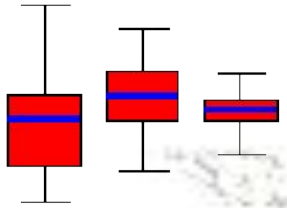
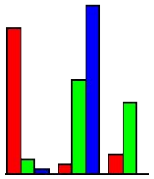
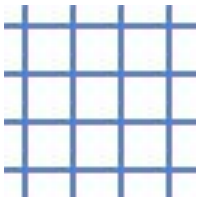
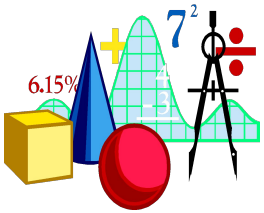






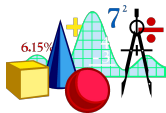


Таблицы сопряженности

Стат. методы в психологии (Радчикова Н.П.)

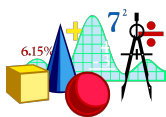
Цели

- **Вспомнить, что такое таблицы сопряженности**
- **Вспомнить, какую статистику можно для них считать**



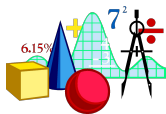
ТАБЛИЦЫ СОПРЯЖЕННОСТИ

- **Таблицы сопряженности – это совместное распределение двух переменных.**
- **Строки таблицы образуются значениями одной переменной.**
- **Столбцы таблицы образуются значениями второй переменной.**



ТАБЛИЦЫ СОПРЯЖЕННОСТИ

- **В клетке таблицы (на пересечении строки и столбца) указывается частота совместного появления соответствующих значений.**
- **Суммы частот по строке или по столбцу называются маргинальными частотами.**
- **Распределения маргинальных частот представляют собой одномерное распределение переменных.**



Проводим исследование:

X – семейное положение – НП

Y – занятость - ЗП

Собранные данные выглядят примерно так:

Испытуемый	Занятость	Семейное положение
1. Анна К.	Работает	Замужем
2. Галина Б.	Работает	Разведена
3. Татьяна В.	Не работает	Не замужем
…	…	…

Таким образом представленные данные не дают нам много информации.

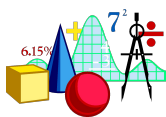
Можно их сгруппировать в виде таблиц:

по занятости:

Занятость	Частота	Проценты
Работает	98	49.0
Не работает	102	51.0
Всего	200	100

и по семейному положению:

Семейное положение	Частота	Проценты
Замужем	35	17.5
Никогда не была замужем	125	62.5
Разведена	15	7.5
Вдова	25	12.5
Всего	200	100



А можно и по двум переменным сразу:

Занятость (Y)	Семейное положение (X)				
	Не зам.	Зам.	Развед.	Вдова	Всего
Работает	21	60	11		
Не работает	14	65	4	19	
	35	125	15		

По строкам обычно идет зависимая переменная

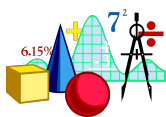
По столбцам обычно приводится независимая переменная

тельная таблица и называется таблицей сопряженности

Проценты в таблице сопряженности можно считать тремя способами:

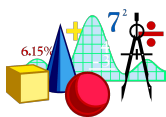
- **по столбцам, т.е. по независимой переменной**

Занятость (Y)	Семейное положение (X)			
	Не зам.	Зам.	Развед.	Вдова
Работает	**60%**	**48%**	**73.3%**	**24%**
Не работает	**40%**	**52%**	**26.7%**	**76%**
Всего	100%	100%	100%	100%



по строкам, т.е. по зависимой переменной

Занятость (Y)	Семейное положение (X)				
	Не зам.	Зам.	Развед.	Вдова	Всего
Работает	**21.4%**	**61.2%**	**11.2%**	**18.6%**	100%
Не работает	**13.7%**	**63.7%**	**3.8%**	**6.1%**	100%



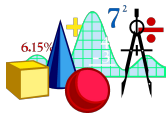
по всей таблице сразу:

Занятость (Y)	Семейное положение (X)				
	Не зам.	Зам.	Развед.	Вдова	
Работает	**10.5%**	**30%**	**5.5%**	**3%**	
Не работает	**7%**	**32.5%**	**2%**	**9.5**	
					100%

ТАБЛИЦЫ СОПРЯЖЕННОСТИ

для шкал наименований

для шкал порядка



ТАБЛИЦЫ СОПРЯЖЕННОСТИ

для шкал наименований

χ^2 Пирсона,
коэффициент сопряженности С,
V Крамера,

Ф
χ^2 МакНемара,
критерий Фишера
критерий Ятса (Yates)
...

для таблиц 2х2

для шкал порядка

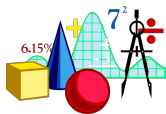
+

τ Кендалла,
Гамма (G),
ρ Спирмена,
d Соммера

ТАБЛИЦЫ СОПРЯЖЕННОСТИ

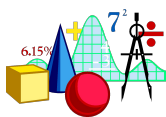
для шкал наименований

для шкал порядка



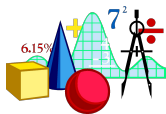
СТАТИСТИЧЕСКИЕ КРИТЕРИИ ДЛЯ ТАБЛИЦ СОПРЯЖЕННОСТИ

Проверяют, есть ли зависимость в распределении одной переменной от распределения по другой переменной.



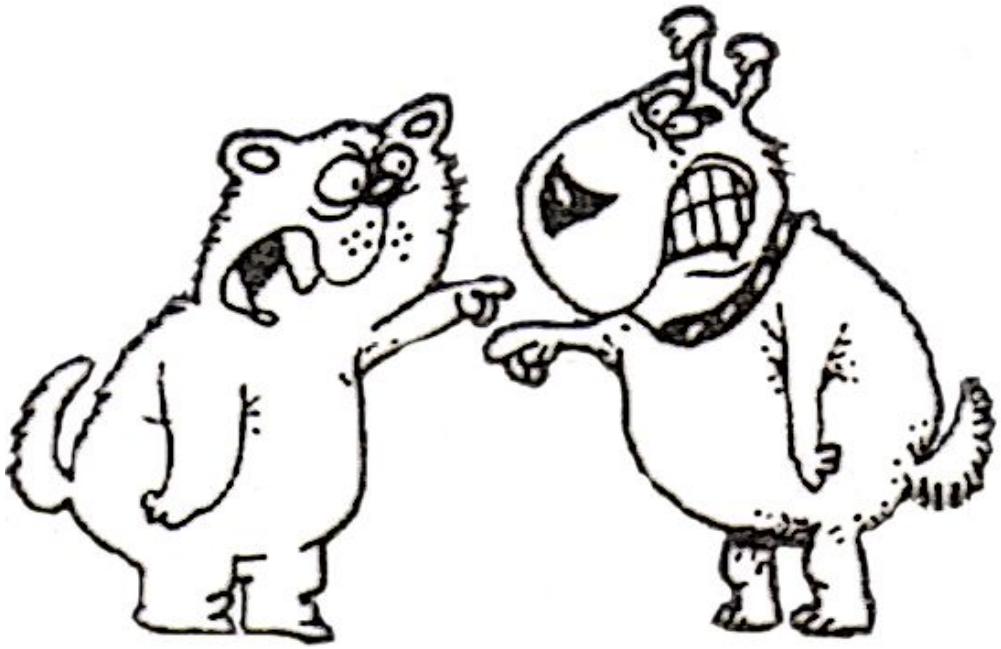
Межгрупповая схема ⇒ χ^2 Пирсона

Интраиндивидуальная схема ⇒ χ^2 МакНемара



$\chi 2$ Пирсона

**Пример: мы хотим проверить, правда ли,
что мужчины больше любят собак,
а женщины - кошек**



Было опрошено 550 человек. Результаты опроса представлены в таблице:

Любимое животное (Y)	Пол (X)		всего
	муж	жен	
Собака	**125**	**225**	350
Кошка	**75**	**125**	200
всего	200	350	550

Мы можем проверить, зависит ли предпочтение домашнего животного (распределение по переменной Y) от пола

Подсчет критерия χ2

(Пирсона)

$$\chi^2 = \sum_{i=1}^{k} \frac{(f_o - f_e)^2}{f_e}$$

f_o - эмпирическая частота,

f_e - теоретическая частота,

k=r*c,

r- число строк в таблице,

c –число столбцов в таблице,

df=(r-1)(c-1).

Как определить теоретическую частоту?

Для выделенной ячейки:

Следовательно, вероятность быть мужчиной и предпочитать собак равна

(200/550)*(350/550).

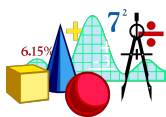
Умножив все это на количество испытуемых (550), получим теоретическую частоту для выделенной клетки:

(200/550)*(350/550)*550=127,3.

Подсчитав таким образом теоретические частоты для всех клеток, находим

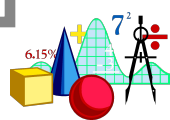
$$\chi^2=0,18;\ p=0,67$$

Следовательно, предпочтение домашнего животного не зависит от пола: мужчины и женщины одинаково любят собак.



Ограничения критерия χ2

- Если *теоретическая частота* клеток маленькая, то вычисления могут быть не точны. Сейчас общепринятым является правило, что когда df>1 *теоретическая частота* должна быть равна или больше 5 по крайней мере в 80% клеток.



$\chi 2$ МакНемара (McNemar)

Увы! Только для таблиц 2*2.

Тот критерий применяется, чтобы определить, произошли ли изменения после какого-либо условия. Данные обычно представляются в виде таблицы:

		после	
		II	I
до	I	A	B
	II	C	D

Получается, что A+D – это число изменений

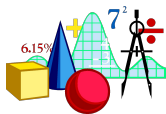
Подсчет критерия χ2

(МакНемара)

$$\chi^2 = \frac{(A-D)^2}{A+D}$$

Ограничения:

A+D должно быть не меньше 10!



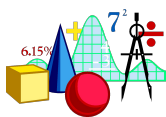
Пример: в телестудии проводятся дебаты, нужна ли смертная казнь. Зрители, сидящие в зале, опрашиваются до начала дебатов и в конце передачи.

		после дебатов	
		За смертную казнь	Против смертной казни
до	Против смертной казни	13	28
	За сметную казнь	27	7

χ^2=1,25; p=0,26. Следовательно, можно сделать вывод, что приглашенные ораторы были одинаково успешны в отстаивании своих точек зрения: мнения зрителей существенно не изменились

Что делать, если таблица большей размерности, а схема – интраиндивидуальная?

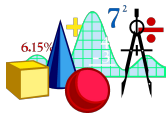
Для случая, когда условий больше (до дебатов, после дебатов, через год после дебатов…), можно использовать Q-критерий Кочрена (Кохрена), но только если данные представлены как дихотомические переменные (да/нет, за/против,…)



Что делать, если таблица большей размерности, схема – интраиндивидуальная, а данные не дихотомические?

Не проводить такие исследования!

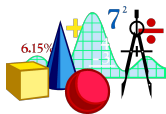
МЕРЫ ЗАВИСИМОСТИ ДЛЯ ТАБЛИЦ СОПРЯЖЕННОСТИ



Меры зависимости для шкал наименований

Все эти меры не имеют знака и не показывают направление отношений.

В программе STATISTICA можно посчитать три таких меры



Коэффициент φ

- употребляется в основном с таблицами 2x2
- меняется от 0 (когда переменные независимы) до 1 (когда они абсолютно зависимы)

$$\phi = \sqrt{\frac{\chi^2}{N}}$$

Коэффициент сопряженности C (или Ф)

- разработан для использования с квадратными таблицами размера больше, чем 2x2
- меняется от 0 (когда переменные независимы) до $\sqrt{(k-1)/k}$, где k - число строк (столбцов)

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

V Крамера

- **можно употреблять для любых таблиц - квадратных и прямоугольных**
- **меняется от 0 (когда переменные независимы) до 1 (когда они абсолютно зависимы)**

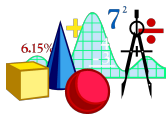
$$V = \sqrt{\frac{\chi^2}{N \cdot \text{Minimum}(r-1, c-1)}}$$

где c – число строк,
r – число столбцов таблицы.

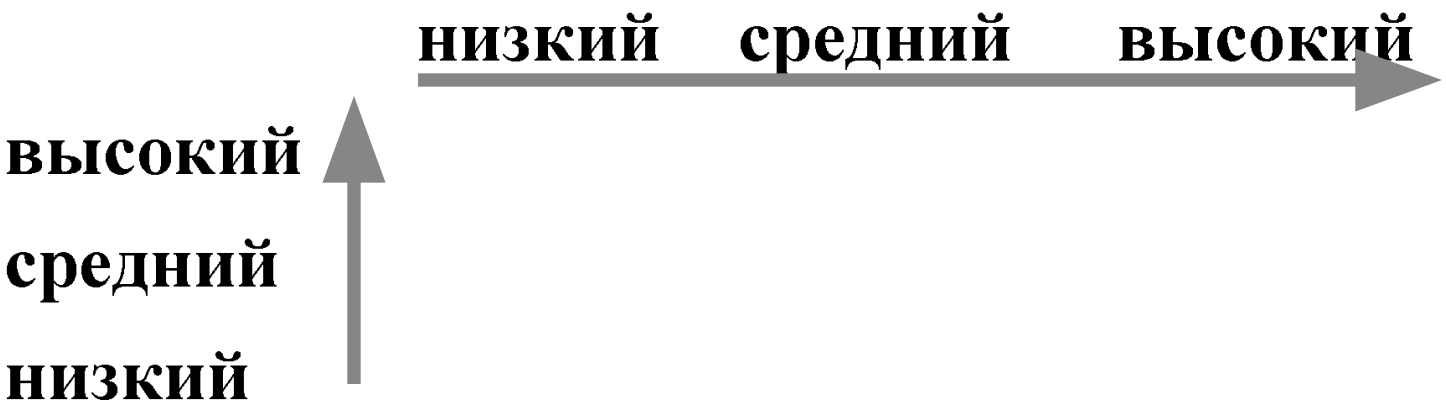
ТАБЛИЦЫ СОПРЯЖЕННОСТИ

для шкал наименований

для шкал порядка



- **В таблице сопряженности можно представлять и порядковые данные.**
- **Обычно они перечисляются слева направо (от меньшего к большему) и сверху вниз (от большего к меньшему):**



Доход (Y)	Возраст (X) молодой	немолодой	старый
высокий	A		D
средний		B	
низкий			C

Согласованная пара **- это пара, где оба члена ранжированы в одном порядке по двум направлениям.**

Доход (Y)	Возраст (X) молодой	немолодой	старый
высокий	A		D
средний		B	
низкий			C

Несогласованная пара **- это пара, где оба члена ранжированы в противоположном порядке по двум направлениям.**

Доход (Y)	Возраст (X) молодой	немолодой	старый
высокий	A		D
средний		B	
низкий			C

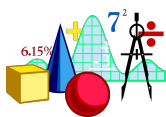
Связанная пара **- это пара, где оба члена ранжированы одинаково по крайней мере по одному направлению.**

Доход (Y)	Возраст (X)		
	молодой	немолодой	старый
высокий	20	6	2
средний	5	30	2
низкий	1	4	10

Если в таблице преобладают несогласованные пары, то зависимость между переменными отрицательная.

Доход (Y)	Возраст (X)		
	молодой	немолодой	старый
высокий	1	6	10
средний	5	30	2
низкий	20	4	2

Если в таблице преобладают согласованные пары, то зависимость между переменными положительная.



Меры зависимости

$$G = \frac{C - D}{C + D}$$

$$d_{yx} = \frac{C - D}{C + D + T_y}$$

$$\tau = \frac{C - D}{\sqrt{(C + D + T_y)(C + D + T_x)}}$$

C- число согласованных пар,

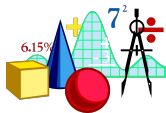
D - число несогласованных пар,

T_x - число пар, связанных по X

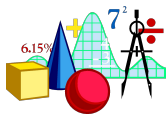
T_y = число пар, связанных по У

✷ **Меры зависимости**
для шкал порядка имеют знак

✷ **τ Кендалла всегда меньше 1,**
если таблица не квадратная



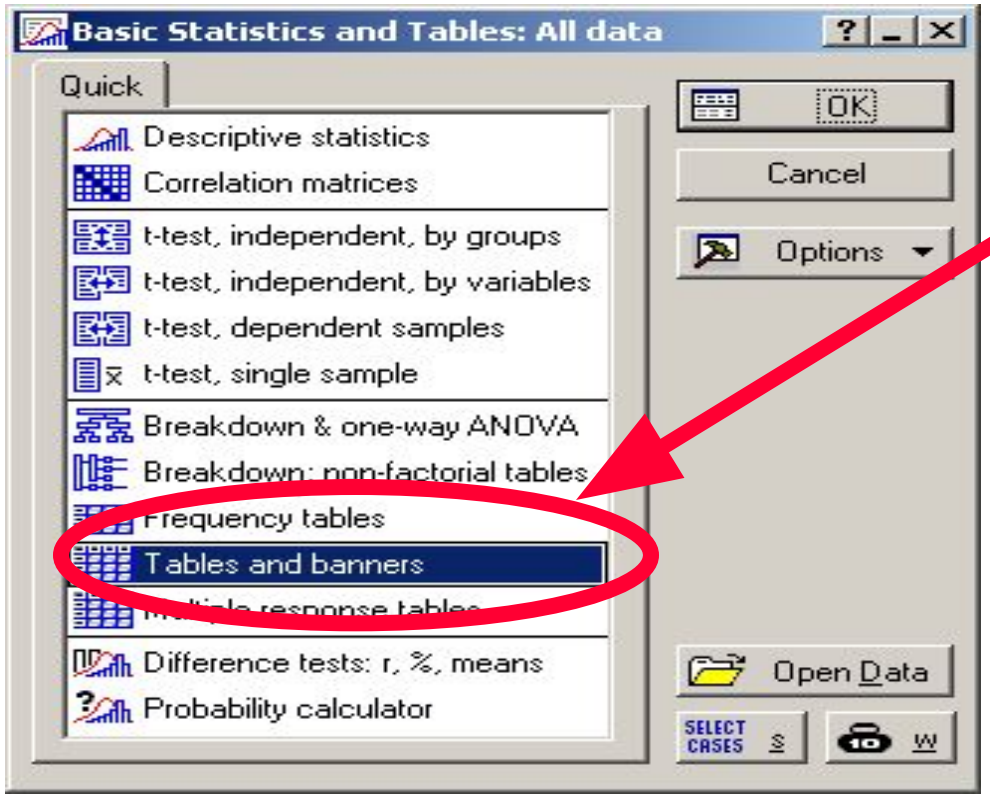
STATISTICA не знает, какая шкала была использована: определить подходящий критерий или меру зависимости - полностью ваша проблема (и ответственность)



Представление данных

Посчитать статистику для таблиц сопряженности можно в модуле

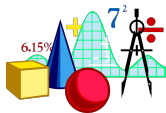
Basic Statistics/ Tables and Banners



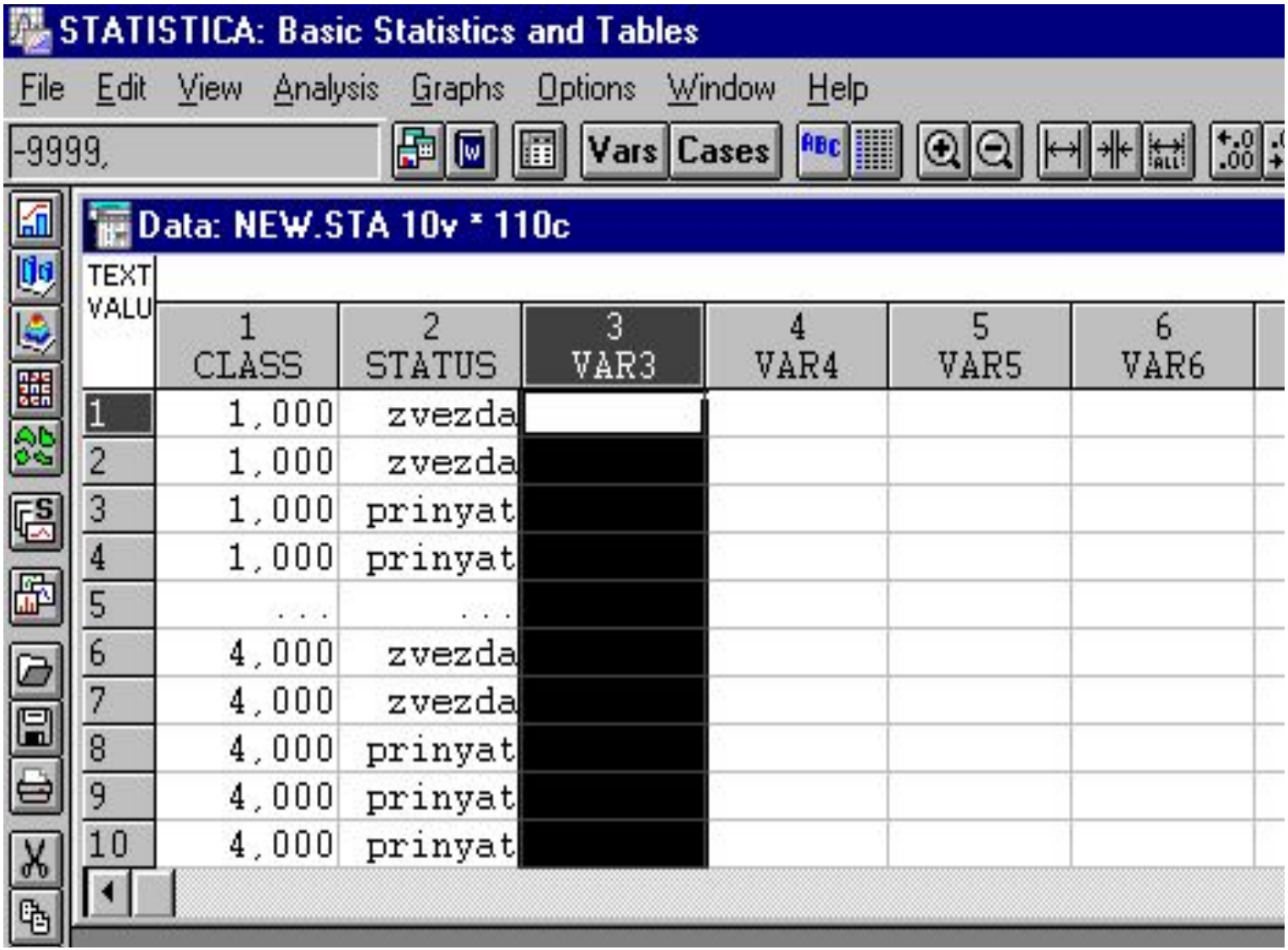
Представление данных

Исходные данные:

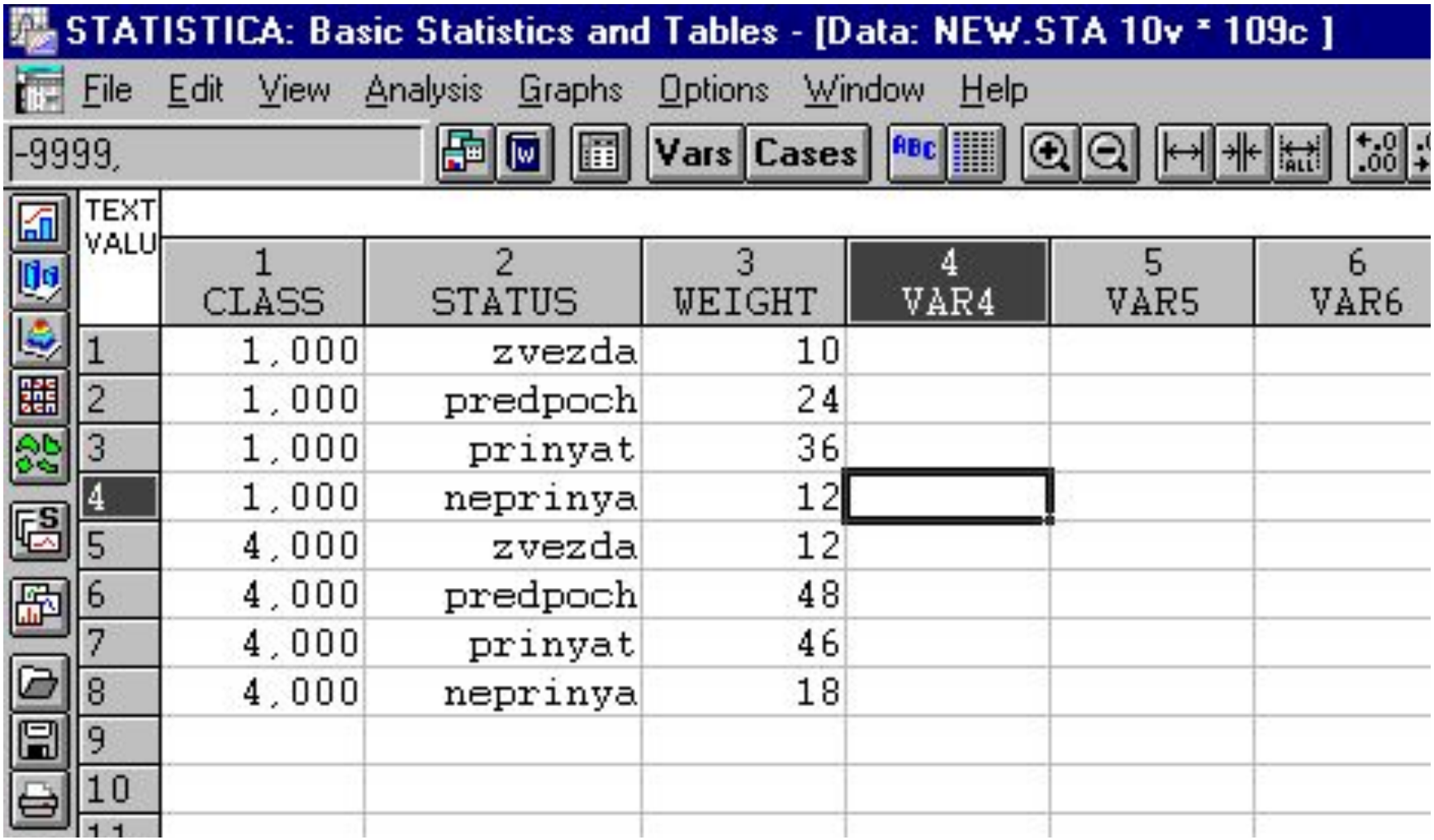
	звезда	принятый	…	…
1-й класс	10	36	…	
4-й класс	12	42	…	



Представление данных



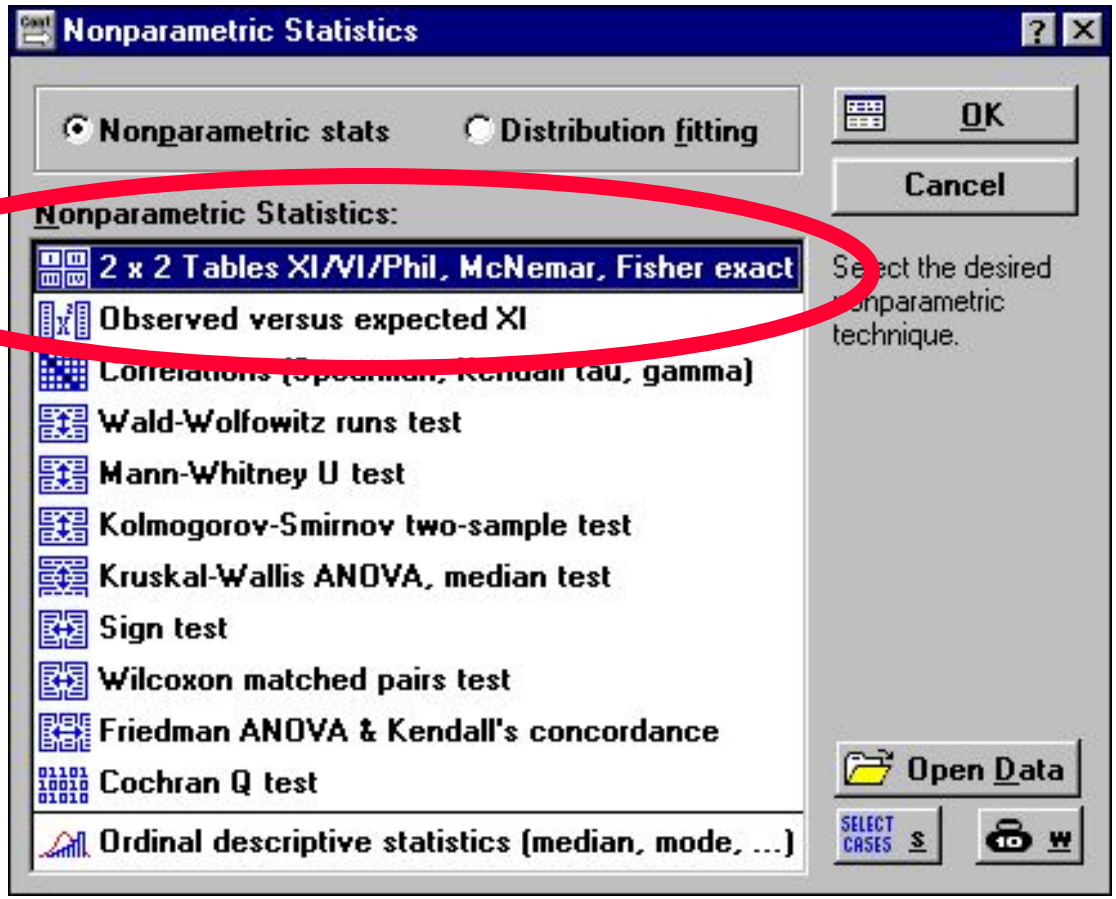
Представление данных



	1 CLASS	2 STATUS	3 WEIGHT	4 VAR4	5 VAR5	6 VAR6
1	1,000	zvezda	10			
2	1,000	predpoch	24			
3	1,000	prinyat	36			
4	1,000	neprinya	12			
5	4,000	zvezda	12			
6	4,000	predpoch	48			
7	4,000	prinyat	46			
8	4,000	neprinya	18			
9						
10						

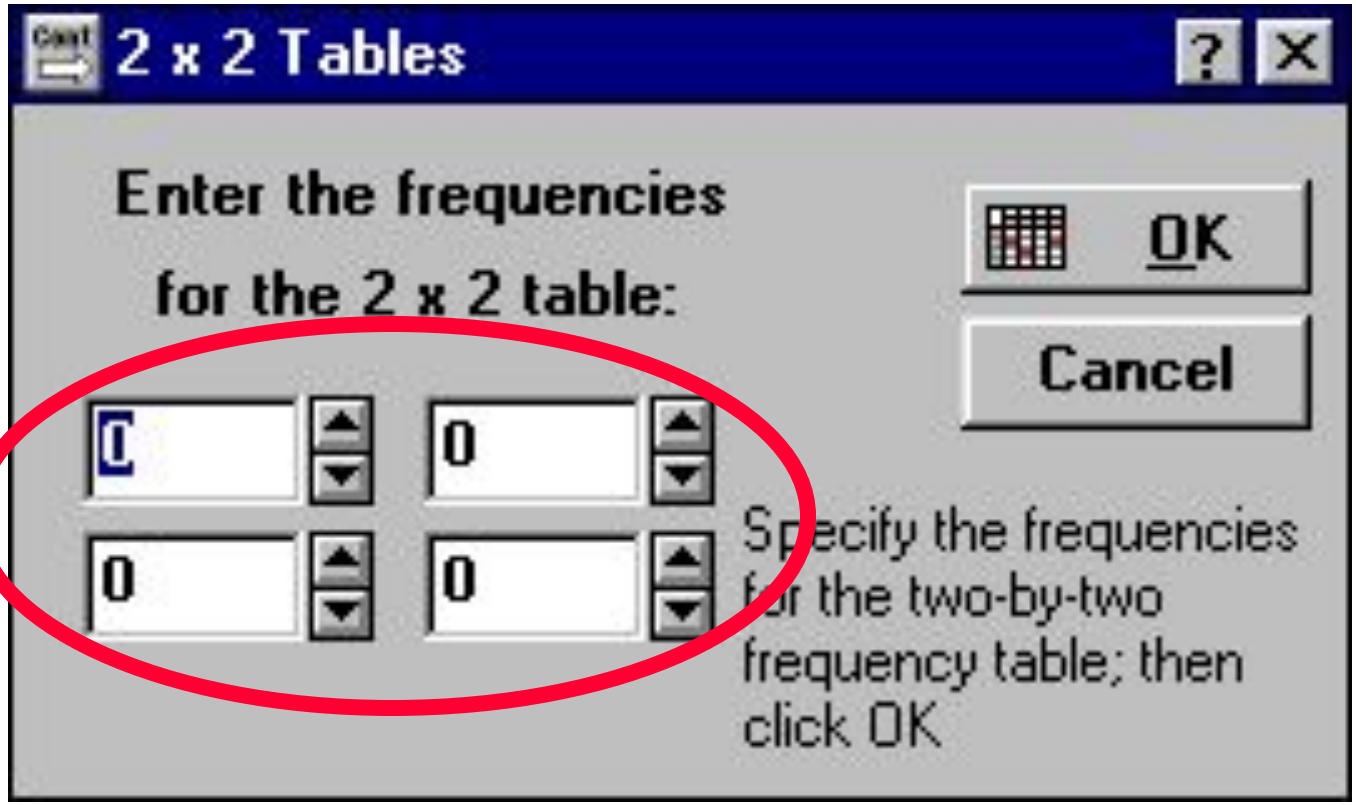
Представление данных

Для таблиц размером 2x2 есть еще модуль в Nonparametrics/Distrib.



Представление данных

Остается только ввести цифры…

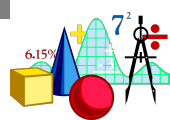


Представление данных

И получаем всю статистику!

2 x 2 Table (new.sta)

NONPAR STATS	Column 1	Column 2	Row Totals
Frequencies, row 1	2	32	34
Percent of total	2,410%	38,554%	40,964%
Frequencies, row 2	34	15	49
Percent of total	40,964%	18,072%	59,036%
Column totals	36	47	83
Percent of total	43,373%	56,627%	
Chi-square (df=1)	32,96	p= ,0000	
V-square (df=1)	32,56	p= ,0000	
Yates corrected Chi-square	30,42	p= ,0000	
Phi-square	,39710		
Fisher exact p, one-tailed		p= ,0000	
two-tailed		p= ,0000	
McNemar Chi-square (A/D)	8,47	p= ,0036	
Chi-square (B/C)	,02	p= ,9020	

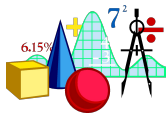


Самостоятельная работа

К следующему занятию прочитать:

Савина и Ванг. Выбор и принятие решений: риск и социальный контекст// ПЖ, ….

(есть в электронном виде)



Можно передохнуть!