

Тема 1. Метод построения таблиц сопряженности

- ▣ **Таблицы сопряженности признаков** (или перекрестные таблицы) обычно строятся в случае, когда анализируются две качественные переменные и ставится вопрос о влиянии одной из них на другую.

Тема 1. Метод построения таблиц сопряженности

- ▣ **Критерий независимости χ^2** (χ^2 Пирсона; Хи-квадрат Пирсона; Pearson Chi-Square). Непараметрический критерий для сравнения ожидаемых и наблюдаемых частот (как правило, для таблиц сопряженности). Отвечает на вопрос: отличаются ли ожидаемые частоты модели от наблюдаемых

Тема 1. Метод построения таблиц сопряженности

Примеры задач

Постановка вопроса	Зависимая переменная	Независимая переменная
1. Зависит ли выбор товара покупателями от района их проживания?	Категории товаров	Район проживания
2. Зависит ли охват населения рекламой от возрасты?	Средства рекламы	Возрастные группы

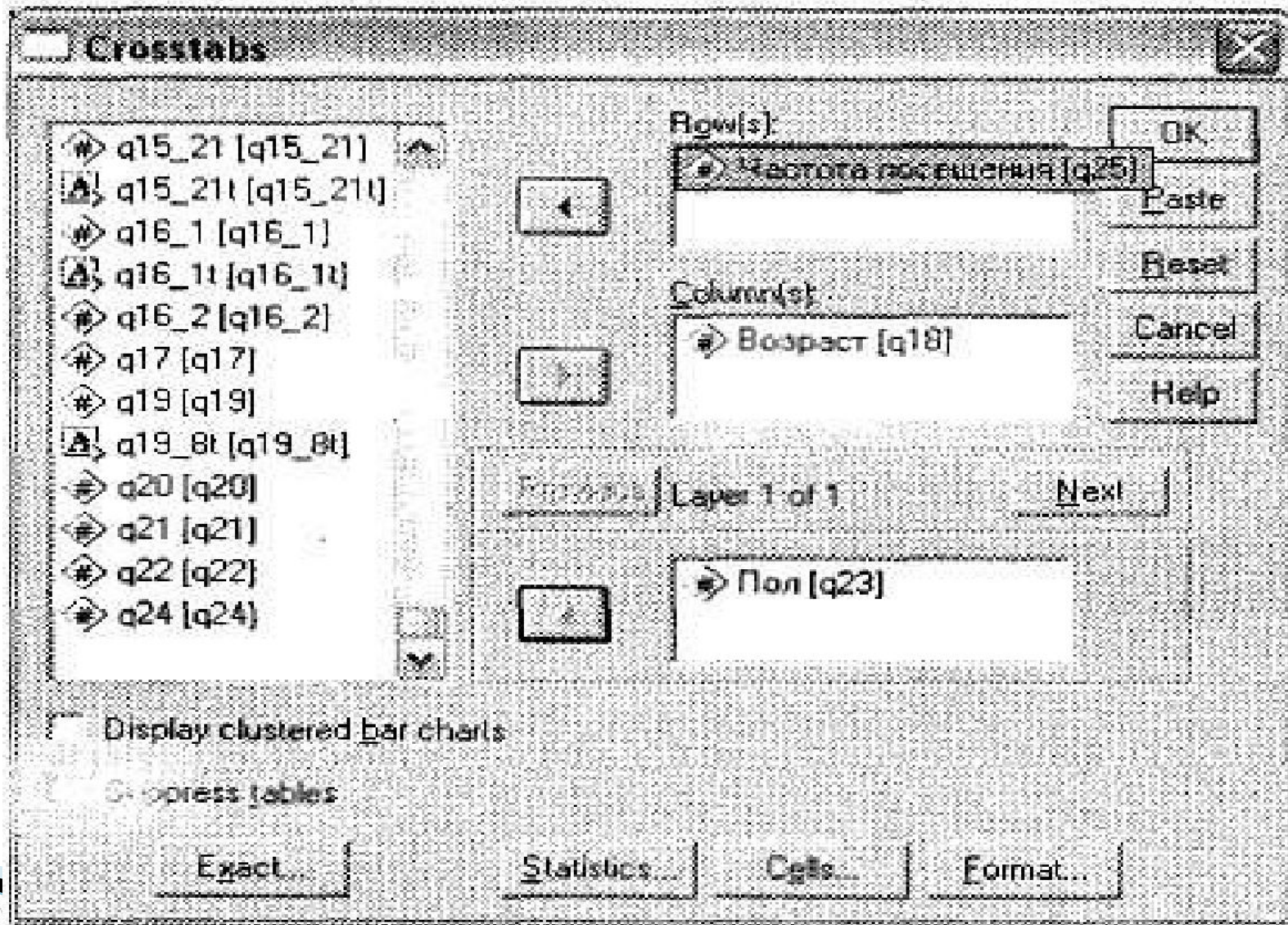
Тема 1. Метод построения таблиц сопряженности

- В результате маркетингового исследования, посвященного исследованию потребительских предпочтений посетителей развлекательного центра, оказалось, что средняя частота посещения центра составляет приблизительно 12 раз в месяц. Также были получены данные о распределении среди посетителей центра мужчин и женщин различных возрастных групп. В ходе подготовительного этапа анализа были сформированы, в частности, три одновариантные переменные:
- 1) частота посещения центра (q25);
- 2) возраст респондентов (q18);
- 3) пол респондентов (q23).

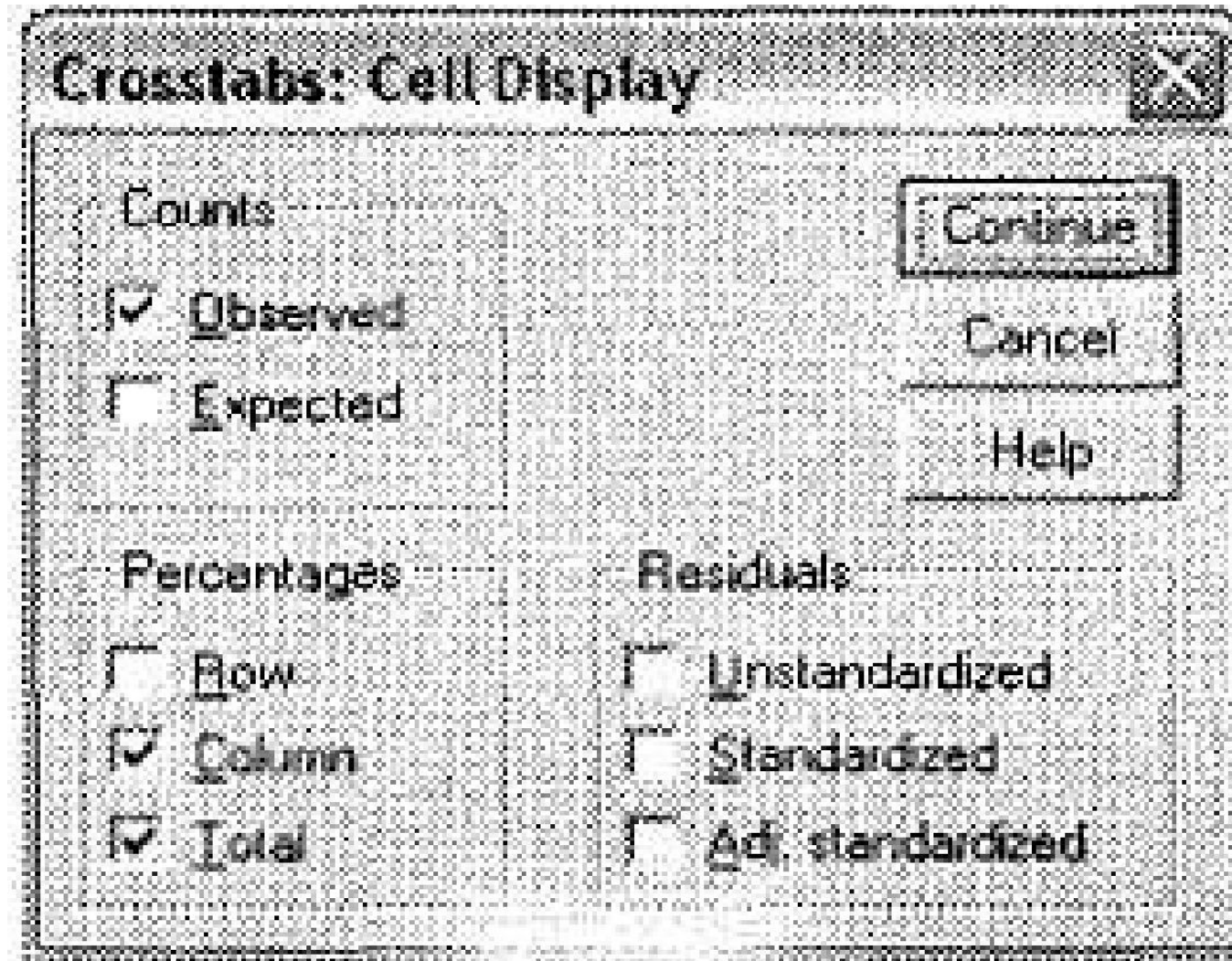
Тема 1. Метод построения таблиц сопряженности

- Построить перекрестное распределение частоты посещения развлекательного
- центра в разрезе возраста и пола респондентов. Рассчитать среднюю частоту посещения
- центра различными целевыми группами потребителей.
- 2. Определить, влияет ли на частоту посещения центра возраст потребителей.
- Установить статистическую значимость зависимости между частотой посещения и возрастом.

Тема 1. Метод построения таблиц сопряженности



Тема 1. Метод построения таблиц сопряженности



Частота посещения * Возраст * Пол Crosstabulation

				Возраст							Total	
				От 18 до 25 лет	От 26 до 30 лет	От 31 до 35 лет	От 36 до 40 лет	От 41 до 45 лет	От 46 до 50 лет	От 51 до 55 лет		Старше 55 лет
Пол Мужчины	Частота посещения	Каждый день	Count	5			2	4	1	1	2	15
			% within Возраст	8,1%			4,9%	10,0%	2,9%	3,7%	3,3%	4,5%
			% of Total	1,9%			,8%	1,2%	,3%	,3%	,8%	4,5%

	Total	Count	62	37	31	41	40	35	27	60	333	
		% within Возраст	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	
		% of Total	18,6%	11,1%	9,3%	12,3%	12,0%	10,5%	8,1%	18,0%	100,0%	
Женщины	Частота посещения	Каждый день	Count	1		3	3	1	2		1	11
			% within Возраст	1,6%		7,7%	7,0%	2,6%	4,4%		1,3%	2,9%
			% of Total	,3%		,8%	,8%	,3%	,5%		,3%	2,9%

		1-2 раза в месяц	Count	3	3	1	3	6	1	1	8	26
			% within Возраст	5,3%	8,4%	2,8%	7,0%	15,8%	2,2%	3,1%	10,3%	6,9%
			% of Total	,8%	,8%	,3%	,8%	1,6%	,3%	,3%	3,1%	6,9%

	Total	Count	67	47	38	43	38	45	32	77	378
		% within Возраст	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%
		% of Total	15,1%	12,4%	10,3%	11,4%	10,1%	11,9%	8,5%	20,4%	100,0%

Тема 1. Метод построения таблиц сопряженности

- В ячейках данной таблицы находятся искомые частоты посещения развлекательного центра каждой из анализируемых целевых групп
- опрошенных. Например, первая ячейка показывает, что 5 (строка Count) респондентов-мужчин в возрасте от 18 до 25 лет посещают развлекательный центр каждый день. Это составляет 8,1% (подстрока % within Возраст) от общего количества мужчин в возрасте от
- 18 до 25 лет, ответивших на три предложенных вопроса, или 1,5% (подстрока % of Total)
- от общего числа мужчин, ответивших на вопросы (это число 333, оно представлено на пересечении строки и столбца Total в первой части таблицы Мужчины).

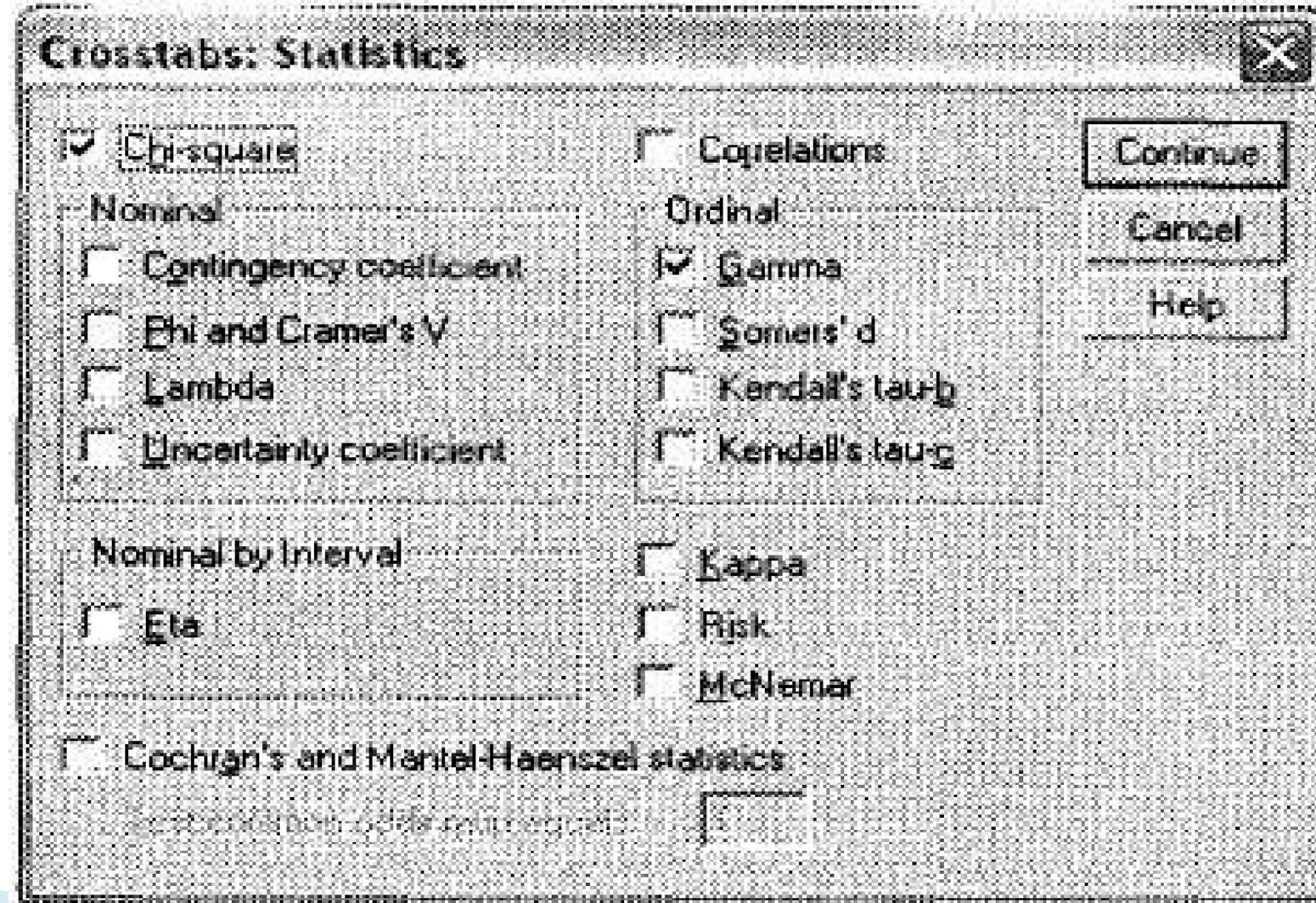
Тема 1. Метод построения таблиц сопряженности

- Строка Total показывает, сколько всего мужчин из каждой возрастной группы ответили на вопрос о частоте посещения центра (в нашем случае 62 респондента-мужчины в
- возрасте от 18 до 25 лет). Столбец Total показывает, сколько всего мужчин, посещающих
- развлекательный центр с различной частотой, ответили на вопрос о возрасте (в нашем
- случае 15 респондентов-мужчин, посещающих центр каждый день).
- Вторая часть таблицы Женщины построена аналогичным образом. Как вы видите,
- 15,8% женщин в возрасте от 41 до 45 лет посещают развлекательный центр 1-2 раза в месяц.

Тема 1. Метод построения таблиц сопряженности

- «Действительно ли существует статистически значимая зависимость между тремя анализируемыми переменными или показанные в табл. 4.2 различия в частотах посещения центра вызваны влиянием случайных факторов (то есть как
- таковой зависимости нет)?».
- Выявить статистическую значимость зависимостей между переменными позволяют критерий и сопутствующие тесты.

Тема 1. Метод построения таблиц сопряженности



Тема 1. Метод построения таблиц сопряженности

Chi-Square Tests

Пол		Value	df	Asymp. Sig. (2-sided)
Мужчины	Pearson Chi-Square	56,048 ^a	28	,001
	Likelihood Ratio	56,557	28	,001
	Linear-by-Linear Association	3,532	1	,060
	N of Valid Cases	333		
Женщины	Pearson Chi-Square	46,844 ^b	28	,014
	Likelihood Ratio	47,776	28	,011
	Linear-by-Linear Association	5,500	1	,019
	N of Valid Cases	378		

a. 22 cells (55,0%) have expected count less than 5. The minimum expected count is ,32.

b. 23 cells (57,5%) have expected count less than 5. The minimum expected count is ,17.

Тема 1. Метод построения таблиц сопряженности

Symmetric Measures

Пол			Value	Asymp. Std. Error ^a	Approx. T ^b	Approx. Sig.
Мужчины	Ordinal by Ordinal	Gamma	,080	,067	1,193	,233
	N of Valid Cases		333			
Женщины	Ordinal by Ordinal	Gamma	,148	,057	2,592	,010
	N of Valid Cases		378			

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

Symmetric Measures

Пол			Value	Approx. Sig.
Мужчины	Nominal by	Phi	,410	,001
	Nominal	Cramer's V	,205	,001
	N of Valid Cases		333	
Женщины	Nominal by	Phi	,352	,014
	Nominal	Cramer's V	,178	,014
	N of Valid Cases		378	

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

Тема 1. Метод построения таблиц сопряженности

- между тремя анализируемыми переменными — возрастом, полом и частотой посещения респондентами развлекательного центра — есть слабые, но статистически значимые зависимости. Вместе с тем было установлено, что больше
- половины (55 %) ячеек в перекрестной таблице имеют ожидаемые частоты меньше 5 — из
- чего следует вывод о неприменимости теста с 2 и сопутствующих асимптотических тестов (Gamma и Cramer's V) в нашем случае. В принципе мы ответили на второй пункт задачи и можем сказать, что различия, выявленные в ходе перекрестного
- анализа действительно имеют место и являются статистически значимыми.

Тема 1. Метод построения таблиц сопряженности

- Построение перекрестных распределений **по многовариантным переменным** на примере **двух многовариантных вопросов** из маркетингового исследования московского рынка сметаны.

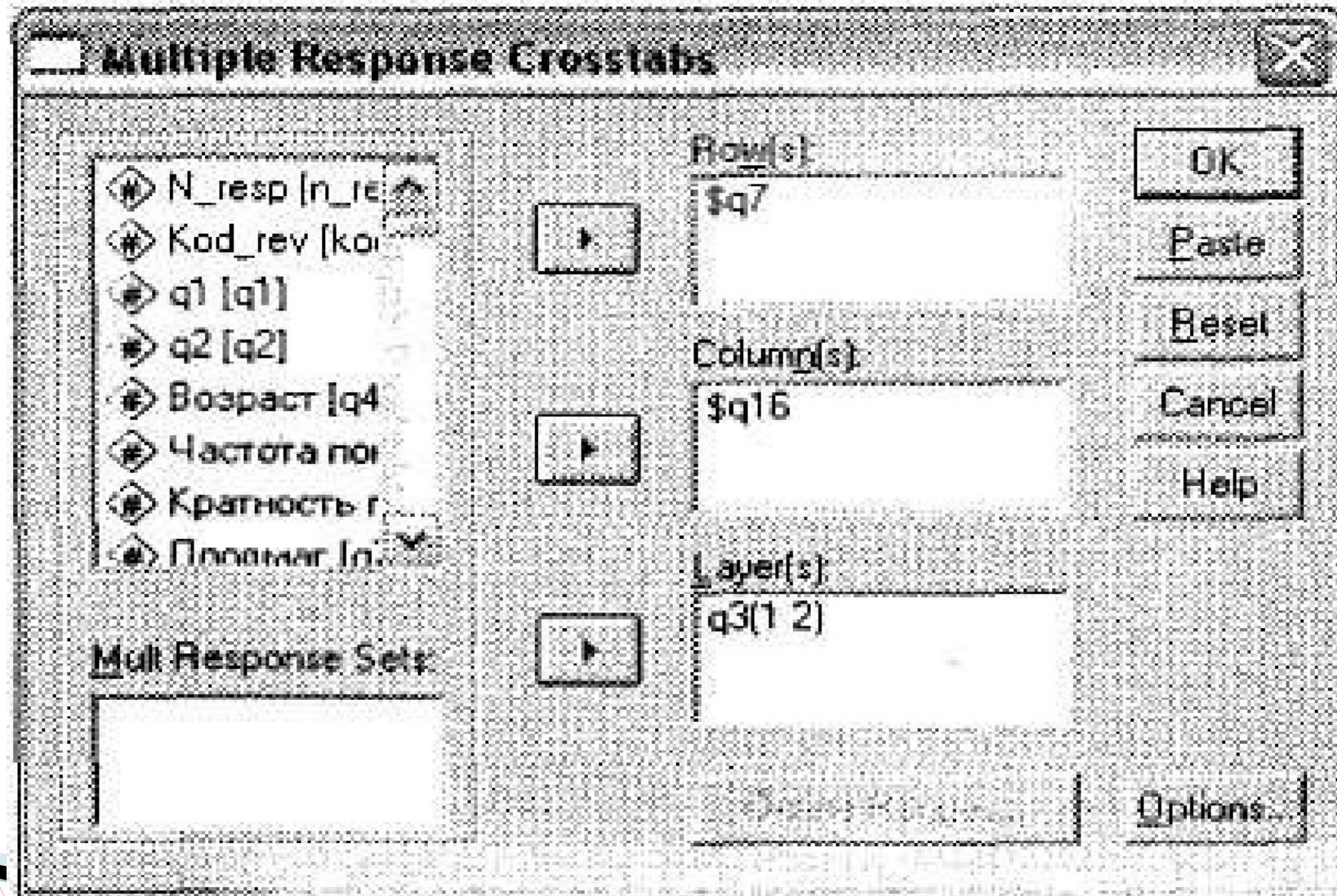
Тема 1. Метод построения таблиц сопряженности

- Первый вопрос: **Где Вы покупаете сметану?** (q7) с вариантами ответа:
 - ■ продмаг (q7_1);
 - ■ рынок (q7_2);
 - ■ супермаркет (q7_3);
 - ■ палатка (q7_4);
 - ■ универсам (q7_5).
- Второй вопрос: **Какую сметану Вы предпочитаете?** с вариантами ответа:
 - ■ в упаковке (q16_1);
 - ■ развесную (q16_2).

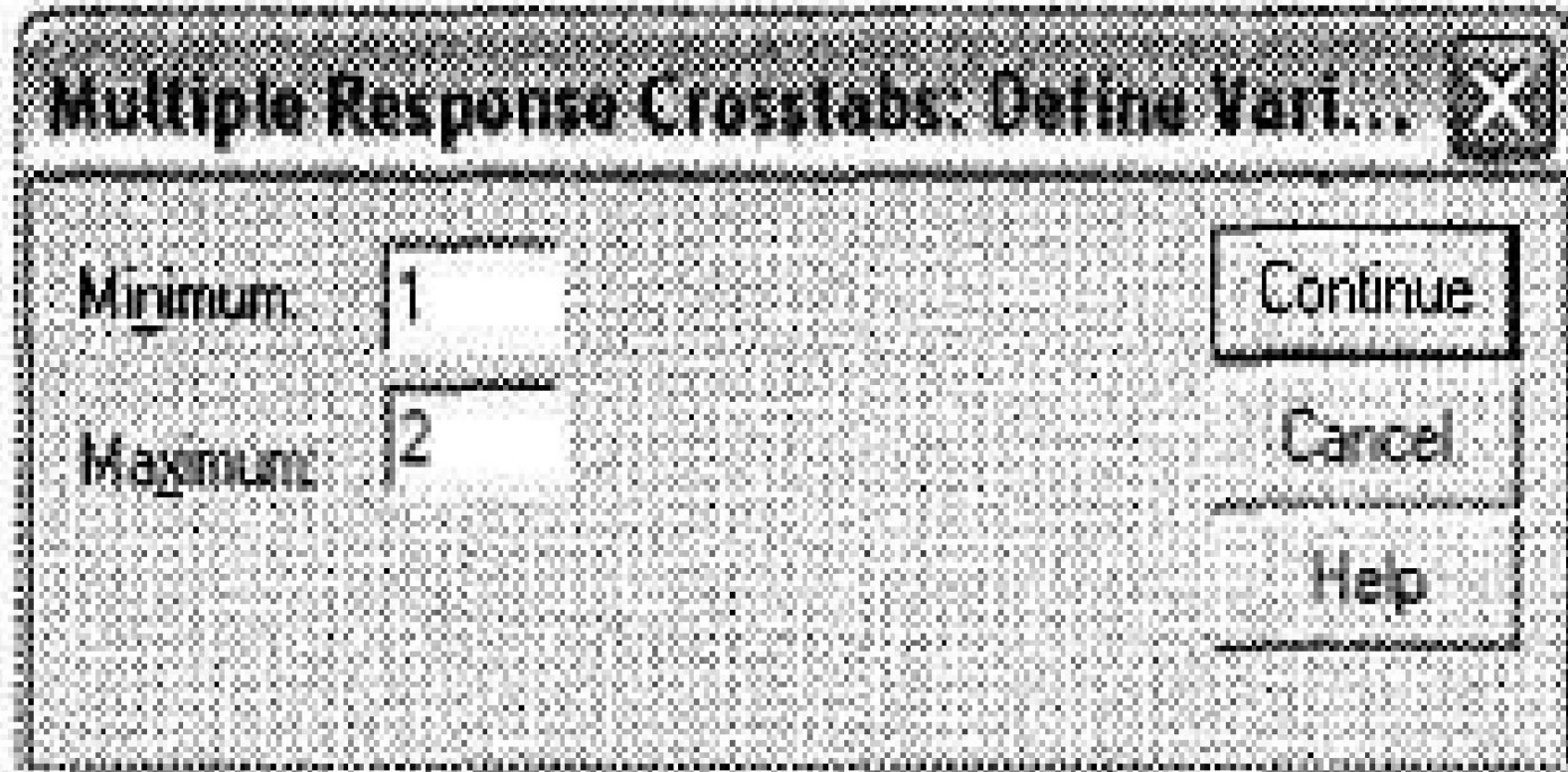
Тема 1. Метод построения таблиц сопряженности


- «Зависят ли предпочтения респондентов в отношении сметаны (упакованной или развесной) от места совершения покупки?»»

Тема 1. Метод построения таблиц сопряженности



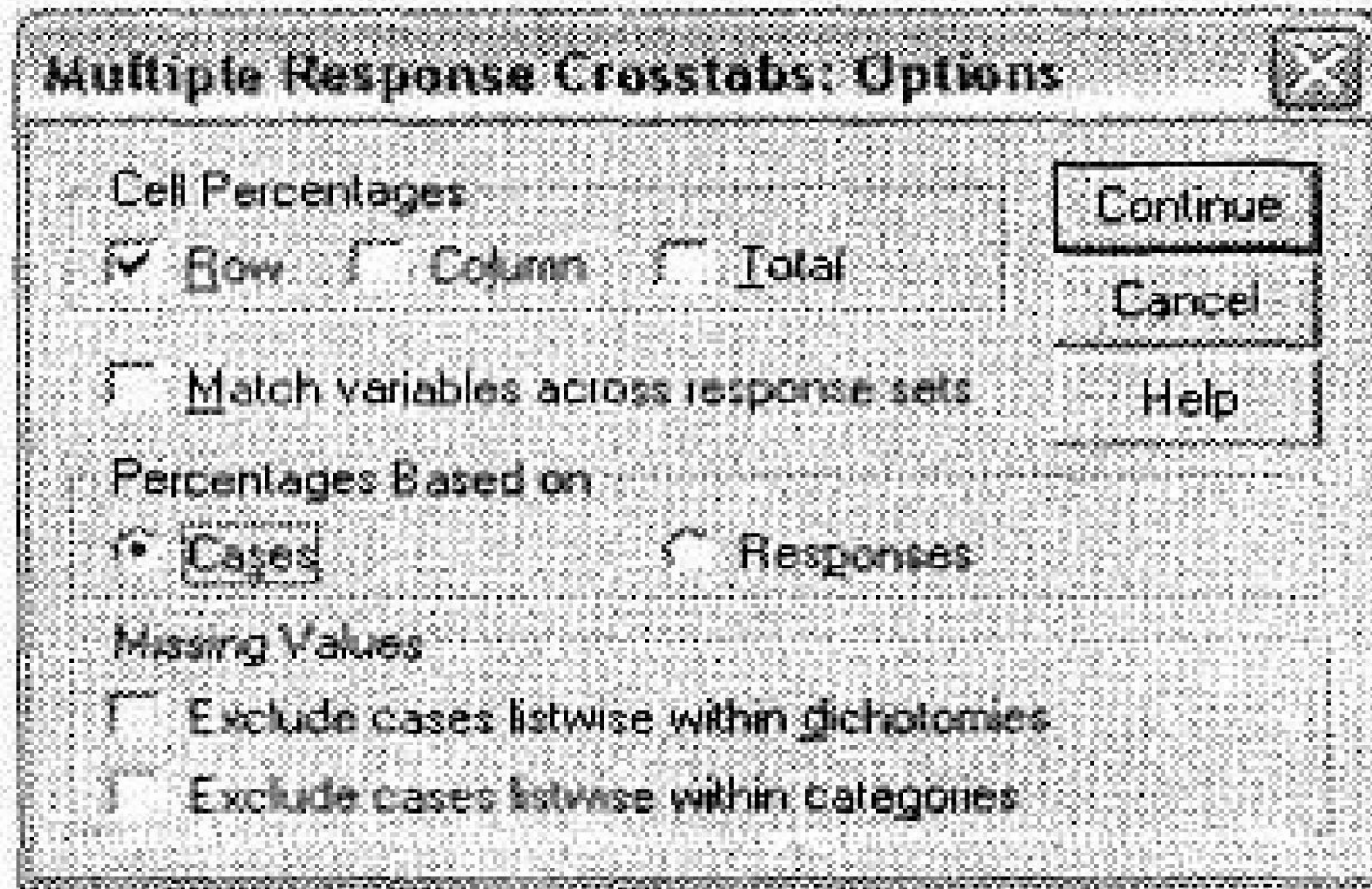
Тема 1. Метод построения таблиц сопряженности



Multiple Response Crosstabs: Define Vari... 

Minimum:	<input type="text" value="1"/>
Maximum:	<input type="text" value="2"/>

Тема 1. Метод построения таблиц сопряженности



Тема 1. Метод построения таблиц сопряженности

*** C R O S S T A B U L A T I O N ***

\$Q7 (tabulating 1) Место покупки сметаны
 by \$Q16 (tabulating 1) Предпочтения сметаны
 by Q3 Пол
 Category = 1 Мужчины

		\$Q16		
		Count	Сметана Развесна	
		Row pct	в упаков я сметан	Row
			ка а	Total
			Q16_1	Q16_2
\$Q7				
	Q7_1	71	23	94
Пропанг		75,5	24,5	51,9
	Q7_2	33	9	42
Рынок		78,6	21,4	23,2
	Q7_3	35	12	47
Супермаркет		74,5	25,5	26,0
	Q7_4	26	6	32
Палатка		81,3	18,8	17,7
	Q7_5	11	3	14
Универсам		78,6	21,4	7,7
	Column	143	38	181
	Total	79,0	21,0	100,0

Percents and totals based on respondents

Тема 1. Метод построения таблиц сопряженности

*** C R O S S T A B U L A T I O N ***

\$Q7 (tabulating 1) Место покупки сметаны
 by \$Q16 (tabulating 1) Предпочтения сметаны
 by Q3 Пол
 Category = 2 Женщины

		\$Q16			
		Count	Сметана	Развесна	
		Row pct	в упаков	а сметан	Row
			ка	а	Total
			Q16_1	Q16_2	
\$Q7					
	Q7_1		276	115	391
Продукт		70,6	29,4		51,5
	Q7_2		196	52	248
Рынок		79,0	21,0		32,7
	Q7_3		124	69	193
Супермаркет		64,2	35,8		25,4
	Q7_4		102	25	127
Палатка		80,3	19,7		16,7
	Q7_5		35	14	49
Универсам		71,4	26,6		6,5
	Column		564	195	759
	Total		74,3	25,7	100,0

Percents and totals based on respondents

940 valid cases; 63 missing cases