

برازش یک مدل مکان-آماري بر بیماری سرطان پوست به

منظور حمايت از تصميم گيري

زهره معصومي کارشناس ارشد مهندسي سيستم هاي اطلاعات جغرافيايي (GIS) دانشگاه خواجه نصيرالدين طوسي

Email: z_massoomy@yahoo.com

دکتر محمد سعدي مسگري عضو هيئت علمي گروه مهندسي سيستم هاي اطلاعات جغرافيايي (GIS) دانشگاه خواجه نصيرالدين طوسي

Email: smesgari@yahoo.com

چکیده

مطالعه جغرافيايي منطقه در بررسی بیماری ها از اهميت خاصی برخوردار است زیرا وقوع بسیاری از بیماری ها به شرايط جغرافيايي منطقه وابسته است. وجود موانع طبيعي نظير کوهستان ها و درياها، از یک طرف و وابستگي بعضی از عوامل بیماری به شرايط محيطی و اقليمي خاص، از طرف ديگر باعث محدود شدن بعضی از بیماری ها به مناطق خاصی از کره ی زمین گردیده است بنابراین جغرافيا و مکان در بررسی بیماری ها اهميت خاصی پیدا می کنند.

با استفاده از مدل سازی پدیده ی بیماری، می توان عوامل اصلي ایجاد بیماری در یک منطقه را به همراه میزان اهميت هر عامل شناسایی و تعیین کرده و برای پیشگیری و درمان آن بیماری، بیشترین هزینه و زمان را صرف عوامل مهم تر استخراج شده از مدل کرد. در این تحقیق هدف برازش یک مدل مکان-آماري بر بیماری سرطان پوست بوده است. برای این کار در ابتدا مدل مناسبی برای بیماری با توجه به ذات بیماری و نحوه ی وقوع آن و نیز با توجه به نحوه ی جمع آوری داده تعیین گردید. بدنبال آن، عوامل ایجاد کننده ی بیماری با انجام مطالعات لازم استخراج شده اند. در مرحله بعد داده های مربوط به این عوامل، جمع آوری شده و با توجه به ذات داده ها و مدل مورد نظر، پردازش و آماده سازی شدند. پس از آماده سازی اطلاعات، متغیر تعداد سرطان پوست در واحد مکانی شهرستان به عنوان متغیر وابسته (متغیر پاسخ) و عوامل بیماری به عنوان متغیرهای مستقل وارد مدل شده و مدل مناسب بر عوامل ایجاد بیماری برازنده شده است. نتیجه ی استخراج شده از برازش مدل، یک فرمول مشخص و رتبه بندی عوامل ایجاد بیماری سرطان پوست است.

واژگان کلیدی: GIS، آنالیزهای مکانی، آنالیزهای آماری، مدل، سرطان پوست.

مقدمه

در مطالعات مربوط به بیماری، در برخی موارد یافتن ارتباط میان بیماری و عوامل ایجاد کننده‌ی آن، با استفاده از توابع آنالیز ساده‌ی GIS مانند توابع همپوشانی لایه‌ها و بافر امکان پذیر است ولی در موارد پیچیده‌تر استفاده از روش‌های آماری در مدل کردن بیماری ضروری می‌نماید [5]. برازش مدل به بیماری منجر به فرمولی مشخص خواهد شد که رابطه موجود میان بیماری به عنوان متغیر پاسخ و عوامل به وجود آورنده آن به عنوان متغیرهای مستقل را بیان خواهد کرد. همچنین با استفاده از این فرمول می‌توان به بررسی وضعیت بیماری در زمان حال و حتی آینده پرداخت زیرا تاثیر تغییر هر یک از متغیرهای مستقل، بر روی متغیر وابسته با استفاده از فرمول قابل بررسی است. با استفاده از مدل کردن بیماری می‌توان اهداف مختلفی را دنبال کرد که از جمله‌ی آنها می‌توان به موارد ذیل اشاره کرد:

- تشخیص عوامل تأثیرگذار بر بیماری، از این طریق می‌توان درجه‌ی اهمیت عوامل بیماری را بدست آورده و آنها را رتبه‌بندی نمود.
- پیش‌بینی گسترش بیماری، از طریق مدل بدست آمده و تغییر هر یک از پارامترهای بوجود آورنده‌ی آن، می‌توان درجه‌ی تأثیر هر یک از پارامترها را بر روی بیماری به صورت واضح و قابل لمس مشاهده نمود.
- اولویت‌بندی بودجه برای مبارزه با عوامل بیماری، با مشخص شدن درجه‌ی اهمیت هر یک از عوامل تأثیرگذار بر بیماری، اولویت‌های اقتصادی برای پیشگیری نیز مشخص می‌شوند. به این ترتیب صرفه‌جویی اقتصادی با اختصاص دادن بودجه‌ی بیشتر به عامل مهم‌تر، عملی می‌شود.

در این تحقیق هدف، برازش یک مدل مکان-آماري بر بیماری سرطان پوست و اولویت‌بندی عوامل ایجاد کننده‌ی این بیماری می‌باشد.

از آنجاییکه مکان عامل بسیار مهمی در وقوع برخی از بیماری‌ها می‌باشد، بنابراین نیاز به توابع تجزیه و تحلیل مختلف GIS در بیشتر مراحل مدلسازی ضروری می‌نماید.

این تحقیق، در سه قسمت انجام شده است. در فاز اول، چارچوب اطلاعاتی مربوط به بیماری سرطان پوست با توجه به عوامل بوجود آورنده‌ی آن طراحی شده است. در فاز دوم، داده‌های مورد نیاز برای مدل کردن بیماری با در نظر گرفتن چارچوب اطلاعاتی بیماری، جمع آوری شده و پردازش‌های لازم بر روی داده‌ها برای ورود به مدل انجام گرفته است. در فاز سوم نیز مدل مناسب به داده‌ها برازش داده شده و کفایت آن بررسی شده است.

۱- طراحی چارچوب اطلاعاتی

برای طراحی چارچوب اطلاعاتی، کلیه اطلاعات و مستندات مربوط به علل بروز سرطان پوست در وزارت بهداشت، لیست عوامل سرطان‌زا که از طرف NCI¹ آمریکا منتشر شده است، نظرات متخصصین پوست و نمونه‌های موردی انجام شده در سازمان بهداشت جهانی و دیگر ارگان‌های بین‌المللی استخراج شده و مورد بررسی قرار گرفته است. در بررسی و ارزیابی مستندات مذکور، دستور کار ذیل مدنظر بوده است:

¹ National Cancer Institute

• بررسی علل اصلی ایجاد و علل افزایش دهنده‌ی این بیماری و تعیین موجودیت‌های مکانی و غیر مکانی مورد نیاز،

- شناسایی نقش موجودیت‌های مکانی انتخاب شده
- شناسایی و کشف اقلام توصیفی مورد نیاز و قابل انتساب به هر موجودیت،
- شناسایی ویژگی‌های هر قلم توصیفی شناسایی شده
- روش مورد نظر برای مدل کردن بیماری، روش آماری است. بنابراین لازم است پردازش‌های مربوط به مدل بر روی اقلام توصیفی، با واحد مکانی یکسان انجام گیرد. بنابراین واحد مکانی مورد نیاز برای تمامی عوامل، پلیگون شهرستان‌های ۱۸ استان کشور می‌باشد. زیرا داده‌های مربوط به بیماری که متغیر پاسخ بوده و مدل بر اساس آن برآزاندن خواهد شد، به صورت تعداد مرگ و میر در اثر بیماری در شهرستان‌های ۱۸ استان کشور ثبت شده است. به این ترتیب موجودیت‌های مورد نیاز برای ورود به مدل استخراج شده اند.

۲- جمع آوری و پردازش داده‌ها

در GIS، پردازش به مجموعه عملیاتی گفته می‌شود که بر روی داده‌های خام اعمال می‌شود تا برای ورود به تابع تجزیه و تحلیلی آماده گردند [۳]. هر پردازش شامل یک یا چند ورودی و یک یا چند خروجی می‌باشد. نوع پردازشی که بر روی داده‌ها اعمال می‌گردد به ساختار داده‌های خام ورودی (برداری یا رستری) و ساختار داده‌ای مورد نیاز جهت ورود به تابع تجزیه و تحلیلی بستگی دارد.

برای اینکه مدل پردازش داده شده به صورت درست و بهینه عمل کند، باید داده‌های ورودی آن با انجام مجموعه عملیات پردازشی موجود در GIS، آماده‌سازی شوند. انتخاب روش آماده‌سازی و پردازش‌های مرتبط به آن، بستگی به نوع داده‌های ورودی و نحوه‌ی عملکرد مدل دارد. در این تحقیق با توجه به نوع مدل انتخاب شده و تعیین نوع ورودی و خروجی آن، موارد ذیل، به عنوان پردازش‌های مطلوب جهت آماده‌سازی لایه‌ها استفاده شده است.

- تبدیل داده‌های نقطه‌ای به پلیگونی با استفاده از روش‌های انترپلاسیون
- استخراج میانگین مقادیر پیکسل‌ها در یک محدوده‌ی بسته با استفاده از توابع Focal
- ارتباط اطلاعات توصیفی با لایه‌های اطلاعات مکانی

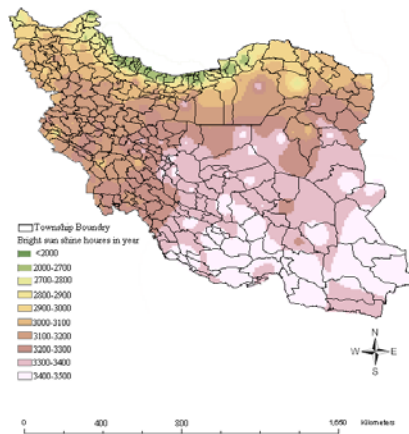
در اینجا روش آماده‌سازی هر لایه و پردازش‌هایی که بر روی هر گروه از موجودیت‌ها انجام شده است به صورت مجزا تشریح گردیده است.

۲-۱- شرایط جوی

به طور کلی محیط به مجموعه‌ای از عوامل و شرایط خارجی و تأثیرات وارده ناشی از آنها بر زندگی یک موجود زنده اطلاق می‌گردد. طبق این تعریف محیط شامل هوا و اقلیم، آب و خاک، شرایط جوی و روابط بین آنها و کلیه‌ی موجودات زنده می‌باشد. بدون شک شرایط جوی و آب و هوا نیز می‌توانند عوامل بروز برخی از بیماری‌ها باشند. با توجه به بررسی‌های انجام شده و نظر متخصصین پوست، دمای بالا، خشکی هوا، فشار پایین جوی و قرار گرفتن در معرض نور خورشید، در منطقه‌ی مورد مطالعه مواردی هستند که تأثیرات نامطلوبی بر روی پوست دارند و در صورتیکه پوست شخص بدون رعایت اصول بهداشتی دائماً در برابر این عوامل قرار گیرد، آسیب پذیر خواهد بود. بنابراین میانگین دمای سالیانه، میانگین رطوبت سالیانه، میانگین فشار جوی سالیانه و تعداد ساعات آفتابی در سال و میزان انرژی خورشیدی در منطقه برای بررسی تأثیر عوامل جوی بر روی سرطان پوست انتخاب شده‌اند.

مشاهدات مربوط به پارامترهای جوی در ایستگاه‌های سینوپتیک هر ۳ ساعت و در ایستگاه‌های اقلیم‌شناسی هر ۱۲ ساعت اندازه‌گیری شده و به ثبت می‌رسد. در این تحقیق از مشاهدات سالانه‌ی مربوط به این ایستگاه‌ها برای استخراج میانگین دما، فشار، رطوبت، انرژی خورشیدی و تعداد ساعات آفتابی استفاده شده است. مراحل تبدیل و پردازش داده‌ها به شکل مورد نظر عبارتند از:

- استخراج میانگین دما، فشار، رطوبت و انرژی خورشیدی سالیانه همچنین تعداد ساعات آفتابی در سال از فایل‌های مشاهداتی ایستگاه‌های سینوپتیک سازمان هواشناسی کشور، اختصاص یک کد منحصری‌فرد به هر ایستگاه و انتساب اطلاعات توصیفی به واحد مکانی مربوطه (نقطه‌ی ایستگاه سینوپتیک) با استفاده از کد منحصری‌فرد،
 - تولید فایل مربوط به اطلاعات موجود با فرمت نقطه‌ای و انترپولاسیون و برازش یک سطح رستری به هر یک از لایه‌ها با استفاده از متد IDW
 - استخراج میانگین دما، فشار، رطوبت، انرژی خورشیدی و تعداد ساعات آفتابی در سال با توجه به تقسیمات کشوری در سطح شهرستان و سطح رستری تشکیل شده.
- به عنوان مثال، شکل ۱ سطح رستری مربوط به تعداد ساعات آفتابی، تهیه شده با روش IDW را نشان می‌دهد.



شکل ۱- سطح رستری مربوط به تعداد ساعات آفتابی، تهیه شده با روش IDW، به همراه مرز سیاسی شهرستان‌های کشور

۲-۲- مشاغل

محیط کار و زندگی انسان می‌تواند به صورت مستقیم یا غیرمستقیم بر روی سلامت وی تأثیرگذار باشد. بعضی از آمارهای WHO^۲، گویای این واقعیت است که یک میلیون انسان سالانه در اثر تماس غیرایمن با موادی که با آنها در تماس هستند دچار مرگ شده و یا از کار افتاده می‌گردند لذا مدیریت ایمن محیط کار و رعایت بهداشت حرفه‌ای یک انتخاب نیست بلکه یک ضرورت و اولویت است [۱].

با توجه به لیست عوامل ایجادکننده‌ی سرطان پوست که از طرف NCI و وزارت بهداشت، درمان و آموزش پزشکی اعلام شده است، مشاغل زیر به عنوان مشاغلی که در معرض خطر ابتلا به سرطان پوست قرار دارند، معرفی شده‌اند:

- مشاغلی که با آفتاب تماس زیاد دارند،

² World Health Organization

• مشاغلی که با قطران، مشتقات نفت، ذغال سنگ و سرب در تماس مستقیم هستند. کشور ما یکی از غنی‌ترین کشورها در زمینه معادن است. واضح است پتانسیل وجود مواد مختلف در کشور و پراکندگی آن به مکان بستگی دارد. بنابراین در هر منطقه با توجه به مواد معدنی موجود، فعالیت‌های معدنی، استخراج و ساخت انجام می‌گیرد. برای بررسی ارتباط مشاغل نامبرده و میزان مرگ و میر در اثر سرطان پوست، از طبقه‌بندی بین‌المللی فعالیت‌های اقتصادی (I.S.I.C)³ در سه قسمت معادن، استخراج، ساخت و تولید استفاده شده است. برای هر یک از مشاغل با در نظر گرفتن کدهای I.S.I.C مرتبط، تعداد شاغلین آن کد در واحد مکانی شهرستان استخراج شده است.

۲-۳- تغذیه

زمانی تصور می‌شد که رژیم‌های غذایی نقشی در مرگ و میرهای ناشی از سرطان ندارند. امروزه محققان ثابت کرده‌اند که رژیم غذایی روزانه‌ی افراد نقش مهمی در پیشگیری، بروز و درمان انواع سرطان‌ها ایفا می‌کند. طبق آمار منتشره از سوی WHO حدود یک‌سوم سرطان‌هایی که در نهایت منجر به مرگ می‌شود با تغذیه در ارتباط هستند. با توجه به نظر متخصصین در این زمینه و نیز با توجه به اطلاعات موجود، ارقام بررسی شده در این تحقیق عبارتند از: متوسط هزینه‌ی کل سالیانه مصرفی هر خانواده برای میوه و سبزی، گوشت قرمز، گوشت حیوانات دریایی، گوشت پرندگان و دخانیات که با استفاده از یک تقریب در سطح شهرستان محاسبه شده است.

۲-۴- کود شیمیایی

اگر کودهای شیمیایی در محصولات کشاورزی، بیش از اندازه مورد استفاده قرار گیرند، به بافت‌های بدن آسیب رسانده و باعث بروز انواع سرطان خواهند شد. طبق بررسی‌های انجام شده در NCI آمریکا نیز، مصرف کودهای شیمیایی یکی از عوامل ابتلا به انواع سرطان می‌باشد. سرطان پوست نیز یکی از انواع سرطان‌هایی است که متأثر از این عامل است. بنابراین بررسی میزان مصرف کودهای شیمیایی با توجه به سطح زیر کشت محصولات کشاورزی، ضروری است.

برای بررسی میزان کود مصرفی در هر شهرستان، مراحل زیر طی شده است:

- سطح زیر کشت محصولات کشاورزی هر شهرستان استخراج شده است،
- میزان کود مصرفی در هر شهرستان استخراج شده است،
- نسبت کود مصرفی به سطح زیر کشت محصولات کشاورزی در هر شهرستان بدست آمده است.

۲-۵- عوامل زمین شناسی

رابطه‌ی بین ژئوشیمی یا ترکیب شیمیایی زمین و محیط زیست که شامل گیاهان و جانداران و انسان می‌شود، سال‌هاست که مورد توجه واقع شده و در این زمینه آگاهی‌های فزاینده‌ای بدست آمده است. برای مثال، کمبود یا افزایش برخی از عناصر مشخص در سنگ‌ها، آب و خاک، از مهم‌ترین عوامل بروز بعضی بیماری‌ها در گیاهان، حیوانات و انسان‌ها به شمار می‌آیند. رابطه‌ی بین ژئوشیمی و سلامت انسان بسیار پیچیده به نظر می‌رسد. سلامت و بهداشت انسان، ممکن است تحت تأثیر عناصر شیمیایی گوناگونی که از طریق غذا، آب آشامیدنی و هوا وارد بدن می‌شوند، قرار گیرد. بر اساس آخرین اطلاعات ارائه شده توسط کمیته‌ی علمی سازمان ملل در زمینه‌ی اثرات پرتوهای اتمی موجود در خاک، متوسط پرتوگیری هر شخص از منابع پرتوزای طبیعی معادل $2/4$ میلی سیورت در سال برآورد گردیده است. حال آنکه متوسط پرتوگیری وی از کلیه‌ی فعالیت‌های هسته‌ای شامل ریزش‌های ناشی از انفجارات اتمی و همچنین

³ International Standard Industrial Classification

پرتوگیری‌های پزشکی (تشخیص و درمان) در اثر کاربرد مواد پرتوزا و دستگاه‌های پرتوساز، حدود ۰/۸ میلی سیورت در سال تخمین زده شده است. بدین لحاظ منابع طبیعی پرتوزا عامل اصلی در پرتوگیری انسان و موجودات زنده محسوب می‌گردند [۴].

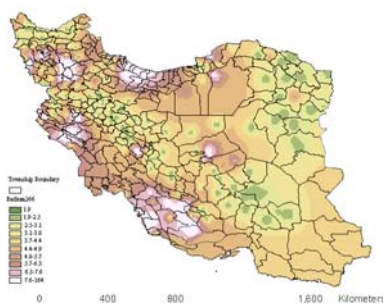
یکی از مهم‌ترین رادیونوکلئیدهای طبیعی موجود در زنجیره‌ی غذایی انسان و به ویژه آب رادیوم (^{226}Ra) است که سابع‌کننده‌ی پرتوهای آلفا می‌باشد. پرتوزایی این عنصر در بدن در اثر بلع آن بسیار خطرناک است زیرا این رادیونوکلئید عنصری استخوان‌گرا (Bone seeker) می‌باشد و به دلیل نیمه‌عمر بالا در استخوان‌ها باقی‌مانده و بر اثر تابش پرتوهای آلفا خطرات جدی را برای سلامتی انسان از جمله سرطان مغز استخوان و سایر سرطان‌ها به وجود می‌آورد [۲].

با توجه به توضیحاتی که داده شد، همچنین تحقیقات انجام شده در زمینه بررسی ارتباط میان ابتلا به سرطان و وجود مواد پرتوزا در طبیعت، توسط مؤسسه‌ی بین‌المللی سرطان، مطالعه‌ی رفتار مربوط به پرتوزایی رادیوم 226 در آب‌های آشامیدنی ایران اهمیت بسزایی یافته و یکی از اهداف این مطالعه قرار گرفته است.

کنترل و بررسی پرتوزایی آب‌های ایران از سال ۱۳۶۵ به صورت پروژه‌ی مشترک با آژانس بین‌المللی انرژی اتمی اجرا گردیده است و نتایج حاصل از آن به صورت مقالات در کنفرانس‌ها و مجلات بین‌المللی ارائه شده است. از سال ۱۳۷۴ از کلیه‌ی استان‌ها نمونه‌برداری آب آشامیدنی بعمل آمده است. از آنجاییکه منابع تأمین‌کننده‌ی آب آشامیدنی ممکن است از یک چاه خصوصی که آب مورد نیاز یک خانواده را تأمین می‌کند تا مجموعه‌ای از چاه‌ها که توسط سیستم شبکه‌ی شهری، آب مصرفی میلیون‌ها انسان را تأمین می‌کند تغییر نماید. بنابراین نمونه‌برداری‌ها در مرحله‌ی اول از شبکه‌ی آبرسانی شهری که دارای منابع تغذیه‌ی متفاوت از قبیل: چاه، چشمه، قنات و رودخانه می‌باشد بعمل آمده و در مواردی که دسترسی به شبکه آبرسانی وجود نداشته است از آب چاه‌ها، چشمه‌ها و یا قنات‌های مورد استفاده‌ی مردم نمونه‌برداری بعمل آمده است.

برای تبدیل داده‌های موجود این نمونه‌برداری، به صورت مورد نظر یعنی غلظت این ماده در محدوده‌ی شهرستان، مراحل زیر طی شده است:

- غلظت اندازه‌گیری شده در نقاط نمونه‌برداری وارد فایل‌های جدولی شده است،
 - هر نقطه‌ی نمونه‌برداری به مکان متناظر خود منتسب شده است،
 - سطح رستری مربوط به داده‌ها با استفاده از روش IDW تولید شده است.
- شکل ۲ سطح رستری مربوط به غلظت عنصر رادیوم 226 در آب‌های آشامیدنی ایران را نشان می‌دهد.



شکل ۲- سطح رستری مربوط به غلظت عنصر رادیوم 226 در آب‌های آشامیدنی ایران، به همراه مرز سیاسی شهرستان‌ها

۳- تعیین مدل برازش سرطان پوست

در این قسمت از تحقیق بر اساس خصوصیات آماری بیماری و عوامل ایجادکننده آن مدل مناسب انتخاب شده و به داده‌ها برازنده شده است. با مطالعات به عمل آمده مدل مناسب، مدل GLM⁴ می‌باشد. مدل GLM در سال ۱۹۷۲ توسط Nelder و Wedderburn فرموله شده و پس از آن مورد استقبال متخصصین آمار و ریاضی قرار گرفته است [6]. این مدل آماری زمانی بکار می‌رود که تابع توزیع احتمال متغیر پاسخ، با احتمال قابل قبولی معلوم باشد. تعداد مرگ و میر در اثر سرطان پوست در واحد مکان با احتمال قابل قبولی دارای تابع توزیع احتمال پواسون می‌باشد. زیرا اولاً ماهیت این داده‌ها حالت باینری دارد و به صورت شمارشی و گسسته است. ثانیاً احتمال رخداد این پدیده در فراوانی‌های پائین، بیشتر است. همچنین احتمال بالای پواسون بودن تابع توزیع احتمال این جامعه با استفاده از تست Kolmogronov-Smirnov بدست آمده است.

۳-۱- انتخاب متغیرها برای ورود به مدل

با استفاده از یک گروه متغیر مستقل مشخص، می‌توان مدل‌های متفاوتی ایجاد کرد. وقتی تعداد متغیرهای مستقل کم باشد، می‌توان تمامی مدل‌های ممکن را بررسی کرد. اما در صورتیکه تعداد متغیرهای مستقل زیاد باشد، بررسی تمامی مدل‌هایی که می‌توان با آنها ساخت نیازمند صرف زمان زیادی است. برای حل این مشکل، روش‌های مختلفی ارائه شده است که به محاسبات متعدد نیاز ندارد. این روش‌ها به صورت متوالی، متغیرهایی را به مدل اضافه یا کم می‌کنند. و تصمیم‌گیری در مورد ورود یک متغیر به مدل و یا خروج یک متغیر از مدل بر اساس مقدار باقیمانده‌های مدل صورت می‌گیرد. همانطور که گفته شد، در اینجا متغیر پاسخ تعداد مرگ و میر در اثر سرطان پوست در واحد مکانی شهرستان و پارامترها (متغیرهای مستقل)، عوامل به وجود آورنده بیماری هستند.

روش‌های انتخاب متغیرها برای ورود به مدل به سه دسته‌ی اصلی تقسیم می‌شوند [7]:

• انتخاب رو به جلو^۵

در این روش، در وهله‌ی اول یک معیار آماری برای ورود متغیر به مدل تعریف می‌شود و فرض بر این است که هیچ متغیر مستقلی در مدل وجود ندارد. اولین متغیری که وارد مدل می‌شود، متغیری است که بزرگ‌ترین ضریب وابستگی با متغیر پاسخ را دارد. متغیر بعدی که وارد مدل می‌شود متغیری است که در صورت وجود متغیر اول، دارای بزرگ‌ترین ضریب وابستگی با متغیر پاسخ باشد. عملیات زمانی خاتمه می‌یابد که معیار محاسبه شده‌ی هیچکدام از متغیرها، از معیار ورود به مدل بزرگ‌تر نباشند و یا هیچ متغیری برای ورود به مدل باقی نماند [7].

• روش حذف رو به عقب^۶

در این روش ابتدا یک معیار آماری برای حذف متغیر از مدل تعریف شده و تمامی متغیرهای مستقل وارد مدل می‌شوند. سپس معیار آماری هر متغیر، محاسبه می‌شود. کوچک‌ترین معیار به دست آمده با مقدار معیار آماری از پیش تعیین شده مقایسه می‌شود. اگر مقدار محاسبه شده از مقدار از پیش تعیین شده کمتر باشد، متغیر از مدل حذف می‌شود. عملیات حذف متغیرها تا زمانی ادامه پیدا می‌کند که کوچک‌ترین مقدار ضریب وابستگی محاسبه شده، از معیار حذف متغیر بزرگ‌تر باشد [7].

• روش انتخاب قدم به قدم متغیرها^۷

⁴ Generalized Linear Models

⁵ Forward

⁶ Backward Elimination

⁷ Stepwise

این روش معمول‌ترین و با دقت‌ترین روش انتخاب متغیر محسوب می‌شود. روش انتخاب قدم به قدم متغیر، تعمیم یافته‌ی روش انتخاب رو به جلو است. با این تفاوت که هر بار بعد از وارد کردن یک متغیر به مدل، متغیرهایی که تاکنون وارد مدل شده‌اند ولی پس از ورود متغیر جدید پیش‌بینی‌کننده‌ی معنی‌داری نیستند، از مدل خارج می‌شوند. در واقع متغیرهایی که اهمیت‌شان با اضافه شدن متغیرهای دیگر کاهش می‌یابد، از مدل حذف می‌گردند [7]. عملیات فوق تا زمانی ادامه پیدا می‌کند که متغیری باقی نماند که در شرط ورود به مدل صدق کند.

برای انتخاب عوامل اصلی ایجادکننده‌ی بیماری سرطان پوست، هر سه روش انتخاب رو به جلو، حذف رو به عقب و قدم به قدم استفاده شده است. لازم به ذکر است میزان معنی‌داری هر متغیر مستقل در مدل با استفاده از تست آماری تعیین می‌شود. متغیرهای بدست آمده از این سه روش به ترتیب اهمیت‌شان در جدول ۱ آورده شده‌اند.

۳-۲- بررسی کفایت^۸ مدل

مانند سایر روش‌های برازش مدل، در روش GLM نیز آنالیز باقیمانده‌ها در برازش مدل بسیار مهم است. زیرا باقیمانده‌ها چگونگی برازش مدل را نشان می‌دهند. روش‌های مختلفی برای بررسی کفایت مدل وجود دارد که تمامی آنها بر اساس باقیمانده‌های مدل عمل می‌کنند. روش غالب برای بررسی باقیمانده‌ها در مدل GLM، باقیمانده‌ی Deviance است. ملاک Deviance یک آماره است که دارای تابع توزیع احتمال کای اسکور با درجه‌ی آزادی معادل تعداد پارامترهای موجود در مدل، می‌باشد. از آنجاییکه Deviance بر مبنای خطای مدل بنا شده است، هر چه مقدار این آماره کمتر باشد بهتر است [7].

همانطور که گفته شد، در این تحقیق سه مدل با استفاده از سه روش انتخاب متغیر انتخاب رو به جلو، حذف رو به عقب و انتخاب قدم به قدم، بدست آمده است. در این سه مدل، تمامی عوامل معنی‌دار بوده و مستقل از هم می‌باشند. در جدول ۱ این مدل‌ها از نظر میزان باقیمانده‌ی Deviance بدست آمده مورد مقایسه قرار گرفته‌اند. همانطور که ملاحظه می‌شود، باقیمانده‌ی Deviance مدل به دست آمده از انتخاب پارامتر به روش قدم به قدم، از دو روش دیگر کمتر است بنابراین این مدل، به عنوان مدل مناسب انتخاب شده است.

جدول ۱- پارامترهای معنی‌دار موجود در سه مدل بدست آمده به ترتیب معنی‌داری

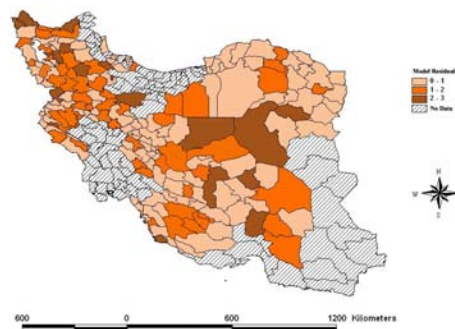
روش بدست آوردن پارامترها	پارامترهای موجود در مدل (به ترتیب معنی‌داری)	معیار Deviance برای پذیرفته شدن مدل	معیار Deviance مدل	نتیجه
انتخاب رو به جلو	دما، فشار، اثر متقابل ^۹ تعداد ساعات آفتابی و انرژی خورشیدی، انرژی خورشیدی، تعداد ساعات آفتابی، تعداد کشاورزان در منطقه، تماس با سرب	۱۴۱،۳۵۴۹	۲۳۹،۷۸۹۷	تست پذیرفته می‌شود.
حذف رو به عقب	دما، فشار، اثر متقابل تعداد ساعات آفتابی و انرژی خورشیدی، انرژی خورشیدی، تعداد ساعات آفتابی، تماس با سرب، تعداد کشاورزان در منطقه، مصرف میوه و سبزی، مصرف کود شیمیایی فسفات	۱۳۹،۶۹۷۹	۲۳۶،۳۳۳۱	تست پذیرفته می‌شود.
انتخاب قدم به قدم	دما، فشار، اثر متقابل تعداد ساعات آفتابی و انرژی خورشیدی، انرژی خورشیدی، تعداد ساعات آفتابی، تعداد کشاورزان در منطقه، مصرف میوه و سبزی، تماس با سرب، مصرف کود شیمیایی پتاس، مصرف دخانیات	۱۳۸،۷۲۰۳	۲۳۳،۴۶۳	تست پذیرفته می‌شود.

^۸ Adequacy

^۹ گاهی اوقات میان متغیرها همبستگی ذاتی وجود دارد. این پدیده غالباً زمانی رخ می‌دهد که پدیده‌های مورد مطالعه، پدیده‌های طبیعی باشند و یک همبستگی طبیعی در آنها موجود باشد [8]. این اثر، اثر متقابل نام دارد. وقتی دو پدیده X_1 و X_2 دارای اثر متقابل باشند، (با ثابت نگه داشتن تمام متغیرهای موجود در مدل) تغییر یکی، دیگری را تحت تأثیر قرار می‌دهد. اثر متقابل در مدل خود را به صورت حاصل ضرب دو عامل نشان می‌دهد.

۴- نتیجه گیری و پیشنهادات

- نتایج حاصل از فرآیند مدل‌سازی بیماری نسبت به عوامل ایجاد کننده ی آن، به شرح ذیل می‌باشد.
- استفاده از روش انتخاب پارامتر قدم به قدم، دارای دقت بیشتری است و معنی‌داری پارامترها در حضور یا عدم حضور یکدیگر، بهتر مورد بررسی قرار می‌گیرد. ولی وقتی تعداد پارامترهای ورودی به مدل زیاد است (مانند این تحقیق)، این کار بسیار وقت گیر و پیچیده خواهد بود. با اینکه دقت روش‌های انتخاب رو به جلو و یا حذف رو به عقب به اندازه‌ی روش قدم به قدم بالا نیست، ولی این روش‌ها سرعت بیشتری در انتخاب پارامترها دارند.
 - تمامی عوامل جوی در حضور هم معنی‌دار هستند. یعنی با حذف هر کدام از آنها، بقیه نیز از مدل خارج شده و باقیمانده‌ی مدل به شدت افزایش می‌یابد.
 - دقت مدل وابسته به دقت داده‌های ورودی به مدل است. با تغییر هر کدام از ورودی‌ها، نتایج بسته به میزان معنی‌دار بودن آن عامل در مدل، تغییر می‌کند.
 - شکل ۳ باقیمانده‌های مربوط به مدل به دست آمده از روش قدم به قدم را نشان می‌دهد. طبق بررسی‌های به عمل آمده بر روی باقیمانده‌های مدل، باقیمانده‌ها در مناطقی که آمار سرطان در آنها بسیار بالاتر و یا بسیار پایین‌تر از مناطق دیگر می‌باشد، بالاتر است. علت این امر را می‌توان برازش مدل سراسری به کل ناحیه دانست. زیرا در این حالت مدل برای بهترین برازش به داده‌ها باید از تمامی نقاط سطح عبور کند و به این ترتیب تغییرات عمده در مدل از بین می‌رود.



شکل ۳- باقیمانده‌های مربوط به مدل به دست آمده از روش قدم به قدم

- عوامل اصلی ایجاد سرطان پوست به ترتیب اهمیت‌شان در مدل عبارتند از: دما، فشار، اثر متقابل تعداد ساعات آفتابی و انرژی خورشیدی، انرژی خورشیدی، تعداد ساعات آفتابی، تعداد کشاورزان در منطقه، مصرف میوه و سبزی، تماس با سرب، مصرف کود شیمیایی پتاس، مصرف دخانیات.
- با توجه به نتایج فوق و بررسی جزئیات تحقیق، پیشنهادات و راهکارهای ذیل برای ادامه و بهبود زمینه‌ی تحقیقی فوق ارائه می‌گردد:
- عوامل زمین‌شناسی در منطقه نقش مهمی در ابتلای به بیماری‌ها دارند. چگونگی ورود این مواد به چرخه‌ی زندگی افراد و ایجاد اختلال در آن، خود عامل مهمی است که نیازمند مدل‌سازی است. به عنوان مثال بررسی اینکه چگونه رادیوم وارد چرخه‌ی غذایی افراد شده و به آنها آسیب می‌رساند، خود جای بحث فراوان و مدل‌سازی پیچیده دارد.
 - در این پروژه به دلیل کمبود داده در مورد عوامل سایر انواع سرطان‌ها، سرطان پوست انتخاب شده است. لذا بسیار ضروری می‌باشد که به تحلیل مکانی بیماری‌های شایع‌تری مثل سرطان معده و یا سرطان ریه نیز پرداخته

شود. این امر مستلزم صرف هزینه و اهتمام مسئولین در رابطه با جمع‌آوری داده‌های مختلف مربوط به عوامل این بیماری‌هاست.

منابع

- [۱]. بیگل هول آر، بونیتا آر، کجلستروم تی، ترجمه جانقربانی محسن، ژیانپور مطهره، مبانی اپیدمیولوژی، چاپ اول، انتشارات معاونت پژوهشی وزارت بهداشت و درمان آموزش پزشکی، تهران، ۱۳۷۱.
- [۲]. غضبان فریدون، زمین‌شناسی زیست-محیطی، چاپ دوم، انتشارات دانشگاه تهران، تهران، ۱۳۸۲.
- [۳]. گریم اف، بونهام کارتر، سیستم‌های اطلاعات جغرافیایی برای دانش‌پژوهان علوم زمین (مدل‌سازی به کمک GIS)، گروه اطلاعات زمین‌مرجع، سازمان زمین‌شناسی و اکتشافات معدنی کشور، ۱۳۷۹.
- [۴]. گزارش طرح تحقیقاتی بررسی پرتوزایی رادیوم ۲۶۶ در آب‌های آشامیدنی ایران، نظام ایمنی هسته‌ای کشور، امور حفاظت در برابر اشعه، انتشارات سازمان انرژی اتمی ایران، ۱۳۸۰، صفحات ۵-۳.
- [5]. Burrough P, Geographic Information Systems for Natural Resources Assessment, Oxford University Press, New York, 1986.
- [6]. German Rodriguez Princeton university's leature notes, Generalized Linear Models Theory, <http://data.princeton.ed/notes>, 2001
- [7]. Montgomery D, Peck E, Vining G, Introduction To Linear Regression Analysis, Third Edition, Wiley Interscience Publication, NewYork, 2001.
- [8]. Wen Hsiang W, Generalized Linear Models, Tunghai university Department of statistics, Tunghai, 2000.