

# مدل رگرسیون فضایی ضریب متغیر برای عوامل خطر نسبی بیماران سرطان مری

کیما کاظمی<sup>۱</sup>، محسن محمدزاده<sup>۲</sup>

تاریخ دریافت: ۹۹/۱۱/۱۰

تاریخ پذیرش: ۹۹/۱۲/۲۷

چکیده:

در روش‌های متداولی که برای مدل‌بندی داده‌های بقای فضایی و عوامل مؤثر بر آن استفاده می‌شوند، اغلب فرض بر آن است که ضرایب متغیرهای تبیینی در مناطق مختلف ناحیه مورد مطالعه تأثیر یکسانی بر زمان بقا دارند و معمولاً همبستگی فضایی داده‌ها از طریق یک اثر تصادفی در مدل لحاظ می‌شود؛ اما در بسیاری از مسائل کاربردی عوامل مؤثر بر زمان بقا و خطر نسبی، در مناطق مختلف اثرات یکسانی ندارند. در این مقاله، اثرات عوامل مؤثر بر خطر نسبی که در مناطق مختلف یکسان نیستند، مورد بررسی قرار می‌گیرد. برای این منظور مدل‌های رگرسیون فضایی و ضریب متغیر فضایی معرفی و نحوه برآورد بیزی پارامترهای آن‌ها ارائه می‌شود. سپس سه مدل رگرسیون کلاسیک، رگرسیون فضایی و رگرسیون فضایی ضریب متغیر برای مدل‌بندی عوامل مؤثر بر خطر نسبی مازاد بیماران سرطان مری استفاده شده و خطر ناشی از عوامل مختلف مورد بررسی و ارزیابی قرار می‌گیرد.

**واژه‌های کلیدی:** داده‌های بقای فضایی، مدل ضریب متغیر فضایی، مدل مخاطرات فضایی.

## ۱ مقدمه

توسعه داده شد. [۸] در مدل‌بندی نسبت‌های نوزاد فوت‌شده قبل و بعد از تولد، ضرایب عوامل خطر را بسته به محل سکونت مادر متغیر در نظر گرفت و از یک مدل ضریب متغیر فضایی<sup>۳</sup> برای تحلیل داده‌ها استفاده کرد. [۶] برای مدل‌بندی پذیرش نوآوری‌های فنی در بخش کشاورزی برزیل یک مدل رگرسیون خطی فضایی در نظر گرفتند و نشان دادند که مقادیر پارامترهای مدل بسته به محل کشاورزی متفاوت هستند.

در اینجا مدل‌های تعمیم‌یافته‌ای مطرح می‌شوند که ضرایب آن‌ها به صورت توابعی هموار از متغیرهایی دیگر تغییر می‌کنند که تعدیل‌کننده اثر<sup>۴</sup> نامیده می‌شوند. فرض کنید توزیع متغیر پاسخ  $Y$  به پارامتر

$$\eta = \beta_0 + X_1\beta_1(R_1) + \dots + X_p\beta_p(R_p) \quad (1)$$

وابسته است که در آن متغیرهای  $R_1, \dots, R_p$  ضرایب متغیرهای تبیینی  $X_1, \dots, X_p$  را از طریق توابع نامشخص  $\beta_1(\cdot), \dots, \beta_p(\cdot)$  تعدیل می‌کنند. در مدل ضریب متغیر (۱)، وابستگی بین  $\beta_j(\cdot)$  و  $R_j$  نشان‌دهنده نوع خاصی از اثر متقابل بین هر  $R_j$  و  $X_j$  است که در بعضی از حالات قابل تشخیص نیستند و در حالاتی دیگر می‌توان  $R_j$  را یک متغیر خاص مانند زمان در نظر گرفت.

یک حالت عمومی از کاربرد این نظریه در رده مدل‌های خطی تعمیم‌یافته است که در آن  $\eta$  پیش‌گوی خطی است و به میانگین  $\mu = E(y)$  از طریق یک

در بسیاری از کاربردها از مدل‌های رگرسیونی برای توضیح متغیر پاسخ بر اساس متغیرهای تبیینی در منطقه‌ای مانند  $D$ ، با پذیره وابستگی فضایی استفاده می‌شود. اغلب فرض می‌شود ضرایب رگرسیونی مدل در سراسر منطقه ثابت هستند؛ اما گاهی در عمل ضرایب در سطوح محلی یا منطقه‌ای در فضای مورد مطالعه تغییر می‌کنند. [۷] و [۵] مدل‌های رگرسیونی محلی را در زمینه مدل‌سازی فضایی عوامل مؤثر بر قیمت مسکن در سطح ناحیه مورد بررسی قرار دادند، اما همه این نویسندگان فرض نسبتاً محدودکننده‌ای را برای ضرایب در نظر گرفته‌اند که در واحدهای فضایی مختلف، ثابت هستند. سطح این واحدها با استفاده از متغیرهای تبیینی به صورت اتورگرسیو شرطی مدل‌سازی می‌شود که در آن تعیین مقیاس، همواری سطوح و سطح مکان‌های مجزا از مسائل مورد توجه آن است. اگر بتوان در مدل‌بندی داده‌های فضایی به ضرایب رگرسیونی اجازه تغییر برحسب موقعیت فضایی آن‌ها داده شود و برای هر کدام یک سطح فضایی منظور شود، مدل‌هایی کارا تر حاصل خواهند شد.

گاهی در مدل‌بندی داده‌های فضایی، ضرایب رگرسیونی برای متغیرهای تبیینی در طول زمان نیز تحت تأثیر عوامل محیطی تغییر می‌کنند. یک رهیافت مدل‌سازی انعطاف‌پذیر، برای مدل رگرسیون با ضریب متغیر توسط [۱۳]

<sup>۱</sup> گروه آمار، دانشکده علوم ریاضی، دانشگاه تربیت مدرس

<sup>۲</sup> گروه آمار، دانشکده علوم ریاضی، دانشگاه تربیت مدرس (نویسنده مسئول) mohsen\_m@modares.ac.ir

<sup>۳</sup>Spatial Vrying Coefficient

<sup>۴</sup>Effect Modifiers

گسترش پایگاه‌های داده‌های فضایی در سال‌های اخیر استفاده از مدل‌های بقای فضایی نیز به شدت مورد توجه قرار گرفته است. دلیل آن هم نقش مؤثر و مهم اطلاعات مکانی در پیشگویی تابع بقا است. [۱۱] زمان بقای بیماران مبتلا به سرطان خون را با توجه به همبستگی فضایی میان زمان‌های بقا مورد بررسی قرار دادند. [۹] زمان بقای فضایی افرادی که هم‌زمان مبتلا به چند سرطان بودند را از طریق مدل‌های بقای بیز سلسله‌مراتبی مدل‌بندی کردند. [۲] مدل‌های شکنندگی و خطرهای متناسب کاکس در برازش داده‌های بقا با وجود عوامل خطر ناشناخته را بررسی و کارایی این دو مدل هنگامی که منبع اثر عوامل خطر ناشناخته همبستگی فضایی داده‌های بقا است، مورد بررسی قرار دادند. [۳] نحوه کاربست مدل بقای فضایی برای تحلیل داده‌های مربوط به زمان ابتلا به بیماری سرکوسپوریوز در باغات زیتون را نشان دادند. [۱] مدل‌های بقای کاکس و شکنندگی را برای تحلیل داده‌های سرطان مری بکار گرفتند. [۱۰] یک مدل بیز سلسله‌مراتبی را برای لحاظ کردن اثر همبستگی فضایی-زمانی در داده‌های سرطان سینه در ایالت آیووا را مورد مطالعه قرار دادند.

### ۳ معرفی مدل‌ها

فرض کنید متغیر پاسخ  $d_{\epsilon}$  تعداد مرگ و میر گروهی از بیماران مبتلا به سرطان مری در منطقه مورد مطالعه است که توسط یک مدل خطی تعمیم یافته با توزیع احتمال پواسون به صورت  $d_{\epsilon} \sim Poisson(\mu_{\epsilon})$  مدل‌بندی شده است که در آن  $\mu_{\epsilon}$  تعداد مرگ و میر مورد انتظار است که در یک مدل بقای نسبی به صورت  $\mu_{\epsilon} = d_{\epsilon}^* + y_{\epsilon} \exp(\eta_{\epsilon})$  مدل‌بندی می‌شود، به طوری که  $d_{\epsilon}^*$  نشان‌دهنده تعداد مرگ و میرهای غیرسرطانی مورد انتظار،  $y_{\epsilon}$  زمان در خطر فرد و  $\eta_{\epsilon}$  خطر مازاد<sup>۵</sup> است. این خطر نیز از یک مدل پیروی می‌کند که انواع آن در ادامه معرفی می‌شود. تعداد مرگ و میر مورد انتظار بیماران غیر سرطانی بر اساس نرخ مرگ و میر جامعه در ناحیه مورد نظر محاسبه می‌شود. **الف- مدل رگرسیون کلاسیک:** مدل رگرسیون بر حسب خطر مازاد به صورت

$$\eta_{\epsilon} = \alpha + \beta'X + v_i, \quad i = 1, \dots, r; \quad (5)$$

مدل‌بندی می‌شود که در آن  $\alpha$  عرض از مبدأ،  $\beta$  بردار  $p \times 1$  بعدی ضرایب رگرسیونی،  $X$  بردار  $p \times 1$  بعدی از متغیرهای تبیینی،  $v_i$  مانده‌های تصادفی و  $r$  تعداد مناطق در ناحیه مورد مطالعه است. برای تحلیل بیزی این مدل، با فرض آنکه عبارت  $v_i$  دارای توزیع نرمال با میانگین صفر و پارامتر دقت  $\tau_v$  است، برای پارامترهای  $\alpha$ ،  $\beta$  توزیع‌های پیشین نرمال با میانگین صفر و پارامترهای دقت  $\tau_{\alpha}$  و  $\tau_{\beta}$ ، در نظر گرفته شده است که هر یک از ابرپارامترها دارای توزیع  $\Gammaamma(1, 0/5)$  هستند. در این صورت از طریق فرم نمایی اجزای

تابع پیوند، مانند  $\eta = g(\mu)$ ، مرتبط است. در ساده‌ترین حالت  $Y$  دارای توزیع نرمال با میانگین  $\mu = g(\mu) = \eta$  است. مدل ضریب متغیر (۱) را می‌توان به صورت

$$Y = \beta_0 + X_1\beta_1(R_1) + \dots + X_p\beta_p(R_p) + \epsilon \quad (2)$$

نوشت که در آن  $E(\epsilon) = 0$  و  $Var(\epsilon) = \sigma^2$ . مدل لگاریتم خطی یک مدل رایج دیگر است که در آن  $\eta = \log \mu$  و  $Y$  دارای توزیع پواسون است. مدل لوژیستیک خطی نیز دارای تابع پیوند

$$\eta = g(\mu) = \log(\mu/(1-\mu)) \quad (3)$$

است که در آن متغیر پاسخ  $Y$  دارای توزیع دو جمله‌ای است. پارامترهای  $\beta_1(\cdot), \dots, \beta_p(\cdot)$  در مدل (۱) با مینیمم کردن عبارت میانگین توان دوم خطای  $E[Y - \sum_{j=1}^p X_j\beta_j(R_j)]^2$  برآورد می‌شوند. با اعمال شرط بر روی هر  $R_j$  شروط کافی برای حل مسئله به صورت

$$E[X_j\{Y - \sum_{k=1}^p X_k\beta_k(R_k)\} | R_j] = 0 \quad j = 1, \dots, p,$$

هستند که جواب آن‌ها برای هر  $\beta_j(\cdot)$  به صورت

$$\beta_j(R_j) = \frac{E[X_j\{Y - \sum_{k \neq j} X_k\beta_k(R_k)\} | R_j]}{E(X_j^2 | R_j)} = \frac{E[X_j\{Y - \sum_{k \neq j} X_k\beta_k(R_k)\} / X_j | R_j]}{E(X_j^2 | R_j)} \quad (4)$$

حاصل می‌شود [۱۴]. عبارت (۴) یک نسبت از دو امیدریاضی شرطی است و می‌تواند به صورت یک میانگین موزون شرطی در نظر گرفته شود که در آن وزن‌های شرطی توسط عبارت  $X_j^2$  موزون شده‌اند. عبارت مخرج کسر (۴) این اطمینان را می‌دهد که وزن‌ها منجر به انتگرالی برابر یک شوند. برای هر تابع  $\beta_j(\cdot)$ ،  $j = 1, \dots, p$ ، یک عبارت مشابه وجود دارد و مجموعه این  $p$  عبارت باید به صورت هم‌زمان برای  $\beta_j(\cdot)$  حل شود. در این صورت می‌توان هر تابع  $\beta_j(\cdot)$  را با یک روش تکراری، با هموارسازی  $\{Y - \sum_{k \neq j} X_k\beta_k(R_k)\} / X_j$  روی  $R_j$  با وزن‌های  $X_j^2$  برآورد کرد.

### ۲ داده‌های بقای فضایی

بسیاری از روش‌های معمول تحلیل داده‌های بقا، بر اساس استقلال داده‌ها است؛ اما این فرض همیشه برقرار نیست. اگر همبستگی داده‌های بقا به موقعیت یا فاصله بین واحدهای آماری مرتبط باشد، داده‌های بقای فضایی نامیده می‌شوند. با پیشرفت و گسترش ابزارهای ثبت اطلاعات مکانی و

<sup>5</sup>Excess Risk

<sup>6</sup>Relative Excess Risk

سپس هم‌گرایی زنجیرها با استفاده از نمودار اثر یا روش تشخیص هم‌گرایی گلن-روبین<sup>۸</sup> مورد بررسی قرار می‌گیرد. تصادفی بودن مجموعه داده‌ها نیز از طریق نمودارهای خودهمبستگی داده‌ها در تأخیرهای مختلف بررسی و فاصله<sup>۹</sup> مناسب برای انتخاب مقادیر زنجیر مشخص می‌شود.

برای ارزیابی مدل‌های برازش داده‌شده از ملاک اطلاع انحراف، به صورت  $DIC = \overline{D(\theta)} + p_D$  استفاده می‌شود که در آن  $\overline{D(\theta)}$  میانگین پسینی انحراف،  $D(\theta) = -2 \log f(y|\theta)$  انحراف و  $p_D$  تعداد پارامترهای مؤثر در مدل است. مقدار  $p_D$  با  $\overline{D(\theta)} - D(\hat{\theta})$  تقریب زده می‌شود که در آن  $\hat{\theta}$  برآورد پارامتر  $\theta$  است [۱۲]. مقادیر کوچک این ملاک نشان‌دهنده برازش بهتر مدل است.

## ۴ تحلیل داده‌های سرطان مری

خاک شمال ایران از خاک‌های آتش‌فشانی است که دارای غلظت نترات بسیار بالایی است. این ماده از طریق محصولات کشاورزی و آب وارد چرخه تغذیه‌شده و ابتلا به سرطان را در این مناطق افزایش داده است. افزایش بسیار نگران‌کننده سرطان معده و مری در شمال ایران باعث شده در هر ۱۰۰ هزار نفر ۵۰ مرد و زن به این بیماری مهلک مبتلا شوند. در این مقاله زمان بقای افراد مبتلا به سرطان مری در استان گلستان مورد بررسی قرار می‌گیرد. در این مطالعه اطلاعات ۲۲۳ نفر از ساکنین استان گلستان که طی سال‌های ۱۳۸۶ تا ۱۳۸۷ به سرطان مری مبتلا شده‌اند ثبت و زمان بقا و علت فوت این افراد تا سال ۱۳۹۰ بررسی و گزارش شده است. [۴] عوامل مؤثر بر زمان بقای این افراد را مورد بررسی قرار دادند و برای هر فرد هشت عامل سن، جنس (مرد و زن)، شغل سرپرست خانوار (کارگر و غیر کارگر)، محل زندگی (شهر و روستا)، سطح تحصیلات، قومیت (سیستانی، ترکمن، فارس و سایر)، وضعیت تأهل (مجرد و متأهل) و سطح بیماری (پیشرفته و خفیف) را به‌عنوان متغیرهای تبیینی پیشنهاد کردند.

جدول ۱: برآورد پارامترهای مدل کاکس نیمه پارامتری

عامل	پارامتر	برآورد	انحراف معیار	خطر نسبی	p-مقدار	$R_G^2$
محل زندگی (شهر - روستا)	$\beta_1$	-۰/۳۹۳	۰/۱۸۳	۰/۶۷۵	۰/۰۳۲۴	
شغل (غیر کارگر - کارگر)	$\beta_2$	-۰/۴۲۳	۰/۱۲۶	۰/۶۵۵	۰/۰۰۰۸	۰/۵۰۲۶
سطح بیماری (پیشرفته - خفیف)	$\beta_3$	-۱/۲۶۹	۰/۱۲۶	۰/۲۸۱	< ۰/۰۰۰۱	

مدل رگرسیون، خطر مازاد نسبی<sup>۶</sup> (RER) مرگ‌ومیر توسط مقادیر  $\exp(\alpha)$ ،  $\exp(\beta)$ ،  $\exp(v_i)$  حاصل می‌شوند. **ب-مدل رگرسیون فضایی:** مدل رگرسیون فضایی برحسب خطر مازاد به صورت

$$\eta_i = \alpha + \beta X + u_i + v_i, \quad i = 1, \dots, r, \quad (6)$$

مدل‌بندی می‌شود که در آن  $u_i$  و  $v_i$  به ترتیب اثرهای تصادفی ساختاریافته فضایی و اثرهای تصادفی ساختار نیافته است. در مدل رگرسیون خطی ساده پارامترهای  $\alpha$  و  $\beta$  ثابت فرض می‌شوند در حالی که در اینجا از ناحیه‌ای به ناحیه‌ای دیگر در حال تغییر هستند و هر یک باید  $p$  بار از طریق مینیم کردن میانگین توان دوم خطاها برآورد شوند. برای تحلیل بیزی این مدل، نیز توزیع‌های پیشین برای  $\alpha$ ،  $\beta$  و ابرپارامترها مانند رگرسیون کلاسیک در نظر گرفته می‌شود. برای  $u_i$  یک فرایند اتورگرسیو شرطی<sup>۷</sup> (CAR) با ابرپارامتر دقتی که دارای توزیع  $\text{Gamma}(1, 0.05)$  است در نظر گرفته شده است. در این صورت از طریق فرم نمایی اجزای مدل رگرسیون فضایی، خطر مازاد نسبی مرگ‌ومیر توسط مقادیر  $\exp(\alpha)$ ،  $\exp(\beta)$ ،  $\exp(u_i + v_i)$  حاصل می‌شوند. **ج-مدل ضریب متغیر فضایی:** مدل ضریب متغیر فضایی برحسب خطر مازاد به صورت

$$\eta_i = \alpha + (\beta + \delta_i)X + u_i + v_i, \quad i = 1, \dots, r, \quad (7)$$

در نظر گرفته می‌شود که با اضافه کردن یک اثر تصادفی فضایی  $\delta_i$  برای هر یک از متغیرهای تبیینی در مدل رگرسیون فضایی (۶) باعث انعطاف‌پذیرتر شدن مدل می‌شود. برای تحلیل بیزی این مدل، علاوه بر توزیع‌های پیشین در نظر گرفته شده برای دو مدل قبل، برای اثر تصادفی فضایی  $\delta_i$  در هر ناحیه یک توزیع پیشین اتورگرسیو شرطی گاوسی چند متغیره با ماتریس دقت  $\Sigma$  و توزیع  $\text{Wishart}(Q, k)$  در نظر گرفته می‌شود که در آن  $Q$  یک ماتریس  $k \times k$  است که همبستگی بین  $k$  متغیر را نشان می‌دهد. برای برازش بیزی مدل‌ها با استفاده از زنجیر مارکوفی مونت کارلو (MCMC) و از طریق نرم‌افزار WinBUGS مقادیر تصادفی از توزیع‌های پسین تولید می‌شود.

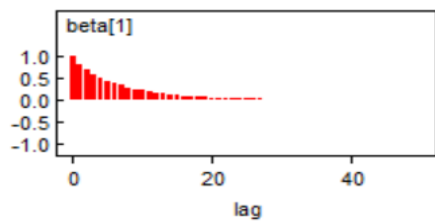
<sup>7</sup>Conditionally Autoregressive

<sup>8</sup>Gelman-Rubin

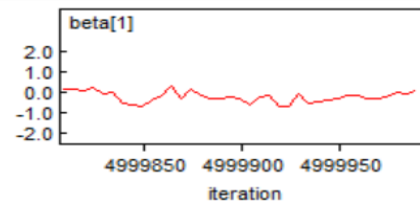
<sup>9</sup>Thin

۱ گزارش شده است. با توجه به ملاک  $R_G^2$  مدل خطرهای متناسب کاکس نیمه‌پارامتری توانسته ۵۰/۲۶ درصد پراکندگی زمان بقای افراد را از طریق متغیرهای تبیینی بیان کند. همان‌طور که ملاحظه می‌شود بیشترین اثر را متغیر محل زندگی با خطر نسبی ۰/۶۷۵ برای افراد داشته است. ریسک خطر افرادی که در روستا زندگی کرده‌اند کمتر از یک بوده است و افراد روستایی ۰/۶۷ ریسک خطر کمتری نسبت به افراد شهرنشین داشته و از بقای بیشتری برخوردار بوده‌اند. هم‌چنین افراد با شغل کارگر ۰/۶۵ ریسک خطر کمتری نسبت به افراد دارای سایر شغل‌ها داشته‌اند و افراد با سطح بیماری خفیف ۰/۲۸ ریسک خطر کمتری نسبت به افراد با سطح بیماری پیشرفته داشته‌اند.

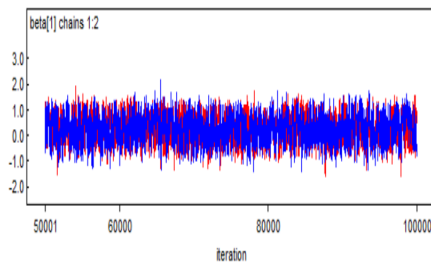
برای تعیین مؤثرترین عوامل یا متغیرهای تبیینی، ابتدا با در نظر گرفتن تمام هشت عامل، مدل خطرهای متناسب کاکس نیمه‌پارامتری به داده‌ها برازش داده شد. پس از حذف متغیرهای تبیینی بی‌اثر، متغیرهای محل زندگی ( $X_1$ )، شغل سرپرست خانوار ( $X_2$ ) و سطح بیماری ( $X_3$ ) که با  $p$ -مقدار کمتر از ۰/۰۵ معنی‌دار تشخیص داده شدند، در ادامه برای تحلیل داده‌ها استفاده شده است. حالت شهری، غیرکارگر و پیشرفته به‌عنوان مرجع انتخاب شدند. تابع خطر به صورت  $h(t|X) = h_0(t)e^{\beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3}$  در آن تابع خطر پایه و  $\beta_1$ ،  $\beta_2$  و  $\beta_3$  پارامترهای اثرات متغیرهای تبیینی منظور شده در مدل هستند. برآورد پارامترهای این مدل در جدول



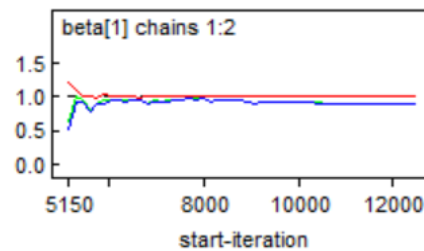
(ب)



(الف)



(د)



(ج)

شکل ۱: الف- نمودار اثر ب- نمودار خودهمبستگی ج- نمودار گلن-روبین د- نمودار اثر بعد از داغیدن، برای پارامتر  $\beta_1$ .

سه مدل رگرسیون کلاسیک، رگرسیون فضایی و رگرسیون ضریب متغیر فضایی با در نظر گرفتن توزیع‌های پیشین مطرح شده در بخش ۳ و با استفاده از نرم‌افزار WinBugs، داغیدن ۵۰۰۰، فاصله ۱۰ و تکرار ۲۰۰۰۰ به داده‌ها برازش داده شده‌اند. تمام اندازه‌های همگرایی، مدل قابل قبولی را برای تمام پارامترها نشان می‌دهند. به‌عنوان نمونه، نمودارهای اثر و خودهمبستگی و نزدیکی مقادیر آماره آزمون گلن-روبین به یک، برای برآورد پارامتر  $\beta$ ، در شکل ۱ بیانگر همگرایی الگوریتم MCMC و سرعت کاهش همبستگی داده‌ها است که بر اساس آن فاصله انتخاب داده‌ها ۲۰ اختیار شده است.

سه مدل رگرسیون کلاسیک، رگرسیون فضایی و رگرسیون ضریب متغیر فضایی با در نظر گرفتن توزیع‌های پیشین مطرح شده در بخش ۳ و با استفاده از نرم‌افزار WinBugs، داغیدن ۵۰۰۰، فاصله ۱۰ و تکرار ۲۰۰۰۰ به داده‌ها برازش داده شده‌اند. تمام اندازه‌های همگرایی، مدل قابل قبولی را برای تمام پارامترها نشان می‌دهند. به‌عنوان نمونه، نمودارهای اثر و خودهمبستگی و نزدیکی مقادیر آماره آزمون گلن-روبین به یک، برای برآورد پارامتر  $\beta$ ، در شکل ۱ بیانگر همگرایی الگوریتم MCMC و سرعت کاهش همبستگی داده‌ها است که بر اساس آن فاصله انتخاب داده‌ها ۲۰ اختیار شده است.

<sup>10</sup>Relative Spatially Varying Coefficient

جدول ۲: مقادیر  $RER$  برای پارامترهای سه مدل رگرسیون کلاسیک، فضایی و ضریب متغیر

مدل	پارامتر	میانگین	انحراف استاندارد	%۲.۵	میان	%۷۵.۵
کلاسیک	$\alpha$	۱/۳۰۴	۰/۱۷۵۵	۰/۹۸۳۲	۱/۲۹۶	۱/۶۶۱
	$\beta_1$	۰/۷۲۰۵	۰/۱۲۴۳	۰/۵۱۱۸	۰/۷۱۰۳	۰/۹۹۱۷
	$\beta_2$	۰/۵۹۸۳	۰/۱۶۰۷	۰/۳۲۵۹	۰/۵۸۴	۰/۹۵۶
	$\beta_3$	۱/۴۲۷	۰/۵۴۶۸	۰/۶۴۷۳	۱/۳۳۳	۲/۷۵۱
فضایی	$\alpha$	۰/۸۹۹۹	۰/۳۶۵۵	۰/۳۴۷۵	۰/۸۴۷۸	۱/۷۵۵
	$\beta_1$	۱/۱۷۸	۰/۷۴۱۹	۰/۳۳۴۱	۰/۹۹۶۴	۳/۱۰۳
	$\beta_2$	۱/۳۱۴	۰/۷۷۸۳	۰/۳۵۱۴	۱/۱۴۱	۳/۳۰۴
	$\beta_3$	۱/۷۴۶	۰/۳۷۱	۰/۳۴۵۲	۱/۴۱۷	۳/۰۳۳
ضریب متغیر	$\alpha$	۰/۸۸۵۶	۰/۱۹۵۱	۰/۴۶۴۹	۰/۸۹۷۳	۱/۲۵۶
	$\beta_1$	۱/۵۹۷	۰/۷۵۷۴	۰/۵۶۸۵	۱/۴۱۵	۳/۱۵۸
	$\beta_2$	۱/۶۰۳	۰/۷۰۲۲	۰/۵۷۴۷	۱/۴۷۳	۳/۰۴۳
	$\beta_3$	۱/۹۸	۰/۸۵۱	۰/۶۰۶۶	۱/۹۵	۳/۷۰۲

جدول ۳: مقادیر  $RER$  برای  $u_i + v_i$ 

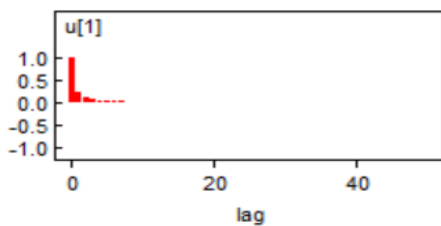
$i$	مدل رگرسیون فضایی	مدل ضریب متغیر
۱	۰/۶۷۳	۰/۷۰۰
۲	۰/۲۲۴	۰/۲۷۸
۳	۰/۸۷۳	۰/۸۸۷
۴	۲/۷۱۴	۲/۴۸۶
۵	۰/۷۵۹	۰/۶۱۸
۶	۰/۷۶۰	۰/۶۳۸
۷	۰/۷۵۹	۰/۵۸۱
۸	۰/۵۴۳	۰/۶۱۹
۹	۱/۶۹۸	۱/۰۲۵
۱۰	۲/۷۱۴	۱/۶۳۵
۱۱	۱/۳۷۲	۰/۹۷۵
۱۲	۱/۲۶۵	۰/۹۵۶

محاسبه شده‌اند. به علاوه ملاک DIC به همراه تعداد پارامترهای مؤثر برآورد شده برای هر سه مدل در جدول ۵ ارائه شده‌اند. همان‌طور که ملاحظه می‌شود،

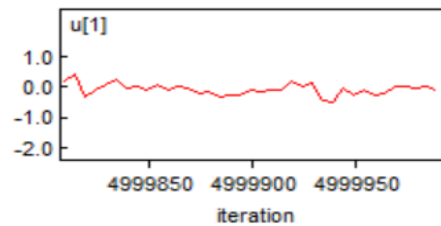
مقادیر میانگین یا برآورد  $RER$  به همراه بازه‌های باورمندی آن‌ها برای سه مدل رگرسیون کلاسیک، رگرسیون فضایی و رگرسیون ضریب متغیر

جدول ۴: مقادیر  $RSVC$  برای  $\delta_i + u_i + v_i$

$i$	پایه	خفیف	روستایی	کارگر	روستایی پیشرفته	پیشرفته کارگر	روستایی کارگر	پیشرفته روستایی کارگر
۱	۰٫۶۷۷	۰٫۷۰۱	۰٫۷۶۵	۰٫۲۹۲	۰٫۸۵۰	۰٫۷۰۶	۰٫۷۵۸	۰٫۸۴۳
۲	۰٫۳۴۸	۰٫۲۰۳	۰٫۱۹۵	۰٫۱۱۶	۰٫۱۹۹	۰٫۱۷۶	۰٫۲۱۰	۰٫۲۲۱
۳	۰٫۸۴۲	۰٫۸۱۵	۰٫۸۳۰	۰٫۳۹۸	۰٫۷۷۰	۰٫۸۴۶	۰٫۸۳۱	۰٫۷۷۴
۴	۲٫۲۵۰	۲٫۳۸۸	۲٫۴۶۰	۱٫۴۷۷	۲٫۳۲۸	۲٫۵۵۱	۲٫۴۶۵	۲٫۳۱۸
۵	۰٫۶۹۰	۰٫۴۵۰	۰٫۷۰۴	۰٫۴۰۱	۰٫۵۲۶۲	۰٫۴۲۵	۰٫۷۰۵	۰٫۵۵۳
۶	۰٫۹۹۸	۱٫۰۱۷۰	۰٫۶۸۰	۰٫۲۵۰	۰٫۷۴۳	۰٫۵۷۸	۰٫۴۸۷	۰٫۵۸۹
۷	۰٫۹۹۳	۱٫۰۱۷	۰٫۶۵۴	۰٫۲۴۶	۰٫۷۴۷	۰٫۵۶۴	۰٫۴۵۸	۰٫۵۵۹
۸	۰٫۶۴۴	۰٫۶۱۹	۰٫۶۶۹	۰٫۲۱۸	۰٫۷۶۴	۰٫۶۰۵	۰٫۶۵۵	۰٫۷۳۱
۹	۱٫۴۲۴	۰٫۹۸۳	۱٫۶۶۱	۰٫۹۲۳	۱٫۱۶۱	۰٫۹۲۲	۱٫۶۸۰	۱٫۲۰۹
۱۰	۱٫۹۹۸	۲٫۵۳۰	۱٫۶۹۲	۰٫۹۲۷	۱٫۸۸۲	۲٫۸۳۱	۱٫۹۳	۲٫۰۳۰
۱۱	۱٫۱۷۹	۰٫۷۹۳	۱٫۲۵۵	۰٫۷۰۲	۰٫۹۳۰	۰٫۷۴۰	۱٫۲۳۴	۰٫۹۷۰
۱۲	۱٫۲۰۶	۱٫۲۸۷	۰٫۸۳۱	۰٫۶۲۰	۰٫۹۳۸	۱٫۳۶۳	۰٫۹۳۶	۱٫۰۳۰



(ب)



(الف)

شکل ۲: الف- نمودار اثر، ب- نمودار خودهمبستگی، برای اثر ساختار یافته  $u_i$ .

در مدل ضریب متغیر بالاتر است. همچنین خطر نسبی مرگ و میر بالاتری در حالت شهری در مدل کلاسیک و حالت روستایی در مدل رگرسیون فضایی و ضریب متغیر و حالت پیشرفته در هر سه مدل مشاهده می‌شود که در نتیجه از بقای کمتری برخوردار هستند؛ بنابراین افراد کارگری که در مناطق روستایی زندگی می‌کنند در مدل کلاسیک از بقای بیشتری برخوردار هستند، اما در مدل رگرسیون فضایی افراد کارگری که در مناطق شهری هستند بقای بیشتری دارند. در حالی که در مدل ضریب متغیر کلاً متفاوت است و افراد غیر کارگر در مناطق شهری بقای بیشتری داشته‌اند.

مقدار ملاک DIC برای مدل رگرسیون فضایی نسبت به مدل کلاسیک و مدل ضریب متغیر کاهش یافته و این مدل بهتر عمل کرده است. مدل رگرسیون فضایی و مدل ضریب متغیر نسبتاً مثل هم عمل کرده و هر دو نسبت به مدل کلاسیک دارای مقدار DIC کمتری هستند. برتری مدل رگرسیون فضایی بر مدل رگرسیون کلاسیک به خاطر همبسته بودن داده‌ها قابل انتظار است. ولی برتری آن نسبت به مدل رگرسیون ضریب متغیر فضایی قابل تأمل است که می‌تواند صرفاً دو حالتی بودن متغیرهای تبیینی و کم بودن تعداد داده‌ها عاملی مؤثر باشد. همان‌طور که در جدول ۵ ملاحظه می‌شود، خطر نسبی مرگ و میر در حالت غیر کارگر در مدل کلاسیک و رگرسیون فضایی و حالت کارگر

جدول ۵: میانه و بازه باورمندی ۹۵ درصد برای RER سه مدل

عامل‌ها	کد	تعداد	مدل رگرسیون کلاسیک	مدل رگرسیون فضایی	مدل ضریب متغیر فضایی
مرحله بیماری	کارگر	۷۱	۰/۷۱۰(۰/۵۱۲, ۰/۹۹۲)	۰/۹۹۶(۰/۳۳۴۱, ۳/۱۰۳)	۱/۴۱۵(۰/۵۶۸, ۳/۱۵۸)
	غیر کارگر	۱۵۲	۱/۰۰	۱/۰۰	۱/۰۰
محل زندگی	روستا	۱۵۲	۰/۵۸۴(۰/۳۲۵۹, ۰/۹۵۶)	۱/۱۴۱(۰/۳۵۱۴, ۳/۳۰۴)	۱/۴۷۳(۰/۵۷۵, ۳/۰۴۳)
	شهری	۷۱	۱/۰۰	۱/۰۰	۱/۰۰
شغل	خفیف	۴۳	۱/۳۳۳(۰/۶۴۷, ۲/۷۵۱)	۱/۴۱۷(۰/۳۴۵, ۵/۰۳۳)	۱/۹۵(۰/۶۰۷, ۳/۷۰۲)
	پیشرفته	۱۸۰	۱/۰۰	۱/۰۰	۱/۰۰
			۱۱۴/۸۳۴	۷۴/۳۷۷	۷۶/۹۲۲
			۳/۸۴۹	۱۱/۰۱۶	۱۱/۵۵۴
					DIC
					PD

ترکمن، گرگان و کلاله دارای ریسک خطر بالاتری نسبت به افراد شهری با شغل‌های دیگر هستند. بقیه مناطق با داشتن  $RSVC < 1$  افراد روستایی کارگر ریسک خطر کمتری نسبت به افراد شهری غیر کارگر داشته‌اند. شهر گنبد با  $RSVC = 2.46$  بیشترین ریسک خطر را برای افراد روستایی کارگر داشته است. در شکل ۳-۵، در حالی که عامل‌ها دوتایی در نظر گرفته شوند، در حالی که افراد روستایی دارای مرحله بیماری پیشرفته باشند، در شهرهای گنبد، بندر ترکمن و گرگان دارای ریسک خطر بالاتری نسبت به افراد شهری با مرحله خفیف بیماری هستند. بقیه مناطق با داشتن  $RSVC < 1$  افراد روستایی پیشرفته ریسک خطر کمتری نسبت به افراد شهری خفیف داشته‌اند. شهر گنبد با  $RSVC = 2.32$  بیشترین ریسک خطر را برای افراد روستایی با مرحله بیماری پیشرفته داشته است. در شکل ۳-۶، در حالی که عامل‌ها دوتایی در نظر گرفته شوند، در حالی که افراد کارگر دارای مرحله بیماری پیشرفته باشند، در شهرهای گنبد، بندر ترکمن و آق‌قلا دارای ریسک خطر بالاتری نسبت به افراد غیر کارگر با مرحله خفیف بیماری هستند. بقیه مناطق با داشتن  $RSVC < 1$  افراد کارگر پیشرفته ریسک خطر کمتری نسبت به افراد غیر کارگر خفیف داشته‌اند. شهر بندر ترکمن با  $RSVC = 2.83$  بیشترین ریسک خطر را برای افراد روستایی با مرحله بیماری پیشرفته داشته است.

## ۵ بحث و نتیجه‌گیری

مدل رگرسیون ضریب متغیر تعمیمی از مدل‌های رگرسیون تعمیم یافته است که به ضرایب اجازه می‌دهد تا اثر متغیرهای تبیینی در هر عامل تغییر کند. مدل رگرسیون فضایی، نابرابری‌های فضایی کلی را از طریق اثرات تصادفی ساختاریافته و ساختاریافته با فرض اینکه در تمام مناطق ثابت هستند، تعیین

نمودارهای اثر و خودهمبستگی برای اثر ساختاریافته  $u_i$  در شکل ۲ بیانگر همگرایی الگوریتم MCMC و سرعت کاهش همبستگی داده‌ها است که بر اساس آن فاصله انتخاب داده‌ها ۵ اختیار شده است.

نقشه‌های پهنه‌بندی مقادیر  $RSVC = \exp(u_i + v_i + \delta_i)$  مندرج در جدول ۴، در شکل ۳ رسم شده‌اند. همان‌طور که ملاحظه می‌شود در حالی که فقط عامل محل زندگی در نظر گرفته شود، شهرهای گنبد، گرگان، بندر ترکمن و کلاله بیماران روستایی ریسک خطر بالاتری نسبت به بیماران شهرنشین داشته‌اند. ۸ ناحیه دیگر با داشتن  $RSVC < 1$  افراد روستایی ریسک خطر کمتری نسبت به افراد شهرنشین داشته‌اند. در نهایت ریسک خطر افراد روستایی از افراد شهری کمتر است. شهر گنبد با  $RSVC = 2.46$  بیشترین ریسک خطر را برای افراد روستایی داشته است. در شکل ۳-ب، در حالی که فقط عامل مرحله بیماری در نظر گرفته شود، افراد در شهرهای گالیکش، علی‌آباد، گنبد، بندر ترکمن و آق‌قلا در مرحله بیماری پیشرفته ریسک خطر بالاتری نسبت به حالت خفیف دارند. ۷ ناحیه دیگر با داشتن  $RSVC < 1$  افراد با مرحله بیماری خفیف ریسک خطر کمتری نسبت به افراد با مرحله بیماری پیشرفته داشته‌اند. در نهایت ریسک خطر افراد با سطح بیماری خفیف از افراد پیشرفته کمتر است. شهر بندر ترکمن با  $RSVC = 2.53$  بیشترین ریسک خطر را برای افراد با مرحله بیماری پیشرفته داشته است. در شکل ۳-ج، در حالی که فقط عامل شغل بیماران در نظر گرفته شود، افراد کارگر در شهر گنبد ریسک خطر بالاتری نسبت به شغل‌های دیگر دارند. ۱۱ ناحیه دیگر با داشتن  $RSVC < 1$  افراد کارگر ریسک خطر کمتری نسبت به افراد با شغل‌های دیگر داشته‌اند. در نهایت ریسک خطر افراد کارگر از افراد غیر کارگر کمتر است. در شکل ۳-د، در حالی که عامل‌ها دوتایی در نظر گرفته شوند، در حالی که افراد روستایی دارای شغل کارگر باشند، در شهرهای گنبد، بندر

مهم در بقای بیماران دخیل است لازم است تحلیل داده‌های بقا با مدل‌های ضریب متغیر فضایی-زمانی انجام شود.

می‌کند. بنابراین یک تناسب مشابه با داده‌ها توسط مدل ضریب متغیر و مدل رگرسیون فضایی ارائه شد. با توجه به تحقیقات انجام شده در زمینه شیوع سرطان مری در استان گلستان، زمان بقای افراد مبتلا به سرطان مری در این استان با مدل ضریب متغیر فضایی مدل‌بندی شد. طبق بررسی‌های انجام شده سطح بیماری در زمان تشخیص سرطان بسیار مهم است و به میزان پیشرفته بودن مرحله بیماری افزایش و در نتیجه زمان بقای فرد کاهش خواهد یافت. هم‌چنین بیمارانی که کارگر بوده یا در روستا زندگی کرده‌اند بقای بیشتری نسبت به بیماران شهری یا شغل‌های دیگر داشته‌اند. تحلیل بقای بیماران سرطان مری با مدل ضریب متغیر فضایی انجام شد و از آنجا که زمان نیز به‌عنوان یک عامل

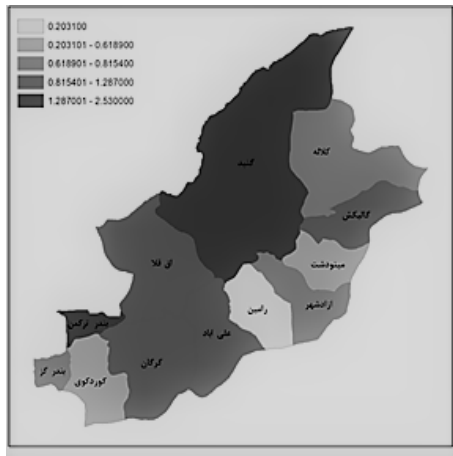
## تقدیر و تشکر

نویسندگان از سردبیر و هیئت تحریریه محترم مجله به دلیل چاپ این ویژه‌نامه به‌افتخار خدمات ارزنده و یادبود استاد فرهیخته جناب آقای دکتر محمدرضا مشکانی تشکر می‌نمایند. از حمایت قطب علمی تحلیل داده‌های وابسته فضایی و فضایی-زمانی دانشگاه تربیت مدرس نیز قدردانی می‌شود.

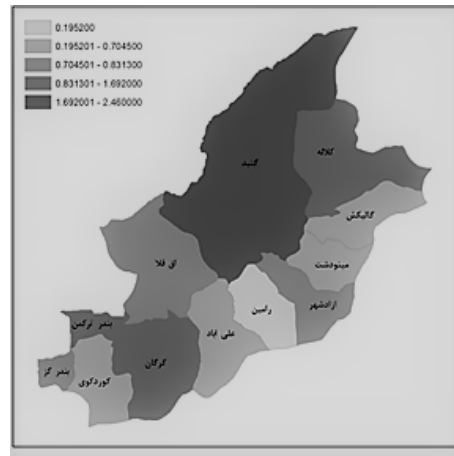
## مراجع

- [۱] آبیاری، آ.، محمدزاده، م.؛ و مترجم، ک. (۱۳۹۵)، مدل‌های بقای کاکس و شکنندگی برای تحلیل داده‌های سرطان مری، نشریه اندیشه آماری، ۲۱(۱)، ۶۴-۵۷.
- [۲] مترجم، ک. محمدزاده، م.؛ و آبیاری، آ. (۱۳۹۴)، مدل‌بندی فضایی داده‌های بقا، پژوهش‌های ریاضی (نشریه علوم دانشگاه خوارزمی)، ۱(۲)، ۶۵-۷۴.
- [۳] مترجم، ک.، محمدزاده، م.؛ و آبیاری، آ. (۱۳۹۴)، عملکرد مدل‌های کاکس و شکنندگی برای داده‌های بقای فضایی، مجله مدل‌سازی پیشرفته ریاضی، ۲(۲)، ۱۱۱-۱۱۷.
- [۴] نورافکن، ز.، یاور، پ.، روشندل، غ.، خلیلی، د.، بهنام‌پور، ن.؛ و زائری، ف. (۱۳۹۲)، برآورد میزان بقای مبتلایان به سرطان مری و برخی عوامل مرتبط با آن در استان گلستان در سال ۱۳۸۷، مجله اپیدمیولوژی ایران، ۹، ۱۱-۱۸.
- [5] Agarwal, D. K., Gelfand, A. E., Sirmans, C. F., and Thibadeau, T. G. (2003), Nonstationary Spatial House Price Models, *Journal of Statistical Planning and Inference*.
- [6] Assuncao, R.M., Assuncao, J. J. and Lemos, M. B. (1998), Induced Technical Change: A Bayesian Spatial Varying Parameter Model. In *Proceedings of XVI Latin American Meeting of the Econometric Society*. Catholic University of Peru: Lima.
- [7] Assuncao, J. J., Gamerman, D., and Assungao, R. M. (1999), Regional Differences In Factor Productivities of Brazilian Agriculture, A Space-Varying Parameter Approach, *Technical Report*, Universidade Federal do Rio de Janeiro, Statistical Laboratory.
- [8] Congdon, P. (1998), A Multilevel Model for Infant Health Outcomes: Maternal Risk Factors and Geographic Variation. *Journal of the Royal Statistical Society, Series D*, **47**, 159-182.
- [9] Diva, U., Banerjee, S. and Dey, D. K. (2007), Modeling Spatially Correlated Survival Data for Individuals with Multiple Cancers, *Statistical Modeling*, **7**, 191-213.
- [10] Hanson, T., Jara, A. and Zhao, L. (2012), A Bayesian Semi-parametric Temporally Stratified Proportional Hazards Model with Spatial Frailties, *Bayesian Analysis*, **7**, 147-188.
- [11] Henderson, R., Shimakura, S. and Gorst, D. (2002), Modeling Spatial Variation in Leukemia Survival Data, *Journal of the American Statistical Association*, **97**, 965-972.

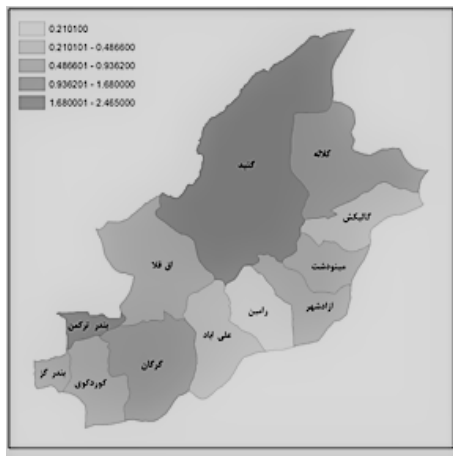




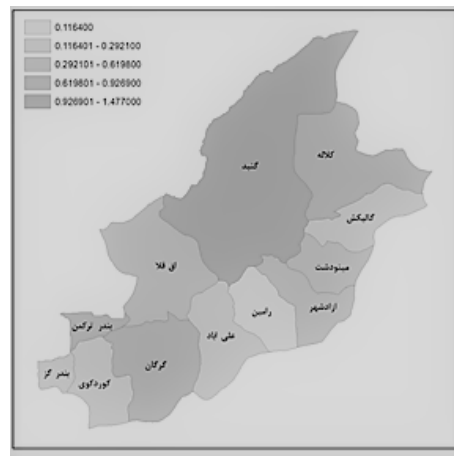
(ب)



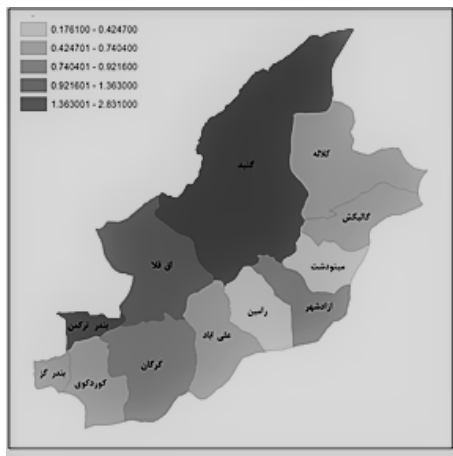
(الف)



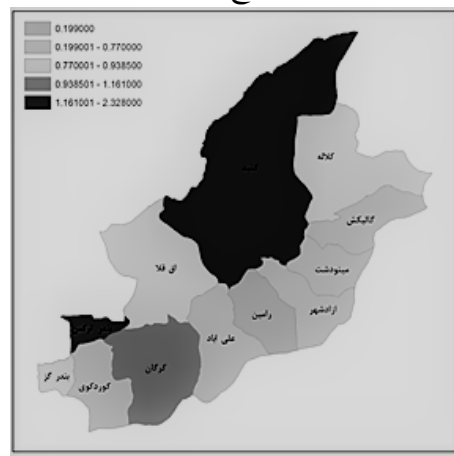
(د)



(ج)



(ر)



(ه)

شکل ۳: پهنه‌بندی برآورد ضریب متغیر فضایی نسبی  $RSVC$  برای مدل رگرسیون ضریب متغیر در حالت الف- روستایی، ب- پیشرفته، ج- کارگر، د- روستایی کارگر، ه- روستایی پیشرفته، ر- پیشرفته کارگر.

[12] Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and Van DerLinde, A. (2002), Bayesian Measures of Model Complexity and Fit (with discussion), *Journal of the Royal Statistical Society*, Series B, **64**, 583-639.

- [13] O'Hagan, A. and Gamerman, D. (1991), Dynamic Bayesian Models for Survival Data, *Applied Statistics*, Serie B, **40**, 63-79.
- [14] West, M., Harrison, P. J. and Migon, H. S. (1985), Dynamic Generalized Linear Models and Bayesian Forecasting (with discussion), *Journal of the American Statistical Association*, **80**, 73-97.