

## مدل‌بندی رگرسیونی شکل از طریق مثلثی کردن

میثم مقیم بیگی، موسی گل‌علی‌زاده

گروه آمار، دانشگاه تربیت مدرس

چکیده: با توجه به تعریف کندال از شکل به‌عنوان نقطه‌ای در فضای اَبَر کره، در این مقاله مدل‌بندی رگرسیونی شکل در این فضا مورد مطالعه قرار می‌گیرد. همچنین به منظور سهولت در مدل‌بندی، روش مثلث‌بندی شکل با استفاده از دو نقطه شاخص خاص پیشنهاد می‌شود که عملکرد مناسبی در مقایسه با رویکردهای دیگر دارد. مثلث‌بندی نه‌تنها مدل‌بندی رگرسیونی شکل را آسان می‌نماید بلکه توانایی بازسازی ساختار هندسی اشیاء با استفاده از ابزارهای ساده محاسباتی را دارد. نوآوری روش پیشنهادی مقاله حاضر در استفاده از متغیر تبیینی مبتنی بر شکل اشیاء است که تغییرات هندسی متغیر پاسخ را به‌خوبی توصیف می‌کند. مقایسه و ارزیابی روش پیشنهادی با مدل انطباق پروکراستس کامل بر اساس معیار مجموع توان دوم خطا انجام و عملکرد دو مدل در تحلیل داده‌های پیکربندی جمجمه موش‌های آزمایشگاهی مورد بررسی قرار می‌گیرد.

واژه‌های کلیدی: رگرسیون کروی، آمار شکل، مثلث‌بندی، انطباق پروکراستس، فضای نااقلیدسی.

## ۱ مقدمه

گاهی اوقات شکل به‌عنوان منبع ارزشمندی از اطلاعات می‌تواند برای تحلیل آماری مورد استفاده قرار گیرد. انجام این مهم مستلزم حفظ ساختار هندسی آن است چراکه در غیر اینصورت بررسی تناظر بین اشیاء در یک قالب استاندارد ممکن نخواهد بود. در سالیان اخیر، موضوع مورد علاقه محققین، مدل‌بندی رگرسیونی داده‌های شکل بوده است. یکی از رویکردهای مرسوم برای نیل به این هدف، در نظر گرفتن مدل‌های معمول در فضای اقلیدسی برای مجموعه‌ای از اشکال است که با توجه به تعاریف ارائه شده از فضای شکل، به‌نظر می‌رسد این مدل‌ها از کارایی مطلوبی برخوردار نباشند. یکی از روش‌های مناسب جایگزین و تا حدی منطقی‌تر، انجام این تحلیل‌ها در فضای اصلی مورد مطالعه یعنی فضای شکل است.

بنا به **کندال (۱۹۷۷)** شکل، همه اطلاعات باقی‌مانده از یک پیکربندی پس از حذف اثرات مکان، مقیاس و دوران است. در دیدگاه ایشان فضای شکل یک فضای کروی با بعد بالا است. به‌طور دقیق‌تر و با پیروی از او، شکل یک مثلث معادل نقطه‌ای روی کره معمولی با بعد دو ( $S^2$ )، و شکل یک مربع، هم ارز نقطه‌ای روی کره با بعد سه ( $S^3$ ) است. نوعی دیگری از نگاه به اشکال هندسی که توسط **کندال (۱۹۸۴)** و **بوکشتین (۱۹۸۶)** معرفی شد مثلث‌بندی<sup>۱</sup> است. در این رویکرد با اختیار سه نقطه از هر شکل تمامی مثلث‌های ممکن به‌دست می‌آید و این مجموعه مثلث‌ها مبنای تحلیل آماری قرار می‌گیرد. می‌توان ملاحظه کرد که چنین روشی حالت خاصی از سنگفرش ورونی<sup>۲</sup> است که به‌عنوان یکی از ابزارهای ریخت‌شناسی<sup>۳</sup> در برخی از تحلیل‌های آماری ماترب با مباحث فرایندهای تصادفی و الگوهای نقطه‌ای به‌کار گرفته شده است. جزئیاتی از این موضوع را می‌توان در **استویان و استویان (۱۹۹۴)** مطالعه کرد. حال اگر شخصی مثلث‌بندی روی یک شی را انجام و سپس مختصات مثلث‌های حاصل از طریق مختصات **کندال (۱۹۸۴)** بدست آورد او با تعدادی نقطه روی کره که بین آن‌ها همبستگی وجود دارد، روبه‌رو خواهد شد. آنگاه، مدل‌بندی رگرسیونی شکل بر اساس موقعیت این نقاط و سپس بازخوانی ارتباط اشکال موضوع تحقیق آماری است که مقاله حاضر سعی در پاسخ به آن دارد.

مدل‌های رگرسیونی بر روی کره در ابعاد بالا قبلاً توسط **چانگ (۱۹۸۶)** و **ریوست (۱۹۸۹)** ارائه شده‌اند. آن‌ها مدل‌های دورانی را روی کره معرفی کردند و به بررسی برخی از ویژگی‌های این مدل‌ها پرداختند. اخیراً،

<sup>1</sup>Triangulation

<sup>2</sup>Voronoi tessellation

<sup>3</sup>Morphology

**فیشبوق و همکاران (۲۰۱۳)** مدلی رگرسیونی در فضای شکل که از تعمیم مدل رگرسیونی خطی ساده به دست می‌آید را معرفی کردند. آن‌ها با استفاده از مدل رگرسیون خطی و ژئودزیک شکل به برآورد پارامترهای مدل خود پرداختند. تقریباً با همین نگاه، **کائو و همکاران (۲۰۱۴)** استفاده از مدل رگرسیونی چندمتغیره برای داده‌های شکل روی صفحه را پیشنهاد کردند. آن‌ها با معرفی یک تابع رگرسیون برداری به پیش‌بینی شکل داده‌های پیکربندی صورت انسان پرداختند. **هینکل و همکاران (۲۰۱۴)** با گسترش رگرسیون چندجمله‌ای در منفی‌د ریمانی به معرفی مدل‌های رگرسیونی برای داده‌های شکل پرداختند. آن‌ها همچنین به معرفی مدل ژئودزیک و اسپلاینی براساس چندجمله‌ای‌ها در منفی‌د کره پرداخته و آن‌ها را با استفاده از داده‌های واقعی مقایسه کردند. بررسی مدل‌های رگرسیونی شکل که در آن متغیر پاسخ و پیشگو هر دو شکل (بر اساس تعریف کندال) هستند موضوع جالب توجه‌ای در مدل‌بندی آماری است. این مهم در طی سالیان گذشته مورد توجه محققان مختلف در مسائل کاربردی بوده است. به نظر می‌رسد مدل انطباق پروکراستس که **کندال (۱۹۸۴)** برای تصویر کردن یک پیکربندی بر روی پیکربندی دیگر به کار گرفت، خود نوعی از مدل رگرسیونی شکل باشد. لازم به اشاره است که مدل انطباق پروکراستس با تغییر مکان، دوران و تغییر مقیاس، یک پیکربندی را روی پیکربندی دیگر منطبق می‌کند. اما از آنجایی که چنین مدلی تغییرات موجود در شکل را مورد بررسی قرار نمی‌دهد کارایی لازم برای مدل‌بندی رگرسیونی شکل بر اساس شکل دیگری را ندارد. از این‌رو نیاز به معرفی یک مدل رگرسیونی کارا برای شکل یکی از دغدغه‌های محققین علاقه‌مند در این حوزه بوده و هست. پژوهش حاضر کوششی برای پاسخ به این نیاز است.

برای ارائه نتایج مقاله حاضر، در بخش ۲ به معرفی مدل رگرسیونی پروکراستس پرداخته می‌شود. در بخش ۳ روش مثلثی کردن شکل و مدل رگرسیون کروی مربوطه برای مثلث تشریح می‌شود. تحلیل یک مثال واقعی با رویکرد مورد اشاره در این مقاله در بخش ۴ می‌آید. علاوه بر این، مقایسه آماری مناسبی بین چند مدل نیز انجام خواهد شد.

## ۲ انطباق پروکراستس

یکی از راه‌های مشخص کردن شکل یک شیء استفاده کردن از نقاطی است که بر روی سطح خارجی شیء قرار می‌گیرند. انتخاب این نقاط باید به گونه‌ای باشد که تا حد امکان اطلاعات کلی موجود در شیء را بیان کنند. واضح است که هرچه تعداد این نقاط بیشتر باشد، اطلاعات بیشتری از شیء در دسترس خواهد بود. در آمار شکل

به این مجموعه نقاط متناهی که هندسه شیء را مشخص می‌کنند، نقاط شاخص و به مجموعه تمام نقاط شاخصی که برای یک شیء در نظر گرفته می‌شود، پیکربندی آن شیء می‌گویند. به علاوه، به منظور مشخص کردن پیکربندی در روی صفحه با  $L$  نقطه شاخص، از یک ماتریس  $۲ \times L$  بُعدی استفاده می‌کنند. با تعریف پیکربندی در مجموعه اعداد مختلط می‌توان این ماتریس را به صورت یک بردار مختلط  $۱ \times L$  در نظر گرفت.

به ازای  $n, \dots, ۱, i$ ، مجموعه‌ای از پیکربندی‌های دو بُعدی مستقل  $X_i = (x_1, \dots, x_k)^T$  و  $Y_i = (y_1, \dots, y_k)^T$  در مجموعه اعداد مختلط با ویژگی  $Y_i^* \mathbf{1}_k = 0 = X_i^* \mathbf{1}_k$  که در آن  $Y_i^*$  ترانزپوز مزدوج  $Y_i$  است را در نظر بگیرید. فرض کنید هدف، انطباق  $X_i$  بر روی  $Y_i$  یا به عبارتی دیگر یافتن یک مدل رگرسیونی مناسب برای  $Y_i$  بر حسب  $X_i$  است. یک معیار مناسب برای انطباق  $X_i$  به روی  $Y_i$  استفاده از تبدیلات تشابه و تفاوت است. برای جزئیات بیشتر در این زمینه به فصل سوم **درآیند و ماردیا (۱۹۹۸)** مراجعه شود. با بازخوانی نمادگذاری اعداد مختلط و با پیروی از **درآیند و ماردیا (۱۹۹۸)**، یکی از معادلات رگرسیونی مناسب به صورت

$$\begin{aligned} Y_i &= (a + ib)\lambda_k + \beta e^{i\theta} X_i + \epsilon \\ &= [\lambda_k, X_i]A + \epsilon_i \\ &= X_i^d A + \epsilon_i, \end{aligned} \quad (۱)$$

است که در آن  $A = (A_1, A_2)^T = (a + ib, \beta e^{i\theta})^T$  بردار دو بُعدی از پارامترهای مختلط،  $a + ib$  نمایانگر انتقال،  $\beta > 0$  معرف مقیاس و  $0 \leq \theta < 2\pi$  نشانگر دوران هستند. به علاوه  $X_i^d = [\lambda_k, X_i]$  یک ماتریس طرح  $۲ \times k$  و  $\epsilon_i$  بردار خطای مختلط  $۱ \times k$  بُعدی است. به منظور ارائه یک انطباق مناسب، باید برآورد پارامتر  $A$  با می‌نیم‌سازی تابع مجموع توان دوم خطاهای<sup>۴</sup>

$$\sum_{i=1}^n D^2(Y_i, X_i) = \sum_{i=1}^n \epsilon_i^* \epsilon_i = (Y_i - X_i^d A)^* (Y_i - X_i^d A)$$

به دست آید. در ادبیات آمار چندمتغیره، برآورد معادله رگرسیونی در این مسئله به انطباق کامل پروکراستس  $X_i$  روی  $Y_i$  معروف است. به عنوان مثال **ماردیا و همکاران (۱۹۷۹)** را ببینید. می‌توان نشان داد که برآورد ماتریس  $A$  که با  $\hat{A}$  نمایش داده خواهد شد عبارتست از:

$$\hat{A} = (\hat{a} + i\hat{b}, \hat{\beta}e^{i\hat{\theta}})^T = \arg \inf \sum_{i=1}^n \epsilon_i^* \epsilon_i = \arg \inf \sum_{i=1}^n (Y_i - X_i^d A)^* (Y_i - X_i^d A).$$

<sup>۴</sup>Sum of squared errors

به‌طور دقیق‌تر، برآورد پارامترها در انطباق کامل پروکراستس به‌صورت

$$\hat{a} + i\hat{b} = 0,$$

$$\hat{\theta} = \arg\left(\sum_{i=1}^n x_i^* y_i / n\right) = -\arg\left(\sum_{i=1}^n y_i^* x_i / n\right),$$

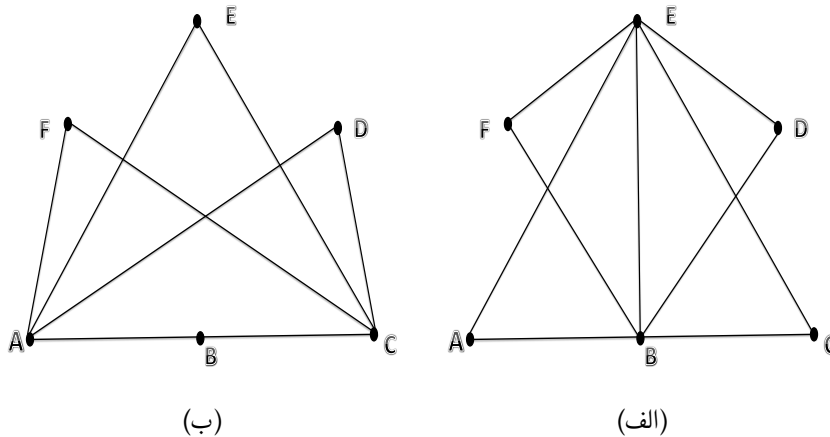
$$\hat{\beta} = \left(\sum_{i=1}^n x_i^* y_i y_i^* x_i\right)^{1/2} / \left(\sum_{i=1}^n x_i^* x_i\right)$$

هستند (درآیدن و ماریا، ۱۹۹۸). از آن‌که مدل رگرسیونی (۱) بر اساس پارامترهای مکان، دوران و مقیاس نوشته شده است، بنابراین نمی‌تواند تغییرات موجود در شکل‌ها را به‌درستی نشان دهد. به‌عبارت دیگر چنانچه تغییرات متغیر شکل پاسخ خیلی چشم‌گیرتر از تغییرات متغیر شکل پیش‌گو باشد، مدل رگرسیونی (۱) نمی‌تواند آن‌را به‌خوبی توصیف کند. به‌طور دقیق‌تر برآورد شکل متغیر پاسخ، شکلی است که از جابجایی، دوران و تغییر مقیاس شکل پیش‌گو به‌دست می‌آید و در نتیجه تا حد زیادی شبیه همان شکل پیش‌گو است. در خیلی از مواقع این موضوع مد نظر محقق نیست بلکه ایشان در پی مدلی است که تغییرات هر کدام از مولفه‌های درونی و بیرونی متغیر پاسخ را بر اساس چنین مولفه‌هایی در متغیر پیش‌گو توصیف کند. به‌نظر می‌رسد مثلث‌بندی پیکربندی‌ها و سپس توجه به هر یک از آن‌ها به‌عنوان افزایش از شکل شیء نهایی رویکرد مناسب‌تری برای نیل به موضوع مدل‌بندی رگرسیونی شکل باشد. چون شکل مثلث‌ها در دیدگاه کندال (۱۹۷۷) هم ارزش نقاطی روی کره است مدل پیشنهادی را رگرسیون کروی نامیده‌ایم. در ادامه جزئیاتی از این مدل معرفی می‌شود.

### ۳ رگرسیون کروی

همانگونه که اشاره شد نمایش شکل (با ادبیاتی که کندال (۱۹۸۴) معرفی کرد) اشیاء با هر تعداد اضلاع روی  $S^2$  مستلزم مثلثی کردن پیکربندی است. بیان این نکته ضروری است که مثلثی کردن شکل به روش‌های مختلفی انجام می‌گیرد که برخی از آن‌ها در زلدیچ و همکاران (۱۹۸۹)، بوکشتین (۱۹۹۱) و راتو (۲۰۰۰) اشاره کرد. از روش‌های مثلثی کردن یک پیکربندی استفاده از یک محور اصلی است. در این روش دو نقطه شاخص روی یک پیکربندی به‌عنوان محور اصلی مثلث‌بندی انتخاب شده و نقطه شاخص سوم از میان سایر نقاط پیکربندی انتخاب می‌شود. دو نمونه از مثلثی کردن یک شکل بر اساس یک محور اصلی که برگرفته از بوکشتین (۱۹۹۱) است، در شکل ۱ به نمایش درآمده است. همان‌طور که ملاحظه می‌شود شکل ۱-ب بر اساس محور اصلی  $AC$  و شکل ۱-الف بر اساس محور اصلی  $BE$  مثلثی شده‌اند.

به‌عنوان یک ایده جدید می‌توان از روش مثلث‌بندی پیکربندی‌ها برای مدل‌بندی رگرسیونی اشیاء استفاده کرد. در این مقاله ما ابتدا پیکربندی‌ها را با استفاده از روش مثلث‌بندی براساس محور اصلی مثلث‌بندی کرده و سپس با استفاده از تبدیل کندال (۱۹۸۴) مثلث‌ها را به روی کره منتقل می‌کنیم. در انتها نیز با بکارگیری مدل رگرسیونی کروی که در ادامه معرفی خواهیم کرد مثلث‌ها را مدل‌بندی می‌کنیم. مهمترین ویژگی روش مثلثی کردن اشیاء آن است که پس از پیش‌گویی شکل مثلث‌ها به راحتی می‌توان شکل کلی را با استفاده از عکس تبدیلات کندال بازسازی کرد. به‌طور دقیق‌تر، ابتدا شکل با استفاده از یک محور مثلثی شده سپس نقاط متناظر آن‌ها روی کره  $S^2$  مشخص می‌شوند. با چنین رویکردی به شکل اشیاء، می‌توان مدل رگرسیونی کروی که نمایانگر یک مدل رگرسیون مناسب برای بیان ارتباط بین مجموعه نقاطی از کره که متغیرهای پیشگو هستند با مجموعه دیگری از نقاط در همین کره که متغیرهای پاسخ خواهند بود، بنا کرد. از آنجایی که این روش به رگرسیون تک تک نقاط شکل به‌طور جداگانه می‌پردازد به‌نظر می‌رسد تغییرات شکل نیز در آن اعمال شود. از دیگر ویژگی‌های این روش آن است که چنانچه محقق علاقه‌مند به بررسی تغییرات مقیاس (اندازه) شکل باشد، می‌تواند چنین تغییراتی را از طریق اندازه محور اصلی در شکل اولیه شیء محاسبه کرد.



شکل ۱: مثلثی کردن شکل با استفاده از یک محور اصلی الف: محور اصلی  $BE$ . ب: محور اصلی  $AC$ .

به‌طور کلی، به‌ازای  $n, \dots, 1, i$  و  $d \geq 3$ ، برای بردار متغیرهای پیشگوی  $\mathbf{x}_i = (x_{1i}, \dots, x_{di})^T$  و متغیرهای پاسخ  $\mathbf{y}_i = (y_{1i}, \dots, y_{di})^T$ ، یک مدل رگرسیونی ساده مدلی به‌صورت  $\mathbf{y}_i = \mu(\mathbf{x}_i) + \varepsilon_i$  است که در آن  $\varepsilon_i$  یک بردار از متغیر تصادفی با میانگین بردار صفر و ماتریس کوواریانس  $\Sigma_i$  است. همان‌طور که ملاحظه می‌شود میانگین متغیرهای پاسخ در واقع میانگین شرطی  $\mu(\mathbf{x}_i) = E(\mathbf{y}_i | \mathbf{x}_i)$  است که می‌تواند

صورت پارامتری، نیمه پارامتری یا حتی ناپارامتری داشته باشد (برای مشاهده جزئیات بیشتر به [سبر و لی \(۲۰۱۲\)](#) مراجعه شود). یکی از ساده‌ترین مدل‌های رگرسیونی روی کره مدل دورانی است که  $\mu(\mathbf{x}_i)$  صورت پارامتری دارد ([چانگ، ۱۹۸۶](#)). اگر  $SO(d)$  مجموعه کلیه ماتریس‌های متعامد با دترمینان یک باشد، آن‌گاه در مدل دورانی پارامتری تابع  $\mu(\mathbf{x}_i)$  به صورت  $A\mathbf{x}_i$  در نظر گرفته می‌شود که در آن  $A \in SO(d)$  ماتریسی  $d \times d$  بعدی با مقادیر حقیقی متعلق به گروه متعامد یا یک زیر مجموعه از آن است. لذا مدل رگرسیونی کروی به صورت

$$\mathbf{y}_i = A\mathbf{x}_i + \varepsilon_i \quad (۲)$$

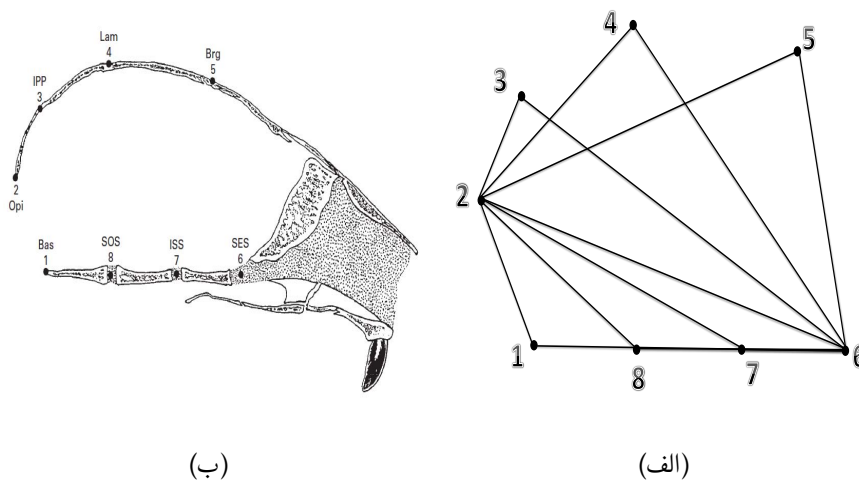
است، که در آن  $A$  یک دوران صلب بوده و  $\varepsilon_i$  متغیری تصادفی معرف خطا روی کره است. یکی از مناسب‌ترین توزیع‌های مورد استفاده روی کره که بتواند تغییرات  $\varepsilon_i$ ها را توصیف کند توزیع فون میزس-فیشر است که اولین بار توسط [ماردیا \(۱۹۷۲\)](#) معرفی شد. نحوه برآورد پارامتر  $A$  در مدل رگرسیونی (۲) به وسیله [مکنزی \(۱۹۵۷\)](#) و [استیفنز \(۱۹۷۹\)](#) و با استفاده از روش کمترین توان‌های دوم خطا تشریح شد که جزئیاتی از آن در ادامه می‌آید.

فرض کنید  $C = E[\sum_{i=1}^n \mathbf{y}_i \mathbf{x}_i^T] \in \mathbb{R}^{d \times d}$ . تجزیه ویژه مقدار پیراسته<sup>۵</sup> این ماتریس به صورت  $C = U\Lambda V^T$  است که  $U, V \in SO(d)$  و  $\Lambda$  ماتریس قطری با درایه‌های  $|\lambda_d| \geq \lambda_{d-1} \geq \dots \geq \lambda_1$  هستند. آن‌گاه یک دوران بهینه به صورت  $\hat{A} = UV^T$  است. [چانگ \(۱۹۸۶\)](#) نشان داد که این برآوردگر، یک برآوردگر سازگار قوی بوده و دارای توزیع مجانبی نرمال است. همچنین او نشان داد که تبدیل نمایی معکوس  $\hat{A}^T \hat{A}$  به طور مجانبی دارای توزیع نرمال چندمتغیره است. توجه شود که نتایج حاصل به این شرط وابسته است که ماتریس  $\frac{1}{n} \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^T$  به یک ماتریس تمام رتبه  $\Sigma$  همگرا باشد. [شین و همکاران \(۲۰۰۱\)](#) به این نکته اشاره کردند که رتبه  $\Sigma$  را می‌توان تا  $d - 1$  پایین آورد که در حقیقت یک شرط لازم برای سازگاری قوی  $\hat{A}$  است.

<sup>5</sup>Modified singular value decomposition

## ۴ تحلیل داده‌های جمجمه موش

در بخش ۲ مدل رگرسیونی پروکراستس کامل، تشریح و در بخش ۳ روشی بر مبنای مدل‌بندی رگرسیونی شکل از طریق مثلث‌بندی پیکربندی و همچنین استفاده از رگرسیون کروی پیشنهاد شد. به‌کارگیری این دو مدل برای داده‌های واقعی و همچنین مقایسه آن‌ها می‌تواند معیاری برای ارزیابی عملکرد روش پیشنهادی باشد. به‌همین منظور، در این مقاله زیرمجموعه‌ای از داده‌های شکل جمجمه ۲۰ موش آزمایشگاهی که اولین بار توسط بوکشتین (۱۹۹۱) مورد تحلیل قرار گرفت انتخاب می‌شود و مدل‌بندی رگرسیونی شکل بر اساس دو روش ارائه شده انجام خواهد گرفت. این اشکال مربوط به پیکربندی جمجمه موش‌ها در روز ۷-ام و ۱۵۰-ام از تولد آن‌ها است. نمونه‌ای از جمجمه موش‌ها در شکل ۲-ب به نمایش درآمده است. بوکشتین (۱۹۹۱) توصیه کرد



شکل ۲: الف: نحوه مثلث‌بندی با استفاده از نقاط شاخص ۲ و ۶. ب: نمونه‌ای از جمجمه موش‌های آزمایشگاهی.

به‌منظور مثلثی کردن شکل با استفاده از یک محور، دو مسئله در نظر گرفته شود. نکات مورد اشاره ایشان عبارتند از: محور اصلی تا حد امکان از نقاط شاخص دور از هم انتخاب شود؛ محورهای انتخابی از درون پیکربندی عبور کند. با در نظر گرفتن این نکات یک انتخاب مناسب برای محور اصلی مثلث‌بندی، محور به‌دست آمده از خط واصل بین دو نقطه ۲ و ۶ است. با توجه به محور اصلی انتخاب شده نحوه مثلث‌بندی جمجمه‌ها در شکل ۲-الف نشان داده شده است. البته باید یادآوری کرد که رویکرد بوکشتین مثلث‌بندی پیکربندی‌ها روی



صفحه و تحلیل آن در صفحه بوده است. با توجه به این‌که فضای شکل فضای ابرکره است لذا ما بر این باوریم که مثلث‌ها باید در فضای کره بررسی شوند. بنابراین با تلفیق روش بوکشتین (۱۹۹۱) و کندال (۱۹۸۴) و استفاده از مدل دورانی کروی به مدل‌بندی رگرسیونی داده‌های موش آزمایشگاهی می‌پردازیم.

به‌منظور اثبات ادعای بوکشتین در عملکرد مناسب نقاط دورتر در تحلیل‌های آماری، می‌توان تمامی مثلث‌بندی‌های ممکن از شکل را انجام و مدل رگرسیونی (۲) را برای مثلث‌ها به‌کار برد. بر اساس محورهای مختلف، مثلث‌بندی‌های گوناگونی به‌وجود می‌آید که مقدار مجموع توان دوم خط‌های آن‌ها در جدول ۱ ارائه شده است. همانگونه که از این جدول مشاهده می‌شود، مقدار مجموع توان دوم خط برای مدل رگرسیونی (۲)

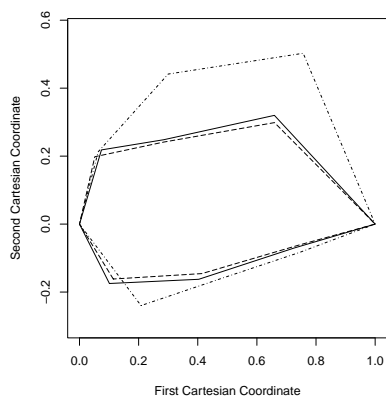
جدول ۱: مقدار مجموع توان دوم خط با انتخاب جفت نقاط شاخص به‌عنوان محور اصلی مثلث‌سازی

نقاط شاخص	۱	۲	۳	۴	۵	۶	۷	۸
۱	-	۹۹,۰۰	۱۳,۷۶	۱۵,۳۱	۳,۶۰	۴,۱۸	۱۰,۵۲	۳۴,۰۸
۲	-	-	۳۴,۰۸	۲۰,۹۱	۲,۹۳	۱,۵۷	۲,۸۶	۵,۴۰
۳	-	-	-	۱۳۷,۲۶	۵,۳۴	۱,۶۲	۲,۶۸	۳,۸۷
۴	-	-	-	-	۱۱,۲۲	۵,۶۰	۸,۴۴	۹,۳۶
۵	-	-	-	-	-	۴۲,۸۲	۳۰,۶۰	۴,۸۲
۶	-	-	-	-	-	-	۲۷,۹۷	۸,۱۷
۷	-	-	-	-	-	-	-	۴۵,۸۵

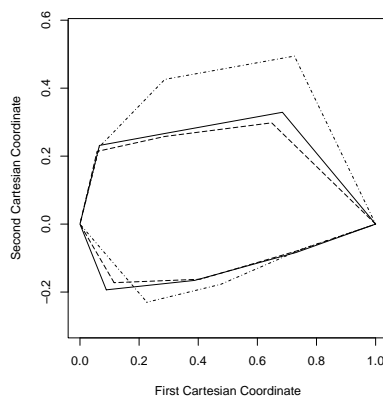
بر اساس مثلث‌سازی شکل بر پایه محور اصلی و با استفاده از دو نقطه شاخص ۲ و ۶ کمترین و برابر ۱/۵۷ است. پس از برازش مدل (۱) مقدار برآورد پارامترها برابر  $\hat{\theta} = -۰/۰۰۶$  و  $\hat{\beta} = ۰/۲۱$  به‌دست آمدند. همچنین مجموع توان دوم خط برای مدل (۱) برابر ۱۳,۹۰ است که بسیار بیشتر از مجموع توان دوم خطای مدل پیشنهادی است. پس از انجام رگرسیون برای داده‌های شکل، به‌منظور ارزیابی عملکرد روش پیشنهادی‌مان از آزمون گودال (۱۹۹۱) برای بررسی میزان انطباق مقادیر واقعی پاسخ (ساختار هندسی شکل حجمه در روز ۱۵-ام) با مقادیر پیش‌گویی (حاصل از انجام مدل رگرسیون کروی شکل) متناظر آن‌ها در همان روز استفاده کردیم. مقدار آماره این آزمون برابر  $۱۰^{-۴} \times ۱/۶$  و  $p$ -مقدار متناظر خیلی نزدیک به یک شد. بنابراین فرضیه برابری میانگین شکل مبتنی بر مقادیر واقعی و پیش‌گویی در سطح ۵٪ تایید می‌شود. همچنین این آزمون نشان می‌دهد که مدل رگرسیون کروی شکل پیشنهاد شده در این مقاله از دقت بالایی برخوردار است.

واضح است که استفاده از معیارهای مرسوم رگرسیونی مانند ضریب تعیین برای این امر ترجیح داده می‌شود. اما به دلیل این‌که روش‌هایی برای محاسبه این معیار و کمیت‌های دیگر رگرسیونی تاکنون در حوزه رگرسیون شکل معرفی نشده است بررسی چنین امری به تحقیقات آتی در این حوزه موکول می‌شود.

برای نمایش کاربردی از مدل پیشنهادی‌مان در این مقاله پیش‌گویی ساختار هندسی مجموعه دو موش (شماره ۱ و ۲) در روز ۱۵۰-ام با استفاده از اطلاعات موجود در متغیر تبیینی (روز ۷-ام) را مدنظر قرار دادیم. اگرچه تغییر شکل مجموعه متأثر از عوامل متعددی می‌باشد اما انتظار می‌رود مدل رگرسیون کروی شکل پیشنهادی در این مقاله نیز بتواند ویژگی‌های هندسی مربوط به آن را توصیف کند. نمایشی از ساختار هندسی مجموعه دو موش انتخابی در روزهای ۷-ام و ۱۵۰-ام همراه با پیش‌گویی شکل مجموعه در روز ۱۵۰-ام در نمودار ۳ آمده است. همان‌طور که ملاحظه می‌شود انطباق نسبتاً مناسبی بین مقادیر واقعی و پیش‌گویی (هر دو برای روز ۱۵۰-ام) وجود دارد.



(ب)



(الف)

شکل ۳: شکل مجموعه دو موش شماره الف: ۱ و ب: ۲ در روز ۷-ام (نقطه-خط)، روز ۱۵۰-ام (خطوط بریده) و پیش‌گویی ساختار هندسی آن‌ها (خطوط ممتد).

## بحث و نتیجه‌گیری

اغلب مدل‌بندی‌های رگرسیونی شکل در فضای مماسی یا روی صفحه انجام می‌گیرند. با توجه به این‌که فضای شکل یک ابر کره است، در نظر گرفتن مدل رگرسیونی مناسب در همین فضا ضروری به نظر می‌رسد. اگرچه مدل‌بندی رگرسیونی در فضای شکل به دلیل ماهیت غیراقلیدسی فضای مشاهدات با دشواری‌هایی روبه‌رو است، اما تسهیلات فراهم‌شده توسط این مدل قابلیت‌های منحصر بفردی خواهد داشت. در این مقاله پیشنهاد شد که یکی از رویکردهای ساده‌سازی صورت مسئله، مثلث‌بندی شکل هندسی اشیاء و استفاده از فضای کره دوبعدی است. این ایده کمک کرد که پس از انجام تحلیل‌ها بر روی مثلث‌ها، شکل به راحتی بازسازی شود. مدل رگرسیون کروی ویژگی بارز قرار داشتن روی کره تحت تبدیلات متعامد را داشت. عملکرد مدل پیشنهادی بر روی مجموعه داده واقعی اعتبار آن را در مقایسه با مدل‌های مرسوم قبلی نشان داد. با الگوبرداری از مقاله حاضر، تعمیم مدل رگرسیونی کره بر روی ابرکره‌های دیگر، یافتن مدل‌های نیرومندتر و ارتباط رگرسیونی اشکال با متغیرهای دودویی و طبقه‌بندی‌شده می‌توانند موضوعات مناسبی برای تحقیقات آتی در حوزه مدل‌بندی رگرسیونی کروی باشند.

## مراجع

- Bookstein, F. L. (1986), Size and Shape Spaces for Landmark Data in Two Dimensions (With Discussion), *Statistical Sciences*, **1**, 181-242.
- Bookstein, F. L. (1991), *Morphometric Tools for Landmark Data: Geometry and Biology*, Cambridge University Press, New York.
- Cao, X., Wei, Y., Wen, F. and Sun, J. (2014), Face Alignment by Explicit Shape Regression, *International Journal of Computer Vision*, **107**, 177-190.
- Chang, T. (1986), Spherical Regression, *Annals of Statistics*, **14**, 907-924.
- Dryden, I. L. and Mardia, K. V. (1998), *Statistical Shape Analysis*, John Wiley, Chichester.

- Fishbaugh, J., Prastawa, M., Gerig, G. and Durrleman, S. (2013), Geodesic Shape Regression in the Framework of Currents, In Proceedings of *International Conference on Information Processing in Medical Imaging*, 718-729.
- Goodall, C. (1991), Procrustes Methods in the Statistical Analysis of Shape, *Journal of Royal Statistical Society: Series B*, **53**, 285-339.
- Hinkle, J., Fletcher, P.T. and Joshi, S. (2014), Intrinsic Polynomials for Regression on Riemannian Manifolds, *Journal of Mathematical Imaging and Vision*, **50**, 32-52.
- Kendall, D. G. (1977), The Diffusion of Shape, *Advances in Applied Probability*, **9**, 428-430.
- Kendall, D. G. (1984), Shape Manifolds, Procrustean Metrics, and Complex Projective Spaces, *Bulletin of the London Mathematical Society*, **16**, 81-121.
- MacKenzie, J. K. (1957), The Estimation of an Orientation Relationship, *Acta Cryst.*, **10**, 61-62.
- Mardia, K.V., (1972), *Statistics of Directional Data*, Academic Press, London.
- Mardia, K. V., Kent, J. T., and Bibby, J. M. (1979), *Multivariate Analysis*, Academic Press, London.
- Rao, C. R. (2000), A Note on Statistical Analysis of Shape Through Triangulation of Landmarks, *Proceedings of the National Academy of Sciences*, **97**, 2995-2998.
- Rivest, L. P. (1989), Spherical Regression for Concentrated Fisher-von Mises Distribution, *Annals of Statistics*, **17**, 307-317.
- Seber, G. A. and Lee, A. J. (2012), *Linear Regression Analysis*, John Wiley & Sons, New Jersey.

- Shin, H. H., Takahara, G. K., and Murdoch, D. J. (2001), Uniqueness, Consistency and Optimality in Spherical Regression Experiments, *Statistics and Probability Letters*, **54**, 61-65.
- Stephens, M. A. (1979), Vector Correlation, *Biometrika*, **66**, 41-48.
- Stoyan, D., and Stoyan, H. (1994), *Fractals, Random Shapes, and Point Fields: Methods of Geometrical Statistics*, John Wiley & Sons, Chichester.
- Zelditch, M. L., Debry, R. W., and Straney, D. O. (1989), Triangulation-Measurement Schemes in the Multivariate Analysis of Size and Shape, *Journal of Mammalogy*, **70**, 571-579.

# **Regression Modelling of Shape Through Triangulation**

**Moghimbeygi, M., Golalizadeh, M.**

Department of Statistics, Tarbiat Modares University, Tehran, Iran.

## **Abstract:**

Recalling the definition of shape as a point on hyper-sphere, proposed by Kendall, the regression model is studied in this paper. In order to simplify the modeling, the triangulation via two landmarks is proposed. The triangulation not only simplifies the regression modelling of the shapes but also provides straightforward computation procedure to reconstruct geometrical structure of the objects. Novelty of the proposed method in this paper is on using the predictor variable, based upon the shape, which suitably describes the geometrical variability of the response. The comparison and evaluation of the proposed methods with the full Procrustes matching through the mean square error criteria are done. Application of two models for the configurations of rat skulls is investigated.

**Keywords:** spherical regression, Shape analysis, Triangulation, Procrustes matching, Non-Euclidean space.

**Mathematics Subject Classification (2010):** 62F30, 62J99.