پیش‌بینی سری زمانی تعداد معلولیت‌های مربوط به حوادث ناشی از کار برای بیمه شدگان تأمین اجتماعی بین سال‌های ۱۳۷۹ تا ۱۳۸۹ در ایران با استفاده از روش تحلیل باکس جنکینز

مهدی ایمانی: کارشناسی ارشد آمار زیستی، گروه آمار و ریاضی، دانشکده مدیریت و اطلاع رسانی پزشکی، دانشگاه علوم پزشکی تهران، تهران، ایران.istat@live.com
***دکترمسعود صالحی:** استادیار آمار زیستی، گروه آمار و ریاضی، مرکز تحقیقات علوم مدیریت و اقتصاد سلامت دانشکده مدیریت و اطلاع رسانی پزشکی، دانشگاه علوم پزشکی تهران، تهران، ایران. (*نویسنده مسئول) salehi74@yahoo.com
آغا فاطمه حسینی: کارشناس ارشد و مربی آمار زیستی، گروه آمار و ریاضی، دانشکده مدیریت و اطلاع رسانی پزشکی، دانشگاه علوم پزشکی تهران، تهران، ایران.
hosseini_f@tums.ac.ir

تاریخ دریافت: ۹۰/۱۲/۱۷ تاریخ پذیرش: ۹۱/۳/۲۰

چکیده

زمینه و هدف: کنترل وقوع حوادث در محل کار موضوعی است که در تمامی کشورهای جهان مورد توجه قرار گرفته است. پیامدهای مالی وقوع این حوادث و زیان‌های اقتصادی که از این طریق به شرکت‌های مربوطه تحمیل می‌شود، تنها یکی از جنبه‌های کوچک این آسیب است؛ حال آنکه با در نظر گرفتن آسیب‌های غیر اقتصادی اما نامحسوس به جامعه، این گونه ضررها در حاشیه قرار خواهند گرفت. هدف از انجام این پژوهش برازش مدل باکس جنکینز(Box-Jenkins Model) به سری زمانی تعداد معلولیت‌های مربوط به حوادث ناشی از کار و برآورد مقادیر گمشده سری حین برازش این مدل می‌باشد.

روش کار: این پژوهش طولی (از نوع سری های زمانی) به منظور مدل‌بندی، برآورد و پیش‌بینی سری زمانی تعداد معلولیت‌های مربوط به حوادث ناشی از کار ,برای بیمه‌شدگان سازمان تأمین اجتماعی بین سال‌های ۱۳۷۹ تا ۱۳۸۹ خورشیدی در کشور ایران انجام شده است. دو روش هموارسازی اسپلاین (Spline Smoothing) به منظور درون‌یابی و برآورد مقادیر گمشده سری و روش باکس جنکینز به منظور انجام بهترین پیش‌بینی روی سری مورد استفاده قرار گرفته‌اند.

یافته‌ها: روش هموارسازی اسپلاین با استفاده از تعدیلاتی برای سری‌های فصلی, به منظور برآورد مقادیر گمشده, عملکردی قابل قبول روی سری مذکور نشان داد. به کار گیری تحلیل باکس جنکینز نشان داد که مدل سریزمانی مناسب برای داده‌های معلولیت، به صورت $SARIMA(0,1,1) \times (0,1,1)_{12}$ است. این مدل بیان می‌کند که سری زمانی معلولیت‌های ناشی از کار دارای اثر فصلی یک ساله، با میانگین متحرک مرتبه اول برای هر دو صورت معمولی و فصلی است.

نتیجه‌گیری: نحوه برآورد داده‌های گمشده در این سری زمانی با در نظر گرفتن خصوصیاتی مانند طویل بودن بازه داده‌های گمشده و اینکه این مقادیر پشت سر هم و بدون نقاط میانی قابل اتکا در سری اتفاق افتاده اند، روش خوبی برای برون رفت از حالت وجود چنین فواصلی در داده ها است. با نگاهی به شاخص‌های نیکویی برازش مدل می‌توان گفت که خطای برآورد و پیش‌بینی مدل نسبتاً پایین است و بنابراین مدل به خوبی به سری برازش داده شده است و بنابراین می‌توان نتیجه گرفت که مدل باکس جنکینز مدلی قابل اتکا برای برازش به داده‌های مشابه خواهد بود.

کلیدواژه‌ها: حوادث ناشی از کار، مقادیر گمشده، هموارسازی اسپلاین، سری‌های زمانی، مدل‌های باکس جنکینز.

مقدمه

فرآیند صنعتی شدن پدیده‌ای جبری است که بشریت را تحت سلطه خویش گرفته است و چنانچه قواعد بازی در آن به درستی رعایت نشود و کل فرآیند، قانونمند نگردد، تبعات سوء این سرنوشت محتوم، یکی پس از دیگری نمایان می شود. اما چنانچه با صنعت به طور آگاهانه برخورد شود و شناخت کافی از زیر و بم های آن به دست آید، جبر حاکم بر آن، به اختیار بدل می‌شود(۱).

حوادث ناشی از کار در کلیه کشورهای جهان رخ می‌دهند. بر اساس آماری که سازمان بین‌المللی کار ارائه داده است، روزانه پنج هزار نفر از کارگران جهان بر اثر حوادث ناشی از کار جان خود را از دست می‌دهند و سالانه تقریباً ۲۷۰ میلیون فقره از حوادث گوناگون در محل کار رخ می‌دهد که

www.SID.ir

۳۳۵ مورد آن کشنده است. این اطلاعیه زیان اقتصادی ناشی از این حوادث را سالانه یک تریلیون و دویست و پنجاه میلیارد دلار اعلام کرده است (۲). آمار حوادث در کشور ما نیز بالاست. از این رو شایسته است که ابعاد این معضل اجتماعی شناسایی و از گسترش قلمرو آن جلوگیری شود. با کسب شناخت و آگاهی از میزان حوادث ناشی از کار و اثرات آن، زمینه لازم برای سیاست‌گذاری‌های اصولی و مؤثر برای پیشگیری از حوادث و در نتیجه کاهش میزان آن ها فراهم خواهد آمد (۱).

حوادث ناشی از کار طبق آخرین تعریف سازمان تأمین اجتماعی ایران عبارت از حوادثی است که درحین انجام وظیفه و به سبب آن برای بیمه شده اتفاق می افتد. براساس این تعریف حوادثی که به یکی از صورت‌ها یا علل زیر رخ دهد ناشی ازکارمحسوب می شود:

- دراوقاتی که بیمه شده درکارگاه یا مؤسسات وابسته یا ساختمان ها ومحوطه آن مشغول به کارباشد.
- درزمانی که به صورت مامور درخارج از کارگاه انجام وظیفه می‌کند.
- دراوقات عادی رفت و برگشت بیمه شده بین منزل وکارگاه.
- دراوقات مراجعه به درمانگاه یا بیمارستان بابت معالجات درمانی وتوانبخشی.
- درحین اقدام برای نجات سایر بیمه شدگان ومساعدت به آنان (۳).

بیشترین تعداد افراد حادثه دیده در محدوده سنی ۲۰ تا ۲۹ سالگی قرار دارند. این نتیجه به علت بالا بودن تعداد شاغلان کشور در این محدوده سنی است؛ که فعال‌ترین قشر تولید کننده کشور را تشکیل می‌دهند (۴). عواملی مانند کم سوادی و بی‌تجربگی، بی باکی بیش از حد و بی‌احتیاطی از جمله وسایل بروز حوادث در سنین پایین است. با افزایش سن و به علت افزایش تجربه کاری و محتاط‌تر شدن شاغلان، تعداد حوادث نیز کاهش می یابد (۵). افراد کم سواد یا بی سواد بیشتر از افراد با تحصیلات بالا دچار حادثه می شوند. تعداد حادثه دیدگان متأهل بیشتر از حادثه دیدگان مجرد است که دلیل عمده این امر فزونی تعداد کارکنان متاهل بر کارکنان مجرد در محیط های کاری است (۳).

اطلاعات حوادث ناشی از کار که معمولاً در بسیاری از کشورها به صورت سالانه انتشار می‌یابند، در همهٔ نقاط جهان استاندارد نیست. به خصوص در کشورهای در حال توسعه اطلاعات قابل اتکایی در مورد حوادث ناشی از کار به علت نبود سیستم ثبت و اطلاع رسانی درست وجود ندارد. در این کشورها تعداد حوادث معمولاً کمتر از مقدار واقعی ثبت می‌شوند؛ اما همین اعداد هم هنوز به عنوان زیربنایی برای بررسی امنیت شغلی استفاده می شوند (۶-۸). تنها اطلاع از علت وقوع چنین حوادثی برای کاهش خطر آن ها کافی نیست. داده‌های مورد استفاده در این پژوهش نیز، از این قاعده مستثنا نیستند و در برخی ماه های سال و در بعضی سال ها دارای مقادیر گمشده یا مقادیر کمتر از حد انتظار بوده اند ,که با استفاده از روش‌های دقیق آماری برآورد شده‌اند.

با توجه به این که تا کنون پژوهشی در خصوص بررسی روند رخداد معلولیت های حوادث ناشی از کار در ایران انجام نشده است، هدف از انجام این پژوهش مدل‌بندی سری زمانی تعداد این معلولیت ها به منظور ایجاد دورنمای کلی از روند وقوع آن ها و پیشنهاد روش های پیشگیری در راستای کاهش وقوع معلولیت‌ها در سال‌های آتی است.

روش بررسی

این پژوهش یک مطالعه طولی از نوع سری زمانی است. جامعه آماری این پژوهش را کلیه افراد بیمه شده، شاغل در حرف متفاوت تحت پوشش سازمان تامین اجتماعی، که طی سالیان مختلف برای آن ها حوادث ناشی از کار رخ داده است, تشکیل می‌دهند.

نمونه مورد بررسی این تحقیق شامل افرادی است که با حادثه ناشی از کار مواجه شده (دچار معلولیت دائم یا موقتی شده‌اند)، خود فرد یا بازمانده برای دریافت غرامت به شعب مختلف سازمان تامین اجتماعی در سراسر کشور طی

Archive of SID
www.SID.ir

سال‌های ۱۳۷۹ تا ۱۳۸۹ مراجعه کرده و اطلاعات آنان در سیستم مکانیزه بیمه سازمان تامین اجتماعی ثبت شده است. به عبارت دیگر از داده های یک مقطع زمانی یازده ساله از این جامعه به عنوان نمونه استفاده شده است. داده‌های این پژوهش از کارگاه‌های تحت پوشش سازمان تأمین اجتماعی ایران که در دفتر آمار و محاسبات اقتصادی و اجتماعی این سازمان موجود است، حاصل شده‌اند.

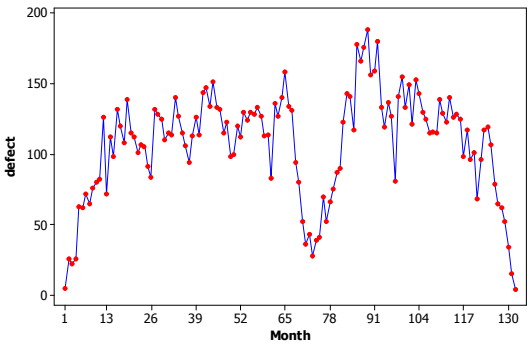
اطلاعات مورد استفاده در این مطالعه شامل متغیرهایی هستند که به صورت زیر تعریف می‌شوند:

- تعداد معلولیت ها: عبارت است از تعداد موارد معلولیت ناشی از کار ثبت شده در سازمان تامین اجتماعی است که به صورت ماهیانه تجمیع شده اند. در اینجا با استفاده از فراوانی تعداد مشاهدات منجر به معلولیت دایم، کلی یا جزیی در هر ماه، عددی خاص به هر ماه از سال تخصیص یافته است. این عدد تعداد معلولیت‌ها در آن ماه می‌باشد.
- زمان: منظور از متغیر زمان در این مطالعه، هر یک از ماه‌های مختلف سال‌های مورد مطالعه است.
- داده‌های گمشده: در این پژوهش عبارت «گمشده» به فراوانی تعداد حوادث در ماهی گفته می شود که کاهش زیادی را نشان داده و از قابلیت اطمینان بالایی برخوردار نباشند. بنابراین در برخورد با این مشاهدات، آن‌ها را از مطالعه کنار گذاشته و به عنوان داده ثبت نشده، برآورد گردیده‌اند.

داده ها در تمامی طول دوره یازده ساله مورد بررسی، به صورت روزانه ثبت شده‌اند و با شمارش تعداد حوادث در هر ماه، سری زمانی تعداد رخداد ماهانه حوادث به صورت شکل ۱ ترسیم می‌شوند.

همان طور که دیده می‌شود، داده‌ها حوالی سال ۱۳۸۵ و اواخر سال ۱۳۸۹ دارای مقادیر خارج از حدود انتظار می باشند. بنابراین به عنوان مقادیر ثبت نشده، مورد برآورد قرار گرفته شد.

روش های آماری: برازش مدل باکس جنکینز به داده‌ها شامل دو مرحله می‌باشد. ابتدا مقادیر



شکل ۱- سری زمانی خام تعداد معلولیت‌ها در دوره ۱۱ ساله مورد بررسی

گمشده در داده‌ها برآورد شده و سپس مدل مذکور به سری برآورد شده، برازش داده شده است.

برآورد مقادیر گمشده: برآورد مقادیر گمشده در سری‌های زمانی با استفاده از روش‌های هموارسازی رگرسیونی ناپارامتری قابل انجام است. اگرچه در استفاده از روش های رگرسیونی در مورد سری‌های زمانی که در آن مشاهدات بر خلاف داده های رگرسیونی از یکدیگر مستقل نیستند و خصوصیت خودهمبستگی (Autocorrelation) در آن ها وجود دارد، نیاز به تعدیلاتی جهت مناسبت برای چنین داده‌هایی احساس می‌شود. روش استفاده شده برای برآورد مقادیر در این پژوهش، هموارسازی اسپلاین درجه سوم Cubic Spline (Smoothing) برای داده‌های خودهم بسته است. روش اسپلاین, منحنی اسپلاین درجه سومی را به مشاهدات برازش می‌دهد. یک اسپلاین درجه سوم، قسمتی از یک تابع شامل توابع توأم چندجمله‌ای درجه سوم است که در آن تابع، مشتق اول و مشتق دوم آن، هر سه پیوسته هستند (۹-۱۱).

سری مورد بررسی در این پژوهش ,تغییرات فصلی آشکاری را داراست. بنابراین در برآورد مقادیر گمشده، باید این خصوصیت سری در نظر گرفته شود. اگر x_t یک سری زمانی فصلی با s مشاهده در هر دوره فصلی باشد، روش‌های کلاسیک تجزیه سری را به چهار مؤلفه تقسیم خواهند نمود: روند (Trend)، تغییرات دوره‌ای (Cyclic component)، تغییرات فصلی

Archive of SID
www.SID.ir

(Seasonal Component) و تغییرات تصادفی (Irregular Component). روند و تغییرات دوره‌ای معمولاً به شکل مؤلفه روند-دوره ترکیب می‌شوند. شکل ابتدایی تجزیه کلاسیک سری به صورت زیر است:

$$x_t = TC_t S_t I_t \qquad (1)$$

که در آن TC_t مؤلفه روند-دوره، S_t مؤلفه فصلی با دوره s با میانگین یک و I_t مؤلفه تصادفی با میانگین یک فرض شده‌اند.

نقاط درونی بازه داده‌ها با استفاده از عملگر سری زمانی فصلی تعدیل یافته, برآورد شده است. پس از تجزیه، سری زمانی اصلی را می‌توان به شکل تعدیل یافته فصلی زیر نشان داد:

$$\tilde{x}_t = x_t / S_t = TC_t I_t \qquad (2)$$

برای نقاط انتهای بازه مشاهدات سری، عملگر پنجره زمانی متحرک استفاده شده است. این عملگر آماره‌ها را برای مجموعه‌ای از مقادیر در یک پنجره زمانی متحرک محاسبه می‌نماید. عملگر پنجره زمانی متحرک مرکزی که در اینجا مورد استفاده قرار گرفته است، آماره مورد نظر را برای n مقدار x_i برای مشاهدات $t - (n-1)/2 \le i \le t + (n-1)/2$ محاسبه می‌نماید. در اینجا پنجره مورد نظر برای ۲۵ مشاهده انتخاب شده است تا دو دوره فصلی در پنجره جای گیرد (۱۲-۱۵).

در برازش هموارساز اسپلاین به داده‌های سری زمانی، تبدیلات زیادی وجود دارد. با توجه به ماهیت سری زمانی مورد بررسی در این پژوهش، از میان انبوه تبدیلات موجود، دو تبدیل تجزیه کلاسیک سری های فصلی تعدیل یافته و میانگین متحرک مرکزی انتخاب شده اند. دلیل این انتخاب، روند فصلی موجود در سری مذکور و ضرورت برآورد داده های گمشده بر طبق این روند فصلی، و همین طور داده‌های از دست رفته سه ماهه آخر سال ۱۳۸۹ و عدم توانایی تبدیل فصلی در برآورد داده های گمشده در انتهای سری ها بوده است که ناچار به استفاده از تبدیل میانگین متحرک مرکزی برای برآورد سه ماهه آخر سری شده‌ایم.

تحلیل باکس جنکینز: مهم ترین مکانیسم‌هایی که می توانند برای تولید داده ها در نظر گرفته شوند، فرآیندهای میانگین متحرک (Moving average)، فرآیندهای خودبازگشت (Autoregressive) و ترکیب این دو فرآیند می‌باشند. سری‌هایی که خصوصیت هر دو فرآیند را دارا باشند، با فرایند میانگین متحرک خودبازگشتی (Auto Regressive Moving Average) شناخته می شوند. مدل این فرآیند را می‌توان به صورت:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} \qquad (3)$$

نشان داد. در این مورد، هر دو فرآیند میانگین متحرک و خودبازگشت از مرتبه ۲ هستند.

با توجه به اینکه روش باکس جنکینز، تحلیل در دامنه زمانی داده های سری های ایستا است، لازم است تا ناایستایی سری مورد بررسی قرار گیرد و در صورت ناایستا بودن آن، تبدیلات مورد نیاز روی سری اعمال شود. آزمون‌های دیکی فولر تعمیم یافته (Augmented Dickey-Fuller Test) و دیکی فولر فصلی (Seasonal Dickey-Fuller Test)، آزمون‌هایی هستند که بدین منظور سری را مورد آزمایش قرار می دهند. پدیده ناایستایی معمولاً در اثر وجود پنج عامل ایجاد می شود: مقادیر پرت (Outliers)، گام تصادفی (Walk Random)، کشیدگی(Drift)، روند و تغییرات واریانس؛ که هر یک تبدیلات مخصوص به خود را برای بی‌اثر شدن در سری می‌طلبد. همچنین ناایستایی فصلی، عامل دیگری است که باید در کنار دیگر عوامل مورد بررسی قرار گیرد (۱۶و۱۷).

پس از بررسی سری از لحاظ ایستایی، پارامترهای مدل برآورد می شود. این امر با استفاده از نمودارهای همبستگی نگار (Auto Correlation Function) ACF و (Partial Auto Correlation Function) PACF و همین طور برآوردگرهای دقیق تر مدل قابل انجام است. به این ترتیب فرآیند مناسب برای برازش به سری زمانی انتخاب می‌شود.

مرحله آخر مربوط به پیش‌بینی سری زمانی برای محدوده زمانی معینی در انتهای بازه داده‌ها است. در اینجا مقادیر مورد انتظار برای تعداد

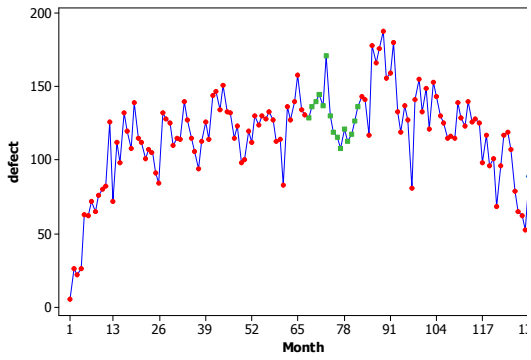
Archive of SID
www.SID.ir

حوادث منجر به معلولیت که در ماه های سال ۱۳۹۰ رخ خواهد داد، مورد پیش‌بینی قرار می‌گیرند.

محدودیت های آماری پژوهش: سری زمانی مورد بررسی در این پژوهش دارای داده های گمشده فراوانی است. منظور از گمشده بودن داده ها در یک ماه خاص، عدم وجود تعداد مشاهدات ثبت شده در آن ماه نیست؛ بلکه در حوالی سال ۱۳۸۵، سیستم ثبت داده‌های سازمان تأمین اجتماعی دستخوش تغییرات بنیادی شده که پیامد آن، عدم ثبت دقیق و کامل موارد رخداد حادثه در بسیاری از استان‌ها بوده است. بنابراین، تعداد مشاهدات در این بازه زمانی بسیار کمتر از تعداد واقعی آن‌ها به ثبت رسیده است. این داده ها با روش های بیان شده در بندهای قبلی، برآورد شده و برای ساخت و تحلیل سری زمانی معلولیت ها به کار گرفته شده اند.

یافته‌ها

سری زمانی تعداد حوادث منجر به معلولیت (همان طور که انتظار می‌رود) دارای روند و تغییرات فصلی مشهود می‌باشد. بنابراین خودهمبستگی در این سری زمانی بسیار قوی است. در اینجا رویکرد رگرسیونی محض به داده‌ها بدون در نظر گرفتن نقش خودهمبستگی بین مشاهدات صحیح نمی‌باشد. از طرفی همان طور که



شکل ۲–سری زمانی تعداد معلولیت‌ها پس از حذف و برآورد مقادیر از دست رفته. (در این شکل، نقاط سبز رنگ، برآوردهای تبدیل فصلی و نقاط آبی رنگ، برآوردهای تبدیل میانگین متحرک مرکزی هستند.)

دیده می‌شود و پیش از این نیز ذکر شد، داده‌ها در سال ۱۳۸۵ به هیچ وجه قابل استناد نیستند (دقیقاً بین ماه های آبان ۱۳۸۴ تا دی ۱۳۸۵). بدیهی است در برازش روش اسپلاین به داده ها، نباید نقش اثرات فصلی موجود در سری زمانی را برای برازش داده های از دست رفته، نادیده گرفت. در غیر این صورت برآورد داده ها در بازه زمانی مذکور به صورت کاملاً خطی و رگرسیونی محض خواهد بود.

در شکل ۲ نقاط برآورد شده مقادیر سال ۱۳۸۵ به کمک روش اسپلاین در سری زمانی اولیه جایگزین شده است.

در برازش مدل های باکس جنکینز به سری های زمانی، به طور کلی نیازمند تفکیک روند برازش به سه مرحله هستیم: شناسایی، برآورد و پیش‌بینی. هر مرحله دارای خصوصیات مخصوص به خود است که در زیر به آن ها پرداخته شده است.

مرحله شناسایی با اطلاعات توصیفی مختصر از سری آغاز شده است. سری زمانی مورد بررسی در این پژوهش، دارای میانگین ۱۱۶/۴۳۹۷ مورد معلولیت در هر ماه و انحراف از معیار ۳۰/۸۹۲۱۴ است. آزمون های خودهمبستگی مربوط به لزوم برازش مدل های باکس جنکینز به سری برای ۲۴ فاصله زمانی ابتدایی سری، همگی مقادیر احتمال معنی‌دار داشته‌اند. بنابراین خودهمبستگی در سری اثبات و استفاده از مدل های باکس جنکینز مناسب به نظر می‌رسد.

جدول ۱مربوط به آزمون های دیکی فولر و دیکی فولر فصلی، قبل و بعد از اختلاف گذاری های مورد نیاز روی سری است.

جدول ۱ شامل ۴ بخش است. این بخش‌ها به ترتیب مقادیر مربوط به آزمون‌های دیکی فولر و دیکی فولر فصلی را قبل و بعد از اختلاف‌گذاری‌های مورد نیاز ,نشان می‌دهند. همان طور که در این جدول دیده می‌شود، وجود گام تصادفی در سری، عاملی است که باعث ناایستایی آن شده است. ضمنا سری از لحاظ فصلی نیز ناپایدار است. این ناایستایی از روی مقادیر احتمال مربوط به این دو آزمون (دو بخش اول جدول) تأیید شده است.

Archive of SID
www.SID.ir

جدول ۱- مقادیر آماره‌های مربوط به آزمون دیکی فولر و دیکی فولر فصلی، قبل و بعد از اختلاف گذاری

		لگ	همبستگی	مقدار احتمال	T	مقدار احتمال	F	Pr>F
مقادیر آزمون دیکی فولر پیش از	گام تصادفی	۰	-۱/۶۱۹۳	۰/۳۷۷۴	-۰/۸۴	۰/۳۵۳۴		
اختلاف‌گذاری	کشیدگی	۰	-۳۴/۶۷۵۸	۰/۰۰۱۱	-۴/۹۷	۰/۰۰۰۱	۱۲/۴۰	۰/۰۰۱۰
	روند	۰	-۳۵/۲۴۲۷	۰/۰۰۱۵	-۴/۸۲	۰/۰۰۰۷	۱۲/۲۸	۰/۰۰۱۰
مقادیر آزمون دیکی فولر فصلی پیش از	گام تصادفی	۰	-۲/۹۶۶۶	۰/۳۳۷۹	-۱/۰۲	۰/۱۸۶۴		
اختلاف‌گذاری	کشیدگی	۰	-۹۲/۷۴۵۴	۰/۰۰۱۳	-۱۰/۹۲	<۰/۰۰۰۱		
مقادیر آزمون دیکی فولر پس از	گام تصادفی	۰	-۱۶۶/۲۴۶	۰/۰۰۰۱	-۱۶/۷۴	<۰/۰۰۰۱		
اختلاف‌گذاری	کشیدگی	۰	-۱۶۶/۳۶۳	۰/۰۰۰۱	-۱۶/۶۹	<۰/۰۰۰۱	۱۳۹/۲۸	۰/۰۰۱۰
	روند	۰	-۱۶۶/۵۳۱	۰/۰۰۰۱	-۱۶/۶۵	<۰/۰۰۰۱	۱۳۸/۶۵	۰/۰۰۱۰
مقادیر آزمون دیکی فولر فصلی پس از	گام تصادفی	۰	-۱۶۰/۵۸۸	۰/۰۰۰۱	-۱۸/۵۴	<۰/۰۰۰۱		
اختلاف‌گذاری	کشیدگی	۰	-۱۶۰/۶۶۴	۰/۰۰۰۱	-۱۸/۴۶	<۰/۰۰۰۱		

جدول ۲- آزمون‌های مربوط به برابری مقادیر پارامتر مدل با مقدار صفر

پارامتر	برآورد	خطای استاندارد	T	مقدارp-	لگ
میانگین	-۰/۶۷۵۶۱	۰/۳۵۳۹۲	-۱/۹۱	۰/۰۵۸۷	۰
MA1,1	۰/۴۶۶۰۱	۰/۰۸۲۸۸	۵/۶۲	<۰/۰۰۰۱	۱
MA2,1	۰/۷۱۸۳۳	۰/۰۷۰۴۷	۱۰/۱۹	<۰/۰۰۰۱	۱۲

مقادیر احتمال در دو قسمت بعدی در جدول نشان می‌دهند که سری پس از یک مرحله اختلاف‌گذاری معمولی و یک مرحله اختلاف‌گذاری فصلی به ایستایی مورد نیاز برای برازش مدل باکس جنکینز رسیده است. بنابراین مقادیر d و D را در توابع I(d) و I(D) برابر یک در نظر می‌گیریم.

با استفاده از نمودارهای همبستگی نگار توابع خودهمبستگی و خودهمبستگی حاشیه‌ای، هم‌گرایی سری تأیید می‌شود. پارامترهای مدل در جدول ۲ برآورد شده است.

همان طور که در جدول بالا مشاهده می‌شود، رتبه های اول برای پارامترهای میانگین متحرک و میانگین متحرک فصلی به شدت تأیید می‌شود. آزمون t در تمامی سطرهای جدول, برابری برآورد پارامتر مورد نظر با مقدار صفر را بررسی می کند. مقادیر احتمال مربوط به برابری این مقادیر با مقدار صفر، فرض صفر را حتماًرد می‌نمایند. همچنین نحوه ورود مؤلفه فصلی به مدل, هنگامی که به صورت ضربی باشد، اندکی شاخص های نیکویی برازش مدل را بهبود می بخشد. بنابراین مدل نهایی برای برازش به سری زمانی، مدل:

$$SARIMA(0,1,1) \times (0,1,1)_{12}$$

می‌باشد.

جدول ۳ حاوی شاخص‌های نیکویی برازش برای مدل بالا است.

شاخص های مندرج در جدول بالا، بهترین حالت برازش ممکن مدل باکس جنکینز به سری زمانی مورد بررسی را نشان می دهد. همان طور که در بالا ذکر شد، هنگامی که مؤلفه فصلی به صورت ضربی وارد مدل شود، معیار اطلاع آکاییک :AIC Akaike information criterion را به مقدار ۲ واحد بهبود می‌بخشد.

به عنوان آخرین گام در برآورد مدل، مدل مناسب برای برازش به سری به صورت زیر به دست آمده است.

$$(1-B)(1-B^{12})Y_t = -0.67561 + (1-0.46601B)(1-0.71833B^{12})a_t$$

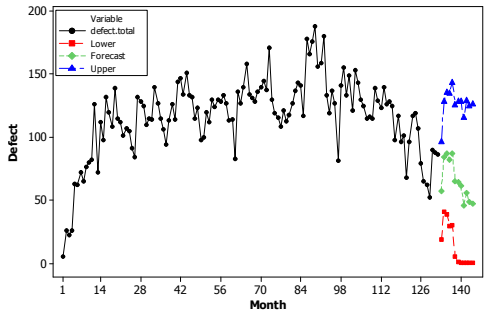
پس از مشخص شدن بهترین مدل برای برازش به سری و برآورد پارامترهای مربوط به آن، پیش بینی سری با استفاده از مدل تعیین شده برای یک سال آتی پس از آخرین مشاهده به صورت شکل ۳ است.

با توجه به پیش بینی سری زمانی که در شکل ۴ نشان داده شده است، روند نزولی‌ای که از سال ۱۳۸۶ آغاز شده است، در پیش‌بینی سری نیز مشهود می‌باشد. بنابراین تعداد حوادث ناشی از کار که به معلولیت افراد ختم شده است، پس از طی مسیری صعودی به سمت سال ۱۳۸۵، روندی

Archive of SID
www.SID.ir

جدول ۳- شاخص‌های نیکویی برازش مدل باکس جنکینز

برآورد ثابت	-۰/۶۷۵۶۱
واریانس برآورد	۳۹۰/۶۳۲۱
خطای استاندارد برآورد	۱۹/۷۶۴۴۲
AIC	۱۰۵۰/۸۳۳
SBC	۱۰۵۹/۱۷
هاتعداد باقی مانده	۱۱۹



شکل ۳- نمودار پیش‌بینی یک ساله سری زمانی تعداد معلولیت‌ها

نزولی را تا سال ۱۳۸۹ در پیش گرفته است. سپس در پیش‌بینی انجام شده نیز، مقادیر مورد انتظار برای سال ۱۳۹۰ کاهش قابل توجهی را نسبت به سال قبل از خود نشان می‌دهد.

بحث و نتیجه‌گیری

سری زمانی حوادث ناشی از کار منجر به معلولیت افراد، خصوصیات ویژه‌ای را نسبت به دیگر متغیرهای ممکن مورد بررسی در زمینه حوادث ناشی از کار داراست. یکی از جنبه‌های ویژه این سری این است که هنگامی که سخن از نتیجه این حوادث به میان می‌آید، عوامل دیگری به جز علل وقوع یک حادثه ناشی از کار نیز در ایجاد این گونه سری‌ها دخیل خواهند شد. نگرشی که در این پژوهش به این داده‌ها داشته‌ایم، نگرشی تک متغیره، صرفاً به منظور بررسی چگونگی روند تکامل این سری در طول زمان بوده است.

هنگامی که یک حادثه در محیط کار روی می‌دهد، بسیاری عوامل کوتاه مدت و بلند پس از آن، نتیجه این حادثه را رقم می‌زنند. عواملی که از همان لحظه پس از وقوع حادثه توسط اطرافیان فرد حادثه دیده آغاز و تا زمان گذراندن دوره نقاهت بیماری و حتی درمان‌های روانی پس از حادثه ادامه می‌یابد. در بخش دیگر مطالعه که تعداد حوادث ناشی از کار در طول زمان بررسی گردید، نتیجه این گونه بود که وقوع حوادث در این بازه ۱۱ ساله زمانی، روندی ثابت و یکنواخت داشته است. با در نظر گرفتن این واقعیت و نتایج به دست آمده در این مطالعه، روند نزولی حوادث منجر به معلولیت افراد، بسیار امیدوار کننده به نظر می رسد؛ زیرا با در نظر گرفتن سیر ثابت حوادث منجر به فوت، می‌توان گفت پس از سال ۱۳۸۵، تعداد حوادث ناشی از کار منجر به معلولیت، نسبت به مقادیر منجر به بهبودی کامل، کاهش قابل توجهی داشته است.

نحوه برآورد داده‌های گمشده در این سری زمانی با در نظر گرفتن خصوصیاتی مانند طویل بودن بازه داده‌های گمشده و اینکه این مقادیر پشت سر هم و بدون نقاط میانی قابل اتکا در سری اتفاق افتاده اند، روش خوبی برای برون رفت از حالت وجود چنین فواصلی در داده ها است. خصوصاً که چنین شکست‌هایی در سری‌های زمانی‌ای که نحوه ثبت و حتی ایده ثبت آن‌ها تازه ظهور و تکامل نیافته‌اند، بسیار اتفاق می‌افتند. همان طور که پیش از این ذکر شد، ثبت حوادث ناشی از کار در کشورهای در حال توسعه چندان دقیق نیست و احتمال وجود چنین اشتباهاتی در ثبت و نگهداری آن‌ها وجود دارد. بنابراین روش های برآورد چنین شکست هایی در داده ها باید دقیق باشد و گسترش این گونه روش‌ها می‌تواند در بالا بردن دقت برازش مدل‌های مورد نظر به داده‌ها بسیار مفید باشد.

با نگاهی به شاخص های نیکویی برازش مدل می‌توان گفت که خطای برآورد و پیش‌بینی مدل آن چنان پایین نیست، بنابراین نمی توان مدعی شد که مدل به خوبی به سری برازش داده شده است. اما از آنجا که این حالت بهترین حالت موجود برای برازش به داده ها است و همچنین جداول مربوط به آزمون خودهمبستگی نیز مدل ARIMA (Autoregressive Integrated Moving Average) را برای برازش به سری تأیید می کنند، نتایج به دست آمده از طریق برازش این مدل قابل اتکا هستند.

Archive of SID
www.SID.ir

می توان این تحقیق را به عنوان اولین نمونه تحلیل‌های آماری روی داده های حوادث ناشی از کار، چگونگی پیش‌بینی این داده‌ها و همین طور برآورد نقاط غیر قابل اتکا در آن‌ها در نظر گرفت. بنابراین، رویکرد دیگری به مقادیر پیش بینی شده سری زمانی در این مطالعه، تشخیص عوامل احتمالی در جهت کاهش هر چه بیشتر تعداد این معلولیت‌ها در طول زمان است. پیش‌بینی‌های انجام شده در این پژوهش را می توان به عنوان تأثیر استراتژی سلامت و ایجاد عوامل پیشگیرانه در زمینه حوادث ناشی از کار در یک دوره کوتاه مدت در ایران در نظر گرفت. عوامل زیادی در این مورد دخیل هستند، اما از آنجا که هیچگاه به طور دقیق در کشور بررسی و تایید نشده‌اند، با نگاهی به پژوهش های انجام شده در جهان، به چند مورد آن ها اشاره شده می شود:

- تشویق شرکت های کوچک و متوسط به ایجاد عوامل پیش گیرانه به منظور کاهش وقوع حوادث ناشی از کار.
- تبلیغ و توسعه قوانینی مبنی بر اجبار شرکت‌ها برای اضافه شدن بندهایی به قراردادها جهت جدی گرفتن پیشگیری از وقوع حوادث در محل کار.
- توسعه طرح هایی برای ارتقا سطح آگاهی کارگران کم سواد و بی‌سواد.
- در نظر گرفتن تسهیلات مالی و تشویق شرکت‌هایی که پیشگیری از وقوع حوادث ناشی از کار را در دستور کار خود قرار می‌دهند.
- ملزم نمودن کارگران به رعایت استانداردهای مورد نیاز کار و متعهد نمودن آن ها به عدم انجام فعالیت‌های پرخطر.
- ایجاد معیاری برای کیفیت فعالیت افراد، بر پایه عوامل انسانی و اقتصادی به منظور الگوسازی و آموزش انجام فعالیت بهینه در محل کار (۶).

طبق پیش بینی تعداد حوادث منجر به معلولیت در سال ۱۳۹۰، کاهش محسوسی در روند وقوع تعداد حوادث منجر به معلولیت دیده می شود. این امر به خودی خود نشان از عملکرد بهتر عوامل دخیل در درمان افراد حادثه دیده دارد،به خصوص اینکه این کاهش به نفع حوادث منجر به بهبودی کامل در سری به وجود آمده است. بنابراین می توان همین روند را در سال ۱۳۹۰ نیز ادامه داد و امید به کسب نتایجی مانند این داشت.

تقدیر و تشکر

این مقاله بخشی از پایان نامه آقای مهدی ایمانی در مقطع کارشناسی ارشد رشته آمار زیستی به راهنمایی آقای دکتر مسعود صالحی و مشاوره سرکار خانم آغا فاطمه حسینی در سال ۱۳۹۰ و کد 90-02-136-13712 می باشد که با حمایت دانشگاه علوم پزشکی و خدمات بهداشتی درمانی تهران اجرا شده است.

منابع

1. Zahedi SH. Ravabet e sanati, Nezam e ravabet e kar (ba takid bar havades e nashi az kar). Tehran: Markaz e nashr e daneshgahi; 1383. [Persian]
2. Benavides FG. Occupational injury, a public health priority. Revista espanola de salud publica. 2006; 80(5): 553-565.
3. Mosayyeb zadeh S. Havades e nashi az kar va rah haye pishgiri az borooz e anha. Tehran: Ketabkhaneye sazman e tamin e ejtemaei; 1389. [Persian]
4. Sanjari A. Amuzesh e hefazat e imeni, barasi e havades e nashi az kar. Tehran: Ketabkhaneye sazman e tamin e ejtemaei; 1389. [Persian]
5. Hamalainen P, Saarela K.L, Takala J. Global trend according to estimated number of occupational accidents and fatal work-related diseases at region and country level. Journal of safety research. 2009; 40: 125-139.
6. Carnero MC, Pedregal DJ. Modeling and forecasting occupational accidents of different severity levels in spain. Reliability engineering and system safety. 2010;95: 1134-41.
7. Bevilacqua M, Ciarapica FE, Giacchetta G. Industrial and occupational ergonomics in the petrochemical process industry: Aregression trees approach. Accident analysis and prevention 2008;40: 1468-79.
8. Burgherr P, Hirschberg S. Severe accident risks in fossil energy chains: A comparative analysis. Energy.2008; 33: 538-53.
9. Keden B, Fokianos K. Regression models for time series analysis. New Jersey: John Wiley & Sons Inc.; 2002.
10. Eubank RL. Nonparametric regression and spline smoothing. 2nd ed. New York: Marcel Dekker, Inc.; 1999.
11. MacNab YC, Dean CB. Autoregressive spatial

Archive of SID
www.SID.ir

smoothing and temporal spline smoothing for mapping rates. Biometrics. 2001; 57: 949-56.

12. SAS/STAT 9.2 User's Guide. 2nd ed. North Carolina: SAS Institute, Inc.; 2009.

13. Yaffee R, McGee M. Introduction to time series analysis and forecasting with applications of SAS and SPSS. New York: Academic Press, Inc.; 1999.

14. Fan J, Yao Q. Nonlinear time series: nonparametric and parametric methods. New York: Springer-Verlag.; 2003.

15. Gurrin LC, Scurrah KJ, Hazelton ML. Tutorial in biostatistics: spline smoothing with linear mixed models. Statistics in Medicine. 2005; 24: 3361-81.

16. Brockwell PJ, Davis RA. Time series: theory and methods. 2nd ed. New York, NY 10013, USA: Springer Science +Business Media, LLC.; 2006.

17. Maharaj EA. Comparison of non-stationary time series in the frequency domain. Computational Statistics & Data Analysis. 2002; 40: 131-41.

Archive of SID

www.SID.ir

Original Article

Forecasting number of work-related injuries time series with Box-Jenkins Models for registered insured in SSIO between 2000 and 2010 in Iran

Mehdi Imani, MSc**.** Biostatistics, Department of Statistics and Mathematics, School of Health Management and Information Sciences, Tehran University of Medical Sciences, Tehran, Iran. mail@sezox.com
***Masoud Salehi,** PhD**.** Assistant Professor of Biostatistics, Department of Statistics and Mathematics, School of Health Management and Information Sciences, Management Science and Health Economics Research Center, Tehran University of Medical Sciences, Tehran, Iran.(*Corresponding author). salehi74@yahoo.com
Agha Fatemeh Hosseini, MSc**.** Biostatistics, Department of Statistics and Mathematics, School of Health Management and Information Sciences, Tehran University of Medical Sciences, Tehran, Iran. hosseini_f@tums.ac.ir

Abstract

Background: Controlling occurrence of accidents in work place has been an interesting subject in all countries worldwide. Financial consequences of these accidents and their economic losses imposed on the involved companies is only one of the insignificant aspects of such damages and when the non-economic but intangible losses to the society are taken into consideration ,these economic damages will be marginalized. Purpose of this research is fitting the box-Jenkins model to time series of total number of accidents in work place and estimation of series' missing values during fitting of this model.

Methods: This longitudinal (time series) study, intends to model, estimate and forecast time series of total number of work place accidents for the insured people between 2000 and 2010 in Iran. Spline non-parametric regression methods to find the best interpolation and estimation of the series' missing value as well as box-Jenkins prediction method to find the best prediction on series have been used.

Results: Smoothing spline method using some adjusts for seasonal time series in order to estimate missing values, shows better performance on the series. Then, suitable box-Jenkins model, $\mathrm{SARIMA}(0,1,1) \times (0,1,1)_{12}$, was fitted to the series. Goodness of fit criterion of model, AIC, for prediction of the time series was equal to 1050.833.

Conclusion: Estimation of the missing data in time series with regard to characteristics such as long period of missing data and exposure of values in tandem without any reliable middle points, is an appropriate method in order to leave situation of such gaps in the data. Looking at the goodness of fit of the model can be said that the error estimation and prediction model is rather low and hence, the model is well fitted to the series. Therefore, box-Jenkins model will be reliable for fitting to similar data.

Keywords: Work-related accidents, Estimation of missing values, Spline Smoothing, Time series, Box-Jenkins models.

Archive of SID

www.SID.ir