سیزدهمین کنفرانس آمار ایران
۴-۲ شهریور ۱۳۹۵

کالیبدن رگرسیون در حضور خطای اندازه‌گیری برای تحلیل داده‌های شکل

نقی همتی[1] موسی گل‌علیزاده
گروه آمار دانشگاه تربیت مدرس

چکیده:داده‌های شکل با توجه به تعدد منابع خطاها اغلب در معرض ابتلا به خطای اندازه‌گیری قرار دارند. مسئله‌ی دخالت خطای اندازه‌گیری در تحلیل شکل بویژه در انطباق پروکراستس اشکال از تحقیقات نو در آمارشکل است. در نگاه سنتی به تحلیل آماری شکل معمولا خطای اندازه‌گیری در نظر گرفته نمی‌شود که نادیده گرفتن آن در صورت وجود، باعث بروز مشکلات زیادی از قبیل اریبی برآوردگرها می‌شود. در این حالت برآوردگرهای حاصل از عدم دخالت خطای اندازه‌گیری برآوردگرهای ناپخته نامیده می‌شوند. ثابت شده است که در انطباق پروکراستس داده‌های شکل، برآوردگر ناپخته برای پارامترهای مقیاس و دوران اریب است. در مقاله حاضر برای تصحیح اریبی و بهبود دقت برآوردگرهای ناپخته، روش کالیبدن رگرسیون پیشنهاد و خواص مفید آن ارائه شده است. به‌علاوه، با انجام مطالعات شبیه‌سازی عملکرد برآوردگرهای ناپخته و کالیبدن رگرسیون مورد مقایسه قرار گرفته است.

واژه‌های کلیدی:تحلیل پروکراستس، خطای اندازه‌گیری، توزیع نرمال مختلط، روش کالیبدن رگرسیون.
کد موضوع‌بندی ریاضی (۲۰۱۰): $62H99$

۱ مقدمه

تحلیل آماری شکل شاخه جدیدی از آمار چند متغیره است که در بسیاری از علوم کاربردی مانند زیست‌شناسی، پزشکی، اخترشناسی و زمین‌شناسی مورد توجه قرار گرفته است. امروزه با توسعه سریع تکنولوژی دسترسی به اطلاعات هندسی اشیاء آسان و تحلیل شکل آن‌ها بسیار مهم شده است. بنا به کندال (۱۹۸۴)، لغت "شکل" در ادبیات شکل به عنوان تمام اطلاعات هندسی‌ای که بعد از حذف مکان، مقیاس و دوران از یک شیء باقی می‌ماند، تعریف می‌شود. در اینجا مطالعات اشکال (پیکربندی‌های هندسی[2]) در نقاط شاخص[3] خلاصه و با ماتریس ارائه می‌شوند. به عنوان مثال پیکربندی هندسی در $\mathbb{R}^m$ با L نقطه‌ی شاخص با ماتریس $\mathbf{X}_{L\times m}$ نمایش داده می‌شود.

---

[1] نقی همتی: naghi.hemmati@modares.ac.ir

A-10-178-1

[2] Geometrical configurations

[3] Landmark

تعریف یک فاصله‌ی مناسب بین دو یا چند شکل برای محاسبه‌ی میانگین و بررسی تغییرات شکل دارای اهمیت زیادی است. روش‌هایی که در آمار اقلیدسی برای تحلیل‌های آماری استفاده می‌شوند بطور مستقیم در آمار شکل کمتر مورد استفاده قرار می‌گیرند. لذا برای تحلیل آماری اشکال به روش‌هایی نیاز است که تعمیمی از روش‌های سنتی آمار چند متغیره هستند. یکی از روش‌هایی که در آمار شکل زیاد مورد استفاده قرار می‌گیرد، روش پروکراستس است. این روش که براساس روش حداقل مربعات برای انطباق اشکال بنا شده است، از مفیدترین ابزارهای علمی در تحلیل آماری شکل با نقاط شاخص است. جزئیاتی از این روش در ماردیا و همکاران (۱۹۷۹) موجود است. توجه کنید که تحلیل پروکراستس عادی[۳] (OPA) برای انطباق دو پیکربندی مورد استفاده قرار می‌گیرد. براساس تعریف انطباق پروکراستس عادی، یک رابطه بین روش‌های پروکراستس و مدل رگرسیونی خطی ساده وجود دارد. محاسبه‌ی یک فاصله غیر اقلیدسی بین دو شکل $\mathbf{X}$ و $\mathbf{Y}$ با استفاده از تحلیل پروکراستس می‌تواند با مدل خطی در قالب نمادگذاری با اعداد مختلط فرمول‌بندی شود (درایدن و ماردیا، ۱۹۹۸).

از طرفی باید توجه داشت که اطلاعات شکل با توجه به تعدد منابع خطا می‌تواند به راحتی در معرض ابتلا به خطای اندازه‌گیری قرار گیرد. به عنوان مثال در عکسبرداری‌های پزشکی، منابع خطای ممکن به جز خطای تصادفی ذاتی، شامل خطاهای ناشی از دستگاه اسکن و تنوع بین اپراتورها خواهد بود. با این حال معمولا خطای اندازه‌گیری در تحلیل‌های سنتی آمار شکل در نظر گرفته نمی‌شود. بنا به فولر (۱۹۸۷) در مطالعات گذشته راجع به رگرسیون خطی معمولی نشان داده شد که نادیده گرفتن خطای اندازه‌گیری می‌تواند استنباط آماری مربوطه را خدشه‌دار کند. لذا، برآوردگری که از نادیده گرفتن خطای اندازه‌گیری بدست می‌آید، برآوردگر ناپخته[۵] نامیده شده است. این برآوردگر اریب است و لذا تصحیح اریبی رویکردی است که به عنوان هدف در مطالعه رگرسیون با حضور خطای اندازه‌گیری مدنظر قرار می‌گیرد. بیان این نکته ضروری است که اکثر محققینی که راجع به اریبی برآوردگر ناپخته تحقیقاتی داشته‌اند مانند گلسر (۱۹۸۱)، کارول و گالو (۱۹۸۲) و استفانسکی (۱۹۸۵) توجه خود را تنها معطوف به مدل‌هایی با متغیرهای تصادفی حقیقی‌مقدار کرده‌اند. با این حال، بدلیل ماهیت متغیرهای تصادفی مورد مطالعه در آمارشکل آنها می‌توانند مختلط نیز باشند. لذا، بررسی تاثیر خطای اندازه‌گیری در آمارشکل که با اعدا مختلط توصیف می‌شوند (داده‌های دوبعدی شکل) و همچنین ارزیابی روش تصحیح اریبی کالیبدن رگرسیون در این حوزه مدنظر این مقاله قرار گرفته است. تقریب کالیبدن رگرسیون[۶] یکی از روش‌های تصحیح اریبی برای مدل رگرسیونی تحت تأثیر خطای اندازه‌گیری است که برای تصحیح اریبی از الگوریتم کالیبدن استفاده می‌کند. این الگوریتم به عنوان یک دیدگاه کلی توسط کارول و استفانسکی (۱۹۹۰) پیشنهاد شد.

ادامه مقاله حاضر به صورت زیر تدوین شده است. در بخش ۲ مدل خطای اندازه‌گیری برای داده‌های شکل معرفی و اریبی برآوردگر ناپخته بررسی می‌شود. در بخش ۳ روش کالیبدن رگرسیون برای برآورد پارامترهای مدل رگرسیونی مختلط در حضور خطای اندازه‌گیری ارائه می‌گردد. ارزیابی روش کالیبدن رگرسیون در قالب مطالعات شبیه‌سازی در بخش ۴ ارائه می‌شود.

۲ مدل خطای اندازه‌گیری برای داده‌های شکل دو بعدی

برای یک پیکربندی دوبعدی، می‌توان مکان نقاط شاخص را با یک عدد مختلط نشان داد. فرض کنید X و Y دو پیکربندی مورد علاقه، هرکدام شامل L (≥ ۳) نقطه‌ی شاخص باشد. واضح است که با نمایش عدد مختلط، $\mathbf{X}$ و $\mathbf{Y}$ عناصری L بعدی درفضای

---

[۳] Ordinary Procrustes Analysis

[۵] Naive

[۶] Regression calibration

مختلط $\mathbf{C}^L$ هستند. بنابراین می‌توان از توزیع‌های مختلط برای مدل‌بندی پیکربندی‌های دو بعدی استفاده کرد. بنا به گوودمن (۱۹۶۳) یکی از مهمترین توزیع‌های آماری در این حالت توزیع نرمال مختلط است. خانواده متغیرهای تصادفی نرمال مختلط از متغیرهای تصادفی مختلطی تشکیل شده است که بخش‌های حقیقی و مجازی آنها تواماً نرمال هستند. فرض کنید $\mathbf{U} = (u_1, \ldots, u_L)^T$ و $\mathbf{V} = (v_1, \ldots, v_L)^T$ بردارهای تصادفی در $\mathbf{R}^L$ هستند طوری که $(\mathbf{U}, \mathbf{V})^T = (u_1, \ldots, u_L, v_1, \ldots, v_L)^T$ یک بردار تصادفی نرمال $2L$ بعدی است. آنگاه بردار تصادفی مختلط $\mathbf{X} = \mathbf{U} + i\mathbf{V}$ از توزیع نرمال مختلط پیروی می‌کند. معمولا کلمه نرمال مختلط با نماد اختصاری CN نوشته شده و پیروی $\mathbf{X}$ از توزیع CN برای حالتی که بخش‌های حقیقی و مجازی متغیر مختلط $\mathbf{X}$ مستقل از هم هستند به صورت

$$\mathbf{X} \sim \mathrm{CN}(\mu, 2\sigma^2)$$

نمایش داده می‌شود که در آن μ میانگین متغیر تصادفی مختلط $\mathbf{X}$ و $2\sigma^2$ واریانس این متغیر است. توجه شود که در اینجا واریانس هر کدام از بخش‌های متغیر تصادفی مختلط $\mathbf{X}$ برابر σ^2 در نظر گرفته شده است.

برای انطباق $\mathbf{X}$ بر روی $\mathbf{Y}$ از طریق OPA، مدل خطی

$$Y = \beta_0 1_L + \beta_1 X + \epsilon \qquad (۱.۲)$$

مد نظر قرار می‌گیرد که در آن β_0 پارامتر انتقال، β_1 پارامتر مقیاس و دوران، 1_L برداری $L \times 1$ از یک‌ها و ϵ خطای تصادفی با میانگین صفر است. بر اساس مدل ارائه شده در (۱.۲)، انطباق X بر روی Y می‌تواند به عنوان یک مسئله‌ی حداقل مربعات فرمول‌بندی شود (درایدن و ماردیا، ۱۹۹۸). توجه کنید که $\mathbf{X}$ و $\mathbf{Y}$ هر دو به عنوان داده‌های مشاهده شده در نظر گرفته می‌شوند. اما مدل (۱.۲) فقط عبارت خطای تصادفی ϵ برای $\mathbf{Y}$ را شامل می‌شود و خطای اندازه‌گیری ممکن در $\mathbf{X}$ را نادیده می‌گیرد. در حضور خطای اندازه‌گیری برای پیکربندی $\mathbf{X}$، مسئله انطباق دو پیکربندی می‌تواند به صورت مدل خطی مختلط با خطای اندازه‌گیری به ازای $l = 1, \cdots, L$ به صورت

$$\begin{cases} Y_l = \beta_0 + \beta_1 X_l + \epsilon_l \\ W_l = X_l + U_l \end{cases} \qquad (۲.۲)$$

نوشته شود که در آن $\mathbf{X} = (X_1, \ldots, X_L)^T$ و $\mathbf{Y} = (Y_1, \ldots, Y_L)^T$ متعلق به فضای مختلط $\mathbf{C}^L$ هستند.

توجه کنید که $\mathbf{W} = (W_1, \ldots, W_L)^T \in \mathbf{C}^L$ مشاهده آلوده به خطای $\mathbf{X}$، $\mathbf{U} = (U_1, \ldots, U_L)^T \in \mathbf{C}^L$ خطای اندازه‌گیری و $\epsilon = (\epsilon_1, \ldots, \epsilon_L)^T \in \mathbf{C}^L$ عبارت خطای تصادفی برای $\mathbf{Y}$ است. فرض می‌شود متغیرهای $\mathbf{X}$، $\mathbf{U}$ و ϵ از یکدیگر مستقل‌اند و به ازای $l = 1, \cdots, L$، $\epsilon_l \sim CN(0, 2\sigma_\epsilon^2)$ و $U_l \sim CN(0, 2\sigma_u^2)$. همچنین برای سادگی فرض می‌شود $Re(U_l)$ و $Im(U_l)$ ناهمبسته و U_lها مستقل‌اند. علاوه بر این، $\Theta = (\beta, \sigma_\epsilon^2)^T$ پارامترهای نامعلوم مدل هستند که در آن $\beta = (\beta_0, \beta_1)^T$.

۱.۲ برآوردگر ناپخته

در این زیربخش، خواص برآوردگر ناپخته β حاصل از نادیده گرفتن خطای اندازه‌گیری تحت مدل (۲.۲) مطالعه می‌شود. از انطباق $\mathbf{X}$ روی $\mathbf{Y}$ با روش OPA، برآوردگر β به صورت $\widehat{\beta} = (\widehat{\beta_0}, \widehat{\beta_1})^T = (\mathbf{X}_D^* \mathbf{X}_D)^{-1} \mathbf{X}_D^* \mathbf{Y}$ بدست می‌آید که در آن $\mathbf{X}_D = \begin{bmatrix} 1_L & \mathbf{X} \end{bmatrix}$ ماتریس طرح مختلط $L \times 2$ و $\mathbf{X}_D^*$ ترانهاده مزدوج مختلط $\mathbf{X}_D$ است. در اکثر موارد پیکربندی واقعی $\mathbf{X}$ مشاهده نمی‌شود و به جای

آن پیکربندی W مشاهده می‌شود. به زبانی دقیق‌تر می‌توان نوشت: $\mathbf{W} = \mathbf{X} + \mathbf{U}$. حال تحت مدل (۲.۲) برآوردگر ناپخته‌ی β برای پیکربندهای مشاهده شده به صورت

$$\widehat{\beta}_{naive} = (\widehat{\beta}_{0,W}, \widehat{\beta}_{1,W})^T = (\mathbf{W}_D^* \mathbf{W}_D)^{-1} \mathbf{W}_D^* \mathbf{Y} \qquad (۳.۲)$$

است که در آن $\mathbf{W}_D$ ماتریس طرح برای پیکربندی مشاهده شده است.

برای مطالعه‌ی اثرات خطای اندازگیری در روش OPA، فرض می‌شود نقاط شاخصی از توزیع نرمال مختلط برای $\mathbf{X}$ به ازای $l = 1, \cdots, L$ وجود دارد طوری که

$$X_l \sim CN(\mu_x, 2\sigma_x^2).$$

هم‌چنین فرض می‌شود متغیرهای تصادفی مختلط نرمال $\{\mathbf{X}, \mathbf{U}, \mathbf{W}\}$ دو به دو مستقل هستند. با توجه به این اطلاعات، لم ۱.۲ که از در و همکاران (۲۰۱۴) اقتباس شده، ارائه می‌شود.

لم ۱.۲. تحت فرض های نرمال مختلط برای X_l، U_l و ϵ_l به ازای $l = 1, \cdots, L$ اگر $\mu_X = (\mu_1, \cdots, \mu_L)^T = \mu 1_L$، آنگاه

$$E(\widehat{\beta}_{0,naive}) = \beta_0 + (1 - \frac{\sigma_x^2}{\sigma_x^2 + \sigma_u^2})\mu\beta_1 = \beta_0 + (1 - \lambda)\mu\beta_1,$$

$$E(\widehat{\beta}_{1,naive}) = \frac{\sigma_x^2}{\sigma_x^2 + \sigma_u^2}\beta_1 = \lambda\beta_1$$

به کمیت $\frac{\sigma_x^2}{\sigma_x^2+\sigma_u^2}$ نرخ قابلیت اعتماد[۷] می‌گویند و با λ نشان می‌دهند. اگر نرخ قابلیت اعتماد بزرگ (نزدیک یک) باشد خطای اندازگیری کوچک و اریبی برآوردگر ناپخته در قدر مطلق کوچک است. همچنین اگر μ_l به ازای تمام مولفه‌ها برابر صفر باشد آنگاه $E\left(\hat{\beta}_{0,naive}\right) = \beta_0$ که نشان دهنده‌ی نااریبی $\hat{\beta}_{0,naive}$ تحت فرض صفر بودن میانگین X است. چون $\lambda \in [0, 1]$، تساوی دوم در لم (۱.۲) یک اثر کاهشی از خطای اندازگیری را روی برآوردگر β_1 نشان می‌دهد. در انطباق دو پیکربندی، این اثر کاهشی به معنی کم برآورده کردن پارامتر مقیاس است.

۲.۲ روش کالیبدن رگرسیون برای داده‌های شکل دو بعدی

کالیبدن رگرسیون یک روش تقریبی است که می‌تواند برای تصحیح اریبی خطای اندازگیری مورد استفاده قرار گیرد. این روش ساده و کارا برای هر مدل رگرسیونی به اندازه کافی تقریب دقیقی را فراهم می‌کند (کارول و همکاران، ۲۰۰۶). در ادامه نحوه اعمال این روش برای مدل رگرسیونی مختلط تشریح می‌شود.

فرض کنید هدف مطالعه این است که پیکربندی $\mathbf{X}$ از طریق مدل رگرسیونی مختلط

$$\mathbf{Y} = \beta_0 + \beta_1 \mathbf{X} + \epsilon$$

روی پیکربندی $\mathbf{Y}$ منطبق شود که در آن β_0 و β_1 پارامترهای مدل هستند. در اکثر موارد علمی بنا به دلایلی پیکربندی درست $\mathbf{X}$ مشاهده نمی‌شود ولی به جای آن پیکربندی $\mathbf{W}$ مشاهده می‌شود به قسمی که $\mathbf{W} = \mathbf{X} + \mathbf{U}$. مدل رگرسیونی براساس پیکربندی‌های مشاهده شده $(\mathbf{Y}, \mathbf{W})^T$ به مدل رگرسیونی ناپخته معروف است و ارتباط بین متغیرها معمولا به صورت

$$\mathbf{Y} = \beta_0 + \beta_1 \mathbf{W} + \epsilon$$

[۷] Reliability rate

نمایش داده می‌شود که برآوردهای حاصل از آن به برآوردهای اریب منجر خواهد شد. در اینصورت برای تصحیح اریبی می‌توان الگوریتم کالیبدن رگرسیون را به صورت زیر بکار برد:

الف) رگرسیون مختلط پیکربندی $\mathbf{X}$ روی پیکربندی $\mathbf{W}$ انجام گیرد.

ب) پیکربندی $\mathbf{X}$ برآورد شده از مدل رگرسیونی مختلط به جای پیکربندی $\mathbf{X}$ مشاهده نشده جایگذاری شود و تحلیل استاندارد برای دستیابی به برآورد پارامترها صورت پذیرد.

ج) نتایج خطای استاندارد به منظور دستیابی به برآورد پیکربندی در مرحله الف تصحیح شود.

حال با فرض اینکه واریانس خطای اندازگیری (σ_u^2) معلوم است، تابع کالیبدن یا بهترین تقریب خطی پیکربندی $\mathbf{X}$ به شرط پیکربندی $\mathbf{W}$ به صورت

$$E\left(\mathbf{X}|\mathbf{W}\right) \approx \hat{\mu}_w + \frac{\hat{\sigma}_x^2}{\hat{\sigma}_x^2 + \sigma_u^2}\left(\mathbf{W} - \hat{\mu}_w\right)$$

بدست می‌آید که در آن

$$\hat{\mu}_w = \frac{\sum_{l=1}^{L} W_l}{L},$$
$$\hat{\sigma}_x^2 = \hat{\sigma}_w^2 - \sigma_u^2.$$

روش کالیبدن رگرسیون در آمارشکل با توجه به مدل‌بندی مختلط داده‌های شکل دوبعدی قابل اجرا است. نحوه‌ی عملکرد این روش برای مدل رگرسیونی مختلط در مطالعه شبیه‌سازی در بخش ۳ مورد مطالعه قرار خواهد گرفت.

۳ مطالعه شبیه سازی

هدف از این شبیه‌سازی مقایسه‌ی دو روش کالیبدن رگرسیون و OPAی ناپخته برای برآورد پارامترهای مدل (۲.۲) است. بدین منظور برای تولید پیکربندی صحیح $(\mathbf{X})$، ابتدا μ_x از توزیع نرمال مختلط تولید می‌شود. به طور دقیق‌تر، قسمت‌های حقیقی و مجازی μ_x به ازای $l = 1, \cdots, L$ از توزیع نرمال با میانگین صفر و واریانس ۸ تولید و سپس $Re\left(\mathbf{X}\right)$ و $Im\left(\mathbf{X}\right)$ به طور مستقل از $N(Re\left(\mu_x\right), \frac{\sigma_x^2}{2})$ و $N(Im\left(\mu_x\right), \frac{\sigma_x^2}{2})$ شبیه‌سازی می‌شوند. همچنین $Re\left(\epsilon_l\right)$ و $Im\left(\epsilon_l\right)$ به طور مستقل از $N(0, \frac{\sigma_\epsilon^2}{2})$، $Re\left(\mathbf{U}_l\right)$ و $Im\left(\mathbf{U}_l\right)$ به صورت مستقل از $N(0, \frac{\sigma_u^2}{2})$ و در آخر، پیکربندی‌های $\mathbf{W}$ و $\mathbf{Y}$ براساس تساوی‌های $\mathbf{W} = \mathbf{X} + \mathbf{U}$ و $\mathbf{Y} = \beta_0 + \beta_1\mathbf{X} + \epsilon$ بدست می‌آیند. مقادیر واقعی پارامتر به صورت $\beta = (\beta_{00} + \beta_{01}i, \beta_{10} + \beta_{11}i)^T = (1 + 2i, 2 + i)^T$ و $\sigma_\epsilon^2 = 2$ در نظر گرفته شده‌اند. داده‌ها به ازای λهای مختلف با تعداد نقاط شاخص $L = 30$ به ازای $M = 1000$ بار شبیه‌سازی شده و نتایج حاصل از برآورد پارامترها به روش‌های متفاوت در جدول ۱ آمده است که در آن نمادهای Naive و RC به ترتیب نماینده روش‌های ناپخته و کالیبدن رگرسیون هستند. همانطورکه ملاحظه می‌شود وقتی که λ به کوچکی ۰/۵ است، نتایج در جدول ۱ اریبی معنی داری را در برآورد ناپخته پارامترها نشان می‌دهند. با افزایش λ، خطای اندازگیری کاهش و اریبی کمتر می‌گردد. نکته حائز اهمیت این است که روش RC اریبی حاصل در برآوردگر ناپخته را تا حدود زیادی برطرف کرده و برای بهبود دقت برآوردگر ناپخته موثر واقع شده است.

جدول ۱: میانگین برآوردهای ناپخته (Naive) و کالیبدن رگرسیون (RC) در حالت میانگین‌های شکل نابرابر با $L = ۳۰$ نقطه شاخص انتخاب شده از توزیع نرمال مختلط برای پارامترهای مدل با مقادیر واقعی پارامتر $\sigma_\epsilon^2 = ۲$، $\beta_0 = (\beta_{00} = ۱, \beta_{01} = ۲)^T$ و $\beta_1 = (\beta_{10} = ۱, \beta_{11} = ۲)^T$ به ازای مقادیر متفاوت نرخ قابلیت اعتماد (λ)

پارامتر	روش برآورد	نرخ قابلیت اعتماد $\lambda = ۰/۵$	$\lambda = ۰/۸$	$\lambda = ۰/۹۵$
β_{00}	Naive	(۰/۸۱۱) ۱/۰۸۲	(۰/۵۷۲) ۱/۰۵۱	(۰/۳۳۸) ۱/۰۱۳
	RC	(۱/۳۶۸) ۱/۰۰۰	(۰/۶۶۶) ۱/۰۰۷	(۰/۳۵۰) ۱/۰۰۰
β_{01}	Naive	(۰/۷۷۰) ۲/۰۶۰	(۰/۵۶۱) ۲/۰۳۲	(۰/۳۰۵) ۱/۹۹۷
	RC	(۱/۳۰۳) ۱/۹۸۶	(۰/۶۵۶) ۲/۰۰۴	(۰/۳۱۵) ۱/۹۹۲
β_{10}	Naive	(۰/۱۴۵) ۱/۰۰۶	(۰/۱۲۱) ۱/۶۰۲	(۰/۰۷۵) ۱/۹۰۱
	RC	(۰/۲۹۱) ۲/۰۱۲	(۰/۱۵۱) ۲/۰۰۲	(۰/۰۷۹) ۲/۰۰۱
β_{11}	Naive	(۰/۱۵۱) ۰/۳۹۸	(۰/۱۲۱) ۰/۷۹۷	(۰/۰۷۰) ۰/۹۵۲
	RC	(۰/۳۰۳) ۰/۹۹۷	(۰/۱۵۱) ۰/۹۹۶	(۰/۰۷۳) ۱/۰۰۳
σ_ϵ^2	Naive	(۸/۴۰۷) ۳۹/۵۴۶	(۳/۸۷۷) ۲۰/۳۸۵	(۱/۱۹۵) ۶/۰۸۷
	RC	(۸/۴۰۷) ۳۹/۵۴۶	(۳/۸۷۷) ۲۰/۳۸۵	(۱/۱۹۵) ۶/۰۸۷

بحث و نتیجه‌گیری

انطباق پروکراستس دو پیکربندی برای داده‌های شکل دو بعدی می‌تواند به چشم یک مدل رگرسیون خطی مختلط نگاه شود که در آن برآورد پارامترهای مقیاس، دوران و انتقال با برآورد ضرایب مدل رگرسیونی معادل است. در این مقاله خطای اندازگیری براساس همین دیدگاه، برای انطباق پروکراستس مدل‌بندی شده است. در صورت وجود خطای اندازگیری، برآوردگر ناپخته برای پارامترهای مقیاس، دوران و انتقال اریب است. روش کالیبدن رگرسیون که اریبی برآوردگر ناپخته را کاهش می‌دهد، برای بهبود دقت این برآوردگر بکار گرفته شد. در مطالعه شبیه‌سازی نشان داده شد که روش تصحیح اریبی مورد استفاده عملکرد مناسب‌تری در مقایسه با روش مرسوم ناپخته دارد.

مراجع

Carroll, R. J., Ruppert, D., Stefanski, L. A., and Crainiceanu, C. M. (2006), *Measurement Error in Nonlinear Models*, Chapman and Hall, Boca Raton.

Carroll, R. J., and Stefanski, L. A. (1990), Approximate Quasilikelihood Estimation in Models with Surrogate Predictors, *Journal of the American Statistical Association*, **85**, 652–663.

Carroll, R. J., and Callo, P. P. (1982), Some Aspects of Robustness in the Functional Errors-in-Variables Regression Model. *Communications in Statistics-Theory and Methods*, **11**, 2573–2585.

Dryden, I. L. and Mardia, K. V. (1998), *Statistical Shape Analysis*. John Wiley and Sons, Chichester.

Du, J., Dryden, I. L., and Huang, X. (2014), Size and Shape Analysis of Error-Prone Shape Data, *Journal of the American Statistical Association*, **110**, 368-379.

Fuller, W. A. (1987), *Measurement Error Models*, John Wiley and Sons, New York.

Gleser, L. J. (1981), Estimation in a Multivariate "Errors in Variables" Regression Model: Large Sample Results, *The Annals of Statistics*, **9**, 24–44.

Goodman, N. R. (1963), Statistical Analysis Based on a Certain Multivariate Complex Gaussian Distribution (an Introduction), *The Annals of Mathematical Statistics*, **34**, 152–177.

Kendall, D. G. (1984), Shape Manifolds, Procrustean Metrics and Complex Projective Spaces. *Bulletin of the London Mathematical Society*, **16**, 81–121.

Mardia, K. V., Kent, J. T., and Bibby, J. M. (1979), *Multivariate Analysis*. Academic Press, London.

Stefanski, L. A. (1985), The Effects of Measurement Error on Parameter Estimation, *Biometrika*, **72**, 583–592.