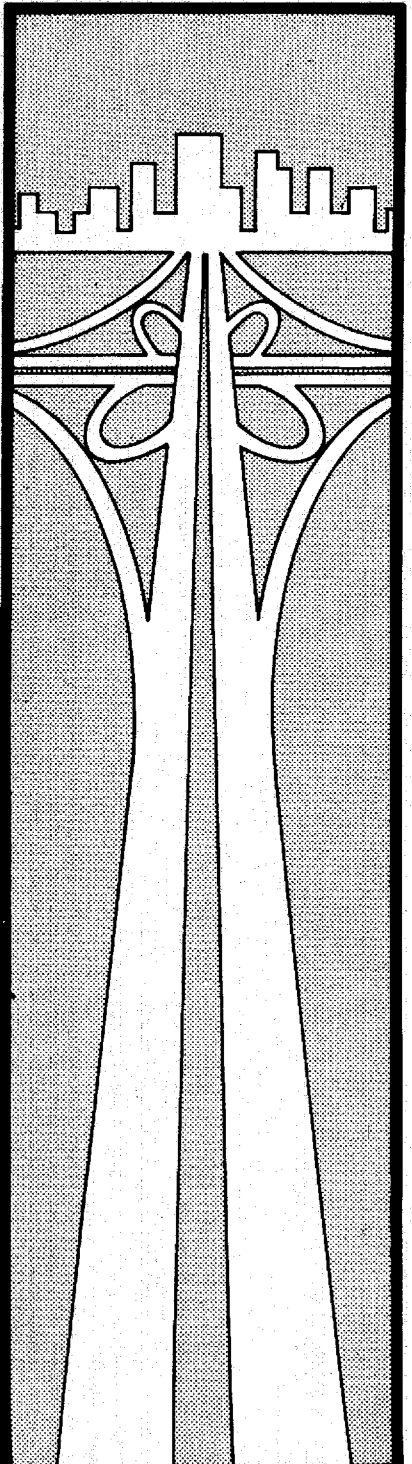
SVERCL

HIGHWAY PLANNING TECHNICAL REPORT



JULY 1973
NUMBER 31

SAMPLING SURVEYS FOR ESTIMATING LOCAL, RURAL AND URBAN VEHICLE - MILES OF TRAVEL

U. S. DEPARTMENT OF TRANSPORTATION
Federal Highway Administration

U.S. DEPARTMENT OF TRANSPORTATION
FEDERAL HIGHWAY ADMINISTRATION
WASHINGTON, D.C. 20590

SAMPLING SURVEYS FOR ESTIMATING LOCAL RURAL AND URBAN VEHICLE-MILES OF TRAVEL

By
Richard R. Bodle, Highway Engineer
Office of Highway Planning
Program Management Division
Methods Branch

INTRODUCTION

Reliable estimates of vehicle-miles of travel (VMT) on each highway system within a State continue to be required for many purposes. Traffic data provide a basis for producing independent estimates of vehicle-miles on the major highway systems in a State. The reliability of State estimates of vehicle-miles (VMT) has become increasingly important in recent years. This results partially from the increasing emphasis on safety programs and the need to determine the effectiveness of safety programs in different States. In developing new highway programs and financing procedures it is important that the estimates of travel for rural and urban areas provide a reliable basis for comparison. A sound documented basis for the VMT values for each State is essential for meaningful comparisons and forecasts.

The most desirable method for satisfying this need for statewide VMT estimates would be to have representative traffic counts on every section of highway, both urban and rural. This, of course, would be prohibitively expensive. In addition, many highway departments do not have counting programs on roads and streets that are not on Federal-aid or State systems. Traditionally, total statewide travel is estimated on the basis of fuel consumption and a single average miles per gallon figure; the VMT for the major systems are computed from traffic counts. The remainder becomes the local road and street travel because many States do not have reliable estimates of travel for local rural roads and local city streets based on recent traffic counts. With the lack of data, considerable judgment is necessary to arrive at the average miles per gallon figure.

TWO SAMPLING APPROACHES

Recognizing the need for reliable VMT estimates on local rural roads and city streets, two probability sampling procedures were developed in late 1969 and early 1970. The reader should understand that these procedures were intended only to fill the void of VMT data on local rural roads and city streets. They were not intended to be used in obtaining statewide or citywide VMT estimates on all systems. These sampling procedures, which may be applied to rural systems and urban streets, are based on different types of samples--the sample area and the link-day. The two approaches will be described in this section.

AREA SAMPLING

For rural areas, the basic plan is to divide the State (or other study area) into sampling areas containing about 50 miles of local rural roads each. A variation in length not exceeding 5 miles is acceptable. A total of 104 sample areas are selected for sample counts in a year. The probability of a particular sampling area being selected in any week will depend on the total number of sampling areas in the State. It is important that each sample area contain about equal mileage (50 miles) of local roads and that it not be divided by travel barriers such as a river without a bridge or a mountain ridge. The boundaries of the selected sample areas should be readily identifiable in the field, such as marked political boundaries, or natural or manmade physical boundaries such as rivers or power lines.

Two sample areas are randomly selected for each week within the year, and counters set out for a one-week period at every fifth milepoint on the local rural roads within the selected area. With this procedure it is possible to select the same sampling area more than once but for different weeks of the year. The table below illustrates this selection procedure.

County Name	Miles of Local rural road	Number of Sampling Areas	Sample Areas in County	Random Digits 1st Week	2nd Week	3rd Week
Able	783.2	16	1-16			7
Baker	412.9	8	17-24		22	
Charley	1144.6	23	25-47	43	31	36
Douglas	929.3	18	48-65	50		

County Able contributes sampling areas 1 through 16 to the total. County Baker contributes sampling areas 17 through 24 to the total. The random number 50 actually selects the third sampling area defined in Douglas County. The third sample area of the county is the fiftieth in the listing sequence. This sample area concept is intended to concentrate the time and location of counts sufficiently for efficient field operation while providing for probability sampling.

The sampling procedure for local urban streets is similar to that described above with two exceptions. Urban sampling areas should be defined to contain about 5 miles of local streets with a permissible variation of plus or minus a half mile. Counters should be placed at every half milepost point rather than every fifth milepost point.

This area sampling procedure selects both the locations and the time periods for obtaining counts. Further details on sample selection and calculation of relative errors are described in Appendix A.

LINK-DAY SAMPLING

This statistical procedure is based upon a sampling of 24-hour volume counts on city streets. The studies to be described were designed to yield statewide urban VMT within $\pm$ 5 percent of the actual VMT at the 95 percent confidence level. The results did not quite attain that level. The number of sample counts may be reduced with a corresponding reduction in precision if required by budgetary constraints. The procedure may be applied to a single urban area or a group of urban areas.

The design involves stratified random sampling of 24-hour volumes utilizing the link-day as the basic sampling unit. A link in a street system is considered to be one street intercepted by two other streets. If a street link has abrupt volume changes due to abutting land uses, the link should be further divided into two or more links. Links are stratified into volume groups. This stratification must be done on the basis of best information available or by personal evaluation. In many cases, some form of volume information will be available for many of the local roads and streets. Maintenance personnel are another source of information useful in stratifying the links. A "link-day" is a 24-hour volume on the given link. Based on the desired accuracy of the VMT estimate and assumed values for the standard deviations of the strata, the required number of sample counts (n) can be determined from the following formulas:

$$n = S^2 \div E^2 \qquad (1)$$

$$n = \frac{(\Sigma\ NiSi)^2}{(NE)^2 + \Sigma\ NiSi^2} \qquad (2)$$

Where:

- n = total number of 24-hour sample counts.
- E = desired sample error of urban VMT estimate.
- S = assumed overall standard deviation of link vehicle-miles based upon variation in link volumes.
- Si = assumed standard deviation of the link vehicle-miles based on variation in link volumes in volume group "i".
- N = overall total number of link-days from which a sample may be drawn.
- Ni = total number of link-days from which a sample may be drawn from volume group "i". If there are 5100 links in a volume group, the total number of link-days from which the sample could be selected is 5100 x 365 or 18.61×10^5 link-days.

Formula (1) is used to select samples from volume group 1-999 and formula (2) for the remaining volume groups. Investigations showed that volume group 1-999 had many more links than any of the remaining groups, thus overweighting the estimate of the needed sample if all volume groups were considered as one universe. Therefore, data should be analyzed as two separate universes.

In order to utilize formulas (1) and (2), values for S and Si must be estimated. In making a first year's study design, it is assumed that the lengths of the links in any volume group are approximately constant. The variability in vehicle-miles will then be proportional to the variation in the volumes. It is further assumed that the volumes are normally distributed within each volume group. A range of ± 36 around the group mean would then account for almost 100 percent of the links. An estimate of Si for a volume group could then be obtained by dividing the range (R) of the volume group by 6. Because of the variability of daily vehicle counts about the ADT, a safer estimate of Si would be obtained by dividing the range by 5 when VMT estimates are for weekdays only. If the VMT estimates are to include weekdays, the range should be divided. by 4.5 (Si = Ri ÷ 4.5).

Since the object is to estimate annual urban VMT, the sample counts will be distributed throughout the year on links of the street system for which the estimates are to be made. If desired, sampling may be limited to local city streets. If L represents the total number of urban street links, the population of link-days to be sampled (N) equals 365L. If the VMT estimate is for non-holiday weekdays only, we can assume 250 non-holiday weekdays in the universe (N = 250L).

An illustration of the detailed sample selection procedure is contained in Appendix B.

PILOT STUDIES - Area Sampling

Colorado

During the period October 10, 1970, to September 30, 1972, the Colorado Department of Highways, conducted an area sampling study to estimate vehicle-miles of travel on local rural roads and city streets. The rural portion of the study was conducted during the first year with the urban portion following during the second year.

As described previously, this sampling procedure requires obtaining 7-day volume counts in each of two sample areas each week of the year. Scheduling of men and equipment became a problem in

Colorado since travel time between sample areas was not provided for. This was solved by taking volume counts for six days-setting out the counters on Tuesday and picking them up the following Monday. The total 7-day volume was calculated from the familiar expression:

$$V = 5A + B + C$$

where V = total 7-day volume
A = average weekday as determined from the weekday portion of the 6-day count
B = Saturday volume
C = Sunday volume

The random sample selection procedure determines both the sample areas and the weeks they are to be counted. In a few cases this resulted in sample areas for consecutive weeks being prohibitively far apart to accomplish the required travel in the allotted time. In such cases, the pair of sampling areas for a week was switched with those for another week. The table below illustrates.

Original Schedule			Revised Schedule		
Calendar Week	Schedule Week	Sample Areas	Calendar Week	Schedule Week	Sample Areas
1	1	A + B	1	1	A + B
2	2	C + D	2	3	E + F
3	3	E + F	3	4	G + H
4	4	G + H	4	2	C + D

The total number of areas in Colorado from which the sample areas were selected was 1326 rural and 1363 urban. Table 1, prepared by the Colorado Department of Highways, summarizes the Colorado results. The relative error for the rural portion was a disappointing 31.1 pereent. The relative error for the urban local streets was 8.3 percent with an overall urban plus rural error of 11.7 percent. The formulas used to calculate vehicle-miles and relative errors are explained in Appendix A.

Oregon

The Oregon State Highway Division conducted a county road vehicle-mile study from September 29, 1969, to September 28, 1970, using the area sampling concept. In this study machine counts were

TABLE 1

STATE DEPARTMENT OF HIGHWAYS
DIVISION OF HIGHWAYS — STATE OF COLORADO
Planning and Research Division

1970—1972 VEHICLE-MILES OF TRAVEL STUDY
Calculation of the Relative Error

Study Week Number	RURAL								
	X_{1R}	X_{2R}	$\lvert X_{1R} - X_{2R}\rvert$	$Y_{1R} = \lvert X_{1R} - X_{2R}\rvert \times 3,315$	YN_R	Relative Error $= \frac{Y_{1R}}{YN_R} \times 100$	X_{1U}	X_{2U}	$\lvert X_{1U} - X_{2}$
1	1,399	588	811	2,688,465	6,590,220	40.8	7,930	15,986	8,056
2	1,217	4,499	3,282	10,879,830	18,948,540	57.4	5,649	11,492	5,843
3	1,442	812	630	2,088,450	7,472,010	28.0	7,804	34,210	26,406
4	100,779	10,552	90,227	299,102,505	369,065,580	81.0	146,813	402,129	255,316
5	713	625	88	291,720	4,435,470	6.6	256,855	18,514	238,341
6	1,385	1,295	90	298,350	8,884,200	3.4	23,914	32,463	8,549
7	3,495	875	2,620	8,685,300	14,486,550	60.0	183,566	109,680	73,886
8	1,069	22,744	21,675	71,852,625	78,943,410	91.0	31,853	71,980	40,127
9	2,250	8,877	6,627	21,968,505	36,889,320	60.0	160,007	4,627	155,380
10	2,160	1,923	237	785,655	13,538,460	5.8	80,263	46,327	33,936
11	592	1,146	554	1,836,510	5,761,470	31.9	13,676	248,374	234,698
12	975	181	794	2,632,110	3,832,140	68.7	18,379	37,160	18,781
13	8,398	937	7,461	24,733,215	30,948,840	79.9	22,322	8,532	13,790
14	1,145	239	906	3,003,390	4,587,960	65.5	46,346	133,424	87,078
15	2,187	284	1,903	6,308,445	8,194,680	77.0	90,718	18,744	71,974
16	3,305	1,141	2,164	7,173,660	14,738,490	48.7	187,284	107,202	80,082
17	1,119	16,863	15,744	52,191,360	59,610,330	87.6	75,938	227,328	151,390
18	550	221	329	1,090,635	2,559,180	42.6	69,341	111,773	42,432
19	727	1,063	336	1,113,840	5,933,850	18.8	177,396	145,039	32,357
20	895	978	83	275,145	6,212,310	4.4	236,485	153,349	83,136
21	988	16,722	15,734	52,158,210	58,708,650	88.8	157,118	30,921	126,197
22	2,560	2,676	116	384,540	17,357,340	2.2	11,220	13,045	1,825
23	0	1,107	1,107	3,669,705	3,673,020	99.9	101,370	134,247	32,877
24	4,192	1,433	2,759	9,146,085	18,650,190	49.0	37,384	133,826	96,442
25	1,031	288	743	2,463,045	4,375,800	56.3	144,045	50,291	93,754
26	0	398	398	1,319,370	1,319,370	100.0	74,645	21,407	53,238
27	276	2,322	2,046	6,782,490	8,612,370	78.8	75,623	219,038	143,415
28	598	5,011	4,413	14,629,095	18,597,150	78.7	234,878	196,960	37,918
29	447	985	538	1,783,470	4,747,080	37.6	207,487	8,956	198,531
30	6,944	780	6,164	20,433,660	25,605,060	79.8	353,592	2,604	350,988
31	3,022	845	2,177	7,216,755	12,822,420	56.3	122,636	39,905	82,731
32	0	1,071	1,071	3,550,365	3,553,680	99.9	103,134	9,309	93,825
33	22,262	500	21,762	72,141,030	75,456,030	95.6	18,603	17,728	875
34	1,140	3,014	1,874	6,212,310	13,770,510	45.1	290,068	86,393	203,675
35	354	1,265	911	3,019,965	5,370,300	56.2	76,835	95,463	18,628
36	4,640	175	4,465	14,801,475	15,965,040	92.7	283,786	17,514	266,272
37	879	5,648	4,769	15,809,235	21,646,950	73.0	193,689	89,017	104,672
38	99	109	10	33,150	689,520	4.8	202,216	119,362	82,854
39	130,238	3,247	126,991	420,975,165	442,506,090	95.1	230,381	68,022	162,359
40	2,817	89	2,728	9,043,320	9,633,390	93.4	22,370	34,529	12,159
41	840	918	78	258,570	5,827,770	4.4	296,405	177,084	119,321
42	605	15,112	14,507	48,090,705	52,105,170	92.3	18,604	16,704	1,900
43	3,290	2,505	785	2,602,275	19,213,740	13.5	22,500	119,209	96,709
44	34,505	825	33,680	111,649,200	117,118,950	95.3	31,970	203,984	172,014
45	1,168	552	616	2,042,040	5,701,800	35.8	80,978	46,050	34,928
46	1,171	378	793	2,628,795	5,138,250	51.2	217,023	141,020	76,003
47	3,158	1,240	1,918	6,358,170	14,579,370	43.6	80,152	63,004	17,148
48	469	883	414	1,372,410	4,481,880	30.6	77,804	76,649	1,155
49	2,254	895	1,359	4,505,085	10,442,250	43.1	71,298	125,761	54,463
50	6,527	1,388	5,139	17,035,785	26,241,540	64.9	227,036	45,931	181,105
51	5,004	1,132	3,872	12,835,680	20,340,840	63.1	12,650	18,655	6,005
52	4,214	752	3,462	11,476,530	16,462,290	69.7	40,587	93,822	53,235
Total or Average					1,762,346,820	31.1			

X = Total Count in One Area for One Week.

Number of Areas Rural = 1,326.

Number of Areas Urban = 1,363.

Constant Rural = $\frac{5(1,326)}{2}$ = 3,315.

Constant Urban = $\frac{5(1,363)}{20}$ = 341.

Relative Error of 52 Weeks Rural =

$$3,315\sqrt{\frac{1}{52}\left|X_{1R} - X_{2R}\right|^2} \div YN_R \times 100.$$

Relative Error of 52 Weeks Urban =

$$341\sqrt{\frac{1}{52}\left|X_{1U} - X_{2U}\right|^2} \div YN_U \times 100.$$

Relative Error of 52 Weeks Total =

$$\sqrt{\frac{1}{52} Z_1^{\,2}} \div Z_2 \times 100.$$

TABLE 1 (continued)

STATE DEPARTMENT OF HIGHWAYS
DIVISION OF HIGHWAYS — STATE OF COLORADO
Planning and Research Division

1970—1972 VEHICLE-MILES OF TRAVEL STUDY

Calculation of the Relative Error

URBAN			TOTAL			
$_{1U} = \lvert X_{1U} - X_{2U} \rvert \times 341$	YN_U	Relative Error $= \frac{Y_{1U}}{YN_U} \times 100$	$Z_1 = \sqrt{3{,}315^2 \lvert X_{1R} - X_{2R} \rvert^2 + 341^2 \lvert X_{1U} - X_{2U} \rvert^2}$	$Z_2 = YN_R + YN_U$	Relative Error $= \frac{Z_1}{Z_2} \times 100$	Study Week Number
2,747,096	8,149,377	33.7	3,843,745	14,739,597	26.1	1
1,992,463	5,840,455	34.1	11,060,768	24,788,995	44.6	2
9,004,446	14,316,952	62.9	9,243,466	21,788,962	42.4	3
87,062,756	187,052,668	46.5	311,515,860	556,118,248	56.0	4
81,274,281	93,833,009	86.6	81,274,804	98,268,479	82.7	5
2,915,209	19,211,485	15.2	2,930,436	28,095,685	10.4	6
25,195,126	99,924,256	25.2	26,650,118	114,410,806	23.3	7
13,683,307	35,382,117	38.7	73,143,917	114,325,527	64.0	8
52,984,580	56,099,717	94.4	57,358,355	92,989,037	61.7	9
11,572,176	43,136,224	26.8	11,598,815	56,674,684	20.5	10
80,032,018	89,294,219	89.6	80,053,086	95,055,689	84.2	11
6,404,321	18,925,255	33.8	6,924,112	22,757,395	30.4	12
4,702,390	10,512,819	44.7	25,176,266	41,461,659	60.7	13
29,693,598	61,257,309	48.5	29,845,102	65,845,269	45.3	14
24,543,134	37,299,858	65.8	25,340,913	45,494,538	55.7	15
27,307,962	100,346,786	27.2	28,234,485	115,085,276	24.5	16
51,623,990	103,338,571	50.0	73,409,634	162,948,901	45.1	17
14,469,312	61,715,277	23.4	14,510,357	64,274,457	22.6	18
11,033,737	109,870,067	10.0	11,089,814	115,803,917	9.6	19
28,349,376	132,836,617	21.3	28,350,711	139,048,927	20.4	20
43,033,177	64,074,630	67.2	67,619,029	122,783,280	55.1	21
622,325	8,269,321	7.5	731,546	25,626,661	2.9	22
11,211,057	80,287,515	13.0	11,796,377	83,960,535	14.0	23
32,886,722	58,340,489	56.4	34,134,840	76,990,679	44.3	24
31,970,114	66,219,992	48.3	32,064,852	70,595,792	45.4	25
18,154,158	32,729,719	55.0	18,202,038	34,049,089	53.5	26
48,904,515	100,406,758	48.7	49,372,601	109,019,128	45.3	27
12,930,038	147,149,480	8.8	19,524,249	165,746,630	11.8	28
67,699,071	73,753,293	91.8	67,722,558	78,500,373	86.3	29
119,686,908	121,373,787	98.6	121,418,650	146,978,847	82.6	30
28,211,271	55,386,868	50.9	29,119,707	68,209,288	42.0	31
31,994,325	38,315,293	83.5	32,190,711	41,868,973	76.9	32
298,375	12,380,129	2.4	72,141,647	87,836,159	82.0	33
69,453,175	128,280,108	54.1	69,730,454	142,050,618	49.1	34
6,352,148	58,711,225	10.8	7,033,489	64,081,525	11.0	35
90,798,752	102,667,975	88.4	91,997,266	118,633,015	77.5	36
35,693,152	96,332,751	37.1	39,037,584	117,979,701	33.1	37
28,253,214	109,578,385	25.8	28,253,233	110,267,905	25.6	38
55,364,419	101,681,163	54.4	424,600,174	544,187,253	78.0	39
4,146,219	19,388,675	21.4	9,948,505	29,022,065	34.3	40
40,688,461	161,342,399	25.2	40,689,282	167,170,169	24.3	41
647,900	12,031,201	5.4	48,095,069	64,136,371	75.0	42
32,977,769	48,288,364	68.3	33,080,282	67,502,104	49.0	43
58,656,774	80,402,007	73.0	126,119,629	197,520,957	63.9	44
11,910,448	43,284,791	27.5	12,084,233	48,986,591	24.7	45
25,917,023	122,003,493	21.2	26,050,002	127,141,743	20.5	46
5,847,468	48,780,407	12.0	8,638,241	63,359,777	13.6	47
393,855	52,630,882	0.7	1,427,806	57,112,762	2.5	48
18,571,883	67,148,195	27.7	19,110,484	77,590,445	24.6	49
61,756,805	93,013,846	66.4	64,063,413	119,255,386	53.7	50
2,047,705	26,125,984	7.8	12,997,991	46,466,824	28.0	51
18,153,135	45,800,889	39.6	21,476,662	62,263,179	34.5	52
	3,564,523,052	8.3		5,326,869,872	11.7	

adjusted to vehicle counts on the basis of manual classification counts made concurrently with the machine counts. Classification counts were conducted throughout each sample area during hours and at locations thought to be most representative of all traffic in the sample area. Complete 24-hour classification counts were not taken at any recorder location. A total of 497 areas were identified from which the sample areas were selected.

Field data were collected in 47 weeks of the 52-week period. During five additional weeks data were collected in only one sample area due to bad weather, mishandled mail or illness. VMT, standard error and relative error were computed on the basis of the 47 weeks when data were collected in two sampling areas.

The relative error for the Oregon study was 26.9 percent which is in the same order of magnitude as the results in Colorado. The results are summarized in Table 2. The calculations are similar to the Colorado study and are described in Appendix A. The high relative error for some weeks was due to the inclusion of some high volume roads in the sample of predominately low volume roads. In some areas there was quite a variation in the volumes counted, especially near large urban areas.

PILOT STUDIES - Link-Day Sampling

Wisconsin

In 1969, an urban VMT study was conducted in the city of Oshkosh in conjunction with the Fox Valley Transportation Study. Field work was conducted between March 31 and June 13, and between September 2 and October 31. As designed, the study produced estimates of weekday VMT for the sampling period on all streets. The study was not limited to local streets as was the case in Colorado and Oregon. The number of samples by volume group was calculated as described above and in Appendix B. Due to budgetary constraints and a desire for a minimum sample to enable an evaluation of a stratum's contribution to sampling error, the number of samples was altered as shown in Table 3.

Table 3

Volume Group	Miles of Streets	Number of Calculated Samples	Number Samples Used
1-499	8838	553	104
500-999	1720	87	48
1000-2999	1729	74	73

TABLE 2

Oregon

Week	Relative Error (Percent)	Week	Relative Error (Percent)
1	14.7	27	32.8
2	19.4	28	68.3
3	32.2	29.	99.1
4	41.1	30	8.5
5	34.4	31	91.5
6	88.7	32	71.9
7	48.0	33	20.7
8	56.0	34	11.0
9	70.0	35	60.7
10	43.5	36	10.8
11	98.3	37	*
12	72.5	38	55.2
13	12.9	39	1.3
14	*	40	97.8
15	*	41	54.7
16	86.7	42	18.8
17	51.5	43	54.6
18	99.5	44	98.8
19	8.7	45	93.5
20	7.2	46	80.8
21	71.2	47	81.4
22	67.5	48	70.6
23	2.4	49	*
24	13.6	50	*
25	77.3	51	35.5
26	23.3	52	6.0
		Total	26.9

* Only one sample area surveyed during these weeks.

Table 3 (continued)

Volume Group	Miles of Streets	Number of Calculated Samples	Number Samples Used
3000-3999	856	17	50
4000-5499	816	23	50
5500-6999	819	28	50
7000-8999	427	21	50
9000-10999	261	16	44
11000-11999	206	3	30
Over 12,000	277	17	30
Total	15,949	839	529

The sample was a stratified random sample of street link-time periods within a 20-week study period. Within each week, three 48-hour counting periods were defined as:

Monday-Tuesday
Tuesday-Wednesday
Wednesday-Thursday

Because of two 1-day holidays within the study period, two of the 60 possible time periods were not used. The 58 remaining time periods covered 78 non-overlapping weekdays. Due to the use of 48-hour rather than 24-hour counts the computation of VMT, standard error and relative error differed somewhat from that described in Appendix B. The overall relative error for this study was 6.9 percent. The results are summarized in Table 4.

Table 4

Volume Group	Population of Link-Days	Coefficient of Variation (%)
1-499	22330	52.4
500-999	3480	20.5
1000-2999	5162	10.7
3000-3999	2436	7.2
4000-5499	2262	10.5
5500-6999	2610	7.4
7000-8999	1450	7.4
9000-10,999	1102	9.0
11,000-11,999	464	17.3
12,000 +	580	24.1
Total	41876	6.9

The results indicate that the first and last strata were the major contributors to sampling variability. These large values arise from non-homogeneous strata components. The State's analysis showed that although the internal strata do not contribute as much to sampling variability as the two extremes, many of these internal strata had proportionately more incorrect link assignments than the two extremes. It is apparent that errors in link assignments to volume groups were not as great in absolute value for the interior strata as for the exterior strata (first and last volume groups). It should be noted that although the first and last strata contributed large relative errors, these strata accounted for only 23 percent of the study area's total estimate of VMT. Also of note is the fact that link assignments were made on the basis of ADT, while the sampling was conducted over a relatively short period of time. ADT may be a poor approximation of the seasonal volumes, especially for the "large error" links in the two extreme strata. Any subsequent study in the Oshkosh area would benefit through a more accurate assignment of links to volume groups.

Idaho

In 1958, Idaho conducted a link-day sampling study for estimating urban VMT in 15 cities with the population of 5,000 or more. The study covered a period of 218 weekdays. Out of a universe of 17,149 links, 2,338 were selected to be counted. Counts were actually made for 24-hour duration on 2,163 links. Out of the total universe of links, 4,369 were on streets with ADT volumes of 1,000 or more. Of the 2,163 sample counts, 1,976 were on streets with volumes of 1,000 or more. Table 5 shows the volume groups, universe of links, number of sample counts and relative error.

Table 5

Volume Group	No. of Links in Universe	Number of Sampled Links	Coefficient of Variations (%)
0-999	12,780	187	8.9
1,000-1,999	1,178	280	6.9
2,000-2,999	681	164	8.8
3,000-4,999	1,031	479	5.3
5,000-6,999	566	272	5.9
7,000-9,999	500	361	5.2
10,000-13,499	259	204	5.5
13,500-18,999	135	172	5.0
19,000 +	19	44	11.0
Over 999	4,369	1,976	6.9
Total	17,149	2,163	

The procedure for selecting the number of links to be counted resulted in a prohibitively large sample for the 0-999 volume group. It was decided that a maximum of 200 sample counts would be made for this volume group--187 counts were actually made. Table 6 shows the man-hours and cost expended on the study. With a total cost of $27,533 to obtain 2,163 counts and the analysis, the average cost per count was $12.73.

Table 6

IDAHO VMT STUDY
Cost and Man-Hour Requirements

Work Item	Man-Hours	Cost
Color code traffic volumes on maps	33	$ 103
Number and describe links for volume groups	82	257
Scale link mileages by volume groups	15	47
Prepare and select link-days sample	82	256
Develop recording forms	18	56
Key-punch link-days	8	16
Sort link-days	2	4
Computer-straight listing	-	25
Schedule field operations	196	614
Develop map guides for field use	15	47
Obtain traffic volumes	7,389	23,128
Tabulate traffic volumes	806	2,523
Scale sample link miles	77	241
Compute vehicle-miles and error limits	45	141
Summarize results	24	75
	8,792	$27,533

The estimate of 564,762,112 vehicle-miles for the 218 day period had a relative error of 6.9 percent which is similar to the Wisconsin results. (The fact that the same relative error was obtained in two studies has no statistical significance.)

COMMENTS

Table 7 summarizes the pilot studies in the four States discussed earlier.

Table 7

State	Sample Type	Urban or Rural	Highways Counted	Relative Error (%)
Colorado	Area	Rural	Local Roads	31.1
	Area	Urban	Local Streets	8.3
Oregon	Area	Rural	Local Roads	26.9
Wisconsin	Link-Day	Urban-Oshkosh	All city streets	6.9
Idaho	Link-Day	Urban-all over 5,000 pop.	All city streets	6.9

Although the rural results were not statistically precise, VMT estimates with a measurable relative error were obtained. In the Colorado study, the results indicated that previous estimates of rural local road VMT may have been too low. Both the Colorado and Oregon studies showed where improvements in the design of rural studies were needed. These include:

1. The area concept, as designed, requires a much larger sample than originally recommended to attain the desired precision.

2. The high relative error appears due to inclusion of some high volume roads in the sample of predominately low volume roads. In Oregon, counts among the 10 that were made in an area could be quite variable, especially near large urban areas.

3. It appears that some stratification by volume groups is warranted as an additional feature of the area sampling concept. This could be achieved by either of the following:

    a. Stratify the sample areas into volume groups-low, medium and high. Count the areas selected from each of the groups during only a portion of the year.

    b. Identify local roads carrying high volumes and sample them on a road section basis. The remainder of the local roads may be sampled on an area basis.

Due to the great difference in geography between eastern and western Oregon, it was thought that a geographic stratification of samples would improve the relative error. This concept was tested with the data by dividing the sample areas into two groups. This resulted in an insignificant improvement in relative error from 26.9 to 26.1 percent.

Data for the Wisconsin urban VMT study were obtained during only a portion of the year. The link-day study design stratified links on the basis of annual average daily traffic (ADT). Large seasonal variation on some of the links would account for large errors in some of the volume groups. This indicates that when a VMT study is to cover only a portion of a year, links should be stratified on the basis of average traffic for the season only.

In the Wisconsin study the 12,000 and over volume group seems to contribute excessive variability. The unknown value at the upper end of the stratum may have differed from the smallest stratum value by a considerable amount. If this is the case, stratification was cut off too soon. This problem was reduced somewhat in Idaho where links were stratified up to 19,000 and over volume.

In reviewing the results of the link-day samples it should be noted that the objective was a statewide (urban area wide) VMT estimate with a measurable reliability. We were not so concerned with relative error resulting in each volume stratum although they are reported in the summary tables. If reliability within strata were to become a governing factor, the study designs would have to be changed radically.

Appendix A

Area Sampling

Introduction

This probability sampling procedure is suggested as a basis for estimating vehicle-miles of travel on road categories for which sufficient travel data are not available. In most States, this includes the category of local roads and city streets. Although the basic principle is the same in both cases, a separate description of the sampling procedure is provided for local rural road travel and for city street travel. While statewide applications are described, the same procedures can be used to provide estimates for smaller areas as a county, a group of counties, a city, or a group of cities.

The procedure involves a sampling plan to collect traffic data over a one-year period, Available time may be insufficient to collect data for an entire year. In such a case, the sample plan may be applied in the period available to produce estimates for that period. These estimates for a shortened period may then be factored to an annual basis.

Procedure for local rural roads

The basic plan is to divide the entire State (or other area) into sampling areas with about 50 miles of local rural roads in each sampling area. A variation in length not exceeding 5 miles is acceptable. Each of the sampling areas has 52 chances of selection for each of the 52 weeks of the year. Each selected area is chosen for a specified week within the year, and counters are put out for a 1-week period at every fifth milepost point on the local rural roads within the selected area.

The total local rural road mileage in each county, determined from office records, should be divided by 50 to define the number of sampling areas into which the county should be divided, prior to the selection of sample areas. For example, a county with 523 miles of local rural roads should be divided into 10 sampling areas. However, determination of selected area boundaries should not be made at this point. Only the number of areas into which a county could be divided if needed should be determined at this point.

The number of sampling areas in each county should be listed in any desired sequence and progressively summed. Two different random numbers less than or equal to the total number of such areas in the State (or other desired region) should be determined to select two sample areas for the first week of data collection. An independent selection of two random numbers should be made to select two sample areas for the second week of data collection. This procedure should be continued for the 52 weeks of data collection. With this procedure, it is possible to select the same sampling area more than once but for different weeks of the year. A total of 104 sample areas are selected for the year.

The artificial example in the following table illustrates the procedure thus far.

County Name	Miles of Local rural road	Number of Sampling Areas	Progres-sive totals	Week	Week	Week	
Able	783.2	16	16			7	
Baker	412.9	8	24		22		etc.
Charley	1144.6	23	47	43	31	36	
Douglas	929.3	18	65	50			

County Able contributes sampling areas 1 through 16 to the total. County Baker contributes sampling areas 17 through 24 to the total, and so on. The random numbers are placed on the line of the county contributing that numbered area to the total. As an example, the random number 50 actually selects the third sampling area to be defined in Douglas County. That third sample area of the study is the fiftieth in the listing sequence.

In States with many counties, not every county will contribute to the sample. Where there are few counties, some counties will contribute several sample areas, and the same sample areas may be selected more than once.

On maps of counties containing selected sample areas, draw boundaries for all sampling areas within those counties, not just for selected areas. Number sequentially all the sampling areas in the counties containing selected areas starting in the northwest corner, proceeding in any regular sequence until all the sampling areas in those counties are numbered. Then determine the selected sample areas.

The intent of the sample area concept is to concentrate the time and location of counts sufficiently for efficient field operation while providing for probability sampling. It is important that each sample area contain about equal mileage (50 miles) of local roads and that it not be divided by travel barriers such as a river without a bridge or a mountain ridge.

If local roads are in a grid pattern at 1 mile spacing, the sample areas would be 5 miles square. In most counties there are diagonal and meandering local roads, as well as a variety of intervening non-local roads so that the actual area containing 50 miles of local road may be 10 miles by 10 miles or larger in sparsely settled rural areas or less than 3 miles by 3 miles in developed areas outside suburban fringes. Generally, it is practical to scale off 4 or 5 sections of east-west roads 4 to 8 miles long starting at the north-west corner of the county, with each succeeding section directly south of the preceding section. The north-south roads and diagonal roads are then scaled off until the desired 50 mile total is achieved. Adjustment is usually necessary, but as additional sample areas are scaled it becomes easier to select arrangements which require minimum final adjustment.

In this process it is important that all mileage in the local road inventory be accounted for, including stubs, cul de sacs, etc.

The boundaries of the selected sample areas should be readily identifiable in the field, such as marked political boundaries of township lines, or natural or man-made physical boundaries such as rivers or power lines.

Once the boundaries of selected sample areas have been determined, the location of counters should be established in the office. The local roads within a selected sample area should be listed in the following sequence: east-west roads in sequence from south to north; north-south roads in sequence from west to east. Diagonals should be arbitrarily assigned to an appropriate place in the sequence. The mileage should be listed and summed. Every fifth milepost point starting with milepoint two should be located on the map. If the sample areas have been defined with good judgment, 10 counting points will be located with each count representing 5 miles of roads within the sample area.

The procedure that has been described selects both the locations and the time periods for obtaining counts.

Since each one week count will represent the vehicles passing over 5 miles of local road on the average, multiplication of the count by 5 is an estimate of vehicle-miles for the week yielded by the counter. If all the one week counts in a sample area are added and then multiplied by 5, an estimate of vehicle-miles for the week in the sample area is obtained. If only one sample area was selected each week, multiplication of the one week area estimate by N (the total number of areas from which the sample area was selected) would yield an estimate of vehicle-miles for one week for the State (region). However, two areas are to be selected for each week. Therefore the average of the two areas should be calculated and that average multiplied by N. The preceding description may be symbolized by the following expression:

$$X'_h = \frac{5N}{2} \sum_{i=1}^{2} \sum_{j=1}^{10} X_{hij} = \frac{5N}{2} \left[\sum_{j=1}^{10} X_{h1j} + \sum_{g=1}^{10} X_{h2j} \right]$$

Where:

X_{hij} = The one week count obtained at counter j in area i during week h, the week under consideration.

N = The total number of areas in the State (region) from which the sample areas were selected.

X'_h = The estimated vehicle-miles for week h.

The above estimate should be calculated for each week's counts and summed to get an annual total for local rural roads.

If some counts are missed in whole or in part in a sample area because of equipment malfunction or for other reasons, estimates should be obtained from the counts obtained at the other count points in the sample area during the same week.

Procedure for local urban streets

The procedures for local urban streets parallel that for local rural roads with two exceptions. Urban sampling areas should be defined as containing about 5 miles of local urban streets with a permissible variation of plus or minus a half mile. Also, counters should be placed at every half milepost point rather than every fifth milepost point. The expression $\frac{5N}{2}$ becomes $\frac{5N}{20}$.

Alternate procedures

The procedure described can be modified to achieve increased efficiency in use of manpower and equipment depending on the particular area and type of highway for which estimates are required. The size of the sampling area can be increased or decreased, to include more or less mileage. Certain other parts of the procedure are critical for valid results. Due to daily and seasonal variation, count periods must be selected so that each day of the week and each week of the year has equal probability of selection. This requires that an appropriate number of counts be obtained during winter.

Where areas are snowed in, or it can be positively determined by other means that no vehicles passed during the count period, valid zero counts should be recorded. The importance of accurate counts on low volume roads should be evident to an experienced, well trained field crew. On mileage where ADT is 5 vehicles, an error of one vehicle is equivalent to 20 percent. The proportional difference between a zero count, and a count of one or two vehicles is theoretically infinite. With two areas each week, each 1-week count represents just under 1 percent of the total sample. On mileage where daily volumes during the sampling week are 25 or less, any counts of the field crew vehicles should be subtracted before the value is used in computing the total for the State or other area for which vehicle-miles are to be estimated. The same basic sample of 104 seven day counts should be obtained for any area, highway system, or other category, containing 104 road sections or more, for which a reliable vehicle-mile estimate is desired. The reliability of the estimate cannot be determined until after the standard deviation of the counts is calculated. Based on typical low volume traffic variability, the procedure described is expected to produce a standard error of 3 percent or less. Thus, there would be 95 percent assurance that a difference of 6 percent or more in vehicle-miles between areas, systems or other categories estimated in this way would be meaningful.

The standard error of the estimated vehicle-miles for a 1-week period either on local rural roads or on local city streets is given by the absolute value of equation (1a) or (1b) as appropriate. (The absolute value is bracketed by the vertical lines.)

$$\text{rural:} \; - \frac{5N}{2} \left| \sum_{j=1}^{10} X_{h1j} - \sum_{j=1}^{10} X_{h2j} \right| = s_{x_h'} \qquad (1a)$$

$$\text{urban:} \; - \frac{5N}{20} \left| \sum_{j=1}^{10} X_{h1j} - \sum_{j=1}^{10} X_{h2j} \right| = s_{x_h'} \qquad (1b)$$

The relative error of the estimate is:

$$\frac{s_{x'_h}}{x'_h} = \frac{\left| \sum_{j=1}^{10} X_{h1j} - \sum_{g=1}^{10} X_{h2j} \right|}{\sum_{j=1}^{10} X_{h1j} + \sum_{j=1}^{10} X_{h2j}} \qquad (2)$$

An estimate of vehicle-miles on local rural roads and on local city streets for week h may be symbolized as:

$$X'_h = \frac{5N^{(a)}}{2} \left[\sum_{i=1}^{2} \sum_{j=1}^{10} X^{(a)}_{hij} \right] + \frac{5N^{(b)}}{20} \left[\sum_{i=1}^{2} \sum_{j=1}^{10} X^{(b)}_{hij} \right] \qquad (3)$$

Where:

$N^{(a)}$ = The total number of rural sampling areas from which the rural sample was selected.

$N^{(b)}$ = The total number of urban sampling areas from which the urban sample was selected.

$X^{(a)}_{hij}$ = The 1-week count obtained at counter j in rural area i during week h, the week under consideration.

$X^{(b)}_{hij}$ = The 1-week count obtained at counter j in urban area 1 during week h, the week under consideration.

The standard error of this estimate is:

$$s_{x'_h} = \sqrt{\left(\frac{5N^{(a)}}{2}\right)^2 \left(\sum_{j=1}^{10} X^{(a)}_{hij} - \sum_{j=1}^{10} X^{(a)}_{hij} \right)^2 + \left(\frac{5N^{(b)}}{20}\right)^2 \left(\sum_{j=1}^{10} X^{(h)}_{hij} - \sum_{j=1}^{10} X^{(b)}_{h2j} \right)^2} \qquad (4)$$

The relative error of the estimate is $s_{x'_h} \div x'_h$ where $s_{x'_h}$ is given by equation (4) and x'_h is given by equation (3).

An estimate of vehicle-miles either on local rural roads or on local city streets for R weeks, the period of the study, (R = 52 if the study extends over one year), is given by:

$$\text{rural: - } X' = \frac{5N}{2} \sum_{h=1}^{R} \sum_{i=1}^{2} \sum_{j=1}^{10} X_{hij} \qquad (5a)$$

$$\text{urban: - } X' = \frac{5N}{20} \sum_{h=1}^{R} \sum_{i=1}^{2} \sum_{j=1}^{10} X_{hij} \qquad (5b)$$

The standard error of this estimate is:

$$\text{rural: - } s_{x'} = \frac{5N}{2} \sqrt{\sum_{h=1}^{R} \left(\sum_{j=1}^{10} X_{h1j} - \sum_{j=1}^{10} X_{h2j} \right)^2} \qquad (6a)$$

$$\text{urban: - } s_{x'} = \frac{5N}{20} \sqrt{\sum_{h=1}^{R} \left(\sum_{j=1}^{10} X_{h1j} - \sum_{j=1}^{10} X_{h2j} \right)^2} \qquad (6b)$$

The relative error of the estimate is given by the value yielded by equation (6a) divided by the value yielded by equation (5a) or by equation (6b) divided by equation (5b) as appropriate.

An estimate of vehicle-miles on local rural roads and on local city streets for R weeks of the study is given by:

$$X' = \frac{5N^{(a)}}{2} \left[\sum_{h=1}^{R} \sum_{i=1}^{2} \sum_{j=1}^{10} X^{(a)}_{hij} \right] + \frac{5N^{(b)}}{20} \left[\sum_{h=1}^{R} \sum_{i=1}^{2} \sum_{j=1}^{10} X^{(b)}_{hij} \right] \qquad (7)$$

Where: $N^{(a)}$, $N^{(b)}$, $X^{(a)}_{hij}$, $X^{(b)}_{hij}$ are defined as for equations (3) and (4).

The standard error of this estimate is:

$$s_{x'} = \sqrt{\left(\frac{5N^{(a)}}{2}\right)^2 \left[\sum_{h=1}^{R} \left(\sum_{j=1}^{10} X^{(a)}_{h1j} - \sum_{j=1}^{10} X^{(a)}_{h2j}\right)^2\right] + \left(\frac{5N^{(b)}}{20}\right)^2 \left[\sum_{h=1}^{R} \left(\sum_{j=1}^{10} X^{(b)}_{h1j} - \sum_{j=1}^{10} X^{(b)}_{h2j}\right)^2\right]} \qquad (8)$$

The relative error of the estimate is given by the value yielded by equation (8) divided by the value yielded by equation (7).

Following is a completely artificial example devised to illustrate the calculations required by the preceding formulas. Let it be assumed that a State has about 40,000 miles of local rural roads and about 8,000 miles of local city streets. The rural portion of the State is divided into 800 sampling areas with about 50 miles of local rural roads in each area. The urban portion of the State is divided into 1,600 sampling areas with about 5 miles of local city streets in each area. For the illustration, let it be assumed that the following data were collected during a 4-week period, the life of the study.

Week	Counter	Rural		Urban	
		Area 1	Area 2	Area 1	Area 2
1	1	10	50	50	20
	2	25	100	100	10
	3	40	20	75	50
	4	55	10	125	70
	5	15	30	150	40
	6	75	10	30	50
	7	120	50	10	80
	8	100	40	110	20
	9	20	50	90	40
	10	40	40	60	20
	Total	500	400	800	400
2	1	10	50	30	150
	2	15	70	50	80
	3	20	45	80	130
	4	20	95	110	70
	5	40	70	20	160
	6	25	100	60	180
	7	50	50	120	110
	8	40	60	80	130
	9	15	20	90	100
	10	15	40	60	190
	Total	250	600	700	1300

Week	Counter	Rural Area 1	Rural Area 2	Urban Area 1	Urban Area 2
3	1	20	100	100	40
	2	25	40	170	60
	3	30	30	140	60
	4	40	60	70	100
	5	10	90	60	70
	6	60	30	150	20
	7	30	20	100	60
	8	25	80	50	60
	9	50	60	100	80
	10	10	90	60	50
	Total	300	600	1000	600
4	1	200	30	90	50
	2	150	15	80	200
	3	40	20	110	150
	4	30	55	50	100
	5	70	20	40	130
	6	60	30	100	150
	7	50	25	70	250
	8	60	30	90	170
	9	100	40	130	80
	10	40	35	140	220
	Total	800	300	900	1500

N (rural) = 800

N (urban) = 1600

Estimated vehicle-miles:

Rural:

Week 1: $\frac{5(800)}{2}$ (500 + 400) = 1,800,000

Week 2: $\frac{5(800)}{2}$ (250 + 600) = 1,700,000

Week 3: $\frac{5(800)}{2}$ (300 + 600) = 1,800,000

Week 4: $\frac{5(800)}{2}$ (800 + 300) = 2,200,000

Weeks 1-4: 7,500,000

(If the study had extended over 52 weeks, and if the rural volumes were of the order of magnitude found in the first two weeks, the estimated rural vehicle-miles would be of the order of magnitude of 97,500,000)

Estimated vehicle-miles:

Urban:

Week 1: $\frac{5(1600)}{20}$ (800 + 400) = 480,000

Week 2: $\frac{5(1600)}{20}$ (700 + 1300) = 800,000

Week 3: $\frac{5(1600)}{20}$ (1000 + 600) = 640,000

Week 4: $\frac{5(1600)}{20}$ (900 + 1500) = 960,000

Weeks 1-4: 2,880,000

(If the urban study had extended over 52 weeks, and if the urban volumes were of the order of magnitude found in the first four weeks, the estimated urban vehicle-miles would be of the order of magnitude of 37,440,000)

Standard errors:

Rural (equation 1a and equation 6a)

Week 1: $\frac{5(800)}{2}\left|500 - 400\right| = 2000(100) = 200,000$

Week 2: $\frac{5(800)}{2}\left|250 - 600\right| = 2000(350) = 700,000$

Week 3: $\frac{5(800)}{2}\left|300 - 600\right| = 2000(300) = 600,000$

Week 4: $\frac{5(800)}{2}\left|800 - 300\right| = 2000(500) = 1,000,000$

Weeks 1-4: $\frac{5(800)}{2}\sqrt{(100)^2+(350)^2+(300)^2+(500)^2} = 2000\ (687.4) = 1,374,800$

(With the same assumptions as for the estimates of vehicle-miles, the standard error for 52 weeks of the rural sample will be of the order of magnitude of $2000(\sqrt{13})(687.4)$, which equals approximately 5,000,000)

Urban (equation 1b and equation 6b)

Week 1: $\frac{5(1600)}{20}\left|800 - 400\right| = 400(400) = 160,000$

Week 2: $\frac{5(1600)}{20}\left|700 - 1300\right| = 400(600) = 240,000$

Week 3: $\frac{5(1600)}{20}\left|1000 - 600\right| = 400(400) = 160,000$

Week 4: $\frac{5(1600)}{20}\left|900 - 1500\right| = 400(600) = 240,000$

Weeks 1-4: $\frac{5(1600)}{20}\sqrt{(400)^2+(600)^2+(400)^2+(600)^2} = 400(721.11) = 288,444$

(With the same assumptions, the standard error of 52 weeks of urban sampling will be approximately 1,040,000)

Relative errors

Rural

Week 1: $\frac{200,000}{1,800,000} = .1111 = 11.1\%$

Week 2: $\frac{700,000}{1,700,000} = .4117 = 41.2\%$

Week 3: $\frac{600,000}{1,800,000} = .3333 = 33.3\%$

Week 4: $\frac{1,000,000}{2,200,000} = .4545 = 45.5\%$

Weeks 1-4: $\frac{1,374,800}{7,500,000} = .1833 = 18.3\%$

Weeks 1-52: $\frac{5,000,000}{97,500,000} = .0512 = 5.1\%$

Urban

Week 1: $\frac{160,000}{480,000} = .3333 = 33.3\%$

Week 2: $\frac{240,000}{800,000} = .3000 = 30.0\%$

Week 3: $\frac{160,000}{640,000} = .2500 = 25.0\%$

Week 4: $\frac{240,000}{960,000} = .2500 = 25.0\%$

Weeks 1-4: $\frac{288,444}{2,880,000} = .1001 = 10.0\%$

Weeks 1-52: $\frac{1,040,000}{37,440,000} = .0277 = 2.8\%$

If the data items don't change radically as the study progresses, the example indicates that the relative sampling error of the study total is expected to decrease as the time period covered by the study increases.

Estimated vehicle-miles:

Rural and urban

Week 1: 1,800,000 + 480,000 = 2,200,000

Week 2: 1,700,000 + 800,000 = 2,500,000

Week 3: 1,800,000 + 640,000 = 2,440,000

Week 4: 2,200,000 + 960,000 = 3,160,000

Weeks 1-4: 10,300,000

Standard errors

Rural and urban (equations 4 and 8)

Week 1:
$$\sqrt{\left[\frac{5(800)}{2}\right]^2 (500-400)^2 + \left[\frac{5(1600)}{20}\right]^2 (800-400)^2} =$$

$$\sqrt{65600000000} = 256125$$

Week 2:
$$\sqrt{\left[\frac{5(500)}{2}\right]^2 (250-600)^2 + \left[\frac{5(1600)}{20}\right]^2 (700-1300)^2} =$$

$$\sqrt{547600000000} = 740000$$

Week 3:
$$\sqrt{\left[5\ \frac{(800)}{2}\right]^2 (300-600)^2 + \left[\frac{5(2600)}{20}\right]^2 (1000-600)^2} =$$

$$\sqrt{385600000000} = 620967$$

Week 4:
$$\sqrt{\left[\frac{5(800)}{2}\right]^2 (800-300)^2 + \left[\frac{5(1600)}{20}\right]^2 (400-1500)^2} =$$

$$\sqrt{1057600000000} = 1028397$$

Weeks 1-4:
$$\sqrt{(656 + 5476 + 3856 + 10576)\ 10^8} = 1434016$$

Relative errors

Rural and urban

Week 1: $\frac{256,125}{2,200,000}$ = .1164 = 11.6%

Week 2: $\frac{740,000}{2,500,000}$ = .2960 = 29.6%

Week 3: $\frac{620,967}{2,440,000}$ = .2544 = 25.4%

Week 4: $\frac{1,028,397}{3,160,000}$ = .3254 = 32.5%

Weeks 1-4: $\frac{1,434,016}{10,300,000}$ = .1392 = 13.9%

(If the study extends over a 52 week period, it is expected that the relative sampling error will not exceed 3 percent of the estimate of vehicle-miles.)

Appendix B

Link-Day Sampling

Introduction

This statistical sampling procedure is based upon a sampling of 24-hour volume counts. The procedure was designed to yield estimates of urban vehicle-miles of travel (VMT) within $\pm$ 5 percent of the actual VMT at the 95 percent confidence level. The results obtained in pilot studies did not quite attain this goal. The procedure may be used to obtain estimates in a single urban area, or for statewide estimates of urban VMT. The following description and sample problem illustrate the mechanics of selecting the locations and dates for the sample counts. The resulting sample sizes are not intended to indicate the amount of counting required in any particular State.

The Technique

This technique involves stratified random sampling of 24-hour volumes utilizing the link-day as the basic sampling unit. A link in a street system is considered to be one street intercepted by two other streets. If a street link has abrupt volume changes due to abutting land uses, the link should be further divided into two or more links. A "link-day" is a 24-hour period for a given link. Based on the desired accuracy of the this VMT estimates and assumed values for the standard deviations of the strata, the required number of sample counts (n) can be determined from the following formulas:

$$n = S^2 \div E^2 \qquad (1)$$

$$n = \frac{(\Sigma N_i \; S_i)^2}{(NE)^2 + \Sigma Ni \; Si^2} \qquad (2)$$

Where: n = total number of 24-hour sample counts

E - desired error of the urban VMT estimate

S = assumed overall standard deviation of link vehicle-miles based upon variation in link volumes

S_i = assumed standard deviation of the link vehicle-miles based on variation in link volumes in volume group "i"

N = overall total number of link-days from which a sample may be drawn

N_i = total number of link-days from which a sample may be drawn from volume group "i." If there are 5100 links in a volume group, the total number of link-days from which the sample could be selected is 5100 x 365 or 18.61×10^5 link-days.

Formula (1) will be used for samples from volume group 1-999 and formula (2) for the remaining volume groups. Originally it was intended to apply the random sampling technique to all volume groups. However, investigations showed that volume group 1-999 had many more links than any of the remaining groups, thus overweighting the estimate of the needed sample size, n. It was decided to analyze the data as two separate universes. The use of the two formulas requires the following assumptions:

1. Error of urban VMT estimate (E) - The error of the VMT estimate should be fine enough to provide statewide estimates of urban VMT which are sensitive to annual change. The desirable error selected for use in specific urban areas for transportation planning purposes is $\pm$ 5 percent at the 95 percent confidence level ($1\sigma = 2.5$ percent). In order for statewide estimates from year-to-year to produce meaningful trend values, it seems appropriate to adopt the same accuracy goal for statewide estimates of urban VMT. Individual States may determine that somewhat lesser accuracy with the corresponding decrease in sample size is acceptable.

2. Standard deviation of link-day vehicle-miles (S or S_i) - In order to utilize formulas (1) and (2), we must estimate the standard deviation of the vehicle-miles contributed by each link-day within each volume group. This is the "S_i" in formula (2). To determine the sample design for the first year, it may be assumed that the lengths of the links in any volume group are approximately constant. The variability in the vehicle-miles will then be proportional to the variation in the volumes. It may be further assumed, for the first year's design, that the volumes are approximately normally distributed within each volume group. A range of $\pm$ 3σ around the group mean

would then account for almost 100 percent of the links. An estimate of the standard deviation, S_i, for a volume group could then be obtained by dividing the range (R) of the volume group by 6, However, because of the variability of daily vehicle counts about the ADT, a safer estimate of S_i would be obtained by dividing the range by 5 ($S_i = R_i \div 5$) when the sample is designed to estimate urban VMT for weekdays only. If the VMT estimate is to include weekends, the range should be divided by 4.5 ($S_i = R_i \div 4.5$). In volume group 1-999, the effective range of volume is approximately 500 rather than 1000, therefore a value of S=115 could be used ($S = 500 \div 4.5 = 111$). These procedures for estimating S and S_i are useful for designing the first year's sample. In succeeding years, a State may calculate estimates of S and S_i from the sample results.

3. Number of samples (n) - In this procedure "n" refers to the number of 24-hour volume counts that must be taken on the selected links (L) on the days determined in the sampling. Since the basic sampling unit is a "link-day," the random sampling of each volume group will select both the link to be counted and the day on which the count will be taken.

4. Identifying the population to be sampled - The object of this sampling is to produce an estimate of annual urban VMT. Therefore the sample counts will be distributed throughout the year on links of the street systems for which the estimates are to be made. If L represents the total number of urban street links, the population of link-days to be sampled (N) equals 365L (N = 365L). If the VMT estimate is for non-holiday weekdays only, we can assume 250 non-holiday weekdays. Then N = 250L.

Illustrative Problem

The enclosed table and computation sheet illustrate how a sample size can be calculated and distributed among several volume groups of street links. As a first step all urban street links in the State to be considered in the estimate must be identified and classified into volume groups. The table of the enclosed problem shows volume groups that can be used. The volume groups for different States may vary. The lower volume groups will probably remain similar; however, different combinations of the higher volume groups may be advisable. The final selection of volume groups should be based on the following rule: The ratio of the highest VMT for a group to the lowest VMT for a group shall be no greater than two. In the illustrative problem the volume group 10,000-13,499 has the largest estimated VMT of 4,112,500 while the volume group 2000-2999 has the lowest VMT of 2,000,000. Since the ratio is only slightly greater than two the volume groupings will be used. The volume group under 1000 ADT need not be included in testing the rule since it will be sampled separately.

Figures for the "Estimated daily VMT" column may be based on estimated traffic volumes. Hopefully, much of these data have been gathered for cities over 50,000 population in the course of current transportation studies. Traffic data gathered from any State, county, or local source for the smaller cities may be used. Lacking sufficient data, street links may be assigned to volume groups based on judgment. The question arises of how to treat a selected link when it is found that the counted volume falls outside its assigned volume group. In this case the analysis is completed using the initial volume group assignment. The number of these instances should be minor. For subsequent years' estimates, the assignment of links to volume groups should be revised to reflect the result of actual volume counts.

The table and computation sheet illustrate the computation of sample size: 830 counts for links under 1000 ADT and 1770 counts for links of 1000 ADT and over. Once the total sample size for the higher volume groups has been calculated, it can be distributed among the several volume groups according to the formula:

$$n_i = \frac{N_i\ S_i}{\Sigma N_i\ S_i}\ n$$

where n_i = sample counts in volume group "i." The remaining terms have been previously defined. The last column in the table shows how the 2600 total sample would be distributed among the volume groups. If the computation of the required sample for the volume groups under 1000 ADT results in an n less than 50, in order to achieve stability in the estimate of standard error from the sample, a minimum of 50 sample counts is recommended. If the desirable standard deviation had been 5 percent rather than 2.5 percent, the number of sample counts for the higher volume groups would have been approximately 442 (1770 divided by 4). The links and the days on which the counts will be taken are chosen randomly. Sample counts are selected from all days throughout the year. If the VMT estimate is to be for weekdays only, weekends and holiday weekdays should be excluded from the sample selected.

In selecting the sample of link-days, each link will have the chance of being selected for each of 365 days. In other words, each link in a volume group will appear 365 times in the population to be sampled. In the illustrative problem, volume group 3000-4999 includes 5500 links. The total population from which the sample is drawn is 5500 x 365 or 20.07×10^5 link-days. Numbers from 1 to 2,007,000 are assigned to the links with 1 through 365 assigned to the first link, 366 through 730 to the second link, etc. From a random number table, 26 numbers less than 20.07×10^5 are selected. These numbers identify the selected sample links and the days on which the counts should be taken.

When the 24-hour volumes have been obtained the data are analyzed by volume group. The product of the volume and the sample link length ($y_{ij} \cdot l_{ij}$) is the vehicle-miles of travel (X_{ij}) for the sample link-day. The total estimated VMT for the volume group is obtained by multiplying the average vehicle-miles per sample link ($\sum_{i=1}^{n} X_{ij} \div n_i$) by the population of link-days in the volume group (N_i). Summation of the individual group estimates results in the overall statewide estimate of urban vehicle-miles of travel. From these data the actual error of the estimate obtained may be calculated and desirable changes in the sample for subsequent years made. The formula below may be used to calculate the standard error from the sample:

$$s^2_{X'} \doteq \sum_{i=1}^{m} \left(\frac{N_i}{n_i}\right)^2 \frac{1}{(n_i - 1)} \left[n_i \sum_j X^2_{ij} - \left(\sum_j X_{ij}\right)^2\right]$$

Relative Error = $s_{X'} \div X'$

where: y_{ij} = 24-hour count of jth link-day in the ith stratum

l_{ij} = length of link associated with the y_{ij} count

$X_{ij} = y_{ij} \cdot l_{ij}$ = observed VMT resulting from a sample count

$X' = \sum_{i=1}^{m} \frac{N_i}{n_i} \sum_{j=1}^{n_i} X_{ij}$ = estimate of total vehicle-miles

Any State conducting a study of this type should make this calculation. A review of the volume data will show any links that should be reassigned to other volume groups.

Illustrative Problem

Calculate Sample Size

Volume Groups Over 999

Average Link Loading = 22,665,900 ÷ 44,600 = 508.20 vehicle-miles per link

$E @ 1\sigma = .025\ (508.20) = 12.70$ = absolute VMT error per link

$NE = (162.77 \times 10^5)\ 12.70 = 20.67 \times 10^7$ = absolute error on all links

$$n = \frac{(\Sigma N_i S_i)^2}{(NE)^2 + \Sigma N_i S_i^2} = \frac{(86.97 \times 10^8)^2}{(20.67 \times 10^7)^2 + (88.21 \times 10^{11})}$$

$$= \frac{7563.78 \times 10^{16}}{(427.25 \times 10^{14}) + (.09 \times 10^{14})} = \frac{7563.78 \times 10^{16}}{427.34 \times 10^{14}}$$

n = 1770

Volume Group 1-999

Average Link Loading = 6,400,000 ÷ 80,000 = 80.00

$E @ 1\sigma = .05\ (80.00) = 4.0$

$S = 115$

$$n = S^2 \div E^2 = \frac{13,225}{16.0}$$

$n = 827$ **USE n = 830**

Total Number of Samples = 2600

Illustrative Problem

Volume group	Range (R)	No. of links (L_i)	Miles of streets	Estimated daily VMT	$S_i = R/4.5$*	S_i^2 ($X10^4$)	$N_i = 365L_i$ ($X10^5$)	$N_i S_i$ ($X10^8$)	$N_i S_i^2$ ($X10^{11}$)	Number of samples
1-999	1,000	80,000	8,000	6,400,000	115	1.32	292.00	32.58	3.85	830
1,000-1,999	1,000	14,000	1,400	2,100,000	225	5.06	51.10	11.50	2.59	234
2,000-2,999	1,000	8,000	800	2,000,000	225	5.06	29.20	6.57	1.48	134
3,000-4,999	2,000	5,500	550	2,200,000	445	19.80	20.07	8.93	3.97	182
5,000-6,999	2,000	4,000	400	2,400,000	445	19.80	14.60	6.50	2.89	132
7,000-9,999	3,000	4,100	410	3,485,000	670	44.89	14.96	10.02	6.71	204
10,000-13,499	3,500	3,500	350	4,112,500	780	60.84	12.77	9.96	7.77	203
13,500-18,999	5,500	3,500	250	4,062,500	1,225	150.06	12.77	15.64	19.16	318
19,000 +	11,000	2,000	200	2,305,900	2,445	597.80	7.30	17.85	43.64	363
Total over 999		44,600	4,360	22,665,900			162.77	86.97	88.21	1,770
Total		124,600	12,360	29,065,900			454.77	119.55	92.05	2,600

* For Range (R)
- 1000: $S_i = 1000/4.5 = 222$, Use 225
- 2000: $S_i = 2000/4.5 = 444$, Use 445
- 3000: $S_i = 3000/4.5 = 666$, Use 670
- 3500: $S_i = 3500/4.5 = 778$, Use 780
- 5500: $S_i = 5500/4.5 = 1222$, Use 1225
- 11000: $S_i = 11{,}000/4.5 = 2444$, Use 2445