TECHNICAL REPORT STANDARD TITLE PAGE

1. Report No. FHWA/TX/84/26+233-2F	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle AN IMPROVED BRIDGE SAFETY INDEX FOR NARROW BRIDGES		5. Report Date August 1983 6. Performing Organization Code
7. Author(s) R. B. V. Gandhi and R. L. Lytton		8. Performing Organization Report No. Research Report 233-2F
9. Performing Organization Name and Address Texas Transportation Institute Texas A&M University System College Station, Texas 77843		10. Work Unit No. 11. Contract or Grant No. Study 2-18-78-233
12. Sponsoring Agency Name and Address Texas State Department of Highways and Public Transportation Transportation Planning Division P.O. Box 5051; Austin, Texas 78763		13. Type of Report and Period Covered Final - September 1977 August 1983 14. Sponsoring Agency Code

15. Supplementary Notes

Work done in cooperation with FHWA, DOT.
Study Title: Priority Treatment of Narrow Bridges

16. Abstract

In this report, a new bridge safety index is developed based upon an extensive statistical study of accident data on 78 bridges. A total of 655 accidents were recorded at these bridges over the six-year period between 1974 and 1979. Cluster analysis showed that 26 of the bridges could be grouped into a "less-safe" class and the remaining 52 bridges culstered into the "more-safe" group. Assigning a score of Ø to the less-safe group and 1 to the more-safe group, logistic regression analysis was used to develop an equation for the probability that a bridge is safe. The R^2 of the equation is Ø.53 and includes the following variables: bridge width, bridge length, ADT, vehicle speed, grade continuity, shoulder reduction, and traffic mix. A sensitivity analysis shows that bridge safety may be improved most by reducing vehicle speed and secondly by increasing the bridge width. Recommendations are made for the data that should be collected and the number of bridges that should be represented in future studies.

The Bridge Safety Index developed in this study is considerably better than all that have been proposed previously since it is based objectively upon accident rates and has a much better correlation with accident rates than any previously proposed Bridge Safety Index. Consistent use of this index in setting funding priorities for projects to improve bridge safety should result in a reduction of accident rates, property damage, injuries and fatalities at bridges.

17. Key Words Narrow Bridges, Bridge Safety Index, Funding Priorities	18. Distribution Statement No Restricitions. This document is available to the public through the National Technical Information Service, Springfield, Virginia 22161

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	146	

Form DOT F 1700.7 (8-69)

AN IMPROVED BRIDGE SAFETY INDEX FOR NARROW BRIDGES

by

R. B. V. Gandhi
R. L. Lytton

Research Report Number 233-2F

Priority Treatment of Narrow Bridges
Study No. 2-18-78-233

conducted for the
State Department of Highways
and Public Transportation

in cooperation with the
U. S. Department of Transportation
Federal Highway Administration

by the

TEXAS TRANSPORTATION INSTITUTE
Texas A&M University
College Station, Texas

August 1983

ABSTRACT

In this report, a new bridge safety index is developed based upon an extensive statistical study of accident data on 78 bridges. A total of 655 accidents were recorded at these bridges over the six-year period between 1974 and 1979. Cluster analysis showed that 26 of the bridges could be grouped into a "less-safe" class and the remaining 52 bridges clustered into the "more-safe" group. Assigning a score of Ø to the less-safe group and 1 to the more-safe group, logistic regression analysis was used to develop an equation for the probability that a bridge is safe. The R^2 of the equation is Ø.53 and includes the following variables: bridge width, bridge length, ADT, vehicle speed, grade continuity, shoulder reduction, and traffic mix. A sensitivity analysis shows that bridge safety may be improved most by reducing vehicle speed and secondly by increasing the bridge width. Recommendations are made for the data that should be collected and the number of bridges that should be represented in future studies.

The Bridge Safety Index developed in this study is considerably better than all that have been proposed previously since it is based objectively upon accident rates and has a much better correlation with accident rates than any previously proposed Bridge Safety Index. Consistent use

of this index in setting funding priorities for projects to improve bridge safety should result in a reduction of accident rates, property damage, injuries and fatalities at bridges.

SUMMARY

A Bridge Safety Index should be able to indicate reliably which bridges are more likely to cause accidents and what can be improved to reduce the accident rate at those bridges. This report documents the development of an objective bridge safety index using an extensive statistical study of accident data on 78 bridges. A total of 655 accidents were recorded at these bridges over a six-year period between 1974 and 1979. A statistical technique known as cluster analysis showed that 26 of the bridges clustered into a "less-safe" class and the remaining 52 bridges clustered into the "more-safe" group.

The study then investigated a number of characteristics of the bridge, the approach roadway geometrics, the traffic, and the driving environment including signs, pavement markings, and distractions to determine which of these contributed to the assignment of a bridge into either the "more-safe" or "less-safe" bridge class. A total of 20 such characteristics were studied some of which must be rated subjectively but tables, graphs, and nomographs are provided for that purpose.

Assigning a score of 0 to the less-safe group and 1 to the more-safe group, logistic regression analysis was used to develop an equation for the probability that a bridge is

safe. The final equation includes the following variables: bridge width, bridge length, average daily traffic (ADT), vehicle speed, the degree of continuity of the approach and departure grade, the reduction of the shoulder width from the approach roadway to the bridge, and the traffic mix, which is primarily based upon the percent trucks in the traffic stream. The new equation has a correlation coefficient (R) with accident rate of 0.53 which is considerably better than that of the Bridge Safety Index developed by TTI and reported in NCHRP Report 203, "Safety at Narrow Bridge Sites." The latter equation only had a correlation coefficient of 0.23. No perfect correlation is expected since accidents also depend upon the driver, which was not considered in this study.

A sensitivity analysis shows that bridge safety may be improved most by reducing vehicle speed, secondly by increasing bridge width, and thirdly by reducing the number of trucks in the traffic stream.

It is considered possible to improve still more on the Bridge Safety Index that is developed in this report if more data are available. It is estimated that over 200 bridges would be needed for that purpose. Recommendations are also made on the additional data that should be collected in the event that it is considered desirable to improve the Bridge Safety Index of this report.

The Bridge Safety Index developed in this study is considerably better than all that have been proposed previously since it is based objectively upon accident rates and has a much better correlation with accident rates than any previously proposed Bridge Safety Index. Consistent use of this index in setting funding priorities for projects to improve bridge safety should result in a reduction of accident rates, property damage, injuries and fatalities at bridges.

IMPLEMENTATION STATEMENT

The Bridge Safety Index presented in this report can be used together with project cost to determine a benefit-cost ratio for proposed bridge safety improvement projects. Because this Bridge Safety Index is based objectively on accident rate data, consistent use of this index by the Texas State Department of Highways and Public Transportation to set funding priorities for projects to improve bridge safety should result in a reduction of accident rates, property damage, injuries and fatalities at bridges.

DISCLAIMER

TABLE OF CONTENTS

TABLE OF CONTENTS (cont'd)

TABLE OF CONTENTS (cont'd)

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

A bridge may require replacement or rehabilitation for a variety of reasons including structural, geometric, and functional obsolescence, all of which may contribute to a reduction in safety for the driving public. Oglesby and Hicks (1) state that on the Federal-aid highway system, of the 240,000 bridges inventoried, there are about 9000 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. Engineering News Record (2) reported that one in six U.S. highway bridges is deficient and tabulated the percentage of deficient bridges in each state. Iowa led the list with 39%. U.S. News reported in an article (3) that weak bridges are a growing hazard on the highways. Several articles appeared on highway hazards in the Better Roads Journal (4, 5, 6). The latest Better Roads inventory showed that there are nearly 90,000 substandard bridges in the U.S. To bring even a fraction of the bridges up to the modern design standards involves billions of dollars. Until recently, no federal aid was available to rehabilitate all of these deficient bridges. However, after the enactment of the Surface Transportation Act of 1978, there was a dramatic increase in funds for highway bridge replacement and rehabilitation programs (7). For the fiscal years 1979 to 1982, $4.2 billion were authorized for bridge rehabilitation. The Better Roads Journal (8, 9) also reported on funding increases and rehabilitation programs undertaken in different states.

The American Association of State Highway and Transportation Officials (AASHTO) published several manuals and documents to standardize the design and maintenance of the highway system (10, 11, 12). All of the new highways and bridges are generally built to these modern standards. Special committees of the AASHTO have dealt with bridge safety and maintenance problems (13, 14, 15). Since the Highway Safety Act was passed by the U.S. Congress in 1973, the U.S. Department of Transportation placed high priority on highway and bridge safety. The National Highway Safety Needs Study (16, 17) laid emphasis on improved guard rail design and bridge widening for improving safety at bridges. They also initiated periodic bridge inventory and appraisal programs in all states on a standardized basis (18). The highway safety programs are evaluated periodically and reported to Congress (19).

In addition to structurally unsound bridges, there are several other bridges in the U.S. which are structurally adequate but narrow in width compared to the approach roadway width. Narrowing of the roadway on the bridge creates a significant accident potential for the driving public. These accidents result from the impact of vehicles on bridge abutments, approach guard rails and bridge railings and from collisions with oncoming vehicles due to 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 (20) 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 has been an ongoing research activity of the Texas Transportation Institute (21) and a Bridge Safety Index (BSI) has been formulated to distinguish between "more safe" and "less safe" bridges on the basis of several factors related to the bridge and the approach roadway. 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 which affect safety, and analysis using modern statistical techniques such as cluster, discriminant and factor analysis and multivariate regression were used.

The procedure that was followed in developing a Bridge Safety Index required seven basic steps:

1. Review the literature on bridge safety and determine what other studies have shown to be significant

contributing factors to enhance bridge safety.

2. Examine and make critical analysis of the readily available information as tabulated at the Texas Transportation Institute (TTI) with a view to selecting relevant and important variables to be included in the Bridge Safety Index (BSI) model;
3. Collect additional data on bridge geometrics and accidents to improve the existing BSI model;
4. Analyze the final data set to ascertain the potential relationships between variables;
5. Develop an enhanced safety model to classify narrow bridges as "more safe" or "less safe" when certain bridge geometrics and conditions are known. The model should be developed as objectively as possible using variables that contribute significantly to accident rate and those that can be readily improved to increase bridge safety;
6. Obtain a Bridge Safety Index from the model that can be used to identify a potentially hazardous narrow bridge;
7. Develop a methodology that can be used in general for obtaining safety models for other types of bridges, as well as other roadside structures that effectively narrow the roadway.

The consistent use of the Bridge Safety Index in helping to set priorities for projects to improve bridge safety should reduce accident rates, minimize property damage and human injury, and save lives.

CHAPTER 2

LITERATURE REVIEW

A significant amount of research has been conducted in the United States on highway related accidents with a view to improve public safety on the highway system. Although the United States has one of the best highway systems in the world, some bottlenecks still exist in the form of narrow bridges, which pose a danger to high speed traffic. The literature review presented here relates to the safety problems at narrow bridges and what is being done to solve them.

Defining a Narrow Bridge

A survey of narrow bridges (21) conducted by the Texas Transportation Institute (TTI) showed that different States had different criteria for defining narrow bridges. The questionnaire summary indicates that a large number of state bridges, 7,211 in number, are considered narrow if they are 18 feet or less in width, and a large number of city and municipal bridges, 7,905 in number, are considered deficient if they are 16 feet or less. TTI drew the following conclusions from the survey:

1. The lower limit for two-way operations generally appears to fall in the range of 16 to 20 feet.

2. In general, a bridge is considered narrow if the clear roadway width on the bridge is equal to or less than the approach roadway width.

Southwest Research Institute of San Antonio (SWRI) in its study on narrow bridges (22) defined narrow bridges as follows:

1. One-lane, 18 feet or less in width,
2. Two-lane, 24 feet or less in width,
3. Total approach width greater than total bridge width (curb to curb) and bridge shoulder is less than 50% of approach roadway shoulder (i.e., >50% shoulder reduction).

According to Johnson (23), any bridge that changes driver's behavior with regard to speed or lateral positioning of the vehicle can be considered "narrow". He developed equations to estimate driver behavior in terms of bridge perceptual parameters such as bridge width, length and initial speed.

In general, it appears that anything less than a 24-foot clear bridge width for a two-way bridge operation or a reduction of the shoulder on the bridge or any factor that causes changes in a driver's lateral position and speed defines a "narrow bridge condition".

Factors Which Affect Safety at Bridges

Bridges pose a high accident potential especially when the roadway is narrowed on the bridge. Common types of accidents are collisions with guard rails on approach roadways or collisions with bridge railings. If the bridge width is narrow, drivers have a tendency to move in closer to the centerline to draw safely away from the bridge railing, but instead risk a collision with oncoming cars. Several investigators have carefully studied these accidents and suggested remedial measures.

Hilton (24) conducted an extensive study of bridge accidents in Virginia and found that several geometric characteristics predominate at the accident sites. Some of the salient characteristics found were pavement transitions at bridge approaches, approach roadway curvature to the left, narrow bridge roadway widths, intersections adjacent to bridges and combinations of these and other geometric factors. He also found that the severity index (a ratio of the proportion of persons killed to the proportion of all accidents on the highway system in Virginia) was high for the bridge accidents on interstate highways.

Gunnerson (25) studied 72 bridges in Iowa over a 12-year period and found 65 bridges had a width of less than 24 feet which was the approach roadway width. Others had a 30-foot width which was actually 6 feet wider than the approach roadway width. He concluded that if the approach roadway is widened but the bridge is not, there is an increase in the accident rate, whereas on control bridges, where the bridge is also widened, the accident rate decreases.

Raff (26) conducted an extensive study of accident data from 15 states on rural highways covering 5000 miles. The routes were grouped into a large number of short, homogeneous sections and accident rates of the groups were compared. The study included both highways and bridges. Some of the factors included in the study were number of lanes, average daily traffic, degree of curvature, pavement and shoulder widths and sight distance. Traffic volume was found to have a strong effect on accident rates on most highway sections except at two-lane curves and intersections. At bridges and underpasses there was found to be great value in increasing the roadway width several feet over the approach roadway.

Shelby and Tutt (27) evaluated the effect of two-lane bridge widths on the lateral placement of vehicles. Lateral placement of vehicles was measured at several bridge sites both on the road and the approach roadway. The sites included lane widths in the range of 11 to 19 feet. It was established that bridge width has a definite influence on the lateral placement of vehicles. Although the conclusions were not definitive, it appeared that an average driver needs a bridge lane width of about 20 feet in order to cross the bridge with little or no deviation in lateral position from that on the approach roadway.

Brown and Foster (28) conducted an investigation of bridge accidents on rural highways in New Zealand. The variables considered were: day and night; horizontal alignment of approaches (straight approach, left curved approach and right curved approach); place of impact (within the bridge or several points in the approaches); and width of bridge or approaches.

The researchers found that night time accidents are more frequent than day time ones and recommended reflectorization techniques for night visibility. They also found that 60 percent of bridge accidents occurred within the left hand approach, 20 percent occurred within the right hand approach and the remaining took place within the bridge structure. Furthermore, 70 percent of the bridge accidents took place where the bridge width was less than 79% of the approach roadway. The researchers recommended properly designed guard rails to "deflect the vehicles back on to the highway and/or away from the end posts."

As early as 1941, Walker (29) studied bridge width and its influence on transverse positioning of vehicles on the bridge. On the basis of transverse positioning of freely moving vehicles on the bridges, he found that an 18-foot pavement with 3-foot shoulders or a total roadway width of 24 feet, requires a concrete bridge width of 26 to 28 feet in order to make the driver feel secure. The greatest width of a bridge required for a 22-foot pavement was found to be 30.6 feet. Sidewalks apparently added nothing to the effective roadway width on the bridge because transverse position occurred at a fixed distance from the curb or from a parapet if there was no curb.

King and Plummer (30) evaluated the lateral vehicle placement on a simulated bridge of variable width and concluded that there should be a minimum shoulder width of 4 to 6 feet on the bridge for safe traffic operations. Taragin and Eckhardt (31) determined from their study that the shoulders on highways should be at least 4 feet wide to provide maximum lateral clearance between vehicles travelling in opposite directions. They suggested further research on the effect of shoulder types and widths on accident rates. In a further study, Taragin (32) found that a relationship exists between vehicle speeds and lateral positions of vehicles on highways with paved shoulders.

Huelke and Gikas (33) considered non-intersectional fatalities a problem in roadway design. Sixty percent of the accidents were single car, off-road collisions with fixed objects such as trees, utility poles, bridge abutments and guard rails.

Roberts (34) studied the effect of bridge shoulder widths with curbs on the operational characteristics of vehicles. No strong relationship was observed between bridge shoulder widths and accidents. Vehicles moved farther away from the edge under all curbing conditions. There was strong evidence that outside shoulders 6 feet wide would not seriously affect the operational characteristics of vehicles on bridges.

A computerized inventory and priority analysis for roadside obstacles was carried out by Cunard and others in Michigan (35). They observed that 26% of all fatal accidents in Michigan during the study year involved collisions with fixed objects such as trees, guard rails, etc. They recommended a priority-ranking system of roadside obstacles based on the severity index of injuries caused for improving the roadside environment.

Presence of bumps at the approach and exit ends of bridges produce a traffic hazard. Hopkins and Deen (36) studied this problem and stated that this was due to the settlement of embankments on which approach roadways are built. They suggested that a careful analysis of consolidation of soils should be made in advance so that preventive measures could be taken. Bridge deck deterioration is another problem which can create hazardous conditions on the bridge. Several investigators studied this problem and suggested methods of repair and the effects of traffic vibrations during repairs (37, 38).

An article in the Indian Concrete journal (39) deals with methods of inspection and maintenance of highway bridges using suitable equipment. The authors suggested a periodic and systematic in-depth inspection of all bridges in a planned way and prompt remedial action to be taken in case defects are found.

Methods Employed to Improve Safety at Bridges

Highway engineers have been improving their designs and taking preventive measures to increase safety at bridge sites. These efforts have resulted, among other things, in increased widths for new bridges, better geometric features to guide the drivers safely through the bridge, improved guard rails to deflect the cars safely from the barriers, improved signs, markings and use of proper lighting or reflective markers. Some of these developments are discussed in detail in this section.

As early as 1947, the chief of the Bridge Division of Public Roads Administration, Washington, D.C. described the significant changes occurring in modern bridge design to improve safety (40). Some of the improvements mentioned were extending the full width of the roadway on to the bridge, wider curbs and streamlined railings. Godwin (41) described how old wooden bridges in California are replaced by small concrete slab culverts or pipe-arch installations. A bridge parapet for safety was developed at General Motors Proving Ground (42) which consists of a tube railing on a concrete parapet. Tests at the proving ground showed that the design met many safety requirements, although it was somewhat more expensive.

Zuk et.al. (43) , described various methods of modifying historic bridges in Virginia for contemporary use. Many of these old bridges were narrow and had low load carrying capacities, making them targets for replacement. On a case by case basis, the bridges were investigated for their potential for strengthening and widening. Some were explored for non-vehicular uses such as museums for historic preservation.

In the article "New Highway Bridge Dimensions for Safety", Clary (44) emphasized that safety must be the first priority in design followed by cost, durability and aesthetics. He suggested that maximum clearances must be provided between a roadway and the exposed parts of bridges such as piers and abutments. Bridge railings must be of adequate strength and continuous with approach guard rails. There must be adequate transition between the rigid bridge railing and the flexible approach guard rail. Any obstruction such as wide curb which might cause vaulting if struck by a vehicle must be removed as far as practicable. Openings between twin bridges must be eliminated where feasible or they must be adequately protected by a guard rail.

Under the sponsorship of the National Cooperative Highway Research Program (NCHRP) considerable research was conducted for improved designs of guard rails and bridge rails to reduce severity of accidents. The main objective was to deflect the vehicle away from the fixed object with minimum impact, thus reducing accident severity. Several reports (45, 46, 47, 48) were issued by the NCHRP on guard rail design and evaluation which indicate considerable progress in the development of safe guard rails. Under the same program Olson and others (49, 50) developed design criteria and guide lines for bridge rail systems.

Michie and Bronstad (51) gave guide lines for location, selection and maintenance of traffic barriers. They classified traffic barriers into two types: (1) longitudinal systems (guard rails, bridge rails and median barriers) and (2) crash cushion systems (to prevent head on impact with fixed objects near bridges and off-ramp areas). Examples of crash cushions include steel barrel configurations, entrapment nets and an array of containers filled with sand or water. They suggested that accidents frequently may actually increase if these traffic barriers are not properly selected and placed.

Nordlin et.al. (52), reported on five full scale vehicle impact tests on California Type 20 bridge rail. They found that the rail will retain and redirect a 4900 lb. passenger car impacting at speeds up to 65 mph and approach angles up to 25 degrees, with little or no damage to the rail. Occupant injuries varied from minor with seat belt, to moderate with no seat belt. Other researchers (53, 54) also reported on energy absorbing guard rails to improve safety of the drivers. Post et.al. (55), discussed the cost-effectiveness of two types of guard rail-bridge rail transition improvements. The findings showed that a double W-beam type used in Nebraska was slightly superior to the AASHTO stiff post system.

Hunter et.al. (54), gave methodology for ranking roadside hazard correction programs. Initial runs on the system indicated that use of transition guard rails at bridge ends and tree removal at certain locations in North Carolina were promising. A study undertaken in Texas (57) showed that there are ways other than widening to improve safety at narrow bridges. Good results were reported by blending the approach rail smoothly with the bridge rail and by using pavement markings to guide the driver on to the narrow bridge.

The Texas Transportation Institute (21) suggested fourteen alternative treatments to improve safety at narrow bridges and indicated how these could be used in various situations. Some of the fourteen treatments suggested are changing approach grades, realigning the roadway, installing smooth bridge rails and guard rails, placing edge lines, installing narrow bridge signs, and advisory speed signs. Three NCHRP reports (58, 59, 60) give methods for evaluating highway safety improvements and determining the cost-effectiveness of the programs. Rehabilitation and replacement of bridges on secondary highways and local roads were discussed in two recent reports which were also published by the NCHRP (61, 62). Several types of repair and rehabilitation procedures for correcting common structural and functional deficiencies in highway bridges and bridge decks were included in the two reports. Another report of the NCHRP (63) gives guidelines for bridge approach design to eliminate rough riding characteristics at approaches to bridges.

Tamburri (64) and others discussed the effectiveness of minor improvements in reducing accident rates. The improvements included warning flashing beacons, safety lighting installations in reducing night accidents, protective guard rails and various delineation devices. Protective guard rails were found to be very effective at narrow bridges. The New Jersey State Highway Department (65) found that low level lighting produced by fluorescent lights mounted on bridge railings was very effective. An article in Better Roads (66) considers the Danish practice of wide edge and center line pavement markings using thermoplastic materials as outstanding. Beaton and Rooney (67) studied raised reflective markers for lane delineation. They found fully beaded button markers more effective than wedge markers in rainy weather at night on concrete roads. They found wedge markers more durable on asphalt roads than the button type. Turner and others (68) found that full width paved shoulders reduce accident rates on rural highways and especially on two lane roads.

The Manual on Uniform Traffic Control Devices (69) describes standard signs for posting "NARROW BRIDGE" for two way bridges with widths 16 to 18 feet and "ONE LANE BRIDGE" if the bridge width is less than 16 feet for light traffic or less than 18 feet for heavy commercial traffic. Additional signs such as "NO PASSING ZONE", "YIELD", etc. for advance warning of dangerous structures are also sometimes recommended. A dynamic sign system (traffic actuated) to alert motorists to the presence of narrow bridges at an experimental main facility did not show much difference between dynamic and normal signing in terms of speed and lateral placement (70). The role of signs in a highway information system as well as several aspects in the design were analyzed in detail in an NCHRP report (71).

Relating Accidents to Highway and Bridge Features

To initiate corrective treatments for improving safety at bridges, it is important to know which factors contribute to the accident potential at bridges. Several researchers attempted to relate roadway and bridge characteristics with accident rates but achieved only a limited success because of the complexity of the problem. Highway accidents involve three principal elements: the driver, the vehicle, and the highway. Highway features are only a part of the cause of accidents. Driver behavior also plays an important part. A major reason for the limited success of previous studies to relate bridge and roadway characteristics to accident rates is the non-uniformity in reporting and collecting accident data. This is especially so of data collected in the past.

One of the earliest studies of bridge accidents was conducted by Raff (26) for the Bureau of Public Roads in 1954. Accident data were collected from 15 states on different types of road sections covering curves, straight portions (tangents), structures, railroad crossings and different grades (slopes). He first encountered the problem of combining detailed data from different states but he tried three different approaches to combine the data and obtained three types of accident rates. His analysis showed that traffic volume was found to have 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 (72) attempted to relate accidents at bridges to roadway geometrics at bridge approaches. They studied eleven independent variables, some of which are average daily traffic volume, bridge width, width of approaching pavement, sight distance, curveline, height of bridge rail, length of the bridge and travelling speed. For the purpose of studying conditions in the driving environment, the data were classified into several categories including two-lane and four-lane rural highways. Multivariate regression and stepwise regression procedures were used in developing models. The 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 are not linear but can be expressed by a logarithmic transformation. On two lane roadways sight distance (the greatest distance that a driver can clearly see ahead while driving on the highway in order to spot an obstacle on the road) was found to be important for night driving, while the degree of curvature became critical during day time.

Kihlberg and Tharp (73) conducted a study to relate accident and severity rates for various highway types and to various geometric elements of the highway. For the statistical analysis of the data, the highway segments were arranged into 15 ADT groups. The presence of geometric elements (curves, grades, intersections and structures) increased the accident rate on highways. The presence of combinations of geometric elements generated higher accident rates than the presence of individual elements. Grade and curvature affected accident rates only when they changed from zero value to 4 percent value. The geometric elements did not affect the severity rates.

Turner (75), using bridge accident data from Texas and Alabama developed a probability table which predicts the number of accidents per million vehicles for various combinations of roadway width and bridge relative width. His statistical investigation did not uncover a unique combination of variables to predict accident rates conclusively at specific structures. He attributed this to the effects of minor variables not included in the bridge data and to the fact that bridge accidents are complex with multiple contributing aspects. He also developed a cost-effectiveness methodology for the relative evaluation of various bridge safety treatments.

A comprehensive analysis of safety specifically at narrow bridges was conducted by TTI for the NCHRP (21). They identified ten important factors related to approach roadway, bridge geometry, traffic and roadside distractions. They developed a linear model combining these factors and called it the Bridge Safety Index. The Bridge Safety Index (BSI) could be expressed as:

$$BSI = F_1 + F_2 + F_3 + \ldots\ldots + F_{10}$$

where

F_1 is a function of clear bridge width determined by entering Figure 1 with the clear bridge width.

F_2 is the ratio of the bridge width to the approach roadway width, a measure of the relative constriction of lateral movement as a vehicle travels from the approach lane on to the bridge.

F_3 is the approach guard rail and bridge rail structural factor and attempts to define the safety aspects of the rail and the contribution to bridge perspective that the approach rail offers to an

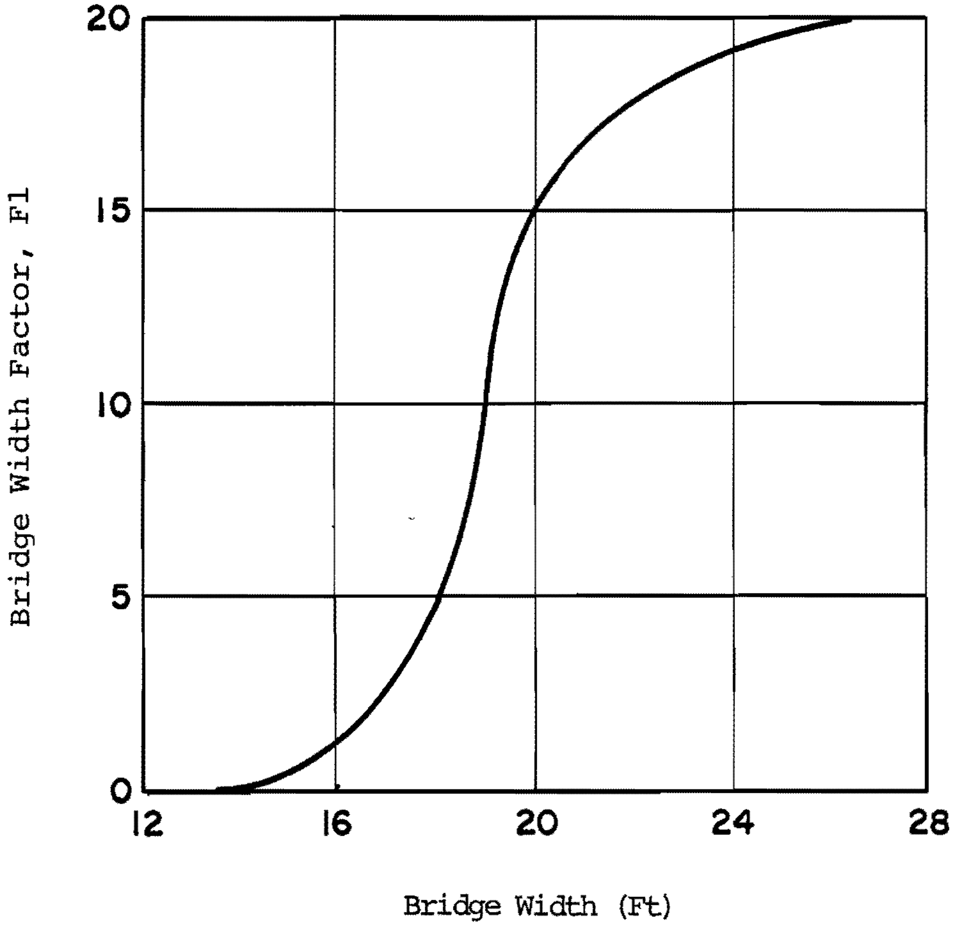

Figure 1. Weighting of Bridge Width Factor (F_1).

oncoming driver.

F_4 is the ratio of approach sight distance (ft) to 85% approach speed (mph) and indicates the time within which a driver may prepare for the bridge crossing.

F_5 is a measure of the approach curvature and is equal to 100 + tangent distance to the curve (ft)/curvature (degrees).

F_6 is grade continuity (%) and denotes the average grade throughout the bridge zone and the algebraic difference in approach and departing grades.

F_7 is shoulder reduction (%) and is defined as the percentage that the shoulder width on the approach roadway is reduced as it is carried across the bridge.

F_8 is the ratio of volume to capacity and is an indirect way of accounting for the number of conflicts on the bridge.

F_9 is the traffic composition. If the traffic composition includes relatively high percentage of large truck traffic (> 10%) narrow bridges can become critically narrow.

F_{10} is the distraction and roadside activities and is considered the least objective of all of the factors proposed.

Ivey et.al. (21) considered the first three factors to be 4 times more important than the factors F_4-

F_{10}. Table 1 gives the evaluation of factors F_2-F_{10}.

In this first BSI model of Ivey, et.al. (21), the factors F_1, F_2, and F_3 are rated from 0 to 20 while the factors F_4 to 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. They suggested that the model was preliminary and would be improved as more data and information became available from different states. Newton (76) developed a manual for field evaluation of bridges for the TTI study and its extension. Tseng (77) collected additional data for the TTI study with a view to improve the BSI model. He added two new factors F_{11} (paint marking) and F_{12} (warning signs or reflectors) to the study of BSI as defined below.

F_{11} is a factor that deals with paint markings and is defined as the combination of centerline, no passing zone stripes, edge lines, and diagonal lines on the shoulder of the pavement.

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 or black-white panels on the bridge ends.

These factors were evaluated subjectively and given values ranging from one to five. Factor F_{11} may be

Table 1. Factors Used to Determine Bridge Safety Index

	Factor Rating for F_2 and F_3				
Factor	0	5	10	15	20
F_2	$\leq$ 0.8	0.9	1.0	1.1	1.2
F_3	Critical	Poor	Average	Fair	Excellent
	Factor Rating for F_4 - F_{10}				
	1	2	3	4	5
F_4	$\leq$ 5	7	9	11	14
F_5	$\leq$ 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 Dis-continuities	Non-Uniform	Normal	Fairly Uniform	Uniform
F_{10}	Continuous	Heavy	Moderate	Few	None

defined in terms of the condition of centerline and no passing zone stripes, edge lines and diagonal lines on the shoulder of the pavement. For example, using the nomogram given in Figure 2 one can obtain a value of four for a marginal center line, adequate edge line and a marginal diagonal line.

The F_{12} factor can be defined in terms of narrow bridge signs, speed signs, reflectors on the bridge and black and white panels on the bridge ends. A value of five corresponds to an excellent condition of warning signs, four for fair, three for average, two for poor and one for no signs. Table 2 indicates the evaluation of F_{12}. Engineering judgement was used to convert the observed estimation into one of the descriptive terms. F_{11} and F_{12} together are considered to be effective in reducing the lateral movement of the vehicle and controlling speed, thereby contributing to the reduction of accidents. Tseng's BSI factors ranged from 1 to 5 with a maximum possible BSI of 60.

Tseng used discriminant analysis to categorize bridges into safe or hazardous, and used stepwise regression analysis to determine the factors which significantly affect bridge safety. He found only two of the twelve factors, F_1 and F_6, to be significant. He considered that more data are necessary to get meaningful results.

Luyanda and Smith (78) conducted a multivariate

TABLE 2. Evaluation of F_{12}

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

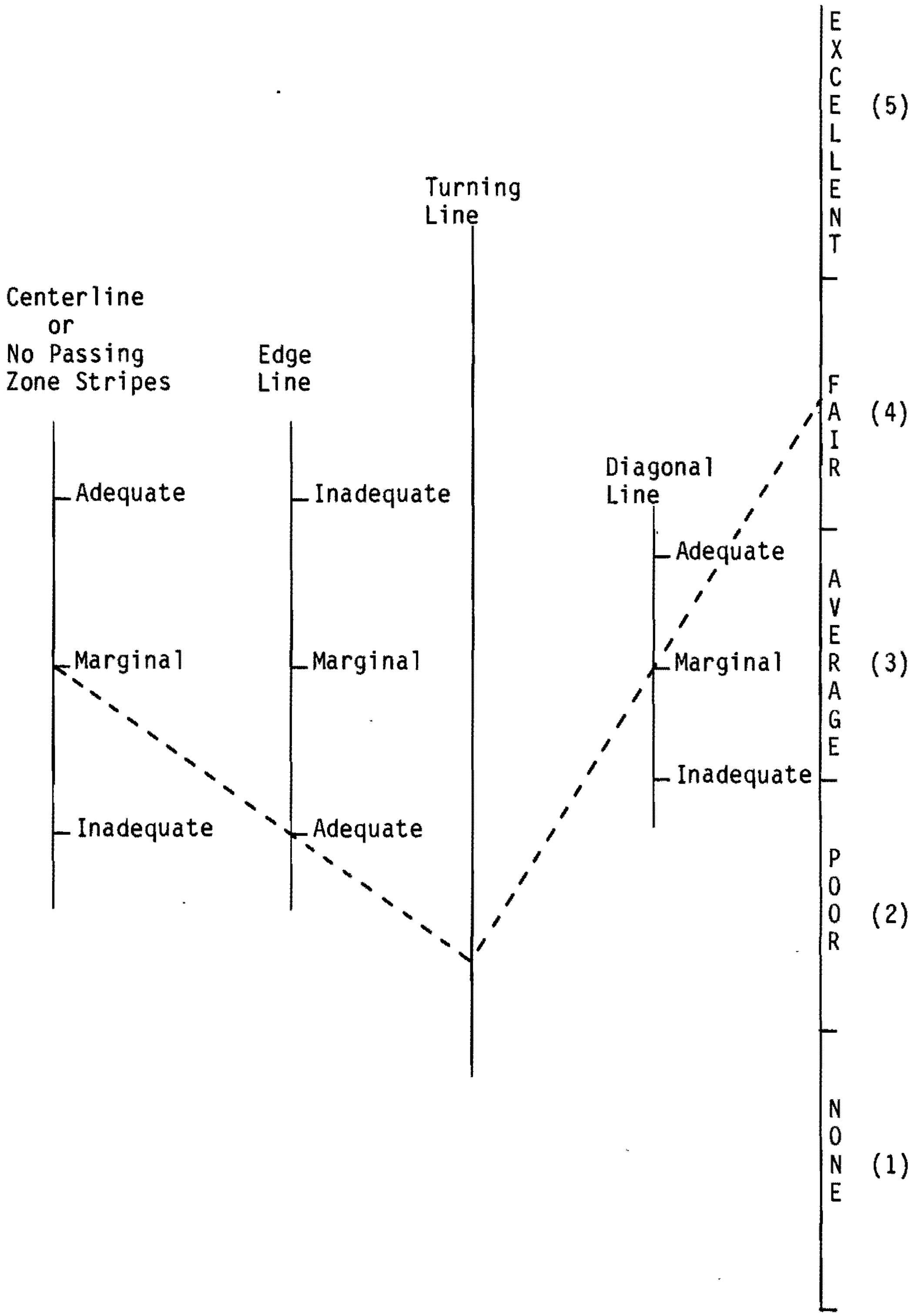

Figure 2. Nomogram Used to Determine F_{11}

statistical analysis to relate highway accidents to highway conditions with regard to intersections. They were able to divide rural intersections and segments into groups using cluster analysis and used discriminant analysis to identify the variables affecting accidents in each group of intersections.

Southwest Research Institute (79) conducted an extensive study to evaluate the effectiveness of measures for reducing accidents and accident severity at narrow bridge sites. Environmental and accident data were collected from five states using the FHWA bridge inventory and accident files. Also, 125 accidents at bridge sites were investigated in depth. Extensive statistical analyses were conducted to relate bridge characteristics to accidents. Some of the variables studied are bridge length, lane width, curveline, sight distance, percent shoulder reduction, speed limit reduction, average daily traffic, percent trucks, signing, and roadside distractions. Their main problem was dealing with inconsistencies of data in individual state files which required major screening and code transformation. Bridge narrowness, as defined in terms of shoulder reduction had a significant effect on accident rates for two lane undivided structures. A general lack of positive relationship existed between individual bridges and approach characteristics and accident severity. As expected, ADT was the most predominant operational factor affecting accident frequency. BSI was found to be significant only for the accident rate on divided bridges. The authors considered discriminant analysis to be a reliable tool in distinguishing between hazardous and nonhazardous bridges. They were not able to evaluate the counter measure effectiveness with the data collected in the study. Many of the problems encountered in the study were associated with the quality of the available data. Accuracy of accident locations was another problem area requiring further attention.

Summary

The literature review yielded a wealth of information regarding the factors which 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 (curved or straight), sight distance, type and location of guard rails, transition of guard rails to bridge rails and traffic factors are all considered very important. The researchers were not completely successful in developing a model relating accident rate at the bridges to all of the pertinent features mentioned above. 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. Hence some researchers used probabilities to estimate accidents at bridges, others used multivariate statistics. Thus, there is much scope for improving the bridge accident model using suitable statistical techniques.

CHAPTER 3

PRELIMINARY ANALYSIS OF DATA

In Chapters 3 and 4, a brief description is presented of the statistical analysis that was made in developing a new Bridge Safety Index. It was the objective of this development to obtain the new index in as objective a manner as possible by using observed accident rates as the indicator of bridge safety.

A number of statistical terms are used in these two chapters with which the reader may not be familiar. In order to become better acquainted with these terms, six appendices have been prepared which describe a number of the statistical methods that were used in the development of the new Bridge Safety Index. These six appendices are as follows:

Appendix A - Regression and Correlation Analysis and Allied Concepts and Procedures

Appendix B - Factor Analysis

Appendix C - Cluster Analysis

Appendix D - Discriminant Analysis

Appendix E - Testing for Normality

Appendix F - Logistic Regression

There are some terms that are used in the discussion that follows immediately that are defined here. A "regressor" is what is commonly called an "independent variable" in a regression equation. "Multicollinearity" is a distortion of the coefficients in a regression equation that occurs when two of the "regressors" are closely correlated with each other. "Variance Inflation Factors" are calculated numbers that are used to detect the variables that may contribute to "multicollinearity". The "level of significance", which is denoted by α, is a measure of the strictness of the statistical test that is applied to the variables. The usual value of α is 5% or 10% with the larger number indicating a more severe test.

In Chapter 3, a preliminary analysis is described in which the variables previously identified by Tseng (77) were tested for their correlation, multicollinearity, level of significance, and so on. It was found that some of these variables are significant indicators of potentially high accident rates and suggested that other variables might be found that would permit the development of a better Bridge Safety Index. The collection and analysis of additional data are carried out and the results are described in Chapter 4.

In developing his Bridge Safety Index model, Tseng (77) compiled a computer data file consisting of the variables F_1 to F_{12} and accident rate which were made readily available. As described in the literature review the F-factors are defined as follows:

F_1 = clear bridge width

F_2 = bridge lane width/approach lane width

F_3 = guard rail and bridge rail structure

F_4 = approach sight distance/85% approach speed

F_5 = 100+ tangent distance to curve/curvature

F_6 = grade continuity

F_7 = shoulder reduction

F_8 = volume/capacity ratio

F_9 = traffic mix

F_{10} = distractions and roadside activities

F_{11} = paint markings

F_{12} = warning signs and reflectors.

Correlation and factor analyses were conducted on these data and the data were searched for signs of multicollinearity. All possible combinations of the variables were tried starting with one variable models and ending with a full model of 12 variables in order to investigate the nature of the contribution of the variables toward a good fit in the regression model.

A description of the statistical procedures that were used and the results of the analysis follows.

Correlation Analysis

An analysis of the correlations between the variables was done as a routine procedure. It should be noted that

	F_1	F_2	F_3	F_4	F_5	F_6	F_7	F_8	F_9	F_{10}	F_{11}	F_{12}	AR
F_1		*	*					*		*		*	*
F_2			*				*					*	
F_3													
F_4					*		*			*		*	
F_5										*	*		
F_6													*
F_7									*	*			
F_8									*	*			
F_9													
F_{10}											*		
F_{11}													
F_{12}													
AR													

FIGURE 3. Matrix of Significant Correlations of F-factors of Tseng's data at α = Ø.1Ø (Note: AR = Accident Rate)

there are no high correlations between the F - variables (F_1 - F_{12}) nor between the accident rate and the F's. On the other hand, when there are a fairly large number of regressors, no pair of correlations may be large. The independent variable F_1 corresponding to bridge width had the highest correlation of -0.43398 with the accident rate, the variable F_6 corresponding to grade continuity had the next highest correlation of -0.26237. Both F_1 and F_6 have statistically significant correlations with accident rate at $\alpha = 0.10$.

Examining the correlations in Figure 3 between F-variables related to bridge geometrics with $\alpha = 0.10$, it is seen that F_1 and F_2 are significantly correlated. This is expected since F_2 is a function of clear bridge width. It should also be noted that F_2 has a significant correlation with F_7 (shoulder reduction). Also, F_4 and F_5 are significantly correlated possibly because the curvature of the bridge and sight distance are related. Correlations between the subjective factors (F_3, F_9, F_{10}, F_{11}, F_{12}) with each other and with objective factors are not readily interpretable.

<u>Multicollinearity Diagnostics Variance-Inflation Factors (VIF)</u>

Variance Inflation Factors (VIF) for the variables F_1 - F_{12} are given in Table 3. VIF's greater than

1Ø indicate multicollinearity. Because the variance inflation factors for all of the variables are less than 1.5Ø, no multicollinearity is indicated.

TABLE 3. VARIANCE INFLATION FACTORS

Variable	Variance Inflation Factors
F_1	1.359Ø31
F_2	1.594948
F_3	1.144575
F_4	1.6Ø3319
F_5	1.35819Ø
F_6	1.1Ø3444
F_7	1.759364
F_8	1.419933
F_9	1.273589
$F_{1Ø}$	1.454376
F_{11}	1.15278Ø
F_{12}	1.411476

Variable Selection with R^2 as Indicator

It is well known that the variance of the predicted variable increases with the number of unnecessary predictor variables included in the model. A model needs to predict well with all of the predictor variables that are included. At the same time having as few independent variables as possible that effectively predict is considered a desirable quality in a regression model.

The R^2 is an indication of a good model because the higher R^2 usually means a better fit. The R^2 increases with the number of regressors and if an additional regressor does not increase the value of R^2 substantially it can be deleted if it is not otherwise important for practical reasons. To get the best model in regression sometimes several models are considered and the 'best' as judged from a practical and feasible point of view is accepted. The 'R square' procedure of the Statistical Analysis System (SAS) obtains all possible regressions for a dependent variable when the regressors are known and when the behavior of many models is to be investigated.

In the stepwise procedure used by Tseng (77), it was noted that the model with all variables had an R^2 of 0.2589 and the linear regression model that has F_1 (bridge lane width) and F_6 (grade continuity) has an R^2 of 0.226. Adding 10 more variables did not improve R^2 much.

All possible regression models were obtained starting with a single predictor variable and going up to the full model with 12 variables. It is seen from the models with one variable that F_1 has the maximum R^2, the next being F_6 and in the models with two variables. The model with F_1 and F_6 has the best R^2 as noted by Tseng. With only 5 variables, F_1, F_4, F_5, F_6, and F_8 yield an R^2 of 0.25. The maximum R^2 in a model of 6 predictors with the predictor variables of F_1, F_2, F_4, F_5, F_6, F_8 yields an R^2 of 0.255. When the variables F_{12}, F_{10}, F_3, F_{11} and F_5 are added successively, the R^2 does not change for all practical purposes. The full model with the twelve variables yields an R^2 of only Ø.259. Hence it is observed that adding 6 more regressor variables yielded only a 1.5% increase in R^2.

Factor Analysis

Factor Analysis was done on the data using the Factor Procedure from SAS with the principal axis method (Appendix B). Communality is the proportion of commonness that a given variable shares with others. It is found that out of the 12 factors that are possible the first four together account for about 56% of all of the variability. Adding 5 more factors accounts for about 89% of the variability. The remaining 3 factors account for less than 5% of the variability for each one of them. Table 4 gives factor loadings for each factor.

TABLE 4. Factor Loadings for Bridge F-Data

F Variables	Factor 1	Factor 2	Factor 3	Factor 4
F_1	Ø.95754	-Ø.ØØ129	-Ø.16686	-Ø.Ø5757
F_2	Ø.Ø9618	-Ø.Ø281Ø	-Ø.Ø3473	-Ø.21725
F_3	Ø.Ø8423	Ø.Ø4279	-Ø.ØØ7Ø2	-Ø.Ø3919
F_4	Ø.ØØ985	Ø.21933	Ø.Ø1753	-Ø.14114
F_5	-Ø.ØØ177	Ø.96477	Ø.Ø5619	Ø.Ø5633
F_6	Ø.Ø6474	Ø.Ø6874	-Ø.ØØØ76	-Ø.Ø2232
F_7	-Ø.Ø5976	Ø.Ø5924	Ø.Ø2739	Ø.94184
F_8	-Ø.16813	Ø.Ø5756	Ø.96445	Ø.Ø2592
F_9	Ø.Ø3Ø6Ø	Ø.Ø2234	Ø.1213Ø	Ø.1Ø69Ø
$F_{1Ø}$	-Ø.Ø931Ø	Ø.Ø76Ø1	Ø.193Ø2	Ø.1Ø498
F_{11}	Ø.Ø6583	Ø.Ø8Ø32	Ø.Ø1644	Ø.Ø415Ø
F_{12}	-Ø.12231	Ø.ØØ165	Ø.Ø3Ø68	-Ø.Ø6967

The most important variable in Factor 1 is F_1 with a factor loading of 0.95754. F_5 is the most important variable in Factor 2 and F_8 contributes the most to Factor 3. Factor 4 is heavily affected by F_7. If it is desirable to reduce the original variables to only four, it would be necessary to keep only F_1, F_5, F_8, and F_7.

This analysis is very useful when there are many variables. Since we have only twelve variables, it is not difficult to keep all the variables and use the complete information to develop an appropriate model. Factor Analysis has some subjectivity in the choice of the number of factors to be retained and the number of original variables.

Conclusions About Existing Data

Analysis of the existing data yielded the information that F_1 and F_6 are the most important variables and that F_2, F_6, F_7, F_9, and F_{11} are important factors as seen from the factor analysis and the R^2 procedure. No obvious multicollinearity could be diagnosed. None of the correlations between variables were high. It appeared that a better model of bridge safety could be constructed by investigating a number of other relevant variables.

CHAPTER 4

COLLECTION OF ADDITIONAL DATA AND ANALYSIS OF THE TOTAL DATA

Tseng (77) obtained the data for developing a new Bridge Safety Index from the data collected by TTI in the years 1978-1979 at 78 bridge sites in 15 districts in the state of Texas. These data were collected by T. M. Newton (76) with the cooperation and assistance of the district engineers and other district personnel. Only bridges on two-way, two-lane roadways were included. The Bridge Inventory and Inspection File (BRINSAP) was the source of the list of the narrow bridges and some bridge geometrics (101). The Texas Brinsap file consists of several items such as structure number and inventory route length. A sample of the information available is given below in Table 5.

TABLE 5. Some Items From BRINSAP File of Texas

Field	Item number	Item description
2	8	Structure number
11	11	Mile Point
24	24	Federal-Aid-System
29	29	ADT
32	32	Approach Roadway width
48	49	Structure length
51	51	Sidewalk or no

Accident rate was calculated from the accident data for the years 1974-1979 which were available in the form of computer output. Other information about each accident was not available at this stage but only the accident rate given by

$$y_1 = \frac{Y}{(\text{No. of Years})(1000\ \text{ADT})}$$

where y_1 = accident rate per 1000 vehicles

Y = number of accidents in the 6 years 1974-1979.

The conditions of the driving environment for each accident, including the type of traffic control and alignment were not readily available as a data set. It was considered that this information would be useful in assessing the relationship between accident rate and bridge geometrics and characteristics. If the variability of accident rate due to environmental conditions is significant, then this matter should be taken into consideration in developing a safety index. More information was needed and additional data were collected.

Additional Data Collected

Additional information was collected both from the printed computer output and bridge evaluation data available on the 78 narrow bridges. Since it was felt that several statistical techniques work better with continuous data instead of discrete data such as the F-values, the data on the bridge characteristics of bridge lane width, approach lane width, relative width (the difference between bridge lane width and approach lane width), approach sight distance, 85% approach speed, tangent distance to curve, curvature, actual grade continuity, actual shoulder reduction in feet and actual percent reduction in shoulder, volume and capacity at each bridge were noted down and tabulated. Individual values of the 85% approach speed and sight distance were not available for fourteen of the seventy-eight bridges. In order to make use of all the data in the analysis, 55 mph, which is the weighted average value of the 85% approach speed, was substituted for the missing values and 770 feet, a reasonable value, was substituted for sight distance.

From the available accident data, alignment, curvature, type of traffic control, light conditions, weather condition, surface condition, and road condition were noted. The number killed and the number injured at each accident were also recorded so that the accident can be categorized as being fatal, with injuries only or with property damage. Month, day and time of day were collected but only light condition was used in this study since that is what effects driver visibility.

From the BRINSAP file, item #24 was noted down to determine whether the road was rural or urban. Length of the bridges was noted down (item #49) for each bridge as well as sidewalk information on the bridges (item #51).

All of the information collected on the 78 bridges and 655 accidents was tabulated and made ready as a computer file. Several pertinent variables that may have a meaningful relationship with accident rate were collected and tabulated for the purpose of developing an improved BSI model. Some of the analyses performed are discussed below.

Testing Statistically for Effect of Type of Bridge

Before cluster analysis was done information about the road type was gathered using BRINSAP files (101). Item #24 yielded information about the type of bridge.

It was noted down whether the bridge was in a rural area or urban area and listed as R for rural and U for urban in the data files. 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, urban or rural, as the independent variable did not show significant differences (Table 6). The analysis of variance is given below.

TABLE 6. Analysis of Variance for Type of Bridge

Source	Degrees of Freedom	Sum of Squares	F	R^2	Pr>F*
Model (type)	1	0.0339			
Error	76	4.6807	0.55	0.007	0.4601

*Pr>F = probability that a random F value would be greater than or equal to the observed value.

Testing Statistically for the Effect of Sidewalks

Item #51 in the BRINSAP file gives information about whether the bridges have sidewalks or not. Often sidewalks may imply a curb and the effect on accidents of having or not having a curb was of interest.

There were 12 bridges with sidewalks and 66 without sidewalks. In the data file, 1 is a code for having a sidewalk and Ø for not having a sidewalk. An analysis of variance did not indicate a significant difference in accident rate between bridges with a sidewalk and bridges without a sidewalk (Table 7).

TABLE 7. Analysis of Variance for Sidewalks

Source	Degrees of Freedom	Sum of Squares	F	R^2	Pr>F
Model (sidewalk)	1	Ø.Ø3Ø86	Ø.5Ø	Ø.ØØ65	Ø.4813
Error	76	4.6838			

Because there was no significant difference between them, bridges from rural and urban areas were clustered together.

A Histogram of the Accident Rate

A histogram of the accident rate shown in Figure 4 indicates possible groups with high and low accident rates. The characteristics of the bridges that had an accident rate

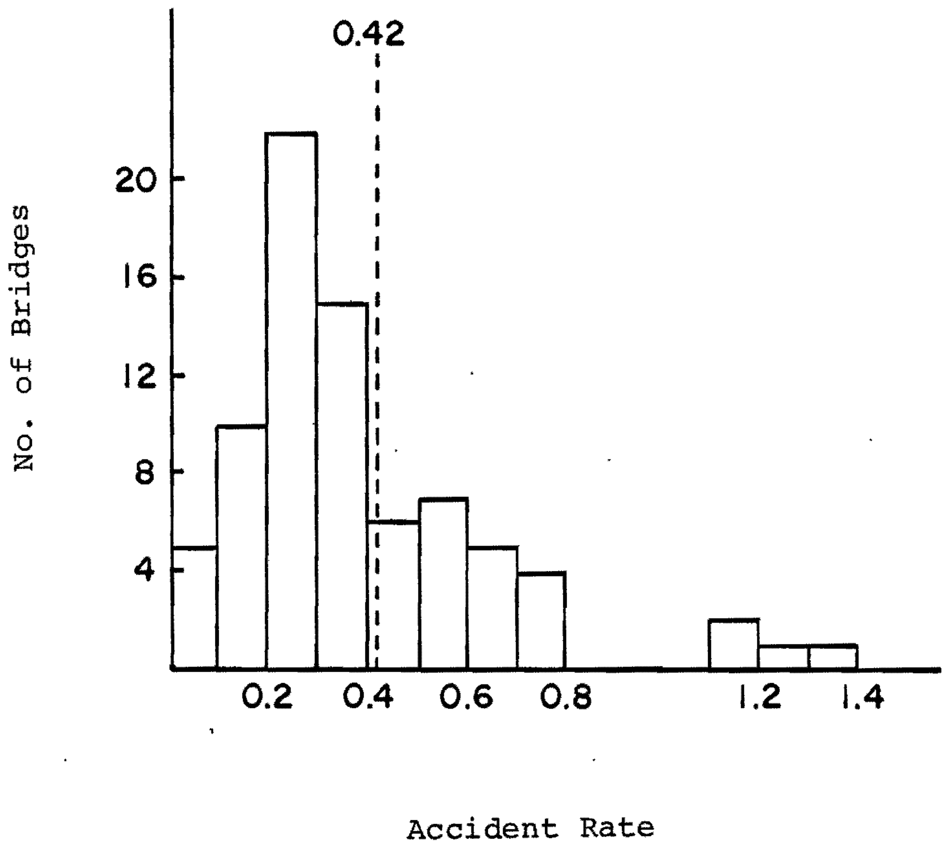

Figure 4. A Histogram of the Accident Rate.

of more than 1.0 were not observed to have any particular pattern. Cluster analysis gave the demarcation point between high and low accident rates to be about 0.4166.

Cluster Analysis of the Data

At the outset, an effort was made to cluster the bridges into two groups using accident rate and other variables, either continuous or discrete or both continuous and discrete. In clustering, the investigator accepts a cluster that is reasonable from a practical viewpoint since clustering is to some extent subjective and there are different clustering algorithms which give different clusters. After several attempts in partitioning, the cluster with 26 bridges in the "less-safe" group and 52 in the "more-safe" groups was accepted. This was given by the 'cluster' procedure of SAS using accident rate as the variable. Further analyses were made with these clusters. Since there is only one response variable, an analysis of variance was conducted on the groups obtained from cluster analysis with the following results.

TABLE 8. Analysis of Variance for Cluster Groups

Source	Degrees of Freedom	Sum of Squares	F	Pr>F
Model (groups)	1	2.98501200	131.16	0.0001
Error	76	1.7296939	0.55	

From Table 8, it is seen that the groups are very significantly different with regard to accident rate. The cluster yielded the 26 bridges into the less-safe group and the group identification variable is given by Z=Ø whereas the 52 more-safe bridges have the group identification given by Z=1. These values Z=Ø and Z=1 were used for further analysis to differentiate between the two groups. Table 9 shows the bridges in the less-safe groups (Z=Ø) as given by the cluster analysis. The full list of bridges may be found in Appendix G.

TABLE 9. Identification (ID) Numbers of the Bridges in Less-Safe Group

#	ID	Accident Rate
1	2F	1.111ØØØ
2	22F	1.1111Ø
3	1ØH	1.Ø753Ø
4	12I	1.1994Ø
5	2H	Ø.555Ø
6	11E	Ø.555Ø
7	9G	Ø.5477Ø
8	18G	Ø.54Ø5Ø
9	16G	Ø.53Ø3Ø
1Ø	1ØD	Ø.5866Ø
11	11D	Ø.625ØØ
12	3F	Ø.625ØØ
13	16F	Ø.6154Ø
14	12G	Ø.6395Ø
15	9F	Ø.44Ø2Ø
16	13F	Ø.45Ø5Ø
17	18B	Ø.4487Ø
18	19E	Ø.4166Ø
19	13B	Ø.4726Ø
2Ø	2ØH	Ø.4762Ø
21	4B	Ø.463ØØ
22	1ØF	Ø.8454Ø
23	13D	Ø.6838Ø
24	21H	Ø.6838Ø
25	3C	Ø.7575Ø
26	3G	Ø.7143Ø

The remaining 52 bridges were in the more safe group. It should be noted that the boundary between high and low accident rates is 0.4166. When accident rate is lower than 0.4166 the bridge is classified into the more-safe group. The mean of the accident rate for the less-safe group is 0.65974 whereas the mean for the more safe group was 0.24476.

A Correlation Analysis

A correlation analysis was done on the data set compiled. The two matrices of correlation are given in Figure 5 and 6 and indicate significant correlations at an $\alpha = 0.10$.

Additional notation in Figures 5 and 6 is given as: BW = Bridge width, L = Bridge length, SD = Sight distance, GC = Grade continuity, RW = Relative width, SRN = Shoulder reduction, $F_{20} = (F_{11} + F_{12})/2$ (average), SP = Speed, y_1 = accident rate. Figure 6 includes the variable F_{20} which is the average of F_{11} and F_{12}.

At the outset, it was observed that none of the correlations are very large. Bridge width is significantly correlated to relative width which is somewhat a function of bridge width. Using relative width and bridge width together in a model has to be approached with caution. Bridge width and sight distance are highly correlated but it is not particularly meaningful. The distraction factor

	BW	ADT	L	SP	RW	SD	GC	SRN	F_3	F_5	F_9	F_{10}	F_{11}	F_{12}
BW					*	*			*			*	*	*
ADT												*		
L								*				*		
SP							*			*		*	*	*
RW								*	*	*				
SD								*						
GC													*	
SRN												*		
F_3														
F_5												*		
F_9														
F_{10}													*	
F_{11}														
F_{12}														
y_1	*		*											

*Indicates significance at $\alpha = 0.10$.

Figure 5. Matrix of significant correlations including F_{11} and F_{12}.

	BW	ADT	L	SP	RW	SD	GC	F_6	F_7	F_{20}	F_3	F_5	F_9	F_{10}
BW					*	*					*			*
ADT														*
L									*					*
SP												*		*
RW									*	*	*	*		
SD									*					
F_6														
F_7													*	
F_{20}														*
F_3														
F_5														
F_9														
F_{10}														
y_1	*		*					*						

* Indicates significance at $\alpha = 0.10$.

Figure 6. Matrix of significant correlations including the variable F_{20}.

(F_{10}) is significantly correlated with many variables such as average daily traffic (ADT), length, speed and grade continuity. Accident rate is significantly correlated only with bridge width, length, and F_6. It is somewhat related (significant at $\alpha = 0.15$) with ADT.

Multicollinearity Diagnostics for Total Data

The regression procedure in SAS was run on the continuous variables with the collinearity diagnostics given in Table 10.

All the variance inflation factors are less than ten and thus no multicollinearity is indicated. It is noted that relative width, which is the difference of bridge lane width and approach lane width, did not have a high variance inflation factor.

TABLE 10. Variance Inflation Factors (Total Data)

Variable	Variance Inflation Factor
Bridge width	0.905679
Average daily traffic	1.471238
Length	1.280139
Speed	1.468393
Relative width	1.520630
Sight distance	1.580017
Actual grade continuity	1.244776
Actual shoulder reduction	1.493000
F_3	1.208593
F_5	1.418598
F_9	1.209635
F_{10}	1.858220
F_{11}	1.229094
F_{12}	1.354638

Simple Linear Regression with Some of the Independent Variables

The regression procedure was applied to the variables actual shoulder reduction, sight distance, bridge width, relative width, speed, length, grade continuity, shoulder reduction as percent and the F-factors taking only one variable at a time to get an idea of the relationship of accident rate with the regressors. The coefficient of the regressor variable and its R^2 for the simple one variable regression are tabulated in Table 11.

It is observed from the signs of the regression coefficients that more bridge width leads to a smaller accident rate and a greater length means a higher accident rate, both of which are expected. The positive regression coefficient of speed indicates that more accidents occur with higher speed. The negative sign of the regression coefficient of sight distance shows that if the bridge can be seen at a greater distance there are fewer accidents. It is seen that the regression coefficients of relative width (the difference between bridge lane width and approach lane width) is positive which is not expected. In addition, the positive regression coefficient of shoulder reduction in percent implies that more shoulder reduction leads to more accidents. Actual shoulder reduction has a positive regression coefficient meaning that more reduction leads to more accidents. Average daily traffic does not have the

TABLE 11. Regression Coefficients
in Simple Linear Regression

Variable	Coefficients	R^2	Statistical Significance
Bridge width	-0.01969802	0.112169	**
Length	9.615403E-05	0.133414	**
Speed	0.00169350	0.000699	
Relative width	0.00305447	0.000204	
Sight distance	-3.6029814E-06	0.001113	
ADT	-0.01339770	0.026979	
Shoulder reduction in percent	0.00029671	0.001679	
Grade continuity (F_6actual)	0.01067415	0.027403	
Actual shoulder reduction (F_7actual)	0.00096289	0.000247	
F_1	-0.23163689	0.155898	**
F_2	-0.01087596	0.155898	
F_3	-0.02036228	0.009125	
F_4	-0.00480595	0.000151	
F_5	-0.02237066	0.017692	
F_6	-0.04128947	0.053350	*
F_7	-0.00601092	0.001216	
F_8	0.01816129	0.002336	
F_9	0.00599381	0.000603	
F_{10}	-0.00013498	0.00000	
F_{11}	-0.3377434	0.018361	
F_{12}	0.04454014	0.019059	
F_{20}	-0.00980679	0.000566	

* significant at α = 0.10
** very significant at α = 0.10

expected regression coefficient.

It should be noted that all the F-factors are expected to have negative regression coefficients since they are safety factors and greater safety generally implies fewer accidents. It can be seen that with the exception of F_{12} all of the factors have negative regression coefficients for this particular data set. The factor F_1 (bridge width) is highly significant and F_6 (grade continuity) is significant in the models. The remaining F-factors are not statistically significant but are possibly useful in the model even if they are not very significant because some of them can lead to increased safety, are much less expensive than widening a bridge and may save a few human lives. The marking (F_{11}) and signing (F_{12}) F-factors may affect accident rates by helping to lower traffic speed.

Individual F's and factors are investigated in simple linear regression models because sometimes a combined model due to possible interrelationships of the variables does not lead to correct interpretations.

Stepwise Regression of the Total Data

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 (Appendix A). The analysis was conducted using the data for bridge and roadway characteristics of 78 bridge sites and 655 accidents with accident rate as the dependent variable. The following table gives the regression coefficients of the model with accident rate as the response variable yielding an R^2 of Ø.6547.

TABLE 12. Regression Coefficients of Stepwise Regression

Variable	Coefficient	Prob>F
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

From Table 13, it can be seen that the first three variables to enter, length, F_6, and bridge width in that order contributed well and accounted for about 49.9% of the variability. Comparing the three variable model with the final 21 variable model, we find 75% of the contribution to R^2 is made by the three variables. Some more of the stepwise regression results are given below.

TABLE 13. 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

A 12 variable model has an R^2 of 0.653 when rounded to 3 decimal places. This is 99.65% of the R^2 of the all-variable model. The environmental factors of light condition, road condition, surface condition, and weather condition are in the model as indicator variables (1 for the normal safe condition and 0 for a non-normal condition) and did not prove to be significant in the stepwise model. Alignment, traffic control, and curvature were also considered in the model but did not contribute much to accident rate. As can be seen from the final table, Table 13, bridge width, ADT, speed, length, F_6 (grade continuity), F_{10} (distractions), F_{11} (markings), F_{12} (signs), relative width, and sight distance were significant but F_{10} (distractions) and F_{12} as well as relative width and sight distance do not have the expected signs even though in the single variable regression, it was noted that F_{10} and F_{11} had the correct signs indicating less accidents with a larger F-value.

Relative width has a positive regression coefficient in the 21 variable model and in the simple linear regression model, which is not satisfactory. Sight distance has a positive coefficient in the larger model although it had the correct negative sign in the simple variable regression model. Some of the F-factors like F_2 and F_{10} had the correct negative coefficients in the linear simple regression models but have positive coefficients in the final model. Multicollinearity could not be diagnosed but might be hidden. As many of the variables as possible that are shown to be important in this procedure will be included in the models to be explored. Also, most of the subjective F's that are important and can be easily improved will be considered.

Regression Analysis Using the SAS R-Square Procedure

To understand the interrelationships of the variables, a model including the factors bridge lane width, average daily traffic, speed, length, sight distance, relative width, F_3 (rail condition), F_7 (shoulder reduction), F_9 (traffic mix), $F_{1\emptyset}$ (distractions), F_{11} (markings), and F_{12} (signs) was used. The maximum number of variables to be included in a model was limited to twelve. Only the 78 bridges with their geometrics and F-factors were considered but not the 655 accidents with accident rate as the dependent variable.

It is seen that length, bridge width, F_6 and ADT had the highest values of R^2, in that order, when one variable models were considered. Bridge width, length and F_6 together yielded a model with an R^2 of Ø.276 which was the highest value in the three variable models. In most of the models with more than 6 variables, relative width and sight distance were included besides bridge width, length and F_6.

The best 12 variable model had an R^2 of Ø.374 as compared with an R^2 of Ø.351 for the best 8 variable model showing an increase of only 6.3% by adding four more variables. The factor F_{12} was included in several of the models as contrasted with F_{11} which occurred mostly in the eleven and twelve variable models.

The contribution of the variables and their combinations are noted and will be considered in the development of an improved safety index model.

Factor Analysis of the Total Data

Two factor analysis models were developed using the data on the bridge and roadway characteristics. The first one included the variables bridge width, average daily traffic, length, speed, relative width, sight distance and the variables F_1 to F_{12} inclusive. In the second model, all of the variables of the first model were considered except that the variable F_{20}, which was the average of F_{11} and F_{12}, replaced these two variables. Tables 14 and 15 present the results. It can be seen from the results of the Factor Analysis for the first model that the first 6 factors accounted cumulatively for about 66% of the variability. The next 4 factors accounted for less than 6% of the variability each and the remaining seven factors together accounted for less than eleven percent of the variability of the data. Six of the eigen-values (see Appendix B) are greater than one and these six factors were retained.

It is seen from Table 14 that speed and F_4 and F_5 contributed heavily in the first factor while the loadings of relative width, F_2 and F_7 are of larger magnitude in Factor 2. The ADT and F_8 variables appeared to be important in the third factor whereas bridge width and sight distance contributed strongly to Factor 4. The variable F_9 has a high loading in Factor 5 and length is important in Factor 6. The variables F_6,

TABLE 14. Factor Loadings For Total Data with F_{11} and F_{12}

Variable	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6
Bridge W	0.15306	0.21023	-0.14944	0.78248	0.21985	0.09734
ADT	0.09752	-0.04829	-0.91343	-0.01441	0.02083	0.01463
Speed	0.76859	-0.01066	0.02414	0.06665	-0.03093	-0.03601
Length	0.10540	0.05974	0.03534	0.01389	-0.03888	-0.87433
Sight Dist.	0.01778	-0.15958	0.21174	0.77317	-0.08396	-0.04750
Relative W	-0.14160	0.93791	-0.01818	0.07348	0.09073	-0.03371
F_2	0.02678	0.93353	-0.01134	0.04514	0.05323	0.03623
F_3	0.12754	0.30612	-0.00084	0.44371	-0.29469	-0.10144
F_4	0.80377	0.15507	0.00572	0.03514	0.03587	-0.24583
F_5	0.67962	-0.16056	0.03754	0.09156	0.03177	0.10271
F_6	0.32864	0.13656	0.00765	0.16264	-0.35351	0.20088
F_7	-0.06069	-0.55599	0.00262	0.32858	0.47862	0.31844
F_8	0.08527	-0.09668	0.88395	0.01147	0.14421	-0.04549
F_9	0.06173	0.13033	0.21454	-0.06927	0.77040	0.11309
F_{10}	0.38117	-0.02269	0.54512	0.08926	0.08304	0.48977
F_{11}	0.35576	-0.03933	-0.02873	0.36545	-0.08246	0.24735
F_{12}	-0.15393	-0.38958	0.18224	-0.28536	-0.44448	0.18131

TABLE 15. Factor Loadings For Total Data with F_{20}

Variable	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6
Bridge W	0.18204	0.19540	-0.15408	0.80143	0.11623	0.16213
ADT	0.11197	-0.04832	-0.91806	-0.00724	0.01475	0.03024
Speed	0.76618	-0.01293	0.02562	0.06218	-0.11591	-0.01553
Length	0.12622	0.03845	0.05340	0.01325	0.03150	-0.85180
Sight Dist.	0.03691	-0.15311	0.18464	0.80762	-0.10419	-0.03074
Relative W	-0.11255	0.93723	-0.02747	0.11101	0.13076	-0.03048
F_2	0.04080	0.93232	-0.01798	0.07068	0.06659	0.03286
F_3	0.11040	0.30001	-0.01085	0.44026	-0.28522	-0.15507
F_4	0.83029	0.14508	0.00066	0.06077	0.02724	-0.20567
F_5	0.68873	-0.16314	0.02670	0.10755	-0.03323	0.12422
F_6	0.25960	0.14112	0.02281	0.09347	-0.45280	0.14105
F_7	-0.03897	-0.57218	0.01966	0.32202	0.36651	0.39951
F_8	0.08898	-0.10071	0.89179	0.01370	0.11487	-0.01698
F_9	0.15250	0.12157	0.21520	0.00470	0.72817	0.25026
F_{10}	0.35887	-0.01272	0.54702	0.07211	-0.05705	0.52323
F_{20}	0.11462	-0.26722	0.10620	0.03830	-0.52250	0.31047

F_{10}, F_{11} and F_{12} contribute little to the first six factors.

Adding variable F_{20} instead of F_{11} and F_{12} did not result in F_{20} contributing substantially since none of the loadings of F_{20} in any of the six factors were large, as illustrated by the factor loadings in Table 15.

It is noted that some of the continuous variables are related to the F's because some of the F's are functions of bridge measurements. Most of the subjective variables F_3, F_{10}, F_{11}, and F_{20} did not indicate a strong contribution to the variability but they are still important because they can be more easily improved than bridge geometrics like bridge width, and may help to reduce traffic speed.

Testing for Normality of the Variables

Since discriminant analysis assumes normality of the independent variables, several variables were tested for normality with the univariate procedure of SAS (80) using the characteristics of the 78 bridges.

TABLE 16. Results of the Normality Tests

Variable	Reject or Not Reject Normality
Bridge width	Reject
ADT	Reject
Relative width	Reject
Speed	Reject
Shoulder reduction (%)	Reject
Sight distance	Reject
F_1	Reject

It is seen from Table 16 that all of the variables are non-normal.

Discriminant Analysis

Discriminant analysis was used to determine the distinguishing characteristics between the 26 less-safe bridges and the 52 more-safe bridges determined by cluster analysis. Some of the variables are non-normal and most of the F's are discrete and hence violate an assumption of discriminant analysis which assumes that the independent variables are normally distributed. Numerous combinations of the variables were tried as well as a stepwise discriminant procedure, some of which are discussed below. The goodness of the model is indicated by the percent of bridges classified correctly. In most combinations of variables in these models, the covariance matrices within each of the two groups of bridges were not equal and a quadratic model was more appropriate. The model with variables F_3, F_5, F_6, F_9, F_{10}, F_{11}, F_{12}, ADT, length, relative width, bridge width, shoulder reduction, speed, and sight distance yielded a quadratic discriminant function that classified 100% of the less-safe and 96.15% of the more-safe bridges correctly. It classified the bridges very well with only 2 bridges being misclassified. This function gives an equation having 113 terms. It is noted that more coefficients were determined than the number of data points (bridges). Also, the coefficients in the quadratic function require constraints to be imposed on them so that they will always have the appropriate sign from a practical viewpoint. The fourteen variable quadratic discriminant model will need 14 constraints on it which is not very practical. When the number of variables were reduced most models still needed a quadratic discriminant function.

When the above 14 variable model was approximated by a linear function it classified 57.69% of the less-safe bridges correctly and 92.31% of the more-safe bridges correctly. But the coefficients of several of the F-variables as well as relative width and sight distance were not in conformity with a model for safety index.

A quadratic model eliminating all the F-factors from the 14 variable model and including only the 7 continuous variables classified 84.62% of the less-safe bridges and 82.69% of the more-safe bridges correctly. When it was reduced to a linear discriminant function, it could classify only 50% of the less-safe bridges and 92.31% of the more-safe bridges. In addition the coefficients of sight distance and relative width were not satisfactory.

A quadratic model with the variables bridge width, length, average daily traffic, speed, F_6, F_7 and F_9 classified 65.38% of the less-safe bridges correctly and 88.46% of the more-safe bridges correctly. When it was reduced to a linear function, it could classify correctly only 42.3% of the less-safe bridges correctly with F_7 and ADT having unacceptable signs in the linear model.

Because of the difficulty of applying constraints to the quadratic discriminant function and because the assumption of normally distributed variables on which discriminant analysis is based is not satisfied, other methods were investigated.

Logistic Regression

It was observed that discriminant analysis requires the assumptions of multivariate normal distribution which is not true with respect to most of the variables.

Logistic regression was considered since it yields the degree of safety directly, is a good model for 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 and must conform to what is known by experience to be safe. After many trials the following model was accepted as the best possible.

TABLE 17. Logistic Regression Model of Bridge Safety

Variable X_i	Coefficient	Prob.	R
Intercept	-1.78999897	.3936	
Bridge width	0.44123886	.0000	0.294
ADT	-0.10753546	.0055	-0.080
Speed	-0.24633482	.0000	-0.207
Length	-0.00101675	.0000	0.260
F_9	0.95457213	.0000	0.204
F_6	0.56696522	.0000	0.212
F_7	0.33232235	.0014	0.095

The model is given by Probability of Safety = exp(y)/1+ exp(y)

where y = [-1.78999897 + 0.44123886 (Bridge width)

-0.10753546 (Average Daily Traffic)

-0.2463382 (Speed)

-0.00101675 (Length) + 0.95457213 (F_9)

+0.56696522 (F_6) + 0.33232235 (F_7)]

The equation obtained has an 'R' of 0.624. This 'R' is not the same as but is akin to the multiple correlation coefficient. The fraction of concordant pairs of predicted probabilities and responses is 0.906 out of a maximum possible value of 1.0. The rank correlation between predicted probability and actual probability is 0.81, which indicates the goodness of the model. All of the 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 the R-values in

Table 17, it is apparent that bridge width is the most important variable in the determination of safety index, the second most important variable is bridge length. ADT and F_7 are less important variables.

Sensitivity Analysis of the Logistic Model

The model is sensitive to changes in the variables as explained below. The model is of the form

$$P = \frac{\exp(y)}{1 + \exp(y)}$$

$$y = \beta_0 + \Sigma \beta_i x_i$$

the partial derivative of the probability of safety, P, with respect to a variable x_i will have the sign of its regression coefficient, β_i. For a given characteristic of a bridge, x_i, the sensitivity of the Safety Index is observed as the x_i are varied. An example of sensitivity analysis is given below.

In Table 18 the basic observation (observation 1) has the values 30 feet for bridge width; 4300 vehicles per day as Average Daily Traffic (KADT=4.3); speed is 50 mph; bridge length as 125 feet and F_9, F_6, F_7 are equal to 3, 1 and 1 respectively and the bridge has a Safety Index of 0.909202. Changes in the probability are noted when one of the variables are increased or decreased with the remaining variables being held constant.

TABLE 18. Sensitivity Analysis

OBS	Bridge ID	Bridge W	KADT	Speed	Length	F_9	F_6	F_7	Safety Index
1	C2	30	4.3	50	125	3	1	1	0.909202
2	C2	32	4.3	50	125	3	1	1	0.960320
3	C2	30	5.3	50	125	3	1	1	0.899926
4	C2	29	4.3	50	125	3	1	1	0.865611
5	C2	30	3.3	50	125	3	1	1	0.917698
6	C2	30	4.3	50	1125	3	1	1	0.783670
7	C2	30	4.3	50	50	3	1	1	0.915304
8	C2	30	4.3	50	125	3	5	1	0.989766
10	C2	30	4.3	50	125	5	1	1	0.985415
11	C2	30	4.3	50	125	2	1	1	0.794025
12	C2	30	4.3	50	125	3	2	1	0.946389
13	C2	30	4.3	20	125	3	1	1	0.999938
14	C2	30	4.3	60	125	3	1	1	0.460231
15	C2	30	4.3	50	125	3	1	5	0.974250
16	C2	30	4.3	50	125	3	1	3	0.951134

In Observation 2 an increase of bridge width (widening the road) by 2 feet increases the safety to 0.960320 and decreasing it to 29 in Observation 4, decreased the safety index to 0.865611. When ADT is increased to 5.3 in Observation 3, the probability of safety decreased to 0.899926 while the probability of safety increased to 0.917697 when the ADT is decreased to 3.3 in Observation 5. The model is very sensitive to changes in bridge length. When length is increased to 1125 in Observation 6, the probability of safety is reduced to 0.783670 while it increased to 0.915303 when the length is decreased to fifty feet in Observation 7. It is similarly observed that when F_6 is increased to 5 from 1 the safety index increases to 0.989766 and when F_6 is increased to 2 the safety index is raised to 0.946389 from the basic value of 0.909202 showing that improved grade continuity leads to improved safety. Similar results can be shown with regard to F_7 and F_9. A decrease of speed to 20 mph in Observation 13 gives a safety probability of 0.999938 making the bridge appear to be very safe.

The logistic model appears to be the most acceptable model for bridge safety and has good sensitivity with respect to each of the seven variables that are included in the model.

Sensitivity Index

Table 19 was obtained by changing each of the variables by 10%, one at a time. The Sensitivity Index may be defined as the change in Safety Index in percent divided by the change in a given variable in percent. Table 20 gives the sensitivity index for each of the variables and ranks them.

Table 20 shows that speed is the most sensitive variable followed by bridge width, traffic mix (F_9), grade continuity (F_6), ADT, shoulder reduction (F_7), and bridge length in that order.

TABLE 19. Data for Sensitivity Index

Observation	Bridge W	Length	Speed	ADT	F_6	F_7	F_9	Safety Index
1	30	125	50	4.30	1.0	1.0	3.0	0.909202
2	33	125	50	4.3	1.0	1.0	3.0	0.974110
3	30	125	50	4.73	1.0	1.0	3.0	0.905312
4	30	125	55	4.3	1.0	1.0	3.0	0.745026
5	30	125	50	4.3	1.1	1.0	3.0	0.913775
6	30	125	50	4.3	1	1.1	3.0	0.911909
7	30	125	50	4.3	1	1	3.3	0.930235
8	30	137.5	50	4.3	1	1	3	0.908148

TABLE 20. Sensitivity Index

Variable	Sensitivity Rank	% change of variable	Sensitivity Index % change in Safety Index / % change of variable	
		+	+	-
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

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

Using the data from 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 Ø.52 with accident rate as the depedent variable. All of the variables were significant except F_7 and F_9. It is of interest to note that Tseng's model consisted of 12 variables and was developed using the bridge and roadway characteristics of the 78 bridges alone with accident rate as the dependent variable. Out of the 12 variables, only 2 variables, F_1 and F_6 were significant in his model. In view of the fact that R^2 increases with the number of variables included, the present model can be considered an improvement.

A linear discriminant model with all of the same variables classified 42.3% of the unsafe bridges correctly. A quadratic discriminant model with the same variables classified 65.38% of the unsafe bridges as unsafe correctly and 88.46% of the safe bridges correctly. The logistic regression model with the same variables had a rank correlation (between predicted and observed probabilities) of Ø.81 out of a maximum possible value of 1.Ø. 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 a number that is bounded between Ø and 1.

The above discussion indicates that the variables used in the model do have a substantial relationship to accident rate. Accidents are also related to vehicle and driver variations which could not be considered due to lack of information. Discriminant analysis, although a robust procedure in general, did not yield a linear model that is interpretable with good classification results. Since some of the variables are non-normal, logistic regression is preferred (1ØØ) and the final model was developed with a logistic function.

Comparison of Safety Index Obtained with the Previous Safety Indices

Table 21 gives the correlation matrix of BSI One (Ivey), BSI Two (Tseng), Safety Index obtained by Logistic Regression (BSI Three) and Accident Rate.

TABLE 21. Correlation Coefficients

	Accident Rate	Safety Index	BSI One	BSI Two
Accident Rate	1.ØØØØØ	-Ø.53115	-Ø.22882	-Ø.2Ø43Ø
Safety Index	-Ø.53115	1.ØØØØØ	Ø.38Ø19	Ø.4ØØ16
BSI One	-Ø.22882	Ø.38Ø19	1.ØØØØØ	Ø.71371
BSI Two	-Ø.2Ø43Ø	Ø.4ØØ16	Ø.71371	1.ØØØØØ

It should be noted that the Safety Index obtained in this report has a much higher correlation of -0.53115 with accident rate as against correlations of -0.22882 of BSI one and -0.20430 of BSI Two. The Safety Index (BSI Three) shows an improvement of 265% in its correlation with accident rate. All three of the BSI's have significant correlation with accident rate and with each other.

CHAPTER 5

SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

SUMMARY

As a preliminary investigation of the factors that contribute to bridge safety, several statistical analyses were conducted on Tseng's data which were already assembled in a computerized data base (77). Regression and correlation analyses confirmed Tseng's observation that the factors F_1 (clear bridge width) and F_6 (grade continuity in percent) contributed most to a prediction of the accident rate at bridges. Factor analysis revealed that factors F_4 (ratio of approach sight distance to 85% approach speed), F_9 (traffic mix) and F_2 (ratio of bridge lane width to approach lane width) are also contributing factors but appear to be less important than the two mentioned above, F_1 and F_6.

The preliminary investigation showed that more information and additional data were needed in order to develop an improved bridge safety index model. The F-factors were numbers between Ø and 5 and some of them were obtained from ratios of two variables. Because, the actual measurements of bridge and approach roadway geometrics are more amenable to parametric methods in statistical analyses, additional data were collected for several variables such as actual bridge width and length, and information was obtained for each of the 655 accidents during the years 1974 to 1979 at the 78 bridge sites. Additional data concerning the environmental conditions at the time of the accident, type of traffic control, alignment curvature, number injured, and number killed were also collected to create a new computer data set.

Detailed statistical analyses were conducted on the data set. Analysis of variance indicated that the accident rate did not depend upon whether the bridge is in a rural or urban environment. Regression and correlation analyses indicated that bridge width, length, average daily traffic, sight distance, and grade continuity were important factors in predicting accident rate. Multicollinearity was not observed in these analyses. Environmental factors did not appear to contribute significantly to accident rate. Factor analysis indicated that the variables F_6 (grade continuity), F_{10} (distractions near bridge site), F_{11} (paint markings), and F_{12} (warning signs) were less important than the others.

After understanding the interrelationships of the variables, an attempt was made to develop an improved model that would include as many of the statistically significant variables as possible as well as the subjective F-factors such as F_3 (guard rail condition), F_6 (grade continuity), and F_{12} (signs) which have been known by engineers to be important in reducing accident rates. Some of the subjective factors are more easily corrected and improved than the bridge geometrics such as the length of the bridge. The bridges were divided objectively into two groups, one more safe than the other, using cluster analysis on the observed accident rates. Discriminant and logistic regression procedures were used on the cluster groups to develop an equation that would classify a bridge into one of the two groups (more-safe group or less-safe group). Discriminant analysis, did not yield an easily interpretable linear model. Logistic regression, which does not assume that the variables are normally distributed as does discriminant analysis, yielded a safety index directly. The final model is given by:

$$\text{Safety Index} = \exp(y)/1+\exp(y) \qquad (10)$$

where

$$y = -1.79+0.44\ (\text{Bridge Width}) - 0.11\ (\text{ADT}) - 0.25\ (\text{Speed}) - 0.001\ (\text{Bridge Length}) + 0.95\ (F_9) + 0.57\ (F_6) + 0.33\ (F_7)$$

where

F_6 = grade continuity, factor,

F_7 = shoulder reduction, factor,

F_9 = traffic mix.

CONCLUSIONS

The conclusions reached in this study are as follows:

1. An enhanced BSI model was developed through a statistical approach using standard statistical analyses, classification and discrimination techniques bearing in mind at all times the experience and expertise of the bridge engineers as revealed by previous studies. The model is considered enhanced because: (i) the bridges were divided into "more safe" and "less safe" groups objectively with the use of cluster analysis and not arbitrarily as in previous studies; (ii) the model uses only 7 independent variables as against 12 variables in the previous BSI model; (iii) with approximately only half of the number of variables the new model yields more than twice the R^2 compared to the previous model which is considered to be an excellent improvement; (iv) the Safety Index developed yields a higher correlation coefficient of -0.53 with accident rate as compared with a correlation of -0.20 of the previous model; and (v) the logistic model used does not require the assumption that the variables are normally distributed as with discriminant analysis. It was observed that most of the variables in this data set are non-normal.

2. The final model fits the data well as 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 because of this can be used readily to establish priorities for improvement or repairs of bridges.

4. The important variables that are related to accident rate appear to be the bridge width, bridge length, vehicle speed, traffic mix and grade continuity. The final model adopted gives the probability that a bridge is safe 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 average daily traffic.

5. The model is sensitive to improvements in bridge conditions and results in higher probabilities of safety when the above mentioned factors are improved.

6. According to the model, bridge safety may be improved by decreasing vehicle speed. This can be done by posting appropriate signs at and before the bridge or perhaps by using texture or rumble strips for speed reduction. Although the other factors of bridge width, length, shoulder reduction, grade continuity, average daily traffic and traffic mix are not as 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.

Recommendations for Future Investigations

It is possible to make further improvements in the Bridge Safety Index that is reported here. A better model will allow more accurate determinations of the cost effectiveness of the use of funds that are allocated for bridge safety improvement. The following recommendations will result in an improved Bridge Safety Index.

1. Most of the bridges in the data set given in this report have bridge widths greater than 24 feet. It is necessary to collect more data on bridges having widths less than 24 feet. A better model may result as a consequence.
2. More comprehensive accident data for each bridge site as well as actual measurements of all variables involved in the F-factors should be collected. This would allow the analysis to use the constituent variables of the ratios directly.
3. No information was readily available about the type of injuries (incapacitating, nonincapacitating, or possible injury). If more information can be collected on these aspects, 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.

4. It is considered essential to obtain another carefully taken sample of bridges to validate the conclusions and the model developed in this study. Such a sampling plan should include information on the bridge characteristics of length, bridge width, relative width (the difference between bridge width and approach roadway width) and average daily traffic. This information can be obtained from the BRINSAP files. If three levels of each of these variables are considered and two replicates are included for each variable combination, 162 bridges are needed. If an additional 50 bridges are included for validating the developed model (a total of 212 bridges), then, the statistical reliability of the model would be improved.
5. If, in a future model, sight distance and relative width can be included as additional variables, the resulting safety index may yield a higher correlation with accident rate.

REFERENCES

1. G. H. Oglesby and R. G. Hicks, Highway Engineering, John Wiley & Sons, New York, 1982.

2. "One in Six Highway Bridges is Deficient", Engineering News Record, Vol. 198, No. 10, 1977, pp. 18-21.

3. "Weak Bridges: Growing Hazard on the Highways", U.S. News and World Report, Vol. 84, No. 1, 1978, p. 72.

4. "Highway Hazards: What States Are Doing About Them", Better Roads, Vol. 48, No. 11, 1978, p. 26.

5. "What's Going On", Better Roads, Vol. 29, No. 8, 1959, p. 12.

6. "Better Roads Exclusive Bridge Inventory", Better Roads, Vol. 50, No. 5, 1980, p. 33.

7. K. S. Bowers, "Bridge Rehabilitation Programs", Traffic Quarterly, Vol. 34, No. 2, 1980, pp. 263-270.

8. "$51 Billion Trust Fund Extension", Better Roads, Vol. 48, No. 11, 1978, p. 15.

9. "Bridge Problems Still Exist, But Help Is On The Way", Better Roads, Vol. 48, No. 11, 1978, p. 18.

10. American Association of State Highway Officials, Standard Specifications for Highway Bridges, Washington, D.C., AASHO, 1973.

11. American Association of State Highway Officials, A Policy on Geometric Design of Rural Highways, Washington, D.C., AASHO, 1965.

12. American Association of State Highway Officials, A Policy on Design Standards for Stopping Sight Distance, Washington, D.C., AASHO, 1971.

13. American Association of State Highway Officials, Manual for Maintenance Inspection of Bridges, Washington, D.C., AASHO, 1970.

14. W. J. Wilkes, "Bridges and Safety", American Association of State Highway Officials, Committee Proceedings, 1967.

15. A. M. White, "Safety and Bridges", American Association of State Highway Officials, Committee Proceedings, 1965.

16. "The National Highway Safety Needs Study", U.S. Department of Transportation, Washington, D.C., March 1976.

17. "The National Highway Safety Needs Report", U.S. Department of Transportation, Washington D.C., April 1976.

18. "Bridge Inspector's Training Manual", U.S. Department of Transportation, Washington D.C., 1979.

19. "An Evaluation of the Highway Safety Program", A Report to the Congress from the Secretary of Transportation, U.S. Department of Transportation, Washington D.C., July 1977.

20. "Narrow Bridges - Driver Dilemmas", Hearing before the Subcommittee on Investigations and Review of the Committee on Public Works, Ninety-Third Congress, Washington D.C., June 1973.

21. D. L. Ivey, R. M. Olson, N. E. Walton and G. D. Weaver, "Safety at Narrow Bridge Sites", NCHRP Report No. 203, Transportation Research Board, Washington D.C., June 1979.

22. L. R. Calcote and K. K. Mabe, "Evaluation of the Effectiveness of Measures for Reducing Accidents and Accident Severity at Highway Narrow Bridge Sites", Vol. I, Report No. SWR1-03-4941, FHA, Washington D.C.

23. W. C. Johnson, "An Analysis of Driver Behavior at Narrow Bridges", Ph.D. Dissertation, Texas A&M University, 1977.

24. M. H. Hilton, "Some Case Studies of Highway Bridges Involved in Accidents", HRR, No. 432, 1973.

25. H. E. Gunnerson, "Iowa Narrow Bridge Accident Study", Highway Research Abstracts, Vol. 31, No. 7, 1961, pp. 23-24.

26. M. S. Raff, "Interstate Highway Accident Study", Highway Research Board, Bul. 74, Publication 265, 1953, pp. 18-43.

27. M. D. Shelby and P. R. Tutt, "Vehicle Speed and Placement Survey", Highway Research Board, Bul. 170, 1958, pp. 33-39.

28. J. V. Brown and J. Foster, "Bridge Accidents on Rural Highways in New Zealand: Analysis and Appraisal", ARRB Proceedings, Vol. 3, Part 1, 1966, pp. 638-646.

29. W. P. Walker, "Influence of Bridge Widths on Transverse Positions of Vehicles", HRB Proceedings, Vol. 21, 1941, pp. 361-365.

30. L. E. King and R. W. Plummer, "Lateral Vehicle Placement and Steering Wheel Reversals on a Simulated Bridge of Variable Width", HRR, No. 432, 1973, pp. 61-73.

31. A. Taragin and H. G. Eckhardt, "Effects of Shoulders on Speed and Lateral Placement of Motor Vehicles", HRB Proceedings, Vol. 32, 1953, pp. 371-378.

32. A. Taragin, "Driver Behavior as Related to Shoulder Type and Width on Two-Lane Highways", HRB, Bul. 170, 1958, pp. 54-55.

33. D. F. Huelke and P. W. Gikas, "Non-Intersectional Automobile Fatalities - A Problem in Roadway Design", HRR, No. 152, pp. 103-104.

34. R. R. Roberts, "Effect of Bridge Shoulder Width on Traffic Operational Characteristics", TRR, No. 601, 1976, pp. 73-74.

35. R. A. Cunard, D. D. Perkins and T. K. Dalta, "Computerized Inventory and Priority Analysis for Roadside Obstacles", TRR, No. 679, 1978, pp. 10-12.

36. T. C. Hopkins and R. C. Deen, "The Bump at the End of the Bridge", HRR, No. 302, 1970, pp. 72-76.

37. W. M. Moore, "Detection of Bridge Deck Deterioration", HRR, No. 451, 1973, p. 53.

38. D. G. Manning, "Effects of Traffic-Induced Vibrations on Bridge Deck Repairs", NCHRP, No. 86, 1981.

39. G. Venkatesulu, P. Kumar and S. C. Motwani, "Inspection and Maintenance of Highway Bridges", Indian Concrete Journal, Vol. 55, No. 1, 1981, pp. 5-21.

40. R. Archibald, "Significant Changes in Modern Bridge Design", Roads and Sheets, Vol. 90, No. 2, 1947, pp. 57-58.

41. M. H. Godwin, "Small Spans Replace Old Bridges", Better Roads, Vol. 18, No. 9, 1948, pp. 25-26.

42. L. C. Lundstrom, P. C. Skeels, B. R. Englund, and R. A. Rogers, "A Bridge Parapet Designed for Safety," HRR, No. 83, 1965.

43. W. Zuk, H. H. Newlon, and W. T. McKeel, "Methods of Modifying Historic Bridges for Contemporary Use", Report No. FHWA-VA-80-48, 1980.

44. J. N. Cleary, "New Highway Bridge Dimensions for Safety", Transportation Engineering Journal, Vol. 97, No. TE2, 1971, pp. 371-381.

45. N. J. Deleys and R. R. McHenry, "Highway Guardrails - A Review of Current Practice", NCHRP No. 36, 1967.

46. J. D. Michie and L. R. Calcote, "Location, Selection, and Maintenance of Highway Guardrails and Median Barriers", NCHRP No. 54, 1968.

47. J. D. Michie, L. R. Calcote, and M. E. Bronstad, "Guardrail Performance and Design", NCHRP No. 115, 1971.

48. J. D. Michie and M. E. Bronstad, "Guardrail Crash Test Evaluation - New Concepts and Designs", NCHRP No. 129, 1972.

49. R. M. Olson, E. R. Post, and W. F. McFarland, "Tentative Service Requirements for Bridge Rail Systems", NCHRP No. 86, 1970.

50. R. M. Olson, D. L. Ivey, E. R. Post, R. H. Gunderson, and A. Cetiner, "Bridge Rail Design-Factors, Trends and Guidelines", NCHRP No. 149, 1974.

51. J. D. Michie and M. E. Bronstad, "Location, Selection, and Maintenance of Highway Traffic Barriers", NCHRP No. 118, 1971.

52. E. F. Nordlin, J. H. Woodstrom, R. P. Hackett, and J. J. Folson, "Dynamic Tests of the California Type 20 Bridge Barrier Rail", HRB No. 343, 1971, pp. 57-66.

53. G. W. Walker and C. Y. Warner, "Crash Test Evaluation of Strong-Post, Energy-Absorbing Guardrail Using a Lapped W-Beam for Transitions and Median Barriers", HRB No. 386, 1972, pp. 78-87.

54. J. L. Beaton, E. F. Nordlin, and R. N. Field, "Dynamic Tests of Corrugated Metal Beam Guardrail", HRB No. 174, 1967.

55. E. R. Post, P. T. McCoy, T. J. Wipf, and R. J. Ruby, "Cost-Effectiveness of Guardrail-Bridgerail Transition Improvements: Double W-Beam Versus Decreased Post Spacing", TRR No. 769, 1980, pp. 37-43.

56. W. W. Hunter, F. M. Council, A. K. Dutt, D. G. Cole, "Methodology for Ranking Roadside Hazard Correction Programs", TRR No. 672, 1978.

57. D. L. Woods, B. Bohuslav, and C. J. Keese, "Remedial Safet Treatment of Narrow Bridges", Traffic Engineering, Vol. 46, No. 3, 1976, pp. 11-15.

58. J. C. Laughland, L. E. Haffner, J. W. Hall, and D. R. Clough, "Methods for Evaluating Highway Safety Improvements", NCHRP No. 162, 1975.

59. J. C. Glennon, "Roadside Safety Improvement Programs on Freeways", NCHRP No. 148, 1974.

60. R. Jorgenen Associates, Inc., "Cost and Safety Effectiveness of Highway Design Elements", NCHRP No. 197, 1978.

61. The National Research Council, "Bridges on Secondary Highways and Local Roads", NCHRP No. 22, 1980.

62. University of Virginia Civil Engineering Department and Virginia Department of Highways and Transportation, "Rehabilitation and Replacement of Bridges on Secondary Highways and Local Roads", NCHRP No. 243, 1981.

63. American Association of State Highway Officials and The Bureau of Public Roads, "Bridge Approach Design and Construction Practices", NCHRP No. 2, 1969.

64. T. H. Tamburri, C. J. Hammer, Jr., J. C. Glennon, and A. Lew, "Evaluation of Minor Improvements", HRR No. 257.

65. "Low-Level Lighting for Bridges is Tried Out", Better Roads, Vol. 29, No. 3, 1959.

66. J. M. Dale, "Pavement Marking - Danish Style", Better Roads, Vol. 40, No. 2, 1970, pp. 29-31.

67. J. L. Beaton and H. A. Rooney, "Raised Reflective Markers for Highway Lane Lines", HRR No. 105.

68. D. S. Turner, D. B. Fambro, and R. O. Rogness, "Effects of Paved Shoulders on Accident Rates for Rural Texas Highways", TRR No. 819.

69. Federal Highway Administration, "Manual on Uniform Traffic Control Devices for Streets and Highways", U.S. DOT, 1978.

70. J. S. Kosiol, Jr., "Maine Facility Research Summary: Dynamic Sign Systems for Narro Bridges", Report No. FHWA-RD-78-33, 1978.

71. G. F. King and H. Lunenfeld, "Development of Information Requirements and Transmission Techniques for Highway Users", NCHRP No. 123, 1971.

72. J. Behnam and J. G. Laguros, "Accidents and Roadway Geometrics at Bridge Approaches", Public Works, Vol. 104, No. 4, 1973, pp. 66-70

73. J. K. Kihlberg and K. J. Thorp, "Accident Rates as Related to Design Elements of Rural Highways", NCHRP No. 47, 1968.

74. J. A. Deacon, C. V. Zegeer, and R. C. Deen, "Identification of Hazardous Rural Highway Locations", TRR 543, 1975, pp. 16-33.

75. D. S. Turner, "A Technique to Select Bridge Safety Treatments", Ph.D. Dissertation, Texas A&M University, 1980.

76. T. M. Newton, "A Method of Field Evaluation of Narrow Bridges for Priority Indexing", Report No. FHWA-TX-79-20+233-1, 1979.

77. K. Tseng, "Evaluation of Bridge Safety Index in Texas", Unpublished Report at Texas A&M University, 1982.

78. F. Luyanda, R. W. Smith, M. Padron, P. Resto, J. Gutierrez, J. Haddock, and L. Fernandez, "Multivariate Statistical Analysis of Highway Accidents and Highway Conditions", Draft Phase Report for U.S. Department of Transportation, 1982.

79. L. R. Calcote and K. K. Mak, "Evaluation of the Effectiveness of Measures for Reducing Accidents and Accident Severity at Highway Narrow Bridge Sites", Vol. I, Final Report No. SWRI-03-4941.

80. Barr, A. J., Goodnight, J. H., Sall, J. P., and Helwig, J. T., A User's Guide to SAS 79, SAS Institute, Inc., Sparks Press, Raleigh, North Carolina, 1979.

81. Rudolf J. Freund and Ramon C. Liltell, "SAS for Linear Models", SAS Institute Inc., Box 8000, Cary, N.C. 27511-8000, 1981.

82. John Neter, William Wasserman, Michael H. Kutner, "Applied Linear Regression Models", Richard Irwin Inc., Homewood, Illinois, 1983.

83. Douglas C. Montgomery and Elizabeth A. Peck, "Introduction to Linear Regression Analysis", John Wiley & Sons, New York.

84. Samprit Chatterjee and Bertram Price, "Regression Analysis by Example", John Wiley & Sons, New York, 1977.

85. Andrew L. Comrey, "A First Course in Factor Analysis", Academic Press, N.Y., 1973.

86. Harry H. Harman, "Modern Factor Analysis", University of Chicago Press, 1960.

87. Richard A. Johnson and Dlan W. Wichern, "Applied Multivariate Statistical Analysis", Prentice Hall Inc., Eaglewood Cliffs, New Jersey, 1982.

88. Anderberg, M. R., "Cluster Analysis for Applications", Academic Press, New York, 1973.

89. A. D. Gordon, Classification, Chapman and Hall, New York, 1981.

90. Elisa T. Lee, "Statistical Methods for Survival Data Analysis", Lifetime Learning Publications, Belmont, California, 1980.

91. Frank Harrell, "The Logist Procedure", SAS Supplemental Library User's Guide, 1980 Edition, SAS Institute Inc., Cary, N.C., 1980.

92. Frank Harrell, "Logist Procedure: Support Type A", Box 3337, Duke University Medical Center, Durham, North Carolina, 1983.

93. Fisher, R. A., "Statistical Methods for Research Workers", Hasney Publishing Company Inc., New York, 1948.

94. James Press S., "Applied Multivariate Analysis", Holt, Rinehart and Winston, Inc., New York, 1979.

95. T. W. Anderson, "An Introduction to Multivariate Statistical Analysis", John Wiley and Sons, Inc., New York, 1958.

96. P. A. Lachenbrunch, "Discriminant Analysis", Hafner Press, New York, 1975.

97. R. A. Eisenbeis and R. B. Avery, "Discriminant Analysis and Classification Procedures", Lexington Books, Massachusetts, 1972.

98. James D. Kuoke, "Discriminant Analysis with Discrete and Continuous Variables", Biometrika 38, 191-200, 1982.

99. Yvonne M. M. Bishop, Stepen E. Frenberg, and Paul W. Holland, "Discrete Multivariate Analysis, Theory and Practice", The MIT Press, Cambridge, Massachusetts, 1975.

100. S. J. Press and Wilson, "Choosing Between Logistic Regression and Discriminant Analysis", Journal of American Statistical Association, 73, pp. 699-705, 1978.

101. Bridge Inventory, Inspection and Appraisal Program (BRINSAP), Manual of Procedures, State Department of Highways and Public Transportation, 1977.

102. D. A. Belseley, E. Kuh, and R. E. Welsch, "Regression Diagnostics", John Wiley and Sons, New York, 1980.

103. A. M. Mood, F. A. Graybill, and D. C. Boes, Introduction to the Theory of Statistics, McGraw-Hill Book Company, New York, 1974.

104. L. H. Koopmans, An Introduction to Contemporary Statistics, Duxberry Press, Boston, Massachusetts, 1981.

105. W. W. Daniel, Applied NonParametric Statistics, Houghton Mifflin Company, Boston, 1978.

106. S. S. Shapiro and M. B. Wilk, "An Analysis of Variance Test for Normality (Complete Sampler), Biometrika, Vol. 52, 1965, pp. 591-611.

107. N. H. Nie et.al., Statistical Package for the Social Sciences, McGraw-Hill Book Company, New York, 1975.

APPENDICES

APPENDIX A

REGRESSION AND CORRELATION ANALYSIS AND ALLIED CONCEPTS AND PROCEDURES

Regression analysis may broadly be defined as a statistical technique for analyzing and modeling the relationship between variables. The relationship is often expressed in the form of an equation, the simplest example being the linear regression model

$$y = \beta_0 + \beta_1 x + \varepsilon \qquad \text{(A-1)}$$

with one independent variable, where

y = dependent variable

x = independent variable

ε = statistical error, and

β_0, β_1 = unknown constants.

To avoid confusion with concepts of statistical independence, the x's are often referred to as regressor or predictor variables and y as the response variable. The multiple linear regression model with more than one regressor can be written as

$$y = \beta_0+\beta_1 x_1+\beta_2 x_2+\ldots+\beta_k x_k+ \varepsilon \qquad \text{(A-2)}$$

with similar notation as Equation (A-1). The above model is said to be linear because it is linear in the parameters β_1, $\beta_2 \ldots \beta_k$ and not because y is related linearly to the x's. Estimation of the unknown parameters (β's) is one of the most important objectives of regression analysis.

Limitations of Regression models:

The usual assumptions of classical regression given in all standard text books need to be valid for the model to be effective. In addition, it should be noted that the regression equation is an approximation to the true relationship between variables and is valid only over the region of the regressors in the given data. Outliers or bad values in data should be investigated as they can disturb the fit of the least squares regression model. The slope of the line in the linear regression model is strongly influenced by remote values of x though all points have equal weight in determining the height of the line. Even when a regression analysis indicates a strong relationship between two variables, expectation of discovering cause and effect relationships is very limited.

Polynomial Regression

Polynomial Regression with one independent variable is given by

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \ldots + \beta_n x^n \qquad \text{(A-3)}$$

A redefinition of variables makes a polynomial amenable to the usual linear regression procedures. A polynomial function is very useful since it is simply a sum of linear, quadratic, cubic or higher terms whose corresponding shapes are well known.

Coefficient of Multiple Determination (R^2)

R^2 measures the proportion of total variation in the dependent variable explained by the regression model. R^2 varies between 0 and 1 and is given by the equation

$$R^2 = \frac{SSR}{SST} \tag{A-4}$$

where

SSR = Sum of squared errors due to regression, and

SST = Total sum of squared errors

F-Statistic

In the model, the F-statistic is given by the ratio of the mean squared error due to regression to the mean squared error due to residual variation. A larger F-value leads to a higher statistical significance of the model.

Correlation Analysis

Correlation can be used to assess and test the linear relationship between any two variables x and y. The simple correlation 'r' between x and y is given by

$$r = \frac{\sum_{i=1}^{n} (x_i-\bar{x})(y_i-\bar{y})}{\left[\sum_{i=1}^{n} (x_i-\bar{x})^2 \sum_{i=1}^{n} (y_i-\bar{y})^2\right]^{1/2}} \tag{A-5}$$

when there are n pairs of observations of (x_i, y_i) and x and y are the arithmetic means of x's and y's.

The 'corr' procedure of the Statistical Analysis System (SAS) yields correlation coefficients between variables and significance probabilities. A high correlation between two regressors should be noted since using both of them together in a model may lead to multicollinearity problems. When there are several regressors in the model, none of the correlations may be very large.

Multicollinearity, Variance Inflation Factors and Condition Indices

Multicollinearity is a high degree of linear relationship among independent variables which makes interpretation of partial regression coefficients difficult. The existence of multicollinearity implies that the independent variables are strongly related to each other which makes it almost impossible to vary one of the variables while keeping the other variables constant. Two regressors are said to be orthogonal when there exists no linear relationship between them. Multicollinearity is said to exist when the regressors are nonorthogonal to each other. In other words when there is a linear or near linear dependence between two independent variables, the problem of multicollinearity is indicated (83).

The multiple regression model, in matrix notation, can be written as

$$\underline{y} = X\,\underline{\beta} + \underline{\varepsilon} \tag{A-6}$$

where

$\underline{y}$ = the vector of responses

X = the matrix of the independent variables

$\underline{\beta}$ = the vector of the unknown parameters, and

$\underline{\varepsilon}$ = the vector of random errors distributed normally and independently with a mean of zero and variance of σ^2.

Let it be further assumed that all the independent and dependent variables are scaled to unit length by the transformations

$$w_{ij} = \frac{x_{ij}-\bar{x}_j}{s_{jj}^{\frac{1}{2}}} \qquad \begin{array}{l} i = 1, 2, \ldots, n \\ j = 1, 2, \ldots, k \end{array} \tag{A-7}$$

and

$$\dot{y}_i = \frac{y_i-\bar{y}}{s_{jj}^{\frac{1}{2}}} \qquad i = 1, 2, \ldots, n \tag{A-8}$$

where S_{jj} is the corrected sum of squares of the regressor x_J. The regression model now reduces to the equation

$$\dot{y}_i = b_1 w_{i1} + b_2 w_{i2} + \ldots + b_k w_{ik} + \varepsilon_i$$
$$i = 1, 2, \ldots n$$

yielding the vector of least squares regression coefficient as

$$\underline{b} = (W'W)^{-1} W' \underline{\dot{y}} \tag{A-9}$$

W'W matrix is in the form of a correlation matrix (83). Let the

i^{th} column of the X matrix be denoted by $\underline{x}_i$ so that the matrix is

$$X = [\underline{x}_1 \underline{x}_2 \ldots \underline{x}_m].$$

The vectors $\underline{x}_1$, $\underline{x}_2, \ldots \underline{x}_m$ are said to be linearly dependent if there is a set of constants c_1, c_2, - c_m such that

$$c_1\underline{x}_1 + c_2\underline{x}_2 + \ldots c_m\underline{x}_m = 0 \qquad \text{(A-10)}$$

where not all c_i are equal to zero.

If the above equation is true then $(X'X)^{-1}$ does not exist. If it is approximately true for some subset of the column of X giving rise to near linear dependency, then there exists the problem of multicollinearity. It can be said that multicollinearity is a form of ill conditioning of the X'X matrix (83).

Multicollinearity Diagnostics (83)

Examination of the correlation matrix

Very often, an examination of the off diagonal terms in the X'X matrix gives indication of the existence of multicollinearity. The absolute value of r_{iJ} will be large if the variables x_i and x_j are nearly linearly dependent. Examining simple correlation indicates near linear dependency between two regressors only. When there are more than two regressors involved in a linear dependency none of the pairwise correlations may be large.

Variance Inflation Factors

The variance of β_i is given by

$$V(\beta_i) = c_{ii}\sigma^2 \qquad \text{(A-11)}$$

where

c_{ii} = a diagonal element of the matrix $c=(X'X)^{-1}$ and

σ^2 = the error mean square.

It can be shown that

$$c_{ii} = \frac{1}{1-R_i^2} \quad i = 1, 2, \ldots n$$

where R^2_i is the coefficient of multiple determination for regression of x_i on the remaining m-1 regressors. If there exists considerable multicollinearity between x_i and the remaining regressors then R_i is close to unity and $(1-R^2_i)$ is close to zero. This results in a very large c_{ii} and makes the variance of β_i given by Equation (11) become very inflated. $c_{ii} = (1-R^2_i)^{-1}$ is often called the Variance Inflation Factor (VIF). If VIF exceeds 5 or 10, multicollinearity is indicated.

Examination of Eigen Values and Singular Values

The eigenvalues of an n x n matrix X are the n roots of the equation $|X - \lambda I| = 0$ where I is the identity matrix (n x n). Each eigenvalue gives rise to a corresponding eigenvector. Some times the characteristic roots or eigenvalues of X'X (λ_i, i=1,2,...n) are used to measure the extent of multicollinearity in the data. If there are linear dependencies, then one or more eigenvalues will be small. Sometimes a condition number of X'X is defined by

$$K = \frac{\lambda_{maximum}}{\lambda_{minimum}} \tag{A-12}$$

where

$\lambda_{maximum}$ = the largest eigenvalue

$\lambda_{minimum}$ = the smallest eigenvalue

Condition numbers between 100 and 1000 imply moderate to strong multicollinearity and if K is greater than 1000, it implies severe multicollinearity.

Eigen system analysis also helps identify the nature of the linear dependencies in data. The X'X matrix can be written as

$$X'X = L\Delta L'$$

where is an m x m diagonal matrix whose main diagonal elements are the eigenvalues of X'X (λ_j, j=1,...m) and L is an m x m orthogonal matrix whose columns are the eigenvectors of X'X. If an eigenvalue

λ_i is very close to zero, indicating near linear dependency in the data, the elements of its associated eigenvector $\underline{\ell}_i$ (where $\underline{\ell}_1, \underline{\ell}_2 - \underline{\ell}_m$ are the columns of L) indicate the linear dependency from the equation when

$$\sum_{i=1}^{m} \alpha_i \underline{\ell}_i = 0 \qquad \text{(A-13)}$$

where ℓ_i are constants (i=1,2,...m)

A singular value decomposition can be done on the X matrix (nxm) by writing it in the form

X = VDL'

where

V = n x m matrix

L = m x m matrix, and

D = m x m diagonal matrix with non-negative diagonal elements μ_j (j=1,2,...m).

Then the μ_j are known to be the singular values of X. The value of the singular values reflects the malconditioning of x-matrix. There will be one small singular value for each near linear dependency (83).

$$y_i = \frac{\mu_{max}}{\mu_i} \qquad i=1,2,\ldots m \qquad \text{(A-14)}$$

are defined to be the condition indices of the X matrix by Belsley, Kuh and Welsch (102). The condition number of X is given by the largest value of y_i. A large condition index greater than 30 indicates multicollinearity.

The covariance matrix of β can be written as

$$V(\hat{\underline{\beta}}) = \sigma^2(X'X)^{-1} = \sigma^2(L\Delta L') \tag{A-15}$$

The variance of the i-th regression coefficient is the i-th diagonal element of this matrix given by

$$v(\hat{\beta}_i) = \sigma^2 \sum_{j=1}^{m} \frac{\ell^2_{ij}}{\mu^2_j} = \sigma^2 \sum_{j=1}^{m} \frac{\ell^2_{ij}}{\lambda_j} \tag{A-16}$$

where ℓ_{ij} is the typical element of the matrix L. Apart from σ^2, the i-th diagonal element of $L\Delta^{-1}L'$ is the i-th variance inflation factor $(VIF)_i$. Hence

$$(VIF)_i = \sum_{j=1}^{m} \frac{\ell^2_{ij}}{\mu^2_j} = \sum_{j=1}^{m} \frac{\ell^2_{ij}}{\lambda_j} \tag{A-17}$$

A small eigenvalue or singular value will inflate the variance of β_i. According to Belsley, Kuh and Welsch (102) variance decomposition proportions should be used to measure multicollinearity which are given by

$$\pi_{ji} = \frac{\ell^2_{ij}/\mu^2_j}{(VIF)_i} \qquad (i, j = 1, 2, \ldots, m) \qquad \text{(A-18)}$$

If the π_{ij} are arrayed in an m x m matrix Π then the elements of each column of Π represent the proportion of the variance of each β_j contributed by the i-th eigenvalue or singular value. Multicollinearity is suspected if a high proportion of the variance of two or more regression coefficients correspond to one small singular value or eigenvalue. Variance decomposition factors of more than 0.50, suggest multicollinearity.

If the F-statistic for the overall model is significant but the t-statistic for the individual β's is not significant, multicollinearity is sometimes indicated as well as when adding or deleting an independent variable produces significant changes in estimates of β_i. At times the signs of β_i are not what they are expected to be from a practical point of view, one has to look for multicollinearity among regressors.

Some Possible Reasons for Multicollinearity and Methods for Dealing with the Problem

Reasons

(i) Data collection methods sometimes lead to multicollinearity if only a subspace of the regressors defined by the equation

$\Sigma \ell_i x_i = 0$ is sampled.

(ii) At times constraints on the model or population being sampled cause multicollinearity irrespective of the sampling method employed.

(iii) Multicollinearity is often the result of the choice of the model as in the case of a polynomial regression model in which say x and x^2 both occur as regressor variables.

(iv) When there are more regressors than observations, multicollinearity may result. The source of the multicollinearity needs to be recognized before it can be removed.

Methods

(i) The best method of dealing with multicollinearity problems is suggested to be collection of additional data in a manner designed to break up the multicollinearity in existing data (83).

(ii) When two highly correlated regressors are used in the model resulting in multicollinearity, redefining the regressors may remove the problem. For example, if x_i, x_j and x_k are nearly linearly dependent one can use a function of the three variables as a regressor and still preserve the information contained in the original regressors. Another approach is to eliminate some of the variables if it is practical and desirable.

(iii) One of the methods suggested in the presence of multicollinearity is to use ridge regression or principal component regression instead of the method of least squares estimates.

Multicollinearity and Diagnostics in SAS (80)

The approach of SAS in the 'Proc Reg' procedure follows that of Belsley, Kuh and Welsch (102). "The 'collin' option on the model statement yields a collinearity analysis of the output. The X'X matrix is scaled so as to have ones on the diagonal. If 'collinoint' is requested at first, the intercept variable, β_0, is adjusted out at the beginning of the analysis. After this the eigenvalues and eigenvectors are extracted. The analysis is reported with eigenvalues of X'X instead of eigenvalues of X. The eigenvalues of X'X are the squares of the eigenvalues of the X matrix.

The condition indices are the square roots of the ratio of the largest eigenvalue to each individual eigenvalue. The largest condition index is the condition number of the scaled X matrix. When this number is large, the matrix is said to be ill conditioned. For each variable, the SAS option 'Reg' prints the proportion of the variance of the estimate accounted for by each principal component. A collinearity problem occurs when a component associated with a high condition index contributes strongly to the variance of two or more variables (80).

R-Square Procedure of SAS

This procedure does all possible regressions for one or more response variables and a set of regressors. The output consists of the R^2 value for each model starting with a single variable regression model and ending with the full model consisting of all of the regressors in the set (80).

Stepwise Regression Methods (80, 81, 82, 83, 84)

The R-Square Procedure of SAS can yield all possible regressions with the variables specified starting with a model with a single regressor and going up to a model with all the regressors included in the data. Since the R-Square procedure is expensive where there are many regressors, stepwise regression methods are considered which evaluate only a small number of subset regression models by either adding or deleting regressors one at a time. Stepwise procedures can be classified into the three broad categories: (i) forward selection (ii) backward elimination and (iii) stepwise regression which is a combination of (i) and (ii).

Forward selection begins with the best one variable equation and adds additional variables. The best one independent variable model is obtained by finding the regressor that has the largest sample correlation with the response variable y which will also have the largest F-statistics for testing the significance of regression. This regressor is entered if the F-statistic exceeds a predetermined F-value. The second independent variable to enter the model will be the one which has the strongest correlation with the response variable y after adjusting for the effect of the first regressor and which yields the largest partial F-statistics that exceed the F-value to enter and so on. The procedure comes to an end when the particular F-statistic at a stage or step does not exceed the predetermined F-value or when all the regressors have been exhausted and there are no more independent values to be considered for inclusion in the model.

Backward elimination begins with the full model with all the regressors and drops one at a time using the criterion of their contribution to the reduction of the sum of squared errors. The partial F-statistic is computed for each regressor on the assumption that it was the last to enter the model and is compared with a predetermined F-value, say, F-drop. The variable with the smallest partial F-statistic that is less than 'F-drop' is dropped from the model. Now the model has one less regressor than when the procedure was started. The partial F-statistics for this model can be calculated and the procedure repeated until the smallest F-value is greater than or equal to the predetermined F-drop value.

Stepwise regression is a combination of forward selection and backward elimination procedures. It is primarily a forward selection procedure with the option of dropping a variable at each step which it has in common with the backward elimination procedure. The stepwise regression procedure has a predetermined F-keep and F-drop value with the provision that different levels of significance can be assumed for entering or dropping variables. At each step the contribution of all regressors already in the model is reassessed through their partial F-statistics to see whether any are redundant or not. The variable whose partial F-statistic is less than the F-drop value can be removed from the model. Thus a variable already in the model does not necessarily stay there in the end (80-84).

Stepwise Procedure with the Option of Maximum R Improvement (MAXR) in SAS

This procedure, due to James H. Goodnight, is considered by SAS (80) to be superior to the stepwise method and almost equal to R-Square procedure which yields all possible regressions. This procedure does not settle on a single model but searches for the "best" one variable model, the "best" two variable model and so on. With the 'MAXR' procedure all switches of variables are evaluated for improvement in R^2 before any switch is made (80).

APPENDIX B

FACTOR ANALYSIS

The main aim of factor analysis is to achieve economy in the description of an entity or individual where the term individual stands for objects such as persons, bridges, or businesses when several characteristics or variables are known about the individual. Factor analysis is concerned with the resolution of the set of variables in terms of a smaller number of 'factors'. This resolution begins with the analysis of the correlation matrix of the set of variables. The resulting solution consists of factors which, although fewer in number than the number of original variables, will still contain most of the essential information. A given matrix of correlations can be factored in an infinite number of ways. One of the preferred types of factor solutions is the statistically optimal method of principal axes. This method determines factors in sequence such that the factors extracted account for the most variance of the variables at each successive stage.

Notation, Model, and Definitions of Some Terms Used in Factor Analysis (86)

Let n be the total number of entities or individuals and p be the number of variables. Let the index i be used to denote the i-th individual and J the J-th variable and x_{Ji} the value of a variable

x_J for the i-th individual. Let the standardized value of the J-th variable for the i-th individual be z_{Ji}

where

$$z_{ji} = \frac{(x_{ji}-\overline{x}_j)}{\sigma_j}$$

$$\overline{x}_j = \sum_{i=1}^{n} x_{ji}\sigma_j \quad \text{, and}$$

$$\sigma^2_j = \sum_{i=1}^{n} (x_{ji}-\overline{x}_j)^2$$

n = the total number of individuals

The set of all values of z_{ji} (i=1,2,...,n) constitute the statistical variable z_j in standard form. Factor analysis aims to express z_j linearly in terms of the hypothetical factor constants. Factors may be common factors which deal with two or more variables or may be unique factors that are present only in a particular single variable of a set. All of the variables including factors are assumed to be in standard form with a mean of 0 and a variance of 1 and the factors are assumed to be uncorrelated.

The model for the factor analysis can be written as

$$z_{ji} = a_{j1}F_{1i} + a_{j2}F_{2i} + \dots a_{jm}F_{mi} + a_jU_{ji} \quad (i,J=1,2,\dots n) \qquad \text{(B-1)}$$

where the F's are factors and U's are unique factors of the j-th variables and the a's are coefficients.

It can be shown by considering the variance of both sides of Equation (19) above that the contribution of the factor F_J to the total variance is given by

$$var(F_j) = \sum_{i=1}^{n} a^2_{ij} \tag{B-2}$$

The communality of the i-th variable is defined to be the sum of squares of the common factor coefficients and is given by

$$h^2_i = a^2_{i1}+a^2_{i2}+...+a^2_{im} \tag{B-3}$$

The unique factor U_i can be decomposed into the variance due to the specific factor (specific variance or specificity) and variance due to errors in measurement (error variance or unreliability). The complement of the error variance is known as the reliability of the variable. With the unique factor decomposed, the model of factor analysis can be written as

$$z_i = a_{i1}F_1+a_{i2}F_2+...+a_{im}F_m+b_iS_i+c_iE_i \quad (i=1,2,...,n) \tag{B-4}$$

yielding the relationship

$$1 = h_i^2 + b_i^2 + c_i^2 \tag{B-5}$$

when variances are considered on both sides of Equation (B-4), since the z's, F's, s's, are all standard normal variables.

The basic problem of factor analysis is to determine the coefficients a_{iJ} of the common factors. The coefficients a_{iJ} are usually known as the 'loadings' of the factors.

The fundamental idea of factor analysis is that the correlation matrix R can be decomposed into a product of a matrix A and its transpose and the fundamental equation is given by

$$R = AA' \tag{B-6}$$

where

R = the correlation matrix

A = the matrix of factor loadings with a typical element as a_{iJ}, and

A' = the transpose of A

Factor analysis consists of three steps which are (a) preparation of correlation matrix (b) extraction of initial factors (c) rotation of factors to reach the terminal solution. The first step involves calculating the appropriate measurement of association between the relevant variables and is relatively easy. The second step in factor analysis is to obtain the factor constants on the basis of the correlation matrix. Factors are extracted by more than one method but only the principal factor method is described here.

The principal factor method begins with extracting the first factor F_1 such that its contribution to the communalities of the variables is as great as possible. Then the residual correlations after the first factor extraction are obtained. The second factor F_2 which is orthogonal to or independent of F_1 is obtained such that it contributes most to the residual communality and so on until all of the communality is exhausted. The first factor is obtained so that the sum of the contributions of the factor to the total communality is given by

$$S_1 = a^2_{11}+a^2_{21}+\ldots+a^2_{n1} \tag{B-7}$$

is a maximum under the conditions.

$$r_{ij} = \sum_k a_{ik}a_{jk} \quad (i, j=1,2,\ldots n) \tag{B-8}$$

where the correlation $r_{ij} = r_{ji}$ and r_{ii} is the communality h_i^2 of the variable z_i.

S_1 is maximized using Lagrange multipliers. It can be shown that the maximization leads to the conditions given by

$$\begin{bmatrix} h_1^2-\lambda & r_{12} & r_{13} \cdots & r_{1n} \\ r_{21} & h_2^2-\lambda & r_{23} \cdots & r_{2n} \\ \vdots & & & \\ r_{n1} & r_{n2} & r_{n3} & h_n^2-\lambda \end{bmatrix} = 0 \tag{B-9}$$

which is the characteristic equation. The largest root λ_1 leads to solution of a_{ij} for the first factor.

The first factor residuals are given by

$$(r_{ij})_{Residual} = r_{ij} - a_{i1}\, a_{j1} \tag{B-10}$$

It can be shown that Factor analysis does not yield a unique solution to the matrix equation R = AA'. There are an infinite number of methods of decomposing R as a product of the factor matrices, a matrix and its inverse.

For example, let the A matrix or the matrix of unrotated loadings be given by the 4 x 2 matrix A

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \\ a_{31} & a_{32} \\ a_{41} & a_{42} \end{bmatrix} \qquad \text{(B-11)}$$

and a matrix (2x2) be given by

$$\Delta = \begin{bmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{bmatrix} \qquad \text{(B-12)}$$

The product $A\Delta = B$ given by

$$B = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \\ b_{41} & b_{42} \end{bmatrix} = A\Delta \qquad \text{(B-13)}$$

If R = AA', the value of BB' is

$$BB' = (A\Delta)(A\Delta)' = A\Delta\Delta'A' = AA' \qquad \text{(B-14)}$$

since $\Delta\Delta' = \begin{bmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{bmatrix} \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix}$

$$= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I$$

where I is the identity matrix of dimension 2.

The second coefficients of factor F_2 are obtained by considering the matrix of residual correlations and maximizing the quantity

$$S_2 = a_{12}^2 + a_{22}^2 + \ldots + a_{n2}^2 \quad \text{(B-15)}$$

which is the sum of the contribution of F_2 to the residual communality and so on. It can be shown that the second largest eigenvalue of the original correlation matrix R leads to the solution of the second factor and so on.

The third step involves rotation of the factor loadings matrix obtained initially into a mathematically equivalent matrix which yields factor constants that are more useful for scientific purposes than unrotated factor constants. It was seen that since R = AA' = BB' irrespective of the value of ϕ this step is not difficult.

There are several methods of rotation. The 'varimax' method of combined rotation used in this study simplifies the columns of the factor loadings of ease factor in order to make interpretation a simple matter. The factor loadings can be seen as the correlations between the variables and the hypothetical factor. The variables that have high loadings for each factor are studied carefully. If a variable does not have high loadings for the first few factors that account for most of the variability of the data, it can be eliminated if needed.

In this study, in the 'Factor Procedure' of SAS, the options of the principal axis method and rotation by the varimax method were used.

APPENDIX C

CLUSTER ANALYSIS

In cluster analysis no assumption is made about the group structure. Grouping a set of objects or entities is done on the basis of similarity or distances. For the purpose of clustering the distance 'd' between 2 m-dimensional observations $\underline{x} = (x_1, x_2, \ldots x_m)'$ and $\underline{y} = (y_1, y_2, \ldots y_m)'$ is given by the distance, d, where

$$d^2 = (x_1-y_1)^2 + (x_2-y_2)^2 + \ldots (x_m-y_m)^2 \qquad \text{(C-1)}$$

Clustering is subjective to some extent and the researcher is expected to know enough about the background of the problem to be able to differentiate between bad and good groupings. There are many ways of clustering a set of objects with different clustering algorithms. The 'Cluster' Procedure of SAS performs an agglomerative hierarchial cluster analysis and the 'Fastclus' Procedure can be used for large data sets with only two or three pairs over the data. In the Fastclus procedure the observations belong to one and only one cluster. Fastclus uses Arderberg's (88) 'nearest centroid sorting'. A very brief discussion of the agglomerative hierarchial clustering method and centroid methods is given below.

In hierarchial clustering agglomerative hierarchial methods or a divisive hierarchial method may be used. Agglomerative hierarchial procedures begin with the individual entities with as many clusters as objects. The nearest or most similar objects are first grouped and then the initial groups are again merged according to similarities or distances. Thus, eventually all subgroups are fed into one single group or cluster. Divisive hierarchial methods work in the reverse way. The total set of objects initially start as one group and then divide into two subsets such that the entities in one group are as distant from the second group as possible. These subsets are again divided into dissimilar subsets and so on until finally there are as many groups or clusters as there are entities.

There are generally four steps necessary for agglomerative hierarchial clustering. The first step starts with n clusters, one for each object and a nxn symmetric matrix of distances or similarities is calculated. Let the distance matrix be denoted by $D = \{d_{ij}\}$. Step two is a search of the distance matrix for the nearest or most similar pair of clusters. The distance between the two nearest clusters C_1 and C_2 can be denoted by $d_{C_1C_2}$. The third step is to merge clusters C_1 and C_2. Denote the newly formed cluster by (C_1, C_2). Now the distance matrix can be updated by deleting the rows and columns corresponding to cluster C_1 and C_2 and by adding a row and column giving the distances between cluster (C_1, C_2) and the remaining clusters. Steps two and three are repeated n-1 times to obtain a single cluster in the end. The identity of clusters that are merged into and the distances or similarities at which they are merged can be recorded.

From the cluster procedure of SAS, one can obtain at each step the (i) number of clusters (ii) the maximum distance between the observations in a cluster which is called the maximum diameter of a cluster (iii) the number of distances within clusters less than the maximum diameter (iv) the total number of distances less than the maximum diameter and (v) the ratio of steps (ii) and (v). A good grouping is indicated when a local peak is seen for the value of the ratio in step (v).

In the 'Fastclus' procedure, a set of point cluster seeds are selected by an initial guess as the means or centroids of the clusters. Each object in the set is assigned to the nearest seed to yield temporary clusters whose means replace the initial seeds. The process is repeated until no further changes take place in the cluster. In the cluster procedure, Euclidean distances are used as the basis to assign observations to groups.

APPENDIX D

DISCRIMINANT ANALYSIS

Discriminant analysis is concerned with analysis of groups of populations or data sets. One of the purposes of discriminant analysis is to construct classification schemes based on a set of m variables to classify a new member to one of the predetermined groups. The analysis can also be used to test for mean group differences and to describe the overlap among groups. The usual assumptions of discriminant analysis are the following (97).

(1) The groups must be discrete and identifiable,

(2) One should be able to describe each observation by a set of measurements on the m variables, and

(3) These discriminating variables must have a multivariate normal distribution. (Testing for normality is given in Appendix E)

The aim of discriminant analysis in this study was to develop a classification model to categorize an object on the basis of the profile of its characteristics to one of the groups determined by the cluster analysis. The theoretical approach to this problem of deciding the group to which an object belongs is to undertake that action which minimizes the average loss (risk) due to misclassification. The word 'loss' is used in place of 'error' and the loss function is the measure of error. The risk function is the average loss for a given loss function (103).

In the case of two populations π_1 and π_2, if $L(a_i, \pi_j)$ denotes the loss associated with taking action a_i when state of nature π_j prevails, the loss matrix can be written as

Loss Matrix for Classification

Action	State of nature z belongs to	z belongs to
a_1	0	C_{12}
a_2	C_{21}	0

where c_{ij} is the cost of misclassifying an object which actually belongs to the i-th population into the J-th population (i, j=1,2). If ρ denotes the risk ρ = Expected Loss = $\varepsilon[L(a_i, \pi_J)]$. The rule of classification is developed so as to minimize the risk (94).

The 'Discrim' Procedure of SAS (80) computes linear discriminant functions for classifying observations into two or more groups on the basis of numerical values. The classification criterion is by a measure of generalized squared distance (ρ) which can be based on either the within group variances or the pooled covariance matrix after prior probabilities of the groups are accounted for. A new observation object is placed in the class from which it has the smallest generalized squared distance. The generalized squared distance $D^2{}_i(\underline{z})$ for a new object $\underline{z}(m\times1)$ when costs of misclassification are ignored is given by

$$D_i^2(\underline{z}) = (\underline{z}-\underline{x}_i)'S_i^{-1}(\underline{z}-\underline{x}_i) + \ln|S_i| - 2\ln|P_i| \qquad \text{(D-1)}$$

where

$\underline{z}$ = the vector of variables for v, the new object to be classified,

$\underline{x}_i$ = the vector of means of variables in group i,

S_i = the covariance matrix for group i,

S_i^{-1} = the inverse of S_i

$|S_i|$ = the determinant of S_i, and

P_i = the prior probability of assignment to group i given by the ratio of the number of observations in group i to the sum of the number of observations in all the groups.

When the covariance matrices of all groups are not equal, a quadratic discriminant function is the result. If the covariance matrices are equal, a linear discriminant function is obtained.

The classification rule of assigning an observation based on the probability of its belonging to a particular group gives the same results as that based on assigning to a group based on the smallest generalized distance. The posterior probability of an observation $\underline{z}$ belonging to group i (i=1,2,... groups) is given by

$$P_i(\underline{z}) = \frac{\exp[-0.5\ D_i^2(\underline{z})]}{\sum_k \exp[-0.5\ D_i^2(z)]} \qquad \text{(D-2)}$$

If setting i = w produces the smallest value of $D^2_i(\underline{z})$ or the largest value of $P_i(\underline{z})$ then the observation is assigned to population or group w.

The discriminant function developed can be judged or evaluated as to its performance by the number of misclassifications that result when the function is used. Fewer misclassifications indicate a better discriminating function. The SAS 'Discrim' Procedure yields the number of misclassified observations for both the linear and quadratic discriminant functions. When the within-covariance matrices are equal, the linear discriminant function can be calculated from the SAS output which also prints posterior probabilities of membership in a group. For the quadratic discriminant function, SAS yields classification results, but not coefficients of the function. A program using the Proc Matrix procedure (80) can be used to obtain the quadratic function and posterior probabilities when covariance matrices are not equal. A quadratic discriminant function in this study included 98 quadratic terms, 14 linear terms and a constant. The model was not found desirable because there were more parameters estimated than the number of observations and it was difficult to interpret.

One of the assumptions of discriminant procedure is that each of the variables has a normal distribution. The variables can be tested for normality using the 'univariate' procedure of SAS discussed in the following Appendix (E).

APPENDIX E

TESTING FOR NORMALITY

The univariate procedure of SAS tests for the normality of a variable when the 'normal' option is specified. SAS tests the data against a normal distribution with mean and variance equal to the sample mean and variance. The Kolmogorov-Smirnov D-Statistic is used when the sample size n is greater than 50 and the W-statistic when n is less than 50 to test the null hypothesis that the data values are a sample from a normal distribution. The D and W statistics are briefly discussed later in this appendix. The associated probability that the data comes from a normal distribution is also printed. The output also consists of a stem and leaf plot, box plot and normal probability plot which are all also briefly described below to help decide whether the sample comes from a normal distribution.

Stem and Leaf Plot

The Stem and Leaf Plot forms a picture of the frequency distribution that is almost the same as the histogram except that it is rotated by 90 degrees. It is a pictorial representation of the distribution that is adequate for exploratory purposes (104).

Box Plot

A Box Plot is a graphical display that indicates the center or location of the median. The box consists of the middle 50% of the observations. The position of the median line indicates the skewness if any. Since the distribution mean and median are at the same position in a normal distribution, a discrepancy from normal distribution can be observed by the position of the median in the box. The outliers or extreme points are also graphed. The number of outliers also indicates non-normality (104).

Kolmogorov-Smirnov Test

The Kolmogorov-Smirnov D-Statistic is often used to test the goodness of fit of continuous data. The test involves two cumulative distribution functions - the hypothesized and the observed. Let F(x) be used to denote the probability that X is less than or equal to x, or $F(x) = P(X \leq x)$.

If $F_0(x)$ represents the cumulative hypothesized distribution and S(X) the cumulative sample distribution, the aim of the Kolmogorov-Smirnov test is to determine whether the disagreement between $F_0(x)$ and S(x) is enough to doubt the hypothesis that the sample comes from the hypothesized distribution (105). The test statistics are given by

$$D = \sup |S(x) - F_0(x)|$$

which is read as "D equals the supreme over all x_i of the absolute value of the difference $S(x) - F_0(x)$". Graphically D is the greatest vertical distance between S(x) and $F_0(x)$. The critical values for decision are obtained from tables given in standard non-parametric text books.

Normal Probability Plot

Normal probability graph paper is designed so that the plot of a cumulative normal distribution is a straight line. Substantial departures from a straight line indicate that the distribution is non-normal (83).

W-Statistic

The W-statistic is used for testing normality when the sample size is less than or equal to 50. Shapiro and Wilk (106) discuss an analysis of variance test for normality with the W-statistic which is obtained by dividing the square of an appropriate linear combination of the sample order statistics by the usual symmetric estimate of a variance. The W-statisic is defined by

$$W = \frac{b^2}{S^2} \qquad \text{(E-1)}$$

where

$y' = (y_1, \quad y_2, \quad \ldots \quad y_n)$ is a vector of ordered random observation, and

$$S^2 = \sum_{i=1}^{n} (y_i - y)^2 \qquad \text{(E-2)}$$

where $x_1 \leq x_2 < \ldots x_n$

denote an ordered random sample of size n from the normal distribution with a mean of 0 and a variance of 1 such that

$$E(x)_i = m_i \quad (i=1,2,\ldots m)$$

and b is the best linear unbiased estimate of the slope of a linear regression relating the ordered observations, y_i, to the expected values m_i of the standard normal order statistics. The constant b is given by

$$b = a_n(y_n - y_1) + \ldots + a_{k+2}(y_{k+2} - y_k) \qquad \text{(E-3)}$$

where the coefficients a, and the percentile points of distribution of W, are obtained from tables (106). Small values of W indicate non-normality.

APPENDIX F

LOGISTIC REGRESSION

Some of the terms used in logistic regression are explained here.

Likelihood Function (103)

Consider n random variables x_1, x_2, ... x_n. Their likelihood function may be defined as their joint density given by

$$f_{x_1, x_2, \dots x_n}(x_1, x_2, \dots x_n; \theta)$$

which can be taken to be the function of θ.

If the sample is taken from the density $f(x;\theta)$, then the likelihood function is given by

$$f(x_1;\theta)f(x_2;\theta) \dots f(x_n;\theta) = \pi f(x_i;\theta)$$

The likelihood function yields the 'likelihood' that the random variable θ assumes a particular value x_1, x_2, ... x_n. It is the value of a density function and a probability for discrete random variables.

Maximum Likelihood Estimators (103)

Let (x_1', x_2', ... x_n') be the observed values in a sample which supposedly has come from a population with joint density given by

$$f_{x_1,\ldots x_n}(x_1, x_2, \ldots x_n;\theta)$$

where θ is unknown.

The aim is to know θ or the density from which this particular sample is 'most likely' to have come. We wish to find the value of θ, say θ_0 which maximizes the likelihood function $L(\theta;x', x', \ldots x')$ sometimes the logarithm of the likelihood given by $\log L(\theta)$ is considered since $L(\theta)$ and $\log L(\theta)$ have their maxima for the same value of θ.

The maximum likelihood estimator is the solution of the equation given by

$$\frac{dL(\theta)}{d\theta} = 0 \qquad \text{(F-1)}$$

If there are ρ parameters to be considered, the likelihood function is given by

$$L(\theta;x_1,\ldots,x_n)L(\theta_1,\theta_2,\ldots,\theta\rho) = \prod_{i=1}^{n} f(n_i,\theta_1,\theta_2,\ldots,\theta\rho) \qquad \text{(F-2)}$$

and maximum likelihood estimators are obtained by solving the equations

$$\frac{\partial L(\theta_1,\theta_2,\ldots,\theta\rho)}{\partial\theta_i} = 0 \quad (i=1,2,\ldots,\rho) \qquad \text{(F-3)}$$

It is often better to work with the logarithm of the likelihood function.

Likelihood Ratio

Likelihood ratio is the ratio of two likelihood functions or their logarithms.

Logistic Regression

Often when the response variable is an indicator variable, the shape of the response function is not linear but curvilinear (82). A graphical representation of the response function which is found to be adequate when the dependent variable is binary is given in Figure (7). As can be seen from Figure (7) this response function is shaped like a letter S and has asymptotes at 0 and 1.

Logistic regression does not require the assumption that the variables are normally distributed as is the case with discriminant analysis. Also costs of misclassification do not enter the picture as they do in discriminant analysis. In this study, two groups of bridges were determined as belonging to 'more-safe' and 'less-safe' categories by cluster analysis. If z is an indicator variable with z=1 for bridges in the more-safe group and with z=0 for bridges in the less-safe group, it is possible to fit a logistic regression model. This model can be used to predict the probability of a bridge being 'safe' because in this case, a '1' represents a safe bridge. The probability that is calculated directly by the model can serve as an index of safety and is always between 0 and 1.

Let P_i be denoted as the probability of a bridge being safe. In order to relate the probability of safety of the dichotomous (categorical) response variable y to the m independent variables x_i (i=1,2,...m), the following model is used:

$$P_i = y/1+y, \text{ and} \tag{F-4}$$

$$1-P_i = 1/1+y \tag{F-5}$$

where

y_i = $\exp \sum_{j=1}^{m} b_j x_{ij}$

b_j = unknown coefficients that need to be determined.

The quantity λ_i is defined as

$$\lambda_i = \ln(p_i/(1-P_i) = \sum_{j=1}^{m} b_j x_{ij} \tag{F-6}$$

The variable λ_i is called the logistic transform of p_i and Equation (F-6) gives the linear logistic model. The probability of safety can be obtained from Equation (F-4) or (F-6). The b_i coefficients can be estimated by using the maximum likelihood method. The logarithm of the likelihood function can be obtained as

$$LL(b_1,b_2,\ldots,b_m) =$$

$$\sum_{j=1}^{m} b_j t_j - \sum_{i=1}^{n} \ln[1+\exp(\sum_{j=1}^{m} b_j x_{ij})] \tag{F-7}$$

where y_1, y_2, ... y_n are the dichotomous observations on the n objects (90) and $t_J = \sum_{i=1}^{n} x_{ij} \; y_i$. The maximum likelihood estimates of the b_j's that maximize the logarithm of the likelihood function are obtained by solving iteratively the following m equations simultaneously:

$$t_j - \sum_{i=1}^{n} \frac{x_{ij} \exp\left(\sum_{j=1}^{m} b_j x_{ij}\right)}{1 + \exp\left(\sum_{j=1}^{m} b_j x_{ij}\right)} = 0 \qquad (j=1,2,\ldots,m) \tag{F-8}$$

The 'logit' procedure in SAS due to Harrell (91, 92) solves the above Equation (F-8) iteratively and computes the maximum likelihood estimates for the regression coefficients. This model can also be fitted in a stepwise technique based on a strategy that determines the best variable to be added at any given stage with very few calculations. If the stepwise option is not used, the maximum likelihood estimates for the parameters associated with each independent variable are calculated and printed. The output includes of the model likelihood ratio, the model Chi-Square, and the degrees of freedom of the model. The model log likelihood ratio Chi-Square is defined to be twice the difference in the log likelihood of the current model from the likelihood based on the intercept alone. A statistic 'R' which indicates the predictive ability of the model is also obtained in the output. The 'R' statistic is similar to the multiple correlation coefficient after a correction is made to account for the number of parameters estimated. It is obtained from the equation

$$R^2 = \frac{(\text{model Chi-Square} - 2m)}{[-2L(\theta)]} \tag{F-9}$$

where m is the number of variables in the model excluding the intercept and L(θ) is the maximum log-likelihood with only intercept in the model.

"Partial R's" are also computed for each independent variable in the model and are given by

$$R = \left[\frac{\text{MLE Chi-Square} - 2}{[-2L(\theta)]}\right]^{0.5} \qquad \text{(F-10)}$$

and carry the sign of the corresponding regression coefficient (92).

The output of the logistic procedures contains the fraction of the "concordant pairs of predicted probabilities and responses." In other words, all pairs of observations having different values of y are considered and the number of pairs are counted in which the observation with the larger y has a higher predicted probability than does the observation with the smaller y. The ratio of the number of pairs in which the predicted probabilities are concordant with the dependent variable value to the total number of possible pairs gives the fraction of observations that are concordant. If this value is denoted by 'c' and the fraction of discordant pairs is denoted by 'd' (d=1-c if there are no ties); an index of rank correlation between the predicted probabilities and dependent variable is given by c-d. This rank correlation is very similar to Kendall's Rank correlation and is actually Somer's D-Statistic, both of which are discussed below. These measures indicate the goodness of the model.

Kendall's Tau Rank Correlation Coefficient and Somer's D

Kendall's Tau is a non-parametric statistics that yield a measure of correlation between two variables x and y. It is based on the ranks of observations and assumes values between -1 and +1. Two pairs of observations say (x_1, y_1) and (x_2, y_2) are said to be concordant if the difference between x_1 and x_2 is in the same direction as the difference between y_1 and y_2. There is said to be concordance if when $x_1 > x_2$, then $y_1 > y_2$ or when $x_1 < x_2$ then we have also $y_1 < y_2$. The pairs are said to be discordant if the directions of difference are not same. The procedure begins by arranging x-ranks in order from 1 to n and pair with the corresponding of y ranks. Then the strategy consists of obtaining the number of consistencies or concordance say, c, and the number of discrepancies or discordance say, d, in the paired ranks. The ratio of the number of concordant pairs to the total number of possible consistencies in the case of perfect concordance yields Kendall's rank correlation coefficient.

There is an adjustment in calculating the value of Kendall's Tau in case of ties in the rank (92). Somer's D-Statistic is similar to Kendall's Tau except slightly different in adjusting for ties (107).

APPENDIX G

LISTS OF BRIDGE DATA AND SITE LOCATIONS

Table 1G. List of Bridge Sites

DISTRICT	BRIDGE I.D.	COUNTY	BRIDGE NO.
2	C	Tarrant	220
2	D	Wise	249
2	E	Erath	73
2	F	Erath	73
2	G	Hood	112
2	H	Erath	73
9	A	Bell	14
9	F	Bell	14
9	G	Bell	14
10	D	Henderson	108
10	F	Anderson	001
10	G	Gregg	93
10	H	Van Zandt	234
10	I	Anderson	001
10	C	Henderson	108
11	C	Trinity	228
11	D	Trinity	228
11	E	Houston	114
11	F	San Augustine	203

Table 1G. List of Bridge Sites (cont'd)

DISTRICT	BRIDGE I.D.	COUNTY	BRIDGE NO.
12	A	Brazoria	20
12	G	Harris	102
12	I	Brazoria	20
12	J	Brazoria	20
12	K	Harris	102
12	L	Montgomery	170
13	A	Calhoun	29
13	B	Colorado	45
13	F	Dewitt	62
13	D	Fayette	76
13	E	Gonzales	90
13	C	Dewitt	62
14	A	Bastrop	11
14	C	Llano	150
14	E	Bastrop	11
14	F	Hays	106
14	D	Williamson	246
14	H	Williamson	246
14	I	Williamson	246
16	D	Refugio	196
16	E	Refugio	196

Table 1G. List of Bridge Sites (cont'd)

DISTRICT	BRIDGE I.D.	COUNTY	BRIDGE NO.
16	F	San Patricio	205
16	G	San Patricio	205
18	B	Denton	61
18	C	Kaufman	130
18	D	Navarro	175
20	B	Hardin	101
20	F	Jasper	122
20	H	Liberty	146
20	I	Hardin	101
3	B	Wichita	243
3	C	Wilbarger	244
3	D	Wilbarger	244
3	E	Wilbarger	244
3	F	Young	252
3	G	Young	252
3	H	Montague	169
4	A	Hartley	104
4	B	Hemphill	107
4	D	Oldham	180
4	E	Oldham	180
4	F	Randall	191

Table 1G. List of Bridge Sites (cont'd)

DISTRICT	BRIDGE I.D.	COUNTY	BRIDGE NO.
7	A	Coke	141
7	G	Tom Green	226
7	H	Tom Green	226
7	I	Tom Green	226
21	A	Kenedy	66
21	B	Duval	67
21	D	Hidalgo	109
21	G	Webb	240
21	H	Zapata	253
21	I	Zapata	253
22	A	Kinney	136
22	B	Kinney	136
22	C	Maverick	159
22	D	Uvalde	232
22	E	Val Verde	233
22	F	Val Verde	233
22	H	Zavala	254

Table 2G. F_1 - F_{12} Rating Data for all 78 Bridges

Bridge Site	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12
2C--Tarrant 220	5	5	5	5	5	1	1	4	3	3	1	2
2D--Wise 249	5	3.75	2.50	5	5	5	1	5	2	5	3	2
2E--Erath 73	5	5	1.25	5	5	2	1	4	4	2	3	3
2F--Erath 73	4.5	2.50	0.00	4	4	3	1	5	3	4	3	3
2G--Hood 112	5	5	2.50	5	5	5	1	4	2	2	3	2
2H--Erath 73	4.5	2.50	3.75	5	5	5	1	5	4	4	3	3
9A--Bell 14	5	1.25	1.25	5	5	1	2	4	3	4	3	3
9F--Bell 14	5	1.25	2.50	5	5	1	3	5	2	3	1	2
9G--Bell 14	5	3.75	0.00	3	1	5	1	2	1	4	1	2
10D--Henderson 108	5	2.50	1.25	5	5	5	2	4	2	2	1	2
10F--Anderson 001	4.25	1.25	0.00	5	5	2	5	5	4	4	1	2
10G--Gregg 93	5	1.25	1.25	5	5	5	5	4	5	4	3	2
10H--Van Zandt 234	2.5	2.50	0.00	5	3	1	5	5	4	4	1	2
10I--Anderson 001	5	3.75	2.50	5	5	5	1	4	1	1	1	2
10C--Henderson 108	5	2.50	1.25	5	5	5	2	4	2	2	1	2
11C--Trinity 228	5	2.50	1.25	5	4	5	1	4	3	3	1	1
11D--Trinity 228	5	2.50	1.25	5	5	5	1	4	3	3	2	2
11E--Houston 114	5	3.75	1.25	5	5	5	5	5	5	5	1	2
11F--San Aug 203	5	2.50	1.25	5	5	5	2	5	4	4	1	1
12A--Brazoria 20	5	3.75	1.25	5	1	5	1	4	4	4	3	2
12G--Harris 102	5	2.50	0.00	5	5	1	1	2	2	1	1	3
12I--Brazoria 20	5	3.75	1.25	5	5	1	1	4	4	3	1	3
12J--Brazoria 20	5	1.25	0.00	5	5	5	5	4	3	3	1	2
12K--Harris 102	5	1.25	0.00	5	5	5	5	4	4	4	3	2

Table 2G. (Continued)

Bridge Site	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12
12L--Montgmry 170	5	2.50	1.25	5	5	5	5	4	2	2	3	2
13A--Calhoun 29	5	2.50	1.25	5	5	5	2	5	3	5	4	1
13B--Colorado 45	5	2.50	0.00	5	1	4	1	4	2	2	2	3
13F--Dewitt 62	5	2.50	5	5	5	5	1	4	3	1	1	2
13D--Fayette 76	5	2.50	1.25	5	3	4	1	4	3	2	1	2
13E--Gonzales 90	5	2.50	2.50	5	5	5	1	4	4	1	1	2
13C--Dewitt 62	5	2.50	3.75	5	4	5	1	4	3	1	2	2
14A--Bastrop 11	5	1.25	1.25	5	2	3	1	4	5	2	1	3
14C--Llano 150	5	2.50	1.25	5	5	3	1	5	3	1	2	1
14E--Bastrop 11	4.5	1.25	1.25	5	5	3	5	5	5	5	1	2
14F--Hays 106	5	2.50	1.25	1	1	1	5	5	3	1	1	3
14D--Wmson 246	5	5	2.50	5	5	3	5	5	5	5	1	2
14H--Wmson 246	5	5	2.50	5	5	4	1	5	4	4	1	2
14I--Wmson 246	5	5	5	5	2	5	1	4	4	3	1	2
16D--Refugio 196	5	2.50	3.75	5	3	3	3	4	2	4	3	2
16E--Refugio 196	5	2.50	5	5	5	5	3	4	2	4	4	2
16F--Sanpatricio 205	3.75	2.50	3.75	5	5	5	1	5	3	4	2	3
16G--Sanpatricio 205	3.75	2.50	3.75	5	5	5	1	5	3	4	2	3
18B--Denton 61	5	2.50	3.75	5	5	5	1	4	2	2	1	2
18C--Kaufman 130	5	3.75	2.50	5	1	4	1	4	2	3	3	3
18D--Navarro 175	5	5	0.00	5	5	5	1	5	5	5	1	3
20B--Hardin 101	5	2.50	1.25	5	5	5	2	4	2	4	2	2
20F--Jasper 122	5	2.50	1.25	5	5	5	2	4	2	4	2	2
20H--Liberty 146	3.75	5	0.00	5	5	5	1	5	2	4	2	2

Table 2G. (Continued)

Bridge Site	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12
20I--Hardin 101	5	3.75	1.25	5	5	5	1	4	3	3	2	2
3B--Wichita 243	4.5	1.25	0.00	5	4	5	1	5	4	3	1	5
3C--Wilbarger 244	4.5	1.25	2.50	5	3	4	1	5	2	3	1	4
3D--Wilbarger 244	4.5	1.25	0.00	5	3	4	1	5	2	3	1	3
3E--Wilbarger 244	4.5	1.25	0.00	5	5	5	1	5	2	3	1	3
3F--Young 252	4.5	0.00	0.00	4	5	3	1	5	2	4	1	5
3G--Young 252	3.75	2.50	0.00	5	3	2	1	5	2	4	1	2
3H--Montaque 169	4.5	1.25	0.00	5	5	2	1	5	3	4	2	3
4A--Hartley 104	5	3.75	3.75	5	5	5	1	5	2	3	1	2
4B--Hemphill 107	5	5	2.50	5	1	5	1	5	3	4	1	2
4D--Oldham 180	5	2.50	2.50	5	1	2	1	5	2	4	1	2
4E--Oldham 180	5	3.75	1.25	5	4	3	1	5	2	4	1	2
4F--Randall 191	5	5	0.00	5	3	2	1	4	4	2	1	2
7A--Coke 41	5	2.50	2.50	5	5	4	2	5	3	4	3	1
7G--Tom Green 226	3.75	1.25	0.00	2	1	5	5	5	3	2	1	3
7H--Tom Green 226	5	3.75	2.50	4	3	5	1	3	4	2	1	1
7I--Tom Green 226	5	2.50	1.25	5	1	2	2	4	4	2	1	1
21A--Kenedy 66	5	2.50	1.25	5	5	4	4	4	2	4	2	3
21B--Duval 67	5	2.50	5	5	5	5	5	5	4	5	1	2
21D--Hidalgo 109	5	3.75	0.00	5	3	5	1	5	4	3	3	2
21G--Webb 240	5	2.50	1.25	4	1	2	1	4	3	2	1	2
21H--Zapata 253	5	3.75	2.50	5	5	4	1	5	3	4	4	2
21I--Zapata 253	5	3.75	2.50	5	5	4	1	4	3	2	3	2
22A--Kinney 136	5	1.25	2.50	5	5	4	3	5	4	3	3	2

Table 2G. (Continued)

Bridge Site	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12
22B--Kinney 136	5	2.50	1.25	5	5	3	3	5	4	5	4	2
22C--Maverick 159	5	2.50	0.00	5	5	5	3	5	4	3	4	2
22D--Uvalde 232	5	3.75	0.00	5	5	4	1	5	5	4	1	2
22E--Val Verde 233	5	5	1.25	5	5	2	1	5	4	1	1	1
22F--Val Verde 233	5	3.75	1.25	5	1	4	1	5	4	2	2	1
22H--Zavala 254	5	3.75	0.00	5	5	5	1	5	4	4	1	2

Table 3G. Data for New Bridge Safety Index

BRIDGE NO.	BRIDGE ID	BRIDGE WIDTH, FT	KADT, 1000 VEH/DAY	SPEED, MPH	LENGTH, FT	F9	F6	F7
1	C2	30.00	4.30	50	125	3	1	1
2	D2	28.00	3.20	60	910	2	5	1
3	E2	30.00	7.30	55	240	4	2	1
4	F2	22.00	0.90	55	255	3	3	1
5	G2	30.00	7.60	55	825	2	5	1
6	H2	22.00	0.90	55	1105	4	5	1
7	A9	26.00	3.92	55	720	3	1	2
8	F9	25.80	2.65	60	3632	2	1	3
9	G9	23.50	2.13	45	316	1	5	1
10	D10	28.00	5.14	60	2401	2	5	2
11	F10	21.75	1.38	60	819	4	2	5
12	G10	26.00	4.70	55	450	5	5	5
13	H10	19.30	0.62	55	50	4	1	5
14	I10	28.00	4.51	55	1723	1	5	1
15	C10	28.00	5.14	55	1920	2	5	2
16	D11	23.80	4.00	55	2884	3	5	1
17	E11	29.00	0.90	55	282	5	5	5
18	F11	30.00	1.70	55	640	4	5	2
19	A12	27.70	7.14	55	1040	4	5	1
20	G12	28.00	17.20	55	1261	2	1	1
21	I12	25.60	6.67	55	2416	4	1	1
22	J12	23.40	6.99	55	75	3	5	5
23	K12	30.75	14.78	60	303	4	5	5
24	L12	25.80	6.52	60	1647	2	5	5
25	A13	30.00	3.30	55	520	3	5	2
26	B13	25.50	6.70	55	1041	2	4	1
27	F13	23.70	3.70	55	1295	3	5	1
28	D13	25.60	7.80	55	1414	3	4	1
29	E13	23.80	5.60	55	1426	4	5	1
30	C13	23.70	3.70	55	782	3	5	1
31	C11	23.70	4.00	55	962	3	5	1
32	A14	26.00	7.50	40	1124	5	3	1
33	C14	24.00	7.50	40	876	3	3	1
34	E14	22.00	0.80	55	150	5	3	5
35	F14	24.00	3.20	40	150	3	1	5
36	D14	28.00	1.80	60	510	4	4	1
37	H14	28.00	1.80	60	501	4	4	1
38	I14	27.50	3.68	55	411	4	5	1
39	D16	29.80	5.50	60	247	2	3	3

Table 3G. (cont'd)

BRIDGE NO.	BRIDGE ID	BRIDGE WIDTH, FT	KADT, 1000 VEH/DAY	SPEED, MPH	LENGTH, FT	F9	F6	F7
40	E16	30.00	5.50	60	601	2	5	3
41	F16	21.80	2.30	55	1997	3	5	1
42	G16	21.90	2.20	55	798	3	5	1
43	B18	24.00	7.80	55	804	2	5	1
44	C18	26.00	3.70	55	849	2	4	1
45	D18	26.00	2.80	55	482	5	5	1
46	B20	28.40	7.47	60	303	2	5	2
47	F20	28.40	7.47	55	1604	2	5	2
48	H20	21.30	1.40	55	275	2	5	1
49	I20	28.00	7.00	60	607	3	5	1
50	B03	22.00	2.40	55	397	4	5	1
51	C03	24.00	2.20	55	1852	2	4	1
52	D03	24.00	2.20	55	362	2	5	1
53	E03	24.00	2.20	50	161	2	5	1
54	F03	22.00	1.60	55	839	2	3	1
55	G03	20.00	0.70	55	1063	2	2	1
56	H03	22.00	1.60	55	2036	3	2	1
57	A04	30.00	2.00	55	440	2	5	1
58	B04	27.00	3.60	55	2926	3	5	1
59	D04	27.80	1.50	55	250	2	2	1
60	E04	27.00	1.60	55	1627	2	3	1
61	F04	30.00	5.20	55	334	4	2	1
62	A07	30.20	1.70	55	886	3	4	2
63	G07	21.60	2.10	50	335	3	5	5
64	H07	30.00	11.00	55	1130	4	5	1
65	I07	30.00	4.30	55	1143	4	2	2
66	A21	39.90	3.80	55	100	2	4	4
67	B21	44.00	1.50	55	205	4	5	5
68	D21	30.00	1.50	55	3181	4	5	1
69	G21	25.50	3.90	55	189	3	2	1
70	H21	27.60	2.00	55	1190	3	4	1
71	I21	27.60	3.20	55	1850	3	4	1
72	A22	34.90	2.20	55	1431	4	4	3
73	B22	38.00	2.20	55	66	4	3	3
74	C22	31.00	1.60	55	120	4	5	3
75	D22	27.30	1.80	55	341	5	4	1
76	E22	27.70	1.00	55	1310	4	2	1
77	F22	30.00	1.50	55	5462	4	4	1
78	H22	25.60	2.00	55	1505	4	5	1

Table 4G. Range of Variables in New Bridge Safety Index

Range of Variables	Bridge Width, Ft	KADT 1000 Veh/Day	Speed, MPH	Length, Ft	F_9	F_6	F_7
Maximum	44.0	17.20	60	5462	5	5	5
Minimum	19.3	0.70	40	50	1	1	1