Speed distribution profile of traffic data and sample size estimation

N. Nezamuddin, University of Texas, Joshua L. Crunkleton, URS Corporation, Philip J. Tarnoff, Center for Advanced Transportation Technology, University of Maryland, and S. Travis Waller, University of Texas.

Sample size estimation is fundamental to traffic engineering analysis. An iterative procedure using standard deviation estimates is the most reliable method of sample size estimation, but practitioners find it cumbersome. In the past, attempts have been made to simplify this process and render it a one-step exercise. To this end, the Institute of Transportation Engineers (ITE) manuals provide standard deviation values for spot speeds for roadways classified by annual average daily traffic (AADT) volumes. This study obtained new estimates of standard deviation using a more granular, operational-level data collected by traffic detectors. It validated the earlier ITE estimates for moderate traffic conditions on freeways, but found the ITE estimates to be inadequate for arterials and for very low and heavy traffic conditions on freeways. This study also found a consistent U-shaped relationship between the standard deviation of speed and traffic volume, which alludes to an inherent speed distribution profile of traffic data. These speed distribution profiles are robust since they exist across different a) locations, b) roadway types (freeway and arterial), and c) aggregation intervals (5-, 15-, and 60-minutes). The U-shaped curves are also validated through existing traffic safety literature.

INTRODUCTION

Sample size is calculated using the standard deviation of the performance measure under study. If standard deviation for the given measure is known, required sample size can be calculated in a single step using a standard formula. Otherwise, an iterative procedure using estimates of the standard deviation is used to calculate sample size. In this procedure, an initial number of data points for the given measure are collected and its standard deviation is estimated. Using this standard deviation, required sample size is calculated for the desired accuracy and confidence level. If the calculated sample size is less than the initial number of data points taken, then the data taken initially is sufficient to estimate the performance measure with the given accuracy and confidence level. Otherwise, the procedure is repeated iteratively until the calculated sample size and the degree of freedom of *t*-statistic converge (Equation 1).

If the number of data points used for calculations is thirty or more, the estimator of the performance measure (ie, average speed) can be assumed to follow a normal distribution, and therefore z-statistic for the given confidence level can be used for sample size calculation. Otherwise, Student's *t*-statistic, whose value itself depends on the initial sample size, should be used. This further complicates the calculation.

Naturally, traffic engineers and field practitioners find the sample size calculation process cumbersome. In the past, attempts have been made to simplify this process and even render it a one-step exercise. The ITE manuals provide standard deviation values for spot speeds on different roadways classified by average annual daily traffic (AADT) volumes. Unfortunately, these values, which were suggested almost three decades ago, are based on very aggregate characteristics of traffic data (ie, AADT volume) and are inadequate for traffic operations and intelligent transportation system (ITS) applications today. Another method present in literature is borrowed from quality control theory and makes use of the *average range* of a sample instead of its standard deviation. This method assumes a simple relationship between these two quantities, which may not necessarily hold true for all datasets.

This study estimated standard deviation of speed for different locations using large amount of detector data and found a consistent speed distribution profile in traffic data. These profiles hold at different aggregation intervals (1-minute, 15-minute, and 1-hour) and across different locations for both freeways and arterials. Also, these values are better suited for traffic operations and real-time applications than those based on AADT volumes.

CONTACT DETAILS
Corresponding author N Nezamuddin is a Graduate Research Assistant in the Department of Civil, Architectural & Environmental Engineering at The University of Texas at Austin. He can be contacted by email at nezam@mail.utexas.edu

BACKGROUND

Sample size estimation methods

The minimum sample size required to estimate a variable with an accuracy of $\pm\varepsilon$ units at a certain confidence level is given by this standard formula (1):

$$n \geq \left(\frac{t_{\alpha/2, n-1} \times s}{\varepsilon} \right)^2 \qquad (1)$$

Where:
s = Sample standard deviation,
ε = Error tolerance
$t_{\alpha/2, n-1}$ = Student's *t*-statistic for (1-α)*100% confidence level and (n-1) degree of freedom

This is the fundamental equation for calculating sample size. However, it is not the simplest method in practice because it involves several iterations with different values of the *t*-statistic each time. If the variable is normally distributed or a large sample (≥ 30 data points) is available to calculate the standard deviation, the *z*-statistic can be used instead of the *t*-statistic.

For travel time and delay studies, the *ITE Manual of Traffic Engineering Studies* and the *Manual of Transportation Engineering Studies* suggest a simpler method of sample size calculation which uses the average range of data instead of its standard deviation (2, 3). This method is borrowed from quality control theory where 3σ (99.7% confidence level) control limits for the mean are determined using the average range in data (4). The average range, $\bar{R}$, is calculated by taking average over the absolute differences between sequential data points (ie, second minus first, third minus second, etc). The formula for sample size is given as:

$$n = \left(\frac{z \times \bar{R}}{d \times \varepsilon} \right)^2 \qquad (2)$$

Where:
z is the *z*-statistic for a given confidence level and d is the ratio of $\bar{R}$ to standard deviation (σ). The value of *d* can be obtained from existing literature (4). The sample size at other confidence levels can be calculated using:

$$n_2 / n_1 = \left(z_2 / z_1 \right)^2 \qquad (3)$$

Where, z_1 and z_2 are normal deviates at corresponding confidence levels.

Validity of the normal distribution assumption is critical in using Equation 3 – and this assumption is questionable because in most practical cases the recommended sample size is less than thirty (2, 3, 4). Overall, this method greatly simplifies the iterative procedure of sample size calculation and is easier to work with since calculating average range of a sample is easier than calculating its standard deviation. It makes the assumption, however, that a simple ratio between average range and standard deviation exists. Li et al (5) have found this method to be the least satisfactory when compared with 'hybrid' and 'modified' methods, which are described below.

Quiroga and Bullock (6) noted that the aforementioned method of calculating $\bar{R}$ is biased since the moving ranges are correlated through their common speed and its value tends to be lower than the true value. They put forth a 'hybrid' formulation which retains the use of $\bar{R}$ because of its intuitive and simple-to-use nature, and compensates it by changing the *z*-statistic in Equation 2 to the *t*-statistic (since $t > z$). The authors also note that there are systematic numerical errors in the sample size tables provided in Robertson (3) and Oppenlander (4).

Li et al (5) reverted to the use of *standard deviation* from *average range* for accuracy purpose. Use of the *z*-statistic instead of the *t*-statistic is also suggested for ease of calculation and an adjustment factor (ε_n) is introduced to balance the discrepancy (since $t > z$). The 'modified' equation is:

$$n = \left(\frac{z \times \sigma}{\varepsilon} \right)^2 + \varepsilon_n \qquad (4)$$

Based on numerical results, the value of ε_n is recommended as 2, 3, and 4 for confidence levels 90%, 95%, and 99%, respectively. Another work suggested a constant value of 2 for ε_n in situations with less than 30 observations (7).

Within the standard deviation-based methods, there are two different formulations depending upon the specification of error tolerance. When the error tolerance is specified in absolute values, Equation 1 is used. If the error tolerance is specified in percent values (α), then a different formulation is used:

$$n = \left(\frac{t \times s}{\varepsilon} \right)^2 = \left(\frac{t \times s}{\alpha/100 \times \bar{x}} \right)^2 = \left(\frac{t \times c_v}{\alpha/100} \right)^2 \qquad (5)$$

Where:
s = Sample standard deviation,
ε = Error tolerance,
$\bar{x}$ = Sample mean
c_v = Coefficient of variation, $\left(s/\bar{x} \right)$

Sample size calculation requires three specifications (confidence level, error tolerance, and measure of variability) and different studies provide these specifications differently depending upon data availability and ease of calculation. Table 1 contains a summary of studies dealing primarily with sample size calculation methods and how they provide the above mentioned specifications.

Estimates of variability

Oppenlander (1963) estimated standard deviation of spot speed to describe its variability for various roadway types classified by AADT volume. Average standard deviations for different roadways ranged from 4.16 to 5.31 mph, and a single value of 5.0 mph was suggested for *any highway in any traffic area* (2, 3, 8).

National Cooperative Highway Research Program (NCHRP) Report 398 – *Quantifying Congestion* – used a large data set of travel time for arterials and freeways to calculate the average and the 85th percentile values of the coefficient of variation (c_v) for sample size calculation purposes (7). Such results are very useful for traffic studies. However, traffic data for different conditions (peak hours, off-peak hours, etc), which are likely to have different variability, were mixed together in the calculation. Turner and Holdener (9) have performed more focused analysis by calculating the 85th percentile of c_v for travel time using only peak-hour data for Houston. Such focused analyses are better suited for real-time applications. Moreover, attempts to estimate population c_v using large datasets by Lomax et al (7) and Turner and Holdener (9) have better justification for using the *z*-statistic in the sample size calculation.

In the present work, similar analyses have been performed to estimate standard deviation of speed at different traffic volume levels, which were subsequently used

Table 1: Summary of Sample Size Calculation Methods

Specifications / Literature	Confidence level		Error Tolerance		Measure of Variability		
	t-statistic	z-statistic	±ε units	α %	σ	$\bar{R}$	c.v.
Oppenlander (1963, 1976)		✓	✓			✓	
Turner et al. (1995)		✓		✓			✓
Lomax et al. (1997)		✓	✓		✓		
Quiroga and Bullock (1998)	✓		✓			✓	
Li et al. (2002)		✓	✓		✓		

to calculate sample size for traffic studies. The results presented here are part of a larger effort by the NCHRP Special Study 20-7, *Guide to Benchmarking Operations Performance Measures*, to provide definition and measurement guidelines for 11 operations performance measures developed by the National Transportation Operations Coalition (NTOC). The reports of this study can be found on the University of Maryland's Center for Advanced Transportation Technology Web site (10).

SPEED DISTRIBUTION PROFILE OF TRAFFIC DATA

In order to establish a possible trend in the variability of speed with traffic volume levels, traffic detector data was collected at 27 locations in the Washington, D.C. metropolitan area. Speed, volume, and occupancy data for arterials and freeways were obtained at intervals of 15-minutes and 5-minutes, respectively. 15-minute and 5-minute average speeds obtained from detectors were grouped into hourly volume bins with a range of 200 vehicles per hour per lane (ie, Group 1: 0 – 200 veh/hr/ln; Group 2: 200 – 400 veh/hr/ln, etc). For a given location, standard deviation of speed was calculated for each volume group using at least five weekdays' worth of data. For simplicity's sake, in order to calculate hourly volume, time intervals were always chosen as 8:00am – 8:59am, 2:00am – 2:59am, etc.

In cases where two or more 15-minute entries were missing from the arterial data for a given hour, that hour was discarded from the analysis. If only one 15-minute entry was missing in a given hour, hourly volume was calculated by extrapolation. Since only hourly volume *range*, not actual volume, is used in determining the appropriate volume bin, the overall analysis is expected to be insensitive to this adjustment. For freeway data, data were discarded from analysis if less than nine 5-minute entries were present for a given hour. Otherwise, extrapolated hourly volume was used for classification into volume bins.

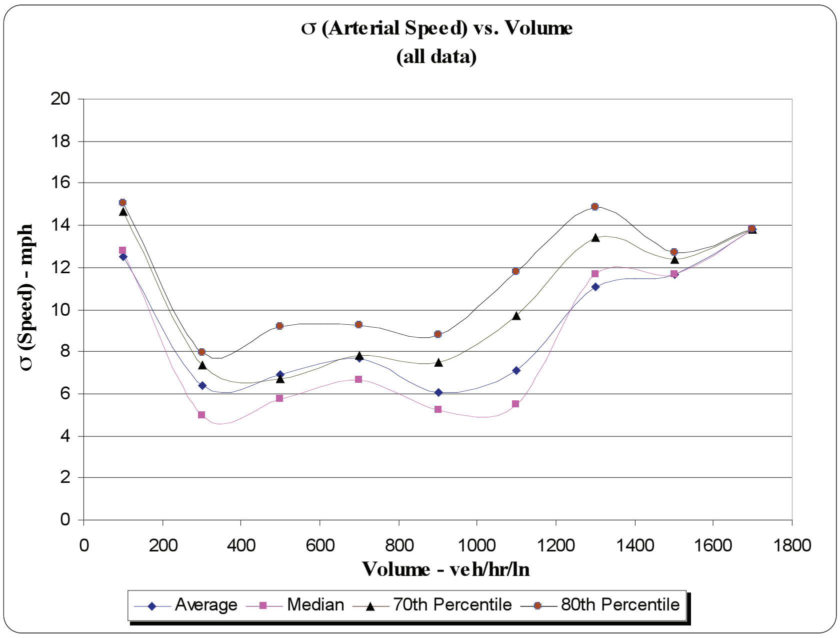
Arterials

Data for arterials was obtained at 20 locations in Northern Virginia. These locations were near intersections, but were sufficiently upstream of the intersections so that they can be assumed to be representative of mid-block traffic conditions. Individual lane data at these locations were also available. Standard deviations of speed were calculated separately for each location.

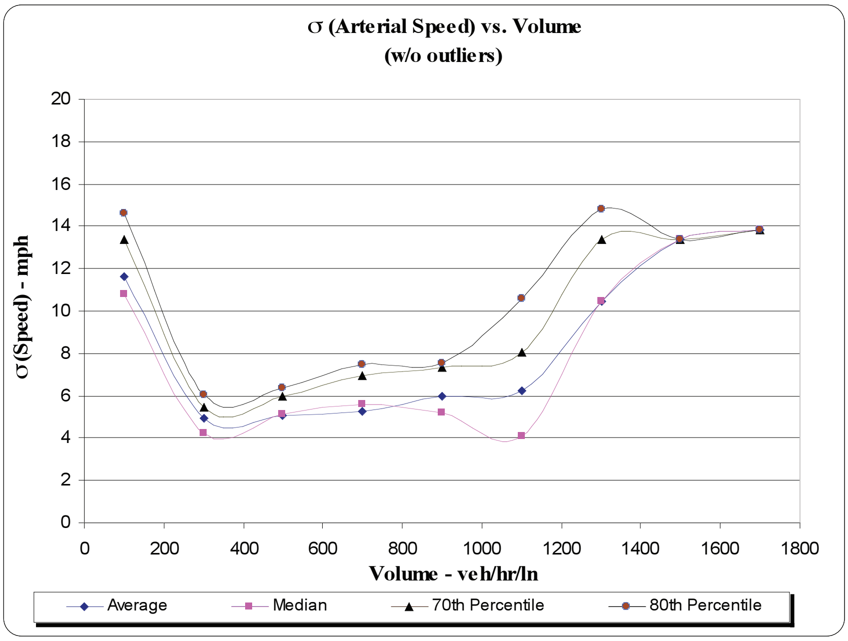
The majority of the locations (80%) had identical speed distribution trends. Several commonly-used summary statistics (average, median, 70th percentile and 80th percentile) for standard deviation values across locations were calculated. These statistics, calculated for both the complete set of data points and after discarding outlier locations, are presented in Figure 1. The general trend between the standard deviation of speed and per lane hourly volume can be described as follows:

> *For volumes less than 200 veh/hr/ln, the standard deviation is characterized by very high values. For higher volume ranges, however, the standard deviation value decreases by half and remains roughly stable throughout the 200 – 1000 veh/hr/ln volume ranges. For volumes greater than 1000 veh/hr/ln, the standard deviation value once again increases. This data showcases a U-shaped relation between standard deviation of speed and hourly volume.*

Both of the plots below (Figures 1 (a) and (b)) follow a general U-shaped relation. Curves in 1(b) show the distribution of speed variability for different traffic levels, which is a significant improvement over a single deterministic value of speed variability for *all* traffic conditions suggested by the ITE manuals.



(a) Variability of speed versus hourly volume plotted using all data.



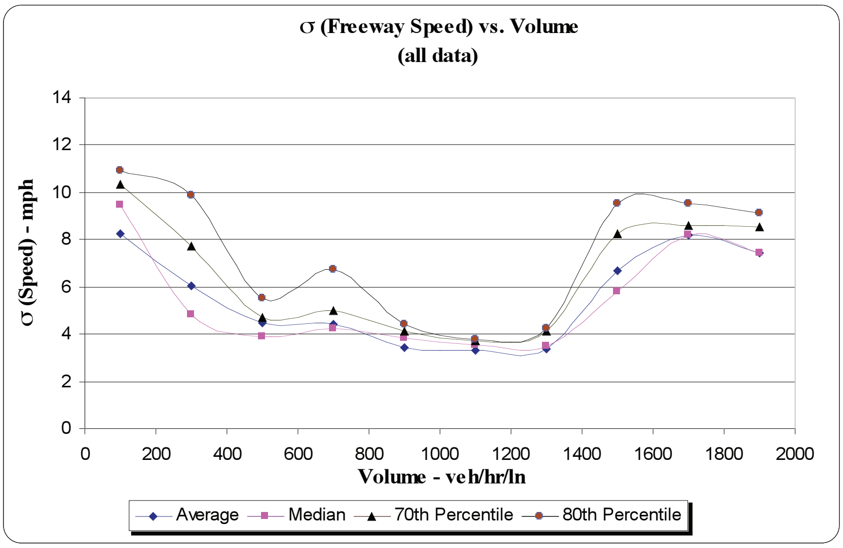
(b) Variability of speed versus hourly volume plotted after discarding outlier data.

Figure 1: Variability of arterial speed with hourly volume.

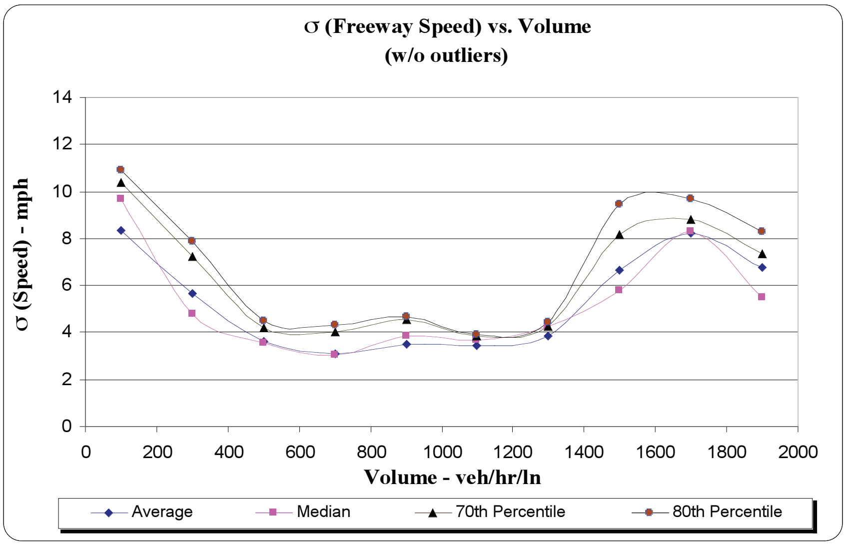
Freeways

Traffic detector data were also obtained at several locations on I-270 and I-695 in Washington, D.C. – Baltimore area. Standard deviations for 5-minute average speeds were calculated in a manner similar to that for arterials. Summary statistics for standard deviation over several locations are plotted in Figure 2.

The relationship between the speed variability and hourly volume was again found to be a U-shaped curve. Few observations were recorded above 1800 veh/hr/ln and the standard deviation of speed decreased in this range. This may be due to the fact that, after exhibiting high speed variability just before the critical density con-



(a) Variability of speed versus hourly volume plotted using all data.



(b) Variability of speed versus hourly volume plotted after discarding outlier data.

Figure 2: Variability of freeway speed with hourly volume.

dition, traffic speed again becomes uniform (ie, low standard deviation of speed) in near-jam conditions where there is little room for maneuvering.

The ITE manuals suggest a uniform standard deviation value of 5 mph for spot speed *on any highway type in any traffic area* (2, 8). It can be seen from Figures 1 and 2 that this recommendation is suitable only for intermediate traffic levels (500 – 1300 veh/hr/ln) on freeways. The above recommendation underestimates the speed variability on arterials (see Figure 1 (b)), and is inadequate for very light and heavy traffic conditions on freeways (see Figure 2 (b)).

Since freeway data was collected at a small interval of 5-minutes, it was checked whether the characteristic U-shaped curve regenerates itself at a higher aggregation interval. Aggregation of speed for higher intervals (15-minute and 1-hour) was performed by taking both simple averages and weighting by volume. Standard deviation values of speed for these two methods matched up to the first decimal point for most volume groups. The results obtained for 15-minute and 1-hour intervals also exhibited a similar trend, and attest to an underlying U-shaped relation between the speed variability and traffic volume. Variability of speed versus volume at different aggregation intervals is presented in Figure 3 for one location, I-270 at Comus Rd.

SAMPLE SIZE

The *Manual of Traffic Engineering Studies* suggests a permitted error of ± 2.0 to ± 4.0 mph for sample size calculation for traffic operations purposes (2). As an example, sample size for arterial and freeway speeds were calculated using 70th percentile standard deviation values shown in Figures 1 (b) and 2 (b).

Since standard deviation estimations are based on large amount of detector data, the t-statistic in Equation 1 has been replaced by the *z*-statistic. Sample sizes for a permitted error of ± 4.0 mph at 90% and 95% confidence levels (CL) are presented in Figures 4 and 5 for arterials and freeways, respectively.

CONCLUDING REMARKS

Explanation of U-shaped Curve

Traffic speed is characterized by high variability for light traffic conditions for both arterials and freeways (see Figures 1 and 2). This may be due to the fact that during such conditions drivers are free to choose a wide range of speeds. During late night and early morning hours, some drivers may be driving at the posted speed limit or below, while a small portion of drivers may be cruising along the roadway at very high speeds. A small number of observations during light traffic conditions may also lead to a high standard deviation of speed.

For intermediate traffic ranges, standard deviation of speed decreases and remains fairly constant. This reflects the stable condition of traffic for such volume ranges. During these conditions, there is enough space available for drivers to maintain their desired speed, though not as much as during low traffic conditions. As a result, traffic remains fairly stable and smooth flowing.

With high volume ranges, speed variability again increases sharply. This represents the unstable nature of traffic before critical density (11). At such times of high density, any small perturbation quickly propagates itself and often causes the transition from free-flow to jam-density.

This characteristic U-shaped curve is thoroughly validated by the existing traffic safety literature. High accident rates correlate positively with high standard deviation of speed (12, 13). One of the earliest reports investigating relationship between the two, *Speed Variance and Its Influence on Accidents*, found that accident rate (measured as number of accidents per 100 million vehicle miles of travel) was proportional to the square of speed variance (14). This implies that during very light and heavy traffic conditions, when standard deviation of speed is greater, accident rates should be higher, too. Therefore, relationship between accident rates and traffic volume level should be a U-shaped curve. Such a relationship is well known among traffic safety researchers and is extensively reported (15, 16, 17, 18, 19).

Relevance of sample size issue

With the advent of ITS technologies, large amounts of traffic data can be collected with minimal effort in most circumstances. Although today there is no dearth of data, the issue of sample size is still very relevant, if not critical. Even with more data in hand, traffic agencies need information regarding sufficient data requirements. These data requirements are used to base their calculations of certain traffic characteristics in order to report on variable message signs (VMS) or for dissemination by other means.

There are a number of circumstances in which sample size is extremely important, including:

a) Travel times during short duration events (eg, incidents): States are anxious to post travel times on VMSs, and one of the most important points of the present work is to provide guidance related to how long they should wait (sample the change in speeds) until they have a high enough level of confidence to post the message of longer travel times.

b) Measurement using temporary instruments: There are instances where measurements are made based on temporary instrumentation. States often use automatic traffic recorders (ATRs) to collect statewide traffic and speed counts. Depending on the hourly volume, the minimum sample size to estimate speed can be calculated using standard deviation curves shown in Figures 1 (b) and 2 (b).

c) Traffic signal timing: There are instances when traffic studies are conducted for traffic signal timing purposes on arterials. These counts are made by individuals, manually over a single 15 minute to one hour period. Again, standard deviation estimates provided in this study can be useful.

Standard deviation curves presented in Figures 1 (b) and 2 (b) show the range of standard deviation variability and can serve as a general guideline for locations where data cannot be collected locally due to technical or budgetary constraints. In these circumstances, it is better to select a value of standard deviation using different percentile curves for a given traffic level than pick a constant value of 5 mph for *any highway type in any traffic area* for *all* traffic conditions.

Limitations

In this study, standard deviations are calculated for 5-minute and 15-minute average speeds and not for individual speeds of vehicles. But real life detector data is also reported as an average over a time interval (30-seconds, 5-minutes, etc) and individual vehicular speeds are not stored. As a result, traffic models based on detector data (eg, crash-prediction models, travel-time estimation models, etc) often use averaged data instead of individual vehicular data. Therefore, although standard deviation values presented in this paper are not actual standard deviation values of individual vehicular speed *per se*, they are still useful because of their use in modeling. Oh et al have found 5-minute standard deviation of speed as the most important factor in predicting accidents (20). Pande and Abdel-Aty have also used it in their crash prediction models (21).

Conclusions

This study presented estimates of the standard deviation of speed based on inherent speed distribution profiles of arterial and freeway traffic data. These speed profiles have been found to exist across different locations and at different aggregation intervals. The characteristic U-shaped curve is identifiable even in the presence of small amounts of outlier data, and it is verified by the traffic safety literature. All of the above observations point to the robustness of the speed distribution profile and the standard deviation estimates of speed.

To the knowledge of the authors, this is the first study of its kind after Oppenlander (8), which found its way in the *ITE Manual of Traffic Engineering Studies* and the *ITE Manual of Transportation Engineering Studies*. The standard deviation value of 5 mph suggested by Oppenlander (8) is suitable only for intermediate traffic volumes on freeways. It underestimates the variability of traffic speed during very light and heavy traffic conditions on freeways, and for all traffic conditions on arterials. The new estimates have been provided for operational level traffic conditions (eg, hourly volume) as opposed to planning level traffic conditions (eg, AADT volume) provided in previous literature. The authors believe that the present results are a significant improvement over the earlier estimate of a single value of 5 mph for all circumstances, and these results are more suitable for ITS applications.

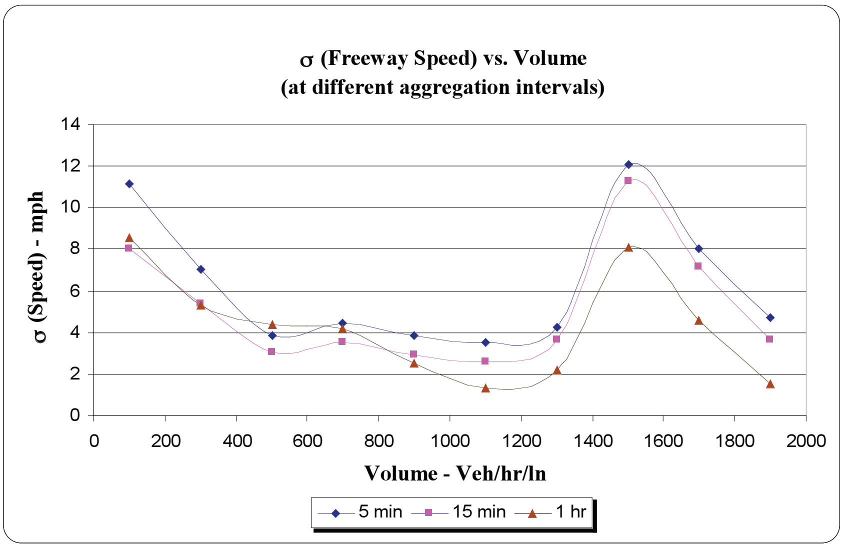


Figure 3: Variability of freeway speed at different aggregation intervals.

ACKNOWLEDGMENT

This study was conducted as a part of the National Cooperative Highway Research Program (NCHRP) research project 20-7 – *Guide to Benchmarking Operations Performance Measures* – at the University of Maryland's Center for Advanced Transportation Technology (CATT). The authors acknowledge the support of the Federal Highway Administration/National Transportation Operations Coalition. Sincere thanks are extended to Mr. Ta-Cheng Hsu of the Virginia Department of Transportation and Mr. Nikola Ivanov of CATT for their assistance with obtaining data for this study.

The first two authors were affiliated with the University of Maryland, College Park at the time of this study.

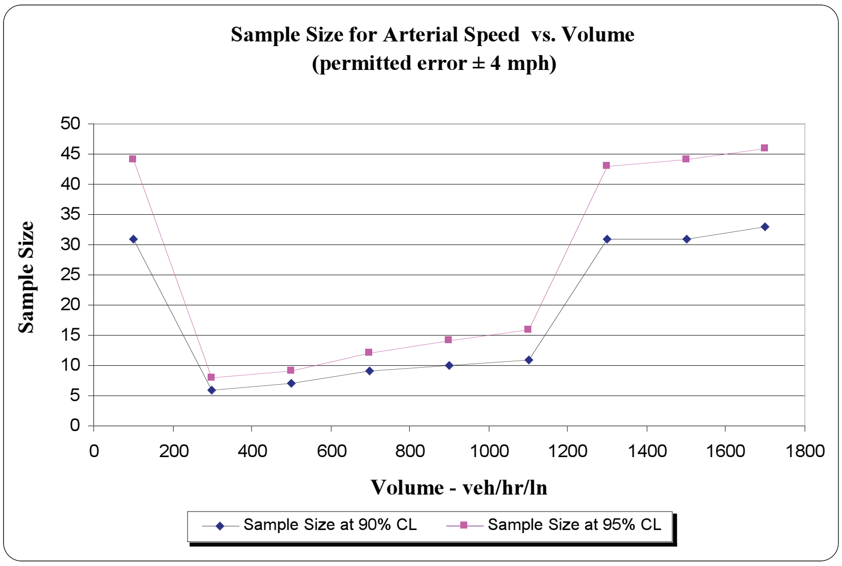


Figure 4: Sample size estimates for arterial speed.

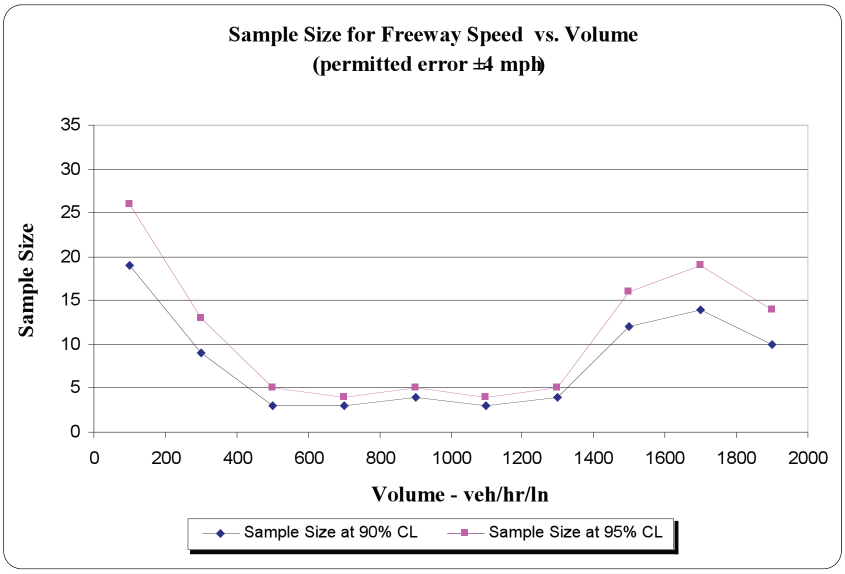


Figure 5: Sample size estimates for freeway speed.

REFERENCES

1. Devore, J.L. Probability and Statistics for Engineering and the Sciences. Brooks/Cole Publishing Co., Belmont, CA, 2004.

2. Box, P.C., and J.C. Oppenlander. Manual of Traffic Engineering Studies, 4th ed., Institute of Transportation Engineers, Washington, D.C., USA, 1976.

3. Robertson, H.D. (Ed.). ITE Manual of Transportation Engineering Studies. Prentice Hall, Englewood Cliffs, N.J., 1994.

4.Oppenlander, J.C. Sample Size Determination for Travel Time and Delay Studies. Traffic Engineering, Vol. 46, No. 9, Institute of Traffic Engineers, Washington D.C., 1976, pp. 25–28.

5. Li, S., K. Zhu, B.H.W. van Gelder, J. Nagle, and C. Tuttle. Reconsideration of Sample Size Requirements for Field Traffic Data Collection Using GPS Devices. In Transportation Research Record: Journal of the Transportation Research Board, No. 1804, TRB, National Research Council, Washington, D.C., 2002, pp. 17–22.

6. Quiroga, C.A., and D. Bullock. Determination of Sample Sizes for Travel Time Studies. ITE Journal, Vol. 68, No. 8, Institute of Transportation Engineers, Washington, D.C., 1998, pp. 92–98.

7. Lomax, T., S. Turner, G. Shunk, H.S. Levinson, R.H. Pratt, P.N. Bay, and G.B. Douglas. Quantifying Congestion, Volume 2: User's Guide. NCHRP Report 398, TRB, National Research Council, Washington, D.C., 1997.

8. Oppenlander, J.C. Sample Size Determination for Spot-Speed Studies at Rural, Intermediate, and Urban Locations. In Highway Research Record: Journal of the Highway Research Board, No. 35, HRB, Washington, D.C., 1963, pp. 78–80.

9. Turner, S. M., and D.J. Holdener. Probe Vehicle Sample Sizes for Real-Time Information: The Houston Experience. In proceedings of the 6th International Vehicle Navigation and Information Systems Conference, Seattle, Washington, 1995, pp. 3-10.

10. Guide to Benchmarking Operations Performance Measures, NCHRP 20-7. http://www.catt.umd.edu/research/nchrp-guide.html. Accessed December 7, 2009.

11. Traffic Flow Theory: A State of the Art Report. Federal Highway Administration. http://www.tfhrc.gov/its/tft/tft.htm. Accessed July 29, 2009.

12. Lave, C. Speed, Coordination, and 55 mph Limit. The American Economic Review, Vol. 75, No. 5, 1985, pp. 1159-1164.

13. Thorton, M., and R. W. Lyles. Freeway Speed Zones: Safety and Compliance Issues. In Transportation Research Record: Journal of the Transportation Research Board, No. 1560, TRB, National Research Council, Washington, D.C., 1996, pp. 65–72.

14. Garber, N.J., and R. Gadirau. Speed Variance and Its Influence on Accidents. Publication ED312438. AAA Foundation for Traffic Safety, Washington, DC., 1988.

15. Gwynn, D. W. Relationship of Accident Rate and Accident Involvements with Hourly Volumes. Traffic Q., Vol. 21, No. 3, 1967, pp. 407–418.

16. Ceder, A., and M. Livenh. Relationship between Road Accident and Hourly Traffic Flow-I and II. Accident Analysis & Prevention, Vol. 14, No. 1, 1982, pp. 19-44.

17. Frantzeskakis, J. M., and D. I. Iordanis. Volume-to-Capacity Ratio and Traffic Accident on Interurban Four-lane Highways in Greece. In Transportation Research Record: Journal of the Transportation Research Board, No. 1122, TRB, National Research Council, Washington, D.C., 1987, pp. 29–38.

18. Hall, J. W., and O. J. Pendleton. Relationship Between V/C Ratio and Accident Rates. Publication FHWA-HPR-NM-88-02. FHWA, U.S. Department of Transportation, 1989.

19. Martin, J-L. Relationship between Crash Rate and Hourly Traffic Flow on Interurban Motorways. Accident Analysis & Prevention, Vol. 34, No. 5, 2002, pp. 619-629.

20. Oh, J-S., C. Oh, S. G. Ritchie, and M. Chang. Real-time Estimation of Accident Likelihood for Safety Enhancement. Journal of Transportation Engineering, Vol. 131, No. 5, ASCE, 2005, pp. 358 – 363.

21. Pande, A., and M. Abdel-Aty. Comprehensive Analysis of Relationship between Real-Time Traffic Surveillance Data and Rear-End Crashes on Freeways. In Transportation Research Record: Journal of the Transportation Research Board, No. 1953, TRB, National Research Council, Washington, D.C., 2006, pp. 31– 40.