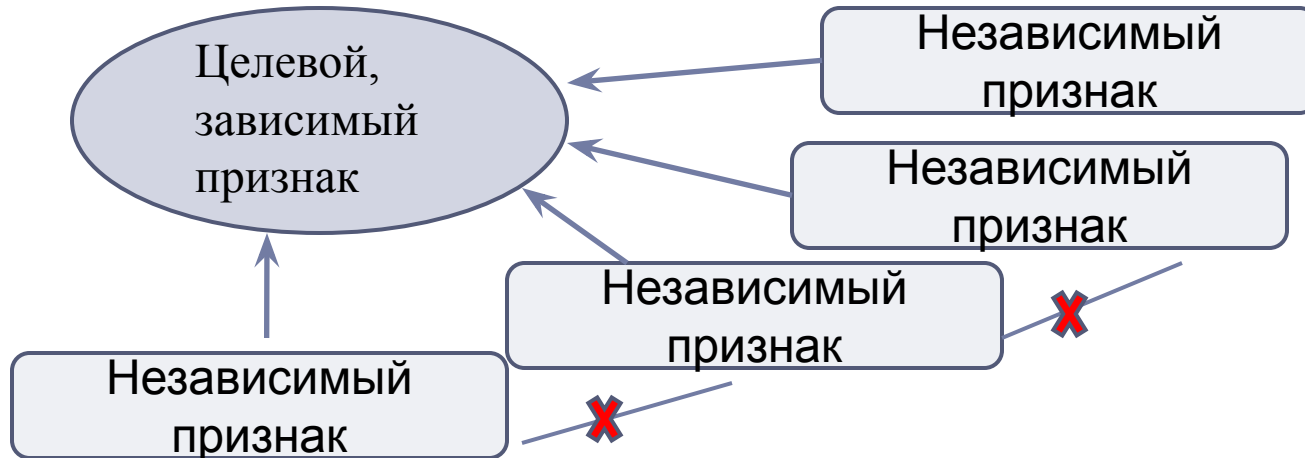


Меры связи, основанные  
на статистической зависимости.  
Коэффициент Юла

Курс «Анализ социологических данных», магистратура, 1 г.о.

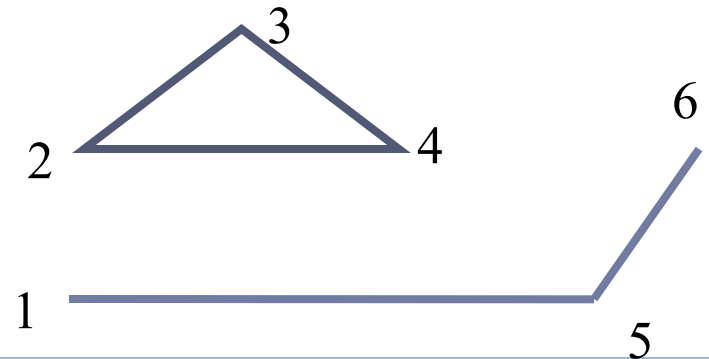
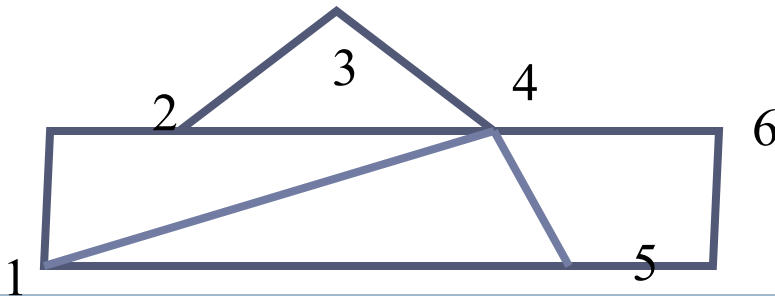
© Пузанова Ж.В., 2017

# Логика использования парных коэффициентов связей



I.  $K_3 > K_2 > K_4 > K_1 \rightarrow$  гипотеза о причинно-следственных СВЯЗЯХ

II. Множество независимых признаков: граф структуры взаимосвязей



# Анализ данных в таблицах сопряженности 2x2

Например: опрошены студенты социологи

	Уверены в устройстве на работу	Не уверены в устройстве на работу	Итого
Магистры	15 (a)	35 (d)	50
Бакалавры	110 (c)	340 (b)	450
Итого	125	375	500

Коэффициент Юла:

$$Q = \frac{ab - cd}{ab + cd}$$

**Есть ли статистическая закономерность между этими признаками?**

Если наблюдается статистическая *независимость*, то ...

доля магистров в группе «уверенных» должна равняться доле магистров в группе «неуверенных», т.е.

$$\frac{a}{a+c} = \frac{d}{d+b} \quad \frac{a}{a+d} = \frac{c}{c+b}$$

, т.е.  $a(d+b) = (a+c)d \rightarrow ab = cd$ ; и  $a(c+b) = (a+d)c \rightarrow ab = cd$

**$ab - cd$**  - мера отклонения от статистической независимости;

**Знаменатель** – для нормирования (изменения от  $-1$  до  $+1$ )



Свойства коэффициента Юла:  $Q = \frac{ab - cd}{ab + cd}$

---

- $Q = 1$ , если  $c = 0$  или  $d = 0$
- $Q = -1$ , если  $a = 0$  или  $b = 0$
- $Q = 0$ , если  $ab = cd$ ,

	У	Н
Маг	<b>a</b>	<b>d</b>
Бак	<b>c</b>	<b>b</b>

т.е. наблюдается статистическая независимость

$$Q = 0,1397 \sim 0,14$$



# Использование Коэф. Юла в сравнительном контексте

Специальность	уверенность в работе по специальности					
	5	4	3	2	1	n <sub>io</sub>
СОЦИОЛОГ	3	15	15	10	7	50
ПОЛИТОЛ	5	30	30	20	15	100
МО	7	23	30	45	45	150
ИСТОРИК	3	7	10	15	15	50
ФИЛОСОФ	1	8	7	5	4	25
ГМУ	6	42	8	55	14	125
n <sub>oj</sub>	25	125	100	150	100	500

Социол

Политол

МО

Истор

Филос

ГМУ

+ Y	- Y
18	32
132	318

+ Y	- Y
35	65
115	285

+ Y	- Y
30	120
120	230

+ Y	- Y
10	40
140	310

+ Y	- Y
9	16
141	334

+ Y	- Y
48	77
102	273

$$Q_c = 0,15 ; Q_{\Pi} = 0,14; Q_{MO} = - 0,35; Q_{Ис} = - 0,29; Q_{\Phi} = 0,14; Q_{ГМУ} = 0,25$$



# Сравнение по коэффициенту:

---

- $Q_c = 0,15$  ;
- $Q_{\Pi} = 0,14$ ;
- $Q_{\text{МО}} = - 0,35$ ;
- $Q_{\text{ИС}} = - 0,29$ ;
- $Q_{\Phi} = 0,14$ ;
- $Q_{\text{ГМУ}} = 0,25$
- Коэффициенты упорядочиваются по силе связи (без учета прямая или обратная зависимость), т.о.:
- $Q_{\text{МО}} > Q_{\text{ИС}} > Q_{\text{ГМУ}} > Q_c > Q_{\Pi} > Q_{\Phi}$

---

□  
▶ **Связь (в данном контексте) – статистическая зависимость**

# Понятие детерминации для локальной связи

□ Если **A**, то **B** - правило детерминации:

□ одно свойство оказывает влияние на другое:  $A \rightarrow B$ .

□ *Например: Социолог  $\rightarrow$  Уверенность в работе.*

□ Характеристики детерминации:

□ **Интенсивность :**

$$I(C \rightarrow Y) = \frac{a}{a + d}$$

□ **Емкость:**

$$C(C \rightarrow Y) = \frac{a}{a + c}$$

□ Для социологов:  $a = 18$ ;  $a + d = 50$ ;  $a + c = 150$

□  $I(C \rightarrow Y) = 36\%$ ,  $C(C \rightarrow Y) = 12\%$

+ Y	- Y
18	32
132	318

□ *Интерпретация детерминации (схема):*

Интенсивность	$I \sim 0$	$I \sim 1$	$I \sim 0$	$I \sim 1$
Емкость	$C \sim 0$	$C \sim 0$	$C \sim 1$	$C \sim 1$
Детерминация	Неточная и неполная	Точная, но неполная	Неточная, но полная	Точная, полная

# Непосредственная – опосредованная связь

□  $Q_c = 0,15$  – указывает статистическую независимость

□ для всех

для девушек

для юношей

	+ Y	- Y	
C	18	32	50
HC	132	318	450
	150	350	500

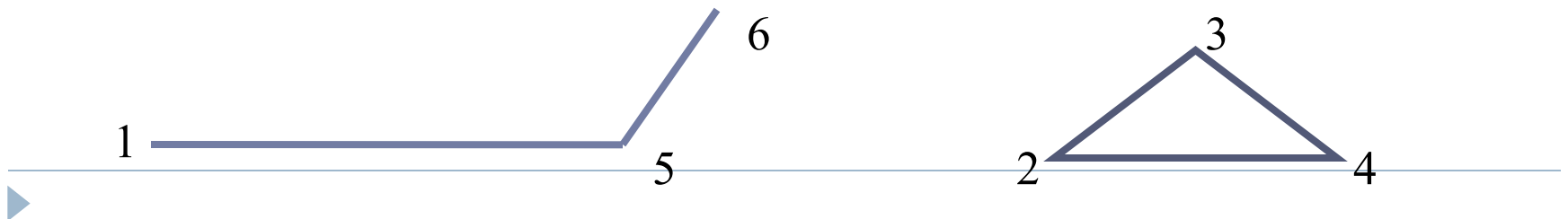
	+ Y	- Y	
C	12	8	20
HC	12	234	246
	24	242	266

	+ Y	- Y	
C	6	24	30
HC	120	84	204
	126	108	234

$$Q_{ж} = \frac{12 * 234 - 12 * 8}{12 * 234 + 12 * 8} \sim 0,93$$

$$Q_{м} = \frac{6 * 84 - 120 * 24}{6 * 84 + 120 * 24} \sim - 0,70$$

□  $Q_c = 0,15$  – ложное отсутствие корреляционной связи





# Меры связи, основанные на $\chi^2$

- Основывается на статистической независимости (как и коэф. Юла)
- Для любой ячейки таблицы (r, s) в случае статистической независимости должно выполняться соотношение:

□ 
$$n_{ij} = \frac{n_{oj} n_{io}}{n_{oo}}$$
, т.к.  $\frac{n_{ij}}{n_{oj}} = \frac{n_{io}}{n_{oo}}$ , например:

$$n_{14} = \frac{125 * 50}{500} = 12,5$$

Специальность (i)	уверенность в работе по специальности (j)					n io
	1	2	3	4	5	
СОЦИОЛОГ	7	10	15	15	3	50
ПОЛИТОЛ	15	20	30	30	5	100
МО	45	45	30	23	7	150
истор	15	15	10	7	3	50
филос	4	5	7	8	1	25
ГМУ	14	55	8	42	6	125
n oj	100	150	100	125	25	500

# Реальные и теоретические частоты (для $\chi^2$ )

Спец-ть (i)	уверенность в работе по специальности (j)										n <sub>io</sub>
	1	1 <sub>т</sub>	2	2 <sub>т</sub>	3	3 <sub>т</sub>	4	4 <sub>т</sub>	5	5 <sub>т</sub>	
СОЦИОЛОГ	7	<b>10</b>	10	<b>15</b>	15	<b>10</b>	15	<b>12,5</b>	3	<b>2,5</b>	50
ПОЛИТОЛ	15	<b>20</b>	20	<b>30</b>	30	<b>20</b>	30	<b>25</b>	5	<b>5</b>	100
МО	45	<b>30</b>	45	<b>45</b>	30	<b>30</b>	23	<b>37,5</b>	7	<b>7,5</b>	150
истор	15	<b>10</b>	15	<b>15</b>	10	<b>10</b>	7	<b>12,5</b>	3	<b>2,5</b>	50
филос	4	<b>5</b>	5	<b>7,5</b>	7	<b>5</b>	8	<b>6,25</b>	1	<b>1,25</b>	25
ГМУ	14	<b>25</b>	55	<b>37,5</b>	8	<b>25</b>	42	<b>31,25</b>	6	<b>6,25</b>	125
n <sub>oj</sub>	100		150		100		125		25		500

$$n_{ij} = \frac{n_{oj} n_{io}}{n_{oo}}$$

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^s \frac{(n_{ij} - n_{ij}^t)^2}{n_{ij}^t}$$

►  $\chi^2 = \frac{(7-10)^2}{10} + \frac{(10-15)^2}{15} + \dots + \frac{(6-6,25)^2}{6,25} = \dots$

# Коэффициент взаимной сопряженности Пирсона (на основе критерия $\chi^2$ )

---

- $\chi^2_t$  – табличное значение коэффициента (находим по таблице, при заданном уровне значимости (возможной ошибки  $\alpha = 0,1; 0,05; 0,01$ ) и числе степеней свободы  $(r-1)(s-1)$  (где  $r$  – число строк,  $s$  – число столбцов в таблице)
- **ВЫВОД** о статистической значимости: Если  $\chi^2 > \chi^2_t$  при заданном уровне значимости (возможной ошибки  $\alpha = 0,1; 0,05; 0,01$ ) и числе степеней свободы  $(r-1)(s-1)$  – признаки являются статистически зависимыми...
- Наиболее часто используемый коэффициент на основе распределения  $\chi^2$  :

$$C = \sqrt{\frac{\chi^2}{\chi^2 + N}}$$

- $C$  – приближается к **0** или **1**
- 



## Домашнее задание (на любом примере из исследования практикума):

---

- Выявить локальные связи на основе использования коэффициента Юла. **Описать результаты**
- Вычислить интенсивность и емкость детерминации для нескольких свойств (в этом же примере). **Описать результаты.**
- Вычислить значение хи-квадрат и сравнить с табличным **при различных уровнях значимости** (НАЙТИ ТАБЛИЦУ ХИ-КВАДРАТ САМОСТОЯТЕЛЬНО). **Описать результаты**



# Примеры для Д/З (если не нашли примеры в Практикуме)

- Нравится ли Вам в целом обучение в РУДН:

	курс			
	1	2	3	4
Да, нравится	60	53	30	25
Скорее да, чем нет	22	30	53	35
И да, и нет	8	7	5	20
Скорее нет, чем да	5	6	7	14
Нет, не нравится	5	5	5	7

	факультет								
	инж	физмат	эколог	аграр	мед	филол	юрид	экон	ФГСН
Да, нравится	39	28	15	23	37	29	14	18	47
Скорее да, чем нет	43	20	15	17	35	32	20	23	32
Затрудняюсь ответить	12	14	14	10	13	11	13	12	17
Скорее нет, чем да	14	15	14	13	13	11	11	11	11
Нет, не нравится	11	10	10	10	10	12	10	10	10
Всего									

