مجله تخصصی اپیدمیولوژی ایران؛۱۳۹۱، دوره ۸، شماره۳: صفحات ۲۱-۱۲.

مقاله پژوهشی

روندهای فصلی و الگوهای قابل توجیه در داده‌های کشوری نظام مراقبت بیماری سرخک: رویکردهای شناسایی و حذف

منوچهر کرمی[۱]، حمید سوری[۲]، یداله محرابی[۳]، علی اکبر حقدوست[۴]، محمد مهدی گویا[۵]

[۱] استادیار اپیدمیولوژی، گروه آمارزیستی و اپیدمیولوژی، دانشکده بهداشت، دانشگاه علوم پزشکی همدان، ایران

[۲] استاد اپیدمیولوژی، مرکزتحقیقات ارتقای ایمنی و پیشگیری از مصدومیت‌ها و گروه اپیدمیولوژی، دانشگاه علوم پزشکی شهید بهشتی، ایران

[۳] استاد آمار زیستی، گروه اپیدمیولوژی، دانشکده بهداشت، دانشگاه علوم پزشکی شهید بهشتی، ایران

[۴] دانشیار اپیدمیولوژی، مرکز تحقیقات مدلسازی سلامت و گروه اپیدمیولوژی و آمار زیستی، دانشکده بهداشت، دانشگاه علوم پزشکی کرمان، ایران

[۵] استاد بیماری‌های عفونی، مرکز مدیریت بیماری‌های واگیر، وزارت بهداشت، درمان و آموزش پزشکی، ایران

نویسنده رابط: حمید سوری، گروه اپیدمیولوژی، دانشکده بهداشت، دانشگاه علوم پزشکی شهید بهشتی، تهران، ایران. تلفن: ۲۲۴۳۱۹۹۳، پست الکترونیک: hsoori@yahoo.com

تاریخ دریافت:۹۰/۹/۵؛ پذیرش: ۹۱/۳/۶

مقدمه و هدف: شناسایی و حذف الگوهای قابل توجیه، شامل روندهای ماهانه، فصلی و سالانه، آثار روزهای هفته، روزهای تعطیل و پایان هفته به دلیل اختلال در اعلام هشدارهای واقعی مبنی بر وجود تغییرات در روند بیماری‌ها و نیز ایجاد هشدارهای کاذب بسیار اهمیت دارد. مطالعه حاضر با هدف شناسایی و حذف الگوهای قابل توجیه داده‌های کشوری نظام مراقبت بیماری سرخک در ایران انجام گردید.

روش کار: برای شناسایی الگوهای قابل توجیه از داده‌های کشوری، موارد مشکوک به سرخک طی سال‌های ۱۳۸۷ تا۱۳۸۹، از نمودارهای خطی و میانگین متحرک، توابع خودهمبستگی و خودهمبستگی جزئی و شاخص متوسط پایان هفته استفاده شد. برای حذف الگوهای قابل توجیه نیز از روش‌های هموارسازی میانگین‌های متحرک و هموارسازی نمایی هولت- وینترز استفاده شد.

نتایج: نتایج استفاده از روش‌های مختلف برای شناسایی الگوهای قابل توجیه بیانگر وجود الگوهای سالانه، ماهانه و به ویژه آثار روزهای هفته و تعطیل در داده‌های بیماری سرخک است. همچنین، یافته‌ها نشانگر کارایی روش هموارسازی میانگین‌های متحرک در کاهش انحراف از پیش فرض نرمال‌بودن و عملکرد مناسب روش هموارسازی نمایی هولت- وینترز در حذف الگوهای فصلی، شامل آثار روزهای هفته، ماه و سال است.

نتیجه‌گیری: یافته‌های مطالعه حاضر بر اهمیت بررسی داده‌های خام نظام مراقبت بیماری سرخک، پیش از به کارگیری روش‌های کشف طغیان تأکید می‌کند و استفاده از روش هموارسازی هولت وینترز برای حذف الگوهای فصلی را، با وجود سادگی و کارایی مناسب روش میانگین‌های متحرک در کاهش انحراف از پیش‌فرض نرمال‌بودن از داده‌های خام نظام مراقبت سرخک توصیه می‌کند.

واژگان کلیدی: روند فصلی، طغیان، هموارسازی، سرخک، نظام مراقبت سندرومیک

مقدمه

نظام‌های مراقبت، افزون بر استفاده از داده‌های موارد ابتلا به بیماری که دارای تأیید تشخیص آزمایشگاهی هستند، از دیگر منابع داده‌های بالینی و غیر بالینی مانند موارد مشکوک به یک بیماری تظاهرات بالینی خاص نیز، با هدف تشخیص زود هنگام طغیان‌ها و مراقبت از بیماری‌ها، استفاده می‌کنند (۱). به طور معمول، پایش داده‌های مربوط به یک سندروم خاص یاموارد‌مشکوک به یک بیماری خاص با الگوریتم‌ها یا روش‌های مختلف کشف طغیان مانند الگوریتم‌های مجموع تراکمی (Cumulative Sums= CuSums) و میانگین متحرک نمایی (Exponentially Weighted Moving Average= EWMA) صورت می‌گیرد (۲). چنین داده‌هایی در طول زمان دو الگوی کلی از خود نشان می‌دهند. دسته اول، الگوهای قابل توجیه (Explainable Patterns) هستند که خود شامل روندهای فصلی، اثرات روزهای هفته، اثرات روزهای تعطیل، اثرات روزهای پایان هفته و برخی اثرات خاص دیگر مربوط به عوامل خارجی می‌باشند (۲). دسته دوم الگوهای غیرقابل توجیه (Unexplainable Patterns) هستند که شامل هرگونه تغییر در روند یعنی افزایش یا کاهش در مقدار واقعی داده‌ها در طول زمان به دلایلی غیر از الگوهای قابل توجیه می‌باشد (۳).

[Downloaded from irje.tums.ac.ir on 2024-08-02]

به کارگیری الگوریتم‌ها به ویژه مدل‌های زمانی مبتنی بر چارت، از جمله الگوریتم CuSum روی داده‌های خام، به ایجاد هشدارهای کاذب (False Alarms) منجر می‌گردد. مهم‌ترین علت ایجاد این هشدارها، وجود الگوهای قابل توجیه در داده‌ها، مانند آثار روزهای هفته، تعطیلی‌ها و روندهای فصلی هستند. برای حذف الگوهای قابل توجیه یاد شده از داده‌ها، روش‌های پیش پردازش مختلفی استفاده می‌شوند که در دو دسته مبتنی بر مدل‌ها (Model based methods) و مبتنی بر داده‌ها (Data driven methods) قرار می‌گیرند (۴). روش‌های هموارسازی نمایی هولت- وینترز و میانگین‌های متحرک جزو مدل‌های مبتنی بر داده ها هستند.

در حال حاضر، نظام مراقبت بیماری سرخک در کشور ایران با هدف حذف این بیماری برقرار است. این نظام همه موارد مشکوک به سرخک، یعنی افراد دارای علایم بالینیِ تب، بثورات جلدی ماکولوپاپولر و یکی از علایم سرفه یا آبریزش بینی یا قرمزی ملتحمه را،به صورت روزانه به سطوح بالاتر نظام گزارش می‌کند (۵). از آنجا که کشف به هنگام طغیان یا هرگونه تغییر در روند بیماری مهم‌ترین هدف نظام‌های مراقبت است، نظام مراقبت از بیماری سرخک، همان‌گونه که در بالا گفته شد، از رویکرد سندرومیک برای شناسایی طغیان‌ها استفاده می‌کند و برای پاسخ به هنگام به طغیان‌های احتمالی منتظر تأیید آزمایشگاهی موارد مشکوک به سرخک نمی‌ماند. با این حال، تصمیم‌گیری براساس داده‌های گزارش شده روزانه موارد مشکوک به سرخک به صورت خام، همانند دیگر منابع داده‌ای مورد استفاده در نظام‌های مراقبت سندرومیک که دارای روندها و الگوهای خاصی است، ممکن است به هدر رفتن منابع منجر شود. برای مثال، کاهش تعداد موارد گزارش شده مشکوک به سرخک در اواخر یک هفته یا طی روزهای تعطیل و متعاقب افزایش در تعداد موارد گزارش شده، ممکن است دست اندرکاران نظام مراقبت را به پاسخ به این هشدار کاذب وادار کند. از طرفی، ابزارهای نظام مراقبت به ویژه نظام‌های مراقبت سندرومیک، برای شناسایی به‌هنگام طغیان‌های احتمالی و یا هرگونه تغییر در روند بیماری‌ها، دارای پیش‌فرض‌هایی هستند که وجود چنین روندها و الگوهایی قابل توجیه در داده‌های استفاده شده برای آن‌ها، به انحراف از این پیش فرض‌ها منجر می‌شوند. برای مثال، در صورتی که داده‌های خام مورد استفاده برای به‌کارگیری الگوریتم CuSum یا دیگر نمودارهای کنترل بدون حذف الگوهای قابل توجیه با هدف کشف به هنگام طغیان‌های احتمالی در یک نظام مراقبت به کار رود، به دلیل رعایت نشدن پیش فرض‌های مربوط، از جمله نرمال بودن، نه تنها مانع شناسایی به هنگام طغیان‌ها و تغییر در روند بیماری‌ها می‌شوند، بلکه به ایجاد هشدارهای کاذب و در نتیجه، هدر رفتن منابع نظام مراقبت منجر می‌گردند (۳-۲). از این رو، آشکاراست که شناسایی و حذف الگوهای قابل توجیه در داده‌های مورد استفاده در نظام‌های مراقبت،به دلایل یاد شده، یعنی اختلال در اعلام هشدارهای واقعی مبنی بر وجود تغییرات واقعی در روند بیماری‌ها یا همان الگوهای غیر قابل توجیه و نیز، ایجاد هشدارهای کاذب بسیارمهم است. از طرفی عملکرد نامناسب یک نظام مراقبت، به دلیل استفاده از روش‌های کشف طغیان بدون لحاظ کردن الگوهای توجیه‌پذیر و رعایت پیش فرض‌های مربوط به هدر رفتن منابع انسانی و مالی منجر خواهد شد. بنابراین، مطالعه حاضر با هدف شناسایی الگوهای قابل توجیه در داده‌های کشوری نظام مراقبت بیماری سرخک در ایران، طی یک دوره سه ساله و نیز حذف این الگوها، با استفاده از دو روش هموارسازی نمایی هولت- وینترز و میانگین‌های متحرک انجام شده است.

روش کار

۱) داده‌های استفاده شده

با توجه به اهداف مطالعه حاضر، از داده‌های کشوری موارد مشکوک به سرخک طی سال‌های ۱۳۸۷ تا ۱۳۸۹ برای شناسایی الگوهای قابل توجیه، شامل روندهای ماهانه، فصلی و سالانه، آثار روزهای هفته، روزهای تعطیل و پایان هفته استفاده شد. این داده‌ها شامل ۴۸۳۴ مورد از افراد دارای علایم بالینی تب، بثورات جلدی ماکولوپاپولر و یکی از علایم سرفه یا آبریزش بینی یا قرمزی ملتحمه هستند که به صورت روزانه از سوی سطوح محیطی نظام مراقبت به مرکز مدیریت بیماری‌های واگیر وزارت بهداشت گزارش شده‌اند.

۲) روش‌های استفاده شده برای شناسایی الگوهای قابل توجیه

۱-۲) اثر روزهای هفته (Day of week effect)

به منظور شناسایی اثر روزهای هفته در داده‌های موارد مشکوک سرخک، از نمودار خطی(Line Plot)، نمودار میانگین متحرک(Moving Average Chart)، تابع خودهمبستگی (Autocorrelation Function) و تابع خودهمبستگی جزئی (Partial Autocorrelation Function) استفاده شدکه در ادامه، به شرح هر یک از آن‌ها پرداخته شده است.

- نمودار خطی: برای مشاهده آثاراحتمالی روزهای هفته با بزرگ‌نمایی در مقیاس هفتگی از این نمودار استفاده شد. تکرار

[Downloaded from irje.tums.ac.ir on 2024-08-02]

یک الگوی خاص در روزهای خاص، براساس مشاهده نمودار،نشان‌گر وجود اثر آن روز خاص در هفته خواهد بود.

- نمودار میانگین متحرک: از این نمودار نیز با بزرگ‌نمایی در مقیاس هفتگی برای مشاهده آثاراحتمالی روزهای هفته استفاده شد.

- تابع خودهمبستگی: برای نمایش وابستگی داده‌ها و اثر روزهای هفته از تابع خودهمبستگی با تأخیر (Lag) ۷ روز و ضرایب آن استفاده شد. مقادیر بزرگ‌تر همبستگی و معنی‌دار بودن آن نشان‌گر وجود اثر روزهای هفته است. آزمون به کار رفته برای مستقل بودن و نبود همبستگی (White Noise)، آماره کیو (Portmanteau (Q) Statistics) است (۶).

- تابع خودهمبستگی جزئی: افزون بر تابع خودهمبستگی، برای نمایش اثر روزهای هفته از تابع خودهمبستگی جزئی با تأخیر ۷ روز و ضرایب آن استفاده شد. مقادیر بزرگ‌تر همبستگی نشان‌گر وجود اثر روزهای هفته است.

۲-۲) اثر روزهای تعطیل و پایان هفته (Holiday Effect)

برای شناسایی اثر روزهای تعطیل و پایان هفته در داده‌های موارد مشکوک سرخک از شاخص متوسط پایان هفته (Weekend Mean)،یعنی متوسط موارد گزارش شده در پایان هفته (روزهای پنج شنبه و جمعه) و روزهای تعطیل رسمی در سالنامه هجری خورشیدی کشور و اختلاف آن با متوسط گزارش در روزهای کاری هفته (شنبه تا چهارشنبه)، استفاده شد.

۲-۳) دیگرالگوهای قابل توجیه

۲-۳-۱) نمودار خطی: برای مشاهده آثاراحتمالی ماهانه و سالانه نیز از این نمودار با بزرگ‌نمایی در مقیاس ماهانه و کلی استفاده گردید.

۲-۳-۲) تابع خودهمبستگی: برای نمایش وابستگی داده‌ها و روندهای ماهانه و سالانه از تابع خودهمبستگی،به ترتیب با تأخیر ۳۰ روز و ۳۶۵ روز استفاده شد. مقادیر بزرگ‌تر همبستگی و معنی‌دار بودن آن، نشان‌گر وجود الگوهای ماهانه و سالانه است. آزمون به کاررفته برای مستقل بودن و نبود همبستگی (White Noise)، آماره کیو (Portmanteau (Q) Statistics) است.

۲-۴) انحراف از پیش فرض نرمال بودن

انحراف از پیش فرض نرمال بودن مربوط به الگوریتم‌های زمانی استفاده شده برای شناسایی به هنگام طغیان‌های احتمالی با استفاده از شاخص‌های چولگی (Skewness) و کشیدگی (Kurtosis)ارزیابی شد. برای هموارسازی داده‌ها و در نتیجه، کاهش انحراف از پیش فرض نرمال بودن از روش میانگین‌های متحرک (Moving Averages) با دوره ۳ و ۷ روزه، استفاده شد.

نرم‌افزار استفاده شده برای شناسایی الگوهای قابل توجیه، نرم‌افزار SPSSنگارش ۱۷ است و برای ترسیم نمودارهای خطی از نرم‌افزار STATAنگارش ۱۰ استفاده شده است.

۳) روش استفاده شده برای حذف الگوهای قابل توجیه

در روش هموارسازی نمایی هولت- وینترز (Holt- Winter'sexponential smoothing) تصور می‌شود که داده‌ها متشکل از سه جزء سطح (Lt)، روند (Tt) و فصلی بودن (St) هستند که برای پیش‌بینی مقادیر در واحد زمانی بعدی که با k نمایش داده می‌شود، از فرمول زیر استفاده می‌گردد:

$$\hat{y}_{t+k} = (L_t + kT_t)S_{t+k-M},$$

که در آن M تعداد فصل‌ها است (برای تعیین دوره‌ای بودن داده در هفته برابر با ۷ خواهد بود) و سه جزء سطح (Lt)، روند (Tt) و فصلی بودن (St)، به شرح زیر محاسبه می‌شوند:

$$\begin{aligned} L_t &= \alpha\frac{Y_t}{S_{t-m}} + (1-\alpha)(L_{t-1}+T_{t-1}) \\ T_t &= \rho(L_t - L_{t-1}) + (1-\rho)T_{t-1} \\ S_t &= \gamma\frac{Y_t}{L_t} + (1-\gamma)(S_{t-M}), \end{aligned}$$

در فرمول‌های بالا α، β، γ پارامترهای هموارسازی(Smoothing) هستند که مقادیر بین صفر تا یک را به خود اختصاص می‌دهند.

در این مطالعه، از روش تجمعی و ضربی هموارسازی نمایی هولت- وینترز برای حذف الگوهای قابل توجیه داده‌های کشوری نظام مراقبت بیماری سرخک استفاده شد. مقادیر پارامترهای α، β، γ مورد استفاده در این مطالعه، با توجه به توصیه منابع موجود (۷) به ترتیب ۰/۴، ۰ و ۰/۱۵ بودند. با این حال، افزون بر مقادیر یادشده، از مقادیر بهینه شده پارامترهای α، β، γ با استفاده از نرم‌افزار XLSTAT(۸)نگارش ۲۰۱۱ برای دستیابی به مناسب‌ترین و کاراترین روش هموارسازی نیز استفاده شده است.

یافته‌ها

مشاهده نمودارهای خطی داده‌های خام موارد گزارش شده مشکوک به سرخک و نمودار میانگین متحرک مربوط،نشان‌گر وجود آثارروزهای هفته و پایان هفته است (نمودارهای یک و دو). در جدول شماره ۱ نیز مقادیر خودهمبستگی و خودهمبستگی جزئی همراه با آماره کیو (Portmanteau (Q) Statistics) آورده شده است.

[Downloaded from irje.tums.ac.ir on 2024-08-02]

همان‌طور که در جدول مشاهده می‌شود، مقادیر خودهمبستگی برای آثارروزهای هفته (تأخیر ۷ و ۱)، وجود الگوهای ماهانه (تأخیر ۳۰) و سالانه (تأخیر ۳۶۵) همگی با مقدار P کمتر از ۰/۰۰۱ معنی‌دار شده است که نشان‌گر وابسته بودن داده‌های موارد گزارش شده مشکوک به سرخک است. در جدول شماره ۲ میانگین و فاصله اطمینان ۹۵ درصد متوسط تعداد موارد مشکوک به سرخک، طی سال‌های ۱۳۸۷ تا ۱۳۸۹ برای شناسایی آثارروزهای هفته، تعطیل و پایان هفته

جدول شماره ۱- مقادیر خودهمبستگی و خودهمبستگی جزئی برای تأخیرهای ۱، ۳۰، ۳۶۵ و ۷ همراه با برخی از ضرایب تأخیر ۷ در داده‌های موارد گزارش شده مشکوک به سرخک طی دوره سه ساله

تأخیر (Lag)	تابع خودهمبستگی (ac)	تابع خودهمبستگی جزئی (pac)	آماره کیو* (Portmanteau (Q) Statistics)	P
۱	۰/۵۴۱	۰/۵۴۱	۳۲۱/۳	<۰/۰۰۱
۷	۰/۴۷۱	۰/۱۲۵	۱۸۱۴/۶	<۰/۰۰۱
۱۴	۰/۵۰۳	۰/۱۰۶	۳۶۰۹/۷	<۰/۰۰۱
۲۱	۰/۴۵۱	۰/۰۴۳	۵۰۶۲/۱	<۰/۰۰۱
۲۸	۰/۳۹۷	- ۰/۰۰۱	۶۴۶۲/۳	<۰/۰۰۱
۳۰	۰/۴۰۸	۰/۰۲۷	۶۸۳۲/۶	<۰/۰۰۱
۳۶۵	۰/۱۰۲	۰/۰۲۰	۱۴۵۷۴	<۰/۰۰۱

*Portmanteau (Q) Statistics for white noise

(پنج شنبه و جمعه) نشان داده شده است. فواصل اطمینان مربوط به میانگین در روزهای کاری هفته (شنبه تا چهارشنبه) پس از حذف روزهای تعطیل رسمی، یعنی ۴/۶ مورد (۹۵٪ **CI:** ۴/۳-۴/۸) و مقایسه چشمی آن با مقدار متناظر در روزهای پایان هفته، یعنی ۴/۱۳ مورد (۹۵٪ **CI:** ۳/۷-۴/۵) ، نشان‌گر اثر روزهای پایان هفته بر تعداد موارد گزارش شده و کاهش غیر واقعی موارد گزارش شده مشکوک به سرخک است. به رغم آن، مشاهده موارد گزارش شده روزانه با استفاده از نمودارهای خطی با بزرگ‌نمایی در مقیاس هفتگی نشان‌گر آن است که تعداد موارد گزارش شده مشکوک به سرخک در اوایل هفته، بالا می‌رود و سپس در میانه هفته، یعنی روز سه شنبه، به اوج می‌رسد و در روزهای پایان هفته کاهش می‌یابد. هرچند این مقدار در روز جمعه، به طور متوسط بالاتر از روزهای چهارشنبه و پنج‌شنبه است که دلایل احتمالی آن در قسمت بحث،بررسی شده است. مشاهده نمودار خطی (نمودار شماره ۱) و نمودار میانگین متحرک (نمودار شماره ۲) با بزرگ‌نمایی در مقیاس هفتگی تأییدکننده اثر روزهای هفته است. افزون بر این، متوسط تعداد موارد گزارش شده مشکوک به سرخک در روزهای تعطیل رسمی کشور (تمام روزهای تعطیل به جز جمعه‌ها)، یعنی ۳/۴۲ و مقایسه آن با دیگرروزهای هفته و حتی روزهای پایان هفته، بیان‌گر پایین بودن گزارش تعداد موارد گزارش شده مشکوک به سرخک در روزهای تعطیل رسمی کشور است (جدول شماره ۲).

جدول شماره ۳ شاخص‌های ارزیابی کارایی روش‌های میانگین‌های متحرک و هموارسازی نمایی هولت- وینترز در حذف الگوهای قابل توجیه داده‌های کشوری نظام مراقبت بیماری سرخک، طی دوره سه ساله را نشان می‌دهد. همان‌طور که در جدول مشاهده می‌شود، مقادیر چولگی و کشیدگی داده‌های خام دوره سه ساله، بیان‌گر مصداق نداشتن پیش فرض نرمال بودن هستند. پس از هموارسازی با روش ساده میانگین‌های متحرک با دوره ۷ روزه بهبود یافته و مقدار چولگی از ۱/۸۲ به ۱/۵۰ و کشیدگی از ۷/۷۹ به ۲/۳۱ رسیده است. در رابطه با روش هموارسازی نمایی هولت- وینترز، مقدار چولگی از ۱/۸۲ به ۱/۵۱ و کشیدگی از ۷/۷۹ به ۲/۴۴، با استفاده از روش ضربی کاهش یافته است. مقادیر خودهمبستگی نیز با استفاده از روش تجمعی هموارسازی نمایی هولت- وینترز برای تأخیرهای هفتگی، ماهانه و سالانه پس از هموارسازی با این روش کاهش یافته است. این کاهش نشان‌گر نقش روش تجمعی در حذف الگوهای فصلی شامل آثارروزهای هفته، ماه و سال است. با این حال، مقادیر متناظر برای روش میانگین‌های متحرک نه تنها کاهش نداشته، بلکه افزایش داشته است. روند داده‌های خام و هموار شده موارد گزارش شده مشکوک به سرخک طی دوره سه ساله در کشوربا استفاده از روش میانگین‌های متحرک با دوره ۷ روزه، در نمودار شماره ۳

[Downloaded from irje.tums.ac.ir on 2024-08-02]

نشان داده شده است.

کارایی این روش در هموارسازی که به پیشگیری از اعلان هشدارهای کاذب، به دنبال استفاده از الگوریتم‌های کشف طغیان،منجر می‌گردد، در نمودار آشکاراست. با این حال، مقادیر خودهمبستگی در داده‌های هموارشده نسبت به خام، کاهش پیدا نکرده است.

جدول شماره ۲- میانگین و فاصله اطمینان ۹۵ درصد برای متوسط تعداد موارد مشکوک به سرخک طی سال‌های ۱۳۸۷ تا ۱۳۸۹

روزهای هفته (تعداد روزها طی دوره)*	**میانگین (فاصله اطمینان ۹۵ درصد)**
شنبه (۱۴۶ روز)	۴/۹ (۴/۳ -۵/۵)
یک شنبه (۱۴۹ روز)	۴/۸ (۴/۲ -۵/۴)
دوشنبه (۱۴۵ روز)	۴/۶۱ (۴/۱۰ -۵/۱۲)
سه شنبه (۱۴۵ روز)	۴/۷ (۳/۹۸ -۵/۴۵)
چهارشنبه (۱۴۳ روز)	۳/۹ (۳/۳ -۴/۵)
روزهای کاری هفته (شنبه تا چهارشنبه ۷۲۸ روز)	۴/۶ (۴/۳۵ -۴/۸۹)
پنج شنبه (۱۴۵ روز)	۳/۸ (۳/۳ -۴/۳)
جمعه (۱۵۷ روز)	۴/۴ (۳/۸ -۴/۹)
روزهای پایان هفته(پنج شنبه و جمعه ۳۰۲ روز)	۴/۱ (۳/۷ -۴/۵)
روزهای تعطیل (جمعه و سایر روزهای تعطیل رسمی) (۲۲۰ روز)	۴/۱ (۳/۶ -۴/۵)
روزهای تعطیل رسمی (بجز جمعه) (۶۵ روز)	۳/۴ (۲/۶ -۴/۱)

*روزهای تعطیل رسمی کشور(از شنبه تا پنج‌شنبه) در نظر گرفته نشده است.

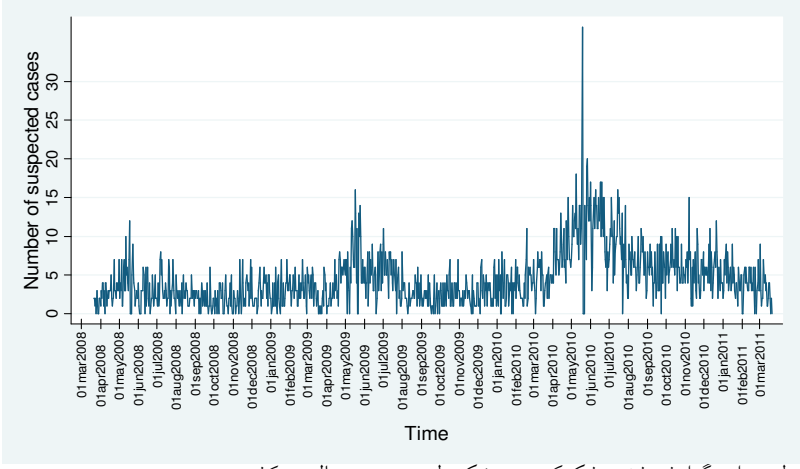
جدول شماره ۳- شاخص‌های ارزیابی کارایی روش هموارسازی نمایی هولت- وینترز در حذف الگوهای قابل توجیه از داده‌های کشوری نظام مراقبت بیماری سرخک، همراه با مقادیر خودهمبستگی مربوط

نام شاخص		داده‌های خام (۱۳۸۹- ۱۳۸۷)	داده‌های هموارشده با روش تجمعی هولت وینتر*	داده‌های هموارشده با روش ضربی هولت وینتر†	داده‌های هموارشده با روش میانگین‌های متحرک (۷ روز)
میانگین و انحراف معیار (Mean± SD)		۴/۴ ±۳/۵	۴/۴۵ ±۳/۸۶	۴/۶۲ ±۲/۷۱	۴/۴۲ ±۲/۶۶
میانگین و انحراف معیار متوسط پایان هفته(Mean± SD)		۴/۱۳ ±۳/۲۳	۳/۸۶ ±۳/۵۵	۴/۶۱ ±۲/۷۷	۴/۳۵ ±۲/۷۰
چولگی		۱/۸۲	۱/۴۵	۱/۵۱	۱/۵۰
کشیدگی		۷/۷۹	۵/۸۱	۲/۴۴	۲/۳۱
مقادیر خودهمبستگی	تأخیر ۷ (هفتگی)	۰/۴۷۱	۰/۴۱۰	۰/۸۷۸	۰/۷۹۴
	تأخیر ۳۰ (ماهانه)	۰/۴۰۸	۰/۲۹۸	۰/۶۸۸	۰/۶۶۳
	تأخیر ۳۶۵ (سالانه)	۰/۱۰۲	۰/۰۸۸	۰/۱۷۵	۰/۱۲۹

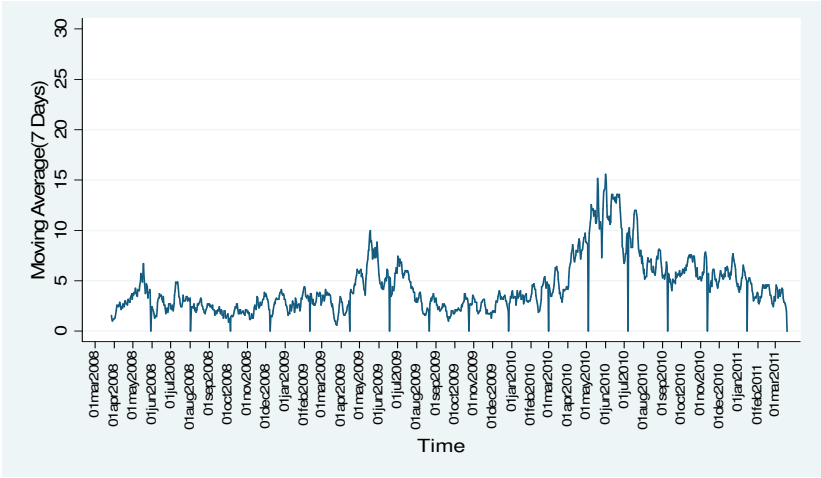
*Holt- Winters exponential smoothing (Additive)

†Holt- Winters exponential smoothing (Multiplicative)

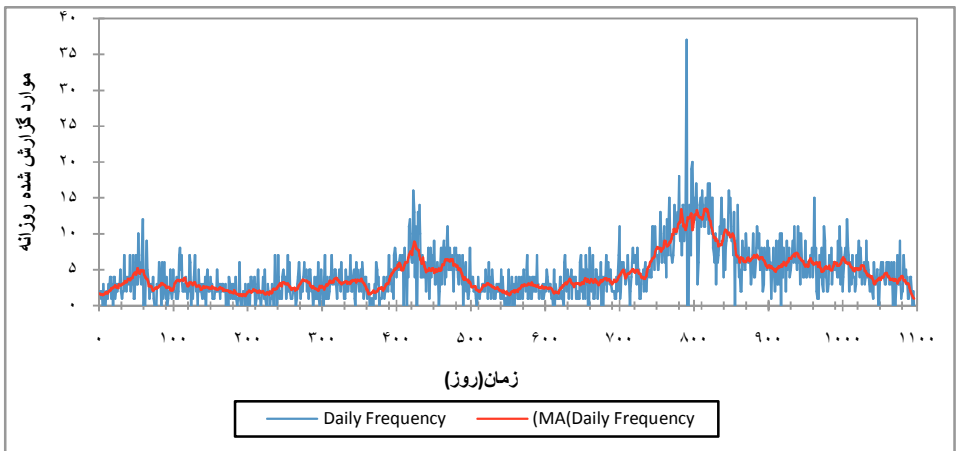
[Downloaded from irje.tums.ac.ir on 2024-08-02]



نمودار شماره ۱- نمودار خطی موارد گزارش شده مشکوک به سرخک طی دوره سه ساله در کشور



نمودار شماره ۲- نمودار خطی میانگین متحرک (۷ روز) موارد گزارش شده مشکوک به سرخک طی دوره سه ساله در کشور



نمودارشماره ۳- روند داده‌های خام و هموار شده موارد گزارش شده مشکوک به سرخک طی دوره سه ساله در کشوربا استفاده از روش میانگین‌های متحرک (۷روز).

[Downloaded from irje.tums.ac.ir on 2024-08-02]

بحث

از آنجا که هدف مطالعه حاضر شناسایی و حذف الگوهای قابل توجیه از داده‌های نظام مراقبت بیماری سرخک در کشور ایران است، در این قسمت، ابتدا نتایج کلیدی مرتبط با شناسایی الگوهای یادشده و انحراف از پیش فرض نرمال بودن بیان شده و در ادامه، به بحث درباره حذف الگوهای قابل توجیه با استفاده از دو روش میانگین‌های متحرک و هموارسازی نمایی هولت- وینترز پرداخته شده است. همان‌طور که در قسمت یافته‌ها اشاره شد، نتایج استفاده از روش‌های مختلف برای شناسایی الگوهای قابل توجیه،بیان‌گر وجود الگوهای سالانه، ماهانه و به ویژه آثار روزهای هفته و تعطیل در داده‌های کشوری موارد گزارش شده سرخک است. اگرچه بر اساس دانش نویسنده،مطالعه‌های مشابه داخلی در کشور وجود ندارد، نتایج مطالعه لوتز و همکارانش (۳) که با هدف شناسایی و حذف الگوهای قابل توجیه از داده‌های مربوط به فروش بدون نسخه داروی گلودرد و تعداد ویزیت‌های انجام شده به دلیل گاستروآنتریت انجام شد، با مطالعه حاضر همخوانی دارد. همچنین، نمودارهای تابع همبستگی و خودهمبستگی تأییدکننده وجود الگوهای یاد شده‌اند. در یک مطالعه دیگر که کولیر و همکارانش (۹) در سال ۲۰۱۰ میلادی با هدف ارزشیابی اخبار سلامت برای کمک به اعلان خودکار هشدار، با استفاده از داده‌های جست و جو شده روزانه از یک پایگاه اطلاعاتی انجام دادند، نتایج بیان‌گر کاهش ناگهانی در شمار جست وجوهای انجام شده به دلیل آثارروزهای هفته بود. افزون بر این، تأکید و توصیه مطالعه‌های دیگر (۱۲-۱۰) درباره ضرورت درنظرگرفتن نقش الگوهای فصلی مانند الگوهای ماهانه و آثار روزهای هفته در داده‌های استفاده شده در نظام‌های مراقبت، نیز بیان‌گر وجود این الگوهاست.

یکی از مهم‌ترین یافته‌های موجود در مطالعه حاضر که غیرمنتظره بود، روند متوسط تعداد موارد گزارش شده مشکوک به سرخک است. همان‌طور که در نمودار شماره ۵ و جدول شماره ۳ مشاهده می‌شود، موارد گزارش شده، مطابق انتظار،اوایل هفته بالا می‌رود و سپس،اواسط هفته، یعنی سه‌شنبه، به اوج خود می‌رسد و در روزهای پایان هفته کاهش می‌یابد. اما آنچه که نمی‌شد پیش‌بینی کرد، بالا بودن این مقدار در روز جمعه، نسبت به روزهای چهارشنبه و پنج‌شنبه بود. به نظر می‌رسد دو دلیل زیر از مهم‌ترین دلایل این یافته است:

دلیل نخست، اجرای برنامه‌های معمول و مطابق دیگر روزهای هفته، از جمله حضور مطابق برنامه و با اطلاع پیشین کارکنان محیطی نظام مراقبت، برای پایش موارد مشکوک به سرخک است. به سخن دیگر، روال‌مندشدن فعالیت‌های نظام مراقبت، مانند دیگر روزهای هفته، دلیل احتمالی افزایش تعداد موارد گزارش شده سرخک است.

دلیل دوم که ریشه در ویژگی‌های نظام مراقبت سرخک دارد، به هنگام نبودن و تأخیر احتمالی در گزارش موارد مشکوک به سرخک از سطوح محیطی نظام مراقبت به سطوح بالاتر نظام است. در واقع، ممکن است بخشی از موارد گزارش شده در روز جمعه مربوط به روز یا روزهای پیش باشد. گفتنی است که پذیرفتن دلیل دوم،مهم‌ترین محدودیت این مطالعه را، یعنی نامشخص بودن زمان تأخیر در گزارش موارد مشکوک به سرخک، توسط سطوح محیطی نظام مراقبت به سطوح بالاتر نمایان می‌سازد.

نتایج استفاده از روش‌های مختلف برای آزمون پیش فرض نرمال بودن داده‌ها بیان‌گر انحراف از این پیش فرض مهم است. مصداق نداشتن این فرضیه در داده‌های مطالعه حاضر به ایجاد مشکلاتی مهم، از جمله ایجاد هشدارهای کاذب و غیر واقعی منجر خواهد شد. از این رو، به نظر می‌رسد که دست اندرکاران نظام مراقبت باید هنگام استفاده از نمودارهای کنترل مانند نمودار شوهارت، CuSumو EWMA این مهم را بررسی کنند و در صورت انحراف از پیش فرض نرمال بودن الگوهای قابل توجیه در داده‌ها را حذف کنند. استفاده از روش میانگین‌های متحرک، به دلیل سادگی و عملکرد مناسب در کاهش انحراف از پیش‌فرض نرمال بودن پیشنهاد می‌گردد. نتایج حاصل از منبع ۳ که با هدف شناسایی و حذف الگوهای قابل توجیه از داده‌های مربوط به تعداد ویزیت‌های انجام شده به دلیل گاستروآنتریت انجام شد، بیان‌گر انحراف از پیش فرض یادشده بود (۳) که با استفاده از روش نسبت میانگین‌های متحرک، داده‌ها هموارسازی شد.

یافته‌های مطالعه درباره کارایی روش هموارسازی نمایی هولت- وینترز در حذف الگوهای قابل توجیه از داده‌های کشوری نظام مراقبت بیماری سرخک نشان‌گر کارایی بهتر روش ضربی هموارسازی نمایی هولت- وینترز با پارامترهای بهینه نسبت به روش تجمعی در بهبود مقادیر چولگی و کشیدگی داده‌های خام است. به رغم این، باید یادآور شدکه روش ساده میانگین‌های متحرک، نسبت به روش‌های یادشده در کاهش مقادیر چولگی و کشیدگی، به عنوان شاخص‌های ارزیابی پیش فرض نرمال بودن، از عملکردی بهتر برخوردار بوده است. این یافته با نتایج مطالعه

[Downloaded from irje.tums.ac.ir on 2024-08-02]

لوتز و همکارانش (۳) که چند روش هموارسازی را مقایسه کرده است، همخوانی دارد. در مقابل، روش تجمعی هموارسازی نمایی هولت- وینترز با پارامترهای بهینه نسبت به روش ضربی در حذف الگوهای فصلی (هفتگی، ماهانه و سالانه) مقادیر خودهمبستگی نیز با استفاده از روش تجمعی هموارسازی نمایی هولت- وینترز برای تأخیرهای هفتگی، ماهانه و سالانه از عملکردی بهتر برخوردار است به طوری که مقادیر خودهمبستگی به میزانی در خورِ توجه در داده‌های هموارشده، نسبت به داده‌های خام، کاهش یافته است. این کاهش که نشان‌گر نقش مناسب روش تجمعی در حذف الگوهای فصلی، شامل آثارروزهای هفته، ماه و سال است، با نتایج مطالعه لوتز و همکارانش (۳) همخوان است. اگرچه در مطالعه لوتز و همکارانش به ضربی یا تجمعی بودن روش هولت- وینترز مورد استفاده و نیز مقادیر پارامترهای مربوط اشاره نشده، منبع ۶ مقادیر پارامترهای α، β، γ را به ترتیب ۰/۴، ۰ و ۰/۱۵ توصیه نموده‌است (۷)، نویسندگان در این مطالعه کاراترین مقادیر را با بهره‌گیری از پارامترهای یاد شده و مقادیر بهینه شده آن‌ها گزارش کرده‌اند.

استفاده از رویکردهای مختلف شناسایی الگوهای فصلی و توجیه‌پذیر در داده‌های یک دوره سه ساله موارد گزارش شده مظنون به سرخک، به عنوان یکی از منابع داده مورد استفاده در نظام مراقبت از بیماری سرخک در کشور، یکی از مهم‌ترین نقاط قوت مطالعه حاضر است. از سویی، به نظر می‌رسدکه به‌رغم به کارگیری و مقایسه دو روش پرکاربرد هموارسازی برای پیش پردازش داده‌ها در این مطالعه، مقایسه عملکرد رویکردهای مختلف دیگر هموارسازی، مانند روش‌های مبتنی بر مدل‌ها، از جمله پرسش هایی است که این مطالعه به تنهایی قادر به پاسخ دادن به آن نیست.

نتیجه‌گیری

به طور کلی، اهمیت شناسایی الگوهای قابل توجیه و حذف آن‌ها به دلیل پیامدهای به‌کارگیری الگوریتم‌های کشف طغیان در نظام‌های مراقبت در مطالعه‌های مختلف و منابع مربوط (۷،۱۳،۲) آورده شده است. براساس نتایج این مطالعه، اگرچه کارایی روش میانگین‌های متحرک در پیش پردازش داده‌های این مطالعه و سادگی استفاده از آن با کمک نرم‌افزارهای شناخته شده مانند اکسل، آشکاراست، اما این روش در حذف الگوهای فصلی، عملکرد مطلوب ندارد.

تشکر و قدردانی

نویسندگان لازم می‌دانند تا از یکایک کارکنان نظام مراقبت بیماری سرخک در تمام سطوح در کشور، به ویژه دست‌اندکاران این نظام در اداره بیماری‌های قابل پیشگیری با واکسن و قرنطینه- مرکز مدیریت بیماری‌ها، جناب آقای دکتر محسن زهرایی و سرکار خانم اعظم صبوری برای در اختیار گذاشتن داده‌ها و نظرهای علمی و ارزنده‌شان سپاسگزاری و قدردانی کنند.

منابع

1- Mandl KD, Overhage JM, Wagner MM, Lober WB, Sebastiani P, Mostashari F, et al. Implementing syndromic surveillance: a practical guide informed by the early experience. J Am Med Inform Assoc. Mar-Apr 2004; 11: 141-50.

2- Chen H, Zeng D, Yan P. Infectious Disease Informatics Syndromic Surveillance for Public Health and Bio Defense, Integrated Series in Information Systems. Springer Science and Business Media, LLC 2010.

3- Lotze TH, Murphy S, Shmueli G. Implementation and Comparison of Preprocessing Methods for Biosurveillance Data. Advances in Disease Surveillance. 2008;1.

4- Zeng D, Chen H, Chavez C, lober WB, Thurmond MC. *Infectious* Disease Informatics and Biosurveillance, Integrated Series in Information Systems: Springer Science+Business Media; 2011.

5- Zahraei SM, Dadras MN, Saborio A .National guideline for measles surveillance (Elimination Phase). Ministry of Health, Tehran, Iran: Arvij; 2009.

6- Wegman, EG. Time Series Analysis:Theory, Data Analysis and Computation. Fairfax: George Mason University; 1996.

7- Burkom HS, Murphy SP, Shmueli G. Automated time series forecasting for biosurveillance. STATISTICS IN MEDICINE.2007; 26: 4202- 18.

8- *XLSTAT* [computer program]. New York: Addinsoft; 2011.

9- Collier N. What's unusual in online disease outbreak news? *J* Biomed Semantics. 2010; 1: 2.

10- Yahav I, Shmueli G. Algorithm combination for improved performance in biosurveillance systems. Lect Notes Comput Sc. 2007; 4506: 91-102.

11- Kleinman KP, Abrams AM, Kulldorff M, Platt R. A model-adjusted space-time scan statistic with an application to syndromic surveillance. Epidemiol Infect. Jun 2005; 133: 409-19.

[Downloaded from irje.tums.ac.ir on 2024-08-02]

12- Chan TC, King CC, Yen MY, Chiang PH, Huang CS, Hsiao CK. Probabilistic daily ILI syndromic surveillance with a spatio-temporal Bayesian hierarchical model. PLoS One. 2010; 5: e11626.

13- Elbert Y, Burkom HS. Development and evaluation of a data-adaptive alerting algorithm for univariate temporal biosurveillance data. Statistics in Medicine. 2009; 28: 3226-48.

[Downloaded from irje.tums.ac.ir on 2024-08-02]

Iranian Journal of Epidemiology 2012; 8(3): 12-21.

Original Article

Detecting and Removing the Explainable Patterns of the Daily Counts of Suspected Cases of Measles as a Prediagnostic Data Source in Iran

Karami M[1], Soori H[2], Mehrabi Y[3], Haghdoost AA[4], Gouya MM[5]

1- Assistant Professor of Epidemiology, Department of Biostatistics and Epidemiology, Hamadan University of Medical Sciences, Hamadan, Iran
2- Professor of Epidemiology, Safety Promotion and Injury Prevention Research Center, Faculty of Public Health, Shahid Beheshti University of Medical Sciences, Tehran, Iran
3- Professor of Biostatistics, Faculty of Public Health, Shahid Beheshti University of Medical Sciences, Tehran, Iran
4- Associate Professor of Epidemiology and Biostatistics, Research Centre of Modeling in Health & Department of Epidemiology and Biostatistics, School of Health, Kerman University of Medical Sciences, Kerman, Iran
5- Director General, Center for Disease Control, Ministry of Health & Medical Education, Tehran, Iran
Corresponding author: Soori H., hsoori@yahoo.com

Background & Objectives: Knowledge of the presence of seasonal trends and other explainable patterns in the prediagnostic data sources and removing such patterns before applying outbreak detection methods seem very important. This study aimed to detect and remove the explainable patterns such as seasonality, day-of-week (DOW) and holiday effects of the daily counts of suspected cases of measles in Iran.

Methods: Data on daily counts of suspected cases of measles as a pre-diagnostic data source were obtained from Iranian national surveillance system between 21 March 2008 and 20 March 2011. We used lines plot, moving average chart, autocorrelation and partial autocorrelation functions for detecting explainable patterns. Moving average (MA) and Holt- Winters (HW) exponential smoothing method are used for removing explainable patterns.

Results: Our findings indicate the presence of seasonality, DOW effect, holidays and weekend effects in the daily counts of suspected cases of measles. The good performance of HW exponential smoothing technique in removing seasonal patterns is evident. MA technique showed better performance regarding assumption violation on outbreak detection methods.

Conclusion: Because of the presence of explainable patterns in the daily counts of suspected cases of measles, considering such patterns before applying outbreak detection algorithms is very important. Implementing both MA (7 days) techniques for its simplicity as a pre- processing method and HW method for its efficacy in removing seasonal patterns is recommended.

Keywords: Seasonality, Outbreak, Smoothing, Measles, Public health surveillance

[Downloaded from irje.tums.ac.ir on 2024-08-02]