MPRA
Munich Personal RePEc Archive

A measure of association (correlation) in nominal data (contingency tables), using determinants

Thomas Colignatus

Thomas Cool Consultancy & Econometrics

20. March 2007

Online at http://mpra.ub.uni-muenchen.de/2662/
MPRA Paper No. 2662, posted 10. April 2007

A measure of association (correlation) in nominal data (contingency tables), using determinants

Thomas Colignatus, April 10 2007
http://www.dataweb.nl/~cool

■ Summary

Nominal data currently lack a correlation coefficient, such as has already been defined for real data. A measure can be designed using the determinant, with the useful interpretation that the determinant gives the ratio between volumes. With M a $m \times n$ contingency table with $m \geq n,$ and A = Normalized[M], then $A'A$ is a square $n \times n$ matrix and the suggested measure is r = Sqrt[det[$A'A$]]. With M an $n_1 \times n_2 \times \ldots \times n_k$ contingency matrix, then pairwise correlations can be collected in a $k \times k$ matrix $\mathbf{R}$ so that the overall correlation is $R = f[\mathbf{R}]$. An option is to use R = Sqrt[1 - det[$\mathbf{R}$]]. However, for both nominal and cardinal data the advisable choice for such a function f is to take the maximal multiple correlation within $\mathbf{R}$.

Introduction

With $x_i \in \mathbb{R}^n$ real data vectors, $i = 1, \ldots, k$, the correlation matrix $\mathbf{R}$ contains the pairwise correlations between x_i and x_j. Let $\mathcal{R}$ = det[$\mathbf{R}$] the determinant of $\mathbf{R}$ and let $\mathcal{R}_{i,j}$ be that specific co-factor. Then the multiple correlation coefficient for the OLS regression of the first variable on the others is $(R^2)_{1.2,..,k} = 1 - \mathcal{R} / \mathcal{R}_{1,1}$, see Johnston (1972:134). This approach not only relies on various typographic forms of the letter “r” but also on the ratio scale of the data (real variables). For nominal data such a correlation measure apparently is lacking. For example for a political poll on Republican or Democrat support also categorized by sex (perhaps not only constituents but also candidates) we cannot apply the Pearson formula for correlation since categorical values such as Male versus Female aren’t numbers, making it difficult to find some “average value” and to square them.

This paper suggests a measure of association (or correlation) between nominal data. It relies on the use of determinants. The general idea is that determinants measure the change in volume when mapping a body from one space to another, and one can find a specific determinant that lies between 0 and 1. Since the suggested measure is new and has not been tested in statistical practice, it is hoped that theorists of linear algebra and statisticians working with nominal data look into the potential value of this suggestion.

It must be mentioned as well that the author has limited time and resources. He has used the literature in the list of references but there obviously exists more literature.

An other important limitation is that a correlation coefficient only gives some rough indication of association. For the specific kind of association one would design a model and perform a test. However, the correlation coefficient has some psychological value for research on real data and undoubtedly will have the same effect for nominal data as well. A strong association for example might cause more curiosity as to what the true model is.

An example

Real data are a researcher's paradise since you can do anything with real data. For example, when we have data on temperatures {-2.4, 4.3, 10.5, 21.5, 29.9} in degrees Celsius and associated sales of icecream {10.5, 11.3, 8.7, 50.4, 70.8} in kilo's then it is easy to determine a correlation. A correlation value generally lies between -1 and +1 and in this case we find that temperature and icecream sales are highly and positively correlated, so that we are motivated to determine the true model that captures their relationship.

```
Correlation[{-2.4, 4.3, 10.5, 21.5, 29.9}, {10.5, 11.3, 8.7, 50.4, 70.8}]
```

0.928301

The situation is entirely different for nominal data. This kind of data only gives categories that do not allow straightforward calculation. Consider for example two shops selling both blue and green hats, and a customer visiting both shops. The customer tries all hats and scores them as "fit" or "no fit". The resulting data are nominal since they only count the cases and there isn't an "average colour" or "average fit". To determine associations, the nominal data can be collected in cross tabulations a.k.a. contingency tables. The following example is from Kleinbaum et al. (2003:277). When we want to analyze these data and express "how much association" there is between the data, and read the statistical reference guides for help, we discover that nominal data and in particular these contingency tables lack a standard coefficient of correlation. The contingency table is meant to analyze association, and the data lie there on the disection table right in front of us, but we cannot do anything with them, to great frustration.

```
TableForm[mat = {{{5, 1}, {8, 2}}, {{2, 8}, {1, 5}}},
 TableHeadings → {{"Shop1", "Shop2"}, {"Green", "Blue"} , {"Fit", "No fit"}}]
```

		Green	Blue
Shop1	Fit	5	8
	No fit	1	2
Shop2	Fit	2	1
	No fit	8	5

Contingency tables are much used in psychology, epidemiology or experimental economics, and the frustration that we feel with respect to above table is just a small example of what these researchers must experience every day. Statistical analyses in psychology, epidemiology and experimental economics quickly proceed with more complex approaches like the χ^2test on statistical independence, which tests are not only more complex but also require levels of statistical significance that tend to say little about the strength of association. When we collect sufficiently large numbers of data then a low association may still turn statistically significant. The most dubious research outcome is when we start out with small numbers and a suggestion of high association that motivates us to collect more numbers, with the result that we find a low association that however differece from independence in a manner that is statistically significant: what to make of that ? Perhaps the true test in that case is to see whether the result also differs significantly from the originally found high association - but to discuss that we actually need a measure of association.

It is also interesting to observe that a common notion in real data analysis is that "correlation doesn't mean causation". However, this expression has little use in nominal data - since there is no standard measure of correlation.

This paper presents a new proposal for correlation in contingency tables. We will develop this measure below, starting with a 2 × 2 table, generalize this to a $n \times n$ table, then $m \times n$, and finally $n_1 \times n_2 \times ... \times n_k$.

For the hat shop contingency table above we find for the overall association:

```
NominalCorrelation[mat] // N
```

0.648564

Leading ideas

The general idea is that the determinant of a matrix measures the change in volume when mapping a body from one space to another. We can also find a specific determinant that lies between 0 and 1, so that we get a normalized volume ratio that can be interpreted as correlation. Note that the correlation coefficient for real data has the geometric interpretation of the cosine of the angle between the vectors. For nominal data we thus don't find a cosine but the important point is that there is a meaningful interpretation. Perhaps the true position should be the other way around that the standard is a volume ratio and that real data don't quite follow that standard but use a cosine instead. Perhaps there is a way that the volume ratio can be translated into a cosine interpretation. For now it suffices that we consider volumes only. Since the suggested coefficient is new and has not been tested in statistical practice, it is hoped that theorists of linear algebra and statisticians working with nominal data look into the potential value of this suggestion.

With respect to the statistical reference guides - and see in particular the accessible sources on the internet Becker (1999), Garson (2007) and Losh (2004) - it must be observed that we find that there are actually various possible measures of association for contingency tables. This is another way of saying that there is no standard. The multitude of measures does not present a Land of Cockaign but rather a tropical forest or maze. We find Phi, (Pearson's) Contingency Coefficient, Tschuprow's T, Cramer's V, Lambda, (Theil's) Uncertainty Coefficient and some tetrachoric correlation. **Appendix A** reviews these measures and rejects them for various reasons. A key point is that some measures are not symmetric. Another key point is that the most promising measures Phi, (Pearson's) Contingency Coefficient and Cramer's V all depend upon the χ^2statistic that is not elementary. Cramer's V is the most promising measure but is limited to the $m \times n$ case, and thus cannot handle the hat shop example above. However, it appears that the proposed measure for nominal correlation is equal to Cramer's V for the 2 × 2 case. We might start with that case and develop the story from there. However, it is better to develop the new measure by its own logic. **Appendix B** discusses the use of Cramer's V in relation to the χ^2test and shows its complexity. Readers who are versed in statistics and want to start on familiar grounds might consider starting there.

When we compare this situation with the one for real data then it appears that it is not common for real data to summarize the correlation matrix into one single number. The aforementioned correlation $(R^2)_{1.2,..,k} = 1 - \mathcal{R} / \mathcal{R}_{1,1}$ between the first and the other variables should not be confused with the overall association between a set of variables irrespective of order. Apparently, real data are such a paradise for analysis that the question about an overall correlation is not urgent. The issue is more urgent for nominal data since one quickly gets higher order contingency tables. It appears useful to define such an overall measure for both for nominal and real data. For real data, the idea is further developed in **Appendix C**. Essentially the point is the choice of some f such that overall correlation within the data is $R = f[\mathbf{R}]$. One option is to say $R^2 = 1 - \mathcal{R} = 1 - \det[\mathbf{R}]$ since off-diagonal zero's mean no correlation while that determinant becomes 1. It appears better to base an overall correlation upon the multiple correlations $(R^2)_{i.2,..,k}$. For both types of data, the best f to take is the highest value of any multiple correlation in the data. The reason is that we are used to think in terms of such multiple correlations. Having a single high multiple correlation is sufficient to say that the whole block of data shows high association. PM 1. We may also think about separating data into subsets. Once the set of data with that maximal correlation has been located, it can be separated from the other data, that may contain another maximum. This reminds of canonical correlation methods, but is not the same. PM 2. This generalization procedure might also be used for Cramer's V but has not been implemented.

Analysis of the example

As shown above, the hat shop case has a nominal correlation of 0.649. Since this is closer to 1 than to 0, we conclude that there is quite some association in these data. If we would select a statistical significance level then we could decide on statistical independence. But given this amount of correlation we certainly become more interested in what the relation between these variables is.

The example of the hat shops has been taken since it is an example of the Simpson paradox. In each separate shop the green hats fit relatively better, but for both shops combined the blue hats fit relatively better. For Shop 1 the fit / no-fit odds for green hats are 5/1 and for blue hats 8/2, thus the odds ratio (5/1) / (8/2) = 5/4. For Shop 2 we find the fit / no-fit odds ratio (1/4) / (1/5) = 5/4 too. For the two shops separately the odds ratio is above 1 but for the total (7/9 versus 9/7) it is below 1. The dispersion over the two shops is a confounder.

```
OddsRatio /@ Append[mat, Plus @@ mat]
```

$\left\{\frac{5}{4}, \frac{5}{4}, \frac{49}{81}\right\}$

The further analysis depends upon the case at hand. When the tables only concern the problem of fitting hats then it makes sense to eliminate the confounder, add the two tables, and find a small association.

```
NominalCorrelation[Plus @@ mat] // N
```

0.125

When these data tables don't concern hat shops but represent another kind of problem, then it might not be sensible to merely add the subtables. In that case the difference between 0.125 and 0.649 helps us to consider that there indeed is some relation, and, if the problem is serious enough, we might grow convinced that it could be worth while to find the true model. For example, we might do a meta-analysis on the findings of the separate shops, aggregating the problem in such a way that the overall direction reflects the individual ones. Or, for example in voting theory, if a bill must pass both Houses in Parliament (Congress and Senate) then we might accept that it is OK that it passes in both while it would not pass in a joint session.

Thus, this example shows that the notion of nominal correlation coefficient might help in the analysis. To understand the measure, let us start with the 2 × 2 case. Note that **Appendix D** discusses conventional approaches to the 2 × 2 case.

The 2 × 2 case

■ The layout

The 2 × 2 case is much discussed in epidemiology. Contingency tables generally are presented with table-headings and border-sums. Calculations are normally done with the inner matrix. The following is an example where men and women are dieting or not. Observed frequencies are *a* to *d*. We wonder whether the behaviour of the groups differs.

```
DiseaseTestMatrix[] = mat2 = Bordered2DPrSolve[ a  b  □
                                                c  d  □  ];
                                                □  □  □
```

```
DiseaseTestMatrix[Table, "men", "dieting"]
```

	men	¬ men	Tested
dieting	a	b	$a+b$
¬ dieting	c	d	$c+d$
Sum	$a+c$	$b+d$	$a+b+c+d$

For analysis, we rely on the inner matrix only.

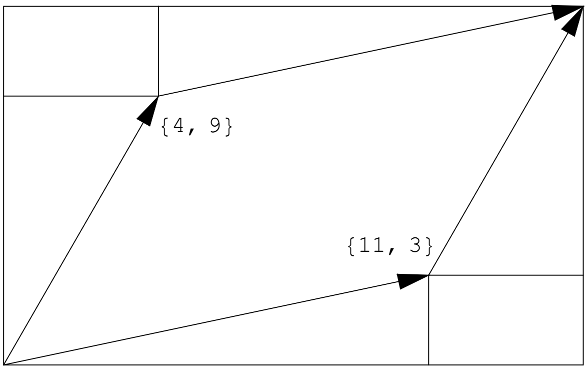
```
mat3 = StripMatrix[mat2]
```

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

■ An insight: diagram of the 2 × 2 case

The 2 × 2 matrix $\{\{a, b\}, \{c, d\}\}$ contains two row vectors $\{a, b\}$ and $\{c, d\}$ that together span a parallelepiped. When we draw a diagram of this, we find that the parallelepiped is contained in a rectangle with sides $(a + c)$ and $(b + d)$ which are the column sums of the matrix. The following gives a numerical example (a routine defined for this), and recall that we discuss nonnegative matrices:

```
ShowDet[{{11, 3}, {4, 9}}];
```



The total area of the rectangle is given by $(a + c)(b + d)$ while the area of the parallelepiped can be found by substraction of the small rectangles and triangles, thus $(a + c)(b + d) - 2\, bc - 2 * (\frac{1}{2}ab) - 2 * (\frac{1}{2}cd) = ad - bc$. This latter value is the determinant of the matrix.

```
(a + c) (b + d) - 2 b c - 2 (1/2 a b) - 2 (1/2 c d) // Simplify
```

$a\,d - b\,c$

```
Det[{{a, b}, {c, d}}]
```

$a\,d - b\,c$

When we take the ratio of the areas $cr = (ad - bc) / ((a + c)(b + d))$ then we find a number between -1 and 1.

Note also that the determinant $ad - bc$ also holds for the dual (transposed) matrix, giving a ratio rr.

Since it is arbitrary which variable influences the other, a more robust measure is the geometric average $\sqrt{cr * rr}$. The numerator remains *ad* - *bc* but the denominator becomes $\sqrt{}((a+c)(b+d)(a+b)(c+d))$. This gives us a "standardized volume ratio".

```
FullSimplify[CorrelationPr2By2[mat2], Assumptions → {a ≥ 0, b ≥ 0, c ≥ 0, d ≥ 0}]
```

$$\frac{a\,d-b\,c}{\sqrt{(a+b)\,(a+c)\,(b+d)\,(c+d)}}$$

We can easily check that a diagonal matrix with $b = c = 0$ gives outcome +1 and with $a = d = 0$ gives outcome -1. Nominal data have no natural order, but one cannot avoid an order of presentation and the sign of the correlation in this case reflects that.

■ Statistical independence means zero correlation

As might already have been obvious from the properties of determinants, algebraic independence means that this measure shows zero association. The following routine constructs a matrix by multiplying the marginals. We can multiply with the total number of observations.

```
mat4 = PrTable[t, p] n
```

$$\begin{pmatrix} n\,p\,t & n\,(1-p)\,t \\ n\,p\,(1-t) & n\,(1-p)\,(1-t) \end{pmatrix}$$

```
FullSimplify[CorrelationPr2By2[mat4], Assumptions → {t ≥ 0, p ≥ 0}]
```

0

Next step: square matrices

The 2 × 2 case can easily be extended to the $n \times n$ case. The determinant is only defined for square matrices. We directly get a measure if we apply the paradigm that we put the determinant in the numerator and the square root of the products of the border sums in the denominator.

In normalizing a matrix with the products of sums of columns and rows, it may be noted that the latter relate to determinants of diagonal matrices. Let A be any matrix. Let detr(A) = det(diag(A 1)) and detc(A) = det(diag(A' 1)). These detr and detc are just the products of the diagonals but there may be an advantage to write them in this manner. The determinant of a normalized matrix can be resolved in subterms using the properties of the determinant with respect to multiplications of rows and columns. The suggested measure for square matrice M then gives det(M) / $\sqrt{\text{detr}(M)\ \text{detc}(M)}$.

The case $n = 3$ is already a big deviation from $n = 2$. The following example is from Mood & Graybill (1963:325). Given the categories, and possibly an implied order related to the way of presentation, one might interprete the negative association as "the more capable the less poorly clothed". The correlation measure shows little *overall* association though (note the emphasis on "overall").

```
mat5 = ( {"Dull", "Intelligent", "Very Capable"},
         "Very well clothed"   81   322   233
         "Well clothed"       141   457   153
         "Poorly clothed"     127   163    48 );
```

	"Dull"	"Intelligent"	"Very Capable"
"Very well clothed"	81	322	233
"Well clothed"	141	457	153
"Poorly clothed"	127	163	48

```
SquareMatrixNormedDet[Take[mat5, -3, -3]] // N
```

-0.0285548

The low overall association does not preclude that there might be some stronger association when we aggregate over subgroups. However, in that case we need a theory that can handle $m \times n$ matrices.

```
mat6 = mat5 . Transpose[{{1, 0, 0, 0}, {0, 1, 1, 0}, { 0, 0, 0, 1}}]
```

Null	Dull + Intelligent	Very Capable
Very well clothed	403	233
Well clothed	598	153
Poorly clothed	290	48

```
SquareMatrixNormedDet[Take[mat6, -3, -2]] // N
```

SquareMatrixNormedDet::dim : Must be a n by n matrix More…

Next step: *m* × *n* matrices

Take an arbitrary $m \times n$ contingency table, say $A \in \mathbb{R}^{m,n}$, not necessarily integers. With $f[x] = A\, x$, we have $f: \mathbb{R}^n \to \mathbb{R}^m$. A property in linear algebra is that if S is a subset of $\mathbb{R}^n$ with Volume[S] then Volume[f[S]] = Sqrt[det[$A'.A$]] Volume[S]. Hence Sqrt[det[$A'.A$]] gives a volume ratio. For a problem like the hat shops, one can interprete f as taking combinations of the nominal categories, where a combination is not quite an addition (though it is in linear algebra). Thus there is an interpretation that makes the volume ratio an interpretable summary of association. Note that there is also a dimensionality adjustment, since a 2D surface has zero 3D volume; however, each space has a unit metric, e.g. as $1\, m^2$ or $1\, m^3$.

To make the measure robust we include the following: (1) A is normalized by dividing column-wise and row-wise by the square roots of the sums of the columns and rows respectively, (2) of $A'.A$ and $A.A'$ we take the one with lower size, since the larger one will give a zero determinant. It makes sense to consider only the smaller matrix since all variation in the higher dimension is only relevant with respect to the smaller dimension.

Hence: let M be the $m \times n$ contingency table and $m \geq n$. The suggested VolumeRatio or NominalCorrelation measure is

$$r = \text{Sqrt}[\det[A' A]] \text{ with } A = \text{Normalized}[M]$$

PM 1. For normalization, let dr = diag[M.1] and dc = diag[1'.M] with diag[...] the diagonal matrix. Then $A = dr^{-\frac{1}{2}}\, M\, dc^{-\frac{1}{2}}$. We can check that if M is square, then the $m \times n$ measure reduces to the measure already defined for square measures, with the only loss of the sign (due to the square root of the squares). PM 2. Note that this is a special kind of normalization. Repeated application results into different values. PM 3. Given that we are still considering nonnegative matrices, the only condition for the division would seem to be that every row or column contains at least one non-zero element. A row or column with only zero's would cause the determinant to be zero too so that division can be left out of consideration. It might be too simple to merely drop such a row or column. PM 4. For a calculation procedure, it seems most efficient to first check the diagonal. Only for zero elements on the diagonal it is necessary to check whether also the row or column are all zero. PM 5. In OLS the coefficient of determination relates to the correlation between explained variable y and explanation $\hat{y}$. In this present case for nominal data we consider an overall correlation. See **Appendix C**.

We now can tackle above case.

```
NominalCorrelation[Take[mat6, -3, -2]] // N
```

0.208759

Finally $n_1 \times n_2 \times \ldots \times n_k$

■ In general

Let us take x_i as nominal data, $i = 1, \ldots, k$, and let n_i be the number of nominal categories in the i^{th} variable, then a contingency matrix M is of size $n_1 \times n_2 \times \ldots \times n_k$. When we can define an association measure $r_{i,j}$ between two variables then we can collect all of those in a $k \times k$ correlation matrix **R.** Subsequently, we can define an overal correlation as $R = f[\mathbf{R}]$. For this *f,* we best take the highest value of any multiple correlation in the data.

For nominal data, the crux is to find a good $r_{i,j}$. The relation between two variables x_i and x_j can be considered in two ways. (1) One way is to sum out all other variables, giving $S = M_{i,j}$, the border matrix of x_i and x_j. It may well be that some third variable in some category has most influence, but when we consider only two variables then the influence of the other variables and the manner of influence might be considered to no longer apply. (2) The other approach is to hold that all variation in the submatrices that generate the border matrix is important too. Consider all matrices $S_p = M_{i,j}(p)$ used in the summing procedure to create $S = M_{i,j} = \sum_{p=1}^{n[i,j]} M_{i,j}(p)$ and note that $p = 1, \ldots, n[i, j] = n_1 \times n_2 \times \ldots \times n_k / (n_i \times n_j)$. With our measure r on an arbitrary $m \times n$ contingency table, we determine the r_p for each S_p, and determine a weighted average $\bar{r}_p$, using the total number of observations in S_p as the weight. We thus distinguish between aggregating correlations for the same two variables given a total border sum (for which we take the weighted average) and aggregating correlations for k variables (for which we take the maximal value of any multiple correlation). Each method has its own reason.

■ Using bordersums

Reconsider the shop case, its dimensions (1) Shop, (2) Colour and (3) Fitness, and its border matrices. For example, the border matrix for Shop versus Colour (the first two dimensions) is created by summing over Fitness (the third dimension).

```
TableForm[mat = {{{5, 1}, {8, 2}}, {{2, 8}, {1, 5}}},
 TableHeadings → {{"Shop1", "Shop2"}, {"Green", "Blue"} , {"Fit", "No fit"}}]
```

		Green	Blue
Shop1	Fit	5	8
	No fit	1	2
Shop2	Fit	2	1
	No fit	8	5

```
BorderMatrices[mat]
```

$$\left\{\{1, 2\} \to \begin{pmatrix} 6 & 10 \\ 10 & 6 \end{pmatrix}, \{1, 3\} \to \begin{pmatrix} 13 & 3 \\ 3 & 13 \end{pmatrix}, \{2, 3\} \to \begin{pmatrix} 7 & 9 \\ 9 & 7 \end{pmatrix}\right\}$$

When we use only these border matrices to determine the association between the variables then we neglect the variation that is in the submatrices. Application of the NominalCorrelation measure to these bordermatrices gives the elements of the total correlation matrix.

```
nc[bm, Mat] = NominalCorrelationMatrix[mat, Method → BorderMatrices]
```

$$\begin{pmatrix} 1 & \frac{1}{4} & \frac{5}{8} \\ \frac{1}{4} & 1 & \frac{1}{8} \\ \frac{5}{8} & \frac{1}{8} & 1 \end{pmatrix}$$

By default, MultipleSquared[mat] only considers the maximal multiple correlation in this correlation matrix.

```
nc[mrs, bm] = {Correlation →  (MultipleRSquared[nc[bm, Mat]] // Sqrt // N)}
```

{Correlation → 0.648564}

An alternative is to take the average of all multiple correlations. One may experiment with such functions but it will be noted quickly that a single high correlation within a block of data might be sufficient to say that there is high correlation within that data.

```
MultipleRSquared[nc[bm, Mat], Function → Average] // Sqrt // N
```

0.540495

PM. Note the outcome of this alternative measure of overall correlation:

```
nc[Det, bm] = {Correlation →  ((1 - Det[nc[bm, Mat]]) // Sqrt // N)}
```

{Correlation → 0.655506}

■ Using inner submatrices

When we want to account for the variation in the inner submatrices, then we can determine all submatrices that are used in the sum for a border matrix, determine each separate VolumeRatio, and then add these outcomes. It will be sensible in this addition to use the weights given by the numbers of observations in each submatrix. Note that there is an element of arbitrariness in using the weighted sum. We might also weigh e.g. the *squared* measures of association, or use a geometric average, or take the maximal value, and so on. For the time being, the simple weighted average seems wise. It will also be instructive to see how the procedure works. In the following, the "VR" stands for the VolumeRatio measure, i.e. the NominalCorrelation, and the "Add" stands for taking the weighted sum.

```
NominalCorrelationMatrix[mat, Method → Show]
```

$$\begin{pmatrix} 1 & \text{Add}\left[\left\{\text{VR}\left[\begin{pmatrix} 5 & 8 \\ 2 & 1 \end{pmatrix}\right], \text{VR}\left[\begin{pmatrix} 1 & 2 \\ 8 & 5 \end{pmatrix}\right]\right\}\right] & \text{Add}\left[\left\{\text{VR}\left[\begin{pmatrix} 5 & 1 \\ 2 & 8 \end{pmatrix}\right], \text{VR}\left[\begin{pmatrix} 8 & 2 \\ 1 & 5 \end{pmatrix}\right]\right\}\right] \\ \text{Add}\left[\left\{\text{VR}\left[\begin{pmatrix} 5 & 8 \\ 2 & 1 \end{pmatrix}\right], \text{VR}\left[\begin{pmatrix} 1 & 2 \\ 8 & 5 \end{pmatrix}\right]\right\}\right] & 1 & \text{Add}\left[\left\{\text{VR}\left[\begin{pmatrix} 5 & 1 \\ 8 & 2 \end{pmatrix}\right], \text{VR}\left[\begin{pmatrix} 2 & 8 \\ 1 & 5 \end{pmatrix}\right]\right\}\right] \\ \text{Add}\left[\left\{\text{VR}\left[\begin{pmatrix} 5 & 1 \\ 2 & 8 \end{pmatrix}\right], \text{VR}\left[\begin{pmatrix} 8 & 2 \\ 1 & 5 \end{pmatrix}\right]\right\}\right] & \text{Add}\left[\left\{\text{VR}\left[\begin{pmatrix} 5 & 1 \\ 8 & 2 \end{pmatrix}\right], \text{VR}\left[\begin{pmatrix} 2 & 8 \\ 1 & 5 \end{pmatrix}\right]\right\}\right] & 1 \end{pmatrix}$$

Actually doing the calculation gives this correlation matrix.

```
nc[All, Mat] = NominalCorrelationMatrix[mat, Method → All]  // N
```

$$\begin{pmatrix} 1. & 0.221917 & 0.61807 \\ 0.221917 & 1. & 0.0413449 \\ 0.61807 & 0.0413449 & 1. \end{pmatrix}$$

In this numerical example, using all variation in the submatrices, the total measure appears to be no different from the one using only the border matrices.

```
nc[mrs, All] = {Correlation → (MultipleRSquared[nc[All, Mat]] // Sqrt // N)}
```

{Correlation → 0.648564}

PM. Note the outcome of this alternative measure of overall correlation.

```
nc[Det, All] = {Correlation → ((1 - Det[nc[All, Mat]]) // Sqrt // N)}
```

{Correlation → 0.649327}

■ Review

We calculated the nominal correlations using both the border matrices or all submatrices, and using only the maximal multiple RSquared or the determinant measure. The following table collects above results in a summary overview.

```
heading = TableHeadings → {{mrs, Det}, {bm, All}};
```

```
InsideTable[Set, nc, heading]
```

```
InsideTable[Show, Correlation]
```

	bm	All
mrs	0.648564	0.648564
Det	0.655506	0.649327

It appears that the Det measure is more sensitive to the use of the inner submatrices, while the maximal RSquared is not affected.

■ The default

As already shown above, the default definition uses the inner submatrices and the maximal Multiple RSquared.

```
NominalCorrelation[mat] // N
```

0.648564

See **Appendix E** for an example in higher dimensions. See **Appendix F** for a note on the Frobenius theorems on nonnegative (contingency) matrices. See **Appendix G** for a note on causality, and how this paper originated. See **Appendix H** for a list of the routines used here.

Conclusion

The crux in this development lies in the $m \times n$ case. It covers the lower orders, and it forms the core for the $n_1 \times n_2 \times \ldots \times n_k$ generalization. Since the $m \times n$ case has a sound interpretation, the overall interpretation is sound. There is an element of arbitrariness in the methods of aggregation in the upward generalization. Namely there is the use of a weighted average for submatrices of pairwise correlations, and the use of the maximal element of multiple correlations in the total correlation matrix. Though arbitrary as this seems, each step has merit, and as a standard it is well defined. The link to Cramer's V in the 2 × 2 case seems to be only a happenstance, though additional research might show a deeper cause.

The suggested measure has a useful interpretation as the volume ratio, with values between 0 and 1. Next to the χ^2 scores and tests currently in use, the suggested measure has the added value of indicating the overall strength of association. That a deviation from independence is statistically significant or not, at some level of significance, need not be the most meaningful message when researching an issue.

Given this measure of correlation, we must say (and we can do it now, with some relief that we can do so) that the normal caveats apply, i.e. that correlation is no causation, and that correlation itself doesn't say much about the actual model.

Appendix A: Measures of association mentioned by common resources

There seems to be no standard and satisfactory measure of association for nominal data. Apart from the official books mentioned in the list of references, like Mood & Graybill (1963) and Kleinbaum et al. (2003), also the resources on the internet mentioned there, notably Becker (1999), Garson (2007) and Losh (2004), have been used to find measures for the association in nominal data. See Cool, Th. (1999, 2001), "The Economics Pack, Applications for *Mathematica*", and the website update, for an implementation of the LLR and Pearson tests, in the CrossTable package, and for some implementations in the Life Sciences packages. The following measures have been found.

(1) Fisher's exact test does a test and does not provide a measure of association. Similarly for the likelihood ratio test and the Pearson approximation. The measure then is yes/no with respect to passing the test.

(2) The Odds Ratio depends upon direction (column-wise versus row-wise) and does not seem to be generalizable. For the 2 by 2 table, taking the default column direction:

```
OddsRatio[]
```

$$\frac{a\,d}{b\,c}$$

(3) "The tetrachoric correlation coefficient is essentially the Pearson product-moment correlation coefficient between the row and column variables, their values for each observation being taken as 0 or 1 depending on the category it falls into"

(4) Phi (Cramer's V in non-square tables): "Also in 2-by-2 tables, phi is identical to the correlation coefficient. In larger tables, where phi may be greater than 1.0, there is no simple intuitive interpretation, which is a reason why phi is often used only for 2-by-2 tables." (Garson (2007))

PM 4.1. The statement “Also in 2-by-2 tables, phi is identical to the correlation coefficient” is confusing since that correlation coefficient is not clearly defined for nominal data. As with the "tetrachoric" measure, one takes {1, 0} assignments to the nominal values, but are these also the values for Phi ?

PM 4.2. There are different ways to determine a χ^2 value. The Pearson test statistic is $(o - t)^2 / t$, with *o* the observed and *t* the theoretical frequency. For a 2 by 2 table the theoretical frequency might come from the hypothesis of independence. Both other hypotheses are possible too.

PM 4.3 Cramer’s V is the most useful of all these possible measures. Yet its interpretation is the χ^2 from the hypothesis of statistical independence, and one wonders whether this captures the intuition of correlation as given by the real variables.

Cramer’s V, defined for a $m \times n$ matrix, is the square root of the Pearson χ^2 value divided by the sample size *p* times *q*, where *q* is the smaller of (m - 1) and (n - 1). Thus V = $\sqrt{\chi^2 / (p\, q)}$. For a 2 × 2 case we find $q = 1$, and then V reduces to Phi (that always takes $q = 1$).

See **Appendix B** for a longer discussion.

(5) The Contingency Coefficient, Pearson’s C = SQRT[X^2/(X^2 + n)]. “There is no easily intuited interpretation of C or C*, though C* may be viewed as the association between two variables as a percentage of their maximum possible variation. Pearson viewed C as a nominal approximation of Pearsonian correlation r.” (Garson (2007))

(6) The Uncertainty Coefficient, UC or Theil’s U - an asymmetric measure.

(7) Hoeffding’s Dependence Coefficients - not looked into.

(8) Eta is an asymmetric correlation coefficient, and its dependent variable would be interval scaled.

Appendix B: Relation to the χ^2 measure

▪ In general

A standard procedure in the analysis of contingency tables is to perform the Pearson test on independence, which is an approximation of the log-likelihood ratio test. Independence arises when the cells in the matrix are mere products of the marginals, given in the border sums. With o the observed data and t the theoretical frequencies, derived from those products given by the hypothesis of independence, the Pearson test statistic is $\Sigma\ (o - t)^2\ /\ t$. When we introduce additional assumptions on the distribution then we can perform a test. A possible assumption is a multinomial distribution and for larger numbers this gets closer to the multivariate normal, such that the Pearson test statistic would have a χ^2distribution. Mood & Graybill (1963:314) clarify the procedure: "In casting about a test which may be used when the sample is not large, we may inquire how it is that a test criterion comes to have a unique distribution for large samples when the distribution actually depends on unknown parameters which may have any values in certain ranges. The answer is that the parameters are not really unknown; they can be estimated, and their estimates approach their true values as the sample size increases." The routine Test in The Economics Pack contains the Pearson χ^2 test and explains the various intermediate steps. In output, the rule **Do → Accept | Reject** expresses whether the hypothesis of independence is accepted or rejected at the stated significance level. When we simplify the output of this routine then it appears that there are -1. terms that are reals and better should be treated as integers; this replacement is not done in the part that is printed or that is contained within some expressions.

■ The 2 × 2 case

```
test = Test[Chi2, Pearson, {{a, b}, {c, d}}] /.
   Times[-1. , x_] → Times[-1, x] // Simplify
```

Pearson::consis : The o and t input may not add up to the same

Observed

$$\begin{matrix} a & b \\ c & d \end{matrix}$$

Theoretical

$$\begin{matrix} \frac{(a+b)(a+c)}{a+b+c+d} & \frac{(a+b)(b+d)}{a+b+c+d} \\ \frac{(a+c)(c+d)}{a+b+c+d} & \frac{(b+d)(c+d)}{a+b+c+d} \end{matrix}$$

Pearson (o − t)^2 / t

$$\begin{matrix} \frac{(a+b+c+d)\left(a-\frac{1.(a+b)(a+c)}{a+b+c+d}\right)^2}{(a+b)(a+c)} & \frac{(a+b+c+d)\left(b-\frac{1.(a+b)(b+d)}{a+b+c+d}\right)^2}{(a+b)(b+d)} \\ \frac{(a+b+c+d)\left(c-\frac{1.(a+c)(c+d)}{a+b+c+d}\right)^2}{(a+c)(c+d)} & \frac{(a+b+c+d)\left(d-\frac{1.(b+d)(c+d)}{a+b+c+d}\right)^2}{(b+d)(c+d)} \end{matrix}$$

$\Big\{\text{ArrayDepth} \to 2,\ \text{BorderSums} \to \begin{pmatrix} a+b & c+d \\ a+c & b+d \end{pmatrix},$

$\text{Chi2PValue} \to 1. - 1.\, Q\left(0.5, 0., \frac{0.5\left(\frac{(b\,c-1.\,a\,d)^2}{(a+b)(b+d)} + \frac{(b\,c-1.\,a\,d)^2}{(a+c)(c+d)} + \frac{(a\,d-1.\,b\,c)^2}{(a+b)(a+c)} + \frac{(a\,d-1.\,b\,c)^2}{(b+d)(c+d)}\right)}{a+b+c+d}\right),$

$\text{DegreesOfFreedom} \to 1,\ \text{Dimensions} \to \{2, 2\},$

$\text{Do} \to \text{If}\left[1.\, Q\left(0.5, 0., \frac{0.5\left(\frac{(b\,c-1.\,a\,d)^2}{(a+b)(b+d)} + \frac{(b\,c-1.\,a\,d)^2}{(a+c)(c+d)} + \frac{(a\,d-1.\,b\,c)^2}{(a+b)(a+c)} + \frac{(a\,d-1.\,b\,c)^2}{(b+d)(c+d)}\right)}{a+b+c+d}\right) \geq 0.95, \text{Reject}, \text{Accept}\right],$

$\text{MarginalPr} \to \left\{\text{MarginalPr}(1) \to \left\{\frac{a+b}{a+b+c+d}, \frac{c+d}{a+b+c+d}\right\}, \text{MarginalPr}(2) \to \left\{\frac{a+c}{a+b+c+d}, \frac{b+d}{a+b+c+d}\right\}\right\},$

$\text{NumberOfObservations} \to a+b+c+d,\ \text{Partition} \to \{1, 2\},\ \text{ProbabilityMatrix} \to \begin{pmatrix} \frac{(a+b)(a+c)}{(a+b+c+d)^2} & \frac{(a+b)(b+d)}{(a+b+c+d)^2} \\ \frac{(a+c)(c+d)}{(a+b+c+d)^2} & \frac{(b+d)(c+d)}{(a+b+c+d)^2} \end{pmatrix},$

$\text{SignificanceLevel} \to 0.05,\ \text{Test} \to \text{Pearson},\ \text{TestStatistic} \to \frac{(a+b+c+d)(b\,c-a\,d)^2}{(a+b)(a+c)(b+d)(c+d)}\Big\}$

The key object of interest is the test statistic. We can select it from above output and give it a name that expresses its distribution under the null hypothesis of independence.

```
chi2 = (TestStatistic /. test)
```

$$\frac{(a+b+c+d)(b\,c-a\,d)^2}{(a+b)(a+c)(b+d)(c+d)}$$

The most useful existing measure for association in crosstables is Cramer's V, defined for a $m \times n$ matrix, as the square root of the χ^2 value divided by the sample size p times q, where q is the smaller of (m - 1) and (n - 1). Thus V = $\sqrt{\chi^2 / (p\,q)}$. For a 2 × 2 case we find q = 1, and then V reduces to Phi (that always takes q = 1).

Cramer's V then is (for the 2 × 2 case):

```
Sqrt[chi2 / (NumberOfObservations /. test)]
```

$$\sqrt{\frac{(b\,c-a\,d)^2}{(a+b)\,(a+c)\,(b+d)\,(c+d)}}$$

Thinking about association within a 2 × 2 table along this route is not simple and plunges us into the deep waters of statistical hypothesis testing. We can recognize that the 2 × 2 case has the same form as the suggested measure for Nominal Correlation, yet, the reasoning behind it is quite different.

■ A 3 × 3 case

With respect to above 3 × 3 case, Cramer’s V suggests some modest association.

```
CramersV[Take[mat5, -3, -3]]
```

0.197584

This CramerV routine has been defined using the Test routine, its results have been stored and are still available. We find that the hypothesis of independence would be rejected at the standard 5% significance level.

```
Do /. Results[CramersV]
```

Reject

Note however this crucial observation, Cool (2001:368) discussing some other examples: “The χ^2 test is sensitive to the number of observations per degree of freedom. (... When ...) there are more observations per degree of freedom, (...) the test is stronger. (In the same manner, by increasing the sample size, acceptance tends to turn into rejection, and by reducing the size, rejection can turn into acceptance.)” Thus a statistically significant deviation from independence in a χ^2 test does not imply that there would be much of an association.

The Cramer V outcome of 0.197584 differs importantly from the outcome of Nominal Correlation of 0.0285548. There seems little to be said about this. Different measures, different results. But one cannot evade the impression that Cramer’s V is too sensitive to the quirks of χ^2 testing.

■ A higher dimensional case

Cramer’s V is only defined for $m \times n$ matrices. But the χ^2 test can handle higher dimensions. We can run the test on the hat shop case and find that it tells us little about the amount of association. The hypothesis of independence is rejected at the 5% significance level. And then ? The χ^2 p-value does not impress as a good measure for association either.

```
test = Test[Chi2, Pearson, mat]
```

Observed

5	8
1	2
2	1
8	5

Theoretical

4.	4.
4.	4.
4.	4.
4.	4.

Pearson (o − t)^2 / t

0.25	4.
2.25	1.
1.	2.25
4.	0.25

$$\left\{\text{ArrayDepth} \to 3,\ \text{BorderSums} \to \begin{pmatrix} 16 & 16 \\ 16 & 16 \\ 16 & 16 \end{pmatrix},\ \text{Chi2PValue} \to 0.000107511,\right.$$

$$\text{DegreesOfFreedom} \to 1,\ \text{Dimensions} \to \{2, 2, 2\},\ \text{Do} \to \text{Reject},$$

$$\text{MarginalPr} \to \{\text{MarginalPr}(1) \to \{0.5, 0.5\},\ \text{MarginalPr}(2) \to \{0.5, 0.5\},\ \text{MarginalPr}(3) \to \{0.5, 0.5\}\},$$

$$\text{NumberOfObservations} \to 32,\ \text{Partition} \to \{1, 2, 3\},\ \text{ProbabilityMatrix} \to \begin{pmatrix} \{0.125, 0.125\} & \{0.125, 0.125\} \\ \{0.125, 0.125\} & \{0.125, 0.125\} \end{pmatrix},$$

$$\left.\text{SignificanceLevel} \to 0.05,\ \text{Test} \to \text{Pearson},\ \text{TestStatistic} \to 15.\right\}$$

Appendix C: Overall correlation between real variables

■ In general

With $x_i \in \mathbb{R}^n$ real data vectors, $i = 1, ..., k$, the correlation matrix **R** contains the pairwise correlations between x_i and x_j. Let $\mathcal{R}$ = det[**R**] the determinant of **R** and let $\mathcal{R}_{i,j}$ be that specific co-factor. Then the (squared) multiple correlation coefficient for the OLS regression of the first variable on the others is $(R^2)_{1.2,..,k} = 1 - \mathcal{R} / \mathcal{R}_{1,1}$, see Johnston (1972:134). It is not customary to use a measure for overall correlation between real variables, or to summarize the correlation matrix by one number. However, in the light of the discussion of nominal variables, it is not insensible to do so. A possible measure is $R^2 = 1 - \mathcal{R}$. For example, when all variables are uncorrelated then the off-diagonal elements in **R** are zero and det[**R**] = 1, and then we see an overall correlation of $R^2 = 0$. Similarly, when two variables are fully correlated in a 2 × 2 case **R** = {{1, -1}, {-1, 1}} then the determinant is zero and R^2 = 1 - det[**R**] = 1 reflects that overall correlation. However, the choice of $R^2 = 1 - \mathcal{R}$ might be less informative when we are not used to interpreting such values. Given that we are most accustomed to the (squared) multiple correlation coefficient, we can consider the whole series $(R^2)_i = (R^2)_{i.2,..,k} = 1 - \mathcal{R} / \mathcal{R}_{i,i}$, $i = 1, ..., k$, and take the maximal value. In the case that for one i that $\mathcal{R}_{i,i} = 1$ then this again collapses to $1 - \mathcal{R}$. It seems to make most sense to say that the correlation within a block of data is given by the maximal value that we can find in it. Of course, we might consider a normal or geometric average. Considering the options it seems that using $R^2 = 1 - \mathcal{R}$ or such averages lose out against simply taking the maximal value.

■ Derivation of 0 ≤ det[R] ≤ 1

As said, a possible measure for overall correlation is $R^2 = 1 - \mathcal{R}$ where $\mathcal{R}$ = det[**R**]. Whatever we said about taking the maximal multiple correlation as the superior measure, there can still be value in just this determinant measure. The measure is valid if we can show that $0 \leq R^2 = 1 - \mathcal{R} \leq 1$ or 0 ≤ det[**R**] ≤ 1.

The correlation matrix is by definition a symmetric and positive semi-definite matrix, so that det[**R**] ≥ 0. What about the upper bound, such that $R^2 = 1 - \mathcal{R} \geq 0$ or det[**R**] ≤ 1 ? It took me a moment to find the following proof. Since **R** is a symmetric and positive semi-definite matrix, its determinant is given by the product of all its (nonnegative) eigenvalues λ_i, while the sum of those is also given by the trace of the diagonal, i.e. the dimension, see Johnston (1972:105-109). Thus we have det[**R**] = $\lambda_1 ... \lambda_k$ while $k = \lambda_1 + ... + \lambda_k$ and $\lambda_i \geq 0$. For singular **R** we trivially have $\mathcal{R} = 0$. For regular **R** and $k \leq 2$ we trivially have $0 \leq \mathcal{R} \leq 1$. For example for $k = 2$, we know that the correlation r between two variables satisfies $-1 \leq r \leq 1$ and the determinant is:

```
Det[{{1, r}, {r, 1}}]
```

$1 - r^2$

At issue is only regular **R** and $k > 2$. A step in the proof is that the maximal determinant is given by $\lambda_i = \lambda$. We might show this by calculus, maximizing $\mathcal{R}$ subject to the given constraints, but it is simpler to assume that all eigenvalues are at their maximal value except for the first two. Then we get $\mathcal{R} = \lambda_1 \lambda_2 \lambda^{n-2}$. Since the sum is given, the difference between λ and λ_1 is reflected in λ_2, thus $\mathcal{R} = (\lambda - x)(\lambda + x)\lambda^{n-2} = (\lambda^2 - x^2)\lambda^{n-2} = \lambda^n - X$, where X is a nonnegative amount. Thus the determinant is maximal when $x = 0$, and then all $\lambda_i = \lambda$. Then also $k = k\,\lambda$, or $\lambda = 1$, and $\mathcal{R} = 1$.

This proof can be supported by a drawing in the 2D plane where $2 = \lambda_1 + \lambda_2$ and their product finds a maximum in $\lambda_1 = \lambda_2 = 1$, meaning that $r = 0$.

Hence, an option is to use R = Sqrt[1 - det[**R**]]. Yet its usefulness must show in practice.

PM 1. Det[**R**] is also used in tests on (multi-) collinearity (Farrar-Glauber). PM 2. Testing that all correlations are zero gives a test statistic - n Log[det[**R**]]. PM 3. It can also be mentioned that there are other measures for overall correlation, notably in the area of entropy. This however seems to lead too far from the correlation paradigm.

■ Numerical example: Klein I model

The Klein I model is a good example, see Theil (1971:432). Data and regressions can be found at UCLA ATS (2007).

```
lis = Partition[TextToMatrix["
1920 39.8 12.7 28.8 2.7 180.1 44.9 2.2 2.4 3.4
1921 41.9 12.4 25.5-0.2 182.8 45.6 2.7 3.9 7.7
1922 45.0 16.9 29.3 1.9 182.6 50.1 2.9 3.2 3.9
1923 49.2 18.4 34.1 5.2 184.5 57.2 2.9 2.8 4.7
1924 50.6 19.4 33.9 3.0 189.7 57.1 3.1 3.5 3.8
1925 52.6 20.1 35.4 5.1 192.7 61.0 3.2 3.3 5.5
1926 55.1 19.6 37.4 5.6 197.8 64.0 3.3 3.3 7.0
1927 56.2 19.8 37.9 4.2 203.4 64.4 3.6 4.0 6.7
1928 57.3 21.1 39.2 3.0 207.6 64.5 3.7 4.2 4.2
1929 57.8 21.7 41.3 5.1 210.6 67.0 4.0 4.1 4.0
1930 55.0 15.6 37.9 1.0 215.7 61.2 4.2 5.2 7.7
1931 50.9 11.4 34.5-3.4 216.7 53.4 4.8 5.9 7.5
1932 45.6 7.0 29.0-6.2 213.3 44.3 5.3 4.9 8.3
1933 46.5 11.2 28.5-5.1 207.1 45.1 5.6 3.7 5.4
1934 48.7 12.3 30.6-3.0 202.0 49.7 6.0 4.0 6.8
1935 51.3 14.0 33.2-1.3 199.0 54.4 6.1 4.4 7.2
1936 57.7 17.6 36.8 2.1 197.7 62.7 7.4 2.9 8.3
1937 58.7 17.3 41.0 2.0 199.8 65.0 6.7 4.3 6.7
1938 57.5 15.3 38.2-1.9 201.8 60.9 7.7 5.3 7.4
1939 61.6 19.0 41.6 1.3 199.9 69.5 7.8 6.6 8.9
1940 65.0 21.1 45.0 3.3 201.2 75.7 8.0 7.4 9.6
1941 69.7 23.5 53.3 4.9 204.5 88.4 8.5 13.8 11.6", Number], 10];
```

```
{year, cons, profit, wpriv, invest,
   klag, xprod, wgov, govt, taxes} = Transpose[lis] ;
```

Currently, we just take the example of the regression of consumption on profits, lagged profits and the total wage sum (table 16.4). According to the references, the OLS estimate gives an R-Square of 0.9810. We can verify this as follows:

```
dat = {cons, profit, Lag[profit], wpriv + wgov} // Transpose // Rest;
```

```
cm = CorrelationMatrix[dat]
```

$$\begin{pmatrix} 1. & 0.715338 & 0.65205 & 0.982703 \\ 0.715338 & 1. & 0.769128 & 0.634156 \\ 0.65205 & 0.769128 & 1. & 0.579332 \\ 0.982703 & 0.634156 & 0.579332 & 1. \end{pmatrix}$$

The routine MultipleRSquared[**R**, i] calculates RSquared[i] = 1 - $\mathcal{R}$ / $\mathcal{R}_{i,i}$. Taking $i = 1$ gives us the squared multiple correlation coefficient of the OLS regression of the first variable, consumption, on the other three variables including a constant. The value we get fits the one reported in the literature.

```
MultipleRSquared[cm, 1]
```

0.981008

If we want to summarize the correlation matrix into one number, we might consider 1 - det[cm]. Note that the following still is a squared value.

```
1 - Det[cm]
```

0.995522

However, since we are more used to consider multiple correlations, the routine MultipleSquared[mat] by default puts out the maximal value of those. The routine allows the option Function → ... to use on the list of multiple correlations and the default value is Function → Max.

```
MultipleRSquared[cm]
```

0.981008

PM. While evaluating MultipleRSquared[mat] the routine stores all possible RSquared[i] in MultipleRSquared[List].

```
MultipleRSquared[List]
```

{0.981008, 0.719033, 0.627152, 0.976314}

Appendix D: Other relations for the 2 × 2 case

■ 1. Just recall

This is just to recall that the 2 × 2 case is generated both by the Volume Ratio measure and Cramer's V.

■ 2. Epidemiology

However, in epidemiology, the following approach is possible. Let $p_0 = a / (a + c)$, $p_1 = b / (b + d)$ and p the pooled probability found in the sum column. Then a test statistic for $p_0 = p_1$ differs from the VolumeRatio measure only in $\sqrt{N}$ with $N = a + b + c + d$.

$$\frac{\text{p0} - \text{p1}}{\sqrt{\text{p}\,(1 - \text{p})\,\left(\frac{1}{\text{a} + \text{c}} + \frac{1}{\text{b} + \text{d}}\right)}} \text{ /. } \{\text{p0} \to \text{a} / (\text{a} + \text{c}),\ \text{p1} \to \text{b} / (\text{b} + \text{d}),\ \text{p} \to (\text{a} + \text{b}) / (\text{a} + \text{b} + \text{c} + \text{d})\};$$

```
FullSimplify[%, Assumptions → {a ≥ 0, b ≥ 0, c ≥ 0, d ≥ 0}]
```

$$\frac{a\,d - b\,c}{\sqrt{\frac{(a+b)\,(a+c)\,(b+d)\,(c+d)}{a+b+c+d}}}$$

■ 3. Assigning values {-1, 1}

For a 2 × 2 table, the suggested measure collapses to the Pearson coefficient of correlation when the nominal values are replaced by {1, -1}.

Note that values {1, 0} are allowed in logic and that operations are defined on them, see Colignatus (2007a).

(i) A variable gets assigned the values {1, -1} rather than {1, 0} or True | False as in logic. A reason to avoid zero is that it might needlessly destroy information. A reason to use 1 versus -1 is that equal numbers of observations might be thought to balance each other, with an average outcome of zero.

(ii) The data points {1, 1}, {1, -1}, {-1, 1}, {-1, -1} generate x = {1, 1, -1, -1} and y = {1, -1, 1, -1} with frequencies {*a, c, b, d*}. Indeed, we might as well give a lists of {1, 1}, b lists of {1, -1} etcetera.

(iii) Thus we get the normal correlation between x = {1, 1, -1, -1} and y = {1, -1, 1, -1}with frequencies {*a, c, b, d*}.

(iv) Using formal parameters {*a, c, b, d*} and simplification shows that the determinant is used in the numerator and the row sums and column sums in the denominator. The values 1 and -1 assigned to the nominal data do not occur any more, at least, not in an obvious manner.

It may be mentioned that it was this relation that caused the author to investigate the issue into the direction that resulted into the general measure suggested above. This case was generalized first to $n \times n$, then $m \times n$, then $n_1 \times n_2 \times \ldots \times n_k$ (and the latter first to bordermatrices and then to the inner submatrices).

Appendix E: An example in a higher dimension

The following shows an example of the Nominal Correlation measure in a higher dimension. The example shows that the algorithm is straightforward. It also indicates that, especially considering higher dimensions, a weighted average is a sensible choice.

```
mat2 = Table[i + 10 j + 100 k + 1000 m, {i, 2}, {j, 3}, {k, 4}, {m, 2}]
```

$$\begin{pmatrix} \begin{pmatrix} 1111 & 2111 \\ 1211 & 2211 \\ 1311 & 2311 \\ 1411 & 2411 \end{pmatrix} & \begin{pmatrix} 1121 & 2121 \\ 1221 & 2221 \\ 1321 & 2321 \\ 1421 & 2421 \end{pmatrix} & \begin{pmatrix} 1131 & 2131 \\ 1231 & 2231 \\ 1331 & 2331 \\ 1431 & 2431 \end{pmatrix} \\ \begin{pmatrix} 1112 & 2112 \\ 1212 & 2212 \\ 1312 & 2312 \\ 1412 & 2412 \end{pmatrix} & \begin{pmatrix} 1122 & 2122 \\ 1222 & 2222 \\ 1322 & 2322 \\ 1422 & 2422 \end{pmatrix} & \begin{pmatrix} 1132 & 2132 \\ 1232 & 2232 \\ 1332 & 2332 \\ 1432 & 2432 \end{pmatrix} \end{pmatrix}$$

This is the border matrix for dimension 3 and 4 (that should have sizes m = 4 and n = 2).

```
BorderMatrix[mat2, {3, 4}]
```

$$\begin{pmatrix} 6729 & 12729 \\ 7329 & 13329 \\ 7929 & 13929 \\ 8529 & 14529 \end{pmatrix}$$

These are the 4 by 2 submatrices used in the summation of the border matrix.

```
TabledBorderMatrix[mat2, {3, 4}]
```

$$\begin{pmatrix} \text{Mat}\left(\begin{pmatrix} 1111 & 2111 \\ 1211 & 2211 \\ 1311 & 2311 \\ 1411 & 2411 \end{pmatrix}\right) & \text{Mat}\left(\begin{pmatrix} 1121 & 2121 \\ 1221 & 2221 \\ 1321 & 2321 \\ 1421 & 2421 \end{pmatrix}\right) \\ \text{Mat}\left(\begin{pmatrix} 1131 & 2131 \\ 1231 & 2231 \\ 1331 & 2331 \\ 1431 & 2431 \end{pmatrix}\right) & \text{Mat}\left(\begin{pmatrix} 1112 & 2112 \\ 1212 & 2212 \\ 1312 & 2312 \\ 1412 & 2412 \end{pmatrix}\right) \\ \text{Mat}\left(\begin{pmatrix} 1122 & 2122 \\ 1222 & 2222 \\ 1322 & 2322 \\ 1422 & 2422 \end{pmatrix}\right) & \text{Mat}\left(\begin{pmatrix} 1132 & 2132 \\ 1232 & 2232 \\ 1332 & 2332 \\ 1432 & 2432 \end{pmatrix}\right) \end{pmatrix}$$

Check that it fits.

```
(% // Add) /. Mat → Identity
```

$$\begin{pmatrix} 6729 & 12729 \\ 7329 & 13329 \\ 7929 & 13929 \\ 8529 & 14529 \end{pmatrix}$$

These are all the submatrices.

```
VolumeRatioMatrix[mat2, Method → Show]
```

$$\left(\begin{array}{l}
1 \\[6em]
\text{Add}\left[\begin{pmatrix} \text{VR}\left[\begin{pmatrix} 1111 & 1121 & 1131 \\ 1112 & 1122 & 1132 \end{pmatrix}\right] & \text{VR}\left[\begin{pmatrix} 2111 & 2121 & 2131 \\ 2112 & 2122 & 2132 \end{pmatrix}\right] & \text{VR}\left[\begin{pmatrix} 1211 & 1221 & 1231 \\ 1212 & 1222 & 1232 \end{pmatrix}\right] & \text{VR}\left[\begin{pmatrix} 2211 & 2221 & 22 \\ 2212 & 2222 & 22 \end{pmatrix} \\ \text{VR}\left[\begin{pmatrix} 1311 & 1321 & 1331 \\ 1312 & 1322 & 1332 \end{pmatrix}\right] & \text{VR}\left[\begin{pmatrix} 2311 & 2321 & 2331 \\ 2312 & 2322 & 2332 \end{pmatrix}\right] & \text{VR}\left[\begin{pmatrix} 1411 & 1421 & 1431 \\ 1412 & 1422 & 1432 \end{pmatrix}\right] & \text{VR}\left[\begin{pmatrix} 2411 & 2421 & 24 \\ 2412 & 2422 & 24 \end{pmatrix} \end{pmatrix}\right. \\[6em]
\text{Add}\left[\begin{pmatrix} \text{VR}\left[\begin{pmatrix} 1111 & 1211 & 1311 & 1411 \\ 1112 & 1212 & 1312 & 1412 \end{pmatrix}\right] & \text{VR}\left[\begin{pmatrix} 2111 & 2211 & 2311 & 2411 \\ 2112 & 2212 & 2312 & 2412 \end{pmatrix}\right] & \text{VR}\left[\begin{pmatrix} 1121 & 1221 & 1321 & 1421 \\ 1122 & 1222 & 1322 & 1422 \end{pmatrix}\right] \\ \text{VR}\left[\begin{pmatrix} 2121 & 2221 & 2321 & 2421 \\ 2122 & 2222 & 2322 & 2422 \end{pmatrix}\right] & \text{VR}\left[\begin{pmatrix} 1131 & 1231 & 1331 & 1431 \\ 1132 & 1232 & 1332 & 1432 \end{pmatrix}\right] & \text{VR}\left[\begin{pmatrix} 2131 & 2231 & 2331 & 2431 \\ 2132 & 2232 & 2332 & 2432 \end{pmatrix}\right] \end{pmatrix}\right] \\[6em]
\text{Add}\left[\begin{pmatrix} \text{VR}\left[\begin{pmatrix} 1111 & 2111 \\ 1112 & 2112 \end{pmatrix}\right] & \text{VR}\left[\begin{pmatrix} 1211 & 2211 \\ 1212 & 2212 \end{pmatrix}\right] & \text{VR}\left[\begin{pmatrix} 1311 & 2311 \\ 1312 & 2312 \end{pmatrix}\right] \\ \text{VR}\left[\begin{pmatrix} 1411 & 2411 \\ 1412 & 2412 \end{pmatrix}\right] & \text{VR}\left[\begin{pmatrix} 1121 & 2121 \\ 1122 & 2122 \end{pmatrix}\right] & \text{VR}\left[\begin{pmatrix} 1221 & 2221 \\ 1222 & 2222 \end{pmatrix}\right] \\ \text{VR}\left[\begin{pmatrix} 1321 & 2321 \\ 1322 & 2322 \end{pmatrix}\right] & \text{VR}\left[\begin{pmatrix} 1421 & 2421 \\ 1422 & 2422 \end{pmatrix}\right] & \text{VR}\left[\begin{pmatrix} 1131 & 2131 \\ 1132 & 2132 \end{pmatrix}\right] \\ \text{VR}\left[\begin{pmatrix} 1231 & 2231 \\ 1232 & 2232 \end{pmatrix}\right] & \text{VR}\left[\begin{pmatrix} 1331 & 2331 \\ 1332 & 2332 \end{pmatrix}\right] & \text{VR}\left[\begin{pmatrix} 1431 & 2431 \\ 1432 & 2432 \end{pmatrix}\right] \end{pmatrix}\right]
\end{array}\right.$$

```
NominalCorrelationMatrix[mat2] // N
```

$$\begin{pmatrix} 1. & 1.42186\times 10^{-6} & 0.0000194491 & 0.0000834515 \\ 1.42186\times 10^{-6} & 1. & 0. & 0.00136274 \\ 0.0000194491 & 0. & 1. & 0.0186297 \\ 0.0000834515 & 0.00136274 & 0.0186297 & 1. \end{pmatrix}$$

And the total measure of association comes from taking the determinant from the latter matrix.

```
MultipleRSquared[%] // Sqrt
```

0.0186796

Or directly:

```
NominalCorrelation[mat2] // N
```

0.0186796

Appendix F: A note on the Frobenius theorems

A contingency matrix is a nonnegative matrix, so that the Frobenius theorems apply. It is not quite clear how the eigenvectors come into play. There might be a relation here. The Frobenius eigenvalue $\lambda \geq 0$ is a real value that is at least as large as the absolute value of any other (possibly complex) eigenvalue. (Takayama (1974: 375)). Thus det(A) = $\lambda_1 ... \lambda_n \leq \lambda^n$ and thus, if $\lambda > 0$ (which is definitely the case when A is indecomposable) then the implied Frobenius ratio measure is FrobRatio(A) = det(A) / $\lambda^n \leq 1$. It is not clear however how this relates to the notion of the volume ratio.

With square matrix A, if $A\, x = \lambda\, x$ for vector $x \neq 0$ and scalar λ, then we say that λ is an eigenvalue and x its eigenvector. Let $r = A$ 1 the row sums and $c =$ 1' A the column sums. Then 1' $A\, x = \lambda$ 1' x or $\lambda = c'x$ / 1'x = $c'x^*$, when 1' $x \neq 0$, and $x^* = x$ / 1'x a normalized vector. Since A and A' have the same eigenvalues, there is also a $y'\, A = \lambda\, y'$ or $\lambda = y'r$ / y' 1. When $x \neq 0$ (everywhere) and in particular $x > 0$ then there is also the possibility of a change in dimensions such that D^{-1}= diag(x) and B = D A D^{-1} such that B 1 = λ 1.

In general det(A) = $\lambda_1 ... \lambda_n$. Collecting all eigenvalues on the diagonal in Λ, zero everywhere else, and all eigenvectors in X (also using spanning vectors for higher multiplicity) then $A\, X = X\, \Lambda$.

Appendix G: On inference and causality

This discussion originated from considering the links between logic (Colignatus (2007a)) and causality (Pearl (2000)). Suppose that rain is a cause and wetness of streets is an effect. Is the observation that the streets are wet a good predictor of what was the cause ?

mat =

"Observation count"	"It rains"	"It doesn't rain"	"Total"
"The streets are wet"	25	3	□
"The streets are not wet"	0	□	□
"Total"	□	□	100

;

```
mat = Headed2DTableSolve[mat]
```

$$\begin{pmatrix} \text{Observation count} & \text{It rains} & \text{It doesn't rain} & \text{Total} \\ \text{The streets are wet} & 25 & 3 & 28 \\ \text{The streets are not wet} & 0 & 72 & 72 \\ \text{Total} & 25 & 75 & 100 \end{pmatrix}$$

Instead of remembering all these 100 cases either individually or by frequency distribution, the memory processing unit might save on storage and retrieval costs by adopting a general rule (induction) that “If it rains then the streets are wet”. This can become a general rule for which we can use a truthtable. The truthtable tests the condition whether the frequency is zero or non-zero, with entries True or False in the table.

In the course of considering these issues, the author noted that the common notion “correlation doesn’t mean causation” has little use for such nominal data - since there is no standard measure of correlation. But now there is:

```
CorrelationPr2By2[Take[mat, -3, -3]] // N
```

0.92582

PM. The author may now continue considering the links between logic and causality. His hypothesis is that causality cannot be inferred from common statistics and has to do with the model and the order of calculation (time’s arrow). In logic, there is a difference between implication ⇒ and inference ⊢. In causal models, there is a difference between equality == and assignment = (using *Mathematica*'s notation; other notations use = for equality and := for assignment). What causality is in Nature, inference is in the Mind. The overall umbrella would be the difference between statics and dynamics.

Appendix H: Routines

This discussion uses The Economics Pack, Cool (2001). The proper routines are all defined with the more technical term “VolumeRatio” but for statistical usage the term “NominalCorrelation” will be more suitable.

This analysis started out with a small case, clarified and simplified by *Mathematica*. The subsequent steps in generalization benefitted from that testing and prototyping environment as well. The availability of linear algebra routines and matrix manipulations was essential. The routines rely on symbolic operations. The environment also allowed the quick creation of user friendly routines, that not only have a clear logical structure but also come with help support and all. Finally, these routines also form building blocks that can be used immediately within other routines. All in all, there is yet another reason to thank the makers of *Mathematica*.

```
?CorrelationPr2By2
```

```
CorrelationPr2By2[{{n11, n12}, {n21, n22}}] gives the measure
  of association for a contingency table of two binary nominal variables
CorrelationPr2By2[mat] may also take a 3x3 matrix but then
  the borders are seen as sum totals, and dropped
CorrelationPr2By2[Definition, mat] gives the original definition
  without simplification based upon the nonnegative values
Let C (cause) be the column variable and E (effect) the row variable. In logic,
  the variables take values {1, 0}. Here it is better to take {1, -1} so that
  equal numbers of observations give a zero mean. Output then is the normal
  Pearson CorrelationPr[{1, 1, -1, -1}, {1, -1, 1, -1}, {n11, n21, n12, n22}].
See SquareMatrixNormedDet for larger {n, n} and VolumeRatio for
  {n, m} contingency tables (Cramer's V is in the Chi2 package)
```

?SquareMatrixNormedDet

```
SquareMatrixNormedDet[mat] gives Det[NormalizedMatrix[mat]] for square mat normalized by
  Sqrt[rs * cs] where the latter are the products of the row and column sums. See
  VolumeRatio. That measure can be simplified for square matrices with a sign retained
  for the direction. For 2 by 2 matrices, see CorrelationPr2By2 in Cool`Statistics`Common`
```

?NormalizedMatrix

```
NormalizedMatrix[mat] divides the elements of a n by m matrix by
  the squares of the row and column sums. Note that repeated application
  doesn't generate the same result. Is primarily used in VolumeRatio
```

?NominalCorrelation

```
NominalCorrelation[...] translates into VolumeRatio[...]
```

?NominalCorrelationMatrix

```
NominalCorrelationMatrix[...] translates into VolumeRatioMatrix[...]
```

?VolumeRatio

```
VolumeRatio[mat] does (1) ma = NormalizedMatrix[mat],(2) m = ma.ma' or m = ma'.ma
  whichever has smaller dimensions and nonzero determinant, (3) output Sqrt[Det[m]].
  PM. The absolute value of the determinant of real vectors gives the volume of the
  parallelepipedum created by those vectors. Let f: Rn -> Rm be defined by matrix
  A, so that f[x] = A x, and let S be a subset of Rn, then volume[f[S]] = Sqrt[Det[
  A'.A]]* volume[S]. Hence Sqrt[Det[m]] gives a normalized volume ratio. PM. If
  mat is a contingency table with nonnegative numbers, for nominal variables, then
  VolumeRatio gives a measure of association comparable to a correlation coefficient
VolumeRatio[mat] gives Sqrt[MultipleRSquared[VolumeRatioMatrix[mat]]] if mat is more-
  dimensional (MatrixQ[mat] is false). In that latter case the Method option is passed
  on to VolumeRatioMatrix, and the Function option passed on to MultipleRSquared
```

?VolumeRatioMatrix

```
VolumeRatioMatrix[mat] for a {n1, ..., nm} matrix gives a m by
  m correlation matrix, containing the VolumeRatio[m[ni, nj]] measures of
  association. There are three ways to obtain that latter individual measure:
Method -> BorderMatrices sums the other dimensions
Method -> All (default) determines all individual k
  by p matrices in the lower dimensions, determines their association,
  and gives the sum, weighted by total numbers in those submatrices
Method -> Show shows that latter method
```

?TabledBorderMatrix

```
TabledBorderMatrix[mat, {i, j}] decomposes the {i, j} border matrix into the submatrices
  that cause its sum value. These submatrices are indicated with label Mat. PM
  Check the same outcome as BorderMatrix by using (% // Add) /. Mat -> Identity
```

?MultipleRSquared

```
MultipleRSquared[mat, i] = 1 - Det[mat] /  CoFactor[mat, i, i], and gives R^2[i;
  2, ..., i-1, i+1, ..., n] or the squared multiple correlation coefficient from
  the OLS regression of ith (dependent) variable upon the other (independent)
  variables, provided that mat is the CorrelationMatrix between all variables
MultipleRSquared[mat] takes a function of values for all i;
  such results can be found in MultipleRSquared[List]
Options are Function -> (default Max) and Range -> (default True);
  the latter for a test whether mat indeed has values -1 <= # <= 1
A special case is Function -> Det, which just returns 1 - Det[mat]
```

Literature

Colignatus is the name of Thomas Cool in science.

Becker, L.A. (1999), “Measures of Effect Size (Strength of Association)”, http://web.uccs.edu/lbecker/-SPSS/glm_effectsize.htm, Retrieved from source

Cool, Th. (1999, 2001), “The Economics Pack, Applications for *Mathematica*”, http://www.dataweb.nl/~cool, ISBN 90-804774-1-9, JEL-99-0820

Colignatus, Th. (2006), “On the sample distribution of the adjusted coefficient of determination (R2Adj) in OLS”, http://library.wolfram.com/infocenter/MathSource/6269/

Colignatus, Th. (2007a), “A logic of exceptions”, http://www.dataweb.nl/~cool, ISBN 978-90-804774-4-5

Colignatus, Th. (2007b), “Voting theory for democracy”, 2nd edition, http://www.dataweb.nl/~cool, ISBN 978-90-804774-5-2

Garson, D. (2007), “Nominal Association: Phi, Contingency Coefficient, Tschuprow's T, Cramer's V, Lambda, Uncertainty Coefficient”, http://www2.chass.ncsu.edu/garson/pa765/assocnominal.htm, Retrieved from source

Johnston J. (1972), “Econometric methods”, McGraw-Hill

Kleinbaum, D.G., K.M. Sullivan and N.D. Barker (2003), “ActivEpi Companion texbook”, Springer

Losh, S.C. (2004), “Guide 5: Bivariate Associations and Correlation Coefficient Properties”, http://edf5400-01.fa04.fsu.edu/Guide5.html, Retrieved from Source

Mood, A.M. and F.A. Graybill (1963), “Introduction to the theory of statistics”, McGraw-Hill

Pearl, J. (2000), “Causality. Models, reasoning and inference”, Cambridge

Simon, R. (2007), “Lecture Notes and Exercises 2006/07”, http://www.maths.lse.ac.uk/Courses/MA201/, Retrieved from source

Takayama A. (1974), “Mathematical economics”, The Dryden Press

Theil H. (1971), “Principles of econometrics”, North-Holland

UCLA ATS (2007), “SAS Textbook Examples. Econometric Analysis, Fourth Edition by Greene. Chapter 16: Simultaneous Equations Models”, http://www.ats.ucla.edu/stat/SAS/examples/greene/chapter16.htm, Retrieved from source

(Other) websites

http://en.wikipedia.org/wiki/Contingency_table
http://post.queensu.ca:8080/SASDoc/getDoc/en/procstat.hlp/corr_sect26.htm
http://en.wikipedia.org/wiki/Fisher%27s_exact_test