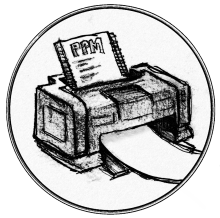
Chi-Squared Tests

These are simple tests that compare categorical variables… to other categorical variables. This is different from a t-test. There are no continuous variables here. Your independents and dependents are both categorical. For example, who is more likely to own a pet snake: men or women? Let's say you ask 200 men and 200 women. Within your 400-subject sample, the frequency of snake ownership among men is 2.5% and the ownership among women is 0.9%. That's a pretty big difference in your *sample*, but is it big enough to indicate a difference in the *population*? The chi-squared test is an inferential statistic. It is used to make inferences about the larger population.



Examples of Questions for Chi-Square Tests:

Do football players and baseball players differ in divorce rate?

Are cigarette smokers more likely to get lung cancer?

What is the difference in STD frequency between professional and amateur hockey players?

Do republicans and democrats differ in astrological sign?

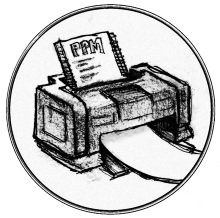


Table 1: *Subject baseline characteristics*

	Total	Men	Women	Sig.
N	45	23	22	
Age (years)	20.2 ± 0.7	21.0 ± 0.5	22.3 ± 0.9	p = 0.412
BMI (kg/m^2)	26.5 ± 2.8	27.4 ± 0.8	26.0 ± 0.6	p = 0.518
GPA	2.9 ± 0.7	2.6 ± 1.0	3.2 ± 0.5	**p = 0.048**
Nightly sleep (hours)	7.5 ± 1.4	7.1 ± 1.2	7.6 ± 1.9	p = 0.525
Employed (%)	22.5%			
Weekly work (hours)	4.0 ± 5.2	8.5 ± 4.1	2.3 ± 6.6	p = 0.126
Academic Scholarship (%)	12.9%			
Athletic Scholarship (%)	11.5%			

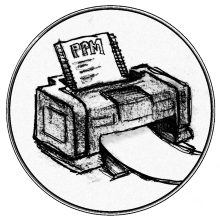
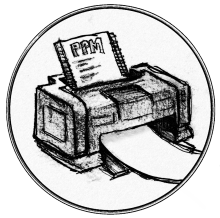


Table 1: *Subject baseline characteristics*

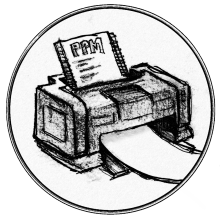
	Total	Men	Women	Sig.
N	45	23	22	
Age (years)	20.2 ± 0.7	21.0 ± 0.5	22.3 ± 0.9	p = 0.412
BMI (kg/m²)	26.5 ± 2.8	27.4 ± 0.8	26.0 ± 0.6	p = 0.518
GPA	2.9 ± 0.7	2.6 ± 1.0	3.2 ± 0.5	**p = 0.048**
Nightly sleep (hours)	7.5 ± 1.4	7.1 ± 1.2	7.6 ± 1.9	p = 0.525
Employed (%)	22.5%			
Weekly work (hours)	4.0 ± 5.2	[illegible] ± 4.1	2.3 ± 6.6	p = 0.126
Academic Scholarship (%)	12.9%			
Athletic Scholarship (%)	11.5%			

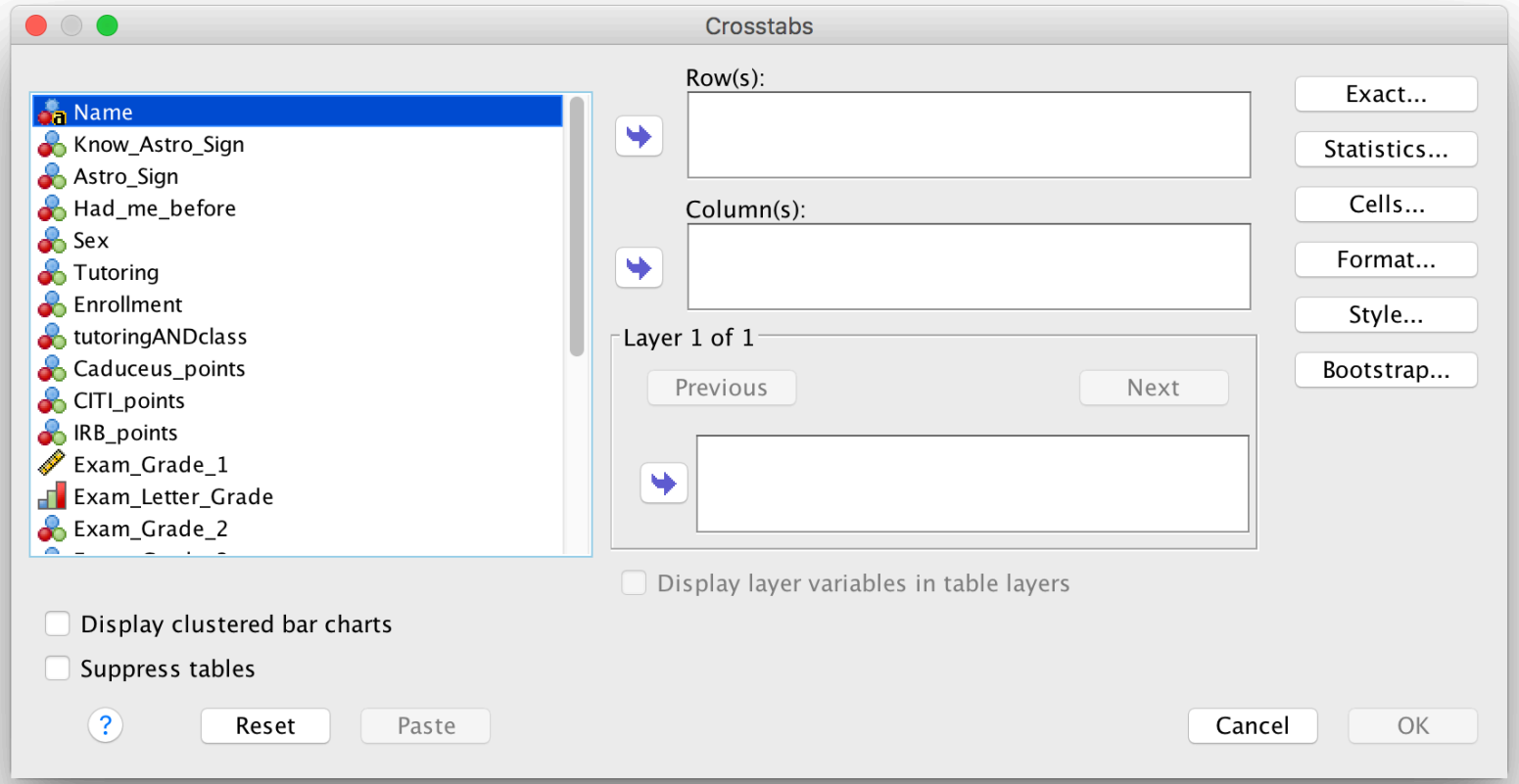
SPSS Statistics File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help

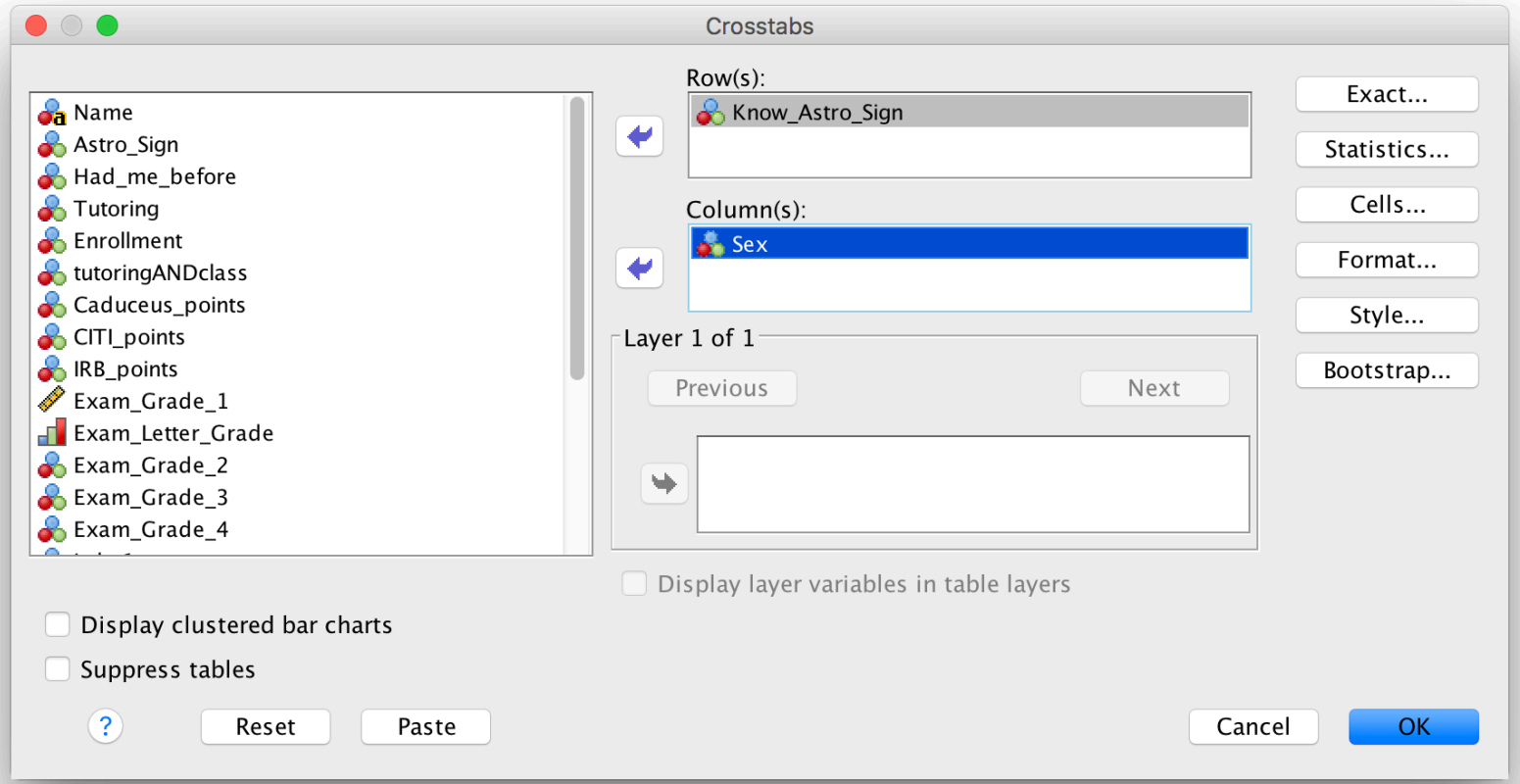


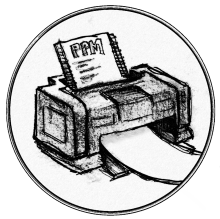
SPSS Statistics File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help

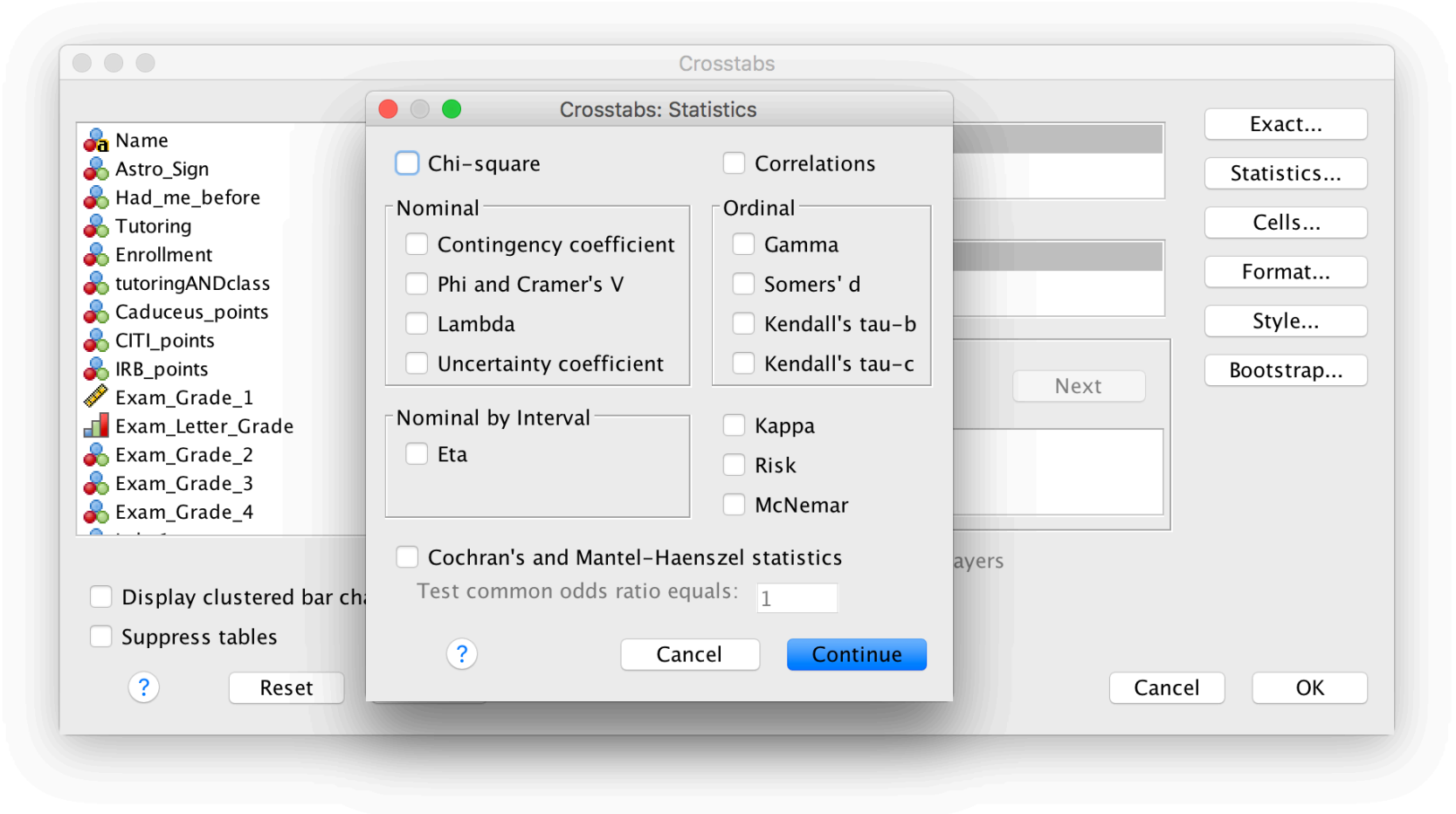
- Reports ▶
- Descriptive Statistics ▶
  - Frequencies...
  - Descriptives...
  - Explore...
  - Crosstabs...
  - TURF Analysis
  - Ratio...
  - P-P Plots...
  - Q-Q Plots...
- Compare Means ▶
- General Linear Model ▶
- Generalized Linear Models ▶
- Mixed Models ▶
- Correlate ▶
- Regression ▶
- Loglinear ▶
- Classify ▶
- Dimension Reduction ▶
- Scale ▶
- Nonparametric Tests ▶
- Forecasting ▶
- Survival ▶
- Multiple Response ▶
- Simulation...
- Quality Control ▶
- ROC Curve...
- Spatial and Temporal Modeling... ▶

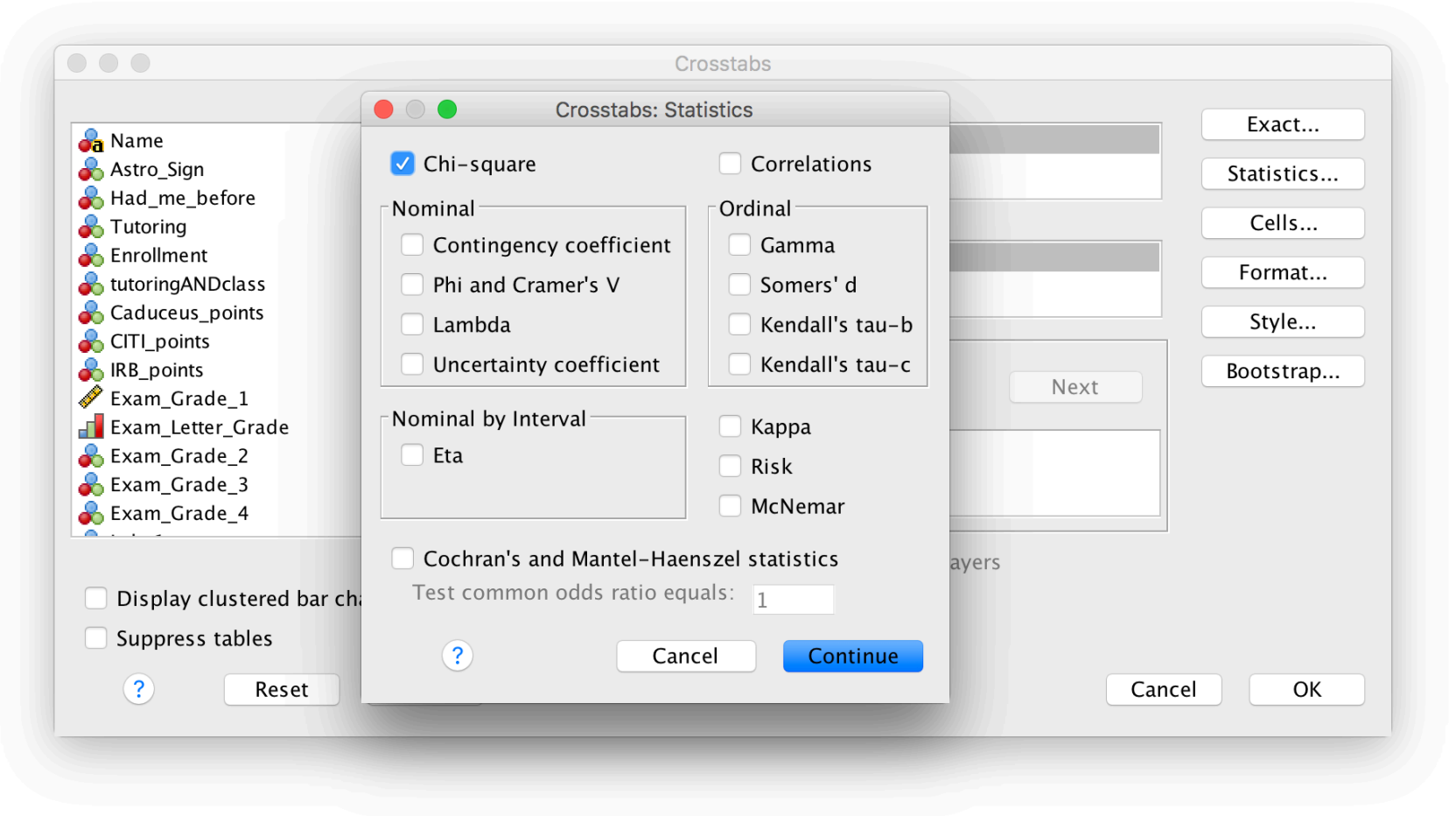


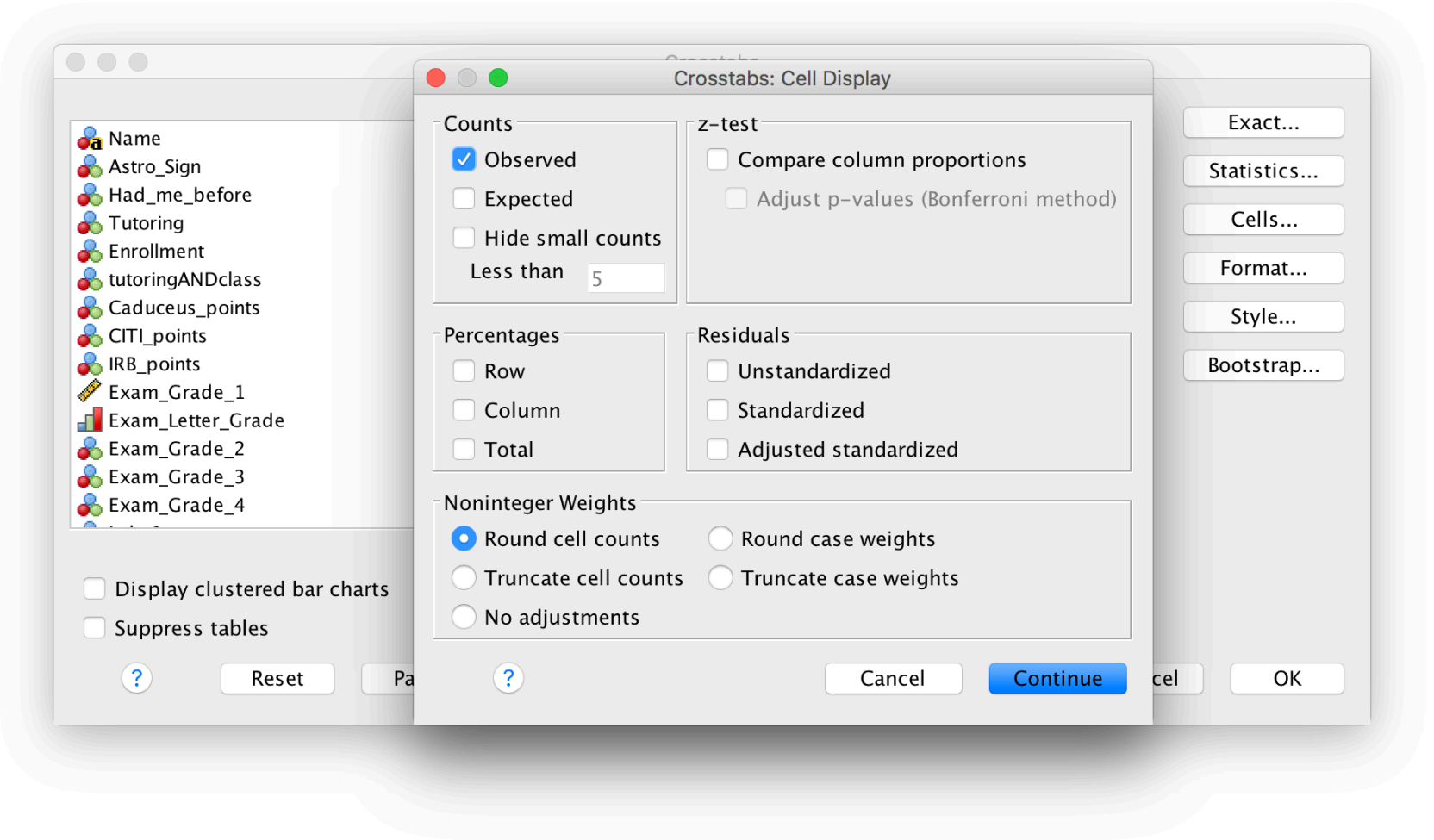


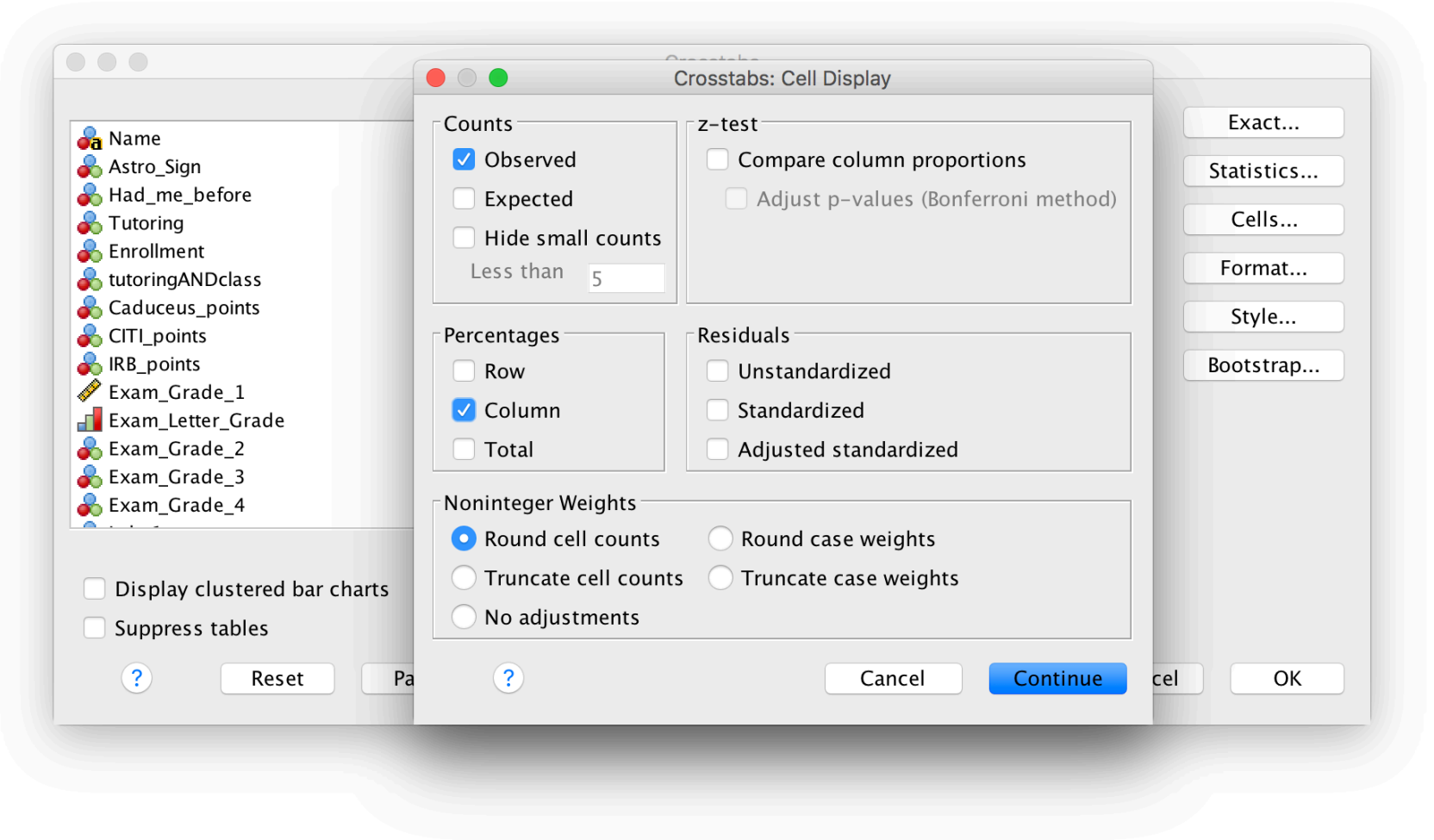












Crosstabs

Know_Astro_Sign * Sex Crosstabulation

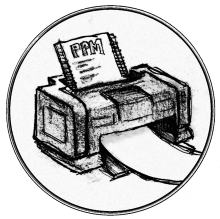
			Sex		
			Male	Female	Total
Know_Astro_Sign	Does not know	Count	3	5	8
		% within Sex	33.3%	20.8%	24.2%
	Knows	Count	6	19	25
		% within Sex	66.7%	79.2%	75.8%
Total		Count	9	24	33
		% within Sex	100.0%	100.0%	100.0%

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	.557[a]	1	.456		
Continuity Correction[b]	.084	1	.772		
Likelihood Ratio	.534	1	.465		
Fisher's Exact Test				.651	.374
Linear-by-Linear Association	.540	1	.462		
N of Valid Cases	33				

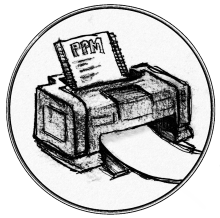
Table 1: *Subject baseline characteristics*

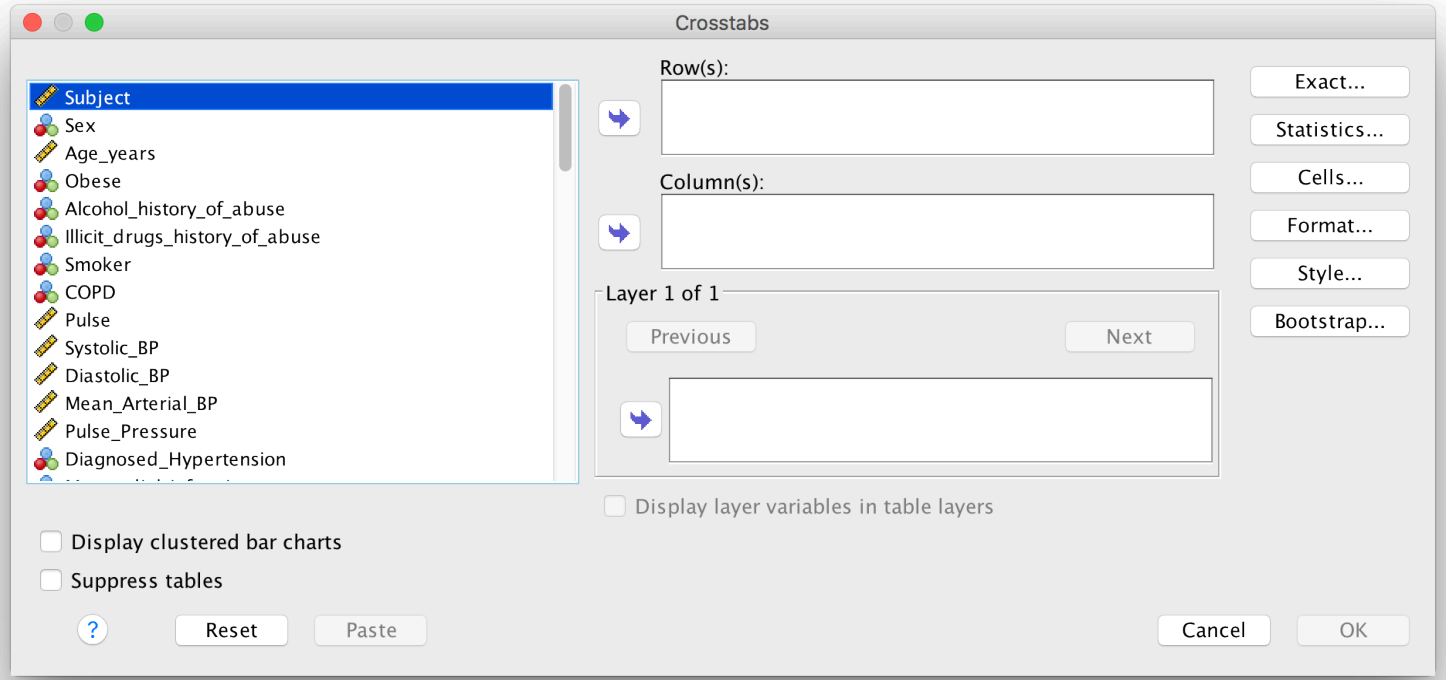
	Total	Men	Women	Sig.
N	45	23	22	
Age (years)	20.2 ± 0.7	21.0 ± 0.5	22.3 ± 0.9	p = 0.412
BMI (kg/m^2)	26.5 ± 2.8	27.4 ± 0.8	26.0 ± 0.6	p = 0.518
GPA	2.9 ± 0.7	2.6 ± 1.0	3.2 ± 0.5	**p = 0.048**
Nightly sleep (hours)	7.5 ± 1.4	7.1 ± 1.2	7.6 ± 1.9	p = 0.525
Employed (%)	22.5%	24.0%	19.7%	p = 0.099
Weekly work (hours)	4.0 ± 5.2	8.5 ± 4.1	2.3 ± 6.6	p = 0.126
Academic Scholarship (%)	12.9%	13.2%	12.6%	p = 0.884
Athletic Scholarship (%)	11.5%	11.9%	11.3%	p = 0.768

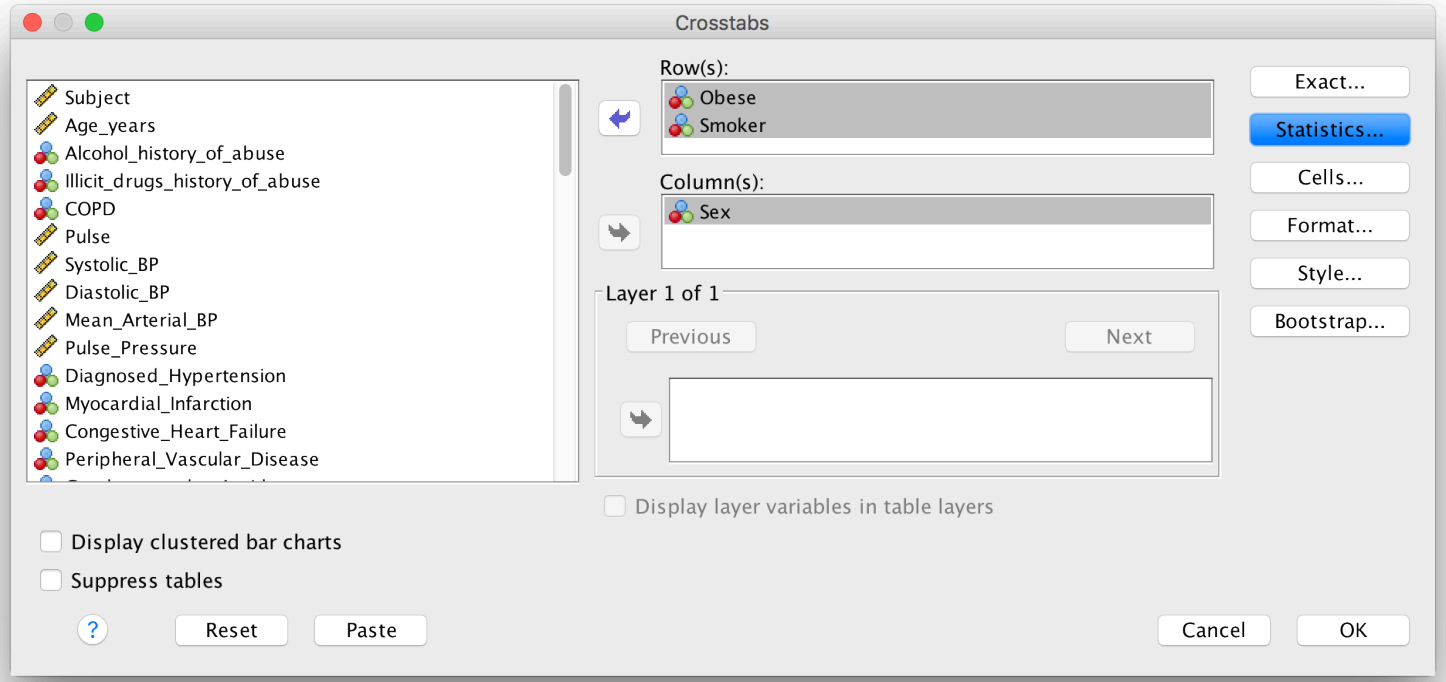


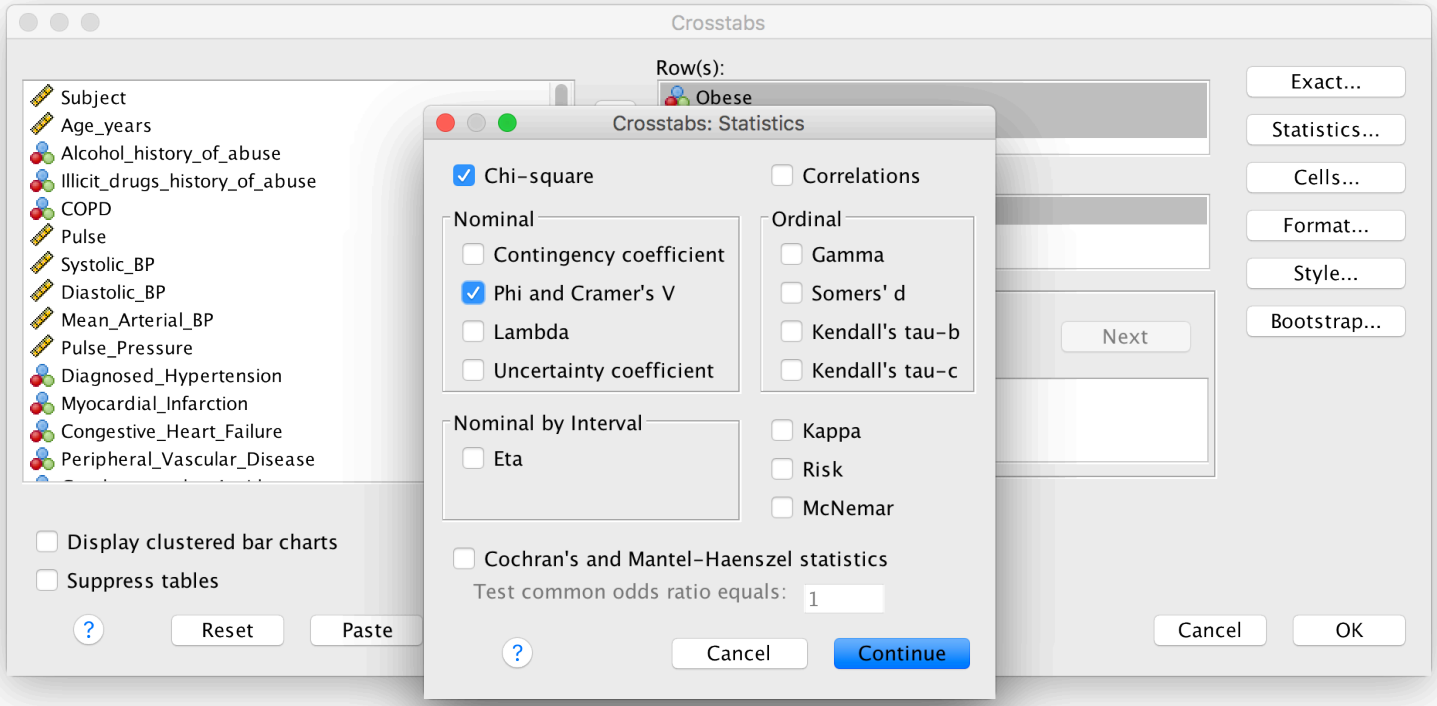
Chi-Squared Value. Two binary measurements. What's the distribution between these? Is it the same percentage in each group (expected) or is there a difference between groups (observed)?

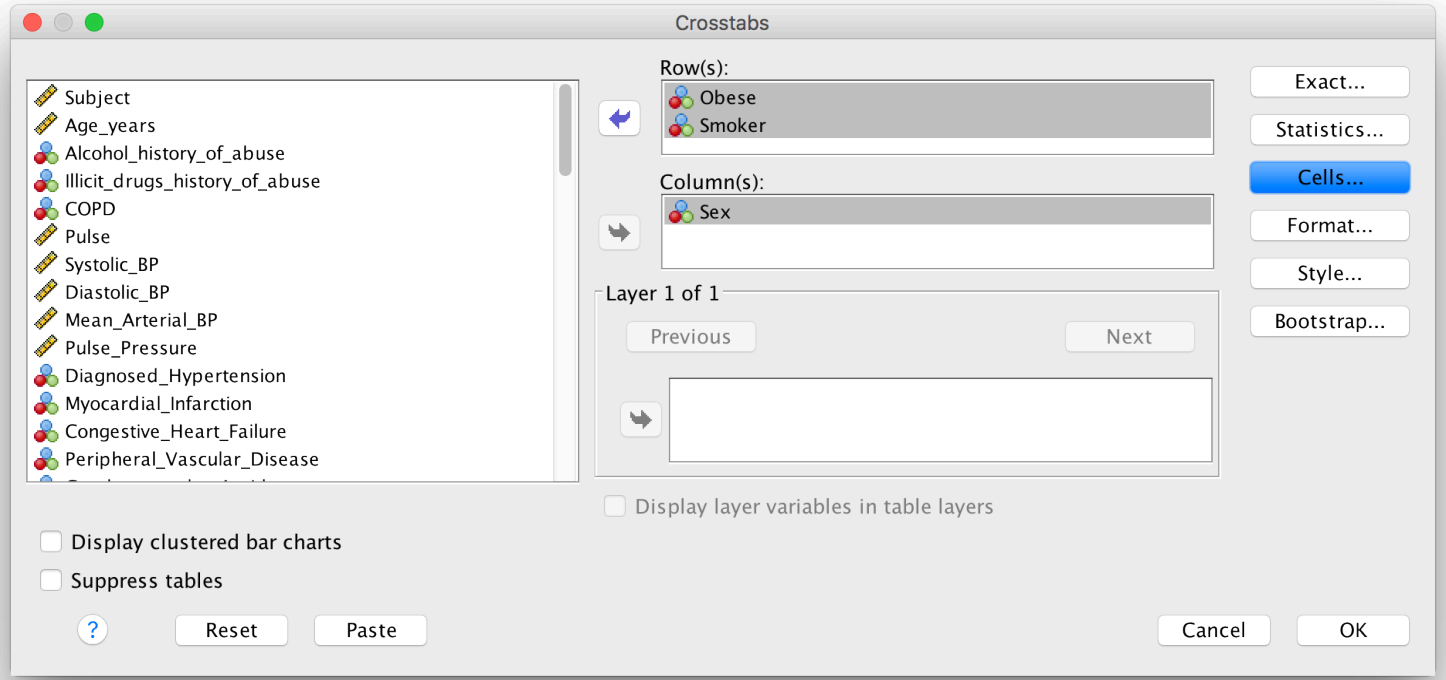
$$\text{Sum of } \frac{(\text{Observed Values} - \text{Expected Values})^2}{\text{Expected Values}}$$

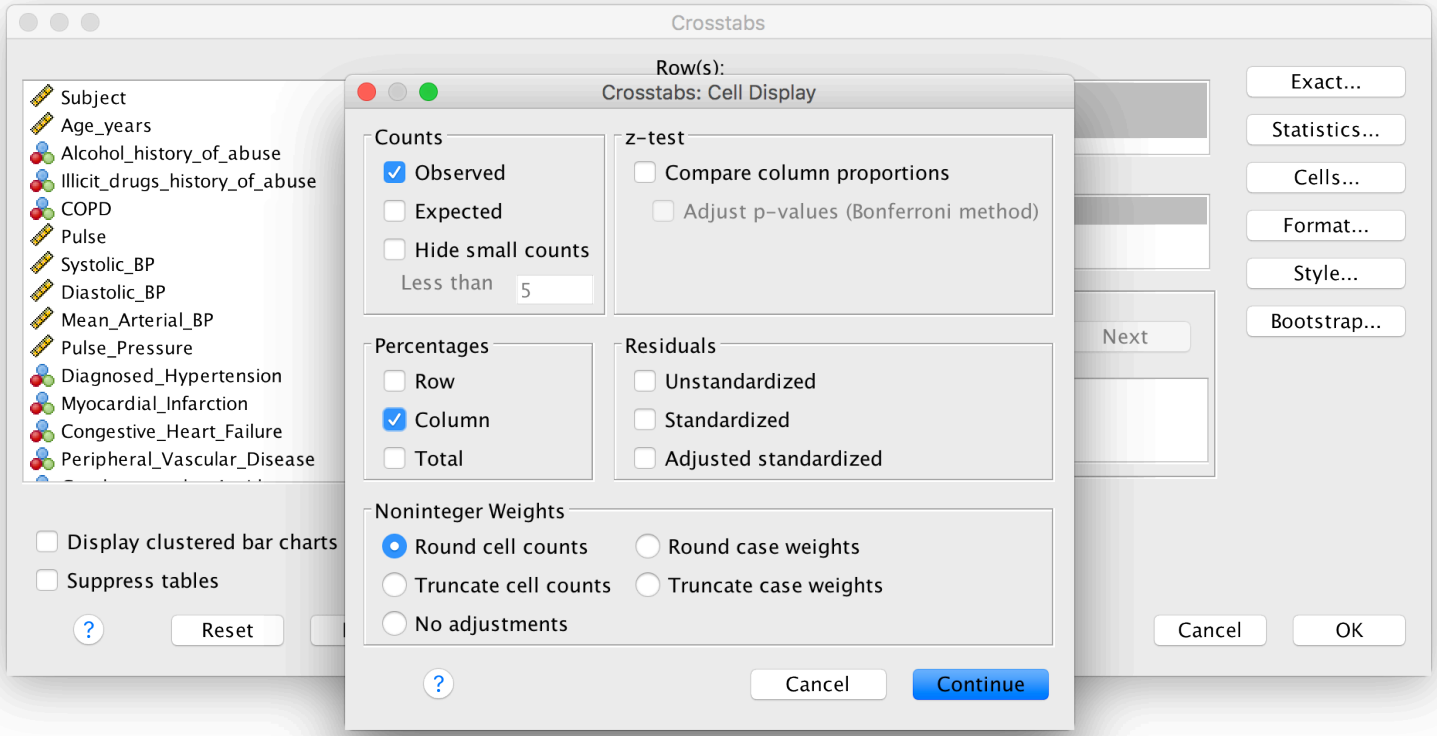


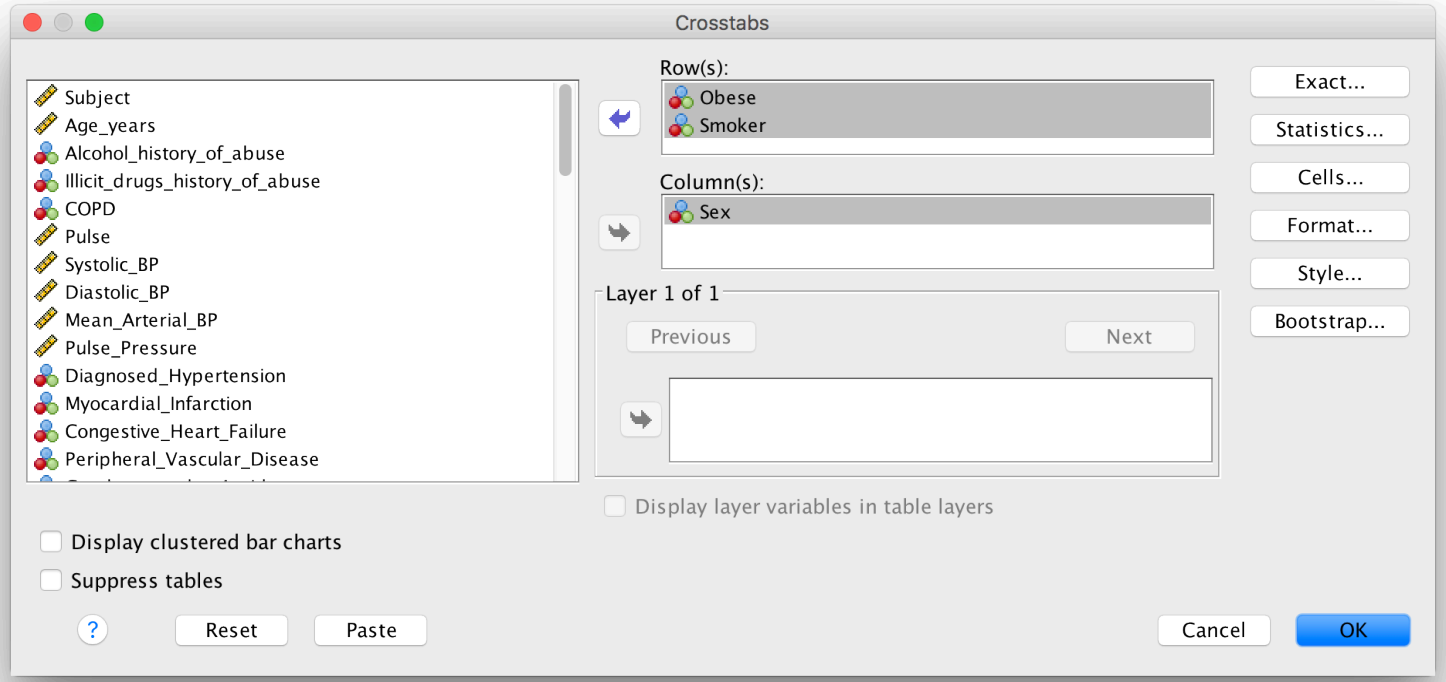












Obese * Sex

Crosstab

			Sex		
			Male	Female	Total
Obese	0	Count	1080	849	1929
		% within Sex	83.7%	83.6%	83.7%
	1	Count	211	166	377
		% within Sex	16.3%	16.4%	16.3%
Total		Count	1291	1015	2306
		% within Sex	100.0%	100.0%	100.0%

Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	.000	.994
	Cramer's V	.000	.994
N of Valid Cases		2306	

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	.000[a]	1	.994		
Continuity Correction[b]	.000	1	1.000		
Likelihood Ratio	.000	1	.994		
Fisher's Exact Test				1.000	.519
Linear-by-Linear Association	.000	1	.994		
N of Valid Cases	2306				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 165.94.

b. Computed only for a 2x2 table

Smoker * Sex

Crosstab

			Sex		
			Male	Female	Total
Smoker	0	Count	881	831	1712
		% within Sex	68.2%	81.9%	74.2%
	1	Count	410	184	594
		% within Sex	31.8%	18.1%	25.8%
Total		Count	1291	1015	2306
		% within Sex	100.0%	100.0%	100.0%

Symmetric Measures

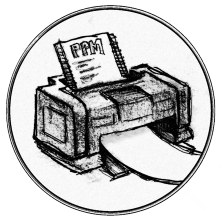
		Value	Approximate Significance
Nominal by Nominal	Phi	-.155	.000
	Cramer's V	.155	.000
N of Valid Cases		2306	

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	55.204[a]	1	.000		
Continuity Correction[b]	54.493	1	.000		
Likelihood Ratio	56.539	1	.000		
Fisher's Exact Test				.000	.000
Linear-by-Linear Association	55.180	1	.000		
N of Valid Cases	2306				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 261.45.

b. Computed only for a 2x2 table



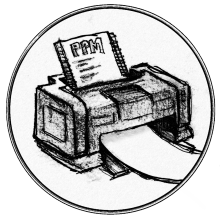
When examining the association between two dichotomous variables, the correlation is called Phi.

When examining the association between two categorical variables that have more than two categories, Cramer's V is the appropriate statistic.

Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	-.155	.000
	Cramer's V	.155	.000
N of Valid Cases		2306	

If you want the correlation of one continuous variable to another continuous variable, use the Pearson correlation coefficient (bivariate correlation table); if you want the correlation of a continuous variable and a dichotomous variable, that's a point-biserial correlation (also bivariate table).



Remember: Chi-squared tests aren't controlling for anything. What is the difference in lung cancer frequency among people who do and do not keep matches in their pockets? Let's say the chi-square is significant: there's a five-fold increase in the frequency of lung cancer in the matches-in-pocket group (p=0.002). Does that mean matches *cause* lung cancer (at a 99.8% confidence level)? Or is causality a different question altogether?

Table 1: *Subject baseline characteristics*

	Total	Men	Women	Sig.
N	45	23	22	
Age (years)	20.2 ± 0.7	21.0 ± 0.5	22.3 ± 0.9	p = 0.412
BMI (kg/m²)	26.5 ± 2.8	27.4 ± 0.8	26.0 ± 0.6	p = 0.518
GPA	2.9 ± 0.7	2.6 ± 1.0	3.2 ± 0.5	**p = 0.048**
Nightly sleep (hours)	7.5 ± 1.4	7.1 ± 1.2	7.6 ± 1.9	p = 0.525
Employed (%)	22.5%	24.0%	19.7%	p = 0.099
Weekly work (hours)	4.0 ± 5.2	8.5 ± 4.1	2.3 ± 6.6	p = 0.126
Academic Scholarship (%)	12.9%	13.2%	12.6%	p = 0.884
Athletic Scholarship (%)	11.5%	11.9%	11.3%	p = 0.768