An Improved Bridge Safety Index for Narrow Bridges

B.V. RADHAKUMARI GANDHI, ROBERT L. LYTTON, and BIMAN DAS

ABSTRACT

An improved safety index model is developed for narrow bridges by using data collected on 78 bridges in Texas by the Texas Transportation Institute. Cluster analysis was used to classify the bridges into more-safe and less-safe groups. Correlation and factor analyses were used to find the interrelationships between the independent variables of bridge and approach roadway characteristics. Stepwise regression was used to find the most important variables related to accident rate. On the groups developed by cluster analysis, logistic regression was used to develop a model that is a function of several variables that were found to be significantly related to accident rate. The enhanced safety index model consists of the following variables: bridge width, length, average daily traffic, and speed, as well as the subjective safety factors F_6 (grade continuity), F_7 (shoulder reduction), and F_9 (traffic mix). The model was arrived at scientifically by using more objective procedures of classification and by having a stronger correlation with the accident rate than the previous models used by other researchers. The fewer variables used yield a much higher R^2 when accident rate is used as the response variable in multiple regression. It is sensitive to changes or improvements in the constituent factors. The model yields the fraction of concordant pairs of predicted probabilities and responses as 0.91 and a high rank correlation of 0.81 between predicted probability and response, which indicates the goodness of the model. It also gives the safety index directly and can be used to identify a potentially hazardous narrow bridge.

Many bridges built on U.S. highways before 1930 pose a safety hazard because they are structurally and geometrically deficient with respect to modern high-speed, high-volume traffic. Oglesby and Hicks (1) state that, on the Federal-Aid highway system alone, of the 240,000 bridges recently inventoried there are about 9,000 structurally obsolete and 31,000 functionally obsolete bridges. Several articles have been written in professional journals and news magazines discussing the gravity of this problem. The Better Roads inventory (2) noted that there are close to 90,000 substandard bridges in the United States. To bring even a fraction of the bridges to modern design standards involves billions of dollars.

In addition to structurally unsafe bridges, there are several other bridges in the United States that are structurally safe but narrow in width compared to the approach roadway width. Narrowing of the roadway on the bridge imposes a significant accident potential on the driving public. These accidents result from the impact of vehicles on bridge abutments, approach guardrails, and bridge railings and from the collisions with oncoming vehicles because of the narrowness of the bridge. Public awareness of the narrow bridge problem escalated after two major accidents at narrow bridges in New Mexico and Texas took a toll of 28 lives. These accidents resulted in a subcommittee hearing (3) in the U.S. Congress from June 12 to 14, 1973, and the narrow bridge problem attained nationwide attention.

A comprehensive analysis of safety at narrow bridges was conducted recently at the Texas Transportation Institute (4), and a bridge safety index (BSI) was formulated to distinguish between safe and less-safe bridges on the basis of several subjective factors related to the bridge and the approach roadway characteristics. The research reported here is related to the improvement of the BSI model for better classification of narrow bridges. Additional data were collected on factors that affect safety, and analyses using modern statistical techniques such as cluster, discriminant, and factor analysis and logistic regression were used to develop and arrive at an improved BSI model.

REVIEW OF RELATED STUDIES

A survey of narrow bridges (4) conducted by the Texas Transportation Institute (TTI) noted that different states had different criteria for defining narrow bridges. The questionnaire summary indicates that a large number of state bridges (7,211) are considered narrow if they are 18 ft or less in width, and a large number of city and municipal bridges (7,905) are considered deficient if they are 16 ft or less. Southwest Research Institute (SwRI) of San Antonio, Texas, in its study on narrow bridges (5), defined narrow bridges as 18 ft or less in width for one lane and 24 ft or less in width for two lanes. According to Johnson (6), any bridge that changes driver behavior with regard to speed or lateral positioning of the vehicle can be considered narrow. It appears, in general, that anything less than a 24-ft clear bridge width for a two-way bridge operation, or reduction of shoulder on the bridge or parameters that causes changes in drivers' lateral position and speed, leads to a narrow bridge condition.

One of the earliest studies of bridge accidents was conducted by Raff (7) for the Bureau of Public Roads in 1954. His analysis noted that traffic volume was found to have a major effect on accident rates. For roads carrying the same amount of traffic, sharp curves had higher accident rates than flat curves. Extra width in relation to the approach pavement definitely reduced accident hazard on bridges. Behnam and Laguros (8) attempted to relate accidents at bridges to roadway geometrics at bridge approaches. Multivariate regression and stepwise regression procedures were used in developing the models. The predictor models indicated that average daily traffic (ADT) was one of the most significant variables and that the relationship between traffic accidents and geometric elements of a roadway were not linear but could be expressed by logarithmic transformation. On two-lane roadways sight distance was found to be important for night driving, whereas the degree of curvature became critical during the day time. Using bridge accident data from Texas and Alabama, Turner (9) developed a probability table that predicts the number of accidents per million vehicles for various combinations of roadway width and bridge relative width (the difference between bridge width and approach roadway width).

A comprehensive analysis of safety specifically at narrow bridges was conducted by TTI for NCHRP (4). Data were collected at 25 bridge sites throughout the United States. Ten important factors related to approach roadway, bridge geometry, traffic, and roadside distractions were identified. Ivey et al. (4) developed a linear model combining these factors and called it the bridge safety index (BSI). The BSI was expressed as the sum of the factors F_1 to F_{10}.

F_1 is a function of clear bridge width. The value of F_1 is determined by looking up the value on the graph in Figure 1. The clear bridge width was measured in the field perpendicular to the centerline of the highway.

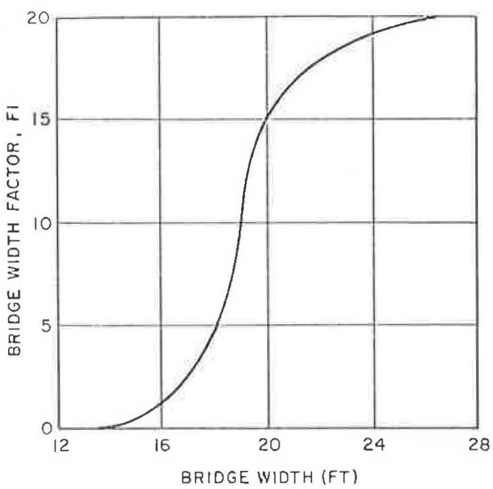

FIGURE 1 Weighting of bridge width factor (F_1).

F_2 is the ratio of the bridge lane width to the approach roadway lane width. The ratio of bridge lane width to the approach lane width, expressed as a decimal, is used in Table 1 to determine F_2. For example, if the ratio is ≥ 1.2, F_2 is given a value of 20.

F_3 is related to the approach guardrail and bridge rail structural factor. The approach guardrail, transition from the approach guardrail, and the bridge rail are inspected to determine if each meets currently acceptable standards. The nomogram shown in Figure 2 is used to convert from the word description to a quantitative value for the F_3 factor reading.

F_4 is related to the ratio of approach sight distance (feet) to 85 percent approach speed (miles per hour). The approach sight distance is measured from the point where the bridge is clearly discernable to the nearer end of the bridge. The 85th percentile approach speed is determined by radar measurement or from any reliable source. The ratio is used in Table 1 to determine the F_4 factor rating.

F_5 is related to the ratio of 100 ft + tangent distance to the curve (feet) ÷ curvature (degrees). This ratio was found for both approaches, and the smaller ratio was used in Table 1 to determine the F_5 factor rating.

F_6 is related to the grade continuity factor (percent) and denotes average grade throughout the bridge zone and the algebraic difference in approach and departing grades. F_6 grade continuity is the sum of the average of the grades approaching and leaving the bridge plus the absolute value of the difference in the two grades. This sum is used in Table 1 to determine the F_6 factor rating (e.g., if this grade continuity is equal to 10, F_6 is evaluated as being equal to 1).

F_7 is related to shoulder reduction (percent) on the bridge compared to approach roadway. At the

TABLE 1 Factors Used to Determine Bridge Safety Index

Evaluated Factor For:	Assigned Factor Rating for F_2 and F_3 in the model				
	0	5	10	15	20
F_2	≤0.8	0.9	1.0	1.1	1.2
F_3	Critical	Poor	Average	Fair	Excellent
	Assigned Factor Rating for F_4 - F_{10} in the model				
	1	2	3	4	5
F_4	≤ 5	7	9	11	14
F_5	≤10	60	100	200	300
F_6	10	8	6	4	2
F_7	100	75	50	25	None
F_8	0.5	0.4	0.3	0.1	0.05
F_9	Wide Discontinuities	Non-Uniform	Normal	Fairly Uniform	Uniform
F_{10}	Continuous	Heavy	Moderate	Few	None

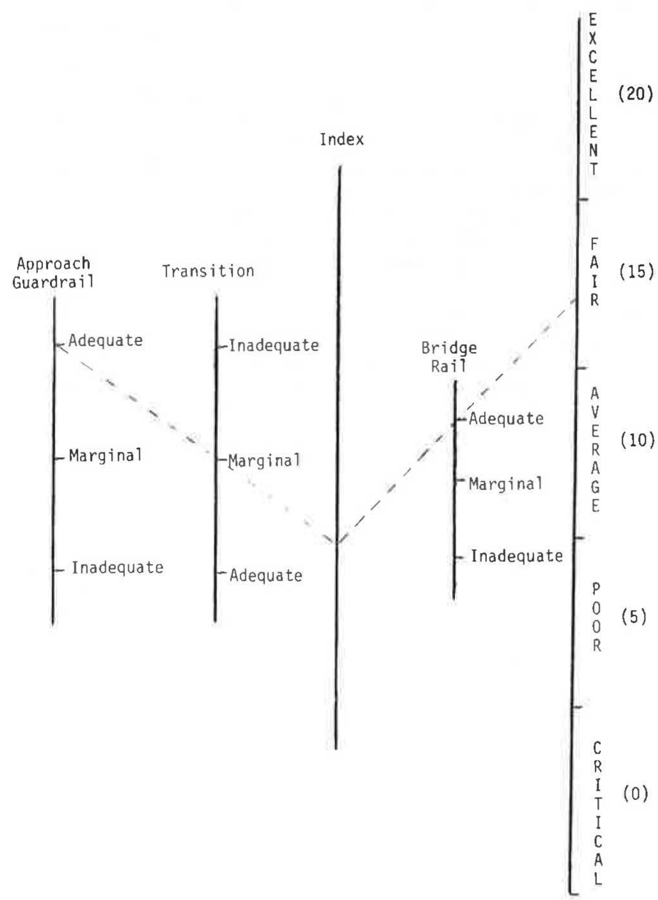

FIGURE 2 Nomogram used to determine F_3.

time lane widths were measured, shoulder width was also measured, and the percentage that the shoulder width on the approach is reduced was noted. This percentage is used in Table 1 to determine the value of factor F_7. For example, if there is a 50 percent reduction, F_7 is given a value of 3.

F_8 is related to the ratio of volume to capacity and is an indirect way of accounting for the number of conflicts on the bridge. Annual average daily traffic (AADT) for the bridge was determined by using a current traffic map or by a physical count. The capacity of the bridge was also determined by taking into consideration that the basic capacity of a two-lane road is 2,000 vehicles per hour and will never exceed 48,000 vehicles per day. The ratio of volume to capacity is used in Table 1 to determine the F_8 factor rating.

F_9 is a factor related to the traffic composition. If the traffic composition includes a relatively high percentage of trucks, narrow bridges can become critically narrow. The traffic mix was estimated as having wide discontinuities or as being nonuniform, normal, fairly uniform, or uniform. This description of the traffic mix was arrived at by inspection and by interviews with local people. The descriptive term is entered into Table 1 to get a value for F_9. For example, if the traffic mix is uniform, F_9 is evaluated as 5.

F_{10} is the distraction and roadside activities factor and its evaluation is similar to F_9. Any unusual activity or environment can distract the driver. Distractions and roadside activities were determined by inspection to be continuous, heavy, moderate, few, or none. The description is entered into Table 1 to evaluate F_{10}.

Ivey et al. (4) considered the first three factors to be 4 times more important than the factors F_4 through F_{10}. The data in Table 1 give the evaluation of factors F_2 through F_{10}. In this first BSI model, the factors F_1, F_2, and F_3 are rated from 0 to 20, whereas the factors F_4 through F_{10} are rated from 1 to 5. The most ideal bridge site conditions would produce a BSI of 95, and critically hazardous sites would have values of less than 20. Ivey et al. (4) suggested that this first BSI model was preliminary and would be improved as more data and information became available from different states. Tseng et al. (10) attempted to improve the first BSI model. They used data collected in 1978 and 1979 by TTI at 78 bridge sites where corrective treatments were recommended in 15 districts of the State of Texas. Information was available for these 78 bridges for the 10 factors F_1 through F_{10} as described previously. For the purpose of statistical analysis, Tseng et al. divided the first three factors by four so that all factors had a maximum value of 5. Tseng et al. (10) added to the analysis two new factors, F_{11} and F_{12}.

F_{11} is a factor that deals with paint markings and is related to the combined effect of centerline, no-passing-zone stripes, edge lines, and diagonal lines on the shoulder of the pavement. Figure 3 shows a way of determining the factor F_{11} in the field. A check mark indicates the conditions of the centerline, edge line, and diagonal lines as adequate, marginal, or inadequate. Figure 4 was then used to arrive at a judgment of the overall condition being excellent, fair, average, poor, or none. The F_{11} factor was evaluated as 5 for excellent, 4 for fair, and so on.

F_{12} is a factor involving warning signs or reflectors and is defined in terms of narrow bridge signs, speed signs, reflectors on the bridge, and black-and-white panels on the bridge ends. Figure 5 shows a way of evaluating the F_{12} factor. The condition of the warning signs or reflectors was determined as excellent, fair, average, poor, or none by observing available slides at TTI. Values ranging from 5 to 1 were given to F_{12}, with excellent condition having a rating of a maximum of 5.

F_{11} Painting Marking ______

	Adequate	Marginal	Inadequate
Centerline or No Passing Zone Stripes		✓	
Edge Line	✓		
Diagonal Line		✓	

FIGURE 3 Evaluation of F_{11}.

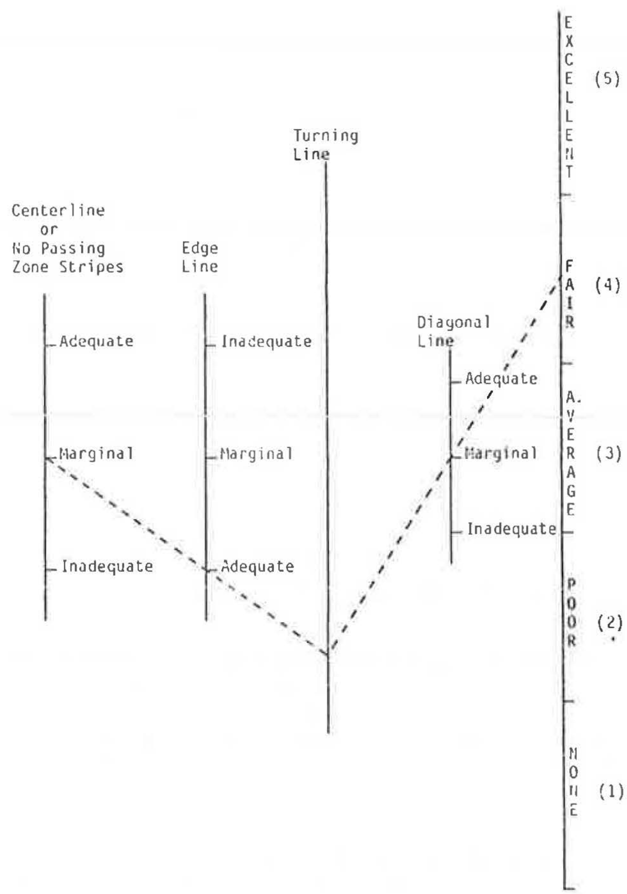

FIGURE 4 Nomogram used to determine F_{11}.

Tseng et al. (10) divided the bridges into two reasonable groups by trial and error and used discriminant analysis to categorize bridges as safe or unsafe and used stepwise regression analysis with accident rate as the response variable to determine the factors that significantly affect bridge safety. They found only 2 of the 12 factors (F_1 and F_6) to be significant.

SwRI (5) conducted an extensive study to evaluate

F_{12} Warning Sign or Reflectors ______	
Excellent	5
Fair	4
Average	3
Poor	2
None	1

FIGURE 5 Evaluation of F_{12}.

the effectiveness of measures for reducing accidents and accident severity at narrow bridge sites. Several statistical analyses were conducted to relate bridge characteristics to accidents. One of their conclusions was that bridge narrowness, as defined in terms of shoulder reduction, had a significant effect on accident rates for two-lane undivided structures.

The literature review yielded a wealth of information regarding the factors that affect safety at bridges in general and narrow bridges in particular. Reduction of the roadway width on the bridge is considered to be the most important factor. Geometric characteristics of the approach road such as alignment, sight distance, type and location of guardrails, transition of guardrails to bridge rails, and traffic factors are all considered important. The researchers were not completely successful in developing a reliable model relating accident rate at the bridges to all of the pertinent features mentioned. Some researchers explained this problem as resulting from variability in accident data. Not only road and bridge features but vehicle and driver characteristics entered into the problem.

RESEARCH APPROACH

The aim of this research effort was to develop an enhanced safety index model for narrow bridges as objectively as possible and to include variables that would have a high degree of contribution to accident rate or that could be readily improved in actual practice and thus make a significant contribution in refining the model. To achieve this objective, readily available information was initially analyzed with modern statistical techniques, and additional data relative to bridge geometrics and accidents were collected as deemed necessary by the results of the preliminary analyses.

The bridge sites considered are the same as those used in the second BSI model. Data were taken from the information file about the 78 narrow bridges in Texas. The accident rate for each selected bridge was obtained from the accident data available at TTI in a computer file. It was gathered from the Texas State Department of Highways and Public Transportation (TSDHPT). Data on 655 accidents at the 78 bridge sites during the 6 years were considered. All of the accident cases at a particular bridge naturally had the same bridge and road characteristics but different accident-related data.

The target was to develop an enhanced safety index model that would classify a narrow bridge into a more-safe or less-safe group. To arrive at a model and predict safety at a bridge site, independent variables such as bridge lane width, length, and approach road width are considered with the dependent variable as the accident rate at a bridge. The accident rate is determined as follows:

$$\begin{aligned} y_1 &= \text{accident rate per 1,000 vehicles} \\ &= 1{,}000\ (\text{Total number of accidents in the 6 years 1974–1979})/6\ (\text{Average daily traffic}) \end{aligned} \quad (1)$$

To choose the proper independent variables to be used for predicting accident rate, correlation and factor analyses were first done. Correlation analysis indicates the measure of association between the variables. Factor analysis investigates the correlation matrix to see which of the variables or variable combinations contributed most to the variability in the data. Factor analysis makes it possible to rearrange or reduce the data to a smaller set of factors or components that may be taken as the source variables accounting for the observed variation. The bridge characteristics of the 78 bridges were used both for factor analysis and for correlation analysis.

Multiple regression is used to analyze the relationship between a dependent variable and a set of independent or predictor variables. It describes and quantifies the relationship between the dependent and independent variables. The collinearity diagnostics in the regression procedure of the Statistical Analysis System (SAS) (11) indicate which of the independent variables are highly correlated. The data from the 655 accidents were used in the stepwise regression procedure, with accident rate as the dependent variable and with bridge and approach roadway characteristics as well as driving environment factors as independent variables. Because the full R^2 procedure is expensive, it was done only on the bridge characteristics on the smaller data set of 78 bridges instead of on the 655 accidents on which stepwise procedure was done. It should be noted that in this research regression is used only to select the relevant variables; it was not used to model safety. The correlation, regression, and factor analyses have been exploratory, and the significant values of different variables are indicative of their importance.

After the independent variables are selected there still remains the problem of classifying given bridges as more safe or less safe. First, the given bridges had to be divided into two distinct groups with different degrees of safety. For this purpose cluster analysis was done with accident rate as the differentiating variable. Once there are two distinct groups in the given data with different degrees of safety, the remaining task is to develop an equation that classifies a bridge into one of the two groups.

To classify a bridge into one of the groups, discriminant analysis or logistic regression can be used. Discriminant analysis yields a function made up of explanatory independent variables that predict the response variable that is categorical and unordered (12). It assumes a multivariate normal distribution for the explanatory variables, which makes it sometimes difficult to apply, although the procedure is considered to be robust with a large sample. An alternative method of classification is by logistic regression in which the assumption of normality of variables is not required (13). The cluster analysis is used to get two distinct groups on which discriminant and logistic procedures are applied to develop an equation that will yield a safety index.

COLLECTION OF ADDITIONAL DATA

Several statistical techniques (such as correlation analysis) were applied to the data that were readily available at TTI to understand the relationships between the variables, which did not reveal much useful information. The researchers believed that more data needed to be collected. The sources of additional information about the 78 sample bridges were (a) the Bridge Inventory, Inspection, and Appraisal Program (BRINSAP) files of TSDHPT (14), (b) the files on bridge and road characteristics available at TTI, and (c) the accident files for the years 1974 through 1979, also available at TTI.

By using BRINSAP microfiche, information was obtained as to whether the bridge was rural or urban, whether the bridge had shoulders, and the length of each bridge. Actual road width, actual shoulder reduction in feet, actual shoulder reduction in percent, actual grade continuity, sight distance, relative width as a difference of bridge lane width and approach lane width, ADT, and 85th percentile approach speed were tabulated manually from the files available at TTI. Factors F_2 and F_8 were ratios. They were each transformed into a nonratio form, and the transformed variables were considered as factors. Instead of the ratio factor F_2, the difference between its numerator and denominator was considered as the variable of relative width. F_8, which is a ratio of average daily traffic to capacity, was considered as ADT because the capacity was the same for most of the bridges. This was done because several statistical procedures do not assume that the independent variables are ratios.

Additional data concerning each accident, which involved environmental factors such as light condition, road condition, surface condition, and time of the accident, were collected. Also collected was information about traffic control, alignment, curvature, number injured, and number killed in each of the 655 accidents during the years 1974 through 1979 on the 78 bridge sites. The relevant information collected about the accidents and the bridge characteristics were made into a data file to be used for statistical analysis. On the data set constructed initially, several statistical procedures were considered that would give an insight into the interrelationships.

BRIEF DESCRIPTION OF STATISTICAL METHODS USED

A brief description of the statistical methods used in the analyses is presented here. Analysis of the results is given in a later section.

Testing Statistically for Effect of Type of Bridge

Before other analyses were done, information about the road type (rural or urban) was gathered by using BRINSAP files (14). There were 69 bridges in rural areas and 9 bridges in urban areas. An analysis of variance done with accident rate as the dependent variable and type of bridge as the independent variable did not indicate significant differences.

Testing Statistically for Effect of Sidewalks

The BRINSAP file (14) also yields information about whether the bridges have sidewalks or not. Often sidewalks may imply a curb, and the effect of curb or no curb on accidents was of interest. There were 12 bridges with sidewalks and 66 without sidewalks. An analysis of variance did not indicate a significant difference in accident rate between bridges with sidewalks and bridges without sidewalks.

Stepwise Regression and R-Square Procedures

Polynomial and logarithmic models were tried at the outset but were not found particularly beneficial for this research. Linear-regression analyses, including the stepwise and R-square procedures of SAS, were used to relate accident rate to bridge geometrics. The coefficient of multiple determination (R^2) measures the total variation in the dependent variable explained by the regression model. The R^2 varies between 0 and 1, and higher values of R^2 indicate a better fit of the model. The stepwise procedure of SAS with the maximum R^2 option is simply a one-variable-at-a-time model selection procedure, and at each step a variable that contributes most to the R^2 is chosen. The R-square procedure of SAS (11) gives all possible regressions but is expensive when there are many variables. The stepwise procedure gives a subset of the full R^2 procedure. In this research stepwise regression was used with the total data including the accident data, and a full R^2 procedure was done on bridge geometrics, conditions, and approach characteristics. Results are discussed in detail in the next section.

Multicollinearity and Variance Inflation Factors

Multicollinearity (15) is a high degree of linear relationship among independent variables that makes interpretation of partial regression coefficients difficult in regression analysis. One formal method of detecting the presence of multicollinearity is by means of variance inflation factors (VIFs). These factors indicate how much the variances of the estimated regression coefficients are inflated. A VIF in excess of 10 indicates multicollinearity. However, for these data, no multicollinearity was observed.

Factor Analysis

Factor analysis (16,17) is a term for a mathematical and statistical technique that is designed to investigate the nature of the relationship between variables in a specified set. The basic problem is to determine whether the n variables in a set exhibit patterns of relationships with one another, such that the set can be broken down into m subsets, each consisting of a group of variables tending to be more highly related to others within the subset than to those in the other subsets.

The single most distinctive characteristic of factor analysis is its data-reduction capability. When the correlation coefficients in a set of variables are known, factor analysis techniques enable researchers to see whether some underlying pattern of relationship exists, such that the data may be rearranged or reduced to a smaller set of factors or components that may be taken as source variables accounting for the observed interrelations in the data. Factor analysis consists of two more steps after the initial step of calculating the appropriate measurement of association between the relevant variables.

The next step in factor analysis is to explore the data-reduction possibilities by constructing a set of new variables on the basis of interrelations exhibited in the data. The new variable may be defined as a mathematical transformation of the original data. A principal component analysis performs the extraction of initial factors. The principal component analysis is a method of transforming a given set of variables into a new set of composite variables or principal components that are orthogonal or uncorrelated to each other and that explain as much of the total variation as possible.

The last step is to rotate the factors to a terminal solution in order to achieve simple factor patterns to ease the interpretation of results. Complex overlapping factor loadings are not useful. The matrix of loadings is rotated to a mathematically equivalent matrix that has only a few high loadings for each factor and low loadings for the rest. By inspecting the high loadings of the factors, the importance of the original variables can be determined. Loadings can be considered as similar to correlation coefficients.

In this study the factors are explored to see which of the original variables contributed to most of the variability in the data and eliminate those that did not contribute. The first four factors that accounted for 66 percent of the total variability did not receive significant contribution from factors F_3, F_{10}, and F_{11}.

Cluster Analysis

Cluster analysis (18,19) is an exploratory procedure to understand the nature of data. It is a statistical technique that objectively groups together objects or variables based on their similarities and differences. In most practical applications of cluster analysis the investigators know enough about the problem of distinguishing good groupings from bad. In this research the objective is to group the bridges into more-safe and less-safe groups as objectively as possible. These groups will be used to develop a criterion to classify a new bridge into one of the groups.

Because type of bridge or having a sidewalk did not make a difference in accident rate, all the bridges were considered together for cluster analysis in this research. Cluster analysis was attempted by using the cluster procedure and Fastclus procedure of SAS. The cluster procedure, designed to help identify clusters of observations that have similar attributes, performs a hierarchial cluster analysis. The technique begins by forming one cluster for each observation in the analysis. The two closest clusters are combined into a single cluster and so on. The Fastclus procedure is designed for disjoint clustering of a large data set and can find good clusters with only two or three passes of the data.

The analysis yielded a reasonable division of the 78 bridges into two clusters that had 52 bridges in the more-safe (less accident-prone) group and 26 in less-safe (more accident-prone) group. The two groups yielded by the cluster procedure were significantly different with regard to accident rate. All bridges that had an accident rate of more than 0.41 (1,000 vehicles per year) classified into the less-safe group.

Discriminant Analysis

Discriminant analysis (20) and classification are multivariate techniques that deal with separation of distinct sets of objects and with allocating a new object into previously defined groups. As opposed to cluster analysis, a classification model is used to categorize an object on the basis of a profile of its characteristics. Fisher (21) developed a solution for the two-group case that is known as the linear discriminant analysis. More generally, when $\underline{Z}$ ($P \times 1$) denotes a vector of observed characteristics of an object that belongs to exactly one of the mutually exclusive populations $\Pi_1, \Pi_2, \ldots, \Pi_k$, then discrimination models are developed to classify the object into one of the population groups.

For the two-group case, when the populations denoted by Π_1 and Π_2 are multivariate normal with mean vectors $\underline{\theta}_1$ and $\underline{\theta}_2$ and covariance matrices are Σ_1 and Σ_2 and the costs of misclassification are given by C_{12} and C_{21}, the rule is to classify $\underline{Z}$ into Π_1 if

$$[(\underline{Z} - \underline{\theta}_2)'\Sigma_2^{-1}(\underline{Z} - \underline{\theta}_2) - (\underline{Z} - \underline{\theta}_1)'\Sigma_1^{-1}(\underline{Z} - \underline{\theta}_1)] \geqslant 2\ell n \; [(|\Sigma_1|/|\Sigma_2|)^{1/2}(P_2/P_1)(C_{12}/C_{21})] \tag{2}$$

If $\Sigma_1 = \Sigma_2$, a linear discriminant function is obtained. Otherwise a quadratic discriminant function is the result. If a linear discriminant function is to discriminate effectively between two groups, then it is expected that their mean values are far apart relative to the variation within the groups. This distance from each individual observation to each of the group centroids is used as a criterion for assigning observations to a particular group.

After a discriminant function is developed, it needs to be evaluated to determine how well it is classifying. One of the ways of gauging the efficiency of the function developed is to find the apparent error rate that is obtained by reclassifying the data used in developing the equation and noting the number of misclassifications. Fewer misclassifications indicate a better discriminatory function.

In this research the objective is to develop a model to classify a given bridge into a more-safe or less-safe group when bridge geometrics and characteristics are known. In this situation discriminant procedures can be applied to the two groups in question if most of the assumptions hold true. Discriminant analysis was attempted by using the Discrim procedure of SAS because it is a robust procedure. Results of discriminant analyses are presented in the next section.

Logistic Regression

Sometimes the dependent variable of interest has only two possible outcomes and therefore can be represented by an indicator variable taking on values 0 and 1. A dependent variable taking on the value of 0 and 1 is considered to be binary or dichotomous and is a categorical variable with two categories, in which case logistic models are of interest.

Logistic regression (22) is often used in survival data analysis. If a particular drug is used on n cases, the survival cases can be considered as successes and the others as failures. Let $Y_i = 1$ when the individual case is a success and $Y_i = 0$ otherwise. Let $x_{i1}, x_{i2}, \ldots, x_{im}$ be the characteristics of the individual case, the variables being qualitative or quantitative. The dependent variable Y_i is to be related to the independent variables x_i. In a situation of successes and failures, the linear logistic model is applicable. The model is given by

$$P_i = \exp\left(\sum_{j=1}^{p} \beta_j X_{ij}\right) / [1 + \exp(\Sigma \beta_j X_{ij})] \tag{3}$$

where $P_i = \text{Prob}(Y_i = 1)$ or $1 - P_i = \text{Prob}(Y_i = 0)$. The logarithm of the quantity $[P_i/(1 - P_i)]$ is a simple linear function of the x_j's given by

$$\log [P_i/(1 - P_i)] = \sum_{j=1}^{p} \beta_j X_{ij} \tag{4}$$

P_i yields the probability of success, which is the probability of being a safe bridge in this research. Harrell's LOGIST procedure (23) is used to develop the final logistic model.

ANALYSES AND RESULTS

Some of the most important results from stepwise regression and discriminant and logistic analysis are discussed in this section.

Stepwise Regression

A stepwise regression procedure was applied with 21 variables in order to determine their contribution to R^2, although it is not the only criterion on which variable selection is made. The data in Tables 2 and 3 give the regression coefficients and steps in order in which the variables entered the regression model, which yielded an R^2 of 0.65.

From the data in Table 3 it can be observed that the first three variables to enter--length, F_6, and bridge width (in that order)--contributed well to R^2 and accounted for about 49.9 percent of the variability. Sight distance was the fourth variable to enter and is significant with the sign of regression coefficient being positive, which implies that the accident rate increases with sight distance, which is not expected. This occurred because there were a lot of missing data for the sight-distance variable. Hence this variable could not be considered in the final model. Environmental factors of light condition, road condition, surface condition, and weather condition did not prove to be significant in the stepwise model. Alignment, traffic control, and curvature were also considered in the model but they did not contribute significantly. Regression was considered here only on an exploratory basis but not for developing the final model.

TABLE 2 Regression Coefficients of Stepwise Regression

Variable	Value	Prob>F[a]
Intercept	0.89847349	
Bridge width	-0.03857594	0.0001
ADT	0.00867674	0.0006
Speed	0.00705234	0.0012
Length	0.00014109	0.0001
F_6	-0.07169020	0.0001
F_9	0.00784998	0.3476
L (light)	0.00740584	0.6017
W (width)	-0.01491271	0.5709
R (road)	0.00487835	0.8376
S (surface)	0.01102201	0.6977
F_2	0.00173885	0.9195
F_3	-0.00511092	0.4982
F_7	-0.00764914	0.3221
F_{10}	0.04080974	0.0001
F_{11}	-0.01802289	0.0426
F_{12}	0.04170525	0.0004
Relative width	0.04365003	0.0063
Sight distance	0.00004724	0.0001
Alignment	-0.02755675	0.1114
Traffic control	0.00236779	0.6330
Curvature	0.01818683	0.4758

[a]Indicates the probability of obtaining this value of F or one larger by chance alone. Smaller numbers indicate higher significance of the coefficients (not obtained by chance).

TABLE 3 Steps of Stepwise Regression

Step	Variable Entered	Variable Replaced By	R^2
1	Length	-	0.268
2	F_6	-	0.450
3	Bridge width	-	0.499
4	Sight distance	-	0.591
5	F_{10}	-	0.606
6	F_{12}	F_{10} replaced by F_2	0.623
7	F_{10}	-	0.632
8	ADT	F_2 replaced by relative width	0.642
9	Speed	-	0.648
10	F_{11}	-	0.651
11	Alignment	-	0.653
12	F_9	-	0.653
13	F_7	-	0.654
14	F_3	-	0.654
15	Curvature	-	0.654
16	Traffic control	-	0.654
17	Light condition	-	0.654
18	Road condition	-	0.654
19	-	Road condition replaced by surface condition	0.655
20	Road condition	-	0.655
21	F_2	-	0.655

Discriminant Analysis

In the groups obtained by the cluster procedure of 26 less-safe bridges and 52 more-safe bridges, a discriminant analysis conducted by the Discrim procedure of SAS was applied. However, some of the variables were found to be non-normal when tested by the univariate procedure of SAS. The discriminant analysis, which assumed that the independent variables are normally distributed, was attempted on an exploratory basis because it was a robust procedure. The linear discriminant functions developed with various variable combinations either classified poorly or did not result in the appropriate coefficient signs. Hence logistic regression was considered because it does not assume normality.

Logistic Regression

Logistic regression was considered because it yields the degree of safety directly, is a good model for the safety index, and the variables need not have a normal distribution. Several models were tried with several combinations of variables. In order to be able to make a meaningful interpretation of the model, the signs of the coefficients of the variables in the model must be appropriate. After examining several, the following was accepted as the final model (Table 4):

$$\text{Probability of safety (safety index)} = \exp(y)\,[1 + \exp(y)] \quad (5)$$

where

$$y = [-1.790 + 0.441\,(\text{bridge width}) - 0.107\,(\text{average daily traffic}) - 0.246\,(\text{speed}) - 0.001\,(\text{length}) + 0.954\,(F_9) + 0.567\,(F_6) + 0.332\,(F_7)] \quad (6)$$

TABLE 4 Logistic Model Coefficients

Variable X_i	β	Probability	Partial R
Intercept	-1.78999897	0.3936	
Bridge width	0.44123886	0.0000	0.294
ADT	-0.10753546	0.0055	-0.080
Speed	-0.24633482	0.0000	-0.207
Length	-0.00101675	0.0000	0.260
F_9	0.95457213	0.0000	0.204
F_6	0.56696522	0.0000	0.212
F_7	0.33232235	0.0014	0.095

The equation obtained has an R (akin to the multiple correlation coefficient) of 0.62. The fraction of concordant pairs of predicted probabilities and responses is 0.91 out of a maximum possible value of 1.0. The rank correlation (which is similar to Kendall's correlation coefficient) between predicted probability and response is 0.81, which indicates the goodness of the model. All the y variables are highly significant. Individual R statistics (partial R's) computed for the logistic model provide a measure of the contribution of the variables and are not to be confused with the regression coefficients. From partial R's given in Table 4, it is apparent that according to the logistic model, bridge width is the most important variable in the determination of the safety index, and the second most important variable is length. ADT and F_7 are less-important variables.

By taking randomly one individual observation and changing each of the independent variables by 10 percent, it was observed that speed is the most sensitive to changes and improvements (Table 5). Percent change in the safety index was highest for speed where percent change for all other variables was a constant 10 percent.

TABLE 5 Sensitivity Index

Variable	Sensitivity Rank	Percentage Change of Variable (+)	Sensitivity Index[a] (%) Plus (+)	Sensitivity Index[a] (%) Minus (-)
Speed	1	10		1.81
Bridge width	2	10	0.71	
F_9	3	10	0.23	
F_6	4	10	0.05	
ADT	5	10		0.04
F_7	6	10	0.03	
Length	7	10		0.01

[a]Sensitivity index = (Percentage change in safety index) ÷ (Percentage change of variable).

Discussion of Regression, Discriminant, and Logistic Models With the Independent Variables Selected in the Final Model

By using the data of 655 accidents and the bridge and roadway characteristics of the 78 bridges, a multiple regression model was fitted to the independent variables bridge width, ADT, length, speed, F_6, F_7, and F_9. This model yielded an R^2 of 0.52 with the dependent variable as accident rate. All the variables were significant, except F_7 and F_9.

A linear discriminant model with all of the same variables classified 42.3 percent of the unsafe bridges correctly. A quadratic discriminant model with the same variables classified 65.38 percent of the unsafe bridges as unsafe correctly and 88.46 percent of the safe bridges correctly. The logistic regression model with the same variables had a rank correlation (between predicted probabilities and responses) of 0.81 out of a maximum possible value of 1.0. The logistic model yielded directly a safety index that is sensitive to changes in the variable and hence was concluded to be the best possible model. The safety index is the number that is bounded between 0 and 1.

SUMMARY AND CONCLUSIONS

The findings of this research are summarized as follows.

1. An enhanced BSI model was developed, bearing in mind at all times the experience and expertise of the bridge engineers, which resulted in the first model. This new model is considered enhanced because (a) the bridges were divided into more-safe and less-safe groups scientifically with the use of cluster analysis and not arbitrarily as in previous research; (b) the model is parsimonious (or has fewer variables and terms) with only 7 independent variables as against 12 variables in the previous BSI model; (c) even with only approximately half the number of variables it yields more than twice the R^2 compared to the previous model; (d) the safety index developed yields a higher and more significant correlation of -0.53 with accident rate as compared with the previous models; and (e) the logistic model used does not need the assumption of normality of variables as compared with the previous model, which is important because most of the variables are found to be non-normal in these data.
2. The goodness of the final model is apparent from the fact that it yields the fraction of concordant probabilities and responses as 0.91 compared to a maximum possible value of 1.00 and has a high rank correlation of 0.81.
3. The developed model yields a safety index directly when the relevant factors are known and can be readily used to establish priorities for improvement or repairs of bridges.
4. The important variables related to accident rate appeared to be bridge width, length of the bridges, speed, traffic mix, and grade continuity. The final model adopted gives the probability of safety of a bridge and includes the variables F_6 (grade continuity), F_7 (shoulder reduction), and F_9 (traffic mix) as well as bridge width, speed, length, and ADT.
5. The model is sensitive to improvements and results in higher probabilities of safety when the factors mentioned in 4 are improved.
6. To improve the safety as per the model, speed limits should be decreased by posting appropriate signs at and before the bridge or consideration may be given to use of rumble strips or similar devices for speed reduction. The other factors--bridge width, length, shoulder reduction, grade continuity, ADT, and traffic mix--are not easy to change; nevertheless, it is possible to make some improvements in these factors.
7. The model yields a safety index that can be used to identify a potentially hazardous narrow bridge.

FUTURE RESEARCH

To realize maximum benefit from an improved safety index model, research is suggested as follows.

1. More comprehensive accident data for each bridge site, as well as actual measurements of all factors involved in the F-factors instead of ratios, should be collected. This would facilitate the anal-

ysis by enabling the use of the constituent variables of a ratio directly.

2. If more information can be collected on injuries incurred during accidents, it may be possible to relate the bridge safety index not only to the accident rates but also to the severity of accidents on bridges.

3. It is necessary to obtain another carefully taken sample of bridges to validate the conclusions and the model developed in this research. In addition, more comprehensive data should be collected about each accident.

4. A cut-off point of probability or norm was not arrived at in this research, although a low probability of safety or low safety index indicates a hazardous bridge. This is another area of interest.

5. Speed was determined to be one of the most influencing factors of the safety index. The techniques suggested to reduce speed, such as advisory signs and rumble strips, were only preliminary. There is much scope to investigate this aspect.

ACKNOWLEDGMENT

This research was conducted under the Cooperative Highway Research Program funded by FHWA and TSDHPT. Their support is gratefully acknowledged.

REFERENCES

1. G.H. Oglesby and R.G. Hicks. Highway Engineering. Wiley, New York, 1982.
2. Better Roads Exclusive Bridge Inventory. Better Roads, Vol. 50, No. 5, 1980, pp. 33.
3. Narrow Bridges--Driver Dilemmas. Hearing Before the Subcommittee on Investigations and Review of the Committee on Public Works, 93rd Congress, U.S. Government Printing Office, June 1973.
4. D.L. Ivey, R.M. Olson, N.E. Walton, G.D. Weaver, and L.W. Furr. Safety at Narrow Bridge Sites. NCHRP Report 203. TRB, National Research Council, Washington, D.C., June 1979, 63 pp.
5. K.K. Mac and L.R. Calcote. Accident Analysis at Highway Narrow Bridge Sites. Final Report. FHWA and NHTSA, U.S. Department of Transportation, Feb. 1983.
6. W.C. Johnson. An Analysis of Driver Behavior at Narrow Bridges. Ph.D. dissertation. Texas A&M University, College Station, 1977.
7. M.S. Raff. Interstate Highway Accident Study. Bull. 74. HRB, National Research Council, Washington, D.C., 1953, pp. 18-43.
8. J. Behnam and J.G. Laguros. Accident and Roadway Geometrics at Bridge Approaches. Public Works, Vol. 104, No. 4, 1973, pp. 66-70.
9. D.S. Turner. A Technique to Select Bridge Safety Treatments. Ph.D. dissertation. Texas A&M University, College Station, 1980.
10. K.-H. Tseng, R.L. Lytton, and B.D. Garrett. An Evaluation of Bridge Safety Index in Texas. Tech. Memorandum. Texas Transportation Institute, Texas A&M University, College Station, Jan. 1983.
11. A.J. Barr, J.H. Goodnight, J.P. Sall, and J.T. Helwig. A User's Guide to SAS 79. SAS Institute, Inc., Sparks Press, Raleigh, N.C., 1979.
12. R.A. Eisenbeis and R.B. Avery. Discriminant Analysis and Classification Procedures. Lexington Books, Lexington, Mass., 1972.
13. S.J. Press and S. Wilson. Choosing Between Logistic Regression and Discriminant Analysis. Journal of American Statistical Association, Vol. 73, 1978, pp. 699-705.
14. Bridge Inventory, Inspection, and Appraisal Program (BRINSAP), Manual of Procedures. Texas State Department of Highways and Public Transportation, Austin, 1977.
15. D.C. Montgomery and E.A. Peck. Introduction to Linear Regression Analysis. Wiley, New York, 1982.
16. N.H. Nie, C.H. Hull, J.G. Jenkins, K. Steinbrenner, and D.H. Bent. Statistical Package for the Social Sciences. McGraw-Hill, New York, 1975.
17. H.H. Harman. Modern Factor Analysis. University of Chicago Press, Chicago, 1960.
18. R.A. Johnson and O.W. Wichern. Applied Multivariate Statistical Analysis. Prentice Hall, Englewood Cliffs, N.J., 1982.
19. M.R. Anderberg. Cluster Analysis for Applications. Academic Press, New York, 1973.
20. S.J. Press. Applied Multivariate Analysis. Holt, Rinehart, and Winston, Inc., New York, 1979.
21. R.A. Fisher. Statistical Methods for Research Workers. Hasney Publishing Company, Inc., New York, 1948.
22. E.T. Lee. Statistical Methods for Survival Data Analysis. Lifetime Learning Publications, Belmont, Calif., 1980.
23. F. Harrell. Logist Procedure: Support Type A. Duke University Medical Center, Durham, N.C., 1983.

Publication of this paper sponsored by Committee on Operational Effects of Geometrics