بررسی ساختار فضایی سرطان پستان براساس مدل آمیخته خطی تعمیم‌یافته پواسون در ایران

منصور رضایی[۱]
عبدالله جلیلیان[۲]
بهزاد مهکی[۳]
مریم ویسمرادی[٤*]

۱- گروه آمار زیستی، مرکز تحقیقات توسعه اجتماعی و ارتقای سلامت، دانشگاه علوم پزشکی کرمانشاه، کرمانشاه، ایران.
۲- گروه آمار، دانشکده علوم پایه، دانشگاه رازی، کرمانشاه، ایران.
۳- گروه آمار زیستی، دانشکده بهداشت، دانشگاه علوم پزشکی کرمانشاه، کرمانشاه، ایران.
٤- گروه آمار زیستی، کمیته تحقیقات دانشجویی، دانشگاه علوم پزشکی کرمانشاه، کرمانشاه، ایران.

*نویسنده مسئول: کرمانشاه، میدان ایثار، بلوار دولت‌آباد، دانشگاه علوم پزشکی کرمانشاه، دانشکده بهداشت، کمیته تحقیقات دانشجویی.
تلفن: ۰۸۳-۳٤۲۹۰۲۷۳
E-mail: maryamveysmorady@gmail.com

چکیده

دریافت: ۱۳۹۷/۱۱/۰٦ ویرایش: ۱۳۹۷/۱۱/۱۳ پذیرش: ۱۳۹۸/۰۳/۲۱ آنلاین: ۱۳۹۸/۰۳/۳۱

زمینه و هدف: سرطان پستان یکی از شایع‌ترین بیماری‌های زنان است و بیش از سایر سرطان‌ها منجر به مرگ‌ومیر آنان می‌شود. روند رو به افزایش سرطان پستان در ایران، ضرورت انجام مطالعات گسترده در این زمینه را روشن می‌سازد. برخی مطالعات نشان داده‌اند که در کشورهای مختلف و حتی مناطق مختلف یک کشور تفاوت خطر بروز سرطان پستان وجود دارد. هدف از پژوهش کنونی تعیین ساختار فضایی بروز سرطان پستان براساس مدل‌های آماری و شناسایی مناطق دارای بروز بالای سرطان پستان در ایران بود.

روش بررسی: این مطالعه اکولوژیک از بهمن ۱۳۹٦تا تیر ۱۳۹۷ در دانشگاه علوم پزشکی کرمانشاه انجام شد. داده‌های مربوط به بیماران مبتلا به سرطان پستان تمام استان‌های ایران در فاصله زمانی سال‌های ۸۸-۱۳۸۳ بررسی شدند. عوامل مورد بررسی در این مطالعه شامل مصرف میوه و سبزیجات، فعالیت فیزیکی، اضافه‌وزن یا چاقی و شاخص توسعه انسانی بودند. برای تحلیل داده‌ها از مدل‌های آمیخته خطی تعمیم‌یافته پواسون معمولی و فضایی استفاده گردید.

یافته‌ها: در هر دو مدل، ارتباط مستقیم و معناداری بین بروز سرطان پستان و شاخص توسعه انسانی یافت شد (P<۰/۰۵). افزون‌بر شاخص توسعه انسانی، عوامل اضافه‌وزن یا چاقی در مدل‌های آمیخته خطی تعمیم‌یافته پواسون فضایی نیز به‌طور مستقیم و معناداری با بروز سرطان پستان ارتباط داشت. در مدل‌های آمیخته خطی تعمیم‌یافته پواسون فضایی با ساختار همبستگی بیز کامل، استان‌های گیلان و آذربایجان شرقی دارای بیشترین و استان کهکیلویه و بویر احمد دارای کمترین خطر بروز سرطان پستان بودند.

نتیجه‌گیری: توزیع سرطان پستان در ایران دارای ساختار فضایی است. یعنی استان‌های هم‌جوار خطرات بروز مشابهی در مورد ابتلا به این بیماری دارند.

کلمات کلیدی: سرطان پستان، توزیع پواسون، عوامل خطر.

مقدمه

با توجه به مطالعات انجام‌شده در زمینه سرطان پستان، اضافه‌وزن یا چاقی، مصرف میوه و سبزیجات، فعالیت بدنی و وضعیت اقتصادی-اجتماعی ازجمله عوامل مؤثر در بروز این بیماری است.[۱و۲] بروز سرطان پستان به‌عنوان متغیر پاسخ، از نوع شمارشی است. در حالتی که متغیر، پاسخ تعداد رخدادهای نادر مانند موارد جدید سرطان در یک جامعه در مدت زمانی خاص باشد فرض می‌شود که متغیر پاسخ دارای توزیع پواسون است.[۳] در بیشتر روش‌های آماری فرض بر این است که مشاهدات در شرایط یکسان و به‌صورت مستقل گردآوری شده‌اند. هرچند فرض استقلال کمک شایانی به تسهیل مبانی نظری می‌کند، اما در برخی موارد ممکن است این فرض ما را از واقعیت دور کند، زیرا موارد زیادی وجود دارد که مشاهدات مستقل نیستند.[٤] درمواردی‌که بین مشاهدات همبستگی وجود دارد،

[Downloaded from journals.tums.ac.ir on 2024-10-02]

فرض استقلال مشاهدات به استقلال شرطی تعدیل و همبستگی بین آن‌ها با اضافه کردن اثرات تصادفی از طریق متغیرهای پنهان به مدل در نظر گرفته می‌شود. این مدل‌ها به‌عنوان تعمیمی از مدل‌های خطی تعمیم‌یافته (Generalized linear models, GLMs) در نظر گرفته می‌شوند که شامل هر دو اثر تصادفی و ثابت هستند و مدل آمیخته خطی تعمیم‌یافته (Generalized linear mixed model, GLMM) نامیده می‌شود.[۵] اگر وابستگی مشاهدات تابعی از فاصله‌ی بین موقعیت‌های مشاهدات باشد، به‌طوری‌که مشاهدات نزدیک به هم وابسته‌تر و مشاهدات دورتر از هم وابستگی کمتری داشته باشند، این‌گونه مشاهدات داده‌های فضایی نامیده می‌شوند. یکی از انواع داده‌های فضایی، داده‌های زمین آماری است. این نوع داده‌ها در موقعیت‌های ثابت و مشخص در ناحیه‌ای پیوسته مشاهده می‌شوند. در این حالت متغیر مورد بررسی ممکن است گسسته یا پیوسته باشد.[۴] هدف از پژوهش کنونی تعیین ساختار فضایی مرتبط با بروز سرطان پستان براساس مدل‌های آماری و شناسایی مناطق دارای بروز بالای سرطان پستان در ایران بود.

روش بررسی

این مطالعه اکولوژیک از بهمن ۱۳۹۶ تا تیر ۱۳۹۷ در دانشگاه علوم پزشکی کرمانشاه انجام شد. داده‌های مورد ارزیابی در این مطالعه، به‌جای سطوح فردی در سطوح گروهی و جمعیتی (استان) اندازه‌گیری شده‌اند. جامعه آماری شامل تمام بیماران سرطان پستان ثبت‌شده‌ی استان‌های کشور (۳۰ استان) در فاصله زمانی بین سال‌های ۱۳۸۳ تا ۱۳۸۸ است. در این پژوهش، نمونه‌گیری انجام نشده است و تمامی موارد مربوط به سرطان پستان که در این دوره زمانی ثبت شده‌اند، مورد استفاده قرار گرفته است.

داده‌های مربوط به بروز سرطان پستان از گزارش‌های سالانه‌ی کشوری ثبت موارد سرطان که توسط مرکز مدیریت بیماری‌های غیرواگیر معاونت بهداشتی وزارت بهداشت، درمان و آموزش پزشکی منتشر شده، استخراج گردیده است. فایل داده‌ها از اطلاعات نظام ثبت سرطان، به‌دست آمده است. در نظام ثبت سرطان اطلاعات مربوط به بروز و مشخصات سرطان‌ها به‌طور مداوم گردآوری و ثبت می‌شود و به‌منظور تجزیه و تحلیل در اختیار پژوهشگران قرار می‌گیرد. عوامل خطر مورد استفاده در این مطالعه عبارت است از متغیر اضافه‌وزن یا چاقی، فعالیت فیزیکی، مصرف میوه و سبزیجات و شاخص توسعه انسانی (Human development index, HDI) که مربوط به سال‌های ۸۸-۱۳۸۳ و ۳۰ استان می‌باشد. متغیر اضافه‌وزن یا چاقی به‌صورت درصد افراد دارای Body mass index (BMI)>۲۵ در هر استان، متغیر فعالیت فیزیکی به‌صورت افراد دارای فعالیت فیزیکی کم‌تر از ۶۰۰ مت-دقیقه در هفته و متغیر مصرف میوه و سبزیجات به‌صورت جمع میانگین تعداد واحد میوه مصرفی روزانه با میانگین تعداد واحد سبزیجات مصرفی در هر استان مشخص شد. این عوامل خطر از گزارش‌های سالانه نظام مراقبت عوامل خطر بیماری‌های غیرواگیر وزارت بهداشت به‌دست آمده است. شاخص توسعه انسانی از گزارش‌های سالیانه بانک مرکزی استخراج شد. این شاخص از امید به زندگی، تحصیلات و تولید سرانه ناخالص ملی به‌دست می‌آید و نسبتی بین ۰ و ۱ است.[۶] در این مطالعه با توجه به این‌که متغیر پاسخ (تعداد موارد مبتلا به سرطان پستان) شمارشی است، برای مدل‌سازی آن از توزیع‌های گسسته‌ای مانند پواسون و دوجمله‌ای منفی می‌توان بهره برد. با توجه به سادگی و انعطاف‌پذیری ریاضی توزیع پواسون در بیشتر مطالعات مربوط به نقشه‌بندی بیماری‌ها از توزیع پواسون استفاده می‌شود.[۷] توزیع پواسون عضو خانواده توزیع‌های نمایی است و به‌همین دلیل می‌توان از مدل‌های آمیخته خطی تعمیم‌یافته برای مدل‌سازی تاثیر جمعیت، متغیرهای کمکی و سایر عوامل پنهان بهره برد. به‌منظور مدل‌بندی پاسخ‌های گسسته فضایی زمین آمار از مدل‌های آمیخته خطی تعمیم‌یافته فضایی استفاده می‌شود و ساختار همبستگی فضایی داده‌ها از طریق متغیرهای پنهان در نظر گرفته می‌شود. در مدل‌های آمیخته خطی تعمیم‌یافته فضایی به‌دلیل نرمال نبودن پاسخ و وجود تعداد زیادی متغیرهای پنهان، توزیع‌های پسین شکل بسته‌ای ندارد. برای حل این مشکل یک روش تحلیل بیزی تقریبی به‌کار گرفته می‌شود که بر اساس تقریب لاپلاس آشیانه‌ای جمع بسته (Integrated nested laplace approximation, INLA) است.[۸] برازش مدل توسط R Project for Statistical Computing, Version 3.4.3 (https://cran.r-project.org) و با استفاده از پکیج INLA انجام شده است. در این مدل‌ها فرض می‌کنیم $x_{it}^{(1)},...,x_{it}^{(4)}$ بیان‌گر مقدار متغیرهای کمی (چاقی، میزان مصرف میوه و سبزیجات، میزان فعالیت فیزیکی، HDI) و p_{it} جمعیت استان، برای استان i ام در سال t ام باشند.

[Downloaded from journals.tums.ac.ir on 2024-10-02]

متغیر پاسخ دارای توزیع پواسون با میانگین λ_{it} است. باید توجه داشت که λ_{it} نرخ بروز سرطان پستان در استان i ام در سال t ام است. با توجه به تأثیر مستقیم جمعیت بر نرخ بروز سرطان از رابطه‌ی

$$\lambda_{it} = p_{it}exp\left(\beta_0 + \beta_1 x_{it}^{(1)} + \cdots + \beta_q x_{it}^{(q)} + \alpha t + z\gamma\right)$$

به‌منظور مدل‌سازی تأثیر متغیرهای کمکی بر نرخ بروز سرطان پستان استفاده شد. ضرایب $\beta_0, \beta_1, \ldots, \beta_q$ اثرهای ثابت عرض از مبدأ و متغیرهای کمکی، α بیان‌گر روند زمانی و γ بیان‌گر اثر تصادفی استان‌ها هستند.[۹]

برای لحاظ کردن وابستگی‌های فضایی و زمانی می‌توان از یک مدل GLMM به شکل زیر بهره گرفت.

$$\lambda_{it} = p_{it}exp\left(\beta_0 + \beta_1 x_{it}^{(1)} + \cdots + \beta_q x_{it}^{(q)} + \alpha t + \eta_{it}\right) \qquad \lambda_{it} = p_{it}exp(\theta_{it})$$

در رابطه بالا، η_{it} اثر تصادفی استان i ام در سال t ام است که بیان‌گر تأثیر عوامل پنهان مشاهده‌نشده‌ای است که منجر به وابستگی‌های فضایی و زمانی در داده‌ها می‌شوند. به‌طور معمول فرض می‌شود که η_{it} یک میدان تصادفی گوسی غیر قابل مشاهده (پنهان) است.[۸] در مؤلفه‌های این مدل از بردارهای تصادفی دارای توزیع نرمال چندمتغیری با بردار میانگین صفر و ماتریس واریانس کوواریانس Σ استفاده شده است. ماتریس واریانس کوواریانس Σ همبستگی فضایی و زمانی را مشخص می‌کند و در حالت کلی می‌تواند هر ماتریس معین نامنفی باشد.[۱۰] یکی از فرض‌هایی که منجر به ساده شدن ساختار ماتریس واریانس کوواریانس Σ می‌شود آن است که همبستگی‌های فضایی و زمانی تفکیک‌پذیر هستند. به‌عبارت‌دیگر فرض می‌شود $\Sigma=\Sigma_{spat}+\Sigma_{temp}$ که در Σ_{spat} و Σ_{temp} به‌ترتیب ماتریس‌های واریانس و کوواریانس فضایی و زمانی هستند.[۱۱] با توجه به این‌که داده‌های موجود تنها یک دوره زمانی ٦ ساله را پوشش می‌دهند و این دوره زمانی برای بررسی وابستگی‌های زمانی در داده‌ها کافی نیست، در این مطالعه از بررسی وابستگی‌های زمانی صرف‌نظر کردیم و بنابراین ماتریس Σ_{temp} برابر با صفر در نظر گرفته می‌شود. برای ساختار همبستگی فضایی Σ_{spat} به چند طریق می‌توان عمل کرد که در زیر به آن‌ها اشاره می‌شود.[۱۱] در ساختار همبستگی فضایی Besag فرض می‌شود که متغیر اثر تصادفی برای هر سال و استان، دارای توزیع نرمال با میانگین متوسط اثر تصادفی استان‌های هم‌جوار و واریانس محدود است. ماتریس همبستگی‌های فضایی بر اساس توزیع‌های شرطی مشخص می‌شود.[۱۲] در مدل ساختار همبستگی فضایی بیز کامل (Besag York Molie, BYM) فرض بر این است که تعداد موارد بیماری در هر ناحیه دوبه‌دو مستقل بوده و از توزیع پواسون پیروی می‌کند. برای نرخ‌های خطر، اثرات تصادفی مناطق خاص به دو جز تقسیم می‌شود. یک جز اثراتی که بین مناطق به‌صورت ساختاری تغییر می‌کند (ساختاریافته فضایی) و یک جز که اثرات به‌صورت غیرساختاری بین مناطق تغییر می‌کند (نوسانات یا خطای تصادفی). ماتریس واریانس کوواریانس ساختار همبستگی فضایی Besag بیان‌شده در بالا، با اضافه شدن ماتریس واریانس بخش ساختاریافته فضایی (σ_b^2) به خطای تصادفی (σ_h^2)، ساختار همبستگی BYM را می‌سازد.[۱۳]

یافته‌ها

تعداد کل موارد سرطان پستان ثبت‌شده در طی سال‌های ۸۸-۱۳۸۳ در ایران ۳۵۹٦۸ مورد بوده است. بیشترین مورد مشاهده‌شده سرطان پستان در بین استان‌ها مربوط به استان‌های تهران (۹۹۲۵ مورد) و اصفهان (۳۱۷۱ مورد) است. استان کهگیلویه و بویراحمد با ۱۱۳ مورد دارای کمترین تعداد موارد سرطان پستان است. نتایج حاصل از برازش مدل آمیخته خطی تعمیم‌یافته پواسون نشان داد که تنها متغیر شاخص توسعه انسانی معنادار شد (جدول ۱).

در مرحله بعد دو متغیر سال بروز و استان را به‌عنوان اثر تصادفی به مدل اضافه شد (جدول ۲).

جدول ۱: نتایج حاصل از برازش مدل آمیخته خطی تعمیم‌یافته پواسون و برآورد پارامترهای مدل

متغیر	برآورد ضرایب	سطح معناداری
عرض از مبدأ*	-۱٤۱/۳۷۹	۰/۰۰۵
اضافه‌وزن یا چاقی	۰/۰۱۱	۰/۱۱۸
فعالیت فیزیکی	۰/۰۰۳	۰/۵۲٦
مصرف میوه و سبزیجات	-۰/۰۲۹	۰/۷۲۷
شاخص توسعه انسانی*	۹/۰۸٦	۰/۰۰۱

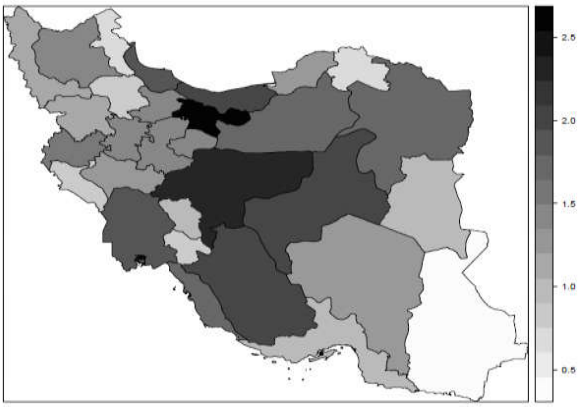
* برازش مدل آمیخته خطی تعمیم‌یافته پواسون در سطح معناداری ۰/۰۵.

[Downloaded from journals.tums.ac.ir on 2024-10-02]

جدول ۲: برآورد، انحراف‌معیار، سطح معناداری و فاصله اطمینان متغیر سال و استان به‌عنوان اثر تصادفی

				فاصله اطمینان ۹۵٪	
اثر باقی‌مانده‌ها	برآورد	انحراف‌معیار	سطح معناداری	حد پایین	حد بالا
سال	۰/۰۱۱	۰/۰۱۸	۰/۵۳۰	۰/۰۰۰	۰/۲۵۵
استان	۰/۰۱۸	۰/۰۲۰	۰/۳۵۱	۰/۰۰۲	۰/۱۵۰

برازش مدل آمیخته خطی تعمیم‌یافته پواسون با اضافه کردن اثرات تصادفی در سطح معناداری ۰/۰۵.



نقشه ۱: نقشه میانگین نرخ بروز سرطان پستان در هر استان

نتایج نشان داد که دو متغیر سال و استان به‌عنوان اثر تصادفی معناداری نشدند. میانگین نرخ بروز سرطان پستان در زنان را در استان‌های کشور روی نقشه مشخص شد (نقشه ۱).

نتایج نشان داد که استان‌های هم‌جوار از هم تأثیر پذیرفته‌اند و از هم متفاوت نیستند، پس ممکن است بین آن‌ها همبستگی فضایی وجود داشته باشد، این موضوع انگیزه‌ای برای برازش مدل آمیخته خطی تعمیم‌یافته پواسون فضایی فراهم کرد.

در چارچوب مدل‌سازی بیزی برای اجرای مدل آمیخته خطی تعمیم‌یافته پواسون فضایی با ساختار همبستگی فضایی BYM، همه پارامترها و اثرات تصادفی به‌عنوان مقادیر نامعلوم در نظر گرفته می‌شوند و باید به آن‌ها توزیع پیشین اختصاص دهیم. ازآنجایی‌که هیچ ایده و دانش پیشین در مورد تحلیل پارامترها نداشتیم از توزیع‌های ناآگاهی بخش به‌عنوان توزیع‌های پیشین استفاده کردیم. در این مطالعه فرض می‌شود پارامترهای $\alpha, \beta_1, \ldots, \beta_4, \beta_6$ دارای توزیع پیشین یکنواخت ناسره و $-log\sigma_v^2$ و $-log\sigma_u^2$ دارای توزیع پیشین لگ گاما با پارامتر شکل ۱ و پارامتر مقیاس ۰/۰۰۰۴ هستند و براساس این پیشین‌ها توزیع‌های پسین را به‌دست می‌آوریم. برای برآورد هر یک از پارامترها توزیع پسین را محاسبه کنیم (جدول ۲).

به‌منظور بررسی معناداری پارامترها از فاصله باورمندی ۹۵٪ بیزی (جدول ۳) استفاده می‌کنیم. با توجه به این فواصل، متغیرهای چاقی و شاخص توسعه انسانی دارای اثرات مثبت معنادار هستند (فاصله باورمندی صفر را شامل نمی‌شود). متغیرهای فعالیت فیزیکی و مصرف میوه و سبزیجات تأثیر معناداری نداشتند و متغیر α که بیانگر روند زمانی است، دارای اثر معنادار مثبت است. هر دو جز ساختاریافته فضایی و نوسانات تصادفی نیز دارای اثرات مثبت معنادار هستند.

نمودار $\exp(\gamma_{it})$ که خطر تعدیل‌شده نسبت به جمعیت و متغیرهای کمکی را نشان می‌دهد بر روی نقشه رسم شد (نقشه ۲). براساس این نقشه استان‌های گیلان و آذربایجان شرقی بیشترین خطر بروز سرطان پستان و پس از آن استان‌های کردستان، سمنان، خراسان جنوبی و کرمان قرار دارند و استان کهکیلویه و بویراحمد، کمترین میزان خطر بروز را دارد.

سؤالی که در اینجا مطرح می‌شود این است که آیا استان‌هایی که در نقشه ۲ دارای بالاترین خطر تعدیل‌شده بروز سرطان پستان بودند، با میانگین کل کشور نیز دارای اختلاف معناداری هستند یا خیر؟ مشاهده می‌شود که در استان‌های گیلان و آذربایجان شرقی خطر نسبی بروز بیماری نسبت به میانگین کل کشور به‌طور معناداری بیشتر است (جدول ٤). با توجه به بازه‌های باورمندی ۹۵٪ بیزی مقدار خطر نسبی برای این دو استان بزرگ‌تر از ۱ است، بنابراین خطر نسبی این دو استان نسبت به میانگین کل کشور بیشتر است.

[Downloaded from journals.tums.ac.ir on 2024-10-02]

جدول ۳: برآورد پارامترهای مدل بیز کامل

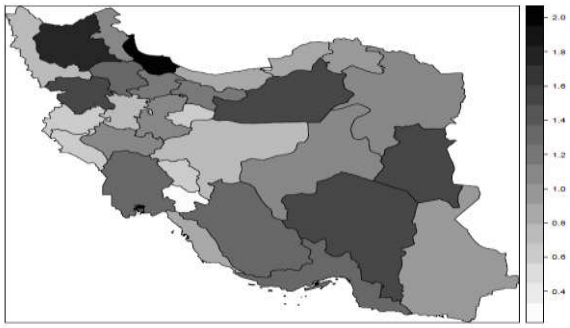
متغیر			فاصله باورمندی ۹۵٪ بیزی	
	میانگین توزیع پسین	انحراف معیار توزیع پسین	چندک ۲/۵٪ توزیع پسین	چندک ۹۷/۵٪ توزیع پسین
عرض از مبدأ	−۱۹۶/۰۵۵۸	۹/۶۵۱۴	−۲۱۵/۰۱۱۵	−۱۷۷/۱۲۸۵
چاقی یا اضافه‌وزن	۰/۰۰۱۱	۰/۰۰۰۶	۰/۰۰۰۰	۰/۰۰۲۳
فعالیت فیزیکی	۰/۰۰۰۲	۰/۰۰۱۰	−۰/۰۰۱۸	۰/۰۰۲۳
میوه و سبزیجات	−۰/۰۰۰۸	۰/۰۱۲۱	−۰/۰۲۴۵	۰/۰۲۳۰
شاخص توسعه انسانی	۰/۱۳۹۱	۰/۰۱۶۰	۰/۱۰۷۸	۰/۱۷۰۴
روند زمانی	۰/۱۴۱۴	۰/۰۰۷۰	۰/۱۲۷۷	۰/۱۵۵۰
نوسانات تصادفی	۰/۱۶۴۴	۰/۰۴۳۹	۰/۰۹۶۸	۰/۲۶۷۹
جز ساختاریافته فضایی	۰/۰۰۱۸	۰/۰۰۳۸	۰/۰۰۰۰	۰/۰۰۹۵

آزمون فاصله باورمندی ۹۵٪ بیزی

جدول ۴: میانگین و بازه‌ی باورمندی پسین برای خطر نسبی تعدیل شده نسبت به جمعیت و متغیرهای کمکی

استان	میانگین پسین	فاصله باورمندی ۹۵٪ بیزی	
		چندک ۲/۵٪ توزیع پسین	چندک ۹۷/۵٪ توزیع پسین
آذربایجان شرقی	۱/۰۹	۱/۰۴	۱/۱۴
گیلان	۱/۱۷	۱/۱۳	۱/۲۲

آزمون فاصله باورمندی ۹۵٪ بیزی



نقشه ۲: نقشه برآورد پسین $\exp(\eta_{it})$

برای بررسی سهم مربوط به بخش خطای تصادفی و بخش ساختاریافته فضایی در ماتریس واریانس کوواریانس Σ_{spat} میانگین و چندک‌های ۲/۵٪ و ۹۷/۵٪ پسین برای کمیت $\frac{\sigma_s^2}{\sigma_h^2+\sigma_s^2}$ محاسبه و به‌ترتیب برابر با ۰/۰۱۱، ۰/۰۰۰۹ و ۰/۰۵۹ می‌باشند. بنابراین می‌توان نتیجه گرفت که کمتر از ٦٪ تغییرات (واریانس) اثر تصادفی η_{it}، به‌واسطه‌ی بخش خطای تصادفی است و سایر تغییرات (واریانس) یعنی ۹۴٪، از بخش ساختاریافته فضایی حاصل می‌شود. پس همبستگی فضایی نقش قابل‌توجهی در تعیین اثر تصادفی η_{it} دارد.

بحث

در مدل آمیخته خطی تعمیم‌یافته پواسون رابطه معنادار و مستقیم شاخص توسعه انسانی بر بروز سرطان پستان تایید می‌شود، به‌عبارت‌دیگر بروز سرطان پستان در استان‌های با شاخص توسعه انسانی بالاتر، بیشتر است. این نتیجه با نتایج مطالعه Ghoncheh و

[Downloaded from journals.tums.ac.ir on 2024-10-02]

همکاران همخوانی داشت.[۱۴] Hajian و همکاران در مطالعه‌ای به این نتیجه رسیدند که اضافه‌وزن، چاقی و چاقی مرکزی با شانس ابتلا به سرطان پستان همراه است درحالی‌که در پژوهش کنونی و در مدل آمیخته خطی تعمیم‌یافته پواسون به‌دلیل تفاوت در جامعه هدف، ارتباط معناداری بین چاقی و بروز سرطان پستان یافت نشد.[۲] در مطالعه Saki و همکاران که با هدف بررسی عوامل خطر سرطان پستان با استفاده از تحلیل مدل‌های درختی انجام گرفت، رابطه معناداری بین عدم فعالیت فیزیکی و خطر بروز سرطان پستان یافت شد اما به‌دلیل تفاوت در جامعه هدف و نیز متفاوت بودن مدل‌های مورد استفاده، این ارتباط در پژوهش کنونی یافت نشد.[۱] در مطالعه کنونی ارتباط معناداری بین مصرف میوه و سبزیجات و بروز سرطان پستان مشاهده نشد درحالی‌که در مطالعه Abbastabar و همکاران ارتباط معناداری بین سرطان پستان و مصرف میوه و سبزیجات مشاهده شد.[۱۵]

در این مطالعه ابتدا وجود اثرات تصادفی در داده‌ها مشاهده شد و سپس با توجه به نتایج، معناداری ساختار همبستگی فضایی برای داده‌ها تایید شد. مؤلفه ساختار همبستگی فضایی نسبت به مؤلفه نوسانات تصادفی نقش بیشتری داشت. به‌این‌معنی که استان‌های هم‌جوار دارای خطرات نسبی بروز مشابه در بیماری‌ها هستند. بیشتر مطالعات انجام‌شده در خصوص سرطان نیز اهمیت توجه به توزیع جغرافیایی این بیماری را تایید کرده‌اند.[۱۶-۱۸] نقشه مربوط به میانگین نرخ بروز سرطان پستان در این مطالعه با نقشه مطالعه Jafari-Koshki که مربوط به خطر نسبی سرطان پستان است، شباهت بسیاری دارد.[۱۸] در این نقشه استان‌های تهران، اصفهان و یزد بیشترین خطر نسبی را دارا هستند.

در مدل آمیخته خطی تعمیم‌یافته پواسون فضایی ارتباط مستقیم معناداری بین شاخص توسعه انسانی با خطر بروز سرطان پستان یافت شد که با نتایج مطالعه Khoshkar و همکاران همخوانی داشت.[۱۸] این موضوع می‌تواند به دلایلی مانند آلودگی هوا و تماس بیشتر با عوامل سرطان‌زا در استان‌هایی با شاخص توسعه انسانی بالاتر باشد، همچنین ممکن است به‌دلیل تشخیص بیشتر موارد سرطان و ثبت دقیق‌تر این موارد در شهرهایی با وضعیت اقتصادی-اجتماعی بهتر رخ دهد.

در مطالعه Khoshkar در مدل BYM با تعدیل عوامل خطر، استان‌های خراسان رضوی، لرستان و همدان دارای بالاترین و استان‌های اردبیل و کهگیلویه و بویراحمد دارای کمترین خطر نسبی می‌باشند و در کل استان‌های جنوب غربی دارای خطر بروز کمتری می‌باشند درحالی‌که در پژوهش کنونی استان‌های گیلان و آذربایجان شرقی بیشترین خطر نسبی بروز سرطان پستان و پس از آن‌ها استان‌های کردستان، سمنان، خراسان جنوبی و کرمان بالاترین خطر نسبی بروز را دارند. استان کهکیلویه و بویراحمد، کمترین میزان خطر نسبی بروز را دارد. اختلاف در نتایج ممکن است به‌دلیل متفاوت بودن تقریب‌های بیزی مورد استفاده برای به‌دست آوردن توزیع‌های پسین در این مطالعات باشد. در پژوهش کنونی ارتباط مستقیم و معناداری بین چاقی یا اضافه وزن و سرطان پستان مشاهده شد اما در مطالعه Khoshkar این ارتباط مشاهده نشد.[۱۸]

توزیع سرطان پستان در ایران دارای ساختار فضایی است. یعنی استان‌های هم‌جوار خطرات بروز مشابهی در مورد ابتلا به این بیماری دارند.

سپاسگزاری: با تشکر از معاونت پژوهشی دانشگاه علوم پزشکی کرمانشاه که هزینه اجرای این طرح را پذیرفتند. این مقاله قسمتی از پایان‌نامه تحت عنوان "بررسی عوامل مرتبط با سرطان پستان در ایران با استفاده از مدل‌های آمیخته خطی تعمیم‌یافته پواسون و پواسون فضایی" دوره کارشناسی ارشد در دانشگاه علوم پزشکی کرمانشاه با کد ۳۰۰۵۲۳۱ در سال ۱۳۹۷ می‌باشد.

References

1. Saki A, Hajizadeh E, Tehranian N. Evaluating the risk factors of breast cancer using the analysis of tree models. *Horizon Med Sci* 2011;17(1):60-8.
2. Hajian K, Gholizadehpasha A, Bozorgzadeh S. Association of obesity and central obesity with breast cancer risk in pre and postmenopausal women. *J Babol Univ Med Sci* 2013;15(3):7-15.
3. Agresti A. An Introduction to Categorical Data Analysis. 3rd ed. Newark, NJ : John Wiley and Sons; 2018.
4. Mohammadzade M. Spatial Statistic and Its Applications. Tehran, Iran: Tarbiat Modares University Press; 2015.
5. Breslow NE, Clayton DG. Approximate inference in generalized linear mixed models. *J Am Stat Assoc* 1993;88(421):9-25.
6. Akbari M, Abachizadeh K, Khayamzadeh M. Iran Cancer Report. Cancer Research Center Shahid Beheshti University Of Medical Sciences. Qom, Iran: Darolfekr; 2008.
7. Baum C. Models for Count Data and Categorical Response Data. Berlin, Germany: Boston College and DIW; 2010.
8. Rue H, Martino S, Chopin N. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *J Royal Stat Soc Series B (Stat Methodol)* 2009;71(2):319-92.

[Downloaded from journals.tums.ac.ir on 2024-10-02]

9. Demidenko E. Mixed Models: Theory and Applications with R. New York, NY: John Wiley and Sons Inc.; 2013.
10. Chung MK. Introduction to random fields [theory on the Internet]. Wisconsin, WI: University of Wisconsin–Madison, Department of Statistics; 2007 [cited 2019 May 15]. Available from: http://pages.stat.wisc.edu/~mchung/teaching/MIA/theories/randomfield.feb.02.2007.pdf
11. Riebler A, Sørbye SH, Simpson D, Rue H. An intuitive Bayesian spatial model for disease mapping that accounts for scaling. *Stat Methods Med Res* 2016;25(4):1145-65.
12. Rue H, Held L. Gaussian Markov Random Fields: Theory and Applications. Boca Raton, FL: Chapman and Hall/CRC; 2005.
13. Besag J, York J, Mollie A. Bayesian image restoration with two applications in spatial statistics (with discussion). *Ann Inst Stat Math* 1991;43: 1–59.
14. Ghoncheh M, Mirzaei M, Salehiniya H. Incidence and Mortality of Breast Cancer and their Relationship with the Human Development Index (HDI) in the World in 2012. *Asian Pac J Cancer Prev* 2015;16(18):8439-43.
15. Abbastabar H1, Hamidifard P, Roustazadeh A, Mousavi SH, Mohseni S, Sepandi M, et al. Relationships between breast cancer and common non- communicable disease risk factors: an ecological study. *Asian Pac J Cancer Prev* 2013;14(9):5123-5.
16. Mahdavifar N, Pakzad R, Ghoncheh M, Pakzad I, Moudi A, Salehiniya H. Spatial analysis of breast cancer incidence in Iran. *Asian Pac J Cancer Prev* 2016;17(S3):59-64.
17. Jafari-Koshki T, Schmid VJ, Mahaki B. Trends of breast cancer incidence in Iran during 2004-2008: a Bayesian space-time model. *Asian Pac J Cancer Prev* 2014;15(4):1557-61.
18. Khoshkar AH, Koshki TJ, Mahaki B. Comparison of Bayesian spatial ecological regression models for investigating the incidence of breast cancer in Iran, 2005- 2008. *Asian Pac J Cancer Prev*;16(14):5669-73.

[Downloaded from journals.tums.ac.ir on 2024-10-02]

TUMJ Tehran University Medical Journal, June 2019; Vol. 77, No. 3: 152-159 Original Article

Spatial structure of breast cancer using Poisson generalized linear mixed model in Iran

Mansour Rezaei Ph.D.[1]
Abdullah Jalilian Ph.D.[2]
Behzad Mahaki Ph.D.[3]
Maryam Veismoradi M.Sc.[4*]

1- Department of Biostatistics, Social Development and Health Promotion Research Center, Kermanshah University of Medical Sciences (KUMS), Kermanshah, Iran.
2- Department of Statistics, Basic Sciences College, Razi University, Kermanshah, Iran.
3- Department of Biostatistics, School of Public Health, Kermanshah University of Medical Sciences (KUMS), Kermanshah, Iran.
4- Department of Biostatistics, Students Research Committee, Kermanshah University of Medical Sciences (KUMS), Kermanshah, Iran.

[*] Corresponding author: Student Research Committee, School of Public Health, Kermanshah University of Medical Sciences, Dawlatabad Blvd., Isar Sq., Kermanshah, Iran.
Tel: +98 83 34290273
E-mail: maryamveysmorady@gmail.com

Abstract

Received: 26 Jan. 2019 Revised: 02 Feb. 2019 Accepted: 11 Jun. 2019 Available online: 21 Jun. 2019

Background: Breast cancer is one of the most common diseases in women and causes more deaths rather than other cancers. The increasing trend of breast cancer in Iran makes clear the need of extensive breast cancer research in this area. Some studies showed that in the variety countries and even in the different areas in one country has different risk of breast cancer incidence and this is a reason that there is a correlation between region of life and risk of breast cancer. The purpose of this study was to determine the spatial structure associated with the incidence of breast cancer based on statistical models and identification of areas with high incidence of breast cancer in Iran.

Methods: This ecological study was conducted in Kermanshah University of Medical Sciences, Iran, from February to July 2018. Data on breast cancer patients in all provinces of Iran (30 provinces) were investigated since 2004 to 2009. Risk factors in this study included fruit and vegetable consumption, physical activity, overweight or obesity, and human development index. In this study, we have used routine and spatial Poisson's generalized linear mixed models for data analysis.

Results: In both routine and spatial models, direct and significant correlation was found between the incidence of breast cancer and the human development index ($P<0.05$). In addition to human development index, overweight or obesity factors were also had direct and significant relationship to the incidence of breast cancer in the spatial Poisson's generalized linear mixed model ($P<0.05$). In the spatial Poisson's generalized linear mixed model with correlation structure of Besag Yorg Molie (BYM), two provinces of Gilan and East Azerbaijan had the highest risk of breast cancer incidence and province of Kohgiluyeh and Boyer Ahmad had the lowest risk of breast cancer incidence.

Conclusion: The results showed that the distribution of breast cancer incidence in Iran has a spatial structure. That is, the adjacent provinces have similar incidences of this disease.

Keywords: breast cancer, poisson distribution, risk factors.

[Downloaded from journals.tums.ac.ir on 2024-10-02]