

## How to run a chi-square 2 way test in SPSS

In this example, we want to test the claim that there is an association between the restrictions on movie watching (no R rated, occasional R rated, and anything) and smoking among children on the east coast. These are the hypotheses:

$H_a$ : There is an association between the restrictions on movie watching (no R rated, occasional R rated, and anything) and smoking among children\* on the east coast.

$H_0$ : There is no association between the restrictions on movie watching (no R rated, occasional R rated, and anything) and smoking among children\* on the east coast.

which (because the response only takes on two values) can be written mathematically as

$H_a$ : At least one  $\pi$  is different than the others

$H_0$ :  $\pi_1 = \pi_2 = \pi_3$

where



$\pi_1$  = the proportion of smokers\* among children on the east coast who are not allowed to watch R rated films

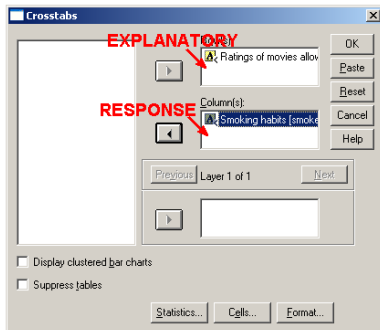
$\pi_2$  = the proportion of smokers\* among children on the east coast who are allowed to watch an occasional R rated film

$\pi_3$  = the proportion of smokers\* among children on the east coast who are allowed to watch anything

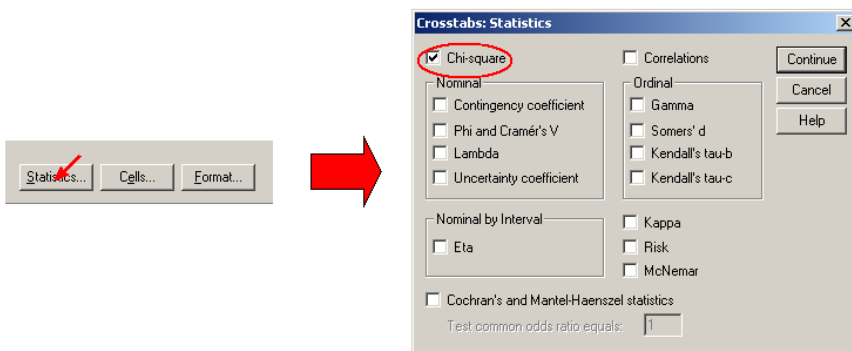
\* = children who could potentially get into the sample given the sampling method. The data set is called "moviesmoke" and can be found on the class website.

1) In SPSS: Go to Analyze → Descriptive Statistics → Crosstabs . . .

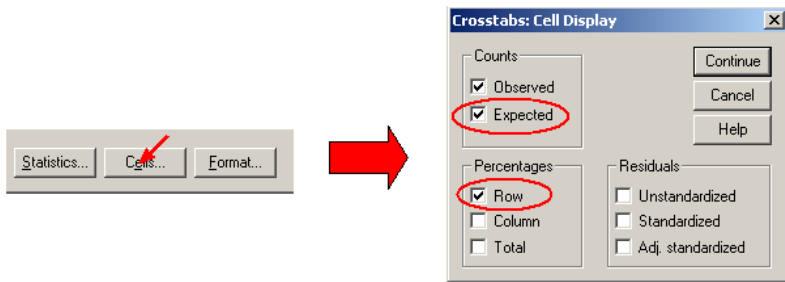
2) Click the explanatory variable name, then click the upper right arrow button, , to move it to the Row(s) list. In this case, the explanatory variable is "ratings of movies allowed by parents". Click the response variable name, then click the lower right arrow button, , to move it to the Column(s) list. In this case, the response is "Smoking Habits"



3) Click the "Statistics" button, and select "Chi-squared", then click "Continue"



4) Click the “Cells” button, and select “Row” and “Expected”, then click “Continue”



5) When you get back, click “OK”. Your output should look like this:

## Crosstabs

### Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Ratings of movies allowed by parents * Smoking habits	8400	100.0%	0	.0%	8400	100.0%

### Ratings of movies allowed by parents \* Smoking habits Crosstabulation

			Smoking habits		Total
			don't smoke	smoke	
Ratings of movies allowed by parents	anything	Count	4125	875	5000
		Expected Count	4206.0	794.0	5000.0
		% within Ratings of movies allowed by parents	82.5%	17.5%	100.0%
	no R rated	Count	698	52	750
		Expected Count	630.9	119.1	750.0
		% within Ratings of movies allowed by parents	93.1%	6.9%	100.0%
	occasional R rated	Count	2243	407	2650
		Expected Count	2229.2	420.8	2650.0
		% within Ratings of movies allowed by parents	84.6%	15.4%	100.0%
Total	Count	7066	1334	8400	
	Expected Count	7066.0	1334.0	8400.0	
	% within Ratings of movies allowed by parents	84.1%	15.9%	100.0%	

### Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	55.300 <sup>a</sup>	2	.000
Likelihood Ratio	65.019	2	.000
N of Valid Cases	8400		

This is the p-value

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 119.11.