

شبیه‌سازی گسترش مکانی سرطان پوست توسط آنالیزهای مکان - آماری

زهرة معصومی^{۱*}، محمد سعدی مسگری^۲ و محمد کریمی^۳

^۱ دانشجوی دکتری سیستم اطلاعات مکانی - دانشگاه صنعتی خواجه نصیرالدین طوسی

^۲ استادیار دانشکده مهندسی نقشه برداری - دانشگاه صنعتی خواجه نصیرالدین طوسی

^۳ استادیار دانشکده مهندسی نقشه برداری - دانشگاه صنعتی خواجه نصیرالدین طوسی

(تاریخ دریافت ۸۶/۱۱/۲۷، تاریخ دریافت روایت اصلاح شده ۸۸/۱۲/۲۲، تاریخ تصویب ۸۹/۶/۹)

چکیده

آمار و احتمال وقوع بیماری سرطان پوست با عوامل متعددی ارتباط دارد که بیشترین این عوامل توزیع مکانی غیر یکنواخت دارند. از سویی دیگر بیماری سرطان یک پدیده احتمالی است و حتی در شرایط مناسب آن نیز به طور حتم پیش نمی‌آید. در نتیجه مدل‌های زمین آماری گزینه مناسبی برای تحلیل مکانی این گونه بیماری‌ها هستند. در این تحقیق، یک مدل زمین آماری برای شبیه‌سازی گسترش مکانی بیماری طراحی و توسعه داده شده است. رابطه مکان-آماری ارائه شده، بیانگر عوامل مهم ایجاد بیماری به همراه درجه اهمیت آنها است. مدل توسعه داده شده می‌تواند به عنوان یک ابزار شبیه‌سازی وقوع بیماری و ابزار کمک به تصمیم‌گیری توسط مدیران مربوطه، به کار رود. به عبارتی دیگر، می‌توان با تغییر هر یک از عوامل و یا تغییر همزمان تعدادی از آنها به میزان دلخواه، نتیجه این گونه تغییرات را در میزان وقوع و توزیع مکانی بیماری ملاحظه و بررسی کرد.

واژه‌های کلیدی: آنالیزهای مکان - آماری، مدل GLM، سیستم‌های اطلاعات مکانی، پیش‌بینی مکانی، سرطان پوست

مقدمه

یکی از نیازهای اصلی تصمیم‌گیرندگان امور بهداشتی، پیش و کنترل بیماری‌ها از نظر پراکندگی مکانی و تعداد افراد مبتلا است. با استفاده از مدل‌سازی بیماری می‌توان عوامل مهم ایجاد بیماری را رتبه‌بندی کرد. از سویی دیگر اگر در این مدل المان مکان وارد شود، مدل قابلیت پیش‌بینی گسترش مکانی بیماری را نیز خواهد داشت.

از میان سه المان اصلی تأثیرگذار بر بیماری یعنی شخص، زمان و مکان، آنالیز المان مکان همواره سخت و زمان‌بر بوده است [۱]. ارتباط مکانی یک بیماری با عوامل ایجادکننده آن (در صورت محدود بودن تعداد عوامل) با

توابع ساده مکانی مانند همپوشانی و حریم‌گذاری قابل تعریف و تفسیر است [۲]. ولی با افزایش تعداد عوامل ایجادکننده بیماری، شبیه‌سازی و مدل‌کردن آن با استفاده از آنالیزهایی که فقط مکانی هستند، کاری بس دشوار است [۳]. بدین منظور لازم است از تکنیک‌های موجود در علوم دیگر همچون آمار، برای بالا بردن کارایی مدل‌های مکانی استفاده کرد. بنابراین با استفاده از تلفیق مدل‌های مکانی و آماری می‌توان از مزایای مدل‌های آماری در آنالیزهای مکانی بهره جست. در نتیجه با استفاده از یک مدل مکان-آماری می‌توان به عوامل مهم ایجاد بیماری و محاسبه درجه

اهمیت هر یک دست یافت. بدین ترتیب اولویت‌ها در پیشگیری از بیماری برای تصمیم‌گیران روشن خواهد شد. این بدان معنی است که تخصیص بودجه و امکانات برای مبارزه با بیماری با برنامه‌ریزی دقیق‌تر انجام خواهد شد. از طرفی با شبیه‌سازی بیماری با استفاده از این نوع مدل‌ها، پیش‌بینی گسترش مکانی بیماری فراهم می‌شود، زیرا با استفاده از مدل به دست آمده می‌توان با تغییر هر یک از عوامل ایجادکننده بیماری در مدل، تغییرات تعداد بیماری و پراکندگی مکانی آن را مشاهده کرد و تصمیم مناسبی برای پیشگیری از بیماری گرفت.

در سال‌های اخیر کارهای محدودی در ارتباط با مدل‌سازی مکان-آماری بیماری‌ها انجام شده است. در بیشتر موارد هدف، به دست آوردن ارتباط میان تعداد محدودی از عوامل ایجاد بیماری (حتی یک عامل) با گسترش مکانی بیماری بوده است. در حالت کلی کارهای انجام‌شده را می‌توان در دو دسته طبقه‌بندی کرد: در دسته اول فقط رابطه مکانی یک عامل ایجادکننده بیماری مانند آلودگی آب، هوا و غیره با تعداد بیماری رخ داده و پراکندگی آن با استفاده از آنالیزهای ساده سیستم اطلاعات مکانی به دست آمده است. برای مثال، تحقیقی برای به

مدل تعمیم یافته خطی (GLM^۱)

GLM در واقع یک روش برای تخمین و محاسبه مدل-های مختلف است که با تئوری رگرسیون‌های خطی و نیز مدل‌های غیرخطی سازگاری دارد [۹]. یک فرض بسیار مهم در GLM این است که تابع متغیر پاسخ، عضوی از خانواده توابع چگالی احتمال نمایی است [۱۰] و [۱۱]. به عنوان مثال توابع چگالی احتمال نرمال، پواسون، معکوس نرمال، گاما و... جزو این خانواده هستند. رابطه عمومی خانواده توابع نمایی به این شکل است:

$$f(y_i, \theta_i, \Phi) = \exp\left\{ \frac{y_i \theta_i - b(\theta_i)}{a(\Phi)} + h(y_i, \Phi) \right\} \quad (1)$$

در این رابطه، $b(\theta)$ و $a(\Phi)$ و $h(y)$ تابعی از عوامل معلوم هستند. θ عامل طبیعی تابع چگالی احتمال $\varphi > 0$ عامل مقیاس تابع است [۱۲].

بنابراین در صورتی که تابع توزیع احتمال متغیر پاسخ (در اینجا تعداد بیماری) با یک احتمال بالا مشخص باشد، می‌توان از این مدل استفاده کرد [۱۳]. تعداد مرگ و میر در اثر سرطان پوست که به نسبت جمعیت واحدهای مکانی استاندارد شده است، با سطح اطمینان قابل قبولی، دارای تابع توزیع احتمال پواسون است (البته باید این مطلب با آزمایش‌های آماری اثبات شود که در قسمت برازش مدل به آن پرداخته شده است). زیرا از طرفی رخداد مرگ در اثر سرطان پوست به شکل باینری^۱ است و از طرف دیگر، احتمال رخداد بیماری در تعداد بالا، کم است. این مشخصات، مشخصات اصلی تابع توزیع احتمال پواسون است [۱۴].

هنگامی که تابع توزیع احتمال پواسون باشد، رابطه (۱) به رابطه (۲) تبدیل می‌شود:

$$f(y_i, \theta_i, \Phi) = \exp\left\{ y_i \log(\mu) - \mu - \log(y_i!) \right\} \quad (2)$$

که در آن μ مقدار میانگین است. بنابراین با مقایسه دو رابطه ۱ و ۲ خواهیم داشت:

$$\theta = \log(\mu), \quad \phi = 1, \quad a(\phi) = 1 = \phi, \quad (3)$$

$$b(\theta) = \mu = e^\theta, \quad h(y, \phi) = -\log(y!)$$

المان‌های GLM

GLM سه المان اصلی دارد:

دست آوردن رابطه تشعشعات الکترومغناطیس و سرطان کودکان در Clorado (Denver) انجام شده است. در این تحقیق با استفاده از آنالیزهای همپوشانی و حریم‌گذاری، تعداد کودکان مبتلا به سرطان که در معرض این تشعشعات قرار داشته‌اند به دست آمده است و مطالعاتی برای به دست آوردن رابطه‌ای میان این دو المان انجام شده است [۴] و [۵]. در تحقیق دیگری رابطه میان آلودگی هوا و بیماری آسم در نیویورک به دست آمده است. به این ترتیب که ابتدا با استفاده از داده‌های به دست آمده از ایستگاه‌های سنجش آلودگی هوا، پهنه‌بندی آلودگی هوا به دست آمده و سپس رابطه پراکندگی مکانی بیماری و مناطق پرخطر آلودگی هوا مورد مطالعه قرار گرفته است [۶]. در دسته دوم، ارتباط میان بیماری و همه المان‌های اصلی ایجادکننده آن به دست آمده است. برای مثال برای بررسی مناطق پرخطر تب زرد و عوامل محیطی و اکولوژیکی ایجادکننده آن از رگرسیون لجستیک استفاده شده است که در آن در نهایت یک رابطه میان المان‌های ایجادکننده بیماری و تعداد بیماری به دست آمده است. همان‌طور که ملاحظه می‌شود، در این تحقیق، هدف تنها پیش‌بینی تعداد بیماری بوده است و نیز اثر متقابل المان‌های ایجادکننده بیماری در نظر گرفته نشده است [۷]. در تحقیق دیگری در انگلستان رابطه میان بیماری سل با دمای سالیانه، متوسط دما، بیشینه دما و رطوبت به دست آمده است. این کار با استفاده از مدل GLM انجام شده و در نهایت ضریب تأثیر هر یک از عوامل به دست آمده است [۸].

در تحقیق حاضر، هدف شبیه‌سازی بیماری سرطان پوست بر اساس عوامل مهم ایجادکننده آن در محدوده ۱۸ استان از کشور و در واحدهای مکانی شهرستان است. به طوری که پراکندگی مکانی بیماری مد نظر بوده و پیش‌بینی آن نیز میسر باشد که این موضوع با استفاده از آنالیزهای سیستم اطلاعات مکانی و تلفیق آن با مدل‌های آماری، محقق شده است. علاوه بر این، در این مدل استقلال المان‌های ورودی به مدل از یکدیگر بررسی شده و اثر المان‌های وابسته وارد مدل شده است. از مدل آماری GLM برای به دست آوردن رابطه آماری میان بیماری و عوامل ایجادکننده آن و از سیستم اطلاعات مکانی برای بررسی پراکندگی مکانی و نیز پیش‌بینی گسترش مکانی بیماری استفاده شده است.

^۱ صفر: عدم رخداد مرگ. یک: رخداد مرگ

انتخاب رو به جلو: این فرایند با فرض اینکه هیچ متغیر مستقلی در مدل وجود ندارد، آغاز می‌شود. هدف، یافتن بهترین زیر مجموعه با ورود متغیرهای مستقل به مدل است. اولین متغیر مستقلی که وارد مدل می‌شود، متغیری است که بالاترین وابستگی را با متغیر پاسخ دارد. فرض کنید این متغیر x_1 است. به علاوه این متغیر بالاترین مقدار آماره $F^{\#}$ را در آزمایش درجه اطمینان دارد.

متغیر مستقل بعدی در صورتی وارد مدل خواهد شد که آماره F آن به مرز مقدار F که از پیش برای مدل تعیین شد و F_{IN} یا $F_{(q-mer)}$ نامیده می‌شود، برسد. متغیر مستقل بعدی که برای ورود به مدل انتخاب می‌شود، متغیری است که بالاترین وابستگی را با Y پس از محاسبه اثر X_1 در Y دارد. این فرایند تا زمانی ادامه پیدا می‌کند که مقدار آماره F هیچ متغیر دیگری به مقدار F_{IN} نرسد [۱۱].

حذف رو به عقب: فرایند این روش در خلاف جهت روش انتخاب رو به جلوست. در این روش، مدل اولیه همه متغیرهای مستقل را دارد [۱۹]. در مرحله بعد آماره F برای هر یک از متغیرهای مستقل در مدل محاسبه می‌شود. کوچک‌ترین مقدار آماره F با مقدار اولیه انتخاب شده F برای مدل که F_{out} نامیده می‌شود، مقایسه شده و در صورتی که کوچک‌ترین F از F_{out} کمتر باشد، متغیر از مدل حذف می‌شود. در مرحله بعد، یک مدل $k-1$ عامل به داده‌ها برازش داده و آماره F برای مدل جدید محاسبه شده و این عملیات تکرار می‌شود. این روش زمانی خاتمه می‌یابد که کوچک‌ترین آماره F مربوط به عوامل از آماره F_{out} بزرگ‌تر باشد.

روش انتخاب قدم به قدم: یکی از پر کاربردترین روش‌های برازش، مدل روش قدم به قدم انتخاب عوامل است. این روش، تعمیم یافته روش انتخاب رو به جلو است؛ با این تفاوت که در اینجا با ورود یک متغیر به مدل، متغیرهایی که تا کنون وارد مدل شده‌اند ولی پس از ورود متغیر جدید پیش‌بینی کننده معنی‌داری نیستند، از مدل خارج می‌شوند. با توجه به نوع پردازش در اینجا دو مرز تعریف می‌شود: F_{IN} و F_{out} [۲۰]. در ابتدا دو متغیر به روش رو به جلو وارد مدل می‌شوند و سپس به روش رو به عقب

المان تصادفی: واضح است که در مدل کردن یک پدیده طبیعی، دسترسی به مدل واقعی، غیر ممکن است. به این ترتیب مدل محاسبه شده در مقایسه با مدل واقعی همواره خطاهای مشاهداتی و محاسباتی دارد. این خطاها، المان تصادفی GLM نامیده می‌شوند [۱۵].

المان سیستماتیک: در ریاضیات وقتی رابطه‌ای میان دو المان برقرار می‌شود، احتمال در رابطه مذکور وجود ندارد و سطح اطمینان رابطه صد در صد است. برای مثال رابطه ۴ نشان‌دهنده رابطه خطی میان متغیر پاسخ Y و متغیر مستقل X در ریاضیات است که در آن β ماتریس ضرایب مجهول است [۱۶]:

$$\beta Y = X_i \quad (4)$$

در صورتی که اگر سطح اطمینان، وارد رابطه شود، المان تصادفی (ε) وارد رابطه خواهد شد:

$$\beta Y = X_i + \varepsilon \quad (5)$$

قسمتی از مدل که با سطح اطمینان صد در صد پذیرفته می‌شود، المان سیستماتیک نامیده می‌شود.

تابع ارتباط^۲: مهم‌ترین خصوصیت مدل‌های GLM گسترش یک مدل خطی برای یک متغیر پاسخ با توجه به متغیرهای مستقل است. با فرض اینکه η_i پیش‌بینی کننده خطی نام دارد؛

$$\eta_i = g[E(y_i)] = g(\mu_i) = x_i \beta \quad (6)$$

که در آن $Y = [y_1, y_2, \dots, y_n]$ ماتریس متغیر پاسخ و $X = [x_1, x_2, \dots, x_n]$ ماتریس متغیرهای مستقل و β ماتریس ضرایب مجهول است. در رابطه ۶، $E(y_i)$ به صورت زیر محاسبه می‌شود:

$$E(Y_i) = g^{-1}(\eta_i) = g^{-1}(X_i \beta) \quad (7)$$

که در آن g تابع ارتباط نامیده می‌شود [۷]، اثبات می‌شود وقتی تابع توزیع احتمال متغیر پاسخ پواسون باشد، تابع ارتباط لگاریتم است [۱۱].

نحوه انتخاب عوامل در مدل GLM

از آنجایی که محاسبه همه مدل‌های موجود پیچیده و بسیار زمان‌بر است، روش‌های متفاوتی برای محاسبه بهترین مدل ارائه شده است [۱۱]. این روش‌ها اغلب به پردازش‌های stepwise - type مشهورند که به سه دسته کلی تقسیم می‌شوند: (۱) انتخاب رو به جلو^۳ (۲) حذف رو به عقب^۴ (۳) روش قدم به قدم^۵ [۱۷] و [۱۸]. در اینجا این روش‌ها به طور خلاصه توضیح داده می‌شوند:

^{۱۱} درجه اطمینان وابستگی میان متغیر پاسخ و متغیرهای مستقل با استفاده از آماره F آزمایش می‌شود که دارای تابع توزیع احتمال فیشر (Fisher) است.

متدولوژی و روش انجام کار

در این قسمت پس از معرفی محدوده مطالعاتی، طراحی پایگاه داده و چارچوب اطلاعاتی مدل و نحوه برازش مدل خواهد آمد.

محدوده مطالعاتی

محدوده مطالعاتی این تحقیق ۱۸ استان کشور است که در آنها مرگ و میر ناشی از چند نوع سرطان از طرف وزارت بهداشت ثبت شده است. واحد مکانی انتخاب شده برای مدل، شهرستان‌های کشور هستند که تعداد آنها در ۱۸ استان ذکر شده، ۹۲ شهرستان است.

طراحی پایگاه داده و چارچوب اطلاعاتی مدل

اولین مرحله برای مدل کردن بیماری، یافتن عوامل اصلی ایجاد سرطان پوست و تصمیم‌گیری درباره چگونگی آماده‌سازی داده‌ها بوده است. برای دستیابی به این هدف در این مرحله، مدارک و اسناد و نیز گزارش‌های سازمان بهداشت جهانی، وزارت بهداشت درمان و آموزش پزشکی، مؤسسه بین‌المللی تحقیقات سرطان و سایر مستندات در دسترس مطالعه و بررسی شده و مشاوره‌های لازم با متخصصان این موضوع به عمل آمده است. خروجی این مطالعات، نشان‌دهنده مهم‌ترین عوامل ایجاد سرطان پوست است که در جدول (۱) دسته‌بندی آنها و نیز واحد مکانی داده‌های موجود و واحد مکانی داده‌های مورد نیاز برای ورود به مدل آمده است.

در نهایت پایگاه داده مورد نیاز برای ورود به مدل طراحی و آنالیزها و پردازش‌های مختلف بر داده‌ها در محیط سیستم اطلاعات مکانی انجام شده است که در قسمت بعدی شرح داده خواهد شد.

پردازش داده‌ها برای ورود به مدل

در این قسمت پردازش‌های انجام شده روی هر دسته از داده‌ها به طور مختصر شرح داده خواهد شد.

المان‌های طبیعی و اقلیم: در این تحقیق، متوسط سالیانه عوامل فشار اتمسفر، دما، رطوبت، انرژی خورشیدی و تعداد ساعات آفتابی برای هر شهرستان و بر مبنای مشاهده‌های ایستگاه‌های کلیماتولوژی و سینوپتیک سازمان هواشناسی کشور، محاسبه شده است. مراحل زیر برای تولید داده‌ها برای ورود به مدل انجام شده است:

معیار حذف بررسی می‌شود. در هر مرحله با استفاده از روش رو به جلو متغیری وارد مدل شده سپس با استفاده از قواعد حذف رو به عقب، همه متغیرهایی که تا کنون وارد مدل شده‌اند، بررسی شده و در صورت وجود معیار لازم، از مدل حذف می‌شوند. این عملیات تا زمانی ادامه پیدا می‌کند که متغیری باقی نماند که در شرط ورود به مدل صدق کند.

جستجوی اثر متقابل میان متغیرهای مستقل

در برخی موارد، اثرات متقابل میان متغیرهای ورودی به مدل وجود دارد. همان طور که توسط برخی از دانشمندان [۱۵] مطرح شده است، این پدیده بیشتر زمانی رخ خواهد داد که متغیرهای ورودی به مدل متغیرهای طبیعی باشند و یک ارتباط طبیعی میان آنها وجود داشته باشد [۲۱]. هنگامی که دو متغیر اثر متقابل داشته باشند، با ثابت در نظر گرفتن همه متغیرهای دیگر، هر تغییر در یکی از آنها باعث تغییر در دیگری خواهد شد. مدل GLM جزو مدل‌های آماری است که قادر است این اثر را وارد مدل کند. این اثر در مدل اغلب با ورود یک متغیر جدید که حاصل ضرب دو عامل است، نشان داده می‌شود [۸] و [۱۱].

آنالیز باقی‌مانده‌ها در مدل

مانند سایر مدل‌ها و فرایند شبیه‌سازی، آنالیز باقی‌مانده‌ها در GLM اهمیت خاصی دارد. باقی‌مانده‌های مدل، نشانگر دقت برازش مدل و دقت انتخاب تابع ارتباط هستند [۹]. باقی‌مانده معمولی در مدل GLM، تفاضل میان مقادیر به دست آمده از مدل و مقادیر واقعی است.

از میان باقی‌مانده‌های موجود، در مدل GLM اغلب باقی‌مانده‌های deviance بررسی می‌شوند. باقی‌مانده‌های deviance طبق رابطه ۸ محاسبه می‌شوند که در آن d_i یک تابع میان y و μ است، که در نقطه i تعریف می‌شود [۲۲].

$$r_{pi} = \sqrt{d_i} \text{sign}(y_i - \hat{y}_i) \quad (8)$$

ملاک deviance یک آماره است که دارای تابع توزیع احتمال کای اسکور با درجه آزادی معادل تعداد عوامل موجود در مدل است. از آنجا که deviance بر مبنای خطاهای مدل بنا شده است، هر چه مقدار این آمار کمتر باشد، بهتر است [۱۱]. در این تحقیق باقی‌مانده‌های deviance برای هر سه روش انتخاب مدل محاسبه شده و نتایج در بخش قبل آورده شده است.

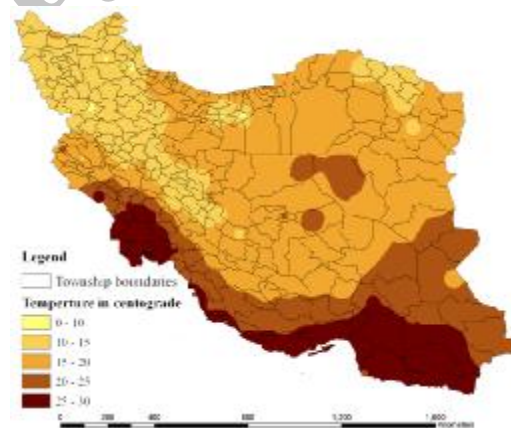
- برای تعریف یک رابطه میان این مشاغل و سرطان پوست، از استاندارد بین‌المللی کلاسه‌بندی صنایع^A (ISIC) استفاده شده است. به عنوان مثال برای آماده‌سازی داده‌های مربوط مشاغلی که در تماس مستقیم با ذغال سنگ هستند، مراحل زیر طی شده است:
- همه کدهای ISIC مربوط به مشاغلی که به نحوی در تماس مستقیم با ذغال سنگ هستند، استخراج شده است.
 - تعداد افراد شاغل در این مشاغل در سطح هر شهرستان استخراج شده است.
 - تعداد استخراج‌شده به شهرستان مربوط در محیط نرم‌افزارهای GIS، نسبت داده شده است.

جدول ۱: مهمترین عوامل ایجاد سرطان پوست و واحد مکانی داده‌های مربوط به آن.

Spatial Unit Of Needed Data	Spatial Unit Of Available Data	Skin Cancer Causative Factors	
Polygon	Point	Solar Energy	Natural Elements
Polygon	Point	Hours of Sunshine and UV radiation	
Polygon	Point	Air Humidity	
Polygon	Point	Air Pressure	
Polygon	Point	Air Temperature	
Polygon	Polygon	Contact with Tar	Life & Work Environment
Polygon	Polygon	Contact with Oil	
Polygon	Polygon	Contact with Coal	
Polygon	Polygon	Contact with Sun	
Polygon	Polygon	Fruit and Vegetable Usage	Nutrition
		Red Meat Usage	
		Poultry Usage	
		Sea food Usage	
		Tobacco and Cigarette Usage	
Polygon	Point	Arsenic and Barium In Water and Air	Geological Elements
Polygon	Polygon	Usage of Fertilisers rate in agriculture lands	Amount Of Chemical Fertilisers' Usage

تغذیه: مطابق آمارهای سازمان بهداشت جهانی، یک سوم مرگ و میرهای ناشی از سرطان وابسته به عادت‌های غذایی و تغذیه مردم است. تنها داده‌های موجود در مورد عادت‌های غذایی و تغذیه مردم، مقدار هزینه‌ای است که هر خانواده به طور متوسط در ماه برای مواد غذایی صرف می‌کنند. با تحقیقات به عمل آمده از منابع علمی، عوامل مهم تأثیرگذار در ابتلا به سرطان پوست که وابسته به

- متوسط سالیانه مشاهده‌ها برای هر ایستگاه محاسبه شده است.
- یک لایه نقطه‌ای که نشان‌دهنده موقعیت ایستگاه‌ها است، تولید شده است.
- اطلاعات مربوط به مقادیر شرایط اقلیمی در محیط نرم‌افزارهای GIS با موقعیت ایستگاه‌ها به هم مرتبط شده است.
- برای هر عامل، با استفاده از روش^Y IDW، یک لایه رستری به میانگین سالانه در ایستگاه‌های مشاهداتی، برازش داده شده است. روش IDW بر اساس این اصل استوار است که نقاطی که فاصله کمتری از هم دارند، بیشتر به هم شبیه هستند [۲۳] و [۲۴].
- برای هر عامل شرایط اقلیمی، لایه رستری تخمین زده شده با لایه پلیگونی مربوط به مرز شهرستان‌ها همپوشانی شده است و برای هر شهرستان متوسط مقادیر پیکسل‌های رستری با توابع موجود محاسبه شده و این مقدار به واحد مکانی شهرستان نسبت داده شده است. شکل (۱) نقشه مربوط به دما را که با روش IDW ایجاد شده است، نشان می‌دهد.

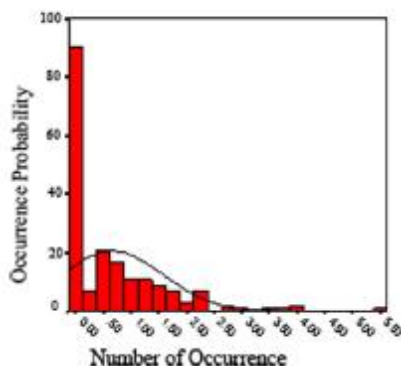


شکل ۱: نقشه رستری مربوط به دما، تولید شده توسط روش IDW.

محیط کار و زندگی: در همه اسناد و مدارک مطالعه‌شده درباره بیماری، از جمله اسناد و مدارک مرکز تحقیقات سرطان، سازمان بهداشت جهانی و وزارت بهداشت، درمان و آموزش پزشکی، مشاغل زیر جزو مشاغلی بوده‌اند که در معرض ابتلا به سرطان پوست قرار دارند: مشاغلی که در آن افراد در معرض آفتاب شدید قرار دارند که مهم‌ترین آنها کشاورزان و ماهیگیران هستند و نیز مشاغلی که در آن افراد با قطران، ذغال سنگ، نفت، و سرب در تماس هستند.

برازش مدل

همان طور که در بخش قبل گفته شد، در این تحقیق مدل GLM برای تعریف یک رابطه میان سرطان پوست و عوامل ایجادکننده آن، به کار رفته است. درستی انتخاب تابع توزیع احتمال مناسب از طریق آزمایش kolmogronov-Smirnov به دست آمده است. با استفاده از این آزمایش، درجه اطمینان برازش داده‌های مربوط به مرگ و میر در اثر سرطان پوست بر تابع توزیع احتمال یواسون، ۹۳ درصد بوده است.



شکل ۲: نمودار تابع توزیع احتمال سرطان پوست.

در این تحقیق از روش برآورد بیشترین احتمال (MLE)^۸ برای به دست آوردن ضرایب مجهول استفاده شده است. با فرض اینکه $l(\beta)$ تابعی است که برآورد بیشترین احتمال آن مدنظر است.

$$l(\beta) = \log [f(y, \theta, \phi)] = \log \{f[y, h(x\beta), \phi]\} \quad (9)$$

با فرض زیر (رابطه ۱۰)

$$U(\beta) = \frac{\partial l(\beta)}{\partial \beta} = \begin{bmatrix} \frac{\partial l(\beta)}{\partial \beta_1} \\ \frac{\partial l(\beta)}{\partial \beta_2} \\ \vdots \\ \frac{\partial l(\beta)}{\partial \beta_p} \end{bmatrix} = \begin{bmatrix} U_1(\beta) \\ U_2(\beta) \\ \vdots \\ U_p(\beta) \end{bmatrix} \quad (10)$$

و نیز

$$A(\beta) = -\frac{\partial^2 l(\beta)}{\partial \beta^2} = -\frac{\partial U(\beta)}{\partial \beta} = \begin{bmatrix} -\frac{\partial^2 l(\beta)}{\partial \beta_1 \partial \beta_1} & -\frac{\partial^2 l(\beta)}{\partial \beta_1 \partial \beta_2} & \dots & -\frac{\partial^2 l(\beta)}{\partial \beta_1 \partial \beta_p} \\ -\frac{\partial^2 l(\beta)}{\partial \beta_2 \partial \beta_1} & -\frac{\partial^2 l(\beta)}{\partial \beta_2 \partial \beta_2} & \dots & -\frac{\partial^2 l(\beta)}{\partial \beta_2 \partial \beta_p} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{\partial^2 l(\beta)}{\partial \beta_p \partial \beta_1} & -\frac{\partial^2 l(\beta)}{\partial \beta_p \partial \beta_2} & \dots & -\frac{\partial^2 l(\beta)}{\partial \beta_p \partial \beta_p} \end{bmatrix} = \begin{bmatrix} A_{11}(\beta) & A_{12}(\beta) & \dots & A_{1p}(\beta) \\ A_{21}(\beta) & A_{22}(\beta) & \dots & A_{2p}(\beta) \\ \vdots & \vdots & \ddots & \vdots \\ A_{p1}(\beta) & A_{p2}(\beta) & \dots & A_{pp}(\beta) \end{bmatrix} \quad (11)$$

تغذیه است، عبارتند از: مصرف میوه و سبزیجات (با اثر کاهش)، مصرف گوشت قرمز، مصرف حیوانات دریایی (با اثر کاهش)، مصرف گوشت سفید (با اثر کاهش) و مصرف دخانیات. تنها داده‌های موجود در این زمینه نمونه‌گیری سالانه هزینه درآمد خانوار توسط سازمان مدیریت و برنامه‌ریزی کشور است که در آن متوسط هزینه هر خانوار برای اقلام ذکر شده موجود است.

مصرف کود شیمیایی: افزایش بی‌رویه استفاده از کودهای شیمیایی باعث ایجاد بسیاری از بیماری‌ها از جمله سرطان پوست خواهد شد [۲۵]. در این تحقیق با توجه به نظر متخصصان، مصرف کودهای شیمیایی پتاس، نترات و فسفات مورد بررسی قرار گرفته است. برای ورود این داده‌ها به مدل، از نسبت مصرف کود شیمیایی در شهرستان‌ها به سطح زیرکشت در آنها استفاده شده است. در پایان این مقدار به واحدهای مکانی مربوطه نسبت داده شده است.

المان‌های ساختاری زمین: رابطه میان ژئوشیمی و محیط شامل سلامت جمعیت، گیاهان و حیوانات، مدت‌ها است که مورد تحقیق دانشمندان علوم مختلف است [۲۶] و [۲۷]. از میان مواد مختلف، باریوم و آرسنیک به عنوان عوامل اصلی ایجاد سرطان پوست شناخته شده‌اند.ⁱⁱⁱ داده‌های موجود از این عناصر تنها میزان انحلال باریوم در آب آشامیدنی مردم است که در کل کشور نمونه‌گیری شده است. برای ورود اطلاعات مربوط به باریوم به مدل، پردازش‌های زیر انجام شده است:

- غلظت باریوم در هر ایستگاه سنجش، با اختصاص یک کد منحصر به فرد به ایستگاه، در جداول ثبت شده است.
- با استفاده از مختصات نقاط نمونه‌گیری، لایه نقطه‌ای مربوط به ایستگاه‌ها ایجاد شده است (به هر ایستگاه یک کد منحصر به فرد نسبت داده شده است).
- اطلاعات توصیفی نسبت داده شده است.
- با استفاده از روش IDW، مقادیر محاسبه شده و یک لایه رستری ایجاد شده است.
- نتایج با لایه مرز شهرستان‌ها همپوشانی شده و متوسط مقادیر مربوط به پیکسل‌ها در محدوده شهرستان محاسبه شده است.

ⁱⁱⁱ لازم به ذکر است، باریوم یک عنصر رادیواکتیو است که در صورت ورود به بدن در طولانی مدت با ایجاد اشعه‌های رادیواکتیو، باعث بروز سرطان می‌شود.

$$f(y) = [\exp(+0.2279 * X23) + 0.1709] \\ \times \exp(-0.0015 * X25) - 0.3620 \\ \times \exp(+7.4129 * X15) - 0.8833 \\ \times \exp(+0.0083 * X24) - 0.9384 \\ \times \exp(+137.0908 * X1) - 0.1001 \\ \times \exp(-0.000001 * 34 * X18) + 0.0063 \\ \times \exp(+111.2451 * X14) + 0.0204 \times \\ \exp(+0.00002504 * X6) - 0.0058 \\ \times \exp(+0.000005914894 * X17) + 0.0055] \\ \sqrt{\exp(-0.00244 * X15 * X24) + 0.8642} \quad (16)$$

جدول ۲: پارامترهای انتخابی توسط مدل و ضریب کاهش یا افزایش آنها.

Variable	Variable description	Sign of coefficient
X1	Contact with sun exposure	Positive
X6	Usage of potassium fertiliser in cultivated area	Positive
X14	Number of factory workers in contact with lead per township population	Positive
X15	Solar energy per area unit	Positive
X17	Average smoking expenses per family	Positive
X18	Average fruit/vegetable expenses per family	Negative
X23	Annual average temperature	Positive
X24	Annual bright sunshine hours	Positive
X25	Annual average atmosphere pressure	Negative
X15*X24	Interactive effect of annual hours of sunlight and solar energy	Negative

اولویت بندی المانها

عوامل منتج شده از به ترتیب سطح اطمینان آنها در مدل به دست آمده از روش قدم به قدم، به شرح ذیل است: دما، فشار (اثر منفی)، اثر متقابل میان انرژی خورشیدی و تعداد ساعاتهای آفتابی، انرژی خورشیدی، تعداد ساعاتهای آفتابی، تماس با آفتاب، مصرف میوه و سبزی (اثر منفی)، تماس با سرب، مصرف کود شیمیایی پتاس و مصرف دخانیات.

همان طور که در المانها دیده می شود، پس از جستجو میان متغیرها، معنی داری اثر متقابل عوامل متوسط انرژی خورشیدی و متوسط تعداد ساعاتهای آفتابی ثابت شده است. این وابستگی به این دلیل است که یکی از عواملی که میزان انرژی خورشیدی را تعریف می کند، تعداد ساعاتهای آفتابی است. بنابراین دو پدیده به طور طبیعی، وابسته هستند.

برای مقایسه سایر روشهای انتخاب عوامل در مدل، المانهای به دست آمده از آنها به ترتیب اولویت در اینجا آورده شده است:

روش انتخاب رو به جلو: دما، فشار، اثر متقابل میان

اگر $\hat{\beta}$ برآورد کننده بیشترین احتمال باشد، باید رابطه ۱۲ برقرار باشد.

$$U(\hat{\beta}) = 0 \quad (12)$$

بنابراین با توجه به موضوع مقدار میانگین می توان رابطه ۱۳ را داشت:

$$-U(\beta_0) = U(\hat{\beta}) - U(\beta_0) = \\ \frac{\partial U(\beta^*)}{\partial \beta} (\hat{\beta} - \beta_0) = -A(\beta^*) (\hat{\beta} - \beta_0) \quad (13)$$

همچنین برای هر $\beta^* \in [\beta_0, \hat{\beta}]$ خواهیم داشت:

$$\hat{\beta} - \beta_0 = A^{-1}(\beta^*) U(\beta_0) \Leftrightarrow \hat{\beta} = \beta_0 + A^{-1}(\beta^*) U(\beta_0) \quad (14)$$

بنابراین در این مرحله با توجه به رابطه ۱۳، ضرایب مجهول برآورد می شوند. لازم به ذکر است این کار با استفاده از تکرار به روش Newton-Raphson انجام شده که در آن تکرار با برقراری شرط رابطه ۱۴، پایان می یابد.

$$|\hat{\beta}_{N+1} - \hat{\beta}_N| < \varepsilon \quad (15)$$

که در آن ε یک عدد بسیار کوچک است.

در نهایت با برآورد ضرایب با استفاده از نرم افزاری که بدین منظور طراحی شده است، یک فرمول آماری استخراج شده است که در آن، سطح اطمینان وجود هر یک از عوامل ایجاد بیماری در مدل و نیز ضرایب آنها در مدل، ذکر شده است.

لازم به ذکر است در این تحقیق همه سه روش انتخاب عامل در مدل اجرا شده که نتایج آنها در بخش نتایج آورده خواهد شد.

در پایان با استفاده از مدل به دست آمده و توسعه^{۱۰} طراحی شده در محیط نرم افزارهای مربوط به سیستم اطلاعات مکانی نتایج تحلیل شده اند.

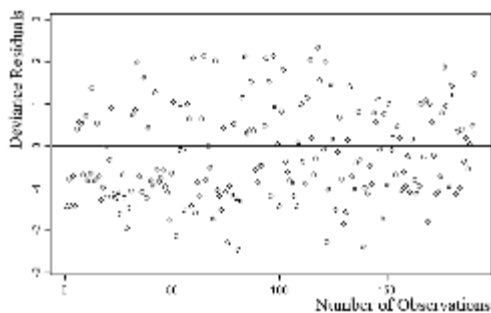
نتایج

رابطه آماری به دست آمده

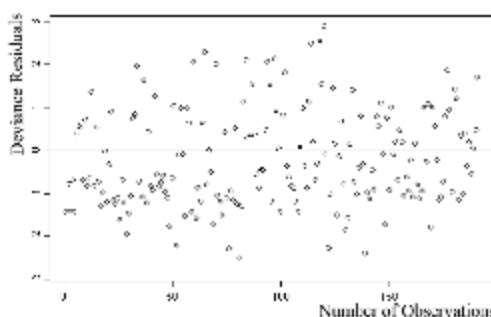
نتایج حاصل از روش قدم به قدم، کمترین باقیمانده را در میان سایر روشهای انتخاب عوامل مدل داشته است و رابطه (۱۵)، رابطه آماری^{iv} منتج از مدل قدم به قدم را نشان می دهد. عوامل رابطه (۱۵) در جدول (۲) توضیح داده شده اند:

^{iv} لازم به ذکر است در روابط آماری باقیماندهی رابطه قابل تحلیل است.

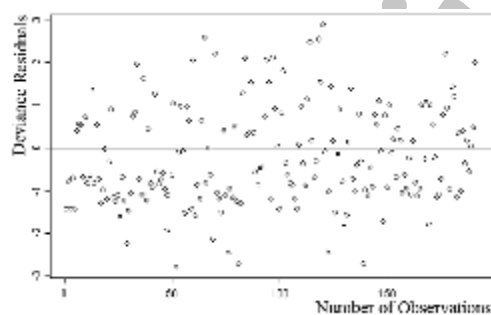
بیشتر باشد، برازش دقت بالاتری دارد. همچنین در نمودار باقی‌مانده‌ها الگویی نباید وجود داشته باشد. در واقع نمودار باقی‌مانده‌ها حول صفر، باید پراکندگی تصادفی داشته باشد.



شکل ۳: باقی‌مانده‌های Deviance روش قدم به قدم.



شکل ۴: باقی‌مانده‌های Deviance روش حذف رو به عقب.



شکل ۵: باقی‌مانده‌های Deviance روش انتخاب رو به جلو.

مقادیر باقی‌مانده‌های مدل به دست آمده از روش قدم به قدم، روی نقشه منطقه مورد مطالعه در شکل (۶) نشان داده شده است. بررسی بر نقشه باقی‌مانده‌ها، نشان‌دهنده بالا بودن این مقادیر در مناطقی است که تعداد بیماری بسیار بالاتر و یا بسیار پایین‌تر از سایر مناطق است. دلیل این موضوع برازش تنها یک مدل بر کل منطقه مورد مطالعه است، زیرا مدل باید از همه تغییرات در منطقه به بهترین شکل عبور کند.

انرژی خورشیدی و تعداد ساعت‌های آفتابی، انرژی خورشیدی، تعداد ساعت‌های آفتابی، تماس با آفتاب و تماس با سرب.

روش حذف رو به عقب: دما، فشار، اثر متقابل میان انرژی خورشیدی و تعداد ساعت‌های آفتابی، تماس با سرب، تماس با آفتاب، اثر منفی مصرف میوه و سبزی، مصرف کود فسفات، مصرف دخانیات.

همان طور که ملاحظه می‌شود، در این دو مدل، تعداد عوامل معنی‌دار کمتر است و این موضوع از آنجا ناشی می‌شود که در این دو مدل با حذف یک متغیر به مدل دیگر نمی‌توان آن را اضافه کرد و یا افزودن متغیرها به صورت تک تک انجام می‌شود و نمی‌توان معنی‌داری یک متغیر را در صورت حضور (یا عدم حضور) متغیرهای دیگر به خوبی روش قدم به قدم بررسی کرد.

آنالیز باقی‌مانده‌های به دست آمده از سه روش انتخاب عامل در مدل

مدل‌های به دست آمده از سه روش انتخاب عامل در مدل، بر اساس باقی‌مانده‌های deviance مقایسه و بررسی شده‌اند که نتایج آن در جدول (۳) آمده است.

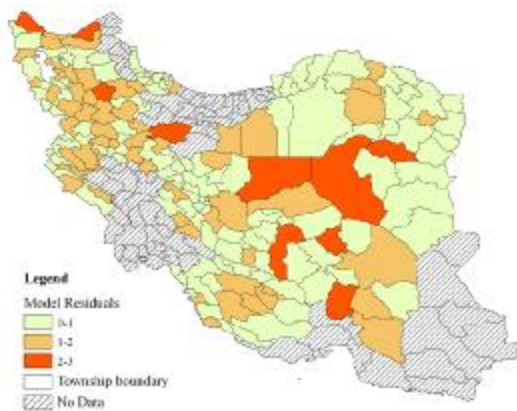
همان طور که در جدول (۳) نمایش داده شده است، باقی‌مانده‌های deviance روش قدم به قدم از سایر روش‌ها کوچک‌تر است. بنابراین این مدل مناسب‌ترین مدل برای نمایش پراکندگی مکانی سرطان پوست در منطقه مورد مطالعه است. بنابراین با اینکه انتخاب عوامل در روش قدم به قدم با دشواری‌هایی مواجه است، ولی مدل به دست آمده از این روش، کفایت بیشتری نسبت به دو روش دیگر دارد.

جدول ۳: مقایسه باقی‌مانده Deviance سه روش انتخاب پارامتر.

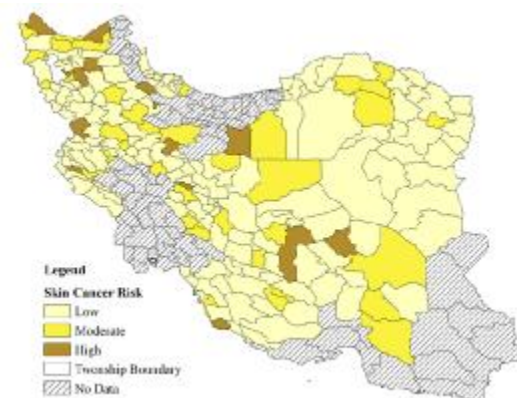
Method of parameter selection	Acceptance level of deviance	Result
Forward selection	141.3549	Accepted
Backward elimination	139.6969	Accepted
Stepwise selection	138.7203	Accepted

شکل‌های (۳) و (۴) و (۵) نشان‌دهنده پراکندگی بهتر باقی‌مانده‌های deviance روش قدم به قدم نسبت به دو روش دیگر است. هر چه پراکندگی باقی‌مانده‌ها حول صفر

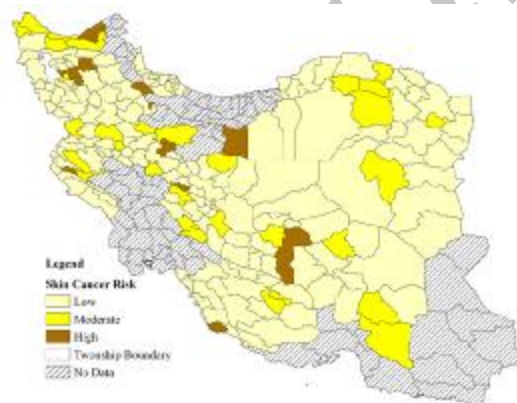
تصمیمات لازم برای مبارزه با بیماری را به عمل آورد.



شکل ۶: نقشه‌ی باقیمانده‌ها در روش قدم به قدم.



شکل ۷: نقشه‌ی احتمال سرطان پوست مربوط به نتایج حاصل از مدل.



شکل ۸: نقشه‌ی احتمال سرطان پوست مربوط به نتایج حاصل از مدل برای حالتی که مصرف میوه و سبزی ۲۰٪ افزایش یافته است.

نتیجه گیری

در این تحقیق با استفاده از تلفیق مدل آماری GLM و امکانات سیستم اطلاعات مکانی، بیماری سرطان پوست به نحوی مدل شده است که قابلیت پیش‌بینی گسترش مکانی

پیش‌بینی بیماری با تغییر عوامل ایجاد آن

مدل به دست آمده، یک رابطه میان تعداد مرگ و میر در اثر سرطان پوست در واحد مکانی شهرستان و عوامل ایجادکننده آن تعریف می‌کند. بنابراین مدل می‌تواند برای مشاهده تغییرات تعداد و پراکندگی مکانی مرگ و میر در اثر سرطان پوست با تغییر هر یک از عوامل ایجادکننده آن به کار رود. مشاهده این تغییرات کمک بزرگی برای تصمیم گیرندگان امور بهداشتی در زمینه پیشگیری بیماری است.

قبل از هر گونه تغییرات در عوامل، لازم است میزان مرگ و میر در اثر بیماری توسط مدل نهایی به دست آمده محاسبه شود. همان‌طور که ملاحظه شد، اختلاف قابل قبولی میان تعداد مرگ و میر به دست آمده از مدل و تعداد واقعی وجود دارد. تعداد تخمین زده شده بیماری از مدل در شکل (۷) نشان داده شده است. در حقیقت شکل (۷)، نشان‌دهنده نقشه احتمال تعداد مرگ و میر به دست آمده از مدل است. نقشه احتمال یک روش مکان - آماری برای نمایش داده‌های پلیگونی آمار بیماری است که در آن با استفاده از روش‌های آماری مناطق تحت خطر بیماری و مکان‌هایی که بیماری در آنها شیوع پیدا کرده است، به دست می‌آید. توضیحات مربوط به این روش، دلیل استفاده و مزایای آن در [۲۸] و [۲۹] وجود دارد.

به عنوان مثال فرض بر این است که با فرهنگ‌سازی و تبلیغات مؤثر، مصرف میوه و سبزی ۲۰ درصد افزایش یابد. برای نمایش این تغییرات، میزان مصرف میوه و سبزی برای واحدهای مکانی مدل /درصد افزایش داده می‌شود. سپس مقادیر جدید مصرف میوه و سبزی و مقادیر قبلی عوامل دیگر وارد مدل شده و مقدار جدید مرگ و میر از مدل محاسبه می‌شود. مقادیر جدید مرگ و میر با استفاده از نقشه‌های احتمال در شکل (۸) نشان داده شده است.

با توجه به نتیجه به دست آمده و مقایسه نقشه‌های احتمال ۷ و ۸، در بسیاری از شهرستان‌ها میزان مرگ و میر پایین آمده است. به دلیل تعداد بالای شهرستان‌ها، در اینجا تغییرات مربوط به هر شهرستان آورده نشده است. ولی با ملاحظات به عمل آمده با افزایش ۲۰ درصد مصرف میوه و سبزی، به طور متوسط تعداد مرگ و میر ۸ درصد کاهش یافته است.

بنابراین همان‌طور که در تحلیل بالا دیده شد، تصمیم‌گیران بهداشت با تغییر هر یک از عوامل بیماری می‌توانند تغییرات بیماری را از نظر تعداد و مکان ملاحظه کرده و

کشت تنها داده موجود بوده و وارد مدل شده است، در صورتی که تأثیر این عامل ابتدا روی گیاهان و سپس از این روش روی انسان است. و این عامل بر دقت مدل تأثیرگذار است. می‌توان در یک کار دقیق‌تر این موارد را نیز با روش‌هایی نظیر مدل‌سازی‌های دیگر و یا آنالیزهای آماری حل کرد.

از آنجا که مدل برازش داده شده به شکل سراسری بوده است، باقی‌مانده‌ها در مناطقی که تعداد بسیار بالا و یا بسیار پایین بیماری را دارند، بالا است زیرا مدل، سعی در برازش کلی و در همه منطقه را دارد و به این دلیل نرم شده است تا به طور یکنواخت از همه مناطق عبور کند.

به عنوان کارهای آینده برای بهتر کردن دقت مدل می‌توان از مدل‌های محلی استفاده کرد. به این ترتیب که مناطقی که از نظر شرایط اقلیمی به هم نزدیک‌تر هستند را در گروه‌های جداگانه مدل‌سازی کرد. در این صورت دقت برازش مدل به طور قابل توجهی افزایش خواهد یافت.

بیماری بر اساس تغییر عوامل ایجاد بیماری را دارد. اصلی‌ترین نتایج به دست آمده از این تحقیق به این شرح است:

همان‌طور که ملاحظه شد، نتایج انتخاب عامل به روش قدم به قدم بهتر و دقیق‌تر از سایر روش‌ها است. مزیت دیگر این روش، بررسی بهتر معنی‌داری عوامل در حضور و یا نبود یکدیگر است. ولی وقتی تعداد عوامل ورودی به مدل زیاد است (مانند این تحقیق)، این کار بسیار وقت‌گیر و پیچیده خواهد بود. با اینکه دقت روش‌های انتخاب رو به جلو و یا حذف رو به عقب به اندازه روش قدم به قدم بالا نیست، ولی این روش‌ها سرعت بیشتری در انتخاب عوامل دارند.

مشکل اساسی در مدل‌سازی در این تحقیق، نبود سنخیت داده‌های مختلف از منابع مختلف یا واقعیت‌المان مورد نظر بوده است. به عنوان مثال در مصرف کودهای شیمیایی، در اینجا میزان مصرف کود به نسبت سطح زیر

مراجع

- 1 - Melnick, A. L. and Fleming, D. W. (1999). "Modern geographic information systems: promise and pitfalls." *Journal of Public Health Management and Practice*, Vol. 5, PP. 8-10.
- 2 - Malczewski, J. (1999). *GIS and Multicriteria Decision Analysis*. 1st Ed. Jhon Wiley & Sons, USA, PP. 15-16.
- 3 - Gordon, A. and Womersley, J. (1997). "The use of mapping in public health medicine." *Journal of Public Health Medicine*, Vol. 2, No. 3, PP. 139-147.
- 4 - Wertheimer, N. and Leeper, E. (1997). "Electric wiring configuration and childhood cancer." *American Journal of Epidemiology*, No. 109, PP. 273-284.
- 5 - Wartenberg, D. (1993). "Identification and characterization of populations living near high-voltage transmission lines: A pilot study." *Environmental Health Perspectives*, No, 101, PP. 626-632.
- 6 - Weisner, C. G. (1994). *Predict asthma and outdoor air pollution in Brooklyn and Queens: A Geographic Analysis*, Master thesis, Departement of Geography, Hunter College, City university of New York, New York.
- 7 - Glass, G. E. (2000). *Geographic Information Systems in infectious disease epidemiology theory and practice*, 1st Ed. Apsen Publishers Inc, Gaithersburg, MD, PP. 231- 254.
- 8 - Dirk, U. P. et al. (2008). *Spatial Analysis in Epidemiology*. 1st.Ed. Oxford University Press, USA, New York, PP. 97-100.
- 9 - Gill, J. (2001). *Generalized Linear Models*. 1st Ed. Sage Publication Inc, London, PP. 51-66.
- 10 - Hible, J. M. and Hadin, J. N. (2001). *Generalized linear Models and Extensions*. Sata Press, USA, PP. 7-13.
- 11 - Montgomery, D, Peck, E. and Vining, G. (2001). *Introduction to Linear Regression Analysis*. 2nd Ed. Wiley Interscience Publication, NewYork, PP. 320-410.
- 12 - Rento, M., Potter, E. and Cavenaghi, S. (2002). "A Bayesian space varying parameter model applied to estimating." *Statistics in Medicine*, John Wiley & Sons, Vol. 21, No. 14, PP. 2057-2075.

- 13 - Lindesey, J. K. (1997). *Applying Generalized Linear Models*, 2nd Ed. Springer, New York, USA, PP. 210-211.
- 14 - Winkelmann, R. (2003). *Econometric Analysis of Count Data*. Springer, New York, USA, PP. 8-10.
- 15 - Ripley, B. D. (2004). *Spatial Statistics*. 1st Ed. Willey Interscience, University of London, PP. 29-75.
- 16 - Burrough, P. (1986). *Geographic Information Systems for Natural Resources Assessment*. 2nd Ed. Oxford University Press, New York.
- 17 - Nelder, J. A. and McCullagh, P. (1989). *Generalized Linear Models*, CRC Press, Washington, D. C., PP. 466-487.
- 18 - Roa, C. R. and Toutenburg, H. (1999). *Linear Models: Least Sqaears and Alternatives*. Springer, New York, USA, PP. 164-394.
- 19 - Weiss, R. E. (2005). *Modelling Longitudinal Data*. 1st Ed. Springer, New York, USA, PP. 151-152.
- 20 - Farhrmeir, L. and Tutz, G. (2001). *Multivariate Statistical Modelling Based on Generalized Linear Models*. 1st Ed. Spinger, New York, USA, PP. 140-145.
- 21 - Piegorsch, W. W. and Bailer, A. J. (2005). *Analyzing Environmental Data*. 1st Ed. John Wiley and sons, PP. 104-105.
- 22 - Miller, A. J. (2002). *Subset Selection in Regression*, CRC Press, Washington, D.C., PP. 39-46.
- 23 - Clarke, K. C. (1996). *An Epidemiology and Geographic Information Systems: A Review and Discussion of Future Directions*. 2nd Ed. Emerging infectious Diseases 2, Oxford University Press, New York, PP. 85-92.
- 24 - Goodchild, M. F., Steyeart, L. T., Parks, B. O., Johnston, C., Maindment, M. C. and Glendinging, S. (1996). *GIS and Environmental Modelling: Progress and Research Issues*. 2nd Ed. John wiley and sons, USA, PP. 290- 300.
- 25 - Ghazban, F. (2003). *Environmental Geology*. 2nd Ed. Tehran University Publication, Tehran, PP. 60-90, "In Farsi".
- 26 - Gatrell, A. C. (2001). *Geographies of Health: An Introduction*, Blackwell, USA, PP. 215-255.
- 27 - Skidmore, A. (2002). *Environmental Modelling with GIS and Remote Sensing*, 1st Ed. Taylor & Francis, London, PP. 8-55.
- 28 - Mesrai, M.S. and Massoomy, Z. (2006). "Detection of lung cancer clusters in Iran using GIS and statistical analyses." *International Journal of Geoinformatics*, Vol. 2, No. 4, PP. 59-63
- 29 - Tim, U. S. (1995). "The application of GIS in environmental health sciences: opportunities and limitations." *Environmental Research*, Vol. 71, No.2, PP.75-88.

واژه های انگلیسی به ترتیب استفاده در متن

- | | |
|---------------------------------|--------------------------------------------|
| 1 - Generalized Linear Model | 2 - Link Function |
| 3 - Forward Selection | 4 - Backward Elimination |
| 5 - Stepwise Regression | 6 - World Health Organization |
| 7 - Inverse Distance Weighted | 8 - International Standard Industrial Code |
| 9 - Maximum Likelihood Estimate | 10 - Extension |
| 11 - Smooth | |