



Эконометрика-1

Филатов Александр Юрьевич

(Главный научный сотрудник, доцент ШЭМ ДВФУ)

alexander.filatov@gmail.com

<http://vk.com/alexander.filatov>, <http://vk.com/baikalreadings>

Лекции 2.1-2.2

Множественные связи.

Порядковые и категоризованные переменные

Линейная зависимость

2

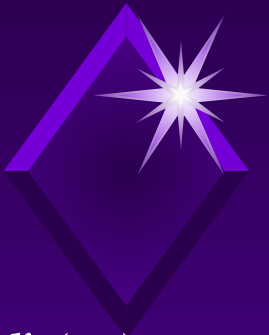
от нескольких объясняющих переменных

Парные коэффициенты корреляции $r_{yx}^{(i)}$ не учитывают влияние на эту связь других переменных $x^{(j)}$. Следовательно, **необходим измеритель связи, очищенный от опосредованного влияния других переменных**, т.е. дающий оценку тесноты связи между y и $x^{(i)}$ при условии, что остальные переменные зафиксированы на некотором постоянном уровне.

Предположение: простой (линейный) характер влияния всех остальных переменных на y : $y = \theta_0 + \theta_1 x^{(1)} + \dots + \theta_p x^{(p)} + \varepsilon(X)$.

Обозначим для удобства $y \equiv x^{(0)}$.

$$R = \begin{pmatrix} 1 & r_{01} & r_{02} & \dots & r_{0p} \\ r_{10} & 1 & r_{12} & \dots & r_{1p} \\ \dots & \dots & \dots & \dots & \dots \\ r_{p0} & r_{p1} & r_{p2} & \dots & 1 \end{pmatrix} \begin{array}{l} R_{ij} - \text{алгебраическое дополнение для } r_{ij} \text{ в} \\ \text{определителе корреляционной матрицы.} \\ R_{ij} = (-1)^{i+j} \det A_{ij}, \text{ матрица } A_{ij} \text{ получена из } R \\ \text{вычеркиванием } i\text{-строки и } j\text{-столбца.} \end{array}$$



Частные (очищенные) коэффициенты корреляции

3

$r_{ij(-ij)} = \frac{-R_{ij}}{(R_{ii}R_{jj})^{1/2}}$ – частный коэффициент корреляции, коэффициент корреляции между переменными $x^{(i)}$ и $x^{(j)}$ при фиксированных значениях всех остальных переменных.

Случай трех переменных:

$$\hat{R} = \begin{pmatrix} 1 & r_{01} & r_{02} \\ r_{01} & 1 & r_{12} \\ r_{02} & r_{12} & 1 \end{pmatrix} \quad r_{01(2)} = \frac{r_{01} - r_{02}r_{12}}{\sqrt{(1 - r_{02}^2)(1 - r_{12}^2)}}$$

Свойства частных коэффициентов корреляции:

Проверка гипотезы о наличии/отсутствии связи, а также построение доверительного интервала для частного коэффициента корреляции k -порядка (при исключении влияния k переменных) происходит по тем же формулам, что и для парного коэффициента корреляции с единственным отличием: **объем выборки уменьшается на k .**

Численные примеры

4

Пример 1:

$n = 37$ – число исследуемых предприятий легкой промышленности,

$x^{(0)} \equiv y$ – качество ткани (в баллах),

$x^{(1)}$ – среднеемесячное число профилактических наладок автоматич. линии,

$x^{(2)}$ – среднеемесячное число обрывов нити.

$$\hat{R} = \begin{pmatrix} 1 & 0,105 & 0,024 \\ 0,105 & 1 & 0,996 \\ 0,024 & 0,996 & 1 \end{pmatrix} \quad r_{01(2)} = \frac{0,105 - 0,024 \cdot 0,996}{\sqrt{(1 - 0,024^2)(1 - 0,996^2)}} = 0,908,$$
$$r_{02(1)} = \frac{0,024 - 0,105 \cdot 0,996}{\sqrt{(1 - 0,105^2)(1 - 0,996^2)}} = -0,907.$$

Связь есть, что согласуется с профессиональными представлениями!

При нахождении доверительного интервала корректируем $n = 37 - 1 = 36$.

$$z = \text{ФИШЕР}(0,908) - \frac{0,908}{2(36-1)} = 1,502, \quad z \in [1,161; 1,843],$$
$$z \in \left[1,502 - \frac{1,96}{\sqrt{36-3}}; 1,502 + \frac{1,96}{\sqrt{36-3}} \right], \quad r \in [0,8214; 0,9511] \text{ при } \gamma = 0,95.$$

Численные примеры

5

Пример 2:

$n = 20$ – число лет метеонаблюдений,

$x^{(0)} \equiv y$ – урожайность кормовых трав,

$x^{(1)}$ – весеннее количество осадков,

$x^{(2)}$ – накопленная за весну сумма активных (выше $+5,5^\circ\text{C}$) температур.

$$\hat{r}_{01} = 0,80, \quad \hat{r}_{02} = -0,40, \quad \hat{r}_{11} = -0,56.$$

$$\hat{R} = \begin{pmatrix} 1 & 0,80 & -0,40 \\ 0,80 & 1 & -0,56 \\ -0,40 & -0,56 & 1 \end{pmatrix}$$

$$r_{01(2)} = \frac{0,80 - 0,56 \cdot 0,40}{\sqrt{(1 - 0,56^2)(1 - 0,40^2)}} = 0,759,$$

$$r_{02(1)} = \frac{-0,40 + 0,80 \cdot 0,56}{\sqrt{(1 - 0,80^2)(1 - 0,56^2)}} = 0,097.$$

Связь со второй переменной не отрицательная, а слабая положительная, что согласуется с профессиональными представлениями!

$$r_{01(2)} \in [0,448; 0,898], \quad r_{02(1)} \in [-0,376; 0,526] \quad \text{при } \gamma = 0,95.$$



Множественный коэффициент корреляции

6

Множественный коэффициент корреляции – коэффициент корреляции между y и линейной функцией регрессии, т.е. между y и наилучшей линейной комбинацией переменных $x^{(1)}, \dots, x^{(p)}$ – той, для которой значение коэффициента корреляции максимально.

$$R_{y.X} = r(y, f(X)) = r\left(y, \theta_0 + \theta_1 x^{(1)} + \dots + \theta_p x^{(p)}\right)$$

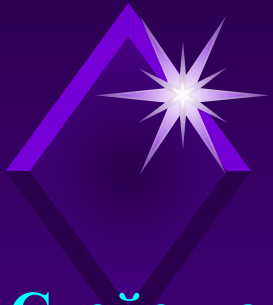
Свойства множественного коэффициента корреляции:

1. При предположении о линейности связи

$$K_d(y; X) = R_{y.X}^2 = 1 - \frac{D\varepsilon}{Dy}.$$

2. Вычисление множественного коэффициента корреляции по корреляционной матрице:

$$R_{y.X}^2 = 1 - \frac{|R|}{|R_{00}|}.$$



Множественный коэффициент корреляции

7

Свойства множественного коэффициента корреляции:

3. Вычисление МКК по частным коэффициентам корреляции:

$$R_{y.X}^2 = 1 - \left(1 - r_{01}^2\right) \left(1 - r_{02(1)}^2\right) \left(1 - r_{03(12)}^2\right) \dots \left(1 - r_{0p(123\dots(p-1))}^2\right)$$

$$\hat{R}_{y.(x^{(1)}, x^{(2)})}^2 = 1 - \left(1 - \hat{r}_{01}^2\right) \left(1 - \hat{r}_{02(1)}^2\right) = 1 - \left(1 - 0,105^2\right) \left(1 - (-0,9068)^2\right) = 0,8243.$$

$$\hat{R}_{y.(x^{(1)}, x^{(2)})}^2 = 1 - \left(1 - \hat{r}_{02}^2\right) \left(1 - \hat{r}_{01(2)}^2\right) = 1 - \left(1 - 0,024^2\right) \left(1 - 0,9079^2\right) = 0,8243.$$

4. МКК мажорирует все парные и частные КК, характеризующие статистическую связь: $R_{y.X}^2 \geq r_{0j(I_j)}^2$, где I_j – любое подмножество $\{1, \dots, p\}$, не содержащее j .

5. Присоединение новой переменной не может уменьшить величины R вне зависимости от порядка присоединения:

$$R_{y.x^{(1)}}^2 \leq R_{y.(x^{(1)}, x^{(2)})}^2 \leq R_{y.(x^{(1)}, x^{(2)}, x^{(3)})}^2 \leq \dots \leq R_{y.(x^{(1)}, x^{(2)}, \dots, x^{(p)})}^2.$$

Проверка гипотезы о наличии множественной линейной связи

Гипотеза о статистической независимости y и $x^{(1)}, \dots, x^{(p)}$ $H_0: R_{y.X} = 0$.

1. Выбираем уровень значимости α .

Типичные значения $\alpha = 0,05$; 0,1; 0,01, 0,001.

2. Вычисляем эмпирическое значение критерия:

$$F_{\text{эмп}} = \frac{\hat{R}_{y.X}^2}{1 - \hat{R}_{y.X}^2} \frac{n - p - 1}{p}.$$

3. Вычисляем критическую точку:

$$F_{\text{крит}} = F(\alpha; p; n - p - 1) = \text{ФРАСПОБР}(\alpha; p; n - p - 1).$$

4. Сравниваем эмпирическое и критическое значение и делаем вывод:

Если $F_{\text{эмп}} > F_{\text{крит}}$, то гипотеза H_0 об отсутствии множественной линейной связи отвергается при уровне значимости α , связь есть.



Корреляционный анализ порядковых переменных

9

$x^{(1)}, \dots, x^{(p)}$ – порядковые переменные (обозначающие порядковое место в ряду, отсортированному по соответствующему показателю).

Типовые задачи:

1. Анализ структуры упорядочений.

1. Точки разбросаны равномерно, нет согласованности между переменными.
2. Часть из p переменных близки между собой.
3. Часть из n объектов близки между собой.

2. Анализ совокупной согласованности переменных.

Исследование степени согласованности мнений экспертов.

3. Построение единого группового упорядочения объектов, т.е. ранжировки $x^{(0)}$, минимально удаленной от $x^{(1)}, \dots, x^{(p)}$.

Объединенные ранги:

Если есть неразличимые по некоторому свойству объекты, им всем приписывается единый ранг, равный среднему арифметическому.



Ранговый коэффициент корреляции Спирмена

10

Базовая формула:

$$\hat{r}_{kj}^{(S)} = 1 - \frac{6}{n^3 - n} \sum_{i=1}^n \left(x_i^{(k)} - x_i^{(j)} \right)^2.$$

Формула для случая объединенных рангов:

$$\hat{r}_{kj}^{(S)} = \frac{1/6 \left(n^3 - n \right) - \sum_{i=1}^n \left(x_i^{(k)} - x_i^{(j)} \right)^2 - T^{(k)} - T^{(j)}}{\sqrt{\left(1/6 \left(n^3 - n \right) - 2T^{(k)} \right) \left(1/6 \left(n^3 - n \right) - 2T^{(j)} \right)}},$$

$$T^{(k)} = \frac{1}{12} \sum_{t=1}^{m^{(k)}} \left(\left(n_t^{(k)} \right)^3 - n_t^{(k)} \right), \quad m^{(k)} - \text{число групп объединенных рангов,}$$

$n_t^{(k)} - \text{число элементов в каждой групп.}$

Свойства коэффициента Спирмена:

$$\hat{r}_{kj}^{(S)} = 1, \text{ если } x_i^{(k)} = x_i^{(j)}, i = 1, \dots, n,$$

$$\hat{r}_{kj}^{(S)} = -1, \text{ если } x_i^{(k)} + x_i^{(j)} = n + 1, i = 1, \dots, n,$$

$$\hat{r}_{kj}^{(S)} \in (-1; 1) \text{ в остальных случаях.}$$



Численные примеры

Пример 1:

10 инвестиционных проектов, проранжированных 2 экспертами.

$$x^{(1)} = (1\ 2\ 3\ 4\ 5\ 6\ 7\ 8\ 9\ 10)$$

$$x^{(2)} = (2\ 3\ 1\ 4\ 6\ 5\ 9\ 7\ 8\ 10)$$

$$\begin{aligned} \hat{r}_{12}^{(S)} &= 1 - \frac{6}{990} (1 + 1 + 4 + \dots + 0) = \\ &= 1 - \frac{84}{990} = 0,915. \end{aligned}$$

Пример 2:

10 стран, проранжированных по уровню жизни и качеству

институтов, $x^{(1)} = (1\ 2,5\ 2,5)(4,5\ 4,5)(6,5\ 6,5) 8\ (9,5\ 9,5)$

$x^{(2)} = ((1,5\ 1,5)(4,5\ 4,5\ 4,5\ 4,5)(8\ 8\ 8) 10)$

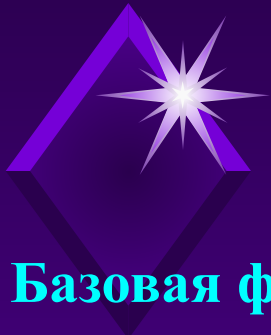
$$T^{(1)} = \frac{1}{12} (2^3 - 2) \cdot 4 = \frac{24}{12} = 2,$$

$$T^{(2)} = \frac{1}{12} \left((2^3 - 2) + (4^3 - 4) + (3^3 - 3) \right) = 7,5,$$

$$\hat{r}_{12}^{(S)} = \frac{\frac{990}{6} - (0,25 + 1 + \dots + 0,25) - 2 - 7,5}{\sqrt{\left(\frac{990}{6} - 2 \cdot 2 \right) \left(\frac{990}{6} - 2 \cdot 7,5 \right)}} = 0,911.$$

Недостатки коэффициента Спирмена:

1. Недостаточная изученность статистических свойств.
2. Невозможность построения частных коэффициентов корреляции.
3. Необходимость полного пересчета при добавлении объекта.



Ранговый коэффициент корреляции Кендалла

12

Базовая формула:

$$\hat{r}_{kj}^{(K)} = 1 - \frac{4\nu(x^{(k)}, x^{(j)})}{n(n-1)}, \quad \nu(x^{(k)}, x^{(j)}) - \text{минимальное число обменов соседних элементов переменной } x^{(j)} \text{ для ее приведения к виду } x^{(k)}.$$

Свойства коэффициента Кендалла:

$$\begin{aligned} \hat{r}_{kj}^{(K)} &= 1, \text{ если } x_i^{(k)} = x_i^{(j)}, i = 1, \dots, n, \\ \hat{r}_{kj}^{(K)} &= -1, \text{ если } x_i^{(k)} + x_i^{(j)} = n + 1, i = 1, \dots, n, \nu = n(n-1)/2, \\ \hat{r}_{kj}^{(K)} &\in (-1; 1) \text{ в остальных случаях.} \end{aligned}$$

Расчет числа обменов неудобен, ν – также число инверсий (число расположенных в разном порядке пар элементов из $x^{(k)}$ и $x^{(j)}$).

Удобно произвести сортировку данных по одной из переменных!

$$\begin{pmatrix} 4 \\ 3 \\ 2 \\ 1 \end{pmatrix} \rightarrow \begin{pmatrix} 4 \\ 3 \\ 1 \\ 2 \end{pmatrix} \rightarrow \begin{pmatrix} 4 \\ 1 \\ 3 \\ 2 \end{pmatrix} \rightarrow \begin{pmatrix} 1 \\ 4 \\ 3 \\ 2 \end{pmatrix} \rightarrow \begin{pmatrix} 1 \\ 4 \\ 2 \\ 3 \end{pmatrix} \rightarrow \begin{pmatrix} 1 \\ 2 \\ 4 \\ 3 \end{pmatrix} \rightarrow \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix} \Leftrightarrow \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix} \begin{pmatrix} 4 \\ 3 \\ 2 \\ 1 \end{pmatrix} \begin{matrix} 3 \\ 2 \\ 1 \\ 0 \end{matrix} \quad \nu = 6, \quad \hat{r}_{12}^{(K)} = 1 - \frac{4 \cdot 6}{4 \cdot 3} = -1.$$



Ранговый коэффициент корреляции Кендалла

13

Формула для случая объединенных рангов:

$$\hat{r}_{kj}^{*(K)} = \frac{\hat{r}_{kj}^{(K)} - \frac{2(u^{(k)} + u^{(j)})}{n(n-1)}}{\sqrt{\left(1 - \frac{2u^{(k)}}{n(n-1)}\right)\left(1 - \frac{2u^{(j)}}{n(n-1)}\right)}}, \quad u^{(k)} = \frac{1}{2} \sum_{t=1}^{m^{(k)}} n_t^{(k)} (n_t^{(k)} - 1).$$

Пример 1:

$$x^{(1)} = (1\ 2\ 3\ 4\ 5\ 6\ 7\ 8\ 9\ 10)$$

$$x^{(2)} = (2\ 3\ 1\ 4\ 6\ 5\ 9\ 7\ 8\ 10)$$

$$1\ 1\ 0\ 0\ 1\ 0\ 2\ 0\ 0\ 0$$

$$\nu(x^{(1)}, x^{(2)}) = 5,$$

$$\hat{r}_{12}^{(K)} = 1 - \frac{4 \cdot 5}{90} = 0,778.$$

Пример 2:

$$x^{(1)} = (1\ (2,5\ 2,5)(4,5\ 4,5)(6,5\ 6,5)\ 8\ (9,5\ 9,5))$$

$$x^{(2)} = ((1,5\ 1,5)(4,5\ 4,5\ 4,5\ 4,5)(8\ 8\ 8)\ 10)$$

$$u^{(1)} = \frac{1}{2}(4 \cdot 2 \cdot 1) = 4, \quad u^{(2)} = \frac{1}{2}(2 \cdot 1 + 4 \cdot 3 + 3 \cdot 2) = 10,$$

$$\hat{r}_{12}^{(K)} = \frac{90}{\sqrt{\left(1 - \frac{2 \cdot 4}{90}\right)\left(1 - \frac{2 \cdot 10}{90}\right)}} = 0,818.$$

Проверка гипотезы о наличии связи между порядковыми переменными

Связь есть, если $|\hat{r}^{(S)}| > t\left(\frac{\alpha}{2}; n-2\right) \sqrt{\frac{1 - (\hat{r}^{(S)})^2}{n-2}}$

или $|\hat{r}^{(K)}| > u\left(1 - \frac{\alpha}{2}\right) \sqrt{\frac{2(2n+5)}{9n(n-1)}}$.

$0,915 > \text{СТЮДРАСПОБР}(0,025; 8) \sqrt{(1 - 0,915^2)/8}$, **$0,915 > 0,392$** .

$0,778 > \text{НОРМСТОБР}(0,975) \sqrt{\frac{2 \cdot (2 \cdot 10 + 5)}{9 \cdot 10 \cdot (10 - 1)}}$, **$0,778 > 0,487$** .

Неравенства утверждают, что связь есть при уровне значимости $\alpha = 0,05$.

Доверительный интервал для коэффициента Кендалла

$$r_{kj}^{(K)} \in \left[\hat{r}_{kj}^{(K)} - u\left(\frac{1+\gamma}{2}\right) \sqrt{\frac{2}{n} \left(1 - (\hat{r}_{kj}^{(K)})^2\right)}; \hat{r}_{kj}^{(K)} + u\left(\frac{1+\gamma}{2}\right) \sqrt{\frac{2}{n} \left(1 - (\hat{r}_{kj}^{(K)})^2\right)} \right]$$

Интервал приближенный, формулу использовать для больших выборок!



Связь между несколькими порядковыми переменными

15

Коэффициент конкордации:

$$\hat{W}(m) = \frac{12}{m^2(n^3 - n)} \sum_{i=1}^n \left(\sum_{j=1}^m x_i^{(k_j)} - \frac{m(n+1)}{2} \right)^2, \quad n - \text{число объектов,}$$

$m - \text{число переменных,}$
 $k_1, \dots, k_m - \text{номера переменных.}$

$$\hat{W}(m) = \frac{\sum_{i=1}^n \left(\sum_{j=1}^m x_i^{(k_j)} - \frac{m(n+1)}{2} \right)^2}{\frac{1}{12} m^2 (n^3 - n) - m \sum_{j=1}^m T^{(k_j)}} - \text{при наличии объединенных рангов.}$$

Свойства коэффициента конкордации:

$$W(m) \in [0; 1],$$

$W(m) = 1$ при полном совпадении переменных,

$W(m) = 0$, когда распределение случайно.

Коэффициент конкордации не может быть отрицательным:

если $x_i^{(2)} = n - x_i^{(1)} + 1$, $x_i^{(3)} = n - x_i^{(1)} + 1$, то $x_i^{(2)} = x_i^{(3)}$.



Численный пример

16

Ранжировка 10 инвестиционных проектов, осуществленная 3 экспертами.

$x^{(1)}$	$x^{(2)}$	$x^{(3)}$	Σ	$\Sigma - m\bar{x}$
1	2,5	2	5,5	-11
4,5	1	1	6,5	-10
2	2,5	4,5	9	-7,5
3	4,5	4,5	12	-4,5
4,5	4,5	4,5	13,5	-3
7,5	8	4,5	20	3,5
6	9	8	23	6,5
9	6,5	8	23,5	7
7,5	10	8	25,5	9
10	6,5	10	26,5	10

$$\begin{array}{l|l} 2 & T^{(1)} = \frac{1}{12}(2^3 - 2) \cdot 2 = 1, \\ 2 & T^{(2)} = \frac{1}{12}(2^3 - 2) \cdot 3 = 1,5, \\ 4 & T^{(3)} = \frac{1}{12}((4^3 - 4) + (3^3 - 3)) = 7. \end{array}$$

$$\hat{W}(3) = \frac{((-11)^2 + (-10)^2 + \dots + 10^2)}{\frac{1}{12} \cdot 3^2 \cdot (10^3 - 10) - 3 \cdot (1 + 1,5 + 7)} = 0,828.$$



Проверка гипотезы о наличии связи между несколькими порядковыми переменными

17

Связь есть, если $m(n-1)\hat{W}(m) > \chi^2(\alpha; n-1) = \text{ХИ2ОБР}(\alpha; n-1)$.

Пример 1:

$$\chi_{\text{эмп}}^2 = 3 \cdot 9 \cdot 0,828 = 22,35, \quad \chi_{\text{крит}}^2 = \text{ХИ2ОБР}(0,05; 9) = 16,92.$$

$22,35 > 16,92 \Rightarrow$ связь между 3 переменными есть при $\alpha = 0,05$.

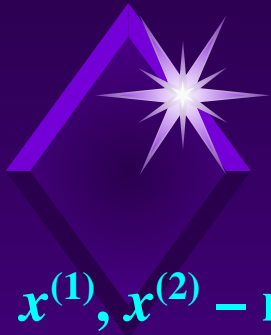
Пример 2:

$$m = 28, \quad n = 13, \quad \hat{W}(28) = 0,08$$

$$\chi_{\text{эмп}}^2 = 28 \cdot 12 \cdot 0,08 = 26,88, \quad \chi_{\text{крит}}^2 = \text{ХИ2ОБР}(0,05; 12) = 21,03$$

$26,88 > 21,03 \Rightarrow$ связь между 28 переменными есть при $\alpha = 0,05$.

Замечание: при большом количестве переменных даже малого значения коэффициента конкордации достаточно для вывода о наличии связи.



Корреляционный анализ категоризованных переменных

18

$x^{(1)}, x^{(2)}$ – категоризованные переменные (переменные, описываемые конечным числом состояний).

пол, социальная страта, сезон, фирма-производитель, ...

Таблица сопряженности:

	1	2	...	m_2		
1	n_{11}	n_{12}	...	n_{1m_2}	$n_{1\bullet} = \sum n_{1j}$	$w_{1\bullet} = n_{1\bullet}/n$
2	n_{21}	n_{22}	...	n_{2m_2}	$n_{2\bullet} = \sum n_{2j}$	$w_{2\bullet} = n_{2\bullet}/n$
...
m_1	n_{m_11}	n_{m_12}	...	$n_{m_1m_2}$	$n_{m_1\bullet} = \sum n_{m_1j}$	$w_{m_1\bullet} = n_{m_1\bullet}/n$
	$n_{\bullet 1} = \sum n_{j1}$	$n_{\bullet 2} = \sum n_{j2}$...	$n_{\bullet m_2} = \sum n_{jm_2}$	n	
	$w_{\bullet 1} = n_{\bullet 1}/n$	$w_{\bullet 2} = n_{\bullet 2}/n$...	$w_{\bullet m_2} = n_{\bullet m_2}/n$		

Статистическая независимость переменных: $\frac{n_{ij}}{n} \approx \frac{n_{i\bullet}}{n} \frac{n_{\bullet j}}{n}$, $w_{ij} \approx w_{i\bullet} w_{\bullet j}$.

Чем больше отклонение, тем больше показатель связи: $\delta_{ij} = w_{ij} - w_{i\bullet} w_{\bullet j}$.



Случай тесной связи

19

и независимости переменных

$x^{(1)}$ – пол (муж/жен), $x^{(2)}$ – уровень зарплаты (высокая/низкая), $n = 100$.

50	0	50	0,5
0	50	50	0,5
50	50	100	
0,5	0,5		

$$\Delta = \begin{pmatrix} 0,25 & -0,25 \\ 0,25 & 0,25 \end{pmatrix}.$$

Максимально тесная связь, знание значения одной переменной позволяет восстановить значение другой.

25	25	50	0,5
25	25	50	0,5
50	50	100	
0,5	0,5		

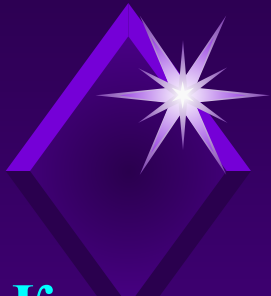
$$\Delta = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

Полное отсутствие связи, знание значения одной переменной не позволяет сделать никаких выводов о значении другой.

28	42	70	0,7
12	18	30	0,3
40	60	100	
0,4	0,6		

$$\Delta = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

Полное отсутствие связи, знание значения одной переменной не позволяет сделать никаких выводов о значении другой.



Квадратичная сопряженность – характеристика тесноты связи

20

Квадратичная сопряженность: два способа расчета:

$$\hat{X}^2 = n \sum_{i=1}^{m_1} \sum_{j=1}^{m_2} \frac{\delta_{ij}^2}{w_{i \cdot} w_{\cdot j}} = n \left(\sum_{i=1}^{m_1} \sum_{j=1}^{m_2} \frac{n_{ij}^2}{n_{i \cdot} n_{\cdot j}} - 1 \right), \quad \hat{X}^2 \in [0; +\infty)$$

Проверка гипотезы о наличии связи:

$$\hat{X}^2 > X^2(\alpha; (m_1 - 1)(m_2 - 1)) \Rightarrow \text{связь есть при уровне значимости } \alpha.$$

Коэффициент Крамера:

Недостатком квадратичной сопряженности является неограниченность ее значений: при $n \rightarrow \infty$ $X^2 \rightarrow \infty$. Следовательно, желательно построить другой показатель, находящийся в привычном диапазоне $[0; 1]$.

$$\hat{C} = \sqrt{\frac{\hat{X}^2}{n \min\{m_1 - 1; m_2 - 1\}}} \in [0; 1]$$



Численный пример

Зависимость оплаты труда (низкая; средняя; высокая) **от образования** (неполное среднее; среднее; среднее специальное; высшее; высшее со степенью), $n = 300$.

14	28	13	18	3	76	0,25
11	48	35	51	17	162	0,54
1	10	7	19	25	62	0,21
26	86	55	88	45	300	
0,09	0,29	0,18	0,29	0,15		

Равномерное распределение

7	22	14	22	11
14	46	30	48	24
5	18	11	18	9

$$\Delta = \begin{pmatrix} 0,025 & 0,021 & -0,013 & -0,014 & -0,028 \\ -0,010 & 0,005 & 0,018 & 0,012 & -0,024 \\ -0,015 & -0,026 & -0,015 & 0,003 & 0,052 \end{pmatrix} \frac{\delta_{ij}^2}{w_{i \cdot} w_{\cdot j}} = \begin{pmatrix} 0,028 & 0,006 & 0,000 & 0,003 & 0,021 \\ 0,002 & 0,000 & 0,003 & 0,001 & 0,007 \\ 0,012 & 0,011 & 0,006 & 0,000 & 0,088 \end{pmatrix}$$

$$\hat{X}^2 = 300 \cdot 0,188 = 56,48, \quad \hat{C} = \sqrt{\frac{56,48}{300 \min\{4; 2\}}} = 0,307.$$

$\chi_{\text{крит}}^2 = \text{ХИ2ОБР}(0,001; 4 \cdot 2) = 26,12$, **56,48 > 26,12** \Rightarrow связь есть при $\alpha=0,001$.



*Спасибо
за внимание!*

alexander.filatov@gmail.com

<http://vk.com/alexander.filatov>, <http://vk.com/baikalreadings>