مجله علوم آماری، پاییز و زمستان ۱۳۹۴
جلد ۹، شماره ۲، ص ۱۴۹-۱۶۷

رگرسیون چندکی بیزی با تاوان لاسو و لاسوی تطبیق‌پذیر برای داده‌های طولی دودویی

علی آقامحمدی، سکینه محمدی
گروه آمار، دانشگاه زنجان

تاریخ دریافت: ۱۳۹۳/۳/۱۱ تاریخ آخرین بازنگری: ۱۳۹۳/۱۱/۴

چکیده : در بسیاری از مطالعات علوم پزشکی برای بیان سیر بیماری و تأثیر درمان از مطالعات طولی استفاده می‌شود، که در آن پاسخ‌ها به طور مکرر در طول زمان اندازه‌گیری می‌شوند. اما گاهی این پاسخ‌ها دو حالته و گسسته هستند. در این مقاله مدل رگرسیون چندکی با تاوان لاسو و لاسوی تطبیق‌پذیر برای داده‌های طولی با پاسخ‌های دوحالته ارائه شده و هر دو روش از دیدگاه آمار بیزی مورد تحلیل قرار می‌گیرد. با توجه به این‌که در هر دو روش توزیع‌های پسینی پارامترها به شکل بسته قابل حصول نیستند، توزیع‌های پسینی شرطی کامل پارامترها محاسبه شده و از الگوریتم نمونه‌گیری گیبز برای استنباط استفاده می‌شود. برای مقایسه کارایی روش‌های ارائه شده با روش‌های متداول، مطالعهٔ شبیه‌سازی انجام شده و در پایان نیز نحوهٔ کاربست مدل‌ها در قالب مثال کاربردی شرح داده خواهد شد.

واژه‌های کلیدی : رگرسیون چندکی دودویی، تاوان لاسو، تاوان لاسوی تطبیق‌پذیر، داده‌های طولی، نمونه‌گیری گیبز، استنباط بیزی.

آدرس الکترونیک مسئول مقاله: علی آقامحمدی، Aghamohammadi.ali@znu.ac.ir
کد موضوع‌بندی ریاضی (۲۰۱۰): ۶۲J۹۹، ۶۲F۱۵

۱ مقدمه

در بسیاری از مطالعه‌های مربوط به علوم پزشکی و اجتماعی، متغیر پاسخ برای هر فرد در چندین نوبت متوالی مشاهده می‌شود . برای مثال یک پزشک، سلامت بیماران را به طور هفتگی ارزیابی می‌کند تا دریابد، آیا داروی جدید موفقیتی داشته است یا خیر. به مطالعه‌ای که اندازه‌گیری مربوط به یک صفت در طول زمان مورد بررسی قرار می‌گیرد، مطالعهٔ طولی[۱] گویند. ممکن است فاصله‌های زمانی که هر واحد آزمایشی مورد مشاهده قرار می‌گیرد، یکسان نباشد. مهم‌ترین مزیت این نوع مطالعات تفکیک اثر زمان از تفاوت‌های فردی است که در مطالعه‌های مقطعی (پاسخ برای هر واحد آزمایشی، تنها یک بار مشاهده می‌شود) امکان‌پذیر نیست. لذا مطالعه‌های طولی نسبت به مطالعه‌های مقطعی دارای دقت بیشتری هستند (غلامی فشارکی و همکاران، ۱۳۹۲). گاهی پاسخ‌های ثبت شده به صورت کیفی، به‌ویژه دو حالته هستند. به‌عنوان مثال تأثیر یا عدم تأثیر یک داروی خاص روی نوعی بیماری. بدیهی است تحلیل این نوع داده‌ها با توجه به نوع متغیر پاسخ که دودویی است، روش‌های خاص خود را می‌طلبد. یکی از روش‌های تحلیل این متغیرها، رگرسیون چندکی[۲] دودویی است که مانند رگرسیون چندکی در مقایسه با روش رگرسیون میانگین دارای دو مزیت مهم است. نخست این که رگرسیون چندکی دودویی نسبت به ناپایداری واریانس و داده‌های دورافتاده حساس نیست و دوم این که اطلاعات جزئی‌تری نسبت به تأثیر متغیرهای توصیفی در چندک‌های مختلف توزیع متغیر پاسخ ارائه می‌دهد. به همین دلیل مطالعاتی نیز در زمینه تحلیل داده‌های طولی با پاسخ‌های دودویی صورت گرفته است. فیزماریس و لارد (۱۹۹۳) روشی بر پایهٔ درستنمایی برای تحلیل داده‌های طولی با پاسخ‌های دو حالته ارائه کردند. سپس گانکلوز (۲۰۰۲) این روش را برای تحلیل داده‌های طولی گسسته تعمیم داد. گانکلوز و آزلاینی (۲۰۰۸) استفاده از زنجیرهای مارکوفی را در روش درستنمایی مورد مطالعه قرار دادند. در ادامه گانکلوز و همکاران (۲۰۱۲) روش رگرسیون لوژستیک را برای تحلیل داده‌های طولی با پاسخ‌های دودویی ارائه کردند. در مطالعهٔ

[۱] Longitudinal study

[۲] Quantile regression

داده‌های طولی علاوه بر اثرات ثابت که تغییرات درون گروه‌ها (تغییرات در طول زمان) را کنترل می‌کند، اثرات تصادفی بین متغیرهای پاسخ نیز در مدل لحاظ می‌شود که در واقع تغییرات بین گروهی را کنترل می‌کند. در بسیاری از مسائل، تعداد مشاهدات از متغیر پاسخ نسبت به مشاهدات انجام گرفته در طول زمان از متغیر مربوطه، زیاد است. بنابراین در این مدل‌ها تعداد اثرات تصادفی (تعداد پارامترها) نیز خیلی زیاد شده و لذا تعبیر و تفسیر مدل مشکل می‌شود. کوئنکر (۲۰۰۴) رگرسیون چندکی را برای تحلیل این داده‌ها با افزودن تاوان لاسو روی اثرات تصادفی را مورد بررسی قرار داد که با ایجاد تاوان، اثرهای تصادفی را به سمت صفر منقبض[۳] کرده و اثرهای کم اهمیت را از مدل حذف و مدلی تنک[۴] ایجاد می‌کند.

اولین بار تیبشیرانی (۱۹۹۶) رگرسیون لاسو که با ایجاد تاوان لاسو روی پارامترها انجام می‌شود را ارائه کرد. سپس زو (۲۰۰۶) رگرسیون لاسوی تطبیق‌پذیر[۵] را به منظور گسترش رگرسیون لاسو پیشنهاد کرد. در این روش برخلاف تاوان لاسو که برای کلیهٔ ضرایب مدل اندازهٔ تاوان را یکسان در نظر می‌گیرد، اندازهٔ تاوان‌های مختلفی برای ضرایب رگرسیونی متفاوت لحاظ می‌شود. بررسی‌های زو (۲۰۰۶) مشخص کرد که در رگرسیون با تاوان لاسوی تطبیق‌پذیر اریبی در برآورد پارامترها در مقایسه با روش لاسو کمتر است. الحمزوی و همکاران (۲۰۱۲) نیز با تعریف توزیع پیشینی به‌صورت لاپلاس متقارن روی ضرایب رگرسیونی، مدل رگرسیون چندکی با تاوان لاسوی تطبیق‌پذیر را برای داده‌های مقطعی مورد مطالعه قرار دادند. در این مقاله هدف بررسی رگرسیون چندکی با تاوان لاسو و لاسوی تطبیق‌پذیر روی اثرات تصادفی در داده‌های طولی با پاسخ‌های دو حالته از دیدگاه آمار بیزی است. برای این منظور تاوان لاسو و لاسوی تطبیق‌پذیر روی اثرات تصادفی در داده‌های طولی دودویی از طریق تعریف توزیع‌های پیشینی مناسب در نظر گرفته می‌شود و هر دو روش از دیدگاه آمار بیزی مورد تحلیل و ارزیابی قرار می‌گیرند. در بخش ۲ مطالب کلی در مورد مدل رگرسیون چندکی برای

---

۳ Shrink

۴ Sparse

۵ Adaptive Lasso

داده‌های طولی بیان می‌شود. در بخش‌های ۳ و ۴ مدل رگرسیون چندکی دودویی با تاوان لاسو و لاسوی تطبیق‌پذیر از دیدگاه آمار بیزی ارائه می‌شود. در بخش ۵ در مطالعه‌ای شبیه‌سازی، کارایی مدل‌های ارائه شده با مدل‌های دیگر مورد مقایسه قرار می‌گیرد. بخش ۶ نیز شامل تحلیل داده‌های واقعی و ارزیابی کارایی مدل‌ها است.

۲ تعریف مدل

مدل رگرسیون خطی

$$y_i = x_i'\beta + \epsilon_i, \qquad i = ۱, \ldots, n$$

را در نظر بگیرید، که در آن x_i بردار مربوط به متغیرهای توصیفی، β یک بردار $k \times ۱$ بعدی از پارامترها و ϵ_i مولفهٔ خطا است. در مدل رگرسیونی میانگین، هدف استنباط در مورد متوسط مقدار متغیر پاسخ به‌ازای مقادیر متفاوت متغیرهای توصیفی است، لذا $E(y_i|x_i) = x_i'\beta$ مورد توجه است. اما در رگرسیون چندکی، تابع چندک شرطی

$$Q_\tau(y_i|x_i) = x_i'\beta_\tau \qquad i = ۱, \ldots, n$$

مورد توجه است، که در آن $Q_\tau(\cdot)$ معکوس تابع توزیع تجمعی متغیر پاسخ y_i به شرط معلوم بودن بردار x_i است. به عبارت دیگر $x_i'\beta_\tau$ چندک τام شرطی متغیر y_i را نشان می‌دهد. در رگرسیون معمولی میانگین توزیع خطاها صفر در نظر گرفته می‌شود، اما در رگرسیون چندکی، چندک τام ϵ_iها برابر صفر هستند، یعنی $\int_{-\infty}^{۰} f_\tau(\epsilon_i)d\epsilon_i = \tau$ برای $i = ۱, \ldots, n$. در رگرسیون چندکی، ضرایب رگرسیونی یعنی پارامتر β_τ از مینیمم کردن رابطهٔ

$$\Sigma_{i=۱}^{n} \rho_\tau(y_i - x_i'\beta_\tau), \qquad (۱)$$

نسبت به β_τ برآورد می‌شود، که در آن $\rho_\tau(u) = \frac{|u|+(۲\tau-۱)u}{۲}$ تابع زیان است (کوئنکر و ماکادو، ۱۹۹۹). چون رابطهٔ (۱) نسبت به β_τ در مبدأ مشتق پذیر نیست، راه‌حل تحلیلی برای برآورد ضرایب رگرسیونی وجود ندارد. برای رفع این مشکل

کوئنکر و ماکادو (۱۹۹۹) استفاده از توزیع لاپلاس نامتقارن را پیشنهاد کردند. متغیر تصادفی Y را دارای توزیع لاپلاس نامتقارن[۶] (ALD) با پارامترهای τ، σ و μ گویند و با نماد $Y \sim ALD(\mu, \sigma, \tau)$ نشان می‌دهند، هر گاه تابع چگالی آن به‌صورت

$$f(y|\mu, \sigma, \tau) = \frac{\tau(1-\tau)}{\sigma} \exp\{-\rho_\tau(\frac{y-\mu}{\sigma})\},$$

باشد، که در آن $0 < \tau < 1$ پارامتر چولگی، σ پارامتر مقیاس و $-\infty < \mu < +\infty$ پارامتر مکانی است. با فرض این که $Y_i \sim ALD(x_i'\beta_\tau, \sigma, \tau)$، آن‌گاه تابع درستنمایی به‌صورت

$$L(\beta_\tau|y, x) \propto (\frac{1}{\sigma})^n \exp\{-\Sigma_{i=1}^n \rho_\tau(\frac{y_i - x_i'\beta_\tau}{\sigma})\}$$

به‌دست می‌آید، که ماکسیمم کردن آن در حضور پارامتر مزاحم σ با مینیمم کردن رابطهٔ (۱) نسبت به β_τ معادل است. به همین دلیل در بسیاری از موارد، تحلیل رگرسیون چندکی با استفاده از این توزیع مورد بررسی قرار می‌گیرد.

مدل رگرسیون چندکی برای داده‌های طولی با اثرات تصادفی به صورت

$$y_{ij} = x_{ij}'\beta_\tau + \alpha_i + \epsilon_{ij}; \quad i = 1, \ldots, n, \; j = 1, \ldots, n_i, \; \Sigma_i n_i = N$$

تعریف می‌شود، که در آن y_{ij}، jامین پاسخ اندازه‌گیری شده روی iامین واحد آزمایشی و x_{ij} بردار مربوط به متغیرهای توصیفی و β_τ بردار $k \times 1$ بعدی از پارامترها و α_i اثر تصادفی مربوط به پاسخ y_{ij} را نشان می‌دهد. تابع چندک شرطی در این مدل به صورت $Q_\tau(y_{ij}|x_{ij}) = x_{ij}'\beta_\tau + \alpha_i$ بیان می‌شود. با فرض تاوان لاسو روی اثرات تصادفی به صورت $p(\alpha) = \Sigma_{i=1}^n |\alpha_i|$، برآورد پارامترهای β_τ و $\alpha = (\alpha_1, \ldots, \alpha_n)$ از دیدگاه آمار بسامدی از مینیمم کردن تابع زیان

$$\Sigma_{i=1}^n \Sigma_{j=1}^{n_i} \rho_\tau(y_{ij} - x_{ij}'\beta_\tau - \alpha_i) + \lambda \Sigma_{i=1}^n |\alpha_i|$$

نسبت به α و β_τ به‌دست می‌آید (کوئنکر، ۲۰۰۴)، که در آن λ را پارامتر تاوانیدن گویند و میزان انقباض به سمت صفر پارامترهای α_i را کنترل می‌کند. هر چه مقدار

[۶] Asymmetrice Laplace Distribution

آن بزرگتر باشد، انقباض به صفر نیز بیشتر و اثرهای تصادفی زیادی از مدل حذف خواهند شد. در این رهیافت برآورد β_τ تحت تأثیر اندازهٔ λ است که معمولاً از روش اعتبارسنجی متقابل[۷] محاسبه می‌شود.

۳ رگرسیون چندکی دودویی با تاوان لاسو

یکی از روش‌های معمول برای تعریف مدل‌های رگرسیونی با پاسخ‌های دودویی استفاده از متغیرهای پنهان[۸] است (بنویت و همکاران، ۲۰۱۳). مدل رگرسیون چندکی دودویی در داده‌های طولی به‌صورت

$$y^*_{ij} = x'_{ij}\beta_\tau + \alpha_i + \epsilon_{ij},\ y_{ij} = g(y^*_{ij}),\ \ i = ۱,\dots n,\ j = ۱,\dots n_i, \qquad (۲)$$

تعریف می‌شود، که در آن $g(\cdot)$ تابع پیوند[۹] و به‌صورت $g(y^*_{ij}) = I(y^*_{ij} > ۰)$ تعریف شده ($I(\cdot)$ تابع مشخصه است) و y_{ij}، jامین پاسخ مشاهده شده برای iامین واحد آزمایشی و y^*_{ij} متغیر پنهان و غیر قابل مشاهده است. چون چندک τام ϵ_{ij}ها برابر صفر فرض می‌شوند، $x'_{ij}\beta_\tau + \alpha_i$ چندک τام y^*_{ij} خواهد بود. از آنجایی‌که تابع پیوند یک تابع یکنوا است، $g(x'_{ij}\beta_\tau + \alpha_i)$ چندک τام متغیر $y_{ij} = g(y^*_{ij})$ است. بنابراین در این مدل تابع زیان با تاوان لاسو روی اثرات تصادفی به‌صورت

$$\sum_{i=۱}^{n}\sum_{j=۱}^{n_i}\rho_\tau(y^*_{ij} - x'_{ij}\beta_\tau - \alpha_i) + \lambda\sum_{i=۱}^{n}|\alpha_i| \qquad (۳)$$

به‌دست می‌آید. با فرض این‌که $\epsilon_{ij} \sim ALD(۰,\sigma,\tau)$ و توزیع پیشینی برای α_i به‌صورت $\pi(\alpha_i|\lambda,\sigma) = \frac{\lambda}{۲\sqrt{\sigma}}\exp\{-\frac{\lambda|\alpha_i|}{\sqrt{\sigma}}\}$ در نظر گرفته شود، به‌ازای $\nu = \frac{\lambda}{\sqrt{\sigma}}$، توزیع پسینی α به‌صورت

$$\pi(\alpha|y,x,\sigma,\lambda,\beta_\tau) = \\ \sigma^{-N}\exp\{-\sigma^{-۱}\sum_{i=۱}^{n}\sum_{j=۱}^{n_i}\rho_\tau(y^*_{ij} - x'_{ij}\beta_\tau - \alpha_i)\}\prod_{i=۱}^{n}\frac{\nu}{۲}\exp\{-\nu|\alpha_i|\} \qquad (۴)$$

---

۷ Cross validation

۸ Hidden variable

۹ Link function

به‌دست می‌آید. مینیمم کردن تابع هدف (۳) نسبت به α و β_τ معادل ماکسیمم کردن تابع درستنمایی (۴) در حضور پارامتر مزاحم σ است. بنابراین می‌توان توزیع لاپلاس نامتقارن را برای این مدل نیز مورد استفاده قرار داد. چون استفاده از تابع درستنمایی توزیع لاپلاس نامتقارن به دلیل وجود قدرمطلق آسان نیست، در تحلیل‌های بیزی معمولاً از شکل آمیختهٔ این توزیع استفاده می‌شود. اگر $u \sim ALD(0, \sigma, \tau)$ و متغیرهای تصادفی z و e به ترتیب دارای توزیع نرمال استاندارد و نمایی با میانگین $\frac{\sigma}{\tau(1-\tau)}$ و مستقل از هم باشند، آنگاه

$$u = k_1 e + \sqrt{2\sigma e} z$$

که در آن $k_1 = (1 - 2\tau)$ (کازومی و کابایاشی، ۲۰۱۱). با استفاده از این خاصیت توزیع لاپلاس نامتقارن و با توجه به این که $\epsilon_{ij} \sim ALD(0, \sigma, \tau)$، مدل رابطهٔ (۲) به‌صورت

$$y^*_{ij} = x'_{ij}\beta_\tau + \alpha_i + k_1 e_{ij} + \sqrt{2\sigma e_{ij}} z_{ij} \qquad (۵)$$

خواهد بود، که در آن z_{ij} و e_{ij} به ترتیب دارای توزیع نرمال استاندارد و نمایی با میانگین $\frac{\sigma}{\tau(1-\tau)}$ و مستقل از هم هستند. چون توزیع پیشینی α_iها لاپلاس متقارن است، می‌توان آن را به‌صورت توزیع آمیخته از دو توزیع

$$\alpha_i | s_i \sim N(0, s_i), \qquad s_i | \nu^2 \sim Exp(\frac{\nu^2}{2}) \qquad (۶)$$

نوشت، که در آن $Exp(\frac{\nu^2}{2})$ توزیع نمایی با میانگین $\frac{2}{v^2}$ همان توزیع آمیختگی است (پارک و کسلا، ۲۰۰۸). چون هدف تحلیل بیزی است، برای ابرپارامتر ν^2 (به جای ν) توزیع پیشینی مزدوج نمایی به‌صورت $\pi(\nu^2|\phi) = \phi \exp\{-\phi\nu^2\}$ و برای σ و β_τ به ترتیب وارون گاما و توزیع نرمال چند متغیره در نظر گرفته می‌شود. حال با توجه به مدل رابطهٔ (۵)، (۶) و توزیع‌های پیشینی تعریف شده، مدل سلسله مراتبی را می‌توان به‌صورت

$$y_{ij} = g(y^*_{ij}), \quad y^*_{ij} | e_{ij}, \beta_\tau, \sigma, \alpha_i \sim N(x'_{ij}\beta_\tau + \alpha_i + k_1 e_{ij}, 2\sigma e_{ij}),$$
$$e_{ij} | \sigma \sim Exp(\frac{\tau(1-\tau)}{\sigma}), \quad \beta_\tau | b_0, B_0 \sim N_k(b_0, B_0), \quad \alpha_i | s_i \sim N(0, s_i),$$

$$s_i|\nu^{۲} \sim Exp(\frac{\nu^{۲}}{۲}),\ \nu^{۲}|\phi \sim Exp(\phi),\ \pi(\phi) \propto \frac{۱}{\phi},\ \sigma|c_۰, d_۰ \sim IG(c_۰, d_۰),$$

در نظر گرفت، که در آن $IG(a,b)$ توزیع وارون گاما با پارامترهای a و b است. با توجه به شکل سلسله مراتبی فوق، توزیع پیسینی کلیه پارامترها و متغیرهای پنهان به‌صورت

$$\begin{aligned}
&\pi(y^*, \beta_\tau, \sigma, s, \nu^{۲}, \alpha, \phi|y, x) \propto \prod_{i=۱}^{n}\prod_{j=۱}^{n_i} \frac{۱}{\sqrt{۴\sigma\pi e_{ij}}} \\
&\times \exp\{-\frac{۱}{۴\sigma e_{ij}}(y^*_{ij} - x'_{ij}\beta_\tau - \alpha_i - k_۱ e_{ij})^{۲}\} \\
&\times \{I(y_{ij} = ۰)I(y^*_{ij} \le ۰) + I(y_{ij} = ۱)I(y^*_{ij} > ۰)\} \\
&\times (\frac{۱}{\sigma})^N \exp\{-\Sigma_{i=۱}^{n}\Sigma_{j=۱}^{n_i}\frac{e_{ij}}{\sigma}\} \times \prod_{i=۱}^{n}\frac{۱}{\sqrt{۲\pi s_i}}\exp\{-\frac{\alpha_i^{۲}}{۲s_i}\}\frac{\nu^{۲}}{۲}\exp\{-\frac{\nu^{۲}s_i}{۲}\} \\
&\times \exp\{-\frac{۱}{۲}(\beta_\tau - b_۰)B_۰^{-۱}(\beta_\tau - b_۰)\}\exp\{-\phi\nu^{۲}\} \times (\frac{۱}{\sigma})^{c_۰+۱}\exp\{-\frac{d_۰}{\sigma}\}
\end{aligned}$$

به‌دست می‌آیند، که در آن $s = (s_۱, \ldots, s_n)$. ملاحظه می‌شود که توزیع پسینی کامل پارامترها و متغیرهای پنهان توزیع شناخته شده‌ای نیست. اما با اندکی محاسبات جبری توزیع‌های پسینی شرطی کامل متغیرهای پنهان و پارامترها به‌صورت

$$\pi(y^*_{ij}|y_{ij}, e_{ij}, \alpha, \beta_\tau) = \begin{cases} N(x'_{ij}\beta_\tau + \alpha_i + k_۱ e_{ij}, ۲\sigma e_{ij})I(y^*_{ij} > ۰), y_{ij} = ۱ \\ N(x'_{ij}\beta_\tau + \alpha_i + k_۱ e_{ij}, ۲\sigma e_{ij})I(y^*_{ij} \le ۰), y_{ij} = ۰ \end{cases}$$

$$\pi(e_{ij}|.) \sim GIG(\frac{۱}{۲}, \hat{\gamma}_{ij}, \hat{\delta}_{ij})$$

به‌دست می‌آیند، که در آن $\hat{\gamma}^{۲}_{ij} = \frac{(y^*_{ij} - x'_{ij}\beta_\tau - \alpha_i)^{۲}}{۲\sigma}$، $\hat{\delta}^{۲}_{ij} = \frac{k_۱^{۲}}{۲\sigma} + \frac{۲\tau(۱-\tau)}{\sigma}$ و $GIG(r,m,n)$ نشان‌دهندهٔ توزیع تعمیم یافته وارون نرمال[۱۰] با پارامترهای r، m و n است. تابع چگالی این توزیع به‌صورت

$$f(x|r,m,n) \propto x^{r-۱}\exp\{-\frac{۱}{۲}(m^{۲}x^{-۱} + n^{۲}x)\}, x > ۰, -\infty < r < \infty, m, n > ۰$$

است، که با استفاده از تابع $rgig()$ در پکیج $ghyp$ در نرم افزار R می‌توان از این توزیع نمونهٔ تصادفی تولید کرد (لیوزی و بریمن، ۲۰۱۴).

[۱۰] Generalized Invers Gaussian

توزیع شرطی کامل σ نیز به‌صورت

$$\pi(\sigma|\cdot) \sim IG(\nu_1, w_1)$$

است، که در آن $w_1 = \sum_{i=1}^{n}\sum_{j=1}^{n_i}(\frac{(y_{ij}^{*}-x_{ij}'\beta_\tau-\alpha_i-k_1e_{ij})^2}{4e_{ij}} + \tau(1-\tau)e_{ij}) + d_0$ و $\nu_1 = \frac{3N}{2} + c_0$. همچنین $\pi(\alpha_i|.) \sim N(\bar{\mu}_i, \tilde{\sigma}_i)$، که در آن

$$\bar{\mu}_i = \tilde{\sigma}_i(\frac{1}{2\sigma}\sum_{j=1}^{n_i}\frac{(y_{ij}^{*}-x_{ij}'\beta_\tau-k_1e_{ij})}{e_{ij}}), \quad \tilde{\sigma}_i = (\frac{1}{s_i}+\frac{1}{2\sigma}\sum_{j=1}^{n_i}\frac{1}{e_{ij}})^{-1}$$

$$\pi(\phi|\cdot) \sim Exp(\nu^2), \quad \pi(s_i|\cdot) \sim GIG(\frac{1}{2}, \sqrt{\alpha_i^2}, \sqrt{\nu^2}),$$

به‌علاوه $\pi(\nu^2|\cdot) \sim G(n+1, \frac{\sum_{i=1}^{n}s_i}{2} + \phi)$، که در آن $G(a,b)$ نشان دهندهٔ توزیع گاما با پارامترهای a و b است. بالاخره توزیع پسینی شرطی کامل β_τ به‌صورت

$$\pi(\beta_\tau|\cdot) \sim N_k(Bb, B)$$

به‌دست می‌آید، که در آن $b = \frac{1}{2\sigma}\sum_{i=1}^{n}\sum_{j=1}^{n_i}\frac{x_{ij}(y_{ij}^{*}-\alpha_i-k_1e_{ij})}{e_{ij}} + B_0^{-1}b_0$ و $B^{-1} = \frac{1}{2\sigma}\sum_{i=1}^{n}\sum_{j=1}^{n_i}\frac{x_{ij}x_{ij}'}{e_{ij}} + B_0^{-1}$. حال با مشخص شدن توزیع پسینی پارامترها و متغیرهای پنهان می‌توان با روش نمونه گیری گیبز نمونه‌های کافی تولید کرده و استنباط‌ها را انجام داد.

۴ رگرسیون چندکی دودویی با تاوان لاسوی تطبیق‌پذیر

در این روش به منظور بسط روش قبل اندازهٔ تاوان‌های متفاوتی روی هر یک از اثرات تصادفی مدل در نظر گرفته می‌شود. بنابراین تابع زیان در این مدل به صورت

$$\sum_{i=1}^{n}\sum_{j=1}^{n_i}\rho_\tau(y_{ij} - g(x_{ij}'\beta_\tau + \alpha_i)) + \sum_{i=1}^{n}\lambda_i|\alpha_i| \qquad (۷)$$

تعریف می‌شود، که در آن عبارت سمت راست را تاوان لاسوی تطبیق‌پذیر گویند (زو، ۲۰۰۶). اگر در این مدل توزیع پیشینی برای α_iها به‌صورت $\pi(\alpha_i|\lambda_i,\sigma) = \frac{\lambda_i}{2\sqrt{\sigma}}\exp\{-\frac{\lambda_i|\alpha_i|}{\sqrt{\sigma}}\}$ تعریف شود و قرار دهیم، $\nu_i = \frac{\lambda_i}{\sqrt{\sigma}}$، آنگاه توزیع پسینی بردار α به‌صورت

$$\pi(\alpha|y,x,\sigma,\nu,\beta_\tau) =$$

$$(\sigma)^{-N}\exp\{-\sigma^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n_i}\rho_\tau(y_{ij}^*-x_{ij}'\beta_\tau-\alpha_i)\}\prod_{i=1}^{n}\frac{\nu_i}{2}\exp\{-\nu_i|\alpha_i|\}$$

به‌دست می‌آید. مینیمم کردن رابطهٔ (۷) با ماکسیمم کردن توزیع پسینی α در حضور پارامتر مزاحم σ معادل است. لذا در این مدل نیز می‌توان از توزیع لاپلاس نامتقارن استفاده کرد. اگر توزیع پیشینی α_i به شکل آمیخته از دو توزیع

$$\alpha_i|s_i\sim N(0,s_i),\quad s_i|\nu_i^2\sim Exp(\frac{\nu_i^2}{2})$$

در نظر گرفته شود، با استفاده از شکل آمیختهٔ توزیع لاپلاس نامتقارن، مدل را می‌توان مانند بخش ۳ به‌صورت سلسله مراتبی زیر نوشت:

$$\begin{aligned}
&y_{ij}=g(y_{ij}^*),\quad y_{ij}^*|\beta_\tau,\sigma,e,\alpha_i\sim N(x_{ij}'\beta_\tau+\alpha_i+k_1e_{ij},2\sigma e_{ij})\\
&e_{ij}|\sigma\sim Exp(\frac{\tau(1-\tau)}{\sigma}),\quad \beta_\tau|b_0,B_0\sim N_k(b_0,B_0),\quad s_i|\nu_i^2\sim Exp(\frac{\nu_i^2}{2})\\
&\alpha_i|s_i\sim N(0,s_i),\ \nu_i^2|\phi\sim Exp(\phi),\ \pi(\phi)\propto\frac{1}{\phi},\ \sigma|c_0,d_0\sim IG(c_0,d_0).
\end{aligned}$$

توزیع پسینی کامل پارامترها و متغیرهای پنهان در این مدل نیز به شکل بسته قابل حصول نیستند. اما توزیع پسینی شرطی کامل پارامترها و متغیرهای پنهان قابل محاسبه‌اند. توزیع‌های پسینی y^*، β_τ، α و σ مانند بخش ۳ می‌باشد. اما توزیع‌های پسینی برای سایر پارامترها در این مدل به‌صورت

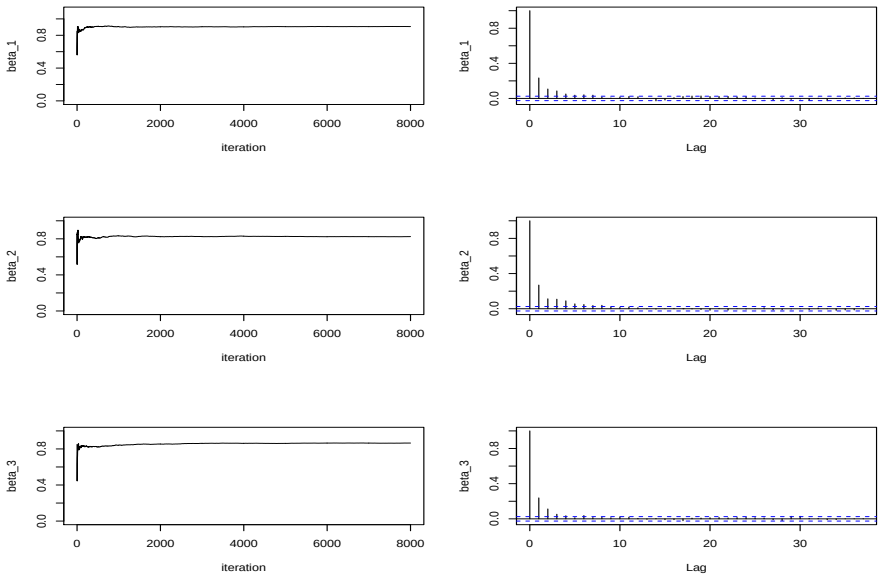
$$\begin{aligned}
&\pi(\nu_i^2|\cdot)\sim G(2,\frac{s_i}{2}+\phi),\quad \pi(\phi^2|\cdot)\sim G(n,\Sigma_{i=1}^n\nu_i^2),\\
&\pi(s_i|\cdot)\sim GIG(\frac{1}{2},\sqrt{\alpha_i^2},\sqrt{\nu_i^2}),
\end{aligned}$$

به‌دست می‌آیند. در این روش نیز با مشخص شدن توزیع پسینی هر یک از پارامترها و متغیرهای پنهان می‌توان با روش نمونه گیری گیبز، نمونه‌های کافی تولید کرده و استنباط‌ها را انجام داد.

۵ مطالعهٔ شبیه‌سازی

در این بخش برای مقایسه کارایی روش‌های ارائه شده با روش‌های متداول دیگر مطالعه‌ای شبیه‌سازی انجام می‌پذیرد. مدل به‌صورت

$$y_{ij}^*=x_{ij}^1\beta_1+x_{ij}^2\beta_2+x_{ij}^3\beta_3+x_{ij}^4\beta_4+x_{ij}^5\beta_5+x_{ij}^6\beta_6+\alpha_i+\epsilon_{ij}$$



شکل ۱: نمودار اثر و تابع خودهمبستگی در روش رگرسیون چندکی با تاوان لاسو ($\tau = ۰/۵$)

در نظر گرفته شده است، که در آن $i = ۱,\ldots,۵۰$ و $j = ۱,\ldots,۵$. توجه شود که تعداد متغیرهای پاسخ در مقایسه با متغیرهای توصیفی زیاد است. متغیرهای توصیفی یعنی x_{ij}^kها به صورت $x_{ij}^k \sim N(۰, ۱)$ برای $k = ۱,\ldots,۶$ تولید کرده و توزیع اثرهای تصادفی به‌صورت $\alpha_i \sim N(۰, ۴)$ فرض شده است. برای β_τ نیز بردارهای

$$\beta_\tau = (۵, ۰, ۰, ۰, ۰, ۱), \quad \beta_\tau = (۳, ۱/۵, ۰, ۰, ۰, ۱),$$
$$\beta_\tau = (۰/۸۵, ۰/۸۵, ۰/۸۵, ۰/۸۵, ۰/۸۵, ۰/۸۵)$$

در نظر گرفته شده است، که در آن بردار اول متناظر با یک مدل کاملاً تنک است. برای مولفهٔ خطا نیز توزیع‌های نرمال استاندارد، تی با سه درجه آزادی و توزیع لاپلاس متقارن در نظر گرفته شده است. به منظور کاهش حساسیت مدل به توزیع‌های پیشینی برای پارامتر β_τ توزیع پیشینی را $N_۶(۰, ۱۰۰I)$ و برای ابرپارامتر σ توزیع پیشینی تخت[۱۱] به صورت $IG(۰/۰۱, ۰/۰۱)$ فرض می‌شود. روش

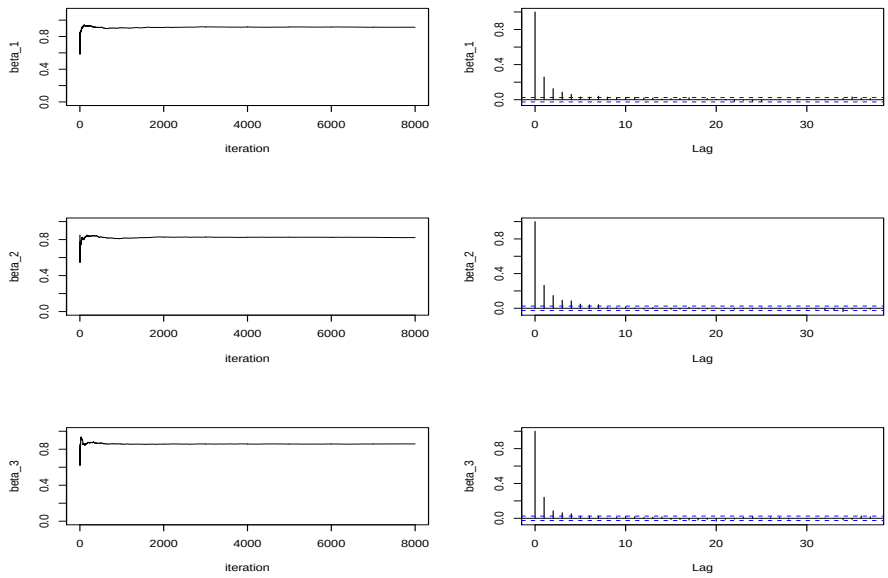
[۱۱] Flat

رگرسیون چندکی دودویی با تاوان لاسو (BL) و روش رگرسیون چندکی دودویی با تاوان لاسوی تطبیق‌پذیر (BAL) برای داده‌های طولی، با دو روش بیزی و بسامدی مورد مقایسه قرار می‌گیرد. روش اول (بنویت و همکاران، ۲۰۱۳) مدل رگرسیون چندکی دودویی با تاوان لاسو از دیدگاه آمار بیزی برای داده‌های مقطعی است که تاوان لاسو روی اثرات ثابت اعمال شده و اثرات تصادفی در مدل لحاظ نمی‌شود، این مدل به اختصار مدل BA نامیده می‌شود. روش دوم (گانکلوز و همکاران، ۲۰۱۲) مدل رگرسیون لوژستیک برای داده‌های طولی با پاسخ‌های دودویی (به اختصار RLG) است که پکیج محاسبهٔ آن (bild) در نرم‌افزار R موجود است.

برای برآورد پارامترها با روش نمونه‌گیری گیبز، ۸۰۰۰ نمونه از توزیع‌های پسینی شرطی کامل پارامترها تولید کرده و ۲۰۰۰ نمونهٔ اول به عنوان دورهٔ همگرای مطلوب کنار گذاشته شده است. برای ارزیابی همگرایی زنجیر و تعیین تقریبی تعداد نمونه‌ها به‌عنوان همگرایی مطلوب از آمارهٔ $\hat{R}$ که توسط گلمن و همکاران (۱۹۹۵) ارائه شده استفاده می‌شود. برای کلیه روش‌ها تقریباً بعد از ۲۰۰۰ نمونهٔ اولیه، همگرایی مطلوب رخ می‌دهد که شکل ۱ نمودارهای اثر و تابع خود همبستگی نمونه‌ای زنجیر را برای روش رگرسیون چندکی با تاوان لاسو و شکل ۲ این نمودارها را برای روش رگرسیون چندکی با تاوان لاسوی تطبیق‌پذیر برای $\beta_\tau = (۰/۸۵, ۰/۸۵, ۰/۸۵, ۰/۸۵, ۰/۸۵, ۰/۸۵)$ نشان می‌دهند. هر دو نمودار به خوبی همگرایی زنجیرهای مارکف تولید شده از توزیع‌های پسینی ضرایب را نشان می‌دهند. با توجه به نمودارهای اثر نیز در هر دو روش می‌توان گفت که تقریباً بعد از ۲۰۰۰ تکرار، همگرایی مطلوب رخ داده است. جدول ۱ برآورد پارامتر β_τ برای چهار روش را در مقایسه با مقدار واقعی β_τ نشان می‌دهد. همان‌طور که مشاهده می‌شود روش‌های ارائه شده عملکرد بهتری در برآورد پارامتر β_τ در مقایسه با مقدار واقعی دارند. برای مقایسهٔ کارایی چهار روش عنوان شده و به منظور لحاظ کردن تغییرپذیری نتایج حاصل از شبیه‌سازی، ۲۰۰ مرتبه روند شبیه‌سازی را تکرار کرده و سه معیار اریبی[۱۲] پارامترها، متوسط قدرمطلق خطا[۱۳] و مجذور

[۱۲] Bias

[۱۳] Mean absolute error



شکل ۲: نمودار اثر و تابع خودهمبستگی در روش رگرسیون چندکی با تاوان لاسوی تطبیق‌پذیر ($\tau = ۰/۵$)

متوسط توان دوم خطا[۱۴] به‌صورت

$$
\begin{aligned}
Bias(\hat{\beta_k}) &= \frac{۱}{۲۰۰}\sum_{i=۱}^{n}(\hat{\beta}_{ik} - \beta_k), \\
MAE(\hat{\beta_k}) &= \frac{۱}{۲۰۰}\sum_{i=۱}^{n}|\hat{\beta}_{ik} - \beta_k|, \\
RMSE(\hat{\beta_k}) &= \sqrt{\frac{۱}{۲۰۰}\sum_{i=۱}^{n}(\hat{\beta}_{ik} - \beta_k)^۲}
\end{aligned}
$$

محاسبه می‌شود. واضح است هر چه اندازهٔ این معیارها کوچک باشد، حاکی از کارایی بالای مدل می‌باشد. چون نوشتن تمام مقادیر با توجه به وجود چهار مدل، سه روش برآورد برای سه توزیع خطا و هجده پارامتر در قالب جدول امکان‌پذیر نیست، بنابراین نتایج در قالب نمودارهای جعبه‌ای خلاصه شده که نتایج آن در شکل‌های شمارهٔ ۳، ۴ و ۵ ارائه شده است. هر چه نمودارهای جعبه‌ای کوچکتر و

[۱۴] Root mean square error

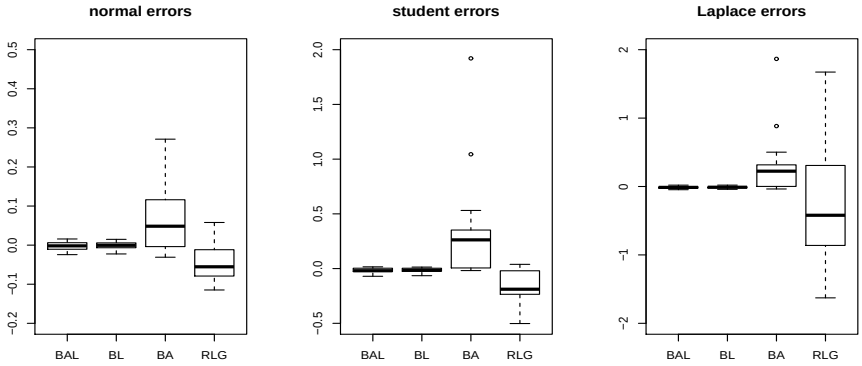
جدول ۱: میانگین پسینی پارامترها با توزیع خطاهای نرمال و $\tau = ۰/۵$

مدل	$\beta_۱$	$\beta_۲$	$\beta_۳$	$\beta_۴$	$\beta_۵$	$\beta_۶$
β_{True}	۵	۰	۰	۰	۰	۱
BAL	۵/۰۲۵	−۰/۰۰۲	۰/۰۴۸	۰/۰۱۴	۰/۰۶۸	۰/۸۶۳
BL	۵/۰۱۲	−۰/۰۰۷	۰/۰۵۵	−۰/۰۰۲	۰/۰۸۶	۰/۸۶۱
BA	۴/۵۶۴	−۰/۲۲۱	۰/۱۹۸	۰/۳۳۰	−۰/۳۸۶	۰/۶۸۲
RGL	۴/۸۳۰	۰/۰۹۱۳	۰/۰۳۳	−۰/۲۵۴	۰/۰۰۱	۱/۰۳۱
β_{True}	۳	۱/ ۵	۰	۰	۰	۱
BAL	۳/۰۸۵	۱/۵۲۳	−۰/۲۰۲	۰/۰۴۶	−۰/۰۴۲	۱/۰۵۷
BL	۳/۱۷	۱/۵۷	−۰/۰۷۴	۰/۰۸۶	−۰/۱۴۶	۱/۲۲
BA	۲/۱۱۵	۱/۲۲	−۰/۰۵۸	−۰/۳۵۰	−۰/۱۸۳	۰/۹۹۶
RGL	۳/۰۷۱	۰/۰۹۱۳	۰/۰۳۳	−۰/۲۵۴	۰/۰۰۱	۱/۰۹۱
β_{True}	۰/۸۵	۰/۸۵	۰/۸۵	۰/۸۵	۰/۸۵	۰/۸۵
BAL	۰/۸۹۰	۰/۸۸۱	۰/۸۷۳	۰/۹۴۴	۰/۸۰۴	۰/۸۲۴
BL	۰/۸۸۹	۰/۸۷۴	۰/۸۶۷	۰/۹۴۳	۰/۸۰۳	۰/۸۱۷
BA	۰/۹۲۷	۰/۸۷۳	۰/۶۶۸	۰/۸۲۳	۰/۵۱۳	۰/۶۷۲
RGL	۰/۷۶۵	۱/۰۰۱	۰/۶۳۹	۰/۹۷۶	۱/۱۲۰	۰/۴۵۰

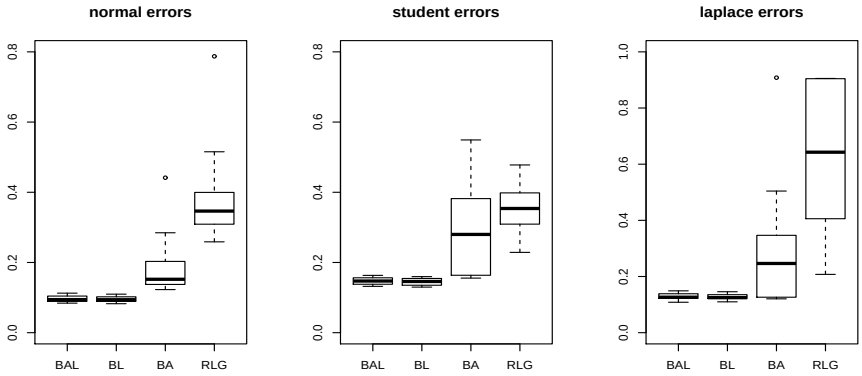
نزدیک صفر باشد، نشان دهندهٔ اریبی کمتر و برآورد بهتر پارامترها است. شکل ۳ عملکرد روش‌ها را از نقطه نظر اریبی مورد مقایسه می‌دهد. با توجه به این شکل دو روش BAL و BL در تمامی توزیع‌های فرض شده برای خطاها عملکرد بهتری نسبت به سایر روش‌ها دارد. اما این دو روش تقریباً عملکرد یکسانی در برآورد پارامترها دارند. شکل‌های ۴ و ۵ نیز عملکرد روش‌ها را به لحاظ معیارهای RMSE و MAE مورد بررسی قرار می‌دهد. با توجه به این دو شکل نیز مشاهده می‌شود که همانند نمودارهای اریبی، در این نمودارها نیز روش‌های ارائه شده برآورد بهتری از پارامترها در تمامی توزیع‌های فرض شده برای خطا دارند.

۶ تحلیل داده‌های واقعی

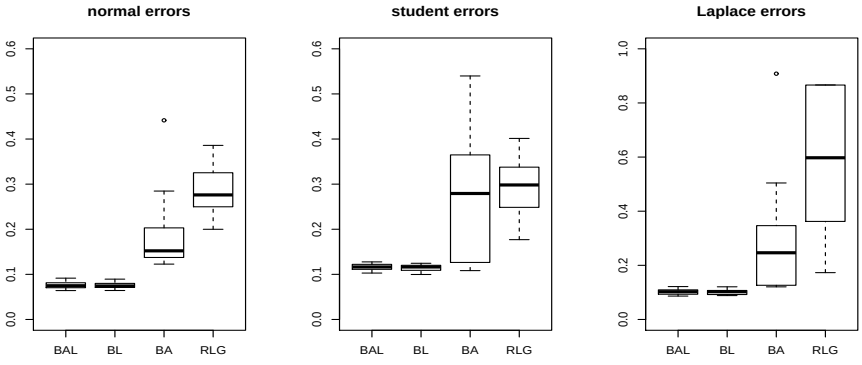
داده‌های واقعی شامل رشد ۵۰ کودک در شهرستان زنجان، از بدو تولد تا دوسالگی است که از درمانگاه‌های شهرستان به تصادف انتخاب شده‌اند. هدف بررسی تأثیر



شکل ۳: عملکرد روش‌ها به لحاظ معیار اریبی در چندک ۰/۵



شکل ۴: عملکرد روش‌ها به لحاظ معیار RMSE در چندک ۰/۵



شکل ۵: عملکرد روش‌ها به لحاظ معیار MAE در چندک ۰/۵

برخی عوامل بر رشد کودک در طول زمان می‌باشد. این عوامل عبارت‌اند از قد، وزن و دور سر کودک که در فواصل زمانی بدو تولد، ۲، ۴، ۶، ۹، ۱۲ ، ۱۸ و ۲۴ ماهگی کودکان اندازه‌گیری شده است. متغیر پاسخ نیز به صورت صفر و یک در همین فواصل زمانی است که عدد یک نشان دهندهٔ رشد مناسب و عدد صفر نشان‌دهندهٔ رشد نامناسب کودک است. برای برآورد پارامترها با استفاده از روش نمونه‌گیری گیبز، ۸۰۰۰ نمونه از توزیع های پسینی شرطی کامل پارامترها تولید کرده و ۲۰۰۰ نمونهٔ اول را به عنوان دورهٔ همگرایی مطلوب کنار می‌گذاریم. در جدول ۲ برآورد ضرایب رگرسیونی برای سه چندک $\tau = (۰/۲۵, ۰/۵, ۰/۷۵)$ برای سه روش بیزی محاسبه شده است. همچنین با مدل برازش شده، متغیرهای پاسخ y_{ij}ها برای کلیه داده‌ها که تعداد آن‌ها برابر ۴۰۰ مشاهده است، برآورد شده و تعداد پاسخ‌های برآورد شده که با پاسخ‌های واقعی مطابقت دارد در سطرهای آخر جدول۲ (سطر مربوط به $\hat{y}_{True}$) ارائه شده است. در جدول۳ نیز برآورد پارامترها و تعداد پاسخ‌های برآورد شده که با پاسخ‌های واقعی مطابقت دارد با روش RLG آمده است. با توجه به این دو جدول ملاحظه می‌شود که هر دو روش ارائه شده در چندک‌های پایین بخصوص برای چندک ۰/۵ از دو مدل دیگر عملکرد بهتری داشته‌اند. ولی برای چندک بالا ($\tau = ۰/۷۵$) مدل غیر بیزی RLG عملکرد بهتری نسبت به مدل‌های بیزی دارد.

بحث و نتیجه‌گیری

در مطالعات طولی علاوه بر اثرات ثابت که تأثیر هر یک از متغیرهای توصیفی را بر متغیر پاسخ بیان می‌کند، اثرات تصادفی نیز در مدل لحاظ می‌شود. این اثرات تغییرات بین متغیرهای پاسخ (تغییرات بین گروهی) را کنترل می‌کند. در مطالعاتی که تعداد مشاهدات از متغیر پاسخ زیاد است، لذا اثرات تصادفی نیز زیاد شده و پارامترهای زیادی وارد مدل شده و بنابراین دقت مدل کم و تفسیر آن مشکل می‌شود. در این مقاله با ایجاد تاوان لاسو و لاسوی تطبیق‌پذیر روی اثرات تصادفی آن‌ها را به سمت صفر منقبض کرده و اثرات کم اهمیت از مدل حذف شدند. هر دو

جدول ۲: برآورد ضرایب رگرسیونی در چندک‌های متفاوت

مدل	۰/۲۵	۰/۵	۰/۷۵
BAL			
β_1	۰/۰۱۶۷	۰/۰۰۹۶	۰/۰۲۱۴
β_2	۰/۱۲۸۹	۰/۰۸۲۰	۰/۱۵۳۴
β_3	۰/۰۰۲۳۱	۰/۰۱۶۴	۰/۰۰۵۷
$\hat{y}_{True}$	۳۰۰	۳۵۰	۲۵۰
BL			
β_1	۰/۰۱۵۸	۰/۰۰۹۰	۰/۰۰۰۳۵
β_2	۰/۱۲۲۲	۰/۰۸۲۶	۰/۰۵۰۵
β_3	۰/۰۰۳۶۱	۰/۰۱۶۷	۰/۰۲۷۷
$\hat{y}_{True}$	۳۰۰	۳۵۰	۲۵۰
BA			
β_1	۰/۱۵۰۲	۰/۰۱۲۸	۰/۰۱۴۹
β_2	۰/۲۱۰۰	۰/۰۲۴۱	۰/۹۰۶۲
β_3	۰/۲۵۴۱۱	۰/۰۳۱۰۷	۰/۰۲۲۶
$\hat{y}_{True}$	۲۴۹	۲۸۵	۲۷۵

جدول ۳: برآورد ضرایب رگرسیونی با استفاده از مدل رگرسیون لوژستیک

مدل	β_1	β_2	β_3	$\hat{y}_{True}$
RGL	۰/۰۲۳۶	۰/۰۴۹۹	۰/۰۴۹۵	۲۸۵

نوع تاوان با تعریف توزیع‌های پیشینی لاپلاس متقارن که تابع چگالی آن در نقطه صفر نوک‌دار است، در مدل لحاظ شده و هر دو مدل از دیدگاه آمار بیزی مورد بررسی قرار گرفتند. نتایج حاصل از شبیه‌سازی و تحلیل داده‌های واقعی نشان داد که روش‌های ارائه شده در چندک ۰/۵ نسبت به روش بیزی (بنویت و همکاران، ۲۰۱۳) که اثرات تصادفی و تاوان روی آن‌ها را در مدل لحاظ نمی‌کند و همچنین روش بسامدی که اثرات تصادفی را در مدل لحاظ می‌کند، ولی هیچ گونه تاوانی روی آن‌ها قرار نمی‌دهد، در برآورد ضرایب رگرسیونی عملکرد بهتری دارد. نتایج حاصل از تحلیل داده‌های واقعی نشان داد که روش‌های ارائه شده در چندک‌های ۰/۲۵ و ۰/۵ دقت بالایی نسبت به دو روش دیگر دارند، اما در چندک بالا ($\tau = ۰/۷۵$) رهیافت بسامدی تا حدودی از روش‌های بیزی عملکرد بهتری

دارد. اما از آنجایی‌که مدل‌های ارائه شده بر اساس روش رگرسیون چندکی است، لذا در مقایسه با مدل بسامدی که بر اساس روش رگرسیون میانگین است، نسبت به داده‌های پرت و توزیع خطاها استوار است. بنابراین استفاده از مدل‌های ارائه شده در تحلیل داده‌های واقعی توصیه می‌شود.

تقدیر و تشکر

نویسندگان مقاله از داوران و سردبیر محترم مجله به خاطر ارایه پیشنهادات ارزنده که سبب بهبود مقاله گردید، کمال تشکر و قدردانی را دارند.

مراجع

غلامی فشارکی، م.، کاظم نژاد، ا. و زایری، ف. (۱۳۹۲)، تحلیل دو سطحی با اثرات تصادفی چوله نرمال و مدل‌بندی داده های طولی، مجله علوم آماری، ۷، ۲۴۸-۲۳۳.

Alhamzawi, R., Yu, K. and Benoit, D. F. (2012), Bayesian Adaptive Lasso Quantile Regression, *Statistical Modeling*, **12**, 279-297.

Benoit, D. F., Alhamzawi, R. and Keming, Y. U. (2013), Bayesian Lasso Binary Quantile Regression, *Computational Statistics*, **28**, 2861-2873.

Fitzmaurice, G. M. and Laird, N. M. (1993), A Likelihood-Based Method for Analyzing Longitudinal Binary Respones, *Biometrika*, **80**, 141-151.

Gelman, A., Carlin, J. B., Stern, H. S. and Rubin, D. B. (1995), *Bayesian Data Analysis*, Chapman and Hall, London.

Goncalves, M. H. (2002), *Likelihood Methods for Discrete Longitudinal Data*, Ph.D. Thesis, University of Lisbon.

Goncalves, M. H. and Azzalini, A. (2008), Using Markov Chains for Marginal Modeling of Binary Longitudinal Data in Exact Likelihood Approach, *Metron*, **LXVI**, 157-181.

Goncalves, M. H., Cabral, M. S. and Azzalini, A. (2012), The R Package Bild for the Analysis of Binary Longitudinal Data, *Journal of Statistical Software*, **46**, 1-17.

Koenker, R. (2004), Qauntile Regression for Longitudinal Data, *Journal of Multivariate Analysis*, **91**, 74-89.

Koenker, R. and Machado, J. (1999), Goodness of Fit and Related Inference Processes for Quantile Regression, *Journal of the American Statistical Assosation*, **94**, 1296-1310.

Kozumi, H. and Kabayashi, G. (2011), Gibbs Sampling Methods for Bayesian Qauntile Regression, *Journal of Statistical Computation and Simulation*, **81**, 1565-1578.

Luethi, D. and Breyman, W. (2014), A Pakage on the Generalized Hyperbolice Distribution and Its Special Casses, R Pakage Version 1.5.6, URL http://www.r-project.org

Park, T. and Casella, T. (2008), The Bayesian Lasso, *Journal of the American Statistical Assosation*, **103**, 681-686.

Tibshirani, R. (1996), Regression Shrinkage and Selection Via the Lasso, *Journal of the Royal Statistical Society*, B, **58**, 267-288.

Zou, H. (2006), The Adaptive Lasso and Its Oracle Properties, *Journal of the American Statistical Assosation*, **101**, 1418-1429.