مجلّه‌ی پژوهش‌های آماری ایران
پژوهشکده‌ی آمار
دوره‌ی ۵، شماره‌ی ۱، بهار و تابستان ۱۳۸۷، صص ۱۹–۳۱

مؤلفه‌های پیوند جزئی در جدول‌های پیشایندی چندبعدی و تحلیل آماری آن‌ها

سید کامران قریشی[†,*] و محمدرضا مشکانی[‡]

[†] دانشگاه رازی
[‡] دانشگاه شهید بهشتی

چکیده. نتایج عرضه‌شده در این مقاله ملهم از ایده‌ی گودمن و کروسکال (۱۹۵۴) هستند که: «زمانی که بیش از دو متغیر رسته‌ای وجود داشته باشند طبیعی خواهد بود که به پیوند جزئی بین دو متغیر به معنایی با متوسط گرفتن از اثر بقیه‌ی متغیرها بیندیشیم». در این مقاله با استفاده از مدل پیوند $\mathrm{RC}(M_۱, M_۲)$ مؤلفه‌های (چندجمله‌ای متعامد) پیوند جزئی بین هر جفت از متغیر رسته‌ای سازنده‌ی جدول پیشایندی چندبعدی را تعریف می‌کنیم. چون این کمّیت‌ها اغلب مجهول‌اند، روش‌های استنباط آماری پارامترهای مذکور را از دیدگاه بیزی و فراوانی‌گرا و توزیع‌های مجانبی آن‌ها را بررسی می‌کنیم. همچنین با کاربست مقتضی آن‌ها نشان می‌دهیم که نتایج عرضه‌شده برای هر مؤلفه‌ی متعامد دلخواه (غیر چندجمله‌ای متعامد) نیز صادق‌اند. در پایان، نتایج را برای تحلیل دو جدول سه‌بعدی از داده‌های واقعی به کار می‌بندیم.

واژگان کلیدی. پیوند جزئی و چندگانه؛ روندهای چندجمله‌ای متعامد؛ مدل‌های پیوند؛ معیارهای پیوند.

۱ مقدمه

در تحلیل جدول‌های پیشایندی، که در زمینه‌های تحقیقاتی اعم از پزشکی، جامعه‌شناسی و ... به دست می‌آیند، مطالعه‌ی رابطه‌ی بین متغیرهای سازنده‌ی جدول از اهمیت ویژه‌ای برخوردار است. برای این مهم

[*] نویسنده‌ی عهده‌دار مکاتبات

از معیارها و مدل‌هایی استفاده می‌شود که معیارهای پیوند و مدل‌های پیوند نامیده می‌شوند. مدل‌های پیوند، که معرف اندازه‌ی کلی پیوندند، اجزاء پیوند بین متغیرهای رسته‌ای را بیان می‌کنند. در اندازه‌گیری کلی پیوند فقط به شدّت و ضعف پیوند توجه می‌شود، در صورتی که در اندازه‌گیری اجزاء پیوند علاوه بر تعیین مؤثر بودن یک عامل، میزان آن نیز اندازه‌گیری می‌شود. اگر دو متغیر سازنده‌ی یک جدول پیشایندی دو بعدی، سطح سواد و درآمد باشند. جالب خواهد بود که تخمینی از الگوی تأثیر میزان سواد بر درآمد وجود داشته باشد.

انواع معیارهای پیوند در گودمن و کروسکال (۱۹۵۴) و گودمن (۱۹۷۹) و لاییترا (۱۹۸۳) و مدل‌های پیوند را در گودمن (۱۹۷۲، ۱۹۷۹، ۱۹۸۵، ۱۹۹۱) و قریشی و مشکانی (۲۰۰۶، ۲۰۰۸) می‌توان یافت.

در این مقاله به کمک مدل پیوند $\mathrm{RC}(M_۱, M_۲)$، قریشی و مشکانی (۲۰۰۶)، مؤلفه‌های پیوند جزئی بین دو متغیر رسته‌ای (در حالی که بقیّه‌ی متغیرها، متغیرهای کنترلی فرض می‌شوند) را به مؤلفه‌های چندجمله‌ای متعامد تجزیه می‌کنیم. نقش و سهم هر مؤلفه‌ی جزئی، کمک فراوانی به محققان در تحلیل جدول‌های پیشایندی با ابعاد بزرگ‌تر از ۲ خواهد کرد. اجازه دهید تا بحث را با تعریف مؤلفه‌های پیوند جزئی پیش بریم.

تعریف ۱ متغیرهای رسته‌ای $X_۱$، $X_۲$، …، X_r، $(r \geqslant ۳)$، را که به ترتیب دارای $I_۱$، $I_۲$، …، I_r سطح و در آن‌ها حد اقل دو متغیر $X_۱$ و $X_۲$ از نوع ترتیبی باشند، در نظر بگیرید. فرض کنید

$$\pi_{i_۱ i_۲ i_۳ i_۴ \cdots i_r} = \Pr(X_۱ = i_۱, X_۲ = i_۲, \ldots, X_r = i_r)$$

و

$$\pi_{..i_۳ i_۴ \cdots i_r} = \Pr(X_۳ = i_۳, X_۴ = i_۴, \ldots, X_r = i_r).$$

اگر مؤلفه‌های پیوند بین دو متغیر رسته‌ای ترتیبی $X_۱$ و $X_۲$ را برای $I_۳ \times I_۴ \times \cdots \times I_r$ جدول جزئی، در سطوح دیگر متغیرها و بر اساس احتمال‌های شرطی

$$\pi_{i_۱ i_۲ / i_۳ i_۴ \cdots i_r} = \frac{\pi_{i_۱ i_۲ i_۳ i_۴ \cdots i_r}}{\pi_{..i_۳ i_۴ \cdots i_r}}$$

محاسبه کنیم میانگین موزون هر مؤلفه‌ی محاسبه‌شده با وزن‌های $\pi_{..i_۳ i_۴ \cdots i_r}$ را مؤلفه‌ی پیوند جزئی بین $X_۱$ و $X_۲$ و متناظر با آن مؤلفه می‌نامیم.

با این تعریف ملاحظه می‌کنیم که تعداد مؤلفه‌های پیوند جزئی بین دو متغیر رسته‌ای ترتیبی $X_۱$ و $X_۲$

برابر $(I_{۲}-۱)(I_{۱}-۱)$ خواهد بود. اگرچه اصطلاح مؤلفه‌های پیوند ممکن است تعاریف گوناگونی داشته باشد، لیکن به‌دلیل تفسیرپذیر بودن مؤلفه‌های پیوند چندجمله‌ای متعامد استفاده از آن‌ها در عمل جذابیت بیش‌تری دارد.

در بخش ۲، ابتدا در مورد مؤلفه‌های پیوند در جدول‌های جزئی دوبعدی بحث می‌کنیم. سپس مؤلفه‌های پیوند جزئی را تعریف خواهیم کرد. بخش ۳ به تحلیل فراوانی‌گرای معیارهای پیوند جزئی اختصاص دارد. تحلیل بیزی مؤلفه‌های پیوند جزئی در بخش ۴ گنجانده شده است. در بخش ۵ یافته‌های این مقاله را به کمک دو مثال واقعی بررسی خواهیم کرد. بخش ۶ بخش نهایی این مقاله بوده و در آن به بحث و نتیجه‌گیری می‌پردازیم.

۲ مؤلفه‌های پیوند جزئی

اجازه دهید تا بدون کاستن از کلیّت مطلب بحث را با سه متغیر رسته‌ای X (با I سطح)، Y (با J سطح)، و Z (با L سطح)، پیش بریم. فرض کنید حد اقل متغیرهای X و Y از نوع ترتیبی بوده و ما علاقمند به پیوند جزئی بین این دو متغیر هستیم. برای این منظور، ابتدا مؤلفه‌های پیوند بین متغیرهای رسته‌ای X و Y را در جدول‌های جزئی دوبعدی (X,Y) به‌ازای سطوح متغیر رسته‌ای Z به دست آورده، سپس مؤلفه‌های موزون (در سطوح مختلف متغیر Z) را محاسبه می‌کنیم. مؤلفه‌های حاصل، مؤلفه‌های پیوند جزئی خواهند بود که به‌عنوان شاخص‌هایی در تحلیل جدول‌های پیشایندی به کار می‌روند.

احتمال تعلق گرفتن یک مشاهده به خانه‌ی (i,j,l) جدول پیشایندی را با π_{ijl} و بسامدهای مورد انتظار خانه‌ی مذکور را با $m_{ijl}=n\pi_{ijl}$ نمایش می‌دهیم که در آن n مجموع بسامدهای مشاهده‌شده یا اندازه‌ی نمونه است. همچنین احتمال شرطی $\pi_{ij|l}=\frac{\pi_{ijl}}{\pi_{..l}}$ را احتمال تعلق یك نمونه به خانه‌ی (i,j) از جدول جزئی $Z=l$ و مشروط به آن تعریف می‌کنیم. بر اساس این نمادگذاری، مدل پیوند موزون $\mathrm{RC}(M_{۱},M_{۲})$ را در جدول جزئی $Z=l$ به‌صورت زیر در نظر می‌گیریم:

$$\ln m_{ijl}=\lambda_{.(l)}+\lambda^X_{i(l)}+\lambda^Y_{j(l)}+\sum_{k=۱}^{I-۱}\sum_{k'=۱}^{J-۱}\phi^{(l)}_{kk'}\mu_{ik}\nu_{jk'}. \qquad (۱)$$

قیدهای این مدل عبارت‌اند از

$$\sum_i w_i\mu_{ik}=\circ,\quad \sum_i w_i\mu^{۲}_{ik}=۱,\quad \sum_j w'_j\nu_{jk'}=\circ,\quad \sum_j w'_j\nu^{۲}_{jk'}=۱,$$

$$\sum_i w_i \mu_{ik}\mu_{ik'} = ۰, \quad \sum_j w'_j \nu_{jk}\nu_{jk'} = ۰, \quad \sum_i w_i = ۱, \quad \sum_j w'_j = ۱.$$

در این جا کمّیت‌های μ_{ik} و $\nu_{ik'}$ ضرایب مقایسه‌ی چندجمله‌ای‌های متعامد و متناظر با مؤلفه‌های روند چندجمله‌ای kام و k'ام می‌باشد. w_i و w'_j وزن‌های معلوم منتسب‌شده به سطح iام متغیر رسته‌ای X و jام متغیر رسته‌ای Y و در نهایت $\phi^{(l)}_{kk'}$ پارامتر پیوند ذاتی در جدول جزئی lام می‌باشد. در ادامه خواهیم دید که هرگاه وزن‌های w_i و w'_j مساوی یک باشند تحلیل موزون پیوند جزئی به تحلیل غیر موزون تبدیل می‌شود. برای اطلاع از چگونگی تحلیل مدل‌های پیوند موزون $\mathrm{RC}(M)$ می‌توان به گودمن (۱۹۹۱) مراجعه کرد. برای شباهت بیش‌تر مدل (۱) به مدل لگ-خطی قرار می‌دهیم

$$\lambda^{XY}_{ij(l)} = \sum_{k=۱}^{I-۱}\sum_{k'=۱}^{J-۱} \phi^{(l)}_{kk'}\mu_{ik}\nu_{jk'}. \qquad (۲)$$

با استفاده از روابط (۱) و (۲) و قیدهای مربوط، مجموع توان دوم موزون انحراف‌ها (SSD) برای $\lambda^{XY}_{ij(l)}$ برابر با

$$\mathrm{WSS}_{xy|Z=l} = \sum_{i=۱}^{I}\sum_{j=۱}^{J} w_i w'_j (\lambda^{XY}_{ij(l)})^۲ = \sum_{k=۱}^{I-۱}\sum_{k'=۱}^{J-۱} (\phi^{(l)}_{kk'})^۲,$$

و مجموع توان دوم موزون انحراف‌های مؤلفه‌ی $k \times k'$ام یعنی $\mathrm{WSS}_{xy|Z=l,k\times k'}$ برابر با

$$\mathrm{WSS}_{xy|Z=l,k\times k'} = (\phi^{(l)}_{kk'})^۲, \qquad k = ۱, \cdots, I-۱, \quad k' = ۱, \cdots, J-۱$$

خواهند بود، (قریشی و مشکانی، ۲۰۰۶ و ۲۰۰۸).

همچنین از رابطه‌ی (۲) و قیدهای متناظر، پیوند ذاتی $k \times k'$امین مؤلفه‌ی پیوند بین (X, Y)، در جدول جزئی $Z = l$، به‌صورت زیر به دست می‌آید:

$$\phi^{(l)}_{kk'} = \sum_{i=۱}^{I}\sum_{j=۱}^{J} w_i w'_j \mu_{ik}\nu_{jk'} \ln m_{ijl}. \qquad (۳)$$

به‌راحتی می‌توان تحقیق کرد که مساوی صفر بودن همزمان همه‌ی این کمّیت‌ها شرط لازم و کافی برای استقلال جزئی بین X و Y است. بر اساس رابطه‌ی (۳)، $k \times k'$امین مؤلفه‌ی پیوند جزئی را، که برابر میانگین موزون $\phi^{(l)}_{kk'}$ها است، به‌صورت

$$\psi_{kk'} = \sum_{i=1}^{L} \pi_{..l}\phi_{kk'}^{(l)} \qquad (4)$$

تعریف می‌کنیم. اگر توان دوم این کمّیت‌ها را به‌عنوان تعریف غیر دقیق از SSD مؤلفه‌ی $k \times k'$ام در نظر بگیریم، سهم $k \times k'$امین مؤلفه‌ی پیوند جزئی بین متغیرهای X و Y برابر با

$$\kappa_{kk'}^{2} = \frac{\psi_{kk'}^{2}}{\sum_{u=1}^{I-1}\sum_{u'=1}^{J-1}(\psi_{uu'})^{2}}, \quad k = 1,\cdots,I-1,\ k' = 1,\cdots,J-1 \qquad (5)$$

به دست خواهد آمد. به‌وضوح می‌بینیم که کمّیت‌های $\kappa_{kk'}^{2}$ اعدادی در بازه‌ی $[0,1]$ بوده، شدّت پیوند $k \times k'$امین مؤلفه را اندازه‌گیری می‌کنند. چون در عمل اغلب جهت پیوند مهم است، معیار پیوند جزئی $k \times k'$امین مؤلفه‌ی پیوند را به‌صورت زیر تعریف می‌کنیم که عددی در بازه‌ی $[-1,1]$ است.

$$\kappa_{kk'} = \frac{\psi_{kk'}}{\sqrt{\sum_{u=1}^{I-1}\sum_{u'=1}^{J-1}(\psi_{uu'})^{2}}}, \quad k = 1,\cdots,I-1,\ k' = 1,\cdots,J-1. \qquad (6)$$

بر اساس کمّیت‌های $\psi_{kk'}$ یا $\kappa_{kk'}$ می‌توان مؤلفه‌های چندجمله‌ای مؤثر در پیوند جزئی را شناسایی کرد و برای پرسش‌های گوناگون در خصوص ساختار پیوند جزئی بین X و Y پاسخ‌های مناسب فراهم نمود.

لازم است به این نکته اشاره کنیم که تعریف ارائه‌شده از پیوند جزئی در بالا کاملاً متفاوت از تعریفی است که در مقاله‌ی قریشی و مشکانی (۲۰۰۹) ارائه شده است.

همان‌طور که می‌دانیم ضرایب مقایسه‌ی متعامد چندجمله‌ای، حالتی خاص از ضرایب مقایسه‌ی متعامدند. یعنی تجزیه‌ی (۲) برای هر $I-1$ مقایسه‌ی متعامد سطری و هر $J-1$ مقایسه‌ی متعامد ستونی برقرار است. با این وجود مقدار WSS_{xy} همواره ثابت است. این نتیجه به این معناست که همواره می‌توان سهم هر مؤلفه‌ی دلخواه سطری $(\mu_1^*,\cdots,\mu_I^*)$ و هر مؤلفه‌ی دلخواه ستونی $(\nu_1^*,\cdots,\nu_J^*)$از ساختار پیوند را تعیین کرد و در صورت نیاز آزمون‌هایی در خصوص آن‌ها انجام داد. معیار پیوند جزئی برای دو مؤلفه‌ی نام‌برده عبارت است از

$$\kappa^{*} = \frac{\psi^{*}}{\sqrt{\sum_{u=1}^{I-1}\sum_{u'=1}^{J-1}(\psi_{uu'})^{2}}}, \qquad (7)$$

که در آن

$$\psi^* = \sum_{l=1}^{K} \pi_{..l} \left(\sum_{i=1}^{I} \sum_{j=1}^{J} w_i w'_j \mu_i^* \nu_j^* \ln m_{ijl} \right), \tag{8}$$

و $\psi_{kk'}$ با رابطه‌ی (۴) داده می‌شود.

۳ تحلیل فراوانی‌گرای معیارهای پیوند جزئی

در این بخش برآوردهای ماکسیمم درست‌نمایی پارامترهای $\psi_{kk'}$ و $\kappa_{kk'}$ و توزیع مجانبی آن‌ها و در نتیجه بازه‌های اطمینان متناظر را ارائه خواهیم کرد. به‌دلیل خاصیت ناوردایی برآورد ماکسیمم درست‌نمایی با جای‌گذاری بسامدهای مشاهده‌شده‌ی جدول پیشایندی n_{ijl} به‌جای بسامدهای مورد انتظار $m_{ijl} = n\pi_{ijl}$، اگرستی (۲۰۰۲)، می‌توان برآوردهای ماکسیمم درست‌نمایی معیارهای مذکور را به دست آورد. یعنی:

$$\hat{\phi}_{kk'}^{(l)} = \sum_{i=1}^{I} \sum_{j=1}^{J} w_i w'_j \mu_{ik} \nu_{jk'} \ln n_{ijl}, \tag{9}$$

$$\hat{\psi}_{kk'} = \sum_{l=1}^{L} \frac{n_{..l}}{n} \hat{\phi}_{kk'}^{(l)}, \tag{10}$$

$$\hat{\kappa}_{kk'} = \frac{\hat{\psi}_{kk'}}{\sqrt{\sum_{u=1}^{I-1} \sum_{u'=1}^{J-1} (\hat{\psi}_{uu'})^2}}. \tag{11}$$

همان‌طور که می‌دانیم برآوردهای ماکسیمم درست‌نمایی دارای توزیع مجانبی نرمال‌اند. لذا توزیع مجانبی $\hat{\psi}_{kk'}$ و $\hat{\kappa}_{kk'}$ را در غالب دو گزاره‌ی زیر ارائه می‌کنیم. برای این گزاره‌ها فرض می‌کنیم n_{ijl} نمونه‌ای تصادفی از توزیع چندجمله‌ای با اندازه‌ی نمونه‌ی n و احتمال‌های π_{ijl} $(i = 1, \ldots, I; j = 1, \ldots, J; l = 1, \ldots, L)$ باشند. اثبات این گزاره‌ها با استفاده از روش‌دلتا، پس از مشتق‌گیری‌های لازم و ساده کردن عبارت‌ها، به دست می‌آیند که به‌دلیل وضوح مطلب از ارائه‌ی جزئیات بیش‌تر خوداری می‌کنیم.

گزاره‌ی ۱ توزیع مجانبی $\hat{\psi}_{kk'}$ تعریف‌شده با رابطه‌ی (۱۰) عبارت است از

$$\hat{\psi}_{kk'} \sim N(\psi_{kk'}, \sigma^{۲}_{\hat{\psi}_{kk'}})$$

که در آن

$$\sigma^{۲}_{\hat{\psi}_{kk'}} = \sum_{i=۱}^{I}\sum_{j=۱}^{J}\sum_{l=۱}^{L}\pi_{ijl}\left(\phi^{(l)}_{kk'} - \frac{\pi_{..l}}{\pi_{ijl}}w_i w'_j \mu_{ik}\nu_{jk'}\right)^{۲} - \left\{\sum_{i=۱}^{I}\sum_{j=۱}^{J}\sum_{l=۱}^{L}\pi_{ijl}\left(\phi^{(l)}_{kk'} - \frac{\pi_{..l}}{\pi_{ijl}}w_i w'_j \mu_{ik}\nu_{jk'}\right)\right\}^{۲}. \tag{۱۲}$$

گزاره‌ی ۲ توزیع مجانبی $\hat{\kappa}_{kk'}$ تعریف‌شده با رابطه‌ی (۱۱) عبارت است از:

$$\hat{\kappa}_{kk'} \sim N(\kappa_{kk'}, \sigma^{۲}_{\hat{\kappa}_{kk'}}).$$

که در آن

$$\sigma^{۲}_{\hat{\kappa}_{kk'}} = \sum_{i=۱}^{I}\sum_{j=۱}^{J}\sum_{l=۱}^{L}\pi_{ijl}q^{۲}_{ijl} - \left(\sum_{i=۱}^{I}\sum_{j=۱}^{J}\sum_{l=۱}^{L}\pi_{ijl}q_{ijl}\right)^{۲} \tag{۱۳}$$

$$q_{ijl} = \frac{\frac{\partial\psi_{kk'}}{\partial\pi_{ijl}} - \kappa_{kk'}\sum_{u=۱}^{I-۱}\sum_{u'=۱}^{J-۱}\frac{\partial\psi_{uu'}}{\partial\pi_{ijl}}\kappa_{uu'}}{\sqrt{\sum_{u=۱}^{I-۱}\sum_{u'=۱}^{J-۱}\psi^{۲}_{uu'}}}.$$

با استفاده از روابط (۱۱) و (۱۳)، یک بازه‌ی اطمینان $(۱-\alpha)$ درصد برای $\kappa_{kk'}$ با رابطه‌ی زیر داده می‌شود:

$$\kappa_{kk'} : \hat{\kappa}_{kk'} \pm z_{\frac{\alpha}{۲}}\hat{\sigma}_{\hat{\kappa}_{kk'}}.$$

برآورد ماکسیمم درست‌نمایی κ^*، در رابطه‌ی (۷)، برابر است با $\hat{\kappa}^*$، که با جای‌گذاری بسامدهای مشاهده‌شده به‌جای بسامدهای مورد انتظار به دست می‌آید. با توجه به گزاره‌ی ۲، توزیع مجانبی $\hat{\kappa}^*$ با لم ۱ داده می‌شود.

لم ۱ تحت شرایط لازم برای گزاره‌ی ۲، توزیع مجانبی $\hat{\kappa}^*$ نرمال با میانگین κ^* و واریانس

$$\sigma^{۲}_{\hat{\kappa}_{kk'}} = \sum_{i=۱}^{I}\sum_{j=۱}^{J}\sum_{l=۱}^{L}\pi_{ijl}q^{۲}_{ijl} - \left(\sum_{i=۱}^{I}\sum_{j=۱}^{J}\sum_{l=۱}^{L}\pi_{ijl}q_{ijl}\right)^{۲}$$

است که در آن

$$q^*_{ijl} = \frac{\frac{\partial\psi^*}{\partial\pi_{ijl}} - \kappa^*\sum_{u=۱}^{I-۱}\sum_{u'=۱}^{J-۱}\frac{\partial\psi_{uu'}}{\partial\pi_{ijl}}\kappa_{uu'}}{\sqrt{\sum_{u=۱}^{I-۱}\sum_{u'=۱}^{J-۱}\psi^{۲}_{uu'}}}.$$

با توجه به لم ۱ یک بازه‌ی اطمینان مجانبی $(۱-\alpha)$ درصد برای κ^* با در نظر گرفتن برآوردهای $\hat{\kappa}^*$ و $\hat{\sigma}^{۲}_{\hat{\kappa}^*}$ عبارت است از

$$\kappa^* : \hat{\kappa}^* \pm z_{\frac{\alpha}{۲}}\hat{\sigma}_{\hat{\kappa}^*}.$$

۴ تحلیل بیزی مؤلفه‌های پیوند جزئی

در تحلیل بیزی مؤلفه‌های جزئی از چگالی‌های پیشینی متفاوت برای پارامترهای مدل پیوند می‌توان استفاده کرد. این چگالی‌ها ممکن است برای پارامترهای مدل پیوند یا بسامدهای مورد انتظار خانه‌های جدول مورد استفاده قرار گیرند. در هر صورت، محصول نهایی به کارگیری هر چگالی پیشینی چگالی پسینی بسامدهای مورد انتظار m_{ijl} خواهد بود. با توجه به این‌که مدل پیوند (۱) یک مدل اشباع‌شده است، با در دست بودن بسامدهای مورد انتظار، m_{ijl}، پارامترهای پیوند ذاتی $\phi^{(l)}_{kk'}$ به‌طور یکتا به دست می‌آیند. بنا بر این، به کمک چگالی پسینی بسامدهای مورد انتظار می‌توان چگالی پسینی $\phi^{(l)}_{kk'}$ را به دست آورد. در حقیقت چگالی حاصل، چگالی انواع روندهای جزئی خواهد بود. اگر چه چگالی پسینی $\phi^{(l)}_{kk'}$ برای هرگونه مطالعه‌ی بیزی به کار می‌رود لیکن، در بسیاری از کاربردهای عملی میانگین و واریانس پسینی مربوط به آن کافی خواهد بود. روش محاسبه‌ی عددی این دو کمیّت به‌صورت زیر قابل انجام است.

فرض کنیم، $m^{(t)}_{ijl}$، $t=۱,\ldots,N$، نمونه‌ی تصادفی حاصل از چگالی پسینی به اندازه‌ی N باشد. در آن صورت طبق روابط

$$\phi^{(l)(t)}_{kk'} = \sum_{i=۱}^{I}\sum_{j=۱}^{J}w_iw'_j\mu_{ik}\nu_{jk'}\ln m^{(t)}_{ijl},$$

جدول ۱. بسامدهای انواع آسیب‌ها در بین ۳۶۶ نفر به‌تفکیک سن و جنسیت

سن (سال)	جنسیت	انواع آسیب‌ها H.T.	M.T.	O.T.
۰–۲۰	مرد	۵۸	۱۳	۷۱
	زن	۱۸	۴	۵
۲۰–۴۰	مرد	۵۱	۳۲	۱۳
	زن	۱۵	۵	۴
۴۰ <	مرد	۱۳	۶	۱۹
	زن	۱۰	۲۰	۹

$$\psi_{kk'}^{(t)} = \sum_{i=1}^{L} \frac{m_{..l}^{(t)}}{m_{...}^{(t)}} \phi_{kk'}^{(l)(t)},$$

و

$$\kappa_{kk'}^{(t)} = \frac{\psi_{kk'}^{(t)}}{\sqrt{\sum_{u=1}^{I-1}\sum_{u'=1}^{J-1}(\psi_{uu'}^{(t)})^2}},$$

مقادیر $\phi_{kk'}^{(l)(t)}$ و $\kappa_{kk'}^{(t)}$ نمونه‌ی تصادفی حاصل از چگالی پسینی متناظر خواهند بود. با استفاده از قانون ضعیف اعداد بزرگ، برای N به اندازه‌ی کافی بزرگ، برآورد میانگین و واریانس پسینی $\kappa_{kk'}$ به‌صورت زیر به دست می‌آید:

$$E(\kappa_{kk'}|n) \simeq \bar{\kappa}_{kk'} = \frac{1}{N}\sum_{t=1}^{N}\kappa_{kk'}^{(t)}. \tag{۱۴}$$

$$\mathrm{var}(\kappa_{kk'}|n) \simeq S_{\kappa_{kk'}}^2 = \frac{1}{N-1}\sum_{t=1}^{N}\left(\kappa_{kk'}^{(t)} - \bar{\kappa}_{kk'}\right)^2. \tag{۱۵}$$

چون برآوردهای پیوند ذاتی تابعی از برآوردهای ماکسیمم درست‌نمایی هستند، بنا بر این، برآوردهای بیزی $\kappa_{kk'}$ دارای توزیع مجانبی نرمال‌اند. بنا بر این، با استفاده از آماره‌های $\bar{\kappa}_{kk'}/\sqrt{S_{\kappa_{kk'}}^2}$، به‌عنوان آماره‌ی آزمون، و محاسبه‌ی کمّیت‌های مذکور به روش بیزی می‌توان روندهای مؤثر در پیوند را تعیین نمود.

جدول ۲. بسامدهای مشاهده‌شده از ۵۱۴۴ نفر به‌تفکیک سه متغیر مصرف روزانه‌ی الکل، درآمد و محل سکونت

مصرف روزانه‌ی الکل (واحد مصرف)	درآمد (دلار)	محل سکونت ۰	۱	۲	۳	۴
< ۱	$۰ - ۵۰$	۲۱	۱۳	۴۴	۱۱۴	۱۹۰
	$۵۰ - ۱۰۰$	۵۷	۵۸	۷۸	۲۶۸	۲۸۷
	$۱۰۰ - ۱۵۰$	۲۶	۳۵	۲۵	۱۰۴	۸۳
	$۱۵۰ <$	۹۹	۸۳	۸۲	۳۲۶	۳۴۶
$۱ - ۲$	$۰ - ۵۰$	۲۰	۱۰	۲۵	۳۸	۵۷
	$۵۰ - ۱۰۰$	۵۲	۶۳	۸۲	۲۰۰	۱۷۰
	$۱۰۰ - ۱۵۰$	۴۳	۸۲	۴۹	۱۵۸	۱۰۵
	$۱۵۰ <$	۱۰۵	۱۳۳	۱۳۲	۳۳۵	۲۲۴
$۲ <$	$۰ - ۵۰$	۶	۲	۵	۱۳	۸
	$۵۰ - ۱۰۰$	۱۹	۲۱	۱۹	۵۱	۵۱
	$۱۰۰ - ۱۵۰$	۲۲	۳۴	۱۴	۴۶	۲۸
	$۱۵۰ <$	۸۲	۸۰	۳۹	۱۱۲	۷۰

۵ کاربرد

در این بخش نتایج حاصل از گزاره‌های ۱ و ۲ را برای دو مجموعه داده‌ی واقعی تحقیق خواهیم کرد. جدول ۱، داده‌های حاصل از بررسی میزان آسیب‌ها در حوادث غیر مترقبه در فصل اول سال ۱۳۸۵ است که بر اساس یک نمونه‌ی ۳۶۶تایی از اورژانس بیمارستان طالقانی کرمانشاه به دست آمده است. در این جدول انواع آسیب‌های غیر مترقبه، در یک تقسیم‌بندی کلی، به سه دسته‌ی آسیب‌های ناحیه سر و گردن (H.T.)، آسیب‌های چندگانه‌ی بدن (M.T.)، و آسیب‌های دیگر (O.T.) تقسیم شده‌اند. علاوه بر این، سن و جنس مصدومین نیز درج شده است. در این‌جا با فرض وجود ترتیب بین سطوح متغیر رسته‌ای انواع آسیب‌ها، به این معنی که آسیب‌های (H.T.) مهم‌ترین و آسیب‌های (O.T.) کم‌اهمیت‌ترین آسیب‌ها هستند، علاقمندیم مؤلفه‌های پیوند جزئی بین متغیر مذکور و سن مصدوم را برای دو جنس مختلف به دست آوریم.

جدول ۲، (اندرسن ۱۹۹۱)، شامل اطلاعات ۵۱۴۴ فرد است که بر اساس میزان مصرف الکل، درآمد و محل سکونت بنا شده است. تشخیص و تحلیل مؤلفه‌های پیوند جزئی بین متغیرهای میزان مصرف روزانه‌ی الکل و درآمد در سطوح مختلف محل سکونت ممکن است ابعاد گوناگونی از پیوند بین این متغیرها را در اختیار ما قرار دهد.

جدول ۳. برآوردهای بیزی سهم‌های جزئی همراه با انحراف معیارهای آن‌ها برای داده‌های جدول ۲

روندها	$L\times L$	$L\times Q$	$L\times C$	$Q\times L$	$Q\times Q$	$Q\times C$
$k_{kk'}$	۰٫۸۱۲	−۰٫۴۰۰	−۰٫۱۹۹	−۰٫۱۷۹	۰٫۲۴۰	۰٫۱۴۴
S.E.$(\hat{k}_{kk'})$	(۰٫۰۴۳)	(۰٫۰۶۵)	(۰٫۰۸۲)	(۰٫۰۸۸)	(۰٫۰۷۸)	(۰٫۰۶۴)
مقدار $- P$	۰٫۰۰۰	۰٫۰۰۰	۰٫۰۱۵	۰٫۰۴۲	۰٫۰۰۲	۰٫۰۲۴

جدول ۴. برآوردهای بیزی سهم‌های جزئی همراه با انحراف معیارهای آن‌ها برای داده‌های جدول ۱

روندها	$L\times L$	$L\times Q$	$Q\times L$	$Q\times Q$
$k_{kk'}$	۰٫۱۷۲	−۰٫۴۱۴	۰٫۶۴۶	۰٫۵۴۵
S.E.$(\hat{k}_{kk'})$	(۰٫۱۳۷)	(۰٫۱۸۲)	(۰٫۱۰۴)	(۰٫۱۴۴)
مقدار $- P$	۰٫۲۰۹	۰٫۰۲۳	۰٫۰۰۰	۰٫۰۰۰

همان‌طور که در بالا توضیح داده شد هدف از تحلیل داده‌های جدول ۱ یافتن الگویی برای آسیب به‌عنوان تابعی از سن با در نظر گرفتن اثر جنس مصدوم و نیز هدف از تحلیل داده‌های جدول ۲ یافتن الگوی مصرف روزانه‌ی الکل به‌عنوان تابعی از درآمد با در نظر گرفتن اثر محل سکونت است.

برای سادگی نمایش انواع روند (چندجمله‌ای‌های متعامد) از نمادهای: روند خطی $= L$، روند درجه‌ی دوم $= Q$، و درجه‌ی سوم $= C$، استفاده می‌کنیم. فرض کنید وزن‌های منتسب‌شده به سطرها و ستون‌ها مساوی و برابر یک است. همچنین احتمال خانه‌های جدول سه‌بعدی از نوع توزیع چندجمله‌ای و چگالی پیشینی دیریکله با ابرپارامترهای مساوی یک (یا چگالی یکنواخت) باشد. با $N = ۱۰۰۰$ در رابطه‌های (۱۴) و (۱۵) میانگین و انحراف معیار روندهای جزئی $\kappa_{kk'}$ مربوط به جدول‌های ۱ و ۲ به‌ترتیب در جدول‌های ۳ و ۴ داده شده است.

جدول ۳ نشان می‌دهد که هر مؤلفه‌ی جزئی بین میزان مصرف روزانه‌ی الکل (X) و درآمد Y در سطوح مختلف محل سکونت تا چه اندازه در ساختار پیوند بین این دو متغیر رسته‌ای مؤثر است. از بین همه‌ی مؤلفه‌ها، مؤلفه‌ی «خطی در خطی» $(L\times L)$ بیش‌ترین سهم ۰٫۸۱۲ را داراست. این به این معناست که با افزایش خطی در درآمد، مصرف روزانه‌ی الکل به‌صورت تابعی خطی از درآمد افزایش می‌یابد.

اگرچه تمام مؤلفه‌های جزئی تا حدودی معنادار هستند ولی در سطح معنی‌داری ۰٫۰۱ سه مؤلفه‌ی «خطی در درجه‌ی سوم» $(L\times C)$، «درجه‌ی دوم در خطی» $(Q\times L)$ و «درجه‌ی دوم در درجه‌ی سوم» $(Q\times C)$ را می‌توان از ساختار پیوند حذف کرد.

جدول ۴ نشان می‌دهد که هر مؤلفه‌ی جزئی بین مصدومیت X و انواع آسیب‌ها Y برای زنان و مردان تا چه اندازه در ساختار پیوند بین این دو متغیر رسته‌ای مؤثر است. از بین همه‌ی مؤلفه‌ها، مؤلفه‌ی «درجه‌ی دوم در خطی» $(Q \times L)$ بیش‌ترین سهم ۰٫۶۴۶ را داراست. این نتیجه به این معنی است که در هر سطح از سن، نوع مصدومیت به‌صورت خطی تغییر می‌کند در حالی که برای هر سطح مصدومیت، سن با الگوی درجه‌ی دوم تغییر می‌کند. توجه داریم که در این‌جا فرض بر مرتب بودن سطوح متغیر رسته‌ای Y است. در این جدول نیز مؤلفه‌های «خطی در خطی» $(L \times L)$ و «خطی در درجه‌ی دوم» $(L \times Q)$ در سطح معنی‌داری ۰٫۰۱ از ساختار پیوند حذف می‌شود.

۶ بحث و نتیجه‌گیری

همان‌طور که قبلاً اشاره شد تجزیه‌ی مؤلفه‌های پیوند جزئی در هر جدول پیشایندی که حد اقل دو متغیر تشکیل‌دهنده‌ی آن از نوع ترتیبی باشند امکان‌پذیر است. اگر چه تحلیل پیوند جزئی بین متغیرهای رسته‌ای مهم است، تحلیل و تعیین مؤلفه‌های پیوند چندگانه نیز از اهمیت ویژه‌ای برخوردار است. برای تعیین مؤلفه‌های پیوند چندگانه بین یک متغیر رسته‌ای با بقیّه‌ی متغیرها، ابتدا جدول چندبعدی را به‌صورت یک جدول دوبعدی بازسازی می‌کنیم به‌طوری که متغیر مورد نظر در سطر و همه‌ی ترکیب‌های سطوح بقیّه‌ی متغیرها با رعایت ترتیب در ستون قرار گیرند. اگر مؤلفه‌های پیوند را بر اساس مدل (۱) فقط برای جدول پیشایندی حاصل به کار ببریم به تجزیه‌ای برای ساختار پیوند چندگانه به مؤلفه‌های چندجمله‌ای متعامد دست خواهیم یافت. پیوند جزئی و چندگانه در اینجا دقیقاً معادل همبستگی جزئی و چندگانه برای داده‌های پیوسته است.

اگر در همه‌ی روابط داده‌شده تا کنون وزن‌های w_i و w'_j را مساوی یک بگیریم مؤلفه‌های پیوند جزئی ناموزون به دست خواهند آمد. به هر حال در بسیاری از موارد کاربردی، در صورتی که وزن‌ها به‌نحوی مؤثر و معقول انتخاب شده باشند، مدل‌های موزون برازش بهتری نسبت به مدل‌های ناموزون خواهند داشت.

مرجع‌ها

Agresti, A. (2002). *Categorical Data Analysis.* Wiley, New Jersey.

Andersen, E.B. (1991).*The Statistical Analysis of Categorical Data.* Springer, New York.

Ghoreishi, S.K. and Meshkani, M.R. (2006). Bayesian analysis of assocition (BANOAS) in contingency tables with ordinal and interval variables. *J. Stat. Theory Appl.* **5**, 363-372.

Ghoreishi, S.K. and Meshkani, M.R. (2008). Asymptotic maximum likelihood and bayesian analysis of shares of various weighted trends in association models in contingency tables. *J. Stat. Theory Appl.* **7**, 229-243.

Ghoreishi, S.K. and Meshkani, M.R. (2009). Multiple and partial association componenets: Asmptotic distribution and application. *J. Stat. Theory Appl.* **8**, 245-252.

Goodman, L.O. (1972). Some multiplicative models for the analysis of cross-classified data. *Proceeding of the sixth Brekeley Symposium on Mathematical Statistics and Probability* eds. L. le Cam, J. Neyman, and E. L. Scott, Brekely: Univ. of California press, **1**, 649-696.

Goodman, L.O. (1979). Simple models for the analysis of association in cross-classifications having ordered categories. *J. Amer. Statist. Assoc.* **74**, 537-552.

Goodman, L.O. (1985). The analysis of cross-classification data having ordered andor unordered categories: association models, correlation models and asymmetry models for contingency tables with or without missing entries. *Ann. Statist.* **13**, 10-69.

Goodman, L.O. (1991). Measures, models, and graphical displays in the analysis of croos-classified data. *J. Amer. Statist. Assoc.* **86**, 1085-1111.

Goodman, L.O. and Kruskal, W.H. (1954). Measures of association for cross-classification. *J. Amer. Statist. Assoc.* **49**, 723-764.

Goodman, L.O. and Kruskal, W.H. (1954). *Measures of Association for Cross-Classifications.* Springer-Veriag, New York.

Leibetrau, A.M. (1983). *Measures of Association.* Sage, California.

سید کامران قریشی
گروه آمار، دانشکده‌ی علوم،
دانشگاه رازی،
کرمانشاه، ایران.
رایانشانی: *atty_ghoreishi@yahoo.com*

محمدرضا مشکانی
گروه آمار، دانشکده‌ی علوم ریاضی،
دانشگاه شهید بهشتی،
تهران، ایران.
رایانشانی: *m-meshkani@sbu.ac.ir*