DISTRIBUTION OF VEHICLE SPEEDS AND TRAVEL TIMES

DONALD S. BERRY AND DANIEL M. BELMONT
UNIVERSITY OF CALIFORNIA

1. Summary

Data on speeds and travel times of motor vehicles are essential to traffic engineers responsible for design and operation of streets and highways. The usefulness of the data, however, is related to procedures followed in assembly and analysis.

This paper outlines different methods for analyzing distributions of vehicle speeds and travel times, investigates the relationship between speeds and travel times, sets forth applications to preliminary data, suggests which techniques of analysis are best suited to the requirements of the engineers, and points out the need of further study.

The Highway Research Board of the National Research Council is now organizing a Committee on Speed Characteristics to assemble and analyze data on motor vehicle speeds under different physical, traffic, and environmental conditions. This paper is intended as a preliminary study to assist the new Committee in its planning.

2. Definitions

Spot speed—A spot speed is the speed, in miles per hour, of a vehicle as it passes a given location on a street or highway.

Travel time—The total time required to traverse a given distance, including all traffic stops and delays.

Running time—The total time required to traverse a given distance, excluding the stopped time.

Over-all speed—The total distance traversed, divided by the total travel time, expressed in miles per hour.

Running speed—The total distance traversed, divided by the running time, expressed in miles per hour.

Design speed—The highest continuous speed at which individual vehicles can travel with safety upon a highway when weather conditions are favorable, traffic density is low, and the highway design features are the governing conditions.

10 *mph pace*—The 10 mph speed range containing the largest percentage of the vehicles, in a distribution of spot speeds at a location.

3. Variables

Spot speeds and travel times of motor vehicles may vary because of different *physical* factors (curvature, grade, sight distance, frequency of intersections, and roadside development), different *traffic* factors (volume of traffic, percent of through traffic, turning traffic, proportion of trucks, parking conditions, and volume of pedestrians), and different *environmental* factors (section of the country, driver characteristics, weather, season, visibility, enforcement practices, speed limits, and other traffic controls).

The speed characteristics at any one location may change from time to time because of the effect of one or more factors such as changes in traffic volumes, weather, or visibility. Speeds may also change as the result of changes or a combination of changes in speed limits, parking regulations, enforcement, or other traffic control measures.

4. Spot speeds

Spot speeds are usually obtained by timing vehicles over a short test course [1]. Speeds of every vehicle are usually taken during the period of the study, except when using some types of timing equipment at times of heavy traffic volume.

Spot speed data are useful to highway and traffic engineers in selecting speed limits for speed zones [2], evaluating the effectiveness of changes in physical and traffic conditions, and in determining whether prevailing speeds are too fast for conditions. The data are used also in timing traffic signals, locating warning signs, and selecting the appropriate design speed for new highways [3].

Engineers strive to provide highways and traffic control measures which, for similar traffic volumes, result in the most compact speed distribution. The less the dispersion, the more nearly uniform the speeds and the fewer are the number of hazardous passing maneuvers.

The three portions of a speed distribution curve of primary interest to traffic officials are the central portion (between the 85th and 15th percentiles), the high speed group (above the 85th percentile), and the slow group (below the 15th percentile).

The central portion contains the main group of drivers who usually conform voluntarily to traffic regulations. Maximum speed limits for speed zones are usually set after taking into consideration the 85th percentile speed and other factors such as tolerance in enforcement.

Engineers and enforcement officials consider that the 15 percent of the motorists exceeding the 85th percentile speed is the group which will not conform voluntarily to reasonable control measures and which needs enforcement action.

The 85th percentile speed is also used in making computations of needed sight distance for the prevailing speeds at blind intersections. Furthermore, the timing of traffic signals along a street is fixed to accommodate the central portion of the speed distribution curve.

The 85th–15th percentile range, $P_{85} - P_{15}$, approximates the reasonably straight, steep portion of the cumulative speed distribution curve as plotted on coordinate paper. Above and below this range the curve tends to decrease rapidly in slope.

The lower portion of the curve (below the 15th percentile speed) represents vehicles which may obstruct movement when there are few opportunities for passing because of heavy traffic volume or restricted sight distance. The precise distribution of this slower group is seldom of interest to engineers, except perhaps in the special case of truck speeds on grades.

The upper portion of the curve is of interest in the selection of the appropriate design speed for a new highway. Curvature, superelevation, and minimum non-passing sight distance are designed for a uniform maximum speed. Design engineers sometimes refer to the 98th percentile [3] as the desirable definition point for the design speed, but it is probable that knowledge of the speed distribution up to the 90th or 95th percentile is sufficient for design purposes.

5. Analysis of spot speed distributions

In the analysis of spot speed distributions, the method used should (1) require a minimum of computation, (2) reveal clearly the important features of the distributions, and (3) provide comparisons easily understood by persons who may not be familiar with statistical terms.

This third requirement is especially important in speed control, since most officials and citizens are also motorists, and have their *own* opinions on proposed changes in speed regulations. Traffic engineers thus must be prepared to support their recommendations by presenting statistical data that can be easily understood by city councilmen, and interested citizens.

Current Method—Usual analysis procedure used by engineers [2] includes the following steps:

(a) Compute the arithmetic mean,

(b) Identify the 85% speed, and the percent of vehicles exceeding the speed limit (or 5 mph above limit),

(c) Identify the 10 mph pace, and the percent of vehicles traveling within the pace.

(These steps yield columns 5, 9 and 11 of table I.)

The percent of vehicles within the 10 mph pace is used as a measure of the dispersion. Distributions with a high percentage of vehicles within the pace are preferred. The percent within the pace may be determined easily from the array of data—no plotting of curves is necessary. The resulting figures are readily understood by laymen.

The 10 mph pace, however, is suited best to distributions of rural highway speeds. In urban areas, speeds are lower, and the 10 mph range usually covers too great a part of the curve.

Suggested Technique for Engineers—(a) Identify the 85th, 50th and 15th percentile values from an array of the data (or from a graph of the cumulated percentages). (b) Determine the 85th–15th percentile range.

The median (50th percentile) is as good a central measure as is the mean speed, and easier to obtain. For certain economic applications, the mean should be used, but for most purposes the median has at least as good theoretical justification. Actually, the difference between mean and median (in speed distributions) tends to be very small indeed, as may be seen in table I, columns 5 and 6.

The range $P_{85} - P_{15}$ is the actual speed in miles per hour, between the extreme speeds of the central 70% group. It is a measure easy to understand, and perhaps even more suitable than is the pace for use in reports to the city council or the

TABLE I

ANALYSIS OF SPOT SPEED DISTRIBUTIONS

Case No.	Location and Conditions	1	2	3	4	5	6	7	8	9	10	11
		$R^*/F\sigma$ est				Mean mph	Median mph	σ est mph	$P_{85}-P_{15}$ mph	P_{85} mph	σ est / Median	% in 10 mph Pace
		93–7	85–15	70–30	93–50							
1	Typical high speed two-lane rural highway [6]											
a	Lowest traffic volume	1	1.00	0.94	0.97	48.5	47.6	8.14	16.8	55.6	0.17	47
b	1800 vehicles per hour	1	0.97	0.85	0.98	31.8	31.8	2.34	4.7	34.0	0.07	96
2	Two-lane rural highway (eight stations in Washington)	1	0.96	0.91	0.91	48.7	49.1	6.95	13.9	55.6	0.14	56
3	Four-lane divided highway (Calif.)	1	0.95	0.91	0.95	50.4	50.6	6.78	13.4	57.1	0.13	59
4	Two-lane rural highway. (Missouri) [2]											
a	Day	1	0.98	0.97	0.94	53.0	52.4	9.66	19.5	62.0	0.18	44
b	Night	1	0.96	1.02	0.94	49.5	49.4	9.90	19.7	58.3	0.20	39
5	40 mph zone (Erie Co., N.Y.) [2]											
a	Before zoning	1	1.00	0.94	1.12	(38)	38.1	8.54	17.6	47.9	0.22	48
b	After zoning	1	1.00	0.98	0.86	(38)	37.8	5.08	10.5	42.4	0.13	70
6	35 mph rural curve (Calif.)											
a	400′ ahead of curve	1	1.05	1.15	1.07	36.0	35.9	5.12	11.1	41.3	0.14	64
b	At center of curve	1	0.93	0.95	1.00	26.9	27.1	4.41	8.5	31.2	0.16	75
7	Two-lane suburban street (Colusa Ave., El Cerrito, Calif.)	1	1.01	0.88	1.11	30.2	29.9	4.00	8.4	34.6	0.13	79
8	Two-lane through street (Fulton St., Berkeley, Calif., 50% free flowing)	1	1.03	1.02	1.11	23.5	23.3	3.46	7.4	27.4	0.15	85
9	Congested two-lane rural highway [3]	1	0.88	0.87	1.21	28.9	28.7	4.91	9.0	34.0	0.17	73
10	Four-lane through street (Telegraph Ave., Berkeley, Calif., 90% free flowing)	1	0.99	0.93	1.01	25.9	25.8	4.00	8.2	29.9	0.16	80

* Ratio of indicated percentile range to the corresponding range of the normal distribution having the observed $P_{93}-P_{07}$ range. Data in order of increasing speeds. σ est $= (P_{93}-P_{07})/2.95$.

police department, or for release to the press. As a measure of dispersion it has the advantage over the pace of suggesting immediately the total spread of the speed distribution. It states explicitly the spread around the median of 70% of the speeds, and may be expanded in usefulness by the following rule of thumb: 95% of the speeds lie within the median $\pm(P_{85} - P_{15})$, (since $P_{85} - P_{15}$ is roughly twice the standard deviation σ; compare columns 7 and 8 of table I). It thus affords a fair estimate of σ for use in entering formulas for computing needed size of sample, etc.

The above technique gives a picture of the speed distribution sufficient for the usual engineering purposes. A more complete picture, lending itself to more careful analysis, is provided by the following proposed method.

6. Percentile method of analysis

The dispersion and skewness of speed distributions can be determined quickly utilizing a percentile method of estimation. Steps in this method are as follows:

(1) Plot the cumulative frequency curve, speed vs. percent, preferably on probability paper.

(2) Estimate standard deviation σ by computing

$$\frac{P_{93} - P_{07}}{2.95}.$$

(3) List the median.

(4) Complete following table:

Percentile Points	93–7	85–15	70–30	93–50
Normal Deviate $=F$	2.95	2.07	1.05	1.48
Observed Percentiles				
Difference of Percentiles $=R$				
$R/(F\sigma)$	1			

Values of the indicated percentiles are noted from the cumulative frequency curve or an array of data, and their differences are recorded as R. The factor F is the distance between the indicated percentiles in the standard normal curve, $N(x; 0, 1)$, so that everywhere $R/F\sigma = 1$ for normally distributed data. Thus, $R/F\sigma$ is the ratio of the observed range R to that of the corresponding percentile points of a normal curve having the same standard deviation. Values of $R/F\sigma < 1$ in the 70–30 column indicate that the central 40% of the observed data are more compact than in the normal curve, that is, that the observed curve is leptokurtic. In the 93–50 column, $R/F\sigma$ indicates skewness about the median. Values greater than 1 mean that there is positive skewness (toward the higher speeds).

The table serves as a quick means of comparing the shapes of observed distributions with the normal curve and with each other. The method requires a sample big enough to yield a curve reasonably smooth in the regions of the required percentiles. One or two hundred observations usually suffice.

It will be observed that the method of estimating σ, and the use of the median, cut computing labor almost to the vanishing point. (Raw data as frequently gathered are already arrayed.)

The estimation of σ as $(P_{93} - P_{07})/2.95$ has been discussed by Mosteller [5, p. 392] for samples of a normal population. It may often safely be used also for highly abnormal distributions, if minimum error in the estimate is not essential. It is easy to describe a distribution in which the error would be colossal, but it may be worth noting that for the rectangular, semicircular, and even right triangular distribution, σ lies between $(P_{93} - P_{07})/2.95$ and $(P_{94} - P_{06})/3.11$. Perhaps, an excellent all-around estimate is afforded by $(P_{93.3} - P_{06.7})/3.00$.

The method of analysis uses (and describes) only that portion of an observed distribution which lies between the 7th and the 93rd percentiles. For many engineering purposes, this central bulk of the data is of more importance than are the tails. In such cases, the above estimate of σ may be a more valuable measure of dispersion than is the true standard deviation (and the median may be a better central measure than is the mean). Another advantage in restricting the fundamental percentile range lies in the fact that many available curves are smooth and well defined within this range but become ragged beyond it. With the best technique, to obtain an estimate of the 98th percentile requires a 50% larger sample than is needed for an equally good estimate of the 93rd percentile (see Sample size below).

7. Applications to preliminary data

Table I, columns 1 to 7, shows the application of the above method in the analysis of spot speed distributions for several typical conditions on urban, suburban, and rural highways.

Values of $R/F\sigma$ are generally close to 1 which suggests that the speed distributions are roughly normal. This conclusion has been reached by Greenshields [3] and is supported by the generally linear graph found on plotting on probability paper, and by a reasonable value of P in applying the χ^2 test. There is a tendency for the distribution to be peaked in the center (column 3) and to display some skewness (column 4). These deviations from normality are not serious enough to invalidate many applications of normal theory. For example, $\bar{X} + 1.037\sigma$ (the normal value of the 85th percentile) agreed with the observed 85th percentile within 0.1σ (or about 0.6 mph) in 15 cases out of 16 examined.

Where speed is moderately limited by traffic volume, curves, or speed limits, the spot speed distribution remains roughly normal. The principal differences from open road conditions are of course lower average speed and smaller dispersion. There may be also a significant increase in concentration near the center, as in cases 1b and 6b. Columns 4 and 5 suggest that distributions of high mean speed are skewed toward the high speeds. (This tendency would probably be more marked if more of the tail were included in column 4.)

Effect of Traffic Volume—The U.S. Bureau of Public Roads [6] reports that there is a linear decrease of mean speed as traffic volume increases, when other conditions are equal and the critical traffic density is not exceeded. Further analyses of their data reveal that the standard deviation σ of the speed distributions also decreases approximately linearly, with increases in volume. The coefficient of variation $\sigma/\bar{X}$ decreases along a mild hyperbolic arc.

When volume exceeds the practical road capacity, the speed distribution may become so heavily skewed toward the higher speeds that all semblance of normality is lost. Case 9 of table I is a good example.

Since traffic volume so materially affects speed distributions, it is apparent that studies of the effects of speed control measures should be made under comparable volume conditions.

Other Factors—The effects of some other factors which influence speed distributions have been the subjects of scattered studies, but not enough material appears to have been published to justify important generalizations.

Posting of speed limits seems often to result in lower dispersion, with or without decrease in average speed (see table I, case 5, and reference [9], but conflicting reports have been made [8, p. 6]). The effect of highway curves depends of course upon the safe speed of the curve. If deceleration is required, dispersion should decrease, but the coefficient of variation need not (table I, cases 6a, 6b, columns 7 and 10). Posting of advisory speed signs at curves may substantially decrease the coefficient of variation [7, p. 18].

8. Travel time

Travel times are usually obtained by utilizing test vehicles which "float" with traffic, so as to simulate average traffic characteristics. Each driver is instructed to stay with traffic, and to pass as many cars as pass his vehicle. Travel times are also obtained by taking data on all the vehicles composing a traffic stream. In this procedure, the license number of each vehicle and the times at which it enters and leaves a test section are recorded.

Relationships between travel times obtained by the two methods have been reported by Berry and Green [4].

Travel time studies furnish valuable data on the effectiveness of changed conditions, such as a new traffic signal system, one-way streets, rerouting of bus lines, changes in parking regulations, or widening or rebuilding of the street or highway. Travel time studies provide a quantitative measure of congestion and, in accordance with the test car procedure for obtaining the data, also furnish information on the causes and the amounts of the delay. Results are also used in computing reductions in travel time arising from completed or proposed physical improvements or changes in traffic controls.

The mean travel time is the value most commonly used in comparisons of the travel time performance of a street or highway under the existing physical, traffic, and environmental conditions. It has the advantage of permitting direct computation of total travel time for all traffic.

The median travel time, however, has its own advantage, since it corresponds to the median over-all speed whereas the mean travel time is not directly convertible to the mean over-all speed. Percentile ranges of travel times also are useful as a measure of dispersion, since they can be converted easily to over-all speed ranges. Such ranges can be used in determining the timing for coordinated traffic signal systems.

9. Relation of travel time to spot speeds

The following analysis applies primarily to open highway conditions in which vehicles are not required to stop because of traffic signals, turning traffic, or other factors frequently found in urban areas.

Assume:

1. Over a given unit course each vehicle maintains a constant speed.
2. The speeds are symmetrically distributed around their mean.

The travel times, being reciprocals of speeds, will not be symmetrically distributed. Figure 1 shows a normal speed curve, a, compared with the corresponding travel time curve, b. The latter is more concentrated around its mode (and

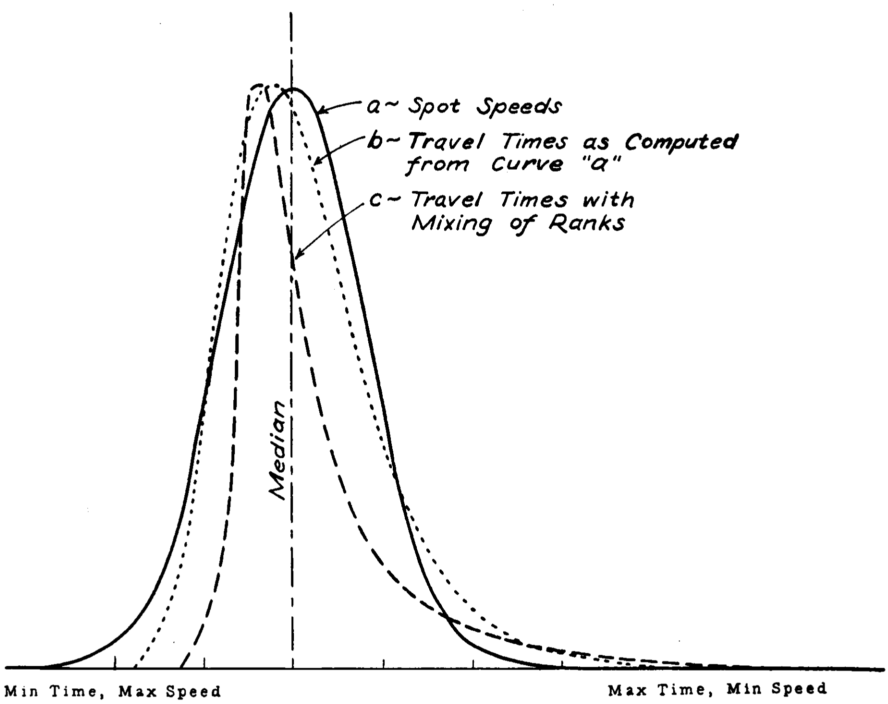

FIGURE 1

Typical distributions of speeds and travel times (open highway conditions)

median), and skewed toward the higher times. The extent of such deviation depends only upon the coefficient of variation of the normal curve.

In practice, vehicles do not maintain a constant speed. The nature of a typical actual time distribution may perhaps be derived from the following considerations.

Define the average speed v of each vehicle as

$$v = \frac{1}{s} \int_0^s \left(\frac{d s}{d t} \right) d s , \qquad \text{where } s = \text{distance}, \; t = \text{time} .$$

Assume

(a) The travel time for each vehicle is given by

$$t_1 = \frac{s}{v} ,$$

(b) The spot speeds are normally distributed,

(c) Vehicles maintain their respective ranks—that is, each remains at the same distance (in σ) from the spot mean at all points.

Under these conditions, the average speeds v will be normally distributed, and the time curve will be as shown in figure 1.

(If, in two normal curves, points of the same σ distances from the mean are averaged together, the result will be a normal curve of average $\bar{x}$ and σ. Thus any number of normal spot speed curves would compose into a normal over-all speed curve, if ranks are maintained.)

Actually, the assumptions (a), (b) and (c) are all suspect. The problem is to estimate the influence of the errors each introduces.

Assumption (a) calculates the time as

$$t_1 = \frac{s^2}{\int_0^s \left(\frac{ds}{dt}\right) ds}$$

whereas actually the time is given by

$$t = \int_0^s \left(\frac{dt}{ds}\right) ds .$$

To estimate the error introduced by use of t_1, suppose that the speed of a vehicle varies linearly over the distance, rising from v_0 to cv_0 in the distance s_1. Then

$$\frac{ds}{dt} = \frac{v_0 (c-1) s + v_0 s_1}{s_1}$$

and

$$t_1 = \frac{2 s_1}{v_0 (c+1)}$$

while the actual time would be

$$t = \int_0^s \left(\frac{ds}{dt}\right) ds = \frac{s_1 \log c}{v_0 (c-1)} .$$

Hence

$$\frac{t}{t_1} = \frac{c+1}{c-1} \frac{\log c}{2} .$$

We have:

c	t/t_1
1	1
1.5	1.01
2	1.04
5	1.21
10	1.42

Hence if the speed of a vehicle fluctuates fairly smoothly, with the maximum speed not more than 2 or 3 times the minimum, the calculation of travel time as t_1 should be reasonably accurate. For typical open highway conditions, assumption (a) is probably satisfactory. Assumption (b) as we have seen, holds also fairly well. Typical open highway spot speed distributions are roughly normal, only mildly skewed.

Assumption (c) is the poorest. Vehicles do not maintain their respective ranks at all points. Mixing of ranks has no effect on the mean value of the over-all speeds. It will however, reduce their standard deviation, since

$$2\left(\bar{x}-\frac{x_1+x_2}{2}\right)^2<(\bar{x}-x_1)^2+(\bar{x}-x_2)^2.$$

If this mixing is symmetric around the over-all mean speed, the resulting over-all speed distribution will remain symmetric (although it need not remain normal).

Practically, however, it is easier for slower vehicles to maintain their speed than it is for the faster vehicles. More mixing of rank would be expected, with wider rank changes, on the high speed side of the mean.

Thus, under probable mixing conditions, the over-all speeds would have the same mean as in the absence of rank mixing but would be compressed toward the mean, and especially so on the high speed side. Their variance would be smaller, and skewness probably greater, and toward the low speeds.

The corresponding travel times would have smaller variance and probably much greater skewness toward the large times, than would times for the fixed rank condition.

A comparison of distribution curves for the two conditions is shown schematically in figure 1. Curve b represents travel times arising from normal spot speeds (curve a) with ranks remaining fixed. Curve c is a typical travel time distribution under rank mixing conditions, displaying smaller variance and greater positive skewness.

On streets having delays because of stops for traffic signals or other traffic conditions it is difficult to relate travel times to spot speeds. The general effect of such delays on the travel time curve may be suggested by considering the probable over-all speed distribution. As delays increase, the average speed decreases, and the speed distribution tends to become increasingly skewed toward the higher speeds (table I, cases 7, 8, and 9). This offsets to some extent the tendency of travel times to be skewed toward the larger times, and in extreme cases may result in a travel time curve roughly symmetric about the median.

10. Applications to preliminary data

Table II shows applications of the same methods of analysis used in table I. Case 1 in table II shows comparisons of speeds and travel times for a four-lane divided rural highway. Case 1-a represents the average of two "typical" spot speed distributions. Case 1-b, the travel times by direct conversion of these average spot speeds, is skewed toward the larger travel times. Case 1-c, showing the actual travel times for a 7.6 mile section, has greater skewness toward the large travel times, and a lower dispersion, as expected. Case 1-d shows that over-all speeds for this high speed open highway are skewed toward the lower speeds, and have a lower dispersion than the spot speed distribution in case 1-a.

Cases 2, 3, and 4 of table II show distributions of travel times and over-all speeds for sections of heavily traveled streets controlled by traffic signals [4]. Case 2 (Wilshire Boulevard) has a flexible progressive signal system with low delay. Drivers have more opportunities to select their own speeds, than in either case 3 or 4

TABLE II

ANALYSIS OF TRAVEL TIME AND OVER-ALL SPEED DISTRIBUTION

Case No.	Location and Conditions	Type of Distrib.	1	2	3	4	5	6	7	8	9	10
			R^* / $F\sigma$ est				Mean	Median	σ est	$P_{85}-P_{15}$	P_{85}	σ est / Median
			93–7	85–15	70–30	93–50						
1	Four-lane divided highway 7.6 miles on US 40 near Fairfield, Calif.											
a	Average of two spot speed distributions	←	1	0.95	0.94	1.00	52.2 mph	52.3 mph	8.59 mph	17.0 mph	60.8 mph	0.16
b	Travel times by direct conversion of spot speeds 1-a	←	1	0.92	0.88	1.24	8.68 min	8.66 min	1.52 min	2.88 min	10.3 min	0.17
c	Actual travel times	Time	1	0.95	0.80	1.42	9.37 min	8.98 min	1.38 min	2.71 min	10.8 min	0.15
	Actual over-all speeds	Speed	1	1.01	0.92	0.76	49.4 mph	50.4 mph	6.71 mph	14.1 mph	56.0 mph	0.13
2	Wilshire Blvd. (Los Angeles) [4] 1.5 miles Westbound 4:30–6:00 P.M.	Time	1	0.72	0.60	1.31	4.59 min	4.40 min	1.26 min	1.89 min	5.47 min	0.29
		Speed	1	0.91	0.74	1.06	21.2 mph	20.8 mph	4.78 mph	9.0 mph	25.7 mph	0.23
3	Potrero Ave. (San Francisco) [4] 1.0 miles Southbound 4:30–6:00 P.M.	Time	1	0.98	1.08	1.23	4.61 min	4.48 min	0.65 min	1.32 min	5.33 min	0.15
		Speed	1	1.03	1.10	1.02	13.1 mph	13.1 mph	1.73 mph	3.7 mph	14.7 mph	0.13
4	Broadway (Oakland) [4] 1.5 miles Northbound											
	5:00–6:00 P.M.	Time	1	1.02	0.99	1.10	7.29 min	7.19 min	1.87 min	3.93 min	9.23 min	0.26
		Speed	1	1.04	0.99	1.25	13.2 mph	12.5 mph	3.36 mph	7.2 mph	16.8 mph	0.27
	2:00–4:00 P.M.	Time	1	0.95	0.90	1.41	5.64 min	5.29 min	1.42 min	2.80 min	7.22 min	0.27
		Speed	1	1.02	1.04	0.92	16.8 mph	16.9 mph	3.76 mph	7.9 mph	20.3 mph	0.22

* Ratio of indicated percentile range to the corresponding range of the normal distribution having the observed $P_{93} - P_{07}$ range. Data in order of increasing times and speeds. σ est $= (P_{93} - P_{07})/2.95$.

where delays and congestion are greater. On Potrero Avenue (case 3) there are traffic signals at every block, and drivers have less chance to pick their own speed. Case 4 (Broadway) represents heavy traffic delay on a street with few traffic signals (5–6 P.M.), and also the same street during times of lighter traffic, with less delay (2–4 P.M.).

Case 4 (5–6 P.M.) with the heaviest congestion, shows the least skewness in travel time distribution (column 4). This is in accordance with the analysis given above. For cases of lesser delay, the travel times are markedly more skewed (and the corresponding over all speeds far less skewed toward high speeds).

The range $P_{85} - P_{15}$ of the over all speeds furnishes significant information on dispersion. The very low value for case 3 is caused by the characteristics of the signal system, which here severely restrict speeds.

In comparing travel time distributions, $P_{85} - P_{15}$ is directly useful only if the courses being considered are of equal length. The ratio of the range to the median affords a more useful measure.

11. Sample size

In estimating the minimum number of observations required to find the mean speed (within a given range, and with a given confidence) the usual procedure applies. Because of the observed normality of spot speed data, it is also reasonable to apply normal theory to the size of sample required for the estimation of percentiles. One formula suffices for both cases, provided that the sample size is "large" (say 100 or greater):

$$\text{Minimum sample size } N = \frac{v^2\sigma^2(2+u^2)}{2d^2}$$

where v = normal deviate corresponding to desired confidence,
u = normal deviate corresponding to the percentile being estimated,
d = permitted error in the estimate,
σ = standard deviation of the population.

For example: If $\sigma = 6$, to estimate the 75th percentile within 1 mph, with 90% confidence:

$$N = \frac{(1.64)^2(6^2)(2+.67^2)}{2(1)^2} = 120\,.$$

σ is usually unknown, but can often be estimated closely enough to permit the formula to be a useful guide. After the observation has been taken, its standard deviation may be used in the formula to check the adequacy of N.

The formula is based on the distance from the mean to the required percentile. In skewed distributions, the mean is usually more centrally located in, for example, the interval $(P_{15} - P_{85})$ than is the median. Hence the formula applies quite closely where skewness occurs in the moderate amount characteristic of spot speed distributions.

12. Derivation of sample size formula

The population is assumed to be normal and the sample size "large." To estimate the percentile p, within d mph, let

m_s = sample mean, σ_s^2 = sample variance,
m = population mean, σ^2 = population variance.

The estimate of p will be

$$m_s + u\sigma_s$$

where u is the normal deviate (mean 0, variance 1) corresponding to the percentile p.

$$m_s = m + \frac{\alpha\sigma}{\sqrt{N}}, \qquad \text{and} \qquad \sigma_s = \sigma + \frac{\beta\sigma}{\sqrt{2N}},$$

where α and β are independent normal deviates (mean 0, variance 1), since the sample mean and standard deviation may here be taken to be independently normally distributed. It is required that the probability,

$$P\{ m + u\sigma - d \leqq m_s + u\sigma_s \leqq m + u\sigma + d \} = c$$

or

$$P\left\{ -d \leqq \frac{\alpha\sigma}{\sqrt{N}} + \frac{u\beta\sigma}{\sqrt{2N}} \leqq d \right\} = c,$$

or

$$P\left\{ \left| \alpha + \frac{u\beta}{\sqrt{2}} \right| \leqq \frac{d\sqrt{N}}{\sigma} \right\} = c.$$

But $\left(\alpha + \frac{u\beta}{\sqrt{2}}\right)$ is itself a normally distributed variable, of mean 0 and variance $1 + \frac{u^2}{2}$. Hence

$$P\left\{ \left| \alpha + \frac{u\beta}{\sqrt{2}} \right| \leqq \frac{d\sqrt{N}}{\sigma} \right\} = P\left\{ | v | = \frac{d\sqrt{2N}}{\sqrt{2 + u^2}\sigma} \right\} = c,$$

where v is the normal deviate (mean 0, variance 1) corresponding to the confidence c. Solving for N,

$$N \geqq \frac{v^2\sigma^2(2 + u^2)}{2d^2}.$$

13. Summary

The use of the range $P_{85} - P_{15}$ is recommended to engineers in reporting the dispersion of speeds and travel times of motor vehicles. A more thorough description of distributions uses also the ranges $P_{93} - P_{07}$, $P_{70} - P_{30}$, and $P_{93} - P_{50}$. These permit easy study of the general shape of the distribution, including its skewness and the relative compactness of various portions.

The speeds of vehicles past a point on a highway tend to have a roughly normal distribution except when traffic volume exceeds the practical capacity of the highway. Study of the relation between spot speeds and travel times reveals that travel times for free moving traffic should be heavily skewed toward the greater times, with this skewness decreasing as delays increase. This conclusion is supported by the data of tables I and II.

14. Further study

There is a need for further study of the effect of various physical, traffic, and environmental conditions on distributions of speeds and travel times. When much more data are available on the effect of the different factors, under different traffic volume conditions, it may be possible to select standards for 85th–15th percentile ranges which can be regarded as desirable in measuring the effectiveness of traffic control programs.

The Committee on Speed Characteristics, of the Highway Research Board, in establishing procedure for collection of data on speed characteristics, should consider the feasibility of covariance studies to isolate the effect of each important variable.

REFERENCES

[1] *Manual of Traffic Engineering Studies*, Association of Casualty and Surety Executives, 1945, pp. 20–23, 65–70.

[2] *Speed Regulation*, National Safety Council, Chicago, 1941, pp. 53–56.

[3] D. W. LOUTZENHEISER, "Percentile speeds on existing highway tangents" and discussion by GREENSHIELDS, *Proc. Hwy. Res. Bd.*, Vol. 20 (1940), pp. 372–392. Case 9 of table I is curve 4–0 of p. 390.

[4] D. S. BERRY and F. H. GREEN, "Over-all speeds of motor vehicles in urban areas," *Proc. Hwy. Res. Bd.*, Vol. 29 (1949), pp. 311–318.

[5] F. MOSTELLER, "On some useful 'inefficient' statistics," *Annals of Math. Stat.*, Vol. 17 (1946), pp. 377–408.

[6] *Highway Capacity Manual*, Bureau of Public Roads, U.S. Government Printing Office, 1950, pp. 29–34.

[7] *Report of Special Study on Speed Zoning, 1938 Report of Committee on Speed and Accidents*, National Safety Council, 1938, pp. 1–47.

[8] C. C. WILEY, C. A. MATYAS, and J. C. HENBERGER, "Effect of speed limit signs on vehicular speeds," mimeographed, Dept. Civil Engineering, University of Illinois (1949), pp. 1–19.

[9] D. S. BERRY, "Speed control," *Proceedings, Convention Group Meetings, Amer. Assoc. of State Hwy. Officials*, Vol. 34 (1948), pp. 361–365.