International Mathematical Forum, 3, 2008, no. 34, 1655 - 1658

On Maximum Value of Correlation Coefficient

Narges Abbasi

Department of Statistics, Payame Noor University
Shiraz Centre 71365-944, Shiraz, Iran
Abbasi@spnu.ac.ir

Abstract

Among the methods of analysis of qualities of random variables, the study of relations is the most significant. Many criteria in the measurement of the extend of correlations exit, but some of them don't attain values ± 1 when two variables are perfectly correlated. In this article, the upper bound of correlation coefficient is considered. There are different methods to show that the upper bound occurs in perfectly correlated between the two variables.

Keywords: Correlation Coefficient,log-linear Method, Methods of Analysis, Qualities Random Variables

1 Introduction

The $n \times n$ contingency table for two variables(factors), A and B, arranged as follows:

	A_1	$\cdots$	A_n	total
B_1	f_{11}	$\cdots$	f_{n1}	$f_{1.}$
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
B_n	f_{1n}	$\cdots$	f_{nn}	$f_{n.}$
total	$f_{.1}$	$\cdots$	$\cdots$	f

where f_{ij} is the number of observations belonging to ith stage of A and jth stage of B, and f is the number of total observations. The analysis of such two-dimensional tables generally involves testing for the independence of the two variables using the familiar chi-squared statistic. Three and higher-dimensional tables are now routinely analyzed using log-linear models(Everitt (1992)). Usually, the test statistics for testing independence indicate the measure of correlation of factors. The chi-square statistic, $\chi^2 = \sum_{i,j} \frac{(f_{ij}-e_{ij})^2}{e_{ij}}$, is used for determination of correlation; where e_{ij} is the expected frequency of ith stage

of A and jth stage of B under independence. The value of χ^2 is positive, while the value of independence lies between -1 and $+1$ (or zero and $+1$). Several measurements of correlation were introduced by Pearson(1975), Goodman and Kurskal(1972), and Yule(1900). One of them, $C = \sqrt{\frac{\chi^2}{\chi^2+f}}$, named contingency coefficient. It shows that the measure of association of factors and attained values between zero and less than one. The big value of C indicates the high degree of association. In this article, we obtain the upper bound for C, by different methods.

2 The Maximum Value

Notice to following formula is very useful

$$\begin{aligned} \chi^2 + f &= \sum_{i=1}^{n}\sum_{j=1}^{n}\frac{f_{ij}^2}{e_{ij}} \\ &= f\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{f_{ij}^2}{f_{i.}f_{.j}} \end{aligned}$$

and

$$\begin{aligned} C^2 &= \frac{\chi^2}{\chi^2+f} \\ &= \frac{f\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{f_{ij}^2}{f_{i.}f_{.j}} - f}{f\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{f_{ij}^2}{f_{i.}f_{.j}}} \\ &= 1 - \frac{1}{\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{f_{ij}^2}{f_{i.}f_{.j}}} \end{aligned} \tag{1}$$

The problem equals to determining the minimum upper bound for $\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{f_{ij}^2}{f_{i.}f_{.j}}$.

2.1 Simple Method

By simple method, we show that the maximum value of C is $\frac{\sqrt{2}}{2}$ for 2×2 contingency table. Let $0 < a,\ b,\ p,\ q < 1$ and $a = \frac{f_{11}}{f_{.1}},\ b = \frac{f_{12}}{f_{.2}}\ p = \frac{f_{11}}{f_{1.}}$, and $q = \frac{f_{21}}{f_{2.}}$. Again perhaps new notation we have

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{f_{ij}^2}{f_{i.}f_{.j}} = \sum_{i=1}^{n}\sum_{j=1}^{n}\frac{f_{ij}}{f_{i.}}\frac{f_{ij}}{f_{.j}}$$

$$\begin{aligned} &= pa + (1-p)b + q(1-a) + (1-b)(1-q) \\ &= (p-q)(a-b) + 1 \\ &\leq pa + 1 \\ &\leq 2 \end{aligned}$$

therefore

$$\begin{aligned} C^2 &= 1 - \frac{1}{\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{f_{ij}^2}{f_{i.}f_{.j}}} \\ &\leq 1 - \frac{1}{2} = \frac{1}{2} \end{aligned}$$

or

$$C \leq \frac{\sqrt{2}}{2}.$$

This method becomes very complicate when the dimensional of contingency table arise.

2.2 Using the Expectation

Reviewing on formula of χ^2, we can see that

$$\begin{aligned} \sum_{i=1}^{n}\sum_{j=1}^{n}\frac{f_{ij}^2}{f_{i.}f_{.j}} &= \sum_{i=1}^{n}\sum_{j=1}^{n}\frac{f_{ij}}{f_{i.}}\frac{f_{ij}}{f_{.j}} \\ &= \sum_{i=1}^{n}\mathbf{E}_{B|A=A_i}(\frac{f_{ij}}{f_{.j}}) \\ &\leq \sum_{i=1}^{n}1 \quad (since \frac{f_{ij}}{f_{.j}} \leq 1) \\ &= n. \end{aligned}$$

where $\mathbf{E}_{j|i}$ is conditional expectation of B. So, For any $n \times n$ contingency table the maximum value of contingency coefficient is $\sqrt{\frac{n-1}{n}}$. This method shows that if the number of stages of A or B be $m \leq n$, then the minimum value changes to $\sqrt{\frac{m-1}{m}}$.

2.3 Model of Association

Another method: the minimum value of χ^2 (zero) occurs when two random variables, on our survey, are independent and the maximum value of χ^2 related to perfectly correlated of variables, that is the value of correlation coefficient equals to near one. In this case, For two quantities variables, there is a linear

relationship with probability one. (Like two random variables X and Y, with $P(Y = aX + b) = 1$. Obviously, the number of options of X and Y must be equal.) This association on qualities factors stated on frequencies:

$$f_{jj} = \frac{f}{n}; \quad j = 1, 2, \ldots, n, \quad i \neq j$$

or

$$\begin{aligned} \frac{f_{ij}}{f_{i.}} \frac{f_{ij}}{f_{.j}} &= 1; \quad i = j = 1, 2, \ldots, n \\ \frac{f_{ij}}{f_{i.}} \frac{f_{ij}}{f_{.j}} &= 0; \quad i = j \end{aligned}$$

With this means, the maximum value of $\sum\sum \frac{f_{ij}^2}{f_{i.}f_{.j}}$ equals n, and the result will be obtained.

3 Sakoda Coefficient

By multiply $\sqrt{\frac{n}{n-1}}$, the maximum value of C changes to one. The new measure of association, $S = C\sqrt{\frac{n}{n-1}}$, named Sakoda Coefficient. It applies for any two dimensional contingency tables, $n \times m$ $(n = min(n, m))$.

4 References

[1] Bishop, Y. M. M., Fienberg, S. E. And Holland. P. W. (1975). *Discrete Multivariate Analysis: Theory and Practice*, M. I. T. Press, Cambridge, Mass.
[2] Everitt, B. S. (1992), *The Analysis of Contingency Tables*, 2nd edition, Chapman and Hall.
[3] Goodman, L. A., and Kruskal, E. H. (1972). *Measures of Association For Cross-Classification*, IV. J. Amer. Statist. Assoc., 67, 415-421.
[4] Yule, G. U. (1900). *On the Association of Attributes in Statistics*, Phil. Trans., A. 194, 257-319.

Received: March 16, 2008