مهندسی صنایع و مدیریت شریف، (زمستان ۱۳۹۲)
دوره‌ی ۱ - ۲۹، شماره‌ی ۴، ص. ۱۴۵-۱۵۷، (یادداشت فنی)

Research Note

# تخمین هم‌زمان سطوح چندپاسخی وابسته به‌هم برای داده‌های طبقه‌بندی شده

رضا کامران‌راد (دانشجوی کارشناس ارشد)
مهدی بشیری* (دانشیار)
گروه مهندسی صنایع، دانشگاه شاهد

هدف این نوشتار، تخمین رابطه‌ی بین متغیرهای کنترلی و متغیرهای پاسخ از نوع داده‌های طبقه‌بندی شده و دارای وابستگی با استفاده از یک روش ابتکاری است. در این نوشتار با استفاده از مدل لگاریتم خطی[1]، آزمایش‌هایی با بیش از یک متغیر پاسخ طبقه‌بندی شده تحلیل و مدل‌سازی شده است. برای تخمین پارامترهای مدل رگرسیون لجستیک برای پاسخ‌های وابسته (دو متغیر پاسخ وابسته)، از یک روش ابتکاری غیرخطی تکرارپذیر با هدف بیشینه‌کردن تعداد انطباق‌ها استفاده شده است. مقایسه‌ی نتایج حاصل از روش ابتکاری با نتایج به دست آمده در حالت استقلال متغیرها و یکی از روش‌های موجود ــ برای مثال‌های فرضی با داده‌های شبیه‌سازی شده و یک مطالعه‌ی موردی با اندازه‌های متفاوت ــ نشان می‌دهد که روش ابتکاری پیشنهادی در مقایسه با یکی از روش‌های موجود و روش تخمین جداگانه ضرایب متغیرهای پاسخ براساس شاخص میزان انطباق و شاخص حداکثر درست‌نمایی از عملکرد مناسب برخوردار است.

**واژگان کلیدی:** بهینه‌سازی سطوح چندپاسخی، مدل لگاریتم خطی، رگرسیون لجستیک، تخمین پارامتر، تعداد انطباق.

kamranrad@shahed.ac.ir
bashiri@shahed.ac.ir

## ۱. مقدمه

امروزه در بسیاری از مطالعات آماری و بازاریابی، پیش‌بینی وضعیت آینده با درنظرگرفتن شرایط موجود برای بهبود و توسعه‌ی فعالیت‌های حاضر و باقی‌ماندن در عرصه‌ی رقابت بسیار حیاتی است. باید توجه داشت که بسیاری از مسائل موجود در عالم واقع به‌صورت چندپاسخی و دارای ماهیت گسسته (طبقه‌بندی شده) است و نیز یک نوع وابستگی بین پاسخ‌ها وجود دارد. در این حالت هریک از پاسخ‌ها را می‌توان با در نظرگرفتن تعداد طبقه‌بندی و ماهیت طرح آزمایش به‌ترتیب باینری، اسمی و ترتیبی[2] تعریف کرد. برای تحلیل مسائلی با داده‌های طبقه‌بندی شده می‌توان از شیوه‌های مختلفی همچون رگرسیون لجستیک[3]، آزمون کای‌دو پیرسون، SN-Ratio تاگوچی و... بهره گرفت. اگر تعداد متغیر پاسخ طبقه‌بندی شده یکی باشد، از مدل رگرسیون لجستیک استفاده می‌شود درحالی که مدل لگاریتم خطی برای مسائلی با حداقل دو پاسخ طبقه‌بندی شده به کار می‌رود. همچنین هنگامی که داده‌های آزمایش ترکیبی از تعداد مختلفی پاسخ طبقه‌بندی شده باشد، از تبدیل لجستیک چندمتغیره[1] استفاده می‌شود. در جدول ۱ دسته‌بندی این شیوه‌ها برای مسائل مختلف با پاسخ‌های پیوسته و طبقه‌بندی شده نشان داده شده است. براین اساس، رگرسیون لجستیک یکی از شیوه‌های کاربردی برای بهینه‌سازی مسائلی با پاسخ طبقه‌بندی شده است. یکی از جدیدترین کاربردهای این شیوه، بهینه‌سازی طراحی نوع محصول است که در سال ۲۰۰۹ برای بهینه‌سازی طراحی و تولید شیر آشپزخانه کاربرد یافته است.[2] در ادامه کاربرد سایر روش‌های بیان شده در جدول ۱ مورد بحث قرار خواهد گرفت.

در این نوشتار پس از بررسی تحقیقات انجام‌شده در زمینه‌ی داده‌های طبقه‌بندی شده، مسائل تک‌پاسخی و چندپاسخی طبقه‌بندی می‌شود و مسائل چندپاسخی با داده‌های پیوسته نیز مورد بحث قرار می‌گیرد. سپس در بخش سوم مفاهیم و مدل لگاریتم خطی برای تحلیل مسائل چندپاسخی با داده‌های طبقه‌بندی شده معرفی می‌شود. در بخش چهارم روش ابتکاری پیشنهادی برای تخمین ضرایب مدل رگرسیون همزمان چند پاسخ با بیشترین شاخص انطباق ارائه می‌شود. به‌منظور بررسی عملکرد روش پیشنهادی، سه مثال عددی با داده‌های شبیه‌سازی شده و اندازه‌های متفاوت در بخش پنجم مورد تحلیل و مقایسه قرار خواهد گرفت. در بخش ششم برای نشان‌دادن درستی روش ابتکاری، نتایج حاصله با نتایج یکی از روش‌های تخمین موجود در ادبیات مقایسه می‌شود؛ و نهایتاً در بخش هفتم روش پیشنهادی برای تحلیل یک مطالعه‌ی واقعی مورد استفاده قرار گرفته است.

## ۲. مروری بر مطالعات پیشین

چنان که در بخش اول اشاره شد، در عالم واقع بسیاری از متغیرهای موجود ماهیتاً طبقه‌بندی شده‌اند و نمی‌توان آن‌ها را با استفاده از شیوه‌های تحلیل داده‌های پیوسته[4] مورد بررسی قرار داد. به‌همین دلیل روش‌های تحلیل داده‌های طبقه‌بندی شده معرفی

* نویسنده مسئول
*تاریخ: دریافت ۱۳۹۰/۶/۶، اصلاحیه ۱۳۹۰/۱۰/۱۳، پذیرش ۱۳۹۱/۲/۱۷.*

جدول ۱. دسته‌بندی انواع روش‌های تحلیل مسائل باتوجه به نوع متغیرهای پاسخ.

| چند پاسخه | | تک پاسخه | | نوع مسئله |
|---|---|---|---|---|
| پیوسته | طبقه‌بندی شده | پیوسته | طبقه‌بندی شده | |
| تحلیل واریانس چندمتغیره | کای‌دو پیرسون | تحلیل واریانس | کای‌دو پیرسون | بررسی اثرگذار بودن عامل کنترلی |
| مدل چندپاسخی درجه ۱ و ۲ | مدل لگاریتم خطی | رگرسیون خطی | رگرسیون لجستیک | پیش‌بینی و بهینه‌سازی |
| تابع زیان تاگوچی | | | تاگوچیSN-Ratio | بهینه‌سازی با هدف کاهش میزان انحراف از محدوده فنی (با در نظر گرفتن میانگین و ماتریس واریانس کوواریانس) |
| تابع مطلوبیت | | | | بهینه‌سازی با هدف تعیین رضایت‌مندی و صرفا نزدیک‌شدن به‌مقادیر هدف هریک از توابع |

و توسعه داده شد. جدول ۱ جایگاه تحقیق حاضر در میان سایر مطالعات مشابه را نشان می‌دهد. ابتدا تحقیقات انجام‌شده در زمینه‌ی مسائلی با پاسخ‌های پیوسته بیان خواهد شد.

## ۱.۲. بررسی مطالعات انجام‌شده در زمینه داده‌های پیوسته

در سال ۲۰۱۰ اثربخشی روش‌های پیوسته‌ی مختلف برای بهینه‌سازی مسائل چندپاسخی مستقل با استفاده از رگرسیون چندگانه بررسی شد.[۳] در این تحقیق پنج روش پیوسته برای بهینه‌سازی مسائل چندپاسخی ارائه شده و کارایی آن‌ها مورد مقایسه قرار گرفته است. در نهایت مشخص شد که شیوه‌رگرسیون چندگانه مبتنی بر روش نسبت وزنی S-N [۵] به‌عنوان یک روش پیوسته، دارای بیشترین اثربخشی در یافتن جواب بهینه مسائل چندپاسخی خواهد بود. محققین دیگر نیز کاربرد شیوه‌ی اندازه‌گیری آنتروپی براساس روش تاگوچی برای حل مسائل بهینه‌سازی چندپاسخی همبسته (مسئله‌ی جوشکاری نقطه‌یی) را ارائه کردند.[۴] آن‌ها در مقاله‌ی خود از شیوه‌ی‌تحلیل مؤلفه‌های اصلی[۶] برای حذف همبستگی موجود میان پاسخ‌ها و تبدیل آنها به پاسخ‌های ناهمبسته استفاده کردند. در نهایت، روش تاگوچی موجب استخراج یک فرایند بهینه برای تولید مطلوب شد.[۴] در سال ۲۰۰۸ شیوه‌ی‌داده‌کاوی مسائل چندپاسخی پویا در طراحی آزمایش تاگوچی مورد استفاده قرار گرفت.[۵] در این پژوهش رویکرد داده‌کاوی برای حل مسائل طراحی پارامتر پویا با پاسخ‌های چندگانه ارائه شد و از شبکه‌ی عصبی مصنوعی برای ایجاد مدل تابع پاسخ استفاده شد. همچنین تابع مطلوبیت برای ارزیابی روش‌های اندازه‌گیری مسائل چندپاسخی و شبیه‌سازی تبرید برای تعیین بهترین مجموعه فاکتورها از میان مدل تابع پاسخ ایجاد شده مورد استفاده قرارگرفت. یکی دیگر از ابزارهای تحلیل مسائل چندپاسخی پیوسته، تحلیل واریانس چندمتغیره است. تحلیل واریانس چندمتغیره در تعیین زیرمجموعه‌ی متغیرهایی با توان تشخیص بالا ــ که امروزه موارد استفاده‌ی زیادی دارد ــ به کار گرفته می‌شود. در فرایند انتخاب زیرمجموعه‌ی متغیرها دو کار انجام می‌شود: جست‌وجوی متغیرها توسط روش‌های جست‌وجو، و تخمین تابع ارزیابی. روش‌های جست‌وجو به سه دسته تقسیم می‌شوند: ۱. جست‌وجوی کامل؛ ۲. جست‌وجوی ابتکاری؛ ۳. جست‌وجوی تصادفی. محققین در مطالعات خود توابع ارزیابی را به پنج گروه تقسیم کرده‌اند: فاصله‌ی متغیرها، محتوای اطلاعات، وابستگی، سازگاری و سطح نرخ خطا.[۶] همچنین با ترکیب تحلیل واریانس چندمتغیره و طراحی آزمایش‌ها یک فرایند دومرحله‌یی برای انتخاب متغیرها و وجوه تأثیرگذار ارائه کرده‌اند.[۷] محققین دیگری نیز در مقاله‌یی با عنوان بهینه‌سازی پارامترهای فرایند با استفاده از تابع زیان تاگوچی وزنی، نسبت به تعیین مجموعه پارامترهای بهینه از میان مهم‌ترین فاکتورها برای تعیین درآمد محصول اقدام کردند. آن‌ها با در نظرگرفتن متغیرهای کنترلی و پاسخ چندگانه مدل تاگوچی را برای فرایندهای کلی‌تر تعمیم دادند که به «تابع زیان وزنی» معروف است.[۸]

## ۲.۲. مسائل تک‌پاسخی با داده‌های طبقه‌بندی‌شده

سابقه‌ی استفاده از روش‌های تحلیل داده‌های طبقه‌بندی شده به بیش از یکصد سال می‌رسد، جایی که افرادی همچون کارل پیرسون[۷]، جرج یول[۸]، فیشر[۹] و... (۱۹۰۰-۱۹۵۱) با معرفی توزیع کای‌دو برای جداول توافقی[۱۰]، نسبت احتمالات تا مدل‌های لجیت برای انواع داده‌های طبقه‌بندی شده به‌نام رگرسیون لجستیک، گام بلندی در زمینه‌ی داده‌های طبقه‌بندی شده برداشتند.[۹] ارائه‌ی مدل لجیت تجمعی توسط والکر و دانکن[۱۱] (۱۹۶۷)[۱۰] و نسبت احتمالات توسط کولاژ[۱۲] (۱۹۸۰)[۱۱] از مهم‌ترین فعالیت‌های انجام‌شده در زمینه‌ی داده‌های طبقه‌بندی شده‌ی تک‌پاسخی است. مطابق جدول ۱، رگرسیون لجستیک و SN-Ratio تاگوچی از ابزارهای تحلیل و بهینه‌سازی مسائل تک‌متغیره با پاسخ طبقه‌بندی شده‌اند که کاربردهای فراوانی در زمینه‌ی بهداشت و درمان، بازاریابی، طراحی محصول و... دارد. در ادامه برخی از کاربردهای این روش‌ها معرفی می‌شود. در بررسی تطبیقی روش‌های بهینه‌سازی داده‌های طبقه‌بندی شده، پژوهش‌گران با مقایسه‌ی سه روش تحلیل تجمعی تاگوچی[۱۳]، طرح امتیازبندی نیر[۱۴]، و طرح امتیازبندی احتمالی وزنی جنگ و گو[۱۵]، به بیان مزایا و معایب هریک از این سه روش پرداختند.[۱۲] برای بررسی اثر ذهنیت مصرف‌کنندگان بر طراحی محصول با استفاده از رگرسیون لجستیک، ابتدا نیازمندی‌های و افکار مشتریان درمورد محصولی خاص بررسی شد و سپس این نظرات برای برقراری ارتباط منطقی بین طراحی محصول و نیازهای مشتریان به طراح منتقل گردید.[۱۳] محققین از مدل رگرسیون لجستیک ترتیبی

برای تعیین نوع محصول بهینه بهره گرفتند.[۱۴] همچنین از این مدل برای تعیین ارتباط بین زمان بلوغ افراد و متغیرهای کنترلی ــ نظیر تغذیه، جنسیت، محیط زندگی و... ــ استفاده کردند.[۱۵] از طرف دیگر، در مقاله‌یی با عنوان «بررسی پروفایل برای پاسخ باینری»، نحوه‌ی تخمین پارامترهای مدل رگرسیون لجستیک باینری برای مسائل تک‌پاسخی ارائه شد.[۱۶] در مطالعه‌یی که برای بهینه‌سازی متغیر پاسخ در مدل رگرسیون لجستیک ترتیبی با استفاده از روش‌های ابتکاری و فرااابتکاری انجام شد، با استفاده از تعیین بهترین سطح متغیرهای کنترلی، بهترین احتمال وقوع برای پاسخی با بالاترین رتبه‌ی تخصیصی برآورد شد.[۱۷] در ادامه، پژوهش‌های انجام شده در زمینه‌ی مدل‌های چندپاسخی طبقه‌بندی شده معرفی خواهد شد.

## ۳.۲. بررسی مطالعات انجام‌شده برای مسائل چندپاسخی طبقه‌بندی شده

گودمن[۱۶] مدل‌های ارتباطی را با عنوان نسبت متقاطع محلی[۱۷] برای جداول I*J بعدی ارائه کرد؛ این مدل با نام مدل سطر ـ ستون[۱۸] شناخته می‌شود.[۱۸] بعد از گودمن، محققین دیگری نسبت به معرفی یک مدل حاشیه‌یی برای داده‌های طبقه‌بندی شده با عنوان «لجیت تجمعی حاشیه‌یی و نسبت متقاطع فراگیر[۱۹]» اقدام کردند که با استفاده از این مدل‌ها می‌توان مقادیر احتمالات توأم را برای پاسخ‌های وابسته محاسبه کرد.[۱۹،۲۰] پس از آن، با استفاده از مدل‌های محلی گودمن و فراگیر دیل، یک مدل ارتباطی برای متغیرهای پاسخ ترتیبی ارائه شد.[۲۱] در سال ۱۹۹۴ نیز یک مدل لجستیک چندمتغیره برای دو و سه متغیر پاسخ باینری ارائه شد، و در سال ۱۹۹۵ نیز این مدل برای چند متغیر پاسخ گسسته‌ی اسمی و ترتیبی توسعه داده شد.[۲۲] در سال ۲۰۰۲، با به کارگیری مدل لگاریتم خطی برای جداول توافقی با دو، سه و چهار متغیر پاسخ طبقه‌بندی شده، نحوه‌ی محاسبه‌ی احتمالات توأم برای پاسخ‌های وابسته نشان داده شد.[۲۳] پس از آن در سال ۲۰۰۴، یک رویکرد نظری برای متغیرهای تصادفی طبقه‌بندی شده‌ی چندمتغیره با الگوهای همبستگی معرفی شد.[۲۴] در این تحقیق پژوهش‌گران نحوه‌ی محاسبه‌ی همبستگی میان پاسخ‌های وابسته و مقادیر احتمالات توأم را ارائه دادند.

در پژوهش حاضر یک روش ابتکاری برای تخمین همزمان پارامترهای مدل لگاریتم خطی، با هدف تحلیل مسئله‌ی چندپاسخی با پاسخ‌های وابسته است. یکی از خلاءهای تحقیقاتی موجود در زمینه‌ی مسائل چندپاسخی طبقه‌بندی شده، داده‌های وابسته است. در جدول ۲ پژوهش‌های انجام شده در زمینه‌ی انواع مسائل با داده‌های پیوسته و طبقه‌بندی شده ارائه شده است. در بخش بعد به معرفی مدل‌های لگاریتم خطی می‌پردازیم.

# ۳. مدل لگاریتم خطی

چنان‌که در بخش‌های پیشین اشاره شد، برای تحلیل مسائلی با حداقل دو متغیر پاسخ طبقه‌بندی شده، می‌توان از مدل لگاریتم خطی استفاده کرد. اگر تعداد متغیرهای پاسخ طرح آزمایش دوتا باشد، آنگاه می‌توان دو حالت زیر را در نظر گرفت.

## ۱.۳. مدل مستقل[۲۰]

فرض کنید آزمایش مورد نظر دارای دو متغیر پاسخ با $i$ و $j$ طبقه‌بندی باشد. آنگاه اگر دو متغیر پاسخ از یکدیگر مستقل باشند مدل کلی لگاریتم خطی به‌صورت رابطه‌ی ۱ خواهد بود:

$$\log \mu_{ij} = \lambda + \lambda_i^X + \lambda_j^Y \qquad (۱)$$

برای بیان دقیق مفهوم $\lambda, \lambda_i^X, \lambda_j^Y$ ارائه‌ی توضیحاتی در رابطه با جدول توافقی در مدل لگاریتم خطی ضرورت می‌یابد. جدول توافقی جدولی است که تعداد مشاهدات در هر طبقه از متغیرهای پاسخ را نشان می‌دهد. با استفاده از این مشاهدات می‌توان مقادیر احتمالات هر سلول را محاسبه کرد. احتمال سلول، در واقع همان احتمال وقوع مشاهدات در هر طبقه از یک متغیر پاسخ است. علاوه برآن می‌توان مقادیر احتمال هر سطر یا ستون جدول احتمالی را محاسبه کرد. به احتمال سطر $i$ و ستون $j$، اثرات سطر $i$ام و ستون $j$ام گفته می‌شود. لذا مقدار مشاهده انتظاری برای هر سلول از جدول احتمالی، در حالتی که متغیرها از یکدیگر مستقل‌اند، به‌صورت رابطه‌ی ۲ محاسبه می‌شود:

$$\mu_{ij} = n\,\Pi_{i+}\Pi_{+j} \qquad (۲)$$

که در آن $n$ بیانگر تعداد کل مشاهدات، $\Pi_{i+}$ اثر سطر $i$ام، و $\Pi_{+j}$ اثر سطر $j$ام است. نحوه‌ی محاسبه‌ی اثرات (احتمال) هر سطر یا ستون، تقسیم تعداد مشاهدات هر سطر یا ستون بر تعداد کل مشاهدات است. حال اگر از طرفین رابطه‌ی ۲ لگاریتم گرفته شود، آنگاه $\log \mu_{ij}$ وابسته به تعداد مشاهدات، احتمال سطر $i$ و ستون $j$ خواهد بود. در نتیجه رابطه‌ی ۲ را می‌توان به‌صورت رابطه‌ی ۱ برای اثر سطری $\lambda_i^X$ و اثر ستونی $\lambda_j^Y$ نوشت، طوری که $\lambda_i^X$ بیانگر اثر طبقه‌بندی در سطر $i$ام باشد؛ هرچه مقدار این پارامتر بیشتر باشد، مقدار مشاهده‌ی انتظاری در آن سطر بزرگ‌تر خواهد بود. به‌طور مشابه، این موضوع برای $\lambda_j^Y$ نیز صادق است.[۲۱] نکته‌ی قابل توجه آن که در مدل لگاریتم خطی از مقادیر امید انتظاری به‌جای مقادیر احتمالات برای تحلیل مدل استفاده می‌شود، چرا که در محاسبه‌ی مقادیر احتمالات توأم از امید انتظاری استفاده می‌شود. حال اگر متغیرهای پاسخ به هم وابسته باشند از روش دیگری که در ادامه بیان می‌شود باید استفاده کرد.[۲۱]

## ۲.۳. مدل اشباع‌شده[۲۱]

درصورتی که وابستگی معناداری بین دو متغیر پاسخ وجود داشته باشد، مدل لگاریتم خطی به‌صورت رابطه‌ی ۳ خواهد بود:

$$\log \mu_{ij} = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_{ij}^{XY} \qquad (۳)$$

$\{\lambda_{ij}^{XY}\}$ مقدار ارتباط بین دو متغیر است و بیانگر انحراف از استقلال متغیرهاست.

یکی از موارد مهم در بهینه‌سازی مسائل چندپاسخی، تعیین مقادیر احتمالات توأم برای وقوع هم‌زمان پاسخ‌های طبقه‌بندی شده است. به‌همین منظور و برای تعیین احتمالات توأم در هر یک از تیمارهای طرح آزمایش می‌توان از روابط ۴ و ۵ استفاده کرد. اگر متغیرهای پاسخ $X$ و $Y$ دارای $a$ و $b$ طبقه‌بندی باشد، آنگاه مقادیر احتمالات توأم در حالت مدل اشباع شده عبارت است از:

$$\Pi_{ij} = \frac{\mu_{ij}}{\sum\sum \mu_{ab}} \qquad (۴)$$

مفهوم رابطه فوق آن است که مقدار احتمال توأم دو متغیر پاسخ که در طبقه‌بندی $i$ و $j$ قرار دارند، به‌صورت تقسیم امید انتظاری همان طبقه از یک تیمار به امید انتظاری تمامی طبقه‌بندی‌های همان تیمار از طرح آزمایش محاسبه می‌شود، لذا احتمالات توأم در هر تیمار از طرح آزمایش دارای محدودیت‌های $\Pi_{ij} \geq 0$ و $\sum_i \sum_j \Pi_{ij} = 1$ خواهد بود.[۲۱]

جدول ۲. مروری بر مطالعات انجام شده در زمینه داده‌های طبقه‌بندی شده.

| نوع مسئله | محققین | سال پژوهش | مدل ارائه شده | کاربرد |
|---|---|---|---|---|
| داده پیوسته | [۳] | ۲۰۱۰ | ارائه‌ی مدل رگرسیون چندگانه مبتنی بر روش تاگوچی | بهینه‌سازی مسائل چندپاسخی پیوسته |
| | [۴] | ۲۰۰۹ | - | استفاده از شیوه‌ی تحلیل مولفه‌های اصلی برای حذف همبستگی میان پاسخ‌ها |
| | [۵] | ۲۰۰۸ | ارائه شیوه‌ی داده‌کاوی برای مسائل چندپاسخی پویا (داده‌ی پیوسته) | تعیین مدل تابع پاسخ و تعیین بهترین مجموعه فاکتورها |
| | [۶] | ۱۹۹۷ | تعیین توابع ارزیابی | انتخاب بهترین مجموعه متغیرها |
| | [۷] | ۲۰۰۴ | - | ترکیب تحلیل واریانس چند متغیره و طراحی آزمایشات یک فرایند دو مرحله‌یی برای انتخاب متغیرها |
| | [۸] | ۲۰۰۹ | تابع زیان تاگوچی وزنی | تعیین مجموعه پارامترهای بهینه |
| تک پاسخی با داده‌های طبقه‌بندی شده | [۱۲] | ۲۰۰۶ | - | مطالعه‌ی تطبیقی بر روش‌های بهینه‌سازی داده‌های طبقه‌بندی شده |
| | [۱۳] | ۲۰۰۸ | - | بررسی اثر ذهنیت مصرف کنندگان در طراحی محصول |
| | [۲] | ۲۰۰۹ | - | بکارگیری شیوه رگرسیون لجستیک ترتیبی برای طراحی محصول (شیر آب) |
| | [۱۴] | ۲۰۰۹ | - | بکارگیری شیوه رگرسیون لجستیک ترتیبی برای طراحی محصول |
| | [۱۵] | ۲۰۰۹ | - | بکارگیری شیوه رگرسیون لجستیک در تعیین ارتباط بین بلوغ افراد و متغیرهای کنترلی |
| | [۱۶] | ۲۰۰۹ | ارائه‌ی روش تخمین پارامتر در مدل رگرسیون لجستیک باینری | بکارگیری تخمین پارامتر در بررسی پروفایل |
| | [۱۷] | ۲۰۱۱ | - | بکارگیری الگوریتم‌های ابتکاری و فرا‌ابتکاری برای بهینه‌سازی متغیر پاسخ در مدل رگرسیون لجستیک ترتیبی |
| چند پاسخی با داده‌های طبقه‌بندی شده | [۱۸] | ۱۹۸۱ | نسبت متقاطع محلی | احتمالات توام |
| | [۱۹] | ۱۹۸۶ | مدل لجیت تجمعی حاشیه‌یی | احتمالات توام |
| | [۱] | ۱۹۹۴ | مدل لجستیک چندمتغیره | تحلیل مسائل دو و سه پاسخه باینری |
| | [۲۰] | ۱۹۹۷ | نسبت متقاطع فراگیر | احتمالات توام |
| | [۲۱] | ۱۹۹۸ | مدل ارتباطی برای پاسخ ترتیبی | امید انتظاری برای پاسخ‌های همبسته |
| | [۲۲] | ۱۹۹۵ | مدل لجستیک چندمتغیره برای پاسخ‌های باینری، اسمی و ترتیبی | - |
| | [۲۳] | ۲۰۰۲ | مدل لگاریتم خطی برای جداول احتمالی | اثربخشی پاسخ‌های طبقه‌بندی شده برای آزمایشات |
| | [۲۴] | ۲۰۰۴ | رویکرد تئوری برای متغیرهای طبقه‌بندی شده | ایجاد نمونه‌های تصادفی طبقه‌بندی شده ترتیبی همبسته با الگوهای ساختاری همبستگی |
| | پژوهش حاضر | - | رویکرد ابتکاری برای تحلیل مسائل چندپاسخی طبقه‌بندی شده | تخمین همزمان پارامترهای رگرسیونی برای مدل چندپاسخی طبقه‌بندی شده |

### ۳.۳. مدل لگاریتم خطی تعمیم‌یافته

مدل عمومی لگاریتم خطی معمولی که در روابط ۱ و ۳ برای مدل‌های مستقل و اشباع بیان شد به‌صورت:

$$\log \mu_{ij} = X\beta \qquad (۵)$$

است که در آن $X$ ماتریس متغیرهای کنترلی است و $\beta$ نیز پارامترهای مدل است و باید تخمین زده شود، چرا که مقادیر پارامتر در هر طرح آزمایش نامشخص است. از رابطه‌ی ۵ درمی‌یابیم که $\mu_{ij} = \exp(X\beta)$. با توجه به رابطه‌ی ۵ مدل لگاریتم خطی تعمیم‌یافته طبق رابطه‌ی ۶ تعریف می‌شود.

$$C\log(A\mu) = X\beta \qquad (۶)$$

هنگامی که ماتریس $A$ و $C$ یکه باشند، آنگاه مدل لگاریتم معمولی در رابطه‌ی ۵ حاصل می‌شود. لذا با توجه به رابطه‌ی ۴ و ۵، برای محاسبه‌ی مقادیر احتمالات توأم می‌توان از رابطه‌ی ۷ استفاده کرد:

$$\Pi_{ij} = \frac{\exp(\lambda + \lambda_i^X + \lambda_j^Y + \lambda_{ij}^{XY})}{\sum_a \sum_b (\lambda + \lambda_a^X + \lambda_b^Y + \lambda_{ab}^{XY})} = \frac{\exp(X\beta)}{\sum_a \sum_b (X\beta)} \qquad (۷)$$

در ادامه پس از بیان مسئله، چگونگی تخمین پارامتر برای بهینه‌سازی مدل تشریح خواهد شد.[۲۱]

## ۴. تعریف مسئله

چنان که در بخش‌های پیشین اشاره شد، هدف از این تحقیق، تخمین ضرایب مدل رگرسیون هم‌زمان سطوح چندپاسخی برای داده‌های طبقه‌بندی شده است، به‌طوری که وابستگی معناداری بین متغیرهای پاسخ وجود داشته باشد. یکی از روش‌های بهینه‌سازی مدل چندپاسخی وابسته، تخمین بهینه‌ی پارامترهای مدل با هدف بیشینه‌سازی تعداد انطباق‌هاست. منظور از انطباق در بهینه‌سازی آن است که مقادیر وقوع هریک از احتمالات توأم پاسخ‌های وابسته‌ی مشاهده شده در هر تیمار از طرح آزمایش از سایر احتمالات توأم برای پاسخ‌های مشاهده نشده بیشتر باشد. به‌عنوان نمونه طرح آزمایشی با دو متغیر پاسخ وابسته که به‌صورت باینری تعریف شده‌اند را در نظر بگیرید. اگر در تیمار اول مقادیر مشاهده شده برای دو متغیر پاسخ به‌ترتیب ۰ و ۱ باشد، آنگاه انطباق درصورتی رخ خواهد داد که احتمال وقوع مشاهده توأم (۰,۱) از حالات (۰,۱)، (۱,۱) و (۰,۰) بیشتر باشد. لذا با استفاده از این شیوه، پارامترهای مدل طوری تخمین زده می‌شود که تعداد انطباق‌ها برای کل تیمارهای طرح آزمایش بیشینه‌شود. در ادامه روش ابتکاری برای تخمین پارامترهای مدل ارائه خواهد شد.

### ۱.۴. روش ابتکاری تخمین پارامتر

رابطه‌ی ۴ و ۵ را در نظر بگیرید. با توجه به آن که متغیرهای پاسخ وابسته در این تحقیق دوتا هستند و به‌صورت باینری تعریف می‌شوند، باید برای هریک از تیمارهای طرح آزمایش چهار مقدار امید انتظاری و سپس چهار مقدار احتمال توأم محاسبه شود تا وجود انطباق در هر تیمار مورد بررسی قرارگیرد. به‌عبارت دیگر در هر تیمار نیاز به تخمین چهار نوع پارامتر مختلف برای محاسبه‌ی بهترین مقدار امید انتظاری و احتمالات توأم خواهیم داشت. بدین‌منظور و برای تخمین پارامتر از یک روش تکرارپذیر استفاده می‌شود. گام‌های تخمین پارامتر در روش ابتکاری به‌شرح زیر است.

#### ۱.۱.۴. گام‌های روش ابتکاری

۱. در مرحله‌ی اول می‌توان یک مقدار اولیه‌ی تصادفی برای پارامترهای مدل در نظرگرفت، یا با استفاده از روش حداقل مربعات خطا[۲۲] یک مقدار اولیه برای هریک از چهار پارامتر محاسبه کرد.

۲. نحوه‌ی حرکت پارامتر به سمت بهبود و نزدیک‌شدن به نقطه‌ی بهینه براساس روش تخمین پارامتر[۱۶] است اما شرط توقف الگوریتم این تحقیق با شرط توقف روش یادشده متفاوت است، به‌طوری که نقطه‌ی پایانی الگوریتم روش ابتکاری آن است که اگر پس از یک سری تکرار مشخص تعداد انطباق‌های مدل از حداکثر مقدار به دست آمده در تکرارهای قبل بیشتر نشود، الگوریتم در همان نقطه متوقف شده و پارامترهای آن نقطه بهترین پارامتر تخمینی خواهد بود. به‌عبارت دیگر می‌توان گفت که تفاوت اصلی روش تخمین پارامتر پیشنهادی با روش یه و همکاران در نوع هدف مسئله است به‌طوری که در روش پیشنهادی، تخمین پارامتر با هدف بیشینه‌کردن تعداد انطباق‌ها انجام می‌شود در حالی که در روش یه و همکاران تخمین پارامتر با هدف بیشینه‌کردن درست‌نمایی[۲۳] انجام شده است.

۳. پس از تخمین پارامتر بهینه، مقادیر مشاهدات انتظاری و سپس احتمالات توأم محاسبه شده و تعداد انطباق‌ها در هر تکرار تعیین می‌شود.

#### ۲.۱.۴. پیش کد الگوریتم ابتکاری

۱. مقادیر متغیرهای کنترلی و پاسخ را تعیین کنید و به‌ترتیب $X$ و $Y(y_1, y_2)$ بنامید.

۲. یک مقدار اولیه‌ی تصادفی برای پارامترهای مدل قرار دهید، یا این مقادیر اولیه را با استفاده از روش حداقل مربعات خطا تخمین زنید. معادله‌ی روش حداقل مربعات خطا عبارت است از:

$$\beta_{ij} = \left(X^T X\right)^{-1} * X^T Y_i, \qquad i,j = 1,2 \qquad (8)$$

که در آن $i$ بیانگر شماره‌ی متغیر پاسخ و $j$ نشان‌دهنده‌ی سطح مورد نظر متغیر پاسخ است.

با توجه به آن که تعداد متغیرهای پاسخ وابسته دوتاست، باید چهار مقدار از احتمالات توأم محاسبه شود. به‌همین دلیل، محاسبه‌ی چهار مقدار امید ریاضی توأم و چهار سری پارامتر لازم است. با استفاده از رابطه‌ی ۸ تنها دو سری از مقادیر اولیه تخمین زده شده که می‌توان سایر مقادیر را ضریبی از پارامترهای تخمین زده شده در نظرگرفت. پس از محاسبه‌ی پارامترهای اولیه، لازم است تابع خطی رگرسیونی $\eta(ij) = X\beta_{ij}$ محاسبه شود، چرا که در مراحل بعد کاربرد خواهد داشت.

۳. پس از محاسبه‌ی پارامترهای اولیه، مقادیر امید ریاضی و واریانس فردی را برای محاسبه‌ی هریک از مشاهدات انتظاری توأم تعیین کنید. مقادیر واریانس‌های محاسبه شده، همان عناصر قطر اصلی ماتریس قطری[۲۴] است. ماتریس قطری، ماتریس مربعی است که عناصر روی قطر اصلی آن عددی خاص و سایر عناصر ماتریس صفر باشد. لازم به ذکر است که تعداد سطرها و ستون‌های ماتریس قطری برابر تعداد تیمارهای طرح آزمایش مورد نظر خواهد بود.

۴. پس از محاسبه‌ی امید ریاضی و واریانس‌های فردی، باید مقادیر شاخص $Q$ را طبق رابطه‌ی ۹ محاسبه کرد. باید توجه داشت که در هر تکرار از الگوریتم ابتکاری، تعداد چهار شاخص $Q$ محاسبه می‌شود.

$$Q(ij) = \eta(ij) + \left(\hat{W}(ij)\right)^{-1}(y(i) - E(ij)) \quad \text{for } i,j = 1,2 \qquad (9)$$

۵. پس از محاسبه‌ی شاخص $Q$، پارامترهای جدید در هر تکرار طبق رابطه‌ی ۱۰ محاسبه می‌شوند.

$$\beta^{(k+1)} = \left(X^T W(ij) X\right)^{-1} X^T W(ij) Q(ij) \quad \text{for } i,j = 1,2 \qquad (10)$$

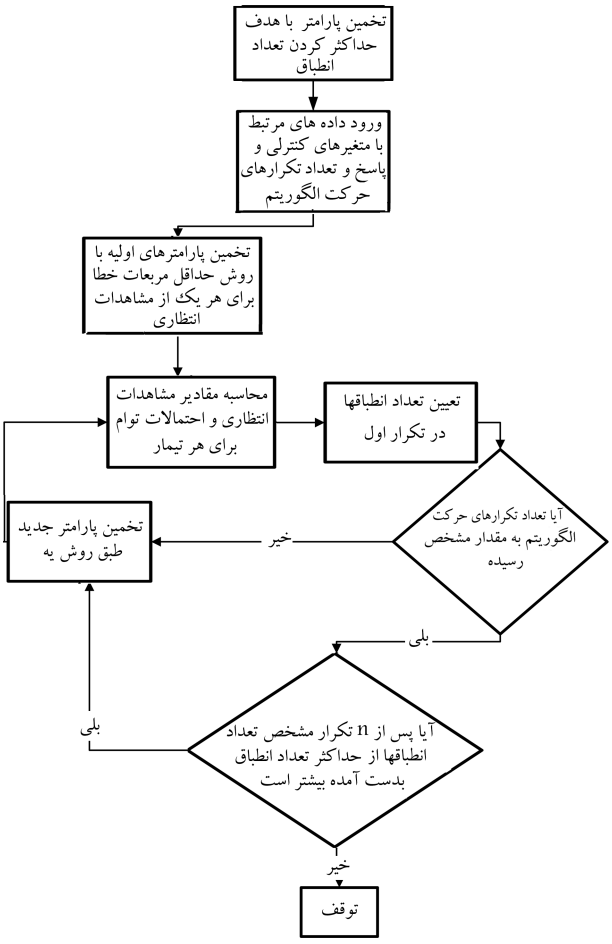
۶. پس از محاسبه‌ی پارامترهای جدید در هر مرحله، باید یک مقدار به $k$ اضافه کرد تا الگوریتم به تکرار بعد برود.

۷. مقادیر مشاهدات انتظاری توأم و در نهایت احتمالات توأم را با استفاده از روابط ۵ و ۷ محاسبه کنید.

۸. در هر تکرار تعداد انطباق را براساس مقادیر احتمالات توأم محاسبه کنید؛ به‌این ترتیب شرط توقف عبارت خواهد بود از: اگر پس از تعدادی تکرار مشخص،

تعداد انطباق‌های مسئله بیشتر از حداکثر تعداد انطباق‌های به دست آمده در تکرارهای قبل نباشد، آنگاه اجرای الگوریتم متوقف شده و ضرایب به دست آمده قابل ارائه است.

ارتباط روابط ۹ و ۱۰ با میزان انطباق‌ها در مقدار امید انتظاری و ماتریس قطری واریانس است که این مقادیر با استفاده از مقادیر احتمالات فردی و پارامترهای تخمینی در هر تکرار محاسبه می‌شود، به‌طوری که خود پارامترها نیز در ارتباط مستقیم با متغیرهای پاسخ (وابسته) تخمین زده می‌شوند. لذا پارامترهای تخمینی نهایی با استفاده از مقادیر امید و واریانس مرحله‌ی ماقبل به‌هنگام می‌شود. برای آشنایی بیشتر با نحوه‌ی تخمین پارامتر در روش ابتکاری، فلوچارت روش پیشنهادی در شکل ۱ نشان داده شده است. تفاوت روش پیشنهادی با روش یه و همکاران[۱۶] را می‌توان در وجود متغیرهای پاسخ وابسته، تابع هدف الگوریتم دانست که همان بیشینه‌سازی تعداد انطباق‌ها به‌جای بیشینه‌کردن درست‌نمایی است. پس از ارائه‌ی فلوچارت الگوریتم ابتکاری، در ضمن مثال‌های عددی، مناسب بودن روش پیشنهادی در مقایسه با حالت استقلال بین متغیرهای پاسخ نشان داده شده است. در این مقاله همچنین برای نشان‌دادن برتری روش پیشنهادی نسبت به روش‌های موجود، مقایسه‌یی بین مقادیر درست‌نمایی بیشینه و تعداد انطباق‌های مدل در روش ابتکاری و روش یه و همکاران انجام خواهد شد؛ بدین‌صورت که پس از تخمین پارامترهای مدل با استفاده از دو روش ابتکاری و موجود، مقادیر بیشینه درست‌نمایی حاصل از دو روش مقایسه می‌شوند تا درستی روش پیشنهادی مشخص شود، سپس برای نشان دادن برتری




شکل ۱. فلوچارت گام‌های روش ابتکاری.

روش پیشنهادی نسبت به روش موجود، مقادیر انطباق‌های حاصل از پارامترهای تخمینی مقایسه خواهد شد.

## ۵. مثال

در این بخش، با استفاده از سه مثال عددی با اندازه‌های متفاوت به بررسی روش ابتکاری پیشنهادی می‌پردازیم. مثال اول شامل پنج متغیر کنترلی با هفت تیمار، در آزمایش دوم ده تیمار به همراه هفت متغیرکنترلی، و در آزمایش سوم نیز پانزده تیمار به‌همراه ده متغیر کنترلی که همگی آزمایشات شامل دو متغیر پاسخ وابسته باینری هستند، در نظرگرفته می‌شوند. مقادیر متغیرهای کنترلی و پاسخ طرح‌های آزمایش در جداول ۳، ۴ و ۵ آمده است. لازم به ذکر است که در این مثال، تعداد عملیات لازم پس از آخرین تکرار با حداکثر تعداد انطباق، پنجاه تکرار است؛ مقادیر این جداول نیز به روش شبیه‌سازی تصادفی رایانه‌یی توسط نرم‌افزار MATLAB تعیین شده‌اند. لذا در این حالت تصادفی بودن مقادیر باقی‌مانده نشان داده می‌شود. همچنین بین متغیرهای پاسخ وابستگی متقابلی وجود داشته که برای شبیه‌سازی آن‌ها از الگوریتم نورتا[۲۵] استفاده شده است. این الگوریتم قادر است متغیرهای باینری وابسته به هم را با میزان وابستگی مشخص تولید کند. در هرصورت در مسائل واقعی لازم است آزمایش‌ها به‌صورت تصادفی اجرا شوند.

لازم به ذکر است که جهت بررسی درستی مقادیر احتمالات توأم در حالت وابستگی بین متغیرهای پاسخ، تنها برای داده‌های آزمایش اول، مقادیر مشاهدات انتظاری آورده می‌شود. این در حالی است که برای سایر آزمایشات، تنها مقادیر پارامترهای تخمینی و احتمالات توأم نشان داده خواهد شد. پس از اجرای الگوریتم در نرم‌افزار MATLAB و پس از انجام ۸۵ تکرار متوالی مقادیر پارامترهای تخمینی، مشاهدات انتظاری و احتمالات توأم برای طرح آزمایش اول محاسبه شده که به‌ترتیب در جداول ۶، ۷ و ۸ آورده شده است. لازم به ذکر است که بهترین نتایج در تکرار ۳۵ حاصل شده است.

از این پس با دادن یک مقدار جدید به متغیر کنترلی می‌توان احتمالات توأم وقوع هرجفت از متغیرهای پاسخ وابسته را محاسبه کرد. چنان که پیش‌تر نیز بیان شد، برای محاسبه‌ی هریک از احتمالات توأم برای چهار حالت ممکن در مسئله‌ی دوپاسخه‌ی وابسته، باید چهار مقدار امید انتظاری توأم محاسبه شود. مقادیر هریک از چهار حالت امید انتظاری برای هفت تیمار طرح آزمایش اول در جدول ۷ نشان داده شده است.

لازم به ذکر است که مقادیر پارامترهای تخمینی سطر اول، همان پارامتر ثابت[۲۶] مدل رگرسیونی بوده و سایر مقادیر از سطر دو تا پنجم نیز، ضرایب متغیرهای کنترلی

جدول ۳. مقادیر متغیرهای کنترلی و پاسخ طرح آزمایش اول.

| تیمار | متغیرهای کنترلی | | | | | متغیرهای پاسخ | |
|---|---|---|---|---|---|---|---|
| | $X1$ | $X2$ | $X3$ | $X4$ | $X5$ | $Y1$ | $Y2$ |
| ۱ | ۱ | ۲ | ۱ | ۲ | ۵ | ۱ | ۰ |
| ۲ | ۴ | ۴ | ۱ | ۵ | ۳ | ۱ | ۱ |
| ۳ | ۳ | ۱ | ۲ | ۳ | ۱ | ۰ | ۰ |
| ۴ | ۱ | ۲ | ۱ | ۲ | ۵ | ۱ | ۱ |
| ۵ | ۲ | ۱ | ۴ | ۱ | ۲ | ۱ | ۰ |
| ۶ | ۲ | ۱ | ۳ | ۲ | ۴ | ۰ | ۱ |
| ۷ | ۳ | ۴ | ۳ | ۵ | ۱ | ۰ | ۱ |

جدول ۴. مقادیر متغیرهای کنترلی و پاسخ طرح آزمایش دوم.

| متغیرهای پاسخ | | متغیرهای کنترلی | | | | | | | تیمار |
|---|---|---|---|---|---|---|---|---|---|
| $Y1$ | $Y2$ | $X7$ | $X6$ | $X5$ | $X4$ | $X3$ | $X2$ | $X1$ | |
| ۱ | ۰ | ۴ | ۱ | ۲ | ۳ | ۱ | ۵ | ۲ | ۱ |
| ۰ | ۱ | ۲ | ۳ | ۴ | ۲ | ۱ | ۵ | ۳ | ۲ |
| ۱ | ۰ | ۳ | ۲ | ۱ | ۴ | ۲ | ۱ | ۳ | ۳ |
| ۱ | ۱ | ۲ | ۱ | ۵ | ۲ | ۳ | ۲ | ۴ | ۴ |
| ۰ | ۱ | ۳ | ۵ | ۲ | ۱ | ۴ | ۳ | ۲ | ۵ |
| ۰ | ۱ | ۱ | ۳ | ۴ | ۲ | ۳ | ۵ | ۲ | ۶ |
| ۱ | ۱ | ۲ | ۲ | ۱ | ۵ | ۳ | ۴ | ۳ | ۷ |
| ۱ | ۱ | ۳ | ۱ | ۲ | ۵ | ۳ | ۲ | ۴ | ۸ |
| ۱ | ۰ | ۵ | ۲ | ۴ | ۱ | ۲ | ۳ | ۳ | ۹ |
| ۰ | ۱ | ۵ | ۱ | ۲ | ۳ | ۴ | ۳ | ۲ | ۱۰ |

جدول ۵. مقادیر متغیرهای کنترلی و پاسخ طرح آزمایش سوم.

| متغیرهای پاسخ | | متغیرهای کنترلی | | | | | | | | | تیمار |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $Y1$ | $Y2$ | $X10$ | $X9$ | $X8$ | $X7$ | $X6$ | $X5$ | $X4$ | $X3$ | $X2$ | $X1$ |
| ۱ | ۰ | ۵ | ۲ | ۱ | ۲ | ۴ | ۳ | ۳ | ۵ | ۴ | ۳ | ۱ |
| ۰ | ۱ | ۴ | ۲ | ۳ | ۲ | ۵ | ۲ | ۱ | ۲ | ۵ | ۲ | ۲ |
| ۱ | ۰ | ۲ | ۳ | ۴ | ۳ | ۲ | ۱ | ۴ | ۵ | ۳ | ۴ | ۳ |
| ۱ | ۱ | ۳ | ۵ | ۲ | ۲ | ۱ | ۵ | ۲ | ۱ | ۲ | ۵ | ۴ |
| ۰ | ۱ | ۳ | ۴ | ۲ | ۲ | ۱ | ۲ | ۵ | ۲ | ۴ | ۳ | ۵ |
| ۰ | ۱ | ۱ | ۲ | ۵ | ۲ | ۳ | ۱ | ۲ | ۴ | ۵ | ۲ | ۶ |
| ۱ | ۱ | ۱ | ۲ | ۵ | ۲ | ۳ | ۲ | ۴ | ۳ | ۴ | ۳ | ۷ |
| ۰ | ۱ | ۵ | ۳ | ۱ | ۳ | ۴ | ۲ | ۴ | ۳ | ۲ | ۵ | ۸ |
| ۱ | ۰ | ۴ | ۱ | ۳ | ۵ | ۳ | ۲ | ۳ | ۵ | ۴ | ۲ | ۹ |
| ۱ | ۱ | ۴ | ۲ | ۱ | ۵ | ۵ | ۳ | ۲ | ۴ | ۴ | ۳ | ۱۰ |
| ۰ | ۱ | ۲ | ۵ | ۳ | ۱ | ۳ | ۴ | ۵ | ۳ | ۴ | ۱ | ۱۱ |
| ۱ | ۱ | ۳ | ۲ | ۴ | ۲ | ۲ | ۴ | ۵ | ۴ | ۱ | ۳ | ۱۲ |
| ۰ | ۰ | ۱ | ۲ | ۴ | ۳ | ۴ | ۲ | ۵ | ۳ | ۵ | ۱ | ۱۳ |
| ۰ | ۱ | ۳ | ۴ | ۱ | ۵ | ۲ | ۴ | ۳ | ۵ | ۳ | ۴ | ۱۴ |
| ۱ | ۰ | ۲ | ۵ | ۳ | ۲ | ۱ | ۳ | ۴ | ۱ | ۳ | ۵ | ۱۵ |

جدول ۶. مقادیر پارامترهای تخمینی روش ابتکاری طرح آزمایش اول در تکرار ۳۵.

| $\beta_{01}$ | $\beta_{10}$ | $\beta_{00} * e^{003}$ | $\beta_{11}$ |
|---|---|---|---|
| ۷٫۱۵۵۷ | −۵۳٫۰۰۳۷ | −۳٫۱۶۱۱ | −۲۸٫۲۷۲۱ |
| −۸٫۴۱۳۶ | ۱۶٫۷۲۰۲ | ۰٫۱۳۲۴ | ۱۳٫۰۱۹۲ |
| −۱۷٫۸۸۹۵ | ۲۱٫۵۶۷۷ | ۰٫۶۴۴۹ | ۹٫۳۸۳۶ |
| −۴٫۱۵۹۵ | ۱٫۶۱۸۱ | ۰٫۴۴۷۴ | −۰٫۷۲۶۸ |
| ۲۰٫۶۶۱۸ | −۲۵٫۴۴۳۱ | −۰٫۳۰۷۸ | −۱۵٫۶۷۴۰ |
| −۰٫۱۱۷۰ | ۸٫۴۸۳۳ | ۰٫۳۸۱۲ | ۵٫۷۰۹۸ |

جدول ۷. مقادیر مشاهدات انتظاری در طرح آزمایش اول در تکرار ۳۵.

| $\mu_{01}$ | $\mu_{10}$ | $\mu_{00}$ | $\mu_{11}$ | تیمار |
|---|---|---|---|---|
| ۰٫۶۳۲۷ | ۱٫۰۰۰۲ | ۰٫۳۳۶۹ | ۰٫۹۸۸۶ | ۱ |
| ۲٫۰۹۹۵ | ۱٫۰۰۰۲ | ۱ | ۰٫۵۳۳۵ | ۲ |
| ۲٫۴۱ * $e^{007}$ | ۰ | ۰ | ۰ | ۳ |
| ۰٫۶۳۲۷ | ۱٫۰۰۰۲ | ۰٫۳۳۶۹ | ٫۹۸۸۶ | ۴ |
| ۰ | ۱٫۰۰۰۳ | ۰٫۰۰۰۷ | ٫۹۸۸۴ | ۵ |
| ۲٫۲۶۲۳ | ۰ | ٫۹۹۷۵ | ۰٫۰۲۹۰ | ۶ |
| ۲٫۹۱۷۴ | ۰ | ٫۹۹۴۷ | ۰ | ۷ |

جدول ۸. احتمالات توأم برای دو متغیر پاسخ وابسته در طرح آزمایش اول در تکرار ۳۵.

| وضعیت انطباق | $\Pi_{01}$ | $\Pi_{10}$ | $\Pi_{00}$ | $\Pi_{11}$ | تیمار |
|---|---|---|---|---|---|
| C | ۰٫۲۱۳۹ | ۰٫۳۳۸۱ | ۰٫۱۱۳۹ | ۰٫۳۳۴۲ | ۱ |
| D | ۰٫۴۵۳۱ | ٫۲۱۵۹ | ۰٫۲۱۵۸ | ۰٫۱۱۵۲ | ۲ |
| D | ۱ | ۰ | ۰ | ۰ | ۳ |
| C | ۰٫۲۱۳۹ | ٫۳۳۸۱ | ۰٫۱۱۳۹ | ۰٫۳۳۴۲ | ۴ |
| C | ۰ | ۰٫۵۰۲۸ | ۰٫۰۰۰۴ | ۰٫۴۹۶۹ | ۵ |
| C | ۰٫۶۸۷۹ | ۰ | ۰٫۳۰۳۳ | ۰٫۰۰۸۸ | ۶ |
| C | ۰٫۷۴۶۰ | ۰ | ۰٫۲۵۴۰ | ۰ | ۷ |

C: انطباق  D: عدم انطباق

اول تا پنجم است. مثلاً مقادیر موجود در ستون $\beta_{11}$ بیان‌گر پارامترهای مدل لگاریتم خطی اول یعنی $\ln \mu_{11}$ خواهد بود. به‌عبارت دیگر، هریک از چهار مقدار سمت چپ مدل عمومی لگاریتم خطی به‌صورت رابطه‌ی ۵، شامل تابع خطی از یک مقدار ثابت و متغیرهای کنترلی با ضرایب نشان داده شده در جدول ۶ است. بنابراین از همین پارامترها برای نوشتن معادله‌ی پیش‌بینی استفاده می‌کنیم. رابطه‌ی رگرسیونی برای طرح آزمایش اول به‌صورت رابطه‌ی ۱۱ نشان داده شده است.

$$\begin{aligned}
\ln \mu_{11} &= -۲۸٫۲۷۲۱ + ۱۳٫۰۱۹۲X_1 + ۹٫۳۸۳۶X_2 \\
&\quad -۰٫۷۲۶۸X_3 - ۱۵٫۶۷۴۰X_4 + ۵٫۷۰۹۸X_5 \\
\ln \mu_{00} &= -۳۱۶۱٫۱ + ۱۳۲٫۴X_1 + ۶۴۴٫۹X_2 \\
&\quad +۴۴۷٫۴X_3 - ۳۰۷٫۸X_4 + ۳۸۱٫۲X_5 \\
\ln \mu_{10} &= -۵۳٫۰۰۳۷ + ۱۶۷۲۰۲X_1 + ۲۱٫۵۶۷۷X_2 \\
&\quad +۱٫۶۱۸۱X_3 - ۲۵٫۴۴۳۱X_4 + ۸٫۴۸۳۳X_5 \\
\ln \mu_{01} &= ۷٫۱۵۵۷ - ۸٫۴۱۳۶X_1 - ۱۷٫۸۸۹۵X_2 \\
&\quad -۴٫۱۵۹۵X_3 + ۲۰٫۶۶۱۸X_4 - ۰٫۱۱۷۰X_5 \qquad (۱۱)
\end{aligned}$$

مقادیر موجود در جدول ۷ برای محاسبه‌ی درستی مقادیر احتمالات توأم که در جدول ۸ آمده، به‌کار می‌رود. چنان که در جدول ۸ مشاهده می‌شود، مقادیر پارامترهای تخمینی در تکرار ۳۵ موجب ۵ انطباق در طرح آزمایش مورد نظر شده است.

برای نمونه با توجه به مقدار احتمالات در اولین سطر جدول ۸، واضح است که بیشترین احتمال مربوط به $\Pi_{10}$ بوده و با توجه به داده‌های اولیه در جدول ۳ پیش‌بینی مقادیر پاسخ‌ها درست بوده و لذا در این مورد انطباق رخ داده است.

برای نشان‌دادن کارایی روش، می‌توان نتیجه‌ی حاصل (تعداد انطباق‌ها) را با حالت استقلال بین متغیرهای پاسخ مقایسه کرد. بدین‌منظور می‌توان مقادیر پارامترها را برای هر پاسخ به‌طور مستقل تخمین زد و سپس مقادیر احتمالات فردی (مستقل) را با استفاده از مدل رگرسیون لجستیک محاسبه کرد. برای محاسبه‌ی احتمالات فردی باید از رابطه‌ی ۱۲ استفاده کرد (مدل رگرسیون لجستیک باینری).[۲۵]

$$\Pi_i = \frac{\exp(X_i\beta)}{۱ + \exp(X_i\beta)} \qquad (۱۲)$$

لازم به ذکر است که در حالت استقلال دو یا بیشتر از دو متغیر، مقادیر احتمالات توأم برابر حاصل‌ضرب تک‌تک احتمالات فردی خواهد بود. مقادیر پارامترهای تخمینی در حالت استقلال، برای دو متغیر پاسخ و پنج متغیر کنترلی با استفاده از نرم‌افزار Minitab در جدول ۹ نشان داده شده است.

حال می‌توان با استفاده از رابطه‌ی ۱۲ مقادیر احتمالات فردی برای هرطبقه از دو متغیر پاسخ را به‌طور جداگانه محاسبه کرد، که نتایج حاصل به شرح جدول ۱۰ است.

در جدول ۱۰ مقادیر احتمالات فردی و توأم متغیرهای پاسخ با فرض استقلال آن‌ها ارائه شده و بیان‌گر تعداد ۲ انطباق در این حالت است.

همچنین می‌توان مقادیر احتمالات فردی و توأم را برای حالتی که متغیرهای پاسخ در طبقه‌بندی دیگر قرار داشته باشند (مثلاً برای تیمار اول به‌جای $Y1 = ۱$ و $Y2 = ۰$ قرار دهید $Y1 = ۰$ و $Y2 = ۱$) محاسبه کرد. هدف از این کار، در نهایت تعیین بهترین طبقه‌بندی برای متغیرهای پاسخ در حالت استقلال است. مقادیر احتمالات فردی و توأم در این حالت در جدول ۱۱ ارائه شده است.

تفاوت جداول ۱۰ و ۱۱ آن است که در جدول ۱۰ مقادیر احتمالات فردی و توأم برای متغیرهای پاسخ با استفاده از داده‌های جدول ۳ محاسبه شده است، در حالی که جدول ۱۱ هر چهار حالت از احتمالات توأم را نشان می‌دهد. مثلاً در سطر اول جدول ۱۰ تنها مقدار $\Pi_{10}$ محاسبه شده، اما در سطر اول جدول ۱۱ تمامی چهار مقدار از احتمالات توأم محاسبه شد. دلیل محاسبه‌ی احتمالات توأم در دو حالت اول و دوم، در واقع مقایسه با مقادیر احتمالات به دست آمده در حالت وابستگی بین متغیرهای پاسخ است که در ادامه به بررسی آن می‌پردازیم. چنان که از جدول ۱۱ برمی‌آید، تعداد انطباق در حالت دوم برابر ۲ است. عدم وجود تعداد انطباق زیاد در حالت استقلال بین متغیرها، نشان‌دهنده‌ی عدم تناسب بین متغیرهای کنترلی و پاسخ آزمایش تعیین شده است؛ به‌عبارت دیگر طرح آزمایش در حالت استقلال بین داده‌ها از تناسب پیوسته برخوردار است. از طرف دیگر با مقایسه‌ی تطبیقی جداول ۱۰ و ۱۱ با جدول ۸، بیشتر بودن احتمالات وقوع هریک از پاسخ‌ها در طبقه‌بندی‌های مربوطه در حالت وابسته نمایان می‌شود. در ادامه مقادیر پارامترهای تخمینی و احتمالات توأم در حالت وابستگی و استقلال بین متغیرهای پاسخ ارائه خواهد شد. دلیل آوردن مثال‌های بیشتر، نشان دادن درستی روش پیشنهادی برای مسائلی با اندازه‌های بزرگ‌تر (به‌عبارت دیگر با تعداد تیمار و متغیرهای کنترلی بیشتر) است. مقادیر پارامترهای تخمینی و احتمالات توأم طرح آزمایش دوم به‌ترتیب در جداول ۱۲ و ۱۳ آمده است.

چنان که ملاحظه می‌شود، مقادیر پارامترهای تخمینی با استفاده از روش ابتکاری منجر به وقوع هفت انطباق شده است. همچنین مقادیر پارامترهای تخمینی برای طرح آزمایش سوم در جدول ۱۴ نشان داده شده است. با استفاده از مقادیر پارامترهای تخمینی، می‌توان احتمالات توأم را برای این طرح آزمایش محاسبه کرد که در جدول ۱۵ آورده شده است.

چنان که از جدول ۱۵ برمی‌آید، تعداد ۱۰ انطباق با استفاده از پارامترهای تخمینی روش ابتکاری حاصل شده است. در ادامه نتایج حاصل از دو روش (با فرض وابستگی و مستقل بودن آن‌ها) در سه طرح آزمایش بیان شده در جدول ۱۶ ارائه شده و سپس در شکل ۲ مورد مقایسه و بررسی قرار گرفته است.

با توجه به شکل ۲ به‌طور واضح برتری روش ابتکاری پیشنهادی نسبت به حالت استقلال بین متغیرهای پاسخ، مشخص است. با مقایسه‌ی نتایج دو روش، روشن است که میزان افزایش تعداد انطباق‌ها در روش ابتکاری برای هر طرح آزمایش به طرح آزمایش با اندازه بزرگ‌تر، بیشتر از میزان افزایش همین انطباق در حالت استقلال متغیرهاست، به‌طوری که با افزایش تعداد تیمارها از ۷ به ۱۵، تعداد انطباق‌ها از ۵ به ۱۰ رسیده، اما در حالت دوم این تعداد از ۲ به ۴ تغییر یافته است.

جدول ۹. پارامترهای تخمینی برای هریک از دو متغیر پاسخ مستقل.

| | ثابت | $\beta_1$ | $\beta_2$ | $\beta_3$ | $\beta_4$ | $\beta_5$ |
|---|---|---|---|---|---|---|
| $(Y1)$ | ۵۹/۴۰۵ | ۲۰/۹۶۷ | ۳۷/۶۳۴ | −۱۵/۰۵۳ | −۴۸/۹۲۵ | −۴/۳۰۱ |
| $(Y2)$ | −۱۳۱/۹۸ | −۱/۶۱۴ | −۷/۷۹۵ | ۱۵/۸۵۹ | ۲۵ | ۱۶/۶۶۷ |

جدول ۱۰. مقادیر احتمالات فردی و توأم با فرض مستقل بودن متغیرهای پاسخ در طرح آزمایش اول.

| وضعیت انطباق | $\Pi_1 * \Pi_2$ | $\Pi_2$ | $\Pi_1$ | $Y2$ | $Y1$ | تیمار |
|---|---|---|---|---|---|---|
| C | ۰/۴۹۹۹ | ۰/۵۰ | ۰/۹۹۹۹ | ۰ | ۱ | ۱ |
| C | ۱/۰ | ۱/۰ | ۱/۰ | ۱ | ۱ | ۲ |
| D | ۰ | ۰ | ۰ | ۰ | ۰ | ۳ |
| D | ۰ | ۰ | ۱/۰ | ۱ | ۱ | ۴ |
| D | ۰ | ۰ | ۱/۰ | ۰ | ۱ | ۵ |
| D | ۰ | ۱/۰ | ۰ | ۱ | ۰ | ۶ |
| D | ۰ | ۱/۰ | ۰ | ۱ | ۰ | ۷ |

جدول ۱۱. مقادیر احتمالات توأم حالت دوم با فرض مستقل بودن متغیرهای پاسخ در طرح آزمایش اول.

| وضعیت انطباق | $\Pi_{01}$ | $\Pi_{10}$ | $\Pi_{00}$ | $\Pi_{11}$ | تیمار |
|---|---|---|---|---|---|
| C | ۰ | ۰/۵ | ۰ | ۰/۵ | ۱ |
| C | ۰ | ۰ | ۰ | ۱ | ۲ |
| D | ۰ | ۰ | ۰ | ۱ | ۳ |
| D | ۰ | ۱ | ۰ | ۰ | ۴ |
| D | ۰ | ۰ | ۰ | ۱ | ۵ |
| D | ۰ | ۰ | ۰ | ۱ | ۶ |
| D | ۰ | ۰ | ۰ | ۱ | ۷ |

جدول ۱۲. مقادیر پارامترهای تخمینی در حالت وابستگی بین متغیرهای پاسخ طرح آزمایش دوم.

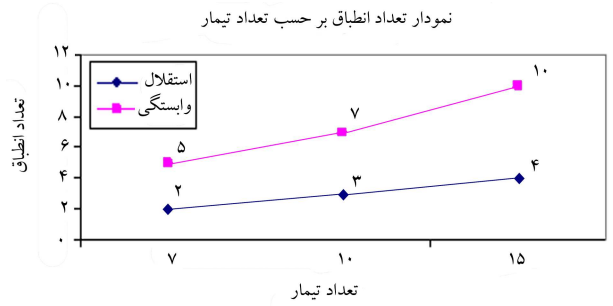
| $\beta_{01}$ | $\beta_{10}$ | $\beta_{00}$ | $\beta_{11}$ |
|---|---|---|---|
| ۹/۵۶۳۲ | −۱۶/۶۳۵۴ | ۷/۵۹۷۷ | −۱۳/۱۳۹۹ |
| ۳/۵۷۸۴ | ۱/۳۸۷۸ | ۳/۳۱۸۳ | ۰/۷۱۷ |
| ۰/۶۱۲۶ | ۱/۲۶۸۵ | ۰/۵۹۴۵ | ۰/۶۱۵ |
| −۰/۱۴۴۸ | ۲/۱۸۸۱ | −۰/۱۴۰۷ | ۱/۱۰۷۴ |
| −۲/۸۴۲ | ۰/۷۸۰۴ | −۲/۵۳۰۵ | ۰/۶۷۰۷ |
| −۳/۰۳۲۲ | ۰/۴۹۸۴ | −۲/۷۲۸۱ | ۰/۶۵۳۷ |
| −۱/۹۲۳۶ | ۰/۵۱۴ | −۱/۶۹۲۲ | ۰/۵۵۶۶ |
| −۰/۷۱۶۸ | −۰/۳۳۹۶ | −۰/۵۶۴۲ | ۰/۱۹۵۱ |

جدول ۱۳. احتمالات توأم برای دو متغیر پاسخ وابسته در طرح آزمایش دوم.

| وضعیت انطباق | $\Pi_{01}$ | $\Pi_{10}$ | $\Pi_{00}$ | $\Pi_{11}$ | تیمار |
|---|---|---|---|---|---|
| C | ۰/۵۲۱۶ | ۰/۰۲۳۷ | ۰/۴۳۲۳ | ۰/۰۲۲۴ | ۱ |
| C | ۰/۰۶۹۳ | ۰/۶۷۵ | ۰/۰۶۹۷ | ۰/۱۸۶۱ | ۲ |
| C | ۰/۵۴۹۵ | ۰/۰۱۷۹ | ۰/۴۱۳۳ | ۰/۰۱۹۲ | ۳ |
| D | ۰/۱۵۵ | ۰/۶۵۳۸ | ۰/۱۰۹۳ | ۰/۰۸۱۹ | ۴ |
| C | ۰/۰۰۷۸ | ۰/۸۷۹۲ | ۰/۰۰۷۹ | ۰/۱۰۵۱ | ۵ |
| C | ۰/۰۰۲ | ۰/۹۶۴۶ | ۰/۰۰۰۲ | ۰/۰۳۵۱ | ۶ |
| D | ۰/۰۱۵۳ | ۰/۹۴۶۷ | ۰/۰۱۲۸ | ۰/۰۲۵۱ | ۷ |
| D | ۰/۰۹۳۳ | ۰/۷۵۵۴ | ۰/۰۷۸۳ | ۰/۰۷۳ | ۸ |
| C | ۰/۳۹۶۶ | ۰/۰۷۷۶ | ۰/۳۸۱۷ | ۰/۱۴۴ | ۹ |
| C | ۰/۰۳۸۳ | ۰/۷۴۶۹ | ۰/۰۳۸۸ | ۰/۱۷۶۱ | ۱۰ |

جدول ۱۴. مقادیر پارامترهای تخمینی در روش ابتکاری طرح آزمایش سوم.

| $\beta_{01}$ | $\beta_{10}$ | $\beta_{00}$ | $\beta_{11}$ |
|---|---|---|---|
| ۵/۲۹۱ | ۸۰۴/۸۲۷۷ | ۵/۳۲۷۶ | ۱۲/۱۵۷۹ |
| ۲/۵۱۱۸ | −۲۹/۵۰۷۶ | ۲/۴۹۰۳ | −۰/۹۲۸ |
| ۰/۷۵۷۴ | −۴۵/۶۷۱۸ | ۰/۷۵۱۱ | −۱/۲۳۳۷ |
| −۰/۰۷۵۵ | −۹/۱۹۰۹ | −۰/۰۷۵۵ | −۰/۱۹۱۹ |
| −۰/۷۶ | −۳۲/۰۱۰۴ | −۰/۷۵۳۹ | −۰/۵۸۲۴ |
| ۰/۷۱۰۸ | −۱۹/۱۰۵۱ | ۰/۷۰۵۶ | −۰/۵۶۱۷ |
| −۱/۳۹۹۶ | −۹/۳۷۴ | −۱/۳۹۱۱ | ۰/۱۳۶۲ |
| −۰/۳۰۰۷ | −۱۷/۵۳۲۴ | −۰/۳۳۰۷ | −۰/۰۷۷۳ |
| −۰/۲۹۵ | −۴۰/۴۴۰۵ | −۰/۲۹۸۲ | −۰/۴۹۰۹ |
| −۳/۱۸۲ | −۱۷/۳۳۸۲ | −۳/۱۶۴ | ۰/۲۸۸۴ |
| −۰/۵۵۱۲ | −۳۷/۹۵۸۰ | −۰/۵۵۴۵ | −۰/۴۲۸۸ |




شکل ۲. مقایسه تعداد انطباق‌های سه طرح آزمایش در دو روش ابتکاری و استقلال متغیرها.

جدول ۱۵. احتمالات توأم برای دو متغیر پاسخ وابسته در طرح آزمایش سوم.

| وضعیت انطباق | $\Pi_{01}$ | $\Pi_{10}$ | $\Pi_{00}$ | $\Pi_{11}$ | تیمار |
|---|---|---|---|---|---|
| C | ۰٫۴۳۱۵ | ۰٫۰۲۴۵ | ۰٫۴۳۰۸ | ۰٫۱۱۳۲ | ۱ |
| C | ۰ | ۱ | ۰ | ۰ | ۲ |
| C | ۰٫۳۷۰۹ | ۰٫۰۰۸۹ | ۰٫۳۷۱۱ | ۰٫۲۴۹۲ | ۳ |
| D | ۰ | ۱ | ۰ | ۰ | ۴ |
| C | ۰ | ۰٫۹۹۹۹ | ۰ | ۰٫۰۰۰۱ | ۵ |
| C | ۰٫۰۰۳۱ | ۰٫۹۹۱۳ | ۰٫۰۰۲۹ | ۰٫۰۰۸۲ | ۶ |
| D | ۰٫۲۱۴۴ | ۰٫۵۵۲۷ | ۰٫۲۱۴۷ | ۰٫۰۱۸۲ | ۷ |
| C | ۰٫۱۱۰۹ | ۰٫۴۳۱۶ | ۰٫۱۱۰۶ | ۰٫۳۴۶۹ | ۸ |
| C | ۰٫۴۴۹۷ | ۰ | ۰٫۴۴۷۶ | ۰٫۱۰۲۷ | ۹ |
| D | ۰ | ۱ | ۰ | ۰ | ۱۰ |
| C | ۰ | ۰٫۲۶۹۲ | ۰ | ۰٫۷۳۰۸ | ۱۱ |
| D | ۰ | ۱ | ۰ | ۰ | ۱۲ |
| D | ۰٫۰۰۹۴ | ۰٫۰۸۸۵ | ۰٫۰۰۹۹ | ۰٫۸۹۲۲ | ۱۳ |
| C | ۰٫۰۶۵۷ | ۰٫۷۸۵۲ | ۰٫۰۶۶ | ۰٫۰۸۳۲ | ۱۴ |
| C | ۰٫۳۶۵۶ | ۰٫۰۵۵۶ | ۰٫۳۶۵۵ | ۰٫۲۱۳۳ | ۱۵ |

جدول ۱۶. جدول مقایسه‌یی نتایج حاصل از دو روش برای سه طرح آزمایش.

| طرح سوم | | طرح دوم | | طرح اول | | روش |
|---|---|---|---|---|---|---|
| عدم انطباق | انطباق | عدم انطباق | انطباق | عدم انطباق | انطباق | |
| ۵ | ۱۰ | ۳ | ۷ | ۲ | ۵ | پیشنهادی |
| ۶ | ۴ | ۷ | ۳ | ۵ | ۲ | موجود |

## ۶. اعتبارسنجی روش پیشنهادی

در این قسمت برای نشان‌دادن درستی روش ابتکاری پیشنهادی، ابتدا مقادیر درست‌نمایی بیشینه‌ی توأم[۲۷] را با استفاده از مقادیر تخمینی پارامترهای مدل در دو روش ابتکاری و موجود (روش یه و همکاران) محاسبه کرده تا کارایی روش پیشنهادی مشخص شود. سپس برای نشان دادن برتری روش ابتکاری نسبت به روش موجود، مقادیر انطباق‌های حاصل از تخمین پارامتر در دو روش مذکور مقایسه می‌شود. مقادیر حداکثر درست‌نمایی توأم و تعداد انطباق‌های حاصل برای دو روش ابتکاری و موجود براساس هریک از چهار مقدار پارامترهای تخمینی برای سه مثال شبیه‌سازی شده به‌صورت جدول ۱۷ و ۱۸ خواهد بود.

یادآور می‌شود که ابتدا با استفاده از هردو روش ابتکاری و موجود، مقادیر پارامترهای مدل تخمین زده می‌شوند. سپس با استفاده از رابطه‌ی حداکثر درست‌نمایی،[۲۶] مقادیر موجود در جدول ۱۷ محاسبه شده‌اند. مطابق جدول ۱۷ مقادیر حداکثر درست‌نمایی حاصل از روش ابتکاری پیشنهادی براساس مقادیر پارامترهای تخمینی جدید، نسبت به روش موجود یه و همکاران (۲۰۰۹)، از اختلاف ناچیزی برخوردار است. همچنین با محاسبه‌ی تعداد انطباق‌های حاصل از مقادیر پارامترهای تخمینی

جدول ۱۷. مقادیر حداکثر درست نمایی توأم در دو روش ابتکاری و موجود.

| روش | مقادیر حداکثر درست نمایی توأم براساس هریک از چهار دسته از پارامترهای تخمینی | | | | |
|---|---|---|---|---|---|
| **مثال اول** | $\beta_{01}$ | $\beta_{10}$ | $\beta_{00}$ | $\beta_{11}$ | |
| ابتکاری | −۲۷٫۶۶۵۲ | −۴۴٫۰۱ | −۱۳٫۵۰۱۷ | −۳۴٫۰۲۳۶ | |
| موجود | −۲۷٫۱۴۴۶ | −۴۳٫۲۳۷۶ | −۱۳٫۰۴ | −۳۴٫۰۱۰۱ | |
| **مثال دوم** | $\beta_{01}$ | $\beta_{10}$ | $\beta_{00}$ | $\beta_{11}$ | |
| ابتکاری | −۱۷٫۷۴۸۳ | −۹٫۸۹۶۴ | −۱۸٫۵۰۲۹ | −۱۵٫۳۰۵ | |
| **موجود** | −۱۶٫۱۷۲۲ | −۹٫۰۹۹ | −۱۸.۳۷۸ | −۱۵٫۲۴۳۱ | |
| **مثال سوم** | $\beta_{01}$ | $\beta_{10}$ | $\beta_{00}$ | $\beta_{11}$ | |
| ابتکاری | −۳۶٫۴۱۷۶ | −۳۶٫۳۰۴۴ | −۳۶٫۲۵۱۳ | −۲۱٫۴۴۵۴ | |
| موجود | −۳۶٫۰۸۱ | −۳۶٫۰۱۸ | −۳۶٫۰۰۲ | −۲۰٫۳۳۱۱ | |

جدول ۱۸. مقادیر انطباق‌های حاصل از دو روش ابتکاری و موجود.

| روش | طرح اول | | طرح دوم | | طرح سوم | |
|---|---|---|---|---|---|---|
| | عدم انطباق | انطباق | عدم انطباق | انطباق | عدم انطباق | انطباق |
| پیشنهادی | ۲ | ۵ | ۳ | ۷ | ۵ | ۱۰ |
| موجود | ۳ | ۴ | ۴ | ۶ | ۶ | ۹ |

در دو روش مذکور و مقایسه‌ی آن‌ها مشخص شده است که تخمین پارامتر با استفاده از روش ابتکاریِ پیشنهادی تعداد انطباق‌های بیشتری را به دست می‌آورد، که این مورد به‌معنای بالاتربودن میزان تناسب مدل در روش ابتکاری است.

با نگاهی به جداول ۱۷ و ۱۸ می‌توان دریافت که روش پیشنهادی به‌دلیل پیروی از روند کلی تخمین پارامتر برای مدل‌های غیرخطی باینری و با داشتن تابع هدف حداکثر تعداد انطباق مدل، هم دارای مقادیر درست‌نمایی بسیار نزدیک به مقادیر بهینه بوده و هم از تعداد انطباق بالایی برخوردار است. لذا در این قسمت درستی روش پیشنهادی و همچنین مزیت این روش نسبت به روش موجود یه و همکاران مشخص شده و می‌توان گفت که روش پیشنهادی، روشی مناسب برای تخمین پارامترهای مسئله‌ی دوپاسخی وابسته‌ی باینری است.

## ۷. مطالعه‌ی موردی

در این قسمت برای نشان دادن کاربرد روش پیشنهادی در مسائل واقعی، به بررسی یک مطالعه‌ی واقعی می‌پردازیم که دارای مشخصاتی مشابه با مثال‌های این پژوهش است و برگرفته از ادبیات تحقیق است.[۲] طبق یک مطالعه‌ی عملی از سه محله‌ی موجود در آمریکا در سال ۱۹۸۸، نظر تعدادی از ساکنین محلات برای ارزیابی عملکرد دولت در زمینه‌ی سلامتی افراد و اجرای قوانین بهداشتی جمع‌آوری شده است. در این مطالعه چهار متغیر کنترلی جنسیت، شرایط اجتماعی، جمعیت محله و استعمال دخانیات که هریک از تعداد سطح مشخصی برخوردارند، مورد سؤال واقع شد. همچنین دو متغیر پاسخ سلامتی و رعایت بهداشت نیز

جدول ۱۹. داده‌های مربوط به مطالعه‌ی موردی.

| متغیرهای پاسخ | | | متغیرهای کنترلی | | |
|---|---|---|---|---|---|
| سلامتی | رعایت بهداشت | جمعیت | اجتماعی | دخانیات | جنسیت |
| ۱ | ۱ | ۱ | ۱ | ۱ | ۱ |
| ۱ | ۱ | ۲ | | | |
| ۰ | ۱ | ۳ | | | |
| ۱ | ۱ | ۱ | ۲ | ۱ | ۱ |
| ۰ | ۰ | ۲ | | | |
| ۰ | ۰ | ۳ | | | |
| ۱ | ۱ | ۱ | ۱ | ۲ | ۱ |
| ۱ | ۱ | ۲ | | | |
| ۱ | ۱ | ۳ | | | |
| ۱ | ۱ | ۱ | ۲ | ۲ | ۱ |
| ۰ | ۱ | ۲ | | | |
| ۱ | ۰ | ۳ | | | |
| ۱ | ۱ | ۱ | ۱ | ۱ | ۲ |
| ۱ | ۱ | ۲ | | | |
| ۰ | ۱ | ۳ | | | |
| ۰ | ۱ | ۱ | ۲ | ۱ | ۲ |
| ۰ | ۱ | ۲ | | | |
| ۰ | ۰ | ۳ | | | |
| ۱ | ۱ | ۱ | ۱ | ۲ | ۲ |
| ۱ | ۱ | ۲ | | | |
| ۱ | ۱ | ۳ | | | |
| ۱ | ۱ | ۱ | ۲ | ۲ | ۲ |
| ۰ | ۱ | ۲ | | | |
| ۰ | ۱ | ۳ | | | |

**جنسیت:** ۱. مرد و ۲. زن. **دخانیات:** ۱. مصرف و ۲. عدم مصرف.
**شرایط اجتماعی:** ۱. مناسب و ۲. نامناسب. **جمعیت:** ۱. کم و ۲. متوسط و ۳. زیاد.

جدول ۲۰. مقادیر پارامترهای تخمینی برای مثال واقعی.

| $\beta_{01}$ | $\beta_{10}$ | $\beta_{00}$ | $\beta_{11}$ |
|---|---|---|---|
| ۴٫۸۵۱۶ | −۰٫۵۴۵۳ | ۴٫۶۹۱۸ | ۰٫۳۹۶۸ |
| −۱٫۲۲۴۲ | ۰٫۴۸۲۴ | −۱٫۱۶۳۵ | ۰٫۳۶۵۱ |
| ۰٫۰۵۲۹ | ۰٫۲۴۷۸ | ۰٫۰۹۱۵ | ۰٫۲۱۷۲ |
| −۱٫۷۷۳۸ | −۰٫۳۳۴۱ | −۱٫۷۴۰۷ | −۰٫۵۷۸۲ |
| −۱٫۰۹۹۴ | −۰٫۳۱۶۰ | −۱٫۱۰۶۸ | −۰٫۴۱۶۷ |

به‌صورت باینری تعریف شده است. داده‌های این بررسی در جدول ۱۹ ارائه شده است.

در ادامه مقادیر پارامترهای تخمینی به‌همراه احتمالات توأم مربوط به هریک از تیمارهای طرح آزمایش در جداول ۲۰ و ۲۱ ارائه شده است.

در ادامه مقادیر بیشینه‌ی احتمالات توأم در هر تیمار از طرح مورد نظر به‌همراه وضعیت انطباق حاصل از پارامترهای تخمینی ارائه شده است.

چنان که در جدول ۲۱ مشخص شده است، تعداد ۱۷ انطباق از ۲۴ تیمار کل طرح حاصل شده است. لذا رویکرد ابتکاری پیشنهادی برای مطالعات واقعی نیز انجام شده و نتایج خوبی را ارائه داده است.

جدول ۲۱. مقادیر احتمالات توأم و وضعیت انطباق برای هریک از تیمارهای مثال واقعی.

| وضعیت انطباق | احتمال متناظر | بیشینه احتمال | تیمار | وضعیت انطباق | احتمال متناظر | بیشینه احتمال | تیمار |
|---|---|---|---|---|---|---|---|
| C | $\Pi_{11}$ | ۰٫۳۷۹۱ | ۱۳ | D | $\Pi_{01}$ | ۰٫۳۷۳۶ | ۱ |
| C | $\Pi_{11}$ | ۰٫۴۵۰۲ | ۱۴ | C | $\Pi_{11}$ | ۰٫۲۹۳۷ | ۲ |
| D | $\Pi_{10}$ | ۰٫۴۹۰۰ | ۱۵ | C | $\Pi_{10}$ | ۰٫۳۵۱۱ | ۳ |
| D | $\Pi_{11}$ | ۰٫۴۵۹۰ | ۱۶ | C | $\Pi_{11}$ | ۰٫۳۱۵۷ | ۴ |
| C | $\Pi_{11}$ | ۰٫۴۷۶۲ | ۱۷ | C | $\Pi_{00}$ | ۰٫۳۹۳۵ | ۵ |
| C | $\Pi_{00}$ | ۰٫۴۸۸۲ | ۱۸ | D | $\Pi_{11}$ | ۰٫۴۴۶۲ | ۶ |
| C | $\Pi_{11}$ | ۰٫۳۹۴۸ | ۱۹ | D | $\Pi_{00}$ | ۰٫۳۵۲۰ | ۷ |
| C | $\Pi_{11}$ | ۰٫۴۵۸۶ | ۲۰ | C | $\Pi_{11}$ | ۰٫۲۷۵۲ | ۸ |
| C | $\Pi_{11}$ | ۰٫۳۹۴۸ | ۲۱ | C | $\Pi_{00}$ | ۰٫۳۵۲۰ | ۹ |
| C | $\Pi_{11}$ | ۰٫۴۶۱۴ | ۲۲ | C | $\Pi_{11}$ | ۰٫۳۳۲۷ | ۱۰ |
| D | $\Pi_{11}$ | ۰٫۴۶۶۲ | ۲۳ | D | $\Pi_{11}$ | ۰٫۴۰۴۲ | ۱۱ |
| C | $\Pi_{10}$ | ۰٫۴۹۸۱ | ۲۴ | C | $\Pi_{01}$ | ۰٫۴۴۶۳ | ۱۲ |

## ۸. نتیجه‌گیری

آنچه که در این نوشتار مورد بررسی و تحلیل قرارگرفت، تخمین هم‌زمان ضرایب مدل رگرسیون برای پاسخ‌های طبقه‌بندی شده وابسته است. به‌این منظور روشی ابتکاری برای تخمین پارامترهای مدل لگاریتم خطی که یکی از روش‌های تحلیل مسائل طبقه‌بندی شده با حداقل دو پاسخ است، ارائه شد. بدین‌منظور مقادیر پارامترهای مدل با توسعه‌ی الگوریتم تخمین پارامتر یه و همکاران (۲۰۰۹) برای مسئله‌ی دوپاسخی وابسته‌ی باینری با هدف بیشینه‌سازی تعداد انطباق‌های مدل براساس محاسبه‌ی احتمالات توأم تخمین زده شدند. برای نشان‌دادن درستی روش ابتکاری پیشنهادی، نتایج حاصل برای سه طرح آزمایش با اندازه‌های مختلف با نتایج حاصل از همین سه طرح آزمایش و در حالت استقلال بین متغیرهای پاسخ و نتایج حاصل از روش موجود یه و همکاران مقایسه شد که نتایج مقایسه‌ی اول، بیانگر برتری روش ابتکاری نسبت به روش استقلال متغیرها براساس مشخص تعداد انطباق‌هاست.

همچنین با مقایسه‌ی دو روش ابتکاری و موجود، مشخص شد که مقادیر درست‌نمایی بیشینه حاصل از پارامترهای تخمینی در روش پیشنهادی از اختلاف ناچیزی نسبت به مقادیر بهینه درست‌نمایی برخوردار است، این در حالی است که با استفاده از پارامترهای روش ابتکاری می‌توان به تعداد انطباق‌های بالاتری (تناسب بیشتر) نسبت به روش موجود دست یافت. بنابراین روش ابتکاری پیشنهادی با تخمین بهترین پارامترهای مدل، قادر است علاوه بر تنظیم حداکثر تعداد انطباق، مقادیر درست‌نمایی بسیار خوبی را برای مدل به دست آورد. لذا با توجه به تمامی نتایج موجود در این پژوهش می‌توان گفت که روش پیشنهادی روشی مناسب برای تخمین همزمان پارامتر مسائل دوپاسخی وابسته طبقه‌بندی شده خواهد بود.

به‌عنوان تحقیقات آتی می‌توان از الگوریتم‌های فراابتکاری برای بهبود جواب مسئله که همان افزایش تعداد انطباق‌های مدل است، استفاده کرد. همچنین می‌توان مسائلی با سه پاسخ طبقه‌بندی شده در حالت‌های مختلف وابستگی بین پاسخ‌ها را در نظرگرفته و به تحلیل و بهینه‌سازی آن‌ها پرداخت.

## پانوشت‌ها

1. log-linear model
2. binary, nominal & ordinal
3. logistic regression
4. metric
5. weighted signal-to-noise ratio
6. principal component analysis
7. karl pearson
8. G. Yule
9. Fisher
10. contingency table
11. walker & duncan
12. Cullagh
13. Taguchi's accumulation analysis
14. Nair's scoring scheme
15. Jeng's weighted probability scoring scheme
16. Goodman
17. local cross-ratios
18. row-column model
19. global cross ratios
20. independence model
21. Saturated Model
22. least square error
23. maximum likelihood
24. diagonal matrix
25. Norta
26. intercept
27. joint maximum likelihood

## منابع (References)


1. Cullagh, M.C. and Nelder, J.A. "Multivariate logistic model", *Journal of the Royal Statistical society, Series B*, **57**, pp. 533-546 (1995).
2. Ezgi Aktar, D., Anagun, A.S., and Koksal, G. "Determination of optimal product styles by ordinal logistic regression versus conjoint analysis for kitchen faucets", *International Journal of Industrial Ergonomics*, **39**, pp. 866-875 (2009).
3. Pal, S. and Gauri, S.K. "Assessing effectiveness of the various performance metrics for multi-response optimization using multiple regression", *Computers & Industrial Engineering*, **59**, pp. 976-985 (2010).
4. Datta, S., Nandi, G. and Bandyopadhyay, A. "Application of entropy measurement technique in grey based Taguchi methodfor solution of correlated multiple response optimization problems: A case study in welding", *Journal of Manufacturing Systems*, **28**, pp. 55-63 (2009).
5. Chang, H.H. "A data mining approach to dynamic multiple responses in Taguchi experimental design", *Expert Systems with Applications*, **35**, pp. 1095-1103 (2008).
6. Dash, M. and Liu, H. "Feature selection for classification", *Intell Data Anal*, **1**, pp. 131-156 (1997).
7. ChinWang, C. and Bernarn, J.C. "Integral DOE and MANOVA techniques for classification feature selection: Using solder joint defects as an example", *The International Journal of Advanced Manufacturing Technology*, **27**, pp. 392-396 (2004).
8. Chang, Y.C., Liu, C.T. and Huang, W.L. "Optimization of process parameters using weighted convex loss functions", *European Journal of Operational Research*, **196**, pp. 752-763 (2009).
9. Agresti, A., *An Introduction to Categorical Data Analysis*, Department of Statistics University of Florida-Gainesville, Florida. Second Edition (2007).
10. Walker, S.H. and Duncan, D.B. "Estimation of the probability of an event as a function of several independent variables", *Biometrica*, **54**, pp. 167-179 (1967).
11. Cullagh, M.C. "Regression model for ordinal data (with discussion)", *J. R. Statis Soc, Series B*, **42**, pp. 109-142 (1980).
12. Wu, F.C. and Yeh, C.H. "A comparative study on optimization methods for experiments with ordered categorical data", *Computers & Industrial Engineering*, **50**, pp. 220-232 (2006).
13. Zhou, F., Wu, D., Yang, X. and Jiao, J. "Ordinal logistic regression for affective product design", *Proceedings of the IEEE IEEM*, **52**, pp. 1986-1990 (2008).
14. Deimirtas, E.A., Anagun, A.S. and Koksal, G. "Determination of optimal product styles by ordinal logistic regression versus conjoint analysis for kitchen faucets", *International Journal of Industrial Ergonomics*, **39**, pp. 866-875 (2009).
15. Huang, B., Biro, F.M. and Dorn, D.D. "Determination of relative timing of pubertal maturation through ordinal logisticnext term modeling: Evaluation of growth and timing parameters", *Journal of Adolescent Health*, **45**, pp. 383-388 (2009).
16. Yeh, A.B., Huwang, L. and Li, Y.M. "Profile monitoring for a binary response", *IIE Transactions*, **41**, pp. 931-941 (2009).
17. Bashiri, M., Kamranrad, R. and Karimi, H. "Response optimization in ordinal logistic regression using heuristic and metaheuristic algorithm", *Journal of Sharif University*, Accepted, 2011 & it will publishe at 2012.
18. Goodman, L.A. "Association models and bivariate normal for contingency tables with orderd categories", *Biometrika*, **68**, pp. 347-355 (1981).
19. Dale, J.R. "Global cross- ratio models for bivariate, discrete, ordered responses", *Biometrics*, **42**, pp. 909-917 (1986).
20. Molenberg, G. and Lesaffer, E. "Marginal modeling of multivariate categorical data using a multivariate placket distribution", *J. Amer. Statist. Assoc*, **89**, pp. 633-644 (1997).
21. Lapp, K., Molenbrghs, G. and Lesaffre, E. "Models for the association between ordinal variables", *Computational Statistics & Data Analysis*, **28**, pp. 387-411 (1998).
22. Glonek, G.F.V. and Cullagh, M.C. "PMultivariate logistic models", *J. Roys. Statist. Ser. B.*, **47**, pp. 533-546 (1995).
23. Agresti, A., *Categorical Data Analysis*, Department of Statistics University of FloridaGainesville, Florida. John Wiley & Sons, Inc., Hoboken, New Jersey (2002).
24. Biswas, A. "Generating correlated ordinal categorical random samples", *Statistics & Probability Letters*, **70**, pp. 25-35 (2004).

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

26. Bashiri, M. and Kamranrad, R. "Parameter estimation for improving association indicators in binary logistic regression", *Journal of Isfahan University,* **2**, pp. 67-89 (2011).

# SIMULTANEOUSLY ESTIMATION OF REGRESSION COEFFICIENTS FOR CORRELATED MULTIPLE RESPONSES WITH CATEGORICAL DATA

**R. Kamranrad**
kamranrad@shahed.ac.ir
**M. Bashiri**(corresponding author)
bashiri@shahed.ac.ir
**School of Industrial**
**Shahed University**

Sharif Industrial Engineering and Management Journal
Volume 29, Issue 2, Page 145-157, Research Note

© Sharif University of Technology

- Received 28 August 2011; received in revised form 3 January 2012; accepted 6 May 2012.

**Abstract**

Regression coefficient estimation of multiple responses is an important problem which has been previously studied. In this paper, a heuristic algorithm has been proposed to estimate regression coefficients of the relationship between control and correlated binary response variables. In this paper, a log-linear model has been used for analyzing experiments with more than one categorical response variable. The considered model in this study is called the saturated log-linear model, because the responses (2 responses) of this research are cross correlated. To estimate the parameters of the logistic regression model for dependent responses, a heuristic iterative nonlinear method is used to maximize the number of concordance.

The proposed heuristic approach is a development of the parameter estimation method (Yeh et al. 2009) that is presented for the univariate binary logistic regression model, which is then applied to estimate the parameters of the log-linear model.

The proposed approach uses the concept of concordance. Concordance means that the joint probability of the occurrence of dependent responses in each treatment is more than other probabilities in the same treatment. Although much research has been undertaken on issues of single and multiple continuous responses and single categorical response problems, this study presents a new approach for simultaneous estimation of the parameters of the log-linear model with correlated categorical responses. Hence, it can be useful in real experimental cases. To indicate the efficiency of the heuristic method, the proposed approach has been compared to existing approaches in some hypothetical examples with simulated data and different sizes (seven, ten and fifteen treatments). Thus, initially, the parameter values of dependent responses were estimated, and then, the model parameters for each response variable were calculated separately. After estimating parameters, the joint probability values were obtained for both cases of dependent and independent response problems.

It should be noted that joint probability values, in the case of independence between variables, are equal to the product of the individual probabilities of two independent responses.

Because the number of concordance in the proposed heuristic method is greater than in the second case (independence between variables), the proposed heuristic method represents a good performance compared to the estimated coefficients of individual variables, on the basis of concordance measurement. Also, the proposed approach has been analyzed for a real case study from the literature, and with the existing method of (Yeh et al. 2009), and analyses show the efficiency of the proposed approach.

**Key Words:** Multi response surfaces, log-linear model, logistic regression, parameter estimation, number of concordance.