Sociology 405/805
January 15, 2004

Example of lambda (λ) as a measure of association

Lambda is a measure of association using the principle of proportional reduction in error. An example of λ and the formula for computing it is contained in section 11.3 of *Introductory Statistics for the Social Sciences*. Here a similar example is used – association between federal and provincial political preference. The data for this example come from the *ssae98.sav* data set and refer to the political preferences of undergraduates at the University of Regina in Fall 1998. The two variables used are *provote* and *fedvote* but, for this example, (a) the uncertain and minor parties have been eliminated and (b) the Reform, Alliance, Saskatchewan, and Progressive Conservative parties have been amalgamated into a category termed "Conservative." Those respondents who said they favoured no political party are retained in this example. The resulting cross-classification of provincial by federal political preference is given in Table 1. The explanation of how to obtain λ follows the table.

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

A. λ **for rows dependent – predicting provincial political preference (PV)**

First, predict PV without examining its relationship with FV, the column variable. The number of errors made in predicting PV when the column totals are ignored is E_{TR}. This is the total number of cases in the table minus the number of cases in the row with the largest frequency of occurrence. That is,

$$E_{TR} = \text{Grand total} - \text{Maximum row total.}$$

Now conduct a similar procedure for each column, or value of FV, using the information about the relationship between FV and PV. For each column j, where $j = 1, 2, \ldots, c$, let the number of errors of prediction be E_{C_j}. For each column of the table,

$$E_{C_j} = \text{Column } j \text{ total} - \text{Maximum count in column } j.$$

Let E_C be the sum of the errors of prediction, E_{C_j}. That is, E_C is the total number of errors of prediction when the columns are used to predict the row results, so

$$E_C = \sum_{j=1}^{c} E_{C_j}.$$

The proportional reduction in error by using the column information is

$$\lambda_R = \frac{E_{TR} - E_C}{E_{TR}}$$

where λ_R is the value of λ when the rows are being predicted.

For this example, the procedure is as follows.

1. **Predict provincial political preference (PV) without knowing anything about federal political preference (FV)**.

   From the row totals, the best guess is that PV will be NDP since more respondents favour NDP than Liberal, Conservative, or None. This prediction will be correct for 164 cases and incorrect for the other $515 - 164 = 351$ cases. Using only the row totals, the overall number of errors of prediction is

   $$E_{TR} = 515 - 164 = 351.$$

2. **Can the predictions of PV be improved if we know which federal political party is favoured by each respondent?** That is, how many errors of prediction will there be if we use federal political preference to predict provincial political preference? In order to calculate this, use each value of FV and, from this, obtain predictions of PV. This involves going through the table column by column, as follows.

   (a) Column 1 – Favour Liberal at federal level. Since more of these favour the Liberal party provincially than any other party, the best prediction is that a respondent in the Liberal column will vote Liberal at the provincial level. Of these 188 federal Liberal supporters, 85 favour Liberals provincially, so no error is made for each of these. But for the other $188 - 85 = 103$ cases an error of prediction is made. For column 1, the number of errors of prediction, using our best method, is

   $$E_{C_1} = 188 - 85 = 103.$$

   (b) Column 2 – Favour NDP at federal level. Of these, more favour the NDP provincially than favour any other party, so the best prediction is to predict that these respondents favour the NDP provincially. Of the 93 federal NDP supporters, 83 favour NDP provincially, so no error is made for each of these. But for the other $93 - 83 = 10$ cases an error of prediction is made. For column 2, the number of errors of prediction, using this method, is

   $$E_{C_2} = 93 - 83 = 10.$$

   (c) Column 3 – Favour a conservative party at federal level. Again, the best prediction is that each of these will favour a conservative party at provincial level, since more of these say they favour a conservative party provincially than any other possibility. Of the 83 federal conservatives, 56 favour a conservative party provincially, so no error is made for each of these. But for the other $83 - 56 = 27$ cases an error of prediction is made. For column 3, the number of errors of prediction is

   $$E_{C_3} = 83 - 56 = 27.$$

(d) Column 4 – Favour no party at federal level. Since more of these say they favour no party provincially than any other possibility, the best prediction for these respondents is that they will favour no party provincially. Of the 151 respondents who favour no party federally, 141 favour no party provincially, so no error is made for each of these. But for the other $151 - 141 = 10$ cases an error of prediction is made. For column 4, the number of errors of prediction is

$$E_{C_4} = 151 - 141 = 10.$$

3. Number of errors of prediction using the information on federal political preference is

$$E_C = \sum_{j=4}^{c} E_{C_j} = 103 + 10 + 27 + 10 = 150.$$

4. Reduction in error. The total number of errors made in predicting provincial political preference using no information about federal political preference is

$$E_{TR} = 351.$$

The number of errors made in predicting provincial political preference using information about federal political preference is

$$E_C = 150.$$

The reduction in the number of errors is

$$E_{TR} - E_C = 351 - 150 = 201$$

and as a proportion of the original number of errors, this is $201/351 = 0.573$. That is,

$$\lambda_C = \frac{E_{TC} - E_R}{E_{TC}} = \frac{351 - 150}{351} = \frac{201}{351} = 0.573.$$

B. λ for column dependent – predicting federal political preference (FV)

$$E_{TC} = \text{Grand total} - \text{Maximum column total.}$$

For each of the rows of the table,

$$E_{R_i} = \text{Row } i \text{ total} - \text{Maximum count in row } i.$$

If there are r rows in the table,

$$E_R = \sum_{i=1}^{r} E_{R_i}.$$

$$\lambda_C = \frac{E_{TC} - E_R}{E_{TC}}$$

For Table 1, the calculations are as follows:

$$E_{TC} = 515 - 188 = 327$$

$$E_{R_1} = 96 - 85 = 11$$

$$E_{R_2} = 164 - 83 = 81$$

$$E_{R_3} = 94 - 56 = 38$$

$$E_{R_4} = 161 - 141 = 20$$

$$E_R = \sum_{i=1}^{4} E_{R_i} = 11 + 81 + 38 + 20 = 150.$$

$$\lambda_C = \frac{E_{TC} - E_C}{E_{TC}} = \frac{327 - 150}{327} = \frac{177}{327} = 0.541.$$

C. Other measures of association for this table

$$\chi^2 = 659.268 \text{ with 9 d.f. and significance less than 0.000.}$$

$$\phi = 1.131$$

$$\text{Cramer's } V = 0.653.$$

$$C = 0.749.$$