METHODOLOGY CORNER

Treating Ordinal Scales As Ordinal Scales

THOMAS R. KNAPP

In a recent article in this journal, Knapp (1990) summarized and tried to resolve the controversy regarding the treatment of ordinal scales as interval scales. He cited Agresti's (1984) book on the analysis of ordinal data as a source of a variety of techniques that have been developed explicitly to handle those bothersome scales that permeate so much of nursing research. The purpose of the present paper is to bring to the attention of nursing researchers a variation of one of the strategies for analyzing ordinal/ordinal relationships that is explained in Agresti's 1984 text and in his later (1990) expansion of that book. Two examples based on hypothetical data in McLaughlin and Marascuilo's (1990) nursing research text will be used for illustrative purposes.

The Examples: In one of the appendices to their book, McLaughlin and Marascuilo (1990) provide an interesting set of data for a hypothetical study of the effect of preoperative teaching on postoperative recovery for a sample of 246 subjects. Included among their variables are three classic ordinal scales: (a) cigarette smoking behavior (never smoked, less than one pack per week, less than one pack per day but more than one pack per week, one or more packs per day); (b) alcohol consumption (never used alcohol, less than one drink per day, more than one drink per day); and (c) severity of illness (mild, moderate, severe). These data provide a basis for a secondary analysis of two research questions: "What is the relationship between cigarette smoking behavior and severity of illness?"; and "What is the relationship between alcohol consumption and severity of illness?" (The relationship between the two independent variables, cigarette smoking behavior and alcohol consumption, is of lesser interest.)

In order to answer the first of those questions, most "liberal" researchers would code the levels of the smoking variable 1, 2, 3, and 4 and the levels of the illness variable 1, 2, and 3; calculate Pearson r; and test it for statistical significance (r =.462; $p < .001$). They would do the same sort of thing for the second research question ($r = .073$; *NS*).

Most "conservative" researchers would display the data for the first question in a 4 x 3 contingency table and carry out a chi-square test of the difference between the corresponding obtained and expected frequencies, perhaps accompanied by a calculation of the associated contingency coefficient ($\chi^2 = 71.80$; $df = 6$; $p < .001$; $C = .475$). They would do likewise for the second question ($\chi^2 = 5.33$; $df = 4$; *NS*).

Both approaches are technically incorrect, for opposite reasons. The "liberal" approach endows the two ordinal scales with interval properties that they do not in fact possess. The "conservative" approach deflates the ordinal scales to nominal scales, since neither chi-square nor the contingency coefficient takes into account the inherent order of the row and column categories. (Moses, Emerson, and Hosseini [1984] take clinical investigators to task for their similarly inappropriate analyses of the differences between two groups.)

A more defensible analysis treats both variables as no more than, and no less than, what they are—ordinal. The following discussion is concerned with one such technique based on the work of Williams (1952), Marascuilo and Levin (1983), Goodman (1984), and Agresti (1984, 1990).

The Analysis: This technique involves a (no more than) four-step process:

1. Set up the i x j cross-classification frequency table (contingency table) for the two ordinal variables, where i is the number of categories for the variable that has the larger number of categories and j is the number of categories for the variable that has the smaller number of categories. (If both variables have the same number of categories, it doesn't matter which is chosen as the row variable and which is chosen as the column variable.)

2. Test that table for independence by either the traditional Pearson chi-square statistic or the likelihood-ratio chi-square statistic, G^2 (the two approaches usually yield very similar results). If chi-square is not significant for some prespecified alpha level, for (i–1)(j–1) degrees of freedom, you may be all done. The hypothesis that the variables are unrelated in the population cannot be rejected, and any measure of the degree of relationship between the variables in the sample is of little interest. (If chi-square "just misses" statistical significance and/or if some of the discrepancies between the obtained and expected frequencies are fairly large, especially in the corners of the table, you may want to go on to the next step. (See the example in Agresti [1990], pp. 267–269.)

3. If chi-square *is* significant, use Goodman's (1984) loglinear approach to test the hypothesis that in the population the uniform association model holds, for "scores" of 1 to i for the row variable and "scores" of 1 to j for the column variable. If the associated chi-square for that test is not significant, for (i–1)(j–1)–1 degrees of freedom, the equally-spaced scores are reasonable choices. Various functions of the local odds ratio (including a Pearson *r* approximation), and their corresponding confidence intervals, can be used to estimate the degree of relationship between the variables in the population from which the sample has been drawn (see Agresti, 1990).

4. If *that* (second) chi-square *is* significant, the uniform association model (equally-spaced scores) does not fit the data very well, and the scores to be assigned to the categories of the two ordinal variables need to be estimated from the data, using a procedure that is similar to general-linear-model canonical correlation analysis. Williams (1952) provided the necessary formulas for doing so, and they have been incorporated in Goodman's work. (See also Marascuilo and Levin [1983], pp. 451–460.)

Applications to the Two Hypothetical Examples: The data in McLaughlin and Marascuilo (1990) for testing the relationship between cigarette smoking behavior and severity of illness are:

	Illness		
Smoking	mild	moderate	severe
never	67	34	7
< pack/week	33	6	11
< pack/day	20	19	13
> pack/day	2	10	24

For these data the Pearson chi-square is 71.80, for $df = 6$. The likelihood-ratio chi-square is 74.96, also for $df = 6$. Both are well beyond the tabled values for the .05, .01, and .001 traditional significance levels (12.59, 16.81, and 22.46, respectively). The independence model therefore does not fit the data very well and the null hypothesis of no relationship in the population is rejected.

But what is the form of the relationship? A test of the goodness-of-fit of Goodman's (1984) uniform association model (see Appendix A for an SPSSX program that can be used to carry out that test) yields a Pearson chi-square of 18.07, for $df = 5$, which is significant at .05 and .01 (but not at .001). Therefore, that model does not fit the data well either, so assigning scores of 1 to 4 for Smoking and 1 to 3 for Illness is not a very good idea.

Williams' (1952) approach involves the calculation of the eigenvalues and eigenvectors of a matrix whose elements are functions of the frequencies in the 4 x 3 cross-tab (see Appendix B). The scores for the Smoking variable that best fit the data are –.74, –.27, .27, and 2.23 (or any nonzero multiplicative transformation thereof). The scores for the Illness variable are –.80, .01, and 1.75 (or, again, any nonzero multiplicative transformation thereof).

The data in McLaughlin and Marascuilo (1990) for testing the relationship between alcohol consumption and severity of illness are:

	Illness		
Alcohol	mild	moderate	severe
never used	66	30	31
< drink/day	31	16	14
> drink/day	25	23	10

For these data the test of independence yields a Pearson chi-square of 5.33, for $df = 4$, which is not statistically significant for any of the usual levels (the tabled value for the .05 level is 9.49), and none of the discrepancies between obtained and expected frequencies are very large. This is sufficient evidence to conclude that the independence model fits the data reasonably well and the null hypothesis of no relationship between Alcohol and Illness cannot be rejected. There is, therefore, little reason to pursue the data any further.

Summary: To investigate the relationship between two ordinal variables, one first tests their independence. If the independence model fits well (nonsignificant chi-square), the variables are said to be unrelated. If the independence model does not fit well (significant chi-square), one then tests for uniform association. If that model fits well (nonsignificant chi-square), the relationship is said to be approximately linear, scores of 1 to i and 1 to j are assigned to the categories of the row and column variables, respectively, and various measures of the direction and the strength of the relationship can be pursued. If that model does *not* fit well (significant chi-square), one uses Williams' method to estimate the scores that should be assigned to the categories of the two variables so that the correlation between them is as high as possible.

APPENDIX A

Goodman's test of uniform association can be carried out by using the loglinear routines in SPSSX or any of the major statistical packages. (You can use BMDP to get a graphical representation of the data through correspondence analysis [Greenacre, 1984].) For SPSSX the following commands are required (the variable names, data, etc. are those of the smoking/illness example in the McLaughlin and Marascuilo text):

```
BEGIN DATA
1 1   67
1 2   34
1 3    7
2 1   33
2 2    6
2 3   11
3 1   20
3 2   19
3 3   13
4 1    2
4 2   10
4 3   24
END  DATA
DATA LIST LIST/SMOKING ILLNESS COUNT*
VALUE LABELS SMOKING 1 'NEVER'
   2 '<PACK/WEEK'
   3 '<PACK/DAY' 4 '>PACK/DAY'/
   ILLNESS 1 'MILD' 2 'MODERATE' 3 'SEVERE'
WEIGHT BY COUNT
COMPUTE UV = SMOKING*ILLNESS
COMPUTE U = SMOKING
LOGLINEAR SMOKING(1,4)
   ILLNESS (1,3) WITH UV/
   DESIGN = SMOKING ILLNESS UV/
   PRINT = ESTIM/
FINISH
```

The expected counts on the first full page of output are the expected frequencies, *given that the uniform association model holds.* They are compared to the observed counts, and both the likelihood ratio chi-square and the Pearson chi-square are calculated and provided at the bottom of that same page. There is a "loss" of one additional degree of freedom for estimating those expected counts from the sample data.

APPENDIX B

To apply Williams' procedure for estimating the "scores" that should be assigned to the row and column categories, you must first create a matrix M of the same size as the contingency table, whose elements are the table entries divided by the square root of the product of the corresponding row and column totals. For example, the element in the first row, first column of the M matrix for the Smoking and Illness data is equal to 67 divided by the square root of 122 times 108, i.e., .5837.

Next, a vector V of i elements is created by taking the square root of the quotient of each row total and the sample size. For the Smoking and Illness data, the first element is equal to the square root of 122/246, i.e., .7042. The remaining elements are determined in like manner.

Next, the matrix M is multiplied by its transpose and the eigenvalues and eigenvectors of that matrix product are determined. The *second largest* eigenvalue is the square of the maximum correlation between the two variables. (The largest eigenvalue will always be equal, within rounding error, to 1.) For the Smoking and Illness data, that correlation, R, is .5058.

Finally, the scores for the row categories are determined by dividing the elements of the eigenvector associated with the second largest eigenvalue by the corresponding elements of the vector V defined above (they are –.7449, –.2741, .2686, and 2.2275 for the Smoking and Illness data); and the scores for the column categories are determined by dividing each row score by R, multiplying those quotients by the corresponding cell frequencies in a given column, summing, and dividing by the column total. For example, the score for the first column category for the Smoking and Illness data is obtained by dividing the –.7749, the –.2741, the .2686, and the 2.2275 by .5058, multiplying *those* by 67, 33, 20, and 2, respectively, summing the four products and dividing that sum by 122. The result is –.7961. The other two column scores are obtained in like manner.

As far as computer packages for Williams' method are concerned, MINITAB can be easily programmed to do all of the calculations, since it has built-in matrix operations. Some of the other packages could be "fooled" into getting the necessary eigenvalues and eigenvectors the same way they get eigenvalues and eigenvectors for factor analyses, discriminant analyses, or canonical correlation analyses. **NR**

Accepted for publication January 2, 1992.

THOMAS R. KNAPP, EdD, is professor of nursing and education, The Ohio State University, Columbus, OH.

References

AGRESTI, A. (1984). *Analysis of ordinal categorical data.* New York: Wiley.

AGRESTI, A. (1990). *Categorical data analysis.* New York: Wiley.

GOODMAN, L. (1984). *The analysis of cross-classified data having ordered categories.* Cambridge: Harvard University Press.

GREENACRE, M. J. (1984). *Theory and applications of correspondence analysis.* New York: Academic Press.

KNAPP, T. R. (1990). Treating ordinal scales as interval scales: An attempt to resolve the controversy. *Nursing Research, 39,* 121–123.

MARASCUILO, L. A., & LEVIN, J. R. (1983). *Multivariate statistics in the social sciences: a researcher's guide.* Monterey, CA: Brooks/Cole.

MCLAUGHLIN, F. E., & MARASCUILO, L. A. (1990). *Advanced nursing and health care research: Quantification approaches.* Philadelphia: Saunders.

MOSES, L. E., EMERSON, J. D., & HOSSEINI, H. (1984). Analyzing data from ordered categories. *New England Journal of Medicine, 311,* 442–448.

WILLIAMS, E. J. (1952). Use of scores for the analysis of association in contingency tables. *Biometrika, 39,* 274–289.