Estimating and Updating Flows on Pedestrian Facilities in the Central Business District

P. N. SENEVIRATNE AND M. JAVID

Pedestrian volume data are needed for a variety of purposes—on the demand side, for examining trends and planning facilities, and on the supply side for evaluating safety and level of service, to name a few. However, lack of flexible analytical tools and data gathering costs limit authorities' ability to examine pedestrian flows and provide the required levels of service. Efforts to develop expansion models and demand models to minimize manual data gathering are also hindered by the lack of information about variation in flow over time, as well as uncertainty in the models. A statistical approach based on Bayesian theory is discussed that can combine available data, analysts' experience (subjective judgment), and short-period (sample) counts, to estimate the expected (mean) flow at a given site at a given time that will serve as input for a particular expansion model. Compared with sample means, Bayesian estimates are much closer to the true mean and, hence, expanded values would be less uncertain. Moreover, the technique could be used to update previous estimates by combining them with newly performed short-term counts.

As cities grow and traffic congestion in central areas becomes acute, planners and engineers will need to look more carefully at the possibility of exploiting the more efficient and practical modes of travel such as public transit and walking. In fact, several studies including Seneviratne and Morrall (*1*) and Bondada (*2*) have shown that, even at present, walking is the most frequently used mode for intra–central business district (CBD) trips. Unfortunately though, apart from isolated examples of aesthetically pleasing environments with priority for pedestrians, the concern in terms of long-term goals and objectives for this mode seems inadequate to encourage continued usage.

Some of the imbalances and inequities in investment can be attributed to the deficiency of information on pedestrian needs, network characteristics, and trip functions. How often are pedestrian volume counts taken on the network links, or how many cities conduct pedestrian origin-destination surveys? Nevertheless, pedestrian flow data are extremely important for purposes of planning and design. On the one hand, these data indicate levels of facility usage and trends that are needed for long-term facility planning and policy formulation. On the other hand, these data form the basis for identifying hazardous (high-conflict) sites, delays, and capacity problems.

Each municipality and provincial transportation agency in Canada has procedures for collecting and analyzing vehicular traffic data. However, pedestrian data are not collected or investigated with the same degree of intensity and enthusiasm.

Department of Civil Engineering, Concordia University, 1455 de Maisonneuve Boulevard West, Montreal, Quebec, Canada H3G 1M8.

Although alarming accident statistics and declining employment levels in CBDs have created a stronger need for analyzing pedestrian movement, the absence of flexible analytical tools, costs associated with manual collection, and lack of proven mechanical or electronic data collection techniques have forced authorities to shy away from the question of pedestrians.

Until the technology is fully developed and automated data gathering and analytical tools become easily accessible, local agencies could take advantage of several statistical techniques to minimize the cost of pedestrian data. A technique for estimating and updating short-term pedestrian flow or the input to expansion models is discussed. These input values are usually estimates of the mean (expected) flow during a short time interval of say 10 or 15 min. Sample data, experience of the analysts (subjective judgment), or historic data can be combined according to Bayesian statistical rules to obtain estimates of mean flow closer to the true mean.

ESTIMATION OF PEDESTRIAN FLOW

Following the vehicular traffic flow theories postulated by Greenshields (*3*) and Greenberg (*4*), researchers of pedestrian movement such as Fruin (*5*), Navin and Wheeler (*6*), and O'Flaherty and Parkinson (*7*) concentrated mainly on evaluating walkway capacities and levels of service. Much of the work on the demand side involved development of multiple regression-type predictive models (*8–10*) based on land-use variables at the specific sites. However, as pointed out by Davis et al. (*11*), transferability and long-term validity of these models are not clearly demonstrated. Davis et al. (*11*) proposed expansion models for crosswalk volumes based on data collected at eight sites in Washington, D.C. These authors found that hourly pedestrian volumes can be predicted from counts as small as 5 min during the middle of the hour being considered. From Seneviratne et al. (*12*), relationships also exist between hourly or daily pedestrian volumes and short-period counts. However, the strength and validity of the relationship between the two variables will be (a) dependent on the variability of volume and other characteristics of the sample sites, and (b) influenced by the position and length of the short-term count interval.

Thus, a certain degree of uncertainty can be expected in the validity of the expansion models, primarily in relation to the form and parameters of the models. For instance, Davis et al. (*11*) obtained logarithmic relations between short-period

counts and volumes for periods of up to 4 hr. However, these expansion models would only be valid for sites with a certain range of volumes because the sample sites were selected according to volume. Moreover, the models are unable to explain 100 percent of the variation and in some instances contain large standard errors of estimate. Thus, uncertainty results from the form of the model. Uncertainty in model parameters, on the other hand, stems from insufficient data. In other words, regression constants and coefficients contain uncertainty because they are estimated from small samples.

Uncertainty also stems from the short-term pedestrian flow itself. For instance, even with a sufficiently large data base, the fundamental uncertainty in pedestrian trip generation rates affects the validity of the estimates. Although fundamental or inherent uncertainty will remain despite sample size, the Bayesian approach can be used in two fashions to minimize some other sources of uncertainty in the estimates from expansion models. First, the Bayesian approach could be used to modify the chosen input value or the short-term (say 10-min) count to bring it closer to the true mean of the 10-min counts during the period under investigation. Alternately, if the underlying data base of the expansion models is available, modification of the expanded short-term counts can be done using the Bayesian approach.

In order to use the Bayesian approach, variations in short-term counts and the underlying probability distributions need to be considered.

DATA

Data from two surveys are used in the discussion and a numerical example given later. The first survey was an extensive pedestrian survey conducted in Calgary, Alberta, in 1983. As a facility usage survey, flows on the entire above-ground and at-grade networks were monitored for 6 hr over two consecutive weeks. Of the 32 midblock sites, flows in 30 were recorded at 5-min intervals for 6 hr (i.e., during 0730 to 0900, 1000 to 1130, 1230 to 1330, 1430 to 1530, and 1600 to 1700 hr). Two sites were used as control sites where the flows were monitored continuously from 0700 to 1900 hr for the entire 2-week period. This large data base enabled analysis of variation in flow over time, as well as the development of expansion models.

The second survey was a limited survey conducted in the CBD of Montreal, Quebec, in 1989. This survey was confined to the noon-hour peak (1200 to 1300 hr) when the pedestrian flows at 10 sites similar in terms of land use and volumes were monitored for 2 weeks.

VARIATION OF PEDESTRIAN FLOW

Because of greater interaction between pedestrians, who have a tendency of walking side by side in pairs or groups, as well as trip generation rates that are influenced by capacities of transportation modes, elevators and escalators, and traffic signals, flows are highly irregular. Studies by Haynes (*13*), Seneviratne and Morrall (*14*), and Pushkarev and Zupan (*9*) have revealed that flows during small intervals are virtually irreproducible from one day to another. Thus, because the input to an expansion model [i.e., the mean (expected) short-term flow] is generally estimated from a few counts taken during the time period under investigation and the standard error is large, the expanded values may be uncertain.

Haynes (*13*) has shown that the standard error of estimate can be minimized by selecting short-term counts taken over intervals greater than 10 min. On the other hand, sample data could be combined with subjective judgment on the basis of experience and historical data to obtain the maximum likelihood estimate of the mean short-term flow. This approach is known as the Bayesian estimation technique.

The procedure for combining data depends entirely on the probability distribution of the flow at the investigated site, which has been found to follow a certain regular pattern during sufficiently large time intervals. For instance, even though short-term fluctuations are virtually unavoidable and of less significance to planners and engineers, Haynes (*13*) found that at midblock locations, the 1-min flow is a normal random variable. Seneviratne and Morrall (*14*) show that the 15-min flow during a.m., noon, and p.m. peaks could be represented by normal distributions.

After fitting three different distributions (i.e., normal, Poisson, and negative binomial), Javid (*15*) found that the 5-min flow at intersections can also be reasonably approximated by a normal distribution as shown in Figure 1. Moreover, Javid (*15*) and Seneviratne and Morrall (*14*) have shown that the means of the 5-min flow during a given hour at similar intersections as well as the means of the 15-min flow at comparable (in terms of variation in flow) midblock sites are normally distributed. In other words, the short-term flow at each sample site i is a normal random variable with mean m_i and variance s_i^2 and the distribution of m_i taken over all sites ($i = 1, \ldots, n$) is also normal in form with mean m' and standard deviation σ'. This satisfies some of the prerequisites for using Bayesian updating (i.e., normal conjugate prior).

BAYESIAN ESTIMATES OF EXPECTED FLOW

The Bayesian approach has had many applications in geotechnical engineering (*16*) in which testing is costly and the design characteristic (soil stability) is highly variable as are pedestrian and traffic flows. Yet, this approach has been used to a lesser extent in transportation planning (*17*).

Described in terms of notations, the analyst's objective should be to refine the mean (expected) flow m_s obtained from a few short-term counts at a site j, which is similar to some sites for which sufficient data are available. Suppose that using experience and available data, m' or σ' can be estimated. Then, these two parameters can be combined with m_s and s_s to derive Bayesian estimates of the mean (m'') and the variance (σ''^2) of short-term flow at site j and these estimates would be closer to the true short-term mean m_j.

Thus, the basic procedure consists of three simple steps. The first step involves finding or assuming the prior distribution of mean pedestrian flow per unit of time during the period of analysis. The prior distribution is essentially the distribution of the means of flow at several similar sites, and assumed to be the probability distribution of flow at the site under investigation. The form and parameters of the prior distribution are estimated from available data or analyst's experience with

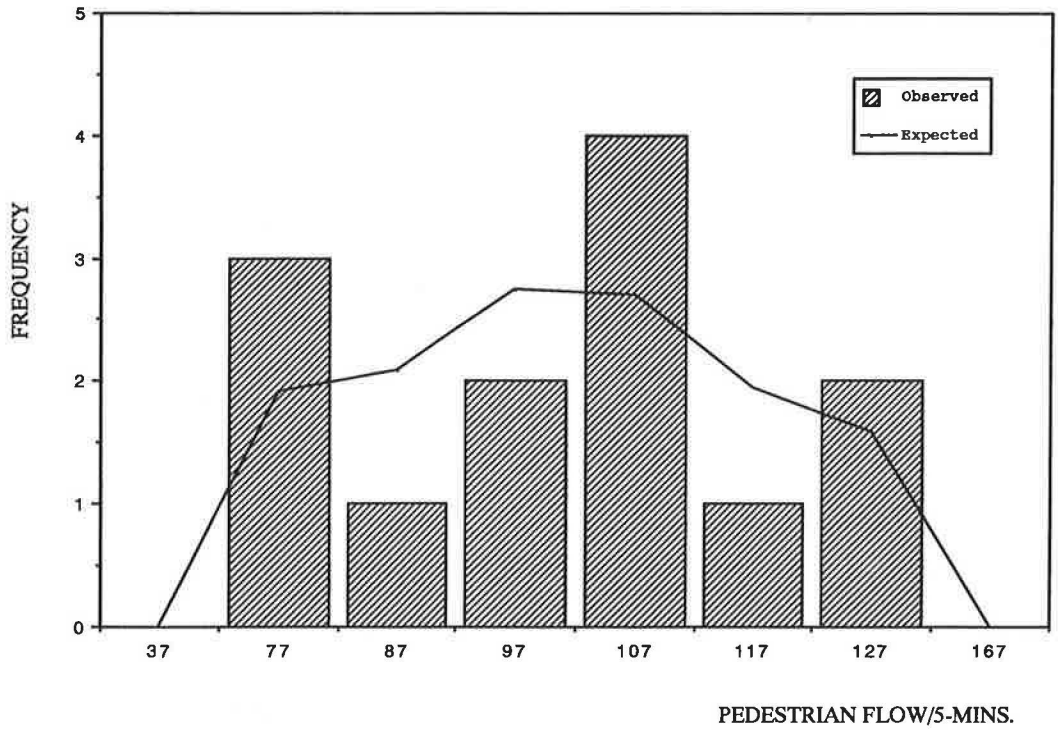

FIGURE 1 Distribution of 5-min flows.

similar sites. As mentioned previously, data from Montreal and Calgary suggest that the prior distribution is normal in form and parameters m' and σ' are obtained from

$$m' = \frac{1}{n} \sum_{\text{all } i} m_i \tag{1}$$

$$\sigma'^2 = \frac{1}{(n-1)} \sum_{\text{all } i} (m_i - m')^2 \tag{2}$$

The next step is to derive the sampling distribution and its parameters. The sampling distribution is the distribution of the few short-term counts at site j. Because the data base in this case is extremely small, an assumption can be made of the form of the distribution. If the distributions of flow at the sites included in the prior distribution are normal, then the distribution of the k ($k << n$) short-term counts $X_j = \{x_1, x_2, \ldots, x_k\}$ can be reasonably assumed to be normal. For example, Montreal and Calgary data were confirmed to have a normal distribution with $N(m_j, s_j)$ and estimates of m_j and s_j usually assumed to be equal to the sample parameters given by the following expressions:

$$m_s = \frac{1}{k} \sum_{\text{all } k} x_k \tag{3}$$

$$s_s^2 = \frac{1}{(k-1)} \sum_{\text{all } k} (x_k - m_s)^2 \tag{4}$$

To obtain the true mean of the 15-min flow (m_j) during the p.m. peak, ideally m_j should be computed from several 15-min observations at the same time over several days. However, because the objective is to minimize data collection efforts, that m_j equals m_s can usually be assumed. The latter, m_s, could be the mean of 15 prorated 1-min counts or a randomly chosen 15-min count during the p.m. peak. Thus, m_s would often differ from m_j.

With the Bayesian approach, m_s could be updated or refined to derive an estimate closer to m_j. In other words, if the forms and parameters of the prior and sample distributions are known, Raiffa and Schlaifer (*18*) have shown that the posterior or updated distribution parameters can be derived using empirical Bayesian analysis. Compared with sample and prior distribution parameters, the updated parameters are shown to be closer to the true parameters.

When prior and sample distribution are normal in form, Howard and Raiffa (*18*) found that the following relationships hold between the posterior distribution parameters and the prior and sample parameters. Thus, the third step is to enter the appropriate parameters obtained from the first two steps into the following expressions:

$$m'' = \frac{m'/(\sigma')^2 + m_s/(s_s^2/k)}{1/(\sigma')^2 + 1/(s_s^2/k)} \tag{5}$$

$$\sigma'' = \left[\frac{1}{1/(\sigma')^2 + 1/(s_s^2/k)}\right]^{1/2} \tag{6}$$

where k, m_s, and s_s are sample size, mean, and standard deviation, respectively, of the sample distribution at Site j. The parameters of the prior and posterior distributions are denoted by (m', σ') and (m'', σ''), respectively.

In order to examine the extent to which sample size and sampling interval will make m'' approach m_j, the 5-min flows during the noon hour at a test site in Montreal were analyzed. The mean 5-min flow (m_j) computed from the 12 counts during

the hour was 74 pedestrians per 5 min and the standard deviation was 11 pedestrians per 5 min. As presented in Table 1 and shown in Figure 2, regardless of the sampling interval and the sample size m'' is generally closer to m_j than is m_s.

NUMERICAL EXAMPLE

The means of 5-min flows during the noon hour from 16 midblock locations in Calgary followed a normal distribution with a mean of 180 persons per 5 min and a standard deviation of 16 persons per 5 min. These two values, which are based on sixty 5-min counts at each site, are the prior distribution parameters m' and σ' that serve as input to Equations 5 and 6.

Suppose an evaluation is needed of the noon-hour volume at a new, previously unsurveyed midblock site similar to the 16 sites used in the prior distribution analysis. This evaluation may be estimated from an expansion model of the form given by Davis et al. (*11*):

$$\text{Average hourly volume} = 19.91\ m^{0.79} \tag{7}$$

where m is the expected middle 5-min flow.

Suppose that because of resource constraints, counts (i.e., x_1, x_2, x_3) could only be obtained during three 5-min intervals commencing at 12:25 p.m. Thus, if the conventional procedure is followed, m would be computed from these three counts. However, if the Bayesian approach is followed, these sample counts can be modified to derive a better estimate of m. With the latter approach, the three counts enable the computation of the sample distribution parameters m_s and s_s from Equations 3 and 4, respectively. Thus, the fundamental

TABLE 1 COMPARISON OF SAMPLE AND POSTERIOR PARAMETERS

data set	k (Sampling Interval)	m_s	s_s	m''	σ''
1	4 (12:00-12:20h)	77	12	75	5
2	4 (12:20-12:40h)	76	10	75	5
3	3 (12:00-12:15h)	69	11	71	5
4	3 (12:15-12:30h)	71	10	72	4
5	3 (12:30-12:45h)	79	12	76	4
6	2 (12:00-12:10h)	78	7	76	4
7	2 (12:10-12:20h)	64	10	68	5
8	2 (12:20-12:30h)	69	10	71	5

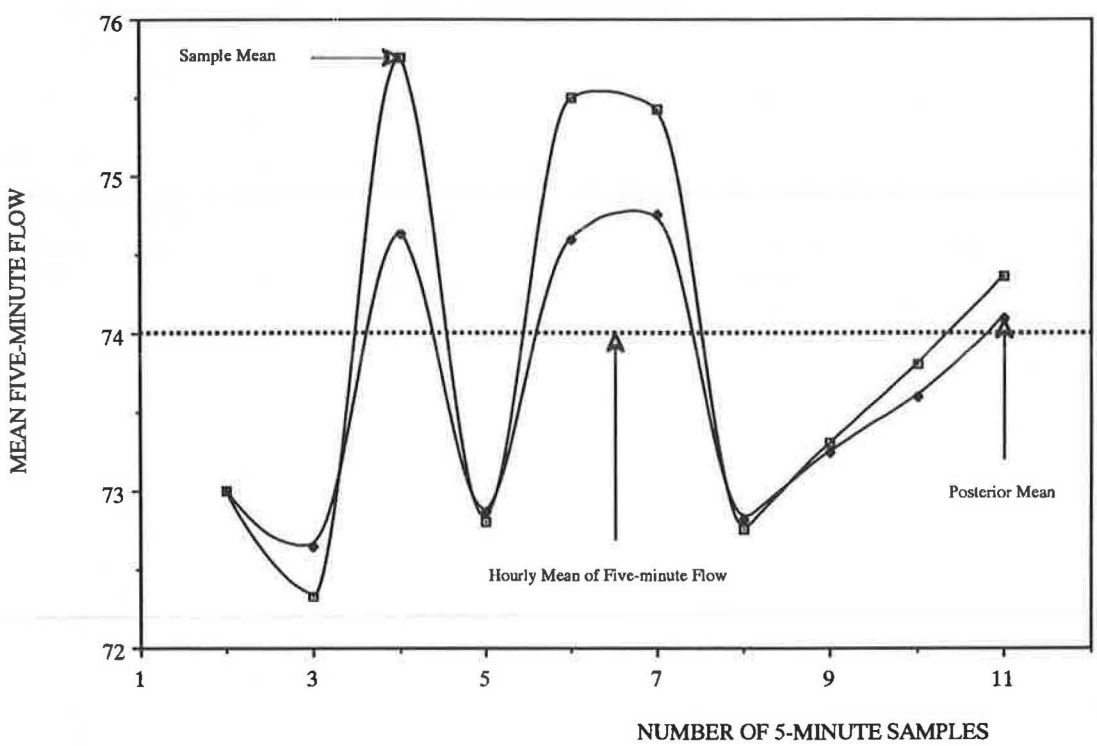

FIGURE 2 Comparison of sample and posterior parameters.

data are $m' = 180$, $\sigma' = 16$, $m_s = 220$, $s_s = 25$, and $k = 3$ (5-min counts).

Substituting these values in Equations 5 and 6 yields posterior parameters or the Bayesian estimates $m'' = 202$ and $\sigma'' = 11$.

When compared with m_s and s_s, m'' and σ'' are closer to the true mean of 200 persons per 5 min and standard deviation of 13 persons per 5 min computed from eleven 5-min counts between 1200 and 1300 hr. This result shows that the Bayesian estimates of the parameters are closer to the true (observed) parameters than the sample parameters and ideally $m = 202$ should be used in Equation 7. In this example, the difference in the expanded value from Equation 7 when using $m = m_s = 220$ is approximately 100 pedestrians per hour.

The alternate procedure requires a knowledge of the standard error of estimate of the expansion model and the mean and standard error of the hourly volumes used in the regression analysis. Accordingly, the latter data would be m' and σ', whereas the standard error of estimate (s_e) of the model would be s_s. The value of m_s would be the average hourly volume obtained from Equation 7 for a given short-term count, in this case 220.

Both procedures are simple and data requirements are minimized (i.e., requires only one 5-min or one 15-min count) while achieving higher accuracy. These procedures could also be used for updating information. For example, next year flow data may need to be updated for the same site. In this case, current posterior parameters will act as prior parameters, and next year another three 5-min counts will need to be taken during the same interval to obtain sample parameters. The next set of posterior parameters (Bayesian estimates) obtained from Equations 5 and 6 will be even closer to the true values than this year's values.

Bayesian approach can also be used as a routine (annual) updating procedure for all sites. If last year's counts are available, the figures could be updated with the aid of a short-term sample by following the previous steps.

CONCLUSION

Given that expansion models can usually explain only a part of the variation, estimation errors could multiply unless input values are accurate. This condition was illustrated in the previous numerical example in which the variability of the short-term counts resulted in a difference in volume of over 100 pedestrians per hour. Even when expansion models are unavailable, Bayesian estimation procedure could be used to derive the required information. For instance, most municipalities may have pedestrian flow data collected during intersection traffic counts in previous years, or may have performed some ad hoc counts at different sites. These data can be combined with sample counts to derive updated flows needed for level-of-service analysis.

Thus, in summary, the empirical Bayesian approach can be used to estimate flows at new sites or update flows at previously surveyed sites. The basic steps to follow at a new site are as follows:

1. Compute, from data available for sites similar in terms of magnitude and variation of flow or assume according to experience, the mean (m') and standard deviation (σ') of, for example, p.m. peak volume likely to exist at the site in question.
2. Perform a sample count over a short period and compute mean (m_s) and standard deviation (s_s).
3. Substitute these values in Equations 5 and 6 to obtain expected volume (m'') and standard deviation (σ''). This value of m'' should be used in the expansion model instead of m_s.

ACKNOWLEDGMENT

This research was funded by the Natural Science and Engineering Research Council of Canada.

REFERENCES

1. P. N. Seneviratne and J. F. Morrall. Analysis of Factors Affecting the Choice of Route of Pedestrians. *Transportation Planning and Technology*, Vol. 10, 1985, pp. 147–159.
2. M. V. A. Bondada. Intra-CBD Secondary Travel Patterns of Downtown Workers. *Transportation Engineering Journal of ASCE*, Vol. 108, No. TE1, Jan. 1982.
3. B. D. Greenshields. A Study of Highway Capacity. *Proc., Highway Research Board, Vol. 14*, National Research Council, Washington, D.C., 1935, pp. 448–477.
4. H. Greenberg. A Mathematical Analysis of Traffic Flow. *Tunnel Traffic Capacity Study*. Port Authority of New York, 1958.
5. J. J. Fruin. *Pedestrian Planning and Design*. Metropolitan Association of Urban Designers and Environmental Planners, Inc., New York, 1971.
6. F. P. D. Navin and R. J. Wheeler. Pedestrian Flow Characteristics. *Traffic Engineering*, June 1969, pp. 30–36.
7. C. A. O'Flaherty and M. H. Parkinson. Movement on a City Centre Footway. *Traffic Engineering & Control*, Feb. 1972, pp. 434–438.
8. G. S. Rutherford. Use of the Gravity Model For Pedestrian Travel Distribution. In *Transportation Research Record 728*, TRB, National Research Council, Washington, D.C., 1979, pp. 53–58.
9. B. Pushkarev and J. Zupan. Pedestrian Travel Demand. In *Highway Research Record 355*, HRB, National Research Council, Washington, D.C., 1971, pp. 37–53.
10. J. Behnam and B. G. Patel. A Method of Estimating Pedestrian Volume in a Central Business District. In *Transportation Research Record 629*, TRB, National Research Council, Washington, D.C., 1977, pp. 37–53.
11. S. E. Davis, L. E. King, and H. D. Robertson. Predicting Pedestrian Crosswalk Volumes. *Transportation Research Record 1168*, TRB, National Research Council, Washington, D.C., 1988, pp. 25–30.
12. P. N. Seneviratne and J. F. Morrall. Estimation of Flow for Classifying and Evaluating Pedestrian Facilities. Presented at the Annual Conference of the Road and Transport Association of Canada, St. John's, Newfoundland, Sept. 1990.
13. R. Haynes. Sampling Time and Sampling Error in Pedestrian Counts. *Traffic Engineering & Control*, Feb. 1977, pp. 72–74.
14. P. N. Seneviratne and J. F. Morrall. Estimation of Pedestrian Flow & Facility Size. Presented at the Road and Transport Association of Canada, Edmonton, Canada, Oct. 1983.
15. M. Javid. *Evaluation of Pedestrian Flow and Safety*. M. Eng. thesis, Dept. of Civil Engineering, Concordia University, Montreal, Quebec, 1989.
16. J. R. Benjamin and C. A. Cornell. *Probability, Statistics and Decision for Civil Engineers*. McGraw-Hill, New York, 1970.
17. H. S. Mahmassani and K. C. Sinha. Bayesian Updating of Trip Generation Parameters. *Transportation Journal of ASCE*, Vol. 107, No. TE5, Sept. 1981.
18. H. Raiffa and R. Schlaifer. *Applied Statistical Design Theory*. Harvard University Press, Cambridge, Mass., 1961.

Publication of this paper sponsored by Committee on Pedestrians.