Five-Minute-Cluster Sampling for Determining Urban Traffic Volumes

WARREN T. ADAMS, *Chairman,*
Research Projects Subcommittee,
Highway Research Board Urban Volume Characteristics Committee

● THE principal objective of the Urban Volume Characteristics Committee has been to establish urban traffic volume characteristics for the varying needs and interests of the cities and states. Principal among these characteristics are those required for planning and designing urban street and highway systems, moving traffic and determining roadway usage. To satisfactorily evaluate these characteristics it is essential to develop the most efficient and economical methods for obtaining and analyzing the necessary urban volume data.

Acceptable sampling methods for gathering the required data for specific uses and under specific conditions must first be established. Time and cost will not permit the determination of the needed urban volume characteristics through obtaining continuous count information in many localities.

Practicable and economical field sampling techniques, as a reliable measure of urban volumes for these needs, will first have to be developed. These must be based on sound statistical principles and procedures. The objective of this study is to show that random sampling can be modified by cluster sampling so as to produce a sampling method with these qualifications.

Most of the urban volume data for this study was furnished by the Traffic Audit Bureau of New York City through the interest and courtesy of its Managing Director, Mr. V. H. Pelz, a member of the Committee. In 1948 the Traffic Audit Bureau conducted a series of traffic counts by five minute periods for 18 hours in 23 cities. The counts were made at principal intersections and the volume recorded by direction. These cities are scattered throughout the United States and represent most of the geographical and climatic regions. In size they range from New York City, Boston and Detroit down to smaller cities like Greenville, South Carolina, Mobile, Alabama, and Cedar Rapids, Iowa.

The tables of raw data were prepared by the Traffic Audit Bureau and in the beginning of 1954 were made available to the Urban Volume Characteristics Committee. The data was tabulated by both 12 and 18 hour periods. Analyses have been made of both the 12 and 18 hour urban volume counts for many of these cities, but the results presented in this paper will be limited to those of the 12 hour period. For the 18 hours, volumes ranged between 3,000 and 21,000, and the ratio of 12 hour to 18 hour totals varied from 70% to 80%.

The cluster sampling method as used permits the preselection of stations so that field observers will be able most economically and efficiently to make observations for four to six stations per hour. Experience has shown that where this number of short counts can be made per hour this method may be more economical and in combination with other factors more practical than machine counts.

Cluster sampling as defined by Yates is "sampling in which the sampling units are aggregates or 'clusters' of natural units". (*1*) Summarized herewith are the principles governing the cluster sampling method used for this study. The Appendix contains a detailed description of the method used.

1. For this study it has been assumed that travel time between adjacent stations would require 5 minutes due to data being recorded by 5 minute intervals.
2. The interval in the first hour for each sample was selected by using a series of random numbers.
3. (a) If the interval drawn for observation in the first hour were an odd 5 minute interval, it was assumed that: (1) travel between stations would be on the even 5 minute intervals in each succeeding hour and (2) volumes for that sample would be observed on whatever odd 5 minute interval should be selected for each hour.

(b) If the interval drawn for observation in the first hour were an even 5 minute interval, then it was assumed that: (1) travel between stations would be on the odd 5 minute intervals in each succeeding hour and (2) volumes for that sample would be observed on whatever even 5 minute interval should be selected for each hour. (Thus for a 6 station schedule per hour each 12 hour cluster sample would be made up either solely of odd 5 minute interval samples or solely of even 5 minute interval samples.)

4. The observation interval selected in each of the succeeding hours was based on a formula which would permit the unbiased selection of the interval by use of random numbers but at the same time limit travel to 5 minutes to permit taking 5 minute samples at 6 stations per hour or 10 minute samples at 4 stations per hour by each field team.

Initially all analyses of this traffic data were made by selecting 5 minute sample intervals of each hour for the 12 hour period. By using a standard table of random numbers, 72 random samples of one 5 minute interval for each of the 12 hours were selected for each observation point count, and the sample means and the standard deviation computed for each of these locations. Then 72 cluster samples of one 5 minute interval for each of the 12 hours were selected for each urban traffic count and the sample means and the standard deviations determined. Later similar random and cluster sample analyses were made for sample intervals of ten consecutive minutes in each hour for those cities where the traffic count at observation points was less than 6,000 vehicles for the 12 hour period.

Table I shows for cities in which the observation point count is greater than 6,000 vehicles for 12 hours the comparative results of the sample mean and the standard deviation for random and cluster samples. This table indicates that there is only a slight difference in the sample means and in the standard deviations computed for these two sample series. For these observed volumes it should be noted that the standard deviation for cluster sampling varies from 4.3% to 7.0% of the sample mean, with most of them between 5.0 and 6.5%. This indicates that, 95 times out of 100, one should expect due to chance a sample mean with an error not greater than twice the standard deviation for that sample.

Table II similarly presents the sample mean and the standard deviation for the random and cluster samples by 10 minute

TABLE I
COMPARISON OF STATISTICAL ANALYSIS OF URBAN TRAFFIC BY RANDOM AND CLUSTER SAMPLING FOR 5 MINUTES IN EACH HOUR
(12 Hour Period—6 AM to 6 PM)

Location	Mean Volume per Hour			Standard Deviation			
				Random sample		Cluster sample	
	Pop.	Random sample	Cluster sample	SD	% SM	SD	% SM
Detroit	1205	1201	1209	76	6.3	74	6.1
Boston	523	526	522	30	5.7	26	5.0
Manhattan (NYC)	501	500	499	26	5.2	28	5.6
Houston	651	642	652	35	5.5	31	4.8
Jersey City	731	730	733	48	6.6	49	6.6
Atlanta	646	646	645	28	4.3	28	4.3
Buffalo	609	606	609	34	5.6	28	4.6
Denver (inbound)	531	530	532	33	6.2	29	5.5
Denver (outbound)	494	490	498	27	5.5	35	7.0
Oklahoma City	501	502	504	36	7.2	31	6.2
Greenville, S. C.	609	615	611	40	6.5	38	6.2

TABLE II
COMPARISON OF STATISTICAL ANALYSIS OF URBAN TRAFFIC BY RANDOM AND CLUSTER SAMPLING FOR 10 MINUTES IN EACH HOUR
(12 Hour Period—6 AM to 6 PM)

Location	Mean Volume per 2 Hours			Standard Deviation			
				Random sample		Cluster sample	
	Pop.	Random sample	Cluster sample	SD	% SM	SD	% SM
Little Rock	722	719	728	66	9.2	63	8.7
Johnstown, Pa.	688	687	687	35	5.1	34	5.0
Helena, Montana	316	317	314	17	5.4	24	7.6
Mobile	568	567	568	33	5.8	35	6.2
Wichita	723	723	724	39	5.4	38	5.3
Cedar Rapids	873	872	875	59	6.8	57	6.5
Sioux Falls (inbound)	438	442	438	26	5.9	27	6.2
Amarillo (inbound)	633	632	633	31	4.9	29	4.6
Amarillo (outbound)	610	609	608	41	6.7	38	6.3
Steubenville	551	554	552	24	4.4	30	5.5
Pocatello (outbound)	423	422	424	26	6.2	22	5.2

periods for cities where the recorded traffic count was less than 6,000 for 12 hours. This table also brings out that the difference between the sample means and the standard deviations for random and cluster samples is small. For these cluster samples the standard deviation ranges from 4.6 to 8.7% of the sample mean. As in Table I most of them are between 5.0 and 6.5%. Therefore, due to chance one should expect 95% of the time a sample mean with an error not exceeding 13% for those with standard deviation of 6.5% and not more than 10% for those with standard deviation of 5.0%.

The travel time between stations observed is a controlling factor in the efficient utilization of this cluster sampling procedure. Generally it is not feasible to make use of this method of obtaining adequate traffic data if the transportation time between stations exceeds 5 minutes. Where volumes are large, that is between 15,000 and 20,000 vehicles per 12 hour period, it is quite possible that the observation period may be reduced to 4 or 3 minutes to obtain similar results. This will depend on other factors as traffic signal cycle length, number of lanes at a specific point, and time of day. If the observation time can be so reduced, the travel time between stations can be correspondingly increased to 6 or 7 minutes. In practice, if four to six stations per hour are being sampled, the observation time, required for acceptable results, at each station may be different, depending on the volume. This would permit greater flexibility in arranging schedules.

Data was not available to attempt to establish the most efficient length of the short count that would give acceptable results. This would, however, vary from place to place. Volume, travel time between stations, the number of stations, and the objectives will largely determine the sample length. For example, 6 minute and 12 minute samples may produce more reliable estimates than 5 and 10 minute samples respectively where travel time can be reduced to 4 or 3 minutes. Moreover, the use of 6 and 12 minute samples would ease the burden of computation.

There apparently are minimum volume limits to the use of this sampling method in relation to its acceptable accuracy. Where volumes are below 3,000 vehicles per 12 hour period, the study indicates it is questionable if adequate samples for reliably estimating volumes can be obtained by 10 minute counts. If the total observation and travel time averages more than 15 minutes per station, it is doubtful that this method can economically be justified in comparison with machines, if volumes only are considered. There are, however, other advantages in addition to cost in using the manual method of obtaining adequate samples. One of these is the necessity for gathering classification, turning, and directional data which cannot be recorded by machines.

The accuracy of this sampling method depends upon the training and interest of the observers. Machine counts, however, are not infallible. In addition to the inaccuracies due to operating factors, some of which may be compensating, there are also the errors due to mechanical failures which frequently may not be caught until an appreciable lapse of time.

This method could be extended to the estimating of traffic volumes for the 24 hours of the day, although the results may be slightly less accurate because of smaller volumes during the hours not covered by this study. Conceivably the method of cluster sampling could also be used in determining average daily total volumes (ADT) by either spreading the few minute counts throughout the year or sampling 24 hour periods or periods of some other durations in various sequences.

In rural traffic count experience it has been found that estimates based on the factorization of single samples is the predominantly practical method at the present time. Experiments with urban traffic counts in the application of the method of expansion of single samples into ADT by means of factors are being carried on by Mr. Petroff of the Bureau of Public Roads.

Limited studies have also indicated that the same principle might be used in estimating transit data and maximum hour volumes. The studies on applying cluster sampling have not been conclusive, but the work that has been done to date has shown that an adequate cluster sampling technique can probably be developed for gathering needed transit data and maximum hour traffic.

The use of short counts for estimating urban volumes is not new. Several studies have been made in this field. Among these are those made in North Carolina under the direction of Mr. J. S. Burch (*2*), in Indiana by Mr. V. H. Pelz of the Traffic Audit Bureau and Mr. Francis White of the Highway Department (*3*), and in Puerto Rico. (*4*) To our knowledge, however, this is the first attempt to use a method that would permit the computation of the statistical probability of the error in estimating urban volumes due to chance only without any additional mathematical assumptions except those involved in the probability theory. Only through an acceptable satisfactory sampling procedure can the necessary data be obtained for determining urban volume characteristics and volume relationships to the factors affecting transportation. At present there is insufficient knowledge to develop the most efficient stratified sampling methods for various purposes. This study clearly indicates, however, that the cluster sampling method can be efficiently and economically used under certain conditions for reliably estimating urban traffic volumes within acceptable limits of error.

REFERENCES

1. Frank Yates, Sc.D., FRS., "Sampling Methods for Censuses and Surveys", London, Charles Griffin & Co. Ltd; New York, Hafner Publishing Co., 1949, page 20.
2. James Burch, "Total Travel in North Carolina Municipalities", 1951 Proceedings of the Highway Research Board.
3. Francis White and V. H. Pelz, "Short Count Results in a Traffic Survey Made in Fort Wayne, Indiana", Traffic Engineering, August 1948.
4. "Application of the Vickery Short Count Method to Urban Traffic Volume Counts in Puerto Rico", not published; prepared in 1951 by Bureau of Highways, Department of Public Roads, Commonwealth of Puerto Rico.

APPENDIX

METHOD OF PRESELECTING STATION NUMBER ORDER FOR FIELD CLUSTER SAMPLING OF TRAFFIC VOLUMES

a. *For an Assumed Maximum Travel Time of Five Minutes between Adjacent Stations*

1. *With five minute sample intervals per hour per station*

This would permit counting at six different stations per hour and is the basis used in obtaining the sample data for the analyses summarized in Table I. In this the following assumptions were also made:

If the odd five minute intervals should be selected for counting traffic, then travel between adjacent stations would be on even five minute intervals;

If even five minute intervals should be selected for counting traffic volumes, then travel time between adjacent stations would be on the odd five minute intervals.

a) Station Preselection Procedure for Counting on *Odd* Five Minute Intervals

Step No. 1, 1st Hour. Using the number series of 1, 2, 3, 4, 5 and 6 to represent the six observation stations, select, by means of a standard table of random numbers, one of these station numbers to determine which station should be checked the 1st five minute period of the first hour. The 3rd, 5th, 7th, 9th and 11th five minute periods in the first hour will be assigned by station number order to the other five stations in the group. (For example, if station No. 2 is drawn by random for the 1st five minute period, station No. 3 will be checked in the 3rd five minute period, station No. 4 in the 5th, station No. 5 in the 6th, station No. 6 in the 9th and station No. 1 in the 11th, shown in Sample A below.

Step No. 2, Succeeding Hours. To keep travel time to a minimum of five minutes, the 1st five minute period in each succeeding hour will be assigned through selecting by use of a standard table of random numbers one of the following three prescribed stations: either the station last checked in the preceding hour (No. 1 in the preceding example) or one of the adjacent stations on either side (No. 6 and 2 in the above example). As in Step No. 1, the 3rd, 5th, 7th, 9th and 11th five minute periods in the second hour will then be assigned to the other five stations in the group by station number order.

For example, it will be assumed:

1) That station 1 was the station checked in the last five minute period of the preceding hour and that stations 6 and 2 are adjacent stations on each side of station 1;
2) That station No. 6 of the three described stations in this example has been selected by a table of random numbers to be checked on the first five minutes of the succeeding hour;
3) Then station No. 1 would be checked in the 3rd five minute period, station No. 2 in the 5th, station No. 3 in the 7th, station No. 4 in the 9th, and station 5 in the 11th five minute period of that hour.

This method of selection will continue from hour to hour in the total period. If the six sta-

5 Min. Obs. Int.	Sample A						Sample B					
	1st	3rd	5th	7th	9th	11th	1st	3rd	5th	7th	9th	11th
Hour	Station preselected for each five minute observation interval											
1	2	3	4	5	6	1	6	1	2	3	4	5
2	6	1	2	3	4	5	6	1	2	3	4	5
3	5	6	1	2	3	4	4	5	6	1	2	3
4	5	6	1	2	3	4	2	3	4	5	6	1
5	4	5	6	1	2	3	1	2	3	4	5	6
6	2	3	4	5	6	1	5	6	1	2	3	4
7	2	3	4	5	6	1	3	4	5	6	1	2
8	6	1	2	3	4	5	1	2	3	4	5	6
9	6	1	2	3	4	5	6	1	2	3	4	5
10	4	5	6	1	2	3	4	5	6	1	2	3
11	3	4	5	6	1	2	3	4	5	6	1	2
12	1	2	3	4	5	6	3	4	5	6	1	2

tion sites are appropriately selected and numbered so that the total travel time between adjacent stations averages five minutes or less, this will be a feasible method for the field staff to obtain counts for the odd five minute periods of each hour that have been thus randomly assigned to the six stations in the group.

b) Station Preselection Procedure for Counting on *Even* Five Minute Intervals

A similar procedure would be followed when the even five minute periods of each hour are used in lieu of the odd five minute periods for observation.

2. *With ten minute sample intervals per hour per station*

This would permit counting at four different stations per hour and is the basis used in obtaining the sample data for the analyses summarized in Table II for those stations in which the 12 hour volume was less than 6,000.

a) Station Preselection Procedure for Ten Minute Interval Samples

5 Min. Obs. Int.	Sample D					
	1st	3rd	5th	7th	9th	11th
Hour	Station preselected for each five minute observation interval					
1	2	3	4	5	6	1
2	1	2	3	4	5	6
3	6	1	2	3	4	5
4	5	6	1	2	3	4
5	4	5	6	1	2	3
6	3	4	5	6	1	2
7	2	3	4	5	6	1
8	1	2	3	4	5	6
9	6	1	2	3	4	5
10	5	6	1	2	3	4
11	4	5	6	1	2	3
12	3	4	5	6	1	2

Step No. 1, 1st Hour. The number series 1, 2, 3 and 4 was used to represent the four observation stations. One of these was selected by means of a standard table of random numbers to be counted in the first and second intervals of the first hour. Then in station order, allowing five minutes for travel between each station, the remaining three stations would be counted respectively in the 4th and 5th, the 7th and 8th, and the 10th and 11th five minute intervals.

If the initial ten minute observation period should consist of the 2nd and 3rd five minute intervals of the first hour, then in order in the hour the remaining three stations would be counted respectively on the 5th and 6th, the 8th and 9th, and the 11th and 12th five minute intervals.

Step No. 2, Succeeding Hours. In Step No. 2 the first station to be counted in the first two five minute intervals in each succeeding hour would be selected by the same method as in Step No. 2 under 1 (a). The other three stations

5 Min. Obs. Int.	Sample C					
	1st	3rd	5th	7th	9th	11th
Hour	Station preselected for each five minute observation interval					
1	5	6	1	2	3	4
2	3	4	5	6	1	2
3	1	2	3	4	5	6
4	5	6	1	2	3	4
5	3	4	5	6	1	2
6	1	2	3	4	5	6
7	5	6	1	2	3	4
8	3	4	5	6	1	2
9	1	2	3	4	5	6
10	5	6	1	2	3	4
11	3	4	5	6	1	2
12	1	2	3	4	5	6

5 Min. Obs. Int.	Sample E											
	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th	11th	12th
Hour	Station preselected for each five minute observation interval											
1	4		5		6		1		2		3	
2		3		4		5		6		1		2
3	2		3		4		5		6		1	
4		1		2		3		4		5		6
5	6		1		2		3		4		5	
6		5		6		1		2		3		4
7	4		5		6		1		2		3	
8		3		4		5		6		1		2
9	2		3		4		5		6		1	
10		1		2		3		4		5		6
11	6		1		2		3		4		5	
12		5		6		1		2		3		4

would be sampled in numerical order for two consecutive five minute intervals, with five minute travel time intervening.

b. For an Assumed Maximum Travel Time of Five Minutes between Any Two of the Stations

1. *With five minute sample intervals per hour per station*

a) Station Preselection Procedure

Step No. 1, 1st Hour. If the maximum travel time between any stations is five minutes, then step No. 1 will be the same as in "a 1".

Step No. 2, Succeeding Hours. In step No. 2, step No. 1 above will be repeated for each hour of the total period. Since the travel time between any two of the stations would not be more than five minutes, any one of the six stations can be chosen by use of the table of random numbers for sampling the first odd or even five minute interval in each hour, rather than be limited to the selection of one of the three prescribed stations as in step No. 2 in "a 1".

2. *With ten minute sample intervals per hour per station*

a) Station Preselection Procedure

Step No. 1, 1st Hour. In this step the first station to be counted would be determined in the same manner as in Step No. 1 under "a 2".

Step No. 2, Succeeding Hours. In step No. 2 as any two of the stations would not be more than five minutes apart, any one of the four stations might be selected by use of the table of random numbers as the first station in each succeeding hour, following the last station counted in the preceding hour. The remaining three stations would be counted in numerical order for two consecutive five minute periods with five minute travel time intervening.

USE OF RANDOM NUMBER SERIES IN THE PRESELECTION OF STATION ORDER

It can be expected that the use of random sampling will give a larger standard deviation most of the time than with stratified sampling when knowledge makes stratification feasible. Since random sampling is dependent on pure chance it can be expected that preselection of station order would vary from sample to sample or from random series to random series. The two samples (A and B) obtained by a table of random numbers for six stations over a 12 hour period with five minute travel time between each five minute observation, as described in "a 1", illustrate this variation.

Several modifications of this method have been tested. In each the primary purpose has been either to simplify procedure or to reduce travel time. Each of these modifications would be one of the possible random samples previously illustrated in Samples A and B. Tests have indicated that the expected results would probably be as acceptable as those obtained in Tables I and II. Illustrations of these are given in Samples C, D and E.

Sample C is an illustration of one simplified method. In this, the observer in the 1st. observation interval of each hour would always record traffic volumes at the next to the last station checked in the preceding hour.

Samples D and E are illustrations of station preselection procedures for reducing total time used for traveling. In Sample D the observer will have a break period of five minutes at the end of each hour, and the station observed in the first odd interval of each hour would always be the station checked in the last odd interval in the preceding hour. If even 5-minute intervals are used for observation, the station checked in the first even interval of each hour would be the same station observed in the last even interval of the preceding hour.

In Sample E, as in Sample D, the station observed in the first interval in each hour would always be the same one checked in the last interval in the preceding hour. In Sample E, however, it will be noted that when the station is observed in the first even five minute interval of each hour, it is the same station checked in the last odd five minute interval of the preceding hour, thus giving a ten minute break in recording. When, however, the station is observed in the first odd five minute interval of each hour, it is the same station checked in the last even five minute interval of the preceding hour, thus requiring ten continuous minutes of recording.