مجله علمی – پژوهشی مدیریت تولید و عملیات
سال دوم، شماره دوم، بهار و تابستان ۱۳۹۰
تاریخ وصول: ۶ / ۱۱/ ۸۹
تاریخ پذیرش: ۲۴/ ۳/ ۹۰
صص: ۲۲-۱

بکارگیری تخیمن پارامتر برای بهبود شاخص‌های ارتباطی در رگرسیون‌لجستیک باینری

مهدی بشیری*[۱]، رضا کامران‌راد[۲]
۱- استادیار گروه مهندسی صنایع دانشگاه شاهد
۲- دانشجوی کارشناسی ارشد مهندسی صنایع دانشگاه شاهد

چکیده

هدف از این مقاله، تخمین پارامترهای مدل رگرسیون لجستیک باینری بمنظور حداکثرسازی تابع لگاریتم درست‌نمایی با بهبود شاخص‌های ارتباطی است. رگرسیون لجستیک نیز همانند سایر روش‌های تخمین پارامتر نیاز به یک معیار برای مناسب بودن تخمین پارامترهای آن دارد که از جمله آنها می‌توان به شاخص‌های ارتباطی اشاره نمود. شاخص‌های ارتباطی، شاخص‌هایی هستند که تعداد وقوع پاسخ موفقیت در مقابل پاسخ شکست را برای تعداد مشخصی آزمایش مستقل برنولی نشان می‌دهد. ماهیت این شاخص‌ها به‌گونه‌ای است که از یک مرحله معینی علی‌رغم بهبود پارامترهای تخمینی، مقدار شاخص تغییر نمی‌کند. لذا نمی‌توان آن شاخص روند بهبود تخمین پارامتر را نشان دهد. در این مقاله به معرفی یک شاخص جدید مبتنی بر شاخص‌های موجود و میزان احتمالات رگرسیون لجستیک پرداخته می‌شود که دقت بالاتری در تشخیص میزان مناسب بودن پارامترهای تخمین زده شده رگرسیون لجستیک دارد. نتایج بررسی یک مثال و مقایسه آن با شاخص ارتباطی نشان‌دهنده دقت بالاتر و مناسب بودن شاخص‌های پیشنهادی دارد.

واژه‌های کلیدی: تخمین پارامتر، رگرسیون لجستیک باینری، شاخص ارتباطی، تعداد تطابق، درجه عضویت

* نویسنده مسؤول: Bashiri@shahed.ac.ir

۱- مقدمه

رگرسیون لجستیک یکی از تکنیک‌های کاربردی برای تحلیل داده‌های طبقه‌بندی شده است. بعنوان نمونه اگر نتیجه آزمایشی را بصورت برد/ باخت تعریف کنیم، در این حالت متغیر پاسخ دیگر پیوسته نبوده، بلکه بصورت طبقه‌بندی شده خواهد بود. یکی از اقسام رگرسیون لجستیک، مدل رگرسیون لجستیک باینری بوده که تعداد طبقه‌بندی‌های متغیر پاسخ در این مدل دوتاست. اگر این تعداد طبقه‌بندی‌ها بیش از دو تا باشد، آنگاه با توجه به جنس متغیر پاسخ (یعنی اسمی و ترتیبی) مدل‌های رگرسیون لجستیک اسمی و ترتیبی حاصل می‌شود. مثال ارائه شده در این پژوهش به صورت یک آزمایشی درشش تیمار (وضعیت) با سه متغیر کنترلی بوده و متغیر پاسخ نیز به صورت صفر و یک بوده (جدول ۳)، لذا مدل مورد استفاده، مدل رگرسیون لجستیک باینری خواهد بود.

همانگونه که از عنوان این پژوهش مشخص است، مسأله مورد نظر در سیستم‌های غیرخطی مانند سیستم باینری بررسی شده که با استفاده از تخمین پارامتر برای مدل رگرسیون لجستیک باینری، می‌خواهیم شاخص‌های ارتباطی که نشان‌دهنده وضعیت سیستم است را بهبود دهیم، اما در عمل مقادیر پارامترها برای یک معادله رگرسیونی نامشخص است[۱]. برای ایجاد این معادله، باید ضرایب متغیرها یا همان پارامترهای مسأله را بطور تخمینی به‌دست آورد. روش‌های مختلفی برای تخمین پارامتر وجود دارد؛ یکی از روش‌های رایج، روش حداقل مربعات خطا[1]ست که در این روش هدف، یافتن مقادیر ضرایب رگرسیونی به منظور حداقل کردن میانگین مربعات خطاست. روش دیگر مورد استفاده، روش حداکثر درست نمایی[2] بوده، که با هدف ماکزیمم‌سازی درست‌نمایی مدل به تخمین پارامترهای مدل می‌پردازد. در واقع، روش تخمین حداقل مربعات خطا، حالت خاصی از روش حداکثر درست‌نمایی است زمانی که خطای مشاهده شده و داده‌ها از توزیع نرمال یا گوسین پیروی نمایند. زیبندگی به‌دست آمده که با یافتن مقادیر پارامتر مدل، بهترین تناسب برای داده را ایجاد می‌کند، تخمین پارامتر و همچنین گاهی بهینه‌سازی نیز گفته می‌شود [۲]. با توجه به آنکه مدل بکار رفته در این مسأله، به صورت غیرخطی است، لذا مقادیر پارامتر با استفاده از روش لگاریتم حداکثر درست‌نمایی تخمین زده می‌شود؛ دلیل استفاده از لگاریتم بخاطر خطی کردن تابع درست‌نمایی و آسان شدن عملیات ریاضی برای رسیدن به رابطه مورد نظر است که گام‌های آن برای مدل رگرسیون لجستیک باینری در این مقاله ارائه شده است. همچنین علاوه بر تعیین مقادیر درست-نمایی، مقادیر شاخص‌های اندازه‌گیری بنحو جدیدی محاسبه می‌شود. ساختار این مقاله بدین صورت است که در بخش بعد به مطالعاتی که در زمینه تخمین پارامتر شورت گرفته اشاره شده است. در بخش سوم چگونگی و گامهای تخمین پارامتر در مدل رگرسیون لجستیک باینری بیان می‌شود. در

1 Least Square Error
2 Maximum Likelihood

بخش چهارم و پنجم به تعریف متغیرهای مسأله و تبیین شاخص‌های مورد استفاده برای تحلیل مدل و روش پیشنهادی تحقیق می‌پردازیم. در ادامه مثالی برای توضیح و روشن شدن روش پیشنهادی ارائه خواهد شد و نهایتاً در بخش هفتم به تحلیل نتایج آزمایش خواهیم پرداخت.

۲- مروری بر مطالعات پیشین

اهمیت ضرایب متغیرها در مدل رگرسیون که نشان‌دهنده میزان تاثیر هر متغیر کنترلی بر متغیر پاسخ است، موجب شده تا روش‌هایی برای تعیین یا تخمین آنها ایجاد و بهبود داده شود. در این بخش مطالعاتی که در زمینه تخمین پارامتر و کاربردهای رگرسیون لجستیک انجام شده، اشاره خواهد شد. در دو دهه اخیر تخمین پارامترهای غیرخطی به کمک کامپیوتر به همراه روش‌های آماری، روش استانداردی برای تحلیل داده‌های کاربردی را ایجاد نمودند(بارد و همکاران، ۱۹۷۴). روش بیان شده توسط بارد[1] (۱۹۷۴) برای محاسبه پارامترهای نامعین مدل براساس حداقل کردن تابع هدف و سپس تعیین پارامتر کافی نبوده و دقت تخمین‌ها باید براساس استنباط آماری و روش تخمین صورت گیرد[۲]. مایرز و همکاران[2] (۱۹۳۷) در کتاب الگوهای خطی تعمیم‌یافته روش‌های مختلف برآورد پارامتر را معرفی نمودند. در این کتاب، الگوهای مختلف برآورد پارامتر در سیستم‌های خطی و غیر خطی شامل رگرسیون لجستیک و پوآسن ارائه شده است[۱۵]. پویتون و همکاران[3] (۲۰۰۶) به بررسی تخمین پارامتر در مدل‌های پویا با زمان پیوسته با استفاده از تحلیل تفاضلی اصلی پرداختند. تحلیل تفاضلی اصلی یکی از تکنیک‌های تخمین پارامتر برای مدل‌های معادله دیفرانسیلی است که در آن توابع پایه برای داده‌های پویا مناسب هستند. مزیت این تحلیل کم بودن محاسبات نسبت به رگرسیون غیر خطی بوده و دارای این عیب است که پارامترهای تخمینی دارای دقت کمتری هستند [۱۷]. کاسکر و همکاران[4] (۲۰۱۰) بهبود تخمین پارامتر را در قلمرو زمانی بررسی نمودند. تخمین پارامتر می‌تواند، بطور مستقیم در قلمرو زمانی با ایزوله کردن نواحی که در آن خطای پیش‌بینی بعنوان خطای پارامترهای مدل پویا در نظر گرفته می‌شوند، انجام شود. بر پایه معادلات تک پارامتری خطای پیش‌بینی، پارامترهای مدل تکی خطا را می‌توان بصورت تکراری تخمین زد [۷]. داچین[5](۲۰۰۳) نحوه تخمین پارامتر را در فرایندهای شیمی و بیوشیمی ارائه نمود [۸]. وان یانگ و همکاران[6](۲۰۱۰) به معرفی روش مقیاسی ابتکاری برای تخمین پارامتر کارا پرداختند. در تخمین پارامتر، ترتیب بزرگی اختلاف بین مقادیر پارامتر برای یافتن بهینگی دشوار است، چون مسأله تخمین پارامتر با اختلاف می‌تواند بد حالت باشد که برای اجتناب از سختی باید فاکتور مقیاسی مناسب پارامتر برای

1 Bard
2 Myers et al.
3 Poyton et al.
4 Cusker et al.
5 Dochain
6 Yang et al.

تخمین پارامتر تعریف شود. از آنجایی که روش مناسبی برای تعیین مجموعه فاکتور مقیاسی وجود ندارد لذا در این مقاله روش ابتکاری جدیدی برای تعیین فاکتور مقیاسی که مسأله تخمین پارامتر را خوش حالت می‌کند ارائه شده است[۱۹]. لیندنر و همکاران[1](۲۰۰۶) طراحی آزمایشات را برای تخمین پارامتر بهینه فرایند جنبشی آنزیم براساس تحلیل ماتریس اطلاعات فیشر انجام دادند[۱۳]. چو و همکاران[2](۲۰۰۹) توسعه‌های اخیر در تخمین پارامتر و ساختارشناسی سیستم‌های بیوشیمی و زیست‌شناسی ارائه کردند[۴]. لو و همکاران[3](۲۰۰۸) الگوریتمی برای تخمین پارامترها در مدل‌های معادلات دیفرانسیل احتمالی ارائه نمودند. این الگوریتم بر پایه رابطه حداقل مربعات وزنی غیرخطی است که در آن تابع هدف براساس میانگین مقادیر متغیرهای پیش‌بینی است [۱۴]. یه و همکاران[4] (۲۰۰۹) به بررسی ارتباط بین تابع پروفایل با متغیرهای کنترلی و پاسخ با استفاده از رگرسیون لجستیک پرداختند. در این مقاله ابتدا با بکارگیری روش حداکثر درست نمایی، ضرایب رگرسیونی برای مدل رگرسیون لجستیک باینری تخمین زده شده و سپس با تشکیل مدل رگرسیونی به بررسی هدف مورد نظر می‌پردازد[۲۰]. پس از بیان تحقیقات انجام شده در زمینه تخمین پارامتر، لازم است انواع روش‌ها و کاربردهای رگرسیون لجستیک نیز اشاره شود. رگرسیون لجستیک تا اواسط دهه ۴۰ میلادی توسعه نیافته بود و تا دهه ۷۰ میلادی نیز کمتر استفاده می‌شد، اما هم اکنون کاربردهای فراوانی از آن دیده می‌شود[۱]. رگرسیون لجستیک یکی از تکنیک‌های کاربردی جهت تحلیل داده‌های طبقه‌بندی شده است. این تکنیک با توجه به نوع و تعداد طبقه-بندی‌های متغیر پاسخ به سه دسته باینری، اسمی و ترتیبی تقسیم می‌شود. مطالعات اولیه در رابطه با این تکنیک در راستای ارائه مدل‌ها و روابط آماری بوده در حالی‌که مطالعات بعدی شامل به‌کارگیری و استفاده از روابط در مسائل ذکر شده است.

والکر[5] و همکاران (۱۹۶۷) مدل لجیت تجمعی را را پیشنهاد دادند[۱۸] که بعدها توسط کولاژ[6] (۱۹۸۰) (۱۹۸۰) توسعه یافت و این مدل، به مدل نسبی احتمالات شهرت پیدا کرد[۶]. فینبرگ[7](۱۹۸۰) یک مدل جایگزین برای مدل نسبی احتمالات ارائه کرد که به مدل نسبت تجمعات نام گذاری شد[۹]. کاکس[8] و همکاران (۱۹۸۴) برای تاثیر میزان دز چهار چهار دارو بر روی سلامتی افراد از مدل رگرسیون لجستیک ترتیبی استفاده کردند[۵]. پیترسون[9] و همکاران (۱۹۹۰) مدل احتمالات جزئی نسبی را برای رگرسیون لجستیک ترتیبی ارائه دادند[۱۶]. این مدل به دو زیر مدل با محدودیت و بی محدودیت تقسیم می‌شود؛ مدل ارائه شده توسط پیترسون و همکاران، توانست محاسبه مقادیر احتمال را ساده‌تر کند. که

1 Lindner
2 Chou
3 Ló et al.
4 Yeh et al.

5 Walker et al.
6 Mc Cullagh
7 Feinberg et al.
8 Cox et al.
9 Peterson et al.

مدل مورد استفاده در محاسبه احتمالات جزیی در بخش تخمین پارامتر، همان مدل احتمالات جزیی نسبی پترسون است. یکی از کاربردهای رگرسیون لجستیک در امور بهداشت و درمان است در این زمینه کارهای فراوانی انجام شده که برخی از آنها عبارتست از، بریزر[1] و همکاران (۱۹۹۱) از رگرسیون رگرسیون لجستیک ترتیبی در پیش بینی تومور روده استفاده کرده‌اند [۳]. آنها برای این کار از ۴۶۱ بیمار استفاده کردند و اذعان داشتند که روش آنها بدرستی توانسته است تومور بیماران انتخابی را پیش‌بینی کند، اما آنها دلیلی برای عمومیت این روش بیان نکردند پس لزومی برای درست بودن این روش بر اساس متغیرهای کنترلی تعیین شده برای بیماران دیگر وجود ندارد. نپ[2] (۱۹۹۹) از نسبت‌های ترتیبی برای ارزیابی شدت بیماری با مقیاس طبقه‌بندی شده از قبیل ضعیف، متوسط و شدید استفاده کرد [۱۱]. بدین منظور هریک از متغیرهای پاسخ نامبرده شده را بترتیب با اعداد ۱ و۲ و ۳ نمایش می‌دهد. عدد ۳ بیانگر وضعیت بحرانی‌تر نسبت به عدد۲ بوده و عدد ۲ نیز نشان‌دهنده حالت بحرانی‌تر از وضعیت ۱ است. با این تعاریف از مدل رگرسیون لجستیک ترتیبی برای محاسبه احتمال وقوع هریک از پاسخ‌ها که نشان‌دهنده وضعیت و نوع بیماری هستند را محاسبه می‌کند. هوانگ[3] و همکاران (۲۰۰۹) روش جدیدی با با استفاده از مدلسازی آماری برای تعیین ارتباط بین ارزیابی بلوغ و پارامتر زمان توسط رگرسیون لجستیک ترتیبی ارائه کردند [۱۰]. از دیگر کاربردهای رگرسیون لجستیک در امور بازاریابی و فروش و صنعت است. لوسونل[4] و همکاران (۲۰۰۶)، تجزیه و تحلیلی برای رضایت مشتریان توسط رگرسیون لجستیک باینری و اسمی و ترتیبی انجام داده‌اند. آنها پس از به‌دست آوردن رابطه رگرسیون، توانستند عامل اثرگذار رضایت مشتری را بیابند[۱۲]. ژو[5] و همکاران (۲۰۰۸)،از رگرسیون لجستیک ترتیبی ترتیبی برای بررسی نقش احساسات مشتریان در طراحی محصول استفاده کردند، بدین‌صورت که ابتدا نیازمندی‌ها و احساسات مشتریان را در مورد محصولی خاص(مطالعه موردی :اتاق کامیون ولوو) بررسی کرده و سپس این نظرات را به طراح، برای برقراری ارتباط منطقی بین طراحی محصول و نیازهای مشتریان منتقل کردند [۲۱]. همچنین مهدی بشیری و همکاران(۱۳۸۹) در مقاله‌ای با "عنوان بهینه‌سازی متغیرهای پاسخ در مدل رگرسیون لجستیک ترتیبی با استفاده از روش‌های ابتکاری و فرا ابتکاری" به تعیین و بهبود متغیرهای کنترلی تاثیرگذار بر رضایت‌مندی دانشجویان از عملکرد سایت کامپیوتر بمنظور بهینه‌کردن متغیر پاسخ با استفاده از مدل رگرسیون لجستیک ترتیبی پرداختند[۱](منبع فارسی). جداول (۱) خلاصه‌ای از مطالعات انجام شده در زمینه تخمین پارامتر و کاربردهای رگرسیون لجستیک را بترتیب سال پژوهش نشان می‌دهد.

1Brazer et al.
2 Knapp
3Huang et al.
4 Lowsonl rt al.
5 Zhu et al.

جدول ۱- بررسی مطالعات پیشین در زمینه تخمین پارامتر

محققان	سال	مطالعات انجام شده و کاربردهای تخمین پارامتر
مایرز و همکاران	۱۹۳۷	ارائه الگوهای مختلف برآورد پارامتر در سیستم‌های خطی و غیر خطی
بارد و همکاران	۱۹۷۴	تخمین پارامترهای غیرخطی به کمک کامپیوتر به همراه روش‌های آماری
داچین	۲۰۰۳	ارائه روش تخمین پارامتر در فرایندهای شیمی و بیوشیمی
پویتون و همکاران	۲۰۰۶	بررسی تخمین پارامتر در مدل‌های پویا با زمان پیوسته با استفاده از تحلیل تفاضلی اصلی
لیندنر و همکاران	۲۰۰۶	انجام طراحی آزمایشات برای تخمین پارامتر بهینه فرایند جنبشی آنزیم
لو و همکاران	۲۰۰۸	ارائه الگوریتمی برای تخمین پارامترها در مدل‌های معادلات دیفرانسیل احتمالی
چو و همکاران	۲۰۰۹	ارائه توسعه‌های اخیر در تخمین پارامتر و ساختارشناسی سیستم‌های بیوشیمی و زیست‌شناسی
یه و همکاران	۲۰۰۹	ارائه روش تخمین پارامتر در سیستم‌های غیرخطی باینری بمنظور تعیین ارتباط بین تابع پروفایل و متغیرهای کنترلی در مدل رگرسیون لجستیک
کاسکر و همکاران	۲۰۱۰	بررسی بهبود تخمین پارامتر را در قلمرو زمانی
وان یانگ و همکاران	۲۰۱۰	معرفی روش مقیاسی ابتکاری برای تخمین پارامتر کارا

جدول ۲- بررسی مطالعات پیشین در زمینه رگرسیون لجستیک

محققان	سال	مطالعات انجام شده و کاربردهای رگرسیون لجستیک	بررسی شاخص‌های ارتباطی
والکر و همکاران	۱۹۶۷	ارائه مدل لجیت تجمعی	-
کولاژ	۱۹۸۰	ارائه مدل نسبی احتمالات	-
فینبرگ	۱۹۸۰	ارائه مدل نسبت تجمعات	-
کاکس و همکاران	۱۹۸۴	بررسی تاثیر میزان دز چهار دارو بر سلامتی افراد با استفاده از مدل‌های رگرسیون لجستیک ترتیبی	-
پیترسون و همکاران	۱۹۹۰	مدل احتمالات جزئی نسبی	-
بریزر و همکاران	۱۹۹۱	پیش بینی بروز تومور روده با به‌کارگیری رگرسیون لجستیک	-
نپ	۱۹۹۹	استفاده از نسبت‌های ترتیبی برای ارزیابی شدت بیماری با مقیاس طبقه‌بندی شده	-
لوسونل و همکاران	۲۰۰۶	معرفی انواع مدل‌های رگرسیون لجستیک و به‌کارگیری تکنیک-های رگرسیون لجستیک در تعیین میزان رضایت‌مندی مشتریان	-
ژو و همکاران	۲۰۰۸	به‌کارگیری رگرسیون لجستیک ترتیبی برای بررسی نقش احساسات مشتریان بر طراحی محصول	-
هوانگ و همکاران	۲۰۰۹	استفاده از مدلسازی آماری برای تعیین ارتباط بین ارزیابی بلوغ و پارامتر زمان توسط رگرسیون لجستیک ترتیبی	-
بشیری و همکاران	۱۳۸۹	بهینه‌سازی متغیر پاسخ در مدل رگرسیون لجستیک ترتیبی با استفاده از روش‌های ابتکاری و فراابتکاری	-
تحقیق حاضر	۱۳۸۹	ارائه شاخص ارتباطی رگرسیون لجستیک باینری با حساسیت بیشتر به تخمین پارامتر در حداکثرسازی لگاریتم درست‌نمایی	*

*: بررسی شاخص‌های ارتباطی در مقاله	- : عدم بررسی شاخص‌های ارتباطی در مقاله

۳- تخمین پارامتر در سیستم‌های غیرخطی باینری

همانگونه که قبلا نیز اشاره شد، روش‌های مختلفی مانند روش حداقل مربعات خطا و حداکثر درست‌نمایی برای تخمین ضرایب رگرسیونی در سیستم‌های خطی وجود دارد. اما بدلیل ماهیت طبقه-بندی متغیرهای پاسخ در مدل‌های رگرسیون لجستیک، استفاده مستقیم از این روش‌ها برای تخمین پارامتر در این مسائل امکان پذیر نمی‌باشد. بطور کلی می‌توان برای تبدیل مسائل غیرخطی به خطی از تبدیل لگاریتم استفاده نمود که به مدل حاصل، مدل لگاریتم طبیعی یا لجیت گفته می‌شود. روش مورد استفاده در این مقاله برای تخمین پارامتر، روش حداکثر درست نمایی است با این تفاوت که از تبدیل لگاریتم برای تحلیل ماهیت طبقه‌بندی متغیرها استفاده شده است که به این روش، روش لگاریتم حداکثر درست‌نمایی[1] گفته می‌شود. در این روش پارامترهای مدل لجستیک بگونه‌ای تخمین زده می-شوند که لگاریتم درست‌نمایی حداکثر شود. در ادامه نحوه تخمین ضرایب با استفاده از روش لگاریتم حداکثر درست نمایی بیان می‌شود.

۳-۱- مفروضات مسأله

فرض کنید n مجموعه آزمایش مستقل وجود دارد و در هر مجموعه بردار متغیرهای پیش بینی p بصورت $x_i=(x_{i1},x_{i2},...,x_{ip})$ است، طوریکه متغیر پاسخ بصورت z_i , $i=1,2,...,n$ تعریف می شود. z_i دارای توزیع برنولی با احتمال موفقیت π_i می باشد. احتمال π_i تابعی از x_i بوده و در مدل رگرسیون لجستیک توسط تابع ارتباطی $g(\pi_i)$ مشخص شده که مطابق رابطه(۱) است.

(۱)

$$g(\pi_i)=\log\frac{\pi_i}{1-\pi_i}=\beta_1x_{i1}+\beta_2x_{i2}+\cdots+\beta_px_{ip}$$

بردار $\beta=(\beta_1,\beta_2,\cdots,\beta_p)^T$ بردار پارامتر مدل است. توجه کنید که $x_{i1}=1$ بطوریکه β_1 مقدار ثابت مدل است. با توجه به این تعاریف مقدار احتمال برای هر مجموعه از مشاهدات طبق رابطه (۲) تعیین می شود.

(۲)

$$\pi_i=\frac{\exp(x_i^T\beta)}{1+\exp(x_i^T\beta)}=\frac{\exp(\eta_i)}{1+\exp(\eta_i)}$$

جایی‌که $\eta_i=x_i^T\beta=\sum_{j=1}^{p}\beta_jx_{ij}$. فرض کنید که داده ها گروهی بوده طوریکه برای مجموعه iام متغیرهای کنترلی، m_i مشاهده وجود دارد. $M=\sum_{i=1}^{n}m_i$ بیانگر تعداد کل مشاهدات است. اگر z_{ij} بیانگر jامین مشاهده در iامین مجموعه از متغیرهای کنترلی باشد، آنگاه $y_i=\sum_{j=1}^{m_i}z_{ij}$ برابر مجموع مشاهدات (یا مجموع تعدادی از آزمایشات مستقل برنولی) انجام گرفته در تیمار iام طرح آزمایش بوده و لذا دارای توزیع دو جمله‌ای (m_i, π_i) با میانگین $m_i\pi_i$ و واریانس $m_i\pi_i(1-\pi_i)=m_i\times\frac{\exp(\eta_i)}{1+\exp(\eta_i)}\times\frac{1}{1+\exp(\eta_i)}$ است.

1 Log- likelihood

فرض کنید داده های گروهی مستقل هستند، لذا تابع درست نمایی توام $y_1, y_2, \cdots, y_n$ بصورت رابطه(۳) تعریف می‌شود.

(۳)
$$L(\Pi; y) = \prod_{i=1}^{n} \binom{m_i}{y_i} \pi_i^{y_i} (1-\pi_i)^{m_i - y_i}$$

طوری که $\Pi = (\pi_1, \pi_2, \cdots, \pi_n)^T$ و $y = (y_1, y_2, \cdots, y_n)^T$ است. از رابطه(۳) لگاریتم گرفته و با درنظر گرفتن $\eta_i = x_i^T \beta = \sum_{j=1}^{p} \beta_j x_{ij} = \log \frac{\pi_i}{1-\pi_i}$، می‌توان تابع لگاریتم درست نمایی را طبق رابطه (۴) تعریف نمود.

$$l(\beta; y) = \sum_{i=1}^{n} \log \binom{m_i}{y_i} + \sum_{i=1}^{n} \sum_{j=1}^{p} y_i \beta_j x_{ij} - \sum_{i=1}^{n} m_i \log\left[1 + \exp\left(\sum_{j=1}^{p} \beta_j x_{ij}\right)\right] \quad (۴)$$

با مشتق گرفتن از رابطه(۴) نسبت به β رابطه(۵) برقرار است.

$$\frac{\partial l(\beta; y)}{\partial \beta} = X^T y - \sum_{i=1}^{n} m_i \pi_i X_i = X^T (y - \mu) \quad (۵)$$

جایی که $\mu = (\mu_1, \mu_2, \cdots, \mu_n)^T$ و $E(y) = (m_1\pi_1, m_2\pi_2, ..., m_n\pi_n)$ یک ماتریس n*p است. طبق رابطه(۵) $X^T(y-\mu) = 0$ بوده که $0 = (0, 0, \cdots, 0)^T$ بردار صفر p بعدی است. در عمل MLE را می توان با تکرار تخمینهای روش حداقل مربعات وزنی نیز تقریب زد. قرار دهید $\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2, \cdots, \hat{\beta}_p)^T$ بیانگر تخمین‌های β بطوریکه $\hat{\beta} = (\hat{\eta}_1, \hat{\eta}_2, \cdots, \hat{\eta}_n)^T = X\hat{\beta}$ و $\hat{\pi}_i = \frac{\exp(\hat{\eta}_i)}{1+\exp(\hat{\eta}_i)}$.

ماتریس وزن $\hat{W} = diag\{m_1\pi_1(1-\pi_1), m_2\pi_2(1-\pi_2), ..., m_n\pi_n(1-\pi_n)\}$ یک ماتریس قطری n*n که دارای قطر اصلی بیان شده و سایر عناصر صفر است. q متغیر وابسته تعدیل شده بصورت $q = (q_1, q_2, \cdots, q_n)^T$ است جایی که $q_i = \hat{\eta}_i + (y_i - m_i\hat{\pi}_i)/[m_i\hat{\pi}_i(1-\hat{\pi}_i)]$. لذا مطابق تعاریف بیان شده رابطه (۶) حاصل می شود.

$$q = \hat{\eta} + \hat{W}^{-1}(y - \hat{\mu}) = X\hat{\beta} + \hat{W}^{-1}(y - \hat{\mu}) \quad (۶)$$

که $\hat{\mu} = (m_1\hat{\pi}_1, m_2\hat{\pi}_2, ..., m_n\hat{\pi}_n)$. هر دو سمت رابطه (۶) را در $X^T\hat{W}$ ضرب کرده لذا رابطه(۷) به‌دست می آید.

$$X^T\hat{W}q = X^T\hat{W}X\hat{\beta} + X^T\hat{W}\,\hat{W}^{-1}(y - \hat{\mu}) = X^T\hat{W}X\hat{\beta} + X^T(y - \hat{\mu}) \quad (۷)$$

طبق فرض $X^T(y-\mu)=0$ معادله (۷) را می توان بصورت رابطه تخمین β تبدیل نمود که معادله (۸) برای تخمین پارامتر بکار گرفته می شود.

(۸)

$$\hat{\beta}=\left(X^T\hat{W}X\right)^{-1}X^T\hat{W}q$$

که مقادیر β پس از چند تکرار و با توجه به شرط توقف مطابق گام‌های زیر تخمین زده می شود[۲۰].

۳-۲- گامهای تخمین پارامتر

۳-۲-۱- تخمین اولیه‌ای از β را به‌دست آورده و $\hat{\beta}^{(0)}$ بنامید. توجه کنید که این مقدار اولیه را می توان بر اساس تخمین حداقل مربعات معمولی به‌دست آورد؛ یعنی

(۹)

$$\hat{\beta}^{(0)}=\left(X^TX\right)^{-1}X^Ty$$

قرار دهید i=0.

۳-۲-۲- بر اساس $\hat{\beta}^{(i)}$، مقادیر $\hat{\eta}^{(i)}$، $\hat{\pi}^{(i)}$ و $\hat{W}^{(i)}$ را محاسبه نمایید.

۳-۲-۳- $\hat{q}^{(i)}=\hat{\eta}^{(i)}+\left(\hat{W}^{(i)}\right)^{-1}(\mathrm{y}-\hat{\mu}^{(\mathrm{i})})$ را به‌دست آورید.

۳-۲-۴- تخمین β را با استفاده از رابطه $\hat{\beta}^{(i+1)}=\left(X^T\hat{W}^{(i)}X\right)^{-1}X^T\hat{W}^{(i)}\hat{q}^{(i)}$ به هنگام کرده و قرار دهید i=i+1.

۳-۲-۵- گام ۲ تا ۴ را تکرار کنید. تا زمانیکه $\frac{\left\|\hat{\beta}^{(i)}-\hat{\beta}^{(i-1)}\right\|}{\left\|\hat{\beta}^{(i-1)}\right\|}\le\alpha$ شود. $\|v\|$ نرم اقلیدسی بردار v و α مقدار ثابت کوچک(مثلا $\alpha=10^{-5}$) است، آنگاه $\hat{\beta}=\hat{\beta}^{(i)}$ تخمینی مطلوب برای β است[۲۰].

٤- متغیرهای مسأله و تبیین شاخص‌های موجود

همان‌گونه که پیشتر نیز بیان شد، هدف از این مقاله تخمین مقادیر پارامترهای مدل رگرسیون لجستیک باینری با استفاده از روش یه و همکاران [۲۰] به منظور بهبود شاخص‌های ارتباطی است. آنچه که نقش تخمین پارامتر را در این تحقیق نمایان می‌سازد، روش پیشنهادی محققین است که به‌طور کامل در انتهای این بخش و بخش پنجم بیان خواهد شد. اگر بخواهیم به‌طور خلاصه نقش تخمین پارامتر را بیان کنیم، می‌توان به حساس نمودن شاخص‌های ارتباطی به روند تغییر و بهبود پارامترهای مدل در روش پیشنهادی نسبت به روش‌های موجود پیشین اشاره نمود. پس از آشنایی با نحوه تخمین پارامتر، حال لازم است متغیرهای مورد نیاز جهت محاسبه شاخص‌های ارتباطی معرفی شوند.

برای تحلیل یک آزمایش می‌توان از مقیاسهای اندازه‌گیری مختلفی استفاده نمود. این مقیاسها به چند دسته مقیاسهای ارتباطی[1]، خطایاب[2] و آماره زیبندگی[3] است. هریک از این مقیاسها دارای ویژگی‌هایی هستند که تناسب یا عدم تناسب مدل را نشان می‌دهند. بطور نمونه، یکی از شاخص‌های موجود در مقیاس ارتباطی، شاخص تطابق[4] است. این شاخص به مقایسه احتمال وقوع پاسخ ۱ (یا موفقیت) در مقابل پاسخ صفر (یا شکست) پرداخته و درصورتیکه احتمال وقوع موفقیت از شکست بیشتر باشد تطابق صورت گرفته است. از آنجایی که هدف اصلی این

1 Association Measures
2 Diagnostic Measuers
3 Goodness-of-fit statistics
4 Concordant

مقاله، تخمین پارامتر برای بررسی شاخص ارتباطی در مدل است، لذا باید به معرفی متغیرها و تمام شاخص‌های موجود در این مقیاس و نحوه بکارگیری آنها پرداخت. برای درک بهتر مفهوم هریک از متغیرها و نوع مسأله مورد نظر، لازم است مقدماتی ارائه شود. یک طرح آزمایش با تعداد تیمارهای مشخص و تعدادی مشاهدات مستقل (متغیرهای پاسخ) را در نظر بگیرید. هر متغیر پاسخ به‌صورت باینری برای هر تیمار تعریف می‌شود. پس از تعریف متغیرهای پاسخ، می‌توان زوج مرتب‌هایی تنها شامل عناصر یک و صفر را که متشکل از عناصر یک پاسخ با پاسخ‌های دیگر است، تعیین نمود. هریک از عناصر زوج مرتب، دارای احتمال وقوع بوده که نحوه محاسبه آن در بخش سوم اشاره گردید. با مقایسه مقادیر احتمالات هر دو عضو از یک زوج مرتب، متغیرهای مسأله تعریف شده که در ادامه به‌طور کامل بیان خواهد شد.

٤-١- جفت[1]

مقیاس ارتباطی جهت اندازه‌گیری وضعیت مدل، نیاز به زوج مرتبهای شامل عناصر صفر و یک دارد. تعداد جفتهای موجود در یک آزمایش، برابر تعداد آزمایشها با پاسخ یک ضرب در تعداد آزمایشهایی با پاسخ صفر است.

٤-٢- شاخص تطابق (nc)

این شاخص میزان توان مدل را برای پیش‌بینی نشان می‌دهد. هرچه مقدار این شاخص بیشتر باشد، توانایی مدل برای پیش‌بینی وقوع موفقیت بیشتر خواهد شد. نحوه محاسبه آن در بالا بیان شده است.

٤-٣- شاخص عدم تطابق[2] (nd)

در صورتی که در یک زوج مرتب شامل عناصر صفر و یک، احتمال وقوع عنصر صفر(شکست) بیشتر از عنصر یک باشد، آنگاه عدم تطابق صورت می‌گیرد. هرچه مقدار این شاخص کمتر باشد، مدل از توانایی بیشتری برای پیش‌بینی برخوردار می‌شود.

٤-٤- گره[3] (nt)

اگر در زوج مرتبه‌ای موجود، احتمال عنصر یک با عنصر صفر برابر باشد، گره صورت می‌گیرد.

٤-٥- شاخص تطابق احتمالی (pc)

مجموع اختلاف مقادیر احتمالی جفت‌هایی که عنصر یک با احتمال بیشتری از عنصر صفر اتفاق می‌افتد. به عبارت دیگر در جفتهایی با احتمال وقوع بیشتر عنصر یک، اختلاف مقادیر احتمالات محاسبه شده و با یکدیگر جمع می‌شوند.

٤-٦- شاخص عدم تطابق احتمالی (pd)

مجموع اختلاف مقادیر احتمالی جفتهایی که عنصر صفر با احتمال بیشتری از عنصر یک اتفاق می‌افتد. در واقع نحوه محاسبه این شاخص، عکس حالت قبلی خواهد بود.

همچنین، با استفاده از این شاخص‌ها، می‌توان شاخص‌های دیگری نیز بنام شاخص‌های ارتباطی، به‌دست آورد. این شاخص‌ها عبارتست از؛

1 Pair

2 Discordant Measures
3 Tied

Goodman- Kukal Gamma ،Somers'D و Kendall's Tau-a است. نحوه محاسبه هریک از این سه شاخص مطابق روابط (۱۰) و (۱۱) و (۱۲) است.

$$Somers'D = \frac{nc - nd}{nc + nd + nt} \quad (۱۰)$$

$$Goodman - KukalGamma = \frac{nc - nd}{nc + nd} \quad (۱۱)$$

$$Kendall'sTau - a = \frac{nc - nd}{0.5 \times N \times (N-1)} \quad (۱۲)$$

هدف از این تحقیق محاسبه این سه شاخص با روابط بیان شده نیست، چراکه این شاخص‌ها پیشتر نیز محاسبه شده‌اند. کاری که در این مقاله انجام شده، ارائه روش جدید برای محاسبه این سه شاخص بوده که در بخش بعدی بیان خواهد شد. سوالی که در اینجا مطرح می‌شود آنست که نقش تخمین پارامتر در تعیین این شاخص‌ها چیست؟ پاسخ این سوال در بخش روش تخمین پارامتر آمده است؛ پس از تخمین ضرایب در هر مرحله، مقادیر احتمالات پاسخ محاسبه می‌شود که مقادیر احتمالات به‌دست آمده در تعیین مقادیر تطابق، عدم تطابق و گره جدید مورد استفاده قرار می‌گیرند. جهت آشنایی بیشتر با نحوه محاسبه شاخص‌های ارتباطی، به ذکر یک مثال عددی برای شاخص Somers'D می‌پردازیم.

فرض کنید در یک طرح آزمایش با دو متغیر پاسخ مستقل و سه تیمار، تعداد سه پاسخ یک و سه پاسخ صفر مشاهده شود. جدول (۳) مقادیر متغیرهای پاسخ و کنترلی، احتمال وقوع پاسخ و جفت‌های تشکیل شده که مقادیر احتمالات هر جفت در زیر آن آورده شده را نشان می‌دهد.

جدول ۳- مقادیر احتمالات و جفت‌های تشکیل شده برای طرح آزمایش مورد نظر

X1	X2	Y	Probability	Pairs
2	4	1	0.50000	(1,0),(1,0),(1,0) (0.5,0.0),(0.5,0.5),(0.5,0.0)
1	2	1	1.00000	(1,0),(1,0),(1,0) (1.0,0.0),(1.0,0.5),(1.0,0.0)
3	2	0	0.00000	(1,0),(1,0),(1,0) (1.0,0.0),(1.0,0.5),(1.0,0.0)
2	4	0	0.50000	
1	2	1	1.00000	
3	2	0	0.00000	

این آزمایش دارای سه تیمار با دو تکرار[1] است که برای نشان دادن مقادیر احتمال وقوع مشاهدات هر تیمار، نتایج دو تکرار در زیر هم در جدول(۳) قرار گرفته است. پس از تخمین نهایی پارامترهای مدل با استفاده از الگوریتم بخش سوم، مقادیر احتمالات

1 Replicate

فردی با استفاده از رابطه (۲) به راحتی محاسبه شده است که این مقادیر را می‌توانید در ستون چهارم جدول (۳) مشاهد کنید. پس از این مرحله لازم است تمامی زوج مرتب‌های متشکل از عناصر صفر و یک تشکیل گردد که تعداد آن برابر حاصل‌ضرب تعداد صفرها در تعداد یک‌های موجود در مشاهدات است. به عبارت دیگر هریک از مشاهدات "یک" در هر تکرار می‌تواند با هر مشاهده "صفر" موجود در همان تکرار یا تکرارهای دیگر تشکیل یک زوج مرتب را دهد. زوج‌های تشکیل شده در ستون Pairs جدول (۳) قابل مشاهده می‌باشد. که دو ردیف اول مربوط به زوج‌های مشاهدات "یک" در تکرار اول و ردیف سوم نیز مربوط به جفتهای مشاهده "یک" در تکرار دوم است. مقادیری که در زیر هریک از جفتها آورده شده نیز همان احتمالات فردی وقوع هر مشاهده از تیمار مورد نظر می‌باشد، که به عنوان نمونه برای اولین زوج، احتمالات ۰/۵ و صفر نوشته شده که در واقع احتمالات مرتبط با زوج مرتب‌های تیمار اول و سوم است.

از ستون سمت راست جدول (۲) تعداد جفت‌ها برابر ۹، تعداد تطابق برابر ۸، تعداد عدم تطابق برابر صفر و تعداد گره برابر یک به‌دست می‌آید. به عبارت دیگر، nc=8 ، nd=0 و nt=1 . و لذا شاخص Somers'D بصورت زیر محاسبه می‌شود.

$$\text{Somers'D} = (8-0)/(8+0+1) = 0.89$$

که با توجه به این شاخص، می‌توان گفت که مدل از تناسب خوبی برخوردار است.

لازم به ذکر است که شاخص‌های بیان شده (روابط ۱۰ و ۱۱ و ۱۲) به‌عنوان روابطی است که پیشتر معرفی و مورد استفاده قرار گرفته است. یکی از مهم‌ترین معایب موجود این شاخص‌ها را می‌توان به عدم حساسیت آنها به تغییر و بهبود پارامترهای تخمینی در هر تکرار تا رسیدن به نقطه توقف اشاره نمود. به عبارت دیگر، با وجود آنکه مقدار پارامترهای مدل رگرسیون لجستیک در هر تکرار در حال بهبود است، اما با توجه به آنکه معیار اندازه‌گیری شاخص‌های ارتباطی برحسب تعداد تطابق است، لذا تعداد تطابق در هر تکرار ثابت مانده و لذا مقدار این شاخص‌ها نیز بدون تغییر باقی می‌ماند. برای برقراری حساسیت این شاخص‌ها به تغییر یا بهبود مقادیر پارامترهای تخمینی، روشی جدید ارائه شده که در بخش بعدی به آن اشاره خواهد شد.

۵-تعریف مسأله و روش پیشنهادی

گام‌های تخمین پارامتر بگونه‌ای است که در هر تکرار از الگوریتم پس از تخمین ضرایب، مقادیر احتمالات برای هر پاسخ محاسبه می‌شود. در اینجا لازم است به بخش سوم برگردیم. هر تیمار از آزمایش می‌تواند m بار تکرار شود که پاسخ هرتکرار می‌تواند یک یا صفر باشد. تمام زوج مرتبهای(جفتهای) ممکن را تشکیل داده و سپس مقادیر احتمال منطبق با هر عنصر از یک جفت را به‌دست می‌آوریم(مقادیر احتمال در هر تکرار محاسبه می‌شود). در اینجا دیگر بجای مقایسه احتمال هر جفت و تعیین مقادیر تطابق، عدم تطابق و گره؛ اختلاف مقادیر احتمالات بین دو عنصر محاسبه می‌شود. یعنی بجای استفاده از تعداد تطابق در رابطه، از مجموع مقادیر اختلاف احتمالاتی که احتمال وقوع عنصر یک بیشتر از احتمال وقوع عنصر صفر است. همچنین بجای محاسبه تعداد عدم تطابق، از مجموع

اختلاف مقادیر احتمالی که احتمال وقوع عنصر صفر بیشتر از عنصر یک باشد،استفاده می‌شود. بعبارت دیگر میزان درجه عضویت احتمال پاسخ به پاسخهای صفر و یک محاسبه شده است. اما تعداد گره‌ها همانند روش اصلی خود محاسبه می‌شود. با این توضیحات نحوه محاسبه شاخص‌های ارتباطی جدید پیشنهادی مطابق با روابط (۱۳) و (۱۴) و (۱۵) خواهد بود.

$$Somers'D\ p = \frac{pc - pd}{pc + pd + nt} \quad (۱۳)$$

$$Goodman - KukalGamma\ p = \frac{pc - pd}{pc + pd} \quad (۱۴)$$

$$Kendall'sTau - a\ p = \frac{pc - pd}{0.5 \times N \times (N-1)} \quad (۱۵)$$

همان‌گونه که از روابط (۱۳) تا (۱۵) مشخص است، محاسبه شاخص‌های ارتباطی جدید برخلاف روش‌های پیشین که براساس تعداد تطابق بوده‌اند، با استفاده از مقادیر احتمال تطابق در هر تکرار برای مدل مورد نظر می‌باشد. مزیت این روش نسبت به روش‌های سابق در حساس نمودن شاخص‌های ارتباطی به تغییرات پارامترهای مدل در هر تکرار تا رسیدن به هدف مورد نظر با توجه به احتمال وقوع تطابق است، در حالی‌که در روش‌های پیشین، با توجه به آنکه تعداد تطابق نسبت به تغییرات پارامترهای مدل از خود حساسیت نشان نداده، لذا مقادیر شاخص‌های ارتباطی نیز ثابت باقی می‌ماند. به عبارت دیگر در روش پیشنهادی، ابتدا در هر تکرار مقادیر پارامتر مدل تخمین زده شده، با استفاده از پارامترهای تخمینی، احتمالات تطابق و عدم تطابق محاسبه شده و از آنجایی که میزان تغییر احتمالات، حساسیت زیادی به تغییر پارامتر داشته و شاخص‌های ارتباطی جدید نیز ارتباط مستقیم با احتمالات محاسبه شده دارد، لذا شاخص‌های ارتباطی جدید در هر تکرار با بهبود پارامتر مدل، بهبود می‌یابد. برای درک بهتر روش بیان شده، همان مقادیر آزمایش جدول (۳) را در نظر بگیرید. با استفاده از نرم‌افزار Matlab مقادیر شاخص‌های اولیه در طی ۳۴ مرحله تخمین پارامتر مقدار ثابت ۰/۸۹ بوده در حالیکه مقدار شاخص‌های پیشنهادی از تکرار اول تا آخرین تکرار در حال افزایش است. جدول (۴) مقایسه‌ای بین مقادیر شاخص‌های اولیه و پیشنهادی را برای ۱۰ تکرار آخر نشان می‌دهد.

جدول ٤- مقایسه شاخص‌های ارتباطی در دو حالت اولیه و پیشنهادی برای ده تکرار آخر

Step	1	2	3	4	5	6	7	8	9	10
Somrer'sD	0.889	0.889	0.889	0.889	0.889	0.889	0.889	0.889	0.889	0.889
Somrer'sDp	0.8545	0.8546	0.8547	0.8548	0.8549	0.8550	0.8551	0.8552	0.8553	0.8554

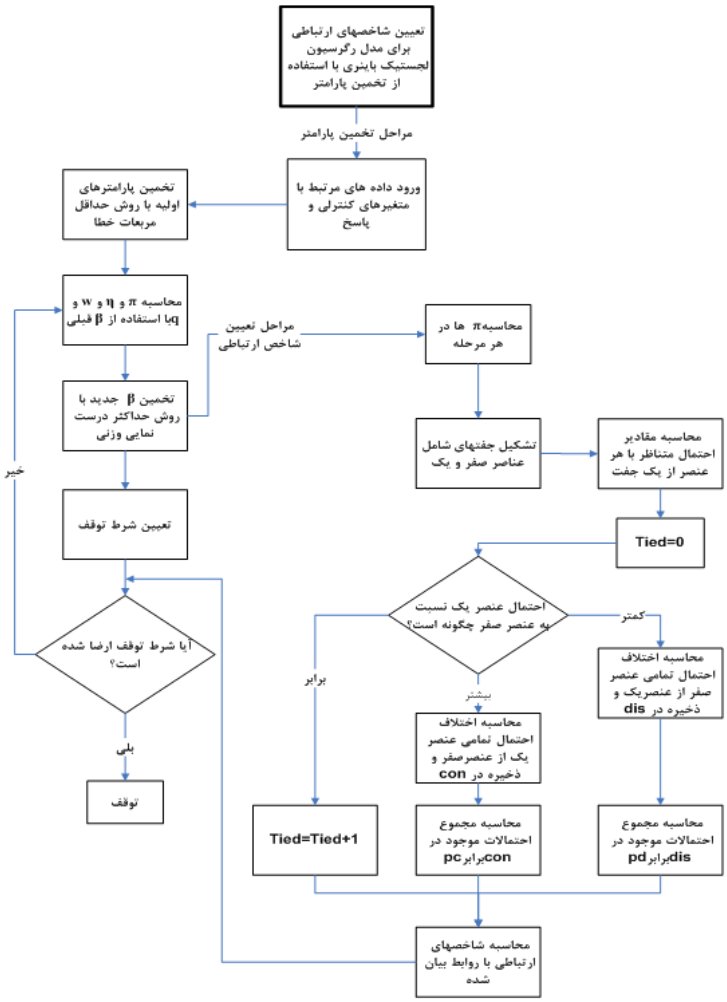
لازم به ذکر است که در بخش بعدی، تمامی مراحل ذکر شده در ضمن یک مثال شبیه‌سازی شده به طور کامل توضیح داده خواهد شد؛ اما برای آشنایی بیشتر با نحوه محاسبه شاخص پیشنهادی مقدار این شاخص در آخرین تکرار بصورت زیر محاسبه شده است.

Somers'Dp= (5.92-0)/ (5.92+0+1) = 0.8554

باید توجه داشت که مقادیر احتمالات ارائه شده در جدول(۳) بصورت رند شده بوده، اما مقدار شاخص پیشنهادی براساس مقادیر رند نشده محاسبه گردید. به عبارت دیگر مقدار ۵/۹۲ مجموع اختلاف احتمالات ۶ زوج مرتبی است که احتمال وقوع پاسخ یک بیشتر از پاسخ صفر است. همان‌گونه که از جدول (۴) مشخص شده است، مقادیر شاخص‌های ارتباطی پیشنهادی برخلاف شاخص‌های اولیه در ده تکرار آخر، در حال بهبود است. این بدان معناست که شاخص‌های پیشنهادی دارای حساسیت زیادی نسبت به تغییرات مقادیر پارامترهای تخمینی در هر تکرار تا رسیدن به نقطه بهینه است. در حالی‌که شاخص‌های اولیه با وجود تغییرات در مقدار پارامتر و مقدار احتمال وقوع پاسخ هر تیمار، هیچ حساسیتی از خود نشان نمی‌دهد و این نشان‌دهنده ضعف این شاخص‌هاست، چراکه می‌بایست با بهبود مقادیر پارامترهای تخمینی، شاخص‌های ارتباطی مدل(برتری پاسخ موفقیت نسبت به شکست) نیز بهبود یابد.

شکل(۱) گام‌های تعیین شاخص‌های ارتباطی را نشان می‌دهد. این فلوچارت از دو بخش تخمین پارامتر (در سمت چپ نمودار) و تعیین شاخص‌های ارتباطی (در سمت راست نمودار) در روش پیشنهادی تشکیل شده است. بخش تخمین پارامتر از مطالعات پیشین یه و همکاران [۲۰] و بخش تعیین شاخص‌های ارتباطی به عنوان روش پیشنهادی محققین یا جنبه نوآوری تحقیق به حساب می‌آید.

مجله علمی – پژوهشی مدیریت تولید و عملیات
سال دوم، شماره دوم، بهار و تابستان ۱۳۹۰
تاریخ وصول: ۶ / ۱۱/ ۸۹
تاریخ پذیرش: ۲۴/ ۳/ ۹۰
صص: ۲۲-۱



شکل ۱- فلوچارت روش پیشنهادی برای تعیین شاخص‌های ارتباطی

جدول ۵- مقادیر متغیرهای کنترلی و پاسخ در هر تکرار

متغیر پاسخ	متغیرهای پاسخ باینری در هر تکرار	متغیرهای کنترلی

* نویسنده مسؤول: Bashiri@shahed.ac.ir

x1	x2	x3	x4	z1	z2	z3	z4	y
1	1	0	1	1	0	0	1	2
1	3	2	2	0	0	1	0	1
1	0	0	2	1	1	1	0	3
1	3	2	1	0	0	0	0	0
1	2	1	1	1	0	0	0	1
1	2	3	1	1	0	1	0	2

مقادیر پارامترهای تخمینی و احتمالات پاسخ برای هر تیمار با استفاده از نرم‌افزار MATLAB پس از ۷ تکرار با توجه به شرط توقف حاصل شده است. این مقادیر در جدول(۶) و (۷) آمده است.

جدول ٦- مقادیر پارامترهای تخمینی

1	-1.4922	1.2004	0.5288	0.53570	0.53571	0.53574	0.53576
2	-0.1184	-1.6575	-1.3013	-1.3347	-1.33510	-1.33513	-1.33514
3	-0.0128	0.7363	0.4882	0.5067	0.50680	0.50681	0.50683
4	0.1953	0.3813	0.4470	0.4794	0.48010	0.4802	0.4805

جدول ۷- مقادیر احتمالات هر تیمار

1	0.8465	0.1954	0.4810	0.4192	0.4208	0.4208	0.4208
2	0.6858	0.1851	0.1770	0.1815	0.1830	0.1830	0.1830
3	0.9616	0.2494	0.8768	0.8058	0.8168	0.8170	0.8170
4	0.5905	0.1574	0.1281	0.1242	0.1218	0.1217	0.1217
5	0.7382	0.1756	0.2695	0.2424	0.2409	0.2409	0.2409
6	0.8688	0.1719	0.6167	0.4593	0.4665	0.4665	0.4665

همانطور که بیان شد، هدف از تخمین پارامتر در این مسأله حداکثر کردن تابع لگاریتم درست نمایی بوده، که مقادیر جدول (۸) گویای این مطلب است.

جدول ۸- مقادیر تابع لگاریتم درست‌نمایی

Step	1	2	3	4	5	6	7
Log-Likelihood	-17.0421	-12.9516	-12.6713	-12.6690	-12.6690	-12.6690	-12.6690

همانگونه که از جدول(۸) مشخص است، مقادیر پارامترها بگونه‌ای تخمین زده شدند که تابع لگاریتم درست‌نمایی در هر تکرار مقدار بهتری را به‌دست آورد.

در این مرحله پس از تشکیل زوج‌های صفر و یک و تعیین احتمال متناظر با هریک از عناصر، مقادیر شاخص‌های تطابق، عدم تطابق و گره در ۷ تکرار بصورت جدول (۹) حاصل شده است.

جدول ۹- مقادیر شاخص‌های تطابقی

Step	1	2	3	4	5	6	7
concordant	97	89	97	97	97	97	97
discordant	21	29	21	21	21	21	21
Tied	17	17	17	17	17	17	17

با توجه به جدول(۹) مقادیر شاخص‌های ارتباطی طبق جدول (۱۰) به‌دست می‌آید.

جدول ۱۰- مقادیر شاخص‌های ارتباطی

Step	1	2	3	4	5	6	7
Somrer'sD	0.5630	0.4444	0.5630	0.5630	0.5630	0.5630	0.5630
Goodman	0.6441	0.5085	0.6441	0.6441	0.6441	0.6441	0.6441
Kendalls	0.2574	0.2174	0.2574	0.2574	0.2574	0.2574	0.2574

همانگونه که از جدول(۱۰) مشخص شده، با استفاده از روابط (۱۱) تا (۱۳) مقادیر شاخص‌های ارتباطی بهبود نیافته‌اند. حال مقادیر مجموع شاخص‌های تطابقی احتمالی با استفاده از روش بیان شده(اختلاف احتمال بین عناصر صفر و یک یا میان درجه عضویت به پاسخ موفقیت) محاسبه شده که در جدول(۱۱) بیان شده است.

جدول ۱۱- مقادیر مجموع شاخص‌های تطابقی احتمالی

Step	1	2	3	4	5	6	7
SumConcordantp	19.1818	4.0789	40.8160	34.8063	35.4906	35.5023	35.5029
SumDiscordantp	2.3264	0.6849	6.0444	4.8151	4.9121	4.9132	4.9135
Tied	17	17	17	17	17	17	17

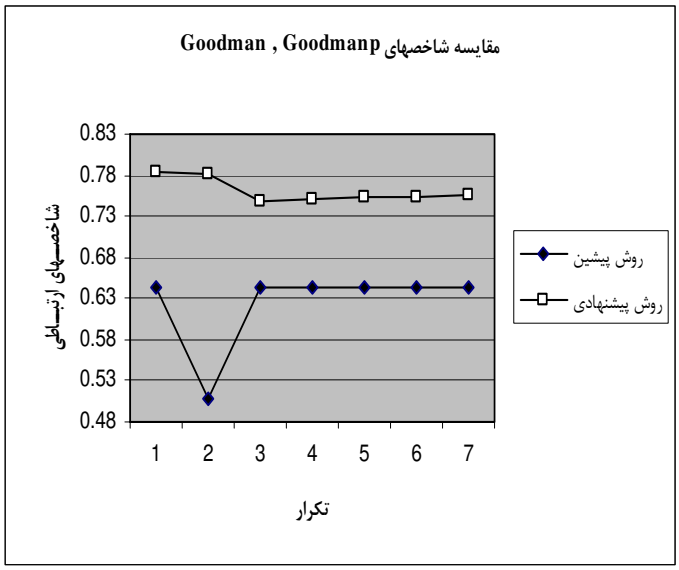
مقادیر دو سطر اول بدینصورت به‌دست آمده که پس از تعیین اختلاف احتمال متناظر با هریک از زوجهای صفر و یک، مجموع این اختلافات برای هر تکرار محاسبه شده‌اند. با استفاده از مقادیر جدول(۱۱)، مقادیر شاخص‌های ارتباطی احتمالی جدید که طبق روابط (۱۴) تا (۱۶) به‌دست آمده اند که در جدول (۱۲) نشان داده شده است.

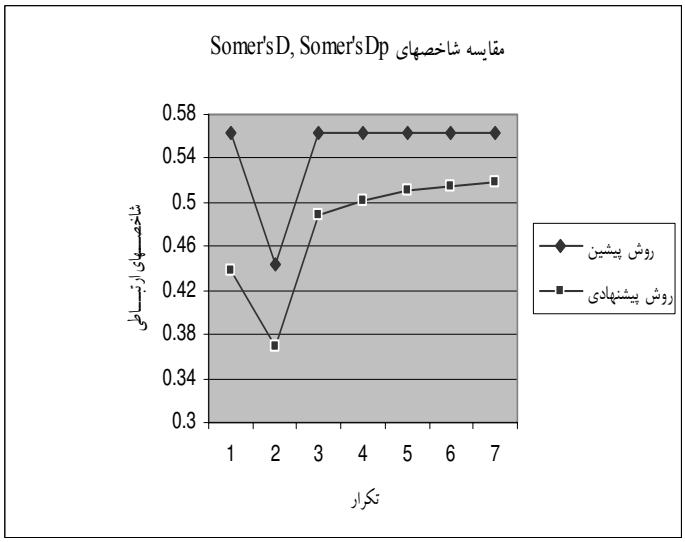
جدول ۱۲- مقادیر شاخص‌های ارتباطی احتمالی

Step	1	2	3	4	5	6	7
Somrer'sDp	0.4377	0.3695	0.4878	0.5023	0.5103	0.5150	0.5180
Goodmanp	0.7837	0.7803	0.7490	0.7519	0.7533	0.7542	0.7548
Kendallsp	0.0611	0.0123	0.1260	0.1087	0.1108	0.1109	0.1110

مقادیر جدول(۱۱) نشان می‌دهد که روش ارائه شده، روش مناسبی جهت افزایش شاخص‌های ارتباطی با هدف حداکثرسازی تابع لگاریتم درست‌نمایی است. بعبارت دیگر می‌توان با استفاده از مفهوم مقدار درجه عضویت به پاسخ موفقیت به‌جای استفاده از تعداد تطابق (یا تعداد وقوع موفقیت با احتمال بیشتر نسبت به شکست) مقادیر شاخص‌های ارتباطی را به‌دست آورد. شکل(۲) مقایسه‌ای از شاخص‌های ارتباطی Somer'sD و Goodman موجود در جداول (۹) و (۱۱) را بطور نمونه نشان می‌دهد.

مجله علمی – پژوهشی مدیریت تولید و عملیات
سال دوم، شماره دوم، بهار و تابستان ۱۳۹۰
تاریخ وصول: ۶ / ۱۱/ ۸۹
تاریخ پذیرش: ۲۴/ ۳/ ۹۰
صص: ۲۲-۱

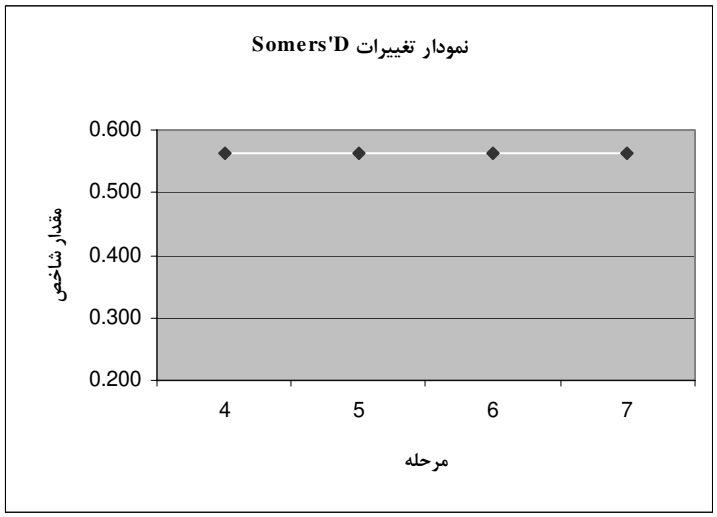


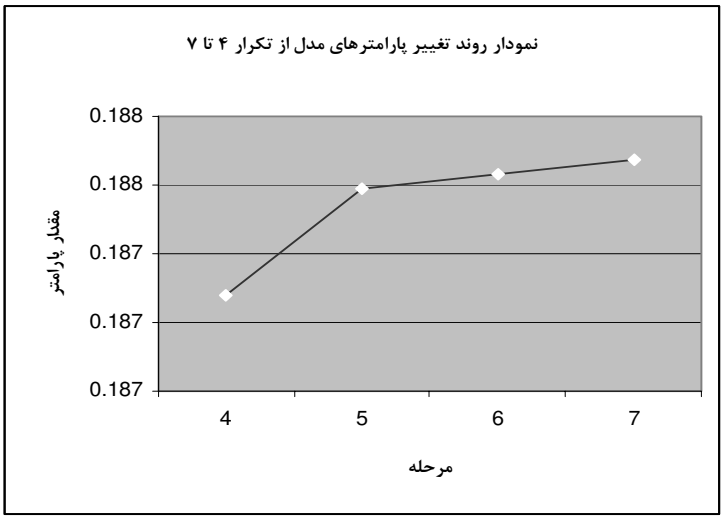


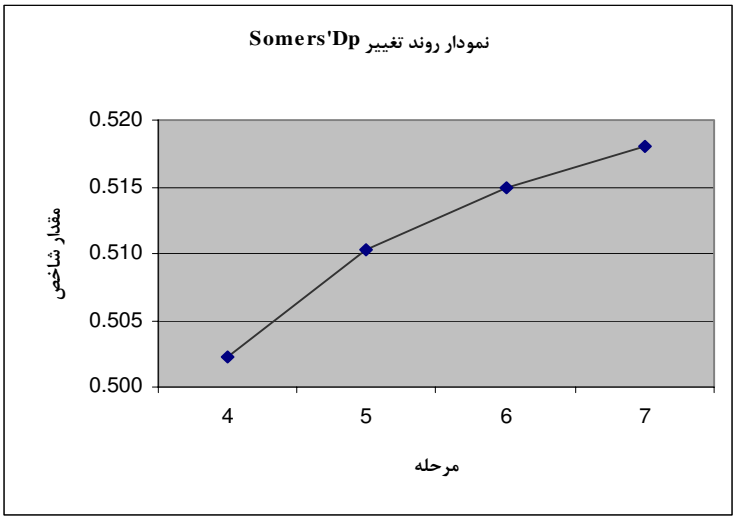
* نویسنده مسؤول: Bashiri@shahed.ac.ir

شکل ۲- مقایسه شاخص‌های ارتباطی در دو روش پیشنهادی و پیشین

همان‌گونه که از شکل(۲) پیداست، برخلاف شاخص‌های اولیه(جدول ۹)، شاخص‌های تعیین شده با روش پیشنهادی (استفاده از درجه عضویت) از تکرار سوم به بعد افزایش می‌یابند و این افزایش به معنای بهبود وضعیت سیستم در هر تکرار است بعبارت دیگر با بهبود پارامترهای مدل به‌منظور افزایش تابع لگاریتم درست‌نمایی، شاخص‌های ارتباطی نیز افزایش می‌یابند. لازم به ذکر است که دلیل افت شدید شاخص‌ها از تکرار اول به تکرار دوم، مقدار اولیه پارامترهاست ($\beta^{(0)}$) که ابتدا با روش حداقل مربعات خطا تخمین زده شده و معیار تخمین ($\beta^{(1)}$) است، اما در تکرارهای بعدی با روش حداکثر درست‌نمایی مقادیر پارامترها تخمین زده می‌شود؛ که این تغییر الگوریتم(روش) موجب نوسان شدید اولیه شده است. شکل(۳) یک مقایسه عددی را بین روش پیشنهادی با روش‌های ارائه شده پیشین انجام داده و بیانگر مزیت روش پیشنهادی نسبت به روش‌های سابق است. این مقایسه تطبیقی از مرحله ۳ تا مرحله ۷ با توجه به تغییرات پارامترهای مدل انجام شده است.







شکل ۳- نمودارهای تطبیقی دو روش پیشنهادی و روش پیشین با در نظر گرفتن روند تغییرات پارامترهای مدل

همان‌گونه که از شکل (۳) مشخص است، روند تغییرات Somer'sD هیچ حساسیتی نسبت به تغییرات پارامتر نداشته و این روند کاملا ثابت است، در حالیکه روند تغییرات Somer'sDp با توجه به روند صعودی تغییر پارامتر، صعودی است. لذا با استناد براین نمودار، می‌توان بهتر بودن روش پیشنهادی را نسبت به روش پیشین اثبات نمود.

۷- نتیجه‌گیری و جمع‌بندی

آنچه که در این تحقیق بررسی شد، استفاده از تابع لگاریتم درست‌نمایی برای تخمین پارامترهای مدل رگرسیون لجستیک باینری و بهبود شاخص‌های ارتباطی بوده است. شاخص‌های ارتباطی نشان‌دهنده وضعیت سیستم از نظر تعداد و احتمال پیشامد موفقیت نسبت به پیشامد شکست است که هرچه مقدار آن بیشتر باشد، سیستم از وضعیت بهتری برخوردار است. استفاده از روابط تعریف شده قبلی نتوانسته(مفهوم " تعداد ") موجب بهبود شاخص‌های ارتباطی در تکرارهای مورد نیاز تا رسیدن به نقطه همگرایی پارامترهای مدل رگرسیون لجستیک باینری شود که این مورد از جداول (۹) و (۱۰) کاملا مشخص شده است. اما با استفاده از مفهوم درجه عضویت و استفاده از مجموع اختلاف احتمالات پیشامدهای موفقیت و شکست در هر تکرار می‌توان شاخص‌های تطابق و ارتباطی را بهبود داده و با این مفهوم احتمال وقوع موفقیت را حداکثر نمود. به عبارت دیگر در تخمین پارامترها، شاخص مناسب بودن تخمین، همان شاخص‌های ارتباطی است. اما این شاخص‌ها از مرحله‌ای به بعد کارایی خود را از دست می‌دهند، لذا با استفاده از شاخص ارتباطی احتمالی به دنبال افزایش کارایی تخمین پارامترها بودیم که این موارد در شکل (۳) کاملاً مشخص است. پیشنهاداتی که می‌توان برای پژوهش‌های آتی ارائه نمود، شامل بکارگیری توام تخمین پارامتر و بهبود متغیرهای کنترلی به منظور روش ارائه شده در مقالات پیشین و رویکرد پیشنهادی در این پژوهش را با یکدیگر تلفیق نموده و در حین تخمین پارامتر جهت ماکزیمم‌سازی لگاریتم درست‌نمایی،

شاخص‌های ارتباطی نیز به عنوان تابع هدف حـداکثر گردند.

منابع

[۱]. بشیری، مهدی.، کامران‌راد، رضا.، کریمی، حسین. (۱۳۸۹). بهینه‌سازی متغیرهای پاسخ در مدل رگرسیون لجستیک ترتیبی با استفاده از روش‌های ابتکاری و فراابتکاری. مجله علمی و پژوهشی شریف، پذیرفته شده و در حال چاپ.

[1]. Agresti, A. (2007). An Introduction to Categorical Data Analysis. Department of Statistics University of FloridaGainesville, Florida. Second Edition.

[2]. Bard. (2010). Nonlinear Parameter Estimation. Developments in Petroleum Science, Vol 57, pp 197-301.

[3]. Brazer, S. R., Pancotto, F. S., Long, T. L., and Harrell, F. E. (1991). "Using ordinal logistic regression to estimate the likelihood of colorectal neoplasia", Journal of Clinical Epidemiology , Vol. 44, pp. 1263-1270.

[4]. Chou, I., Voit, E.O.(2009). Recent developments in parameter estimation and structure identification of biochemical and genomic systems. Mathematical Biosciences. Vol 219. pp 57–83.

[5]. Cox,C and Chuang, C.A. (1984) "Comparison of Chi-square patitioning and two logit analysis of ordinal pain data from a pharmecutical study", Statical Medicine, Vol. 3, pp. 273-285.

[6]. Cullagh, M. C. (1980). "Regression model for ordinal data (with discussion)" , J R Statis Soc Series B.

[7]. Cusker,R.mc., Currier, T., Danai,K.(2010). Improved parameter estimation by noise compensation in the time-scaledomain. Signal Processing . Vol 91, pp 72-84.

[8]. Dochain, D. (2003). State and parameter estimation in chemical and biochemical processes: a tutorial. Journal of Process Control. Vol 13, pp 801–818.

[9]. Feinberg, B. (1980). "Analysis of cross-classified data", Cambridge : MIT Press.

[10]. Huang, B., Biro, F. M., & Dorn, L. D. (2009)."Determination of Relative Timing of Pubertal Maturation through Ordinal Logistic Modeling: Evaluation of Growth and Timing Parameters", Journal of Adolescent Health , Vol .45, pp. 383-388.

[11]. Knapp, T. R. (1999). "Focus on quantitative methods: The analysis of the data for two-way contingency tables". Research in Nursing and Health, Vol. 22, pp. 263–268.

[12]. Lawsonl, C., and Montgomery, D., C. (2006) "logistic regression analysis of costumer satisfaction data", Quality and Reliability Engineering International, Vol. 22, pp. 971-984.

[13]. Lindner, P.F.O., Hitzmann, B. (2006). Experimental design for optimal parameter estimation of an enzyme kinetic process based on the analysis of the Fisher information matrix. Journal of Theoretical Biology. Vol 238, pp 111–123.

[14]. Ló, B.P., Haslam, A.J., Adjiman, C.S. (2008). An algorithm for the estimation of parameters in models with stochastic differential equations. Chemical Engineering Science. Vol63, pp 4820 – 4833.

[15]. Myers, R.H., Montgomery, D.C., Vening, G.G. (1937). Generalized linear Models with Applications in Engineering and The Sciences. Translated by: Niioumand, H.A. Ferdowsi University of Mashhad Publication. No 445.

[16]. Peterson, B. L., and Harrell, F., E.(1990). "Partial proportional odds models for ordinal response variable", Appl Stat, Vol. 39. pp. 205-217.

[17]. Poyton, A.A., Varziri, M.S., McAuley, K.B., McLellan, P.J., Ramsay, J.O. (2006). Parameter estimation in continuous-time dynamic models using principal differential analysis. Computers and Chemical Engineering.Vol.30, pp 698–708.

[18]. Walker, S.H., and Duncan, D., B. (1967). "Estimation of the probability of an event as a function of several independent variables", Biometrica, Vol. 54, pp. 167-179.

[19]. Yang, K.W., Lee, T.Y. (2010). Heuristic scaling method for efficient parameter

estimation. chemical engineering research and design. Vol 88. pp 520–528.

[20]. Yeh, A.B., Huwang, L., Li, Y.M. (2009). Profile monitoring for a binary response. IIE Transactions . Vol 41, pp 931–941.

[21]. Zhou, F., Wu, D., Yang, X., and Jiao, J. (2008). "Ordinal logistic regression for affective product design", Proceedings of the IEEE IEEM, pp. 1986-1990.