STATISTICAL TECHNIQUES IN THE FIELD OF TRAFFIC ENGINEERING AND TRAFFIC RESEARCH

T. W. FORBES
UNIVERSITY OF CALIFORNIA AT LOS ANGELES

1. Introduction

Although the use of mathematical statistics has not been very extensive in traffic engineering until fairly recently, there is coming to be a greater use of statistical methods and techniques in this field and especially in the field of traffic research. An attempt will be made in this paper to review briefly the types of statistical techniques that have been used as well as to indicate certain types of problems where such techniques will probably be of greater use in the future. It has not been possible to cover all of the available studies so that no claim is made of completeness. Many other studies of equal merit could have been used as illustrations of methods in place of those selected.

2. Traffic surveys—estimates of central tendency and sampling

Of first importance historically and also in the extent of general use is the use of statistical techniques in connection with traffic surveys. Here the important questions to be answered are: How can the best estimate of average daily traffic volume be obtained? How large a sample is necessary? How should the sample be selected to give the most accurate estimate for the least cost? Similar questions also apply to origin and destination, parking and other types of surveys. The problems involved are essentially those of obtaining the best estimate of central tendency and involve the general problem of obtaining an adequate and a representative sample.

Before the days of the automatic traffic counting devices cooperative surveys were carried on by various states and the Federal Bureau of Public Roads to investigate traffic on Federal aid highways. These involved the obtaining of counts of daily traffic at certain stations on these highways. Later as the attempt was made to apply traffic engineering methods to city traffic the need for more information at a smaller cost led to the development and use of what has come to be known as the "short count method" of making traffic surveys [16].

Although the method was proposed on a basis of logic and common sense rather than of statistical theory, it did represent an attempt to obtain a representative sample. It was noted from 24 hour counts, for instance, that "the percentage of total daily traffic occurring in any given hour is approximately constant at different points along the same route" and that "the total volume of traffic does not

On leave from the University of Hawaii.

vary materially between normal weekdays, Monday to Friday inclusive." It was therefore held unnecessary to count all of the 24 hours or all of the days of the week. The familiar short count method was evolved which took one hour samples at various different stations and corrected these in terms of the 24 hour count at some nearby "master station."

Although the short count method became popular due to its economy, it was greeted by some skepticism which appears to have stimulated studies of counts of different length. Also, of course, sampling theory was being tested. Shelton [24] checked the adequacy of different length counts by means of complete 24 hour data from the Holland tunnel traffic for the year 1933. Sixteen watches of eight hours each were selected from the data and also spaced one hour counts were obtained covering the same traffic in such a way as to make a total of 70, 62, 35, 28, 24 and 19 hours of record. It was shown that all of the spaced one hour records down to 24 hours gave a more accurate estimate of central tendency, standard deviation and range of variation than did the 16 watches of eight hours each. It was pointed out that although it was "surprising to find that the measures obtained from 24 separate hours are more stable than those of 16 watches of eight hours" this statement was based on the results of a study of actual data and not merely on the theory of sampling. It was also pointed out, however, that these were counts distributed throughout the period of the year and not merely a single count at one station, as in the case of the original short count method.

A more detailed report on the variation of the daily volume of traffic based on records of the Holland tunnel and the George Washington Bridge was given by Shelton [25] in which it was pointed out that the increase of the number of separate and independent observations was the largest factor in the precision of estimate.

Cherniack [5] presented an elaborate analysis of the traffic at various Hudson River crossings and indicated not only variations in daily and monthly average traffic but also indicated the standard deviation around the median (figure 1).

On the basis of patterns for days of the week for different crossings (figure 2) and hourly vehicle traffic patterns (figure 3), he concluded that the use of the "borrowed" patterns as a basis for the correction factor for short counts may lead to error and that the pattern for every important facility should be determined in order to obtain accurate estimates.

Shelton [26] also reported from a study of Michigan data wide changes of the coefficient of variation between hours of the day, days of the week and months of the year which also differed depending on highway location (figures 4, 5, and 6).

Vickery [28] summarized the studies of Shelton and Cherniack and discussed the whole question of survey method. He pointed out that there may be two errors in the short count method, one of which is the error in the count, and the second of which is the error in the expansion factor. Also, when several counts are made, and the "method of quadrature" is used, there will be not only random errors, but also "errors of quadrature" resulting from the systematic diurnal and seasonal cycles mentioned above. He pointed out that periodic spacing, and an increase in the number of counts will reduce the error of quadrature. Also if the characteristics of the traffic cycles are known, the spacing of counts should be allocated in accordance with principles of stratified sampling as set forth by Neyman. He pointed out

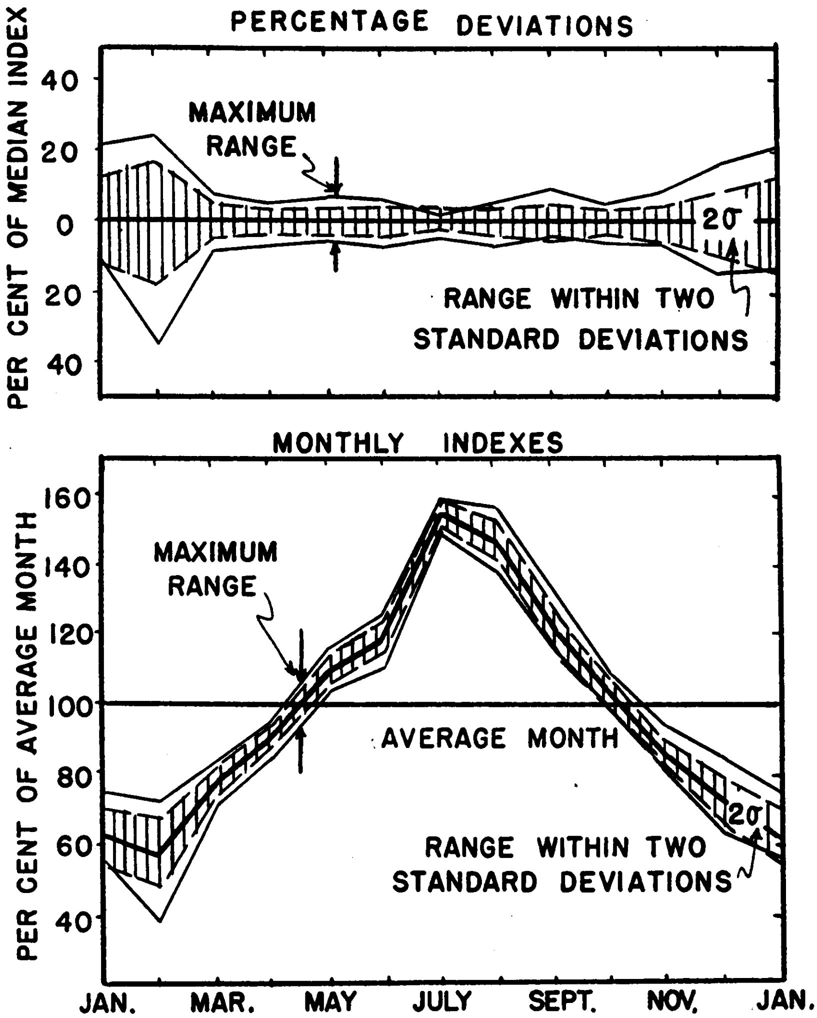


FIGURE 1

Median monthly indexes and maximum and standard deviations for lower Hudson River crossings vehicular traffic. (Cherniack, 1936, figure 5, p. 257.)

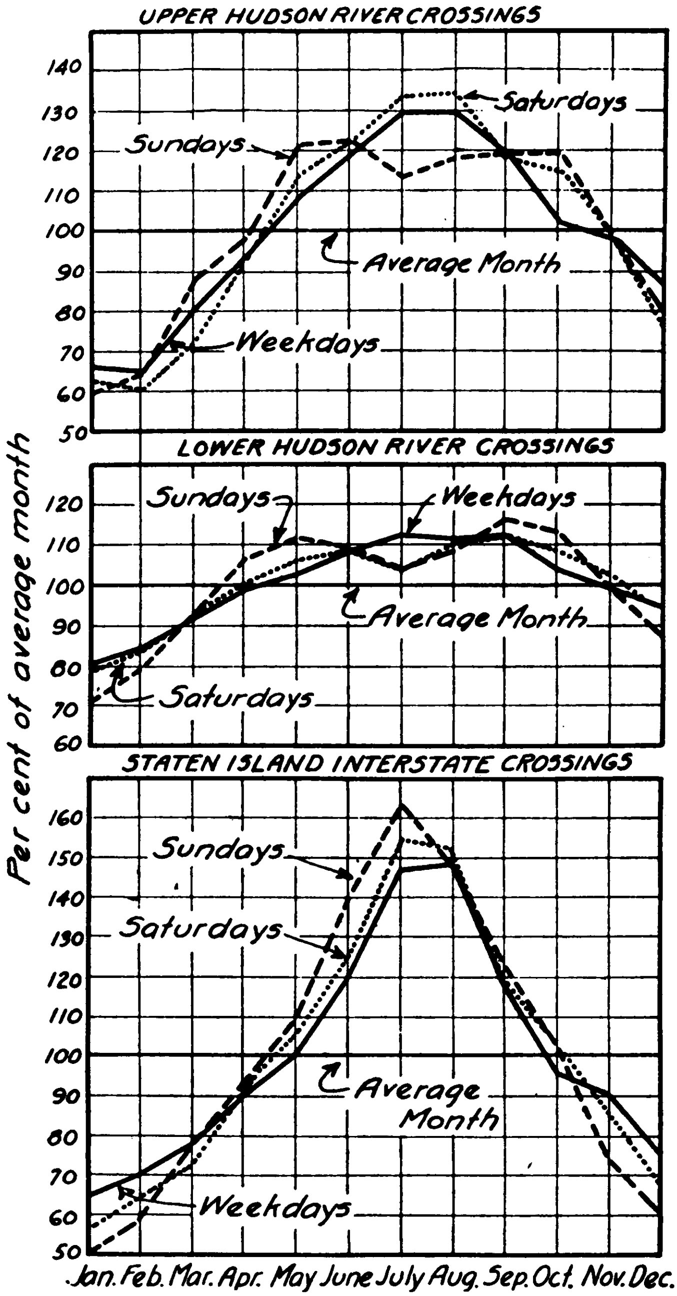


FIGURE 2

Patterns for groups of New York crossings for weekdays, Saturdays and Sundays. (Cherniack, 1936, figure 6, p. 258.)

that Neyman's formula says essentially that the sample should be allocated to strata in proportion to the product of the standard deviation of the stratum and to the number of cases in the stratum. Neyman's method of "double sampling" was recommended as useful. This method uses a preliminary random sample to obtain certain control information by which a stratified sample may be designed for the second part of the survey, and an estimate of the best allocation of funds between the preliminary survey and the final survey may be made. However, Neyman [18] pointed out that in some cases the double sampling method is of advantage whereas in others it may not be and that certain characteristics of the sample must be known in the first place in order to estimate the best allocation of funds between the two parts of the double sample.

Vickery [27] and Swain [23] reported the use of "extremely short counts" of five minute duration spaced thirty minutes apart for estimating city street traffic. Spacing of the counts was based on "judgment and observation of local conditions." The practical limitation of the method, especially in nonurban areas, was said to be the time and expense of transportation of personnel between the counting stations which may be some miles apart. Swain placed observers on the tops of tall buildings to eliminate "the time lost between stations" and in other areas used observers who traveled between stations (which were located close together) by means of motorcycles. Thus practical application was made of the principle of multiple short counts spaced over the diurnal cycle.

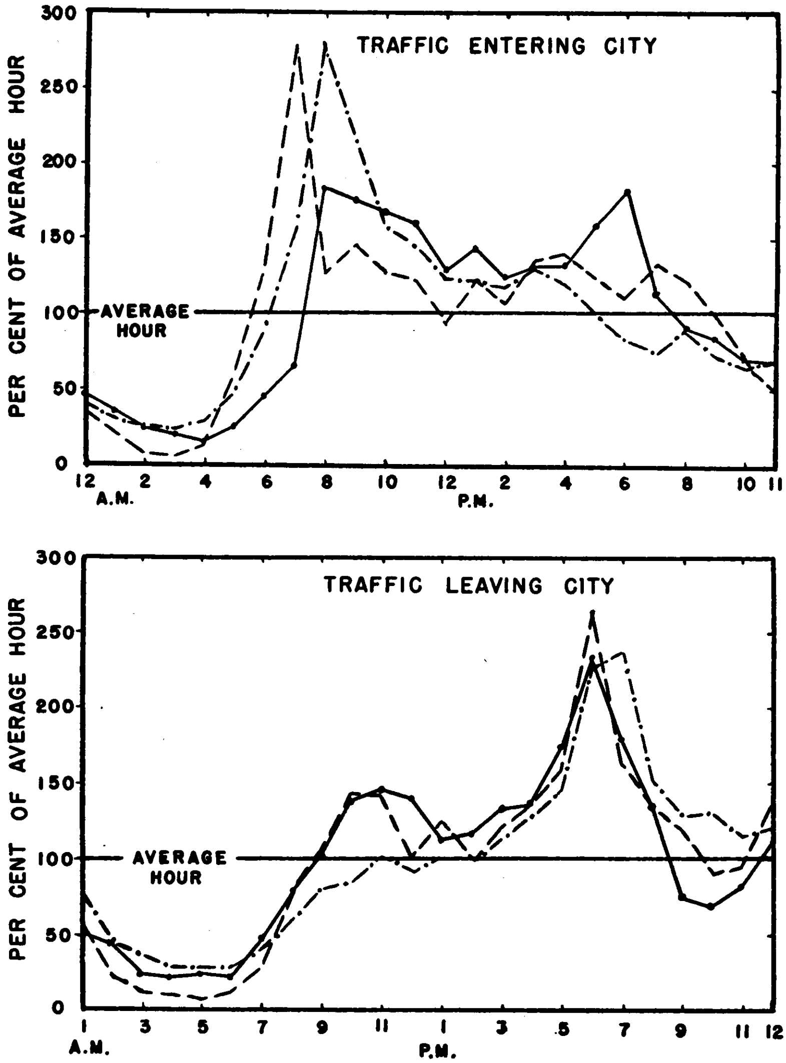
A further analysis of traffic patterns and the effects of spaced short counts was reported by Cunningham [4]. He found that counts between 1 and 8 P.M. showed the least deviation from the 24 hour annual mean. This is logical in comparison with his other counting periods when viewed in terms of the portion of the diurnal traffic cycle which was covered.

As to the effect of shortening the duration of counts, the average error of estimate did not increase markedly until the length of the sample was reduced below fifteen minutes. These comparisons were based on a 28 count schedule so arranged as to maintain a 13 day spacing between successive counts. Figure 7 shows distributions of deviations for 3 lengths of counts.

Vickery [28] also gave consideration to various other types of surveys in which the sampling problem and the problem of estimating a total count or an average is involved. It is thus seen that the problem of obtaining the best estimate of measures of central tendency has received considerable attention. The daily, weekly, and seasonal variations of traffic have been rather fully investigated for various different kinds of traffic conditions and these must be taken into consideration in any statistical treatment of highway traffic data.

3. Problems where individual driver vehicle variability is of major importance

In contrast to the problems discussed above, there are many problems where the variation of performance of different drivers and vehicles cannot be ignored. An index of this variation in performance may, in fact, be of fundamental importance. Under this heading come such questions as these. What sight distances are necessary for the design of safe passing distances on highways? How are vehicle speeds related to different highway conditions and design considerations? What is the



Based on traffic as follows:

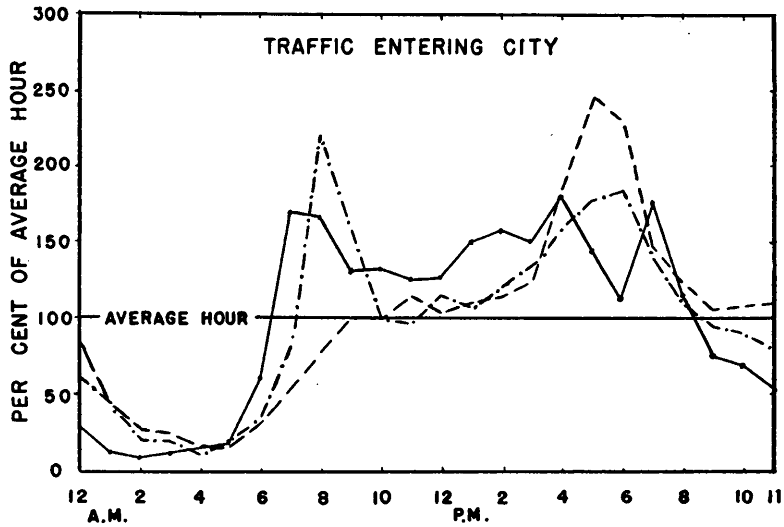
Holland Tunnel, Wednesday May 15, 1935

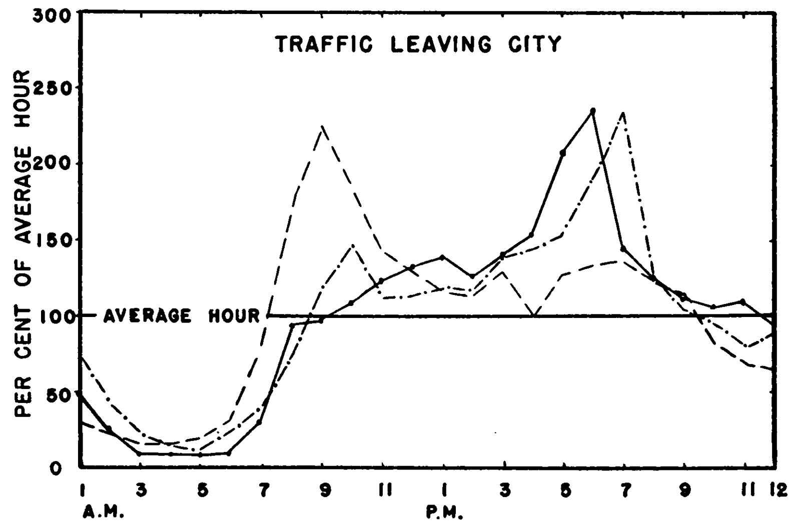
Sumner Tunnel, Wednesday May 15, 1935

Manhattan Bridge, Tuesday June 18, 1935

FIGURE 3a

Hourly vehicular traffic patterns for selected facilities. (Cherniack, 1936, figure 16, p. 265)





Based on traffic as follows:

Posey Tube, August 19–25, 1934

Queensboro Bridge, Tuesday May 28, 1935

Geo. Washington Bridge, Wednesday May 22, 1935

FIGURE 3b

Hourly vehicular traffic patterns for selected facilities. (Cherniack, 1936, figure 16, p. 265) effect of driver reaction to congestion on vehicle spacings and delay? These and a wide variety of other problems arise out of the abilities, limitations and habits of the driving public. In such studies the main problem is to be able to take account of the basic individual differences found by observation or measurement and still to be able to treat the data in such a way as to show any systematic relationships which exist and to draw valid conclusions.

In a report on passing distances observed on the highway, Matson and Forbes [14] used a simple presentation to show both the distributions of distance values and the relation of these to the speeds at which the observations were made. As shown in figure 8 this allowed presentation of mean distances in relation to speed, while

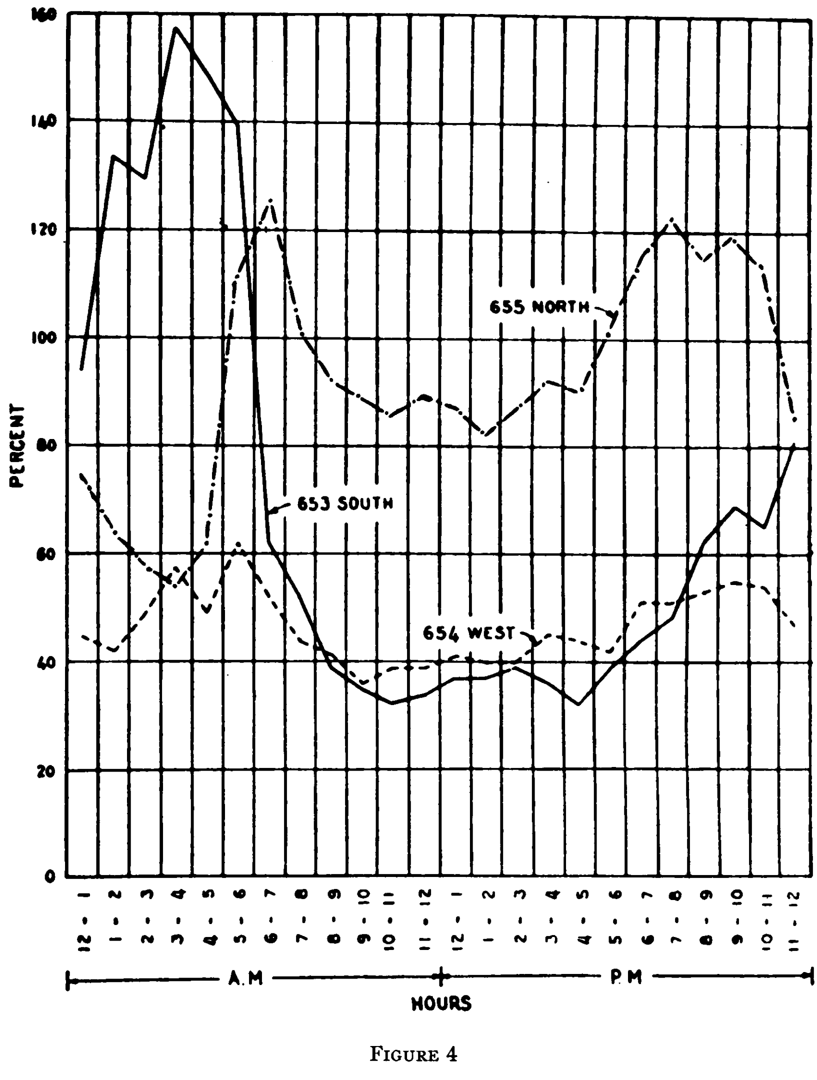


FIGURE 4

Coefficient of variation of highway traffic volume by hours for year 1936–37. (Shelton, 1939, figure 1, p. 347—Michigan data.)

an 80 percentile (or other percentile) line could be shown in case the reader desired to design for a large or smaller proportion of the driving public. This type of presentation shows clearly the number of cases and whether the distributions are smooth and symmetrical or skewed or discontinuous.

In a study of legibility distances of highway destination signs, Forbes and Holmes [6] used a similar presentation of the distances at which sign letters of

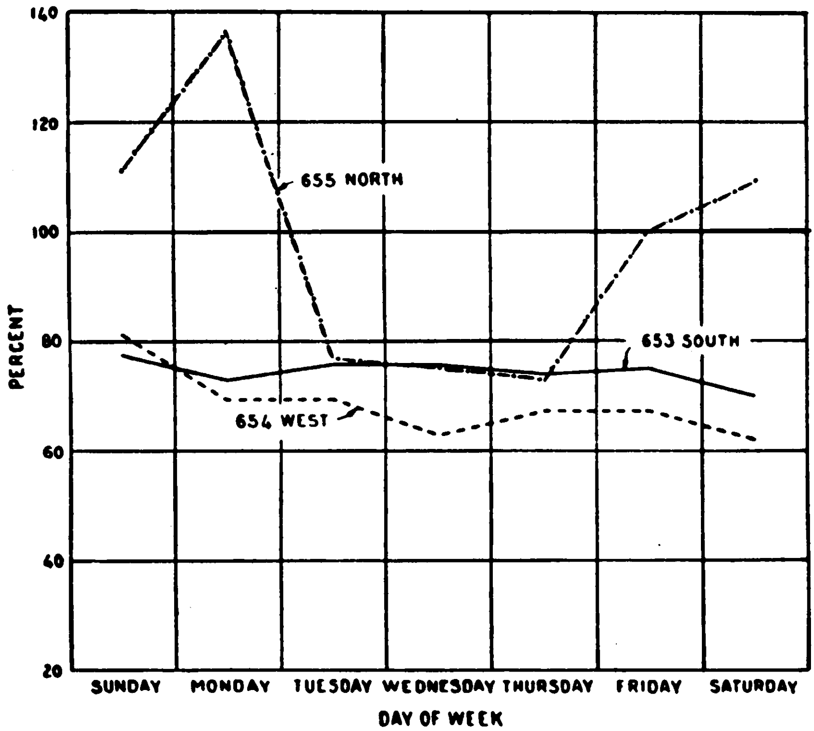


FIGURE 5

Coefficient of variation of highway traffic volume by days of week for year 1936–37. (Shelton, 1936, figure 6, p. 360—Michigan data.)

given sizes could be read by large groups of observers. As shown in figure 9, the lines of relationship for 80 percentile and 95 percentile points as well as for medians were drawn in.

Loutzenheiser [12] presented an analysis of speeds on different sections of highway tangent and determined a speed rating for the highway. He used cumulative percentile distribution curves in presenting his data and utilized the ratio:

$$K = \frac{\text{percentile speed}}{\text{average speed}},$$

using 90, 95 and 98 percentile speeds to compare different stretches of highway having the same average speed. He then made comparisons between two, three and four lane roads, to determine a method of rating highways in terms of a speed percentile.

Cumulative per cent frequency distribution curves have also been used by Normann [19] in presenting data from large scale highway capacity studies and by others. The use of cumulative frequency distributions of course has the advantage that it allows easy smoothing of curves and where the samples are large, as in the case of the studies quoted, it is quite justified. However, it should be pointed out that in some studies where small samples of data are presented in cumulative per cent frequency distributions, unless the actual points observed are plotted and

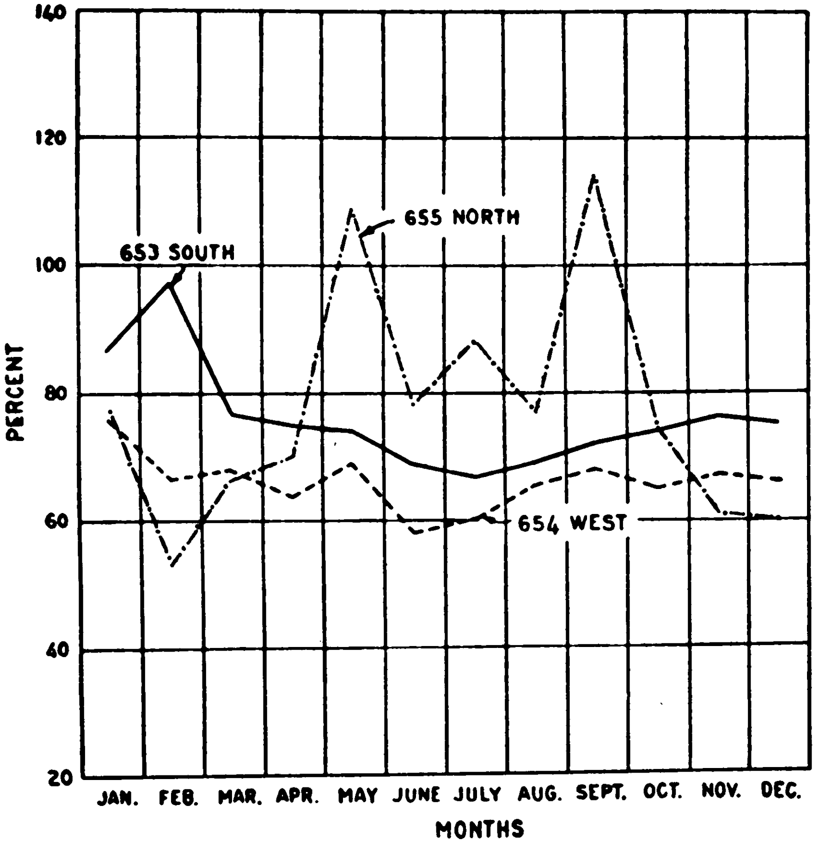


FIGURE 6

Coefficient of variation of highway traffic volume by months for year 1936–37. (Shelton, 1939, figure 9, p. 361—Michigan data.)

number of cases indicated, the use of such curves may lead to erroneous conclusions.

The cumulative frequency distribution allows easy reading of any desired percentile direct from the plotted curve and therefore has advantages in certain studies.

4. Accident records, exposure factors and expectancy indexes for highways

The statistical techniques which have been used in studies of accident records of drivers are many and various. These have included the correlation of accident records with various kinds of test scores and other indexes of driver ability or condition by means of the Pearsonian correlation coefficient. They have also involved comparison of the accident frequency distributions of various groups of drivers with the Poisson distribution. The Poisson exponential function has been shown to

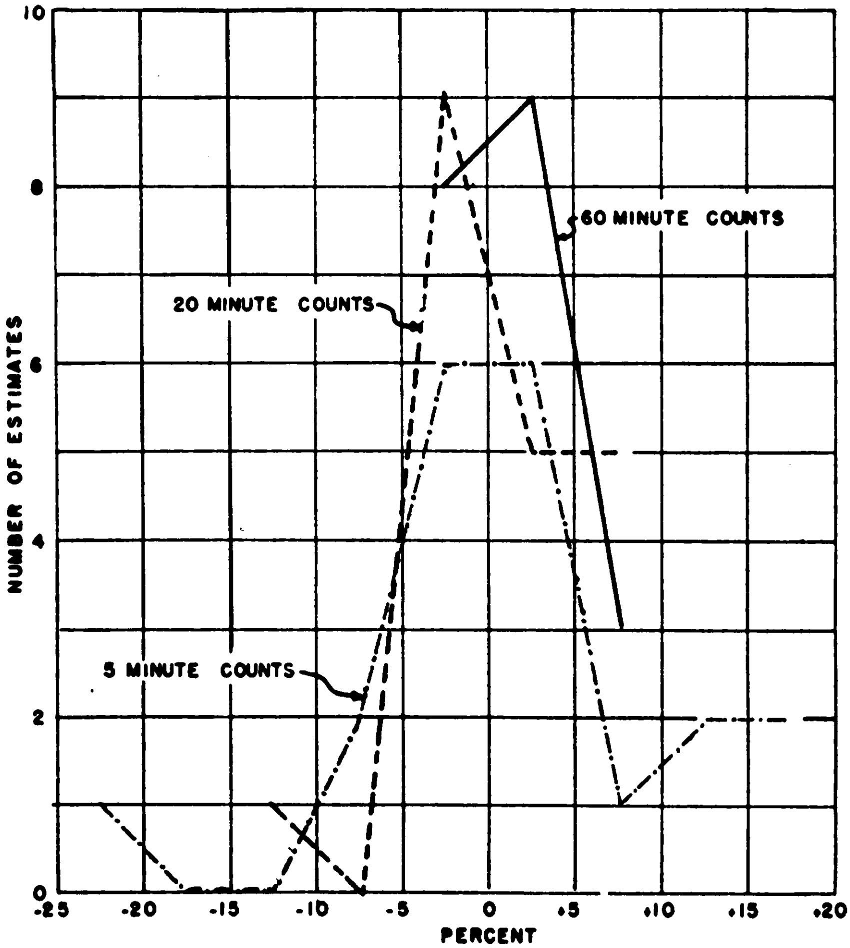


FIGURE 7

Frequency distribution of the percentage deviations from the 24 hour annual mean for 5, 20 and 60 minute counts. (Cunningham, 1945, figure 5, p. 340.)

represent pure chance under conditions where the probability of occurrence is relatively small. For a recent discussion of the Poisson distribution as applied to the frequency of accidents, see Mintz and Blum [15].

A somewhat different approach was used by Kraft and Forbes [11][1] in a study of streetcar operators where an accident "expectancy" index was developed. This was possible due to the fact that the company had careful records not only of the

[1] This analysis was completed in the Fall of 1941 but publication was delayed by events. For a preliminary report, see T. W. Forbes, *Proc. Amer. Psychol. Assn.*, 1941.

accidents themselves but also of the route, time of day, mileage operated and similar other factors for each man. Since the different operators worked on different routes during the 5 years for which the records were available, it was possible to obtain the average hazard factor for each of the streetcar routes as well as for hazards for different phases of the diurnal and seasonal cycles from the city traffic engineers records of traffic. By the use of master cards, gang punching and the use of the automatic multiplying punch of the Hollerith machine sorting and tabulating equipment it was possible to calculate mechanically an expected accident figure for each man. Figure 10 shows the distribution of actual accidents compared with

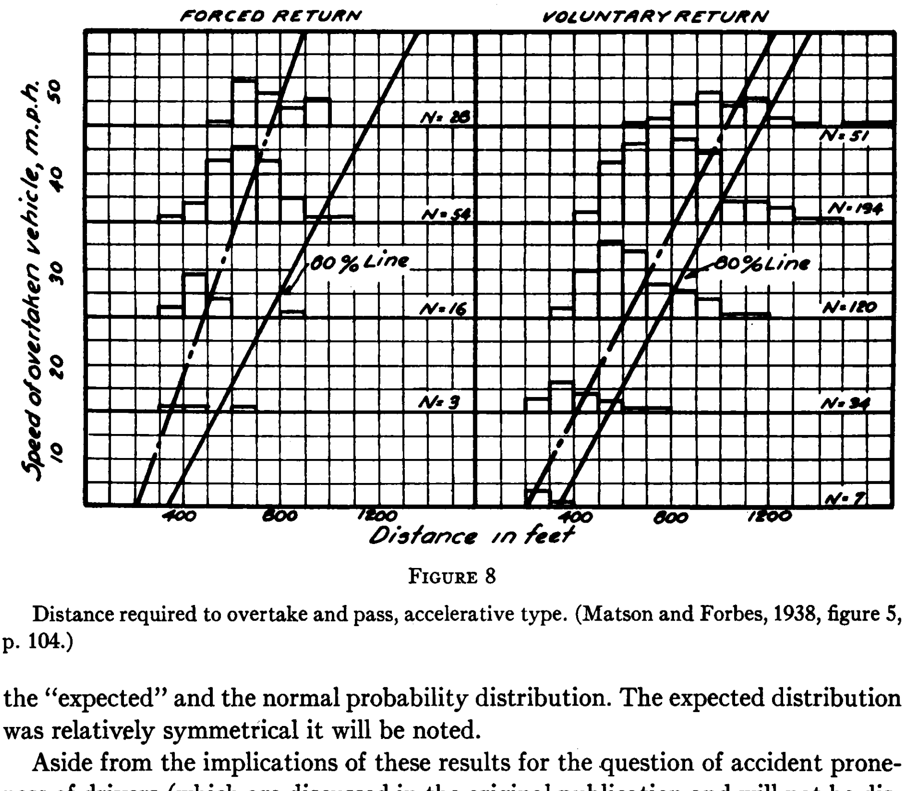


FIGURE 8

Distance required to overtake and pass, accelerative type. (Matson and Forbes, 1938, figure 5, p. 104.)

the "expected" and the normal probability distribution. The expected distribution was relatively symmetrical it will be noted.

Aside from the implications of these results for the question of accident proneness of drivers (which are discussed in the original publication and will not be discussed here) it is of interest to our present discussion that the same type of approach may be applied to an accident analysis of highways, that is, it should be possible to develop an accident "expectancy" figure for highways of different lane widths and for various features of highways, such as tangents, curves and intersections. Such expectancy figures would represent the relative hazard which these characteristics present. Such expectancy figures would evaluate, therefore, these features from a safety standpoint and could also be used to compare different highways with each other.

An attempt to start this in a rough way was made in the report of the Committee on Analysis of Accident Data, American Road Builders Association [2]. Samples of data from 12 different states were collected and analyzed separately.

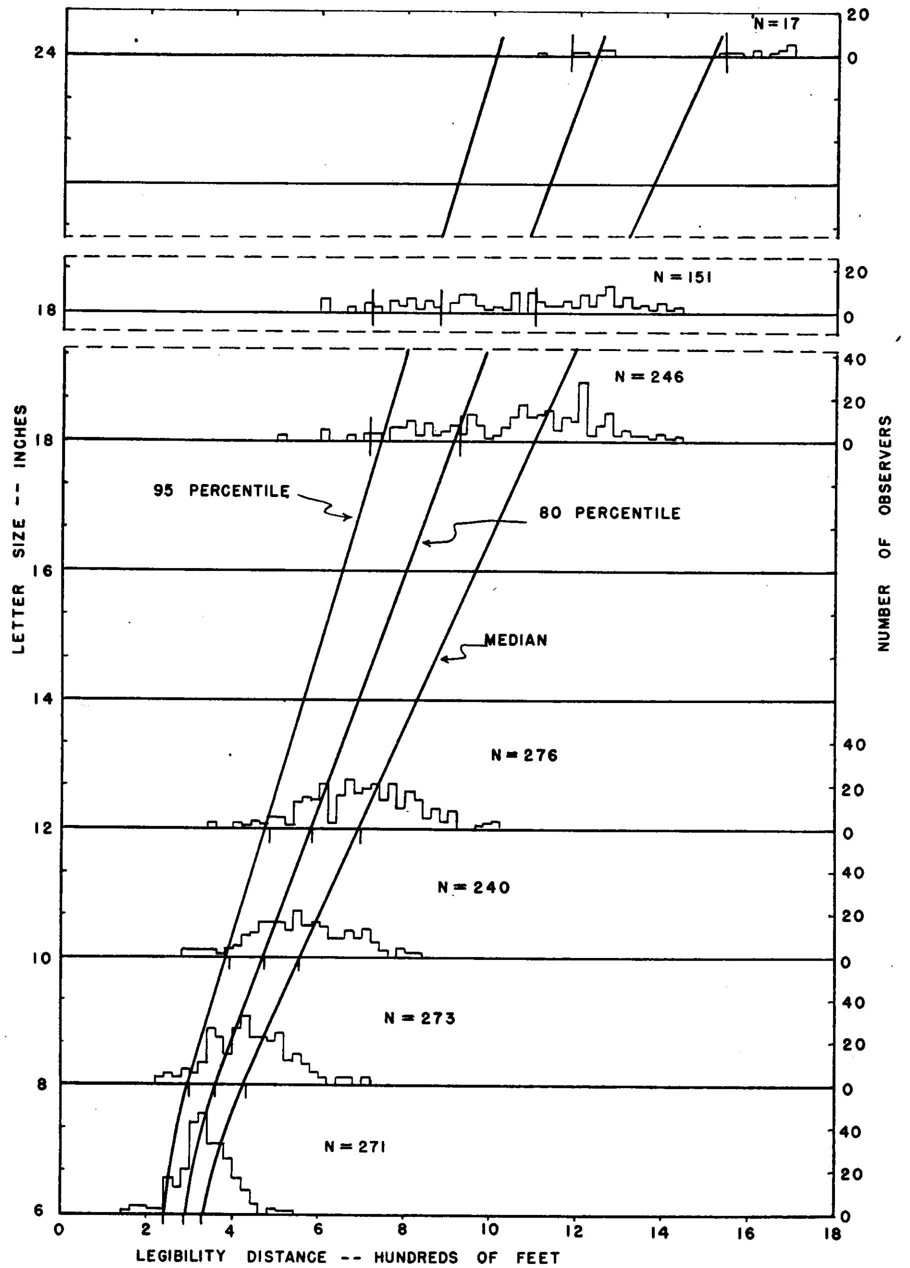


FIGURE 9

Daylight legibility distances, black on white, series D (rectangular) letters, unreflectorized. (Forbes and Holmes, 1939, figure 2, p. 326.)

It was necessary to do this since it was found that the accident rates from different states varied considerably, probably due to the well known difference of completeness of reporting and differences in the interpretation of accident definitions. Analysis was limited to two lane highways since the number of samples of multilane highways was too small to furnish reliable results.

Although coefficients were not computed, a plot of accidents against vehicle miles (in tens of millions) showed fairly good correlation in the data from certain states but much more variability around the trend line in the data from other states (figure 11).

The trend, however, suggested that, as would be assumed, exposure increases in proportion to mileage traveled by the given car and also in proportion to the

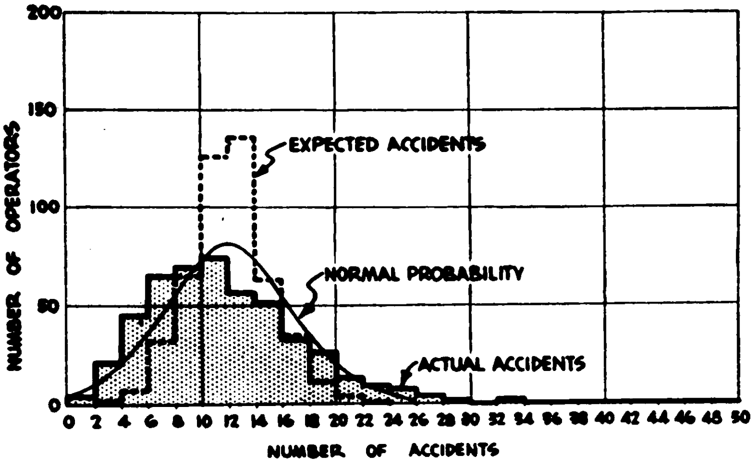


FIGURE 10

Distributions of actual and "expected" accidents of 482 streetcar operators. (Kraft and Forbes, 1944, figure 1, p. 283.)

number of vehicles on the highway. If so, dividing the number of accidents on each stretch of highway by vehicle miles should eliminate the effect of these two factors and result in a horizontal line of relationship. Figures 12 and 13 show such plots to which trend lines have been fitted by inspection. It will be noted that, although the trend is horizontal and a line could be fitted by least squares or average ordinates, the variability of one of the two samples was so great that the error of estimate would be rather large. This scatter seemed to indicate that other factors which differed between different segments of highway must be evaluated.

Thus the scatter of accident rates from different stretches of highway may be of such large relative proportion as to indicate other factors of importance even when a constant relationship has been obtained in a line of average ordinates.[2]

Since that time Baldwin [3] has analyzed accident rate of highway segments by

[2] Jelenik, in an appendix to the ARBA committee report tells of an analysis of Chicago Park District streets into 46 sections of $\frac{1}{2}$ mile each on which a multiple correlation analysis of car miles and intersections per mile was made. Both were found to be related to accidents by the formula $x_1 = -8.6 + 14.3x_2 + 1.6x_3$ where x_1, x_2, and x_3 were accidents, car miles and intersections per mile, respectively.

lane width, tangent, degree of curvature and other factors. He applied a correction factor for each state and then apparently combined data from various parts of the country. By Hollerith punched card technique the large amount of data was classified into categories and accident rates were reported, but no index of variability was reported.

Jorgensen [9] also noted the need for an index of hazard for highway design factors. He reported an analysis of Connecticut highways in terms of traffic

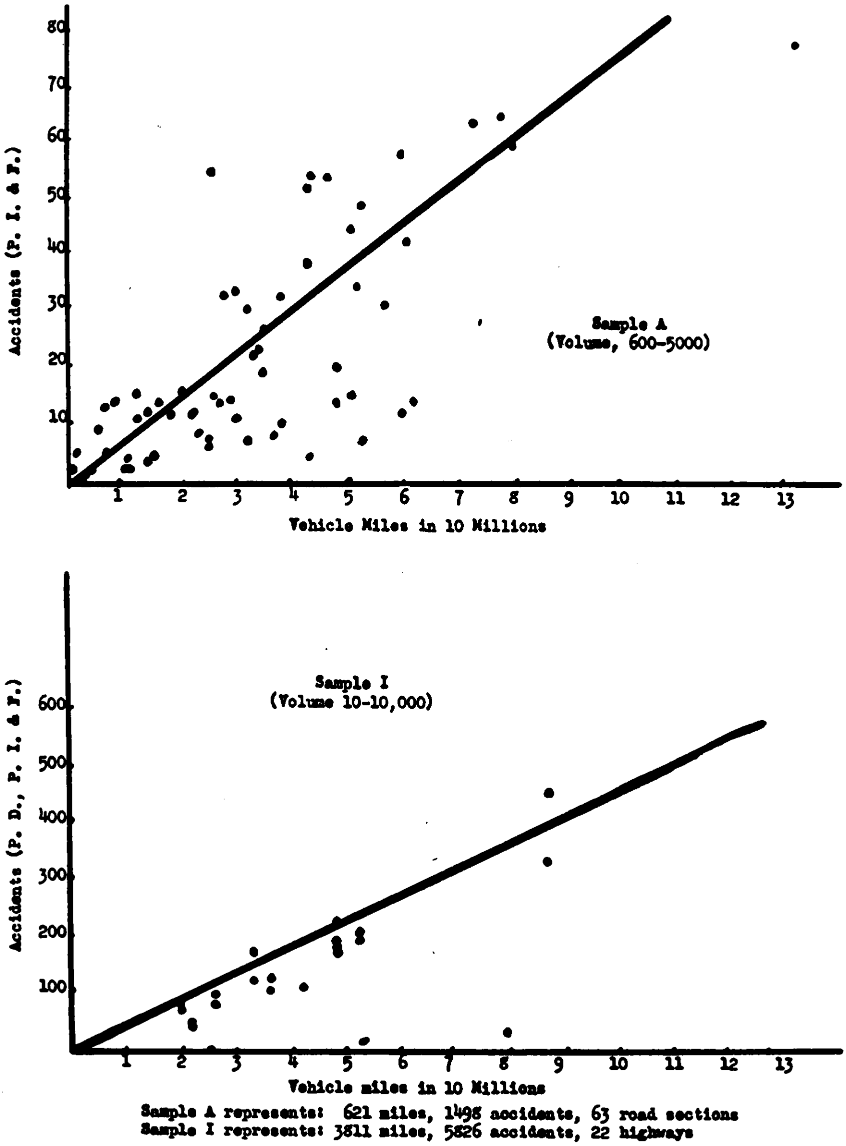


FIGURE 11

Relation of accidents to vehicle miles for samples A and I. Two lane highway segments from two states. (ARBA Committee on Analysis of Accident Data, 1941, figure 3.)

volume, and design standards met, showing a lower accident rate to be related to modern design standards. Again an indication of variability would be enlightening.

It is of interest, also, that the technique of multiple correlation has been used in an attempt to determine the relative importance of car registrations, mileage and population of different counties for accident records of drivers [13]. This suggests another approach which might perhaps be applied to highways. In this case, the zero order correlation coefficients were calculated and the beta weights were derived for the multiple correlation of these factors with the accident record criterion. The factors used were population, registrations and over all estimated vehicle

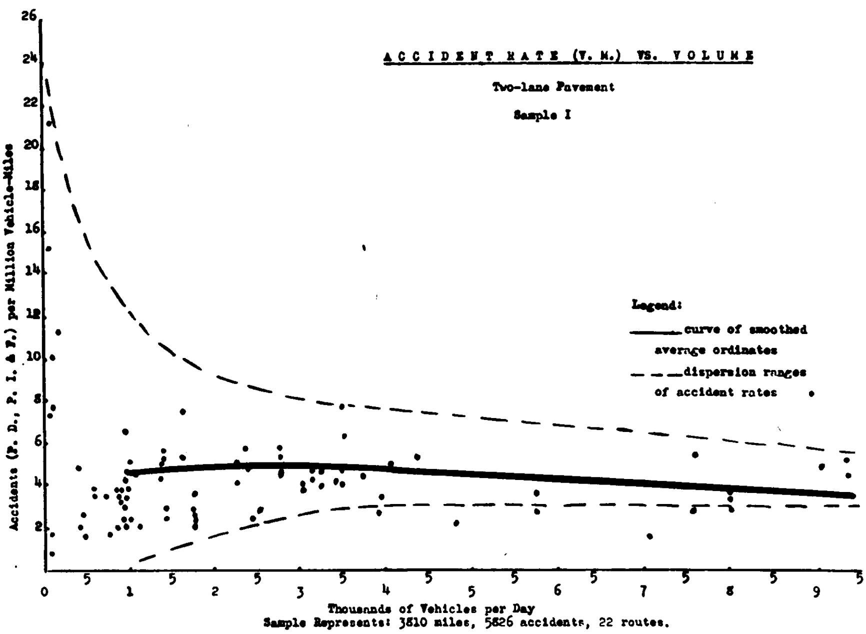


FIGURE 12

Accident rate and volume for sample I. Two lane highway segments. (ARBA Committee on Analysis of Accident Data, 1941, figure 6.)

mileage for counties in Iowa. Records of fatal accidents by counties were used as the criterion. The beta weights were proposed as the best measure of relative importance of the factors.

5. Study of vehicle spacings, delays at intersections and opportunity to cross

Several investigators have attempted to apply the frequency distribution obtained from the Poisson exponential function[3] to the description of traffic volumes and vehicle spacing. Kinzer [10] developed a speed distribution curve on the basis

[3] The Poisson exponential function is the approximate expansion of the probability $(p + q)^n$ when p is small with respect to q. That is, it gives the probability of occurrence of different numbers of events due to "chance" when the over all probability of occurrence is small. (For derivation, see Rietz [22].)

of the Poisson distribution assuming random distribution of vehicles along the length of the highway. No test against actually observed data was reported.

On the other hand Adams [1] made a similar application based on the assumption that equal time intervals are equally likely to contain an equal number of events (in this case vehicles which arrive). On this basis, he applied the Poisson equation

$$P = N e^{-m} \frac{m^x}{x!},$$

where P is the probable number of time periods in which x vehicles arrive, N is the

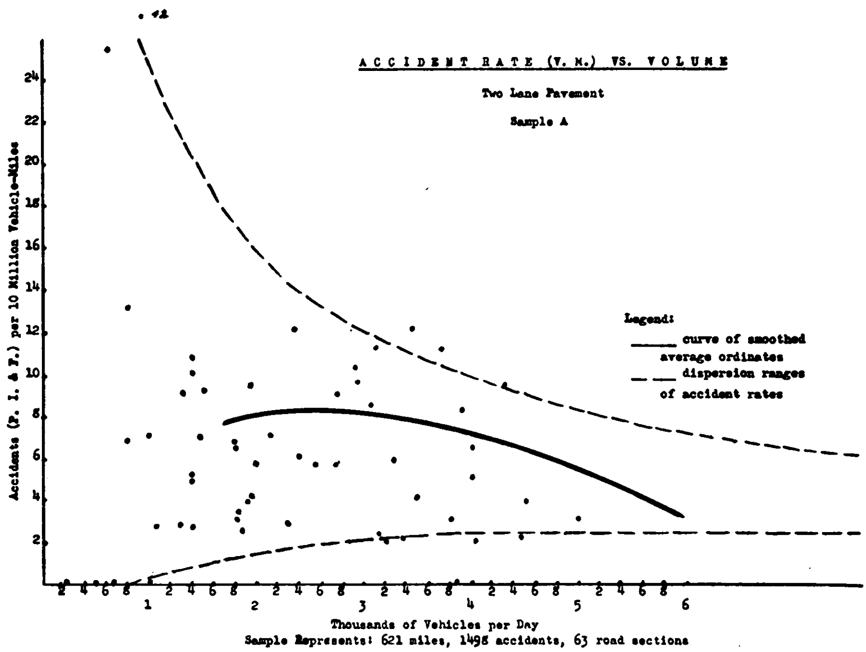


FIGURE 13

Accident rate and volume for sample A. Two lane highway segments. (ARBA Committee on Analysis of Accident Data, 1941, figure 7.)

total number of vehicles and m is the average number of vehicles which arrive during a given time. This distribution was tested against observed data from two streets. Observations were made on "free flowing traffic." A rather good fit was obtained under these conditions but the author pointed out that when traffic volume increased to a certain point, the Poisson distribution no longer fitted, presumably due to the interference of vehicles with each other.

Adams also reported that the distribution of intervals of time between vehicles is given by the expression ae^{-Nt} for the interval between vehicles, where a is the number of intervals and N is the number of vehicles per second. It will be noted that this expression is the first term of the Poisson exponential function, that is, it is the probable number of time intervals in which exactly zero vehicles arrive. This expression also apparently gave good fit when tested against actual observations under the conditions used by Adams. Adams also suggested various applications of these formulas to traffic problems involving such questions as delay and opportunity to cross.

In the United States a further and more detailed application of the Poisson formula to the arrival of vehicles and to spacings was reported by Greenshields, Schapiro and Ericksen [8], also partially reported by Greenshields [7]. Attention

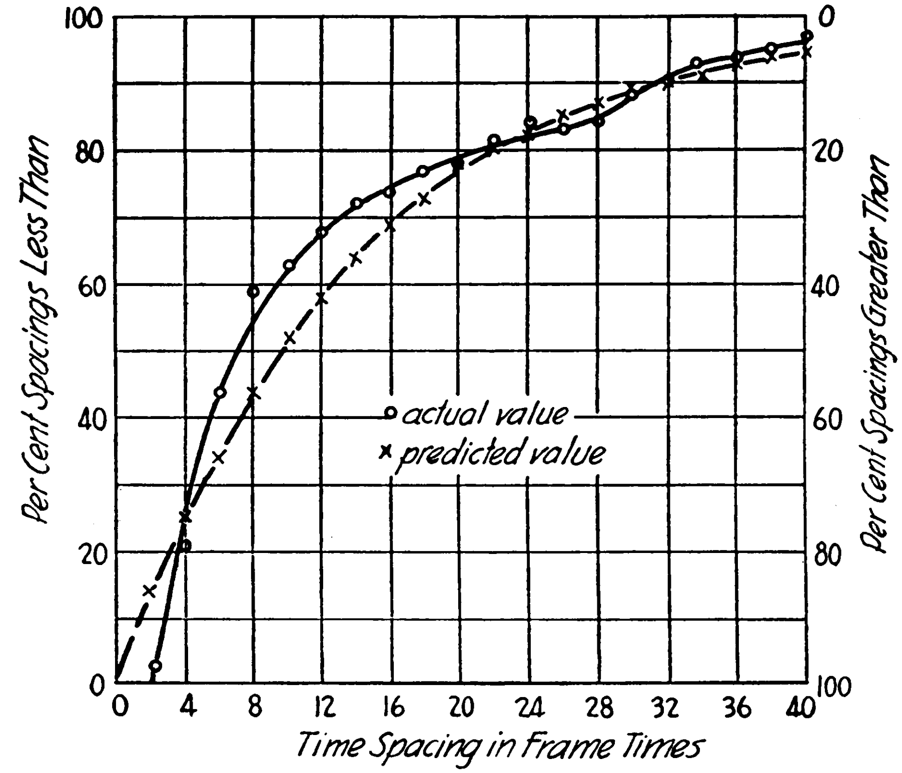


FIGURE 14

Cumulative distribution of time spacings between vehicles. Traffic volume 378 v.p.h. 174 spacings recorded frame time—60/88 second. (Greenshields, Schapiro, Ericksen, 1947, figure 43, p. 82.)

was called to Adams' paper and essentially the same formulas were applied to Normann's data on vehicle spacings in highway traffic and to observations made in city traffic in eastern cities by the authors. A fairly good fit was obtained with Normann's data except for time intervals under two seconds. However, for the three samples of spacings between cars obtained under city conditions the fit was not satisfactory according to the authors. Although no mathematical test of goodness of fit was reported, figure 14 shows an example of the fit obtained. However, by using the formula for the number of vehicles arriving during given intervals (the complete Poisson distribution) a better fit was obtained. This is shown in figure 15. The authors admitted that the fit was not too satisfactory for observed spacings but concluded "in general however the Poisson theory which describes the irregular spacings between vehicles gives results sufficiently accurate to have practical importance." They then proceeded with the development of calculations

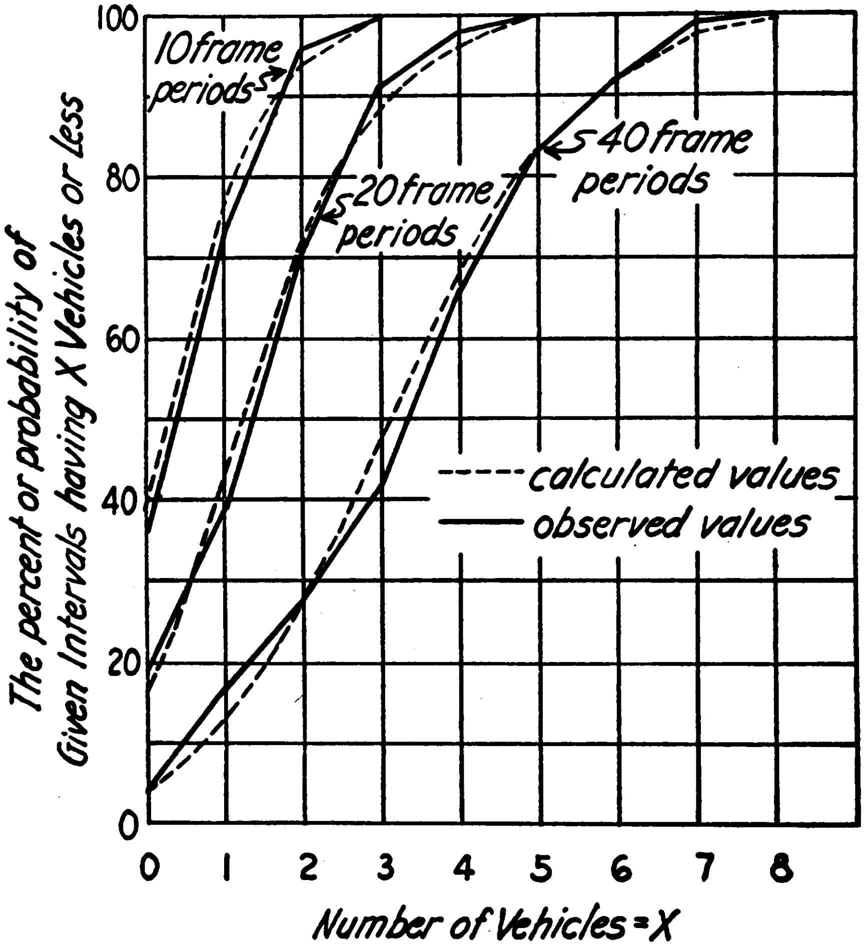


FIGURE 15

Probability that given time intervals will have "X" vehicles or less. 345 vehicles. Traffic volume 490 v.p.h. (Greenshields, Schapiro and Ericksen, 1947, figure 45, p. 84.)

of delay at intersections as an illustration of how application may be made to traffic problems.

The work of Greenshields *et al.* has been used as a basis by a number of others including Raff [21] who reported a formula for the probability of delay occurring to cars approaching a stop street from a side street. Observed data from one sample of data for a given volume of traffic on a side street was presented together with the curve computed from the formula. Although the fit seems reasonably good no mathematical goodness of fit test apparently was made and it is not clear whether or not the plotted data are those from which one of the constants of the equation was derived.

Since the use of such a function as the Poisson can contribute to the ease of solution of many problems, it would seem desirable to have a clearer demonstration of its applicability. More goodness of fit tests using both data from free flowing open highway conditions and from more congested conditions on both highways and in cities should be made for this purpose. It will be noted that the application to the estimation of delay at intersections is one which involves a considerable degree of city congestion.

6. "Before and after" studies

A well known technique for evaluating the effectiveness of changes in design, regulations, signing, street and highway lighting, and other conditions is the "before and after" comparison. Most such studies involve either accident records, speeds of vehicles, the time for drivers to react or similar measures. In a great many cases these comparisons have been made without any statistical test to determine whether the differences obtained could have been due to chance factors.

In this type of problem, we ask of our statistics the question, "What is the probability of the measured difference (before and after the change of conditions) being the result of chance factors?" In answering this question we find two general cases: (1) the case in which we must deal with the average change in the behavior of two different groups of vehicles and drivers and (2) the case in which we find a difference in the behavior of each individual of the same group of vehicles and drivers which are submitted to the two conditions of operation.

In the first category would fall the familiar (before and after) accident record study of intersection redesign. Answering the question as to chance occurrence in such studies is a difficult one since typically only a very small number of accidents is involved either "before" or "after." It is our suggestion that since it has been shown elsewhere that accidents may occur in Poisson fashion, this function might be used as a basis of estimating the probability of occurrence. It would first be necessary to obtain sufficient data on intersections of certain classifications to test whether the Poisson function applies.

Another type of problem is one in which a change of conditions is introduced and it is desired to test the effects of this on, let us say, the speed of vehicles. Ordinarily the change is one which may take some time and therefore it is necessary to compare the speeds of vehicles on this highway some days or weeks preceding the change with that of some days or weeks after the change. In the simplest case, observations may be made before the change on a group of vehicles for a given hour on a certain day of the week and compared with a set of observations on another group of vehicles for the same time and day of the week after the change. Average speeds may then be computed for the two groups and the reliability of the difference between the means tested by the well known ratio $(M_1 - M_2)/\sigma$, where M_1 and M_2 are the means in question and σ is the standard error of the difference.

Where the two samples are small, Fisher's t-function can be used to estimate the probability that the difference obtained is due to chance in terms of an estimate of the population variance. (For a simple discussion of these techniques see Peters and Van Voorhis [20, chaps. 5-6].)

It should be noted that in the use of the formula for the standard error of the difference there is a term which involves the correlation between the two variables. Where two different sets of vehicles are involved, it would be assumed that the two sets of measurements are uncorrelated causing this term in the formula to vanish. This will not be so in case 2, however.

The second case is one in which the same group of individual vehicles and drivers is exposed to the two different conditions. In this case we know to start with that there will be characteristic variations between individuals which we would like to eliminate. The question is whether the effect of the experimental condition on each individual is great enough to be "real," that is, not due to chance variation. Here then, we desire to use "the individual as his own control." The technique which tests the obtained difference must do so in terms of the individuals themselves. Or if the whole sample is used, it must take account of the correlation which exists due to the fact that the individuals are the same ones.

As an example, it might be desired to measure the effect of two different types of pavement on succeeding stretches of the same highway on the speeds of vehicles running over that highway. If speed determinations are made on the same vehicles and if they can be identified, it will be possible to obtain speed differences for each vehicle. We may then obtain a distribution of such velocity differences, the mean of these differences, and standard error of this mean by the usual formulas. Or we may use the difference of the means technique using a formula which takes into account the correlation between the variables.

It was suggested above that the observations be made at the same time of day and at the same day of the week. This example was used in order to eliminate the necessity of treating the diurnal and weekly traffic volume cycles which were treated in the first part of the paper and which are now so well known. However, it may be desirable to make comparisons at different times of day in order to see whether or not any obtained differences are found in heavy as well as light traffic conditions. In other words, to make the most general types of comparison different phases of the traffic cycle should be sampled. In order to test the differences obtained from such samples it may be necessary to make comparisons for three or more times a day on each week day. Average speeds may be computed for each time of day for each day of the week and mean differences between the "before and after" series of values obtained. Reliability of the differences may then be tested by the method of variance analysis which gives an estimate of the probability of their being due to chance. By this method, the proportion of the variance between days and between times of day may be obtained, and compared with the remainder. (For a recent discussion of variance analysis of this sort, see McNemar [17, chap. 14].)

It should also be noted that answering the question as to the probability of the obtained difference being due to chance does not tell *why* the difference is obtained. If the experiment is not carefully set up it is quite possible to obtain a difference of means which passes the usual test of reliability but at the same time which may arise from some other change of conditions than the one which we thought we were measuring. The use of statistical comparison therefore can never make up for poor experimental logic nor substitute for good experimental planning. And for routine traffic engineering procedures, differences measured must be great enough to be of practical importance first of all.

7. Summary

A number of studies from the highway traffic engineering field were cited to illustrate the following points: Considerable attention has been given to sampling problems and obtaining the best estimate of certain group measures in connection with traffic surveys. Other problems require the use of techniques for the comparison of means of uncorrelated scores, correlated scores or individual differences. Expectancy figures for accident rates of highway segments have been attempted but variability does not seem to have been taken into sufficient account. The Poisson exponential function has been applied by a number of investigators both to accident distributions and to volume, spacing and delay problems, although more satisfactory tests of goodness of fit are desirable.

It was suggested that further application could be made of such techniques as expectancies from the Poisson function and analysis of variance as well as wider use of statistical tests of the reliability of differences and goodness of fit. Applications were suggested to "before and after" accident, speed and similar studies.

Although for routine traffic engineering procedure any differences measured in such studies must, first of all, be great enough to be of practical importance, tests of statistical reliability are of importance thereafter. And for research purposes such reliability determinations are of primary importance for sound results.

REFERENCES

[1] W. F. Adams, "Road traffic considered as a random series," *Jour. Inst. Civil Engrs.*, Vol. 4 (1936), pp. 121–130.

[2] American Road Builders Association, *Report of Committee on Analysis of Accident Records*, mimeographed. Also *Proc. A. R. B. Assn.* (1941), pp. 161–182.

[3] D. M. Baldwin, *The Relation of Highway Design to Traffic Accident Experience*, mimeographed, presented to Committee of Planning and Traffic Engineering, American Association of State Highway Officials, Dec. 1946.

[4] H. E. Cunningham, "Analysis of short count methods in measuring traffic flow," *Proc. Hwy. Res. Bd.*, Vol. 25 (1945), pp. 329–342.

[5] N. Cherniack, "Methods of estimating vehicular traffic volume with the aid of traffic patterns," *Proc. Hwy. Res. Bd.*, Vol. 16 (1936), pp. 253–272.

[6] T. W. Forbes and R. S. Holmes, "Legibility distances of highway destination signs in relation to letter heights, letter width and reflectorization," *Proc. Hwy. Res. Bd.*, Vol. 19 (1939), pp. 321–335.

[7] B. D. Greenshields, "Application of statistical sampling methods to traffic performance at urban intersections," *Proc. Hwy. Res. Bd.*, Vol. 26 (1946), pp. 377–389.

[8] B. D. Greenshields, D. Schapiro and E. L. Ericksen, "Traffic performance at urban intersections," *Technical Report Bureau of Hwy. Traffic, Yale University*, No. 1 (1947), pp. 1–152.

[9] R. E. Jorgensen, *Accident Analyses for Program Planning*, mimeographed, presented at Hwy. Res. Bd., 1949.

[10] J. P. Kinzer, "Application of the theory of probability to problems of highway traffic," a digest by Lloyd F. Rader, *Proc. Inst. Traf. Engrs.* (1934), pp. 118–124.

[11] M. A. KRAFT and T. W. FORBES, "Evaluating the influences of personal characteristics on the traffic accident experience of transit operators," *Proc. Hwy. Res. Bd.*, Vol. 24 (1944), pp. 278–291.

[12] D. W. LOUTZENHEISER, "Percentile speeds on existing highways," *Proc. Hwy. Res. Bd.*, Vol. 20 (1940), pp. 372–389.

[13] G. I. LAUER, MAX CUTLER and A. P. LAUER, "Exposure risks as a criterion of traffic accident hazards in Iowa," *Proc. Iowa Acad. Sci.*, Vol. 52 (1945), pp. 261–263.

[14] T. M. MATSON and T. W. FORBES, "Overtaking and passing requirements as determined from a moving vehicle," *Proc. Hwy. Res. Bd.*, Vol. 18 (1938), pp. 100–112.

[15] ALEXANDER MINTZ and M. L. BLUM, "A re-examination of the accident proneness concept," *Jour. Applied Psychology*, Vol. 33 (1949), pp. 195–211.

[16] MILLER MCCLINTOCK, "Short count traffic surveys and their application to traffic design," *Portland Cement Association*, revised, Chicago (1935). First edition about 1934.

[17] QUINN MCNEMAR, *Psychological Statistics*, Wiley, New York, 1949.

[18] J. NEYMAN, "Contribution to the theory of sampling human populations," *Jour. Amer. Stat. Soc.*, Vol. 33 (1938), pp. 101–116.

[19] O. K. NORMANN, "Preliminary results of highway capacity studies," *Public Roads*, Vol. 19 (1939), pp. 225–232, 240–241.

[20] C. C. PETERS and W. R. VAN VOORHIS, *Statistical Procedures and Their Mathematical Bases*, McGraw-Hill, New York, 1940.

[21] M. S. RAFF, "Space-time relationships at 'stop' intersections," *Proc. Inst. Traf. Engrs.* (1949), pp. 42–49.

[22] H. L. RIETZ, *Mathematical Statistics*, Banta, Wisconsin, 1936.

[23] R. O. SWAIN, "A practical application of observing traffic by means of extremely short counts," mimeographed, *State Hwy. Department, Austin, Texas* (1939); abstract in *Traffic Engineering*, Vol. 9, No. 18 (1939).

[24] W. A. SHELTON, "Discussion of 'Toll Bridge Traffic Patterns,'" *Proc. Hwy. Res. Bd.*, Vol. 14 (1934), pp. 399–409.

[25] ———, "Methods of estimating highway traffic volume," *Proc. Hwy. Res. Bd.*, Vol. 16 (1936), pp. 239–252.

[26] ———, "Dispersion of highway traffic volume by time periods," *Proc. Hwy. Res. Bd.*, Vol. 19 (1939), pp. 347–361.

[27] C. W. VICKERY, "On estimating street traffic by means of extremely short counts," mimeographed, *State Hwy. Department, Austin, Texas* (1939).

[28] ———, "Improvement of methods used in the highway planning surveys," *Proc. Hwy. Res. Bd.*, Vol. 21 (1941), pp. 88–109.