص توده‌ی بدنی (BMI) مادر، شاخص توده‌ی بدنی (BMI) پدر، جنسیت نوزاد (پسر یا دختر)، نوع شیر تغذیه نوزاد (شیر مادر یا شیر خشک یا شیر مادر و شیر خشک)، وضعیت زایمان (طبیعی یا سزارین)، وضعیت اقتصادی (برحسب درآمد ماهیانه، ۸۰۰ هزار تا ۱میلیون و ۱۰۰ هزار تومان یا ۱ میلیون و ۱۰۰ هزار تا ۱ میلیون ۴۰۰ هزار تومان یا ۱ میلیون و ۴۰۰ هزار تومان و بیش از آن)،

[36] Green
[37] Skuladottir
[38] Victora

تعداد دفعات شیردهی در شبانه روز، مدت زمان حاملگی (به هفته)، رتبه تولد، میانگین مصرف گروه نان و غلات توسط مادر در هفته (تعداد سهم)، میانگین مصرف گروه شیر و لبنیات توسط مادر در هفته (تعداد سهم)، میانیگن مصرف گروه گوشت، حبوبات، تخم مرغ و مغزها توسط مادر در هفته (تعداد سهم)، میانگین مصرف گروه میوه‌ها توسط مادر در هفته (تعداد سهم)، میانگین مصرف گروه سبزیجات توسط مادر در هفته (تعداد سهم)،میانگین مصرف گروه متفرقه مواد غذایی توسط مادر در هفته (تعداد سهم) است. جدول ۱، میزان همبستگی اسپیرمن دو به دوی متغیرهای وزن، قد و دور سر را به همراه مقدار احتمال آزمون همبستگی آن‌ها نشان می‌دهد. با توجه به مقادیر احتمال می‌توان گفت همبستگی دو به دو متغیرهای در سطح ۰۰۱/۰ معنی دار است و با توجه به مقادیر همبستگی ارتباط خطی بین متغیرها قوی و مستقیم است.

جدول ۱: ضریب همبستگی و مقدار احتمال آزمون اسپیرمن دو به دو متغیرهای وابسته

	وزن-قد	وزن-دور سر	قد-دور سر
ضریب هم بستگی	۹۲۲/۰	۹۱۸/۰	۹۱۱/۰
P-مقدار	<۰۰۰۱/۰	<۰۰۰۱/۰	<۰۰۰۱/۰

انتخاب توزیع مناسب برای حاشیه‌ها

با توجه به جدول ۲، انتخاب توزیع نرمال برای وزن (در سطح معنی داری، $P = ۱/۰$) برای ۶ زمان اندازه‌گیری وزن مناسب است ولی برای قد و دور سر مناسب نیست.

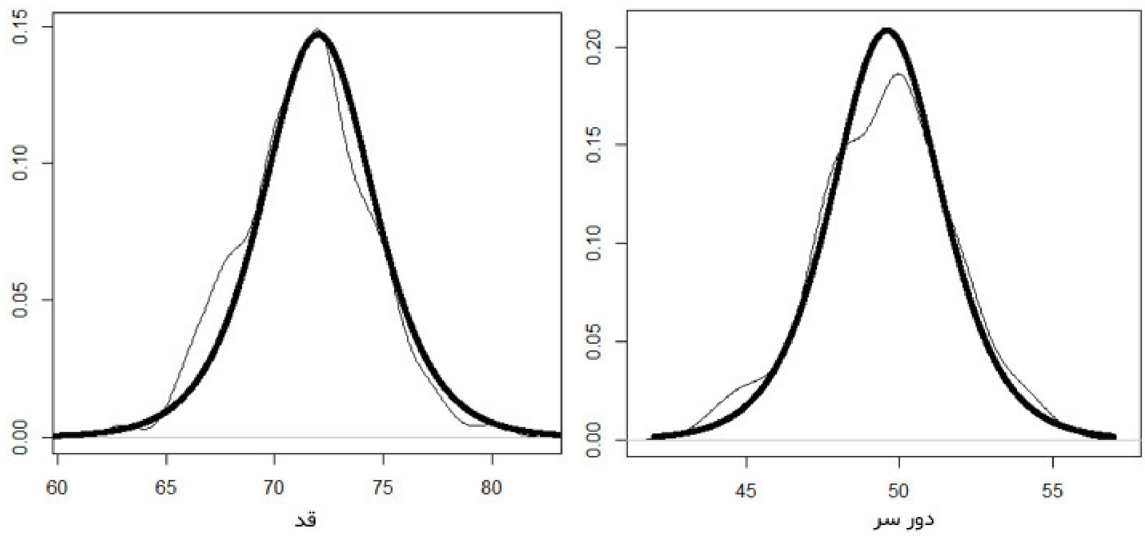
جدول ۲: مقدار احتمال آزمون کلموگروف اسمیرونف برای فرض نرمال بودن متغیرهای وابسته

	وزن	قد	دور سر
بدو تولد	۲۰/۰	< ۰۰۰۱/۰	< ۰۰۰۱/۰
۲ ماهگی	۱۹/۰	< ۰۰۰۱/۰	< ۰۰۰۱/۰
۴ ماهگی	۲۰/۰	< ۰۰۰۱/۰	< ۰۰۰۱/۰
۶ ماهگی	۱۰/۰	۰۰۷/۰	< ۰۰۰۱/۰
۷ ماهگی	۲۰/۰	۰۸۳/۰	< ۰۰۰۱/۰
۹ ماهگی	۲۰/۰	۰۰۵/۰	< ۰۰۰۱/۰

جهت یافتن توزیع مناسب برای این دو متغیر از نرم افزار Easyfit استفاده شده است. از آنجایی که مشاهدات در طول زمان اندازه‌گیری شده‌اند هر متغیر را در زمان‌های مختلف بررسی کرده و در نهایت توزیعی برای متغیر حاشیه‌ای انتخاب شده است که در اکثر زمان‌های بررسی، برازش بهتری به مشاهدات داشته‌اند. توزیع لجستیک به عنوان توزیع مناسب برای هر دو متغیرهای قد و دور سر انتخاب شده است. جدول ۳ ، مقدار احتمال آزمون کلموگروف اسمیرنوف را برای برازش توزیع لجستیک به این دو متغیر نشان می‌دهد که با مقایسه مقادیر احتمال جدول ۲ با جدول ۳ می‌توان گفت توزیع لجستیک نسبت به توزیع نرمال برازش مناسب‌تری داشته است. در شکل ۲ برآورد ناپارامتری تابع چگالی برای مقادیر استاندارد شده دور سر و قد به همراه توزیع لجستیک برازش شده به آنها آورده شده است که نشان دهنده مناسب بودن برازش توزیع لجستیک به داده‌ها می‌باشد.

جدول ۳: مقدار احتمال آزمون کلموگروف اسمیرونف برای برازش توزیع لجستیک به متغیرهای قد و دور سر

	قد	دور سر
بدو تولد	۰۹/۰	۰۸/۰
۲ ماهگی	۱۰/۰	۰۷/۰
۴ ماهگی	۰۸/۰	۰۶/۰
۶ ماهگی	۱۲/۰	۱۳۳/۰
۷ ماهگی	۲۰/۰	۱۶/۰
۹ ماهگی	۱۳/۰	۱۱/۰



شکل ۲: برآورد ناپارامتری تابع چگالی برای مقادیر استاندارد شده دور سر و قد (منحنی کم رنگ) در مقایسه با چگالی استاندارد شده توزیع لجستیک (منحنی پر رنگ)

انتخاب بهترین تابع مفصل برای برازش مدل‌های دو متغیره طولی

مدل‌های دو متغیره با استفاده از توابع مفصل نرمال، فرانک، کلایتون، FGM و حاصلضرب برای متغیرهای قد، وزن و دور سر به صورت دو به دو برازش داده شد که مقدار معیار اطلاع آکائیک (AIC) تمام این مدل‌ها در جدول ۴ آمده است.

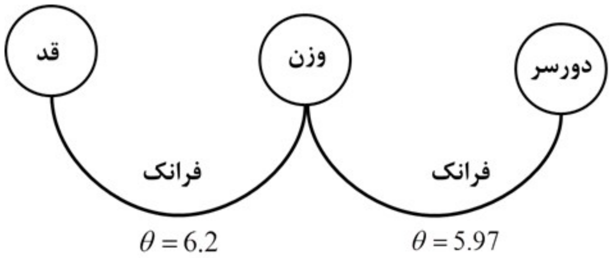
جدول ۴: معیار آکائیک (AIC) برای مدل‌های دو متغیره با توابع مفصل مختلف

تابع مفصل	قد–وزن	وزن–دور سر	قد–دور سر
فرانک	۷/۴۲۵۰	۱/۳۴۵۶	۷/۴۹۳۰
نرمال	۲/۴۲۵۴	۹/۳۴۹۲	۶/۴۹۴۲
کلایتون	۷/۴۳۲۰	۲/۳۵۳۱	۷/۴۹۶۳
FGM	۱/۴۳۸۹	۸/۳۵۹۸	۰/۵۰۲۶
حاصلضرب	۹/۴۵۷۷	۰/۳۸۰۰	۰/۵۱۹۲

مقدار کمتر این شاخص نشان دهنده برازش بهتر مدل است. بنابراین تابع مفصل فرانک برای برازش مدل‌های دو متغیره طولی نسبت به سایر توابع مفصل مناسب‌تر بوده است. همچنین با توجه به جدول ۴، بیشترین مقدار شاخص آکائیک برای تابع مفصل حاصلضرب و برای هر سه مدل دو متغیره است. لذا می‌توان گفت مدل سازی با در نظر گرفتن وابستگی میان متغیرها مناسب تر از مدل سازی براساس فرض استقلال بوده است.

همانطور که در بخش سوم بیان شد با استفاده از روش واین می‌توان مدل‌ها با مفصل‌های دو متغیره را به هم پیوند داد و مدل‌های چند متغیره را ساخت. در این تحقیق با استفاده از رابطه (۵.۳) مدل سه متغیره را به روش واین برای برازش داده‌های طولی رشد نوزاد مورد استفاده قرار گرفت. از آن جایی که طبق جدول ۴ بهترین مفصل برای مدل‌های دو متغیره مفصل فرانک بوده است. بنابراین مدل سه متغیره به کمک روش واین را می‌توان به صورت شکل ۱ نشان داد.

در شکل ۳ متغیر میانی از آن جهت وزن انتخاب شده است که طبق جدول ۱، همبستگی بین وزن با دو متغیر دیگر بیشتر از سایر همبستگی‌های دو گانه است که در اینجا θ برآورد پارمتر وابستگی مفصل فرانک برای مدل‌های دو متغیره است. برازش مدل سه متغیره هم به روش واین و هم به روش تابع مفصل نرمال سه متغیره انجام شد، مقدار معیار آکائیک روش واین ۸/۶۰۸۲ و برای مدل سه متغیره حاصل از تابع مفصل نرمال ۹/۶۰۰۸ برآورد شده است. بنابراین مدل حاصل از تابع مفصل سه متغیره نرمال برای برازش مدل‌ها مناسب‌تر است. در ادامه برآورد پارمترها با استفاده از هر دو روش ارایه می‌شود.

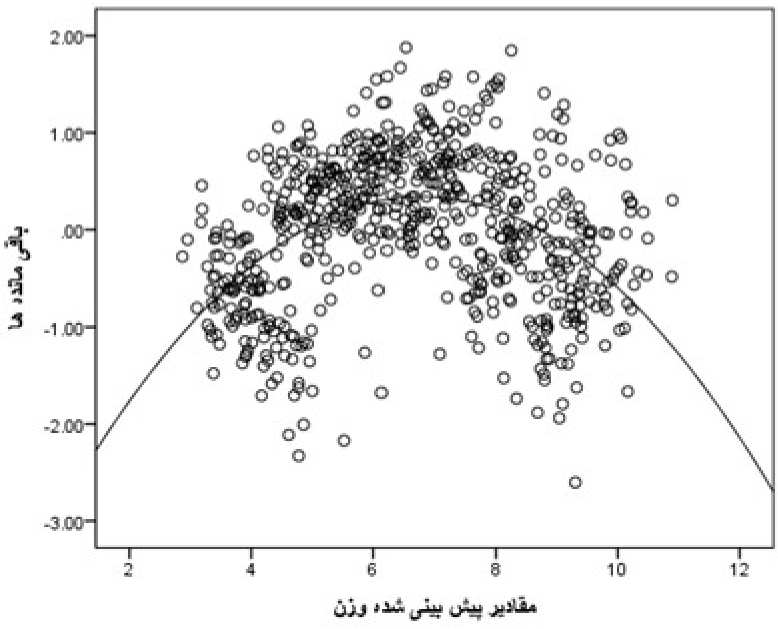


شکل ۳: مدل واین سه متغیره با استفاده از مفصل فرانک

برآورد پارامترها

برآورد پارامترهای مدل در سه مرحله انجام گرفت. در مرحله اول همه متغیرهای مستقل وارد مدل سه متغیره شده و متغیرهایی که مقادیر احتمال آن‌ها کمتر از ۰/۰۵ بود در مدل نگه داشته شد و سایر متغیرهای غیر معنی‌دار از مدل حذف شدند. بنابراین برای وزن، متغیرهای جنسیت (P=۰/۰۲۶) و تعداد دفعات شیردهی در شبانه روز (P=۰/۰۳۱)، برای قد، متغیرهای جنسیت (P<۰/۰۰۱)، سن مادر (P<۰/۰۰۱)، سن پدر (P<۰/۰۰۱)، مصرف گروه سوم موادغذایی توسط مادر (P=۰/۰۲) و وضعیت اقتصادی خانواده (P=۰/۰۳۴)، برای دور سر، متغیرهای جنسیت (P<۰/۰۰۰۱)، سن مادر (P=۰/۰۰۳) ، سن پدر (P=۰/۰۰۲)، مصرف گروه اول موادغذایی توسط مادر (P=۰/۰۳۵) و تعداد دفعات شیردهی توسط مادر (P=۰/۰۰۲) در این مرحله در مدل ماندند و سایر متغیرها (در سطح معنی‌داری ۰/۰۵) روی شاخص‌های رشد کودک موثر نبوده و از مدل حذف شدند.

با توجه به شکل ۴، ۵ و ۶ که نمودار مقادیر پیش‌بینی شده[39] وزن، قد و دور سر را در مقابل مقادیر باقی مانده‌های[40] مدل سه متغیره طولی با مفصل نرمال نشان می‌دهد. می‌توان گفت بین متغیرهای مستقل و متغیرهای وابسته یک رابطه درجه دوم نیز می‌تواند وجود داشته باشد.

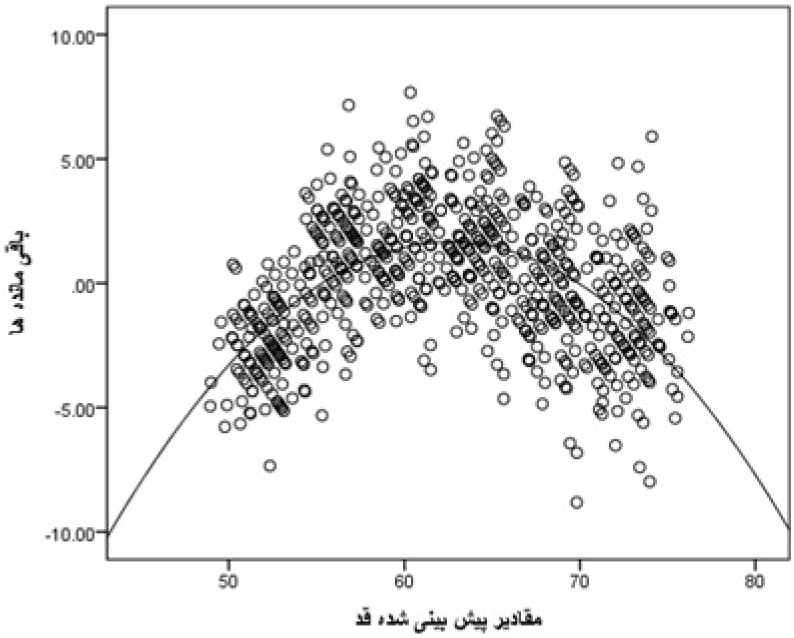


شکل ۴: نمودار مقادیر پیش بینی شده وزن در مقابل باقی مانده‌های وزن در مدل سه متغیره طولی

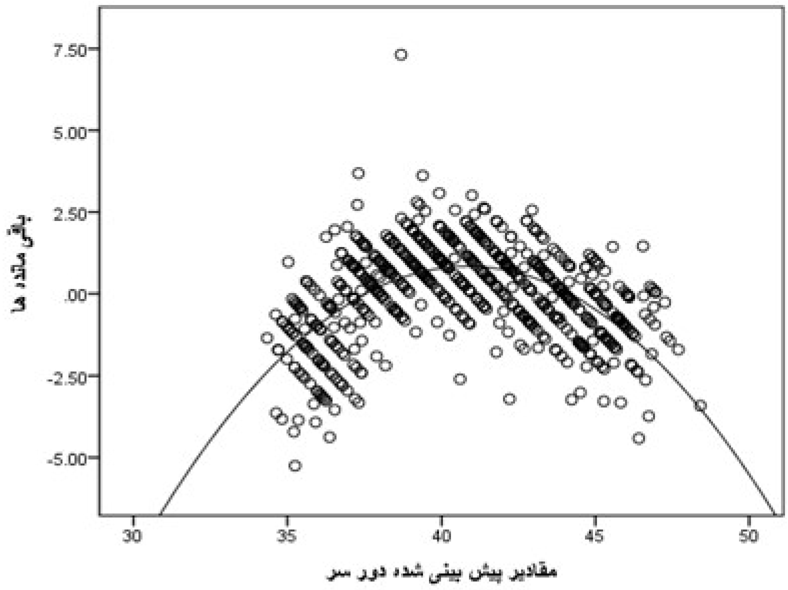
بنابراین در مرحله دوم، کلیه متغیرهای مستقل کمی به صورت توان دوم نیز وارد مدل شده و آن دسته از متغیرهایی که توان دوم آن‌ها در مدل معنی دار بودند به مدل اضافه شدند که در این مرحله توان دوم متغیرهای مصرف گروه اول مواد غذایی توسط مادر برای متغیر دورسر (P=۰/۰۰۶) و توان دوم مصرف گروه سوم مواد غذایی توسط مادر (P=۰/۰۰۵) به مدل اضافه شدند.

در مرحله سوم، مدل سه متغیره با متغیرهای مستقل معنی دار از مرحله اول و دوم برازش شد که نتایج آن در جدول ۵ آمده است. با توجه

[39]Predicted values
[40]Residual values



شکل ۵: نمودار مقادیر پیش بینی شده قد در مقابل باقی مانده‌های قد در مدل سه متغیره طولی



شکل ۶: نمودار مقادیر پیش بینی شده دور سر در مقابل باقی مانده‌های دور سر در مدل سه متغیره طولی

به اینکه مدل سه متغیره حاصل از تابع مفصل نرمال نسبت به روش واین از معیار آکائیک کمتری برخوردار بود تفسیر نتایج براساس این مدل مطرح می‌شود.

نوع تغذیه با شیر مادر نسبت به شیر خشک بر شاخص‌های رشد وزن (P=۰/۶۴۵)، قد (P=۰/۳۹۲) و دور سر (P=۰/۱۹۵) تاثیری نداشت. همچنین نوع تغذیه با شیر مادر و شیر خشک نسبت به شیر خشک به تنهایی، بر رشد شاخص‌های وزن (P=۰/۴۷۴)، قد (P=۰/۴۶۲) و دور سر (P=۰/۲۵۳) تاثیری نشان نداد. به همین خاطر از مدل نهایی حذف شدند.

با توجه به جدول ۵، می‌توان گفت رشد وزنی پسران بهتر از دختران است (P=۰/۰۲۸). به ازای یک واحد افزایش در تعداد دفعات شیردهی توسط مادران با فرض ثابت نگه داشتن سایر متغیرها رشد وزنی نوزاد ۰/۰۱۸ کیلوگرم بهتر می‌شود (P=۰/۰۲۱).

رشدی قدی پسران بهتر از دختران است (P<۰/۰۰۰۱). به ازای افزایش یک سال سن مادران در شروع بارداری رشد قدی نوزادان به میزان ۰/۰۷۱ کمتر می‌شود (P=۰/۰۰۱) و به ازای یک واحد در افزایش سن پدران ۰/۰۹۶ سانتی متر رشد قدی نوزادان بهتر می‌شود (P<۰/۰۰۰۱). نوزادانی که در خانورهای با درآمد کمتر از این مقدار یک میلیون ۴۰۰ هزار تومان زندگی می‌کنند نسبت به نوزادان

جدول ۵: برآورد پارمتر مدل نهایی برای هر دو روش مفصل سه متغیره و واین

متغیرهای مستقل			مدل با مفصل سه متغیره نرمال			مدل به روش واین		
			برآورد	انحراف استاندارد	مقدار احتمال	برآورد	انحراف استاندارد	مقدار احتمال
وزن	عرض از مبدا		۸۲۷/۰	۴۵۲/۰	<۰۰۰۱/۰	۷۷۱/۰	۴۱۵/۰	۰۰۶/۰
	جنسیت	پسر دختر	۲۶۰/۰	۱۱۷/۰	۰۲۸/۰	۲۵۰/۰	۱۱۵/۰	۰۳۲/۰
	تعداد دفعات شیردهی		۰۱۸/۰	۰۰۸/۰	۰۲۱/۰	۰۲۱/۰	۰۰۷/۰	۰۰۴/۰
قد	عرض از مبدا		۸۳۰/۳۵	۰۰۳/۴	<۰۰۰۱/۰	۲۵۱/۳۶	۰۵۴/۴	<۰۰۰۱/۰
	مصرف گروه سوم مواد غذایی توسط مادر		-۵۵۴/۳	۸۱۷/۱	۰۴۱/۰	-۵۵۶/۴	۸۴۳/۱	۰۱۵/۰
	توان دوم		۵۳۲/۰	۲۸۵/۰	۰۲۱/۰	۶۸۱/۰	۲۸۹/۰	۰۲۰/۰
	سن مادر		-۰۷۱/۰	۰۲۱/۰	۰۰۱/۰	-۰۷۳/۰	۰۲۱/۰	۰۰۱/۰
	سن پدر		۰۹۶/۰	۰۲۲/۰	<۰۰۰۱/۰	۰۸۵/۰	۰۲۲/۰	۰۰۰/۰
	جنسیت	پسر دختر	۳۹۸/۱	۲۴۶/۰	<۰۰۰۱/۰	۳۲۳/۱	۲۳۶/۰	<۰۰۱/۰
	درآمد ماهیانه (تومان)	۸۰۰۰۰۰>	-۹۱۲/۰	۳۰۱/۰	۰۰۳/۰	-۸۸۳/۰	۳۰۷/۰	۰۰۵/۰
		۱۱۰۰۰۰۰-۸۰۰۰۰۰	-۰۶۱/۰	۳۴۱/۰	۸۵۹/۰	-۲۱۶/۰	۳۵۹/۰	۵۴۸/۰
		۱۴۰۰۰۰۰-۱۱۰۰۰۰۰	-۱۷۷/۱	۴۱۶/۰	۰۰۶/۰	-۰۳۲/۱	۴۱۷/۰	۰۱۵/۰
		۱۴۰۰۰۰۰<						
دور سر	عرض از مبدا		۴۴۴/۳۱	۸۷۳/۱	<۰۰۰۱/۰	۳۶۴/۲۹	۹۴۸/۱	<۰۰۰۱/۰
	مصرف گروه اول مواد غذایی توسط مادر		-۰۵۵/۳	۰۰۸/۱	۰۰۳/۰	-۴۳۱/۳	۰۹۷۹.	۰۰۱/۰
	توان دوم		۴۹۶/۰	۱۶۴/۰	۰۰۳/۰	۵۳۳/۰	۰/۵۵۹	۰۰۱/۰
	سن مادر		-۰۲۲/۰	۰۱۰/۰	۰۲۹/۰	-۰۲۳/۰	۰۰۹/۰	۰۱۵/۰
	سن پدر		۰۳۰/۰	۰۱۱/۰	۰۰۶/۰	۰۳۱/۰	۰۱۰/۰	۰۰۲/۰
	جنسیت	پسر دختر	۹۵۹/۰	۱۵۱/۰	<۰۰۱/۰	۸۹۰/۰	۰/۱۴۷	<۰۰۱/۰
	تعداد دفعات شیردهی		۰۳۴/۰	۰۱۲/۰	۰۰۶/۰	۰۴۳/۰	۰۱۳/۰	۰۰۱/۰

از خانورهای با درآمد بیش از یک میلیون و چهارصد هزار تومان در ماه نسبت رشد قدی کمتری دارند (برای کمتر از ۸۰۰۰۰۰ تومان درماه (P=۰۰۳/۰) و برای بین ۱۱۰۰۰۰۰ تا ۱۴۰۰۰۰۰ تومان در ماه (P=۰۰۶/۰). اگر مادری به جای مصرف چهار سهم از گوشت، حبوبات، تخم مرغ و مغزها روزی پنج سهم مصرف کند نوزاد او به طور میانگین ۲۳۴/۱ سانتی متر رشد قدی بهتری خواهد داشت (P=۰۴۱/۰) و درجه دوم این متغیر (P=۰۲۱/۰).

رشد دور سر پسران بهتر از دختران است (P<۰۰۰۱/۰). به ازای افزایش یک سال سن مادران در شروع بارداری رشد قدی نوزادان به میزان ۰۲۲/۰ کمتر می‌شود (P=۰۲۹/۰) و به ازای یک واحد در افزایش سن پدران ۰۳/۰ سانتی متر رشد قدی نوزادان بهتر می‌شود (P=۰۰۶/۰). به ازای هر واحد افزایش در تعداد دفعات شیردهی توسط مادران با فرض ثابت نگه داشتن سایر متغیرها رشد دورسر نوزاد ۰۳۴/۰ سانتی متر بهتر می‌شود (P=۰۰۶/۰). اگر مادری به جای مصرف چهار سهم از نان و غلات روزی پنج سهم مصرف کند نوزاد او به طور میانگین ۴/۱ سانتی متر رشد قدی بهتری خواهد داشت (P=۰۰۳/۰) برای عبارت خطی و برای ضریب درجه دوم این متغیر (P=۰۰۳/۰) بدست آمده است.

۵ بحث و نتیجه‌گیری

در این مطالعه با استفاده از تابع مفصل نرمال، فرانک، کلایتون و FGM به مدل‌سازی دو متغیره مشاهدات طولی پرداخته و سپس مدل‌های دومتغیره به کمک روش واین به مدل‌های سه متغیره طولی تعمیم داده شد و در ادامه با مدل حاصل از تابع مفصل سه متغیره نرمال مقایسه شد. استفاده از توابع مفصل جفتی واین موجب می‌شود که علاوه بر انعطاف‌پذیری در انتخاب توزیع‌های مختلف برای حاشیه‌ها، متغیرهای پاسخ بتوانند به‌صورت دو به دو ساختارهای همبستگی متفاوتی را اختیار کنند و بهترین ساختار همبستگی بین متغیرها برای مدل‌سازی داده‌ها مورداستفاده قرار گیرد (روسکن و همکاران، ۲۰۱۷). علی‌رغم اینکه در این مطالعه مدل سه متغیره با تابع مفصل نرمال به‌عنوان مدل نهایی معرفی شد، در بسیاری از کاربردها خصوصاً مطالعات علوم پزشکی با وجود تعداد زیاد متغیرهای وابسته روش واین به علت ساختار ساده‌تر و انعطاف‌پذیری بیشتر در ساخت مدل چندمتغیره می‌تواند نسبت به ساختارهای پیچیده‌تر مورد توجه قرار گیرد (بیراکدر[41]، ۲۰۱۷). در این مطالعه کاربردی از روش واین جهت توسعه مدل با بیش از دو متغیر پاسخ وابسته ارائه شد. همچنین استفاده از آن را برای سایر مطالعات کاربردی پیشنهاد می‌کنیم. از مزیت‌های به‌کارگیری تابع مفصل در مطالعات طولی می‌توان به این مورد اشاره کرده که، توسعه مدل‌هایی با چند متغیر پاسخ و با اندازه‌گیری طولی با استفاده از روش‌های کلاسیک مانند رگرسیون چند متغیره طولی، نسبت به استفاده از تابع مفصل

[41] Bairakdar

بسیار پیچیده‌تر خواهد بود. علی‌رغم اینکه مفروضاتی مانند متغیر پاسخ با توزیع احتمالی خاص، در مدل‌های کلاسیک نشان از انعطاف‌پذیری کمتر این مدل‌ها نسبت به استفاده از تابع مفصل خواهد بود(شی، ۲۰۱۸). اگرچه امروزه استفاده از مدل‌های حاشیه‌ای برای پاسخ‌های طولی غیر نرمال و چوله محبوبیت زیادی در بین محققین پیدا کرده است، با توجه به ساختار این مدل‌ها توسعه پاسخ‌های چند متغیره در این مدل‌ها به‌مراتب پیچیده‌تر از به‌کارگیری تابع مفصل یا استفاده از روش واین خواهد بود (سفیدی و همکاران، ۲۰۲۲). تفاوت مطالعه حاضر در قیاس با مطالعات مذکور گذشته، به کارگیری چند تابع مفصل دیگر به غیر از نرمال و ارائه مدل سه متغیره طولی است. در پایان پیشنهاد می‌شود، سایر محققان از اثر تصادفی دارای توزیع غیر نرمال استفاده کرده نتایج آن را با این مطالعه مقایسه کنند. همچنین استفاده از سایر تابع مفصل‌ها به غیر از توابع مفصل به کاربرده شده در این مطالعه می‌تواند مورد توجه قرار گیرد. همچنین مدل به کارگرفته شده در این مطالعه می‌تواند در مطالعات علوم پزشکی غیر از مطالعه رشد نوزاد مانند مطالعات کوهورت و کارآزمایی بالینی نیز به کارگرفته شود.

فهرست منابع

[1] Aas, K. Czado, C. Frigessi, A. Bakken, H., 2009. Pair-copula constructions of multiple dependence. *Insurance: Mathematics and economics*, **2** (44) (2009), pp. 182-98. https://doi.org/10.1016/j.insmatheco.2007.02.001.

[2] Amirhakimi, G., 2015. A longitudinal growth study from birth to maturity for weight, height and head circumference of normal Iranian children compared with western norms: A standard for growth of Iranian children, *Iranian Journal of Medical Sciences*, **28** , pp. 9-16.

[3] Bairakdar, R., 2017. *Modeling Nested Copulas with GLMM Marginals for Longitudinal Data*, Doctoral dissertation, Concordia University.

[4] Bedford, T. Cooke, R. M., 2002.Vines: A new graphical model for dependent random variables, *Annals of Statistics* , **30**(4), pp. 1031 - 1068. doi: 10.1214/aos/1031689016.

[5] Behrman, R., Kliegman, R. and Jenson H., 2000. *Nelson textbook of pediatrics*, 16th Edition, WB Sauders Company, Philadelphia.

[6] De Leon, A. Zhu Y., 2008. ANOVA extensions for mixed discrete and continuous data, *Computational Statistics& Data Analysis*, **52**, pp. 2218-2227. https://doi.org/10.1016/j.csda.2007.07.018.

[7] De Leon, A.R. Wu, B., 2011. Copula-based regression models for a bivariate mixed discrete and continuous outcome, *Statistics in medicine*, **2**(30), pp. 175-85. https://doi.org/10.1002/sim.4087.

[8] Diggle, P., 2000. *Analysis of longitudinal data*, Oxford University Press, United States.

[9] Fitzmaurice, G. Davidian, M. Verbeke, G. Molenberghs, G., 2008. *Longitudinal data analysis*, CRC Press, United States.

[10] Genest, C. MacKay, J., 1986. The joy of copulas: bivariate distributions with uniform marginals,*The American Statistician*,**4** (40), pp. 280-3. https://doi.org/10.1080/00031305.1986.10475414.

[11] Genest, C. Nešlehová, J., 2007. A primer on copulas for count data, *Astin Bulletin*, **2** (37), pp. 475-515. https://doi.org/10.2143/AST.37.2.2024077.

[12] Green, C. J., 2001. Fibre in enteral nutrition, *Clinical Nutrition*, **20**, pp. 23-39. https://doi.org/10.1054/clnu.2001.0425.

[13] Gueorguieva, R.V. Agresti A., 2001. A correlated probit model for joint modeling of clustered binary and continuous responses, *Journal of the American Statistical Association*, **455**, pp. 1102-12. https://doi.org/10.1198/016214501753208762.

[14] Jiryaie, F. Withanage, N. Wu, B. de Leon, A., 2016. Gaussian copula distributions for mixed data, with application in discrimination, *Journal of Statistical Computation and Simulation*, **9** (86), pp. 1643-1659. https://doi.org/10.1080/00949655.2015.1077386.

[15] Joe, H., 1997. *Multivariate models and multivariate dependence concepts*, CRC Press, United States.

[16] Johnson, R.A. Wichern, D.W. , 2002. *Applied multivariate statistical analysis*, Prentice hall Upper Saddle River, NJ, United States.

[17] Kim, J. M., Liao, S. M., Jung, Y. S., 2007. Mixture of D-vine copulas for modeling dependence, *Computational Statistics & Data Analysis*, **64**, pp. 1-19. https://doi.org/10.1016/j.csda.2013.02.018.

[18] Kole, E. Koedijk, K. Verbeek, M. , 2007. Selecting copulas for risk management, *Journal of Banking & Finance*, **8** (31), pp. 2405-2423. https://doi.org/10.1016/j.jbankfin.2006.09.010.

[19] Kolev, N. Paiva, D., 2009. Copula-based regression models: A survey, *Journal of statistical planning and inference*, **11** (139) , pp. 3847-3856. https://doi.org/10.1016/j.jspi.2009.05.023.

[20] Laird, N. M. Ware, J. H., 1982. Random-effects models for longitudinal data, *Biometrics*, **38** (4), pp. 963-974.

[21] Lee, Y. Nelder, J. A., 1996. Hierarchical generalized linear models, *Journal of the Royal Statistical Society Series B (Methodological)*, **58** (4), pp. 619-656. https://doi.org/10.1111/j.2517-6161.1996.tb02105.x.

[22] Lin, H., 2020. *D-Vine Pair-Copula Models for Longitudinal Binary Data*, Doctoral dissertation, Old Dominion University.

[23] Nelsen R. B., 1991. *Copulas and association. Advances in probability distributions with given marginals*, Springer, New York.

[24] Nelsen, R., 2006. *An introduction to copulas*, Springer, New York.

[25] Nikoloulopoulos, A.K. Joe, H., 2015. Factor copula models for item response data, *Psychometrika*, **80**, pp. 126-130

[26] Pinheiro, J.C. Chao, E.C., 2006. Efficient Laplacian and adaptive Gaussian quadrature algorithms for multilevel generalized linear mixed models, *Journal of Computational and Graphical Statistics*, **15**, pp. 58-81. https://doi.org/10.1198/106186006X96962.

[27] Radice, R. Marra, G. Wojtys, M., 2013. Copula Regression Spline Models for Binary Outcomes With Application in Health Care Utilization, *Statistics and Computing*, **26**, pp.981-995. doi:10.1007/s11222-015-9581-6.

[28] Reyhani, T. Ajam, M., 2000. The comparative study of the 0-6 month childeren growth curve using formula and breast ffeding in Gonabad city, *Iranian Academic Center for Education, Culture and Research*, **6**, pp. 49-55.

[29] Roy, M.,*Conditional Dependence in Joint Modelling of Longitudinal Non-Gaussian Outcomes*, University of Calgary, 2016.

[30] Nai Ruscone, M., & Osmetti, S. A, 2017. Modelling the dependence in multivariate longitudinal data by pair copula decomposition,*InSoft methods for data science*, **27**, 373-380. doi:10.1007/978-3-319-42972-4-46.

[31] Sabeti, A. Wei, M. Craiu, R. V., 2014. Additive models for conditional copulas,*Stat*, **3** (1), pp. 300-312. https://doi.org/10.1002/sta4.64.

[32] Sefidi, S., Ganjali, M., 2022. Analysis of ordinal and continuous longitudinal responses using pair copula construction, *METRON*, **80**, pp. 255-280. doi:10.1007/s40300-022-00231-2.

[33] Shi, P., & Zhao, Z. 2018. Predictive modeling of multivariate longitudinal insurance claims using pair copula construction, *arXiv preprint arXiv:1805.07301.*.

[34] Sklar, A., 1959. Distribution functions of n dimensions and margins, *Publications of the Institute of Statistics of the University of Paris*, **8**, 229-31.

[35] Skuladottir, A. Thome, M. Ramel, A., 2005. Improving day and night sleep problems in infants by changing day time sleep rhythm: a single group before and after study, *International journal of nursing studies*, **42**, pp. 843-850. doi: 10.1016/j.ijnurstu.2004.12.004.

[36] Song, P. X. K, Li, M. Yuan, Y., 2009. Joint regression analysis of correlated data using Gaussian copulas,*Biometrics*, **65**, pp. 60-8. doi:10.1016/j.ijnurstu.2004.12.004.

[37] Toutounchi, P., 2009. The weight to age growth chart in 5 years old children and its risk factors in Tehran, Iran, *Iranian Academic Center for Education, Culture and Research*, **8**, pp. 67-73. doi:20.1001.1.16807626.1387.8.1.9.6.

[38] Trivedi, P. K. Zimmer, D.M., 2007. *Copula modeling: an introduction for practitioners*, Now Publishers Inc, United States.

[39] Victora, C.G. Bahl, R. Barros, AJ. França, G.V. Horton, S. Krasevec, J., 2016. Breastfeeding in the 21st century: epidemiology, mechanisms, and lifelong effect, *The Lancet*, **387**, pp. 475-90. doi:https://doi.org/10.1016/S0140-6736(15)01024-7.

[40] Withanage, N. N. K. P. , 2013. *Methods and Applications in the Analysis of Correlated Non-Gaussian Data*, University of Calgary.

[41] Wu, B. de Leon, A. R., 2014. Gaussian copula mixed models for clustered mixed outcomes, with application in developmental toxicology, *Journal of Agricultural, Biological, and Environmental Statistics*, **19**, pp. 39-56. doi:10.1007/s13253-013-0155-9.

[42] Xue Kun Song, P., 2000. Multivariate dispersion models generated from Gaussian copula, *Scandinavian Journal of Statistics*, **27**, pp. 1534-1561. https://doi.org/10.1111/1467-9469.00191.

[43] Yang, L., Czado, C., 2022. Two part Dvine copula models for longitudinal insurance claim data, *Scandinavian Journal of Statistics*, **49**, pp. 305-320. https://doi.org/10.1111/sjos.12566.

[44] Zimmer, D.M. Trivedi, P. K., 2006. Using trivariate copulas to model sample selection and treatment effects: application to family health care demand, *Journal of Business & Economic Statistics*, **24**, pp. 63-76. https://doi.org/10.1198/073500105000000153.

Modeling Multivariate Longitudinal Data Using Vine Pair Copula Constructions

Mohammad Sadegh Loeloe,(1) Mohammad Reza Akhoond,(2) [1] Kambiz Ahmadi Angali,(1) Fatemeh Borazjani(3)

(1) Department of Biostatistics, Faculty of Health, Ahvaz Jundishapur University of Medical Sciences, Ahvaz, Iran.

(2) Department of Statistics, Mathematical Sciences and Computer Faculty, Shahid Chamran University of Ahvaz, Ahvaz, Iran.

(3) Nutrition and Metabolic Disease Research Center, Ahvaz Jundishapur University of Medical Science, Ahvaz, Iran.

Communicated by: Rahim Chinipardaz

Received: 20 January 2023 Accepted: 23 December 2023

Abstract: In some medical studies, we may have several measurements on each patient. In such a situation, one method is to use random effects in data modeling. Sometimes these longitudinal data may be measured for several response variables, in this case, although the responses can be modeled separately, such an approach reduces the power and efficiency in estimating the effects of auxiliary variables in the response variable. In the analysis of such data, in addition to the analysis of the dependence between repeated measures related to each of the response variables, the dependence between the responses should also be considered. Among the methods that have attracted the attention of many researchers for multivariate data modeling in recent years is data modeling using a copula function. One of the most important advantages of using the copula function compared to the longitudinal multivariate modeling of the data in the classic way is that, in addition to the normal distribution, any other distribution other than the normal can be considered as marginal distributions. Also, marginal distributions can even have different distributions. In situations where the data has a multivariate structure, one of the ways to form multivariate distributions is to use the vine pair-copula function. In this study, we form a multivariate longitudinal structure by using the vine pair copula functions and compare these models with the model obtained from the fitting of the multivariate normal copula function. Then we will introduce the best model using the Akaike information criterion and at the end we will use the presented model on the data of the estimation of the effect of nutrition on growth.

Keywords: Longitudinal measurments, Multivariate normal copula function, Vine pair copula, Infant growth, Infant nutration.

[1]Corresponding author

E-mail address: (M.R. Akhoond) mr.akhoond@scu.ac.ir