

## مدل بندی داده های آمیخته طولی پیوسته و ترتیبی تحت مکانیسم گمشدگی انصراف

سجاد نوریان<sup>۱</sup>

گروه آمار، دانشگاه قم

تاریخ دریافت: ۱۳۹۶/۵/۲۵ تاریخ پذیرش: ۱۳۹۶/۱۱/۱۴

**چکیده:** در برخی از مطالعات طولی، به ویژه در زمینه های اجتماعی، اقتصادی، پزشکی و غیره ممکن است به طور هم زمان دو پاسخ مورد علاقه با دو مقیاس اندازه گیری متفاوت وجود داشته باشد که این دو پاسخ می توانند با یکدیگر همبسته باشند. از طرفی با توجه به ماهیت مطالعه های طولی، هر یک از پاسخ های مرتبط با یک آزمودنی در طول زمان نیز می توانند با یکدیگر همبسته هستند. پس باید در تحلیل این دسته از داده ها دو نوع ساختار همبستگی به طور هم زمان در نظر گرفته شود. در یک مطالعه ی طولی، ممکن است برای برخی آزمودنی ها به هر دلیلی (مانند تغییر مکان، فوت و ...) انصراف رخ دهد و اطلاعات آن ها موجود نباشد. در این حالت مدل بندی هم زمان داده های آمیخته ی طولی با پدیده ی انصراف، مطلوب تر از مدل بندی جداگانه ی هر یک از آن ها است. در این مقاله، مدل بندی ریاضی این گونه از داده ها تحت مکانیسم گمشدگی انصراف با استفاده از رویکرد بیزی ارائه می شود. از مطالعه شبیه سازی و تحلیل مثال کاربردی برای ارزیابی عملکرد مدل پیشنهادی استفاده شده است. این مدل، مدل های ارائه شده برای داده های کامل را به عنوان یک حالت خاص شامل می شود. همچنین در تحلیل مثال کاربردی نیز چند آزمون برای انتخاب بهترین مدل برازش داده شده به داده ها انجام شده است.

**واژه های کلیدی:** پاسخ های آمیخته، مدل اتورگرسیو، مدل لوژستیک تجمعی، مدل بندی توأم، مطالعات طولی، مکانیسم انصراف.

رده بندی ریاضی (۲۰۱۰): ۶۲H۱۵، ۶۲M۰۲.

## ۱- مقدمه

به‌طور معمول در مطالعات مقطعی و طولی اغلب یک پاسخ موردعلاقه وجود دارد و پژوهشگر به استنباط بر روی مشخصه یا مشخصه‌هایی از آن پاسخ برحسب یک یا چند متغیر تبیینی می‌پردازد. اما در برخی از مطالعات با این حالت مواجه هستیم که بیش از یک پاسخ موردعلاقه وجود دارد که مقیاس این پاسخ‌ها نیز می‌توانند متفاوت باشند. به‌عنوان مثال در یک مطالعه، ممکن است محقق علاقه‌مند به مطالعه مرحله‌ی بیماری و همچنین نمره افسردگی بیماران برحسب سایر متغیرهای تبیینی باشد. در این مثال محقق با دو پاسخ مواجه است که به ترتیب دارای مقیاس‌های ترتیبی و نسبتی هستند و همچنین به ترتیب در رده‌ی متغیرهای گسسته و پیوسته قرار می‌گیرند. طبیعی است که این دو پاسخ می‌توانند در هر زمان به یکدیگر وابسته باشند. از طرفی به دلیل جمع‌آوری داده‌ها در طول زمان، دنباله‌ی پاسخ‌های هر کدام از دو متغیر برای هر شخص به یکدیگر وابسته است. بنابراین در تحلیل این‌گونه داده‌ها باید دو نوع ساختار همبستگی، یکی برای هر پاسخ در طول زمان و دیگری برای دو پاسخ مرتبط با هر شخص در هر زمان را در نظر گرفت.

برای مدل بندی توأم پاسخ‌های گسسته و پیوسته چندین روش تاکنون پیشنهاد شده است. آلکین و تات [۱] استفاده از مدل مکانی عام را پیشنهاد کردند که در آن توزیع توأم دو متغیر به‌صورت حاصل ضرب یک توزیع چندجمله‌ای برای متغیر گسسته در یک توزیع شرطی چند متغیره نرمال برای متغیر پیوسته به شرط مقدار متغیر گسسته نوشته می‌شود. یانگ و همکاران [۲] مدلی مشابه را برای پاسخ‌های همبسته‌ی پواسون و پیوسته پیشنهاد کردند. روش دیگری که توسط کاکس و ورموت [۳] پیشنهاد شد، از عکس ترتیب بالا برای تجزیه توزیع توأم دو پاسخ آمیخته استفاده می‌کند، بدین معنا که توزیع توأم دو پاسخ آمیخته را به‌صورت حاصل ضرب یک توزیع چند متغیره برای متغیر پیوسته در یک توزیع شرطی برای متغیر گسسته به شرط متغیر پیوسته تجزیه می‌کند. بهرامی سامانی و گنجعلی [۴] و بهرامی سامانی و همکاران [۵] به ارائه یک مدل توأم بر اساس رهیافت متغیر پنهان پرداختند. نوریان و همکاران [۶] ضمن استفاده از دو روش بالا به معرفی پارامتر پیوند برای تعیین ساختار همبستگی دو متغیر آمیخته در هر زمان پرداختند، اما آن‌ها صرفاً مدل را در حالتی که مجموعه‌ی کامل داده‌ها در دسترس است در نظر گرفتند و هیچ اشاره‌ای به انصراف آزمودنی‌ها که یک مسئله معمول در تحلیل داده‌های طولی است نکردند.

لازم به ذکر است که تحت مکانیسم گمشدگی انصراف، مجموعه‌ی کامل داده‌ها در دسترس نخواهد بود و کاربر با مجموعه‌ای از داده‌ها مواجه است که بعد داده‌ها یک تابع نزولی از زمان اندازه‌گیری داده‌ها است. به‌عبارت‌دیگر در هر زمان، تعداد مشاهدات در حالتی که انصراف رخ

ندهد برابر تعداد مشاهدات در زمان قبل؛ و اگر انصراف رخ دهد کمتر از آن است. بنابراین لازم است که مدل بندی این دسته از داده‌ها تحت مکانیسم انصراف توسعه یابد.

در این مقاله با ارائه یک مدل ریاضی به بهبود مدل معرفی شده توسط آن‌ها تحت مکانیسم گمشدگی انصراف پرداخته می‌شود. در بخش ۲، مطالب مربوط به مدل‌سازی ریاضی شامل مدل بندی توزیع توأم پاسخ‌های همبسته بر اساس توزیع حاشیه‌ای برای متغیر پیوسته و توزیع شرطی برای متغیر ترتیبی به شرط متغیر پیوسته و همچنین مدل مربوط به توزیع پاسخ‌ها تحت حالت انصراف ارائه می‌شود. معرفی توزیع پیشین برای پارامترهای مدل و تابع درست‌نمایی پسین در بخش ۳ انجام می‌شود. در بخش ۴، به معرفی خلاصه مثال کاربردی، برازش مدل پیشنهادی به آن، اجرای دو آزمون درباره مجموعه‌ای از پارامترهای مدل (به‌طور هم‌زمان) و انتخاب بهترین مدل از بین آن‌ها با استفاده از معیارهای بیزی پرداخته می‌شود. سرانجام در بخش ۵، برخی نتایج ارائه می‌شود.

## ۲- مدل بندی ریاضی

### ۲-۱- مدل آمیخته انتقالی برای داده‌های کامل

فرض کنید پاسخ‌های ترتیبی و پیوسته برای شخص  $i$ -ام در زمان  $t$  را به ترتیب با  $Z_{it}$  و  $Y_{it}$  نمایش دهیم و برای هر شخص مجموعه‌ای از پاسخ‌ها در یک دوره از زمان جمع‌آوری شده باشد و نیز فرض کنید که پاسخ ترتیبی  $Y_{it}$  بتواند در مجموعه  $\{1, 2, \dots, J\}$  مقدار بگیرد. برای پاسخ‌های ترتیبی از یک مدل لوژستیک تجمعی استفاده می‌شود که در آن پاسخ قبلی را به‌عنوان یک متغیر تبیینی به‌منظور بررسی میزان تأثیر آن بر پاسخ فعلی در نظر می‌گیرد. میزان پیوند بین دو پاسخ همبسته‌ی پیوسته و گسسته نیز با پارامتر پیوند  $\gamma_t$  در مدل لوژستیک وارد شده است. برای مدل بندی پاسخ پیوسته نیز از یک مدل اتورگرسیو مرتبه‌ی اول استفاده شده است. در این صورت مدل بندی توأم این دو متغیر را می‌توان در زمان اول،  $t = 1$ ، به‌صورت

$$\begin{cases} \text{logit}[P(Y_{i1} \leq b | Z_{i1}, \alpha_{ib}, \boldsymbol{\beta}_1)] = \alpha_{ib} + \mathbf{X}'_{i1} \boldsymbol{\beta}_1 + \gamma_1 Z_{i1}, \\ Z_{i1} | [\mu_1, \boldsymbol{\theta}_1, \phi_1] = \mu_1 + \mathbf{X}'_{i1} \boldsymbol{\theta}_1 + \varepsilon_{i1}, \end{cases}$$

و در زمان‌های بعد،  $t > 1$ ، به‌صورت

$$\begin{cases} \text{logit}[P(Y_{it} \leq b | Y_{i,t-1} = a, Z_{it}, \alpha_{ib}, \boldsymbol{\beta}_t, \xi_t)] = \alpha_{ib} + \mathbf{X}'_{it} \boldsymbol{\beta}_t + \gamma_t Z_{it} + a \xi_t, \\ Z_{it} | [Z_{i,t-1}, \mu_t, \boldsymbol{\theta}_t, \phi_t] = \mu_t + \mathbf{X}'_{it} \boldsymbol{\theta}_t + \phi_t Z_{i,t-1} + \varepsilon_{it}, \end{cases} \quad (1)$$

نمایش داد. دلیل تفاوت دو مدل در دو زمان، این است که در زمان اول، هیچ‌گونه پاسخ قبلی وجود ندارد که بتوان باوجود آن از مدل‌های انتقالی استفاده کرد. در مدل (۱) برای پاسخ ترتیبی به شرط پاسخ پیوسته از یک مدل لوژستیک مرتبه اول استفاده شده است که در آن  $\alpha_{ib}$ ،  $b = 1, 2, \dots, J-1$  را پارامترهای نقطه‌ی برش یا آستانه‌ای می‌نامند. برای متغیر پنهان  $Y_{it}^*$   $Y_{it} = j$  خواهد بود اگر و فقط اگر  $\alpha_{ij} < Y_{it}^* < \alpha_{i,j-1}$ . بنابراین به منظور رعایت ساختار ترتیبی و همچنین شناساپذیر بودن، این پارامترها باید در شرط  $-\infty \leq \alpha_{i1} \leq \alpha_{i2} \leq \dots \leq \alpha_{i,J-1} \leq \infty$  صدق کنند (مک‌کاینلی و همکاران [۱۷]). همچنین ماتریس متغیرهایی کمکی،  $\beta_i$  بردار ضرایب رگرسیونی برای متغیرهای کمکی روی پاسخ ترتیبی،  $\gamma_i$  نشان دهنده میزان وابستگی پاسخ ترتیبی در زمان  $t$  به زمان  $t-1$  و  $\gamma_i$  نشان دهنده پیوند بین پاسخ گسسته و پاسخ پیوسته در زمان  $t$  است. همچنین برای پاسخ پیوسته از یک مدل اتورگرسیو مرتبه اول استفاده شده است که در آن پارامتر عرض از مبدأ در زمان  $t$ ،  $\theta_i$  بردار ضرایب رگرسیونی برای متغیرهای کمکی روی پاسخ پیوسته،  $\phi_i$  نشان دهنده میزان وابستگی پاسخ پیوسته  $Z_{it}$  به پاسخ زمان قبل آن یعنی  $Z_{i,t-1}$  است. سرانجام  $\varepsilon_{it}$  نیز خطای تصادفی است که فرض می‌شود دارای توزیع نرمال با میانگین صفر و واریانس ثابت  $\sigma_i^2$  است. تابع درستمایی برای این مجموعه از داده‌ها در حالت داده‌های کامل (یعنی در حالتی که مجموعه داده‌ها به صورت کامل در دسترس است) به صورت زیر به دست می‌آید:

$$\begin{aligned} L &= \prod_{i=1}^N f(Y_{i1}, Y_{i2}, \dots, Y_{iT}, Z_{i1}, Z_{i2}, \dots, Z_{iT}) \\ &= \prod_{i=1}^N f(Y_{i1}, Z_{i1}) f(Y_{i2}, Z_{i2} | Y_{i1}, Z_{i1}) \dots f(Y_{iT}, Z_{iT} | Y_{i1}, Y_{i2}, \dots, Y_{i,T-1}, Z_{i1}, Z_{i2}, \dots, Z_{i,T-1}) \\ &= \prod_{i=1}^N \left[ f(Y_{i1}, Z_{i1}) \prod_{t=2}^T f(Y_{it}, Z_{it} | Y_{i,t-1}, Z_{i,t-1}) \right] \\ &= \prod_{i=1}^N \left[ f(Z_{i1}) P(Y_{i1} | Z_{i1}) \prod_{t=2}^T f(Z_{it} | Z_{i,t-1}) P(Y_{it} | Y_{i,t-1}, Z_{i,t-1}) \right] \end{aligned}$$

دلیل تساوی آخر این است که در این مدل فرض می‌شود که داده‌ها از یک مدل انتقالی مرتبه‌ی اول پیروی می‌کنند. همچنین در این تساوی توابع جرم احتمال به صورت زیر به دست می‌آیند:

$$\begin{aligned} P(Y_{i1} = y_{i1} | Z_{i1}) &= P(Y_{i1} \leq y_{i1} | Z_{i1}) - P(Y_{i1} \leq y_{i1} - 1 | Z_{i1}), \\ P(Y_{it} = y_{it} | Y_{i,t-1} = a, Z_{it}) &= P(Y_{it} \leq y_{it} | Y_{i,t-1} = a, Z_{it}) \\ &\quad - P(Y_{it} \leq y_{it} - 1 | Y_{i,t-1} = a, Z_{it}). \end{aligned}$$

### ۲-۲- مدل آمیخته برای داده‌های با مکانیسم گمشدگی انصراف

هنگامی که برای یک آزمودنی خاص انصراف رخ می‌دهد، تنها پاسخ‌هایی که تا قبل از آن زمان جمع‌آوری شده‌اند در دسترس خواهند بود. به عبارت دیگر تمامی پاسخ‌های آزمودنی در آن زمان و زمان‌های بعدی غیرقابل مشاهده می‌شوند و در نتیجه برای این آزمودنی یک مجموعه داده‌ی ناکامل یا ناقص وجود دارد. انصراف برای هر کدام از آزمودنی‌ها و در هر زمان از زمان‌های جمع-آوری داده می‌تواند رخ دهد. در این حالت تابع درستنمایی را می‌توان با در نظر گرفتن داده‌های مشاهده شده در هر زمان به صورت زیر نوشت:

$$L = \left[ \prod_i^{N_1} f(Y_{i1}, Z_{i1}) \right] \times \left[ \prod_i^{N_2} f(Y_{i2}, Z_{i2} | Y_{i1}, Z_{i1}) \right] \times \dots \times \left[ \prod_i^{N_T} f(Y_{iT}, Z_{iT} | Y_{i,T-1}, Z_{i,T-1}) \right]$$

$$= \left[ \prod_i^{N_1} f(Z_{i1}) P(Y_{i1} | Z_{i1}) \right] \times \left[ \prod_i^{N_2} f(Z_{i2} | Z_{i1}) P(Y_{i2} | Y_{i1}, Z_{i2}) \right] \times \dots$$

$$\times \left[ \prod_i^{N_T} f(Z_{iT} | Z_{i,T-1}) P(Y_{iT} | Y_{i,T-1}, Z_{iT}) \right]$$

که در آن  $N_1$  تعداد همه افرادی است که پاسخ‌های آن‌ها در زمان اول جمع‌آوری شده است (تعداد کل آزمودنی‌ها) و  $N_t, t = 2, 3, \dots, T$ ، تعداد افرادی است که برای آن‌ها هر دو پاسخ ترتیبی و پیوسته در زمان‌های دوم تا  $T$ -ام مشاهده شده است. لازم به ذکر است که در هر زمان  $t$ ، مقدار  $N_{t-1} - N_t$  برابر با تعداد آزمودنی‌هایی است که برای آن‌ها در این زمان انصراف رخ داده است.

### ۳- تحلیل بیزی

در این بخش به انجام استنباط بر روی پارامترهای مدل (۱) از طریق رهیافت بیزی پرداخته می‌شود. چون پارامترهای نقطه‌ی برش دارای ساختار غیر نزولی نسبت به یکدیگر هستند، استفاده از توزیع چند متغیره نرمال به عنوان توزیع پیشینی برای آن‌ها پیشنهاد نمی‌شود ([۹-۱۸])؛ بنابراین به جای آن از توزیع‌های نرمال مستقل دم‌بریده استفاده می‌شود. برای لحاظ کردن ساختار مارکوفی مرتبه‌ی اول بر روی پارامترهای ضرایب رگرسیونی از توزیع‌های پیشین زیر در زمان اول و زمان‌های بعدی استفاده می‌شود؛

$$\beta_1 \sim N(0, \Sigma_\beta); \beta_t | \beta_{t-1} \sim N(\beta_{t-1}, \Sigma_\beta); \theta_1 \sim N(0, \Sigma_\theta); \theta_t | \theta_{t-1} \sim N(\theta_{t-1}, \Sigma_\theta);$$

که در آن  $\Sigma_\beta$  و  $\Sigma_\theta$  ماتریس‌های کوواریانس قطری  $K$  بعدی هستند که به صورت  $\Sigma_\beta = \sigma_\beta^2 \mathbf{I}$  و  $\Sigma_\theta = \sigma_\theta^2 \mathbf{I}$  تعریف می‌شوند که در آن  $\sigma_\beta^2 = (\sigma_{\beta_1}^2, \sigma_{\beta_2}^2, \dots, \sigma_{\beta_K}^2)'$  و

[۱۰] در هر زمان از پیشین‌های زیر استفاده می‌شود:

$$\alpha_{t1} \sim TN(\mu, \sigma^2); -\infty < \alpha_{t1} < \alpha_{t2},$$

$$\alpha_{tb} \sim TN(\mu, \sigma^2); \alpha_{t,b-1} < \alpha_{tb} < \alpha_{t,b+1}; b = 2, \dots, J-2,$$

$$\alpha_{t,J-1} \sim TN(\mu, \sigma^2); \alpha_{t,J-2} < \alpha_{t,J-1} < +\infty,$$

که در آن  $TN(\mu, \sigma^2)$  توزیع نرمال دم‌بریده پارامترهای  $\mu$  و  $\sigma^2$  است. برای سایر پارامترهای مدل یعنی  $\xi_t$ ،  $\gamma_t$ ،  $\mu_t$  و  $\phi_t$  نیز از توزیع  $N(\cdot, \sigma^2)$  و برای واریانس خطاها،  $\sigma_t^2$ ، نیز از توزیع کای-دو با یک درجه آزادی به‌عنوان توزیع پیشین استفاده شده است. با داشتن داده‌های  $N$  آزمودنی و امکان رخ داده انصراف برای برخی از آزمودنی‌ها، توزیع پسینی توأم متناسب است با

$$\begin{aligned} L = & \left[ \prod_i^{N_1} f(Z_{i1}) P(Y_{i1} | Z_{i1}) \right] \times \left[ \prod_i^{N_2} f(Z_{i2} | Z_{i1}) P(Y_{i2} | Y_{i1}, Z_{i2}) \right] \times \dots \\ & \times \left[ \prod_i^{N_T} f(Z_{iT} | Z_{i,T-1}) P(Y_{iT} | Y_{i,T-1}, Z_{iT}) \right] \times \pi(\alpha) \\ & \times \left( \prod_{t=2}^T \pi(\beta_t | \beta_{t-1}) \times \pi(\theta_t | \theta_{t-1}) \times \pi(\xi_t) \times \pi(\gamma_t) \times \pi(\mu_t) \times \pi(\phi_t) \times \pi(\sigma_t^2) \right). \end{aligned} \quad (2)$$

#### ۴- مثال کاربردی

##### ۴-۱- مطالعه شبیه‌سازی

در این بخش، یک مطالعه شبیه‌سازی به‌منظور بررسی کارایی مدل ارائه شده در برآورد پارامترها اجرا شده است. در شبیه‌سازی شده از مدل (۱) به همراه دو متغیر کمکی و ناوردای زمانی  $X_{vit}$  و  $X_{vit}$ ،  $t=1, 2, 3$ ، استفاده شده است که در آن  $X_{vit}$  یک متغیر دودویی (مانند نشانگر استفاده از دارو) و  $X_{vit}$  یک متغیر پیوسته (مانند وزن) هستند که به ترتیب از توزیع برنولی با احتمال موفقیت پارامتر  $0/5$  و نرمال با میانگین  $80$  و واریانس  $5$  تولید شده‌اند. برای پارامترهای مدل (۱) مقادیر زیر در نظر گرفته شده است:  $\beta_1 = (-0/40, -0/35, -0/30)$ ،  $\beta_2 = (-0/20, -0/05, -0/27)$ ،  $\alpha_1 = (3, 4)$ ،  $\alpha_2 = (1, 4)$ ،  $\alpha_3 = (9, 12)$ ،  $\theta_1 = (-0/06, -0/03, -0/05)$ ،  $\theta_2 = (0/60, -0/10, 0/06)$ ،  $\theta_3 = (-3, -2, -1)$ ،  $\xi = (-5, -2)$  و  $\phi = (0/8, 0/9)$  و  $\gamma = (0/3, 0/4, 0/5)$ . همچنین فرض شده است که در هر

زمان خطاهای تصادفی،  $\varepsilon_{it}$ ، از توزیع نرمال استاندارد انتخاب شده‌اند. در زمان اول ۲۰۰ داده (مربوط به ۲۰۰ آزمودنی) تولید شد، سپس در زمان‌های دوم و سوم، با استفاده از توزیع دو جمله‌ای با پارامترهایی ۲۰۰ و ۰/۲۰، مدل انصراف را بر آزمودنی‌ها اعمال کردیم. به عبارت دیگر در هر زمان، برای ۲۰ درصد از آزمودنی‌ها به تصادف انصراف رخ داده است. در شبیه‌سازی انجام شده در زمان‌های اول، دوم و سوم به ترتیب ۲۰۰، ۱۶۱ و ۱۲۶ آزمودنی در مطالعه حضور داشتند. پس از انتخاب توزیع  $TN(0, 10000)$  برای پارامترهای نقطه‌ی برش، توزیع  $TN(0, 10000)$  برای پارامتر پیوند  $(\gamma_t)$  و پارامترهای  $\mu_t$ ،  $\xi_t$  و  $\phi_t$ ، توزیع‌های  $N(\beta_{t-1}, 10000)$  و  $N(\theta_{t-1}, 10000)$  (با در نظر گرفتن  $\beta_0 = \theta_0 = 0$ ) به ترتیب برای ضرایب رگرسیونی  $\beta_t$  و  $\theta_t$  و سرانجام توزیع  $\mathcal{X}_{(1)}^2$  برای واریانس خطاهای تصادفی  $(\sigma_t^2)$  به‌عنوان توزیع پیشینی، برآورد پارامترهای مدل با استفاده از بسته نرم‌افزاری OpenBUGS با ۲۰۰,۰۰۰ تکرار پس از انتخاب زمان داغیدن برابر ۳۰۰,۰۰۰ (جهت اطمینان از حصول همگرایی نتایج) و انتخاب یک نمونه از هر ۱۰۰ نمونه در جدول (۱) آورده شده است.

همان‌طور که ملاحظه می‌شود مقادیر برآورد شده، تقریباً به مقادیر واقعی نزدیک هستند و خطای نسبی برآوردگرها نیز حدود ده درصد است. با توجه به این جدول و همچنین انحراف معیار برآوردگرها (که در اینجا گزارش نشده است) می‌توان نتیجه گرفت که زنجیر مارکوفی یک زنجیر همگن است و همچنین پارامترهای مدل شناسا پذیرند.

**جدول (۱):** مقدار واقعی (Tr)، مقدار برآورد شده پارامترها (Est) و خطای نسبی برآوردگر (Re) در مطالعه شبیه‌سازی

$\sigma_t^2$	$\phi_t$	$\theta_{1t}$	$\theta_{2t}$	$\mu_t$	$\gamma_t$	$\xi_t$	$\beta_{1t}$	$\beta_{2t}$	$\alpha_{1t}$	$\alpha_{2t}$	Par	t
۱/۰۰	---	۰/۶۰	-۰/۰۶	-۳/۰۰	۰/۳۰	---	-۰/۲۰	-۰/۴۰	۴/۰۰	۳/۰۰	Tr	۱
۰/۹۳	---	۰/۶۳	-۰/۰۷	-۳/۳۷	۰/۳۳	---	-۰/۲۱	-۰/۳۷	۴/۲۴	۳/۲۸	Est	
۰/۰۷	---	۰/۰۵	۰/۱۰	۰/۱۲	۰/۱۰	---	۰/۰۵	۰/۰۸	۰/۰۶	۰/۰۹	Re	
۱/۰۰	۰/۸۰	-۰/۱۰	-۰/۰۳	-۲/۰۰	۰/۴۰	-۵/۰۰	-۰/۰۵	-۰/۳۵	۴/۰۰	۱/۰۰	Tr	۲
۰/۹۳	۰/۸۳	-۰/۱۱	-۰/۰۳	-۱/۸۱	۰/۴۳	-۵/۴۴	-۰/۰۴	-۰/۳۱	۳/۶۴	۰/۸۹	Est	
۰/۰۷	۰/۰۴	۰/۱۰	۰/۰۰	۰/۱۰	۰/۰۸	۰/۰۹	۰/۲۰	۰/۱۱	۰/۰۹	۰/۱۱	Re	
۱/۰۰	۰/۹۰	۰/۰۶	-۰/۰۵	-۱/۰۰	۰/۵۰	-۲/۰۰	-۰/۲۷	-۰/۳۰	۱۲/۰۰	۹/۰۰	Tr	۳
۰/۹۶	۰/۹۸	۰/۰۶	-۰/۰۴	-۱/۱۴	۰/۵۶	-۱/۸۲	-۰/۳۰	-۰/۲۴	۱۰/۷۲	۸/۰۷	Est	
۰/۰۴	۰/۰۹	۰/۰۰	۰/۱۰	۰/۱۴	۰/۱۲	۰/۰۹	۰/۱۱	۰/۱۳	۰/۱۱	۰/۱۰	Re	

## ۲-۴- داده‌های آزمون انفرادی پیشرفت پی‌بادی

داده‌های آزمون انفرادی پیشرفت پی‌بادی (از این به بعد، پی‌بادی) داده‌هایی هستند که با استفاده از انجام مصاحبه‌هایی از هر مادر و کودکش به صورت یک سال در میان در فاصله‌ی سال‌های ۱۹۸۶ تا ۱۹۹۲ جمع‌آوری شده است. در این مطالعه، دو پاسخ موردعلاقه عبارت بودند از (۱) رفتار غیراجتماعی کودکان ( $Y_{ii}$ ) که به صورت یک متغیر ترتیبی با سطوح ۱ تا ۳ برحسب اظهارات مادر و پاسخی که او به برخی سؤالات می‌دهد، اندازه‌گیری شده است و (۲) توانایی کودکان در خواندن لغات ( $Z_{ii}$ ) که طبق آزمونی که از کودک گرفته می‌شود به صورت نمره و یک متغیر کمی جمع‌آوری شده است.

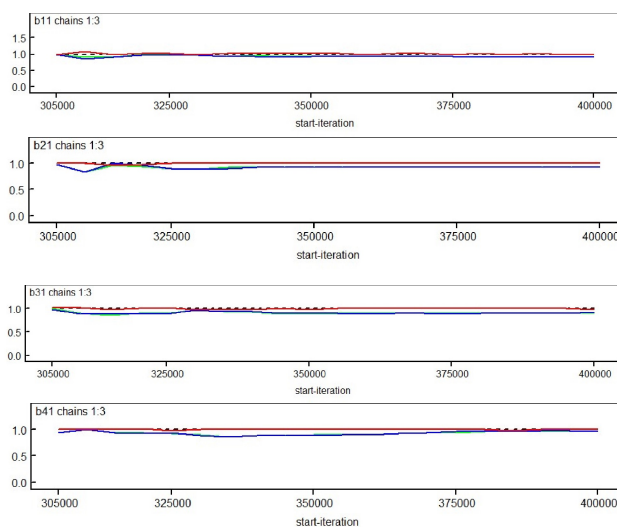
در زمان (ویزیت) اول جمع‌آوری داده‌ها ۴۰۵ زوج (مادر و کودک) حضور داشتند که تعداد زوج‌هایی که هر دو پاسخ آن‌ها در زمان‌های دوم، سوم و چهارم مشاهده شده است به ترتیب برابر ۳۶۲، ۲۵۴ و ۲۲۱ نفر می‌باشد. به عبارت دیگر تنها ۲۲۱ زوج بودند که از ابتدا تا انتها در مطالعه حضور داشتند و همه پاسخ‌های توأم آن‌ها (به همراه سایر اطلاعات کمکی) جمع‌آوری شد. در این مطالعه چند متغیر کمکی مانند جنسیت کودکان ( $X_1$ )، سن مادران ( $X_2$ ) و سن فرزندان نشان ( $X_3$ ) نیز جمع‌آوری شده است که در صورت معنادار بودن ضرایب رگرسیونی مرتبط، از آن‌ها به عنوان متغیرهای تبیینی در مدل استفاده می‌شود. نوریان و همکاران [۶] نشان دادند که در هر زمان هر کدام از دو پاسخ به طور معناداری به پاسخ زمان قبل وابسته است. بنابراین آن‌ها از مدل انتقالی برای تحلیل این پاسخ‌ها استفاده کردند. اما هیچ اشاره‌ای به نحوه‌ی استفاده از این مدل در حالتی که در داده‌ها گمشدگی رخ دهد، نکردند. به عبارت دیگر، مدل ارائه شده توسط آن‌ها تنها در حالت داده‌های کامل کاربرد دارد، اما با استفاده از مدل ارائه شده در این مقاله می‌توان به خوبی به تحلیل داده‌ها هم در حالت داده‌های کامل و هم در حالت وجود انصراف پرداخت. بنابراین، مدل ارائه شده توسط نوریان و همکاران [۶] حالت خاصی از مدل ارائه شده در این مقاله است که در هیچ زمانی از زمان‌های اندازه‌گیری داده‌ها انصرافی رخ نداده است و در نتیجه  $N_1 = N_2 = \dots = N_T$ . در واقع آن‌ها [۶] تنها اطلاعات مربوط به ۲۲۱ زوج که داده‌های آن‌ها کامل بود را در نظر گرفته بودند، اما در این مقاله با استفاده از مدل (۱) و تابع درست‌نمایی (۲) که یک مدل انتقالی توأم مرتبه اول برای در نظر گرفتن انصراف در داده‌ها است، از همه‌ی اطلاعات موجود تا زمان انصراف استفاده و به بهبود مدل آن‌ها پرداخته می‌شود. همچنین در این مقاله با استفاده از شبیه‌سازی درستی عملکرد مدل در برآورد پارامترها و نیز شناساپذیر بودن آن‌ها به صورت تجربی بررسی شده است، در حالی که در مقاله‌ی آن‌ها این کار مغفول مانده بود. این مدل برای تحلیل داده‌های پی‌بادی در زمان اول و در زمان‌های بعدی به صورت زیر نوشته می‌شود:



$$\begin{cases} \text{logit}[P(Y_{i1} \leq b | Z_{i1}, \alpha_{ib}, \beta_{11})] = \alpha_{ib} + \beta_{11}X_{i1} + \beta_{12}X_{i2} + \beta_{13}X_{i3} + \gamma_1 Z_{i1}, \\ Z_{i1} | [\mu_1, \theta_1, \phi_1] = \mu_1 + \theta_{11}X_{i1} + \theta_{12}X_{i2} + \theta_{13}X_{i3} + \varepsilon_{i1}, \\ \text{logit}[P(Y_{it} \leq b | Y_{i,t-1} = a, Z_{it}, \alpha_{ib}, \beta_t, \xi_t)] = \alpha_{ib} + \beta_{t1}X_{it} + \beta_{t2}X_{it} \\ \quad \beta_{t3}X_{it} + a\xi_t + \gamma_t Z_{it}, \\ Z_{it} | [Z_{i,t-1}, \mu_t, \theta_t, \phi_t] = \mu_t + \theta_{t1}X_{it} + \theta_{t2}X_{it} + \theta_{t3}X_{it} + \phi_t Z_{i,t-1} + \varepsilon_{it}. \end{cases}$$

#### ۲-۴- تحلیل بیزی داده‌های پی‌بادی

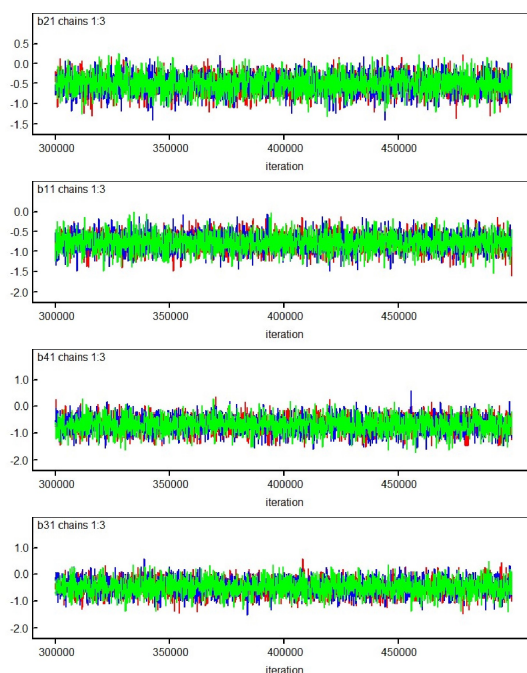
برای پارامترهای نقطه‌ی برش از توزیع  $TN(0, 10000)$ ، برای پارامتر پیوند  $(\gamma_t)$  و پارامترهای  $\mu_t$ ،  $\xi_t$  و  $\phi_t$  از توزیع  $N(0, 10000)$  استفاده می‌شود، برای ضرایب رگرسیونی  $\beta_t$  و  $\theta_t$ ، به ترتیب از توزیع  $N(\beta_{t-1}, 10000)$  و  $N(\theta_{t-1}, 10000)$  که در آن در  $\beta_0 = \theta_0 = 0$  و سرانجام برای واریانس خطاهای تصادفی  $(\sigma_t^2)$  از توزیع  $\chi^2_{(1)}$  به‌عنوان توزیع پیشینی استفاده می‌شود.



شکل (۱): نمودار گل‌من روبین پارامترهای  $\beta_{11}$ ،  $\beta_{21}$ ،  $\beta_{31}$  و  $\beta_{41}$  در مدل (۱)

دلیل انتخاب واریانس بزرگ در توزیع‌های نرمال، انتخاب پیشین‌هایی با آگاهی بخشی پایین است. برآورد پارامترهای مدل با استفاده از بسته نرم‌افزاری OpenBUGS با ۳۰۰,۰۰۰ تکرار پس از انتخاب زمان داغیدن برابر ۲۰۰,۰۰۰ (جهت اطمینان از حصول همگرایی نتایج) و انتخاب یک نمونه از هر ۱۰۰ نمونه به دست آمده و در جدول (۲) آورده شده است. معیار گل‌من-روبین

[۱۱] برای پارامترهای مدل پس از زمان داغیدن بین ۰/۹۹ و ۱/۰۷ می‌باشد که نشان‌دهنده‌ی همگرایی زنجیر مارکوفی است. در شکل (۱) به‌عنوان نمونه، نمودار همگرایی گلن روبین و در شکل (۲) نمودار اثر ضرایب متغیر جنسیت در مدل لوژستیک در سه بار تکرار زنجیر آورده شده است که هیچ نشانه‌ای از عدم همگرایی زنجیر مارکوفی را نمایش نمی‌دهند. باوجود اینکه واریانس توزیع پیشینی پارامترها بسیار بزرگ (۱۰,۰۰۰) در نظر گرفته شده است، اما ملاحظه می‌شود که انحراف معیار برآوردگرهای پسینی بسیار پایین است. کم بودن انحراف معیار برآوردگرهای پسینی یکی دیگر از نشانه‌های شناساپذیر بودن پارامترها است.



شکل (۲): نمودار سری زمانی پارامترهای  $\beta_{11}$ ,  $\beta_{21}$ ,  $\beta_{31}$  و  $\beta_{41}$  در مدل (۱)

آزمونی که ابتدا در نظر گرفته می‌شود، آزمون همگنی است. یعنی به ازای هر پارامتر  $\delta_{ti}$  داشته باشیم:

$$H_0: \delta_{\tau_i} = \delta_{\tau_i} = \delta_{\tau_i}. \quad (3)$$

در واقع فرضیه صفر را به‌صورت واضح می‌توان به‌صورت زیر نوشت:

$$H_0: \begin{cases} \alpha_{t1} = \alpha_{t1} = \alpha_{f1}; & \alpha_{t2} = \alpha_{t2} = \alpha_{f2}; & \beta_{t1} = \beta_{t1} = \beta_{f1}; & \beta_{t2} = \beta_{t2} = \beta_{f2} \\ \beta_{t3} = \beta_{t3} = \beta_{f3}; & \theta_{t1} = \theta_{t1} = \theta_{f1}; & \theta_{t2} = \theta_{t2} = \theta_{f2}; & \theta_{t3} = \theta_{t3} = \theta_{f3} \\ \xi_t = \xi_t = \xi_f; & \gamma_t = \gamma_t = \gamma_f; & \mu_t = \mu_t = \mu_f; & \phi_t = \phi_t = \phi_f \\ \sigma_t^2 = \sigma_t^2 = \sigma_f^2 \end{cases}$$

جدول (۲): برآورد پارامترهای مدل (۱) (Est) و انحراف معیار آن‌ها (Sd)

پارامترهای مدل اتورگرسیو						پارامترهای مدل لوژستیک							
$\sigma_t^2$	$\phi_t$	$\theta_{t3}$	$\theta_{t2}$	$\theta_{t1}$	$\mu_t$	$\gamma_t$	$\xi_t$	$\beta_{t3}$	$\beta_{t2}$	$\beta_{t1}$	$\alpha_{t2}$	$\alpha_{t1}$	$t$
۱/۸۳ (۰/۰۴)	...	۰/۸۷ (۰/۰۶)	۰/۰۳ (۰/۰۲)	-۰/۰۶ (۰/۰۸)	-۴/۱۴ (۰/۶۴)	۰/۳۶ (۰/۱۴)	...	-۰/۶۹ (۰/۲۲)	۰/۰۶ (۰/۰۵)	-۰/۷۸ (۰/۲۲)	۶/۳۳ (۱/۷۷)	۱/۷۵ (۱/۷۳)	۱
۱/۵۵ (۰/۰۵)	۰/۸۵ (۰/۰۶)	-۰/۲۴ (۰/۰۸)	۰/۰۷ (۰/۰۲)	-۰/۰۷ (۰/۰۹)	۱/۹۰ (۰/۷۵)	۰/۲۲ (۰/۱۱)	-۱/۳۵ (۰/۲۴)	۰/۰۴ (۰/۱۸)	۰/۰۳ (۰/۰۶)	-۰/۵۱ (۰/۲۴)	۳/۴۳ (۱/۷۷)	-۰/۵۱ (۱/۷۷)	۲
۱/۸۸ (۰/۰۵)	۰/۸۴ (۰/۰۵)	-۰/۰۷ (۰/۰۸)	۰/۰۳ (۰/۰۲)	۰/۱۰ (۰/۰۹)	۱/۳۱ (۰/۷۷)	۰/۴۳ (۰/۱۳)	-۱/۴۴ (۰/۲۷)	-۰/۴۰ (۰/۲۳)	-۰/۰۲ (۰/۰۸)	-۰/۴۷ (۰/۲۹)	۷/۴۷ (۲/۴۶)	۲/۸۶ (۲/۴۲)	۳
۲/۰۲ (۰/۰۵)	۰/۸۹ (۰/۰۵)	-۰/۱۴ (۰/۰۹)	۰/۰۴ (۰/۰۳)	-۰/۰۳ (۰/۱۰)	۱/۲۲ (۰/۸۵)	۰/۲۰ (۰/۱۳)	-۱/۹۱ (۰/۳۲)	۰/۱۷ (۰/۲۳)	۰/۰۱ (۰/۰۹)	-۰/۷۰ (۰/۳۰)	۴/۳۶ (۲/۵۴)	۰/۱۲ (۲/۵۲)	۴

اگر فرضیه صفر و در نتیجه همگنی برقرار باشد، آنگاه نتیجه می‌شود که ضرایب رگرسیونی، پارامتر پیوند بین دو پاسخ آمیخته و همچنین واریانس خطاهای تصادفی ناوردای زمانی هستند. بنابراین در مدل بندی توأم پاسخ‌های آمیخته، احتمال انتقال پاسخ ترتیبی از سطح  $a$  به سطح  $b$  در هر زمان دلخواه  $t, t > 1$ ، و همچنین مقدار پیش‌بینی برای پاسخ پیوسته، برای همه‌ی افرادی که دارای مقادیر یکسانی در متغیرهای کمکی هستند، برابر است. چون مدل ارائه شده یک مدل انتقالی مرتبه اول (به شرط پاسخ قبلی) است، بنابراین فرضیه‌های همگنی روی پارامترهای زمان اول اعمال نشده‌اند. برآورد پارامترها تحت فرضیه‌ی همگنی داده شده در رابطه‌ی (۳) در جدول (۳) آمده است.

با توجه به برآورد پارامترها در هر دو مدل، ملاحظه می‌شود که ضرایب رگرسیونی متغیر کمکی سن مادر ( $X_7$ ) در مدل لوژستیک مربوط به پاسخ ترتیبی،  $\beta_{7t}$ ، و متغیر جنسیت ( $X_1$ ) در مدل اتورگرسیو مربوط به پاسخ پیوسته،  $\theta_{1t}$ ، در هیچ‌کدام از زمان‌ها معنی‌دار نیست. بنابراین فرضیه دیگری که می‌تواند در مدل همگن اعمال شود، صفر بودن این ضرایب است. این فرضیه را می‌توان (به همراه سایر فرضیه‌های مربوط به همگنی) به صورت زیر نوشت:

$$H_0' : \begin{cases} \alpha_{t1} = \alpha_{t1} = \alpha_{t1} ; \alpha_{t2} = \alpha_{t2} = \alpha_{t2} & ; \beta_{t1} = \beta_{t1} = \beta_{t1} ; \beta_{t2} = \beta_{t2} = \beta_{t2} = \beta_{t2} = 0 \\ \beta_{t3} = \beta_{t3} = \beta_{t3} ; \theta_{t1} = \theta_{t1} = \theta_{t1} = \theta_{t1} = 0 ; \theta_{t2} = \theta_{t2} = \theta_{t2} ; \theta_{t3} = \theta_{t3} = \theta_{t3} \\ \xi_t = \xi_t = \xi_t ; \gamma_t = \gamma_t = \gamma_t & ; \mu_t = \mu_t = \mu_t ; \phi_t = \phi_t = \phi_t \\ \sigma_t^2 = \sigma_t^2 = \sigma_t^2 \end{cases}$$

برآورد پارامترها تحت این فرضیه؛ که مدل مرتبط با آن را مدل (۳) می‌نامیم؛ در جدول (۴) آمده است.

جدول (۳): برآورد پارامترهای مدل تحت فرضیه‌ی  $H_0$  (Est) و انحراف معیار آن‌ها (Sd)

پارامترهای مدل اتورگرسیو										پارامترهای مدل لوژستیک					
$\sigma_t^2$	$\phi_t$	$\theta_{t2}$	$\theta_{t3}$	$\theta_{t1}$	$\mu_t$	$\gamma_t$	$\xi_t$	$\beta_{t2}$	$\beta_{t3}$	$\beta_{t1}$	$\alpha_{t2}$	$\alpha_{t1}$	t		
۱/۸۲ (۰/۰۴)	...	۰/۸۷ (۰/۰۶)	۰/۰۳ (۰/۰۲)	-۰/۰۶ (۰/۰۷)	-۴/۱۶ (۰/۱۴)	۰/۳۷	...	-۰/۶۸ (۰/۲۰)	۰/۰۶ (۰/۰۶)	-۰/۷۷ (۰/۲۲)	۶/۲۹ (۲/۰۳)	۱/۷۰ (۱/۹۹)	Est ۱ (Sd)		
۱/۷۱ (۰/۰۳)	۰/۷۶ (۰/۰۲)	-۰/۱۱ (۰/۰۴)	۰/۰۶ (۰/۰۱)	-۰/۰۳ (۰/۰۵)	۱/۳۵ (۰/۴۵)	۰/۲۵ (۰/۰۶)	-۱/۴۹ (۰/۱۵)	-۰/۰۵ (۰/۱۲)	۰/۰۱ (۰/۰۴)	-۰/۵۲ (۰/۱۵)	۴/۷۵ (۱/۳۲)	۰/۶۲ (۱/۳۰)	Est ۲ (Sd)		
۱/۷۱ (۰/۰۳)	۰/۷۶ (۰/۰۲)	-۰/۱۱ (۰/۰۴)	۰/۰۶ (۰/۰۱)	-۰/۰۳ (۰/۰۵)	۱/۳۵ (۰/۴۵)	۰/۲۵ (۰/۰۶)	-۱/۴۹ (۰/۱۵)	-۰/۰۵ (۰/۱۲)	۰/۰۱ (۰/۰۴)	-۰/۵۲ (۰/۱۵)	۴/۷۵ (۱/۳۲)	۰/۶۲ (۱/۳۰)	Est ۳ (Sd)		
۱/۷۱ (۰/۰۳)	۰/۷۶ (۰/۰۲)	-۰/۱۱ (۰/۰۴)	۰/۰۶ (۰/۰۱)	-۰/۰۳ (۰/۰۵)	۱/۳۵ (۰/۴۵)	۰/۲۵ (۰/۰۶)	-۱/۴۹ (۰/۱۵)	-۰/۰۵ (۰/۱۲)	۰/۰۱ (۰/۰۴)	-۰/۵۲ (۰/۱۵)	۴/۷۵ (۱/۳۲)	۰/۶۲ (۱/۳۰)	Est ۴ (Sd)		

جدول (۴): برآورد پارامترهای مدل تحت فرضیه‌ی  $H_0'$  (Est) و انحراف معیار آن‌ها (Sd)

پارامترهای مدل اتورگرسیو										پارامترهای مدل لوژستیک					
$\sigma_t^2$	$\phi_t$	$\theta_{t2}$	$\theta_{t3}$	$\theta_{t1}$	$\mu_t$	$\gamma_t$	$\xi_t$	$\beta_{t2}$	$\beta_{t3}$	$\beta_{t1}$	$\alpha_{t2}$	$\alpha_{t1}$	t		
۱/۸۲ (۰/۰۴)	...	۰/۸۶ (۰/۰۶)	۰/۰۳ (۰/۰۲)	...	-۴/۲۰ (۰/۶۰)	۰/۳۸ (۰/۱۵)	...	-۰/۶۸ (۰/۲۳)	...	-۰/۷۶ (۰/۲۲)	۷/۷۶ (۱/۴۷)	۳/۲۰ (۱/۳۹)	Est ۱ (Sd)		
۱/۷۲ (۰/۰۳)	۰/۷۶ (۰/۰۲)	-۰/۱۱ (۰/۰۴)	۰/۰۶ (۰/۰۱)	...	۱/۳۴ (۰/۴۵)	۰/۲۶ (۰/۰۶)	-۱/۴۷ (۰/۱۵)	-۰/۰۵ (۰/۱۲)	...	-۰/۵۲ (۰/۱۵)	۴/۴۹ (۰/۸۸)	۰/۸۱ (۰/۸۴)	Est ۲ (Sd)		
۱/۷۲ (۰/۰۳)	۰/۷۶ (۰/۰۲)	-۰/۱۱ (۰/۰۴)	۰/۰۶ (۰/۰۱)	...	۱/۳۴ (۰/۴۵)	۰/۲۶ (۰/۰۶)	-۱/۴۷ (۰/۱۵)	-۰/۰۵ (۰/۱۲)	...	-۰/۵۲ (۰/۱۵)	۴/۴۹ (۰/۸۸)	۰/۸۱ (۰/۸۴)	Est ۳ (Sd)		
۱/۷۲ (۰/۰۳)	۰/۷۶ (۰/۰۲)	-۰/۱۱ (۰/۰۴)	۰/۰۶ (۰/۰۱)	...	۱/۳۴ (۰/۴۵)	۰/۲۶ (۰/۰۶)	-۱/۴۷ (۰/۱۵)	-۰/۰۵ (۰/۱۲)	...	-۰/۵۲ (۰/۱۵)	۴/۴۹ (۰/۸۸)	۰/۸۱ (۰/۸۴)	Est ۴ (Sd)		

اگر مدل را در حالت کلی، تحت فرضیه‌ی (۳) و تحت فرضیه‌ی (۴) را به ترتیب مدل‌های ۱، ۲ و ۳ بنامیم، آنگاه برای مقایسه‌ی این مدل‌ها و انتخاب بهترین مدل از معیارهای اطلاع انحراف  $(DIC)$  [۱۲]، اطلاع آکائیک  $(AIC)$  [۱۳]، اطلاع بیزی  $(BIC)$  [۱۴] و عامل بیزی [۱۵] استفاده می‌شود. در جدول (۵) مقدار سه معیار اول برای مقایسه این سه مدل آمده است. مقدار  $DIC$  به صورت  $DIC = Dbar + pD = Dhat + 2pD$  تعریف می‌شود. اگر انحرافه‌ی  $D(\theta)$  برابر منفی دو برابر لگاریتم درست‌نمایی باشد،  $D(\theta) = -2 \times \log(\text{likelihood})$ ، آنگاه  $Dbar$  از میانگین‌گیری از  $D(\theta)$  روی نمونه‌های انتخاب شده  $\theta$  و  $Dhat$  با محاسبه مقدار  $D(\theta)$  در میانگین  $\theta$ ‌های انتخاب شده به دست می‌آیند. به عبارت دیگر  $Dbar$  میانگین پسینی انحرافه و  $Dhat$  مقدار انحرافه در میانگین پسینی پارامتر است.  $pD$  نیز تعداد مؤثر پارامترهای مدل است و به عنوان اندازه‌ای از پیچیدگی مدل تعریف می‌شود که از تفاضل میانگین پسینی انحرافه و انحرافه میانگین پسینی به دست می‌آید،  $pD = Dbar - Dhat$ . اگر  $k$  تعداد پارامترهای برآورد شده در مدل و  $\hat{L}$  مقدار ماکسیمم تابع درست‌نمایی مدل باشد آنگاه مقدار  $AIC$  به صورت  $AIC = 2k - 2 \times \log(\text{likelihood})$  تعریف می‌شود. سرانجام برای مقایسه دو مدل  $M_1$  و  $M_2$  اگر  $\pi(M_1)$  و  $\pi(M_2)$  توزیع‌های پیشینی دو مدل و  $Y$  مجموعه‌ی داده‌ها باشند، آنگاه مقدار عامل بیزی به صورت

$$B_{M_1, M_2} = \frac{f(Y | M_1)}{f(Y | M_2)} = \frac{\frac{\pi(M_1 | Y)}{\pi(M_1)}}{\frac{\pi(M_2 | Y)}{\pi(M_2)}}$$

تعریف می‌شود.

در جدول (۵) مقدار سه معیار اول برای مقایسه این سه مدل آمده است.

جدول (۵): مقدار معیارهای  $DIC$ ،  $AIC$  و  $BIC$  در مدل‌های ۱ تا ۳

مدل‌ها	$DIC$	$AIC$	$BIC$
مدل (۱)	۴۶۹۴/۲۲	۱۰۴۳۰	۱۱۳۷۰
مدل (۲)	۴۶۶۲/۸۷	۱۰۳۵۰	۱۰۸۵۰
مدل (۳)	۴۶۵۶/۸۹	۱۰۳۳۰	۱۰۷۵۰

با توجه به جدول (۵) ملاحظه می‌شود که مقدار  $DIC$  در مدل (۳) از مقدار متناظر در سایر مدل‌ها کمتر است. بنابراین این مدل که همانند مدل (۲) یک مدل همگن است و نسبت به آن

ممسک‌تر است، به سایر مدل‌ها ترجیح داده می‌شود. مقدار عامل بیزی برای مقایسه دوه‌دوی مدل‌ها نیز در جدول (۶) داده شده است. عدد هر خانه مقدار عامل بیزی برای مقایسه مدل داده شده در سطر مرتبط با مدل داده شده در ستون مرتبط است. مقدار عامل بیزی داده شده برای مقایسه مدل‌های (۱) و (۳) برابر  $7/14$  است که با توجه به [۱۵]، به‌طور بسیار قوی ارجحیت مدل (۳) را نسبت به مدل (۱) در برازش بهتر داده‌ها تأیید می‌کند. مقدار عامل بیزی برای مقایسه مدل‌های (۲) و (۳) برابر  $3/84$  است. با توجه به [۱۳] این مقدار نیز ارجحیت مدل (۳) نسبت به مدل (۲) در برازش مناسب داده‌ها را به‌طور قوی می‌پذیرد. بنابراین با توجه به نتایج داده شده در جدول‌های ۵ و ۶ نتیجه می‌شود که از بین سه مدل موجود، مدل (۳) به‌عنوان بهترین مدل برای برازش به داده‌ها انتخاب می‌شود.

جدول (۶): مقدار عامل بیزی برای مقایسه دوه‌دوی مدل‌های ۱ تا ۳

مدل (۱)	مدل (۲)	مدل (۳)
مدل (۱)	۰/۵۴	۰/۱۴
مدل (۲)	۱	۰/۲۶
مدل (۳)	۳/۸۴	۱

#### ۶- نتیجه‌گیری کلی

مدل بندی پاسخ‌های همبسته با دو مقیاس مختلف در داده‌های طولی یکی از موضوعات کاربردی است که می‌تواند موردعلاقه محققین این حوزه و سایر حوزه‌های مشابه مانند آمار فضایی، تحلیل بقاء، اقتصادسنجی و غیره باشد. در این حالت دو نوع ساختار همبستگی بین داده‌ها، در هر زمان و در طول زمان، وجود دارد. همچنین در مطالعات طولی، انصراف برخی آزمودنی‌ها از ادامه مطالعه، یکی از چالش‌های پیش رو است.

در این مقاله، به مدل بندی داده‌های طولی آمیخته ترتیبی و پیوسته در حالت وجود انصراف با استفاده از رهیافت بیزی پرداخته شد. برآورد پارامترهای مدل در برازش آن به داده‌های پی‌بندی از طریق رهیافت بیزی به دست آمد. برای مدل بندی انصراف از افزای تابع درست‌نمایی به توابع درست‌نمایی در هر زمان استفاده شد. توزیع توأم پاسخ‌های ترتیبی پیوسته و ترتیبی در هر زمان نیز به‌صورت حاصل ضرب توزیع حاشیه‌ای متغیر پیوسته در توزیع شرطی متغیر ترتیبی به‌شرط متغیر پیوسته دسته‌بندی شد. فرضیه‌ی همگنی پارامترهای مدل بر اساس معیارهای بیزی  $AIC$ ،  $DIC$ ،  $BIC$  و عامل بیزی آزمون شد و مشخص گردید که مدل همگن ممسک، مدل (۳)، بهترین مدل در بین سه مدل موجود است.

توسعه این مدل در حالت وجود اثرهای تصادفی، وجود بیش از دو متغیر پاسخ با مقیاس‌های اندازه‌گیری مختلف، وجود سایر مکانیسم‌های گمشدگی و همچنین بررسی شناساپذیر بودن پارامترهای مدل به صورت نظری نیز می‌تواند به عنوان تحقیقات پیش روی محققین این حوزه در نظر گرفته شود.

#### منابع

- [1] Olkin, L. and Tate, R.F. (1961). Multivariate correlation models with mixed discrete and continuous variables, *Annals of Mathematical Statistics*, **32**(2), 448-456.
- [2] Yang, Y., Kang, J., Mao, K., & Zhang, J. (2007). Regression models for mixed Poisson and continuous longitudinal data, *Statistics in Medicine*, **26**(20), 3782-3800.
- [3] Cox, D. R. and Wermuth, N. (1992). Response models for mixed binary and quantitative variables, *Biometrika*, **79**(3), 441-461.
- [4] Bahrami Samani, E. and Ganjali, M. (2008). A multivariate latent variable model for mixed continuous and ordinal responses, *World Applied Sciences Journal*, **3**(2), 294-299.
- [5] Bahrami Samani, E., Ganjali, M. and Khodaddadi, A. (2008). A latent variable model for mixed continuous and ordinal responses, *Journal of Statistical Theory and Applications*, **7**(3), 337-349.
- [6] Noorian, S., Ganjali, M., and Bahrami Samani, E. (2016). A Bayesian test of homogeneity of association parameter using transition modelling of longitudinal mixed responses, *Journal of Applied Statistics*, **43**(10), 1850-1863.
- [7] McKinley, T. J., Morters, M., and Wood, J. L. (2015). Bayesian model choice in cumulative link ordinal regression models. *Bayesian Analysis*, **10**(1), 1-30.
- [8] Johnson, V. E. and Albert, J. (1999). *Ordinal Data Modeling*, New York: Springer-Verlag.
- [9] Congdon, P. (2014). *Applied Bayesian modelling*, New York: John Wiley & Sons.
- [10] Sung, M., Soyer, R. and Nhan, N. (2007). Bayesian analysis of non-homogeneous Markov chains: Application to mental health data, *Statistics in Medicine*, **26**(15), 3000-3017.

- 
- [11] Gelman, A., and Rubin, D. B. (1992). Inference from Iterative Simulation Using Multiple Sequences, *Statistical Science*, **7**, 457–511.
- [12] Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **64**(4), 583-639.
- [13] Akaike, H. (1998). Information theory and an extension of the maximum likelihood principle, In *Selected Papers of Hirotugu Akaike* (pp. 199-213). Springer New York.
- [14] Schwarz, G. (1978). Estimating the dimension of a model, *The Annals of Statistics*, **6**(2), 461-464.
- [15] Kass, R. E. and Raftery, A. E. (1995). Bayes factors, *Journal of the American Statistical Association*, **90**(430), 773-795.



## Modeling Mixed Continuous and Ordinal Longitudinal Data Under Drop-out Mechanism

Sajad Noorian

Department of Statistics, University of Qom, Qom, Iran

### Abstract

In some longitudinal studies, especially in social, economic, medical and other fields, there may be two interested responses with two different scales at a time where they may be correlated with each other. Also, considering the nature of longitudinal studies, each of the responses associated with a subject over time can also be correlated. So two correlation structure should be considered simultaneously in the data analysis. In a longitudinal study, some subjects may not be available for any reason (such as displacement, death and others). In a longitudinal study, some subjects may withdraw for any reason (such as displacement, death, etc.) and their information is not available. In this case, joint modeling of longitudinal data and drop-out event is more desirable than separate modeling of either one. In this paper, the mathematics modeling of this type of data under drop-out mechanism is presented using Bayesian approach. A Simulations study and a real data analysis is used to evaluate the performance of the proposed model. This model includes the presented models for complete data as a special case when there is no drop-out in the data set. Also, some tests for choosing the best fitted model to data are performed in the real data analysis.

**Keywords:** Mixed Responses, Autoregressive Model, Cumulative Logistic Model, Joint Modeling, Longitudinal Studies, Drop-out Mechanism.

**Mathematics Subject Classification (2010):** 62H15, 62M02.