APPLICATION OF REGRESSION TECHNIQUES TO HIGHWAY SYSTEMS SAFETY

by

P. D. Kiser

Thesis submitted to the Graduate Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of
MASTER OF SCIENCE
in
Civil Engineering

APPROVED:

Robert J. Popper, Chairman

Richard D. Walker

Valiant Mah

May 1975

Blacksburg, Virginia

ACKNOWLEDGEMENTS

The author would like to thank those who helped to make this study a reality; these include Dr. Robert Jay Popper for his constant guidance and many helpful suggestions, Dr. Valiant Mah for his statistical counseling, the Civil Engineering Department for the much needed Graduate Teaching Assistantship, and the author's fellow students, many as they are, especially those in 105 Norris.

A special offer of appreciation goes to Dr. Olin K. Dart of Louisiana State University for his help in providing the data used in this study.

The author would also like to pay a special heartfelt tribute to his wife, Penny, for her typing and editing of this study.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES (cont.)

CHAPTER ONE

INTRODUCTION

Since the passage of the Highway Safety Acts of 1966, billions of dollars have been spent in improving the highway system in the United States. The main goal was to reduce the constant trend in the death toll of 50,000 lives and more than 2.0 million injured in highway accidents every year (1). These figures have undergone little or no significant decrease for nearly ten years, until recently, when the speed limits were reduced to 55 mph. Along with the fatalities and injuries are accident costs, nearly 10 billion dollars a year, that are largely paid by the private citizen.

The primary concern here is the lack of any definite planning structure to deal with the highway safety problems. If a highway safety program is to be successful, goals must be stated and pursued. Baker states that "operating without goals in highway safety means that there is no limit to the resources which can go into the program, since there is no limit to the level of safety that could be required." (2) Goals should be in terms of lives saved and economic losses reduced. Setting goals, resources and mobility requirements will permit the best allocation among the several programs of government related to preservation of life and property.

The problems of allocating resources to highway safety projects stem from the fact that much of the money is spent on alleviating the "cause" of the accidents without really knowing what is the cause. Baker comments on this by saying that:

> "All accidents for practical purposes, have more than one contributing factor. Attempts to determine "the" cause, or programs directed to eliminating "unsafe" conditions are based upon a single cause-and-effect assumption; the best way to reduce the occurrence of highway accidents is to remove the cause. In reality an accident is a "breakdown" or failure of the highway transport system; the system fails to get people and goods from one point to another with safety."(2)

From this standpoint highway safety is a systems problem and should be tackled as such. When a systems approach is applied to this problem the complexities involved become quite apparent and reveal the problem as being more than a simple task of identifying a "cause".

As it is financially impossible to deal with all the weaknesses of the highway system, the key to the most rapid improvement is the identification and correction of factors having the greatest effect. The need for a systems approach to highway safety is evident simply due to the fact that the highway system is a very large and complex one, requiring an approach or methodology that will identify those factors contributing to traffic accidents. Such an identification process should be a rigorous application of procedures that will enable the analyst to deal with the large amount of statistics that would stem from a systems analysis of highway safety.

One procedure commonly used today for accident studies is multiple regression analysis. This statistical procedure would seem to be a potentially good approach for analyzing traffic accidents. However, regression analysis in past studies has often been misused and improperly evaluated, apparently due to a misunderstanding of the limits and capabilities of the procedure. With the use of high speed computers regression analysis has become a valuable tool in the analysis of increasingly larger variable and data sets. Among these large variable sets are those defining various elements of the highway system. These highway system elements, namely the human, the vehicle, and the roadway/environment, are described by a large number of variables, many of which can be statistically analyzed using regression techniques. Therefore, to use regression techniques successfully a great deal of thought and effort must be directed toward the variable selection and data collection process. Since it is impractical, if not impossible to incorporate every variable of the highway system into a regression analysis, it is necessary that the analyst choose specific areas of the highway system that will benefit most from the use of regression techniques. As with any highway safety study, the variable selection is a very crucial step in the analysis.

The following are objectives for this study:

1. Review of past and present works that involve a systems approach to highway safety and those studies incorporating regression techniques in accident analysis,
2. Presentation of the accident estimation problem from a total systems viewpoint through identification of each main component and the information available to describe each component,
3. Development of a rigorous regression analysis procedure, incorporating variables from the various components identified in the systems approach,
4. Demonstration of the capabilities of a rigorous regression procedure that analyzes the roadway/environment component of the

highway system; identifying the most significant variables that contribute to accidents.

CHAPTER TWO

LITERATURE REVIEW

Prior to the undertaking of a systems approach of highway safety incorporating regression techniques, it is desirable to review those works and studies that have attempted similar approaches and/or techniques.

The purpose here is to discuss the extent of previous research in dealing with highway safety as a systems problem, i.e., to point out the discrepancies and inadequacies of those studies attempting to explain highway accidents from the use of regression techniques. Therefore, the basis of this study is to expand and improve accident estimation where previous studies have been unsuccessful in doing so.

The accident studies reviewed in this study were somewhat limited in their scope of a systems approach to accident analysis. It was therefore necessary to sufficiently identify and define the highway system before regression techniques could be applied.

The main elements of the highway system used for analysis in this study are as follows:

1. Human
2. Vehicle
3. Roadway/Environment

The highway system basically involves the transporting of people and goods from one point to another over a roadway network. It is the interaction and variation of the above elements composing the system that lead to highway safety problems. Each of the three elements is composed of a number of variables, many of which have a direct relationship to the occurrence of traffic accidents. Some of the relavant factors within the three elements include the driver's condition, behavior and training for the human element; vehicle type, condition and crashworthiness for the vehicle element and roadway geometry, roadside conditions and traffic conditions for the roadway/environment element.

A review of past and present literature reveals that most of the previous analyses have centered around the roadway/environment element and often just around a few variables of this specific element. The main reason for this is simply a lack of data for the other elements. Perhaps another reason this occurs is due to the problem of deciding on pertinent information to be collected for drivers or vehicles.

In searching for studies or works dealing with a systems approach to highway safety it was found that only one work existed which came close to doing this. This was the publication by Arthur D. Little, Inc. (10) concerning traffic safety. It is perhaps the most comprehensive work available on the highway system as a whole. The information is actually a literature review of the following factors,

1. human factors,
2. environmental factors,
3. vehicular factors (before and during impact),
4. loss-limiting factors (post-accident), and
5. regulatory and legal factors.

An important objective of this work was to point out that highway safety involves many interrelated factors, thereby cautioning those who would look only at one or two of the five factors listed previously in studying traffic accidents. The description and definition of the many factors composing the highway system were taken from the Little (10) work and used as the basic structure for the highway system discussion in Chapter Three.

In recent years traffic accidents have been studied and analyzed by a variety of methods. Some of these methods included simple ranking procedures to more complex statistical analyses.

Presently much attention is given to the statistical approaches; due in part to a better understanding of these methods plus the fact that they provide more reliable results for decision makers in the allocation of funds for highway safety improvement projects.

As for the use of regression analysis in accident studies, it has been used for identifying significant variables as well as for the predictive power it posses. However, the misuse of regression techniques may have produced undesirable and often useless results. (9), (12), (13).

There are several reasons for the continued use of regression analysis in accident studies. Besides the fact that regression computer packages are readily available, groups of variables of various sizes can be analyzed with valuable information provided for each variable analyzed. For instance regression results can provide the analyst with those variables having the most significant relationship with accidents and whether or not they contribute in reducing or increasing accident rates. Regression techniques are also capable of producing sufficient results for small sets of data as well as the large data sets. Perhaps the most distinct advantage of regression analysis is the ability to predict certain values that may exist when presented with specific conditions.

The following are summaries of several studies that used regression analysis in accident studies. Comments are provided on the uses of the regression technique in each study.

One study (11) which evaluated signalized and non-signalized intersections inocrporated the "traffic conflicts technique" developed by General Motors Research Laboratories.

> "The basic premise of the conflicts technique is that the number of evasive maneuvers and brake-light indications can be used both to estimate the number of accidents that will occur over a given period of time and to evaluate the operational problems of the subject intersection." (11)

Multiple regression was used with variables falling into the following six classes:

1. Average daily traffic,
2. Conflict type,
3. Environment (rural, retail strip, residential),
4. Intersection type,
5. Major to minor route volume ratio, and
6. Intersection control (signalized vs non-signalized).

Although general equations were given for signalized and unsignalized intersections there was little to indicate the values of these equations since no *R^2 values were provided.

Snyder (13) used a different approach to accident analysis by employing "environmental" characteristics into the Regression model. These characteristics include the percentage of developed frontage, percentage of commercial frontage, percentage of residential frontage, number of land use category changes, percentage of regional land developed, vehicle density, employment density, population density, residential density, value of homes, rent and percentage of population between 16 and 24. Also included as variables were road type and traffic volumes.

Although the results found by the author were good as far as the R^2 values were concerned there is some doubt as to the actual relationship between accidents and the environmental variables. The use of regression equations containing such variables is questionable when considering possible methods of reducing the accident rate.

The environmental variables the author has chosen for his regression model are perhaps better suited for predicting highway user trips rather than highway accidents. Also, it is unclear as to what type of regression model the author used; whether it was linear or non-linear in form.

A study by Schultz, Berg, and Oppenlander (12) evaluated various protection devices at rail-highway grade crossings. The protective devices included painted crossbucks, reflecterized crossbucks, flashers and gates. The authors incorporated a linear regression model into

*R^2 = multiple correlation coefficient (proportion of total variation about the mean Y explained by the regression)

the study using the protective devices as "dummy" variables in the equation. The final regression equations also included the number of tracks, lane width, trains per day and average daily traffic. The R^2 value for the equation, including the above variables (dummy variables, also), was 0.193, which gives the indication that the equation was useless for predictive or explanatory purposes. A nomograph was prepared from the study results to evaluate rail-highway grade crossings for determining the best protective device that prevented accidents. It should be pointed out the authors selected crossings with and without accident experience.

Much of the highway accident research being conducted at present deals with relationships between accidents and highway geometry or design elements.

It has been envisioned that results from these studies could provide information to government agencies in allocating funds for improvement of the critical highway geometry sections that show high accident potential.

A study by Dart and Mann, (4), attempted to identify roadway geometry variables contributing most to accidents and to predict the accident potential of a certain section of Louisiana highways. The variables studied were percentage of trucks, traffic volume ratio, lane width, shoulder width, pavement cross slope, horizontal alignment, vertical alignment, percentage of continuous obstructions, marginal obstructions per mile, and traffic access points per mile. The above variables, their squares and first order interactions were used in regression analysis (linear model) to produce equations for accidents occurring on wet and dry roads, accidents occurring at night and during daytime and for fatalities and injuries.

The following variables were found to be most significant:

1. Traffic volume and pavement cross slope,
2. Traffic conflicts and traffic volume,
3. Lane width and traffic conflicts,
4. Traffic volume and horizontal alignment,
5. Shoulder width and horizontal alignment, and
6. Traffic volume and trucks.

The R^2 value for the total accidents equation was 0.46 with the F-ratios being significant at the 0.05 level.

As a whole the study is very informative with a solid and large enough data base for regression. There is some doubt as to the accuracy of the curves given in the study where several of the independent variables are plotted against accidents (dependent variable). There appears to be too few points to justify the curves given.

Kihlberg and Tharp (9) performed a study similar to the one discussed above. However, due to a lack of sufficient data on each geometric variable the authors used a different regression approach.

Traffic volume data from road segments (all were 0.3 mile segments) were grouped by gradient, curvature, intersections, and structures (bridges) and combinations of these. Each of these groups were labeled according to number of lanes, median, and access control. The dependent variable, accidents, was classified by accident type and severity.

The following regression model was used for analysis:

$$\log A = a + b_1 \log T + b_2 \log^2 T$$

A = accidents

a = constant

b_1, b_2 = regression coefficients

T = average daily traffic

The authors justified use of this model by stating that traffic volume and accidents are more or less directly related which implies as traffic volume increases, accidents increase. Although the data base for this study was fairly large (data collected for 3 states), it failed to provide information on specific elements of the roadway system. Therefore, the results from this study were too general to be of use to a highway improvement program.

The review of the preceding studies indicate the differences and misuses that exist in the use of regression analysis. Some of the problems lie in the type of variables used in estimating accidents. Studies (13) have been conducted using variables that had little or no apparent relationship with accidents or two few variables, but gave accepted results (high R^2 values). This leads to the problem of basing study results solely on the calculated R^2 values (coefficients of multiple correlation) without regard to the types of variables and their regression coefficients. Studies (12) have also been published that gave R^2 values so low that the regression equation was useless for practical purposes. Another misuse or perhaps underuse of regression analysis is the use of very simple first degree equations, (i.e.,

$Y = b_0 + b_1 x_1 + b_2 x_2$) without considering any transformations of the variables to improve the estimating quality of the equation. The simple equation above would be sufficient for those (x) variables having a definite linear relationship with the Y variables, however, this does not always exist for real world situations in the highway system.

The regression methodology developed in this study deals with the above mentioned problems in a step by step procedure that provides the analyst with graphical analysis of the data before and during the statistical analysis. Independent and dependent variable relationships can be examined and improved, the variables in the regression equation can be examined and needed changes made at various stages of the procedure, and a statistical method is given to select transformations of the variables that will improve the regression equation. Also covered is a discussion on the examination of the final regression equation. A demonstration of this methodology is given in Chapter Five.

The summaries on the above studies indicate there is no clearcut method, even for regression analysis, in estimating accidents. There has been, in the past, misuse of regression stemming mostly from lack of understanding of the strengths and weaknesses of regression analysis. Although it is a valuable tool for accident estimation there is no reason to believe regression analysis is the only approach available for the analyst, however its advantages of multivariate analysis and its predictive potential make it a valuable tool in highway safety improvement programs.

CHAPTER THREE

HIGHWAY SYSTEMS SAFETY ANALYSIS

The highway transportation system as it exists today is one of many complexities and problems. The role of safety in this system requires an understanding of a wide variety of social, economic, political, psychological, legal and physiological as well as engineering factors that are related to the driver, the vehicle and the roadway. The system is not only characterized by its complex nature, but also by the high degree of interrelationships among the previously mentioned factors.

A systems approach to highway safety requires the identification of the main elements as well as the identification of the variables associated with each element. A necessary step in any analysis, the identification and definition of significant variables will often times rely a great deal on the personal knowledge and experience of the analyst. Of course, a realistic approach to systems analysis would use input from several disciplines associated with the human-vehicle-roadway/environment elements.

A. Systems Definition

The highway system can be characterized by three main elements; these are the human, the vehicle, and the roadway/environment elements. The questions of safety concerns all three, for it is from the interaction and combination of the three that accidents will occur.

There are those who will argue that highway fatalities are rare events and that the death rate (deaths per 100,000,000 vehicle miles driven) is decreasing (11 deaths per 100,000,000 vehicle miles in 1945 to 5 deaths per 100,000,000 vehicles miles in 1971). (1) Although this may be true, the fact remains that inadequacies still exist throughout the highway system that need identification and improvement.

The following, based on the work by A. D. Little, Inc. (10), is the basic structure of the highway system. Included are factors that may not be directly linked to accidents as contributors, but may be descriptors of accident experience severity, or type. Since vehicles come under strict operation and condition codes from state and federal government the variables for the vehicle element may be used more for describing accident severity than for contributing to the occurrence of accidents.

1. <u>Human Element</u> - the human element is usually thought of only as the driver but may include pedestrians as well. Pedestrian highway deaths account for about 20% of the total highway deaths each year. (1) However, it is the driver as a contributor of accidents with which most studies are concerned.

In response to the statement that driver error is responsible for from 80 to 90 percent of all accidents;

> "Driving involves the performance of a complex perceptual-motor skill with the driver responding to and interacting with a large set of stimuli. From this point of view, the driver's behavior and his likelihood of error can be modified by changing the stimuli, whether these be conditions which give rise to driver fatigue and a consequent deterioration in performance; visual arrangements which produce error in judging distance from the car ahead; highway signs which lead a driver to slow down dangerously in a high speed lane; medical conditions which impair performance, or others. In terms of this analysis, the driver, while controlling the vehicle, may be no more a contributor to accidents than the

vehicle, road and legal and regulatory system." (10)

This statement points out the fact that although drivers do have control over their speeds and directions they are never able to perform under ideal conditions at all times. Therefore their reactions to situations will change with changes in the roadway, traffic conditions, weather conditions and so forth. With this in mind, one can see that drivers are capable of maintaining safe maneuvers only so long as the other elements in the system provide no abrupt condition changes.

a. Biographical Variables - these are variables referred to above as being descriptors rather than contributors. Examples of these would be age and sex per exposure rate, marital status and traffic violation record. The exposure rate is the estimated amount of travel and is usually expressed in 100,000,000 miles of travel. This is also used when expressing accident rates. (Accidents, fatalities or injuries per 100,000,000 vehicle-miles).

b. Psychological Variables - little is known about the relationship between the psychological conditions of certain types of drivers and accidents. Studies have been conducted which show that some drivers may possess "accident proneness" traits. (7), (10). Collection of data for such variables would be difficult if not impossible. It is hoped continuing research will reveal more about the personality of highway users.

c. Physiological Variables - this category could include most all variables of a medical nature. These include diseases, physical impairment, drug usage (alcohol included) and mental illness. It is theorized that accidents due to suicides are caused from a complication of the above mentioned variables especially, drug usage and mental illness.

d. Driving Skill Variables - although much work has been done in determining driving skill using simulators as well as actual driving observation, there is no scientific evidence connecting driving ability with accident occurrence. It has been suggested, (10) that refinement of techniques be made to improve the quality of results dealing with driving skill.

3. Driver Education Variables - Driver training is conceived as an aid in contributing to highway safety. The main concern is the ability to properly train the thousands of young people who reach the driving age each year. Variables from this category might include type and length of training and driving experience.

2. Vehicle Element - The vehicle is that element of the highway system that transports persons through the system with a relatively high degree of freedom as to time and direction of travel. The identification of vehicular factors can be separated into two categories; one concerning the proper functioning of the vehicle for safe travel and the other that deals with occupant protection in the event of an accident.

a. Visual Variables - these variables deal with the driver's ability to observe at all times those operations ahead and behind him during his own functioning of the vehicle. This would be those parts of the vehicle such as windshields, side windows, windshield pillars and headlights. Driver size and age may be an important factor to consider with these variables.

b. Vehicle Design Variables - these pertain to the interior design features of vehicles, such as foot and hand controls, sound and weather proofing, and others that are necessary for comfortable operation of the vehicle.

c. Dynamic Control Variables - the dynamic aspect of these variables include the steering capabilities of the vehicle, effects of vehicle loading and especially the power-to-weight ratio of vehicles.

d. Vehicle Condition Variables - these variables include the same features of the vehicle that must be checked for state vehicle inspections. This usually involves any part of the vehicle necessary for safe operation at all times. Especially important for consideration are tire type and condition and brake type and condition.

e. Impact Variables - these variables relate to events where vehicles are involved in an accident and subjected to large, sudden forces causing distortion of the vehicle structure. Door lock failure, bumper strength and others fit into this group. Although this type of variable is not considered in the iniation of an accident they become very important during accident occurrence.

f. Vehicle Occupant Protection Variables - these variables also deal with the occurrence of the accident and not as contributors of accidents. They could be thought of more as contributors to the severity of injury. The main concern is the protection of the passenger after the initial impact and when the occupant is still in a state of motion inside the vehicle. Variables in this group include types of steering column assembly, shock absorption qualities of interior fixtures and devices and restraint systems, such as seat belts and harnesses.

3. Roadway/Environment Element - The roadway/environment element of the highway system provides a passageway used by the combination of driver and vehicle, as well as information for traveling the passageway. Due to the numerous studies and information collected for this element (among other reasons), more funds are allocated for its improvement of safety than the combination of the first two elements. The amount of control exercised over most of the variables in this category make it a logical choice for attention from engineers and decision makers.

a. Roadway Variables - these variables identify the physical part of the roadway and shoulder. Included in this set of variables would be dimensions of road and shoulder, number of lanes, topography, roadway surface type and condition, roadway defects, shoulder type and etc.

b. Roadside Variables - variables in this group would be those obstructions or change of slope that exist from the shoulder's edge away from the road surface. Obstructions would include guard rails, trees, utility poles, curbs, median barriers, etc. Change of slope indicates any abrupt change of slope from the shoulder's edge that is not protected by guard rail or similar devices.

c. Traffic Conflicts Variables - this group of variables would cover those areas where interference would occur in the traffic flow. This could be intersections, driveways, bridges, constrictions, tunnels and etc.

d. Communicational Variables - included in this group are those variables dealing with the motorist's ability to discern the roadway ahead as well as identify various directional and control information provided by signs, signals, markings, and roadway delineation. Some of the variables would include levels of illumination, glare from headlights or sunlight, types of signing, signaling, markings, delineation and placement, etc.

e. Traffic Control Variables - these variables deal with those devices used for controlling and regulating traffic speed and flow. Such devices would be signs (one way, speed limit, stop signs, etc.), traffic signals and others.

f. Environmental Variables - this group of variables covers meteorological conditions affecting the roadway and driver, and the levels of natural light present during the driving task. Important information would concern the atmospheric conditions such as degrees of rainfall, snowfall, dust, etc. and the amount of available light such as that present at dawn and dusk.

B. A Systems Viewpoint of Highway Accidents

Haddon (15), in discussing highway safety, states that any analysis of highway accidents should be divided into three consecutive phases. These three phases represent the before, during and after periods of a highway accident. The following is a description of the three phases (15):

> "The First, or 'Initiation' Phase is the period before a crash or near-crash. In it operate the many driver, pedestrian, vehicle, highway, and other factors that lead to a

crash or to its avoidance.

The Second or 'Crash' Phase is the brief interval of the crash itself. In it the forces of impact - depending on their characteristics and on the degree to which they have been anticipated by appropriate vehicle and crash design - do their work, harmlessly or with damage to people and to the vehicle and other structure involved.

The Third, or 'Clean-up' Phase is the after-the-crash period in which the present slowness and inadequacy of emergency medical care and transportation and other factors commonly lead to completely unnecessary death, disability, and prolonged medical care."

The three phases, Pre-Crash, Crash, and Post-Crash are shown in Figure 1 (16) in a matrix form. Each cell in the matrix represents a certain group of variables that relate to that phase of accident occurrence for a specific element.

The "Pre-Crash" phase is the time period before the accident or near-accident. In this phase are the many variables from the main elements of the system that contribute in initiating those actions leading to the accident. In the "Crash" phase the vehicle and roadway/environment play a large part in protecting or injurying the human element involved as well as influencing property damage. The human element, whether driver or pedestrian, in the crash phase has little or no control over bodily or vehicular actions. The "Post-Crash" phase deals with limiting further injury and damage, and the immediate care of those involved in accidents. Little work has been accomplished in the post-crash phase area leaving much of the responsibility to local governments, medical facilities and civic groups.

Since the highway system is such a complicated and immense system the only approach for the analysis, for safety purposes, would be the

	Pre-Crash	Crash	Post-Crash
Human	Cell 1	Cell 2	Cell 3
Vehicle	Cell 4	Cell 5	Cell 6
Roadway/ Environment	Cell 7	Cell 8	Cell 9

Figure 1. Matrix for Highway Accident Analysis

dissection into those subsystems that identify the main system elements. For statistical analysis, such as the regression techniques used in this study, this dissection of the system can be somewhat of an advantage. Taking variables from one or several of the cells, regression analysis can be used to identify the significant variables that relate to traffic accidents. Knowing what the most important variables are in each cell, decision makers will have a guideline for setting safety goals and allocating funds for safety programs.

C. Problem Definition and Variable Selection Phase

The specific statement of the problem is the most important phase of any problem-solving procedure. It is necessary that the analyst be exact in specification of his problem. The problem definition must be to the point and, for regression analysis, both the response (dependent) and predictor (independent) variables must be clearly identified.

To avoid a problem statement that is too general in nature, the analyst must have some knowledge of the variables he will be working with. Instead of a problem statement where the analyst asks, "What causes accidents?", he may ask "What are the effects of roadway geometry on accident rates for two-lane rural roads?". The previously defined highway accident matrix allows the analyst to view a finite aspect of the system, thus allowing him to make specific, and not general, problem statements.

Since regression techniques are applied in this study, the variable selection consists of deciding what cell or cells from the traffic accident matrix will be used and what will be used for the dependent (response) and independent variables.

There should be no restraint in the initial selection of variables before the data collection takes place. This will provide a large group from which the analyst can eventually narrow down and use in the regression procedure. Draper and Smith (6) state, "The important point to remember is that the screening of variables should never be left to the sole discretion of any statistical procedure, including the multiple regression procedures."

The following is a descriptive outline of variable groups that may influence highway safety. It is developed to aid in the variable selection phase. Some of the groups contain variables that are categorical (non-numerical) in form, but nonetheless should be considered for analysis purposes. The variables represented in these groups would be independent variables in a regression equation. Following this outline is a listing for dependent (response) variables.

The variables are grouped according to highway element (human-vehicle-roadway/environment) and phase (pre-crash, crash, and post-crash). The outline starts at cell 1 for the human element and continues through cell 9 for the roadway/environment element.

Shown in the appendix, Figures (38), (39), (40) and (41), is an accident report form for the state of Virginia (17). This form gives some indication of what information is currently collected for highway accidents.

Cell 1 - Human: Pre-Crash Phase

1. Physical or Physiological Failures

a. Heart Attack

b. Stroke

c. Falling Asleep

2. Mental Performances

a. Physical

(i) Inattention

(ii) Distraction

b. Psychological

(i) Improper maneuvers

(ii) Excessive speed

(iii) Improper evasive action

(iv) Misjudgement

c. Information Processing

(i) Recognition

(ii) Decision

(iii) Action

(iv) Reaction

3. Conditions

a. Physiological

(i) Drinking

(ii) Other intoxication

(iii) Fatigue

(iv) Handicapped

b. Psychological

(i) Emotional stress

(ii) Alcohol or drug problem

(iii) Mental disorders

4. Training/Experience

a. User inexperience

b. Vehicle unfamiliarity

c. Area/road unfamiliarity

d. Training in educational system

Cell 2 - Human: Crash Phase

1. Human Tolerance

a. Age effects

(i) Brittle bones, etc.

(ii) Heart condition

2. Human Kinematics

a. Seat belt submarining

b. Occupant impact behavior

(i) Bracing

(ii) Sleeping

(iii) Panicked state

3. Occupant Use of Restraint Systems

a. Seat belt

b. Seat and shoulder belt

c. Doors locked

d. Helmets on motorcycle users

Cell 3 - Human: Post-Crash Phase

1. Occupant Extrication

a. Handling of victims during extrication

b. Panic behaviors during extrication

2. Emergency Services

a. Ambulance service arrival

b. First aid

(i) At scene

(ii) En-route

(iii) At hospital

3. Treatments

a. In hospital care

b. Autopsy

Cell 4 - Vehicle: Pre-Crash Phase

1. Vehicle Alterations

a. Non standardization

b. Control panel layout

c. Special modifications

d. Driver incompatibility

e. Special controls

2. Vehicle Condition and Maintenance

a. Negligent or improper inspection

b. Vehicle handling and stability problems

c. Improper maintenance

3. Vehicle Component Failure, Degradation or Maladjustment

a. Brake failure

b. Tire failure

c. Steering linkage, etc.

Cell 5 - Vehicle: Crash Phase

1. Design

   a. Standardization

      (i) Interior

      (ii) Exterior

   b. Impact Tests

2. Damage

   a. Doors opening

   b. Gas tank rupture

   c. Hood penetrating windshield

3. Crash-worthiness

   a. Energy absorbing interior parts

   b. Shatter proof windshield

4. Injury Reduction

   a. Seat restraints

   b. Head rest

   c. Air bags

Cell 6 - Vehicle: Post-Crash Phase

1. Post Crash Hazards

   a. Fuel spillage

   b. Fire-ignition

   c. Non-flammable interiors

2. Egress Problems

   a. Doors jammed

b. Trapped passengers

3. Emergency Equipment

Cell 7 - Roadway/Environment: Pre-Crash Phase

1. Right-of-way (Roadway and Roadside)

a. Regional topography

(i) Type of terrain, hill, trees, buildings

b. Features and Appurtances

(i) Adequacy of design standards

(ii) Signs - break-away, fixed

(iii) Obstacles - trees, unprotected guardrail

(iv) Shoulder maintenance

c. Geometrics

(i) Vertical and horizontal curves

(ii) Road width; lane width

(iii) Pavement cross slope

(iv) Sight distance

d. Traffic Controls

(i) Signs - clear, understandable, consistent

(ii) Signals - timing, etc.

(iii) Marking, delineation

e. Accident History

f. Maintenance

g. Artificial illumination

2. Environmental

a. Regional Variations

b. (i) Urban, suburban, rural

(ii) Population

(iii) Law enforcement

b. Weather

(i) Rain, snow, sleet, dust

(ii) Pavement surface condition (skid resistance)

c. Visibility

(i) Dark, dawn, dusk, bright, etc.

d. Traffic Characteristics

(i) Volume

(ii) Speed - mean, distribution

(iii) Composition

Cell 8 - Roadway/Environment: Crash Phase

1. Restraining or Guiding Devices

a. Absence of guardrail or median barrier

b. Break away sign supports

c. Cushioning vegatation

d. Energy absorbing barriers

2. Injury Producing Agents

a. Exposed guardrail ends

b. Concrete sign supports

c. Tree too close to roadway

Cell 9 - Roadway/Environment: Post-Crash Phase

1. Post Accident Traffic Control

a. Other collisions

b. Inadequate warning signals

2. Debris, Hazard Control, Clean-up

3. Environmental Repair

4. Investigate; Follow-up

Dependent (response) variables used in regression techniques will deal with accidents in one form or another. Actually, better and more often used, are rates, whether they be accident or severity rates. These might be expressed as accidents per length of roadway.

1. Accidents can be classified as total accidents, accidents occurring in rural or urban areas, during inclement types of weather and etc. Accidents may also be classified according to severity such as fatalities, degree of injury and etc.

2. Rates are in vehicle miles driven and are usually expressed as 100,000,000 vehicle miles, and

3. Length of roadway is the desired length of roadway section for analysis purposes. Some studies have used the same length throughout, such as increments of 0.10 miles and some studies have used variable section lengths dependent on roadway geometry, number of accidents and etc.

As it would be impractical to list every variable associated with each element of the highway system, the preceding outline provides a solid base for selecting variables from any part of the highway system being examined by a highway or traffic engineer.

The variables used in this study relate to Cell 7 of the roadway/ environment element and are considered as potential contributors in the pre-crash phase of accident occurrence. The response variable used, total accident rate, is expressed as a function of the various roadway, roadside and traffic variables chosen for the study. This is shown in equation 1 below.

$$Y = f(x_a, x_b, x_c) \qquad (1)$$

where;

Y = Computed accident rate (accident/100 million vehicle miles)

x_a = Roadway geometry variables

x_b = Roadside friction variables

x_c = Traffic variables

The various roadway/environment variables mentioned above have been used in previous studies (9), (11) due to the fact that data for these can usually be collected without much difficulty. Many state highway agencies keep records and computer data files on these variables as well as other variables for the driver and vehicle.

Variables from the human element, which are largely driver variables, have not been incorporated in any extent into regression analyses due to data requirements. There is certainly data available that could possibly be incorporated into regression analysis in some form or another. Figure (41) in the appendix gives an example of what driver information is collected by the state of Virginia for highway accidents. Although useful information may exist for certain human and vehicle variables that are not currently or previously used in accident analysis, there are variables from these two elements that may prove to be very important but data cannot be collected for them. These may include the condition of driver and vehicle immediately prior to the accident, driver psychological and personality factors and etc.

The incorporation of large groups of variables from the various elements into regression analysis may be hampered by the nonsuitability of the variables as well as the unavailability in data collection. As a tool in aiding the analysis of variables, such as the screening of variables, the regression methodology in Chapter Four was developed.

Chapter Four

REGRESSION METHODOLOGY FOR ACCIDENT ESTIMATION

When considering the safety problems of the highway system from a systems analysis viewpoint, the need arises for a procedure or methodology that will effectively handle the large amounts of data that result from such a large system. The matrix for highway accidents from the previous chapter can provide smaller scale data sets from the overall system and can be analyzed by regression analysis for accident estimation and identification of important variables as accident contributors. The regression methodology developed in this chapter can be applied as a statistical tool in the systems analysis approach to highway safety. The methodology provides a rigorous and helpful approach in the useage of regression techniques.

Before the final selection of regression analysis as the statistical tool for this methodology several techniques were evaluated for possible use. The first approach was using only the correlation matrix for the variables. The correlation matrix only gave the interaction existing between any two variables and not between the variable set as a whole, as was desired. Factor analysis was also reviewed as a possible analysis. This approach was first used for the explanation of psychological theories. Its main purpose was to obtain quantitative measures of the contribution of each variable in the data set. Since the variables used in this study were readily quantifiable as well as many other variables from the highway system, this approach was not considered for further use.

Regression analysis, incorporating the use of "dummy" variables, was chosen because of its strength and flexibility in identifying the significant variables as contributors to highway accidents. The use of "dummy" variables in regression analysis allows for the analysis of non-numerical or categorical data as well as the quantifiable data. A discussion of the use of dummy variables is given in the part B of this chapter. Since the purpose was one of finding the relationship of highway accidents with a specific set of variables from the highway system it was felt that multivariate regression would provide the most satisfying results.

A further requirement after regression has been chosen for the analysis, is the selection of a best regression equation. There is no statistical procedure for doing this, thus making it necessary to use personal judgement. Draper and Smith (6) in their discussion of searching for the best regression equation look at the following procedures:

(1) all possible regressions

(2) backward elimination

(3) forward selection

(4) stepwise

(5) two variations on the four previous methods and

(6) stagewise regression.

Of the six procedures, Draper and Smith found stepwise regression to be the best for multivariate analysis. The stepwise procedure involves the re-examination at every stage of the regression of the variables incorporated into the model in previous stages. A variable which may have been the best single variable to enter at an early stage may, at a later stage, be nonessential because of the relationships between it and variables in the regression. This process continues until no more variables will be admitted to the equation.

Although Draper and Smith recommend its use, they warn that stepwise regression can be easily abused; citing that "sensible judgement is still required in the initial selection of variables and in the critical examination of residuals." (6)

With this knowledge, stepwise regression was chosen for analysis purposes in this study. The availability of high speed computers and statistical computer packages makes possible a comprehensive approach to regression analysis using a large number of variables. The computer package used for this study was the Statistical Analysis System, developed by Barr and Goodnight. (3) This package was available through the Computer Center of VPI&SU and provided a variety of statistical procedures including the stepwise regression techniques used for demonstration purposes in this study. The Calcomp Plotter, also available in the VPI&SU Computer Center was used for constructing the scatter diagrams.

A. Dummy Variables in Regression Analysis

Dummy variables used in regression analysis provide the analyst with a single method of handling discontinuous and/or non-linear variables. Categorical variables such as sex, age, vehicle type, weather conditions, etc., can be properly treated without violating the basic assumptions made in using regression techniques.

The dummy variable technique can be applied to introduce non-linear independent variables into a regression model. These non-linear variables are changed to a dummy format by grouping them into a number of discrete classes, each indicating a value or range of values which the original variable falls within. A dummy variable is then assigned to each of the discrete classes. For example, the variable in this study that was used as a dummy variable was lane width. When the relationship between accident rate and lane width was examined, four distinct data groupings were found, one for each lane width (9ft., 10ft., 11ft., 12ft.) in the data set. By indicating only the dummy class into which an observation falls, errors often made in recording data of the explanatory variables become less troublesome in their effect on the response variable. This technique helps to eliminate the biases that are usually involved in assuming the form of functional relationship between the variables. This is because it allows for the identification of non-linear relationships without any previous assumption concerning the character of non-linearity. Therefore, assumptions such as linearity do not have to be made.

Dummy Variable Technique

In the dummy variable technique a value of one is assigned to the dummy variable corresponding to the class within which the value of the observed independent variable falls; all other dummy variables take on the value of zero. An example of this would be driver sex. As can be seen, the values for male and female drivers involved in traffic accidents can be expressed in a simple matrix form.

MALE	1	0
FEMALE	0	1

Thus, a listing of the data would show a value of 1,0 for male drivers in accidents and a value of 0,1 for female drivers in accidents.

The estimating equation generally takes the form

$$Y = b_o + \sum_{i=1}^{n} b_i x_i + \sum_{k=1}^{s} \sum_{j=1}^{c} b_{jk} Z_{jk} \qquad (2)$$

where: Y = the estimate of the dependent variable

b_o = regression constant

b_i = regression coefficient for the ith linear dependent variable denoted by x_i

b_{jk} = coefficient of the jth dummy variable in the kth dummy variable set

Z_{jk} = 1, if observed variable falls into the jth class and 0 if not in the jth class

n = total number of linear variables

s = total number of dummy variable sets

c = number of classes within a dummy variable set

The least squares criterion can then be used to produce a regression surface where the sum of the squared deviations of all the observations about it is a minimum. The solution of the developed "normal" equations yield the desired estimates of this intercept and the regression coefficients. (8) However the introduction of all classes of each set of dummy variables into the regression model will cause the normal equations to become indeterminate. This occurs because there are more coefficients than there are independent normal equations based on the least squared criterion.

To remove the indeterminancy, one of the coefficients in each of the dummy variable groups is set equal to zero and then the regression is performed. In a report by Hobeika (8) a method is shown which evaluates the regression coefficents of dummy variables. This method is derived from two constraint approaches, one in which a coefficient measures the net effect

other approach obtains regression coefficients which indicate differences from the overall mean of the dependent variable rather than from the individual classes.

Even though the two approaches produce different regression coefficients when they are each imposed on the normal equations it was shown that they differ only by a constant. Therefore a relationship was developed from the coefficients of the two approaches to produce an adjusted regression coefficient. This relationship is as follows (8):

$$\overset{*}{b}_{jk} = b_{jk} - \sum_{j=1}^{c} P_{jk} b_{jk} \qquad (3)$$

where:

$\overset{*}{b}_{jk}$ = the adjusted partial regression coefficient of the jth dummy variable in the kth dummy variable set indicating the influence of the jth dummy class in the kth dummy variable set on the response variable with respect to the overall mean.

b_{jk} = the uncorrected regression coefficient indicating the relative influence of the jth dummy class in the kth dummy variable set on the response variable as compared to that of the omitted dummy class in the set.

P_{jk} = a weighting indicating the fraction of the sample in the jth dummy class in the kth dummy variable set.

c = total number of classes in the set.

This approach is demonstrated in the following chapter by using lane width as the dummy variable set. One disadvantage of this dummy variable approach is that one cannot perform the statistical t-test used for testing whether a given partial regression coefficient is significantly different from some stated B_0 (usually taken as zero).

The dummy variable technique can be extended one step further. A beta coefficient can be determined for each group to estimate that group's contribution in the overall model. This holds only when more than one dummy variable group is being analyzed. The following beta relationship is defined as follows (8):

$$B_s = \frac{1}{S_y} \sqrt{\frac{\sum_{j-1}^{c} n_j B_j^2}{\sum_{j=1}^{c} n_j}}$$

where: B_s = beta coefficient of dummy variable set S

S_y = standard deviation of the dependent variable

n_j = number of observations in each class of dummy variable sets

B_j = adjusted partial regression coefficient

The largest beta coefficient found for a specific variable group signifies that it has the greatest effect on the response variable.

Many times a given class of dummy variable group or set will not have a sufficient F-level to enter the model. The first classes chosen may not have been appropriate and this requires a regrouping of classes, usually into a smaller number than first used. However, after eventually reducing the number of classes to the minimum of two and the class will not enter the equating, other means will have to be taken. Often the altering of the range of each class will have an effect on the class entering the equation.

It is possible to select classes or regroup them by analyzing the simple correlation matrix. It is often appropriate to break the given set into many classes and restructure the range of classes after viewing the correlation matrix. Classes that tend to have correlation coefficents in the same range and are correlated the same (positive or negative correlation) can many times be grouped into one class.

It should be pointed out that regression coefficients for dummy variables must undergo the same scrutiny as those for regression models with linear models. The direction and numerical value for these coefficients should be logical for the variables and data used. In summing up the use of dummy variables Suits (14) states,

> "One occasionaly encounters suspicion of dummy variables and a feeling that somehow something not quite respectable is involved in their use . . . Perhaps part of the trouble lies in the use of the term "dummy" variable. There is nothing artificial about such variables; indeed in a fundamental sense they are more properly scaled than conventionally measured variables."

B. Data Analysis Procedure for Accident Estimation

Following the variable selection and collection of data it will be necessary to select the best regression model for the data. The following procedure, diagrammed in Figure 2, was developed to aid the analyst in applying regression techniques to highway safety analysis. This procedure will enable the analyst to examine the regression variables graphically and statistically. It also allows for improvements in the estimating quality of the regression equation through the use of dummy variables and data grouping.

1. Construct Scatter Diagrams for Individual Varaibles

Scatter Diagrams are the graphical representations of the dependent variable, Y plotted against the independent variables, x_i, where i=1, 2, 3 . . . These diagrams will give some indication of the relationship that exists between the variables, i.e., a straight line or curve relationship. However, as the case may be in many real world situations, clearcut relationships do not always exist and it may be necessary to alter the variables through transformation to improve the (x,y) relationships.

Transformation of the variables will have to be made if no linear relationship exists between the dependent and independent variables.

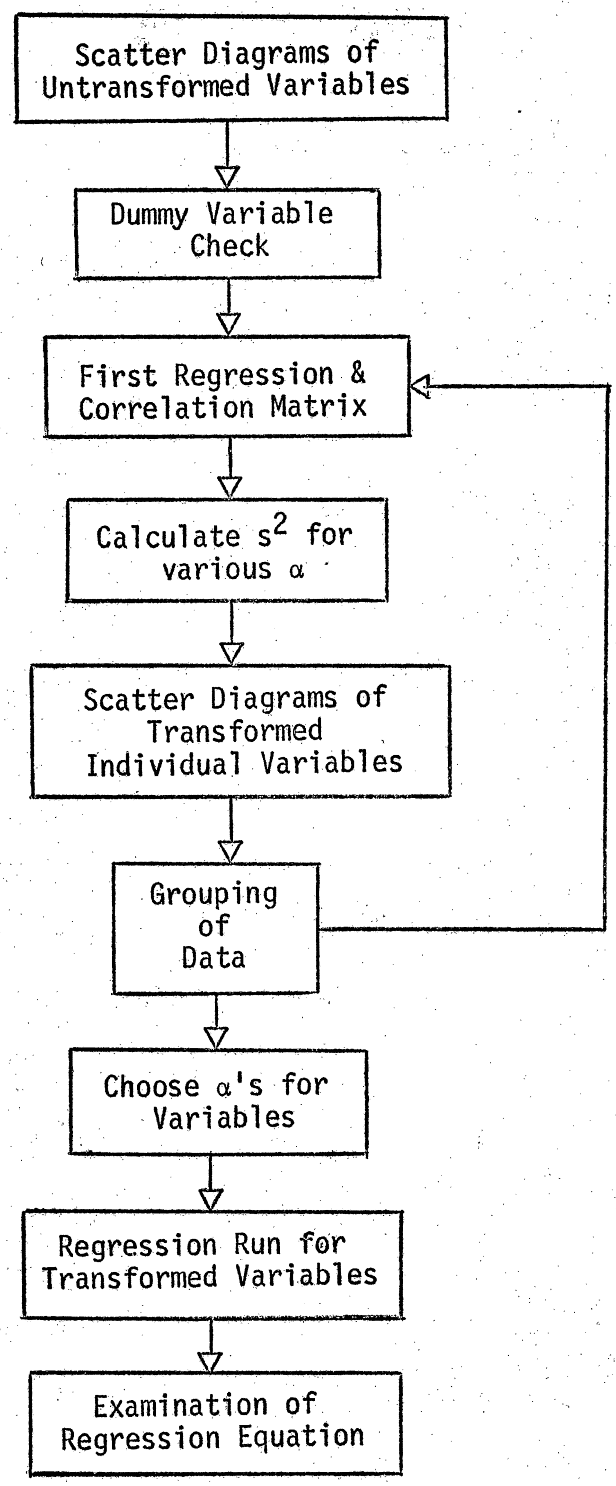


Figure 2. Regression Methodology Diagram

Choosing which variables are to be substituted for dummy variables requires more discretion than for transformation. Only those variables which are categorical and/or nonlinear in nature as explained in part A previously, should be changed to dummy variables. At this point in the procedure it may not be possible to tell whether or not a variable will fall into this classification; therefore it may be necessary to make a decision only after transformation of the variables has been made.

2. Dummy Variable Check

After the scatter diagrams have been constructed, the analyst should look for those plots that show distinct groupings of data points. The analyst should consider the independent variable of such plots as possible dummy variables.

3. First Regression and Correlation Matrix

The correlation matrix should first be computed for gaining some insight into the reasonableness of an analytical solution. The analyst should be looking at the correlation between the dependent and independent variables and whether or not any x-variables are correlated to each other. It is desirable to have a high correlation, approaching 1.00 between the dependent and independent variables. This will depend on the variables chosen and the type of data collected for those variables.

Using those variables (untransformed at this point) finally decided upon for use in the analysis, the first regression run should be made for each dependent variable if there is more than one. These first regression runs will inform the analyst which variables will need to be transformed and which cross products of the variables could possibly be made.

It will be necessary to check the list of variables in the order in which they entered the regression equation in order to determine what contribution they make to the R^2 value and what F-value each has. This is important because a cutoff point may have to be made including or excluding variables from the regression equation. This will in most cases depend on the number of variables entered into the regression equation because of their F-values. Past studies have used cutoff points where variables contributed less than one percent to the R^2 value. This decision is usually left to the analyst or will be done within the particular computer program used.

The combinations or cross products of the variables, [$x_k=(x_i \cdot x_j)$ where k = new variable, i,j = initial variables], should be considered for those individual variables found in the final equation of the first regression run. Dart and Mann (4) explain the use of cross products of the variables from the roadway component of the highway system by stating that,

> "The large number of highly significant cross products means not necessarily that one variable depends on the other but that elements of the roadway are generally built as compatible units. For example, very seldom would one suggest today that a roadway be built with a high degree of adequacy for surface width but a low degree of adequacy for sight distance."

If cross products are to be used in the regression analysis they should be made before transformation of the variables are made, following the first regression on the individual variables.

4. Calculate s^2 for Various α

This step involves the procedure for selection of the "best" transformation, α, where $Y = b_0 + b_1 x_1^{\alpha} + b_2 (x_1 x_2)^{\alpha}$. . . for each variable and cross product variable from the previous step. The purpose of the transformation, α, is to improve the degree of estimation of each variable thereby improving the estimation quality of the entire regression equation. As mentioned previously, if an (x,y) graphical representation does not show a clear cut relationship, either linear or curved, then some transformation of the variables may be necessary.

The object here is to produce the largest possible square ratio or F-ratio. (See analysis of variance table below)

Source	df	SS	MS	F-ratio
Regression	1	SSR	MSR	$\frac{MSR}{s^2}$
Residual	n-2	SSE	s^2	
Total	n-1	SST		

where; n = number of observations in the data set,

df = degrees of freedom; this number indicates how many independent pieces of information involving (n) independent observations are needed to compile the sum of squares,

SS = sum of squares

MS = mean square,

SSR = sum of squares due to regression,

SSE = sum of squares about regression,

SST = sum of squares about the mean or, SSR + SSE,

MSR = regression mean square or $\frac{SSR}{df}$, and

s^2 = residual mean square or $\frac{SSE}{df}$.

This can be accomplished by choosing α in such a way that MSR is large and s^2 is small. It should be pointed out that transformation is being made only on x and not y so that the mean squares are examined instead of the ratios of mean squares. For a given SST, SSE is minimal when SSR is maximal. Therefore s^2 for various α's is needed for each variable in the data set. This method of choosing a transformation was proposed and demonstrated by Draper and Hunter (5) for a simple two variable regression equation. The same basic technique is to be applied to polynomial regression analysis in addition to simple linear regression analysis. This is demonstrated in Chapter Five.

The s^2 should be calculated for each variable at several different α levels to obtain a sufficient group for analysis purposes. Draper and Hunter (5) used α levels of $-2.0<\alpha<1.5$ for their example. For the demonstration in chapter 5 this was expanded to include $\alpha=2.0$ which set the limits at $-2.0<\alpha<2.0$. It may be necessary to go beyond these limits until a minimal value of s^2 can be found; just how far is left to the discretion of the analyst and his knowledge of statistical procedures.

It may be useful at this point to construct the plot of s^2 vs α to obtain a more accurate transformation. An example plot of s^2 vs α is given in Chapter Five. However it is felt that α's rounded off to the nearest 0.5 will provide a sufficient transformation for the variables and therefore selection of the α's at increments of 0.5 and choosing the smallest s^2 value would be as easy and sufficient as plotting s^2 vs α.

5. Construct Scatter Diagram for Transformed Individual Variables

A further check of the data following step 4 will be the construction of scatter diagrams for the transformed independent variables versus the dependent variables. These scatter diagrams will be for the purpose of ensuring that the transformations made on the variables are reasonable and show an improvement toward linearity in the relationship between x and y. It should be pointed out that although scatter diagrams are found for transformed individual variables, the same could apply to transformed cross product variables. If the analyst has any doubts about the relationship between the dependent and cross product variable, the scatter diagram can be constructed and viewed for those variables.

If no change can be seen between scatter diagrams of untransformed variables and transformed variables, i.e., no improvement in the linearity, then the variable should be considered as a dummy variable or else be dropped form the analysis. A recheck of the correlation matrix for variables may be helpful in deciding what significance to place on the variable for further use in the regression analysis. A low correlation with the dependent variable may indicate little or no contribution to the equation and/or R^2 value and therefore it should be deleted from analysis.

One advantage of the scatter diagrams at this point is the fact that plots of variables that are non-categorized will sometimes show a distinct break in the data, for example, a variable may have most of its observations with the same value and the rest of the observations may fall into a distinct or random pattern. This may indicate to the analyst that regression calculations should be made for groups of data for a variable.

A further advantage of the scatter diagrams would enable the analyst to substitute transformed variables for dummy variables. If it was found that transformation would not improve the relationship with the dependent variable they may be deleted from the analysis.

6. Grouping of Data

If, following the plotting of the transformed individual variables, there appears to be any grouping of data points for the independent variables the analyst should consider a separate regression analysis for each group. Situations for grouping of data arising in highway accident analysis may be straight, flat roads versus curved, hilly sections of roadway and etc. Therefore regression equations could be obtained for different characteristics of the roadway and for different parts of a state or region in which those roads exists. If grouping is made the analyst should begin the analysis of each group at step 3.

7. Choose Transformations

If the scatter diagrams from step 5 show an improvement in the two-variable relationship and no distinct breaks or groupings appear in the plot, then the appropriate α should be chosen for each variable to be included in the next regression calculation. As mentioned in step 5, scatter diagrams may not be needed for cross product variables, therefore the minimum α found for these variables can be the α used for the regression calculation and no grouping of the data will occur among these variables.

8. Regression Run for Transformed Variables

This run should include all those transformed variables from step 7 and the dummy variables, if any, for the data base as a whole. The regression run for transformed variables from the grouped data will also be made at this point. This will include as many groups as the analyst wishes to use.

9. Examination of the Regression Equation

Once the final regression run is made, it will be necessary to examine and justify those values found. The following is a list of values that should be considered when performing regression analysis:

a. R^2 value - This value explains the total variation about the mean Y explained by the regression equation. Expressed as a percentage, a high R^2 value approaching 1.00 is obviously sought. Each significant variable entered into the regression equation should be looked at to determine how much the variable contributes in explaining the variation. Variables contributing little to the R^2 value, for example less than 1%, are of limited value and should be removed, if this is not done by the computer program used.

As Draper and Smith (6) state,

> "The addition of a new variable may increase the accuracy of the estimate of the response. This is because the reduction in the residual sum of squares may be less than the original residual mean square. Since one degree of freedom is removed from the residual

> degrees of freedom as well, the resulting mean square may get larger."

Since the residual mean square is a ratio of residual sum of squares to the degrees of freedom, a reduction in the degrees of freedom would indeed produce a large mean value and thus a larger R^2 value. As variables are added and the degrees of freedom are reduced then the R^2 increases because the mean square value increases. The authors of the preceding quote caution the analyst to be sure that an improvement in R^2 due to adding a new term to the model has some real significance and is not due to the fact that the number of parameters in the model is getting close to the saturation point - that is, the number of observations. This is a particular danger when there are repeat observations. The following is a simple example of the problems with repeat observations which occur for 5 different lane widths in groups of ten repeat observations. A simple regression equation is sought relating lane width to total accidents. There are essentially five pieces of information, represented by five mean values, and fourty-five error degrees of freedom for pure error, nine at each point. Therefore a regression model containing a variable for each lane width, or a five variable model will provide a perfect fit to the five means and may give a very large value of R^2, especially if the experimental error is small compared with the spread of the five means. A regression model containing only one variable representing lane width would probably give a low, unexceptable R^2 value but the equation would be a true representation of the relationship existing between total accidents and lane width.

b. Standard Error of the Estimate - A check should be made to determine if the standard error of the estimate decreases or increases from the first to final step of the regression procedure. The standard error of the estimate is often expressed as a percentage of the mean response. A significant decrease in this value shows an improvement in the accuracy of the estimation. Some computer programs or packages may only give the error of the estimate for the final equation, therefore a manual calculation may be needed to check if the value increased or decreased.

c. B_i values - These values include the B_o value or constant term in the regression equation and coefficients, B_i . . . B_n, and precede the variables in the regression equation. The B_o term should be examined for its numerical size and directional significance. Large constants, especially negative ones, should be viewed with caution. The size of the constant term may also depend on the type of data used. The same scrutiny should be given the variable coefficients to determine whether they are reasonable and logical for the variables they represent. The size and direction of the coefficient should also depend on the data used.

d. Correlation and Covariance Matrices - If there is a serious doubt about the R^2 values and/or the B_i values then a close look is needed at the correlation and covariance matrices. Both give indications of the relationships that exist between the variables. Highly correlated variables (those approaching a correlation value of 1.00) will often times give biased information and therefore it is important to look at the correlations for each variable to determine whether or not a variable or variables should be transformed, combined and/or deleted from the analysis. The covariance matrix also gives a degree of dependency of one variable or another. The smaller the covariance the more independent variables are of each other.

It would be wise to look at the R_{xy} correlation coefficients, for the final variables in the equation, for these measure the association between each x and Y. (R_{xy} is the correlation coefficient for x and Y, the values fall between 0 and 1) However, the fact that the correlation R_{xy} is nonzero implies only that there is association between the values x_i and Y_i, i = 1, 2 . . . n and does not itself imply any sort of causal relationship between x and Y. This false assumption can lead to erroneous conclusions.

The methodology developed in this chapter is a tool to aid in the analysis of the wide variety of variables that compose the highway system. Categorical and non-categorical (numerical) data can be analyzed at the same time and the results will allow the analyst to determine the significance of each variable of each type from the same regression equation.

The ability to evaluate the regression variables graphically and statistically enhances the quality of the final regression equation for accident estimation. Also, the method for choosing transformations is an improvement over a trial and error type method.

The overall effectiveness of this regression methodology can only be evaluated by actual application. This is accomplished in Chapter

Five, where dummy variables and data grouping are also used.

Chapter Five

REGRESSION METHODOLOGY APPLICATION

The application of the regression methodology that was developed in the previous chapter analyzes data from the roadway/environment element of the highway system for cell 7, or the pre-crash phase of the highway accident matrix. Therefore the application identifies those roadway/environment variables that contribute significantly to highway accident occurrence.

The regression methodology applied to the chosen data was performed firstly to illustrate the capabilities and strengths of the developed methodology and secondly to test several hypotheses that were made about the data used. These involved the use of dummy variables and analysis of grouped data. Both of these were easily tested using the developed regression methodology from Chapter Four.

The variables used in this application were those chosen as variables for the study from Louisiana from which the data was obtained. The study, entitled "Relationship of Rural Highway Geometry to Accident Rates in Louisiana", was conducted by Dr. Lawrence Mann, Jr, and Dr. Olin K. Dart of the Louisiana State University. (4) Table 1, gives a listing of accidents and enforcement in Louisiana for a 10 year period. The study by Dart and Mann (4) covers the time period from 1962 through 1966.

A. Regression Variables

The following independent variables were used as defined by Dart and Mann (4):

Table 1

LOUISIANA HIGHWAY ACCIDENTS AND ENFORCEMENT FROM 1957 TO 1966

Year	Rural Accidents	Rural Fatalities	Automobile Fatalities	Death Rate per 100 mvm[a]	Hazardous Citations	Ratio of Citations to Accidents
1957	17,052	568	817	7.3	-	-
1958	16,301	608	845	7.4	-	-
1959	16,972	560	803	7.2	45,501	2.68
1960	15,623	594	841	7.4	51,699	3.31
1961	15,732	576	779	6.7	63,351	4.02
1962	17,299	546	770	6.4	71,876	4.16
1963	19,638	624	846	8.3	67,017	3.41
1964	23,934	791	1,037	9.7	68,667	2.82
1965	25,904	857	1,106	9.0	78,992	3.05
1966	29,916	917	1,232	9.4	89,061	2.98

[a]Accidents per 100 million vehicle-miles

1. Average daily traffic (ADT) - total 2-way traffic volume for a 24-hour period. Variable x_1.
2. Traffic volume ratio - the ratio of the peak hour traffic volume (2-way) for a highway section to its corresponding service volume at level of service B as defined in the Highway Capacity Manual. Variable x_2.
3. Lane width - width of the lane in feet measured from the center line to the pavement edge. Variable x_3.
4. Shoulder width - width of the shoulder, in feet, measured from the pavement edge to the shoulder edge, generally indicated by a change in slope between the shoulder and the side slope. Variable x_4.
5. Pavement cross slope - slope of the pavement from the centerline to the shoulder, measured in feet per foot of pavement width. Variable x_5.
6. Horizontal alignment - the percentage of the length of a highway section that has a horizontal highway curvature in excess of 3 degrees. Variable x_6.
7. Vertical Alignment - the percentage of the length of a highway section that has a highway gradient in excess of 3 percent. Variable x_7.
8. Continuous obstruction - the percentage of the total length of a highway section that has some roadside feature or obstacle that runs for more than a few feet on either or both sides of the roadway. Such a feature would be a deep

roadside ditch or steep side slope that presents an obstacle to a vehicle's safely leaving the roadway in an emergency at posted highway speed. Variable x_8.

9. Marginal obstructions - the total number of discrete objects on both sides, such as a driveway embankment, culvert, roadway culvert headwall, tree or telephone pole, per mile of a highway section within the cleared right-of-way. This is not to be confused with the marginal obstruction within 6 ft. of the pavement edge used in capacity analysis. Variable x_9.

10. Traffic conflicts - the total number of traffic access points on both sides per mile of highway section. These access points include only minor road intersections (intersection with major roads were considered as break points between study sections) and principle access driveways to abutting property along the highway section. Variable x_{10}.

11. Trucks - percentage of trucks found in the average daily traffic. Variable $_{11}$.

Table 2, (4) gives the value ranges for the independent variables.

The dependent variable, accident rates, was grouped by Dart and Mann (4), according to fatalities, injuries and also to account for the time of day and the road surface condition. Each group was based on exposure rates of 100 million vehicle miles each (MVM). The groups were as follows:

1. Accidents (Total),
2. Injuries,

Table 2

VALUE RANGES FOR INDEPENDENT VARIABLES

Variable	Value range
Traffic	
Trucks	2 to 33 percent of traffic
Traffic volume	190 to 11,933 vehicles per day
Traffic volume ratio	0.04 to 2.12
Width	
Lane width	9 to 12 ft
Shoulder width	1 to 12 ft
Cross slope	0.000 to 0.038 ft/ft
Alignment	
Horizontal alignment	0 to 57.5 percent of length > 3 deg
Vertical alignment	0 to 48.0 percent of length > 3 percent
Roadside friction	
Continuous obstructions	0 to 100 percent of section length
Marginal obstructions	0 to 34 per mile
Traffic conflicts	0 to 46 per mile

3. Fatalities,
4. Accidents during the day,
5. Accidents during the night,
6. Accidents on wet roads, and
7. Accidents on dry roads.

Total accidents per 100 MVM was used as the dependent variable in the regression application.

B. Methodology Application

The regression methodology was applied for three separate analyses. The first analysis (Section 1) started with the same variable set and number of observations as used by Dart and Mann (4). The purpose for this analysis was to use the methodology to show the improvements made in the estimating quality of the calculated regression equation over that of Dart and Mann.

The second analysis (Section 2) involved the use of lane width as a dummy variable. Although lane width was not found to be a significant variable in accident occurrence, it demonstrated the dummy variable technique and some of its advantages. This technique allows the analyst to determine the influence of each dummy variable class, i.e., lane width, on the dependent variable (accident rate). Such information could be valuable to highway designers. It should be noted that the same variables and transformations used in Section 1 were also used for Section 2. Lane width was the only variable changed.

For the third analysis (Section 3) a grouping of the data was made for horizontal and vertical alignment. Approximately 53% of the data observations showed little or no horizontal or vertical alignment (<5%), therefore roadway sections that were relatively straight and flat were grouped separately from those sections having hills and curves. The hypothesis tested in this analysis was that drivers on straight, flat roads would require less driving ability and actions than for curved, hilly roads, therefore roadway/environment variables for straight, flat roads would give a better explanation for accident occurrence than the hilly, curved roads.

The initial analysis of the total data set included attempts to match the results found by Dart and Mann. This proved to be successful and provided an interesting observation about the data set. It was discovered that data for the four-lane roads played a very significant part in the R^2 values. Since only 8 of the 246 roadway sections were four-lane roads a decision was made to exclude them and analyze only the two-lane roadway sections. The R^2 value for the two-lane sections produced an R^2 value approximately 50% lower than that for the total roadway sections. This could possibly be explained by the much higher average daily traffic and accident rates on the four-lane roadway sections. This is interesting to note and may indicate that separate analyses be made for two-lane and four-lane roads for future studies.

Section 1 (Total Data Set)

This analysis gives the step by step regression procedure developed in the previous chapter for the total data set. Also included in this section is the dummy variable check and the dummy variable matrix for lane width. The analysis for the dummy variable is given in the following section.

1. Scatter Diagrams - Figures (3) through (13) show the plots of each individual independent variable versus the dependent variable or total accidents. It can be seen from these plots that the points are distributed in a rather diffused fashion showing little or no definite relationship between the two variables. This indicates that some sort of alteration of the variables is needed, either by transformation or by the use of dummy variables. The plot of lane width versus total accidents clearly shows four distinct groupings of observations; therefore lane width was converted into a dummy variable for the second regression equation.

2. Dummy Variable Check - As pointed out in step 1, lane width versus total accidents shows four distinct groups, therefore it was possible to use lane width in the regression equation as a dummy variable. This was not done for the first regression calculation; but such analyses were performed for the second equation. The purpose in doing this was to compare the two and determine the effect on the results of the dummy variables. The following matrix shows how the lane width variable was converted to a dummy variable:

	9 ft	10 ft	11 ft	12 ft
9 ft	1	0	0	0
10 ft	0	1	0	0
11 ft	0	0	1	0
12 ft	0	0	0	1

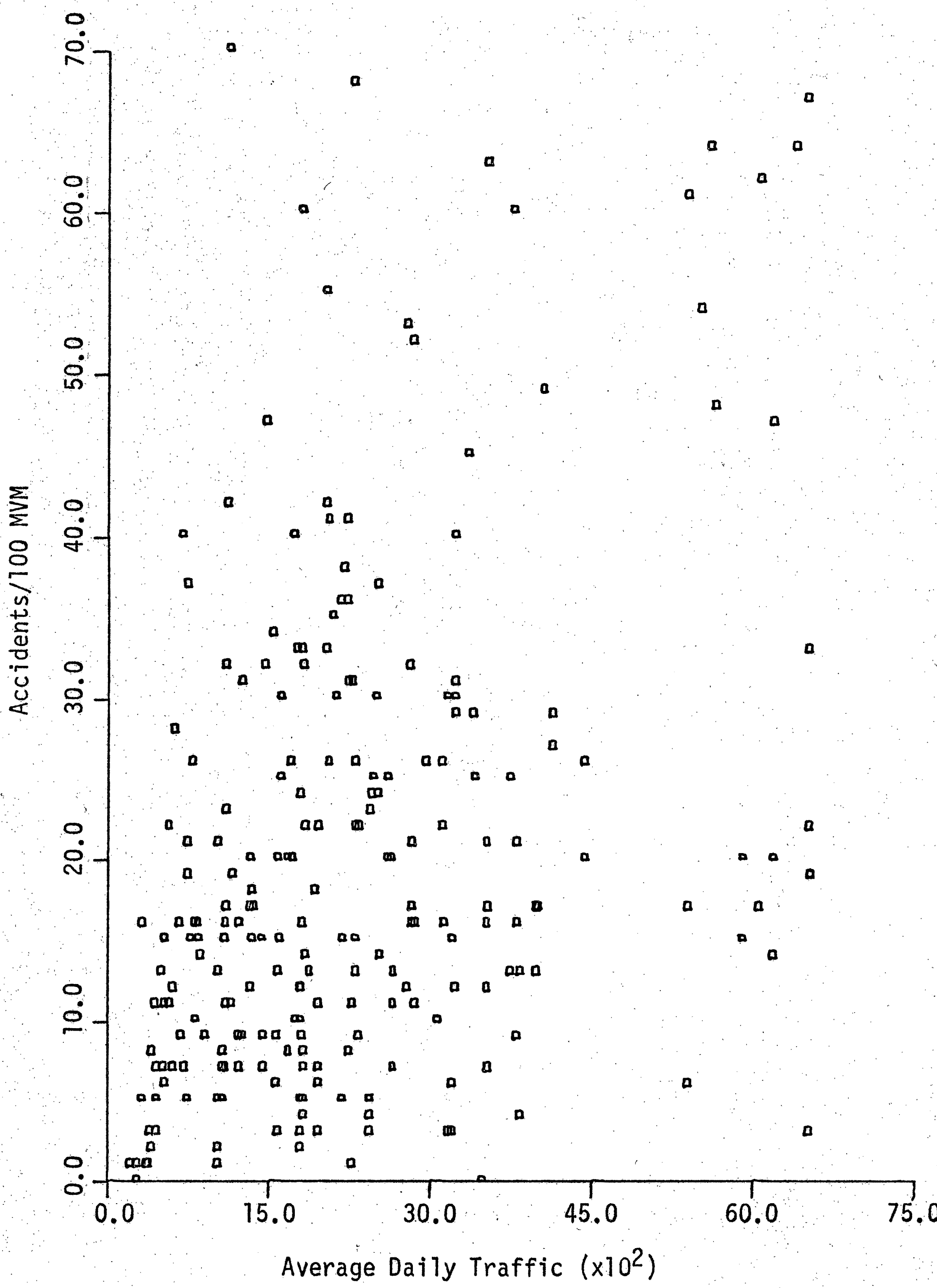


Figure 3. Accident Rate Versus Average Daily Traffic

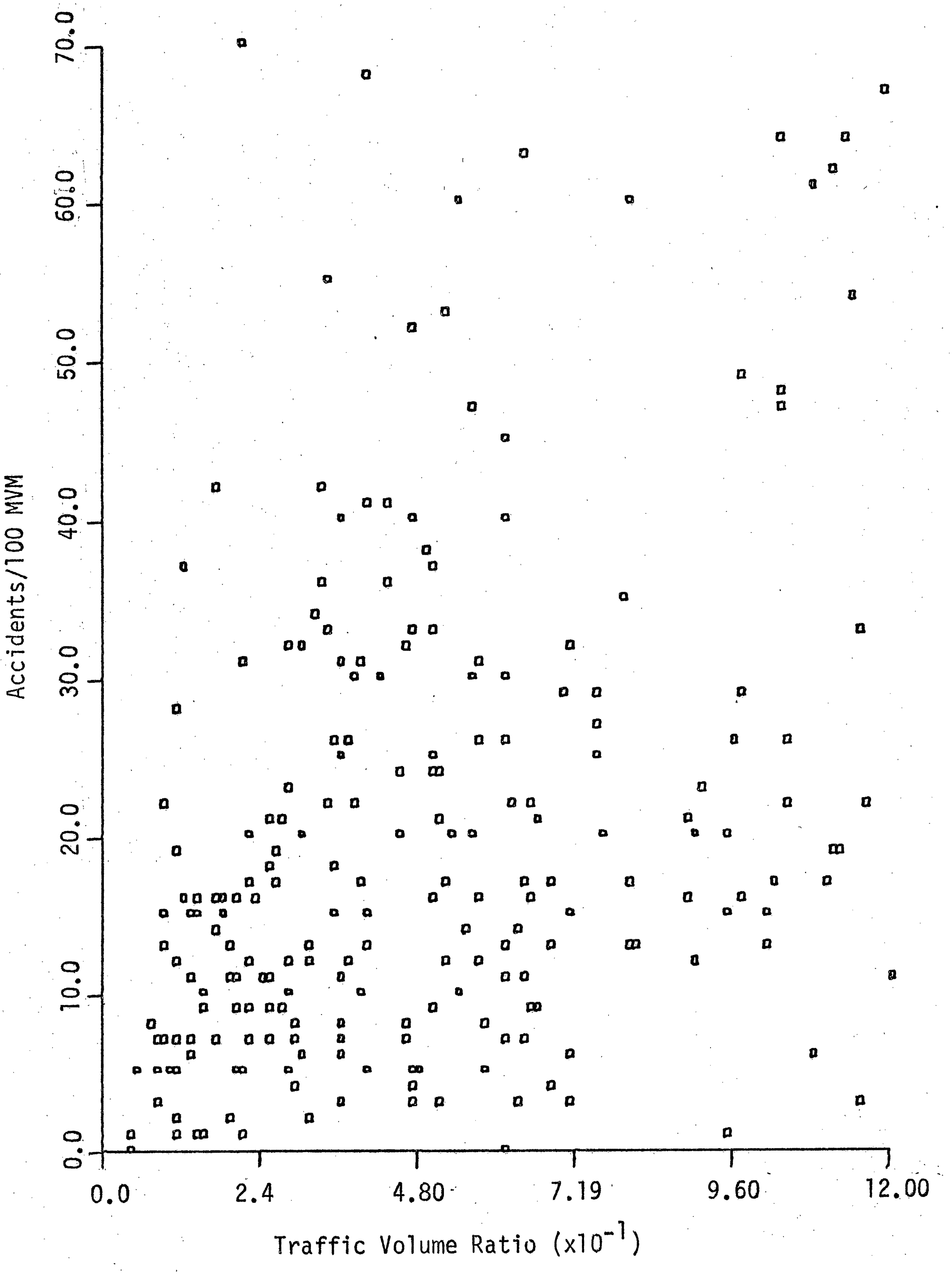


Figure 4. Accident Rate Versus Traffic Volume Ratio

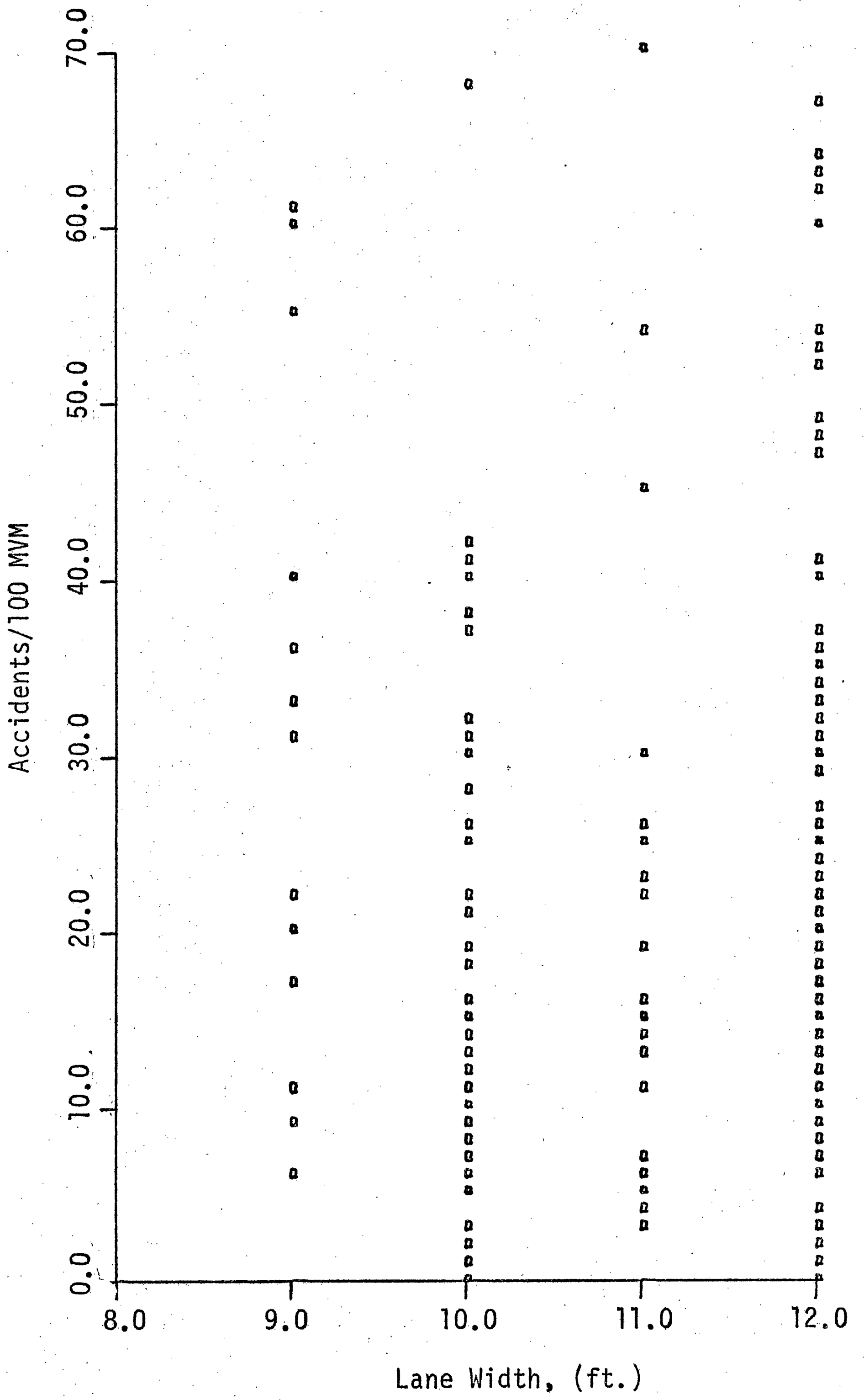


Figure 5. Accident Rate Versus Lane Width

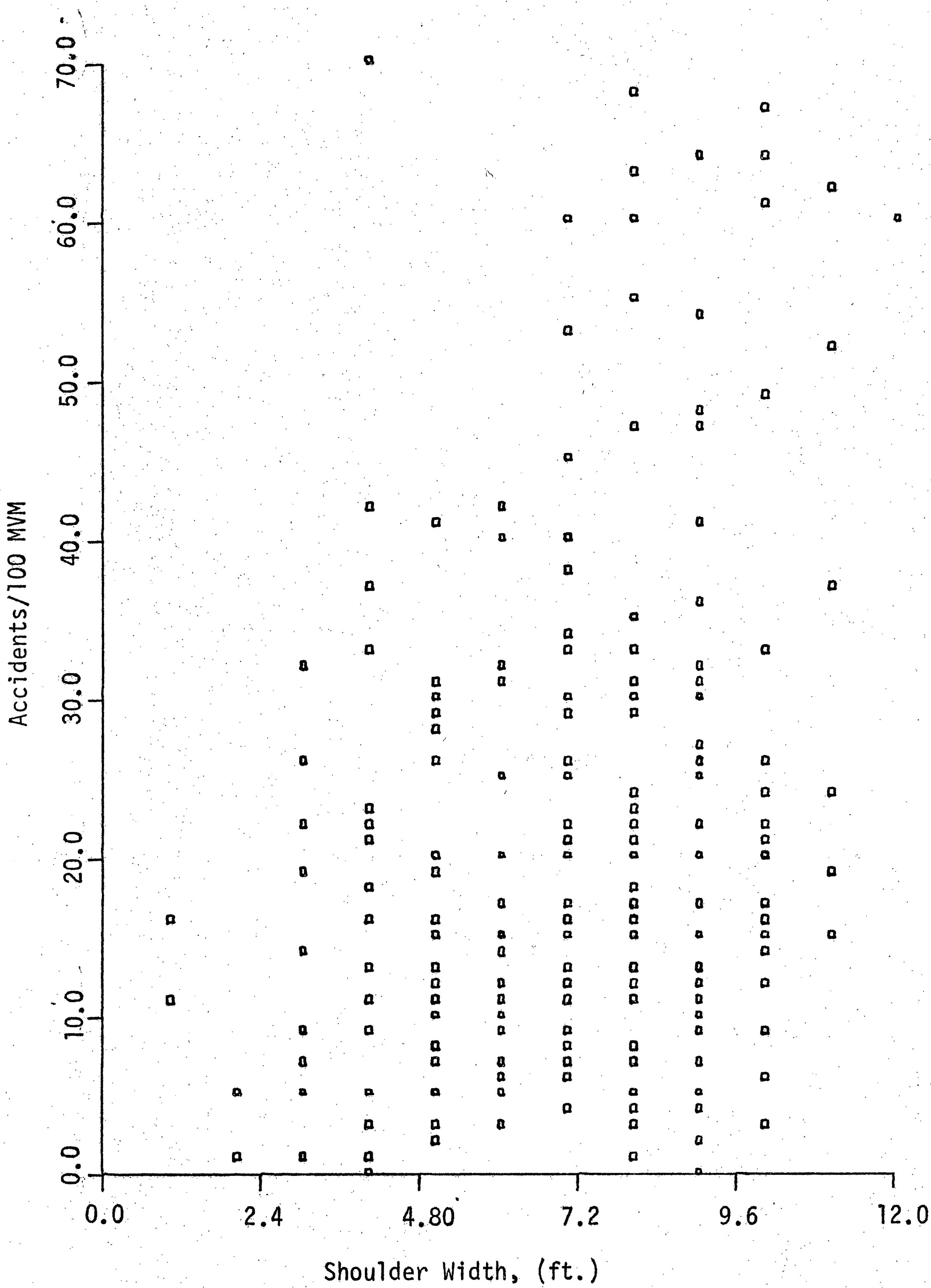


Figure 6. Accident Rate Versus Shoulder Width

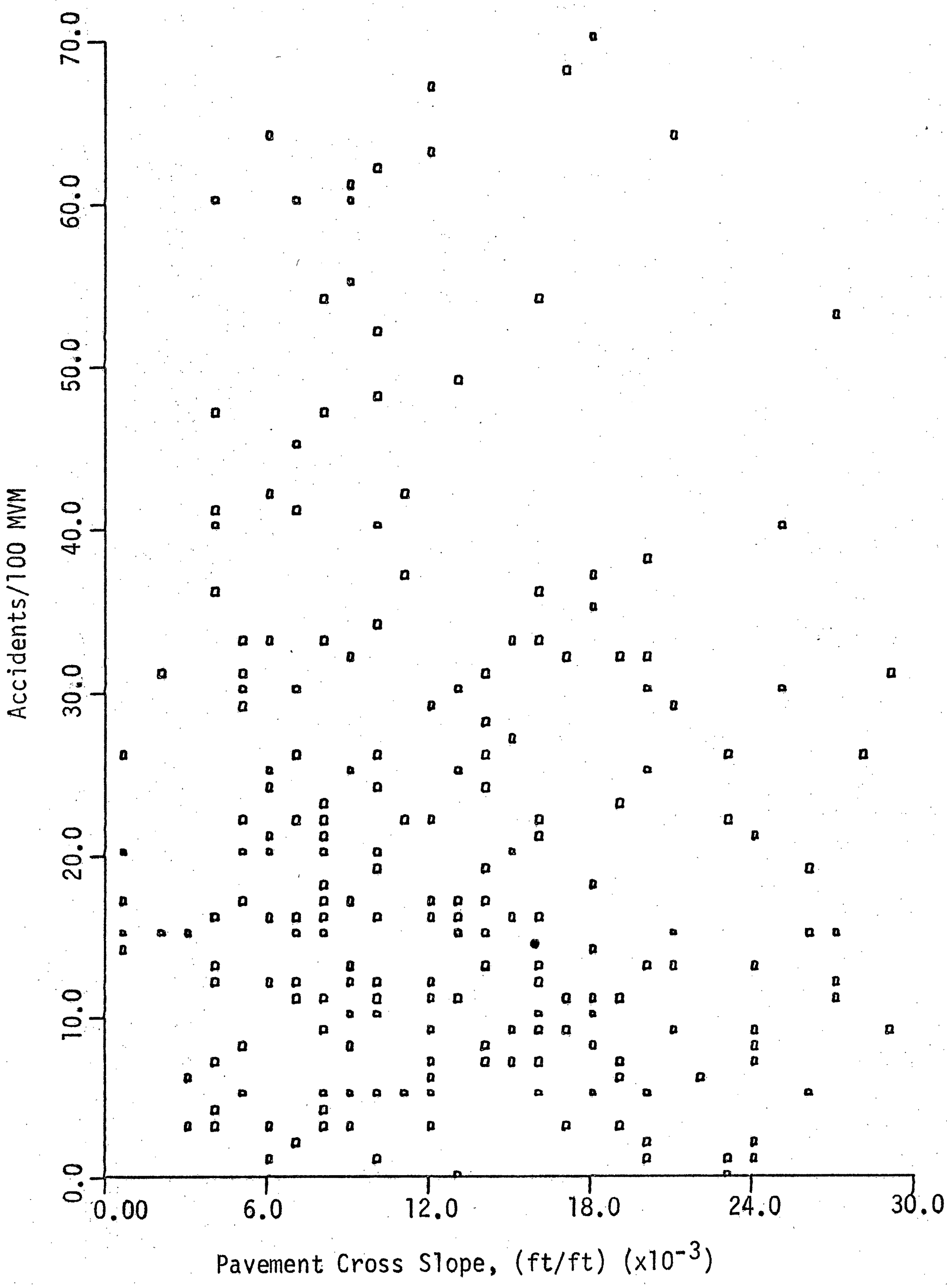


Figure 7. Accident Rate Versus Pavement Cross Slope

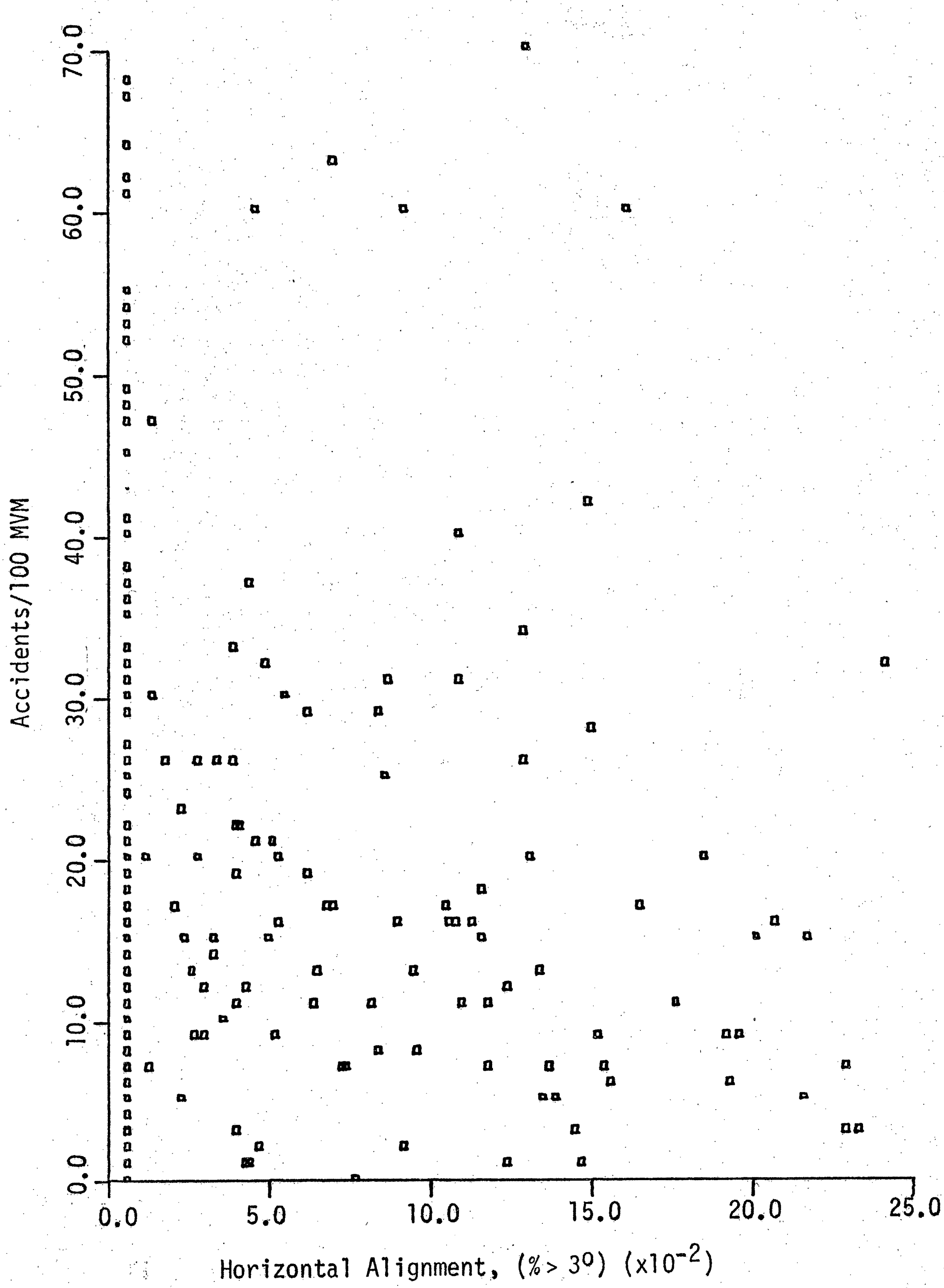


Figure 8. Accident Rate Versus Horizontal Alignment

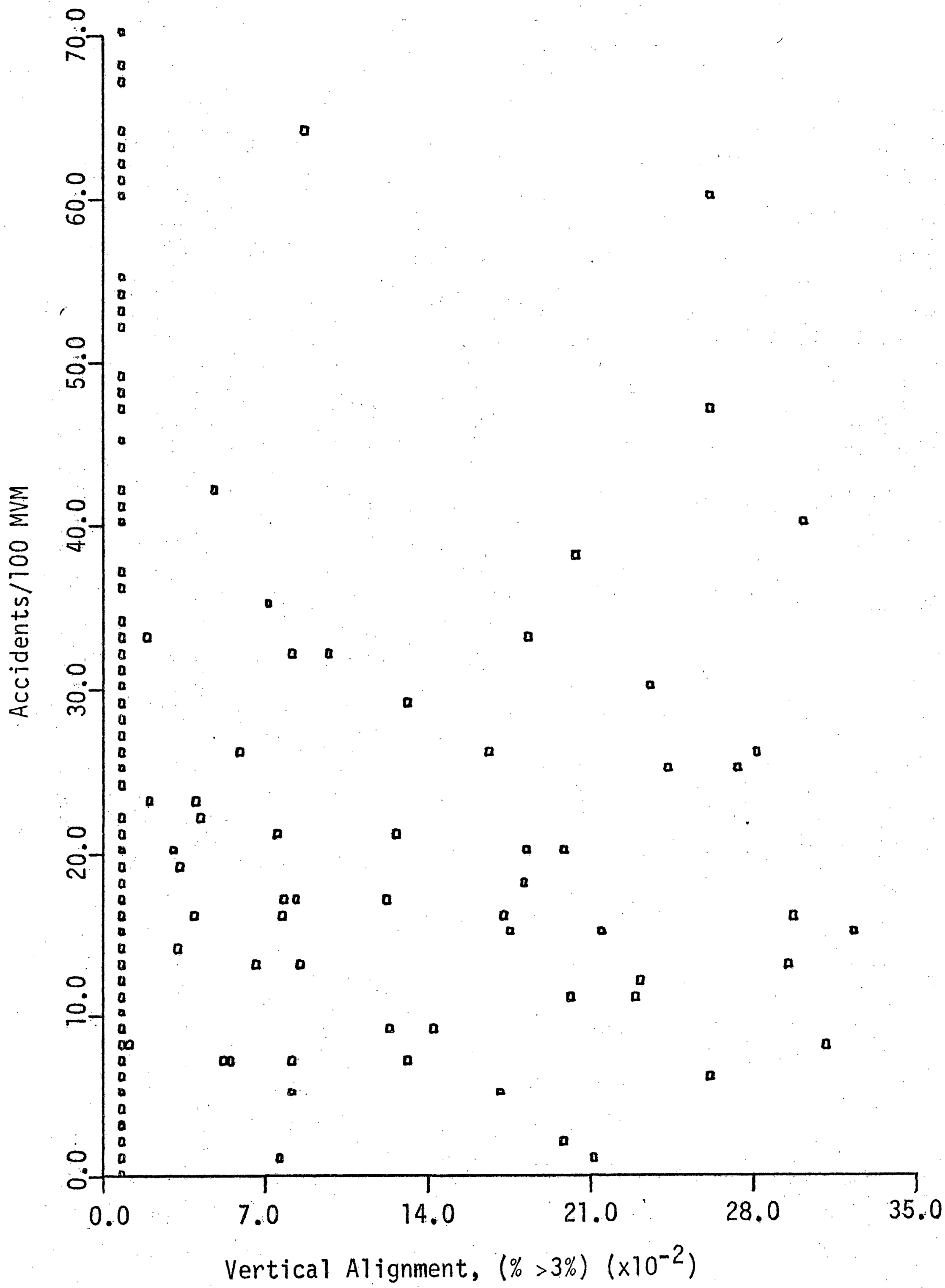


Figure 9. Accident Rate Versus Vertical Alignment

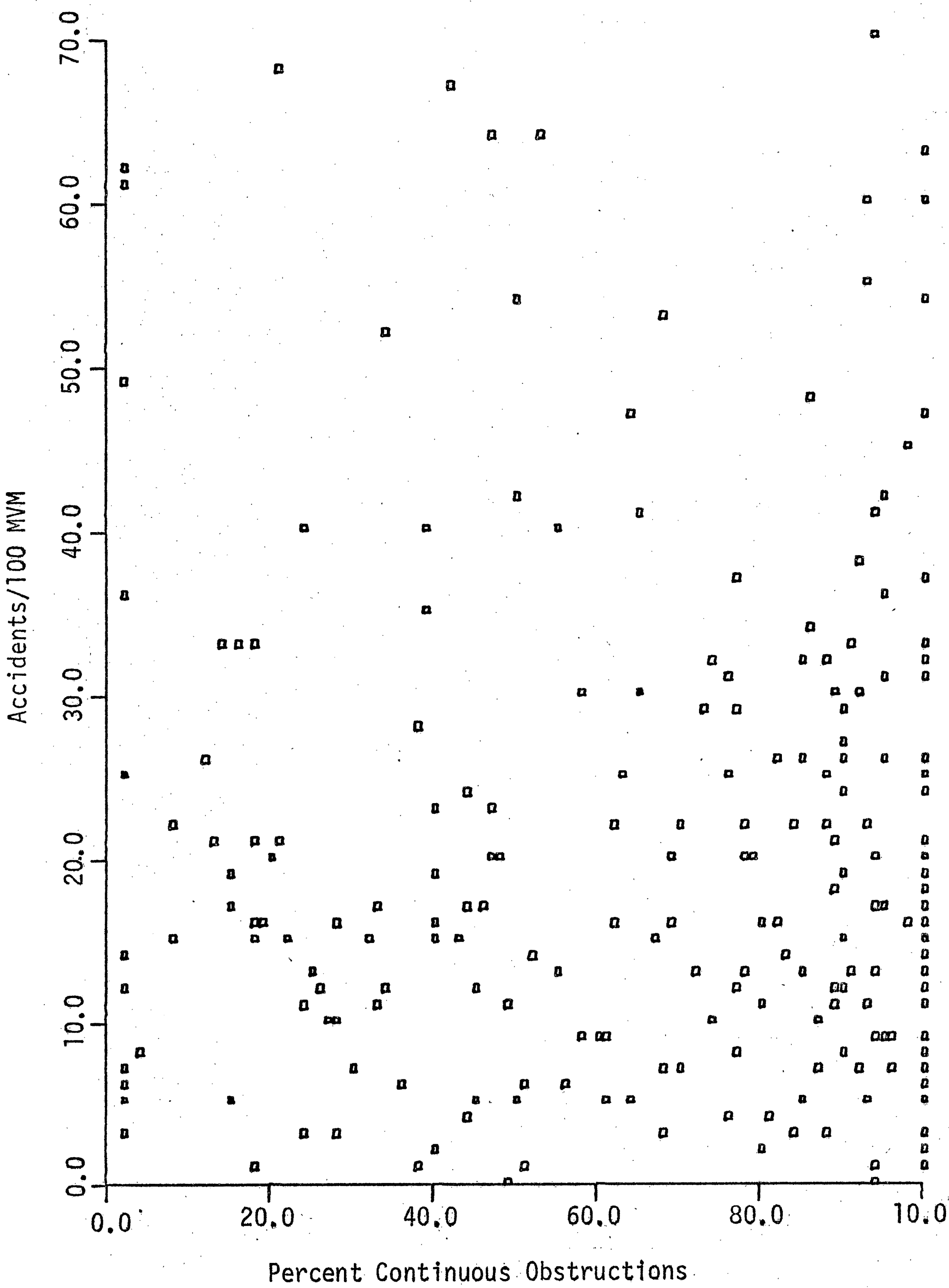


Figure 10. Accident Rate Versus Percent Continuous Obstructions

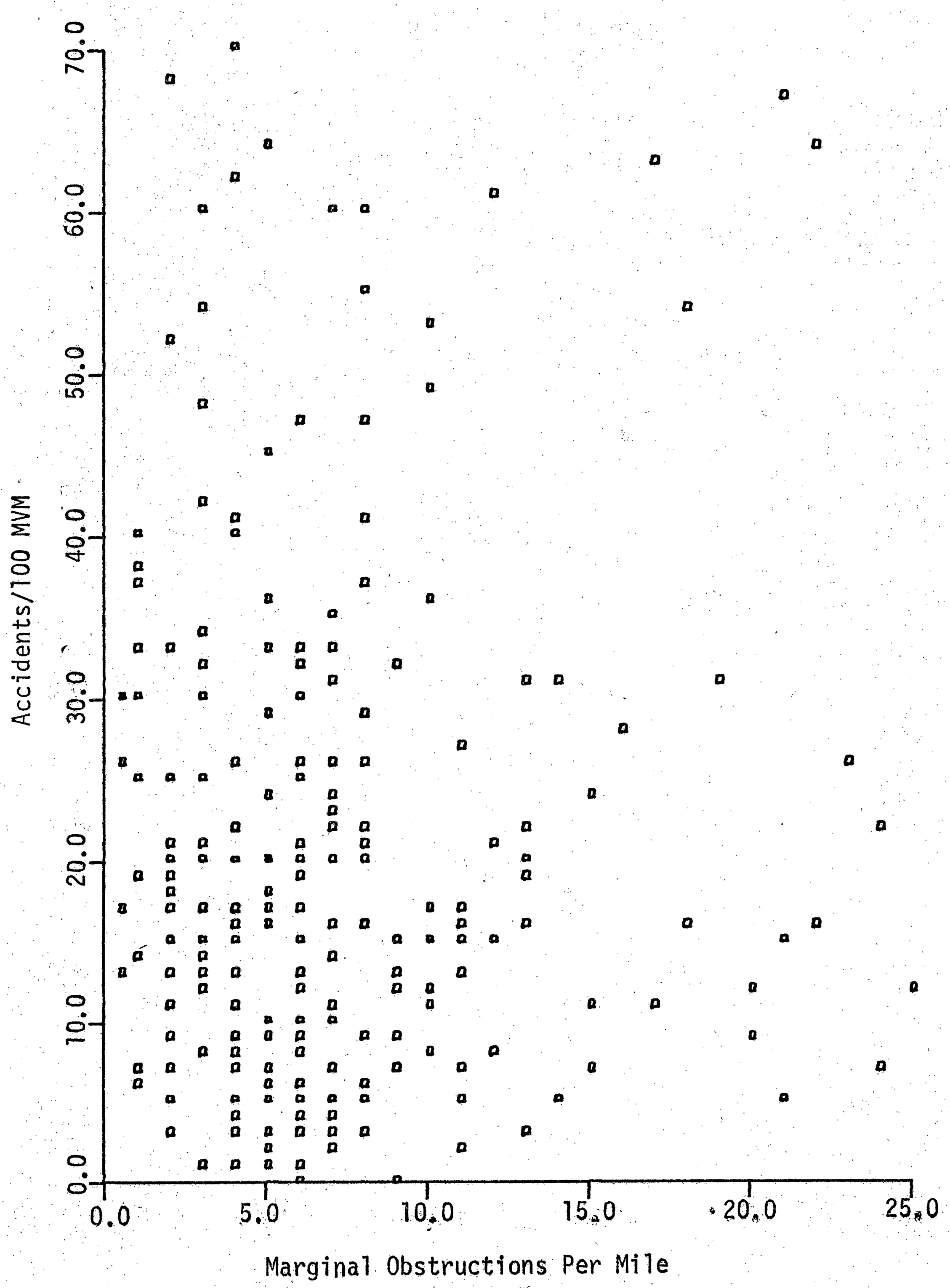


Figure 11. Accident Rate Versus Marginal Obstructions Per Mile

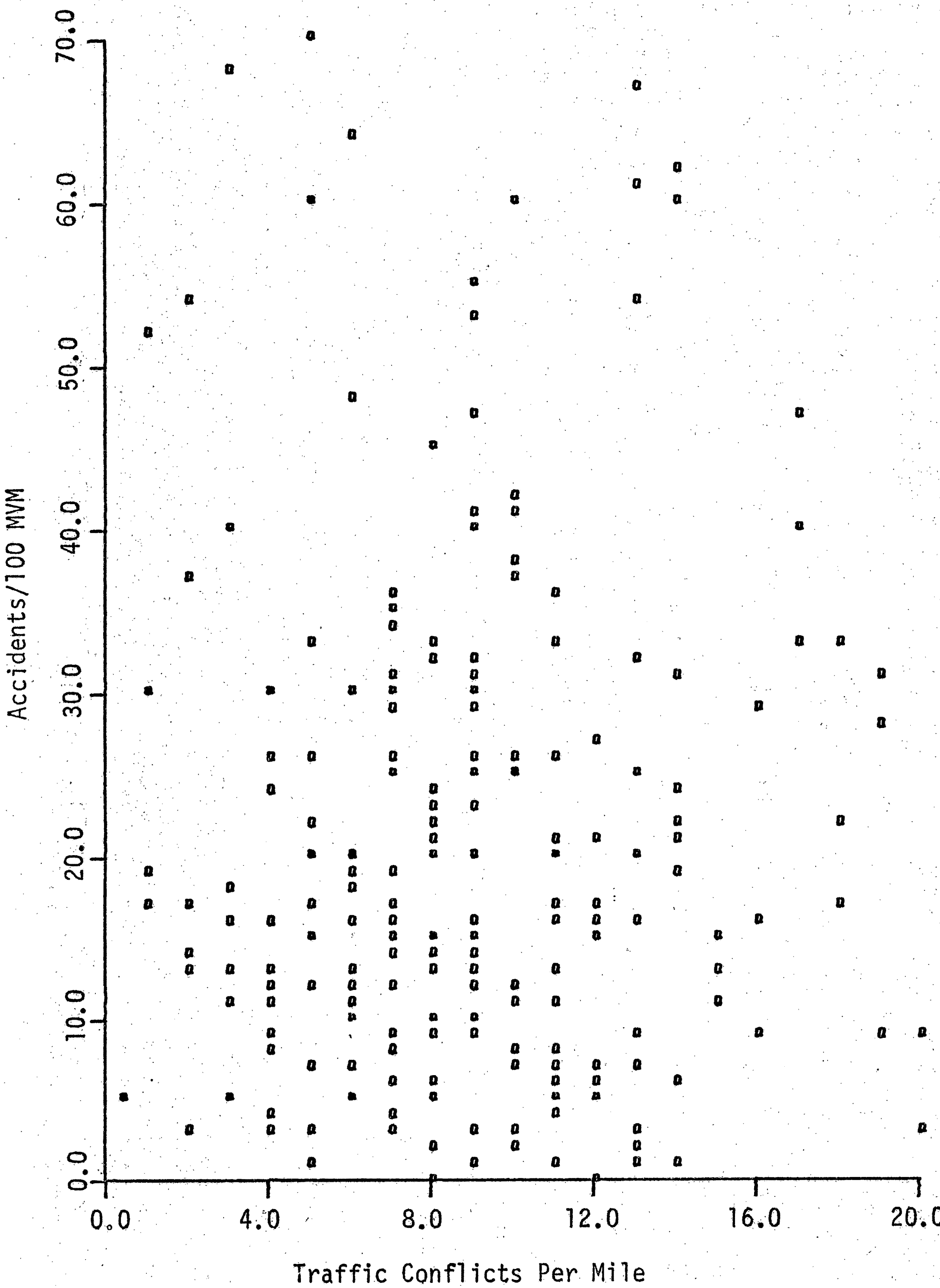


Figure 12. Accidents Rate Versus Traffic Conflicts Per Mile

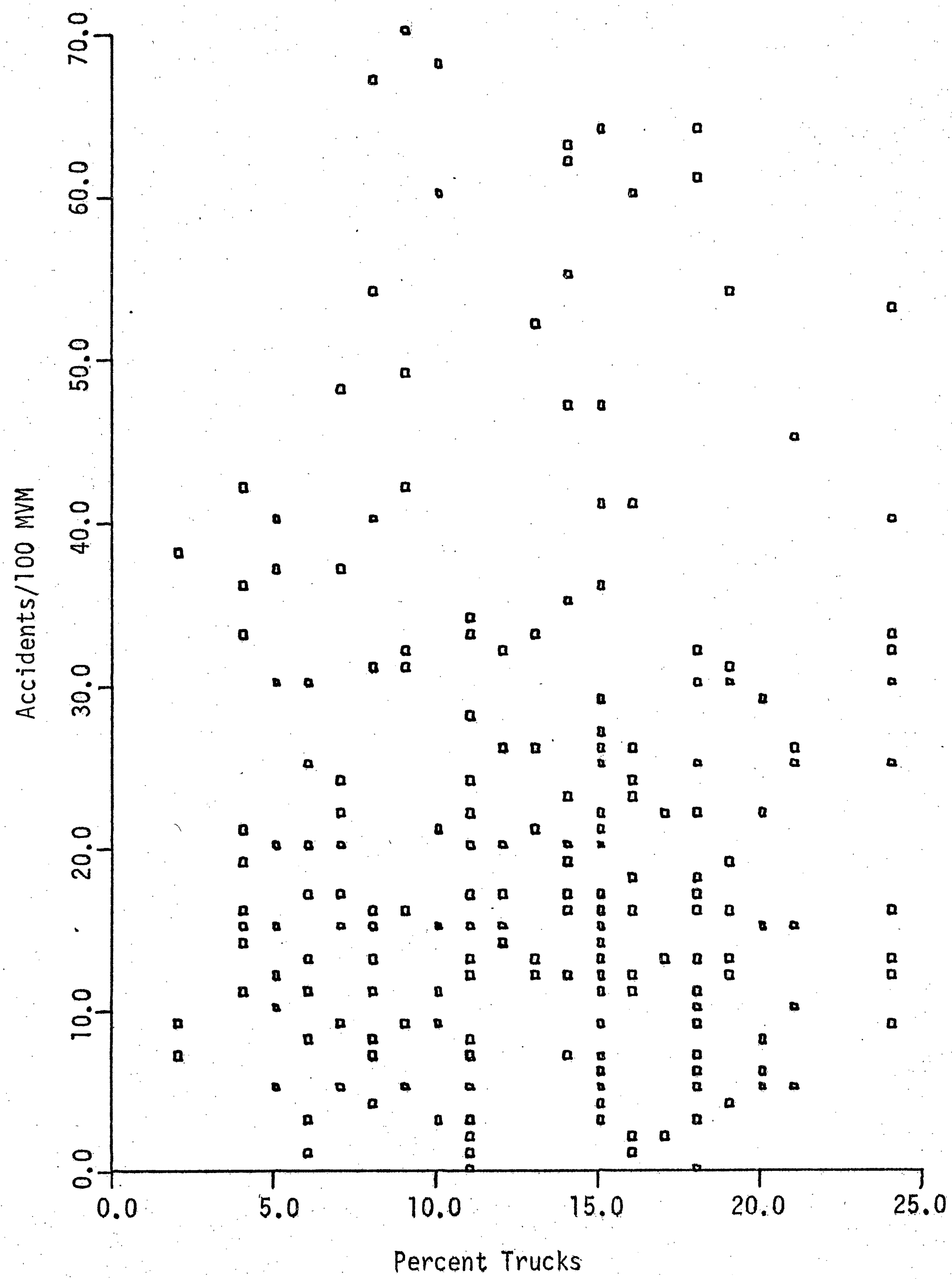


Figure 13. Accident Rate Versus Percent Trucks

Each value of lane width on the data cards was replaced by three zeros and a one according to the specific lane width.

3. First Regression and Correlation Matrix - The first regression calculations (same equation with and without dummy variables) included all the individual independent variables listed in the beginning of this chapter and total accidents as the dependent variable. The first regression revealed that average daily traffic (ADT) contributed considerably more to the R^2 value than did traffic volume ratio. Since the correlation between ADT and traffic volume ratio was high, 0.898, the decision was made to drop traffic volume ratio from the analysis and use ADT as the traffic variable. ADT also showed a higher correlation than traffic volume ratio with total accidents. The R^2 value for the first regression equation, including ADT and excluding traffic volume ratio, and no cross product terms was $R^2 = 0.3195$. Horizontal and vertical alignment were dropped from the analysis as individual variables because of their small contribution to the R^2 value. They were retained and used in the first order interactions or cross product variables. The individual and cross product variables made a total of 53 variables.

4. Calculation of s^2, residual mean square, for various α - Tables (3) through (8) give the s^2 value for each of the 53 variables from step 3. Tables (9) and (10) are s^2 values for α's beyond the limits of $-2.0<\alpha<2.0$ for some of the variables. The minimum s^2 for each variable was then chosen, corresponding to a specific value for α. It should be mentioned here that for some of the variables the s^2 never

Table 3. s^2, Residual Mean Square, 246 Observations

VAR \ α	-2.0	-1.5	-1.0	-0.5	LOG(x)	0.5	1.0	1.5	2.0
*ADT	634.18	621.53	602.90	576.42	543.44	510.90	486.34	471.88	465.54
LAN	653.97	653.23	652.52	651.81	651.13	650.47	649.83	649.22	648.63
SHD	660.23	655.48	648.33	641.09	637.39	640.41	650.66	660.27	664.44
CRS	665.87	665.87	665.87	665.87	665.87	665.87	665.87	665.87	665.87
CON	661.06	662.36	662.55	662.89	663.05	663.14	663.21	663.28	663.34
MRG	655.69	656.88	659.14	661.29	662.89	663.93	664.60	665.02	665.25
TRF	665.08	664.52	663.08	661.06	658.71	656.41	654.37	652.96	652.58
TRK	665.04	664.47	663.69	662.86	662.25	662.06	662.35	663.06	663.99

*For variable abbreviation list see page 93.

Table 4. s^2, Residual Mean Square, 246 Observations

VAR \ α	-2.0	-1.5	-1.0	-0.5	LOG(x)	0.5	1.0	1.5	2.0
ADT-LAN	634.87	622.79	605.32	580.61	543.11	517.51	492.19	476.47	469.33
ADT-SHD	662.30	656.72	640.59	606.15	543.36	528.92	520.48	537.47	565.98
ADT-CRS	634.17	621.51	602.86	576.36	543.40	510.76	441.06	471.69	465.37
ADT-HRZ	634.08	620.15	600.45	673.80	543.40	509.95	665.71	469.32	462.32
ADT-VER	628.59	616.63	599.06	573.74	543.40	509.38	663.33	472.26	467.44
ADT-CON	642.29	631.18	613.12	585.67	543.38	524.16	612.03	525.30	545.01
ADT-MRG	651.17	646.41	638.84	626.26	543.15	587.86	591.82	579.59	587.38
ADT-TRF	641.15	632.29	617.16	591.63	542.89	503.53	466.93	447.60	452.18
ADT-TRK	641.53	633.57	620.80	599.52	543.07	518.65	470.49	438.14	428.07

Table 5. s^2, Residual Mean Square, 246 Observations

VAR \ α	-2.0	-1.5	-1.0	-0.5	LOG(x)	0.5	1.0	1.5	2.0
LAN-SHD	659.24	653.69	645.64	638.05	643.97	638.23	649.37	659.93	664.42
LAN-CRS	654.02	653.27	652.53	651.81	651.01	650.42	665.65	649.12	648.50
LAN-HRZ	662.37	662.32	662.33	662.41	652.01	662.75	662.94	663.34	663.70
LAN-VER	664.59	664.54	664.54	664.57	652.82	664.68	661.93	664.86	664.96
LAN-CON	665.99	666.06	666.11	666.15	654.92	666.18	664.85	666.17	666.16
LAN-MRG	658.22	657.79	658.14	659.93	665.42	664.81	665.77	666.14	666.18
LAN-TRF	664.77	664.18	662.79	660.14	652.36	651.11	646.92	643.48	643.23
LAN-TRK	665.01	664.22	662.92	661.22	659.55	658.33	657.86	658.15	659.03
SHD-CRS	660.22	655.47	648.32	641.07	636.73	640.36	641.50	660.25	664.43

Table 6. s^2, Residual Mean Square, 246 Observations

VAR \ α	-2.0	-1.5	-1.0	-0.5	LOG(x)	0.5	1.0	1.5	2.0
SHD-HRZ	657.54	652.53	646.37	641.13	636.56	642.96	664.97	660.95	664.62
SHD-VER	658.21	653.89	648.57	643.85	637.38	644.39	664.95	660.77	664.51
SHD-CON	657.69	653.75	650.03	647.93	637.23	649.68	664.01	658.19	662.26
SHD-MRG	662.56	664.07	665.50	666.18	665.59	664.09	663.66	662.23	662.49
SHD-TRF	657.15	654.05	649.57	643.69	641.27	633.55	634.81	637.85	644.45
SHD-TRK	661.99	658.78	653.84	647.77	652.62	641.21	647.25	647.25	663.25
CRS-HRZ	661.99	662.02	662.06	662.10	662.20	662.19	659.14	662.32	662.39
CRS-VER	661.27	661.19	661.10	661.01	660.83	660.81	654.40	660.59	660.48
CRS-CON	661.07	661.51	661.92	662.30	662.91	662.94	661.92	663.43	663.62

Table 7. s^2, Residual Mean Square, 246 Observations

VAR \ α	-2.0	-1.5	-1.0	-0.5	LOG(x)	0.5	1.0	1.5	2.0
CRS-MRG	659.59	657.92	656.89	657.72	661.19	662.99	663.65	665.29	665.50
CRS-TRF	665.08	664.83	664.11	662.64	659.74	657.24	662.66	652.63	652.54
CRS-TRK	665.04	664.47	663.69	662.87	662.21	662.06	663.65	663.05	663.98
HRZ-VER	658.29	658.35	658.42	658.50	658.38	658.74	664.29	659.10	659.33
HRZ-CON	659.15	659.68	660.17	660.59	661.60	661.24	662.68	661.63	661.74
HRZ-MRG	661.24	659.34	657.62	657.45	661.20	661.90	656.85	664.64	665.03
HRZ-TRF	665.31	665.11	664.56	663.48	659.89	659.87	663.37	657.48	657.90
HRZ-TRK	665.14	664.70	664.15	663.65	665.93	663.45	663.77	664.35	664.96
VER-CON	656.94	657.43	657.94	658.46	659.81	659.51	663.76	660.53	661.02

Table 8. s^2, Residual Mean Square, 246 Observations

VAR \ α	-2.0	-1.5	-1.0	-0.5	LOG(x)	0.5	1.0	1.5	2.0
VER-MRG	658.24	656.39	655.15	655.91	660.87	662.73	662.06	664.43	664.64
VER-TRF	665.24	665.13	664.72	663.79	659.99	659.89	662.42	655.53	654.87
VER-TRK	664.53	664.20	663.88	663.70	665.19	663.96	659.91	664.92	665.44
CON-MRG	663.03	660.88	658.49	657.33	661.60	660.68	660.95	664.05	664.76
CON-TRF	664.94	664.67	664.26	663.86	659.63	663.78	665.76	664.26	664.46
CON-TRK	665.87	665.72	665.49	665.20	663.80	664.76	666.13	665.10	666.53
MRG-TRF	666.15	666.18	666.00	665.75	665.70	666.14	666.12	666.16	666.18
MRG-TRK	666.08	666.16	666.14	665.78	665.98	666.05	666.18	665.82	665.43
TRF-TRK	663.91	663.20	661.76	659.14	652.88	650.45	654.11	640.40	636.87

Table 9. s^2, Residual Mean Square, 246 Observations

VAR \ α	-4.0	-3.5	-3.0	-2.5	2.5	3.0	3.5	4.0
LAN	**	**	**	**	648.07	647.54	647.63	646.55
CON	662.28	662.09	661.74	660.59	**	**	**	**
ADT-LAN	**	**	**	**	468.19	470.54	478.93	478.93
ADT-HRZ	**	**	**	**	461.33	463.72	467.56	471.75
LAN-CRS	**	**	**	**	647.92	647.36	646.83	646.32
SHD-MRG	658.07	658.99	660.04	661.21	663.20	664.04	**	**
CRS-HRZ	661.88	661.90	661.93	661.96	**	**	**	**

Table 10. s^2, Residual Mean Square, 246 Observations

VAR \ α	-4.0	-3.5	-3.0	-2.5	2.5	3.0	3.5	4.0
CRS-VER	**	**	**	**	660.37	660.25	660.14	660.03
CRS-CON	659.16	659.64	660.13	660.61	**	**	**	**
HRZ-VER	658.19	658.20	658.22	658.25	**	**	**	**
HRZ-CON	656.72	657.35	657.97	658.54	**	**	**	**
VER-CON	655.29	655.66	656.05	656.48	**	**	**	**
MRG-TRK	**	**	**	**	665.22	665.21	665.32	665.48

reached a lowest value but rather decreased at a slow rate up to α=4.0 or α=-4.0. Since the change in s^2 was small a value of 4.0 or -4.0 was used for α instead of extending it further. Figure (14) is an example of a plot for s^2 versus α if a more accurate value of α is desired. For the demonstration purposes in this study, s^2 values were chosen at intervals of 0.5 and proved to be sufficient.

5. Scatter Diagrams for Transformed Individual Variables - As shown in Figures (15) through (21), the transformation of the variables has produced a change in the x,Y relationship for some of the variables and not for others, such as lane width. The plot of lane width versus total accidents has the same distinct grouping of points regardless of the transformation used; therefore it is desirous to change lane width to a dummy variable.

6. Grouping of the Data - The decision to group the data came after re-examination of the first scatter diagrams, especially Figures (8) and (9) for horizontal and vertical alignment, respectively. Not clearly shown on these plots, is the fact that about half of the observations show little or no horizontal and vertical alignment. The column of points on the plot just to the right of the Y axis for both horizontal and vertical alignment actually represent a combination of over 100 observations. Therefore two groups of data were analyzed. One for relatively straight, flat roadway section (0-5% horizontal and vertical alignment greater than 3 degrees curvature and 3% grade, respectively) and the other group for roadway sections with curves and hills (>5%). The steps for this analysis are found in Section 3.

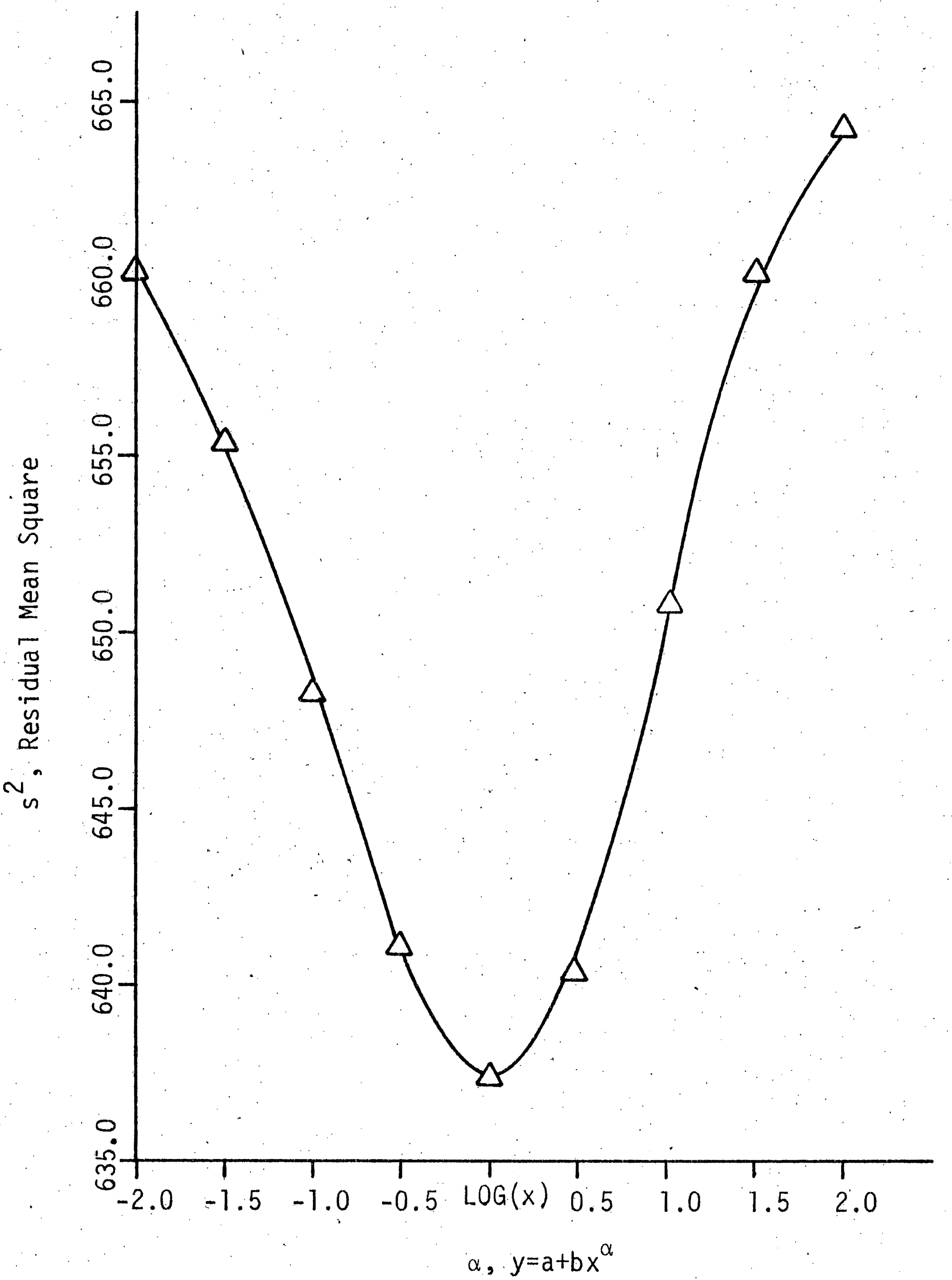


Figure 14, s^2 Versus α, Shoulder Width

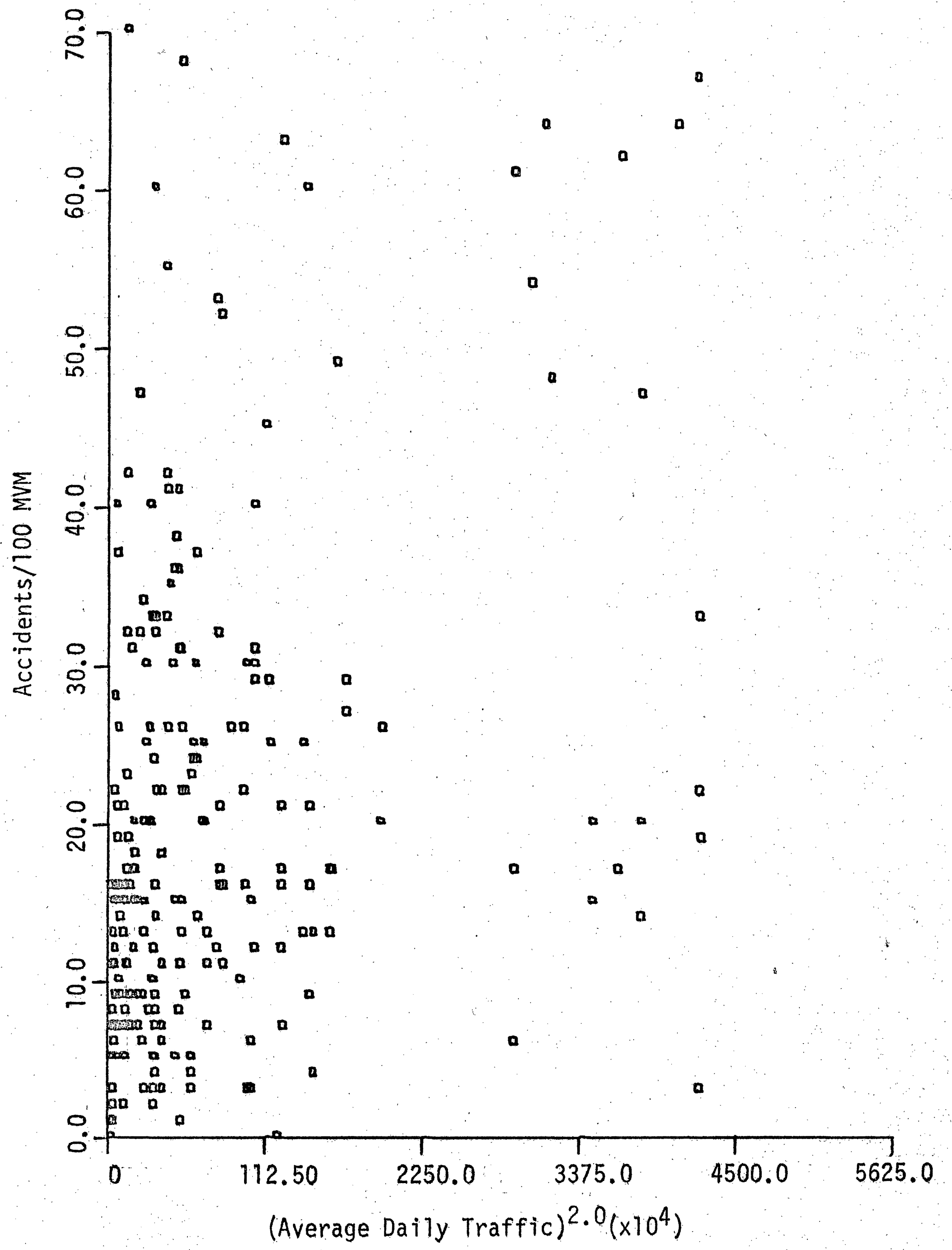


Figure 15. Accident Rate Versus (Average Daily Traffic)$^{2.0}$

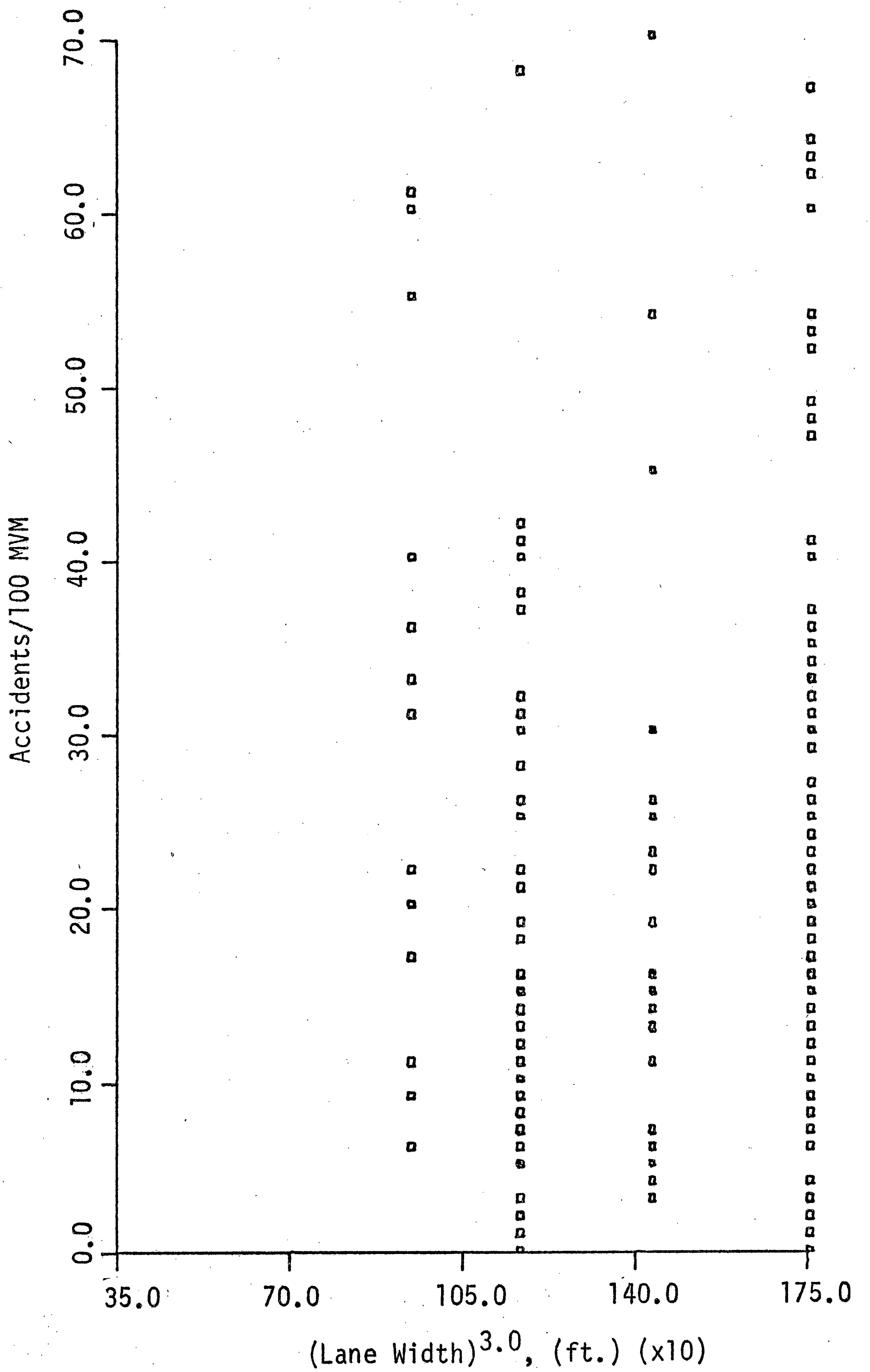


Figure 16. Accident Rate versus (Lane Width)$^{3.0}$

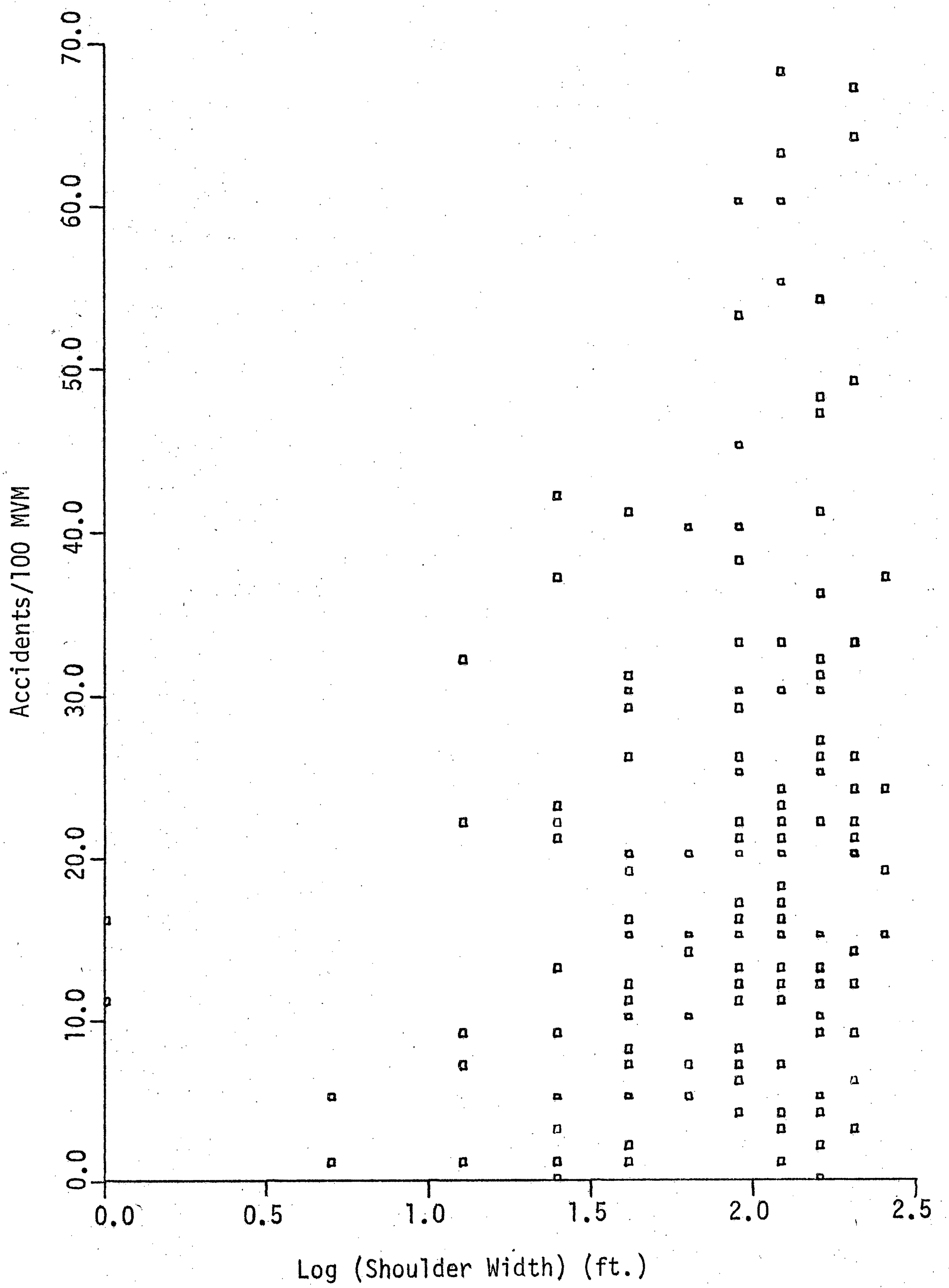


Figure 17. Accident Rate Versus Log (Shoulder Width)

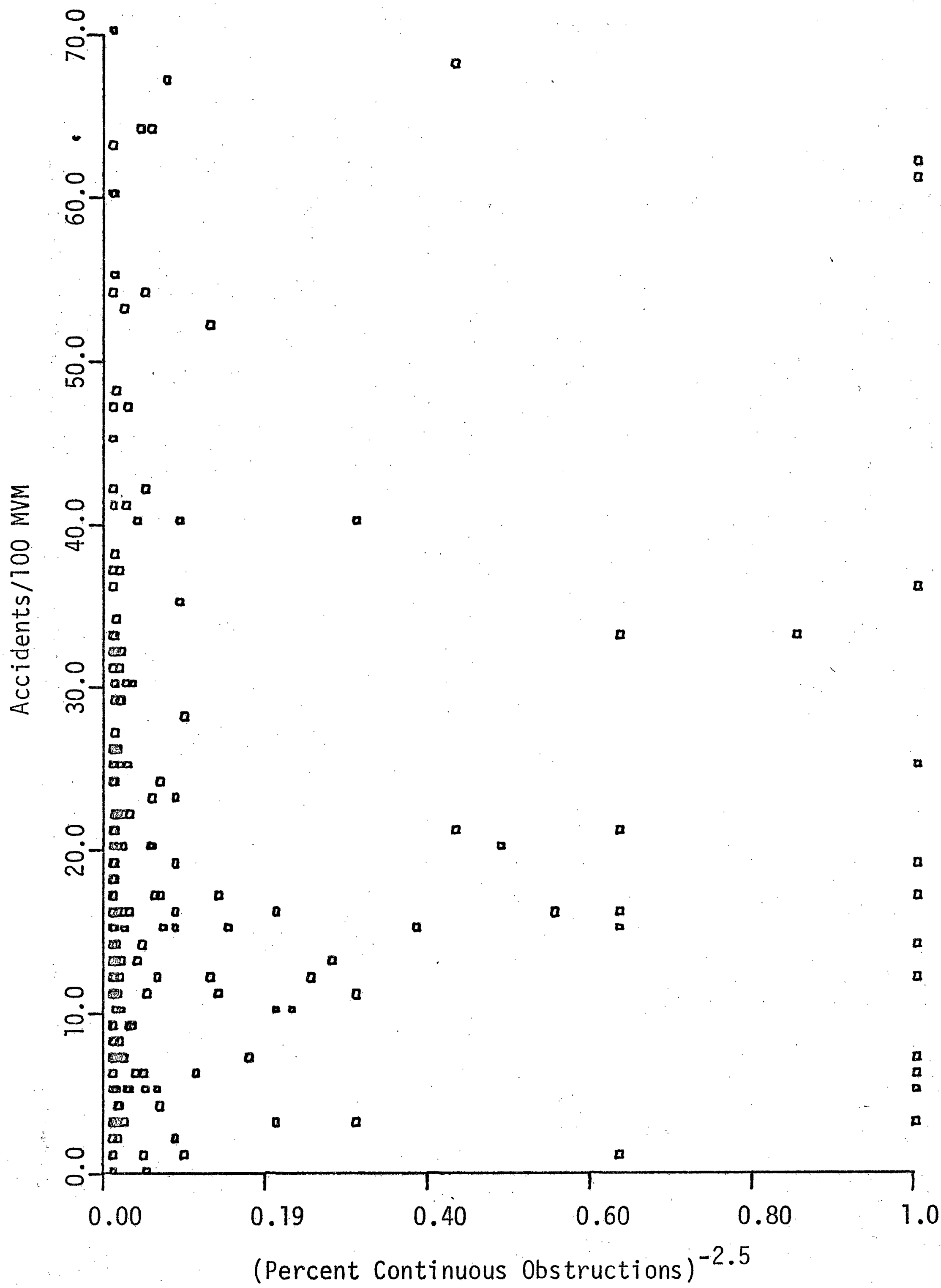


Figure 18. Accident Rate Versus (Percent Continuous Obstructions)$^{-2.5}$

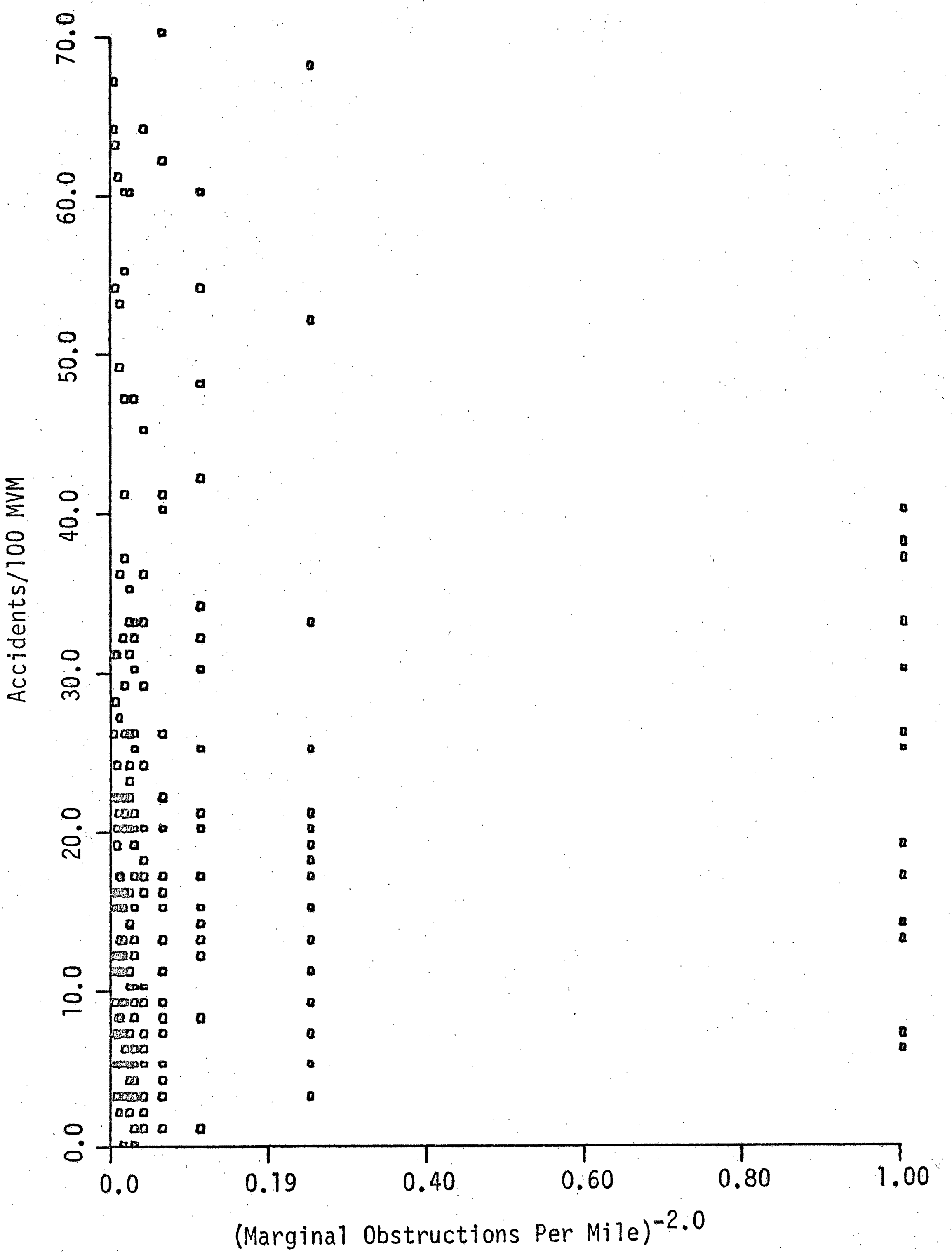


Figure 19. Accident Rate versus (Marginal Obstructions Per Mile)$^{-2.0}$

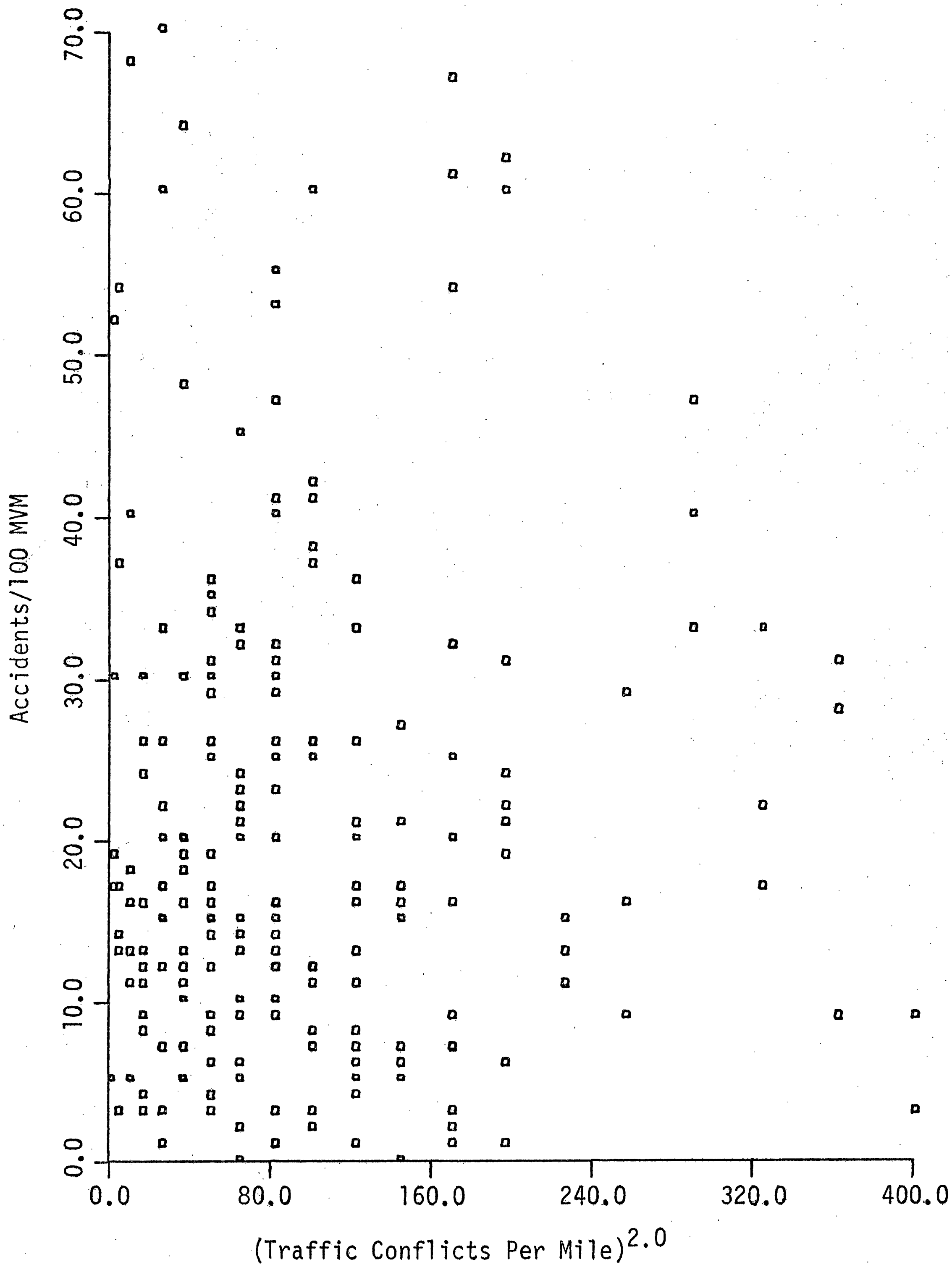


Figure 20. Accident Rate Versus (Traffic Conflicts Per Mile)$^{2.0}$

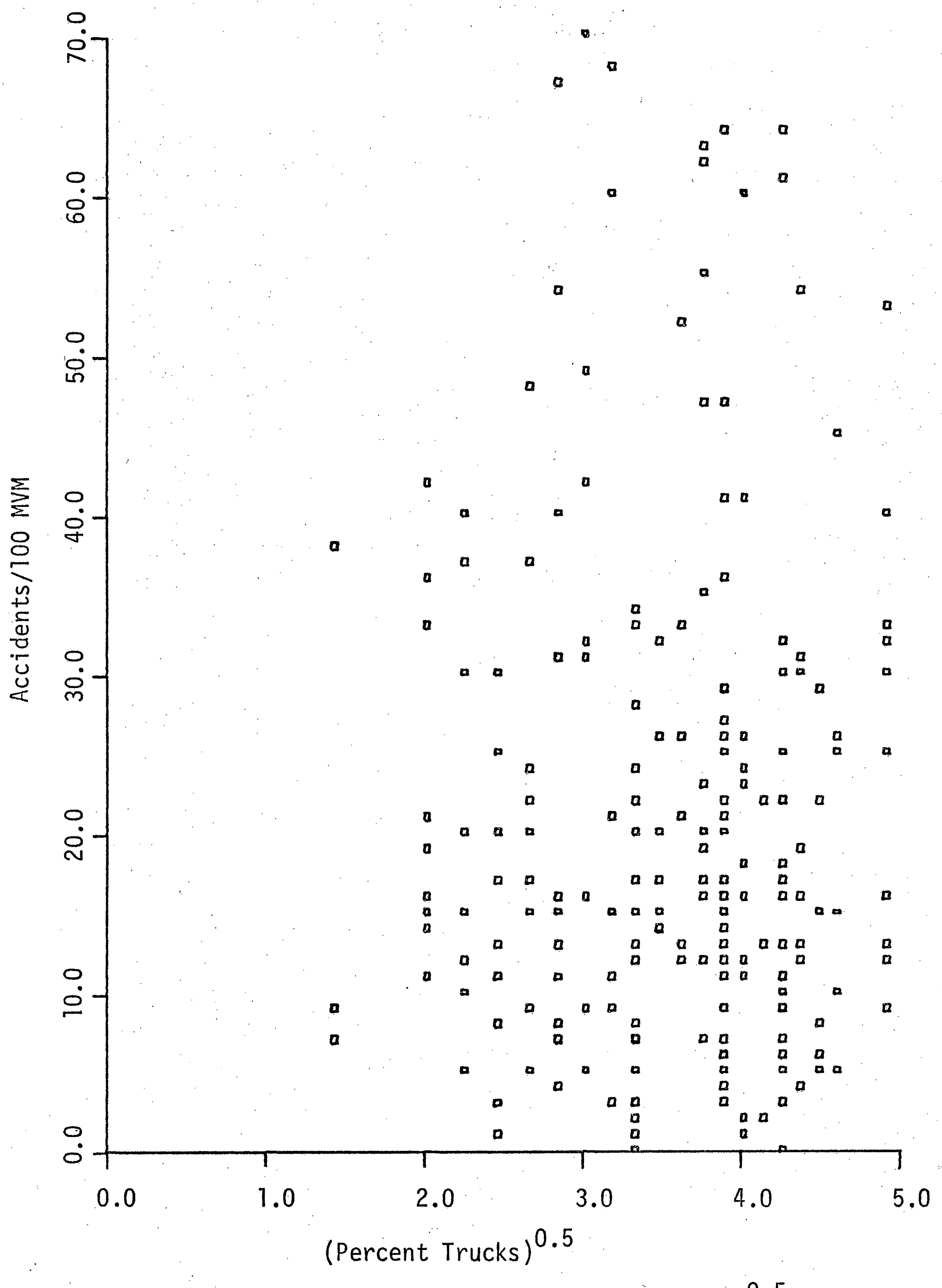


Figure 21. Accident Rate Versus (Percent Trucks)$^{0.5}$

7. Choose Transformations - Table 11 is a list of the variables and their respective transformations, α, and variable (x) values as they were used for the final regression calculations. The same transformations were used for the equations with and without the dummy variable of lane width. The abbreviated names of the variables are as follows:

1. Average Daily Traffic - ADT
2. Lane Width - LAN
3. Shoulder Width - SHD
4. Cross Slope - CRS
5. Horizontal Alignment - HRZ
6. Vertical Alignment - VER
7. Continuous Obstructions - CON
8. Marginal Obstructions - MRG
9. Traffic Conflicts - TRF
10. Trucks - TRK
11. Total Accidents - TOTACC

8. Final Regression Calculations and Results - The variables and coefficients for the regression equation found for the total data set without the dummy variable are listed as follows:

$R^2 = .5251$ F-ratios all significant at the 0.05 level

Y = total accidents per 100 million vehicle miles

Table 11. Regression Variables and Transformations

Variable	α	Variable	α	Variable	α	Variable	α
ADT(x_1) *	2.0	ADT-CON(x_1x_8) *	1.0	SHD-HRZ(x_4x_6) *	log	HRZ-CON(x_6x_8) *	-4.0
LAN(x_3) *	3.0	ADT-MRG(x_1x_9) *	log	SHD-VER(x_4x_7) *	log	HRZ-MRG(x_6x_9) *	-0.5
SHD(x_4) *	log	ADT-TRF(x_1x_{10}) *	1.5	SHD-CON(x_4x_8) *	log	HRZ-TRF(x_6x_{10}) *	0.5
CRS(x_5) *	1.0	ADT-TRK(x_1x_{11}) *	2.0	SHD-MRG(x_4x_9) *	-4.0	HRZ-TRK(x_6x_{11}) *	0.5
CON(x_8) *	-2.5	LAN-SHD(x_3x_4) *	-0.5	SHD-TRF(x_4x_{10}) *	0.5	VER-CON(x_7x_8) *	-4.0
MRG(x_9) *	-2.0	LAN-CRS(x_3x_5) *	4.0	SHD-TRK(x_4x_{11}) *	0.5	VER-MRG(x_7x_9) *	-1.0
TRF(x_{10}) *	2.0	LAN-HRZ(x_3x_6) *	log	CRS-HRZ(x_5x_6) *	-4.0	VER-TRF(x_7x_{10}) *	2.0
TRK(x_{11}) *	0.5	LAN-VER(x_3x_7) *	log	CRS-VER(x_5x_7) *	4.0	VER-TRK(x_7x_{11}) *	-0.5
ADT-LAN(x_1x_3) *	2.5	LAN-CON(x_3x_8) *	log	CRS-CON(x_5x_8) *	-4.0	CON-MRG(x_8x_9) *	-0.5
ADT-SHD(x_1x_4) *	1.0	LAN-MRG(x_3x_9) *	-1.5	CRS-MRG(x_5x_9) *	-1.0	CON-TRF(x_8x_{10}) *	log
ADT-CRS(x_1x_5) *	1.0	LAN-TRF(x_3x_{10}) *	2.0	CRS-TRF(x_5x_{10}) *	2.0	CON-TRK(x_8x_{11}) *	log
ADT-HRZ(x_1x_6) *	2.5	LAN-TRK(x_3x_{11}) *	1.0	CRS-TRK(x_5x_{11}) *	0.5	MRG-TRF(x_9x_{10}) *	log
ADT-VER(x_1x_7) *	2.0	SHD-CRS(x_4x_5) *	log	HRZ-VER(x_6x_7) *	-3.0	MRG-TRK(x_9x_{11}) *	3.0
						TRF-TRK($x_{10}x_{11}$) *	2.5

SOURCE		B-VALUES
Mean		4.70458413
ADT	(x_1)	-0.00000214
ADT-CON	(x_1x_8)	0.46320312
ADT-TRF	(x_1x_{10})	0.00000088
ADT-TRK	(x_1x_{11})	0.00004018
LAN-TRF	(x_3x_{10})	-0.00012934
LAN-TRK	(x_3x_{11})	-5.73362821
SHD-MRG	(x_4x_9)	5439.79670409

The B-values for x_1, x_1x_{10} and x_1x_{11} are very small but this is due to the large numbers generated by the various transformations on the variables. One disadvantage is that some computer packages for regression analysis do not make calculations and print results for such small numbers. The B-value for the mean in the above equation represents those accidents that would occur or be unexplained by the variables used if all the independent variables would have a zero value, providing this could be physically possible. For the variables in this study all but two could have a zero value; these two are lane width and average daily traffic. Therefore the above regression equation fitted to the data would have its Y-intercept at 4.70458413 accidents per 100 MVM.

Although the variables in the regression equation explain 52.51% of the variation in accident rates there is still about 48% of the variation in accident rates explained by the driver, the vehicle or some other variable not included from the roadway/environment element.

Listed below are the preceding variables in the order of their importance in contributing to the R^2 value and also in their importance to the pre-crash phase of accident occurrence.

1. Average daily traffic and trucks
2. Average daily traffic and traffic conflicts
3. Shoulder width and marginal obstructions
4. Average daily traffic
5. Average daily traffic and continuous obstrctions
6. Lane width and traffic conflicts
7. Lane width and trucks

It appears that average daily traffic is an important variable when combined with the effects of trucks and traffic conflicts (driveways and intersections), also the lane and shoulder width have a significant effect when combined with trucks and traffic conflicts as well as marginal obstructions. The cross product variables of lane width combined with traffic conflicts and trucks, respectively show a reducing effect on the accident rate, i.e., as the variable values increase the accident rate will tend to decrease. The effects of increasing lane width are more obvious than the effects of increased traffic conflicts and presence of trucks in traffic.

It is interesting to note, according to the regression coefficients, that accidents will tend to decrease as traffic volume increases. At first glance this seems illogical, however most of the data is for rural two-lane roads where approximately 30.0% of the fatalities occurring on rural roads in the U.S. were single vehicle run-off-the-road type accidents. Therefore the increased presence of traffic on rural roads apparently effects the speed and driving behavior of drivers thus reducing accident potential to some degree.

It is useful to compare the results of the Dart and Mann (4) study which used the same data but with a less formalized regression approach. Their final regression equation contained individual and cross product variables with none of the individual squared variables appearing in the equation. Traffic volume ratio was used as the traffic variable instead of average daily traffic as was used in this study. The following is a list of the variables, coefficients and R^2 value found by Dart and Mann (4):

$R^2 = 0.46$ F-ratios significant at the 0.05 level

Y = total accidents per 100 million vehicle miles

SOURCE	B-VALUES
Mean	41.32
Trucks	-1.23
Traffic Volume ratio - (TVR)	-0.54
Cross Slope	-0.67
(TVR) (Trucks)	0.03
(TVR) (Cross Slope)	0.03
(TVR) (Horizontal Alignment)	-0.0009
(TVR) (Traffic Conflicts)	0.026
(Lane Width) (Traffic Conflicts)	-0.12
(Shoulder Width) (Horizontal Alignment)	0.009

Some of the differences between the two equations are apparent since different variables were used for traffic volume and also because of the degree of transformation used. Dart and Mann (4) used squared terms and cross products of the individual variables. The individual variables of trucks and cross slope did not appear in the regression equation for this study. The reasons are not quite obvious and perhaps may be due to the degree of transformation as mentioned previously.

Section 2 (Use of Dummy Variables)

Since the only change in the analysis from Section 1 was the conversion of lane width to a dummy variable, only the final results will be given for this analysis. The same cross products and transformations were used for the dummy variable analysis as with the first analysis in Section 1. The purpose of this section was to determine the influence of each dummy variable class on the dependent variable.

The variables and coefficients for the regression equation using lane width as a dummy variable are given below:

$R^2 = .5291$ F-ratios all significant at 0.05 level

Y = total accidents per 100 million vehicle miles

SOURCE	B-VALUES
Mean	4.17975227
ADT (x_1)	-0.00000251
LANE 9 (x_A)	9.25332156
LANE 10 (x_B)	-1.80426626
LANE 11 (x_C)	-4.58737133
ADT-CON (x_1x_8)	0.44241526
ADT-TRF (x_1x_{10})	0.00000061

-cont.

-cont-

SOURCE	B-VALUES
ADT-TRK (x_1x_{11})	0.00004112
LAN-TRK (x_3x_{11})	-4.69445710
SHD-MRG (x_4x_9)	5654.56340020

The dummy variable classes x_A, x_B, x_C for 9ft., 10ft., and 11ft respectively were introduced into the first equation using the same cross products and transformations; only the individual variable of lane width was replaced. In comparing the equation in Section 1 with the preceding equation it is seen that the same variables appear in both with little variation in the regression coefficients. The R^2 values were also very nearly the same with the difference being explained by the presence of the lane width dummy variables in the preceding equation.

Since lane width is a discrete variable, as shown in the scatter diagrams, it was felt that the equation using lane width as a dummy variable would perhaps be the more logical predictor for accident rates and also lead to better regression fits. One of the advantages of using dummy variables, as stated in Chapter Four, is that estimates can be made of the contribution of all classes of a given dummy variable set. The B-values for the dummy variable classes in the preceding equation measures the net effect that each lane width has on the dependent variable or accident rate when compared to the excluded class for 12 ft lane widths. Therefore the analyst can interpret the lane width coefficients as follows. Roads with 9ft lane widths will have 9.25 (rounding off) more accidents per 100 MVM occurring on them than for roads with 12ft lanes. Those roads having 10ft lane widths will have 1.80 less accidents per 100 MVM occurring on them than for roads with 12 ft lanes. Similarly, 11ft lane roads will have 4.59 less accidents per 100 MVM than for 12 ft lane roads. It may be interesting to note that 6 of the 8 four-lane roadway sections had lane widths of 12 ft.

The analyst can also make estimates of the influence of each dummy variable class on the accident rate with respect to the overall mean. This can be accomplished by using equation (4.1) from the previous chapter which calulates partial regression coefficients for each class. This is done for lane width as follows:

$$B_j = b_j - \sum_{j=1}^{c} P_j b_j$$

where: B_j = Adjusted regression coefficient of the dummy variable class. This indicates the influence of the dummy variable class on the dependent variable with respect to the overall mean,

b_j = B-value for the dummy variable class,

P_j = ratio of the number of observations in dummy variable class j to the total number of observations in the data set, and

c = number of dummy variable classes.

Lane Width (ft)	# obs	P_j	b_j	$P_j b_j$	B_j
9	16	0.0650	9.25	0.60	9.83
10	82	0.3333	-1.80	-0.60	-1.22
11	31	0.1260	-4.59	-0.58	-4.01
12	117	0.4756	0	0	0.58
Total	246			-0.58	

From the above table the roadway sections having 9ft lane widths can be interpreted as having 9.83 more accidents per 100 MVM than the average lane width. There will be 1.22 less accidents per 100 MVM occurring on roads with 10ft lane width than for the average. Likewise 11ft lane roads will have 4.01 less accidents per 100 MVM than for the average and 12ft lane roads will have 0.58 more accidents per 100 MVM than the average. This indicates that roadways with 11ft lane widths will have lower accident rates than the other lane widths. The only problem with this interpretation is the fact that an average lane width may not physically exist. Although the scatter diagrams show four distinct groupings of points, it is not known if the lane widths were rounded to a whole number when the data was collected.

Section 3 (Data Group for Straight, Flat Roadway Sections)

As discussed previously the total data set was grouped according to straight, flat roads and curved, hilly roads. The R^2 value found for the first equation - for straight, flat roads was 0.4547; the R^2 value of the equation for curved, hilly roads was 0.1369. The difference between the two initial values supports the hypothesis that the roadway/environment variables give a better estimation of accident occurrence on straight, flat roads due to driver actions for these types of roads. It is theorized that since drivers require less information to operate their vehicles on straight, flat roads, the majority of drivers will drive in a similar manner. Therefore the roadway/environment variables as they interact with driver variables produce a higher R^2 value in explaining accident rate variation for straight, flat roads.

For application purposes the roadway sections for straight, flat roadways were further analyzed. The same steps were followed as in Section 1.

1. Scatter Diagrams - Figures (22) through (29) give the plots of accident rates versus the individual independent variables. Horizontal and vertical alignment were not included in this analysis simply because there was little or no alignment values in the data chosen. There were 131 observations for the straight, flat road group.

2. Dummy Variable Check - Since dummy variables were used in Section 2 they were not used for this analysis. None of the other variables appeared as likely candidates for dummy variable conversion.

3. First Regression and Correlation Matrix - The R^2 value, as mentioned above, was 0.4547 for the first regression calculations without transformations or cross product terms. The correlations between the dependent and independent variables was improved over that for the total data group.

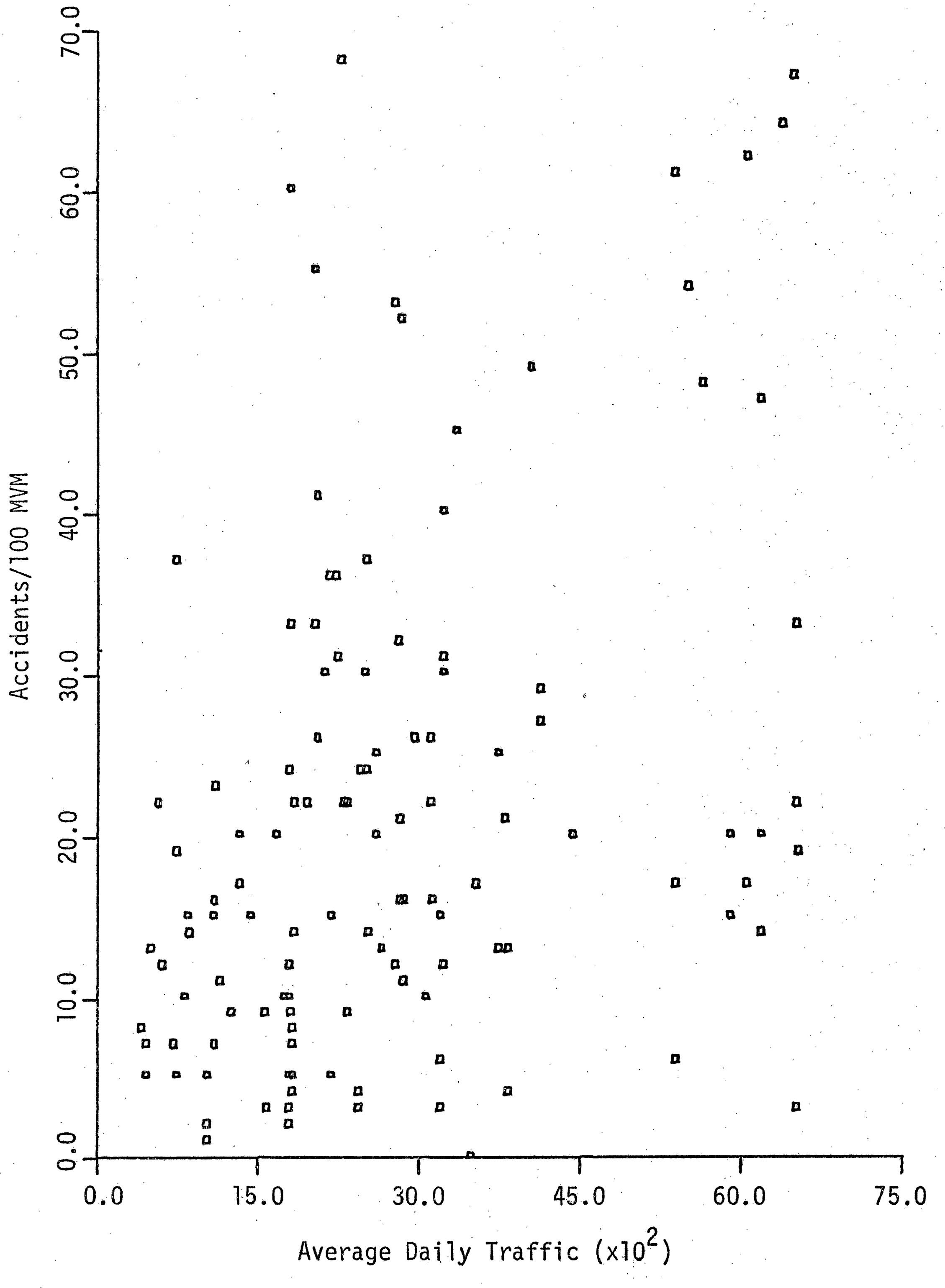


Figure 22. Accident Rate Versus Average Daily Traffic

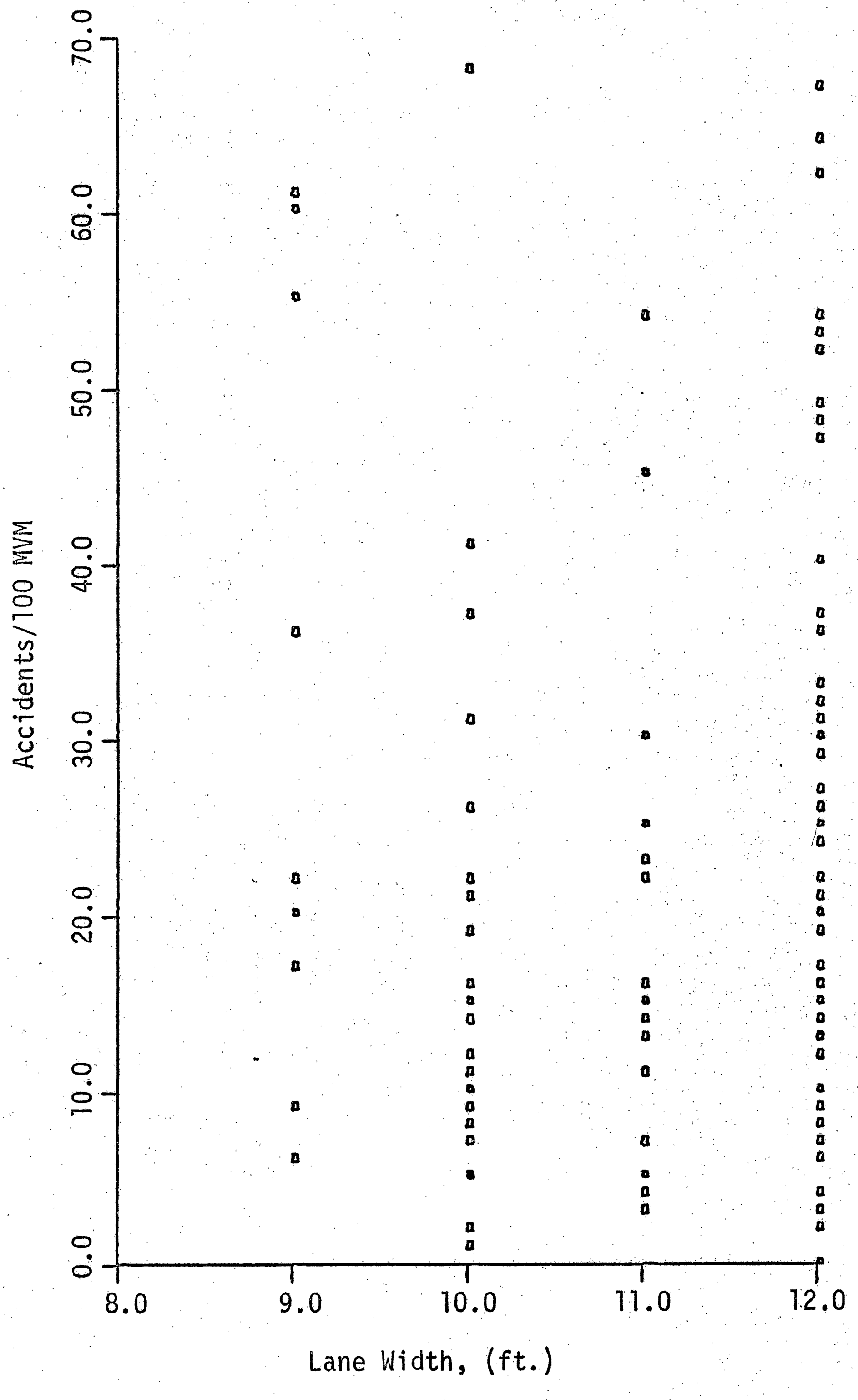


Figure 23. Accident Rate Versus Lane Width

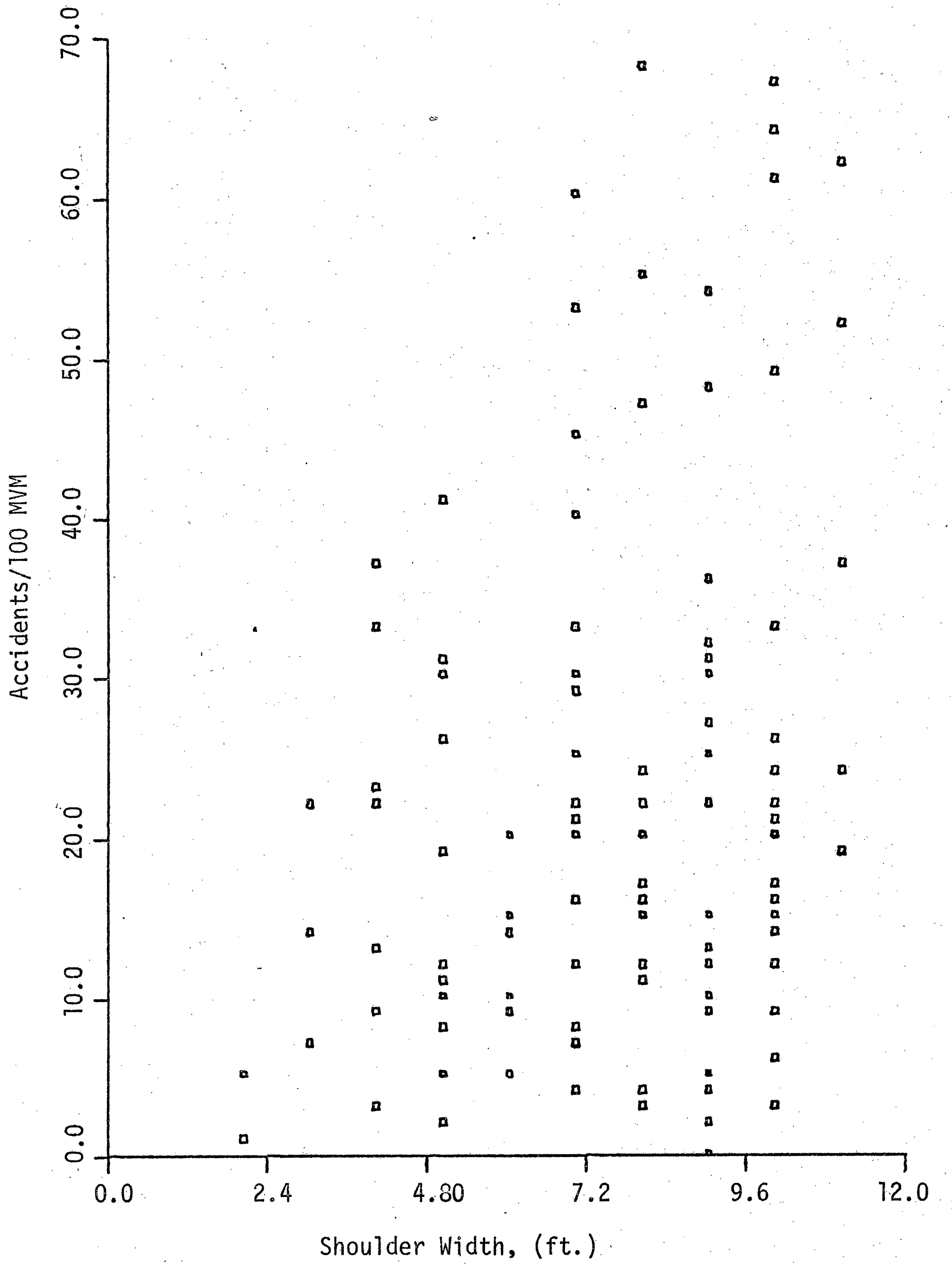


Figure 24. Accident Rate Versus Shoulder Width

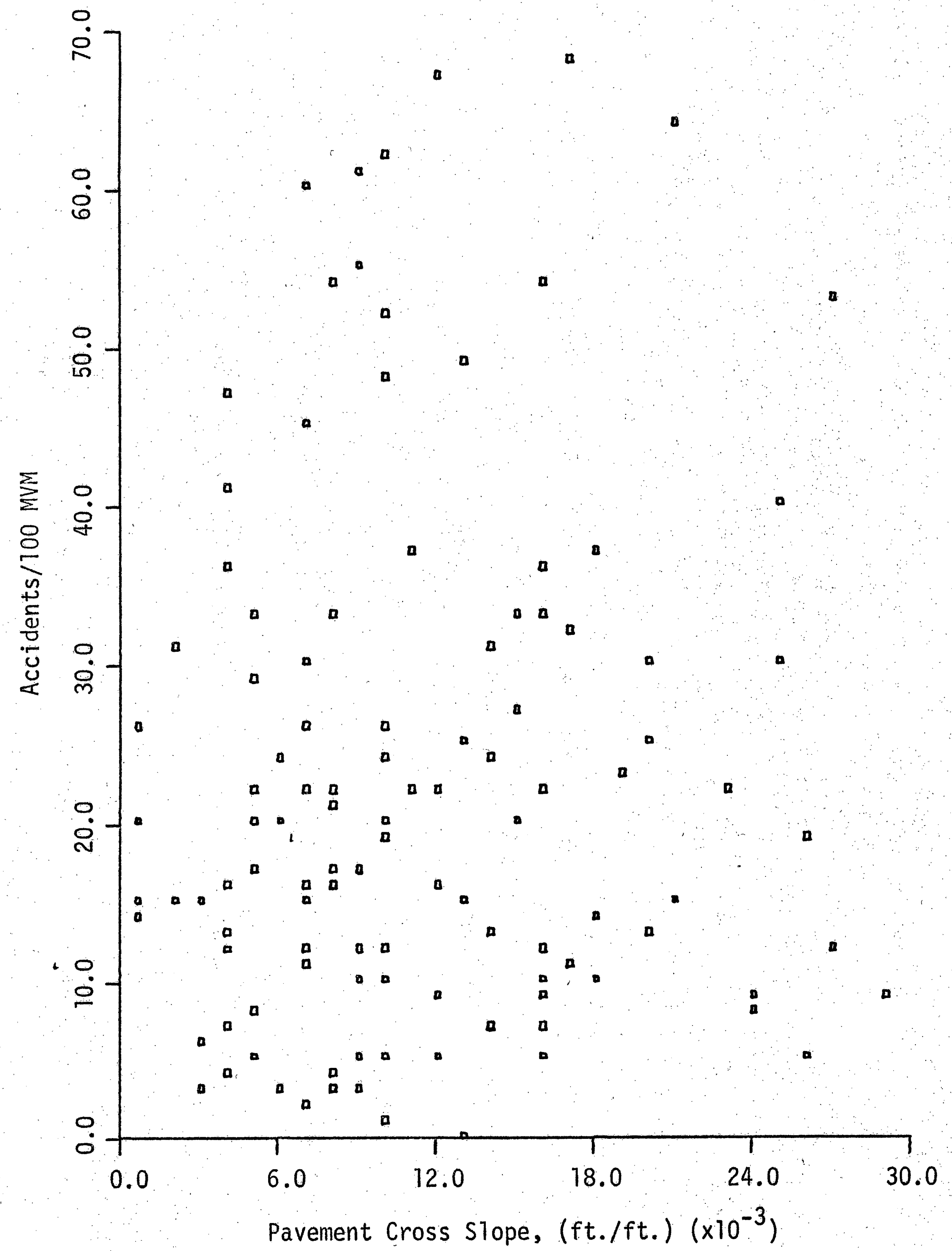


Figure 25. Accident Rate Versus Pavement Cross Slope

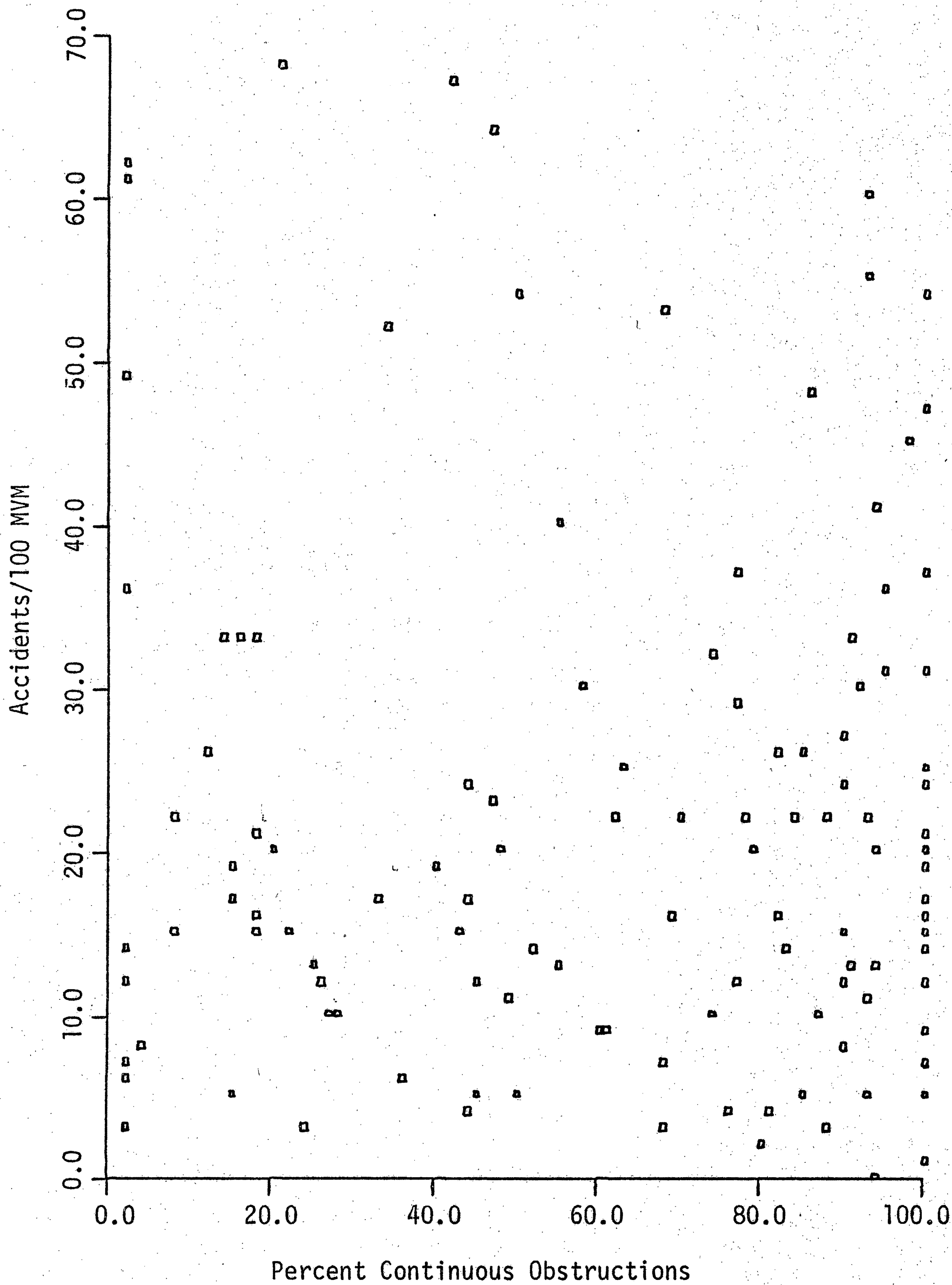


Figure 26. Accident Rate Versus Percent Continuous Obstructions

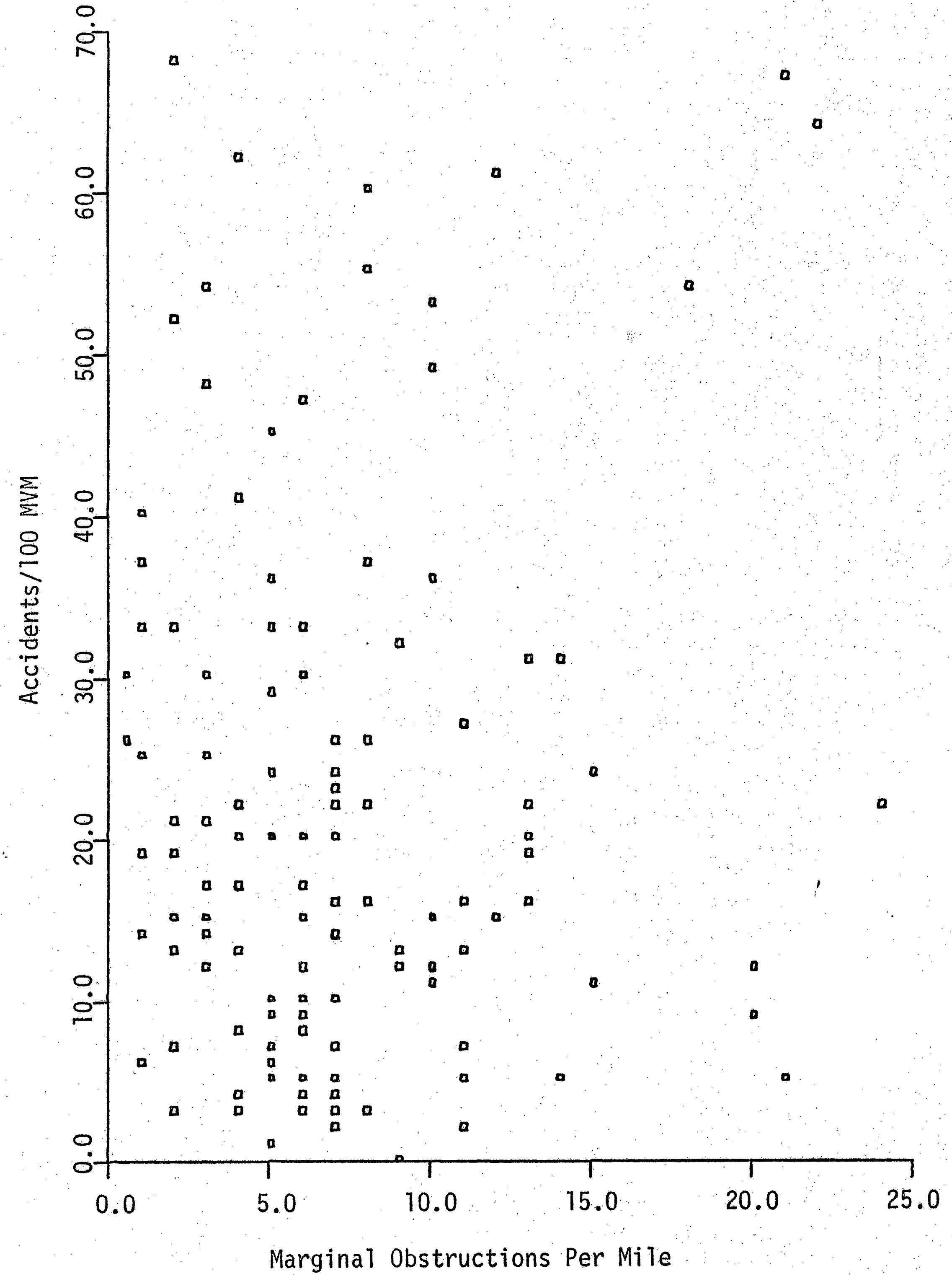


Figure 27. Accident Rate Versus Marginal Obstructions Per Mile

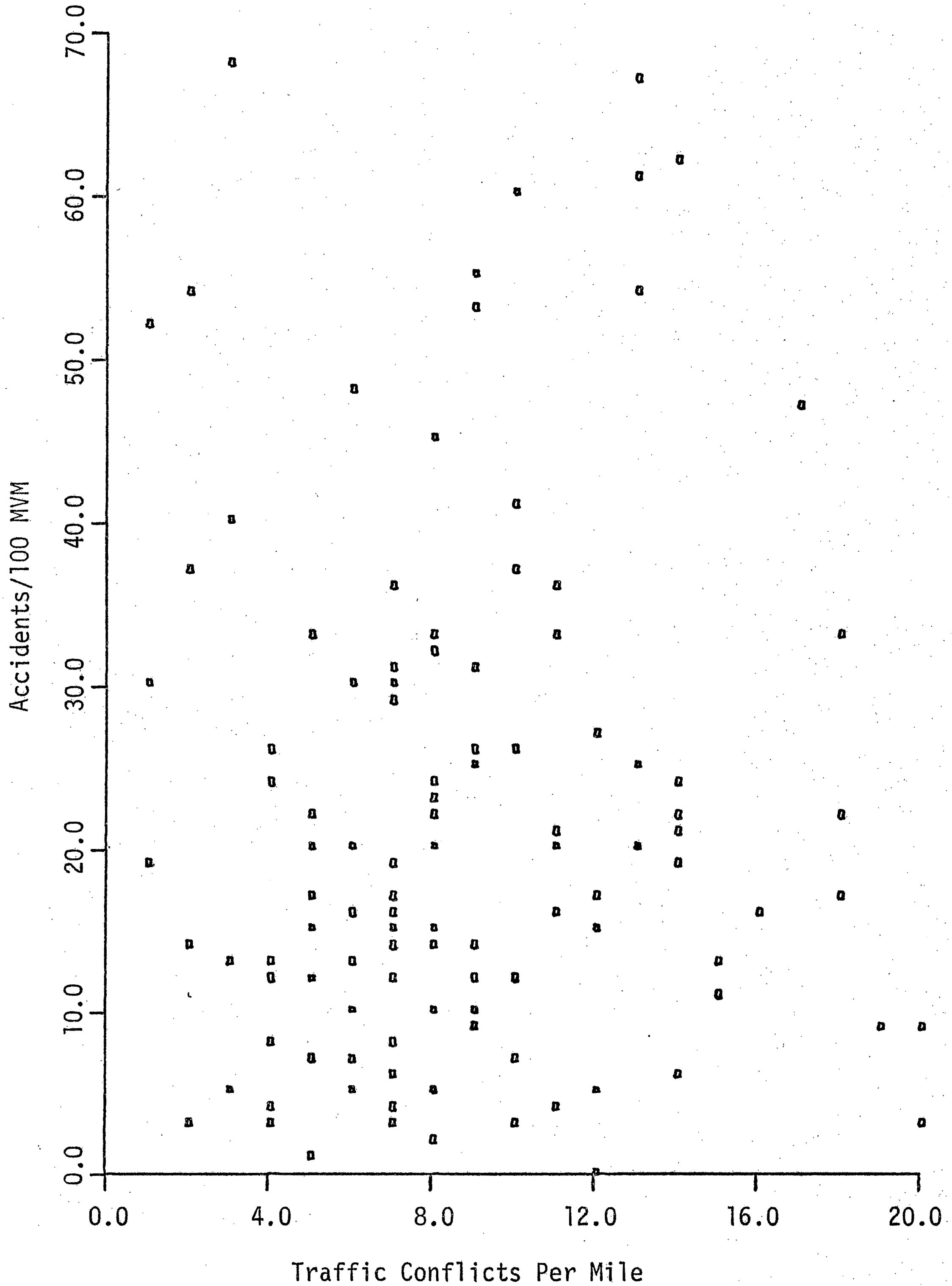


Figure 28. Accident Rate Versus Traffic Conflicts Per Mile

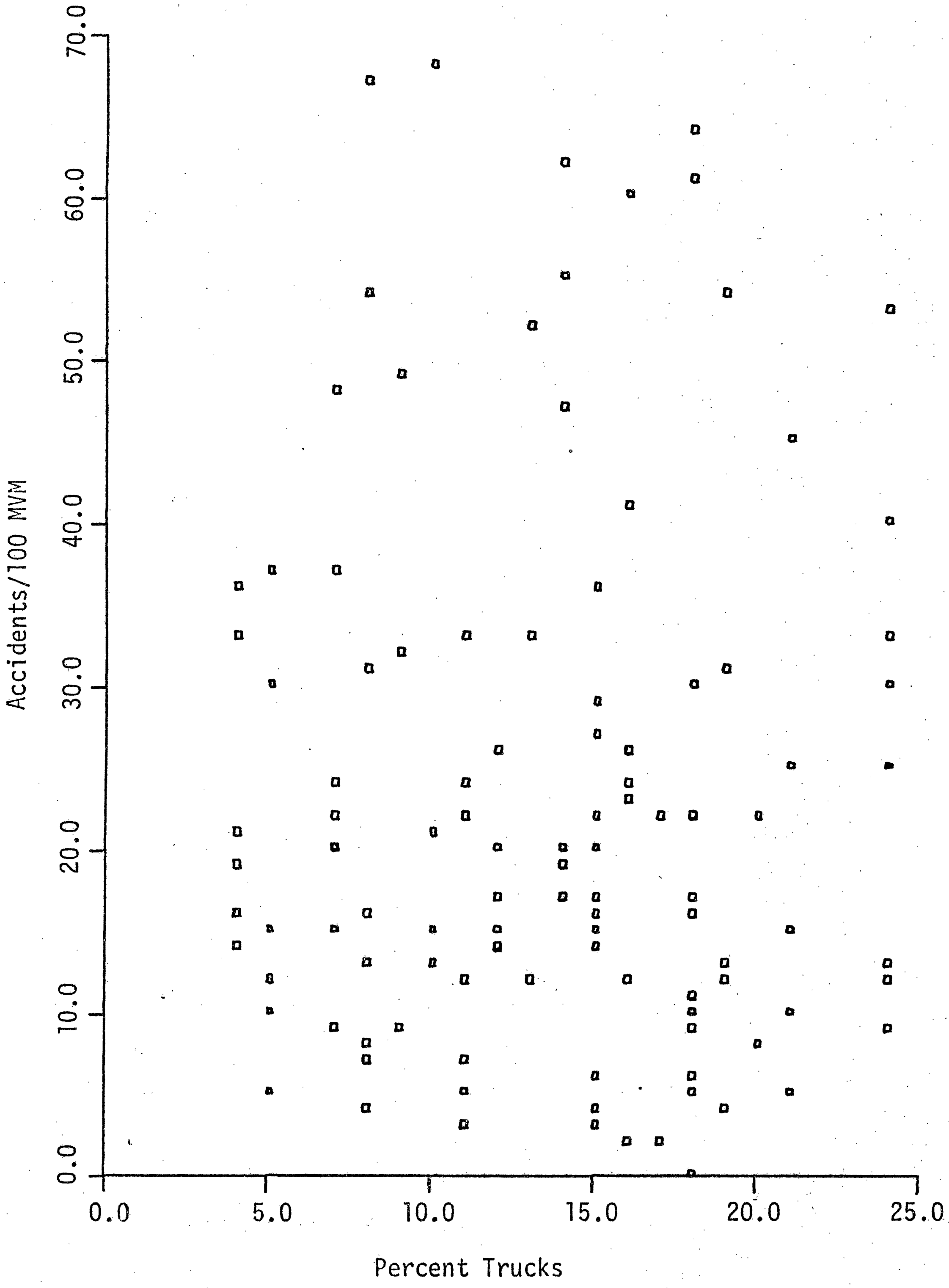


Figure 29. Accident Rate Versus Percent Trucks

4. Calculation of s^2, residual mean square, for various α - Tables (12) through (18) give the s^2 values for each of the individual variables and also for the cross product terms of these variables. Tables (19) through (21) are the s^2 values for those α's beyond the limits of $-2.0<\alpha<2.0$.

5. Scatter Diagrams for Transformed Individual Variables - Figures (30) through (37) show the relationship between the accident rate and the transformed individual variables.

6. Choose Transformations - Table 22 gives the individual and cross product variables and the transformations used in the final regression calculations. The same variable name abbreviations were used as specified in step 7 of Section 1.

7. Final Regression Calculations and Results - Following are the variables and regression coefficients that were found as accident contributors from the roadway/environment element for the straight, flat road sections.

$R^2 = 0.6474$ F-ratios all significant at 0.05 level

Y = total accidents per 100 million vehicle miles

SOURCE	B-VALUES
Mean	2174.12825899
CRS (x_5)	-2195.36683671
ADT-MRG (x_1x_9)	0.00000062
ADT-TRF (x_1x_{10})	0.00000069
ADT-TRK (x_1x_{11})	0.00000110

-cont-

Table 12, s^2, Residual Mean Square, 131 Observations

VAR \ α	-2.0	-1.5	-1.0	-0.5	LOG(x)	0.5	1.0	1.5	2.0
*ADT	827.93	810.80	781.07	733.64	670.49	604.65	550.29	513.58	494.02
LAN	853.24	852.61	851.99	851.39	850.80	850.24	849.70	849.18	848.68
SHD	838.74	834.11	830.20	828.60	831.69	841.73	854.40	861.70	864.12
CRS	853.16	853.19	853.20	853.21	853.21	853.21	853.21	853.22	853.22
HRZ	860.63	860.67	860.69	860.70	860.70	860.70	860.71	860.71	860.71
VER	854.68	854.51	854.43	854.43	854.40	854.38	854.37	854.36	854.35
CON	848.41	850.93	851.39	852.24	852.70	852.97	853.19	853.40	853.61
MRG	862.51	862.90	863.12	863.99	864.64	864.88	864.88	864.80	864.75
TRF	864.89	863.32	858.14	850.25	840.39	829.27	816.10	800.16	781.63

*For variable abbreviation list see page 93.

Table 13, s^2, Residual Mean Square, 131 Observations

VAR \ α	-2.0	-1.5	-1.0	-0.5	LOG(x)	0.5	1.0	1.5	2.0
TRK	862.51	861.16	858.78	856.65	856.51	857.90	859.88	861.89	863.52
ADT-LAN	830.85	815.66	788.80	744.29	669.96	614.46	557.00	517.99	497.75
ADT-SHD	842.62	832.25	810.94	768.22	670.33	645.53	632.31	658.89	702.91
ADT-CRS	827.90	810.75	780.99	733.49	670.43	604.33	450.33	513.19	493.67
ADT-HRZ	827.67	810.15	779.93	732.08	670.43	603.32	864.74	512.72	493.34
ADT-VER	827.80	810.63	780.79	733.09	670.43	603.27	818.98	512.37	493.15
ADT-CON	843.16	827.07	795.99	744.38	670.38	619.48	785.34	587.77	611.95
ADT-MRG	859.07	852.49	837.35	804.96	670.12	708.90	717.96	704.35	722.19

Table 14, s^2, Residual Mean Square, 131 Observations

VAR \ α	-2.0	-1.5	-1.0	-0.5	LOG(x)	0.5	1.0	1.5	2.0
ADT-TRF	854.64	846.89	827.70	783.40	670.15	594.38	510.38	475.34	481.52
ADT-TRK	863.18	861.92	854.39	811.58	670.49	626.16	544.46	492.17	472.14
LAN-SHD	838.47	832.99	828.26	826.16	835.62	839.36	853.14	861.36	864.06
LAN-CRS	853.12	852.47	851.82	851.19	850.68	849.98	842.98	848.86	848.33
LAN-HRZ	854.24	853.62	853.01	852.41	850.77	851.26	861.18	850.19	849.69
LAN-VER	852.23	851.46	850.69	849.93	850.56	848.43	852.60	846.99	846.30
LAN-CON	860.57	860.56	860.61	860.71	857.70	861.05	855.59	861.45	861.65
LAN-MRG	864.44	864.14	863.98	864.28	864.65	864.76	863.90	862.30	861.16
LAN-TRF	864.90	864.27	861.41	854.43	826.98	821.72	795.59	760.89	721.81

Table 15, s^2, Residual Mean Square, 131 Observations

VAR \ α	-2.0	-1.5	-1.0	-0.5	LOG(x)	0.5	1.0	1.5	2.0
LAN-TRK	862.23	860.54	857.49	854.39	850.12	853.77	855.28	857.18	859.11
SHD-CRS	838.74	834.09	830.15	828.51	832.84	841.62	790.94	861.68	864.11
SHD-HRZ	839.20	834.59	830.68	829.07	832.91	842.34	863.68	861.90	864.18
SHD-VER	838.65	833.89	829.82	828.01	832.77	841.06	842.08	861.55	864.07
SHD-CON	846.59	847.30	848.56	850.40	836.74	856.05	842.72	862.40	863.97
SHD-MRG	864.65	864.23	863.44	862.04	860.33	856.30	854.21	849.22	847.10
SHD-TRF	854.42	850.83	844.41	834.17	816.74	806.76	793.30	784.42	776.89
SHD-TRK	862.52	860.63	854.91	843.86	831.46	838.52	849.23	859.15	863.34
CRS-HRZ	861.52	861.52	861.53	861.53	861.54	861.53	859.65	861.54	861.54

Table 16, s^2, Residual Mean Square, 131 Observations

VAR \ α	-2.0	-1.5	-1.0	-0.5	LOG(x)	0.5	1.0	1.5	2.0
CRS-VER	852.67	852.64	852.62	852.59	852.51	852.54	863.95	852.48	852.46
CRS-CON	848.57	849.35	850.14	850.93	852.37	852.50	862.81	853.97	854.65
CRS-MRG	864.23	863.73	863.31	863.44	864.34	864.78	863.58	864.73	864.67
CRS-TRF	864.89	864.55	862.54	857.38	845.37	833.97	804.57	794.47	769.95
CRS-TRK	862.50	861.16	858.77	865.63	859.13	857.86	841.61	861.86	863.50
HRZ-VER	864.83	864.83	864.83	864.83	864.84	864.84	864.07	864.84	864.84
HRE-CON	848.10	848.88	849.67	850.47	852.05	852.04	861.20	853.49	854.16
HRZ-MRG	864.20	863.69	863.23	863.35	864.33	864.75	861.78	864.73	864.75
HRZ-TRF	864.89	864.57	862.62	857.59	845.40	835.12	860.16	798.16	775.71

Table 17, s^2, Residual Mean Square, 131 Observations

VAR \ α	-2.0	-1.5	-1.0	-0.5	LOG(x)	0.5	1.0	1.5	2.0
HRZ-TRF	864.89	864.57	862.62	857.59	845.40	835.12	860.16	798.16	775.71
HRZ-TRK	862.48	861.13	858.81	856.81	861.23	858.09	864.32	861.91	863.49
VER-CON	849.32	850.05	850.80	851.54	852.78	852.98	863.84	854.33	854.95
VER-MRG	864.22	863.73	863.32	863.47	864.34	864.82	836.72	864.65	864.57
VER-TRF	864.89	864.53	862.47	847.20	845.33	833.61	859.87	794.35	770.13
VER-TRK	862.50	861.15	858.73	856.52	857.75	857.62	850.67	861.61	863.31
CON-MRG	864.76	864.25	863.07	861.68	864.26	861.64	855.09	862.88	862.99

Table 18, s^2, Residual Mean Square, 131 Observations

VAR \ α	-2.0	-1.5	-1.0	-0.5	LOG(x)	0.5	1.0	1.5	2.0
CON-TRF	864.67	864.09	862.92	861.23	844.67	858.76	854.49	859.74	860.85
CON-TRK	863.29	863.07	862.90	862.99	854.06	863.73	864.48	864.62	864.88
MRG-TRF	864.58	864.50	864.68	864.80	857.45	860.59	862.73	864.05	864.84
MRG-TRK	864.16	864.25	864.31	864.15	864.23	861.53	859.21	856.19	854.61
TRF-TRK	863.56	863.20	861.25	853.06	847.26	815.99	790.40	768.73	743.05

Table 19, s^2, Residual Mean Square, 131 Observations

VAR \ α	-4.0	-3.5	-3.0	-2.5	2.5	3.0	3.5	4.0
ADT	* *	* *	* *	* *	487.68	489.69	495.92	503.67
LAN	* *	* *	* *	* *	848.22	847.78	847.36	846.98
CRS	853.19	853.19	853.18	853.15	* *	* *	* *	* *
HRZ	860.67	860.67	860.65	860.62	* *	* *	* *	* *
CON	850.73	850.32	849.56	847.64	* *	* *	* *	* *
MRG	863.35	863.66	864.06	864.54	* *	* *	* *	* *
TRF	* *	* *	* *	* *	738.37	709.56	681.53	655.47
ADT-LAN	* *	* *	* *	* *	492.21	495.88	504.09	513.73
ADT-HRZ	* *	* *	* *	* *	487.17	489.33	495.69	503.52

Table 20, s^2, Residual Mean Square, 131 Observations

VAR \ α	-4.0	-3.5	-3.0	-2.5	2.5	3.0	3.5	4.0
ADT-VER	* *	* *	* *	* *	486.91	488.60	494.17	501.01
ADT-TRK	* *	* *	* *	* *	474.32	487.52	504.67	522.32
LAN-HRZ	* *	* *	* *	* *	849.20	848.75	848.32	847.91
LAN-VER	* *	* *	* *	* *	845.62	844.96	844.32	843.70
LAN-MRG	* *	* *	* *	* *	860.69	860.88	861.47	862.23
LAN-TRF	* *	* *	* *	* *	680.26	639.80	603.64	573.68
SHD-MRG	* *	* *	* *	* *	846.17	846.12	846.17	846.12
SHD-TRF	* *	* *	* *	* *	771.36	767.37	764.51	762.47

Table 21, s^2, Residual Mean Square, 131 Observations

VAR \ α	-4.0	-3.5	-3.0	-2.5	2.5	3.0	3.5	4.0
SHD-TRF	* *	* *	* *	* *	771.36	767.37	764.51	762.47
CRS-VER	* *	* *	* *	* *	852.43	852.40	852.37	852.34
CRS-CON	845.78	846.42	847.10	847.82	* *	* *	* *	* *
HRZ-CON	845.29	845.94	846.62	847.35	* *	* *	* *	* *
HRZ-TRF	* *	* *	* *	* *	751.65	727.04	702.97	680.34
VER-CON	846.57	847.22	847.89	848.59	* *	* *	* *	* *
VER-TRF	* *	* *	* *	* *	743.82	716.54	689.54	663.97
MRG-TRK	* *	* *	* *	* *	854.28	854.90	856.04	857.33
TRF-TRK	* *	* *	* *	* *	717.69	693.75	671.96	652.57

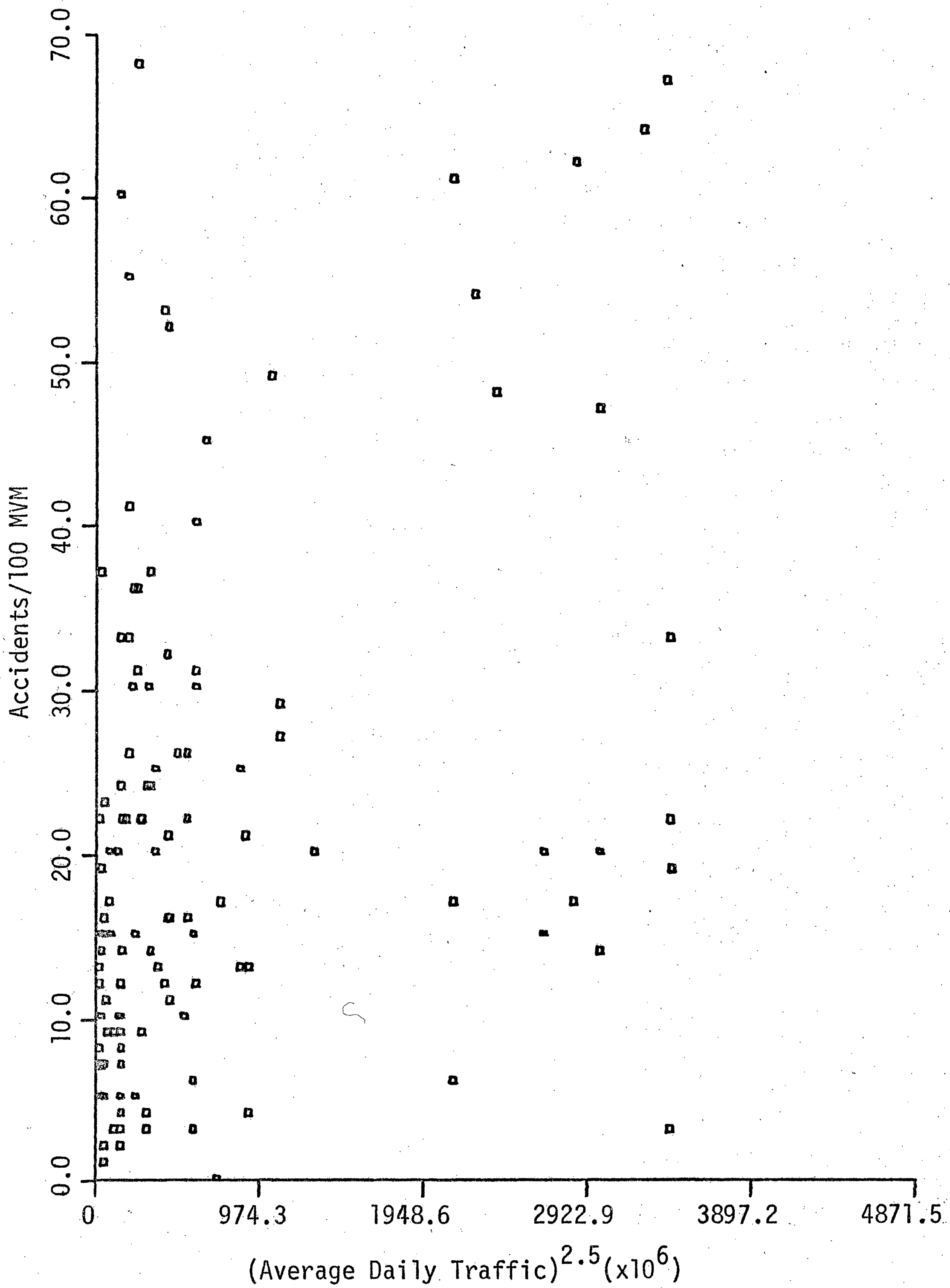


Figure 30. Accident Rate Versus (Average Daily Traffic)$^{2.5}$

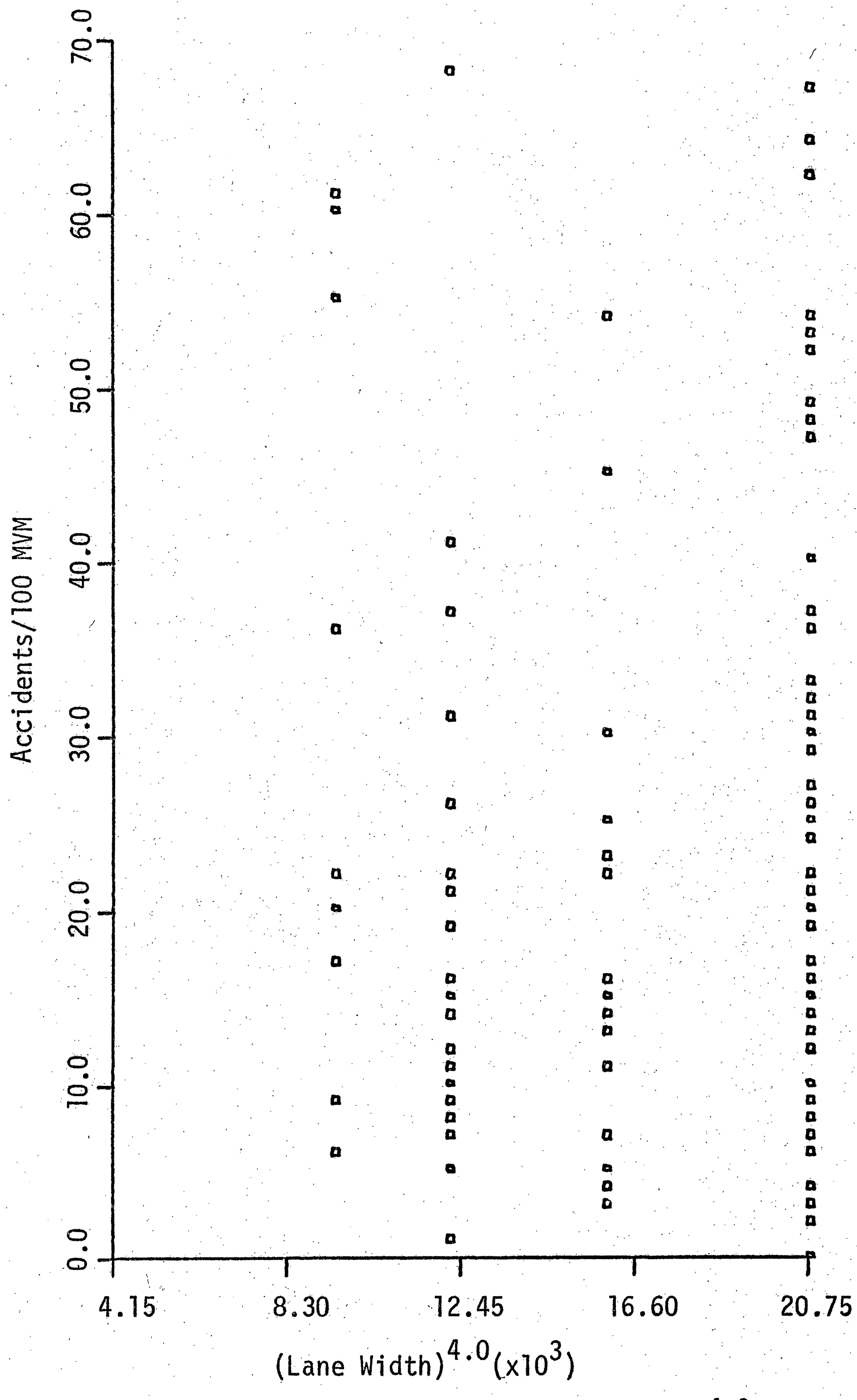


Figure 31. Accident Rate Versus (Lane Width)$^{4.0}$

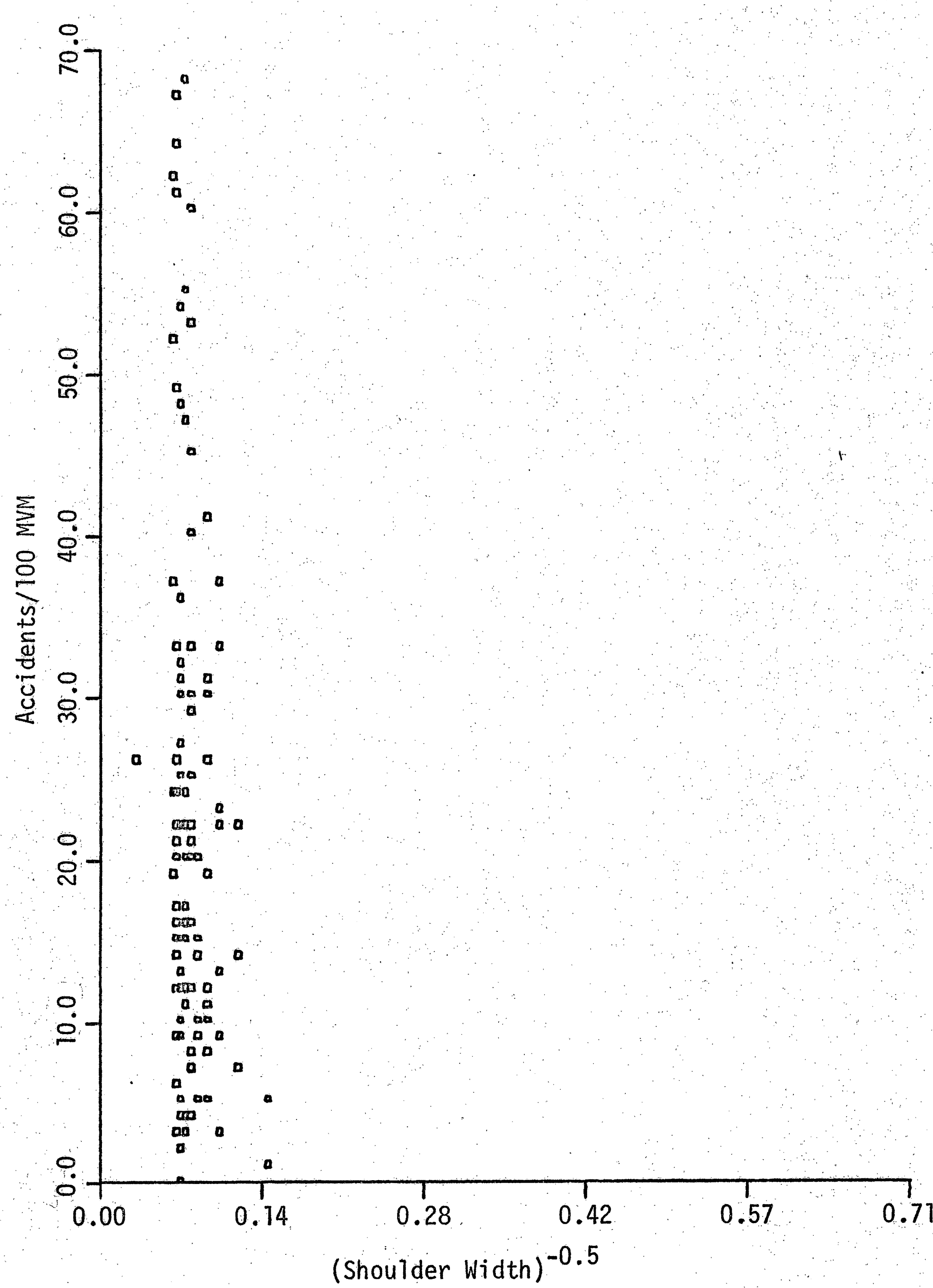


Figure 32. Accident Rate Versus (Shoulder Width)$^{-0.5}$

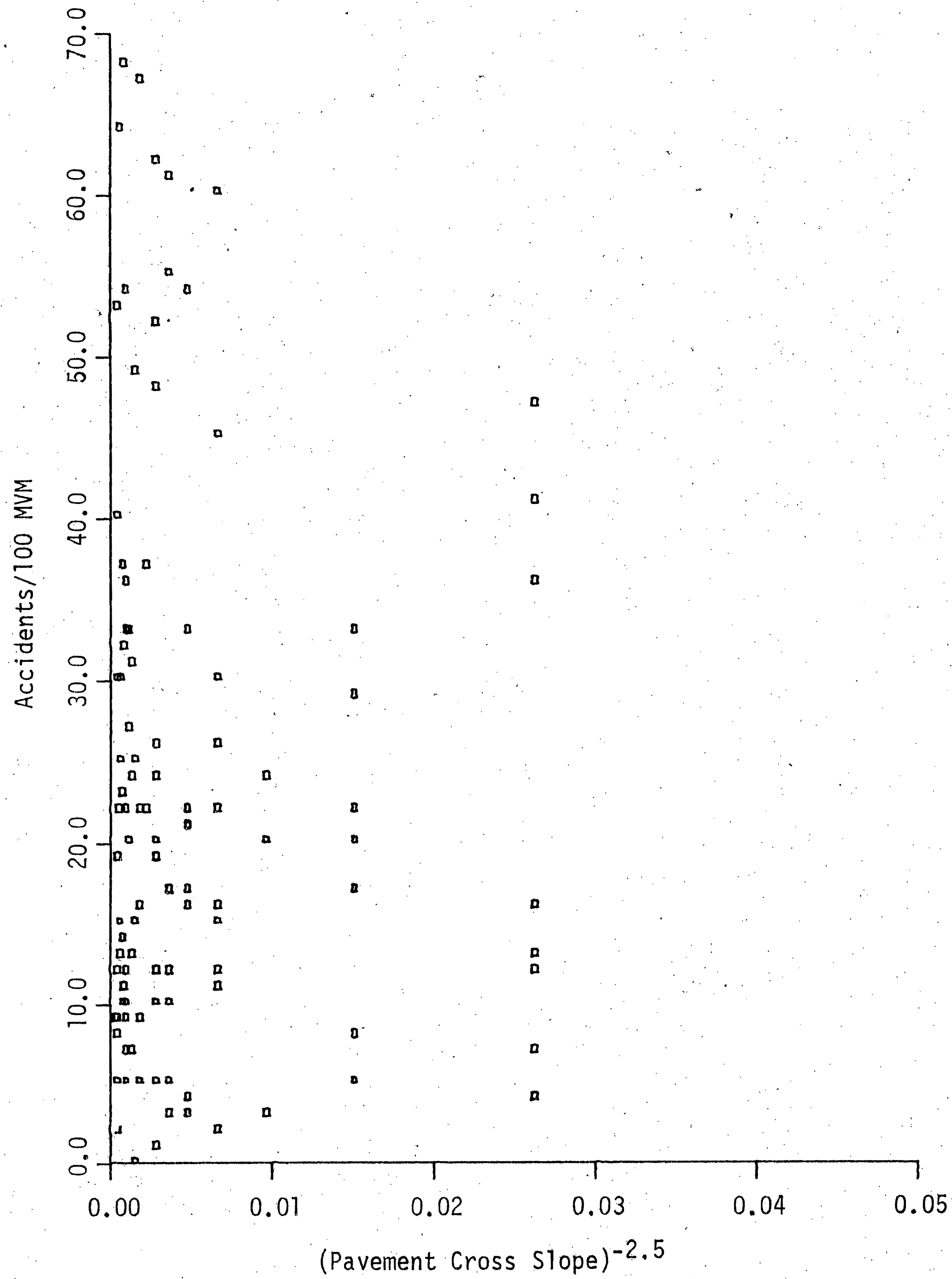


Figure 33. Accident Rate Versus (Pavement Cross Slope)$^{-2.5}$

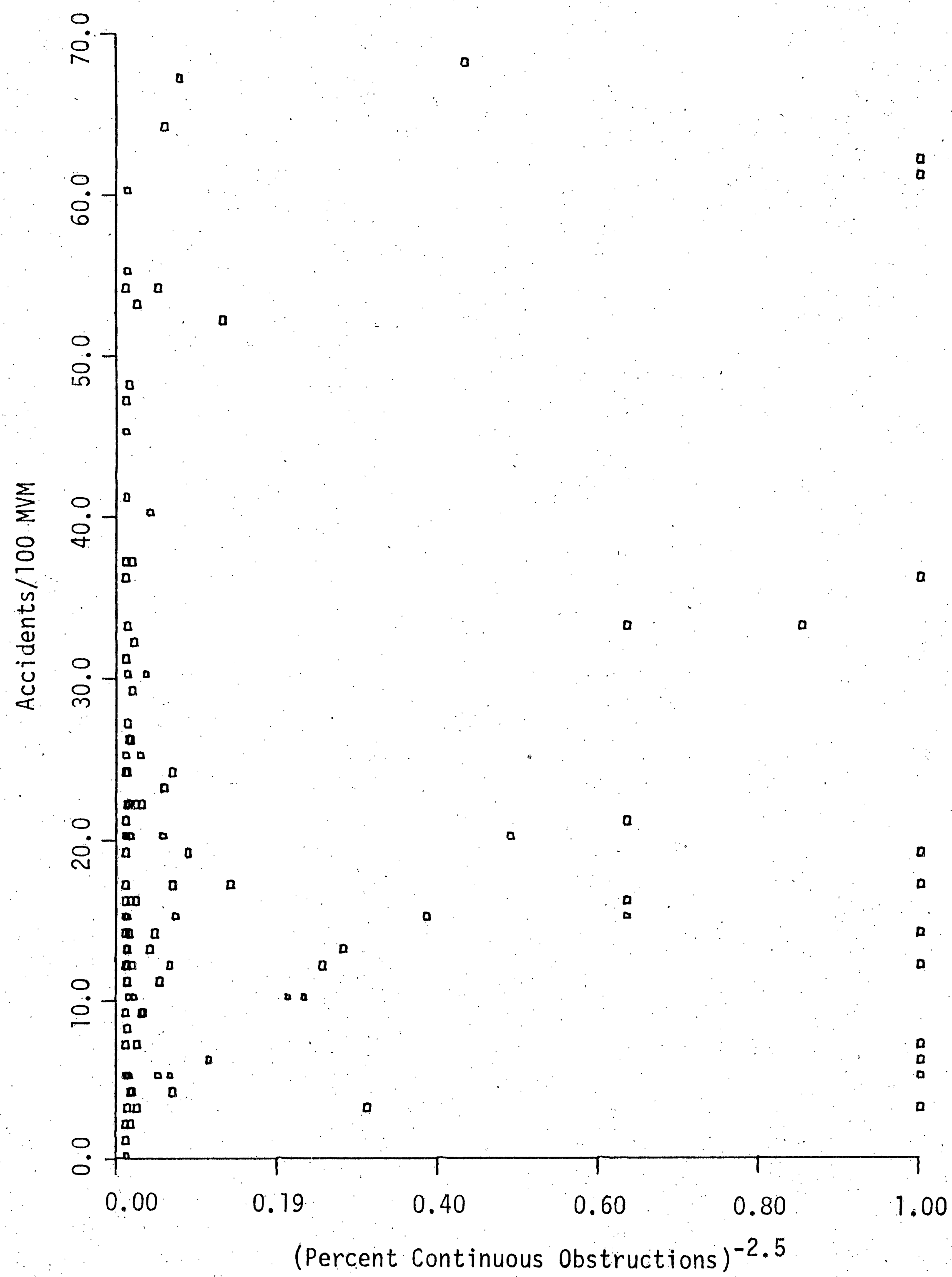


Figure 34. Accident Rate Versus (Percent Continuous Obstructions)$^{-2.5}$

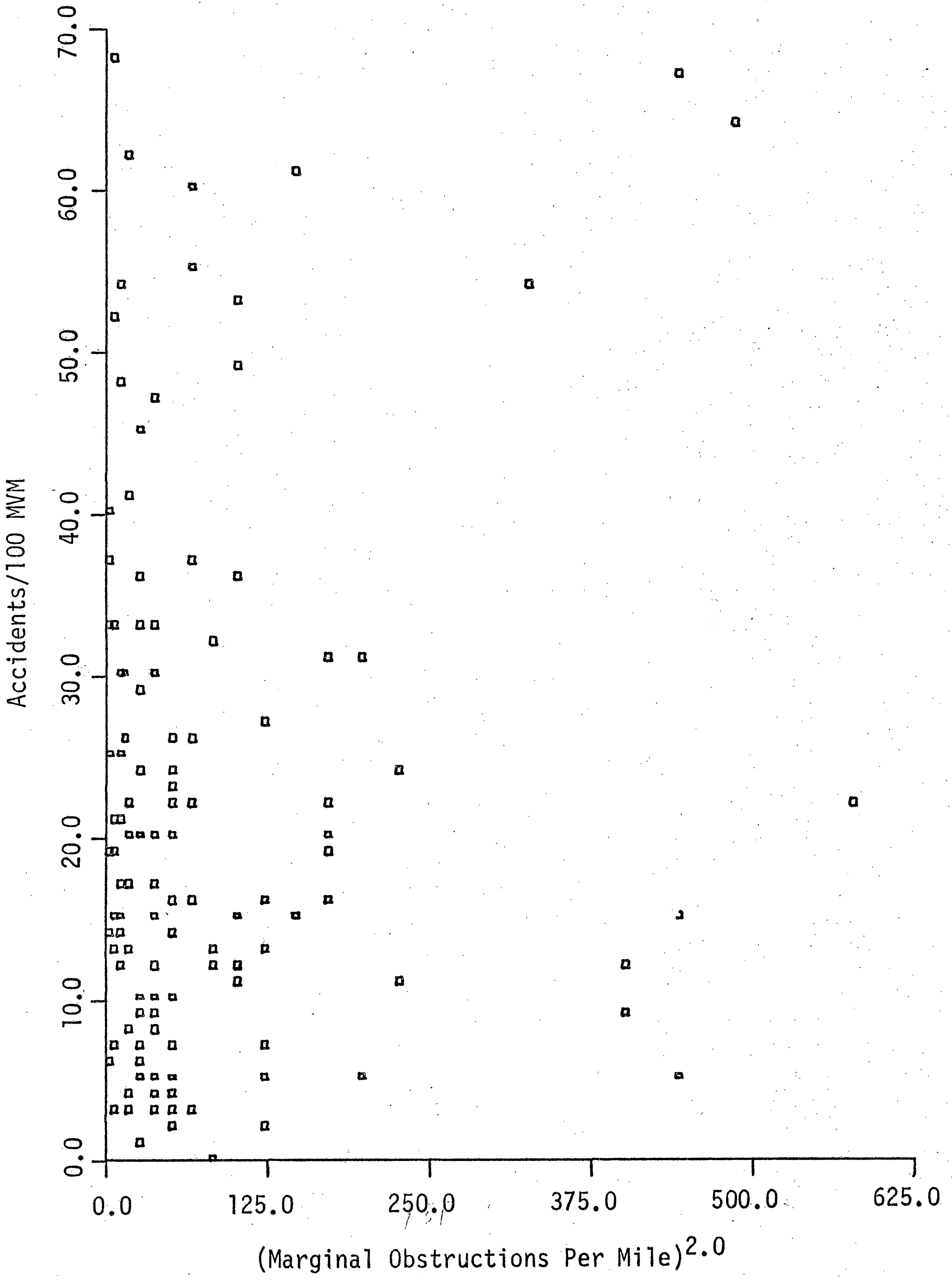


Figure 35. Accident Rate versus (Marginal Obstructions Per Mile)$^{2.0}$

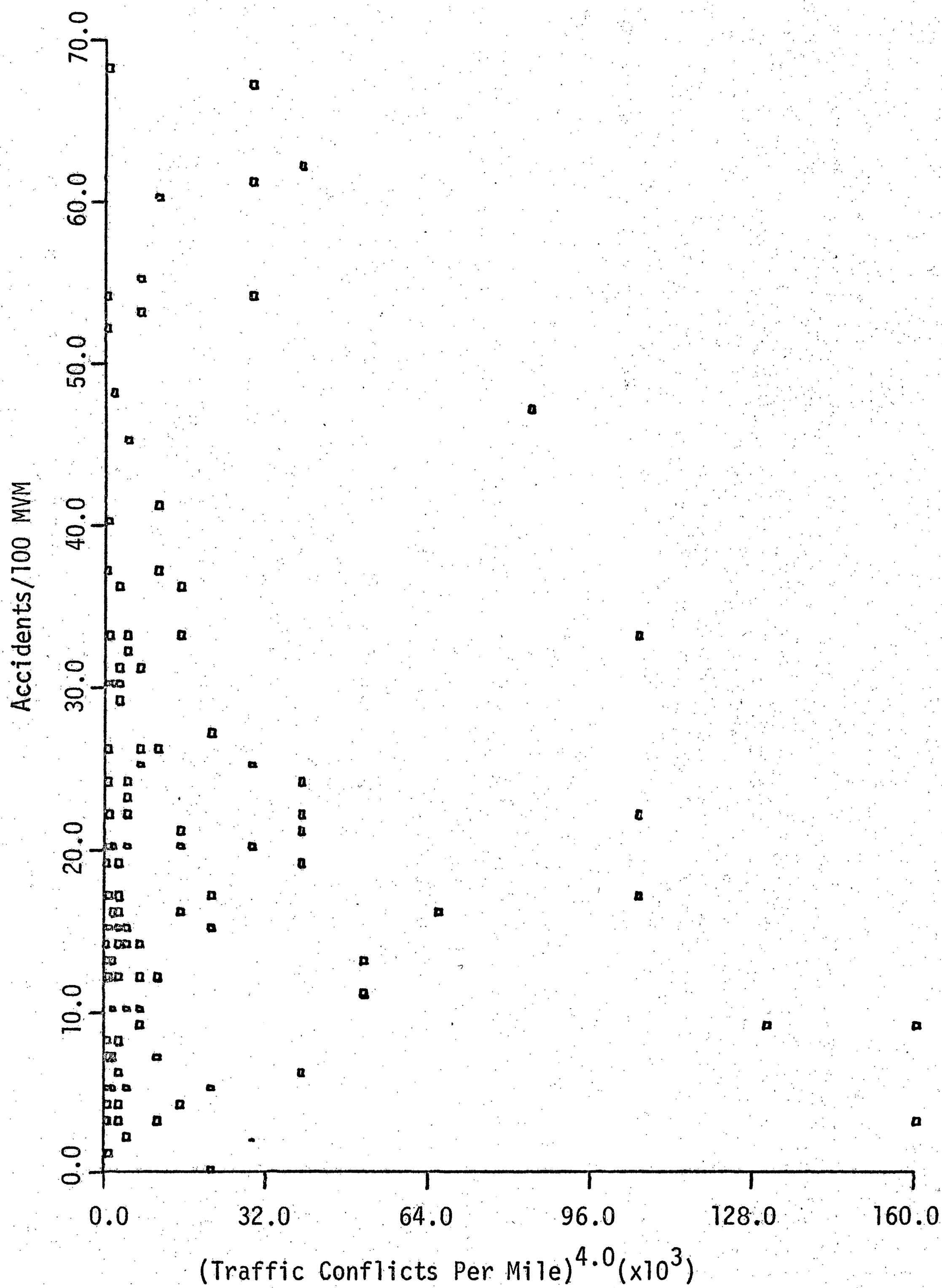


Figure 36. Accident Rate Versus (Traffic Conflicts Per Mile)$^{4.0}$

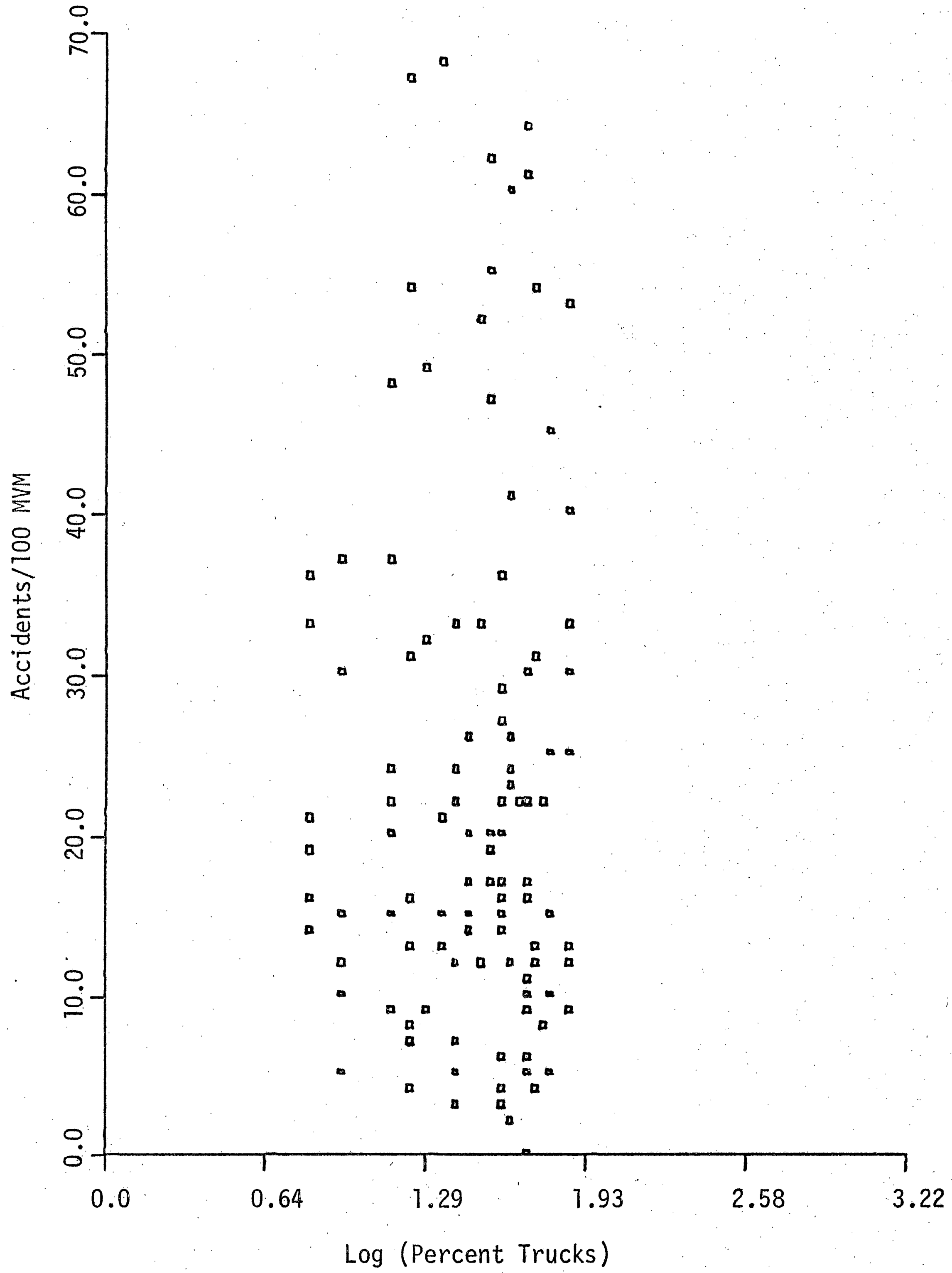


Figure 37. Accident Rate Versus Log (Percent Trucks)

Table 22. Regression Variables and Transformations

Variable		α	Variable		α	Variable		α	Variable		α
ADT(x_1)	*	2.5	ADT-CON(x_1x_8)	*	1.5	SHD-HRZ(x_4x_6)	*	-0.5	HRZ-CON(x_6x_8)	*	-4.0
LAN(x_3)	*	4.0	ADT-MRG(x_1x_9)	*	1.5	SHD-VER(x_4x_7)	*	-0.5	HRZ-MRG(x_6x_9)	*	1.0
SHD(x_4)	*	-0.5	ADT-TRF(x_1x_{10})	*	1.5	SHD-CON(x_4x_8)	*	log	HRZ-TRF(x_6x_{10})	*	4.0
CRS(x_5)	*	-2.5	ADT-TRK(x_1x_{11})	*	2.0	SHD-MRG(x_4x_9)	*	3.0	HRZ-TRK(x_6x_{11})	*	-0.5
CON(x_8)	*	-2.5	LAN-SHD(x_3x_4)	*	-0.5	SHD-TRF(x_4x_{10})	*	4.0	VER-CON(x_7x_8)	*	-4.0
MRG(x_9)	*	-2.0	LAN-CRS(x_3x_5)	*	1.0	SHD-TRK(x_4x_{11})	*	log	VER-MRG(x_7x_9)	*	1.0
TRF(x_{10})	*	4.0	LAN-HRZ(x_3x_6)	*	4.0	CRS-HRZ(x_5x_6)	*	1.0	VER-TRF(x_7x_{10})	*	4.0
TRK(x_{11})	*	log	LAN-VER(x_3x_7)	*	4.0	CRS-VER(x_5x_7)	*	4.0	VER-TRK(x_7x_{11})	*	1.0
ADT-LAN(x_1x_3)	*	2.5	LAN-CON(x_3x_8)	*	1.0	CRS-CON(x_5x_8)	*	-4.0	CON-MRG(x_8x_9)	*	1.0
ADT-SHD(x_1x_4)	*	1.0	LAN-MRG(x_3x_9)	*	2.5	CRS-MRG(x_5x_9)	*	-1.0	CON-TRF(x_8x_{10})	*	log
ADT-CRS(x_1x_5)	*	1.0	LAN-TRF(x_3x_{10})	*	4.0	CRS-TRF(x_5x_{10})	*	2.0	CON-TRK(x_8x_{11})	*	log
ADT-HRZ(x_1x_6)	*	2.5	LAN-TRK(x_3x_{11})	*	log	CRS-TRK(x_5x_{11})	*	1.0	MRG-TRF(x_9x_{10})	*	log
ADT-VER(x_1x_7)	*	2.5	SHD-CRS(x_4x_5)	*	-0.5	HRZ-VER(x_6x_7)	*	1.0	MRG-TRK(x_9x_{11})	*	2.5
									TRF-TRK($x_{10}x_{11}$)	*	4.0

-cont- SOURCE	B-VALUES
LAN-TRF (x_3x_{10})	0.00000001
CRS-TRF (x_5x_{10})	-0.09063018
CON-TRK (x_8x_{11})	31.45303286

The above variables in the regression equation explain 64.74% of the variation in the accident rate which shows an increase in estimating quality over equation for the total data set; hence the advantage in grouping the data to find a more sufficient explanation of accident rates. It is interesting to note that an increase in the pavement cross slope (x_5) value has a decreasing effect on accident rate. This supports the findings of Dart and Mann (4), and their theory that the flatter the road surface the more likely is the possibly of hydroplaning during high rainfall.

The above regression approach of using the grouped data could be valuable to a highway department in analyzing various geographical regions. Decisions could be made in allocating improvement funds for the most crucial roadway/environment components of the system for reducing accident frequency. The data grouping also enables the analyst in dismantling the complexities of the highway system.

The following is a listing of the variables for the straight, flat road regression equation according to their hierarchy of importance in explaining accident rates.

1. Average Daily Traffic and Trucks
2. Lane Width and Traffic Conflicts

3. Cross Slope and Traffic Conflicts
4. Average Daily Traffic and Marginal Obstructions
5. Continuous Obstructions and Trucks
6. Average Daily Traffic and Traffic Conflicts
7. Cross Slope

The above hierarchy of variables suggest that average daily traffic and traffic conflicts are significantly linked with accident occurrence on straight, flat roads. An initial solution would be to limit roadway access and perhaps alleviate traffic volume by four laning the roadway. Whether this would be cost-effective is another part of the solution to highway safety problems.

The results from the application of regression methodology in this chapter could provide valuable information to organizations such as state highway departments. These organizations have a certain amount of control over many of the variables from the roadway/environment and the regression methodology from this study would aid them in allocating funds to improve the system. The use of dummy variables and data grouping also have their significance in the analysis of discrete variables such as lane width and the ability to analyze data for regions with certain geometric characteristics.

Chapter Six

CONCLUSIONS AND RECOMMENDATIONS

Highway safety from a systems viewpoint is a very logical approach to a very complex but not unsolvable problem. This study defines highway safety as a systems problem and establishes a rigorous statistical methodology which can be incorporated into the analysis of those failures occurring in the highway system.

The work performed for this study has lead to the following conclusions and recommendations:

1. The highway accident matrix used for this study provides a useful and disaggregated approach to safety problems in the highway system and should be incorporated into future highway accident studies
2. The regression methodology developed in this study provides a rigorous and uncomplicated approach for highway accident analysis. The availability of statistical computer packages enables the analyses of large data sets composed of numerical as well as non-numerical data.
3. The technique used to determine transformation of the regression variables proved to be a straight forward and reliable method for improving the estimating quality of the regression equation. This technique removes much of the guesswork from the transformation selection process, therefore it should be applied to those highway accident studies where regression analysis is used.

4. The dummy variable technique discussed and demonstrated in this study was shown to be an important tool of regression analysis in analyzing variables whose relationship with accidents was either not clear or could not be graphically determined. This technique is an important aspect of regression analysis for analyzing non-numerical variables from the human and vehicle elements.
5. The importance of analyzing specific groups of a particular data set was shown with separate analyses of straight, flat and curved, hilly roadway sections. This is a valuable analysis practice to incorporate into regression techniques.
6. Further analysis and data is needed for the driver and vehicle components of the highway system in determining their roles in highway safety.

BIBLIOGRAPHY

1. Accident Facts, National Safety Council, 1972.

2. Baker, R. F., The Highway Risk Problem: Policy Issues in Highway Safety, John Wiley & Sons, Inc., New York, N.Y., 1971.

3. Barr, A. S. and Goodnight, J. H., Statistical Analysis System, Department of Statistics, North Carolina State University, Raleigh, North Carolina, August, 1972.

4. Dart, O. K. and Mann, L. Jr., "Relationship of Rural Highway Geometry to Accident Rates in Louisiana," Highway Research Record, No. 312, 1970.

5. Draper, N. R. and Hunter, W. G., "Transformations," Technometrics, Vol. 11, No. 1, February, 1969.

6. Draper, N. R. and Smith, H., Applied Regression Analysis, John Wiley & Sons, Inc., New York, N.Y., 1966.

7. Haddon, W. Jr., Luchman, E. A. and Klein, D., Accident Research, Harper & Row, New York, N.Y., 1964.

8. Hobeika, A. G., unpublished report on dummy variables in regression analysis.

9. Kihlberg, J. K. and Tharp, K. J., "Accident Rates As Related to Design Elements of Rural Highways," National Cooperative Highway Research Program, Report No. 47, 1968.

10. Little, Arthur D., Inc., The State of the Art of Traffic Safety, Praeger Publishers, New York, 1970.

11. Paddock, R. O., "The Traffic Conflicts Technique: An Accident Prediction Method," Transportation Research Record, No. 486, 1974.

12. Schultz, T. G., Berg, W. D., and Opperlander, J. C., "Evaluation of Rail Highway Grade Crossing Protection in Rural Areas," Highway Research Record, No. 272, 1969.

13. Snyder, J. C., "Environmental Determinant of Traffic Accidents: An Alternate Model," Transportation Research Record, No. 486, 1974.

14. Suits, D. B. "Use of Dummy Variables in Regression Equations," Journal of American Statistical Association, Vol. 52. Dec., 1957.

15. Traffic Safety - A National Problem, Symposium sponsored by the National Academy of Engineering, The Eno Foundation, April, 1966.

16. Virginia Department of Highways and Transportation, unpublished report on accident analysis, 1975.

17. Virgina Division of Motor Vehicles, "Motor Vehicle Accident Report Form," Richmond, Va.

APPENDIX

TIME	DATE OF ACCIDENT (6-7) MONTH (8-9) DATE (10) YEAR	DAY OF WEEK	(11) HOUR .... A.M. .... P.M.

DO NOT WRITE IN THIS SPACE

NO. .... CA. .... LET. .... LET. .... TYPE. ....

(1-2-3-4-5)

NO .... CODED BY .... TYPE ....

FAT P.I. PD 50 P.D.U.

T.T. B.V.S. NSP OFF

LOCATION

COUNTY .... CITY OR TOWN .... (14-15-16)

(17)

IF ACCIDENT OCCURRED IN RURAL AREA INDICATE DISTANCE IN MILES AND TENTHS OF MILE FROM NEAREST TOWN. USE TWO DISTANCES AND TWO DIRECTIONS IF NECESSARY.

.... MILES NORTH
.... MILES SOUTH
.... MILES EAST
.... MILES WEST

OF ☐ LIMITS OF .... CITY OR TOWN

ACCIDENT HAPPENED ON (18-19-20) .... (21)
GIVE NAME OF STREET OR HIGHWAY NUMBER (U.S. OR STATE). IF NO HIGHWAY NUMBER, IDENTIFY BY NAME

CHECK AND COMPLETE ONE

☐ AT ITS INTERSECTION WITH .... NAME OF INTERSECTING STREET OR HIGHWAY NUMBER
OR IF
☐ NOT AT INTERSECTION

.... FEET NORTH OF ....
.... FEET SOUTH
.... FEET EAST
.... FEET WEST

URBAN - LOCATE TO NEAREST INTERSECTING STREET. HOUSE NUMBER. BRIDGE. RAILROAD CROSSING. ALLEY. DRIVEWAY. UNDERPASS. NUMBERED TELEPHONE POLE. OR OTHER IDENTIFYING LANDMARK.

RURAL - LOCATE TO NEAREST INTERSECTION. SHOW EXACT DISTANCE USING TWO DIRECTIONS AND TWO DISTANCES IF NECESSARY. (22)

ROAD

ALIGNMENT (CHECK ONE) (23)

1 ___ STRAIGHT-LEVEL
2 ___ CURVE-LEVEL
3 ___ GRADE-STRAIGHT
4 ___ GRADE-CURVE
5 ___ HILLCREST STRAIGHT
6 ___ HILLCREST-CURVE
7 ___ DIP-STRAIGHT
8 ___ DIP-CURVE

SURFACE CONDITION (CHECK ONE) (24)

4 ___ DRY
5 ___ WET
6 ___ SNOWY
7 ___ ICY
8 ___ MUDDY
9 ___ OILY

DEFECTS (CHECK ONE)

12 ___ UNDER REPAIR
0 ___ LOOSE MATERIAL
1 ___ HOLES, RUTS, BUMPS
2 ___ SOFT OR LOW SHOULDER
3 ___ NO DEFECTS

TRAFFIC CONTROL (CHECK ONE) (25)

4 ___ OFFICER OR WATCHMAN
1 ___ STOP AND GO LIGHT
2 ___ STOP SIGN
3 ___ SLOW OR WARNING SIGN
7 ___ RAILROAD GATES OR SIGNALS
12 ___ TRAFFIC LANES MARKED
5 ___ NO PASSING LINES
0 ___ YIELD SIGN
9 ___ ONE WAY ROAD OR STREET
6 ___ RAILROAD WATCHMAN
8 ___ NO TRAFFIC CONTROL

KIND OF LOCALITY (CHECK ONE) (26)

12 ___ BUSINESS OR INDUSTRIAL
0 ___ RESIDENTIAL DISTRICT
1 ___ SCHOOL, CHURCH OR PLAYGROUND ZONE
2 ___ OPEN COUNTRY

LIGHT (CHECK ONE)

4 ___ DAYLIGHT
5 ___ DUSK
6 ___ DAWN
7 ___ DARKNESS-STREET OR HIGHWAY LIGHTED
8 ___ DARKNESS-STREET OR HIGHWAY NOT LIGHTED

WEATHER (CHECK ONE) (27)

1 ___ CLEAR
2 ___ CLOUDY
3 ___ FOG
4 ___ MIST
5 ___ RAINING
6 ___ SNOWING
7 ___ SLEETING
8 ___ SMOKE-DUST

SURFACE

___ CONCRETE
___ BLACKTOP
___ BRICK
___ GRAVEL
___ DIRT
___ SPECIFY OTHER

Part A - Virginia Accident Report Form

NOTE! DO NOT WRITE IN BOXES.

VEHICLES

Your Vehicle-No. 1 (28) (29) (30) Office Use Only

Make ___ Type (Sedan, Truck, Taxi, Bus, etc.) ___ Year ___ Was Vehicle Insured? ___ Yes or No ___ Vehicle License Plate ___ Year ___ State ___ Number ___ Ins. Code

(31) (32) (33)

Driver's Name ___ First ___ Middle Initial ___ Last ___ Street or R.F.D. ___ City and State ___ Born (Mo., Day, Yr.) ___ Sex ___ Race ___

(35)

Driver's Occupation ___ (34) Carpenter, Clerk, etc. ___ Driving Experience ___ Years ___ Driver's License ___ State ___ Number ___ ☐ Chauffeur ☐ Operator ☐ Beginner ___ Soc. Sec. ___ Number

Owner's Name ___ First ___ Middle Initial ___ Last ___ Street or R.F.D. ___ City and State ___ Estimated Speed Before Acc. ___ M.P.H. ___ (36-37) ___ Speed Limit ___ M.P.H. ___ Maximum Safe Speed ___ M.P.H.

Driver's License ___ State ___ Number ___ Was Safety Belt Installed? Yes ☐ No ☐ ___ Was Belt in use? Yes ☐ No ☐

State Parts of Vehicle Damaged ___ Approximate Cost to Repair Vehicle $___

Other Vehicle-No. 2 (38) (39) (40) Office Use Only

Make ___ Type (Sedan, Truck, Taxi, Bus, etc.) ___ Year ___ Was Vehicle Insured? ___ Yes or No ___ Vehicle License Plate ___ Year ___ State ___ Number ___ Ins. Code

(41) (42) (43)

Driver's Name ___ First ___ Middle Initial ___ Last ___ Street or R.F.D. ___ City and State ___ Born (Mo., Day, Yr.) ___ Sex ___ Race ___

(45)

Drivers Occupation ___ (44) Carpenter, Clerk, etc. ___ Driving Experience ___ Years ___ Driver's License ___ State ___ Number ___ ☐ Chauffeur ☐ Operator ☐ Beginner ___ Soc. Sec. ___ Number

For Other Vehicles Use Another Form

Owner's Name ___ First ___ Middle Initial ___ Last ___ Street or R.F.D. ___ City and State ___ Estimated Speed Before Acc. ___ M.P.H. ___ (46-47) ___ Speed Limit ___ M.P.H. ___ Maximum Safe Speed ___ M.P.H.

Total Vehicles Involved ___

Driver's License ___ State ___ Number ___ Was Safety Belt Installed? Yes ☐ No. ☐ ___ Was Belt in use? Yes ☐ No ☐

State Parts of Vehicle Damaged ___ Approximate Cost to Repair Vehicle $___

Damage to Property Other than Vehicles ___ (48)

Name Object, Show Ownership, and State Nature of Damage

Approximate Cost to Repair $___

INJURED

Name ___ Address ___ ☐ Driver ☐ Passenger ☐ Pedestrian ☐ Bicyclist ___ In Vehicle No. ___ (51)

Age ___ Sex ___ (50) Nature and Extent of Injury ___

(49) Race ___ ___ Was Person Killed ___

Name ___ Address ___ ☐ Driver ☐ Passenger ☐ Pedestrian ☐ Bicyclist ___ In Vehicle No. ___ (54)

Total Injured ___

Age ___ Sex ___ (53) Nature and Extent of Injury ___

(52) Race ___ ___ Was Person Killed ___

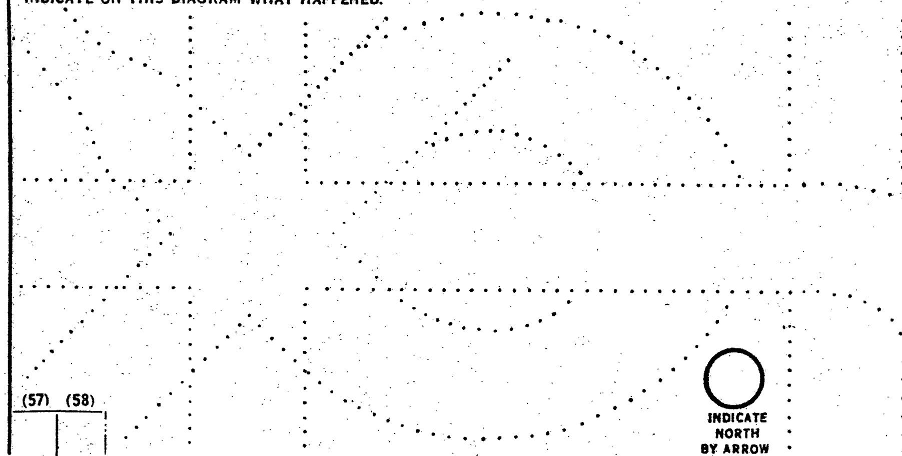
NOTE! DO NOT WRITE IN BOXES.

Part B - Virginia Accident Report Form

LOCATION IN VEHICLE OF PERSONS KILLED AND INJURED.
(CHECK ONE FOR EACH PERSON KILLED AND INJURED)

VEHICLE 1 (55) KILLED	VEHICLE 2 (56) KILLED		VEHICLE 1 (55) INJURED	VEHICLE 2 (56) INJURED
1___	___	FRONT LEFT	___	___
2___	___	FRONT CENTER	___	___
3___	___	FRONT RIGHT	___	___
4___	___	REAR LEFT	___	___
5___	___	REAR CENTER	___	___
6___	___	REAR RIGHT	___	___

INDICATE ON THIS DIAGRAM WHAT HAPPENED.



INSTRUCTIONS

1. FOLLOW DOTTED LINES TO DRAW OUTLINE OF ROADWAY AT PLACE OF ACCIDENT.
2. NUMBER EACH VEHICLE AND SHOW DIRECTION OF TRAVEL BY ARROW. 1 2
3. USE SOLID LINE TO SHOW PATH BEFORE ACCIDENT 2 , DOTTED LINE AFTER ACCIDENT 2
4. SHOW PEDESTRIAN BY:
5. SHOW RAILROAD BY:
6. SHOW DISTANCE AND DIRECTIONS TO LANDMARKS: IDENTIFY LANDMARKS BY NAME OR NUMBER.

DESCRIBE WHAT HAPPENED

(REFER TO VEHICLES BY NUMBER) ....

USE THIS SPACE FOR LISTING ADDITIONAL INJURED PERSONS. ALSO EXPLAIN QUESTIONS NOT FULLY ANSWERED BY CHECKING IN THE SPACES PROVIDED AND REASON FOR SUBMITTING IF SUPPLEMENTARY. ....

IF MORE SPACE IS NEEDED USE ANOTHER FORM OR A SHEET OF PAPER THE SAME SIZE. ....

(59) OFFENSES CHARGED DRIVER? ....

Part C - Virginia Accident Report Form

DRIVERS' ACTIONS INDICATED
(CHECK ONE FOR EACH DRIVER)

DRIVER
1 2

(60) (63)
- 1__ __ EXCEEDED SPEED LIMIT
- 2__ __ EXCEEDED SAFE SPEED BUT NOT SPEED LIMIT
- 3__ __ OVERTAKING ON HILL
- 4__ __ OVERTAKING ON CURVE
- 5__ __ OVERTAKING AT INTERSECTION
- 6__ __ IMPROPER PASSING OF SCHOOL BUS
- 7__ __ CUTTING IN
- 8__ __ OTHER IMPROPER PASSING
- 9__ __ WRONG SIDE OF ROAD-NOT OVERTAKING

(61) (64)
- 12__ __ DID NOT HAVE RIGHT-OF-WAY
- 0__ __ FOLLOWING TOO CLOSE
- 1__ __ FAILED TO SIGNAL OR IMPROPER SIGNAL
- 2__ __ IMPROPER TURN-WIDE RIGHT TURN
- 3__ __ IMPROPER TURN-CUT CORNER ON LEFT TURN
- 4__ __ IMPROPER TURN FROM WRONG LANE
- 5__ __ OTHER IMPROPER TURNING
- 6__ __ IMPROPER BACKING
- 7__ __ IMPROPER START FROM PARKED POSITION
- 8__ __ DISREGARDED OFFICER OR WATCHMAN
- 9__ __ DISREGARDED STOP-GO-LIGHT

(62) (65)
- 12__ __ DISREGARDED STOP OR YIELD SIGN
- 0__ __ DISREGARDED SLOW SIGN
- 1__ __ FAILED TO STOP AT THROUGH HIGHWAY-NO SIGN
- 2__ __ DROVE THROUGH SAFETY ZONE
- 3__ __ FAILED TO SET OUT FLARES OR FLAGS
- 4__ __ FAILED TO DIM HEADLIGHTS
- 5__ __ DRIVING WITHOUT LIGHT
- 6__ __ IMPROPER PARKING LOCATION
- 7__ __ OTHER VIOLATIONS
- 8__ __ NO VIOLATIONS

MISCELLANEOUS
(CHECK ONE ITEM FOR EACH DRIVER, IF APPLICABLE)

DRIVER
1 2

(66)
- 12__ __ AVOIDING PEDESTRIAN
- 0__ __ AVOIDING OTHER VEHICLE
- 1__ __ AVOIDING ANIMAL
- 2__ __ SKIDDING-BEFORE APPLYING BRAKES
- 3__ __ SKIDDING-AFTER APPLYING BRAKES
- 4__ __ CROWDED OFF ROADWAY
- 5__ __ HIT AND RUN
- 6__ __ CAR RAN AWAY-NO DRIVER
- 7__ __ BLINDED BY LIGHTS

PEDESTRIAN ACTIONS (CHECK ONE)

(67-68)
- 01__ CROSSING AT INTERSECTION-WITH SIGNAL
- 02__ CROSSING AT INTERSECTION-AGAINST SIGNAL
- 03__ CROSSING AT INTERSECTION-NO SIGNAL
- 04__ CROSSING AT INTERSECTION-DIAGONALLY
- 05__ CROSSING NOT AT INTERSECTION-RURAL
- 06__ CROSSING NOT AT INTERSECTION-URBAN
- 07__ COMING FROM BEHIND PARKED CARS
- 08__ GETTING OFF OR ON SCHOOL BUS
- 09__ PLAYING IN ROADWAY
- 10__ GETTING OFF OR ON OTHER VEHICLE
- 11__ HITCHING ON VEHICLE
- 12__ WALKING IN ROADWAY WITH TRAFFIC-SIDEWALKS AVAILABLE
- 13__ WALKING IN ROADWAY WITH TRAFFIC-SIDEWALKS NOT AVAILABLE
- 14__ WALKING IN ROADWAY AGAINST TRAFFIC-SIDEWALKS AVAILABLE
- 15__ WALKING IN ROADWAY AGAINST TRAFFIC-SIDEWALKS NOT AVAILABLE
- 16__ WORKING IN ROADWAY
- 17__ STANDING IN ROADWAY
- 18__ LYING IN ROADWAY
- 19__ NOT IN ROADWAY

CONDITION OF DRIVERS AND PEDESTRIAN
(CHECK ONE FOR EACH DRIVER AND PEDESTRIAN)

DRIVER
1 2 PED

(69) (71) (73)
- 9__ __ __ NO DEFECTS
- 0__ __ __ EYESIGHT DEFECTIVE
- 1__ __ __ HEARING DEFECTIVE
- 2__ __ __ OTHER BODY DEFECTS
- 3__ __ __ ILL
- 4__ __ __ FATIGUED
- 5__ __ __ APPARENTLY ASLEEP
- 6__ __ __ OTHER HANDICAP

(CHECK ONE FOR EACH DRIVER AND PEDESTRIAN)

(70) (72) (74)
- 12__ __ __ HAD NOT BEEN DRINKING
- 0__ __ __ DRINKING-OBVIOUSLY DRUNK
- 1__ __ __ DRINKING-ABILITY IMPAIRED
- 2__ __ __ DRINKING-ABILITY NOT IMPAIRED
- 3__ __ __ DRINKING-NOT KNOWN WHETHER IMPAIRED

VEHICLE CONDITION
(CHECK ONE FOR EACH VEHICLE)

VEHICLE
1 2

(75) (76)
- 8__ __ NO DEFECTS
- 1__ __ LIGHTS DEFECTIVE
- 2__ __ BRAKES DEFECTIVE
- 3__ __ STEERING DEFECTIVE
- 4__ __ PUNCTURE OR BLOWOUT
- 5__ __ WORN OR SLICK TIRES
- 6__ __ MOTOR TROUBLE
- 7__ __ OTHER DEFECTS

WHAT DRIVERS WERE DOING
(CHECK ONE FOR EACH DRIVER)

DRIVER
1 2
- __ __ GOING STRAIGHT AHEAD
- __ __ MAKING RIGHT TURN
- __ __ MAKING LEFT TURN
- __ __ MAKING U TURN
- __ __ SLOWING OR STOPPING
- __ __ STARTING IN TRAFFIC LANE
- __ __ STARTING FROM PARKED POSITION
- __ __ STOPPED IN TRAFFIC LANE
- __ __ PARKED
- __ __ BACKING
- __ __ Passing

DRIVER VISION OBSCURED
(CHECK ONE FOR EACH DRIVER)

DRIVER
1 2
- __ __ RAIN, SNOW, ETC. ON WINDSHIELD
- __ __ WINDSHIELD OTHERWISE OBSCURED
- __ __ VISION OBSCURED BY LOAD ON VEHICLE
- __ __ TREES, CROPS, ETC.
- __ __ BUILDING
- __ __ EMBANKMENT
- __ __ SIGNBOARD
- __ __ HILLCREST
- __ __ PARKED VEHICLES
- __ __ MOVING VEHICLES
- __ __ .... SPECIFY OTHER
- __ __ NOT OBSCURED

Part D - Virginia Accident Report Form

The vita has been removed from the scanned document

APPLICATION OF REGRESSION TECHNIQUES
TO HIGHWAY SYSTEMS SAFETY

by

P. D. Kiser

(ABSTRACT)

The highway system was classified into its main elements and the variables affecting each element were identified and shown how they fit into a matrix description of highway accidents. A statistical methodology was provided and applied that can effectively be used in analyzing the system of accident related variables and determining their relationship with highway accidents.

Several different regression equations were calculated which use variables from the roadway/environment element of the highway system. These equations were calculated using "dummy" variable techniques as well as the analysis of grouped data, and comments were provided on the use of each technique. The equations were useful in providing information about the roadway/environment to the analyst or decision maker in relationship to highway safety.