CHAPTER 14
ORDINAL MEASURES OF CORRELATION: SPEARMAN'S RHO AND GAMMA

Chapter 13 introduced the concept of correlation statistics and explained the use of Pearson's Correlation Coefficient when working with interval level data. This chapter introduces two additional measures of correlation called Spearman's Rho and Gamma. These statistical tools are employed when researchers seek to evaluate the level of correlation between variables representing ordinal measures.

Ordinal data comes in two general types. The first type of ordinal data is often called continuous ordinal data. This type of data is found in situations where the researcher has a relatively detailed set of rankings where the cases have been ranked in a relatively broad range of categories and where the number or tied rankings is relatively small. Researchers working with this type of ordinal data use Spearman's Rho to determine the level of correlation that exists between variables.

The second type of ordinal data is referred to as a collapsed ordinal variable. Collapsed variables place all observations into response categories which are clearly ordinal in nature, but which are limited in number (no more than five or six). This type of ordinal measure is extremely common in survey research and is exemplified by the Agree-Disagree questions often employed by researchers. Respondents are asked to respond to a statement in one of five response categories that range from strongly agree to strongly disagree. There is a clear order to the response categories which indicates that the data is ordinal in nature, but the limited number of response options creates a data set with an extremely large number of observations that are tied in the overall rankings. When working with this type of collapsed ordinal variables, Gamma is the appropriate statistical tool for exploring the level of correlation that exists between

variables.

Spearman's Rho

Spearman's Rho (r_s) produces a rank order correlation coefficient that is similar to the

correlation coefficient produced by the Pearson's Correlation Coefficient test. This test evaluates

the degree to which individuals or cases with high rankings on one variable were observed to

have similar rankings on another variable. Calculation of the statistic is a relatively simple and

straightforward procedure.

In some cases, researchers will work with data that has already been ranked. In others,

the first step in the process of calculating Spearman's Rho will involve assigning ranks. It is

extremely important to remember that this process is conducted in a slightly different manner for

Spearman's rho than it was for the Mann Whitney U statistic. With Spearman's rho, the *highest*

value is assigned a rank of 1 and ranks are assigned separately for each variable. A solution

matrix is created once ranks have been assigned to each case on both of the variables under

consideration. Tied scores are ranked in the same way that is used with Mann Whitney U. This

means that each of the tied scores is assigned a rank equal to the average of all the tied positions.

For example, if a pair of scores are tied for the 2nd and 3rd ranks, both scores are assigned a rank

of 2.5 $\left(\dfrac{2+3}{2} = 2.5 \right)$. The next highest score is then assigned a rank of 4. That solution

matrix includes values for each of the variables and two additional columns that represent the

difference between the ranking for the first and second variable (X and Y) and the squared

difference in rankings. An example of a solution matrix for Spearman's Rho is provided in

Figure 14:1:

Figure 14:1
Solution Matrix – Spearman's Rho

Respondent	X	Y	D $(X-Y)$	D^2 $(X-Y)^2$
1	1	5	-4	16
2	3	1	2	4
3	5	2	3	9
4	2	4	-2	4
5	4	3	1	1
				$\sum D^2 = 34$

Once the solution matrix has been completed, Spearman's rho is calculated using the

formula: $r_s = 1 - \dfrac{6\sum D^2}{n(n^2-1)}$. The value for the sum of differences squared is calculated from

the total of values in the final column of the solution matrix. The value for n is equal to the

number of cases or respondents included in the data set (5 in the example above). The final step

in calculating the value of the statistic is substitution and simplification.

$$r_s = 1 - \frac{6\sum D^2}{n(n^2-1)}$$

$$r_s = 1 - \frac{6(34)}{5(5^2-1)}$$

$$r_s = 1 - \frac{204}{5(25-1)}$$

$$r_s = 1 - \frac{204}{120}$$

$$r_s = 1 - 1.7$$

$$r_s = -.7$$

The result of this process is a rank order correlation coefficient of -.7 which provides an initial indication of the existence of a relatively strong negative correlation between the two variables. To determine if the obtained value of Spearman's Rho is sufficient to be deemed statistically significant, consult Appendix L at the back of the text. Appendix L lists the critical values for Spearman's Rho at both the .05 and .01 levels of significance. Looking at the table, you will find that the critical value for rejecting the null hypothesis at the .05 level with a sample size of 5 is 1. Since the obtained value for the correlation coefficient falls within the range of values between 1 and -1, the null hypothesis must be accepted and we must conclude that there is not a statistically significant correlation between the two variables.

Gamma

The gamma statistic is used when working with ordinal level data that is ranked in a small number of response categories. It is designed to determine how effectively a researcher can use information about an individuals ranking on one variable to predict that individuals ranking on the other. An obtained value of +1 for gamma indicates the presence of a perfect correlation between the two variables. This means that an individual who ranked above another on one variable would always rank above them on the other variable as well. In contrast, an obtained value of -1 indicates the presence of a perfect negative correlation. This means that a person who ranked above another on one variable would always rank below them on the other.

The gamma statistic computes the level of association between the two variables based on two sums that are obtained by constructing a cross tabulation and applying two formulas. The first sum represents the Number of Agreements (N_a). This sum is determined by identifying the number of cases that are ranked in the same relative position on both variables. The second

sum represents the Number of Inversions (N_i). This sum is determined by identifying the

number of cases that ranked differently on the two variables (higher on one and lower on

another). To compute Gamma, begin with the construction of a cross tabulation that represents

the observed values for each case under consideration on both of variables. When constructing

the cross tabulation, the highest ranking should be at the top among the rows and at the left

among the columns. Figure 14:2 provides an example of one of these cross tabulations.

Figure 14:2

Cross Tabulation of Relative Rankings

	High	Medium	Low	Total
High	5	7	20	32
Medium	10	8	15	33
Low	15	6	10	31
Total	30	21	45	96

A pair of cases is represented by taking any one case from a cell and comparing to any other case

in the table. As an example, 1 pair of cases can be identified by looking at 1 of the 5 cases from

the cell representing those ranking high in both categories (We'll call this person subject A) and 1

of the 8 cases in the cell representing those ranking in the medium category on both variables

(Subject B). The pair of cases identified in the example represents a case of agreement. Just as

Subject A is ranked above Subject B on one variable, Subject A is also ranked above Subject B

on the other variable. Gamma computes the total number of pairs that fit into the category of

agreement by multiplying the cell frequency of a particular cell by the sum of frequencies for all

cells that represent agreement. Using the cross tabulations constructed at the first stage of the

process, the researcher begins with the upper left cell and multiplies its frequency by the sum of

all cells located both *below and to the right*. This procedure is repeated for each cell. The results

are then added together to yield the value for N_a. Figure 14:3 demonstrates this process.

Figure 14:3
Number of Agreements

Cell	Number of Agreements	Contribution to N_a
HH (High/High)	5 (8+15+6+10)	195
HM	7 (15+10)	175
HL	20 (0)	0
MH	10 (6+10)	160
MM	8 (10)	80
ML	15 (0)	0
LH	15 (0)	0
LM	6 (0)	0
LL	10 (0)	0
		610

A similar procedure is followed to determine the number of inversions or disagreements. This

represents the number of pairs of observations where an individual rank differently on each of the

variables (Subject A is higher than B on one variable and lower on the other). Calculation of the

number of inversions begins in the upper right cell of the cross tabulation (High/Low in the

example) and multiplies the number of cases in the cell by the sum of all cells *below and to the*

left of that cell. The results are then added together to yield a total for N_i. The procedure is

demonstrated in Figure 14:4.

Cell	Number of Inversions	Contribution to N_i
HL	20 (8+10+15+6)	780
HM	7 (10+15)	175
HH	5 (0)	0
ML	15 (6+15)	315
MM	8 (15)	120
MH	10 (0)	0
LL	10 (0)	0
LM	6 (0)	0
LH	15 (0)	0
		1390

The values generated by the process described provides all of the information needed to calculate

the gamma statistic. The formula for gamma is: $G = \dfrac{N_a - N_i}{N_a + N_i}$. Simply substitute and solve

for gamma.

$$G = \frac{N_a - N_i}{N_a + N_i}$$

$$G = \frac{610 - 1390}{610 + 1390}$$

$$G = \frac{-720}{2000}$$

$$G = -.39$$

The initial indication produced by the calculation of gamma is that there is a moderate negative correlation between the two variables. To determine the significance of a gamma statistic, a z-score is calculated based upon the following formula: $z = G\sqrt{\dfrac{N_a + N_i}{N(1 - G^2)}}$

Therefore the calculation of a z-score for the example we have been addressing would follow the following process.

$$z = G\sqrt{\dfrac{N_a + N_i}{N(1 - G^2)}}$$

$$z = -.39\sqrt{\dfrac{610 + 1390}{96(1 - (-.39^2))}}$$

$$z = -.39\sqrt{\dfrac{2000}{96(1 - .15)}}$$

$$z = -.39\sqrt{\dfrac{2000}{96(.85)}}$$

$$z = -.39\sqrt{\dfrac{2000}{81.6}}$$

$$z = -.39\sqrt{24.51}$$

$$z = -.39(4.95)$$

$$z = -1.93$$

The obtained value for the z-score is then compared to the critical values of z to determine if the correlation is statistically significant. The critical value for z at the .05 level is ± 1.96. At the .01 level the critical value is ± 2.58. In the example, the obtained value for z falls within the range defined by -1.96 and +1.96. Therefore, the null hypothesis is accepted and we must conclude that there is no significant relationship between the two variables.

The two statistics introduced in this chapter provide tools necessary for determining the level of correlation and significance of association between two ordinal variables. Spearman's Rho is employed when there is a relatively broad range of rankings and relatively few ties exist. Gamma provides a mechanism for evaluating correlation when the range of rankings is narrowly constrained and large numbers of ties are produced. In chapter 15, a new statistic called Cramer's V is explained which provides a measure of association that may be employed when working with nominal data.

Exercises
Chapter 14

1. Using the data provided below, construct rankings for each variable and calculate the rank order correlation coefficient using Spearman's Rho. Determine if the correlation is statistically significant. Show all work. Draw a conclusion related to the correlation.

Subject	X	Y
1	82	80
2	73	75
3	95	89
4	66	76
5	84	79
6	89	73
7	51	69
8	82	89
9	75	77
10	90	85
11	51	71
12	89	69
13	34	51
14	50	60
15	87	74

2. Using the data provided below, construct rankings and calculate the rank order correlation coefficient using Spearman's Rho. Determine the significance of the correlation. Show all work.

X = 82, 73, 95, 66, 84, 89, 51, 82, 75, 90, 60, 81, 34, 49, 82, 95, 49

Y = 76, 83, 89, 76, 79, 73, 52, 89, 77, 85, 48, 69, 51, 25, 74, 60, 50

3. A researcher wants to determine if the length of time that it takes an individual to commute to work in the morning is related to the level of job satisfaction that the individual expresses on a survey. Using the data provided below, compute a value for gamma. Is the obtained value for gamma statistically significant? Show all work and draw appropriate conclusions.

Commuting Time (Minutes)				
Job Satisfaction	60+	30-59	15-29	Under 15
Very Satisfied	8	12	25	22
Somewhat Satisfied	9	20	23	11
Dissatisfied	13	18	17	7
n=185				

4. A researcher is interested in the relationship between the amount of reading an individual does and social class. 105 children were sampled about the number of books they had read over the past year. Using the following data, compute Gamma. Determine if there is a statistically significant relationship between the variables. Show work and draw appropriate conclusions.

Socioeconomic Status				
Job Satisfaction	Upper	Upper Middle	Lower Middle	Lower
More than one	10	9	6	4
One	8	10	8	12
None	5	6	12	15
n=105				