مجله علوم آماری، پاییز و زمستان ۱۳۹۵

جلد ۱۰، شماره ۲، ص ۲۸۱-۲۹۷

DOI: 10.18869/acadpub.jss.10.2.281

تأثیر تابع وزن در برآورد پارامترهای رگرسیونی تحت نمونه‌گیری موزون

مهتاب طرهانی، سید محمد رضا علوی

گروه آمار، دانشگاه شهید چمران اهواز

تاریخ دریافت: ۱۳۹۳/۵/۱۳ تاریخ آخرین بازنگری: ۱۳۹۴/۹/۷

چکیده : در نمونه‌گیری موزون به عنوان تعمیمی از نمونه‌گیری تصادفی هر مشاهده، با احتمالی متناسب با یک تابع نامنفی از آن ثبت می‌شود. در این مقاله مدل رگرسیونی نرمال تحت نمونه‌گیری موزون برای یک وزن کلی پیشنهادی مطالعه می‌گردد. پارامترهای مدل در دو حالت معلوم و نامعلوم بودن پارامتر وزن برآورد می‌شوند و با استفاده از شبیه‌سازی کارایی برآوردها هنگامی که شکل بسته‌ای برای آنها به‌دست نمی‌آید مطالعه می‌شوند. به‌عنوان یک کاربرد، داده‌های تعداد نسخه پزشکان متخصص طرف قرارداد سازمان تأمین اجتماعی اهواز این مدل تحلیل می‌شوند.

واژه‌های کلیدی : برآورد، تابع وزن، ضرایب رگرسیونی، شبیه سازی، نمونه‌گیری موزون.

° . آدرس الکترونیک مسئول مقاله: سید محمدرضا علوی، alavi_m@scu.ac.ir

° . کد موضوع‌بندی ریاضی (۲۰۱۰): ۶۲J۰۵، ۶۰E۰۵

۱ مقدمه

هر چند بسیاری از استنباط‌های آماری مبتنی بر تصادفی بودن نمونه‌ها است، اما موارد زیادی وجود دارند که مشاهده‌های ثبت شده، یک نمونه تصادفی از جامعه اصلی نیستند. اگر سازوکار ثبت مشاهده‌ها به‌گونه‌ای باشد که مشاهده با مقدار y با احتمالی متناسب با یک تابع نا منفی $w(y,\alpha)$ ثبت شود، به سازوکار ثبت مشاهده‌ها نمونه‌گیری موزون و به این تابع نامنفی تابع وزن گفته می‌شود. α پارامتر تابع وزن است. واضح است که اگر این تابع نسبت به y ثابت باشد، نمونه‌گیری تصادفی است. بنابراین نمونه‌گیری موزون یک حالت تعمیم یافته از نمونه‌گیری تصادفی است.

در مدل رگرسیونی خطی نرمال توزیع شرطی متغیر پاسخ، Y، نرمال با میانگین $\mathbf{x}'\beta = \beta_۰ + \beta_۱ x_۱ + \ldots + \beta_k x_k$ و واریانس ثابت $\sigma^۲$ است. به‌عبارت‌دیگر، $\mathbf{Y}|\mathbf{x} \sim \mathbf{N}(\mathbf{x}'\beta,\, \sigma^۲)$، که درآن $\mathbf{x} = (۱, x_۱, \ldots, x_k)'$ بردار متغیرهای کمکی و $\beta = (\beta_۰, \ldots, \beta_k)$ بردار ضرایب رگرسیونی است. هنگامی‌که بردار مشاهده‌های $\mathbf{Y} = (Y_۱, \ldots, Y_n)$، یک نمونه تصادفی از Y است، برآورد ماکسیمم‌درستنمایی ضرایب مدل رگرسیونی نرمال با فرض پررتبه ستونی بودن ماتریس طرح $\mathbf{X} = (\mathbf{1}, \mathbf{x}_۱, \ldots, \mathbf{x}_k)$ به‌صورت $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\mathbf{X}'\mathbf{Y}$ می‌باشد که در آن $\mathbf{x}_i = (x_{i۱}, \ldots, x_{in})'$ و $\mathbf{1}$ برداری با مولفه‌های یک است.

اگر هر مشاهده y_i تحت وزن $w(y_i,\alpha)$ ثبت شود، به $(y_۱, \ldots, y_n)$ یک نمونه موزون مشاهده شده از Y تحت وزن $w(y,\alpha)$ گفته می‌شود. این مشاهده‌ها یک نمونه تصادفی از متغیر جدیدی به نام نسخه موزونی Y تحت وزن $w(y,\alpha)$ است و با Y^w نمایش داده می‌شود. به توزیع Y^w، توزیع موزون Y گفته می‌شود. درصورتی‌که $E[w(Y,\alpha)]$ موجود باشد، تابع چگالی احتمال Y^w، بنا به قانون بیز به‌صورت

$$f^w(y;\theta,\alpha) = \frac{w(y,\alpha)f(y;\theta)}{E[w(Y,\alpha)]} \qquad (۱)$$

است. نمونه‌گیری تصادفی حالت خاص نمونه‌گیری موزون است، هنگامی که تابع وزن نسبت به y ثابت باشد. اگر تابع وزن به‌صورت $w(y,\alpha) = y^\alpha$ باشد،

نمونه‌گیری را اریب‌اندازه مرتبه α گویند. گاهی به نمونه‌گیری اریب اندازه مرتبه اول نمونه‌گیری اریب طول نیز گفته می‌شود. این نوع نمونه‌گیری برای مثال در علوم مربوط به حیاط وحش و بوم شناسی هنگامی که نمونه‌ها با عکس‌های هوایی شناسایی می‌شوند، به‌کار می‌رود. به‌عنوان مثال در نمونه‌گیری مربعی که ناحیه مورد مطالعه به مربع‌های یکسان تقسیم می‌شود و تعداد حیوان‌ها در هر مربع با چشم توسط افراد ماهر شمارش می‌شود، به‌دلیل اینکه مربعی که حیوان زیادتری داشته باشد شانس بیشتری برای انتخاب دارد، حیوان‌های شمارش شده یک نمونه اریب اندازه می‌باشند (پاتیل، ۲۰۰۲). توزیع‌های موزون برای اولین بار توسط رائو (۱۹۶۵) به‌طور رسمی معرفی گردید. اخیراً محققین زیادی از جمله گوپتا و کوندا (۲۰۰۹)، علوی و چینی پرداز (۲۰۰۹)، چاکرا براتی (۲۰۱۰)، کریمی و علوی (۲۰۱۴) و علوی (۲۰۱۷) از توزیع‌های موزون در زمینه‌های مختلف استفاده کرده‌اند. یکی از موقعیت‌های مهم در توزیع‌های موزون هنگامی است که توزیع تحت وزن مورد نظر پایا باشد یعنی شکل تابع چگالی توزیع موزون و توزیع اصلی یکسان باشد. در چنین موقعیت‌هایی اگر پارامتر وزن معلوم باشد، استنباط پارامترهای مدل با اندکی تغییر در روش‌های معمول قابل حصول است (علوی و چینی‌پرداز، ۲۰۰۹ و علوی، ۲۰۱۷) اما هنگامی که پارامتر وزن نامعلوم باشد پارامتر وزن و پارامترهای مدل با مشکل غیر قابل تشخیص شدن مواجه می‌شوند (زادکرمی، ۲۰۰۸). بنابر (۱) اگر مشاهدات $(y_1, \ldots, y_n)$ یک نمونه موزون از مدل رگرسیونی نرمال باشند، تابع چگالی احتمال موزون متغیر پاسخ به‌صورت

$$f^w(y;\beta,\alpha) = \frac{w(y,\alpha)}{E[w(Y,\alpha)]}\frac{1}{\sqrt{2\pi}\sigma}exp\{-\frac{1}{2\sigma^2}(y-x'\beta)^2\}$$

خواهد بود. کریستوبال و آلکالا (۲۰۰۰) برآوردگرهای رگرسیونی ناپارامتری برای داده‌های اریب‌طول را به‌دست آوردند. علوی (۲۰۰۸) مدل رگرسیونی موزون تحت وزن نمایی با پارامتر وزن معلوم را مطالعه کرد. چن (۲۰۰۹) مدل رگرسیون نیمه پارامتری را برای داده‌های اریب اندازه به‌کار برد.

دربخش ۲ با معرفی یک وزن پیشنهادی جدید، مدل رگرسیونی نرمال موزون تحت این وزن مطالعه و برآورد ماکسیمم درستنمایی ضرایب رگرسیونی وقتی

پارامترهای وزن معلوم باشند، تعیین می‌شود. این تابع وزن تعمیمی از وزن‌هایی است که در ادبیات توزیع‌های موزون استفاده می‌شود. بخش ۳ به مطالعه برآورد پارامترهای مدل اختصاص می‌یابد هنگامی‌که پارامترهای وزن نامعلوم باشند. در بخش ۴ با استفاده از شبیه‌سازی خواص برآوردها هنگامی که شکل‌های بسته برای برآوردها به‌دست نمی‌آیند مطالعه می‌شوند و در نهایت در بخش ۵ یک مدل رگرسیونی برای داده‌های تعداد نسخه‌های پزشکان طرف قرارداد سازمان تأمین اجتماعی اهواز که توسط کریمی و علوی (۲۰۱۴) به‌عنوان یک نمونه موزون استفاده شده‌اند، تحلیل می‌شوند. برآورد پارامترها تحت سه وزن مختلف محاسبه و با نمونه‌گیری تصادفی مقایسه می‌شوند.

۲ برآورد پارامترهای مدل رگرسیونی نرمال موزون

فرض کنید مشاهده‌های $(y_1, \dots, y_n)$ در مدل رگرسیونی نرمال تحت وزن پیشنهادی

$$w(y, \alpha, \beta, \sigma^2) = e^{(\alpha_1 y + \alpha_2 y^2)} y^{\alpha_3} \phi(\alpha_4 \frac{y - x'\beta}{\sigma}) \tag{2}$$

ثبت شده باشند، که در آن $\alpha = (\alpha_1, \alpha_2, \alpha_3, \alpha_4)$ بردار پارامترهای وزن، $\phi(\cdot)$ تابع توزیع نرمال استاندارد و σ^2 واریانس مدل رگرسیونی نرمال است. دلیل ارائه این وزن جدید آن است که در بردارنده گونه‌های مختلف وزن‌هایی است که در ادبیات توزیع‌های موزون به کار می‌روند. واضح است که $E[w(Y, \alpha, \beta, \sigma^2)] < \infty$. تابع چگالی متغیر پاسخ مشاهده‌های ثبت شده به‌صورت

$$f^w(y; \alpha, \beta, \sigma^2) = \frac{w(y, \alpha, \beta, \sigma^2; x)}{c(\alpha, \beta, \sigma^2; x) \frac{1}{\sigma} \varphi(\frac{y - x'\beta}{\sigma})} \tag{3}$$

است، که در آن $E[w(Y, \alpha, \beta, \sigma^2)] = c(\alpha, \beta, \sigma^2; \mathbf{x})$ و ϕ تابع چگالی احتمال نرمال استاندارد است. قابل توجه است که به‌ازای $\alpha_4 \neq 0$، تابع وزن نیز به $\mathbf{x}$ بستگی دارد.

حالات خاص مدل (۲) عبارتند از:

الف– به‌ازای $\alpha = \mathbf{0}$، مدل رگرسیونی خطی معمولی به‌دست می‌آید که

$Y^w \sim N(\mathbf{x}'\beta, \sigma^{۲})$.

ب– برای $\alpha = (۰, ۰, ۰, \alpha_۴)$، مدل رگرسیونی خطی نرمال چوله (سارتوری، ۲۰۰۶) حاصل می‌شود، یعنی $Y^w \sim SN(\mathbf{x}'\beta, \sigma^{۲}, \alpha_۴)$ که نماد SN بیانگر توزیع چوله-نرمال است.

ج– برای $\alpha = (\alpha_۱, ۰, ۰, ۰)$، مدل رگرسیونی خطی نرمال موزون نمایی درجه اول به‌دست می‌آید، یعنی $Y^w \sim N(\mathbf{x}'\beta + \alpha_۱\sigma^{۲}, \sigma^{۲})$ (علوی و چینی‌پرداز، ۲۰۰۹).

د– برای $\alpha = (\alpha_۱, \alpha_۲, ۰, ۰)$، مدل رگرسیونی خطی نرمال موزون نمایی درجه دوم با چگالی زیر به‌دست می‌آید:

$$Y^w \sim N\left(\frac{x'\beta + \alpha_۱\sigma^{۲}}{۱ - ۲\alpha_۲\sigma^{۲}}, \frac{\sigma^{۲}}{۱ - ۲\alpha_۲\sigma^{۲}}\right). \qquad (۴)$$

ه– برای $\alpha = (۰, ۰, \alpha_۳, ۰)$، مدل موزون منجر به مدل رگرسیونی خطی نرمال اریب اندازه مرتبه $\alpha_۳$ می‌شود (علوی و چینی‌پرداز، ۲۰۰۹).

اگر $(y_۱, x_۱), \ldots, (y_n, x_n)$ مقادیر مشاهده شده یک نمونه موزون تحت وزن پیشنهادی (۲) با پارامترهای معلوم از مدل رگرسیون (۳) باشند، لگاریتم تابع درستنمایی موزون عبارت است از

$$\begin{aligned} \ell^w(\alpha, \beta, \sigma^{۲}) &= \sum \log\{w(y_i, \alpha, \beta, \sigma^{۲}; x_i)\} - \sum \log\{c(\alpha, \beta, \sigma^{۲}; x_i)\} \\ &- n\log(\sigma) - \frac{۱}{۲\sigma^{۲}} \sum (y_i - x_i'\beta)^{۲} \end{aligned}$$

روشن است که برای $\alpha = ۰$، هنگامی که ماتریس طرح $\mathbf{X}$ پر رتبه ستونی باشد، برآورد ضرایب رگرسیونی $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-۱}\mathbf{X}'\mathbf{Y}$ خواهد شد.

ز– برای $\alpha = (۰, ۰, ۰, \alpha_۴)$ برآورد ماکسیمم درستنمایی $(\beta, \sigma^{۲})$ شکل بسته‌ای ندارد و به روش‌های عددی محاسبه می‌شوند (سارتوری، ۲۰۰۶).

ح– برای $\alpha = (\alpha_۱, ۰, ۰, ۰)$ و $\sigma^{۲}$ معلوم، با توجه به این‌که:

$$E(Y^w|x) = (\beta_۰ + \alpha_۱\sigma^{۲}) + \beta_۱ x_۱ + \ldots + \beta_k x_k,$$

با تعریف $z_i = y_i - \alpha_۱\sigma^{۲}$ برآورد پارامترهای رگرسیونی به‌صورت $(\hat{\beta_۰}, \ldots, \hat{\beta_k})' = (\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\mathbf{X}'\mathbf{Z}$ به‌دست می‌آید، که در آن

$\hat{\beta_\circ} = \bar{y} - \sum_{i=1}^{k} \hat{\beta}_i \bar{x}_i - \alpha_1 \sigma^2$ و $Z = (z_1, \ldots, z_n)'$.
بنابراین در مقایسه با رگرسیون معمولی فقط عرض از مبدأ خط به اندازه $-\alpha_1\sigma^2$ تغییر می‌کند و ضرایب متغیرها مانند حالتی است که نمونه تصادفی است (علوی، ۲۰۰۸).

برای حالتی‌که σ^2 نامعلوم است، برآوردهای ماکسیمم درستنمایی (β, σ^2) شکل بسته‌ای ندارند و از روش‌های عددی باید محاسبه شوند.
ط ـ برای $\alpha = (\alpha_1, \alpha_2, 0, 0)$ و σ^2 معلوم، با توجه به (۴)، وقتی $\alpha_2 \leq \frac{1}{2\sigma^2}$ شکل بسته‌ای برای $\hat{\beta}$ به‌صورت

$$\hat{\beta} = (X'X)^{-1} X' [Y^w (1 - 2\alpha_2\sigma^2) - \alpha_1 \sigma^2 \mathbf{1}]$$

به‌دست می‌آید، که اولین مولفه $\hat{\beta}$ عبارت‌است از:

$$\hat{\beta_\circ} = (1 - 2\alpha_2\sigma^2) - \sum_{i=1}^{k} \hat{\beta}_i \bar{x}_i - \alpha_1 \sigma^2.$$

هنگامی‌که σ^2 نامعلوم باشد برآورد ماکسیمم درستنمایی (β, σ^2) شکل بسته‌ای ندارد و به‌روش‌های عددی به‌دست می‌آید.
ی ـ برای $\alpha = (0, 0, \alpha_3, 0)$ که α_3 عددی صحیح و مثبت باشد، برآوردهای ماکسیمم درستنمایی (β, σ^2) شکل بسته‌ای ندارند و به روش‌های عددی محاسبه می‌شوند. این برآوردهای ماکسیمم درستنمایی سازگار و به‌طور مجانبی دارای توزیع نرمال هستند و واریانس آنها با استفاده از معکوس ماتریس اطلاع فیشر محاسبه می‌شود (کسلا و برگر، ۲۰۰۲).

۳ برآورد پارامترها با پارامترهای وزن نامعلوم

در صورتی‌که پارامترهای وزن نامعلوم باشند در حالت کلی وقتی که توزیع موزون پایا نباشد برآوردهای ماکسیمم درستنمایی (β, σ^2) شکل بسته‌ای ندارند و به روش‌های عددی محاسبه می‌شوند. اگر توزیع موزون تحت وزن مورد نظر پایا باشد، ممکن است بعضی از پارامترها با مشکل غیر قابل تشخیص شدن مواجه

شوند. بنابراین نیاز به یک سری اطلاعات اضافی در مورد آن پارامترها داریم. یک راه حل این مسئله در بخش قبل با معلوم فرض کردن پارامترهای وزن بررسی شد. راه حل دیگر استفاده از روش ماکسیمم درستنمایی تعمیم یافته است که با در نظر گرفتن یک توزیع پیشین مناسب روی پارامترهای نامعلوم تحت شرایطی می‌توان پارامترهای وزن و مدل را با رهیافت بیزی برآورد کرد (زادکرمی، ۲۰۰۸).
بدون از دست دادن کلیت مسئله مدل رگرسیون خطی ساده در نظر گرفته می‌شود یعنی $\beta = (\beta_0, \beta_1)'$. فرض کنید توزیع Y^w تحت وزن $w(y,\alpha)$ پایا و $\pi(\beta,\alpha,\sigma^2)$ توزیع پیشین توأم پارامترها باشد. توزیع پسین یا تابع درستنمایی تعمیم یافته به‌صورت

$$\begin{aligned} GL(\beta,\alpha,\sigma^2) &= f^w(y|\beta,\alpha,\sigma^2)\pi(\beta,\alpha,\sigma^2)m^{-1}(y) \\ &= L(\beta,\alpha,\sigma^2)\pi(\beta,\alpha,\sigma^2)m^{-1}(y) \end{aligned}$$

است، که در آن

$$\begin{aligned} m(y) &= \int\int\int f^w(y|\beta,\alpha,\sigma^2)\pi(\beta,\alpha,\sigma^2)d\beta d\alpha d\sigma^2, \\ f^w(y|\beta,\alpha,\sigma^2) &= \prod_{i=1}^{n} f^w(y_i|\beta,\alpha,\sigma^2). \end{aligned}$$

با ماکسیمم کردن $GL(\beta,\alpha,\sigma^2)$ می‌توان پارامترهای مدل را برآورد کرد. برای مثال با توجه به این‌که توزیع موزون تحت وزن نمایی $w(y,\alpha_1,\alpha_2) = e^{\alpha_1 y+\alpha_2 y^2}$ پایا است، ابتدا حالت ساده $\alpha_2 = 0$ و $\sigma^2 = 1$ مطالعه می‌شود. یعنی:

$$Y_i^w \sim N(\beta_0 + \beta_1 x_i + \alpha_1, 1).$$

و تابع درستنمایی به‌صورت

$$L^w(\beta_0,\beta_1,\alpha_1,\sigma^2,y_i) = (2\pi)^{-n/2}e^{-1/2\sum(y_i-\beta_0-\beta_1 x_i-\alpha_1)^2}$$

است. چون α_1 و β_0 غیر قابل برآورد هستند، با درنظر گرفتن

$$\pi(\beta_0,\beta_1,\alpha_1) = \pi_0(\beta_1|\beta_0,\alpha_1)\pi_1(\beta_0|\alpha_1)\pi_2(\alpha_1),$$

و معرفی توزیع‌های پیشین به‌صورت

$$\pi_2(\alpha_1) \sim N(0,1), \pi_1(\beta_0|\alpha_1) \sim N(\alpha_1, 1), \pi_0(\beta_1|\beta_0, \alpha_1) = 1,$$

تابع درستنمایی تعمیم یافته برابر است با

$$GL(\theta) = (2\pi)^{-\frac{n+2}{2}} e^{-\frac{1}{2}\sum(y_i-\beta_0-\beta_1 x_i-\alpha_1)^2 - \frac{(\beta_0-\alpha_1)^2}{2} - \frac{\alpha_1^2}{2}} m^{-1}(y),$$

که در آن $\theta' = (\beta_0, \beta_1, \alpha_1)$ و

$$m(y) = \int\int\int (2\pi)^{-\frac{n+2}{2}} e^{-\frac{1}{2}\sum (y_i-\beta_0-\beta_1 x_i-\alpha_1)^2 - \frac{(\beta_0-\alpha_1)^2}{2} - \alpha_1{}^2/2} d\alpha_1 d\beta_0 d\beta_1$$

با مشتق گرفتن از لگاریتم این تابع نسبت به پارامترها و معادل صفر قرار دادن آنها برآوردهای ماکسیمم درستنمایی تعمیم یافته عبارتند از:

$$\begin{aligned}
\hat{\alpha_1} &= \frac{\sum y_i + \frac{n(1-n)}{n+1}\bar{y} - \frac{2}{n+1}\sum {x_i}^2 \frac{y_i - \frac{n}{n+2}\bar{y}}{\sum (x_i)^2 - \frac{n^2}{n+1}\bar{x}^2}}{\frac{5n-1}{n+1} - \frac{4n^2\bar{x}^2}{(n+1)^2[\sum x_i^2 - \frac{n^2}{n+1}\bar{x}^2]}},\\
\hat{\beta_0} &= \frac{n\bar{y} - \hat{\beta_1} n\bar{x} - (n-1)\hat{\alpha_1}}{n+1},\\
\hat{\beta_1} &= \frac{\sum x_i[y_i - \frac{n}{n+1}\bar{y} - \frac{2\hat{\alpha_1}}{n+1}]}{\sum x_i^2 - \frac{n^2 x^2}{n+1}}.
\end{aligned}$$

هنگامی که σ نامعلوم باشد، مشابه روش قبل و با در نظر گرفتن پیشین $\pi(\sigma|\beta_0, \alpha_1) = 1$ علاوه بر پیشین‌های معرفی شده، برآوردهای زیر حاصل می‌شوند:

$$\hat{\beta_0} = \frac{\bar{y} - \hat{\sigma}^2\hat{\alpha_1} - \hat{\beta_1}\bar{x} + \frac{\hat{\alpha_1}}{n}}{\frac{1}{n} + \frac{1}{\hat{\sigma^2}}},$$

$$\hat{\beta_1} = \frac{\sum x_i[((\frac{1}{n} + \frac{1}{\hat{\sigma^2}})y_i - \bar{y}) + \hat{\alpha_1}(\hat{\sigma}^2 - \frac{\hat{\sigma}^2}{n} - 1 - \frac{1}{n}))}{\sum_{i=1}^n x_i^2(\frac{1}{n} + \frac{1}{\hat{\sigma}^2}) - n\bar{x}^2},$$

$$\hat{\alpha_1} = \frac{\bar{y} - \bar{x}\sum \frac{x_i y_i}{\sum x_i^2} + b[\frac{\bar{y}}{\alpha} - \frac{n\bar{x}^2}{\alpha}(\alpha y_i - \bar{y})]}{(\hat{\sigma}^2 + \frac{2}{n}) - \frac{(\hat{\sigma}^2 - n\hat{\sigma}^2)(1-n\hat{\sigma}^2)}{n(n+\hat{\sigma^2})} - \frac{\bar{x}\hat{\sigma}^2}{\sum x_i^2}[1 + \frac{1-n\hat{\sigma}^2}{n+\hat{\sigma}^2}] + \frac{n\bar{x}^2\hat{\sigma}^2(1-\frac{1}{n}-1-\frac{1}{n})}{a}\frac{b}{\bar{x}^2 a - n\bar{x}^2}}$$

که در آن $a = 1\hat{\sigma}^{2} + 1/n$ و $b = \frac{\bar{x}}{\sum x_i^{2}} + 1 - \frac{1}{n}$. برآورد واریانس نیز به‌صورت

$$\hat{\sigma^{2}} = \frac{-1 + \sqrt{1 + \frac{4\hat{\alpha_1}^{2}}{n}\sum_{i=1}^{n}(y_i - \hat{\beta_0} - \hat{\beta_1}x_i)^{2}}}{2\hat{\alpha_1}^{2}}$$

به‌دست می‌آید، که به دلیل وابسته بودن معادله‌ها به هم، برآوردها به روش‌های عددی قابل حصول هستند.

برای $\alpha = (\alpha_1, \alpha_2, 0, 0)$ و α_1 و α_2 نامعلوم، ابتدا σ^{2} را معلوم و برای راحتی برابر یک فرض می‌کنیم. بنابراین برآورد پارامترها از معادله‌های زیر به روش‌های عددی به‌دست می‌آیند:

$$\begin{aligned}
\hat{\beta_0} &= \frac{(1-2\hat{\alpha_2})\bar{y} - \hat{\beta_1}\bar{x} - \hat{\alpha_1} + \frac{1}{n}\hat{\alpha_1}(1-2\hat{\alpha_2})}{1 + \frac{1}{n}(1-2\hat{\alpha_2})},\\
\hat{\beta_1} &= (1-2\hat{\alpha_2})\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y}) - \hat{\alpha_1}\frac{1}{n}}{\sum_{i=1}^{n}(x_i-\bar{x})^{2}},\\
\hat{\alpha_1} &= \frac{(1-2\hat{\alpha_2})\bar{y} - \hat{\beta_1}\bar{x} - \hat{\beta_0}(1-(\frac{1-2\hat{\alpha_2}}{n}))}{\frac{2}{n}(1-2\hat{\alpha_2})}\frac{n}{(1-2\hat{\alpha_2})}\\
&+ \sum_{i=1}^{n}(y_i + \frac{\hat{\alpha_1} - \hat{\beta_0} - \hat{\beta_1}x_i}{(1-2\hat{\alpha_2})})^{2} + 2\frac{(1-2\hat{\alpha_2})}{(1-2\hat{\alpha_2})}\sum_{i=1}^{n}(\frac{\hat{\alpha_1} + \hat{\beta_0}\hat{\beta_1}x_i}{(1-2\hat{\alpha_2})^{2}})\\
&\times (y_i + \frac{-\hat{\alpha_1} - \hat{\beta_0} - \hat{\beta_1}x_i}{1-2\hat{\alpha_2}}) = 0.
\end{aligned}$$

اگر σ^{2} نامعلوم باشد، به‌طور مشابه و با معرفی پیشین $\pi(\sigma|\beta_0, \alpha_1, \alpha_2) = 1$ به روش‌های عددی برآوردهای ماکسیمم درستنمایی تعمیم یافته به‌دست می‌آید.

برای $\alpha = (0, 0, \alpha_3, 0)$ و α_3 نامنفی، تابع چگالی موزون با وزن اریب اندازه به‌صورت

$$f^{w}(y; x'\beta, \sigma^{2}) = \frac{y^{\alpha_3}}{\mu_{\alpha_3}}\frac{1}{\sigma\sqrt{2\pi}}e^{\frac{-1}{2\sigma^{2}}(y - x'\beta)^{2}},$$

است، که $E(Y^{\alpha_3}) = \mu_{\alpha_3}$ گشتاور نامرکزی مرتبه α_3 برای مدل رگرسیون نرمال است و در حالت کلی ممکن است به $\mathbf{x}'\beta$ وابسته باشد. واضح است که در این حالت توزیع پایا نیست، بنابراین با روش‌های عددی می‌توان برآوردهای ماکسیمم درستنمایی را به‌دست آورد.

برای $\alpha = (۰, ۰, ۰, \alpha_۴)$ مدل رگرسیونی چوله نرمال حاصل می‌شود که پارامترهای آن توسط سارتوری (۲۰۰۶) برآورد شده است.

۴ مطالعه شبیه‌سازی

برای شبیه سازی مدل رگرسیون موزون تحت وزن اریب اندازه به‌دلیل پیچیدگی شبیه‌سازی از این مدل از نمونه‌گیری با احتمال متناسب با اندازه به روش لاهیری با در نظر گرفتن تابع وزن اریب اندازه به‌عنوان متغیر اندازه استفاده شده است (سمپات، ۲۰۰۵). برای این منظور ابتدا N مقدار x از یک توزیع معلوم مانند نرمال یا پواسن تولید کرده سپس برای هر x، از توزیع نرمال با میانگین معلوم $\beta_۰ + \beta_۱ x$ و انحراف معیار معلوم σ یک مشاهده y تولید می‌شود. این N مقدار تولید شده را به‌عنوان یک جامعه متناهی درنظر گرفته و با $y_۱, \ldots, y_N$ نمایش داده می‌شوند. از این جامعه متناهی به روش لاهیری نمونه‌ای با جایگذاری به حجم n با احتمال متناسب با اندازه (تابع وزن) انتخاب می‌شود. طبق این روش ابتدا متغیر اندازه(w_i)، متناسب با تابع وزن در نظر گرفته می‌شود یعنی $w_i = kw(y_i)$ و ضریب تناسب k طوری انتخاب می‌شود که تمام w_i ها عدد صحیح باشند. ماکسیمم w_i ها با M نشان داده می‌شود. روش لاهیری طی دو گام انجام می‌شود:

گام۱: عدد تصادفی T بین یک تا N و عدد تصادفی دیگر R بین یک تا M انتخاب می‌شود.

گام ۲: اگر $R \leq W_T$ باشد واحد T ام جامعه متناهی، یعنی y_T انتخاب می‌شود **گام‌های ۱ و ۲** آن‌قدر تکرار می‌شوند تا واحدی از جامعه انتخاب شود. ثابت می‌شود که احتمال انتخاب واحد T ام متناسب با w_T است (سمپات، ۲۰۰۵). با این روش می‌توان یک نمونه با جایگذاری به حجم n از جامعه متناهی انتخاب کرد. اگر N به اندازه کافی بزرگ باشد انتظار می‌رود که نمونه حاصل یک نمونه موزون با وزن $w(y)$ از مدل رگرسیونی نرمال باشد.

برای وزن نمایی با توجه به پایا بودن توزیع، در هر بار از مدل رگرسیونی نرمال با میانگین $\beta_۰ + \beta_۱ x$ و انحراف معیار ۲ نمونه‌های موزون به حجم n شبیه‌سازی

شده است. در هر کدام از نمونه‌های موزون شبیه سازی شده، برآوردها به روش ماکسیمم درستنمایی با استفاده از دستور $optim$ در نرم افزار R به‌دست آمده و بعد از B بار تکرار شبیه‌سازی، متوسط و واریانس برآوردها به‌عنوان امید ریاضی و واریانس شبیه سازی برآوردها محاسبه گردید.

نتایج شبیه‌سازی با تکرار $B = ۱۰۰۰۰$ در جدول ۱ خلاصه شده است. برای وزن اریب اندازه مرتبه دو نتایج شبیه‌سازی به‌دلیل استفاده از روش لاهیری و زمان بر بودن اجراها با $B = ۱۰۰۰$ تکرار در جدول ۲ نشان داده شده است. محاسبات با تقریب کمتراز ۰/۰۱ گرد شده‌اند. همان‌گونه که ملاحظه می‌شود برآوردهای پارامترها تقریبا نااریب هستند و واریانس و اریبی برآوردها با افزایش تعداد نمونه‌ها کاهش می‌یابد. همچنین برآوردها نسبت به تغییر مقادیر اولیه حساسیت خاصی نداشته‌اند. لذا در مثال کاربردی برای برآورد پارامترها به روش عددی از مقادیر اولیه متفاوتی استفاده نشد.

۵ مثال کاربردی

در این بخش به بررسی نتایج مربوط به برآورد پارامترهای مدل رگرسیونی خطی ساده تحت نمونه‌گیری تصادفی و موزون با سه تابع وزن برای ۷۹ داده کامل از ۸۴ داده‌ای که توسط کریمی و علوی (۲۰۱۴) استفاده شده‌اند، پرداخته می‌شود. آنها فقط متوسط تعداد نسخه ماهانه ۸۴ پزشک را تحلیل کردند. به‌دلیل عدم ارائه حق‌الزحمه ۵ پزشک از ۷۹ داده کامل استفاده شده است. در این داده‌ها حق‌الزحمه (x) برحسب میلیون تومان به عنوان متغیر مستقل و تعداد نسخه پزشک متخصص در سال (y) به عنوان متغیر وابسته در نظر گرفته‌شده است.

این داده‌ها یک نمونه موزون ۷۹ تایی از ۲۱۶ پزشک متخصص طرف قرارداد سازمان تامین اجتماعی اهواز در سال ۱۳۸۸ هستند که به روش زیر از داروخانه‌ها به‌دست آمده است:
با انتخاب یک نمونه تصادفی از داروخانه‌ها به‌تصادف از هر داروخانه چند بیمار انتخاب شد. سپس پزشکان معالج این بیماران به عنوان پزشکان منتخب از جامعه

جدول ۱: مقادیر اولیه، میانگین و واریانس برآورد پارامترهای دو مدل رگرسیونی تحت وزن نمایی درجه اول

اندازه نمونه	مدل	مقادیر اولیه	میانگین برآورد	واریانس برآورد
n	$\alpha_1, \sigma, \beta_1, \beta_0$	$\alpha_1, \sigma, \beta_1, \beta_0$	$\alpha_1, \sigma, \beta_1, \beta_0$	$\alpha_1, \sigma, \beta_1, \beta_0$
۵۰	۲،۲،۳،۱	۱/۵،۵،۶،۱/۵	۱/۹۹،۱/۹۵،۳،۱/۶۲	۰/۲۴،۰/۰۴،<۰/۰۱،۱/۰۱
۱۰۰	۲ ،۲ ،۳ ، ۱	۱، ۱، ۱، ۱	۲/۰۶، ۱/۹۷، ۳، ۱/۳۰۳	۰/۰۵، ۰/۰۲، <۰/۰۱، ۰/۲۵
۵۰۰	۲ ، ۲ ، ۳ ، ۱	۱/۵ ، ۳ ، ۴ ، ۰/۵	۲/۱۲ ، ۱/۹۹ ، ۳ ، ۰/۵۸	۰/۰۰۲ ، ۰/۰۰۴ ،<۰/۰۱ ، <۰/۰۱
۵۰	۲ ، ۲ ، ۳ ، ۱	۱ ، ۴ ، ۵ ، ۰/۵	۲/۲۶ ، ۱/۹۵ ، ۳ ، ۰/۶۵	۰/۲۴ ، ۰/۰۰۴ ، <۰/۰۱ ، ۰/۰۰۸
۱۰۰	۲ ، ۲ ، ۳ ، ۱	۲/۵ ، ۳ ، ۴ ، ۱/۵	۱/۶۹ ، ۱/۶۷ ، ۳ ، ۱/۴۸	۰/۰۸ ، ۰/۰۲ ، <۰/۰۱ ، ۰/۰۰۲
۵۰۰	۲ ، ۲ ، ۳ ، ۱	۱ ، ۱ ، ۰ ، ۰/۶	۲/۱ ، ۱/۹۹ ، ۳ ، ۰/۸۲	۰/۰۲ ، <۰/۰۱، <۰/۰۱ ، <۰/۰۱
۵۰	۲ ، ۲ ، ۳ ، ۱	۱ ، ۶ ، ۷ ، ۰	۲/۳ ، ۱/۹۵ ، ۳ ، ۰/۵۲	۰/۲۳ ، ۱/۰۴ ، <۰/۰۱ ، ۰/۰۶
۱۰۰	۲ ، ۲ ، ۳ ، ۱	۱ ، ۴ ، ۵ ، ۰/۵	۲/۱۸ ، ۱/۹۸ ، ۳ ، ۰/۶۲	۰/۱۱ ، ۰/۰۲ ، <۰/۰۱ ، ۰/۶۲
۵۰۰	۲ ، ۲ ، ۳ ، ۱	۲/۵ ، ۴ ، ۶ ، ۱/۵	۱/۹ ، ۱/۹۹ ، ۳ ، ۱/۴۹	۰/۰۲ ، <۰/۰۱ ، <۰/۰۱ ، <۰/۰۱
۵۰	۲، ۲ ، −۸ ، ۴	۱ ، ۳ ، ۰ ، ۳	۲/۲۸ ، ۱/۹۵ ، −۸ ، ۳/۵۷	۰/۲۲ ، ۰/۰۴ ، <۰/۰۱ ،۰/۰۷
۱۰۰	۲، ۲ ، −۸ ، ۴	۱ ، ۳ ، ۰ ، ۳	۲/۳ ، ۱/۹۸ ، −۸ ، ۳/۱۳	۰/۱۱ ، ۰/۰۲ ، <۰/۰۱ ،۰/۰۱
۵۰۰	۲، ۲ ، −۸ ، ۴	۳، ۵ ، ۱۰ ، ۴/۵	۱/۸۸ ، ۱/۹۹ ، −۸ ، ۴/۵۵	۰/۰۲ ، <۰/۰۱ ، <۰/۰۱ ، <۰/۰۱ ،
۵۰	۲، ۲ ، −۸ ، ۴	۱ ، ۵ ، ۸ ، ۳/۵	۲/۸۱ ، ۱/۹۵ ، −۸ ، ۳/۹۶	۰/۲۲ ، ۰/۰۴ ، <۰/۰۱ ، ۰/۱۲
۱۰۰	۲، ۲ ، −۸ ، ۴	۱ ، ۵ ، ۸ ، ۳/۵	۲/۱ ، ۱/۹۸ ، −۸ ، ۳/۹۳	۰/۱ ، ۰/۰۲ ، <۰/۰۱ ، ۰/۱۲
۵۰۰	۲، ۲ ، −۸ ، ۴	۱ ، ۵ ، ۸ ، ۳/۵	۲/۱۲، ۱/۹۹ ، −۸ ، ۳/۹۵	۰/۰۲ ، <۰/۰۱، <۰/۰۱ ، <۰/۰۱
۵۰	۲، ۲ ، −۸ ، ۴	۰/۵ ، ۲ ، ۳ ، ۴/۵	۱/۹۷ ،۱/۹۵،−۸، ۴/۷۴	۰/۱۸،۰/۰۴ ،<۰/۰۱، ۰/۰۷
۱۰۰	۲، ۲ ، −۸ ، ۴	۰/۵ ، ۲ ، ۳ ، ۴/۵	۱/۹۱، ۱/۹۸،−۸، ۴/۶۳	۰/۰۹،۰/۰۲، <۰/۰۱،<۰/۰۱
۵۰۰	۲، ۲ ، −۸ ، ۴	۱، ۱، ۱، ۶	۲/۱۵ ، ۲/۰۰، −۸، ۳/۴۴	۰/۰۲، <۰/۰۱، <۰/۰۱،۰/۲۳

نماد« <۰/۰۱» در جدول بیانگر مقدار کمتر از ۰/۰۱ است.

جدول ۲: مقادیر اولیه، میانگین و واریانس برآوردپارامترهای دو مدل رگرسیونی تحت وزن اریب اندازه مرتبه دوم

اندازه نمونه	مدل	مقادیر اولیه	میانگین برآورد	واریانس برآورد
n	σ, β_1, β_0	σ, β_1, β_0	σ, β_1, β_0	σ, β_1, β_0
۵۰	۲،۲،۱	۱/۹،۱/۹،۰/۹	۱/۹۵، ۲،۰/۹۹	۰/۰۲، <۰/۰۱،۳/۲۳
۱۰۰	۲،۲،۱	۲/۱ ، ۲/۱ ، ۱/۱	۱/۹۸، ۲، ۱/۰۱	۰/۰۲، <۰/۰۱، ۱/۵۵
۵۰۰	۲،۲،۱	۳ ، ۵ ، ۲	۲، ۲ ، ۰/۹۸	<۰/۰۱ ،<۰/۰۱ ، ۰/۳۶
۵۰	۲،۲،۱	۱/۵ ، ۱ ، ۰/۷	۱/۹۵ ، ۲ ، ۰/۹۷	۰/۰۴ ، <۰/۰۱ ، ۳/۱۹
۱۰۰	۲،۲،۱	۲/۳ ، ۴ ، ۱/۳	۱/۹۷، ۲ ، ۱/۰۲	۰/۰۲ ، <۰/۰۱ ،۱/۴۹
۵۰۰	۲،۲،۱	۲/۲، ۳ ، ۱/۵	۱/۹۹ ، ۲ ، ۱/۰۱	<۰/۰۱ ، <۰/۰۱ ، ۰/۳۷
۵۰	۲،۲،۱	۲/۵، ۳/۳ ، ۱/۴	۲/۰۱، ۲ ، ۰/۶۴	۰/۰۵، <۰/۰۱ ، ۰/۱۳
۱۰۰	۲،۲،۱	۴ ، ۵ ، ۲	۱/۹۷ ، ۲ ، ۱/۰۳	<۰/۰۱، <۰/۰۱ ، ۱/۴۴
۵۰۰	۲،۲،۱	۰/۱، −۱ ، ۰/۵	۱/۹۹ ، ۲ ، ۰/۹۸	<۰/۰۱ ، <۰/۰۱ ، ۰/۳۶
۵۰	۴ ، ۴ ،−۴	۴/۱، ۵ ، −۴/۱	۳/۸۸ ، ۴ ، −۳/۷۴	۰/۱۵، <۰/۰۱ ،۱۲/۹۶
۱۰۰	۴ ، ۴ ،−۴	۴/۲، ۴/۵ ، −۳/۹	۳/۹۴ ، ۴ ، −۴	۰/۰۸، <۰/۰۱ ،۶/۴۶
۵۰۰	۴ ، ۴ ،−۴	۴/۳، ۵ ، −۴/۳	۳/۹۹ ، ۴ ، −۴	۰/۰۲ ، <۰/۰۱ ، ۱/۴۶،
۵۰	۴ ، ۴ ،−۴	۴/۳، ۳ ، −۴/۳	۳/۸۸، ۴ ، −۳/۹۹	۰/۱۵، <۰/۰۱ ، ۱۳/۰۲
۱۰۰	۴ ، ۴ ،−۴	۴/۷، ۳/۵ ، −۳/۷	۳/۹۵، ۴ ، −۳/۹۲	۰/۰۸، <۰/۰۱ ، ۵/۳۸
۵۰۰	۴ ، ۴ ،−۴	۴/۴ ، ۵/۷ ، −۴/۴	۳/۹۸ ،۴ ، −۳/۹۴	۰/۰۲ ، <۰/۰۱ ، ۱/۳۸
۵۰	۴ ، ۴ ،−۴	۳، ۲ ، −۳/۵	۳/۸۹،۴، −۳/۹۹	۰/۱۵،<۰/۰۱، ۱۳/۵۹
۱۰۰	۴ ، ۴ ،−۴	۵ ، ۲/۵ ، ۴/۸	۳/۹۵،۴، −۳/۸۷	۰/۰۸، <۰/۰۱،۶/۰۶
۵۰۰	۴ ، ۴ ،−۴	۳/۵، ۳، −۳	۳/۹۹، ۴، −۳/۹۶	۰/۰۲، <۰/۰۱،۱/۳۵

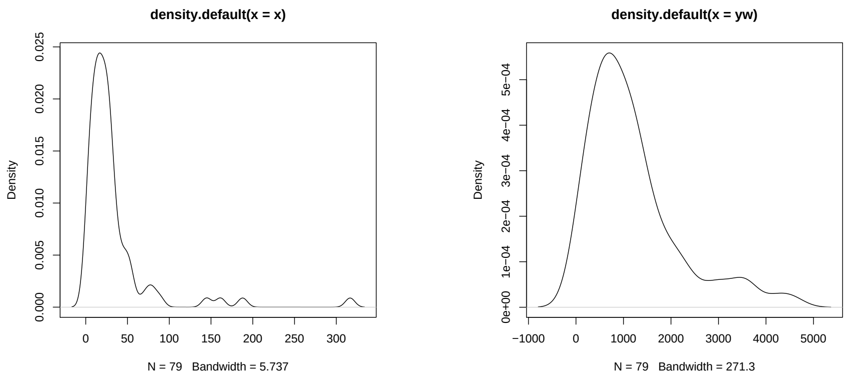
نماد« <۰/۰۱» در جدول بیانگر مقدار کمتر از ۰/۰۱ است.

درنظر گرفته شدند و اطلاعات آنها از سازمان تامین اجتماعی ثبت گردید. باتوجه‌به اینکه پزشکی که تعداد بیمار بیشتری را معاینه می‌کند شانس بیشتری برای انتخاب در نمونه دارد، می‌توان تعداد نسخه پزشکان منتخب به این روش را یک نمونه موزون در نظر گرفت. این داده‌ها درجدول ۳ آمده است. خلاصه‌ای از آن‌ها در جدول ۴ ارائه شده است. همان‌طورکه در شکل ۳ ملاحظه می‌شود، نمودار چگالی‌ها چوله به راست بودن توزیع داده‌ها را نشان می‌دهند.

جدول ۳: داده‌های پزشکان متخصص طرف قرارداد سازمان تامین اجتماعی اهواز

y	x	y	x	y	x
۹۷	۱۹۰۱۲۰۲	۷۲	۱۴۸۹۶۰۰۱	۱۱۹	۱۵۶۴۲۰۰
۲۳۷	۵۵۸۱۱۰۰	۲۱۴	۱۴۹۷۳۵۰۰۱	۱۳۸۱	۳۰۹۶۲۹۶۱
۴۱۱	۹۵۷۴۶۰۰	۳۱۰	۷۰۱۱۹۰۰۱۱	۲۸۳	۶۳۹۴۵۰۰۱
۴۹۲	۱۱۲۳۵۷۰۰	۴۴۹	۱۱۹۲۴۰۷۰۰	۴۴۸	۱۰۲۰۱۸۰۰
۵۷۹	۱۳۰۳۸۹۰۰	۵۷۵	۸۷۰۲۴۰۰۱۱	۵۱۶	۱۱۸۸۷۷۵۰
۶۵۳	۱۵۱۹۴۹۰۰	۶۴۶	۱۱۴۷۶۳۷۰۰۱	۶۰۸	۱۳۹۳۰۷۰۰
۷۰۳	۱۶۲۵۳۳۰۰	۶۷۲	۱۱۵۲۸۸۰۰۱	۶۶۱	۱۴۹۸۹۱۰۰۱
۷۲۴	۳۰۳۳۹۵۴۰	۷۱۹	۱۱۶۷۵۳۱۰۰۱	۷۱۰	۱۶۲۸۷۶۰۰
۸۱۸	۱۸۶۳۴۷۰۰	۷۹۰	۱۱۱۷۹۶۳۴۰۰	۷۶۶	۱۶۷۳۱۹۸۸
۹۲۴	۲۲۲۳۶۶۰۰	۹۰۱	۱۱۱۹۹۸۷۴۳۶	۸۶۸	۱۹۶۹۴۵۵۶
۱۰۹۷	۲۵۰۸۳۱۰۰	۹۹۵	۲۲۶۹۱۹۰۰	۹۸۸	۱۲۲۱۶۲۷۰۰۱
۱۲۳۴	۲۸۱۱۱۳۰۰	۱۲۰۶	۲۷۲۴۸۰۶۰	۱۱۰۸	۲۵۲۷۴۲۰۰
۱۴۷۰	۳۳۸۲۴۷۰۰	۱۲۷۰	۲۹۰۰۸۰۰۰	۱۲۴۰	۲۸۱۹۹۵۰۰
۱۷۰۵	۳۸۹۰۱۱۰۰	۱۵۲۸	۴۱۳۱۸۹۰۰	۱۵۵۹	۲۵۷۴۰۶۰۰
۱۹۸۲	۴۵۳۹۳۶۰۰	۱۸۲۵	۴۱۴۵۵۰۰	۱۷۵۰	۴۰۰۸۶۹۰۰
۲۳۲۱	۵۲۸۳۵۲۴۴	۲۱۸۷	۴۹۵۶۲۳۲۴	۲۰۶۲	۵۱۵۷۱۷۹۸
۲۹۷۴	۶۹۳۳۰۱۰۰	۳۵۷۳	۷۶۸۱۲۴۰۰	۲۴۴۷	۵۵۸۵۵۱۰۰
۱۳۳۹	۳۰۴۰۴۵۰۰	۲۸۲۳	۳۱۶۵۷۹۹۰۰	۲۱۰۵	۴۸۲۵۰۳۰۰
۱۳۸۵	۳۱۶۳۴۴۰۰	۱۱۰۱	۲۵۳۵۸۳۴۰	۱۲۵۸	۲۸۶۴۰۵۰۰
۱۰۴۹	۲۴۱۰۸۰۰۰	۱۰۸۲	۲۵۰۸۳۱۰۰	۱۱۱۰	۲۵۵۳۸۸۰۰
۱۳۰۱	۲۹۹۱۹۴۰۰	۱۰۵۹	۲۴۱۲۲۷۰۰	۱۲۴۸	۲۹۹۹۵۴۲۰
۳۴۸۲	۷۹۶۶۹۱۰۰	۱۵۸۶	۳۶۷۱۸۸۵۰	۱۴۶۰	۳۳۲۴۱۶۰۰
۴۱۹۹	۱۸۷۸۳۱۷۰۰	۳۱۳۲	۱۴۴۷۶۰۲۵۲	۳۶۱۹	۱۶۱۳۴۵۹۴۰
۷۴۸	۱۷۴۲۹۳۰۰	۷۴۴	۱۷۰۴۷۱۰۰	۴۵۵۳	۸۸۸۴۱۰۸۸
۲۶۰	۶۰۶۱۳۰۰	۱۶۵	۳۹۷۸۸۰۰	۶۱	۱۹۵۴۱۲۰۰
۴۲۷	۹۶۹۷۱۰۰	۳۸۵	۷۸۰۲۴۰۰	۲۷۰	۶۲۰۸۳۰۰
				۵۱۸	۱۲۰۵۴۰۰۰

معمولاً برای همگرایی سریع‌تر، مقادیر اولیه پارامترها را می‌توان برآوردهای کمترین توان دوم که نیازی به توزیع ندارند انتخاب کرد (یعنی با فرض تصادفی



(الف) (ب)

شکل ۱: نمودار چگالی الف: تعداد نسخه‌پزشکان و ب: حق الزحمه پزشکان

بودن نمونه، به روش معمول ضرایب رگرسیونی برآورد شدند و از آنها به‌عنوان مقادیر اولیه در مدل‌های با نمونه‌گیری موزون استفاده شد. مقادیر اولیه پارامترهای وزن نیز صفر درنظر گرفته شدند) بنابه خاصیت مجانبی برآوردهای ماکسیمم درستنمایی، توزیع مجانبی آنها نرمال است لذا با تقسیم هر برآورد بر انحراف معیار آن آماره Z برای هر ضریب رگرسیونی (که توزیع تقریبی نرمال استاندارد دارد) محاسبه گردیدند. پارامترهای برآورد شده مدل رگرسیون ساده تحت نمونه‌گیری تصادفی و موزون با وزن‌های چوله، اریب اندازه و نمایی در جدول ۵ آمده است. با توجه به مقادیر آکاییک AIC یا BIC مناسب‌ترین وزن نمونه‌گیری، وزن چوله است.

جدول ۴: آماره های توصیفی حق‌الزحمه و تعداد نسخه‌های پزشکان

متغیر	اندازه نمونه	میانگین	انحراف معیار	دامنه داده‌ها
x	۷۹	۳۳/۴	۴۵/۸۹۰۶۵	۳۱۶/۰۱۵۷
y	۷۹	۱۲۰۴	۹۸۵/۵۵۳۸	۵۴۳۴

جدول ۵: برآورد پارامترهای مدل رگرسیونی تحت نمونه‌گیری تصادفی و موزون

نمونه‌گیری	تصادفی	اریب‌اندازه	موزون نمایی	موزون چوله
برآورد β_0	۶۶۶/۷	۶۶۶/۸	۵۱۵/۶	۲۷۸/۸
آماره آزمون Z	۷/۲۷	۷/۳۶	۵/۹۳	۳/۰۳
p- مقدار	$<$ ۰/۰۰۱	$<$ ۰/۰۰۱	$<$ ۰/۰۰۱	۰/۰۰۳
برآورد β_1	۱۶/۰۸	۱۶/۰۸	۱۶/۰۸	$<$ ۰/۰۰۱
آماره آزمون Z	۹/۹۱	۲۰/۷۹	۹/۳۱	۹/۸۸
p- مقدار	$<$ ۰/۰۰۱	$<$ ۰/۰۰۱	$<$ ۰/۰۰۱	$<$ ۰/۰۰۱
برآورد واریانس مدل	۴۳۲۴۸۲	۶۴۹/۶	۷۰۰/۳	۷۷۱/۵
برآورد پارامتر وزن	---	۰/۰۰۰۲	۰/۰۰۰۳	۰/۹۶
معیار AIC	۱۲۵۳/۴	۱۲۵۵/۴	۱۲۵۶/۲	۱۲۱۶/۴
معیار BIC	۱۲۶۰/۵	۱۲۶۴/۵	۱۲۶۵/۷	۱۲۲۵/۹

بحث و نتیجه‌گیری

در این مقاله برآورد پارامترهای مدل رگرسیونی خطی تحت نمونه‌های موزون با یک وزن جدید بررسی شد. تابع وزن نمایی، اریب‌اندازه و چوله حالات خاصی از این وزن پیشنهادی بودند. برآورد پارامترهای مدل رگرسیونی نرمال تحت نمونه‌های موزون اریب اندازه و نمایی مطالعه گردید. در مورد وزن نمایی، با فرض معلوم بودن پارامتر پراکندگی، شکل بسته‌ای برای پارامترها به روش ماکسیمم درستنمایی تعمیم یافته برآورد شد. برآورد پارامترهای رگرسیونی در دیگر موارد به روش‌های عددی به‌دست آمدند که با مطالعه شبیه‌سازی نشان داده شد که این برآوردها تقریبا نااریب و سازگار هستند. داده‌های مربوط به حق‌الزحمه و تعداد نسخه پزشکان متخصص طرف قرارداد سازمان تامین اجتماعی شهر اهواز در سال ۸۸ به روش نمونه‌گیری موزون تحلیل شدند و نتایج نشان می‌دهد که مناسب‌ترین وزن برای داده‌ها وزن چوله با پارامتر چولگی ۰/۹۶ است.

تشکر و قدردانی

نویسندگان مقاله نهایت تشکر و قدردانی را از داوران و سردبیر محترم نشریه، که با رهنمودهای ارزنده خود باعث بهتر شدن مقاله گردیده‌اند، دارند.

مراجع

Alavi, S. M. R. (2008), *Statistical Inference for Weighted Disrtibution*, Ph.D. Thesis of Shahid Chamran University of Ahvaz, Ahvaz, Iran.

Alavi, S. M. R. (2017), A Generalized Class form-invariant Bivariat Weighted Distributions, *Communications in statistics-Theory and Methods*, **46**, 2193-2201.

Alavi, S. M. R. and Chinipardaz, R. (2009), Form-Invariance under Weighted Sampling, *Statistics*, **43**, 81-90.

Casella, G. and Berger, R. L. (2002), *Statistical Inference*, Duxbury Press, USA.

Chakrobarty, S. (2010), On some Distributional Properties of the Family of Weighted Generalized Poisson Distribution, *Communications in statistics-Theory and Method*, **39**, 2767-2788.

Chen, Y. Q. (2009), Semiparametric Regression in Size-Biazed Sampling, *Biometrika*, **66**, 149-158.

Cristobal, J. and Alcala, J. (2000), Nonparametric Regression Estimatore for Length Biased Data, *Journal of Statistical Planning and Inference*, **89** , 145-168.

Gupta, R. D. and Kundu, D. (2009), A New Class of Weighted Distributions, *Statistics*, **43**, 621-634.

Karimi, M. and Alavi, S. M. R. (2014), The Effect of Weight Function on Hypothesis Testing in Weighted Sampling, *Journal of Applied Statistics*, **41**, 2493-2503.

Patil, G. P. (2002), Weighted Distributions, *Encyclopedia of Environmetrics*, **4**, 2369-2377.

Rao, C. R. (1965), On discrete Distributions arising out of Methods of Ascertainment, *Sankhya*,**27**, 311-324.

Sartori, N. (2006), Bias Privention of Maximum Liklihood Estimates for ScalarSkew Normal and Skew T Distributions, *Journal of Statistical Planning and Inference*, **136**, 4259-4275.

Sampath, S. (2005), *Sampling Theory and Methods*, Alpha ScienceInternational, Harrow, England.

Zadkarami, M. R. (2008), On Indentifibility in Weighted Aistributions Using Generalized Maximum Liklihood Estimations, *Journal of the Iranian Statistical Society*, **7**, 73-8.