مجله‌ی بررسی‌های آمار رسمی ایران
سال ۲۷، شماره‌ی ۱، بهار و تابستان ۱۳۹۵، صص ۱۷ – ۳۵

مدل‌بندی تعداد ادعاهای خسارت در بیمه‌ی ورزش

مریم کریمی باجه باج[†]، آتوسا گودرزی[†] و زهرا رضایی قهرودی[‡,*]

[†] مؤسسه‌ی آموزش عالی بیمه‌ی اکو
[‡] پژوهشکده‌ی آمار

چکیده: هدف اصلی بیمه‌های ورزشی، شناسایی، تعیین مخاطره و همچنین کاهش هزینه‌های درمانی باشگاه‌ها و بازیکنان می‌باشد. همچنین شرکت‌های بیمه در جهت ارایه‌ی پوشش‌های بیمه‌ای مناسب، بر اساس تحلیل خسارت‌ها، حق بیمه را محاسبه می‌نمایند. با توجه به این‌که در صنعت بیمه ادعاهای خسارت صفر (عدم پرداخت خسارت) به‌دلایل زیادی از جمله عدم گزارش‌دهی در خصوص ادعاهای کوچک زیاد است. فراوانی صفر در نمونه‌ها به وفور وجود داشته و تورم عدد صفر در داده‌ها بروز می‌نماید، در این مقاله با استفاده از مدل‌های خطی تعمیم‌یافته (GLMs)[۱]، فراوانی ادعاهای خسارت یک فصل تیم‌های باشگاه سپاهان در فاصله‌ی سال‌های ۱۳۹۲ تا ۱۳۹۳ با استفاده از مدل‌های پواسونی، پواسونی صفرمتورم، دوجمله‌ای منفی و دوجمله‌ای منفی صفرمتورم، مورد بررسی و تحلیل قرار می‌گیرد و سپس با استفاده از معیار AIC مناسب‌ترین مدل برای این داده‌ها ارایه می‌شود. یافته‌های پژوهش حاکی از آن است که توزیع دوجمله‌ای منفی و پواسونی برازش مناسب‌تری را ارایه می‌دهند. همچنین، نتایج مؤید این امر می‌باشد که متغیر کمکی موقعیت بازیکن در زمین بر فراوانی صدمه‌ی واردشده به وی تأثیر معنی‌داری دارد.

واژگان کلیدی: مدل پواسونی و دوجمله‌ای منفی صفرمتورم؛ بیمه‌ی ورزش؛ مدل‌های خطی تعمیم‌یافته؛ ادعای خسارت.

---

* نویسنده‌ی عهده‌دار مکاتبات
دریافت: ۱۳۹۵/۱/۲۸، پذیرش: ۱۳۹۵/۶/۳۰.

۱- مقدمه

با توجه به ماهیت فعالیت‌های ورزشی، مخاطره‌ی مصدومیت همواره سلامت ورزشکاران و حیات ورزشی آن‌ها را تهدید می‌کند. بیمه‌ی ورزشی در جهت ارایه‌ی پوشش‌های کامل بیمه‌ای برای بازیکنان مصدوم و یا از کار افتاده، نقش اساسی و انکارناپذیری در پوشش کامل خسارت‌ها ایفا می‌کند. ارایه‌ی پوشش‌های کامل بیمه‌ای سبب آرامش خاطر بازیکنان و دست‌اندرکاران حوزه‌ی ورزش می‌شود و این امر با انجام مطالعات در زمینه‌ی مخاطرات ورزشی، صدمات ورزشکاران و محاسبات حق بیمه‌ی صحیح در جهت ارایه‌ی پوشش بیمه‌ای کامل برای ورزشکاران امکان‌پذیر می‌شود.

باشگاه‌های ورزشی سالانه هزینه‌ی هنگفتی را صرف درمان بازیکنان مصدوم می‌نمایند و عدم اطلاع و آموزش مسئولین باشگاه‌های ورزشی ایران از شیوه‌های نوین مدیریت مخاطره و جبران خسارت ناشی از مصدومیت‌های جدی ورزشکاران، مشکلات متعددی را متوجه جامعه‌ی ورزشی ایران نموده است.

در بسیاری از کشورها، بیمه‌ی ورزشی در زمینه‌های متنوعی چون بیمه‌ی مربیان، بیمه‌های مسئولیت، بیمه‌ی تماشاگران، بیمه‌ی ورزشکاران حرفه‌ای، غیر حرفه‌ای و ... ارایه می‌شود. به‌عنوان مثال شرکت اسپورتزکاور[۲] یکی از شرکت‌های بیمه‌ی ورزشی فعال در کشورهای استرالیا، انگلستان و بسیاری کشورهای اروپایی است که خدمات بیمه‌ی ورزشی را به میلیون‌ها نفر ورزشکار در سراسر دنیا ارایه می‌دهد. در کشور کانادا، همه‌ی ورزشکاران، مربیان، مدیران و مسئولانی که اعضای سازمان ملی ورزش، سازمان‌های ورزشی استان‌ها، مرکز ورزشی کانادا، تیم‌ها، لیگ و یا باشگاه‌های ورزشی باشند، می‌توانند تحت پوشش بیمه‌ی ورزش (CAIP)[۳] قرار گیرند.

انواع پوشش‌های بیمه‌ای عرضه‌شده عبارت‌اند از: بیمه‌های پایه‌ای[۴] که مصدومیت‌های طولانی‌مدت را تحت پوشش قرار می‌دهند، پوشش‌های فوق‌العاده[۵] که مصدومیت‌های کوتاه‌مدت و بلندمدت را شامل می‌شود، بیمه‌ی ورزشکاران حرفه‌ای که مخاطره‌ی مصدومیت و مخاطره‌های مالی قراردادهای هنگفت آن‌ها را تحت پوشش قرار می‌دهد و سایر انواع پوشش‌های بیمه‌ای که برای یک تیم ورزشی ارایه می‌شود. همه‌ی این موارد حاکی از تنوع و وسعت انواع بیمه‌های ورزشی است که بازیکنان و باشگاه‌ها با توجه به

نیاز خود می‌توانند به دلخواه پوشش بیمه‌ای مناسب را برای ورزشکاران و یا تیم انتخاب کنند.

در ایران تنها بیمه‌ی ورزشی که به جامعه‌ی ورزشکاران ارایه می‌شود، بیمه‌ای است که از طرف فدراسیون پزشکی ورزشی طراحی شده است. همه‌ی ورزشکاران ایرانی باید تحت پوشش این بیمه بوده و باشگاه‌ها موظف به خرید این بیمه‌نامه هستند. نرخ حق بیمه‌ی ورزشی اعلام‌شده توسط فدراسیون پزشکی ورزشی برای تمام قهرمانان و ورزشکاران حرفه‌ای و غیر حرفه‌ای در سال ۱۳۹۴، مبلغ ۱۵۰ هزار ریال بوده است و سقف تعهدات هزینه‌ی درمان و غرامت فوت در سال ۱۳۹۴ به ترتیب ۶۵ و ۱۲۰ میلیون ریال بوده است. مبلغ پایین حق بیمه و پوشش قشر عظیم ورزشکاران ایرانی از مزایای این نوع بیمه می‌باشد، ولی طولانی بودن زمان پرداخت خسارت و پایین بودن خسارت‌های پرداختی و معرفی بیمارستان‌های محدود طرف قرارداد، همگی بر این نکته اشاره دارد که این بیمه، توانایی پوشش کامل و کافی در پرداخت هزینه‌های درمانی و یا جراحی و مشکلات مالی دوران از کار افتادگی به‌خصوص برای ورزشکاران رده‌های بزرگسال و بازیکنان حرفه‌ای را ندارد.

با توجه به این‌که در صنعت بیمه به‌خصوص بیمه‌ی ورزشی، ادعاهای خسارت صفر (عدم پرداخت خسارت) به‌دلایل عدیده از جمله عدم گزارش‌دهی ادعاهای کوچک به‌ویژه در مواردی که خسارت دریافتی (در قیاس با افزایش حق بیمه‌ی سال بعد) قابل توجه نباشد، فراوانی صفر در نمونه‌ها زیاد رخ می‌دهد و تورم عدد صفر در داده‌ها پدید می‌آید، که در این موارد از مدل‌های دوجمله‌ای منفی، پواسونی صفرمتورم یا مدل دوجمله‌ای منفی صفرمتورم برای تحلیل داده‌ها استفاده می‌شود.

طی دهه‌های اخیر استفاده از مدل‌های خطی تعمیم‌یافته و براورد پارامترها به روش بیشینه درستنمایی نقش بسزایی در تحلیل‌های رگرسیونی فراهم آورده است [۸]. در سال ۱۹۹۵ وینکلمن و زیمرمن روش بیشینه درستنمایی را برای برازش مدل دوجمله‌ای منفی توسعه دادند [۹]. برای تحلیل داده‌هایی که دارای فراوانی زیاد صفر بوده و شکل توزیع آن‌ها دارای دم‌های کشیده و سنگین است، استفاده از مدل‌های خطی تعمیم‌یافته برای پاسخ‌های گسسته با در نظر گرفتن توزیع پواسونی و برنولی، نامناسب است. در این مواقع استفاده از مدل‌های تحلیلی صفرمتورم (ZI)[۶] کارایی بیش‌تری دارد که در این خصوص، لمبرت و هال مدل پواسونی صفرمتورم و هیلب مدل دوجمله‌ای منفی صفرمتورم را معرفی

نموده‌اند ([۷]، [۳] و [۴]). در سال ۱۹۹۴ هیلبرن از مدل‌های صفرمتورم در مدل‌بندی داده‌های بیش‌پراکنده در داده‌های شمارشی استفاده کرد [۵]. ایسمیل و جمین با استفاده از مدل‌های پواسونی و دوجمله‌ای منفی تلاش‌هایی را جهت بهبود بیش‌پراکندگی به‌عمل آوردند [۶]. همچنین دی ژان و هلر از مدل‌های خطی تعمیم‌یافته برای تعیین فراوانی و شدت خسارت استفاده کردند [۱]. فلاین و فرانسیس با استفاده از مدل‌های صفرمتورم به مدل‌بندی خسارت‌های ثبت‌شده در تصادفات برای تعیین حق بیمه‌ی اتومبیل پرداختند [۲].

در این مقاله با استفاده از مدل‌های خطی تعمیم‌یافته، فراوانی ادعاهای خسارت یک فصل تیم‌های باشگاه سپاهان با استفاده از مدل‌های پواسونی، پواسونی صفرمتورم، دوجمله‌ای منفی و دوجمله‌ای منفی صفرمتورم، تحلیل و مدل مناسب ارایه می‌شود.

۲- معرفی داده‌های مصدومین بازیکنان باشگاه سپاهان

ادعاهای مصدومین بازیکنان فوتبال باشگاه سپاهان در رده‌های پیشکسوتان و بزرگسالان، جوانان و نوجوانان و امید و نونهالان که تحت پوشش بیمه بوده‌اند، برای سال ۱۳۹۲-۱۳۹۳ از طریق بخش آمار پزشکی باشگاه سپاهان گرداوری شده است. با توجه به اطلاعات جدول ۱ مشاهده می‌شود فراوانی ادعاهای صفر (به معنای عدم درخواست خسارت) ۶۹ درصد است. با محاسبه‌ی واریانس و میانگین مجموعه داده‌ها مشاهده می‌شود، واریانس داده‌ها بیش‌تر از میانگین است، بنا بر این، ممکن است استفاده از توزیع پواسونی برای این نوع داده‌ها به‌دلیل این‌که میانگین و واریانس آن‌ها یکسان نیست (در صورت معنی‌داری اختلاف میانگین و واریانس)، مناسب نباشد و در این مواقع باید از توزیع پواسونی صفرمتورم، دوجمله‌ای منفی و دوجمله‌ای منفی صفرمتورم استفاده شود.

۳- مدل پواسونی و دوجمله‌ای منفی متورم‌شده

مدل‌های خطی تعمیم‌یافته، تعمیمی از مدل رگرسیون خطی هستند که داده‌های آن دارای توزیع نرمال نیست. این مدل رگرسیون رابطه‌ی بین متغیر وابسته و مستقل را مشخص می‌کند. مدل‌های خطی تعمیم‌یافته از ۳ مؤلفه تشکیل شده‌اند.

جدول ۱- فراوانی ادعاهای خسارت فوتبالیست‌ها در تیم باشگاه سپاهان: ۱۳۹۲-۱۳۹۳

تعداد خسارت	فراوانی	درصد فراوانی
۰	۱۳۲	۰/۶۹۰۰
۱	۴۹	۰/۲۵۷۰
۲	۵	۰/۰۲۶۴
۳	۳	۰/۰۱۵۸
۴	۱	۰/۰۰۵۰

مجموع تعداد خسارت: ۱۹۰

میانگین تعداد خسارت: ۰/۳۸

واریانس تعداد خسارت: ۰/۴۵

مؤلفه‌ی اول، مؤلفه‌ی تصادفی مدل‌های خطی تعمیم‌یافته شامل متغیر پاسخ y با مشاهدات مستقل $(y_1, \dots, y_N)$ از توزیعی در خانواده‌ی نمایی طبیعی با تابع احتمال یا تابع جرم احتمال به‌صورت:

$$f(y_i, \theta_i) = a(\theta_i) b(y_i) \exp[y_i Q(\theta_i)] \qquad (1)$$

است که در آن $Q(\theta_i)$ پارامتر طبیعی است و توزیع‌های مهم زیادی از جمله پواسونی و دوجمله‌ای را می‌توان از این خانواده نام برد. صورت از رابطه‌ی (۱) به‌صورت زیر است:

$$f(y_i; \theta_i, \Phi) = \exp\left\{\frac{[y_i\theta_i - b(\theta_i)]}{a(\Phi)} + c(y_i, \Phi)\right\} \qquad (2)$$

که در آن Φ پارامتر پراکندگی و θ_i پارامتر طبیعی است. زمانی که Φ معلوم باشد، رابطه‌ی (۲) تبدیل به (۱) می‌شود به‌طوری که در آن $Q(\theta_i) = \frac{\theta_i}{a(\Phi)}$، $a(\theta_i) = \left[\frac{-b(\theta_i)}{a(\Phi)}\right]$ و $b(y) = \exp[c(y, \Phi)]$ است.

مؤلفه‌ی دوم، مؤلفه‌ی سیستماتیک مدل‌های خطی تعمیم‌یافته است که بردار $(\eta_1, \dots, \eta_N)$ را به‌واسطه‌ی یک مدل خطی، به متغیرهای تبیینی مرتبط می‌کند که در رابطه‌ی (۳) تعریف شده است و در آن x_{ij} مقدار پیش‌گویی‌کننده‌ی jامین متغیر کمکی برای iامین مشاهده است.

$$\eta_i = \sum_{j=1}^{P} \beta_j x_{ij} \qquad i = 1, \dots, N \qquad (3)$$

مؤلفه‌ی سوم، تابع ربط مانند g است که متغیر تصادفی و مؤلفه‌ی سیستماتیک را به هم مرتبط می‌کند.
اگر برای $i = ۱, \dots, N$ فرض کنیم $\mu_i = E(y_i)$، مدل خطی تعمیم‌یافته، μ_i را به وسیله‌ی تابع ربط g به مؤلفه‌ی سیستماتیک به‌صورت $\eta_i = g(\mu_i)$ مرتبط می‌کند. بنا بر این خواهیم داشت:

$$g(\mu_i) = \sum_{j=۱}^{P} \beta_j \, x_{ij} \qquad i = ۱, \dots, N \tag{۴}$$

زمانی که فراوانی صفر در نمونه زیاد باشد، در واقع مجموعه داده‌ها دارای تورم یا انباشتگی در صفر است، که برای تحلیل این نوع داده‌ها از توزیع صفرمتورم استفاده می‌شود. واریانس در توزیع صفرمتورم بزرگ‌تر از میانگین است و بیش‌پراکندگی بروز می‌نماید. در حالی که در توزیع پواسون واریانس و میانگین برابر است. بنا بر این، چنانچه در مجموعه داده‌هایی، واریانس بیش‌تر از میانگین باشد، این اختلاف نشانه‌ی بیش‌پراکندگی قلمداد می‌شود ([۷]).
در توزیع پواسونی صفرمتورم (ZIP)[۷] فرض می‌شود توده‌ی ادعاها در نتیجه‌ی یک فرایند دو بخشی شامل صفرهای ساختاری و ادعاهای تصادفی شمارشی تولید شده است. از طرفی، در برخی موارد رگرسیون پواسونی صفرمتورم نامناسب است که در این موارد ممکن است مدل دوجمله‌ای منفی صفرمتورم کارکرد مناسبی داشته باشد.
فرض کنید متغیر وابسته‌ی شمارشی y از توزیع آمیخته که در k نقطه متورم شده است، پیروی کند. در این صورت:

$$P(Y = y) = \varphi_۱ + \varphi_۲ e^{-\lambda_۲} + \cdots + \varphi_k e^{-\lambda_k},$$

$$P(Y = y) = \varphi_۲ \frac{\lambda_۲^{y} e^{-\lambda_۲}}{y!} + \cdots + \varphi_k \frac{\lambda_k^{y} e^{-\lambda_k}}{y!} \quad y = ۱, ۲, \dots \tag{۵}$$

که در آن k تعداد مؤلفه‌های ترکیبی، λ_i میانگین مؤلفه‌ی iام و φ_i وزن‌های ترکیبی مؤلفه‌ی iام است که به‌صورت زیر تعریف می‌شود:

$$\sum_{i=۱}^{k} \varphi_i = ۱ \qquad \circ < \varphi_i < ۱ \qquad i = ۱, \dots, k,$$

در حالت خاص $(k = ۲)$، توزیع پواسونی صفرمتورم به‌صورت زیر قابل تعریف است:

$$P\left(Y=\circ\right)=\varphi+\left(۱-\varphi\right)e^{-\lambda},$$

$$P(Y=y)=(۱-\varphi)\frac{\lambda^{y}e^{-\lambda}}{y!}\qquad y=۱,۲,\ldots \tag{۶}$$

میانگین و واریانس متغیر پواسونی صفرمتورم عبارت است از:

$$\begin{aligned}E(Y)&=\sum_{y=\circ}^{\infty}y\times p(Y=y)\\&=\circ\times p\left(Y=\circ\right)+\sum_{y=۱}^{\infty}y\times\left(۱-\varphi\right)\frac{\lambda^{y}e^{-\lambda}}{y!}=(۱-\varphi)\lambda\end{aligned}$$

$$\begin{aligned}Var(Y)&=E\left(Y^{۲}\right)-E^{۲}(Y)\\&=\circ\times p\left(Y=\circ\right)+\sum_{y=۱}^{\infty}y^{۲}\times\left(۱-\varphi\right)\frac{\lambda^{y}e^{-\lambda}}{y!}-\left(\left(۱-\varphi\right)\lambda\right)^{۲}\\&=\left(۱-\varphi\right)\times\left[\lambda+\lambda^{۲}\right]-\left[\left(۱-\varphi\right)\lambda\right]^{۲}=(۱-\varphi)\lambda(۱+\varphi\lambda)\end{aligned}$$

توزیع دوجمله‌ای منفی صفرمتورم ترکیبی از متغیر برنولی برای صفرمتورم و دوجمله‌ای منفی است که به‌صورت زیر تعریف می‌شود:

$$p\left(Y=\circ\right)=\varphi+(۱-\varphi)NB(\circ,r,p)$$

$$p(Y=y)=\left(۱-\varphi\right)NB(y,r,p),\qquad y=۱,۲,\ldots \tag{۷}$$

که در آن $NB(y,r,p)$ تابع چگالی توزیع دوجمله‌ای منفی متغیر پاسخ y با پارامترهای r و p است که به‌صورت زیر نمایش داده می‌شود:

$$NB(y,r,p)=\binom{y+r-1}{y}p^{r}\left(۱-p\right)^{y}$$

میانگین و واریانس متغیر دوجمله‌ای منفی صفرمتورم عبارت است از:

$$\begin{aligned}E(Y)&=\sum_{y=\circ}^{\infty}y\times p(Y=y)\\&=\circ\times P\left(Y=\circ\right)+\sum_{y=۱}^{\infty}y\times\left(۱-\varphi\right)\binom{y+r-1}{y}p^{r}\left(۱-p\right)^{y}\\&=\left(۱-\varphi\right)\left[r\frac{۱-p}{p}\right]\end{aligned}\tag{۸}$$

$$
\begin{aligned}
Var(Y) &= E\left(Y^{۲}\right) - E^{۲}(Y) \\
&= ۰ \times P\left(Y = ۰\right) + \sum_{y=۱}^{\infty} y^{۲} \times \left(۱ - \varphi\right)\binom{y+r-1}{y} p^{r}\left(۱ - p\right)^{y} \\
&\quad - \left[\left(۱ - \varphi\right)\left[r\frac{۱-p}{p}\right]\right]^{۲} \\
&= \left(۱ - \varphi\right)\left[r\frac{۱-p}{p}\right]\left(۱ + \left[r\frac{۱-p}{p}\right][\varphi + \mathrm{p}]\right) \qquad (۹)
\end{aligned}
$$

چنانچه براورد پارامتر بیش‌پراکندگی φ در مدل پواسونی و دوجمله‌ای منفی صفرمتورم برابر صفر باشد و یا براورد این پارامتر در این مدل‌ها معنی‌دار نباشد، مدل پواسونی و دوجمله‌ای منفی صفرمتورم و مدل‌های پواسون و دوجمله‌ای معمولی، عملکرد یکسانی دارند.

زمانی که هدف بررسی اثر متغیرهای کمکی مانند موقعیت بازیکن در زمین فوتبال، محل صدمه در بدن بازیکن، سن و جنس بازیکن، نوع درمان و... بر تعداد صدمات و خسارت وارده بر بازیکنان باشد، می‌توان علاوه بر براورد پارامتر پراکندگی، تأثیر این متغیرها را بر اساس مدل رگرسیونی زیر بررسی نمود.

با فرض اثرگذاری موقعیت بازیکن در زمین فوتبال بر تعداد خسارات ثبت‌شده، مدل پواسونی صفرمتورم به‌صورت زیر معرفی می‌شود.

$$\mu_i = E(y_i) = \eta_i = b_{۰} + b_{۱} \times position_i, \qquad i = ۱, ۲ \dots, N$$

که در آن $b_{۱}$ اثر متغیر کمکی موقعیت بازیکن و $\lambda = \exp(\eta_i)$ است. دو معیار سنجش برازش و انتخاب بهترین مدل معیار اطلاع آکائیکه (AIC) و معیار اطلاع بیزی (BIC) است که به‌صورت زیر تعریف می‌شود:

$$AIC = -۲\ln l + ۲p$$

$$BIC = -۲\ln l + ۲p(\ln(N))$$

که در آن p تعداد پارامترهای مدل، N اندازه‌ی نمونه و $\ln l$ لگاریتم تابع درستنمایی است.

جدول ۲ ادعاهای خسارت فوتبالیست‌ها به تفکیک موقعیت بازیکن در تیم باشگاه سپاهان در سال مورد بررسی را نشان می‌دهد. با توجه به تعداد کم نمونه‌های این مطالعه، در مدل‌بندی، متغیر کمکی موقعیت بازیکن به ۲ رده تبدیل شده است که در آن موقعیت بازیکنان حمله یا دفاع یک رده و هافبک یا دروازبان رده‌ی دیگر در نظر گرفته شده است. بنا بر این در مدل‌بندی، متغیر دو حالتی موقعیت بازیکن «حمله یا دفاع» نسبت به «هافبک یا دروازبان» بر فراوانی خسارت واردشده به بازیکن بررسی می‌شود.

جدول ۲- ادعاهای خسارت به‌تفکیک موقعیت بازیکن در تیم باشگاه سپاهان: ۱۳۹۲-۱۳۹۳

موقعیت بازیکن	فراوانی خسارت	
	بدون آسیب (ادعای صفر)	آسیب (ادعای ۱ وبیش‌تر)
حمله	۳۶	۱۹
هافبک	۳۹	۱۲
دفاع	۲۰	۲۶
دروازبان	۳۷	۱
مجموع	۱۳۲	۵۸

۴- نتایج مدل‌بندی ادعاهای خسارت یک فصل بازیکنان تیم‌های باشگاه سپاهان

با توجه به ادعاها و نتایج جدول ۱، ۶۹ درصد داده‌ها را ادعاهای صفر تشکیل می‌دهد. بنا بر این، مدل‌های معرفی‌شده (با و بدون در نظر گرفتن متغیر کمکی)، بر داده‌ها برازش داده شد. مقادیر لگاریتم درستنمایی برای مدل‌های مختلف پواسونی، دوجمله‌ای منفی، مدل پواسونی صفرمتورم و دوجمله‌ای منفی صفرمتورم در جدول ۳ ارایه شده است. همان‌طور که مشاهده می‌شود مقادیر AIC در مدل دوجمله‌ای منفی و پواسونی با در نظر گرفتن متغیر کمکی کم‌تر از بقیه‌ی مدل‌ها است.

در جداول ۴ و ۵، به‌ترتیب نتایج براورد پارامترهای مدل‌های برازش داده‌شده بدون در نظر گرفتن متغیر کمکی موقعیت بازیکن در زمین و با در نظر گرفتن این متغیر در مدل با استفاده از نرم‌افزار SAS ارایه شده است. بر اساس نتایج جدول ۴، در این مجموعه

داده‌ها، r، پارامتر توزیع دوجمله‌ای منفی و b_0 پارامتر ثابت مدل، معنی‌دار نیست. همچنین پارامتر پراکندگی (φ) در مدل پواسونی صفرمتورم و دوجمله‌ای منفی صفرمتورم معنی‌دار نیست. بر اساس نتایج جدول ۵، پارامترهای r، φ و b_0 معنی‌دار نیست. همچنین متغیر کمکی موقعیت بازیکن «موقعیت حمله یا دفاع» نسبت به «هافبک یا دروازبان» بر فراوانی خسارت واردشده به بازیکن معنی‌دار است و مدل با در نظر گرفتن متغیر کمکی، AIC کم‌تری نسبت به بقیه‌ی مدل‌ها و روش‌ها دارد. لذا با توجه به نتایج به‌دست آمده می‌توان نتیجه گرفت که به‌دلیل معنی‌دار نبودن اختلاف میانگین و واریانس در این مجموعه داده‌ها، مدل پواسونی به‌عنوان بهترین مدل انتخاب شده است.

جدول ۳- مقادیر لگاریتم درستنمایی برای مدل‌های مختلف

مدل (بدون در نظر گرفتن متغیر کمکی)	AIC	**مدل** (با در نظر گرفتن متغیر کمکی)	AIC
پواسون	۳۰۹/۸	پواسون	۲۹۸/۳
دوجمله‌ای منفی	۳۰۹/۸	دوجمله‌ای منفی	۳۰۱/۶
پواسونی صفرمتورم	۳۰۹/۸	پواسونی صفرمتورم	۳۰۲/۰
دوجمله‌ای منفی صفرمتورم	۳۰۹/۸	دوجمله‌ای منفی صفرمتورم	۳۰۳/۶

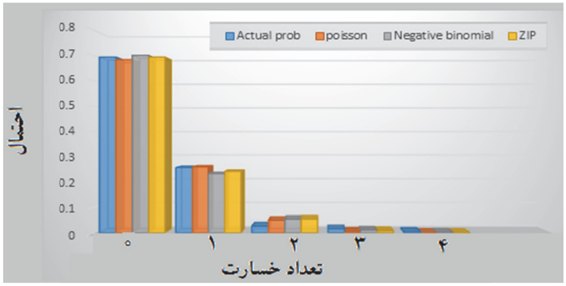
در شکل ۱، احتمال رخداد خسارت فوتبالیست‌های باشگاه سپاهان طی سال‌های ۱۳۹۲-۱۳۹۳ بر اساس مقادیر واقعی، احتمال‌های محاسبه‌شده بر اساس برازش مدل‌های پواسونی، دوجمله‌ای منفی و پواسونی صفرمتورم ارایه شده است. همان‌طور که ملاحظه می‌شود مدل پواسونی و پواسونی صفرمتورم احتمال‌های نزدیک‌تری به مقادیر واقعی را براورد نموده است.

جدول ۴- براورد پارامترهای مدل‌های پواسونی، پواسونی صفرمتورم، دوجمله‌ای منفی و دوجمله‌ای منفی صفرمتورم بدون در نظر گرفتن متغیر کمکی

پارامتر		b_0	میانگین	واریانس	r	φ
پواسون	E.S.	-۰/۹۷	۰/۳۷۸	۰/۳۷۸	-	-
	S.D	۰/۱۱۸	۰/۰۴۵	۰/۰۴۵	-	-
	p-value	<۰/۰۰۰۱	<۰/۰۰۰۱	<۰/۰۰۰۱	-	-
پواسن صفرمتورم	E.S.	-۰/۸۰	۰/۳۷۸	۰/۴۰۵	-	۰/۱۵۶
	S.D	۰/۲۵۸	۰/۰۴۶	۰/۰۶۵	-	۰/۱۹۶
	p-value	<۰/۰۰۲	<۰/۰۰۰۱	<۰/۰۰۰۱	-	<۰/۴۲۵
دوجمله‌ای منفی	E.S.	-۰/۹۷	۰/۳۷۹	۰/۴۳۴	۰/۳۸۱	-
	S.D	۰/۱۲۶	۰/۰۴۸	۰/۰۷۸	۰/۳۳۳	-
	p-value	<۰/۰۰۰۱	<۰/۰۰۰۱	<۰/۰۰۰۱	۰/۲۵۴	-
دوجمله‌ای منفی صفرمتورم	E.S.	-۰/۹۷	۰/۳۸	۰/۳۷۸	۰/۳۸	۰/۰۰۰
	S.D	۰/۱۲۶	۰/۰۴۷	۰/۰۴۷	۰/۳۳	۰/۰۰۰۳
	p-value	<۰/۰۰۰۱	<۰/۰۰۰۱	<۰/۰۰۰۱	۰/۲۵	۰/۹۹۵

جدول ۵: برآورد پارامترهای مدل‌های پواسونی، پواسونی صفرمتورم، دوجمله‌ای منفی و دوجمله‌ای منفی صفرمتورم با در نظر گرفتن متغیر کمکی

پارامتر		b_0	موقعیت بازیکن	میانگین	واریانس	r	φ
پواسون	E.S.	۰/۳۲۶	−۱/۰۱۸	۰/۵۰۰	۰/۵۰۰	−	−
	S.D	۰/۳۸۰	۰/۳۰۶	۰/۰۶۵	۰/۰۶۵	−	−
	p−value	۰/۳۹۳	۰/۰۰۱۱	<۰/۰۰۰۱	<۰/۰۰۰۱	−	−
پواسن صفرمتورم	E.S.	۰/۵۸۲	−۱/۱۴۷	۰/۵۰۰	۰/۵۳۰	−	۰/۱۲۰۷
	S.D	۰/۵۸۳	۰/۳۷۷	۰/۰۶۷	۰/۹۷	−	۰/۱۹۶۶
	p−value	۰/۳۱۹	۰/۰۰۲۷	<۰/۰۰۰۱	<۰/۰۰۰۱	−	۰/۵۳۹۹
دوجمله‌ای منفی	E.S.	۰/۳۲۶	−۱/۰۱۸	۰/۵۰۰	۰/۵۵۲	۰/۲۰۸	−
	S.D	۰/۳۹۳	۰/۳۱۴	۰/۰۶۸	۰/۱۰۷	۰/۲۷۵	−
	p−value	۰/۴۰۹	۰/۰۰۱۴	<۰/۰۰۰۱	<۰/۰۰۰۱	۰/۴۵۱	−
دوجمله‌ای منفی صفرمتورم	E.S.	۰/۳۲۶	−۱/۰۱۸	۰/۵۰۰	۰/۵۰۰	۰/۲۰۸	۰/۰۰۰
	S.D	۰/۳۹۳	۰/۳۱۴	۰/۰۶۸	۰/۰۶۸	۰/۲۷۵	۰/۰۰۱
	p−value	۰/۴۰۸	۰/۰۰۱۴	<۰/۰۰۰۱	<۰/۰۰۰۱	۰/۴۵۰۶	۰/۹۹۹۳



شکل ۱- احتمال رخداد واقعی و براوردشده‌ی خسارت فوتبالیست‌های باشگاه سپاهان بر اساس مدل‌های مختلف: ۱۳۹۲-۱۳۹۳

۵- بحث و نتیجه‌گیری

با توجه به این‌که در صنعت بیمه به ویژه در بیمه‌های خودرو و بیمه‌های ورزشی، ادعاهای خسارت صفر (عدم پرداخت خسارت) به‌دلایل زیادی از جمله عدم گزارش‌دهی در خصوص ادعاهای کوچک (به‌ویژه ادعاهایی که آثار آن بر حق بیمه‌ی سال آتی بیش‌تر از منافع مربوط به دریافت خسارت در سال جاری باشد) زیاد است، فراوانی صفر در نمونه‌ها به تعداد زیاد وجود داشته و تورم عدد صفر در داده‌ها بروز می‌نماید. در این موارد از مدل‌های پواسونی صفرمتورم یا مدل دوجمله‌ای منفی صفرمتورم برای تحلیل داده‌ها استفاده می‌شود لذا در قیمت‌گذاری بیمه‌های ورزشی، مدل‌های خطی تعمیم‌یافته نقش مهمی در علم محاسبات فنی بیمه ایفا می‌نماید.

در این مقاله فراوانی ادعاهای صدمات واردشده به بازیکنان تیم فوتبال سپاهان در سال ۱۳۹۲-۱۳۹۳، با استفاده از مدل‌های پواسونی، پواسونی صفرمتورم، دوجمله‌ای منفی و دوجمله‌ای منفی صفرمتورم، مورد بررسی و تحلیل قرار گرفت. همچنین به‌منظور بررسی تأثیر متغیرهای کمکی بر مدل‌بندی ادعاهای صدمات وارد شده، تأثیر متغیر کمکی موقعیت بازیکن به‌عنوان متغیر پیش‌گو وارد مدل شد.

نتایج نشان می‌دهد با و بدون در نظر گرفتن متغیر کمکی، توزیع دوجمله‌ای منفی و پواسونی بهترین برازش را دارد و متغیر کمکی موقعیت بازیکن در زمین نیز بر فراوانی خسارت وارد شده به بازیکن معنی‌دار است.

با توجه به عدم گسترش بیمه‌های ورزشی در کشور (علی‌رغم ظرفیت بالای این رشته‌ی بیمه‌ای در کشورهای پیشرفته) و با عنایت به افزایش تقاضای بیمه برای شرکت‌ها در حوزه‌ای که توان پرداخت حق بیمه‌ی بالایی دارد (به ویژه در رشته‌ی فوتبال)، این مطالعه زمینه‌ای جدید را برای توسعه‌ی تقاضای بیمه در ایران، معرفی می‌نماید. شایان ذکر است تنوع بازار بیمه و معرفی رشته‌های جدید، از جمله سیاست‌های مورد تأکید بیمه‌ی مرکزی جمهوری اسلامی ایران است.

ارزش یافته‌های این مطالعه از آنجا ناشی می‌شود که گزینش مدل مناسب در جهت پیش‌بینی‌ها و ارزیابی‌های مربوط به تعیین حق بیمه اهمیت اساسی دارد. این پژوهش مدل مناسب برای شرکت‌های فعال در رشته‌ی بیمه‌های ورزشی را مدل‌های صفرمتورم معرفی می‌کند. ضمناً، متغیرهای تأثیرگذار بر تعداد خسارات (نظیر موقعیت بازیکن) نیز مورد بررسی و ارزیابی قرارگرفت و این متغیرها می‌توانند برای راه‌اندازی این نوع بیمه‌ها توسط شرکت‌های بیمه‌ی علاقمند، مورد بهره‌برداری قرار گیرند.

با توجه به این‌که ضریب نفوذ بیمه در ایران پایین‌تر از کشورهای با توان اقتصادی مشابه می‌باشد، این مطالعه با طرح موضوعی جدید می‌تواند ملاک تصمیم‌گیری و سیاست‌گذاری برای بیمه‌ها قرار گیرد. در پایان پیش‌نهاد می‌شود، بیمه‌ی مرکزی جمهوری اسلامی ایران با ورود به این حوزه و تدوین ضوابط و مقررات، امکان فعالیت شرکت‌ها در این رشته را ایجاد نماید.

توضیحات

1. Generalized Linear Models
2. Sports Cover
3. Canadian Athlete Insurance Program
4. Basic
5. Superior
6. Zero- Inflated
7. Zero- Inflated Poisson

مرجع‌ها

[1] De Jong, P. and Heller, G. (2008). *Generalized Linear Models for Insurance Data*. Cambridge.

[2] Flynn, M. and Francis, L.A. (2010). *More Flexible GLMs Zero-Inflated Models and Hybrid Models*. Casualty Actuarial Society E-Forum.

[3] Hall, D.B. (2000). Zero-Inflated Poisson and Binomial Regression with Random Effects: A Case Study. *Biometrics*, **56**, 1030-1039.

[4] Hilbe, J.M. (2007). *Negative Binomial Regression*. Cambridge University Press.

[5] Heilbron, D.C. (1994). Zero-Altered and Other Regression Models for Count Data with Added Zeros. *Biometrical Journal*, **36**, 531-547.

[6] Ismail, N. and Jemain, A.A. (2007). Handling Over Dispersion with Negative Binomial and Generalized Poisson Regression Models. CAS Winter Forum including the Ratemaking Call Papers.

[7] Lambert, D. (1992). Zero-Inflated Poisson Regression with an Application to Defects in Manufacturing. *Technometrics*, **34**, 1-14.

[8] Nelder, J.A. and Wedderburn, R.W.M. (1972). Generalized Linear Models. *Journal of the Royal Statistical Society* A, **137**, 370-384.

[9] Winklemann, R. and Zimmermann, K.F. (1995). Recent Developments in Count Data Modeling: Theory and Application. *Journal of Economic Surveys*, **9**, 1-21.

پیوست

B.1 Poisson model via Proc NLMIXED

```
proc nlmixed data=claims1;
parms b0=0;
eta=b0;
lambda = exp(eta);
model clm_freq~poisson(lambda);
estimate 'lambda' lambda;
title 'Poisson model via Proc NLMIXED';
```

```
run;
```

2 Negative binomial model via Proc NLMIXED

```
proc nlmixed data=claims1;
parms b_0=0 ;
eta = b_0;
mu = exp(eta);
y = clm_freq;
loglike = (lgamma(y + (1/k)) - lgamma(y + 1) -
lgamma(1/k) + y*log(k*mu) - (y + (1/k))*log(1 + k*mu));
 model y ~ general(loglike);
predict mu out=out1(keep=clm_freq pre);
contrast 'k = 0' k - 0;
estimate 'exp(b_0)' exp(b_0);
estimate 'mean' mu;
estimate 'k' k;
estimate 'variance' mu + k*mu**2;
title 'Negative Binomial model via Proc NLMIXED';
run;
```

C.1 ZIP via proc Nlmixed

```
 proc nlmixed data=claims1;
parms p_0=0.57 bll_0=0.5;
bounds 0<p_0<1;
eta = bll_0;
lambda = exp(eta);
y = clm_freq;
if y=0 then loglike = log(p_0 + (1 - p_0)*exp(-lambda));
else loglike = log(1 - p_0) + y*log(lambda) - lambda -
lgamma(y + 1);
model y ~ general(loglike);
contrast 'p_0 - 0' p_0 - 0;
estimate "p_0" p_0;
estimate "Expected zeros=exp(-lambda)" exp(-lambda);
estimate 'Conditional Poisson Mean (lambda)' lambda;
estimate 'ZIP Mean (1-p_0)*lambda' (1 - p_0)*lambda;
estimate 'ZIP Var(1-p_0)*lambda*(1+lambda-(1-
p_0)*lambda)'
(1 - p_0)*lambda*(1 + lambda - (1 - p_0)*lambda);
estimate 'theta=p_0/(1-p_0)' p_0/(1 - p_0);
predict p_0 out=pred_zi(keep=pred);
predict lambda out=pred(keep=clm_freq pred);
predict (1 - p_0)*lambda out=out1(keep=clm_freq pred);
title 'Zero-Inflated Poisson (ZIP) distribution';
run;
```

C.2 ZINB via Proc Nlmixed

```
proc nlmixed data=claims1;
PARMS a0 = 0 b0 = 0;
zpart= a0;
infprob= 1/(1+exp(zpart));
eta_nb= b0;
mean = exp(eta_nb);
p0 = infprob+(1-infprob)*exp(-
(clm_freq+(1/k))*log(1+k*mean));
p_else = (1-infprob)*exp(lgamma(clm_freq+(1/k))-
lgamma(clm_freq+1)-
lgamma(1/k)+clm_freq*log(k*mean)-
(clm_freq+(1/k))*log(1+k*mean));
if clm_freq = 0 then loglike = log(p0);
else loglike = log(p_else);
model clm_freq~ general(loglike);
estimate "Estimated proportion of 'extra' zeros (theta)"
infprob;
estimate 'Estimated Conditional Poisson Mean (mean)'
mean;
estimate 'ZINB Mean(1-infprob)*mean' (1-infprob)*mean;
estimate 'ZINB Var (1-
infprob)*mean*(1+(mean*infprob))'(1-
nfprob)*mean*(1+(mean*infprob));
title 'Zero-inflated Negative Binomial ZINB
Distribution';
run;
```

D.1 Poisson SAS code:

```
proc nlmixed data=claims2;
parms a0= 0 a1=0 ;
eta=a0+a1*po;
lambda= exp(eta);
model clm_freq~poisson(lambda);
estimate 'eta' eta;
estimate 'lambda' lambda;
title'poisson regression';
run;
```

D.2 Negative Binomial regression SAS code:

```
proc nlmixed data= claims2;
parms a0= 0 a1 =0;
linpinfl= a0+a1*po;
lambda =exp(linpinfl);
```

```
II = lgamma(clm_freq+(1/k))-lgamma(clm_freq+1)-
lgamma(1/k) +clm_freq*log(k*lambda)-
(clm_freq+(1/k))*log(1+k*lambda);
model clm_freq~general(II);
predict mu out=out1(keep=clm_freq pre);
contrast 'k = 0' k - 0;
estimate 'linpinfl'linpinfl;
estimate 'mean'exp(linpinfl) ;
estimate 'k' k;
estimate 'variance' lambda+ k*lambda**2;
title 'Negative Binomial model via Proc NLMIXED';
run;
```

D.3 ZIP regression SAS code:

```
proc nlmixed data=claims2;
parms a0= 0 a1= 0 b0= 0.5 b1=0  ;
linpinfl= a0+a1*po;
eta= b0+b1*po;
lambda= exp(eta);
infprob= 1/(1+exp(linpinfl));
if clm_freq =0 then ll=log(infprob+(1-infprob)*exp(-
lambda));
else ll= log((1-infprob))+clm_freq*log(lambda)-
lgamma(clm_freq+1)-lambda;
model clm_freq ~general(ll);
ESTIMATE 'infprob 1/(1+exp(linpinfl))'
1/(1+exp(linpinfl));
estimate 'Conditional Poisson Mean (lambda)' lambda;
estimate 'ZIP Mean (1-infprob)*lambda' (1 -
infprob)*lambda;
estimate 'ZIP Var(1-infprob)*lambda*(1+lambda-(1-
infprob)*lambda)'
(1 - infprob)*lambda*(1 + lambda - (1 -
infprob)*lambda);
title'ZIP via nlmixed';
run;
```

D.4 ZINB regression SAS code:

```
proc nlmixed data=claims2;
parms a0=0 a1=0 b0=0 b1= 0;
linpinfl=a0+a1*po;
infprob = 1/(1+exp(linpinfl));
eta_nb =b0+b1*po;
lambda= exp(eta_nb);
```

```
p0 =infprob+(1-infprob)*exp(-
(clm_freq+(1/k))*log(1+k*lambda));
p_else =(1-infprob)*exp(lgamma(clm_freq+(1/k))-
lgamma(clm_freq+1)-lgamma(1/k)+clm_freq*log(k*lambda)-
(clm_freq+(1/k))*log(1+k*lambda));
if clm_freq =0 then loglike= log(p0); else loglike
=log(p_else);
model clm_freq~general(loglike);
ESTIMATE 'infprob' 1/(1+exp(linpinfl));
estimate "Estimated proportion of 'extra' zeros (theta)"
infprob;
estimate 'Estimated Conditional Poisson Mean (lambda)'
lambda;
estimate 'ZINB Mean(1-infprob)*lambda' (1-
infprob)*lambda;
estimate 'ZINB Var (1-
infprob)*lambda*(1+(lambda*infprob))'(1-
infprob)*lambda*(1+(lambda*infprob));
title'ZINB VIA NLMIXED';
run;
```

مریم کریمی باجه باج
کارشناسی ارشد علوم محاسبات و برنامه‌ریزی بیمه (اکچوئری)
تهران، خیابان شهید دکتر بهشتی، خیابان خالد اسلامبولی (وزرا)، کوچه‌ی هفتم، شماره‌ی ۱۴
رایانشانی: actvary.mk@gmail.com

آتوسا گودرزی
دکتری علوم اقتصادی
تهران، خیابان شهید دکتر بهشتی، خیابان خالد اسلامبولی (وزرا)، کوچه‌ی هفتم، شماره‌ی ۱۴
رایانشانی: atousagoodarzi@yahoo.com

زهرا رضایی قهرودی
دکتری آمار
تهران، خیابان شهید فاطمی، خیابان باباطاهر، خیابان سرتیپ فکوری، شماره‌ی ۱۴۵، پژوهشکده‌ی آمار.
رایانشانی: z_rezaei@srtc.ac.ir