

کنفرانس روش‌های مدرن
در قیمت‌گذاری‌های بیمه‌ای
و آمارهای صنعتی



مجموعه چکیده مقالات



کتابچه چکیده مقالات کنفرانس

روش‌های مدرن در قیمت‌گذاری‌های بیمه‌ای و آمارهای صنعتی

۱۲ تا ۱۴ شهریور ۱۳۹۶

دانشگاه بوعلی سینا

همدان

گردآوری و ویرایش

دکتر رحیم محمودوند، دانشگاه بوعلی سینا

دکتر فرزانه صفوی‌منش، دانشگاه شهید بهشتی

دکتر سید مرتضی نجیبی، دانشگاه شیراز

کمیته علمی کنفرانس

- دکتر ابراهیم امینی سرشت، دانشگاه بوعلی سینا
- پروفیسور هیربد آسا، دانشگاه لیورپول، انگلستان
- پروفیسور سورن اسموسن، دانشگاه آرهوس دانمارک
- دکتر طاهر اکین، دانشگاه ایالتی تگزاس، آمریکا
- دکتر ترزا اولیویرا، دانشگاه آبرتا، پرتغال
- دکتر مجید چهکندی، دانشگاه بیرجند
- دکتر پائولو رودریگوئز، دانشگاه ایالتی باہیا برزیل
- دکتر لاله سمرقندی، دانشگاه ریرسون، کانادا
- دکتر اصغر سیف، دانشگاه بوعلی سینا
- دکتر دیرک سورمان، دانشگاه فنی دورتموند، آلمان
- پروفیسور رفیق سویر، دانشگاه جورج واشینگتن آمریکا
- دکتر جواد فردمال، دانشگاه علوم پزشکی همدان
- دکتر لوکا فریگوا، دانشگاه کگلیری، ایتالیا
- دکتر مجید صادقی فر، دانشگاه بوعلی سینا
- دکتر فرزانه صفوی منش، دانشگاه شهید بهشتی
- پروفیسور احسان صوفی، دانشگاه ویسکانسین میلوواکی آمریکا
- پروفیسور بیژن ظهوری زنگنه، دانشگاه صنعتی شریف
- پروفیسور عادل محمدپور، دانشگاه صنعتی امیرکبیر
- دکتر رحیم محمودوند، دانشگاه بوعلی سینا
- دکتر ایوز لورنت گریز، بیمه بالویزه، بازل، سوئیس
- دکتر کریستینا لورز، شرکت پروگنیستا جی ام بی اچ آلمان
- دکتر ابراهیم نصیرالاسلامی، دانشگاه بوعلی سینا

کمیته اجرایی کنفرانس

- دکتر ابراهیم امینی سرشت، دانشگاه بوعلی سینا
- اعظم دهقانی، دانشگاه شهید بهشتی
- دکتر رسول سلیمانی، دانشگاه بوعلی سینا
- دکتر اصغر سیف، دانشگاه بوعلی سینا
- دکتر مجید صادقی فر، دانشگاه بوعلی سینا
- دکتر ابراهیم نصیرالاسلامی، دانشگاه بوعلی سینا
- دکتر رحیم محمودوند، دانشگاه بوعلی سینا
- دکتر سید جمال میرکمالی، دانشگاه اراک
- دکتر سید مرتضی نجیبی، دانشگاه شیراز

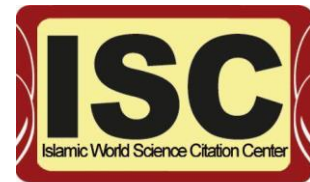
پیشگفتار

این کتابچه شامل خلاصه مقاله‌های فارسی کنفرانس روش‌های مدرن در قیمت‌گذاری‌های بیمه‌ای و آمارهای صنعتی است که پس از داوری و ویرایش و با همت و تلاش اعضای کمیته علمی، کمیته برگزاری و داوران علمی تهیه شده است. علاوه بر این جلد دیگری شامل خلاصه مقاله‌های انگلیسی به همین صورت تهیه شده است. در هر جلد مقاله‌ها بر حسب ترتیب الفبایی عناوین مرتب شده‌اند. در مجموع ۸ سخنرانی کلیدی، ۸۴ سخنرانی و ۳۳ پوستر از طرف شرکت‌کنندگان در کنفرانس ارائه شده است که بصورت جداگانه در این کتابچه آمده است. تمام تلاش برگزارکنندگان کنفرانس بر این بوده است که خلاصه مقاله‌های چاپ شده از کیفیت مطلوبی برخوردار باشند. اما با توجه به محدودیت زمان و کندی مراحل داوری اصلاحیه‌های پیشنهادی داوران برای خلاصه مقاله‌های پذیرفته شده به جز ویرایش‌های صوری امکان‌پذیر نبوده است. لذا مسئولیت محتوای مقاله‌ها بر عهده نویسندگان آنها می‌باشد.

در اینجا لازم است از حمایت‌های انجمن آمار ایران، انجمن محاسبات بیمه و مالی ایران، پژوهشکده بیمه مرکزی ج.ا.ا، موسسه عالی پژوهش تامین اجتماعی، موسسه بین‌المللی توسعه دانش فردای ایرانیان، انجمن بین‌المللی آمارهای صنعتی و بازرگانی، دانشگاه آرهوس دانمارک، گروه مهندسان مالی و بیم‌سنجی ایران، انجمن حرفه‌ای صنعت بیمه ایران، تیم تحلیل علمی داده‌ها (سدا)، بخش آمارهای صنعتی در انجمن بین‌المللی تحلیل‌های بیزی و پایگاه استنادی علوم جهان اسلام قدردانی شود.

برگزارکنندگان کنفرانس

حامیان کنفرانس



پایگاه استنادی علوم جهان اسلام



سخنرانان مدعو

- Hansjörg Albrecher, University of Lausanne, Switzerland
- Søren Asmussen, Aarhus University, Denmark
- Alireza Edalati, Actuary at VHV Insurance AG, Germany
- Amin Hassanzadeh, Shahid Beheshti University, Iran
- David Rios Insua, Rey Juan Carlos University, Spain
- Gholamreza Nakhaeizadeh, Karlsruhe Institute of Technology, Germany
- Jens Perch Nielsen, CASS Business School, United Kingdom
- Amir Safari, Insurance Research Center, Iran
- Ali Shojaie, University of Washington

فهرست مطالب

۱	چکیده سخنرانی‌ها.....
۲	استنباط بیزی روی قابلیت اعتماد مدل تنش-مقاومت برای داده‌های فیبرهای کربن.....
۶	استنباط روی پارامترهای مدل دوعاملی داده‌های پانل تک‌جامعه‌ای.....
۱۰	انتخاب متغیر و پیشگویی در داده‌های بزرگ توسط یک روش غیرخطی.....
۱۴	برآورد ذخیره با به‌کارگیری باورمندی و بوت‌استرپ در روش نردبان زنجیری.....
	برآورد ذخایر در رشته درمان گروهی و رشته شخص ثالث مطالعه موردی: بیمه‌های درمان گروهی شرکت بیمه البرز و شخص ثالث
۱۶	بیمه پاسارگاد ۱۳۹۰-۱۳۹۵.....
۲۰	برآوردگرهای استوار مدل‌های ARMA در قلمروهای زمان و فرکانس.....
۲۴	بررسی آماری تغییرات دبی در حوزه سد اکباتان همدان.....
۲۵	برآورد نقطه‌ای پارامتر تنش-مقاومت در توزیع نمایی دوپارامتری براساس نمونه‌ی مجموعه‌ی رتبه دار رکوردی.....
۲۹	بررسی تأثیر شوک‌های پولی بر شاخص‌های حسابداری، اقتصادی و بازار در شرکت‌های بورس اوراق بهادار تهران.....
۳۰	بررسی روند تغییرات سرانه دستمزد در صنایع ایران.....
۳۱	بهبود عملکرد الگوریتم‌های پردازش تصاویر سیاه و سفید بر اساس شباهت در یادگیری ماشین.....
۳۲	بررسی عوامل موثر بر کیفیت سود با تأکید بر اندازه شرکت نمونه مورد مطالعه: شرکت‌های بیمه فعال در بازار بیمه ایران.....
۳۶	تحلیل مدل‌های چندسطحی بیزی و کاربرد آن در صنعت.....
۳۷	توزیع لاپلاس چوله اصلاح شده و کاربرد آن.....
۴۱	خسارت‌های با توزیع پارتو در مدل‌های مخاطره شرکت‌های بیمه.....
۴۵	خلاصه‌سازی و تحلیل منحنی‌های رشد صنایع ایران با استفاده از رویکردی مبتنی بر تجزیه و تحلیل مؤلفه‌های اصلی.....
۴۹	دیدگاه‌ها در مورد روش‌های بیزی و داده‌های بزرگ.....
۵۲	روش جدید بازسازی داده‌های گم‌شده در رگرسیون.....
۵۶	روش‌های توصیف و تحلیل مه‌داده‌ها در آمارهای ثبتي.....
۶۰	روشی جدید در برآورد ارزش در معرض مخاطره و احتمال پیشامد زیان کرانگین بر اساس فرایند نقطه‌ای خودانگیزی.....
۶۴	سیستم پرداخت حق بیمه وسایل نقلیه در زمان سوخت‌گیری.....
۶۸	شناسایی عوامل موثر بر حوادث ناشی از کار در کارگاه‌های صنعتی با تأکید بر رگرسیون لسو.....
۷۲	غربالگری متغیرها در داده‌های بزرگ با استفاده از همبستگی فاصله‌ای.....

۷۶.....	قرارداد بیمه اتکایی چندلایه‌ای بهینه بر اساس تابع زیان اعوجاج.....
۸۰.....	کاربرد متن کاوی در تعیین قیمت خودروهای کارکرده.....
۸۱.....	کاربرد نمونه‌گیری معکوس در صنعت.....
۸۵.....	کاهش تعداد پارامترها در مدل‌سازی ذخایر ادعا.....
۸۹.....	کشش قیمتی در گروه‌های همگن ریسک؛ مطالعه موردی.....
۹۳.....	مدل‌بندی ریسک حوادث فاجعه‌آمیز در بیمه عمر: مطالعه موردی آن در ایران.....
۹۴.....	معیار نیکوئی برازش جدید در مواجهه با مه‌داده‌های درآمد.....
۹۸.....	مکانیزم متعادل خودکار در سیستم‌های بازنشستگی عمومی.....
۹۹.....	نرخ‌گذاری بر اساس مدل‌های آمیخته‌ی خطی باورمندی با ادعاهای دم‌سنگین.....
۱۰۱.....	نرخ‌گذاری کارای مبتنی بر طبقه‌بندی ریسک به منظور قیمت‌گذاری منصفانه.....
۱۰۶.....	بررسی عملکرد توسعه‌های روش تتا در پیش‌بینی سری‌های زمانی و کاربرد آن در پیش‌بینی داده‌های صنعت بیمه کشور.....
۱۰۸.....	مشخصه‌سازی خانواده توزیع‌های مکانی و مقیاسی با استفاده از آنتروپی شانون آماره‌های مرتب.....
۱۰۹.....	بکارگیری داده‌های تلماتیک در بهبود فرآیند انتخاب ریسک و قیمت‌گذاری بیمه اتومبیل.....
۱۱۳.....	معیار ریسک بر اساس آنتروپی تجمعی ماکزیمم آماره‌ی ترتیبی.....
۱۱۷.....	چکیده پوسترها
۱۱۸.....	آزمون ایستایی داده‌های فضایی- زمانی.....
۱۱۹.....	الگوهای پرداخت وابسته به ذخایر در بیمه‌های عمر و درمان.....
۱۲۰.....	اندازه نادرستی در آماره‌های ترتیبی.....
	بررسی تجمع هم‌زمان سرمایه‌گذاری مستقیم خارجی در خدمات و صنایع تولیدی با استفاده از
۱۲۱.....	رویکرد آمار فضایی و اقتصاد فضایی.....
۱۲۲.....	برآورد مدل رگرسیون وقتی ضرایب مدل تابعی از یک متغیر کمکی است.....
۱۲۳.....	بررسی مدل‌های بیزی پواسون چندمتغیره در نرخ‌گذاری بیمه.....
۱۲۴.....	بررسی اهمیت نگاه آینده‌پژوهانه به شاخص‌های ارزیابی در توسعه آموزش عالی.....
۱۲۵.....	پوشش اتکایی مازاد-زیان بهینه تحت معیارهای بقای توأم، مطالعه موردی: داده‌های آتش‌سوزی دانمارک.....
۱۲۶.....	تحلیل تکنیکی طیفی (SSA) و کاربرد آن در هواشناسی.....
۱۲۷.....	تعیین حق بیمه خالص بهینه برای سیستم‌های پاداش جریمه.....
۱۲۸.....	تعیین عوامل مؤثر بر دوره بیکاری در ایران.....
۱۲۹.....	تغییرات مشخصه‌های آماری بارش در ایستگاه همدان.....

- ۱۳۰.....حق بیمه باورمندی تحت مدل آمیخته ارلانگ برای نرخ گذاری بیمه اتکایی محصولات کشاورزی.....
- ۱۳۱.....داده کاوی و شبکه های بیزی.....
- ۱۳۲.....رویکرد قانون بیمه اجباری مصوب ۱۳۹۵ در حمایت از حقوق بیمه گذاران و زیان دیدگان ناشی از حوادث رانندگی.....
- ۱۳۳.....روش های برآورد ارزش منصفانه تعهدات بیمه ای.....
- ۱۳۴.....طرح سانسور پیش رونده ی پیوندی نوع I در مطالعه طول عمر یک سیستم صنعتی از دو جامعه نمای.....
-کاربرد داده کاوی در صنعت برق، بررسی خصوصیات پیشنهاددهندگان برتر و الگویابی آن در شرکت برق منطقه ای مازندران و گلستان.....
- ۱۳۵.....
- ۱۳۶.....کاربرد مدل های آمیخته خطی تعمیم یافته در آمار اکچوئری.....
- ۱۳۷.....گونه تعمیم یافته مدل پاسخ تصادفیده.....
- ۱۳۸.....گذری بر بازار بیمه اتومبیل در ایران.....
- ۱۳۹.....محاسبات فنی بیمه ای و تأثیر آن در جدول اضافه نرخ بیمه نامه عمر.....
- ۱۴۰.....مدل معادلات ساختاری جریمه ریح.....
- ۱۴۱.....مشخصه سازی توزیع پارتو و بتا-پارتو بر اساس مشاهدات نزدیک رکورد.....
- ۱۴۲.....مقایسه دو روش برآورد در توزیع نمایی تعمیم یافته.....
- ۱۴۳.....نیرومندی ارزش در معرض ریسک تحت توزیع های مختلف.....
- ۱۴۴.....ویژگی های خانواده مفصل مارشال-الکین.....
- ۱۴۵.....نام نامه.....

چکیده سخنرانی‌ها

استنباط بیزی روی قابلیت اعتماد مدل تنش-مقاومت برای داده

فیبرهای کربن

اکبر آبروش^۱، مسعود گنجی^۲ و بهداد مصطفائی^۳
^{۱،۲،۳} گروه آمار و کاربردها، دانشگاه محقق اردبیلی، اردبیل،

Akbar.abravesht@uma.ac.ir
mganji@uma.ac.ir
Behdad.mostafayi@gmail.com

چکیده مختصر

در این مقاله، داده‌ی فیبرهای کربن با استفاده از توزیع با تابع مخاطره توانی (DPHF) مدل بندی شده و سپس یک استنباط بیزی روی قابلیت اعتماد مدل تنش-مقاومت (R) برای آن انجام می‌گیرد. ابتدا براورد بیز R تحت زیان توان دوم خطا و به روش مونت کارلو در مدل DPHF بدست می‌آید، در ادامه، یک بازه‌ی باورمند بیزی نیز برای R ساخته می‌شود. در پایان، مقادیر براورد شده‌ی R برای داده‌ی مقاومت فیبرهای کربن بدست می‌آید.

کلمات کلیدی

بازه‌ی باورمند بیزی، براورد بیز، توزیع با تابع مخاطره توانی، داده‌ی فیبرهای کربن، روش مونت کارلو، قابلیت اعتماد مدل تنش-مقاومت.

۱- مقدمه

فرض کنید متغیرهای تصادفی X و Y به ترتیب مقاومت و تنش یک سیستم باشند. اگر در یک سیستم میزان مقاومت بیشتر از میزان تنش باشد، سیستم بدون مشکل کار خواهد کرد ولی زمانی که میزان مقاومت بیشتر از میزان تنش باشد سیستم دچار مشکل خواهد شد. به عبارت بهتر $R = P(X > Y)$ می‌تواند یک شاخص برای اندازه‌گیری قابلیت اعتماد یک سیستم باشد. این مفهوم، R در علوم کاربردی و مهندسی قابلیت اعتماد مدل تنش-مقاومت نامیده می‌شود. در برخی از مطالعات اخیر روش‌های بیزی برای استنباط روی R به کار رفته است. سینگ و همکاران (۲۰۱۲) براورد بیز R را در توزیع گاما بدست آوردند. روبیو و استیل (۲۰۱۳) یک استنباط بیزی روی R را در یک رده‌ی خاص از توزیع‌ها انجام دادند. اصغر زاده و همکاران (۲۰۱۴) یک براوردگر بیز برای R در توزیع نمایی تعمیم یافته بدست آوردند. رضایی و همکاران (۲۰۱۵) نیز یک استنباط بیزی روی R در توزیع پارتوی تعمیم یافته ارائه کردند. هم‌چنین، کیناچی (۲۰۱۴) یک استنباط کلاسیک روی R در DPHF انجام داده است. ما در این مقاله یک استنباط بیزی روی R در DPHF انجام می‌دهیم.

۲- توزیع با تابع مخاطره توانی

متغیر تصادفی مثبت Z دارای توزیع با تابع مخاطره توانی با پارامترهای (α, θ) است هرگاه تابع چگالی، $f(z)$ و تابع توزیع، $F(z)$ ، آن بصورت زیر باشد

$$f(z) = \alpha z^{\theta-1} e^{-\frac{\alpha}{\theta} z^{\theta}}, \theta > 0, \alpha > 0, \quad (1)$$

$$F(z) = 1 - e^{-\frac{\alpha}{\theta} z^\theta}, \theta > 0, \alpha > 0. \quad (2)$$

برخی از توزیع‌های معروف آماری مانند ریلی، وایبل و نمایی حالت‌های خاصی از این توزیع هستند. (موگدادی، ۲۰۰۵) با یک انتگرال‌گیری ساده می‌توان نشان داد که اگر X و Y دو متغیر تصادفی مستقل و دارای توزیع با تابع مخاطره توانی به ترتیب با پارامترهای (α, θ) و (α, θ) باشند، در این صورت داریم

$$R = P(X > Y) = \frac{\alpha_2}{\alpha_1 + \alpha_2}. \quad (3)$$

۳- مدل‌بندی داده‌های فیبرهای کربن

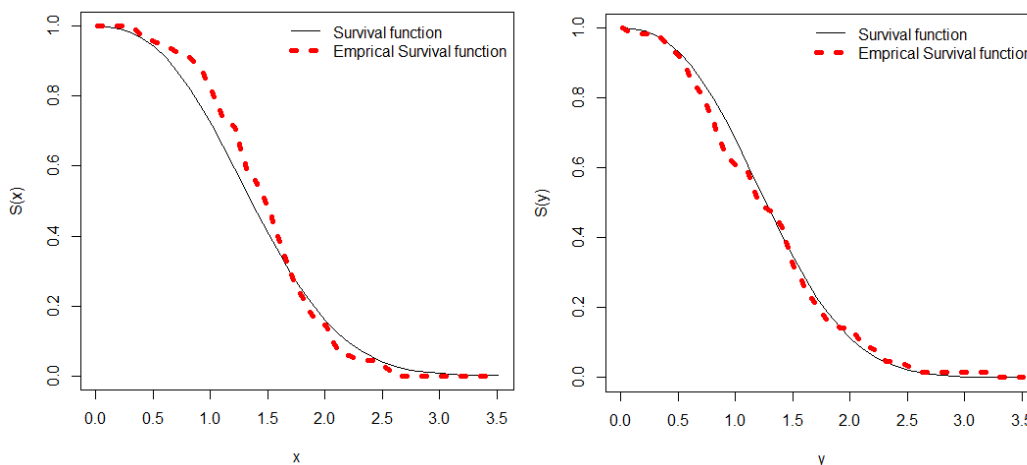
داده‌های فیبرهای کربن اولین بار توسط بادار و پریست (۱۹۸۲) گزارش شد. این داده شامل میزان مقاومت برای یک رشته از فیبرهای کربنی است که هر یک متشکل از ۱۰۰۰ الیاف فیبر کربن هستند. این رشته‌های فیبر تحت کشش‌های با طول ۱، ۱۰، ۲۰ و ۵۰ میلی‌متر مورد آزمون قرار گرفت. فرض کنید متغیرهای تصادفی X و Y به ترتیب میزان مقاومت فیبرهای کربن تحت کشش ۱۰ میلی‌متر و ۲۰ میلی‌متر باشند. رقب و همکاران (۲۰۰۸) داده‌ی مربوط به X و Y را با استفاده از توزیع نمایی تعمیم یافته سه پارامتری مدل‌بندی کردند. در این تحقیق نیز این داده توسط تابع توزیع با تابع مخاطره توانی مدل‌بندی می‌شود.

با استفاده از آزمون کولموگروف-اسمیرنوف فرض این‌که دو متغیر X و Y دارای توزیع‌های با تابع مخاطره توانی به ترتیب با پارامترهای (α, θ) و (α, θ) هستند، تایید می‌شود. پارامترهای توزیع، آماره آزمون و p -مقدار (معنی داری) برای این دو متغیر با نرم افزار آماری R محاسبه شده و در جدول ۱ گزارش شده‌اند. توجه کنید که پارامتر θ در اینجا برای هر دو متغیر مقداری معلوم و برابر ۲٫۵ است.

جدول ۱: نتایج خروجی آزمون کولموگروف-اسمیرنوف برای مدل‌بندی داده با DPHF

متغیر تصادفی	پارامترهای توزیع (α, θ)	آماره آزمون	p -مقدار
X	(۰٫۸۱، ۲٫۵)	۰٫۱۲۷۷۴	۰٫۲۱۳۴
Y	(۰٫۹۶، ۲٫۵)	۰٫۱۲۵۱۸	۰٫۲۸۰۴

با توجه به این‌که هر دو p -مقدار بزرگتر از ۰٫۰۵ هستند لذا در سطح اطمینان ۹۵ درصد می‌پذیریم که هر دو متغیر دارای توزیع با تابع مخاطره توانی هستند. برای این‌که مناسب بودن مدل آماری واضح‌تر دیده شود، نمودارهای کاپلان-میر برای داده‌های فیبرهای کربن در شکل ۱ رسم شده است.



شکل ۱: نمودار کاپلان-میر برای داده‌های فیبرهای کربن

۴- برآوردگر بیز R

فرض کنید X_1, \dots, X_n و Y_1, \dots, Y_m دو نمونه تصادفی مستقل از توزیع با تابع مخاطره توانی به ترتیب با پارامترهای (α, θ) و (ν, θ) باشند. با در نظر گرفتن تابع زیان توان دوم خطا، برآوردگر بیز R برابر امید پسین R خواهد بود. امید پسین R بصورت زیر تعریف می‌شود

$$E[R | x, y] = \int_0^1 r \cdot \pi(r | x, y) dr, \quad (4)$$

که در آن $\pi(r | x, y)$ توزیع پسین R است.

در مسأله حاضر این امید پسین دارای فرم بسته‌ای نیست و بنابراین مقدار آن به روش مونت کارلو بدست می‌آید. با توجه به این روش

هرگاه $\Gamma_1, \dots, \Gamma_k$ یک نمونه تصادفی از $\pi(r | x, y)$ باشد، آنگاه $\sum_{i=1}^k r_i / k$ یک تقریب برای $E[R | x, y]$ خواهد بود. لذا کافی است

$\pi(r | x, y)$ را بدست آورده و با استفاده از الگوریتم متروپلیس-هستینگز از آن نمونه تولید کرد.

در این مقاله، بدلیل نداشتن اطلاعات زیاد پیرامون پارامترها و همچنین برخی ویژگی‌های خوب تابع توزیع پیشین جفریز (جفریز، ۱۹۴۶)، از این پیشین استفاده می‌شود. بعد از اعمال این پیشین توزیع پسین $\alpha_1 | x_1, \dots, x_n$ و $\alpha_2 | y_1, \dots, y_m$ به ترتیب برابر گاما با پارامترهای $(n, \sum_{i=1}^n x_i^\theta / \theta)$ و $(m, \sum_{j=1}^m y_j^\theta / \theta)$ خواهد بود. حال با استفاده از روش تبدیل، توزیع پسین R، $\pi(r | x, y)$ ، بصورت زیر بدست می‌آید،

$$\pi(r | x, y) = \frac{v^n w^m \Gamma(n+m) r^{n-1} (1-r)^{m-1}}{\Gamma(n)\Gamma(m) (rv + (1-r)w)^{n+m}}, \quad 0 \leq r \leq 1, \quad (5)$$

که در آن

$$v = \sum_{j=1}^m y_j^\theta / \theta, \quad w = \sum_{i=1}^n x_i^\theta / \theta \quad (6)$$

و $\Gamma(\cdot)$ تابع گاما است. حال به راحتی می‌توان روش مونت کارلو را با استفاده از الگوریتم متروپلیس-هستینگز برای این مسأله بکار برد.

۵- بازه باورمند بیزی R

بازه‌های باورمند بیزی در واقع نسخه‌ی بیزی بازه‌های اطمینان در روش‌های آماری کلاسیک هستند. به عبارت دیگر، بازه‌ی C را یک

بازه‌ی باورمند $100(1-\alpha)\%$ برای R گویند هرگاه $P(R \in C) = 1-\alpha$. برای مسأله اخیر، یک بازه‌ی باورمند

بیزی برای R بصورت زیر بدست می‌آید.

$$\left(1 / \left[1 + \frac{nv}{mw} F_{1-\alpha/2}(2n, 2m) \right], 1 / \left[1 + \frac{nw}{mv} F_{\alpha/2}(2n, 2m) \right] \right), \quad (7)$$

که در آن $F_{\alpha/2}(2n, 2m)$ برابر چندک $\alpha/2$ -م توزیع F فیشر است.

۶- نتایج

در این بخش، از نرم افزار آماری R استفاده شده و برآوردگر بیز و بازه‌ی باورمند ۹۵٪ بیزی برای R با استفاده از مطالب بیان شده در

بخش‌های قبل بدست می‌آید. نتایج بطور خلاصه در جدول ۲ گزارش شده‌اند.

جدول ۲: برآوردگر بیز و بازه‌ی باورمند ۹۵٪ بیزی برای R

بازه‌ی باورمند ۹۵٪ بیزی	برآوردگر بیزی
(0.4496, 0.6186)	۰.۵۵۶۳

مراجع

- [1] Asgharzadeh, A., Valiollahi, R., and Raqab, M. Z., (2014), Estimation of $\Pr(Y < X)$ for the Two parameter of Generalized Exponential Records, Communications in Statistics - Simulation and Computation, DOI: 10.1080/03610918.2014.964046
- [2] Badar, M. G., Priest, A. M. (1982). Statistical aspects of fiber and bundle strength in hybrid composites. In: Hayashi, T., Kawata, K., Umekawa, S., eds. Progress in Science and Engineering Composites, Tokyo: ICCM-IV, 1129-1136.
- [3] Jeffreys, H. , (1946), An invariant form for the prior probability in estimation problems, Proceedings of the Royal Society of London, Series A, Mathematical and Physical Sciences, 186, 453-461.
- [4] Kinaci, I., (2014), Estimation of $P(Y < X)$ for distribution having power hazard function, Pak. J.Statist, 30, 57-70.
- [5] Mugdadi, A. R.(2005), The least squares type estimation of the parameters in the power hazard function, Appl. Math. Comput., 169, 737-748.
- [6] Raqab, M. Z., Madi, M. T., Kundu, D., (2008), Estimation of $P(Y < X)$ for the three-parameter generalized exponential distribution, Communications in Statistics - Theory and Methods, 37:18, 2854-2864.
- [7] Rezaei, S., Alizadeh Noughabi, R., Nadarajah, S., (2015), Estimation of stress-strength reliability for the generalized pareto distribution based on progressively censored samples. Ann. Data. Sci., 2(1), 83-101.

استنباط روی پارامترهای مدل دوعاملی داده‌های پانل تک‌جامعه‌ای

احمد ملک‌زاده^۱، متین دالوندپور^۲

^۱گروه آمار دانشگاه صنعتی خواجه نصیرالدین طوسی، تهران
malekzadeh@kntu.ac.ir (admalekzadeh@yahoo.com)
matindalvandpur@gmail.com

چکیده مختصر

داده‌های پانل با مدل کردن داده‌های مقطعی در طول زمان، یکی از پرکاربردترین مدل‌های آماری در بررسی‌های اقتصادی است. به‌طور عمده استنباط درباره پارامترهای اینگونه مدل‌ها به‌دلیل پیچیدگی به‌صورت دقیق انجام نمی‌پذیرد و غالباً روش‌های عددی بدین منظور ارائه می‌گردد. در اینجا به استنباط درباره پارامترهای داده‌های پانل دوعاملی در یک جامعه با استفاده از روش بوت‌استرپ پارامتری می‌پردازیم و آن را با روش تعمیم‌یافته ارائه شده توس ژائو (۲۰۰۷) مقایسه می‌نماییم. عملکرد آن را نیز توسط یک مثال واقعی بیان خواهیم کرد.

کلمات کلیدی

داده‌های پانل، بوت‌استرپ پارامتری، روش تعمیم‌یافته، شبیه‌سازی.

۱- مقدمه

داده‌های پانل، ترکیبی از داده‌های مقطعی و سری زمانی هستند، یعنی اطلاعات مربوط به داده‌های مقطعی در طول زمان مشاهده می‌شوند، به همین دلیل آن‌ها را داده‌های ترکیبی یا داده‌های مقطعی-سری زمانی نیز می‌گویند. داده‌های پانل در علوم مختلف از قبیل علوم اقتصادی، علوم اجتماعی، علوم سیاسی، کشاورزی، روان‌شناسی و پزشکی مورد استفاده قرار می‌گیرند. دو مثال معروف از داده‌های پانل در ایالات متحده عبارتند از: بررسی پانلی پویایی‌های درآمد، که اطلاعات آن توسط مؤسسه تحقیقات اجتماعی در دانشگاه میشیگان (PSID) جمع‌آوری شده است و نظرسنجی طولی ملی (NLS)، که مجموعه‌ای از نظرسنجی‌هایی است که توسط اداره آمار کار جمع‌آوری شده است. PSID در سال ۱۹۶۸ با ۴۸۰۰ خانوار آغاز و تا سال ۲۰۰۱ به بیش از ۷۰۰۰ خانوار رسید. در سال ۲۰۰۳، اطلاعات بیش از ۶۵۰۰۰ نفر را جمع‌آوری کرده بود که جمعاً ۳۶ سال از زندگی هر فرد را پوشش می‌داد. فهرستی از متغیرهای PSID عبارتند از درآمد، وضعیت فقر، کمک‌های عمومی به‌صورت غذا یا مسکن، زمان کار خانگی، تحرک جغرافیایی، پس‌زمینه وضعیت اقتصادی اجتماعی و غیره. از سوی دیگر NLS، مجموعه‌ای از تحقیقات انجام شده برای جمع‌آوری اطلاعات در نقاط متعدد در زمان فعالیت در بازار کار و سایر رویدادهای مهم زندگی از چندین گروه از مردان و زنان است. فهرستی از متغیرهای این تحقیق عبارتند از میزان تحصیل و نقل و انتقالات کاری، ازدواج و باروری، سرمایه‌گذاری‌های آموزش، مراقبت از کودک، استفاده از مواد مخدر و غیره.

در تجزیه و تحلیل داده‌های پانل، با توجه به در نظر گرفتن اثر شخص، اثر زمان و یا هر دوی این اثرات، مدل‌های تک‌عاملی و دو‌عاملی تعریف می‌شوند. اگر ماتریس واریانس در این مدل‌ها معلوم باشد، می‌توان استنباط‌های دقیقی انجام داد اما در صورت مجهول بودن ماتریس واریانس، انجام استنباط دقیق کاری پیچیده و دشوار است. در این مواقع می‌توان از روش تعمیم‌یافته، که توسط تسوی و ویراهاندی [2] و همچنین ویراهاندی [4] معرفی شده است و یا روش بوت‌استرپ پارامتری استفاده کرد. ژائو [5] به بررسی مدل پانلی با استفاده از مفهوم استنباط تعمیم یافته پرداخت و ژو و تیان [3] به انجام استنباط در تک جامعه تک‌عاملی با استفاده از روش بوت استرپ پارامتری پرداخت. در این مقاله سعی داریم روش انجام استنباط بر روی پارامترهای اصلی مدل پانلی را با استفاده از روش بوت

۳- ارائه روش جدید استنباط

باتوجه به مطالب فوق داریم $\hat{\beta} \sim N_k(\beta, \Sigma_{\hat{\beta}})$ که در آن $\Sigma_{\hat{\beta}} = \left(\sum_{j=1}^3 \frac{X'Q_jX}{\pi_j^2} \right)^{-1}$ تعریف شده است. بنابراین $c'\hat{\beta} \sim N(c'\beta, c'\Sigma_{\hat{\beta}}c)$ و متعاقب آن $Q^* = (c'\hat{\beta} - c'\beta) / c'\Sigma_{\hat{\beta}}c \sim N(0,1)$ (یعنی π_1^2, π_2^2 و π_3^2) در عمل $\hat{\beta}$ و $\Sigma_{\hat{\beta}}$ مجهول است و امکان استفاده از Q^* بمنظور انجام استنباط بر روی ترکیب خطی $c'\beta$ وجود ندارد، بنابراین پیشنهاد می‌کنیم از آماره زیر که حاصل از جایگزینی π_i^2 با برآوردگر نارایب آنها، یعنی $\hat{\pi}_i^2$ برای $i = 1, 2, 3$ استفاده شود:

$$Q = (c'\hat{\beta} - c'\beta) / c'\hat{\Sigma}_{\hat{\beta}}c \quad (6).$$

جایی که $\hat{\Sigma}_{\hat{\beta}} = \left(\sum_{j=1}^3 \frac{X'Q_jX}{\hat{\pi}_j^2} \right)^{-1}$ و $\hat{\beta} = \left(\sum_{j=1}^3 \frac{X'Q_jY}{\hat{\pi}_j^2} \right)$ توزیع آماره Q بسیار پیچیده است و می‌توان آنرا با استفاده از روش بوت‌استرپ پارامتری تقریب زد. بدین منظور فرض کنید $\hat{\beta}_0$ و $\hat{\Sigma}_{\hat{\beta}_0}$ به ترتیب مقادیر مشاهده شده $\hat{\beta}$ و $\hat{\Sigma}_{\hat{\beta}}$ هستند. Y_B را از توزیع $N(Z(\hat{\alpha}, \hat{\beta}_0)')$ تولید کرده و با استفاده از تعاریف (۴) و (۵) مقادیر π_1^2, π_2^2 و π_3^2 را مجدد برآورد می‌نماییم. و متعاقباً مقادیر $\hat{\beta}$ و $\Sigma_{\hat{\beta}}$ یعنی $\hat{\beta}_B$ و $\Sigma_{\hat{\beta}_B}$ را نیز بدست می‌آوریم و براساس این مقادیر فرم بوت‌استرپ کمیت محوری Q برابر می‌شود با

$$Q^B = (c'\hat{\beta}_B - c'\hat{\beta}_0) / c'\Sigma_{\hat{\beta}_B}c \quad (7).$$

حال اگر Q^B را بعنوان چندک γ ام توزیع Q^B تعریف کنیم، آنگاه فاصله اطمینان $(1 - \alpha)$ درصدی برای $c'\beta$ برابر خواهد بود با

$$\left[c'\hat{\beta}_0 - Q_{1-\alpha/2}^B \sqrt{c'\hat{\Sigma}_{\hat{\beta}_0}c}, c'\hat{\beta}_0 - Q_{\alpha/2}^B \sqrt{c'\hat{\Sigma}_{\hat{\beta}_0}c} \right] \quad (8).$$

نکته: با تعریف $Q = (\hat{\beta} - \beta)' \hat{\Sigma}_{\hat{\beta}}^{-1} (\hat{\beta} - \beta)$ و همچنین $Q^B = (\hat{\beta}_B - \hat{\beta}_0)' (\Sigma_{\hat{\beta}_B}^B)^{-1} (\hat{\beta}_B - \hat{\beta}_0)$ می‌توان فاصله اطمینان و آزمون فرضهای مربوط به بردار پارامتری β را انجام داد.

۴- شبیه‌سازی

بالتاجی و گریفین [1]، معادله‌ی تقاضای بنزین را به صورت زیر در نظر گرفتند

$$\ln\left(\frac{Gas}{Car}\right) = \alpha + \beta_1 \ln\left(\frac{Y}{N}\right) + \beta_2 \ln\left(\frac{PMG}{P_{GMD}}\right) + \beta_3 \ln\left(\frac{Car}{N}\right) \quad (9),$$

که در آن $\frac{Gas}{Car}$ ، مصرف بنزین توسط خودرو، $\frac{PMG}{P_{GMD}}$ ، قیمت واقعی بنزین و $\frac{Car}{N}$ ، موجودی سهام سرانه‌ی اتومبیل را نشان می‌دهند. این داده‌های پانل، مشاهدات سالیانه‌ی هیجده کشور عضو سازمان همکاری و توسعه‌ی اقتصادی را در سال‌های ۱۹۶۰-۱۹۷۸ در بر می‌گیرند. داده‌ها در مقالاتی همانند مقاله‌ی ژائو [5]، قابل مشاهده است. از متغیرهای توصیفی این مطالعه بمنظور انجام شبیه‌سازی استفاده می‌نماییم. بمنظور مقایسه عملکرد روش جدید با روش تعمیم یافته ارائه شده توسط ژائو [5] برای مقادیر مختلف σ_{λ}^2 ، σ_{μ}^2 و σ_V^2 به بررسی احتمال پوشش (CP) و طول فاصله اطمینان (L) ایجاد شده، می‌پردازیم. در جدول ۱ این مقادیر گزارش شده است. نتایج براساس ۵۰۰۰ بار تولید داده می‌باشند. براساس نتایج شبیه‌سازی می‌توان گفت، روش جدید ارائه شده عملکرد بهتری دارد.

جدول ۱: برآورد احتمال پوشش (CP) و متوسط طول (L) دو روش عددی.

روشهای مورد استفاده		GCI		PB	
$\sigma_v^2, \sigma_\mu^2, \sigma_\lambda^2$	c_1, c_2, c_3	CP	L	CP	L
1, 1, 1	1, 1, 1	۰,۹۵۶	۱۱,۲۳	۰,۹۴۹	۱۰,۷۲
	-2, 1, 5	۰,۹۶۲	۱۵,۸۷	۰,۹۵۴	۱۴,۶۵
	1, 5, -8	۰,۹۶۴	۱۹,۴۵	۰,۹۵۱	۱۸,۷۷
2, 5, 8	1, 1, 1	۰,۹۵۳	۲۱,۴۰	۰,۹۴۲	۲۰,۹۰
	-2, 1, 5	۰,۹۵۹	۲۹,۷۱	۰,۹۵۵	۲۸,۸۹
	1, 5, -8	۰,۹۶۶	۳۷,۹۷	۰,۹۴۸	۳۶,۶۳
4, 1, 5	1, 1, 1	۰,۹۵۸	۱۸,۳۳	۰,۹۵۳	۱۷,۲۴
	-2, 1, 5	۰,۹۶۴	۲۵,۷۲	۰,۹۵۱	۲۴,۶۷
	1, 5, -8	۰,۹۶۰	۳۳,۰۸	۰,۹۴۸	۳۲,۱۸

۵- مثال واقعی

برای داده‌های مثال بالتاجی و گریفین یعنی مدل (۹)، مقادیر برآورده شده مربوط به $\pi_1^2, \pi_2^2, \pi_3^2$ و π_4^2 بترتیب برابر است با $\hat{\beta}_0 = (0.048, -0.229, -0.373)'$ همچنین برآورد بردار پارامتری $\beta_0 = (0.048, -0.229, -0.373)'$ نتیجه شده است. حال برای مقادیر مختلف ضرایب (جدول ۲) فاصله اطمینان‌ها را براساس روشهای ارائه شده، بدست می‌آوریم.

جدول ۲: برآورد احتمال پوشش (CP) و متوسط طول (L) دو روش عددی.

c_1, c_2, c_3	PBI		GCI	
	فاصله اطمینان	طول	فاصله اطمینان	طول
(۱, ۰, ۰)	(۰, ۰.۳۹۰۰, ۰.۵۶)	۰, ۰.۱۷	(۰, ۰.۳۷۰۰, ۰.۵۸)	۰, ۰.۲۱
(۱, -۱, ۰)	(۰, ۰.۲۳۴۰, ۰.۳۱۱)	۰, ۰.۷۷	(۰, ۰.۲۴۴۰, ۰.۳۴۷)	۰, ۱.۰۳
(۱, ۰, -۱)	(۰, ۰.۳۶۱۰, ۰.۴۸۷)	۰, ۱.۲۶	(۰, ۰.۳۴۴۰, ۰.۴۷۹)	۰, ۱.۳۵

مراجع

- [1] Baltagi, B. H., Griffin, J. M. (1983). Gasoline demand in the OECD: an application of pooling and testing procedures. *Eur. Econ. Rev.* 22:117-137.
- [2] Tsui, K., Weerahandi, S., (1989), Generalized p-Values in Significance Testing of Hypotheses in the Presence of Nuisance Parameters, *Journal of the American Statistical Association*, 84, 406, 602-607.
- [3] Xu, L., Tian, M., (2016), Parametric Bootstrap Inferences for Panel Data Models, *Communications in Statistics - Theory and Methods*, 46, 11, 18-36.
- [4] Weerahandi, S. (1993). Generalized confidence intervals. *Journal of the American Statistical Association*, 88(423):899-905.
- [5] Zhao, H.B., (2007), Exact tests in panel data using generalized p-Values, *Communication in Statistics - Theory and Methods*, 37, 1, 18-36.
- [6] Swamy, P. A. V. B., Arora, S. S. (1972). The exact finite sample properties of the estimators of coefficients in the error components regression models. *Econometrica* 40:261-275.

انتخاب متغیر و پیشگویی در داده‌های بزرگ توسط یک روش غیرخطی

محبوبه محبی^۱، داود شاهسونی^۲

^۱ دانشجوی کارشناسی ارشد، دانشگاه صنعتی شاهرود،
M.Mohebbi@shahroodut.ac.ir

^۲ هیات علمی گروه آمار، دانشگاه صنعتی شاهرود
dshahsavani@shahroodut.ac.ir

چکیده مختصر

تحلیل داده‌های با بعد بالا در همه‌ی زمینه‌های مختلف علوم، صنعتی و تجاری، همواره با چالش "انتخاب متغیرهای مهم" روبه‌رو است. روش‌های متداول انتخاب متغیر مبتنی بر ساختارهای خطی هستند، اما در بسیاری از موارد و در بعد بالا ممکن است ارتباط ساختاری متغیر پاسخ و متغیرهای توضیحی خطی نباشد. همچنین این روش‌ها، متغیر-هایی را که به تنهایی تأثیر ناچیزی بر متغیر پاسخ دارند، اما در ترکیب با سایر متغیرها که برای پیشگویی ممکن است مفید باشند را نادیده می‌گیرند. در این پژوهش با ارائه الگوریتمی چند مرحله‌ای مبتنی بر رگرسیون موضعی خطی، متغیرهای مهم در قالب اثرات متقابل تأثیرگذار، شناسایی و انتخاب شده و مدل پیشگویی برازش می‌شود. بررسی‌های انجام شده با داده‌های واقعی در صنعت دارویی نشان از کارآمدی الگوریتم در فضای متغیرهای توضیحی با بعد بالا است.

کلمات کلیدی

انتخاب متغیر، رگرسیون ناپارامتری، رگرسیون موضعی، اعتبار سنجی تقاطعی

۱- مقدمه

گسترش روز افزون فناوری اطلاعات، امکان ایجاد مجموعه داده‌هایی با بعد بالا را در تحقیقات حوزه‌های مختلف صنعتی و تجاری و نیز شاخه‌های مختلف علوم همچون زیست‌شناسی، نجوم، ژنتیک و پزشکی پدید آورده است. ساختار گونه‌ای از این داده‌ها به این صورت است که بعد مجموعه متغیرهای توضیحی بسیار بزرگتر از اندازه نمونه بوده و بدین ترتیب چالشی را برای کشف روابط پنهان در داده‌ها و بکارگیری روش‌های متداول آماری و یادگیری ماشین ایجاد می‌کند. از این رو، محققین با مسأله مهمی تحت عنوان "انتخاب متغیر" در هر دو زمینه یادگیری با راهنما (رگرسیون و رده بندی) و یادگیری بدون راهنما (خوشه‌بندی و آنالیز پیوند) مواجه هستند. در مسأله رگرسیون، با هدف پیشگویی متغیر پاسخ (y) برحسب متغیرهای توضیحی، از حدود ۱۵ سال پیش تاکنون، توجه ویژه‌ای به مسئله انتخاب متغیر در چارچوب مدل‌های خطی شده است. این تحقیقات با ارائه روش لاسو در سال ۱۹۹۶ آغاز گشت. روش‌هایی چون Least Angel Regression [۱]، The Group Lasso [۲]، The Relaxed Lasso [۳]، Dantzig Selector [۴]؛ از جمله روش‌های انتخاب متغیر در بعد بالا و مبتنی بر فرض خطی بودن رابطه y (متغیر پاسخ) با X ها (متغیرهای پیشگو) هستند. علی‌رغم کاربرد موفق این روش‌ها در بعد بالا برای بسیاری از مجموعه داده‌ها، مشکل آن است که ارزیابی وجود رابطه خطی در بعد بالا کاری بسیار دشوار است و ممکن است در موارد بسیار زیادی این فرض معنی‌دار نبوده و رابطه غیرخطی حاکم باشد؛ لذا استفاده از روش‌های خطی، موثر نخواهند بود. علاوه بر این، روش‌های متداول انتخاب متغیر منجر به حذف متغیرهایی می‌شوند که به تنهایی ارزش کمی دارند، در حالی که ممکن است در ترکیب با دیگر متغیرها نتایج سودمندی داشته باشند. در این مقاله درصدد

معرفی و کاربرد روشی هستیم که مبتنی بر فرضیه غیر خطی بودن رابطه متغیر پاسخ با متغیرهای پیشگو است و اثرات متقابل مهم را نیز شناسایی می‌کند.

۲- روش شناسی

فرض کنید بردار P متغیره $X_i = (X_{i1}, \dots, X_{ip})$ مشاهده i ام و Y_i مقدار متغیر پاسخ متناظر باشد به طوری که $i \in S = \{1, 2, \dots, n\}$ و $p > n$. همچنین فرض کنید \hat{Y} برآورد ناپارامتری متغیر پاسخ مبتنی بر ساختار رگرسیون موضعی خطی و $E = \sum_{i=1}^n (y_i - \hat{y}^{-i})^2 w_i$ خطای اعتبار سنجی است که در آن \hat{y}^{-i} برآوردگر leave-one-out $[\delta]$ بدست آمده از مجموعه $S - \{i\}$ بر اساس وزن‌های w_i مبتنی بر توابع هسته می‌باشد. در اینجا هدف آن است که بر اساس رگرسیون موضعی خطی، ضمن انجام یک غربالگری، مجموعه کوچکی از متغیرهای توضیحی موثر در قالب اثرات متقابل چند عاملی، شناسایی شده و پیشگویی متغیر پاسخ بوسیله متغیرهای انتخاب شده انجام گیرد. این روش در چند گام، عمل انتخاب متغیر را همزمان با پیشگویی متغیر پاسخ انجام می‌دهد.

۲-۱- الگوریتم

گام اول: رگرسیون موضعی متغیر پاسخ بر روی هر یک از متغیرهای توضیحی بصورت جداگانه انجام شده و بر اساس مینیمم خطای حاصل از پیشگویی (E) که وابسته به ارزیابی تقاطعی است، متغیرها بصورت صعودی مرتب می‌شوند. در اینجا \hat{Y} تابعی از یک متغیر $(\hat{y}(X_j), j = 1, \dots, p)$ است. فرض کنید موثرترین متغیر گام اول که دارای مینیمم خطا است، با نماد $I(1) = \{X_{(1)}^1\}$ نشان داده شود.

گام دوم: در لیست مرتب شده متغیرهای گام اول، تعداد $p_1 = \sqrt{p}$ متغیر از ابتدای لیست که دارای کمترین مقدار خطا هستند، برای ورود به گام دوم انتخاب می‌شوند. سپس کلیه اثرات متقابل دو عاملی این p_1 متغیر، که تعداد آنها برابر با $p_2^* = \binom{p_1}{2}$ است، در نظر گرفته شده و رگرسیون موضعی Y بر روی هر یک از این زوج متغیرها برازش می‌شود. سرانجام از بین p_2^* زوج که بر اساس میزان خطای متناظرشان مرتب شده‌اند، تعداد $p_2 = \sqrt{p}$ زوج از ابتدای لیست، برای ورود به گام بعدی تعیین می‌شوند. فرض کنید اولین و موثرترین زوج متغیر که دارای مینیمم خطا است، با نماد $I(2) = \{X_{(1)}^2, X_{(2)}^2\}$ نشان داده شود که لزوماً شامل $\{X_{(1)}^1\}$ نیست. شایان ذکر است که به ازای هر زوج مذکور، تابع ناپارامتری \hat{Y} ، تابعی از آن دو متغیر توضیحی است.

گام سوم: به ازای هر دو زوج متغیر انتخاب شده گام قبل، زیر مجموعه‌های جدیدی (شامل سه یا چهار متغیر) از اجتماع زوج متغیرها تشکیل شده و رگرسیون موضعی برای هر یک از این زیر مجموعه‌ها برازش می‌شود. برای مثال اگر $\{1, 3\}$ ، $\{6, 8\}$ ، $\{4, 6\}$ سه تا از بهترین زوج متغیرهای گام ۲ باشند، آنگاه $\{1, 3, 4, 6\}$ ، $\{1, 3, 6, 8\}$ ، $\{4, 6, 8\}$ زیر مجموعه‌های کاندید سه و چهارتایی برای گام سوم هستند. سپس در لیست مرتب شده این زیر مجموعه‌ها، مجموعه متغیر $I(3)$ مشخص می‌شود. پس از مشخص شدن $I(4), I(5), \dots$ در گام‌های بعدی، این روال تا زمانی ادامه می‌یابد که

$$\frac{E_{I(l)} - E_{I(l+1)}}{E_{I(l)}} \leq t$$

سرانجام با پایان یافتن آن در گام k ام، موثرترین متغیرها شامل متغیرهای مجموعه

$$I(k) = \{X_{(1)}^k, X_{(2)}^k, \dots, X_{(l)}^k\}, k \leq l \leq 2^{k-1}$$

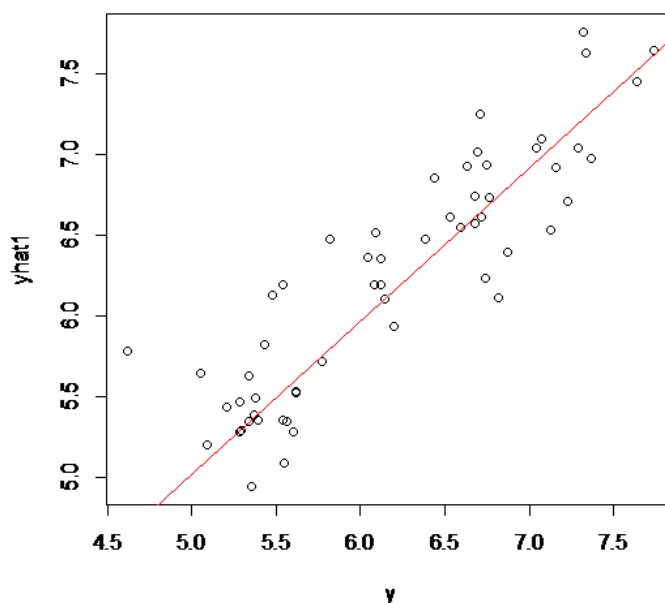
انتخاب شده و مدل نهایی و پیشگویی متغیر پاسخ توسط برازش رگرسیون موضعی خطی بر این زیر مجموعه انجام می‌گیرد. لازم به ذکر است که الگوریتم مذکور به ازای مقادیر بزرگ t بطور سریع پایان می‌یابد و مقادیر کوچک t باعث افزایش تعداد گام‌ها و در نتیجه افزایش تعداد متغیرهای انتخاب شده نهایی می‌شوند.

۳- بررسی کارایی الگوریتم بر روی داده‌های واقعی

در این پژوهش از یک مجموعه داده‌ی واقعی در صنایع دارویی استفاده شده و هدف آن است که از بین انبوهی از ویژگی‌های ساختاری یک دارو، تنها چند ویژگی موثر بر میزان غلظت دارو در بدن بیمار انتخاب شده و یک مدل پیشگویی برای میزان غلظت تعیین شود. داده‌ها شامل $p=1595$ متغیر و مبین ویژگی‌هایی همچون تعداد عناصر تشکیل دهنده ساختار دارو، زاویه پیوند، طول پیوند، بعد مکانی در فضا و... برای $n=57$ بیمار هستند. متغیر پاسخ مربوط به خاصیت بیولوژیکی ساختارهای مذکور (غلظت دارو) در بدن بیمار است. با استفاده از بسته NOVAS در نرم افزار آماری R و با در نظر گرفتن $t=0.1$ ، تعداد ۸ متغیر $X_{295}, X_{337}, X_{461}, X_{493}, X_{774}, X_{882}, X_{1511}, X_{1553}$ تاثیرگذار بر روی غلظت دارو در بدن بیمار از بین ۱۵۹۵ متغیر طی ۵ گام اجرای الگوریتم شناسایی شدند که جزئیات نتایج در جدول ۱ مشخص شده است. همان‌گونه که پیداست، ضرورتی ندارد که بهترین متغیر(های) انتخاب شده در هر گام در مراحل بعدی حضور داشته باشد. سرانجام پیشگویی متغیر پاسخ توسط رگرسیون خطی موضعی بر روی متغیرهای انتخاب شده در گام پنجم انجام شده است (شکل ۱).

جدول ۱- متغیرهای انتخاب شده در هر گام از اجزای الگوریتم

مراحل	اندیس متغیرهای انتخاب شده دارای کمترین مقدار خطا	(E) اندازه‌ی خطا
گام ۱	۱۵۵۳	۰/۲۵۸۲۷۲۹
گام ۲	۲۹۷, ۱۵۵۳	۰/۱۷۳۱۵۶۲
گام ۳	۲۹۷, ۳۳۷, ۶۵۲, ۱۵۵۳	۰/۱۶۷۶۶۰۷۷
گام ۴	۳۰۵, ۳۳۷, ۶۵۲, ۸۸۲, ۹۳۹, ۱۵۵۳	۰/۱۳۴۴۶۳۵
گام ۵	۲۹۵, ۳۳۷, ۴۶۱, ۴۹۳, ۷۷۴, ۸۸۲, ۱۵۱۱, ۱۵۵۳	۰/۱۲۷۶۲۹۵



شکل ۱. نمودار غلظت دارویی پیشگویی شده در برابر غلظت‌های واقعی ($R^2=0/86$)

۴- بحث و نتیجه گیری

انتخاب متغیر و انتخاب ساختار خطی یا غیرخطی بودن ارتباط متغیر پاسخ با متغیرهای توضیحی، دو چالش مهم در داده‌های با بعد بالا است. در این مقاله، استفاده موفق از روش رگرسیون خطی موضعی به عنوان یک روش ناپارامتری منجر به انتخاب تعداد ۸ متغیر توضیحی از بین انبوه متغیرها گردید و عملکرد پیشگویی غیرخطی آن (شکل ۱) با ضریب همبستگی ۰/۸۶ بسیار قابل قبول است. کاهش هزینه محاسباتی موضوع مهم دیگری است که در بطن این روش و اجرای آن در نرم افزار لحاظ شده است.

مراجع

- [1] Efron, B., & Hastie, T. (2004). Least angle regression. *The Annals of statistics*, 407-499.
- [2] Yuan, M., & Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 49-67.
- [3] Meinshausen, N. (2007). Relaxed lasso. *Computational Statistics & Data Analysis*, 374-393.
- [4] Candes, E., & Tao, T. (2007). The Dantzig selector: Statistical estimation when p is much larger than n . *The Annals of Statistics*, 2313-2351.
- [5] Ferraty, F., & Hall, P. (2015). An algorithm for nonlinear, nonparametric model choice and prediction. *Journal of Computational and Graphical Statistics*, 695-714.

برآورد ذخیره با به‌کارگیری باورمندی و بوت‌استرپ در روش نردبان زنجیری

حمیده داریوش همدانی^۱، زهرا عسگری^۲، جلال یادگاری^۳

^۱ هیات علمی دانشگاه شهید بهشتی، تهران
h-hamedani@sbu.ac.ir

^۲ دانشجوی کارشناسی ارشد بیم‌سنجی؛ دانشگاه شهید بهشتی، تهران
z.asgari@mail.sbu.ac.ir

^۳ کارشناس ارشد بیم‌سنجی؛ دانشگاه شهید بهشتی، تهران
yadegari1411@gmail.com

چکیده مختصر

یکی از مهم‌ترین وظایفی که شرکت‌های بیمه به منظور حفظ توانگری مالی خود انجام می‌دهند، تعیین ذخیره‌ای برای مطالبات معوق است. روش‌های مختلفی برای برآورد میزان این ذخایر معرفی شده است که روش نردبان زنجیری از قدیمی‌ترین و ساده‌ترین روش‌ها است. اما این روش به عنوان یک الگوریتم قطعی معرفی می‌شود. به همین دلیل قادر به محاسبه میزان خطا در این روش نیستیم. به منظور دست یافتن به خطای برآورد لازم است که از یک مدل تصادفی استفاده شود. مقالات زیادی در مورد مدل‌های تصادفی که الگوریتم نردبان زنجیری را تولید می‌کنند، منتشر شده است. به عنوان مثال می‌توان مدل مک را که یک مدل تصادفی آزاد توزیع بوده و برآورد یکسانی با الگوریتم نردبان زنجیری می‌دهد را می‌دهد نام برد. اما این مدل‌های تصادفی فقط از داده‌های فردی یک مثلث تاخیر برای برآورد میزان ذخایر استفاده می‌کنند. ما علاقه‌مند به برآورد ذخایر، با استفاده از اطلاعات پیشین هستیم. بنابراین، از روش نردبان زنجیری بیزی استفاده می‌کنیم. در این روش فرض می‌شود که عامل‌های نردبان زنجیری، دارای توزیع پیشینی بوده و با استفاده از این توزیع پیشین می‌توان عامل‌های نردبان زنجیری و در نتیجه مقدار ذخایر را برآورد کرد. در بسیاری از این حالت‌ها که پیدا کردن یک شکل بسته برای برآوردگر بیز امکان‌پذیر نیست، می‌توان از روش‌های عددی مثل MCMC استفاده کرد و یا کلاس برآوردگرهای ممکن را به کلاس کوچکتری که تابع خطی از مشاهدات هستند، محدود کرد که به آن برآوردگر باورمندی گفته می‌شود. مزیت استفاده از برآوردگر باورمندی این است که به جای توزیع پیشین، فقط به دو گشتاور اول و دوم آن نیاز است. این گشتاورها را می‌توان از روی داده‌های یک داکتیمان با مخاطره‌های مشابه به دست آورد.

در این مقاله، طی یک مطالعه موردی و با استفاده از داده‌های واقعی یک شرکت بیمه، ذخایر مربوط به بیمه شخص ثالث پنج شهر بزرگ (تهران، اصفهان، مشهد، شیراز و ساری) را با کمک برآوردگرهای باورمندی و با روش نردبان زنجیری محاسبه کرده و میانگین توانهای دوم خطای شرطی را برای هر دو دیدگاه بلند مدت و دیدگاه کوتاه مدت بدست می‌آوریم. نتیجه مهم این پژوهش آن است که میانگین توانهای دوم خطای شرطی که با استفاده از باورمندی به دست آمده است، برای هر دو دیدگاه بلند مدت و کوتاه مدت، از میانگین توانهای دوم خطای شرطی که با استفاده از نردبان زنجیری به دست آورده‌ایم، کمتر است. بنابراین برآوردگرهای باورمندی در داده‌های مربوط به شهرهای

بزرگ ایران، بهتر از برآوردهای نردبان زنجیری هستند. همچنین با استفاده از داده‌های این پنج شهر، تکنیک بوت استرپ را برای بدست آوردن ذخیره و میانگین توانهای دوم خطای شرطی به کار برده ایم. مقدار ذخایر برآورد شده با روش بوت استرپ، برای همه‌ی شهرها، بیشتر از ذخایری هستند که با روش باورمندی به دست آمده اند. علاوه بر زیاد بودن مقدار ذخایر به دست آمده، مقدار میانگین توانهای دوم خطای شرطی نیز در این روش، نسبت به دو روش باورمندی و نردبان زنجیری بیشتر است. در نتیجه، تکنیک بوت استرپ در این حالت، عملکرد ضعیفی داشته و تصحیح آن به منظور کسب خروجی‌های مناسب‌تر، موضوع پژوهش‌های بعدی خواهد بود.

کلمات کلیدی

ذخیره‌سازی ادعا، نردبان زنجیری، بوت استرپ، میانگین توانهای دوم خطای شرطی، برآوردهای باورمندی.

مراجع

- [1] Bühlmann, H., De Felice, M., Gisler, A., Moriconi, F., and Wüthrich, M. (2009), "Recursive credibility formula for chain ladder factors and the claims development result." ASTIN Bulletin, 39, 275–306.
- [2] Bühlmann, H. and Gisler, A. (2005), A Course in Credibility Theory and its Applications, Springer, Berlin.
- [3] De Vylder.(1982), "Estimation of IBNR claims by credibility theory," Insurance: Math. Econom,1, 35–40.
- [4] Gisler, A. and Wüthrich, M. (2008), "Credibility for the chain ladder reserving method," ASTIN Bulletin, 38, 565–600.
- [5] Mack, T. (1993), "Distribution-free calculation of the standard error of chain ladder reserve estimates." ASTIN Bulletin, 32, 213–225
- [6] Mack, T. (2000), "Credible claims reserves: The Benktander method," ASTIN Bulletin, 30,333- 347.
- [7] Merz, M. and Wüthrich, M. (2008), "Modelling the claims development result for solvency purposes," CASE-Forum, 542–568.

برآورد ذخایر در رشته درمان گروهی و رشته شخص ثالث مطالعه موردی: بیمه‌های درمان گروهی شرکت بیمه البرز و شخص ثالث بیمه پاسارگاد ۱۳۹۵-۱۳۹۰

امیر الفت^۱، زهره مسلمانی^۲

^۱ واحد تحقیق و توسعه، شرکت بیمه البرز، تهران،
amir_olfat@yahoo.com

^۲ گروه محاسبات فنی، شرکت بیمه پاسارگاد، تهران
z.mosalmanni@gmail.com

چکیده مختصر

در صنعت بیمه اغلب خساراتی که در یک سال مشخص اتفاق می‌افتد، در همان سال نهایی نمی‌شوند و یا حتی گزارشی به بیمه‌گر در خصوص خسارت ارجاع نمی‌شود که به هر حال باید در آینده جبران شوند. این بدان معناست که برای خساراتی که وجود دارند اما مقادیر احتمالی آنها نامعلوم است باید ذخیره فنی نگهداری شود. محاسبه تعهدات معوق کلیه خسارت‌های اتفاق افتاده و تعیین ذخایر آنها از امور شرکت‌های بیمه بوده که با روش‌های گوناگونی در حال اجراست. یکی از روش‌های آشنا در این حوزه روش نردبان زنجیره‌ای است که به علت عدم پیچیدگی محاسباتی می‌تواند مورد استفاده قرار گیرد. روش دیگر استفاده از مدل‌های خطی تعمیم یافته در ساختار نردبان زنجیری است که از دقت بالاتری نسبت به روش مذکور برخوردار است. در این مقاله تلاش بر آن شده است که ضمن بیان این دو روش، با اعمال آن‌ها بر روی داده‌های خسارت درمان گروهی بیمه البرز و شخص ثالث بیمه پاسارگاد طی پنج سال اخیر کارایی و عملکرد آن‌ها را با روش‌های پیش‌بینی مورد استفاده در حال حاضر مقایسه نماییم.

کلمات کلیدی

ذخیره IBNR، روش نردبان زنجیره‌ای، مدل‌های خطی تعمیم یافته، ذخایر فنی، بیمه‌های درمان گروهی، بیمه شخص ثالث.

۱- مقدمه

در ادبیات حسابداری مؤسسات بیمه، ذخایر فنی از مهم‌ترین بخش‌ها و عناصری است که همواره مورد تأکید بوده و به آن پرداخته شده است. شورای عالی بیمه در راستای اجرای ماده ۶۱ قانون تأسیس بیمه مرکزی ج.ا.ا و آیین‌نامه مصوب شماره ۵۸ شورای عالی بیمه و اصول حاکمیت شرکتی با هدف ایجاد شفافیت برای بیمه‌گذاران و صاحبان سهام مؤسسات بیمه، چگونگی لحاظ کردن ذخایر فنی در حساب‌های مؤسسات بیمه را تحت عنوان آیین‌نامه ذخایر فنی مؤسسات بیمه تصویب کرده است. در برخی موارد اجرای موارد فوق همواره با اشکالاتی از سوی نهاد ناظر، مدیران و کارشناسان صنعت بیمه مواجه بوده است. از جمله چالش‌های موجود در آیین‌نامه فعلی، مغایرت آن در برخی موارد نسبت به استاندارد حسابداری ۲۸ با استناد به گزارش سازمان حسابرسی، آمار و اطلاعات بخش نظارت مالی بیمه مرکزی ج.ا.ا. است. عدم استفاده از روش‌ها و مدل‌های استاندارد آکچوئری در برآورد

ذخایر فنی، عدم وجود سازوکار لازم جهت نظارت بر کفایت ذخایر فنی شرکت‌ها و عدم شفافیت و صحت در ارقام مربوط به ذخایر فنی در ترازنامه‌های شرکت‌های بیمه باعث تضییع حقوق بیمه‌گذاران و سهامداران شرکت‌ها می‌شود. با این وجود و با توجه به شرایط شرکت‌های بیمه و چالش‌های پیش‌رو، برای محاسبه ذخایر فنی باید از سازوکاری مناسب‌تری استفاده شود و روش‌ها و مدل‌های استاندارد در ذخیره‌گیری مورد بازنگری قرار گرفته و از علوم اکچوئرال و نرم افزارهای محاسباتی بروز بهره گرفته شود.

بر اساس قوانین موجود، شرکت‌های بیمه موظفند در پایان هر دوره مالی، ذخایر فنی خود را در چارچوب آیین‌نامه مصوب شماره ۵۸ شورای عالی بیمه و الحاقیه‌های بعدی آن، با عنوان «ذخایر فنی موسسات بیمه» محاسبه و در صورت‌های مالی افشاگری نمایند. ذخایر فنی در واقع اقلام تعیین کننده در ترازنامه و صورت حساب سود و زیان است که با احتساب موارد زیر محاسبه می‌شود: ذخایر فنی حق بیمه، ذخایر فنی حق بیمه برگشتی، ذخیره فنی تکمیلی، ذخایر خسارات معوق، ذخایر فنی خسارات طبیعی، ذخیره فنی ریاضی، ذخیره فنی مشارکت در منافع بیمه‌گذاران عمر و ذخیره خسارات واقع شده ولی اعلام نشده که بخشی از ذخایر خسارات معوق است. حصول اطمینان نسبت به شناسایی هزینه‌ها و بدهی‌های مرتبط با خسارات در دوره مالی مربوطه حائز اهمیت است. خسارت، ناشی از حوادث تحت پوششی است که در طول دوره بیمه‌نامه اتفاق می‌افتد. در پایان سال مالی، بیمه‌گران به‌ویژه در بازارهای کشورهای توسعه‌یافته، ذخیره‌ای برای خسارت‌هایی که بیمه‌گر از آنها اطلاع ندارد در نظر می‌گیرند. مواردی پیش می‌آید که خسارت واقع شده و تحت پوشش بیمه قرار دارد ولی بیمه‌گر در زمان بستن حساب‌ها از آنها بی‌اطلاع است. برخی از این موارد به قرار زیر است:

- خسارت در روزهای پایانی سال اتفاق افتاده و بیمه‌گذار فرصت نکرده است تا به بیمه‌گر اعلام کند.
- بیمه‌گذار از تحت پوشش بیمه‌ای بودن خسارت اطمینان ندارد.
- بیمه‌گذار از تحقق خطر منجر به خسارت بی‌خبر است.
- بیمه‌گذار قادر نیست که به بیمه‌گر اطلاع بدهد
- فرایند تعیین و اعلام مبلغ خسارت از سوی بیمه‌گذار یا سازمان‌های واسطه (بیمارستان‌ها، دادگاه‌ها و ...) به زمان بیشتری نیاز دارد.

بدهی مربوط به خسارت واقع شده‌ای که پرداخت نشده است (شامل خسارت گزارش نشده (IBNR))، باتوجه به تجربیات در خصوص خسارت واقع شده در گذشته و آخرین اطلاعات موجود برآورد و شناسایی می‌شود. برای برآورد مبلغ خسارت یاد شده، نه تنها تجربیات گذشته بلکه تغییر شرایط از قبیل آخرین حوادث فاجعه‌آمیز، تغییر در آئین‌نامه‌ها و تغییر در حجم و ترکیب خطرات تحت پوشش بیمه باید در نظر گرفته شود. بیمه‌گران با استفاده از تجربیات گذشته خود و با توجه به میزان و ترکیب پرتفوی شرکت بیمه، میزان ذخیره خسارت معوق را تعیین می‌کنند.

۲- توصیف داده‌های بیمه‌ای

همانطور که واضح است یکی از دلایل اصلی رشد معوقات صناعات بیمه، ناشی از معوقات در دو رشته درمان و شخص ثالث است که شرکت بیمه را موظف به نگهداری ذخیره IBNR در این دو رشته می‌کند. با توجه به سرفصل‌های صورت‌های مالی عبارت ذخیره خسارت معوق شامل خسارت‌های اعلام شده پرداخت نشده و خسارت‌های اعلام نشده می‌باشد. لازم به ذکر است داده‌های استفاده شده در این مقاله مجموع خسارت‌های پرداختی است که پس از برآورد این داده‌ها نیز دارای همین ماهیت می‌باشد که شامل خسارت معوق و IBNR است.

در اینجا برای ارزیابی روش‌های مطرح شده در بخش قبل، داده‌های خسارت بیمارستانی در رشته درمان گروهی شرکت بیمه البرز و همچنین داده‌های مربوط به خسارت شخص ثالث در بیمه پاسارگاد را مورد بررسی قرار داده ایم.

بدین منظور گزارش‌های مربوط به خسارت‌های بیمارستانی بخش درمان گروهی از سال ۱۳۹۰ تا سال ۱۳۹۵ بر مبنای سال وقوع خسارت و سال پرداخت خسارت تهیه گردید. شایان ذکر است بیمه‌های درمان گروهی شاخه‌ای از بیمه‌های اشخاص است و خسارت‌های این رشته به دو بخش بیمارستانی و پاراکلینیکی تقسیم می‌گردد. با توجه به اینکه اغلب معوقات مربوط به خسارت‌های

بیمارستانی می‌شود بنابراین در جدول شماره ۲ نردبان زنجیره ای مربوط به خسارت های معوق تجمعی برای این رشته ارائه گردیده است.

جدول ۲: خسارت درمان بیمه البرز (ارقام به میلیون ریال)

سال پرداخت خسارت						سال صدور
ششم	پنجم	چهارم	سوم	دوم	اول	
۴۴۳،۰۱۵	۴۴۳،۰۱۵	۴۴۳،۰۱۵	۴۴۲،۹۶۱	۴۳۹،۱۹۴	۳۳۰،۰۸۷	۹۰
	۴۴۳،۷۴۹	۴۴۳،۷۴۹	۴۴۳،۷۱۴	۴۴۳،۲۱۳	۳۲۳،۲۷۴	۹۱
		۵۴۵،۳۱۴	۵۴۴،۹۰۰	۵۴۲،۸۸۵	۳۹۹،۸۹۶	۹۲
			۱،۴۶۹،۴۹۵	۱،۳۳۴،۳۲۸	۷۶۳،۹۰۸	۹۳
				۲،۲۳۳،۸۹۱	۱،۴۶۴،۶۴۵	۹۴
					۱،۷۳۷،۰۶۸	۹۵

همانطور که در جدول شماره ۲ دیده می‌شود در بیمه های درمان شرکت بیمه البرز اکثر خسارت ها در دو سال اول پرداخت شده و درصد کمتری در سال‌های سوم به بعد پرداخت می‌شود. دلیل رشد خسارت از سال ۹۳ به ۹۴ ابلاغ آئین نامه‌ی بیمه های سلامت از سوی وزارت بهداشت و تغییر در تعرفه خدمات درمانی است. هدف ما برآورد مقادیر خالی جدول فوق به بهترین شکل ممکن است. در رشته شخص ثالث که یکی از انواع بیمه های مسئولیت است، مسئولیت راننده را در مقابل اشخاص دیگر (ثالث) بیمه می‌کند. منظور از شخص ثالث تمامی اشخاصی است که در اثر حوادث و وسائل موتوری زمینی دچار زیان های بدنی و یا مالی می‌شوند. اعم از اینکه داخل و یا خارج از وسیله نقلیه بیمه شده باشند، به غیر از راننده مسبب حادثه. داده های مورد استفاده در این مقاله مربوط به خسارت های پرداختی به صورت تجمعی در سال‌های ۹۰ تا ۹۵ رشته شخص ثالث شرکت بیمه پاسارگاد می‌باشد. نظیر جدول شماره ۲ داده های خسارت پرداختی بیمه‌ی شخص ثالث در جدول شماره ۳ آمده است.

جدول ۳: خسارت شخص ثالث بیمه پاسارگاد (ارقام به میلیون ریال)

سال پرداخت خسارت						سال صدور
ششم	پنجم	چهارم	سوم	دوم	اول	
۸۶۰،۴۱۷	۸۵۷،۶۴۳	۸۳۱،۴۷۲	۷۵۷،۸۱۵	۵۸۰،۷۰۶	۲۵۵،۹۴۸	۹۰
	۱،۳۶۴،۶۵۵	۱،۳۴۳،۸۵۴	۱،۱۹۹،۷۴۹	۷۴۶،۴۰۷	۱۳۱،۹۸۵	۹۱
		۱،۸۹۱،۹۷۴	۱،۷۴۰،۸۴۸	۱،۰۷۶،۸۷۳	۲۱۷،۰۳۰	۹۲
			۲،۱۹۰،۵۱۶	۱،۴۱۳،۳۱۱	۲۵۵،۳۳۸	۹۳
				۱،۸۳۲،۳۴۰	۳۸۸،۴۸۳	۹۴
					۴۵۴،۳۲۶	۹۵

با توجه به جدول فوق به خوبی بلند مدت بودن پرداخت خسارت در این رشته نمایان است. همچنین از عوامل موثر در افزایش سالیانه خسارات در این رشته می‌توان به پرداخت دیه یوم الادا در این خسارات اشاره کرد. این تصمیم که از سوی نهادهای ناظر به صورت سالیانه اتخاذ می‌گردد تاثیرات خود را بر روی داده‌های زیان این رشته می‌گذارد.

مراجع

- [۱] بیمه مرکزی جمهوری اسلامی ایران. آئین نامه شماره ۵۸.
- [۲] بیمه مرکزی جمهوری اسلامی ایران. استاندارد حسابداری شماره ۲۸
- [3] Rob Kaas, Marc Goovaerts, Jan Dhaene and Michel Denuit;2008- Modern Actuarial Risk Theory Using R (second edition)-chapter 10
- [4] Gillian Z. Heller P;2008- Generalized Linear Models for Insurance Data

برآوردهای استوار مدل‌های ARMA در قلمروهای زمان و فرکانس

مسعود یارمحمدی^۱، مهناز لشنی^۲

^۱گروه آمار، دانشگاه پیام نور، تهران،

masyar@pnu.ac.ir

mah.lashane60@gmail.com

چکیده مختصر

هدف اصلی از این تحقیق تعمیم روشهای رگرسیونی استوار به مدل‌های سربهای زمانی $ARMA(p,q)$ و کاربرد آن برای برآورد پارامترها در قلمرو زمان و فرکانس بر اساس تابع اتوکواریانس استوار باقیمانده (RA) می باشد. با استفاده از روشهای شبیه سازی کارایی این برآوردها را در مقایسه با روش کمترین توان دوم خطا مورد بحث و بررسی قرار می گیرد.

کلمات کلیدی

نقاط دور افتاده، استوار سازی، مدل اتورگرسیو و میانگین متحرک، تابع اتوکواریانس استوار باقیمانده.

۱- مقدمه

هدف اصلی این تحقیق، تعمیم روشهای رگرسیون استوار به مدل‌های $ARMA(p,q)$ و به نوعی کاهش اثر نقاط دور افتاده در برازش بهترین الگو می باشد، روسیو و لروی (۱۹۸۷). فرض کنید X_t یک سری زمانی مانا و معکوس پذیر $ARMA(p,q)$ به صورت

$$\phi_p(B)(X_t - \mu) = \theta_q(B)Z_t \quad (1)$$

که در آن $Z_t \sim N(0, \sigma^2)$ ، میانگین X_t و $\phi_p(B)$ و $\theta_q(B)$ به ترتیب عملگرهای اتورگرسیو و میانگین متحرک اند. به طور کلی در مشاهدات جمع آوری شده سری زمانی نقاط دور افتاده می تواند به صورت جمع پذیر (AO) و ابداعی (IO) رخ دهند. حال با توجه به امکان وقوع بیشتر نقاط دور افتاده از نوع اول توجه خود را بر روی این دسته از داده ها معطوف کرده و نخست به معرفی آن می پردازیم. سری زمانی X_t با وجود نقطه دور افتاده جمع پذیر از یک فرآیند $ARMA$ نبوده و بجای آن فرآیند $X_t = W_t + V_t$ می باشد، که در آن W_t یک فرآیند $ARMA$ با $Z_t \sim N(0, \sigma^2)$ و V_t یک دنباله مستقل از متغیرها و مستقل از دنباله W_t می باشد. متغیرهای V_t دارای توزیع H به صورت $H = (1 - \varepsilon)\delta_0 + \varepsilon G$ می باشد، که در آن δ_0 دارای توزیع تباهیده در مبداء و G یک توزیع دلخواه است.

۲- برآورد اتوکواریانس استوار باقیمانده ها (RA)

در یک سری زمانی با یک فرآیند $ARMA(p,q)$ داده شده در (۱) برآورد کمترین توان دوم برای پارامترهای $\lambda = (\phi, \theta, \mu)$ با

مینیمم کردن $\sum_{t=p+1}^n Z_t^2$ بدست می آید. به طوریکه Z_t باقیمانده های مدل بوده و عبارت است از:

$$Z_t = \theta_q^{-1}(B)\phi_p(B)(X_t - \mu) \quad (2)$$

که برای $t < p+1$ فرض می‌شود که $Z_t = 0$. این برآوردها به نقاط دور افتاده AO در مقایسه با نقاط دور افتاده IO حساس تر بوده و کسر کوچکی از این نقاط می‌توانند به برآوردهایی با اریبی بالا منجر شوند. ایده اصلی مربوط به برآوردهای RA نمایش برآوردهای کمترین توان دوم (LS) در شکلی شامل برآوردهای کوواریانس باقیمانده ها و سپس استوار ساختن برآوردهای (LS) با استوار کردن برآوردهای کواریانس به روش مرسوم است. از رابطه (۲) برآوردهای LS که از مینیمم کردن مجموع توان دوم خطا بدست می‌آید را در نظر می‌گیریم. با مشتق گیری از این عبارت $p+q+1$ معادله بشرح زیر برای برآوردهای LS حاصل می‌شود.

$$\sum_{t=p+1}^n Z_t \left(\frac{dZ_t}{d\phi_j} \right) = \sum_{t=p+1}^n Z_t \phi_p^{-1}(B) Z_{t-j} = 0 \quad 1 \leq j \leq p$$

$$\sum_{t=p+1}^n Z_t \left(\frac{dZ_t}{d\theta_j} \right) = \sum_{t=p+1}^n Z_t \theta_q^{-1}(B) Z_{t-j} = 0 \quad 1 \leq j \leq q$$

$$\sum_{t=p+1}^n Z_t \left(\frac{dZ_t}{d\mu} \right) = \left[-\left(1 - \sum_{i=1}^p \phi_i\right) / \left(1 + \sum_{j=1}^q \theta_j\right) \right] \sum_{t=p+1}^n Z_t = 0 \Rightarrow \sum_{t=p+1}^n Z_t = 0 \quad (3)$$

از بسط سریهای $\phi_p^{-1}(B)$ و $\theta_q^{-1}(B)$ در توانهایی از B به ترتیب با استفاده از ضرایب s_j و g_i داریم:

$$\left(\frac{1}{n}\right) \sum_{t=p+1+j}^n Z_t \sum_{h=0}^{\infty} s_h Z_{t-j-h} = 0 \quad 1 \leq j \leq p$$

$$\left(\frac{1}{n}\right) \sum_{t=p+1+j}^n Z_t \sum_{h=0}^{\infty} g_h Z_{t-j-h} = 0 \quad 1 \leq j \leq q \quad \sum_{t=p+1}^n Z_t = 0 \quad (4)$$

فرمولهای که در معادلات (۴) با اندیس $t=p+1+j$ ظاهر شده اند معادل آنهایی است که در معادلات (۳) آمده است ولی با شرط اولیه اینکه اگر $i < p+1$ ، $Z_t = 0$ است. این نتیجه با توجه به اینکه داریم:

$$t < p+1+j \Rightarrow \phi_p^{-1}(B) Z_{t-j} = \sum_{h=0}^{\infty} s_h Z_{t-j-h} = 0$$

صادق است. لذا با استفاده از شرط اینکه $Z_t = 0$ برای $t < p+1$ و تعویض ترتیب مجموعها داریم:

$$\sum_{h=0}^{n-j-p-1} s_h \gamma'(h+j) = 0 \quad 1 \leq j \leq p$$

$$\sum_{h=0}^{n-j-p-1} g_h \gamma'(h+j) = 0 \quad 1 \leq j \leq q \quad \sum_{t=p+1}^n Z_t = 0 \quad (5)$$

که در آن $\gamma'(k) = \sum_{t=p+1}^{n-k} (Z_t Z_{t+k}) / n$ یک کلاس از برآوردهای استوار با در نظر گرفتن اتوکوواریانس های استوار باقیمانده

به صورت $\gamma'_R(k) = \frac{1}{n} \sum_{t=p+1}^{n-k} \eta\left(\frac{Z_t}{\hat{\sigma}}, \frac{Z_{t+k}}{\hat{\sigma}}\right)$ تعریف می‌شود. بوستوس و یوهای (۱۹۸۶) با شبیه سازی مقدار $\hat{\sigma}$ را برابر

$\hat{\sigma} = \text{Median}(|Z_{p+1}|, |Z_{p+2}|, \dots, |Z_n|) / 0.6745$ محاسبه و پیشنهاد نموده است. انتخاب تابع η را به دوروش انجام می‌گیرد:

(۱) نوع مالوس با انتخاب: $\eta_M(u, v) = \psi(u)\psi(v)$

(۲) نوع همپل با انتخاب: $\eta_H(u, v) = \psi(uv)$

به طوریکه ψ تابعی فرد و پیوسته می باشد. تابع η به روشهای دیگری ممکن است انتخاب شوند. بعنوان مثال خانواده هوبر

$$\psi_{H,c}(u) = \text{Sign}(u) \text{Min}(|u|, c), \quad 2 \leq c \leq 3$$

داده شده که $\text{Sign}(u)$ تابع علامت می باشد. قابل ذکر است که اگر $\eta(u, v) = \psi(u)v$ آنگاه برآورد RA مجانباً معادل با M -برآود ها می باشد. لذا برآوردهای RA فرآیندهای $ARMA(p, q)$ از حل معادلات

$$\sum_{h=0}^{n-j-p-1} S_h \gamma'_R(h+j) = 0 \quad 1 \leq j \leq p$$

$$\sum_{h=0}^{n-j-p-1} g_h \gamma'_R(h+j) = 0 \quad 1 \leq j \leq q$$

$$\sum_{t=p+1}^n \psi\left(\frac{Z_t}{\hat{\sigma}}\right) = 0$$

بدست می آیند. اگر توزیع فرآیند Z_t ها متقارن باشد، توزیع مجانبی برآوردهای RA توسط بوستوس و یوهای (۱۹۸۶) مورد بررسی قرار گرفته است.

۱-۲- محاسبه برآوردهای RA

اگر $Z_t^* = \psi\left(\frac{Z_t}{\hat{\sigma}}\right)\hat{\sigma}$ و $X_t^* = \hat{\mu} + \hat{\phi}_p^{-1}(B)\hat{\theta}_q(B)Z_t^*$ تعریف شده و ψ از خانواده هوبر باشد، در این صورت سری زمانی X_t^* می تواند به عنوان سری X_t وینزوری (استوار) شده محسوب گردد. این خاصیت پیشنهاد یک الگوریتم تکراری محاسباتی که با برآوردهای اولیه از پارامترها آغاز شده و هر زمان اثرات غیرنرمال بودن از داده ها را زدوده و برآورد جدیدی از پارامترها تا همگرایی کامل بدست آورد، هوبر (۱۹۸۱).

۲-۲- نتایج شبیه سازی

نتایج مربوط به یک شبیه سازی برای یک مورد ارایه می شود. اگر $X_t = W_t + V_t$ و $W_t = 0.4W_{t-1} + Z_t$ به طوری که $Z_t \sim N(0, \sigma^2)$ و V_t متغیرهای تصادفی $i.i.d$ مستقل از Z_t و دارای توزیع $H = (1-\varepsilon)N(0, \sigma^2) + \varepsilon G$ ، به طوریکه $G \approx N(0, \tau^2)$. نتایج مربوط به میانگین توان دوم خطا (MSE) برای نمونه هایی با اندازه ۱۰۰ برای داده های شبیه سازی شده در حضور و عدم حضور داده دور افتاده از نوع جمع پذیر در جدول (۱) ارایه شده است. تعداد تکرار برای انجام این شبیه سازی ۵۰۰ می باشد. روشن است که برآوردهای استوار از مقدار MSE کمتر و کارایی بیشتری برخوردار است.

جدول (۱)

Estimate	$\varepsilon = 0$			$\varepsilon = 0.1, \tau^2 = 9$		
	Mean	MSE	EFF	Mean	MSE	EFF
Lease Squares	0.389	0.78	1	0.223	4.32	1
RA based $\eta_{HH,c}$	0.387	0.79	0.987	0.346	3.38	1.278
RA based $\eta_{MH,c}$	0.388	0.80	0.975	0.357	3.15	1.371

۳- تابع چگالی طیف استوار یک سری زمانی

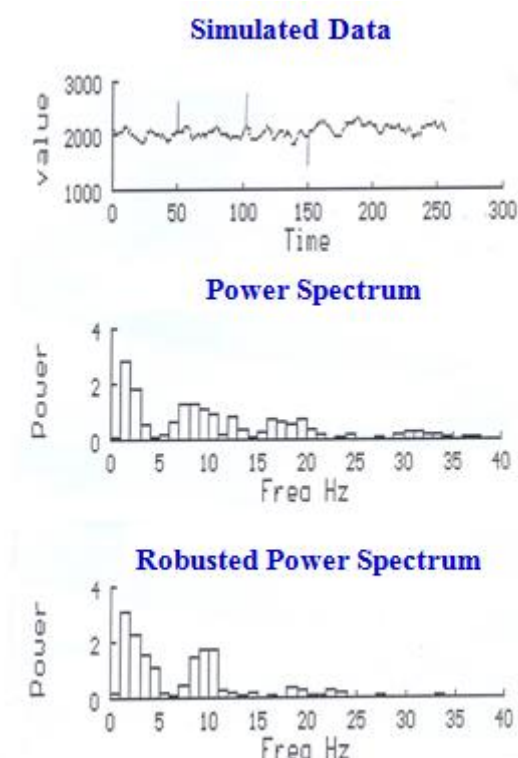
در این بخش کاربرد تابع اتوکوریانس استوار باقیمانده (برآوردهای RA) را در یافتن تابع چگالی طیف استوار را برای داده های شبیه سازی شده نشان می دهیم. اگر X_t فرایندی منای حقیقی مقدار $ARMA(p, q)$ و $\hat{\theta}_q^r(B)$ و $\hat{\phi}_p^r(B)$ از جایگزینی

برآوردهای استوار بدست آمده از روش برآورد اتوکواریانس استوار باقیمانده ها (RA) باشند. برآورد تابع چگالی طیفی برای این فرایند عبارت است از:

$$\hat{f}(\omega_j) = \frac{\hat{\sigma}_a^2}{2\pi} \frac{\hat{\theta}_q^r(e^{-i\omega}) \hat{\theta}_q^r(e^{i\omega})}{\hat{\phi}_q^r(e^{-i\omega}) \hat{\phi}_q^r(e^{i\omega})}$$

۳-۱ نتایج شبیه‌سازی

در این بخش با استفاده از روش‌های شبیه‌سازی ضرورت به‌کاربردن روش‌های استوار سازی تابع چگالی طیف استوار برای سری زمانی شبیه سازی شده در بخش ۲-۲ در دو حالت حضور و عدم حضور داده دور افتاده از نوع جمع پذیر در شکل (۱) نشان داده شده است.



شکل ۱: نمودار سری زمانی شبیه‌سازی شده، توان طیف کلاسیک و استوار

مراجع

- [1] Bustos, O. H. and Yohai, V. J. (1986), Robust Estimates of ARMA models, J. AM. Stat. Ass. , Vol.81, No. 393, 155-168.
- [2] Huber, P. (1981).Robust Statistics. Wiley, New York.MR0606374.
- [3] Rousseeuw P. J. and Leroy, A. M. (1987), Robust regression and outlier detection, John Wiley, and Sons.

بررسی آماری تغییرات دبی در حوزه سد اکباتان همدان

امید حمیدی^۱، زهره مریانجی^۲

^۱ گروه علوم پایه، دانشگاه صنعتی همدان، همدان،
omid_hamidi@hut.ac.ir

^۲ گروه جغرافیا، دانشگاه سید جمال الدین اسدآبادی، اسدآباد
z.maryanaji@gmail.com

چکیده مختصر

در این مطالعه پارامترهای هواشناسی (حداقل، حداکثر، متوسط دما و بارش سالانه) حوزه سد اکباتان و روابط آنها با میزان دبی به منظور تعیین روندهای زمانی بلندمدت داده‌ها و تغییرات احتمالی آن، طی دوره بیست ساله مورد تجزیه و تحلیل واقع شد. بدین منظور میزان همبستگی هر یک از سری داده‌های (حداقل، حداکثر، متوسط دما و بارش سالانه) با میزان دبی مورد بررسی قرار گرفت سپس هر یک از سری‌ها بر اساس آزمون‌های آماری منکنندال بررسی شدند. با توجه به روش ناپارامتریک منکنندال، داده‌های دما سالانه دارای روند افزایشی معنا داری هستند در حالیکه داده‌های بارش دارای نوسانات و جهش‌های بسیاری است اما در آن روند معناداری دیده نمی‌شود بر این اساس داده‌های دبی که در بین این پارامترها بیشترین همبستگی را با بارش دارد نیز دارای نوسان است و مانند بارش روند معناداری در آن دیده نمی‌شود. در بین پارامترهای حرارتی بیشترین همبستگی مربوط به متوسط دما می‌باشد. نتایج این مطالعه نشان می‌دهد که دمای حوزه مورد نظر گرمتر شده است، از کاربرد نتایج استفاده در پهنه‌بندی و پیش‌بینی‌های اقلیمی و همچنین برنامه‌ریزی و مدیریت منابع آب منطقه می‌باشد.

کلمات کلیدی

حوزه سد اکباتان، آزمون منکنندال، روند. ضریب همبستگی.

برآورد نقطه‌ای پارامتر تنش-مقاومت در توزیع نمایی دوپارامتری بر اساس نمونه‌ی مجموعه‌ی رتبه‌دار رکوردی

امینه صادق‌پور^۱، احمد نزاکتی^۲، سیدمهدی صالحی^۳

^۱ دانشجوی دکتری، گروه آمار دانشکده ریاضی، دانشگاه صنعتی شاهرود

Amineh_64@yahoo.com

^۲ دانشیار گروه آمار دانشکده ریاضی، دانشگاه صنعتی شاهرود

Anezakati591@gmail.com

^۳ استادیار گروه آمار دانشکده علوم، دانشگاه نیشابور

salehi2sms@gmail.com

چکیده مختصر

با توجه به اهمیت قابلیت اعتماد تنش-مقاومت در علوم مختلف به ویژه در صنعت، پزشکی و مهندسی و لزوم استفاده از روش مناسب در جمع‌آوری مشاهدات، برای دستیابی به میزان قابلیت اعتماد بالا، در کمترین زمان و با حداقل هزینه، در این مقاله با در نظر گرفتن روش نمونه‌گیری مجموعه‌ی رتبه‌دار رکوردی به برآورد نقطه‌ای پارامتر تنش-مقاومت با در نظر گرفتن توزیع نمایی دو پارامتری به عنوان توزیع جامعه‌ی آماری می‌پردازیم.

کلمات کلیدی

تنش-مقاومت، نمونه‌ی مجموعه‌ی رتبه‌دار رکوردی، برآوردگر درست‌نمایی ماکسیمم، برآوردگر ناریب با کمترین واریانس، توزیع نمایی دو پارامتری

۱- مقدمه

قابلیت اعتماد تنش-مقاومت در حقیقت احتمال سالم ماندن مؤلفه یا به عبارتی احتمال عدم شکست می‌باشد که به صورت $\mathbb{R} = \Pr(X < Y)$ نشان داده می‌شود. که در آن متغیر تصادفی X بیان‌گر فشار وارد شده بر یک مؤلفه و متغیر تصادفی Y مقاومت در برابر فشار می‌باشد. برای مثال فرض کنید متغیر تصادفی Y نشان‌دهنده‌ی مقاومت یک ساختمان در هنگام زمین‌لرزه و متغیر تصادفی X فشار ناشی از زمین‌لرزه باشد، آن‌گاه \mathbb{R} می‌تواند احتمال طراحی یک ساختمان ایمن باشد. با توجه به کاربردهای فراوان قابلیت اعتماد تنش-مقاومت در علوم مختلف، علاقه‌مندیم میزان قابلیت اعتماد با در نظر گرفتن حداقل زمان و کمترین هزینه تا حد امکان نزدیک به یک باشد. لذا برای به‌دست آوردن برآوردهای به اندازه کافی دقیق برای \mathbb{R} ، نیازمند استفاده از روش مناسب در جمع‌آوری مشاهدات نمونه می‌باشیم. صالحی و احمدی (۲۰۱۴)، به معرفی طرحی برای جمع‌آوری آماره‌های رکوردی که در بعضی موارد جایگزین مناسبی برای طرح رکوردهای معمولی می‌باشد، پرداخته‌اند. در این طرح با تعداد آزمایش‌های کمتر نسبت به طرح رکوردهای معمولی، به سطح اطلاع فیشر داده‌شده در رابطه با پارامتر مجهول جامعه دست یافته‌اند. این طرح برگرفته از ساختار نمونه‌ی رتبه‌دار می‌باشد و تحت عنوان نمونه‌ی مجموعه‌ی رتبه‌دار رکوردی ($RRSS$)، معرفی شده است. در این طرح، با در نظر گرفتن π دنباله‌ی مستقل، آمین نمونه در صورت مشاهده‌ی آمین رکورد متوقف می‌شود. در حقیقت در این طرح، آخرین رکورد از هر دنباله تنها داده‌ی در دسترس می‌باشد.

$$\begin{aligned} 1 &: U_{(1)1} && \rightarrow U_{1,1} = U_{(1)1} \\ 2 &: U_{(1)2} \quad U_{(2)2} && \rightarrow U_{2,2} = U_{(2)2} \\ &\vdots \quad \vdots \quad \vdots \quad \ddots && \vdots \quad \vdots \\ n &: U_{(1)n} \quad U_{(2)n} \quad \dots \quad U_{(n)n} && \rightarrow U_{n,n} = U_{(n)n} \end{aligned} \quad (1)$$

اگر قرار دهیم $U_{i,i} = U_{(i),i}$ ، $i=1, \dots, m$ ، آن‌گاه بردار $U = (U_{1,1}, U_{2,2}, \dots, U_{m,m})$ را یک نمونه‌ی مجموعه‌ی رتبه‌دار رکوردی به اندازه m می‌گوییم. قابل ذکر است رکوردهای حاصل از طرح مذکور مستقل از هم و این امکان وجود دارد که مرتب نباشند.

فرض کنید جامعه‌ی آماری تحت بررسی دارای تابع چگالی $f(\cdot)$ و تابع توزیع احتمال $F(\cdot)$ باشد و $u = (u_{1,1}, u_{2,2}, \dots, u_{m,m})$ مقدار مشاهده شده‌ی $U = (U_{1,1}, U_{2,2}, \dots, U_{m,m})$ از $RRSS$ بالا باشد، در این صورت با استفاده از تابع چگالی حاشیه‌ای رکوردهای معمولی (آرنولد و همکاران، ۱۹۹۸)، تابع چگالی توأم U به صورت زیر به دست می‌آید.

$$f_U(u, \theta) = \prod_{i=1}^m \frac{\{-\log \bar{F}(u_{i,i}, \theta)\}^{i-1}}{(i-1)!} f(u_{i,i}, \theta), \quad (2)$$

صالحی و احمدی (۲۰۱۵)، با در نظر گرفتن توزیع نمایی به عنوان توزیع جامعه آماری به برآورد پارامتر تنش-مقاومت براساس $RRSS$ پرداخته‌اند. از جمله کارهای انجام شده در زمینه تنش-مقاومت براساس داده‌های ترتیبی می‌توان به بکلیزی (۲۰۱۴)، بصیرت و همکاران (۲۰۱۵) اشاره نمود. در این مقاله با در نظر گرفتن مدل $RRSS$ و با فرض اینکه متغیرهای تصادفی X و Y مستقل و هم توزیع با توزیع نمایی دو پارامتری می‌باشند، به برآورد نقطه‌ای پارامتر تنش-مقاومت می‌پردازیم.

۲- برآوردگر درست‌نمایی ماکسیمم

می‌دانیم تابع چگالی توزیع نمایی دو پارامتری به صورت زیر می‌باشد

$$f(x|\mu, \theta) = \frac{1}{\theta} e^{-\frac{1}{\theta}(x-\mu)}, \quad x > \mu, \quad -\infty < \mu < \infty, \quad \theta > 0 \quad (3)$$

که در آن μ و θ به ترتیب پارامتر مکان و مقیاس می‌باشند. فرض می‌کنیم $r = (r_{1,1}, r_{2,2}, \dots, r_{n,n})$ بردار مشاهدات $R = (R_{1,1}, R_{2,2}, \dots, R_{n,n})$ نمونه‌ای به حجم n از رکوردهای بالای نمونه‌ی مجموعه‌ی رتبه‌دار با توزیع $E(\mu_1, \theta_1)$ و $s = (s_{1,1}, s_{2,2}, \dots, s_{m,m})$ نیز بردار مشاهدات $S = (S_{1,1}, S_{2,2}, \dots, S_{m,m})$ نمونه‌ای به حجم m از رکوردهای بالای مجموعه‌ی رتبه‌دار با توزیع $E(\mu_2, \theta_2)$ می‌باشند. با جایگذاری رابطه (۳) در (۲) برآورد درست‌نمایی ماکسیمم پارامترهای (μ_1, θ_1) و (μ_2, θ_2) به صورت زیر به دست می‌آید.

$$\hat{\mu}_1 = \min\{r_{i,i} \mid 1 < i < n\} \quad \text{و} \quad \hat{\theta}_1 = \frac{\sum_{i=1}^n (r_{i,i} - \hat{\mu}_1)}{N} \quad (4)$$

$$\hat{\mu}_2 = \min\{s_{j,j} \mid 1 < j < m\} \quad \text{و} \quad \hat{\theta}_2 = \frac{\sum_{j=1}^m (s_{j,j} - \hat{\mu}_2)}{M} \quad (5)$$

که در آن $N = \frac{n(n+1)}{2}$ و $M = \frac{m(m+1)}{2}$. لذا با انجام محاسباتی برآورد درست‌نمایی ماکسیمم پارامتر تنش-مقاومت را به صورت زیر به دست می‌آوریم.

$$\hat{R}_{ML} = \begin{cases} 1 - \frac{\hat{\theta}_1}{\hat{\theta}_1 + \hat{\theta}_2} e^{-\frac{1}{\hat{\theta}_1}(\hat{\mu}_2 - \hat{\mu}_1)} & \hat{\mu}_2 - \hat{\mu}_1 \geq 0 \\ \frac{\hat{\theta}_2}{\hat{\theta}_1 + \hat{\theta}_2} e^{-\frac{1}{\hat{\theta}_2}(\hat{\mu}_1 - \hat{\mu}_2)} & \hat{\mu}_2 - \hat{\mu}_1 < 0 \end{cases} \quad (6)$$

حال با در نظر گرفتن حالت‌های خاص برآورد درست‌نمایی ماکسیمم پارامتر تنش-مقاومت را محاسبه می‌نماییم.
حالت اول: $(\mu_1 = \mu_2 = \mu)$

در صورتی که پارامتر مکان در دو جامعه برابر باشد، برآورد درست‌نمایی ماکسیمم پارامترهای $(\mu, \theta_1, \theta_2)$ براساس $RRSS$ بالا برابر است با

$$\hat{\mu} = \min\{r_{i,i}, s_{j,j}, \quad 1 \leq i \leq n \quad 1 \leq j \leq m\} \quad (7)$$

$$\hat{\theta}_1 = \frac{1}{N} \sum_{i=1}^n (r_{i,i} - \hat{\mu}) \quad (8)$$

$$\hat{\theta}_2 = \frac{1}{M} \sum_{j=1}^m (s_{j,j} - \hat{\mu}) \quad (9)$$

لذا

$$\hat{R}_{ML} = \frac{\hat{\theta}_2}{\hat{\theta}_1 + \hat{\theta}_2} \quad (10)$$

حالت دوم: $(\theta_1 = \theta_2 = \theta)$

در صورتی که پارامترهای مقیاس دو جامعه برابر باشند، برآورد درست‌نمایی ماکسیمم (μ_1, μ_2, θ) براساس $RRSS$ بالا به صورت زیر به دست می‌آید.

$$\hat{\mu}_1 = \min\{r_{i,i}, 1 \leq i \leq n\} \quad (11)$$

$$\hat{\mu}_2 = \min\{s_{j,j}, 1 \leq j \leq m\} \quad (12)$$

$$\hat{\theta} = \frac{\sum_{i=1}^n (r_{i,i} - \hat{\mu}_1) + \sum_{j=1}^m (s_{j,j} - \hat{\mu}_2)}{N + M} \quad (13)$$

که به سادگی نتیجه می‌شود

$$\hat{R}_{ML} = \begin{cases} 1 - \frac{1}{\hat{\theta}} e^{-\frac{1}{\hat{\theta}}(\hat{\mu}_2 - \hat{\mu}_1)} & \hat{\mu}_2 - \hat{\mu}_1 \geq 0 \\ \frac{1}{\hat{\theta}} e^{-\frac{1}{\hat{\theta}}(\hat{\mu}_2 - \hat{\mu}_1)} & \hat{\mu}_2 - \hat{\mu}_1 < 0 \end{cases} \quad (14)$$

۳- برآوردگر ناریب با کمترین واریانس

در این بخش برآوردگر ناریب با کمترین واریانس پارامتر تنش-مقاومت \mathfrak{R} را در حالتی که پارامترهای مکان در دو جامعه برابر با مقدار معلوم μ باشند را محاسبه می‌نماییم. با توجه به این که $\left(\sum_{i=1}^n R_{i,i}, \sum_{j=1}^m S_{j,j} \right)$ آماره بسنده‌ی کامل برای (θ_1, θ_2) می‌باشد و با در نظر گرفتن $I(R_{1,1} < S_{1,1})$ به عنوان یک برآوردگر ناریب برای پارامتر \mathfrak{R} ، به طوری که $I(A)$ تابع نشانگر پیشامد A می‌باشد. می‌توانیم با استفاده از قضیه‌ی راثو-بلکول و لهما و شفه (لهمن و کسلا ۱۹۹۸، را ببینید)، برآوردگر $UMVU$ پارامتر \mathfrak{R} را با در نظر گرفتن سه حالت به صورت زیر محاسبه نماییم.

حالت اول: $\min(m, n) \geq 2$

در این حالت با انجام محاسباتی داریم

$$\hat{R}_{UMVU} = F_W \left(\frac{\sum_{j=1}^m (S_{j,j} - \mu)}{\sum_{i=1}^n (R_{i,i} - \mu)} \right) \quad (15)$$

که $F_W(\cdot)$ بیان‌گر تابع توزیع تجمعی متغیر تصادفی W می‌باشد و $W = \frac{\sum_{j=2}^m (S_{j,j} - \mu)}{\sum_{i=2}^n (R_{i,i} - \mu)}$ آماره فرعی و بنا به

قضیه باسو، مستقل از آماره بسنده‌ی کامل می‌باشد.

حالت دوم: $(n = 1, m \geq 2)$

واضح است در این حالت $\left(R_{1,1}, \sum_{j=1}^m S_{j,j} \right)$ آماره بسنده‌ی کامل برای $(\theta_1, \theta_2)^T$ می‌باشد و همانند حالت قبل داریم

$$\hat{R}_{UMVU} = 1 - \left(1 - \frac{R_{1,1} - \mu}{\sum_{j=1}^m (S_{j,j} - \mu)} \right)^{M-1} \quad (16)$$

حالت سوم: $(n \geq 2, m = 1)$

در این حالت $\left(\sum_{i=1}^n R_{i,i}, S_{1,1} \right)$ آماره بسنده کامل و همانند حالت‌های قبل داریم

$$\hat{R}_{UMVUE} = 1 - \left(1 - \frac{S_{1,1} - \mu}{\sum_{i=1}^n (R_{i,i} - \mu)} \right)^{N-1} \quad (17)$$

مراجع

- [1] Baklizi, A. (2014), Interval estimation of the stress-strength reliability in the two parameter exponential distribution based on records, *Journal of Statistical Computation and Simulation*, 84, 2670-2679.
- [2] Basirat, M., Baratpour, S. and Ahmadi, J. (2015), *Statistical inferences for stress-strength in the proportional hazard models based on progressive Type-II censored samples*. *Journal of Statistical Computation and Simulation*, 85, 431-449.
- [3] Arnold, B. C. Balakrishnan, N. and Nagaraja, H. N. (1998), *Records*, John Wiley and Sons, New York.
- [4] Salehi, M. Ahmadi, J. (2014), Record ranked set sampling scheme. *METRON*, 72, 351-365.
- [5] Salehi, M., Ahmadi, J. (2015), Estimation of stress-strength using record ranked set sampling scheme from the exponential distribution, *Filomat*, 29, 1149-1162.

بررسی تأثیر شوک‌های پولی بر شاخص‌های حسابداری، اقتصادی و بازار در شرکت‌های بورس اوراق بهادار تهران

جعفر عبدالمالکی

گروه حسابداری، دانشگاه آزاد اسلامی، ملایر،
abdolmaleke@gmail.com

چکیده مختصر

پژوهش حاضر به منظور بررسی تأثیر شوک‌های پولی بر شاخص‌های حسابداری، اقتصادی و بازار در شرکت‌های بورس اوراق بهادار تهران انجام گرفت، این تحقیق از حیث هدف تحقیق، از جمله تحقیقات کاربردی و بر حسب طبقه بندی بر مبنای روش از نوع تحقیقات همبستگی است. ابزار گردآوری اطلاعات، اسناد کاوی و بانک‌های اطلاعاتی می باشد و داده‌های مورد نیاز شامل سود عملیاتی، هزینه حقوق و دستمزد، هزینه استهلاک، دارایی مشهود، جمع داراییها، حقوق صاحبان سهام، درآمد و قیمت سهام شرکت‌های نمونه آماری در اطلاعات مالی مندرج در صورت‌های مالی حسابرسی شده آنها خواهد بود. جامعه آماری تحقیق، شامل کلیه شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران می باشد. با توجه به دوره زمانی ۶ ساله تحقیق (از ابتدای سال ۱۳۸۹ پایان سال ۱۳۹۴)، شرکت‌هایی انتخاب گردیده اند که حداقل در ابتدای سال ۱۳۸۹ به عضویت بورس اوراق بهادار تهران در آمده باشند و هم چنین دوره مالی آن‌ها منتهی به ۲۹ اسفند ماه باشد. روش نمونه‌گیری به صورت مرحله‌ای و با حذف سیستماتیک بوده است که تعداد ۹۹ شرکت به عنوان حجم نمونه انتخاب شد. یافته‌ها نشان می‌دهد که شوک پولی انبساطی بر نرخ بازده دارایی شرکت‌های بورس اوراق بهادار تهران تأثیر معنا داری دارد. همچنین نتایج نشان داد که شوک پولی انقباضی بر نرخ بازده دارایی شرکت‌های بورس اوراق بهادار تهران تأثیر معنا داری دارد. یافته‌ها نشان می‌دهد که شوک پولی انبساطی بر ارزش افزوده اقتصادی شرکت‌های بورس اوراق بهادار تهران تأثیر معنا داری دارد. یافته‌ها نشان می‌دهد که شوک پولی انقباضی بر ارزش افزوده بازار شرکت‌های بورس اوراق بهادار تهران تأثیر معنا داری دارد. یافته‌ها نشان می‌دهد که شوک پولی انقباضی بر ارزش افزوده بازار شرکت‌های بورس اوراق بهادار تهران تأثیر معنا داری دارد.

کلمات کلیدی

شوک‌های پولی، شاخص‌های حسابداری، اقتصادی، بازار، شرکت‌های بورس اوراق بهادار تهران.

بررسی روند تغییرات سرانه دستمزد در صنایع ایران

سیما فرهادی^۱، آیت اله موسوی^۲

^۱ محقق، بانک مرکزی جمهوری اسلامی ایران، تهران
sima007farhadi@gmail.com

^۲ رئیس دایره آمار صنعتی اداره آمار اقتصادی، بانک مرکزی جمهوری اسلامی ایران، تهران
sima007farhadi@gmail.com

چکیده مختصر

بخش صنعت در ایران یکی از مولفه‌های اصلی بخش واقعی اقتصاد است. سهم این بخش از تولید ناخالص داخلی طی دهه اخیر حدود ۱۵ درصد بدون نفت و حدود ۱۲ درصد با احتساب نفت بوده است. بانک مرکزی از سال ۱۳۴۳ تاکنون با اجرای طرح بررسی آمار صنعتی، اطلاعاتی از متغیرهای مختلف در حوزه صنعت کشور جمع‌آوری می‌نماید. یکی از این متغیرها سرانه دستمزد می‌باشد که در این مقاله روند تغییرات دستمزد اسمی و واقعی در رشته فعالیت‌های مختلف صنعتی کشور طی دهه اخیر از دیدگاه آماری مورد تجزیه و تحلیل قرار می‌گیرد.

کلمات کلیدی

سرانه دستمزد، ارزش تولید سرانه، دستمزد اسمی، دستمزد واقعی و توزیع آماری.

مراجع

- [۱] طرح بررسی آمار صنعتی، اداره آمار اقتصادی بانک مرکزی، سال‌های ۹۲-۱۳۸۳.
- [۲] کردبچه، حمید و سوری، علی (۱۳۹۰). "منابع تفاوت بین صنعتی دستمزدها در صنایع بزرگ ایران"، فصلنامه مدل‌سازی اقتصادی (سال پنجم، شماره ۲، صفحات ۲۷-۴۲).
- [3] Véronique Genre, Karsten Kohndand and Daphne Momferatou.(2009). Understanding inter-industry wage structures in the euro area . European central bank, working paper series, No 1022, March 2009.

بهبود عملکرد الگوریتم‌های پردازش تصاویر سیاه و سفید بر اساس

شباهت در یادگیری ماشین

سید محمد حسینی^۱، پیمان پهلوانی^۲

^۱دانشگاه تحصیلات تکمیلی علوم پایه زنجان، زنجان،

Seyyedmohammad@iasbs.ac.ir

pahlevani@iasbs.ac.ir

چکیده مختصر

در یادگیری ماشینی به موضوعات مرتبط با یافت شیوه‌ها و الگوریتم‌هایی می‌پردازیم که بر اساس آنها کامپیوترها (در مفهوم کلی) بتوانند توانایی تعلّم و یادگیری را داشته باشند. صنایع پزشکی، رباتیک (مانند وسیله‌های نقلیه بدون سرنشی) و خیلی دیگر از بخش‌های صنعتی و غیر صنعتی در حال استفاده از یادگیری ماشین در کامپیوترها هستند. این کامپیوترها به تدریج و با افزایش داده‌ها، کارایی بهتری در انجام وظیفه مورد نظر پیدا می‌کنند. تصویر یکی از انواع این داده‌ها است و پردازش و تحلیل این داده‌های تصویری نیز از شیوه‌های یادگیری این کامپیوترها می‌باشد. می‌دانیم که برای انجام عملیات پردازش و یادگیری، کامپیوترها باید زمان و انرژی صرف کنند بنابراین هر چقدر الگوریتم و شیوه تحلیل این داده‌ها از پیچیدگی کمتر و دقت بیشتری برخوردار باشد در زمان و انرژی صرفه جویی می‌شود. در مقاله ارائه شده بر روی کاهش پیچیدگی محاسباتی و افزایش دقت تصمیم‌گیری در پردازش تصاویر سیاه و سفید در کامپیوترها و مقایسه آنها با بانک اطلاعاتی پرداخته شده است بطوری که بتوانیم عملیات تصمیم‌گیری و مقایسه با بانک اطلاعاتی را در زمان بلادرنگ انجام دهیم. همچنین روش پیشنهادی باعث کاهش حداقل ۳۰ درصدی عملیات پردازشی در مقایسه با موارد مشابه بدون افت دقت تصمیم‌گیری می‌شود.

کلمات کلیدی

یادگیری ماشین، تصاویر سیاه سفید

بررسی عوامل موثر بر کیفیت سود با تأکید بر اندازه شرکت نمونه مورد مطالعه: شرکتهای بیمه فعال در بازار بیمه ایران

غلامرضا کشاورز حداد^۱، سیف اله امید^۲ و شمس قاسمی^۳

^۱ دانشیار دانشکده مدیریت و اقتصاد، دانشگاه صنعتی شریف، تهران

g.k.haddad@sharif.edu

^۲ دانشجوی کارشناسی ارشد، موسسه آموزشی عالی رجا، قزوین

s.omidi@samaninsurance.ir

^۳ دکتری اقتصاد مالی، بیمه مرکزی جمهوری اسلامی ایران، تهران

shghasemi.tmu@gmail.com

چکیده مختصر

بر اساس مبانی نظری حسابداری و گزارشگری مالی، حسابداری بر مبنای فرض‌های اولیه‌ای استوار است که شالوده اصول، استانداردها و روش‌های شناسایی و اندازه‌گیری را تشکیل می‌دهند از جمله این مفروضات حسابداری، فرض تعهدی می‌باشد. فرض تعهدی شناسایی درآمدها را به محض تحقق و هزینه‌های را به محض تحمل صرفنظر از جریانهای ورودی و خروجی وجه نقد حاصل از آنها را محقق می‌نماید. فرض مذکور مخالف فرض نقدی است. احتساب فرض تعهدی و نقدی در حسابداری بیمه، کیفیت سود گزارش شده در گزارش‌های مالی را متأثر می‌سازد. بنابراین با وجود مسأله مذکور، هدف اصلی این مقاله بررسی کیفیت سود با تأکید بر عامل اندازه شرکت است. با استفاده از تکنیک اقتصاد سنجی با رویکرد داده-های پانل برای ۲۹ شرکت فعال در بازار بیمه ایران طی سال‌های ۱۳۸۶-۱۳۹۳، بر پایه مدل مک نیکولز (۲۰۰۲)، این نتیجه حاصل شد، چهار متغیر خالص جریان وجه نقد عملیاتی سال قبل (با اثر مثبت)، خالص جریان وجه نقد عملیاتی سال جاری (با اثر منفی)، تغییر در فروش شرکتهای بیمه (اثر مثبت) و مجموع دارایی‌های شرکت (با اثر مثبت) تأثیر معنی‌داری بر تغییرات سرمایه در گردش (و یا به عبارتی کیفیت سود شرکتهای بیمه) داشته است و متغیر خالص جریان وجه نقد عملیاتی سال بعد شرکتهای بیمه تأثیر معنی‌داری بر تغییر در سرمایه در گردش آنها در دوره مورد بررسی نداشته است. همچنین تفاوت معنی‌داری بین تغییرات سرمایه در گردش شرکتهای بیمه بزرگ و کوچک وجود داشته است.

کلمات کلیدی

کیفیت سود، بیمه، اندازه شرکت، داده‌های پانل.

۱- مقدمه

استاندارد حسابداری شماره ۲۸ ایران؛ فعالیتهای بیمه عمومی در تیرماه سال ۱۳۸۵ تصویب شد و در مورد کلیه صورت‌های مالی که دوره مالی آنها از تاریخ اول فروردین ۱۳۸۶ و بعد از آن بود، اجرای آن ضروری شد اما عملاً تا تصویب آیین نامه شماره ۵۸ شورای عالی بیمه، اجرای آن به تعویق افتاد. بهر حال، مطابق با استاندارد حسابداری شماره ۲۸ ایران، درآمد حق بیمه باید از تاریخ شروع پوشش بیمه‌ای و به محض اینکه حق بیمه به گونه‌ای اتکاپذیر قابل اندازه‌گیری باشد، شناسایی شود. به دلیل حجم زیاد بیمه نامه‌های صادر

شده، پراکندگی جغرافیایی صدور بیمه‌نامه‌ها، عدم دسترسی به موقع به اطلاعات لازم و سایر محدودیت‌های موجود، تعیین تاریخ شروع پوشش بیمه‌ای در رابطه با هر یک از بیمه‌نامه‌های صادره دشوار و فاقد توجیه اقتصادی است. بنابراین مدیریت امکان پیدا خواهد نمود با ملاحظه شرایطی اقدام به اعمال قضاوت حرفه‌ای در برآورد شروع پوشش بیمه‌نامه نماید، به عبارت دیگر مدیریت بالقوه قادر خواهد بود شناسایی درآمد حق بیمه را تسریع نموده و یا به تعویق بیندازد. همچنین، براساس بند ۲۰ استاندارد مذکور، بدهی بابت خسارت معوق باید برای بیمه‌های مستقیم و اتکایی شناسایی شود. مبلغ خسارت معوق باید بیانگر برآورد معقولی از مخارج لازم برای تسویه تعهد فعلی در تاریخ ترازنامه باشد. بنابراین، این امکان برای مدیریت نیز وجود دارد تا براساس قضاوت حرفه‌ای اقدام به شناسایی هزینه خسارت و برآورد بدهی خسارت نماید. به عبارت دیگر، این امکان وجود دارد تا شناسایی بدهی را بیشتر یا کمتر برآورد نماید. بنابراین با غیر رویکرد حسابداری از فرض نقدی به فرض تعهدی در راستای تغییرات ماهوی در فعالیتهای تجاری، علیرغم تدوین استاندارد حسابداری شماره ۲۸ ایران، به طور بالقوه امکانات بیشتری را در اختیار مدیران برای مدیریت، دستکاری سود قرار می‌دهد که می‌تواند از مهم‌ترین دلایل تغییرات قابل توجه بر کیفیت سود گزارش شده گردد. هم‌چنین شفافیت صورت‌های مالی شرکت-های بیمه و نقش آن در سیاست‌گذاری‌های کلان صنعت بیمه از یک سو و کمک به تصمیم‌گیری سرمایه‌گذاران و فعالان بازار بورس از سوی دیگر، خواهد داشت. مسایل بیان شده بررسی کیفیت سود شرکتهای بیمه ایران و عوامل مؤثر بر تغییرات ایجاد شده در کیفیت سود را ضروری می‌نماید. با توجه به اهمیت و ضرورت بیان شده، هدف اصلی این مقاله بررسی عوامل مؤثر بر کیفیت سود شرکتهای فعال در بازار بیمه ایران با تاکید بر اندازه شرکت (بزرگی و کوچکی) است. به همین منظور این مقاله با طرح فرضیه‌های متناظر به دنبال پاسخ به این پرسش‌هاست که (۱) آیا جریانهای نقدی عملیاتی (سال جاری، سال قبل و سال آتی) شرکتهای بیمه ایران در تغییرات سرمایه در گردش آنها تاثیر دارد؟ و (۲) آیا اندازه شرکتهای بیمه ایران (کوچک و بزرگ) بر کیفیت سود آنها تاثیر دارد؟

۲- ادبیات موضوع

گزارش‌های مالی مهم‌ترین خروجی یک سیستم حسابداری محسوب می‌شود. هدف از گزارش‌های مالی فراهم کردن اطلاعاتی است که می‌تواند برای تصمیمات تجاری سودمند واقع شود (فرزانه و اوستا، ۲۰۱۳). تعاریف متعدد و متفاوتی برای کیفیت سود مطرح شده است. دچو و دیچف (۲۰۰۲) سود را براساس ارتباط اقلام تعهدی با جریانهای نقدی عملیاتی دوره گذشته، حال و آینده تعریف می‌کنند. در این صورت، کیفیت سود ناشی از کیفیت اقلام تعهدی خواهد بود. بال و شیواکومار (۲۰۰۵) مفهوم کیفیت را سودمندی صورت‌های مالی برای سرمایه‌گذاران، اعتباردهندگان، مدیران و تمامی بخش‌های مرتبط با شرکت تعریف می‌کنند. واضح است تعاریف متعدد، گستره‌ای از معیارها، روش‌ها و مدل‌های اندازه‌گیری را برای محققان در بر گرفته است. در بین تحقیق‌های انجام گرفته پیرامون کیفیت سود دو معیار پایداری و اقلام تعهدی کاربرد بیشتری داشته است. در این مقاله نیز سنگ‌بنای مدل تجربی بر معیار اقلام تعهدی استوار شده است. بر این اساس مدل مک نیکولز (۲۰۰۲) که تعدیل یافته مدل دچو و دیچف (۲۰۰۲) بر اساس کل دارایی‌ها است، به شکل زیر مبنای مدل‌سازی تجربی قرار گرفت:

$$\Delta WC_t = b_0 + b_1 CFO_{t-1} + b_2 CFO_t + b_3 CFO_{t+1} + b_4 \Delta Sale_t + \log TA_t + \varepsilon_t$$

در مدل فوق ΔWC_t بیانگر تغییر در سرمایه در گردش، CFO جریان نقد عملیاتی، $\Delta Sale$ تغییر در فروش، $\log TA_t$ لگاریتم طبیعی مجموع داراییها و ε خطای باقیمانده مدل است. محققان بسیاری در خارج و داخل کشور به سنجش کیفیت سود و بررسی رابطه آن سایر مفاهیم، متغیرها و محیط مالی پرداخته‌اند.

۳- روش‌شناسی و نتایج

با توجه به کوتاه بودن دوره زمانی داده‌های مربوط به الگوی نظری پژوهش و سایر مزایای بیان شده از داده‌های پانل (کشاورز حداد، ۱۳۹۵)، به منظور تخمین و برآزش الگوی تجربی از تکنیک اقتصادسنجی با رویکرد داده‌های پانل برای ۲۹ شرکت فعال در بازار بیمه ایران (به استثنای بیمه مرکزی ج.ا.ایران) مذکور طی سال‌های ۱۳۹۳-۱۳۸۶ (۲۳۹) داده آماری استخراج شده از صورت‌های مالی

حسابرسی شده مندرج در کدال و سایر منابع در دسترس) استفاده شده است. در نهایت، آزمون فرضیه‌های تحقیق با استفاده از روش تجزیه و تحلیل رگرسیون در داده‌های پانل انجام گردید.

نخست، با استفاده از آزمون چاو (F لیمر) این نتیجه حاصل شد که داده‌های مدل تجربی این پژوهش، امکان pooled کردن را فراهم می‌کنند. به عبارتی دیگر شرکت‌های بیمه مورد بررسی همگن بوده و برای تخمین مدل می‌توان از روش حداقل مربعات معمولی ساده (OLS) استفاده نمود و دیگر استفاده از روش پانل ضرورت ندارد. برای بررسی اندازه شرکت نیز از بین دو معیار فروش و دارایی-ها، با توجه به اینکه ضریب میانگین دارایی‌ها (با علامت اختصاری Large Mean) در مدل از نظر آماری معنی‌دار بود، از آن به عنوان معیار اندازه شرکت استفاده شد. نتایج حاصل از برآورد مدل (با استفاده از نرم‌افزار STATA) به شکل جدول زیر است:

نتایج برآورد الگوی تجربی برای کلیه شرکتهای بیمه

VARIABLES	(1) Fixed Effect	(2) OLS Regression	(3) OLS Regression
operational cash flow, (CFO_t)	-0.786*** (0.158)	-0.735*** (0.140)	-0.769*** (0.141)
operational cash flow = L, (CFO_{t-1})	0.938*** (0.108)	0.667*** (0.0911)	0.671*** (0.0899)
CFO_{t+1} (operational cash flow = F,	0.309** (0.153)	0.157 (0.120)	0.149 (0.120)
total premium = D, $(\Delta Sale_t)$	0.414*** (0.0466)	0.312*** (0.0320)	0.303*** (0.0320)
Log ta	81,887 (104,276)	92,265* (52,157)	40,211 (58,504)
Large Mean	437,814* (245,159)		356,727* (186,871)
Constant	-1.402e+06 (1.465e+06)	-1.203,719 (734,598)	-500,914 (816,706)
Observations	178	178	178
R-squared	0.554	0.628	0.636
Number of id	29		
Leamer F(28, 143) Stat	1.01(0.45)	NO	
Individual Binary	Yes	NO	
Time Binary	No	NO	

*** p<0.01, ** p<0.05, * p<0.1 Standard errors in parentheses

نتایج حاصل از مدل‌سازی تجربی نشان داد:

۱. چهار متغیر خالص جریان وجه نقد عملیاتی سال قبل (با اثر مثبت)، خالص جریان وجه نقد عملیاتی سال جاری (با اثر منفی)، تغییر در فروش شرکت‌های بیمه (اثر مثبت) و مجموع دارایی‌های شرکت (با اثر مثبت) تاثیر معنی‌داری بر تغییرات سرمایه در گردش (و یا به عبارتی کیفیت سود شرکت‌های بیمه) داشته است و متغیر خالص جریان وجه نقد عملیاتی سال بعد شرکت‌های بیمه تاثیر معنی-داری بر تغییر در سرمایه در گردش آنها در دوره مورد بررسی (۱۳۸۶-۱۳۹۳) نداشته است. بنابراین فرضیه اول این پژوهش برای چهار متغیر توضیحی خالص جریان وجه نقد سال جاری، سال قبل و تغییر در فروش شرکت‌های بیمه در سطح خطای ۱٪ (یا سطح اطمینان ۹۹٪) و مجموع دارایی‌های شرکت در سطح خطای ۱۰٪ (یا سطح اطمینان ۹۰٪) تأیید و برای متغیر خالص جریان وجه نقد عملیاتی سال بعد دارایی‌های شرکت‌های رد می‌شود.

۲. اندازه شرکت بر مقادیر تغییر در سرمایه در گردش آنها در سطح اطمینان ۹۰٪ تاثیر معنی‌دار داشته به طوری که شرکت‌های بزرگ به طور متوسط مقدار تغییر در سرمایه در گردش خود را در هر سال به مقدار ۳۵۶,۷۲۷ افزایش داده‌اند. بنابراین فرضیه "تفاوت معنی‌دار اندازه شرکت بر تغییرات سرمایه در گردش آنها تأیید می‌گردد."

مراجع

[۱] کشاورز حداد، غلامرضا. (۱۳۹۵)، اقتصادسنجی داده‌های خرد و ارزیابی سیاست، تهران: نشر نی چاپ اول.

- [2] McNichols, M. (2002) . Discussion of the Quality of Accruals and Earnings: the Role of Accrual Estimation Errors, *The Accounting Review*, 77 (supplementary) , 61-69.
- [3] Farzaneh,H., Osta,S.(2013).Ownership Structure and the quality of earnings: Evidence from iran, *Science road Journal* ,l(3):129-139.
- [4] Dechow, P. and Dichev, I. (2002) . The Quality of Accruals and Earnings: the Role of Accrual Estimation Errors, *The Accounting Review*, 77 (supplementary) , 35-59.
- [5] Ball, R. and Shivakumar, L. (2005). Earnings quality in UK private firms: comparative loss recognition timeliness. *Journal of accounting and economics*, 39 (1): 83-128.

تحلیل مدل‌های چندسطحی بیزی و کاربرد آن در صنعت

آرزو حبیبی راد^۱، رضا رضایی^۲

^۱ و ^۲ گروه آمار دانشگاه فردوسی مشهد، مشهد،

ahabibi@um.ac.ir

re.rezaiystat@stu-mail.um.ac.ir

چکیده مختصر

الگوهای چندسطحی یکی از پر کاربردترین الگوها در حوزه ی مدل های خطی است که برای مدل بندی داده هایی با ساختار همبستگی درون گروهی به کار می روند. تحلیل چندسطحی روشی کارا برای تجزیه و تحلیل داده های علوم اقتصادی، صنعتی، اجتماعی و پزشکی است و در واقع حالت بسط داده شده از مدل های خطی تعمیم یافته می باشند به طوری که در آن علاوه بر مدل بندی متغیر پاسخ، ضرایب رگرسیونی نیز مدل بندی می شوند و روشی کارا برای داده های طولی و همبسته می باشند. در این مقاله هدف استفاده از تحلیل های بیزی در برازش مدل های چندسطحی برای برآورد پارامترهای مدل و مولفه های واریانس با کمک نرم افزار MLwiN می باشد. همچنین در انتها برای توضیح بیشتر یک مثال کاربردی در بخش صنعت نیز آورده می شود.

کلمات کلیدی

مدل های چند سطحی، تحلیل بیزی، نرم افزار MLwiN.

- [1] Aitkin, M., and Wilson, G. T. (1980). Mixture models, outliers, and the EM algorithm. *Technometrics*, 22(3), 325-331.
- [2] Browne, W. J., and Draper, D. (2006). A comparison of Bayesian and likelihood-based methods for fitting multilevel models. *Bayesian analysis*, 1(3), 473-514.
- [3] Browne, W. J., and Rasbash, J. (2012). MCMC estimation in MLwiN. Centre of Multilevel Modelling, University of Bristol, UK.
- [4] Bryk, A. S., and Raudenbush, S. W. (1992). *Hierarchical linear models: applications and data analysis methods*. Sage Publications, Inc. US.
- [5] Draper, D. (2008). Bayesian multilevel analysis and MCMC. In *Handbook of multilevel analysis* (pp. 77-139). Springer New York.
- [6] Goldstein, H. (1987). *Multilevel models in education and social research*. Oxford University Press, London.
- [7] Goldstein, H. (1986). Multilevel mixed linear model analysis using iterative generalized least squares. *Biometrika* 73: 43-56.
- [8] Longford, N. T. (1987). A fast scoring algorithm for maximum likelihood estimation in unbalanced mixed models with nested random effects. *Biometrika*, 74(4), 817-827.

توزیع لاپلاس چوله اصلاح شده و کاربرد آن

فاطمه آرزومند^۱، مسعود یارمحمدی^۲
^۱ دانشجوی دکتری آمار، دانشگاه پیام نور، تهران،
F_arezoomand2007@yahoo.com
^۲ هیات علمی گروه آمار، دانشگاه پیام نور، تهران
masyar@pnu.ac.ir

چکیده مختصر

در این تحقیق به معرفی شکل جدیدی از توزیع لاپلاس چوله می‌پردازیم. در ابتدا خواص توزیعی آن را مورد توجه قرار می‌دهیم. در ادامه برآورد ماکسیمم درستنمایی پارامترهای آن را به روش عددی و با استفاده از الگوریتم ژنتیک به دست آورده و با انجام روش‌های شبیه‌سازی کارایی آنها را براساس نمونه‌هایی با اندازه‌های مختلف مقایسه می‌کنیم. در پایان براساس یک نمونه داده واقعی نشان می‌دهیم که نتایج حاصل از آزمون نیکویی برازش توزیع لاپلاس چوله اصلاح شده نسبت به فرم متقارن آن به شکل محسوسی بهبود یافته‌اند.

کلمات کلیدی

توزیع لاپلاس، الگوریتم ژنتیک، برآورد ماکسیمم درستنمایی، شبیه‌سازی

۱- مقدمه

تا کنون تلاش‌های زیادی در زمینه گسترش کلاس‌های توزیع‌های آماری صورت گرفته و توزیع‌های زیادی با چولگی و کشیدگی معرفی شده است. به ویژه صورت‌های متنوعی از توزیع لاپلاس چوله معرفی و در زمینه‌های مالی و اقتصادی به کار گرفته شده‌اند (آریال، ۲۰۰۶). در این مقاله قصد ما بر آن است که مدل احتمال لاپلاس اصلاح شده متقارن را به لاپلاس اصلاح شده چوله تعمیم داده و نشان دهیم که انتخاب مناسب‌تری برای مجموعه‌ای از داده‌های واقعی است. توزیع لاپلاس که به توزیع نمایی دوپل نیز مشهور است توسط پیر-سیمون لاپلاس (۱۸۲۷-۱۷۴۹) معرفی شد. این توزیع توسط کاتز و همکاران (۲۰۰۱) به طرز کامل‌تری مورد بحث و بررسی قرار گرفته است. توزیع لاپلاس اصلاح شده متقارن $MCL(\theta, \sigma)$ دارای تابع چگالی احتمال زیر روی $(-\infty, \infty)$ است. (محمودوند و همکاران، ۲۰۱۵)

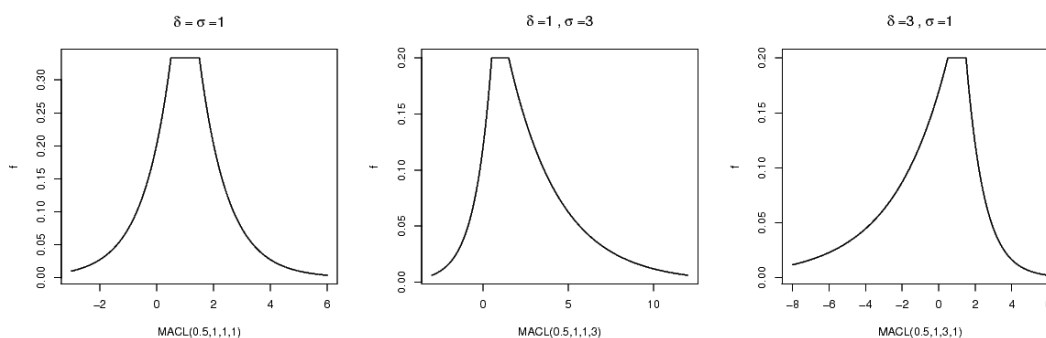
$$f(x, \theta, \sigma) = \frac{1}{3\sigma} \begin{cases} \exp\left\{-\frac{|x-\theta|}{\sigma}\right\} & x < \theta \\ 1 & \theta \leq x < \theta + \sigma \\ \exp\left\{-\frac{|x-\theta-\sigma|}{\sigma}\right\} & \theta + \sigma \leq x, \end{cases}$$

(۱)

که در آن $\theta \in R$ و $\sigma > 0$ هستند. با توجه به مسطح بودن این توزیع در قسمت وسط مد یکتا نیست و انتروپی آن نیز نسبت به فرم کلاسیک توزیع بیشتر است (ون کامپنهوت و کاور، ۱۹۸۱). هدف ما آن است که توزیع لاپلاس اصلاح شده متقارن را به فرم چوله آن $MACL(\theta, \mu, \delta, \sigma)$ تعمیم دهیم. تابع چگالی احتمال $MACL$ به صورت زیر است:

$$f(x, \theta, \mu, \delta, \sigma) = \frac{1}{\mu + \delta + \sigma} \begin{cases} \exp\left\{-\frac{|x-\theta|}{\delta}\right\} & x < \theta \\ 1 & \theta \leq x < \theta + \mu \\ \exp\left\{-\frac{|x-\theta-\mu|}{\sigma}\right\} & \theta + \mu \leq x, \end{cases} \quad (2)$$

که در آن $\theta \in R$ پارامتر مکان، $\mu > 0$ پارامتر سطح و $\sigma, \delta > 0$ پارامترهای مقیاس هستند. در شکل ۱ تابع چگالی احتمال $MACL$ به ازای مقادیر مختلف پارامترها ترسیم شده است.



شکل ۱. نمودار چگالی احتمال توزیع لاپلاس چوله اصلاح شده

قضیه ۱: فرض کنید $X \sim MACL(\theta, \mu, \delta, \sigma)$. در این صورت K -امین گشتاور خام X به صورت زیر است:

$$E(X^k) = \frac{1}{\mu + \delta + \sigma} \times \left[\delta \sum_{t=0}^k \binom{k}{t} t! (-1)^t \theta^{k-t} \delta^t + (\theta + \mu)^{k+1} - \theta^{k+1} + \sigma \left(\sum_{t=1}^k \sum_{j=0}^k \frac{k!}{t! j!} \sigma^{k-t-j} \theta^j + \sum_{t=0}^k \binom{k}{t} t! \sigma^{k-t} \theta^t \right) \right]. \quad (3)$$

قضیه ۲: فرض کنید $X \sim MACL(\theta, \mu, \delta, \sigma)$ چند Q_p به صورت زیر محاسبه می شود

$$Q_p = \begin{cases} \delta \ln \left(\frac{p(\mu + \delta + \sigma)}{\delta} \right) + \theta & x < \theta \\ p(\mu + \delta + \sigma) - \theta + \delta & \theta \leq x < \theta + \mu \\ -\sigma \ln \left(\frac{(\mu + \delta + \sigma)(1-p)}{\sigma} \right) + \theta + \mu & \theta + \mu \leq x. \end{cases} \quad (4)$$

با استفاده از روش تبدیل معکوس می‌توان از این توزیع نمونه‌گیری کرد. کافی است از توزیع یکنواخت نمونه‌گیری و سپس از معادله (۴) برای به دست آوردن چندک‌ها استفاده کرد.

۲- برآورد ماکسیمم درست‌نمایی پارامترهای $MACL(\theta, \mu, \delta, \sigma)$ و شبیه‌سازی

برای نوشتن تابع درست‌نمایی ابتدا تابع چگالی احتمال را به صورت زیر می‌نویسیم

$$f(\theta, \mu, \delta, \sigma) = \frac{1}{\mu + \delta + \sigma} \exp \left\{ \frac{x - \theta}{\delta} \right\}^{I_{(-\infty, \theta)}(x)} \exp \left\{ -\frac{x - (\theta + \sigma)}{\sigma} \right\}^{I_{(\theta + \mu, \infty)}(x)} \quad (5)$$

حال فرض کنید X_1, X_2, \dots, X_n یک نمونه تصادفی از $MACL(\theta, \mu, \delta, \sigma)$ باشد. با استفاده از (۵) لگاریتم تابع درست‌نمایی به صورت زیر نوشته می‌شود

$$l(\theta, \mu, \delta, \sigma) = -n \ln(\mu + \delta + \sigma) - \frac{1}{\delta} \sum_{i=1}^n (x_i - \theta)^- - \frac{1}{\sigma} \sum_{i=1}^n (x_i - \theta - \mu)^+, \quad (6)$$

که در آن $(x - \theta)^- = \max\{-(x - \theta), 0\}$ و $(x - \theta - \mu)^+ = \max\{x - \theta - \mu, 0\}$ و تمامی پارامترها مجهول هستند. با استفاده از الگوریتم ژنتیک (اسکروشا، ۲۰۱۳) پارامترهای مدل را به روش عددی برآورد کرده ایم. از بسته GA در نرم افزار R بدین منظور استفاده شده است. اندازه نمونه برای شبیه‌سازی به ترتیب ۱۰، ۳۰ و ۵۰ در نظر گرفته شده و تعداد تکرارها نیز ۱۰۰۰ است. میانگین و میانگین مربعات خطای برآورد پارامترهای توزیع برای ۱۰۰۰ تکرار در جدول ۱ آمده است. بدیهی است که با افزایش اندازه نمونه، میانگین مربعات خطای برآورد کاهش یافته است.

$(\theta, \mu, \delta, \sigma)$	size	$\hat{\theta}$		$\hat{\mu}$		$\hat{\delta}$		$\hat{\sigma}$	
		Mean	MSE	Mean	MSE	Mean	MSE	Mean	MSE
(1,0.2,0.1,0.2)	10	1.0001	0.0134	0.2606	0.0297	0.0708	0.0059	0.1292	0.0152
	30	0.9939	0.0070	0.2347	0.0260	0.0866	0.0025	0.1749	0.0051
	50	0.9990	0.0046	0.2079	0.0166	0.0930	0.0016	0.1876	0.0026

جدول ۱. ML به ازای حجم نمونه‌های مختلف برای $MACL(\theta, \mu, \delta, \sigma)$

۳- کاربرد توزیع لاپلاس چوله اصلاح شده برای داده‌های واقعی

برای بررسی کارایی توزیع لاپلاس چوله اصلاح شده، داده‌های مربوط به دستمزد روزانه برحسب پوند در کشور انگلستان از سال ۱۲۶۰ تا ۱۹۹۴ (برگرفته از <https://datamarket.com>) مورد استفاده قرار داده‌ایم. مقدار آماره آزمون $K-S$ و p -value برای توزیع لاپلاس چوله اصلاح شده ۰/۰۶۴۳ و ۰/۰۰۴۶ و همین نتایج برای توزیع لاپلاس اصلاح شده متقارن به ترتیب ۰/۰۹۳۳ و ۰/۱۱۲/۰ که نشان دهنده سودمندی استفاده از $MACL$ است.

مراجع

- [1] Aryal, G.R. (2006), Study of Laplace and Related Probability Distributions and Their Applications. Ph.D Thesis, Department of Mathematics, University of South Florida
- [2] Kotz, S. Kozubowski, T. and Podgorski, K. (2001), The Laplace Distribution and Generalizations: A Revisit with New Applications. Springer.
- [3] L.Scrucca, (2013), GA: A Package for Genetic Algorithms in R, Journal of Statistical Science, Volume 53, Issue 4.
- [4] Mahmoudvand, R., Faradmal, J., Abbasi, N., Lurz, K. (2015). A New Modification of the Classical Laplace distribution. Journal of The Iranian Statistical Society. Vol.14.No. 2, pp 93-111.

خسارت‌های با توزیع پارتو در مدل‌های مخاطره شرکت‌های بیمه

ابوذر بازیاری^۱، ارمغان نهاری^۲، مانده به فروز^۳

^۱ استادیار گروه آمار، دانشگاه خلیج فارس بوشهر
ab_bazaryari@yahoo.com

^۲ دانشجوی کارشناسی ارشد آمار، دانشگاه خلیج فارس بوشهر
arm.nahari@gmail.com

^۳ دانشجوی کارشناسی ارشد آمار، دانشگاه خلیج فارس بوشهر
maedeh.bfz@gmail.com

چکیده مختصر

در این مقاله، به بررسی وجود خسارت‌های با توزیع پارتو در مدل‌های مخاطره بیمه پرداخته شده است. نشان داده می‌شود که وجود این نوع خسارت‌ها می‌توانند بیمه‌گر آن شرکت را دچار مشکل کرده و حتی شرکت را تا مرز ورشکستگی پیش ببرد. مدل بیمه اتکایی مازاد خسارت نوعی مدل بیمه است و در این مدل نشان داده می‌شود که برآورد میانگین خسارت‌های رخ داده شده از طرف بیمه‌گذاران، در صورتی که خسارت‌ها دارای توزیع پارتو باشند، برآوردی مناسب برای شرکت بیمه نخواهد بود. به‌عنوان یک کاربرد، داده‌های آتش‌سوزی شرکت بیمه کارآفرین استان بوشهر طی ۹ سال متوالی با سرمایه اولیه ثابت در نظر گرفته شده است. دیده می‌شود که این داده‌ها دارای توزیع دم سنگین خواهند بود و برآورد امید شرطی خسارت‌ها محاسبه شده است.

کلمات کلیدی

احتمال ورشکستگی، توزیع دم سنگین، توزیع پارتو، حق بیمه، مدل مخاطره جمعی بیمه

۱- مقدمه

لاندربرگ (۱۹۲۶)، بسیاری از مفاهیم ریاضی و آماری نظریه مخاطره را مورد بررسی قرار داد. رلسکی و همکاران (۱۹۹۹)، روش‌ها و نتایج جالبی را در نظریه مخاطره بدست آوردند که این روش‌ها در ارتباط با مسائل مربوط به بیمه و دارایی بودند. آسموسن (۲۰۰۰)، بسیاری از مفاهیم مدل‌های مخاطره را مورد بررسی قرار داد و برای هر کدام از این مدل‌ها با روش‌های مختلف احتمال ورشکستگی را محاسبه کرد. گزارشات کامل و دقیق از پیدایش فرایندهای مخاطره شرکت بیمه و بررسی احتمالات ورشکستگی در مدل‌های مختلف، توسط بازیاری و پرهام (۱۳۸۷) و بازیاری (۱۳۹۱) داده شده است. بازیاری (۱۳۹۵)، با یک روش کاملاً جدید احتمال ورشکستگی زمان نامتناهی در مدل مخاطره جمعی را محاسبه کرد. در این مقاله، ابتدا مشکلات با وجود خسارت‌های با توزیع پارتو در مدل مخاطره شرکت بیمه بیان شده است. با توجه به داده‌های بدست آمده از شرکت بیمه کارآفرین استان بوشهر، خواهیم دید که علاوه بر اینکه داده‌ها دارای توزیع پارتو هستند، می‌توانند باعث بروز مشکلات جدی در دریافت و پرداخت حق بیمه و خسارت‌های رخ داده شده شوند.

۲- مدل مخاطره جمعی شرکت بیمه

فرایند مخاطره شرکت بیمه یک فرایند بسیار مهم با چارچوبی تصادفی برای اطلاع از وضعیت سرمایه شرکت بیمه در یک دوره زمانی است. فرض کنید شرکت با سرمایه اولیه ثابت u شروع به فعالیت کند، آنگاه فرایند مخاطره جمعی شرکت بیمه به صورت زیر تعریف می‌شود:

$$R(t) = u + ct - \sum_{k=1}^{N(t)} X_k, \quad (1)$$

در رابطه (۱)، ct مجموع حق بیمه‌های دریافتی تا زمان t ، X_k ، $k = 1, 2, \dots, N(t)$ ، متغیر تصادفی اندازه خسارت رخ داده شده از طرف k امین بیمه‌گذار در بازه $[0, t]$ و $\sum_{k=1}^{N(t)} X_k$ مجموع خسارت‌های رخ داده شده تا زمان t می‌باشد. در واقع فرض می‌شود که شرکت در هر دوره زمانی مبلغ ثابت c را از هر بیمه‌گذار دریافت کند که آن را نرخ ناخالص حق بیمه مخاطره می‌گویند. در مدل مخاطره جمعی، $\{N(t) : t \geq 0\}$ فرایند تعداد خسارت‌های رخ داده شده از طرف بیمه‌گذاران تا زمان t است و معمولاً فرض می‌شود که $\{N(t) : t \geq 0\}$ دارای فرایند پواسون با پارامتر λ باشد. یعنی تساوی $E(N(t)) = \lambda t$ برقرار است.

۳- مشکلات در شرکت بیمه با وجود خسارت‌های دارای توزیع پارتو

۳-۱- برآورد خسارت‌ها در مدل‌های مخاطره بیمه

فرض کنید یک نمونه تصادفی n از بیمه‌گذاران با اندازه خسارت X_k ، $k = 1, 2, \dots, n$ گرفته شده باشد. آنگاه مقدار برآورد میانگین و واریانس خسارت‌ها عبارتند از:

$$\hat{\mu} = \frac{1}{n} \sum_{k=1}^n X_k, \quad \hat{\sigma}^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k - \hat{\mu})^2.$$

خسارت‌ها دارای هر توزیع آماری که باشند، این برآوردگرها ناریب هستند. اما نشان داده می‌شود که $\hat{\mu}$ برآوردی مناسب نیست اگر خسارت‌ها دارای توزیع پارتو باشند. فرض کنید متغیر تصادفی خسارت X_k ، $k = 1, 2, \dots, n$ ، دارای توزیع پارتو $Pa(\alpha, 1)$ ، $\alpha > 1$ ، باشد. با توجه به اینکه در توزیع پارتو پارامتر دوم، پارامتر مقیاسی است، بنابراین بدون اینکه از کلیت مساله کم شود فرض می‌شود که مقدار این پارامتر برابر با یک باشد. همچنین فرض کنید برای مقدار ثابت M ، نامساوی

$$X_n^{\max} = \max_{k \leq n} X_k \leq M,$$

برقرار باشد. ابتدا هدف این است که مقدار امید ریاضی شرطی $\hat{\mu}$ به شرط $X_n^{\max} \leq M$ و نیز خطای نسبی این امید شرطی در مقایسه با مقدار میانگین واقعی توزیع پارتو $Pa(\alpha, 1)$ ، یعنی $(\alpha - 1)^{-1}$ ، محاسبه شود. بدلیل هم‌توزیع بودن متغیرهای تصادفی، خواهیم داشت:

$$\begin{aligned} E[\hat{\mu} | X_n^{\max} \leq M] &= E[\hat{\mu} | X_1 \leq M, X_2 \leq M, \dots, X_n \leq M] \\ &= \frac{\alpha}{\alpha - 1} \frac{1 - (1 + M)^{-(\alpha - 1)}}{1 - (1 + M)^{-\alpha}} - 1. \end{aligned}$$

همچنین

$$\frac{E[\hat{\mu} | X_n^{\max} \leq M] - (\alpha - 1)^{-1}}{(\alpha - 1)^{-1}} = -\alpha \frac{(1 + M)^{-(\alpha - 1)} - (1 + M)^{-\alpha}}{1 - (1 + M)^{-\alpha}}.$$

فرض کنید برای عدد ثابت $p \in (0, 1)$ ، مقدار M طوری انتخاب شود که تساوی $P(X_n^{\max} \leq M) = p$ برقرار باشد. به عبارت دیگر به دلیل هم‌توزیع بودن متغیرهای تصادفی

$$P(X_n^{\max} \leq M) = P(X_1 \leq M, X_2 \leq M, \dots, X_n \leq M) = (P(X_1 \leq M))^n \\ = (1 - (1 + M)^{-\alpha})^n = p,$$

یا به طور معادل تساوی $(1 + M)^{-1} = \left(1 - p^{\frac{1}{n}}\right)^{\frac{1}{\alpha}}$ برقرار می‌باشد.

۲-۳- مدل بیمه‌انکایی مازاد خسارت

فرض کنید که خسارت‌های رخ داده شده از طرف بیمه‌گذاران دارای توزیع پارتو $Pa(\alpha, 1)$ ، $\alpha > 1$ ، باشند. برای متغیر تصادفی خسارت X_k ، $k = 1, 2, \dots, n$ ، مقدار امید شرطی $E[X_k - M | X_k > M]$ عبارت است از:

$$E[X_k - M | X_k > M] = \frac{\int_M^{\infty} x \frac{\alpha}{(1+x)^{\alpha+1}} dx}{\int_M^{\infty} \frac{\alpha}{(1+x)^{\alpha+1}} dx} - M = \frac{M+1}{\alpha-1}. \quad (2)$$

بنابراین

$$E[X_k | X_k > M] = \frac{M + (\alpha + 1)}{\alpha - 1}. \quad (3)$$

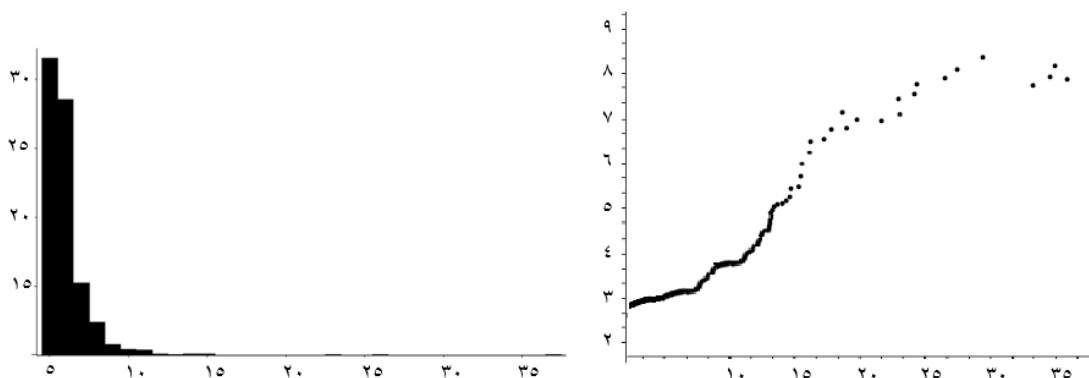
رابطه (۳) یک تابع افزایشی از M است؛ اما، در مدل بیمه‌انکایی مازاد خسارت، عبارت (۲) ممکن است برای شرکت بیمه مشکل‌ساز باشد، چرا که در مدل بیمه‌انکایی مازاد خسارت همواره $M > 0$ بوده و با اختیار مقادیر بزرگ M ، برآورد میانگین خسارت‌ها نیز بزرگ می‌شود. این موجب خواهد شد که طبق قرارداد، بیمه‌گر اتکایی مبلغ زیادتری را از بیمه‌گذاران خود دریافت کند. بنابراین وجود خسارت‌های دارای توزیع پارتو در مدل بیمه‌انکایی مازاد خسارت مناسب نیست.

۴- مثال کاربردی

در این بخش، نتایج به دست آمده در مقاله با مثال واقعی مورد بررسی قرار گرفته‌اند. در مثال، آزمون برازندگی جهت بررسی اینکه آیا واقعاً داده‌ها از توزیع پارتو آمده‌اند یا خیر انجام شده است.

مثال (داده‌های آتش‌سوزی شرکت بیمه کارآفرین استان بوشهر). یک نمونه تصادفی ۲۵۰ تایی از خسارت‌ها که مربوط به داده‌های آتش‌سوزی واحدهای صنعتی و کارگاهی شامل ساختمان، تأسیسات، ماشین‌آلات، مواد اولیه، کالای ساخته‌شده و موجودی‌های انواع کارخانه‌ها و کارگاه‌های فعال در زمینه‌های مختلف صنعتی شهرهای استان بوشهر طی ۹ سال متوالی از ابتدای سال ۱۳۸۶ تا پایان سال ۱۳۹۴ می‌باشد، از شرکت بیمه کارآفرین بوشهر دریافت و در وبسایت http://uplod.ir/gdv8n8cojahj/250-Random_samples_of_claim_sizes_in_Bushehr_province.pdf.htm گزارش داده شده‌اند. این بیمه‌نامه با سرمایه اولیه ثابت بوده و بیمه‌گذار مبلغ ثابتی را برای مدت یکسال پرداخت می‌کند. آزمون نیکویی برازش روی داده‌ها انجام شده، داده‌ها به ۸ رده تقسیم شده و مقدار آماره آزمون عدد ۸/۱۲ به دست آمده است. چون مقدار آماره از مقدار $\chi^2_{(0.95)}(7) = 14.07$ کمتر است، بنابراین فرضیه صفر رد نمی‌شود. به عبارت دیگر در سطح معنی داری $\alpha = 0.05$ داده‌ها دارای توزیع پارتو هستند. نمودارهای هیستوگرام و برآورد امید شرطی خسارت‌ها رسم و در نمودار ۱ نشان داده شده است.

نمودار ۱. هیستوگرام و برآورد امید شرطی خسارت‌ها برای داده‌های آتش‌سوزی استان بوشهر از سال ۱۳۸۶ تا سال ۱۳۹۴



۵- بحث و نتیجه‌گیری

در این مقاله، نشان داده شد که برآورد میانگین خسارت‌های رخ داده شده با توزیع پارتو، برآوردی مناسب در مدل‌های مخاطره نخواهد بود و مدیریت نادرست مدیران شرکت می‌تواند شرکت را در پرداخت خسارت‌های بیمه‌گذاران با مشکل مواجه کند. در این سبب بیمه با محاسبه امید شرطی متغیر تصادفی اندازه خسارت، دیده شد که استفاده از خسارت‌های دارای توزیع پارتو مناسب در برآورد خسارت‌ها نخواهد بود. با توجه به آزمون برازندگی انجام شده در سطح معنی‌داری $\alpha = 0.05$ چون خسارت‌های رخ داده شده برای آتش‌سوزی واحدهای صنعتی و کارگاهی در شرکت بیمه کارآفرین استان بوشهر، دارای توزیع پارتو بودند.

مراجع

- [۱] بازیاری، ا و پرهام، غ. ع. (۱۳۸۷)، پیدایش فرایندهای مخاطره و بررسی مدل‌های مربوط به آن برای محاسبه احتمالات ورشکستگی، نهمین کنفرانس آمار ایران، مجموعه مقالات منتخب، ۴۶-۶۱، دانشگاه اصفهان.
- [۲] بازیاری، ا، (۱۳۹۱)، احتمال ورشکستگی فرآیند مخاطره انفرادی شرکت بیمه با خسارتهای وابسته، مجله علوم آماری، جلد ۶، شماره ۱، ص ۲۱-۳۷، دانشگاه تربیت مدرس تهران.
- [۳] بازیاری، ا، (۱۳۹۵)، تحلیل احتمال ورشکستگی زمان نامتناهی در مدل مخاطره جمعی، پذیرش برای چاپ در مجله علوم آماری.
- [4] Asmussen, S. (2000), *Ruin Probability*, World Scientific, Singapore.
- [5] Lundberg, F. (1926), *Forsakringsteknisk Riskutjamning*, F. Englund's boktryckeri A. B., Stockholm.
- [6] Rolski, T. (2000), *Some Problems in the Theory of Risk*, Tokyo Institute of Technology and University of Wroclaw.

خلاصه‌سازی و تحلیل منحنی‌های رشد صنایع ایران با استفاده از رویکردی مبتنی بر تجزیه و تحلیل مؤلفه‌های اصلی

فرزاد تیموریان^۱، آیت اله موسوی^۲

^۱ محقق اداره آمارهای اقتصادی، بانک مرکزی ج.ا.ا، تهران،

f.teymourian@cbi.ir

^۲ رئیس دایره آمار صنعتی اداره آمارهای اقتصادی، بانک مرکزی ج.ا.ا، تهران

a.mousavi@cbi.ir

چکیده مختصر

بخش صنعت در کشور ما با توجه به نقشی که می‌تواند در نوسازی اقتصاد و رهایی آن از وابستگی به نفت داشته باشد، یکی از مهمترین مولفه‌های بخش واقعی اقتصاد است. تحلیل روند رشد صنعت از جنبه‌های مختلف، عاملی کلیدی در تحلیل‌های اقتصادی و رکن اساسی تصمیم‌سازی‌های اقتصادی است. نماگرهای اقتصادی کوتاه‌مدت نقش مهمی در این زمینه ایفا می‌کنند و در این بین، شاخص‌های بخش صنعت به ویژه شاخص تولید کارگاه‌های بزرگ صنعتی از اهمیت زیادی برخوردارند.

در این مقاله، طرح بررسی آمار صنعتی که توسط اداره آمار اقتصادی بانک مرکزی اجرا می‌شود؛ معرفی می‌گردد و سپس به خلاصه‌سازی و ارائه تحلیلی از منحنی‌های رشد رشته فعالیت‌های مختلف صنعتی در ایران از سال ۱۳۹۰ به بعد و مشخص کردن مولفه‌های اصلی تغییرات شاخص کل تولید کارگاه‌های بزرگ صنعتی کشور پرداخته می‌شود. برای این منظور پس از بررسی مبانی نظری روش تجزیه و تحلیل مولفه‌های اصلی، با به کارگیری روش پیشنهادی جونز و رایس، ۹۶ رشته فعالیت صنعتی به ۱۵ رشته فعالیت صنعتی خلاصه و تحلیل گردید.

کلمات کلیدی

تحلیل مولفه‌های اصلی، رشد صنعت، شاخص تولید کارگاه‌های بزرگ صنعتی.

۱- مقدمه

در بیشتر مسائل کاربردی مشاهدات بر روی تعداد زیادی متغیرهای همبسته انجام می‌گیرد و چون یک وابستگی بین این متغیرها وجود دارد چنین به نظر می‌رسد که بایستی در این گونه مسائل بُعد مساله را بدون از دست دادن اطلاعاتی که این متغیرها در ماتریس کوواریانس دارند کاهش داد. یکی از روش‌های کاهش بُعد روش تجزیه و تحلیل مولفه‌های اصلی است. تجزیه و تحلیل مولفه‌های اصلی یک روش ریاضی است که احتیاج به موجود بودن یک مدل آماری ندارد و به خصوص هیچ فرضی در مورد توزیع متغیرهای اولیه انجام نمی‌گیرد. از سوی دیگر، نمایش تعداد زیادی از منحنی‌ها بر روی یک نمودار به طوری که هر منحنی بر روی سایر منحنی‌ها قرار گیرد یا در قسمت‌های زیادی با یکدیگر تلاقی داشته باشند، از نظر ظاهری ناخوشایند و از جنبه اطلاع‌دهی ضعیف است. در این مقاله ابتدا با استفاده از روش تجزیه و تحلیل مولفه‌های اصلی، از ۹۶ رشته فعالیت صنعتی ایران، پنج مولفه اصلی استخراج شد. سپس با توجه به اینکه هر مولفه اصلی یک ترکیب خطی از متغیرهای اولیه است که ممکن است مقادیر و دامنه تغییرات آن‌ها با متغیرهای اولیه منطبق

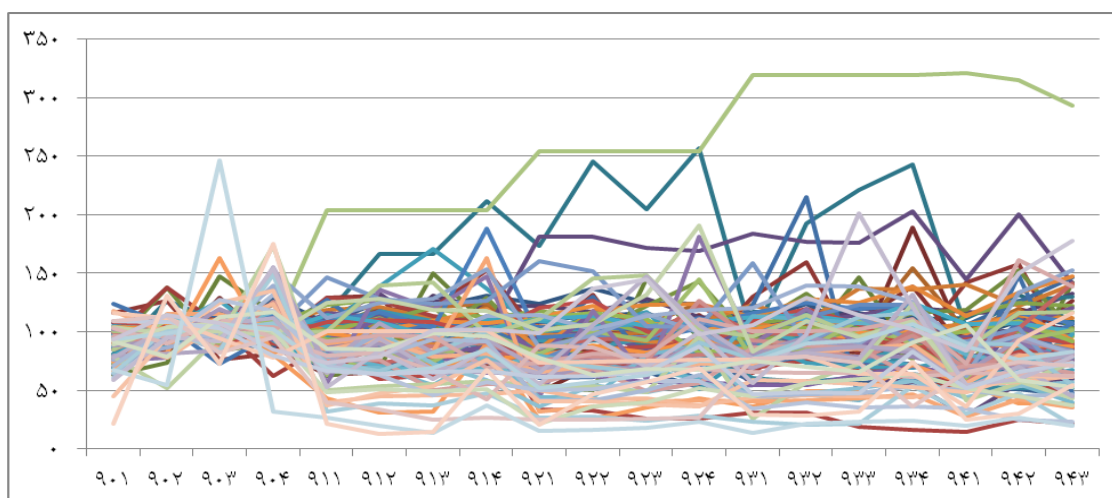
نباشد؛ با به‌کارگیری روش پیشنهادی جونز و رایس، به ازای هر یک از مولفه‌های اصلی تعداد محدودی از متغیرهای اولیه طوری انتخاب می‌شوند که نحوه تغییرات موجود بین همه متغیرها را تا حد زیادی بیان می‌کنند.

۲- معرفی داده‌ها

داده‌های این تحقیق از نتایج طرح بررسی آمار صنعتی که توسط اداره آمار اقتصادی بانک مرکزی جمهوری اسلامی ایران اجرا می‌شود اتخاذ شده است. این طرح که از سابقه‌ای طولانی (بیش از ۵۱ سال) برخوردار است به لحاظ طبقه‌بندی صنایع و ارائه شاخص‌های سه‌ماهه فعالیت کارگاه‌های بزرگ صنعتی (کارگاه‌هایی با تعداد کارکنان ۱۰۰ نفر و بیشتر) در سطح کشور یکتا می‌باشد. در این طرح بر اساس آخرین نسخه طبقه‌بندی استاندارد بین‌المللی فعالیت‌های اقتصادی (ISIC نسخه ۴/۰) تعداد ۹۶ رشته فعالیت صنعتی (سطح چهار رقم ISIC) مورد بررسی قرار می‌گیرد.

۳- طرح مساله

در نمودار شماره (۱)، روند تغییرات شاخص تولید ۹۶ رشته فعالیت صنعتی (کد صنعت) در بازه سه‌ماهه نخست سال ۱۳۹۰ تا سه‌ماهه سوم سال ۱۳۹۴ رسم شده است. با توجه به این که رفتار هر صنعت در انبوهه حاصل از سیاهی ایجادشده توسط سایر منحنی‌ها گم شده است، لذا این نمودار از نظر ظاهری ناخوشایند و از جنبه اطلاع‌دهی ضعیف و موجب سردرگمی است.



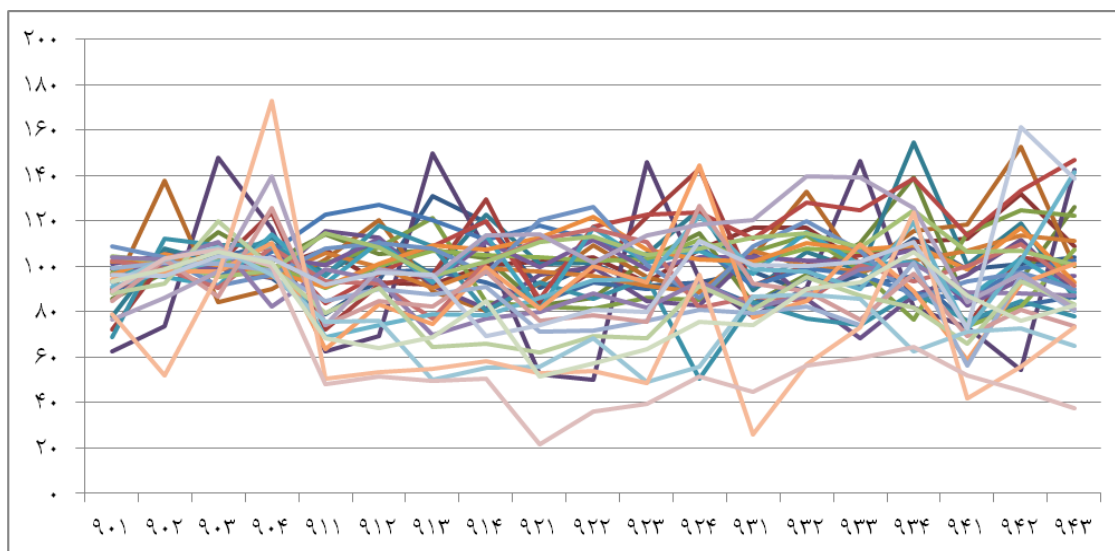
نمودار شماره ۱- منحنی‌های رشد شاخص تولید رشته فعالیت‌های صنعتی (۱۳۹۰-۹۴)

ام. سی. جونز و جی. ا. رایس (۱۹۹۲) با بهره‌گیری از روش تجزیه و تحلیل مولفه‌های اصلی، راه ساده‌ای را برای نمایش بخش عمده اطلاعات پنهان‌شده در نمودارهایی شبیه نمودار شماره (۱) ارائه کرده‌اند. در این روش فقط زیرمجموعه‌ای از منحنی‌های اولیه نمایش داده می‌شوند به طوری که این زیرمجموعه به گونه‌ای انتخاب می‌شود که جنبه‌های مهم تغییرات مجموعه اولیه را به بهترین وجه نمایش دهد.

۴- خلاصه‌سازی با رویکردی مبتنی بر تجزیه و تحلیل مولفه‌های اصلی (روش جونز و رایس)

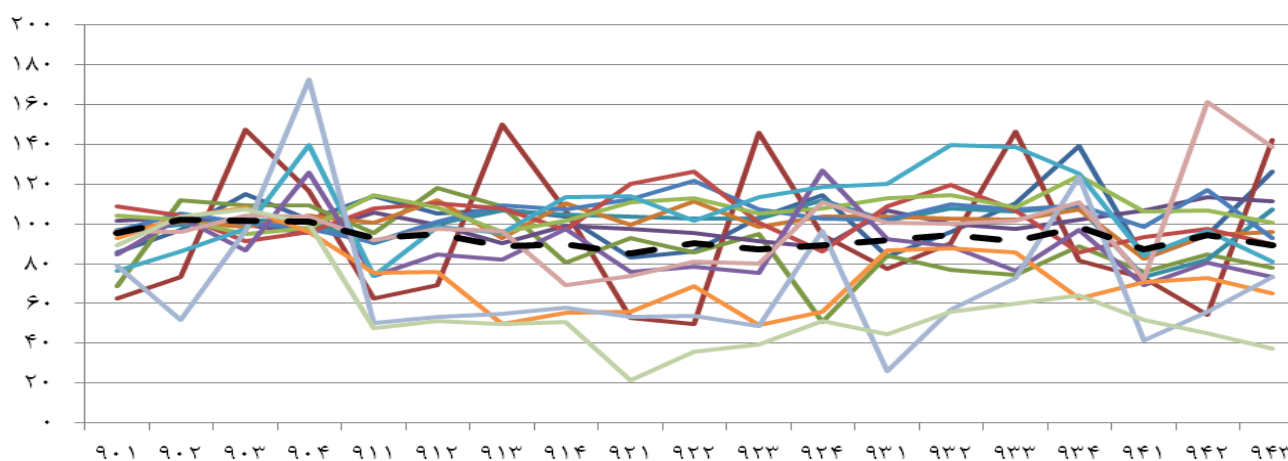
از بین ۹۶ منحنی نمودار شماره (۱)، تعداد محدودی از آن‌ها را می‌بایست طوری انتخاب کنیم که جنبه‌های مهم تغییرات موجود در منحنی‌ها را به بهترین وجه منعکس کند. از بین رشته فعالیت‌های صنعتی مورد بررسی، برخی از صنایع با وجود اثرگذاری بسیار کم در شاخص کل تولید، پراکندگی زیادی دارند که اجرای روش تجزیه و تحلیل مولفه‌های اصلی بر روی این داده‌ها، موجب انحراف نتایج این روش خواهد شد. لذا در مرحله اول، صنایع پرنوسان و در عین حال کم‌اثر (صنایعی که مجموع سهم ارزش افزوده آن‌ها در صنعت کشور کمتر از ۱۰ درصد است) را از تحلیل کنار می‌گذاریم و روش تجزیه و تحلیل مولفه‌های اصلی را روی مجموعه باقی‌مانده که

متشکل از ۳۳ رشته فعالیت (کد صنعت) است اجرا می‌کنیم. نمودار شماره (۲)، تغییرات شاخص تولید در سطح این کدهای صنعت را نشان می‌دهد که مجموعاً بیش از ۹۰ درصد از ارزش افزوده کارگاه‌های بزرگ صنعتی را در اختیار دارند. نمودار شماره (۲) هسته مرکزی منحنی‌های اولیه را نشان می‌دهد. در جدول شماره (۱)، فهرست این رشته فعالیت‌ها آمده است.



نمودار شماره ۲- منحنی‌های رشد شاخص تولید رشته فعالیت‌های صنعتی پس از حذف صنایع کم‌اثر

پس از اجرای روش تجزیه و تحلیل مولفه‌های اصلی و به‌کارگیری روش جونز و رایس، در نهایت ۱۵ رشته فعالیت صنعتی انتخاب می‌شوند که روند تغییرات آن‌ها به همراه روند شاخص کل تولید در نمودار شماره (۳) رسم شده است.



نمودار شماره ۳- منحنی شاخص کل تولید و متغیرهای حاصل از روش جونز و رایس

منحنی خط‌چین روند شاخص کل تولید را نشان می‌دهد. همان‌طور که نمودار فوق نشان می‌دهد نوسانات شاخص کل تولید به خوبی توسط متغیرهای انتخابی پوشش داده شده و رفتار و تغییرات آن توسط متغیرهای انتخابی قابل توضیح و تفسیر است.

۵- نتیجه‌گیری

بررسی روند تغییر تولیدات صنعتی، تجزیه و تحلیل رفتار صنایع مهم و اثرگذار کشور و داشتن نگرشی کلان و در عین حال فارغ از پیچیدگی‌ها و اطلاعات تکراری، مورد نظر تصمیم‌سازان و سیاست‌گذاران حوزه اقتصاد است. در این مقاله رفتار شاخص کل تولید کارگاه‌های بزرگ صنعتی که نوسانات کوتاه‌مدت (سه‌ماهه) تولید صنعتی در کشور را نشان می‌دهد از این منظر مورد مطالعه قرار گرفت که با داشتن کمترین تعداد متغیرهای تاثیرگذار، بیشترین اطلاعات ممکن از وضع صنعت کشور حاصل گردد. بدین منظور تعداد ۹۶ کد صنعت در سطح چهارم رقم دستور العمل ISIC نسخه چهارم، از سه‌ماهه نخست سال ۱۳۹۰ به بعد مورد بررسی قرار گرفت. در این مطالعه از رویکرد آماری پیشنهادی جونز و رایس که مبتنی بر روش تجزیه و تحلیل مولفه‌های اصلی می‌باشد برای کاهش تعداد ۹۶ متغیر مورد بررسی استفاده شده است.

دیدگاه‌ها در مورد روش‌های بیزی و داده‌های بزرگ

الهام خرازی کلجاهی^۱ و حسین جباری خامنه‌ای^۲

^۱ دانشجوی کارشناسی ارشد آمار ریاضی، دانشگاه پیام نور تبریز، تبریز
Mahsaf557m@yahoo.com

^۲ گروه آمار، دانشکده علوم ریاضی، دانشگاه تبریز، تبریز
h_jabbari@tabrizu.ac.ir

چکیده مختصر

محققان در حال حاضر با دنیایی رو به رو هستند که حجم داده‌ها و ابزارهای تحلیلی مانند الگوریتم‌های استنباط بیزی را بهبود می‌بخشند، تا بتوانند با پیشرفت تکنولوژی هماهنگ شوند. روش‌های بیزی برای رشته تحلیل بازاریابی منافع قابل توجهی را به ارمغان آورده‌اند، اما چالش‌های محاسباتی برای داده‌هایی با مقیاس بزرگ همچنان وجود دارد. در این مقاله چندین استراتژی با نمونه‌های خاص با استفاده از رگرسیون درختی افزایشی و انتخاب متغیر، مورد بحث قرار گرفته است.

کلمات کلیدی

آمار بیزی، داده‌های بزرگ، محاسبه مقیاس بیز

۱- مقدمه

پیشرفت در فناوری، با توجه به محاسبات و ارتباطات، منجر به افزایش داده‌های روز افزون شده است که می‌توانند خلاصه‌ای از زندگی مدرن را ارائه دهند. این خلاصه‌ها می‌توانند بینشی در مورد سیستم کلی و جزئیات مربوط به افراد که تقریباً موازی با پیشرفت روش‌های استنباط بیزی است را ارائه دهند.

در این مقاله برخی از نقاط قوت چارچوب مدل سازی بیزی که آن را به خوبی برای بازاریابی تجزیه و تحلیل می‌کند، بحث می‌کنیم. در زمینه ارائه استراتژی که روش‌های بیزی را به صورت محاسباتی کارآمد می‌داند دو نمونه را ارائه می‌دهیم، یکی با استفاده از رگرسیون درختی افزایشی بیزی (*BART*) و دیگری با استفاده از یک جایگزین برای انتخاب متغیر در داده‌های بزرگ و رویکرد دیگر روش‌های بیزی برای مشکلات داده‌های بزرگ اعمال چند جمله‌ای‌ها است. روش‌هایی که به محقق اجازه می‌دهد تا بتوانند با عملکردهای تقریبی احتمال بر محاسبات فرم بسته بپردازد. روش تقریبی برای شمارش داده‌ها به صورت توزیع دو جمله‌ای منفی استفاده می‌شود و همچنین مدل دو جمله‌ای بتا و سپس به طور کلی به مدل‌های انتخابی گسترش می‌یابد که این مقاله مربوط به تحقیقات اخیر در مورد روش‌های مختلف بیزی است که روش‌های تقریبی کلی تری را ارائه می‌دهند. در حالی که علم داده‌های بزرگ باعث ساخت سیستم‌های *CRM* بسیاری از شرکت‌های بزرگ شده است. بسیاری از داده‌ها ارزش اقتصادی مستقیم را به شرکت ارائه نمی‌دهند. بنابراین انتخاب روش‌هایی بیزی برای داده‌های بزرگ می‌تواند مشکل محاسبات پیچیده را شکست دهند.

۲- نوآوری *BART*، چالش آمار بیزی برای داده‌های بزرگ

استنباط بیزی مدل اساسی زیر را فراهم می‌کند:

$$y = f(x) + \epsilon, \quad \epsilon \sim N(0, \sigma^2)$$

در جایی که f شناخته شده نیست، *BART* یک رویکرد مبتنی بر مدل بیزی است.

از مزایای کلیدی روش *BART* عبارتند از:

- (۱) نتایج منطقی اغلب با استفاده از اطلاعات پیشین به دست می‌آید.
- (۲) عمق مدل‌هایی که برای ساخت تابع f استفاده می‌شود، نسبتاً به صورت اتوماتیک تعریف شده به طوری که سطح تعامل بین متغیرها می‌تواند به طور طبیعی کشف شود.
- (۳) هیچ نیازی به تعریف مجموعه‌ای از تغییر در فضای X وجود ندارد.
- (۴) سطح عدم قطعیت از طریق پسین بیزی به دست می‌آید.

انواع برنامه‌های *BART* (از جمله X با ابعاد بزرگ) را با استفاده از اجرای *BART* در پکیج R با مدل‌های بیزی امکان پذیر است. الگوریتم *BART* یک رویکرد *MCMC* است، که با استفاده از محاسبات موازی پیاده‌سازی می‌شود که این رویکرد نیازمند ارسال پیام مکرر بین ماشین‌ها است و با استفاده از رابط پیام‌گذار اجرا می‌شود. داده‌ها را در میان دستگاه‌ها تقسیم می‌کند و *MCMC* را جداگانه از هر زیرمجموعه داده اجرا و سپس نتایج را با استفاده از میانگین وزنی ترکیب می‌کند که این عملیات با استفاده از بسته موازی در R اجرا می‌شود.

۳- تأثیر بیز و داده‌های بزرگ در تحقیقات بازاریابی

انقلاب بزرگ داده‌ها موجب جمع‌آوری و ذخیره‌سازی داده‌های بی‌سابقه‌ای از سوی شرکت‌ها شده است. طوری که صنعت بازاریابی متقاعد به این که طلا در داده‌های بزرگ وجود دارد شده است. داده‌های بزرگ به وسیله ضبط تراکنش‌های بازار از افراد ایجاد می‌شود و امید به درک کامل‌تر از الگوهای مصرف را فراهم می‌کند و اطلاعات مفیدی برای شناسایی مشتریان بالقوه در دست‌های مکمل محصول است. داده‌های بزرگ در مورد پاسخ فردی مشتریان به پیشنهادات بازار در بسیاری دسته‌های مجهول به ما می‌گوید که در بسیاری اقدامات مانند ایجاد مقدماتی برای اندازه‌گیری اثر در قیمت‌گذاری آنلاین و آزمایش‌های تبلیغاتی مفید است. داده‌های بزرگ گسسته هستند و انتخاب‌ها به عنوان رویدادهای بیزی ثبت می‌شوند و هنگامی که مقادیر خرید ثبت می‌شود مقادیر آن بر روی شبکه تعریف شده توسط اندازه‌بسته موجود است.

یکی از مهم‌ترین جنبه‌های داده‌های بزرگ شایع بودن تعداد صفر در داده‌هاست. نسبت بالای صفر نشان می‌دهد که که مصرف‌کنندگان در رفتار بازار خود با هدف هدایت شده و منبع محرمانه هستند. مصرف‌کنندگان اغلب وبسایت‌های مرتبط را نمی‌بینند در مورد بسیاری از مارک‌های رقیب یاد نمی‌گیرند و عمق گزینه‌ها را درک نمی‌کنند، زیرا که توجه آن‌ها محدود به پول و زمان است. در سطح بنیادی این صفرها به لحاظ کیفی از مقادیر مثبت متفاوت است و اطلاعات محدودی در مورد آنچه که در زندگی پاسخ‌دهندگان اتفاق می‌افتد را ارائه می‌دهد. این رویکردها برای تصمیم‌گیری‌های تاکتیکی مفید هستند، بازاریاب‌ها باید به این توجه داشته باشند که داده‌های بزرگ برای تنظیم قیمت‌ها و جستجو برای فرصت‌های متقابل فروش استفاده می‌شود.

کالیبراسیون توابع کاربردی (مثلاً در مدل‌های انتخابی گسسته) یک روش تحقیق پر بار در مدل‌های داده‌های بزرگ است. رویکرد دیگر استفاده از مدل اثرات تصادفی برای کاستن کمبود اطلاعات به علت کم بودن داده‌های بزرگ است. هر دو روش به اجتناب از نفوذ ناپایدار در داده‌ها و سایر آلاینده‌ها کمک می‌کند که منجر به نتایج بدیعی در تجزیه و تحلیل‌های بعدی می‌شود. در بازاریابی مدل‌های ساده‌تر به پیش‌بینی قوی‌تر کمک می‌کند، استفاده از تئوری برای توضیح تنوع در داده‌های بزرگ به دلیل اندازه آن ارزشمند است. استفاده از تئوری و ابزار استنباط آماری پایه بیزی تحلیل و بررسی بازاریابی در قیمت و مکان از نیازهایی است که روش‌های بیزی به دلیل توانایی آن‌ها برای مقابله با مجموعه داده‌های بزرگ و کم عمق و توانایی آن برای تولید دقیق و نتیجه‌گیری نمونه‌های محدود در این موارد بسیار مفید هستند.

با نگاه به آینده، اینترنت باعث شده است که افراد در تحقیقات در مورد مسائل انتخابی و موضوعاتی با هزینه‌های بسیار کم مشارکت داشته باشند. برای مثال نظر سنجی‌های مصرف‌کننده گوگل به محققان اجازه می‌دهد که تعداد کمی از سوالات را برای افراد در وب به ازای پاسخ دهندگان با مبلغ کم ارائه دهند. این نوع تجزیه و تحلیل شبیه حملات هوایی است، به نحوی که یک نظر سنجی سنتی مصرف‌کننده که شامل ده‌ها سوال است و تکمیل آن تا ۳۰ دقیقه طول می‌کشد باز می‌شود و توانایی پاسخ سریع به سوالات خاص به شرکت‌ها اجازه می‌دهد تا تمرکز تحقیقات بازاریابی را در سازمان‌ها غیر متمرکز کرده و تقاضا برای تجزیه و تحلیل داده‌ها را افزایش دهند.

تجزیه و تحلیل بیزی به ویژه برای یکپارچه سازی اطلاعات از طریق استفاده از پارامترهای فوق العاده ای که تغییرات سایر پارامترها را توضیح می‌دهد مناسب است. که نیازها و خواسته‌ها می‌تواند در سطح بالایی از یک مدل سلسله مراتبی با استفاده از نیازها به عنوان متغیرهایی در مدل اثرات تصادفی یکپارچه شوند. این انتقال متغیرها به یک متغیر مشترک در حال حاضر در سطح ابتدائی انجام می‌شود و مزایای بالقوه را با توجه به دسترسی به داده‌های بزرگ از بازار و به صورت تقویت شده با داده‌های نظر سنجی‌های آنلاین و سنتی ارائه می‌دهد که به عوامل تعیین‌کننده تقاضا کمک می‌کند.

۴- نتیجه گیری

داده‌ها در آینده بدون شک پیچیده تر از حال حاضر خواهد بود. فناوری به ما اجازه دسترسی روز افزون به داده‌ها را می‌دهد. زیرا زندگی در جزئیات دقیق تری ثبت می‌شود همانطور که منابع محاسباتی را افزایش می‌دهد تا بتوانیم جهان را اندازه‌گیری کنیم. ابزار تحلیلی آینده توسط تیم‌های تخصصی ایجاد می‌شود که ابزار خاص برای کار در دست دارند. ابزارهای اتوماسیون یا ابزارهای تاکتیکی که به شیوه‌های پهنه به سیستم پاسخ می‌دهند. فرضی که پاسخ سیستم را تغییر نمی‌دهد ساده ترین ابزار است. در تمام این تلاش‌ها روش‌های بیزی یک جعبه ابزار قدرتمند ارائه می‌دهند. روش بیزی اجازه می‌دهد که متخصصان و آزمایش‌هایی که استفاده می‌شود به طور رسمی در یک مدل قرار گرفته و ابزار تصمیم‌گیری عوامل فردی را بر عهده گیرند. استنباط بیزی برای جستجوهای تصادفی در فضای مدل‌های احتمالی امکان‌پذیر است که احتمال پیشین مثبت هستند، عناوین گفته شده در تحقیق مبتنی بر اقتصاد (بازاریابی) به افزایش همکاری بین دانشگاه و صنعت اشاره دارند که می‌تواند به محقق کمک کند تا مشکلات اساسی را حل و بهره‌وری بیشتری را برای اقتصاد ما به ارمغان آورد.

مراجع

- [1] Allenby, G. M., & Rossi, P. E. (1998). Marketing models of consumer heterogeneity. *Journal of econometrics*, 89(1), 57-78.
- [2] Chipman, H. A., George, E. I., & McCulloch, R. E. (2010). BART: Bayesian additive regression trees. *The Annals of Applied Statistics*, 4(1), 266-298.
- [3] Everson, P. J., & Bradlow, E. T. (2002). Bayesian inference for the beta-binomial distribution via polynomial expansions. *Journal of Computational and Graphical Statistics*, 11(1), 202-207.
- [4] Rossi, P. E., Allenby, G. M., & McCulloch, R. (2012). *Bayesian statistics and marketing*. John Wiley & Sons.
- [5] Sutton, A. J., & Abrams, K. R. (2001). Bayesian methods in meta-analysis and evidence synthesis. *Statistical methods in medical research*, 10(4), 277-303.

روش جدید بازسازی داده‌های گم‌شده در رگرسیون

اسحاق الماسی^۱، عادل محمدپور^۲

^۱ گروه ریاضی، دانشکده علوم پایه، دانشگاه ایلام،

i.almasi@aut.ac.ir

^۲ گروه آمار، دانشکده ریاضی و علوم کامپیوتر، دانشگاه امیرکبیر، تهران

adel@aut.ac.ir

چکیده مختصر

در تحلیل آماری معمولاً اطلاعات برخی واحدها یا متغیرها ثبت نشده یا به عبارتی با داده گم‌شده مواجه‌ایم. در این مقاله روش جدیدی را بر اساس قضیه تصویر برای بازسازی یا جانمایی این داده‌ها ارائه می‌دهیم. با یک مطالعه شبیه سازی نشان می‌دهیم که در چه شرایطی روش جدید از برخی روش‌های معمول دیگر عملکرد بهتری دارد. در پایان با استفاده از داده واقعی عملکرد این روش را ارزیابی می‌کنیم.

کلمات کلیدی

قضیه تصویر، آمارهای مرتب، بازسازی، طرح سانسور.

۱- مقدمه

به طور معمول در بیشتر اوقات بنا به دلایلی، اطلاعات برخی از متغیرهای مورد بررسی ثبت نشده یا از دست می‌رود. در چنین شرایطی با مسأله گم‌شده‌گی داده مواجه هستیم. از آنجا که همه روش‌های استاندارد آماری بر پایه مشاهده همه متغیرها بنا شده است، بنابراین وجود داده گم‌شده این روش‌ها را دچار مشکل می‌کند. پیش فرض همه نرم افزارهای آماری بر این اساس است که واحد آماری حاوی داده گم‌شده را از لیست داده‌ها خارج کنیم و بر اساس داده‌های موجود و با فرض داده کامل تحلیل را انجام دهیم. اما این کار در بیشتر اوقات باعث از دست رفتن حجم زیادی از اطلاعات می‌شود. راه حل دیگر جانمایی داده‌های گم شده با مقادیر مناسب مثل میانگین داده‌های مشاهده شده و ... می‌باشد. در این مقاله روش جدیدی برای بازسازی داده گم‌شده بر اساس قضیه تصویر و آمارهای ترتیبی را ارائه می‌دهیم.

۲- روش‌های برخورد با داده‌های گم‌شده

در طی سال‌های اخیر روش‌های متعددی برای جانمایی مقادیر گم‌شده پیشنهاد شده است. یکی از این روش‌ها جانمایی با میانگین کل (MI) هست. در این روش میانگین کل مشاهدات هر متغیر جایگزین مقادیر گم‌شده می‌شود. اگرچه این روش ساده‌ترین نوع جانمایی است اما شاید بتوان گفت غیرجذاب‌ترین نوع جانمایی نیز باشد. جانمایی تمام مقادیر براساس میانگین موجب ایجاد اختلال در توزیع داده‌ها و کم برآورد شدن واریانس می‌شود. اگرچه جانمایی با میانگین روش نسبتاً آسانی است ولی دو اشکال مهم در آن دیده می‌شود، یکی این که شکل توزیع متغیر مورد نظر را دگرگون می‌کند. دوم اینکه چون به جای تعدادی از اعضای نمونه (گم‌شده‌ها) یک مقدار ثابت مانند میانگین جایگزین می‌شود، تغییرات در بین اعضای نمونه کاهش پیدا می‌کند و این سبب می‌شود واریانس حاصل کمتر از واریانس واقعی نشان داده شود. در اینجا منظور از واریانس واقعی، واریانس همه اعضای نمونه به شرط عدم وجود داده گم‌شده می‌باشد. بنابراین جانمایی با مقدار میانگین برای تحلیل‌های ساده‌ای کاربرد دارد که برآوردهای نقطه‌ای چون میانگین یا مجموع بدون توجه به

واریانس مورد نظر باشد. روش دیگر جانهای رگرسیونی (CMI) می‌باشد. در این روش مقادیر گم‌شده بر اساس پیش‌بینی از یک مدل رگرسیونی برای تمامی موارد جایگزین می‌شوند.

۳- جانهای در آماره‌های مرتب

اگر (y_1, \dots, y_n) یک نمونه تصادفی از توزیع دلخواه با تابع توزیع تجمعی F باشد و $(y_{(1)}, \dots, y_{(n)})$ آماره‌های ترتیبی مربوط به این نمونه باشد. فرض می‌کنیم که برخی از آماره‌های ترتیبی سانسور یا گم‌شده باشند و مشاهدات به صورت $(y_{(1)}, \dots, y_{(r)}, y_{(s)}, \dots, y_{(n)})$ باشند که در اینجا $0 \leq r < s \leq n+1$ می‌باشد. در حالت $s = n+1$ سانسور راست و $r = 0$ سانسور چپ خواهیم داشت. الماسی و همکاران (۲۰۱۶) بر اساس قضیه تصویر آماره‌های ترتیبی گم‌شده را به صورت زیر بازسازی کردند

$$\hat{y}_{(l)} = \begin{cases} F^{-1}\left(\frac{l}{s}F(y_{(s)})\right) & r = 0, l = 1, \dots, s-1, \\ F^{-1}\left(\frac{s-l}{s-r}F(y_{(r)}) + \frac{l-r}{s-r}F(y_{(s)})\right) & l = r+1, \dots, s-1, \\ F^{-1}\left(\frac{l-r}{n+1-r} + \frac{n+1-l}{n+1-r}F(y_{(s)})\right) & s = n+1, l = r+1, \dots, n. \end{cases}$$

۴- روش پیشنهادی

مدل رگرسیونی ساده خطی زیر را در نظر بگیرید

$$y_j = \beta x_j + \varepsilon_j, j = 1, 2, \dots, n.$$

فرض می‌کنیم مقادیر $\{x_j\}$ به صورت کامل مشاهده شده باشند اما متغیر پاسخ $\{y_j\}$ را تنها برای m مقدار مشاهده و $n-m$ مقدار دیگر گم‌شده می‌باشند. فرض می‌کنیم مکانیسم این بی‌پاسخی به صورت "گم‌شده‌گی تصادفی" (MAR^۱) باشد، برای دیدن جزئیات بیشتر روین (۱۹۷۶) را ببینید. الگوریتم پیشنهادی برای جانهای مقادیر گم‌شده به صورت زیر می‌باشد.

گام اول: برای اساس مقادیر مشاهده شده، ضریب همبستگی نمونه‌ای را به دست می‌آوریم.

گام دوم: مقادیر $\{x_j\}$ را به صورت غیرنزولی مرتب و رتبه آنها یعنی $r_{x_j} = \text{rank}(x_j)$ را به دست می‌آوریم.

گام سوم: رتبه مقادیر مشاهده شده $y_j, j = 1, \dots, m$ را به صورت زیر به تعیین می‌کنیم

$$r_{y_j} = \begin{cases} r_{x_j} & \rho > 0 \\ n+1 - r_{x_j} & \rho < 0 \end{cases}$$

گام چهارم: مقادیر گم‌شده $y_j, j = 1, \dots, m$ را به صورت زیر بازسازی می‌کنیم

$$\hat{y}_{(j)} = \begin{cases} F^{-1}\left(\frac{1}{2}F(y_{(1)})\right) & r_{y_j} = 1 \\ F^{-1}\left(\frac{1}{2}F(y_{(r_{y_j}-1)}) + \frac{1}{2}F(y_{(r_{y_j})})\right) & r_{y_j} = 2, \dots, m \\ F^{-1}\left(\frac{r_{y_j} - m}{n - m + 1} + \frac{n - r_{y_j} + 1}{n - m + 1}F(y_{(m)})\right) & r_{y_j} = m + 1, \dots, n. \end{cases}$$

نکته: در مدل رگرسیونی چندگانه با k متغیر مستقل، الگوریتم بالا را به ازای هر متغیر مستقل انجام و در نهایت میانگین k مقدار بازساز را در نظر می‌گیریم.

۵- مطالعه عددی

در این قسمت به صورت شبیه‌سازی، روش پیشنهادی را با روش‌های میانگین و پیش‌بینی رگرسیونی مقایسه می‌کنیم. برای این کار ابتدا یک نمونه $n = 20$ تایی از توزیع نرمال دو متغیره با ضریب همبستگی ρ را تولید می‌کنیم. بدون از دست دادن کلیت مسئله میانگین و واریانس نرمال دو متغیره را به ترتیب ۰ و ۱ در نظر می‌گیریم. این شبیه‌سازی را ۱۰۰۰۰ بار برای مقادیر مختلف $\rho \in \{0.3, \dots, 0.9\}$ تکرار کرده‌ایم. ابتدا برآورد میانگین، انحراف معیار و ضریب همبستگی داده‌های کامل را به دست می‌آوریم که نتایج در جدول ۱ با NM مشخص کرده‌ایم. سپس $m = 10$ مشاهده را به تصادف از متغیر پاسخ $\{y_i\}$ انتخاب و حذف می‌کنیم. با استفاده از روش‌های CMI، MI و روش پیشنهادی PI مقادیر گم‌شده را بازسازی و مقادیر برآورد پارامترها و میانگین مربعات خطا (MSE) را به دست می‌آوریم. با توجه به نتایج این شبیه‌سازی و به روش مشابه برای مقادیر منفی ضریب همبستگی، نتیجه می‌گیریم، روش پیشنهادی برای مقادیر $0.4 < |\rho| < 0.7$ عملکردی بهتری از نظر میانگین مربعات خطا دارد (در جدول ۱، روش بهینه با رنگ آبی مشخص شده است). همچنین برای $|\rho| \leq 0.7$ روش CMI عملکرد بهتری دارد. برای $|\rho| \leq 0.3$ روش‌های MI عملکرد بهتری نسبت به دیگر روش‌ها دارد.

جدول ۱: مقایسه MSE و برآورد پارامترهای مدل رگرسیونی با خطای نرمال استاندارد برای برخی روش‌های مختلف با $n = 20$ و $m = 10$.

		ρ						
روش		0.3	0.4	0.5	0.6	0.7	0.8	0.9
MSE	NM	0.9051	0.8722	0.8223	0.7583	0.6778	0.5696	0.4130
	MI	1.0474	1.0541	1.0511	1.0461	1.0481	1.0493	1.0493
	CMI	1.0693	1.0317	0.9717	0.8952	0.8008	0.6748	0.4879
	PI	1.1337	1.0495	0.9589	0.8751	0.8092	0.7713	0.7832
$\hat{\rho}$	NM	0.2915	0.3894	0.4903	0.5875	0.6887	0.7915	0.8950
	MI	0.1994	0.2627	0.3308	0.3977	0.4664	0.5369	0.6064
	CMI	0.3216	0.4195	0.5355	0.6403	0.7385	0.8344	0.9251
	PI	0.5118	0.5431	0.6062	0.6548	0.6890	0.7880	0.8013
$\hat{\mu}$	NM	-0.0027	0.0016	0.0022	0.0034	-0.0017	0.0042	0.0036
	MI	0.0043	0.0029	0.0033	0.0006	0.0020	-0.0041	0.0049
	CMI	-0.0032	-0.0026	0.0032	0.0008	0.0007	-0.0071	0.0020
	PI	-0.0028	-0.0020	0.0032	0.0016	0.0015	-0.0060	0.0032
$\hat{\sigma}$	NM	0.9869	0.9868	0.9862	0.9856	0.9856	0.9875	0.9854
	MI	0.6714	0.2627	0.6671	0.6689	0.6678	0.6693	0.6674
	CMI	0.8503	0.8695	0.8898	0.8804	0.9250	0.9372	0.9524
	PI	0.9376	0.9331	0.9054	0.8624	0.8316	0.7821	0.7390

۶- داده واقعی

اسنیدکور و کوکران، درصد کرم‌خورده‌گی (y) و تعداد میوه (x) روی ۱۲ درخت را اندازه‌گیری کردند (جدول ۱، ۶، ۹، ۱ اسنیدکور و کوکران ۱۹۶۷). میزان همبستگی بین این دو متغیر -0.88 می‌باشد. به عبارتی رابطه معکوسی بین درصد کرم‌خورده‌گی و تعداد میوه‌ها وجود دارد. در این قسمت با استفاده از روش CMI و PI میزان کرم‌خورده‌گی را بازسازی کرده‌ایم و نتایج به دست آمده در جدول ۲ آمده است. اعدادی که با رنگ آبی مشخص شده است، برآورد بهتری را نسبت به مقدار واقعی ارائه داده است. همانطور که در بخش شبیه‌سازی اشاره کردیم، روش PI در حالت $0.4 < |\rho| < 0.7$ عملکرد بهتری دارد اما در اینجا با توجه به اینکه مقدار ضریب همبستگی برابر -0.88 می‌باشد، همچنان عملکرد قابل قبولی دارد.

جدول ۲: داده‌های مربوط به تعداد میوه و درصد کرم‌خوردگی ۱۲ اصله درخت سیب

شماره درخت	$y: 100 \times$ درصد میوه‌های کرم‌خورده	x : تعداد میوه (۱۰۰ میوه)	CMI	PI
۱	59.0	8.0	56.14	56.70
۲	58.0	6.0	58.16	51.07
۳	56.0	11.0	53.10	53.83
۴	53.0	22.0	41.96	44.12
5	50.0	14.0	50.06	51.56
6	45.0	17.0	47.02	49.53
7	43.0	18.0	46.01	47.67
8	42.0	24.0	39.93	36.43
9	39.0	19.0	45.00	45.88
10	38.0	23.0	40.94	39.21
11	30.0	26.0	37.91	32.87
12	27.0	40.0	23.73	26.74

مراجع

- [1] Almasi, I., Mohammadpour, A., Mohammadi, M. (2016). Best Linear Unbiased Interpolation of Order Statistics, Communications in Statistics-Simulation and Computation. DOI:10.1080/03610918.2015.1109656.
- [2] Rubin, D. B. (1976). Inference and missing data. Biometrika, 63, 581-592.
- [3] Snedecor, G., Cochran, W. G. (1967). Statistical Methods, 6th ed. Ames: Iowa State University Press.

روش‌های توصیف و تحلیل مه‌داده‌ها در آمارهای ثبتي

۱ساجده مرادنيا، آعین اله پاشا

۱دانشکده ریاضی و علوم کامپیوتر، دانشگاه خوارزمی، تهران،

moradniasajedeh1371@gmail.com

paasha@khu.ac.ir

چکیده مختصر

مه داده‌ها دارای سه مولفه‌ی اصلی تنوع، حجم و سرعت تغییر هستند و برای ذخیره سازی، تحلیل و نمایش نتایج، مشکلاتی پیش روی آن‌هاست. تنوع بالای مه داده‌ها سبب پررنگ‌تر شدن نقش علم آمار در تحلیل آن‌ها شده است. اما دو مانع در این مسیر وجود دارد:

- ۱- این داده‌ها به دلیل حجم و پیچیدگی زیادی که دارند، از ظرفیت ابزار تحلیلی خارج هستند. ۲- انجام محاسبات بر روی آن‌ها بسیار زمان‌بر و طولانی است.

این‌ها از جمله چالش‌هایی هستند که آماردانان در این حوزه با آن‌ها مواجه‌اند. در این مقاله سعی شده است که برخی روش‌های آماری و محاسباتی اخیر برای مقابله با مشکلات پیش روی مه داده‌ها معرفی گردد.

کلمات کلیدی

مه داده، بوت استرپ، زیرنمونه‌گیری، مونت کارلوی زنجیر مارکوفی، میانگین لگاریتم درست‌نمایی

۱- مقدمه

لین و خی (۲۰۱۱) برآورد خطی از معادلات را به صورت بلوک‌های جداگانه مطرح کردند و روش افراز و تقسیم را در مه داده‌ها به کار گرفتند. لیانگ و همکاران (۲۰۱۳) روش تقریب تصادفی براساس بازنمونه‌گیری را در این حوزه به کار بردند. لیانگ و کیم (۲۰۱۳) نیز به بیان روش میانگین لگاریتم درست‌نمایی با استفاده از الگوریتم متروپولیس-هستینگس در مونت کارلوی زنجیر مارکوفی پرداختند. همچنین کلایتر و همکارانش (۲۰۱۴) روش نمونه‌های کوچک خودگردان را برای ارزیابی کیفیت در مه داده‌ها پیشنهاد دادند. حال با توجه به اهمیت و نقش علم آمار در این حوزه، به طور مختصر به تشریح هر یک از این روش‌های آماری می‌پردازیم. در بخش دوم، روش معادلات برآورد انباشته شده، در بخش سوم، روش‌های میانگین لگاریتم درست‌نمایی، در بخش چهارم، زیرنمونه‌گیری براساس مونت کارلوی زنجیر مارکوفی و در بخش پنجم، روش نمونه‌های کوچک خودگردان را شرح خواهیم داد. در بخش ششم نیز به معرفی برخی بسته‌های نرم‌افزار R می‌پردازیم و سپس در بخش هفتم با ارائه‌ی نتیجه‌گیری، مقاله را به اتمام می‌رسانیم.

۲- معادلات برآورد انباشته

در این روش با توجه به پژوهش‌های صورت گرفته بر روی پردازش تحلیلی آن‌لاین، به دنبال ارائه‌ی یک روش محاسباتی مناسب و دارای قدرت ذخیره سازی برای برآورد کردن برآورد معادله در مه داده‌ها با استفاده از روش افراز و تقسیم هستیم به طوری که در هر قسمت از مجموعه داده، داده‌ها را به بعدهای کم آماری فشرده می‌کنیم و سپس تقریب برآوردگر را به وسیله‌ی حل معادله‌ی تجمعی در تمامی بخش‌ها محاسبه می‌کنیم. این روش سبب کاهش زمان موردنیاز برای انجام محاسبات شده و بر مشکل محدودیت حافظه که برای مه داده‌ها مطرح است، غلبه می‌کند. در مدل رگرسیون خطی، برآوردگر حداقل مربعات برای ضرایب رگرسیونی β برای کل داده‌ها

می‌تواند به صورت یک میانگین وزنی از برابردگر حداقل مربعات در هر بلوک بیان شود. موفقیت در این روش به خطی بودن معادلات برابرد در β بستگی دارد و معادله‌ی برابرد برای تمام داده‌ها، یک جمع ساده از آن برای تمامی بلوک‌هاست. برای معادلات برابرد غیرخطی، روش تقریب زدن خطی از معادلات برابرد به وسیله‌ی بسط تیلور در هر بلوک پیشنهاد می‌شود. بنابراین، ساده‌کردن معادله‌ی برابرد غیرخطی به نوع خطی، سبب می‌شود که پاسخ‌های تمامی بلوک‌ها به وسیله‌ی میانگین وزنی با یکدیگر ترکیب شوند.

۳- میانگین لگاریتم درست‌نمایی

همانطور که می‌دانیم مدل‌های آماری گاوسی به طور گسترده‌ای برای مدل‌سازی داده‌های فضایی مورد استفاده قرار می‌گیرند اما به هر حال محاسبات و پیاده‌سازی این مدل، یک چالش است چرا که نیازمند معکوس ماتریس کوواریانس است که بسیار پیچیده است، مخصوصاً هنگامی که تعداد داده‌ها بسیار زیاد باشد. این روش درصدد است که با به کارگیری تخمین تصادفی با استفاده از بازنمونه‌گیری به مقابله با این معضل بپردازد. با الگو گرفتن از کوچک کردن واگرایی کولبک-لیبلر، این واگرایی را با استفاده از میانگین محاسبه شده از زیرنمونه‌ها برابرد می‌کنیم که منجر به روش برابرد میانگین لگاریتم درست‌نمایی ماکسیمم می‌شود. در این روش، در هر تکرار، یک زیرنمونه به حجم m از مجموعه داده‌های اصلی انتخاب می‌شود و پاسخ میانگین معادله‌ی امتیازی از روش تقریب تصادفی حاصل می‌شود به طوری که در هر تکرار، برابرد براساس زیرنمونه به روزرسانی می‌شود. از آنجایی که m بسیار کوچکتر از n است و از ایجاد معکوس ماتریس کوواریانس که بزرگ است جلوگیری می‌کند، این روش می‌تواند برا حل مشکل مه داده‌ها مفید باشد.

۴- زیرنمونه‌گیری بر پایه‌ی روش‌های مونت کارلوی زنجیر مارکوفی

هنگامی که چگالی f را بتوان ارزیابی کرد اما به آسانی نتوان از آن نمونه گرفت، برای به دست آوردن نمونه‌های تقریبی و دقیق، روش‌هایی مورد بحث قرار می‌گیرند که از جمله‌ی آن‌ها می‌توان به روش‌های مونت کارلوی زنجیر مارکوفی (MCMC) اشاره کرد. همانطور که می‌دانیم هنگامی که زنجیر تحویل‌ناپذیر و نامتناوب باشد، توزیع آن به توزیع مانای حدی می‌گراید. راهبرد روش MCMC ساختن چنین زنجیری است که توزیع مانای مربوط به آن با توزیع هدف f برابر باشد. از این رو این روش در استنتاج بی‌زی بسیار کاربردی است چرا که در آن f توزیع پسین بی‌زی برای پارامتر است اما هزینه‌ی هنگفت برابرد درست‌نمایی در هر تکرار، از جمله مشکلات موجود بر سر راه مه داده‌ها است. در این روش در پی تحلیل بی‌زی مه داده‌ها با استفاده از الگوریتم متروپولیس-هستینگس هستیم و روش میانگین لگاریتم درست‌نمایی را برای این الگوریتم در MCMC بسط می‌دهیم. نسبت درست‌نمایی برابرد پیشنهادی در الگوریتم متروپولیس، با تقریب میانگین لگاریتم درست‌نمایی براساس k نمونه بوت استرپ به حجم m جایگزین می‌شود. این الگوریتم از به کار بردن تمام داده‌ها در هر تکرار جلوگیری می‌کند و از این جهت می‌توان از این روش در مه داده‌ها استفاده کرد.

۵- نمونه‌های کوچک خودگردان

روش بوت استرپ یک وسیله‌ی ساده و قدرتمند برای ارزیابی کیفیت برابردگراست. در مه داده‌ها که دارای حجم زیادی هستند، این روش برای جلوگیری از محاسبات پیچیده بسیار مناسب است. با استفاده از زیرنمونه‌گیری و انتخاب m از n بوت استرپ، می‌توان هزینه‌های محاسباتی را کاهش داد اما این روش‌ها نیازمند استفاده از اطلاعات پیشین هستند. در این جا روش BLB که ترکیبی از زیرنمونه‌گیری و نمونه‌گیری بوت استرپ است را به عنوان یک وسیله‌ی محاسباتی مناسب با عملکرد قوی برای ارزیابی کیفیت برابردگرا معرفی می‌کنیم. این روش شامل گام‌های زیر می‌باشد:

۱- S زیرمجموعه به اندازه‌ی m از مجموعه داده‌ی اصلی به حجم n انتخاب کنید .

۲- برای هر کدام از S زیرمجموعه، t نمونه‌ی بوت استرپ بگیرید و برآورد نقطه‌ای و مقدار کیفیت آن‌ها را به دست آورید.

۳- سپس S مقادیر به دست آمده از گام دوم را باهم ترکیب کنید تا برآورد کلی حاصل شود (برای مثال میانگین بگیرید).

لازم به ذکر است در زیرنمونه‌ها، m برابر با n^{γ} به طوری که $\gamma \in [0.5, 1]$ پیشنهاد شده است که بسیار کوچکتر از n است و حجم بسیار کمتری را به خود اختصاص می‌دهد.

۶- بسته‌های نرم‌افزار **R**

پردازش مه داده‌ها یکی از موضوعات محاسباتی در دنیای امروز است. نرم افزار **R** بسته‌های افزودنی آن به عنوان محبوب‌ترین نرم-افزار آماری، طیف گسترده‌ای از محاسبات را با کارایی بالا، ارائه می‌دهد. برای هرگونه تحلیل آماری روی مه داده‌ها با استفاده از سیستم مدیریت داده‌ها، نیازمند انجام محاسبات عددی هستیم. خانواده‌ی **bigmemory** برای ایجاد، ذخیره کردن، دسترسی و استفاده از ماتریس‌های حجیم به کار می‌رود. پکیج‌های **biganalytics** برای تجزیه و تحلیل آماری، **bigtabulate** برای عملیات جدول بندی و **bigalgebra** برای اعمال ماتریسی، توابع پیشرفته‌ای را ارائه می‌دهند و استفاده از آن‌ها در محیط‌های موازی، سرعت و کارایی قابل توجهی دارد. به عنوان مثالی دیگر می‌توان به بسته‌ی **ffbase** اشاره کرد که توابع پایه‌ی آماری را برای اعضای **ff** ارائه می‌دهد. **RRE** نیز محصول اصلی انقلاب تجزیه و تحلیل است. همچنین شرکتی است که ابزارهای نرم‌افزار **R** را تولید و پشتیبانی می‌کند. **RRE** بر روی مه داده‌ها و انجام محاسبات در مقیاس‌های بزرگ متمرکز است. در مثالی از داده‌های مربوط به خطوط هوایی در کشور آمریکا به حجم 12GB که شامل جزئیات ورود و خروج پروازها از سال ۱۹۸۷ تا سال ۲۰۰۸ است، مورد مطالعه قرار گرفته و با استفاده از بسته‌های موجود، رگرسیون لوژستیک روی این مه داده‌ها انجام گرفته است. این داده‌ها در دسترس عموم به آدرس (<http://stat-computing.org/dataexpo/2009/>) قرار دارند.

نتایج رگرسیون لوژستیک برای تأخیر در ورود پروازها به شرح زیر است:

	Estimate	Std.Error($\times 10^4$)
(Intercept)	-2.985	9.470
Departure Hour	0.104	0.601
Distance	0.235	4.032
Night	-0.448	8.173
Weekend	-0.177	5.412

این نتایج حاکی از آن است که پروازهای با ساعت خروج دیرتر یا مسافت طولانی‌تر، از احتمال بیشتری برای داشتن تأخیر برخوردارند.

۷- نتیجه‌گیری

چالش‌های حال حاضر مه داده‌ها، بسیار فراتر از قلمرو آمار کلاسیک است و نیازمند داشتن دانش این حوزه، مهارت‌های محاسباتی و فکر و ایده‌های آماری است. در این مقاله به بیان برخی روش‌های آماری برای حل مشکلات مه داده‌ها پرداختیم. به طور کلی این روش‌ها در دو گروه "زیرنمونه‌گیری" و "افراز و تقسیم" قرار می‌گیرند. در این مقاله، از شاخه‌ی زیرنمونه‌گیری، روش‌های میانگین لگاریتم - درستنمایی، زیرنمونه‌گیری بر پایه‌ی روش‌های مونت کارلوی زنجیر مارکوفی و نمونه‌های کوچک خودگردان معرفی شدند. از گروه افراز و تقسیم نیز به بیان روش معادلات برآورد انباشته پرداختیم. لازم به ذکر است که از بیان روش‌های دیگر به دلیل خلاصه‌نویسی نگارشی خودداری شده است.

مراجع

- [1] Chun Wang, Ming-Hui Chen, Elizabet Schifano, Jing Wu, and Jun Yan.) 2015(,“Statistical methods and computing for big data,” Stat Interface, 9(4) : 399-414
- [2] Lin, N. and Xi, R. (2011), “Aggregated Estimating Equation Estimation,” Statistics and Its Interface, 4, 73–83.
- [3] Liang, F., Cheng, Y., Song, Q., Park, J., and Yang, P. (2013), “A Resampling-based
- [4] Stochastic Approximation Method for Analysis of Large Geostatistical Data,” Journal of the American Statistical Association, 108, 325–339.
- [5] Liang, F. and Kim, J. (2013), “A Bootstrap Metropolis–Hastings Algorithm for Bayesian Analysis of Big Data,” Tech. rep., Department of Statistics, Texas A & M University.
- [6] Kleiner, A., Talwalkar, A., Sarkar, P., and Jordan, M. I. (2014), “A Scalable Bootstrap for Massive Data,” Journal of the Royal Statistical Society: Series B (Statistical Methodology), 76, 795–816.

روشی جدید در برآورد ارزش در معرض مخاطره و احتمال پیشامد زیان کرانگین بر اساس فرایند نقطه‌ای خودانگیزی

^۱ سامان سپهوند، محمد قاسم وحیدی اصل

^۱ دانشجوی کارشناسی ارشد دانشگاه شهید بهشتی، گروه بیم سنجی

Saman_sepahvand@yahoo.com

^۲ عضو هیئت‌علمی دانشگاه شهید بهشتی، گروه آمار

m_vahidi_asl@yahoo.com

چکیده مختصر

امروزه مفاهیمی چون سهام، اوراق قرضه، اوراق مشارکت و ابزارهای مالی از این دست برای همگان مفاهیمی آشناست و تقریباً زندگی تمامی افراد، حتی شرایط اقتصادی کشور به‌طور مستقیم و یا غیرمستقیم تحت تأثیر این مفهوم مالی است. بی‌شک فقدان درک و نگرش تخصصی مناسب در انجام مطالعات بازار موجب ایجاد خسارت‌های جبران‌ناپذیری می‌گردد. در این مقاله قصد داریم تا با رویکردی نوین برخی از معیارهای سنجش مخاطره بازارهای مالی ایران را مورد بررسی قرار دهیم. از آنجایی که شاخص‌های مالی و قیمت سهام به‌شدت تحت تأثیر تاریخچه‌ی خود است لذا عدم توجه به این مسئله موجب گمراهی و وقوع بحران‌های مالی می‌شود. وقوع پیشامدهای کرانگین در گذشته و اثرات آن در ایجاد خوشه‌های زیان کرانگین در آینده، می‌تواند مصداقی مناسب از چالش عنوان‌شده باشد. لذا سعی می‌شود تا به کمک فرایندهای نقطه‌ای خودانگیزی نشان‌دار و تابع شدت مرتبط با آن نرخ وقوع پیشامدهای زیان کرانگین در آینده را برآورد و پیش‌بینی نمود. برای برآورد تابع شدت فرایند نقطه‌ای عنوان‌شده از مدل‌های دوره‌ی شرطی خودبازگشتی ACD بهره می‌گیریم، سپس به کمک تابع شدت برآورد شده و توضیحات ارائه‌شده در ادامه‌ی مقاله، برآوردهایی از معیارهای مخاطره VaR و ELP شرطی یک‌روزه را محاسبه می‌کنیم و پس از آن اعتبار برآوردهای حاصل شده را مورد آزمون قرار می‌دهیم. این روش محاسبه معیارهای مخاطره اولین بار توسط حمیدیه و دیگران (۲۰۱۲) [۱] پیشنهاد گردید. در این برآوردها به‌خوبی اثرات دینامیک زمان و رفتار گذشته‌ی فرایند انعکاس داده می‌شود. از خصوصیات مناسب دیگر این برآوردها تفسیرپذیر بودن آن است. درنهایت به کمک مجموعه مشاهدات مالی بازار سهام ایران، معیارهای مخاطره نام برده‌شده محاسبه شده، با چند روش کلاسیک مقایسه می‌گردد.

کلمات کلیدی

فرایند نقطه‌ای خودانگیزی، توزیع تعمیم‌یافته پارتو (GPD)، ارزش در معرض مخاطره (VaR)، احتمال پیشامد زیان کرانگین (ELP)، توزیع‌های دم‌سنگین، مدل‌های دوره‌ی شرطی خودبازگشتی (ACD)، بازده‌های مالی.

۱- مقدمه

به زبان عامیانه پیشامد کرانگین به پیشامدهایی نادر گفته می‌شود که در مباحث بیمه‌ای منجر به رخداد خسارت‌های بسیار بزرگ شود. با این تعریف مقادیر کرانگین موردنظر ما مقادیری مثبت و متعلق به انتهای دم سمت راست توزیع خسارت‌ها است، بنابراین با مدل‌بندی و پیش‌بینی این دسته از پیشامدها می‌توان تا حدود زیادی مخاطرات مالی شرکت‌ها را مدیریت نمود. اهمیت این موضوع زمانی که با محصولات پیچیده و متنوع شرکت‌های مالی روبه‌رو هستیم دوچندان می‌شود، چراکه عدم مدیریت این دسته از محصولات مالی

موجب وقوع بحران‌های مالی شده و برای شرکت‌کنندگان در بازارهای مالی امری بسیار چالش‌برانگیز است. برای سنجش میزان مخاطره‌ی مالی این نوع از محصولات می‌بایست از معیارهای دقیق و مناسب بهره جست، البته هریک از این معیارها دارای نقاط ضعف و قوت مختص به خود است. برای مثال معیار VaR در سطح $q \in (0,1)$ عبارت است از چندک q ام توزیع خسارت‌ها و از آن می‌توان به‌عنوان شاخصی برای سنجش مخاطره‌ی تعهدات ترانزنامه‌ها مانند: "قراردادهای تحویل آتی"، "قراردادهای تاخت"، "قراردادهای پیشرو" و "اختیار معاملات" استفاده نمود. این معیار تنها از طریق توزیع حاشیه‌ای اطلاعات شکل گرفته و دینامیک زمانی مرتبط با بازار را در نظر نمی‌گیرد. این مسئله می‌تواند پیامدهای مهمی را دربر داشته باشد چراکه خوشه‌های زیان کرانگین عمدتاً در مدت‌زمان کوتاهی رخ می‌دهند و می‌توانند منجر به کاهش سریع ارزش معاملات شوند و بحران‌های مالی را در پیش داشته باشند. این پیشامد-های خوشه‌ای از طریق مدل‌هایی که تنها بر مبنای توزیع حاشیه‌ای اطلاعات شکل می‌گیرند به خوبی توضیح داده نمی‌شوند. هدف اصلی این مقاله به‌کارگیری برآوردهای توسعه‌یافته معیارهای مخاطره‌ی مالی پیشنهاد شده توسط حمیدیه و دیگران (۲۰۱۲) [1] برای محاسبه‌ی مقادیر VaR و ELP چندین شرکت فعال در بازار سهام ایران است. با در نظر گرفتن دینامیک زمانی و وابستگی میان مقدار خسارت‌های کرانگین رخ داده شده در گذشته و زمان ورود آن‌ها در آینده می‌توانیم برآوردهای معیار مخاطره را بهبود بخشیم.

۲- تشریح ساختار ریاضی مدل ACD و فرایند نقطه‌ای خودانگیزی

به‌طور کلی سود یا زیان یک دارایی از تغییرات قیمت آن در طول دوره‌ای معین از زمان تعریف می‌شود. هنگامی که ارزش سبد داراستمان مثبت باشد عموماً از شاخص درصد بازده استفاده می‌نماییم. نکته‌ی حائز اهمیت این است که در دوره‌های کوتاه‌مدت شاخص درصد بازده به‌خوبی توسط لگاریتم بازده تقریب زده می‌شود.

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}} \Rightarrow R_t + 1 = \frac{P_t}{P_{t-1}} ; \ln(R_t + 1) = \ln\left(\frac{P_t}{P_{t-1}}\right) ; \text{if } R_t \rightarrow 0 \Rightarrow \ln\left\{\frac{P_t}{P_{t-1}}\right\} \approx \frac{P_t - P_{t-1}}{P_{t-1}} \quad (1)$$

مدل‌های ACD (دوره‌ی شرطی خودبازگشتی) اولین بار در سال ۱۹۹۸ توسط انگل و راسل [2] مطرح شد. از این مدل برای مطالعه و تحلیل اطلاعات مرتبط با زمان انجام معاملات استفاده گردید. این مدل‌ها به‌عنوان یکی از پرکاربردترین مدل‌ها جهت تحلیل و مطالعه‌ی مدت‌زمان سپری شده تا وقوع پیشامدها در علوم‌ی چون اقتصادسنجی شناخته‌شده و تاکنون مورد استقبال فراوان تحلیل‌گران قرار گرفته است. به‌طور کلی از این مدل جهت مطالعه‌ی مدت‌زمان سپری شده تا وقوع پیشامدهای متوالی استفاده می‌شود. به کمک ساختار خاص این مدل قادر هستیم تا با به‌روزرسانی اطلاعات قدرت پیش‌بینی مدل را افزایش دهیم. همچنین از قابلیت‌های این مدل می‌توان به توانایی آن در مدل‌بندی فرایندهای رفتار خوشه‌ای و ساختار بیش‌پراکنده اشاره نمود. شاید یکی از بهترین ویژگی‌های مدل ACD ایجاد شرایط لازم جهت تحلیل و آزمون برخی از مفاهیم مرتبط با ساختار بازارهای مالی است (تسای [۳]). پس از انگل و راسل، مشتقات بسیاری از مدل‌های ACD ایجاد و معرفی شد. مدل‌های Log-ACD یکی از این مشتقات به شمار می‌رود، این مدل برای اولین بار توسط جیوت و باوونس در سال ۲۰۰۰ [4] معرفی شد. در این مقاله از این مدل بهره می‌بریم. فرض کنید که دنباله‌ی $0 = t_0 \leq t_1 \leq t_2 \leq \dots \leq t_n \leq \dots$ نشان دهنده‌ی زمان وقوع پیشامدهایی دلخواه باشند. همچنین $N(t)$ نشان دهنده‌ی تعداد پیشامدها تا زمان t باشد. در این صورت $T = t_{N(T)}$ زمان مشاهده‌ی آخرین ورودی است به‌طوری‌که $0 = t_0 \leq t_1 \leq t_2 \leq \dots \leq t_{N(T)} = T$. حال متغیر تصادفی X_j را به‌صورت $X_j = T_{j+1} - T_j$ تعریف می‌کنیم. به متغیر تصادفی X_j مدت‌زمان سپری شده تا وقوع $j+1$ امین پیشامد متوالی گوئیم و هدف مدل بندی این فرایند تصادفی است. جیوت و باوونس در سال ۲۰۰۰ [4] مدل ذیل را جهت مدل بندی X_j پیشنهاد دادند:

$$X_t = \exp\{\psi_t\} \varepsilon_t, \quad \psi_t = \omega + \sum_{j=1}^p \alpha_j g(x_{t-1}, \varepsilon_{t-1}) + \sum_{j=1}^q \beta_j \psi_{t-1} \Rightarrow E\{X_t | H_{t-1}\} = \exp\{\psi_t\} E(\varepsilon_t) \quad (2)$$

که در آن $\Theta = \{\omega, \alpha_1, \dots, \alpha_p, \beta_1, \dots, \beta_q\}$ پارامترهای مجهول و قابل برآورد مدل هستند.

۳- تفسیر تابع شدت فرایند نقطه‌ای $N(t)$ از طریق تابع مخاطره‌ی $X_{N(t)}$:

با توجه به تابع نرخ مخاطره‌ی مرتبط با فرایند تصادفی $X_{N(t)}$ می‌توان تابع شدت شرطی مرتبط با فرایند نقطه‌ای $N(t)$ را به صورت زیر مطرح نمود. اگر متغیر تصادفی اغتشاش را به صورت $\varepsilon_t \sim \text{Wibull}(1, \gamma)$ در نظر بگیریم داریم:

$$P(X_{N(t)+1} < x | H_t) = P(\exp\{\psi_{N(t)+1}\} \varepsilon_{N(t)+1} < x | H_t) \Rightarrow$$

$$f_{X_{N(t)+1}}(x | H_t) = \gamma \exp\{-\psi_{N(t)+1}\} (\exp\{-\psi_{N(t)+1}\} x)^{\gamma-1} \exp(-x \exp\{-\psi_{N(t)+1}\})^\gamma \quad (3)$$

$$X_{N(t)+1} | H_t \sim \text{Wibull}(e^{-\psi_t}, \gamma) \Rightarrow \lambda(x_t | H_t) = \frac{\frac{\partial}{\partial x} P(X_{N(t)+1} < x | H_t)}{P(X_{N(t)+1} > x | H_t)} = (e^{-\psi_t})^\gamma (x_t)^{\gamma-1} \gamma$$

از طرفی می‌دانیم که:

$$P(\{N(t, t+h) = 0\} | H_t) = P(X_{N(t)} > h | H_t) = \exp\{-h\lambda(t | H_t)\} = \exp\{-(h e^{-\psi_{N(t)}})^\gamma\} \quad (4)$$

۴- مدل بندی بازده‌های مالی به کمک فرایند نقطه‌ای خودانگیزی نشان‌دار

فرض کنید $\{Z_t\}_{t \geq 0}$ منفی لگاریتم بازده‌های مرتبط با دارایی‌های تحت مطالعه باشد به طوری که $t \in \mathbb{N}^+$. در صورتی که ارزش دارایی تحت مطالعه در زمان t را به صورت P_t نمایش دهیم خواهیم داشت $Z_t = -\ln\left(\frac{P_t}{P_{t-1}}\right) \Rightarrow Z_t = -\ln(R_t)$ با در نظر گرفتن یک مقدار مناسب به عنوان مقدار حد آستانه‌ی u برای مقادیر Z_t به تمامی مقادیر $Z_t > u$ مقدار زیان کرانگین گوئیم. زیان‌های کرانگین را به صورت Z_{T_i} نشان می‌دهیم که در آن اندیس T_i نشان‌دهنده‌ی زمان وقوع i امین زیان کرانگین است. بنابراین همواره داریم $\{Z_{T_i} > u; T_i \in \mathbb{N}^+, i \in \mathbb{N}\}$. همچنین متغیر وابسته به Z_{T_i} را به صورت $Y_i = Z_{T_i} - u$ تعریف می‌نماییم. مقدار Y_i نشانه‌ی فرایند نقطه‌ای تحت مطالعه در نظر گرفته می‌شود. متغیر تصادفی دیگری به صورت $X_{i+1} = T_{i+1} - T_i; i \geq 1$ تعریف می‌نماییم. به وضوح $X_i; i \geq 1$ مدت‌زمان سپری‌شده تا وقوع زیان کرانگین i ام متوالی تعریف می‌شود. تاریخچه و رفتار گذشته‌ی فرایند نقطه‌ای نشان‌دار $\{(T_i, Y_i)\}_{i \geq 1}$ را به صورت یک پالایه H_i نشان می‌دهیم و داریم $H_n \equiv \sigma\{(X_j, Y_j); j \leq n\}$. همچنین فرایند نقطه‌ای شمارشی $N(t)$ را به صورت $N(t) \equiv \max\{n \geq 0; T_n \leq t, t > 0\}$ تعریف می‌کنیم. تابع شدت شرطی فرایند تحت مطالعه را به صورت زیر نشان می‌دهیم:

$$\lambda(t | H_t) = \lim_{h \rightarrow 0} \frac{P(N\{t, t+h\} > 0 | H_t)}{h} \quad (5)$$

این تابع شدت به مقادیر مشاهدات $\{(X_i, Y_i)\}_{i \geq 1}$ وابسته است و این نکته اساس ایجاد یک فرایند خودانگیزی است.

۵- برآورد معیار مخاطره‌ی ELEP و VaR شرطی

با توجه به فرایند نقطه‌ای تعریف‌شده‌ی $N(t)$ دو پیشامد زیر معادل یکدیگر هستند:

$$\{N(t, t+1) = 1 | H_t\} = \{Z_{t+1} > u | H_t\} \quad (6)$$

درواقع پیشامد وقوع یک زیان کرانگین مازاد بر u ، در زمان $t+1$ معادل آن است که در زمان $t+1$ فرایند نقطه‌ای $N(t)$ یک ورودی داشته باشد. به کمک معادل بودن دو پیشامد مطرح‌شده داریم:

$$\rho_u(t) = P(\{Z_{t+1} > u\} | H_t) = 1 - P(\{N(t, t+1) = 0\} | H_t) = 1 - \exp\{-\lambda(t | H_t)\} \quad (7)$$

بنابراین به کمک تابع شدت برآورد شده‌ی مرتبط با فرایند نقطه‌ای $N(t)$ می‌توان یک برآورد مناسب جهت محاسبه معیار مخاطره‌ی ELEP تعریف نمود. با توجه به ثبت مشاهدات به‌صورت روزانه می‌توانیم معیار مخاطره‌ی VaR شرطی را برای یک روز آتی محاسبه نماییم. توزیع متغیر تصادفی Y_i (مقادیر زیان مازاد بر u) به‌صورت ذیل است:

$$F_Y(y_i) = P(Y_i < y_i) = P(\{Z_{T_i} - u < y_i\} | Z_{T_i} > u) \quad , \quad Y_i = Z_{T_i} - u | Z_{T_i} > u \quad (8)$$

همچنین با توجه به قضیه‌ی "فیشر- تیپت" توزیع مشاهدات کرانگین Y_i به‌صورت ذیل حاصل می‌شود:

$$P(Y_i < x) = H_{\kappa, \xi}(x) \equiv 1 - \left(1 + \xi \frac{x}{\kappa}\right)_+^{-\frac{1}{\xi}} \quad (9)$$

که در آن $\xi \in \mathbb{R}$, $\kappa > 0$, $(x)_+ = \max(0, x)$. کلاس توزیع‌های GPD جهت مدل بندی Y_i ها کلاسی بسیار منعطف و دارای خواصی مناسب است. پس از محاسبه‌ی مقدار مناسب حد آستانه‌ی u به کمک روش هیل و برآورد پارامترهای $\xi \in \mathbb{R}$, $\kappa > 0$ توزیع GPD، به سراغ محاسبه‌ی معیار مخاطره‌ی شرطی VaR می‌رویم. پیشامد $\{Z_{t+1} > \zeta_q(t)\}$ را می‌توان به‌صورت ذیل تعریف نمود:

$$P(\{Z_{t+1} > \zeta_q(t)\}) = P(\{Z_{t+1} - u > \zeta_q(t) - u\} | Z_{t+1} > u) P(\{Z_{t+1} > u\}) \quad , \quad \zeta_q(t) > u \quad (10)$$

همچنین ثابت نمودیم که دو پیشامد $\{N(t, t+1) = 1 | H_t\} = \{Z_{t+1} > u | H_t\}$ معادل هستند، بنابراین داریم

$$\begin{aligned} P(\{Z_{t+1} > \zeta_q(t)\} | H_t) &= P(\{Z_{t+1} - u > \zeta_q(t) - u\} | Z_{t+1} > u, H_t) P(\{Z_{t+1} > u\} | H_t) \\ &= P(\{Y_{N(t)+1} > \zeta_q(t) - u\} | Y_{N(t)+1} > 0, H_t) P(\{N(t, t+1) = 1\} | H_t) = 1 - q \end{aligned} \quad (11)$$

با توجه به توزیع مقادیر کرانگین Y_i و مستقل در نظر گرفتن آن‌ها از رفتار گذشته‌ی فرایند داریم

$$P(\{Z_{t+1} > \zeta_q(t)\} | H_t) = P(\{Y_{N(t)+1} > \zeta_q(t) - u\}) (1 - P(\{N(t, t+1) = 0\} | H_t)) = 1 - q \quad (12)$$

با جایگذاری توزیع $Y_i \sim GPD(\kappa, \xi)$ داریم:

$$\begin{aligned} P(\{Z_{t+1} > \zeta_q(t)\} | H_t) &= \left(1 + \xi \frac{\zeta_q - u}{\kappa}\right)^{-\frac{1}{\xi}} (1 - P(\{N(t, t+1) = 0\} | H_t)) \\ &= \left(1 + \xi \frac{\zeta_q - u}{\kappa}\right)^{-\frac{1}{\xi}} (1 - \exp\{-\lambda(t | H_t)\}) = 1 - q \end{aligned} \quad (13)$$

بنابراین با این ترفند معیار VaR شرطی را به تابع شدت فرایند نقطه‌ای مرتبط می‌سازیم. درنهایت با حل معادله‌ی بالا داریم

$$\zeta_q(t) = u + \frac{\kappa}{\xi} \left\{ \left(\frac{1 - \exp(-\lambda(t | H_t))}{1 - q} \right)^\xi - 1 \right\} \quad (14)$$

مراجع

- [1] Hamidieh, K., Stoev, S., & Michailidis, G. (2013). Intensity - based estimation of extreme loss event probability and value at risk. *Applied Stochastic Models in Business and Industry*, 29(3), 171-186.
- [2] Engle, R. F., & Russell, J. R. (1998). Autoregressive conditional duration: a new model for irregularly spaced transaction data. *Econometrica*, 1127-1162
- [3] Tsay, R. S. (2009). Autoregressive conditional duration models. In *Palgrave handbook of econometrics* (pp. 1004-1024). Palgrave Macmillan UK.
- [4] Bauwens, L., & Giot, P. (2000). The logarithmic ACD model: an application to the bid-ask quote process of three NYSE stocks. *Annales d'Economie et de Statistique*, 117-149

سیستم پرداخت حق بیمه وسایل نقلیه در زمان سوخت‌گیری

رحیم محمودوند^۱، فاطمه سادات آل حسینی^۲

^۱ گروه آمار، دانشگاه بوعلی سینا، همدان
r.mahmodvand@gmail.com

^۲ کارشناسی ارشد آکچوئری، شرکت بیمه کارآفرین، تهران
alehosseinifateme@gmail.com

چکیده مختصر

پایین بودن ضریب نفوذ بیمه، خسارت‌های جانی و مالی بسیار سنگین ناشی از تصادف وسایل نقلیه، و افزایش چشم‌گیر مصرف سوخت از جمله دغدغه‌هایی هستند که همواره مورد توجه دولتمردان قرار گرفته است. اجباری شدن بیمه شخص ثالث و سهمیه بندی سوخت (در چند سال گذشته) از جمله اقدام‌هایی بوده‌اند که در جهت کنترل پیامدهای این مسائل مطرح شده‌اند. با این وجود، آمارها گویای این مطلبند که این اقدام‌ها موفقیت صد در صدی در نیل به اهداف مطلوب فوق را نداشته و از این رو، مطالعه و اجرای راهکارهای دیگر نیز ضروری خواهد بود. در این مقاله، امکان استفاده از یک سیستم جایگزین برای بیمه‌ی دارندگان وسایل نقلیه موتوری را مورد مطالعه قرار می‌دهیم. در این سیستم، مبلغ حق بیمه، در زمان سوخت‌گیری و با توجه به میزان سوخت‌گیری و سایر عوامل ریسک محاسبه و پرداخت می‌شود. از آنجاکه به کارگیری این سیستم منجر به پرداخت حق بیمه توسط تمامی دارندگان وسایل نقلیه می‌شود، از یک سو می‌تواند باعث افزایش ضریب نفوذ بیمه و از سوی دیگر، کاهش مصرف سوخت خواهد شد.

کلمات کلیدی

حق بیمه، سوخت‌گیری، وسایل نقلیه.

۱- مقدمه

در مباحث بیمه‌ای علاوه بر تعیین نرخ حق بیمه‌ی مناسب، که طبق قانون بیمه شخص ثالث جدید بر اساس خصوصیات خودرو و راننده است، سازوکار جذب مشتریان نیز از اهمیت بالایی برخوردار است. اهمیت این موضوع در رشته‌هایی مانند بیمه وسایل نقلیه در کشور ایران از چند منظر بیشتر از سایر رشته‌های بیمه‌ای است:

- داشتن بیمه شخص ثالث برای وسایل نقلیه اجباری است؛
 - بخش قابل توجهی از پرتفوی صدور شرکت‌های بیمه متعلق به این رشته است؛
 - خسارت‌های ناشی از بروز ریسک در این رشته‌ی بیمه‌ای نسبت به دیگر رشته‌های بیمه‌ای وسیع‌تر است.
- متأسفانه، با وجود الزام قانونی در داشتن بیمه شخص ثالث، همچنان بسیاری از دارندگان وسایل نقلیه در خرید این بیمه‌نامه کوتاهی می‌کنند. جدول ۱ آمار تعداد وسایل نقلیه موتوری و میزان پوشش بیمه‌ای آنها را طی سال‌های ۱۳۸۹ تا ۱۳۹۳ نشان می‌دهد. همانگونه که ملاحظه می‌شود، در این بازه، نزدیک به ۴۰ درصد وسایل نقلیه بدون بیمه شخص ثالث در حال تردد هستند.

جدول ۱- وضعیت استفاده از بیمه وسایل نقلیه در سال‌های ۱۳۸۹ تا ۱۳۹۳ (منبع: بیمه مرکزی)

عنوان	۱۳۸۹	۱۳۹۰	۱۳۹۱	۱۳۹۲	۱۳۹۳
تعداد وسایل نقلیه موتوری (میلیون)	۵/۲۳	۹/۲۵	۳/۲۷	۴/۲۸	۳/۲۹
وسایل نقلیه تحت پوشش بیمه ثالث (درصد)	۵۸	۵۸	۵۹	۶۰	۶۳
وسایل نقلیه تحت پوشش بیمه بدنه (درصد)	۱۳	۱۴	۱۳	۱۳	۱۳

در این مقاله، مروری بر سیستم پرداخت حق بیمه در زمان سوخت‌گیری، به عنوان یک روش جایگزین که منجر به افزایش پوشش بیمه‌ای در بین وسایل نقلیه در حال تردد می‌شود، خواهیم داشت.

۲- ادبیات موضوع

اثرات مخربی که حوادث ناشی از تصادف‌های رانندگی برای کشورها به دنبال دارند، موجب شده است تا متولیان و محققان به معرفی سیستم‌ها و راهکارهای جدید متوسل شوند. نهادهایی مانند پلیس و وزارت راه و شهرسازی در کنترل حوادث نقش بسزایی دارند. از سوی دیگر، استفاده از بیمه که می‌تواند فشار مالی ناشی از تصادف‌ها را برای دارندگان وسایل نقلیه کمتر سازد، به عنوان یک راهکار مورد استفاده قرار می‌گیرد. البته گستردگی خسارت‌های ناشی از وسایل نقلیه در کشور ما به حدی زیاد است که قانونگذار داشتن بیمه شخص ثالث را برای استفاده از وسایل نقلیه اجباری کرده است. هر چند وجود بیمه به عنوان حامی نباید به گونه‌ای باشد که رانندگان در استفاده از وسایل نقلیه جانب احتیاط را رعایت نکنند (عدالتی و همکاران، ۱۳۸۹).

به منظور کنترل رفتار رانندگان توسط بیمه‌گر، سیستم‌های تعدیل‌گری که بعدها به نام سیستم‌های پاداش-جریمه معروف شدند، معرفی و به کار گرفته شده است. در این سیستم‌ها، میزان حق بیمه دارندگان وسایل نقلیه با توجه به سوابق خسارتی آنها تعدیل می‌شود؛ به این شکل که سوابق بدون خسارت منجر به کاهش حق بیمه و سوابق خسارتی منجر به افزایش آن می‌شود. استفاده از چنین سیستم‌هایی مزایایی چون احتیاط بیشتر در رانندگی به منظور برخورداری از تخفیف را می‌تواند به همراه داشته باشد (دنوئیت و همکاران، ۲۰۰۷).

علاوه بر این، در چند سال اخیر سیستم‌های دیگری پیشنهاد شده است که در آنها متناسب با مسافت پیموده شده یا در صورت رانندگی حق بیمه دریافت می‌شود (لیتمن، ۲۰۰۰). از مزایای این سیستم‌ها آن است که دارندگان وسایل نقلیه در صورت استفاده کمتر، حق بیمه کمتری می‌پردازند که این موضوع آنها را به استفاده کمتر (و در صورت امکان، استفاده از وسایل حمل و نقل عمومی) ترغیب می‌کند. به کارگیری این سیستم‌ها نیازمند کنترل مسافت پیموده شده با استفاده از دستگاه‌های مجزا است که امکان دستکاری و تقلب در آن زیاد خواهد بود. با این وجود، در صورت اجرای درست می‌توانند کارآمد و اثرگذار باشند.

تحلیل بالا نشان می‌دهد که در حال حاضر برای محاسبه‌ی حق بیمه‌ی وسایل نقلیه دو رویکرد کلی وجود دارد:

رویکرد اول- در نظر گرفتن یک حق بیمه‌ی ثابت برای کل دوره بیمه: این رویکرد که در اغلب کشورها استفاده می‌شود در عمل ساده‌تر بوده و مزایا و معایبی نیز به همراه دارد. تاکنون تحقیق‌های بسیاری به منظور توسعه و بهبود این رویکرد انجام شده است. برای یک مرور بر برخی از روش‌های به کار گرفته شده، عدالتی و همکاران (۱۳۸۹) را ببینید.

رویکرد دوم- در نظر گرفتن حق بیمه‌ی متغیر برای کل دوره‌ی بیمه: این رویکرد که عمر کمتری دارد و هنوز مقبولیتی در کشورها پیدا نکرده است کمتر مورد توجه محققان بوده و لذا منابع علمی کمی نیز در خصوص آن وجود دارد. دو نمونه سیستم طراحی شده در این زمینه، یکی سیستم‌های پرداخت به میزان رانندگی (PAYD) و دیگری سیستم‌های پرداخت در زمان سوخت‌گیری (PAP) است. برای یک مرور در زمینه سیستم‌های PAYD، پفژن و همکاران (۲۰۱۳) و برای PAP، خازوم (۲۰۰۰) را ببینید.

۳- سیستم‌های پرداخت در زمان سوخت‌گیری

از آنجاکه هیچ وسیله‌ی نقلیه‌ای بدون داشتن سوخت حرکت نمی‌کند، به نظر می‌رسد یک راهکار مناسب، استفاده از اضافه نرخ سوخت در دریافت حق بیمه (یا بخشی از آن) باشد. سیستم‌هایی مشابه این ایده در برخی از کشورها مانند آفریقای جنوبی و برخی از ایالت‌های کانادا و آمریکا به کار گرفته شده است. از مزایای چنین سیستم‌هایی می‌توان به موارد زیر اشاره کرد:

- هیچ وسیله نقلیه در حال حرکتی بدون بیمه نخواهد بود،
- نیازی به اجبار برای خرید بیمه نیست چرا که خود بخود خریداری می‌شود،
- حق بیمه به صورت تقسیط پرداخت می‌شود که منجر به حمایت از قشر ضعیف می‌شود،
- باعث کنترل مصرف سوخت و ترغیب مردم به استفاده از وسایل نقلیه عمومی می‌شود.

فرض کنید c_i میزان پرداختی توسط بیمه‌گذار i به ازای هر واحد سوخت و c میزان هزینه سوخت به ازای هر واحد باشد. آنگاه بایستی مقدار m_i به گونه‌ای تعیین شود که داشته باشیم:

$$c_i = c + m_i$$

بر اساس این معادله کل مبلغ حق بیمه‌ای که بیمه‌گذار i ام پرداخت می‌کند بایستی در رابطه زیر صدق کند:

$$p_i = m_i f_i$$

که در آن f_i مقدار سوختی است که بیمه‌گذار i در طول یکسال سوخت‌گیری می‌کند. به منظور تعیین مقدار m_i می‌توان ابتدا بیمه‌گذاران را با توجه به عوامل ریسک گروه‌بندی کرد، به نحوی که اعضای داخل هر گروه بیشترین شباهت را از نظر ریسک به یکدیگر داشته باشند و گروه‌های مختلف بیشترین اختلاف را با هم دارا باشند. فرض کنید، L گروه g_1, \dots, g_L تشکیل شده باشند. آنگاه یک برآورد شهودی از مقادیر m_i را می‌توان به صورت زیر به دست آورد:

$$m_i = \frac{\sum_{j \in g_i} p_j}{\sum_{j \in g_i} f_j}, \quad i = 1, \dots, L.$$

بر این اساس، هزینه هر واحد سوخت برای بیمه‌گذار به شکل زیر خواهد:

$$c_i = c + \frac{\sum_{j \in g_i} p_j}{\sum_{j \in g_i} f_j}, \quad i = 1, \dots, L.$$

به عنوان یک مثال عددی، صرف نظر از گروه ریسک، کل حق بیمه‌ی تولیدی وسایل نقلیه بر اساس حدود ۳۰ میلیون وسیله نقلیه در حال تردد می‌بایست رقمی در حدود ۱۲ هزار میلیارد تومان باشد. همچنین به طور متوسط معادل حدود ۱۴۰ میلیون لیتر بنزین در روز توسط این وسایل نقلیه مصرف می‌شود. اگر قیمت بر حسب نرخ بنزین در هر لیتر حدود

$$100000 + \frac{120000}{51.1} = 12353$$

باشد، کل مبلغ حق بیمه تولیدی تامین شده و هیچ وسیله نقلیه بدون بیمه‌ای نیز وجود نخواهد داشت.

۴- نتیجه‌گیری

در این مقاله سیستم پرداخت حق بیمه در زمان سوخت‌گیری به عنوان یک سیستم اجرایی به منظور افزایش نفوذ بیمه وسایل نقلیه معرفی شد. در ادامه و در کنار بررسی مزایای این سیستم، رویکرد ساده‌ای نیز در انجام محاسبات فنی و محاسبه‌ی هزینه سوخت پیشنهاد شد. نتایج یک مثال عددی براساس اطلاعات تقریبی از کشور ایران نشان می‌دهد که با افزایش حدود ۲۳۵ تومان به مبلغ بنزین می‌توان تمام وسایل نقلیه را تحت پوشش بیمه قرار داد.

مراجع

- [۱] عدالتی، علیرضا؛ محمودوند، رحیم؛ شکوهی، فرهاد و عباسیانفر، عباسقلی (۱۳۸۹). بررسی سیستم‌های پاداش- جریمه‌ی بیمه‌ی شخص ثالث اتومبیل در سایر کشورها و پیشنهاد سیستمی مناسب برای کشور. طرح پژوهشی، پژوهشکده بیمه مرکزی.
- [2] Denuit, M. Marchal, X., Pitrebois, S. and Walhin, J.F. (2007). Actuarial Modelling of Claim Counts. Wiley, New York.
- [3] Litman, T. (2000). Distance-Based Vehicle Insurance-A Practical Strategy for more Optimal Vehicle Pricing, Chapter in Book: Social Costs and Sustainable Mobility, Springer.
- [4] Khazoom, J.D. (2000). Pay-at-Pump Auto Insurance. Discussion Paper 98-13-Rev, Resources for the Future, Washington, DC. Available at www.rff.org/files/sharepoint/WorkImages/Download/RFF-DP-98-13-REV.pdf.
- [5] PAEFGEN, J., STAAKE, T. and THIESSE, F. (2013) Evaluation and aggregation of pay-as-you-drive insurance rate factors: A classification analysis approach. *Decision Support Systems*, 56, 192-201.

شناسایی عوامل موثر بر حوادث ناشی از کار در کارگاه‌های صنعتی با تأکید بر رگرسیون لسو

حسین علیزاده^۱، کیانا ملک پور^۲، گلستان احسنی^۳

^۱ مرکز انفورماتیک، دانشگاه تهران، تهران
alizadeh.stat68@gmail.com

^۲ مرکز آمار و اطلاعات راهبردی، وزارت تعاون کار و رفاه اجتماعی، تهران،
malekpour_kiana@yahoo.com

^۳ دفتر صنعت، معدن و زیربنایی، مرکز آمار ایران، تهران،
golsan.l.ahsani@gmail.com

چکیده مختصر

این مقاله قصد دارد با استفاده از روش آماری رگرسیون لسو مدل مناسبی را برای تعداد وقوع حادثه ناشی از کار در کارگاه‌های صنعتی ۱۰ کارکن و بیشتر ارائه نماید. از آنجا که این نوع رگرسیون در مقایسه با رگرسیون حداقل مربعات (OLS) مدلی است که دارای واریانس برآورد ضرایب کمتر بوده لذا به نظر می‌رسد مدل مناسبتری باشد. از اینرو در پژوهش حاضر با تکیه بر روش CV علاوه بر مدل رگرسیون حداقل مربعات به بررسی و مقایسه مدل رگرسیون ریج و لسو پرداخته و در نهایت با بهره‌گیری از معیارهای گزینش مدل، رگرسیون لسو را برای این داده‌ها مناسب‌تر تشخیص داده می‌شود.

کلمات کلیدی

حادثه ناشی از کار، رگرسیون لسو، رگرسیون خطی، کارگاه‌های صنعتی ۱۰ کارکن و بیشتر.

۱- مقدمه

یکی از مباحث حایز اهمیت در خصوص بنگاه‌های اقتصادی از جمله کارگاه‌های صنعتی، بهره‌وری نیروی کار و سودآوری بنگاه می‌باشد. چراکه این دو مهم بر بقای کارگاه و رقابت پذیر بودن آن تأثیرگذار هستند. در مباحث مرتبط با بهره‌وری نیروی کار و سودآوری همواره به دو مقوله مهم توجه می‌شود. نخست، هزینه‌هایی که بابت جبران خسارات مالی یا جانی پرداخت می‌شود و دوم مبالغ دریافتی بالقوه‌ای که به سبب رخداد حادثه، بنگاه از دریافت آنها محروم می‌شود. هر دو این موارد در هنگام بروز حادثه دامنگیر بنگاه اقتصادی می‌گردند و از اینرو است که کارفرمایان اقدام به بیمه نمودن کارکنان و بنگاه اقتصادی خود می‌نمایند. در اصل حتی با پرداخت کل خسارات مالی و جانی از سوی شرکت‌های بیمه، درآمد بالقوه‌ای که به علت حادثه از دست می‌رود، جبران‌ناپذیر است. لذا همواره کارفرمایان، شرکت‌های بیمه و سیاستگذاران بازار کار به منظور پیشگیری و پیش بینی احتمال رخداد حوادث ناشی از کار در پی شناسایی عوامل موثر بر این نوع از حوادث بوده‌اند. بر این اساس این مقاله قصد دارد با استفاده از نتایج طرح آمارگیری کارگاه‌های صنعتی ۱۰ کارکن و بیشتر سال ۱۳۹۳ مرکز آمار ایران مدلی مناسب را برای حوادث ناشی از کار ارائه نماید. بدین لحاظ با شناسایی

عوامل احتمالی، برازش مدل رگرسیون خطی، رنج و لسو و مقایسه نتایج حاصل از برازش هر یک از این مدل‌ها، مدل مناسب‌تر را انتخاب نماید.

۲- معرفی داده‌ها

همانطور که در قسمت قبل بیان گردید، داده‌ها مورد استفاده در مقاله حاضر برگرفته از نتایج طرح آمارگیری کارگاه‌های صنعتی ۱۰ کارکن و بیشتر می‌باشد. از آنجایی که نتایج برای سال‌های ۱۳۹۴ و ۱۳۹۵ هنوز منتشر نگردیده به ناچار سال ۱۳۹۳ به عنوان سال مورد بررسی انتخاب شده‌است.

آمارگیری از کارگاه‌های صنعتی ۱۰ کارکن و بیشتر، با هدف شناسایی ویژگی‌های کارگاه‌های صنعتی به صورت سالانه صورت می‌پذیرد. این طرح به صورت تمام‌شماری بوده و فقط کارگاه‌های ۱۰ تا ۴۹ کارکن ۱۳ استان به صورت نمونه انتخاب می‌گردند. در این طرح سوال‌های متنوعی با موضوع‌های حادثه ناشی از کار، آموزش‌های ایمنی، هزینه‌های و دریافتی‌های کارگاه، محصولات آن، ویژگی‌های شاغلان مورد پرسش قرار می‌گیرد.

۳- تعاریف متغیرهای به‌کاررفته

کارگاه صنعتی: مکان ثابتی است که در آن مجموعه‌ای از سرمایه و نیروی کار به منظور تولید یک یا چند محصول صنعتی به کار گرفته شده‌است.

کارگران ساده: به کارگرانی اطلاق می‌شود که شغل مورد تصدی آنها نیاز به تعلیم و کسب مهارت ندارد.

کارگران ماهر: به کارگرانی اطلاق می‌شود که به اعتبار دانش فنی و تجربه ای که کسب کرده‌اند توانایی انجام کارهای فنی را دارند.

تکنسین‌ها: به شاغلانی اطلاق می‌شود که ضمن طی دوره آموزش کاردانی، تجربه و مهارت کافی را بدست آورده‌اند.

مهندسین: به شاغلانی اطلاق می‌شود که دارای تحصیلات عالی حداقل در سطح کارشناسی بوده و در خط تولید مشغول به انجام وظیفه هستند.

شاغلان غیرتولیدی: شامل کارکنان دفتری، اداری، خدمات و امور حمل و نقل می‌باشند که مستقیماً در امر تولید و ساخت دخالت ندارند.

پرداختی بابت کارهای کنتراتی: پرداختی کارگاه به افراد یا کارگاه‌های دیگر به منظور تولید یا تکمیل کالاهای خود است.

پرداختی بابت تعمیرات جزئی ساختمان: پرداختی‌های کارگاه بابت تعمیرات و عملیاتی که برای نگهداری و حفاظت ساختمان کارگاه انجام می‌گیرد.

پرداختی بابت تعمیرات جزئی ماشین آلات، تجهیزات اداری و وسایل نقلیه: پرداختی‌های کارگاه بابت تعمیرات و عملیاتی که برای نگهداری و حفاظت از ماشین آلات، تجهیزات و وسایل نقلیه خود به دیگران می‌پردازد.

اموال سرمایه‌ای: زمین، ساختمان، وسایل نقلیه و همچنین وسایل و تجهیزات متعلق به کارگاه که عمر مفید آنها از یک سال بیشتر است.

هزینه تعمیرات اساسی اموال سرمایه‌ای: هزینه‌های کارگاه بابت تعمیراتی است که عمر مفید، قدرت تولیدی و یا ارزش اموال سرمایه‌ای کارگاه را افزایش می‌دهد.

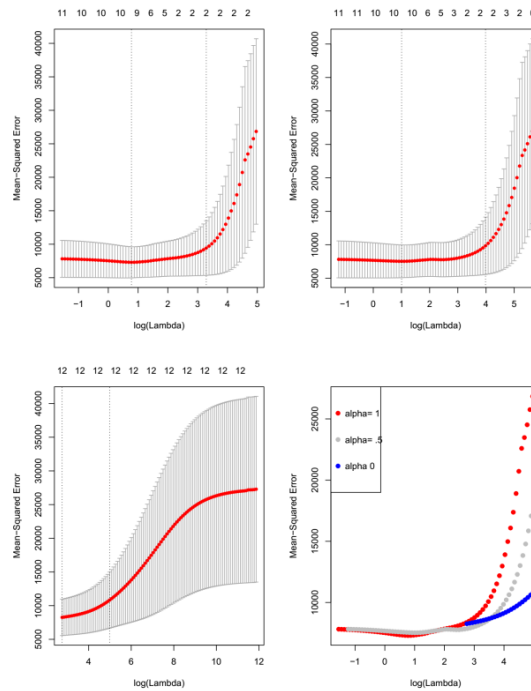
۴- معرفی رگرسیون‌های رنج و لسو

همانطور که مطلعید در رگرسیون حداقل مربعات (OLS) با تعدادی متغیر پیشگو (مستقل) X_i و $i = 1, \dots, N$ متغیر پاسخ (وابسته) روبرو هستیم. هدف از برازش چنین مدل رگرسیونی یافتن مدل خطی بین X ها و Y به طوری که مربع خطای باقی‌مانده‌ها حداقل گردد. به عبارت بهتر برآورد ضرایب خطی در مدل رگرسیونی مذکور با حداقل نمودن مربع خطای باقی‌مانده‌ها حاصل می‌شود. این رگرسیون با دو مساله روبرو می‌باشد. نخست دقت پیش بینی: برآوردهای رگرسیون حداقل مربعات اغلب آریبی کم اما واریانس

زیاد دارد که این مهم خود دقت پیش بینی را تحت الشعاع قرار می‌دهد. دومین مساله تفسیر آن است. به بیان دیگر، در هنگامی که تعداد متغیرهای پیشگو زیاد می‌باشد اغلب انتخاب زیرمجموعه با تاثیرگذاری بالا مد نظر است و یافتن چنین زیرمجموعه‌ای خود مساله مهمی می‌باشد. برای پاسخ به چنین مساله‌ای، با بهره‌گیری از رگرسیون‌های ریج یا لسو چنین زیرمجموعه‌ای شناسایی می‌شود. از اینرو رگرسیون‌های ریج و لسو با کاهش برخی ضرایب تلاش در انتخاب موثرترین زیرمجموعه متغیرهای پیشگو را دارند. در این رگرسیون‌ها اگرچه ممکن است برآوردهای ضرایب نارایب نباشند، واریانس برآوردها کم است. ولی رگرسیون لسو بر رگرسیون ریج مزیت دارد چراکه در رگرسیون ریج علی‌رغم کاهش ضرایب، هیچ‌گاه آنها صفر نخواهند بود و لذا بین ضرایب ضروری و غیرضروری نمی‌توان تمایز قطعی قائل شد. درحالی‌که رگرسیون لسو امکان دارد برای برخی متغیرها، ضریب صفر برآورد کند. همچنین نکته حایز اهمیت در خصوص رگرسیون‌های مذکور این مهم است، در رگرسیون خطی باید زیرمجموعه‌ای متغیرها انتخاب شود که فاقد هیچ گونه هم خطی با هم نبوده و از یکدیگر مستقل باشند اما این موضوع در زمان کار با داده‌های واقعی به سختی محقق می‌گردد. ولی در رگرسیون‌های نامبرده از شرط عدم همخطی صرف نظر شده و اتفاقاً با وجود ارتباط بین متغیرهای مستقل اقدام به برآورد ضرایب و گزینش موثرترین زیرمجموعه از این متغیرها می‌شود.

۵- ارائه مدل

همانطور که قبلاً بیان گردید، با استفاده از اطلاعات مورد پرسش در طرح آمارگیری از کارگاه‌های صنعتی ۱۰ کارکن و بیشتر، عوامل احتمالی تاثیرگذار بر وقوع حادثه ناشی از کار شناسایی و به عنوان متغیرهای پیشگو احتمالی مورد بررسی قرار گرفتند. این متغیرهای پیشگوی احتمالی عبارتند از: وضعیت آموزش ایمنی، وجود واحد بهداشت و ایمنی در کارگاه، نوع فعالیت صنعتی (مانند صنایع غذایی یا تولید مواد شیمیایی)، تعداد شاغلان غیرتولیدی، تعداد کارگران ساده، تعداد کارگران ماهر، تعداد تکنسین‌ها، تعداد مهندسان، متوسط شیفت‌های کاری روزانه، متوسط تعداد روز کاری در ماه، پرداختی کارهای کنتراتی، هزینه تعمیرات ماشین‌آلات اداری، هزینه تعمیرات جزئی (مجموع هزینه‌های پرداختی بابت تعمیرات جزئی ساختمان و تعمیرات جزئی ماشین‌آلات) و هزینه‌های تعمیرات اساسی (مجموع هزینه تعمیرات اموال سرمایه‌ای که توسط شاغلان کارگاه یا دیگران و هزینه پرداختی بابت خرید لوازم جهت تعمیر). سپس با استفاده از نرم‌افزار R سه مدل رگرسیون خطی حداقل مربعات (OLS)، رگرسیون ریج و رگرسیون لسو برآزش داده شدند که معیار MSE آنها به ترتیب غیرقابل محاسبه، ۲۳۴۲،۲۷۵ و ۲۶۹۳،۲۳۸ حاصل گردیده‌است. عدم امکان محاسبه MSE برای رگرسیون خطی حداقل مربعات وجود هم خطی بین متغیرهای پیشگو می‌باشد. از اینرو مدل‌های رگرسیونی ریج و لسو مناسب‌تر هستند. لذا به منظور بررسی دقیق‌تر بین دو رگرسیون مذکور و با توجه به این که MSE هر دو روش ریج و لسو تقریباً نزدیک به هم هستند لذا از طریق نمودار به شناسایی مدل مناسب پرداخته می‌شود.



نمودار (۱): نحوه میل به صفر برآورد ضرایب بر اساس تابع تاوان در مدل‌های رگرسیونی مختلف

با توجه به نمودار بالا، مقدار $\alpha = 1$ مناسب‌تر می‌باشد که مربوط به روش لسو است لذا مدل رگرسیون لسو برای داده‌های مورد نظر به شرح زیر است که بیانگر عوامل موثر بر حوادث ناشی از کار در کارگاه‌های صنعتی ۱۰ کارکن و بیشتر هستند:

$$y = -16.675 + 4.160HU + 0.023SW1 + 0.39SW2 + 11.31WS$$

که در آن، HU: داشتن واحد بهداشت و ایمنی، SW1: کارگران ساده، SW2: کارگران ماهر و WS: متوسط شیفت کاری روزانه است.

مراجع

- [۱] طرح آمارگیری کارگاه‌های ۱۰ کارکن و بیشتر (۱۳۹۳)، مرکز آمار ایران.
- [۲] علیزاده، حسین (۱۳۹۴)، تخصیص دارایی بر اساس رخ داده‌های دمی، پایان‌نامه جهت اخذ درجه کارشناسی ارشد در رشته آمار ریاضی.
- [3] Tibshirani, Robert, (1996), Regression Shrinkage and Selection via the Lasso, *Journal of the Royal Statistical Society*, Volume 58, Issue 1, Pages 267-288.
- [4] Choi, Yunjin, Park, Rina, & Seo, Michael (2012), Lasso on Categorical Data.
- [5] Hastie, Trevor, Tibshirani, Robert, Wainwright, Martin (2015), *Statistical Learning with Sparsity the Lasso and Generalizations*, published by Taylor & Francis Group.

غربالگری متغیرها در داده‌های بزرگ با استفاده از همبستگی فاصله‌ای

محمد کاظمی^۱، داود شاهسونی^۲، محمد آرشی^۳

^۱ گروه آمار، دانشکده علوم ریاضی دانشگاه صنعتی شاهرود
m.kazemie64@yahoo.com

^۲ گروه آمار، دانشکده علوم ریاضی دانشگاه صنعتی شاهرود
dshahsavani@shahroodut.ac.ir

^۳ گروه آمار، دانشکده علوم ریاضی دانشگاه صنعتی شاهرود
m_arashi_stat@yahoo.com

چکیده مختصر

در تحلیل داده‌های با بعد بسیار بالا، شناسایی متغیرهای پیشگوی مؤثر بر پاسخ، از اهمیت بسزایی برخوردار است. در این نوع داده‌ها، ابتدا با استفاده از یک روش غربالگری تعداد متغیرهای پیشگو را کاهش داده و سپس از روش‌های انتخاب متغیر مبتنی بر جریمه برای انتخاب مدل نهایی استفاده می‌کنیم. در این مقاله، برای غربالگری متغیرها از ضریب همبستگی فاصله‌ای بین متغیرهای پیشگو و تابع توزیع حاشیه‌ای متغیر پاسخ استفاده می‌کنیم. این روش نسبت به تشخیص نادرست ساختار مدل نیرومند بوده و همچنین زمانی که متغیر پاسخ دم سنگین یا دارای مقادیر فرین باشد، عملکرد خوبی را از خود نشان می‌دهد. کارایی این روش را با یک مثال شبیه‌سازی نشان می‌دهیم.

کلمات کلیدی

انتخاب متغیر، داده‌های با بعد بسیار بالا، رگرسیون جریمه شده، غربالگری.

۱- مقدمه

با پیشرفت توان محاسباتی کامپیوترها و سایر فناوری‌های نوین، محققین قادرند با هزینه نسبتاً کمی، داده‌هایی با حجم بسیار بالا در اختیار داشته باشند. در این نوع داده‌ها تعداد متغیرهای پیشگو p ، بسیار بزرگتر از تعداد مشاهدات n ، است. تحلیل این داده‌ها چالش‌ها و فرصت‌های جدیدی را پیش روی آماردانان قرار داده است. وقتی تعداد متغیرهای پیشگو در مدل زیاد باشد، تفسیر مدل مشکل‌تر و هزینه محاسبات افزایش می‌یابد. لذا، شناسایی متغیرهای پیشگوی مؤثر بر پاسخ بسیار مهم است. تاکنون روش‌های متعددی برای انتخاب متغیر در داده‌های با بعد بالا و در قالب روش‌های جریمه معرفی شده است. از جمله این روش‌ها می‌توان LASSO (تیبشیرانی ۱۹۹۶) و SCAD (فن و لی ۲۰۰۱) را نام برد. این روش‌ها برای انتخاب متغیر در داده‌های با بعد بالا که p بزرگتر از n است، قابل استفاده می‌باشد اما برای داده‌های با بعد بسیار بالا که p بسیار بزرگتر از n است، با چالش مواجه می‌شوند. برای رفع این مشکل، فن و لیو (۲۰۰۸) استفاده از یک روش غربالگری موسوم به SIS را برای کاهش بعد متغیرها در یک مدل خطی پیشنهاد دادند. آنها از ضریب همبستگی پیرسن حاشیه‌ای متغیرهای پیشگو با متغیر پاسخ به منظور غربالگری استفاده کردند و تعداد متغیرهای پیشگو را از بعد بسیار بزرگ p به بعد نسبتاً متوسط d کاهش داده و سپس از روش‌های انتخاب متغیر فوق برای انتخاب مدل نهایی استفاده کردند. پس از آن، روش SIS به مدل‌های آماری مختلف مانند مدل خطی تعمیم یافته، مدل جمعی ناپارامتری و مدل ضریب متغیر تعمیم داده شد. این روش‌ها مدل مبنا بوده و در صورت نادرست بودن مدل آماری مفروض، ممکن است دارای عملکرد مناسبی نباشند. بنابراین برخی از محققین برای اجتناب از تشخیص نادرست ساختار مدل، روش‌های غربالگری آزاد-مدل را معرفی کردند. در این روش‌ها برای غربالگری متغیرهای پیشگو نیازی به مشخص کردن ساختار مدل نیست.

یکی از این روش‌ها توسط لی و همکاران (۲۰۱۲) بر اساس ضریب همبستگی فاصله‌ای بین متغیرهای پیشگو و متغیر پاسخ معرفی شد، یعنی ابتدا متغیرهای پیشگو را براساس همبستگی فاصله‌ای آنها با متغیر پاسخ رتبه بندی کرده و سپس تعدادی از متغیرهای پیشگو با بیشترین همبستگی فاصله‌ای را برای ورود به مدل انتخاب کردند. در این مقاله، ما برای رتبه بندی متغیرها از ضریب همبستگی فاصله‌ای متغیرهای پیشگو با تابع توزیع حاشیه‌ای متغیر پاسخ استفاده می‌کنیم. این روش تعمیمی از روش لی و همکاران (۲۰۱۲) می‌باشد که نسبت به تشخیص نادرست ساختار مدل نیرومند بوده و همچنین زمانی که متغیر پاسخ دم سنگین یا دارای مقادیر فرین باشد، عملکرد خوبی را در انتخاب متغیرهای مهم نشان می‌دهد. در ادامه کارایی این روش را با استفاده از یک مثال شبیه‌سازی بررسی می‌کنیم.

۲- غربالگری مستقل نیرومند بر اساس همبستگی فاصله‌ای

در این بخش یک روش غربالگری نیرومند با استفاده از همبستگی فاصله‌ای ارائه می‌دهیم. کوواریانس فاصله‌ای بین متغیرهای تصادفی X و Y به صورت زیر تعریف می‌شود:

$$d \operatorname{cov}^2(X, Y) = S_1 + S_2 - 2S_3$$

که

$$S_3 = E\left\{E\left(\left|X - \tilde{X}\right| \mid X\right)E\left(\left|Y - Y\right| \mid Y\right)\right\}, S_2 = E\left\{\left|X - X\right|\right\}E\left\{\left|Y - Y\right|\right\}, S_1 = E\left\{\left|X - X\right|\left|Y - Y\right|\right\}$$

و (\tilde{X}, \tilde{Y}) متغیرهای مستقل و هم‌توزیع با (X, Y) می‌باشند. لذا همبستگی فاصله‌ای بین Y, X به صورت زیر تعریف می‌شود:

$$d \operatorname{corr}(X, Y) = \frac{d \operatorname{cov}(X, Y)}{\sqrt{d \operatorname{cov}(X, X)}\sqrt{d \operatorname{cov}(Y, Y)}}. \quad (1)$$

زکلی^۲ و همکاران (۲۰۰۷) نشان دادند که $d \operatorname{corr}(X, Y) = 0$ اگر و تنها اگر Y, X مستقل باشند و $d \operatorname{corr}(X, Y)$ یک تابع اکیدا صعودی از قدر مطلق ضریب همبستگی پیرسن بین Y, X است. به دلیل وجود این دو ویژگی، لی و همکاران (۲۰۱۲) یک روش غربالگری مستقل مطمئن موسوم به DC-SIS برای رتبه بندی متغیرها براساس همبستگی فاصله‌ای آنها با متغیر پاسخ ارائه دادند. ما برای رتبه بندی متغیرها از مربع ضریب همبستگی فاصله‌ای متغیرهای پیشگو با تابع توزیع حاشیه‌ای متغیر پاسخ، یعنی

$$\omega_k = d \operatorname{corr}^2(X_k, F(Y)), \quad k=1, \dots, p \quad (2)$$

استفاده می‌کنیم که $F(Y) = P(Y \leq y)$. این معیار چند برتری نسبت به روش‌های موجود دارد: الف) $d \operatorname{corr}(X, Y) = 0$ اگر و فقط اگر Y, X مستقل نباشند، بنابراین غربالگری براساس (۲) یک روش آزاد-مدل بوده و برای هر مدلی قابل استفاده است. ب) چون $F(Y)$ برای هر نوع متغیر پاسخ یک تابع کراندار است، می‌توانیم انتظار داشته باشیم که این روش غربالگری برای متغیر پاسخ دم سنگین یا دارای مقادیر فرین عملکرد مناسبی را از خود نشان دهد.

برای پیاده سازی این روش غربالگری، کافی است معیار مطلوبیت حاشیه‌ای (۲) را براساس یک نمونه تصادفی برآورد کنیم. فرض کنید $\{(X_i, Y_i), i = 1, \dots, n\}$ یک نمونه تصادفی از جامعه (X, Y) باشند. ابتدا کوواریانس فاصله‌ای بین X_k و $F(Y)$ را با روش گشتاوری برآورد می‌کنیم:

$$d \operatorname{cov}^2(X_k, F(Y)) = \hat{S}_{k,1} + \hat{S}_{k,2} - 2\hat{S}_{k,3},$$

که

$$\hat{S}_{k,1} = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n |X_{ik} - X_{jk}| |F_n(Y_i) - F_n(Y_j)|,$$

$$\hat{S}_{k,2} = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n |X_{ik} - X_{jk}| \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n |F_n(Y_i) - F_n(Y_j)|,$$

$$\hat{S}_{k,3} = \frac{1}{n^3} \sum_{i=1}^n \sum_{j=1}^n \sum_{l=1}^n |X_{ik} - X_{lk}| |F_n(Y_j) - F_n(Y_l)|,$$

برآوردهای $S_{k,3}, S_{k,2}, S_{k,1}$ و $F_n(y) = \frac{1}{n} \sum_{i=1}^n I(Y_i \leq y)$ تابع توزیع تجربی Y می‌باشند. در نتیجه معیار ω_k به صورت

$$\hat{\omega}_k = dcorr^2(X_k, F(Y)) = \left(\frac{d \operatorname{cov}(X_k, F(Y))}{\sqrt{d \operatorname{cov}(X_k, X_k)} \sqrt{d \operatorname{cov}(F(Y), F(Y))}} \right)^2 \quad (3)$$

برآورد می‌شود. سپس تعدادی از متغیرهای پیشگو که دارای مقادیر ω_k بزرگتر از یک مقدار مشخصی باشند، به عنوان مجموعه متغیرهای مهم انتخاب می‌شوند، یعنی مجموعه $\hat{A} = \{k = 1, \dots, p : \hat{\omega}_k > \nu\}$ که ν مقدار مثبت از پیش تعیین شده است. در عمل انتخاب ν یک چالش است. فن و لیو (۲۰۰۸) به جای استفاده از ν ، انتخاب d متغیر با بزرگترین مقدار ω_k را پیشنهاد دادند.

۳- مثال شبیه‌سازی

در این بخش با استفاده از شبیه‌سازی، عملکرد این روش غربالگری را با SIS و DC-SIS مقایسه می‌کنیم. به منظور سادگی، روش ارائه شده خود را با ROSIS نشان می‌دهیم. در اینجا فرض می‌کنیم تعداد متغیرهای پیشگو، $p = 1000$ ، و حجم نمونه، $n = 200$ ، باشد. دو مدل زیر را در نظر بگیرید:

$$\text{مدل ۱: } Y = X_1 + 2X_2^2 + 3X_3^3 + \varepsilon \quad \text{مدل ۲: } Y = \{1 + \exp(-3X^T \beta)\}^{-1} \varepsilon$$

که $\beta = (3, 1.5, 0, 0, 0, 0, 2, 0, \dots, 0)$ و بردار متغیرهای $X = (X_1, \dots, X_p)^T$ از توزیع $N(0, \Sigma)$ تولید شده‌اند که $\Sigma = (\sigma_{ij})_{p \times p}$ با مولفه‌های $\sigma_{ij} = 0.5^{|i-j|}$ است. همچنین خطای تصادفی ε از دو توزیع نرمال استاندارد و توزیع تی استودنت با یک درجه آزادی تولید شده است. همانطور که می‌بینید، در مدل ۱ تنها سه متغیر X_3, X_2, X_1 و در مدل ۲ متغیرهای X_7, X_2, X_1 واقعا مهم هستند و بقیه ۹۹۷ متغیر دیگر بی اهمیت هستند.

برای مقایسه، هر آزمایش را ۲۰۰ بار تکرار می‌کنیم و سپس عملکرد هر روش را با معیارهای M, P_j و S ارزیابی می‌کنیم که M حداقل اندازه مدل برای در برگرفتن تمام متغیرهای مهم، P_j احتمال تجربی انتخاب متغیر مهم X_j و S احتمال انتخاب تمام متغیرهای مهم برای اندازه مدل داده شده $d = 2 \lceil \frac{n}{\log(n)} \rceil$ است. برای مقایسه دقیق‌تر چندک‌های M را ارائه می‌کنیم. نتایج شبیه‌سازی در جدول ۱ خلاصه شده است.

جدول ۱: نتایج شبیه‌سازی

Model	Error	Method	M					P _j			S
			5%	25%	50%	75%	95%	P ₁	P ₂	P ₃	
Model 1	N(0,1)	SIS	3	4	4	5	12	0.99	1.00	1.00	0.99
		DC-SIS	3	4	4	4	5	1.00	1.00	1.00	1.00
		ROSI	3	3	3	4	4	1.00	1.00	1.00	1.00
	t ₁	SIS	4	7	53	475	935	0.48	0.87	0.98	0.47
		DC-SIS	3	4	4	7	145	0.68	0.94	1.00	0.86
		ROSI	3	3	4	4	8	0.81	0.98	1.00	0.98
Model 2	N(0,1)	SIS	359	605	771	925	983	0.08	0.06	0.08	0.00
		DC-SIS	3	3	4	8	32	1.00	1.00	0.99	0.99
		ROSI	3	5	8	6	44	1.00	1.00	0.98	0.98
	t ₁	SIS	316	607	783	882	981	0.09	0.08	0.09	0.00
		DC-SIS	4	16	66	182	576	0.90	0.86	0.58	0.54
		ROSI	3	5	9	21	64	1.00	1.00	0.96	0.97

مقادیر کوچک M در جدول نشان دهنده عملکرد بهتر است و در بهترین حالت برابر با تعداد متغیرهای مهم است. همچنین در یک روش غربالگری مناسب مقادیر S, P_j باید نزدیک به یک باشند. با توجه به جدول می‌بینیم که در هر دو مدل، SIS برای خطای دم سنگین‌تری استوندت خوب عمل نمی‌کند. حتی زمانی که توزیع خطاها نرمال است، SIS در شناسایی متغیرهای مهم در مدل‌های غیر خطی فوق‌ناتوان است. زمانی که ϵ دارای توزیع نرمال است، DC-SIS و ROSIS رفتار مشابهی دارند، اما برای ϵ غیر نرمال ROSIS خیلی بهتر از DC-SIS عمل می‌کند.

۴- نتیجه‌گیری

در این مقاله یک روش نیرومند برای غربالگری متغیرها در داده‌های بزرگ براساس همبستگی فاصله‌ای بین متغیرهای پیشگو و تابع توزیع حاشیه‌ای متغیر پاسخ ارائه شد. با در نظر گرفتن دو مدل آماری مختلف و جمله خطای جمعی و ضربی نشان دادیم که این روش نسبت به ساختار مدل نیرومند است. همچنین موقعی که خطا دارای توزیع نرمال است، عملکرد این روش و DC-SIS تقریباً یکسان است، اما زمانی که خطا دارای توزیع دم سنگین‌تری استوندت است، این روش نسبت به SIS و DC-SIS عملکرد بهتری را نشان می‌دهد.

مراجع

- [1] Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, 96, 1348-1360.
- [2] Fan, J. and Lv, J. (2008). Sure independence screening for ultrahigh dimensional feature space (with discussion). *Journal of the Royal Statistical Society*, B, 70, 849-911.
- [3] Li, R., Zhong, W., and Zhu, L. (2012). Feature screening via distance correlation learning. *Journal of the American Statistical Association*, 107, 1129-1139.
- [4] Szekely, G. J., Rizzo, M. L., and Bakirov, N. K. (2007). Measuring and testing dependence by correlation of distances. *Annals of Statistics*, 35, 2769-2794.
- [5] Tibshirani, R. (1996). Regression shrinkage and selection via LASSO. *Journal of the Royal Statistical Society*, B, 58, 267-288.

قرارداد بیمه‌ای چندلایه‌ای بهینه بر اساس تابع زیان اعوجاج

علی پناهی بزاز^۱، امیر تیمور پاینده نجف آبادی^۲

^۱ گروه آماردانان بدون مرز، موسسه ناب، دزفول،

al_panahi@sbu.ac.ir, alipanahiazaz@yahoo.com

^۲ هیات علمی گروه آمار، دانشگاه شهید بهشتی، تهران

amirtpayandeh@sbu.ac.ir

چکیده مختصر

در این مقاله قصد داریم تا یکی از ابزارهای بسیار قدرتمند و کاربردی را در طراحی قراردادهای اتکایی بهینه، به نام تابع غرامت حاشیه‌ای، را معرفی کنیم. سپس نحوه طراحی قرارداد اتکایی بهینه با استفاده از تابع غرامت حاشیه‌ای بر اساس مینیمم کردن مقدار در معرض مخاطره و مقدار در معرض مخاطره شرطی توضیح داده می‌شود. در ادامه چگونگی تعمیم قراردادهای طراحی شده به قراردادهای چند لایه‌ای را توضیح داده و الگوریتمی را برای این تعمیم معرفی می‌کنیم. در نهایت قراردادهای بهینه طراحی شده بر اساس مینیمم کردن اندازه مقدار در معرض مخاطره و مقدار در معرض مخاطره شرطی را با استفاده از الگوریتم معرفی شده به قراردادهای چند لایه‌ای تعمیم می‌دهیم.

کلمات کلیدی

قرارداد اتکایی بهینه، قرارداد اتکایی چند لایه‌ای، تابع غرامت حاشیه‌ای، مقدار در معرض مخاطره، مقدار در معرض مخاطره شرطی.

۱- مقدمه

پورس و همکاران (۲۰۱۳) با استفاده از مدل بیمه‌ای اتکایی تجربی که در ونگ (۲۰۰۹) توضیح داده شده، نشان دادند که برای حق بیمه‌ای اتکایی انحراف استاندارد، بیمه‌ای اتکایی یک لایه‌ای زیان‌بس مینیمم‌کننده‌ی امید ریاضی دم شرطی خسارت کلی بیمه‌گر اولیه در بیمه‌ی محصولات کشاورزی است. اوپانگ و لی (۲۰۱۰) یک بیمه‌ی اتکایی چند لایه‌ای را برای دستیابی به توسعه‌ی پایدار در بیمه‌های محصولات کشاورزی در حضور خطرات ذهنی و اخلاقی طراحی کردند. ددو (۲۰۱۲) دو قرارداد اتکایی زیان‌بس و نسبتی را تعمیم داده و یک بیمه‌ی اتکایی چند لایه‌ای را معرفی کرد. کورتس و همکارا (۲۰۱۳) یک بیمه‌ی اتکایی چند لایه‌ای را با تعداد لایه‌های ثابت طراحی کردند. سپس همانند دیگر پژوهشگران برای تعیین دقیق این بیمه‌ی اتکایی مقدار ریسک خاصی را با شرط ثابت بودن حق بیمه اتکایی مینیمم کردند. چی و تن (۲۰۱۳) نشان دادند که بیمه‌ی اتکایی زیان‌بس یک لایه‌ای بیمه‌ی اتکایی بهینه‌ای است که مینیمم‌کننده‌ی مقدار در معرض ریسک و معیار در معرض ریسک شرطی خسارت کلی بیمه‌گر اولیه در کلاس خاصی از بیمه‌های اتکایی محذب هستند. پژوهشگران دیگری نظیر پاسالاکو (۲۰۰۷) و ونگ (۲۰۱۴) نیز تلاشهایی را برای طراحی قراردادهای اتکایی چندلایه‌ای انجام دادند.

۲- تعاریف و مفاهیم مورد نیاز

مقدار در معرض ریسک (Var) متغیر تصادفی نامنفی X ، یا همان خسارت تصادفی، در سطح اطمینان $1 - \alpha$ ، که $0 < \alpha < 1$ ، به صورت زیر تعریف می‌شود،

$$VaR_{\alpha}(X) = \inf\{x \in \mathbb{R} : P(X \leq x) \geq 1 - \alpha\}.$$

مقدار در معرض ریسک شرطی ($CVaR$) متغیر تصادفی نامنفی X ، یا همان خسارت تصادفی، در سطح اطمینان $1 - \alpha$ ، که $0 < \alpha < 1$ ، به صورت زیر تعریف می‌شود

$$CVaR_{\alpha}(X) = \frac{1}{1-\alpha} \int_{\alpha}^1 VaR_q(X) dq$$

تابع پرداخت غرامت حاشیه‌ای (MIF) یکی از ابزارهای جدیدی است که برای حل مسأله‌ی طراحی قراردادهای اتکایی بهینه از آن استفاده می‌شود. این روش به طور دقیق و کامل در مقاله آسا (۲۰۱۵) مطرح شده و با استفاده از آن قرارداد اتکایی بهینه برای مدل مطرح شده در کو و همکاران (۲۰۱۳) البته بدون قید حق بیمه، طراحی شد. حال به معرفی تابع MIF پرداخته و قضایای مورد نیاز ارائه می‌شود. در ابتدا رده‌ی قراردادهای C_{MIF} به صورت زیر تعریف می‌شود:

$$C_{MIF} = \left\{ f : R_+ \rightarrow R_+ \mid f(x) = \int_0^x h(t) dt, 0 \leq h \leq 1 \right\}.$$

فضای توابع پرداخت غرامت به صورت زیر تعریف می‌شود

$$D_{MIF} = \{ h : R_+ \rightarrow R_+ \mid 0 \leq h \leq 1 \}$$

تعریف ۱: برای هر تابع پرداخت غرامت $f \in C_{MIF}$ ، یک تابع پرداخت غرامت حاشیه‌ای مانند $h \in D_{MIF}$ موجود است به طوری که

$$f(x) = \int_0^x h(t) dt, \quad x \geq 0$$

تعریف ۲: فرض کنید g یک تابع غیرکاهشی حقیقی مقدار باشد که از بازه‌ی $[0, 1]$ به بازه‌ی $[0, 1]$ تعریف شده است به طوری که $g(0) = 1 - g(1) = 0$. حق بیمه‌ی اعوجاج به صورت

$$\pi(\theta) = (1 + \theta)\rho(X)$$

تعریف می‌شود که $\theta \geq 0$ و $\rho(X) = \int_0^{\infty} g(\bar{F}_X(t)) dt$

۳- نتایج و مثالها

در این قسمت الگوریتمی را بیان می‌کنیم که قرارداد اتکایی بهینه چند لایه ای با استفاده از آن طراحی می‌شود.

الگوریتم ۱: فرض کنید X_R سهم بیمه‌گر اتکایی از خسارت تصادفی X باشد. با استفاده از مراحل زیر می‌توان یک قرارداد اتکایی چندلایه‌ای که مینیمم‌کننده‌ی CTE_{α} خسارت کلی بیمه‌گر اولیه است طراحی کرد.

گام اول: قرارداد بیمه‌ی اتکایی چندلایه‌ای را با الگوریتم تکرار زیر می‌توان طراحی کرد:

مرحله‌ی اول: برای $k \geq 2$ بازه‌ی $[M_k, +\infty)$ به دو زیربازه‌ی $[M_k, M_{k+1})$ و $[M_{k+1}, +\infty)$ افراز می‌شود و سهم بیمه‌گر اتکایی از خسارت تصادفی X به صورت:

$$f_k(X) = f_{k-1}(X) I_{[0, M_k]}(X) + [f_{k-1}(M_k) + f(X - M_k)] I_{[M_k, \infty)}(X),$$

است که در آن $f_0(X) = f(X) = \max\{X - d_{\alpha}, 0\}$ ؛

مرحله‌ی دوم: در صورت مشاهده‌ی معیار توقف به گام دوم رفته و در غیر این صورت مقدار $k = k + 1$ قرار داده و به مرحله‌ی اول برمی‌گردیم.

گام دوم

مرحله‌ی اول: سهم بیمه‌گر اتکایی در قرارداد بیمه‌ی اتکایی k -لایه‌ای به صورت:

$$X_R = f(X) I_{[0, M_1]}(X) + \sum_{j=1}^{k-1} f_j(X) I_{[M_j, M_{j+1})}(X) + [f_{k-1}(M_k) + f(X - M_k)] I_{[M_k, \infty)}(X)$$

محاسبه می‌شود.

مرحله‌ی دوم) در انتها پارامترهای مجهول تحت یک معیار مناسب، یا یک روش برآورد مناسب، با در نظر گرفتن شرط $E(\max\{X - d_{\alpha}, 0\}) = E(X_R)$ برآورد می‌شوند.

حال با استفاده از دو مثال دو قرارداد اتکایی چند لایه‌ای با استفاده از الگوریتم (۱) طراحی می‌شود. لازم است به این نکته توجه شود که با توجه به گزاره‌های (۳) و (۴) مقاله پاینده و پناهی (۲۰۱۷)، و با توجه به انتقال ناوردا بودن اندازه‌های ریسک مقدار در معرض ریسک و مقدار در معرض ریسک شرطی، قراردادهای چند لایه‌ای طراحی شده نیز بهینه می‌باشند. مثال ۱: در این مثال یک قرارداد اتکایی دو لایه‌ای بهینه طراحی می‌شود. برای این منظور قرارداد:

$$f(X) = \begin{cases} 0 & X \leq a^* \\ X - a^* & a^* < X < a^* + L \\ L & a^* + L \leq X < M_1 \\ f^*(X) & M_1 \leq X \end{cases}$$

حال با استفاده از الگوریتم (۱) و خاصیت انتقال ناوردا بودن اندازه مقدار در معرض ریسک، و در حقیقت با انتقال مبداء مختصات به نقطه‌ی $(M, a^* + L)$ ، قرارداد اتکایی دو لایه‌ای به صورت:

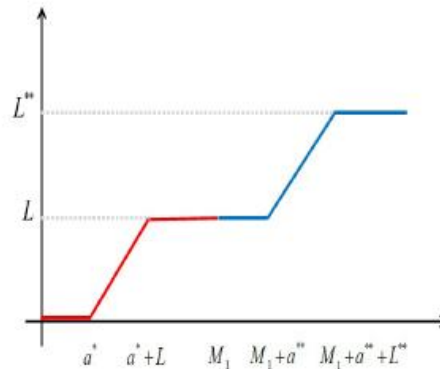
$$X_R^{2-layer} = \begin{cases} 0 & X \leq a^* \\ X - a^* & a^* < X < a^* + L \\ L & a^* + L \leq X < M_1 \\ L & M_1 \leq X \leq M_1 + a^{**} \\ X + L - M_1 - a^{**} & M_1 + a^{**} < X < M_1 + a^{**} + L^{**} \\ L + L^{**} & M_1 + a^{**} + L^{**} \leq X \end{cases}$$

نتیجه می‌شود که در آن a^{**} از معادله‌ی زیر به دست می‌آید و $L^{**} = F_X'^{-1}(\alpha) - a^{**}$

$$\Pi_2(F_X'^{-1}(a^{**})) = \frac{a_2 - 1}{a_2}$$

و $F_X'(t) = \frac{F_X(t) - F_X(M_1)}{1 - F_X(M_1)}$ برای $x \geq M_1$. سطوح سوم و چهارم قرارداد اتکایی طراحی شده در بالا را می‌توان با هم ادغام

کرد، ولی به خاطر روشن بودن روش از انجام این کار اجتناب می‌شود. نحوه طراحی قرارداد اتکایی بالا را می‌توان در شکل زیر مشاهده کرد.



تمودار مربوط به قرارداد اتکایی طراحی شده با استفاده از تابع پرداخت غرامت حاشیه‌ای (MIF) و بر اساس مینیم سازی اندازه‌ی مقدار در معرض ریسک.

۴- نتیجه گیری

با استفاده از الگوریتم (۱) و گزاره‌های (۳) و (۴) مقاله پاینده و پناهی (۲۰۱۷) میتوان قرارداد اتکایی چند لایه ای بهینه بر اساس قراردادهای اتکایی بهینه طراحی شده با استفاده از تابع غرامت حاشیه ای طراحی کرد که علاوه بر دارا بودن خاصیت بهینگی قراردادهای مذکور میتوانند از نظر شاخصهای مهم دیگر از قرارداد اولیه بهتر باشند.

مراجع

- [1] Assa, H. (2015). On Optimal Reinsurance Policy with Distortion Risk Measures and Premiums. *Insurance: Mathematics and Economics*, 61, 70-75.
- [2] Bazaz, A. P., & Payandeh Najafabadi, A. T. (2015). An Optimal Reinsurance Contract from Insurer's and Reinsurer's Viewpoints. *Applications & Applied Mathematics*, 10(2).
- [3] Chi, Y. (2012). Reinsurance Arrangements Minimizing the Risk-Adjusted Value of an Insurer's Liability. *Astin Bulletin*, 42(02), 529-557.
- [4] Cai, J., Fang, Y., Li, Z., & Willmot, G. E. (2013). Optimal Reciprocal Reinsurance Treaties under the Joint Survival Probability and the Joint Profitable Probability. *Journal of Risk and Insurance*, 80, 145-168.
- [5] Chi, Y., Tan, K. S. (2013). Optimal reinsurance with general premium principles. *Insurance: Mathematics and Economics*, 52(2), 180-189.
- [6] Hesselager, O. (1990). Some Results on Optimal Reinsurance in Terms of the Adjustment Coefficient. *Scandinavian Actuarial Journal*. 1990, 80-95.
- [7] Passalacqua, L. (2007). Measuring effects of excess-of-loss reinsurance on credit insurance risk capital. *Giornale dell'Istituto Italiano degli Attuari*, LXX, 81-102.
- [8] Payandeh Najafabadi, A. T. & Panahi Bazaz, A. (2017). An Optimal Multi-layer Reinsurance Policy under Conditional Tail Expectation. *Annals of actuarial science*, 2017(1).
- [9] Tan, K. S., Weng, C., Zhang, Y. (2011). Optimality of General Reinsurance Contracts under CTE Risk Measure. *Insurance: Mathematics and Economics*, 49, 175-187.

کاربرد متن‌کاوی در تعیین قیمت خودروهای کارکرده

سیدجمال میرکمالی

گروه ریاضی دانشگاه اراک، اراک
Mirkamali.sj@gmail.com

چکیده

بیش از هفتاد سال از مقاله مشهور استیونز می‌گذرد که در آن انواع داده به چهار دسته اسمی، ترتیبی، فاصله‌ای، و نسبتی تقسیم‌بندی شد. بر اساس این دسته‌بندی و با توجه به جایگاه و تعداد متغیرهای مورد بررسی انواعی از ابزارهای تحلیل آماری ارائه شد که امکان استنباط از این انواع داده را فراهم کردند. اگرچه این دسته‌بندی هنوز هم پایه بسیاری از تحلیل‌های آماری است، معرفی دسته‌های جدیدی از داده و روش‌های تحلیل آن‌ها نیز ضروری به نظر می‌رسد. داده‌هایی از نوع متن، تصویر، ویدیو، صوت، ایموجی، و... از جمله انواع داده‌ای هستند که در سال‌های اخیر به طور وسیع توسط کاربران اینترنت تولید و در مراکز داده ذخیره شده‌اند.

در این بین داده‌های متنی به دلیل اینکه در جامعه کاربران اینترنت و به طور خاص در شبکه‌های اجتماعی فراگیرتر است از اهمیت بسیار بالایی برخوردار است. به همین دلیل توسعه ابزارهای تحلیل داده‌های متنی ضروری به نظر می‌رسد. تا کنون فعالیت‌های قابل توجهی در این زمینه صورت گرفته است به طوری که در حال حاضر متن‌کاوی به یک میان رشته بین تخصص‌های آمار، زبان‌شناسی، و یادگیری ماشین تبدیل شده است. علی‌رغم پیشرفت‌هایی که در این حوزه صورت گرفته است، ابزارهای تحلیل متون فارسی چندان توسعه نیافته است و نیازمند توجه بیشتر از پیش پژوهشگران به این بخش است.

در سال‌های اخیر سایت‌ها و شبکه‌های اجتماعی فارسی زبان در بخش‌های اقتصادی بخصوص بازار کالاهای دست دوم فعالیت قابل توجهی داشته‌اند. از جمله این سایت‌ها می‌توان به سایت‌های شیپور، دیوار، و باما اشاره کرد. در این مقاله، به عنوان یک مطالعه موردی، تجزیه و تحلیل محتواهای متنی سایت باما را برای قیمت‌گذاری خودروهای کارکرده در نظر گرفته‌ایم. برای این منظور ابزارها و روش‌های استخراج داده‌های متنی را توسعه داده و سپس با استفاده از مدل‌های آماری داده‌های استخراج شده را مورد تحلیل و ارزیابی قرار داده‌ایم و نتایج حاصل از آن را ارائه کرده‌ایم.

کلمات کلیدی

متن‌کاوی، قیمت‌گذاری، سایت باما، درخت رگرسیون، وب‌کاوی، تحلیل مولفه‌های اصلی

کاربرد نمونه‌گیری معکوس در صنعت

خاطره زارعی^۱، بردیا پناه‌بغ^۲

^۱دانشکده ریاضی و علوم کامپیوتر، دانشگاه خوارزمی، تهران،

Zarei.khatereh69@gmail.com

panahbehagh@khu.ac.ir

چکیده مختصر

در مسائل بررسی موقعیت‌های شغلی بنگاه‌های صنعتی برای گرفتن یک نمونه، لیستی از مشاغل را در نظر می‌گیرند و سپس در هر بنگاه مشاغل به صورت تصادفی ساده انتخاب می‌شوند تا زمانیکه r تا از مشاغل انتخاب شود. در عمل این همان نمونه‌گیری معکوس است که یکی از پرکاربردترین طرح‌های نمونه‌گیری است. برآوردگری که برای نمونه‌گیری‌های دنباله‌ای از جمله نمونه‌گیری معکوس استفاده می‌شود برآوردگر مورتی است که برآوردگری ناریب است ولی از کارایی بالایی برخوردار نیست. در این مقاله بر اساس تیله (۲۰۱۶) روشی برای تخمین احتمالات شمول اعضا معرفی می‌گردد که از طریق آن بتوان از برآوردگر کاراتر هارویتز-تامسون برای تخمین پارامتر مجهول جامعه استفاده کرد.

کلمات کلیدی

نمونه‌گیری معکوس، برآوردگر مورتی، برآوردگر هارویتز-تامسون، احتمالات شمول

۱- مقدمه

یکی از اولین طرح‌های نمونه‌گیری معکوس توسط هالدان (۱۹۴۵) معرفی شد که از نمونه‌گیری معکوس برای برآورد فراوانی پیشامدهای نادر استفاده کرد. چیکاگودار (۱۹۶۹) نمونه‌گیری معکوس بدون جایگذاری را بررسی کرد، سمپفورد (۱۹۶۲) طرح نمونه‌گیری معکوس با احتمال نابرابر بدون جایگذاری جریک و ناديو (۲۰۰۷) نمونه‌گیری معکوس با احتمال نابرابر با جایگذاری را معرفی کردند، در سال‌های اخیر صالحی و سیبر (۲۰۰۱) برآوردگر مورتی برای طرح‌های دنباله‌ای و در (۲۰۰۴) مشکل حجم نمونه را بیان کردند و پناه‌بغ (۲۰۱۶) روشی برای برای کنترل حجم نمونه بیان کرد.

در این مقاله سعی می‌شود بر اساس مشکل کاربردی موجود در نمونه‌گیری از مشاغل کشور کانادا (تیله ۲۰۱۶)، نمونه‌گیری معکوس را برای این مشکل سازگار کنیم. برای این منظور هریک از مشاغل ممکن است با استفاده از یک طرح ساده با جایگذاری یا بدون جایگذاری و یا با احتمالات نابرابر بر اساس نمونه‌گیری معکوس انتخاب شود. اما هدف اصلی برآورد میانگین دستمزد برای یک شغل در کل جامعه است که از برآوردگر هارویتز-تامسون استفاده می‌کنیم و برای این منظور مجبور به تخمین احتمالات شمول هستیم.

در فصل ۲، طرح نمونه‌گیری معکوس بر پایه طرح تصادفی ساده با جایگذاری و در فصل ۳ طرح نمونه‌گیری معکوس بر پایه طرح احتمالات نابرابر برای بررسی مشاغل بنگاه‌های صنعتی سازگار شده و احتمالات شمول برآورد می‌شوند. مقاله در فصل ۴ با یک جمع‌بندی به پایان می‌رسد.

نمادها

U : یک جامعه شامل N بنگاه یعنی $U = \{1, 2, \dots, i, \dots, N\}$ (جامعه‌ای از بنگاه‌ها در یک منطقه صنعتی را مشخص می‌کند).

L : لیست مشاغل

M : تعداد مشاغل در لیست یعنی اندازه‌ی L

F_i : لیست مشاغل در بنگاه i ، $F_i \subset L$

D_i : لیست مشاغلی که در بنگاه i وجود دارند، با $D_i \subset L$ ، $D_i \cap F_i = \emptyset$ ، $F_i \cup D_i = L$

MP_i : تعداد مشاغل در بنگاه i یعنی اندازه F_i

r : تعداد مشاغل متمایز بدست آمده در هر بنگاه (تعداد پیروزی‌ها)

X_i : تعداد شکست‌ها قبل از r شغل بدست آمده در بنگاه i

۲- نمونه‌گیری تصادفی ساده با جایگذاری

فرض کنید بنگاه i دارای نسبت p_i از مشاغل در لیست بنگاه است. اگر یک نمونه از مشاغل با جایگذاری از بنگاه i انتخاب شود تا زمانی که r شغل شناسایی شود پس X_i دارای توزیع دوجمله‌ای منفی است.

$$X_i \sim NB(r, p_i)$$

اگر $k \in L$ ، تعداد دفعات تکرار شغل k ام در نمونه گرفته شده از بنگاه i باشد، در یک طرح ساده با جایگذاری با اندازه n ، دارای توزیع چندجمله‌ای است. از این رو

$$p(A_{ik} = a_{ik}; k \in L) = \frac{n!}{M^n} \prod_{k \in L} \frac{1}{a_{ik}!}$$

که در آن $\sum_{k \in L} a_{ik} = n$ و $A_{ik} = 0, 1, 2, \dots, n$

اگر شرایط را روی قسمت مشخصی از جامعه محدود کنیم، توزیع چندجمله‌ای باقی می‌ماند پس:

$$p(A_{ik} = a_{ik}, k \in F_i | \sum_{k \in F_i} A_{ik} = r) = \frac{r!}{(MP_i)^r} \prod_{k \in F_i} \frac{1}{a_{ik}!}$$

با $\sum_{k \in F_i} a_{ik} = r$

در فرایندی که به روش با جایگذاری برای رسیدن به r شغل در بنگاه i انجام می‌دهیم، داریم:

$$E(A_{ik} | X_i) = \frac{r}{MP_i}; k \in F_i$$

$$\frac{X_i}{M - MP_i}; k \in D_i$$

که در نتیجه احتمال شمول $\pi_{k|i}$ به صورت زیر بدست می‌آید:

$$\pi_{k|i} = EE(A_{ik} | X_i) = \frac{r}{MP_i}; k \in F_i$$

که در آن p_i نامعلوم است و می‌توان با استفاده از روش گشتاورها و حل معادله $E(X_i) = X_i$ ، p_i را برآورد کرد:

$$\hat{p}_{i1} = \frac{r}{X_i + r}$$

که همان برآورد درست‌نمایی ماکزیمم پارامتر نیز است که برآوردگری اریب است.

اگر $r \geq 2$ برآوردگر واریانس مینیمم ناریب از p_i بصورت زیر است:

$$\hat{p}_{i2} = \frac{r-1}{X_i + r - 1}$$

اگرچه $\frac{1}{\hat{p}_{i1}}$ برای $\frac{1}{p_i}$ ناریب است.

حال احتمالات شمول تخمین زده شد و می‌توان از برآوردگر هارویتز-تامسون برای برآورد پارامتر مجهول استفاده کرد.

۳- نمونه‌گیری با احتمالات نابرابر با جایگذاری

حال اگر p_{ik} احتمال انتخاب یک شغل در هر مرحله باشد داریم:

$$\sum_{k \in L} p_{ik} = 1$$

اگر p_i مجموعی از p_{ik} های محدود شده از مشاغل در بنگاه i باشد:

$$p_i = \sum_{k \in F_i} p_{ik}$$

در این مورد مجدداً $X_i \sim NB(r, p_i)$ ، بنابراین اگر A_{ik} تعداد دفعاتی باشد که واحد k ام در نمونه تکرار شده است. در یک طرح با احتمال نابرابر با جایگذاری با اندازه n ، مقادیر A_{ik} دارای توزیع چندجمله‌ای است:

$$p(A_{ik} = a_{ik}, k \in L) = n! \prod_{k \in L} \frac{p_{ik}^{a_{ik}}}{a_{ik}!}$$

که در آن $A_{ik} = 0, \dots, n$ و $\sum_{k \in L} a_{ik} = n$.

اگر شرایط را روی یک اندازه ثابت در یک قسمت جامعه محدود کنیم، توزیع چند جمله‌ای باقی می‌ماند بطوریکه:

$$p(A_{ik} = a_{ik}, k \in F_i | \sum_{k \in F_i} A_{ik} = r) = r! \prod_{k \in F_i} \left(\frac{p_{ik}}{p_i} \right)^{a_{ik}} \frac{1}{a_{ik}!}$$

با $\sum_{k \in F_i} a_{ik} = r$

برای رسیدن به r شغل در بنگاه i به روش جایگذاری داریم:

$$E(A_{ik} | X_i) = \frac{r p_{ik}}{p_i}; k \in F_i$$

$$\frac{X_i p_{ik}}{1 - p_i}; k \in D_i$$

مقدار مورد انتظار از A_{ik} بصورت زیر است:

$$\pi_{ik} = EE(A_{ik} | X_i) = \frac{r p_{ik}}{p_i}$$

مجدداً مانند حالت قبل با استفاده از روش گشتاورها و حل معادله $E(X_i) = X_i$ میتوان p_i را برآورد کرد:

$$\hat{p}_{i1} = \frac{r}{X_i + r}$$

که همان برآورد درست‌نمایی ماکزیمم پارامتر است که برآوردگری اریب است.

برآوردگر واریانس مینیمم ناریب بصورت زیر است:

$$\hat{p}_{i2} = \frac{r-1}{X_i + r - 1}$$

حال احتمالات شمول تخمین زده شد و می‌توان از برآوردگر هارویتز-تامسون برای برآورد پارامتر مجهول استفاده کرد.

۴- جمع‌بندی

با روش معرفی شده در این مقاله می‌توان از نمونه‌گیری معکوس و همچنین برآوردگر کارای هارویترز-تامسون در مسائل مربوط به بررسی مشاغل بنگاه‌های صنعتی بهره برد. اگرچه مورد باجایگذاری خیلی رضایت‌بخش نیست، چون انتخاب r شغل باجایگذاری منجر به r شغل جداگانه نمی‌شود از طرفی اگر MP_i کوچک باشد ممکن است نمونه‌گیری طولانی شود بنابراین نمونه‌گیری بدون جایگذاری ترجیح داده می‌شود. برای طرح پایه نمونه‌گیری معکوس، دو طرح باجایگذاری و با احتمالات نابرابر معرفی شدند. از بیان روش بدون جایگذاری به دلیل خلاصه نویسی نگارش خودداری شد که این طرح پایه با کمی محاسبات بیشتر به نمونه‌گیری معکوس اضافه می‌شود. برای این امر می‌توان از چیکاگودار (۱۹۶۹) استفاده کرد.

مراجع

- [1] Tille, Y. (2016). Unequal probability inverse sampling, *Survey Methodology* 42(2).
- [2] Chikkagoudar, M.S. (1969). Inverse sampling without replacement. *Australian Journal of Statistics*, 11, 155-165.
- [3] Sampford, M.R. (1962). Methods of cluster sampling with and without replacement for clusters of unequal sizes. *Biometrika*, 49(1/2), 27-40.
- [4] Greco, L. and Naddeo, N. (2007). Inverse sampling with unequal selection probabilities. *Communications in Statistics, Theory and Methods*, 36, 1039-1048.
- [5] Salehi, M.M., and Seber, G.A.F. (2001). A new proof of Murthy's estimator which applies to sequential sampling. *The Australian and New Zealand Journal of Statistics*, 43, 281-286.
- [6] Salehi M.M. and Seber G.A.F. (2004). A general Inverse Sampling Scheme and its Application to Adaptive cluster sampling. *The Australian and New Zealand Journal of Statistics*, 46, 483-494.
- [7] Panahbehagh, B and Brown, J. (2016) Gap based inverse sampling. *Communications in Statistics, Theory and Methods*, Accepted to publish.

کاهش تعداد پارامترها در مدل‌سازی ذخایر ادعا

زهرا پرچیانلو^۱، محمد ذکایی^۲

^۱ گروه بیم‌سنجی، دانشگاه شهید بهشتی، تهران،
z.parchiyanloo@gmail.com

^۲ گروه بیم‌سنجی، دانشگاه شهید بهشتی، تهران
zokaei@sbu.ac.ir

چکیده مختصر

در این مقاله کاربرد جدیدی از روش پرش برگشت‌پذیر مونت کارلو (RJCMC) را برای کاهش تعداد پارامترها در مدل‌سازی الگوی توسعه ادعا در مثلث زیان‌بس بیان می‌کنیم. در این روش به صورت خودکار از یک منحنی نمایی برای به دست آوردن عامل دمی مثلث توسعه ادعا استفاده می‌کنیم، در حالی که، در روش‌های دیگر با استفاده از برون‌یابی الگوی توسعه ادعا، عامل دمی به دست آورده می‌شود. همچنین با استفاده از روش RJCMC علاوه بر کاهش تعداد پارامترها در مدل و به دست آوردن عامل دمی، توزیع پیش‌بین کامل ادعاهای معوق را به دست می‌آوریم. اندازه‌ای از عدم قطعیت در برآورد ذخایر ادعای معوق که در محاسبات توانگری مالی ۲ مورد نیاز است را ارائه می‌دهیم.

کلمات کلیدی

پرش برگشت‌پذیر مونت کارلو، ذخایر ادعای معوق، پواسون بیش پراکنش، مثلث توسعه ادعا، توانگری مالی

۱- روش‌های ذخیره‌سازی ادعا

قیمت محصولات بیمه‌ای بر خلاف سایر صنعت‌ها، هنگام بستن قرارداد یا حتی هنگامی که ادعای خسارت برای اولین بار گزارش می‌شود، مشخص نیست. در رشته‌های بیمه‌ای دم سبک (مانند بیمه بدنه‌ی اتومبیل و منازل مسکونی) که خسارات به سرعت تسویه می‌شوند، مقدار نهایی ادعا با سرعت بیشتری مشخص می‌شود. در مقابل، در رشته‌های بیمه‌ای دم‌سنگین (مانند مسئولیت عمومی تجاری، غرامت کارکنان یا بیمه جامع مسئولیت) ممکن است سال‌های زیادی طول بکشد تا مقدار نهایی خسارت مشخص شود. بزرگترین بدهی در شرکت‌های بیمه غیر عمر مربوط به ذخایر ادعای معوق پرداخت‌نشده است. به علت تأخیر در پرداخت و ماهیت تصادفی زمان و مقدار زیان پرداختی، علاوه بر اینکه محصولات بیمه‌ای از دیگر محصولات تجاری متمایز شده‌اند، برای مدیریت مالی و مخاطره در شرکت‌های بیمه، نیاز به قوانین مدیریتی متفاوتی نسبت به دیگر شرکت‌های تجاری است. بنابراین شرکت‌های بیمه موظف به ذخیره‌سازی سرمایه لازم برای پرداخت بدهی‌ها در صورت تحقق تعهداتشان هستند، که در آن نباید بیش از سرمایه لازم ذخیره‌سازی صورت بگیرد، به طوری که سهامداران متناسب با سرمایه خود در شرکت بیمه سود دریافت کنند. در واقع در مسئله ذخیره‌سازی ادعا تنها به دنبال تعیین متوسط مقدار ذخایر ادعا نیستیم، بلکه به دنبال تعیین توزیع این ذخایر نیز هستیم. بنابراین داشتن توزیع پیش‌بین مثلث پایین، علاوه بر مدیریت مخاطره، بررسی مخاطره موجود در برآورد ذخایر، می‌توان مقدار سرمایه مورد نیاز برای توانگری مالی ۲ را نیز به دست آورد.

تعیین ذخایر در شرکت‌های بیمه غیر عمر، بر اساس بررسی مثلث توسعه ادعا صورت می‌گیرد. در بیشتر موارد زیان‌های معوق با استفاده از مثلث توسعه ادعا برآورد می‌شود. نمایش معمول داده‌های مربوط به خسارت‌های پرداخت شده بر حسب سال حادثه و سال

توسعه، همان‌گونه که در جدول ۱ دیده می‌شود، به صورت مثلثی است که به آن مثلث توسعه ادعا می‌گویند. منظور از سال حادثه همان سال صدور بیمه نامه است که با i نشان می‌دهیم. سال تأخیر، نشان‌دهنده تعداد سال‌هایی است که تا پرداخت نهایی ادعا تأخیر افتاده است که با j نشان می‌دهیم. در مثلث توسعه ادعا، هر سطر مثلث نشان‌دهنده سال‌های حادثه و هر ستون آن متناظر با سال‌های توسعه است.

جدول ۱. مثلث تأخیر

سال تصادف i ام	سال‌های توسعه j								
	0	1	2	3	4	...	j	...	J
0									
1									
⋮									
$I+1-J$									
⋮									
i									
⋮									
$I-2$									
$I-1$									
I									

مشاهدات به صورت $C_{i,j}, X_{i,j}$
($i+j \leq I$)

مقادیر پیش‌گویی شده
 $C_{i,j}, X_{i,j}$ ($i+j > I$)

این مثلث روند تکامل پرداخت‌های بیمه‌گر را نشان می‌دهد، پرداخت‌هایی که وابسته به سال‌های توسعه و سال حادثه برای ادعای اولیه است. هدف بیم‌سنج‌ها تکمیل کردن این مثلث، از طریق برآورد ذخایر ادعای معوق برای سال‌های حسابرسی آتی است. به عبارت دیگر $X_{i,j}$ مقدار پرداختی برای ادعایی است که در سال حادثه i رخ داده و در سال $i+j$ مورد ارزیابی قرار گرفته است. مجموع ادعای پرداختی برای سال حادثه i بعد از j سال توسعه، با استفاده از رابطه

$$C_{i,j} = \sum_{k=0}^j X_{i,k},$$

به دست می‌آید. در زمان I مثلث توسعه ادعا به دو قسمت تقسیم می‌شود، مثلث بالا که شامل مشاهدات

$$D_I = \{X_{i,j}, i+j \leq I, 0 \leq i \leq I, 0 \leq j \leq J\},$$

و مثلث پایین که شامل ادعاهای معوق

$$D_I^c = \{X_{i,j}, i+j > I, 0 \leq i \leq I, 0 \leq j \leq J\},$$

است، که باید برآورد کنیم.

روش‌های مختلفی برای ذخیره‌سازی ادعا وجود دارد، از جمله می‌توان به روش نردبان زنجیره‌ای، روش بورن هویتر-فرگوسن و مدل پواسون اشاره کرد. با توجه به اینکه روش نردبان زنجیره‌ای و بورن هویتر فرگوسن الگوریتم محاسباتی قوی هستند که با استفاده از آنها شاخص‌های عدم قطعیت مانند خطای برآورد را نمی‌توان محاسبه کرد، بنابراین بیم‌سنج‌ها به معرفی مدل‌های تصادفی که بر پایه‌ی این الگوریتم‌های محاسباتی هستند، روی آوردند.

در بیشتر روش‌های ذخیره‌سازی ادعا، برای هر سطر اثر μ_i و برای هر ستون اثر λ_j در نظر گرفته می‌شود. بنابراین تعداد پارامترها وابسته به تعداد مشاهدات واقع در مثلث توسعه ادعاست. با توجه به اینکه اطلاعات کافی برای برآورد این پارامترها در دسترس نیست،

دقت مدل کاهش پیدا می‌کند. در این مقاله فرض می‌کنیم که مقدار افزایشی ادعا X_{ij} از توزیع پواسون بیش‌پراکنش (ODP) پیروی می‌کند. توزیع پواسون بیش‌پراکنش همانند توزیع پواسون است، که میانگین و واریانس آن به صورت

$$E[X_{i,j}] = m,$$

$$Var[X_{i,j}] = \phi m.$$

است. یک روش معمولی برای کاهش تعداد پارامترها، استفاده از منحنی پارامتری برای برآورد تعدادی از پارامترها است.

در این مقاله با استفاده از روش مطرح شده توسط ورال و همکاران (۲۰۱۲) برای کاهش تعداد پارامترها در مدل، از یک منحنی نمایی نزولی برای برآورد پارامترهای الگوی توسعه ادعا استفاده می‌کنیم. به عبارت بهتر، فرض می‌کنیم که لگاریتم الگوی توسعه ادعا از یک خط پیروی می‌کند اما این امکان وجود دارد که در برخی از نقاط از این خط منحرف شود. در نتیجه لگاریتم الگوی توسعه ادعا چند تکه‌ای خطی است که می‌تواند ترکیبی از چند خط باشد. برای تعیین نقاطی که لگاریتم الگوی توسعه ادعا از این خط منحرف شده است از الگوریتم RJMCMC استفاده می‌کنیم.

الگوریتم RJMCMC یک حالت خاصی از الگوریتم متروپولیس هستینگس است که شامل انتخاب مدل است. فرض می‌کنیم با مسئله انتخاب مدل بیزی روبه‌رو هستیم. مجموعه‌ی شمارای $\{M_k, k \in K\}$ ، مجموعه‌ای از مدل‌هایی است که می‌توان به داده‌های مشاهده شده X برازش داد. با توجه به اینکه زنجیر مارکوف بین مدل‌هایی با ابعاد مختلف پرش می‌کند، بنابراین نمی‌توان از روش گیبز برای تولید نمونه استفاده کرد. گرین (۱۹۹۵) روش نمونه‌گیری RJMCM را پیشنهاد داد. این روش زنجیر مارکوفی را که بین مدل‌هایی با ابعاد مختلف پرش می‌کند، تولید می‌کند. ایده اصلی روش RJMCMC بر پایه تطبیق بعد زنجیر مارکوف تولیدشده در وضعیت جدید با بعد زنجیر مارکوفی است که در وضعیت قبل تولید شده است. این روش از یک متغیر کمکی برای تطبیق بعد زنجیره‌های مارکوف استفاده می‌کند.

در یک نگاه کلی توانگری مالی را می‌توان وضعیتی در نظر گرفت که در آن دارایی‌های شرکت بیمه بیشتر از بدهی‌های آن باشد. اگر شرکت در این وضعیت قرار بگیرد، در این صورت توانگر خواهد بود، بنابراین باید به بررسی مخاطره موجود در برآورد ذخایر ادعای معوق بپردازیم. مخاطره موجود در شرکت بیمه غیر عمر در دو شاخه مخاطره ذخایر و مخاطره حق بیمه مورد بررسی قرار می‌گیرد. مخاطره ذخایر مربوط به تعهدات شرکت‌های بیمه در قراردادهای بسته‌شده در سال‌های قبل است که اغلب به عنوان مخاطره در ذخایر ادعا شناخته می‌شود. مخاطره ذخایر مربوط به ادعاهایی است که قبل از سال حسابرسی جاری اتفاق افتاده‌اند. مخاطره حق بیمه مربوط به مخاطره موجود در قراردادهاست که در سال حسابرسی جاری بسته می‌شوند و یا ادعاهایی که بر اساس قراردادهای قبلی در سال حسابرسی جدید رخ می‌دهند. به عبارت بهتر مخاطره‌هایی که پوشش داده شده، اما هنوز اتفاق نیفتاده است. با توجه به مطالعات اخیر خطای برآورد ذخایر ادعا در شاخه‌های تجاری دم سنگین، زمانی که این برآورد بر اساس بررسی کامل مثلث توسعه ادعا انجام می‌شود، ۲ یا ۳ برابر زمانی است که در یک بازه زمانی یکساله، تحت قوانین توانگری مالی ۲ به برآورد این ذخایر بپردازیم. بنابراین در این قسمت تفاوت استفاده از بازه زمانی یکساله که برای بررسی توانگری مالی ۲ شرکت مورد نیاز است (چشم‌انداز کوتاه‌مدت) را در مقابل بررسی کامل مثلث توسعه ادعا (چشم‌انداز بلندمدت) بررسی می‌کنیم.

۲- نتیجه‌گیری

در این مقاله ما یک رویکرد جدید برای مدل‌سازی مثلث توسعه ادعا را با استفاده از الگوریتم RJMCMC ارائه دادیم. علاوه بر این، توزیع پیش بین کامل مثلث توسعه ادعا و ذخایر ادعای معوق را به دست آوردیم که با استفاده از این توزیع، می‌توان سرمایه مورد نیاز برای توانگری مالی ۲ را محاسبه کرد.

مراجع

- [1] Verrall, Richard J., and Mario V. Wüthrich. "Reversible jump Markov chain Monte Carlo method for parameter reduction in claims reserving." *North American Actuarial Journal* 16.2 (2012): 240-259.
- [2] Verrall, Richard, Ola Hössjer, and Susanna Björkwall. "Modelling claims run-off with reversible jump Markov chain Monte Carlo methods." *ASTIN Bulletin: The Journal of the IAA* 42.1 (2012): 35-58.
- [3] Green, Peter J. "Reversible jump Markov chain Monte Carlo computation and Bayesian model determination." *Biometrika* (1995): 711-732.

کشش قیمتی در گروه‌های همگن ریسک؛ مطالعه موردی

مهناز منطقی‌پور^۱، علیرضا غفاری حدیقه^۲

^۱ دانشجوی دکتری، دانشگاه شهید مدنی آذربایجان، تبریز
manteqipour@azaruniv.ac.ir

^۲ هیات علمی گروه ریاضی، دانشگاه شهید مدنی آذربایجان، تبریز
hadigheha@azaruniv.ac.ir

چکیده مختصر

تعیین نرخ یکی از اولویت‌های پژوهشی در صنعت بیمه است. با محاسبه امید ریاضی خسارت‌ها و افزودن ضریب سربار ریسک به این مقادیر، نرخ نهایی بیمه‌نامه‌ها تعیین می‌شود. در این روش‌ها تابع تقاضای بیمه‌نامه‌ها در نظر گرفته نمی‌شود، در حالی که با در نظر گرفتن این توابع می‌توان نرخ‌ها را به صورت هدف دار تعیین نمود. در این مقاله در گروه‌های متفاوت ریسک در رشته بیمه باربری، تابع تقاضا را برآورد و کشش قیمتی را برای آنها محاسبه می‌کنیم. محاسبات انجام شده بر پایگاه داده‌های این پژوهش نشان می‌دهد ریسک‌های خوب کشش قیمتی پایین‌تری دارند. به عبارتی دیگر، با کاهش یکسان نرخ انواع بیمه‌نامه‌ها، ریسک‌های بد بیش از ریسک‌های خوب جذب می‌شوند.

کلمات کلیدی

حق بیمه، ریسک، تابع تقاضا، کشش قیمتی، گروه همگن ریسک، روش خوشه بندی K-میانگین‌ها.

۱- مقدمه

به طور کلی نرخ بیمه‌نامه‌ها، با توجه به امید ریاضی خسارت‌ها تعیین می‌شود. به این ترتیب که به امید ریاضی برآورد شده‌ی خسارت‌ها با استفاده از ضریب سربار ریسک، مقادیری اضافه می‌شود [1]. شرکت‌های بیمه با توجه به استراتژی‌های خود و شرایط بازار این ضریب سربار را تعیین می‌کنند. به طور رایج، وقتی شرکت‌ها دچار مشکلات کمبود نقدینگی می‌شوند با نرخ‌های پایین‌تری بیمه‌نامه صادر می‌کنند. کاهش نرخ‌ها تعداد مشتریان آن‌ها را افزایش می‌دهد اما این به معنی سودآوری شرکت نیست. افزایش تعداد مشتریان برای یک شرکت بیمه به معنی افزایش تعهدات آن می‌باشد. از طرفی دیگر کاهش نرخ بیمه‌نامه‌ها، درآمد حاصل از صدور یک بیمه‌نامه را کاهش می‌دهد. بنابراین در نرخ‌گذاری‌ها علاوه بر امید ریاضی خسارت‌ها، بررسی تابع تقاضا و کشش قیمتی برای گروه‌های متفاوت ریسک ضروری به نظر می‌رسد.

در این مقاله برای داده‌های رشته بیمه باربری یک شرکت بیمه گروه‌های همگن ریسک را به طور جداگانه مورد بررسی قرار داده‌ایم. برای ساختن گروه‌های همگن ریسک با قابلیت بیمه شوندگی، از روش ارایه شده در [2] استفاده کرده‌ایم و داده‌های ۹ سال یک شرکت بیمه را به ده گروه همگن تقسیم بندی کرده‌ایم. با استفاده از مدل‌های ارایه شده برای تابع تقاضا [3]، ضابطه‌ی تابع تقاضا را در هر گروه تعریف کرده‌ایم. برای برآورد پارمترهای این توابع نیاز داریم تا مقادیر آن‌ها را در چند نقطه به دست آوریم. به این منظور از خوشه‌بندی [4] قیمت‌های ارایه شده توسط شرکت و تعداد اعضای خوشه‌ها استفاده نموده‌ایم و به این ترتیب ۵ نقطه از هر تابع تقاضا را به دست آورده‌ایم. سپس با استفاده از تکنیک رگرسیون غیرخطی، پارامترهای مدل را با استفاده از نرم‌افزار R برآورد کرده‌ایم [5]. با استفاده از توابع تقاضا، ضابطه‌ی کشش قیمتی را در هر گروه همگن به دست آورده‌ایم.

در مقایسه‌ی کشش قیمت در گروه‌های همگن ریسک این نتیجه حاصل می‌شود که کشش قیمتی ریسک‌های خوب کمتر از کشش قیمت ریسک‌های بد است. منظور از ریسک‌های خوب ریسک‌هایی است که نرخ خسارت آن‌ها کمتر است. این به این مفهوم است که

با کاهش یکسان نرخ تمام ریسک‌ها ریسک‌های بد بیشتر از ریسک‌های خوب جذب می‌شوند. بنابراین، کاهش یکسان نرخ‌ها، تحت هر هدفی، تصمیم درست و معقولانه‌ای نیست و شرکت را با حجم بالایی از خسارت‌ها مواجه خواهد کرد. این معضل، در صنعت بیمه ایران توسط شرکت بیمه توسعه تجربه شده است.

۲- گروه‌های همگن ریسک

منظور از گروه‌های همگن ریسک گروه‌هایی از بیمه‌نامه‌ها است که اعضای آنها بیشترین شباهت را به یکدیگر و بیشترین تفاوت را با اعضای سایر گروه‌ها دارند. در نرخ‌گذاری بیمه‌نامه‌ها موضوعی که در تعیین شباهت‌ها حایز اهمیت است نرخ خسارت‌ها است. در روش ارایه شده در [2]، گروه‌های همگن به گونه‌ای ساخته شده‌اند که بیمه‌نامه‌هایی که بیشترین شباهت در نرخ خسارت‌ها را دارند در یک گروه قرار گرفته‌اند. از طرفی دیگر برای آن که بتوان با مطالعه‌ی هر گروه، در مورد رفتار اعضای آن نتیجه‌گیری کرد تعداد اعضای هر گروه، بزرگتر از حدی معین در نظر گرفته شده است. این روش را به اختصار شرح می‌دهیم.

این روش دارای دو مرحله است. در مرحله‌ی اول بر چسب داده‌ها که همان خسارت‌ها می‌باشند، با استفاده از اعداد فازی رتبه‌بندی می‌شوند و به این ترتیب خصوصیات اسمی به اعداد حقیقی تبدیل می‌شوند. سپس با استفاده از روش k -میانگین‌های اصلاح شده، گروه‌ها به گونه‌ای انتخاب می‌شوند که هر گروه حداقل تعداد اعضا را داشته باشد تا بتوان از قانون اعداد بزرگ استفاده کرد. تحلیل گروه‌های به دست آمده نشان داده است که این گروه‌ها همگن هستند. پایگاه داده‌های بیمه‌نامه‌های رشته بیمه‌باربری وارداتی یک شرکت بیمه طی ۹ سال دارای بیش از ۳۵۴۰۰۰ بیمه‌نامه بوده است. این داده‌ها با استفاده از روش مذکور، به ده گروه همگن ریسک تقسیم بندی شده‌اند. میانگین نرخ خسارت‌ها و تعداد اعضای گروه‌ها مطابق جدول ۱ می‌باشد.

جدول ۱. تعداد اعضای و میانگین خسارت گروه‌های همگن ریسک

شماره گروه	تعداد اعضای گروه	میانگین نرخ خسارت‌ها
۱	۶۰۳۴	۰,۰۰۰۰۶۴۲۰
۲	۴۲۶۲۹	۰,۰۰۰۴۱۱۳۴
۳	۳۴۲۴۸	۰,۰۰۰۰۳۳۹۲
۴	۲۳۱۴۹	۰,۰۰۰۰۸۹۴۱
۵	۹۶۹۰	۰,۰۰۰۰۵۳۴۰
۶	۱۴۲۱۶۸	۰,۰۰۰۰۶۵۳۰
۷	۵۹۰۲	۰,۰۰۰۲۶۷۹۵
۸	۱۴۸۹۰	۰,۰۰۰۰۰۳۴۵
۹	۲۱۶۰۷	۰,۰۰۰۲۰۲۱۳
۱۰	۵۴۴۶۷	۰,۰۰۰۳۰۱۳۱

۳- برآورد تابع تقاضا و کشش قیمت

پس از گروه‌بندی داده‌ها به ده گروه همگن ریسک، در هر گروه همگن تابع تقاضا را بر آورد می‌کنیم. ضابطه تقاضای بیمه‌نامه‌ها به صورت

$$d_i(x) = P_i e^{\left(1 - \frac{x}{\mu_i}\right)^{\lambda_i}}$$

تعریف شده است [3]. که در آن $d_i(x)$ ، تابع تقاضای گروه i ام در سطح قیمت x ، λ_i پارامتر کشش قیمت، μ_i امید ریاضی خسارت‌ها و P_i تقاضا در سطح قیمت μ_i است. برای برآورد این تابع در هر گروه نیاز است که چند نقطه از این توابع را داشته باشیم. به این منظور از خوشه‌بندی نرخ‌های ارایه شده توسط شرکت در سال‌های گذشته استفاده کرده‌ایم. در هر گروه همگن تعداد خوشه‌های قیمتی را با استفاده از روش k -میانگین‌ها [4]، برابر با ۵ خوشه در نظر گرفته‌ایم. مرکز هر خوشه، میانگین اعضای خوشه است و تقاضا برای نرخ مرکز هر خوشه را با تعداد اعضای خوشه تقریب زده‌ایم. به این ترتیب در هر گروه همگن ۵ نقطه از تابع تقاضا را برآورد کرده‌ایم. برای برآورد پارامترهای P_i و λ_i ، از تکنیک رگرسیون غیر خطی استفاده نموده‌ایم. برآورد پارامترها توسط رگرسیون غیر خطی نیاز به نقاط شروع اولیه دارد. به منظور به دست آوردن نقاط شروع اولیه، ابتدا توابع تقاضا را به صورت $d_i(x) = P_i^0 e^{(-x)}$ در نظر گرفته‌ایم و با استفاده از تکنیک رگرسیون خطی، نقاط شروع P_i^0 را تقریب زده‌ایم. با استفاده از مقادیر حاصل شده برای P_i^0 و $\lambda_i^0 = 0.5$ ، پارامترهای توابع d_i را برآورد کرده‌ایم که نتایج آن در جدول زیر قابل ملاحظه است.

جدول ۲. مقادیر برآورد شده پارامترهای توابع تقاضا در ده گروه

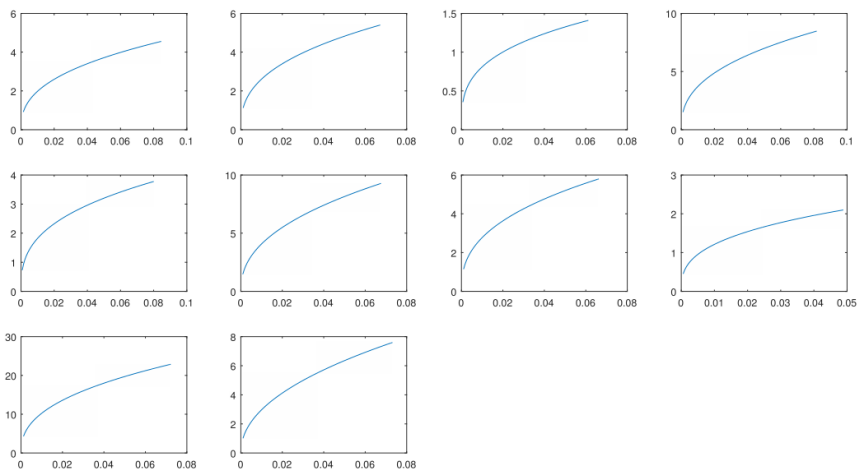
شماره گروه	پارامتر	مقادیر برآورد شده	انحراف معیار	t-value	p-value
۱	P_1	۵۶۱۵۰	۱۶۲۶	۳۴.۵۳	۰.۵۵-۵.۳۴
	λ_1	۰.۳۸۸۷	۰.۰۰۲۸	۱۳۶.۸	۰.۷۵-۸.۶۱
۲	P_2	۴۵۷۹۰	۳۷۶۶	۱۲.۱۶	۰.۰۰۱۲۰
	λ_2	۰.۳۸۱۲	۰.۰۰۳۰۲	۱۲.۶۲	۰.۰۰۱۰۷
۳	P_3	۹۶۹۷	۳۰۷۹	۳۱.۵۰	۰.۵۵-۷.۰۳
	λ_3	۰.۳۰۵۶	۰.۰۰۳۴	۸۸.۱۸	۰.۶۵-۳.۲۱
۴	P_4	۹۶۹۱۰	۱۰۳۹	۹۳	۰.۶۵-۲.۷۴
	λ_4	۰.۳۹۳	۰.۰۰۱۵۶	۲۵۱.۱	۰.۷۵-۱.۳۹
۵	P_5	۲۷۱۰۰	۲۳۰۰	۱۱.۷۸	۰.۰۰۱۳۱
	λ_5	۰.۳۵	۰.۰۱۰۹	۳۱.۸۴	۰.۵۵-۶.۸۱
۶	P_6	۱۱۲۴۰۰	۳۱۴۹	۳۵.۷	۰.۵۵-۴.۸۳
	λ_6	۰.۴۳	۰.۰۰۳۲	۱۳۴.۴	۰.۷۵-۹.۰۸
۷	P_7	۷۴۰۶	۱۰۱۵	۷.۲۹۵	۰.۰۰۵۳۲
	λ_7	۰.۳۹	۰.۰۳۸	۱۰.۲۶۲	۰.۰۰۱۹۷
۸	P_8	۱۷۶۴۰۰	۱۹۰۳۰	۹.۲۶۹	۰.۰۰۲۶۶
	λ_8	۰.۳۴۶۲	۰.۰۰۳۵	۹۷.۶۹۲	۰.۶۵-۲.۳۶
۹	P_9	۴۷۷۱	۸۴۶	۵۶.۴۰	۰.۵۵-۱.۲۳
	λ_9	۰.۴۰۵۴	۰.۰۰۴	۹۹.۷۵	۰.۶۵-۲.۲۲
۱۰	P_{10}	۸۵۹۹	۱۶۱۰۰۰	۵۳.۴۱	۰.۵۵-۱.۴۵
	λ_{10}	۰.۴۷۱۸	۰.۰۶۱۷	۷۶.۳۶	۰.۶۵-۴.۹۵

مقادیر کوچک p-value در جدول ۲. معنا دار بودن مقادیر پارامترها را نشان می‌دهد.

برای توابع محاسبه شده، کشش قیمتی که تغییرات تقاضا به تغییرات قیمت را محاسبه می‌کند از رابطه‌ی زیر محاسبه می‌شود:

$$\left| \frac{x}{d_i} \cdot \frac{\partial d_i(x)}{\partial x} \right| = \left| \frac{\partial \ln[d_i(x)]}{\partial \ln x} \right| = \lambda_i \left(\frac{x}{\mu_i} \right)^{\lambda_i}$$

در شکل ۱. برای ده گروه همگن تابع کشش قیمتی را رسم کرده‌ایم. این نمودار، توابع کشش قیمتی را از چپ به راست و از بالا به پایین برای گروه‌های ۱ تا ۱۰ نشان می‌دهد.



شکل ۱. توابع کشش قیمتی در گروه‌های همگن ریسک، از بالا به پایین و از چپ به راست در گروه‌های ۱ تا ۱۰.

۴- نتایج

نتایج این پژوهش مناسب بودن ضابطه تابع تقاضای ارائه شده برای بیمه‌نامه‌ها را تایید می‌کند. نتیجه دیگر آن است که با مقایسه توابع کشش قیمتی شکل ۱ و نرخ خسارت‌ها در جدول ۱. این نتیجه حاصل می‌شود که ریسک‌های خوب کشش قیمتی کمتری دارند. بنابراین در صورت کاهش یکسان نرخ انواع ریسک‌ها، ریسک‌های بد بیشتر از ریسک‌های خوب جذب می‌شوند و کاهش غیر اصولی نرخ‌ها بیش از آن‌چه تصور می‌شود برای شرکت‌های بیمه مشکل ایجاد می‌نماید. به عنوان مثال، شرکت بیمه توسعه شرکتی است که با انتخاب چنین رویه‌ای دچار بحران مالی شده است.

مراجع

- [1] Mikosch, T., (2009). Non-life insurance mathematics, Springer.
- [2] Manteqipour, M., Ghaffari-Hadigheh, A., Mahmoodvand, R. Safari, A., (2017). Grouping objects to homogenous classes satisfying requisite mass, Accepted for publication in Journal of AI and Data mining.
- [3] Jong, D. , P., Ferris, S., (2006). Adverse selection spirals, Astin Bulletin, 36, 589-628.
- [4] Wu, J. (2012). Advances in K-means clustering: a data mining thinking, Springer Science & Business Media.
- [5] Gentleman R., Hornik K., Parmigiani G., (2011) . Nonlinear Regression with R, Springer, 93—108.

مدل‌بندی ریسک حوادث فاجعه‌آمیز در بیمه عمر: مطالعه موردی آن در ایران

^۱سیده مریم محدث کسائی و ^۲آمین حسن‌زاده

^۱انجمن محاسبات بیمه و مالی ایران
Kassai.sbu@gmail.com

^۲گروه بیم‌سنجی، دانشگاه شهید بهشتی
Ako.hassanzadeh@gmail.com

چکیده مختصر

حوادث فاجعه‌آمیز، در صورت بروز، پیامدهای مالی سنگینی را برای شرکت‌های بیمه می‌توانند به بار آورند. مسأله مورد بررسی در این مقاله مدل‌بندی ریسک حوادث فاجعه‌آمیزی است که یک شرکت بیمه عمر با آن روبه‌رو می‌شود. پایه مدل‌بندی بر اساس مدل پواسن مرکب است و از مدل قله‌های فراتر از آستانه برای ارزیابی توزیع خسارت ناشی از کشته شدن افراد در هر حادثه فاجعه‌آمیز و از توزیع بتا-دوجمله‌ای برای نسبت این خسارت به مشتریان شرکت بیمه، استفاده شده است. برای بررسی نیکویی برازش مدل معرفی شده داده‌های مربوط به تعداد و شدت حوادث فاجعه‌آمیز ایران از سال ۱۹۸۳ تا سال ۲۰۱۶ جمع‌آوری شده و مدل معرفی شده به آن‌ها برازنده شده است. در پایان، با شبیه‌سازی هزینه خسارت سالانه کل ناشی از چنین حوادثی، قیمت بیمه‌نامه عمر مازاد زیان حوادث فاجعه‌آمیز برای شرکت‌های بیمه قابل محاسبه خواهد بود.

کلمات کلیدی

پوشش مازاد زیان حوادث فاجعه‌آمیز، بیمه عمر اتکایی، مدل قله‌های فراتر از آستانه، داده‌های حوادث فاجعه‌آمیز، مدل حوادث فاجعه‌آمیز

مراجع

- [1] Ekhedden, E. and Hossjer, O. (2014), Pricing Catastrophe Risk in Life (Re)Insurance, Scandinavian Actuarial Journal, 4, 352-367.
- [2] Swiss Re (2007-2016), Natural catastrophes and man-made disasters, Sigma, Swiss Reinsurance Company.

معیار نیکوئی برازش جدید در مواجهه با مه‌داده‌های درآمد

شهریار میرزایی

استادیار گروه آمار دانشگاه پیام نور، کرمانشاه
sh_mirzaee@pnu.ac.ir

چکیده مختصر

در برازش مدل‌های آماری به داده‌های انبوه درآمد، معیارهای معمول نیکوئی برازش عملکرد خوبی ندارند. در این مقاله به معرفی معیارهای جدید نیکوئی برازش بر اساس منحنی‌های نابرابری درآمد لورنتس و زنگا می‌پردازیم. همچنین مزیت این ابزارها نسبت به سایر معیارهای معمول در تشخیص توزیع آماری بحث می‌شود.

کلمات کلیدی

معیار نیکوئی برازش، منحنی لورنتس، منحنی زنگا.

۱- مقدمه

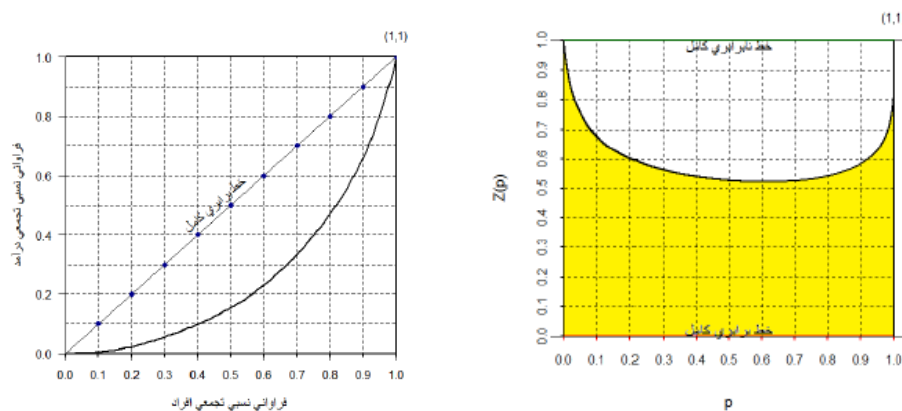
یکی از مسائلی که معمولاً در تحقیقات با آن روبه‌رو می‌شویم، تعیین توزیع داده‌های جمع‌آوری شده است. نیکویی برازش در آمار ارزیابی نزدیکی بین یک مجموعه از مشاهدات و یک مدل فرضی در نظر گرفته شده می‌باشد. توزیعی را که حدس می‌زنیم داده‌ها از آن باشند توزیع برازنده بر داده‌ها و آزمون لازم را آزمون برازندگی می‌نامیم. اغلب روش‌های آماری بر مبنای یک توزیع پیش فرض برای جامعه نتایج را به دست می‌آورند که این مسئله بیانگر ضرورت به‌کارگیری آزمون‌های نیکویی برازش است. در قابلیت اعتماد، اقتصاد، علوم رفتاری و اجتماعی دانستن این که توزیع داده‌ها از چه نوعی است، اهمیت ویژه‌ای دارد زیرا اگر فرض ما نسبت به توزیع جامعه نادرست باشد، نتایج به دست آمده بی‌اعتبار خواهند بود. برای آزمون نیکوئی برازش از آزمون‌های مختلف از جمله کای‌دو، اندرسون دارلینگ، کرامر-فون میزس و نیز معیارهایی بر اساس مفهوم آنتروپی و نظریه اطلاع استفاده می‌شود. ولی از جمله معایب این آزمون‌ها آن است که در مواجهه با مه‌داده‌ها پرتوان نبوده یا مقادیر آماره آزمون آنها دارای مقادیر ناجور خیلی بزرگ یا خیلی کوچک هستند. از این‌رو ایده گسترش کلاس‌های مختلف آزمون‌های نیکویی برازش همواره مورد توجه پژوهشگران بوده است.

در مبحث نابرابری درآمد، چون اغلب با داده‌های انبوه درآمد سروکار داریم، آزمون‌های مختلف نیکویی برازش پرتوان نبوده و جوابگو نمی‌باشند. هدف ما در این مقاله ارائه شکل‌های جدیدی از معیارهای نیکویی برازش بر اساس منحنی‌های نابرابری درآمد است که قابلیت رقابت با سایر آزمون‌های نیکویی برازش را در مواجهه با مه‌داده‌های درآمد دارد.

۲- منحنی‌های نابرابری درآمد

منحنی لورنتس. این منحنی، نموداری است که فراوانی تجمعی نسبی درآمد را در برابر فراوانی تجمعی نسبی جمعیت رسم می‌کند. فرض کنید داده‌های نا منفی درآمد به اندازه n به صورت یک نمونه تصادفی X_1, \dots, X_n از جامعه مورد نظر باشند. این داده‌ها را از کوچک به بزرگ مرتب می‌کنیم و به صورت $X_{1:n} \leq X_{2:n} \leq \dots \leq X_{n:n}$ نمایش می‌دهیم. تابع لورنتس به ازای $k = 0, 1, \dots, n$ در نقطه

همانند نمودار سمت راست شکل ۱ حاصل می‌شود. در مطالعه منحنی لورنتس هدف این است که بدانیم به درصد مشخصی از جمعیت، چند درصد از درآمد کل جامعه اختصاص داده شده است. قابل ذکر است وقتی درآمدهای افراد جامعه برابر باشند، منحنی لورنتس بر نیمساز ربع اول محور مختصات منطبق می‌شود. بنابراین هر گونه فاصله میان خط نیمساز (برابری کامل) و منحنی لورنتس، میزان نابرابری را در جامعه بیان می‌کند.



شکل ۱: الگوئی از منحنی زنگا در سمت چپ و الگوئی از منحنی لورنتس در سمت راست

منحنی نابرابری زنگا. در سال ۲۰۰۷ زنگا شاخصی را بر اساس مقایسه میانگین درآمد پایین جامعه به میانگین درآمد بالای متناظرش معرفی کرد. فرض کنید x_1, \dots, x_n داده‌های درآمد افراد و $x_{1:n} \leq x_{2:n} \leq \dots \leq x_{n:n}$ آماره‌های ترتیبی این درآمدها باشند. در این

صورت میانگین درآمد $100 \times p$ % جامعه را با $M_p^- = \frac{\sum_{i=1}^{[np]} x_{i:n}}{np}$ و میانگین درآمد مابقی افراد جامعه را به صورت

نشان می‌دهند. بنابراین منحنی زنگا مطابق منحنی سمت چپ شکل ۱ براساس نسبت میانگین درآمد این دو زیر

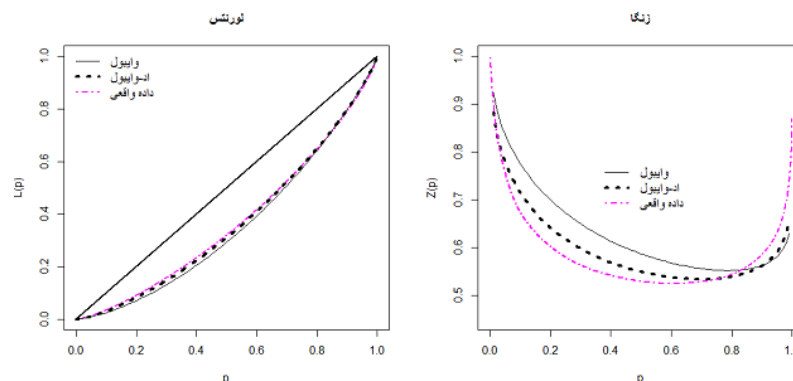
گروه جامعه به ازای هر $p \in (0, 1)$ به صورت $Z(p) = 1 - \frac{M_p^-}{M_p^+}$ تعریف می‌شود.

۳- ارزیابی نیکویی برازش با منحنی‌های نابرابری

برای تبیین مساله، در اینجا، به داده‌های درآمد براساس درآمد ۱۴۸۲۷ خانوار اتریشی در سال ۲۰۰۶ توزیع وایبول (W) با تابع توزیع $G(x, \alpha, \beta) = 1 - \exp(-(\frac{x}{\beta})^\alpha)$ ، $x, \alpha, \beta > 0$ و تعمیم خاصی از مدل وایبول به نام توزیع اد-وایبول (OW) (کوریا، ۲۰۰۶) با

تابع توزیع $F(x, \alpha, \beta, \nu) = \frac{G(x)^\nu}{G(x)^\nu + \bar{G}(x)^\nu}$ برازش می‌دهیم که در آن $\bar{G}(x) = 1 - G(x)$ و ν پارامتر جدید شکل توزیع است.

با رسم و مقایسه توابع توزیع مدل‌های یاد شده، تابع توزیع تجربی داده‌های واقعی و نیز توابع چگالی مدل‌های وایبول و اد-وایبول روی هیستوگرام داده‌های درآمد، به صورت بصری نشان داده می‌شود که توزیع اد-وایبول برازش بهتری به داده‌های درآمد دارد. در شکل ۲ دو منحنی لورنتس و زنگا ترسیم شده‌اند. فاصله بین منحنی نابرابری نظری تحت مدل توزیع درآمد و منحنی نابرابری تجربی مربوط به داده‌های واقعی به‌عنوان یک ملاک نیکویی برازش پیشنهاد می‌شود. از دیدگاه هندسی همان‌گونه که در شکل ۲ نشان داده شده است این ملاک می‌تواند به‌عنوان فاصله بین منحنی نظری و تجربی توزیع درآمد برای منحنی‌های لورنتس و زنگا در نظر گرفته شود. ملاک‌های نیکویی برازش براساس انواع فواصل بین منحنی‌های نظری و تجربی توزیع درآمد لورنتس و زنگا در جدول ۱ گزارش شده‌اند که در آن $\hat{L}(p)$ و $\hat{Z}(p)$ به ترتیب منحنی‌های زنگا و لورنتس توزیع برازش یافته، $L(p)$ و $Z(p)$ منحنی‌های توزیع تجربی داده‌ها هستند. این معیارها در حالت کمترین اختلاف، برابر صفر و در حالت بیشترین فاصله، برابر یک هستند. مقادیر کمتر این معیارها، برازش بهتر مدل را بیان می‌کند. ملاک‌های نیکویی برازش براساس فواصل تعریف شده بین منحنی‌های نظری و تجربی نابرابری درآمد برای توزیع W و OW برای داده‌های درآمد خانوارهای اتریشی در جدول ۲ ارائه شده‌اند. همچنین برای مقایسه بهتر، از مقادیر معیارهای اطلاع آکائیکه (AIC)، اطلاع بیزی (BIC)، اطلاع هان-کوین (HQIC) و نیز اطلاع آکائیکه سازگار (CAIC) استفاده می‌شود (آکائیکه، ۱۹۷۴). این معیارها بر اساس مفهوم آنتروپی برای سنجش نیکویی برازش به کار می‌روند و نشان می‌دهند که استفاده از یک مدل به چه میزان باعث از دست رفتن اطلاعات می‌شود. یافته‌های جدول ۲ نشان می‌دهند که معیارهای جدید نیکویی برازش براساس منحنی‌های نابرابری، علاوه بر اینکه بهنجارند یعنی مقادیر آنها بین صفر و یک قرار دارند، با دیگر معیارهای عمومی نیکویی برازش هماهنگ می‌باشند. بنابراین معیارهای جدید براساس منحنی‌های نابرابری درآمد معتبر به نظر می‌رسند.



شکل ۴: منحنی‌های لورنتس و زنگای برازش یافته به داده‌های درآمد نمونه‌های خانوارهای اتریشی

معیارهای نیکویی برازش براساس منحنی‌های نابرابری

لورنتس	زنگا
$A_1 = \int_0^1 \frac{ \hat{L}(p) - L(p) }{L(p)} dp$	$B_1 = \int_0^1 \frac{ \hat{Z}(p) - Z(p) }{Z(p)} dp$
$A_2 = \int_0^1 \frac{ \hat{L}(p) - L(p) }{\hat{L}(p)} dp$	$B_2 = \int_0^1 \frac{ \hat{Z}(p) - Z(p) }{\hat{Z}(p)} dp$
$A_3 = \left[\int_0^1 \left[\frac{\hat{L}(p) - L(p)}{L(p)} \right]^2 dp \right]^{\frac{1}{2}}$	$B_3 = \left[\int_0^1 \left[\frac{\hat{Z}(p) - Z(p)}{Z(p)} \right]^2 dp \right]^{\frac{1}{2}}$
$A_4 = \left[\int_0^1 \left[\frac{\hat{L}(p) - L(p)}{\hat{L}(p)} \right]^2 dp \right]^{\frac{1}{2}}$	$B_4 = \left[\int_0^1 \left[\frac{\hat{Z}(p) - Z(p)}{\hat{Z}(p)} \right]^2 dp \right]^{\frac{1}{2}}$

جدول ۱: معیارهای نیکویی برازش براساس منحنی‌های نابرابری

مقایسه‌ی معیارهای نیکویی برازش

W			OW		
لورنتس	زنگا	معیارهای اطلاع	لورنتس	زنگا	معیارهای اطلاع
$A_1 = 0.1207$	$B_1 = 0.0978$	$AIC = 313283$	$A_1 = 0.0614$	$B_1 = 0.0409$	$AIC = 312152$
$A_2 = 0.0172$	$B_2 = 0.0089$	$BIC = 313298$	$A_2 = 0.0069$	$B_2 = 0.0040$	$BIC = 312175$
$A_3 = 0.2475$	$B_3 = 0.2793$	$HQIC = 313488$	$A_3 = 0.2479$	$B_3 = 0.1475$	$HQIC = 312159$
$A_4 = 0.2881$	$B_4 = 0.0992$	$CAIC = 313483$	$A_4 = 0.1966$	$B_4 = 0.0463$	$CAIC = 312152$

جدول ۲: مقایسه معیارهای نیکویی برازش برای داده‌های واقعی براساس منحنی‌های نابرابری

مراجع

- [1] Akaike, H. (1974), A new look at the statistical model identification, *IEEE Transactions on Automatic Control*, 196, 716-723.
- [2] Cooray K. (2006), Generalization of the Weibull distribution: the odd Weibull family, *Statistical Modelling*, 6, 265-277.
- [3] Lorenz, M. O. (1905). Methods of measuring the concentration of wealth, *Journal of the American Statistical Association*, 9, 209-219.
- [4] Zenga, M. (2007). Inequality curve and inequality index based on the ratios between lower and upper arithmetic means. *Statistica and Applicazioni*, 5, 3-28.

مکانیزم متعادل خودکار در سیستم‌های بازنشستگی عمومی

شیوا مهدی‌پور قبادلو^۱، امین حسن‌زاده^۲

^۱ گروه بیم‌سنجی، دانشگاه شهید بهشتی، تهران،
mehdipoors45@yahoo.com
am_hassanzadeh@sbu.ac.ir

چکیده مختصر

در دو دهه اخیر، بحث در ماهیت و اصلاح سیستم‌های بازنشستگی بدون اندوخته (PAYG) در دنیا اوج گرفته است. یکی از بزرگ‌ترین ضعف‌های PAYG، این است که حل مشکلات مالی و اقتصادی سیستم، به افزایش نرخ کسورات می‌انجامد، در حالی که ایده اصلی در رابطه با کاهش هزینه، عدم تغییر نرخ کسورات برای آینده نامحدود است. یکی از اصلاحات مورد استفاده برای PAYG، مکانیزم متعادل خودکار (ABM) است. یک ABM با هدف حفظ نقدینگی، توانگری مالی و پایداری سیستم در دراز مدت و بدون مداخله قانونی، زمانی که ریسک‌های جمعیتی مانند، کاهش نیروی کار و ریسک‌های اقتصادی مانند کاهش بودجه ذخایر به دلیل عملکرد بازارهای مالی، وجود داشته باشد، در سیستم اعمال می‌شود. در حالی که متضمن ایفای تعهدات، بدون تغییر در نرخ کسورات است. در این مقاله، به بیان مفاهیم، فواید و ضرورت وجود یک مکانیزم متعادل خودکار در سیستم بازنشستگی PAYG پرداخته شده است. همچنین انواع ABM‌های مورد استفاده در سوئد، کانادا، آلمان، ژاپن و فنلاند، مورد تحلیل و بررسی قرار می‌گیرد.

کلمات کلیدی

پایداری، نرخ کسورات، ارزیابی بیم‌سنجی، توانگری مالی، مزایای بازنشستگی، امید به زندگی، ریسک.

مراجع

- [1] Vidal-Meliá, C., Boado-Penas, M. E. C., & Settergren, O., (2009), Automatic balance mechanisms in pay-as-you-go pension systems, *The Geneva Papers on Risk and Insurance: Issues and Practice*, 34 (2), 287-317.
- [2] Boado-Penas, Mara del Carmen, Inmaculada Dominguez-Fabian, Carlos Vidal-Melia, (2007), Notional Dened Contribution Accounts (NDCs): Solvency and Risk, Application to the Case of Spain, *International Social Security Review*, 60 (4), 105-127.
- [3] Godínez-Olivares, H., Boado-Penas, M. D. C., & Pantelous, A. A., (2016), How to Finance Pensions: Optimal Strategies for Pay-as-You-Go Pension Systems, *Journal of Forecasting*, 35(1), 13-33.
- [4] Godínez-Olivares, H., del Carmen Boado-Penas, M., & Haberman, S., (2016), Optimal strategies for pay-as-you-go pension finance: A sustainability framework, *Insurance: Mathematics and Economics*, 69, 117-126.

نرخ‌گذاری بر اساس مدل‌های آمیخته‌ی خطی باورمندی با ادعاهای دم‌سنگین

هما نوروزی^۱، شیما جلیلی^۲

^۱دانشجوی ارشد اکچوئری، بیمه دانا، تهران،
homa.1368@gmail.com

shimajalili68@gmail.com

چکیده مختصر

مبحث نرخ‌گذاری در حوزه‌های مختلف بیمه، از اهمیت بسیاری برخوردار است. در این راستا، در این مقاله به بررسی توانایی و دقت مهم‌ترین مدل‌های علم بیم‌سنجی در این زمینه با عنوان مدل‌های باورمندی پرداخته شده و نیز پیشگوی جدیدی با دقت بیشتر در زمینه‌ی نرخ‌گذاری بر اساس برآوردگر استوار کارای تصحیح شده برای مدل‌های باورمندی معرفی شده است. همچنین کاربرد این پیشگوی جدید در مدل‌های باورمندی بررسی شده است. این بررسی بر اساس مقایسه‌ی نتایج حاصل از مدل‌های باورمندی استوار با پیشگوی جدید با مشهورترین مدل‌ها در این حوزه، موسوم به مدل‌های باورمندی بولمان-استراب و رگرسیون هاجمیستر انجام شده است. با مطالعه‌ی مجموعه‌ای از داده‌های بیمه‌ی شخص ثالث، کاربرد این مدل در واقعیت نشان داده شده است. از این رو با توجه به ماهیت مفهوم استواری و دم‌سنگین بودن ادعاها در دنیای واقعی بیمه، می‌توان از مدل باورمندی استوار، به‌عنوان مدلی توانمند در مبحث نرخ‌گذاری بیمه‌ای استفاده کرد.

کلمات کلیدی

مدل باورمندی استوار، ادعاهای دم‌سنگین، برآوردگر استوار کارای تصحیح‌شده، مدل‌های آمیخته‌ی خطی، مدل بولمان-استراب، مدل رگرسیون هاجمیستر.

۱- کاربرد مدل باورمندی استوار

در مثالی از داده‌های بیمه‌ی شخص ثالث، ابتدا داده‌ها از نظر عدم وابستگی به زمان با توجه به نمودار سری زمانی چندمتغیره بررسی شده‌اند و وابستگی آن‌ها به زمان مشخص شده است و البته از طریق برازش مدل رگرسیونی و های به‌دست آمده، این موضوع نیز اثبات شده است. درنهایت مدل

$$\log(y_{it}) = x_i \beta + z_i \alpha_i + \varepsilon_{it} v_{it}^{1/2}, i = 1, \dots, 10; t = 1, \dots, 11$$

را بر روی داده‌ها برازش می‌دهیم. نتایج حاصل از این مدل (CALTC-UE) و نتایج مربوط به برازش مدل رگرسیونی هاجمیستر بازمینی‌شده (REML-HR) و مدل باورمندی استوار هم‌واریانس (CALT-E) به‌دست آمده و در جدول زیر با هم مقایسه شده است.

همانطور که تمام اعداد جدول نشان می‌دهد، مدل باورمندی استوار ناهم واریانس (CALTC-UE)، میزان پارامترها را عادلانه‌تر نسبت به سایر مدل‌ها پیش‌بینی می‌کند و این نتیجه‌ای بسیار مطلوب در محاسبه حق‌بیمه است

منطقه‌ی ۱۰	منطقه‌ی ۹	منطقه‌ی ۸	منطقه‌ی ۷	منطقه‌ی ۶	منطقه‌ی ۵	منطقه‌ی ۴	منطقه‌ی ۳	منطقه‌ی ۲	منطقه‌ی ۱	β	مدل
۲۹۳۰۶۳۵٫۷	۲۹۷۴۸۰۹	۳۶۳۹۷۲۳٫۸	۵۴۵۹۶۴۶	۵۸۵۵۶۳۳٫۸	۵۲۷۷۷۱۴٫۸	۴۹۹۲۲۱۳٫۴	۵۲۹۱۰۲۸٫۴	۷۷۷۲۶۶۵٫۴	۱۹۷۹۱۸۸۵	۳۱۹۹۲۹۸	REML - HR
۶۸۸۸۲۰۵	۶۹۱۳۹۳	۷۳۰۱۱۴٫۵	۸۳۶۰۹۸	۸۵۹۱۵۸٫۳	۸۲۵۵۰۳٫۱	۸۰۸۸۷۶٫۹	۸۲۶۲۷۸٫۴	۹۷۰۷۹۶٫۹	۱۶۷۰۷۳۸	۴۴۵۳۸۹	
۱۲٫۴۵	۱۲٫۴۶	۱۳٫۲۳	۱۴٫۶۵	۱۴٫۷۶	۱۴٫۸۰	۱۴٫۲۵	۱۴٫۷۹	۱۵٫۴۵	۱۶٫۶۹	۱۴٫۳۵	CALT - E
۰٫۱۴	۰٫۱۶	۰٫۱۷	۰٫۱۰	۰٫۰۹	۰٫۰۷۵	۰٫۱۲	۰٫۰۷۵	۰٫۰۶۵	۰٫۰۵۶	۰٫۱۱	
۱۲٫۱۷	۱۲٫۲۰	۱۳٫۱۴	۱۴٫۶۴	۱۴٫۷۴	۱۴٫۸۰	۱۴٫۲۲	۱۴٫۷۸	۱۵٫۴۵	۱۶٫۶۲	۱۳٫۰۴۴	CALTC - UE
۰٫۱۸	۰٫۱۹	۰٫۱۹	۰٫۱۰	۰٫۰۹۳	۰٫۰۷۵	۰٫۱۲	۰٫۰۷۶	۰٫۰۶۴	۰٫۰۵۵	۰٫۱۶۳	

۲- بحث و نتیجه‌گیری

با توجه به مثال ارائه شده می‌توان نتیجه گرفت که به‌طور مستقیم، برآوردگرهای بسیار کارا و استوار بدون هیچ ویژگی خاصی برای انتخاب نقاط بریده شده، تأثیر نقاط دورافتاده بر وزن‌های باورمندی که چه درون مخاطره‌ها و چه بین مخاطره‌ها وجود دارند را می‌کاهند و براساس آنها می‌توان به سادگی حق‌بیمه‌های باورمندی را بدون محاسبه‌ی وزن‌های آنها به دست آورد.

مراجع

- [1] Bühlmann, H. and Gisler, A. 2005, A Course in Credibility Theory and its Applications, Springer, New York.
- [2] Gisler, A. and Reinhard, P. 1993, "Robust credibility," ASTIN Bulletin, 23 (1), 117-143.
- [3] Searle, S. R., Casella, G., and McCulloch, C. E. 1992, Variance Components, Wiley Series in Survey Methodology.
- [4] Norberg, R. 1980, "Empirical Bayes credibility," Scandinavian Actuarial Journal, 181-221.
- [5] Dornheim, H. and Brazauskas, V. 2011, "Robust-efficient credibility models with heavy-tailed claims: A mixed linear models perspective," Insurance: Mathematics and Economics, 48, 72-84.
- [6] Klugman, S., Panjer, H., and Willmot, G. 2004, Loss Models: From Data to Decision, 2nd ed, New York: Wiley

نرخ‌گذاری کارای مبتنی بر طبقه‌بندی ریسک به منظور قیمت‌گذاری منصفانه

سمانه عزیزنصیری

کارشناسی ارشد اکچوئری، شرکت بیمه حکمت صبا، تهران،
saziznasiri@gmail.com

چکیده مختصر

نرخ‌گذاری مبتنی بر طبقه‌بندی یکی از مهمترین مراحل در محاسبات نرخ بیمه‌های اموال و حوادث است. این رویکرد، اصلاح نرخ محاسبه شده برای کل پورتفوی بیمه‌گذاران براساس خصوصیات هر بیمه‌گذار در تلاش برای منصفانه نمودن حق‌بیمه در سطح انفرادی می‌باشد. انتخاب نامناسب زمانی به وجود می‌آید که بیمه‌گر نتواند براساس فاکتور ریسکی که سایر بیمه‌گران از آن استفاده می‌کنند، طبقه‌بندی مناسبی انجام دهد. در مقابل انتخاب مناسب برای یک شرکت زمانی است که آن شرکت از فاکتورهای ریسک برای تمییز بین بیمه‌گذاران استفاده می‌نماید. در نظر گرفتن فاکتورهای ریسک در سیستم نرخ‌گذاری منجر به منصفانه‌تر شدن حق‌بیمه‌ها می‌شود. در این مقاله به دنبال در اختیار داشتن سیستم کارای نرخ‌گذاری، کارایی سیستم با در نظر گرفتن فاکتورهای ریسک سن وسیله نقلیه و نوع کاربری در بیمه بدنه اتومبیل محاسبه و تحلیل شد. نتایج نشان می‌دهد تغییرپذیری سیستم نرخ متوسط در بیمه بدنه اتومبیل بسیار بالا و حق‌بیمه‌ها غیرمنصفانه می‌باشد و در نظر گرفتن فاکتورهای ریسک، کارایی سیستم را جهت تمییز منصفانه بین بیمه‌گذاران بهبود می‌بخشد.

کلمات کلیدی

نرخ‌گذاری، سیستم طبقه‌بندی ریسک، کارایی، معرض خطر، نرخ‌گذاری انفرادی ریسک.

۱- مقدمه

چگونه محصولی بدون اطلاع از هزینه‌های صورت گرفته قیمت‌گذاری می‌شود؟ صنعت بیمه روزانه با این چالش روبرو است. در حالی که اکثر صنایع از هزینه‌های مواد اولیه، نیروی کار و حاشیه سود برای محاسبه قیمت محصولات خود مطلع هستند، بیمه‌گران هزینه محصولات خود را در زمان فروش نمی‌دانند. هزینه صحیح ارائه یک محصول در صنعت بیمه تا چندین سال هم می‌تواند نامعلوم باقی بماند تا زمانیکه خسارتی پرداخت شده باشد. بنابراین شرکت‌های بیمه به خصوص اکچوئری‌ها، بر داده‌های گذشته برای پیش‌بینی رفتار آتی به منظور محاسبه نرخ حق‌بیمه و قیمت‌گذاری محصولات بیمه‌ای متکی هستند. داده‌های گذشته فقط از این جهت استفاده می‌شوند که اطلاعات با ارزشی در مورد تخمین هزینه‌های مورد انتظار آینده ایجاد می‌کنند. در هنگام استفاده از داده‌های خسارت گذشته باید دقت شود که برای تبدیل این تجربه به آنچه در آینده انتظار می‌رود، تعدیلاتی لازم است. به عنوان مثال، اگر تورم بر مبالغ خسارت اثر گذاشته باشد،

خسارت‌های آینده بیش از آن چیزی خواهند بود که در گذشته اتفاق افتاده است. اشتباه در تشخیص خسارت باعث کمتر تخمین زدن حق بیمه مورد نیاز جهت حصول سود مورد نظر شرکت بیمه است.

زمانی که در مورد کفایت و مناسب بودن نرخ‌ها صحبت می‌شود، منظور برقراری معادله اساسی بیمه در سطح کلی و جزئی است. تعادل در سطح کلی به این معنا است که این اطمینان ایجاد شود که کل حق بیمه دریافتی به منظور پوشش کل خسارت‌های مورد انتظار، هزینه‌ها و حاشیه سود بیمه‌گر کافی باشد. اگر نرخ‌های پیشنهادی زیاد یا کم باشند، شرکت برای تامین سود باید نرخ‌ها را بطور یکنواخت کم یا زیاد کند. البته در زمان نرخ‌گذاری باید توجه نمود که علاوه بر برقراری تعادل در سطح کلی، برقراری معادله اساسی بیمه در سطح جزئی هم دارای اهمیت است. بیمه‌نامه‌ای که از ریسک بالاتری برخوردار است نسبت به بیمه‌نامه‌ای که ریسک پایین‌تری دارد، باید حق بیمه بیشتری پرداخت نماید. به عنوان مثال، در بیمه آتش‌سوزی، ساختمانی که فاقد تجهیزات اطفای حریق است خسارت مورد انتظار بیشتری دارد نسبت به ساختمانی که چنین تجهیزاتی را دارد. در این شرایط شرکت‌های بیمه با در نظر گرفتن چنین تفاوت ریسکی، حق بیمه را تغییر می‌دهند. اشتباه در درک تفاوت ریسک‌ها، منجر به نرخ‌هایی می‌شود که ممکن است منصفانه نباشند. "نرخ‌گذاری مبتنی بر طبقه‌بندی ریسک"، رویکردی کارآمد برای قیمت‌گذاری منصفانه در محصولات بیمه‌ای می‌باشد.

در این مقاله در بخش اول، اهمیت قیمت‌گذاری منصفانه در کسب و کار بیمه‌ای توضیح داده می‌شود. در بخش دوم سیستم طبقه‌بندی ریسک در بردارنده تعریف سیستم، معیارهای انتخاب فاکتورهای ریسک و محاسبه میزان کارایی سیستم بیان شده است. در آخر نیز نتایج و منابع آورده شده است.

۲- اهمیت قیمت‌گذاری منصفانه

فرض کنید در یک شرکت بیمه، بیمه‌گر برای کلیه ریسک‌ها نرخ میانگینی را در نشر بگیرد در صورتی که سایر شرکت‌های رقیب از متغیرهای نرخ‌گذاری استفاده می‌نمایند به این ترتیب که نرخ‌ها، براساس تفاوت در خسارت مورد انتظار، متفاوت خواهند بود. در این شرایط، شرکت بیمه که با میانگین نرخ کار می‌کند، ریسک‌های خطرناک‌تر را جذب می‌نماید و ریسک‌های کم خطر را بدلیل بالا بودن نرخ از دست می‌دهد. این اتفاق منجر به انتقال توزیع پوتفو به سمت بیمه‌گذاران با ریسک بالا می‌شود بنابراین میانگین نرخ اعمال شده این شرکت را ناکافی می‌نماید از این رو فعالیت شرکت سودآور نخواهد بود. شرکت مجبور است میانگین نرخ خود را افزایش دهد. این افزایش، بیمه‌گذاران با ریسک پایین‌تر را تشویق می‌نماید که شرکت بیمه خود را تعویض نمایند و به سراغ سایر شرکت‌های بیمه که نرخ منصفانه‌تری ارائه می‌دهند بروند. باقی ماندن بیمه‌گذاران با ریسک بالا در شرکت مذکور، باعث می‌شود مجدداً میانگین نرخ آن شرکت سودآور نباشد. این روند منفی تا زمانی که شرکت بیمه نرخ‌های خود را بهبود بخشد، یا ورشکسته شود و یا تصمیم به کاهش تمرکز بر ریسک‌های پرخطر و به دنبال آن افزایش نرخ را نگرفته است، ادامه خواهد یافت. به این فرایند انتخاب نامناسب می‌گویند. البته روند این فرایند به عوامل مختلفی بستگی دارد. اینکه متقاضیان بیمه دانش و آگاهی کامل و دقیقی از تفاوت نرخ در شرکت‌های مختلف بیمه‌ای دارند و اینکه قیمت به تنهایی چقدر بر تصمیم آنها برای خرید بیمه‌نامه تاثیرگذار خواهد بود.

۳- طبقه‌بندی ریسک

اساساً اکچوئری‌ها از طبقه‌بندی ریسک در نرخ‌گذاری زمانی استفاده می‌نمایند که اطلاعات کافی (سابقه خسارت) برای تخمین حق بیمه به صورت انفرادی را نداشته باشند. از این رو به منظور استخراج نرخ حق بیمه، از مکانیزم طبقه‌بندی ریسک استفاده می‌نمایند و بیمه‌گذاران را براساس توانایی ایجاد خسارت مشابه در یک گروه قرار می‌دهند. در اینصورت، اکچوئری نرخ حق بیمه را برای گروه تعیین می‌نماید با این فرض که برای تمام اعضای آن گروه قابل استفاده است. پیشرفت‌های اخیر در صنعت کامپیوتر و فرایندهای تحلیلی، این اجازه را به اکچوئری‌ها داده تا بتوانند مدل‌های پیشرفته خسارت مانند رگرسیون چندگانه، شبکه‌های عصبی و غیره را مدلسازی نمایند. با استفاده از

این مدل‌ها، متوسط خسارت برای گروهی از بیمه‌شدگان که دارای خصوصیات مشابه هستند برآورد می‌شود. مکانیزم نرخ‌دهی علاوه بر طبقه‌بندی، شامل انتخاب پایه معرض خطر، ملاحظات بازار و بیمه‌گری و همچنین نرخ‌دهی ریسک انفرادی است و این عوامل بر مقدار حق بیمه نهایی تاثیرگذار هستند. هدف کلی، تعیین قیمتی منصفانه متناسب با شرایط بیمه‌گذار و پوشش مورد درخواست است. این مهم به طرق مختلفی می‌تواند انجام گیرد که طبقه‌بندی ریسک یک گام از این مکانیزم است که به تعیین قیمتی منصفانه کمک می‌نماید.

۴- انتخاب فاکتورهای ریسک

اولین گام در نرخ‌گذاری بر مبنای طبقه‌بندی، تعیین فاکتورهای ریسک براساس تاثیرپذیری آنها بر ریسک مورد است. به عنوان مثال، فرض کنید شرکت بیمه متوجه شود که در رشته آتش‌سوزی، قدمت ساختمان بر ریسک آتش‌سوزی تاثیرگذار است. بنابراین می‌تواند متغیر سن ساختمان را به عنوان یکی از فاکتورهای تاثیرگذار بر ریسک آتش‌سوزی در زمان تعیین نرخ در نظر بگیرد. هر یک از فاکتورهای ریسک دارای سطوحی هستند که باید اکتوئری میزان تاثیرپذیری (ضرایب) را در هر یک از این سطوح نسبت به سطح پایه محاسبه نماید. در مثال فوق سنین مختلف ساختمان همان سطوح هستند. اولین مرحله در نرخ‌گذاری بر مبنای طبقه‌بندی، انتخاب فاکتورهای ریسک است تا بتوان جمعیت بیمه‌گذاران را براساس ریسک‌های مشابه، به گروه‌های مختلفی تقسیم نمود. به عنوان مثال، در رشته آتش‌سوزی بیمه‌گر باید تصمیم بگیرد که آیا می‌تواند از سن ساختمان، تجهیزات اعلام و اطفاء حریق که جزء مشخصه‌های ریسک هستند، به عنوان فاکتور ریسک در این رشته استفاده کند یا خیر. معیارهای مختلف ارزیابی مناسب بودن فاکتورهای ریسک شامل معیارهای آماری، عملیاتی، اجتماعی و حقوقی می‌باشد.

۵- کارایی سیستم طبقه‌بندی

کارایی به عنوان معیاری برای بررسی صحت سیستم طبقه‌بندی تعریف می‌شود. این معیار به اکتوئری کمک می‌نماید که بدانند تا چه اندازه سیستم طبقه‌بندی خوب کار می‌کند و تا چه اندازه نیاز به بهبود دارد. در این قسمت بررسی می‌شود که چرا و چگونه خسارات بین بیمه‌گذاران تغییر می‌نماید و چگونه می‌توان چنین تغییراتی را اندازه گرفت. اکتوئری ممکن است بدانند به عنوان مثال هزینه‌های گروه الف دو برابر هزینه‌های گروه ب است ولی نمی‌دانند چه میزان تغییرپذیری ممکن است بین اعضای گروه الف و ب وجود داشته باشد. دلیل اصلی طبقه‌بندی ریسک، وجود تغییرپذیری خسارات مورد انتظار از یک بیمه‌گذار به بیمه‌گذار دیگری است. کلید اصلی برای محاسبه کارایی، درک چنین تغییرپذیری می‌باشد. هزینه‌ها تغییر می‌کنند زیرا فراوانی خسارت و شدت خسارت متغیر هستند. در یک سیستم طبقه‌بندی کارآمد، تمام بیمه‌گذاران به درستی طبقه‌بندی می‌شوند و تغییرپذیری آن مشابه تغییرپذیری در جامعه بیمه‌شده است. سیستمی که تغییرپذیری کمتری نسبت به جامعه بیمه‌شده دارد، نمی‌تواند کارآمد باشد چرا که برای دو بیمه‌گذار با شرایط متفاوت، نرخ یکسانی ارائه می‌دهد. یک فاکتور پیچیده، ماهیت تصادفی بیمه است. خسارات نامشخص هستند. اندازه‌گیری تغییرپذیری خسارت با استفاده از حوادثی است که رخ داده‌اند و مربوط به سابقه خسارت می‌باشند. حوادث مشابه احتمالاً دوباره اتفاق نمی‌افتند. در مورد اینکه آیا حوادثی که تاکنون اتفاق افتاده‌اند می‌توانند نماینده خوبی برای حوادث آتی باشند یا خیر نااطمینانی وجود دارد. آینده می‌تواند تغییرپذیری بیشتر یا کمتری نسبت به گذشته داشته باشد.

برای تعیین تغییرپذیری در سیستم طبقه‌بندی انجام گرفته، اطلاعاتی چون ضرایب ریسک طبقات و درصد معرض خطر در هر طبقه مورد نیاز است. فرض می‌شود که ضرایب طبقات، مقادیر مورد انتظار خسارت در هر طبقه می‌باشند. داده‌های مشاهده شده را می‌توان به عنوان نمونه‌ای از فرایند ریسک فرضی در نظر گرفت. مدلسازی می‌تواند به طرق مختلفی صورت گیرد. یکی از مرسوم‌ترین روش‌ها این است که فرض شود، متغیر خسارت در هر یک از بیمه‌نامه‌ها از توزیع پواسن آمیخته تبعیت می‌نماید. بنابراین در نظر بگیرید به ازای هر بیمه‌نامه:

N تعداد خسارت در طول دوره بیمه‌نامه

S مبلغ خسارت به ازای هر خسارت

T خسارت کل در هر بیمه‌نامه

فرض می‌شود که N دارای توزیع پواسن با پارامتر λ باشد. مقادیر زیر را با فرض استقلال تعداد و مبلغ خسارت و برابری میانگین و واریانس در توزیع پواسن ($E(N) = Var(N) = \lambda$) خواهیم داشت:

$$T = \sum_{i=1}^N S_i$$

$$E(T) = E(N) \times E(S)$$

$$Var(T) = E(N)Var(S) + Var(N)E^2(S)$$

$$Var(S) = E^2(S) - E^2(S)$$

$$CV^2(S) = \frac{E(S^2)}{E^2(S) - 1}$$

بنابراین ضریب تغییرات خسارت کل در هر بیمه‌نامه عبارت است از:

$$CV^2(T) = \frac{Var(T)}{E^2(T)} = \frac{E(S^2)}{E^2(S)\lambda} = \frac{1 + CV^2(S)}{\lambda}$$

از این معادله بر می‌آید که ضریب تغییرات خسارات برای فراوانی‌های بیشتر، کوچک‌تر خواهد بود و به ازای مقادیر بیشتر مبلغ خسارت، بزرگتر خواهد بود.

۶- خلاصه و نتیجه‌گیری

در این مقاله، اهمیت سیستم نرخ‌گذاری مبتنی بر طبقه‌بندی به منظور ارائه حق‌بیمه منصفانه در صنعت بیمه ارائه شد. یک از مهمترین مفاهیمی که بر ضرورت وجود یک سیستم طبقه‌بندی تاکید دارد، انتخاب نامناسب می‌باشد. به طوری که در یک بازار رقابتی، استفاده از سیستم طبقه‌بندی نقطه قوت شرکت‌های بیمه تخصصی نسبت به سایر شرکت‌ها می‌باشد و به آنها در مدیریت ریسک و انتخاب مناسب ریسک‌ها کمک می‌نماید. کارایی سیستم‌های نرخ‌گذاری فعلی قابل برآورد است و این برآوردها مبتنی بر اندازه‌گیری تغییرپذیری در خسارات مورد انتظار بین بیمه‌گذاران می‌باشد. اکثر معیارهای موجود در سنجش کارایی سیستم طبقه‌بندی، از معیارهای آماری مانند واریانس و مشابه آن استفاده می‌نمایند. استفاده از معیار واریانس از آنجا که به صورت گسترده در مباحث آماری کاربرد دارد توصیه می‌شود. در این مقاله از معیار واریانس به عنوان معیار کارایی استفاده شده است. به این ترتیب که تغییرپذیری سیستم طبقه‌بندی را با تغییرپذیری کل جامعه بیمه‌شده مقایسه می‌نماید. معیار تغییرپذیری، ضریب تغییرات یعنی نسبت انحراف استاندارد به میانگین است که میزان تغییرپذیری را به ازای هر واحد میانگین محاسبه می‌نماید. کارایی سیستم در نهایت عبارت از نسبت مجذور ضریب تغییرات سیستم طبقه‌بندی به کل جامعه بیمه‌شده است. با استفاده از داده‌های صدور و خسارت بیمه بدنه اتومبیل خودروهای سواری طی سال‌های ۹۲-۹۴، شرایط سیستم نرخ‌گذاری در این رشته بیمه‌ای مورد بررسی قرار گرفت. نتایج نشان می‌دهد که در نظر گرفتن متوسط حق‌بیمه برای کلیه خودروهای سواری از اعتبار بالایی برخوردار نیست و تغییرپذیری بسیار بالایی دارد. از این رو در نظر گرفتن فاکتورهای ریسک تاثیرگذاری که به بهترین نحو بتوانند تغییرپذیری بین بیمه‌گذاران را توصیف نمایند، توصیه می‌شود. در نتایج مشاهده شد که کارایی سیستم نرخ‌گذاری با در نظر گرفتن فاکتور سن وسیله نقلیه و نوع کاربری بهبود می‌یابد. از آنجا که هرچه توزیع ریسک در پورترفوی بیمه‌شده متناسب‌تر باشد، ضریب تغییرات طبقه‌بندی صورت گرفته کمتر خواهد شد لذا میزان بهبود کارایی متناسب با نوع پورترفوی

بیمه‌گر متغیر خواهد بود. پورتنوی مورد مطالعه در فاکتورهای سن وسیله نقلیه و نوع کاربری، توزیع ریسک نسبتاً متناسبی نسبت به ریسک زیرطبقات داشته است لذا این فاکتورها تا حدودی سیستم را بهبود بخشیده‌اند ولی همچنان فاکتورهای ریسک تاثیرگذاری دیگری مانند مشخصات راننده، سیستم خودرو وجود دارند که باید در نظر گرفته شوند تا تغییرپذیری بین بیمه‌گذاران تا جای ممکن در محاسبه حق بیمه لحاظ و حق بیمه منصفانه‌ای ارائه گردد که بدلیل نبود اطلاعات کافی، در این مطالعه لحاظ نشده‌اند. پیشنهاد می‌شود شرکت‌های بیمه به عنوان اولین گام در طراحی سیستم نرخ‌گذاری کارآمد، اقدام به شناسایی و ثبت فاکتورهای ریسک تاثیرگذار به صورت اجباری بنمایند.

مراجع

- [۱] تعرفه بیمه بدنه اتومبیل، آیین‌نامه شماره ۳۳، بیمه مرکزی
- [2] Brown and Gottlieb, 2001, Introduction to Ratemaking and Loss Reserving for Property and Casualty Insurance.
- [3] Casualty Actuarial Society, 2001, Foundations of Casualty Actuarial Science, Fourth Edition.
- [4] SRI International, 1979, Choice of a Regulatory Environment for Automobile Insurance (Final Report).
- [5] Werner, G., Modlin, C., 2010, Basic Ratemaking, Fourth Edition, Casualty Actuarial Society.
- [6] Woll, R.G., 1979, A Study of Risk Assessment, Proceedings of the Casualty Actuarial Society, 66:84-138.

بررسی عملکرد توسعه‌های روش تتا در پیش‌بینی سری‌های زمانی و کاربرد آن در پیش‌بینی داده‌های صنعت بیمه کشور

حمیده مهدوی شیانی^۱، بهزاد منصوری^۲ و رحیم چینی پرداز^۳

^{۱،۲} گروه آمار، دانشگاه شهید چمران اهواز، اهواز
hamidehmahdavi@gmail.com
b.mansouri@scu.ac.ir
chinipardaz_r@scu.ac.ir

چکیده مختصر

هدف از تحلیل سری زمانی، می‌تواند تنها یافتن یک مدل مناسب جهت توصیف ساختار درونی سری زمانی باشد. با این وجود معمولاً بیشتر محققین علاقمند به پیش‌بینی سری‌های زمانی هستند. یکی از روش‌های پیش‌بینی روش تتا است. در این مقاله جدیدترین تعمیم‌های روش تتا بر مبنای فضای حالت که در سال‌های اخیر توسط محققین مطرح شده، بررسی شده و عملکرد این تعمیم‌ها با روش تتا و مدل‌بندی باکس جنکیز با استفاده از یک مجموعه از داده‌های شبیه‌سازی شده مقایسه شده‌اند. در پایان کاربرد این روشها در پیش‌بینی داده‌های صنعت بیمه نشان داده شده است.

کلمات کلیدی

پیش‌بینی، سری‌های زمانی، روش تتا، فضای حالت، داده‌های بیمه.

۱- مقدمه

پیش‌بینی سری‌های زمانی یکی از موضوعات مهم و جذاب در حوزه‌های مختلف علوم اقتصادی، مالی، مدیریتی، مهندسی و ... است. مدل‌های ARIMA و توسعه آنها یعنی مدل‌های SARIMA یکی از شناخته شده‌ترین مدل‌های مورد استفاده در مدل‌بندی و پیش‌بینی سری‌های زمانی می‌باشند. به صورت ساده‌ای این مدل‌ها و زیر شاخه‌های آنها بر پایه این فرض هستند که مدل سری زمانی به صورت خطی بوده و توزیع مشخصی دارد. روش‌های هموارسازی نمایی از دیگر روش‌های موجود برای پیش‌بینی هستند که به خوبی توسعه یافته‌اند. این روش‌ها شامل هموارسازی نمایی ساده، روش براون، روش هالت، روش وینترز و ... بوده که با افزودن پارامتر میرایی از یک سو و مدل‌بندی در قالب فضای حالت از سوی دیگر بسیار توسعه یافته‌اند.

توسعه‌ی روش‌های پیش‌بینی دقیق، قوی و قابل اعتماد برای سری‌های زمانی یک متغیره، زمانی که تعداد زیادی از سری‌های زمانی در مدل سازی و فرایند پیش‌بینی در نظر گرفته می‌شوند، بسیار مهم است. روش تتا به خاطر سادگی آن و عملکرد خوب و شگفت انگیزش توجه محققان را به خود جلب کرده است. این روش توسط آسیکماپولس و نیکولوپولس (۲۰۰۰) معرفی شد و مبتنی بر مفهوم اصلاح خمیدگی مکانی سری زمانی از طریق ضریب تتا است که به طور مستقیم از تفاضل مرتبه‌ی دوم داده‌ها استفاده می‌کند بدین معنی که اگر داده‌های اولیه را با X^{data} نشان دهیم آنگاه تعریف می‌کنیم:

$$X^{new} = \theta \cdot X^{data}$$

که در آن

$$X'' \text{ data} = X_t - 2X_{t-1} + X_{t-2}$$

بنابراین به ازای مقادیر مختلف θ یک مجموعه از سری‌های زمانی جدید که خطوط تنا نامیده می‌شوند، ساخته می‌شوند. هر کدام از خطوط تنا به طور جداگانه برون‌یابی می‌شوند و در نهایت پیش‌بینی‌ها به سادگی ترکیب می‌شوند. آسیکماپولس و نیکولوپولس (۲۰۰۰) برای روش پیشنهادی خود مدلی را ارائه ندادند در نتیجه نمی‌توان برآورد حداکثر درست‌نمایی پارامترها را به دست آورد و فواصل پیش‌بینی تولید کرد. هایندمن و بیلا (۲۰۰۳)، برای روش تنا یک مدل آماری براساس دیدگاه فضای حالت که پیش‌بینی‌هایی برابر با مدل هموارسازی نمایی ساده با رانش (SES-d) داشت را معرفی کردند. این مدل امکان ارائه فواصل پیش‌بینی را فراهم می‌کند. در سال‌های اخیر تعمیم‌های مختلفی از روش تنا در متون جدید پیشنهاد شده‌اند. نیکولوپولس و آسیکماپولس (۲۰۰۵) و پتروپولس و نیکولوپولس (۲۰۱۳) در ارتباط با کاربرد تعداد خطوط بیشتر به منظور استخراج کردن اطلاعات بیشتر از داده‌ها، مطالعاتی انجام داده‌اند. همچنین کنستانتینیدو ای تی ال (۲۰۱۲) و پتروپولس و نیکولوپولس (۲۰۱۳) کاربرد وزن‌های نابرابر در مرحله‌ی ترکیب برای پیش‌بینی‌های نهایی را بررسی کرده‌اند.

اخیرا توماس و نیکولوپولس (۲۰۱۵) یک فرمول بندی نظری جدید را برای کاربرد این روش برای سری‌های زمانی چند متغیره بسط داده و شرایطی را تحت روش تنای دو متغیره بررسی کرده‌اند که انتظار می‌رود بهتر از روش یک متغیره پیش‌بینی کند. فیورسی و همکاران (۲۰۱۶) مدل‌های متفاوتی را برای بهینه‌سازی روش تنا معرفی کردند. آن‌ها روش تنا را از طریق انتخاب بهینه‌ی خط تنایی که نوسان‌های کوتاه مدت را توصیف می‌کند، توسعه داده و ارتباط عملی و نظری بین مدل تازه ارائه شده و روش تنا را مطالعه نمودند. آن‌ها عملکرد بهتر این مدل‌ها نسبت به دیگر مدل‌های هموارسازی نمایی را با به کار بردن یک مجموعه داده‌های حقیقی (M3 competition, مکریداکس و هیبن، ۲۰۰۰) نشان دادند.

در این مقاله ضمن معرفی روش تنا و توسعه‌های آن، ارتباط آن‌ها با مدل‌های فضای حالت مطالعه شده و عملکرد آن‌ها را به کمک شبیه‌سازی با دیگر روش‌های موجود برای پیش‌بینی مقایسه شده است. همچنین کاربرد روش تنا و توسعه‌های آن در پیش‌بینی سری‌های زمانی عملکرد صنعت بیمه نشان داده شده است. این مقاله در پنج بخش سازماندهی شده است. در بخش دوم مقاله به طور مختصر روش تنا معرفی شده و ارتباط آن با روش هموارسازی نمایی با رانش (SES-d) نشان داده شده است. در بخش سوم تعمیم‌های مختلف روش تنا معرفی شده است. بخش چهارم به نحوه برآورد پارامترها اختصاص یافته و سر انجام در بخش پنجم، بخش نهایی، یک مطالعه عددی انجام شده و عملکرد روش‌های پیشنهاد شده با استفاده از شبیه‌سازی با دیگر روش‌ها مقایسه شده‌اند. همچنین در این بخش کاربرد روش تنا و توسعه‌های آن در پیش‌بینی سری‌های زمانی عملکرد صنعت بیمه نشان داده شده است و نتایج تحلیل شده است.

مراجع

- [1] Assimakopoulos, V., and Nikolopoulos, K. (2000). The theta model: a decomposition approach to forecasting. *International Journal of Forecasting*, **16**(4):521-530
- [2] Constantinidou, C., Nikolopoulos, K., Bougioukos, N., Tsiafa, E., Petropoulos, F., and Assimakopoulos, V. (2012). A neural network approach for the theta model. In *Lecture notes in information technology*: Vol. 25. Information engineering (pp. 116–120).
- [3] Fiorucci, J. A., Pellegrini, P. R., Louzada, F., Petropoulos, F., and Koehler, A. B. (2016). Models for optimising the theta method and their relationship to state space models. *International Journal of Forecasting*, **32**:1151-1161.
- [4] Hyndman, R. J., and Billah, B. (2003). Unmasking the theta method. *International Journal of Forecasting*, **19**:287–290.
- [5] Makridakis, S., and Hibon, M. (2000). The M3-competition: results, conclusions and implications. *International Journal of Forecasting*, **16**:451-476.
- [6] Nikolopoulos, K., and Assimakopoulos, V. (2005). Fathoming the theta model. In *25th International symposium on forecasting, ISF, San Antonio, Texas, USA*. unknown.
- [7] Petropoulos, F., and Nikolopoulos, K. (2013). Optimizing Theta model for monthly data. In *Proceedings of the 5th International conference on agents and artificial intelligence*.
- [8] Thomakos, D., and Nikolopoulos, K. (2015). Forecasting multivariate time series with the theta method. *Journal of Forecasting*, **34**:220–229.

مشخصه‌سازی خانواده توزیع‌های مکانی و مقیاسی با استفاده از آنتروپی شانون آماره‌های مرتب

ابراهیم امینی سرشت^۱، مونا شیری^۲

^۱ دانشگاه بوعلی سینا، همدان، ایران

^۲ مدرسه نگاه، همدان، ایران

e.amini64@yahoo.com

چکیده مختصر

هر مجموعه از آماره‌های مرتب از دو توزیع احتمالی متفاوت، ممکن است شامل اطلاع فیشر و اطلاع شانون برابر در مورد پارامتر مجهول باشند. در این مقاله ما به بررسی این سوال اساسی که چگونه می‌توان با استفاده از آنتروپی شانون آماره‌های مرتب از دو توزیع احتمالی، خانواده توزیع‌های مکانی و مقیاسی را مشخص کرد پرداختیم. و شرایط لازم و کافی را برای مشخص نمودن خانواده توزیع‌های مکانی و مقیاسی براساس آنتروپی شانون آماره‌های مرتب بدست می‌آوریم.

کلمات کلیدی

آماره‌های مرتب، آنتروپی شانون، اطلاع فیشر، توزیع‌های مکانی-مقیاسی، مشخصه‌سازی.

بکارگیری داده‌های تلماتیک در بهبود فرآیند انتخاب ریسک و قیمت‌گذاری بیمه اتومبیل

فاطمه نصیری^۱، محسن قره‌خانی^۲

^۱ کارشناس ارشد اکچوئری، شرکت بیمه ملت، تهران، nasiri.fateme24@gmail.com

^۲ دکتری مهندسی صنایع، مدیر اتکایی شرکت بیمه ملت، تهران، gharahkhani@gmail.com

چکیده

بکارگیری حسگرها تأثیر جهانی رشد دهنده‌ای بر صنایع مختلف از جمله صنعت بیمه داشته است. با افزایش استفاده از اینترنت اشیا (IOT)، جریان بسیار عظیمی از داده‌ها تولید می‌شود که این منابع داده‌ای جدید با ترکیب شدن با روش‌های داده‌کاوی می‌توانند فرصت‌های جدیدی را برای خودکار کردن و پشتیبانی از تصمیم‌گیری‌های کسب‌وکار ایجاد کنند.

در فرآیند قیمت‌گذاری و محاسبه حق بیمه در بیمه اتومبیل، مدل‌های کنونی بسیار معمولی هستند و به صورت پهنه برای کاربر نهایی اصلاح نشده‌اند. اما اطلاعات پارامترهای مهمی که اخیراً از سیستم‌های تلماتیک نصب شده بر روی خودروها قابل استخراج است نظیر کیلومتر پیموده شده، رفتار رانندگی و نوع جاده‌های پیموده شده، می‌تواند فرآیند انتخاب ریسک و ظهور مدل‌های جدید تعیین حق بیمه در صنعت بیمه اتومبیل را تسهیل نماید. با وجود این که استفاده از سیستم‌های تلماتیک بیمه‌ای می‌تواند برای بیمه‌گران مزیت رقابتی ایجاد نماید، بیمه مصرف‌محور (UBI) همچنان حوزه‌ای است که به اندازه کافی در آن پژوهش صورت نگرفته است، به خصوص در بازار بیمه ایران که این موضوع کاملاً جدید بوده و بستر لازم برای اجرای آن فراهم نشده است. بدین ترتیب هدف اصلی این مقاله معرفی و ایجاد شناخت بهتر این موضوع از طریق ارائه مفاهیم مهم در حوزه مربوط به بیمه اتومبیل تلماتیک محور، تشریح اصول اساسی این تکنولوژی، معرفی مدل‌های جدید تعیین حق بیمه در بیمه‌های اتومبیل مصرف‌محور و بررسی تأثیر داده‌های تلماتیک در بهبود فرآیند انتخاب ریسک و قیمت‌گذاری بیمه اتومبیل است.

کلمات کلیدی

قیمت‌گذاری بیمه اتومبیل، سیستم‌های تلماتیک بیمه‌ای، اینترنت اشیا (IOT)، بیمه‌های مصرف‌محور (UBI)، PAYD، بیمه مسافت محور (DBI)

۱- مقدمه

۱-۱- استفاده بیمه‌گران از مه داده‌ها

مه داده‌ها یا داده‌های بزرگ عبارتی است که در مورد جمع‌آوری و تحلیل پایگاه‌های داده‌ای بزرگ و دستیابی نتایج مفید حاصل از این داده‌ها به کار می‌رود. تعریف رایج مه داده، تحلیل داده‌های دارای حجم، سرعت و پراکندگی زیاد را دربرمی‌گیرد که روش‌ها و امکانات پردازش داده‌ای معمول و استاندارد برای انجام این کار کافی نیستند. امروزه با در نظر گرفتن سرعت پیشرفت فنی در نرم‌افزار و سخت افزار پردازش داده، مفهوم پایگاه داده‌ای بزرگ هدفی دور از دسترس نیست [۱].

حوزه‌های متعددی در عملیات بیمه‌گری وجود دارند که مه داده می‌تواند در آن‌ها مورد استفاده قرار گیرد. اولین نقش شهودی مه داده در عملیات بیمه‌ای انطباق دقیق‌تر حق‌بیمه‌ها و خسارات در فرآیند قیمت‌گذاری و بیمه‌گری با استفاده از داده‌های سیستم‌های تلماتیک بیمه‌ای است. ابزارهای تلماتیک، داده‌های زمان واقعی را در مورد رفتار رانندگی بیمه‌گذار جمع‌آوری می‌کنند. به علاوه، مه داده‌ها ابزار جدیدی را برای بهبود سوابق خسارت مشتری، کاهش خلاء بین پوشش‌ها، کاهش تقلب و بهبود کارایی فرآیندها فراهم می‌کنند.

۱-۲- اهمیت بیمه بر مبنای سیستم‌های تلماتیک

صنعت بیمه لاقلاً در مورد بیمه اتومبیل به مرحله بلوغ چرخه عمر خود رسیده است. رقابت شدید و کمبود تنوع محصول منجر به کاهش قابل توجه درآمد برای اکثر شرکت‌های بیمه فعال در بازار بیمه می‌شود، به ویژه در بازارهای کمتر توسعه یافته که در آن‌ها بیمه اتومبیل مهم‌ترین بخش از پرتفوی محصولات بیمه‌ای شرکت‌ها را تشکیل می‌دهد، این موضوع ریسک قابل توجهی را بر تحمل‌پذیری کسب‌وکار آن‌ها وارد می‌کند. فشار رو به افزایش رقابت و تلاش برای دستیابی به کاهش قیمت هم برای بیمه‌گر و هم برای بیمه‌گذار منجر به تغییر پارادایم بیمه‌های سنتی با پیشنهاد بیمه‌های جدید مصرف‌محور (که از این به بعد UBI نیز گفته می‌شود) شده است. UBI مبتنی بر هزاران داده استخراج شده از ابزارهای تلماتیک نصب شده درون خودرو نظیر (IVDR) است که این داده‌ها شامل کیلومتر پیموده شده، سرعت، موقعیت مکانی، زمان، کل زمان سفر، نیروی جاذبه (G)، ترمز گرفتن‌ها، دور زدن‌ها و تغییر مسیرها می‌شوند [۲، ۳]. بدین ترتیب، بیمه‌گر قادر به جمع‌آوری داده‌های مربوط به رفتار رانندگی هر مشتری است و حق‌بیمه راننده با توجه به رفتار رانندگی خود او اصلاح می‌شود. این داده‌ها می‌توانند پیش‌بینی کننده باشند و از طریق تحلیل رفتار رانندگان برای شرکت‌های بیمه مزیت‌های رقابتی ایجاد نمایند. به علاوه، ابزارهای تلماتیک بیمه‌ای که کاملاً برای مشتریان شفاف هستند، می‌توانند شناسایی تقلب در فرآیند رسیدگی به خسارت را تسهیل نمایند و یا حداقل به شرکت بیمه فرصت اصلاح حق‌بیمه را با توجه به پروفایل ریسک پویای هر بیمه‌گذار بدهند [۲، ۳]. از طریق این بازخوردها و حتی تنها آگاهی راننده از این موضوع که بیمه‌گر در حال پایش رفتار رانندگی وی است، تصادفات اتومبیل، فوت و آسیب‌دیدگی‌های همراه با آن قطعاً کاهش می‌یابد. بنابراین، استفاده از داده‌های رفتار رانندگی سیستم‌های تلماتیک می‌تواند ارزش افزوده‌ای برای شرکت بیمه با بهبود پروفایل ریسک مشتریان بیمه اتومبیل ایجاد نماید.

محصولات UBI در برخی از بازارها نظیر ایران یا هنوز معرفی نشده‌اند و یا بسیار جدید هستند و برای معرفی آن‌ها به بازارهای آتی، شرکت‌های بیمه مجبور به تلاش بسیار جهت افزایش اقبال عمومی مردم نسبت به آن‌ها خواهند بود. در چند سال آینده، تعداد شرکت‌های بیمه استفاده کننده از سیستم‌های تلماتیک در دنیا با نرخ رشد ترکیبی ۹۰٪ از ۱,۸۵ میلیون

دلار در سال ۲۰۱۲ به ۸۹ میلیون دلار در سال ۲۰۱۷ رشد خواهد کرد [۴]. این داده‌ها نشان می‌دهند که شرکت‌های بیمه از این بازارهای نوظهور منتفع خواهند شد. لذا در این مقاله ارزش داده‌های تلماتیک در بهبود فرآیند انتخاب ریسک، مزایای اجتماعی، اقتصادی و زیست محیطی استفاده از بیمه‌های مصرف محور برای بیمه‌گران و بیمه‌گذاران و مدل‌های قیمت‌گذاری بیمه‌های اتومبیل مصرف محور با استفاده از داده‌های سیستم‌های تلماتیک مورد بررسی قرار می‌گیرد.

۲- مدل‌های پیشگو

در گذشته، روش‌های محاسبه حق‌بیمه اساساً مبتنی بر فاکتورهای عمومی بودند. ویژگی‌های خاص وسیله نقلیه و داده‌های اجتماعی- جمعیت‌شناسی مربوط به رانندگان تنها ورودی برای محاسبات بودند که این رویکرد می‌توانست از طریق وارد کردن سابقه خسارت بیمه‌گذار بهبود یابد. به علاوه، با بکارگیری مدل تشویق و تنبیه (پاداش و جریمه) مبتنی بر خسارات گذشته، بیمه‌گران می‌توانند سطح سازگارتری از میزان ریسک را تعریف کنند. علی‌رغم اقبال گسترده چنین مدل‌هایی، آن‌ها همچنان محدودیت‌هایی در برآورد سطح واقعی ریسک بیمه‌گذار دارند. به طور خاص، در معرض ریسک قرار گرفتن (یعنی این که کجا، چه زمانی و به چه میزان بیمه‌گذار با وسیله نقلیه خود رانندگی می‌کند) همچنان در نظر گرفته نمی‌شود. در صدر فهرست موضوعات کلیدی که بیمه‌گران مشغول به تلاش برای پذیرش یا توسعه برنامه‌های UBI مبتنی بر تلماتیک با آن مواجه می‌شوند، توانایی ساخت مدل‌های هزینه خسارت پیش‌گو است که بتواند رفتارهای نشان دهنده رانندگی ناایمن وسایل نقلیه را شناسایی کند. مدل‌های هزینه خسارت کنونی برای محصولات UBI مبتنی بر تلماتیک به صورت گسترده دو نوع هستند. نوع اول مدل‌ها بر مبنای کل مسافت پیموده شده، ساعات شبانه روز و مجموعه‌ای از پیشامدهای رانندگی از پیش تعریف شده هستند. اما رویکرد دوم بر مبنای جمع‌آوری داده‌های بسیار جزئی و ریزتر در مورد استفاده از وسیله نقلیه بر مبنای ثانیه به ثانیه و سپس استفاده از این جزئیات اضافی برای بررسی و تحقیق در مورد قدرت پیش‌گویی یک میزبان (هاست) و ویژگی‌های رانندگی به صورت بسیار مفهومی است. اگر محقق یک سری از آستانه‌ها را بر مبنای این که مثلاً چند درصد از تغییرات واقعاً نشان دهنده رفتار خاصی هستند، انتخاب کند و سپس اعتبارسنجی کند که کدام پیشامدهای آستانه‌ای بیشترین همبستگی را با خسارت‌های بیمه شده واقعی دارند، می‌تواند پیشامدی را که به قدرت پیش‌گویی مدل هزینه خسارت موجود اضافه می‌کند، شناسایی نماید. این نوع تحقیق و اصلاح پیوسته می‌تواند به ایجاد مدل‌های پیش‌گوتر در طول زمان منجر شود [۵].

چالشی که برای بدنه قانون‌گذاری وجود دارد ایجاد توازن بین میل به حفظ محرمانگی اطلاعات در مقابل ارزش اجازه دادن به مشتریان برای پیوستن به برنامه‌هایی است که از اطلاعات آن‌ها می‌توان برای بهبود مدل قیمت‌گذاری استفاده نمود. این مدل‌ها نیز به نوبه خود به تغییر رفتار مشتریان به سمت ایجاد هزینه خسارت کمتر، مصرف سوخت کمتر و نجات جان انسان‌ها هدایت خواهند کرد. ثابت شده است که استفاده از داده‌های واقعی مربوط به رانندگی به میزان قابل توجهی پیش‌گویی بیشتری از هزینه‌های خسارت نسبت به متغیرهای شاخصی دارد که برای نرخ‌گذاری بیمه اتومبیل مورد استفاده قرار می‌گیرند.

۳- نتیجه‌گیری

از گذشته تاکنون، داده یکی از بزرگترین و با ارزش‌ترین دارایی‌های شرکت بیمه به شمار می‌رود. امروزه به دلیل فشارهای رقابتی و تلاش بیمه‌گر و بیمه‌گذار برای کاهش قیمت محصولات بیمه‌ای که منجر به تغییر پارادایم بیمه‌های سنتی با پیشنهاد بیمه‌های جدید مصرف محور شده است، بیمه‌گران به منظور حفظ سودآوری و رقابت خود در بازار، به دنبال

سرمایه‌گذاری بر روی قابلیت جمع‌آوری، ذخیره‌سازی، مدیریت و تحلیل مقادیر قابل توجهی از داده‌های متغیر برای حل مسائل پیچیده هستند. صنعت بیمه اتومبیل با وجود بیمه‌های مصرف‌محور مبتنی بر سیستم‌های تلماتیک به سرعت آماده تغییر بالقوه کسب‌وکار بیمه و تبدیل شدن به یک صنعت مه‌داده است.

برای این‌که بیمه‌گران در طول زمان بتوانند در بازار جدید UBI تلماتیک رقابت کرده و به عنوان رقیب باقی بمانند، باید قادر به جمع‌آوری و تحلیل داده‌های درست و غنی استخراج شده از سیستم‌های تلماتیک بیمه‌ای باشند که این داده‌ها می‌توانند فرآیند انتخاب ریسک و ظهور مدل‌های جدید تعیین حق‌بیمه در صنعت بیمه اتومبیل را تسهیل نموده و از طریق تحلیل رفتار رانندگان، برای شرکت‌های بیمه مزیت‌های رقابتی ایجاد نمایند. مرتبط نمودن هر چه بیشتر و نزدیک‌تر حق‌بیمه به عملکرد واقعی راننده به بیمه‌گران اجازه می‌دهد تا حق‌بیمه‌ها را با دقت بسیار بالاتری محاسبه کنند.

با وجود این‌که استفاده از برنامه‌های UBI مزایای اجتماعی، اقتصادی و محیط زیستی بسیاری برای شرکت‌های بیمه، مشتریان و جامعه دارند و می‌توانند برای بیمه‌گران مزیت رقابتی ایجاد نماید، بیمه مصرف‌محور (UBI) همچنان حوزه‌ای است که به اندازه کافی در آن پژوهش صورت نگرفته است، به خصوص در بازار بیمه ایران که این موضوع کاملاً جدید بوده و بستر لازم برای اجرای آن فراهم نشده است. بدین ترتیب در این مقاله در راستای ایجاد شناخت بهتر این موضوع، مفاهیم مهم در حوزه مربوط به بیمه اتومبیل تلماتیک محور و انواع آن‌ها معرفی شده و به تشریح اصول اساسی تکنولوژی این سیستم‌ها، معرفی مدل‌های جدید تعیین حق‌بیمه در بیمه‌های اتومبیل مصرف‌محور و بررسی تأثیر داده‌های تلماتیک در بهبود فرآیند انتخاب ریسک و قیمت‌گذاری بیمه اتومبیل پرداخته شده است.

منابع

- [1] ABI Research (2012), ABI Research, 89 Million Insurance Telematics Subscribers Globally by 2017.
- [2] Husnjak, S., Peraković, D., Forenbacher, I., & Mumdziev, M. (2015). Telematics system in usage based motor insurance. *Procedia Engineering*, 100, 816-825.
- [3] Vaia, G., Carmel, E., DeLone, W., Trautsch, H., & Menichetti, F. (2012). Vehicle Telematics at an Italian Insurer: New Auto Insurance Products and a New Industry Ecosystem. *MIS Quarterly Executive*, 11(3).
- [4] Powell, L. (2017). Big Data and Regulation in the Insurance Industry.
- [5] National Association of Insurance Commissioners (2014), Usage Based Insurance and Telematics, 2014.

معیار ریسک بر اساس آنتروپی تجمعی ماکزیمم آماره‌ی ترتیبی

مهسا صالحی^۱، سعید طهماسبی^۲

^۱ گروه آمار، دانشگاه خلیج فارس، بوشهر،
m.salehi72@ymail.com

^۲ گروه آمار، دانشگاه خلیج فارس، بوشهر،
tahmasebi@pgu.ac.ir

چکیده

انحراف معیار $\sigma(X)$ یک معیار ریسک متداول در بیمه است. استفاده از این معیار در برخی از مواقع مناسب نمی‌باشد. لذا معیارهای اندازه‌گیری جدیدی مطرح شده‌اند که در مقایسه با این معیار دقیق‌تر هستند و قابلیت‌های بیشتری نیز دارند. در این مقاله ما به معرفی آنتروپی تجمعی ماکزیمم آماره‌ی ترتیبی به‌عنوان یک معیار جدید ریسک می‌پردازیم و سپس این معیار ریسک پیشنهادی را با سایر معیارهای ریسک متعارف در بیمه مورد مقایسه قرار می‌دهیم.

کلمات کلیدی

آنتروپی تجمعی، برآورد تجربی، ماکزیمم آماره‌ی ترتیبی، معیار ریسک.

۱-مقدمه‌ای بر آنتروپی تجمعی ماکزیمم آماره‌ی ترتیبی

آنتروپی تجمعی متغیر تصادفی نامنفی X توسط دی‌کرشنزو و لانگوبردی (2009) به‌صورت زیر معرفی شده‌است.

$$CE(X) = -\int_0^{\infty} F(x) \log F(x) dx,$$

که در آن، $F(x) = P(X \leq x)$ تابع توزیع تجمعی متغیر تصادفی X است. اگر دنباله‌ای از متغیرهای تصادفی X_1, \dots, X_n را در اختیار داشته‌باشیم و این دنباله را به‌صورت غیرنزولی مرتب کنیم، آن‌گاه $X_{1:n}, \dots, X_{n:n}$ را آماره‌های ترتیبی می‌نامند. واضح است که $X_{1:n}$ مینیمم آماره‌ی ترتیبی و $X_{n:n}$ ماکزیمم آماره‌ی ترتیبی است. آماره‌های ترتیبی در برآورد آنتروپی، قابلیت اطمینان، کنترل کیفیت و ... کاربرد زیادی دارند. براتیور (۲۰۱۰) آنتروپی باقیمانده‌ی تجمعی مینیمم آماره‌ی ترتیبی را معرفی و خواص آن را بررسی نمود. اکنون آنتروپی تجمعی متغیر تصادفی $X_{n:n}$ را می‌توان به‌صورت زیر معرفی کرد.

$$CE(X_{n:n}) = -n \int_0^{\infty} (F(x))^n \log F(x) dx.$$

ابتدا برخی از خواص این اندازه اطلاع را بیان می‌کنیم و سپس در برخی از توزیع‌های مشهور، این معیار پیشنهادی را با سایر معیارهای ریسک متعارف در بیمه مقایسه می‌کنیم.

گزاره ۱. اگر X یک متغیر تصادفی نامنفی با تابع توزیع پیوسته‌ی $F(x)$ و آنتروپی تجمعی $CE(X)$ باشد، آن‌گاه

$$1) CE(X_{n:n}) \leq nCE(X),$$

$$2) CE(X_{n:n}) \geq n \sum_{t=0}^n (-1)^t \frac{n!}{t!(n-t)!} \int_0^{\infty} (\bar{F}(x))^{t+1} dx.$$

اگر X یک متغیر تصادفی نامنفی با میانگین متناهی غیرصفر باشد، آن‌گاه آنتروپی تجمعی استاندارد به صورت $NCE(X) = \frac{CE(X)}{E(X)}$ تعریف می‌شود. مشابه این تعریف $NCE(X_{n:n})$ را به صورت زیر تعریف می‌کنیم.

تعریف ۱. اگر X یک متغیر تصادفی نامنفی با آنتروپی تجمعی ماکزیمم آماره‌ی ترتیبی $CE(X_{n:n})$ باشد، آن‌گاه آنتروپی تجمعی استاندارد ماکزیمم آماره‌ی ترتیبی به صورت $NCE(X_{n:n}) = \frac{CE(X_{n:n})}{E(X_{n:n})}$ تعریف می‌شود.

۲- معیار ریسک بر اساس $CE(X_{n:n})$

انحراف معیار $\sigma(X)$ یک معیار ریسک متداول در بیمه است. رمسای (1995) استفاده از $\sigma(X)$ را برای توزیع‌های چوله‌ی دم‌بلند نامناسب دانست. به همین دلیل ونگ (1998) اندازه ریسک انحراف راست‌دنباله‌ای را به صورت زیر معرفی کرد.

$$D(X) = \int_0^{\infty} \sqrt{F(t)} dt - E(X).$$

اخیرا سارا کوس و توماج (2017)، آنتروپی باقیمانده‌ی تجمعی و حالت تعمیم یافته‌ی آن را به عنوان یک معیار ریسک معرفی کرده‌اند. آنها یکی از محدودیت‌های $\sigma(X)$ را عدم وجود آن برای توزیع‌های دم‌سنگین با میانگین $E(X) < \infty$ و $E(X^2) = \infty$ دانستند. $CE(X_{n:n})$ برخی از خواص $\sigma(X)$ و $D(X)$ را حفظ می‌کند. بدین منظور در ادامه برای برخی از توزیع‌های معروف، $CE(X_{n:n})$ را با اندازه ریسک‌های $\sigma(X)$ و $D(X)$ مقایسه می‌کنیم.

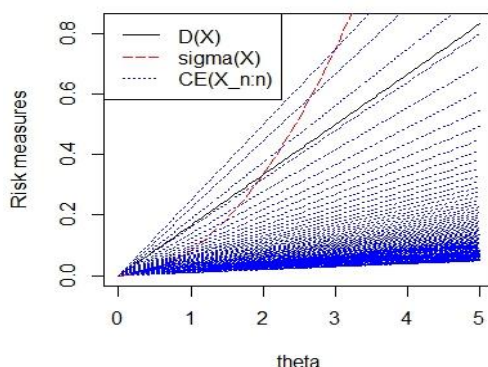
مثال ۱. اگر X دارای توزیع یکنواخت روی بازه‌ی $(0, \theta)$ با تابع توزیع $F(x) = \frac{x}{\theta}$ باشد، آنگاه انحراف معیار، انحراف راست‌دنباله‌ای و

$$.CE(X_{n:n}) = \frac{n\theta}{(n+1)^2} \text{ و } D(X) = \frac{\theta}{6}, \sigma(X) = \frac{\theta^2}{12}$$

در جدول 1 این سه معیار ریسک را برای $\theta = 1$ و مقادیر n مقایسه کرده‌ایم.

$\sigma(X) < D(X) < CE(X_{n:n})$	$n = 1, 2, 3$
$\sigma(X) < CE(X_{n:n}) < D(X)$	$n = 4, \dots, 9$
$CE(X_{n:n}) < \sigma(X) < D(X)$	$n \geq 10$

جدول ۱: مقایسه‌ی $\sigma(X)$ ، $D(X)$ و $CE(X_{n:n})$ در توزیع یکنواخت روی بازه‌ی $(0, 1)$



شکل ۱: مقایسه‌ی معیارهای ریسک $\sigma(X)$ ، $D(X)$ و $CE(X_{n:n})$ در توزیع یکنواخت روی بازه‌ی $(0, \theta)$ برای $0 \leq \theta \leq 5$ و $n = 1, 2, \dots$

علاوه‌براین، $CE(X_{1:n}) > \dots > CE(X_{n:n})$ و زمانی که n به بی‌نهایت میل کند مقدار حد $CE(X_{n:n})$ به سمت صفر میل می‌کند.

مثال ۲. اگر X دارای توزیع نمایی با میانگین $\frac{1}{\lambda}$ باشد، آن‌گاه انحراف معیار، انحراف راست‌دنباله‌ای و آنتروپی تجمعی ماکزیمم آماری

ترتیبی آن به ترتیب برابر است با $D(X) = \sigma(X) = \frac{1}{\lambda}$ و $CE(X_{n:n}) = \frac{n}{\lambda} \left(\frac{\pi^2}{6} - \sum_{j=1}^n \frac{1}{j^2} \right)$. مشهود است که

$$\frac{CE(X_{n:n})}{\sigma(X)} = \frac{CE(X_{n:n})}{D(X)} = r_n, \quad r_n = n \left(\frac{\pi^2}{6} - \sum_{j=1}^n \frac{1}{j^2} \right).$$

همواره $r_n < 1$ است. پس به ازای $n = 1, 2, \dots$ داریم $CE(X_{n:n}) < \sigma(X) = D(X)$. در این توزیع همواره $CE(X_{1:n}) < \dots < CE(X_{n:n})$.

۳- برآورد تجربی آنتروپی تجمعی ماکزیمم آماری ترتیبی

در این بخش برای آنتروپی تجمعی ماکزیمم آماری ترتیبی، یک برآورد تجربی به‌دست آورده و برای توزیع یکنواخت بر بازه‌ی $(0, 1)$ با استفاده از شبیه‌سازی مقدار $CE(X_{n:n})$ را برآورد کرده‌ایم.

تعریف ۲. اگر X یک متغیر تصادفی پیوسته و نامنفی با تابع توزیع $F(x)$ و آنتروپی جمعی $CE(X)$ باشد، آن‌گاه برآورد تجربی آنتروپی جمعی ماکزیمم آماره‌ی ترتیبی X را به صورت زیر تعریف می‌کنیم:

$$\hat{CE}(X_{n:n}) = -n \sum_{j=1}^{n-1} U_{j+1} \left(\frac{j}{n}\right)^n \log\left(\frac{j}{n}\right),$$

که در آن $U_1 = X_{1:n}$, $U_i = X_{i:n} - X_{i-1:n}$, $i = 2, 3, \dots, n$

در توزیع‌های مختلف با استفاده از شبیه سازی مقدار آنتروپی جمعی ماکزیمم آماره‌ی ترتیبی را می‌توان برآورد کرد. بعنوان مثال در یک نمونه‌ی صدتایی از توزیع یکنواخت، برآورد آنتروپی جمعی و آنتروپی جمعی ماکزیمم آماره‌ی ترتیبی برابر است با

$$\hat{CE}(X) = 0.23, \quad \hat{CE}(X_{n:n}) = 0.0088.$$

مراجع

- [1] Baratpour, S. (2010). Characterizations based on cumulative residual entropy of first-order statistics. *Communications in Statistics—Theory and Methods*, 39(20), 3645-3651.
- [2] Di Crescenzo, A., & Longobardi, M. (2009). On cumulative entropies. *Journal of Statistical Planning and Inference*, 139(12), 4072-4087.
- [3] Psarrakos, G., & Toomaj, A. (2017). On the generalized cumulative residual entropy with applications in actuarial science. *Journal of Computational and Applied Mathematics*, 309, 186-199.
- [4] Ramsay, C. M. (1995). Loading gross premiums for risk without using utility theory. *Insurance Mathematics and Economics*, 2(16), 169-170.
- [5] Wang, S. (1998). An actuarial index of the right-tail risk. *North American Actuarial Journal*, 2(2), 88-101.

چکیده پوسترها

آزمون ایستایی داده‌های فضایی - زمانی

علی محمدیان مصمم^۱، الهه شریفی^۲

^۱ دانشیار، گروه آمار، دانشگاه زنجان، زنجان،
a.m.mosammam@znu.ac.ir

^۲ دانشجوی کارشناسی ارشد، دانشگاه زنجان، زنجان،
E_sharifi@znu.ac.ir

چکیده مختصر

بیشتر پدیده‌های تصادفی در علوم زیست‌محیطی و ژئوفیزیک دارای مشاهداتی هستند، که داده‌ها هم از نظر موقعیت مکانی و هم از نظر موقعیت زمانی به هم وابسته‌اند. برای تحلیل این داده‌ها نیاز به فرضیاتی داریم که یکی از مهمترین این فرضیات، فرض ایستایی است. هدف ما در این مقاله معرفی یک آزمون برای ایستایی فرایندهای فضایی-زمانی در قلمروی فرکانس است. ایده اصلی این آزمون این است که تبدیل فوریه یک فرایند تصادفی ایستا تقریباً ناهمبسته خواهد بود؛ درحالی‌که تبدیل فوریه یک فرایند تصادفی نایستا همبسته است.

کلمات کلیدی

تبدیل فوریه، فرایندهای ایستا، تابع چگالی طیفی، فرایندهای نایستا

الگوهای پرداخت وابسته به ذخایر در بیمه‌های عمر و درمان

پریسا دنیادیده^۱، امین حسن‌زاده^۲

^۱ گروه بیم‌سنجی، دانشگاه شهید بهشتی، تهران
p.donyadideh@mail.sbu.ac.ir

^۲ گروه بیم‌سنجی، دانشگاه شهید بهشتی، تهران
am_hassanzadeh@sbu.ac.ir

چکیده مختصر

ذخایر، یکی از مفاهیم مهم در بیمه‌ی عمر است که هریک از انواع آن برای هدف خاصی استفاده می‌شود. ذخایر، شاخصی است که بیمه‌گر به عنوان برآوردی از تعهدات خود به بیمه‌شدگان بیمه‌های عمر و مستمری، افشاء می‌کند. در بیمه‌های عمر پرداخت‌های وابسته به ذخایر رایج هستند، برای مثال در صورتی که بیمه‌گذار خود را بازخرید کند مزایای پرداختی به وی، معمولاً برابر با درصدی از ذخیره انباشته شده در نظر گرفته می‌شود. در بیمه‌ی عمر معمولی محاسبات مربوط به ذخایر پیچیده نیست ولی در مدل‌های چند وضعیتی شرایط بسیار متفاوت است. تحت مدل چند وضعیتی، مزایایی که از طرف بیمه‌گر به بیمه‌گذار پرداخت می‌شود به وضعیت بیمه‌گذار در آن زمان بستگی دارد، از این رو مزایای نوشته شده در قرارداد وابسته به باقی ماندن بیمه‌گذار در یک وضعیت یا انتقال بین وضعیت‌ها است.

تحت فرض مارکوفی، معادله‌ی دیفرانسیل تیل، دینامیک ذخیره انباشته شده را توصیف می‌کند به طوری که می‌توان آن را با استفاده از روش‌های عددی مانند اویلر حل کرد و ذخایر مورد نظر را بدست آورد. با این وجود، هنگامی که مزایا با توجه به ذخایر بیان می‌شود، حل معادلات دیفرانسیل پیچیده می‌شود. نظریه معروف کانتلی تضمین می‌کند که تحت شرایط مناسب، هزینه‌ی بازخرید می‌تواند در محاسبات ذخیره نادیده گرفته شود به شرط آن که مقدار آن برابر ذخیره باشد. این نظریه، هم از دیدگاه آینده‌نگر و هم از دیدگاه گذشته‌نگر صحیح است.

در این مقاله الگوهای پرداخت وابسته به ذخایر در وضعیت بازخریدی بیمه‌گذار بررسی می‌شود. وابستگی پرداخت‌های بازخرید به ذخایر به صورت خطی در نظر گرفته می‌شود، به طوری که فرض می‌کنیم این پرداخت‌ها برابر ذخیره‌ی آینده‌نگر منهای یک هزینه‌ی نسبی و یک هزینه‌ی ثابت β و α است. بدین ترتیب با استفاده از الگوهای پرداخت مزایا که توابعی خطی از ذخایر هستند، عبارت‌های صریحی برای ذخایر مربوط بدست می‌آید.

پرداخت‌هایی که وابستگی غیر خطی با ذخایر دارند، از دیگر الگوهای پرداخت مورد بررسی در این مقاله هستند، به طوری که مزایای پرداخت شده به مقدار بیشینه‌ی ذخیره انباشته شده و مقدار تضمین شده بستگی دارد. برای حل معادله‌ی دیفرانسیل مربوط از روش‌های عددی مناسب استفاده شده است.

کلمات کلیدی

ذخیره ریاضی، بیمه‌ی عمر، مدل‌های چند وضعیتی، فرایند مارکوف، مقدار بازخرید، قضیه‌ی کانتلی.

اندازه نادرستی در آماره‌های ترتیبی

محمد فهیم شریفی^۱، ملیحه عباس نژاد^۲

^۱ دانشجوی کارشناسی ارشد آمار ریاضی، دانشگاه حکیم سبزواری، سبزوار
M.fahim.sharifi67@gmail.com

^۲ استادیار، گروه آمار، دانشگاه حکیم سبزواری، سبزوار
M.abbasnejad@hsu.ac.ir

چکیده مختصر

در این مقاله اندازه نادرستی میان توزیع i -امین آماره ترتیبی و توزیع جامعه را در نظر می‌گیریم و نشان می‌دهیم که اندازه نادرستی، تابع توزیع جامعه را به طور یکتا مشخص می‌کند. همچنین برخی از ویژگی‌های اندازه ارائه شده را مورد بررسی قرار داده و سپس کران‌هایی برای اندازه نادرستی می‌یابیم و میانگین اندازه نادرستی را محاسبه می‌کنیم.

کلمات کلیدی

آنتروپی، فاصله نسبی کولبک-لیبلر، مشخصه‌سازی.

بررسی تجمع هم‌زمان سرمایه‌گذاری مستقیم خارجی در خدمات و صنایع تولیدی با استفاده از رویکرد آمار فضایی و اقتصاد فضایی

سپیده عظیمی

عضو هیات علمی گروه آمار، دانشگاه پیام نور، همدان،
a.azimi@pnu.ac.ir

چکیده مختصر

در این مقاله از آمار فضایی و اقتصاد فضایی جهت تحلیل تجمع هم‌زمان سرمایه‌گذاری مستقیم خارجی (FDI) در خدمات تولیدکننده و FDI در صنایع تولیدی ۲۴ استان چین در طول سال‌های ۲۰۰۴ تا ۲۰۱۱ استفاده شده است. تعامل مشترکی بین تجمع FDI خدمات تولیدکننده و FDI صنایع تولیدی وجود دارد، FDI صنایع تولیدی در دوره‌های تاخیری بطور قابل ملاحظه‌ای تجمع FDI خدمات تولیدکننده در دوره‌های جاری را القاء می‌کند و همچنین FDI خدمات تولیدکننده در دوره‌های تاخیری بطور قابل ملاحظه‌ای باعث تجمع FDI صنایع تولیدی در دوره‌های جاری می‌شود. از نقطه نظر FDI صنایع تولیدی حوضه‌های اقتصاد کلان، صنایع تولیدی سطح بالا و هزینه پایین نیروی کار ارجحیت دارند در حالی که FDI خدمات تولیدکننده حوضه‌های صنایع خدماتی بزرگ مقیاس، صنایع تولیدی با سطح بالای توسعه‌یافتگی و سرمایه انسانی را ترجیح می‌دهد.

کلمات کلیدی

سرمایه‌گذاری مستقیم خارجی، آمار فضایی، اقتصاد فضایی.

برآورد مدل رگرسیون وقتی ضرایب مدل تابعی از یک متغیر کمکی است

محدثه صفری زاده^۱، محمد حسین دهقان^۲

^۱ دانشگاه سیستان و بلوچستان، دانشکده علوم، زاهدان،
Mohade3.safari71@gmail.com

^۲ دانشگاه سیستان و بلوچستان، دانشکده علوم، زاهدان
mhdehghan2000@gmail.com

چکیده مختصر

در این مقاله به طور مختصر به ترکیب، برازش و کاربرد مدل‌های رگرسیونی می‌پردازیم. فرض کنید مدل غیرخطی $Y_i = f(X_i, \theta) + \varepsilon_i$ $i=1, 2, \dots, n$ که در آن برداری از پارامترهای مدل و متغیر کمکی S_j $j=1, 2, \dots, J$ باشد. به ازای مقداری از متغیر S_j ، متغیر پاسخ Y^j یا $\hat{\theta}^j = \begin{pmatrix} \hat{\theta}_1^j \\ \vdots \\ \hat{\theta}_p^j \end{pmatrix}$ را برآورد می‌کنیم. حال مدل خطی یا غیر خطی موجود بین متغیر کمکی S و پارامترها را برآورد می‌کنیم. با جایگذاری این مدل به جای پارامتر مربوطه، یک مدل ترکیبی کلی به دست می‌آوریم که می‌تواند مدل اصلی را به ازای هر مقدار دلخواه از S برآورد نماید. بدین منظور از شبیه‌سازی و نرم‌افزار R استفاده شده است. این روش را برای داده‌های واقعی الکتروشمی بکار برده و مدل ترکیبی مناسب را به دست آورده‌ایم و از مدل حاصل به راحتی می‌توان میزان شدت جریان و بار خارج شده در فواصل زمانی دلخواه محاسبه کرد.

کلمات کلیدی

رگرسیون غیر خطی، برازش مدل، داده‌های الکتروشمی، شبیه‌سازی

بررسی مدل‌های بیزی پواسون چندمتغیره در نرخ‌گذاری بیمه

ساجده کلانتر^۱، مهسا فرخی^۲، محیا پورمصطفی^۳، طاهره دارابیان^۴، فاطمه علی محمدی^۵

^۱ گروه آمار، دانشگاه علامه طباطبایی، تهران،
s.kalantar@atu.ac.ir

^۲ گروه آمار، دانشگاه علامه طباطبایی، تهران،
m.farrokhi@atu.ac.ir

^۳ گروه آمار، دانشگاه علامه طباطبایی، تهران،
mahia.pourmostafa@gmail.com

^۴ گروه آمار، دانشگاه علامه طباطبایی، تهران،
darabian899@gmail.com

^۵ گروه آمار، دانشگاه علامه طباطبایی، تهران،
f.alimohammadi@atu.ac.ir

چکیده مختصر

در این مقاله، ضمن معرفی مدل‌های پواسون چندمتغیره با کواریانس مشترک و کواریانس کامل بین جفت متغیرها، و مدل‌های صفر-افزون؛ ارتباط بین انواع ادعاهای مختلف برای بهبود روش نرخ‌گذاری، ارائه می‌شود. همچنین با در نظر گرفتن مدل‌های صفر-افزون در دو مدل قبلی، مقادیر بزرگ‌تر از صفر و بیش‌پراکنش محاسبه می‌شود. در نهایت در قسمت رویکرد بیزی، امکان برآورد تعداد زیادی پارامتر برای مدل‌های پیچیده و استفاده از اطلاعات قبلی امکان‌پذیر شده و ویژگی‌های خاص مورد نظر محاسبه می‌شوند.

کلمات کلیدی

نرخ‌گذاری بیمه، مدل‌های رگرسیونی پواسون چندمتغیره.

بررسی اهمیت نگاه آینده‌پژوهانه به شاخص‌های ارزیابی در توسعه آموزش عالی

سیده سرور فرهی^۱، فهیمه توانگر^۲، محبت جاوید فر^۳، ناهید دهقانی^۴، محبوبه محسنی فرد^۵

^۱ دانشجوی کارشناسی ارشد مدیریت صنعتی، دانشگاه خلیج فارس، بوشهر
sorour.farrahi@mehr.pgu.ac.ir

^۲ کارشناس ارشد اقتصاد انرژی، دانشگاه خلیج فارس، بوشهر
^۳ کارشناس ارشد مدیریت بازرگانی، دانشگاه خلیج فارس، بوشهر
^۴ دانشجوی کارشناسی ارشد آمار ریاضی، دانشگاه خلیج فارس، بوشهر
^۵ کارشناس مدیریت صنعتی، دانشگاه خلیج فارس، بوشهر

چکیده مختصر

تحولات روزافزون علم و تغییر روند نیازها و اقتضائات جامعه، سبب چالش‌های رقابتی شدیدی در سازمان‌ها و به خصوص دانشگاه‌ها و مراکز آموزش عالی شده است. بنابراین مدیران اجرایی در حوزه آموزش عالی نیز بایستی با داشتن نگاه آینده‌پژوهانه، عملکرد راهبردی خود را با محیط پیچیده و پویای پیرامون خود منطبق کنند تا بتوانند تصمیم‌گیری بهتری در مواجهه با چالش‌های درون‌سازمانی و برون‌سازمانی داشته باشند. در این راستا یکی از ابزارهایی که می‌تواند به طور مؤثر در اختیار مدیران قرار گیرد، شاخص‌های ارزیابی کیفیت و پویایی زیرمجموعه‌های سازمانی در سطوح متفاوت است. اهمیت تعریف شاخص مناسب در اندازه‌گیری متغیرهایی که ذاتاً کیفی می‌باشند اهمیت فراوانی دارد. در این پژوهش تلاش شده است تا این درجه اهمیت را با استفاده از علم آمار و ایجاد نگرشی ساختاریافته به شاخص‌های فرآیندی، پیامدی، درون و برون‌سازمانی نشان داد. اهمیت این نگرش از تعریف تغییرپذیری در علم آمار ساطع می‌شود که در این مقاله به آن پرداخته خواهد شد.

کلمات کلیدی

نگرش آمار راهبردی، آینده‌پژوهی، عملکرد سازمانی، شاخص‌های ارزیابی، تغییرپذیری

پوشش اتکایی مازاد-زیان بهینه تحت معیارهای بقای توأم، مطالعه موردی: داده‌های آتش‌سوزی دانمارک

احمد بیگدلی^۱ حمیده داریوش همدانی^۲

^۱ کارشناس بیمه‌های عمر انفرادی، بیمه نوین، تهران،
ahmad.big2020@gmail.com

^۲ عضو هیئت علمی گروه آمار، دانشگاه شهید بهشتی، تهران،
hamedani@gmail.com

چکیده مختصر

تا به امروز عمدتاً قراردادهای اتکایی صرفاً از دید بیمه‌گر اولیه و منحصرأ بر اساس منافع او تنظیم می‌شده است. در مقاله‌ی حاضر ایده‌ی اصلی، تنظیم چنین قراردادی با در نظر گرفتن منافع هر دو شرکت بیمه اولیه و اتکایی است. مسئله‌ی بهینگی قرارداد بیمه‌ی اتکایی مازاد زیان با سطح نگهداشت و حد بیمه‌نامه را در نظر بگیرید. نشان داده شده است که این مسئله تحت مفروضاتی کلی برای مدل مخاطره و با استفاده از ترکیب دو معیار "مخاطره" و "عملکرد" قابل حل است. تحت مدل مخاطره‌ی مذکور، تابع حق بیمه با استفاده از هر تابع نامنفی و نانزولی، ورود ادعاها با استفاده از فرایند پواسون و شدت ادعاها به وسیله‌ی هر توزیع توأم پیوسته‌ای مدل‌بندی شده است. از "احتمال بقای توأم" به عنوان معیار مخاطره و "سود مورد انتظار به شرط بقای توأم" به عنوان معیار عملکرد استفاده شده است. قرارداد بیمه‌ی اتکایی مازاد زیان "بهینه" ابتدا فقط با استفاده از معیار مخاطره و سپس با استفاده از ترکیب هر دو معیار به دست آمده و در هر حالت مسائل مختلف بهینه‌سازی، شامل مسائل معکوس، مطرح و پاسخ آنها تحت فرض ادعاها‌ی مستقل و با استفاده از داده‌های آتش‌سوزی دانمارک، ارائه شده است.

واژگان کلیدی

بیمه‌ی اتکایی مازاد-زیان بهینه، چندجمله‌ای‌های آپل^۱، احتمال بقای توأم بیمه‌گر اولیه و اتکایی، سود مورد انتظار، داده‌های آتش‌سوزی دانمارک.

^۱ Appell Polynomials

تحلیل تکین طیفی (SSA) و کاربرد آن در هواشناسی

نادر نجاری^۱، نجمه شریفی پناه^۲، محمد ذکایی^۳

^۱گروه آمار، دانشگاه شهید بهشتی، تهران، nader_najari85@yahoo.com

^۲گروه آمار، دانشگاه شهید چمران، اهواز، najmesharifipناه@yahoo.com

^۳گروه آمار، دانشگاه شهید بهشتی، تهران، mo_zokaei@yahoo.com

چکیده مختصر

یکی از مهمترین مسایل در تحلیل سری‌های زمانی و به‌طور عام‌تر مدل‌های آماری، یافتن روشی است که از توانمندی بالایی برخوردار بوده و در عین حال نسبت به تغییرات تصادفی نامنظم (نوفه) استوار باشد. تلاش‌های زیادی در این راستا شده و روش‌های توانمندی نیز معرفی شده‌اند اما بسیاری از آنها مبتنی بر فرض‌های محدود کننده‌ای مانند نرمال و خطی بودن هستند که ممکن است در عمل برقرار نباشند. تحلیل تکین طیفی (مبتنی بر تجزیه مقدار تکین) یکی از روش‌های ناپارامتری قدرتمند برای تحلیل داده‌ها است که برای کاهش سطح نوفه و مدل‌سازی مورد استفاده قرار می‌گیرد و مبتنی بر فرض‌های محدود کننده‌ای مانند نرمال و خطی بودن نیست. در این مقاله نمونه‌ای از کاربرد SSA در حوزه‌ی هواشناسی را بیان کرده و برتری آن را نسبت به مدل SARIMA نشان می‌دهیم.

کلمات کلیدی

تحلیل تکین طیفی، تجزیه‌ی مقدار تکین، سری زمانی، مدل‌سازی، نوفه

تعیین حق بیمه خالص بهینه برای سیستم‌های پاداش جریمه

محمد رضا کردباقری^۱،

mohammadreza366@yahoo.com

علیرضا کردباقری^۲

a.kordbagheri366@gmail.com

سید یونس شریعتمداری^۳

yones.sh@hotmail.com

^{۱،۲،۳} دانشجوی کارشناسی ارشد گروه بیم‌سنجی، دانشگاه شهید بهشتی، تهران،

چکیده مختصر

هدف این مقاله توسعه مدلی مناسب برای محاسبه حق بیمه خالص در سیستم‌های پاداش-جریمه است. مدل ارائه شده، وابسته به تعداد و مقدار ادعای بیمه‌گذاران، در کل سال‌های بیمه‌نامه است. این مطالعه تحقیقی تجربی و پایه‌ای، و با در نظر گرفتن تعداد و مقدار ادعای بیمه‌نامه اتومبیل در دوره‌ای از سال 1388 تا 1395 به اجرا درآمده است. با تقریب حق بیمه به روش بیز، میزان عادلانه بودن حق بیمه‌های این بیمه‌نامه در طول دوره پرداخته می‌شود. این مطالعه براساس روش بیزی و با ترکیب کردن هم‌زمان سه متغیر، تعداد ادعا، مقدار ادعا و زمان بیمه‌نامه، مدل مناسبی برای تعیین حق بیمه فراهم می‌آورد. این مقاله سیستم پاداش-جریمه بهینه را با استفاده از مدل‌های آمیخته متناهی ارائه می‌دهد و توزیع پواسن و گاما به ترتیب برای تعداد ادعاها و مقدار ادعاها انتخاب می‌شود. براساس مدل، رابطه مستقیمی بین افزایش تعداد ادعا و مقدار خسارت با افزایش حق بیمه وجود دارد. این مدل می‌تواند یک ثبات و عدالتی در حق بیمه‌های اتومبیل برای همه بیمه‌گذاران در سطوح مختلف براساس کل ادعاها در یک دوره زمانی به وجود آورد.

کلمات کلیدی

بیمه اتومبیل، سیستم پاداش جریمه، بیزی، پواسن مرکب، توزیع گاما.

تعیین عوامل مؤثر بر دوره بیکاری در ایران

سعید عیسی‌زاده^۱، اعظم قلی‌پور^۲

^۱ عضو هیئت علمی دانشگاه بوعلی سینا، دانشکده علوم اقتصادی و اجتماعی، همدان

isazadeh@gmail.com

^۲ کارشناسی ارشد توسعه اقتصادی و برنامه‌ریزی دانشگاه بوعلی سینا، دانشکده علوم اقتصادی و اجتماعی، همدان

gholipoor.aazam@gmail.com

چکیده مختصر

اشتغال و بیکاری، از جمله موضوع‌های اساسی اقتصاد هر کشوری است، به گونه‌ای که افزایش اشتغال و کاهش بیکاری به عنوان یکی از شاخص‌های توسعه‌یافتگی جوامع تلقی می‌شود. بیکاری پدیده‌ای است که اکثر جوامع با آن روبرو هستند و بیش‌تر کشورهای دنیا به نوعی آن را تجربه کرده‌اند. بیکاری مسئله‌ای کلان است که می‌تواند افراد یک جامعه را تحت تأثیر قرار دهد. اهمیت موضوع بیکاری و آثار و تبعات اقتصادی-اجتماعی آن سبب شده است که سیاست‌های اشتغال و مبارزه با بیکاری، در اکثر کشورها به ویژه کشورهای در حال توسعه در کانون توجه برنامه‌های توسعه‌ی اقتصادی-اجتماعی قرار گیرد. بیکاری موضوعی است که مطالعات بسیاری درباره‌ی آن صورت گرفته است اما آنچه که بیش‌تر اهمیت دارد طول دوره‌ای است که افراد بیکار هستند. دوره‌ی بیکاری مدت زمانی است که افراد در جستجوی شغل هستند و شناسایی عواملی که بر این دوره اثرگذار است می‌تواند کمک فراوانی به کاهش این معضل اجتماعی-اقتصادی نماید. هدف این پژوهش بررسی عوامل تأثیرگذار بر طول مدت بیکاری کارجویان در ایران است و برای همین منظور از داده‌های طرح آمارگیری نیروی کار مرکز آمار ایران سال ۱۳۹۲ استفاده شده است. هم‌چنین برای تخمین داده‌ها از نرم افزارهای SAS9.2 و STATA12 استفاده شده است. نتایج پژوهش نشان می‌دهند که مردان، افراد متأهل، افراد دارای مدرک تحصیلی دیپلم و زیردیپلم و نیز افرادی که ساکن مناطق روستایی هستند دوره‌ی بیکاری کوتاه‌تری را تجربه می‌کنند.

کلمات کلیدی

بیکاری، دوره‌ی بیکاری، بیکاری بلند مدت، روش تحلیل بقا.

تغییرات مشخصه‌های آماری بارش در ایستگاه همدان

امید حمیدی^۱، زهره مریانجی^۲

^۱ گروه علوم پایه، دانشگاه صنعتی همدان، همدان،
omid_hamidi@hut.ac.ir

^۲ گروه جغرافیا، دانشگاه سید جمال الدین اسدآبادی، اسدآباد
z.maryanaji@gmail.com

چکیده مختصر

در این تحقیق مشخصه‌های آماری بارش سالانه ایستگاه همدان به منظور تعیین روندهای زمانی بلندمدت داده‌ها و تغییرات احتمالی اقلیم، طی دوره ۳۰ ساله مورد تجزیه و تحلیل واقع شد. بدین منظور سری‌های داده‌های بارش سالانه بر اساس روش‌های منکندال و منحنی تجمعی باقیمانده‌های نرمال بررسی شدند. با توجه به روش منکندال، تغییرات معنی دار آماری، در سری داده‌های بارش مشاهده نشده است. در روش منحنی تجمعی باقیمانده‌های نرمال، سال ۱۳۸۰ به عنوان محور تقسیم داده‌ها تعیین شده و بر اساس آن، به دو بخش تقسیم گردیدند. فراوانی تعداد دفعات حادث شده در بالا و پایین خط نرمال دو بخش مذکور، به ترتیب ۵۰ و ۵۰ درصد بوده، که این امر نشان‌دهنده عدم وجود روند در سری زمانی بر اساس این روش می‌باشد. بنابراین با توجه به روش منحنی تجمعی باقیمانده‌های نرمال، هیچگونه روند کاهشی یا افزایشی مشاهده نشده است. نتایج گرفت که تغییرات بارش در این ایستگاه صرفاً نوسانات اقلیمی بوده که این امر در اقلیم مناطق نیمه خشک امری کاملاً طبیعی می‌باشد.

کلمات کلیدی

بارش همدان، آزمون‌های منکندال و منحنی تجمعی باقیمانده‌های نرمال شده، نوسانات اقلیمی.

حق بیمه باورمندی تحت مدل آمیخته ارلانگ برای نرخ‌گذاری بیمه اتکایی محصولات کشاورزی

احسان جهانبانی^۱، امیر تیمور پاینده نجف آبادی^۲
^۱گروه آمار، دانشگاه شهید بهشتی، تهران،

e.jahanbani1370@gmail.com
Amirtpayandeh@gmail.com

چکیده مختصر

هدف از این مقاله، بررسی موضوع‌های اساسی پیرامون نرخ‌گذاری بیمه محصولات کشاورزی از دیدگاه بیمه اتکایی است. برای این منظور نسبت هزینه‌ی زیان‌های گذشته (LCR) که برابر نسبت کل خسارت‌های همان سال به کل حق بیمه عملیاتی همان سال است و توزیع آمیخته ارلانگ مطرح می‌شود. یک روش باورمندی تعدیل یافته براساس توزیع آمیخته ارلانگ و LCR وزنی توسعه داده می‌شود. با استفاده از یک مجموعه داده‌ی جامع نشان داده می‌شود که توزیع آمیخته ارلانگ در مقایسه با توزیع‌های دیگر دقیق‌تر است. حق بیمه باورمندی ناهمگن براساس LCR وزنی خیلی محافظه‌کارانه است و یک روش علمی برای قیمت‌گذاری بیمه اتکایی توسعه می‌دهد. چگونگی به کارگیری روش‌ها در عمل با استفاده از داده‌های محصولات کشاورزی ایران انجام گرفته است.

کلمات کلیدی

نرخ‌گذاری، بیمه محصولات کشاورزی، نظریه باورمندی، توزیع آمیخته ارلانگ، نسبت هزینه‌ی زیان‌ها.

مراجع

- [1] Lysa Porth, Wenjun Zhu, Ken Seng Tan, (2014) "A credibility-based Erlang mixture model for pricing crop reinsurance", Agricultural Finance Review, Vol. 74 Iss: 2, pp.162 – 187.

داده‌کاوی و شبکه‌های بیزی

الهام خرازی کلجاهی^۱ و حسین جباری خامنه‌ای^۲

^۱ دانشجوی کارشناسی ارشد آمار ریاضی، دانشگاه پیام نور تبریز، تبریز
Mahsaf557m@yahoo.com

^۲ هیات علمی گروه آمار، دانشگاه تبریز، تبریز
h_jabbari@tabrizu.ac.ir

چکیده مختصر

داده‌کاوی تلاش برای استخراج بینش و آگاهی از انبوه داده‌های بزرگ است. داده‌کاوی با مجموعه‌ای از روش‌های آماری و مدلسازی، می‌تواند روابط پنهان موجود در پایگاه‌های داده را تشخیص دهد. امروزه سازمان‌ها قادرند با هزینه کم اطلاعات وسیعی از وضعیت کسب و کار خود جمع و نگهداری کنند و این باعث ارزش روش‌های داده‌کاوی برای سازمان‌ها شده است. رویکرد‌های متنوعی برای داده‌کاوی وجود دارد، شبکه‌های بیزی یکی از این رویکرد‌ها می‌باشد. شبکه‌های بیزی به صورت گرافیکی توزیع احتمال مجموعه‌ای از متغیرهای تصادفی را بیان می‌کنند. که ساختار شبکه‌های بیزی گرافی جهت دار است که این شبکه‌ها ابزاری قدرتمند برای بازنمایی دانش و استنباط تحت عدم قطعیت در داده‌کاوی هستند.

کلمات کلیدی

داده‌کاوی، شبکه‌های بیزی، دسته‌بندی

رویکرد قانون بیمه اجباری مصوب ۱۳۹۵ در حمایت از حقوق بیمه‌گذاران و زیان‌دیدگان ناشی از حوادث رانندگی

صادق بهرامی فیل آبادی^۱، امید خسروی^۲

^۱ دانشجوی کارشناسی ارشد، دانشگاه حکیم سبزواری، سبزوار
bahrami.sadegh123@gmail.com

^۲ دانشجوی کارشناسی ارشد، دانشگاه حکیم سبزواری، سبزوار
omk1368@gmail.com

چکیده مختصر

نظام حقوقی ایران در زمینه بیمه شخص ثالث و به منظور حمایت از بیمه‌گذاران و زیان‌دیدگان ناشی از حوادث رانندگی از سال ۱۳۳۱ تاکنون پنج دوره قانون‌گذاری را با عناوین مختلف پشت سر گذاشته است. اولین قانون در تاریخ ۱۳۳۱/۸/۸ و قانون بعدی تحت یک لایحه اصلاحی در مورخه ۱۳۳۱/۱۱/۸ در زمان دولت دکتر محمد مصدق به تصویب رسیدند. قوانین بعدی به ترتیب در سالهای ۱۳۴۷، ۱۳۸۷ و ۱۳۹۵ تصویب شدند. در زمینه جبران خسارتهای بدنی تا قبل از سال ۱۳۹۵ معضلاتی وجود داشت به این صورت که از یک سو دارندگان وسایل نقلیه موتوری زمینی با وجود تحصیل بیمه نامه شخص ثالث باز هم محکوم به جبران بخشی از خسارات وارده به زیان‌دیدگان شده و از آنجا که وضعیت مالی مناسبی نداشتند به حبس محکوم می‌شدند و از سوی دیگر جبران خسارت اشخاصی که در نتیجه حوادث رانندگی مجروح یا فوت می‌شدند سالها به تأخیر افتاده و حتی امکان داشت که خسارت آنها به طور کامل جبران نشود. علت اصلی این معضلات؛ سختگیری، تأخیر و سوءاستفاده شرکتهای بیمه و صندوق تأمین خسارتهای بدنی در جبران خسارات بود. قانون جدید بیمه اجباری مصوب ۱۳۹۵ با تحولاتی که در زمینه بیمه شخص ثالث ایجاد کرده تقریباً به تمام این معضلات دامن زده است. از جمله این تحولات که بیشتر در زمینه حمایت از حقوق بیمه‌گذاران و زیان‌دیدگان از حوادث رانندگی بوده است می‌توان به ضرورت جبران خسارت توسط بیمه‌گر و صندوق تأمین خسارتهای بدنی ظرف پانزده یا بیست روز از قطعی شدن میزان خسارت و پیش‌بینی ضمانت اجرای آن، تکلیف صندوق به جبران خسارات ناشی از وسایل نقلیه فاقد بیمه‌نامه، بطلان رضایت زیان‌دیده به بیمه‌گر و صندوق مبنی بر دریافت خسارت کمتر و مواردی از این قبیل اشاره کرد.

کلمات کلیدی

زیان دیده، راننده مسبب حادثه، بیمه‌گر، صندوق تأمین خسارتهای بدنی

روش‌های برآورد ارزش منصفانه تعهدات بیمه‌ای

زهرا ماجدی^۱، ملیحه صمدی^۲

^۱مدیریت طرح و توسعه، شرکت بیمه ملت، تهران،

Majedi.eco@gmail.com

^۲واحد صدور و خسارت، شرکت بیمه آسیا، قزوین،

samadi.eco.62@gmail.com

چکیده مختصر

در دهه‌های اخیر اهمیت و کاربرد ارزش منصفانه برای دستیابی به اطلاعات صحیح و معتبر برای تصمیم‌گیران در صنعت بیمه افزایش یافته است. در این حوزه یکی از چالش‌های موجود، نحوه محاسبه و برآورد ارزش منصفانه بر اساس فرضیات واقع‌بینانه می‌باشد. در این مقاله به مهمترین روش‌های برآورد ارزش منصفانه به همراه مزایا و معایب هر روش پرداخته شده است. رویکردهای مطرح شده در این زمینه به دو دسته رویکردهای مالی و اکچوئری تقسیم می‌شوند. روش‌های برآورد ارزش منصفانه باید تا حد ممکن ساده و قابل فهم بوده و برآوردهای پایدار و معتبر ارائه دهند. همچنین با توجه به متفاوت بودن ماهیت رشته‌های بیمه‌ای، باید از روش برآورد مناسب همان رشته استفاده کرد. به طور کلی، تعدیلات ریسک بر اساس اطلاعات کل صنعت بسیار پایدارتر از تعدیلات ریسک مبتنی بر داده‌های خاص شرکت هستند. همچنین تعدیلات ریسک در یک رشته دارای پایداری کمتری نسبت به تعدیلات ریسک بر اساس تمامی رشته‌ها می‌باشند.

کلمات کلیدی

ارزش منصفانه، CAPM، سربار ریسک، قیمتگذاری.

طرح سانسور پیش‌رونده‌ی پیوندی نوع I در مطالعه‌ی طول عمر یک سیستم صنعتی از دو جامعه‌ی نمایی

عباس رسولی^۱، مریم قاسمی^۲

^۱گروه آمار، دانشگاه زنجان، زنجان،
rasouli@znu.ac.ir

^۲گروه آمار، دانشگاه زنجان، زنجان
Maryam_ghasemi@znu.ac.ir

چکیده مختصر

در این مقاله روش جدید با نام سانسور نوع I پیش‌رونده‌ی پیوندی (JPHCI) برای جمع‌آوری مشاهدات در مطالعات آماری در مباحث طول عمر و قابلیت اعتماد سیستم‌های صنعتی و الکترونیکی معرفی شده است. برآوردگر حداکثر درست‌نمایی و بی‌زی تحت تابع زیان مربع خطا پارامترهای دو جامعه نمایی که از پرکاربردترین توزیع‌ها در آمار کاربردی است را تحت (JPHC-I) بدست آورده و فواصل اطمینان تقریبی و نواحی اعتماد نیز محاسبه می‌شود. در انتها یک مثال عددی واقعی با شبیه‌سازی برای این نتایج ارائه شده است.

کلمات کلیدی

طرح سانسور پیش‌رونده نوع I پیوندی، برآوردگر حداکثر درست‌نمایی، برآوردگر بی‌زی، توزیع نمایی

کاربرد داده‌کاوی در صنعت برق، بررسی خصوصیات پیشنهاددهندگان برتر و الگویابی آن در شرکت برق منطقه‌ای مازندران و گلستان

سیدباقر میراشرفی^۱، مهرزاد رستمی^۲، سیدعلی خالقی^۳

^۱ استادیار گروه آمار دانشکده علوم ریاضی دانشگاه مازندران، بابلسر
b.ashrafi@umz.ac.ir

^۲ کارشناس ارشد شرکت برق منطقه‌ای مازندران و گلستان، واحد فناوری اطلاعات، ساری
mrostami@mazrec.co.ir

^۳ کارشناس ارشد شرکت برق منطقه‌ای مازندران و گلستان، واحد فناوری اطلاعات، ساری
khaleghy.ali@gmail.com

چکیده مختصر

این مقاله در مورد بررسی خصوصیات پیشنهاددهندگان برتر با استفاده از داده‌های موجود در بانک اطلاعاتی سیستم نظام پیشنهادات تهیه و ارائه می‌گردد. بانک اطلاعات از یکپارچه‌سازی بانک‌های اطلاعاتی نظام پیشنهادات، پرسنلی، آموزش و حضور و غیاب به دست آمده است. با توجه به اینکه اطلاعات نظام پیشنهادات به صورت تجمعی ۱۵ ساله می‌باشد لذا با کمک روش‌های داده‌کاوی، مشخصات پیشنهاددهندگان برتر، برای این دوره مشخص می‌گردد. الگوها و شاخص‌های مناسب بر روی داده‌های هر سال نیز می‌تواند پیشنهاددهندگان برتر همان سال را تعیین نماید. به طوریکه شاخص‌ها و الگوهای مناسبی که از طریق کل داده‌های دوره ۱۵ ساله بدست می‌آید نیز می‌توان پیشنهاددهندگان برتر را مشخص نمود.

کلمات کلیدی

پیشنهاددهندگان برتر، داده‌کاوی، معیاره و شاخص‌های پیشنهادات، یکپارچه‌سازی داده‌ها، ارزیابی الگو

کاربرد مدل‌های آمیخته خطی تعمیم‌یافته در آمار اکچوئری

ساجده کلانتر^۱، مهسا فرخی^۲، طاهره دارابیان^۳، محیا پورمصطفی^۴، فاطمه علی محمدی^۵

^۱ گروه آمار، دانشگاه علامه طباطبایی، تهران،
s.kalantar@atu.ac.ir

^۲ گروه آمار، دانشگاه علامه طباطبایی، تهران،
m.farrokhi@atu.ac.ir

^۳ گروه آمار، دانشگاه علامه طباطبایی، تهران،
t.darabian@atu.ac.ir

^۴ گروه آمار، دانشگاه علامه طباطبایی، تهران،
m.pourmostafa@atu.ac.ir

^۵ گروه آمار، دانشگاه علامه طباطبایی، تهران،
f.alimohammadi@atu.ac.ir

چکیده مختصر

با توجه به این‌که در دهه‌های گذشته استفاده از مدل‌های خطی تعمیم‌یافته در آمار مورد توجه بوده است و در مدل‌سازی از یک نمونه‌ای از متغیرهای تصادفی مورد استفاده قرار می‌گیرند بر آن شدیم تا در این مقاله به بررسی تکنیک‌های آماری مورد استفاده برای مدل‌سازی داده‌ها در حوزه‌ی مدل‌های خطی تعمیم‌یافته و کاربرد آن در آمار اکچوئری بپردازیم.

کلمات کلیدی

آمار اکچوئری، مدل‌های خطی تعمیم‌یافته

گونه‌تعمیم‌یافته‌مدل پاسخ تصادفیده

معصومه وقوفی یکتا^۱، سید محمد رضا علوی^۲

^۱ دانشجوی کارشناسی ارشد، گروه آمار، دانشگاه شهید چمران، اهواز ،
Mamajj6457@gmail.com

^۲ گروه آمار، دانشگاه شهید چمران، اهواز
آدرس پست الکترونیکی Alavi_m@scu.ac.ir

چکیده مختصر

معمولاً در نمونه‌گیری پاسخگو به سوال‌های مستقیم درباره خصوصیت حساس پاسخ واقعی نمی‌دهد. لذا در وضعیت‌هایی همچون فرار مالیاتی، تقلب در بیمه، سقط جنین، مصرف مواد مخدر، تقلب در امتحانات و غیره که زمینه پاسخ‌های خجالت‌آور و گناهکارانه وجود دارد، روش پاسخ‌های تصادفیده (RR) برای حفاظت از محرمانگی پاسخ پاسخگو طراحی شده‌اند. در این مقاله ابتدا مدل پاسخ تصادفیده جست وانگ و سینگ (۲۰۰۶) ارائه و سپس گونه‌ای تعمیم‌یافته از مدل جست وانگ و سینگ (۲۰۰۶) را برای برآورد نسبت حساس معرفی می‌کنیم.

کلمات کلیدی

پاسخ تصادفیده، صفت حساس، پاسخ مستقیم، نسبت حساس.

گذری بر بازار بیمه اتومبیل در ایران

فاطمه سادات آل‌حسینی^۱، رحیم محمودوند^۲

^۱ کارشناسی ارشد آکچوئری، شرکت بیمه کارآفرین، تهران
alehosseinifateme@gmail.com

^۲ گروه آمار، دانشگاه بوعلی سینا، همدان
r.mahmodvand@gmail.com

چکیده مختصر

گسترده‌گی و شدت بالای خسارت‌های منجر به جرح و فوت ناشی از تصادفات رانندگی در ایران، پیامدهای ناگوار مالی، انسانی و اجتماعی بسیاری را بر فرد و جامعه تحمیل می‌سازد. گرچه دولتمردان با وضع قوانینی همچون اجباری شدن بیمه شخص ثالث سعی در کاهش اثرات منفی این مسئله کرده‌اند، مطالعه‌ی آمار وسایل نقلیه در حال تردد فاقد پوشش بیمه‌ای مرتبط (همچون، ثالث یا بدنه) و آمار زندانیان بالای محکوم به پرداخت دیه ناشی از تصادفات رانندگی، نشانگر ظرفیت بالا و فاقد پوشش لازم در بازار بیمه اتومبیل کشورمان است. این مقاله به جایگاه بیمه‌های مرتبط با وسایل نقلیه در ایران از منظر آمارهای رسمی می‌پردازد.

کلمات کلیدی

بیمه شخص ثالث، بیمه بدنه، شاخص هرفیندال-هیرشمن.

محاسبات فنی بیمه‌ای و تأثیر آن در جدول اضافه نرخ بیمه‌نامه عمر

مهرداد ترکمان^۱ رویا شریفی اسکندر^۲

^۱ کارشناس ارشد حقوق کیفری و جرم‌شناسی، نماینده شرکت بیمه نوین، همدان
novin4600@gmail.com

^۲ کارشناس ارشد حقوق کیفری و جرم‌شناسی، عضو موسسه حامی نوابغ جوان، تهران
R.sh1986.e@gmail.com

چکیده مختصر

شرکت‌های بیمه به‌منظور تضمین توانایی مالی در ایفای تعهدات آتی، استمرار کسب و کار و حفظ منافع بیمه‌گذاران و سهامداران، به دنبال استفاده از روش‌ها، ابزارها و استانداردهای مناسب و کاربردی مدیریت ریسک هستند. لذا برای دستیابی به این امر در ادبیات بیمه تحت عنوان توانگری مالی دنبال می‌شود که یکی از مباحث مهم در مدیریت ریسک مؤسسات مالی به ویژه شرکت‌های بیمه است. در ادبیات توانگری مالی، محاسبات فنی بیمه‌ها از مهم‌ترین بخش‌ها و عناصری است که همواره مورد تأکید بوده و در شرکت‌های بیمه به آن پرداخته شده است. محاسبات فنی بیمه‌ای یا اکچوئر به معنای فرد متخصصی است که به ارزیابی بروز خطرات مالی احتمالی می‌پردازد. مهارت‌های اکچوئری در ارزیابی ریسک، تعیین تعرفه‌های مناسب و نیز تبصره‌های فنی در بیمه‌های عمر، تنظیم جداول مرگ و میر، تعرفه بیمه‌های عمر، محاسبه ذخایر ریاضی بیمه عمر و هم‌چنین محاسبه جداول بازخرید بیمه‌نامه‌ها و تخفیف سرمایه بیمه عمر (در صورت انصراف بیمه‌گذار از ادامه پرداخت حق بیمه) از وظایف اصلی اکچوئر در یک شرکت بیمه است. شرکت‌های بیمه برای محاسبات فنی بیمه‌ای و چگونگی تأثیر آن در جدول اضافه نرخ بیمه‌نامه عمر سازوکاری برای محاسبه و مدل‌بندی این مورد اتخاذ کرده‌اند. با توجه به شرایط و احتمالات مختلف از قبیل، قد و وزن بیمه‌گذار و یا مشکلات پزشکی و بیماری که هر کدام از این موارد می‌تواند نرخ بیمه‌نامه را با اضافه نرخ بیمه‌نامه عمر مواجه سازد.

کلمات کلیدی

محاسبات فنی بیمه، بیمه عمر، ریسک، اضافه نرخ بیمه‌نامه، اکچوئر.

مدل معادلات ساختاری جریمه ریج

علی سلاجقه،^۱ محمد آرشی،^۲ محمد رحمانی^۳

^۱ دانشجوی ارشد رشته آمارریاضی، دانشگاه صنعتی شاهرود، شاهرود

a.salajeghe@shahroodut.ac.ir

^۲ دانشیار گروه آمار دانشگاه صنعتی شاهرود، شاهرود

m_arashi_stat@yahoo.com

^۳ دانشجوی ارشد رشته آمارریاضی، دانشگاه صنعتی شاهرود، شاهرود

m.rahmani@shahroodut.ac.ir

چکیده مختصر

زمانی که بین متغیرهای توضیحی در یک مدل رگرسیون همخطی وجود داشته باشد، روش ریج در برآورد ضرایب رگرسیونی کاراست. در این مقاله، روش جریمه ریج را برای مدل معادلات ساختاری مورد بررسی قرار دادیم. در این راستا مدل معادلات ساختاری جریمه ریج را برای یک مدل تحلیل عاملی تأییدی با یک عامل پنهان و نه شاخص از مجموعه داده هولزینگر و اسوین فورد (۱۹۳۹) ارائه کردیم.

کلمات کلیدی

مدل معادلات ساختاری، جریمه ریج، تحلیل عاملی تأییدی، درست‌نمایی جریمه شده، کاهش پیچیدگی مدل، همخطی

مشخصه‌سازی توزیع پارتو و بتا-پارتو بر اساس مشاهدات نزدیک رکورد

^۱معصومه اکبری، ^۲احمد پور درویش، ^۳آزهره صفرزاده

^{۱،۲،۳}گروه آمار دانشگاه مازندران، بابلسر، ایران
m.akbari@umz.ac.ir

چکیده مختصر

پارتو یکی از توزیع‌هایی است که دارای کاربرهای فراوانی در مسایل اقتصادی و بیمه‌ای است. تا کنون خواص زیادی از این توزیع بر اساس داده‌های ترتیبی مورد بحث قرار گرفته است. در این مقاله برخی از نتایج مشخصه‌سازی این توزیع بر اساس مشاهدات نزدیک رکورد مطرح و اثبات می‌شود. هم‌چنین نتایجی از مشخصه‌سازی توزیع بتا-پارتو که تعمیمی از توزیع پارتو است، مورد بحث قرار گرفته است.

کلمات کلیدی

توزیع پارتو، توزیع بتا-پارتو، مشخصه‌سازی.

مقایسه دو روش براورد در توزیع نمایی تعمیم یافته

معین زمانی^۱، سعیده حمزه نژاد^۲، مهدی مهدی زاده^۳

^{۱،۲،۳}گروه آمار، دانشگاه حکیم سبزواری، سبزوار

saeede.69mz@yahoo.com

mahdzadeh.m@live.com

moein.zamani169@gmail.com

چکیده مختصر

توزیع‌های گاما و وایبل دو پارامتری در تحلیل داده‌های طول عمر کاربرد فراوان دارند. هر دو توزیع، یک پارامتر شکل و یک پارامتر مقیاس دارد، و بسته به مقدار پارامتر شکل می‌توانند نرخ شکست صعودی یا نزولی داشته باشند. بنابراین، این دو توزیع نسبت به توزیع نمایی (که نرخ شکست آن ثابت است) برتری دارند. توزیع گاما کاربردهای گوناگونی در زمینه‌های غیر از مدل‌سازی طول عمر دارد، و توزیع وایبل در تحلیل داده‌های طول عمر (به ویژه در حضور داده‌های سانسور شده) مورد توجه فراوان قرار گرفته است. یک نقص توزیع گاما اینست که تابع توزیع آن فرم بسته‌ای ندارد جز وقتی که پارامتر شکل آن عدد صحیح باشد. در واقع در توابع مذکور، تابع ناقص گاما ظاهر می‌شود که باید به صورت عددی محاسبه شود. به این دلیل، بیشتر از توزیع وایبل به جای توزیع گاما استفاده می‌شود. این توزیع دارای تابع توزیع و تابع نرخ شکست با فرم بسته‌ای می‌باشد، اما یکی از نقایص آن اینست که همگرایی مجانبی توزیع براوردگرهای درست‌نمایی ماکسیمم به توزیع نرمال بسیار کند است. اخیراً، توزیع نمایی تعمیم یافته به عنوان جایگزین توزیع گاما و وایبل پیشنهاد شده که دارای تابع توزیع با فرم بسته‌ای می‌باشد. براوردگرهای درست‌نمایی ماکسیمم در این توزیع، از طریق روش‌های عددی به دست می‌آیند. روش نیکویی برازش ماکسیمم، روشی جدید برای براورد پارامترهای توزیع‌های آماری است. در این مقاله، با استفاده از شبیه‌سازی به مقایسه عملکرد براوردگرهای درست‌نمایی ماکسیمم و گونه‌ای تغییر یافته از براوردگرهای نیکویی برازش ماکسیمم در مورد توزیع نمایی تعمیم یافته می‌پردازیم. نتایج نشان می‌دهد که روش دوم معمولاً کاراتر است.

کلمات کلیدی

توزیع نمایی تعمیم یافته؛ براورد درست‌نمایی ماکسیمم؛ براورد نیکویی برازش ماکسیمم.

نیرومندی ارزش در معرض ریسک تحت توزیع‌های مختلف

احسان جعفری^۱، امیرتیمور پاینده^۲

^۱ دانشجوی کارشناسی ارشد بیم‌سنجی، دانشگاه شهید بهشتی، تهران

ejafari71.n@gmail.com

^۲ هیات علمی گروه آمار، دانشگاه شهید بهشتی، تهران

amirtpayandeh@sbu.ac.ir

چکیده مختصر

ارزش در معرض ریسک (VaR) یک ابزار کمی است که برای اندازه‌گیری حداکثر زیان بالقوه ارزش‌داری‌های یک سبد در طی یک دوره‌ی معین و با احتمال مشخص مورد استفاده می‌گیرد. برای بدست آوردن ارزش در معرض ریسک، یک برآورد چندکی از انتهای دم چپ توزیع بازده‌های غیر شرطی لازم است. در این پایان‌نامه روش‌های پیش‌بینی ارزش در معرض ریسک (VaR) را تحت فرضیات توزیعی مختلف، و دقت این پیش‌بینی‌ها را مورد بررسی قرار می‌دهیم.

تاکنون محققان زیادی روش‌های جدیدی برای برآورد پارامتر ارائه نموده‌اند. دو مدل برآورد تک متغیره و چند متغیره در مجموعه‌ای از فرضیات توزیعی در مقاله به کار گرفته شده، و دقت خود را در ارائه پیش‌بینی ارزش در معرض ریسک (VaR) در خارج از نمونه به وسیله‌ی روش‌های پیش‌آزمون آماری مشخص می‌کنیم. نتایج حاصل از این آزمون‌ها به استفاده از یک طرح طبقه‌بندی شده بر اساس تعداد پذیرش‌های فرض صفر خلاصه شده است، مشروط بر اینکه فرضیات توزیعی بیشترین دقت را در پیش‌بینی ارزش در معرض ریسک ارائه دهد.

کلمات کلیدی

ارزش در معرض ریسک، دقت پیش‌بینی، توزیع‌ها، پیش‌آزمون‌ها.

ویژگی‌های خانواده مفصل مارشال-الکین

سمانه حمیدی پور^۱، محمد بلبلیان قالیباف^۲

^۱ دانشجوی کارشناسی ارشد آمار ریاضی، دانشگاه حکیم سبزواری، سبزوار

sama85h@gmail.com

^۲ استادیار، گروه آمار، دانشگاه حکیم سبزواری، سبزوار

m.bolbolian@hsu.ac.ir

چکیده مختصر

یکی از مدل‌های پرکاربرد برای تحلیل داده‌های دو متغیره و (چند متغیره) توزیع نمایی دو متغیره مارشال-الکین است. با استفاده از ایده مارشال-الکین در ساختن این توزیع، تاکنون توزیع‌های دو متغیره زیادی معرفی و مطالعه شده است. در این مقاله برخی از ویژگیهای خانواده مفصل دو پارامتری مارشال-الکین معرفی شده و چند اندازه وابستگی در این خانواده مورد بررسی قرار گرفته است.

کلمات کلیدی

تابع مفصل، اندازه‌های وابستگی، مفصل مارشال-الکین

نام‌نامه

ابراهیم امینی سرشت ۱۰۷	رضا رضایی ۳۶
ابوذر بازاری ۴۰	رویا شریفی اسکندر ۱۳۸
احمد ملک‌زاده ۶	زهرا پرچیانلو ۸۴
احسان جعفری ۱۴۲	زهرا عسگری ۱۴
احسان جهانبانی ۱۲۹	زهرا ماجدی ۱۳۲
احمد بیگدلی ۱۲۴	زهرا صفرزاده ۱۴۰
احمد پور درویش ۱۴۰	زهرا مریانجی ۲۴, ۱۲۸
احمد نزاکتی ۲۵	زهرا مسلمانی ۱۶
ارمغان نهاری ۴۰	ساجده کلانتر ۱۲۲, ۱۳۵
اسحاق الماسی ۵۱	ساجده مرادنیا ۵۵
اعظم قلی‌پور ۱۲۷	سامان سپهوند ۵۹
اکبر آبروش ۲	سپیده عظیمی ۱۲۰
الهام خرازی کلجاهی ۴۸, ۱۳۰	سعید طهماسبی ۱۱۲
الهه شریفی ۱۱۷	سعید عیسی‌زاده ۱۲۷
امید حمیدی ۲۴, ۱۲۸	سعیده حمزه نژاد ۱۴۱
امید خسروی ۱۳۱	سمانه حمیدی پور ۱۴۳
امیر الفت ۱۶	سمانه عزیزنصیری ۱۰۰
امیر تیمور پاینده نجف آبادی ۷۵, ۱۲۹	سید محمد حسینی ۳۱
امین حسن‌زاده ۹۲, ۹۷, ۱۱۸	سید محمد رضا علوی ۱۳۶
امینه صادق‌پور ۲۵	سید یونس شریعتمداری ۱۲۶
آرزو حبیبی راد ۳۶	سیدباقر میراشرفی ۱۳۴
آیت اله موسوی ۳۰, ۴۴	سیدجمال میرکمالی ۷۹
بردیا پناه‌بحق ۸۰	سیدمهدی صالحی ۲۵
بهداد مصطفائی ۲	سیده سرور فرهی ۱۲۳
بهزاد منصوری ۱۰۵	سیده مریم محدث کسائی ۹۲
پریسا دنیادیده ۱۱۸	سیف اله امید ۳۲
پیمان پهلوانی ۳۱	سیما فرهادی ۳۰
جعفر عبدالمالکی ۲۹	سیدعلی خالقی ۱۳۴
جلال یادگاری ۱۴	شمسی قاسمی ۳۲
حسین جباری خامنه ای ۴۸, ۱۳۰	شهریار میرزایی ۹۳
حسین علیزاده ۶۷	شیما جلیلی ۹۸
حمیده داریوش همدانی ۱۴, ۱۲۴	شیوا مهدی‌پور قبادلو ۹۷
حمیده مهدوی شیانی ۱۰۵	صادق بهرامی فیل آبادی ۱۳۱
خاطره زارعی ۸۰	طاهره دارابیان ۱۲۲, ۱۳۵
داود شاهسونی ۱۰, ۷۱	عادل محمدپور ۵۱
رحیم چینی پرداز ۱۰۵	عباس رسولی ۱۳۳
رحیم محمودوند ۶۳, ۱۳۷	علی پناهی بزاز ۷۵

محمد فهیم شریفی ۱۱۹	علی سلاجقه ۱۳۹
محمد قاسم وحیدی اصل ۵۹	علی محمدیان مصمم ۱۱۷
محمد کاظمی ۷۱	علیرضا غفاری حدیقه ۸۸
محمد رضا کردباقری ۱۲۶	علیرضا کردباقری ۱۲۶
محیا پورمصطفی ۱۳۵, ۱۲۲	عین اله پاشا ۵۵
مریم قاسمی ۱۳۳	غلامرضا کشاورز حداد ۳۲
مسعود گنجی ۲	فاطمه آرزومند ۳۷
مسعود یارمحمدی ۳۷, ۲۰	فاطمه سادات آل حسینی ۱۳۷, ۶۳
معصومه اکبری ۱۴۰	فاطمه علی محمدی ۱۳۵, ۱۲۲
معصومه وقوفی یکتا ۱۳۶	فاطمه نصیری ۱۰۸
معین زمانی ۱۴۱	فرزاد تیموریان ۴۴
ملیحه صمدی ۱۳۲	فهیمه توانگر ۱۲۳
ملیحه عباس نژاد ۱۱۹	کیانا ملک پور ۶۷
مونا شیری ۱۰۷	گلسان احسنی ۶۷
مهدی مهدی زاده ۱۴۱	مائده به فروز ۴۰
مهرداد ترکمان ۱۳۸	متین دالوندپور ۶
مهرداد رستمی ۱۳۴	محبت جاوید فر ۱۲۳
مهسا صالحی ۱۱۲	محبوبه محبی ۱۰
مهسا فرخی ۱۳۵, ۱۲۲	محبوبه محسنی فرد ۱۲۳
مهناز لشنی ۲۰	محدثه صفری زاده ۱۲۱
مهناز منطقی پور ۸۸	محسن قره‌خانی ۱۰۸
نادر نجاری ۱۲۵	محمد آرشی ۱۳۹, ۷۱
ناهید دهقانی ۱۲۳	محمد بلبلیان قالیباف ۱۴۳
نجمه شریفی پناه ۱۲۵	محمد حسین دهقان ۱۲۱
هما نوروزی ۹۸	محمد ذکایی ۱۲۵, ۸۴
	محمد رحمانی ۱۳۹

