

ارتباط بروز بیماری سل و مهاجرت با استفاده از مدل رگرسیون چندکی در ایران سال ۱۳۸۹

فاطمه سروی^۱، یدا...محرابی^{۲*}، علیرضا ابدی^۳، مهشید ناصحی^۴

۱. کارشناس ارشد آمارزیستی، گروه آمارزیستی، دانشکده پیراپزشکی، دانشگاه علوم پزشکی شهید بهشتی، تهران
۲. استاد آمارزیستی، گروه اپیدمیولوژی، دانشکده بهداشت، دانشگاه علوم پزشکی شهیدبهشتی، تهران
۳. دانشیار آمارزیستی، گروه پزشکی اجتماعی، دانشکده پزشکی، دانشگاه علوم پزشکی شهیدبهشتی، تهران
۴. استادیار اپیدمیولوژی، گروه اپیدمیولوژی، دانشکده بهداشت، دانشگاه علوم پزشکی ایران، تهران

* نشانی برای مکاتبه: تهران- ولنجک- دانشگاه علوم پزشکی شهید بهشتی- دانشکده بهداشت- گروه اپیدمیولوژی، تلفن: ۰۰۹۱۲۳۸۴۹۷۰۷، mehrabi@sbmu.ac.ir

چکیده

سابقه و هدف: مدل رگرسیون چندکی، روشی کارآمد برای پیش بینی و تخمین رابطه متغیرهای تبیینی و نقاط صدکی مختلف توزیع متغیر پاسخ می باشد. در این مطالعه برای بررسی رابطه نرخ مهاجرت و میزان بروز سل از این مدل استفاده شد.

روش کار: مطالعه مقطعی حاضر روی ۱۱۳۲۰ فرد که در سال ۱۳۸۹ مبتلا به بیماری سل شده اند انجام گرفت. داده ها از وزارت بهداشت، درمان و آموزش پزشکی و مرکز آمار ایران دریافت شد. به دلیل ماهیت چولگی داده های میزان بروز سل و هم چنین بررسی چندک های انتهایی توزیع، ارتباط بین میزان بروز سل و نرخ مهاجرت با تعدیل دیگر شاخص ها، با مدل رگرسیون چندکی برآورد شد و برای محاسبه نیکویی برازش از معیار اطلاع آکائیک استفاده شد. برای مدل سازی و تجزیه و تحلیل داده ها، نرم افزار $R(3.0.1)$ و پکیج *Quantreg* به کار گرفته شد.

یافته ها: نتایج مدل رگرسیونی چندکی نشان داد که با افزایش نرخ مهاجرت، میزان بروز سل روند افزایشی داشته است. برای شهرستان هایی که نرخ مهاجرت آن ها در صدک های ۵۰، ۶۰، ۷۰، ۸۰، ۹۵ ام قرار داشت، این ارتباط معنی دار بود.

نتیجه گیری: بیماری سل و مهاجرت با یکدیگر ارتباط داشتند، این ارتباط به خصوص برای شهرستان هایی که نرخ مهاجرت آن ها در صدک های بالایی قرار داشت معنا دار بود.

واژگان کلیدی: مهاجرت، سل، مدل رگرسیون چندکی

مقدمه

محیطی و شرایط اسکان، برای اولین بار با این بیماری برخورد کرده ویا امکان فعال شدن میکروب سل فراهم شده و میزان بروز سل افزایش می- یابد. اگر مهاجرت از یک منطقه با نرخ کم سل به منطقه ای با نرخ زیاد سل صورت بگیرد امکان آلودگی به بیماری سل برای گروه مهاجر افزایش می- یابد و اگر مهاجرت از مناطقی با نرخ بالای سل به مناطقی با نرخ پایین سل صورت گیرد، این گروه مهاجر، میزان بروز سل را برای مناطق مهاجرپذیر افزایش می دهند (۲).

در سال ۱۹۹۹ مرکز مبارزه با سل اتحادیه اروپا گزارش کرد که در برخی کشورهای اروپایی نظیر سوئد، نروژ، دانمارک و اسرائیل، ۵۰ درصد افراد مبتلا به سل از مهاجران بوده است (۳). در سال ۲۰۰۱، جوز لئوپولد و هم کارانش در بررسی توزیع بیماری و مرگ و میر ناشی از بیماری سل را از سال ۱۹۹۴ تا ۱۹۹۸ به این نتیجه رسیدند که بیماری و مرگ و میر حاصل از سل با نرخ مهاجرت ارتباط معنی دار قابل ملاحظه ای دارد. این مطالعه که در شهر سانوآپولو برزیل انجام شد

تقریباً یک سوم جمعیت جهان (یعنی ۲ میلیارد نفر) به میکروب سل آلوده و یا در خطر ابتلا به بیماری سل قرار دارند و هر ساله حدود ۹ میلیون نفر به سل فعال مبتلا شده و ۱/۵ تا ۲ میلیون نفر در اثر این بیماری جان می سپارند. بیش از ۹۰٪ موارد بیماری و مرگ ناشی از سل در کشورهای در حال توسعه رخ می دهد، کشورهایی که ۷۵٪ موارد بیماری در آنها به فعال ترین گروه سنی به لحاظ اقتصادی (یعنی ۱۵ تا ۶۴ سالگی) تعلق دارد. واضح است که سل بجز لطمات اقتصادی، اثرات منفی غیرمستقیم دیگری نیز بر کیفیت زندگی بیماران یا افراد خانواده آنها دارد، و از این نظر دارای مرتبه دهم در بار جهانی بیماری ها است و پیش بینی می شود تا سال ۲۰۲۰ هم چنان جایگاه کنونی خود را حفظ کند و یا تا رتبه هفتم بالا رود (۱). درصد بالاتری از بیماری سل در بین گروه های مهاجر دیده می شود، حتی اگر این گروه ها به مناطقی که از لحاظ بروز سل وضعیت مشابهی داشته باشند مهاجرت کنند، به دلیل تغییر شرایط زیست

رگرسیون چندکی یک روش آماری با قابلیت محاسبه و رسم منحنی های رگرسیونی متفاوت و منطبق با نقاط صدکی مختلف می باشد، که ضمن بیان تصویری کامل تر و جامع تر از داده ها، امکان سنجش ارتباط متغیرهای مستقل با چندک های مورد نظر متغیر وابسته را بدون نیاز به نرمال بودن داده ها و حتی در حضور نقاط دور افتاده فراهم می کند یعنی این رگرسیون نسبت به داده های دور افتاده نیرومند می باشد. از سوی دیگر برخلاف رگرسیون حداقل مربعات که روی میانگین شرطی یعنی پارامتر مکان متمرکز است، رگرسیون چندکی استراتژی منظمی را برای تعیین چگونگی تاثیر متغیرهای مستقل روی مکان و مقیاس و شکل توزیع پیشنهاد می کند (۱۰). برآورد توابع چندک شرطی یا رگرسیون چندکی در سال ۱۹۷۸ توسط کوانکر و باست بررسی شد (۱۱)، در ابتدا این رگرسیون برای داده های پیوسته معرفی گردید ولی به سرعت برای دیگر مجموعه ها و در حیطه های مختلف تعمیم یافت، به عنوان مثال مانسکی در سال ۱۹۷۵ و سال ۱۹۸۵ رگرسیون میانه را برای مدل های چند جمله ای و دودویی تعمیم داد (۱۲)، در سال ۱۹۸۴ و ۱۹۸۶، پاول این رگرسیون را برای داده های سنسور شده تعمیم داد (۱۳)، در سال ۲۰۰۵ سانتز سیلوا و ماکادو این رگرسیون را برای داده های شمارشی تعمیم دادند (۱۴). هدف مطالعه حاضر، تعیین ارتباط بین میزان بروز سل و نرخ مهاجرت در ایران می باشد که برای این منظور از مدل رگرسیون چندکی استفاده شد.

روش کار

در این مطالعه داده های کل افراد مسلول ثبت شده به تعداد ۱۱۳۲۰ نفر به تفکیک شهرستان، مربوط به سال ۱۳۸۹ از اداره کنترل بیماری های سل و جدام وزارت بهداشت دریافت گردید علاوه بر آن شاخص های اجتماعی-اقتصادی شامل نسبت پزشک به تعداد نقاط جمعیتی، نرخ بی کاری، نرخ بی سواد، تراکم خانوار به واحد مسکونی، متوسط درآمد هر خانوار و نرخ مهاجرت و ضریب شهرنشینی از مرکز آمار ایران دریافت شد. با محاسبه نسبت تفاوت بین مهاجران وارد شده و مهاجران خارج شده (خالص مهاجرت) از محدوده مکانی مورد مطالعه به جمعیت محدوده مورد نظر ضرب در ۱۰۰، نرخ مهاجرت به دست می آمد. برخلاف رگرسیون حداقل مربعات که برای برآورد میانگین شرطی از مینیمم کردن مانده ها استفاده می کند، روش رگرسیون چندکی برای برآورد چندک های شرطی از مینیمم کردن قدر مطلق موزون مانده های نامتقارن استفاده می کند. اگر مدل رگرسیون خطی به صورت : $Y_i = x_i' \beta_\tau + \varepsilon_{\tau i}$ باشد و فرض کنیم $Q_Y(\varepsilon_{\tau i} | X_i) = 0$ ، آن گاه تابع چندک شرطی τ ام توزیع Y به شرط متغیرهای تبیینی X به صورت $\hat{Q}_{Y_i}(Y_i | x_i) = x_i' \hat{\beta}(\tau)$ در می آید. برای برآورد پارامترهای رگرسیون چندکی از روش حداقل قدر مطلق انحرافات (LAD) استفاده می شود و پارامتر چندکی یعنی را به این صورت برآورد می شود:

نشان داد که بروز و مرگ و میر ناشی از سل در خارجیانی که در این شهر ساکن شده بودند دو برابر افرادی بوده که در این شهر به دنیا آمده بودند و هم چنین این رقم برای افراد برزیلی که متولد سایر شهرهای برزیل بوده ولی ساکن در شهر سائوپائولو بودند ۸ برابر افرادی بوده که متولد این شهر بوده اند (۴).

در سال ۲۰۰۹، سوپرویا و هم کارانش رابطه بیماری سل را در بین افراد مهاجر از کشورهای دیگر ساکن در شهر بارسلونای اسپانیا در سال های ۲۰۰۶ و ۲۰۰۷ بررسی کردند. شهر بارسلونا مهاجرپذیرترین شهر اسپانیا می باشد که ۱۴/۱ درصد پاکستانی، ۱۰/۱ درصد اکوادوری، ۳۸/۳ درصد مهاجر از آمریکای لاتین و ۱۹/۲ مهاجر از اروپای غربی در سال های ۲۰۰۶ و ۲۰۰۷ داشت. ۶۰ مورد مبتلا به سل در بین مهاجران شناسایی شد که اکثر این بیماران از کشور پاکستان (۲۶/۷٪) و کشور بولیوی (۱۵٪) آمده بودند (۵). تالیوت و هم کارانش در سال ۲۰۰۰ نرخ ابتلا به سل را در بین مهاجران ساکن در ایالت متحده آمریکا از سال ۱۹۹۳ تا سال ۱۹۹۸ بررسی کردند و برای هر صد هزار نفر مهاجر نرخ ابتلا به سل را ۳۲/۹ درصد و برای هر صد هزار نفر متولد آمریکا نرخ ابتلا به سل را ۵/۸ درصد گزارش کردند (۶). گیلیبرت و هم کارانش در سال ۲۰۰۹ مهاجرت روی نرخ ابتلا به سل را در انگلستان با دیگر کشورهای اروپایی مقایسه کردند، از بین ۲۱ کشور اروپایی نرخ ابتلا به سل در سه کشور نروژ، سوئد و انگلستان روند افزایشی داشت، در هر سه کشور تقریباً ۰،۷۵ درصد مبتلایان به سل، در خارج متولد شده بودند و عموماً در کشورهای با نرخ بالای بروز سل متولد شده بودند (۷). در سال ۲۰۱۲ ام سی براید و هم کارانش میزان خطر بروز سل فعال را در مهاجران ایالت ویکتوریا استرالیا بررسی کردند، آن ها دریافتند که بسته به اینکه فرد مهاجر از چه کشوری و از چه ناحیه ای مهاجرت کرده است میزان خطر بروز سل نوسان دارد، برای مثال برای مهاجران آسیایی، آفریقا و اطراف اقیانوس آرام این میزان نسبت به مهاجران بقیه کشورها بیش تر بود و در اولین ۶ سال مهاجرت، در هر صد هزار نفر ۱۰۰ تا ۱۵۰ نفر مبتلا دیده شد (۸).

برای بررسی ارتباط دو متغیر، معمولاً رگرسیون حداقل مربعات به کار گرفته می شود. با رگرسیون حداقل مربعات ارتباط بین متغیرهای کمکی و میانگین پاسخ شرطی را می توان برآورد کرد، اما مثل بسیاری از روش های آماری که دارای محدودیت هایی می باشند، از جمله مفروضاتی برای معتبر بودنشان، رگرسیون حداقل مربعات در صورت وجود داده های پرت یا برقرار نبودن فرض نرمال بودن داده ها، برای تخمین ارتباط بین متغیر کمکی و متغیر پاسخ اعتبار کافی نخواهد داشت. از طرفی رگرسیون حداقل مربعات تنها ارتباط بین متغیرهای کمکی و میانگین پاسخ را بررسی می کند، در حالی که در بسیاری موارد هدف پیدا کردن ارتباط بین متغیرهای کمکی با سایر بخش های توزیع به ویژه چندک های انتهایی توزیع می باشد. لذا روشی جایگزین برای روش حداقل مربعات بسیار ارزشمند می باشد که رگر $\hat{\beta}(\tau)$ کی (Quantile Regression) یکی از این روش ها می باشد

$$\tau) = \min_{\beta \in R^p} \left[\sum_{i \in \{i: y_i \geq x_i' \beta\}} \tau |y_i - x_i' \beta(\tau)| + \sum_{i \in \{i: y_i < x_i' \beta\}} (1 - \tau) |y_i - x_i' \beta(\tau)| \right]$$

$$= \min_{\beta \in R^p} \sum_{i=1}^n \rho_\tau(y_i - x_i' \beta(\tau))$$

جدول ۲. معیار اطلاع آکائیک (AIC) در مدل های برازش داده شده

رگرسیون جداقل مربعات	۹۵	۹۰	۸۰	۷۰	۶۰	۵۰	۴۰	۳۰	۲۰	۱۰	۵	صدک
	۱۸	۷۰	۹۸	۳۱۵۲/۵۸	۱۵۵	۱۱۳	۲۸۰۷/۰۶	۲۷۴۰/۲۷	۲۶۹۳/۴۶	۲۶۶۵/۰۴	۴۰	۲۶۶۸/۸۰
	۳۰۸۲	۳۶۰۸	۳۴۰۹		۳۰۰۲	۲۸۹۱					۲۶۵۳	

بحث

نمودار هیستوگرام چوله به راست بودن داده های بروز سل را نشان می دهد که حتی با تبدیل نیز این چولگی باقی ماند. به دلیل ماهیت چولگی متغیر بروز سل استفاده از مدل رگرسیون جداقل مربعات مناسب نمی باشد، یکی از راهکارها برای بررسی اثر مهاجرت روی داده های چوله بروز سل، استفاده از رگرسیون چندکی است که علاوه بر نیرومند بودن، این امکان را فراهم می کند که در صدک های مختلف، ارتباط میزان بروز سل و نرخ مهاجرت را بررسی کنیم. استفاده از معیار اطلاع آکائیک نیز نیکویی برازش مدل رگرسیون چندکی را در اکثر صدک ها تایید می کند.

به جز صدک ۵ ام که دارای شیب منفی بود، مثبت بودن بقیه شیب ها نشان می دهد که با افزایش نرخ مهاجرت، میزان بروز سل نیز افزایش می یابد. لئوپولد و هم کارانش (۴) و تالبوت و هم کارانش (۶) و گیلبرت و هم کارانش (۷) به نتیجه مشابهی دست یافتند. اما تفاوتی که این مطالعه با مطالعات مشابه دارد در این است که در اکثر مطالعات مشابه، ارتباط میزان بروز سل و نرخ مهاجرت در بین افراد متولد شده کشورها و مهاجران خارجی بررسی شده، بجز کار لئوپولد و هم کارانش که ارتباط نرخ مهاجرت و بروز سل را در بین بومیان ساکن در سائوپولو و مهاجران از دیگر شهرهای برزیل و ساکن در سائوپولو را هم بررسی کردند، و به این نتیجه رسیدند که میزان بروز سل در این مهاجران ۷ برابر می باشد.

مطالعه حاضر بین بروز سل و نرخ مهاجرت در صدک های ۵۰، ۶۰، ۷۰، ۸۰، ۹۵ ارتباط معنی داری را نشان داد. تفسیر ضرایب رگرسیون چندکی، مشابه با رگرسیون جداقل مربعات می باشد، برای شهرستان مرو تپه استان گلستان و حاجی آباد هرمزگان که نرخ مهاجرت آن ها در صدک ۵۰ ام (۰/۱۱-) قرار داشته است، به ازای یک واحد افزایش در نرخ مهاجرت، میزان بروز سل به میزان ۰/۲۶ افزایش داشته است. برای شهرستان کبودرآهنگ استان همدان که نرخ مهاجرت آن در صدک ۶۰ ام (۰/۸۳) قرار داشته است، به ازای یک واحد افزایش در نرخ مهاجرت، میزان بروز سل به میزان ۰/۴۲ افزایش داشته است. برای شهرستان شهرضا اصفهان که نرخ مهاجرت آن در صدک ۸۰ ام (۴/۱۲) قرار داشته است، به ازای یک واحد افزایش در نرخ مهاجرت، میزان بروز سل به میزان ۰/۵ افزایش داشته است. برای شهرستا-

ن جم بوشهر که نرخ مهاجرت آن در صدک ۹۵ ام (۹/۹۱) قرار داشته است، به ازای یک واحد افزایش در نرخ مهاجرت، میزان بروز سل به میزان ۱/۹۱ افزایش داشته است. ویژگی مهم رگرسیون چندکی این است که این ارتباط را برای همه شهرستان ها و برای تمام نقاط صدکی می توان برآورد کرد. رگرسیون چندکی در کنار نقاط قوت، محدودیت هایی هم دارد، محدودیت اول آن این است که به صورت غیر واقعی چندکها را خطی فرض می کند و این باعث می شود که تعیین متغیر پیش گوی مهم و مناسب مشکل شود و دیگر اینکه، وقتی تعداد متغیرهای پیش گو خیلی زیاد یا کم باشند، تقاطع در چندکها ممکن است، به وجود بیاید (۱۵). شناسایی و تعیین عواملی که نرخ ابتلا به سل را افزایش می دهد. می تواند گامی بلند در جهت بهبود عوامل و پیشگیری از بیماری و در مجموع کاهش بار ناشی از بیماری ها بر فرد و نظام سلامت باشد.

پیشنهادات

با توجه به مدل برازش داده شده، ارتباط مهاجرت و میزان بروز سل مشخص شد، اما با توجه به اینکه مطالعه مقطعی می باشد و مربوط به سال ۱۳۸۹ است نیاز به مطالعات گسترده برای تایید این روابط وجود دارد. با توجه به واگیر بودن سل و امکان اپیدمی آن، تشخیص و کنترل علل و عوامل مرتبط با این بیماری در کاهش آن بسیار موثر می باشد، از جمله این عوامل نرخ مهاجرت است که در مطالعه حاضر بررسی گردید. با توجه به این مطالعه و مطالعات مشابه، کنترل بیشتر مسئولان در مهاجرت به داخل کشور از کشورهای همسایه بخصوص پاکستان و افغانستان، و همچنین مهاجرت افراد در داخل کشور از مناطقی با میزان بروز بالای بیماری سل، جهت کاهش بروز این بیماری در کشور، پیشنهاد می شود.

قدردانی

از مرکز آمار ایران و اداره کنترل بیماری سل و جذام معاونت بهداشت، بخاطر در اختیار قرار دادن داده های این پژوهش و از جناب آقای یوسف بشیری و جناب آقای دکتر شریفی و سرکارخانم احمدی همکاران این مراکز صمیمانه تشکر و قدردانی می کنیم. این مقاله برگرفته از پایان نامه کارشناسی ارشد آمارزیستی می باشد.

REFERENCES

1. Nasehi M, Mirhaghaani L. Guideline for control TB. 1th ed Tehran Andishmand Press; 2009:5.
2. EuroTB and the National Coordinators for TB Surveillance in the WHO European Region. Surveillance of tuberculosis in Europe: report on tuberculosis cases reported in 1997. In: Schwoebel V, Antoine D, Veen J, Jean F, editors. Saint-Maurice: Euro TB; 1999. p. 7–95.
3. Leopold J, Antunes F. (2001). The impact of AIDS, immigration and housing overcrowding on tuberculosis deaths in São Paulo, Brazil, 1994-1998. *Social Science & Medicine*. 52 (2001) 1071-1080.
4. Supevia A, Del Bano F, Estive E. (2009). Tuberculosis in an immigrant population: emergency department diagnoses, analyzed by place of origin. *Emergencias* 2009; 21: 410-414.
5. Talbot E, Moore M, McCary E, Binkin N. (2000). Tuberculosis among foreign-born persons in the United States, 1993-1998. *JAMA*, December 13, 2000-vol 284, No. 22.
6. Gilbert R, L., Antoine D, French C, E., Watson J, M., Jones J, A. The impact of immigration on tuberculosis rates in the United Kingdom compared with other European countries. *INT J TUBERC LUNG DIS* 13(5):645–651 © 2009 The Union.
7. McBryde E, Denholm J. (2012). Risk of active tuberculosis in immigrants: effects of age, region of origin and time since arrival in a low-exposure setting. *MJA* 2012; 197: 458–461.
8. Buchinsky M. Recent Advances in Quantile Regression Models: A Practical Guideline for Empirical Research. *The Journal of Human Resources*. 1998; 33(1): 88-126.
9. Melly B. The Theory and Practice of Quantile Regression [dissertation]. University of St Gallen, 2001. p. 1-30.
10. Koenker R, Bassett G. Regression Quantiles. *Econometrica*. 1978; 46(1): 33-50.
11. Manski C. F. (1975), "Maximum Score Estimation of the Stochastic Utility Model of choice," *Journal of Econometrics*, 3, 205-228.
-(1985), "Semiparametric Analysis of Discrete Response: Asymptotic Properties of Maximum Score Estimator," *Journal of Econometric*, 27, 313-333.
12. Powell J L. Least absolute deviations estimation for the censored regression model. *Journal of Econometrics*. 1984; 25(3): 303-325.
-Censored regression quantiles. *Journal of Econometrics*. 1986; 32(1): 143-155.
13. Machado J, Santos Silva J. Quantiles for Counts. *Journal of the American Statistical Association*. 2005; 100(472): 1226-1237.
14. Bjornar Bremnes J. Probabilistic Forecasts of Precipitation in Terms of Quantiles Using NWP Model Output. *Monthly Weather Review* 2004; 132: 338-47.