

تعداد لگ مناسب در مدلسازی تغییرنگار نمایی

محسن محمدزاده : دانشگاه تربیت مدرس

یداله واقعی : دانشگاه بیرجند

چکیده

یکی از مسائل مهم در تجزیه و تحلیل داده‌های فضایی، تعیین ساختار همبستگی داده‌ها است که معمولاً به وسیله تابع تغییرنگار صورت می‌پذیرد. برای برآورد این تابع، فاصله مورد مطالعه به تعدادی لگ تقسیم می‌گردد و در هر لگ مقدار تغییرنگار تجربی محاسبه و مدلی معتبر به مقادیر آن برازش داده می‌شود. از آنجا که دقت مدل برازش داده شده به تعداد لگ‌ها بستگی دارد، تعیین تعداد لگ مناسب برای برآورد تابع تغییرنگار نمایی مورد بررسی قرار می‌گیرد و از آن برای مدلسازی تغییرنگار میزان بروز بیماری سل ریوی در شهرستان‌های کشور استفاده می‌شود. سپس با استفاده از آن میزان این بیماری در تعدادی از شهرستانها تخمین زده می‌شود.

مقدمه

داده‌هایی که نوعاً بر حسب موقعیت قرار گرفتن آنها در فضای مورد مطالعه همبسته باشند و این همبستگی تابعی از فاصله موقعیت آنها باشد، داده‌های فضایی نامیده می‌شوند. در آمار فضایی معمولاً میدان تصادفی $\{Z(t), t \in D\}$ ، $D \subset R^d$ برای مدلسازی داده‌ها استفاده می‌شود که در آن D یک مجموعه اندیس گذار و R^d فضای اقلیدسی d بعدی ($d \geq 1$) است. چنانچه برای هر مقدار صحیح m ، متغیرهای تصادفی $Z(t_1), \dots, Z(t_m)$ دارای توزیع توأم گوسی باشند، میدان تصادفی $\{Z(t), t \in D\}$ گوسی نامیده می‌شود. اگر میانگین میدان تصادفی ثابت و به موقعیت t بستگی نداشته باشد، یعنی $E[Z(t)] \equiv \mu$ ، و واریانس عبارت $Z(t) - Z(t+h)$ فقط تابعی از h به صورت

$$(h) \gamma \text{Var}(Z(t) - Z(t+h)) = 2 \quad (1)$$

باشد، میدان تصادفی مانای ذاتی^۱ نامیده می‌شود. تابع $\gamma(h)$ که ساختار همبستگی داده‌های فضایی را تعیین می‌کند تغییرنگار^۲ می‌نامند. بعلاوه اگر برای هر h ، کوواریانس $Z(t)$ و $Z(t+h)$ تنها تابعی از h باشد، میدان تصادفی $Z(t)$ را مانای مرتبه دوم گویند و تابع $C(h) = \text{Cov}(Z(t), Z(t+h))$ که ساختار همبستگی را در یک

میدان تصادفی مانای مرتبه دوم مشخص می‌کند، هم تغییرنگار^۱ می‌نامند. چنانچه هر یک از توابع $\gamma(h)$ یا $C(h)$ فقط تابعی از اندازه h بوده و به جهت آن بستگی نداشته باشد، میدان تصادفی را همسانگرد^۲ نامند. «ماترون»^[۱] نشان داد $\gamma(\cdot)$ باید واجد خاصیت معین منفی شرطی^۳ باشد، یعنی برای هر تعداد متناهی

$$\sum_{i=1}^m a_i = C \text{ با شرط } \{a_i : i = 1, \dots, m\} \text{ و اعداد حقیقی } \{t_i : i = 1, \dots, m\} \text{ در نامساوی}$$

$$\sum_{i=1}^m \sum_{j=1}^m a_i a_j \gamma(t_i - t_j) \leq 0$$

صدق کند. بعلاوه می‌توان نشان داد هم تغییرنگار $C(\cdot)$ نیز یک تابع معین مثبت^۴ است؛ یعنی برای هر تعداد متناهی موقعیتهای $\{t_i : i = 1, \dots, m\}$ داریم:

$$\sum_{i=1}^m \sum_{j=1}^m a_i a_j C(t_i - t_j) \geq 0$$

در عمل، معمولاً تغییرنگار نامعلوم است و باید بر اساس داده‌های فضایی برآورد شود. «ماترون»^[۲] یک برآورد تجربی برای تغییرنگار بر اساس مشاهدات $\{Z(t_i) : i = 1, \dots, n\}$ به صورت

$$2\hat{\gamma}(h) = \frac{1}{|N(h)|} \sum_{N(h)} (z(t_i) - z(t_j))^2, h \in D \quad (۲)$$

ارائه نمود، که در آن $N(h) = \{(t_i, t_j) : t_i - t_j = h; i, j = 1, \dots, n\}$ مجموعه تمام زوج نقاطی است که در جهت h و به فاصله $|h|$ از یکدیگر قرار دارند و $|N(h)|$ تعداد اعضای متمایز آن است. علیرغم آنکه $2\hat{\gamma}(h)$ یک برآوردگر ناریب برای $\gamma(h)$ است، مستقیماً نمی‌توان آن را در روشهای مختلف تجزیه و تحلیل داده‌های فضایی، از جمله کریگینگ که برای تخمین داده‌های فضایی به کار می‌رود، مورد استفاده قرار داد، زیرا این برآوردگر لزوماً خصوصیت معین منفی شرطی بودن تغییرنگار را ندارد. «کرسی»^[۳] نشان داد این نقیصه می‌تواند سبب منفی شدن میانگین مربع خطای پیش‌بینی شود، بنابراین لازم است یک مدل معتبر معین منفی شرطی به تغییرنگار تجربی برآورد داد.

افزایش تغییرنگار بیانگر کاهش همبستگی داده‌ها است. با توجه به این‌که معمولاً بعد از فاصله‌ای همبستگی داده‌های فضایی ناچیز و حتی مستقل از یکدیگر می‌شوند، مدل تغییرنگار از این فاصله به بعد تغییر چندانی ندارد و به صورت تخت درمی‌آید. این فاصله، دامنه^۵ تغییرنگار نامیده می‌شود. تغییرنگار بعد از دامنه، به مقدار ثابتی گرایش می‌یابد که آستانه^۶ نام دارد. در بعضی مدلها مانند مدل نمایی، تغییرنگار در بی‌نهایت به مقدار ثابت آستانه می‌رسد. دامنه این‌گونه مدلها مقداری است که تغییرنگار نزدیک آستانه، مثلاً به ۹۵٪ آن برسد. البته بعضی مدلها مانند مدل خطی یا توانی اصلاً دارای آستانه نیستند. معیار دیگر تغییرنگار، که معمولاً به دلیل خطاهای نمونه‌گیری ظاهر می‌شود و برابر $c_0 = \lim_{h \rightarrow \infty} \gamma(h)$ است، اثر قطعه‌ای^۷ است. از نظر تئوری اثر قطعه‌ای باید صفر باشد، زیرا اگر دو نمونه از نقطه واحدی گرفته شوند باید خصوصیات یکسانی داشته

۱ - Covariogram
 ۴ - Positive Definite

۲ - Isotropic
 ۵ - Range

۳ - Conditional Negative Definite
 ۶ - Sill
 ۷ - Nugget Effect

باشند، اما به دلایلی از جمله خطای اندازه گیری، یا تغییرات شدید خصوصیت مورد بررسی در نقاط نزدیک به هم، مقدار c_0 گاهی مخالف صفر است.

برای برآورد پارامتری تغییرنگار معمولاً مدل‌های مختلفی از جمله مدل‌های خطی، کروی، نمایی، گوسی، توانی و موجی مورد استفاده قرار می‌گیرند [۴]. به عنوان مثال مدل شبه تغییرنگار نمایی در حالت همسانگرد به این صورت است:

$$\gamma(h) = c_0 + c_e \{1 - \exp(-3|h|/a_e)\}, c_0 \geq 0, a_e \geq 0, c_e \geq 0 \quad (3)$$

و در آن c_0 اثر قطعه‌ای $c_0 + c_e$ آستانه و a_e دامنه شبه تغییرنگار، یعنی نقطه‌ای است که شبه تغییرنگار به 0.95 آستانه خود می‌رسد.

برآزش مدل تغییرنگار تا حدی شبیه برازش یک مدل رگرسیونی است، با این تفاوت که در اینجا مقادیر تغییرنگار در لگ‌های^۱ مختلف مشاهدات مستقل نیستند. برای برازش یک مدل تغییرنگار پارامتری می‌توان همانند [۳] از روش‌های حداقل مربعات معمولی، حداقل مربعات موزون یا درستیابی استفاده کرد. برای این کار فاصله مورد مطالعه به چندین لگ تقسیم می‌شود و با استفاده از زوج نقاطی که در فاصله یک لگ قرار می‌گیرند، برآورد تجربی محاسبه می‌گردد، سپس یک مدل تغییرنگار معتبر به آن برازش داده می‌شود. از آنجا که تعداد لگ‌ها در مقدار برآورد تغییرنگار دخالت دارد، در میزان دقت مدل برازش شده نیز تاثیر می‌گذارد؛ زیرا کم بودن تعداد لگ‌ها سبب کم شدن تعداد نقاط مورد استفاده در برازش و زیاد بودن آن سبب کاهش تعداد ازواج در هر لگ می‌شود. در نتیجه برای هر مسأله لازم است تعداد لگ مناسب مشخص شود. در این مقاله با استفاده از تکنیک شبیه‌سازی، مجموعه داده‌های فضایی همسانگرد با یک مدل تغییرنگار نمایی تولید و تعداد لگ مناسب برای مدل‌سازی تغییرنگار مورد بررسی قرار می‌گیرد.

شبیه‌سازی

برای تولید داده‌های فضایی، یا به عبارت دیگر مقادیر یک میدان تصادفی مانای مرتبه دوم $\{Z(t); t \in D\}$ با تابع هم تغییرنگار معلوم $C(\cdot)$ در n موقعیت معین t_1, \dots, t_n از تجزیه چولسکی [۵] ماتریس معین مثبت $\Sigma = (Cov(Z(t_i), Z(t_j))) = (C(t_i - t_j))$ ، یعنی تجزیه $\Sigma = L'L$ ، به صورت $Z = \mu + L\varepsilon$ ، به صورت $Z = \mu + L\varepsilon$ ، یک بردار از متغیرهای تصادفی می‌شود که در آن L یک ماتریس $n \times n$ پایین مثلثی، $\varepsilon = (\varepsilon(t_1), \dots, \varepsilon(t_n))'$ یک بردار از متغیرهای تصادفی مستقل با میانگین صفر و واریانس واحد است. در این صورت بردار داده‌های فضایی تولید شده $Z = (Z(t_1), \dots, Z(t_n))'$ دارای بردار میانگین μ و ماتریس واریانس - کوواریانس Σ است. چنانچه بردار تصادفی ε گوسی اختیار شود، داده‌های فضایی تولید شده نیز از یک میدان تصادفی گوسی خواهند بود.

۱ - Lags

با استفاده از این روش، یک مجموعه داده فضایی بر یک شبکه منظم 16×16 تایی، در مربعی به اضلاع 500 واحد و بر اساس مدل شبه تغییرنگار نمایی (۳) با پارامترهای $c_0 = 0$ ، $c_e = 100$ و $a_e = 150$ تولید شده و برای تعداد لگهای متفاوت k بین 10 تا 35 مدل تغییرنگار نمایی به روش حداقل مربعات موزون [۶] برازش داده شده است. این عمل را $m=1000$ بار تکرار کرده‌ایم و برای هر یک از مقادیر k ، جذر میانگین انتگرال مجذور خطا به صورت

$$\gamma RMISE = \sqrt{\frac{1}{m} \sum_{j=1}^m \int_0^{500} (2\hat{\gamma}_j(r) - 2\gamma(r))^2 dr} \quad (4)$$

محاسبه شده است، که در آن $\hat{\gamma}_j(\cdot)$ مدل برازش شده در j امین تکرار است. این معیار نشانگر میزان نکویی تغییرنگارهای برازش شده است و مقدار بهینه k با کمینه کردن (۴) به دست می‌آید. لازم به ذکر است که محاسبات اولیه برای تعیین تعداد لگ مناسب بیانگر آن بود که تعداد لگ بهینه کمتر از 10 و بیشتر از 35 نیست، از این رو بررسی مقدار بهینه k در فاصله 10 تا 35 انجام گرفت.

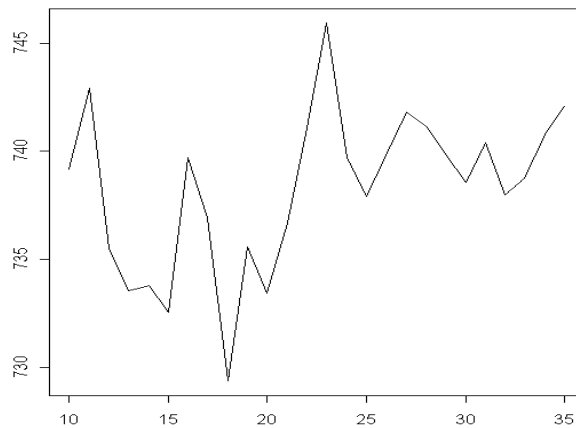
تولید داده‌های تصادفی با برنامه‌نویسی در محیط نرم‌افزار R و به کمک توابع GeoR از [۷] انجام گرفته و معیار RMISE به ازای تعداد لگ‌های مختلف و برای $m=1000$ تکرار محاسبه شده است. جدول ۱ نتایج حاصل از شبیه‌سازی را برای مقادیر مختلف k نشان می‌دهد. چنانکه در شکل ۱ ملاحظه می‌شود، کمترین مقدار RMISE به ازای تعداد لگ ۱۸ حاصل شده است و می‌توان آن را به عنوان تعداد لگ بهینه مورد استفاده قرار داد.

جدول ۱- مقادیر RMISE برای تعداد لگ‌های مختلف

k	۱۰	۱۱	۱۲	۱۳	۱۴	۱۵	۱۶
RMISE	۷۳۹/۱۶	۷۴۳/۹۲	۷۳۵/۵	۷۳۳/۵۴	۷۳۳/۷۶	۷۲۲/۵۶	۷۲۹/۷
k	۱۷	۱۸	۱۹	۲۰	۲۱	۲۲	۲۳
RMISE	۷۳۶/۹	۷۲۹/۳۶	۷۳۵/۵۴	۷۳۳/۴۲	۷۳۶/۶۸	۷۴۱/۰۸	۷۴۵/۹۲
k	۲۴	۲۵	۲۶	۲۷	۲۸	۲۹	۳۰
RMISE	۷۳۹/۷	۷۳۹/۹۲	۷۳۹/۸۲	۷۴۱/۷۸	۷۴۱/۱۶	۷۳۹/۸۶	۷۳۸/۵۵
k	۳۱	۳۲	۳۳	۳۴	۳۵		
RMISE	۷۴۰/۴۲	۷۳۹/۹۸	۷۳۸/۷۶	۷۴۰/۷۸	۷۴۲/۱۲		

مثال کاربردی

گاهی در علوم پزشکی و به خصوص در همه‌گیرشناسی جغرافیایی با داده‌های فضایی روبرویم که لازم است در تجزیه و تحلیل آماری ساختار همبستگی فضایی آنها نیز مدنظر قرار گیرد. به عنوان مثال میزان بروز یک



شکل ۱- نمودار RMISE در مقابل تعداد لگ (k)

بیماری یا میزان مرگ و میر ناشی از یک بیماری، در مناطق مختلف جغرافیایی، یا در شهرستانهای مختلف، یک مجموعه داده فضایی هستند. به طور کلی اغلب شاخصها یا میزانهای بهداشتی مناطق مختلف می‌توانند صدیقی از داده‌های فضایی باشند. یکی از کارهایی که با استفاده از تحلیل فضایی می‌توان انجام داد، تخمین فضایی میزان بیماری است. با توجه به همبستگی فضایی این گونه داده‌ها، تعیین ساختار همبستگی آنها از طریق تابع تغییرنگار اولین مرحله تجزیه و تحلیل آنها به شمار می‌آید.

در این بخش ساختار همبستگی فضایی دو مجموعه داده، شامل میزان بروز (درصد شیوع) بیماری سل ریوی شهرستانهای کشور، در سالهای ۱۳۷۷ و ۱۳۷۸ از طریق تغییرنگار برآورد می‌شود. هر یک از مجموعه داده‌ها با $\{Z(t_1), \dots, Z(t_n)\}$ نشان داده می‌شود، که در آن $n=262$ تعداد شهرستانهای کشور، $t_i = (x_i, y_i)$ طول و عرض جغرافیایی مرکز شهرستان i ام و $z(t_i)$ میزان بروز بیماری در شهرستان i ام در سال موردنظر است که از تقسیم تعداد موارد بیماری ثبت شده در طی یک سال بر جمعیت شهرستان i ام در وسط همان سال به دست آمده است. این داده‌ها توسط اداره کل مبارزه با بیماریهای وزارت بهداشت، درمان و آموزش پزشکی، ثبت و جمع‌آوری شده است.

چون تابع تغییرنگار برای یک میدان تصادفی مانای ذاتی تعریف می‌شود، ابتدا مانایی داده‌ها مورد بررسی قرار می‌گیرد. به دلیل وجود یک روند فضایی در داده‌ها، لازم است این روند از داده‌ها حذف و بر اساس داده‌های بدون روند شده تغییرنگار برآورد شود. روند داده‌ها را می‌توان به صورت یک چندجمله‌ای درجه دو

برحسب موقعیتهای $t_i = (x_i, y_i)$ به صورت

$$\mu(t_i) = a_0 + a_1 x_i + a_2 y_i + a_3 x_i^2 + a_4 y_i^2 + a_5 x_i y_i$$

در نظر گرفت. در این صورت داده‌های بدون روند شده به صورت

$$R(t_i) = z(t_i) - \mu(t_i)$$

حاصل می‌شوند. با توجه به این که

$$\text{Var}(R(t_i) - R(t_i + h)) = \text{Var}(Z(t_i) - Z(t_i + h))$$

است، می‌توان تغییرنگار داده‌های اصلی را بر اساس داده‌های بدون روند شده برآورد نمود. برای هر یک از مجموعه داده‌های بدون روند شده حاصل از میزان بروز بیماری سل، در سالهای ۱۳۷۷ و ۱۳۷۸، تغییرنگار تجربی در ۱۸ لگ محاسبه گردید، سپس به روش حداقل مربعات موزون مدل تغییرنگار نمایی به آنها برازش داده شده است. جدول ۲ برآورد پارامترها و حداقل مجموعه مربعات موزون (MWSS) را برای هر یک از مجموعه داده‌های سال ۱۳۷۷ و ۱۳۷۸ برای دو حالت $c_0 = 0$ و $c_0 \neq 0$ نشان می‌دهد.

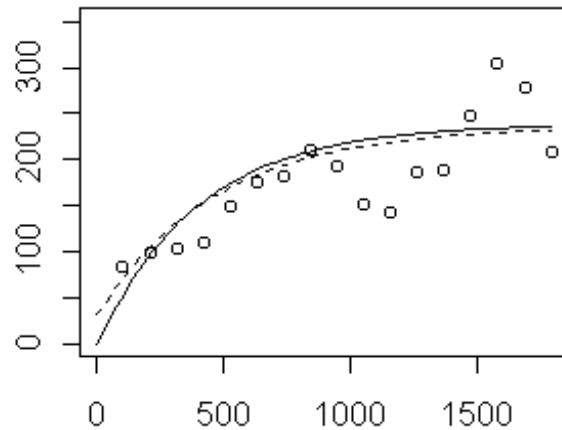
جدول ۲- برآورد پارامترهای مدل تغییرنگار نمایی

سال	اثر قطعه‌ای	\hat{c}_0	\hat{c}_e	\hat{a}_e	MWSS
۱۳۷۷	$c_0 = 0$	۰	۱۰۸/۱۴	۱۶۳/۶۵	۱۲۶/۲۸
	$c_0 \neq 0$	۲۸/۵۰	۸۴/۱۲	۱۰۶۵/۴۴	۸۰/۱۲
۱۳۷۸	$c_0 = 0$	۰	۱۲۲/۱۸	۱۹۸/۸۰	۱۵۳/۸۳
	$c_0 \neq 0$	۲۹/۷۱	۹۶/۷۲	۱۰۳۵/۵۰	۱۰۱/۸۲

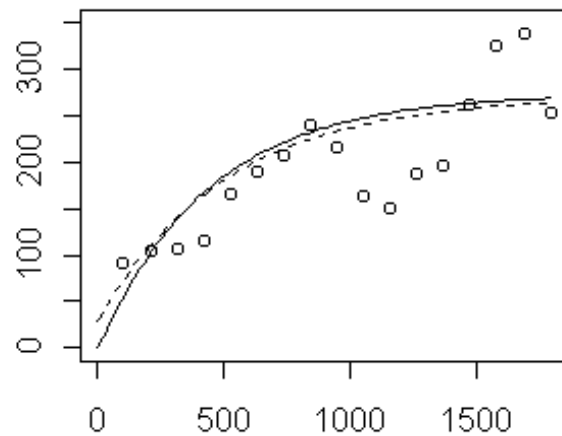
در حالت $c_0 = 0$ ، اثر قطعه‌ای مقدار ثابت صفر در نظر گرفته شده و مقدار آن برآورد نشده است و در حالت دیگر اثر قطعه‌ای نیز همانند دو پارامتر دیگر برآورد شده است. در شکل‌های ۲ و ۳ نقاط نشان دهنده تغییرنگار تجربی هستند و مدل برآورد شده، با فرض $c_0 = 0$ ، با نقطه‌چین و مدل برآورد شده، با فرض $c_0 \neq 0$ ، با خط‌چین نمایش داده شده است.

چنانکه ملاحظه می‌شود حداقل مجموع مربعات موزون در هر دو مجموعه داده برای حالت $c_0 \neq 0$ کمتر از حالت $c_0 = 0$ است، بنابراین مدل برازش شده تحت $c_0 \neq 0$ به داده‌ها بهتر برازش می‌شود، به عبارت دیگر اثر قطعه‌ای تغییرنگار برای این دو مجموعه داده مخالف صفر است.

مدلهای حاصل برای تغییرنگار (با اثر قطعه‌ای) برای تخمین میزان بیماری به روش کریگینگ [۳] در ۲۴ شهرستان که اطلاعی از میزان بیماری آنها در دسترس نبود، مورد استفاده قرار گرفت. جدول ۳ به ترتیب، نام شهرستانها، طول و عرض مرکز شهرستانها، تخمین میزان بروز بیماری سل ریوی (بر حسب ۱۰۰۰۰۰ نفر) و انحراف معیار آنها را برای سالهای ۱۳۷۷ و ۱۳۷۸ نشان می‌دهد.



شکل ۲- تغییرنگار تجربی و مدل برازش شده برای سال ۱۳۷۷



شکل ۳- تغییرنگار تجربی و مدل برازش شده برای سال ۱۳۷۸

بحث و پیشنهاد

برای انتخاب یک مدل تغییرنگار می‌توان مدلهای مختلف را با تعداد لگ ثابت برازش داد و با استفاده از معیار مجموع مربعات موزون باقیمانده‌ها، از میان آنها بهترین مدل را انتخاب کرد. اما به دلیل آن که با افزایش تعداد لگ‌ها مجموع مربعات موزون باقیمانده‌ها افزایش می‌یابد، به کمک این معیار حتی برای یک مجموعه داده خاص، نمی‌توان تعداد لگ مناسب را تشخیص داد.

برای انتخاب تعداد لگ مناسب، اغلب صاحب‌نظران تجزیه و تحلیل داده‌های فضایی فقط به توصیه‌های کلی اکتفا کرده‌اند. از جمله [۸] برای برآورد تغییرنگار توصیه می‌کنند، در هر لگ حداقل ۳۰ زوج متمایز قرار داشته باشد. بنابراین تعداد لگ‌ها باید به گونه‌ای انتخاب شود که بیشتر آنها در این شرط صدق کنند. «کرسی» [۳] معتقد است که طول لگ‌ها باید آن قدر کوچک باشد که ساختار همبستگی را با وضوح بیشتری نشان دهد و از طرف دیگر باید به قدر کافی بزرگ باشد تا برآوردگر تغییرنگار تجربی پایدار شود. «ویستر و

اولیور» [۹] معایب کم یا زیاد بودن تعداد لگها را بیان و تاکید می‌کند که مقدار متعادل آن به تعداد و موقعیت داده‌ها و شکل تغییرنگار بستگی دارد و متوسط فاصله میان زوج داده‌ها را به عنوان یک مقدار اولیه برای طول لگ پیشنهاد می‌کند. با توجه به فقدان قاعده، یا نظریه‌ای کلی برای تعداد لگ بهینه، مطالعه حاضر صورت گرفت. در این مطالعه حجم نمونه تقریباً برابر تعداد داده‌های مربوط به میزان بروز بیماری سل ریوی در شهرستانهای ایران قرار داده شد تا بتوان نتیجه حاصل از شبیه‌سازی را در تجزیه و تحلیل فضایی آنها به کار گرفت. از این رو استفاده از نتیجه این مطالعه در تجزیه و تحلیل داده‌ها کاربر را امیدوار می‌سازد که مدل بهتری به تغییرنگار داده‌ها برآزش دهد و تخمین‌های بهتری بدست آورد. در عین حال، مطالعه‌ای برای بررسی تاثیر حجم نمونه در تعداد لگ بهینه ضروری به نظر می‌رسد. معمولاً محققان و تحلیلگران داده‌های فضایی طول لگها را یکسان می‌گیرند. این امر موجب می‌شود تعداد زوج مشاهدات در لگها متفاوت بوده و میزان دقت برآوردهای حاصل در هر لگ مختلف باشد. اگر طول لگها بگونه‌ای انتخاب شوند که تعداد زوج مشاهدات در آنها حتی‌الامکان یکسان باشند می‌توان تا حدودی به این نقیصه فائق آمد.

جدول ۳- تخمین فضایی میزان بروز بیماری سل ریوی در ۲۴ شهرستان

نام شهرستان	طول	عرض	سال ۱۳۷۷		سال ۱۳۷۸	
			میزان	انحراف معیار	میزان	انحراف معیار
اسکو	۲۱۹/۲	۱۳۰۸/۸	۵/۵۳	۶/۲۵	۴/۷۵	۶/۴۶
جلفا	۱۷۲/۸	۱۴۲۰/۸	۷/۶۶	۷/۱۰	۵/۸۳	۷/۴۲
چالدران	۵۹/۲	۱۴۳۲/۰	۷/۶۹	۷/۰۱	۴/۸۱	۷/۲۷
کوثر	۴۰۸/۰	۱۲۸۴/۸	۴/۳۶	۶/۴۸	۴/۳۹	۶/۷۲
نیر	۳۷۹/۲	۱۳۳۲/۸	۴/۹۷	۶/۴۳	۴/۸۱	۶/۶۷
ایجرود	۴۲۵/۶	۱۱۶۰/۰	۳/۶۷	۶/۵۲	۳/۸۱	۶/۷۶
خرم دره	۵۰۸/۸	۱۱۳۶/۰	۴/۹۰	۶/۲۸	۴/۳۹	۶/۴۸
املش	۵۸۷/۲	۱۴۰۴/۸	۱۳/۹۲	۸/۰۸	۱۴/۴۴	۸/۵۳
رضوانشهر	۴۸۸/۰	۱۲۸۸/۰	۹/۸۰	۶/۳۱	۹/۴۱	۶/۵۳
سیاهکل	۵۶۶/۸	۱۲۴۴/۸	۱۱/۱۸	۶/۰۹	۱۱/۵۵	۶/۲۹
ماسال	۴۸۸/۰	۱۲۶۸/۸	۱۰/۲۶	۶/۱۹	۹/۷۸	۶/۴۰
جویبار	۸۲۷/۲	۱۲۱۹/۲	۸/۹۶	۶/۱۰	۷/۰۲	۵/۹۹
بندرگز	۹۲۰/۰	۱۲۴۴/۸	۱۴/۲۳	۶/۲۲	۱۴/۵۳	۶/۴۲
آزادشهر	۱۰۲۴/۰	۱۲۹۶/۰	۴۱/۹۸	۶/۳۰	۴۳/۸۰	۶/۵۰
آق قلا	۹۶۰/۰	۱۲۸۰/۰	۲۵/۴۱	۶/۴۵	۲۶/۷۶	۶/۶۸
رامیان	۱۰۲۴/۰	۱۲۸۸/۰	۳۹/۷۷	۶/۳۷	۴۱/۵۴	۶/۵۸
کلالة	۱۰۴۸/۰	۱۳۳۲/۸	۴۶/۵۷	۶/۸۴	۴۹/۷۰	۷/۱۱
پاکدشت	۷۶۸/۰	۱۰۸۰/۰	۱۳/۹۷	۶/۳۸	۱۶/۴۸	۶/۶۰
تیران و کرون	۷۲۰/۰	۷۶۸/۰	۱۳/۱۷	۶/۲۷	۴/۴۲	۶/۵۰
ارسنجان	۹۵۸/۴	۴۸۶/۴	۱/۷۶	۶/۸۹	۱/۹۰	۷/۱۸
خرم بید	۹۳۷/۶	۵۶۰/۰	۰/۸۷	۶/۷۱	۱/۴۵	۶/۹۸
جاجرم	۱۱۳۱/۲	۱۲۹۴/۴	۲۸/۲۴	۷/۲۲	۲۹/۰۷	۷/۶۷
ابوموسی	۱۱۸۵/۶	۰۴۵/۴	۱۱/۲۳	۸/۰۸	۱۰/۱۱	۸/۴۹
راور	۱۲۷۲/۰	۶۷۶/۸	۱۰/۴۱	۷/۴۲	۱۰/۲۷	۷/۷۷

این مطالعه با فرض همسانگرد بودن داده‌ها صورت گرفته است. لذا برای حالت ناهمسانگرد، که لازم است یک تغییرنگار جهتی برای تعیین ساختار همبستگی داده‌ها بکار گرفته شود، تعیین تعداد لگ بهینه مسئله‌ای دیگر است.

مراجع

1. G. Matheton, The Theory of Regionalized Variables and Its Application. Cahiers du Centre de Morphologie Mathematique, No.5. Fontainebleau, France (1971).
2. G. Matheron, Traite de Geostatistique Appliquee, I. Tome Memories du Bureau de Recherches Geologiques et Minieres, No. 14, Editions Technip, Paris (1962).
3. N. Cressie, Statistics for Spatial Data, New York, John Wiley (1993).
4. J. Barry, M. Crowder, and P. Diggle, "Parametric Estimation of the Variogram", Technical Report ST- 97- 06, Dep. Maths and Stats, Lancaster University, Lancaster, UK (1997).
5. G. H. Golub, and C. F. Van Loan, Matrix Computations, The Johns Hopkins University press, M. D. Baltimore, (1983).
6. N. Cressie, Fitting Variogram Models by Weighted Least Squares, Mathematical Geology, V.17 (1985) 563-586.
7. Jr. P. J. Ribeiro and P.J. Diggle, "geoR/geoS: Functions for Geostatistical Analysis Using R or S-PLUS" Technical Report: ST-99-09, Dep. of Maths. and Stats. Lancaster University, Lancaster, UK (2000).
8. A. G. Journel, and C. J. Huijberegts, Mining Geostatistics, Academic Press (1978).
9. O. Webster, and M. Oliver, Geostatistics for Environmental Scientists, New York, John Wiley (2001).