Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2008, Article ID 186372, 12 pages
doi:10.1155/2008/186372

Research Article

A Mathematical Tool for Inference in Logistic Regression with Small-Sized Data Sets: A Practical Application on ISW-Ridge Relationships

Tsung-Hao Chen,[1] Chen-Yuan Chen,[2]
Hsien-Chueh Peter Yang,[3] and Cheng-Wu Chen[4]

[1] *Department of Business Administration, Shu-Te University, Yen Chau, Kaohsiung, Taiwan 82445, Taiwan*
[2] *Department of Management Information System, Yung-Ta Institute of Technology and Commerce, Pingtung, Taiwan 90941, Taiwan*
[3] *Department of Risk Management and Insurance, Kaohsiung First University of Science and Technology, Kaohsiung, Taiwan 811, Taiwan*
[4] *Department of Logistics Management, Shu-Te University, Yen Chau, Kaohsiung, Taiwan 82445, Taiwan*

Correspondence should be addressed to Chen-Yuan Chen, chency@mail.ytit.edu.tw

Received 6 October 2007; Revised 3 March 2008; Accepted 26 August 2008

Recommended by Irina Trendafilova

The general approach to modeling binary data for the purpose of estimating the propagation of an internal solitary wave (ISW) is based on the maximum likelihood estimate (MLE) method. In cases where the number of observations in the data is small, any inferences made based on the asymptotic distribution of changes in the deviance may be unreliable for binary data (the model's lack of fit is described in terms of a quantity known as the deviance). The deviance for the binary data is given by D. Collett (2003). may be unreliable for binary data. Logistic regression shows that the P-values for the likelihood ratio test and the score test are both <0.05. However, the null hypothesis is not rejected in the Wald test. The seeming discrepancies in P-values obtained between the Wald test and the other two tests are a sign that the large-sample approximation is not stable. We find that the parameters and the odds ratio estimates obtained via conditional exact logistic regression are different from those obtained via unconditional asymptotic logistic regression. Using exact results is a good idea when the sample size is small and the approximate P-values are <0.10. Thus in this study exact analysis is more appropriate.

1. Introduction

Internal waves refer to the motion at the interface between layers of water of different densities in a stratified water body, such as the ocean. The simplest oceanic density structure, where differences in water density are mostly caused by differences in water temperature or salinity, can be approximated by a two-layer model. Oceanic internal waves typically have wavelengths ranging from hundreds of meters to tens of kilometers, with periods from tens of minutes to tens of hours. In the Andaman and Sulu Sea they can have amplitudes (peak to trough distance) exceeding 50m and in the South China Sea the amplitude can exceed 110 m [1–9]. The mixing and dissipation generated by internal waves have important effects on the cross slope exchange processes, enhancement of bottom stress, and generation of the nepheloid layers. It has recently been proposed that internal waves may make a significant contribution to internal oceanic mixing and hence have an important influence on climatic change. This is why it is necessary to scrutinize the interaction of nonlinear internal solitary waves (ISWs) with the seabed topography [10–17].

Several studies, including both simulations and laboratory experiments, aiming at exploring the mechanisms for the generation, propagation, and evolution of ISWs, have already been carried out. However, since energy dissipation plays such an important and varied role on water and sedimentary movement in coastal seas [18], we need a better fitting and more appropriate model for predicting ISW propagation. A preliminary approach has recently been made in which the effects of weighted parameters on the amplitude and reflection of energy-based ISWs from uniform slopes in a two layered fluid system were investigated [19]. The results are quite consistent with other experimental results, and are applicable to the naturally occurring reflection of ISWs from sloping bottoms. More recently, Chen et al. [20] concluded the goodness-of-fit and predictive ability of the cumulative logistic regression models to be better than that of the binary logistic regression models. However, in cases where the data are so small that there are some observations with proportions close to zero or one, inferences based on the asymptotic distribution of the change in deviance may be unreliable. In point of fact, reports on statistical manipulations related to this theme are rather rare.

The rest of the paper is organized as follows. In Section 2 we describe the experimental set-up and theoretical background needed to understand the hydrodynamic interaction. We also discuss the analysis of the logistic regression model, and introduce the exact conditional logistic model and the hypothesis on which the parameters are based. Section 3 is devoted to a comparison of the conditional exact logistic regression model and the unconditional asymptotic logistic regression model. Finally, some conclusions are made. It is noted that small sample size means that there are some observations with proportions close to zero or one and P-values of less than 0.10, which is an indication that an exact analysis would be more appropriate.

2. Research framework

Experiments were carried out in the laboratory using a two-layer fluid system of fresh and briny water in a 12 m long wave flume (rectangular in cross-section). The upper layer of water in the wave flume consisted of fresh water with a density ρ_1 and a depth H_1, while the lower layer was comprised of brine with a density ρ_2 and a depth H_2. The leading ISW was generated by the lifting of a pneumatic sluice gate at one end of the flume. The wave propagated into the main section of the flume to the left-hand side (LHS) of the gate. The amplitude a and characteristic length L_w of the ISW were predetermined by arranging the step length L and step depth η_0 (see Figure 1). Six ultrasonic probes connected to an amplifier unit and A/D converter, then to a personal computer, gathered and processed digital signals as the ISW propagated along the flume. As the ISW propagated from the RHS (right-hand side) to the LHS of the flume, the first ultrasonic probe (P1) recorded the properties of the incident ISW, the wave amplitude and characteristic length, while the second probe (P2) collected reflected signals showing the wave-obstacle interaction. The methodology for

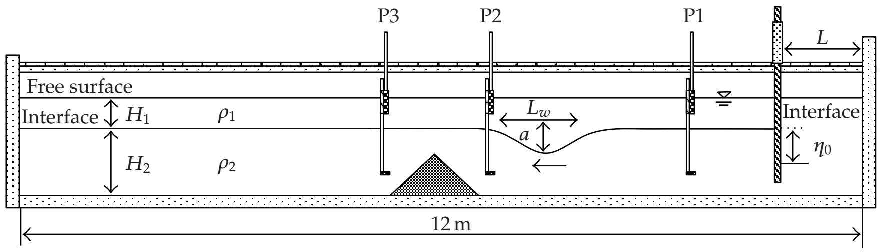

Figure 1: A schematic view showing the set up for ISW propagation in a two-layer fluid system over a single obstacle.

measuring the physical properties related to the propagation and dissipation of the ISW has been reported in detail by Chen et al. [21]. The amplitude-based transmission rate during the wave-ridge interaction was dependent on two factors, ridge height and potential energy.

2.1. Exact conditional logistic regression model

The theoretical basis for the exact conditional logistic regression model was originally laid down by Cox [22], but recent algorithmic advances in computing the exact distributions have made the methodology more practical. Since then Hirji et al. [23] have developed an efficient algorithm for generating the required conditional distributions. Cox and Snell [24] noted that it has been known since the 1970's how to extend the theory of Fisher's exact test to logistic regression models. The interested reader may refer to Mehta and Patel [25] for a useful summary of exact logistic regression. A complete discussion of the exact logistic regression methodology and more detailed applications can be found in a variety of sources [26–30].

Here, let $\pi_i(X)$ represent the probability of "success" for a binary response (Y) for the explanatory variables $(X) = (x_1, x_2, x_3, \ldots, x_k)$. The notation can be simplified by using $\pi_i(X) = E(Y_i \mid x)$ to represent the conditional mean of Y given x when a logistic distribution is utilized:

$$\begin{aligned} \pi_i = E(Y_i) = P &= \frac{1}{1+e^{-(\alpha+\beta_1 X_{i1}+\beta_2 X_{i2}+\cdots+\beta_k X_{ik})}} = \frac{1}{1+e^{-Z_i}}, \\ 1-\pi_i = 1-E(Y_i) = 1-P &= 1-\frac{1}{1+e^{-(\alpha+\beta_1 X_{i1}+\beta_2 X_{i2}+\cdots+\beta_k X_{ik})}} = 1-\frac{1}{1+e^{-Z_i}} = \frac{e^{-Z_i}}{1+e^{-Z_i}}, \end{aligned} \tag{2.1}$$

such that

$$\left(\frac{\pi_i}{1-\pi_i}\right) = e^{Z_i} = e^{(\alpha+\beta_1 X_{i1}+\beta_2 X_{i2}+\cdots+\beta_k X_{ik})}. \tag{2.2}$$

The transformation of $\pi(x_i)$, which is central to this study of logistic regression, is the logit transformation. This transformation is defined as

$$\operatorname{logit}(\pi(x_i)) = \ln\left(\frac{\pi(x_i)}{1-\pi(x_i)}\right) = x_i'\beta, \tag{2.3}$$

where $\beta = (\beta_1, \ldots, \beta_k)'$ is an unknown parameter vector.

The sufficient statistics for the β_j in the unconditional likelihood function are

$$T_j = \sum_{i=1}^{n} y_i x_{ij,\, j=1,\ldots,t+s}, \tag{2.4}$$

where y_i is the realization of Y_i.

If T_0 and T_1 indicate sufficient statistics corresponding to β_0 and β_1, the conditional probability density function of T_1 conditional on T_0 can be formulated as

$$f_{\beta_1}(t_1 \mid t_0) = \frac{C(t)\exp(t_1'\beta_1)}{\sum_u C(u, t_0)\exp(u'\beta_1)}, \tag{2.5}$$

where $C(u, t_0)$ indicates the number of vectors y, such that $y'X_1 = u$ and $y'X_0 = t_0$.

Conditional exact inference involves the generation of the conditional permutational distribution $f_{\beta_1}(t_1 \mid t_0)$ for the sufficient statistics for the parameters. The distribution $f_{\beta_1}(t_1 \mid t_0)$ is called the permutation conditional distribution or exact conditional distribution.

2.2. Testing the hypotheses

According to exact logistic regression (for both the exact score conditional test and the probability test) the parameters for the specified hypothesis are equal to zero. If an effect consists of two or more parameters, then it is hypothesized that all the parameters are simultaneously equal to zero [26, 27].

2.2.1. Exact score conditional test

The null hypothesis is

$$\begin{aligned} H_0 &: \beta_1 = 0, \\ H_1 &: \beta_1 \neq 0. \end{aligned} \tag{2.6}$$

The conditional mean μ_1 and variance matrix R_1 of T_1 (conditional on $T_0 = t_0$) are calculated via the exact conditional scores test. The score statistic is

$$s = (t_1 - \mu_1)' R_1^{-1} (t_1 - \mu_1). \tag{2.7}$$

Now compare this to the score for each member of the distribution

$$S = (T_1 - \mu_1)' R_1^{-1} (T_1 - \mu_1). \tag{2.8}$$

In the null hypothesis, an exact P-value, which is the probability of obtaining a more extreme statistic than the observed one, is assumed.

The result of the P-value is

$$p(t_1 \mid t_0) = \Pr(S \geq s) = \sum_{u \in \Omega} f_0(u \mid t_0), \tag{2.9}$$

where

$$\Omega_s = \{u : \text{there exists } y \text{ with } y'X_1 = u,\ y'X_0 = t_0,\ S(u) \geq s\}. \tag{2.10}$$

A mid P-value, adjusted for the discreteness of the distribution, is assumed for the null hypothesis.

The mid-p statistic is defined as

$$p(t_1 \mid t_0) - \frac{1}{2} f_0(u \mid t_0). \tag{2.11}$$

2.2.2. Probability testing

For small samples, the parameter inference process is carried out using conditional distribution probabilities, such as exact P-values, rather than a crude approximation [29]. For testing the null hypothesis we use

$$\begin{aligned} H_0 &: \beta_1 = 0 \\ H_1 &: \beta_1 \neq 0. \end{aligned} \tag{2.12}$$

Under the null hypothesis, the exact probability test statistic is just $f_{\beta_1=0}(t_1 \mid t_0)$; the corresponding P-value gives the probability of getting a less likely statistic

$$p(t_1 \mid t_0) = \sum_{u \in \Omega_p} f_0(u \mid t_0), \tag{2.13}$$

where

$$\Omega_p = \{u : \text{there exists } y \text{ with } y'X_1 = u,\ y'X_0 = t_0,\ f_0(u \mid t_0) \leq f_0(t_1 \mid t_0)\}. \tag{2.14}$$

3. Analytical results

The effects of the ridge height, the depth of the lower water layer, and the potential energy on the propagation of the ISW are all considered. The results from the laboratory experiments are shown in the data sets. The amplitudes of the incident and reflected waves are also included. The dependent variables for the binary logistic regression model are classified into two groups, weak and strong, based on the amplitude incident rate. When the hypothetical incident rate is >0.5 it is considered strong and when it is <0.5 it is considered weak. The frequencies for the strong and weak levels are 35 and 28, respectively.

3.1. Asymptotic logistic regression model

The methodologies utilized in the asymptotic logistic regression model and the diagnostics of the goodness-of-fit statistics are discussed below.

Table 1: Deviance and Pearson goodness-of-fit statistics.

Deviance and Pearson goodness-of-fit statistics				
Criterion	DF	Value	Value/DF	Pr > Chi-sq
Deviance	50	36.3380	0.7268	0.9260
Pearson	50	60.7842	1.2157	0.1412

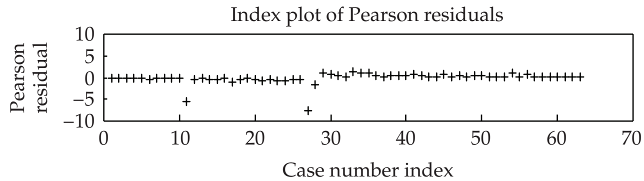

Figure 2: Plot of Pearson residual (Reschi) versus case number index.

3.1.1. Goodness-of-fit statistics

The Pearson Chi-squares test and deviance Chi-squares test are used. The results of the Pearson Chi-square test give a distribution with the degrees of freedom = $\{(r-1)(s-1)-t\}$, where t is the number of explanatory variables, r is the number of response levels, and s is the number of subpopulations.

The goodness-of-fit statistics are shown in Table 1. The dispersion parameter (value/DF), which indicates estimated deviance, is given in the value/DF column. The dispersion parameter is 0.7268 and the Pearson Chi-squares dispersion parameter is 1.2157. Ideally, this value should be very close to 1.00. The values of the Pearson Chi-square and deviance Chi-square statistics are 60.7842 and 36.338, respectively, with 50 degrees of freedom $((2-1)(54-1)-3=50)$. The Pearson Chi-squares value is slightly larger than the degrees of freedom; the P-values for the deviance and Pearson Chi-squares are all larger than 0.05 (0.9260, 0.1412). From this we see that although there is a little over dispersion, this model seems to have an acceptable fit with the data. The overdispersion means that the model still needs to be modified.

3.1.2. Regression diagnostics

There are a number of different ways to plot the regression diagnostics, each directed at a particular aspect of the fit. For examples see Hosmer and Lemeshow [28], and Landwehr et al. [31] who discussed graphical techniques for logistic regression diagnostics. Generally such techniques offer a visual rather than numerical representation that may be more intuitively appealing to some researchers. Index plots are useful for the identification of extreme values [32]. An examination of the index plots of the Pearson residuals (Figure 2) and the deviance residuals (Figure 3) for our data indicates that case 11 and case 27 are poorly accounted for by the model. It can be seen in the index plot of the diagonal elements of the hat matrix (Figure 4) that case 49 is at the extreme point in the design space.

3.1.3. Outliers and influential observations

The values of outliers can be quite substantial and influential. A look at Table 5 shows the advantage of removing such observations from the data (here, case 11, case 27, and case 49), then refitting the newly revised model to the remaining observations.

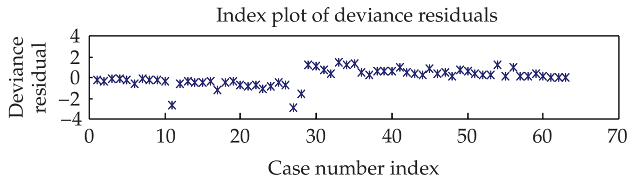

Figure 3: Plot of deviance residual (Resdev) versus case number index.

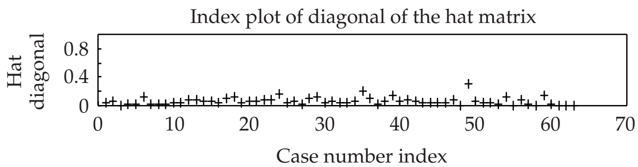

Figure 4: Plot of hat diagonal (Resdev) versus case number index.

The goodness-of-fit statistics are presented in Table 2. The estimates of deviance are shown in the column marked value/DF. The dispersion parameter (value/DF) is 0.3376 and the Pearson Chi-square dispersion parameter is 0.4752. The values of the deviance and Pearson Chi-square are less than the degrees of freedom, while the P-values of the deviance and Pearson Chi-square are all >0.05 (i.e., 1.0000, 0.9993, resp.). These indicate that this model seems to have an acceptable fit with the data.

3.1.4. Testing the global null hypothesis: $\beta = 0$

When testing the null hypothesis for large samples, the explanatory variables have coefficients of zero. According to the Chi-squares analysis, the associated P-values are all approximately zero, suggesting that the explanatory coefficients are all zero.

The results obtained after rerunning the unconditional asymptotic logistic regression after the removal of some of the observations from the data (i.e., case 11, case 27, and case 49) (see Table 3) still contain some unconditional asymptotic results. These results are obtained by deriving the Chi-square statistics while testing for the global null hypothesis ($\beta = 0$) (likelihood ratio, score, and Wald tests). For the likelihood ratio and score tests, the null hypothesis that β is zero is rejected, but not for the Wald test. The seeming discrepancies in P-values obtained between the Wald test and the other two tests are a sign that the large-sample approximation is not stable.

3.2. Exact logistic regression model

Exact logistic regression for binary outcomes can be utilized to provide an exact score test and an exact probability test for hypotheses where the parameters are equal to zero; these tests produce an exact P-value and a mid P-value.

To test whether individual parameter estimates are zero, we also require point estimates of the parameters, an odds ratio that contains two-sided confidence limits, and the P-value.

Table 2: Deviance and Pearson goodness-of-fit statistics.

Criterion	DF	Value	Value/DF	Pr > Chi-sq
Deviance	49	16.5436	0.3376	1.0000
Pearson	49	23.2826	0.4752	0.9993

Table 3: Testing of the global null hypothesis: $\beta = 0$.

Test	Chi-square	DF	Pr > Chi-sq
Likelihood ratio	65.5642	3	<.0001
Score	37.4643	3	<.0001
Wald	7.8052	3	0.0502

3.2.1. Conditional exact tests: $\beta = 0$

The results of exact conditional analysis obtained using the exact logistic regression model are shown in Table 4. The results for the exact score conditional test and the probability test are also reported in this table. For the joint test it is required that all the parameters for the exact statement be simultaneously equal to zero, that is, the null hypothesis is $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$.

In the joint test results an exact P-value of <.0001 is produced; the probability test produces an exact P-value of 0.0023. These test results lead to a rejection that the null hypothesis of $\beta_1 = \beta_2 = \beta_3$ is zero. This shows that the ridge height (X_1), lower layer water depth (X_2), and potential energy (X_3) are significant for the joint exact test.

Given the effects of the ridge height (X_1), lower layer water depth (X_2), and potential energy (X_3), the exact P-value and mid P-value are both <.0001. These results lead to a rejection of the null hypothesis that β_i is zero.To put it another way, ridge height (X_1), lower layer water depth (X_2), and potential energy (X_3) are significant factors associated with the amplitude-based incident rate.

3.2.2. Parameter estimation and odds ratio estimation

Stokes et al. [27] have suggested that large sample theory may not be appropriate for small-sized data. This thus means that tests based on the asymptotic normality of the MLEs may be unreliable. They recommend that when sample sizes are small, with approximate P-values of less than 0.10, it is a good idea to look at the exact results. If the approximate P-values are larger than 0.15, then the approximate methods are probably satisfactory, in the sense that the exact results are likely to agree with them.

Parameter estimates for unconditional asymptotic logistic regression

The analytical results for the estimated maximum likelihood and odds ratios are shown in Tables 5 and 6. The ridge height (X_1), lower layer water depth (X_2), and potential energy (X_3) are all significant factors affecting the amplitude-based incident rate ($P = .0106$, $P = .0053$, and $P = .0067$, resp.).

The fitted unconditional asymptotic logistic regression lines can be stated as

$$\begin{aligned}\operatorname{logit}(\hat{p}) = \ln\left[\frac{\hat{p}}{1-\hat{p}}\right] &= \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 \\ &= -10.9958 - 0.7916 x_1 + 1.1171 x_2 - 0.7211 x_3.\end{aligned} \tag{3.1}$$

Table 4: Conditional exact test results.

Effect	Test	Statistic	P-value Exact	Mid
Joint	Score	37.8649	<.0001	<.0001
	Probability	5.2E-18	0.0023	0.0023
Intercept	Score	7.5473	0.0082	0.0051
	Probability	0.00618	0.0082	0.0051
X_1	Score	22.5898	<.0001	<.0001
	Probability	7.315E-8	<.0001	<.0001
X_2	Score	30.4083	<.0001	<.0001
	Probability	2.96E-11	<.0001	<.0001
X_3	Score	22.0683	<.0001	<.0001
	Probability	2.488E-8	<.0001	<.0001

Table 5: Analysis of MLEs.

Parameter	DF	Estimate	Standard Error	Wald Chi-square	Pr > Chi-sq
Intercept	1	−10.9958	4.7733	5.3067	0.0212
X_1	1	−0.7916	0.3099	6.5232	0.0106
X_2	1	1.1171	0.4008	7.7694	0.0053
X_3	1	−0.7211	0.2659	7.3561	0.0067

Parameter estimation for conditional exact logistic regression

The analytical results of the exact parameter estimates and exact odds ratio estimates are presented in Tables 7 and 8, respectively. The ridge height (X_1), lower layer water depth (X_2), and potential energy (X_3) are all significant factors affecting the amplitude-based incident rate ($P < .0001$). We create a median unbiased estimate instead of the conditional MLE, because the value of the observed sufficient statistic lies at the extreme end of the derived distribution. The implication is that the conditional MLE does not exist. Even though the asymptotic results are unreliable, the exact analysis allows us to conclude that these factors have a significant effect. The fitted conditional exact logistic regression lines can be formulated as

$$\begin{aligned} \operatorname{logit}(\widehat{p}) &= \ln\left[\frac{\widehat{p}}{1-\widehat{p}}\right] = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 \\ &= -4.6013 - 0.6384x_1 + 0.8277x_2 - 0.6120x_3. \end{aligned} \tag{3.2}$$

We can see from Tables 5 and 7 that the parameters obtained from conditional exact logistic regression are smaller than those obtained from unconditional asymptotic logistic regression, but the P-values of the unconditional asymptotic estimates are larger than those of the exact estimates. A comparison of the odds ratio estimates (in Tables 6 and 8) shows that the parameters obtained from the conditional exact logistic regression are different than those obtained from the unconditional asymptotic logistic regression.

Stokes et al. [27] recommended that when sample sizes are small and the approximate P-values are less than 0.10, it is better to look at the exact results. Thus in this study, the small sample size and P-values make exact analysis more appropriate.

Table 6: Odds ratio estimates.

Effect	Point Estimate	95% Wald Confidence limits
X_1	0.453	0.247 ~ 0.832
X_2	3.056	1.393 ~ 6.703
X_3	0.486	0.289 ~ 0.819

Table 7: Exact parameter estimates.

Parameter	Estimate	95% Confidence Limits		P-value
Intercept	−4.6013*	−Infinity	−1.1169	0.0124
X_1	−0.6384*	−Infinity	−0.2851	<.0001
X_2	0.8277*	0.4196	Infinity	<.0001
X_3	−0.6120*	−Infinity	−0.2790	<.0001

NOTE: *indicates a median unbiased estimate.

4. Conclusions

A laboratory experiment is designed to investigate the propagation of an internal solitary wave over a submerged ridge. Analytical methods and a logistic regression model are employed to examine the amplitude-based incident rate. Large sample theory may not be suitable for data with small cell counts. This tends to make tests based on the asymptotic normality of the MLEs unreliable.

The ridge height, lower layer water depth, and potential energy are considered in the regression model. Once a model has been fitted to the observed values of a binary response variable, it is essential to check the validity of the fit. We discuss some methods for exploring the adequacy of the model and some diagnostic methods. The techniques used to examine the adequacy of a fitted unconditional asymptotic logistic regression model and conditional exact logistic regressions are known as diagnostics methods for testing the global null hypothesis. Based on the analytical results we can draw the following conclusions.

(1) The unconditional asymptotic logistic model results lead us to the conclusion that the three explanatory variables (ridge height, lower layer water depth, and potential energy) are significant factors affecting the amplitude-based incident rate. Both deviance and Pearson Chi-square tests are used to examine the goodness-of-fit of the model. The dispersion parameter for the estimate of deviance (value/DF) is 0.7268, and the Pearson Chi-square dispersion parameter is 1.2157. Preferably, this value should be very close to 1.00. The Pearson parameter is slightly larger than the degrees of freedom. We note that there is still a little overdispersion with this model which means that it needs to be modified.

(2) A look at the index plots for the Pearson residuals (Figure 2) and the deviance residuals (Figure 3) shows that case 11 and case 27 are poorly accounted for by the model. In the index plot of the diagonal elements of the hat matrix (Figure 4), case 49 is an extreme point in the design space. After these observations (case 11, case 27, and case 49) are removed from the data, the new revised model is refitted based on the remaining observations. The values of the deviance and Pearson Chi-squares are now less than the degrees of freedom, and the P-values for deviance and Pearson Chi-square are all >0.05 (1.0000, 0.9993, resp.). In other words, this revised model seems to fit the data acceptably well.

Table 8: Exact odds ratios.

Parameter	Estimate	95% Confidence Limits		P-value
X_1	0.528*	0	0.752	<.0001
X_2	2.288*	1.521	Infinity	<.0001
X_3	0.542*	0	0.757	<.0001

NOTE: *indicates a median unbiased estimate.

(3) When testing the global null hypothesis $\beta = 0$, only three Chi-square statistics (likelihood ratio, score, and Wald tests) are generated. The P-values obtained by logistic regression for the likelihood ratio test and score test are both <0.05. However, the null hypothesis is not rejected for the Wald test. The seeming discrepancies in P-values obtained between the Wald test and the other two tests are a sign that the large-sample approximation is not stable.

(4) The results of exact conditional analysis from the exact logistic regression model are shown in Table 4. The ridge height (X_1), lower layer water depth (X_2), and potential energy (X_3) are all significant in the joint results. The ridge height (X_1), lower layer water depth (X_2), and potential energy (X_3) effects are all significant factors affecting the amplitude-based incident rate.

(5) A comparison of the parameters shown in Tables 6 and 8 and the odds ratio estimates in Tables 6 and 8 shows that the parameters and the odds ratio estimates obtained from conditional exact logistic regression are different from those obtained from unconditional asymptotic logistic regression. As recommended by Stokes et al. [27], in cases of small sample sizes where the approximate P-values are less than 0.10, it is a good idea to look at the exact results. For this study, the small sample size and P-values indicate that an exact analysis would be more appropriate.

Acknowledgments

The authors would like to thank the National Science Council of the Republic of China, Taiwan for financial support of this research under Contracts no. NSC 96-2628-E-366-004-MY2 and NSC 96-2628-E-132-001-MY2. They also wish to thank the editor of Mathematical Problems in Engineering, and the three anonymous reviewers for their helpful suggestions on the improvement of this paper.

References

[1] A. R. Osborne and T. L. Burch, "Internal solitons in the Andaman Sea," *Science*, vol. 208, no. 4443, pp. 451–460, 1980.

[2] J. R. Apel, J. R. Holbrook, J. Tsai, and A. K. Liu, "The Sulu Sea internal soliton experiment," *Journal of Physical Oceanography*, vol. 15, no. 12, pp. 1625–1651, 1985.

[3] A. K. Liu, J. R. Holbrook, and J. R. Apel, "Nonlinear internal wave evolution in the Sulu Sea," *Journal of Physical Oceanography*, vol. 15, no. 12, pp. 1613–1624, 1985.

[4] A. K. Liu, Y. S. Chang, M. K. Hsu, and N. K. Liang, "Evolution of nonlinear internal waves in the East and South China Seas," *Journal of Geophysical Research*, vol. 103, no. C4, pp. 7995–8008, 1998.

[5] A. K. Liu and M. K. Hsu, "Internal wave study in the South China Sea using Synthetic Aperture Radar (SAR)," *International Journal of Remote Sensing*, vol. 25, no. 7-8, pp. 1261–1264, 2004.

[6] M. K. Hsu and A. K. Liu, "Nonlinear internal waves in the South China Sea," *Canadian Journal of Remote Sensing*, vol. 26, no. 2, pp. 72–81, 2000.

[7] M. K. Hsu, A. K. Liu, and C. H. Lee, "Using SAR image to study internal waves in the Sulu Sea," *Journal of Photogrammetry and Remote Sensing*, vol. 3, pp. 1–14, 2003.
[8] K. Zeng and W. Alpers, "Generation of internal solitary waves in the Sulu Sea and their refraction by bottom topography studied by ERS SAR imagery and a numerical model," *International Journal of Remote Sensing*, vol. 25, no. 7-8, pp. 1277–1281, 2004.
[9] Q. Zheng, R. D. Susanto, C.-R. Ho, Y. T. Song, and Q. Xu, "Statistical and dynamical analyses of generation mechanisms of solitary internal waves in the northern South China Sea," *Journal of Geophysical Research*, vol. 112, no. 3, Article ID C03021, 2007.
[10] T. W. Kao, F.-S. Pan, and D. Renouard, "Internal solitons on the pycnocline: generation, propagation, and shoaling and breaking over a slope," *Journal of Fluid Mechanics*, vol. 159, pp. 19–53, 1985.
[11] K. R. Helfrich, "Internal solitary wave breaking and run-up on a uniform slope," *Journal of Fluid Mechanics*, vol. 243, pp. 133–154, 1992.
[12] F. Wessels and K. Hutter, "Interaction of internal waves with a topographic sill in a two-layered fluid," *Journal of Physical Oceanography*, vol. 26, no. 1, pp. 5–20, 1996.
[13] H. Michallet and G. N. Ivey, "Experiments on mixing due to internal solitary waves breaking on uniform slopes," *Journal of Geophysical Research*, vol. 104, no. C6, pp. 13467–13477, 1999.
[14] C.-Y. Chen and J. R.-C. Hsu, "Interaction between internal waves and a permeable seabed," *Ocean Engineering*, vol. 32, no. 5-6, pp. 587–621, 2005.
[15] C.-Y. Chen, J. R.-C. Hsu, C.-F. Kuo, H.-H. Chen, and M.-H. Cheng, "Laboratory observations on internal solitary wave evolution over a submarine ridge," *China Ocean Engineering*, vol. 20, no. 1, pp. 61–72, 2006.
[16] C.-Y. Chen, J. R.-C. Hsu, M.-H. Cheng, H.-H. Chen, and C.-F. Kuo, "An investigation on internal solitary waves in a two-layer fluid: propagation and reflection from steep slopes," *Ocean Engineering*, vol. 34, no. 1, pp. 171–184, 2007.
[17] C.-Y. Chen, "An experimental study of stratified mixing caused by internal solitary waves in a two-layered fluid system over variable seabed topography," *Ocean Engineering*, vol. 34, no. 14-15, pp. 1995–2008, 2007.
[18] D. A. Cacchione, L. F. Pratson, and A. S. Ogston, "The shaping of continental slopes by internal tides," *Science*, vol. 296, no. 5568, pp. 724–727, 2002.
[19] C.-W. Chen, C.-Y. Chen, P. H.-C. Yang, and T.-H. Chen, "Analysis of experimental data on internal waves with statistical method," *Engineering Computations*, vol. 24, no. 2, pp. 116–150, 2007.
[20] C.-W. Chen, H.-C. P. Yang, C.-Y. Chen, A. K.-H. Chang, and T.-H. Chen, "Evaluation of inference adequacy in cumulative logistic regression models: an empirical validation of ISW-ridge relationships," *China Ocean Engineering*, vol. 22, no. 1, pp. 43–56, 2008.
[21] C.-Y. Chen, J. R.-C. Hsu, C.-W. Chen, C.-F. Kuo, H.-H. Chen, and M.-H. Cheng, "Wave propagation at the interface of a two-layer fluid system in the laboratory," *Journal of Marine Science and Technology*, vol. 15, no. 1, pp. 8–16, 2007.
[22] D. R. Cox, *The Analysis of Binary Data*, Methuen, London, UK, 1970.
[23] K. F. Hirji, C. R. Mehta, and N. R. Patel, "Computing distributions for exact logistic regression," *Journal of the American Statistical Association*, vol. 82, no. 400, pp. 1110–1117, 1987.
[24] D. R. Cox and E. J. Snell, *Analysis of Binary Data*, vol. 32 of *Monographs on Statistics and Applied Probability*, Chapman & Hall, London, UK, 2nd edition, 1989.
[25] C. R. Mehta and N. R. Patel, "Exact logistic regression: theory and examples," *Statistics in Medicine*, vol. 14, no. 19, pp. 2143–2160, 1995.
[26] R. E. Derr, "Performing exact regression with the SAS system," in *Proceedings of the 25th Annual SAS Users Group International Conference*, Cary, NC, USA, April 2000, paper P254-25.
[27] M. E. Stokes, C. S. Davis, and G. G. Koch, *Categorical Data Analysis Using the SAS System*, SAS Institute, Cary, NC, USA, 2000.
[28] D. W. Hosmer and S. Lemeshow, *Applied Logistic Regression*, John Wiley & Sons, New York, NY, USA, 2nd edition, 2000.
[29] A. Agresti, *Categorical Data Analysis*, Wiley Series in Probability and Statistics, John Wiley & Sons, New York, NY, USA, 2nd edition, 2002.
[30] D. Collett, *Modelling Binary Data*, Chapman & Hall/CRC Texts in Statistical Science Series, Chapman & Hall/CRC, Boca Raton, Fla, USA, 2nd edition, 2003.
[31] J. M. Landwehr, D. Pregibon, and A. C. Shoemaker, "Graphical methods for assessing logistic regression models," *Journal of the American Statisical Association*, vol. 79, no. 385, pp. 61–71, 1984.
[32] SAS Institute, *SAS/STAT User's Guide*, vol. 2, SAS Institute, Cary, NC, USA, 8th edition, 2000.