Journal of Scientific and Human Dimensions
مجلة الأبعاد العلمية والإنسانية
ISSN: 3106-1702
Volume 2, Issue 4, 2026 Pages: 448-453
Journal homepage: https://jshd.com.ly/index.php/jshd/en

Goodness of Fit Statistic for Logistic Regression Models with only Continuous Predictor

Jalal A. Moaiti[1 *], Ali Meelad Abd Allateef [2], Nawal Mohammed Othman Mustafa[3]
[1,2,3] Department of Statistics, Faculty of Sciences Al-Marj, Benghazi University, Libya

إحصائي لاختبار جودة التوفيق للنماذج الخطية اللوجستية في حالة وجود متغير متغير تفسيري متصل وحيد

جلال عبد الله امعيطي سعيد [1*]، علي ميلاد عبد اللطيف [2]، نوال محمد عثمان مصطفي [3]
[3،2،1] قسم الإحصاء، كلية العلوم المرج، جامعة بنغازي، ليبيا

*Corresponding author: jalal.abdullah@uob.edu.ly

Received: August 14, 2026	Accepted: October 25, 2026	Published: November 15, 2026

Copyright: © 2026 by the authors. This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

Abstract:
Logistic regression models have become widely used statistical models to model binary responses. However, assessing the goodness of fit (GOF) in these models, when we have only one continuous predictor is not clear. The deviance and the Pearson chi square tests are unreliable and don't follow chi square distribution. Furthermore, some researches showed that Hosmer and Lemeshow test is slightly conservative. In this research, the Pearson chi square test and its moments were investigated via a simulation study. The simulation showed that the problem is that the variance of this statistic is not stable and doesn't equal to theoretical variance for chi square test. However, to stabilize the variance Fisher approximation to standard normal was used to solve or relax this problem. The approximated standard normal test had slightly stable variance and good type one error rate, compared with chi square test.

Keywords: Logistic regression models, goodness of fit, Fisher Approximation, Variance Stabilization, Simulation Study

الملخص
نماذج الانحدار اللوجستي (Logistic Regression Models) من النماذج الإحصائية واسعة الاستخدام لنمذجة متغيرات الاستجابة الثنائية (Binary Responses)، إلا أن تقييم جودة المطابقة (Goodness of Fit - GOF) في هذه النماذج لا يزال غير واضح عند وجود متغير تفسيري مستمر واحد فقط (Continuous Predictor)، حيث إن اختبار الانحراف (Deviance Test) واختبار مربع كاي لبيرسون (Pearson Chi-square Test) يُعدّان غير موثوقين ولا يتبعان توزيع مربع كاي، علاوة على أن بعض الأبحاث أظهرت أن اختبار هوسمر وليميشو (Hosmer and Lemeshow Test) يميل إلى التحفظ قليلاً. وفي هذا البحث، جرى فحص اختبار مربع كاي لبيرسون وعزومه (Moments) من خلال دراسة محاكاة (Simulation Study)، والتي أظهرت أن الإشكالية تكمن في عدم استقرار تباين (Variance) هذه الإحصاءة وعدم مساواته للتباين النظري لاختبار مربع كاي. ولتثبيت هذا التباين، استخدم تقريب فيشر إلى التوزيع الطبيعي المعياري (Fisher Approximation to Standard Normal) لمعالجة هذه المشكلة او التخفيف من حدتها وحيث اظهر الاختبار الطبيعي المعياري المقرّب استقراراً نسبياً في التباين ومعدلاً جيداً لخطأ النوع الأول (Type I Error Rate) مقارنة باختبار مربع كاي.

الكلمات المفتاحية: الانحدار اللوجستي، ملاءمة النموذج، العزوم، تقريب فيشر، استقرار التباين، دراسة محاكاة.

1. Introduction

In many fields of sciences such as medical, social and other sciences, need to model binary responses on some predictors. These models known as logistic regression models. However, Assessing the goodness of fit for logistic regression models with continuous predictors, using traditional tests like the Pearson chi-square and deviance tests often yield invalid results. The estimated residuals in the logistic models are the differences between the actual response value, which is 0 or 1, and the estimated probabilities of the model. The estimated residuals are fractions between -1 and 1. Therefore, the sum of squares of the residuals will be negligible and cannot be approximated as a chi square distribution with degrees of freedom related to the whole summation (Agresti, 2002). The alternative test which has been commonly used is the Hosmer and Lemeshow (1980) test statistic. This test statistic groups the response variable into subgroups based on the estimated probabilities. That means instead of dealing with single observations, we deal with a frequency of each group so the residuals of the model represent the differences between the actual frequencies and the estimated frequencies. However, a recent study showed this test statistic is slightly conservative; it is not recommended for use in such models (Evans & Hosmer, 2004). Furthermore, it might have low power to detect departure of model fit because it is only based on grouping the response while it ignores the predictors region (Hosmer, Hosmer, le Cessie, & Lemeshow, 1997; le Cessie & van Houwelingen, 1991). Another test statistic was used by Pan and Lin (2005). They proposed graphical and numerical goodness of fit statistics for generalized linear mixed effects models. These statistics used the cumulative sum of the residuals, which have a distribution that can be approximated as a zero-mean Gaussian process under the true model. They generated the realizations of these processes by using Monte Carlo simulation and then compared the observed process of the model visually and analytically to the simulated realization. For large samples, their test statistics gave good type I error rate and power.

In this work a simulation study was applied to investigate the standardized Pearson residual and its moments. The study showed that the problem of this test statistic is that its variance is not stable. To solve this problem, fisher approximation from chi distribution to standard normal distribution was tried. However, this approximation is usually used to stabilize the variance. A simulation study was used on the approximated standard normal test, and it gave slightly stable variance and good type I error rates, compared with standardized chi square test.

2. Logistic regression models

In many practical studies in most sciences, we need to model binary response variables for which the response outcomes are success or failure. For example, one might be interested in modeling the existence of specific disease of some patients on some observed predictors like age, family history, etc. In this case, the response variable would take the value of 1 if the disease exists and 0 if not. One of the statistical models that could be used to deal with binary response data is the logistic regression model. The binary random response can be defined as,

$$y = \begin{Bmatrix} 1 & if\ the\ outcome\ is\ success \\ 0 & if\ the\ outcome\ is\ failure \end{Bmatrix}$$

The above binary random response follows the Bernoulli random variable with probability of success p and probability of failure $(1-p)$. Similarly, the sum of the responses over a sample n would have a binomial distribution. The general form for the logistic regression model can be written as

$$logit\ (\boldsymbol{p}) = \log\left(\frac{\boldsymbol{p}}{1-\boldsymbol{p}}\right) = \boldsymbol{X\beta}. \qquad (1)$$

The right-hand side of the above equation is called the systematic component, where $\boldsymbol{X}$ is a (n x k) design matrix and $\boldsymbol{\beta}$ is a (k x 1) vector of parameters.

The term $logit\ (\boldsymbol{p}) = \log\left(\frac{\boldsymbol{p}}{1-\boldsymbol{p}}\right)$ is a logit transformation from probabilities to a continuous random response; it is called the link function. From equation 2.1, we can write the probability of success vector as

$$\boldsymbol{p} = \frac{e^{\boldsymbol{X\beta}}}{1+e^{\boldsymbol{X\beta}}}.$$

2.1 Estimation

Maximum likelihood. Let $y_1, \ldots, y_m$ be independent random variables such that y_i is the number of successes in the group or class i, n_i is the number of trials in the i^{th} class, and p_i is the probability of success in the i^{th} class. In

that case, y_i would have the binomial distribution with parameters (n_i, p_i). The likelihood function for the i[th] observation could be written as

$$l(p_i; y_i) = \binom{n_i}{y_i} p_i^{y_i}(1-p_i)^{n_i - y_i}.$$

For the independent observations, the likelihood function would be

$$l(\boldsymbol{p}; \boldsymbol{y}) = \prod_{i=1}^{m} \binom{n_i}{y_i} p_i^{y_i}(1-p_i)^{n_i - y_i}.$$

Therefore, the log likelihood function $L(.)$ could be written as

$$L(\boldsymbol{p}; \boldsymbol{y}) = \sum_{i=1}^{m} \left[log \binom{n_i}{y_i} + y_i \log\left(\frac{p_i}{1-p_i}\right) + n_i(1-p_i) \right]. \quad (2)$$

Using Equation 2.1, the log likelihood function in terms of the $X_{i's}$ and $\beta_{i's}$ could be written as follows:

$$L(\boldsymbol{\beta}; \boldsymbol{y}) = \sum_{i=1}^{m} \left[log \binom{n_i}{y_i} + y_i \sum_{i=1}^{k} X_{ij}\beta_j - n_i \log\left(1 + \exp \sum_{i=1}^{k} X_{ij}\beta_j\right) \right].$$

To find the estimators for the coefficients $\boldsymbol{\beta}$, we would derive the log likelihood function with respect to $\boldsymbol{\beta}$ and maximize this function. First, the derivative of the log likelihood function 2.2 with respect to p_i is

$$\frac{\partial L}{\partial p_i} = \sum_{i=1}^{m} \frac{y_i - n_i p_i}{p_i(1-p_i)}.$$

Using the relation $p_i = \frac{e^{\beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik}}}{1 + e^{\beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik}}}$, we can find $\frac{\partial p_i}{\partial \beta_j}$.

By applying the chain rule, we can find the derivative of the log likelihood function with respect to β_j as follows:

$$\frac{\partial L}{\partial \beta_j} = \frac{\partial L}{\partial p_i} \frac{\partial p_i}{\partial \beta_j},$$

were

$$\frac{\partial L}{\partial p_i} = \sum_{i=1}^{m} \frac{y_i - n_i p_i}{p_i(1-p_i)},$$

and

$$\frac{\partial p_i}{\partial \beta_j} = \frac{e^{\beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik}}}{[1 + e^{\beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik}}]^2} X_{ij}.$$

Therefore, the maximum likelihood estimator for the j[th] coefficient could be obtained by solving the following score equations:

$$\frac{\partial L}{\partial \beta_j} = \sum_{i=1}^{m} \frac{y_i - n_i p_i}{p_i(1-p_i)} \frac{e^{\beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik}}}{[1 + e^{\beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik}}]^2} X_{ij} = 0.$$

These equations could be solved by applying the iterative method using a computer program. The same procedure would be applied for all entire parameters.

For the vector of parameters $\boldsymbol{\beta}$, the iterative equations could be derived by using Newton-Raphson method, which is derived by using Taylor expansions around initial parameter for a function $f'(\theta)$ of parameter θ. This procedure is based on the general iterative equation,

$$\theta_{t+1} = \theta_t - \frac{f'(\theta)}{f''(\theta)},$$

where $\theta_{t+1} = \theta_t$ (converges) when $f'(\theta) = 0$.

Using this method, we can derive the iterative equations for the parameters vector as

$$\widehat{\boldsymbol{\beta}}_{\boldsymbol{new}} = \widehat{\boldsymbol{\beta}}_{\boldsymbol{old}} + \left[-L''(\widehat{\boldsymbol{\beta}}_{\boldsymbol{old}})\right]^{-1}.L'(\widehat{\boldsymbol{\beta}}_{\boldsymbol{old}}).$$

The maximum likelihood estimators for the parameters would be efficient, consistent, and asymptotically normally distributed.

2.2 Goodness of fit statistics

There are some types of residuals can be obtained from a logistic regression model, such as Pearson residuals, deviance residuals, standardized Pearson residuals, and standardized deviance residuals. In this work the standardized Pearson residuals was used with transformed standardized Pearson residuals.

The Pearson residuals of the logistic regression models can be as,

$$r_i = \frac{y_i - \hat{\pi}_i}{\sqrt{\hat{\pi}_i(1-\hat{\pi}_i)}},$$

where $y_i \in \{0,1\}$ is the observed outcome and $\hat{\pi}_i$ is the fitted probability. However, the residuals of generalized linear models don't have constant variance like the residuals of regression models, and should be adjusted for the leverage of each point such that,

$$r_i = \frac{y_i - \hat{\pi}_i}{\sqrt{\hat{\pi}_i(1-\hat{\pi}_i)(1-h_i)}},$$

Where h_i is the leverage and it is the i[th] element of the hat matrix,

$$H = W^{\frac{1}{2}} X\,(X^T W X)^{-1} X^T W^{\frac{1}{2}} \quad ; \quad W = diag[\hat{\pi}_i(1-\hat{\pi}_i)].$$

The standardized residuals approximately have standard normal distribution and then the sum of squares of these residuals can be approximated as chi square distribution,

$$S = \sum_i^n r_i^2 \sim \chi_n^2$$

The above sum of squares can be considered as goodness of fit statistic and it has mean and variance,

$$E(S) = n \quad and \quad V(S) = 2n.$$

Fisher's approximation

The fisher's approximation from chi square to standard normal distribution can be used to stabilizing the variance. Variance-stabilizing transformation via the Delta method to transform a Chi-squared distributed random variable into an approximately standard normal distribution.

$$S \sim \chi_n^2$$

Using Fisher's approximation we can approximate,

$$\sqrt{2S} \sim N(\sqrt{2n-1}, 1)$$

$$Z = \sqrt{2S} - \sqrt{2n-1} \sim N(0,1)$$

In this research this approximated statistic will be used as a goodness of fit statistic for the logistic model.

3. Simulation Study

The simulation study was applied by generate data from a simple logistic model with only one continuous predictor. The data were generated from the following model with different sample sizes (20, 40, 80, 100, 150), and a continuous predictor x:

$$\log\left(\frac{p}{1-p}\right) = 1 + 1.2\,x$$

The simulation was applied on 1000 replication for each model and the standardized Pearson statistic and transformed chi to standard normal using Fisher's approximation were collected over the replication. The mean and the variance were calculated for each statistic and the following are the results.

Table 1 Mean and variance of Standardized Pearson sum of squares for different sample sizes.

Sample size	20	40	80	100	150
Mean	17.891	39.607	80.016	99.787	150.74
Variance	26.716	112.85	312.90	150.27	1310.8

From table 1 it is clear that the means of standard Pearson statistic are close to the sample sizes which is the theoretical expected value for chi square distribution. However, the variances are not stable and they gave values are not close to the theoretical variance of chi square distribution (2n).

Table 2 Mean and variance of Transformed Standardized Pearson sum of squares for different sample sizes.

Sample size	20	40	80	100	150
Mean	5.8910	8.8443	12.628	14.103	17.301
Variance	1.0803	0.9934	0.7447	0.8664	1.1213

From Table 2, the means of the transformed test are close to the theoretical mean for the approximation which is $\sqrt{2n-1}$. Furthermore, the results of the variances were slightly stable and close to the theoretical variance for the approximation which is 1.

Table 3 Empirical Type I error rate for Pearson statistic and the transformed statistic

Sample size	20	40	80	100	150
Standardized Pearson sum of squares	0.006	0.032	0.019	0.038	0.034
Transformed Standardized Pearson sum of squares	0.019	0.048	0.039	0.055	0.060

From the results of the above table, the empirical Type I error rates for the Pearson statistic are not close the theoretical type I error rate ($\alpha = 0.05$) due to the instable variances. However, the transformed statistic can improve the results, and it gave Type I error rates slightly close the theoretical type I error rate.

Conclusion

From this study it is concluded that using the Pearson statistic for such models is not recommended. The problem is due to the unstable variance and this statistic is unreliable. However, the transformed Pearson statistic had slight improvement on the results of the moments and Type I error rates. Furthermore, a future study using smoothing techniques on the transformed statistic may be helpful and could improve the performance of this statistic.

References

[1] Agresti, A. (2002). Categorical data analysis (2nd ed.). New York: John Wiley and Sons.

[2] McCullagh, P., & Nelder, J. (1989). Generalized linear models (2nd ed.). New York: Chapman Hall.

[3] Hosmer, D. W., & Lemeshow, S. (1980). A goodness-of-fit test for the multiple logistic regression model. Communications in Statistics, A10, 1043-1069.

[4] Evans, S. R., & Hosmer, D. W. (2004). Goodness of fit tests for mixed effects logistic models characterized by clustering. Communications in Statistics: Theory and Methods, 33(5), 1139–1155.

[5] Hosmer, D., Hosmer, T., le Cessie, S., & Lemeshow, S. (1997). A comparison of goodness-of-fit tests for the logistic regression model. Statistics in Medicine, 16, 965–980.

[6] Pan, Z., & Lin, D. Y. (2005). Goodness-of-fit methods for generalized linear mixed models. Biometrics, 61, 1000-1009.

[7] le Cessie, S., & van Houwelingen, J. (1991). A goodness-of-fit test for binary regression models based on smoothing methods. Biometrics, 47, 1267–1282.

[8] Ailobhio, D. T., & Ikughur, J. A. (2024). A review of some goodness-of-fit tests for logistic regression model. Asian Journal of Probability and Statistics, 26(7), 75-85.

[9] Ma, J., Dhiman, P., Qi, C., Bullock, G., van Smeden, M., Riley, R. D., & Collins, G. S. (2023). Poor handling of continuous predictors in clinical prediction models using logistic regression: a systematic review. Journal of Clinical Epidemiology, 161, 140-151.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of **JSHD** and/or the editor(s). **JSHD** and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.