Navid No 27-90 (2024) 24-33
https://doi.org/10.22038/nnj.2024.80983.1455

Navid No

Journal homepage: https://nnj.mums.ac.ir/

Original Article

Utilization of Fused Lasso Approach in Distributed Lag Models for Estimating the Adjusted Fatality Rate of Covid-19 in Iran

Farideh Khosravi [1], Azadeh Saki [2], Hamed Tabesh [3*]

1. PhD Candidate in Biostatistics, Department of Epidemiology and Biostatistics, School of Health, Mashhad University of Medical Sciences, Mashhad, Iran.
2. Associate Professor of Biostatistics, Department of Epidemiology and Biostatistics, School of Health, Mashhad University of Medical Sciences, Mashhad, Iran.
3. Associate Professor of Biostatistics, Department of Medical Informatics, School of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran.

Corresponding author: tabeshh@mums.ac.ir
Received: 06 July 2024; Revised: 15 August 2024; Accepted: 09 September 2024

Abstract

Background and Aims: Distributed lag models can estimate disease fatality in long-term epidemics by adjusting the delayed effect of the number of daily cases. The fused lasso approach is a regularization method that considers the inherent order between features, such as the time order between the numbers of daily hospitalizations, and can remove the effect of collinearity between features by placing a penalty in the estimation of the parameters of the distributed lag model. Therefore, this study has been carried out by estimating the adjusted mortality of COVID-19 using the lasso regularization approach in distributed lag models.

Materials and Methods: The data relating to the COVID-19 disease (the number of deaths and the number of people identified daily) in Iran from the middle of February 1398 to the middle of May 1402 was prepared from the OurWorldinData database and using R4. 3.3 The distributed lag model was fitted with the fused lasso method.

Results: The adjusted fatality rate of COVID-19 in Iran is 1.72%, with five days delay between new cases diagnosis and deaths. The highest fractions were on the first day and the fifth day after detection.

Conclusion: Using the fused lasso method improved the significance of the coefficients in the distributed lag model. The short time from detection to death in Iran compared to other countries is a sign of poor screening and a lack of adequate equipment and services for the emergency treatment of patients.

Keywords: Fused lasso, distributed lag model, Covid-19, Fatality Rate

Cite this article as: Khosravi F, Saki A, Tabesh H. Utilization of Fused Lasso Approach in Distributed Lag Models for Estimating the Adjusted Fatality Rate of Covid-19 in Iran. Navid No, 2024; 27(90): 24-33. https://10.22038/nnj.2024.80983.1455.

E-ISSN: 2645-5927 / P-ISSN: 2645-5919
Copyright: © 2024 by the author.

Open Access: This is an open- access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Publisher's Note: Mashhad University of Medical Sciences remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

نوید نو، دوره ۲۷، شماره ۹۰، تابستان ۱۴۰۳، صفحه ۳۳-۲۴
https://doi.org/10.22038/nnj.2024.80983.1455

نوید نو
Navid No
Journal homepage: https://nnj.mums.ac.ir/

Mashhad University of Medical Sciences

کمیته تحقیقات دانشجویی معاونت پژوهش و فناوری دانشگاه علوم پزشکی مشهد

مقاله پژوهشی

برآورد میزان کشندگی تعدیل یافته کووید-۱۹ در ایران، با رهیافت لاسو تلفیقی در مدل های با وقفه توزیعی

فریده خسروی[۱]، آزاده ساکی[۲]، حامد تابش[۳*]

۱. دانشجوی دکترای آمارزیستی، گروه اپیدمیولوژی و آمارزیستی، دانشکده بهداشت، دانشگاه علوم پزشکی مشهد، مشهد، ایران.
۲. دانشیار آمارزیستی، گروه اپیدمیولوژی و آمارزیستی، دانشکده بهداشت، دانشگاه علوم پزشکی مشهد، مشهد، ایران.
۳. دانشیار آمارزیستی، گروه انفورماتیک پزشکی، دانشکده پزشکی، دانشگاه علوم پزشکی مشهد، مشهد، ایران.
پست الکترونیک نویسنده مسئول: tabeshh@mums.ac.ir

تاریخ دریافت: ۱۴۰۳/۰۴/۱۶، تاریخ بازنگری: ۱۴۰۳/۰۵/۲۵، تاریخ پذیرش: ۱۴۰۳/۰۶/۱۹

چکیده

مقدمه و هدف: مدل‌های وقفه توزیعی قابلیت برآورد میزان کشندگی یک بیماری در اپیدمی های طولانی مدت را با تعدیل اثر تاخیری تعداد موارد روزانه ابتلا دارا می باشند. رهیافت لاسو تلفیقی یک روش انقباضی است که وجود ترتیب ذاتی بین ویژگی‌ها، همانند ترتیب زمانی بین تعداد بستری‌های روزانه را در نظر می‌گیرد و می تواند اثر همخطی بین ویژگی‌ها را با قرار دادن جریمه در برآورد پارامترهای مدل وقفه توزیعی برطرف نماید. لذا این مطالعه با برآورد مرگ و میر تعدیل شده کووید-۱۹ با استفاده از رهیافت انقباضی لاسو تلفیقی در مدل‌های وقفه توزیعی صورت گرفته است.

مواد و روش‌ها: داده های مربوط به بیماری کووید-۱۹ (تعداد مرگ و میر و تعداد افراد شناسایی به صورت روزانه) در کشور ایران از اول اسفند ماه ۱۳۹۸ تا۲۱ دی ماه ۱۴۰۱ از پایگاه داده *OurWorldinData* تهیه شده و با استفاده از *R4.3.3* مدل وقفه توزیعی با روش لاسو تلفیقی برازش داده شد.

یافته‌ها: میزان کشندگی تعدیل شده کووید-۱۹ در ایران ۱,۷۲ درصد می باشد تعداد موارد ابتلا با وقفه ۵ روزه بر تعداد مرگ تاثیر گذار بود، که بیشترین سهم مربوط به روز اول و روز پنجم پس از شناسایی بود.

نتیجه‌گیری: : به کارگیری روش لاسو تلفیقی باعث بهبود معنی داری ضرایب در مدل وقفه توزیعی شد. زمان کوتاه شناسایی تا مرگ در ایران نسبت به سایر کشورها نشان از غربالگری ضعیف و عدم تجهیزات و خدمات کافی برای رسیدگی اورژانس به مبتلایان بوده است.

کلمات کلیدی
لاسو تلفیقی، مدل وقفه توزیعی، کووید-۱۹، میزان کشندگی

مقدمه

در سال‌های اخیر با ثبت منظم داده‌ها و به وجود آمدن بانک‌های مختلف اطلاعاتی در سراسر دنیا، تمایل به استفاده ازمدل‌های پیشگویی‌کننده افزایش یافته است. هدف اصلی در این مدل‌ها، برازش یک مدل پیشگویی‌کننده آماری برای داده‌های وابسته به زمان براساس اطلاعات گذشته است. امروزه پژوهشگران از کاربردهای مدل‌های رگرسیونی و سری‌های زمانی به طور وسیع در خدمات بهداشتی و درمانی برای پیش‌بینی طغیان بیماری‌ها، تعداد بیماران مراجعه‌کننده، تعداد پرسنل موردنیاز در بخش‌های مختلف درمانی، استفاده می‌کنند (۱). ازآن جایی که یک متغیر سری زمانی می تواند متغیر دیگری را با وقفه زمانی تحت تاثیر قرار دهد، تحلیل گری که از داده های سری زمانی استفاده می‌کند با مشکل اساسی روبرو می شود (۲) . اثر تاخیری بیانگر آن است که تغییرات Y در زمان حال، وابسته به تغییر X در زمان حال و گذشته است

یکی از مشکلات مهم و اساسی در این نوع داده ها وجود هم خطی یا به عبارتی وجود ترکیب خطی متغیری از یک یا چند متغیر دیگر است. از آسیب های مدل در هنگام استفاده از این داده‌ها می توان به عدم پایداری، دقت پیش بینی کم و انتخاب نادرست متغیرها، اشاره نمود. به علاوه این مشکلات زمانی که همبستگی بین متغیرهای پیش بین زیاد باشد، تشدید نیز می شوند.

راهکار اولیه مدل‌سازی این الگوی تأخیری، می‌تواند استفاده از یک مدل رگرسیون خطی با ایجاد متغیرهای مستقل از تأخیرهای سریالی در متغیر پیشگو باشد. از آنجایی که این تأخیرها می‌توانند به‌طور سریالی همبستگی بالایی داشته باشند، استفاده از رگرسیون معمولی مستعد التهاب واریانس به دلیل کاهش درجه آزادی و همچنین مشکل هم‌خطی چندگانه است. یک راه‌حل جایگزین برای این مسئله استفاده از مدل‌های با وقفه توزیعی (۳) است که برای اولین بار توسط آلمون در سال ۱۹۶۵ معرفی شد (۴). این مدل‌ها در دهه‌های بعد به‌طور گسترده در زمینه‌های مختلف از جمله مطالعات محیطی و اپیدمیولوژی مورد استفاده قرار گرفتند و در طی سال‌های بعد برای غلبه بر محدودیت‌های مدل اولیه توسعه داده شده‌اند. به عنوان مثال در بررسی پیامدهای سلامت شووارتز در سال ۲۰۰۰ برای اولین بار در مطالعه‌ای وقفه توزیعی بین آلودگی هوا و مرگ‌ومیر روزانه را بررسی کرد و نشان داد که تأثیر میزان آلاینده‌های هوا روی میزان مرگ با وقفه زمانی ۰ تا ۵ روز همراه است و این اثر بعد از ۵ روز به صفر می‌رسد(۳). در زمینه پاندمی کووید -۱۹ نیز رستاقی و همکاران در سال ۲۰۲۱ با استفاده از مدل‌ها وقفه توزیعی مختلف، میزان مرگ‌ومیر بیمارستانی کووید- ۱۹را با توجه به اثر تأخیری تعداد بستری تخمین زدند. نتایج حاکی تأثیر معنی‌دار میزان بستری بر مرگ‌ومیر بیمارستانی با وقفه ۵ الی ۷ روز است (۵). همچنین در بررسی دیگر بر روی ۳۹ کشور اروپایی در زمینه این بیماری، برآورد میانگین وقفه‌ها و نرخ مرگ و میر تعدیل‌شده با مدل‌های وقفه توزیع‌شده و توزیع فضایی نشان داد که سیاست‌های پیشگیری و پزشکی پس از اولین پیک و همچنین واکسیناسیون در کاهش میزان مرگ و میر کووید- ۱۹ مؤثر است، اما این تأثیرات بین کشورها متفاوت است (۶).

گاسپارینی و همکاران در سال ۲۰۱۰ این مدل‌ها را توسعه دادند و مدل‌های با وقفه توزیعی غیرخطی را به عنوان یک جایگزین انعطاف‌پذیر به منظور توصیف روابط غیرخطی معرفی کردند(۷) . روش های انقباضی به عنوان راهکاری برای کاهش هم خطی، مورد توجه قرار گرفته اند. این روش ها ضرایب رگرسیونی را با اعمال محدودیت روی دامنه تغییرات آن ها برآورد می کنند. اگرچه وجود چنین محدودیت هایی واریانس برآوردگر را کاهش می دهد ولی مقداری اریبی ایجاد می کند، به طوری که می توان امیدوار بود در نهایت میانگین توان دوم خطا کاهش یابد (۸). از جمله برآوردگرهای انقباضی رایج در برآورد پارامترهای مدل رگرسیونی می توان به برآوردگر ریج اشاره نمود که توسط هورل و کنارد معرفی شد. رگرسیون ریج، مقدمه ای برای راهیابی به برآورد و انتخاب متغیر است. این رگرسیون، با

مشکل هم خطی در مدل های خطی مبارزه می کند (۹). با توجه به حضور همه متغیرها در برآوردگر ریج، تفسیرپذیری آن به سادگی امکان پذیر نیست. عضو دیگری از این کلاس، برآوردگر لاسو است. در واقع، برآوردگر لاسو، هم‌زمان هم انتخاب متغیر انجام می‌دهد و هم ضرایب را منقبض می‌کند. هدف رگرسیون لاسو شناسایی متغیرها و ضرایب رگرسیون مربوطه است که منجر به مدلی می شود که خطای پیش بینی را به حداقل می‌رساند . این با تحمیل یک محدودیت بر پارامترهای مدل، که ضرایب رگرسیون را به سمت صفر "کوچک می‌کند" به دست می‌آید، یعنی با مجبور کردن مجموع قدر مطلق ضرایب رگرسیون کمتر از مقدار ثابت (λ) است (۱۰).

از کاربرد های مهم برآوردگرهای لاسو می توان در مدل های تنک (مدل هایی که تعداد پارامترهای صفر آن زیاد باشد) و همچنین هنگامی که بعد فضای پارامتر بیشتر از بعد فضای نمونه باشد، اشاره نمود. به دلیل وجود تعداد زیادی از متغیرها در مدل های با بعد بالا، تفسیر این مدل ها بسیار مشکل است. لذا مسئله انتخاب متغیر نقش بسیار مهمی را در مدل سازی آماری با بعد بالا ایفا می کند. علی رغم مزایای برآوردگر لاسو، عملکرد روش لاسو به عنوان روشی برای انتخاب مدل بهینه در حالتی که مشاهدات شامل داده دور افتاده هستند و توزیع متغیر خطا نرمال در نظر گرفته می شود، ضعیف است (۱۱).

از دیگر کاربردهای این برآوردگرها هنگامی است که مجموعه داده های مطالعات مشابه در عمل برای دستیابی به حجم نمونه بزرگتر و توان بالاتر ترکیب می شوند. چالش اصلی ناشی از ادغام داده ها مربوط به ناهمگونی داده ها از نظر جمعیت مطالعه، طراحی مطالعه یا هماهنگی مطالعه است (۱۲). ناهمگونی بین مطالعه می تواند ناشی از تفاوت در محیط مطالعه، جمعیت، طراحی و پروتکل ها باشد (۱۳, ۱۴). نادیده گرفتن چنین ناهمگونی در تجزیه و تحلیل داده ها ممکن است منجر به برآورد مغرضانه و استنتاج گمراه کننده شود .به طور مثال این اتفاق در مطالعات کارآزمایی بالینی چندمرکزی که داده ها از سایت های گوناگون ترکیب می شوند، رخ می دهد (۱۵).

تکنیک‌های سنتی درمان ناهمگونی داده‌ها شامل استفاده از اثرات متقابل و اثرات تصادفی است که به ویژه هنگام ترکیب مجموعه‌های کوچکتر، از اعتبار کمتری نسبت به دستیابی به قدرت آماری مطلوب یا ارائه یک تفسیر معنادار، برخوردار هستند. مدل به‌دست‌آمده از مطالعات ترکیبی ممکن است به عنوان یک مدل پیش‌بینی مناسب برای هر مطالعه جداگانه، در مورد جمعیت‌های ناهمگن عمل نکند. یک روش خوشه‌بندی تجمعی برای ضرایب رگرسیون در زمینه یکپارچه‌سازی داده‌ها پیشنهاد شده است که به عنوان رویکرد لاسو تلفیقی خوشه‌بندی ضرایب رگرسیون نامیده می‌شودکه برای شناسایی الگوهای ناهمگونی ضرایب رگرسیون در مطالعات (یا مجموعه داده‌ها) و ارائه تخمینی از همه ضرایب رگرسیون به طور همزمان است. هدف از این مطالعه، پیش بینی مرگ و میر ناشی از ویروس کووید۱۹ بر اساس تعدیل مرگ و میرهای رخ داده با در نظر گرفتن یک وقفه ۷ روزه در داده های مربوط به بیماران کووید ۱۹ در کشور ایران با استفاده از روش لاسو تلفیقی است.

روش کار

داده ها:

جهت کاربرد در داده های سلامت، داده های مربوط به بیماری کووید۱۹ (تعداد مرگ و میر و تعداد افراد شناسایی به صورت روزانه) در کشور ایران از اول اسفند ۱۳۹۸ تا۲۱ دی ماه ۱۴۰۱ از پایگاه داده OurWorldinData که اطلاعات مربوط به همه‌گیری کووید ۱۹ را به تفکیک کشورها در دسترس کاربران قرار می‌دهد، تهیه شده است. پس از نهایی شدن و دریافت داده ها فرآیند پیش پردازش داده ها شامل پاکسازی داده ها و آماده سازی جهت استفاده در مدلسازی انجام شده است. در داده های کووید-۱۹ واحد زمانی روز است و وقفه توزیعی بین شناسایی و مرگ برحسب روز گزارش شده است. در مرحله بعد با فراخوانی

داده ها در نرم افزار R تحلیل های توصیفی برای بررسی داده ها انجام شد. با استفاده از نمودار پراکنش خطی بودن ارتباط تعداد افراد شناسایی و مرگ و میر روزانه بررسی شد و مدل وقفه توزیعی با دو رهیافت حداقل مربعات خطا و لاسو تلفیقی برازش داده شد.

روش تحلیل آماری:

شکل کلی یک مدل با اثر تاخیری به صورت زیر است:

$$Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \cdots + \beta_k X_{t-k} + u_t$$

که بر اساس آن می توان گفت اثرX بر Y در زمان t می تواند از k زمان قبل وجود داشته باشد. یا در حقیقت تعداد موارد مرگ در روز t به تعداد موارد ابتلای شناسایی شده در روزهای قبل بستگی دارد. ضریب هر روز نشان دهنده میزان مرگ در آن روز از کل مرگهای رخ داده تا زمان t می باشد و مجموع ضرایب میزان مرگ (کشندگی) تعدیل شده را برآورد می کند.

روش لاسو تلفیقی:

در روش لاسو با استفاده از یک تابع جریمه روی جمع قدرمطلق ضرایب مدل رگرسیونی، تعداد پارامترها کنترل می‌شود. در این حالت، مجموع مربعات خطای رگرسیونی لاسو به صورت زیر نوشته می‌شود:

$$\sum_{i=1}^{N}(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\,\beta_j)^2 + \lambda \sum_{j} |\beta_{ij}|$$

λ به عنوان پارامتر تنظیم کننده است که اگر مقدارش برابر با صفر باشد، مدل به رگرسیون عادی تبدیل شده و همه متغیرها در آن حضور خواهند داشت و اگر مقدار آن افزایش یابد تعداد متغیرهای مستقل در مدل کاهش خواهند یافت. تعیین مقدار برای این پارامتر معمولا توسط روش اعتبارسنجی متقاطع انجام می‌شود.

یک اشکال عمده لاسو این است که اگر بین زیرمجموعه‌ای از پیش‌بینی‌کننده‌ها هم‌خطی وجود داشته باشد، معمولاً تنها یکی را برای نشان دادن کل گروه خطی انتخاب می‌کند.

علاوه بر این، لاسو نمی تواند بیش از n متغیر را زمانی که $p > n$ انتخاب کند (۱۶).

جهت رفع این مشکل می توان از روش لاسو تلفیقی استفاده کنیم.در این روش که در واقع می توان گفت تعمیم روش لاسو است، نوعی متفاوت از ساختار بین ویژگی ها، وجود رتبه بندی ذاتی است. به طور مثال می توان موقعیت های طیف سنجی و داده های زمانی و مکانی را در نظر بگیریم و داریم:

$$\hat{\beta} = argmin \sum_{i}(y_i - \sum_{j} x_{ij}\beta_j)^2$$

$$\sum_{j=1}^{p}|\beta_j| \le S_1,\ \sum_{j=2}^{p}|\beta_j - \beta_{j-1}| \le S_2$$

محدودیت در نظر گرفته شده ضرایب $\boldsymbol{\beta_j}$ را به عنوان تابعی از j در نظر می گیرد. و ضرائبی برآورد می شود که تابع هدف را حداقل کند (۱۷). جریمه شامل دو قسمت است.

الف) جریمه لاسو که در آن $\boldsymbol{\beta_j = 0}$

ب) جریمه ترکیبی که باعث می شود $\boldsymbol{\beta_j}$ برابر با $\boldsymbol{\beta_{j+1}}$ یا $\boldsymbol{\beta_{j-1}}$شود.

در برخی از مطالعات اپیدمیولوژی، زیست محیطی و اقتصاد سلامت اثر مداخله‌ها، مواجهه‌ها و پیشامدها بر روی پیامدهای سلامت مانند بیماری، ناتوانی، بستری و مرگ با وقفه زمانی همراه است. افزایش چشمگیر در داده های جمع آوری شده در بسیاری از زمینه ها خصوصا بیماری کووید ۱۹ با افزایش متناظر در تعداد پیش بینی کننده های موجود در تجزیه و تحلیل داده ها همراه است ، برای تفسیر ساده‌تر فرآیندهای زیربنایی که داده‌ها را تولید می‌کنند، اغلب مایل به داشتن یک مدل نسبتاً صرفه‌جو هستیم.گسترش سریع ویروس کووید ۱۹ از چین به سایر کشورها و شیوع بیماری نیازمند تجزیه و تحلیل اپیدمیولوژیک این بیماری در کوتاه ترین زمان و افزایش آگاهی از مداخلات موثر است. به کارگیری روش ها و تحلیل مناسب داده‌های ثبت شده از مرگ مبتلایان به کووید-۱۹ می‌تواند اطلاعات جامعی برای تصمیم گیری‌های پیشگیرانه

و حفظ سلامت عمومی فراهم آورده و در مدیریت بهتر این اپیدمی کمک شایانی کند.

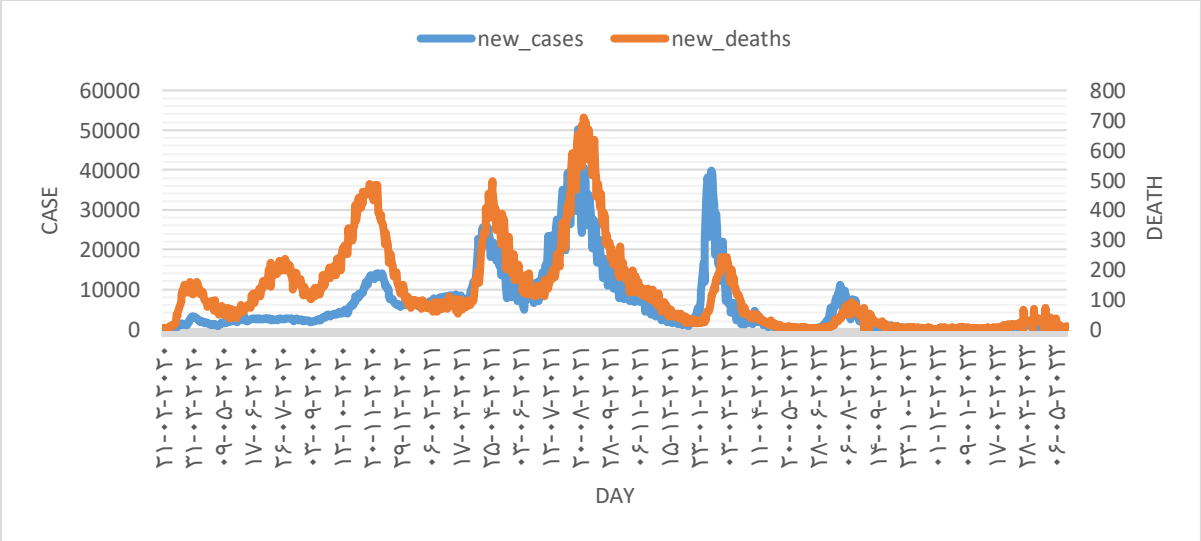
یافته‌ها

داده های مربوط به تعداد موارد جدید و فوت روزانه بر اثر بیماری کووید-۱۹ در ایران از ۲ فوریه سال ۲۰۲۰ (۱۳ بهمن ۱۳۹۸) تا ۲ می ماه ۲۰۲۳ (۱۲ اردیبهشت ۱۴۰۲) است. در مجموع تعداد موارد جدید و فوت ۷۶۱۱۱۳۸ و ۱۴۶۲۳۰ است. در واقع به ازای هر ۵۲ نفر جدید شناسایی شده یک فوت اتفاق افتاده است. میانگین تعداد موارد جدید و فوت به ترتیب در این مدت ۶۴۱۷/۱۸ و ۱۲۳/۳ بوده است. بیشترین مورد جدید شناسایی شده در ۱۹ آگوست ۲۰۲۱(۲۸ مرداد ماه ۱۴۰۰) و بیشترین فوت روزانه در ۲۱ آگوست ۲۰۲۱(۴شهریورماه ۱۴۰۰) رخ داده است. جدول ۱ به بررسی آمارتوصیفی میزان موارد شناسایی شده و مرگ بر اثر کووید-۱۹ در ایران در مدت زمان بیان شده است.

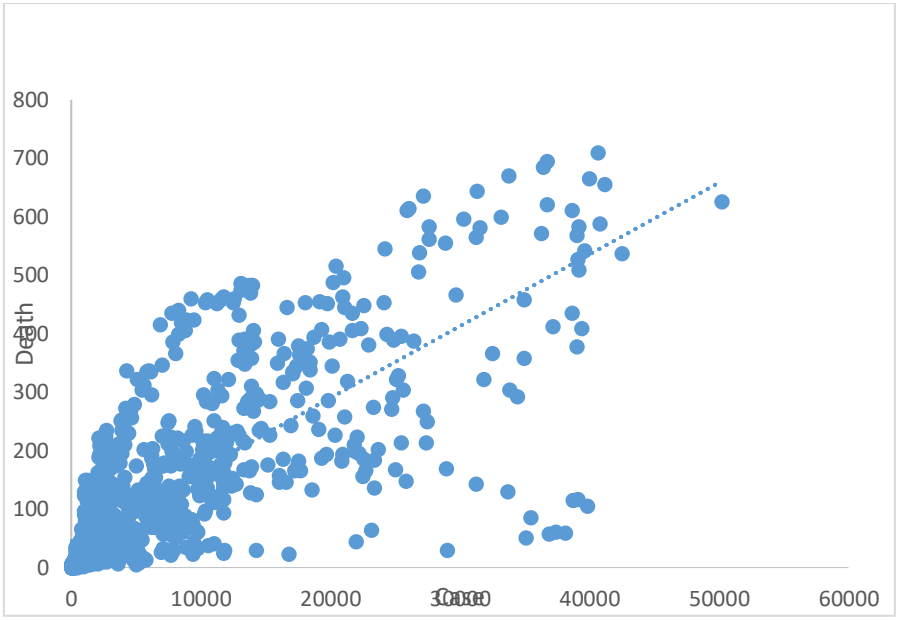
نمودار۱ روند تعداد بستری و مرگ روزانه در اثر ابتلا به کووید-۱۹ در ایران از اول اسفند ماه ۱۳۹۸ الی ۱۲ اردیبهشت ۱۴۰۲ را نشان می‌دهد. بر این اساس میزان مرگ و ابتلا در شهریورماه ۱۴۰۰ همزمان با پیک پنجم به اوج خود رسیده است. با توجه به اینکه بعد از پیک ششم کووید-۱۹ از حالت اضطرار و اپیدمی خارج شده و ثبت موارد ابتلا و مرگ با دقت گذشته انجام نشده است، مطالعه حاضر بر روی اطلاعات ابتلا و مرگ تا پایان پیک ششم یعنی ۲۱ دی ماه ۱۴۰۱ انجام شده است. نمودار۲ پراکنش بین میزان ابتلا و مرگ روزانه را نشان می دهد که نسبتا خطی است.

جدول۱: آمارتوصیفی میزان موارد شناسایی شده و مرگ بر اثر کووید-۱۹ در ایران از اول اسفند ۱۳۹۸ الی ۲۱ دی ۱۴۰۱

متغیر	حداقل	حداکثر	میانگین	انحراف معیار	چارک اول	چارک دوم	چارک سوم
موارد جدید	۰	۵۰۲۲۸	۹۱۲۳/۷۲	۸۹۴۵/۸۵	۲۴۶۸	۶۴۵۹	۱۱۷۵۵
مرگ	۰	۷۰۹	۱۹۳/۸۴	۱۴۳/۹۶	۸۹	۱۴۶	۲۴۱



نمودار ۱: سری زمانی تعداد ابتلا و مرگ روزانه در ایران



نمودار۲: ارتباط بین میزان ابتلا و مرگ روزانه در ایران

جدول۲: نتایج حاصل از برازش مدل خطی وقفه توزیعی با روش حداقل مربعات در داده‌های کووید-۱۹ ایران

تعداد موارد ابتلا	برآورد	خطای معیار برآورد	t-مقدار	p-مقدار
همان روز	۰/۰۰۵۱	۰/۰۰۲۱	۲/۳۶۹	۰/۰۱۸
یک روز قبل	۰/۰۰۰۲	۰/۰۰۲۸	۰/۰۶۱	۰/۹۵۱
دو روز قبل	۰/۰۰۰۹	۰/۰۰۲۸	۰/۳۲۰	۰/۷۴۹
سه روز قبل	۰/۰۰۱۷	۰/۰۰۲۸	۰/۰۵۹۷	۰/۵۵۰
چهار روز قبل	۰/۰۰۱۷	۰/۰۰۲۸	۰/۶۰۹	۰/۵۴۲
پنج روز قبل	۰/۰۰۷۷	۰/۰۰۲۲	۳/۵۹۹	۰/۰۰۰۳

*معناداری در سطح ۰/۰۰۱

جدول ۲ نتایج برازش مدل وقفه توزیعی خطی با رهیافت حداقل مربعات خطا را نشان می دهد. همانطور که در این جدول دیده می شود برخی ضرایب عنی دار نشده اند که به علت همخطی بین تعداد مبتلایان روزانه و التهاب خطای معیار برآوردهاست و غیر قابل قبول می باشد. جدول ۳ نتایج حاصل از به کارگیری روش لاسو تلفیقی در مدل وقفه توزیعی را نشان می‌دهد. با این رهیافت تمامی ضرایب معنی دار شدند و بیشترین سهم اثر ابتلا بر مرگ مربوط به روز اول و پنجم می باشد. و مجموع ضرایب که میزان کشندگی تعدیل شده را برآورد کرده است برابر ۱,۷٪ می شود. هر چند که هر دو رهیافت برآوردهای مشابهی برای اثر تاخیری داده اند اما از لحاظ مهنی داری ضرایب رهیافت لاسو تلفیقی عملکرد بهتری داشته است.

جدول۳: نتایج حاصل از برازش مدل خطی وقفه توزیعی با روش لاسو تلفیقی در داده‌های کووید-۱۹ ایران

تعداد موارد ابتلا	برآورد	خطای معیار برآورد	t-مقدار	p-مقدار
همان روز	۰/۰۰۵۱	۹/۹۴E-۰۵	۵۱/۳۲۹۷	<۰/۰۰۰۱
یک روز قبل	۰/۰۰۰۲	۰/۰۰۰۱۱۱	۱/۸۰۹۴	۰/۰۳۵۲
دو روز قبل	۰/۰۰۰۹	۰/۰۰۰۱۰۲	۸/۸۶۴۲	<۰/۰۰۰۱
سه روز قبل	۰/۰۰۱۶	۰/۰۰۰۱۰۲	۱۵/۶۱۵۸	<۰/۰۰۰۱
چهار روز قبل	۰/۰۰۱۸	۹/۲۵E-۰۵	۱۹/۴۵۶۴	<۰/۰۰۰۱
پنج روز قبل	۰/۰۰۷۷	۸/۸۳E-۰۵	۸۷/۲۴۶۱	<۰/۰۰۰۱

بحث

توزیع اپیدمیولوژیک بیماری کووید-۱۹ از جمله فواصل زمانی از شروع علائم تا تشخیص و گزارش مرگ، جهت توسعه راهبردهای موثر کنترل بیماری مهم است. بر اساس نتایج به نظر ارتباط تعداد موارد جدید و مرگ روزانه بیماری کووید۱۹ خطی است که با مطالعه رستاقی و همکاران در سال ۲۰۲۰ در ایران مطابقت دارد. در این مطالعه با استفاده از مدل‌های خطی با وقفه توزیعی مشخص شد که بین میزان بستری و مرگ روزانه کووید۱۹- ارتباط خطی معنی‌داری و جود دارد (۵). همچنین در مطالعه هادیان‌فر و همکاران در سال ۱۳۹۹ بیان شد نتایج حاصل از مدل غیرخطی با وقفه توزیعی نشان می‌دهد با افزایش میزان بستری روزانه میزان خطر نسبی مرگ در همان روز و روزهای بعد به طور معناداری افزایش می‌یابد به طوری که با عبور میزان بستری ها از ۲۰۰۰ نفر در روز خطر نسبی تجمعی مرگ بیش از یک می‌شود(۱۸). در مطالعه حاضر پس از برازش مدل وقفه توزیعی ضرایب غیر معنی دار به دست آمده است که می‌تواند به دلیل هم‌خطی موجود در متغیر مستقل یعنی تعداد موارد جدید بیماری در روزهای متوالی باشد. جهت رفع این مشکل از روش‌های انقباضی استفاده شده است. همان طور که اشاره شد کنترل پارامترها در روش لاسو با استفاده از یک تابع جریمه روی جمع قدرمطلق ضرایب مدل رگرسیونی انجام می‌شود. مشکل اساسی هنگامی به وجود می‌آید که بین زیرمجموعه‌ای از پیش‌بینی‌کننده‌ها هم‌خطی وجود داشته باشد که برای رفع آن از روش لاسو تلفیقی استفاده می‌شود.. از کاربرد روش لاسو تلفیقی می‌توان به مطالعه Jung و همکاران در سال ۲۰۲۱ جهت شناسایی دقیق ناحیه تومور اشاره نمود. آن‌ها بیان نمودند اتخاذ روش لاسو تلفیقی می‌تواند تقسیم‌بندی نزدیک به آن‌چه توسط متخصصان تولید می‌شود، بدون افزایش قابل توجهی در زمان محاسبات ایجاد کند (۱۹). علاوه بر این، روش لاسو تلفیقی که در واقع تعمیم روش لاسو است، نوعی متفاوت از ساختار بین ویژگی ها، یعنی وجود رتبه بندی ذاتی در داده‌ها را در نظر می‌گیرد. می‌توان گفت داده‌های سری زمانی استفاده شده در این مطالعه از این دسته است. ضرایب به دست آمده از روش لاسو تلفیقی نشان می‌دهد به ترتیب با وقفه ۵ و ۱ روزه بیشترین میزان مرگ روزانه مشاهده شده است. جمع ضرایب به دست آمده از مدل برابر با ۰/۰۱۷۲ است. در مطالعه رستاقی و همکاران نیز میانگین تاثیر وقفه بستری روزانه در بیمارستان بر روی مرگ ۵ روز گزارش شده است که با مطالعه حاضر هم‌خوانی دارد (۵). در مطالعه ای دیگر توسط Guglielmi و همکاران در سال ۲۰۲۳ توزیع وقفه بین ۱۱ تا ۱۳ روز برای کشورهای ایتالیا و اتریش راگزارش کردند (۲۰). پژوهشی در برزیل وقفه بیش از ۸ روز در تشخیص بیماری کووید۱۹ را از عوامل افزایش مرگ بیماران بیان نمود (۲۱). همچنین در مطالعه هادیان‌فر و همکاران میانگین زمان وقفه بیماری کووید - ۱۹ در کشورهای اروپایی در نیمه اول۲۰۲۰ برابر ۲/۹ به دست آمده است (۶).

نتیجه‌گیری

بر اساس مطالعات زمان شناسایی تا مرگ بر اثر کرونا متفاوت است و بستگی به عوامل مختلفی از جمله وضعیت سلامتی فرد، نحوه درمان، واکسیناسیون، و سرعت تشخیص بیماری دارد. شاید بتوان گفت زمان کوتاه‌تر شناسایی تا

مرگ در ایران نسبت به سایر کشورها نشان از غربالگری ضعیف و عدم تجهیزات و خدمات کافی برای رسیدگی اورژانسی به مبتلایان بوده است. البته شایان ذکر است محیط‌های جغرافیایی و زمینه‌های ژنتیکی هم در میزان ابتلا و مرگ ناشی از این بیماری تاثیر دارد.

تشکر و قدردانی

این مقاله حاصل از طرح تحقیقاتی با کد و در معاونت پژوهشی دانشگاه علوم پزشکی مشهد می باشد و نویسندگان از همکاران محترم معاونت پژوهشی قدر دانی می نمایند.

حمایت مالی

این مطالعه منتج طرح تحقیقاتی مصوب دانشگاه علوم پزشکی بوده بوده و بودجه طرح توسط این دانشگاه تامین شده است. (کد طرح: ۴۰۱۲۲۹۶)

ملاحظات اخلاقی

این طرح در کمیته اخلاق دانشگاه علوم گزشکی نشهد با شناسه اخلاق IR.MUMS.FHMPM.REC.1402.070 تایید شده است.

تضاد منافع

بین نویسندگان هیچ‌گونه تعارض منافعی وجود ندارد

مراجع

[1]. Hasanzadeh J, Amiresmaili M, Moosazadeh M, Najafi F, Moradinazar M. Implementing a weather-based early warning system to prevent traffic accidents fatalities. World Appl Sci J. 2013;24(1):113-7.

[2]. Hasanzadeh J, Najafi F, Moradinazar M. How to choose an appropriate model for time series data. Iranian Journal of Epidemiology. 2015;11(1):94-102.

[3]. Schwartz J. The distributed lag between air pollution and daily deaths. Epidemiology. 2000;11(3):320-6.

[4]. Almon S. The distributed lag between capital appropriations and expenditures. Econometrica: Journal of the Econometric Society. 1965:178-96.

[5]. Rastaghi S, Shark NA, Saki A. Application of distribution-delay models to estimating the hospitalized mortality rate of covid-19 according to delay effect of hospitalizations counts. Journal of Biostatistics and Epidemiology. 2021.

[6]. Hadianfar A, Rastaghi S, Tabesh H, Saki A. Application of distributed lag models and spatial analysis for comparing the performance of the COVID-19 control decisions in European countries. Scientific Reports. 2023;13(1):17466.

[7]. Gasparrini A, Armstrong B, Kenward MG. Distributed lag non-linear models. Statistics in medicine. 2010;29(21):2224-34.

[8]. Shadrokh A, Khadembashiri Z, Yarmohammadi M. Regression Modeling Via T-Lasso Bayesian Method. Journal of Advanced Mathematical Modeling. 2021;11(2):365-81.

[9]. Hoerl AE, Kennard RW. Ridge regression: Biased estimation for nonorthogonal problems. Technometrics. 1970;12(1):55-67.

[10]. Ranstam J, Cook J. LASSO regression. Journal of British Surgery. 2018;105(10):1348-.

[11]. Lange KL, Little RJ, Taylor JM. Robust statistical modeling using the t distribution. Journal of the American Statistical Association. 1989;84(408):881-96.

[12]. Tang L, Song PX. Fused lasso approach in regression coefficients clustering: learning parameter heterogeneity in data integration. The Journal of Machine Learning Research. 2016;17(1):3915-37.

[13]. Leek JT, Storey JD. Capturing heterogeneity in gene expression studies by surrogate variable analysis. PLoS genetics. 2007;3(9):e161.

[14]. Liu D, Liu RY, Xie M. Multivariate meta-analysis of heterogeneous studies using only summary statistics: efficiency and robustness. Journal of the American Statistical Association. 2015;110(509):326-40.

[15]. Shekelle PG, Hardy ML, Morton SC, Maglione M, Mojica WA, Suttorp MJ, et al. Efficacy and safety of ephedra and ephedrine for weight loss and athletic performance: a meta-analysis. Jama. 2003;289(12):1537-45.

[16]. Jang W, Lim J, Lazar NA, Loh JM, Yu D. Some properties of generalized fused lasso and its applications to high dimensional data. Journal of the Korean Statistical Society. 2015;44(3):352-65.

[17]. Tibshirani R, Wang P. Spatial smoothing and hot spot detection for CGH data using the fused lasso. Biostatistics. 2008;9(1):18-29.

[18]. Hadianfar A, Rastaghi S, Saki A. Evaluation of the Relative Risk of Covid-19 Mortality Based on the Number of Hospitalizations in Iran using a Log-Linear Distributed Lag Model. Iranian Journal of Epidemiology. 2021;16(5):20-8.

[19]. Jung J, Roh J, Park C-S. Abstract PO-079: Fused LASSO application for gastric cancer image segmentation. Clinical Cancer Research. 2021;27(5_Supplement):PO-079-PO-.

[20]. Guglielmi N, Iacomini E, Viguerie A. Identification of time delays in COVID-19 data. Epidemiologic Methods. 2023;12(1):20220117.

[21]. Cobre AdF, Böger B, Fachi MM, Vilhena RdO, Domingos EL, Tonin FS, Pontarolo R. Risk factors associated with delay in diagnosis and mortality in patients with COVID-19 in the city of Rio de Janeiro, Brazil. Ciencia & saude coletiva. 2020;25:4131-40.