Available online at www.asric.org
ASRIC Journal on Natural Sciences 1 (2021) 30-37

Scaling Outcome Correlated Binary Big Data Using Estimates of Bivariate Dispersion Parameters

Ahmed M. M. Elsayed [1], Nevein N. Aneis[1]

[1]Al-Obour High Institute For Management & Informatics, Department of Basic Sciences, Kilo 21 Cairo-Belbies Road, P.O. Box 27 Al-Obour City, Egypt

Corresponding Author Email: atabl@oi.edu.eg.

Received 31 May 2021; revised 13 July 2021; accepted 23 August 2021

***Abstract*— Dispersion parameter should be the unity in the case of the univariate Bernoulli data. But there may be some deviations if there is a sequence of the Bernoulli outcomes, that may lead to Binomial case. Over (lower) dispersion criterion is happened if the variance of actual response, var(*y*), is more (less) than the nominal variance as a function of the mean, var(*μ*). This paper presents the mathematical form for estimating and modifying the dispersion parameters for the outcome correlated binary (0,1) Big data, with scalar and matrix values, in Bivariate case. The impact of the estimates of dispersion parameter on the outcome correlated binary Big data is indicated.**

In general, the aim is making the dispersion parameters are close or equal to the unity. The purpose is controlling of marginal probabilities of the correlated binary outcomes. Since the increasing of marginals, increases the values of dispersion estimates. We can use these property to decrease the over-dispersion to close to the unity. The R program and its packages, is used to generate and fit the binary correlated Big data. Scaling and Roots techniques that depend on the estimates of dispersion parameters are used to modify the outcome correlated binary data. We have found that Scaling and Roots processes have similar results and good effects, only for binary Big data. Since the manner is different when deal with Small observations.

***Keywords*— VGAM, VGLM, Binary outcomes, Dispersion parameters, Big data, Scaling data, Correlated data.**

I. INTRODUCTION

The estimation of dispersion parameter in the univariate case can be obtained easily using the Pearson's Chi-square or the Deviance function. The over(lower) dispersion can be deduct from the equation: $\text{var}(y) = \phi \ \text{var}(\mu)$ where ϕ is the univariate dispersion parameter. When $\phi > 1$ this implies the over-dispersion, while $\phi < 1$ implies the lower-dispersion (McCullagh and Nelder, 1989). Many studies have devoted the dispersion criteria in Univariate case, namely, when the Binomial data are used. It is difficult to extend these methods to estimate the dispersion parameters in Bivariate case. Because in Bivariate case, the association between the correlated response variables may be happened. So, we must take this association into account when estimate the dispersion parameter. In Independence case, the estimate of dispersion parameter ϕ is performed as in the univariate case.

Some studies have presented attributes of the overdispersion problem as Smith and Heitjan (1993) provided an appropriate statistical tool to detect extra Binomial variation.

Cook and Ng (1997) described Bivariate logistic-normal mixture model for over-dispersed two state Markov processes. Saefuddin et al. (2011) showed the effect of overdispersion on the hypothesis test of Logistic regression. Simple method proposed by William (1982) was used to correct the effect of overdispersion by taking the inflation factor into consideration. When the overdispersion does not occur or very small overdispersion occurs, dispersion parameter ϕ will be approximately equal to zero, so Y_i exactly follows Binomial distribution, $Bin(n_i, \pi_i)$, and $Var(Y_i) = n_i \pi_i (1 - \pi_i)$, Collett (2003). The value of Pearson's Chi-square statistic depends on $\hat{\phi}$ so, iteration process, is needed to find the optimum value, as a test of $\hat{\phi}$.

Davila et al. (2012) introduced a new approach for modeling Multivariate marginals overdispersed Binomial data. They illustrate this approach by analyzing the data using the Gaussian Copula with Beta-Binomial margins.

In order to model the overdispersion, they used the Beta-Binomial model, a generalization of Binomial distribution, Casella and Berger (2002). Hurvich, and Tsai (1989) presented a regression and time series model selection in Small samples. Vectorized Generalized Additive Model (VGAM) introduced by Yee and Wild (1996) and implemented later by Yee and Others (2003-2015). Vectorized Generalized Linear Model (VGLM) is used in the linear case. For fitting the correlated binary data, we used the VGAM(VGLM) conditional functions and the loglikelihood function of Alternative Quadratic Exponetial Form (AQEF) that proposed by Elsayed et al. (2013).

The loglikelihood function for Bivariate AQEF measure is

$$\ell(y,\beta_1,\beta_2,\alpha)=\sum_{i=1}^{n}\left[\beta_1' xy_{1i}+\beta_2' xy_{2i}+\alpha' xy_{1i}y_{2i}-\log(1+e^{\beta_1' x}+e^{\beta_2' x}+e^{\beta_1' x+\beta_2' x+\alpha' x})\right]. \quad (1)$$

Where, the β_1, β_2 is regression parameters, α is the association parameters.

$$\beta_i'=(\beta_{i0}\ \beta_{i1}),\quad \alpha_i'=(\alpha_{i0}\ \alpha_{i1}),\quad \mu_i(x)=\frac{e^{\beta_i' x}}{1+e^{\beta_i' x}},\quad x'=(1\quad x_i),\quad i=1,2. \quad (2)$$

The estimate of dispersion parameters for Bivariate correlated binary data can be obtained using different methods. The first one when the dispersion parameter is scalar, the second one when we have a matrix values of dispersion parameters. These estimates can be extended to Trivariate and Multivariate correlated binary data.

Elsayed (2016) has proposed a new approach for identifying and estimating the dispersion parameters using Equation (1). In this paper, we are used this approach to modify the correlated binary outcomes Big data using Roots and Scaling techniques for VGLM and VGAM algorithms.

This paper can be organized as follows: The dispersion parameters in Bivariate case is presented in Section 2. Numerical results are presented in Section 3. Results discussion are presented in Section 4. Finally, the conclusion is presented in Section 5.

II. DISPERSION PARAMETERS IN BIVARIATE CASE

In this section, we determine the identification and estimation of a fixed value of dispersion parameter ϕ, and also a matrix of dispersion parameters to extend the effect of overdispersion on the analysis of Bivariate correlated binary data.

A. *Scalar Bivariate Dispersion Parameters*

We can use the variance-covariance matrix of $\boldsymbol{Y}_1$ and $\boldsymbol{Y}_2$ to estimate a scalar dispersion parameter, ϕ, in Bivariate binary outcomes. The estimator of ϕ, for n observations, is

$$\hat{\phi}=\frac{1}{n-p}\sum_{i=1}^{n}\begin{bmatrix} y_{1i}-\hat{\mu}_{1i} & y_{2i}-\hat{\mu}_{2i}\end{bmatrix}\begin{bmatrix} V(\hat{\mu}_{1i}) & Cov(\hat{\mu}_{1i},\hat{\mu}_{2i}) \\ Cov(\hat{\mu}_{2i},\hat{\mu}_{1i}) & V(\hat{\mu}_{2i})\end{bmatrix}^{-1}\begin{bmatrix} y_{1i}-\hat{\mu}_{1i} \\ y_{2i}-\hat{\mu}_{2i}\end{bmatrix}. \quad (3)$$

Where p is number of parameter estimates.

Under independence, this quantity follows, approximately χ^2_{n-p}.

Estimator of ϕ in the independence case is

$$\hat{\phi}=\frac{1}{n-p}\sum_{i=1}^{n}\left[\frac{(y_{1i}-\hat{\mu}_{1i})^2}{\hat{\mu}_{1i}(1-\hat{\mu}_{1i})}+\frac{(y_{2i}-\hat{\mu}_{2i})^2}{\hat{\mu}_{2i}(1-\hat{\mu}_{2i})}\right]. \quad (4)$$

B. Matrix Of Bivariate Dispersion Parameters

Now, we use different values for dispersion parameters, $\phi_{11}, \phi_{22}, \phi_{12}$ and ϕ_{21}, here, $\phi_{12} = \phi_{21}$.
Estimators of dispersion parameters matrix are:

$$\begin{bmatrix} \hat{\phi}_{11} & \hat{\phi}_{12} \\ \hat{\phi}_{21} & \hat{\phi}_{22} \end{bmatrix} = \frac{1}{n-p} \begin{bmatrix} \sum_{i=1}^{n} \frac{(y_{1i} - \hat{\mu}_{1i})^2}{V(\hat{\mu}_{1i})} & \sum_{i=1}^{n} \frac{(y_{1i} - \hat{\mu}_{1i})(y_{2i} - \hat{\mu}_{2i})}{Cov(\hat{\mu}_{1i}, \hat{\mu}_{2i})} \\ \sum_{i=1}^{n} \frac{(y_{2i} - \hat{\mu}_{2i})(y_{1i} - \hat{\mu}_{1i})}{Cov(\hat{\mu}_{2i}, \hat{\mu}_{1i})} & \sum_{i=1}^{n} \frac{(y_{2i} - \hat{\mu}_{2i})^2}{V(\hat{\mu}_{2i})} \end{bmatrix}. \quad (5)$$

We can correct the correlated data using the estimates of dispersion parameters $\hat{\phi}_{11}, \hat{\phi}_{22}$ as:

$$\left.\begin{aligned} Var(Y_{1i}) = \varphi_{11i} V(\mu_{1i}) \Leftrightarrow Var\left(\frac{Y_{1i}}{\sqrt{\varphi_{11i}}}\right) = V(\mu_{1i}), \\ Var(Y_{2i}) = \varphi_{22i} V(\mu_{2i}) \Leftrightarrow Var\left(\frac{Y_{2i}}{\sqrt{\varphi_{22i}}}\right) = V(\mu_{2i}), \end{aligned}\right\} \quad (6)$$

III. NUMERICAL RESULTS

In this section, we generated correlated binary Big data using the "bindata" package in R program in two cases. First case we generate 10,000 observations with a marginal probabilities of the first and second correlated binary variables Y_1 and Y_2 of 0.5 and 0.6 respectively, and the joint probability between two correlated binary variables of 0.25. Second case we generated 500,000 observations with the same conditions. To measure the effect of dispersion parameters on the fitting of correlated binary Big data, first we will find the estimates of dispersion parameters, then will use it to modify the correlated binary outcomes.

Then we make another fit to detect the effect of dispersion parameters this is done using the VGAM package and the "vglm" function. Finally, as a regression problem we must need to generate the covariates also, so we can generate one or more independent variables. In our cases, three covariates are sufficient to use it. We are generated a three independent variables X_1, X_2 and X_3, using the "rpois" function with parameters λ = 3,4 and 5 respectively.

A. *Generating 10,000 Observations*

In this subsection, first we estimate the dispersion parameters $\hat{\phi}_{11}$, $\hat{\phi}_{22}$ for two correlated binary variables Y_1 and Y_2 using the equation (5).

The estimated values are $\hat{\phi}_{11} = 0.7289757$, $\hat{\phi}_{22} = 0.9566033$.
In the next two subsections, we apply the VGLM and VGAM algorithms to fit the binary Big data.

1) Vglm Algorithm

In this subsection, we are used the VGLM technique to fit the two correlated variables Y_1, Y_2 with three covariates x_1, x_2, x_3 and the conditional function "loglinb2" as a family technique, we have the regression equation: $\text{cbind}(y_1, y_2) \sim x_1 + x_2 + x_3, \text{family=loglinb2}$.

Table (1) presents the obtained results in this case.

Table 1

VGLM Results before Modifying the Correlated Data (n = 10,000).

Coefficeints	Estimate	Std. Error	Zvalue	Pvalue
Intercept:1	0.497186	0.078773	6.312 2.7	6e-10 ***
Intercept:2	0.943342	0.079396	11.881	2e-16 ***
Intercept:3	-0.847908	0.041987	-20.195	2e-16 ***
$x_{1:1}$	-0.001035	0.011848	-0.087	0.930
$x_{1:2}$	-0.008716	0.012122	-0.719	0.472
$x_{2:1}$	0.002418	0.010198	0.237	0.813
$x_{2:2}$	0.003468	0.010448	0.332	0.740
$x_{3:1}$	-0.001658	0.009120	-0.182	0.856
$x_{3:2}$	-0.010792	0.009331	-1.157	0.247
Loglikelihood = -13418.07	AIC = 26854.14	BIC = 26919.03		

Now we can modify the correlated outcome binary Big data using two methods.

First one by divide the correlated data by each square root of the estimate of dispersion parameter as:

$$\text{new}y_1 = \frac{y_1}{\sqrt{\hat{\phi}_{11}}},\ \text{new}y_2 = \frac{y_2}{\sqrt{\hat{\phi}_{22}}}$$

Re-estimate the dispersion parameters, we have: $\hat{\phi}_{11} = 0.7842686$, $\hat{\phi}_{22} = 0.9379244.$

Table (2) presents the obtained results in this case.

Table 2

VGLM Results after Modifying the Data using Roots Technique (n = 10,000).

Coefficients	Estimate	Std. Error	Zvalue	Pvalue
Intercept:1	1.031456	0.082052	12.571	2e-16 ***
Intercept:2	1.238264	0.082818	14.952	2e-16 ***
Intercept:3	-1.094477	0.044791	-24.435	2e-16 ***
$x_{1:1}$	-0.001400	0.012130	-0.115	0.908
$x_{1:2}$	-0.009159	0.012333	-0.743	0.458
$x_{2:1}$	0.003013	0.010441	0.289	0.773
$x_{2:2}$	0.003833	0.010631	0.361	0.718
$x_{3:1}$	-0.002194	0.009337	-0.235	0.814
$x_{3:2}$	-0.011376	0.009493	-1.198	0.231
Loglikelihood = -13119.44	AIC = 26256.87	BIC = 26321.76		

Second one: by scaling the correlated outcomes and divide it by the square root of the estimates of dispersion parameters for each variable:

$$\text{new}y_1 = \frac{\dfrac{y_1 - \text{Min}(y_1)}{\text{Max}(y_1) - \text{Min}(y_1)}}{\sqrt{\hat{\phi}_{11}}},\ \text{new}y_2 = \frac{\dfrac{y_2 - \text{Min}(y_2)}{\text{Max}(y_2) - \text{Min}(y_2)}}{\sqrt{\hat{\phi}_{22}}}$$

The new estimates of dispersion parameters are: $\hat{\phi}_{11} = 0.7842686$, $\hat{\phi}_{22} = 0.9379244$.

The obtained results in this case are similar to the results of the first case. This can be arised only in the case of Binary Big data.

2) VGAM Algorithm

In this subsection, we will use the VGAM technigue to fit the correlated outcome binary Big data with each coveriate using the formula : $\text{cbind}(y_1, y_2) = s(x_i), \text{family=binomialff}$, $i = 1, 2, 3$.

For the variable X_1, we have:

Dispersion Parameter for binomialff family = 1.

Residual deviance = 27246.69 , Loglikelihood = -13623.35 .

Approximate Chi-squares for Nonparametric Effects

Coefficients	Chisq	Pvalue	Coefficients	E(y_1)	E(y_2)
s(x_1):1	1.8992	0.59313	Intercept	-0.0236345616	0.465219810
s(x_1):2	6.9189	0.07237	s(x_1)	0.0006750673	-0.009068381

For the variable X_2, we have:

Dispersion Parameter for binomialff family = 1.

Residual deviance = 27247.65 , Loglikelihood = -13623.83

Approximate Chi-squares for Nonparametric Effects

Coefficients	Chisq	Pvalue	Coefficients	E(y_1)	E(y_2)
s(x_2):1	4.0055	0.260755	Intercept	-0.028402565	0.425116981
s(x_2):2	4.4297	0.219104	s(x_2)	0.001705123	0.003213371

For the variable X_3, we have:

Dispersion Parameter for binomialff family = 1.

Residual deviance = 27247 , Loglikelihood = -13623.5

Approximate Chi-squares for Nonparametric Effects

Coefficients	Chisq	Pvalue	Coefficients	E(y_1)	E(y_2)
s(x_3):1	4.0041	0.260523	Intercept	-0.0238968151	0.49059366
s(x_3):2	3.8802	0.275484	s(x_3)	0.0004585959	-0.01052149

B. Generating 500,000 Observations

In this section, we follow the same way as in n =10,000 case, just we change n to be 500,0000 observations.

1) Vglm Algorithm

The estimates of dispersion parameters for the two correlated binary outcomes before modifying these data are: $\hat{\phi}_{11} = 0.7389803$, $\hat{\phi}_{22} = 0.935542$

Table (3) presents the obtained results in this case.

Table 3

VGLM Results before Modifying the Data (N = 500,000).

Coefficeints	Estimate	Std. Error	Zvalue	Pvalue
Intercept:1	0.4982789	0.0110077	45.266	2e-16 ***
Intercept:2	0.8344309	0.0110897	75.244	2e-16 ***
Intercept:3	-0.8391459	0.0059161	-141.842	2e-16 ***
$x_{1:1}$	-0.0011770	0.0016671	-0.706	0.480
$x_{1:2}$	-0.0005859	0.0017009	-0.344	0.730
$x_{2:1}$	0.0011614	0.0014455	0.803	0.422
$x_{2:2}$	0.0015757	0.0014751	1.068	0.285
$x_{3:1}$	0.0018118	0.0012903	1.404	0.160
$x_{3:2}$	0.0002754	0.0013165	0.209	0.834
Loglikelihood = -672913.4	AIC = 1345845	BIC =1345945		

After modification using Roots technique, we have the estimates:

$\hat{\phi}_{11} = 0.6531706$, $\hat{\phi}_{22} = 0.9084052$

Table (4) presents the obtained results in this case.

Table 4

VGLM Results after Modifying the Data using Roots Technique (*N* = 500,000).

Coefficeints	Estimate	Std. Error	Zvalue	Pvalue
Intercept:1	1.0293541	0.0114983	89.522	2e-16 ***
Intercept:2	1.1583903	0.0116052	99.817	2e-16 ***
Intercept:3	-1.0965583	0.0063398	-172.965	2e-16 ***
$x_{1:1}$	-0.0014467	0.0017090	-0.847	0.3973
$x_{1:2}$	-0.0007272	0.0017360	-0.419	0.6753
$x_{2:1}$	0.0014501	0.0014821	0.978	0.3279
$x_{2:2}$	0.0017740	0.0015056	1.178	0.2387
$x_{3:1}$	0.0022135	0.0013231	1.673	0.0943
$x_{3:2}$	0.0004558	0.0013437	0.339	0.7344
Loglikelihood = -655906.3	AIC = 1311831	BIC = 1311931		

Using the scaling method, to modify the data, we obtained the estimates of dispersion parameters as: $\hat{\phi}_{11} = 0.7944488$, $\hat{\phi}_{22} = 0.9084052$.

As in the n =10,000 observations, the VGLM results in this case are the same as n = 500,000. This can be arised only in Binary Big data.

2) Vgam Algorithm

As in the previous subsections, we obtained the next results for 500,000 observations:

For the variable X_1, we have:

Dispersion Parameter for binomialff family =1.
Residual deviance = 1366441, Loglikelihood = -683220.5.

Approximate Chi-squares for Nonparametric Effects

Coefficients	Chisq	Pvalue	Coefficients	E(y_1)	E(y_2)
s(x_1):1	7.2587	0.064097	Intercept	0.006258790	0.4034463123
s(x_1):2	4.6824	0.196549	s(x_1)	-0.001056217	-0.0003439202

For the variable X_2 , we have:

Dispersion Parameter for binomialff family =1.
Residual deviance = 1366445, Loglikelihood = -683222.3.

Approximate Chi-squares for Nonparametric Effects

Coefficients	Chisq	Pvalue	Coefficients	E(y_1)	E(y_2)
s(x_2):1	3.4349	0.32925	Intercept	-0.0003177063	0.397069469
s(x_2):2	4.1404	0.24668	s(x_2)	0.0008509257	0.001335797

For the variable X_3 , we have:

Dispersion Parameter for binomialff family = 1.
Residual deviance = 1366427, Loglikelihood = -683213.4.

Approximate Chi-squares for Nonparametric Effects

Coefficients	Chisq	Pvalue	Coefficients	E(y$_1$)	E(y$_2$)
s(x_3):1	5.9847	0.112357	Intercept	-0.005693456	0.4029313497
s(x_3):2	18.5355	0.000339	s(x_3)	0.001756106	-0.0001054667

IV. RESULTS DISCUSSION

The results in Section 3 can be summarized and combined for two the cases of correlated outcome binary Big data as follow:

A. VGLM

In general, the standard errors tend to be lower when we are modifying small data. For the two cases of correlated binary Big data, the standard errors are increasing after modifying the data using Roots or Scaling techniques for all estimators.

n = 10,000					
Case	$\hat{\phi}_{11}$	$\hat{\phi}_{22}$	Loglikelihood	AIC	BIC
Before	0.7289757	0.9566033	-13418.07	26854.14	26919.03
Roots	0.7842686	0.9379244	-13119.44	26256.87	26321.76
Scaling	0.7842686	0.9379244	-13119.44	26256.87	26321.76

For n =10,000, the value of $\hat{\phi}_{11}$ is increasing after modifying (Roots and Scaling) the data. In the other hand the value of $\hat{\phi}_{11}$ is decreasing after modifying (Roots and Scaling) the data. All values of Loglikelihood, AIC and BIC is decreasing after modifying the data. This reflects that these measures are good after modifying the binary correlated outcomes.

n = 500,000					
Case	$\hat{\phi}_{11}$	$\hat{\phi}_{22}$	Loglikelihood	AIC	BIC
Before	0.7389803	0.9355420	-672913.4	1345845	1345945
Roots	0.6531706	0.9084052	-655906.3	1311831	1311931
Scaling	0.7944488	0.9084052	-655906.3	1311831	1311931

For n = 500,000, the values of $\hat{\phi}_{11}$ and $\hat{\phi}_{22}$ are decreasing after modifying the data using Roots process, but $\hat{\phi}_{11}$ is increasing after modifying the data using Scaling process. This means that the scaling process is good for $\hat{\phi}_{11}$ when increasing the number of observations. Also, as in n =10,000, the values of Loglikelihood, AIC and BIC is decreasing after modifying the data. This also reflects that these measures are good after modifying the binary correlated outcomes Big data.

B. VGAM

Approximate Chi-squares for Nonparametric Effects (*n* = 10,000)

Coefficients	Chisq	Pvalue	Coefficients	Chisq	Pvalue	Coefficients	Chisq	Pvalue
s(x_1):1	1.8992	0.59313	s(x_2):1	4.0055	0.260755	s(x_3):1	4.0041	0.260523
s(x_1):2	6.9189	0.07237	s(x_2):2	4.4297	0.219104	s(x_3):2	3.8802	0.275484

Measures	x_1	x_2	x_3
Residual deviance	27246.69	27247.65	27247
Loglikelihood	-13623.35	-13623.83	-13623.5

Approximate Chi-squares for Nonparametric Effects (*n* = 500,000)

Coefficients	Chisq	Pvalue	Coefficients	Chisq	Pvalue	Coefficients	Chisq	Pvalue
s(x_1):1	7.2587	0.064097	s(x_2):1	3.4349	0.32925	s(x_3):1	5.9847	0.112357
s(x_1):2	4.6824	0.196549	s(x_2):2	4.1404	0.24668	s(x_3):2	18.5355	0.000339 ***

Measures	x_1	x_2	x_3
Residual deviance	1366441	1366445	1366427
Loglikelihood	-683220.5	-683222.3	-683213.4

Dispersion parameter for binomialff family =1. For the two correlated binary Big data, the independent variable X_3 has the lowest residual deviance, this reflects the importance of this variable to the model. Also, has a significant effect with 2nd additive predictor in the case of 500,000 observations. But there are not one of the other independent variables X_1 , X_2 have significant effects, with 5% significant level. For Loglikelihood value, the independent variable X_1 has the lowest value in the case of 10,000 observations. While the independent variable X_3 has the lowest value in the case of 500,000 observations.

V. CONCLUSION

This paper presents the mathematical form for estimating and modifying the dispersion parameters for the outcome correlated binary (0,1) Big data, with scalar and matrix values, in Bivariate case. The effect of dispersion estimates on the outcome correlated binary Big data is indicated. The marginals of two correlated binary outcomes variables effect on the values of estimates of dispersion parameters. Using these property, we can motivate the tends of estimates to close to the unity. The program R and its packages are used to generate and fit the Big data.

Roots and Scaling methods are used to modify the outcome correlated binary Big data. We have found that Scaling and Roots processes have similar results, and good effects only for binary Big data. Since the manner is different when deal with small obsevations.

ACKNOWLEDGMENTS:

For Al-Obour High Institute For Management & Informatics.

REFERENCES

[1] McCullagh, P. and Nelder, J. (1989). Generalized linear models (second edition), Chapman and Hall, London, United Kingdom.
[2] Smith, P. and Heitjan, F. (1993). Testing and adjusting for departures from nominal dispersion in generalized linear models, Applied Statistics 42, 1 : pp 31–34.
[3] Cook R. and Ng, E. (1997). A logistic-bivariate normal model for overdispersed two-state Markov process, Biometrics 53, 1 : 358-364.
[4] Saefuddin, A. Setiabudi, A. and Achsani, N. (2011). The effect of overdispersion on regression based decision with application to Churn Analysis on Indonesian Mobile Phone Industry, European Journal of Scientific Research 60, 4 : 584-592.
[5] William, D. (1982). Extra-binomial variation in logistic linear models, Applied Statistics 31 : 144-148.
[6] Collett, D. (2003). Modeling binary data (Second edition), Chapman and Hall, London, United Kingdom.
[7] Davila, E., Lopez L. and Dias, G. (2012). A statistical model for analyzing interdependent complex of plant pathogens, Revista Colombiana de Estadistica Numero especial en Bioestadistica 35, 2 : 255-270.
[8] Casella, G. and Berger , R. (2002). Statistical inference (second edition), Duxbury Press, Florida, United States.
[9] Hurvich, C. and Tsai, C. (1989). Regression and time series model selection in small samples, Biometrika, 76 : 297–307.
[10] Yee, T. and Wild, C. (1996). Vector generalized additive models. Journal of the Royal Statistical Society, Series B, Methodological, 58 : 481–493.
[11] Yee, T. and Hastie, T. (2003). Reduced-rank vector generalized linear models. Statistical Modelling, 3 :15–41.
[12] Yee, T. (2004). A new technique for maximum-likelihood canonical Gaussian ordination. Ecological Monographs, 74 : 685–701.
[13] Yee, T. (2006). Constrained additive ordination. Ecology, 87 : 203–213.
[14] Yee, T. and Stephenson, A. (2007). Vector generalized linear and additive extreme value models. Extremes, 10 : 1–19.
[15] Yee, T. (2008). The VGAM Package. R News, 8 : 28–39.
[16] Yee, T. (2010). The VGAM package for categorical data analysis. Journal of Statistical Software, 32 : 1–34.
[17] Yee, T. (2014). Reduced-rank vector generalized linear models with two linear predictors. Computational Statistics and Data Analysis, 71 : 889–902.
[18] Yee, T. (2015). Vector Generalized Linear and Additive Models: With an Implementation in R. New York, USA: Springer.
[19] Elsayed, A. Islam, M. and Alzaid, A. (2013). Estimation and test of measures of association for correlated binary data, Bulletin of the Malaysian Mathematical Sciences Society 2, 36, 4 : 985-1008.
[20] Elsayed, A. (2016). A new approach for dispersion parameters. Journal of Applied Mathematics and Physics (JAMP) 4, 8 : 1554-1566.