The Empirical Distribution of Wald, Score, Likelihood Ratio, Hosmer-Lemeshow (HL), and Deviance for a Small Sample Logistic Regression Model

Krisada Lekdee and Lily Ingsrisawang

Abstract— **The objective of this research involved the investigation of the empirical distribution of the statistics that were used to examine whether or not the logistic regression model fit the data when the sample size was small. These statistics were Wald, Score, Likelihood Ratio, Hosmer-Lemeshow (HL), and Deviance. The simulation study was employed. The data were simulated for the logistic regression models with 2 dummy and 2 continuous variables with small sample sizes of 30, 50, and 100. For each sample size 1,000 simulation runs were made. The empirical distributions of those statistics (upper tails only) were compared with the Chi-square distributions. The upper tail of the distribution is an important segment since it is used for hypothesis testing. The levels of significance (α) were set at 0.01, 0.05, and 0.10. The study found that, for the sample sizes of 30, 50, and 100, the distribution of HL was closest to the Chi-square distribution at all levels of significance. If the HL was used and assumed to have the Chi-square distribution, the level of significance would change only slightly.**

***Index Terms*— Deviance, Hosmer-Lemeshow (HL), Likelihood Ratio, logistic regression, Score, small sample, Wald**

I. Introduction

A logistic regression model is used to find the relation between a dependent variable whose value is either 0 or 1 and independent variables whose values can be either continuous or discrete numbers. More than one independent variable can be in the logistic regression model. An example of medical research using the logistic regression model was the study of the death of cancer patients within 5 years after the treatment. The value of the dependent variable is 1 if the patient died and 0 if the patient survived. The researcher was interested in factors or independent variables that can predict the probability or risk of death in the patient. The functional form of the logistic model is as follows.

Let y have a Bernoulli distribution and the probability that an event will occur is P($Y=1$)= P , not occur is P($Y=0$) = $1-P$, $E(Y)=P$, and $V(Y)=1-P$.

Manuscript received January 5, 2010.

K. Lekdee is a Ph.D. (Statistics) student at the Faculty of Science, Kasetsart University (corresponding author, e-mail: krisadal@ hotmail.com).

L. Ingsrisawang is with the Department of Statistics, Faculty of Science, Kasetsart University (e-mail: fscilli@ku.ac.th).

Define $\log it(P)=\log(\frac{P}{1-P})=\beta_0+\beta_1 X_1+\beta_2 X_2+...+\beta_p X_p$ where $X_1, X_2,.., X_p$ are independent variables and $\beta_0,\beta_1,\beta_2,..,\beta_p$ are regression coefficients. The logistic model to predict the probability or risk that an interesting event will occur is

$$P(Y=1\,|\,X_1,X_2,..,X_p)=\frac{\exp(\beta_0+\beta_1X_1+\beta_2X_2+...+\beta_pX_p)}{1+\exp(\beta_0+\beta_1X_1+\beta_2X_2+...+\beta_pX_p)}$$

.

A maximum likelihood and an iterative method are employed to estimate the parameters in the logistic regression model. There are several iterative methods for solving nonlinear equations, but the most popular and efficient ones are Newton-Raphson and Fisher's Scoring methods [1]. There are several statistics to be used for the logistic model evaluation, such as Wald, Score, Likelihood Ratio, Hosmer-Lemeshow (HL), and Deviance. These statistics have asymptotic Chi-square distributions [2], [3], [4], [5]. That is when a sample size is large, its distribution is close to the Chi-square distribution. In some applications the sample size may be small, so its distribution is questionable that whether or not it is close to the Chi-square distribution.

There is a wide range of research that uses the logistic model to analyze the data and a wide range study of the statistics that are used for the model evaluation. For example, Chen and et. al. [6] studied a mathematic tool for inference in the logistic regression with small-sized data sets. They found that with the small sample size of 54, large sample theory should not be used since it was not reliable. The null hypothesis of global test was rejected when using the Likelihood Ratio and the Score statistics, but the Wald and the Deviance gave the opposite conclusion, accepting the null hypothesis. Pulkstenis and Robinson [7] studied the goodness-of-fit tests for the logistic regression model with continuous covariates. They found that the distribution of Deviance was far from the Chi-square distribution, so they proposed the new statistic for examining how well the model fitted the data. The method used to develop the new statistic was similar to the method that Hosmer and Lemeshow [4] used to develop the Hosmer-Lemeshow statistic. A variety of simulations were performed comparing the new statistic to the Hosmer-Lemeshow statistic. Kramer [8] studied the distribution of the HL by simulating the logistic model with 20 continuous and 3 dummy variables with the sample size of 50,000. He found that the distribution of the HL is not significantly different from the Chi-square distribution.

However, the study of the distribution of the 5 statistics, Wald, Score, Likelihood Ratio, Hosmer and Lemeshow (HL), and Deviance at the same time, comparing with the Chi-square distribution, when a sample size is small has not presented yet. This paper will present the empirical distribution of those statistics compared with the Chi-square distribution when the sample is small. The result of the study will provide information to researchers to select the suitable statistic to assess the logistic model when a sample size is small.

II. Study Method

A simulation was employed for this study. The data were simulated for the logistic regression model with the dependent, Y, which had a Bernoulli distribution with parameter P, the probability that an event would occur. There were 4 independent variables, 2 dummy variables (X_1, X_2) and 2 continuous variables (Z_1, Z_2). The two continuous independent variables were assumed to have normal distributions. There were 3 situations which are the sample sizes of 30, 50, and 100. For each situation 1,000 Monte Carlo simulation runs were made. The levels of significance were set at 0.01, 0.05, and 0.10. The SAS program supported by the department of statistics, Kasetsart university, was used for the simulation. The processes of the study were as follows;

1. Generated random values for the independent dummy variable (X_1, X_2) which were assumed to have Bernoulli distributions whose parameter values were 0.4 and 0.6 respectively. The functions in SAS used for generating values for X_1 and X_2 were RAND ('BERNOULLI',0.40) and RAND('BERNOULLI',0.60) respectively.
2. Generated random values for the independent continuous variables (Z_1, Z_2) which were assumed to have normal distributions. Given that Z_1 had normal distribution with mean 0.5 and variance 0.1 and Z_2 had normal distribution with mean 0.5 and variance 0.5. The functions used for generating the values for Z_1 and Z_2 were 0.5+RANNOR(-1)*0.1 and 0.5+RANNOR(-1)*0.5 respectively.
3. Given that $\beta_0 = -1.4$, $\beta_1 = 1$, $\beta_2 = 0.5$, $\beta_3 = 1$, $\beta_4 = 1$
4. Let xbeta = $\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 Z_1 + \beta_4 Z_2$, so xbeta = -1.4 + X_1 +0.5 X_2 + Z_1 + Z_2
5. Calculated phat = $P(Y = 1 | X_1, X_2, Z_1, Z_2)$ = exp(xbeta)/(1+exp(xbeta)) according to the form of the logistic regression model. The value of phat was between 0 and 1.
6. Generated random values for dependent variable Y. The value of Y was ether 0 or 1. The random values were generated from a uniform distribution. If the generated random number was more than phat, then the Y was equal to 1. If the generated random number was less than phat, then the Y was equal to 0. The function used to generate the random number was RANUNI (-1).
7. Generated the values for X_1, X_2, Z_1, Z_2, and Y according to 1- 6 until the numbers of the values for each variable were equal to the sample sizes of 30, 50, and 100.
8. Used PROC LOGISTIC and PROC GENMOD to estimate the parameters in the logistic regression model. Recorded the values of the statistics, Wald, Score, Likelihood Ratio, HL, Deviance, from each simulation run.
9. For each situation 1,000 simulation runs were made, so 1,000 values for each statistic were recorded.
10. Analyzed the distribution of each statistic as follows;

10.1 We made a histogram and a QQ plot for each statistic and compared them with the Chi-square distribution. Compared the histogram and the QQ plot of the Wald, Score, Likelihood ratio with the Chi-square distribution with 4 degrees of freedom (the number of independent variables). Compared the histogram and the QQ plot of the HL with the Chi-square distribution with 8 degrees of freedom, and compared the histogram and the QQ plot of the Deviance with the distribution of the Chi-square with the degree of freedom equal to the difference between the number of observations and the number of parameters.

10.2 At the upper tail of its empirical distribution, recorded the critical values of each statistic from the quantile values at 0.99, 0.95 and 0.90 for α = 0.01, 0.05, 0.10 respectively. We compared its empirical critical values with the Chi-square critical values with the same degree of freedom. The Chi-square values were obtained from the CINV(i-i/1000,DF) function where i = 1, 2, …, 1,000. They were the Chi-square values at α = 0.001, 0.002, 0.003, …,0.999 respectively.

10.3 We used the empirical critical values of the 5 statistics to find the Chi-square critical p-values when the 5 statistics were assumed to have the Chi-square distribution.

III. Result

1. Histogram and QQ Plot

The histogram and QQ Plot of each statistic showed that when the sample sizes were 30, 50, and 100, the distribution of HL was closest to the Chi-square distribution. The histograms and the QQ Plots were shown in Fig. 1-6.

2. Comparing the empirical critical values of each statistic with the Chi-square critical values

The empirical critical values of the Wald , Score, Likelihood Ratio, HL, and Deviance compared with the critical values of the Chi-square at α = 0.01, 0.05, and 0.10 were shown in table I.

The table I showed that among those statistics, when the sample sizes were 30, 50, and 100, the empirical distribution of HL was closest to the Chi-square distribution at all levels of significance. For hypothesis testing of the logistic model fitting to the data, if each statistic was assumed to have the Chi-square distribution, the level of significance that was set for testing would change. For example, for the sample size of 30, if the empirical HL was used for testing hypothesis, the null hypothesis would be rejected at α = 0.01, when the HL was greater than 18.60. If the HL was used and assumed to have the Chi-square distribution, the null hypothesis would be rejected when the HL was greater than 20.09 which caused the level of significance change from 0.01 to 0.003 as shown in table II.

Table II showed that if all statistics were assumed to have the Chi-square distribution, the significant level would change. For the sample sizes of 30, 50, and 100, if the HL was used and assumed to have the Chi-square distribution, the level of significance would change only slightly.

3. Percentages of rejecting null hypothesis when each statistic was assumed to have the Chi-square distribution

The percentages of rejecting the null hypothesis at α = 0.01, 0.05, and 0.10 when each statistic was assumed to have the Chi-square distribution were shown in table III.

Table III showed that if the HL was assumed to have the Chi-square distribution, for testing hypothesis for the sample sizes of 30, 50, and 100, the percentage of rejecting the null hypothesis would be closest to α at any level of significance.

IV. Conclusion and Discussion

The simulation was employed to study of the empirical distribution of the Wald, Score, Likelihood Ratio, Hosmer-Lemeshow (HL), and Deviance which were used for assessing the logistic regression model when the sample size was small. The study found that for the sample sizes of 30, 50, and 100, the distribution of HL was closest to the Chi-square distribution at all levels of significance. Hosmer and Lemeshow [4] developed the Hosmer–Lemeshow test to be a commonly used test for assessing the goodness of fit of a logistic regression model. The test is similar to a Chi-square goodness of fit test and has the advantage of partitioning the observations into groups of approximately equal size, and therefore there are less likely to be groups with very low observed and expected frequencies. The observations are grouped into deciles based on the predicted probabilities. Hosmer and Lemeshow [4] recommended sample sizes should be greater than 400. This study found that even though the sample size was not large, the distribution of HL was still close to the Chi-square distribution. The reason was that only the upper tails of them are compared, not the whole distribution. However if the HL was used and assumed to have the Chi-square distribution, the change of the level of significance should be considered.

Competing Interests

The authors declare that they have no competing interests.

References

[1] T. Smith, B. Smith, and W. Honner. (2009, Sept 27). PROC GENMOD for Analysis of Correlated Outcome Data Using the LOGIT Link Function. Available: http://www.lexjansen.com/wuss/2004/data_analysis/i_das_proc_genmod_for_analys.pdf..

[2] A. Agresti. Categorical Data Analysis. 2nd ed. New Jersey: John Wiley & Sons, 2002.

[3] D. Collett. Modelling Binary Data. 2nd ed. Florida: Chapman and Hall/CRC, 2003.

[4] D.W. Hosmer and S. Lemeshow. Applied Logistic Regression, 2nd ed. New York: John Wiley and Sons, 2000.

[5] J.E. Corrente, S.G. Gomes, and A. Paula. "Correction Factor for the Deviance Distribution of Proportion Data." Mat Estat., 2002, 20:175-193.

[6] T_H. Chen, C_Y. Chen, H-C.P. Yang, and C-W. Chen. "A Mathematic Tool for Inference in Logistic Regression with Small-Sized Data Sets: A Practical Application on ISW-Ridge Relationships." Mathematical Problems in Engineering, 2008, ID 186372: 1-12.

[7] E. Pulkstenis and T.J. Robinson. "Two goodness-of-fit tests for logistic regression models with continuous covariates." Stat Med., 2002, 21(1): 79-93.

[8] A.A. Kramer. (2008, Sept 30). Using ODS to Perform Simulations on Statistics for SAS Procedures. Available: http://analytics.ncsu.edu/sesug/2005/SA05_05.PDF.

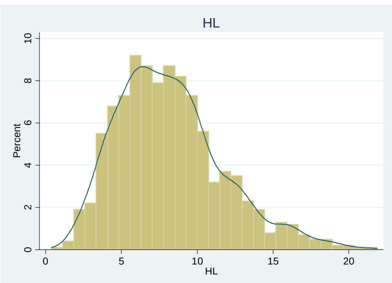

Fig. 1 Histogram of HL, n = 30

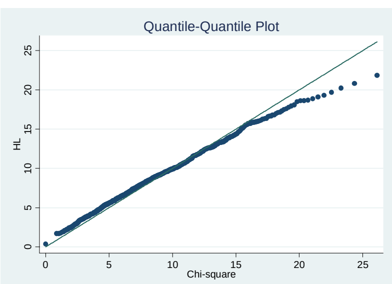

Fig. 2 QQ Plot of HL, n = 30

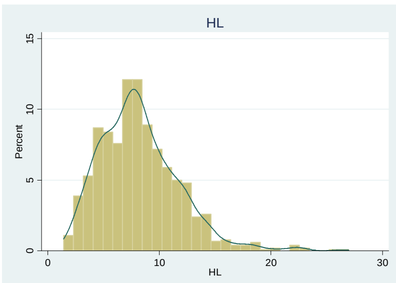

Fig. 3 Histogram of HL, n = 50

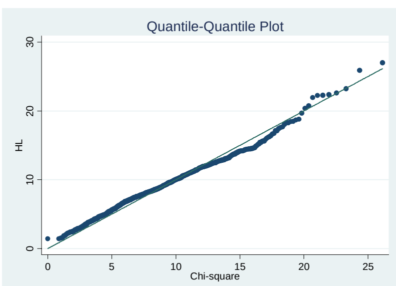

Fig. 4 QQ Plot of HL, n = 50

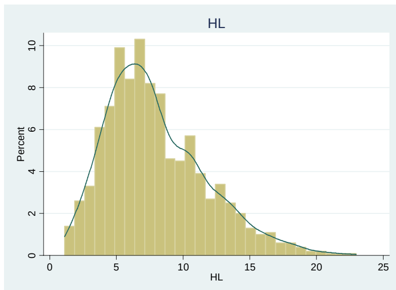

Fig. 5 Histogram of HL, n = 100

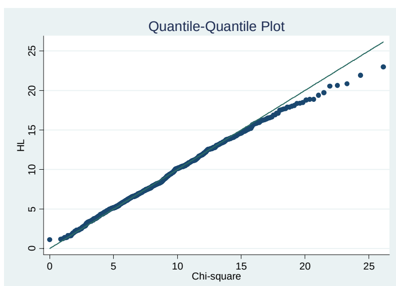

Fig. 6 QQ Plot of HL, n = 100

ISBN: 978-988-18210-5-8
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

Table I Comparison of the empirical critical values of each statistic with the Chi-square critical values at α = 0.01, 0.05, and 0.10

n=30	α	Chi-square DF=4	Wald DF=4	Score DF=4	Likelihood Ratio DF=4	Chi-square DF=8	HL DF=8	Chi-square DF=25	Deviance DF=25
	0.01	13.28	8.52	16.34	23.30	20.09	18.60	44.31	40.23
	0.05	9.49	7.65	13.22	17.17	15.51	15.11	37.65	38.86
	0.10	7.78	7.23	11.79	15.02	13.36	12.90	34.38	38.00
n=50	α	Chi-square DF=4	Wald DF=4	Score DF=4	Likelihood Ratio DF=4	Chi-square DF=8	HL DF=8	Chi-square DF=45	Deviance DF=45
	0.01	13.28	12.95	21.03	26.79	20.09	20.38	69.96	67.40
	0.05	9.49	11.52	16.60	20.24	15.71	14.41	61.66	66.10
	0.10	7.78	10.76	14.70	17.62	13.36	12.72	57.51	65.24
n=100	α	Chi-square DF=4	Wald DF=4	Score DF=4	Likelihood Ratio DF=4	Chi-square DF=8	HL (DF=8)	Chi-square DF=95	Deviance DF=95
	0.01	13.28	21.84	29.98	35.17	20.09	18.79	129.97	134.88
	0.05	9.49	19.20	24.62	28.35	15.71	15.07	118.75	132.67
	0.10	7.78	17.67	21.97	24.61	13.36	13.26	113.04	130.88

Table II Change of the level of significance when each statistic was assumed to have the Chi-square distribution

n=30	α	Chi-square DF=4	Wald DF=4	Score DF=4	Likelihood Ratio DF=4	Chi-square DF=8	HL DF=8	Chi-square DF=25	Deviance DF=25
	0.01	0.01	0.000	0.047	0.150	0.01	0.003	0.01	0.000
	0.05	0.05	0.002	0.233	0.351	0.05	0.045	0.05	0.119
	0.10	0.10	0.038	0.391	0.486	0.10	0.084	0.10	0.382
n=50	α	Chi-square DF=4	Wald DF=4	Score DF=4	Likelihood Ratio DF=4	Chi-square DF=8	HL DF=8	Chi-square DF=45	Deviance DF=45
	0.01	0.01	0.004	0.159	0.245	0.01	0.010	0.01	0.000
	0.05	0.05	0.203	0.400	0.470	0.05	0.030	0.05	0.321
	0.10	0.10	0.412	0.555	0.607	0.10	0.075	0.10	0.620
n=100	α	Chi-square DF=4	Wald DF=4	Score DF=4	Likelihood Ratio DF=4	Chi-square DF=8	HL DF=8	Chi-square DF=95	Deviance DF=95
	0.01	0.01	0.392	0.538	0.591	0.01	0.005	0.01	0.137
	0.05	0.05	0.715	0.763	0.793	0.05	0.042	0.05	0.701
	0.10	0.10	0.827	0.845	0.857	0.10	0.095	0.10	0.898

Table III Percentages of rejecting the null hypothesis when each statistic was assumed to have the Chi-square distribution

n	α	Wald (%)	Score (%)	Likelihood Ratio (%)	HL (%)	Deviance (%)
30	0.01	0.0	4.7	15.0	0.3	0.0
	0.05	0.2	23.3	35.1	4.5	11.9
	0.10	3.8	39.1	48.6	8.4	38.2
50	0.01	0.4	15.9	24.5	1.0	0.0
	0.05	20.3	40.0	47.0	3.0	32.1
	0.10	41.2	55.5	60.7	7.5	62.0
100	0.01	39.2	53.8	59.1	0.5	13.7
	0.05	71.5	76.3	79.3	4.2	70.1
	0.10	82.7	84.5	85.7	9.5	89.8

ISBN: 978-988-18210-5-8
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)