The topic for this week is describing the relationship between two *categorical* variables, such as whether someone smokes (yes or no) and whether they have ever been divorced (yes or no). (See page 195 for this example.) As in earlier chapters, usually one of the variables will be designated as the explanatory variable, and the other as the response variable. For instance, we would designate smoking or not as the explanatory variable because we would be interested in whether smokers are more likely to get divorced. It is not always clear which variable is which. When it is clear, a two-way table of counts, called a *contingency table,* is constructed by putting the categories of the explanatory variable as the rows and the categories of the response variable as the columns. (See pages 194-195.) A contingency table can be constructed in any case, but it won't matter which variable is used for the rows and which is used for the columns.

We are interested in describing the relationship between the two variables using things like *conditional percentages* and *relative risk*. These terms are described in Sections 6.1 and 6.2 of the book. We are also interested in determining if a relationship observed in the *sample* on which we have collected data represents a real relationship in the *population* from which the sample was taken. We do this by carrying out a test to determine if the relationship is *statistically significant.* This topic is covered in Section 6.4 and is just the first of many instances of "hypothesis testing" that will be covered in the class.

A very important aspect of the week's material is learning how you can be misled by results and methods like the ones covered in this week's material. That's the focus of Section 6.3 and parts of Sections 6.2 and 6.4, as well as the "Warnings 2" page of Unit A9 and "Warnings 1" page of Unit C10 in CyberStats, listed in the table below as part of your focus for the week.

The sample quiz, homework problems assigned and CyberStats focus should give you a clear idea of what to emphasize in your work. One specific note is that I won't ask you to compute a chi-square statistic, but I will want you to be able to read and interpret computer output (shown in the book) and understand how to interpret the results. You should also understand and know how to compute "expected counts" for a chi-square test.

Reading and Study Assignment for this week:

Book Chapter or CyberStats Unit	Focus on:
Chapter 6: Relationships between Categorical Variables	All, but especially topics covered by assigned exercises, and the material above in these notes.
Unit A9: Describing Categorical Data	Warnings 2, Example 1
Unit C10: Two-way Tables for Count Data	Warnings 1, Example 1, Self Assess Quiz

Interactivities to play with:
Unit C10:
Work through the answers for Example 2. If you put anything in the boxes and press "submit" you will see the correct answers.

Exercises to hand in: Chapter 6: #3, 4, 13b, 14b, 21, 27, 34, 40, 41, 42ab, 63