MEASURES OF ASSOCIATION FOR MULTIVARIATE CONTINGENCY TABLES

James M. Sakoda, Brown University

Need for a Measure of Association

In analyzing multidimensional contingency tables the goodness of fit of various models is generally tested via Pearson or likelihood ratio chi square. The acceptance or rejection of a model on the basis of a significance test alone runs the risk of allowing the number of cases to determine, at least in part, the number of parameters deemed to be significant. As in other test situations, judgment of the existence of a relationship should be dependent on the strength of the relationship as well as its statistical significance. If a sizable relationship is indicated the acceptable significance level might be raised to .10, say, to avoid rejecting a potentially meaningful source of variation. Conversely, effects which are extremely small, even though statistically significant, might be eliminated from a model. Measures of association are also useful in comparing tables with different numbers of cases.

Several measures of association for contingency tables have been developed for two-way tables, and one of the problems is to select an appropriate one for use with higher dimensions. We conclude that $\underline{X}^2$ divided by the maximum $\underline{X}^2$ for a table serves as a suitable basis for a measure of association. A second problem is the application of the chosen measure of association to higher dimensioned tables. In some situations multidimensional tables can be related to two-way tables. In the multiple correlation type of situation a dependent variable can be related to a combination of categories of independent variables by using a two-way table. In the partial correlation situation, two-way tables can be averaged over a set of control variables. For higher order effects, such as a three-way or four-way effect or for combinations of effects, reduction to two-way tables is not possible. The task is then to find the maximum $\underline{X}^2$ for higher order effects so that they can be compared with the obtained $\underline{X}^2$ for a given effect. Goodman (1971) suggested using a proportional reduction in $\underline{X}^2$ as a method of calculating multiple or partial correlation coefficients. The approach suggested here differs from his in that higher order effects are analyzed in terms of the maximum $\underline{X}^2$ rather than an arbitrarily-selected empirical $\underline{X}^2$.

The Choice Among Measures of Association

For analysis of multidimensional tables the most convenient measures of association are those based on $\underline{X}^2$, since data analysis is performed using $\underline{X}^2$. It is possible to partition higher order $\underline{X}^2$ into their component parts and to relate $\underline{X}^2$ for two-way tables to higher order ones. Goodman (1971), for example, suggests using a proportional reduction in $\underline{X}^2$ as a method of calculating multiple and partial correlation coefficients. $\underline{X}^2$ is suitable with either ordered or unordered categories. Measures of association requiring ordered categories, such as Kendall's <u>tau</u>, Somer's <u>D</u> and Goodman and Kruskal's <u>gamma</u> are too specialized for routine contingency table analysis, since they apply only to certain tables. Moreover, <u>gamma</u> and its 2 x 2 table version, Yule's <u>Q</u>, have tendencies to be high in comparison to other coefficients when marginals are distributed unevenly.

Another advantage of $\underline{X}^2$-based measures of association is their symmetric nature, requiring only a single measure regardless of the direction of relationship or prediction. There are a number of asymmetric measures of association which are developed on different bases and which are meaningful in different ways. These are Goodman and Kruskal's proportional reduction of errors of prediction measures <u>lambda</u> and <u>tau</u>, Margolin and Light's analysis of variance measure BSS/TSS of proportion of row variation attributable to column variation, and the proportional reduction of uncertainty measure based on information theory. <u>Lambda</u> cannot be recommended for tables with uneven marginals since a zero coefficient results when the largest frequencies in each column fall in one row and other measures of association show a relationship. This fault is not shared by <u>tau</u>, even though it is also a measure developed on the principle of proportional reduction in errors of prediction. On the other hand, <u>tau</u> is numerically identical to Margolin and Light's (1974) BSS/TSS measure, showing that proportional reduction in error can be quite similar to proportion of explained variation. According to Bishop, Fienberg and Holland (1975: 391), BSS/TSS and <u>tau</u> involve a Pearson $\underline{X}^2$-like expression, and when the row sums are equal, they are equal to ϕ^2 / (I-1) and hence to $\underline{X}^2$ / N(I-1). N(I-1) is maximum $\underline{X}^2$ when I ≤ J. The relative reduction in uncertainty measure utilizes likelihood statistics to express uncertainty, which is "variance-like" (Hays, 1973). These measures, except for their asymmetric nature (some have symmetric versions), have a great deal in common both meaningfully and numerically, with Cramér's $\underline{V}^2$, which represents $\underline{X}^2$ divided by maximum $\underline{X}^2$. The interpretation of these measures is not any easier than the interpretation of $\underline{X}^2$-based measures, as is sometimes claimed. In fact, these measures produce very small coefficients generally in comparison with measures such as $\underline{V}$ or the contingency coefficient which resemble the Pearson $\underline{r}$ rather than $\underline{r}^2$ as these measures do.

One of the oldest $\underline{X}^2$-based measures of association is Karl Pearson's mean square contingency or contingency coefficient:

$$\underline{C} = \sqrt{\frac{\phi^2}{1 + \phi^2}} = \sqrt{\frac{\underline{X}^2}{\underline{X}^2 + \underline{N}}},$$

where ϕ^2 is estimated by $\underline{X}^2$ / $\underline{N}$. "Karl Pearson showed that, if the items are capable of interpretation as a quantitatively ordered series, if the distributions are normal, and if the regression is rectilinear, $\underline{C}$ becomes identical with $\underline{r}$ as the number of categories is indefinitely increased." (Peters and vanVoorhis, 1940: 392). In other words, $\underline{C}$ is an estimate of the Pearson $\underline{r}$, but can be applied even when categories are unordered and the relationships are not linear. Its shortcoming is that its maximum value does not reach unity. But maximum ϕ^2 is the minimum of $\underline{I}$-1 or $\underline{J}$-1 and $\underline{C}_{max}$ can be calculated:

$$\underline{C}_{max} = \sqrt{\frac{\min(\underline{I}-1,\underline{J}-1)}{\min(\underline{I},\underline{J})}} .$$

It is possible to correct $\underline{C}$ to achieve unity by calculating $\underline{C}/\underline{C}_{max}$, although it is not a standard practice. Tschuprow proposed the use of

$$T = \sqrt{\frac{\phi^2}{\sqrt{(I-1)(J-1)}}}$$

but achieved a maximum of 1 only for square tables. Its use is therefore not recommended.

Cramér (1946) suggested norming ϕ^2 by dividing it by its maximum value:

$$\underline{V} = \sqrt{\frac{\phi^2}{\min(\underline{I}-1,\underline{J}-1)}} .$$

Estimating ϕ^2 by $\underline{X}^2/\underline{N}$,

$$\underline{V} = \sqrt{\frac{\underline{X}^2}{\underline{N}\min(\underline{I}-1,\underline{J}-1)}} .$$

The denominator term is maximum $\underline{X}^2$ for a two-way $\underline{I}$ x $\underline{J}$ table so that $\underline{V}^2$ can be given a proportion of maximum variation due to interaction interpretation. $\underline{V}$ has an acceptable interpretation via $\underline{V}^2$ and unlike $\underline{C}$ and $\underline{T}$ varies between 0 and 1. It is our choice as a suitable measure for application to higher order tables, with $\underline{C}/\underline{C}_{max}$ a second possibility.

Maximum X^2

According to Cramér (1946: 443), the maximum $\underline{V}^2$ of unity is obtained "when and only when each row (when $\underline{r} \le \underline{s}$) or each column (when $\underline{r} \ge \underline{s}$) contains one single element different from zero." An example of an arrangement of cell frequencies for a maximum $\underline{V}^2$ is shown in Fig. 1 for a 3 x 4 table with I < J. Cramér's condition can be expressed as

$$x_{ij} = x_{+j} .$$

30	-	-	-	30
-	15	5	-	20
-	-	-	10	10
30	15	5	10	60

Fig. 1. A 3 x 4 Table with Cramér's $\underline{V}$ = 1.0.

We start with the formula for $\underline{X}^2$:

$$\underline{X}^2 = \underline{N}\left(\Sigma\Sigma \frac{x_{ij}^2}{x_{i+}\,x_{+j}} - 1\right) .$$

With cancellation of x_{ij} and x_{+j},

$$\underline{X}^2_{max} = \underline{N}\left(\sum_{ij}\frac{x_{ij}}{x_{i+}} - 1\right) .$$

By definition $\Sigma x_{ij} = x_{i+}$. Therefore,

$$\underline{X}^2_{max} = \underline{N}\,(\Sigma 1-1) = \underline{N}\,(\underline{I}-1) ,$$

or when $\underline{J} \le \underline{I}$,

$$\underline{X}^2_{max} = \underline{N}\,(\underline{J}-1) .$$

Hence,

$$\underline{X}^2_{max} = \underline{N}\min(\underline{I}-1,\underline{J}-1) .$$

For example,

$$\underline{X}^2 = 60\left(\frac{30^2}{30\cdot30} + \frac{15^2}{20\cdot15} + \frac{5^2}{20\cdot5} + \frac{10^2}{10\cdot10} - 1\right)$$
$$= 60(\ 1 \quad + \quad 1 \quad + \quad 1 \quad -1)$$
$$= 60(3-1) = 120 .$$

Analogue of Multiple Correlation

The analogue to the multiple correlation, in which Variables 2, 3 and 4, for example, are related to Variable 1, the dependent variable, can be set up by means of a two-way table. The row variable is Variable 1 with $\underline{I}$ categories and the column variable consists of all possible combinations of categories of Variables 2, 3 and 4 with $\underline{J}$ x $\underline{K}$ x $\underline{L}$ categories in all. In terms of the loglinear model this table represents an independence model (1 x 234), which tests all effects of 2, 3 and 4 on 1: (12), (13), (14), (123), (124), (134), (1234). The model can be run on a program like ECTA by fitting 1 and 234. With $\underline{X}^2$ for the table available it is a simple matter to compute Cramér's $\underline{V}$ and it can be treated as an analogue to the multiple correlation coefficient:

$$\underline{V}_{1\cdot234} = \sqrt{\frac{\underline{X}^2(1\text{x}234)}{\underline{N}\min(I-1,JKL-1)}} .$$

In Fig. 2 is shown a 3 x 2 x 4 table arranged to give a maximum $\underline{X}^2$ for Variable 1 against 2 and 3 combined. In filling the cells the distinctions between categories of Variable 2 and 3 are ignored. Maximum $\underline{X}^2$ is equal to $\underline{N}$ x (3-1) or 240.

10	-	-	20	-	-	10	-	40
-	20	-	-	10	-	-	20	50
-	-	10	-	-	20	-	-	30
10	20	10	20	10	20	10	20	120

Fig. 2. A 3 x 2 x 4 Table Arranged for Maximum $\underline{X}^2$(1x23).

Partial Correlation

Partial correlation is generally defined as a measure of association between two variables holding constant the effects of a third variable. For continuous variables the effects of a third variable can be removed by taking the residualized score and correlating these. For discrete variables categories are generally unordered and this approach cannot be used. Instead, the alternative of setting up separate subtables for each level of the third variable is used (Agresti, 1977). For each two-way table a measure of association is calculated and these are weighted and averaged to obtain an overall measure of partial association. When using $\underline{X}^2$ it is necessary either to assume that higher order interactions do not exist or to remove the effects.

Given a three-way table, we set up $\underline{I}$ x $\underline{J}$ tables for relationships for Variables 1 and 2 for each level of Variable 3. For each table

$$\underline{V}^2_k = \frac{\underline{X}^2_k}{\underline{N}_k \min(\underline{I}-1,\underline{J}-1)} .$$

For an overall measure each $\underline{V}^2_k$ can be weighted by $\underline{N}_k/\underline{N}$, the proportion of the total number of cases in each subtable:

$$\underline{V}^2 = \sum_k \frac{\underline{N}_k}{\underline{N}} \frac{\underline{X}^2_k}{\underline{N}_k \min(\underline{I}-1,\underline{J}-1)} .$$

The $\underline{N}_k$'s cancel out and

$$\underline{V}^2 = \frac{\Sigma \underline{X}^2_k}{\underline{N} \min(\underline{I}-1,\underline{J}-1)} .$$

The numerator is the $\underline{X}^2$ for the partial association model, (1x2|3) and tests the effects (12) (123), and can be obtained by fitting (12), (13). The denominator term is the maximum $\underline{X}^2$ value for the partial correlation problem. From the numerator the higher order interaction must be removed, leaving only the (12) effect. $\underline{X}^2$(123) can be obtained by fitting (12), (13), (23). $\underline{X}^2$(1x2|3)-$\underline{X}^2$(123) leaves $\underline{X}^2$(12). Hence,

$$\underline{V}_{12\cdot 3} = \sqrt{\frac{\underline{X}^2(12)}{\underline{N} \min(\underline{I}-1,\underline{J}-1)}} .$$

The interpretation of $\underline{X}^2$(12) is the conditional test for (1x2|3), given no three-factor effect (Bishop, Fienberg and Holland, 1975: 171). When higher order interaction exists it is desirable to examine subtables individually.

In Fig. 3 is shown an analysis of the partial association of Preference x Use given Temperature and Softness in the Reis-Smith data. Higher order interactions are calculated by fitting (12)(134)(234). The interactions are significant at the .153 level and $\underline{V}$ of .089 and $\underline{C}/\underline{C}_{max}$ of .126 indicate the existence of an appreciable amount of higher order interaction.

Model	Fitted Parameters	$\underline{X}^2$	df	p	$\underline{V}$	$\underline{C}/\underline{C}_{max}$
(1x2|34)	(134)(234)	27.81	6	.000	.166	.232
(123)(124) (1234)	(12)(134) (234)	8.05	5	.153	.089	.126
(12)		9.76	1	.000	.140	.196

Fig. 3. Analysis of Partial Association for the Reis-Smith Data

For the individual subtables the $\underline{V}_{kl}$'s are

.122	.261	.225
.005	.108	.202

and they show how the partial $\underline{V}_{12\cdot 34}$ is only an average and cannot reflect the full range of variations among subtables.

Higher Order Interactions

To apply Cramér's $\underline{V}$ to higher order interactions it is necessary to find the maximum $\underline{X}^2$ corresponding to them. In Fig. 4 is shown a 3 x 2 x 2 table with frequencies arranged internally to obtain a maximum (123) interaction.

10	-	-	10	20
-	10	10	-	20
-	10	10	-	20
10	20	20	10	60

Fig. 4. A 3 x 2 x 2 Table with Maximum (123) Effect.

$\underline{X}^2$(123) obtained by fitting (12)(13)(23) is given by the ECTA computer program as 60 for the Pearson version and 76.38 for the likelihood ratio one. Evidently, the maximum $\underline{X}^2$ for three-way interaction is given by $\underline{N}$ times the minimum of the three, $\underline{I}$-1, $\underline{J}$-1, $\underline{K}$-1. Hence,

$$\underline{V}_{123} = \sqrt{\frac{\underline{X}^2(123)}{\underline{N} \min(\underline{I}-1,\underline{J}-1,\underline{K}-1)}} .$$

The formula for the maximum $\underline{X}^2$ applies to the Pearson $\underline{X}^2$ and only approximately to the likelihood ratio $\underline{X}^2$. Hence, it is prudent to use Pearson $\underline{X}^2$ for the numerator terms in calculating Cramér's $\underline{V}$.

In Fig. 5 an example of maximum three-way effect for a 3 x 3 x 4 table is shown. $\underline{X}^2$(123) for this table is 240 and the likelihood ratio $\underline{X}^2$ is 263.67. The 240 agrees with the formula:

$$\underline{X}^2_{max} = 120\ (3-1) = 240 .$$

10	-	-	-	-	10	-	10	-	-	10	-
-	10	-	10	-	-	-	-	10	-	-	10
-	-	10	-	10	-	10	-	-	10	-	-

Fig. 5. A 3 x 3 x 4 Table with Maximum (123) Effect.

By analogy maximum X^2 for four-way interaction can be calculated as

$$X^2_{max} = N \min(I-1, J-1, K-1, L-1) \ .$$

In Fig. 6 is shown an example of maximum four-way effect for a 3 x 2 x 2 x 2 table.

10	-	-	10	-	10	10	-
-	10	10	-	10	-	-	10
-	10	10	-	10	-	-	10

Fig. 6. A 3 x 2 x 2 x 2 Table with Maximum (1234) Effect.

Pearson X^2 for this table is 120, which agrees with the formula; likelihood ratio X^2 is 152.76. An indication that this is indeed the maximum value is shown by the fact that other separate effects are zero.

To set up a table for maximum X^2 for a three-way effect one can use a latin square with the number of treatment equal to the smallest dimension. Each treatment appears only once in each row and column. Each treatment is then set up as a separate table as in Fig. 4 or 5. If there are additional rows, columns or blocks, one of the rows, columns or blocks is arbitrarily duplicated. In Fig. 4 the last row is a replicate and in Fig. 5 the last block is. For maximum four-way effects not only the rows and columns are arranged in latin-square form, but also the tables themselves. For example, in Fig. 6 Table 1 is followed by Table 2 and then Table 2 by Table 1. This forms a latin square of the form 1, 2, 2, 1.

There is no reason why Cramér's V cannot be applied to models representing a combination of effects. It would seem reasonable that the maximum X^2 would be determined by the component with the highest maximum. For example, given a model representing $X^2(123) + X^2(124) + X^2(1234)$ for a 3x3x3x4 table, the largest maximum would be for $X^2(124)$. This is $N \times \min(I-1, J-1, L-1)$ or $N \times 2$.

The calculation of C/C_{max} is possible if X^2 and X^2_{max} are available. If variables are basically continuous in nature, although tabled as discrete categories, and if an estimate of the Pearson r is desired, C/C_{max} can be calculated.

Summary

For application to multidimensional contingency tables X^2-based measures of association are the most convenient. Of the available measures based on X^2 for the two-way table Cramér's V is the most appropriate. It is applicable to both ordered and unordered categories, it is a symmetric measure, and V^2 can be interpreted as X^2 divided by the maximum possible X^2. Maximum X^2, which is given as $N \min(I-1, J-1)$ is easy to calculate. Cramér's V can be applied to the multiple correlation situation and the analogue of the partial correlation coefficient. It can also be applied to the three-way, four-way and other higher order interactions, as well as to X^2 based on a combination of effects. A second candidate is C/C_{max}.

References

Agresti, Alan (1977). Considerations in Measuring Partial Association for Ordinal Categorical Data. Journal of the American Statistical Association, 72, 37-45.

Bishop, Yvonne M. M., Fienberg, Stephen E. and Holland, Paul W. (1975). Discrete Multivariate Analysis, Theory and Practice. Cambridge, Mass., The MIT Press.

Costner, Herbert (1965). Criteria for Measures of Association. American Sociological Review, 30, 341-353.

Cramér, H. (1946). Mathematical Methods of Statistics. Princeton, N.J., Princeton University Press.

Goodman, Leo A. (1971). The Analysis of Multidimensional Contingency Tables: Stepwise Procedures and Direct Estimate Methods of Building Models for Multiple Classifications. Technometrics, 13, 33-61.

Hays, William L. (1973). Statistics for the Social Sciences. New York, N.Y., Holt, Rinehart and Winston.

Margolin, Barry H. and Light, Richard J. (1974). Analysis of Variance for Categorical Data II: Small Sample Comparisons with Chi Square and Other Competitors. Journal of the American Statistical Association, 69, 755-764.

Peters, Charles C. and van Voorhies, Walter R. (1940). Statistical Procedures and Their Mathematical Bases. New York, N.Y., McGraw-Hill.