



A Different Attitude in Analyzing 3-Way Contingency Tables Using Conditional Independence

Abdi, M.¹, Madadi, M.²

¹Department of Statistics, Higher Education Complex of Bam, Bam, Iran.

²Department of Statistics, Shahid Bahonar University of Kerman, Kerman, Iran.

Corresponding author: M. Abdi, me.abdi@bam.ac.ir

Received: 23/9/2022 Revised: 16/6/2023 Accepted and Published Online: 20/6/2023.

Introduction

Categorical data are widely utilized in various fields, especially in social and medical sciences. One of the most widely used are contingency tables. Various methods have been proposed to analyze these tables, including odds ratio, log-linear models, logistic regression, logit model, etc., but most of these are asymptotical and are not efficient enough in analyzing tables with small frequencies. In many statistical texts, the log-linear model is used to analyze three-way contingency tables. The number of tests in the log-linear model to achieve analysis is large, and many of them cannot be analyzed. Also, tables with a small sample size cannot be analyzed with the log-linear model or when a large number of cells have zero frequency in the table, the use of log-linear models is not recommended. So, this paper is presented to facilitate this issue and increase the simplicity of the analysis, which leads to analysis without the use of log-linear models.

Material and Methods

In this paper, three-way contingency tables are analyzed using conditional independence properties. The types of independence for the three variables and how they related to each other, are studied. Conditional independence is tested through Fisher's exact test and a simple algorithm is proposed for the analysis of three-way contingency tables.

Results and Discussion

In this paper, a new approach using conditional independence and using Fisher's exact test, for analyzing three-way contingency tables is proposed. It is shown that the types of independence that are tested in log-linear models can be studied without using them and only by using conditional independence tests, in this way, the analysis of contingency tables with small frequency is also possible. A simple algorithm provided for the analysis, first tests the conditional independence of two variables at all levels of the third variable using Fisher's exact test. If all three conditional independence is accepted, the result is mutual independence, and if two of the hypotheses of conditional independence are accepted, the independence of two variables is obtained from one.

Conclusion

Analysis of three-way contingency tables is possible using log-linear models, odds ratio, logistic regression, Fisher's exact test, etc. Most of them are performed asymptotically, and for their implementation, a large sample size is needed, which is not always possible. In addition, the log-linear model in contrast to the saturated model with hierarchical tests that exist in statistical software may sometimes provide different answers and confuse the researcher. Also, the use of log-linear models is not applicable when facing zero frequency in the contingency tables. In this paper, a simple algorithm applying Fisher's exact test and using the concept of conditional independence was proposed. This algorithm can be an alternative to log-linear models in the analysis of these tables.

Keywords: Conditional Independence, Contingency Table, Fisher Exact Test, Log-Linear Model.

Mathematics Subject Classification (2010): 62H17, 60A99.



©The Author(s). The Publisher is Iranian Statistical Society.

This is an open access article distributed under the terms and conditions of [CC BY-NC 4.0](https://creativecommons.org/licenses/by-nc/4.0/)

رویکردی متفاوت در تحلیل جدول‌های پیش‌بینی سه طرفه با استفاده از استقلال شرطی

معراج عبدی^۱، محسن مددی^۲

^۱ گروه آمار، مجتمع آموزش عالی بم

^۲ گروه آمار، دانشگاه شهید باهنر کرمان

نویسنده مسئول: معراج عبدی، me.abdi@bam.ac.ir

تاریخ دریافت: ۱۴۰۱/۷/۱ تاریخ بازنگری: ۱۴۰۲/۳/۲۶ تاریخ پذیرش و انتشار: ۱۴۰۲/۳/۳۰

چکیده:

برای تحلیل روابط بین متغیرها در جدول‌های پیش‌بینی سه طرفه، آزمون‌های مختلفی وجود دارد که بیشتر این آزمون‌ها مجانبی هستند و در تحلیل جداولی با فراوانی‌های کوچک کارایی کافی را ندارند. در این مقاله در رویکردی جدید با استفاده از استقلال شرطی و همچنین بکارگیری آزمون دقیق فیشر، جدول‌های پیش‌بینی سه طرفه تحلیل می‌شوند. نشان داده می‌شود انواع استقلال که در مدل‌های لگ خطی مورد بررسی قرار می‌گیرند را می‌توان بدون استفاده از این مدل‌ها و تنها با بکارگیری استقلال شرطی و آزمون‌های استقلال دو متغیر بررسی کرد، به این ترتیب تحلیل جدول‌های پیش‌بینی با فراوانی کوچک نیز امکان‌پذیر می‌گردد. در انتها با ارائه چند مثال واقعی این بحث تشریح می‌شود. واژه‌های کلیدی: آزمون دقیق فیشر، استقلال شرطی، جدول پیش‌بینی، مدل لگ خطی. کد موضوع بندی ریاضی (۲۰۱۰): 60A99، 62H17.

۱ مقدمه

داده‌های گسسته (رسته‌ای) در علوم مختلف بخصوص در علوم اجتماعی و پزشکی بسیار مورد استفاده قرار می‌گیرند. یکی از پرکاربردترین آن‌ها، داده‌های جدول بندی متقاطع یا جدول‌های پیش‌بینی است. در تحلیل جدول‌های پیش‌بینی، تأثیر متغیرها بر روی هم و انواع استقلال‌های بین آنها مورد بررسی قرار می‌گیرند. برای تحلیل این



جدول‌ها روش‌های مختلفی پیشنهاد و ارائه شده‌اند، از جمله نسبت بخت‌ها، مدل‌های لگ خطی، رگرسیون لجستیک، مدل لجیت و غیره (اگرستی، ۲۰۱۳، ۲۰۱۹). برخی از روش‌ها مختص جدول‌های پیشابندی دو طرفه هستند و برخی برای جدول‌های پیشابندی سه طرفه و مراتب بالاتر نیز به کار می‌روند.

اکثر روش‌های تحلیل جدول‌های پیشابندی به صورت مجانبی انجام می‌شوند و پیش فرض این‌گونه آزمون‌ها حجم نمونه بزرگ است. جمع‌آوری تعداد زیاد نمونه همواره ممکن نیست و بایستی برای تحلیل به سراغ آزمون‌های دقیق رفت. یکی از مهمترین روش‌های تحلیل دقیق جدول‌های پیشابندی آزمون دقیق فیشر است. این آزمون در جدول‌های پیشابندی دوطرفه به کار می‌رود و با توجه به ماهیت توزیع آماره آزمون در جدول‌های 2×2 و $m \times k$ ، به ترتیب از تابع توزیع تجمعی فوق هندسی و فوق هندسی تعمیم‌یافته استفاده می‌شوند. محققان زیادی به مقایسه آزمون دقیق فیشر با سایر آزمون‌های دیگر پرداخته‌اند. در یک مطالعه گسترده دی آگوستینو و همکاران (۱۹۸۸)، آزمون دقیق فیشر را با ۲۲ آزمون استقلال دیگر مقایسه کرده و به این نتیجه رسیده‌اند که آزمون دقیق فیشر نسبت به سایر آزمون‌ها یک آزمون محافظه کارانه است. همچنین اولیویرا و همکاران (۲۰۱۸) و اورهموفا و همکاران (۲۰۱۹) این موضوع را تایید کرده‌اند که آزمون دقیق فیشر یک آزمون محافظه کارانه است. نکته مهم دیگر که در مقالات فوق به آن اشاره شده، توان آزمون دقیق فیشر است که در مقایسه با برخی آزمون‌ها کمتر است که دلیل آن ساختار گسسته این آزمون است. البته هرچه حجم نمونه بیشتر شود توان آزمون نیز افزایش می‌یابد، اما برای انجام محاسبات زمان بیشتری صرف می‌گردد (اولیویرا و همکاران، ۲۰۱۸؛ اورهموفا و همکاران، ۲۰۱۹).

در بسیاری از متون آماری برای تحلیل جدول‌های پیشابندی سه طرفه از مدل لگ خطی استفاده می‌شود (اندرسن، ۱۹۹۴، ۱۹۹۷؛ اگرستی، ۲۰۱۳، ۲۰۱۹). قابل ذکر است که وقتی تعداد سلول‌های با فراوانی صفر در جدول زیاد باشد، استفاده از مدل‌های لگ خطی پیشنهاد نمی‌شود. اگرستی (۲۰۱۳) روشی برای تحلیل جدول‌های پیشابندی $2 \times 2 \times 2$ براساس نسبت بخت‌های شرطی ارائه و روش اجرا را تنها برای آزمون‌های استقلال شرطی بیان نموده است اما چگونگی ارتباط آنها با انواع دیگر استقلال به طور دقیق و روشن بیان نشده است و نیاز به بحث مفصل دارد که در این مقاله به آن پرداخته می‌شود. اگرستی (۲۰۱۳) همچنین برای تحلیل جدول‌های پیشابندی $k \times 2 \times 2$ از آزمون کوکران-منتال-هانزل استفاده کرده است و آزمون دقیق استقلال شرطی را نیز در جدول‌های $k \times 2 \times 2$ ارائه نموده است که از نظر محاسباتی بسیار دشوار بوده و برای محاسبات نیاز به نرم افزارهای آماری است که در مراتب بالاتر k زمان بسیار زیادی برای تحلیل نیاز دارد. قریشی (۱۳۹۳) تحلیل بیزی جدول‌های پیشابندی با استفاده از چگالی‌های پیشین با ساختار معین را مورد مطالعه قرار دادند.

در بخش ۲ انواع مختلف استقلال برای سه متغیر تصادفی در حالت گسسته ارائه و روابط بین آنها بیان می‌شوند. در بخش ۳ مدل‌های لگ خطی و نحوه اجرای آزمون‌های استقلال به طور مختصر بررسی می‌شوند. سپس یک رویکرد متفاوت با به کارگیری استقلال‌های شرطی در تحلیل جدول‌های پیشابندی سه طرفه با ارائه یک الگوریتم کاربردی پیشنهاد می‌شود. در انتها در بخش ۴ مثال‌های عددی برای بررسی پیشنهاد ارائه شده مورد تحلیل قرار می‌گیرند.

۲ انواع استقلال برای سه متغیر تصادفی

فرض کنید X, Y, Z متغیرهای تصادفی گسسته به ترتیب با تکیه‌گاه‌های $S_Y = \{y_1, \dots, y_k\}$ و $S_Z = \{z_1, \dots, z_s\}$ باشند و نماد $\perp$ برای استقلال متغیرها استفاده شود.

تعریف ۱. متغیرهای تصادفی X, Y, Z توأماً مستقل هستند و با نماد $X \perp Y \perp Z$ نشان داده می‌شود، هرگاه

$$f_{X,Y,Z}(x, y, z) = f_X(x)f_Y(y)f_Z(z); \quad \forall x, y, z$$

که در آن $f_{X,Y,Z}(x, y, z)$ توزیع توأم و $f_X(x), f_Y(y)$ و $f_Z(z)$ توزیع‌های حاشیه‌ای سه متغیر تصادفی هستند.

تعریف ۲. متغیرهای تصادفی دو به دو مستقل هستند، یعنی $X \perp Y, X \perp Z$ و $Y \perp Z$ هرگاه به ترتیب

$$f_{X,Y}(x, y) = f_X(x)f_Y(y); \quad \forall x, y$$

$$f_{X,Z}(x, z) = f_X(x)f_Z(z); \quad \forall x, z$$

$$f_{Y,Z}(y, z) = f_Y(y)f_Z(z); \quad \forall y, z$$

تعریف ۳. متغیر تصادفی X مستقل از (Y, Z) است، $X \perp (Y, Z)$ ، هرگاه

$$f_{X,Y,Z}(x, y, z) = f_X(x)f_{Y,Z}(y, z); \quad \forall x, y, z$$

به طور مشابه $Y \perp (X, Z)$ و $Z \perp (X, Y)$ نیز تعریف می‌شوند.

تعریف ۴. اگر X مستقل از Y به شرط Z باشد، یعنی $X \perp Y | Z$ آنگاه

$$f_{X,Y|Z}(x, y | z) = f_{X|Z}(x | z)f_{Y|Z}(y | z); \quad \forall x, y, z$$

یا

$$f_{X,Y,Z}(x, y, z) = \frac{f_{X,Z}(x, z)f_{Y,Z}(y, z)}{f_Z(z)}; \quad \forall x, y, z$$

به طور مشابه $Y \perp Z | X$ و $X \perp Z | Y$ نیز تعریف می‌شوند.

برخی از مفاهیم استقلال‌های حاشیه‌ای و شرطی را می‌توان در دیوید (۱۹۹۸) و لویرفیدو (۲۰۱۰) و مراجع ذکر شده در آن‌ها یافت.

قضیه ۱. استقلال توأم، استقلال دو به دو را نتیجه می‌دهد، اما عکس آن لزوماً برقرار نیست.

برهان: به وضوح با جمع‌بندی روی هر متغیر در تعریف ۱، استقلال دو متغیر دیگر اثبات می‌شود.

قضیه ۲. اگر $X \perp Y \perp Z$ آنگاه $X \perp (Y, Z)$ ، $Y \perp (X, Z)$ و $Z \perp (X, Y)$. عکس قضیه زمانی برقرار خواهد بود که حداقل دو رابطه از روابط $X \perp (Y, Z)$ ، $Y \perp (X, Z)$ و $Z \perp (X, Y)$ برقرار شوند.

برهان: با توجه به قضیه (۱)، چون استقلال توأم، استقلال دو به دو را نتیجه می‌دهد، بنابر تعریف استقلال دو به دو به راحتی برای $X \perp (Y, Z)$ نتیجه می‌شود $f_{X,Y,Z}(x, y, z) = f_X(x)f_{Y,Z}(y, z)$. به همین صورت برای $Y \perp (X, Z)$ و $Z \perp (X, Y)$ نیز می‌توان قضیه را اثبات کرد. برعکس اگر

$$X \perp (Z, Y) \Rightarrow f_{X,Y,Z}(x, y, z) = f_X(x)f_{Y,Z}(y, z), \quad (۱)$$

$$Y \perp (X, Z) \Rightarrow f_{X,Y,Z}(x, y, z) = f_Y(y)f_{X,Z}(x, z), \quad (۲)$$

چون $Y \perp (X, Z)$ در اینصورت $Y \perp Z$ است، لذا $f_{Y,Z}(y, z) = f_Y(y)f_Z(z)$. بنابراین با جایگذاری در رابطه (۱)، $f_{X,Y,Z}(x, y, z) = f_X(x)f_Y(y)f_Z(z)$ حاصل می‌گردد و لذا نتیجه قضیه یعنی $X \perp Y \perp Z$ اثبات گردید.

قضیه ۳. اگر $X \perp (Y, Z)$ آنگاه $X \perp Y$ و $X \perp Z$. عکس قضیه لزوماً برقرار نیست.

برهان: چون $X \perp (Y, Z)$ بنابراین $f_{X,Y,Z}(x, y, z) = f_X(x)f_Y(y)f_Z(z)$ ، حال با جمع‌بندی روی متغیر Z در طرفین تساوی نتیجه می‌شود که $f_{X,Y}(x, y) = f_X(x)f_Y(y)$ ، لذا $X \perp Y$. همچنین با جمع‌بندی روی متغیر Y در طرفین تساوی $X \perp Z$ حاصل می‌شود.

قضیه ۴. $X \perp (Y, Z)$ اگر و فقط اگر $X \perp Y$ و $X \perp Z$.

برهان: اگر $X \perp (Y, Z)$ ، بنا بر تعریف ۴، $f_{X,Y,Z}(x, y, z) = f_X(x)f_{Y,Z}(y, z)$ ، همچنین با توجه به قضیه ۳، با جایگذاری نتایج زیر در رابطه فوق‌الذکر

$$X \perp Y \Rightarrow f_X(x) = \frac{f_{X,Y}(x, y)}{f_Y(y)}, \quad X \perp Z \Rightarrow f_X(x) = \frac{f_{X,Z}(x, z)}{f_Z(z)},$$

حاصل به صورت

$$f_{X,Y,Z}(x, y, z) = \frac{f_{X,Y}(x, y)f_{Y,Z}(y, z)}{f_Y(y)} \Rightarrow X \perp Z \mid Y,$$

$$f_{X,Y,Z}(x, y, z) = \frac{f_{X,Z}(x, z)f_{Y,Z}(y, z)}{f_Z(z)} \Rightarrow X \perp Y \mid Z,$$

به دست می‌آیند. برای حالت عکس طبق تعریف ۴

$$X \perp Z \mid Y \Rightarrow f_{X,Y,Z}(x, y, z) = \frac{f_{X,Y}(x, y) f_{Y,Z}(y, z)}{f_Y(y)}, \quad (۳)$$

$$X \perp Y \mid Z \Rightarrow f_{X,Y,Z}(x, y, z) = \frac{f_{X,Z}(x, z) f_{Y,Z}(y, z)}{f_Z(z)}. \quad (۴)$$

لذا چون طرفین (۳) و (۴) یکسان هستند، رابطه $f_{X,Z}(x, z) f_Y(y) = f_{X,Y}(x, y) f_Z(z)$ برقرار است و با جمع بندی دو طرف این رابطه روی متغیرهای Y و Z

$$f_X(x) f_Y(y) = f_{X,Y}(x, y) \Rightarrow X \perp Y, \quad (۵)$$

$$f_{X,Z}(x, z) = f_X(x) f_Z(z) \Rightarrow X \perp Z, \quad (۶)$$

بدست آمده و در نتیجه با قرار دادن رابطه (۵) در (۳) و رابطه (۶) در (۴) نتیجه می‌شود که $X \perp (Y, Z)$.

قضیه ۵. $X \perp Z$ و $X \perp Y \mid Z$ اگر و فقط اگر $X \perp (Y, Z)$.

برهان: برای اثبات مانند قضیه ۴ عمل می‌شود.

فرع ۱. با توجه به قضیه ۴، اگر $X \perp Y \mid Z$ و $X \perp Z \mid Y$ آنگاه $X \perp (Y, Z)$ و در نتیجه بنا بر قضیه ۳، $X \perp Y$ و $X \perp Z$. همچنین اگر $X \perp Y \mid Z$ ، $X \perp Z \mid Y$ و $X \perp Z$ آنگاه $Y \perp Z \mid X$ و $X \perp Y \perp Z$.

۳ مدل‌های لگ خطی و تحلیل جدول پیشابندی سه طرفه

جدول پیشابندی سه طرفه جدولی است که فراوانی‌ها را به تفکیک سطوح سه متغیر کیفی A ، B و C به ترتیب با m ، k و s سطح نمایش می‌دهد. جدول پیشابندی سه طرفه غالباً به صورت بلوکی با m سطر، k ستون و s لایه نشان داده می‌شود. داده‌های رسته‌ای با سه طرح مختلف نمونه‌گیری ۱- طرح نمونه‌گیری پواسنی، ۲- طرح نمونه‌گیری چندجمله‌ای و ۳- طرح نمونه‌گیری حاصلضرب چندجمله‌ای گردآوری می‌شوند. طرح‌های ۲ و ۳ از طرح ۱ بدست می‌آیند و هر سه طرح نمونه‌گیری منجر به تحلیل یکسانی می‌شوند، برای سادگی به بررسی طرح نمونه‌گیری پواسنی می‌پردازیم. در مدل پواسنی تعداد مشاهدات هر سلول جدول $X_{ij\ell}$ ، متغیری تصادفی از توزیع پواسن با پارامتر $\lambda_{ij\ell}$ است. $X_{ij\ell}$ ‌ها مستقل از هم فرض می‌شوند، به این ترتیب در یک جدول پیشابندی سه طرفه، متغیرهای تصادفی $X_{111}, \dots, X_{mks}$ دارای توزیع توأم

$$f(x_{111}, \dots, x_{mks}) = \prod_{i=1}^m \prod_{j=1}^k \prod_{\ell=1}^s \frac{\lambda_{ij\ell}^{x_{ij\ell}} e^{-\lambda_{ij\ell}}}{x_{ij\ell}!} \quad (۷)$$

است. داده‌های گردآوری شده از این طرح‌ها براساس یک مدل بنیادی یکسان مطرح می‌شود (بیرج، ۱۹۶۳). فرض‌های پارامتری در یک چارچوب پارامتری مشترک با نام پارامترگذاری لگ خطی که اساسا پارامترگذاری مجدد بر مبنای پارامترهای کانونی در خانواده نمایی است، ارائه می‌شود (اندرسن، ۱۹۹۴، ۱۹۹۷). در این مدل λ_{ijl} را با معادله لگ خطی

$$\ln \lambda_{ijl} = \theta^{(0)} + \theta_i^{(1)} + \theta_j^{(2)} + \theta_\ell^{(3)} + \theta_{ij}^{(12)} + \theta_{i\ell}^{(13)} + \theta_{jl}^{(23)} + \theta_{ijl}^{(123)} \quad (8)$$

$$i = 1, \dots, m, j = 1, \dots, k, \ell = 1, \dots, s$$

بیان می‌کنند، که در آن $\theta_{ijl}^{(123)}$ اثر متقابل مرتبه سوم (اثر متقابل سه متغیر A ، B و C)، $\theta_{ij}^{(12)}$ (اثر متقابل A و B)، $\theta_{i\ell}^{(13)}$ (اثر متقابل A و C) و $\theta_{jl}^{(23)}$ (اثر متقابل B و C) اثرات متقابل مرتبه دوم، $\theta_i^{(1)}$ ، $\theta_j^{(2)}$ و $\theta_\ell^{(3)}$ را اثر متغیرهای A ، B و C و $\theta^{(0)}$ را اثر کلی می‌نامند. برای کنترل برابری تعداد پارامترهای دو مدل (۷) و (۸)، قیدهای خطی تعریف می‌شوند تا تعداد پارامترها در هر دو مدل برابر با mks باشد (اندرسن، ۱۹۹۷). لازم است برای

جدول ۱. فهرست آزمون‌های مدل لگ خطی در جدول پیش‌بینی ۳ طرفه

شماره آزمون	فرض	تفسیر استقلالی
۱	$H_{123} : \theta_{ijl}^{(123)} = 0$	تفسیر ندارد
۲	$H_{12} : \theta_{ijl}^{(123)} = \theta_{ij}^{(12)} = 0$	$A \perp B \mid C$
۳	$H_{13} : \theta_{ijl}^{(123)} = \theta_{i\ell}^{(13)} = 0$	$A \perp C \mid B$
۴	$H_{23} : \theta_{ijl}^{(123)} = \theta_{jl}^{(23)} = 0$	$B \perp C \mid A$
۵	$H_{12,13} : \theta_{ijl}^{(123)} = \theta_{ij}^{(12)} = \theta_{i\ell}^{(13)} = 0$	$A \perp (B, C)$
۶	$H_{12,23} : \theta_{ijl}^{(123)} = \theta_{ij}^{(12)} = \theta_{jl}^{(23)} = 0$	$B \perp (A, C)$
۷	$H_{13,23} : \theta_{ijl}^{(123)} = \theta_{i\ell}^{(13)} = \theta_{jl}^{(23)} = 0$	$C \perp (A, B)$
۸	$H_{12,13,1} : \theta_{ijl}^{(123)} = \theta_{ij}^{(12)} = \theta_{i\ell}^{(13)} = \theta_i^{(1)} = 0$	$A \perp (B, C), A = u$
۹	$H_{12,23,2} : \theta_{ijl}^{(123)} = \theta_{ij}^{(12)} = \theta_{jl}^{(23)} = \theta_j^{(2)} = 0$	$B \perp (A, C), B = u$
۱۰	$H_{13,23,3} : \theta_{ijl}^{(123)} = \theta_{i\ell}^{(13)} = \theta_{jl}^{(23)} = \theta_\ell^{(3)} = 0$	$C \perp (A, B), C = u$
۱۱	$H_{12,13,23} : \theta_{ijl}^{(123)} = \theta_{ij}^{(12)} = \theta_{i\ell}^{(13)} = \theta_{jl}^{(23)} = Q = 0$	$A \perp B \perp C$
۱۲	$H_1 : Q = \theta_i^{(1)} = 0$	$A \perp B \perp C, A = u$
۱۳	$H_2 : Q = \theta_j^{(2)} = 0$	$A \perp B \perp C, B = u$
۱۴	$H_3 : Q = \theta_\ell^{(3)} = 0$	$A \perp B \perp C, C = u$
۱۵	$H_{1,2} : Q = \theta_i^{(1)} = \theta_j^{(2)} = 0$	$A \perp B \perp C, A = B = u$
۱۶	$H_{1,3} : Q = \theta_i^{(1)} = \theta_\ell^{(3)} = 0$	$A \perp B \perp C, A = C = u$
۱۷	$H_{2,3} : Q = \theta_j^{(2)} = \theta_\ell^{(3)} = 0$	$A \perp B \perp C, B = C = u$
۱۸	$H_{1,2,3} : Q = \theta_i^{(1)} = \theta_j^{(2)} = \theta_\ell^{(3)} = 0$	$A \perp B \perp C, A = B = C = u$

تحلیل جدول‌های پیش‌بینی سه طرفه و حتی مراتب بالاتر، فرض‌هایی که باید آزمون شوند به صورت سیستماتیک و مرتب ردیف شوند. با توجه به mks پارامتر کانونی مدل، فرض‌های زیادی وجود دارند که ممکن است پژوهشگر در

ابتدا به آزمون آنها وسوسه شود. بسیاری از این فرض‌ها غیر قابل توجیه هستند یا از آزمون‌های دیگر نتیجه می‌شوند. لذا فرض‌ها به صورت سلسله مراتبی مطرح و آزمون می‌شوند. برای تحلیل جدول‌های پیشابندی با روش لگ خطی، آن دسته از فرض‌هایی که تعبیرهای با معنای استقلالی دارند مورد بررسی قرار می‌گیرند (اندرسن، ۱۹۹۷). جدول ۱ کلیه آزمون‌های مدل لگ خطی را برای متغیرهای تصادفی A ، B و C نشان می‌دهد. منظور از $u = A$ در جدول ۱ این است که میانگین سطوح متغیر A یکسان است و دارای توزیع یکنواخت (همگن) است.

بنا بر جدول ۱، برای تحلیل جدول پیشابندی با مدل لگ خطی (۸) بایستی حداکثر ۱۸ آزمون انجام شود، هر چند با ۱۱ آزمون اول می‌توان به قوی‌ترین نوع استقلال دست یافت. این فرض‌ها سلسله مراتبی هستند، به این معنی که هریک از فرض‌ها، فرضیه‌های قبلی را در خودش دارد. برای مثال اگر فرضیه شماره ۲ پذیرفته شود فرضیه شماره ۱ نیز پذیرفته شده و نیاز به آزمون مجدد آن وجود ندارد.

در جدول‌های پیشابندی با انجام آزمون‌های ۱ تا ۱۱ در جدول ۱ حداکثر ۱۱ آزمون برای تحلیل جدول بایستی انجام شود تا به قوی‌ترین نوع استقلال یعنی استقلال توأم رسید. قابل ذکر است با تعداد بالای آزمون‌ها خطای توأم آزمون‌ها نیز افزایش می‌یابد. علاوه بر این کلیه این آزمون‌ها به طور مجانبی انجام می‌شوند و مدل لگ خطی برای تعداد نمونه کم کارایی ندارد. بخصوص هنگامی که تعداد سلول‌های با فراوانی صفر در جدول زیاد باشد، استفاده از مدل‌های لگ خطی پیشنهاد نمی‌شود. ضمن اینکه آزمون‌های لگ خطی در مقابل مدل اشباع با آزمون‌های سلسله مراتبی که در نرم افزارهای آماری وجود دارند برخی اوقات ممکن است جواب متفاوت ارائه دهند و موجب سردرگمی پژوهشگر شوند. اما با استفاده از قضیه ۴ و همچنین به کارگیری آزمون دقیق فیشر می‌توان یک جدول سه طرفه را به طور کامل تحلیل کرد. بنابراین پیشنهاد می‌شود برای تحلیل یک جدول پیشابندی سه طرفه از الگوریتم زیر استفاده کرد.

الگوریتم ۱. فرض کنید متغیرهای تصادفی A, B, C در جدول پیشابندی سه طرفه با ابعاد به ترتیب m, k و s تعریف شوند،

گام ۱- با استفاده از آزمون دقیق فیشر، استقلال شرطی A و B در هر یک از سطوح متغیر C آزمون می‌شود. به همین ترتیب استقلال شرطی A و C در هر یک از سطوح متغیر B و استقلال شرطی B و C در هر یک از سطوح متغیر A بررسی می‌شود.

گام ۲- اگر برای حداقل یکی از سطوح متغیر C ، فرض استقلال در گام ۱ رد شود، فرض $A \perp B \mid C$ رد می‌شود. به همین ترتیب استقلال شرطی $B \mid A \perp C$ و استقلال شرطی $A \mid B \perp C$.

گام ۳- با توجه به قضیه ۴ و فرع ۱، می‌توان یک جدول پیشابندی سه متغیره را به صورت زیر تحلیل کرد:

- اگر هر سه استقلال شرطی پذیرفته شود، استقلال توأم نتیجه می‌شود.
- اگر دو تا از فرض‌های استقلال شرطی پذیرفته شود، استقلال دو متغیر از یک متغیر حاصل می‌گردد.

توجه داریم که وجود استقلال شرطی $A \perp B \mid C$ مستلزم پذیرش استقلال A و B در همه سطوح متغیر C است، همچنین در صورت رد استقلال‌های شرطی ممکن است استقلال‌های دو به دو وجود داشته باشد. همانطور که

۱۱۲ رویکردی متفاوت در تحلیل جدول‌های پیش‌بینی سه طرفه

در مقدمه اشاره شد آزمون دقیق فیشر دارای محاسبات پیچیده‌ای است بنابراین این الگوریتم برای حجم نمونه زیاد دارای سرعت عمل کمتری خواهد بود.

۴ تحلیل داده‌های رسته‌ای واقعی

در این بخش تحلیل داده‌های رسته‌ای برای تشریح نتایج ارائه می‌شود. کد برنامه نرم‌افزار R و خروجی‌های نرم‌افزار SPSS از طریق نویسنده مسئول مقاله قابل دسترس خواهد بود.

مثال ۱. (اندرسن، ۱۹۹۴) در سال ۱۹۶۸ در کشور دانمارک ۷۱۵ کارگر منتخب از کارگران و سرکارگران صنایع به سوال‌هایی درباره میزان رضایت شغلی و نحوه مدیریت کارخانه‌ها پاسخ داده‌اند. پس از انجام مصاحبه و جمع‌آوری اطلاعات، نتایج در جدول ۲ خلاصه گردید. طبق مراحل الگوریتم ۱، با استفاده از نرم افزار IBM

جدول ۲. وضعیت رضایت شغلی کارگران، سرکارگران و کیفیت مدیریت کارخانه.

رضایت کارگر (A)			
کیفیت مدیریت کارخانه (C)	رضایت سرکارگر (B)	پایین	بالا
ضعیف	پایین	۱۰۳	۸۷
	بالا	۳۲	۴۲
قوی	پایین	۵۹	۱۰۹
	بالا	۷۸	۲۰۵

SPSS Statistics، نسخه ۲۶، در گام اول، نتایج آزمون استقلال‌های شرطی جزئی، برای هر یک از سطوح عامل شرط با استفاده از آزمون دقیق فیشر بدست آمده است. با توجه به نتایج جدول ۳ فرض استقلال $A \perp B \mid C$ در سطح معنی‌داری ۵ درصد پذیرفته شده است، بدین معنی که رضایت کارگران (A) مستقیماً تحت تأثیر رضایت سرکارگران (B) قرار نمی‌گیرند بلکه هر دو آنها وابسته به کیفیت مدیریت کارخانه هستند. بنابراین عملکرد مدیریت موجب رضایت یا نارضایتی کارگران و سرکارگران می‌شود. فرض‌های استقلال $A \perp C \mid B$ و $B \perp C \mid A$ در سطح معنی‌داری ۵ درصد رد شده‌اند. در این مثال آخرین آزمون پذیرفته شده $A \perp B \mid C$ است. در جدول ۳ نتایج آزمون‌های استقلال دو به دو $A \perp B$ ، $A \perp C$ و $B \perp C$ نیز آمده که نشان از رد این فرض‌ها دارد.

برای تحلیل جدول ۲ از طریق مدل‌های لگ خطی بایستی تعداد بیشتری آزمون به صورت سلسله مراتبی انجام و نتایج استقلال بین متغیرها را تعیین کرد. نتایج حاصل از این روش نیز با نتایج روش پیشنهادی یکسان است و آخرین فرض پذیرفته شده $A \perp B \mid C$ است. ملاحظه می‌شود آزمون دقیق فیشر که در جدول ۳ استفاده شده است یک آزمون دقیق است و فرض استقلال $A \perp B \mid C$ را با مقدار معنی‌داری بالایی در سطح معنی‌داری ۵ درصد پذیرفته است. این در حالی است که در مدل لگ خطی که مجانبی است، فرض استقلال $A \perp B \mid C$ در

جدول ۳. نتایج آزمون استقلال شرطی (آزمون دقیق فیشر) در سطوح عوامل در مثال ۱.

آزمون	مقدار معنی داری
$A \perp B$	< 0.001
$A \perp B C_1$	۰/۱۳۲
$A \perp B C_2$	۰/۱۱۲
$A \perp C$	< 0.001
$A \perp C B_1$	< 0.001
$A \perp C B_2$	۰/۰۱۱
$B \perp C$	< 0.001
$B \perp C A_1$	< 0.001
$B \perp C A_2$	< 0.001

آزمون سلسله مراتبی دارای آماره آزمون ۵/۳۲۲ با مقدار معنی داری ۰/۰۲۱ رد شده است، در حالی که همین آزمون در مقابل مدل اشباع دارای آماره آزمون ۵/۳۸۷ با مقدار معنی داری ۰/۰۶۸ بوده و فرض استقلال $A \perp B | C$ را به صورت ضعیف پذیرفته است. نتایج نشان داد که آزمون لگ خطی برای یک فرضیه دو نتیجه متفاوت ارائه می دهد که باعث سردرگمی پژوهشگر می شود، در حالی که روش پیشنهاد شده در این مقاله که از آزمون دقیق فیشر استفاده می کند فرض استقلال $A \perp B | C$ را با توان بیشتری پذیرفته است.

مثال ۲. (اندرسن، ۱۹۹۴) تعداد افراد فوت شده در تصادفات ترافیکی در جاده های اصلی و فرعی در طی دو سال، با توجه به وضعیت جاده که آیا دارای محدودیت سرعت ۹۰ کیلومتر بوده یا خیر در جدول ۴ نشان داده شده است. هدف بررسی اثرات ممکن محدودیت سرعت در جاده های مختلف است. طبق نتایج جدول ۵ دو فرض استقلال

جدول ۴. تعداد سوانح رانندگی منجر به فوت طی دو سال.

نوع جاده (C)		
سال (A)	حداکثر سرعت (B)	جاده اصلی
سال اول	۹۰ کیلومتر در ساعت آزاد	۸
		۴۲
سال دوم	۹۰ کیلومتر در ساعت آزاد	۱۱
		۳۷
		۴۵
		۶۹

$A \perp B | C$ و $A \perp C | B$ در سطح معنی داری ۵ درصد رد نمی شوند لذا فرض های استقلال شرطی فوق را می پذیریم، بنابر الگوریتم ۱ با پذیرفتن این دو استقلال شرطی فرض استقلال $A \perp (B, C)$ رد نخواهد شد. بدین معنی که نوع جاده در تعداد تصادفات دو سال تاثیری ندارد و همچنین حداکثر سرعت نیز در تعداد تصادفات منجر

جدول ۵. نتایج آزمون‌های استقلال شرطی (آزمون دقیق فیشر) در سطوح عوامل در مثال ۲.

آزمون	مقدار معنی‌داری
$A \perp B$	۰/۱۹۳
$A \perp B \mid C_1$	۰/۳۲۱
$A \perp B \mid C_2$	۰/۲۷۵
$A \perp C$	۰/۴۳۶
$A \perp C \mid B_1$	۰/۴۴۹
$A \perp C \mid B_2$	۰/۴۵۱
$B \perp C$	۰/۰۰۲
$B \perp C \mid A_1$	۰/۰۱۳
$B \perp C \mid A_2$	۰/۰۴۸

به فوت در دو سال اثری نداشته است. با توجه به عدم رد فرض $A \perp (B, C)$ ، ممکن است علاقه مند به آزمون برابری متوسط تعداد تصادفات در دو سال باشیم که به راحتی با استفاده از آزمون برابری نسبت‌های دو جمله‌ای با مقدار معنی‌داری ۱۰٪ نتیجه می‌شود که تعداد تصادفات منجر به مرگ در دو سال یکسان نیست و تفاوت دارند. فرض استقلال $B \perp C \mid A$ در سطح معنی‌داری ۵ درصد را نیز نمی‌توان پذیرفت. لذا بین نوع جاده و محدودیت سرعت استقلال وجود ندارد و نوع جاده در کاهش تصادفات رانندگی اثر دارد. در جدول ۵، آزمون‌های استقلال دو به دو $A \perp B$ و $A \perp C$ نیز پذیرفته شده و $B \perp C$ رد شده است. برای فرض استقلال $A \perp (B, C)$ با استفاده از مدل لگ خطی، مقدار آماره آزمون ۳/۱۳۲ با درجه آزادی ۳ و مقدار معنی‌داری ۰/۳۷۲ محاسبه شده است که نشان از پذیرش فرض استقلال $A \perp (B, C)$ دارد.

مثال ۳. (شاهکار، ۱۳۸۷) برای بررسی تأثیر تحصیلات و شغل مادر در وضعیت تحصیلی فرزندان، اطلاعات جدول ۶ جمع‌آوری شده‌اند،

جدول ۶. بررسی تأثیر تحصیلات و مشغله مادر در وضعیت تحصیلی فرزندان.

معدل فرزند (C)				
شغل مادر (A)	تحصیلات مادر (B)	زیر ۱۵	بین ۱۵ تا ۱۷	بالتر از ۱۷
خانه‌دار	دبیرستانی و پایین‌تر	۳۷	۵۳	۳۸
	دانشگاهی	۲	۱۰	۱۵
شاغل	دبیرستانی و پایین‌تر	۳	۱۱	۱۵
	دانشگاهی	۵	۱۳	۲۹

جدول ۷. نتایج آزمون‌های استقلال شرطی (آزمون دقیق فیشر) در سطوح عوامل در مثال ۳.

آزمون	مقدار معنی‌داری
$A \perp B$	< 0.001
$A \perp B \mid C_1$	0.001
$A \perp B \mid C_2$	0.001
$A \perp B \mid C_3$	< 0.001
$A \perp C$	0.001
$A \perp C \mid B_1$	0.037
$A \perp C \mid B_2$	0.012
$B \perp C$	< 0.001
$B \perp C \mid A_1$	0.014
$B \perp C \mid A_2$	0.052

طبق نتایج جدول ۷ دو فرض استقلال $A \perp B$ و $A \perp C \mid B$ در سطح معنی‌داری 0.001 رد نمی‌شوند و فرض‌های استقلال شرطی را می‌پذیریم، لذا بنابر الگوریتم ۱ با پذیرفتن این دو استقلال شرطی فرض استقلال $C \perp (A, B)$ رد نخواهد شد. بدین معنی که در این مطالعه در سطح معنی‌داری 0.001 وضعیت تحصیلی فرزندان با شغل و تحصیلات مادر ارتباطی نداشته است. اما چون فرض استقلال $A \perp B \mid C$ در سطح معنی‌داری 0.001 را نمی‌توان پذیرفت، لذا در سطوح مختلف معدل فرزندان بین شغل و تحصیلات مادر ارتباط قوی وجود دارد. با توجه به عدم رد فرض $C \perp (A, B)$ ، ممکن است علاقه مند به آزمون برابری نسبت‌های مختلف سطوح معدل فرزندان باشیم که به راحتی با استفاده از آزمون برابری نسبت‌های کای دو با مقدار معنی‌داری 0.001 نتیجه می‌شود که سطوح معدل فرزندان یکسان نیست و تفاوت دارند. خروجی مدل‌های لگ خطی نشان داد که آخرین فرض پذیرفته شده $A \perp C \mid B$ است که متفاوت از نتیجه بدست آمده از روش پیشنهادی است. با توجه به جدول ۱، آخرین فرض پذیرفته شده از روش پیشنهادی مقاله یک مرحله نسبت مدل لگ خطی جلوتر رفته و مدل $C \perp (A, B)$ را آزموده است که عدم رد آن را نشان داده است. اما با استفاده از مدل لگ خطی این امکان وجود ندارد.

مثال ۴. (اگرستی، ۱۹۹۶) داده‌ها برگرفته از یک مطالعه عمومی در سال ۱۹۹۱ است که در آن سه سوال از ۴۲۳ فرد سفید پوست پرسیده شده است.

- ۱- آیا موافق جداسازی مدارس سیاه‌پوستان و سفیدپوستان هستید؟
- ۲- اگر دوست سیاه‌پوست شما برای ریاست جمهوری کاندیدا شده باشد و صلاحیت این شغل را داشته باشد، آیا به او رأی می‌دهید؟
- ۳- در طول چند سال گذشته، آیا اعضای خانواده شما دوست سیاه‌پوست خود را برای صرف شام به منزل دعوت

۱۱۶ رویکردی متفاوت در تحلیل جدول‌های پیش‌بینی سه طرفه

کرده‌اند؟

جواب سوال بصورت سه گزینه "بله"، "خیر" و "نظری ندارم" طراحی و نظرات ۴۲۳ نفر در جدول ۸ خلاصه شده است. با توجه به اینکه صفرهای جدول زیاد هستند، استفاده از مدل لگ خطی پیشنهاد نمی‌شود و لذا روش ارائه

جدول ۸. بررسی نظرات سفیدپوستان درخصوص سیاه‌پوستان.

دعوت به شام در منزل (C)			
ریاست جمهوری (A)	جداسازی مدارس (B)	بله	خیر
بله	بله	۴۱	۶۵
	خیر	۷۱	۱۵۷
	نظری ندارم	۱	۱۷
خیر	بله	۲	۵
	خیر	۳	۴۴
	نظری ندارم	۱	۰
نظری ندارم	بله	۰	۳
	خیر	۰	۱۰
	نظری ندارم	۰	۱

جدول ۹. نتایج آزمون‌های استقلال شرطی (آزمون دقیق فیشر) در سطوح عوامل در مثال ۴.

آزمون	مقدار معنی‌داری
$A \perp B$	۰/۰۳۲
$A \perp B \mid C_1$	۰/۱۲۱
$A \perp B \mid C_2$	۰/۰۰۹
$A \perp B \mid C_3$	۱/۰۰۰
$A \perp C$	< 0.001
$A \perp C \mid B_1$	۰/۰۳۷
$A \perp C \mid B_2$	< 0.001
$A \perp C \mid B_3$	۰/۰۱۶
$B \perp C$	۰/۰۰۹
$B \perp C \mid A_1$	۰/۰۲۴
$B \perp C \mid A_2$	۰/۰۱۴
$B \perp C \mid A_3$	۰/۰۳۸

شده در این مقاله، برای تحلیل جدول به کار می‌رود. طبق نتایج جدول ۹، تنها فرض‌های پذیرفته شده در جدول فوق

در سطح معنی‌داری ۰/۰۵، استقلال متغیرهای A و B در سطوح اول و سوم متغیر C را نشان می‌دهد، بنابراین کلیه آزمون‌های استقلال شرطی در این سطح معنی‌داری رد می‌شوند و در نتیجه ارتباط بین هر سه متغیر مشهود است، هرچند در سطح معنی‌داری ۰/۰۱۰ فرض استقلال $A \perp C \mid B$ پذیرفته شده است.

بحث و نتیجه‌گیری

تحلیل جدول‌های پیشابندی سه طرفه با استفاده از مدل‌های لگ خطی، نسبت بخت‌ها، رگرسیون لجستیک، آزمون دقیق فیشر و غیره امکان‌پذیر است. اکثر روش‌های تحلیل جدول‌های پیشابندی به صورت مجانبی انجام می‌شوند و برای اجرای آن‌ها نیاز به حجم نمونه بزرگ است که همواره مقدور نیست. ضمن اینکه آزمون‌های لگ خطی در مقابل مدل اشباع با آزمون‌های سلسله مراتبی که در نرم افزارهای آماری وجود دارند برخی اوقات ممکن است جواب متفاوت ارائه دهند و موجب سردرگمی پژوهشگر شوند. همچنین استفاده از مدل‌های لگ خطی در هنگام مواجهه با فراوانی صفر در جدول‌های پیشابندی کارایی لازم را ندارند. در این مقاله با معرفی یک الگوریتم ساده، برای تحلیل جدول‌های پیشابندی سه طرفه از آزمون دقیق فیشر و مفهوم استقلال‌های شرطی استفاده شد که جایگزینی برای مدل‌های لگ خطی در تحلیل این جدول‌هاست. در ادامه این کار می‌توان با استفاده از رویکرد ارائه شده در مقاله، جدول‌های پیشابندی مراتب بالاتر را تحلیل کرد، که از اهداف بعدی نویسندگان مقاله است.

تقدیر و تشکر

نویسندگان مقاله از سردبیر محترم، داوران ارجمند و ویراستار نکته‌سنج مجله که نظرات و پیشنهادات ارزشمند ایشان موجب بهبود مقاله گردید نهایت تشکر و قدردانی را دارند.

مراجع

- شاهکار، غ. (۱۳۸۷)، روش‌های چندمتغیره گسسته (رشته آمار)، انتشارات دانشگاه پیام نور، تهران.
- قریشی، ک. (۱۳۹۳)، چگالی‌های پیشین با ساختار معین برای تحلیل بیزی جدول‌های پیشابندی کامل و ناقص. مجله علوم آماری، ۸ (۲)، ۱۸۵-۲۰۵.
- Agresti, A. (1996), *An Introduction to Categorical Data Analysis*, Second Edition, John Wiley and Sons, New York.
- Agresti, A. (2013), *Categorical Data Analysis*, Third Edition, John Wiley and Sons, Hoboken, New Jersey.

- Agresti, A. (2019), *An Introduction to Categorical Data Analysis*, Third Edition, John Wiley and Sons, Hoboken, New Jersey.
- Andersen, E. B. (1994), *The Statistical Analysis of Categorical Data*, Springer-Verlag, Berlin Heidelberg.
- Andersen, E. B. (1997), *Introduction to the Statistical Analysis of Categorical Data*, Springer-Verlag, Berlin Heidelberg.
- Birch, M. W. (1963), Maximum Likelihood in Three-Way Contingency Tables, *Journal of the Royal Statistical Society, Series B*, **25**, 220–233.
- D’Agostino, R. B., Chase, W. and Belanger, A. (1988), The Appropriateness of Some Common Procedures for Testing the Equality of Two Independent Binomial Populations, *The American Statistician*, **42**, 198–202.
- Dawid, A. P., (1998), Conditional Independence, In: Kotz, S., Read, C.B., Banks, D.L. (Eds.), *Encyclopedia of Statistical Sciences*, Update, vol. 2. Wiley Interscience, pp. 146–155.
- Ehwerhemuepha, L., Sok, H. and Rakovski, C. (2019), A More Powerful Unconditional Exact Test of Homogeneity for $2 \times c$ Contingency Table Analysis, *Journal of Applied Statistics*, **46**(14), 2572–2582.
- Loperfido, N. (2010), A Note on Marginal and Conditional Independence, *Statistics and Probability Letters*, **80**, 1695–1699.
- Oliveira, N. L., Pereira, C. A. D. B., Diniz, M. A. and Polpo, A. (2018), A Discussion on Significance Indices for Contingency Tables under Small Sample Sizes, *PLoS ONE*, **13**(8): e0199102. <https://doi.org/10.1371/journal.pone.0199102>.