A Multivariate Analysis of Vehicular Speeds on Four-Lane Rural Highways

ROBERT H. WORTMAN,
Department of Civil Engineering, University of Illinois, Urbana

This research is basically a continuation of the multivariate studies of vehicular speeds which were conducted by Oppenlander at the University of Illinois. It was limited, however, to a study of vehicular speeds on 4-lane highways in rural areas. The purpose of the investigation was to evaluate the variables present on 4-lane highways which represent the driver, the vehicle, the roadway, traffic conditions and environmental characteristics.

For the evaluation of the variables, mathematical models were devised for the purpose of predicting mean spot-speeds. These models produced equations which were developed by the use of (a) multiple correlation and linear regression analysis and (b) factor analysis.

Speeds on 4-lane rural highways throughout Illinois were studied during the summer of 1962 to obtain data for the evaluation of the mathematical models. Thirty-eight variables were measured at 83 study sites to represent the conditions at each location. The variables which contributed to the greatest variation in mean spot-speed were determined by the use of multiple correlation and linear regression analysis. These variables were out-of-state passenger cars, minimum sight distance, posted speed limit, and the number of roadside establishments. Because the effect of posted speed limit on vehicular velocities has not been conclusively established, a second set of variables was determined in which speed limit was not considered. This list of variables included combination trucks, out-of-state passenger cars, minimum sight distance, median type, the presence of access control, the number of roadside establishments, and the number of access points.

Data were then analyzed by factor analysis. The principal axis factor solution resulted in the generation of 11 interpretable common factors. Of these factors only the stream friction and the traffic stream composition components were found to be statistically significant. The analysis of these two factors also verified the selection of the important variables by the use of the multiple correlation and linear regression analysis.

Seven additional sites were sampled to test the validity of the speed prediction equations developed. The observed mean speeds were compared to the estimated velocities obtained by the equations. The results of these comparisons confirmed that mean spot-speed could be predicted with a reasonable degree of statistical accuracy and that this estimation could be based on a limited number of variables or common factors.

Paper sponsored by Committee on Characteristics of Traffic Flow and presented at the 43rd Annual Meeting.

TRANSPORTATION has had a great influence on the economic, sociological, cultural, and political aspects of our society as well as other societies in the past. No one mode of transportation has dominated, but modes have developed to meet the needs of the people in accordance with technological advances. Although space travel may come in the future, the present time is really the age of the automobile.

It can be said that one of the most striking features of the modern American economy and way of life is the dependence on highway transportation. In fact, highway transportation industries have become so important in the United States today that one out of every seven employed persons work in this field (1).

The increase in automobile registrations, especially the rapid rise since World War II, has brought about problems which until recent years have been of little concern to the general public. In order to provide for the demands created by this rapidly expanding mode of transportation, highway and traffic engineers have had to initiate research projects so that they might obtain information which would enable them to continue the insurance of a safe, economic, and convenient form of transportation. One such project was the Vehicular Speed Regulation Research Project, IHR-53, which was inaugurated for the following purposes:

1. Determining the factors involved in regulating vehicular speeds;
2. Evaluating these factors and establishing warrants for the regulation of vehicular speeds;
3. Developing procedures for the application of these warrants; and
4. Developing methods and devices for obtaining maximum compliance with speed regulation.

There has been a tendency among the automobile manufacturers to build a product that is capable of extremely high speeds; however, the roadways on which these vehicles travel are often not constructed to accommodate such high velocities. In the interest of safety, it is mandatory that some regulation of speed be imposed. A recent publication stated that over the years traffic engineers have evolved a rationale for speed zoning which is somewhat as follows:

> 1. Motorists govern their speeds more by traffic and roadway conditions than by indicated speed regulations. The majority of motorists will select a speed based on roadway and traffic conditions which is reasonable and safe for them. Thus indicated speeds which are obviously higher or lower than those called for by the roadway and traffic conditions will be ignored by the majority of motorists.
>
> 2. Speed limits, to be effective, must be enforceable. This means that a speed limit must be such that a majority of motorists will observe it voluntarily, and enforcement can be directed toward the minority.
>
> 3. Any speed limit is reasonable for the roadway and traffic conditions for which it was set. Since this is generally for fair weather and off-peak volumes, it will be unreasonably high for extreme weather and traffic conditions, and low for more favorable conditions.
>
> 4. Speed limits based on studies of prevailing speeds, the character of the road, the extent and character of development along the margins of the roadway, and the accident history of the roadway tend to reduce the spread in speeds, from the highest to the lowest, and thereby result in smoother traffic flow. This smoother traffic flow results in the reduction of accidents.
>
> 5. Accidents are not related as much to speed (measured by average speed or the speeds at or below which some percentage of the vehicles travel) as to the spread in speeds from the highest to the lowest. In other words, accidents result from differences in speeds rather than from speed as a measure. (2)

Before and after studies of a number of sections of highway which have been speed-zoned bear out the general validity of these beliefs. However, one review of the influence of speed limits found that in some cases a change in speed limit produced a change in the pattern of vehicular speeds, while other cases showed no significant effect on the velocities (6).

Prior to developing new warrants for establishing speed limits, it is necessary to evaluate the variables that represent the roadway and traffic conditions. This would give an insight into the criteria for the establishment of speed zones based on safe and reasonable vehicular speeds. One study, using information taken at locations on 2-lane highways in Illinois, found that 8 variables accounted for 62 percent of the variation in mean spot-speed in rural areas. These 8 variables were out-of-state vehicle, truck combination, degree of curve, gradient, minimum sight distance, lane width, roadside establishment, and total volume (12). The variables would probably not be the same for multi-lane facilities due to the change in roadway design and environment.

It was the purpose of this investigation, therefore, to evaluate the variables which represent the driver, the vehicle, the roadway, traffic conditions, and environmental characteristics that are found on 4-lane rural highways. The variables were analyzed through the development of mathematical models by which mean velocities could be predicted.

THEORY

Oppenlander developed the concept of the time-rate of traffic flow that was used in this study (12). It was the basic intent of this report to expand on this theory by presenting mathematical models for describing mean spot-speeds on 4-lane highways.

Conceptual Model

It was conceived that vehicular speeds were a function of the various travel conditions present in and adjacent to the traffic stream (12). A variation in the types and levels of travel features of the roadway will cause a variation in the vehicular speeds. Thus high speeds are associated with roadways that present few restrictions to the vehicular flow, and decreasing speeds result with an increase in the restrictive variables.

A simplified two-dimensional model of this concept (Fig. 1) shows the speed-volume relationship. The upper limit on this model represents unrestricted flow conditions; the lower limit represents non-free flow conditions. By using the upper limit, it can be seen that an increase in volume causes a decrease in speed. The lower limit de-

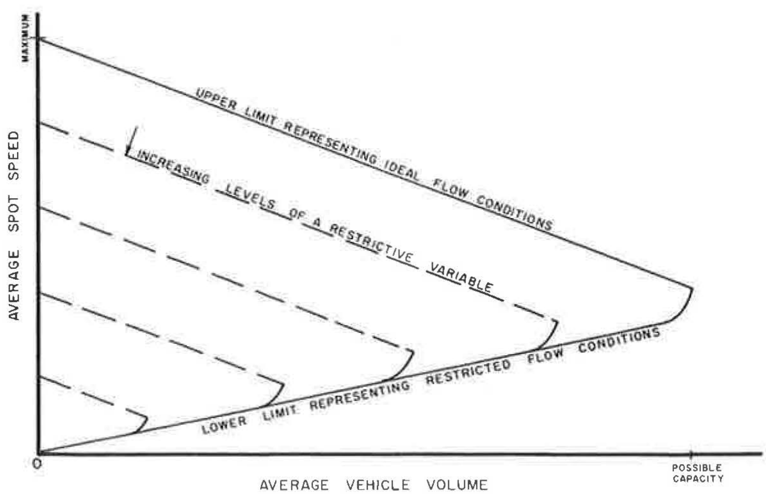

Figure 1. Schematic representation of simplified traffic-flow model.

picts highway congestion in which the vehicles are not allowed to travel at chosen speeds and are under the influence of other vehicles. Various restricted flow conditions are shown by the dashed lines which represent the upper limits under these restricted circumstances. A more specific graphical representation of this concept was presented in a paper by Normann in which the speed-volume relationship was also plotted (11). Varying levels of highway speed were used in this case to restrict the upper limits.

Because average vehicular spot-speed is not easily and completely defined by one variable, it would be reasonable, therefore, to relate mean speed to the variables which best define it. To accomplish this, mean speed was related to variables which were represented by a hyperplane of k + 1 dimensions, where k equals the number of restricting variables. As in the two dimensional case, upper and lower limits were formed to cover the entire range of values that the travel restrictions could possibly have. These limits were also designated as hyperplanes of k + 1 dimensions. This produces a polytope which geometrically represents the comprehensive traffic-flow model. Average spot speed for a given set of variables was represented by a point on or within the surface of the polytope.

Mathematical Models

Because vehicular speed is influenced by variables which are related to the driver, the vehicle, the roadway, the environment, and traffic, it was necessary to express all of these variables mathematically. Multiple linear regression equations seemed to be the best method of presentation.

Multiple linear regression equations may take many forms of which the following three were used:

$$S = a + b_1x_1 + b_2x_2 + \dots + b_nx_n \qquad (1)$$

in which

S = predicted mean spot-speed,
b = regression coefficient,
x = independent variables,
n = number of variables, and
a = intercept.

$$\frac{S}{s_{speed}} = a' + B_1 \frac{x_1}{s_1} + B_2 \frac{x_2}{s_2} + \dots + B_n \frac{x_n}{s_n} \qquad (2)$$

in which

S = predicted mean spot-speed,
s = standard deviation,
a′ = intercept,
B = standardized regression coefficient,
n = number of variables, and
x = independent variable.

$$S = \bar{S} + s\,(c_1F_1 + c_2F_2 + \dots + c_mF_m) \qquad (3)$$

in which

S = predicted mean spot-speed,
$\bar{S}$ = grand mean of sampled spot speeds,
s = standard deviation of spot speeds,
c = common factor regression coefficient,
F = common factor, and
m = number of common factors.

In each equation mean spot-speed is used as the dependent variable. The prediction of speed is based on the assumption that the relationship of the dependent variable to the independent variable is linear. Eqs. 1 and 2 were used in the multiple linear regression analysis; Eq. 3 was used in the factor analysis to relate the common factors to speed.

PROCEDURE

The next step in the study was to obtain data for the development of the proposed mathematical models.

Design of Experiment

Because drivers are influenced in the selection of vehicular speeds by a large number of variables, it was desired to determine and utilize all of the variables that might affect the driver in some manner. However, due to limitations on time, equipment, and personnel certain variables were not considered. A review of literature and past studies provided the following list of variables which were used for this study (the mathematical terms by which each variable was expressed are also shown in the units that were assigned for analytic purposes, measured or computed):

1. Light truck (two axles with single tires)-percent;
2. Single-unit truck (two or more axles with dual tires on one or more axles)—percent;
3. Truck combination (tractor with one or more trailers)--percent;
4. Commercial bus—percent;
5. Vehicle traveling on inside lane (lane next to median)—percent;
6. Vehicle traveling in a queue (a vehicle was said to be in a queue if its speed was controlled by and the same as the vehicle ahead of it)--percent;
7. Out-of-state passenger car—percent;
8. Passenger car with occupants other than driver--percent;
9. Female driver—percent;
10. Volume on the inside lane of travel—veh/hr;
11. Total one-way volume in the direction of travel—veh/hr;
12. Degree of curve—deg;
13. Rate of superelevation—feet/foot;
14. Gradient in direction of travel—percent;
15. Length of grade (measured from the P. I. to the speed site)—ft;
16. Minimum sight distance—ft;
17. Accident rate for 1961 (total number of fatal, personal-injury, and property-damage-only accidents counted for ½ mi in advance of and ½ mi beyond the speed site)—number/mile/year;
18. Day of the week (Monday was represented by 1, Tuesday by 2, Wednesday by 3, Thursday by 4, and Friday by 5. Saturday and Sunday were not used);
19. Time period (the day was broken into four time periods as follows: 7:31 a.m.—to 10:00 a.m.—represented by 1, 10:01 a.m. to 12:30 p.m.—2, 12:31 p.m. to 3:00 p.m.—3, 3:01 p.m. to 5:30 p.m.—4);
20. Cloud cover (1 if clear, 2 if overcast);
21. Drizzle (0 if no, 1 if yes);
22. Rain (0 if no, 1 if yes);
23. Pavement surface (0 if wet, 1 if dry);
24. Presence of large advertising signs (0 if no, 1 if yes);
25. Presence of centerline pavement markings (0 if no, 1 if yes);
26. Presence of edgeline pavement markings (0 if no, 1 if yes);
27. Posted speed limit—miles per hour;
28. Lane width--ft;
29. Shoulder width—ft;
30. Absolute median width—ft;

31. Median type (no median was indicated by 0, a raised median by 1, and a depressed median by 2);
32. Presence of access control (0 if no, 1 if yes);
33. Presence of curb or gutter on median side of traffic flow (0 if no, 1 if yes);
34. Presence of curb or gutter on shoulder side of traffic flow (0 if no, 1 if yes);
35. Number of roadside establishments (counted for ½ mi in advance of and ½ mi beyond the speed site)—number/mile;
36. Number of friction points, which included at-grade intersections, at-grade railroad crossings, pedestrian crossings, and school crossings (measured for ½ mi in advance of and ½ mi beyond the speed site)—number/mile;
37. Number of access points, which included all intersections, driveways, and other points of access to various forms of land use (measured for ½ mi in advance of and ½ mi beyond the speed site)—number/mile; and
38. Mean spot-speed—mph.

Generally, each speed site was represented by two sets of data, one for each direction of travel. Except for certain variables which were applicable to the overall speed site, the data was compiled for each direction of travel individually.

The complete factorial design for arranging the treatments to be tested would require the use of every combination of the different levels of the variables. However, due to the large number of variables which were to be measured in this study, it would have been difficult, if not impossible, to obtain such a set of data. As a result, the following criteria were set as a guide in the selection of the sites:

1. Each study site was as homogenous as possible for a distance of ½ mile in advance of and ½ mile beyond the study site;
2. A speed site on a horizontal curve was located near the center of the curve;
3. A study location was selected near the middle of a grade having a 400-ft minimum length;
4. Locations with and without curb or gutters were chosen;
5. Roadside development ranged from 0 to 10 buildings per mile;
6. Sites were selected with and without medians;
7. Roadways were selected on which access was or was not controlled; and
8. Study locations with and without large advertising signs were specified.

With these criteria as a guide, 83 study sites were sampled for the evaluation of the mathematical models. Seven additional locations were sampled to provide data for the verification of the results of the proposed models.

Conduct of the Studies

The actual collection of the data at the 90 locations was made by the Illinois Division of Highways during the summer of 1962. Due to the nature of some of the information that was required, certain measurements were obtained from drawings and records available in the district offices. The remaining information was measured at the speed site at the time of the study.

The radar speed meter was set up adjacent to the traffic lanes in the direction of travel that was studied; therefore, two meter setups were required when both directions of travel were studied at the same site. The observers and the speed meter were concealed from the view of the drivers throughout the conduct of the study. The observers noted the speed of each vehicle as it passed the study site, because it was not the intent to sample only the free-flowing vehicles. This provided information so that volume could be used as one of the variables.

Past studies on vehicular speeds have shown that a sample size of 150 vehicles would be of adequate size for the determination of mean spot-speed at each study site (13, 14). This sample size was based on a desired accuracy of less than 2 mph. Radar speed meters are generally accurate to ± 2 percent of the measured speed; therefore, any greater degree of accuracy obtained by a larger sample size would be unrealistic.

Analysis of the Data

Once the data collection had been completed, the next step was to evaluate the proposed mathematical models. The data were summarized for each site and punched on cards so that computers could be used for the mathematical analysis. From this point the investigation was divided into two phases:

1. Evaluation of the model by multiple correlation and linear regression analysis; and
2. Evaluation of the model by factor analysis.

The original intent was to consider as many variables as possible; however, this was impractical due to limitations of time and personnel. Certain weather variables (snow, fog, etc.) were not considered in the sample. There were no measurements taken during the hours of darkness. Because the drivers were not interviewed, variables pertaining to details on the vehicle operator were also not obtained.

Multiple Linear Regression Analysis.—New computers at the University of Illinois made it possible to use new approaches in the multivariate investigation of speeds on 4-lane highways. Until this year, computers and computer programs had not been

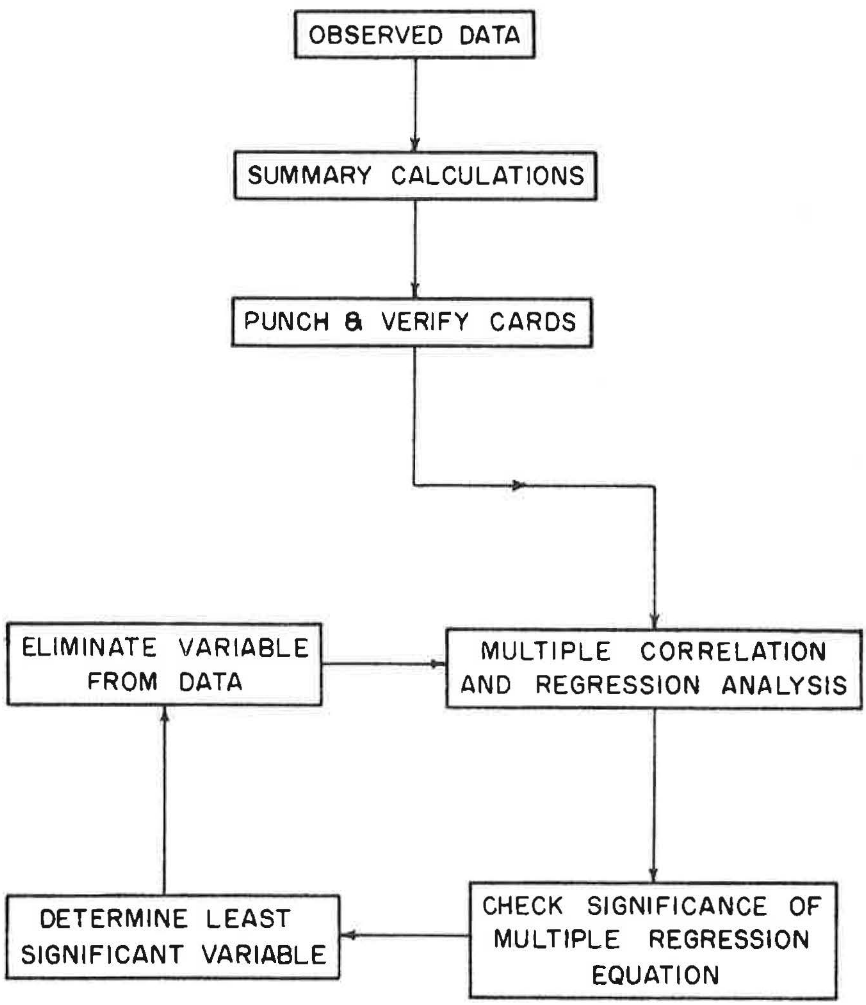

Figure 2. Multiple correlation and linear regression flow diagram.

available to personnel on this project to undertake correlation and regression analysis using the many variables which had been measured. A schematic flow diagram for the analysis which used the IBM 7090 computer is shown in Figure 2.

For this analysis, the summarized data for each site were fed into the computers on card input. The average speed at each site was used as the dependent variable, and the remaining 37 variables were designated as independent variables. The multiple correlation and regression analysis on these variables was performed by the computer, which gave the following information as output:

1. The mean of each variable;
2. The standard deviation of each variable;
3. The correlation matrix which gives the correlation of each variable with every other variable;
4. The standardized regression coefficients or the beta weights;
5. The regression coefficients;
6. The multiple correlation coefficient;
7. The standard error of estimate; and
8. The dependent variable intercept.

A multiple linear regression equation containing the large number of variables would be difficult to use and conducive to error, and it would contain many variables which are insignificant. These variables could generally be explained by other variables or else had little correlation with vehicular speed. The next step in the analysis, therefore, was to eliminate the least significant variables.

A review of the methods of elimination of variables revealed that many conflicting procedures are used. Many of these methods were derived for use without the aid of the modern electronic computer. Great controversy exists among statisticians concerning the best method. It was decided to use a mathematical method of elimination and to check this systematic elimination by a method based on logic and judgment.

One authority states that the importance of individual variables may be compared by their net regression coefficients. The size of the regression coefficients, however, varies with the units in which each variable is stated. They may be made more comparable by expressing each variable in terms of its own standard deviation, using beta coefficients (5). The beta coefficients are also known as standardized regression coefficients. Since the beta coefficients were given in the computer output, it was easy to compare the importance of each variable. This does not imply, however, that the importance of each variable will remain relative to other variables should a variable be eliminated or added.

DuBois proposed a method of elimination of variables in which the variable with the highest beta weight was extracted. The correlation and regression analysis was then performed again. The variable with the highest beta weight was again extracted. This cycle was repeated until the desired number of variables were extracted. The extracted variables were then used as the predictors of the dependent variable (4).

A second method of elimination based on beta weights was found to give better results. This procedure was similar to the one proposed by DuBois except that the variable with the lowest beta coefficient was eliminated. The correlation and regression analysis was then performed again. Again the variable with the lowest beta weight was eliminated. This technique was continued until the desired number of variables were eliminated. The remaining variables were then used to predict mean speed. This method appeared to be superior to other techniques because predictors obtained by this method gave higher multiple correlation coefficients and lower standard errors of estimate when compared to an equal number of predictors obtained by other methods; and this procedure compensated for cases where an independent variable was highly correlated with other independent variables. High association with other variables tends to reduce the beta coefficient until one of the interrelated variables is eliminated.

As each variable was eliminated, the correlation of this variable with other variables was checked to insure that the elimination was logical. It was determined to be a logical elimination if there was low correlation with speed, there were high correlations with other variables, or there was a combination of these two considerations. The standard error of estimate and the multiple correlation coefficient were analyzed after each computer run to determine if it was desirable to eliminate another variable.

Factor Analysis.—Factor analysis was originally used in psychology to analyze the interrelationships among intelligence or personality tests. Basically, factor analysis is a technique used to take a large number of operational indices, and reduce them to a smaller number of conceptual variables. Underlying the use of factor analysis is the idea that if one has a large number of variables which are intercorrelated, the interrelationships may be due to the presence of one or more underlying factors which are related to the variables in varying degrees.

The basic mechanics of factor analysis are thoroughly covered in many textbooks; therefore, it is not the intent of this study to explain factor analysis but rather present how it can be used in analyzing a set of data (6, 7). The flow diagram for the factor analysis data processing is shown in Figure 3.

The input data were the same for the factor analysis as were used in the multiple linear regression analysis. The IBM 7090 computer was also used for this investigation. The data was analyzed by the use of programs made available through the University of Illinois Statistical Services Unit (15).

The independent variables were first used as the computer input data. From this information the standard scores and the correlations of each variable with each other variable were obtained. The principal axis factors were then computed. This method of factoring was chosen because the correlation matrix was condensed into the smallest number of orthogonal factors by this method, and each factor extracts the maximum amount of variance (i.e., the sum of the squares of the factor loadings is maximized on each factor) and gives the smallest possible residuals. (7) It should be noted that after a factor pattern has been obtained, its adequacy as a description of the variables is determined by "removing the factors." This is done by forming the reproduced correlations from the pattern and subtracting them from the corresponding observed correlations. The resulting differences are known as the "residual correlations." (8)

The next step in the factor analysis was the rotation of the axes which obtains meaningful factors that are as consistent (or invariant) as possible from analysis to analysis (7). The varimax method of rotation was chosen, since when the variance is at a maximum, the factor has the greatest interpretability (8).

It was necessary to compute a factor-variable coefficients matrix so that the weights of each variable within each factor could be determined. This information was obtained by the solution of the matrix equation:

$$D = A'R^{-1} \qquad (4)$$

in which

D = factor-variable coefficient matrix,
A' = transposed varimax rotated matrix, and
R^{-1} = inverted correlation matrix.

The factor-scores or the expression of the data at each of the study sites in terms of the factors instead of the variables were computed by

$$E = DZ \qquad (5)$$

in which

E = factor-scores matrix,
D = factor-variable coefficient matrix, and
Z = standard scores matrix.

To use the factor-scores matrix in the computer program, it was easier to use Eq. 5 in the transposed form

$$E' = Z'D' \qquad (6)$$

in which

E′ = transposed factor-scores matrix,
Z′ = transposed standard scores matrix, and
D′ = transposed factor-variable coefficient matrix.

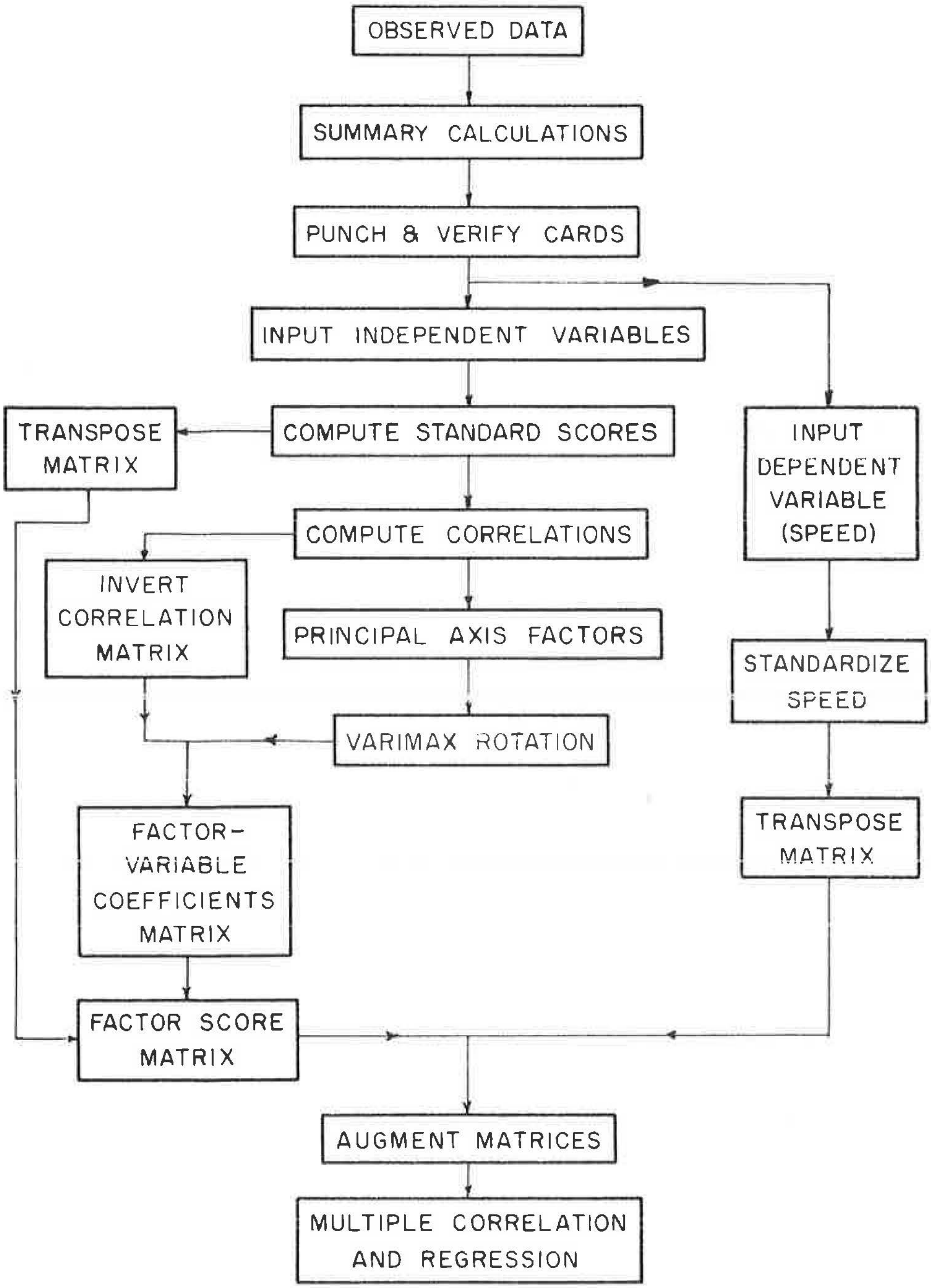

Figure 3. Factor Analysis flow diagram.

Speed data for each site were used as input for the computer and standardized. The standardized speeds were used to augment the factor-scores matrix. Multiple correlation and linear regression analysis of the factors was performed using the information contained in this one matrix.

RESULTS*

Because the two proposed mathematical models were evaluated in independent manners, it is necessary to discuss the results of each technique individually. It should be kept in mind that multiple correlation and linear regression was used as the principal multivariate solution, and factor analysis was used basically to substantiate the selection of the most important variables even though this method produced a mathematical model also.

To simplify the writing of the results, variables were designated by a number and factors were designated by an alphabetic letter. The following assumptions were made for the purpose of the analysis of the data: (a) the samples were taken randomly from normal populations, (b) homoscedasticity; and (c) linear relationships of the variables. (3)

Multiple Linear Regression Analysis

Multiple linear regression is one of the most useful tools in multivariate analysis techniques. In the case of this study, it provided an overall picture which displayed the importance of each independent variable and its relation to vehicular speed. As a result it was found to be possible to explain the major variation in mean spot-speed on 4-lane highways by using a few independent variables.

The analysis was begun by computing a multiple linear regression equation using all of the 37 independent variables. The resulting equation gave a multiple correlation coefficient of 0.9655 and a standard error of estimate of 1.61 mph. This meant that only 6.78 percent of the variation in mean spot-speed was unexplained by this equation. The next step was to eliminate the variables which explained the least variation in the speed. This was accomplished by the previously discussed mathematical technique. This system of variable elimination was checked by the use of the correlation matrix* and judgment. After the systematic elimination of the variables the following equation resulted:

$$S = 20.51 + 0.1147X_7 + 0.0005X_{16} + 0.4333X_{27} - 0.4072X_{35} \qquad (7)$$

in which

S = mean spot-speed,
X_7 = out-of-state passenger car,
X_{16} = minimum sight distance,
X_{27} = posted speed limit, and
X_{35} = number of roadside establishments.

Eq. 7 produced a multiple correlation coefficient of 0.9032 and a standard error of estimate of 2.66 mph, indicating that about 81.57 percent of the variation in mean speed has been explained. It is apparent that the elimination of the 33 variables did not greatly affect the accuracy of the prediction.

The use of posted speed limit as one of the predictors is of some controversy in the field of traffic, because reviews of literature have shown that there are conflicting reports as to the effect of speed zoning in rural areas (9, 12). The speed limits at the study sites were, with very few exceptions, statewide limits of 70 mph for controlled-access highways and 65 mph for highways with no control of access. Because no con-

*The following supporting data are available from the Highway Research Board at cost of Xerox reproduction and handling: principal axis factors matrix, summary of means and standard deviations, the correlation matrix, varimax rotated factor matrix, and the factor-variable coefficient matrix—Supplement XS-2 (Highway Research Record 72), 12 pages.

clusive information concerning the effect of such a speed limit could be found, it was decided to use posted speed limit as one of the independent variables. An analysis of only the sites where the statewide speed limit was in effect produced basically the same results.

There may be cases where the use of an equation which does not contain posted speed limit as a variable would be of benefit; hence another equation was developed:

$$S = 45.05 + 0.2046X_3 + 0.0930X_7 + 0.0009X_{16} + 1.7419X_{31} - 1.4728X_{32} - 0.4449X_{35} - 0.0511X_{37} \qquad (8)$$

in which

S = predicted mean spot-speed,
X_3 = combination trucks,
X_7 = out-of-state passenger car,
X_{16} = minimum sight distance,
X_{31} = median type,
X_{32} = presence of access control,
X_{35} = number of roadside establishments, and
X_{37} = number of access points.

The standard error of estimate for Eq. 8 was 3.05 mph and the multiple correlation coefficient was 0.9000.

The major part of the variables used in Eqs. 7 and 8 dealt with design features of the roadway, whereas the remaining variables represented the traffic stream flow. Volume did not occur as an important variable in this study, because volumes about 530 vph were not obtained in the sample. Based on present information, this volume would have little effect on average speeds (11).

Factor Analysis

Major emphasis was placed on the evaluation of the data by multiple linear regression, because the resulting equation is much easier to deal with than the equation derived by the use of factor analysis. Factor analysis was used to verify the variables obtained by the multiple correlation and linear regression technique of elimination.

Kaiser states that the following four distinguishable bases have been established for the determination of the number of factors to generate in an analysis:

1. Algebraic criterion of necessity,
2. Statistical criterion of significance,
3. Criterion of reliability, and
4. Criterion of meaningfulness.

Of these four criteria, he recommends that the last be used because the factors that have an eigenvalue (latent root) greater than one can be meaningful and reliably interpreted (10). The eigenvalues are determined from the principal axis factors matrix*, in which they are the sum of the squares of the factor coefficients of the variables for each factor. The summary of the latent roots for the factors is given in Table 1. The criterion of using factors with eigenvalues greater than one resulted in the generation of 11 factors which accounted for 73 percent of the variance.

The varimax rotated factors* provide the greatest interpretation of the factors. Arbitrary names were given to the factors depending on the variables which were the most important within each factor. The following is a list of the factors with their letter designations, descriptive names, and the major component variables and their respective factor coefficients:

A—Stream friction (representing the friction of traffic flow):

6—Vehicles traveling in a queue, 0.7424

*See footnote page 11.

TABLE 1

SUMMARY OF LATENT ROOTS

Factor	Latent Root
A	7.2841
B	3.7955
C	3.0399
D	2.6836
E	1.9530
F	1.8721
G	1.6897
H	1.3757
I	1.1980
J	1.1182
K	1.0503

27—Posted speed limit, -0.6765
28—Lane width, -0.6934
29—Shoulder width, -0.5864
31—Median type, -0.5399
34—Curb or gutter on shoulder side, 0.6597
35—Roadside development, 0.8458
36—Friction points, 0.6941
37—Access points, 0.8350

B—Roadway surface (describing the condition or nature of the roadway surface that seems to be the most prevalent):

13—Superelevation, 0.7363
22—Rain, 0.8966
23—Pavement surface, -0.6765
25—Centerline pavement marking, -0.5679
33—Curb or gutter on median side, 0.5651

C—Volume (containing variables which relate the volume characteristics):

5—Vehicle traveling on the inside lane, 0.9139
10—Volume on the inside lane, 0.9094
11—Total volume, 0.5575
18—Day of the week, 0.4230
30—Median width, -0.3853

D—Vehicle description (variables that describe the vehicle and its occupants):

2—Single-unit truck, -0.6883
7—Out-of-state passenger car, 0.4227
8—Passenger cars with occupants, 0.7043
11—Total volume, 0.4821
32—Access control, 0.5155

E—Marginal clearance (not clearly defined; therefore, named for variables representing the edgeline and shoulder width):

20—Cloud cover, 0.7215
26—Edgeline pavement marking, 0.8143
29—Shoulder width, 0.4608

F—Lane friction (depicting internal friction of travel flow due to lane and median width, and median type):

16—Minimum sight distance, -0.6553
18—Day of the week, -0.4020
28—Lane width, 0.4924
30—Median width, -0.4119
31—Median type, -0.4664

G—Safety (not representing a definite grouping of variables, but named for variable with the greatest coefficient):

4—Commercial bus, -0.4850
17—Accident rate, -0.6620
19—Time period, 0.5594

H—Driver distraction (named for the most important variable):

12—Degree of curve, 0.6834
24—Outside advertising, -0.7299
36—Friction points, -0.4316

I—Weather (inclement weather played major role):

19--Time period, -0.4391
21--Drizzle, -0.7309
23--Pavement surface, 0.4879
30--Median width, -0.3414
31--Median type, -0.3430

J--Traffic stream composition (representing the physical make-up of the traffic stream):

1--Light truck, 0.6345
3--Truck combination, -0.7012
7--Out-of-state passenger car, -0.7182
9--Female driver, 0.6989
15--Grade length, 0-.4723
37--Access points, 0.3315

K—Vertical resistance (gradient is the major variable):

14—Gradient, -0.7895
16—Minimum sight distance, 0.2613
30—Median width, -0.2604
31--Median type, -0.2614

In order to express the generated factors as a function of the original variables, it was necessary to compute the factor-variable coefficients. Multiplication of this matrix by the standard scores matrix results in the factor score matrix.

The factor scores matrix was then augmented with the standardized mean speeds for each study site. Multiple correlation and linear regression was performed on the data contained in this combined matrix. The analysis was accomplished by the use of data which was in standard score form; hence the resulting regression coefficients were also the correlation of speed with each of the factors. Table 2 lists these correlations. The coefficients were analyzed to determine which were significant at the 95 percent level. Only the stream friction and traffic stream composition factors were found to be significant; therefore, the following multiple linear regression equation was written to express mean spot-speed in terms of common factors:

$$S = 52.67 + 6.192\,(-0.6099F_A - 0.4826F_J) \tag{9}$$

TABLE 2

CORRELATION OF MEAN SPEED WITH THE FACTORS

Factor	Correlation Coefficient
A	-0.6099
B	-0.0729
C	-0.2157
D	0.2099
E	0.0928
F	-0.1757
G	0.1396
H	-0.0405
I	0.0142
J	-0.4826
K	0.0013

in which

S = predicted mean spot-speed,
F_A = stream friction factor, and
F_J = traffic stream composition factor.

The accuracy of the prediction by Eq. 9 was not as great as either of the two equations that were developed through the use of multiple correlation and linear regression, because the multiple correlation coefficient was 0.7746 and the standard error was 3.92 mph in this case. This prediction accuracy was not required because the factors were used to verify the more important variables obtained in the previous analysis.

If the stream friction and the traffic stream composition factors are evaluated, the variables used in the multiple linear regression equations (with the exception of minimum sight distance) are also contained in the significant factors. This provided an excellent check on the accuracy and suitability of the mathematical method of determining the important variables.

With the aid of the factor-variable coefficient matrix*, it was possible to express the factors in terms of the variables. The following equations were written to compute the significant factors by the use of the more important study variables contained in each of the factors:

$$F_A = 0.2019Z_6 - 0.1268Z_{27} - 0.1552Z_{28} - 0.1808Z_{29} - 0.1064Z_{31} + 0.1384Z_{34} + 0.2037Z_{35} + 0.1342Z_{36} + 0.1692Z_{37} \quad (10)$$

$$F_J = 0.2078Z_1 - 0.2835Z_3 - 0.2475Z_7 + 0.2800Z_9 - 0.2125Z_{15} + 0.0199Z_{37} \quad (11)$$

in which

F = common factor, and
Z = standard score of the variables.

The standard scores of the variables may be calculated by

$$Z = (X - \bar{X})/s \quad (12)$$

in which

Z = standard score,
X = observed value of the variable,
$\bar{X}$ = mean value of the variable, and
s = standard deviation of the variable.

The procedure for predicting mean spot-speed is not as efficient as the methods and equations developed in multiple linear regression; however, it can also be used to make the speed prediction.

Verification of Results

The three equations that were developed for the prediction of mean speed would be of little use unless their validity was proven. Seven of the 90 sites that were sampled

TABLE 3

RESULTS OF VERIFICATION STUDY

	Obs. Mean	Variables Speed Limit Included		Variables Speed Limit Excluded		Factors		
		Est. Mean	Diff.	Est. Mean	Diff.	Obs. Mean	Est. Mean	Diff.
1	55.18	51.74	+3.44	52.94	+2.24	55.18	54.18	+1.00
2	54.24	52.33	+1.91	52.87	+1.37	54.24	53.71	+0.53
3	48.32	50.15	-1.83	49.07	-0.75	48.32	52.16	-3.84
4	52.67	52.05	+0.62	52.15	+0.52	52.67	57.14	-4.47
5	54.24	53.26	+0.98	55.34	-1.10	54.24	50.88	+3.36
6	55.33	52.18	+3.15	54.23	+1.10	55.33	52.02	+3.31
7	56.76	58.57	-1.81	57.33	-0.57	56.76	56.93	-0.17

*See footnote page 11.

were used for verification. These locations were not included in the development of the equations so that an independent comparison of mean speeds at these locations could be made without the introduction of a bias.

Mean spot-speeds for each of the verification sites were computed by Eqs. 7, 8, and 9. These predictions were then compared to the actual observed velocities at these sites (Table 3). The difference between the actual and the estimated speeds for each equation were statistically analyzed at the 95 percent level; there was no significant difference between the actual and the estimated speeds for any of the equations. Therefore, mean spot-speeds on 4-lane rural highways can be estimated by the use of a mathematical model with a reasonable degree of accuracy.

CONCLUSIONS

Due to limitations of this study of vehicular speeds on rural 4-lane highways, the conclusions concerning the prediction of velocities are valid only for the conditions as found at the study locations. It should be noted, however, that generalizations concerning statistical techniques would also be true for other cases.

1. The prediction of mean spot-speed can be accomplished by use of multiple linear regression equations. The equations may use variables or common factors as the independent speed predictors.
2. Many of the variables account for only a minor portion of the variance in average speed; therefore, the mean velocity can be prognosticated by use of a few independent variables.
3. Use of beta weights provided a satisfactory mathematical approach as well as a logical technique for the elimination of the variables which produced the least effect on the variance in mean spot-speed. This method of variable elimination, when compared to other techniques, determined variables which gave higher multiple correlation coefficients and lower standard errors of estimate.
4. The variables defining out-of-state passenger cars, minimum sight distance, posted speed limit, and the number of roadside establishments explained to a statistically acceptable degree the variation in mean spot-speed. When the posted speed limit was not considered in the analysis, the average speed was predicted with a similar degree of accuracy by the variables which represented combination trucks, out-of-state passenger cars, minimum sight distance, median type, presence of access control, the number of roadside establishments, and the number of access points.
5. The stream friction and the traffic stream composition factors accounted for the major portion of the variation in average speed.
6. Use of new computers has allowed the inclusion of many independent variables in multiple linear regression analysis, and provided a more convenient method of analyzing and predicting speeds than the use of common factors.

RECOMMENDED FURTHER RESEARCH

The following suggestions are recommended for additional and further research:

1. Because this study was limited to rural 4-lane highways, the investigation should be continued to urban locales and highways of more than 4 lanes. It is not expected that variables would be of the same importance in a different set of conditions; hence the principal variables describing vehicular speed in rural areas would not necessarily account for the major portion of the variance in speed as a result of this change in surroundings.
2. The measurement of additional variables dealing with the road surface, the driver, and the vehicles would probably explain more of the variance in mean spot-speed. At the present time, it is difficult to measure many of these variables; therefore, as techniques and devices become available for their appraisal, they should be included in multivariate studies.
3. The unlimited opportunities for improvement of statistical techniques and analyses as a result of the development of new computers should be explored.
4. It was found that the statewide posted speed limit had an effect on the average

spot-speed. This contradicts some of the present theories concerning the effect of speed limits. A more detailed investigation of this variable is warranted.

5. It would be logical to determine a maximum safe vehicular speed based on the traffic stream components, roadway conditions, and the driver's capabilities prior to the establishment of warrants for speed limits. The major problem would be to define "maximum safe velocity." This investigation provided a method of determining the variables which have the greatest effect on mean speed; therefore, further research is needed to incorporate this technique of resolving variables with the maximum safe speed for computation of a safe, economical, and convenient speed limit.

ACKNOWLEDGMENTS

Appreciation is expressed to John E. Baerwald, Professor of Traffic Engineering, for his guidance in the preparation of this thesis. A. Keith Stonecipher and Jerry D. Brand also contributed assistance throughout this research investigation.

The Illinois Division of Highways gave aid in the collection of the data; the Statistical Services Unit at the University of Illinois provided assistance in the analysis of the data; Henry F. Kaiser, Associate Professor of Psychology, University of Illinois, gave advice concerning factor analysis.

This report is part of the work program of the Vehicular Speed Regulation Project, IHR-53, sponsored by the Department of Commerce, Bureau of Public Roads, and the State of Illinois, Division of Highways.

This research project is under the general administrative supervision of Ellis Danner, Program Director, John E. Baerwald, Project Supervisor, and A. Keith Stonecipher, Project Investigator, all of the University of Illinois. The Advisory Committee members are Charles H. Bowman and William E. Kappauf of the University of Illinois, William E. Chastain, Sr. and Warren L. Kessler of the Illinois Division of Highways, and Steiner Silence and I. James Ulak of the Bureau of Public Roads.

REFERENCES

1. Automobile Manufacturers Association, "Automobile Facts and Figures 1962." Detroit (1962).
2. Automotive Safety Foundation, "Traffic Control and Roadway Elements." Washington, D. C. (1963).
3. Blalock, Hubert M., "Social Statistics." McGraw-Hill (1960).
4. DuBois, Philip H., "Multivariate Correlation Analysis." Harper and Brothers (1957).
5. Ezekiel, Mordecai, and Fox, Karl A., "Methods of Correlation and Regression Analysis." 3rd Ed., Wiley (1959).
6. Fisher, Ewen Stuart, "An Appraisal of a Theory of Vehicular Speed Regulation." Master's thesis, Univ. of Illinois (Jan. 1963).
7. Fruchter, B., "Introduction to Factor Analysis." VanNostrand (1954).
8. Harman, H. H., "Modern Factor Analysis." Univ. of Chicago Press (1960).
9. Jester, L. R., "The Effect of Various Factors on Vehicle Speeds." Joint Highway Research Project, Purdue Univ. (1963).
10. Kaiser, Henry F., "The Application of Electronic Computers to Factor Analysis." Paper read at the American Psychological Association, Cincinnati (Sept. 7, 1959).
11. Normann, O.K., "Research to Improve Tomorrow's Traffic." Traffic Engineering, Vol. 29, No. 7, pp. 11-21 (April 1959).
12. Oppenlander, J. C., "Multivariate Analysis of Vehicular Speeds." Ph. D. thesis, Univ. of Illinois (April 1962).
13. Oppenlander, J. C., "Sample Size Determination for Spot-Speed Studies at Rural, Intermediate, and Urban Locations." Highway Research Record No. 35, pp. 78-80 (1963).

14. Oppenlander, J. C., Bunte, W. F., and Kadakia, P. L., "Sample Size Requirements for Vehicular Speed Studies." HRB Bull. 281, pp. 68-86 (1961).

Computer Programs

15. "Stat Pack." Statistical Services Unit, Univ. of Illinois.