BULLETIN *of the*
MALAYSIAN MATHEMATICAL
SCIENCES SOCIETY
http://math.usm.my/bulletin

Bull. Malays. Math. Sci. Soc. (2) **36***(4) (2013), 985–1008*

Estimation and Test of Measures of Association for Correlated Binary Data

[1]AHMED M. EL-SAYED, [2]M. ATAHARUL ISLAM AND [3]ABDULHAMID A. ALZAID
[1,2,3]King Saud University, College of Science, Department of Statistics and OR, P.O.Box 2455, Riyadh 11451, Saudi Arabia
[2]Department of Applied Statistics, East West University, Aftabnagar, Dhaka 1212, Bangladesh
[1]atabl@ksu.edu.sa, [2]mataharul@yahoo.com, [3]alzaid@ksu.edu.sa

Abstract. This paper provides the estimation and test procedures for measures of association in the correlated binary data. Several measures of association are proposed for bivariate Bernoulli data during the past decades but the estimation and test procedures for most of these measures are not developed yet. In this paper, the inferential procedures for the measures of association are demonstrated. The generalized linear model approach (GLM) is employed for bivariate Bernoulli variables and the measures of association are estimated and appropriate test procedures are suggested. An alternative to the quadratic exponential form (QEF) is proposed to improve the normalization process. In this paper, different methods of measuring association between bivariate Bernoulli variables are compared. For comparison, we use a simulation procedure which indicates that all the measures of association and their test procedures provide almost similar results. However, the GLM and the proposed alternative quadratic exponential form (AQEF) models display slightly better performance.

2010 Mathematics Subject Classification: 62H20, 62F03, 62F10, 62N03

Keywords and phrases: Correlated binary data, bivariate Bernoulli outcomes, generalized linear models (GLMs), maximum likelihood estimators (MLEs), likelihood ratio test (LRT), deviance test, dispersion parameter, BB package.

1. Introduction

The dependence in outcome variable in non-normal situations gained importance during the recent past due to the wide range of applications to various fields of research. Some methods had been proposed in the past to measure the association in correlated binary data.

Marshall and Olkin [19] explained how some bivariate distributions can be generated by the bivariate Bernoulli distribution. Gourieroux *et al.* [11] showed the quadratic exponential model to be unique in obtaining the maximum likelihood estimates of mean and covariance parameters. For any member of the family, the estimators are consistent and asymptotically normal under regularity conditions. This procedure is referred to as the pseudo-maximum likelihood estimation to emphasize the distinction between the score generated and actual

Communicated by V. Ravichandran.
Received: May 22, 2012; *Revised:* August 1, 2012.

sampling of the likelihood functions. Dale [8] expressed the association between components in terms of global cross-ratios, cross-product ratios of quadrant probabilities, for each double dichotomy of the response table of probabilities into quadrants. The generalized estimating equations (GEE) proposed by Liang and Zeger [16], Zeger and Liang [26] have generated considerable attention in the last two decades and several extensions have been developed. Bonney [4] expressed the likelihood of a set of binary dependent outcomes, with or without explanatory variables, as a product of conditional probabilities each of which is assumed to be logistic, this model is called the regressive logistic model. Zhao and Prentice [27] employed a pseduo-maximum likelihood for analyzing correlated binary responses. Their parametrization is based on a simple pairwise model in which the association between responses is modeled in terms of correlations. Fitzmaurice and Liard [9] discussed a likelihood-based method, based on a multivariate model and used the conditional log odds-ratios. With this approach, the higher-order associations can be incorporated in a natural way. Cessi and Houwelingen [5] presented the logistic regression for binary data in such a way that the marginal response probabilities are logistic too. They used the odds ratio and tetrachoric correlation and compared between them as association measures for the dependence between correlated observations. Cox and Wermuth [6] studied the joint distribution of p binary variables in the quadratic exponential form containing only the mean effects and two-factor interactions in the log probabilities. They have some approximate versions of marginalized forms of the distribution. Glonek and McCullagh [10] have given a general definition when data are comprised of several categorical responses together with categorical or continuous predictors observed, particularly suitable for relating the joint distribution of the responses to predictors. Also, they have used a computational scheme for performing maximum likelihood estimation for data sets of moderate size. Heagerty [13] developed a general parametric class of serial dependence models that permits the likelihood based marginal regression analysis of binary response data. Lovison [17] proposed a matrix-valued Bernoulli distribution, based on the log linear representation introduced by Cox [7], for the multivariate Bernoulli distribution with correlated components. Islam *et al.* [14] developed a new simple procedure to take account of the bivariate binary model with covariate dependence. The model is based on the integration of conditional and marginal models.

This paper provides the estimation and test procedures for various measures of association. A generalized linear model for bivariate Bernoulli data is proposed in this study and is compared with the alternative procedures. For estimation, the likelihood and pseudo-likelihood methods are used. Also, for testing the parameter for the measure of association, we employ the likelihood ratio test (LRT). The goodness of fit of the proposed models are compared using the deviance function.

In this paper, the major works on the measures of association stemming from the bivariate Bernoulli data are presented in Sections 2 to 9. Each section includes joint probability mass function, the log likelihood function, estimation of the association parameter, and the testing of hypothesis for dependence in the bivariate Bernoulli outcomes. These estimation and test procedures have been proposed under a general framework using the likelihood methods. It is noteworthy that in Section 9, we introduce an alternative to the measure based on the quadratic exponential form to make it more realistic in terms of defining the underlying pseudo likelihood function by modifying the normalizing procedure. In Section 10, a numerical comparison among all the measures of association have been demonstrated using a simulation study.

2. The Marshall and Olkin measure

Marshall and Olkin [19] showed that some distributions can be generated by using the bivariate Bernoulli distribution. If there are two correlated binary variables Y_1 and Y_2 that follow Bernoulli distributions, each of which taking the value of either 0 or 1, then it must be that (y_1, y_2) can take only the four possible values $(0,0),(0,1),(1,0),(1,1)$. Table below displays notations for the joint, conditional and marginal distributions for correlated variables Y_1 and Y_2.

Table 1. Joint, conditional and marginal probabilities for correlated binary variables Y_1, Y_2

Outcomes	$Y_2 = 0$	$Y_2 = 1$	Total
$Y_1 = 0$	p_{00}	p_{01}	$1 - p_1$
$Y_1 = 1$	p_{10}	p_{11}	p_1
Total	$1 - p_2$	p_2	1

From Table 1, we can express the joint probability mass function for the two variables Y_1 and Y_2 as

$$f(y_1, y_2) = p_{00}^{(1-y_1)(1-y_2)} p_{01}^{(1-y_1)y_2} p_{10}^{y_1(1-y_2)} p_{11}^{y_1 y_2}, \quad y_1, y_2 = 0, 1, \tag{2.1}$$

with the constraint $\sum_{r=0}^{1} \sum_{s=0}^{1} p_{rs} = 1$. It is evident from Table 1, the marginal probabilities can be expressed as

$$p_1 = p_{10} + p_{11}, \qquad p_2 = p_{01} + p_{11}. \tag{2.2}$$

The joint probabilities can be expressed in terms of the marginal and conditional probabilities as follows

$$Pr(Y_i = s, Y_j = r) = Pr(Y_i = s \mid Y_j = r) Pr(Y_j = r), \qquad i, j = 1, 2, \quad r, s = 0, 1, \tag{2.3}$$

or directly from Table 1, we have

$$p_{10} = p_1 - p_{11}, \qquad p_{01} = p_2 - p_{11}, \qquad p_{00} = 1 - p_{10} - p_{01} - p_{11}. \tag{2.4}$$

Also, the conditional probabilities can be shown as

$$Pr(Y_i = s \mid Y_j = r) = \frac{Pr(Y_i = s, Y_j = r)}{Pr(Y_j = r)}, \qquad i, j = 1, 2, \qquad r, s = 0, 1. \tag{2.5}$$

We can define the covariance between Y_1 and Y_2 from Table 1 as

$$\mathrm{Cov}(Y_1, Y_2) = \sigma_{12} = p_{11} - p_1 p_2 = p_{11} p_{00} - p_{01} p_{10}, \qquad -\infty \leq \sigma_{12} \leq \infty. \tag{2.6}$$

The correlation between Y_1 and Y_2, as a measure of association, is

$$\mathrm{Corr}(Y_1, Y_2) = \rho = \frac{p_{11} - p_1 p_2}{\sqrt{p_1 p_2 (1 - p_1)(1 - p_2)}}, \qquad -1 \leq \rho \leq 1. \tag{2.7}$$

where ρ takes 0, when $\sigma_{12} = 0$ or $p_{11} = p_1 p_2$, this means that Y_1 and Y_2 are independent. For binary responses the cross-ratio reduces to the odds ratio. So, we can use the odds ratio as measure of association. Using Table 1, the odds ratio can be defined as

$$\psi = \frac{Pr(Y_2 = 1 \mid Y_1 = 1)}{Pr(Y_2 = 1 \mid Y_1 = 0)} = \frac{p_{11}}{p_{10}} \div \frac{p_{01}}{p_{00}} = \frac{p_{00} p_{11}}{p_{10} p_{01}} = \frac{p_{11}(1 - p_1 - p_2 + p_{11})}{(p_1 - p_{11})(p_2 - p_{11})}, \quad \psi \geq 0. \tag{2.8}$$

The variables Y_1 and Y_2 are independent if $\psi = 1$, positive association if $\psi > 1$ and negative association if $\psi < 1$.
The relationship between the correlation ρ and the odds ratio ψ can be determined using (2.7) and (2.8) as

(2.9)
$$\psi = \frac{(p_1 p_2 + \rho\sqrt{p_1 p_2(1-p_1)(1-p_2)})(1-p_1-p_2+p_{11})}{(p_1-p_{11})(p_2-p_{11})}, \qquad \psi \geq 0,$$
$$\rho = \frac{\psi(p_1-p_{11})(p_2-p_{11}) - p_1 p_2(1-p_1-p_2+p_{11})}{(1-p_1-p_2+p_{11})\sqrt{p_1 p_2(1-p_1)(1-p_2)}}, \qquad -1 \leq \rho \leq 1.$$

From the equation (2.7), the joint probability p_{11} can be defined as

(2.10)
$$p_{11} = p_1 p_2 + \rho\sqrt{p_1 p_2(1-p_1)(1-p_2)}, \qquad p_{11} \geq 0.$$

2.1. Estimation

Let us define the cell frequencies by $n_{rs}(r,s=0,1)$ and the total sample size is $n = \sum_{r=0}^{1}\sum_{s=0}^{1} n_{rs}$. So, we can display these frequencies in Table 2 as:

Table 2. Observed cell frequencies from a bivariate Bernoulli distribution outcomes

Outcomes	$Y_2 = 0$	$Y_2 = 1$	Total
$Y_1 = 0$	n_{00}	n_{01}	$n - n_1$
$Y_1 = 1$	n_{10}	n_{11}	n_1
Total	$n - n_2$	n_2	n

In this section we use the invariant property of the maximum likelihood estimators. The MLEs of marginal probabilities are

(2.11)
$$\hat{p}_1 = \frac{n_1}{n}, \qquad \hat{p}_2 = \frac{n_2}{n}.$$

The MLEs of joint probabilities are

(2.12)
$$\hat{p}_{rs} = \frac{n_{rs}}{n}, \qquad \hat{p}_{rs} \geq 0, \qquad r,s = 0,1.$$

If Y_1 and Y_2 are independent, then

(2.13)
$$\hat{p}_{11} = \hat{p}_1 \hat{p}_2 = \frac{n_1 n_2}{n^2}.$$

The MLE of correlation ρ is

(2.14)
$$\hat{\rho} = \frac{\hat{p}_{11} - \hat{p}_1\hat{p}_2}{\sqrt{\hat{p}_1\hat{p}_2(1-\hat{p}_1)(1-\hat{p}_2)}}, \quad -1 \geq \hat{\rho} \geq 1.$$

The MLE of odds ratio ψ is

(2.15)
$$\hat{\psi} = \frac{\hat{p}_{11}(1-\hat{p}_1-\hat{p}_2+\hat{p}_{11})}{(\hat{p}_1-\hat{p}_{11})(\hat{p}_2-\hat{p}_{11})}, \quad \hat{\psi} \geq 0.$$

As mentioned before, the independence between Y_1 and Y_2 can be observed if $\hat{\psi} = 1$. We can take the natural logarithm of $\hat{\psi}$ and take its expectation to get $E(\log\hat{\psi}) = \log\psi$. Then, the asymptotic variance of $\log\hat{\psi}$, (See Agresti [1]), is

$$\operatorname{Var}(\log\hat{\psi}) = \sum_{r=0}^{1}\sum_{s=0}^{1}\frac{1}{n_{rs}} = \left[\frac{1}{n_{00}} + \frac{1}{n_{01}} + \frac{1}{n_{10}} + \frac{1}{n_{11}}\right]. \tag{2.16}$$

So, $\log\hat{\psi}$ is approximately distributed as $N[\log\psi, \operatorname{Var}(\log\hat{\psi})]$. The normal approximation can be used to obtain the confidence interval

$$\log\hat{\psi} \pm Z_{\frac{\alpha}{2}}\sqrt{\operatorname{Var}(\log\hat{\psi})}. \tag{2.17}$$

Then, we can exponentiate it to obtain a confidence interval for odds ratio ψ.

2.2. Test of hypothesis

In this subsection, we use three tests. The first one is for testing the independence or dependence of the two variables Y_1 and Y_2 using the likelihood ratio test (LRT) and comparing it with the Chi-square with one degree of freedom. The second one is for testing the adequacy of the model using the deviance test and comparing it with the Chi-square with $(n-p)$ degrees of freedom, where p is the number of parameters estimated. The third one is used to estimate the dispersion parameter ϕ as a goodness of fit measure. In this case, we expect this estimate close to one. But for Bernoulli data, the estimate $\hat{\psi}$ can be more than one indicating the over-dispersion. It can be shown from the exponential family form, and using the following relationship

$$\operatorname{Var}(Y) = \operatorname{Var}(\mu)\phi, \qquad \operatorname{Var}(Y) = \mu(1-\mu), \tag{2.18}$$

if $\phi \geq 1$, then $\operatorname{Var}(Y) \geq \operatorname{Var}(\mu)$, where, μ is $E(Y)$. So, using the joint function (2.1), we can get, for n observations, the log-likelihood function as

$$\ell(y_i;p) = \sum_{i=1}^{n}\Big(y_{00i}\log p_{00} + y_{01i}\log p_{01} + y_{10i}\log p_{10} + y_{11i}\log p_{11}\Big). \tag{2.19}$$

Using the log-likelihood function (2.19) and the estimate $\hat{p}_{11}$ under H_0 which could be changed according to the value ρ_0 from the equation (2.10), we can test the independence or specified values of the correlation or odds ratio for the two variables Y_1 and Y_2. The null hypothesis can be expressed as $H_0 : \rho = \rho_0$ or $H_0 : \psi = \psi_0$ against the alternative hypothesis $H_1 : \rho \neq \rho_0$ or $H_1 : \psi \neq \psi_0$. Using the log-likelihood function (2.19), we can use the likelihood ratio test (LRT) as

$$LRT = -2\Big[\ell(y_i;\psi_0,\rho_0) - \ell(y_i;\hat{\psi},\hat{\rho})\Big] \quad \sim \quad \chi^2_1. \tag{2.20}$$

The deviance function as a way of assessing the goodness of fit for the model which was proposed by McCullagh and Nelder [18], for the univariate case, we can extend it in the bivariate case as follows:

$$\begin{aligned} D &= 2\Big[\ell(y_i,y_i) - \ell(y_i;\hat{p})\Big] \\ &= 2\sum_{i=1}^{n}\Big(y_{00i}\log\frac{y_{00i}}{\hat{p}_{00}} + y_{01i}\log\frac{y_{01i}}{\hat{p}_{01}} + y_{10i}\log\frac{y_{10i}}{\hat{p}_{10}} + y_{11i}\log\frac{y_{11i}}{\hat{p}_{00}}\Big) \sim \chi^2_{n-p} \end{aligned} \tag{2.21}$$

where,

$$\ell(y_i;y_i)=\sum_{i=1}^{n}\Big(y_{00i}\log y_{00i}+y_{01i}\log y_{01i}+y_{10i}\log y_{10i}+y_{11i}\log y_{11i}\Big),$$

is the log-likelihood function for the saturated model evaluated at observed values y_i, and

$$\ell(y_i;\hat{p})=\sum_{i=1}^{n}\Big(y_{00i}\log\hat{p}_{00}+y_{01i}\log\hat{p}_{01}+y_{10i}\log\hat{p}_{10}+y_{11i}\log\hat{p}_{11}\Big),$$

is the log-likelihood function for the model of interest evaluated at maximum likelihood estimates $\hat{p}_{rs}(r,s=0,1)$.
The estimate of dispersion parameter ϕ is

$$\hat{\phi}=\frac{1}{n-p}\sum_{i=1}^{n}\sum_{j=1}^{2}\frac{y_{ji}-\hat{p}_j}{\hat{p}_j(1-\hat{p}_j)},\qquad \hat{\phi}\geq 1. \tag{2.22}$$

The square root of the of the dispersion parameter ϕ is called the scale parameter.

3. The Dale measure

Based on the Marshall and Olkin measure [19], Dale [8] presented a flexible class of measure for the bivariate, discrete, ordered responses. The Global cross-ratio (GCR) models exploit the ordering of the marginal response variables, since the association between them is defined in terms of quadrant probabilities. The GCR may be interpreted as a ratio of odds on conditional events. The joint probability mass function for the two variables, Y_1 and Y_2, as shown in (2.1). Using Table 1 and equation (2.8), the joint probability p_{11} can be expressed in terms of p_1, p_2 and ψ as

$$p_{11}=\frac{\psi(p_1-p_{11})(p_2-p_{11})}{1-p_1-p_2+p_{11}}=\frac{\psi(p_1p_2-p_1p_{11}-p_2p_{11}+p_{11}^2)}{1-p_1-p_2+p_{11}}. \tag{3.1}$$

With some algebraic manipulation on (3.1), we have

$$p_{11}^2(1-\psi)+p_{11}[1+(p_1+p_2)(\psi-1)]-\psi p_1p_2=0, \tag{3.2}$$

setting

$$A=1-\psi,\qquad B=1+(p_1+p_2)(\psi-1),\qquad C=-\psi p_1p_2, \tag{3.3}$$

using the relationship: $\dfrac{-B\pm\sqrt{B^2-4AC}}{2A}$, we have

$$p_{11}=\begin{cases}\frac{1}{2}(\psi-1)^{-1}[a-\sqrt{a^2+b}], & \psi\neq 1\\ p_1p_2, & \psi\neq 1,\end{cases} \tag{3.4}$$

where, $a=1+(p_1+p_2)(\psi-1)$ and $b=4\psi(1-\psi)p_1p_2$.
The other joint probabilities can be obtained as

$$p_{10}=p_1-p_{11},\qquad p_{01}=p_2-p_{11},\qquad p_{00}=1-p_1-p_2+p_{11}. \tag{3.5}$$

One of the drawbacks of such formulation (3.4), is that it employs a single root from a quadratic equation of p_{11}. The argument behind that is the value of p_{11} can never be negative and odds ratio satisfies $\psi\geq 0$. But the same assumptions are also true for some of the values of other root. Therefore it seems better to have the form that uses the possible root which satisfies the same assumptions as

$$p_{11} = \begin{cases} \frac{1}{2}(\psi-1)^{-1}[a \pm \sqrt{a^2+b}], & \psi \neq 1 \\ p_1 p_2, & \psi \neq 1, \end{cases} \tag{3.6}$$

substituting by the values of a and b in the equation (3.4), we get

$$p_{11} = \begin{cases} \frac{1}{2}(\psi-1)^{-1}[1+(p_1+p_2)(\psi-1) \\ -\sqrt{[1+(p_1+p_2)(\psi-1)]^2+4\psi(1-\psi)p_1p_2}], \quad \psi \neq 1 \\ p_1p_2, \quad \psi \neq 1. \end{cases} \tag{3.7}$$

3.1. Estimation

The log-likelihood function, for n observations, is

$$\ell(y_i;p) = \sum_{i=1}^{n}\sum_{r=0}^{1}\sum_{s=0}^{1} y_{rsi}\log p_{rs}, \qquad r,s = 0,1. \tag{3.8}$$

Taking the first order derivative of the log-likelihood function (3.8), with respect to p_{10}, p_{01} and p_{11}, and put and equating to zero, we have

$$\begin{aligned} c\frac{\partial \ell(y_i;p)}{\partial p_{10}} &= \sum_{i=1}^{n}\left(\frac{y_{10i}}{p_{10}} - \frac{y_{00i}}{p_{00}}\right) = 0, \\ \frac{\partial \ell(y_i;p)}{\partial p_{01}} &= \sum_{i=1}^{n}\left(\frac{y_{01i}}{p_{01}} - \frac{y_{00i}}{p_{00}}\right) = 0, \\ \frac{\partial \ell(y_i;p)}{\partial p_{11}} &= \sum_{i=1}^{n}\left(\frac{y_{11i}}{p_{11}} - \frac{y_{10i}}{p_{10}} - \frac{y_{01i}}{p_{01}} + \frac{y_{00i}}{p_{00}}\right) = 0. \end{aligned} \tag{3.9}$$

Solving the estimating equations (3.9) and using the equation (3.5), the estimates $\hat{p}_1, \hat{p}_2$ and $\hat{p}_{11}$ can be obtained, and then we can get the estimates $\hat{p}_{10} = \hat{p}_1 - \hat{p}_{11}, \hat{p}_{01} = \hat{p}_2 - \hat{p}_{11}, \hat{p}_{00} = 1 - \hat{p}_1 - \hat{p}_2 + \hat{p}_{11}$. The estimate $\hat{\psi}$ can be determined using the equation (2.8). These estimates are convenient for the correlation and odds ratio, just for the independence case, specially for large samples. Alternatively, to avoid the effect of ignorance of differentiation of the log-likelihood function (3.9) with respect to p_{00}, and also because of the fact that the model is related to the Marshall and Olkin procedure, we can get all the previous estimates by the same procedure as of the Marshall and Olkin measure as shown before.

3.2. Test of hypothesis

We can use the equations (2.20) to test for the independence or specified values of odds ratio as a measure of association for the two variables Y_1 and Y_2. The null hypothesis in this case can be expressed as $H_0 : \psi = \psi_0$ against the alternative hypothesis $H_1 : \psi \neq \psi_0$. The estimate $\hat{p}_{11}$ under H_0 should be changed according to the value of ψ in the equations (3.7). The equation (2.21) can be used for the deviance function similar to the Marshal and Olkin measure. Finally, the equation (2.22) is also used to determine the estimate of dispersion parameter ϕ.

4. The Cessi and Houwelingen measure

Cessi and Houwelingen [5] proposed different measures of association for correlated binary data such as tetrachoric correlation and odds ratio. The joint probability mass function for the two variables Y_1 and Y_2 can be expressed as shown in the equation (2.1).

4.1. Estimation

The log-likelihood function for n observations is as shown in (3.8). Using the relationship (3.5), we can differentiate the log-likelihood function with respect to p_1, p_2 and p_{11}; this yields

$$\begin{aligned} c\frac{\partial \ell(y_i;p)}{\partial p_1} &= \sum_{i=1}^{n}\left(\frac{y_{10i}}{p_1-p_{11}} - \frac{y_{00i}}{1-p_1-p_2+p_{11}}\right)=0, \\ \frac{\partial \ell(y_i;p)}{\partial p_2} &= \sum_{i=1}^{n}\left(\frac{y_{01i}}{p_2-p_{11}} - \frac{y_{00i}}{1-p_1-p_2+p_{11}}\right)=0, \\ \frac{\partial \ell(y_i;p)}{\partial p_{11}} &= \sum_{i=1}^{n}\left(\frac{y_{11i}}{p_{11}} - \frac{y_{10}}{p_1-p_{11}} - \frac{y_{01i}}{p_2-p_{11}} + \frac{y_{00i}}{1-p_1-p_2+p_{11}}\right)=0. \end{aligned} \tag{4.1}$$

Solving the estimating equations (4.1), we can obtain directly the estimates $\hat{p}_1, \hat{p}_2$ and $\hat{p}_{11}$. Alternatively, we can use the Marshall and Olkin procedure to estimate all parameters to avoid the differentiation of the log-likelihood function with respect to p_{00}.

4.2. Test of hypothesis

To test the independence or specified values of the odds ratio, by the null hypothesis H_0 : $\psi = \psi_0$, we can use the LRT as in equation (2.20). The estimate $\hat{p}_{11}$ under H_0 should be changed according to ψ in the equations (2.8). The deviance function as in the equation (2.21) can be used to determine the adequacy of the model, the difference is made by employing the relationship (3.5) to get the deviance function

$$\begin{aligned} D = 2\sum_{i=1}^{n}\Bigg(& y_{00i}\log\frac{y_{00i}}{1-\hat{p}_1-\hat{p}_2+\hat{p}_{11}} + y_{01i}\log\frac{y_{01i}}{\hat{p}_2-\hat{p}_{11}} \\ & + y_{10i}\log\frac{y_{10i}}{\hat{p}_1-\hat{p}_{11}} + y_{11i}\log\frac{y_{11i}}{\hat{p}_{11}}\Bigg) \sim \chi^2_{n-p}. \end{aligned} \tag{4.2}$$

Finally, we can use the equation (2.22) to estimate the dispersion parameter ϕ.

The dependence between two variables Y_1 and Y_2, can be quantified in different ways. So, in the next three subsections we will explain the odds ratio, tetrachoric correlation as measures of association and compare between them as shown in the following subsections.

4.3. Odds ratio

The first method is to characterize the association in Table 1 by the odds ratio. This measure is used by, for example, Dale [8]. Since, the odds ratio as shown in (2.8) is restricted, $\psi \geq 0$, and we will take $\log\psi = \psi_{12}$ to overcome this restriction. The joint probability p_{11} can be expressed in terms of marginal probabilities p_1, p_2 and ψ as shown in the equation (3.7).

The test statistic for testing whether or not $\psi = 1$ equivalently $\log\psi = \psi_{12} = 0$ is derived as

$$W = \frac{\left[\sum_{i=1}^{n}(y_{1i}-\hat{p}_1)(y_{2i}-\hat{p}_2)\right]^2}{\sum_{i=1}^{n}\hat{p}_1\hat{p}_2(1-\hat{p}_1)(1-\hat{p}_2)} \sim \chi_1^2. \tag{4.3}$$

If there is independence, we would expect W to be around one, whereas if there is no independence, we expect W to be larger [See the results in Table 7]. The score statistic W has a disadvantage that it is used only for the independence case, so the LRT is better than the score statistic W, because the LRT deals with both the independence and non-independence cases.

4.4. Tetrachoric correlation

The second method as a measure of association is a tetrachoric correlation. The use of this measure goes back to Pearson [22]. The multivariate generalization was introduced by Ashford and Sowden [2]. Cessi and Houwelingen [5] followed their approach but used the logistic marginals instead of the probit marginals. The general idea assumes that the outcomes (y_1, y_2) are realizations of a pair of latent (hidden) continuous variables Z_1 and Z_2, where Z_1 and Z_2 are bivariate standard normal distributions with correlation ρ. The variables Y_1 and Y_2 takes 1, if $Z_j < g_j$ with $g_j = \Phi^{-1}(p_j), j = 1,2$, where Φ is the standard normal cumulative distribution function. This means that

$$p_{11} = Pr(Z_1 < g_1, Z_2 < g_2) = \int_{-\infty}^{g_1}\int_{-\infty}^{g_2} f(t_1,t_2)dt_2dt_1, \tag{4.4}$$

where,

$$f(t_1,t_2) = \frac{1}{2\pi\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2(1-\rho^2)}(t_1^2+t_2^2-2\rho t_1t_2)\right\} \tag{4.5}$$

is the joint density function of the bivariate standard normal distribution, with tetrachoric correlation ρ as a measure of dependence between Y_1 and Y_2. Stuart and Ord [23] showed how p_{11} can be evaluated by Hermite polynomials, and the first-order derivative of p_{11} with respect to ρ is $f(g_1, g_2, \rho)$.

The score statistic to test whether or not $\rho = 0$, is quite easy, since $\frac{\partial p_{11}}{\partial\rho}|_{\rho=0} = f(g_1)f(g_2)$, where $f(g)$ is the univariate standard normal density function. This yields a score statistic

$$U = \sum_{i=1}^{n}\frac{(y_{1i}-\hat{p}_1)(y_{2i}-\hat{p}_2)f(g_1)f(g_2)}{\hat{p}_1\hat{p}_2(1-\hat{p}_1)(1-\hat{p}_2)}, \text{ with } \mathrm{Var}(U) = \sum_{i=1}^{n}\frac{f^2(g_1)f^2(g_2)}{\hat{p}_1\hat{p}_2(1-\hat{p}_1)(1-\hat{p}_2)}. \tag{4.6}$$

Testing whether or not $\rho = 0$ can be done, (see Cessi and Houwelingen [5]), by a score statistic

$$M = \frac{U^2}{\mathrm{Var}(U)} \sim \chi_1^2. \tag{4.7}$$

Similar to the score statistic W, we expect that M should be around one, if there is independence according to the expressions (4.6), whereas if there is lack of independence, we expect M to be larger. The score statistic M also has the same disadvantage as the score statistic W that both of them is used just for the independence case.

4.5. Relationship between odds ratio and tetrachoric correlation

Comparing between the two measures of association in the last two subsections, by considering the estimate of $\tilde{\psi}_{12}$ is approximately given by

$$\tilde{\psi}_{12} = \frac{\left(\sum_{i=1}^{n}(y_{1i}-\hat{p}_1)(y_{2i}-\hat{p}_2)\right)}{\sum_{i=1}^{n}\hat{p}_1\hat{p}_2(1-\hat{p}_1)(1-\hat{p}_2)}. \tag{4.8}$$

Also, the estimate of tetrachoric $\tilde{\rho}$ is approximately given by

$$\tilde{\rho} = \frac{\left(\sum_{i=1}^{n}(y_{1i}-\hat{p}_1)(y_{2i}-\hat{p}_2)w(\hat{p}_1)w(\hat{p}_2)\right)}{\sum_{i=1}^{n}\hat{p}_1\hat{p}_2(1-\hat{p}_1)(1-\hat{p}_2)w^2(\hat{p}_1)w^2(\hat{p}_2)}, \quad w(\hat{p}_j) = \frac{\Phi^{-1}(\hat{p}_j)}{\hat{p}_j(1-\hat{p}_j)}, \quad j=1,2. \tag{4.9}$$

Both approximations are weighted by the cross-products $(y_{1i}-\hat{p}_1)(y_{2i}-\hat{p}_2)$. The approximate relationship between $\tilde{\rho}$ and $\tilde{\psi}_{12}$ is given, [Cessi and Houwelingen [5]], by

$$\tilde{\psi}_{12} = (1.7)^2\tilde{\rho} \tag{4.10}$$

This relationship in our study is true only in the independence case [see Table 7].

5. The Teugels measure

Based on the Marshall and Olkin measure [19], Teugels [24] established the multivariate but vectorized versions for Bernoulli and the binomial distributions using the concept of Kronecker product for matrix calculus. The multivariate Bernoulli distribution entails a parameterized model that provides an alternative to the traditional log-linear model for binary variables. If $Y_j (j=1,2)$ is a sequence of Bernoulli random variables, where

$$Pr(Y_j=1)=p_j \quad \text{and} \quad Pr(Y_j=0)=q_j, \quad 0\le q_j = 1-p_j \le 1, \quad j=1,2. \tag{5.1}$$

The joint probabilities can be displayed same as in the equation (2.1). The expected values of Y_1 and Y_2 are $E(Y_1)=p_1$, $E(Y_2)=p_2$, respectively. Also, the covariance between them is $\sigma_{12} = E\left[(Y_1-p_1)(Y_2-p_2)\right]$. Also, we can use $E(Y_1Y_2)=p_{11}=\sigma_{12}+p_1p_2$.
Solving for p_{00}, p_{01}, p_{10} and p_{11} we get the following relations

$$p_{00}=q_1q_2+\sigma_{12}, \quad p_{10}=p_1q_2-\sigma_{12}, \quad p_{01}=q_1p_2-\sigma_{12}, \quad p_{11}=p_1p_2+\sigma_{12}. \tag{5.2}$$

The correlation between Y_1 and Y_2 as a measure of association is

$$\rho = \frac{\sigma_{12}}{\sqrt{p_1p_2q_1q_2}}, \quad -1\le\rho\le 1. \tag{5.3}$$

where ρ takes 0 when $\sigma_{12}=0$, this means that Y_1 and Y_2 are independent. The odds ratio can be expressed as shown below using the equations (2.8) and (5.2)

$$\psi = \frac{(q_1q_2+\sigma_{12})(p_1p_2+\sigma_{12})}{(p_1q_2-\sigma_{12})(q_1p_2-\sigma_{12})}, \quad \psi\ge 0. \tag{5.4}$$

Substituting (5.3) in (5.4), we have the relationship between ρ and ψ as

$$\psi = \frac{(q_1q_2+\rho\sqrt{p_1p_2q_1q_2})(p_1p_2+\rho\sqrt{p_1p_2q_1q_2})}{(p_1q_2-\rho\sqrt{p_1p_2q_1q_2})(q_1p_2-\rho\sqrt{p_1p_2q_1q_2})}, \quad \psi\ge 0. \tag{5.5}$$

From (5.5), it can be shown that the variables Y_1 and Y_2 are independent if $\rho=0$ or $\psi=1$.

5.1. Estimation

For n observations, we can use the log-likelihood function (2.19), to derive the first derivatives with respect to p_1, p_2 and p_{11}, and put it equal to zero. Using the equation (5.2), we have the following estimating

$$
\begin{aligned}
c\frac{\partial \ell(y_i; p, \sigma_{12})}{\partial p_1} &= \sum_{i=1}^{n}\Big(\frac{y_{10i}}{p_1q_2-\sigma_{12}} - \frac{y_{00i}}{q_1q_2+\sigma_{12}}\Big) = 0,\\
\frac{\partial \ell(y_i; p, \sigma_{12})}{\partial p_2} &= \sum_{i=1}^{n}\Big(\frac{y_{01i}}{q_1p_2-\sigma_{12}} - \frac{y_{00i}}{q_1q_2+\sigma_{12}}\Big) = 0,\\
\frac{\partial \ell(y_i; p, \sigma_{12})}{\partial p_{11}} &= \sum_{i=1}^{n}\Big(\frac{y_{11i}}{p_1p_2+\sigma_{12}} - \frac{y_{10i}}{p_1q_2-\sigma_{12}} - \frac{y_{01i}}{q_1p_2-\sigma_{12}} + \frac{y_{00i}}{q_1q_2+\sigma_{12}}\Big) = 0.
\end{aligned} \tag{5.6}
$$

Solving the score equations (5.6), the estimates $\hat{p}_1, \hat{p}_2, \hat{q}_1, \hat{q}_1$ and $\hat{\sigma}_{12}$ can be obtained and then the estimates $\hat{p}_{11}, \hat{p}_{10}, \hat{p}_{01}$ and $\hat{p}_{00}$ can be determined using the relationship (5.2). These estimates provide very good measures of the correlation and the odds ratio in the independence case specially with large samples. Alternatively, the Marshall and Olkin procedure can provide similar estimates as well.

5.2. Test of hypothesis

To test the independence or specified values of the association between two variables Y_1 and Y_2 by the null hypothesis $H_0 : \sigma_{12} = \sigma_0$ against the alternative hypothesis $H_1 : \sigma_{12} \neq \sigma_0$, we can use the LRT as in equation (2.20). An estimate $\hat{p}_{11}$ under H_0 should be changed according to σ_{12} in the equations (5.2).
The deviance function as in the equation (2.21) can be used to determine the adequacy of this model, the difference is made by employing the relationships (5.2) to obtain

$$
\begin{aligned}
D = 2\sum_{i=1}^{n}\Big(y_{00i}\log\frac{y_{00i}}{\hat{q}_1\hat{q}_2+\hat{\sigma}_{12}} + y_{01i}\log\frac{y_{01i}}{\hat{q}_1\hat{p}_2-\hat{\sigma}_{12}} + y_{10i}\log\frac{y_{10i}}{\hat{p}_1\hat{q}_2-\hat{\sigma}_{12}}\\
+ y_{11i}\log\tfrac{y_{11i}}{\hat{p}_1\hat{p}_2+\hat{\sigma}_{12}}\Big) \sim \chi^2_{n-p}.
\end{aligned} \tag{5.7}
$$

Finally, we can use the equation (2.22) to obtain the estimate of the dispersion parameter ϕ.

6. The Bonney's Measure

The likelihood of a set of binary dependent outcomes, in Bonney's measure [4], with or without explanatory variables, is expressed as a product of conditional probabilities each of which is assumed to be logistic function. The logistic regression model provides a simple but relatively unknown parametrization of the multivariate distribution. This model is largely expository and is intended to motivate the development and usage of the regressive logistic model. Let us define the following conditional log odds

$$
\theta_1 = \eta = \log\frac{p_1}{1-p_1}, \qquad \theta_2 = \eta + \gamma_1 Z_1, \qquad p_1 = \frac{e^{\theta_1}}{1+e^{\theta_1}} = \frac{e^{\eta}}{1+e^{\eta}}. \tag{6.1}
$$

where η and γ_1 are well-known measures of dependence. These parameters η and γ_1 can take any values from $-\infty$ to ∞ and $Z_1 = 2Y_1 - 1$ coded $Z_1 = -1$ if $Y_1 = 0$ and $Z_1 = 1$ if

$Y_1 = 1$. So, if $\gamma_1 = 0$, then Y_1 and Y_2 are independent. The joint probability for the two binary dependent variables Y_1 and Y_2 is

$$f(y_1, y_2) = \prod_{j=1}^{2} \frac{e^{\theta_j y_j}}{1+e^{\theta_j}}. \tag{6.2}$$

Thus, the joint mass function of Y_1 and Y_2 can be expressed as products of ordinary logistic functions. To see the relationship of the γ_1 in this model to a well-known measures of dependence (the odds ratio ψ), consider a pair of dependent binary observations (y_1, y_2) without explanatory variables. From (6.1) and (6.2) we have

$$\begin{aligned} cp_{11} &= \frac{e^{\eta}}{1+\eta} \times \frac{e^{\eta+\gamma_1}}{1+e^{\eta+\gamma_1}}, \\ p_{10} &= \frac{e^{\eta}}{1+e^{\eta}} \times \frac{1}{1+e^{\eta+\gamma_1}}, \\ p_{01} &= \frac{1}{1+e^{\eta}} \times \frac{e^{\eta-\gamma_1}}{1+e^{\eta-\gamma_1}}, \\ p_{00} &= \frac{1}{1+e^{\eta}} \times \frac{1}{1+e^{\eta-\gamma_1}}. \end{aligned} \tag{6.3}$$

Using (6.3), and substituting in (2.8), then we have

$$\psi = e^{2\gamma_1}, \quad \gamma_1 = \frac{1}{2}\log\frac{p_{00}p_{11}}{p_{10}p_{01}} = \frac{1}{2}\log\psi = \frac{1}{2}\psi_{12}, \quad \eta = \frac{1}{2}\log\frac{p_{11}p_{01}}{p_{00}p_{10}}, \tag{6.4}$$

and, hence, γ_1 is one-half the natural logarithm of the odds ratio ψ. Note that if $\gamma_1 = 0$, then Y_1 and Y_2 are independent. Note that for Cessi and Houwelingen measure [5], the approximate relationship between $\tilde{\psi}_{12}$ and $\tilde{\rho}$ is given by $\tilde{\psi}_{12} = (1.7)^2\tilde{\rho}$, then we can derive the relation between $\tilde{\psi}$ and $\tilde{\rho}$ is $\tilde{\psi} = e^{(1.7)^2\tilde{\rho}}$. Also, for the measure based on the regressive model, [4], the relationship between γ_1 and ψ_{12} is $\gamma_1 = \frac{1}{2}\psi_{12}$, then we get $\psi = e^{2\gamma_1}$ and also $\psi_{12} = 2\gamma_1$. Finally, the relationship between γ_1 and $\tilde{\rho}$ is $\gamma_1 = 1.445\tilde{\rho}$. According to the conditional log odds interpretation for canonical parameters we have $\theta_2 = \theta_1 + \psi_{12}y_1$, but for the measure based on regressive model, we have $\theta_2 = \theta_1 + \gamma_1(2y_1 - 1)$. So, for $\gamma_1 = \dfrac{1}{2}\psi_{12}$, then $\theta_2 = \theta_1 + \psi_{12}(y_1 - \dfrac{1}{2}) = \theta_1 + \psi_{12}y_1 - \gamma_1$.

6.1. Estimation

For n observations, using the joint probability function (6.2), the log-likelihood function is

$$\ell(y_i; \eta, \gamma_1) = \sum_{i=1}^{n}\sum_{j=1}^{2}\Big(y_{ji}\theta_{ji} - \log(1+e^{\theta_{ji}})\Big). \tag{6.5}$$

Substituting by $Z_1 = 2Y_1 - 1$ and the values θ_1 and θ_2 from (6.1) in the log-likelihood function (6.5), and then taking the first derivative with respect to η and γ_1, and put it equal to zero, we have

$$\begin{aligned} \frac{\partial\ell(y_i;\eta,\gamma_1)}{\partial\eta} &= \sum\Big(y_{1i} + y_{2i} - \frac{e^{\eta}}{1+e^{\eta}} - \frac{e^{\eta+\gamma_1(2y_{1i}-1)}}{1+e^{\eta+\gamma_1(2y_{1i}-1)}}\Big) = 0, \\ \frac{\partial\ell(y_i;\eta,\gamma_1)}{\partial\gamma_1} &= \sum\Big((2y_{1i}-1)\Big(y_{2i} - \frac{e^{\eta+\gamma_1(2y_{1i}-1)}}{1+e^{\eta+\gamma_1(2y_{1i}-1)}}\Big)\Big) = 0. \end{aligned} \tag{6.6}$$

The estimates $\hat{\eta}$ and $\hat{\gamma}_1$ can be derived by solving the equations (6.6), and then using the equations (6.3) to obtain the estimates $\hat{p}_{11}, \hat{p}_{10}, \hat{p}_{01}$ and $\hat{p}_{00}$. The estimate $\hat{\psi}$ can be obtained by the relationship (6.4).

6.2. Test of hypothesis

Under $H_0 : \gamma_1 = \gamma_0$ the estimate $\hat{p}_1$ and then the estimate $\hat{\eta}$ can be obtained using the first equation of (6.6) as

$$\hat{p}_1 = \frac{1}{n}\sum_{i=1}^{n}\Big(\frac{y_{1i}+y_{2i}}{2} - \gamma_0(2y_{1i}-1)\Big), \tag{6.7}$$

where, $\hat{p}_1 = \dfrac{e^{\hat{\eta}}}{1+e^{\hat{\eta}}}$ and $\hat{\eta} = \log\dfrac{\hat{p}_1}{1-\hat{p}_1}$. To test the independence or specified values of the association parameter of the variables Y_1 and Y_2, by the null hypothesis $H_0 : \gamma_1 = \gamma_0$ against the alternative hypothesis $H_1 : \gamma_1 \neq \gamma_0$, using the log-likelihood function (6.5), we can use the LRT as

$$LRT = -2\Big(\ell(y_i;\hat{\eta},\gamma_0) - \ell(y_i;\hat{\eta},\hat{\gamma}_1)\Big) \quad \sim \quad \chi^2_1. \tag{6.8}$$

The deviance function as a way of assessing the goodness of fit for the model can be expressed as

$$D = 2\sum_{i=1}^{n}\Big(y_{1i}+y_{2i} - \hat{\theta}_1 y_{1i} - \hat{\theta}_2 y_{2i} - \log\frac{(1+e^{y_{1i}})(1+e^{y_{2i}})}{(1+e^{\hat{\theta}_{1i}})(1+e^{\hat{\theta}_{2i}})}\Big) \quad \sim \quad \chi^2_{n-p}. \tag{6.9}$$

The estimate of dispersion parameter ϕ can be obtained as in (2.22).

7. The generalized linear model (GLM)

Let us define the two binary variables Y_1 and Y_2, and put the joint probability function of Y_1 and Y_2 in the form of the marginal and conditional probabilities such that

$$f(y_1,y_2) = Pr(Y_2 = y_2 \mid Y_1 = y_1) \times Pr(Y_1 = y_1). \tag{7.1}$$

Supposing that
(7.2)

$$\theta_1 = \log\frac{p_1}{1-p_1}, \quad \theta_2 = \log\frac{p_2}{1-p_2}, \quad \theta_3 = \log\psi, \quad p_1 = \frac{e^{\theta_1}}{1+e^{\theta_1}}, \quad p_2 = \frac{e^{\theta_2}}{1+e^{\theta_2}}, \quad \psi = e^{\theta_3},$$

the marginal probability mass function of Y_1 can be expressed as

$$Pr(Y_1 = y_1) = \Big(\frac{e^{\theta_1}}{1+e^{\theta_1}}\Big)^{y_1}\Big(\frac{1}{1+e^{\theta_1}}\Big)^{1-y_1} = \frac{e^{\theta_1 y_1}}{1+e^{\theta_1}}. \tag{7.3}$$

According to the conditional log odds interpretation (Heagerty and Zeger[12] and Heagerty [13]), the conditional probability of $(Y_2 = y_2)$ given that $(Y_1 = y_1)$ is

$$Pr(Y_2 = y_2 \mid Y_1 = y_1) = \Big(\frac{e^{\theta_2+\theta_3 y_1}}{1+e^{\theta_2+\theta_3 y_1}}\Big)^{y_2}\Big(\frac{1}{1+e^{\theta_2+\theta_3 y_1}}\Big)^{1-y_2} = \frac{e^{\theta_2 y_2+\theta_3 y_1 y_2}}{1+e^{\theta_2+\theta_3 y_1}}, \tag{7.4}$$

where $E(Y_2 = y_2 \mid Y_1 = y_1) = \dfrac{e^{\theta_2+\theta_3 y_1}}{1+e^{\theta_2+\theta_3 y_1}}$.

Then, using the equations (7.1),(7.3) and (7.4), the joint probability mass function of the two binary variables Y_1 and Y_2 is

$$f(y_1,y_2)=\Big(\frac{p_1}{(1-p_1)(1-p_2+p_2\psi)}\Big)^{y_1}\Big(\frac{p_2}{1-p_2}\Big)^{y_2}\psi^{y_1y_2}(1-p_1)(1-p_2). \tag{7.5}$$

If $\psi=1$, then we have complete independence.

7.1. Estimation

Using the notations of the expression (7.2), the expression (7.5) can be written in the exponential family form

$$\begin{aligned} f(y_1,y_2)=\exp\Big\{&\theta_1y_1+\theta_2y_2+\theta_3y_1y_2-\log[1+e^{\theta_1}]-\log[1+e^{\theta_2}]\\ &-y_1(\log[1+e^{\theta_2+\theta_3}]-\log[1+e^{\theta_2}])\Big\}. \end{aligned} \tag{7.6}$$

For n observations, the log-likelihood function can be written as

$$\begin{aligned} \ell(y_i;\theta_1,\theta_2,\theta_3)=\sum_{i=1}^{n}\Big\{&\theta_1y_{1i}+\theta_2y_{2i}+\theta_3y_{1i}y_{2i}-\log[1+e^{\theta_1}]-\log[1+e^{\theta_2}]\\ &-y_{1i}(\log[1+e^{\theta_2+\theta_3}]-\log[1+e^{\theta_2}])\Big\}. \end{aligned} \tag{7.7}$$

The MLEs of the parameters are obtained by setting the first derivative for (7.7) with respect to the parameters θ_1,θ_2 and θ_3, to zero and we have

$$\frac{\partial\ell(y_i;\theta_1,\theta_2,\theta_3)}{\partial\theta_1}=\sum_{i=1}^{n}\Big(y_{1i}-\frac{e^{\theta_1}}{1+e^{\theta_1}}\Big)=0, \tag{7.8}$$

$$\frac{\partial\ell(y_i;\theta_1,\theta_2,\theta_3)}{\partial\theta_2}=\sum_{i=1}^{n}\Big(y_{2i}-\frac{e^{\theta_2}}{1+e^{\theta_2}}\Big)-\sum_{i=1}^{n}y_{1i}\Big(\frac{e^{\theta_2+\theta_3}}{1+e^{\theta_2+\theta_3}}-\frac{e^{\theta_2}}{1+e^{\theta_2}}\Big)=0, \tag{7.9}$$

$$\frac{\partial\ell(y_i;\theta_1,\theta_2,\theta_3)}{\partial\theta_3}=\sum_{i=1}^{n}\Big(y_{1i}y_{2i}-y_{1i}\frac{e^{\theta_2+\theta_3}}{1+e^{\theta_2+\theta_3}}\Big)=0. \tag{7.10}$$

Solving the equations (7.8), we get the estimate

$$\hat{p}_1=\frac{1}{n}\sum_{i=1}^{n}y_{1i}. \tag{7.11}$$

Substituting by the estimate $\hat{p}_1$ in the equation (7.9), we have

$$\frac{\partial\ell(y_i;\theta_1,\theta_2,\theta_3)}{\partial\theta_2}=\sum_{i=1}^{n}y_{2i}-np_2-\sum_{i=1}^{n}y_{1i}y_{2i}+p_2\sum_{i=1}^{n}y_{1i}=0, \tag{7.12}$$

then we obtain the estimate

$$\hat{p}_2=\frac{\sum_{i=1}^{n}y_{2i}-\sum_{i=1}^{n}y_{1i}y_{2i}}{n-\sum_{i=1}^{n}y_{1i}}. \tag{7.13}$$

Also, the estimates $\hat{\theta}_1,\hat{\theta}_2$ and $\hat{\theta}_3$ can be derived by solving the equations (7.8), (7.9) and (7.10) directly. Alternatively, using (7.10) and the estimate $\hat{\theta}_2=\log\frac{\hat{p}_2}{1-\hat{p}_2}$, we get the estimate $\hat{\theta}_3$ as

$$\hat{\theta}_3=\log\frac{(1-\hat{p}_2)\sum_{i=1}^{n}y_{1i}y_{2i}}{\hat{p}_2(\sum_{i=1}^{n}y_{1i}-\sum_{i=1}^{n}y_{1i}y_{2i})}. \tag{7.14}$$

Then, using (7.14) we can obtain the estimate $\hat{\psi} = e^{\hat{\theta}_3}$. The estimate $\hat{p}_{11}$ can be obtained using the equation (3.7), and the estimates of joint probabilities can be obtained as $\hat{p}_{10} = \hat{p}_1 - \hat{p}_{11}, \hat{p}_{01} = \hat{p}_2 - \hat{p}_{11}$ and $\hat{p}_{00} = 1 - \hat{p}_{10} - \hat{p}_{01} - \hat{p}_{11}$.

7.2. Test of hypothesis

Under the null hypothesis $H_0 : \theta_3 = \log\psi_0$, the estimate $\hat{p}_1$ does not change as in the equation (7.11) but the estimate $\hat{p}_2$ can be determined by solving the equation (7.9). Substituting by the value ψ_0 into the equation (7.9), and solving it for $\hat{p}_2$ we have

$$\hat{p}_2 = -\frac{1}{2a}(b - \sqrt{b^2 + 4ac}), \tag{7.15}$$

where,

$$a = \sum_{i=1}^{n} y_{1i} + \psi_0(n - \sum_{i=1}^{n} y_{1i}) - n, \quad b = n + (\psi_0 - 1)(\sum_{i=1}^{n} y_{1i} - \sum_{i=1}^{n} y_{2i}), \quad c = \sum_{i=1}^{n} y_{2i},$$

then we obtain the estimate $\tilde{\theta}_2 = \log\dfrac{\tilde{p}_2}{1 - \tilde{p}_2}$. On the other hand, in the case of independence, $\log\psi_0 = 0$, also the estimate $\hat{p}_1$ does not change as in the equation (7.11), and using the equation (7.9) to obtain the estimate $\hat{p}_2 = \frac{1}{n}\sum y_{2i}$.
To test for the independence or specified values of the odds ratio of the two variables Y_1 and Y_2 by the null hypothesis $H_0 : \theta_3 = \log\psi_0$ against the alternative hypothesis $H_1 : \theta_3 \neq \log\psi_0$ we can use the LRT as

$$LRT = -2\Big(\ell(y_i; \hat{\theta}_1, \tilde{\theta}_2, \log\psi_0) - \ell(y_i; \hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3)\Big) \quad \sim \quad \chi_1^2. \tag{7.16}$$

The deviance function may be employed to assess the goodness of fit for the model and we can express it as

$$D = 2\Big(\ell(y_i; y_i) - \ell(y_i; \hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3)\Big) \quad \sim \quad \chi_{n-p}^2 \tag{7.17}$$

where

$$\ell(y_i; \hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3) = \sum_{i=1}^{n} \Big\{ \hat{\theta}_1 y_{1i} + \hat{\theta}_2 y_{2i} + \hat{\theta}_3 y_{1i} y_{2i} - \log[1 + e^{\hat{\theta}_1}] - \log[1 + e^{\hat{\theta}_2}] \\ - y_{1i}(\log[1 + e^{\hat{\theta}_2 + \hat{\theta}_3}] - \log[1 + e^{\hat{\theta}_2}]) \Big\}$$

is the log-likelihood function for the model of interest evaluated at maximum likelihood estimates $\hat{\theta}_j (j = 1, 2, 3)$, and

$$\ell(y_i; y_i) = \sum_{i=1}^{n} \Big\{ y_{1i} + y_{2i} + y_{1i} y_{2i} - \log[1 + e^{y_{1i}}] - \log[1 + e^{y_{2i}}] \\ - y_{1i}(\log[1 + e^{y_{2i} + y_{1i} y_{2i}}] - \log[1 + e^{y_{2i}}]) \Big\}$$

denotes the log-likelihood function for the saturated model evaluated at observed value y_i.
The estimate of the dispersion parameter ϕ as in the equation (2.22).

8. The quadratic exponential form (QEF)

A model of quadratic exponential form is parameterized in terms of marginal means and pairwise correlations for the regression analysis of correlated binary data. Zhao and Prentice [27] used the pseudo-maximum likelihood method using a special case termed as the multiplicative model. The score estimating functions for the mean and correlation parameters are expressed in simple form under the quadratic exponential family. The quadratic exponential form of Y_1 and Y_2 can be written as

$$f(y_1,y_2)=\frac{1}{\Delta}\exp\Big(\theta_1 y_1+\theta_2 y_2+\theta_3 y_1 y_2+c(y)\Big),\quad -\infty\le\theta_1,\theta_2,\theta_3\le\infty, \tag{8.1}$$

where, $\Delta=\Sigma\exp\Big(\theta_1 y_1+\theta_2 y_2+\theta_3 y_1 y_2+c(y)\Big)$ is the normalizing constant. The summation is over all four possible values of Y_1 and Y_2. The parameters θ_1,θ_2 and θ_3 are canonical parameters which can be specified as

$$\begin{aligned}\theta_1&=\log\frac{p_1}{1-p_1}, & p_1&=\frac{e^{\theta_1}}{1+e^{\theta_1}},\\ \theta_2&=\log\frac{p_2}{1-p_2}, & p_2&=\frac{e^{\theta_2}}{1+e^{\theta_2}},\\ \theta_3&=\log\psi, & \psi&=e^{\theta_3}.\end{aligned} \tag{8.2}$$

The joint function of Y_1 and Y_2, with $c(y)=0$, can be expressed as

$$f(y_1,y_2)=\frac{1}{\Delta}\exp\Big(\theta_1 y_1+\theta_2 y_2+\theta_3 y_1 y_2\Big). \tag{8.3}$$

8.1. Estimation

The log-likelihood function, for n observation, is

$$\ell(y_i;\theta_1,\theta_2,\theta_3)=\sum_{i=1}^{n}\Big(\theta_1 y_{1i}+\theta_2 y_{2i}+\theta_3 y_{1i}y_{2i}-\log\Delta\Big), \tag{8.4}$$

where, $\Delta=\sum\exp\Big(\theta_1 y_1+\theta_2 y_2+\theta_3 y_1 y_2\Big)$. The marginal and joint parameters can be written, using (8.3), as

$$\begin{aligned}p_1&=\sum_{y_1,y_2}\frac{y_1}{\Delta}\exp\Big(\theta_1 y_1+\theta_2 y_2+\theta_3 y_1 y_2\Big),\\ p_2&=\sum_{y_1,y_2}\frac{y_2}{\Delta}\exp\Big(\theta_1 y_1+\theta_2 y_2+\theta_3 y_1 y_2\Big),\\ p_{11}&=\sum_{y_1,y_2}\frac{y_1 y_2}{\Delta}\exp\Big(\theta_1 y_1+\theta_2 y_2+\theta_3 y_1 y_2\Big),\quad p_{11}=\sigma_{12}+p_1p_2,\sigma_{12}=\mathrm{Cov}(Y_1,Y_2).\end{aligned} \tag{8.5}$$

Using the equation (8.5) and taking the first derivatives for the log-likelihood function (8.4) with respect to θ_1,θ_2 and θ_3, and put it equal to zero, we have

$$\sum_{i=1}^{n}y_{1i}-np_1=0,\qquad \sum_{i=1}^{n}y_{2i}-np_2=0,\qquad \sum_{i=1}^{n}y_{1i}y_{2i}-np_{11}=0. \tag{8.6}$$

From the equations (8.6), the maximum likelihood estimates $\hat{p}_1, \hat{p}_2$ and $\hat{p}_{11}$ can be obtained as

$$\hat{p}_1 = \frac{1}{n}\sum y_{1i}, \quad \hat{p}_2 = \frac{1}{n}\sum y_{2i}, \quad \hat{p}_{11} = \frac{1}{n}\sum y_{1i}y_{2i}. \tag{8.7}$$

The estimates of the odds are $\hat{\theta}_1 = \log\frac{\hat{p}_1}{1-\hat{p}_1}$ and $\hat{\theta}_2 = \log\frac{\hat{p}_2}{1-\hat{p}_2}$ and the estimate of the covariance $\hat{\sigma}_{12} = \hat{p}_{11} - \hat{p}_1\hat{p}_2$. The other estimates of the joint probabilities can be obtained as $\hat{p}_{10} = \hat{p}_1 - \hat{p}_{11}, \hat{p}_{01} = \hat{p}_2 - \hat{p}_{11}$ and $\hat{p}_{00} = 1 - \hat{p}_{10} - \hat{p}_{01} - \hat{p}_{11}$.

8.2. Test of hypothesis

Similar to the GLM procedure, we can test the independence or specified values of the odds ratio of two variables, Y_1 and Y_2, by the null hypothesis $H_0 : \theta_3 = \log\psi_0$ against the alternative hypothesis $H_1 : \theta_3 \neq \log\psi_0$ using the LRT as

$$LRT = -2\Big(\ell(y_i, \hat{\theta}_1, \hat{\theta}_2, \log\psi_0) - \ell(y_i, \hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3)\Big) \quad \sim \quad \chi_1^2. \tag{8.8}$$

The estimates $\hat{p}_1, \hat{p}_2$ and $\hat{p}_{11}$ under H_0 cannot be changed according to (8.6) and (8.7). This is considered as a disadvantage for this measure and will have effect on the results of the LRT, especially with higher order associations (See Table 10). The deviance can be used to assess the goodness of fit of the proposed model which can be expressed as

$$D = 2\Big(\ell(y_i; y_i) - \ell(y_i; \hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3)\Big) \quad \sim \quad \chi^2_{n-p} \tag{8.9}$$

where,

$$\ell(y_i; \hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3) = \sum_{i=1}^{n}\Big(\hat{\theta}_1 y_{1i} + \hat{\theta}_2 y_{2i} + \hat{\theta}_3 y_{1i}y_{2i} - \log\Delta\Big),$$

is the log-likelihood function for the model of interest evaluated at maximum likelihood estimates $\hat{\theta}_j (j = 1,2,3)$, $\Delta = \sum\exp\Big(\hat{\theta}_1 y_1 + \hat{\theta}_2 y_2 + \hat{\theta}_3 y_1 y_2\Big)$, the summation is over all four possible values of Y_1 and Y_2, and

$$\ell(y_i; y_i) = \sum_{i=1}^{n}\Big(y_{1i} + y_{2i} + y_{1i}y_{2i} - \log\Delta_y\Big),$$

denotes the log-likelihood function for the saturated model evaluated at observed values y_i, $\Delta_y = \sum\exp\Big(y_1 + y_2 + y_1 y_2\Big)$, the summation is over all four possible values of Y_1 and Y_2.

The estimate of the dispersion parameter ϕ is same as in the equation (2.22).

9. The alternative quadratic exponential form (AQEF)

In this section, we propose a new form for the bivariate quadratic exponential form, called the alternative quadratic exponential form (AQEF). This form overcomes some problems with the existing form of the quadratic exponential form (QEF). Also, this new form can use the ML procedure for estimation with a more natural way to express the normalization. On the other hand, the QEF needs some assumptions for estimation. For this reason, the pseudo-likelihood procedure is used in QEF for the estimation.

To make the quadratic exponential form (QEF) more realistic and effective measure, let us make the following modifications in the existing form (8.3). The joint function of Y_1 and Y_2 can be put in the form

$$f(y_1,y_2)=\exp\Big(\theta_1 y_1+\theta_2 y_2+\theta_3 y_1 y_2-A(\theta_1,\theta_2,\theta_3)\Big),\qquad -\infty\le\theta_1,\theta_2,\theta_3\le\infty. \tag{9.1}$$

Since $\sum_{y_1=0}^{1}\sum_{y_2=0}^{1}f(y_1,y_2)=1$, then we can obtain

$$A(\theta_1,\theta_2,\theta_3)=\log(1+e^{\theta_1}+e^{\theta_2}+e^{\theta_1+\theta_2+\theta_3}).$$

So, the joint function (9.1) becomes

$$f(y_1,y_2)=\exp\Big(\theta_1 y_1+\theta_2 y_2+\theta_3 y_1 y_2-\log(1+e^{\theta_1}+e^{\theta_2}+e^{\theta_1+\theta_2+\theta_3})\Big). \tag{9.2}$$

When Y_1 and Y_2 are independent, $\theta_3=0$, then the equation (9.2) is

$$f(y_1,y_2)=\exp\Big(\theta_1 y_1+\theta_2 y_2-\log(1+e^{\theta_1})-\log(1+e^{\theta_2})\Big). \tag{9.3}$$

where,

$$\log(1+e^{\theta_1}+e^{\theta_2}+e^{\theta_1+\theta_2})=\log(1+e^{\theta_1})+\log(1+e^{\theta_2}).$$

So, in this case, we consider the alternative quadratic exponential form (AQEF) as an special case of GLM, when

$$c(y)=y_1\Big[\log(1+e^{\theta_2})-\log(1+e^{\theta_2+\theta_3})\Big]=0.$$

9.1. Estimation

For n observations, the log-likelihood function can be written as

$$\ell(y_i;\theta_1,\theta_2,\theta_3)=\sum_{i=1}^{n}\Big(\theta_1 y_{1i}+\theta_2 y_{2i}+\theta_3 y_{1i}y_{2i}-\log[1+e^{\theta_1}+e^{\theta_2}+e^{\theta_1+\theta_2+\theta_3}]\Big). \tag{9.4}$$

Taking the first derivative for (9.4) with respect to θ_1,θ_2 and θ_3, and putting it equal to zero, respectively, we obtain

$$\begin{aligned}
c\frac{\partial\ell(y_i;\theta_1,\theta_2,\theta_3)}{\partial\theta_1}&=\sum_{i=1}^{n}\Big(y_{1i}-\frac{e^{\theta_1}+e^{\theta_1+\theta_2+\theta_3}}{1+e^{\theta_1}+e^{\theta_2}+e^{\theta_1+\theta_2+\theta_3}}\Big)=0,\\
\frac{\partial\ell(y_i;\theta_1,\theta_2,\theta_3)}{\partial\theta_2}&=\sum_{i=1}^{n}\Big(y_{2i}-\frac{e^{\theta_2}+e^{\theta_1+\theta_2+\theta_3}}{1+e^{\theta_1}+e^{\theta_2}+e^{\theta_1+\theta_2+\theta_3}}\Big)=0,\\
\frac{\partial\ell(y_i;\theta_1,\theta_2,\theta_3)}{\partial\theta_3}&=\sum_{i=1}^{n}\Big(y_{1i}y_{2i}-\frac{e^{\theta_1+\theta_2+\theta_3}}{1+e^{\theta_1}+e^{\theta_2}+e^{\theta_1+\theta_2+\theta_3}}\Big)=0.
\end{aligned} \tag{9.5}$$

Solving equations (9.5), we get the estimates $\hat{\theta}_1,\hat{\theta}_2$ and $\hat{\theta}_3$, then we can obtain the estimates $\hat{\psi}=e^{\hat{\theta}_3},\hat{p}_1=\dfrac{e^{\hat{\theta}_1}}{1+e^{\hat{\theta}_1}},\hat{p}_2=\dfrac{e^{\hat{\theta}_2}}{1+e^{\hat{\theta}_2}}$, $\hat{p}_{11}$ can be obtained using the equation (3.7) and the other estimates of joint probabilities $\hat{p}_{10},\hat{p}_{01}$ and $\hat{p}_{00}$ can be obtained as the QEF procedure.

9.2. Test of hypothesis

Under $H_0 : \theta_3 = \log \psi_0$, we can solve the first two equations from the estimating equations (9.5) for θ_1 and θ_2 to obtain the estimates $\tilde{\theta}_1$ and $\tilde{\theta}_2$ under H_0. These estimates improve the existing QEF procedure and we obtain deviance for first or higher associations more effectively (See Table 11). Similar to the QEF procedure, we can test the independence or specified values of the odds ratio of the two variables Y_1 and Y_2 by the null hypothesis $H_0 : \theta_3 = \log \psi_0$ against the alternative hypothesis $H_1 : \theta_3 \neq \log \psi_0$ and we can use the LRT as

$$LRT = -2\Big(\ell(y_i, \tilde{\theta}_1, \tilde{\theta}_2, \log \psi_0) - \ell(y_i, \hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3)\Big) \quad \sim \quad \chi^2_1. \tag{9.6}$$

The deviance function can be employed for assessing the goodness of fit of the proposed model which can be expressed as

$$D = 2\Big(\ell(y_i; y_i) - \ell(y_i; \hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3)\Big) \quad \sim \quad \chi^2_{n-p} \tag{9.7}$$

where

$$\ell(y_i; \hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3) = \sum_{i=1}^{n} \Big(\hat{\theta}_1 y_{1i} + \hat{\theta}_2 y_{2i} + \hat{\theta}_3 y_{1i} y_{2i} - \log[1 + e^{\hat{\theta}_1} + e^{\hat{\theta}_2} + e^{\hat{\theta}_1 + \hat{\theta}_2 + \hat{\theta}_3}]\Big),$$

is the log-likelihood function for the model of interest evaluated at maximum likelihood estimates $\hat{\theta}_j (j = 1, 2, 3)$, and

$$\ell(y_i; y_i) = \sum_{i=1}^{n} \Big(y_{1i} + y_{2i} + y_{1i} y_{2i} - \log[1 + e^{y_{1i}} + e^{y_{2i}} + e^{y_{1i} + y_{2i} + y_{1i} y_{2i}}]\Big),$$

denotes the log-likelihood function for the saturated model evaluated at observed value y_i,

The estimate of the dispersion parameter ϕ is same as in the equation (2.22).

10. Numerical examples

For numerical examples we use the Leisch *et al.* technique [15], with bindata package of R program, to generate the multivariate binary data with given pairwise joint probabilities for the correlated binary variables , Y_1 and Y_2. Depending on the parameter values in Table 3, these conditions must be satisfied: $p_{11} \leq \min(p_1, p_2)$ and $p_1 + p_2 - p_{11} \leq 1$, also $p_{11} \geq \max(p_1 + p_2 - 1, 0)$. We will use different values of $n = 50, 500$ with 500 replicates and different values of the odds ratio $\psi = 0.5, 1, 2$. The estimates for the measure based on regressive model and the alternative quadratic exponential form (AQEF) are solved using R program with BBsolve or multiStart functions in BB package [25], for solving a large system of non-linear equations with initial values 0's for all estimates. The test statistics for deviance are compared with χ^2_{n-p} where p is the number of estimated parameters to determine the most powerful model. Also the LRT is compared with χ^2_1 to test for the independence or specified value of the measure of association of the two variables Y_1 and Y_2. For all tests, we used $\alpha = 0.05$.

The estimates of the joint probabilities $\hat{p}_{rs} (r, s = 0, 1)$, the correlation $\hat{\rho}$ and the odds ratio $\hat{\psi}$ for all measures are presented in Tables 4-11. The dispersion parameter $\hat{\psi}$ is determined to choose the best measure which has the lowest value of this estimate. The likelihood ratio test (LRT) indicates the test for independence or specified values of the measure of association via determining the number times out of 500 we accept H_0. The deviance test shows the

Table 3. Parameters values used to generate the correlated binary data

Odds Ratio	p_1	p_2	p_{11}
$\psi = 0.5$	0.6	0.4	0.2
$\psi = 1$	0.6	0.5	0.3
$\psi = 2$	0.6	0.6	0.4

adequacy of all models via how many times we accept the null hypotheses of good fit out of 500 for each measure. For the Cessi and Houwelingen measure (Table 7), we observed the values of the score statistics, W and M, which should be around 1 if there is independence and more than 1 if there is deviation from independence. Also the approximate relationship between the odds ratio and tetrachoric correlation, $\tilde{\psi}_{12} = (1.7)^2\tilde{\rho}$, which is valid only in the independence case. Hence, this is a trivial relationship which does not appear to hold in case of dependence in the bivariate Bernoulli variables. For the measure based on regressive model (Table 8), we can interpret, in addition to the other estimates, the estimates of dependence measure ($\hat{\gamma}_1$). Finally, Table 11 demonstrates that we have some better results for the alternative quadratic exponential form (AQEF) especially in the case of higher association comparing with the results of an conventional quadratic exponential form (QEF).

Table 4. The results for Marshall and Olkin (M. O.) measure

Odds Ratio	$\psi = 0.5$	$\psi = 0.5$	$\psi = 1$	$\psi = 1$	$\psi = 2$	$\psi = 2$
No. of Observations	$n = 50$	$n = 500$	$n = 50$	$n = 500$	$n = 50$	$n = 500$
$\hat{p}_{11}$	0.1988	0.1994	0.3020	0.2989	0.4043	0.4004
$\hat{p}_{10}$	0.4046	0.4012	0.3012	0.3016	0.1979	0.2001
$\hat{p}_{01}$	0.1982	0.2009	0.1969	0.2010	0.1962	0.1997
$\hat{p}_{00}$	0.1984	0.1985	0.1999	0.1985	0.2016	0.1998
$\hat{\rho}$	-0.1705	-0.1711	0.0045	-0.0052	0.1781	0.1668
$\hat{\psi}$	0.5715	0.4988	1.2459	0.9954	2.6433	2.0432
LRT	476	467	465	473	464	475
D. Test	500	500	500	500	500	500
$\hat{\phi}$	1.2356	1.1670	1.2361	1.1671	1.2369	1.1671

Table 5. The results for Dale measure in the independence case

Estimates	$\hat{p}_{11}$	$\hat{p}_{10}$	$\hat{p}_{01}$	$\hat{p}_{00}$	$\hat{\rho}$	$\hat{\psi}$	LRT	D. Test	$\hat{\phi}$
$n = 50$	0.2092	0.1991	0.2964	0.2953	0.0107	1.1368	496	476	15.2419
$n = 500$	0.2024	0.2006	0.2998	0.2972	0.00004	1.0077	495	500	1.4423

Table 6. The results for Teugels measure in the independence case

Estimates	$\hat{p}_{11}$	$\hat{p}_{10}$	$\hat{p}_{01}$	$\hat{p}_{00}$	$\hat{\rho}$	$\hat{\psi}$	LRT	D. Test	$\hat{\phi}$
$n = 50$	0.1991	0.2005	0.3043	0.2961	-0.0088	1.0449	496	476	14.7686
$n = 500$	0.1996	0.2006	0.3006	0.2992	-0.0025	0.9963	498	500	1.4191

Table 7. The results for Cessi and Houwelingen measure based on M.O. procedure

Odds Ratio	$\psi = 0.5$	$\psi = 0.5$	$\psi = 1$	$\psi = 1$	$\psi = 2$	$\psi = 2$
No. of Observations	$n = 50$	$n = 500$	$n = 50$	$n = 500$	$n = 50$	$n = 500$
$\tilde{\rho}$	-0.1737	-0.1711	0.0021	-0.0043	0.1204	0.1113
$\tilde{\psi}_{12}$	-0.7287	-0.7143	0.0192	-0.0214	0.7579	0.6964
$\tilde{\psi}_{12} = (1.7)^2\tilde{\rho}$	-0.5021	-0.4946	0.0062	-0.0124	0.3480	0.3216
W	2.4061	15.688	1.0742	1.0097	2.6072	14.9366
M	2.4061	15.688	1.0742	1.0097	2.6072	14.9366

Table 8. The results for Bonney measure

Odds Ratio	$\psi = 0.5$	$\psi = 0.5$	$\psi = 1$	$\psi = 1$	$\psi = 2$	$\psi = 2$
No. of Observations	$n = 50$	$n = 500$	$n = 50$	$n = 500$	$n = 50$	$n = 500$
$\hat{p}_{11}$	0.2095	0.2079	0.2943	0.2985	0.3998	0.3998
$\hat{p}_{10}$	0.2999	0.3026	0.2526	0.2526	0.1955	0.1933
$\hat{p}_{01}$	0.2965	0.2998	0.2504	0.2516	0.2069	0.2060
$\hat{p}_{00}$	0.1941	0.1897	0.2027	0.1971	0.1978	0.2009
$\hat{\rho}$	-0.1970	-0.2057	-0.0195	-0.0197	0.1570	0.1682
$\hat{\psi}$	0.4838	0.4364	0.9868	0.9294	2.0710	2.0293
$\hat{\gamma}_1$	-0.3630	-0.4146	-0.0066	-0.0366	0.3640	0.3538
LRT	493	482	493	420	491	390
D. Test	500	500	500	500	500	500
$\hat{\phi}$	1.3319	1.2529	1.3089	1.2261	1.2567	1.1784

Table 9. The results for GLM measure

Odds Ratio	$\psi = 0.5$	$\psi = 0.5$	$\psi = 1$	$\psi = 1$	$\psi = 2$	$\psi = 2$
No. of Observations	$n = 50$	$n = 500$	$n = 50$	$n = 500$	$n = 50$	$n = 500$
$\hat{p}_{11}$	0.2630	0.2604	0.2992	0.3009	0.3375	0.3411
$\hat{p}_{10}$	0.3406	0.3402	0.3034	0.2996	0.2647	0.2594
$\hat{p}_{01}$	0.2373	0.2428	0.1962	0.2023	0.1553	0.1589
$\hat{p}_{00}$	0.1591	0.1566	0.2012	0.1972	0.2425	0.2406
$\hat{\rho}$	-0.1647	-0.1714	0.0039	-0.0052	0.1711	0.1672
$\hat{\psi}$	0.5715	0.4988	1.2459	0.9954	2.6433	2.0432
LRT	476	467	465	473	464	475
D. Test	500	500	500	500	500	500
$\hat{\phi}$	1.3147	1.1759	1.3178	1.1733	1.3213	1.1735

Table 10. The results for QEF measure

Odds Ratio	$\psi = 0.5$	$\psi = 0.5$	$\psi = 1$	$\psi = 1$	$\psi = 2$	$\psi = 2$
No. of Observations	$n = 50$	$n = 500$	$n = 50$	$n = 500$	$n = 50$	$n = 500$
$\hat{p}_{11}$	0.1988	0.1994	0.3020	0.2990	0.4043	0.4004
$\hat{p}_{10}$	0.4046	0.4012	0.3012	0.3015	0.1979	0.2001
$\hat{p}_{01}$	0.1982	0.2009	0.1969	0.2010	0.1962	0.1997
$\hat{p}_{00}$	0.1984	0.1985	0.1999	0.1985	0.2016	0.1998
$\hat{\rho}$	-0.1705	-0.1711	0.0045	-0.0052	0.1781	0.1668
$\hat{\psi}$	0.5715	0.4988	1.2459	0.9954	2.6433	2.0432
LRT	439	329	500	500	387	279
D. Test	499	500	487	500	382	16
$\hat{\phi}$	1.2356	1.1670	1.2361	1.1671	1.2369	1.1671

Table 11. The results for AQEF measure

Odds Ratio	$\psi = 0.5$	$\psi = 0.5$	$\psi = 1$	$\psi = 1$	$\psi = 2$	$\psi = 2$
No. of Observations	$n = 50$	$n = 500$	$n = 50$	$n = 500$	$n = 50$	$n = 500$
$\hat{p}_{11}$	0.2706	0.2848	0.2898	0.2947	0.3126	0.3037
$\hat{p}_{10}$	0.3728	0.3717	0.2927	0.2978	0.2158	0.2123
$\hat{p}_{01}$	0.1875	0.2006	0.2003	0.1976	0.2175	0.2130
$\hat{p}_{00}$	0.1691	0.1429	0.2172	0.2099	0.2541	0.2710
$\hat{\rho}$	-0.1076	-0.1440	0.0162	0.0118	0.1282	0.1477
$\hat{\psi}$	0.6557	0.5443	1.0996	1.0531	1.7527	1.8303
LRT	469	457	454	463	492	467
D. Test	496	500	492	500	497	436
$\hat{\phi}$	1.2218	1.1284	1.2758	1.1755	1.3174	1.2342

11. Conclusions

It is evident from the results that all the measures of association considered in this study and their test procedures provide almost similar results, but the generalized linear model and the proposed alternative quadratic form approaches may be considered to have better performance. From Table 4 to Table 11 we observe the following features of the measures of bivariate Bernoulli variables based on their estimation and test results:

- For all measures, the estimates of joint probabilities, correlations and odds ratios are almost close to their parameter values especially with large sample size.
- For the LRT, the results are similar for most of the measures and tend to accept H_0.
- For the deviance test, the results are similar for all measures and indicate good fits.
- The deterestimates of dispersion parameter ϕ are similar for all measures, except for Dale and Teugels mesures at $n = 50$ in the independence case $\psi = 1$, and all estimates exceed one. This indicates the over dispersion case for the binary data. The AQEF measure in Table 11 displays the lowest value of $\hat{\phi}$ with $n = 50$ and $\psi = 0.5$.

- For the Cessi and Houwelingen measure, the score statistics W and M are close to 1 if there is independence and this deals with the theoretical assumption, but the approximate relationship between the log odds ratio and correlation could not be demonstrated for all cases.
- A comparison between the quadratic exponential form (QEF) and the alternative quadratic exponential form (AQEF) demonstrates that for higher values of the odds ratio, the QEF is not a good measure as displayed by both the LRT and deviance tests for large sample size but the AQEF overcomes this disadvantage as shown in Table 11.
- For the measure based on the regressive approach, the results in Table 8 shows that the LRT is good for $n = 50$, this is because the assumption for η that is fixed for two parameters θ_1 and θ_2.

In sum, we may conclude that the GLM and proposed AQEF models display better performance.

Acknowledgement. This project was supported by NPST-Saudi Arabia through the research project 11-MAT1856-02. The authors are grateful to the anonymous referees for their helpful comments.

References

[1] A. Agresti, *Categorical Data Analysis*, second edition, Wiley Series in Probability and Statistics, Wiley-Interscience, New York, 2002.

[2] J. R. Ashford and R. R. Sowden, Multivariate probit analysis, *Biometrics* **26** (1970), 535–546.

[3] R. R. Bahadur, A representation of the joint distribution of responses to n dichotomous items, in *Studies in Item Analysis and Prediction*, 158–168, Stanford Univ. Press, Stanford, CA.

[4] G. E. Bonney, Logistic regression for dependent binary observations, *Biometrics* **43** (1987), 951–973.

[5] S. L. Cessie and J. C. Houwelingen, Logistic regression for correlated binary data, *J. Royal Statist. Soc.* **43** (1994), no. 1, 95–108.

[6] D. R. Cox and N. Wermuth, A note on the quadratic exponential binary distribution, *Biometrika* **81** (1994), no. 2, 403–408.

[7] D. R. Cox, The analysis of multivariate binary data, *Appl. Statist.* **21** (1972), 113–120.

[8] J. R. Dale, Global cross-ratio models for bivariate, discrete ordered responses, *Biometrics* **42** (1986), 909–917.

[9] G. M. Fitzmaurice and N. M. Laird, A likelihood-based method for analyzing longitudinal binary responses, *Biometrika* **80** (1993), no. 1, 141–151.

[10] G. F. V. Glonek and P. McCullagh, Multivariate logistic models, *J. Royal Statist. Soc. Ser. B (Methodological)* **57** (1995), no. 3, 533–546.

[11] C. Gouriéroux, A. Monfort and A. Trognon, Pseudomaximum likelihood methods: Theory, *Econometrica* **52** (1984), no. 3, 681–700.

[12] P. J. Heagerty and S. L. Zeger, Marginalized multilevel models and likelihood inference, *Statist. Sci.* **15** (2000), no. 1, 1–26.

[13] P. J. Heagerty, Marginalized transition models and likelihood inference for longitudinal categorical data, *Biometrics* **58** (2002), no. 2, 342–351.

[14] M. A. Islam, R. I. Chowdhury and L. Briollais, A bivariate binary model for testing dependence in outcomes, *Bull. Malays. Math. Sci. Soc. (2)* **35** (2012), no. 4, 845–858.

[15] F. Leisch, A. Weingessel and J. Hornik, On the generating of correlated artificial binary data, *Working Paper Series* **13** (1998), 1–14.

[16] K. Y. Liang and S. L. Zeger, Longitudinal data analysis using generalized linear models, *Biometrika* **73** (1986), no. 1, 13–22.

[17] G. Lovison, A matrix-valued Bernoulli distribution, *J. Multivariate Anal.* **97** (2006), no. 7, 1573–1585.

[18] P. McCullagh and J. A. Nelder, *Generalized Linear Models*, Monographs on Statistics and Applied Probability, Chapman & Hall, London, 1983.

[19] A. W. Marshall and I. Olkin, A family of bivariate distributions generated by the bivariate Bernoulli distribution, *J. Amer. Statist. Assoc.* **80** (1985), no. 390, 332–338.

[20] L. Martin, *Coercive Cooperation: Explaining Multilateral Economic Sanctions*, Princeton: Princeton University Press, 1992.

[21] J. Palmgren, Regression models for bivariate binary responses, *Technical Report* **101**, Department of Biostatistics, School of Public Health and Community Medicine, Seattle, 1989.

[22] K. Pearson, Mathematical contribution to the theory of evolution: VII, On the correlation of characters not quantitively measurable, *Phil. Trans. R. Soc. Lond. A* **195** (1900), 1–47.

[23] A. Stuart and J. K. Ord, Kendall's Advanced Theory of Statistics, vol. 2, p.659. London: Arnold, 1991.

[24] J. L. Teugels, Some representations of the multivariate Bernoulli and binomial distributions, *J. Multivariate Anal.* **32** (1990), no. 2, 256–268.

[25] R. Varadhan and P. D. Gilbert, BB: An R package for solving a large system of nonlinear equations and for optimizing a high-dimensional nonlinear objective function, *J. Statist. Soft.* **32** (2009), no. 4, 1–26.

[26] S. L. Zeger and K. Liang, Longitudinal data analysis for discrete and continuous outcomes, *Biometrics* **42** (1986), 121–130.

[27] L. P. Zhao and R. L. Prentice, Correlated binary regression using a quadratic exponential model, *Biometrika* **77** (1990), no. 3, 642–648.