Appraisal of Sample Size Based on Phoenix O-D Survey Data

ARTHUR B. SOSSLAU and GLENN E. BROKKE, Highway Research Engineers, Division of Highway Planning, U.S. Bureau of Public Roads

The objective of this research was to determine the relationship between home-interview type of O-D survey sampling rates and the error that can be expected in volumes accumulated from a survey. The data analyzed were taken from the home-interview phase of the 1957 Phoenix-Maricopa County traffic study.

Inasmuch as individual zone-to-zone movements have insufficient volume to be reliable indicators, a method called "trip tracing" was employed to accumulate individual zone-to-zone trip volumes for statistical study. By this method, trips are traced across the survey area in a straight line between origin and destination and accumulated in ¼-mi sections of a grid system superimposed over the survey area. An electronic computer was utilized for the trip-tracing procedure and for much of the statistical computations required for the study.

The trip data from the Phoenix-Maricopa survey were systematically sampled to provide ½, ⅓, and ⅒ subsamples of the original 1-in-15 dwelling-unit sample. The expanded results of the original survey and each of the subsamples were individually processed through the trip tracing program. This procedure provided the data necessary to determine the accuracy that could be expected in any volume due to sampling.

The results indicate that the accuracy of accumulated trip volumes is considerably less than the accuracy predicted by a purely theoretical approach. However, the results agree qualitatively with theory in that the accuracy of trip volumes varies with the square root of sample size and very nearly with the square root of the volume. The results of this research should allow the selection of a sample rate commensurate with the funds available and the degree of accuracy required.

●TRAVEL BY individuals has the characteristic of being habitual. Furthermore, the travel habits of different individuals are similar for the many types of trips generally made. For these reasons, sampling procedures, such as the home-interview type of O-D survey developed by the Bureau of Public Roads, can be used to determine travel information for a metropolitan area.

The accuracy of information developed from a home-interview survey, and indeed from any type of sampling procedure, is to an important degree dependent on the sampling rate used. Sampling rates employed in O-D surveys have usually been based on population and area. In a large, densely populated area a smaller sampling rate is used than in a small, less densely populated area. The sampling rate determined is based, predominantly, on the number of interviews needed to provide a reliable representation of the over-all travel in the city. The zone-to-zone movements developed, however, have been used to present a picture of travel desires and to aid in the location of needed transportation facilities.

The results of an O-D survey are also accumulated on, or assigned to, proposed highway facilities for design considerations. How accurate are these accumulated values that have been expanded from the survey sample to represent average daily movements? The purpose of this study was to determine the relationship between the home-interview type of O-D survey sampling rates and the accuracy of volumes that are accumulated therefrom.

Measures of the accuracy of accumulated volumes have been determined for various sampling rates and predictive equations developed for determining the sampling rate needed for a desired accuracy at any volume. Because sample rate affects the accuracy of volumes used for design purposes, an estimate can be made of the frequency that facilities so designed will be loaded beyond capacity or conversely, the frequency that such facilities will have more lanes than are actually required.

SOURCE OF DATA

The data used for this study were developed from the home-interview phase of the 1957 Phoenix-Maricopa County, Ariz., traffic survey.

The survey area covered 225 sq mi and contained a population of 397,000 persons. Of this, the City of Phoenix occupied 36.3 sq mi and included 192,500 persons. The survey area was divided into 135 zones, which were smaller in area in the downtown section of Phoenix and larger in the outlying areas.

The sample rate used in the home-interview phase of the O-D survey was 1 in every 15 dwelling units, or 6.67 percent. This rate resulted in a total sample of 8,743 dwelling units. The trips analyzed were those having both origin and destination within the survey area. All trips, regardless of mode of travel, were considered. The total expanded number of internal trips was 839,398. Separated by mode, it was as follows:

Automobile drivers	548,439
Automobile, truck and taxi passengers	233,536
Bus passengers	57,423

The survey personnel had checked the expanded household data against the 1950 decennial census and the 1953 special census with satisfactory results. Also the expanded travel data had been compared with two screenline ground counts made during the survey. The survey data accounted for 89.3 percent of the vehicles crossing the screenlines from 6 a.m. to 10 p.m. on an average weekday.

METHOD OF STUDY

Individual zone-to-zone movements obtained from a home-interview type of O-D survey cannot be used as the basis for sound conclusions regarding sample size due to the predominance of small volume movements between many zones (1). Any sample size, within the limits of economic feasibility, would be too small to produce accurate expanded values for these small movements; but the individual zone-to-zone movements do not necessarily have to be accurate if their summation reasonably represents travel accumulations throughout a city.

A method referred to as trip-tracing (1), has provided the means for determining and checking distribution of travel throughout an area. By this method zone-to-zone movements are traced across a city in a straight line between origin and destination. After the equation of the straight line between the two zones is obtained, the points of intersection of this line with previously established gridlines are determined. The volume of trips being traced is added to volumes previously accumulated, for other zone-to-zone movements, at the section of the gridline being crossed. After all zone-to-zone movements have been similarly traced across the city, the result is the accumulated number of trips per section of gridline.

The length of the sections of gridline, in which the volumes are accumulated, is chosen in accordance with the size of the accumulations desired. If large accumulated volumes are desired, large sections are used, and vice versa.

A grid system was superimposed over the Phoenix-Maricopa County traffic study area (Fig. 1). The southwest corner of the city was designed as origin of the rectangu-

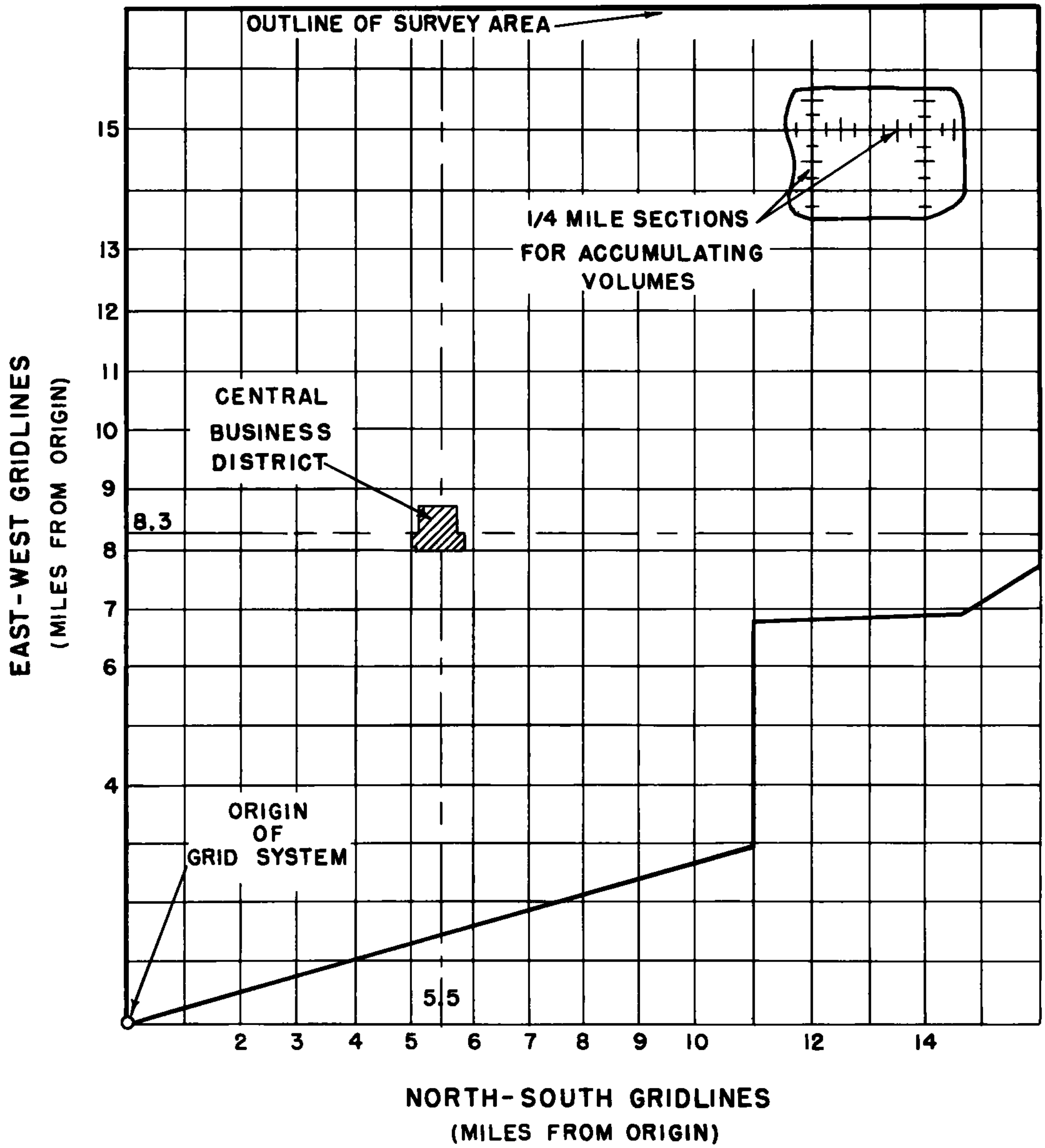

Figure 1. Design of grid system.

lar coordinate system. North-south gridlines were placed at 2, 3, 4, 5, 5.5, 6, 7, 8, 9, 10, 12, and 14 mi from the origin and east-west gridlines at 4, 6, 7, 8, 8.3, 9, 10, 11, 12, 13, and 15 mi from the origin. The closer spacing of gridlines was at the more densely populated part of the study area. Each gridline was broken into ¼-mi sections for volume accumulation purposes. The north-south gridlines were 17 mi long and the east-west lines, 16 mi long. This resulted in 1,520 ¼-mi sections (17 mi x 4 sections per mile x 12 gridlines) + (16 mi x 4 sections per mile x 11 gridlines). The volumes accumulated on the ¼-mi sections ranged from 0 to 35,000 trips.

Features of the procedure that have been described are as follows:

1. The result of the trip-trace accumulation is a spatial distribution of trips throughout the city representing the travel desires of the population. The trips are traced in

such a manner that each zone-to-zone movement is made by the most desirable path—a straight line.

2. Long trips are weighted more heavily than short trips because more gridlines are crossed. This is analogous to a long trip using more of a road network than a shorter trip.

3. The resulting accumulation of trips presents a picture of travel analogous to accumulations of trips on a street system.

The trip-tracing procedure is too lengthy for manual computation methods. It is, however, readily adaptable to computation on an electronic computer of medium size. Therefore, a computer program was developed to trace the Phoenix-Maricopa County trips across the study area. Computer programs were also developed to handle the

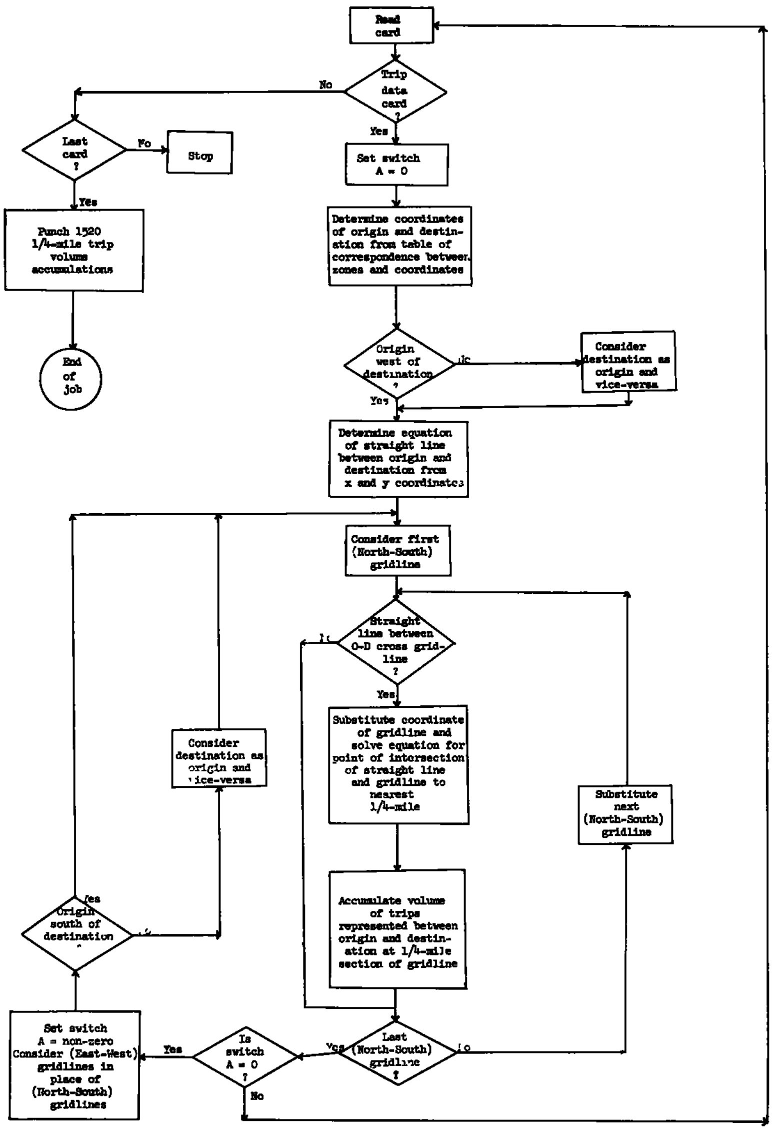

Figure 2. Generalized electronic computer flow chart for tracing trips across a grid system.

various statistical computations that were necessary for this study. A simplified flow chart of the trip-tracing program is shown in Figure 2. Generally, the following computations are carried out:

1. From the coordinates of the two zones being handled, determine the equation of the straight line passing between them.
2. Determine, to the nearest quarter of a mile, the points of intersection of the line and the north-south and east-west gridlines.
3. Add the number of trips being traced to the volumes previously accumulated in computer memory slots representing the 1,520 ¼-mi sections.

All of the internal Phoenix-Maricopa County home-interview trip cards, representing the expanded results of the 1-in-15 dwelling-unit sample, were processed through the trip-tracing program to determine the spatial distribution of trips throughout the survey area. The original deck of cards was then sorted by sample number into 2 one-half subsamples, each representing a 1-in-30 dwelling-unit sample. Each one-half subsample was then separately run through the computer to determine the spatial distribution of the two 1-in-30 dwelling-unit samples. Similarly, the original deck was systematically stratified by sample number into 3 one-third subsamples, and 10 one-tenth subsamples, and each subsample processed through the trip-tracing program.

STATISTICAL PROCEDURE

Running of the total Phoenix-Maricopa County traffic survey data resulted in accumulated average daily volumes, in the 1,520 ¼-mi sections, ranging from 0 to 35,000 person-trips. Similarly, each of the subsamples processed through the trip-tracing program resulted in accumulated volumes in each section. It should be noted, that, as each subsample zone-to-zone movement was traced, the volume was expanded to represent actual movement. For example, as each one-third subsample trip card was processed, the trip factor on the card was multiplied by 3.

The data resulting from the trip-tracing program were analyzed by comparing, on a ¼-mi section basis, the expanded subsample accumulations against the total sample accumulations. It was not, however, considered practical to report the error for each section. Instead, the 1,520 ¼-mi sections were stratified into 15 volume groups, and the individual errors were accumulated and summarized for each volume group. Such a process produced, for each of the three subsamples tested, 15 errors and the average volume at which the error occurred. The range, the number of sections, and the average volume at each volume group are given in Table 1. The sections of gridline were stratified into volume groups in accordance with the volumes accumulated in the sections from the expanded total sample.

The summarization of results, per volume group, for each subsample consisted of determining the differences in volumes accumulated per section from the original sample and the subsample, squaring the differences and accumulating the results of the squaring. The resulting summation was then divided by the number of sections in the volume group and the square root of the quotient taken. The result of this procedure is the root-mean-square error (RMS error) of the subsample as compared with the total sample. The equation for determining this error is:

$$\text{RMS error} = \sqrt{\frac{\sum_{i=1}^{i=n} (V_{ss} - V_s)^2}{n}} \tag{1}$$

in which

RMS error = root-mean-square error
V_{ss} = volume accumulated in section i from subsample
V_s = volume accumulated in section i from total sample
n = number of sections in volume group

TABLE 1

VOLUME GROUP SUMMARY

Volume Group No.	Volume Group Range (Average Daily Person Trips)			Number of Sections Having Volumes Within Range[1]	Mean Volume of Group[2] (Average Daily Person Trips)
1	000.1	-	199.9	268	91.1
2	200.0	-	399.9	88	283.2
3	400.0	-	599.9	69	486.7
4	600.0	-	799.9	45	682.0
5	800.0	-	999.9	50	893.3
6	1,000.0	-	1,499.9	80	1,244.5
7	1,500.0	-	1,999.9	56	1,763.4
8	2,000.0	-	2,499.9	58	2,257.4
9	2,500.0	-	3,499.9	86	2,998.6
10	3,500.0	-	4,499.9	62	4,001.4
11	4,500.0	-	5,499.9	42	5,020.9
12	5,500.0	-	7,999.9	67	6,759.3
13	8,000.0	-	9,999.9	41	8,961.8
14	10,000.0	-	14,999.9	34	12,401.3
15	15,000 and over			28	21,689.9

[1] The number of sections does not equal 1,520 because some sections did not have any trips traced across them.
[2] The mean volume was determined from the total sample.

The RMS error is comparable, statistically, to the standard deviation of a group of values around their mean. For example, if the RMS error for a one-third subsample volume compared with the original sample volume is 50 person-trips, one would make little error by assuming that two-thirds of the expanded volumes obtained from the subsample would lie within 50 person-trips of the total sample volume.

RMS errors were developed for each subsample. That is, a RMS error was developed for each of the one-half subsamples, for each of the one-third subsamples, and for each of the one-tenth subsamples. Little is gained by reporting and analyzing each subsample error. Therefore, a mean error was determined from the 3 one-third subsamples, from the 2 one-half subsamples, and from the 10 one-tenth subsamples.

A RMS error was computed for each volume class, and the "percent root-mean-square error" was determined by dividing the numerical error by the average volume of the volume class being considered. The results of the one-third subsample comparison are given in Table 2.

RESULTS

The results of the trip-tracing program and the statistical procedure used are the percent root-mean-square (percent RMS) errors for the one-half, the one-third, and the one-tenth subsamples, each measured against the total Phoenix-Maricopa County sample. These results are given in Table 3 for each volume group.

As was to be expected, the percent RMS error for any particular volume group is invariably greatest for the smallest subsample rate, and vice versa. For example, for an average volume of 1,763 trips, volume group number 7, the percent RMS error for the one-half subsample is 15.3; for the one-third subsample, 23.5; and for one-tenth subsample 49.4. In addition, the percent RMS error, for each subsample, decreases as the volume increases. These decreases in percent RMS error, as volume increases, approximately follow a straight line if plotted on logarithmic paper.

It should be understood that the one-half subsample errors are in reality the percent RMS errors between a 1-in-30 and a 1-in-15 dwelling-unit sample. Likewise, the

TABLE 2

RESULTS OF ONE-THIRD SUBSAMPLE COMPARISON WITH TOTAL SAMPLE

Volume Group No.	Mean Volume of Group (Total Sample)	Mean Volume[1] 1st One-Third Subsample	Mean Volume[1] 2d One-Third Subsample	Mean Volume[1] 3d One-Third Subsample	Mean Error[2] 1st One-Third Subsample	Mean Error[2] 2d One-Third Subsample	Mean Error[2] 3d One-Third Subsample	RMS Error[3] 1st One-Third Subsample	RMS Error[3] 2d One-Third Subsample	RMS Error[3] 3d One-Third Subsample	Percent RMS Error[4] 1st One-Third Subsample	Percent RMS Error[4] 2d One-Third Subsample	Percent RMS Error[4] 3d One-Third Subsample	Average RMS Error of Three Samples: Absolute Error $\frac{\text{RMS Error 1st + 2d + 3d}}{3}$	Average RMS Error of Three Samples: Percent Error $\frac{\text{Absolute Error}}{\text{Original Sample Volume}}$
1	91.1	83.2	90.6	99.4	-7.9	-0.4	+8.3	77 5	75 1	78.9	85.07	82.44	86.61	77.2	84.74
2	283.2	231.2	309.5	308.9	-52.0	+26.3	+25.7	160.5	153.5	168.3	56.67	54.20	59.43	160.8	56.78
3	486.7	392.9	545.0	522.2	-93.8	+58 3	+35 5	205.6	237.4	225.7	42.24	48.78	46.37	222.9	45.80
4	682.0	623 2	710 9	711 9	-58.8	+28.9	+29.9	307.8	253.6	295.6	45 13	37 18	43.34	285.7	41 89
5	893.3	749.9	977.1	952 7	-143 3	+83.8	+59.5	300 5	306.4	267.8	33.64	34 30	29.98	291.6	32.64
6	1,244.5	1,153 8	1,325 3	1,254.2	-90 6	+80.9	+9 7	308.9	335.2	349.7	24 82	26.93	28.10	331 3	26.62
7	1,763.4	1,772 5	1,883.2	1,634.4	+9.1	+119.8	-128.9	390.8	427.0	424.9	22 16	24 21	24 10	414.2	23 49
8	2,257.4	2,123 7	2,631.0	2,287.5	-133 7	+373.6	+30 1	470 6	537.9	391.4	20 85	23.83	17.34	466 6	20.67
9	2,998.6	2,886 6	3,118.3	2,990.9	-112.0	+119.7	-7 7	513 6	578 4	503.6	17.13	19.29	16.79	531.9	17.74
10	4,001.4	3,938 8	3,995.6	4,069.8	-62.6	-5 8	+68.4	666.1	654 7	644.7	16.65	16.36	16.11	655.2	16.37
11	5,020.9	4,903.9	5,136.4	5,022 5	-117 0	+115 5	+1.6	727.0	667.7	729.5	14 48	13.30	14.53	708.1	14.10
12	6,759.3	6,490.2	6,628 3	7,159.4	-269.1	-131.0	+400 1	774.5	751 7	922.5	11.46	11.12	13.65	816.2	12.08
13	8,961.8	8,896.1	8,923 1	9,066 4	-65 7	-38.8	+104.5	878.5	780.4	910.2	9.80	8.71	10.16	856.4	9.56
14	12,401.3	12,285 2	12,461 4	12,457.2	-116.1	+60.1	+55 9	909.7	913 7	1,156 3	7.34	7.37	9.32	993.2	8.01
15	21,689 9	21,555.2	22,046.2	21,468 2	-134 7	+356.3	-221 7	1,741 0	1,808 9	1,655.6	8.03	8.34	7.63	1,735.2	8.00

[1] Mean volume $= \Sigma \frac{\left(\text{Segment volumes in volume group}\right)}{\text{Number of sections}}$

[2] Mean error $= \Sigma \left[\frac{\left(\text{Subsample volume} - \text{Original sample volume}\right)}{\text{Number of sections}} \right]$

[3] RMS error $= \sqrt{\Sigma \left[\frac{\left(\text{Subsample volume} - \text{Original sample volume}\right)^2}{\text{Number of sections}} \right]}$

[4] Percent RMS Error $= \frac{\text{RMS error}}{\text{Mean volume}}$

one-third subsample rate is actually a 1-in-45 dwelling-unit sample and the one-tenth subsample a 1-in-150 dwelling-unit sample, the error in each case being measured against the 1-in-15 sample.

The values given in Table 3 can be used in the following manner. If it is desired to estimate trips from a home-interview survey at the 4,000 average daily person-trip level, volume group number 10, for some specific design purpose, the use of a 1-in-30 dwelling-unit sample would produce a volume that is within 11.7 percent of the value that would have been obtained with a 1-in-15 sample two-thirds of the time. Likewise, the use of a 1-in-45 dwelling-unit sample would result in a volume that is within 16.4 percent of the value obtained by a 1-in-15 sample two-thirds of the time. If the probability of being within the 1-in-15 sampling rate volume 95 percent of the time is desired, two times the percent RMS error would be used. An expectancy of 99 percent would require three times the percent RMS error.

By using the values given in Table 3, the expected results of a 1-in-30, a 1-in-45, and a 1-in-150 dwelling-unit sample can be compared with that of a 1-in-15 sample. However, the prime purpose of this study was to determine the error between the volume determined from any dwelling-unit sample and the actual volume, the actual volume being the average daily person-trips measured over the study period. Evidently, only through an overwhelming expenditure of time and money could every person in a city be interviewed every day during the study period. However, through statistical procedures, using the comparisons of the 1-in-30, the 1-in-45, and the 1-in-150 dwelling-unit sampling rate with the 1-in-15 sample, an estimate of the error between any size sample and the total population can be determined.

The error between a volume determined from any of the subsamples and the true volume consists of two parts: (1) the error between the subsample volume and the total Phoenix-Maricopa County sample volume, and (2) the error between the total sample volume and the true volume. In statistical computations, for the analysis of variance, the total variance of a group of samples is equal to the "between sample" variance plus the "within sample" variance:

$$\sigma^2_{\text{total}} = \sigma^2_{\text{within}} + \sigma^2_{\text{between}} \quad (2)$$

TABLE 3

PERCENT ROOT-MEAN-SQUARE ERROR OF SUBSAMPLE VOLUME AS MEASURED AGAINST TOTAL SAMPLE VOLUME

Volume Group No.	Percent Root-Mean-Square Error of Subsample Volume Measured Against Total Sample Mean Volume of Group		
	One-Half Original Sample	One-Third Original Sample	One-Tenth Original Sample
1	67.5	84.7	193.0
2	41.4	56.8	120.7
3	28.4	45.8	88.2
4	27.6	41.9	80.2
5	23.1	32.6	67.7
6	18.0	26.6	55.7
7	15.3	23.5	49.4
8	13.4	20.7	45.0
9	14.2	17.7	36.6
10	11.7	16.4	33.8
11	10.2	14.1	29.0
12	9.3	12.1	26.4
13	6.7	9.6	21.4
14	5.4	8.0	18.5
15	3.8	8.0	16.0

in which

σ^2 = variance

The percent RMS errors computed for this study, as mentioned previously, are statistically comparable to the standard deviation of a group of values about their mean. Therefore, an equation for relating percent RMS errors, comparable to Eq. 2 is:

$$E^2_{ss-o} = E^2_{ss-s} + E^2_{s-o} \tag{3}$$

in which

E_{ss-o} = percent RMS error of subsample volume measured against true volume

E_{ss-s} = percent RMS error of subsample volume measured against total sample volume

E_{s-o} = percent RMS error of total sample volume measured against true volume

An equation for estimating the error for a sample from the error found for another independently selected sample is:

$$E_{ss-o} = E_{s-o}\sqrt{\frac{N_s}{N_{ss}}} \tag{4}$$

in which

N_s = number of interviews taken in original survey

N_{ss} = number of interviews represented in subsample

Assuming that this is an acceptable approximation to the situation being considered and substituting Eq. 4 into Eq. 3:

$$\frac{N_s}{N_{ss}} \quad E^2_{s-o} = E^2_{ss-s} + E^2_{s-o}$$

or

$$E^2_{s-o} = \frac{E^2_{ss-s}}{\frac{N_s - N_{ss}}{N_{ss}}}$$

therefore,

$$E_{s-o} = \frac{E_{ss-s}}{\sqrt{\frac{N_s}{N_{ss}} - 1}} \tag{5}$$

Considering Eq. 5 for determining the percent RMS error between the total sample volume and the actual volume from the error determined from the one-tenth subsample, Eq. 5 then becomes:

$$E_{s-o} = \frac{E_{1/10-s}}{\sqrt{\frac{N_s}{1/10\ N_s} - 1}} = \frac{E_{1/10-s}}{\sqrt{10-1}}$$

or,

$$E_{s-o} = \frac{E_{1/10-s}}{3} \tag{6}$$

Likewise, the equations for determining the error of the total sample from the one-half and one-third subsample errors are:

$$E_{s-o} = \frac{E_{1/2-s}}{\sqrt{1}} = E_{1/2-s} \tag{7}$$

$$E_{s-o} = \frac{E_{1/3-s}}{\sqrt{2}} = \frac{E_{1/3-s}}{1.41} \tag{8}$$

Eqs. 6, 7, and 8 were used to determine independent estimates of the percent RMS errors in the total sample volumes, measured from the actual volume, from the one-tenth, one-half, and one-third subsample errors, respectively. If each of these independent calculations produced consistent estimates of the RMS error for the 1-in-15 sample, it appears reasonable to assume that other sampling rates would be equally consistent. These independent estimates of the error in the total sample are given in Table 4.

A comparison of the estimated percent RMS errors of the 1-in-15 dwelling-unit sample, for each volume group, shows little variation. The mean of the three estimates was, therefore, determined and plotted on logarithmic paper (Fig. 3). A least-squares fit determined from the points is also shown in Figure 3. The equation of the developed line is:

$$\text{percent RMS error of 1-in-15 dwelling-unit sample} = \frac{629.0}{\text{volume}^{0.4884}} \tag{9}$$

The coefficient of correlation developed is approximately one, indicating an almost perfect functional relationship between the two variables considered. In other words, the variation in percent RMS error is explained almost entirely by the variation in

TABLE 4

ESTIMATED PERCENT ROOT-MEAN-SQUARE ERROR OF ORIGINAL SAMPLE VOLUME AS MEASURED AGAINST POPULATION VOLUME

Volume Group No.	Estimated Percent Root-Mean-Square Error from One-Half Sample	One-Third Sample	One-Tenth Sample	Mean
1	67.5	59.9	64.3	63.9
2	41.4	47.3	40.2	43.0
3	28.4	32.4	29.4	30.1
4	27.6	29.7	26.7	28.0
5	23.1	23.1	22.6	22.9
6	18.0	18.8	18.6	18.5
7	15.3	15.9	16.5	15.9
8	13.4	14.7	15.0	14.4
9	14.2	12.6	12.2	13.0
10	11.7	11.6	11.3	11.5
11	10.2	10.0	9.7	10.0
12	9.3	8.6	8.8	8.9
13	6.7	6.8	7.1	6.9
14	5.4	5.7	6.2	5.8
15	3.8	5.7	5.4	5.0

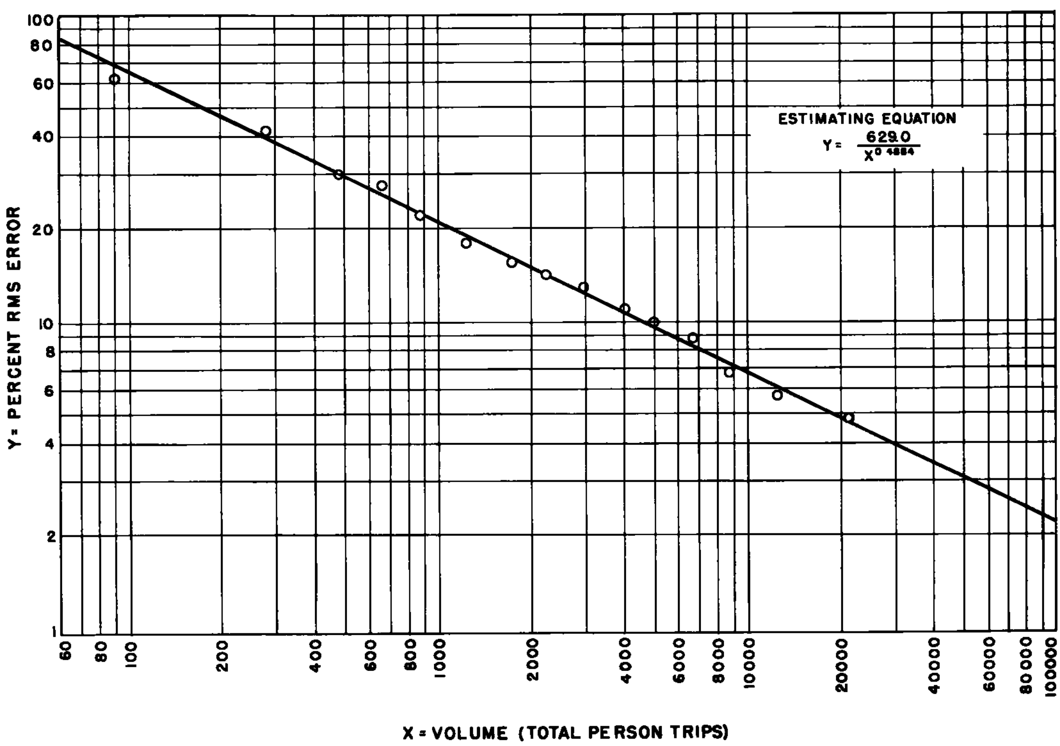

Figure 3. Relation of percent root-mean-square error and total person trips for 1-in-15 dwelling-unit sample.

volume. Because this is a logarithmic relationship the standard error of estimate for the line is a constant percent of the estimated value rather than a constant percent error. The error, which is 5.65 percent of the estimated values, means that at a volume of 1,000 trips, the standard error is about 1.2 percent (5.65 percent x 21 percent) and at 10,000 trips the standard error is about 0.4 percent (5.65 percent x 6.8 percent).

Eq. 4 can be used, in the following form, for determining the relationship between percent RMS error and volume for any sample rate:

$$E_{i-o} = E_{1/15-o} \cdot \sqrt{N_i} \tag{10}$$

in which

E_{i-o} = error of any sample i

$E_{1/15-o}$ = error of 1-in-15 dwelling-unit sample

N_i = number of times the sampling rate of survey i is less than 1 in 15. For example, if 1-in-60 rate is used, N_i would be 4. If the rate was 1 in 5, N_i would be 1/3.

If a 1-in-1 sample were taken, every person in the city interviewed once, the equation for determining percent RMS error would be:

$$\text{percent RMS error} = \frac{629.0 \times \sqrt{N_i}}{\text{volume}^{0.4884}}$$

in which

N_i = 1/15

therefore

$$\text{percent RMS error for 1-in-1 dwelling-unit sample} = \frac{162.4}{\text{volume}^{0.4884}} \tag{11}$$

It should be noted that the error for a 1-in-1 dwelling-unit sample is correctly not zero, because a 1-in-1 sampling rate is not a 100 percent sample for determining average daily traffic during the survey period. Every person in a city would have to be interviewed about his travel for every day during the survey in order to obtain the universe of travel for that period.

Eq. 11 can be used for determining the equation for percent RMS error at any sample rate. Simply multiply Eq. 11 by the square root of the denominator of the sample rate ratio used. For example, if a 1-in-30 home-interview sample rate is used, multiply Eq. 11 by the square root of 30. Figure 4 shows predictive lines for estimating percent RMS errors for volumes between 100 and 100,000 person-trips per day for various sampling rates.

COMPARISON OF RESULTS

A theoretical approach to the problem of estimating the accuracy of various sample sizes relies on the estimation of the standard deviation in volumes determined from the samples. The theory states that the expected deviation (σ) is expressed as follows:

$$\sigma = \sqrt{(\text{volume}) \times \left(\frac{\text{percent sample}}{100}\right) \times \left(1 - \frac{\text{percent sample}}{100}\right)}$$

Inasmuch as the probable volume ($\overline{V}$) is equal to the volume obtained from the survey times the sample rate:

$$\frac{\sigma}{\overline{V}} = \sqrt{\frac{(\text{volume}) \times \left(\frac{\text{percent sample}}{100}\right) \times \left(1 - \frac{\text{percent sample}}{100}\right)}{(\text{volume}) \times \left(\frac{\text{percent sample}}{100}\right)}}$$

or

$$\text{percent } \sigma = 100 \sqrt{\frac{(100 - \text{percent sample})}{(\text{volume}) \times (\text{percent sample})}}$$

A comparison of the errors predicted by the preceding theory with the relationship developed in this paper is given in Table 5 for various volumes and sample sizes. It can be seen that the observed root-mean-square errors are from 1.7 to 1.9 times as great as the error predicted by theory. This difference may be due to nonsampling errors such as response and coding errors. For example, a study made in Cincinnati, Ohio, shows that the respondent reports too few trips in some cases and too many trips in other cases.

TABLE 5

COMPARISON OF OBSERVED ERRORS AND THEORETICAL SAMPLING ERRORS

Sample Rate (%)	Volume	Observed Error (%)[1]	Theoretical Sampling Error (%)[2]	Observed Error / Theoretical Sampling Error
1	100	171.31	99.50	1.7
	1,000	55.64	31.46	1.8
	10,000	18.07	9.95	1.8
	100,000	5.87	3.15	1.9
3	100	98.91	56.87	1.7
	1,000	32.21	17.98	1.8
	10,000	10.43	5.69	1.8
	100,000	3.39	1.80	1.9
4	100	85.65	48.99	1.7
	1,000	27.82	15.49	1.8
	10,000	9.04	4.90	1.8
	100,000	2.94	1.55	1.9
5	100	76.60	43.60	1.8
	1,000	24.90	13.80	1.8
	10,000	8.10	4.36	1.8
	100,000	2.62	1.38	1.9
10	100	51.50	30.00	1.7
	1,000	16.70	9.50	1.8
	10,000	5.43	3.00	1.8
	100,000	1.76	0.95	1.9

[1] This is the percent root-mean-square error as developed in this paper.
[2] This is the theoretical percent standard deviation error.

As Table 5 shows, the observed error is almost two times as great as the theoretical error. This means that in order to maintain a desired degree of accuracy, it is necessary to increase the sampling rate to almost four times the rate indicated by the theoretical standard deviation computation.

USE OF RESULTS

The curves plotted in Figure 4 have been developed for total person-trips and can be utilized to determine the sample rate to be used when desiring an estimate of volume with a desired degree of accuracy. For example, if it is desired to estimate an average daily volume at the 10,000 person-trip level, and be within 8 percent of the true value 95 percent of the time, from Figure 4 at volume equal to 10,000 and percent RMS error equal to 4 percent (8 percent $\div$ 2 for 95 percent confidence), it is found that a

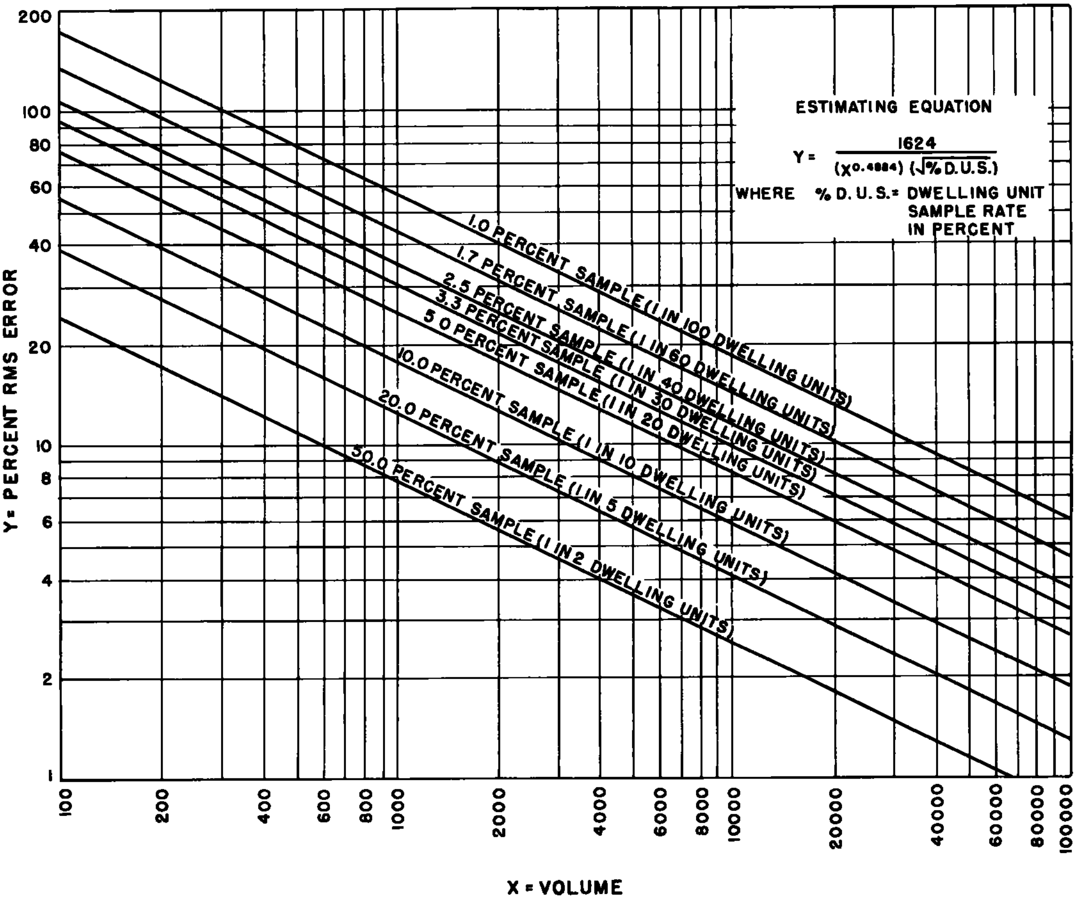

Figure 4. Relation of percent root-mean-square error and volume for various dwelling unit sample rates.

1-in-5 sample should be taken. Similarly, the curves can be used to estimate the error in accumulated volumes after the results of an O-D survey are obtained. For rates not plotted, Eq. 11 would be used in the following form:

$$\text{Dwelling-unit sample rate in percent} = \left[\frac{1,624}{(\text{percent RMS error})\left(\text{volume}^{0.4884}\right)}\right]^2 \qquad (12)$$

For the problem explained, a dwelling-unit sample rate of 20.4 percent would result from the use of Eq. 12 indicating that 1 out of every 5 dwelling units should be interviewed.

More often than not, the highway engineer is interested in the number of vehicle-trips rather than the number of person-trips—vehicle-trips being the figure used for highway design purposes. Therefore, the question is: Can the curves developed for person-trips be used as an indicator of error for vehicle-trips?

The volume of automobile vehicle-trips throughout a city is less than the volume of person-trips, but is similarly distributed. The errors, if developed between the sub-samples and total sample for automobile vehicle-trips, should not, therefore, be any different from the errors determined for all person-trips. That is, a percent RMS error for 10,000 automobile vehicle-trips should be no different from the error determined for 10,000 person-trips. The curves presented in Figure 4 and the equations developed can therefore be used for total person-trips, automobile vehicle-trips, bus-passenger trips, truck- and taxi-passenger trips, and automobile passenger-trips. The results of this study can also be applied to any home-interview type of O-D survey be cause accumulation of trips and not individual zone-to-zone movements are being compared.

REFERENCE

1. Brokke, G. E., and Mertz, W. L., "Evaluating Trip Forecasting Methods with an Electronic Computer." Public Roads, 30:4, 77-87 (Oct. 1958).