Channel Scouring Potential Using Logistic Analysis

Yeou-Koung Tung

1985
Journal Article WWRC-85-47

In

Journal of Hydraulic Engineering

Volume 111

Yeou-Koung Tung
Wyoming Water Research Center
and
Department of Statistics
University of Wyoming
Laramie, Wyoming

Channel Scouring Potential Using Logistic Analysis

By Yeou-Koung Tung,[1] A. M. ASCE

Abstract: Water resource engineers often have to relate qualitative dependent variables to one or more independent variables, which may or may not be quantitative. In such circumstances, the use of conventional regression analysis would encounter a number of difficulties. This paper introduces a statistical method called logistic regression which is specially developed for such conditions. The method is applied to a hydraulic problem of relating scouring potential in a channel to depth and velocity of flow. Whether or not the methodology could become a useful addition in water resources engineering analyses further investigations and applications are necessary.

Introduction

The evaluation and assessment of channel scouring potential has been one of the main concerns of hydraulic engineers. Methods developed for evaluating the scouring potential range from simple empirical-based diagrams, e.g., the famous Shields diagram (13), to sophisticated physical-based probabilistic models (7). Due to the facts that the distribution of bed material is not uniform, the channel geometry varies from one location to another, the flow is nonuniform and unsteady, and due to other factors, the occurrence of scouring in the channel is a random phenomenon.

Water resource engineers often have to relate qualitative dependent variables to one or more independent variables, which may or may not be quantitative. In the case of channel scouring, the dependent variable is the occurrence or nonoccurrence of scouring, which is binary and qualitative, depending on a number of attributes such as the particle size of bed material, flow velocity, depth of water, etc. (4,12,13). Because of the binary nature of the dependent variable, the procedures used by most water resource engineers for the assessment of scouring potential in the streams are empirical with little scientific basis. The well-known Shields diagram and other similar graphical representations (3) for the assessment of the occurrence of stream scouring are of this type. Sometimes, this type of diagram is applied by engineers for stable channel design.

It is important to realize that diagrams so developed can be misleading in the sense that they fail to illustrate the random characteristics of the process. Therefore, the use of Shields diagram or other diagrams of a similar nature for stable channel design could lead to unsatisfactory results if care is not exercised. However, it would be ideal and realistic if certain explicit probabilistic statements were attached to the diagrams so that the users could be aware of the random nature of the process and design a stable channel with a reliability (or risk) level best suited for what is to be designed. Furthermore, explicit presentations of probabilistic statements enable designers to consider the trade-off between risk and benefit and to perform optimal risk-based design.

Hughes (8) is one of the few water resource engineers, who proposed a technique for analyzing binary data in his channel scouring study. However, the technique developed by Hughes is not applicable to cases where there are three or more independent variables and the functional relationship between variables is unknown, which generally is the case in real world problems. The generalization and application of Hughes' methodology to other problems would be difficult, and no mathematical function can be derived to relate the probability of occurrence of channel scouring to contributing factors.

This paper presents a methodology to evaluate the potential of channel scouring as function of a number of attributes by logistic analysis.

Why Logistic Regression?

When a dependent variable is related to a number of independent variables, regression analysis techniques are usually applied. Commonly applied regression analysis techniques are appropriate to use only when both independent variable and explanatory variables are quantitative and continuous. For cases where the dependent variable and some of the independent variables are qualitative, ordinary regression analysis is no longer applicable. To analyze a dichotomous (binary) qualitative variable as a function of a number of explanatory variables, special techniques must be used if the analysis is to be performed adequately. A dependent variable with dichotomous nature is normally measured by 0 or 1 indicating the nonoccurrence or occurrence of an event. In the case of channel scouring, we can use 1 to indicate the occurrence of scouring and 0 otherwise.

Consider a dichotomous random variable y which has a value of 1 if event E occurs, and 0 otherwise. Let $\mathbf{x}_{k\times 1}$ denote a k-vector of variables which have influence on the occurrence or nonoccurrence of event E. In statistical data analysis, the vector $\mathbf{x}$ is used as explanatory variables for y. Suppose we have made a set of n independent observations of $(y, \mathbf{x})$, i.e., $(y_i, \mathbf{x}_i)$, $i = 1, 2, \ldots, n$, and wish to derive a functional relationship between y and x's. Let us adopt the standard regression model

$$y_i = \mathbf{x}_i'\boldsymbol{\beta} + \epsilon_i, \quad \text{for} \quad i = 1, 2, \ldots, n \qquad (1)$$

in which ϵ_i = random error term characterized by distributional properties, and $\boldsymbol{\beta}$ = k-vector of unknown model parameters. However, it is because the very nature of dependent variable y in Eq. 1 being dichotomous and discrete some modifications and special treatment of variable y are necessary. The pitfalls of using ordinary least squares method in regression analysis, when dependent variable is dichotomous, and the evolution of logistic analysis are described in the following:

1. From physical standpoint, the value of dependent variable y can only take a value of either 1 or 0, representing the occurrence or non-

[1]Asst. Prof., Wyoming Water Research Center and Dept. of Statistics, Univ. of Wyoming, Laramie, Wyo.

Note.—Discussion open until July 1, 1985. To extend the closing date one month, a written request must be filed with the ASCE Manager of Journals. The manuscript for this paper was submitted for review and possible publication on November 16, 1983. This paper is part of the *Journal of Hydraulic Engineering*, Vol. 111, No. 2, February, 1985. ©ASCE, ISSN 0733-9429/85/0002-0194/$01.00. Paper No. 19499.

occurrence of an event. Any values other than 1 and 0 assigned to y are physically meaningless. Ordinary regression analysis does not have the capability to handle problems of this nature.

2. It is generally easier to treat a continuous variable than a discrete one. Dependent variables of dichotomous nature would pose computational difficulties in data analysis. To get around such problems we will deal with the conditional probability of occurrence of event E, i.e., $Pr\{y = 1|\mathbf{x}'\boldsymbol{\beta}\}$, instead of y itself where $Pr\{\ \}$ is probability. To shorten the notation, let $\xi = Pr\{y = 1|\mathbf{x}'\boldsymbol{\beta}\}$. By doing this, the problem is transformed to regress ξ on independent variables x's rather than y on x's as originally stated and dependent variable ξ is continuous between 0 and 1. Furthermore, we modify the theory demanding that ξ will approach to 1 as the value of $\mathbf{x}'\boldsymbol{\beta}$ becomes larger and hypothesize that the relationship between $\mathbf{x}'\boldsymbol{\beta}$ and ξ is monotonic. Then the true probability function should generally have a cumulative distribution function (CDF) of S-shape since ξ must lie between 0 and 1 and be nondecreasing. If we plot ξ verses $\mathbf{x}'\boldsymbol{\beta}$, then their relationship should appear as the solid curve as shown in Fig. 1, with its upper and lower asymptotes being 1 and 0, respectively. However, even dependent variable is made continuous to which ordinary regression analysis is less restrictive to apply. The result of unconstrained regression analysis, which would appear as a straight line, has poor approximation for large or small values of $\mathbf{x}'\boldsymbol{\beta}$, and, indeed, violates the condition that the function lies between 0 and 1 for extreme values of the argument (see Fig. 1).

3. With the previous modifications and under the assumptions of regression model in Eq. 1, for fixed explanatory variables $\mathbf{x}_i$, y_i is a Bernoulli random variable with $E(y_i|\mathbf{x}_i) = \xi_i$ and $\mathrm{Var}(y_i|\mathbf{x}_i) = \mathrm{Var}\ (\epsilon_i) = \xi_i(1 - \xi_i)$, where $E(|)$ and $\mathrm{Var}(|)$ are the conditional expectation and variance of y, respectively. Since $\mathrm{Var}(\epsilon_i)$ depends on ξ_i which, in turn, depends on $\mathbf{x}_i$, the ξ_i are heteroscedastic. That is, the variance of regression model

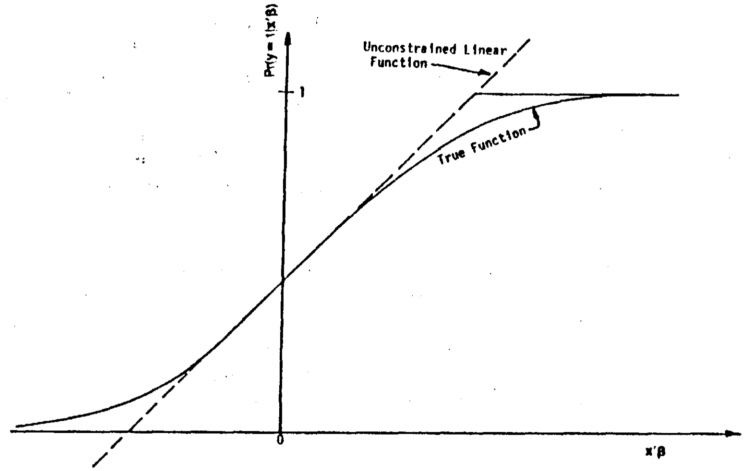

FIG. 1.—Linear Approximation to Probability Function (Nerlove and Press, 1973)

depends on the level of dependent variable. Under such circumstance the use of ordinary least-squares procedure generates inefficient estimators and imprecise predictions. It has been suggested that generalized (or weighted) least squares can be used to remove the heteroscedasticity. Unfortunately, there is no guarantee that predicted values of y_i will lie within the interval (0,1) for all i.

An excellent review on the limitations of using ordinary least-squares regression analysis applied to the problem of analyzing binary data is given by Nerlove and Press (11).

Logistic Analysis

To avoid the technical difficulties described previously when ordinary regression analysis is used to treat binary data, a special data analysis procedure is presented herein. Let us consider the problem of relating the probability of occurrence of event, ξ, to a number of explanatory variables and Eq. 1 is the model to be assumed. If one wants a relationship in which ξ_i is a nondecreasing function of stimulus level $\mathbf{x}_i'\boldsymbol{\beta}$, the following model may be used

$$\xi_i = F(\mathbf{x}_i'\boldsymbol{\beta}), \quad i = 1, \ldots, n \qquad (2)$$

where $F(\mathbf{x}_i'\boldsymbol{\beta})$ denotes a CDF. In general, any distribution function $F(t)$ or function with values lying between 0 and 1 can be candidates for use. The choice of the form of the function, $F(t)$, to be used in the analysis depends largely upon the mathematical properties, computational tractability, and feasibility of implementation.

One of the earliest methods to analyze binary data is called "probit analysis." In probit analysis, the functional form of $F(t)$ in Eq. 2 is the CDF of the standardized normal distribution. Finney (2) applied probit analysis to the problem of analyzing quantal (binary) responses in bioassay. For the probit analysis to be useful, it requires that data be grouped and there are several observations in each group. Also, there is computational difficulty associated with the numerical integration of the standard normal distribution (11).

Another method developed is called logistic analysis where $F(t)$ in Eq. 2 is taken to be the CDF of the standard logistic distribution; that is

$$F(t) = \frac{1}{1 + e^{-t}}, \quad -\infty < t < \infty \qquad (3)$$

After logistic transformation, Eq. 3 can be expressed as

$$\ln\left(\frac{\xi}{1 - \xi}\right) = \mathbf{x}'\boldsymbol{\beta} \qquad (4)$$

in which ξ = probability of the occurrence of event E conditioned upon the vector of explanatory variable $\mathbf{x}$. The left-hand-side of Eq. 4 is called a logit function. As can be seen, the problem is reduced to performing regression analysis to estimate the values of unkown parameters β's using logit function. The value of logit function is continuous over the entire domain of real variables. Various methodologies such as weighted least square logit (10), minimum chi-square (1), minimum logit chi-square

(10), discriminant analysis (6), and maximum likelihood logit (11) have been proposed to perform logistic regression. Note that minimum chi-square and minimum logit chi-square are two different methods.

For the purpose of demonstrating the concept of logistic analysis the minimum chi-square method is applied to analyze channel scouring data in Hughes article (8). The comparison of performance for various methods is undertaken and is beyond the scope of the paper.

Minimum Chi-Square Method

Without loosing generality, following descriptions of minimum chi-square method assume that total data points have been grouped. Combining Eqs. 2 and 3 the conditional probability of the occurrence of event E, given a vector of explanatory variable $\mathbf{x}$ for group l, can be written as

$$\xi_l(\beta) = Pr(Y_l|n_l,\ \mathbf{x}_l'\beta) = \frac{1}{\{1 + \exp(-\mathbf{x}_l'\beta)\}} \quad \text{(5a)}$$

or, more explicitly

$$\xi_l(\beta) = \frac{1}{\{1 + \exp(-\beta_0 - \beta_1 x_{l1} - \ldots - \beta_k x_{lk})\}} \quad \text{(5b)}$$

in which n_l = total number of observations in group l, Y_l = the number of occurrence of event E in group l, i.e., $Y_l = \Sigma Y_i$, $i \in$ group l and $\mathbf{x}_l$ = vector of independent variables with values representative to group l. For example, x_{lk} can be calculated using the average value of the kth independent variable in group l. Since Y_l is the sum of independent dichotomous random variable so it has a binomial distribution with parameters n_l and ξ_l. The statistic, Q, in Eq. 6

$$Q = \sum_{l=1}^{G} \frac{[Y_l - n_l\xi_l(\beta)]}{n_l\xi_l(\beta)\,[1 - \xi_l(\beta)]} \quad \text{(6)}$$

approximately has chi-square distribution in which G = total number of groups. The method of minimum chi-square is to determine the values of the β's such that the value of statistic Q is minimized. In case where data are ungrouped, $G = n$, $n_l = 1$, and $Y_l = y_l$.

The problem of determining the values of β's for minimizing Q falls within the realm of unconstrained nonlinear programming and the technique used herein is called Davidon-Fletcher-Powell (DFP) method. The DFP method is a quasi-Newton method which simultaneously generates the search directions of the conjugate gradient method while updating the Hessian matrix. It requires the computation of the gradient of Q with respect to β's, i.e., $\partial Q/\partial \beta_j$, at each iteration for purpose of determining the search direction and updating the Hessian. For a more detailed description of the method, readers are referrred to Luenberger (5).

Application

Application of logistic analysis described previously to evaluate channel scouring potential is presented in this section. Data obtained from

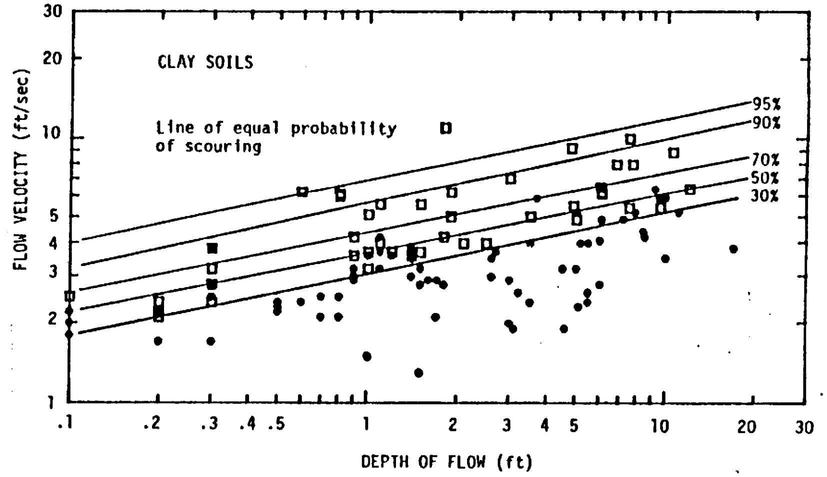

FIG. 2.—Channel Scouring Potential for Clay Soils Derived by Logistic Analysis Using Hughes' Data

field investigations were provided by Hughes (personal communication, 1979) and they are representative of sandy-silt, silty-clay, and clay soils. The data are plotted as scatter diagrams on log-log paper shown in Figs. 2–4. In each scatter diagram a square indicates the occurrence of scouring and a circle represents the nonoccurrence of scouring. The state of channel conditions shown in Figs. 2–4 was extracted from Keeley (9) and Hughes' judgment based on field observations (Hughes, personal communication, 1984). For a hydraulic engineer, once the scatter dia-

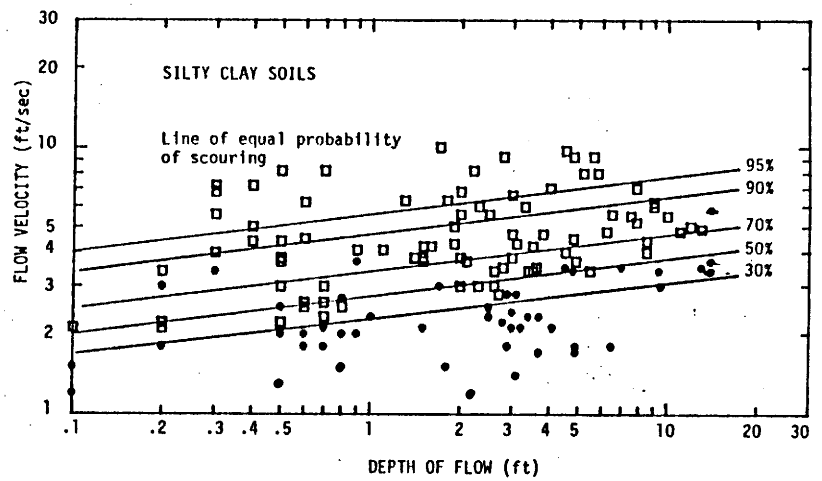

FIG. 3.—Channel Scouring Potential for Silty-Clay Soils Derived by Logistic Analysis Using Hughes' Data

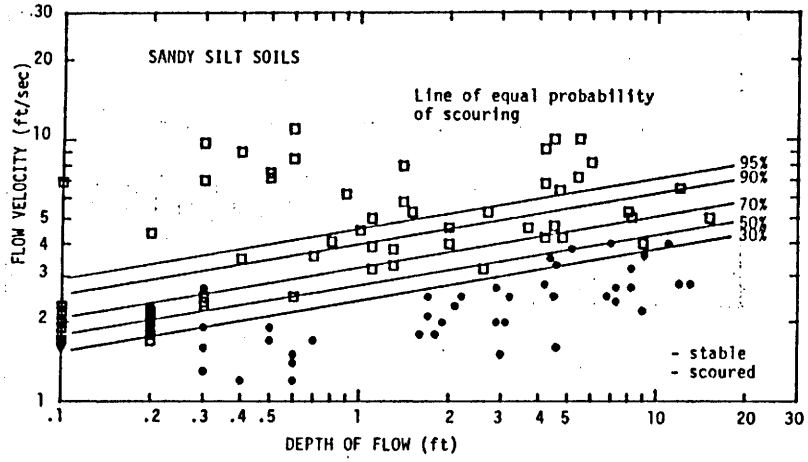

FIG. 4.—Channel Scouring Potential for Sandy-Silt Soils Derived by Logistic Analysis Using Hughes' Data

gram is prepared, the next step is to analyze the data and draw, if possible, a line which distinctively separates the incidence of erosion and nonerosion. Apparently, in this case, it is not possible. Therefore, it is argued that, in terms of evaluating channel scouring potential, it would be more reasonable to derive a probabilistic measure under various conditions instead of deriving a curve which best separates the occurrence and nonoccurrence of scouring and using it as a criterion in the assessment of scouring in the channel.

Based on Hughes' data, for a given soil type the average flow velocity and water depth are the two dominant factors affecting the scouring potential in the channels. The corresponding logistic regression model can be written as

$$Pr\{y_i = 1|x_{i1}, x_{i2}\} = \frac{1}{\{1 + \exp(-\beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2})\}} \quad \text{(7)}$$

in which $Pr\{|\}$ is the conditional probability of channel scouring depending on flow velocity, x_{i1}, and water depth, x_{i2}, for the ith observation. Similar to probit analysis, using minimum chi-square method in logistic regression requires that data are grouped and there are several observations in each group. The requirement to have several observations in each group could impose difficulty on determing proper interval boundaries for each independent variable if large number of groups are to be formed. This difficulty would gradually vanish as number of groups become smaller, but the problem of loss of information content would emerge. In this application, logistic analyses were performed under which the Hughes' data are both ungrouped and grouped into 16 cells. Two different logistic models are examined in this paper. One uses the flow velocity and water depth of original scale and the other uses the log-transformed flow velocity and water depth as independent variables.

TABLE 1.—Results of Logistic Analysis by Minimum Chi-Square Method Using Flow Velocity and Water Depth of Original Scale

Soil type (1)	Model Parameters $\hat{\beta}_0$ (2)	$\hat{\beta}_1$ (3)	$\hat{\beta}_2$ (4)	Total number of observations n (5)	Number of incidences of scouring n_1 (6)	Number of misclassifications (7)	Value of Q (8)
Sandy-silt[a]	−3.670	1.677	−0.381			18	53.42
Sandy-silt[b]	−3.533	1.815	−0.800	106	57	23	6.41
Silty-clay[a]	−3.706	1.511	−0.287			21	62.00
Silty-clay[b]	−3.728	1.583	−0.325	133	85	23	7.64
Clay soil[a]	−2.960	0.926	−0.313			24	73.23
Clay soil[b]	−2.962	0.889	−0.327	112	43	23	6.18

[a]Ungrouped data.
[b]Grouped data.

Hughes (8) used log-transformed flow velocity and log-transformed water depth because the relation between the two variables is assumed to be a power function. The values of estimated model parameters β's in Eq. 7 with both ungrouped and grouped data for different soil types using flow velocity and water depth of original scale and log-transformed scale are given in Columns 2–4 of Tables 1 and 2, respectively. The estimated model parameters values using ungrouped and grouped data are very close, except $\hat{\beta}_2$ for sandy-silt soil in Table 1.

In ordinary regression analysis the commonly used measure of model performance is the correlation coefficient. However, in logistic regressions the value of correlation coefficient is very small. It is because even the model may fit the probability very well, unless the probability is near 0 or 1, the outcome of y_i is not explained very well. Therefore, correlation coefficient does not provide useful information on the performance of a logistic model. After all, it is the capability of a model to predict the outcome of a discrete random event is of our concern.

To evaluate the performance of the two logistic models, a useful criterion is the number of incidences of misclassification by the two models.

TABLE 2.—Result of Logistic Analysis by Minimum Chi-Square Method Using Flow Velocity and Water Depth of Log-Transformed Scale

Soil type (1)	Model Parameters $\hat{\beta}_0$ (2)	$\hat{\beta}_1$ (3)	$\hat{\beta}_2$ (4)	Total number of observations n (5)	Number of incidences of scouring n_1 (6)	Number of misclassifications (7)	Value of Q (8)
Sandy-silt[a]	−6.327	6.183	−1.187			13	44.29
Sandy-silt[b]	−6.339	6.249	−1.276	106	57	13	5.09
Silty-clay[a]	−4.149	4.100	−0.563			14	67.13
Silty-clay[b]	−4.598	4.894	−0.855	133	85	14	10.43
Clay soil[a]	−6.287	4.884	−1.102			22	66.08
Clay soil[b]	−6.288	4.885	−1.126	112	43	23	2.64

[a]Data ungrouped.
[b]Grouped.

The general discriminant function for a logistic model can be expressed as (10)

$$\lambda(x_i) = \ln\left(\frac{\xi_i}{1-\xi_i}\right) = \ln\left(\frac{\pi_1}{\pi_0}\right) + \sum_{j=1}^{k} \hat{\beta}_j (x_{ij} - \bar{x}_j^*) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (8)$$

in which $\hat{\beta}_j$ = the estimate of jth parameter, x_{ij} = the value of jth independent variable of ith observation, and k = number of independent variables in logistic model. In this particular application $k = 2$. Each $\bar{x}_j^*$ is a weighted mean of the values of the jth independent variable. This average is

$$\bar{x}_j^* = \frac{1}{2} m_{j0} + \frac{1}{2} m_{j1} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (9)$$

with m_{j0} being the mean of the subset of n_0 values of x_{ij} for which $y_i = 0$, m_{j1} being the mean of the subset of n_1 values of x_{ij} for which $y_i = 1$, and $n_1 + n_0 = n$. We should keep in mind that, in logistic regression, even though y (a binary variable) was regressed on x, it is the logit function $\lambda(x)$, and not the conditional probability ξ, that is linear in x (see Eq. 4). From Eq. 4 the general discriminant function, Eq. 8, for predicting event outcomes can be derived. The first term in the right-hand-side of Eq. 8, $\ln(\pi_1/\pi_0)$, is the ratio of prior probabilities of occurrence to nonoccurrence of scouring. This means that π_1 is the probability of occurrence of channel scouring before measurements of flow characteristics were made, and $\pi_0 = 1 - \pi_1$ is the probability of no scouring occurring. If the values of π_1 and π_0 are unknown, they can be estimated by n_1/n and n_0/n, respectively. The schematic diagram of classification with two populations using Eq. 8 is illustrated in Fig. 5. To predict the occurrence or nonoccurrence of scouring at a given flow condition, we can base the value of discriminant function, Eq. 8, using the following rules: scouring will occur if $\lambda(x) < 0$; scouring will not occur if $\lambda(x) > 0$; and the chance is 50/50 if $\lambda(x) = 0$.

The total number of data points, number of incidences of scouring, and number of misclassifications by the two logistic models for different soil types are also given in Columns 5–7 of Tables 1 and 2. The average percentage of misclassification for the logistic model using ungrouped data with flow velocity and water depth of original scale is 21%, while 16% for the model using log-transformed scale. Similar results are observed using grouped data. Therefore, the model using log-transformed independent variables for sandy-silt and silty-clay soils out-performs that of using original scale. For clay soil, performance of the two models is indistinguishable. Furthermore, the results of using ungrouped data provide slightly better performance in classification than that of using grouped data. The values of statistic Q in Eq. 6 for various bed materials in two different models considered are given in the last column of Tables 1 and 2. The performance of logistic models based on their ability to predict the discrete outcome, occurrence or nonoccurrence of scouring, is generally consistent with the value of Q, i.e., less number of misclassification corresponding with a smaller value of Q, except for silty-clay case (Table 2). Exact causes for such inconsistency are yet to be explored. However, it might result from the effect of estimation of (π_1/

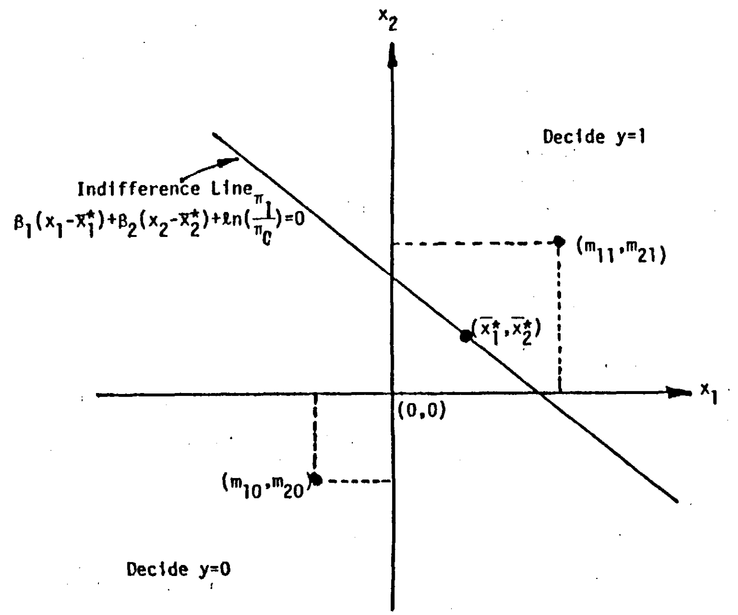

FIG. 5.—Classification of Binary Response with Two Independent Variables

π_0) in Eq. 8, data grouping procedure, the behavior of data set, and performance of search technique for optimization because the problem is nonlinear and global optimality cannot be assured.

Once the values in the logistic model are derived, an evaluation of channel scouring potential, conditioned upon the flow velocity and water depth, can easily be made. A series of curves of different potential levels of scouring can be imposed on the scatter diagrams as shown in Figs. 2–4 which are derived based on the results using ungrouped data. Take sandy-silt soils as an example: the corresponding velocities at which the channels have scouring probability of 0.95, 0.9, 0.7, 0.5, and 0.3, at a depth of 2 ft (0.6 m), are 5.2, 4.6, 3.8, 3.1, and 2.7 ft/sec (1.6, 1.4, 1.2, 0.95, and 0.82 m/s), respectively. Scatter diagrams with explicit probabilistic statements or derivations of functional expressions of scouring potential provides additional information to the designer and this enhances risk-based design.

Summary and Conclusions

Logistic regression is applicable when the dependent variable and some of the independent variables are qualitative. It also provides a means to evaluate the probability of a random event whose occurrence is affected by a number of attributes. To demonstrate the usefulness of the analysis in water resources engineering, the method is applied to relate scouring potential in a channel as function of velocity and depth of flow using

Hughes' data. However, the method requires further investigations to ascertain its applicability in practical design of stable channel.

Acknowledgment

The author wishes to thank Dr. Hughes at the University of Colorado for his generosity in providing the channel scouring data used in this paper. Sincere gratitude is extended to the reviewers for their constructive criticism.

Appendix I.—References

1. Berkson, J., "Maximum Likelihood and Minimum Chi-Square Estimates of the Logistic Function," *Journal of American Statistical Association*, Vol. 50, 1955.
2. Finney, D. J., *Probit Analysis*, Cambridge University Press, Cambridge, England, 1947.
3. Gessler, J., "Beginning and Ceasing of Sediment Motion," in *River Mechanics*, edited by H. W. Shen, Water Resources Publication, Fort Collins, Colo., 1977.
4. Graf, W. H., *Hydraulics of Sediment Transport*, McGraw-Hill Book Co., New York, N.Y., 1971.
5. Luenberger, D. G., *Introduction to Linear and Nonlinear Programming*, Addison-Wesley, Reading, Mass., 1973.
6. Haggstrom, G., "Postscript on Logistic Regression," *Memorandum*, Apr. 9, 1974, The Rand Corporation.
7. He, M., and Qiwei, H., "Stochastic Model of Incipient Sediment Motion," *Journal of the Hydraulics Division*, ASCE, Vol. 108, No. HY2, Feb., 1982.
8. Hughes, W. C., "Scour Velocities in Ephemeral Channels," *Journal of the Hydraulics Division*, ASCE, Vol. 106, No. HY9, Sept., 1980.
9. Keeley, J. W., "Soil Erosion Studies in Oklahoma," *Parts I and II*, U.S. Bureau of Public Roads, Oklahoma Div., 1961.
10. Morris, C., and Ralph, J. E., "Introduction to Data Analysis and Statistical Inferences," *The Rand Paper Series*, P-5819, The Rand Corporation, Santa Monica, Calif., 1978.
11. Nerlove, M., and Press, S. J., "Univariate and Multivariate Log-Linear and Logistic Models," *Rand Publications*, R-1306-EDA/NIH, Dec., 1973.
12. Simons, D. B., and Senturk, F., *Sediment Transport Technology*, Water Resources Publication, Fort Collins, Colo., 1976.
13. Vanoni, V. A., Brooks, N. H., and Kennedy, J. F., "Lecture Notes on Sediment Transportation and Channel Stability," *Report*, No. KH-R-1, California Institute of Technology, 1971.

Appendix II.—Notation

The following symbols are used in this paper:

$E(|)$ = conditional expectation;
$F(\,)$ = cumulative distribution function;
G = number of groups in data set;
n = total number of observations;
n_1 = number of observations with $y = 1$;
n_0 = number of observations with $y = 0$;
Pr = conditional probability;
Q = error function having a chi-square distribution;
Var (|) = conditional variance;
$\mathbf{x}$ = vector of independent variables in regression model;
$\bar{x}^*$ = weighted mean of independent variable;
y = dependent variable in regression model;
$\boldsymbol{\beta}$ = vector of parameter in regression model;
ϵ = random error;
$\lambda(\,)$ = logit function;
ξ = probability of occurrence of event;
π_0 = prior probability of no scouring; and
π_1 = prior probability of scouring.