viding for-hire transportation to the public. It has had a rather checkered career in that it began in 1963 as part of the Census of Transportation, was transferred to the Census of Selected Services in 1967 and 1972, and returned to the fold in 1977.

The survey presents data only on that portion of the motor carrier industry not subject to economic regulation by the ICC and the public warehousing industry. As a result, it has a rather limited audience, primarily intended to fill a void of information needed for the national accounts.

It is being conducted under substantially the same methodology as the previous surveys except that certain questions on expenses and fringe benefits were added. Also, the categories of revenue freight equipment were made comparable to the definitions used by ICC.

There are currently no plans to expand this survey in 1982. However, the Advisory Committee on Gross National Product Data Improvement recommended that there should be a "census of transportation industries with a full complement of questions on inputs and outputs by relevant subindustries." The major obstacle in implementing this recommendation is the restriction in Title 13, U.S. Code, under which the Census Bureau operates, that prohibits the Census Bureau from collecting data for means of transportation for which statistics are required by law to be filed with a designated body (3).

CONCLUSION

The Census of Transportation was originally designed to fill in major data gaps without unnecessary duplication. Therefore, the general structure and objectives of the transportation program differ from those of the orthodox economic censuses, largely because of the availability of various types of transportation data from other government and private sources (4).

While the individual surveys have been expanded and improved on over the years, the goal of the Census of Transportation has not changed since it was conceived nearly 20 years ago. It would appear that the original objective has been substantially achieved. Policymakers, planners, and other users have much more information than existed 20 years ago and are able to make decisions more effectively.

However, there is still much left to do. In this day when questions about energy and transportation are critical, perhaps a new goal should be espoused--the goal of providing an integrated, comprehensive body of statistics that serve rather than confuse the user community. The current state of transportation statistics is characterized by a whole series of conflicting pieces of information, a veritable jigsaw puzzle. New approaches must be conceived if we are to be able to put the pieces together and develop a body of statistics that accurately describe the real world.

REFERENCES

1. Travel During 1977, Census of Transportation. Bureau of the Census, U.S. Department of Commerce, Final Rept., TC77N, 1979.
2. A. R. Tupek and A. M. Perez. Can U.S. Businesses Take a Sample of Their Own Records? Presented at the American Statistical Association Meeting, Washington, DC, Aug. 13-16, 1979.
3. Gross National Product Data Improvement Project Report. Office of Federal Statistical Policy and Standards, U.S. Department of Commerce, Oct. 1977.
4. D. E. Church. Plans for the 1963 Census of Transportation. Transportation Division, Bureau of the Census, U.S. Department of Commerce, Dec. 10, 1962.

Evaluation of the FHWA Vehicle Classification and Automobile-Occupancy Sampling Manual

R. CRAIG HUPP AND CARMINE PALOMBO

The Federal Highway Administration recently sponsored the development of a how-to manual for conducting automobile-occupancy and/or vehicle classification studies. A draft version of the manual, Guide for Estimating Urban Vehicle Classification and Occupancy, was field tested by the Southeast Michigan Council of Governments in cooperation with the Michigan Department of Transportation. This paper describes how the guide was used to set up a regional vehicle classification and automobile-occupancy study. Study procedures and results are discussed. Overall, the guide was found to be an excellent how-to manual for vehicle classification and automobile-occupancy studies. Perhaps the largest contribution of the guide is the short-count sampling approach it recommends. This approach results in a significant survey cost saving with no loss of accuracy.

The U.S. Department of Transportation continuously reviews and evaluates the kinds of data required to support national and local decisions on transportation. In support of this process the Transportation Research Board established the Advisory Committee on Urban Transportation Data-Reporting Needs and Requirements in 1975. The committee identified vehicle type and automobile occupancy as two important traffic variables that describe the highway system and recommended that the Federal Highway Administration (FHWA) develop and test sampling methods to obtain these data (1). The FHWA implemented this recommendation by contracting with Peat, Marwick, Mitchell, and Company to develop a sampling manual entitled Guide for Estimating Urban Vehicle Classification and Occupancy (2). In addition, the FHWA worked with several metropolitan planning organizations (MPOs) to have them test the procedures described in the guide.

The Southeast Michigan Council of Governments (SEMCOG), the MPO for the southeast Michigan region, in conjunction with the Michigan Department of Transportation (MDOT), was one of two MPOs that tested the guide. This paper describes how the guide was used in setting up a vehicle-classification and automobile-occupancy survey and what the results were of the survey that SEMCOG conducted. This paper places emphasis on the practical aspects of developing the survey-sampling plan and analyzing the results. The statistical approach on which the guide is based and that is the key to this economical approach to data collection is not discussed in detail here. The statistical approach incorporated in the guide is summarized in Field Data and Sampling Procedures for Measuring Regional Vehicle Classification and Occupancy (3).

The guide presents procedures to collect automobile-occupancy and vehicle-classification data and to estimate vehicle miles of travel and person miles of travel. The travel estimates are developed by combining vehicle-classification and automobile-occupancy data with traffic-count data from a regional traffic-counting program. These data collected on an areawide basis can be used to support five technical analyses:

1. Evaluate the effectiveness of short-range transportation programs,
2. Assess transportation-related air quality,
3. Assess the energy efficiency of travel,
4. Validate urban transportation planning models, and
5. Monitor general trends in traffic and travel characteristics.

The procedures in the guide can also be used on a corridor or project basis as part of a before-and-after study or as part of project planning.

DEVELOPING THE SAMPLING PROGRAM

The guide stresses the need to clearly state the survey objectives before a sampling plan is developed. The principal objective of SEMCOG's study was to obtain up-to-date regional data for automobile occupancy and vehicular classification. At the time the study was undertaken the most recent data for these items were from a 1965 origin-destination survey. In addition, these data will be used to track trends in ridesharing. Finally, these data will be used to improve SEMCOG's vehicular-emissions and fuel-consumption models and to check the results of SEMCOG's model recalibration process. Obviously, another important objective of the study was to evaluate the guide for its clarity, content, and usefulness in performing vehicle-classification and automobile-occupancy surveys. These considerations are discussed more fully below.

The sampling program consists of the specification of the desired tolerance and confidence level of the estimate, the location-stratification scheme, the selection of sampling locations, and the selection of categories under which these data will be collected. The key to developing a good sampling plan is to use prior knowledge of the characteristics of the data that are to be collected. Prior data can be used to design a program that samples the minimum number of sites required to estimate desired population parameters with a given accuracy and precision.

Definitions

The following definitions were used in the vehicle-occupancy portion of the study:

1. One-person passenger vehicles--includes non-commercial pickup trucks, vans, and private automobiles and taxis;
2. Two-person passenger vehicles;
3. Three-person passenger vehicles; and
4. Passenger vehicles carrying four or more passengers.

The following definitions were used in the vehicle-classification portion of the study:

1. Panel and pickup trucks that have less than 1-ton cargo-carrying capacity, defined as pickup trucks and vans that were obviously commercial and that either contained side markings or construction equipment;
2. Other single-unit trucks, defined as having the cab and cargo area within the same frame served by two or three axles;
3. All truck combinations not otherwise classified (this classification consisted mainly of tractor-trailer trucks);
4. Motorcycles and/or motor scooters, with or without sidecars;
5. Buses, not counted;
6. Motor homes, classified as other single-unit trucks; and
7. Small pickup trucks, classified in the same way as larger pickup trucks (presence of commercial equipment or markings meant inclusion in the panel-and-pickup under 1-ton category; otherwise, they were considered as passenger vehicles subject to the occupancy criteria).

Stratification of Sampling Sites

When a sample population group is known or suspected to include subgroups that have more homogeneous characteristics than the population as a whole, a stratified sampling plan designed to sample each subgroup can provide both a more efficient estimate (in terms of sample size) of average characteristics of the population as a whole and more information about each subgroup within the population. Sample size is a direct function of the variation of the population measured in terms of the variance or standard deviation of the population parameter being sampled. Sampling by subgroup reduces the variation and hence the sample size needed to obtain a given level of accuracy.

For example, existing MDOT and SEMCOG data indicated that the proportion of trucks varies significantly by functional class. It was suspected that it also varied by area type (urban or rural), although sufficient data were not available to verify this assumption. Because of this, SEMCOG decided to use a stratified sampling approach for vehicle-classification sampling. This dictated that a stratified approach also be used for automobile occupancy because automobile-occupancy data would be collected at the same time.

In addition, other planning efforts at SEMCOG had indicated a need to know whether the proportion of trucks was the same on high-volume roads as on low-volume roads. It was decided to stratify sites based on traffic volumes as well as functional class and area type in order to determine whether this were the case. Arterial links were stratified by volume to determine whether the proportion of trucks varied by traffic volume. An average daily volume of 35 000 vehicles/day was selected as a stratification cutoff point because a review of traffic volumes in the region indicated that most of the roads of regional significance were carrying volumes in excess of 30 000 to 35 000 vehicles/day. This stratification applied only to urban arterials because all freeways in the region carry more than 35 000 vehicles/day and rural arterials do not.

It should be noted that, of the three stratifications (functional class, area type, and traffic volume), the first was selected because of known variations in the population while the other two were selected to test for suspected variations. Note also that, even though these stratifications were developed to improve sampling for truck travel, they are also relevant to automobile occupancy. Intuitively, urban freeways would be expected to have a greater proportion of work-trip travel and, hence, lower occupancies during peak periods than other combinations of functional class and area type. Conversely, rural freeways would be expected to have higher occupancies because of recreational travel and because longer trip lengths promote ridesharing. Thus the stratification for trucks was expected to provide useful information about automobile occupancy within a given tolerance and confidence level.

Calculation of Sample Size

Sample size is a function of the tolerance and level of confidence desired in the sample estimate and the variation in the parameter being estimated. For this study two parameters were estimated, average automobile occupancy and proportion of trucks. Whichever parameter requires the larger sample size for a given tolerance and level of confidence will determine the total sample required. In this study, the unit being sampled was the one-way link day, where the link is a unit distance on the highway network.

The guide provides the following formula to calculate the minimum sample size required to estimate population parameters within a given tolerance and confidence level.

$$N = (Z^2 \cdot S^2)/D^2 \qquad (1)$$

where

N = minimum sample size in one-way link days;
Z = normal variate for the (1-α) level of confidence, two-tailed test where α is the level of confidence;
S = composite standard deviation of the sample; and
D = tolerance or acceptable difference between the estimated value of the population parameter and the true value.

The composite standard deviation of the sample S is a function of the manner in which the population parameter varies as a function of several other variables, including the following:

1. The variation across link days within a season,
2. The variation from season to season,
3. The variation across time periods during a day as a result of short counts,
4. The variation between lanes where the short-count procedure includes sampling between lanes,
5. The variation introduced by human error in the survey process, and
6. Other sources.

The composite standard deviation is equal to the square root of the sum of the squares of the standard deviations attributed to each of these sources of variation.

Depending on the sampling program, some of these sources of variance will not apply. For example, the seasonal variation factor was not relevant because the SEMCOG survey was conducted during only one season and an annually adjusted average automobile occupancy or proportion of trucks was not desired. Variation introduced by human error was ignored, although it does exist. Variation across traffic lanes was also omitted on SEMCOG's sample-size estimate because it was not significant. Hence, the only two sources of variation that SEMCOG considered in calculating the minimum sample size were variation across link days and variation across time periods during the day.

The guide provides default values for these standard deviations to be used where there are insufficient local data to permit their calculation. SEMCOG used the default data values. As discussed in the section on results, the composite standard deviation (SD) observed in the data SEMCOG collected was typically smaller than the composite standard deviation obtained from the guide's default values. The default values used are noted below:

Measure	Variation Source	Default Values Used	Composite SD
Proportion of trucks	Location and day	0.040	0.041
	Within day	0.009	
Average automobile occupancy	Location and day	0.063	0.065
	Within day	0.017	

The desired tolerance and confidence levels represent a trade-off between the desired accuracy of the estimate and data-collection costs. SEMCOG specified a 95 percent confidence level (α). This means a sample size was chosen that would result in the estimated average automobile occupancy or proportion of trucks falling within the desired levels of tolerance about the true value of the estimated parameter 19 times in 20.

SEMCOG established maximum tolerance levels D for both average automobile occupancy and proportion of trucks at ±0.03 for freeways locations and ±0.04 for arterial locations. That is, the average automobile-occupancy estimate will be within ±0.03 persons/automobile for freeways or ±0.04 persons/automobile for arterials of the true value that would be found if all vehicles passing all locations were counted. For vehicle classification, total proportion of trucks would be within ±0.03 trucks/total vehicles for freeways and ±0.04 trucks/total vehicles for arterials of the true value. In relative terms, assuming the true average automobile occupancy is 1.30 persons/automobile, specified maximum tolerance of the estimate will be approximately ±3 percent of the average occupancy. For the proportion of trucks the relative tolerance would be greater. Assuming the proportion of trucks on arterials and freeways is 0.05 and 0.15, respectively (i.e., 5 percent and 15 percent of total traffic), the relative tolerance would be ±80 percent and ±20 percent for arterials and freeways, respectively.

However, because these parameters (i.e., average automobile occupancy and proportion of trucks) do not have the same level of variance, different sample sizes would be required to achieve the same tolerance levels. Because the two parameters were being sampled at the same time, the sample that had the most variance determined total sample size. Sample size for both automobile occupancy and proportion of trucks was calculated for each survey cell. In each case the automobile-occupancy parameter required the larger sample size. In terms of the total sample required, the minimum sample required to estimate automobile occupancy was 69 one-way link days and, for proportion of trucks, 28 one-way link days.

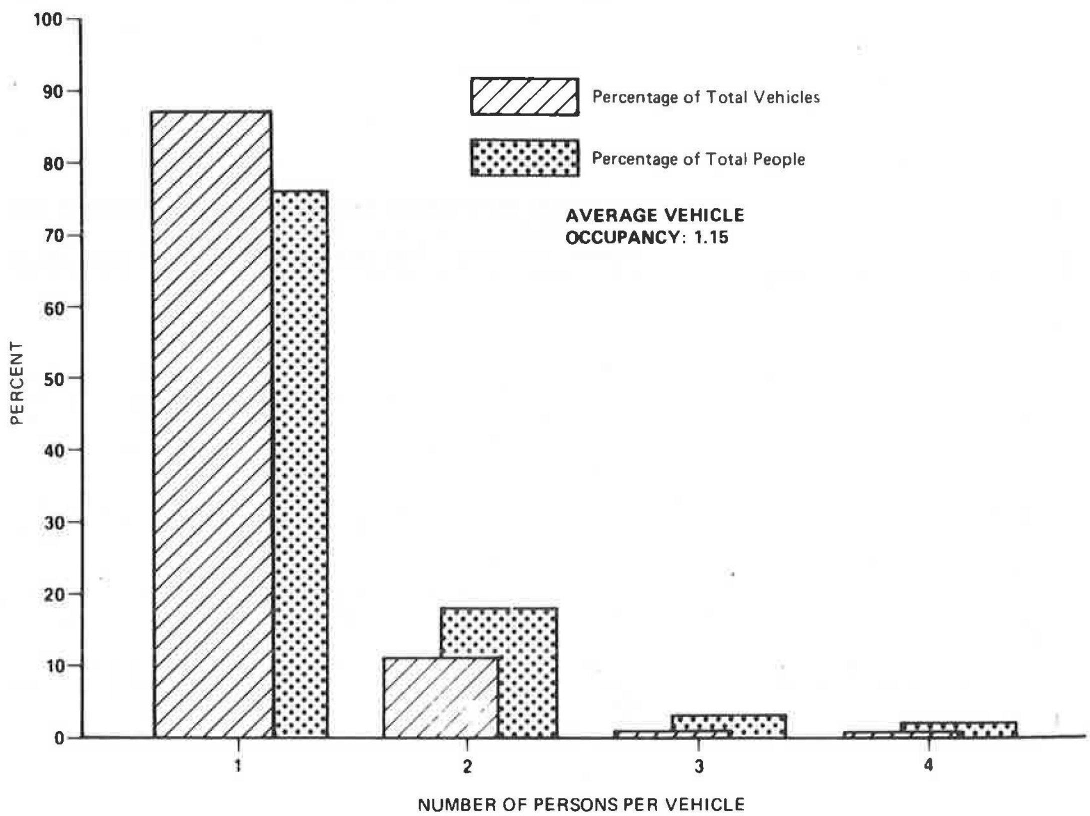
Because more than the minimum sample size was collected for proportion of trucks, the final

Table 1. Study results.

Time Period	Stratification Cell	Automobile Occupancy: Average Occupancy	Composite SD	Tolerance[a]	Vehicle Classification: Proportion of Trucks	Composite SD	Tolerance[a]
7:00-9:00 a.m.	Urban freeway	1.13	0.02	0.01	0.16	0.03	0.02
	Rural freeway	1.24	0.04	0.02	0.26	0.04	0.02
	Urban arterial[b]	1.21	0.08	0.04	0.05	0.02	0.01
	Urban arterial[c]	1.14	0.03	0.01	0.08	0.03	0.01
	Rural arterial[b]	1.25	0.12	0.07	0.13	0.04	0.02
	Total	1.17		0.015	0.11		0.0086
11:00 a.m.-1:00 p.m.	Urban freeway	1.28	0.04	0.02	0.24	0.04	0.02
	Rural freeway	1.45	0.05	0.03	0.32	0.05	0.03
	Urban arterial[b]	1.41	0.07	0.03	0.09	0.03	0.01
	Urban arterial[c]	1.34	0.05	0.02	0.09	0.03	0.01
	Rural arterial[b]	1.43	0.07	0.04	0.15	0.05	0.04
	Total	1.36		0.016	0.14		0.0096
2:00-6:00 p.m.	Urban freeway	1.24	0.02	0.01	0.18	0.02	0.01
	Rural freeway	1.35	0.04	0.03	0.25	0.03	0.02
	Urban arterial[b]	1.39	0.04	0.02	0.06	0.02	0.01
	Urban arterial[c]	1.31	0.04	0.02	0.06	0.02	0.01
	Rural arterial[b]	1.43	0.07	0.04	0.10	0.03	0.02
	Total	1.33		0.012	0.10		0.0061

[a]At 95 percent confidence level. [b]Less than 35 000 vehicles/day. [c]More than 35 000 vehicles/day.

Figure 1. Distribution of automobile occupancy, urban freeway, morning peak hours.



tolerances for this parameter were smaller than those for average automobile occupancy. Final tolerances for proportion of trucks were ±0.019 for freeway locations and ±0.025 for arterials. The final sampling matrix is shown below:

Area Type	Functional Class: Freeway	Arterial (vehicles/day): ≤35 000	≥35 000
Urban	18	11	11
Rural	18	11	X

Sampling Methods

The unit sampled was the one-way link day. Hence, a two-way street represents two possible sampling sites. The guide (2) suggests using short-count data-collection techniques to reduce data-collection costs. This technique involves the periodic sampling of traffic characteristics. The most practical way to accomplish this is to sample lanes one at a time on high-volume multilane facilities. For example, on a six-lane urban freeway, a single surveyor would cover one travel direction and would collect data from the three lanes in that direction for 15 min each. By comparison, an alternative approach frequently used by MDOT involves six surveyors (one for each lane), a supervisor, and, possibly, a backup person to continuously count all lanes in both directions. The cost advantages of the one-way, short-count approach are obvious.

Selecting the Sample Links

The next step was to select a random sample of highway links for each cell of the sampling matrix. The highway network used for sampling was SEMCOG's computer-coded regional highway network, which included all highway links classed as minor arterials or above. Approximately 5000 miles of streets and highways are included in this network. SEMCOG's computerized file for this network represents each road in the system as a series of links of varying length that have as their termini intersections with other roads in the regional system. For each link, a variety of jurisdictional, physical, functional, and operational data are maintained, including functional class, area type, and volume.

To ensure the validity of the random sampling process, it was necessary to consider that the highway links were made up of a series of sublinks. Each sublink was 0.16 km (0.1 mile) long, and had an equal probability of being selected. In practical terms, this meant weighting the link selection process so that the probability of selecting a link was in direct proportion to its length. That is, a link 0.8 km (0.5 mile) long should have a probability of being selected that is 5 times greater than a link only 0.16 km (0.1 mile) long. To accomplish the weighted link-selection process a computer program was written that analyzed each link in the computerized highway network, assigned it to the proper survey cell, and then randomly sampled the sublinks in the survey cell.

PERFORMING THE SURVEY

The survey was performed as a joint effort by SEMCOG and MDOT. MDOT provided survey crews to cover all freeway locations. SEMCOG retained a consultant to survey all arterial locations. Some 69 locations were surveyed.

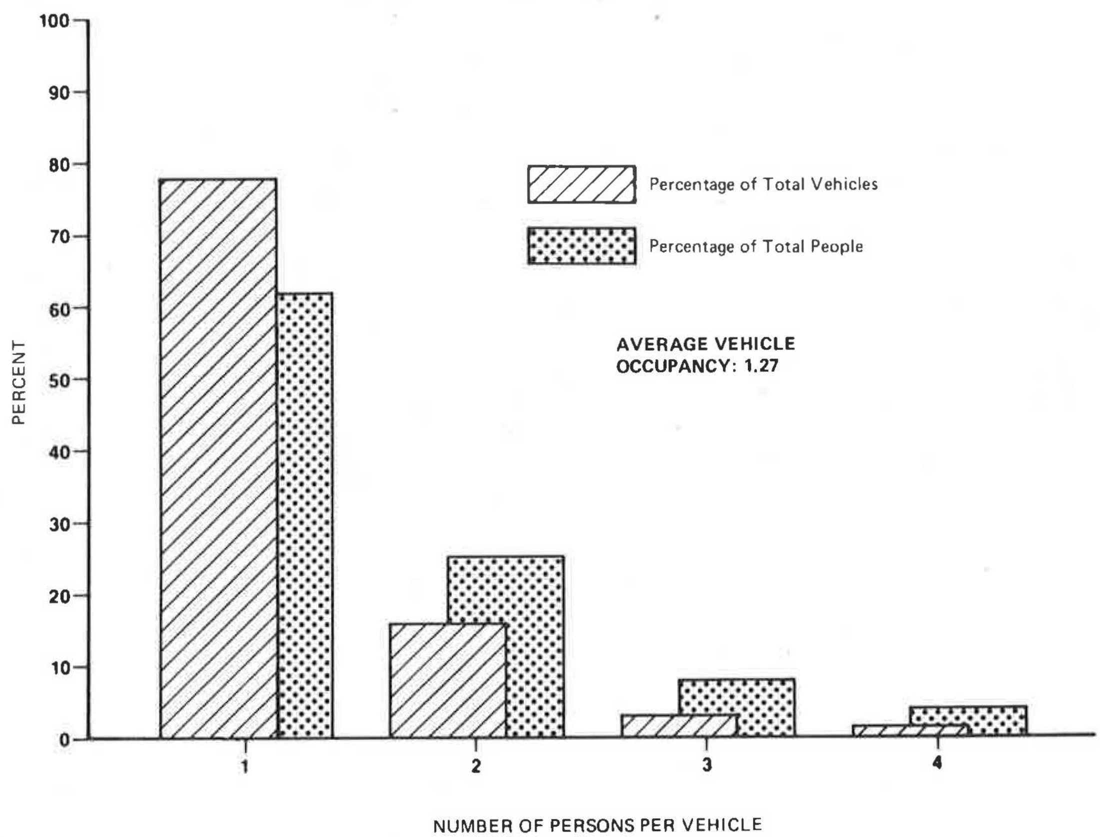
Data were collected for the periods 7:00-9:00 a.m., 11:00 a.m.-1:00 p.m., and 2:00-6:00 p.m. Each lane was counted for 15 min, followed by a 5-min interval to record the counts and reset the counters to zero.

The table below summarizes the survey costs, including SEMCOG staff time to administer the consultant contract. By using part-time labor, the consultant was able to survey at one-half of MDOT's cost per site because MDOT used full-time department survey crews.

Item	Cost ($)
Consultant costs (data collection on arterials, 33 link days)	2 750
Estimated MDOT costs (data collection on freeways, 36 link days)	6 000
SEMCOG staff costs (including staff time for administration of contract, coding, and computer programming, estimated at approximately 600 h)	7 500
Total	16 250

Had MDOT's usual practices for recording vehicle classification and occupancy data been used, it is estimated that the costs of the study would have been substantially higher. In conducting

Figure 2. Distribution of automobile occupancy, rural arterials, morning peak hours.



conventional studies, MDOT uses three-person crews working 10-h days at an estimated $325 per crew per day, including vehicles, meals, and overnight accommodations. Each person is responsible for monitoring one lane; hence, on a six-lane freeway, two crews would be needed. By calculating costs in this manner, estimated MDOT costs for surveying the freeway locations would have been $11 700. Actual costs to MDOT were approximately $6000. Consultant costs, in comparison, would have been approximately $3000 for the freeway locations. This represents an added saving available due to the use of part-time survey crews.

RESULTS

The average passenger automobile-occupancy and proportion-of-truck data including the tolerance and observed composite standard deviation are given in Table 1. In general, the tolerances and composite standard deviation are smaller than default data used by SEMCOG in applying the guide's formula.

In general, the average passenger-vehicle occupancy figures are inversely proportional to the traffic volumes on the particular roadway. That is, the higher occupancy figures were recorded on the lower-volume facilities and the lower occupancy figures were recorded on the higher-volume facilities. Higher occupancy figures were recorded in the more rural counties than in the other counties.

For all facility types monitored, the recorded average passenger-automobile occupancy was lowest during the morning peak period (7:00-9:00 a.m.) and highest during the midday period (11:00 a.m.-1:00 p.m.). For each of the three time periods monitored, the recorded average automobile occupancy was lowest on urban freeways and highest on rural arterials. In the more urban counties, the recorded average occupancy on arterials was higher than on freeways for all three time periods monitored. In the more rural counties, the recorded average automobile occupancy was higher on freeways than on arterials.

For all facility types monitored, the recorded proportion of trucks was highest during the midday period (11:00 a.m.-1:00 p.m.). For each of the three time periods monitored, the recorded proportion of trucks was highest on rural freeways and lowest on urban arterials.

The number of vehicles is shown in Figures 1 and 2 as a percentage of the total number of passenger vehicles monitored during the morning peak hours for both urban freeways and rural arterials with 24-h two-way traffic volumes of less than 35 000 vehicles per day. The percentage of total people carried during these periods by passenger vehicles is also shown. These facility types were chosen because the study indicated that the lowest average vehicle occupancy was recorded on urban freeways and the highest average vehicle occupancy was recorded on rural arterials.

The overwhelming majority of vehicles monitored during the three time periods on these facility types were one-passenger vehicles. However, it should be noted that the average occupancy figure or the percentage of one-passenger vehicles is a deceptive measure of the extent of ridesharing. As shown in Figure 1, although approximately only 12 percent of all vehicles were multioccupant vehicles, they carried over 25 percent of the people. In other words, even at an average automobile occupancy of 1.15 persons/vehicle, 25 percent of all people were sharing rides. In the case of rural arterials, Figure 2 shows that approximately 38 percent of all people were sharing rides during the morning peak periods. In effect, there was 50 percent more ridesharing on rural arterials during morning peak periods than on urban freeways at the same time. In the off-peak period on rural arterials (not shown), more than 55 percent of all people were sharing rides.

EVALUATION OF THE GUIDE

SEMCOG found the guide to be a straightforward how-to manual for collecting automobile-occupancy and vehicle-classification data. The overall sampling approach is clearly explained. The default values for standard deviation appear to be excellent based on SEMCOG's results. Any planner or traffic engineer should be able to use the guide with little or no reference to other materials.

Perhaps the largest contribution of the guide is the short-count sampling approach. This affords a dramatic saving in the survey cost at little loss in accuracy. It reduces survey costs to such an extent that SEMCOG expects to continue its automobile-occupancy and vehicle-classification survey on an annual basis. The tolerances obtained for the average automobile-occupancy estimates at the 95 percent confidence level are so small for most sampling cells that changes over time in average automobile occupancies on the order of 3-5 percent can be detected.

REFERENCES

1. Advisory Committee on Urban Transportation Data-Reporting Needs and Requirements. Proposed Urban Transportation Data-Reporting Requirements for States and Metropolitan Planning Organizations: A Report to the U.S. Department of Transportation. National Academy of Sciences, Washington, DC, 1976.
2. Peat, Marwick, Mitchell, and Co. Guide for Estimating Urban Vehicle Classification and Occupancy. Federal Highway Administration, Sept. 1978.
3. R. A. Ferlis. Field Data-Collection and Sampling Procedures for Measuring Regional Vehicle Classification and Occupancy. TRB, Transportation Research Record 701, 1979, pp. 1-6.