Selçuk J. Appl. Math.
Special Issue. pp. 3-17, 2012

**Selçuk Journal of
Applied Mathematics**

A Comparison between Entropy-Based Association Measures and other Qualitative Association Measures

Atıf Evren, Elif Tuna

Yıldız Technical University, Faculty of Arts & Scince, Department of Statistics, Esenler 34210 Istanbul, Turkiye
e-mail: aevren2006@gmail.com, elfztrk@gmail.com

⋆ *Presented in 3^{rd} National Communication Days of Konya Eregli Kemal Akman Vocational School, 28-29 April 2011.*

Abstract. There are various statistics to measure the degree of association between qualitative variables in literature. Among them, some to mention are Pearson p (the coefficient of contingency), phi-square, Tschuprow's contingency coefficient, and Cramér's contingency coefficient. In addition, statistics derived from the concept of entropy like mutual information, Kullback-Leibler divergence and Jeffreys divergence can also be used in measuring association.

Key words: Measures of association, coefficients of contingency, Kullback Leibler divergence, Jeffreys divergence.
2000 Mathematics Subject Classification: 94A17, 62G10.

1. Introduction

In literature, there are some statistics developed to measure the degree of association between qualitative variables. These statistics are mainly derived from the chi-square value calculated for a contingency table. Besides, some statistics based on entropy measures are widely used to measure qualitative association. Infact statistical entropy empowers scientists quite a lot in attacking some problems especially when the distribution is in qualitative nature. One can consult [4] and [5] on some applications of entropy in statistics. In this study, we intend to compare entropy-based association measures with other qualitative association measures by means of two different applications.

2. Measures of Association for Qualitative Variables

Suppose the joint frequency distribution of two qualitative variables is summarized by a contingency table. Let the first variable is denoted by X_i(i=1,2,...,n) whereas the second variable is denoted by Y_j(j=1,2,...,m). Suppose also that

X_{ij} $(X = X_i, Y = Y_j)$ represents the values of this joint distribution and also that $f_{ij}$and e_{ij} represent observed and expected frequencies of X_{ij} values . Note that N stands for the total number of observations. Then

$$\sum^{n}\sum_{j=1}^{m} f_{ij} = \sum_{i=1}^{n}\sum_{j=1}^{m} e_{ij} = N \tag{1}$$

$$P(X = X_i) = \pi_i = \frac{1}{N}\sum_{j=1}^{m} f_{ij} \tag{2}$$

$$P(Y = Y_j) = \pi_j = \frac{1}{N}\sum_{i=1}^{n} f_{ij} \tag{3}$$

$$P(X = X_i, Y = Y_j) = \pi_{ij} = \frac{f_{ij}}{N}. \tag{4}$$

The sufficient condition for independence is that for all $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, m$

$$\pi_{ij} = \pi_i.\pi_j. \tag{5}$$

To measure the degree of association, phi-square statistic is defined as

$$\varphi^2 = \sum_{i=1}^{n}\sum_{j=1}^{m} \frac{\pi_{ij}^2}{\pi_i \pi_j} - 1. \tag{6}$$

This measure takes 0, if the variables are independent. The maximum value it can take is q-1. It should be noted that q=min$\{n, m\}$. For this reason the ratio $\frac{\varphi^2}{q-1}$can serve as a "standardized"[1]association index. This statistic is 0, when there is no association between the variables and also it is equal to 1 when there is perfect association between them [9]. The maximum likelihood estimators of the probabilities π_{ij}, π_i and π_j that appear in (5) can be found by maximizing the likelihood function based on a sample of N units. If L represents the likelihood function then

$$L = \prod_{i,j} (\pi_i.\pi_j)^{f_{ij}}. \tag{7}$$

Here $\sum_{i=1}^{n}\pi_i = 1$, $\sum_{j=1}^{m}\pi_j = 1$ and the quantity LogL-$\lambda\sum_{i=1}^{n}\pi_i - \mu \sum_{j=1}^{m}\pi_j$ is maximized for

$$\pi_j = \frac{1}{N}\sum_{i=1}^{n} f_{ij} \tag{8}$$

[1] i.e. its minimum value is 0 and maximum value is 1. The term "standardized" here is used somewhat in a different meaning from "standardized variables" in statistics.

$$\pi_i = \frac{1}{N}\sum_{j=1}^{m} f_{ij}. \tag{9}$$

Thus maximum likelihood estimators satisfy $e_{ij} = N\pi_i.\pi_j$

$$\chi^2 = \sum_{i=1}^{n}\sum_{j=1}^{m}\frac{(f_{ij}-e_{ij})^2}{e_{ij}} \tag{10}$$

fits approximately a chi-square distribution with (n-1)(m-1) degrees of freedom.[2]

$$\chi^2 = \sum_{i=1}^{n}\sum_{j=1}^{m}\frac{(f_{ij}-N\pi_i\pi_j)^2}{N\pi_i\pi_j}. \tag{11}$$

After manipulating algebraically a little bit;

$$\chi^2 = N\left(\sum_{i=1}^{n}\sum_{j=1}^{m}\frac{(\pi_{ij})^2}{\pi_i\pi_j} - 1\right) \tag{12}$$

$$\varphi^2 = \frac{\chi^2}{N}. \tag{13}$$

When the variables are independent this statistic is equal to 0. Because this statistic depends on the number of cells in the contingency table, there is a difficulty in evaluating the numeric values obtained. For that reason further modifications seem necessary [9].

2.1. Some Modifications

To overcome the difficulty just mentioned above, Pearson proposed the following statistic:

$$p = \left(\frac{\varphi^2}{1+\varphi^2}\right)^{1/2}. \tag{14}$$

Here p takes values between 0 and 1. Yet this statistic suffers from the fact that although the variables seem perfectly associated, p can not be equal to 1 exactly. In a multinomial sampling scheme, if $\hat{p}$ represents the maximum likelihood estimator of p, then

[2] Chi-square approach is only valid for limiting cases especially when the number of counts in each cell are not negligible. If a significant number of cell counts is less than 5, this approach may highly be misleading (Keeping, p316).

$$\hat{p} = \left(\frac{\chi^2/N}{1+\chi^2/N} \right)^{1/2} = \sqrt{\frac{\chi^2}{\chi^2/N}}. \tag{15}$$

When the number of rows are equal to that of columns in the contingency table, the maximum value that $\hat{p}$ can reach is $\sqrt{(q-1)/q}$, in other cases it can be less than 1. For this reason some adjustments are proposed in literature. For instance, Sakoda proposed the following[9]:

$$p* = \frac{p}{p_{MAX}} = \left(\frac{q\varphi^2}{(q-1)(1+\varphi^2)} \right)^{1/2}. \tag{16}$$

Here , $p*$ is equal to 0 when the variables are independent and 1 when they are associated perfectly.

2.2. Tschuprow's Contingency Coefficient

Another alternative is Tschuprow's contingency coefficient . Let T denotes this coefficient and is defined as

$$T = \left(\frac{\varphi^2}{\sqrt{(m-1)(n-1)}} \right)^{1/2} \tag{17}$$

Here T takes values between 0 and 1 as the other association measures. Besides it is important to note that it only achieves its maximum value when the contingency table is in a square form.

2.3. Cramér's Contingency Coefficient

As an alternative to p and T statistics, Cramér, proposed the following:

$$v = \left(\frac{\varphi^2}{q-1} \right)^{1/2} = \left(\frac{\chi^2}{N(q-1)} \right)^{1/2}. \tag{18}$$

Here even though the number of columns and the number of rows of the contingency table are not equal to each other, v can still reach its maximum value when there is perfect association. In such a case it takes the value of 1 [9].
It is very hard to determine the probability distributions of p, $p*$,T and v. Yet their distributions are determined by large sampling approach only. Under the assumption of independence $\left(\varphi^2 = 0\right)$, the following tail probabilities for T can still be calculated by the help of following equations:

(19) $$P(T \geq t_0) = P(T^2 \geq t_0^2) = P\left(\left[\frac{\chi^2}{N\sqrt{(n-1)(m-1)}}\right] \geq t_0^2\right)$$

(20) $$P(T \geq t_0) = P\left[\chi^2 \geq N t_0^2 \sqrt{(n-1)(m-1)}\right].$$

Tail probabilities for p, $p*$ and v can be calculated similarly. To calculate the standard errors of these distributions under the assumption , $\varphi^2 \neq 0$ one can refer to [9]. Since these formulations are rather complicated and maybe clumsy, we have preferred to skip these.

3. Shannon Entropy

For a discrete probability distribution, Shannon entropy is defined as

(21) (21) $$H = -\sum_{i=1}^{n} p_i \log p_i$$

The biggest uncertainty is encountered when each outcome is equally likely. In that situation the maximum entropy for discrete cases is as below:

(22) $$H_{MAX} = -\sum_{i=1}^{n} \frac{1}{n} \log(\frac{1}{n}) = \log n$$

In the other extreme (minimum uncertainty or minimum entropy) one can calculate it as

(23) $$H_{MIN} = 0.$$

3.1. Generalizations to Multivariate Cases

For multivariate discrete distributions, the entropy can be found by

(24) (24) $$H(X_1, ..., X_n) = -\sum_{X_1} ... \sum_{X_n} \log\left(f\left(x_1, ..., x_n\right)\right) f\left(x_1, ..., x_n\right)$$

and for multivariate continuous distributions

(25) $$H(X_1, ..., X_n) = -\int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} f\left(x_1, ..., x_n\right) \log\left(f\left(x_1, ..., x_n\right)\right) dx_1..dx_n.$$

3.2. Conditional Entropies

Suppose $f_{X\ /\ Y=y_j}\,(X/Y=y_j)$ and $f_{Y\ /\ X=x_i}\,(Y/X=x_i)$ represent two conditional probability distributions of X and Y given that $Y=y_j$ and $X=x_i$ have occurred respectively. In these cases, the conditional entropies are just the entropies of these conditional distributions. Thus one can formulate them as follows:

$$\begin{aligned}&H(X/Y=y_j)=\\&=-\sum_{i=1}^{n} f_{X=x_i/Y=y_j}(X=x_i/Y=y_j)\log(f_{X=x_i/Y=y_j}(X=x_i/Y=y_j))\end{aligned} \tag{26}$$

(27)

$$\begin{aligned}&H(Y/X=x_i)=\\&=-\sum_{j=1}^{m} f_{Y=y_j/X=x_i}(Y=y_j/X=x_i)\log(f_{Y=y_j/X=x_i}(Y=y_j/X=x_i)).\end{aligned}$$

But these two measure uncertainty only under the assumption that $Y=y_j$ and $X=x_i$ have already occurred respectively. So to investigate dependencies among variables, one can find other formulas for the average situation. From (26) and (27), more appropriate measures can be obtained by

(28)

$$\begin{aligned}&H(X/Y)=\\&=-\sum_{j=1}^{m} f_{Y=y_j}\sum_{i=1}^{n} f_{X=x_i/Y=y_j}(X=x_i/Y=y_j)\log(f_{X=x_i/Y=y_j}(X=x_i/Y=y_j))\end{aligned}$$

or

$$\begin{aligned}&H(X/Y)=\\&=\sum_{j=1}^{m}\sum_{i=1}^{n} f_{X=x_i,Y=y_j}(X=x_i/Y=y_j)\log(f_{X=x_i/Y=y_j}(X=x_i/Y=y_j))\end{aligned}$$

and similarly,

(29)

$$H(Y/X)=\sum_{i=1}^{n}\sum_{j=1}^{m} f_{Y=y_j,X=x_i,}(X=x_i/Y=y_j)\log(f_{Y=y_j/X=x_i}\,(Y=y_j/X=x_i))\,.$$

3.3. Entropy and Statistical Independence

If X and Y are independent, one can end in

$$H(X/Y)=H(X) \tag{30}$$

$$H(Y/X) = H(Y) \tag{31}$$

Entropies of different types of distributions (bivariate, univariate, and conditional distributions) are also related to each other. For example,

$$H(X,Y) = H(Y) + H(X/Y) \tag{32}$$

$$H(X,Y) = H(X) + H(Y/X). \tag{33}$$

3.4. Measure of Mutual Information

A measure of information that one variable gives about the uncertainty of the other is proposed by C.E. Shannon and it is as follows[15]:

$$I(X;Y) = \sum_{j=1}^{m}\sum_{i=1}^{n} P(X = x_i, Y = y_j) \log \frac{P(X = x_i, Y = y_j)}{P(X = x_i)P(Y = y_j)}. \tag{34}$$

For continuous distributions, summation operators in (34) are replaced by integration operators. If X and Y are independent, then $I(X;Y) = 0$.[3] After some algebraic work,

$$I(X;Y) = H(X) + H(Y) - H(X,Y) \tag{35}$$

$$I(X,Y) = H(X) - H(X/Y) \tag{36}$$

$$I(X,Y) = H(Y) - H(Y/X) \tag{37}$$

Here it is important to note that mutual information and entropy are two related concepts.

[3] This agrees with general expectations or intuition. It is natural to conclude that independent variables do not give information about the uncertainties of each other.

3.5. Some Modifications on Mutual Information

In this context, one can refer to Coombs, Daves & Tversky (1970)[4] and Press & Flannery (1988)[5]. Among these modifications offered, the followings are especially important:

$$C_{XY} = \frac{I(X;Y)}{H(Y)} \tag{38}$$

$$C_{YX} = \frac{I(X;Y)}{H(X)} \tag{39}$$

(38) and (39) are not necessarily equal. Therefore a symmetric version is proposed as

$$R = \frac{I(X;Y)}{H(X)+H(Y)}. \tag{40}$$

This is the coefficient of redundancy. It is zero in case of independence, and it takes the value of $\frac{1}{2}$ in case of dependence implying that half of these two variables is redundant. Still another dependency measures are as follows:

$$\frac{I(X;Y)}{\min\{H(X),H(Y)\}}, \frac{I(X;Y)}{H(X,Y)}, \frac{I(X;Y)}{\sqrt{H(X)H(Y)}}$$(Yao(2003)[6], Strehl&Ghosh(2002)[7]).

3.6. Multivariate Generalizations

Suppose the joint probability function of $X_1, X_2, ..., X_n$ be $f(x_1, x_2, ..., x_n)$. The entropy of this joint distribution can be expressed as the sum of entropies of conditional distributions.

$$H\left(X_1, X_2, ..., X_n\right) = \sum_{i=1}^{n} H(X_i/X_{i-1}, ..., X_1). \tag{41}$$

For bivariate distributions

$$H\left(X_1, X_2\right) = H(X_1) + H(X_2/X_1). \tag{42}$$

[4] Coombs, C.H.,Daves,R.M.&Tversky(1970), "Mathematical Psychology: An Elementary Introduction", Prentice-Hall,Englewood Cliffs, NJ

[5] Press, W.H., Flannery, B.P., Teukolsky, S.A.,& Vetterling, W.T.(1988) "Numerical Recipes in C :The Art of Scientific Computing", Cambridge University Press, Cambrige,p.634

Or for trivariate distributions it is straightforward to derive formulas like

$$H\left(X_1, X_2, X_3\right) = H(X_1) + H(X_2, X_3/X_1) \tag{43}$$

$$H\left(X_1, X_2, X_3\right) = H(X_1) + H(X_2/X_1) + H(X_3/X_1, X_2). \tag{44}$$

Similarly whenever Z is given, the conditional measure of information between X and Y can be written as

$$I(X;Y/Z) = H(X/Z) - H(X/Y,Z) \tag{45}$$

$$= E_{P(x,y,z)} \log \frac{P(X,Y/Z)}{P(X/Z)P(Y/Z)}. \tag{46}$$

Also for mutual information measures one can conclude that

$$I(X_1, X_2, ..., X_n; Y) = H(X_1, X_2, ..., X_n)(X_1, X_2, ..., X_n/Y) \tag{47}$$

$$= \sum_{i=1}^{n} H(X_i/X_{i-1}, ..., X_1) - \sum_{i=1}^{n} H(X_i/X_{i-1}, ..., X_1, Y) \tag{48}$$

$$I(X_1, X_2, ..., X_n; Y) = \sum_{i=1}^{n} I(X_i; Y/X_{i-1}, X_{i-2}, ..., X_1). \tag{49}$$

3.7. Kullback-Leibler Divergence[8]

H_1 :Probability function is p.
H_2 :Probability function is q which is different from p $(q \neq p)$.

According to Kullback and Leibler, the divergence between these two hypotheses is

$$D_{KL}(H_1//H_2) = D_{KL}(p//q) = \sum_x p(x) \log \frac{p(x)}{q(x)}. \tag{50}$$

[8] Kullback-Leibler divergence and Jeffreys divergence do not satisfy the requirements in the definition of a metric function. For that reason it is customary to use the term divergence rather than distance.

This statistic can also be evaluated as the measure of error when one adopts q instead of p infact H_1 is true. Besides this statistic can be seen as the average amount of information per observation that supports H_1[8]. As a second example, we consider the following alternatives:

H_1 :X and Y are not independent. (or the joint probability function is $f_{XY}(x,y)$)
H_2 :X and Y are independent (for $\forall(x,y) \in \Re^2$,$f_{XY}(x,y) = f_X(x).f_Y(y)$).

In this test, Kullback-Leibler divergence $D_{KL}(f_{XY}(x,y) \; // \; f_X(x) \; f_Y(y))$ can be evaluated as the average amount of information per observation that supports (H_1). If the bivariate distribution of (X,Y) is jointly continuous

(51)

$$D_{KL}(f_{XY}(x,y)//f_X(x)f_Y(y)) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)\log\left[\frac{f_{XY}(x,y)}{f_X(x)f_Y(y)}\right]dxdy$$

if $D_{KL} = 0$ then the following statements are identical:

1) The amount of information from sample that supports H_1 is zero.
2) When the variables are independent, the amount of information that one can obtain for one variable by observing the other variable is zero.

When (X,Y) is jointly and normally distributed, Kullback-Leibler divergence is found as

$$D_{KL}(f_{XY}(x,y)//f_X(x)f_Y(y)) = -\frac{1}{2}\log(1-\rho^2). \tag{52}$$

In bivariate normal distribution, Kullback-Leibler divergence is a function of linear correlation coefficient ρ . Of course, this result agrees with intuition.

3.8. Jeffreys Divergence

A symmetric version of Kullback-Leibler divergence is proposed by Jeffreys. This measure is

$$D_J(p//q) = \sum_x \left[(p(x) - q(x))\log\frac{p(x)}{q(x)}\right]. \tag{53}$$

Here p and q respresents two discrete probability functions. To investigate the degree of dependence between two continuous variables Jeffreys divergence can be formulated as

$$(54)\quad D_J(...//...) = \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty}(f_{XY}(x,y) - f_X(x)f_Y(y))\log\left[\frac{f_{XY}(x,y)}{f_X(x)f_Y(y)}\right]dxdy.$$

3.9. Asymptotic Properties of D_{KL} and D_J

The asymptotic properties of D_{KL} and D_J are analysed thoroughly. One can refer to [8] to have an overall idea of this topic. Suppose that the likelihood function based on a sample of n units obtained from a qualitative distribution is given by

$$L(\tilde{p}) = p_1^{f_1}....p_k^{f_k}$$

where f_i (i=1,2,. . . ,k) , is the frequency of the category A_i. $(\sum\limits_{i=1}^{k} f_i = n)$ Let the null and alternative hypotheses are defined as

$H_1 : \tilde{p} = \tilde{p}_0$
$H_2 : p_i \neq p_{i0}$ (at least for one i)

and the test statistic or the likelihood ratio be

$$(55)\qquad \Lambda = \frac{L(\tilde{p}_0)}{L(\overset{\frown}{\tilde{p}})} = \prod_{i=1}^{k}\frac{p_{0i}^{f_i}}{(f_i/n)^{f_i}}.$$

Here $\overset{\frown}{\tilde{p}}$; whose components are computed as $\hat{p}_i = \dfrac{f_i}{n}$; is the maximum likelihood estimate of the probabilities vector $\tilde{p}$. Based on (55) one can calculate the test statistic

$$(56)\qquad -2\log\Lambda = -2\sum_{i=1}^{k} f_i\left(\log(p_{0i}) - \log\left(\frac{f_i}{n}\right)\right).$$

Here the distribution of $-2\log\Lambda$ is a chi-square distribution with (k-1) degrees of freedom, asymptotically. k-1 is the number of parameters whose values can be estimated freely under the assumption $H_1 : \tilde{p} = \tilde{p}_0$. Besides it can also be shown that under the validity of H_1 , the distributions of $-2\log\Lambda$ and χ^2 are equal asymptotically [10]. In addition, the statistic

$$(57)\qquad 2n\hat{I} = 2n\left[\int f(x,\theta)\log\frac{f(x,\theta)}{f(x,\theta_2)}d\lambda(x)\right]_{\theta=\hat{\theta}}$$

(under the validity of H_1) fits a chi-square distribution with k (the number of components of the parameter vector) degrees of freedom asymptotically. $f(x,\theta)$ is the joint probability density function having multiple parameters. $\hat{\theta}$ is assumed to be consistent, asymptotically multivariate normal, and efficient random estimator of θ . Finally, θ_2 represents the parameter vector specified by H_1 and $\lambda(x)$ is a probability measure. Similarly

$$n\hat{I} = 2n\left[\int (f(x,\theta) - f(x,\theta_2)) \log \frac{f(x,\theta)}{f(x,\theta_2)} d\lambda(x)\right]_{\theta=\hat{\theta}} \tag{58}$$

fits a chi-square distribution with k degrees of freedom asymptotically. A more detailed discussion on this topic can be found in [9].

3.10. Multinomial Distributions

To test the dependence of two variables in a contingency table, we suppose

$H_1 : p_{ij} \neq p_i\ p_j$ at least one (i,j) (i=1,2,. . . ,n ; j=1,2,. . . ,m)
$H_2 : p_{ij} = p_i\ p_j$ for all (i,j) (i=1,2,. . . ,n ; j=1,2,..,m)

$$\sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij} = 1, \quad p_{ij} > 0, \quad p_i = \sum_{j=1}^{m} p_{ij}, \quad p_j = \sum_{i=1}^{n} p_{ij}$$

$$D_{KL}(H_1//H_2) = \sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij} \log \frac{p_{ij}}{p_i p_j} \tag{59}$$

$$D_J(H_1//H_2) = \sum_{i=1}^{n}\sum_{j=1}^{m} (p_{ij} - p_i p_j) \log \frac{p_{ij}}{p_i p_j} \tag{60}$$

4. Application 1

Figures on people older than 60 years according to Turkish population statistics in 2007 are taken from [6] for illustration. The contingency table is formed by categorizing people according to their gender and age. The aim here is to investigate the dependency of gender and age of the people older than 60 years old in Turkish population. The related distribution and summarizing qualitative association statistics are given in Table1, Table2 and Table 3. As can be concluded easily, there is not a significant association between these two variables.

Table1. Older Turkish Population in 2007 categorized according to their age and gender

Age Group	**Male**	**Female**	**Total**
60-64	981178	1086536	2067714
65-69	781165	917418	1698583
70-74	629241	743836	1373077
75-79	441289	628672	1069961
80-84	212383	366496	578879
85-89	58552	123636	182188
90+	27473	70014	97487
Total	3131281	3936608	7067889

Table2: Association statistics based on chi-square

Chi-square	50415.88
Phi-square	0.007
Pearson p	0.119
Sakoda	0.168
Tschuprow	0.053
Cramér	0.084

Table3: Association statistics based on entropy

H(X)	2.402
H(Y)	0.991
H(X,Y)	3.387
I(X,Y)	0.005
C(X,Y)	0.005
C(Y,X)	0.002
Redundancy	0.001
Kullback-Leibler divergence	0.005
Jeffreys divergence	0.011

5. Application 2

Table 4 is taken from [1]. It is a distribution related to the performance scores of students coming from some selected public and private schools in a special entering examination. In this example, the association between the school type that the students graduate from and the score of that entering exam is studied. Although exam scores are taken on a continuous scale, these scores are categoried as indicated in Table 4. The distributions of students exam scores and the type of the school they graduate are given in Table 4. The summarizing statistics for the association between these two variables are given in Table 5 and Table 6. All association statistics (whether they are based on chi-square value or entropy measures) agree in general.Yet the statistics based on entropy are lower than those found by chi-square value. The reason for this difference should probably lie in the fact that in entropy based statistics one has to deal with logarithmic scales. Therefore this difference should have been originated from the different methods applied in transforming frequencies.

Table 4: The joint distribution of school type and exam scores of some selected students

X/Y	0-275	276-350	351-425	426-500	Total
Private school	6	14	17	9	46
Public school	30	32	17	3	82
Total	36	46	34	12	128

X=schooltype, Y=examscore

Table5: Association statistics based on chi-square

Chi-square	17.286
Phi-square	0.135
Pearson p	0.345
Sakoda	0.49
Tschuprow	0.28
Cramér	0.367

Table6: Association statistics based on entropy

H(X)	0.942
H(Y)	1.873
H(X,Y)	2.717
I(X,Y)	0.098
C(X,Y)	0.052
C(Y,X)	0.104
Redundancy	0.035
Kullback-Leibler divergence	0.099
Jeffreys divergence	0.207

6. Conclusion

For a detailed exposition of concepts derived from statistical entropy and their applications in statistics and probability one can consult [2], [11], [12] and [13]. In this study, we tried to emphasize that entropy based association measures can also be used in determining the degree of qualitative association between variables. Entropy-based association measures can easily be adapted to contingency tables as well as other statistics used in qualitative association. To compare, if the variables are independent, then all these measures (whether they are based on entropy measures or on other measures such as chi-square values) produce similar results. On the other hand, if the variables are associated to some extent, entropy-based measures and other measures differ or diverge to some moderate extent. The reason for this difference might lie in the fact that in entropy-based measures one uses logarithmic transformations of frequencies (or probabilities) which probably brings a serious scale change. Finally entropy-based measures can easily be adapted to multivariate distributions which is a positive factor for these measures.

References

1. Conover, W.J., Practical Nonparametric Statistics, Wiley Series in Probability and Statistics , Third Edition, 230-234, 1999

2. Cover, T.M.; Thomas, J.A., Elements of Information Theory, Wiley Interscience (Second Edition), Hoboken, New Jersey, 2006

3. Everitt,B.S., The Cambridge Dictionary of Statistics , Cambridge University Press (Third Edition), Cambridge, 2006

4. Evren A. , Entropinin İstatistik'teki Bazı Uygulamaları, II. Ulusal Konya Ereğli Kemal Akman Meslek Yüksek Okulu Tebliğ Günleri, 13-14 Mayıs 2010, Sayı 2: No:1-7, 414-428, 2010

5. Evren A, İstatistik'te Entropiye Dayalı Uyum Ölçülerinin Diğer Uyum Ölçüleri ile Kıyaslanması, 7. İstatistik Günleri Sempozyumu, Bildiri Tam Metinleri Kitabı, , Orta Doğu Teknik Üniversitesi, Ankara, 58-67,28-30 Haziran 2010

6. İstatistiklerle Türkiye 2008, Türkiye İstatistik Kurumu,Ankara, 2008.

7. Keeping,E.S., Introduction to Statistical Inference,Dover Publications, New York, 1995, 314-315

8. Kullback, S., Information Theory and Statistics, Dover Publications, New York, 8-100,1996

9. Liebetrau, A.M., Measures of Association, Series Quantitative Applications in the Social Sciences, a Sage University Paper, 3-16, USA,1983

10. Lindgren,B.W., Statistical Theory , Chapman&Hall/CRC,USA,366, 1993

11. Rényi, A., Probability Theory, Dover Publications, New York,2000

12. Rényi, A., Foundations of Probability, Dover Publications, New York, 2000

13. Reza,Fazlollah M., An Introduction to Information Theory, Dover Publications, New York, 1994

14. Upton, G.; Cook, I., Oxford Dictionary of Statistics, Oxford University Press (Second edition), NewYork, 2006

15. http://en.wikipedia.org/wiki/Mutual_information