Chi-Square Test for Association/Independence

In a study of heart disease in male federal employees, researchers classified 356 volunteer subjects according to their socioeconomic status (SES) and their smoking habits. There were three categories of SES: high, middle, and low. Individuals were asked whether they were current smokers, former smokers, or had never smoked, producing three categories for smoking habits as well. Here is the two-way table that summarizes the data.

Observed counts for smoking and SES

	SES			
Smoking	High	Middle	Low	Total
Current	51	22	43	116
Former	92	21	28	141
Never	68	9	22	99
Total	211	52	93	356

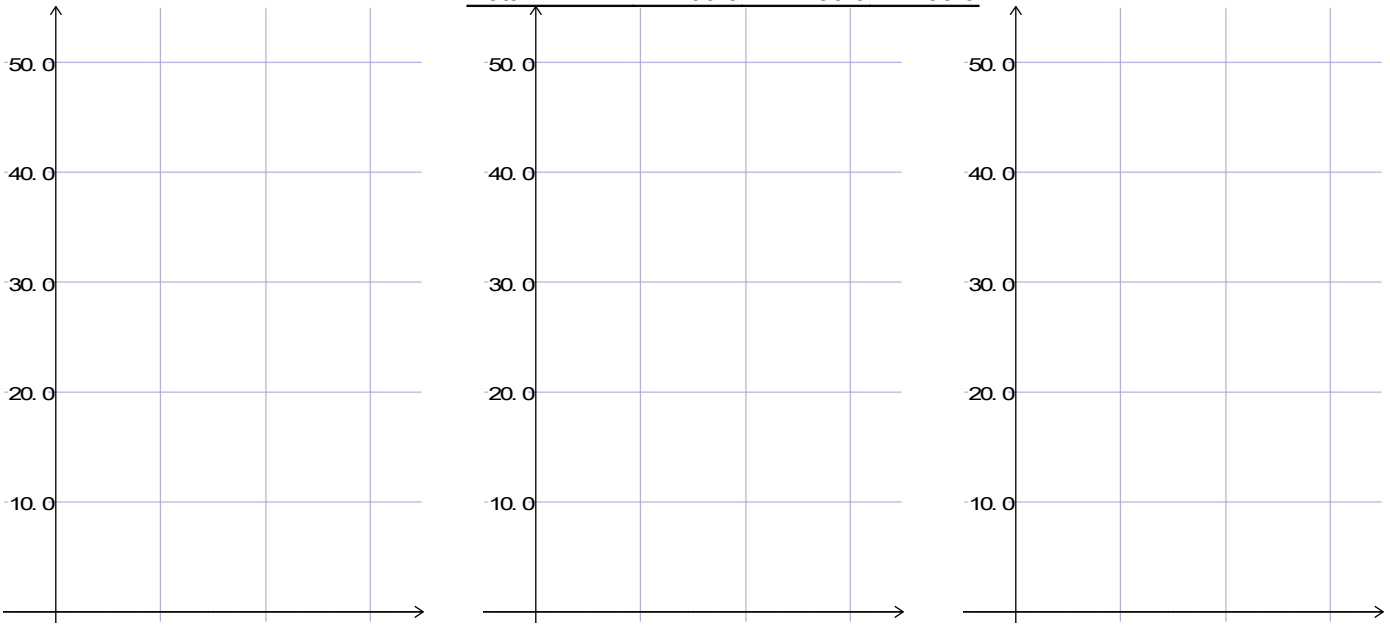
1. How is the construction of this two-way table different from the two-way tables we made for the chi-square tests of homogeneity of populations?

2. The researchers wanted to know if an association exists between the two categorical variables. Which of the two variables would we consider "explanatory" in our analysis, and which is the "response" variable?

3. Construct bar-graphs that examine the explanatory-response relationship between SES and smoking. To do this you will need to compute the marginal distributions for smoking habits in each SES. Begin by completing the table below.

Marginal Distribtutions of Smoking Habits

	SES		
Smoking	High	Middle	Low
Current			
Former			
Never			
Total	100.0	100.0	100.0



4. What do you observe about the relationship of SES and smoking from your graphs?

5. We wish to test if the differences above are "statistically significant". What are the hypotheses we are testing?

6. We can use a chi-square test, but we need to know the expected cell counts for the two-way table.

Expected counts for smoking and SES

	SES		
Smoking	High	Middle	Low
Current			
Former			
Never			

7. Check your conditions for inference. Is it appropriate in this case to use the chi-square test?

8. Perform the test and state your conclusions.