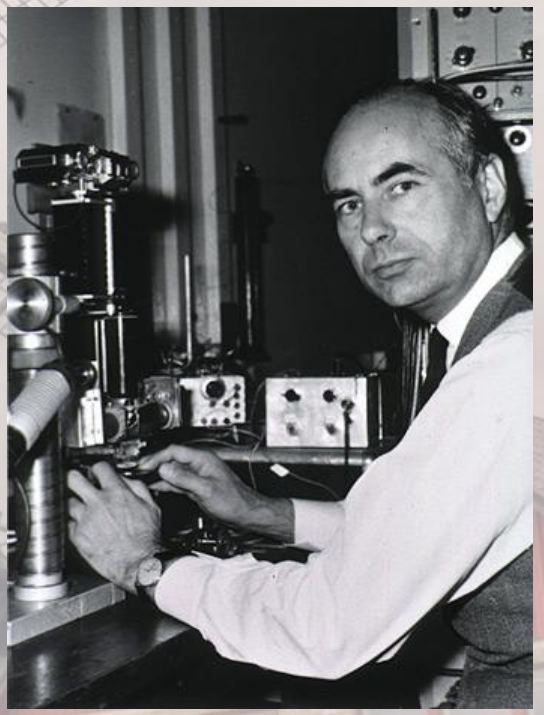
КЛІОМЕТРИКА

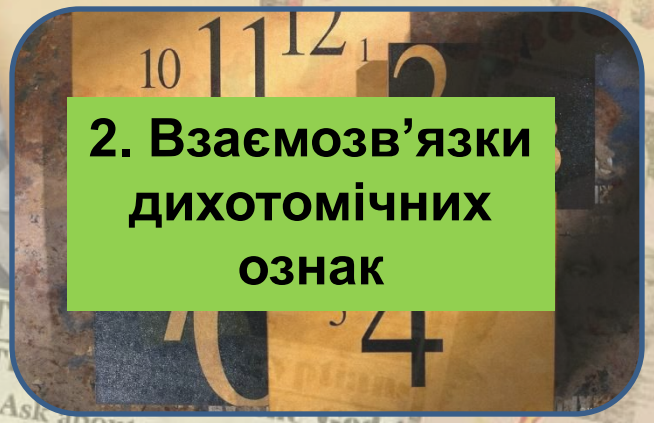
Розробник:
доктор історичних наук,
професор,
завідувач кафедри історії України
СВЯТЕЦЬ Юрій Анатолійович

Математика схожа на млин:
якщо ви засиплете
до неї зерна пшениці,
то одержите борошно,
якщо ж засиплете висівки,
то висівки й отримаєте
(Андру Філлінг Хакслі,
англійський фізіолог)

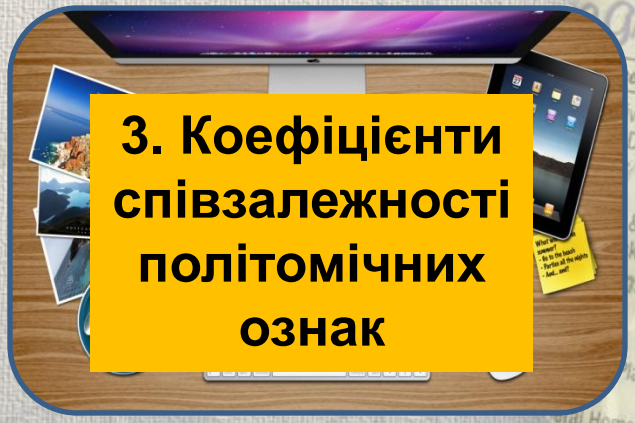


Тема: ОЦІНЮВАННЯ ВЗАЄМОЗВ’ЯЗКІВ НОМІНАЛЬНИХ ОЗНАК

1. Класифікація ознак



2. Взаємозв’язки дихотомічних ознак



3. Коефіцієнти співзалежності політомічних ознак

ОСНОВНА ЛІТЕРАТУРА

Количественные методы в исторических исследованиях: [учеб. пособ.] / под ред. И. Д. Ковальченко. – М.: Высш. шк., 1984. – 384 с.

Миронов, Б. Н. История в цифрах. Математика в исторических исследованиях: [учеб. пособ.] / Б. Н. Миронов; под ред. И. Д. Ковальченко. – Л.: Наука, 1991. – 168 с.

Святець, Ю. А. Кліометрика. Формально-кількісні та математико-статистичні методи: [підручник] / Ю. А. Святець. –Д.: Вид-во Дніпропетр. ун-ту, 2003. – 384 с.

Славко, Т. И. Математико-статистические методы в исторических исследованиях: [учеб. пособ.] / Т. И. Славко; под ред. И. Д. Ковальченко. – М.: МГУ, 1981. – 234 с.

ДОДАТКОВА ЛІТЕРАТУРА

Математика

Аптон, Г. Анализ таблиц сопряженности [Текст] / Г. Аптон; пер. с англ. и предисл. Ю. П. Адлера. – М.: Финансы и статистика, 1982. – 143 с.

Миркин, Б. Г. Группировка в социально-экономических исследованиях [Текст] / Б. Г. Миркин. – М.: Финансы и статистика, 1985. – 223 с.

Миркин, Б. Г. Анализ качественных признаков и структур [Текст] / Б. Г. Миркин. – М.: Статистика, 1980. – 319 с.

Флейс, Дж. Статистические методы для изучения таблиц долей и пропорций [Текст] / Дж. Флейс; под ред. Ю. Н. Благовещенского. – М.: Финансы и статистика, 1989. – 319 с.

ДОДАТКОВА ЛІТЕРАТУРА

Історія

Арутюнян, Ю. В. Опыт применения частотного анализа качественных признаков в этносоциологическом исследовании [Текст] / Ю. В. Арутюнян, Л. Г. Бадалян // Комплексные методы в изучении исторических процессов: [сб. ст.] / отв. ред.: В. Е. Полетаев, И. Л. Корнаковский. – М.: Ин-т истории СССР АН СССР, 1987. – С. 92 –108.

Мурашев, А. А. К вопросу об использовании метода группировок при изучении промышленного производства капиталистической России (Опыт количественного анализа) [Текст] / А. А. Мурашев // Источниковедение массовых источников: [сб. ст.] / ред. кол.: С. В. Воронкова (отв. ред.) [и др.]. – М.: Изд-во Моск. ун-та, 1988. – С. 87 – 99.

Хвостова, К. В. Некоторые вопросы применения количественных методов при изучении социально-экономических явлений средневековья (по данным византийских источников XIII – XIV вв.) [Текст] / К. В. Хвостова // Математические методы в исторических исследованиях: [сб. ст.] / ред. кол.: И. Д. Ковальченко (отв. ред.) [и др.]. – М.: Наука, 1972. – С. 15 –30.

Web-ресурси

Количественные методы в исторических исследованиях: [учеб. пособ.] / под ред. И.Д. Ковальченко. – М.: Высш. шк., 1984: [Електрон. ресурс] / Спосіб доступу: URL: http://www.scribd.com/doc/36882712/Количественные-методы-1984

И.Д.Ковальченко. Методы исторического исследования: 2-е изд. / И.Д. Ковальченко. – М., 2003 : [Електрон. ресурс] / Спосіб доступу: URL: http://aik-sng.ru/node/273

Про що Ви довідаєтеся?

Чи можливо виявити зв’язки між якостями?

Як виявити дискримінацію жінок у праві здобувати освіту?

Скільки клітинок у “чотириклітинковій таблиці“ співзалежності?

Чому коефіцієнт співзалежності “присвячений” студентові?

Як розуміти поняття “міцний взаємозв’язок”?

У темі будуть розглянуті:

терміни:

дихотомія

співзалежність

номінальні ознаки

“чотириклітинкова” таблиця

коефіцієнти асоціації та контингентності

коефіцієнт хі-квадрат

персоналії

Крамер, Карл Гаральд

Стьюдент (Госсет, Вільям Сейлі)

Чупров, Олександр Олександрович

Юл, Джордж Едні

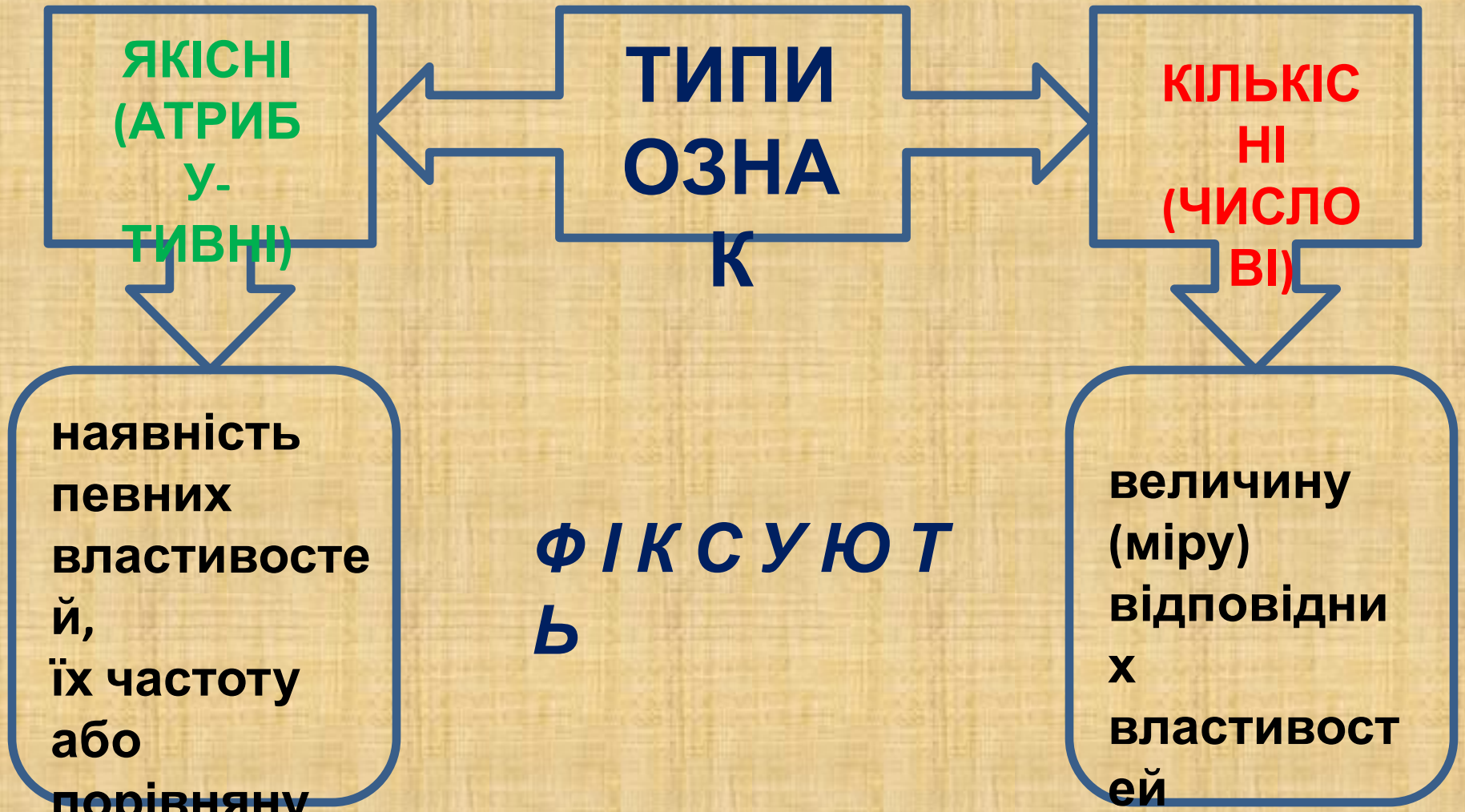
1. КЛАСИФІКАЦІЯ ОЗНАК

… двозначні, приблизні та псевдалі визначення нагадують класифікацію, яку доктор Франі Кун приписує одній китайській енциклопедії під назвою “Небесна імперія благодатних знань”. На її давніх сторінках написано, що тварини поділяються на а) таких, які належать Імператорові, б) набальзамованих, в) приручених, г) ссавців, д) сирен, е) казкових, є) окремих псів, ж) включених до цієї класифікації, з) таких, які несамовито бігають, і) незліченних, к) намальованих найтоншим пензлем з шерсті верблюда, л) інших, м) таких, які розбили вазу, н) схожих віддаль на мух.

Хорхе Луїс Борхес
Аналітична мова
Джона Вілкінса

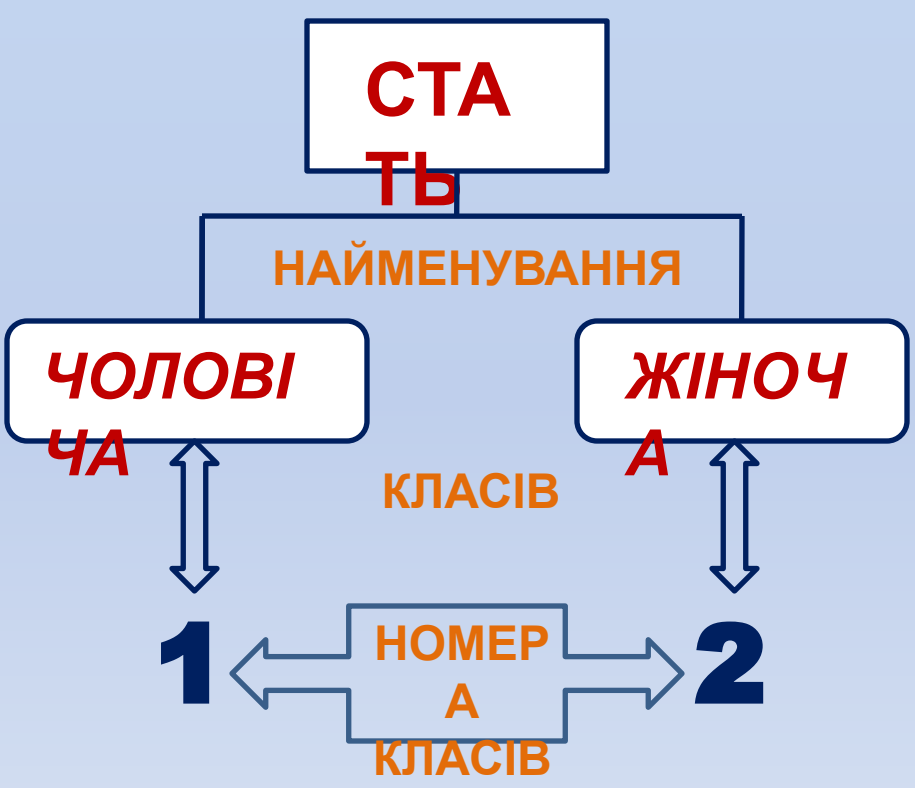
When Cicadas cry

КЛАСИФІКАЦІЯ ШКАЛ


ТИПИ ОЗНАК
ЯКІСНІ (АТРИБУ-ТИВНІ)
КІЛЬКІСНІ (ЧИСЛОВІ)
наявність певних властивостей, їх частоту або порівняну інтенсивність
ФІКСУЮТЬ
величину (міру) відповідних властивостей

КЛАСИФІКАЦІЯ ОЗНАК

ПРИКЛАДИ НОМІНАЛЬНИХ ОЗНАК



ПРИКЛАДИ НОМІНАЛЬНИХ ОЗНАК

ДИХОТОМІЧНІ

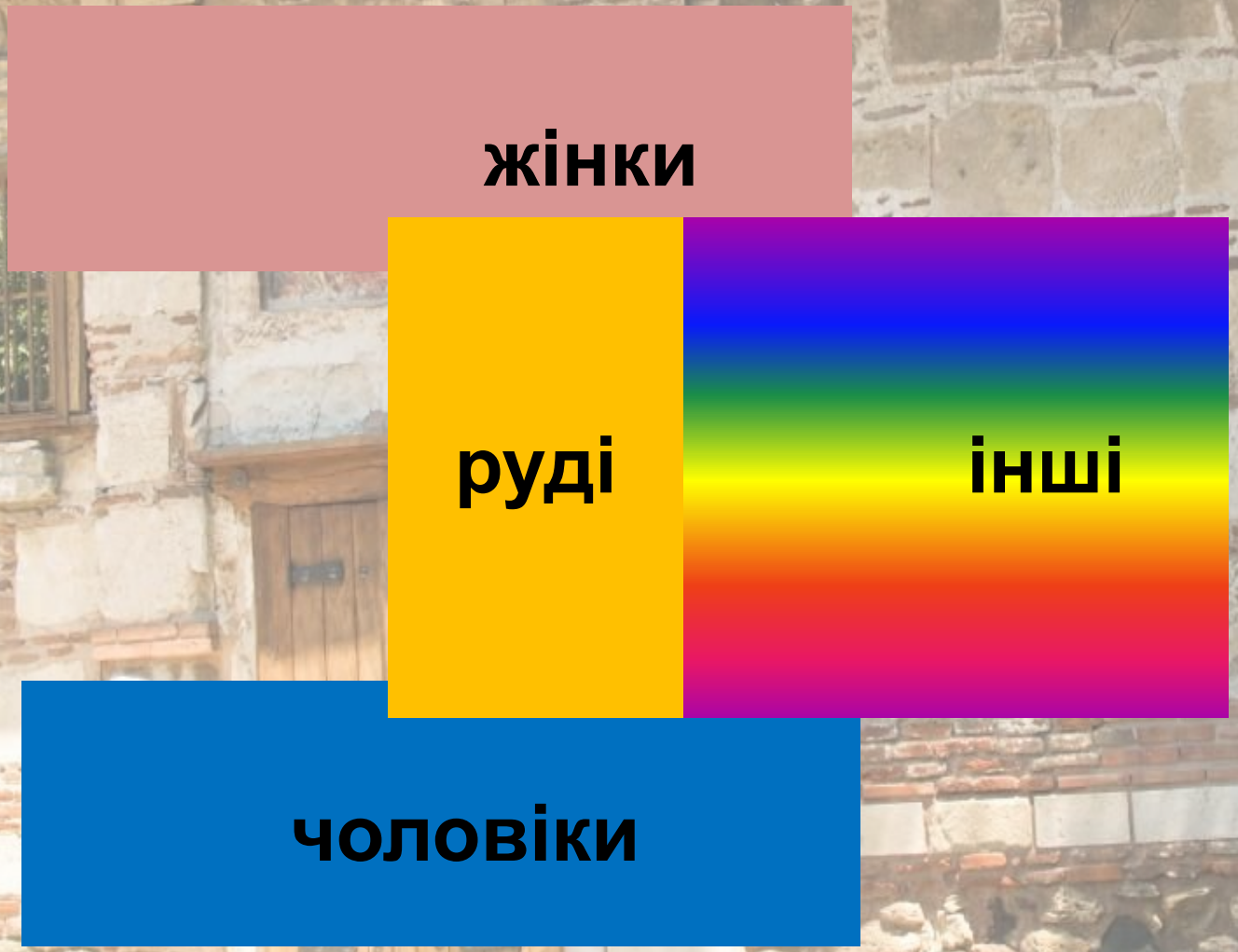
- стать
- шлюбний стан
- письменність
- судимість

ПОЛІТОМІЧНІ

- національність
- професія
- соціальна верства
- освіта
- партійність
- науковий ступінь

Дихотомічні (бінарні, альтернативні) ознаки

2. ВЗАЄМОЗВ’ЯЗКИ ДИХОТОМІЧНИХ ОЗНАК



“ЧОТРИКЛІТИНКОВА” ТАБЛИЦЯ СПІВЗАЛЕЖНОСТІ

Ознаки	B	¬B	Сума
A	a	b	a+b
¬A	c	d	c+d
Сума	a+c	b+d	a+b+c+d

Джордж Едні ЮЛ
(18.02.1871 – 26.06.1951)

КОЕФІЦІЄНТ АСОЦІАЦІЇ
(1900 рік)

$$Q = \frac{ad - bc}{ad + bc}$$

КОЕФІЦІЄНТ КОНТИНГЕНТНОСТІ

$$\varphi = \frac{ad - bc}{\sqrt{(a+b)(a+c)(b+d)(c+d)}}$$

ПРИКЛАД

Розподіл купців, міщан, почесних громадян та інших міських станів за статтю та письменністю

Ознаки	письменні	неписьменні	сума
жінки	1497	2544	4041
чоловіки	2109	1711	3820
сума	3606	4255	7861

$$Q = \frac{ad - bc}{ad + bc} = \frac{1497 \cdot 1711 - 2544 \cdot 2109}{1497 \cdot 1711 + 2544 \cdot 2109} = \frac{2561367 - 5365296}{2561367 + 5365296} = \frac{-2803929}{7926663} = -0{,}35$$

$$\varphi = \frac{ad - bc}{\sqrt{(a+b)(a+c)(b+d)(c+d)}} = \frac{1497 \cdot 1711 - 2544 \cdot 2109}{\sqrt{4041 \cdot 3606 \cdot 4255 \cdot 3820}} = \frac{-2803929}{15390004{,}62} = -0{,}18$$

КЛЮЧОВІ ЗНАЧЕННЯ КОЕФІЦІЄНТІВ СПІВЗАЛЕЖНОСТІ

- **$Q = 0$ – статистична незалежність ознак:**

$a=b=c=d$ -- усі об’єкти розподілені в однаковій кількості між усіма можливими поєднаннями ознак;

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

або $\frac{a}{c} = \frac{b}{d}$ -- однакові пропорції між ознаками

$Q = 1$ -- повний прямий взаємозв'язок:

$b = 0, c \neq 0$ основному значенню ознаки A

$c = 0, b \neq 0$ відповідає винятково основне

$b = 0, c = 0$ значення ознаки B

$Q = -1$ -- повний зворотний взаємозв'язок:

$a = 0, d \neq 0$ основному значенню ознаки A

$d = 0, a \neq 0$ відповідає винятково

$a = 0, d = 0$ альтернативне значення ознаки B

Кваліфікаційні інтервали для коефіцієнтів співзалежності

$0,9 \le |Q| \le 1,0$ -- тісний взаємозв’язок;

$0,7 < |Q| < 0,9$ -- взаємозв’язок середньої сили;

$0,5 \le |Q| \le 0,7$ -- слабкий взаємозв’язок;

$0,3 < |Q| < 0,5$ -- дуже слабкий взаємозв’язок;

$0,0 \le |Q| \le 0,3$ -- дуже слабкий взаємозв’язок.

3. Коефіцієнти співзалежності політомічних ознак

Ознака A	Ознака B						
	B_1	B_2	…	B_j	…	B_n	Сума
A_1	x_{11}	x_{12}	…	x_{1j}	…	x_{1n}	$\sum_{j=1}^{n} x_{1j}$
A_2	x_{21}	x_{22}	…	x_{2j}	…	x_{2n}	$\sum_{j=1}^{n} x_{2j}$
…	…	…	…	…	…	…	…
A_i	x_{i1}	x_{i2}	…	x_{ij}	…	x_{in}	$\sum_{j=1}^{n} x_{ij}$
…	…	…	…	…	…	…	…
A_m	x_{m1}	x_{m2}	…	x_{mj}	…	x_{mn}	$\sum_{j=1}^{n} x_{mj}$
Сума	$\sum_{i=1}^{m} x_{i1}$	$\sum_{i=1}^{m} x_{i2}$		$\sum_{i=1}^{m} x_{ij}$		$\sum_{i=1}^{m} x_{in}$	$N = \sum_{i=1}^{m}\sum_{j=1}^{n} x_{ij}$

КОЕФІЦІЄНТ СТЬЮДЕНТА (хі-квадрат)

Вільям Сейлі ГОССЕТ (Стьюдент) (13.06.1876 – 16.10.1937)

$$\chi^2 = N\left[\sum_{i=1}^{m}\sum_{j=1}^{n}\frac{x_{ij}^2}{X_{i\cdot}X_{\cdot j}} - 1\right]$$

$$X_{i\cdot} = \sum_{i=1}^{m} x_{ij} \qquad X_{\cdot j} = \sum_{j=1}^{n} x_{ij}$$

ПРИКЛАД

Соціальний стан	Тип господарств		Сума
	відробітковий	капіталістичний	
Селяни	5	25	30
Купці, міщани	10	40	50
Поміщики	50	10	60
Сума	65	75	140

$$\chi^2 = N\left[\sum_{i=1}^{m}\sum_{j=1}^{n}\frac{x_{ij}^2}{X_{i\cdot}X_{\cdot j}} - 1\right] =$$

$$= 140\cdot\left[\frac{5^2}{65\cdot 30} + \frac{25^2}{75\cdot 30} + \frac{10^2}{65\cdot 50} + \frac{40^2}{75\cdot 50} + \frac{50^2}{65\cdot 60} + \frac{10^2}{75\cdot 60} - 1\right] =$$

$$= 140\cdot[1{,}411 - 1] = 140\cdot 0{,}411 = 57{,}54$$

ПЕРЕВІРКА ГІПОТЕЗИ НА НАЯВНІСТЬ СПІВЗАЛЕЖНОСТІ

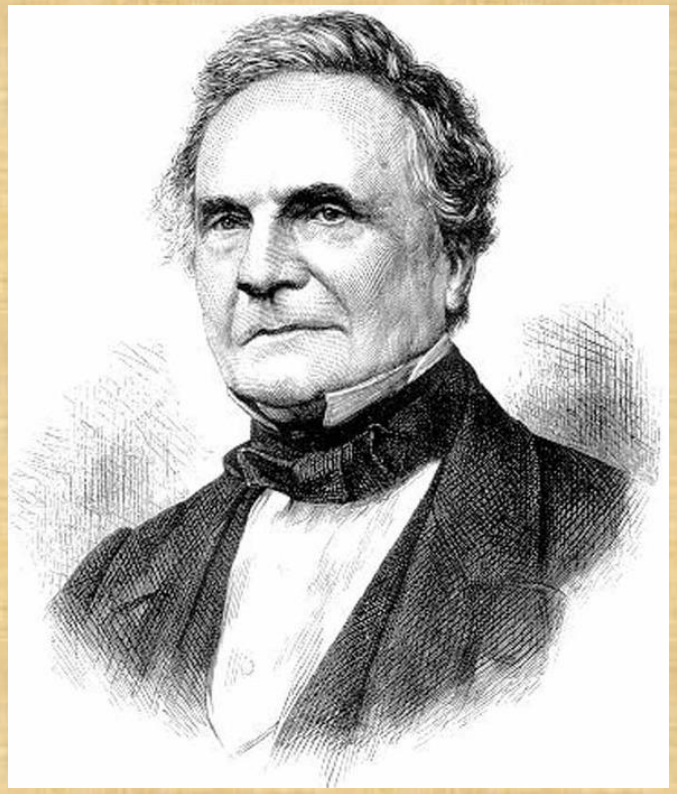
Таблиця критичних значень коефіцієнта Стюдента χ^2

Число степенів вільності, v	Рівень значущості, α	
	0,01	**0,001**
1	6,635	10,827
2	9,210	**13,815**
2	11,345	16,266
4	13,277	18,467
5	15,086	20,515

Правило. Якщо $\chi^2 > \chi^2_{кр}$, за рівня значущості α (ймовірності похибки) та кількості
степенів вільності v (v=(*m*-1)(*n*-1)), то це вказує на наявність статистичного взаємозв'язку між парою політомічних ознак.

Якщо α=0,001, а v=(*m*-1)(*n*-1)=(3-1)(2-1)=2, то $\chi^2_{кр}$ =13,815.
Отримане χ^2 = 57,54, що набагато більше за $\chi^2_{кр}$ = 13,815, що вказує на наявність
взаємозв'язку між соціальною верствою та типом господарства

КОЕФІЦІЄНТ ЧУПРОВА (1918 рік)



Олександр Олександрович ЧУПРОВ
(05.02.1874 – 19.04.1926)

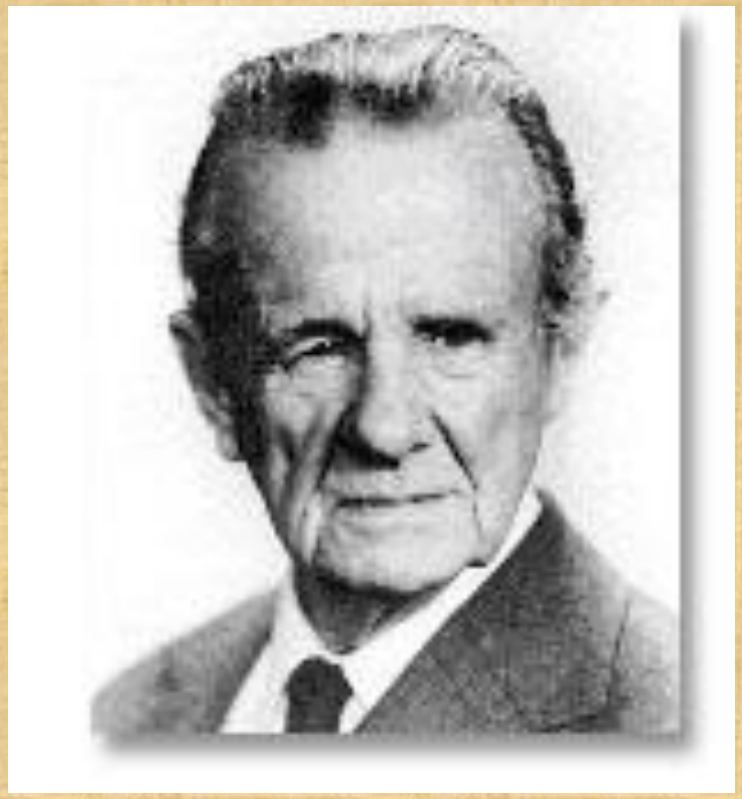
$$T^2 = \frac{\chi^2}{N \cdot \sqrt{\nu}} = \frac{\chi^2}{N \cdot \sqrt{(m-1)(n-1)}}$$

N – загальна кількість об’єктів;

m – кількість рядків у таблиці співзалежності (варіантів значень ознаки А);

n – кількість стовпців у таблиці співзалежності (варіантів

КОЕФІЦІЄНТ КРАМЕРА (1927 рік)



Гаральд КРАМЕР
(25.09.1893 – 05.10.1985)

$$K^2 = \frac{\chi^2}{N \cdot \min(m-1; n-1)}$$

N – загальна кількість об’єктів;

m – кількість рядків у таблиці співзалежності (варіантів значень ознаки А);

n – кількість стовпців у таблиці співзалежності (варіантів

ПРИКЛАД

Соціальний стан	Тип господарств		Сума
	відробітковий	капіталістичний	
Селяни	5	25	30
Купці, міщани	10	40	50
Поміщики	50	10	60
Сума	65	75	140

$$T^2 = \frac{\chi^2}{N \cdot \sqrt{(m-1)(n-1)}} = \frac{57{,}54}{140 \cdot \sqrt{(3-1)(2-1)}} = \frac{57{,}54}{140 \cdot \sqrt{2}} = \frac{57{,}54}{140 \cdot 1{,}414} = 0{,}291$$

$$K^2 = \frac{\chi^2}{N \cdot \min(m-1; n-1)} = \frac{57{,}54}{140 \cdot \min(3-1;2-1)} = \frac{57{,}54}{140} = 0{,}411$$

ВИПАДКИ ЗАСТОСУВАННЯ КОЕФІЦІЄНТІВ

- **коефіцієнт Чупрова** – за умови, що $n = m$

  (тільки за таких обставин T^2 може сягнути максимально можливого значення 1);

- **коефіцієнт Крамера** – в разі, якщо $n \neq m$

ВЛАСТИВОСТІ КОЕФІЦІЄНТІВ СПІВЗАЛЕЖНОСТІ

- $T^2 = K^2$ в разі, якщо $m = n$
- $T^2 < K^2$ за умови, що $m \neq n$
- обидва коефіцієнти дають значення в інтервалі від 0 до 1

ДЯКУЮ ЗА
УВАГУ!