ISSN 2079-0023 (print)
ISSN 2410-2857 (online) Системний аналіз, управління та інформаційні технології

УДК 519.235

О. Н. БЕЗМЕНОВА, И. П. ГАМАЮН, Н. И. БЕЗМЕНОВ

ПОКАЗАТЕЛИ СТЕПЕНИ СВЯЗИ МЕЖДУ НОМИНАЛЬНЫМИ ПРИЗНАКАМИ

Рассматривается проблема оценивания степени связи между параметрами, измеряемыми в шкале наименований. Предложены показатели степени связи, основанные на максимизации вероятности правильного предсказания значения одного параметра по значению другого. Построены показатели, ориентированные на линейную зависимость от вероятностей отдельных значений, а также показатели с эллиптической зависимостью. На основе предложенных показателей сформированы коэффициенты связи. Получены оценки предложенных показателей степени связи, базирующиеся на принципе максимального правдоподобия. Сделаны выводы о возможности использования предложенных показателей степени связи при решении задач группировки параметров и классификации объектов.

Ключевые слова: степень связи, параметр, шкала наименований, показатель степени связи, коэффициент связи, линейная зависимость, эллиптическая зависимость, принцип максимального правдоподобия.

Введение. Несмотря на широкое применение методов группировки параметров и классификации объектов (кластерного анализа) общепринятого определения групп и классов нет [1]. Большинство разработчиков этих методов интуитивно понимают то, что элементы одной группы (или одного класса) должны быть ближе друг к другу, чем к другим элементам, однако особенности этого отношения в общем случае явно указать нельзя. Разработано довольно много показателей степени связи между параметрами (см., например, [2–4]), причем для количественных параметров чаще всего ориентируются на использование коэффициента корреляции [5–7].

Цель статьи – разработка коэффициентов для оценивания степени связи между параметрами, измеряемыми в шкале наименований.

Постановка задачи. Пусть имеется множество параметров $S=\{X_j \mid j=\overline{1,N}\}$, описывающих некоторый объект, над которым произведено n наблюдений. В ходе каждого наблюдения фиксируются значения параметров X_j, $j=\overline{1,N}$. Пусть результаты наблюдений над некоторым объектом представлены в виде матрицы наблюдений $U=[x_{ij}]_{n\times N}$, i-я строка которой содержит результаты i-го наблюдения над множеством S параметров объекта, а j-й столбец составляют реализации параметра X_j в n наблюдениях. При оценивании степени связи между параметрами будем рассматривать только попарные связи. Без нарушения общности в качестве пары параметров рассмотрим X_1 и X_2.

Пусть имеются два параметра X_1 и X_2, измеряемые в шкале наименований, и пусть количества различных значений, которые могут принять эти параметры, равны соответственно n_1 и n_2. Построим показатели степени связи для таких параметров.

Решение. Рассмотрим параметры X_1 и X_2 как случайные величины и будем полагать, что известно их совместное распределение:

Рассмотрим параметры X_1 и X_2 как случайные величины и будем полагать, что известно их совместное распределение:

$$p_{ij},\ i=\overline{1,n_1},\ j=\overline{1,n_2}.$$

Связь между параметрами X_1 и X_2 будем оценивать возможностью предсказания значения параметра X_2 по значению параметра X_1 (либо наоборот). В этом случае максимальная связь будет существовать только в случае, когда любому значению параметра X_1 будет соответствовать одно и только одно значение параметра X_2, то есть для любого i ($i=\overline{1,n_1}$) выполняется соотношение

$$p_{ij}=\begin{cases} p_{i\bullet} & \text{при } j=j^{0i}, \\ 0 & \text{при } j=\overline{1,n_2},\ j\neq j^{0i}, \end{cases}$$

где $p_{i\bullet}=\sum_{j=1}^{n_2} p_{ij}$.

Минимальной связи будет соответствовать случай невозможности предсказания значения случайной величины X_2 по значению величины X_1, то есть при выполнении условия

$$p_{ij}=\frac{p_{i\bullet}}{n_2},\ j=\overline{1,n_2},\ i=\overline{1,n_1}.$$

Исходя из изложенного выше, связь между параметрами X_1 и X_2 можно оценивать по максимальной вероятности предсказания по значению параметра X_1 значения параметра X_2, а в качестве показателя степени зависимости X_2 от X_1 рассматривать сумму этих вероятностей:

$$\zeta_1(X_1\to X_2)=\sum_{i=1}^{n_1}\max_{j=\overline{1,n_2}} p_{ij}.$$

Отметим, что величина $\zeta_1(X_1\to X_2)$ принимает свои значения в отрезке $\left[\frac{1}{n_2},1\right]$.

Аналогично можно построить показатель степени зависимости параметра X_1 от параметра X_2:

$$\zeta_1(X_1\leftarrow X_2)=\sum_{j=1}^{n_2}\max_{i=\overline{1,n_1}} p_{ij}.$$

Этот показатель принимает свои значения в отрезке $\left[1/n_1, 1 \right]$.

При отсутствии информации о направлении возможной связи между двумя параметрами возможно использование агрегированного показателя, в качестве которого может быть использован один из следующих показателей:

$$\varsigma_1(X_1 : X_2) = \frac{\zeta_1(X_1 \to X_2) + \zeta_1(X_1 \leftarrow X_2)}{2};$$

$$\varsigma_2(X_1 : X_2) = \max\left(\zeta_1(X_1 \to X_2) + \zeta_1(X_1 \leftarrow X_2)\right);$$

$$\varsigma_3(X_1 : X_2) = \min\left(\zeta_1(X_1 \to X_2) + \zeta_1(X_1 \leftarrow X_2)\right).$$

Поскольку различные значения одного параметра в разной степени оказывают влияние на значение другого параметра, желательным является учет этого факта в показателе степени связи. Учет того, в какой степени каждое возможное значение одного параметра определяет значение другого параметра, обеспечивают такие показатели:

$$\zeta_2(X_1 \to X_2) = \frac{1}{n_1} \sum_{i=1}^{n_1} \frac{\max\limits_{j=\overline{1,n_2}} p_{ij}}{p_{i\bullet}}; \qquad (1)$$

$$\zeta_2(X_1 \leftarrow X_2) = \frac{1}{n_2} \sum_{j=1}^{n_2} \frac{\max\limits_{i=\overline{1,n_1}} p_{ij}}{p_{\bullet j}}; \qquad (2)$$

$$\varsigma_4(X_1 : X_2) = \frac{\zeta_2(X_1 \to X_2) + \zeta_2(X_1 \leftarrow X_2)}{2};$$

$$\varsigma_5(X_1 : X_2) = \max\left(\zeta_2(X_1 \to X_2) + \zeta_2(X_1 \leftarrow X_2)\right);$$

$$\varsigma_6(X_1 : X_2) = \min\left(\zeta_2(X_1 \to X_2) + \zeta_2(X_1 \leftarrow X_2)\right).$$

Величина $p_{\bullet j}$ в выражении (2) определяется по формуле:

$$p_{\bullet j} = \sum_{i=1}^{n_1} p_{ij}.$$

Пусть теперь имеется выборка из n наблюдений над парой указанных параметров X_1 и X_2 и пусть величина m_{ij} ($i = \overline{1,n_1}$, $j = \overline{1,n_2}$) определяет, сколько раз в выборке i-му значению параметра X_1 соответствует j-е значение параметра X_2. Тогда для величины p_{ij} имеем следующую оценку максимального правдоподобия:

$$\hat{p}_{ij} = \frac{m_{ij}}{n}, \; i = \overline{1,n_1}, \; j = \overline{1,n_2}.$$

Выборочными оценками максимального правдоподобия значений $p_{i\bullet}$ и $p_{\bullet j}$ являются следующие величины:

$$\hat{p}_{i\bullet} = \sum_{j=1}^{n_2} \hat{p}_{ij}, \; i = \overline{1,n_1};$$

$$\hat{p}_{\bullet j} = \sum_{i=1}^{n_1} \hat{p}_{ij}, \; j = \overline{1,n_2}$$

или иначе

$$\hat{p}_{i\bullet} = \frac{m_{i\bullet}}{n}, \; i = \overline{1,n_1};$$

$$\hat{p}_{\bullet j} = \frac{m_{\bullet j}}{n}, \; j = \overline{1,n_2},$$

где $m_{i\bullet} = \sum_{j=1}^{n_2} m_{ij}$, $m_{\bullet j} = \sum_{i=1}^{n_1} m_{ij}$.

Тогда можно получить следующие оценки максимального правдоподобия для введенных выше показателей степени связи:

$$\hat{\zeta}_1(X_1 \to X_2) = \frac{1}{n} \sum_{i=1}^{n_1} \max_{j=\overline{1,n_2}} m_{ij};$$

$$\hat{\zeta}_1(X_1 \leftarrow X_2) = \frac{1}{n} \sum_{j=1}^{n_2} \max_{i=\overline{1,n_1}} m_{ij};$$

$$\hat{\varsigma}_1(X_1 : X_2) = \frac{\hat{\zeta}_1(X_1 \to X_2) + \hat{\zeta}_1(X_1 \leftarrow X_2)}{2};$$

$$\hat{\varsigma}_2(X_1 : X_2) = \max\left(\hat{\zeta}_1(X_1 \to X_2) + \hat{\zeta}_1(X_1 \leftarrow X_2)\right);$$

$$\hat{\varsigma}_3(X_1 : X_2) = \min\left(\hat{\zeta}_1(X_1 \to X_2) + \hat{\zeta}_1(X_1 \leftarrow X_2)\right);$$

$$\hat{\zeta}_2(X_1 \to X_2) = \frac{1}{n_1} \sum_{i=1}^{n_1} \frac{\max\limits_{j=\overline{1,n_2}} m_{ij}}{m_{i\bullet}};$$

$$\hat{\zeta}_2(X_1 \leftarrow X_2) = \frac{1}{n_2} \sum_{j=1}^{n_2} \frac{\max\limits_{i=\overline{1,n_1}} m_{ij}}{m_{\bullet j}};$$

$$\hat{\varsigma}_4(X_1 : X_2) = \frac{\hat{\zeta}_2(X_1 \to X_2) + \hat{\zeta}_2(X_1 \leftarrow X_2)}{2};$$

$$\hat{\varsigma}_5(X_1 : X_2) = \max\left(\hat{\zeta}_2(X_1 \to X_2) + \hat{\zeta}_2(X_1 \leftarrow X_2)\right);$$

$$\hat{\varsigma}_6(X_1 : X_2) = \min\left(\hat{\zeta}_2(X_1 \to X_2) + \hat{\zeta}_2(X_1 \leftarrow X_2)\right).$$

Для сопоставления степени связи для различных пар параметров желательным является наличие у показателей степени связи следующих свойств:

♦ принятие значений в интервале от 0 до 1 включительно;

♦ принятие значения 1 при жесткой зависимости двух параметров;

♦ принятие нулевого значения при отсутствии связи между параметрами.

Все рассмотренные выше показателей степени связи обладают двумя первыми свойствами, но не обладают третьим. Осуществим их преобразование, направленное на обеспечение выполнения третьего свойства с сохранением первых двух.

Поскольку величина $\zeta_1(X_1 \to X_2)$ принимает свои значения в отрезке $\left[1/n_2 , 1 \right]$, вычтем из нее ее минимальное значение, равное $1/n_2$, и разделим полученное выражение на максимальное значение, которое оно может принять.

Получаем следующий результат:

$$\frac{\zeta_1(X_1 \to X_2) - 1/n_2}{\max\left(\zeta_1(X_1 \to X_2) - 1/n_2 \right)} =$$

$$= \frac{\zeta_1(X_1 \to X_2) - 1/n_2}{\max \zeta_1(X_1 \to X_2) - 1/n_2} = \frac{\sum_{i=1}^{n_1} \max\limits_{j=\overline{1,n_2}} p_{ij} - 1/n_2}{1 - 1/n_2}$$

Итак, величина

$$\zeta_3(X_1 \to X_2) = \frac{\sum_{i=1}^{n_1} \max\limits_{j=\overline{1,n_2}} p_{ij} - 1/n_2}{1 - 1/n_2}$$

обладает свойствами коэффициента степени связи.

Аналогично можно ввести коэффициент влияния

$$\zeta_3(X_1 \leftarrow X_2) = \frac{\sum_{j=1}^{n_2} \max\limits_{i=\overline{1,n_1}} p_{ij} - 1/n_1}{1 - 1/n_1}, \qquad (3)$$

а также следующие агрегированные коэффициенты связи

$$\varsigma_5(X_1 : X_2) = \frac{\zeta_2(X_1 \to X_2) + \zeta_2(X_1 \leftarrow X_2)}{2}; \qquad (4)$$

$$\varsigma_6(X_1 : X_2) = \max\left(\zeta_2(X_1 \to X_2), \zeta_2(X_1 \leftarrow X_2) \right); \qquad (5)$$

$$\varsigma_7(X_1 : X_2) = \min\left(\zeta_2(X_1 \to X_2), \zeta_2(X_1 \leftarrow X_2) \right). \qquad (6)$$

Выборочными оценками максимального правдоподобия коэффициентов (3)–(6) является следующие показатели:

$$\hat{\zeta}_3(X_1 \to X_2) = \frac{\frac{1}{n}\sum_{i=1}^{n_1} \max\limits_{j=\overline{1,n_2}} m_{ij} - 1/n_2}{1 - 1/n_2};$$

$$\hat{\zeta}_3(X_1 \leftarrow X_2) = \frac{\frac{1}{n}\sum_{j=1}^{n_2} \max\limits_{i=\overline{1,n_1}} m_{ij} - 1/n_1}{1 - 1/n_1};$$

$$\hat{\varsigma}_8(X_1 : X_2) = \frac{\hat{\zeta}_2(X_1 \to X_2) + \hat{\zeta}_2(X_1 \leftarrow X_2)}{2};$$

$$\hat{\varsigma}_9(X_1 : X_2) = \max\left(\hat{\zeta}_2(X_1 \to X_2), \hat{\zeta}_2(X_1 \leftarrow X_2) \right);$$

$$\hat{\varsigma}_{10}(X_1 : X_2) = \min\left(\hat{\zeta}_2(X_1 \to X_2), \hat{\zeta}_2(X_1 \leftarrow X_2) \right).$$

Сформируем теперь коэффициенты связи на базе показателей (1)–(2). Для этого определим сначала диапазон изменения показателя связи $\zeta_2(X_1 \to X_2)$.

Наименьшее значение этого показателя достигается в случае невозможности предсказания значения параметра X_2 по значению параметра X_1. Этой ситуации соответствует случай равновероятности всех возможных значений X_2 для любого из возможных значений X_1, то есть при

$$p_{ij} = \frac{1}{n_1 n_2}, \; j = \overline{1,n_2}, \; i = \overline{1,n_1}.$$

При этом

$$\max_{j=\overline{1,n_2}} p_{ij} = \frac{1}{n_1 n_2}, \; i = \overline{1,n_1};$$

$$p_{i\bullet} = \sum_{j=1}^{n_2} p_{ij} = \sum_{j=1}^{n_2} \frac{1}{n_1 n_2} = \frac{1}{n_1}, \; i = \overline{1,n_1}.$$

Следовательно,

$$\zeta_2(X_1 \to X_2)_{\min} = \frac{1}{n_1} \sum_{i=1}^{n_1} \frac{1/n_1 n_2}{1/n_1} = \frac{1}{n_1} \cdot n_1 \cdot \frac{1}{n_2} = \frac{1}{n_2}.$$

В то же время максимальное значение показателя $\zeta_2(X_1 \to X_2)$ достигается в том случае, когда значение параметра X_1 с вероятностью 1 определяет значение параметра X_2, то есть если выполняется соотношение

$$\max_{i=\overline{1,n_1}} p_{ij} = p_{i\bullet}, \; i = \overline{1,n_1}.$$

Таким образом,

$$\zeta_2(X_1 \to X_2)_{\max} = \frac{1}{n_1} \sum_{i=1}^{n_1} \frac{p_{i\bullet}}{p_{i\bullet}} = 1.$$

Итак,

$$\zeta_2(X_1 \to X_2) \in \left[\frac{1}{n_2}, 1 \right].$$

Для перехода к интервалу [0, 1] вычтем из $\zeta_2(X_1 \to X_2)$ значение $\zeta_2(X_1 \to X_2)_{\min}$ и разделим результат на $\zeta_2(X_1 \to X_2)_{\max} - \zeta_2(X_1 \to X_2)_{\min}$, после чего выполним некоторые преобразования:

$$\frac{\zeta_2(X_1\to X_2)-\zeta_2(X_1\to X_2)_{\min}}{\zeta_2(X_1\to X_2)_{\max}-\zeta_2(X_1\to X_2)_{\min}}=\frac{\frac{1}{n_1}\sum_{i=1}^{n_1}\frac{\max\limits_{j=\overline{1,n_2}}p_{ij}}{p_{i\bullet}}-\frac{1}{n_2}}{1-\frac{1}{n_2}}=$$

$$=\frac{\left(\frac{1}{n_1}\sum_{i=1}^{n_1}\frac{\max\limits_{j=\overline{1,n_2}}p_{ij}}{p_{i\bullet}}-\frac{n_1}{n_2}\right)}{1-\frac{1}{n_2}}=\frac{n_2}{n_1(n_2-1)}\sum_{i=1}^{n_1}\left(\frac{\max\limits_{j=\overline{1,n_2}}p_{ij}}{p_{i\bullet}}-\frac{1}{n_2}\right)=$$

$$=\frac{n_2}{n_1(n_2-1)}\sum_{i=1}^{n_1}\left(n_2\frac{\max\limits_{j=\overline{1,n_2}}p_{ij}}{p_{i\bullet}}-1\right).$$

В результате получаем следующий коэффициент степени связи:

$$\zeta_4(X_1\to X_2)=\frac{n_2}{n_1(n_2-1)}\sum_{i=1}^{n_1}\left(n_2\frac{\max\limits_{j=\overline{1,n_2}}p_{ij}}{p_{i\bullet}}-1\right).$$

Его выборочная оценка может быть представлена следующим образом:

$$\hat{\zeta}_4(X_1\to X_2)=\frac{n_2}{n_1(n_2-1)}\sum_{i=1}^{n_1}\left(n_2\frac{\max\limits_{j=\overline{1,n_2}}m_{ij}}{m_{i\bullet}}-1\right).$$

Аналогично можно построить коэффициент связи, отражающий влияние параметра X_2 на параметр X_1:

$$\zeta_4(X_1\leftarrow X_2)=\frac{n_1}{n_2(n_1-1)}\sum_{j=1}^{n_2}\left(n_1\frac{\max\limits_{i=\overline{1,n_1}}p_{ij}}{p_{j\bullet}}-1\right),$$

а также агрегированные коэффициенты связи:

$$\varsigma_8(X_1:X_2)=\frac{\zeta_4(X_1\to X_2)+\zeta_4(X_1\leftarrow X_2)}{2};$$

$$\varsigma_9(X_1:X_2)=\max\bigl(\zeta_4(X_1\to X_2),\,\zeta_4(X_1\leftarrow X_2)\bigr);$$

$$\varsigma_{10}(X_1:X_2)=\min\bigl(\zeta_4(X_1\to X_2),\,\zeta_4(X_1\leftarrow X_2)\bigr).$$

Их выборочные оценки определяются следующими выражениями:

$$\hat{\zeta}_4(X_1\leftarrow X_2)=\frac{n_1}{n_2(n_1-1)}\sum_{j=1}^{n_2}\left(n_1\frac{\max\limits_{i=\overline{1,n_1}}m_{ij}}{p_{j\bullet}}-1\right),$$

$$\hat{\varsigma}_8(X_1:X_2)=\frac{\hat{\zeta}_4(X_1\to X_2)+\hat{\zeta}_4(X_1\leftarrow X_2)}{2};$$

$$\hat{\varsigma}_9(X_1:X_2)=\max\bigl(\hat{\zeta}_4(X_1\to X_2),\,\hat{\zeta}_4(X_1\leftarrow X_2)\bigr);$$

$$\hat{\varsigma}_{10}(X_1:X_2)=\min\bigl(\hat{\zeta}_4(X_1\to X_2),\,\hat{\zeta}_4(X_1\leftarrow X_2)\bigr).$$

Если рассматривать введенные выше показатели степени связи между двумя параметрами как функции, аргументами которых является величина $\max\limits_{j=\overline{1,n_2}}p_{ij}$ (или $\max\limits_{i=\overline{1,n_1}}p_{ij}$), то можно отметить линейность этих функций. Рассмотрим несколько иной принцип построения показателя степени связи между двумя параметрами.

Как и ранее, в качестве аргумента функции для показателя, отражающего возможность предсказания значения параметра X_2 по значению параметра X_1, возьмём величину $\max\limits_{j=\overline{1,n_2}}p_{ij}$, $i=\overline{1,n_1}$, которая изменяет свое значение на интервале от $\frac{p_{i\bullet}}{n_2}$ (случай невозможности предсказания) до $p_{i\bullet}$ (случай абсолютной определенности) включительно.

Рассмотрим эллипс, описываемый следующим уравнением:

$$\frac{\left(\max\limits_{j=\overline{1,n_2}}p_{ij}-\frac{p_{i\bullet}}{n_2}\right)^2}{\left(p_{i\bullet}-\frac{p_{i\bullet}}{n_2}\right)^2}+\frac{(y_i-1)^2}{1^2}=1,\ i=\overline{1,n_1}.$$

Графическое отображение этого эллипса приведено на рис. 1 (для обеспечения наглядности на рисунке использованы различные шкалы по двум осям координат).

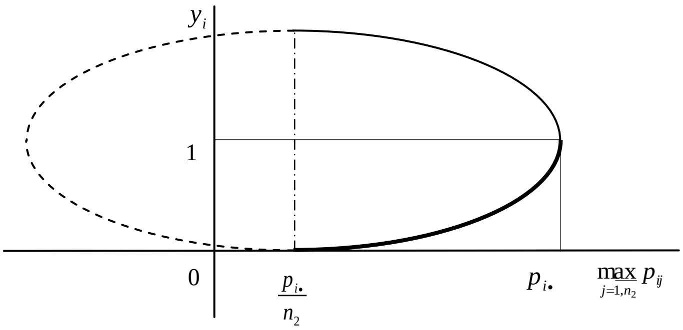

Рис. 1 – Первый вариант формирования составляющей коэффициента связи

Отметим, что аргумент $\max\limits_{j=\overline{1,n_2}} p_{ij}$ в данном случае существует только на отрезке $\left[\dfrac{p_{i\bullet}}{n_2},\, p_{i\bullet}\right]$. Поэтому часть эллипса слева от его вертикальной оси симметрии физически не может существовать.

Получим выражение для y_i:

$$y_{i\,(1,2)} = 1 \mp \sqrt{1 - \frac{\left(\max\limits_{j=\overline{1,n_2}} p_{ij} - \dfrac{p_{i\bullet}}{n_2}\right)^2}{\left(p_{i\bullet} - \dfrac{p_{i\bullet}}{n_2}\right)^2}}\,,\; i = \overline{1,n_1}\,.$$

Поскольку требованием к коэффициенту связи является принятие им значений из отрезка числовой оси от 0 до 1 с установлением факта отсутствия связи при равенстве его 0 и факта наличия полной связи при 1, нас интересует нижняя часть правой половины эллипса, представленного на рис. 1. Запишем выражение для этой части эллипса и выполним некоторые преобразования этого выражения:

$$y_{i\,(1)} = 1 - \sqrt{1 - \frac{\left(\max\limits_{j=\overline{1,n_2}} p_{ij} - \dfrac{p_{i\bullet}}{n_2}\right)^2}{\left(p_{i\bullet} - \dfrac{p_{i\bullet}}{n_2}\right)^2}} = 1 - \sqrt{1 - \frac{\left(n_2 \max\limits_{j=\overline{1,n_2}} p_{ij} - p_{i\bullet}\right)^2}{p_{i\bullet}^2 (n_2 - 1)^2}} =$$

$$= 1 - \frac{1}{n_2 - 1}\sqrt{(n_2 - 1)^2 - \frac{\left(n_2 \max\limits_{j=\overline{1,n_2}} p_{ij} - p_{i\bullet}\right)^2}{p_{i\bullet}^2}} =$$

$$= 1 - \frac{1}{n_2 - 1}\sqrt{(n_2 - 1)^2 - \left(\frac{n_2 \max\limits_{j=\overline{1,n_2}} p_{ij} - p_{i\bullet}}{p_{i\bullet}} - 1\right)^2} =$$

$$= 1 - \frac{n_2}{n_2 - 1}\sqrt{\left(1 - \frac{1}{n_2}\right)^2 - \left(\frac{\max\limits_{j=\overline{1,n_2}} p_{ij} - p_{i\bullet}}{p_{i\bullet}} - \frac{1}{n_2}\right)^2}$$

Тогда коэффициент связи, отражающий степень влияния параметра X_1 на параметр X_2, можно сформировать как среднее арифметическое всех показателей $y_{i\,(1)}$:

$$\zeta_5(X_1 \to X_2) = 1 - \frac{n_2}{n_1(n_2 - 1)}\sum_{i=1}^{n_1}\sqrt{\left(1 - \frac{1}{n_2}\right)^2 - \left(\frac{\max\limits_{j=\overline{1,n_2}} p_{ij} - p_{i\bullet}}{p_{i\bullet}} - \frac{1}{n_2}\right)^2}\,.$$

Выборочная оценка этого показателя имеет следующий вид:

$$\hat{\zeta}_5(X_1 \to X_2) = 1 - \frac{n_2}{n_1(n_2 - 1)}\sum_{i=1}^{n_1}\sqrt{\left(1 - \frac{1}{n_2}\right)^2 - \left(\frac{\max\limits_{j=\overline{1,n_2}} m_{ij} - p_{i\bullet}}{m_{i\bullet}} - \frac{1}{n_2}\right)^2}\,.$$

Рассмотрим теперь эллипс, задаваемый таким уравнением (см. также рис. 2):

$$\frac{\left(\max\limits_{j=\overline{1,n_2}} p_{ij} - \dfrac{p_{i\bullet}}{n_2}\right)^2}{\left(p_{i\bullet} - \dfrac{p_{i\bullet}}{n_2}\right)^2} + \frac{(y_i - 0)^2}{1^2} = 1\,,\; i = \overline{1,n_1}\,.$$

Как и у эллипса на рис. 1, у данного эллипса аргумент $\max\limits_{j=\overline{1,n_2}} p_{ij}$ физически также может принимать значения только из отрезка $\left[\dfrac{p_{i\bullet}}{n_2},\, p_{i\bullet}\right]$.

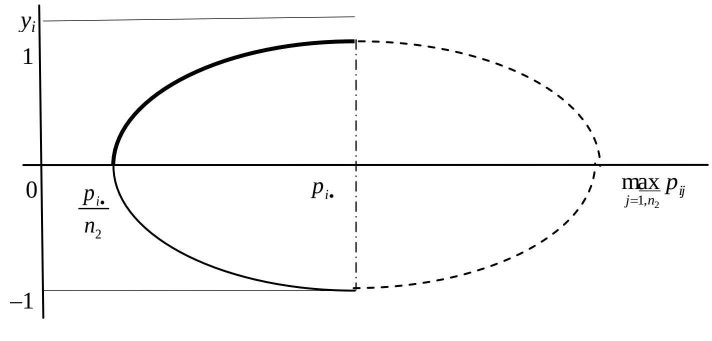

Рис. 2 – Второй вариант формирования составляющей коэффициента связи

В этом случае имеем: $y_{i\,(1,2)} = \mp\sqrt{1 - \dfrac{\left(\max\limits_{j=\overline{1,n_2}} p_{ij} - p_{i\bullet}\right)^2}{\left(p_{i\bullet} - \dfrac{p_{i\bullet}}{n_2}\right)^2}}\,,\; i = \overline{1,n_1}\,.$

ISSN 2079-0023 (print)
ISSN 2410-2857 (online)

В качестве составляющей коэффициента связи будем рассматривать только верхнюю часть левой половины эллипса:

$$y_{i(2)} = \sqrt{1 - \frac{\left(\max\limits_{j=\overline{1,n_2}} p_{ij} - p_{i\bullet}\right)^2}{\left(p_{i\bullet} - \dfrac{p_{i\bullet}}{n_2}\right)^2}}\ .$$

Выполнив преобразования, аналогичные тем, которые осуществлялись при формировании показателя степени связи $\zeta_5(X_1 \to X_2)$, получаем следующее выражение для $y_{i(2)}$:

$$y_{i(2)} = \sqrt{\left(1 - \frac{1}{n_2}\right)^2 - \left(1 - \frac{\max\limits_{j=\overline{1,n_2}} p_{ij}}{p_{i\bullet}}\right)^2}\ ,\ i = \overline{1,n_1}\ .$$

Используя эти выражения, получаем следующий коэффициент связи:

$$\zeta_6(X_1 \to X_2) = \frac{n_2}{n_1(n_2-1)} \sum_{i=1}^{n_1} \sqrt{\left(1 - \frac{1}{n_2}\right)^2 - \left(1 - \frac{\max\limits_{j=\overline{1,n_2}} p_{ij}}{p_{i\bullet}}\right)^2}\ .$$

Его выборочная оценка имеет вид:

$$\hat{\zeta}_6(X_1 \to X_2) = \frac{n_2}{n_1(n_2-1)} \sum_{i=1}^{n_1} \sqrt{\left(1 - \frac{1}{n_2}\right)^2 - \left(1 - \frac{\max\limits_{j=\overline{1,n_2}} m_{ij}}{m_{i\bullet}}\right)^2}$$

Как компромиссный по отношению к коэффициентам связи $\zeta_5(X_1 \to X_2)$ и $\zeta_6(X_1 \to X_2)$ может быть рассмотрен коэффициент связи, построенный на среднем арифметическом $\bar{y}_i$ величин $y_{i(1)}$ и $y_{i(2)}$, $i = \overline{1,n_1}$ (см. рис. 3),

$$\zeta_7(X_1 \to X_2) = \frac{\zeta_5(X_1 \to X_2) + \zeta_6(X_1 \to X_2)}{2}$$

с выборочной оценкой

$$\hat{\zeta}_7(X_1 \to X_2) = \frac{\hat{\zeta}_5(X_1 \to X_2) + \hat{\zeta}_6(X_1 \to X_2)}{2}\ .$$

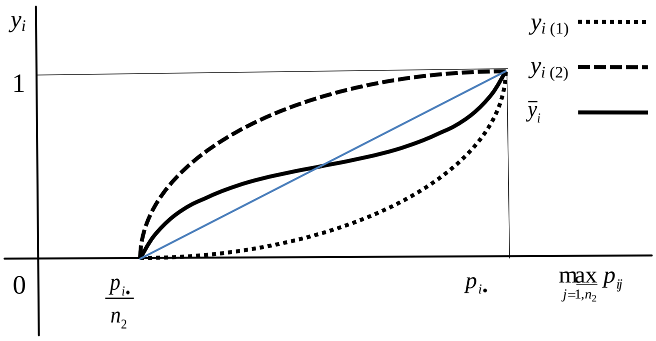

Рис. 3 – Составляющая агрегированного коэффициента связи

Аналогично могут быть построены коэффициенты степени связи, отражающие влияние параметра X_2 на параметр X_1:

$$\zeta_5(X_1 \leftarrow X_2) = 1 - \frac{n_1}{n_2(n_1-1)} \sum_{j=1}^{n_2} \sqrt{\left(1 - \frac{1}{n_1}\right)^2 - \left(\frac{\max\limits_{i=\overline{1,n_1}} p_{ij} - p_{\bullet j}}{p_{\bullet j}} - \frac{1}{n_1}\right)^2}\ ,$$

$$\zeta_6(X_1 \leftarrow X_2) = \frac{n_1}{n_2(n_1-1)} \sum_{j=1}^{n_2} \sqrt{\left(1 - \frac{1}{n_1}\right)^2 - \left(1 - \frac{\max\limits_{i=\overline{1,n_1}} p_{ij}}{p_{\bullet j}}\right)^2}\ ,$$

$$\zeta_7(X_1 \leftarrow X_2) = \frac{\zeta_5(X_1 \leftarrow X_2) + \zeta_6(X_1 \leftarrow X_2)}{2}$$

и их выборочные оценки:

$$\hat{\zeta}_5(X_1 \leftarrow X_2) = 1 - \frac{n_1}{n_2(n_1-1)} \sum_{j=1}^{n_2} \sqrt{\left(1 - \frac{1}{n_1}\right)^2 - \left(\frac{\max\limits_{i=\overline{1,n_1}} m_{ij} - m_{\bullet j}}{m_{\bullet j}} - \frac{1}{n_1}\right)^2}\ ,$$

$$\hat{\zeta}_6(X_1 \leftarrow X_2) = \frac{n_1}{n_2(n_1-1)} \sum_{j=1}^{n_2} \sqrt{\left(1 - \frac{1}{n_1}\right)^2 - \left(1 - \frac{\max\limits_{i=\overline{1,n_1}} m_{ij}}{m_{\bullet j}}\right)^2}\ ,$$

$$\hat{\zeta}_7(X_1 \leftarrow X_2) = \frac{\hat{\zeta}_5(X_1 \leftarrow X_2) + \hat{\zeta}_6(X_1 \leftarrow X_2)}{2}\ .$$

Если направление связи неизвестно, то предлагается использовать такие коэффициенты связи:

$$\varsigma_{11}(X_1 : X_2) = \frac{\zeta_5(X_1 \to X_2) + \zeta_5(X_1 \leftarrow X_2)}{2}\ ,$$

$$\varsigma_{12}(X_1 : X_2) = \max\left(\zeta_5(X_1 \to X_2), \zeta_5(X_1 \leftarrow X_2)\right),$$

$$\varsigma_{13}(X_1 : X_2) = \min\left(\zeta_5(X_1 \to X_2), \zeta_5(X_1 \leftarrow X_2)\right),$$

$$\varsigma_{14}(X_1 : X_2) = \frac{\zeta_6(X_1 \to X_2) + \zeta_6(X_1 \leftarrow X_2)}{2},$$

$$\varsigma_{15}(X_1 : X_2) = \max\left(\zeta_6(X_1 \to X_2), \zeta_6(X_1 \leftarrow X_2)\right),$$

$$\varsigma_{16}(X_1 : X_2) = \min\left(\zeta_6(X_1 \to X_2), \zeta_6(X_1 \leftarrow X_2)\right),$$

$$\varsigma_{17}(X_1 : X_2) = \frac{\zeta_7(X_1 \to X_2) + \zeta_7(X_1 \leftarrow X_2)}{2},$$

$$\varsigma_{18}(X_1 : X_2) = \max\left(\zeta_7(X_1 \to X_2), \zeta_7(X_1 \leftarrow X_2)\right),$$

$$\varsigma_{19}(X_1 : X_2) = \min\left(\zeta_7(X_1 \to X_2), \zeta_7(X_1 \leftarrow X_2)\right)$$

и их выборочные оценки:

$$\hat{\varsigma}_{11}(X_1 : X_2) = \frac{\hat{\zeta}_5(X_1 \to X_2) + \hat{\zeta}_5(X_1 \leftarrow X_2)}{2},$$

$$\hat{\varsigma}_{12}(X_1 : X_2) = \max\left(\hat{\zeta}_5(X_1 \to X_2), \hat{\zeta}_5(X_1 \leftarrow X_2)\right),$$

$$\hat{\varsigma}_{13}(X_1 : X_2) = \min\left(\hat{\zeta}_5(X_1 \to X_2), \hat{\zeta}_5(X_1 \leftarrow X_2)\right),$$

$$\hat{\varsigma}_{14}(X_1 : X_2) = \frac{\hat{\zeta}_6(X_1 \to X_2) + \hat{\zeta}_6(X_1 \leftarrow X_2)}{2},$$

$$\hat{\varsigma}_{15}(X_1 : X_2) = \max\left(\hat{\zeta}_6(X_1 \to X_2), \hat{\zeta}_6(X_1 \leftarrow X_2)\right),$$

$$\hat{\varsigma}_{16}(X_1 : X_2) = \min\left(\hat{\zeta}_6(X_1 \to X_2), \hat{\zeta}_6(X_1 \leftarrow X_2)\right),$$

$$\hat{\varsigma}_{17}(X_1 : X_2) = \frac{\hat{\zeta}_7(X_1 \to X_2) + \hat{\zeta}_7(X_1 \leftarrow X_2)}{2},$$

$$\hat{\varsigma}_{18}(X_1 : X_2) = \max\left(\hat{\zeta}_7(X_1 \to X_2), \hat{\zeta}_7(X_1 \leftarrow X_2)\right),$$

$$\hat{\varsigma}_{19}(X_1 : X_2) = \min\left(\hat{\zeta}_7(X_1 \to X_2), \hat{\zeta}_7(X_1 \leftarrow X_2)\right).$$

Выводы. Разработанные показатели степени связи между параметрами могут быть использованы при решении задач группировки параметров и классификации объектов. Предложенные показатели степени связи ориентированы на номинальную шкалу. Поэтому они могут быть использованы и при анализе систем параметров, которые описываются в порядковых шкалах, для чего достаточно осуществить переход от порядковой шкалы к шкале наименований посредством отказа от учета отношения упорядочения значений, присущего порядковым шкалам, и учитывать только различие значений.

Список литературы: **1.** *Сокал Р. Р.* Кластер–анализ и классификация: предпосылки и основные направления / *Р. Р. Сокал* // Классификация и кластер / ред. *Дж. Вэн Райзин*; перевод с англ. *П. П. Кольцова*; под ред. *Ю. И. Журавлева*. – М. : Мир, 1980. – С. 7–19. **2.** *Кендалл М. Дж.* Статистические выводы и связи : пер. с англ. / *М. Дж. Кендалл, А. Стюарт* / пер. *Л. И. Гальчук, А. Т. Терёхин*; ред. *А. Н. Колмогоров*. – М. : Наука, 1973. – 899 с. **3.** *Елисеева И. И.* Статистические методы измерения связей / *И. И. Елисеева*. – Л. : Изд-во Ленингр. гос. ун-та, 1982. – 136 с. **4.** *Елисеева И. И.* Группировка, корреляция, распознавание образов / *И. И. Елисеева, В. О. Рукавишников*. – М. : Статистка, 1977. – 144 с. **5.** *Вентцель Е. С.* Теория вероятностей / *Е. С. Вентцель*. – М. : Наука, 1969. – 576 с. **6.** *Феллер В.* Введение в теорию вероятностей и ее приложения : в 2 т. : пер. с англ. / *В. Феллер* / пер. *Ю. В. Прохоров*. – М. : Мир, 1984. – Т. 1. – 528 с. **7.** *Гмурман В. Е.* Теория вероятностей и математическая статистика / *В. Е. Гмурман*. – М. : Высш. шк., 2003. – 479 с.

Bibliography (transliterated): **1.** Sokal, R. R. Sokal, R. R. «Klaster-analiz i klassifikacija: predposylki i osnovnye napravlenija.» Edited by Van Rysen, J. Klassifikacija i klaster. Moskow: Mir, 1980. 7–19. Print. **2.** Kendall, M. G. and A. Stuart. *Statisticheskie vyvody i svjazi*. Moskow: Nauka, 1973. Print. **3.** Eliseeva , I. I. *Statisticheskie metody izmerenija svjazej*. Leningrad: Izd-vo Leningr. gos. un-ta, 1982. Print. **4.** Eliseeva, I. I. and V. O. Rukavishnikov. *Gruppirovka, korreljacija, raspoznavanie obrazov*. Moskov: Statistka, 1977. Print. **5.** Ventcel', E. S. *Teorija verojatnostej*. Moskow: Nauka, 1969. Print. **6.** Feller, V. *Vvedenie v teoriju verojatnostej i ejo prilozhenija*. Vol. 1. M: Mir, 1984. Print. **7.** Gmurman, V. E. *Teorija verojatnostej i matematicheskaja statistika*. Moskow: Vyssh. shk., 2003. Print.

Поступила (received) 05.12.2015