

## کاربرد کریگینگ در تهیه نقشه آماری بیماری سل ریوی

دکتر محسن محمدزاده<sup>۱</sup>، دکتر انوشیروان کاظم‌نژاد، دکتر سقراط فقیه زاده، دکتر یدالله... واقعی

### چکیده مقاله

**مقدمه.** معمولاً نقشه‌های بیماری بر اساس اطلاعات حاصل از میزان بروز یک بیماری در تعدادی از مناطق تهیه می‌شوند. عوامل مختلفی از جمله خطاهای اندازه‌گیری و تغییرات سریع میزان بیماری در مناطق مختلف موجب بروز اغتشاش در نقشه بیماریها می‌شود، بگونه‌ای که تعبیر و تفسیر آنها را بعضاً با مشکل اساسی مواجه می‌سازد. برای رفع این مشکل لازم است نقشه‌ها با استفاده از روشهای تخمین میزان بیماری مناطق مختلف هموار شوند.

**روشها.** نظر به اینکه میزانهای بیماری در مناطق مختلف اغلب دارای همبستگی فضایی هستند، در این مقاله نحوه تعیین ساختار همبستگی میزانهای بیماری از طریق تابع تغییرنگار بیان و روش کریگینگ بعنوان بهترین تخمین‌کننده خطی نارایب تشریح شده و با استفاده از آن نقشه آماری هموار میزان بروز بیماری سل ریوی کشور ارائه خواهد شد.

**نتایج.** نقشه آماری هموار بیماری سل ریوی، که بر اساس میزانهای بروز سل ریوی اسمیر مثبت و به روش کریگینگ تهیه شده، روند تغییرات جغرافیایی بیماری را در سطح کشور نشان می‌دهد. بر اساس این نقشه مناطق مرکزی و غربی کشور کمترین میزان بروز بیماری را داشته و بتدریج با نزدیک شده به سمت مرزهای شرقی کشور بر مقدار آن افزوده می‌شود.

**بحث.** در این مقاله روش آماری کریگینگ برای تهیه نقشه آماری بیماریها مورد مطالعه قرار گرفته و برای نمایش کاربرد آن میزانهای بروز بیماری سل ریوی برای ۲۶۲ شهرستان ایران مورد بررسی قرار گرفته است. نظر به اینکه داده‌های سل ریوی اسمیر مثبت، که توسط اداره کل مبارزه با بیماریها جمع‌آوری می‌شود، ممکن است دارای کم‌ثباتی باشد، نتایج عددی و نقشه‌های حاصل تحت تاثیر این پدیده می‌باشند. در عین حال این روش در حیطه‌های مختلف علوم پزشکی مانند همه‌گیری‌شناسی جغرافیایی بیماریها، بهداشت محیط و مهندسی محیط زیست می‌تواند کاربردهای فراوانی داشته باشد.

● واژه‌های کلیدی: نقشه آماری، کریگینگ، تغییرنگار، ناپستایی، سل ریوی.

### مقدمه

معمولاً در همه‌گیری‌شناسی از نقشه‌های بروز یا شیوع بیماریها برای نمایش توزیع جغرافیایی و تجزیه و تحلیل بیماری در بهداشت ملی - منطقه‌ای استفاده می‌شود. یکی از قدیمی‌ترین موارد استفاده از نقشه در همه‌گیری‌شناسی، مطالعه جان اسنو (۱) درباره همه‌گیری وبا در شهر لندن

می‌باشد، که با استفاده از ترسیم موقعیت مرگ افراد و موقعیت جغرافیایی پمپ‌های تأمین آب، تأثیر منابع تأمین آب در انتشار وبا را مورد بررسی قرار داد. نقشه‌های بیماری برای دستیابی به تغییرات جغرافیایی کمبودهای بهداشتی و در نتیجه اختصاص منابع مالی و انسانی و تنظیم سیاستهای بهداشتی مورد استفاده قرار می‌گیرند. همچنین برای مطالعه ارتباط بین بروز بیماری با متغیرهای توضیحی مانند عوامل اقتصادی، اجتماعی و فرهنگی از طریق رگرسیون فضایی (مورد Spatial) بیماری، (Regression) و تحت عنوان تجزیه و تحلیل اکولوژیکی (Ecological Analysis) استفاده قرار می‌گیرند (۳ و ۲). نقشه‌های معمولی میزان بروز یک معمولاً براساس اطلاعات موجود در بعضی یا همه مناطق (موقعیتها) تهیه می‌شوند. این گونه نقشه‌ها در صورت عدم وجود میزان بیماری در بعضی مناطق یا تغییرات خطاهای اندازه‌گیری، ممکن است دارای ناهمواری و نوسانات زیادی باشند به گونه‌ای که تعبیر و تفسیر آنها را دشوار می‌نماید (۴). لذا برای رفع این نواقص لازم است، نقشه آماری به گونه‌ای تهیه شود که هموار و فاقد نوسانهای ناگهانی بوده و بتوان از آن برای تخمین میزان بیماری در مناطق فاقد اطلاع استفاده نمود.

در سه دهه گذشته روشهای مختلفی به منظور تخمین مقدار یک متغیر مانند میزان بیماری در موقعیتی مشخص با استفاده از مشاهدات واقع در (۶ و ۵) که موقعیتهای مجاور توسعه یافته است، که مهمترین آنها عبارتند از: روش عکس فاصله موزون (۵ و ۹)، (Inverse Distance - Weighted) در (۷ و ۸) برای تهیه نقشه آماری بیماری مورد استفاده قرار گرفته است، روش سطوح روند چند جمله‌ای (۱۰ و ۱۱) و (Trend polynomial) Surfaces روش اسپلاین‌ها بیماری بکار برده (Splines) روشهای مبتنی بر بیز (۱۲، ۱۳، ۱۴ و ۱۵). روش کریگینگ از جمله فنون آماری است که برای تخمین فضایی و تهیه نقشه آماری می‌شود و علیرغم روشهای مذکور، ساختار همبستگی فضایی میزانهای بیماری را در تجزیه و تحلیل آنها مورد استفاده قرار می‌دهد. بعلاوه با استفاده از این روش می‌توان در هر موقعیت دلخواه انحراف معیار تخمین میزان بیماری را نیز تعیین و در صورت نیاز نقشه آن را تهیه نمود (۱۶ و ۱۷).

داده‌هایی که نوعاً برحسب موقعیت قرار گرفتن آنها در فضای مورد مطالعه همبسته باشند و این همبستگی تابعی از فاصله موقعیت آنها باشد،

۱ - استادیار گروه آمار، دانشگاه تربیت مدرس

تابع کوواریانس تعمیم یافته (Generalized Covariance Function) داده‌های بدون روند شده استفاده کرد (۲۰).

تخمین فضایی در یک موقعیت مشخص  $(x, y, z)$  و براساس داده‌های  $z(t_1), \dots, z(t_n)$  که به نام دی.جی. کریگ، مهندس معدن آفریقایی جنوبی کریگینگ نامیده شد (۱۹)، یک ترکیب خطی از داده‌ها به صورت  $\sum_{i=1}^n y_i z(t_i)$  است، که ضرایب آن به گونه‌ای تعیین می‌شوند تا این تخمین کننده نااریب و دارای کمترین واریانس باشد (۲۱ و ۱۷). در محاسبه کریگینگ معمولاً برآورد ضرایب  $\lambda$  به گونه‌ای خواهند بود که با مشاهدات نزدیک وزن بیشتر و به مشاهدات دور وزن کمتر اختصاص داده می‌شود. این امر به تبعیت از ماهیت داده‌های فضایی است که با افزایش فاصله، همبستگی آنها کاهش می‌یابد.

در این مقاله برای تهیه نقشه آماری میزان بروز سل ریوی اسمیر مثبت، نقشه ایران به مربع‌هایی به ضلع  $28/8$  کیلومتر تقسیم شده و در هر گره، یعنی راس هر یک از این مربع‌ها تخمین میزان بروز سل ریوی اسمیر مثبت به روش کریگینگ و انحراف معیار آنها نیز محاسبه می‌شوند. سپس با ترسیم توأم دو نمودار تصویری (Image plot) و تراز (Contour plot) و اضافه کردن مراکز شهرستانها و مرزهای کشور به آن، نقشه مورد نظر برای میزان بیماری و انحراف معیار آن تهیه و ارائه می‌شود. انجام محاسبات شامل برازش مدل تغییرنگار و تخمین‌های کریگینگ و تهیه نمودارهای لازم با برنامه‌نویسی در محیط نرم‌افزار آماری SPLUS و با استفاده از مجموعه توابع Geos (۲۲) انجام شده است.

### نتایج

مختصات جغرافیایی مراکز شهرستانها نسبت به یک مبدأ واقع در جنوب غربی کشور به صورت  $t_i = (x_i, y_i)$  بر حسب کیلومتر از نقشه تقسیمات کشوری سال ۱۳۷۸ استخراج شده و از تقسیم تعداد موارد جدید سالیانه سل ریوی اسمیر مثبت بر برآورد جمعیت هر یک از ۲۶۲ شهرستان میزان بروز بیماری سل ریوی اسمیر مثبت به ازای صدهزار نفر محاسبه و با  $z(t_i)$  نشان داده شده است.

بدلیل وجود روند فضایی در مشاهدات با برازش یک چند جمله‌ای درجه دو به فرم  $\mu(t_i) = a + a_1 x_i + a_2 y_i + a_3 x_i^2 + a_4 y_i^2 + a_5 x_i y_i$  داده‌های بدون روند شده بصورت تعیین  $R(t_i) = z(t_i) - \mu(t_i)$  تعیین و تغییرنگار کلاسیک همسانگرد آنها به تبعیت از (۲۳) در ۱۸ لگ محاسبه شده است. برای برازش یک مدل مناسب به تغییرنگار کلاسیک از بین مدل‌های نظری معتبر ارائه شده در (۲۴) یکی را به عنوان بهترین مدل انتخاب نموده‌ایم. برای این منظور از ملاک مجموع مجذورات موزون (WSS) باقیمانده‌های مدل به صورت  $WSS = \sum_{i=1}^{18} w_i (y_i - \hat{y}_i)^2$  استفاده می‌شود، که در آن  $\hat{y}_i$  و  $y_i$  به ترتیب مقادیر مدل نظری و برآورد کلاسیک تغییرنگار در لگ نام و وزن  $w_i$  مطابق (۱۷) به صورت  $N_i$ ، تعداد زوج‌های متمایز در لگ نام می‌باشد. جدول ۱

داده‌های فضایی (Spatial Data) نامیده می‌شوند. چون میزانهای بروز بیماری در مناطق مختلف اغلب از یک ساختار همبستگی پیروی می‌کنند، که به موقعیت آنها بستگی دارد، نوعی داده فضایی هستند و این ساختار همبستگی باید در تجزیه و تحلیل آنها لحاظ گردد در این مقاله نحوه تعیین ساختار همبستگی داده‌ها و استفاده از روش کریگینگ برای تخمین فضایی میزانهای بیماری بیان شده و داده‌های سل ریوی اسمیر مثبت سال ۱۳۷۸ شهرستانهای کشور که توسط اداره کل مبارزه با بیماریها جمع‌آوری شده است، به عنوان یک مطالعه موردی و نمایش فنون و روشهای ارائه شده، مورد بررسی قرار می‌گیرند.

سل ریوی مهمترین و شایع‌ترین نوع بیماری سل است، که میکروب آن از طریق تنفس در ریه شخص وارد می‌شود. یکی از راههای تشخیص این بیماری آزمایش مستقیم خلط است، که نتیجه مثبت این آزمایش، اسمیر مثبت نامیده شده و می‌تواند نشان دهنده ابتلاء فرد به بیماری سل ریوی باشد، اما نتیجه منفی آزمایش دلیل بر سالم بودن فرد نیست. سل ریوی اسمیر مثبت بخش مهمی از موارد بیماری سل را تشکیل می‌دهد، زیرا این بیماران منابع انتقال عفونت سل به دیگران هستند و از لحاظ همه‌گیری‌شناسی و کنترل بیماری حائز اهمیت می‌باشند. لذا برآورد میزان بروز سالانه سل ریوی اسمیر مثبت از جمله مهمترین معیارهای برآورد وضع بیماری سل در هر جامعه می‌باشد (۱۸).

### روشها

داده‌های فضایی معمولاً با نماد  $z(t_1), \dots, z(t_n)$  نمایش داده می‌شوند، که در آن  $N$  تعداد داده‌ها و  $t_i = (x_i, y_i)$  طول و عرض متناظر با موقعیت جغرافیایی مشاهده نام (در فضای دو بعدی) است. ساختار همبستگی داده‌های فضایی با تغییرنگار (Variogram)،

$$\gamma_y(h) = \text{var}(z(t) - z(t+h))$$

تعیین می‌شود. چون این تابع در عمل نامعلوم است، از برآودگر کلاسیک تغییرنگار به صورت

$$\hat{\gamma}_y(h) = \frac{1}{|N(h)|} \sum_{i=1}^{|N(h)|} (z(t_i) - z(t_i+h))^2$$

استفاده می‌شود که در آن  $N(h) = \{(i, j) : t_i - t_j = h\}$  مجموعه تمام زوج داده‌های در فاصله  $h$  و  $|N(h)|$  تعداد زوجهای متمایز این مجموعه می‌باشد (۱۹).

چنانچه داده‌های فضایی دارای روند باشند، نایستا (Nonstationary) خواهند بود و برآودگر کلاسیک تغییرنگار اریب خواهد بود. برای رفع این نقیصه معمولاً تغییرنگار کلاسیک بر اساس داده‌های بدون روند شده محاسبه می‌شود (۱۷). چون تغییرنگار داده‌های اصلی و داده‌های بدون روند شده یکسان نیست و به روش حذف روند از داده‌ها بستگی دارد، می‌توان از

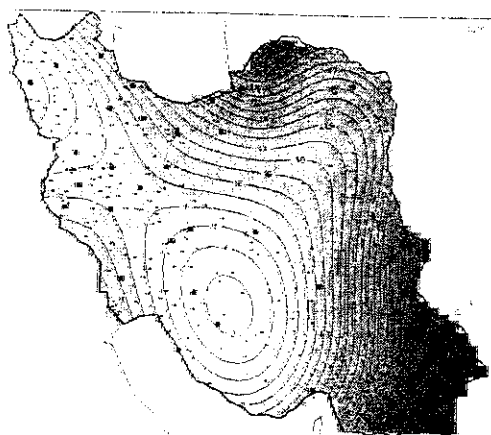
می‌گیرد. مدل مکعبی همسانگرد به صورت زیر می‌باشد که مقادیر برآورد پارامترهای  $c$ ,  $a$ ,  $c$  به ترتیب  $۱۹۶۹/۸۷.۹۵/۳۲.۴۴/۵۴$  می‌باشند.

مقادیر WSS مدل‌های مختلف را نشان می‌دهد. با توجه به اینکه مدل تغییر نگار مکعبی (cubic) از بین هشت مدل معتبر معرفی شده در (۳۴) دارای کمترین مقدار WSS می‌باشد، به عنوان بهترین مدل معتبر مورد استفاده قرار

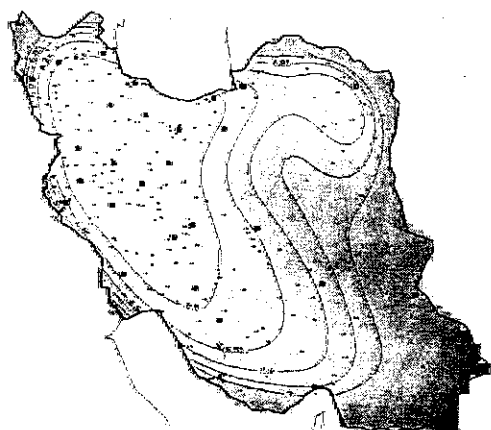
$$\gamma(h) = \begin{cases} 2c + 2c \left[ \sqrt{\left(\frac{|h|}{a}\right)} - 1.75 \left(\frac{|h|}{a}\right)^2 + 3/5 \left(\frac{|h|}{a}\right)^5 - 0.75 \left(\frac{|h|}{a}\right)^7 \right] & \text{اگر } |h| < a \\ 2c + 2c & \text{اگر } |h| \geq a \end{cases} \quad (2)$$

جدول ۱: مجموع مربعات موزون WSS برای مدل‌های تغییر نگار برآزش شده

Cubic	Gneiting	power Exponential	Circular	Spherical	Gaussian	Wave	Matern	Exponential	نام مدل
۵۷/۰۲	۱۵۵/۲۵	۹۴/۲۹	۱۲۶/۵۵	۱۲۲/۸۲	۷۰/۳۲	۵۹/۴۲	۱۰۱/۸۲	۱۰۱/۸۲	WSS



شکل ۱. نقشه میزان بروز سل ریوی



شکل ۲. نقشه انحراف معیار تخمین میزان بروز سل ریوی

با استفاده از مدل مکعبی برآزش شده و روش کریگینگ، میزان بروز سل ریوی اسمیر مثبت و انحراف معیار آن در یک شبکه شطرنجی (با فواصل ۲۸/۸ کیلومتر) تخمین زده شد و در نتیجه نقشه تخمین میزان بیماری و نقشه انحراف معیار آن به کمک نمودارهای تصویری و تراز تهیه شدند. شکل ۱ نقشه میزان بروز بیماری بر اساس مقادیر تخمین و شکل ۲ انحراف معیار آنها را نشان می‌دهند. در نمودار تصویری افزایش تیرگی رنگ نشان دهنده زیاد بودن میزان بروز بیماری (یا انحراف معیار) می‌باشد. نمودار تراز مقدار تخمین میزان بروز بیماری (یا انحراف معیار) در نقاط مختلف کشور را نشان می‌دهد. مثلاً با توجه به شکل ۱ در شهرستان مشهد تخمین میزان بروز سل ریوی اسمیر مثبت ۱۵ در صد هزار نفر می‌باشد و با توجه به شکل ۲ چون مشهد بین خطوط تراز ۶/۹۵ و ۹/۶ واقع شده، انحراف معیار آن حدود ۶/۹۲ می‌باشد. بنابراین افزایش تیرگی رنگ در شکل ۱ از مناطق مرکزی به سمت مرزهای شرقی کشور نشان دهنده کمترین میزان بروز بیماری در مناطق مرکزی است که به تدریج با نزدیک شدن به سمت مرزهای جنوب شرقی و قسمتی از شمال شرق افزایش می‌یابد و بیشترین میزان سل ریوی اسمیر مثبت در مرزهای شرقی استان سیستان بلوچستان و در حدود ۲۸ در صد هزار نفر تخمین زده می‌شود.

همانطور که در شکل ۲ ملاحظه می‌شود، کمترین انحراف معیار مربوط به نیمه غربی کشور است که در این مناطق تجمع شهرستانها بیشتر و در نتیجه داده‌های زیادتری در اختیار هستند. افزایش انحراف معیار تخمین‌ها به سمت شرق می‌تواند ناشی از پراکندگی شهرستانها و در نتیجه کم بودن تعداد داده‌ها و حتی ناشی از تغییرات زیاد بیماری در این مناطق باشد. بطور کلی انحراف معیار تخمین‌ها در سطح کشور ۷ تا ۸ تغییر کرده و نوسان زیادی ندارد و این نشان دهنده آن است که دقت تخمین‌ها در سطح کشور تقریباً یکسان است.

**بحث**

تخمین میزان بیماری در شهرستانهایی که موارد بیماری آنها گزارش نشده یا مخدوش است، از نقاط قوت روش ارائه شده در این مقاله می‌باشد. یکی دیگر از ویژگی‌های روش کریگینگ انطباق تخمین‌ها با مقادیر واقعی مشاهدات است، یعنی مقدار تخمین حاصل از کریگینگ برای هر یک از ۲۶۲ شهرستان مورد مطالعه با میزان بروز بیماری مشاهده شده برای آنها برابر می‌باشد.

**قدردانی و تشکر**

از همکاری صمیمانه واحد سل اداره کل مبارزه با بیماریها که داده‌ها و راهنماییهای فنی در مورد نحوه ثبت و جمع‌آوری آنها را ارائه نموده‌اند و همچنین از آقای دکتر شادرخ ایزدی برای ارائه نظرات تخصصی ارزنده قدردانی و تشکر می‌شود.

در این مقاله روش تخمین فضایی کریگینگ برای تهیه نقشه آماری بیماریها ارائه شد و داده‌های سل ریوی اسمیر مثبت سال ۱۳۷۸ شهرستانهای کشور، به عنوان یک مطالعه موردی، مورد بررسی قرار گرفتند. از جمله نقاط ضعف نتایج حاصل که در نقشه منعکس شده است، مقادیر قابل توجه انحراف معیار تخمین‌ها در سراسر کشور است که دلیل عمده آن معطوف به نحوه جمع‌آوری و دقت داده‌های استفاده شده در این تحقیق می‌باشد که بهبود آن از اراده نگارندگان این مقاله خارج بوده زیرا همانطور که قبلا نیز ذکر شده داده‌های سل ریوی اسمیر مثبت، به علت عدم مراجعه بعضی بیماران به شبکه بهداشت یا عدم تشخیص بیماری تا حدودی تحت تاثیر کم یثتی هستند. اما لحاظ کردن همبستگی فضایی میزان‌های بیماری در تخمین، امکان محاسبه انحراف معیار تخمین در هر نقطه، و امکان

**مراجع**

- 1-Snow J. *On the Mode of Communication of Cholera*. Churchill Livingstone, London, 1954, 2nd edition.
- 2- Lawson A.B. , Cressie N., *Spatial Statistical Methods for Environmental Epidemiology Handbook of statistics*, 2000 18: 357-396.
- 3- Lawson A.B. *Statistical Methods in Spatial Epidemiology* John Wiley, 2001.
- 4- Marshall R.A. *Review of Methods for the Statistical Analysis of Spatial Patterns of Disease Journal of Royal Statistical Society Series A* 1991 , 154, 421-441.
- 5- Ripley, B.D., *Spatial Statistics* New York: John Wiley & Sons 1981 : 44-77.
- 6- Watson, D.F., and Philip G.M. *A Refinement of Inverse Distance Weighted Interpolation Geoprocessing* 1985 , 2: 315-327.
- 7-Pukkala, E. *Cancer Maps of Finland : An Example of Small-Area Based Mapping. Recent Results Cancer Res*, 1989: 114: 208-215
- 8- Glatre, E., *Atlas of Cancer Incidence in Norway 1970-1979. Recent Results Cancer Res*, 1989: 114: 216-226
- 9- Watson, G.S. *Trend Surface Analysis and Spatial Correlation. Geological, Society of America,(special paper)*, 1972, 146: 39-46.
- 10- De Boor C., *A Practical Guide to Splines* New York : Springer-Verlag New York, Inc 1978.
- 11- Linde V.D.A., Witzko K.H. and Jaockel K. H., *Spatial -Temporal Analysis of Mortality Using Splines Biometrics*, 1995 , 51: 1352-1360.
- 12- Clayton, D. and Kaldor, J., *Empirical Bayes Estimates of Age-Standardized Realative Risks for use in Disease Mapping, Biometrics*, 1987, 43: 671-691.
- 13- Clayton, D. and Bernardinelli, L., *Baysian Methods for Mapping Disease Risk; In: Geographical and Environmental Epidemiology: Methods for Small-Area Studies (Eds. Elliott. P., Cuzick, J. English, D. and Stern R.)*, Oxford University Press, Oxford, 1992.
- 14- Bernardinelli, L., Clayton D., Pascutto, C., Montomoli, C, Ghislandi, M. and Songini M., *Baysian Analysis of Space Time Variation in Disease Risk Statistics in Medicine* 1995, 14: 2433-2443.
- 15- Mugglin , A.S, Cressie, N. and Gemmell, I., *Hierarchial Statistical Modeling of Influenza-Epidemic Dynamic in Space and Time, (Tentative title) in preparation*, 2000.

- 16- Carrat F. and Valleron A. J., *Epidemiologic Mapping Using the Kriging Method: Application to an Influenza-Like Illness Epidemic in France*, *American Journal of Epidemiology*, 1992, 135: 1293-1300.
- 17- Cressie, N., *Spatial Statistics for Spatial Data*, New York John Wiley 1993.
- ۱۸- محمدی، منوچهر؛ اپیدمیولوژی بیماری سل از کتاب «اپیدمیولوژی بیماریهای شایع در ایران» دکتر فریدون عزیزی و همکاران، مرکز تحقیقات غدد درون ریز، ۱۳۷۲ صفحه ۳۴۰-۳۱۵.
- 19- Matheron, G., *Principles of Geostatistics*, *Economic Geology* 1963, 58: 1246-1266.
- ۲۰- محمدزاده، محسن، تابع کواریانس تعمیم یافته برای کریگیدن عام؛ اندیشه آماری ۱۳۷۸ شماره دوم، ۱۸-۲۳.
- 21- Webster R. and Oliver M., *Geostatistics for Environmental Scientists*, John Wiley, 2001.
- 22- Ribeiro, Jr.P.J. and Diggle, P.J., *geoR/geoS: Functions for Geostatistical Analysis using R or S-PLUS Technical Report: ST-99-09 Dep. of Maths. And Stats. Lancaster University Lancaster U.K.*, <http://www.maths.lancs.ac.uk/ribeiro/~geos.html>
- ۲۳- محمدزاده، محسن، و واقعی، یدالله، تعداد لگ مناسب در مدلسازی تغییر نگار ارائه شده به مجله علوم دانشگاه تربیت معلم، ۱۳۸۱.
- 24- Barry, J., Crowder, M. and Diggle, P., *Parametric Estimation of the Variogram Technical Report ST-97-06 Dep., Maths. and*