مدل توأم بتا - دوجمله‌ای و ترتیبی با اثرهای تصادفی برای تحلیل پاسخ‌های آمیخته‌ی طولی

سیده صدیقه عظیمی [۱]، احسان بهرامی سامانی [۲]

تاریخ دریافت: ۹۸/۱۲/۱۱
تاریخ پذیرش: ۹۹/۳/۱

چکیده:

تحلیل پاسخ‌های آمیخته‌ی گسسته از موضوعات مهم آماری در شاخه‌های مختلف علوم به شمار می‌آید. از جمله متغیرهای گسسته می‌توان به متغیرهای ترتیبی و متغیرهای دوجمله‌ای بیش پراکنده اشاره کرد. داده‌ی دوجمله‌ای بیش‌پراکنده مجموع آزمایش‌های برنولی ناهمبسته با احتمال موفقیت برابر است. در این مقاله مدل توأم با اثرهای تصادفی برای تحلیل پاسخ‌های آمیخته‌ی دوجمله‌ای بیش‌پراکنده و ترتیبی تحت مطالعه‌ی طولی معرفی می‌گردد. در این مدل فرض می‌شود متغیر پاسخ دوجمله‌ای بیش‌پراکنده از توزیع بتا - دوجمله‌ای پیروی می‌کند و از رویکرد متغیر پنهان برای مدل‌بندی متغیر پاسخ ترتیبی استفاده می‌شود. همچنین پارامترهای مدل با روش ماکسیمم درستنمایی برآورد و با استفاده از روش شبیه‌سازی مونت‌کارلویی برآورد پارامترها ارزیابی می‌شود. در نهایت، کاربست مدل معرفی‌شده در داده واقعی بررسی می‌شود.

واژه‌های کلیدی: روش ماکسیمم درستنمایی، متغیر دوجمله‌ای بیش‌پراکنده، توزیع بتا - دوجمله‌ای، رویکرد متغیر پنهان، داده‌ی *BHPS*.

۱ مقدمه

در برخی از مطالعه‌های آماری، بررسی چندین پاسخ به‌طور توأم صورت می‌گیرد. این داده‌ها می‌تواند پیوسته، گسسته، ترتیبی، دودویی و یا آمیخته‌ای از این نوع متغیرها در نظر گرفته شوند. بهرامی سامانی و همکاران [۴] مدلی برای تحلیل پاسخ‌های آمیخته‌ی پیوسته و ترتیبی بر اساس متغیر پنهان معرفی کردند. به‌بیان‌دیگر، آن‌ها روش هکمن [۷] را برای مدل‌بندی پاسخ‌های دو متغیره و چندمتغیره ترتیبی و پیوسته تعمیم دادند. بهرامی سامانی و طهماسبی‌نژاد [۵] مدلی توأم برای پاسخ‌های همبسته‌ی پیوسته، اسمی و ترتیبی ارائه کردند. همچنین بهرامی سامانی [۳] با تعمیم کار بهرامی سامانی و همکاران [۴] پاسخ‌های آمیخته‌ی ترتیبی، اسمی و پیوسته را به صورت توأم با رویکرد متغیر پنهان و با فرض خطاهای همبسته‌ی بیضی‌گون مدل‌بندی کردند.

از سویی دیگر، مدل‌های آماری مختلفی برای تحلیل داده‌های طولی معرفی‌شده است. در این پژوهش‌ها، مدل‌بندی پاسخ‌های آمیخته‌ی گسسته و ترتیبی تحت مطالعات طولی نیز موردعلاقه است. وارین و ژادو [۱۰] مدل پروبیت چندمتغیره با اثر تصادفی را برای داده‌های آمیخته‌ی طولی برنولی و ترتیبی ارائه نمودند و با استفاده از این مدل به تحلیل داده‌های تعیین شدت میگرن پرداختند. رضیئی و همکاران [۹] پاسخ‌های آمیخته‌ی شمارشی و ترتیبی را با رویکرد متغیر پنهان مدل‌بندی کردند.

یکی از موضوع‌های مهمی که اخیراً موردتوجه محققان آماری قرارگرفته است تحلیل داده‌های گسسته‌ای است که از توزیع دوجمله‌ای پیروی می‌کند. داده‌ی دوجمله‌ای از مجموع پیروزی‌ها در m آزمایش برنولی مستقل و هم توزیع پدید می‌آید. در بسیاری از علوم مختلف با داده‌های دوجمله‌ای روبه‌رو هستیم که بین آزمایش‌های برنولی آن‌ها استقلال برقرار نیست. بنابراین این داده‌ها با حذف استقلال بین آزمایش‌های برنولی، دیگر از توزیع دوجمله‌ای پیروی نمی‌کنند. در حالتی که همبستگی بین آزمایش‌های برنولی مثبت است، واریانس داده‌ها از واریانس آن‌ها نسبت به حالتی که استقلال بین آزمایش‌های برنولی وجود دارد، بیشتر خواهد شد. اصطلاحاً به چنین داده‌هایی، داده‌های دوجمله‌ای بیش‌پراکنده می‌گویند. یکی از توزیع‌های مناسب برای مدل‌بندی داده‌های دوجمله‌ای بیش‌پراکنده، توزیع بتا - دوجمله‌ای است. گریفز [۶] برآوردهای ماکسیمم درستنمایی پارامترهای مدل بتا - دوجمله‌ای را معرفی نمود و ویلیام [۱۲] با استفاده از مدل‌های خطی تعمیم‌یافته، برآورد پارامترهای مدل بتا - دوجمله‌ای را با ترکیبی از برآورد گشتاوری پارامترهای بیش‌پراکنش و برآوردهای شبه درستنمایی پارامترهای مدل رگرسیونی به دست آورد. آیکن ([۱، ۲]) داده‌های دوجمله‌ای بیش‌پراکنده را به ترتیب با فرض این‌که

[۱] گروه آمار دانشگاه شهید بهشتی، تهران، ایران
[۲] گروه آمار دانشگاه شهید بهشتی، تهران، ایران. ehsan_bahrami_samani@yahoo.com

اثر تصادفی دارای توزیع نرمال یا توزیع آمیخته‌ی گسسته است، مدل‌بندی کردند. کیم و لی [۸] به بررسی نیکویی برازش مدل بتا - دوجمله‌ای برای تحلیل داده‌های دوجمله‌ای بیش‌پراکنده پرداختند. اخیراً، وو و همکاران [۱۳] یک مدل بتا - دوجمله‌ای با اثرهای تصادفی را برای تحلیل داده‌های طولی دوجمله‌ای بیش‌پراکنده ارائه و با روش معادلات برآوردگر تعمیم‌یافته پارامترهای این مدل را تحت تابع ربط پروبیت برآورد کردند.

در مطالعاتی که تاکنون برای تحلیل داده‌های دوجمله‌ای بیش‌پراکنده انجام‌شده است، پژوهشی برای مدل‌بندی پاسخ‌های آمیخته‌ی دوجمله‌ای بیش‌پراکنده و ترتیبی تحت مطالعات طولی صورت نگرفته است. لذا در این مقاله، مدل توأم با اثرهای تصادفی برای تحلیل پاسخ‌های آمیخته‌ی طولی دوجمله‌ای بیش‌پراکنده و ترتیبی ارائه می‌کنیم. در این مدل فرض می‌شود متغیر پاسخ دوجمله‌ای بیش‌پراکنده از توزیع بتا - دوجمله‌ای پیروی می‌کند و از رویکرد متغیر پنهان برای مدل‌بندی متغیر پاسخ ترتیبی استفاده می‌شود. به همین منظور، در بخش دوم پس از مرور توزیع بتا - دوجمله‌ای به معرفی مدلی جدید برای تحلیل پاسخ‌های آمیخته‌ی طولی دوجمله‌ای بیش‌پراکنده و ترتیبی با استفاده از اثرهای تصادفی می‌پردازیم. در بخش سوم، برآورد ماکسیمم درستنمایی پارامترهای این مدل را با روش شبیه‌سازی مونت‌کارلویی ارزیابی می‌کنیم. همچنین در بخش چهارم، به تحلیل داده‌ی $BHPS$ با استفاده از مدل معرفی‌شده پرداخته و در نهایت در بخش پنجم به بحث و نتیجه‌گیری می‌پردازیم.

۲ مدل و تابع درستنمایی

در این بخش به معرفی مدل توأم برای تحلیل پاسخ‌های آمیخته‌ی طولی دوجمله‌ای بیش‌پراکنده و ترتیبی می‌پردازیم. توزیع بتا - دوجمله‌ای یک توزیع مناسب برای تحلیل داده‌های دوجمله‌ای بیش‌پراکنده است. این توزیع از ترکیب دو توزیع دوجمله‌ای و بتا حادث می‌گردد. لذا ابتدا این توزیع را مرور می‌کنیم، سپس به معرفی مدل جدید پرداخته و تابع درستنمایی این مدل را ارائه می‌کنیم. فرض کنید متغیر تصادفی Y از توزیع بتا - دوجمله‌ای با حداکثر مقدار موفقیت m و پارامترهای μ و ρ پیروی می‌کند. این متغیر تصادفی دارای تابع جرم احتمال $P(Y=y|\mu,\rho)$ به صورت

$$\binom{m}{y}\frac{\beta\left(y+\mu\left(\rho^{-1}-1\right),m-y+(1-\mu)\left(\rho^{-1}-1\right)\right)}{\beta\left(\mu\left(\rho^{-1}-1\right),(1-\mu)\left(\rho^{-1}-1\right)\right)}$$

است که در آن $y=0,1,\dots,m$ و $0<\mu,\rho<1$. میانگین این توزیع $m\mu$ و پارامتر ρ ضریب همبستگی بین آزمایش‌های برنولی با احتمال موفقیت یکسان است که مجموع آن‌ها متغیر دوجمله‌ای بیش‌پراکنده‌ی Y را تشکیل می‌دهد. در ادامه، به معرفی مدل می‌پردازیم. فرض کنید Y_{it} متغیر پاسخ دوجمله‌ای بیش‌پراکنده برای فرد iام $(i=1,\dots,n)$ در زمان tام $(t=1,\dots,T)$ باشد که از توزیع بتا - دوجمله‌ای با حداکثر مقدار موفقیت m و پارامترهای μ_{it} و ρ پیروی می‌کند. همچنین فرض کنید O_{it} متغیر پاسخ ترتیبی با c_i سطح برای فرد iام در زمان t و متغیر پنهان O^*_{it} متناظر با متغیر پاسخ O_{it} است. پاسخ ترتیبی O_{it} را به صورت

$$O_{it}=\begin{cases}1; & \eta_{i0}<O^*_{it}\le\eta_{i1}\\ 2; & \eta_{i1}<O^*_{it}\le\eta_{i2}\\ \vdots & \\ c_i; & \eta_{ic_i-1}<O^*_{it}\le\eta_{ic_i}\end{cases}$$

در نظر بگیرید که در آن η_{i0}، η_{i1}، ...، η_{ic_i-1} و η_{ic_i} پارامترهای نقاط آستانه هستند به طوری که $-\infty\equiv\eta_{i0}<\eta_{i1}<\dots<\eta_{ic_i-1}<\eta_{ic_i}\equiv\infty$. مدل توأم بتا - دوجمله‌ای و ترتیبی با اثرهای تصادفی برای تحلیل پاسخ‌های آمیخته‌ی طولی Y_{it} و O_{it} به صورت

$$\begin{cases}logit(\mu_{it})=X^{(1)'}_{it}\beta_t+W^{(1)'}_{it}b^{(1)}_i,\\ O^*_{it}=X^{(2)'}_{it}\gamma_t+W^{(2)'}_{it}b^{(2)}_i+\varepsilon_{it},\end{cases}\tag{1}$$

برای $i=1,\dots,n$ و $t=1,\dots,T$ تعریف می‌شود که در آن $X^{(1)}_i=\left(X^{(1)'}_{i1},\dots,X^{(1)'}_{iT}\right)$ و $X^{(2)}_i=\left(X^{(2)'}_{i1},\dots,X^{(2)'}_{iT}\right)$ سطرهای ماتریس‌های طرح و $W^{(1)}_{it}$ و $W^{(2)}_{it}$ زیرمجموعه‌ای از ماتریس‌های طرح هستند. همچنین بردارهای $b^{(1)}_i$ و $b^{(2)}_i$ اثرهای تصادفی هستند که برای در نظر گرفتن همبستگی بین پاسخ‌های آمیخته‌ی طولی در مدل لحاظ می‌شود به طوری که $\left(b^{(1)'}_i,b^{(2)'}_i\right)'\overset{iid}{\sim}MVN(\underline{0},\Sigma_b)$ و نتیجه می‌شود که پاسخ‌های Y_{it} و O_{it} به شرط اثرهای تصادفی $b^{(1)}_i$ و $b^{(2)}_i$ از هم مستقل شوند. از سویی دیگر، بردار $(\varepsilon_{i1},\dots,\varepsilon_{iT})'$ بردار خطاهای مدل است که از توزیع نرمال چندمتغیره با بردار میانگین $\underline{0}$ و ماتریس کوواریانس Σ_ε پیروی می‌کند. همچنین اثرهای تصادفی $b^{(1)}_i$ و $b^{(2)}_i$ از بردار خطاهای مدل مستقل هستند. علاوه بر این، برای شناساپذیری مدل فرض می‌شود ماتریس کوواریانس Σ_ε ثابت و بردار γ_t فاقد عرض از مبدأ است (وانگ [۱۱]). لذا پارامترهای مدل ρ، Σ_b، بردارهای β_t و γ_t برای $t=1,\dots,T$ و بردار η_i برای $i=1,\dots,n$ است. برای به دست آوردن برآورد ماکسیمم درستنمایی پارامترهای مدل لازم است تابع درستنمایی داده‌ها ماکسیمم شود. لذا تابع درستنمایی مدل (۱) برابر است با

$$\begin{aligned}L &= \prod_{i=1}^{n}\int\prod_{t=1}^{T}[P(Y_{it}=y_{it},O_{it}=o_{it}|b^{(1)}_i,b^{(2)}_i,\beta_t,\gamma_t\\ &\quad ,\Sigma_b,\rho,\eta_i)f(b^{(1)}_i,b^{(2)}_i|\Sigma_b)]db^{(1)}_i db^{(2)}_i\\ &= \prod_{i=1}^{n}\int[\prod_{t=1}^{T}[P(Y_{it}=y_{it}|b^{(1)}_i,\beta_t,\Sigma_b,\rho)\\ &\times P(O_{it}=o_{it}|b^{(2)}_i,\gamma_t,\Sigma_b,\eta_i)]f(b^{(1)}_i,b^{(2)}_i|\Sigma_b)]db^{(1)}_i db^{(2)}_i\end{aligned}\tag{2}$$

که در آن $P\left(Y_{it}=y_{it}|b_i^{(1)},\beta_t,\Sigma_b,\rho\right)$ تابع جرم احتمال توزیع بتا - دوجمله‌ای با حداکثر مقدار موفقیت m و پارامترهای μ_{it} و ρ و $f\left(b_i^{(1)},b_i^{(2)}|\Sigma_b\right)$ تابع چگالی احتمال بردار $\left(b_i^{(1)'},b_i^{(2)'}\right)'$ هستند. از سویی دیگر،

$$O_{it}^*|b_i^{(2)},\gamma_t,\Sigma_b,\eta_i\sim N\left(\mu_{it}^*,(\Sigma_b)_{it\times it}\right)$$

که در آن $\mu_{it}^*=x_{it}^{(2)'}\gamma_t+w_{it}^{(2)'}b_i^{(2)}$ و $(\Sigma_b)_{it\times it}$ عنصر $it\times it$ ماتریس Σ_b است. بنابراین داریم

$$\begin{aligned} & P\left(O_{it}=o_{it}|b_i^{(2)},\gamma_t,\Sigma_b,\eta_i\right)\\ = & P\left(\eta_{o_{it}-1}<O_{it}^*\leq\eta_{o_{it}}|b_i^{(2)},\gamma_t,\Sigma_b,\eta_i\right)\\ = & \Phi\left(\frac{\eta_{o_{it}}-\mu_{it}^*}{\sqrt{(\Sigma_b)_{it\times it}}}\right)-\Phi\left(\frac{\eta_{o_{it}-1}-\mu_{it}^*}{\sqrt{(\Sigma_b)_{it\times it}}}\right)\end{aligned}$$

که در آن $\Phi(.)$ تابع توزیع تجمعی توزیع نرمال استاندارد است. تابع درستنمایی ارائه‌شده در رابطه‌ی (۲) فرم صریح و مشخصی ندارد و به دلیل وجود انتگرال‌های بدون فرم بسته به راحتی قابل‌حل نیست. لذا برای رفع این مشکل می‌بایست از روش‌های عددی همچون روش مونت‌کارلویی برای تقریب این انتگرال‌ها استفاده شود. علاوه بر این برای محاسبه‌ی خطای استاندارد برآورد ماکسیمم درستنمایی پارامترها از روش عکس ماتریس اطلاع فیشر استفاده می‌شود.

۳ شبیه‌سازی

در این بخش به ارزیابی مدل معرفی‌شده در بخش ۲ با استفاده از روش شبیه‌سازی مونت‌کارلویی می‌پردازیم. به این منظور، فرض کنید Y_{it} پاسخ دوجمله‌ای بیش‌پراکنده برای فرد iام $(i=1,\ldots,n)$ در زمان tام $(t=1,2)$ باشد که از توزیع بتا - دوجمله‌ای با حداکثر مقدار موفقیت m و پارامترهای μ_{it} و ρ پیروی می کند و O_{it}^* متغیر پنهان متناظر با متغیر پاسخ O_{it} با سه سطح است یعنی

$$O_{it}=\begin{cases}1; & -\infty<O_{it}^*\leq\eta_1\\ 2; & \eta_1<O_{it}^*\leq\eta_2\\ 3; & \eta_2<O_{it}^*\leq\infty\end{cases}.$$

اینک مدل

$$\begin{cases} logit(\mu_{it})=\beta_0+\beta_{1t}x_{it}^{(1)}+b_i,\\ O_{it}^*=\gamma_{1t}x_{it}^{(2)}+b_i+\varepsilon_{it},\end{cases}\qquad(3)$$

را در نظر بگیرید که در آن b_i اثر تصادفی است که برای در نظر گرفتن همبستگی بین پاسخ‌های آمیخته‌ی طولی تعریف می‌شود. برای انجام شبیه‌سازی، بردارهای کمکی $(x_{i1}^{(1)},x_{i2}^{(1)})'$ و $(x_{i1}^{(2)},x_{i2}^{(2)})'$ از توزیع نرمال دو متغیره، اثرهای تصادفی b_i از توزیع نرمال با میانگین صفر و واریانس σ_b^2 و خطاهای ε_{it} از توزیع نرمال استاندارد تولیدشده‌اند. نتایج شبیه‌سازی با مقادیر اولیه‌ی $\beta_0=0/500$، $\beta_{11}=1/000$، $\beta_{12}=1/000$، $\gamma_{11}=1/000$، $\gamma_{12}=1/000$، $\sigma_b^2=2/000$، $\rho=0/500$ و نقاط آستانه $(\eta_1,\eta_2)=(-1/000,1/000)$ در جدول آورده شده است. این شبیه‌سازی با تعداد تکرار مونت‌کارلویی ۱۰۰۰، برای حجم نمونه‌ی $n=50,\ 100$ و مقدار حداکثر موفقیت $m=5,10$ انجام‌شده است. نتایج نشان می‌دهد برای حجم نمونه‌ی مختلف برآورد ماکسیمم درستنمایی پارامترها نزدیک مقدار واقعی است و با افزایش حجم نمونه میزان اریبی برآوردها کاهش می‌یابد. همچنین خطای استاندارد و میانگین توان دوم خطا (MSE) با افزارش حجم نمونه کاهش می‌یابد. کاهش خطای استاندارد سازگاری برآوردهای ماکسیمم درستنمایی پارامترها و کاهش MSE افزایش دقت برآوردها را نشان می‌دهد.

۴ تحلیل داده‌ی BHPS

از سال ۱۹۹۱ میلادی، مطالعه‌ی طولی پانل خانگی ($BHPS$[1]) روی افراد بزرگ‌سال بریتانیایی در انگلیس توسط مرکز مطالعات طولی بریتانیا $ESRC$ با انستیتوی تحقیقات اجتماعی و اقتصادی در دانشگاه اسکس هر ساله انجام می‌شود. در این مطالعه، هدف یافتن عوامل مؤثر بر متغیرهای پاسخ آمیخته‌ی کیفیت زندگی و رضایت از زندگی است. به همین منظور، داده‌ها طی سال‌های ۲۰۰۳ و ۲۰۰۴ میلادی در نظر گرفته‌شده است. متغیر پاسخ رضایت از زندگی یک متغیر ترتیبی است: [۱: اصلاً راضی نیست، ۲: به‌طور متوسط راضی است و ۳: کاملاً راضی است]. این متغیر پاسخ به ترتیب برای زمان اول و دوم با نماد O_{i1} و O_{i2} برای $i=1,\ldots,429$ نمایش داده می‌شود. متغیر پاسخ کیفیت زندگی ($NIALQ$) از مجموع دو متغیر برنولی به دست می‌آید: ۱) متغیر برنولی اول از پاسخ به سؤال "در طول یک ماه چه مقدار از درآمد خود را صرف تفریحات و سرگرمی می‌کنید؟" تعریف می‌شود. پاسخ این سؤال (AM) به صورت [۰: هیچی، ۱: پایین‌تر از ۵۰ پوند و ۲: ۵۰ پوند یا بالاتر] طبقه‌بندی شده است. لذا متغیر برنولی اول را به ترتیب برای زمان اول و دوم به صورت زیر در نظر بگیرید

$$Y_{11}=\begin{cases}1; & AM\neq 0\\ 0; & AM=0\end{cases},\quad Y_{21}=\begin{cases}1; & AM\neq 0\\ 0; & AM=0\end{cases}.$$

[1] $BHPS\ datasets: www.esds.ac.uk/findingData/bhpsTitles.asp$

نتایج شبیه‌سازی برای برآورد ماکسیمم درستنمایی پارامترهای مدل توأم (۱)

پارامتر	مقدار واقعی	$m=۵$					
		$n=۵۰$			$n=۱۰۰$		
		$Est.$	$S.E.$	MSE	$Est.$	$S.E.$	MSE
$\beta_۰$	۰/۵۰۰	۰/۵۰۴	۰/۲۲۷	۰/۰۵۸	۰/۴۹۴	۰/۱۵۸	۰/۰۲۹
$\beta_{۱۱}$	۱/۰۰۰	۱/۰۳۶	۰/۳۱۱	۰/۱۰۲	۰/۹۹۹	۰/۲۱۵	۰/۰۴۹
$\beta_{۱۲}$	۱/۰۰۰	۱/۰۲۲	۰/۳۰۹	۰/۰۹۵	۰/۹۹۸	۰/۲۱۵	۰/۰۵۱
$\gamma_{۱۱}$	۱/۰۰۰	۱/۰۲۳	۰/۲۵۴	۰/۰۷۲	۰/۹۹۰	۰/۱۷۴	۰/۰۳۳
$\gamma_{۱۲}$	۱/۰۰۰	۱/۰۲۹	۰/۲۵۴	۰/۰۶۸	۰/۹۹۰	۰/۱۷۴	۰/۰۳۲
σ_b	۲/۰۰۰	۰/۹۸۶	۰/۲۱۷	۰/۰۵۵	۰/۹۸۵	۰/۱۴۷	۰/۰۲۷
ρ	۰/۵۰۰	۰/۴۹۴	۰/۰۶۳	۰/۰۰۴	۰/۵۰۰	۰/۰۴۴	۰/۰۰۲
$\eta_۱$	−۱/۰۰۰	−۱/۰۱۴	۰/۲۳۸	۰/۰۶۶	−۰/۹۸۳	۰/۱۶۴	۰/۰۳۱
$\eta_۲$	۱/۰۰۰	۱/۰۰۵	۰/۲۳۷	۰/۰۶۵	۰/۹۹۲	۰/۱۶۴	۰/۰۳۲

پارامتر	مقدار واقعی	$m=۱۰$					
		$n=۵۰$			$n=۱۰۰$		
		$Est.$	$S.E.$	MSE	$Est.$	$S.E.$	MSE
$\beta_۰$	۰/۵۰۰	۰/۴۹۷	۰/۲۱۷	۰/۰۵۵	۰/۴۹۶	۰/۱۵۰	۰/۰۲۶
$\beta_{۱۱}$	۱/۰۰۰	۱/۰۱۰	۰/۲۸۶	۰/۰۷۹	۰/۹۹۱	۰/۱۹۹	۰/۰۴۱
$\beta_{۱۲}$	۱/۰۰۰	۱/۰۱۹	۰/۲۸۸	۰/۰۸۷	۱/۰۰۲	۰/۲۰۰	۰/۰۳۹
$\gamma_{۱۱}$	۱/۰۰۰	۱/۰۳۸	۰/۲۵۴	۰/۰۶۹	۰/۹۸۸	۰/۱۷۳	۰/۰۲۹
$\gamma_{۱۲}$	۱/۰۰۰	۱/۰۲۶	۰/۲۵۱	۰/۰۶۳	۰/۹۹۱	۰/۱۷۲	۰/۰۳۴
σ_b	۲/۰۰۰	۰/۹۹۶	۰/۲۰۷	۰/۰۵۵	۰/۹۸۰	۰/۱۴۲	۰/۰۲۵
ρ	۰/۵۰۰	۰/۴۹۴	۰/۰۵۲	۰/۰۰۳	۰/۵۰۱	۰/۰۳۷	۰/۰۰۱
$\eta_۱$	−۱/۰۰۰	−۱/۰۰۱	۰/۲۳۶	۰/۰۶۵	−۰/۹۹۰	۰/۱۶۳	۰/۰۲۸
$\eta_۲$	۱/۰۰۰	۱/۰۲۲	۰/۲۳۷	۰/۰۶۵	۰/۹۹۵	۰/۱۶۳	۰/۳۰۰

۲) متغیر برنولی دوم مربوط به لگاریتم میزان درآمد سال گذشته به واحد هزار پوند (INC) است. این متغیر به صورت تابع نشانگر $\{INC > ۳/۸۴۰\}$ و $\{INC > ۳/۶۹۹\}$ به ترتیب برای زمان اول ($Y_{۱۲}$) و زمان دوم ($Y_{۲۲}$) تعریف می‌شود. بنابراین متغیر پاسخ $NIALQ$ به ترتیب برای زمان اول و دوم به صورت $NIALQ_۱ = Y_{۱۱} + Y_{۱۲}$ و $NIALQ_۲ = Y_{۲۱} + Y_{۲۲}$ تعریف می‌شود. $p-$مقدار آزمون‌های فرض صفر ناهمبسته بودن زوج متغیرهای برنولی ($Y_{۱۱}\ Y_{۱۲}$) و ($Y_{۲۱}, Y_{۲۲}$)، به ترتیب برای زوج متغیر اول و دوم ۰/۰۲۱ و ۰/۰۰۸ به دست آمد که نشان می‌دهد این دو زوج متغیر در سطح معنی‌داری ۰/۰۵ همبسته‌اند. از سویی دیگر، احتمال موفقیت این دو زوج متغیر برنولی با هم برابر هستند زیرا با انجام آزمون‌های فرض صفر اینکه احتمال موفقیت این دو زوج برابر هستند، $p-$مقدارهای ۰/۱۴۰ و ۰/۰۸۱ به دست آمد که نشان می‌دهد نمی‌توان فرض‌های صفر را در سطح ۰/۰۵ رد کرد. لذا متغیر پاسخ $NIALQ$ در زمان‌های اول و دوم یک متغیر دوجمله‌ای است که از مجموع دو متغیر برنولی ناهمبسته با احتمال موفقیت برابر به‌دست‌آمده است. بنابراین به دلیل همبستگی مثبت بین دو زوج متغیرهای برنولی، این متغیر پاسخ یک متغیر دوجمله‌ای بیش‌پراکنده است. در این مطالعه، هدف یافتن عوامل مؤثر بر پاسخ‌های آمیخته‌ی کیفیت زندگی و رضایت از زندگی همچون جنسیت [۱: مرد و ۰: زن]، وضعیت تأهل[۱: متأهل، ۲: بیوه، ۳: مطلقه و ۴: مجرد]، سطح تحصیلات [۱: دیپلم، ۲: کارشناسی، ۳: کارشناسی ارشد و ۴: زیردیپلم] و سن است. ۵۵/۳ درصد افراد موردمطالعه زن هستند و متوسط سن افراد ۴۵ سال و خطای استاندارد آن ۱۷ سال است. علاوه بر این حدود ۶۰ درصد افراد متأهل هستند و ۶۱ درصد افراد دارای مدرک کارشناسی هستند. در ادامه مدل

(۱) را به داده‌های $BHPS$ برازش می‌دهیم. به همین منظور، مدل

$$\begin{cases} logit(\mu_{it}) = \beta_۰ + \beta_۱ Gender_i + \beta_{۲t} Age_{it} \\ \qquad + \beta_{۳t} MS^۱_{it} + \beta_{۴t} MS^۲_{it} + \beta_{۵t} MS^۳_{it} \\ \qquad + \beta_۶ HEQ^۱_i + \beta_۷ HEQ^۲_i + \beta_۸ HEQ^۳_i + b_i, \\ O^*_{it} = \gamma_۱ Gender_i + \gamma_{۲t} Age_{it} \\ \quad + \gamma_{۳t} MS^۱_{it} + \gamma_{۴t} MS^۲_{it} + \gamma_{۵t} MS^۳_{it} \\ \quad + \gamma_۶ HEQ^۱_i + \gamma_۷ HEQ^۲_i + \gamma_۸ HEQ^۳_i + b_i + \varepsilon_{it}, \\ \varepsilon_{it} \overset{iid}{\sim} N(۰, ۰/۱۰۰), \quad b_i \overset{iid}{\sim} N(۰, \sigma^۲_b) \end{cases}$$

را در نظر بگیرید که در آن ε_{it} و b_i برای $i = ۱, \ldots, ۴۲۹$ و $t = ۱, ۲$ از هم مستقل هستند و

$$MS^k_{it}\ (HEQ^k_i) = \begin{cases} ۱; & MS_{it}\ (HEQ_i) = k \\ ۰; & MS_{it}\ (HEQ_i) \neq k \end{cases}, \ k = ۱, ۲, ۳.$$

پارامترهای مدل به روش ماکسیمم درستنمایی برآورد شد. این نتایج نشان می‌دهد با فرض ثابت بودن سایر متغیرها، به‌طور متوسط مردان کیفیت زندگی بالاتری نسبت به زنان دارند. همچنین افراد دارای مدرک دیپلم به‌طور متوسط کیفیت زندگی بهتری نسبت به افراد با سایر سطح تحصیلات دارند. از سویی دیگر، با فرض ثابت بودن سایر متغیرها، افراد با مدرک کارشناسی میزان رضایت بیشتری از زندگی خود نسبت به افراد با سایر مدرک‌های تحصیلی دارند. همچنین رضایت زندگی افرادی که همسرشان فوت‌شده در سطح بالاتری قرار دارد. به علاوه با فرض ثابت بودن سایر متغیرها، افراد مطلقه نسبت به افراد با سایر وضعیت‌های تأهل دارای کیفیت زندگی بهتر و سطح رضایت بالاتری از زندگی هستند.

بحث و نتیجه‌گیری

در این مقاله به معرفی مدل توأم برای تحلیل پاسخ‌های آمیخته‌ی دوجمله‌ای بیش‌پراکنده و ترتیبی تحت مطالعات طولی پرداختیم. در این مدل فرض شد پاسخ دوجمله‌ای بیش‌پراکنده از توزیع بتا - دوجمله‌ای پیروی می‌کند. همچنین متغیر پاسخ ترتیبی بر اساس متغیر پنهان مدل‌بندی شد. نتایج شبیه‌سازی نشان می‌دهد که برآورد ماکسیمم درستنمایی پارامترها نزدیک مقدار واقعی به دست می‌آید. همچنین این نتایج سازگاری برآوردهای ماکسیمم درستنمایی را نیز تأیید می‌کند. در نهایت مدل ارائه‌شده را به داده‌ی $BHPS$ برازش داده و عوامل مؤثر بر پاسخ‌های آمیخته‌ی کیفیت زندگی و رضایت از زندگی شناخته شد. برای مطالعات آتی، مدل ارائه‌شده در این مقاله برای تحلیل پاسخ‌های دارای گمشدگی قابل‌تعمیم است.

مراجع

[1] Aitkin, M. (1995). NPML estimation of the mixing distribution in general statistical models with unobserved random variation, In: Seeber, G.U.H., Francis, B.J., Hatzinger, R., Steckel-Berger G. (Eds.), *Statistical Modelling*. Springer, New York.

[2] Aitkin, M. (1996). A general maximum likelihood analysis of overdispersion in generalized linear models, *Statistics and Computing*, **6**, 251-262.

[3] Bahrami Samani, E. (2013). Local influence in Bayesian elliptically contoured-ordinal model for mixed data, *Application and Applied Mathematics: An international Journal (AAM)*, **8**(2), 391-403.

[4] Bahrami Samani, E., Ganjali, M., and Khodadadi, A. (2008). A latent variable model for mixed continuous and ordinal responses, *Journal of Statistical Theory and Applications*, **7**(3), 337-349.

[5] Bahrami Samani, E. and Tahmasebinejad, ZH. (2011). Joint modeling of mixed correlated nominal, ordinal and continuous responses, *Journal of Statistical Research*, **45**(1), 37-47.

[6] Griffiths, D. (1973). Maximum likelihood estimation for the beta-binomial distribution and an application to the household distribution of the total number of cases of a diseases, *Biometrics*, **29**(4), 637-648.

[7] Heckman, J.J.D. (1978). Endogenous variable in a simultaneous equations system, *Econometrica*, **6**, 931-959.

[8] Kim, J. and Lee, J.H. (2015). The validation of a beta-binomial model for overdispersed binomial data, *Communications in Statistics - Simulation and Computation*, **46**(2), 807-814.

[9] Razie, F., Bahrami, S.E., and Ganjali, M. (2016). Latent variable model for mixed correlated power series and ordinal longitudinal responses with nonignorable missing values, *Communications in Statistics - Theory and Methods*, **46**(12), 5738-5753.

[10] Varin, C. and Czado, C. (2010). A mixed autoregressive probit model for ordinal longitudinal data, *Biostatistics*, **11**(1), 127-138.

[11] Wang, W. (2013). Identifiability of linear mixed effects models, *Electron. J. Stat*, **7**, 244-263.

[12] Williams, D.A. (1996). *Overdispersion in logistic-linear models*, In B. Morgan, Ed., Statistics in Toxicology. Oxford: Clarendon Press.

[13] Wu, H., Zhang, Y., and Long, J.D. (2017). Longitudinal beta-binomial modeling using GEE for overdispersed binomial data, *Statistics in Medicine*, **36**(6), 1029-1040.