

A New Method for Logistic Model Assessment

Di Shu¹ & Wenqing He²

¹ Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Ontario, Canada

² Department of Statistical and Actuarial Sciences, University of Western Ontario, London, Ontario, Canada

Correspondence: Wenqing He, Department of Statistical and Actuarial Sciences, University of Western Ontario, London, ON., N6A 5B7, Canada. E-mail: whe@stats.uwo.ca

Received: August 1, 2017 Accepted: August 15, 2017 Online Published: October 12, 2017

doi:10.5539/ijsp.v6n6p120 URL: <https://doi.org/10.5539/ijsp.v6n6p120>

Abstract

It is well known that the logistic model plays an important role for the analysis of binary outcomes. Most of the existing methods for the assessment of logistic models are constructed based on the distance between the observed and the predicted outcomes. We consider a new method from a different perspective by assessing the distance between two consistent estimators developed under the same logistic model form. The proposed tests are easy to implement and are applicable to both prospective and case-control studies.

Keywords: estimating equations, logistic models, model assessment

1. Introduction

The logistic regression model has become one of the most widely used statistical tools in the analysis of binary outcome data arising from either prospective studies or case-control studies (e.g., Prentice & Pyke, 1979; Breslow & Day, 1980; Hosmer, Lemeshow, & Sturdivant, 2013). Let X be a $p \times 1$ vector of covariates and Y be a binary outcome taking value 1 or 0, where p is a positive integer representing the dimension of the covariates. The logistic model specifies the following relationship between Y and X :

$$\log \left\{ \frac{P(Y = 1|X)}{P(Y = 0|X)} \right\} = \beta_0 + \beta^T X, \quad (1)$$

where β_0 is a scalar parameter and β is a $p \times 1$ vector of parameters. The interpretation of β is the vector of log odds ratios of outcome Y associated with one unit increase in the vector of covariates X . The parameter β_0 can be estimated for prospective studies, but is not estimable in case-control studies (Prentice & Pyke, 1979; Yi, 2017, Ch.7). The inference on β enables investigators to quantify the effects of covariates X on the outcome Y .

The estimation results and interpretations obtained from model (1) can be misleading when model (1) does not reflect the true relationship between X and Y . For example, if the working model (1) fails to include important risk factors for the outcome, or the true link function connecting X and Y is not logit, fitting (1) to the data often leads to invalid inferential results. Therefore, it is essential to assess the appropriateness of the working model (1) before making conclusion based on the model fitting.

There has been extensive research on developing reliable tests for the assessment of logistic models; see Ch.5 of Hosmer et al. (2013) for a comprehensive review. It is common to assess a working model by examining whether the model fits the data well. Most existing tests assess the model by examining summary measures of the distance between the observed and the predicted outcomes. For example, Hosmer and Lemeshow (1980) proposed the so-called Hosmer-Lemeshow goodness-of-fit test based on grouping estimated probabilities. Tsiatis (1980) discussed a score test using the partition of covariates. Osius and Rojek (1992) considered a large sample normal approximation to the Pearson chi-square statistic. Su and Wei (1991) developed a test using cumulative sums of residuals. le Cessie and van Houwelingen (1991, 1995) constructed tests using smoothed residuals. Stukel (1988) proposed a generalized logistic model by introducing additional parameters for an expanded class of models. A test was constructed based on the hypothesis that the additional parameters are equal to zero to shrink the class of models to the desired logistic model. In the context of case-control studies, Qin and Zhang (1997) pointed out that model (1) is equivalent to a two-sample semiparametric model and proposed a goodness-of-fit test by comparing the observed distribution of covariates and the expected counterpart under the assumed model.

In this paper, we propose a new method for the assessment of logistic models. The idea stems from a perspective different from existing methods. We construct the tests by comparing the maximum likelihood estimator for the parameter in (1) and a consistent estimator obtained from a weighted estimating equation, where the weight is a function of covariates and parameters. The intuition behind this approach is as follows. Since the two estimators are both consistent under the same assumption of model (1), a significant distance between them implies that the assumed model (1) may not hold. Our

idea is relevant to Shih (1998) who developed a goodness-of-fit test for the Clayton copula model based on the difference between the unweighted and weighted estimators for the association parameter with the weight depending on the observed data only. In comparison, the weight in our paper also depends on the parameter and is attached to the estimating equation rather than the estimator.

Many existing methods are based on grouping of covariates, or estimated probabilities, which may be unideal in some cases. For instance, it has been documented that test statistics based on grouping can be sensitive to the choice of groups, i.e., test results obtained from different grouping plans can be inconsistent (e.g., D. Hosmer, T. Hosmer, le Cessie, & Lemeshow, 1997). To overcome this shortcoming, we propose tests that are free of grouping, conceptually straightforward and easy to implement. Our methods are applicable to handle both prospective data and case-control data. In our simulations, the proposed tests present similar performance to the widely used Hosmer-Lemeshow (HL) test. In some scenarios, the proposed tests achieve a higher power than the HL test.

The rest of this paper is organized as follows. Section 2 proposes the new assessment method for the logistic regression model, presents theoretical results and describes the test procedures. Section 3 performs simulation studies to assess the performance of the proposed tests in comparison with the HL test. Section 4 illustrates the proposed tests using two examples. Section 5 closes the paper with a discussion.

2. The Proposed Tests

Shih (1998) constructed a simple test to assess the goodness-of-fit for the Clayton copula model by examining the difference between the unweighted and weighted estimators for the association parameter. This test is justified by the fact that if the Clayton copula model holds, then the resulting two estimators should be consistent, thus yielding that the difference converges to zero as sample size goes to infinity. Adapting this scheme, we develop new tests for the assessment of the logistic model.

2.1 Unweighted and Weighted Estimating Equations

Suppose we observe a sample of n subjects. Let X_i and Y_i denote the observed covariates and outcome, respectively, for subject i , where Y_i is a binary outcome taking value 1 or 0. Suppose the relationship between X_i and Y_i is characterized by model (1). The maximum likelihood estimator $(\hat{\beta}_0, \hat{\beta}^T)^T$ of $(\beta_0, \beta^T)^T$ is obtained by solving the likelihood score equations

$$\sum_{i=1}^n \left[\left\{ Y_i - \frac{1}{1 + \exp(-\beta_0 - \beta^T X_i)} \right\} \begin{pmatrix} 1 \\ X_i \end{pmatrix} \right] = \mathbf{0} \tag{2}$$

for $(\beta_0, \beta^T)^T$. It is known that the maximum likelihood estimator $(\hat{\beta}_0, \hat{\beta}^T)^T$ is a consistent, asymptotically normal and efficient estimator for $(\beta_0, \beta^T)^T$ when model (1) is true and certain regularity conditions (e.g., Lehmann & Casella, 1998, Ch.6) hold.

Huang and Wang (1999, 2001) considered the weighted version of estimating equations (2):

$$\sum_{i=1}^n \left[w(\beta_0 + \beta^T X_i) \left\{ Y_i - \frac{1}{1 + \exp(-\beta_0 - \beta^T X_i)} \right\} \begin{pmatrix} 1 \\ X_i \end{pmatrix} \right] = \mathbf{0}, \tag{3}$$

where $w(\beta_0 + \beta^T X_i)$ is a weight function which depends on the covariates and the model parameter. Given a positive weight function $w(\cdot)$, a consistent estimator of $(\beta_0, \beta^T)^T$ can be obtained by solving (3) for $(\beta_0, \beta^T)^T$ by the fact that the expectation of the left hand side of (3) is zero (e.g., Yi, 2017, Ch.2). In particular, Huang and Wang (2001) chose weight functions $w(\beta_0 + \beta^T X_i) = 1 + \exp(-\beta_0 - \beta^T X_i)$ and $w(\beta_0 + \beta^T X_i) = 1 + \exp(\beta_0 + \beta^T X_i)$, and respectively, obtained the following estimating equations:

$$\sum_{i=1}^n \left[\{1 + \exp(-\beta_0 - \beta^T X_i)\} \left\{ Y_i - \frac{1}{1 + \exp(-\beta_0 - \beta^T X_i)} \right\} \begin{pmatrix} 1 \\ X_i \end{pmatrix} \right] = \mathbf{0}, \tag{4}$$

and

$$\sum_{i=1}^n \left[\{1 + \exp(\beta_0 + \beta^T X_i)\} \left\{ Y_i - \frac{1}{1 + \exp(-\beta_0 - \beta^T X_i)} \right\} \begin{pmatrix} 1 \\ X_i \end{pmatrix} \right] = \mathbf{0}. \tag{5}$$

Let $(\hat{\beta}_{0-}, \hat{\beta}_-^T)^T$ be the resulting estimator of $(\beta_0, \beta^T)^T$ obtained by solving (4) for $(\beta_0, \beta^T)^T$, and $(\hat{\beta}_{0+}, \hat{\beta}_+^T)^T$ be the resulting estimator of $(\beta_0, \beta^T)^T$ obtained by solving (5) for $(\beta_0, \beta^T)^T$. Under the assumed logistic model (1) and regularity conditions C1 and C2 of Huang and Wang (2001), both $(\hat{\beta}_{0-}, \hat{\beta}_-^T)^T$ and $(\hat{\beta}_{0+}, \hat{\beta}_+^T)^T$ are consistent and asymptotically normal estimators of $(\beta_0, \beta^T)^T$ (Huang & Wang, 2001).

2.2 The Proposed Test Statistics

We comment that the consistency of $(\hat{\beta}_0, \hat{\beta}^T)^T$, $(\hat{\beta}_{0-}, \hat{\beta}_-^T)^T$ and $(\hat{\beta}_{0+}, \hat{\beta}_+^T)^T$ is basically ensured by the unbiasedness of the corresponding estimating functions in (2), (4) and (5), i.e., these estimating functions have a zero expectation (e.g., Yi, 2017, Ch.2). The zero mean property of the estimating functions in (2) and (3) is guaranteed because the conditional mean of Y_i given X_i is correctly specified as

$$E(Y_i|X_i) = \frac{1}{1 + \exp(-\beta_0 - \beta^T X_i)},$$

which is described by model (1). The involvement of the terms $(1, X_i)^T$ and $w(\beta_0 + \beta^T X_i)$ in (2) and (3) do not alter the zero mean property of the estimating functions in (2) and (3).

The consistency of three estimators $(\hat{\beta}_0, \hat{\beta}^T)^T$, $(\hat{\beta}_{0-}, \hat{\beta}_-^T)^T$ and $(\hat{\beta}_{0+}, \hat{\beta}_+^T)^T$ requires that model (1) holds. When the logistic model (1) does not reflect the truth, we may expect to observe significant difference between the maximum likelihood estimator $(\hat{\beta}_0, \hat{\beta}^T)^T$ and the weighted estimating equation based estimator $(\hat{\beta}_{0-}, \hat{\beta}_-^T)^T$, or significant difference between $(\hat{\beta}_0, \hat{\beta}^T)^T$ and $(\hat{\beta}_{0+}, \hat{\beta}_+^T)^T$. In practice, the estimation of parameter β is often of interest since β provides the odds ratio interpretations. Moreover, for case-control studies, the scalar parameter β_0 , which is related to $P(Y = 1)$ for the population of interest, cannot be estimated. Prentice and Pyke (1979) showed that valid odds ratio estimators can be obtained by fitting logistic model (1) to the case-control data as if the data were collected from a prospective study. Therefore, here we focus on the log odds ratio parameters β and construct two tests based on the distance between $\hat{\beta}$ and $\hat{\beta}_-$, and the distance between $\hat{\beta}$ and $\hat{\beta}_+$, respectively.

Under (1) and suitable regularity conditions, $\hat{\beta}$, $\hat{\beta}_-$ and $\hat{\beta}_+$ are consistent and asymptotically normal estimators of β . Thus, $\sqrt{n}(\hat{\beta} - \beta_-) \xrightarrow{d} N(\mathbf{0}, V_-)$ and $\sqrt{n}(\hat{\beta} - \beta_+) \xrightarrow{d} N(\mathbf{0}, V_+)$ as $n \rightarrow \infty$, where V_- and V_+ are positive definite covariance matrices. The analytic forms of V_- and V_+ can be obtained based on the sandwich type formula (Stefanski & Boos, 2002), and the estimates of V_- and V_+ are available based on the empirical estimates of the corresponding sandwich terms.

The following Theorem reports the two proposed tests.

Theorem: *Suppose the logistic model (1) and suitable regularity conditions hold.*

(a). *The statistic*

$$(\hat{\beta} - \hat{\beta}_-)^T \hat{\Sigma}_-^{-1} (\hat{\beta} - \hat{\beta}_-) \xrightarrow{d} \chi_p^2 \quad \text{as } n \rightarrow \infty,$$

where $\hat{\Sigma}_-$ is a consistent estimator of the covariance matrix $\text{var}(\hat{\beta} - \hat{\beta}_-)$.

(b). *The statistic*

$$(\hat{\beta} - \hat{\beta}_+)^T \hat{\Sigma}_+^{-1} (\hat{\beta} - \hat{\beta}_+) \xrightarrow{d} \chi_p^2 \quad \text{as } n \rightarrow \infty,$$

where $\hat{\Sigma}_+$ is a consistent estimator of the covariance matrix $\text{var}(\hat{\beta} - \hat{\beta}_+)$.

Proof. Since V_- and V_+ are positive definite, they are invertible with rank p . Let V_-^{-1} be the inverse matrix of V_- and let V_+^{-1} be the inverse matrix of V_+ . By the theory of quadratic forms (e.g., Rao, 1973, Ch.3),

$$\{\sqrt{n}(\hat{\beta} - \hat{\beta}_-)^T\} V_-^{-1} \{\sqrt{n}(\hat{\beta} - \hat{\beta}_-)\} \xrightarrow{d} \chi_p^2 \quad \text{as } n \rightarrow \infty$$

and

$$\{\sqrt{n}(\hat{\beta} - \hat{\beta}_+)^T\} V_+^{-1} \{\sqrt{n}(\hat{\beta} - \hat{\beta}_+)\} \xrightarrow{d} \chi_p^2 \quad \text{as } n \rightarrow \infty.$$

Since $n\hat{\Sigma}_-$ is a consistent estimator of V_- and $n\hat{\Sigma}_+$ is a consistent estimator of V_+ , the results in the theorem are immediate.

Remark 1: When the singularity problem of $\hat{\Sigma}_-$ or $\hat{\Sigma}_+$ occurs, we describe two strategies. The first strategy is to simply consider subvectors of $\hat{\beta} - \hat{\beta}_-$ and $\hat{\beta} - \hat{\beta}_+$ such that the resulting covariance matrices are nonsingular. Then the degrees of freedom in (a) and (b) should be modified to be the length of the subvectors. The second strategy is to use the generalized inverse matrix (Rao, 1973, Ch.1) to get around the singularity. The resulting test is still valid, with the degrees of freedom in (a) and (b) replaced by the rank of V_- and V_+ , respectively. The use of generalized inverse matrix for handling singularity was considered by Lin and Wei (1991) in the context of information matrix tests for Cox regression models.

Remark 2: Since the derivation of the analytic forms of the estimators $\hat{\Sigma}_-$ and $\hat{\Sigma}_+$ needs the second derivatives of the likelihood, and the derivatives of the estimating equations, the implementation is complicated. In real applications, we use nonparametric bootstrap method (Efron, 1982) to obtain these two covariance matrices $\hat{\Sigma}_-$ and $\hat{\Sigma}_+$ for convenience.

2.3 Test Procedures

The theorem proposes two tests for assessing the model form (1). We now further elaborate on the detail of the test procedure by using the test statistic in Theorem (a); the test based on Theorem (b) can be conducted in the same manner.

Step 1: Obtain $\hat{\beta}$ and $\hat{\beta}_-$ by solving estimating equations (2) and (4), respectively. In particular, $\hat{\beta}$ can be easily obtained using available software packages for logistic models.

Step 2: Obtain $\hat{\Sigma}_-$ using the nonparametric bootstrap method (Efron, 1982). Specifically, we generate B bootstrap samples by resampling the original data at the individual level with replacement, where B is a user-specified positive integer. For $b = 1, \dots, B$, let $\hat{\beta}(b)$ and $\hat{\beta}_-(b)$ denote the resulting estimators obtained from Step 1 using the b th bootstrap sample. Then $\hat{\Sigma}_-$ is given by the sample covariance matrix of $\{\hat{\beta}(b) - \hat{\beta}_-(b) : b = 1, \dots, B\}$. Specifically, let $\hat{\delta}(b) = \hat{\beta}(b) - \hat{\beta}_-(b)$. Then

$$\hat{\Sigma}_- = \frac{1}{B-1} \sum_{b=1}^B \left\{ \hat{\delta}(b) - \frac{1}{B} \sum_{b=1}^B \hat{\delta}(b) \right\} \left\{ \hat{\delta}(b) - \frac{1}{B} \sum_{b=1}^B \hat{\delta}(b) \right\}^T.$$

Step 3: Calculate the observed value for the test statistic $T_1 = (\hat{\beta} - \hat{\beta}_-)^T \hat{\Sigma}_-^{-1} (\hat{\beta} - \hat{\beta}_-)$. The observed p -value is obtained as $P(\chi_p^2 > T_1)$, where χ_p^2 represents a random variable following a χ^2 distribution with degree of freedom p . For a given significance level α , if the observed p -value is less than α , the test suggests that logistic model (1) may not be appropriate, and the null hypothesis that the model (1) holds is then in doubt.

3. Simulation Studies

We conduct simulation studies to assess the finite sample performance of the proposed tests. The widely used HL test (Hosmer & Lemeshow, 1980) is included in the simulation studies as the benchmark. The HL test is grouping-based method and the number of groups is user-specified. We adopt the commonly used strategy and take 10 groups in our simulations. We use 1000 bootstrap replicates for the proposed tests to estimate the covariance matrices. The sample size is set to be 500 or 1000. For each scenario, 1000 replications are run. For the i th subject, let covariates $X_i = (X_{1i}, X_{2i})^T$ where X_{1i} is drawn from a standard normal distribution, and X_{2i} is generated from a uniform distribution ranging from 0 to 1. The outcome is generated from the model:

$$\text{logit } P(Y_i = 1|X_i) = -0.1 + X_{1i} + X_{2i} + \gamma(X_{1i} + X_{2i})^2, \tag{6}$$

where γ is a parameter for the extra quadratic and interaction terms. The working model we use to fit the simulated data is from (1),

$$\text{logit } P(Y_i = 1|X_i) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i},$$

where no extra quadratic and interaction terms are included.

3.1 Assessment of Test Level

When $\gamma = 0$ in (6), the logistic model form (1) correctly reflects the truth. For this case, we examine the estimated significance levels (i.e., the probability of rejecting model (1) when model (1) is really true) of the proposed tests and the HL test, and report the test results in Table 1 for the significance level $\alpha = 0.01, 0.05$ and 0.1 . The estimated significance levels for all three tests are reasonably close to the nominal significance levels.

Table 1. Estimated significance levels for the Hosmer-Lemeshow test and proposed tests at the significance level $\alpha = 0.01, 0.05$, and 0.1 .

Method	$n = 500$			$n = 1000$		
	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$
Hosmer-Lemeshow	0.010	0.042	0.087	0.009	0.051	0.098
Proposed (1st weight)	0.004	0.038	0.080	0.010	0.043	0.099
Proposed (2nd weight)	0.007	0.034	0.084	0.011	0.059	0.116

3.2 Assessment of Test Power

When $\gamma \neq 0$ in (6), model (1) does not reflect the truth. By setting $\gamma = -0.2$ or -0.5 , we examine the estimated powers (i.e., the probability of rejecting model (1) when model (1) is really not true) of the proposed tests and the HL test at

the significance level $\alpha = 0.05$. Table 2 summarizes the empirical results. The estimated powers are significantly larger with $\gamma = -0.5$ than those with $\gamma = -0.2$, because of the farther departure from the null model. The first proposed test performs similar to the HL test. The second proposed test achieves a significantly higher power than the other two tests when $\gamma = -0.2$. As anticipated, the achieved power gets larger as the sample size increases.

Table 2. Estimated powers for the Hosmer-Lemeshow test and proposed tests at the significance level $\alpha = 0.05$ when the working model lacks quadratic and interaction terms.

Method	$n = 500$		$n = 1000$	
	$\gamma = -0.2$	$\gamma = -0.5$	$\gamma = -0.2$	$\gamma = -0.5$
Hosmer-Lemeshow	0.269	0.967	0.532	1.000
Proposed (1st weight)	0.270	0.801	0.643	1.000
Proposed (2nd weight)	0.508	0.921	0.811	1.000

Now we assess the power of the three methods when the true link function is not logit. To simulate data, we modify (6) by specifying $\gamma = 0$ and replacing the logit link by one of the two link functions: probit and complementary log-log (Clog-log). We still use model (1) to fit the simulated data. Table 3 summarizes the estimated powers of the three tests at the significance level $\alpha = 0.05$. Compared with the complementary log-log link, testing the probit link leads to lower powers as expected, since the probit model is close to the logistic model, and the difference is hard to determine (e.g., Dobson, 2002, Section 7.3). When $n = 500$, the first proposed test and the HL test show similar performance, with higher achieved powers than the second proposed test. Interestingly, when $n = 1000$, the second proposed test shows the highest achieved powers. The estimated powers increase as the sample size increases.

Table 3. Estimated powers for the Hosmer-Lemeshow test and proposed tests at the significance level $\alpha = 0.05$ when the working model misspecifies the link function.

Method	$n = 500$		$n = 1000$	
	probit	Clog-log	probit	Clog-log
Hosmer-Lemeshow	0.041	0.118	0.052	0.324
Proposed (1st weight)	0.032	0.139	0.124	0.276
Proposed (2nd weight)	0.024	0.029	0.199	0.341

4. Examples

We use two examples to illustrate the proposed tests. The HL test is conducted as the benchmark, where we adopt the commonly used 10 groups. One thousand bootstrap replicates are used for the proposed tests.

Example 1. Hosmer et al. (2013) considered a dataset of 100 subjects with age regarded as the covariate and the presence or absence of coronary heart disease taken as the binary outcome. It is of interest to assess model (1) which is used to feature the relationship between the status of coronary heart disease and age. We apply the two proposed tests and the HL test to assess model (1). The observed test statistic for the first proposed test is 0.010 with 1 degree of freedom, and the observed p -value equals 0.920. The observed test statistic for the second proposed test is 0.002 with 1 degree of freedom, and the observed p -value equals to 0.969. The HL test gives the observed test statistic 2.224 with 8 degrees of freedom, and the observed p -value equals 0.973. The three tests produce fairly similar results of p -value and they all suggest no evidence against the use of logistic model (1).

Example 2. Hosmer et al. (2013) discussed a low birth weight study. The dataset includes 189 births to women observed in the obstetrics clinic. The binary outcome is the indicator of low birth weight (< 2500 grams). Consider weight of mother at last menstrual period, age of mother and smoking status during pregnancy as covariates. Our goal is to assess the appropriateness of model (1). The observed test statistic for the first proposed test is 1.314 with 3 degrees of freedom, yielding the observed p -value 0.726. The observed test statistic for the second proposed test is 1.913 with 3 degrees of freedom, giving the observed p -value 0.591. The HL test gives the observed test statistic 7.225 with 8 degrees of freedom,

and the observed p -value equals 0.513. The three tests produce similar results of p -value and they all suggest no evidence against the use logistic model (1).

5. Discussion

In this paper, we construct new tests for the assessment of the logistic model from a different perspective than existing tests. We construct the tests by examining the distance between two consistent estimators developed under the logistic model (1), rather than the distance between the observed outcome and the predicted outcome as considered by most existing methods. The asymptotic results for the proposed tests are established. Our tests are straightforward, easy to implement and applicable for both prospective studies and case-control studies. Test results and odds ratio estimates are given simultaneously. Unlike grouping-based tests which may be sensitive to the choice of groups, the proposed tests are free of grouping. In our limited simulation studies, the proposed tests show similar results to the widely-used HL test. In some scenarios, the proposed tests have higher powers than the HL test. When the sample size is relatively large ($n = 1000$), the second proposed test produces higher powers than the first proposed test.

Defined as the probability of being assigned to receive the treatment/intervention given pre-treatment covariates, the propensity score has been increasingly used to conduct causal inference (Rosenbaum & Rubin, 1983). Logistic regression models are often employed as the propensity score model. It has been documented that misspecifying the propensity score model can lead to severely biased estimated causal effects (e.g. Kang & Schafer, 2007). Our proposed tests may be useful to evaluate the propensity score model in causal inference.

Acknowledgements

The authors would like to thank the editor and two anonymous reviewers for their constructive comments. This research was supported by a grant from the Natural Sciences and Engineering Research Council (NSERC) of Canada.

References

- Breslow, N., & Day, N. E. (1980). *Statistical methods in cancer research, 1, The analysis of case-control studies*, Lyon: IARC.
- Dobson, A. J. (2002). *An introduction to generalized linear models*, 2nd Edition. Chapman & Hall/CRC.
<https://doi.org/10.1201/9781420057683>
- Efron, B. (1982). The jackknife, the bootstrap and other resampling plans. *Philadelphia: Society for industrial and applied mathematics*, 38. <https://doi.org/10.1137/1.9781611970319>
- Hosmer, D. W., & Lemeshow, S. (1980). Goodness-of-fit tests for the multiple logistic regression model. *Communications in Statistics, A10*, 1043-1069. <https://doi.org/10.1080/03610928008827941>
- Hosmer, D. W., 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.
[https://doi.org/10.1002/\(SICI\)1097-0258\(19970515\)16:9<965::AID-SIM509>3.0.CO;2-O](https://doi.org/10.1002/(SICI)1097-0258(19970515)16:9<965::AID-SIM509>3.0.CO;2-O)
- Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression*, 3rd Edition. John Wiley & Sons.
<https://doi.org/10.1002/9781118548387>
- Huang, Y., & Wang, C. Y. (1999). Nonparametric correction to errors in covariates. Technical Report, Fred Hutchinson Cancer Research Center, Seattle.
- Huang, Y., & Wang, C. Y. (2001). Consistent functional methods for logistic regression with errors in covariates. *Journal of the American Statistical Association*, 96, 1469-1482. <https://doi.org/10.1198/016214501753382372>
- Kang, J. D., & Schafer, J. L. (2007). Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data (with discussions). *Statistical Science*, 22, 523-539.
<https://doi.org/10.1214/07-sts227>
- le Cessie, S., & van Houwelingen, J. C. (1991). A goodness-of-fit test for binary regression models, based on smoothing residuals. *Biometrics*, 47, 1267-1282. <https://doi.org/10.2307/2532385>
- le Cessie, S., & van Houwelingen, J. C. (1995). Testing the fit of a regression model via score tests in random effects models. *Biometrics*, 51, 600-614. <https://doi.org/10.2307/2532948>
- Lehmann, E. L., & Casella, G. (1998). *Theory of point estimation*, 2nd Edition. Springer. <https://doi.org/10.1007/b98854>
- Lin, D. Y., & Wei, L. J. (1991). Goodness-of-fit tests for the general Cox regression model. *Statistica Sinica*, 1, 1-17.
- Osious, G., & Rojek, D. (1992). Normal goodness-of-fit tests for multinomial models with large degrees of freedom.

- Journal of the American Statistical Association*, 87, 1145-1152. <https://doi.org/10.2307/2290653>
- Prentice, R. L., & Pyke, R. (1979). Logistic disease incidence models and case-control studies. *Biometrika*, 66, 403-411. <https://doi.org/10.1093/biomet/66.3.403>
- Qin, J., & Zhang, B. (1997). A goodness-of-fit test for logistic regression models based on case-control data. *Biometrika*, 84, 609-618. <https://doi.org/10.1093/biomet/84.3.609>
- Rao, C. R. (1973). *Linear statistical inference and its applications*, 2nd Edition. John Wiley & Sons, Inc. <https://doi.org/10.1002/9780470316436>
- Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70, 41-55. <https://doi.org/10.2307/2335942>
- Shih, J. H. (1998). A goodness-of-fit test for association in a bivariate survival model. *Biometrika*, 85, 189-200. <https://doi.org/10.1093/biomet/85.1.189>
- Stefanski, L. A., & Boos, D. D. (2002). The calculus of M-estimation. *The American Statistician*, 56, 29-38. <https://doi.org/10.1198/000313002753631330>
- Stukel, T. A. (1988). Generalized logistic models. *Journal of the American Statistical Association*, 83, 426-431. <https://doi.org/10.2307/2288858>
- Su, J. Q., & Wei, L. J. (1991). A lack-of-fit test for the mean function in a generalized linear model. *Journal of the American Statistical Association*, 86, 420-426. <https://doi.org/10.2307/2290587>
- Tsiatis, A. A. (1980). A note on a goodness-of-fit test for the logistic regression model. *Biometrika*, 67, 250-251. <https://doi.org/10.1093/biomet/67.1.250>
- Yi, G. Y. (2017). *Statistical analysis with measurement error or misclassification: Strategy, method and application*. Springer. <https://doi.org/10.1007/978-1-4939-6640-0>

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).