POLITICAL SCIENCE RESEARCH METHODS

SEVENTH EDITION

Janet Buttolph Johnson
University of Delaware

H. T. Reynolds
University of Delaware

SAGE | CQPRESS

Los Angeles | London | New Delhi
Singapore | Washington DC

TABLE 13–5
Row Percentages Are Not the Same as Column Percentages

	Gender		
Party Identification Response Category	Male	Female	Total
Strong Democrat	35.0% (69)	65.0% (128)	100% (197)
Weak Democrat	46.8% (87)	53.2% (99)	100% (186)
Independent-leaning Democrat	51.7% (108)	48.3% (101)	100% (209)
Independent	50.9% (59)	49.1% (57)	100% (116)
Independent-leaning Republican	60.0% (84)	40.0% (56)	100% (140)
Weak Republican	54.4% (81)	45.6% (68)	100% (149)
Strong Republican	47.2% (93)	52.8% (104)	100% (197)

Source: 2004 National Election Study.

Note: Numbers in parentheses are frequencies.

identify as "leaning Democrat" (18.6%) . . . and so forth down through all the response categories equals 100 percent. The same is true for women: the total of column percentages sums to 100 percent. It is this arrangement of percentages that allows us to compare the relative frequencies of responses between men and women.

Suppose you asked a computer to give you percentages by row totals or obtain *row percentages*. Table 13–5 suggests what might result and the possible difficulties of interpretation. If you were not careful, you might conclude that there was a huge gender difference on "strong Democrat," 35 percent versus 65 percent. But this is not what the numbers mean. There are 197 strong Democrats in the sample (look in the last column), of which 35 percent are men and 65 percent women. It would be reasonable to say that strong Democrats tend to be composed overwhelmingly of women whereas independents are about half male and half female. Still, if in your mind one variable (e.g., party identification) depends on another variable (e.g., gender) and you want to measure the effect of the latter on the former, make sure the percentages are based on the independent variable category totals.

MEASURING STRENGTH OF RELATIONSHIPS IN TABLES

Do the data in table 13–4 support the hypothesis of a "gender gap"? As we just indicated, a careful examination of the column percentages suggests that the hypothesis has only minimal support. Why? Because a scrutiny of the partisanship distributions by gender does not show much difference. Yet it would be desirable to have a more succinct summary, one that would reveal the strength of the relationship between gender and party identification.

The strength of an association refers to how different the observed values of the dependent variable are in the categories of the independent variable. In the case of cross-classified variables, the strongest relationship possible between two variables is one in which the value of the dependent variable for every case in one category of the independent variable differs from that of every case in another category of the independent variable. We might call such a connection a *perfect relationship*, because the dependent variable is perfectly associated with the independent variable; that is, there are no exceptions to the pattern. If the results can be applied to future observations, a perfect relationship between the independent and dependent variables enables a researcher to predict accurately a case's value on the dependent variable given a known value of X.

A weak relationship would be one in which the differences in the observed values of the dependent variable for different categories of the independent variable are slight. In fact, the weakest observed relationship is one in which the distribution is identical for all categories of the independent variable—in other words, one in which no relationship appears to exist.

To get a better handle on strong versus weak relationships as measured by a cross-tabulation, consider the hypothetical data in tables 13–6 and 13–7. Assume we want to know if a connection exists between people's region of residency and attitudes about continuing the war in Iraq. (The hypothesis might be that southerners and westerners are more favorable than citizens in other parts of the country.) The frequencies and percentages in table 13–6 show no relationship between the independent and dependent variables. The relative frequencies (that is, percentages) are identical across all categories of the independent variable. Another way of thinking about nil relationships is to consider that knowledge of someone's value on the independent variable does not help predict his or her score on the dependent variable. In table 13–6, 48 percent of the easterners pick

TABLE 13–6
Example of a Nil Relationship between Region and Opinions about Keeping Troops in Iraq

	Region			
Opinion	East	Midwest	South	West
Favor keeping troops in Iraq	48% (101)	48% (103)	48% (145)	48% (97)
Favor bringing troops home	52% (109)	52% (111)	52% (158)	52% (106)
Total	100%	100%	100%	100%
N = 930	(210)	(214)	(303)	(203)

Note: Hypothetical responses to the question, "Do you favor keeping a large number of US troops in Iraq until there is a stable government there OR do you favor bringing most of our troops home in the next year?"

TABLE 13–7
Example of a Perfect Relationship between Region and Opinions about Keeping Troops in Iraq

	Region			
Opinion	East	Midwest	South	West
Favor keeping troops in Iraq	0% (0)	0% (0)	100% (303)	100% (203)
Favor bringing troops home	100% (210)	100% (214)	0% (0)	0% (0)
Total *N* = 930	100% (210)	100% (214)	100% (303)	100% (203)

Note: Hypothetical responses to the question, "Do you favor keeping a large number of US troops in Iraq until there is a stable government there OR do you favor bringing most of our troops home in the next year?"

"keeping," but so do 48 percent of the westerners, and for that matter, so do 48 percent of the inhabitants of the other regions. The conclusions are that (1) slightly more than half of the respondents in the survey want American troops brought home and (2) that there is *no* difference among the regions on this point. Consequently, the hypothesis that region affects opinions would not be supported by this evidence.

Now look at table 13–7, in which there is a strong—one might say nearly perfect—relationship between region and opinion. Notice, for instance, that 100 percent of the easterners and Midwesterners favor bringing the troops home, whereas 100 percent of the southerners and westerners have the opposite view. Or, stating the situation differently, knowing a person's region of residence lets us predict his or her response.

Most observed contingency tables, like table 13–5, fall between these extremes. That is, there may be a slight (but not nil) relationship, a strong (but not perfect) relationship, or a "moderate" relationship between two variables. Deciding which is the case requires the analyst to examine carefully the relative frequencies and determine if there is a substantively important pattern. When asked, "Is there a relationship between X and Y?" the answer will usually not be an unequivocal yes or no. Instead, the reply rests on judgment. If you think yes is right, then make the case by describing differences among percentages between categories of the independent variable. If, however, your answer is no, then explain why you think any observed differences are more or less trivial. A little later in the chapter, we present some additional methods and tools that help measure the strength of relationships.

Direction of a Relationship

In addition to assessing the strength of a relationship, one can also examine its "direction." The **direction of a relationship** shows which values of the independent variable are associated with which values of the dependent variable. This is an especially important consideration when the variables are ordinal or have ordered categories such as "high," "medium," and "low" or "strongly agree" to "strongly disagree" or the categories can reasonably be interpreted as having an underlying categorical spectrum, such as "least" to "most" liberal.

Table 13–8 displays the relationship between a scale of political liberalism (call it X) and a measure of opinions about gun control (Y). Both variables have an inherent order. The ideology variable can be thought of as running from lowest to highest liberalism, while responses to the question about firearms might be considered as going from least to most restrictive control.[5]

Take a moment to study the numbers in the table; we guarantee it will pay off in the long run. Start with the "most" liberal category. About two-thirds of respondents in this category (65.4%) are also "most" supportive of restricting gun purchases. That is, there is a tendency for "high" values of ideology to be associated with a "high value" of gun control. Now look in the last column, the "most conservative." You should see that a clear majority of these respondents (63.8%) are in the "least" enthusiastic category of Y, the dependent variable. Here, we have a case of "low" values tending to be linked to "low" values. The middle group (independent thinkers maybe) are more or less split between being for and against making it more difficult for people to buy firearms.

TABLE 13–8
Attitudes toward Gun Control by Liberalism

	Liberalism Scale (X)			
Make It Easier or Harder to Buy a Gun (Y)	Least conservative	Medium (middle of the road)	Most conservative	Total
Least favorable to guns (make it much harder to buy)	65.4% (72)	43.5% (226)	28.2% (50)	43.2% (348)
Medium (make it harder)	14.5% (16)	17.0% (88)	7.9% (14)	14.6% (118)
Most favorable to guns (make it easier to buy plus "same as now")	20.0% (22)	39.5% (205)	63.8% (113)	42.2% (340)
Total	100% (110)	100% (519)	100% (177)	100% (806)

Source: 2004 National Election Study.

Sometimes it helps to draw a sketch of the results. Consider the top row. The percentages decline as the one moves from "least" (66.5%) to "most" (28.2%) conservative. If you plot these numbers on a simple *X-Y* graph with equally spaced intervals for the *X* variable, you can see that the line decreases almost linearly, which can be interpreted simply as "The more conservative a person, the less favorable he or she feels toward stricter gun laws." (The percentage of each category saying "stricter" declines precipitously as one moves from liberals to independents to conservatives.) The upward-sloping line (positive slope) can be interpreted similarly. It shows the percentages in the third row, "make laws easier or keep the same," are plotted on the line that slants upward from left to right, which can be read as "The more conservative (the less liberal) an individual, the less favorable to controls" (see figure 13-3.) In both instances, we see at least monotonic correlation. (If you were to plot the middle row percentages, what would the line look like on the graph?)

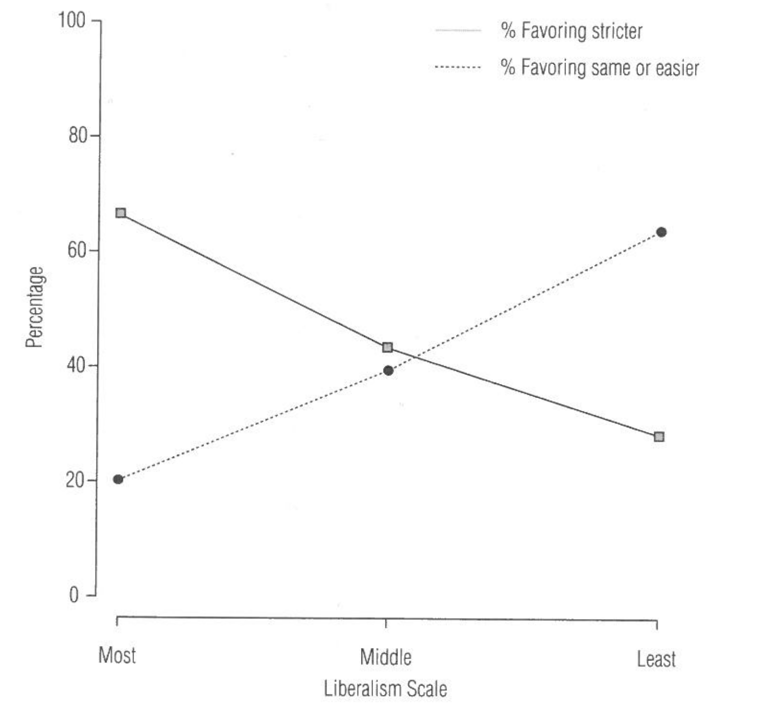
This figure is merely a heuristic device for understanding the data's practical meaning. A more formal method, suitable for presentations and papers, is the "stacked bar chart." Generally speaking, it is better to leave bivariate frequency distributions in tabular form so long as they contain the necessary numbers of cases and information about coding decisions. Interestingly, there are actually quite a few sophisticated graphical tools for summarizing cross-tabulations, even those with more than two variables.

We should add that the association between these two variables, although not perfect by the standards set forth earlier, is quite strong. Why this conclusion? As a preview of things to come, try this thought experiment. Suppose you were asked to predict how Americans would respond to a question about making gun control tougher. In the absence of any other information, you might take the "marginal" distribution of responses to the question in table 13–8 as a first approximation. (The marginal totals are in the rightmost column of the table.) Thus, you could reply, "Well most citizens are either for stricter controls (43.2%) or for leaving things as they are (42.2%) with a smattering of people (14.6%) in between." But suppose that you *also* knew people's political leanings. This knowledge would help you improve your predictions, because the least conservative (most liberal) individuals are apt to want stronger controls while conversely the most conservative (least liberal) respondents by and large favor leaving matters as they stand. So knowing a person's ideology enhances your predictive power. This idea—the proportional reduction in error—underlies several measures of association we will discuss shortly.

Assessing both the strength and type (direction) of a relationship in cross-classification tables requires looking at relative frequencies (percentages) cell by cell. That is not at all a bad practice. But statisticians have developed sophisticated methods for distilling the frequencies down to single numbers or "modeling" them in such a way that hard-to-see features become apparent. We next introduce a few of the ideas.

FIGURE 13–3

Simple Interpretation of Table Percentages: Liberalism and Gun Control



Source: Table 13–8.

Coefficients for Ordinal Variables

So far we have examined the relationship between two categorical variables by inspecting percentages in the categories of the independent variable. To fathom their messages, we have used rough sketches and visual inspection of the tables themselves. However, if the analysis involves many tables or tables that have many cells, another way of summarizing the information is needed. Here we introduce four correlation coefficients for ordinal variables.

TABLE 13–9
Table with Concordant, Discordant, and Tied Pairs

	Variable *X*		
Variable *Y*	High	Medium	Low
High	Alex	Dawn	Gus
Medium		Ernesto	Hera
Low	Carl	Fay	Ike Jasmine

These statistics, much like the descriptive statistics given in chapter 11, represent the data in a table with a single summary number that measures the strength and direction of an association. (You might want to review the introductory section that lists the properties of these indicators.) Among the most common statistics are **Kendall's tau-*b*, Kendall's tau-*c*, Somers' *D*** (two versions), and **Goodman and Kruskal's gamma**—named after the individuals who developed them. Most computer programs calculate these and other coefficients as well. They are similar, but not identical, in how they summarize the contents of a two-way frequency table.

We will not go into the details of their calculation, partly because software makes them so readily available, but instead concentrate on their numerical meaning. Nevertheless, a bit of background won't hurt. Each coefficient compares *pairs* of cases by determining whether those pairs are "concordant," "discordant," or "tied." These can be slippery concepts, so look at table 13–9. It contains nine individuals (cases).

- A *concordant pair* is a pair in which one individual is *higher* on *both* variables than the other case. Alex and Ernesto are concordant because Alex is higher on *Y* and *X*. Alex is also concordant with Fay, Hera, Ike, and Jasmine. There are other concordant pairs such as Dawn-Hera and Ernesto-Ike.
- A *discordant pair* is one in which one case is *lower* on one of the variables but *higher* on the other. Gus, for example, has a higher score on *Y* but a lower score on *X* compared to either Ernesto, Fay, or Carl. Therefore, these pairs "violate" the expectation that as one variable increases, so does the other.
- A *tied pair* is a pair in which both observations have the same value on one or both variables. There are lots of tied pairs in this table: Alex and Dawn are tied on *Y* (they both are in the "high" category"), Alex and Carl are tied on *X* (but not *Y*), and Ike and Jasmine are tied on both *X* and *Y*. (There are several others in the table.).

All of the ordinal coefficients of association (tau-*b*, tau-*c*, Somers' *D*, and gamma) use the probability or number of pairs of different kinds to summarize the relationship in a table. In a population, they measure the probability of a randomly drawn pair of observations being concordant minus the probability of being discordant with respect to *Y* and *X*:

$$\text{Measure} = p_{\text{concordance}} - p_{\text{discordance}},$$

where p means probability. They differ only in whether the probabilities are conditional on the presence or absence of ties. Gamma, for example, is defined as

$$\gamma = p_{C|\text{no ties}} - p_{D|\text{no ties}}.$$

In plain language, it is the probability that a randomly drawn pair will be concordant on *Y* and *X*, given that it is not tied, minus the corresponding probability of discordance. An "excess" of concordant pairs over discordant pairs suggests a positive relationship; if discordant pairs are more likely, then the correlation will be negative.

In samples, the basic comparison made is between the number of concordant and discordant pairs. If both types of pairs are equally numerous, the statistic will be zero, indicating no relationship. If concordant pairs are more numerous, the coefficient will be positive; if discordant pairs are outnumber concordant pairs, the statistic will be negative. The degree to which concordant or discordant pairs predominate, or one kind of pair is more frequent than the other, affects the magnitude of the statistic. Hence, if only the main diagonal were filled with observations, all the pairs would be concordant, and the statistic would be +1—a perfect, **positive relationship** (see table 13–10a). If only the minor (opposite) diagonal were filled with observations, all the pairs would be discordant, and the statistic would be -1—a perfect, **negative relationship** (see table 13–10b).

Gamma can attain its maximum (1 or -1) even if not all of the observations are on the main diagonal because it ignores all tied pairs. The others measures (tau-*b*,

TABLE 13–10
Perfect Positive and Negative Relationships

a. Every Pair Concordant (Perfect Positive Relationship)

	Variable *X*		
Variable *Y*	High	Medium	Low
High	Arthur		
Medium		Candy	
Low			Ed

b. Every Pair Discordant (Perfect Negative Relationship)

	Variable *X*		
Variable *Y*	High	Medium	Low
High			Faith
Medium		Guy	
Low	Hilary		

TABLE 13–11
Perfect Monotonic Relationship

	Variable X			
Variable Y	Very High	Medium High	Medium Low	Very Low
Very High	Abe			
Medium High		Bertha		
Medium Low			Claudio	
Very Low				Darby
Gamma ($\hat{\gamma}$) = 1.0.				

for example) "discount the strength of the relationship by the number of ties in the table."[6] Hence, in table 13–11, gamma would be 1.0, whereas the other coefficients would be slightly less.

In a "real" contingency table, there will be many pairs of all sorts, and counting them can be a nuisance. So we leave their computation to the computer. The formulas for these measures have the same form: one quantity divided by another. The numerator is always the number of concordant minus discordant pairs ($C - D$). The denominators differ, however, in how they handle ties. Gamma ignores tied pairs altogether, whereas the others incorporate them in different ways.[7] To help you understand them, we list a few of their properties.

- Theoretically, all vary between –1 and 1, with 1 indicating a perfect positive (monotonic) correlation and –1 a perfect negative (monotonic) correlation.
- In practice, you will most likely never see one of these coefficients attain these bounds. Indeed, even for strongly related variables, the numerical values will usually be far from 1 or –1. If any of them reaches, say, .4 or .5 in absolute value, there is an association worth investigating.

HOW IT'S DONE

Computing Ordinal Measures of Association

Let C = number of concordant pairs,

D = the number of discordant pairs,

T_X = the number of pairs tied only on X,

T_Y = the number of pairs tied only on Y,

T_{XY} = the number of pairs tied on both X and Y, and

m = the minimum of I or J, where I and J are the numbers of categories of Y and X, respectively.

Gamma: $$\hat{\gamma} = \frac{(C-D)}{(C+D)}$$

Tau-*b*: $$\hat{\tau}_b = \frac{(C-D)}{\sqrt{(C+D+T_Y)}\sqrt{(C+D+T_X)}}$$

Tau-*c*: $$\hat{\tau}_c = \frac{(C-D)}{N^2\left[\frac{(m-1)}{2m}\right]}$$

Somers' D: $$D_{YX} = \frac{(C-D)}{C+D+T_Y}$$

Somers' D: $$D_{XY} = \frac{(C-D)}{C+D+T_X}$$

- Since zero means no correlation, values in the range of –.1 to .1 suggest a weak relationship.
- All will have the same sign.
- The absolute value of gamma ($\hat{\gamma}$) will always be greater than or equal to that of any of the others. The relationships among tau-*b*, tau-*c*, and Somers' D are harder to generalize because they are affected differently by the cross-classification's structure (i.e., number of rows and columns).
- Somers' D is an "asymmetric" measure because its value depends on which variable is considered dependent. Therefore, there are really two possible versions: one, D_{YX}, has Y as the dependent variable, while the other, D_{XY}, treats X as dependent.
- By themselves, the measures are not sufficient to assess how and how strongly one variable is related to another. You should ask the software to calculate all the coefficients *and* spend time visually inspecting the relative frequencies in the table.[8]

The last point is worth emphasizing. None of the coefficients is appropriate if the relationship "curves," in the sense that as X increases so does Y up to a certain point when an increase in X is accompanied by a decrease in Y. Consider table 13–12, which contains four observations. There is a "perfect" association: you tell me a person's value on X, and I will predict exactly her score on Y. Yet the number of concordant pairs (3) equals the number of discordant ones (3), so their difference is zero. This difference ($C - D$) appears in the numerator of all the coefficients, so they would all be nil, implying no relationship. But there is an association; it's just not a correlation.

THE IMPORTANCE OF SCRUTINY. A well-known psychologist and statistician, Robert Abelson, titled a book *Statistics as Principled Argument*. His point was that statistics (either the numbers or the methods) do not speak for themselves. It is

TABLE 13–12
Perfect but Not Monotonic Relationship

	Variable X			
Variable Y	Very High	Medium High	Medium Low	Very Low
Very High				Doris
Medium High	Adele			
Medium Low		Barbara		
Very Low			Connie	

TABLE 13–13
Cross-classification of *Y* by *X* with Majority of Cases in One Row

Y	*X* 1	2	Total
A	550	608	1,158
B	20	7	27
C	12	4	16
Total	582	619	1,201

TABLE 13–14
Comparison of Concordant, Discordant, and Tied Pairs

Type of Pair	Number	Proportion
Concordant	6,130	.01
Discordant	19,540	.02
Tied (on *Y*, on *X*, and *Y* and *X*)	694,930	.96
Total	720,600	.99

Somers' D_{YX} = –.04; Tau-*b* = –.10; gamma = –.52.

Source: Table 13–13.

Note: Proportions do not add to 1.00 due to rounding error.

always necessary to make a case for point of view. Here is an example.

Gamma is one of the most widely reported ordinal coefficients. Its numerical value is always greater than or equal to the tau and Somers' measures, which raises the possibility that an investigator wanting to find large relationship might think gamma gives "the most association for the money." One difficulty, however, is that its computation only uses concordant and discordant pairs and ignores all tied pairs. Yet in some tables, the tied pairs greatly outnumber the concordant and discordant ones. Look at the simple tabulation cross-classification of *Y* by *X* in table 13–13.

By looking carefully at the row totals, you will see that that the vast majority of observations are in the first row. Stated differently, the marginal total is heavily skewed or concentrated in one category. As a consequence, there are not many concordant or discordant pairs compared to ties (see table 13–14).

If someone reported just gamma (–.52), the conclusion might be that a strong *Y*-*X* relationship exists. But we see that about 96 percent of the data have been "thrown out" and that the other coefficients show virtually no relationship.

There are a couple important lessons here. First, always pay attention to the shape of each variable's distribution, a point made emphatically in chapter 11. Social science data sets almost always contain skewed marginal totals on at least a few variables. (The data in this example come from a table analyzed later in the chapter.) Second, try not to rely on just one method, such as ordinal coefficients, to make a substantive claim. In other words, don't rely solely on a coefficient, no matter how convenient and interpretable it is.

Two Examples. Hypothetical data help establish the basic ideas of these ordinal measures of association, but when push comes to shove they do not give much practice understanding actual survey results. Therefore we provide two more tables that explore questions touched on earlier. The first is a cross-tabulation of vote in the 2008 presidential election by self-placement on a seven-point liberalism conservatism scale. The voting variable has only two categories (Barack Obama, Democrat, and John McCain, Republican), but any dichotomous variable (a variable with two categories) can be considered ordinal. You can construe the other variable as measuring the "degree" of conservatism. Since there are 7 × 2 = 14 relative frequencies to scrutinize, measures of (monotonic) correlation

TABLE 13–15
2008 Presidential Vote by Party

	Political Ideology						
	1 Least conservative	2	3	4 Middle	5	6	7 Most conservative
Obama	98.80% (56)	94.83% (167)	80.58% (118)	61.44% (203)	30.70% (60)	10.89% (34)	10.17% (5)
McCain	1.20 (1)	5.17 (9)	19.42 (28)	38.56 (127)	69.30 (134)	89.11 (275)	89.83 (46)
Totals	100% (57)	100% (176)	100% (146)	100% (330)	100% (194)	100% (309)	100% (51)

Question: "Where would you place yourself on this (liberalism-conservatism) scale, or haven't you thought much about this?"

Chi square = 517.99; 6 *df*; gamma = 0.818; tau-*b* = .564; Somers' D_{YX} = .719, λ = .575.

Source: The American National Election Studies (ANES; www.electionstudies.org). The ANES 2008 Time Series Study, Stanford University and the University of Michigan [producers].

may help us decide how closely ideology predicts candidate preference. This table is interpreted exactly like all the others: compare categories of ideology by the percentage in each who voted for, say, Obama.[9]

You should be able to detect a clear-cut pattern: as conservatism increases across the table, the propensity to vote for McCain also increases. Examine the percentages. (Notice, by the way, that the "Least" and "Most" conservative categories have relatively few cases in them. We might have combined those cases with the adjacent categories to improve the precision or reliability of the cell proportion estimates.)

All the measures are "large" by the standards of categorical data analysis. Gamma is 0.82, which indicates a strong positive correlation. Why positive? Consider the two variables as having an order: ideology runs from low to high conservatism. It is also legitimate to think of vote as having a numerical dimension, with Obama arbitrarily used as a low value and McCain as high. Consequently, moving along the columns from left to right, we see "low" values of conservatism associated with "low" values of vote (Obama) and high conservatism scores associated with "high" on voting (McCain). It may seem strange, but a dichotomous or two-category variable can often be interpreted this way. As we mentioned earlier, these numbers seldom get close to their maximums (|1.0|), and values over .4 to .5 indicate a strong correlation. So taken together, these suggest that ideology is highly correlated with voting. Overall, the conclusion is that position on the liberalism-conservatism spectrum predicts voting. Note, however, that since the data show only covariance and not time order or the operation of other variables, we cannot say this is a causal connection.

To wrap up this section, let us look at the second example, which returns to the idea of a gender gap: Are women more liberal than men and, if so, on what issues? Here the response variable is attitudes toward allowing gays to serve in the military. (These data too come from the 2008 ANES study used earlier.) Table 13–16 shows how gender relates to preferences about gays serving in the military. Conventional wisdom might say the women will be somewhat more open to the idea than men will.

The pattern here might be a bit harder to detect. Step back for a second and look at the column totals, as usual. In raw frequencies, there are more women in the sample than men, a common result in public opinion research. Still, there are enough of each gender to make meaningful comparisons. Note first of all that the vast majority of these respondents (55% + 23% = 78%) favor strongly or favor allowing gays to enlist in the military (the last column contains these totals). So right away we sense that there will not be huge sex differences on this issue. But when we look in the body of the table, we see that two-thirds of the women strongly favor lifting the ban on gay military service, and they are joined by 19 percent more who said simply "favor" (rows 4 and 5 of the table). That's 83 percent in favor! By contrast, the corresponding sum among men is 73 percent, a 10 percentage point difference. Note also that fewer than half of the men strongly favor lifting the ban whereas more than a quarter simply favor lifting

TABLE 13–16
Gays in the Military: A Gender Gap?

Gays serve in military?	Male (0)	Female (1)	Totals
(1) Strongly opposed	18.0% (183)	10.8 % (132)	14.1% (315)
(2) Opposed	9.3% (94)	5.8% (71)	7.4% (166)
(3) Favor	28.5% (289)	18.8% (231)	23.2% (520)
(4) Strongly favor	44.2% (449)	64.6% (794)	55.4% (1,244)
Totals	100% (1,015)	100% (1,229)	100% (2,245)

Summary statistics: gamma = .33, tau-b = .19, tau-c = .21, Somers' D = .21, $\hat{\lambda}$ = 0, χ^2 = 94.29 with 3 df.

the ban. So there is a difference in the distribution of men and women in the two categories on the favor side. If you look at the bottom of the table, a similar conclusion emerges. The ordinal coefficients help a bit. They show first a modest to weak correlation—as we saw from the percentages—and second that the relationship is positive.

In this instance, you can think of the variables as having an underlying order. Attitudes toward gays in the military runs from low to high support. Gender can be treated as if it were a numeric variable by letting men be 0 and women 1.[10] So as you move across and down the table, going in effect from low values on X and Y to high values, a slight positive correlation appears. (We place index numbers in parentheses in the table to illustrate the idea, but of course the measures of correlation introduced here do not in any way depend on numerical scale scores.) Beyond saying that there is a limited correlation which the percentages also reveal, these ordinal statistics do not have a common-sense or easily grasped interpretation. The situation improves slightly with the next coefficient.

A Coefficient for Nominal Data

When one or both of the variables in a cross-tabulation are nominal, ordinal coefficients are not appropriate because the identification of concordant and discordant pairs requires that the variables possess an underlying ordering (one value being higher than another). For these tables, different measures of association are employed. Some of the most useful rest on a *proportional-reduction-in-error* interpretation of association. The basic idea is this: you are asked to *predict* a randomly selected person's category or response level on a variable following two rules. Rule 1 requires you to make the guess in the absence of any other prior information (e.g., predict the individual's position on gun control). The other rule lets you know the person's score on a second variable, which you now take into account in making the prediction (e.g., you now know the individual's gender). Since you are guessing in both situations, you can expect to make some errors, but *if* the two variables are associated, then the using the second rule should lead to fewer errors than following the first.

How many fewer errors depends on how closely the variables are related. If there is no association at all, the expected number of errors should be roughly the same, and the reduction will be minimal. If, on the other hand, the variables are perfectly connected, in the sense that there is a one-to-one connection between the categories of the two variables, you would expect no errors by following rule 2. A "PRE measure" gives the **proportionate reduction in errors:**

$$PRE = \frac{(E_1 - E_2)}{E_1},$$

where E_1 is the number of errors made using rule 1 and E_2 is the number made under rule 2.

Suppose for a particular group of subjects the number of rule 1 errors (E_1) predicting variable scores on *Y* is 500. Now, think about these possibilities.

1. *X* has no association with *Y*. Then even using the individuals' *X* scores, the expected number of errors will still be 500, and the proportional reduction in errors will be (500 – 500)/500 = 0. This is the lower limit of a proportion, and it indicates *no* association.
2. Suppose the categories of *X* are uniquely associated with those of *Y* so that if you know *X*, you can predict *Y* exactly. The expected number of errors under rule 2 (E_2) will be zero. Consequently, *PRE* = (500 – 0)/500 = 1.0, the upper boundary for the measure. This means *perfect* association (according to this definition). In the third and last situation
3. Assume that *Y* and *X* have a moderate relationship. The expected number of errors following rule 2 might be, say, 200. Now we have

$$PRE = \frac{(500-200)}{500} = \frac{300}{500} = .6.$$

There is then a 60 percent reduction in prediction errors from knowing the value of *X*, a result that suggests a modest but not complete association.

LAMBDA. Many coefficients of association (e.g., gamma) can be defined in such a way as to lead to a *PRE* interpretation. We describe only one, however: **Goodman and Kruskal's lambda**. Lambda is a proportional-reduction-in-error coefficient. As we did earlier, imagine predicting a person's score on a variable in the absence of any other information ("rule 1"). What exactly would be the best strategy? If you did not know anything, you might ask what proportion of the population had characteristic A, what proportion characteristic B, and so forth for all of the categories of the dependent variable of interest. Let's say B was the most common (modal) category. Then, without other information, guessing that each individual was a B would produce fewer prediction errors than if you picked any other category. Why? Well, suppose there were 10 As, 60 Bs, and 30 Cs in a population of 100. Select a person at random and guess his or her category. If you picked, say, A, you would on average be wrong 60 + 30 = 90 times out of 100 guesses (90% incorrect). If, on the other hand, you chose C, you would be mistaken 10 + 60 = 70 times (70% errors). Finally, if you guessed the modal (most frequent) category, B, your errors would be on average 10 + 30 = 40. By choosing B (the mode), you do indeed make some incorrect predictions but many fewer than if you picked any other category. In sum, rule 1 states that, lacking any other data, your best long-run strategy for predicting an individual's class is to choose the modal one, the one with the most observations.

Now suppose you knew each case's score or value on a second variable, *X*. Say you realized a person fell in (or had property) M of the second variable. Rule 2 directs you to look only at the members of M and find the modal category. Assume that category C is most common among those who are M's. Given that the observation is an M, guessing C would (over the long haul) lead to the fewest mistakes. So rule 2 simply involves using rule 1 *within* each level of *X*.

The key to understanding lambda, a proportional-reduction-in-error type measure of association, lies in this fact: *if Y* and *X* are associated, then the probability of making an error of prediction using rule 1 will be greater than the probability of making an error with rule 2. How much greater? The measure of association, lambda (λ), gives the proportional reduction in error:

$$\lambda = \frac{\left(p_{\text{error 1}} - p_{\text{error 2}}\right)}{p_{\text{error 1}}},$$

where $p_{\text{error 1}}$ is the probability of making a prediction error with the first rule and similarly $p_{\text{error 2}}$ is the likelihood of an error knowing *X*. If the values of *X* are systematically connected to those of *Y*, the errors under the second rule will be less probable than those made under rule 1. In this case, lambda will be greater than zero. In fact, if *no* prediction errors result from rule 2, the probability p_{error2} will be zero, and

$$\lambda = \frac{\left(p_{\text{error 1}} - 0\right)}{p_{\text{error 1}}} = \frac{p_{\text{error 1}}}{p_{\text{error 1}}} = 1.0.$$

But of course if *X* and *Y* are unrelated, then knowing the value of *X* will tell you nothing about *Y*, and in the long run the probability of errors under both rules will be the same. So $p_{\text{error 1}} = p_{\text{error 2}}$ and

$$\lambda = \frac{\left(p_{\text{error 1}} - p_{\text{error 2}}\right)}{p_{\text{error 1}}} = \frac{(0)}{p_{\text{error 1}}} = 0.$$

The upshot is that lambda lies between 0 (no association) and 1.0 ("perfect" association as defined by the prediction rules.) A value of .5 would indicate a 50 percent reduction in errors, which in most situations would be quite a drop and hence suggest a strong relationship. A value of, say, .10—a 10 percent reduction—might signal a weak to nonexistent association. Note that correlation is not an issue here. If there is an *X*-*Y* link of whatever kind, lambda should pick it up. Yet, also remember that lambda does *not* take into account the ordering of the categories.

Again, we emphasize the importance of looking at the whole forest (the overall relationship) and not obsessing over a single tree (a measure of association). These kinds of statistics usually depend to a greater or lesser extent on the marginal distributions of the variables. Take care when a preponderance of observations are piled up in one or two categories.[11] For example, the lambda in table 13–15 is .575, which means knowing a person's ideology allows us to predict vote preference reasonably well; we cut prediction errors by more than 50 percent. This result, of course, agrees with our previous conclusion that voting is closely tied to ideology. (If you want to check another of lambda's characteristics, try scrambling the order of the columns in table 13–15. You should get the same result, .575.)

Testing a Cross-tabulation for Statistical Significance

Before taking up methods for describing relationships between other types of variables, we need to pause to think about on this problem. Apart from the hypothetical data, all of the examples presented so far use sample surveys. As samples go, most are quite large with slightly more than 1,000 cases. Nevertheless, since

FIGURE 13–4
Calculate Lambda

Frequency Table for Computing Lambda

Response (dependent or Y variable)

Decriminalize marijuana (dependent, Y)	Region (independent or X variable): East	North	South	West	Row marginal totals
For	(80)	50	20	10	160
No opinion	30	(90)	30	20	170
Against	10	40	(100)	(60)	(210)
Column margin totals	120	180	150	90	540

Maximum marginal frequency (Modal response category)

Total (N)

Maximum *column* (Y) frequencies (80 + 90 + 100 + 60 = 330)

$E_1 = 540 - 210 = 330$
$E_2 = 540 - 330 = 210$

$$\hat{\lambda} = \frac{(330 - 210)}{330} = .364$$

Source: Hypothetical data.

the totals represent only a tiny fraction of the population, one can always ask, "Do observed relationships reflect true patterns, or did they arise from chance or what is called sampling error?" Chapter 12 introduced concepts for answering that sort of question. Here we apply them to cross-classifications.

Statistical Independence. At this point it is useful to introduce a technical term that plays a large role in data analysis and that provides another way to view the strength of a relationship. Suppose we have two nominal or categorical variables, X and Y. For the sake of convenience, we can label the categories of the first $a, b, c, \ldots$ and those of the second $r, s, t, \ldots$ Let $P(X = a)$ stand for the probability that a randomly selected case has property or value a on variable X, and let $P(Y = r)$ stand for the probability that a randomly selected case has property or value r on Y. These two probabilities are called marginal probabilities and refer simply to the chance that an observation has a particular value (a, for instance) irrespective of its value on another. And, finally, $P(X = a, Y = r)$ stands for the joint probability that a randomly selected observation has both property a *and* property r simultaneously. The two variables are **statistically independent** if and only if the chances of observing a combination of categories is equal to the marginal probability of one category times the marginal probability of the other:

$$P(X = a, Y = r) = [P(X = a)][P(Y = r)] \text{ for all } a \text{ and } r.$$

If, for instance, men are as likely to vote as women, then the two variables—gender and voter turnout—are statistically independent, because, for example, the probability of observing a male nonvoter in a sample is equal to the probability of observing a male times the probability of picking a nonvoter.

In table 13–17, we see that 100 out of 300 respondents are men and that 210 out of the 300 respondents said they voted. Hence, the marginal probabilities are $P(X = m) = 100/300 = .33$ and $P(Y = v) = 210/300 = .7$. The product of these marginal probabilities is $(.33)(.7) = .23$. Also, note that because 70 voters are

HOW IT'S DONE

Calculating Lambda

To calculate lambda, follow these steps. (For an example using hypothetical data, see figure 13–4.)

1. Look at the cross-tabulation with both sets of marginal frequency (not percent) totals displayed.
2. Decide which variable is dependent.
3. Find the maximum marginal total for the dependent variable.
4. Subtract this total from table total, N, to get errors by method 1: N – (maximum frequency) = E_1, the number of predictions errors not knowing the independent variable.
5. In the body of the table, find the maximum frequency within *each* category of the independent variable.
6. Sum the maximums and subtract total from N. Call the result E_2, the number of prediction errors after using knowledge of the independent variable.
7. Calculate lambda:

$$\hat{\lambda} = \frac{(E_1 - E_2)}{E_1}.$$

Note, the numerical value of lambda depends on the choice of independent and dependent variables. Reversing them will usually change $\hat{\lambda}$.

TABLE 13–17
Voter Turnout by Gender

	Gender (*X*)		
Turnout (*Y*)	Male (*m*)	Female (*f*)	Total
Voted (*v*)	70	140	210
Did not vote (*nv*)	30	60	90
Total	100	200	300

Note: Hypothetical data. Cell entries are frequencies.

TABLE 13–18
Voter Turnout by Social Class

	Social Class (*X*)		
Turnout (*Y*)	Upper (*u*)	Lower (1)	Total
Voted (*v*)	100	50	150
Did not vote (*nv*)	50	100	150
Total	150	150	300

Note: Hypothetical data. Cell entries are frequencies.

male, the joint probability of being male *and* voting is 70/300 = .23, the same as the product of the marginal probabilities. Since the same relation holds for all other combinations in this data set, we infer that the two variables in table 13–17 are statistically independent.

Now suppose we had the data shown in table 13–18. There the sample consists of 300 respondents, half of whom voted and half of whom did not. The marginal probabilities of voting and not voting are both 150/300 = .5. It is also clear that the marginal probabilities of being upper- and lower-class equal .5. *If* the two variables were statistically independent, the probability that an upper-class respondent voted would be (.5)(.5) = .25. Similarly, the predicted probability (from these marginal totals) that a lower-class individual did not vote would be (.5)(.5) = .25. But we can see from *observed* cell frequencies that actual proportions of upper- and lower-class voters are .33 and .17, respectively. Since the observed joint probabilities do not equal the product of the marginal probabilities, the variables are not statistically independent. Upper-class respondents are more likely to vote than are lower-class individuals.

In this context, a test for statistical significance is really a test that two variables in a population are statistically independent. The hypothesis is that in the population, the variables are statistically independent, and we use the observed joint frequencies in a table to decide whether or not this proposition is tenable. Generally speaking, the stronger a relationship is, the more likely it is to be statistically significant, because it is unlikely to arise if the variables are really independent. However, even weak relationships may turn out to be statistically significant in some situations. In the case of cross-tabulations, the determination of statistical significance requires the calculation of a statistic called a chi-square, a procedure we discuss next.

Chi-Square Test for Independence. Table 13–19 pertains to civil liberties. It shows by levels of education attainment the degree of agreement with this statement: "Society shouldn't have to put up with those who have political ideas that are extremely different from the majority." The underlying hypothesis is that tolerance of dissent increases with education. By examining the cell proportions and the measures of association, you can surmise that a modest relationship

TABLE 13–19
Opinion on Civil Liberties: Tolerance of Dissent

	Educational Attainment				
Do not put up with extreme differences	Less than high school	High school graduate	Some post–high school education	College graduate or post graduate	Totals
Agree	45.58%	41.12%	23.17%	20.45%	31.7% (312)
Uncertain	7.65%	9.24%	6.80%	7.30%	7.8% (77)
Disagree	46.77%	49.63%	70.03%	72.25%	60.5% (596)
Totals	100% (154)	100% (312)	100% (270)	100% (249)	100% (985)

Chi square = 56.15 with 6 *df.*

Gamma = .32, tau-*b* = .20, tau-*c* = .19, Somers' D_{YX} = .24, lambda = 0, φ = 0.24.

Question: "Now I would like to ask about public affairs. Please indicate whether you agree. Society shouldn't have to put up with those who have political ideas that are extremely different from the majority." ("Agree strongly" and "agree" responses have been combined as have the disagree categories.)

Source: Citizen, Involvement, Democracy Survey, 2006.

exists between the two variables. (You might reinforce your understanding of the coefficients by interpreting them to yourself.) But is the relationship statistically significant? In the population is there really a relationship between tolerance and education?

Whether or not a relationship is statistically significant usually cannot be determined just by inspecting a cross-tabulation alone; instead, a statistic called **chi square** (χ^2) must be calculated. This statistic essentially compares an observed result—the table produced by sample data—with a "hypothetical" table that would occur if, in the population, the variables were statistically independent. Stated differently, the chi square measures the discrepancy between frequencies actually observed and those we would expect to see if there was no population association between the variables. When each observed cell frequency in a table equals the frequency expected under the **null hypothesis** of independence, chi square will equal zero. Chi square increases as the departures of observed and expected frequencies grow. There is no upper limit to how big the difference can become, but if it passes a certain point—a critical value—there will be reason to reject the hypothesis that the variables are independent.

How is chi square calculated? The observed frequencies are shown in the cross-tabulation in table 13–19.) Expected frequencies in each cell of the table are found by multiplying the row and column marginal totals and dividing by the sample size. As an example, consider the first cell in table 13–19.That cell is in the first row, first column of the table, so multiply the row total, 312, by the column total, 154, and then divide by 985, the total sample size in this table. The result is (312 × 154)/985 = 48.78. This is the *expected* frequency in the first cell of the table; it is what we would expect to get in a sample of 985 (with 312 "agrees" and 154 less than high school graduates) *if there is statistical independence in the population*. This is substantially less than the number we actually have, 70, so there is a difference. What about the other cells?

Let's do another example. If there were no association, how many college graduates would we expect to find in the "Disagree" category? Again, find the corresponding marginal totals (here 596 and 249), multiply them, and divide by 985 to get 150.66, the expected number under the null hypothesis. Notice, we keep repeating the phrase "under the . . ." We want to stress that this procedure can be interpreted as measuring the adequacy of a simple model (the model of no association) to these observed data. If the adequacy or fit is good, we say the model partially explains the data, which in turn is a manifestation of the real world. If the assumption of independence is not supported, we wouldn't anticipate that the expected frequencies would equal the observed ones except by chance.

Table 13–20 contains all of the expected frequencies for table 13–19. The overall measure of fit—the observed test statistic—is found by, in effect, comparing observed and expected frequencies. If the sum of differences is relatively small, do not reject the hypothesis of no association. But, if in the aggregate the discrepancy between observed and expected numbers is large, then the model upon which the expected frequencies are calculated is not a summary of the data, and the decision will be to reject the null hypothesis. So what is a large departure from the expected? The statistic is found by subtracting each expected frequency from its observed counterpart, squaring the difference (no minus sign

TABLE 13–20

Observed and Expected Values under Hypothesis of Independence

	Level of Education				
Do not put up with extreme differences	Less than high school	High school graduate	Some post–high school education	College graduate or postgraduate	Totals
Agree	**70** *48.78*	**128** *98.8*	**63** *85.5*	**51** *78.9*	**312**
Uncertain	**12** *12.0*	**29** *24.4*	**18** *21.1*	**18** *19.5*	**77**
Disagree	**72** *93.2*	**155** *188.8*	**189** *163.4*	**180** *150.7*	**596**
Totals	**154**	**312**	**270**	**249**	**985**

Note: Numbers in boldface font are observed frequencies; those in *italics* are expected frequencies under the hypothesis of statistical independence.

will be left), dividing the quotient by the expected frequency, and then adding the results over all the cells of the table. Hence, for table 13–20 we have

$$\chi^2_{obs} = \frac{(70-49)^2}{70} + \frac{(128-99)^2}{128} + \frac{(63-85)^2}{63} \ldots + \frac{(180-150)^2}{180} = 56.15.^*$$

This is the observed chi square, which we compare to a critical value to help decide whether or not to reject the null hypothesis.

Recall that a statistical hypothesis test entails several steps: specify the null and alternative hypothesis, specify a sample statistic and an appropriate sampling distribution, set the level of significance, find critical values, calculate the observed test statistic, and make a decision. A chi-square test of the statistical independence of Y and X has the same general form.

1. Null hypothesis: X and Y are statistically independent.
2. Alternative hypothesis: X and Y are not independent. The nature of the relationship is left unspecified.
3. Sampling distribution: Choose chi square. This distribution is a family each of which depends on *degrees of freedom* (*df*). The **degrees of freedom** equals the number of rows (I) minus 1 times the number of columns (J) minus 1 or $(I - 1)(J - 1)$.
4. Level of significance: Choose the probability (α) of incorrectly rejecting a true null hypothesis.
5. Critical value: The chi-square test is always one-tailed. Choose the critical value of chi square from a tabulation to make the critical region (the region of rejection) equal to α.
6. The observed chi square is the sum of the squared differences between observed and expected frequencies, divided by the expected frequency.
7. Reject the null hypothesis if the observed chi square equals or exceeds the critical chi square; that is, reject if $\chi^2_{obs} \geq \chi^2_{critical}$. Otherwise, do not reject.

For the tolerance and education example, the null hypothesis is simply that the two variables are independent. The alternative is that they are not. (Yes, this is an uninformative alternative in that it does not specify *how* education and political tolerance might be related. This lack of specificity is a major criticism of the common chi-square test. But this is nevertheless a first step in categorical data analysis.) For this test, we will use α = .01 level of significance. To find a critical value, it is necessary to first find the degrees of freedom, which in this case is $(4 - 1)(3 - 1)$, or 6.

*results subject to rounding errors.

TABLE 13–21
Relationship between *X* and *Y* Based on Sample of 300

	Variable *X*			
Variable *Y*	*A*	*B*	*C*	TOTAL
A	30	30	30	90
B	30	30	36	96
C	40	40	34	114
Total	100	100	100	300

$\chi^2 = 1.38$, 4 *df*; $\phi = .07$.

Note: Hypothetical data.

Then we look in a chi-square table to find the value that marks the upper 1 percent (the .01 level) of the distribution (see appendix C). Read down the first column (*df*) until you find the degrees of freedom (6 in this case) and then go across to the column for the desired level of significance. With 6 degrees of freedom, the critical value for the .01 level is 16.81. This means that if our observed chi square is greater than or equal to 16.81, we reject the hypothesis of statistical independence. Otherwise, we do not reject it.

The observed chi square for table 13–20 is 56.15 with 6 degrees of freedom. (*Always report the degrees of freedom.*) Clearly, this greatly exceeds the critical value (16.81), so we would reject the independence hypothesis at the .01 level. Indeed, if you look at the chi-square distribution table, you will see that (for 6 degrees of freedom) 56.15 is much larger than the highest listed critical value, 22.46, which defines the .001 level. So really this relationship is "significant" at the .001 level. We place quote marks around "significant" to reemphasize that all we have done is reject a null hypothesis. We have not necessarily produced a momentous finding. This statement leads to our next point.

The sample size, *N*, and the distribution of cases across the table always have to be taken into account. Large values of chi square occur when the observed and expected tables are quite different and when the sample size upon which the tables are based is large. A weak relationship in a large sample may attain statistical significance, whereas a strong relationship found in a small sample may not. Keep this point in mind. If *N* (the total sample size) is large, the magnitude of the chi-square statistic will usually be large as well, and we will reject the null hypothesis even if the association is quite weak. This point can be seen by looking at tables 13–21 and 13–22. In table 13–21, the chi square of 1.38 suggest that there is virtually no relationship between the categories *X* and *Y*. In table 13–22, which involves a larger sample size but no other difference, the chi-square statistic (13.8) is now statistically significant (at the .05 level). However, the strength of the relationship between *X* and *Y* is still the same as before, namely, quite small.

TABLE 13–22
Relationship between *X* and *Y* Based on Sample of 3,000

	Variable *X*			
Variable *Y*	A	B	C	TOTAL
A	300	300	300	900
B	300	300	360	960
C	400	400	340	1,140
Total	1,000	1,000	1,000	3,000

$\chi^2 = 13.8$, 4 *df*; $\phi = .07$.

Note: Hypothetical data.

The lesson to be drawn here is that when dealing with large samples (say, $N > 1{,}500$), small, inconsequential relationships can be statistically significant.[12] As a result, we must take care to distinguish between statistical and substantive importance. The fact that chi square rapidly inflates with increases in the sample size has led statisticians to propose measures that try to take *N* into account. A simple one, **phi** (ϕ), adjusts the observed chi-square statistic by dividing it by *N* and taking the square root of the quotient. (Because of the division by *N*, the statistic is sometimes referred to as the "mean square contingency coefficient.") Yet, like chi square, phi does not have a readily interpretable meaning, so it is mostly used for comparison. (In ideal situations, phi varies between 0 and 1, but in many bivariate distributions, it can exceed 1.) We see in tables 13–21 and 13–22 that phi does not change even though the chi-square statistic does. So even though we do not use it much in this book, it comes in handy on occasion. If you look back to table 13–19, you will see that phi = .24, indicating once more the weak to moderate relationship between education and political tolerance.

HOW IT'S DONE

The phi Coefficient

Although most software calculates phi as a matter of course, it can be calculated quickly by hand if the observed chi square is available:

$$\varphi = \sqrt{\frac{\chi^2_{\text{obs}}}{N}},$$

where *N* is the sample size.

Generally speaking, the chi-square test is only reliable for relatively large *N*s. Stating exactly how large is difficult because the answer depends on the table's number of rows and columns (or, more formally, its degrees of freedom). Many times, as in a table with many cells (see table 13–20), a sample will be large but the table will contain at least some cells with small frequencies. Very few respondents in the CID study seemed "uncertain," so frequencies in that row are small compared to the others. A rule of thumb directs analysts to be cautious if any cell contains expected frequencies of 5 or fewer, and many cross-classification programs flag these "sparse cells." If you run across this situation, the interpretation of the chi-square value remains the same but should be perhaps advanced with less certainty. Moreover, if the total sample size is less than 20 to 25, alternative procedures are preferable for testing for significance.[13]

Remember: the chi-square statistic in and by itself is not a very good indicator of the strength of an association; rather, it tests the statistical significance of any association that does appear. Assessing relationships is thus a two-step process: (1) measure the strength of the association with percentages, proportions, and coefficients and (2) test to see if the observed results could might have arisen by chance. The first step is the crucial one: make sure the relationship is "worth talking about" and *then* test its significance.

DIFFERENCE-OF-MEANS TESTS

Linear Models

Cross-tabulation is the appropriate analysis technique when both variables are nominal- or ordinal-level measures. When the independent variable is nominal