social class is assigned the lowest score and so forth. If you did not check the coding scheme, you might conclude that class decreases as education increases, when, actually, the opposite is true.

Unfortunately, this source of confusion cannot be avoided when working with ordinal-level variables. Coding schemes will always be arbitrary for these variables, so you need to exercise additional caution when interpreting the direction of ordinal-level variables.

Spearman's Rho (R_s)

In the previous section, we considered ordinal variables that have a limited number of categories (possible values) and are presented in tables. However, many ordinal-level variables have a broad range and many distinct scores. These variables may be collapsed into a few broad categories (such as high, moderate, and low), organized into a bivariate table, and analyzed with gamma. Collapsing scores in this manner may be desirable in many instances but may also obscure or lose some important distinctions between cases. For example, suppose a researcher wished to test the claim that jogging is beneficial, not only physically but also psychologically. Do joggers have an enhanced sense of self-esteem? To deal with this issue, 10 female joggers are evaluated on two scales—the first measuring involvement in jogging and the other measuring self-esteem. Scores are reported in Table 12.9.

These variables could be collapsed by splitting the scores into two values (high and low). There are two potential difficulties with collapsing scores in this way. First, the scores seem continuous and there are no obvious or natural division points in the distribution that would allow us to distinguish between high scores and low ones. Second—and more importantly—grouping these cases into broader categories will lose information. That is, if Wendy and Debbie are placed in the category "high" on involvement, the fact that they had different scores would be lost. If differences like this are important and meaningful, then we should choose a measure of association that permits the retention of as much detail and precision in the scores as possible.

TABLE 12.9 The Scores of 10 Subjects on Involvement in Jogging and a Measure of Self-Esteem

Jogger	Involvement in Jogging (X)	Self-Esteem (Y)
Wendy	18	15
Debbie	17	18
Phyllis	15	12
Stacey	12	16
Evelyn	10	6
Tricia	9	10
Christy	8	8
Patsy	8	7
Marsha	5	5
Lynn	1	2

The Computation of Spearman's Rho

Spearman's rho (r_s) is a measure of association for ordinal-level variables that have a broad range of many different scores and few ties between cases on either variable. Scores on ordinal-level variables cannot, of course, be manipulated mathematically except for judgments of "greater than" or "less than." To compute Spearman's rho, the cases are first ranked from high to low on each variable and then the ranks (not the scores) are manipulated to produce the final measure. Table 12.10 displays the original scores and the rankings of the cases on both variables.

To rank the cases, first find the highest score on each variable and assign it rank 1. Wendy has the high score on X (18) and is thus ranked number 1. On the other hand, Debbie is highest on Y and is ranked first on that variable. All other cases are then ranked in descending order of scores. If any cases have the same score on a variable, assign them the average of the ranks they would have used up had they not been tied. Christy and Patsy have identical scores of 8 on involvement. Had they not been tied, they would have used up ranks 7 and 8. The average of these two ranks is 7.5, and this average of used ranks is assigned to all tied cases. (For example, if Marsha had also had a score of 8, three ranks—7, 8, and 9—would have been used and all three tied cases would have been ranked eighth.)

The formula for Spearman's rho is:

FORMULA 12.2

$$r_s = 1 - \frac{6\Sigma D^2}{N(N^2 - 1)}$$

where:

ΣD^2 = the sum of the differences in ranks, the quantity squared

To compute ΣD^2, the rank of each case on Y is first subtracted from its rank on X (D is the difference between rank on Y and rank on X). A column has been

TABLE 12.10 Computing Spearman's Rho

	Involvement(X)	Rank	Self-Image(Y)	Rank	D	D^2
Wendy	18	1	15	3	−2	4
Debbie	17	2	18	1	1	1
Phyllis	15	3	12	4	−1	1
Stacey	12	4	16	2	2	4
Evelyn	10	5	6	8	−3	9
Tricia	9	6	10	5	1	1
Christy	8	7.5	8	6	1.5	2.25
Patsy	8	7.5	7	7	0.5	0.25
Marsha	5	9	5	9	0	0
Lynn	1	10	2	10	0	0
					$\Sigma D = 0$	$\Sigma D^2 = 22.50$

provided in Table 12.10 to record the differences case by case. Note that the sum of this column (ΣD) is 0. That is, the negative differences in rank are equal to the positive differences, as will always be the case, and you should find the total of this column as a check on your computations to this point. If ΣD is not equal to 0, you have made a mistake either in ranking the cases or in subtracting the differences.

In the column headed D^2, each difference is squared to eliminate negative signs. The sum of this column is ΣD^2, and this quantity is entered directly into the formula. For our sample problem:

$$r_s = 1 - \frac{6\Sigma D^2}{N(N^2 - 1)}$$

$$r_s = 1 - \frac{6(22.5)}{10(100 - 1)}$$

$$r_s = 1 - \frac{135}{990}$$

$$r_s = 1 - 0.14$$

$$r_s = 0.86$$

The Interpretation of Rho

Spearman's rho is an index of the strength of association between the variables; it ranges from 0 (no association) to ± 1.00 (perfect association). A perfect positive association ($r_s = +1.00$) would exist if there were no disagreements in ranks between the two variables (if cases were ranked in exactly the same order on both variables). A perfect negative relationship ($r_s = -1.00$) would exist if the ranks were in perfect disagreement (if the case ranked highest on one variable were lowest on the other and so forth). A Spearman's rho of 0.86 indicates a strong positive relationship between these two variables. The respondents who were highly involved in jogging also ranked high on self-image. These results are supportive of claims regarding the psychological benefits of jogging.

Spearman's rho is an index of the relative strength of a relationship, and values between 0 and ± 1.00 have no direct interpretation. However, if the value of rho is squared, a PRE interpretation is possible. Rho squared (r_s^2) represents the proportional reduction in errors of prediction when predicting rank on one variable from rank on the other variable, as compared to predicting rank while ignoring the other variable. In our example, r_s was 0.86 and r_s^2 would be 0.74. Thus, our errors of prediction would be reduced by 74% if we used the rank on jogging to predict rank on self-image. *(For practice in computing and interpreting Spearman's rho, see problems 12.10 to 12.13. Problem 12.10 has the fewest number of cases and is probably a good choice for a first attempt at these procedures.)*

ONE STEP AT A TIME Computing and Interpreting Spearmans Rho

Step	Operation
To Compute Spearman's Rho	
1.	Set up a computing table like Table 12.10 to organize the computations. In the far left-hand column, list the cases in order, with the case with the highest score on the independent variable (*X*) stated first.
2.	In the next column, list the scores on *X*.
3.	In the third column, list the rank of each case on *X*, beginning with rank 1 for the highest score. If any cases have the same score, assign them the average of the ranks they would have used up had they not been tied.
4.	In the fourth column, list the score of each case on *Y*; in the fifth column, rank the cases on *Y* from high to low. Assign the rank of 1 to the case with the highest score on *Y*, and assign any tied cases the average of the ranks they would have used up had they not been tied.
5.	For each case, subtract the rank on *Y* from the rank on *X* and then write the difference (*D*) in the sixth column. Add up this column. If the sum is not zero, you have made a mistake and need to recompute.
6.	Square the value of each *D* and then record the result in the seventh column. Add this column up to find ΣD^2 and then substitute this value into the numerator of Formula 12.2.
7.	Multiply ΣD^2 by 6.
8.	Square *N* and subtract 1 from the result.
9.	Multiply the quantity you found in step 8 by *N*.
10.	Divide the quantity you found in step 7 by the quantity you found in step 9.
11.	Subtract the quantity you found in step 10 from 1. The result is r_s.
Interpreting the Strength of the Relationship	
1.	Use either or both of the following to interpret strength: a. Use Table 12.2 to describe strength in general terms. b. Square the value of rho and then multiply by 100. This value represents the percentage improvement in our predictions of the dependent variable by taking the independent variable into account.
Interpreting the Direction of the Relationship	
1.	Use the sign of r_s. Be careful when interpreting direction with ordinal-level variables. Remember that coding schemes for these variables are arbitrary and that a positive r_s may mean that the actual relationship is negative and vice versa.

STATISTICS
IN EVERYDAY LIFE

Rankings have become a staple of the media and the Internet, and there seem to be infinite lists of everything imaginable: best places to live, cities with the worst traffic, most important news stories, most popular celebrities, and so forth. These lists are bound to be somewhat arbitrary and subjective but can also provide useful insights and perspectives. A search of the Internet for "top ten" lists provides a

STATISTICS
IN EVERYDAY LIFE

(continued)

multitude of examples—one of the most interesting of which was a list of the most haunted cities in the United States:

Rank	City	Comment
1	New Orleans, LA	A city full of haunted mansions, taverns, and graveyards
2	Savannah, GA	Fits the bill of a haunted city as well as any in the United States
3	Gettysburg, PA	Almost 50,000 died during the battle
4	Chicago, IL	The great fire, gangsters, and "Resurrection Mary"
5	Salem, MA	Witches!
6	Charleston, SC	An older city with plenty of time to accumulate ghost stories
7	Portland, OR	The ghostliest city in the Northwest
8	Athens, OH	One-time home of a "Lunatic Asylum"
9	Key West, FL	Pirates, rumrunners, Ernest Hemingway, and Robert the Doll
10	San Francisco, CA	A rich cultural legacy, disasters, and Alcatraz

More information is available at the website cited below.

Source: http://www.toptenz.net/top-10-most-haunted-cities-in-the-u-s.php.

Applying Statistics 12.3 Spearman's Rho

Five cities have been rated on an index that measures the quality of life. Also, the percentage of the population that has moved into each city over the past year has been determined. Have cities with higher quality-of-life scores attracted more new residents? The table below summarizes the scores, ranks, and differences in ranks for each of the five cities.

Spearman's rho for these variables is:

$$r_s = 1 - \frac{6\Sigma D^2}{N(N^2 - 1)}$$

$$r_s = 1 - \frac{(6)(4)}{5(25 - 1)}$$

$$r_s = 1 - \left(\frac{24}{120}\right)$$

$$r_s = 1 - 0.20$$

$$r_s = 0.80$$

These variables have a strong positive association. The higher the quality-of-life score, the greater the percentage of new residents. The value of r_s^2 is 0.64($[0.80]^2 = 0.64$), which indicates that we will make 64% fewer errors when predicting rank on one variable from rank on the other, as opposed to ignoring rank on the other variable.

City	Quality of Life	Rank	% New Residents	Rank	D	D^2
A	30	1	17	1	0	0
B	25	2	14	3	–1	1
C	20	3	15	2	1	1
D	10	4	3	5	–1	1
E	2	5	5	4	1	1
					$\Sigma D^2 = 0$	$\Sigma D^2 = 4$