مقایسه مدل شبکه عصبی مصنوعی با مدل‌های رگرسیونی داده‌های شمارشی در پیش بینی تعداد دفعات اهدای خون

شیما حقانی: کارشناس ارشد آمار زیستی، دانشکده بهداشت، دانشگاه علوم پزشکی شهرکرد، شهرکرد، ایران. shima_haghani@yahoo.com
***مرتضی سدهی:** استادیار گروه آمار زیستی، دانشکده بهداشت، دانشگاه علوم پزشکی شهرکرد، شهرکرد، ایران (*نویسنده مسئول). sedehi56@gmail.com
سلیمان خیری: دانشیار گروه آمار زیستی، دانشکده بهداشت، دانشگاه علوم پزشکی شهرکرد، شهرکرد، ایران. kheiri@hbi.ir

تاریخ دریافت: ۱۳۹۳/۳/۲۳ تاریخ پذیرش: ۱۳۹۳/۶/۸

چکیده

زمینه و هدف: مدل‌بندی یکی از روش‌های مهم برای تبیین رابطه بین متغیر پاسخ و مستقل می‌باشد. از آنجا که داده‌های مربوط به تعداد دفعات اهدای خون به صورت داده‌های شمارشی (گسسته) می‌باشد، جهت تبیین آن‌ها مناسب‌تر است که از توزیع‌های متغیرهای گسسته مانند پواسن یا دوجمله‌ای منفی استفاده کرد. هدف از انجام این مطالعه تحلیل مدل‌های شمارشی به روش شبکه عصبی و مقایسه آن با روش‌های آماری کلاسیک و انتخاب بهترین روش برای پیش‌بینی تعداد دفعات اهدای خون می‌باشد.

روش کار: در این مطالعه از داده‌های مربوط به اهدای خون که در پایگاه انتقال خون شهرکرد جمع آوری شده است، استفاده گردید و چهار مدل رگرسیونی پواسن، دوجمله‌ای منفی و حالت‌های صفر انبوه آن ها با روش شبکه عصبی مصنوعی مورد مقایسه قرار گرفت. برای آموزش شبکه عصبی مصنوعی از الگوریتم آموزشی BFGS (Broyden–Fletcher–Goldfarb–Shanno algorithm) و برای تعیین مناسب‌ترین مدل از معیار میانگین مربعات خطا (mean-square error) (MSE) استفاده شد. بهترین ساختار شبکه در داده‌های آموزش انتخاب و دقت روش شبکه عصبی با بهترین ساختار در داده‌های آموزش با مدل‌های رگرسیونی کلاسیک مورد مقایسه قرار گرفت تا بهترین روش برای پیش بینی تعداد دفعات مجدد اهدای خون انتخاب گردد.

یافته‌ها: میزان MSE برای مدل های رگرسیونی پواسن، پواسن با صفر انبوه، دوجمله ای منفی و دوجمله ای منفی با صفر انبوه به ترتیب برابر با ۲/۷۱، ۱/۵۴، ۰/۹۴ و ۱/۰۱ و برای روش شبکه عصبی مصنوعی ۱۴:۱۷:۱ با تابع تبدیل تانژانت هایپربولیک هم در لایه میانی و هم در لایه خروجی این معیار ۰/۰۵۶ بدست آمد.

نتیجه‌گیری: نتایج مطالعه نشان داد که، با توجه به میزان MSE می‌توان روش شبکه عصبی مصنوعی را مناسب‌ترین روش با بالاترین دقت جهت پیش‌بینی تعداد دفعات اهدای مجدد خون نسبت به مدل‌های مورد بررسی در این پژوهش دانست.

کلیدواژه‌ها: رگرسیون پواسن، پواسن با صفر انبوه، دوجمله‌ای منفی، دوجمله‌ای منفی با صفر انبوه، شبکه عصبی مصنوعی.

مقدمه

یکی از اساسی‌ترین روش‌ها برای پیدا کردن رابطه بین متغیرها مدل‌بندی می‌باشد. برای پیش‌بینی و ارتباط بین متغیر پاسخ و متغیرهای مستقل از مدلبندی استفاده می‌شود. در مواقعی که داده‌های مربوط به متغیر پاسخ، متغیر شمارشی باشد برای مدل‌بندی آن مناسب‌تر است که از توزیع‌های مربوط به متغیرهای گسسته مانند توزیع پواسن یا دوجمله‌ای منفی استفاده شود (۱). جهت این مدل بندی اگر مقادیر میانگین و واریانس داده‌های متغیر پاسخ با یکدیگر برابر باشند از توزیع پواسن و اگر مقدار واریانس از مقدار میانگین بزرگتر باشد (پراکندگی زیاد متغیر شمارشی) مدل دوجمله‌ای منفی، توزیع مناسبی جهت تبیین متغیر پاسخ شمارشی بر اساس متغیرهای مستقل دیگر می باشد (۵-۲). هنگام کار بر روی داده‌های شمارشی در موضوعات واقعی غالباً با دو مسئله بیش پراکنش و وجود صفرهای زیاد مواجه می‌گردیم. با وجود مقادیر صفر زیاد برای متغیر پاسخ، مقادیر میانگین و واریانس متغیر پاسخ برابر نیستند، بنابراین مدل پواسن، مدل مناسبی برای این نوع داده‌ها نمی‌باشد. در این حالات خاص، استفاده از روشهایی مثل مدل رگرسیون پواسن با صفر انبوه (Zero inflated poisson regression)، رگرسیون دوجمله‌ای منفی با صفر انبوه (Zero inflated

[Downloaded from rjms.iums.ac.ir on 2024-03-29]

(negative binomial regression، مدل هاردل (Hurdle Model) و در نهایت مدل پواسن تعمیم‌یافته (Generalized Poisson Model) پیشنهاد شده است(۹-۶،۴-۲). از طرفی روشهای کلاسیک آماری برای مدلبندی روابط بین متغیرها دارای تعدادی پیش فرض و محدودیت است. از جمله این محدودیت‌ها می‌توان به یکسان بودن واریانس خطاها، در نظر گرفتن یک توزیع پیش فرض مانند نرمال برای متغیرهای پاسخ، خطی بودن رابطه بین متغیرهای وابسته و مستقل اشاره کرد (۱۰). علاوه بر این هیچ یک از این روشها قابلیت مدلبندی روابط پیچیده غیرخطی و اثرات متقابل درجه بالا را ندارند. حساس بودن بیشتر این مدل‌ها به مشاهده‌های گم شده و داده‌های پرت از دیگر محدودیت‌های این مدل‌ها به شمار می‌آیند (۱۱). با توجه به محدودیتهای ذکر شده برای روشهای کلاسیک، نیاز به روشهایی که با محدودیتهای کمتری در این زمینه مواجه باشند، احساس می‌شود. در این بین، شبکه عصبی مصنوعی می‌تواند یکی از مناسب‌ترین روش‌ها باشد. در مقایسه با مدل‌های آماری معمولی، شبکه‌های عصبی ابزار قدرتمندی برای مدلسازی غیرخطی می‌باشند. این روش‌ها به عنوان یک روش جایگزین مدلهای آماری برای داده‌های شمارشی به حساب می‌آیند (۱۲). شبکه‌های عصبی مصنوعی روش نوینی برای مدلسازی هستند که باتوجه به ساختار هوشمند و انعطاف‌پذیرشان رقیب مدلسازی های مرسوم آماری شده‌اند و چه در زمینه نظری و چه در زمینه کاربردی رو به پیشرفت دارند (۱۳). این روش‌هاهیچ فرض اولیه‌ای را بر توزیع داده‌ها تحمیل نمی‌کنند و به خصوص وقتی با ارزشند که ارتباط تابعی بین متغیرهای مستقل و وابسته مشخص نیست و همچنین در حالتی که حجم نمونه بزرگ باشد (۱۴). از دیگر مزایای شبکه‌های عصبی مصنوعی این است که داده‌ها در شبکه به صورت ضمنی پردازش می‌شوند، بر این اساس حتی اگر بخشی از لایه‌های شبکه حذف شوند یا عملکرد صحیحی نداشته باشند احتمال رسیدن به پاسخ صحیح همچنان وجود دارد. تعمیم‌پذیری شبکه این امکان را می‌دهد که مدل، در ارتباط با یک مشاهده جدید نیز پاسخ مناسبی ارائه دهد. از سوی دیگر داده‌های پزشکی از جمله داده‌های حساس هستند و نتایج حاصل از تحلیل آن‌ها باید دارای کمترین مقدار خطا باشد بنابراین باید به دنبال روش‌هایی دقیق برای آنالیز آن‌ها بود. به خصوص در مطالعات بهداشتی و پزشکی که تمرکز برسلامت افراد بوده و پیش‌بینی صحیح متغیر پاسخ بسیار مهم وحیاتی می‌باشد. به کارگیری مدلهای شبکه عصبی برای پیش‌بینی وضعیت این‌گونه برآمدها مناسب به نظر می‌رسد، زیرا این مدلها قادرند با کشف روابط موجود بین داده‌ها، خروجی مناسبی رابا کمترین مقدار خطا تولید نمایند (۱۵).

خون نیزجزئی از سیستم حیاتی بدن است که ویژگی‌های خاص آن باعث می‌شود بتوان با اهدای بخشی‌از این مایع اسرارآمیز جان یک انسان بیمار یا نیازمند را نجات داد. اهمیت این مسئله وقتی بیشترمی‌شود که بدانیم از هر ۳ انسان یک نفر در طول حیات به تزریق خون و فراورده‌های آن نیاز پیدا می‌کند. علی‌رغم تمام پیشرفت‌هایی که امروزه در حوزه‌های مختلف علم پزشکی انجام شده است هیچ گونه جایگزین مصنوعی برای خون پیدا نشده که بتوان با استفاده از آن نیاز بیماران مختلف را مرتفع ساخت و تنها راه رفع نیاز به این ماده حیاتی، خونی است که معمولاً اهدا یا خریداری می‌شود. انسان‌ها می‌توانند در طول دوره زندگی خود چندین مرتبه خون اهدا کنند. اهداکنندگانی که حداقل هر ۶ ماه یک بار خون اهداء می کنند، اهداکنندگان مستمر هستند؛ تعداد دفعات خون دادن این افراد مشخص است و همچنین از سلامت خون آن‌ها نیز اطمینان وجود دارد؛ بنابراین هر چه بتوان شاخص مستمر را افزایش داد، برای سیستم بهداشت مناسب‌تر است. خانم‌ها حداکثر می‌توانند ۳ بار در سال خون بدهند و آقایان نیز هر ۳ ماه یک‌بار امکان اهدای خون دارند (۱۶)؛ بنابراین پیش‌بینی تعداد دفعات اهدای خون جایگاه ویژه‌ای می‌یابد، پس مناسب است به دنبال روشی با دقت بالا جهت پیش بینی این تعداد باشیم.

[Downloaded from rjms.iums.ac.ir on 2024-03-29]

هدف از انجام این مطالعه مقایسه روش‌های کلاسیک آماری با روش شبکه عصبی مصنوعی بر روی تعداد دفعات اهدای مجدد خون در پایگاه انتقال خون شهر کرد بود.

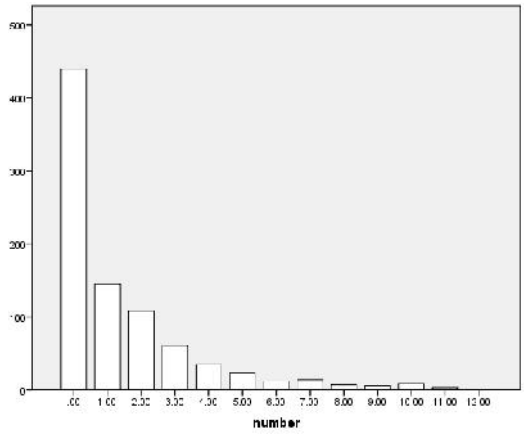
روش کار

این مطالعه بر روی داده‌هایی که به روش توصیفی تحلیلی از نوع مقطعی و از جامعه اهداکنندگان خون در شهرکرد بر روی یک نمونه ۸۶۴ نفری به روش نمونه‌گیری سیستماتیک جمع آوری شده، صورت گرفته است (۱۷).

با توجه به اینکه تعداد دفعات اهدای خون از یک بار به بالا توزیع شده و عدد صفر را در برنمی‌گیرد باید از مدل‌های تغییر یافته که شامل صفر نیستند مانند مدل‌های پواسن بریده شده در نقطه صفر (Zero-Truncated Poisson) و مدل دوجمله‌ای منفی بریده‌شده در نقطه صفر (Zero-Truncated Negative Binomial)، استفاده کرد. در این پژوهش یک متغیر جدید جهت تعداد دفعات اهدای خون تعریف شد. به این صورت که از هر مقدار موجود عدد یک کم شد (تعداد دفعات خون دهی مجدد)، در این صورت توزیع مقادیر تعداد دفعات اهدای خون مجدد دارای مقدار مشاهده صفر نیز گردید. از آنجایی که با این تغییر متغیر تعداد دفعات اهدای خون مجدد شامل تعداد زیادی صفر شد، بنابراین از مدل‌هایی که درصد بیشتری از مشاهده صفر را تبیین می‌کنند مانند توزیع پواسن با صفر انبوه یا توزیع دوجمله‌ای منفی با صفر انبوه استفاده گردید و سپس این مدل‌های رگرسیونی با روش شبکه عصبی مصنوعی مورد مقایسه قرار گرفت. برای مقایسه این دو روش ابتدا داده‌ها به دو قسمت آموزش و آزمون (۷۰ درصد از داده‌ها برای آموزش و ۳۰ درصد برای آزمون) تقسیم شد. برای داده‌های آموزش از یک شبکه عصبی پرسپترون سه لایه استفاده شد که لایه اول شامل متغیرهای ورودی که ۱۴ متغیر بود. لایه دوم، لایه میانی است که در اینجا ۴ تا ۱۸ نرون در این لایه مورد بررسی قرار گرفت که در انتها ۱۷ نرون در لایه میانی بهترین ساختار را نشان داد. لایه سوم که بر اساس نوع متغیر پاسخ تعیین می‌شود و در این مطالعه تعداد دفعات اهدای مجدد خون بود لذا لایه خروجی شبکه با یک نرون در نظر گرفته شد. در این مطالعه از الگوریتم یادگیری (Fletcher– Broyden –algorithm Shanno Goldfarb) BFGS که حالت خاصی از الگوریتم یادگیری شبه نیوتنی می‌باشد به کمک نرم‌افزار STATISTICA استفاده شده است. برای مقایسه چهار مدل رگرسیونی و انتخاب بهترین مدل رگرسیونی از همه داده‌های مربوط به تعداد دفعات اهدای مجدد خون استفاده گردید و برای مقایسه مدل‌ها ملاک اطلاع آکائیک و میانگین مربعات خطا به کار برده شد. معیار اطلاع آکائیک معیاری برای سنجش نیکویی برازش است که مقدار کمتر آن مدل مناسب تر را نشان می‌دهد. برای برازش مدل های رگرسیونی نرم افزار R 3.0.1 به کار برده شد. جهت مقایسه مدل های رگرسیونی و شبکه عصبی مصنوعی از مجموعه داده های آموزش استفاده شد و این مقایسه با معیار میانگین مربعات خطا (MSE) صورت گرفت.

یافته‌ها

از ۸۶۴ فرد مورد مطالعه، ۸۰۱ نفر (۹۲/۷٪) مرد و بقیه زن بودند و ۶۲۳ نفر (۷۲/۱٪) متأهل، ۷۱۰ نفر ساکن شهر بودند (۸۲/۲٪). ۱۹۴ نفر تحصیلات دانشگاهی داشته (۲۲/۵٪) و ۳۰۵ نفر از آن‌ها (۷۷/۵٪) تحصیلاتشان در سطح دیپلم و پایین‌تر بوده است. همچنین ۳۸۶ نفر از افرادی که



نمودار ۱- توزیع فراوانی تعداد دفعات اهدای خون مجدد در واحد های مورد پژوهش

[Downloaded from rjms.iums.ac.ir on 2024-03-29]

در پایگاه انتقال خون شهر کرد اقدام به اهدای خون کردند دارای شغل آزاد بودند (۴۴/۷٪). ۳۶۴ نفر از اهداکنندگان (۴۲/۱٪) دارای گروه خونی O، ۲۸۳ نفر گروه خونی A (۳۲/۸٪)، ۱۵۹ نفرگروه خونی B (۱۸/۴٪) و تنها ۵۸ نفر از افرادی که خون اهدا کردند گروه خونی شان AB بوده است (۶/۷٪). لازم به ذکر است که ۷۹۳ نفر از افراد اهدا کننده RH مثبت داشتند (۹۱/۸٪). میانگین تعداد دفعات اهدای خون مجدد ۱/۴۱ با انحراف معیار ۲/۱۶ بوده است. در این مطالعه ۵۰/۹٪ از افراد (۴۴۰ نفر) مراجعه مجدد به پایگاه انتقال خون برای اهدا نداشته‌اند (نمودار ۱).

در این مطالعه چهار مدل رگرسیونی برای داده‌های شمارشی مورد مقایسه قرار گرفتند. جدول شماره ۱ ضرایب مدل‌های رگرسیونی و مقادیر احتمال معناداری این ضرایب را نشان می‌دهد.

همان‌طور که در جدول شماره ۱ مشاهده می‌شود، با توجه به معیار اطلاع آکائیک رگرسیون دوجمله‌ای منفی و رگرسیون دوجمله‌ای منفی با صفر انبوه عملکرد مشابهی دارند ولی با توجه با معیار میانگین مربعات خطا (MSE) این مقدار برای رگرسیون دوجمله‌ای منفی برابر با ۰/۹۶ و برای رگرسیون دوجمله‌ای منفی با صفر انبوه برابر با ۱/۰۳ شده است.

برای پیدا کردن بهترین ساختار شبکه عصبی مصنوعی شبکه‌هایی با ۴ تا ۱۸ نرون در لایه میانی با تابع تبدیل تانژانت هایپربولیک در لایه خروجی و چهار تابع تبدیل مختلف در لایه میانی توسط نرم افزار STATISTICA ساخته شد و مورد مقایسه قرار گرفت. این چهار تابع تبدیل عبارتند از لجستیک، خطی، نمایی و تانژانت هایپربولیک. جدول شماره ۲ قسمتی از این ساختارها را نشان می‌دهد.

جدول ۱- برآورد ضرایب مدل‌های رگرسیونی تعداد دفعات اهدای خون مجدد در مراجعین به پایگاه انتقال خون شهرکرد

		رگرسیون پواسن		رگرسیون دوجمله‌ای منفی		رگرسیون پواسن با صفر انبوه		رگرسیون دوجمله‌ای منفی با صفر انبوه	
متغیر		برآورد ضریب	مقدار احتمال	برآورد ضریب	مقدار احتمال	برآورد ضریب	مقدار احتمال	برآورد ضریب	مقدار احتمال
	سن	-۰/۰۱۱	*۰/۰۰۱	-۰/۰۱۱	۰/۰۹۱	-۰/۰۰۶	۰/۱۲۲	-۰/۰۱۰	۰/۰۸۵
	جنس	۰/۳۹۹	۰/۱۳۲	۰/۲۸۹	۰/۴۸۸	۰/۵۴۷	*۰/۰۵۴	۰/۳۳۸	۰/۳۸۶
	وزن	۰/۰۲۴	*<۰/۰۰۱	۰/۰۲۸	*<۰/۰۰۱	۰/۰۱۴	*<۰/۰۰۱	۰/۰۲۶	*<۰/۰۰۱
	وضعیت تأهل	-۰/۰۰۴	۰/۹۶۵	-۰/۰۵۰	۰/۷۴۲	-۰/۰۷۴	۰/۴۱۱	-۰/۰۶۳	۰/۶۵۸
	محل اقامت	-۰/۱۶۲	*۰/۰۲۶	-۰/۱۴۸	۰/۲۸۷	-۰/۰۸۸	۰/۲۷۲	-۰/۱۳۵	۰/۳۰۹
	سطح تحصیلات	-۰/۰۷۷	*۰/۰۳۶	-۰/۰۶۷	۰/۳۲۶	-۰/۰۸۵	*۰/۰۴۳	-۰/۰۷۴	۰/۲۶۳
گروه شغلی	خانه‌دار در مقابل بیکار	-۰/۶۹۴	*۰/۰۳۱	-۰/۸۳۲	۰/۱۰۰	-۰/۶۳۱	۰/۰۷۶	-۰/۸۱۹	۰/۰۸۶
	کارمند در مقابل بیکار	-۰/۰۳۸	۰/۷۲۵	-۰/۰۴۳	۰/۸۳۵	-۰/۲۴۷	*۰/۰۲۸	-۰/۱۰۸	۰/۵۷۱
	کارگر-کشاورز-راننده در مقابل بیکار	-۰/۲۳۶	*۰/۰۴۳	-۰/۲۴۸	۰/۲۶۲	-۰/۱۹۶	۰/۱۲۱	-۰/۲۴۹	۰/۲۴۳
	شغل آزاد در مقابل بیکار	-۰/۳۱۸	*<۰/۰۰۱	-۰/۳۰۵	۰/۰۷۹	-۰/۴۱۴	*<۰/۰۰۱	-۰/۳۴۵	*۰/۰۴۳
گروه خونی	ABدر مقابل O	۰/۰۲۶	۰/۸۱۹	۰/۰۱۴	۰/۹۴۷	۰/۱۲۷	۰/۳۱۲	۰/۰۴۹	۰/۸۱۷
	Bدر مقابل O	-۰/۰۹۵	۰/۲۴۱	-۰/۰۸۵	۰/۵۶۷	-۰/۰۶۹	۰/۴۳۸	-۰/۰۷۹	۰/۵۸۱
	Aدر مقابل O	-۰/۱۶۲	*۰/۰۱۶	-۰/۱۸۸	۰/۱۳۰	-۰/۱۳۷	۰/۰۶۸	-۰/۱۷۸	۰/۱۳۵
	Rh	۰/۱۹۶	۰/۰۸۵	۰/۲۱۷	۰/۲۷۵	۰/۲۶۷	*۰/۰۳۴	۰/۲۳۲	۰/۲۱۸
	مقدار ثابت	-۱/۱۶۱	*<۰/۰۰۱	-۱/۳۹۷	*۰/۰۲۶	-۰/۰۳۹	۰/۹۲۳	-۱/۰۶۹	۰/۱۰۳
	معیار AIC	۳۳۷۶/۵		۲۷۴۵/۸		۲۸۹۵/۷		۲۷۴۴/۹۶	
	میانگین مربعات خطا(MSE)	۲/۵۸		۰/۹۷		۱/۴۵		۱/۰۳	

[Downloaded from rjms.iums.ac.ir on 2024-03-29]

جدول ۲- مقادیر مربعات خطابرای شبکه‌های عصبی با ساختار مختلف در داده‌های آموزش مربوط به تعداد دفعات اهدای خون مجددر پایگاه انتقال خون شهرکرد

	تعداد نرون لایه	تابع تبدیل			
ردیف	میانی	خطی	لجستیک	نمایی	تانژانت هایپربولیک
۱	۱۱	۰/۱۹۰۴	۰/۱۰۶۶	۰/۱۳۲۷	۰/۱۰۰۴
۲	۱۲	۰/۱۹۰۴	۰/۰۹۵۴	۰/۱۲۲۴	۰/۰۹۵۶
۳	۱۳	۰/۱۹۰۴	۰/۰۹۲۵	۰/۱۲۲۳	۰/۰۷۹۵
۴	۱۴	۰/۱۹۰۴	۰/۰۸۱۳	۰/۱۱۰۱	۰/۰۸۱۴
۵	۱۵	۰/۱۹۰۴	۰/۰۷۶۱	۰/۱۱۶۲	۰/۰۷۳۰
۶	۱۶	۰/۱۹۰۴	۰/۰۸۴۹	۰/۱۱۱۹	۰/۰۶۱۲
۷	۱۷	۰/۱۹۰۴	۰/۰۷۱۷	۰/۱۰۲۰	*۰/۰۵۶۳
۸	۱۸	۰/۱۹۰۴	۰/۰۷۳۴	۰/۱۰۴۲	۰/۰۶۸۴

جدول ۳- مقایسه عملکرد شبکه عصبی مصنوعی با چهار مدل رگرسیونی مربوط به تعداد دفعات اهدای خون مجدد در پایگاه انتقال خون شهرکرد

	MSE
	آموزش
شبکه عصبی مصنوعی ۱۴:۱۷:۱	۰/۰۵۶
رگرسیون پواسن	۲/۷۱
رگرسیون پواسن با صفر انبوه	۱/۵۴
رگرسیون دوجمله‌ای منفی	۰/۹۴
رگرسیون دوجمله‌ای منفی با صفر انبوه	۱/۰۱

در نهایت مقایسه مدل‌های رگرسیونی و شبکه عصبی مصنوعی بر روی داده های آموزش انجام گرفت. جدول شماره ۳ نتایج عملکرد شبکه عصبی مصنوعی و چهار مدل رگرسیونی را نشان می دهد.

بحث و نتیجه گیری

هدف از انجام این مطالعه مقایسه مدل شمارشی به روش شبکه عصبی با روش‌های آماری کلاسیک بود. در این پژوهش روش شبکه عصبی مصنوعی و روش‌های داده‌های شمارشی از جمله رگرسیون پواسن و دوجمله‌ای منفی و همچنین رگرسیون پواسن با صفر انبوه و رگرسیون دوجمله‌ای منفی با صفر انبوه مورد مقایسه قرار گرفت و روش دقیق‌تر برای پیش‌بینی تعداد دفعات اهدای مجدد خون انتخاب گردید. در مطالعه حاضر میانگین و انحراف معیار تعداد دفعات اهدای مجدد خون به ترتیب ۱/۴۱ و ۲/۱۶ بدست آمد بنابراین مسئله‌ی بیش پراکنشی وجود داشت. همچنین فراوانی در نقطه صفر ۴۴۰ نفر (۵۰/۹٪) بود، بنابراین مقادیر صفر زیاد مشاهده شد.

با توجه به نوع متغیر پاسخ روش رایج جهت مدلبندی و تعیین عوامل مؤثر بر مقادیر تعداد دفعات اهدای مجدد خون استفاده از روش‌های رگرسیونی مربوط به داده‌های شمارشی است. در این مطالعه معیار MSE به عنوان معیاری برای مقایسه مدل ها انتخاب شد. براساس این معیار، مدل با مقدار کمتر MSE پیش بینی بهتری خواهد داشت. باتوجه به این معیار مشاهده شد که روش شبکه عصبی مصنوعی عملکرد بهتری نسبت به روش‌های رگرسیونی داشت. رفیعی و همکاران در سال ۱۳۸۳ مطالعه‌ای مشاهده‌ای - مقطعی تحت عنوان کارایی مدل دوجمله‌ای منفی با انباشتگی بیش از حد انتظار صفر در تحلیل مدت اقامت مادران پس از زایمان انجام دادند. آن‌ها برای انتخاب مناسب‌ترین مدل کلاسیک از آماره انحراف استفاده کردند و در نهایت نتایج نشان داد که مدل دوجمله‌ای منفی با صفر انبوه با کمترین مقدار آماره انحراف مناسب‌ترین مدل می‌باشد (۵). در مطالعه‌ای که Mouatassim و همکاران در سال ۲۰۱۲ تحت عنوان رگرسیون پواسن و رگرسیون پواسن با صفر انبوه: به کارگیری در داده‌های بیمه سلامت خصوصی انجام داده‌اند، رگرسیون پواسن و رگرسیون پواسن با صفر انبوه را برای تعداد درخواست‌های رسیده از بیمه در هر سال، مورد

[Downloaded from rjms.iums.ac.ir on 2024-03-29]

مقایسه قرار دادند. آن‌ها برای مقایسه دو مدل از آزمون وونگ و آزمون تبدیل انتگرال احتمالاتی استفاده کردند و به این نتیجه رسیدند که رگرسیون پواسن با صفر انبوه عملکرد بهتری داشته است (۱۸).

فلاح و همکاران در یک مطالعه طولی در کانادا با عنوان عملکرد رگرسیون پواسن شبکه‌های عصبی در پیش‌بینی تغییرات عمده شناختی، شبکه عصبی مصنوعی را با رگرسیون پواسن بر روی تغییرات شناختی در افراد مسن در یک پیگیری پنج ساله مورد مقایسه قرار دادند و عملکرد این دو روش را با مقایسه میزان MSE آن‌ها بررسی کردند. آن‌هااین میزان را برای رگرسیون پواسن ۸/۳۹ و برای شبکه عصبی مصنوعی با ۲،۳ و ۹ لایه پنهان به ترتیب ۶/۹۶، ۶/۹۴ و ۸/۲۵ بدست آوردند. فلاح و همکاران به این نتیجه رسیدند که شبکه عصبی با هر ساختاری عملکرد بهتری نسبت به رگرسیون پواسن دارد (۱۲).

در مطالعه‌ای که لی ین چنگ تحت عنوان تجزیه تحلیل فراوانی تصادفات آزاد راه: رگرسیون دوجمله‌ای منفی در مقابل شبکه عصبی مصنوعی، انجام داده است تعداد تصادفاتی که در بخشی از بزرگراهی در تایوان در یک دوره زمانی یکساله اتفاق افتاده را مدلبندی کرده است. او به این نتیجه رسید که برای پیش‌بینی یک یا بیشتر از یک تصادف، شبکه عصبی مصنوعی عملکرد بهتری دارد درحالی‌که اگر انباشتگی در مقدار صفر باشد، رگرسیون دوجمله‌ای منفی ۹۷٪ پیش‌بینی درست و شبکه عصبی مصنوعی ۹۰٪ پیش‌بینی درست دارد (۶). هرچند با توجه به اینکه در این مطالعه نیز انباشتگی در نقطه صفر وجود داشته، اما وی این حالت را در نظر نگرفته بود.

در این مطالعه میزان MSE برای مدل‌های رگرسیونی پواسن، پواسن با صفر انبوه، دوجمله‌ای منفی و دوجمله‌ای منفی با صفر انبوه به ترتیب برابر با ۲/۷۱، ۱/۵۴، ۰/۹۴ و ۱/۰۱ بدست آمد، بنابراین رگرسیون دوجمله‌ای منفی در بین مدل‌های مختلف رگرسیونی بهترین مدل برای پیش‌بینی انتخاب گردید؛ اما در مقایسه این روش‌ها با شبکه عصبی مصنوعی باتوجه به مقادیر MSE این نتیجه حاصل شد که شبکه عصبی مصنوعی می‌تواند روش مناسب‌تری برای پیش‌بینی باشد چرا که این میزان برای روش شبکه عصبی مصنوعی ۰/۰۵۶ بدست آمد.

لازم به ذکر است که باتوجه به این که برای انجام این مطالعه ازمجموعه داده های از قبل جمع آوری شده استفاده گردید، امکان بررسی متغیرهای بیشتری که ممکن بود در مدل‌های رگرسیونی معنی دار شوند و باعث بهتر شدن برازش این مدل ها گردند نبود، لذا این مورد می‌تواند از محدودیت های این پژوهش محسوب شود.

منابع

1. Rajaeefard AS, Rafiee M. Modeling of Hospital length of stay based on mixed Poisson distribution. Journal of Babol University of Medical Sciences. 2006; 8(3):36-43. [Persian]
2. Rafii M, Ayatollahi M, Behboodian J. Performance of Zero Inflated Negative binomial model in analysis of Length of stay postpartum Mothers, the sixth year of Lorestan University of Medical Sciences Journal. 2004;23. [Persian]
3. Saffari E, Adnan R. Parameter Estimation on Zero-Inflated Negative Binomial Regression with Right Truncated Data. Sains Malaysiana. 2012;41(11): 1483–1487.
4. Saffari E, Adnan R, Greene W. Hurdle negative binomial regression model with right censored count data. SORT. 2012;36 (2):181-194.
5. Famoye F, Wulu JT, Singh KP. On the generalized poisson regression model with an application to accident data. J Data Sci. 2004; 2:287-95.
6. Chang L. Analysis of Freeway Accident Frequencies: Negative Binomial Regression versus Artificial Neural Network. Safety Science. 2005;43(8): 541-557.
7. Hassan Zadeh F, Kazemi A, Kelishadi R. A Survey on the Relationship of Metabolic Syndrome Components and the Number of Blood Cells Using Count Data Regression Model. Journal of Isfahan Medical School. 2011;29(125). [Persian]
8. Lee AH, Wang K, Yau KKW. Truncated negative binomial mixed regression modeling of ischaemic stroke hospitalizations. Statistics in Medicine. 2003; 22:1129-1139.
9. Lee AH, Xiang L, Fung WK. Sensitivity of score tests for zero-inflation in count data. Statistics in Medicine. 2004; 23:2757-69

[Downloaded from rjms.iums.ac.ir on 2024-03-29]

10. Sedehi M, Mehrabi Y, Kazemnejad A, Hadaegh F. Comparison of Artificial Neural Network, Logistic Regression and Discriminant Analysis Methods in Prediction of Metabolic Syndrome . Iranian Journal of Endocrinology and Metabolism. 2010;11(6).[Persian]

11. Sedehi M, Mehrabi Y, Kazemnejad A, Joharimajd V, Hadaegh F. Artificial Neural Network Disign for Modeling of Mixed Bivariate Outcomes in Medical Research Data .Iranian Journal of Epidemiology. 2010;6(4).[Persian]

12. Fallah N, Mitnitski A , Rockwood K. [Applying neural network Poisson regression to predict cognitive score changes, Journal of Applied Statistics. 2011.38: 2051–2062]

13. Sadat Hashemi M, Kazem Nejad A, Kavehee B. Application of artificial neural networks in predicting pregnancy and compared them with conventional statistical methods. Journal of Shaeed Sdoughi University of Medical Sciences Yazd. 2003;11;10-15. [Persian]

14. Biglarian A, Babayee GH, Azmi R. Application of artificial neural network model to determine the most important predictors of in-hospital mortality after open-heart surgery and comparison with logistic regression model. Modares Journal of Medical Sciences. 2004;7:23-29. [Persian]

15. Biglarian A, Bakhshi E, Rahgozar M, Karimloo M. Comparison of Artificial Neural Network and Logistic Regression in Predicting of Binary Response for Medical Data: the stage of disease in Gastric Cancer. Journal of North Khorasan University of Medical Sciences, 2011; 3 (Biostatistics & Epidemiology Supplementary):22. [Persian]

16. http: www.jamejamonline.ir.

17. Kheiri S. Analysis of return Intervals of blood donors at Shahrekord Blood transfusion center, Report Research, Shahrekord University of Medical Sciences. 2013;950.

18. Mouatassim Y, Ezzahid E. Poisson regression and zero-inflated Poisson regression: application to private health insurance data. Springer. 2012.

[Downloaded from rjms.iums.ac.ir on 2024-03-29]

[Downloaded from rjms.iums.ac.ir on 2024-03-29]

Comparison of artificial neural network model with count data regression models for prediction of blood donation

Shima Haghani, MSC of Biostatistics, School of Public Health, Shahrekord University of Medical Sciences, Shahrekord, Iran. shima_haghani@yahoo.com
***Morteza Sedehi**, PhD, Assistant Professor of Biostatistics, School of Public Health, Shahrekord University of Medical Sciences, Shahrekord, Iran (*Corresponding author). sedehi56@gmail.com
Soleiman Kheiri, PhD, Associate Professor of Biostatistics, School of Public Health, Shahrekord University of Medical Sciences, Shahrekord, Iran. kheiri@hbi.ir

Abstract

Background: Modeling is one of the most important ways for explanation of relationship between dependent and independent response. Since data, related to number of blood donations are discrete, to explain them it is better to use discrete variable distribution like Poison or Negative binomial. This research tries to analyze numerical methods by using neural network approach and compare it by classic statistical methods to choose better way to predict the number of blood donations.

Methods: In this study, data were collected from blood donors at the blood center of the Shahrekord and then four methods were compared by neural network approach. These methods are: Poisson regression model and its zero inflated, Negative binomial models and its zero inflated. To learn neural network approach, (BFGS) Broyden–Fletcher–Goldfarb–Shanno algorithm was used. To choose the best model, mean-square error (MSE) was used. The best network structure in teaching data was chosen and neural network approach resolution was compared by them, to choose the best approach for prediction the number of blood donations.

Results: The MSE for Poisson regression model, Poisson regression with zero inflated, negative binomial and negative binomial with zero inflated are respectively 2.71, 1.54, 0.94 and 1.01. For neural network approach 14:17:1 with activation function of hyperbolic tangent in hidden layer and output layer 0.056 is achieved.

Conclusion: This study showed that catastrophic health care costs of the various groups and households had significant difference, of course, clearly, with the performance of the health system transformation project; changes have taken place in this important indicator and factors affecting it that requires a separate study and research.

Keywords: Catastrophic health care costs, Urban HEART survey, Tehran.

[Downloaded from rjms.iums.ac.ir on 2024-03-29]