مجله علوم آماری، بهار و تابستان ۱۳۹۸
جلد ۱۳، شماره ۱، ص ۹۹ – ۱۱۶
DOI: 10.29252/jss.13.1.99

مدل پواسن دو متغیره با صفر آماسیده با اثر تصادفی چوله‌نرمال و کاربرد آن

مژگان دهقانی، محمدرضا زادکرمی، محمدرضا آخوند

گروه آمار، دانشگاه شهید چمران اهواز

تاریخ دریافت: ۱۳۹۶/۰۲/۱۵ تاریخ پذیرش: ۱۳۹۷/۰۷/۲۳

چکیده: در سال‌های اخیر از رگرسیون پواسن برای مدل‌بندی متغیرهای پاسخ شمارشی استفاده شده است. زمانی که مجموعه داده‌های شمارشی دارای فراوانی بیش از حد در عدد صفر باشند، استفاده از رگرسیون پواسن مناسب نیست. در این مقاله، از دو مدل رگرسیون پواسن با صفر آماسیده و رگرسیون پواسن دو متغیره با صفر آماسیده با اثر تصادفی برای مدل‌بندی پاسخ شمارشی با فراوانی بیش از حد در عدد صفر استفاده شده است. معمولا توزیع اثر تصادفی را نرمال فرض می‌کنند، اما در این مقاله از توزیع چوله نرمال که از انعطاف‌پذیری بیشتری نسبت به توزیع نرمال برخوردار است به‌عنوان توزیع اثر تصادفی استفاده کرده‌ایم. در پایان مدل بدست آمده برای تحلیل داده‌های مربوط به تعداد واحدهای مردودی و نیم‌سال‌های مشروطی دانشجویان دانشگاه شهید چمران اهواز مورد استفاده قرارگرفته است و از روش شبیه‌سازی برای اعتبارسنجی مدل بدست آمده استفاده شده است.

واژه‌های کلیدی: پواسن با صفر آماسیده، اثر تصادفی، چوله نرمال، چوله نرمال دو متغیره.

۱ مقدمه

معمولاً برای بررسی رابطه بین یک متغیر پاسخ شمارشی و متغیرهای توضیحی از مدل رگرسیون پواسن استفاده می‌شود. داده‌های واقعی ممکن است دارای پراکندگی زیاد باشند. یکی از مواردی که باعث پراکندگی در داده‌های شمارشی و عدم برازش مطلوب مدل رگرسیون پواسن به دادها می‌شود وجود صفر آماسیده است. در این حالت می‌توان از مدل‌های رگرسیونی با صفر آماسیده استفاده کرد. لامبرت (۱۹۹۲) رگرسیون پواسن

آدرس الکترونیکی نویسنده مسئول مقاله: محمدرضا زادکرمی، zadkarami@yahoo.co.uk
کد موضوع‌بندی ریاضی (۲۰۱۰): 62E20،

با صفر آماسیده را که امکان مدل‌بندی فراوانی تعداد صفر را فراهم می‌کند پیشنهاد کرد. رگرسیون پواسن با صفر آماسیده در اغلب رشته‌ها به خصوص علوم اجتماعی و اقتصادی، علوم پزشکی و غیره کاربرد فراوانی دارد. در سال‌های اخیر از رگرسیون پواسن با صفر آماسیده برای مدل‌سازی داده‌ها زمانی که متغیرهای پاسخ گسسته و دارای فراوانی زیاد در عدد صفر هستند، استفاده شده است. (لی و همکاران، ۲۰۰۲؛ کریوک و همکاران، ۲۰۰۳؛ زان نوته و فلیشیتی، ۲۰۱۱؛ وانگ و لم، ۲۰۱۳)، همچنین وانگ و همکاران (۲۰۰۳)، لائو و همکاران (۲۰۱۱) و دانگ و همکاران (۲۰۱۴) رگرسیون پواسن دو متغیره با صفر آماسیده را برای مدل‌بندی داده‌های دو متغیره گسسته با فراوانی زیاد در عدد صفر مورد استفاده قرار دادند.

برای کنترل اثر عوامل ناشناخته و غیر قابل اندازه‌گیری در مدل‌های آماری می‌توان از اثر تصادفی استفاده کرد. مدل‌بندی رگرسیون پواسن با صفر آماسیده و اثر تصادفی توسط یو و لی (۲۰۰۱) و فاکس (۲۰۱۳) مورد مطالعه قرار گرفته است. معمولا توزیع اثر تصادفی را نرمال در نظر می‌گیرند اما فرضیه نرمال بودن اثرات تصادفی همواره منجر به نتایج استوار نمی‌شود (وربک و لسافر، ۱۹۹۶؛ ساهو و همکاران، ۲۰۰۳)، در سال‌های اخیر توزیع‌های دیگری که نسبت به توزیع نرمال از انعطاف‌پذیری بیشتری برخوردار هستند، به‌عنوان توزیع اثرات تصادفی پیشنهاد شده‌اند مانند: توزیع ناپارامتری (آیتکن، ۱۹۹۹ و بور و همکاران، ۲۰۰۳)، توزیع t (مارشال و اشپیگل هارتر، ۱۹۹۸ و کانگدان، ۲۰۰۶)، توزیع گاما (ترنر و همکاران، ۲۰۰۱) و توزیع چوله نرمال (رستگاران و زادکرمی، ۲۰۱۵)،

مطالعه در مورد عوامل عدم موفقیت تحصیلی دانشجویان در حین تحصیل در دانشگاه‌ها یکی از اولویت‌های پژوهشی در آموزش عالی است. بررسی تعداد واحدهای مردودی و نیم‌سال‌های مشروطی دانشجویان از جمله روش‌هایی است که می‌توان موفقیت تحصیلی دانشجویان را ارزیابی کرد. اینکه چه عواملی زمینه‌ساز یا شتاب دهنده افت تحصیلی دانشجویان می‌باشند در پژوهش‌های متعددی مورد بررسی قرار گرفته است.

رودباری و همکاران (۱۳۸۹)، جنس، مقطع تحصیلی و دانشکده را از جمله عوامل مؤثر بر پیشرفت تحصیلی دانشجویان دانشگاه علوم پزشکی تهران دانسته‌اند. رودباری و صالحی (۱۳۹۳) عوامل مؤثر بر تعداد دروس مردودی و نیم‌سال‌های مشروطی دانشجویان دانشگاه علوم پزشکی ایران را ارزیابی کردند که معدل و سهمیه ورودی دانشجویان عوامل تأثیرگذار بر تعداد نیم‌سال‌های مشروطی می‌باشند در صورتی که معدل و مقطع تحصیلی دانشجویان را به عنوان عوامل تأثیرگذار بر تعداد دروس مردودی معرفی کرده‌اند. حبیب‌زاده و همکاران (۱۳۹۰) معدل دیپلم، تعداد دیدار با خانواده در ماه، میزان مطالعه به ساعت در روز، داشتن اطلاع از آینده شغلی، فقدان انگیزه، عدم مطالعه منظم و غیربومی بودن را از عوامل مؤثر

بر میزان افت تحصیلی دانشجویان شناسایی کردند. تقربی و همکاران (۱۳۸۸) عوامل مرتبط با وقوع مشروطی در دانش‌آموختگان دانشکده پرستاری و مامایی کاشان را بررسی نمودند که عوامل سهمیه غیر مناطق و جنس مذکر بیشترین خطر وقوع مشروطی را نشان می‌دهند. همچنین شمس و همکاران (۱۹۹۷) نشان دادند که جنسیت در تعداد دروس مردودی یا نیم‌سال‌های مشروطی دانشجویان تأثیرگذار است.

در این مقاله با استفاده از مدل‌های پواسن با صفر آماسیده و اثر تصادفی و پواسن دو متغیره با صفر آماسیده و اثر تصادفی به مدل‌بندی تعداد واحدهای مردودی و نیم‌سال‌های مشروطی دانشجویان دانشگاه شهید چمران اهواز می‌پردازیم. همچنین علاوه بر استفاده از توزیع نرمال به‌عنوان توزیع اثر تصادفی که در مطالعه یو و لی (۲۰۰۱) و فاکس (۲۰۱۳) مورد استفاده قرار گرفته است، توزیع چوله نرمال نیز به‌عنوان توزیع اثر تصادفی مورد استفاده قرار خواهد گرفت. بدین منظور در بخش ۲ ابتدا به بررسی مدل پواسن با صفر آماسیده با اثر تصادفی توزیع چوله نرمال می‌پردازیم و تابع درستنمایی داده‌ها معرفی می‌شود. سپس به بررسی مدل پواسن دو متغیره با صفر آماسیده و توزیع چوله نرمال دو متغیره به‌عنوان توزیع اثر تصادفی پرداخته می‌شود. در بخش ۳ روش پیشنهاد شده برای تحلیل داده‌های تعداد واحدهای مردودی و نیم‌سال‌های مشروطی دانشجویان ورودی سال ۸۸–۸۹ دانشگاه شهید چمران اهواز مورد استفاده قرار می‌گیرد. همچنین در بخش ۴ یک مطالعه برای اعتبارسنجی مدل پیشنهادی ارائه می‌شود. در آخر به بحث در مورد نتایج بدست آمده و نتیجه‌گیری درباره مدل پرداخته خواهد شد.

۲ مدل پواسن با صفر آماسیده با اثر تصادفی

فرض کنید متغیر تصادفی y_i با احتمال p_i مقدار صفر آماسیده را انتخاب می‌کند و با احتمال $۱ - p_i$ مقدار آن از توزیع پواسن با پارامتر μ_i بدست می‌آید، یعنی

$$y_i \sim \begin{cases} ۰ & p_i \\ Poisson(\mu_i) & (۱ - p_i), \end{cases}$$

بنابراین مقدار صفر متغیر تصادفی y_i از دو منبع صفر آماسیده و صفر توزیع پواسن بدست می‌آید. بنابراین رابطه

$$P(Y_i = y_i) = \begin{cases} p_i + (1 - p_i)\exp(-\mu_i) & y_i = 0 \\ (1 - p_i)\frac{\exp(-\mu_i){\mu_i}^{y_i}}{y_i!} & y_i > 0, \end{cases} \quad (1)$$

برقرار است، که در آن $i = 1, \ldots, n_j$، $j = 1, \ldots, J$ و $0 < p_i < 1$، احتمال صفر آماسیده است. میانگین و واریانس این توزیع عبارتست از:

$$E(y_i) = \mu_i(1 - p_i) \quad \mathrm{Var}(y_i) = E(y_i)(1 + \mu_i p_i),$$

دو مدل رگرسیونی توزیع پواسن و مدل صفر آماسیده وجود دارد که به‌ترتیب تابع‌های پیوند آن‌ها به‌صورت

$$\ln(\mu_i) = \sum_{\ell=0}^{m-1} x_{i\ell}\beta_\ell + u_i, \qquad \mathrm{logit}(p_i) = \ln(p_i/(1 - p_i)) = \sum_{\ell=0}^{m-1} z_{i\ell}\gamma_\ell,$$

تعریف می‌شوند، که در آن $\beta' = (\beta_0, \ldots, \beta_{m-1})$ و $\gamma' = (\gamma_0, \ldots, \gamma_{m-1})$ بردارهای $(m \times 1)$ از ضرایب متغیرهای توضیحی و $U = (u_1, \ldots, u_J)$ اثر تصادفی را نشان می‌دهد که فرض می‌شود دارای توزیع نرمال است. بنابراین برای مدل پواسن از تابع پیوند لگاریتم طبیعی و برای مدل صفر آماسیده از تابع پیوند طبیعی لوژستیک استفاده شده است.

توزیع چوله نرمال به‌عنوان تعمیمی از توزیع نرمال در موقعیت‌هایی کاربرد دارد که داده‌ها دارای مقداری چولگی می‌باشند. فرض کنید $\phi(u^*)$ و $\Phi(u^*)$ به ترتیب توابع چگالی و توزیع نرمال استاندارد را نشان می‌دهند. متغیر تصادفی U^* دارای توزیع چوله نرمال است $U^* \sim SN(\lambda)$ اگر دارای تابع چگالی

$$f(u^*) = 2\phi(u^*)\Phi(\lambda u^*)$$

باشد، که در آن λ پارامتر چولگی است. این فرم تابع چگالی چوله نرمال اولین بار توسط آزالینی (۱۹۸۵) معرفی شد.

با اضافه کردن پارامتر مقیاس و پارامتر مکان، تابع چگالی چوله نرمال به شکل

$$f(u^*) = \frac{۲}{\sigma}\phi(\frac{u^* - \eta}{\sigma})\Phi(\lambda(\frac{u^* - \eta}{\sigma})),$$

و با نماد $SN \sim (\eta, \sigma^۲, \lambda)$ تغییر می‌کند. تابع درستنمایی مدل رگرسیون پواسن با صفر آماسیده با اثر تصادفی به‌صورت

$$f(y_j|\beta, \gamma, \sigma, \lambda) = \prod_{i=۱}^{n_j} \int f(y_{ij}|\beta, \gamma, u_j) g(u_j|\sigma, \lambda) du_j,$$

نوشته می‌شود، با استفاده از رابطه (۱) داریم:

$$f(y_{ij}|\beta, \gamma, u_j) = [p_{ij} + (۱ - p_{ij}) f(۰, \mu_{ij})]^{I\{y_{ij}=۰\}} [(۱ - p_{ij}) f(y_{ij}, \mu_{ij})]^{I\{y_{ij}>۰\}},$$

وقتی‌که $f(y_{ij}, \mu_{ij})$ تابع چگالی پواسن با پارامتر μ_{ij} و $g(u_j|\sigma, \lambda)$ تابع چگالی اثر تصادفی است.

۱.۲ مدل پواسن دو متغیره با صفر آماسیده با اثر تصادفی

مدل رگرسیون پواسن دو متغیره با صفر آماسیده برای متغیر تصادفی دو متغیره $Y_i = (Y_{i۱}, Y_{i۲})$ به‌صورت

$$Pr(Y_{ik} = y_{ik}) = \begin{cases} p_{ik} + (۱ - p_{ik}) \exp(-\mu_{ik}) & y_{ik} = ۰, \\ (۱ - p_{ik}) \frac{\exp(-\mu_{ik})\mu_{ik}^{y_{ik}}}{y_{ik}!} & y_{ik} > ۰, \end{cases}$$

تعریف می‌شود، که $k = ۱, ۲$، $i = ۱, \ldots, n_j$، $j = ۱, .., J$ و $۰ < p_{ik} < ۱$، احتمال صفر آماسیده برای مؤلفه kام است. تابع پیوند را برای مدل رگرسیونی به‌صورت

$$\ln(\mu_{i۱}) = \sum_{\ell=۰}^{m-۱} x_{i\ell}\beta_\ell + v_i, \qquad \ln(\mu_{i۲}) = \sum_{\ell=۰}^{m-۱} a_{i\ell}\alpha_\ell + w_i$$

$$\ln(\frac{p_{i۱}}{۱ - p_{i۱}}) = \sum_{\ell=۰}^{m-۱} z_{i\ell}\gamma_\ell, \quad \ln(\frac{p_{i۲}}{۱ - p_{i۲}}) = \sum_{\ell=۰}^{m-۱} s_{i\ell}\delta_\ell,$$

تعریف می‌شود.

وقتی $\beta' = (\beta_0, \dots, \beta_{m-1})$ ، $\alpha' = (\alpha_0, \dots, \alpha_{m-1})$، $\gamma' = (\gamma_0, \dots, \gamma_{m-1})$ و $\delta' = (\delta_0, \dots, \delta_{m-1})$ بردارهای $(m \times 1)$ ضرایب متغیرهای توضیحی و $V = (v_1, \dots, v_J)$ و $W = (w_1, \dots, w_J)$ اثر تصادفی برای مؤلفه‌های متغیر تصادفی دو متغیره Y_i هستند. آزالینی (۲۰۰۵) توزیع دو متغیره چوله نرمال با تابع چگالی

$$f(v^*, \eta, \Sigma, \lambda) = 2\phi_2(v^* - \eta, \Sigma)\Phi_2(\Lambda^T \omega^{-1}(v^* - \eta)),$$

را معرفی کرد، که در آن ϕ تابع چگالی نرمال دو متغیره و Φ تابع توزیع نرمال استاندارد، η پارامتر مقیاس، Σ پارامتر مکانی (ماتریس کوواریانس)، Λ بردار چولگی و ω ماتریس قطری با درایه‌های مثبت است. اگر بردار پارامتر چولگی در تابع چگالی دو متغیره چوله نرمال برابر با بردار صفر در نظر گرفته شود، تابع چگالی دو متغیره نرمال بدست می‌آید. بنابراین اگر برآورد پارامتر چولگی بیانگر صفر بودن پارامتر چولگی باشد، می‌توان توزیع اثر تصادفی را نرمال در نظر گرفت. فرض کنید مؤلفه‌های y_i به شرط اثر تصادفی از یکدیگر مستقل و رابطه

$$f(y_{i1}, y_{i2} | \beta, \alpha, \gamma, \delta, \Sigma, \Lambda) = f(y_{i1} | \beta, \gamma, v_i) f(y_{i2} | \alpha, \delta, w_i)$$

برقرار باشد. بنابراین تابع درستنمایی مدل رگرسیونی پواسن دو متغیره با صفر آماسیده به‌صورت

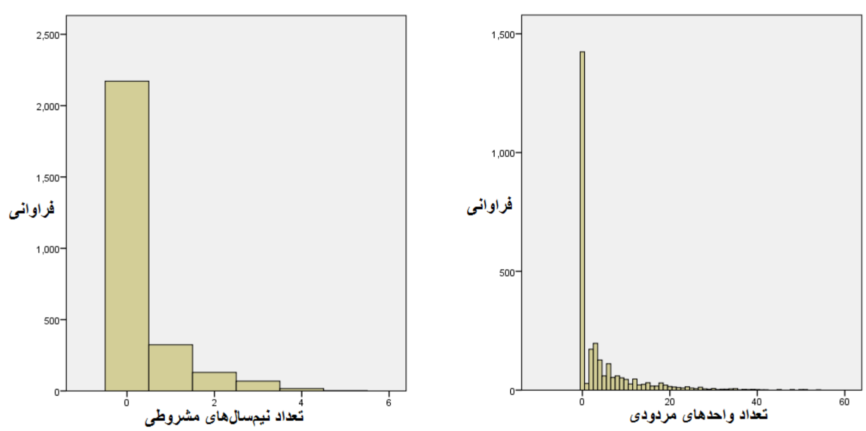
$$f(y_{j1}, y_{j2}, \beta, \alpha, \gamma, \delta, \Sigma, \Lambda) = \prod_{i=1}^{n_j} \int\int f(y_{ij1}|\beta, \gamma, v_j) f(y_{ij2}|\alpha, \delta, w_j) g(v_j, w_j|\Sigma, \Lambda) dv_j dw_j,$$

خواهد بود، که در آن $y_{jk} = (y_{1jk}, \dots, y_{n_j jk})$ ، $k = 1, 2$ است. اگر رابطه $g(v_j, w_j|\Sigma, \Lambda) = g_1(v_j|\sigma_1^2, \lambda_1) g_2(w_j|\sigma_2^2, \lambda_2)$ برقرار باشد در این‌صورت مدل اثر تصادفی را مستقل گوییم.

۳ مثال کاربردی

در این مقاله از داده‌های تعداد واحدهای مردودی و نیم‌سال‌های مشروطی دانشجویان ورودی سال ۸۸–۸۹ دانشگاه شهید چمران اهواز استفاده شده است. جامعه آماری مورد مطالعه شامل دانشجویان شاغل به

تحصیل ورودی سال ۸۸–۸۹ دانشگاه شهید چمران اهواز است. اطلاعات مربوط به ۲۶۹۰ نفر دانشجو انتخاب شده است. با توجه به اطلاعات موجود، تأثیر چهار عامل بر تعداد واحدهای مردودی و نیم‌سال‌های مشروطی مورد بررسی قرار گرفته است. داده‌های جمعیت‌شناسی و آموزشی از طریق حوزه آموزشی دانشگاه شهید چمران اهواز جمع‌آوری شده است. این اطلاعات شامل دانشکده، رشته تحصیلی، مقطع تحصیلی، جنسیت، سهمیه، بومی بودن، تعداد واحدهای مردودی و تعداد نیم‌سال‌های مشروطی بود. با توجه به شکل ۱ مشاهده می‌شود که داده‌ها دارای فراوانی بیش از حد در عدد صفر هستند. از آنجا که تعداد واحدهای مردودی و نیم‌سال‌های مشروطی در بسیاری از دانشجویان مقدار صفر دارند، استفاده از رگرسیون پواسن با صفر آماسیده مناسب است. همچنین به دلیل همبستگی بین دانشجویان یک رشته تحصیلی به دلیل دارا بودن واحدهای درسی یکسان، متغیر رشته تحصیلی به‌عنوان اثر تصادفی به مدل اضافه می‌شود. داده‌ها به کمک نرم‌افزار SAS نسخه ۹.۲، بسته $Nlmixed\ Proc$ و برنامه‌های نوشته شده توسط نویسندگان مورد تحلیل قرار گرفته‌اند. در ابتدا برازش مدل پواسن با صفر آماسیده به تعداد واحدهای مردودی و تعداد نیم‌سال‌های مشروطی بررسی می‌شود و در ادامه به بررسی مدل پواسن با صفر آماسیده، بطور توام برای تعداد واحدهای مردودی و نیم‌سال‌های مشروطی پرداخته می‌شود.



شکل ۱: نمودار بافت‌نگار تعداد واحدهای مردودی (سمت راست) و نیم‌سال‌های مشروطی (سمت چپ)،

۱.۳ مقایسه مدل‌ها

جدول ۱ مقدار معیارهای نیکویی برازش مدل‌ها بر پایه پواسن با صفر آماسیده را نشان می‌دهد. مقدار معیار آکائیک (AIC) و شوارتز (BIC) توسط نرم‌افزار SAS محاسبه شده‌اند. این معیارها به‌ترتیب از روابط $AIC = -۲\ln(L) = ۲k$ و $BIC = -۲\ln(L) + k * \ln(n)$ بدست می‌آیند، که k تعداد پارامترهای مدل، n تعداد مشاهدات و L تابع ماکسیمم درستنمایی مدل برآورد شده است. مدل‌های دارای مقادیر کوچک‌تر معیارهای آکائیک و شوارتز، مدل‌های برازنده‌تری هستند. همان‌طور که در جدول ۱ مشاهده می‌شود براساس معیار آکائیک، مدل پواسن با صفر آماسیده همراه با اثر تصادفی چوله نرمال بهترین مدل برای برازش به داده‌های تعداد واحدهای مردودی و نیم‌سال‌های مشروطی می‌باشد. با توجه به اینکه مدل پواسن با صفر آماسیده با اثر تصادفی چوله نرمال دارای کمترین میزان معیار آکائیک بود، در ادامه برآورد پارامترهای این مدل برای داده‌های واحدهای مردودی و نیم‌سال‌های مشروطی ارائه می‌شود.

جدول ۱: معیارهای نیکویی برازش مدل‌ها بر پایه پواسن با صفر آماسیده

	مدل	$-\ell(\hat{\theta})$	AIC	BIC
	بدون اثر تصادفی	۱۵۳۵۷	۱۵۳۸۹	۱۵۴۸۳
واحدهای مردودی	با اثر تصادفی نرمال	۱۴۷۱۰	۱۴۷۴۴	۱۴۷۴۸
	با اثر تصادفی چوله نرمال	۱۰۹۵۵	۱۱۰۳۱	۱۱۰۳۵
	بدون اثر تصادفی	۳۶۵۷/۱	۳۶۸۹/۱	۳۷۸۳/۴
نیم‌سال‌های مشروطی	با اثر تصادفی نرمال	۳۶۱۰/۷	۳۶۴۴/۷	۱/۳۶۴۸
	با اثر تصادفی چوله نرمال	−۵۰/۹	−۱۴/۹	−۱۱/۳

با توجه به جدول ۲، مقطع تحصیلی کارشناسی‌ارشد نسبت به کارشناسی و سهمیه منطقه یک، دو و سه نسبت به سایر سهمیه‌ها، لگاریتم پارامتر مکان (μ_i) تعداد واحدهای مردودی را کاهش می‌دهند. اما جنسیت مرد نسبت به زن لگاریتم پارامتر مکان (μ_i) تعداد واحدهای مردودی را افزایش می‌دهد. همچنین با توجه به جدول ۲ مقاطع تحصیلی کارشناسی‌ارشد و دکتری نسبت به کارشناسی، غیربومی بودن دانشجویان نسبت به بومی بودن آن‌ها و سهمیه منطقه دو نسبت به سایر سهمیه‌ها، لگاریتم بخت صفر آماسیده را افزایش می‌دهند. اما جنسیت مرد نسبت به زن لگاریتم بخت صفر آماسیده را کاهش می‌دهد. در نظر گرفتن اثر تصادفی در مدل علاوه بر تغییر در کمیت پارامترهای برآورد شده، معنی‌داری آن‌ها را نیز تحت تأثیر قرار داده است، به‌طوری‌که در مدل پواسن با صفر آماسیده بدون اثر تصادفی برخلاف مدل پواسن با صفر آماسیده با اثر تصادفی نرمال و چوله نرمال، مقطع دکتری در مدل‌بندی تعداد واحدهای مردودی

و سهمیه منطقه سه در مدل‌بندی صفر آماسیده معنی‌دار شده‌اند. همچنین در مدل پواسن با صفر آماسیده با اثر تصادفی نرمال برخلاف مدل پواسن با صفر آماسیده بدون اثر تصادفی و با اثر تصادفی چوله نرمال، سهمیه منطقه دو در مدل‌بندی صفر آماسیده معنی‌دار نشده است. برآورد پارامتر چولگی $\hat{\lambda} = ۱۱/۶۵۷۶$ نشانگر این است که توزیع چوله نرمال برای اثر تصادفی نسبت به توزیع نرمال مناسب‌تر است.

جدول ۲: برآورد پارامترهای مدل پواسن با صفر آماسیده همراه با اثر تصادفی چوله نرمال (تعداد واحدهای مردودی)،

مدل	سطح مرجع	متغیرها	برآورد	انحراف معیار	مقدار احتمال
	–	مقدار ثابت	۱/۷۱۲۱	۰/۱۱۱۲	$<$ ۰/۰۰۰۱
	کارشناسی	کارشناسی‌ارشد	−۰/۹۱۵۵	۰/۰۵۴۵	$<$ ۰/۰۰۰۱
		دکتری	۰/۱۵۷۶	۰/۱۲۴۳	۰/۲۴۰۵
پواسن	زن	مرد	۰/۱۹۳۶	۰/۰۲۰۳	$<$ ۰/۰۰۰۱
	بومی	غیربومی	−۰/۰۲۰۵	۰/۰۲۳۶	۰/۴۰۹۱
		منطقه یک	−۰/۲۴۶۳	۰/۰۵۵۴	۰/۰۰۲۲
	سایر سهمیه‌ها	منطقه دو	−۰/۱۶۸۸	۰/۰۳۹۱	۰/۰۰۲۶
		منطقه سه	−۰/۳۱۱۴	۰/۰۳۹۸	$<$ ۰/۰۰۰۱
	–	مقدار ثابت	−۰/۸۱۷۸	۰/۱۸۳۸	۰/۰۰۲۱
	کارشناسی	کارشناسی‌ارشد	۲/۱۵۰۱	۰/۱۳۵۷	$<$ ۰/۰۰۰۱
		دکتری	۱/۴۱۵۱	۰/۳۱۷۵	۰/۰۰۲۱
صفر آماسیده	زن	مرد	−۰/۷۳۰۶	۰/۰۹۷۹	$<$ ۰/۰۰۰۱
	بومی	غیربومی	۰/۴۵۹۳	۰/۱۰۶۰	۰/۰۰۲۵
		منطقه یک	۰/۴۶۰۳	۰/۲۴۷۲	۰/۰۹۹۶
	سایر سهمیه‌ها	منطقه دو	۰/۴۴۵۶	۰/۱۸۸۱	۰/۰۴۵۳
		منطقه سه	۰/۴۴۴۲	۰/۱۸۸۵	۰/۰۸۵۲
اثر تصادفی		λ	۱۱/۶۵۷۶	–	–
		$\sigma^۲$	۰/۲۷۴۸	–	–

با توجه به جدول ۳، مقطع تحصیلی کارشناسی‌ارشد نسبت به کارشناسی، لگاریتم پارامتر مکان تعداد نیم‌سال‌های مشروطی را کاهش می‌دهد. همچنین غیربومی بودن دانشجویان نسبت به بومی بودن آن‌ها لگاریتم بخت صفر آماسیده را افزایش می‌دهد. اما جنسیت مرد نسبت به زن باعث کاهش لگاریتم بخت صفر آماسیده می‌شود. لازم به‌ذکر است که در مدل پواسن با صفر آماسیده با اثر تصادفی چوله نرمال متغیر بومی بودن در قسمت مدل‌بندی تعداد نیم‌سال‌های مشروطی معنی‌دار نیست. این در حالی است که در مدل پواسن با صفر آماسیده بدون اثر تصادفی و با اثر تصادفی نرمال این متغیر معنی‌دار است. برآورد پارامتر چولگی $\hat{\lambda} = ۲۶/۷۷۱۴$ نشانگر این است که توزیع چوله نرمال برای اثر تصادفی نسبت به توزیع نرمال مناسب‌تر است. تعداد نیم‌سال‌های مشروطی نشان دهنده ضعف علمی دانشجو و تعداد واحدهای مردودی

جدول ۳: برآورد پارامترهای مدل پواسن با صفر آماسیده همراه با اثر تصادفی چوله نرمال (تعداد نیم‌سال‌های مشروطی)،

مدل	سطح مرجع	متغیرها	برآورد	انحراف معیار	مقدار احتمال
	–	مقدار ثابت	−۰/۰۰۶۱	۰/۰۷۵۱	۰/۹۳۷۱
	کارشناسی	کارشناسی‌ارشد	−۱/۶۴۲۸	۰/۳۲۰۷	۰/۰۰۰۹
		دکتری	−۰/۴۴۳۴	۰/۵۱۰۰	۰/۴۱۰۰
پواسن	زن	مرد	۰/۱۲۳۵	۰/۱۱۲	۰/۲۹۸۹
	بومی	غیربومی	۰/۲۴۰۹	۰/۱۲۶۳	۰/۰۹۲۸
		منطقه یک	−۰/۰۱۹۹	۰/۱۹۵۴	۰/۹۲۱۲
	سایر سهمیه‌ها	منطقه دو	۰/۱۰۰۰	۱/۰۴۲۴	۰/۹۴۵۸
		منطقه سه	−۰/۲۰۰۰	۱/۳۰۵۶	۰/۸۸۲۰
	–	مقدار ثابت	−۰/۶۳۸۳	۰/۱۰۹۴	۰/۰۰۰۴
	کارشناسی	کارشناسی‌ارشد	−۰/۹۰۶۱	۰/۶۱۱۷	۰/۱۷۶۸
		دکتری	۰/۴۰۴۵	۰/۶۵۶۸	۰/۵۵۵۱
صفر آماسیده	زن	مرد	−۰/۶۱۲۰	۰/۱۷۱۷	۰/۰۰۷۴
	بومی	غیربومی	۰/۶۸۱۴	۰/۱۸۵۴	۰/۰۰۶۳
		منطقه یک	−۰/۲۷۲۵	۰/۳۱۲۳	۰/۴۰۸۴
	سایر سهمیه‌ها	منطقه دو	۰/۱۰۰۰	۰/۹۵۲۸	۰/۹۱۹۰
		منطقه سه	−۰/۲۰۰۰	۰/۸۷۳۴	۰/۸۲۴۶
اثر تصادفی		λ	۲۶/۷۷۱۴	–	–
		σ^2	۰/۰۰۰۰۸۳	–	–

میزان ضعف علمی دانشجو را نشان می‌دهد. هرچه تعداد واحدهای مردودی بیشتر باشد نشان می‌دهد که دانشجو در زمینه‌های بیشتری ضعیف می‌باشد. لذا در اینجا به مقایسه مدل‌های توام برازش داده شده به داده‌های تعداد واحدهای مردودی و نیم‌سال‌های مشروطی می‌پردازیم. جدول ۴ مقدار معیارهای نیکویی مدل‌های برازش داده شده بر پایه پواسن دو متغیره با صفر آماسیده را نشان می‌دهد. همان‌طورکه در جدول ۴ مشاهده می‌شود براساس معیار آکائیک مدل پواسن دو متغیره با صفر آماسیده همراه با اثر تصادفی چوله نرمال مستقل بهترین مدل برای برازش به داده‌ها است. با توجه به اینکه مدل پواسن دو متغیره با صفر آماسیده با اثر تصادفی چوله نرمال مستقل دارای کمترین میزان معیار آکائیک بود، در ادامه برآورد پارامترهای این مدل ارائه می‌شود.

جدول ۴: معیارهای نیکویی برازش مدل پواسن دو متغیره با صفر آماسیده با اثرهای تصادفی مختلف.

مدل	$-\ell(\hat{\theta})$	AIC	BIC
بدون اثر تصادفی	۲۰۴۸۹	۲۰۵۵۳	۲۰۷۶۴
با اثر تصادفی نرمال مستقل	۱۸۳۲۱	۱۸۳۹۰	۱۸۴۱۹
با اثر تصادفی چوله نرمال مستقل	۳۵۱۴/۷	۳۵۸۵/۲	۳۶۱۶/۹
با اثر تصادفی چوله نرمال دو متغیره	۱۰۹۸۹	۱۱۰۶۱	۱۱۰۹۳

با توجه به جدول ۵، مقطع کارشناسی‌ارشد نسبت به کارشناسی و سهمیه منطقه یک، دو و سه نسبت به سایر سهمیه‌ها، لگاریتم پارامتر مکان تعداد واحدهای مردودی را کاهش می‌دهند. اما جنسیت مرد نسبت به زن لگاریتم پارامتر مکان تعداد واحدهای مردودی را افزایش می‌دهد. همچنین مقطع کارشناسی‌ارشد نسبت به کارشناسی باعث کاهش لگاریتم پارامتر مکان تعداد نیم‌سال‌های مشروطی می‌شود. همچنین مقطع کارشناسی‌ارشد و دکتری نسبت به کارشناسی، غیربومی بودن دانشجویان نسبت به بومی بودن و سهمیه منطقه یک، دو و سه نسبت به سایر سهمیه‌ها، لگاریتم بخت صفر آماسیده تعداد واحدهای مردودی را افزایش می‌دهند. اما جنسیت مرد نسبت به زن باعث کاهش لگاریتم بخت صفر آماسیده تعداد واحدهای مردودی می‌شود. غیربومی بودن دانشجویان نسبت به بومی بودن آن‌ها باعث افزایش لگاریتم بخت صفر آماسیده تعداد نیم‌سال‌های مشروطی می‌شود در صورتی که جنسیت مرد نسبت به زن باعث کاهش لگاریتم بخت صفر آماسیده تعداد نیم‌سال‌های مشروطی می‌شود. در نظر گرفتن اثر تصادفی مختلف در مدل معنی‌داری پارامترهای برآورد شده را تحت تأثیر قرار می‌دهد، به‌طوری‌که در مدل پواسن دو متغیره با صفر آماسیده بدون اثر تصادفی متغیر مقطع دکتری و بومی بودن در مدل‌بندی تعداد واحدهای مردودی معنی دار شده است در صورتی که با اضافه شدن اثر تصادفی نرمال مستقل، چوله نرمال مستقل و چوله نرمال دو متغیره به این مدل، این دو متغیر معنی‌دار نمی‌شوند. متغیر مقطع دکتری در مدل‌بندی صفرهای آماسیده (واحدهای مردودی) در مدل پواسن دو متغیره با صفر آماسیده بدون اثر تصادفی، با اثر تصادفی نرمال مستقل و چوله نرمال مستقل معنی‌دار است در صورتی که در مدل با اثر تصادفی چوله نرمال دو متغیره معنی‌دار نیست. همچنین در مدل پواسن دو متغیره با صفر آماسیده بدون اثر تصادفی متغیر جنیست در مدل‌بندی تعداد نیم‌سال‌های مشروطی معنی‌دار شده است این در حالی است که با اضافه شدن اثر تصادفی نرمال مستقل، چوله نرمال مستقل و چوله نرمال دو متغیره، این متغیر معنی‌دار نمی‌شود. همچنین متغیرهای سهمیه منطقه یک، دو و سه در مدل‌بندی تعداد نیم‌سال‌های مشروطی در مدل پواسن دو متغیره با صفر آماسیده همراه با اثر تصادفی چوله نرمال مستقل و چوله نرمال دو متغیره معنی‌دار نیستند. اما مدل پواسن دو متغیره

با صفر آماسیده بدون اثر تصادفی، سهمیه منطقه یک و دو و در مدل پواسن دو متغیره با صفر آماسیده با اثر تصادفی نرمال مستقل سهمیه منطقه دو در مدل‌بندی تعداد نیم‌سال‌های مشروطی معنی‌دار شده‌اند. همچنین متغیرهای جنسیت و غیربومی بودن در مدل‌بندی صفرهای آماسیده (نیم‌سال‌های مشروطی) در مدل پواسن دو متغیره با صفر آماسیده بدون اثر تصادفی معنی‌دار نیستند در صورتی که با اضافه کردن اثر تصادفی نرمال مستقل و چوله نرمال مستقل به مدل این دو متغیر معنی‌دار می‌شوند. و با اضافه کردن اثر تصادفی چوله نرمال دو متغیره متغیر جنسیت معنی‌دار شده است و متغیر غیربومی معنی‌دار نمی‌شود. برآورد پارامتر چولگی $\hat{\lambda} = ۷/۱۱۹۷$ نشانگر مناسب‌تر بودن توزیع چوله نرمال برای اثر تصادفی نسبت به توزیع نرمال است.

جدول ۵: برآورد پارامترهای مدل پواسن دو متغیره با صفر آماسیده با اثر تصادفی چوله نرمال مستقل

مدل	سطح مرجع	متغیرها	برآورد	انحراف معیار	مقدار احتمال
پواسن (واحدهای مردودی)	–	مقدار ثابت	۱/۶۶۱۱	۰/۱۲۱۱	۰/۰۰۰۱ >
	کارشناسی	کارشناسی‌ارشد	−۰/۹۱۳۵	۰/۰۵۴۴	۰/۰۰۰۱ >
		دکتری	−۰/۱۶۱۰	۰/۱۵۲۴	۰/۳۰۶۴
	زن	مرد	۰/۱۹۳۷	۰/۰۲۰۳	۰/۰۰۰۱ >
	بومی	غیربومی	−۰/۰۲۱۴	۰/۰۲۳۶	۰/۳۷۸۴
	سایر سهمیه‌ها	منطقه یک	−۰/۲۴۷۷	۰/۰۵۵۴	۰/۰۰۰۴
		منطقه دو	−۰/۱۷۱۱	۰/۰۳۹۱	۰/۰۰۰۵
		منطقه سه	−۰/۳۱۳۶	۰/۰۳۹۸	۰/۰۰۰۱ >
پواسن (نیم‌سال‌های مشروطی)	–	مقدار ثابت	−۰/۱۱۷۱	۰/۱۸۷۰	۰/۵۳۹۹
	کارشناسی	کارشناسی‌ارشد	−۱/۹۱۷۸	۰/۳۱۳۴	۰/۰۰۰۱ >
		دکتری	−۰/۲۵۶۴	۰/۵۵۳۱	۰/۶۴۹۲
	زن	مرد	۰/۱۵۲۰	۱/۳۸	۰/۱۸۷۳
	بومی	غیربومی	۰/۳۱۹۹	۰/۱۲۵۱	۰/۲۱۱۰
	سایر سهمیه‌ها	منطقه یک	−۰/۴۴۰۷	۰/۲۴۵۵	۰/۰۹۱۶
		منطقه دو	−۰/۳۴۵۹	۰/۱۶۶۳	۰/۰۵۴۰
		منطقه سه	−۰/۲۶۴۳	۰/۱۵۸۸	۰/۱۱۵۵
صفر آماسیده (واحدهای مردودی)	–	مقدار ثابت	−۰/۹۲۴۷	۰/۱۸۴۹	۰/۰۰۰۱
	کارشناسی	کارشناسی‌ارشد	۲/۱۸۹۶	۱۶/۰۳	۰/۰۰۰۱ >
		دکتری	۱/۳۲۷۴	۴/۲۶	۰/۰۰۰۶
	زن	مرد	−۰/۷۲۱۱	۰/۰۹۷۸۶	۰/۰۰۰۱ >
	بومی	غیربومی	۰/۴۴۱۳۶	۳/۹۰	۰/۰۰۱۳
	سایر سهمیه‌ها	منطقه یک	۰/۷۱۹۰	۰/۲۴۹۵	۰/۰۱۰۸
		منطقه دو	۰/۵۶۱۶	۲/۹۷	۰/۰۰۹
		منطقه سه	۰/۵۶۰۱	۰/۱۸۹۳	۰/۰۰۹۲

جدول ۶: برآورد پارامترهای مدل پواسن دو متغیره با صفر آماسیده با اثر تصادفی چوله نرمال مستقل (ادامه)

مدل	سطح مرجع	متغیرها	برآورد	انحراف معیار	مقدار احتمال
صفر آماسیده (نیم‌سال‌های مشروطی)	–	مقدار ثابت	۰/۴۷۰۵	۰/۲۱۵۹	۰/۰۴۴۶
	کارشناسی	کارشناسی‌ارشد	−۱/۴۶۲۰	۰/۸۱۹۷	۰/۰۹۳۵
		دکتری	۰/۴۸۹۹	۰/۶۵۷۶	۰/۴۶۷۱
	زن	مرد	−۰/۵۶۵۴	۰/۱۸۳۷	۰/۰۰۷۲
	بومی	غیربومی	۰/۸۳۴۰	۰/۱۹۴۲	۰/۰۰۰۶
	سایر سهمیه‌ها	منطقه یک	−۰/۵۸۲۵	۰/۳۸۶۵	۰/۱۵۱۳
		منطقه دو	−۰/۰۴۳۶	۰/۲۳۴۵	۰/۸۵۴۸
		منطقه سه	۰/۰۴۲۴	۰/۱۸۰۲	۰/۸۱۶۸
اثر تصادفی		λ	۷/۱۱۹۷	–	–
		σ_1^2	۰/۲۷۱۵	–	–
		σ_2^2	۰/۰۳۲۵	–	–

۴ مطالعه شبیه‌سازی

برای بررسی سازگاری و کارایی برآورد پارامترهای رگرسیون پواسن با صفر آماسیده زمانی که توزیع اثر تصادفی چوله نرمال باشد، از روش شبیه‌سازی استفاده می‌کنیم. در این مطالعه شبیه‌سازی نمونه‌هایی به حجم ۲۰۰، ۵۰۰ و ۱۰۰۰تایی با تکرار ۲۰۰ از توزیع چوله نرمال با پارامترهای $\mu = ۰$، $\sigma^۲ = ۰/۵$ و $\lambda = ۴$ تولید شده است. در مرحله دو از رابطه‌های

$$\mu_i = \exp(\beta_۰ + \beta_۱ X_۱ + \beta_۲ X_۲ + u_i)$$

$$p = \frac{\exp(\gamma_۰ + \gamma_۱ X_۱ + \gamma_۲ X_۲)}{۱ + \exp(\gamma_۰ + \gamma_۱ X_۱ + \gamma_۲ X_۲)}$$

به‌ترتیب μ_i و p را تولید می‌کنیم. در اینجا متغیر $X_۱$ از توزیع نرمال استاندارد و متغیر $X_۲$ از توزیع برنولی با پارامتر، ۰/۵ هستند. با قرار دادن $\beta_۰ = ۱$، $\beta_۱ = ۰/۵$، $\beta_۲ = ۰/۱$، $\gamma_۰ = ۰/۱$، $\gamma_۱ = ۰/۵$ و $\gamma_۲ = ۰/۵$ نمونه‌هایی به حجم ۲۰۰، ۵۰۰ و ۱۰۰۰ تایی با تکرار ۲۰۰ از توزیع پواسن با صفر آماسیده تولید شدند. اعتبارسنجی براساس معیار اریبی $Bias = \frac{۱}{m}\sum_{\ell=۱}^{m}(\hat{\beta}_\ell - \beta)$ و میانگین توان دوم خطا $MSE = \frac{۱}{m}\sum_{\ell=۱}^{m}(\hat{\beta}_\ell - \beta)^۲$ صورت می‌پذیرد، در اینجا $\hat{\beta}_\ell$ مقدار برآورد پارامتر، β مقدار واقعی آن

و m تعداد تکرار است.

جدول ۷: اریبی و میانگین توان دوم خطای برآورد پارامترهای مدل پواسن با صفر آماسیده و اثر تصادفی چوله‌نرمال

پارامتر	حجم نمونه		
	۲۰۰	۵۰۰	۱۰۰۰
β_1	۰/۰۶۷۵(۰/۰۲۵۱)	۰/۰۶۰۰(۰/۰۲۳۶)	۰/۰۵۹۳(۰/۰۲۱۷)
β_2	۰/۰۳۳۵(۰/۰۲۰۱)	۰/۰۳۰۸(۰/۰۱۹۸)	۰/۰۲۹۰(۰/۰۱۸۰)
β_3	۰/۱۲۵۷(۰/۰۰۴۱)	۰/۱۱۰۴(۰/۰۰۳۹)	۰/۰۹۴۰(۰/۰۰۳۸)
γ_0	۰/۰۳۴۷(۰/۰۲۳۷)	۰/۰۳۱۱(۰/۰۲۱۹)	۰/۰۲۹۱(۰/۰۱۹۹)
γ_1	۰/۲۴۱۶(۰/۰۱۵۲)	۰/۲۰۷۴(۰/۰۱۴۹)	۰/۱۹۷۰(۰/۰۱۳۷)
γ_2	۰/۰۴۹۰(۰/۰۳۱۱)	۰/۰۴۶۹(۰/۰۲۹۳)	۰/۰۴۴۱(۰/۰۲۷۹)
λ	−۰/۰۵۸۱(۰/۰۱۹۶)	−۰/۰۵۵۰(۰/۰۱۷۹)	−۰/۰۵۳۱(۰/۰۱۶۱)
σ^2	−۰/۰۲۲۱(۰/۰۰۹۱)	−۰/۰۲۰۳(۰/۰۰۸۶)	−۰/۰۱۸۹(۰/۰۰۶۷)

همچنین برای بررسی سازگاری و کارایی برآورد پارامترهای رگرسیون پواسن دو متغیره با صفر آماسیده زمانی که توزیع اثر تصادفی چوله نرمال مستقل باشد، مانند آنچه در مدل یک متغیره گفته شد عمل می‌کنیم. به این صورت که نمونه‌هایی به حجم ۲۰۰، ۵۰۰ و ۱۰۰۰تایی با تکرار ۲۰۰ از توزیع چوله نرمال با پارامترهای $\mu = 0$، $\sigma^2 = 0/3$ و $\lambda = 4$ تولید شده است. در مرحله دو از رابطه‌های

$$\mu_{i1} = \exp(\beta_{10} + \beta_{11}X_1 + \beta_{12}X_2 + u_i),$$

$$\mu_{i2} = \exp(\beta_{20} + \beta_{21}X_1 + \beta_{22}X_2 + u_i),$$

$$p_1 = \frac{\exp(\gamma_{10} + \gamma_{11}X_1 + \gamma_{12}X_2)}{1 + \exp(\gamma_{10} + \gamma_{11}X_1 + \gamma_{12}X_2)},$$

$$p_2 = \frac{\exp(\gamma_{20} + \gamma_{21}X_1 + \gamma_{22}X_2)}{1 + \exp(\gamma_{20} + \gamma_{21}X_1 + \gamma_{22}X_2)},$$

μ_{i1}، μ_{i2}، p_1 و p_2 را تولید می‌کنیم. در اینجا متغیر X_1 از توزیع نرمال استاندارد و متغیر X_2 از توزیع برنولی با پارامتر، ۰/۵ هستند. با قرار دادن $\beta_{10} = 0/1$، $\beta_{11} = 0/5$، $\beta_{12} = 0/1$، $\beta_{20} = 1$، $\beta_{21} = 0/5$، $\beta_{22} = 0/1$، $\gamma_{10} = 0/1$، $\gamma_{11} = 0/5$، $\gamma_{12} = 0/5$، $\gamma_{20} = 0/1$ ، $\gamma_{21} = 0/5$ و $\gamma_{22} = 0/5$ نمونه‌هایی به حجم ۲۰۰، ۵۰۰ و ۱۰۰۰ تایی با تکرار ۲۰۰ از توزیع پواسن دو متغیره با صفر آماسیده تولید شدند.

جدول ۸: اریبی و میانگین توان دوم خطای برآورد پارامترهای (مدل پواسن دو متغیره با صفر آماسیده واثر تصادفی چوله نرمال مستقل)

پارامتر	حجم نمونه ۲۰۰	۵۰۰	۱۰۰۰
β_{10}	(۰/۰۱۴۸)۰/۱۸۵۰	(۰/۰۱۱۲)۰/۱۷۰۱	(۰/۰۰۹۶)۰/۱۵۹۹
β_{11}	(۰/۰۱۷۴)۰/۲۵۴۱	(۰/۰۱۷۱)۰/۲۴۰۹	(۰/۰۱۶۹)۰/۲۲۰۱
β_{12}	(۰/۰۰۲۳)۰/۰۳۵۰	(۰/۰۰۲۱)۰/۰۳۳۹	(۰/۰۰۱۸)۰/۰۳۲۱
β_{20}	(۰/۰۰۴۴)۰/۰۴۶۹	(۰/۰۰۴۰)۰/۰۴۵۱	(۰/۰۰۳۸)۰/۰۴۴۰
β_{21}	(۰/۰۰۳۴)۰/۰۶۲۱	(۰/۰۰۳۳)۰/۰۶۰۹	(۰/۰۰۳۰)۰/۰۵۹۱
β_{22}	(۰/۰۰۵۳)۰/۰۸۱۱	(۰/۰۰۵۰)۰/۰۷۱۹	(۰/۰۰۴۸)۰/۰۶۳۵
γ_{10}	(۰/۰۰۹۱)۰/۰۵۱۸	(۰/۰۰۹۰)۰/۰۴۹۷	(۰/۰۰۸۸)۰/۰۴۷۴
γ_{11}	(۰/۰۱۱۱)۰/۱۶۶۰	(۰/۰۱۰۰)۰/۱۶۰۱	(۰/۰۰۹۷)۰/۱۵۹۳
γ_{12}	(۰/۰۲۶۷)۰/۰۳۳۸	(۰/۰۲۶۵)۰/۰۳۲۱	(۰/۰۲۶۲)۰/۰۳۰۹
γ_{20}	(۰/۰۰۶۳)۰/۰۴۸۱	(۰/۰۰۶۱)۰/۰۴۷۵	(۰/۰۰۵۹)۰/۰۴۵۹
γ_{21}	(۰/۰۱۸۵)۰/۲۳۲۸	(۰/۰۱۷۹)۰/۲۱۰۵	(۰/۰۱۵۶)۰/۱۹۳۹
γ_{22}	(۰/۰۲۰۵)۰/۰۳۳۰	(۰/۰۲۰۰)۰/۰۳۱۹	(۰/۰۱۹۸)۰/۰۳۰۱
λ	(۰/۰۰۵۵)−۰/۰۴۵۶	(۰/۰۰۵۱)−۰/۰۴۳۸	(۰/۰۰۴۸)−۰/۰۴۱۸
σ_1^2	(۰/۰۰۷۲)−۰/۱۹۴۸	(۰/۰۰۷۱)−۰/۱۹۰۱	(۰/۰۰۶۹)−۰/۱۸۰۷
σ_2^2	(۰/۰۰۹۴)−۰/۲۱۴۲	(۰/۰۰۹۳)−۰/۲۰۷۹	(۰/۰۰۹۰)−۰/۱۹۶۰

با توجه به نتایج جدول‌های ۷ و ۸ مشاهده می‌شود که با افزایش حجم نمونه اریبی و میانگین توان دوم خطا تمام پارامترها کاهش می‌یابد و به سمت صفر میل می‌کند. با این حال نرخ کاهش برای تمام پارامترها یکسان نمی‌باشد. همچنین می‌توان گفت در مدل رگرسیونی پواسن با صفر آماسیده با اثر تصادفی چوله نرمال پارامترهای λ، σ^2 و در مدل رگرسیونی پواسن دو متغیره با صفر آماسیده با اثر تصادفی چوله نرمال مستقل پارامترهای λ، σ_1^2 و σ_2^2 کم برآورد شده‌اند.

بحث و نتیجه‌گیری

در این مطالعه مدل رگرسیونی برای تحلیل داده‌های گسسته، زمانی که فراوانی بیش از حد در عدد صفر وجود داشته باشد ارائه شد. برای این کار از رگرسیون پواسن با صفر آماسیده و همچنین برای در نظر گرفتن عوامل ناشناخته و غیرقابل اندازه‌گیری از اثر تصادفی با توزیع چوله نرمال استفاده و نتایج بدست آمده را با زمانی که مدل فاقد اثر تصادفی یا دارای اثر تصادفی نرمال است از نظر معیار آکائیک مورد مقایسه قرار

گرفت. که براساس معیار آکائیک بدست آمده در بین مدل‌های یک متغیره برازش شده، مدل رگرسیون پواسن با صفر آماسیده با اثر تصادفی چوله نرمال و همچنین در بین مدل‌های دو متغیره، مدل رگرسیون پواسن دو متغیره با صفر آماسیده با اثر تصادفی چوله نرمال مستقل، کمترین مقدار آکائیک و در نتیجه برازش بهتری نسبت به سایر مدل‌ها به داده‌ها بود.

مراجع

تقربی، ز.، فخاریان، ا.، میرحسینی، ف.، رسولی نژاد، ا.، اکبری، ح. و عاملی، ح. (۱۳۸۸)، عوامل مرتبط با وقوع مشروطی در دانش‌آموختگان دانشکده پرستاری و مامایی کاشان، مجله آموزش در علوم پزشکی، **۹**، ۲۱-۲۹.

حبیب‌زاده، ش.، علیزاده، ه.، پورفرضی، ف. قاسمی، ا. و امینی ملکی، ط. (۱۳۹۰)، بررسی میزان افت تحصیلی و عوامل مؤثر بر آن در دانشجویان، مجله سلامت و مراقبت، شماره **۳**، ۲۹-۳۳.

رودباری، م.، احمدی، آ. و عبادی فرد آذر، ف. (۱۳۸۹)، تعیین عوامل مؤثر بر پیشرفت تحصیلی دانشجویان علوم پزشکی تهران (پردیس همت) در سال تحصیلی ۸۹-۸۸، نشریه طب و تزکیه، **۱۹**، ۳۷-۴۸.

رودباری، م.، صالحی، م. (۱۳۹۳)، به کارگیری رگرسیون پواسنی و رگرسیون دوجمله‌ای منفی با انباشتگی در صفر برای مدل‌سازی داده‌های آموزشی، مجله علوم پزشکی رازی، شماره ۱۱۹، ۱۸-۲۴.

Aitkin, M. (1999), Meta-Analysis by Random Effects Modelling in Generalized Linaer Models, *Statistics in Medicine*, **18**, 2334-2351.

Azzalini, A. (1985), A Class of Distributions which Includes the Normal Ones, *Scandinavian Journal of Statistics*, **12**, 171-178.

Azzalini, A. (2005), The Skew-Normal Distribution and Related Multivariate Families, *Scandinavian Journal of Statistics*, **32(2)**, 159–188.

Burr, D., Doss, H., Cooke, G. E. and Goldschmidt-Clermont, P. J. (2003), A Meta-Analysis of Studies on the Association of the Platelet PlA Polymorphism of Glycoprotein IIIa and Risk of Coronary Heart Disease, *Statistics in Medicine*, **22**, 1741–1760.

Carrivick, P. J. W., Lee, A. H. and Yau, K. K. E. (2003), Zero-Inflated Poisson Modeling to Evaluate Occupational Safety Interventions, *Safety Science*, **41**, 53-63.

Congdon, P. (2006), A Model for Non-Parametric Spatially Varying Regression Effects, *Computational Statistics and Data Analysis*, **50**, 422-445.

Dong, C., Richards, S. H., Clarke, D. B., Zhou, X. and Ma, Z. (2014), Examining Signalized Intersection Crash Frequency Using Multivariate Zero-Inflated Poisson Regression, *Safety Science*, **70**, 63-69.

Fox, J. P. (2013), Multivariate Zero-Inflated Modeling with Latent Predictors: Modeling Feedback Behavior, *Computational Statistics and Data Analysis*, **68**, 361-374.

Lambert, D. (1992), Zero-Inflated Poisson Regression, with an Application to Defects in Manufacturing, *Technometrics*, **34**, 1-14.

Lao, Y., Wu, Y. J., Corey, J. and Wang, Y. (2011), Modeling Animal-Vehicle Collisions Using Diagonal Inflated Bivariate Poisson Regression, *Accident Analysis and Prevention*, **43**, 220-227.

Lee, A. H., Stevenson, M. R., Wang, K. and Yau, K. K. W. (2002), Modeling Young Driver Motor Vehicle Crashes: Data with Extra Zeros, *Accident Analysis and Prevention*, **34**, 515-521.

Marshal, E. C. and Spiegelhalter, D. J. (1998), Comparing Instiutional Performance Using Markov Chain Monte Carlo Methods, *Statistical Analysis of Medical Data: New Developmets*, 229-249.

Rastegaran, A and Zadkarami, M. R. (2015), A Skew-Normal Random Effects Model for Longitudinal Ordinal Categorical Responses with Missing Data, *Journal of Applied Statistics*, **42**, 114-126.

Sahu, S. K., Dey, D. K. and Branco, M. (2003), A New Class of Multivariate Skew Distributions with Applications to Bayesian Regression Models, *The Canadian Journal of Statistics*, **31**, 129-150.

Shams, B., Farshidfard, M. and Hassanzadeh, A. (1997), The Comparison of Demographic and Personality Characteristics in Fail and Succeed Students of Isfahan University of medical sciences, *Reaearch in Medical Science*, **4**, 222-226.

Turner, R. M., Omar, R. Z. and Thompson, S. G. (2001), Bayesian Methods of Analysis for Cluster Randomized Trials with Binary Outcome Data, *Statistics in Medicine*, **20**, 453-472.

Verbeke, G. and Lesaffre, E. (1996), A Linear Mixed-Effects Model with Heterogeneity in the Random-Effects population, *Journal of the American Statistical Association*, **91**, 217-221.

Wang, K., Lee, A. H., Yau, K. K. W. and Carrivick, P. J. W. (2003), A Bivariate Zero-Inflated Poisson Regression Model to Analyze Occupational Injuries, *Accident Analysis and Prevention*, **35**, 625-629.

Wong, K. Y. and Lam, K. F. (2013), Modeling Zero-Inflated Count Data Using a Covariate Dependent Random Effect Model, *Statistics in Medicine*, **32**, 1283-1293.

Yau, K. K. W. and Lee, A. H. (2001), Zero-Infated Poisson Regression with Random Effects to Evaluate an Occupational Injury Prevention Programme, *Statistics in Medicine*, **20**, 2907-2920.

Zaninotto, P. and Falaschetti, E. (2011), Comparison of Methods for Modelling a Countoutcome with Excess Zeros: Application to Activities of Daily Living (ADL-s), *Epidemiol Community Health*, **65**, 205-210.