The Performance of Hosmer-Lemeshow Test in Case of the Incorrect Model

Mohammed Elsadig Mahmoud Mohammed

Al Neelain University

GCNU Journal

ISSN: 1858-6228

Volume 16 - 2021
Issue: 03

Graduate College
Al Neelain University

Graduate College Journal – NU
Vol.16 -2021, No. (3)
ISSN: 1858-6228, http://www.neelain .edu.sd

مجلة الدراسات العليا - جامعة النيلين
المجلد (16) –2021، العدد (3)
الرقم الدولي الموحد للدوريات: 1858-6228

The Performance of Hosmer-Lemeshow Test in Case of the Incorrect Model

Mohammed Elsadig Mahmoud Mohammed
Faculty of Mathematical Sciences and Statistics, Al Neelain University, Khartoum, Sudan.
Corresponding author E-mail: Mohammedelsadig534@gmail.com

Abstract

The fact that Hosmer-Lemeshow test is based on formation of groups for variables values poses a number of questions. One of these is how many groups should be formed? Will a different number of groups change the final result? Another is to what extent the power of the test is affected by factors such as sample size and population distribution characteristics? The main aim of this paper is to examine the performance of Hosmer-Lemeshow test when the fitted logistic model is the incorrect model under some factors such as the changing number of groups, sample size and population distribution that is expected to affect its performance to see whether its performance in case of incorrect model better than its performance in the case of correct model. This is accomplished through techniques of analysis and simulation using RStudio package.The analytical approach is composed of conduction of Hosmer-Lemeshow test, formation of groups, etc, while the simulation approach is entirely based on the ideas of data generation based on specified circumstances, sample selection steps, iterations steps and so on. The results concluded that when 10 groups are formulated the value of the test statistic is increased with sample size and when the number of groups is changed the test performance is affected by changing the number of groups especially when the sample size is small. Moreover, when the simulation technique is used to check for the effect of repeated sample and to see whether the covariate's distribution and the sample size will affect the power of the test or not and to what extent, it has been revealed that the power of the test increases with the increase in both the sample size and the variance value and accordingly its performance in case of the incorrect model and through its interaction with the control factors is better compared to its performance in the case of the correct model.

Keywords: Hosmer-Lemeshow Test, Repeated sample, Incorrect Model, Simulation

Introduction

The statistical analysis of dichotomous outcome variables is often interpreted with the use of logistic regression methods (Kleinbaum, 1994). The logistic model is widely used in public health, medicine, epidemiology and other fields. Logistic regression sometimes called the logistic model or logit model, analyzes the relationship between one or more independent variables and a categorical dependent variable, and estimates the probability of occurrence of an event by fitting data to a logistic curve. The goal of an analysis using this method is the same as that of any model-building technique used in statistics: to find the best fitting model to use it to determine relations or for prediction purposes.

Suppose that we have n binary observations of the form$y_i, i = 1,2, \dots, n$. Let Y denote a dichotomous outcome variable, which may assume values "1" if the event occurs and "0" otherwise. Let the vector x'= $(x_1,x_2,\dots,x_p)$ denote a set of p predictor variables. Let the conditional probability that the outcome is present be denoted by $P(Y = 1|x) = \pi(x)$. The logistic model which relates the probability of the event occurring to the predictor variables x is given by:

$$P(Y = 1|x) = \frac{e^{\beta_0+\beta_1 x_1+\cdots+\beta_p x_p}}{1+e^{\beta_0+\beta_1 x_1+\cdots+\beta_p x_p}} = \pi(x) \quad (1)$$

And thus

$$P(Y = 0|x) = 1 - \pi(x)$$

After performing the logit transformation on π(x) in equation (1) we obtain the following multiple logistic regression models:

$$g(x) = logit[x(\pi)] = ln\left[\frac{\pi(x)}{1-\pi(x)}\right] = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p \quad (2)$$

The variables in $x'(x_1, x_2, \dots x_p)$ can be discrete, continuous or binary. In this research, the situation to be considered is the univariate situation where $p = 1$ predictor variable, and thus $\boldsymbol{x'} = (\boldsymbol{x_1})$.

Consequently, equation (1) becomes:

$$\pi(x) = \frac{e^{\beta_0+\beta_1 x}}{1+e^{\beta_0+\beta_1 x}} \quad (3)$$

While the logit of the simple logistic regression model is given by the equation:

$$g(x) = ln\left[\frac{\pi(x)}{1-\pi(x)}\right] = \beta_0 + \beta_1 x \quad (4)$$

Taking into consideration the term *logit*, it is defined as the natural logarithm of the odds – the term that logistic regression derives its name from. The importance of this transformation is that $g(x)$ has many of the desirable properties of a linear regression model. The logit, $g(x)$ is linear in its parameters, may be continuous, and may range from -∞ to ∞, depending on the range of x. The inverse logit function in equation (3) give us the probabilities of events we need, while the logit function in equation (4) gives us the linear function that relates outcomes to the variables.

The linear logistic model is a member of the class of generalized linear models (Nelder and Wedderburn, 1972). This class of generalized models allows $\pi(x)$ to be related to the linear component $(\beta_0 + \beta_1 x)$ by the use of a logistic link function. The link function is the function of dependent variable that yields a linear function of the independent variables. In the case of a linear regression model, it is the identity function, since the dependent variable, by definition, is linear in the parameters. In the logistic regression model the link function is the logit transformation $g(x)$.

After fitting a model to the observed data, one of the essential steps is to investigate how will the proposed model fits the observed data. One method which is used to determine the suitability of the fitted logistic model is goodness of fit test. In logistic regression model there are many goodness of fit tests proposed all have individual advantages and disadvantages (Tsiatis, 1980).

Hosmer and Lemeshow (1980) has introduced a test for goodness of fit for logistic regression model depending on dividing the variables values into groups according to the estimated probabilities obtained from the fitted logistic model and then compares observed and expected probabilities within these groups. The idea applies the concept of the contingency table but creates the table based on a partition of the estimated probability of positive response $\hat{\pi}_i$ into 10 groups.

For the $y = 1$row, estimates of the expected values are obtained by summing the estimated probabilities over all subjects in a group. For the $y = 0$row, the estimated expected value is obtained by summing, over all subjects in the group, one minus the estimated probability.

These groups are often referred to as "deciles of risk". This term comes from health sciences research where the outcome $y = 1$often represents the occurrence of some disease (Hosmer and Lemeshow, 2000). The first group contains approximately n/G subjects having the smallest estimated probabilities, the second group contains approximately n/G subjects having the second smallest estimated probabilities, and the last group contains approximately n/G subjects having the largest estimated probabilities. Where n represents the size of the sample.

A formula defining the calculation of Hosmer-Lemeshow test statistic is as follow:

$$\hat{C} = \sum_{k=0}^{1}\sum_{g=1}^{G}\frac{\left(o_{kg} - e_{kg}\right)^2}{e_{kg}}$$

Where:o_{kg} is the observed frequency of subjects who have had the event occur and not occur in each group g $(g = 1,2,\cdots,G)$. e_{kg} is the expected frequency of subject who have had the event occur and not occur in each group g.

In a previous paper written by the same author under the title of: (The performance of Hosmer – Lemeshow test in case of the correct model) the focus was on the performance of the Hosmer – Lemeshow goodness of fit test when the logistic regression model is the correct one, where it examined the performance under factors such as changing the number of groups, sample size and population distribution. The results of analysis and simulation showed that the test performance was not affected by changing those factors.

Objectives

The main aim of this paper is to examine the performance of Hosmer-Lemeshow test when the fitted logistic model is the incorrect model under some factors expected to affect its magnitude such as the changing number of groups, sample size and population distribution to see whether its performance in case of incorrect model better than its performance in the case of correct model.

Method and Materials

The data were generated using the RStudio System for Windows (version 1.1.444). A single predictor variable was initially generated using a random function in RStudio. The logistic model fit was the incorrect model. This means that our model is incorrectly specified. To examine the performance of Hosmer-Lemeshow test, two aspects of the model are considered; the distribution of the covariate, i.e. various distributions, and the values of the coefficients, the coefficients are chosen to assure sufficiency to allow estimation.

The most effective way of achieving the aforementioned goal was to design a layout in a factorial arrangement, so that patterns or differences might be discerned between variations of the factors.

The factors under consideration that were predetermined to vary are the following:

1. The distribution of the predictor variable X (three levels: standard normal or normal or uniform)
2. The value of the variance of the predictor variable (two levels: 1 or 2)
3. The sample size of the generated data sets (four levels: 50, 100, 200, 500)
4. The parameters used to generate the logistic data (two parameters: $\beta_0 = 0, \beta_1 = 1$)

To determine the behavior of the test statistic on data sets having predictor variables with a skewed distribution, the predictor variable was generated under the chi-squared distribution with:

1. A value of variance of 2 or 4.
2. Values of parameters as:$\beta_0 = -0.85, -2.5, \beta_1 = 1$.These parameters are used to generate logistic data.

Tables I, II and III best illustrate the overall layout of the factorial arrangement as well as the one additional special case.

Table I. Factorial arrangement for generated data

Standard Normal Distribution			
Variance (x)=1			
$\beta_0 = 0$		$\beta_1 = 1$	
N 50,100,200,500			
Normal Distribution			
Variance (x)=1		Variance (x)=2	
$\beta_0 = 0$	$\beta_1 = 1$	$\beta_0 = 0$	$\beta_1 = 1$
n 50,100,200,500		n 50,100.200.500	
Uniform Distribution			
Variance (x)=1		Variance (x)=2	
$\beta_0 = 0$	$\beta_1 = 1$	$\beta_0 = 0$	$\beta_1 = 1$
N 50,100,200,500		n 50,100,200,500	

Table II. Special case for generated data from skewed distribution

Chi-Squared Distribution			
Variance (x)=2		Variance (x)=4	
$\beta_0 = -0.85$	$\beta_1 = 1$	$\beta_0 = -2.5$	$\beta_1 = 1$
N 50,100,200,500		n 50,100,200,500	

Table III. Enter simulation design

Settings	Covariate Distribution	Sample Size	β_0	β_1
1	N (0 , 1)	50	0	1
		100	0	1
		200	0	1
		500	0	1
2	N (0.73 , 1)	50	0	1
		100	0	1
		200	0	1
		500	0	1
3	N (0.73 , 2)	50	0	1
		100	0	1
		200	0	1
		500	0	1
4	U (-1 , 2.46)	50	0	1
		100	0	1
		200	0	1
		500	0	1
5	U (-1.72 , 3.18)	50	0	1
		100	0	1
		200	0	1
		500	0	1
6	$X^2_{(1)}$	50	-0.85	1
		100	-0.85	1
		200	-0.85	1
		500	-0.85	1
7	$X^2_{(2)}$	50	-2.5	1
		100	-2.5	1
		200	-2.5	1
		500	-2.5	1

The generated random predictor variables were controlled by a seed value, which was arbitrarily chosen. The given seed was used to obtain the first observation in the stream of the random numbers and reproduce results i.e.it produces the same sample again and again (RStudio Software, 2009). When we generate random numbers without set seed function, it will produce different samples at different time of execution. For each setting illustrated in Table III, a data set with 1000 observations was generated to build 4 separate data set of size 50,100,200 and 500.

Let $x_{s1}, x_{s2}, \dots, x_{sn}$ denote the elements of a randomly generated random variable, which will be taken to be the predictor variable in the setting 2 up to 5. Moreover, Let x_{sn}~ N(0.73,σ_v^2) for $s = 2,3$ denote the settings for the predictor variable generated from a normal distribution, such that x_{sn} has a mean of 0.73, and variance σ_v^2. Let x_{sn}~$U(a,b)$ for $s = 4,5$ denote the settings for the predictor variable generated from a uniform distribution on the interval between a and b, such that x_{sn}also has a mean of 0.73, and variance σ_v^2. The sample size of the data set is denoted by: n = 50, 100, 200, and 500. Also $\sigma_v^2 = v$, where v = 1, 2 corresponds to the variance size. The simulation process is explained in details as follows:

1. Taking into considerationx_{sn}~$N(0.73, \sigma_v^2), x_{sn}$ the data generation process is implemented as follows:

$$u = 0.73, \sigma^2 = \sigma_v^2, seed = se$$

$$x1_{sn} = u + \left[\sqrt{\sigma_v^2}\right] \times N_{se}(0,1)$$

$$x_{sn} = round(x1_{sn}) which\ rounds x1_{sn} to\ one\ decimal\ point$$

Where:$N_{se}(0,1)$: is a randomly generated number from the standard normal distribution and the seed value is used to obtain the first observation in the stream of the random number.

2. Regarding x_{sn}~$U(a,b), x_{sn}$the data generation was carried out through a transformation process of a random variable which was generated from a Uniform distribution on the interval between a and b as follows:

$$u = 0.73, \sigma^2 = \sigma_v^2, seed = se$$

$$x1_{sn} = a + (b-a) \times U_{se}(0,1)$$

$$x_{sn} = round(x1_{sn}) which\ rounds x1_{sn} to\ one\ decimal\ point$$

Where$U(0,1)$: represents a randomly generated number from the Uniform distribution on the interval (0,1).

3. In order to generate x_{sn} such that it has a mean of 0.73 and a variance$\sigma_v^2 = v$, a and b had to be solved from following equations:

$$\text{Mean of a uniform random variable } \rightarrow \frac{a+b}{2} = 0.73$$

$$\text{Variance of a uniform random variable } \rightarrow \frac{(b-a)^2}{12} = \sigma_v^2$$

$$\text{Resulting in } a = -1, b = 2.46 \text{ where } \sigma_v^2 = 1$$

$$a = -1.72, b = 3.18 \text{ where } \sigma_v^2 = 2$$

4. Considering the special case for settings 6 and 7, the variance values are: $\sigma_v^2 = 2$ and 4 respectively.
5. Within the same step where x_i's were generated, the probability of an event occurring as a result of the x_i's was calculated according to the logistic model:

$$P(Y = 1|X = x_i) = \frac{e^{\beta_0+\beta_1 x_i}}{1+e^{\beta_0+\beta_1 x_i}} = \pi(x_i),$$

Where: $\beta_0 = 0, \beta_1 = 1$ and $\beta_0 = -0.85, -2.5, \beta_1 = 1$for the special chi-squared case. The terms$\beta_0$andβ_1 partially control the proportion of events$\pi(x_i)$. In the simulation process, the outcome variable y is generated by comparing an independently generated $U(0,1)$ random variable u, to the true logistic probability value, using the rule $y = 1$ if $u \leq \pi(x_i)$ and $y = 0$ otherwise (Bernard et al. (2018)).

6. The process outlined above was performed 1000 times resulting in a data set with 1000 observations denoted by:

$$\{(x_1, y_1), (x_2, y_2), \cdots (x_{1000}, y_{1000})\}.$$

7. The next step was to partition the 1000 observation into four individual data sets of sizes 50,100,200 and 500 in the following manner:

Table IV. Size and elements of the generated bivariate observations

Size of data set	Elements of generated data set
50	$\{(x_1, y_1), (x_2, y_2), \cdots, (x_{50}, y_{50})\}$
100	$\{(x_{51}, y_{51}), (x_{52}, y_{52}), \cdots, (x_{150}, y_{150})\}$
200	$\{(x_{251}, y_{251}), (x_{252}, y_{252}), \cdots, (x_{450}, y_{450})\}$
500	$\{(x_{451}, y_{451}), (x_{452}, y_{452}), \cdots, (x_{950}, y_{950})\}$

8. The process described above was replicated 1000 times for each data set in each scenario, producing 4000 data sets generated for each setting. Therefore, there were a total of 5 settings ×4000 data sets = 20000 data sets generated, this is in addition to 2 settings × 4000 data sets = 8000 for the data sets having chi-squared predictor variables. Therefore in total, there were 28000 data sets created, and analyzed using logistic regression.

In all simulation, initially, a sample of size $n = 50, 100, 200$ and 500 is generated as values of the covariate followed by the generation of the outcome variable to follow a logistic model with X^2 as covariate when the covariate distribution is from *Standard Normal*, and with X^{-2} when the covariate distribution is from *Normal* or *Uniform*, and with $X^{-1}$when the covariate is from *Chi-squared* distribution, but the model is fitted continually with linear X as covariate (Hjort, 1988), taking into consideration that the fitted model is incorrectly specified.

Univariate logistic regression using the GLM (generalized linear model) function in RStudio was applied to the generated data to determine the parameter estimates. The GLM function uses the maximum likelihood algorithm (RStudio Software, 2009) to compute the parameter estimates of $\beta_0$and β_1. After the incorrect logistic regression model has been fitted for each setting, the outcome y and model fitted probabilities are passed to the hoslem.test function, choosing $g = 10$ groups.

Results

After following the implementation of the above described steps, the results are stated as:
First let's calculate the test statistic without repeated sample to check how it will perform. Table V below shows the Hosmer-Lemeshow goodness of fit test statistic for each incorrect fitted model. The results from Table V show that, in all settings the value of the test statistic is increased with sample size, and the test never gives significant evidence of poor fit when the sample size is 50. But in most settings, with a sample size of 100 or more, the test gives us evidence of poor fit. Then the result becomes misleading when the setting is *Standard Normal* and the sample size is 100 for the test gives evidence of good fit for this data set. But if we use group numbers as:$g < 10$ or $10 < g < 15$ we will findthe evidence of poor fit for the test.

Table V. Value of the $\hat{C}$ statistic, Degree of freedom and P-value for the Incorrect Model for each setting.

Covariate Distribution	Sample Size	$\hat{C}$	df	p-value
N (0 , 1)	50	12.849	8	0.1172
	100	13.212	8	0.1048
	200	22.630	8	0.0039
	500	80.986	8	0.0000
N (0.73 , 1)	50	7.4711	8	0.4868
	100	21.058	8	0.0070
	200	30.507	8	0.0002
	500	72.557	8	0.0000
N (0.73 , 2)	50	10.694	8	0.2196
	100	18.424	8	0.0183
	200	36.431	8	0.0001
	500	91.010	8	0.0000
U (-1 , 2.46)	50	12.519	8	0.1295
	100	20.557	8	0.0084
	200	21.250	8	0.0065
	500	24.764	8	0.0017

U (-1.72 , 3.18)	50	13.028	8	0.1109
	100	21.169	8	0.0067
	200	59.142	8	0.0000
	500	59.226	8	0.0000
$X^2_{(1)}$	50	10.844	8	0.2107
	100	16.171	8	0.0400
	200	17.652	8	0.0240
	500	49.257	8	0.0000
$X^2_{(2)}$	50	8.1861	8	0.4155
	100	33.220	8	0.0006
	200	43.686	8	0.0000
	500	71.938	8	0.0000

Taking into consideration different numbers of groups, the main objective here becomes to see how the test's p-value changes in case of the incorrect model. These numbers of groups can be:$g = 5, g = 6$, up to $g = 15$. Table VI shows the p-value for Hosmer-Lemeshow statistic computed from the different number of groups.

From the bellow table we note that, when the sample size is 50 and the variance equal to 1 and the distributions are: the *Standard Normal* distribution, *Normal* distribution and *Uniform* distribution, the final result of the test is not affect by the change of the number of groups, meaning that, the p-value in all groups do not give evidence of poor fit, but when the variance change to 2 the final result of the test is affected by the change of some groups most of which are less than 10 groups, and this is true when the distribution is *Normal* distribution, but when the distribution is *Uniform* distribution that result is affected only by the change of some groups which are less than 10 groups.

In the case of *Chi-squared* distribution and when the sample size is 50 and the variance equal to 2, the p-value when the number of groups is 5 or 9 was 0.05 and this is equal to the value of α. In this case, we reject the null hypothesis, and this is a sign of poor fit. This indicates that, the final result of the test is affected by the change of some groups which are less than 10 groups. Meanwhile, if the variance changes to 4, the change in the number of groups did not affect the final result of the test. Whereas, the p-value in all groups did not give evidence of poor fit.

When the sample size is 100, and the variance is equal to 1 then, the final result of the test is affected by the change of some groups most of which are more than 10 groups, whereas the p-value in those groups gives evidence of good fit. This is true in case of *Normal* distribution or *Uniform distribution*. But when the variance changes to 2 and when the distribution is *Normal*, it is not affected by the variance change. On the contrary, when the distribution is *Uniform*, then the variance change helps to detect the lack of fit in those groups. In the case of *Standard Normal* distribution the final result of the test is affected by the change in the number of groups which are less than 10, and most number of groups which are more than 10 groups whereas, its change also helped to detect the lack of fit.

In the case of *chi-squared* distribution, and when the sample size is 100 and the variance value equals to 2, the final result of the test is clearly affected by changing whole number of groups which are less than and more than 10 groups whereas, the p-value in those groups give evidence of good fit, and when the variance change to 4 then its change helps to detect the lack of fit in all groups which are less than 10 groups and some of the groups which are more than 10 groups.

When the sample size is 200 or 500 and the variance equal to 1, and also, when the variance changes to 2 then we find evidence of poor fit in the all groups and this means that the final result of the test is not affect by changing the number of groups and the variance.

In the case of *Chi-squared* distribution, we find evidence of poor fit in all groups except when $g = 15$, which means that changing the number of groups does not have a significant effect on the final result of the test, and this is true when the sample size is 200 and the variance equals to 2, and when the variance changes to 4 then we find evidence of poor fit in the all groups which means that changing the variance reflects lack of fit in that group. And when the sample size is 500 and the variance equals to 2 or changed to 4 then, changing the number of groups is not affecting the final result of the test, and the p-value shows evidence of poor fit in all groups.

To finish, let's check how the test performs in repeated sample, since we want to see whether the covariate's distribution and the sample size will affect the power of the test. Table VII below presents the power; the percent of rejection of the hypothesis of test fit at the $\alpha = 0.05$ level

Table VI. P-value of the Hosmer-Lemeshow $\hat{C}$ statistic for incorrect model computed from different number of groups

Covariate Distribution	Sample Size	$g = 5$	$g = 6$	$g = 7$	$g = 8$	$g = 9$	$g = 10$	$g = 11$	$g = 12$	$g = 13$	$g = 14$	$g = 15$
N (0 , 1)	50	0.8487	0.5531	0.6245	0.7211	0.5730	0.1172	0.2831	0.3943	0.4892	0.3941	0.3991
	100	0.0292	0.0045	0.0120	0.0185	0.0071	0.1048	0.0269	0.0372	0.0436	0.0299	0.1084
	200	0.0014	0.0011	0.0003	0.0002	0.0005	0.0039	0.0034	0.0011	0.0060	0.0031	0.0164
	500	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
N (0.73 , 1)	50	0.3293	0.4292	0.5309	0.4124	0.4715	0.4868	0.4071	0.5318	0.1919	0.5791	0.6654
	100	0.0596	0.1113	0.0210	0.0187	0.0233	0.0070	0.2425	0.2064	0.1251	0.0836	0.0350
	200	0.0047	0.0040	0.0009	0.0007	0.0002	0.0002	0.0018	0.0023	0.0009	0.0001	0.0001
	500	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
N (0.73 , 2)	50	0.0577	0.0342	0.0389	0.0156	0.0217	0.2196	0.1615	0.1496	0.0416	0.1575	0.2094
	100	0.0098	0.1449	0.0433	0.0527	0.0566	0.0183	0.2494	0.2011	0.1104	0.0584	0.0386
	200	0.0006	0.0006	0.0001	0.0000	0.0000	0.0001	0.0000	0.0002	0.0005	0.0002	0.0002
	500	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
U (-1 , 2.46)	50	0.2073	0.1374	0.1152	0.2042	0.2254	0.1295	0.2945	0.1252	0.1752	0.1935	0.3454
	100	0.0214	0.1158	0.0144	0.0124	0.0146	0.0084	0.1557	0.0498	0.0919	0.1355	0.1892
	200	0.0070	0.0395	0.0075	0.0206	0.0042	0.0065	0.0009	0.0145	0.0016	0.0025	0.0026
	500	0.0042	0.0001	0.0009	0.0006	0.0077	0.0017	0.0010	0.0008	0.0002	0.0001	0.0009
U (-1.72,3.18)	50	0.0115	0.1436	0.0220	0.0619	0.0886	0.1109	0.2420	0.3819	0.2244	0.2354	0.0933
	100	0.0589	0.0069	0.0148	0.0136	0.0210	0.0067	0.0109	0.0028	0.0100	0.0033	0.0483
	200	0.0000	0.0001	0.0003	0.0000	0.0000	0.0000	0.0000	0.0006	0.0005	0.0000	0.0000
	500	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
$X^2_{(1)}$	50	0.0464	0.0989	0.0608	0.1318	0.0540	0.2107	0.2894	0.0678	0.1475	0.0576	0.2559
	100	0.0585	0.1671	0.0664	0.1135	0.3217	0.0400	0.1859	0.1829	0.0578	0.1698	0.2759
	200	0.0024	0.0016	0.0096	0.0398	0.0431	0.0240	0.0088	0.0195	0.0349	0.0506	0.0850
	500	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
$X^2_{(2)}$	50	0.7057	0.7232	0.3220	0.1511	0.1877	0.4155	0.2898	0.4891	0.2750	0.2217	0.3345
	100	0.0029	0.0128	0.0026	0.0026	0.0020	0.0006	0.0003	0.0001	0.0867	0.0327	0.0634
	200	0.0000	0.0000	0.0000	0.0000	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	500	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

.

Table VII. Simulated percent of rejection at $\alpha = 0.05$ for incorrect model for each of the settings

Distribution\ Sample Size	N (0 , 1)	N (0.73,1)	N (0.73,2)	U(-1,2.46)	U(-1.72,3.18)	$X^2_{(1)}$	$X^2_{(2)}$
50	28.7	19.8	30.8	13.8	25.8	19.6	46.6
100	68.4	47.3	74.5	37.8	65.7	60.4	85.5
200	97.3	89.2	99.6	81.4	98.5	97.9	99.7
500	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Table VII reflects the results of the different combinations of sample sizes and type of population used in the analysis. It indicates that, from 1000 simulations, the Hosmer-Lemeshow test gives significant evidence of poor fit in all settings. First, let's look at the test performance from the sample sizes, in all settings, with a sample size of 50 and 100, the test has low power to detect poor or lack of fit, except when the covariate distribution is from *Chi-squared* with variance equal to 4, the power is then 85.5 percent for sample of size 100. High power is attained for sample of size 200 or up, and for all settings the power increases rapidly with sample size. For every setting, the power is 100 percent for sample of size 500, so the test does not detect the poor fit more. Table VII also reveals that the variance affects the test performance, since the power in all settings increases when the variance changed between 1 and 2 or 2 and 4.

Discussion

Taking into consideration that the main objective of this paper was to examine the performance of Hosmer-Lemeshow test when the fitted logistic model is the incorrect model along with various circumstances; accordingly the results came out pointing to several things. In case of using 10 groups and in all the various settings, the test statistic increased with the sample size, and it did not give any evidence of poor fit, and more precisely when the sample size is 50. In most settings, the test gives evidence of poor fit in case of sample size of 100 or more. On the contrary, if the selected distribution is the *Standard Normal* accompanied with sample size of 100, the result is misleading, since the test clearly gives evidence of good fit for this data set. This may be due to chance and further investigation is needed.

Moreover, if the number of groups is changed, Hosmer-Lemeshow test is affected by the sample size whereas, the power of the test increased with the increase in the sample size. Also, the test is affected by the change in: number of groups (10 & more), population distribution, especially to the *Uniform* distribution, small sample size, and variance which helped very much to detect lack of fit. In the case of *chi-squared* distribution the final result of the test is affected by changing all the number of groups which are less than and more than 10 groups, but when the variance changed then its change helped to detect the lack of fit. And when the sample size was large, the result of the test remained unaffected by any change in number of groups or variance.

Accordingly simulation process was then applied to check for how well the test will perform in repeated sample and to see whether the covariate's distribution and the sample size will affect the power of the test. Considering thousand numbers of iterations and generally, the results indicated that, in all settings, Hosmer-Lemeshow test gives significant evidence of poor fit. Except for the case of using *Chi-Squared* distribution along with a variance value equal to 4 and a sample size of 100 or more, the detected power was a high one(85.5 per cent).Thus, one can say that the performance of Hosmer – Lemeshow test in case of incorrect model and through its interaction with the control factors, is relatively better compared to its performance in the case of the correct model (presented by previous preceding work), whereas, it's use as a criterion to detect lack of fit is better than using it as a criterion to detect goodness of fit.

Conclusion

Finally, we can say that in the case of the incorrect logistic regression model and when Hosmer-Lemeshow test is used the results concluded that the test is affected by changing the number of groups especially when the sample size was small. Moreover, when the simulation technique is used to check for the effect of repeated sample and to see whether the covariate's distribution and the sample size will affect the power of the test or not and to what extent, it has been revealed that the power of the test increases with the increase in both the sample size and the variance value. The final result of the test does not affect by the changing number of groups or variance value when the sample size is large, and accordingly its performance in case of the incorrect model and through its interaction with the control factors, is relatively better compared to its performance in the case of the correct model. Whereas, it's use as a criterion to detect lack of fit is better than using it as a criterion to detect goodness of fit.

Recommendation

Based on the results of this paper, the following recommendations are available to serve as guidelines for the use of Hosmer-Lemeshow test:

1. The Hosmer-Lemeshow statistic $\hat{C}$ should be used to confirm the lack of fit of the model after using other goodness of fit tests.
2. The results of this paper are not certified and not applicable because they depend entirely on simulations which are methods that, in general, not exact. Simulations yield an empirical result, i.e. numbers, which are valid for the particular simulated experiment only. The second reason is that choosing of the cutoff points in groups may be inappropriate in all settings. Nevertheless, the results may be appropriate in situations that resemble those of the conducted simulation process.

References

1. Bernard P. Zeigler, Alexandre Muzy and Ernesto Kofman (2018). Theory of Modelling and Simulation, 3rd Edition. Chapman & Hall: New York
2. Hjort, N. L. (1988). Estimating the logistic regression when the model is incorrect. Technical Report, Norwegian Computing Centre, Oslo, Norway
3. Hosmer, D.W., and Lemeshow, S. (2000). Applied Logistic Regression, 2nd Edition. Wiley: New York
4. Kleinbaum, D. G. (1994). Logistic Regression: A Self-Learning Text. Springer-Verlag, New York.
5. Nelder, J.A., and Wedderburn, R.W.M. (1972). Generalized linear models. Journal of the Royal Statistical Society Series A 135, 761-768
6. RStudio Software (2009). RStudio Guide, Version 1.1,444. RStudio Inc.
7. Tsiatis, A. A. (1980). A note on a goodness-of-fit test for the logistic regression model. Biometrika, 67, 250-251.