

شناسایی و حذف روند های فصلی و الگوهای قابل توجیه در دادهای نظام مراقبت بیماری مننژیت

خبثات پژوهی: دانشجوی کارشناسی ارشد اپیدمیولوژی، گروه آمار زیستی و اپیدمیولوژی، دانشکده بهداشت، دانشگاه علوم پزشکی همدان، همدان، ایران. pazhouhi.h@gmail.com

* **دکتر منوچهر کریمی:** استادیار و متخصص اپیدمیولوژی، مرکز تحقیقات مدل سازی بیماری های غیرواگیر و گروه آمار زیستی و اپیدمیولوژی، دانشکده بهداشت، دانشگاه علوم پزشکی همدان، همدان، ایران (نویسنده مسئول). ma.karami@umsha.ac.ir

دکتر نادر اسماعیل نسب: دانشیار و متخصص اپیدمیولوژی، گروه اپیدمیولوژی و آمار زیستی، دانشکده بهداشت، دانشگاه علوم پزشکی سنج، سنج، ایران. esmailnasab@yahoo.com

دکتر عباس مقیم بیگی: دانشیار و متخصص آمار زیستی، مرکز تحقیقات مدل سازی بیماری های غیرواگیر و گروه آمار زیستی و اپیدمیولوژی، دانشکده بهداشت، دانشگاه علوم پزشکی همدان، همدان، ایران. moghimb@yahoo.com

محمد فریادرس: دانشجوی کارشناسی ارشد اپیدمیولوژی، گروه آمار زیستی و اپیدمیولوژی، دانشکده بهداشت، دانشگاه علوم پزشکی همدان، همدان، ایران. feryadresmohammad@gmail.com

تاریخ پذیرش: ۹۲/۱۲/۲۴

تاریخ دریافت: ۹۲/۱۱/۲۶

چکیده

زمینه و هدف: تشخیص به موقع طغیان های ناشی از بیماری مننژیت مستلزم شناسایی و حذف الگوهای قابل توجیه در داده ها از اهمیت بالایی برخوردار است. این مطالعه با هدف شناسایی و حذف الگوهای قابل توجیه در داده های سندروم تب و علائم نورولوژیک (موارد مشکوک به مننژیت) در استان همدان انجام شده است.

روش کار: مطالعه حاضر یک مطالعه سری زمانی است. داده های موارد گزارش شده مشکوک به مننژیت در استان همدان طی سال های ۱۳۸۹ تا ۱۳۹۱ مورد تجزیه و تحلیل قرار گرفت. از نمودارهای خطی، چند جمله ای موضعی و میانگین متحرک، توابع خودهمبستگی و خودهمبستگی جزئی، شاخص متوسط پایان هفته و آماره ناپارامتریک من-کندل برای شناسایی الگوهای قابل توجیه در داده های موارد مشکوک به مننژیت استفاده گردید. روش های هموارسازی سری فوریه و LOWESS (Locally Weight Regression and Smoothing Scoter plot) نیز جهت حذف الگوهای قابل توجیه به کارگرفته شد.

یافته ها: نتایج نمودار چند جمله ای موضعی، روش های توابع خودهمبستگی و خودهمبستگی جزئی، شاخص متوسط پایان هفته و آماره ناپارامتریک من-کندل حاکی از وجود الگوهای قابل توجیه (سالانه، ماهیانه، اثرات روزهای هفته و اثر تعطیلات) در داده های مشکوک به بیماری مننژیت است. هر دو روش حذف الگوهای قابل توجیه در حذف این الگوها در داده های موارد مشکوک به مننژیت کارایی داشتند. به طور کلی، روش LOWESS در حذف الگوهای قابل توجیه و روش سری فوریه در بهبود نرمال بودن داده ها عملکردی بهتری داشتند.

نتیجه گیری: نتایج مطالعه وجود الگوهای قابل توجیه در داده های موارد مشکوک به مننژیت را تایید و ضرورت هموارسازی و حذف این الگوها برای شناسایی صحیح و به موقع طغیان ها را آشکار می سازد.

کلیدواژه ها: مننژیت، طغیان، الگوهای فصلی، هموارسازی، نظام مراقبت.

مقدمه

نمودارهای کنترل و الگوریتمها برای نظارت بر داده های سری زمانی برای تشخیص "علل خاص" و یا اختلالات استفاده می کند. در نظام های مراقبت، چنین ناهنجاری هایی به طور بالقوه نشان دهنده طغیان می باشند (۳). داده های سندروم تب و علائم نورولوژیک یا داده های مشکوک به مننژیت مانند همه داده های پیش تشخیصی می توانند دو الگو در خود داشته باشند. یکی الگوهای قابل توجیه (Explainable patterns)

مننژیت به لحاظ میزان مرگ و میر، عوارض و توانایی ایجاد همه گیری یکی از نگران کننده ترین بیماری های عفونی است (۱). کنترل طغیان های مننژیت با وجود نظام مراقبت پویا و گزارش دهی به موقع امکان پذیر است (۲). نظام مراقبت با هدف تشخیص به موقع طغیان ها از داده های سندرومیک استفاده می کند. نظام مراقبت سندرومیک همانند محیط های صنعتی، از

دسته مبتنی بر مدل‌ها (Model based methods) قرار می‌گیرند. سری فوریه و LOWESS از روش داده محور می‌باشند که در آن مطالعه مورد استفاده قرار گرفتند (۸-۱۰). این مطالعه به منظور شناسایی و حذف روندها و الگوهای قابل توجیه مننژیت جهت شناسایی به‌هنگام اپیدمی و کاهش مرگ و میر و عوارض ناشی از بیماری مننژیت می‌باشد.

روش کار

در این پژوهش داده‌های انبوه کلیه موارد سندروم تب و علائم نورولوژیک (موارد مشکوک به مننژیت) گزارش شده در نظام مراقبت بیماری مننژیت استان همدان از سال ۱۳۸۳ الی ۱۳۹۱ می‌باشد. داده‌های مربوط به سال‌های ۱۳۸۳ الی ۱۳۸۸ به صورت ماهیانه مورد استفاده قرار گرفتند. همچنین از تاریخ ۱۳۸۹/۱/۱ الی ۱۳۹۱/۱۲/۳۰ داده‌های موارد گزارش شده سندروم تب و علائم نورولوژیک با توالی روزانه مورد بررسی قرار گرفتند.

به منظور شناسایی الگوهای قابل توجیه شامل اثر روزهای هفته (Day of week effect)، روزهای تعطیل (Holiday effect)، اثر پایان هفته (Weekend effect) و سایر الگوهای زمانی در داده‌های سندروم تب و علائم نورولوژیک از روش‌های نمودار خطی (Line plot)، نمودار میانگین متحرک (Moving average chart)، نمودار چندجمله‌ای موضعی (Local polynomial plot)، تابع خود همبستگی (Autocorrelation function)، تابع خود همبستگی جزئی (Partial autocorrelation function)، متوسط پایان هفته و آماره من-کندل استفاده شد (۱۰ و ۱۱).

نمودار خطی: برای مشاهده اثرات احتمالی روزهای هفته با بزرگ نمایی در مقیاس هفتگی از این نمودار استفاده شد. تکرار یک الگوی خاصی در روزهای خاص براساس مشاهده نمودار دال بر وجود اثر آن روز خاص در هفته خواهد بود. جهت مشاهده اثرات احتمالی ماهیانه و سالیانه از این نمودار نیز با بزرگ نمایی در مقیاس ماهیانه و کلی

یعنی الگوهای وابسته به زمان، اثرات تعطیلات، روزهای هفته و آخر هفته و الگوی دوم، الگوهای غیر قابل توجیه (Unexplainable patterns) است که شامل کاهش و یا افزایش احتمالی داده‌ها است (۴).

در ایران نیز همانند سایر کشورها بیماری مننژیت از سال ۱۳۶۰ جز بیماری‌های قابل گزارش‌دهی است به طوری که در حال حاضر نظام مراقبت از مننژیت به صورت گزارش‌دهی فوری می‌باشد (۵). نظام مراقبت بیماری مننژیت از داده‌های بالینی و غیربالینی نظیر موارد مشکوک بیماری مننژیت و تاییدیه آزمایشگاهی با هدف تشخیص زود هنگام طغیان‌ها و مراقبت از بیماری‌ها استفاده می‌کند. با این وجود، تصمیم‌گیری براساس داده‌های گزارش شده روزانه موارد مشکوک به مننژیت به صورت خام همانند سایر منابع داده‌ای مورد استفاده در نظام‌های مراقبت سندرومیک که دارای روندها و الگوهای خاصی است ممکن است منجر به ایجاد هشدارهای کاذب (False Alarms) می‌گردد. مهم‌ترین علت ایجاد این هشدارها، وجود الگوهای قابل توجیه در داده‌ها هستند (۶).

مفروضاتی مانند مستقل بودن و توزیع شناخته شده داده‌ها برای به کارگیری ابزارهای پایش نظام مراقبت به ویژه نظام‌های مراقبت سندرومیک نظیر الگوریتم‌های مجموع تراکمی (CUSUM) و میانگین متحرک وزن داده‌شده نمایی (EWMA) با هدف شناسایی بهنگام طغیان‌های احتمالی و یا هرگونه تغییر در روند بیماری‌ها بسیار مهم‌اند و از مفروضات روش‌های تشخیص طغیان هستند (۷). در صورتی که این مفروضات رعایت نگردد باعث هشدارهای کاذب می‌گردد. از طرفی وجود الگوهای قابل توجیه و غیرقابل توجیه در داده‌ها بهداشتی منجر به نقض مفروضات و در نتیجه بالا رفتن هشدارهای کاذب می‌شود (۸)؛ بنابراین، به منظور افزایش اثربخشی پایش در نظام‌های مراقبت، داده‌های سلامت نیاز به شناسایی و حذف الگوهای قابل توجیه می‌باشد. این فرآیند حذف الگوهای قابل توجیه در ادبیات روش پایش پردازش نامیده می‌شود. روش‌های پایش پردازش در دو

جهت هموارسازی سری زمانی ابتدا سری زمانی به مختصات فوریه تبدیل می‌شود و سپس فرکانس‌های بالاتر را حذف می‌کند. در نهایت مختصات فوریه به سری هموار زمانی تبدیل می‌شود.

LOWESS (Locally Weight Regression and Smoothing Scoter plot): نوعی از رگرسیون ناپارامتریک است که به منظور تعیین یک خط صاف یا منحنی از نمودار پراکندگی داده‌ها است. برای اولین بار کلویلند در سال ۱۹۷۹ این روش را معرفی کرد. از این روش برای پیش بینی و هموارسازی داده‌ها استفاده می‌گردد. در این مطالعه برای حذف الگوهای قابل توجیه از آن استفاده شد. برای جلوگیری از هموارسازی بیش از حد در روش LOWESS از پهنای باند $0/3$ استفاده شد. پهنای باند هرچه بیشتر باشد هموارسازی بیش از حد بیشتر رخ می‌دهد (۱۲). برای شناسایی الگوهای قابل توجیه و حذف الگوهای قابل توجیه به روش سری فوریه از نرم افزارهای XLSTAT و برای رسم نمودارها و حذف الگوهای قابل توجیه با روش LOWESS از نرم‌افزار STATA نگارش ۱۱ استفاده گردید.

یافته‌ها

نمودار خطی و نمودار میانگین متحرک کارآیی لازم جهت شناسایی الگوهای قابل توجیه را نداشتند (اشکال ۱ و ۲). به نظر می‌رسد مهم ترین دلیل آن موارد صفر روزانه باشد. دلایل احتمالی دیگر آن در قسمت بحث بررسی شده است. نمودار چندجمله‌ای موضعی با فاصله اطمینان ۹۵٪ نشان داد که بیشترین موارد مشکوک به مننژیت در وسط هفته رخ داده است و در پایان هفته (پنج‌شنبه و جمعه) کاهش داشته است (شکل ۳). مقادیر میانگین و فاصله اطمینان بیانگر وجود اثرات روزهای هفته، پایان هفته و تعطیلات می‌باشد (جدول ۱). همان‌طور که در جدول ۱ مشاهده می‌شود، میانگین موارد در طول هفته افزایش می‌یابد به طوری که بیشترین موارد مشکوک به مننژیت را در روز چهارشنبه با میانگین $1/66$ مورد و فاصله اطمینان ۹۵ درصد

استفاده گردید.

نمودار میانگین متحرک: جهت مشاهده اثرات احتمالی روزهای هفته با نمودار میانگین متحرک از دوره ۷ استفاده گردید.

نمودار چندجمله‌ای موضعی: نمودار چندجمله‌ای موضعی مقادیر هموارسازی موارد روزانه را بر مبنای رگرسیون ناپارامتریک نشان می‌دهد. جهت مشاهده اثرات احتمالی روزهای هفته از فاصله اطمینان ۹۵٪ استفاده شد.

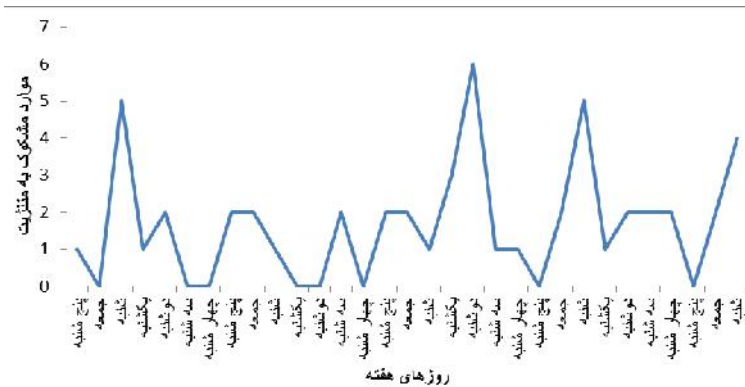
تابع خودهمبستگی و خود همبستگی جزئی: برای اثر روزهای هفته از تابع خود همبستگی با تاخیر (lag) ۷ روز و ضرایب آن استفاده شد. تاخیر ۳۰ روز و ۳۶۵ روز از تابع خود همبستگی جهت نمایش وابستگی داده‌ها و روند های ماهیانه و سالیانه استفاده شد. مقادیر بزرگ تر همبستگی و معنی دار بودن آن دال بر وجود اثر روزهای هفته و اثر روند و سالیانه می‌باشد. آزمون مورد استفاده برای مستقل بودن و نبود همبستگی (White Noise)، آماره کیو (Portmanteau (Q) Statistics) است.

آماره ناپارامتریک من-کندل (Mann-Kendall): آزمون جهت تعیین میزان، جهت و معنی داری روند به کار برده می‌شود و برای روندهای خطی و غیر خطی مناسب است. جهت شناسایی روندهای فصلی و غیرفصلی از این آماره استفاده شد.

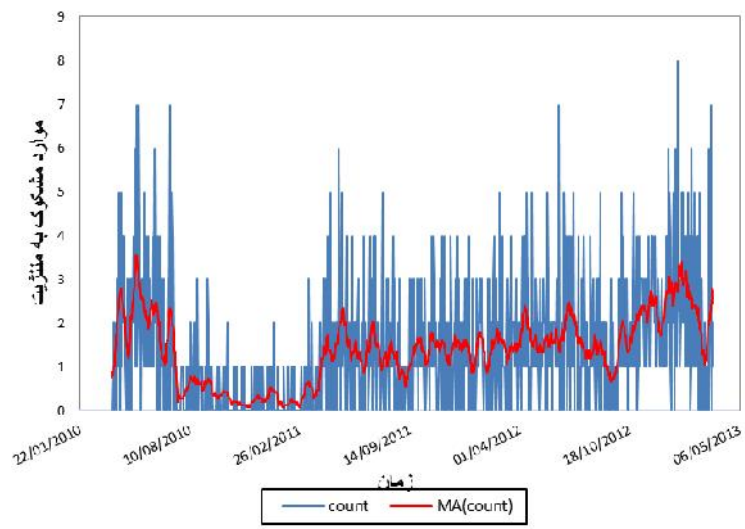
انحراف از پیش فرض نرمال بودن: انحراف از پیش فرض نرمال بودن مربوط به الگوریتم های زمانی مورد استفاده برای شناسایی به هنگام طغیان های احتمالی با استفاده از شاخص های چولگی (Skewness) و کشیدگی (Kurtosis) مورد ارزیابی قرار گرفت.

جهت شناسایی الگوهای قابل توجیه در داده‌های مشکوک به مننژیت روش های هموارسازی سریه فوریه و LOWESS به کار گرفته شد.

روش سری فوریه: سری فوریه عبارت است از بسط تابع تناوبی $f(x)$ در قالب جملاتی از جمع نامتناهی کسینوس‌ها و سینوس‌ها. در واقع سری فوریه بر کاربرد روابط تعامد توابع سینوسی و کسینوسی تاکید دارد. یکی از کاربردهای سری فوریه هموارسازی داده‌های سری زمانی می‌باشد.



شکل ۱- نمودار خطی بزرگ نمایی شده موارد مشکوک به مننژیت در مقیاس یک ماه

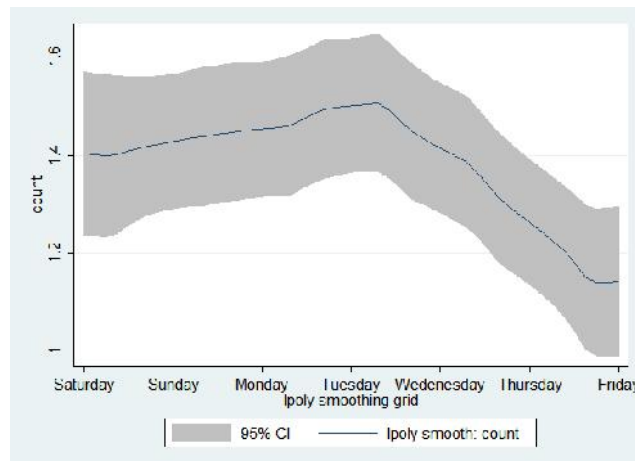


شکل ۲- نمودار میانگین متحرک (دوره ۷) مشکوک به مننژیت از سال ۱۳۸۹ الی ۱۳۹۱

این معناداری به منزله وابستگی داده‌ها در الگوهای قابل توجه می‌باشد (جدول ۲). نتایج شاخص‌های کارآیی روش‌های سری فوریه و LOWESS نشان دادند که هر دو روش کارآیی در حذف الگوهای قابل توجه داده‌های مشکوک به مننژیت داشته‌اند به طوری که چولگی و کشیدگی داده‌های خام روزانه مشکوک به مننژیت بعد از هموارسازی با سری فوریه به ترتیب از $1/21$ و $4/35$ به $0/06$ و $4/22$ کاهش پیدا کرده است، این دو شاخص بعد از هموارسازی با روش LOWESS به $0/92$ و $4/29$ کاهش پیدا کرده است. با این وجود سری فوریه در بهبود توزیع نرمال داده‌ها عملکرد بهتری از LOWESS داشته است (جدول ۳). همان‌طور که در جدول ۳ مشاهده می‌کنید مقادیر خودهمبستگی در تاخیرهای ۷ روز (اثر روزهای سالانه) برای داده‌های ماهیانه مشکوک به مننژیت با مقدار احتمال کمتر از $0/001$ معنادار شده‌اند.

($1/41-1/92$) به بالاترین حد خود می‌رسد و در پایان هفته (پنج‌شنبه و جمعه) به یک باره به میانگین $1/14$ مورد (فاصله اطمینان ۹۵ درصد: $0/99-1/29$) کاهش پیدا می‌کند. این به منزله وجود اثر پایان هفته در داده‌ها می‌باشد. این امر در مورد روزهای تعطیل رسمی با میانگین $1/09$ (فاصله اطمینان ۹۵ درصد: $0/83-3/35$) نیز صدق می‌کند.

شکل ۳ نیز بیانگر اثر روزهای هفته و پایان هفته در داده‌های مشکوک به مننژیت می‌باشد. مقادیر خودهمبستگی و خودهمبستگی جزئی در تمامی تاخیرهای ۱، ۷ و ضرایب ۷ (اثر روزهای هفته) برای داده‌های روزانه مشکوک به مننژیت و تاخیرهای ۱ (اثر ماهیانه)، ۳ (اثر فصلی) و ۱۲ (اثر سالانه) برای داده‌های ماهیانه مشکوک به مننژیت با مقدار احتمال کمتر از $0/001$ معنادار شده‌اند.



شکل ۳- نمودار چند جمله ای موضعی موارد مشکوک به مننژیت طی دوره سه ساله

می باشد. در اشکال ۴ و ۵ نیز به صورت گرافیکی به مقایسه روش های مختلف حذف الگوهای قابل توجیه در داده ها پرداخته شده است.

جدول ۱- مقادیر میانگین و فاصله اطمینان ۹۵ درصد موارد گزارش شده مشکوک به مننژیت در روزهای هفته طی سال های ۱۳۸۹ تا ۱۳۹۱

روزهای هفته	تعداد روز	میانگین (فاصله اطمینان ۹۵٪)
شنبه	۱۴۳	۱/۴۱ (۱/۱۶ و ۱/۶۶)
یکشنبه	۱۴۶	۱/۴ (۱/۱۷ و ۱/۶۲)
دوشنبه	۱۴۶	۱/۵۵ (۱/۲۸ و ۱/۸۲)
سه شنبه	۱۴۷	۱/۴۳ (۱/۲۱ و ۱/۶۶)
چهارشنبه	۱۴۷	۱/۶۶ (۱/۴۱ و ۱/۹۲)
پنجشنبه	۱۴۷	۱/۱۴ (۰/۹۴ و ۱/۳۳)
جمعه	۱۵۶	۱/۱۵ (۰/۹۴ و ۱/۳۶)
روزهای کاری هفته (شنبه تا چهارشنبه)	۷۲۹	۱/۴۹ (۱/۳۸ و ۱/۶۰)
روزهای پایان هفته (پنج شنبه و جمعه)	۲۹۶	۱/۱۴ (۰/۹۹ و ۱/۲۹)
روزهای تعطیل (جمعه و سایر روزهای تعطیل رسمی)	۲۲۰	۱/۱۴ (۰/۹۷ و ۱/۳)
روزهای تعطیل رسمی (بجز جمعه)	۶۴	۱/۰۹ (۰/۸۳ و ۱/۳۵)

مشکوک به مننژیت به ترتیب از ۰/۲۳ و ۰/۲۵ بعد از پیش پردازش با روش سری فوریه به ۰/۱۲ و ۰/۰۴ کاهش پیدا کرده است، مقادیر همبستگی بعد از پیش پردازش با روش LOWESS به ۰/۰۰ و ۰/۰۶ کاهش پیدا کرده است. این نتایج به منزله عملکرد مطلوب LOWESS در حذف اثر روزهای هفته می باشد.

عملکرد روش های هموارسازی را در داده های ماهیانه موارد مشکوک به مننژیت در جدول ۴ آورده شده است. همان طور که مشاهده می شود، مقادیر همبستگی تاخیر ۱ ماه، تاخیر ۳ ماه و تاخیر ۱۲ ماه به ترتیب جهت بررسی اثرات ماهیانه، فصلی و سالیانه برای داده های ماهیانه در جدول ۴ گزارش شده است. نتایج حاصل از آماره من-کندل بیانگر آن است که این آماره در داده های روزانه و ماهیانه مشکوک به مننژیت با مقادیر احتمال کمتر از ۰/۰۰۱ معنادار شده است. این نتایج به منزله وجود الگوهای فصلی و روند در داده های روزانه و ماهیانه مشکوک به مننژیت می باشد (جداول ۳ و ۴). همان طور که در جداول مشاهده می گردد بعد از پیش پردازش داده ها با روش های سری فوریه و LOWESS هیچ کدام از آماره فصلی و روند من-کندل در داده های روزانه و ماهیانه مشکوک به مننژیت معنادار نشده اند. این یافته به منزله حذف الگوهای قابل توجیه در داده های مشکوک به مننژیت بعد از هموارسازی

جدول ۲- مقادیر خود همبستگی و خود همبستگی جزئی در تاخیرهای ۱ و ۳۰ با همراه تاخیر ۷ و برخی ضرایب آن برای موارد روزانه مشکوک به مننژیت و تاخیرهای ۱، ۳ و ۱۲ برای موارد ماهیانه مشکوک به مننژیت

داده	تأخیر (Lag)	تابع خودهمبستگی (ac)	تابع خودهمبستگی جزئی (pac)	آماره کیو* (Portmanteau (Q) Statistics)	p
روزانه	۱	۰/۲۳	۰/۲۳	۹۹/۶۳	<۰/۰۰۱
	۷	۰/۲۳	۰/۰۹	۳۳۰/۲۴	<۰/۰۰۱
	۱۴	۰/۲۱	۰/۰۹	۵۱۱/۱۳	<۰/۰۰۱
	۲۱	۰/۲۵	۰/۱	۶۰۴/۱۷	<۰/۰۰۱
	۲۸	۰/۲۸	۰/۱۳	۶۱۹/۴۸	<۰/۰۰۱
ماهیانه	۳۰	۰/۲۵	۰/۰۹	۶۸۳۲/۶۴	<۰/۰۰۱
	۱	۰/۹	۰/۹۳	۹۰/۵۸	<۰/۰۰۱
	۳	۰/۶۶	۰/۱۳	۲۰۶/۶۱	<۰/۰۰۱
	۱۲	۰/۴۸	۰/۱۲	۴۶۳/۱۷	<۰/۰۰۱

جدول ۳- ارزیابی کارایی روش های حذف الگوهای قابل توجیه در داده های روزانه مشکوک به مننژیت طی سال های ۱۳۸۹ الی ۱۳۹۱

داده های خام	میانگین (انحراف معیار)	میانگین پایان هفته (انحراف معیار)	چولگی	کشیدگی	مقادیر خودهمبستگی			آماره من-کندال	
					تأخیر ۷	تأخیر ۳۰	تأخیر ۳۶۵	الگوی فصلی	
								تأخیر کندال	P
داده های خام	$1/37 \pm (1/44)$	$1/37 \pm (1/27)$	۱/۲۱	۴/۳۵	۰/۲۳	۰/۲۵	N.A.*	۰/۲۲	<۰/۰۰۱
سری فوریه	$0/0 \pm (0/8)$	$-0/09 \pm (0/8)$	۰/۰۶	۴/۲۲	۰/۱۲	۰/۰۴	N.A.	۰/۰۱	۰/۶۱
LOWESS	$0/01 \pm (1/26)$	$-0/23 \pm (1/12)$	۱/۰۴	۴/۲۹	۰/۰۰	۰/۰۶	N.A.	۰/۰۶	۰/۹۹

* N.A: non application

جدول ۴- ارزیابی کارایی روش های حذف الگوهای قابل توجیه در داده های ماهیانه مشکوک به مننژیت طی سال های ۱۳۸۳ الی

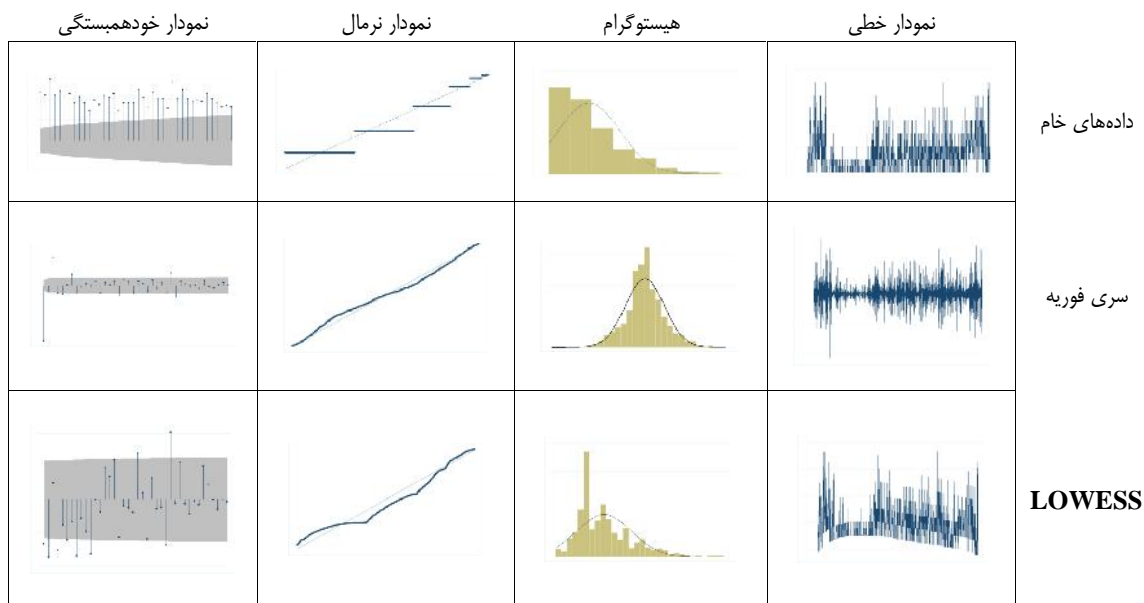
داده های خام	میانگین (انحراف معیار)	چولگی	کشیدگی	مقادیر خودهمبستگی			آماره من-کندال		
				تأخیر ۱	تأخیر ۳۰	تأخیر ۳۶۵	الگوی فصلی		
							تأخیر کندال	P	تأخیر کندال
داده های خام	$15/31 \pm (22/85)$	۱/۴۲	۳/۷۱	۰/۹۵	۰/۶۵	۰/۴۸	۰/۶۸	<۰/۰۰۱	۰/۶۲
سری فوریه	$0/0 \pm (3/89)$	۰/۰۶	۳/۶	-۰/۰۴	۰/۲۳	۰/۳۴	۰/۰۷	۰/۱۴	۰/۱۲
LOWESS	$0/22 \pm (11/25)$	۱/۲۲	۳/۲۳	۰/۳۴	-۰/۳۸	۰/۳	-۰/۰۷	۰/۸۶	۰/۶۹

بحث و نتیجه گیری

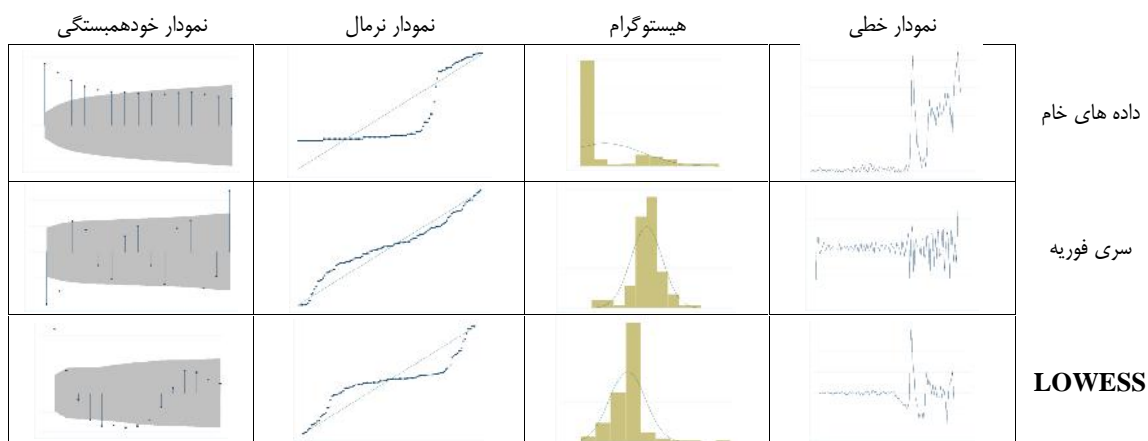
و در نتیجه بالا رفتن هشدارهای کاذب می شود؛ بنابراین، به منظور افزایش اثربخشی نظارت در نظام های مراقبت، داده های سلامت نیاز به پیش پردازش دارند.

مطالعه حاضر که با هدف پیش پردازش داده های نظام مراقبت مشکوک به مننژیت در استان همدان انجام شد بیانگر وجود الگوهای قابل توجیه در داده های مشکوک به مننژیت است. این یافته با نتایج مطالعه داخل کشور کرمی و همکاران (۴) با هدف شناسایی و حذف الگوهای قابل توجیه در داده های مشکوک به سرخک همخوانی دارد. وجود الگوهای قابل توجیه در داده ها سنדרومیک

در نظام های مراقبت سنדרومیک با توجه به حجم بالای داده های جمع آوری شده و منابع مختلف مستقیم و غیرمستقیم با بیماری از نمودارهای کنترل و الگوریتم های مجموع تراکمی و میانگین های متحرک وزن داده شده نمایی برای پایش بر داده های سری زمانی جهت تشخیص طغیان استفاده می شود (۱۳). وجود الگوهای قابل توجیه (اثرات روزهای هفته، تعطیلات، پایان هفته، الگوی فصلی، روندها و دیگر الگوهای زمانی) و غیرقابل توجیه در داده ها بهداشتی منجر به نقض مفروضات نمودارهای کنترل و الگوریتم ها می گردد



شکل ۴- نمودارهای شناسایی الگوهای قابل توجه در داده روزانه مشکوک به منتزیت قبل هموارسازی در ردیف ۱ و بعد از هموارسازی در ردیف‌های ۲ و ۳



شکل ۵- نمودارهای شناسایی الگوهای قابل توجه در داده ماهیانه مشکوک به منتزیت قبل هموارسازی در ردیف ۱ و بعد از هموارسازی در ردیف‌های ۲ و ۳

داده‌ها تاکید دارند. در مطالعه حاضر نشان داده شد که نمودارهای خطی و میانگین متحرک کارآیی شناسایی اثرات روزهای هفته و پایان هفته را نداشتند. این نتایج با نتایج مطالعات کرمی و همکارانش (۴) و لوتز و همکارانش (۱۰) همخوانی ندارد. به نظر می‌رسد، دلیل تفاوت مشاهده شده ناشی از موارد زیر باشد. اولین دلیل می‌تواند از موارد کم روزانه باشد. به طوری که میانگین روزانه در این مطالعه ۱/۳۷ مورد در روز می‌باشد، ولی در مطالعه کرمی و همکارانش میانگین روزانه ۴/۴ مورد در روز و در مطالعه لوتز و همکارانش (۱۰) میانگین روزانه از ۱۷ تا ۹۰۰ مورد در روز متغیر بوده است. دومین

در مطالعه خارجی لوتز و همکاران (۱۰) با هدف پیش پردازش داده‌های فروش داروها و موارد ویزیت گاستروانتریت نیز تایید شده است که با مطالعه حاضر همخوانی دارد. یافته‌های مطالعه حاضر با یافته‌های مطالعه‌ی متس و همکاران (۱۴) با هدف "ارزبایی داده‌های شکایت گزارش شده برای پیگیری زمان نزدیک به واقعی از بیماری‌های مسری جهت تشخیص شیوع بیماری و یا سایر خوشه‌های بیماری غیر منتظره با استفاده بازدیدکنندگان روزانه از بخش اورژانس" انجام شده، همخوانی دارد. علاوه بر این مطالعه‌های (۱۵-۱۸) دیگری با شناسایی الگوهای زمانی در زمینه‌های مختلف بر حضور الگوهای زمانی در

(Over smoothing) از پهنای باند پایین (۰/۳) استفاده شد. این پهنای باند محافظه کارانه هرگونه تورش را در هموارسازی رفع می کند. عملکرد مطلوب LOWESS در حذف الگوهای قابل توجیه در نتایج مطالعه نشان داد که می توان به این روش در نظام مراقبت به ویژه در جمعیت های کم و بیماری هایی با شیوع کم توجه بیشتری نمود. با دانش نویسنده از روش سری فوریه در هیچ مطالعه ای در زمینه هموارسازی داده ها استفاده نشده است. با توجه به آنالیز هارمونیک و توابع سینوسی و کسینوسی و وجود این توابع در الگوهای فصلی به کارگرفته شد که این روش نیز عملکرد نسبتاً مطلوبی در حذف الگوهای قابل توجیه داشت.

در این مطالعه تاکید بر روی ویژگی زمانی داده ها بود و کلیه ابزارها و روش ها در این مطالعه به منظور تحلیل و بررسی تک متغیره استفاده شده است ولی استفاده از خصوصیت مکانی-زمانی می تواند در پیش پردازش داده ها نیز موثر باشد. لذا، امید است در مطالعات آینده بر روی متغیرهای مکانی و مکانی-زمانی با روش های چند متغیره در پیش پردازش داده ها استفاده گردد.

با توجه به مصداق نداشتن مفروضات به کارگیری الگوریتم ها شامل مستقل بودن و داشتن توزیع یکسان داده ها در داده های نظام مراقبت بیماری مننژیت، اهمیت پیش پردازش در شناسایی طغیان های احتمالی این بیماری دارای اهمیت فراوانی است. نتایج این مطالعه نشان داد که آماره من-کندال در تحلیل روند و الگوی فصلی روش مناسبی است. به طور کلی روش LOWESS در حذف الگوهای قابل توجیه و روش سری فوریه در بهبود نرمال بودن داده ها عملکردی بهتری داشتند.

تقدیر و تشکر

نویسندگان مقاله از جناب آقای شجاعیان مسئول بیماری مننژیت در معاونت بهداشتی استان همدان و همچنین معاونت تحقیقات و فناوری به خاطر حمایت مالی طرح شماره ۹۲۰۴۰۴۱۰۷۲ تشکر می نمایند. این مقاله برگرفته از پایان نامه کارشناسی ارشد اپیدمیولوژی

دلیل می تواند تعداد موارد بیش از حد صفر موارد روزانه در داده ها باشد. سومین دلیل می تواند از تاخیر در ارسال گزارش موارد و عدم گزارش دهی به موقع باشد. به نوعی این سه دلیل از مهم ترین محدودیت های این مطالعه می باشند. همان طور که انتظار داشتیم نتایج حاصل نمودار چندجمله ای موضعی، اختلاف میانگین روزهای هفته و مقادیر همبستگی نشان دادند که اثرات روزهای هفته، پایان هفته و تعطیلات در داده ها وجود دارد. این اثرات می تواند به دلایلی نظیر کاهش نیروی انسانی در تعطیلات و آخر هفته، تمایل بیماران که مراجعه خود را به اوایل هفته موکول کنند و به هنگام نبودن ارسال گزارش ها باشد.

از نکات قوت مطالعه استفاده از روش آماره من-کندال در شناسایی الگوهای فصلی و روند می باشد. این آماره در مقابل خودهمبستگی و نرمال نبودن داده ها برای تشخیص روندها و الگوهای فصلی از قدرت بالایی برخوردار است (۱۹).

نتایج حاصل از نمودارهای هیستوگرام و نرمال (اشکال ۴ و ۵ ردیف اول) و شاخص های چولگی و کشیدگی (جدول ۳ و ۴) مطالعه حاضر نشان داد پیش شرط نرمال بودن در داده های مشکوک به مننژیت مهیا نیست. این یافته ها با یافته های مطالعات کرمی و همکارانش (۴) و لوتر و همکارانش (۱۰) همخوانی دارد. با این وجود پیش پردازش داده های مشکوک به مننژیت جهت استفاده در نمودار کنترل و الگوریتم ها تشخیص طغیان با اهمیت می باشد. به منظور پردازش داده ها از دو روش LOWESS و سری فوریه استفاده شد که بر اساس دانش نویسنده در داخل کشور اولین نمونه هموارسازی داده ها با این دو روش آماری است.

روش LOWESS روشی نو در حذف الگوهای زمانی داده هاست. این روش به ما امکان هموارسازی غیرخطی را فراهم می دهد. اساس آن یک هموارسازی چندجمله ای محلی می باشد. تجزیه و تحلیل با این روش آسان است و حجم نمونه کمی نیاز دارد که با توجه به حجم کم نمونه مطالعه حاضر روش مناسبی می باشد (۱۶). در این مطالعه جهت پرهیز از هموارسازی بیش از حد

effects of air pollutants and pollen types on asthma and COPD hospital emergency visits in the industrial and Mediterranean Spanish city of Cartagena. *Allergol Immunopathol.* 2012;40(4):231-7.

16. Hafen RP, Anderson DE, Cleveland WS, Maciejewski R, Ebert DS, Abusalah A, et al. Syndromic surveillance: STL for modeling, visualizing, and monitoring disease counts. *BMC Med Inform Decis Mak.* 2009;9(1):21.

17. Schwartz J, Spix C, Touloumi G, Bacharova L, Barumamdazadeh T, Le Tertre A, et al. Methodological issues in studies of air pollution and daily counts of deaths or hospital admissions. *J Epidemiol Community Health* 1996;50(Suppl 1):S3-11.

18. Stieb DM, Judek S, Burnett RT. Meta-analysis of time-series studies of air pollution and mortality: effects of gases and particles and the influence of cause of death, age, and season. *J Air Waste Manag Assoc.* 2002;52(4):470-84.

19. Hirsch RM, Slack JR. A nonparametric trend test for seasonal data with serial dependence. *Water Resour Res.* 1984;20(6):727-32.

دانشگاه علوم پزشکی همدان است.

منابع

1. Thigpen MC, Whitney CG, Messonnier NE, Zell ER, Lynfield R, Hadler JL, et al. Bacterial meningitis in the United States, 1998-2007. *N Engl J Med.* 2011;364(21):2016-25.
2. Clarke C, Mallonee S. State-based surveillance to determine trends in meningococcal disease. *Public Health Rep.* 2009;124(2):280-7.
3. Yahav I. A data analytical framework for improving real-time, decision support systems in healthcare [PhD thesis]. College Park: University of Maryland 2010.
4. Karami M, Soori H, Mehrabi Y, Haghdoost A, Gouya M. Detecting and removing the explainable patterns of the daily counts of suspected cases of measles as a prediagnostic data source in Iran. *IRJE.* 2012;8(3):12-21. Persian.
5. Esteghamati A, Asgari F, Godarzi N. *Guideline of Menengits Surveillance.* 4eds. Tehran: Chakame ava; 2004. Persian.
6. Karami M, Soori H, Mehrabi Y, Haghdoost AA, Gouya MM. Real time detection of a measles outbreak using the exponentially weighted moving average: does it work? *J Res Health Sci.* 2012;12(1):25-30.
7. Lotze TH, Shmueli G. How does improved forecasting benefit detection? An application to biosurveillance. *Int J Forecast.* 2009;25(3):467-83.
8. Shmueli G, Burkom H. Statistical challenges facing early outbreak detection in biosurveillance. *Technometrics.* 2010;52(1).
9. Borkowf CB, Albert PS, Abnet CC. Using lowess to remove systematic trends over time in predictor variables prior to logistic regression with quantile categories. *Stat Med.* 2003;22(9):1477-93.
10. Loze T, Murphy P, Shmueli G. Implementation and comparison of preprocessing methods for biosurveillance data. *ADS.* 2008;6(1).
11. Hamed KH, Ramachandra Rao A. A modified Mann-Kendall trend test for autocorrelated data. *J Hydrol.* 1998;204(1):182-96.
12. Brownstein JS, Kleinman KP, Mandl KD. Identifying pediatric age groups for influenza vaccination using a real-time regional surveillance system. *Am J Epidemiol.* 2005;162(7):686-93.
13. Fienberg SE, Shmueli G. Statistical issues and challenges associated with rapid detection of bio-terrorist attacks. *Stat Med.* 2005;24(4):513-29.
14. Mathes R, Ito K, Matte T. Assessing syndromic surveillance of cardiovascular outcomes from emergency department chief complaint data in New York City. *PLoS One.* 2011;6(2).
15. Cirera L, García-Marcos L, Giménez J, Moreno-Grau S, Tobías A, Pérez-Fernández V, et al. Daily

Seasonal trends and explainable patterns of meningitis: Data preprocessing on fever and neurological symptoms syndrome

Kabat Pazhouhi, MSc student in Epidemiology, Department of Biostatistics and Epidemiology, Hamadan University of Medical Sciences, Hamadan, Iran. pazhouhi.h@gmail.com.

***Manoochehr Karami**, PhD. Assistant Professor of Epidemiology, Modeling of Non-communicable Diseases Research Center and Department of Biostatistics & Epidemiology, Hamadan University of Medical Sciences, Hamadan, Iran (*Corresponding author). ma.karami@umsha.ac.ir

Nader Esmailnasab, PhD. Associate Professor of Epidemiology, Department of Epidemiology and Biostatistics, Kurdistan University of Medical Sciences, Kurdistan, Iran. esmailnasab@yahoo.com

Abbas Moghim Biygi, PhD. Associate Professor of Biostatistics, Modeling of Non-communicable Diseases Research Center and Department of Biostatistics and Epidemiology, Hamadan University of Medical Sciences, Hamadan, Iran. moghimb@yahoo.com

Mohammad Fariadras, MSc student in Epidemiology, Department of Biostatistics and Epidemiology, Hamadan University of Medical Sciences, Hamadan, Iran. feryadresmohammad@gmail.com.

Abstract

Background: Addressing and removing explainable patterns from syndromic data is required to timely detection of outbreaks. This study aims to detect and remove explainable patterns of fever and neurological symptoms syndrome as suspected meningitis which has been occurred in Hamadan province.

Methods: We analyzed data on reported cases of suspected meningitis in Hamadan province, between March 2010 and March 2013. Line, local polynomial and moving averages (MA) charts, autocorrelation and partial autocorrelation functions, mean differences and nonparametric Mann- Kendall statistics were used to identify explainable patterns in the data on suspected cases of meningitis. Fourier series and LOWESS (Locally Weight Regression and Smoothing Scoter plot) was used to remove such patterns.

Results: Local polynomial charts, autocorrelation and partial autocorrelation functions, mean differences and nonparametric Mann- Kendall statistics indicated the presence of explainable patterns include Day-of- Week (DOW), weekend, holiday effects, seasonality and temporal trend in the syndromic data of fever and neurological symptoms. Overall, LOWESS in removing explainable patterns and Fourier series in improve of normality data have had better performance.

Conclusion: Results showed the presence of explainable patterns in the data of suspected cases of meningitis. Accordingly, timely and accurately detection of meningitis' outbreak required data smoothing.

Keywords: Meningitis, Outbreak, Seasonal patterns, Smoothing, Surveillance system.