Sociology 405/805
January 20, 2004

Goodman and Kruskal tau (τ)

These notes give an example of how to calculate the **Goodman and Kruskal tau** measure of association. The example used is the same as that used for calculating λ in the notes of January 14, 2004 – predicting provincial political preference from knowledge of federal political preference.

The Goodman and Kruskal tau has symbol τ, the Greek letter tau. τ is obtained by using the principle of proportional reduction in error. Like λ, this is a directional measure of assocation, with different values depending on whether rows or columns of a cross-classification table are dependent. τ has limits of zero (no association) and 1 (complete or perfect association). It can be obtained using the crosstabulation procedure of SPSS – when requesting statistics for a crosstabulation, check λ and the output prints out the values of both λ and τ, along with their statistical significance.

The most straightforward explanation I have found for τ is on pages 41-45 of H. T. Reynolds, *Analysis of Nominal Data*, Sage Quantitative Applications in the Social Sciences, number 7.

The data for this example is in Table 1. The explanation of how to calculate τ follows the table.

τ for rows dependent

The principle underlying τ is to determine the proportional reduction in error by using two rules. The rules are a little more complex than in the case of λ in that the aim is to "preserve the original distribution" (Reynolds, p. 41). The method is to predict the actual frequencies in the table for each value of the dependent variable.

Rule 1 is to determine the number of errors of prediction without using the information on the independent variable. This results in E_1, the number of errors of prediction for each value of the dependent variable, using only the row totals. For Rule 2, the information about the independent variable (in each column) is used to predict the frequencies for each value of the dependent variable. This results in a second set of error of prediction, E_2.

Table 1: Cross Classification of Provincial and Federal Political Preference

Provincial Political	Federal Political Preference (FV)				
Preference (PV)	Liberal	NDP	Conservative	None	Total
Liberal	85	2	8	1	96
NDP	63	83	12	6	164
Conservative	28	7	56	3	94
None	12	1	7	141	161
Total	188	93	83	151	515

The proportionate reduction of errors, defined as τ, is

$$\tau = \frac{E_1 - E_2}{E_1}$$

Formula

If R_i is the row total for row i of the table and n is the total number of cases in the table, the number of errors using Rule 1 is

$$E_1 = \sum_i \left(\frac{n - R_i}{n} \times R_i \right).$$

Then for each column j, use the same procedure to and add all the errors and term this E_{2j}. That is, for column j, the number of errors of prediction is:

$$E_{2j} = \sum_j \left(\frac{C_j - O_{ij}}{C_j} \times O_{ij} \right).$$

where Cj is the column total for column j and O_{ij} is the number of cases in row i and column j. The sum of these E_{2j} across all j columns is

$$E_2 = \sum_j E_{2j}.$$

The value of τ is the proportional reduction in error:

$$\tau = \frac{E_1 - E_2}{E_1}$$

Example – τ for provincial political preference (PV) as dependent variable and federal political preference (FV) as independent variable

A. Rule 1

For Liberal, row 1, the number of errors of prediction is

$$\frac{515 - 96}{515} \times 96 = 78.1$$

For NDP, row 2, the number of errors of prediction is

$$\frac{515 - 164}{515} \times 164 = 111.8$$

For Conservative, row 3, the number of errors of prediction is

$$\frac{515 - 94}{515} \times 94 = 76.8$$

For None, row 4, the number of errors of prediction is

$$\frac{515 - 161}{515} \times 161 = 110.7$$

The total number of errors of prediction for PV is the sum of the errors for the four rows,

$$E_1 = 78.1 + 111.8 + 76.8 + 110.7 = 377.4$$

B. Rule 2

Beginning with column 1, support for the Liberal party at the federal level, the errors of prediction for the four rows are:

$$\frac{188 - 85}{188} \times 85 = 46.6$$

$$\frac{188-63}{188} \times 63 = 41.9$$
$$\frac{188-28}{188} \times 28 = 23.8$$
$$\frac{188-12}{188} \times 12 = 11.2$$

Total number of Rule 2 errors for the first column is

$$E_{21} = 46.6 + 41.9 + 23.8 + 11.2 = 123.5.$$

For column 2, support for the NDP at the federal level, the errors of prediction for the four rows are:

$$\frac{93-2}{93} \times 2 = 2.0$$
$$\frac{93-83}{93} \times 83 = 8.9$$
$$\frac{93-7}{93} \times 7 = 6.5$$
$$\frac{93-1}{93} \times 1 = 1.0$$

Total number of Rule 2 errors for the second column is

$$E_{22} = 2.0 + 8.9 + 6.5 + 1.0 = 18.4.$$

For column 3, support for Conservative parties at the federal level, the errors of prediction for the four rows are:

$$\frac{83-8}{83} \times 8 = 7.2$$
$$\frac{83-12}{83} \times 12 = 10.3$$
$$\frac{83-56}{83} \times 56 = 18.2$$
$$\frac{83-7}{83} \times 7 = 6.4$$

Total number of Rule 2 errors for the third column is

$$E_{23} = 7.2 + 10.3 + 18.2 + 6.4 = 42.1.$$

For column 4, support for none of the parties at the federal level, the errors of prediction for the four rows are:

$$\frac{151 - 1}{151} \times 1 = 1.0$$

$$\frac{151 - 6}{151} \times 6 = 5.8$$

$$\frac{151 - 3}{151} \times 3 = 2.9$$

$$\frac{151 - 141}{151} \times 141 = 9.3$$

Total number of Rule 2 errors for the fourth column is

$$E_{24} = 1.0 + 5.8 + 2.9 + 9.3 = 19.0.$$

For Rule 2, the total number of errors of prediction for PV is the sum of the above errors:

$$E_2 = E_{21} + E_{22} + E_{23} + E_{24} = 123.5 + 18.4 + 42.1 + 19.0 = 203.0$$

C. Calculation of τ

$$\tau = \frac{E_1 - E_2}{E_1} = \frac{377.4 - 203.0}{377.4} = \frac{174.4}{377.4} = 0.462$$

Last edited on January 23, 2004.