Business Analytics: Data Analysis and Decision Making, 6e
CHAPTER 3: Finding Relationships among Variables

Answers to Conceptual Questions

Note to Instructors: Student answers will vary. The responses here are intended to provide general guidance in terms of concepts that could be discussed.

C.1. When analyzing a possible relationship between two categorical variables, the joint counts (i.e., the counts of joint categories of the two variables) are all you have to go on. Specifically, it is useful to show these joint counts as percentages of row totals or percentages of column totals. If they are shown as percentages of row totals, for example, where the percentages in each row add to 100%, then the rows should be nearly the same if there is no relationship. Otherwise, you can study the differences between the rows to understand the type of relationship.

C.2. As in the previous problem, the two rows will be nearly the same, such as the following, if the variables aren't related:

		Variable 1	
		No	Yes
Variable	No	65%	35%
2	Yes	65%	35%

But they could be as follows, indicating that those in the Yes category for Variable 2 are more likely to be in the Yes category for Variable 1 than those in the No category for Variable 2:

		Variable 1	
		No	Yes
Variable	No	65%	35%
2	Yes	30%	70%

Or they could be as follows, which is the opposite.

		Variable 1	
		No	Yes
Variable	No	60%	40%
2	Yes	35%	65%

And with only two categories per variable (Yes and No), these are the only two possible forms of dependence, corresponding (roughly speaking) to positive correlation and negative correlation.

C.3. You can imagine two columns, one for monthly sales and one for monthly advertising. If you found the correlation between the two columns, it would indicate the effect of advertising on sales *in that month*. If you think the effect is delayed, you should create one or more lagged columns of advertising (easy with StatTools) that "pushes down" the advertising column by one or more rows. Then you can check the correlation between the sales column and any of the lagged advertising columns.

C.4. You can proceed exactly as in Example 3.5 (frozen lasagna triers). Put any demographic variable in the Rows area, put the "Have Bought" Yes/No variable in the Columns area, and summarize Have Bought by count in the Values area, displayed as percentages of row totals.

C.5. Because the correlations mentioned are across various achievement measures, these low correlations indicate that the measures themselves don't appear to be measuring the same thing. For example, if one measure is of reading ability and another is of mathematical ability, the low correlation would discredit the possibility that the best readers tend to be the best at math. Such a finding might or might not be surprising.

C.6. The pivot tables discussed in this chapter don't allow you to drill down like this, at least not easily. However, the OLAP technology discussed in Chapter 17 makes it fairly easy. In fact, one of the main reasons for the development of this OLAP technology is to allow such drilling down through hierarchies in pivot tables.

C.7. This is another case of Example 3.5 (or #4 above), where you want to see which variables drive Win/Lose outcomes. If one of these drivers is continuous, you would group on it in the pivot table.

C.8. Causality is a common misconception in correlation studies. For example, if the daily returns of stock X and stock Y are positively correlated, as they probably would be if X and Y are in the same industry, it is unlikely that high stock X returns are *causing* high stock Y returns. It is much more likely that the movements in the overall market (or in that industry) are causing both stocks' movements. (Do a Web search for "spurious correlation" to learn more.) And even in cases where X and Y causally related, it isn't always clear which is causing which.

C.9. This is very likely to be impossible. We are told that A and B vary inversely with one another. Suppose C also varies inversely with A, so that C decreases when A increases. But B also decreases when A increases. So it's hard to see how C could be inversely related to B, i.e., how C could increase when B decreases. Actually, it can be shown mathematically that some correlation matrices, such as the one below, can never be valid, i.e., no set of observations could ever result in them.

	A	B	C
A	1		
B	-0.8	1	
C	-0.7	-0.9	1

C.10. This is a difficult question. Probably the best answer is that it makes more sense economically to check whether *changes* in the rates of the currencies are correlated. For example, does an increase in the dollar tend to go along with a decrease in the Euro? You could instead check the correlations between the original series, but such correlations are probably not as meaningful. This would check, say, whether high values of the dollar *through history* have been associated with low values of the Euro. But if this turns out to be the case, it's probably because the *changes* are correlated.