

BASIC programs for computation of the Goodman-Kruskal gamma coefficient

THOMAS O. NELSON

University of Washington, Seattle, Washington

In this article, I provide programs in the BASIC language for computing the Goodman-Kruskal gamma coefficient in three situations: (1) when the data consist of two scores for each of N individual persons/items, as in a correlational situation; (2) when the data are arrayed in an ordered 2×2 contingency table, as in a cross-classification situation; or (3) when the data consist of two conditional probabilities (e.g., conditional probability of saying "old" given that the actual state is *old*, and conditional probability of saying "old" given that the actual state is *new*), as in an absolute-judgment or detection situation.

For many situations, Goodman and Kruskal's (1954) gamma coefficient, G , has several advantages over other measures of correlation, over other measures of association, and over other measures of detection accuracy (for discussions, see Freeman, in press; Nelson, 1984, in press; Reynolds, 1977). G is being used increasingly often in psychological research (e.g., Linehan, Chiles, Devine, Laffaw, & Egan, in press; Metcalfe, 1986; Nelson, Leonesio, Landwehr, & Narens, 1986; Nelson, McSpadden, Fromme, & Marlatt, in press; Shimamura & Squire, in press; Vesonder & Voss, 1985). Although G can be computed via the SPSS statistical package (2nd ed., pp. 228-230), researchers using microcomputers may prefer programs written in BASIC. The present article is a tutorial that provides several programs in Applesoft BASIC; generalizations to other versions of BASIC should be straightforward. For people who want additional information about G , relevant references are mentioned.

CORRELATIONAL SITUATIONS

For many correlational situations, G is preferable to Pearson r for any of three reasons. First, Pearson r requires an interval scale for each of the variables being correlated (i.e., Pearson r is a measure of the degree of linear relationship); however, G is appropriate when either of the variables is conceptualized as being on an ordinal scale, such as rating scales (see Surber, 1984) or scales that are monotonically (but not necessarily linearly) related to a theoretical construct of interest (Townsend & Ashby, 1984). Second, if there are ties on one of the variables, then the maximum possible value of Pearson r (or Spearman rho) is less than 1.0, regardless of the degree of relationship between the two variables on the nontied items; this may be undesirable in many situations (e.g., if an M -place Likert rating scale is used to evaluate N items, with

$M < N$, then ties are inevitable). The researcher might prefer a measure of correlation that can still attain a maximum of 1.0 when there is a perfect relationship among the nontied items. Third, the interpretation of Pearson r is typically in terms of the percentage of variance in one variable that is accounted for by the other variable (defined by r^2). However, this interpretation may be problematic, both because the choice of r versus r^2 is currently controversial (Ozer, 1985) and because when the variables are not interval-scaled then the concept of variance is meaningless (see Townsend & Ashby, 1984).

Computation of G for Correlational Situations

Suppose there is a $2 \times N$ data array with two entries for each of N individuals (e.g., the entries might be the individual's scores on two dependent variables, with any numerical scores allowable). Then the BASIC program for G is:

```
130 INPUT "HOW MANY ROWS ARE IN THE  
TWO-COLUMN ARRAY?"; N  
140 DIM A(N)  
150 DIM B(N)  
160 REM ENTER DATA FOR THE 2 x N ARRAY  
170 FOR I = 1 TO N  
180 PRINT "ENTER THE TWO ENTRIES FOR A  
GIVEN ROW (INDIVIDUAL)"  
190 INPUT "FIRST ENTRY IS: "; A(I)  
200 INPUT "SECOND ENTRY IS: "; B(I)  
210 NEXT I  
220 REM COMPUTATION OF GAMMA  
230 R = 0  
240 W = 0  
250 FOR I = 1 TO N  
260 FOR J = 1 TO N  
270 IF A(I) > A(J) AND B(I) > B(J) THEN R =  
R + 1
```

Preparation of this article was partially supported by NIMH Grant No. 32205. Correspondence should be addressed to Thomas O. Nelson, Psychology (NI-25), University of Washington, Seattle, WA 98195.

```

280 IF A(I) > A(J) AND B(I) < B(J) THEN W =
W + 1
290 NEXT J
300 NEXT I
310 GAMMA = (R - W) / (R + W)
320 PRINT "GAMMA = "; GAMMA
330 REM IF OUTPUT IS "?DIVISION BY ZERO
ERROR IN 310" THEN GAMMA IS INDETER-
MINATE BECAUSE OF NO NON-TIED DYADS
340 END

```

Equations for determining the statistical significance and/or confidence interval for the obtained value of G are given in Reynolds (1977, section 3.4), as are equations for computing partial gamma correlations (section 4.3.2).

CROSS CLASSIFICATION

G can be computed as a measure of association for any ordered $R \times C$ cross-classification table having $R \geq 2$ and $C \geq 2$. For instance, consider the 2×2 array shown in Table 1. The observed frequencies of the four kinds of joint events are designated by a , b , c , and d .

Computation of G for a 2×2 Table

The BASIC program is:

```

400 INPUT "FREQUENCY OF ROW = 1 AND
COLUMN = 1 IS: "; A
410 INPUT "FREQUENCY OF ROW = 1 AND
COLUMN = 0 IS: "; B
420 INPUT "FREQUENCY OF ROW = 0 AND
COLUMN = IS: "; C
430 INPUT "FREQUENCY OF ROW = 0 AND
COLUMN = 0 IS: "; D
440 GAMMA = ((A * D) - (B * C)) / ((A * D) +
(B * C))
450 PRINT "GAMMA = "; GAMMA
460 END

```

The variance of the obtained G , given that all cell entries are greater than zero (Reynolds, 1977, Eq. 2.10), is equal to:

$$.25(1-G^2)[(1/a) + (1/b) + (1/c) + (1/d)].$$

Equations for computing G for $R \times C$ tables where $R > 2$ and/or $C > 2$ are available (e.g., Goodman & Kruskal,

1954; Mueller, Schuessler, & Costner, 1970). Also, a table of significance for G for a 2×2 table, along with a computer program for determining the significance of G for $R \times C$ tables where $R > 2$ and/or $C > 2$, is given in Krinsky and Bronn (1986). Finally, the present computer program gives the same result as the SPSS program; this program slightly underestimates the population value of gamma, and a correction is available in Nelson (1984, Eq. 14).

G AS A MEASURE OF DETECTION ACCURACY

Nelson (in press) described situations for which G was recommended over the traditional measures from signal detection theory, such as d' . For instance, suppose that the data from an observer were arrayed in Table 1 in the following way: The row variable indicates whether the trial actually did or did not contain a signal (1 or 0, respectively), and the column variable indicates whether the individual said "yes" or "no" (1 or 0, respectively). Then in terms of the cell frequencies in Table 1, the probability of a hit, h is equal to

$$a/(a+b),$$

and the probability of a false alarm, f , is equal to

$$c/(c+d).$$

Then (see Nelson, 1984, Eq. 8) G can be computed from h and f via the following program:

```

500 INPUT "THE PROBABILITY OF A HIT IS: "; H
510 INPUT "THE PROBABILITY OF A FALSE
ALARM IS: "; F
520 GAMMA = (H - F) / (H + F - (2 * H * F))
530 PRINT "GAMMA = "; GAMMA
540 END

```

For detection situations in which multiple ROC points are available, G can be computed by using the program for correlations given above. A probabilistic interpretation of G as a measure of detection accuracy is given in Nelson (in press); alternative interpretations of G are given in Reynolds (1977, pp. 74 and 84). These interpretations can also be applied to the aforementioned correlational and cross-classification situations (e.g., Nelson, Leone-sio, Landwehr, & Narens, 1986).

REFERENCES

- FREEMAN, L. (in press). Order-based statistics and monotonicity: Measuring ordinal association. *Journal of Mathematical Sociology*.
- GOODMAN, L., & KRUSKAL, W. (1954). Measures of association for cross classifications. *Journal of the American Statistical Association*, *49*, 732-764.
- KRINSKY, R., & BRONN, S. (1986). *The Goodman-Kruskal gamma correlation table of statistical significance* (Tech. Rep.). Psychology Department, University of Southern Colorado. Manuscript under review.
- LINEHAN, M., CHILES, J., DEVINE, R., LAFFAW, J., & EGAN, K. (in

Table 1
Example of a 2×2 Table

Row Variable	Column Variable		
	1	0	Σ
1	a	b	a+b
0	c	d	c+d

Note—Cell entries (a , b , c , d) are frequencies of the joint events defined by the row and column variables. The total number of observations is equal to $a+b+c+d$.

- press). Interpersonal problems of parasuicides versus suicide ideators and non-suicidal psychiatric patients. *Journal of Consulting & Clinical Psychology*.
- METCALFE, J. (1986). Feeling of knowing in memory and problem solving. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **12**, 288-295.
- MUELLER, J., SCHUESSLER, K., & COSTNER, H. (1970). *Statistical reasoning in sociology*. New York: Houghton Mifflin.
- NELSON, T. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin*, **95**, 109-133.
- NELSON, T. (in press). ROC curves and measures of discrimination accuracy: A reply to Swets. *Psychological Bulletin*.
- NELSON, T., LEONESIO, R., LANDWEHR, R., & NARENS, L. (1986). A comparison of three predictors of an individual's memory performance: The individual's feeling of knowing vs. the normative feeling of knowing vs. base-rate item difficulty. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **12**, 279-287.
- NELSON, T., MCSPADDEN, M., FROMME, K., & MARLATT, B. (in press). Effects of alcohol intoxication on metamemory and on retrieval from long-term memory. *Journal of Experimental Psychology: General*.
- OZER, D. (1985). Correlation and the coefficient of determination. *Psychological Bulletin*, **97**, 307-315.
- REYNOLDS, H. T. (1977). *The analysis of cross-classifications*. New York: Macmillan.
- SHIMAMURA, A., & SQUIRE, L. (in press). Memory and metamemory: A study of the feeling-of-knowing phenomenon in amnesic patients. *Journal of Experimental Psychology: Learning, Memory, & Cognition*.
- SURBER, C. (1984). Issues in using quantitative rating scales in developmental research. *Psychological Bulletin*, **95**, 226-246.
- TOWNSEND, J., & ASHBY, F. (1984). Measurement scales and statistics: The misconception misconceived. *Psychological Bulletin*, **96**, 394-401.
- VESONDER, G., & VOSS, J. (1985). On the ability to predict one's own responses while learning. *Journal of Memory & Language*, **24**, 363-376.

(Manuscript received for publication April 14, 1986.)