

مفهوم و کاربرد مدل رگرسیون Joinpoint در تحلیل روند زمانی مطالعات اپیدمیولوژیک (مورد: سرطان کبد ناشی از مصرف الکل در ایران)

سرطان کبد ناشی از مصرف الکل در ایران

محمدرضا بلوچ حسنخانی^۱، علی اکبر حق دوست^۲، یونس جهانی^۳

- ۱- کارشناسی ارشد آمار زیستی، پژوهشگر، مرکز تحقیقات مدل سازی در سلامت، پژوهشکده آینده پژوهی در سلامت، دانشگاه علوم پزشکی کرمان، کرمان، ایران
- ۲- دکترای اپیدمیولوژی، استاد، مرکز تحقیقات مراقبت HIV و عفونت های آمیزشی، پژوهشکده آینده پژوهی در سلامت، دانشگاه علوم پزشکی کرمان، کرمان، ایران
- ۳- دکترای آمار زیستی، استاد، مرکز تحقیقات مدل سازی در سلامت، پژوهشکده آینده پژوهی در سلامت، دانشگاه علوم پزشکی کرمان، کرمان، ایران

چکیده

اطلاعات مقاله

تاریخ دریافت

۱۴۰۲/۰۱/۲۶

تاریخ پذیرش

۱۴۰۲/۰۶/۱۲

نویسنده رابط

یونس جهانی

ایمیل نویسنده رابط

u.jahani@kmu.ac.ir

نشانی نویسنده رابط

کرمان، ابتدای محور هفت باغ علوی، پردیس دانشگاه علوم پزشکی کرمان، پژوهشکده ی آینده پژوهی در سلامت، مرکز تحقیقات مدل سازی در سلامت

واژگان کلیدی: تحلیل روند، رگرسیون

Joinpoint، الکل، سرطان کبد، میزان

مرگومیر استاندارد شده سنی

مقدمه و اهداف: تجزیه و تحلیل روند زمانی عواملی مانند میزان بیماری و مرگ از مولفه های مهم در برنامه ریزی های مربوط به سلامت هر جامعه است. تحلیل روند زمانی، این امکان را فراهم می کند تا تغییرات در طول زمان با دقت بیشتری تفسیر شوند. این مطالعه با هدف بررسی عملکرد رگرسیون Joinpoint جهت تحلیل روند انجام شد.

روش کار: در این مطالعه سعی شده است ابتدا با زبانی ساده، مفهوم مدل رگرسیون Joinpoint ارائه شود و سپس با اجرای این مدل بر روی داده های میزان مرگومیر ناشی از سرطان کبد به دلیل مصرف الکل طی ۳۰ سال در ایران کاربرد آن ارائه گردد.

یافته ها: نتایج تحلیل روند نشان داد، میزان مرگومیر استاندارد شده سنی ناشی از سرطان کبد به دلیل مصرف الکل طی ۳۰ سال در ایران (۱۹۹۰ تا ۲۰۱۹) به طور میانگین هر سال ۰/۸ درصد کاهش یافته است و پیش بینی نیز نشان داد این روند کاهشی ادامه دارد.

نتیجه گیری: به طور کلی مهمترین مزیت رگرسیون Joinpoint نسبت به سایر مدل ها، شناسایی زمان هایی است که تغییرات قابل توجهی در روندها رخ داده است. بنابر نتایج حاصل، میزان مرگومیر ناشی از سرطان کبد به دلیل مصرف الکل طی ۳۰ سال در ایران را می توان به پنج دوره با میزان تغییرات متفاوت تقسیم کرد.

مقدمه

تعیین روند و شناخت تغییرات میزان بیماری و مرگومیر دیدی پویا از وضعیت سلامت یک جامعه ارائه می دهد. انجام مطالعات طولی در بازه های زمانی طولانی مدت، با توجه به وجود بحران های مالی و اقتصادی، تغییرات گسترده محیطی، اجتماعی و سیاسی طی سال های مختلف اطلاعات ارزشمندی را برای برنامه ریزی تخصیص منابع و ارزیابی استراتژی های پیشگیری و مدیریت سلامت در اختیار سیاست گذاران سلامت قرار می دهد تا بتوانند جهت کنترل مداخلات علمی و موثری را طراحی و اجرا کنند (۱). معمولاً تجزیه و تحلیل روند زمانی برای اهداف متعددی مورد استفاده قرار می گیرد.

مهمترین اهداف تجزیه و تحلیل روند زمانی

۱. نظارت بر بهداشت عمومی و تشخیص اینکه آیا یک سطح وضعیت سلامت و شاخص خدمات با گذشت زمان افزایش یا کاهش یافته است و ارزیابی سرعت افزایش یا کاهش آن (۲).
۲. شناسایی روندهای گذشته و زمان هایی که تغییرات قابل توجهی در روندها رخ داده است، امکان برآورد روندها را با برون یابی آخرین روندها از مشاهدات را برای پیش بینی آینده فراهم می کند.
۳. زمینه مطالعات فراوان دیگری را برای محققین فراهم می کند و اطلاعات ارزشمندی را برای ساختن فرضیات علمی برای مطالعات بعدی فراهم می نماید.

باید مدل خطی انتخاب شود و وقتی که متغیر وابسته از توزیع نمایی یا توزیع پواسون پیروی می‌کند، باید مدل خطی لگاریتمی انتخاب شود.

فرض کنید برای مجموعه‌ی مشاهدات $(x_1, y_1), \dots, (x_n, y_n)$ ها نشان دهنده‌ی متغیر مستقل (زمان) به ترتیب صعودی، y ها نشانگر متغیر پاسخ (میزان‌های مربوط به هر زمان) و n نشان‌دهنده فاصله زمانی مطالعه است.

مدل خطی معادله رگرسیونی به صورت (۱) است:

$$E(y|x) = \beta_0 + \beta_1 x + \delta_1(x - \tau_1)^+ + \dots + \delta_k(x - \tau_k)^+ \\ = \beta_0 + \beta_1 x + \sum_{i=1}^k \delta_i(x - \tau_i)^+ \quad (1)$$

که در آن k تعداد نقاط شکست، τ_1, \dots, τ_k نشان دهنده‌ی نقاط شکست نامعلوم و β_0 پارامتر ثابت، β_1 ضریب رگرسیون و $\delta_1, \dots, \delta_k$ ضرایب رگرسیون قطعه k ام را نشان می‌دهند. مفهوم علامت بالا نویس مثبت این است که چنانچه عبارت داخل پرانتز $[(x - \tau_k)]$ بزرگتر از صفر باشد، آنگاه آن عبارت وارد مدل می‌شود و در غیر این صورت از مدل حذف می‌گردد (۴).

و مدل لگاریتمی معادله رگرسیونی به صورت (۲) است:

$$E(y|x) \\ = e^{\beta_0 + \beta_1 x + \delta_1(x - \tau_1)^+ + \dots + \delta_k(x - \tau_k)^+} \quad (2)$$

که در فرمول e پایه طبیعی است.

روش مدل سازی

یکی از متداول‌ترین روش‌های مورد استفاده جهت برآورد پارامترهای مدل رگرسیون Joinpoint روش حداقل مربعات است. برای تعیین محل نقاط شکست و برآورد پارامترها با استفاده از این روش، از تکنیک‌های جستجوی شبکه‌ای لرمین (Grid Search) و الگوریتم پیوسته هادسون (Hudson's Algorithm) استفاده می‌شود (۵، ۶). تفاوت عمده بین دو روش در زمان محاسبه است که روش هادسون کندتر عمل می‌کند و نرم افزار Joinpoint به صورت پیش فرض از جستجوی شبکه‌ای استفاده می‌کند.

روش جستجوی شبکه‌ای یک "شبکه" از تمام مکان‌های ممکن برای نقاط شکست مشخص شده ایجاد می‌کند و سپس مجموع مربعات خطا (SSE) و میانگین مربعات خطا (MSE) را در هر یک از این نقاط محاسبه می‌کند و نقطه شبکه‌ای دارای کمترین MSE است، به عنوان نقطه شکست برای توابع قطعه‌بندی انتخاب می‌شود. در نهایت، براساس نقاط شکست و توابع بازه‌های قطعه‌بندی انتخابی، پارامترهای معادله نظیر β_0, β_1 و $\delta_1, \dots, \delta_k$ برآورد می‌شوند.

چرا برای تحلیل روند زمانی از رگرسیون Joinpoint

استفاده کنیم؟

از مهمترین دلایل برای به کارگیری رگرسیون Joinpoint جهت تحلیل روند زمانی، شناسایی زمان‌هایی که تغییرات قابل توجهی در روندها رخ داده است. به عبارتی دیگر تشخیص تعداد و مکان (زمان) تغییر در روند است.

مطالعات زیادی به ویژه در اپیدمیولوژی انجام می‌شود که نشان دهد روند تغییرات میزان بیماری و مرگ‌ومیر در طول زمان چگونه است. در این نوع مطالعات متغیر اصلی زمان است و به دنبال دیدن اثرات زمان بر بروز و یا فراوانی پدیده‌هایی همچون بیماری‌ها هستیم اما همیشه این ارتباط خطی نیست و گاهی در طول زمان نقاطی وجود دارند که الگوی روند را تغییر می‌دهند مانند نقطه شروع پاندمی کووید ۱۹ که می‌تواند همه روندهای گذشته را به ناگهان تغییر دهد. علاوه بر این، برای برخی از موقعیت‌ها که با تغییر مواجهه ریسک فاکتورها، انجام برنامه‌های غربالگری یا دیگر مداخلات روبرو هستیم، برای ما مهم است که قادر به تشخیص زمانی که تغییرات معنی‌دار روی می‌دهد باشیم. از طرفی، در نظر گرفتن روند کل برای یک دوره زمانی با مشاهدات گسترده می‌تواند تغییرات روندها را در دوره‌های زمانی کوتاه‌تر پنهان کند (۲، ۳). بنابراین رگرسیون Joinpoint یکی از نگرانی‌های شایع که تشخیص تغییر در روند زمانی است را رفع نموده است. یکی دیگر از دلایل به کارگیری رگرسیون Joinpoint جهت تحلیل روند، تفسیر ساده و قابل فهم‌تر نتایج است. به طوری که می‌توان از کمیت درصد تغییرات سالانه (Annual percent change: APC) برای تفسیر شیب‌های رگرسیونی استفاده کرد. به طور کلی اهداف اصلی مدل رگرسیون Joinpoint، استنباط در مورد وجود یا عدم وجود نقطه شکست، برآورد تعداد و مکان (زمان) نقاط شکست و برآورد پارامترهای قطعه‌های رگرسیونی است. از این رو، این مطالعه با هدف بررسی عملکرد رگرسیون Joinpoint جهت تحلیل روند زمانی انجام شد.

روش کار

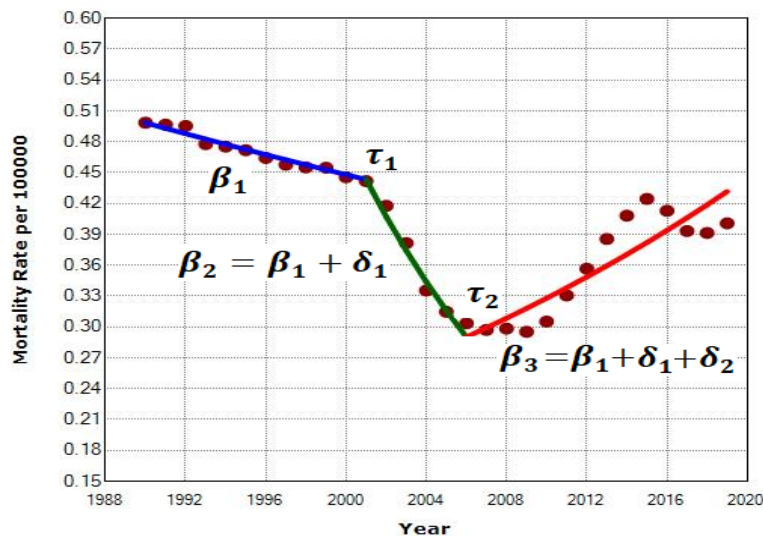
معرفی مدل

مدل رگرسیون Joinpoint روشی است که داده‌ها را به زیر مجموعه‌هایی که هر کدام دارای روند خطی منحصر به فرد خود هستند تقسیم می‌کند که بخش‌های خطی توسط "نقاط شکست" از هم جدا می‌شوند. بنابراین برای هر بخش یک تابع رگرسیون خطی با پارامترهای متفاوت وجود دارد.

مدل‌های رگرسیون Joinpoint شامل مدل‌های خطی $(y=xb)$ و لگاریتمی $(\ln y=xb)$ می‌باشد. اگر متغیر وابسته از توزیع نرمال پیروی کند و اندازه نمونه داده بزرگ باشد (معمولاً بیشتر از ۱۰۰)

شبکه با تولید یک شبکه گسترده‌تر و آزمون تمام موقعیت‌ها، به صورت جامع و کامل برای برازش بهترین مدل عمل می‌کند. به عنوان مثال، برای برازش مدل و برآورد پارامترها با ۲ نقطه شکست، ۳ معادله خطی وجود دارد (شکل ۱). هر قطعه یک معادله خطی فرض می‌شود، در قطعه اول بعد از رسیدن به نقطه شکست اول، شیب خط به یکباره تغییر می‌کند و لذا باید β_1 تغییر کند و مقداری به آن اضافه گردد که معادله خطی دوم با شیب β_2 به وجود می‌آید و در قطعه دوم نیز بعد از رسیدن به نقطه شکست دوم همین رویه ادامه پیدا می‌کند.

در روش هادسون، برای برازش بهترین مدل، آزمون‌های پیوسته بین مقادیر مشاهده شده صورت می‌گیرد. در هر مرحله، یک زیر مدل مناسب از داده‌ها به دست می‌آید که دارای کمترین مربعات خطا (SSE) در نقاط مشاهده شده است، سپس با ترکیب این زیر مدل‌ها، یک منحنی کامل برای ساختار به دست می‌آید که مجموع مربعات خطای (MSE) آن کمینه شده است. به طور کلی، روش هادسون بر اساس یک رویکرد تکاملی و مدل‌سازی تدریجی عمل می‌کند، در حالی که روش جستجوی



شکل شماره ۱- خروجی مدل رگرسیون Joinpoint با حداکثر ۲ نقطه شکست

مقایسه بین مدل‌هایی با تعداد نقاط شکست متفاوت است (۴). تعیین تعداد نقاط شکست به عرض بازه زمانی داده‌های تحقیق (تعداد مشاهدات) بستگی دارد. اولین مقایسه بین مدل بدون نقطه شکست و مدل با تعداد نقاط شکست برابر با حداکثر عدد ثابت دلخواه است. مدل نهایی، مدلی است که کمترین پارامتر را دارد، که افزودن یک پارامتر بیشتر (نقطه شکست) برای آن منجر به بهبود قابل توجهی نمی‌شود. به عنوان مثال، برای تعیین حداکثر ۲ نقطه شکست، باید یک فرضیه صفر مبنی بر عدم نقطه شکست را در مقابل فرضیه ۲ نقطه شکست آزمون کنیم.

$$H_0: E(y|x) = \beta_0 + \beta_1 x \quad (K_a = 0)$$

$$H_1: E(y|x) = \beta_0 + \beta_1 x + \delta_1(x - \tau_1)^+ + \delta_2(x - \tau_2)^+ \quad (K_b = 2)$$

K_b یک واحد کم می‌شود (آزمون فرضیه با ۰ نقطه شکست در مقابل فرضیه ۱ نقطه شکست). و این رویه تا زمانی که $K_a = K_b = k$ و به یک تعداد نقاط شکست برسیم ادامه پیدا می‌کند.

انتخاب بهترین مدل

از آزمون جایگشت (Permutation Test) و معیار اطلاع بیز (Bayesian Information Criterion: BIC) برای انتخاب بهترین مدل از بین مدل‌های موجود استفاده می‌شود (۴، ۷). این دو روش دو رویکرد متفاوت برای انتخاب تعداد بهینه نقاط شکست هستند و به دلیل مشکلات تشخیص ناپذیری مدل‌ها آزمون جایگشت ترجیح داده می‌شود و نرم افزار Joinpoint به صورت پیش فرض از این آزمون استفاده می‌کند. بهترین مدل (تعداد نقاط شکست) توسط آزمون جایگشت با استفاده از نمونه‌گیری مجدد مونت کارلو تعیین می‌شود که شامل

آزمون با کمترین تعداد نقاط شکست (K_a) و بیشترین تعداد نقاط شکست (K_b) شروع می‌شود. اگر فرضیه صفر رد شود به K_a یک واحد اضافه می‌شود (آزمون فرضیه با ۱ نقطه شکست در مقابل فرضیه ۲ نقطه شکست) و اگر فرضیه صفر رد نشود از

همچنین تشخیص مدل بهینه از طریق معیار اطلاع بیز (BIC) به آسانی از طریق دو مرحله، محاسبه BIC برای هر مدل قطعه‌ای و انتخاب مدل با کمترین BIC به عنوان مدل بهینه انجام می‌شود (۷). معادله محاسبه BIC برای یک مدل با k نقطه شکست به صورت (۳) است:

$$BIC(K) = \ln\left(\frac{SSE(K)}{n}\right) + \left(\frac{2k+2}{n}\right) \times \ln(n) \quad (3)$$

میانگین کل این تغییرات است که برای توصیف روندهای خطی براساس هر دوره محاسبه می‌شوند (۸). جهت برآورد APC فرض می‌شود میزان‌ها در سال x با درصد ثابتی از میزان در سال بعد $x+1$ تغییر می‌کند. میزان‌هایی که سالانه با یک درصد ثابتی تغییر می‌کنند با مقیاس لگاریتمی رابطه خطی دارد، بنابراین از مدل رگرسیونی $\log(Y_x) = \beta_0 + \beta_1 x$ استفاده می‌شود که در آن $\log(Y_x)$ لگاریتم طبیعی میزان y در سال x می‌باشد (۹). بنابراین APC از سال x تا سال $x+1$ به صورت (۴) است:

$$APC = \frac{Y_{x+1} - Y_x}{Y_x} \times 100 = \frac{e^{\beta_0 + \beta_1(x+1)} - e^{\beta_0 + \beta_1 x}}{e^{\beta_0 + \beta_1 x}} \times 100 = (e^{\beta_1} - 1) \times 100 \quad (4)$$

عنوان میانگین وزنی APC برآورد شده در هر بخش با استفاده از طول قطعه به عنوان وزن (w_i) برآورد می‌شود و به صورت (۵) است:

$$AAPC = \left(e^{\frac{\sum w_i APC_i}{\sum w_i}} - 1 \right) \times 100 \quad (5)$$

۲۰۰۶ تقسیم می‌شود با شیب‌های به ترتیب برابر با $\beta_1 = -0.010$ ، $\beta_2 = -0.010$ و $\beta_3 = 0.030$ است. بنابراین APC ها برای هر بخش به صورت زیر برآورد می‌شوند:

$$\begin{aligned} APC_1 &= (e^{-0.010} - 1) \times 100 = -1.00\% \\ APC_2 &= (e^{-0.084} - 1) \times 100 = -8.06\% \\ APC_3 &= (e^{0.030} - 1) \times 100 = 3.04\% \end{aligned}$$

بنابراین AAPC به صورت زیر برآورد می‌شود: همچنین طول هر بخش به ترتیب برابر با ۱۱، ۵ و ۱۳ سال است.

$$AAPC = \left(e^{\frac{(-11 \times 0.01) + (-5 \times 0.0806) + (13 \times 0.0304)}{11+5+13}} - 1 \right) \times 100 = -0.40\%$$

به دلیل اینکه آزمون جایگشت زمان بر است، برای ملاحظه سرعت محاسبات و پایداری نتایج، تعداد آزمون جایگشت توصیه شده توسط نرم‌افزار و پیش‌فرض سیستم ۴۴۹۹ بار است. و به دلیل وجود آزمون‌های چندگانه، از روش بونفرونی برای اصلاح سطح معناداری آماری استفاده می‌شود.

که در آن SSE مجموع مربعات خطا مدل رگرسیون با k نقطه شکست، $2k+2$ تعداد پارامترهای مدل با k نقطه شکست و n تعداد مشاهدات است.

شاخص‌های APC و AAPC

درصد تغییر سالانه (Annual percent change: APC) و میانگین درصد تغییر سالانه (Average annual percent change: AAPC) شاخص‌های اصلی نتیجه مدل رگرسیون Joinpoint هستند. APC بیانگر مقدار افزایش یا کاهش درصد تغییرات در هر سال است که در صورت منفی بودن، روند کاهشی و در صورت مثبت بودن، روند افزایشی را نشان می‌دهد و AAPC منیز

وقتی که هیچ نقطه شکستی وجود ندارد یعنی تغییری در روند وجود ندارد، بنابراین APC ثابت و برابر با AAPC خواهد بود. در صورتی که تغییری در روند وجود داشته باشد کل دوره توسط نقاط شکست به بخش‌هایی تقسیم می‌شود. در نتیجه، AAPC به

به عنوان مثال، رگرسیون jointpoint دو نقطه شکست را در سال‌های ۲۰۰۱ و ۲۰۰۶ مشخص می‌کند (شکل ۱)، بنابراین کل دوره در سه بخش ۲۰۰۱-۲۰۰۶، ۲۰۰۶-۲۰۰۱ و ۲۰۱۹-۲۰۰۱

پیش‌بینی

سال قبل تغییر می‌کند (۲، ۱۰). معمولاً استفاده از مدل رگرسیون jointpoint برای پیش‌بینی‌های کوتاه مدت استفاده می‌شود.

به عنوان مثال، اگر آخرین میزان بروز یا مرگ‌ومیر در دسترس برای سال ۲۰۱۹، ۱۰۰ در هر ۱۰۰۰۰۰ و مقدار APC برآورد شده آخرین قطعه برابر ۱٪ باشد، میزان پیش‌بینی شده برای سال‌های ۲۰۲۰ و ۲۰۲۱ به ترتیب برابر ۱۰۱ و ۱۰۲/۰۱ در هر ۱۰۰۰۰۰ خواهد بود.

پیش‌بینی یک رویداد آینده، فرآیند پیچیده‌ای است که در معرض عدم قطعیت‌های بزرگ است و از بسیاری جنبه‌ها مشکوک است. با این حال، کسب اطلاعات در مورد روندهای آینده، حتی اگر نامشخص یا نادقیق باشد، مفید است. ابتدا مدل رگرسیون jointpoint با استفاده از میزان‌های مورد نظر در سال‌های قبل برازش می‌شود. سپس با استفاده از آخرین میزان بروز یا مرگ-ومیر در دسترس و مقدار APC برآورد شده از آخرین قطعه مدل رگرسیون jointpoint می‌توان مقادیر برای آینده پیش‌بینی کرد. بدین صورت که میزان بروز یا مرگ‌ومیر با درصد ثابتی از میزان

$$\begin{aligned} \text{Prediction rate (2020)} &= \text{Latest year rate (2019)} \times \text{APC (\%)} = 100 \times 1.01 = 101 \\ \text{Prediction rate (2021)} &= \text{Prediction rate (2020)} \times \text{APC (\%)} = 101 \times 1.01 = 102.01 \end{aligned}$$

متعلق به موسسه اندازه‌گیری و ارزشیابی سلامت (Institute for Health Metrics and Evaluation: IHME) دانشگاه واشنگتن دریافت شد (۱۲).

به کارگیری مدل

ابتدا میزان مرگ‌ومیر استاندارد شده سنی ناشی از سرطان کبد به دلیل مصرف الکل به عنوان متغیر وابسته و متغیر سال (زمان) برای ۳۰ سال (۲۰۱۹-۱۹۹۰) به عنوان متغیر مستقل در نظر گرفته شدند. برای پیدا کردن نقاط شکست، حداکثر ۴ نقطه شکست در روند در نظر گرفته شد که نرم افزار مدل‌های مختلف را به صورت بدون نقطه‌ی شکست، دارای ۱ نقطه‌ی شکست، دارای ۲ نقطه‌ی شکست، دارای ۳ نقطه‌ی شکست و دارای ۴ نقطه‌ی شکست بر روی داده‌ها بررسی کرد و آزمون‌های مختلف برای معناداری تعداد نقاط شکست براساس داده‌ها را انجام داد و بهترین مدل انتخاب شد و ضرایب مربوط به مدل منتخب و نمودار ارائه گردید. همچنین به دلیل سادگی و ملموس بودن نتایج حاصل شده از مدل لگاریتمی با واریانس ثابت استفاده شد و سطح معناداری آزمون‌ها ۰/۰۵ در نظر گرفته شد.

شناسه اخلاق

این مطالعه در شورای پژوهشی دانشگاه علوم پزشکی کرمان بررسی و با شناسه اخلاق IR.KMU.REC.1402.008 مصوب گردید.

یافته‌ها

برازش داده‌های میزان مرگ‌ومیر استاندارد شده سنی ناشی از سرطان کبد به دلیل مصرف الکل برای کشور ایران طی سال‌های

نرم افزار Joinpoint Regression

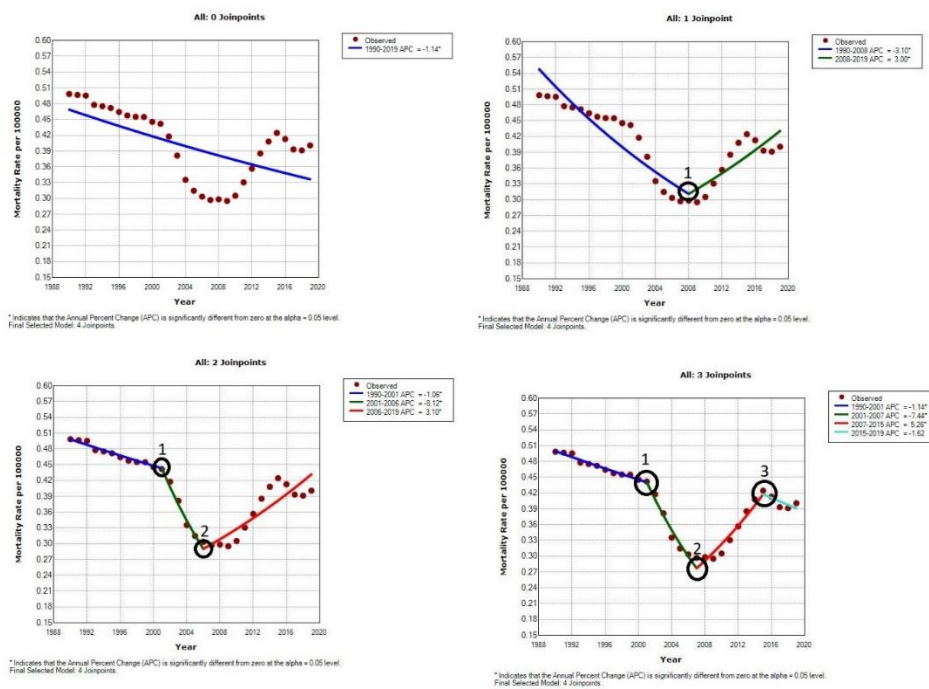
نرم افزار Joinpoint Regression به صورت اختصاصی برای انجام تحلیل رگرسیون Joinpoint طراحی شده است و به صورت رایگان از طریق موسسه ملی سرطان (National Cancer Institute: NCI) در دسترس است (۱۱) و NCI نیز با استفاده از این نرم افزار به تحلیل روند بروز و مرگ‌ومیر سرطان می‌پردازد. نرم افزار قابلیت انجام رگرسیون Joinpoint برای متغیر وابسته به صورت خطی و لگاریتمی را دارد. تعیین تعداد نقاط شکست به عرض بازه زمانی داده‌های تحقیق (تعداد مشاهدات) بستگی دارد. پیش فرض نرم افزار حداقل ۰ و حداکثر ۵ است. کاربر می‌تواند براساس نیاز خود آن را تعریف کند. با این حال، اگر نقاط شکست بیش از حد وجود داشته باشد، عملیات بسیار زمان بر خواهد بود. بنابراین، حداکثر تعداد نقاط شکست توصیه شده توسط نرم افزار براساس تعداد مشاهدات (۰ تا ۶)، (۷ تا ۱۱)، (۱۲ تا ۱۶)، (۱۷ تا ۲۱)، (۲۲ تا ۲۶) و (۲۷ تا و بیشتر) به ترتیب ۰، ۱، ۲، ۳، ۴ و ۵ است. هنگام انجام تجزیه و تحلیل روند، اگر تغییرات نسبتاً پایدار باشد و صعود یا نزول بزرگی وجود نداشته باشد، بیانگر آن است که واریانس داده‌های تحقیق ثابت است، اما اگر افزایش یا افت ناگهانی در مدت زمان کوتاهی وجود داشته باشد، باید از گزینه خطای استاندارد یا پواسون استفاده کرد.

مجموعه داده

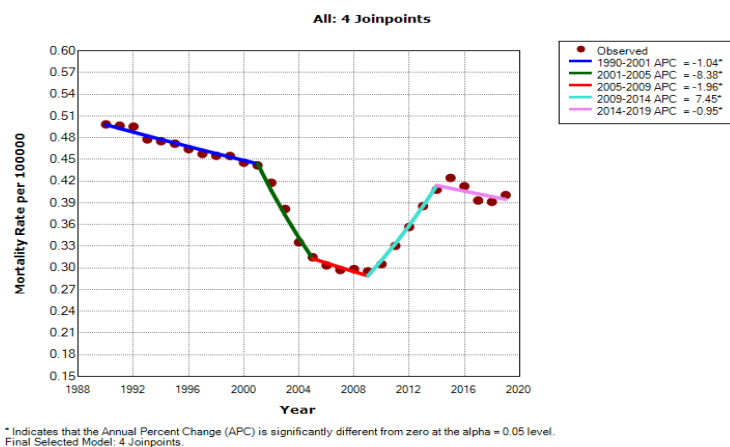
به منظور به کارگیری مدل، داده‌های مربوطه از آمارهای میزان مرگ‌ومیر استاندارد شده سنی ناشی از سرطان کبد به دلیل مصرف الکل برای کل جمعیت ایران طی سال‌های ۲۰۱۹-۱۹۹۰ از سایت بار جهانی بیماری (Global Burden of Disease: GBD)

به طور کلی میزان مرگ‌ومیر استاندارد شده سنی ناشی از سرطان کبد به دلیل مصرف الکل طی دوره بررسی (۱۹۹۰-۲۰۱۹)، به طور میانگین سالانه ۰/۸ درصد کاهش یافته است. همچنین بیش‌ترین افزایش و کاهش در روند مرگ‌ومیر به ترتیب مربوط به دوره چهارم (۲۰۰۹-۲۰۱۴) با ۷/۵ درصد و دوره دوم (۲۰۰۵-۲۰۰۱) با ۸/۴ درصد در سال می‌باشد (جدول ۱). در نهایت، پیش‌بینی میزان کل مرگ‌ومیر استاندارد شده سنی ناشی از سرطان کبد به دلیل مصرف الکل در ایران تا سال ۲۰۳۰ انجام شد (نمودار ۲).

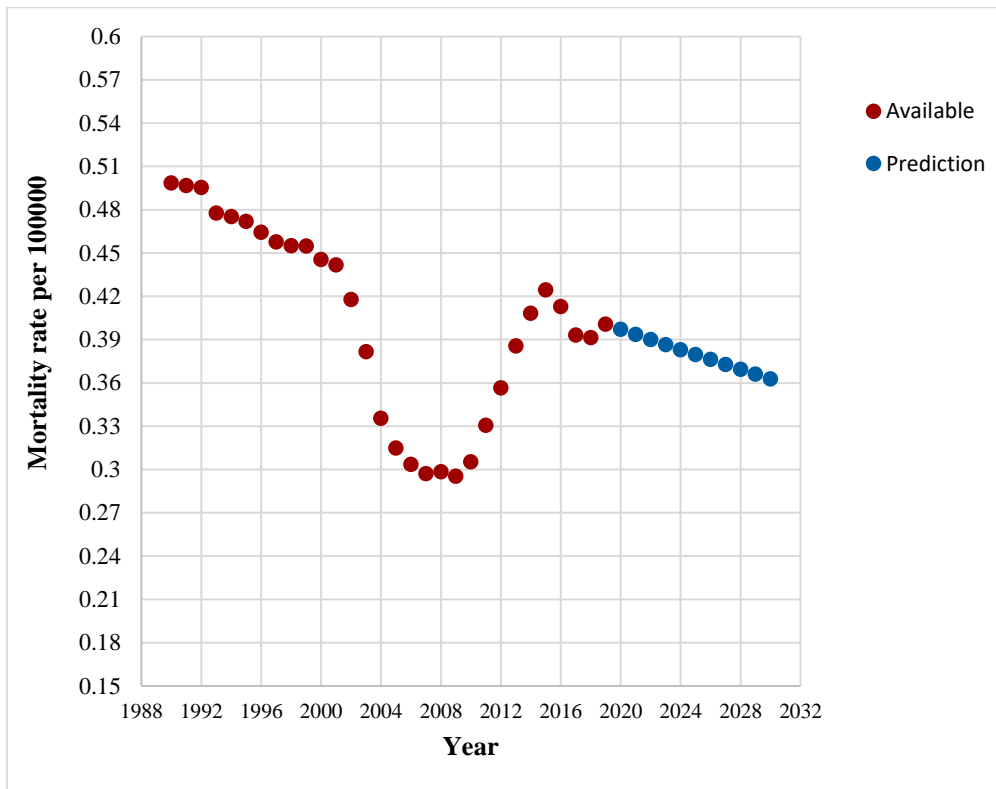
با استفاده از مدل رگرسیون Joinpoint و با بکارگیری روش جستجوی شبکه‌ای لرمین وجود چهار نقطه شکست را در سال‌های ۲۰۰۱، ۲۰۰۵، ۲۰۰۹ و ۲۰۱۴ نشان داد که منجر به پنج دوره زمانی شد (نمودار ۱ و شکل ۲). میزان مرگ‌ومیر استاندارد شده سنی ناشی از سرطان کبد به دلیل مصرف الکل طی سال‌های ۱۹۹۰ تا ۲۰۰۰ هر سال ۱/۰ درصد کاهش یافته، طی سال‌های ۲۰۰۱ تا ۲۰۰۴ هر سال ۸/۴ درصد کاهش یافته، طی سال‌های ۲۰۰۵ تا ۲۰۰۸ هر سال ۲/۰ درصد کاهش یافته، طی سال‌های ۲۰۰۹ تا ۲۰۱۳ هر سال ۷/۵ درصد افزایش یافته و طی سال‌های ۲۰۱۴ تا ۲۰۱۹ هر سال ۰/۹ درصد کاهش داشته است.



شکل شماره ۲- اجرای مدل رگرسیون Joinpoint با نقاط شکست مختلف جهت انتخاب مدل بهینه



نمودار شماره ۱- روند تغییرات کلی میزان مرگ‌ومیر استاندارد شده سنی ناشی از سرطان کبد به دلیل مصرف الکل در ایران طی سال‌های ۱۹۹۰-۲۰۱۹ (مدل بهینه)



نمودار شماره ۲- پیش‌بینی میزان مرگ‌ومیر استاندارد شده سنی ناشی از سرطان کبد به دلیل مصرف الکل در ایران تا سال ۲۰۳۰ (۲۰۳۰-۲۰۲۰)

جدول شماره ۱- تجزیه و تحلیل روند داده‌های میزان مرگ‌ومیر استاندارد شده سنی ناشی از سرطان کبد به دلیل مصرف الکل در ۱۰۰۰۰۰ نفر جمعیت در ایران طی سال‌های ۱۹۹۰-۲۰۱۹

P-value	Test Statistic	APC (95 % CI)	قطعه (بازه زمانی)	نقطه شکست
<۰/۰۰۱	-۸/۳	-۱/۰ (-۱/۳, -۰/۸)	۱۹۹۰-۲۰۰۰	۲۰۰۱
<۰/۰۰۱	-۹/۴	-۸/۴ (-۱۰/۲, -۶/۶)	۲۰۰۱-۲۰۰۴	۲۰۰۵
۰/۰۴۸	-۲/۱	-۲/۰ (-۳/۹, -۰/۱۰)	۲۰۰۵-۲۰۰۸	۲۰۰۹
<۰/۰۰۱	۱۲/۳	۷/۵ (۶/۱, ۸/۸)	۲۰۰۹-۲۰۱۳	۲۰۱۴
۰/۰۳۶	-۲/۳	-۰/۹ (-۱/۸, -۰/۱)	۲۰۱۴-۲۰۱۹	-
<۰/۰۰۱	-۳/۶	-۰/۸ (-۱/۲, -۰/۴)	۱۹۹۰-۲۰۱۹	-

جدول شماره ۲- میزان مرگ‌ومیر استاندارد شده سنی ناشی از سرطان کبد به دلیل مصرف الکل و و پیش‌بینی آن در ۱۰۰۰۰۰ نفر جمعیت در ایران طی سال‌های ۱۹۹۰-۲۰۳۰

میزان رشد (%)		پیش‌بینی		در دسترس			
۲۰۱۹-۲۰۳۰	۲۰۱۹-۲۰۲۵	۲۰۳۰	۲۰۲۵	۲۰۱۹	۲۰۱۰	۲۰۰۰	۱۹۹۰
-۹/۵	-۵/۳	۰/۳۶	۰/۳۸	۰/۴۰	۰/۳۱	۰/۴۵	۰/۵۰

کمک کند تا دینامیک روند زمانی یک بیماری را بهتر درک کنند و از این اطلاعات برای برنامه‌ریزی بهینه استفاده کنند. ایده اصلی این مدل، ایجاد رگرسیون تقسیم‌بندی شده براساس ویژگی‌های زمانی توزیع بیماری یا مرگ است که باعث تقسیم زمان تحقیق

بحث

مدل رگرسیون Joinpoint به دلیل قابلیت تشخیص تغییرات معنادار در روند زمانی، به عنوان یک ابزار قدرتمند در مطالعات اپیدمیولوژیک استفاده می‌شود. این مدل می‌تواند به محققان

جستجو شبکه‌ای لرمین وجود چهار نقطه شکست یا تغییر روند را در سال‌های ۲۰۰۱، ۲۰۰۵، ۲۰۰۹ و ۲۰۱۴ نشان داد که منجر به پنج دوره زمانی شد. همچنین نتایج نشان می‌دهد تنها در دوره چهارم شناسایی شده (۲۰۰۹-۲۰۱۴) روند افزایشی چشم‌گیری وجود دارد، بنابراین مطالعات تکمیلی لازم است تا دلیل افزایش در طی این ۵ سال را بررسی کنند.

لازم به ذکر است از آنجایی که مصرف الکل در ایران غیرقانونی است و به دلیل اعتقادات مذهبی و هنجارهای فرهنگی مصرف الکل پدیده‌ای زشت محسوب می‌شود بنابراین ممکن است آمارهای ثبت شده مربوط به مرگ ناشی از سرطان کبد به دلیل مصرف الکل کامل و دقیق نباشند، در نتیجه، سیاست‌گذاران سلامت باید در مورد روند نزولی مشاهده شده طی ۳۰ سال با در نظر گرفتن شرایط ذکر شده برای کنترل این بیماری برنامه‌ریزی نمایند.

نتیجه‌گیری

به طور کلی، استفاده از مدل رگرسیون Joinpoint امکان توصیف دقیق روند و بررسی تغییرات در روندها را طی سال‌های طولانی فراهم می‌کند، به طوری که سال‌های تقویمی را که در آن تغییرات آماری معنی‌داری در روندها رخ داده است را مشخص می‌کند و تصویر واضح‌تری از روندهای واقعی میزان‌ها در دوره‌های زمانی ارائه می‌دهد. نتایج مثال مورد بررسی در این مطالعه نشان داد، میزان مرگومیر استاندارد شده سنی ناشی از سرطان کبد به دلیل مصرف الکل طی سال‌های ۱۹۹۰ تا ۲۰۱۹ دارای چهار نقطه شکست یا تغییر روند در سال‌های ۲۰۰۱، ۲۰۰۵، ۲۰۰۹ و ۲۰۱۴ بوده است و به طور میانگین طی دوره مورد بررسی هر سال ۰/۸ درصد کاهش یافته است و پیش‌بینی نیز نشان داد این روند کاهشی ادامه دارد.

به فواصل مختلف از طریق چندین نقطه شکست و انجام مدل‌سازی روند در هر بازه مشخص شده است. در نهایت ارزیابی ویژگی‌های تغییر بیماری یا مرگ در بازه‌های زمانی مختلف به صورت دقیق‌تر صورت می‌گیرد. این روش از زمانی که پیشنهاد شد به طور گسترده در زمینه اپیدمیولوژی و به ویژه در زمینه تحقیقات در مورد بروز و مرگومیر سرطان مورد استفاده قرار گرفته است.

هنگام انجام تجزیه و تحلیل روند بیماری، بازه زمانی تحقیق نباید خیلی کوتاه باشد. معمولاً باید حداقل ۵ سال متوالی و ترجیحاً بازه زمانی بیش از ۱۰ سال متوالی باشد. اگر بازه زمانی کوتاه باشد، نتایج تحقیق تا حد زیادی تحت تاثیر عواملی مانند تنوع در توزیع بیماری و یکپارچگی و کیفیت داده‌های پایش قرار می‌گیرد و دستیابی به ویژگی‌های تغییر روند عینی را دشوار می‌کند. هنگام تفسیر نتایج تجزیه و تحلیل روند بلند مدت، لازم است که مکان‌های نقاط شکست و تغییرات روند در فواصل زمانی مختلف با سایر ویژگی‌های اپیدمیولوژیک مانند وقوع و توسعه بیماری، سیاست‌ها و اقدامات پیشگیری و کنترل ترکیب شود و به دنبال عوامل و روندهای ممکن در تغییرات روند باشیم. باید توجه داشت که توضیح منطقی راجع به نتایج تحلیل روند ارائه شود و از ترویج کورکورانه نتایج روند مدل خودداری شود. به عبارتی، مدل رگرسیون Joinpoint در یک مطالعه اکولوژیکی استفاده شده است، بنابراین رابطه علیتی قابل برقراری نیست و نتایج نیاز به تایید بیشتر با داده‌های سطح فردی دارند. تنها به عنوان یک نتیجه، می‌توان در مورد روندهایی که توسط داده‌ها مورد تایید قرار گرفته است فرضیه بیان کرد و باید از ادعاهای علی خودداری شود.

نتایج حاصل از این پژوهش حاکی است که میزان مرگومیر استاندارد شده سنی ناشی از سرطان کبد به دلیل مصرف الکل در ایران طی ۳۰ سال اخیر روند نزولی داشته است. برآزش داده‌ها با استفاده از مدل رگرسیون Joinpoint و با به‌کارگیری تکنیک

References

1. Bowling A. Research methods in health: investigating health and health services. 4th ed. McGraw Hill; Open University Press; 2014, 538P.
2. Jiang Z, Qiu Z, Hatcher J. Joinpoint trend analysis of cancer incidence and mortality using Alberta data. *Cancer Surveillance, Surveillance and Health Status Assessment, Alberta Health Services* 1. 2010, 45P.
3. Edwards BK, Ward E and Kohler BA et al. Annual report to the nation on the status of cancer, 1975-2006, featuring colorectal cancer trends and impact of interventions (risk factors, screening, and treatment) to reduce future rates. *Cancer* 2010; 116(3): 544-73.
4. Kim, H.J., et al.; Permutation tests for joinpoint regression with applications to cancer rates. *Statistics in medicine*, 2000; 19: 335-351.

5. Lerman, P.M. Fitting Segmented Regression Models by Grid Search. *Applied Statistics*. 1980; 29: 77-84.
6. Hudson, D. Fitting segmented curves whose join points have to be estimated. *Journal of the American Statistical Association*. 1966; 61: 1097-1129.
7. Zhang, N. R. & Siegmund, D. O. A modified Bayes information criterion with applications to the analysis of comparative genomic hybridization data. *Biometrics* 2007; 63: 22-32.
8. Balooch Hasankhani M, Naziri M, Roudbari M. Analysis of the incidence and mortality rate of Alzheimer's and other dementias during the last 30 years in Iran. *Koomesh* 1400; 24: 49-56
9. Clegg LX, Hankey BF, Tiwari R, Feuer EJ, Edwards BK. Estimating average annual per cent change in trend analysis. *Stat Med*. 2009 Dec 20;28(29):3670-82. doi: 10.1002/sim.3733. PMID: 19856324; PMCID: PMC2843083.

10. Khorrami Z, Balooch Hasankhani M, Khezri M, Jafari-Khounigh A, Jahani Y, Sharifi H. Trends and projection of incidence, mortality, and disability-adjusted life years of HIV in the Middle East and North Africa (1990–2030). *Scientific Reports*. 2023 Aug 24;13(1):13859.

11. Download Joinpoint Desktop Software (cancer.gov)-
<https://surveillance.cancer.gov/joinpoint/download>

12. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2020. Available from [https://vizhub.healthdata.org/gbd-results/...](https://vizhub.healthdata.org/gbd-results/)

Tehran University of
Medical Sciences

Original Article

The Concept and Application of the Joinpoint Regression Model in Analyzing the Time Trend in Epidemiological Studies (Case: Liver Cancer due to Alcohol Use in Iran)

Mohammadreza Balooch Hasankhani¹, Aliakbar Haghdoost², Yunes Jahani³

1- MS in Biostatistics, Modeling in Health Research Center, Institute for Futures Studies in Health, Kerman University of Medical Sciences, Kerman, Iran

2- Professor of Epidemiology, HIV/STI Surveillance Research Center, Institute for Futures Studies in Health, Kerman University of Medical Sciences, Kerman, Iran

3- Professor of Biostatistics, Modeling in Health Research Center, Institute for Futures Studies in Health, Kerman University of Medical Sciences, Kerman, Iran

Article Information

Received

15 April 2023

Accepted

03 September 2023

Corresponding author

Yunes Jahani

Corresponding author E-maily.jahani@kmu.ac.ir**Keywords:**

Trend analysis, Joinpoint regression, Alcohol, Liver cancer, Age-standardized mortality rate

Abstract

Background and Objectives: Time trend analysis of factors such as disease and mortality rates is a crucial component of health planning for any community. It allows for a more accurate interpretation of changes over time. This study was conducted to examine the performance of the Joinpoint regression model in analyzing time trends.**Methods:** This study aims to first provide a simplified understanding of the Joinpoint regression model and then demonstrate its application on data regarding the 30-year trends of liver cancer mortality due to alcohol consumption in Iran.**Results:** The results of the time trend analysis indicate that the age-standardized mortality rate of liver cancer due to alcohol use consumption has decreased by an average of 0.8% per year over the 30-year period in Iran (1990 to 2019). The projections also suggest that this declining trend will continue.**Conclusion:** In general, the main advantage of the Joinpoint regression model over other models is its ability to identify periods where significant changes in trends have occurred. Based on the results, the mortality rate of liver cancer due to alcohol use consumption over the 30-year period in Iran can be divided into five periods with different rates of change.

Copyright © 2023 The Authors. Published by Tehran University of Medical Sciences.

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International license (<https://creativecommons.org/licenses/by-nc/4.0/>). Non-commercial uses of the work are permitted, provided the original work is properly cited.