

# 卫生统计学实习

何平平

北京大学公共卫生学院  
流行病与卫生统计学系

**Tel: 82801619**

# 实习五

## 分类变量资料的统计推断

第226～235页

# 一、率的抽样误差与标准误

## (一) 定义

在抽样研究中，由于抽样造成的样本率与总体率之间的差异或者样本率之间的差异，称为率的抽样误差。

(二) 计算：率的抽样误差大小用率的标准误来衡量。

$$\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}}$$

一般情况下，由于我们研究的是样本， $\pi$ 未知，所以常用 $p$ 代替 $\pi$ ，得到率的标准误的估计值：

$$s_p = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{pq}{n}}$$

# 一、率的抽样误差与标准误

## （三）性质

- 率的标准误，与样本含量的平方根成反比。
- 在实际工作中，减小率的抽样误差的有效方法是增大样本含量。

## 二、总体率的置信区间的估计

（一）定义：按照预先给定的概率（通常取95%）来估计未知总体率所在的范围。

（二）计算：当 $n$ 足够大，且样本率 $p$ 和 $(1-p)$ 均不太小时，如 $np$ 与 $n(1-p)$ 均大于5时，采用正态近似法。总体率95%置信区间为：

$$p \pm 1.96S_p$$

### 三、完全随机设计的两样本率比较的假设检验

(一) 假设检验的目的：推断  $\pi_1 = \pi_2$  ？

$$H_0: \pi_1 = \pi_2 \quad H_1: \pi_1 \neq \pi_2$$

(二) 方法：

$$\chi^2 = \sum \frac{(A - T)^2}{T}$$

当  $\chi^2 \geq \chi^2_{0.05,1} = 3.84$  时， $P \leq 0.05$ ，拒绝  $H_0$ 。  
当  $\chi^2 < \chi^2_{0.05,1} = 3.84$  时， $P > 0.05$ ，接受  $H_0$ 。

### 三、完全随机设计的两样本率比较的假设检验

#### (三) $\chi^2$ 检验的适用条件及四格表专用公式

	+	—	合计
第一组	$a$	$b$	$a+b$
第二组	$c$	$d$	$c+d$
合计	$a+c$	$b+d$	$n$

### 三、完全随机设计的两样本率比较的假设检验

#### (三) $\chi^2$ 检验的适用条件及四格表专用公式

(1)  $n \geq 40$  且  $T \geq 5$ ,

$$\chi^2 = \sum \frac{(A-T)^2}{T} = \frac{(ad-bc)^2 n}{(a+b)(c+d)(a+c)(b+d)}$$

(2)  $n \geq 40$  且  $1 \leq T < 5$ ,

$$\chi^2 = \sum \frac{(|A-T|-0.5)^2}{T} = \frac{(|ad-bc|-n/2)^2 n}{(a+b)(c+d)(a+c)(b+d)}$$

(3)  $n < 40$  或  $T < 1$ ,

Fisher 精确概率法

注意: **a**  $T_{RC} = \frac{n_R n_C}{n}$     **b**  $\chi^2$ 检验是否需要校正, 关键要考核  $T_{min} \geq 5$ ?



## 四、配对设计的两样本率比较的假设检验

(一) 假设检验的目的：推断  $\pi_1 = \pi_2$  ?  
或者  $B=C$ ?

$$\begin{array}{ll} H_0: \pi_1 = \pi_2 & H_1: \pi_1 \neq \pi_2 \\ \text{或者 } H_0: B=C & H_1: B \neq C \end{array}$$

(二) 方法：

$$\chi^2 = \frac{(b - c)^2}{(b + c)}$$

当  $\chi^2 \geq \chi^2_{0.05,1} = 3.84$  时， $P \leq 0.05$ ，拒绝  $H_0$ 。

当  $\chi^2 < \chi^2_{0.05,1} = 3.84$  时， $P > 0.05$ ，接受  $H_0$ 。

## 四、配对设计的两样本率比较的假设检验

### (三) $\chi^2$ 检验的适用条件及四格表专用公式

第一组	第二组		合计
	+	-	
+	$a$	$b$	$a+b$
-	$c$	$d$	$c+d$
合计	$a+c$	$b+d$	$n$

1. 当  $b+c > 40$ , 
$$\chi^2 = \frac{(b-c)^2}{(b+c)}$$

2. 当  $b+c \leq 40$ , 
$$\chi^2 = \frac{(|b-c|-1)^2}{(b+c)}$$

**SPSS**软件中提供了精确概率法。

## 五、完全随机设计的多个样本率比较的假设检验

(一) 假设检验的目的：推断  $\pi_1 = \pi_2 = \dots = \pi_k$  ?

$H_0$ :  $\pi_1 = \pi_2 = \dots = \pi_k$

$H_1$ : 各总体率不等或不全相等

(二) 方法:

$$\chi^2 = \sum \frac{(A-T)^2}{T} = n \left( \sum \frac{A^2}{n_R n_C} - 1 \right)$$

$$\nu = (R-1) (C-1)$$

当  $\chi^2 \geq \chi^2_{0.05, \nu}$  时,  $P \leq 0.05$ , 拒绝  $H_0$  。

当  $\chi^2 < \chi^2_{0.05, \nu}$  时,  $P > 0.05$ , 接受  $H_0$  。

## 五、完全随机设计的多个样本率比较的假设检验

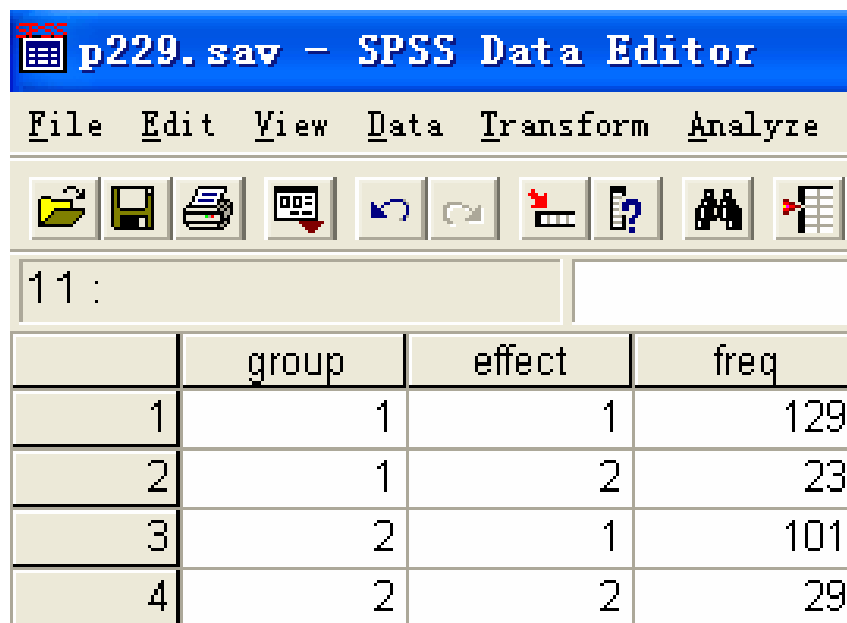
### (三) 注意事项

1. 当  $P \leq 0.05$ ，拒绝  $H_0$  时，**总的说来** 各组总体率有差别，**但并不意味着任何两组总体率都有差别**：可能是任何两者间都有差别，也可能其中某两者间有差别，而其它组间无差别。目前尚无公认的进一步两两比较的方法（可考虑采用 **Logistic** 回归）。
2. 当有1个格子的理论数小于1或者1/5以上格子的理论数小于5时，不宜采用卡方检验，推荐**精确概率法**（SPSS10.0以上已经有行×列表的精确概率法）。

## 六、SPSS13.0软件操作 (卡方检验)

### (一) 完全随机设计的两样本率比较的假设检验

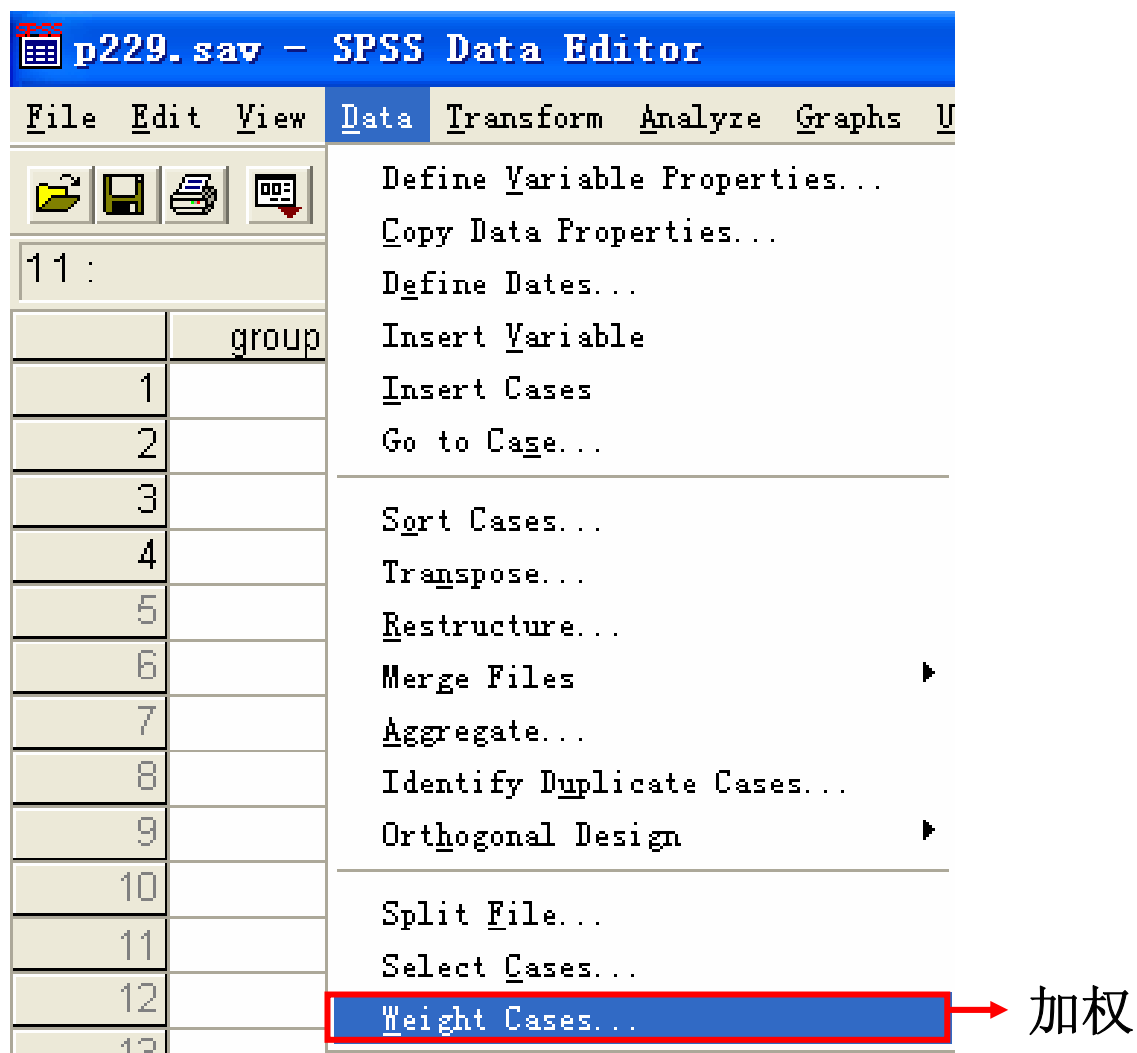
**例1** 见第229页例9-5。 **group**:组别, 1=旧剂型, 2=新剂型; **effect**:疗效, 1=治愈, 2=未愈; **freq**: 频数。



	group	effect	freq
1	1	1	129
2	1	2	23
3	2	1	101
4	2	2	29

## (一) 完全随机设计的两样本率比较的假设检验

### 例1



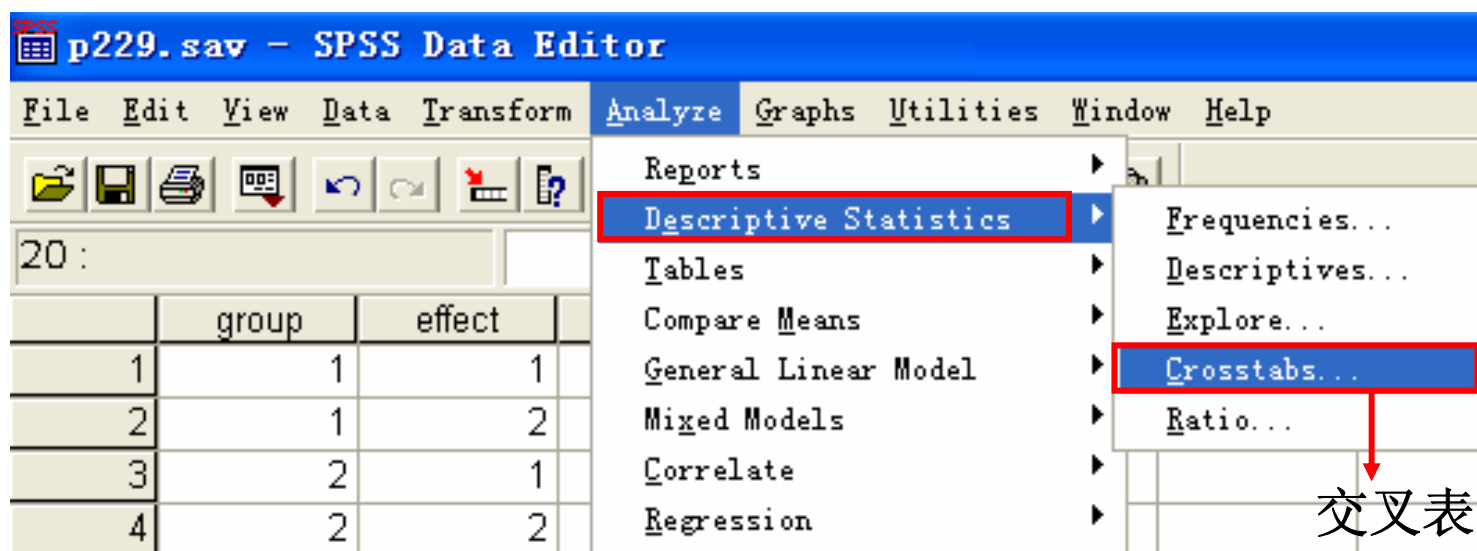
## （一）完全随机设计的两样本率比较的假设检验

### 例1



## （一）完全随机设计的两样本率比较的假设检验

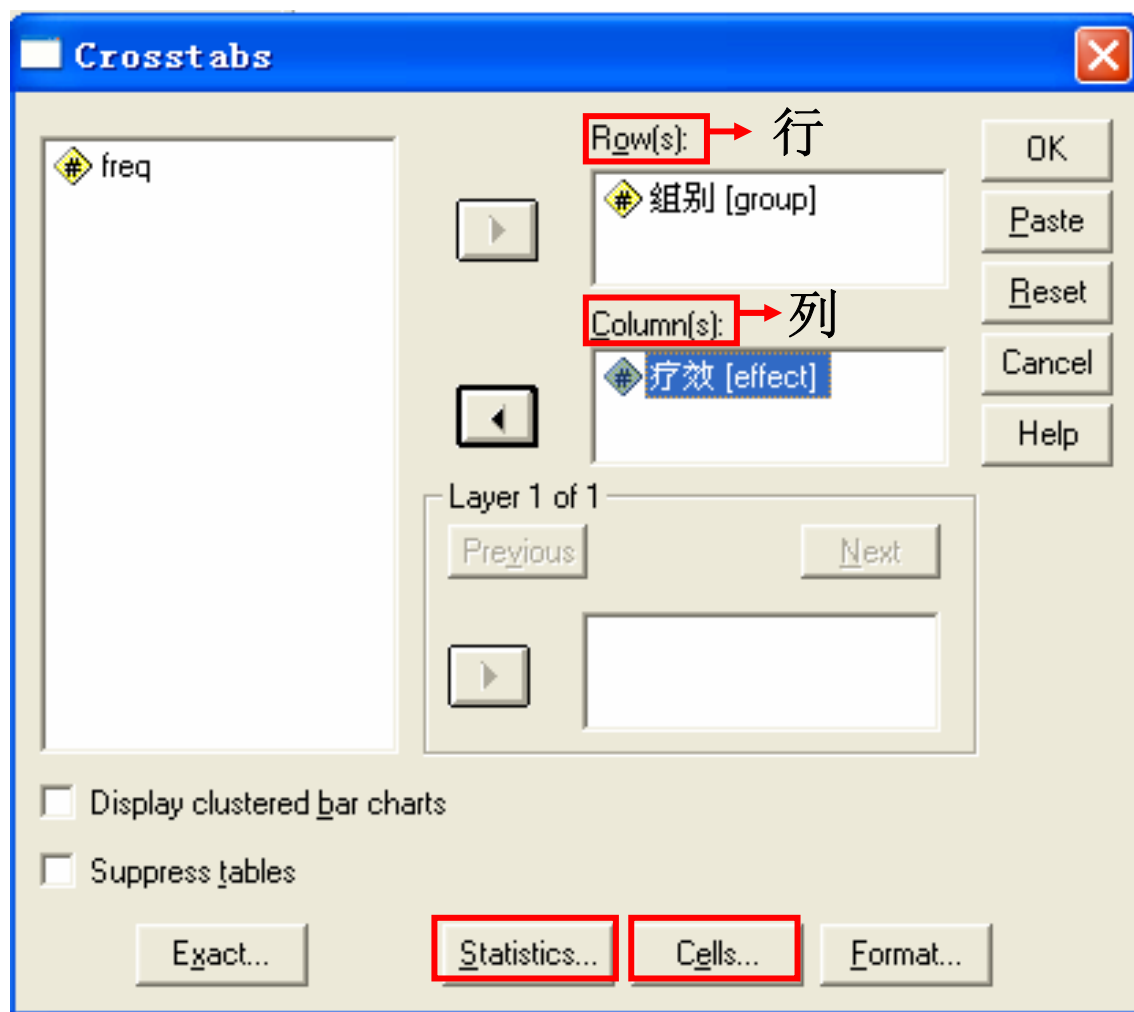
### 例1





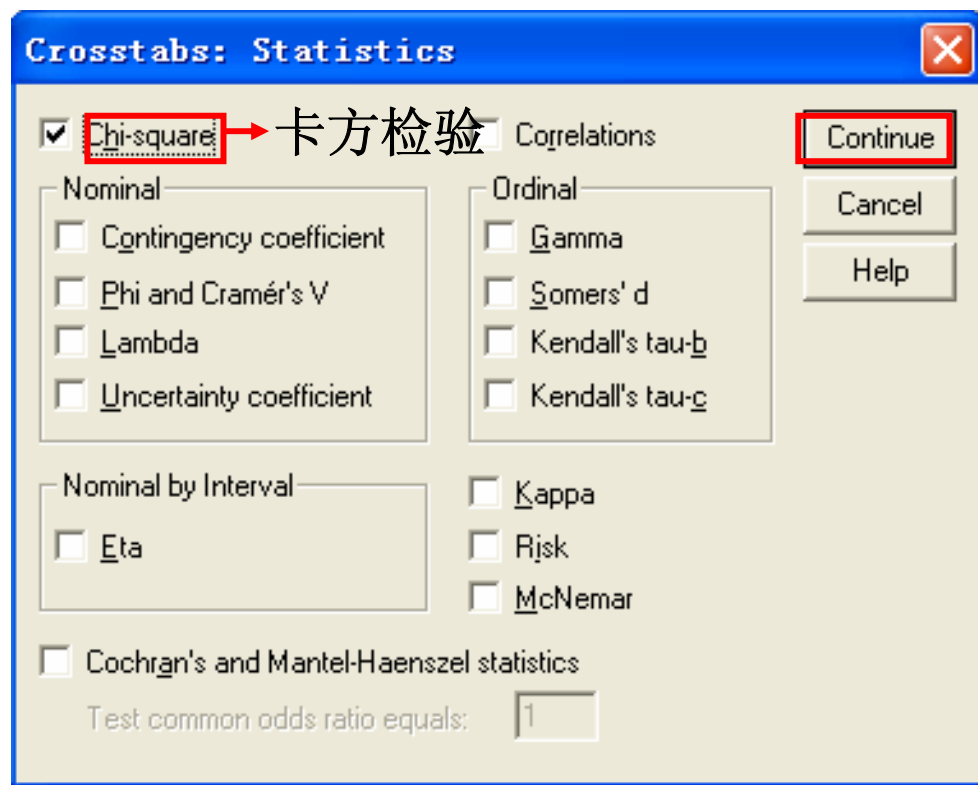
## (一) 完全随机设计的两样本率比较的假设检验

### 例1



## （一）完全随机设计的两样本率比较的假设检验

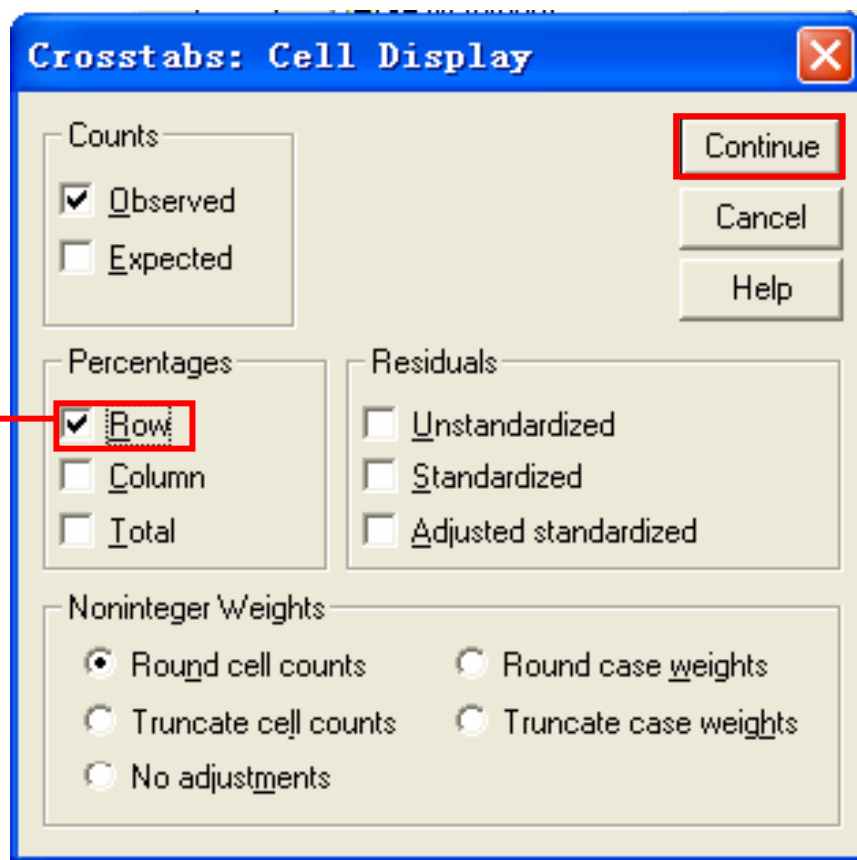
### 例1



## （一）完全随机设计的两样本率比较的假设检验

### 例1

行百分比



组别 \* 疗效 Crosstabulation

			疗效		Total
			治愈	未愈	
组别	旧剂型	Count	129	23	152
		% within 组别	84.9%	15.1%	100.0%
	新剂型	Count	101	29	130
		% within 组别	77.7%	22.3%	100.0%
	Total	Count	230	52	282
		% within 组别	81.6%	18.4%	100.0%

治愈率

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	2.399 <sup>b</sup>	1	.121	.127	.082
Continuity Correction <sup>a</sup>	1.946	1	.163		
Likelihood Ratio	2.393	1	.122		
Fisher's Exact Test					
Linear-by-Linear Association	2.391	1	.122		
N of Valid Cases	282				

a. Computed only for a 2x2 table

b. 0 cells (.0%) have expected count less than 5. The minimum expected count is 23.97.

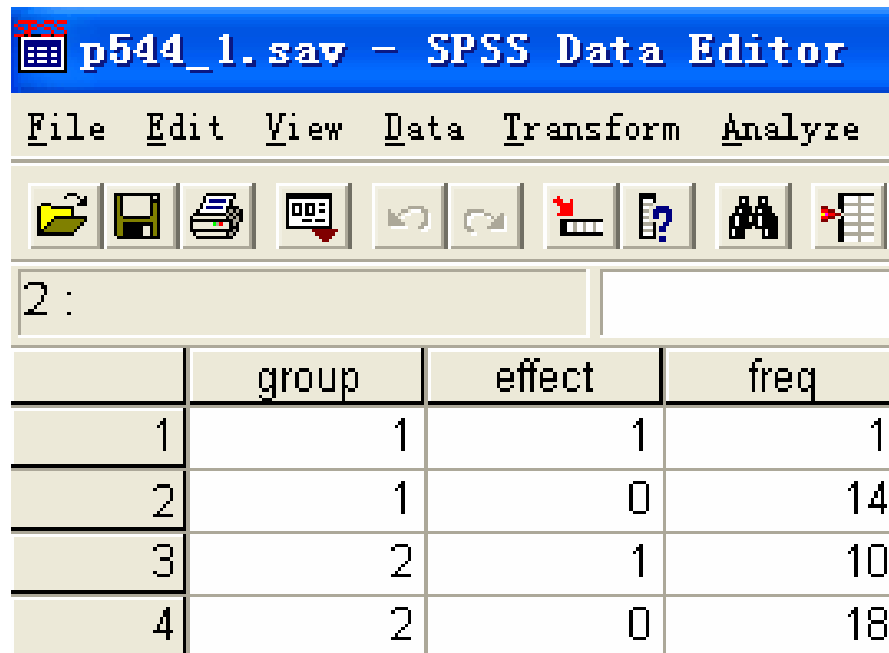
0格理论数小于5，最小理论数为23.97。

未校正的  
Pearson卡方  
检验

P值

## （一）完全随机设计的两样本率比较的假设检验

**例2** 见第544页计算分析题1。**group**:组别, 1=新防护衣, 2=旧防护衣; **effect**:患病情况, 1=患病, 0=未患病; **freq**:频数。(SPSS软件操作步骤同例1)



The screenshot shows the SPSS Data Editor window for a file named 'p544\_1.sav'. The window has a menu bar (File, Edit, View, Data, Transform, Analyze) and a toolbar with icons for opening, saving, printing, and other functions. Below the toolbar, there is a grid of data. The first column contains row numbers 1 through 4. The second column is labeled 'group' and contains values 1, 1, 2, 2. The third column is labeled 'effect' and contains values 1, 0, 1, 0. The fourth column is labeled 'freq' and contains values 1, 14, 10, 18.

	group	effect	freq
1	1	1	1
2	1	0	14
3	2	1	10
4	2	0	18

分组 \* 患病情况 Crosstabulation

			患病情况		Total
			未患病	患病	
分组	新防护衣	Count	14	1	15
		% within分组	93.3%	6.7%	100.0%
	旧防护衣	Count	18	10	28
		% within分组	64.3%	35.7%	100.0%
Total	Count	32	11	43	
	% within分组	74.4%	25.6%	100.0%	

患病率

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	4.329 <sup>b</sup>	1	.037	P值	.038
Continuity Correction <sup>a</sup>	2.938	1	.087		
Likelihood Ratio	5.056	1	.025		
Fisher's Exact Test					
Linear-by-Linear Association	4.228	1	.040	.065	
N of Valid Cases	43				

a. Computed only for a 2x2 table

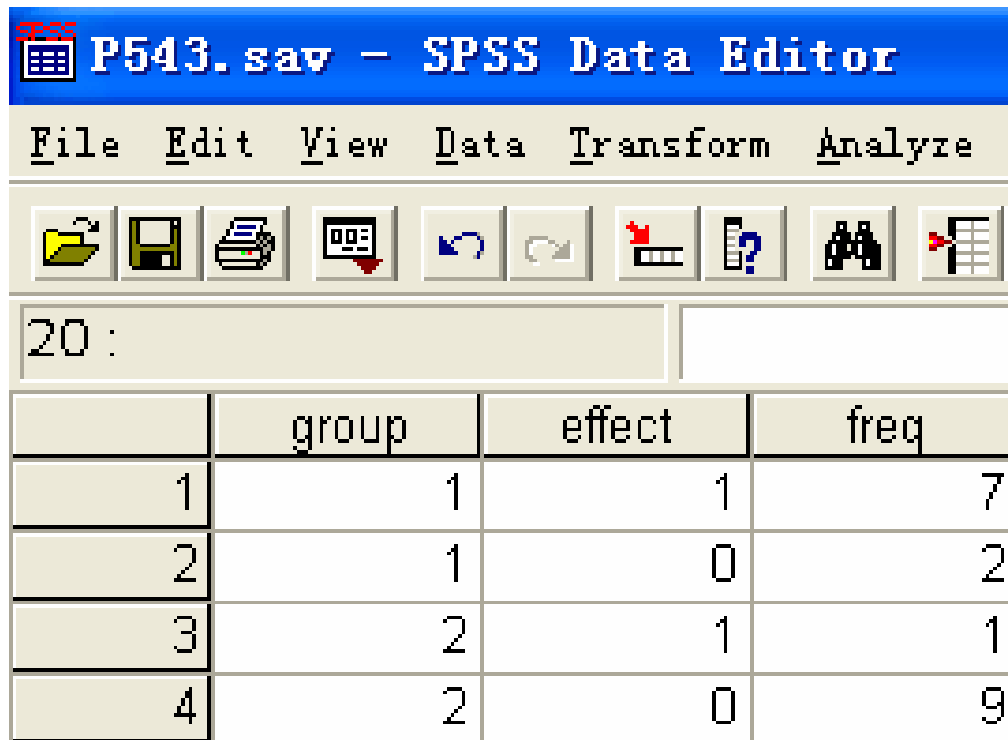
b. 1 cells (25.0%) have expected count less than 5. The minimum expected count is 3.84.

1格理论数小于5，最小理论数为3.84

连续性校正的卡方检验

## （一）完全随机设计的两样本率比较的假设检验

**例3** 见第543页选择题8。**group**:组别, 1=A药, 2=B药;  
**effect**:治愈情况, 1=治愈, 0=未愈; **freq**: 频数。(SPSS  
软件操作步骤同例1)



The screenshot shows the SPSS Data Editor window for a file named 'P543.sav'. The window has a menu bar with 'File', 'Edit', 'View', 'Data', 'Transform', and 'Analyze'. Below the menu bar is a toolbar with icons for opening, saving, printing, and other functions. The data is displayed in a table with 4 rows and 4 columns. The first column is labeled '20:' and contains row numbers 1 through 4. The second column is labeled 'group' and contains values 1, 1, 2, 2. The third column is labeled 'effect' and contains values 1, 0, 1, 0. The fourth column is labeled 'freq' and contains values 7, 2, 1, 9.

20:	group	effect	freq
1	1	1	7
2	1	0	2
3	2	1	1
4	2	0	9

### 例3

分组 \* 患病情况 Crosstabulation

			患病情况		Total
			未愈	治愈	
分组	A药	Count	2	7	9
		% within 分组	22.2%	77.8%	100.0%
	B药	Count	9	1	10
		% within 分组	90.0%	10.0%	100.0%
Total	Count		11	8	19
	% within 分组		57.9%	42.1%	100.0%

$n < 40$

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	8.927 <sup>b</sup>	1	.003		
Continuity Correction <sup>a</sup>	6.363	1	.012		
Likelihood Ratio	9.828	1	.002		
Fisher's Exact Test				.005	.005
Linear-by-Linear Association	8.457	1	.004		
N of Valid Cases	19				

Fisher  
精确概  
率法

P值

a. Computed only for a 2x2 table

b. 2 cells (50.0%) have expected count less than 5. The minimum expected count is 3.79.

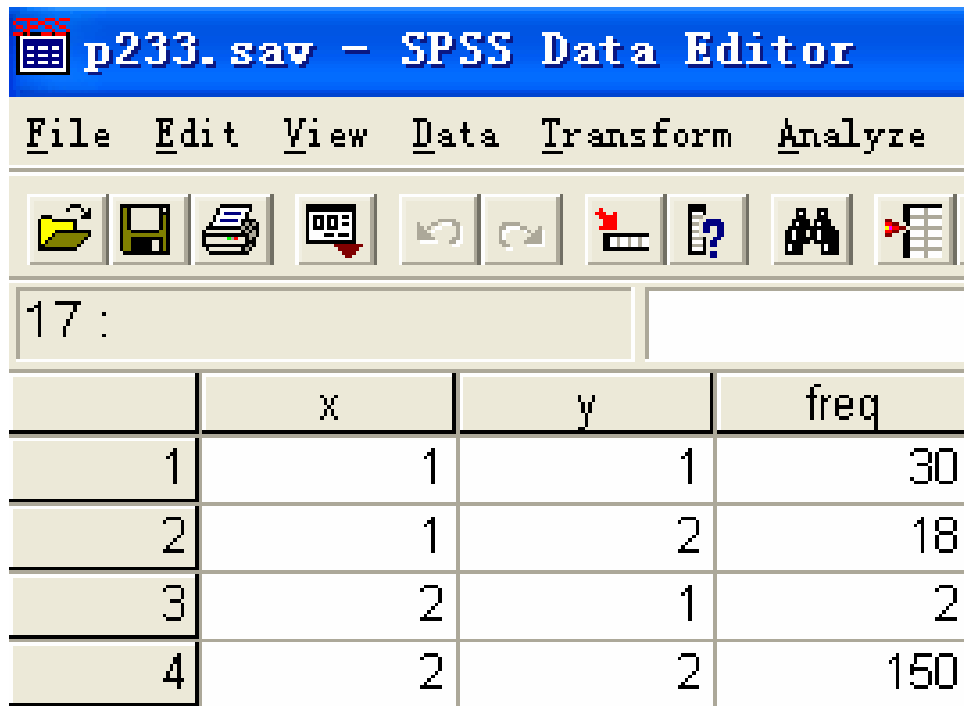
2格理论数小于5，最小理论数为3.79



## （二）配对设计的两样本率比较的假设检验

**例4** 见第233页例9—8。**X**:酶标法, 1=阳性, 2=阴性;  
**Y**:反向血凝法, 1=阳性, 2=阴性; **freq**: 频数。

（先加权, 同例1）

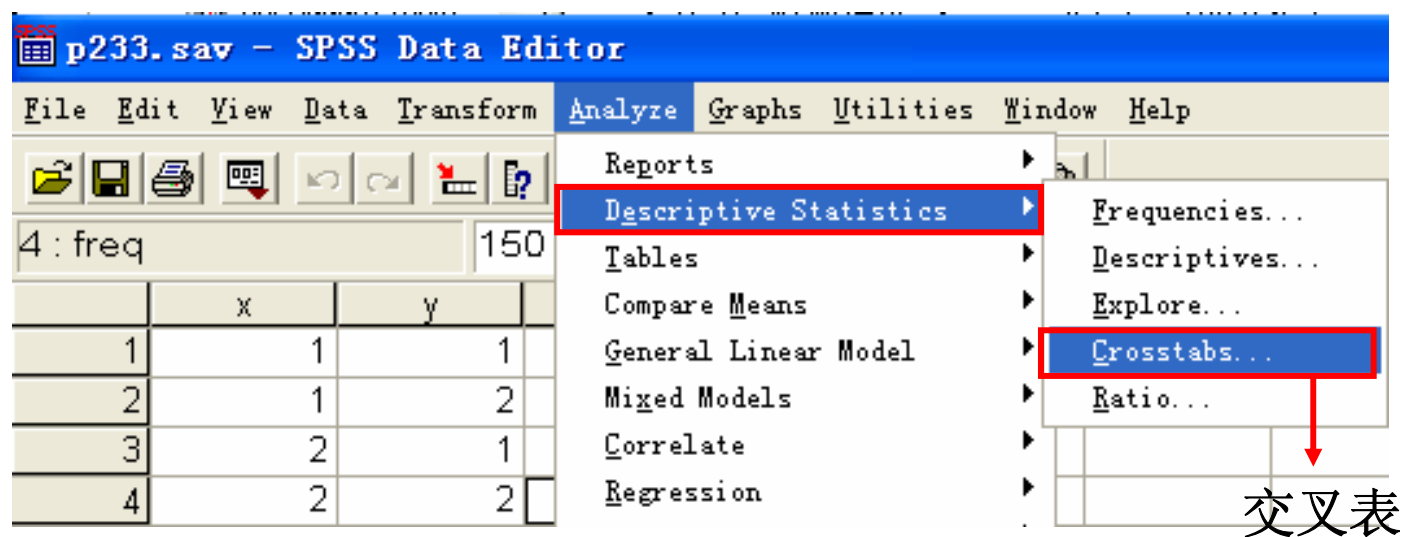


The screenshot shows the SPSS Data Editor window titled "p233.sav - SPSS Data Editor". The menu bar includes File, Edit, View, Data, Transform, and Analyze. Below the menu is a toolbar with icons for opening, saving, printing, and other functions. The data grid shows a table with 4 rows and 4 columns. The first column is labeled "17:" and contains row numbers 1 through 4. The second column is labeled "x" and contains values 1, 1, 2, 2. The third column is labeled "y" and contains values 1, 2, 1, 2. The fourth column is labeled "freq" and contains values 30, 18, 2, 150.

	x	y	freq
1	1	1	30
2	1	2	18
3	2	1	2
4	2	2	150

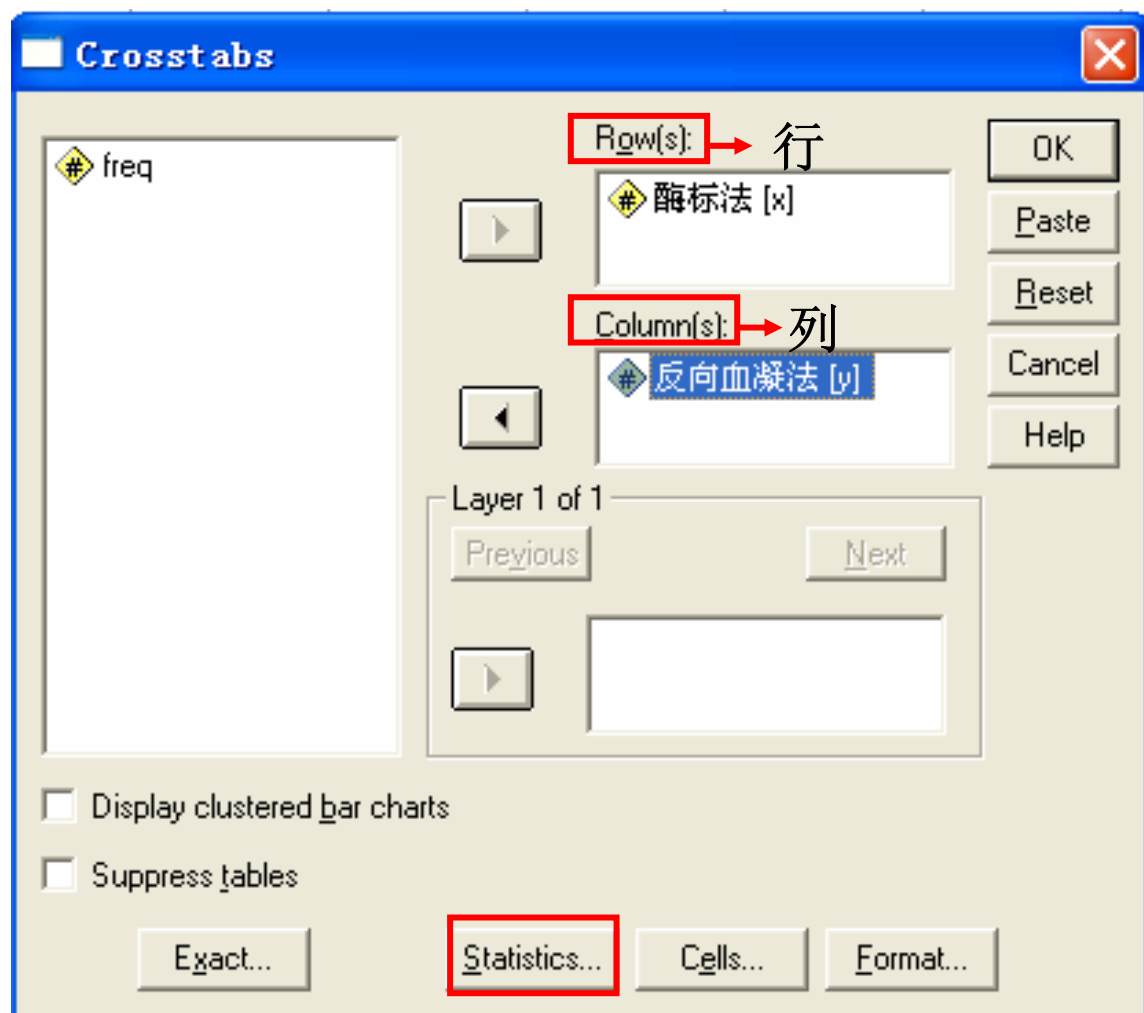
## (二) 配对设计的两样本率比较的假设检验

### 例4 方法一 (SPSS菜单: Crosstabs)



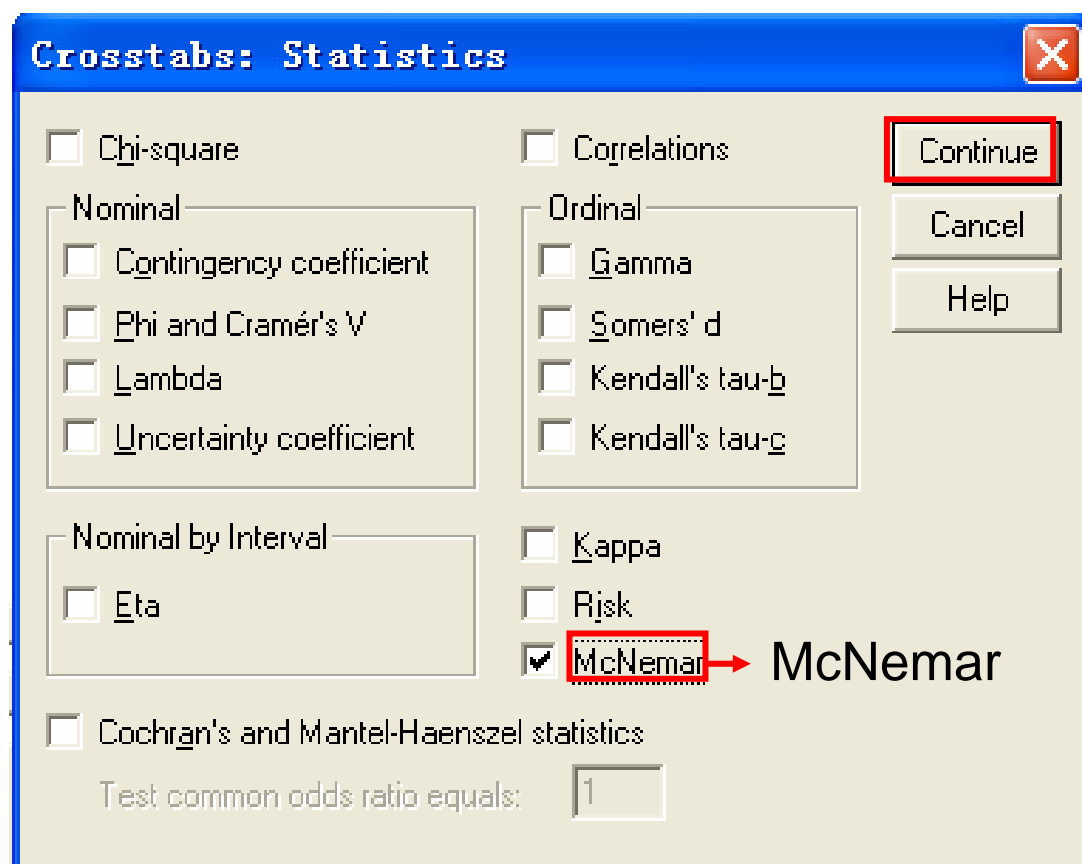
## (二) 配对设计的两样本率比较的假设检验

### 例4 方法一 (SPSS菜单: Crosstabs)



## (二) 配对设计的两样本率比较的假设检验

### 例4 方法一 (SPSS菜单: Crosstabs)



## (二) 配对设计的两样本率比较的假设检验

### 例4 方法一 (SPSS菜单: Crosstabs)

酶标法 \* 反向血凝法 Crosstabulation

Count		反向血凝法		Total
		阳性	阴性	
酶标法	阳性	30	18	48
	阴性	2	150	152
Total		32	168	200

Chi-Square Tests

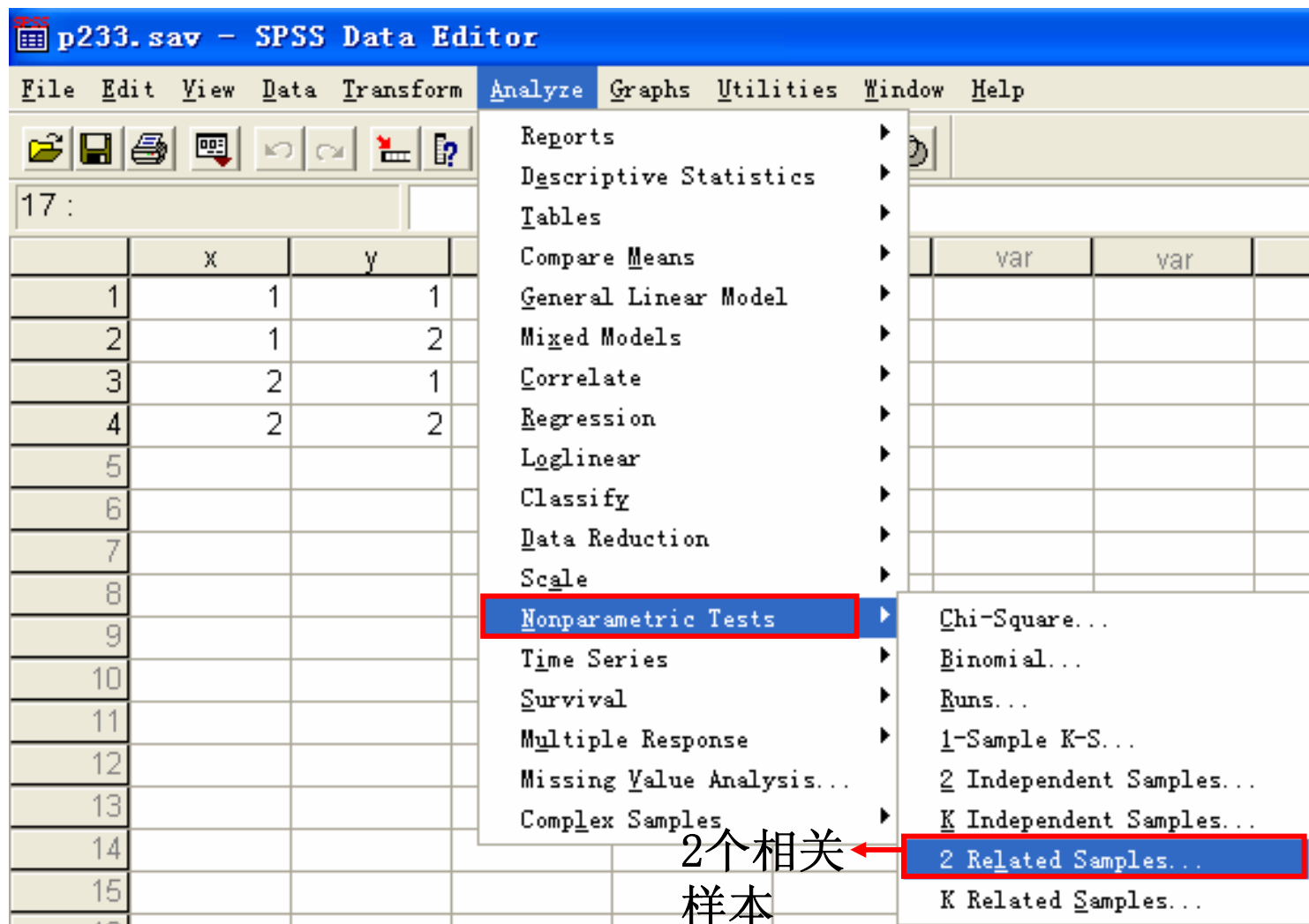
	Value	Exact Sig. (2-sided)
McNemar Test		.000 <sup>a</sup>
N of Valid Cases	200	

精确概率法

a. Binomial distribution used.

## （二）配对设计的两样本率比较的假设检验

例4 方法二 （**SPSS菜单：Nonparametric Tests**）**推荐**



The screenshot shows the SPSS Data Editor window with the file name 'p233.sav'. The 'Analyze' menu is open, and the 'Nonparametric Tests' option is selected. A sub-menu is displayed, showing '2 Related Samples...' as the chosen option. A red box highlights the '2 Related Samples...' option, and a red arrow points to it from the text '2个相关样本' (2 related samples).

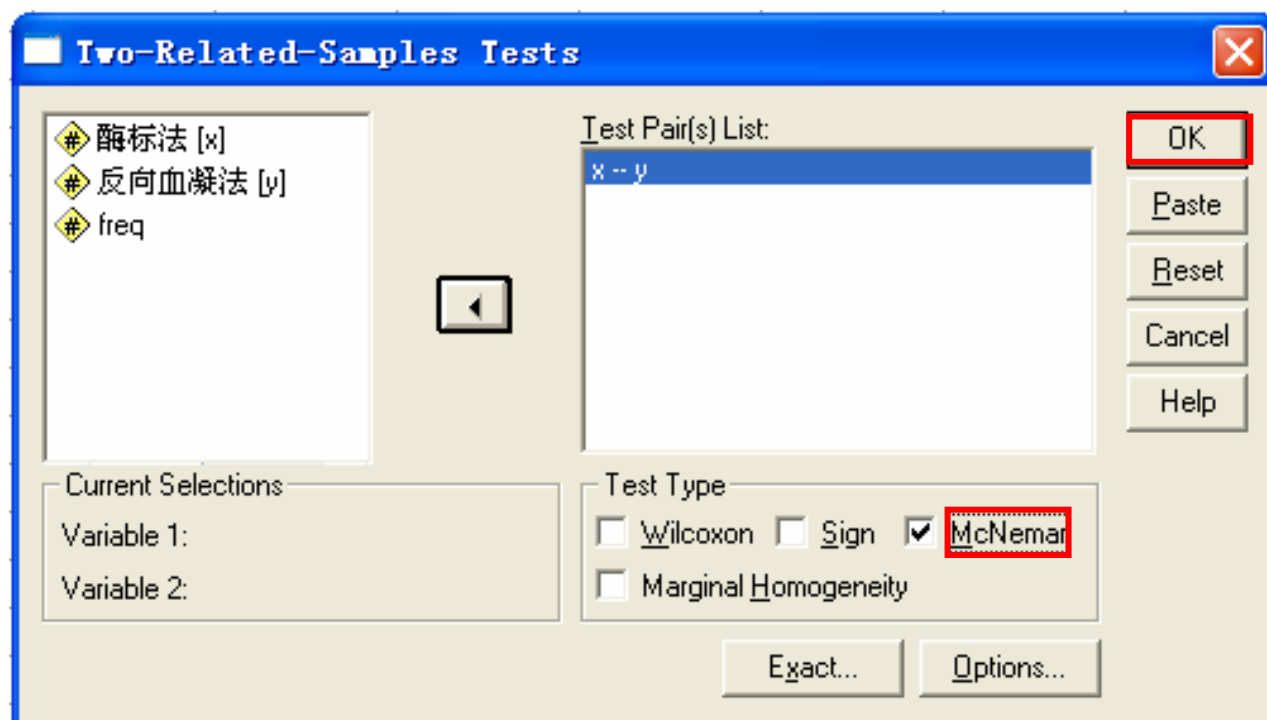
SPSS Data Editor window showing the 'Analyze' menu path: **Analyze > Nonparametric Tests > 2 Related Samples...**

The data table contains the following values:

	x	y
1	1	1
2	1	2
3	2	1
4	2	2
5		
6		
7		
8		
9		
10		
11		
12		
13		
14		
15		

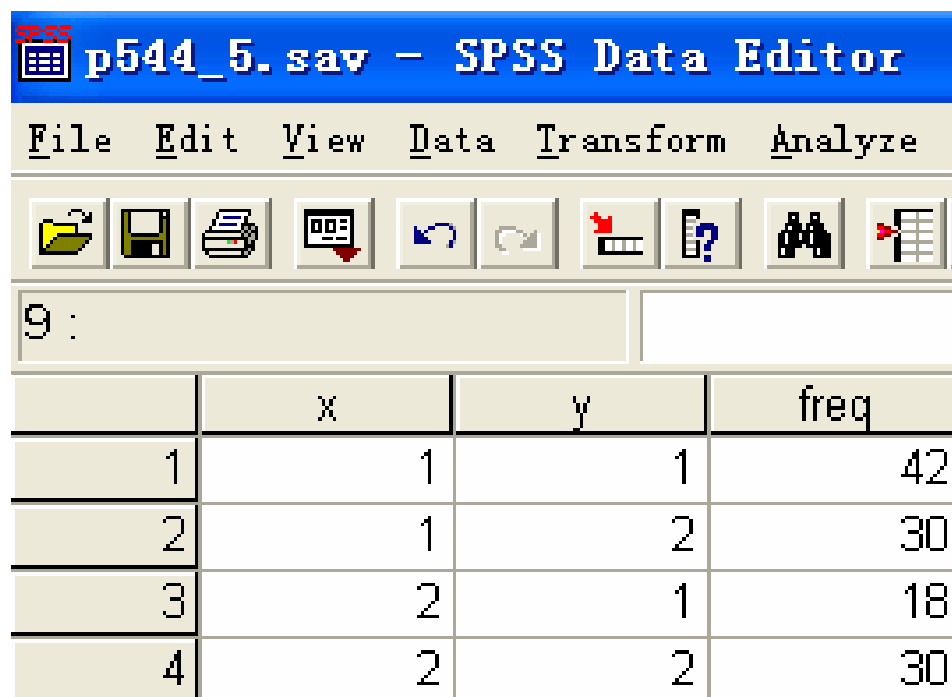
## （二）配对设计的两样本率比较的假设检验

例4 方法二 （**SPSS菜单：Nonparametric Tests**）**推荐**



## (二) 配对设计的两样本率比较的假设检验

**例5** 见第544页计算分析题5。**X**:甲法, 1=检出, 2=未检出; **Y**:乙法, 1=检出, 2=未检出; **freq**: 频数。  
(**SPSS**软件操作步骤与**例4**中方法二相同)



	X	Y	freq
1	1	1	42
2	1	2	30
3	2	1	18
4	2	2	30



## (二) 配对设计的两样本率比较的假设检验

### 例5

#### McNemar Test

#### Crosstabs

甲 & 乙

甲	乙	
	1	2
1	42	30
2	18	30

Test Statistics<sup>b</sup>

	甲 & 乙
N	120
Chi-Square <sup>a</sup>	2.521
Asymp. Sig.	.112

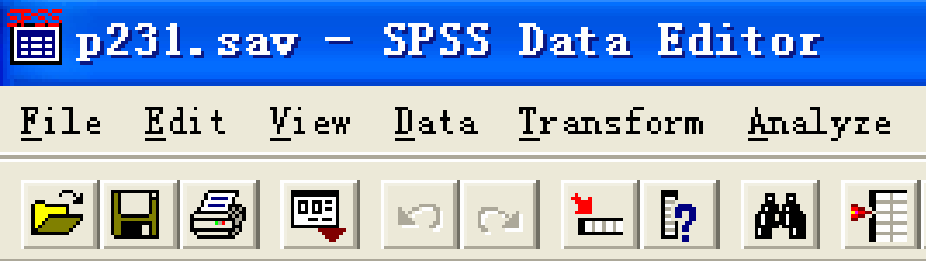
a. Continuity Corrected

b. McNemar Test

→  $\chi^2$ 检验及P值

### (三) 完全随机设计的多个样本率比较的假设检验

**例6** 见第231页例9—6。**group**:组别, 1=高氟区, 2=干预区, 3=低氟区;**effect**:1=患龋, 2=未患龋;**freq**:频数。(SPSS软件操作步骤与**例1**相同)



9 :

	group	effect	freq
1	1	1	45
2	1	2	26
3	2	1	31
4	2	2	15
5	3	1	43
6	3	2	5

### (三) 完全随机设计的多个样本率比较的假设检验

#### 例6

地区 \* 患龋情况 Crosstabulation

			患龋情况		Total
			患龋	未患龋	
地区	高氟区	Count	45	26	71
		% within 地区	63.4%	36.6%	100.0%
	干预区	Count	31	15	46
		% within 地区	67.4%	32.6%	100.0%
	低氟区	Count	43	5	48
		% within 地区	89.6%	10.4%	100.0%
Total	Count		119	46	165
	% within 地区		72.1%	27.9%	100.0%

患龋率

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	10.489 <sup>a</sup>	2	.005
Likelihood Ratio	11.852	2	.003
Linear-by-Linear Association	9.085	1	.003
N of Valid Cases	165		

Pearson  
卡方值

P值

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 12.82.