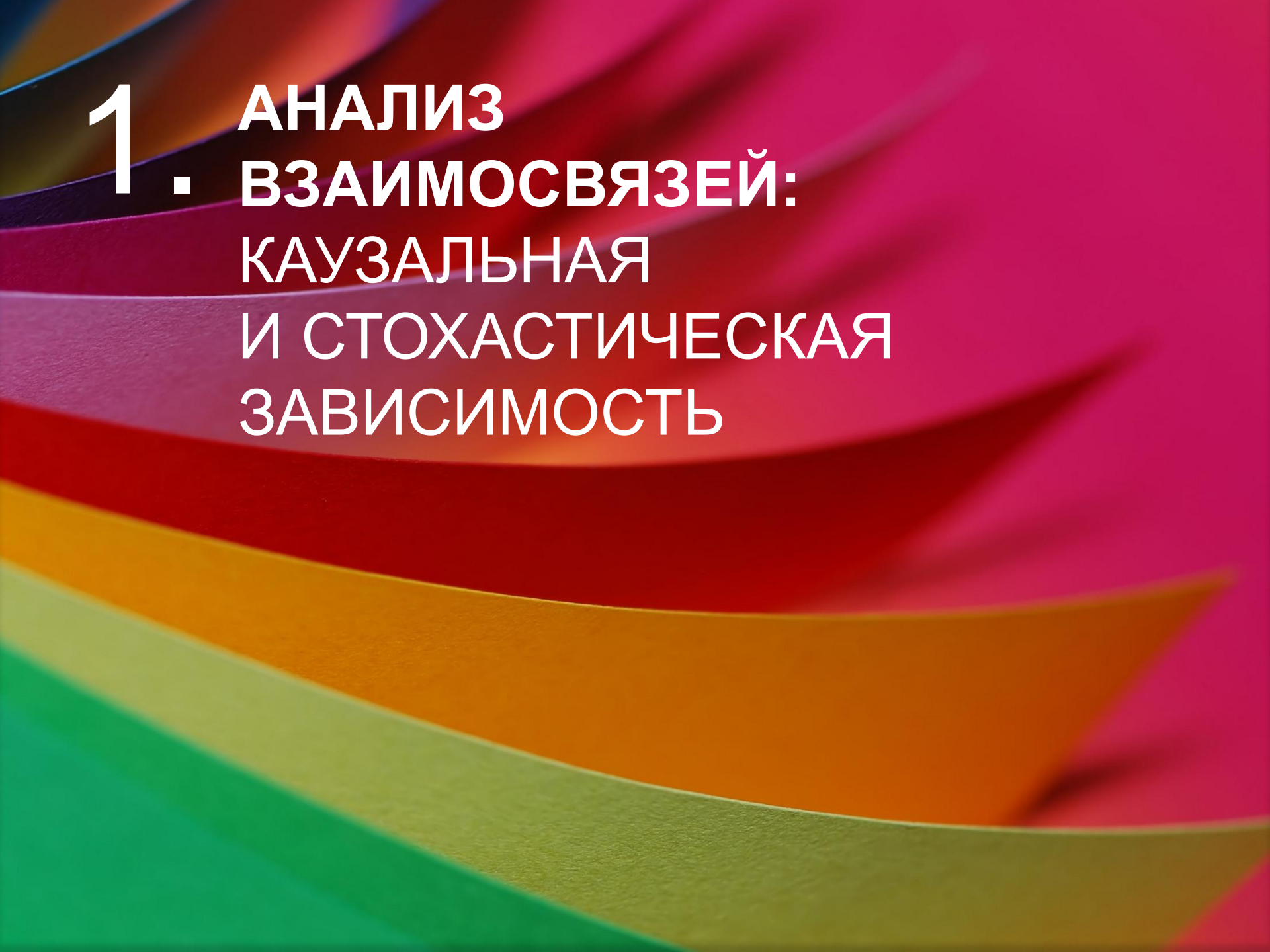


## ТЕМА 4

# АНАЛИЗ ДВУМЕРНОЙ ЗАВИСИМОСТИ. ТАБЛИЦЫ СОПРЯЖЕННОСТИ. КОЭФФИЦИЕНТЫ ПАРНОЙ СВЯЗИ

- 1. Анализ взаимосвязей: каузальная и стохастическая зависимость**
- 2. Таблицы сопряженности: логика построения в SPSS (классический вариант)**
- 3. Проверка наличия взаимосвязи между переменными**
- 4. Альтернативный способ построения таблиц сопряженности**



# 1. АНАЛИЗ ВЗАИМОСВЯЗЕЙ: КАУЗАЛЬНАЯ И СТОХАСТИЧЕСКАЯ ЗАВИСИМОСТЬ

# 1. Анализ взаимосвязей

---


- **Анализ данных** - это прежде всего исследование взаимосвязей, выявление отношений между несколькими признаками.
- Изучение взаимозависимостей социальных явлений начинается с выдвижения теоретических гипотез, содержащих априорные предположения о причинно-следственных связях. Для их проверки применяются **статистические методы**.
- В простейшем случае **каузальную зависимость** можно описать следующим образом:  $X$  определяет  $Y$  (где  $X$  – причина, а  $Y$  — следствие). Суть каузальной зависимости - в ее детерминистском характере. Это означает существование следующих отношений:
  1.  $X$  влияет на  $Y$  ( $X \rightarrow Y$ )
  2. Если  $X$  не существует, то не существует и  $Y$
- Необходимое условие анализа каузальных зависимостей – требование “**при прочих равных условиях**” (т.е. все остальные факторы никак не влияют на  $X$  и  $Y$ ). К сожалению, в социальных исследованиях эти условия, как правило, не выполняются.

# 1. Анализ взаимосвязей

---

- **Пример:** Удовлетворенность доходом (X) и жилищными условиями (Y) ведет к удовлетворенности условиями жизни в целом (Z).
  1. Если исследователь ограничится изучением зависимости между доходом (X) и удовлетворенностью условиями жизни (Z), то он уже не получит каузальную зависимость.
  2. Предположим, что среди опрошенных, которые имеют довольно высокий доход (X), 70% удовлетворены условиями жизни, а 30% не удовлетворены. Следовательно, высокие доходы приводят к удовлетворенности условиями жизни **только с вероятностью 70%** ( $P = 0,7$ ). В этом случае говорят о **стохастической зависимости** (обусловлена воздействием неучтенных факторов).
- На этом примере можно видеть, что связь между двумя или более переменными представляется как стохастическая, если исследователь не принял во внимание все влияющие на нее факторы. **Поэтому после выявления и констатации самого факта существования стохастической зависимости первая задача исследователя – установить, какие дополнительные каузальные факторы не учтены.**
- Изучение статистических зависимостей начинается с определения значимости различий ответов разных групп респондентов и дальнейшего расчёта коэффициентов корреляции. Принципиальное значение при построении подобных коэффициентов приобретает **тип шкалы**, используемый для измерения переменных.



The background of the slide is a vibrant, diagonal rainbow gradient. The colors transition from dark blue and purple at the top left, through green, yellow, and orange, to bright red at the bottom right. Numerous clear, realistic water droplets of various sizes are scattered across the entire surface, giving it a fresh, wet appearance.

# 2. ТАБЛИЦЫ СОПРЯЖЕННОСТИ: ЛОГИКА ПОСТРОЕНИЯ В SPSS (КЛАССИЧЕСКИЙ ВАРИАНТ)

## 2. Таблицы сопряженности: логика построения в SPSS (классический вариант)

Характер отношений между двумя переменными наглядно можно представить разными способами:

1. В случае измерения признаков в **номинальной** или **порядковой** шкале возможные комбинации значений обеих переменных упорядочиваются в форме **таблицы сопряженности**, при этом значения одной переменной образуют строки, а значения другой — столбцы таблицы.
2. После этого определяется частота появления всех возможных комбинаций обоих признаков в собранных данных и значение ее заносится в ячейки, стоящие на пересечении соответствующих строк и столбцов.

Для **количественных данных**, а также порядковых переменных с очень большим числом вариаций исходные значения **предварительно группируются**, соседние значения объединяются в небольшое число интервалов.

Возраст	→	Возрастные группы	Города	→	Тип населенного пункта
18	}	18-24 года	Москва	}	Города-миллионники
19			Санкт-Петербург		
20			Ростов-на-Дону		
21			Самара		
22			Батайск		
23			Химки	}	500 тыс. + 250 тыс. + 100 тыс. + Менее 100 тыс.
24			...		
...					

## 2. Таблицы сопряженности: логика построения в SPSS (классический вариант)

**Пример:** В ходе исследования удовлетворенности условиями жизни населения требуется выявить факторы, влияющие на нее. Естественно предположить, что чем выше уровень дохода на одного члена семьи, тем выше степень удовлетворенности. Данную гипотезу очень легко проверить путем построения **двухмерной таблицы сопряженности**.

Доход, руб.		Удовлетворенность работой			
		Низкая	Средняя	Высокая	Всего
Менее 500.	Частота	100	100	50	250
	Строковый процент	40,0	40,0	20,0	100,0
	Столбцовый процент	38,5	27,0	16,7	26,9
	Общий процент	10,8	10,8	5,4	26,9
500 — 1000.	Частота	100	150	100	350
	Строковый процент	28,6	42,9	28,6	100,0
	Столбцовый процент	38,5	40,5	33,3	37,6
	Общий процент	10,8	16,1	10,8	37,6
1000 — 1500.	Частота	50	100	100	250
	Строковый процент	20,0	40,0	40,0	100,0
	Столбцовый процент	19,2	27,0	33,3	26,9
	Общий процент	5,4	10,8	10,8	26,9
Более 1500р.	Частота	10	20	50	80
	Строковый процент	12,5	25,0	62,5	100,0
	Столбцовый процент	3,8	5,4	16,7	8,6
	Общий процент	1,1	2,2	5,4	8,6
Всего	Частота	260	370	300	930
	Строковый процент	28,0	39,8	32,3	100,0
	Столбцовый процент	100,0	100,0	100,0	100,0
	Общий процент	28,0	39,8	32,3	100,0



## 2. Таблицы сопряженности: логика построения в SPSS (классический вариант)

---

### ПОСТРОЕНИЕ ТАБЛИЦ СОПРЯЖЕННОСТИ В SPSS

#### Для чего создаются?

- Показывают совместное распределение переменных в номинальной и порядковой шкале
- Проверка значимых различий между наблюдаемыми и ожидаемыми частотами
- Выявляют наличие (отсутствие) взаимосвязи между переменными

#### Что находится в окне вывода?

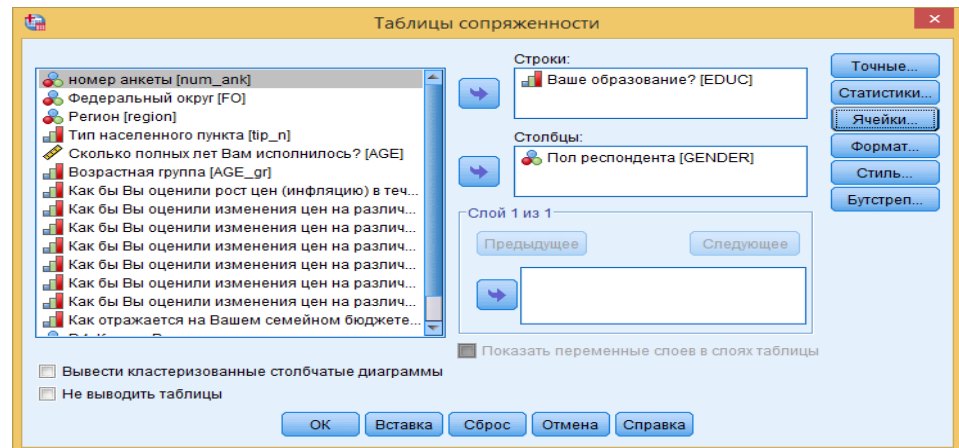
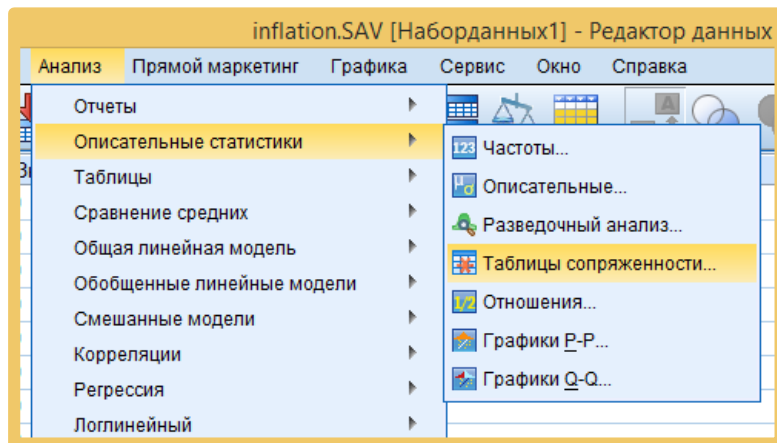
- Строки
- Столбцы

Есть альтернативный способ построения таблиц сопряженности,  
но об этом – в конце лекции

## 2. Таблицы сопряженности: логика построения в SPSS (классический вариант)

### Как построить?

- Меню «Анализ» → «Описательные статистики» → «Таблицы сопряженности».
- Переносим из списка переменных в «Строки» и «Столбцы» необходимые переменные.



## 2. Таблицы сопряженности: логика построения в SPSS (классический вариант)

---

При объяснении причинно-следственных связей – выбор % по строке или по столбцу зависит от исследовательской гипотезы того, что является причиной, а что – следствием:



**Пример:** Оценка уровня инфляции зависит от дохода на одного члена семьи (т.е. различается ли уровень инфляции по группам респондентов с различным материальным положением)? (Файл inflation.sav).

Y – оценка роста цен (инфляции) в течение последнего месяца-двух  
X – доход на одного члена семьи

**Пример:** Оценка уровня инфляции влияет на доход одного члена семьи (**что не логично**).

Y – доход на одного члена семьи  
X – оценка роста цен (инфляции) в течение последнего месяца-двух

## 2. Таблицы сопряженности: логика построения в SPSS (классический вариант)

---

- При анализе таблиц сопряженности выявляется наличие / отсутствие **статистических**, а не причинно-следственных зависимостей.
- Пример: О.Генри («Вождь краснокожих») модель для ответа на вопрос: «Почему дует ветер? – Потому что деревья качаются».



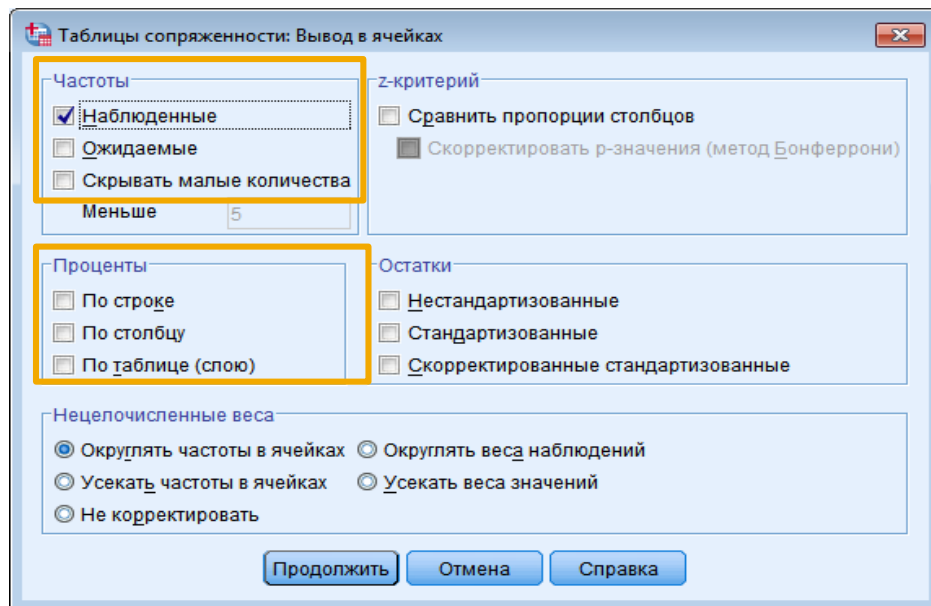
## 2. Таблицы сопряженности: логика построения в SPSS (классический вариант)

- Подразделяя переменные на зависимую и независимую, рекомендуется следовать следующему правилу:

*Если независимая переменная строковая, выбирайте строковые проценты, если независимая переменная столбцовая, выбирайте столбцовые.*

*Это соответствует логике анализа "от причины к следствию".*

- Достоверность выводов зависит от представительности выборки. Если выборка нерепрезентативна, то расчет определенных соотношений на основе имеющихся данных представляется бессмысленным.



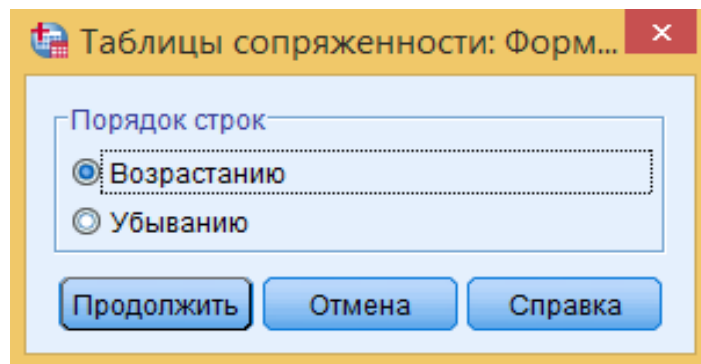
## 2. Таблицы сопряженности: логика построения в SPSS (классический вариант)

---

### ФОРМАТЫ ТАБЛИЦ СОПРЯЖЕННОСТИ

Порядок сортировки строк в таблице сопряженности может быть изменен. Для этого необходимо выполнить следующие действия:

1. Щелкните по кнопке **«Формат»** (Format) в диалоговом окне **«Таблицы сопряженности»** (Crosstabs).
2. В группе **«Порядок строк»** (Row Order) можно выбрать один из следующих вариантов сортировки значений:
  - **«По возрастанию»** (Ascending)
  - **«По убыванию»** (Descending)



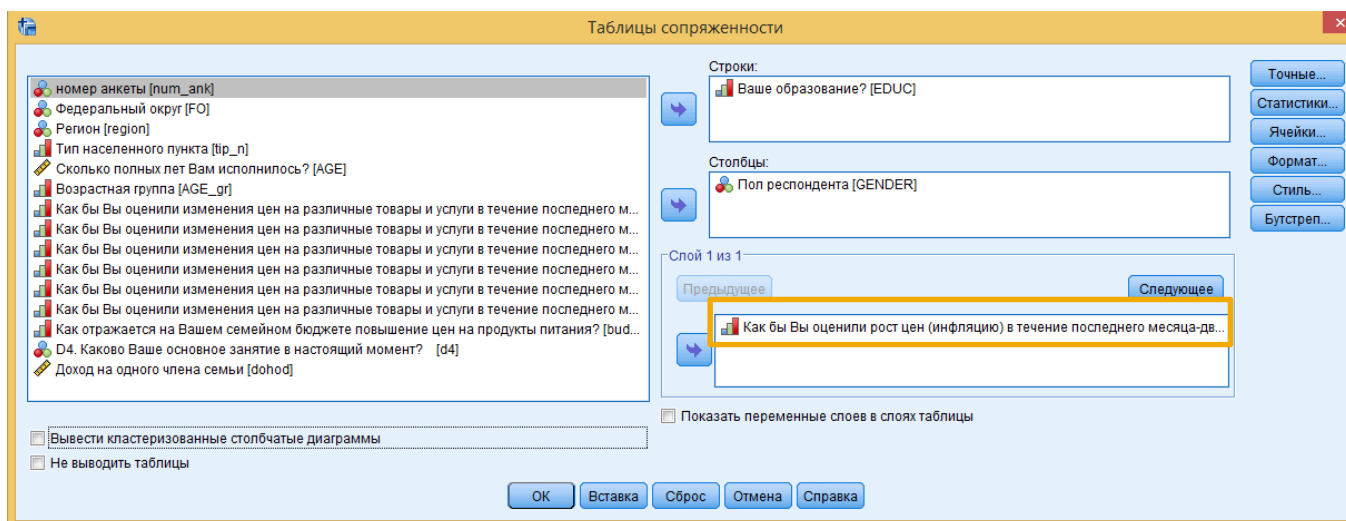
## 2. Таблицы сопряженности: логика построения в SPSS (классический вариант)

### ПРИМЕНЕНИЕ ПЕРЕМЕННЫХ ГРУПП И СЛОЕВ

Есть возможность построить таблицу сопряженности для трех и более переменных.

**Пример:** включить в таблицу уровень образования (EDUC), пол (GENDER) и оценку роста цен в течение последнего месяца-двух (inflation).

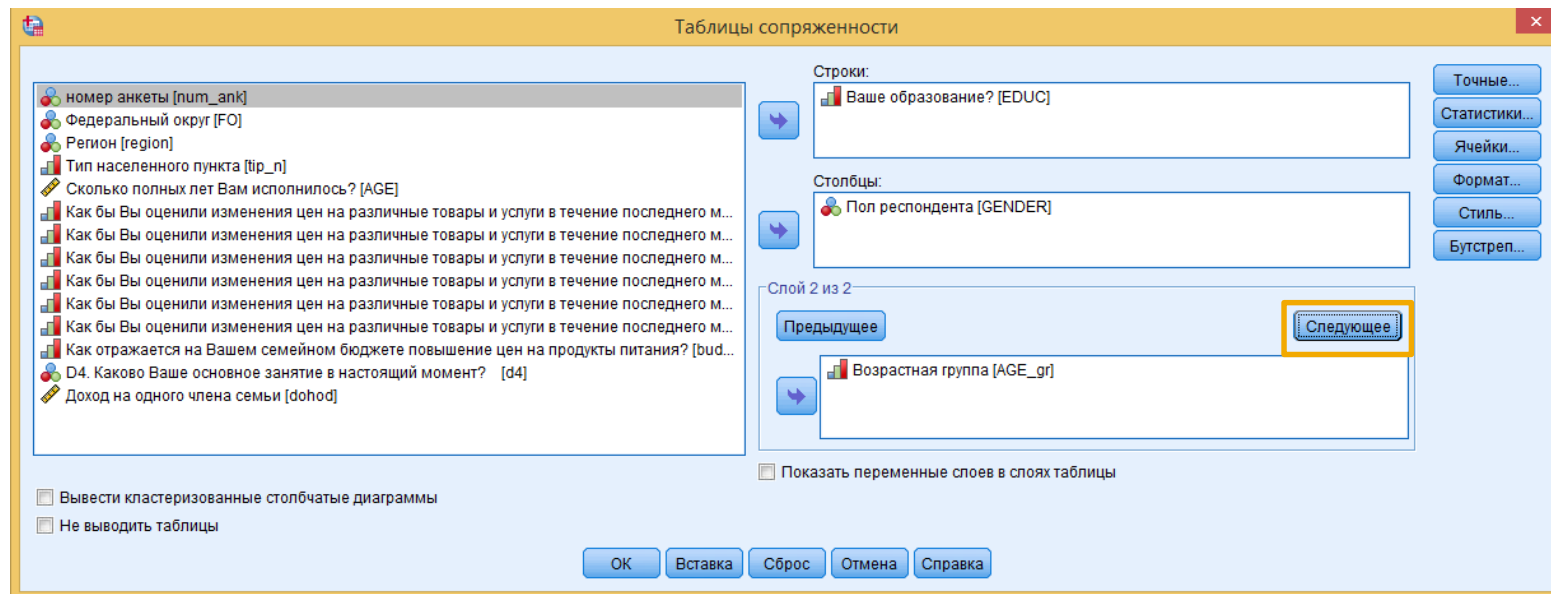
Третью переменную необходимо поместить в поле «Слой». При добавлении в тот же самый слой новой переменной, увеличится количество выводимых таблиц.



## 2. Таблицы сопряженности: логика построения в SPSS (классический вариант)

Можно наращивать слои: группировать по большому количеству переменных.

- Нажать на «Следующее» и добавить в чистое поле новую переменную.
- В таблице сопряженности добавится новый слой.
- Нажать ОК.





## 2. Таблицы сопряженности: логика построения в SPSS (классический вариант)

Два новых слоя

Изначальные переменные

Возрастная группа	Как бы Вы оценили рост цен (инфляцию) в течение последнего месяца-двух?			Пол респондента		Всего	
				Мужской	Женский		
18-24	Инфляция очень высокая	Ваше образование?	Неполное среднее образование или ниже	Количество	1	0	1
				Ожидаемое количество	,5	,5	1,0
			Среднее образование (школа или ПТУ)	Количество	15	10	25
				Ожидаемое количество	13,2	11,8	25,0
			Среднее специальное образование (техникум)	Количество	7	6	13
				Ожидаемое количество	6,9	6,1	13,0
			Незаконченное высшее (с 3-го курса ВУЗа)	Количество	2	2	4
				Ожидаемое количество	2,1	1,9	4,0
			Высшее образование	Количество	2	6	8
				Ожидаемое количество	4,2	3,8	8,0
			Всего	Количество	27	24	51
				Ожидаемое количество	27,0	24,0	51,0
	Инфляция умеренная	Ваше образование?	Неполное среднее образование или ниже	Количество	0	2	2
				Ожидаемое количество	1,0	1,0	2,0
			Среднее образование (школа или ПТУ)	Количество	14	9	23
				Ожидаемое количество	11,1	11,9	23,0
			Среднее специальное образование (техникум)	Количество	13	13	26
				Ожидаемое количество	12,6	13,4	26,0
			Незаконченное высшее (с 3-го курса ВУЗа)	Количество	2	5	7
				Ожидаемое количество	3,4	3,6	7,0
			Высшее образование	Количество	1	3	4
				Ожидаемое количество	1,9	2,1	4,0
			Всего	Количество	30	32	62
				Ожидаемое количество	30,0	32,0	62,0

## 2. Таблицы сопряженности: логика построения в SPSS (классический вариант)

---

Можно построить столбчатую диаграмму.

### Первый способ:

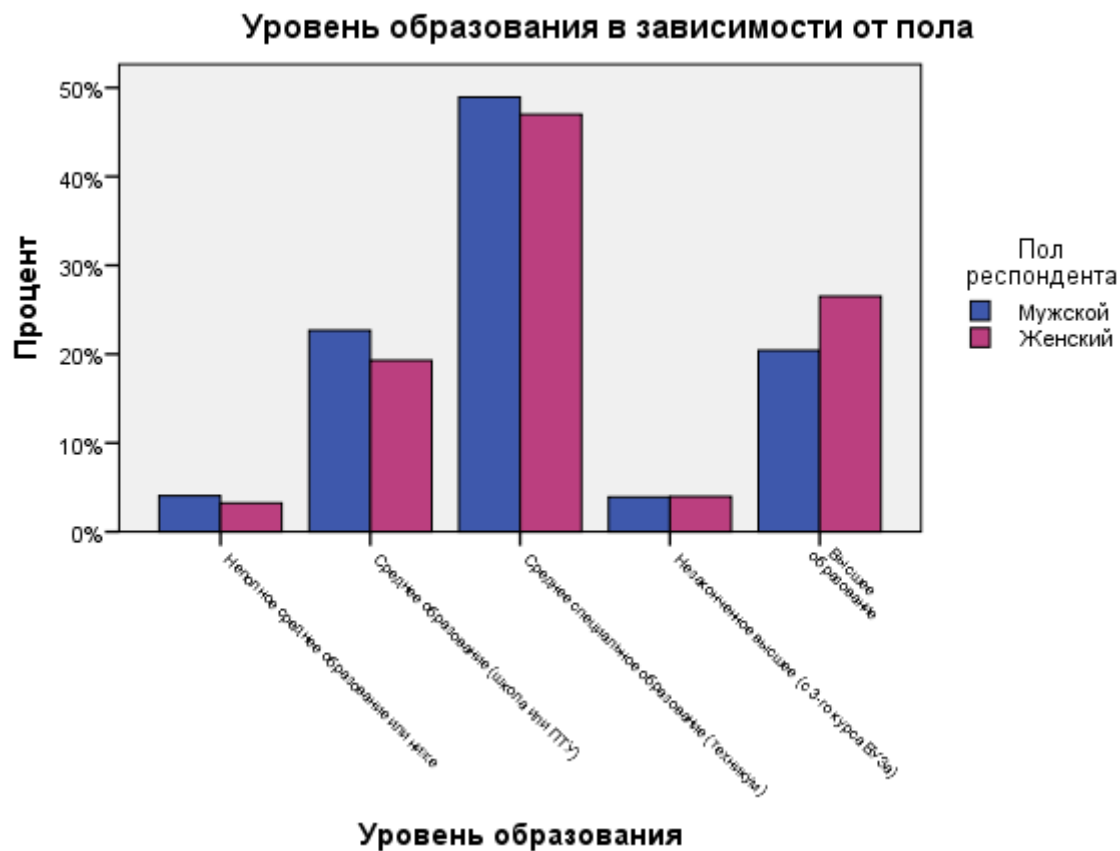
- При создании таблицы сопряженности отметить **«Вывести кластеризованные столбчатые диаграммы»**.
- Однако стоит поменять переменные местами переменные в строке и столбце.

### Второй способ:

- **«Графики» → «Столбцы» → «Кластеризованные»**.
- Нажать **«Задать»**.
- «Столбцы представляют - % наблюдений».
- Добавить переменные в **«Категориальная ось»** и **«Задать кластеры по»**.
- Нажать **ОК**.

## 2. Таблицы сопряженности: логика построения в SPSS (классический вариант)

### Второй способ



## 2. Таблицы сопряженности: логика построения в SPSS (классический вариант)

### Как представлять таблицы сопряженности?

- название и номер
- название переменных и их категорий
- % от переменных (по столбцу или по строке, респондентов, ответивших, ответов и т.д. – особенно при множественном выборе)
- число респондентов в каждом столбце (N)
- источник данных
- в сноске - пояснение, если сумма ответов превышает 100% (сумма превышает 100%, т.к. респондент мог дать несколько ответов на вопрос)

### Когда нецелесообразно использовать таблицы сопряженности?

- Если переменные имеют большое количество категорий
- При малых выборках (недостаточной наполняемости ячеек (минимум = 100 человек, лучше уже не делить на подгруппы))

Уровень измерения		Y variable							
		Число категорий	Nominal	Ordinal			Interval		
			Мало	Мало	Средне	Много	Мало	Средне	Много
X variable	Nominal, Ordinal, Interval	Мало	Bar	Bar	Line	×	Bar	Line	Boxplot
	Interval	Много	×	×	×	×	×	×	Scatterplot



## 2. Таблицы сопряженности: логика построения в SPSS (классический вариант)

---

Исследование таблицы сопряженности, описывающей совместное двухмерное распределение переменных, является первым шагом при изучении отношений между двумя переменными.

Однако с помощью нее **невозможно** оценить или проверить характер и степень их взаимного влияния.

Для этих целей используется специальный статистический инструментарий, который позволяет ответить **на два главных вопроса**:

1. Существует ли связь между рассматриваемыми показателями?
2. Если связь существует, то насколько она сильна?

# 3. ПРОВЕРКА НАЛИЧИЯ ВЗАИМОСВЯЗИ МЕЖДУ ПЕРЕМЕННЫМИ

### 3. Проверка наличия взаимосвязи между переменными

---

В статистических исследованиях особенно большой интерес представляет **гипотеза о независимости переменных друг от друга**, для которых построена таблица сопряженности.

Для ее проверки устанавливают, как выглядело бы совместное распределение переменных  $X$  и  $Y$ , если бы они статистически не зависели друг от друга.

А две переменные по определению независимы, когда вероятность попадания их в любую ячейку данной таблицы сопряженности является результатом произведения вероятностей появления значений обеих переменных этой ячейки.

**Как узнать, есть ли связь между двумя номинальными признаками?**

**Ответ:** Зависимость есть отсутствие независимости. Два события считаются независимыми, если вероятность того, что они произойдут одновременно, равна произведению вероятностей того, что произойдет каждое из них.

### 3. Проверка наличия взаимосвязи между переменными

---

**Пример:** бросаем две монеты. Вероятность выпадения «орла» на каждой из них одинакова:  $P = 0,5$ . В случае отсутствия зависимости между результатами подбрасывания двух монет вероятность одновременного выпадения «орлов» на обеих монетах равна  $0,5 \cdot 0,5 = 0,25$ . Значит, если в большой серии опытов частота совместного выпадения двух «орлов» сильно отличается от 0,25, можно говорить об отсутствии независимости, т.е. о наличии зависимости между бросанием двух монет.

**В социологии:** в массиве 50% мужчин и 30% лиц с высшим образованием. При отсутствии зависимости между полом и образованием мужчин с высшим образованием должно быть  $0,5 \cdot 0,3 = 0,15$ . Можно подсчитать реальную долю мужчин с высшим образованием («**Таблицы сопряженности**» → «**Наблюденные**»), и если эта доля сильно отличается от 0,15, можно говорить, что гипотеза о независимости между полом и наличием высшего образования не подтверждается, т.е. связь есть.

То количество респондентов, которое должно быть в клетках таблицы сопряженности в случае независимости двух событий – **ожидаемая частота**.

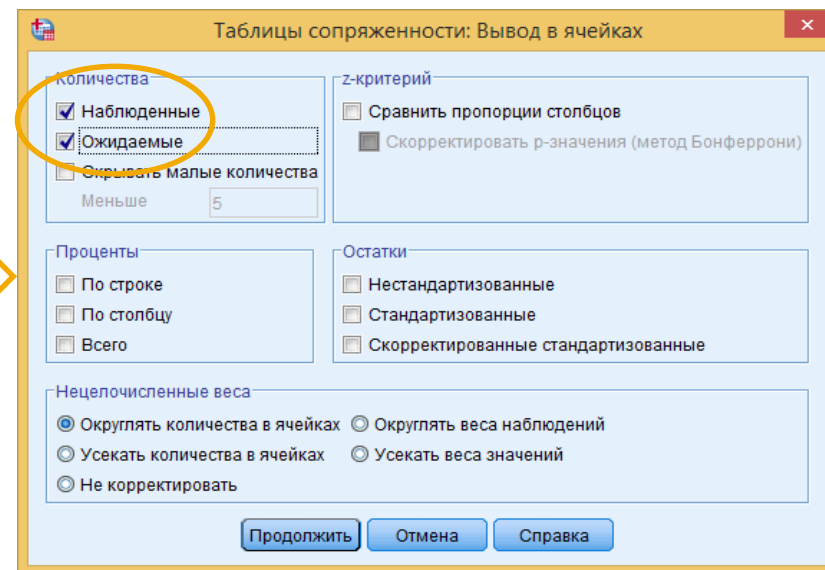
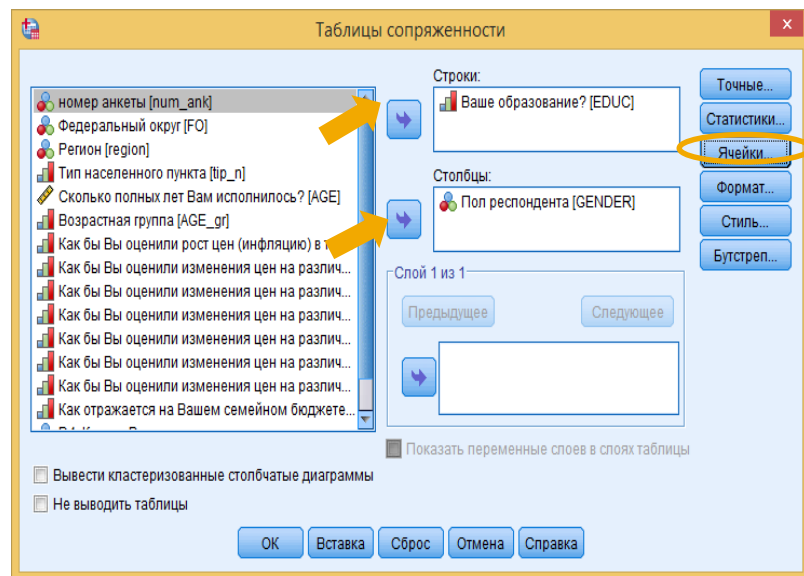
**Пример:** если в выборке 1000 чел., среди которых 50% мужчин и 30% - с в/о, то в случае независимости пола и образования ожидаемая частота в клетке «мужчины с в/о» –  $0,5 \cdot 0,3 \cdot 1000 = 150$ . («**Таблицы сопряженности**» → «**Ячейки**» → «**Ожидаемые**»).

Можно сопоставить реальные частоты и ожидаемые во всех ячейках, и если они разные – модель о независимости переменных не подтверждается.

(Пример: Inflation.sav, GENDER и EDUC).



### 3. Проверка наличия взаимосвязи между переменными



Комбинационная таблица Как бы Вы оценили изменения цен на различные товары и услуги в течение последнего месяца-двух? А) Хлебобулочные изделия \* Пол респондента

			Пол респондента		Всего
			Мужской	Женский	
Как бы Вы оценили изменения цен на различные товары и услуги в течение последнего месяца-двух? А) Хлебобулочные изделия	Значительно снизились	Количество	1	1	2
		Ожидаемое количество	,9	1,1	2,0
	Снизилась, но незначительно	Количество	12	8	20
		Ожидаемое количество	8,9	11,1	20,0
	Не изменились	Количество	248	300	548
		Ожидаемое количество	242,5	305,5	548,0
	Выросли, но незначительно	Количество	283	336	619
		Ожидаемое количество	273,9	345,1	619,0
	Значительно выросли	Количество	107	175	282
		Ожидаемое количество	124,8	157,2	282,0
	Всего	Количество	651	820	1471
		Ожидаемое количество	651,0	820,0	1471,0

### 3. Проверка наличия взаимосвязи между переменными

---

#### ХИ-КВАДРАТ

**Но:** это статистические данные – а вдруг расхождения между ожидаемыми и реальными частотами носят случайный характер? Поэтому механизм проверки гипотезы о независимости переменных немного сложнее:

- Вычисляется степень суммарного расхождения реальных и ожидаемых частот (разности этих частот могут иметь разные знаки, тогда сумма будет равна 0, поэтому суммируются квадраты разностей).
- Складываются не абсолютные, а относительные расхождения частот.

**Критерий ХИ-квадрат Пирсона – это показатель, фиксирующий степень расхождения реальных и ожидаемых частот.**

$$\chi^2 = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i}$$

- если значимость ХИ-квадрат = 0, значит признаки независимы
- чем значимость ХИ-квадрат меньше, тем меньше различия между реальными и ожидаемыми частотами

### 3. Проверка наличия взаимосвязи между переменными

#### ХИ-КВАДРАТ

- **«Число степеней свободы» (df)** – количество значений в итоговом вычислении статистики, способных варьироваться.

$$N=(r-1)(c-1)$$

где **r** – количество столбцов, **c** – количество строк.

- **Пример:** построение таблицы сопряженности между переменными «Пол респондента» и «Ваше образование» в массиве данных inflation.sav.

Комбинационная таблица Пол респондента \* Ваше образование?

			Ваше образование?					Всего
			Неполное среднее образование или ниже	Среднее образование (школа или ПТУ)	Среднее специальное образование (техникум)	Незаконченное высшее (с 3-го курса ВУЗа)	Высшее образование	
Пол респондента	Мужской	Количество	27	151	326	26	136	666
		Ожидаемое количество	24,0	138,5	318,8	26,2	158,5	666,0
	Женский	Количество	27	161	392	33	221	834
		Ожидаемое количество	30,0	173,5	399,2	32,8	198,5	834,0
Всего	Количество		54	312	718	59	357	1500
	Ожидаемое количество		54,0	312,0	718,0	59,0	357,0	1500,0

$$\text{«Число степеней свободы» (df)} = (5 - 1) * (2 - 1) = 4$$

При построении таблиц сопряженности **обязательно проверять существование** связи через хи-квадрат!!! Можно не приводить в таблице, но указать в сноске, что коэффициенты значимы.

### 3. Проверка наличия взаимосвязи между переменными

---

#### ХИ-КВАДРАТ

**Гипотеза о независимости переменных отвергается**, если реальное значение Хи-квадрат больше теоретического (табличного).

- Искомое значение в таблице находится на пересечении столбца с соответствующим значением вероятности и строки с числом степеней свободы.
- **Пример:** критическое значение Хи-квадрат распределения с 4-мя степенями свободы для вероятности 0.25 составляет 5.38527. Это означает, что площадь под кривой плотности хи-квадрат распределения с 4-мя степенями свободы справа от значения 5.38527 равна 0.25.



### 3. Проверка наличия взаимосвязи между переменными

---

#### ХИ-КВАДРАТ

- **Асимптотическая значимость** — вероятность случайности связи. **Чем меньше** эта величина, **тем выше** статистическая значимость (достоверность) связи. Величина  $p \leq 0,05$  свидетельствует о статистически значимом результате, который достоин содержательной интерпретации.
- **Точная значимость** —  $p$ -уровень значимости, вычисляемый точным методом; принимается во внимание, когда условия применения традиционного критерия нарушены (более 25 % ячеек таблицы сопряженности имеют частоту менее 5).
- **Линейно-линейная связь** — статистический критерий, определяющий степень корреляции между переменными.
- **Номинальная по номинальной** — меры связи для двух номинальных переменных.



### 3. Проверка наличия взаимосвязи между переменными

**Пример:** взаимосвязь EDUC и GENDER. Гипотеза о независимости переменных отвергается на уровне значимости  $\text{Sig.}(\alpha)=0,068$ . Это значит, что статистически значимой разницы в уровне образования мужчин и женщин нет, то есть отсутствует связь между полом и уровнем образования (inflation.sav).

Критерии хи-квадрат			
	Значение	ст.св.	Асимптотическая значимость (2-сторонняя)
Хи-квадрат Пирсона	8,750 <sup>a</sup>	4	,068
Отношения правдоподобия	8,807	4	,066
Линейно-линейная связь	8,511	1	,004
Количество допустимых наблюдений	1500		

а. Для числа ячеек 0 (0,0%) предполагается значение, меньше 5. Минимальное предполагаемое число равно 23,98.

Таблица 2.8. Интерпретация результатов расчетов  $\chi^2$ -статистики

Значимость теста	Интерпретация
Меньше 0,05	Гипотеза отвергается
0,05 – 0,1	Трудно сказать
Больше 0,1	Гипотеза принимается

- **Эмпирическое правило:** в таблице должно быть не больше 20% клеток, в которых ожидаемая частота  $<5$  и не должно быть клеток, в которых ожидаемая частота  $<1$ .
- Если обе переменные в таблице являются количественными, то при пометке элемента **Хи-квадрат** рассчитывается критерий линейно-линейной связи.

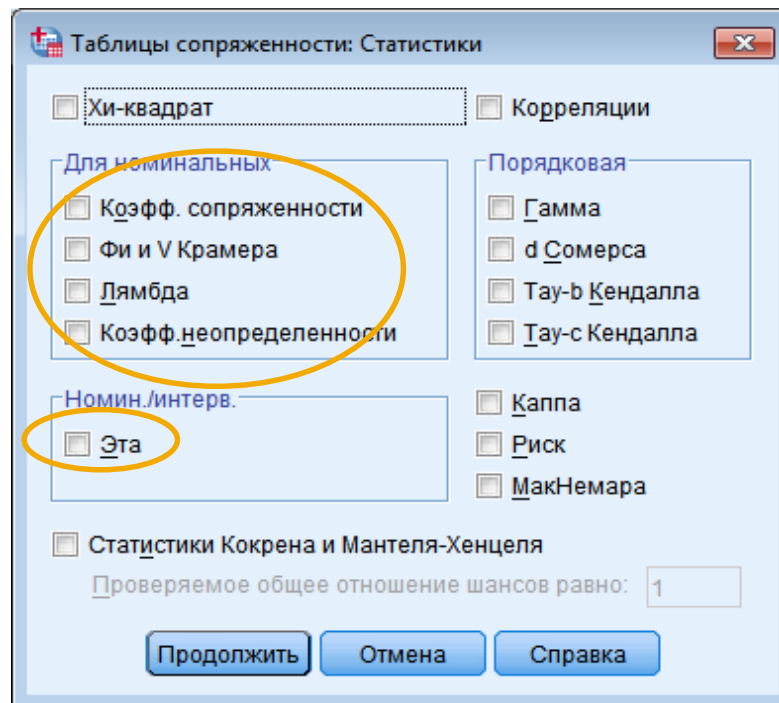
### 3. Проверка наличия взаимосвязи между переменными

Что еще можно посчитать на основе таблицы сопряженности?

**Коэфф. Сопряженности** - мера тесноты связи, основанная на критерии хи-квадрат (меняется между 0 и 1).

**Лямбда** - мера связи, которая отражает относительное снижение ошибки, когда значения независимой переменной используются для предсказания значений зависимой переменной.

**Коэфф.неопределенности** - мера связи, указывающая относительное снижение ошибки в случае, когда значения одной переменной используются для предсказания значений другой.

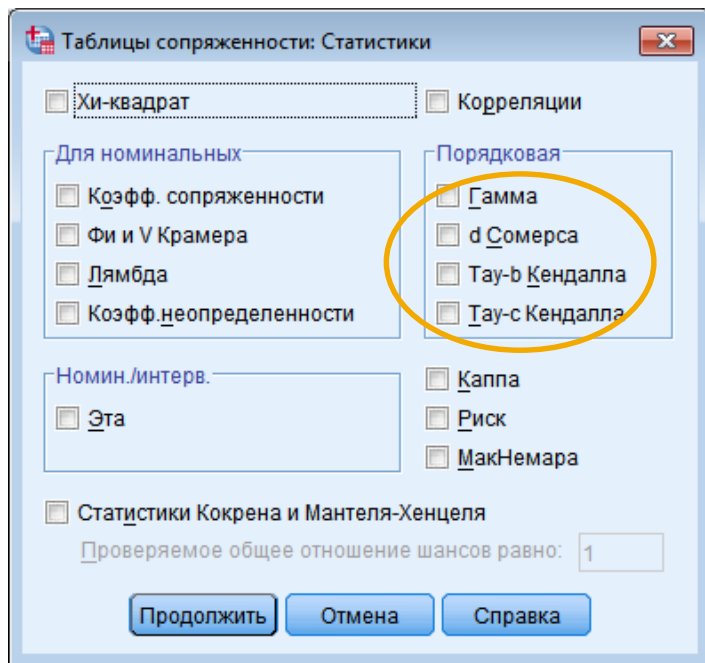


**Эта** - мера связи между переменными строки и столбца, значения которой изменяются от 0 (отсутствие связи) до 1 (сильная связь).



### 3. Проверка наличия взаимосвязи между переменными

Что еще можно посчитать на основе таблицы сопряженности?



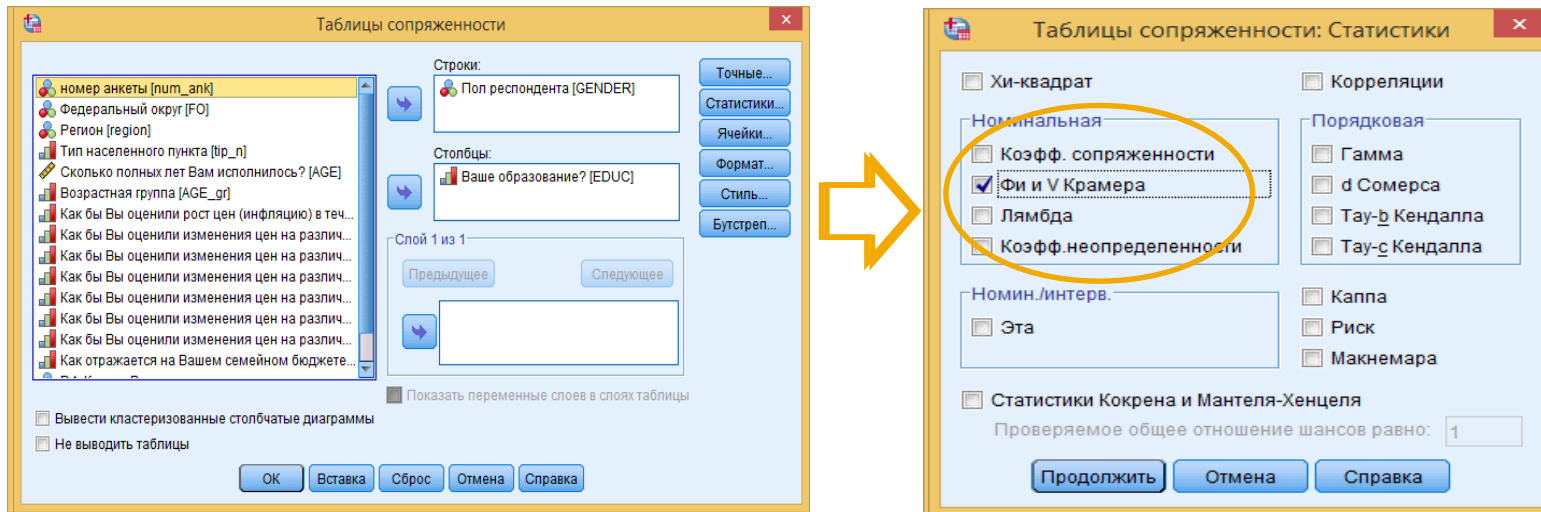
**Тау-с Кендалла** - непараметрическая мера связи для порядковых переменных, игнорирующая возможные совпадения значений (связи).

**Гамма** - симметричная мера связи между двумя порядковыми переменными ( $-1$  и  $1$ ). Значения, близкие по абсолютной величине к  $1$ , указывают на сильную связь переменных. Значения, близкие к  $0$ , говорят о слабой связи или ее отсутствии.

**D Сомерса** - мера связи между двумя порядковыми переменными, ( $-1$  и  $1$ ). Это асимметричное расширение меры гамма, отличающееся только включением числа пар, не имеющих совпадений (связей) по независимой переменной.

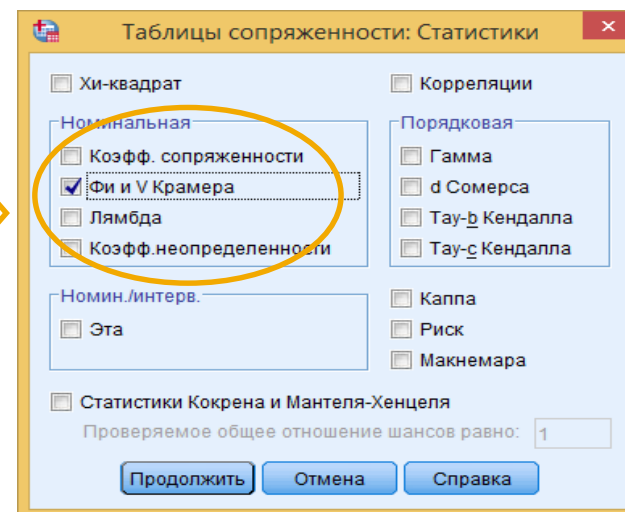
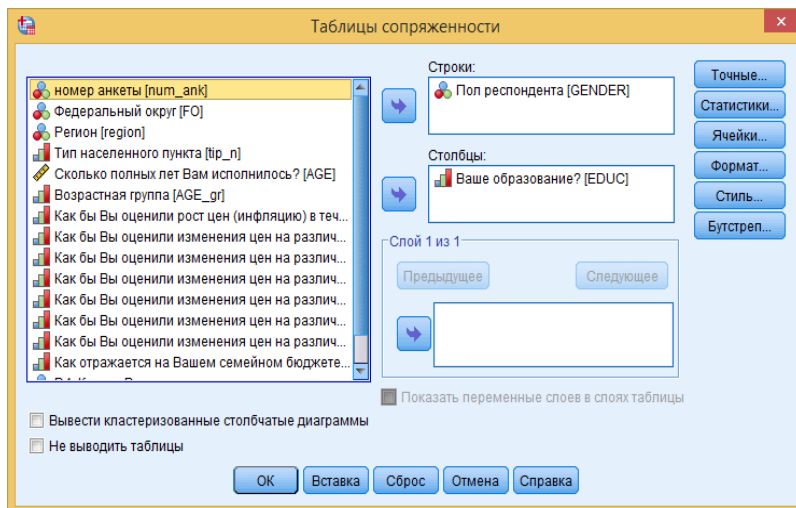
**Тау-б Кендалла** - непараметрическая мера корреляции для порядковых или ранговых переменных, учитывает возможные совпадения значений (связи). Знак коэффициента указывает направление связи, а его модуль - силу связи, причем, чем он больше, тем связь сильнее ( $-1$  и  $+1$ , однако  $-1$  и  $+1$  можно получить только для квадратных таблиц)

### 3. Проверка наличия взаимосвязи между переменными



- Есть много стандартных **коэффициентов связи** для номинальных переменных. Например, коэффициент **Крамера V**, значения –  $[0, 1]$ .  $V=0$  (признаки статистически независимы,  $V=1$  - между ними существует полная связь).
- Коэффициенты связи номинальных признаков всегда положительны, т.к. категории не упорядочены. В случае упорядоченных категорий (порядковые переменные) возможно вычислить не только силу связи, но и ее знак.

### 3. Проверка наличия взаимосвязи между переменными



Симметричные меры

	Значение	Примерная Знач.
Номинал/номинал Фи	,076	,068
В Крамера	,076	,068
Количество допустимых наблюдений	1500	

**Пример:** файл inflation.sav, взаимосвязь между образованием и полом.

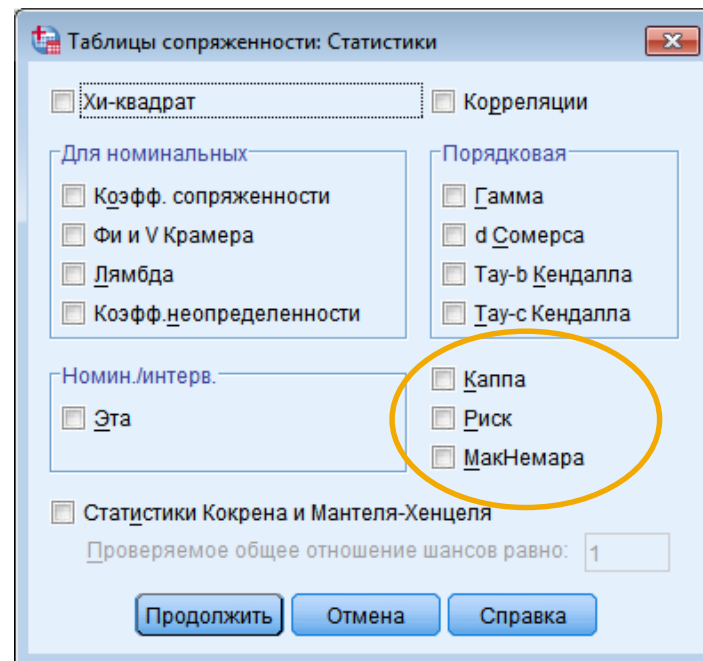
### 3. Проверка наличия взаимосвязи между переменными

**Коэффициент Каппа** - для квадратных таблиц, одинаковые кодировки у переменных;  $K$  – степень согласия.

**Мера риска** - расчет 3 коэффициентов: 2 относительного риска и отношение шансов; характерно для квадратных таблиц, дихотомические переменные.

**Хи-квадрат по Мак-Немару** - характерно для двух независимых дихотомических переменных; ситуации «до» и «после».

**Статистики Кокрена и Мантеля-Хэнцеля** – вычисление отношения шансов в таблицах сопряженности 2X2; рассмотрение отличия категорий переменной по своему отношению шансов от 1 (или другой величины).



### 3. Проверка наличия взаимосвязи между переменными

---

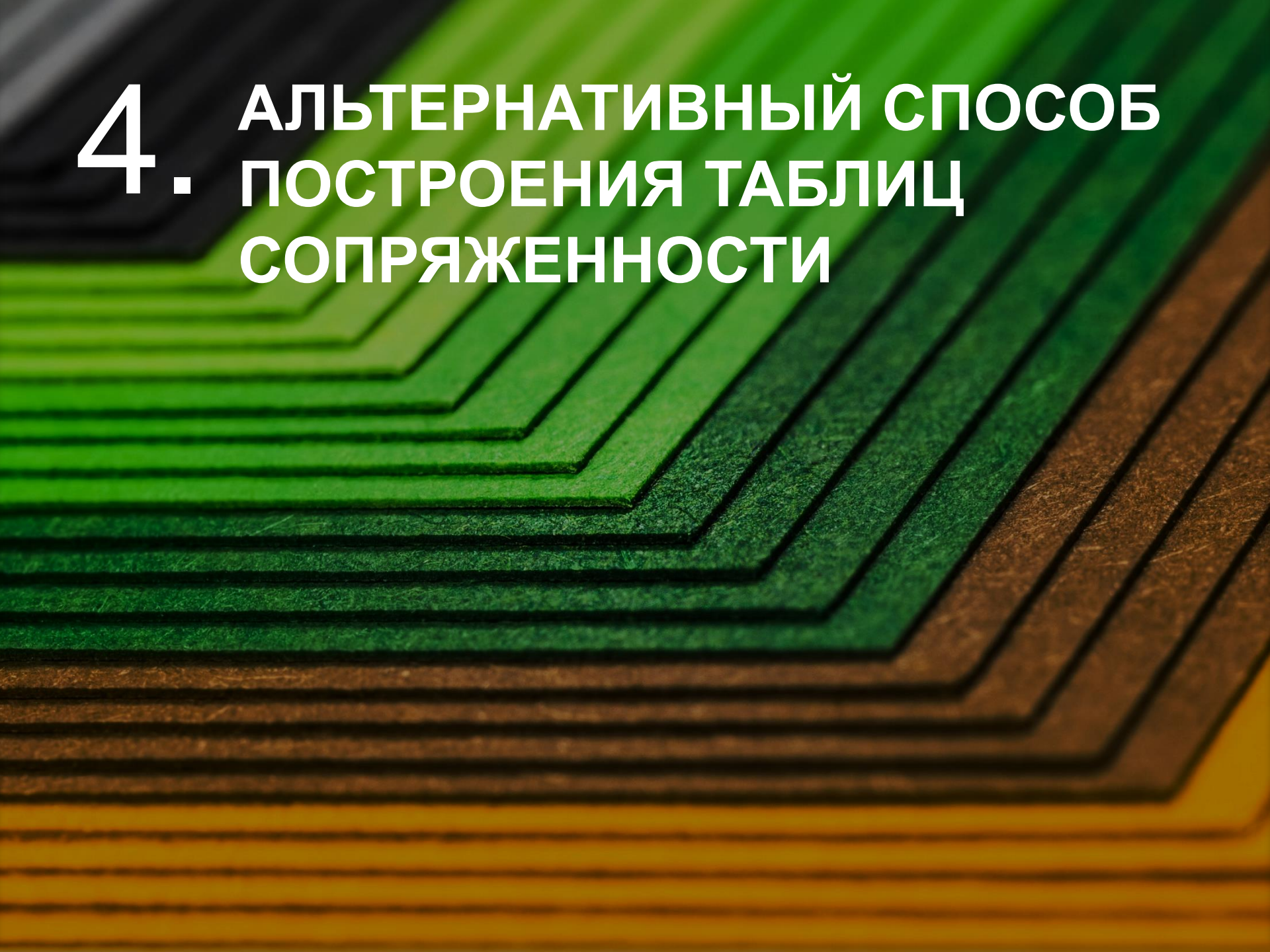
## ПРОБЛЕМА ИНТЕРПРЕТАЦИИ КОЭФФИЦИЕНТОВ

**Вопрос:** Если для одной пары переменных коэффициент Крамера равен 0,2, а для другой – 0,5, то означает ли это, что вторая пара переменных сильнее взаимосвязана, чем первая?

**Ответ:** Большее значение хи-квадрат, коэффициентов Крамера, Пирсона и других означает лишь меньшую вероятность того, что анализируемые переменные независимы. Они ничего не говорят ни о характере выявленной зависимости, ни о ее силе.



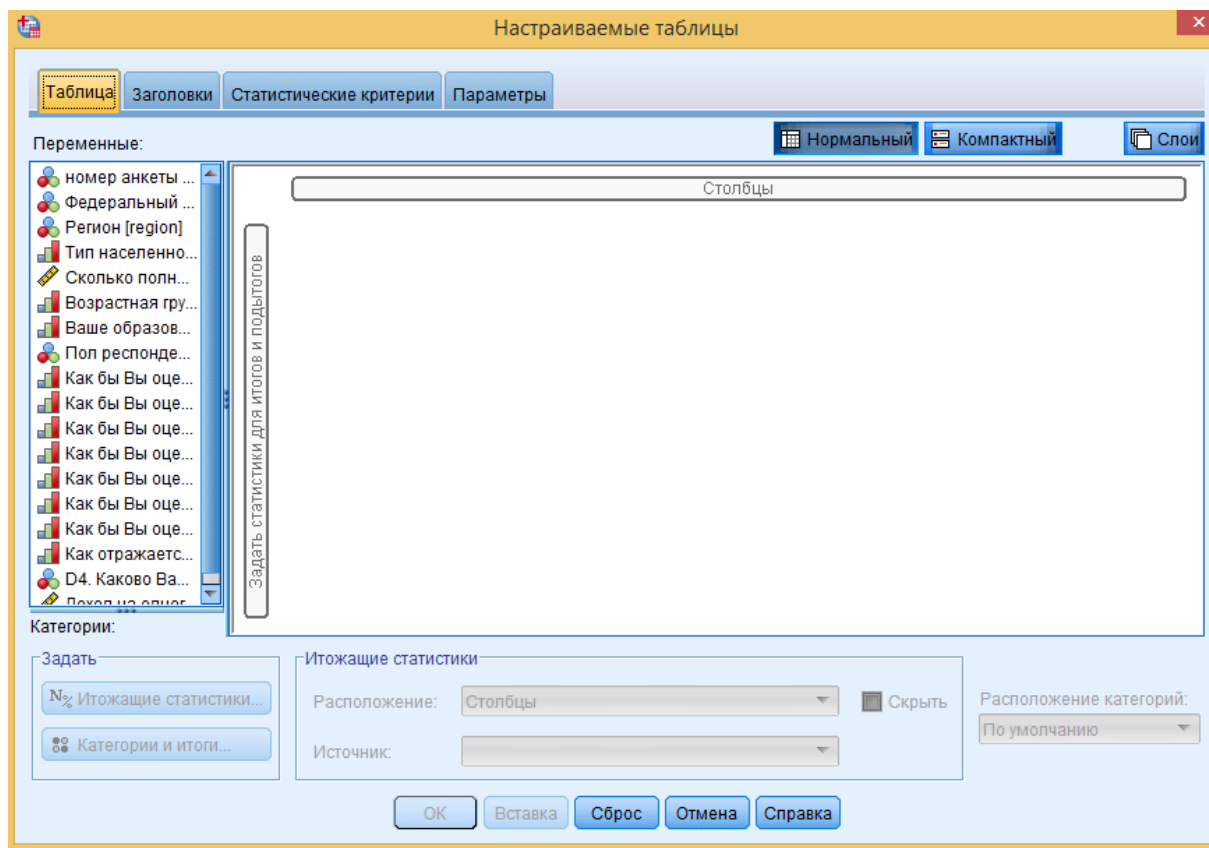
# 4. АЛЬТЕРНАТИВНЫЙ СПОСОБ ПОСТРОЕНИЯ ТАБЛИЦ СОПРЯЖЕННОСТИ



## 4. Альтернативный способ построения таблиц сопряженности

**Настраиваемые таблицы (Custom Tables)** – служат для создания таблиц, готовых к презентации. По сравнению с режимом построения Таблиц сопряженности, Настраиваемые таблицы предоставляют более широкие возможности для построения.

1. Меню «Анализ» → «Таблицы» → «Настраиваемые таблицы».





## 4. Альтернативный способ построения таблиц сопряженности

2. Чтобы создать таблицу, нужно просто перенести переменные в те места в таблице, в которых хотите их видеть:
- Перенесите переменную «Пол» в «Строки», «Ваше образование» в «Столбцы».

Настраиваемые таблицы

Таблица    Заголовки    Статистические критерии    Параметры

Переменные: Нормальный    Компактный    Слои

Задать статистику для итогов и подытогов

		Ваше образование?				
		Неполное ...	Среднее ...	Среднее ...	Незаконче...	Высшее ...
		Количество	Количество	Количество	Количество	Количество
Пол респондент...	Мужской	nnnn	nnnn	nnnn	nnnn	nnnn
	Женский	nnnn	nnnn	nnnn	nnnn	nnnn

Категории:

Задать

N% Итожащие статистики...

Категории и итоги...

Итожащие статистики

Расположение: Столбцы ☐ Скрыть

Источник: Переменные строк

Расположение категорий: По умолчанию

OK    Вставка    Сброс    Отмена    Справка

## 4. Альтернативный способ построения таблиц сопряженности

3. Кликните по кнопке **«Итожащие статистики»**. В данном контекстном меню можно выбрать, какие показатели необходимо вывести в итоговой таблице. Например, вы можете выбрать вывод процентных значений, частот или других статистик.
4. Выберите вывод **«% по строке»**.

Итожащие статистики:

Выбранная переменная: Пол респондента

Статистики:

- Невзвешенное количество
- % по строке
- % по столбцу
- % по таблице
- % по подтаблице

Показать:

Статистика	Метка	Формат	Знаков ...
Количество	Количество	nnnn	0

☐ Задать статистики для итогов и подытогов

Статистики:

- Невзвешенное количество
- % по строке
- % по столбцу
- % по таблице
- % по подтаблице

Показать:

Статистика	Метка	Формат	Знаков ...
Количество	Количество	nnnn	0

Применить к выделенному    Применить ко всем    Закреть    Справка

Настраиваемые таблицы

Статистические критерии    Параметры

Столбцы

		Неполное среднее ...		Среднее образование ...		Средне
		Количество	% по строке	Количество	% по строке	Количество
Пол респондент...	Мужской	nnnn	nnnn,n%	nnnn	nnnn,n%	nnnn
	Женский	nnnn	nnnn,n%	nnnn	nnnn,n%	nnnn

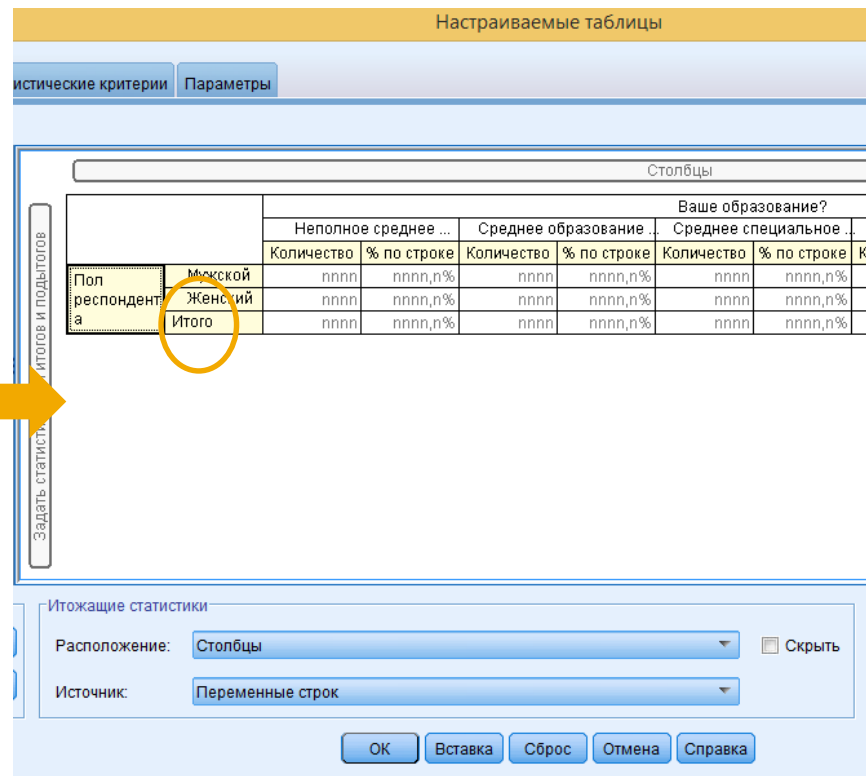
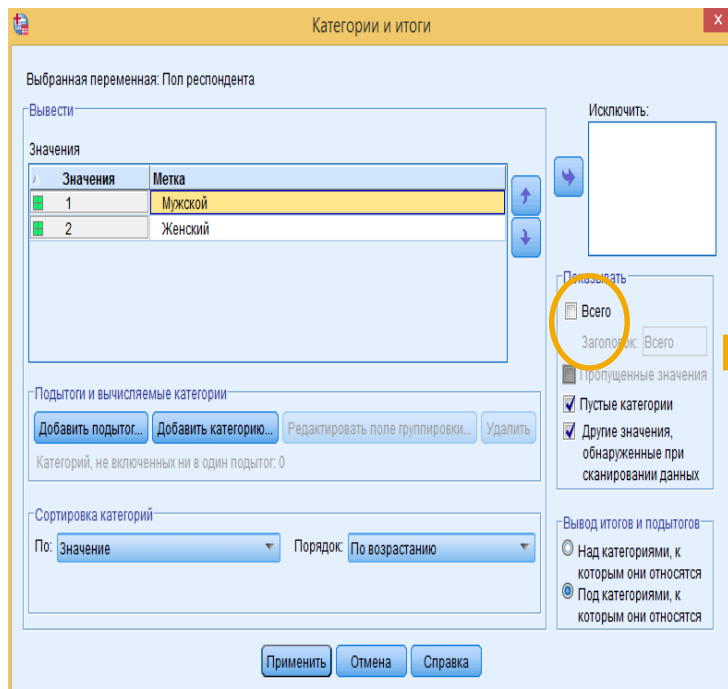
Задать статистики для итогов и подытогов

Итожащие статистики

Расположение: Столбцы

## 4. Альтернативный способ построения таблиц сопряженности

5. По умолчанию, в настраиваемых таблицах не выводятся итоги, но вывести их, а также подытоги, не составляет труда:
- Кнопка **«Категории и итоги»**
  - Чтобы включить общий итог, в группе параметров **«Показывать»** установите флажок **«Всего»**



## 4. Альтернативный способ построения таблиц сопряженности

6. Итоговая таблица содержит данные об уровне образования среди мужчин и женщин:

		Ваше образование?					
		Неполное среднее образование или ниже		Среднее образование (школа или ПТУ)		Среднее специальное образование (техникум)	
		Количество	% по строке	Количество	% по строке	Количество	% по строке
Пол респондента	Мужской	27	4,1%	151	22,7%	326	48,9%
	Женский	27	3,2%	161	19,3%	392	47,0%
	Итого	54	3,6%	312	20,8%	718	47,9%

7. Для более презентабельного представления таблицы, можно ввести названия заголовков:

- **«Анализ» → «Таблицы» → «Настраиваемые таблицы» → вкладка «Заголовки» → введите название «Уровень образования населения»**

		Уровень образования населения					
		Ваше образование?					
		Неполное среднее образование или ниже		Среднее образование (школа или ПТУ)		Среднее специальное образование (техникум)	
		Количество	% по строке	Количество	% по строке	Количество	% по строке
Пол респондента	Мужской	27	4,1%	151	22,7%	326	48,9%
	Женский	27	3,2%	161	19,3%	392	47,0%
	Итого	54	3,6%	312	20,8%	718	47,9%

## Литература по Теме 4

---

- 1. Бююль А., Цеффель П. SPSS: искусство обработки информации. – М., 2005.**
  - Глава 11. Таблицы сопряженности.
- 2. Наследов А. IBM SPSS Statistics 20 и AMOS: профессиональный статистический анализ данных. – СПб., 2013.**
  - Глава 8. Таблицы сопряженности и критерий Хи-квадрат.



---

Для свободного использования в образовательных целях  
Copyright 2017 © Академия НАФИ. Москва  
Все права защищены  
[www.nafi.ru](http://www.nafi.ru)