TTI-2-10-71-167-7

TEXAS
TRANSPORTATION
INSTITUTE

TEXAS
HIGHWAY
DEPARTMENT

COOPERATIVE
RESEARCH

ACCURACY OF TRIP END ESTIMATES FROM THE HOME INTERVIEW SURVEY

in cooperation with the
Department of Transportation
Federal Highway Administration

RESEARCH REPORT 167-7
STUDY 2-10-71-167
URBAN TRAVEL FORECASTING

ACCURACY OF TRIP END ESTIMATES
FROM THE HOME INTERVIEW SURVEY

by

Vergil G. Stover
Study Supervisor

J. D. Benson
Assistant Research Planner

and

Larry J. Ringer
Associate Research Statistician

Research Report 167-7

Urban Travel Forecasting
Research Study Number 2-10-71-167

Sponsored by the
Texas Highway Department
in Cooperation with the
U. S. Department of Transportation
Federal Highway Administration

TEXAS TRANSPORTATION INSTITUTE
Texas A&M University
College Station, Texas

August, 1973

TECHNICAL REPORT STANDARD TITLE PAGE

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.

4. Title and Subtitle	5. Report Date
ACCURACY OF TRIP END ESTIMATES FROM THE HOME INTERVIEW SURVEY	August 1973
	6. Performing Organization Code
7. Author(s)	8. Performing Organization Report No.
Vergil G. Stover, J. D. Benson, and Larry J. Ringer	Research Report 167-7
9. Performing Organization Name and Address	10. Work Unit No.
Texas Transportation Institute Texas A&M University College Station, Texas 77843	11. Contract or Grant No. Research Study 2-10-71-167
	13. Type of Report and Period Covered
12. Sponsoring Agency Name and Address	Interim - September 1972 August 1973
Texas Highway Department 11th & Brazos Austin, Texas 78701	14. Sponsoring Agency Code

15. Supplementary Notes

Research performed in cooperation with DOT, FHWA.
Research Study Title: Urban Travel Forecasting

16. Abstract

This report presents the results of a study of the accuracy of home interview survey data in estimating zonal trip ends. The study is based on 100 percent survey data collected by the Texas Highway Department in three apparently homogeneous, adjacent zones in San Antonio. The general data analysis confirmed the homogeneity of the travel characteristics of the zones. A large number of repeated random samples were drawn at various sampling rates and the results used to verify the basic assumptions and general applicability of a set of theoretical relationships between sample size and the expected error of estimation. The analysis of disaggregate zonal data was directed toward the accuracy of home interview data in estimating the population mean (i.e., the mean trips per dwelling unit) and the population variance (i.e., the variance between dwelling units in trip productivity). The results indicate that, at both the 80 and 95 percent probability levels, disturbingly large error ranges (i.e., a large variance of estimates) may be expected when using traditional sampling rates in estimating the population mean and variance for a given zone. The analysis of aggregate zonal data was directed toward the accuracy of home interview data in estimating the zonal trip ends (i.e., the number of trips produced by the zone). The results, likewise, indicate that, at both the 80 and 95 percent probability levels, disturbingly large error ranges (i.e., a large variance of estimates) may be expected when using traditional sampling rates in estimating the zonal trip ends. The results of both the disaggregate and aggregate zonal analyses provide general guidance in zonal delineation and suggest new approaches to trip generation analysis.

17. Key Words	18. Distribution Statement
Trip Generation, Origin-Destination Surveys, Urban Transportation Studies, Transportation Planning	

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	94	

Form DOT F 1700.7 (8-69)

TABLE OF CONTENTS

ABSTRACT

This report presents the results of a study of the accuracy of home interview survey data in estimating zonal trip ends. The study is based on 100 percent survey data collected by the Texas Highway Department in three apparently homogeneous, adjacent zones in San Antonio. The general data analysis confirmed the homogeneity of the travel characteristics of the zones. A large number of repeated random samples were drawn at various sampling rates and the results used to verify the basic assumptions and general applicability of a set of theoretical relationships between sample size and the expected error of estimation. The analysis of disaggregate zonal data was directed toward the accuracy of home interview data in estimating the population mean (i.e., the mean trips per dwelling unit) and the population variance (i.e., the variance between dwelling units in trip productivity). The results indicate that, at both the 80 and 95 percent probability levels, disturbingly large error ranges (i.e., a large variance of estimates) may be expected when using traditional sampling rates in estimating the population mean and variance for a given zone. The analysis of aggregate zonal data was directed toward the accuracy of home interview data in estimating the zonal trip ends (i.e., the number of trips produced by the zone). The results, likewise, indicate that, at both the 80 and 95 percent probability levels, disturbingly large error ranges (i.e., a large variance of estimates) may be expected when using traditional sampling rates in estimating the zonal trip ends. The results of both the disaggregate and aggregate zonal analyses provide general guidance in zonal delineation and suggest new approaches to trip generation analysis.

Key Words: Trip Generation, Origin-Destination Surveys, Urban Transportation Studies, Transportation Planning

SUMMARY

The purpose of this study was to investigate the accuracy of home interview origin-destination surveys in estimating zonal trip ends. To provide the data base for such a study, the Texas Highway Department conducted home interviews in 100 percent of the dwelling units in three adjacent zones located on the north-central side of San Antonio. The basis for the selection of the zones was their apparent homogeneity and nonunique characteristics. The general appearance of the area typified a lower-middle class neighborhood containing only single family dwelling units.

The analysis of disaggregate data was directed toward the accuracy of home interview data in estimating the population mean (i.e., the mean trips per dwelling unit) and the population variance (i.e., the variance between dwelling units in trip productivity). The results indicate that, at both the 80 and 95 percent probability levels, large error ranges (i.e., a large variance of estimates) may be expected when using traditional sampling rates in estimating the population mean and variance for a given zone.

The analysis of aggregate zonal data was directed toward the accuracy of home interview data in estimating the zonal trip ends (i.e., the number of trips produced by the zone). The results, likewise, indicate that, at both the 80 and 95 percent probability levels, large error ranges may be expected when using traditional sampling rates in estimating the zonal trip ends. The results of both the disaggregate and aggregate zonal analyses provide general guidance in zonal delineation and suggest new approaches to trip generation analysis.

IMPLEMENTATION STATEMENT

The results of this study demonstrate that an extremely large variance of estimates may be expected when using traditional sampling rates in a home interview survey to estimate either the mean trips per dwelling for a zone or the zonal trip ends. Interpreting these variances in terms of the expected error ranges at the 80 and 95 percent probability levels demonstrates the disturbingly large magnitude of these expected error ranges. An experienced transportation study analyst probably can estimate the number of trips within error ranges which are no larger than those expected from expanded survey data using traditional sampling rates. This suggests a new direction for home interview surveys. Instead of a general survey of the entire urban area, a smaller number of home interview surveys might be employed which are specifically directed at monitoring trends, updating urban travel parameters, and for investigating areas which exhibit unique or unusual characteristics.

Based largely on the findings of this study, the Texas Highway Department has discontinued using the traditional home interview survey in its urban transportation studies and has adopted a new "synthetic" study approach. This new synthetic study approach is currently being implemented in the Houston-Galveston Regional Transportation Study (H-GRTS). It has been estimated that the use of this approach in the H-GRTS has resulted in a net savings to THD of approximately $1,000,000, or roughly 75 percent reduction in the estimated cost of the study using the traditional approach.

INTRODUCTION

Over the years, considerable attention has been directed toward the refinement of analysis procedures, modeling techniques, and automatic data processing of urban transportation study data. At one time or another, attention was directed toward a broad range of topics, including: interview procedures and quality control, delineation of traffic assignment zones, trip generation, distribution models, mode split, etc.

The debate on the use of rates versus regression in trip generation analysis and forecasting has raged since the all-too-convenient-to-use computer programs were developed for stepwise regression; this controversy seems to have run "full circle" with the procedures now recommended being not unlike those of 15 or more years ago. Throughout this time, and underlying the bulk of the research and modeling of travel characteristics, a basic assumption has been that the sample data collected in the O-D survey provide a reliable measure of the number of trips generated in a zone. For example, in evaluating trip generation rates and regression models, it has been presumed that statistical measures of dispersion of the differences between the "predicted" and "observed" values were meaningful. In simpler terms, the presumption has generally been that the "observed" values were "correct" and that the "predicted" values deviated from them because of factors not accounted for in the regression model or cross classification scheme.

Previous Research on Sample Size Requirements

Previous research in regard to sample size has utilized data obtained from home interview surveys of various sample sizes. The studies from which these data were obtained generally involved a high degree of quality control--in some cases this quality control and citizen cooperation approached the ideal. However, since the research involved the use of sample data, certain simplifying assumptions had to be made and the results are not, therefore, directly comparable to the research to which this report is directed. Nevertheless, a brief review of selected literature is of some interest as background and comparison for the analyses performed as part of the research reported herein.

In analyzing data from the Rock Hill (South Carolina) Area Transportation Study, Stover and Roberts [(1)] reported that substantial increases in reliability in total trip end estimates resulted from increasing the sample rate from one to five percent, but that only modest increases resulted from further increases in sample size ten to fifteen percent. This research utilized data from a twenty percent sample of the 12,405 dwelling units in an urbanized area having a population of slightly over 40,000. The RHATS Study involved a high degree of quality control and enjoyed an exceptionally high level of public interest and participation.

In their "landmark" research on sample size, Sosslau and Brokke [(2)] utilized data from the 1-in-15 dwelling unit sample collected in the 1957 Phoenix O-D Survey. Their analysis, however, related sample size requirements to root mean square error in assigned link volumes, whereas this report deals with trip end estimates.

Harmelink, Harper, and Edwards [(3)] selected subsamples representing 2.5, five, and ten percent of the total dwelling units from the 12.5 percent home interview survey in Kingston, Ontario (1961, survey area population 63,000). The researchers developed and compared simple as well as multiple regression equations. They concluded that sample size did not materially affect the accuracy of the trip estimating equations. The larger sample sizes, however, resulted in a higher coefficient of correlation and/or a lower standard error, and hence, improved the precision of the estimating models.

Creighton, Hamburg [(7)] utilized sample data from several home interview surveys to focus on the accuracy of sample home interview data in estimating the proportion or percentage of trips having a given attribute. Their analysis utilized a binominal distribution approach with adjustments for the effects of cluster sampling (i.e., the random sampling of dwelling units rather than trips) to estimate sample size requirements. They concluded that higher sampling rates than previously thought necessary are needed in order to obtain reliable estimates of the proportions, or percentages, of trips having selected attributes.

Stover[(10)] investigated the accuracy of home interview data in estimating employment. Comparisons were made of employment as estimated from the home interview survey data (a 1-in-8 dwelling unit sample collected in the 1968 McAllen-Pharr Urban Transportation Study) with the reported employment obtained by field listing (interview) of each employer in the study area. It was concluded that the origin-destination survey does not yield an acceptable estimate of employment for zones, districts, or census tracts.

Various authors[4,5,6] have reported on evaluation of regression models employing aggregate totals (total trips per zone) versus disaggregate totals (trips per household). Such analyses generally have made comparisons of the estimated number of trips with the total number of observed trips. This research report does not deal with the virtues of the aggregate versus disaggregate approaches. However as discussed in the section of this report entitled, "Interpretation and Recommendations," these previous studies are of interest in the necessary redesign of data collection and trip generation analysis procedures suggested by evaluation of the 100 percent survey data.

Site Selection and Data Collection

The purpose of the study is to investigate the accuracy of sample home interview data in estimating trip ends and travel patterns. A complete census of an entire urban area would be ideal for such a study; however, cost of such a data collection effort is prohibitive. A 100 percent interview of a few selected zones could be conducted at a reasonable cost and should be sufficient to provide a useful population base whereby the accuracy of sample data in estimating the attributes of the zones might be studied. The analyses relative to zonal trip end estimation are reported herein, while the analyses relative to travel pattern estimation are reported in Research Report 167-8 entitled, "Accuracy of Travel Pattern Estimates from the Home Interview Survey."

San Antonio was selected as the site for the 100 percent data collection. Collection of the 100 percent data for selected zones in conjunction with an O-D study minimized the cost and provided compatible data for any desired comparison with the normal five percent survey.

Data Collected

The same data were collected in the three zones as were collected in the dwelling units selected in the San Antonio-Bexar County Urban Transportation Study. In each home interview, certain dwelling unit data were collected and recorded on a form similar to that shown in Figure I-1. Trip data were obtained for each resident, five years of age or older, and recorded on a form similar to that shown in

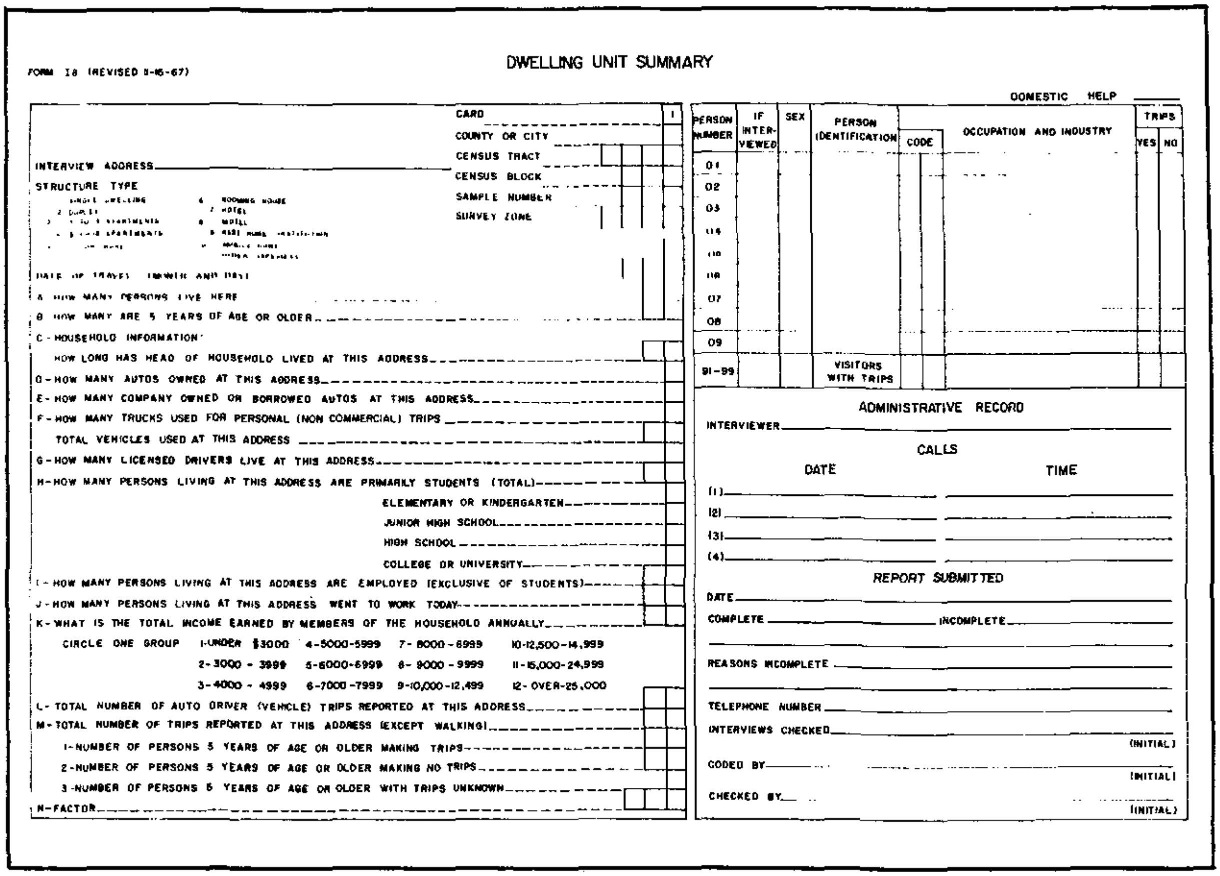

FIGURE I-1: PHOTO REDUCTION OF THE QUESTIONNAIRE USED TO COLLECT THE DWELLING UNIT DATA

Figure I-2. The data collection was performed by the Planning Survey Division of the Texas Highway Department in conjunction with the O-D study in the San Antonio urban area.

Zone Selection

Basically, three alternatives are available in the selection of the three zones for the 100 percent data collection:

- select three zones with obviously different socioeconomic characteristics.
- select three zones in different locations in the urban area with apparently the same socioeconomic characteristics.
- select three adjacent zones with apparently the same socio-economic characteristics.

The latter alternative was chosen since it offered two salient advantages:

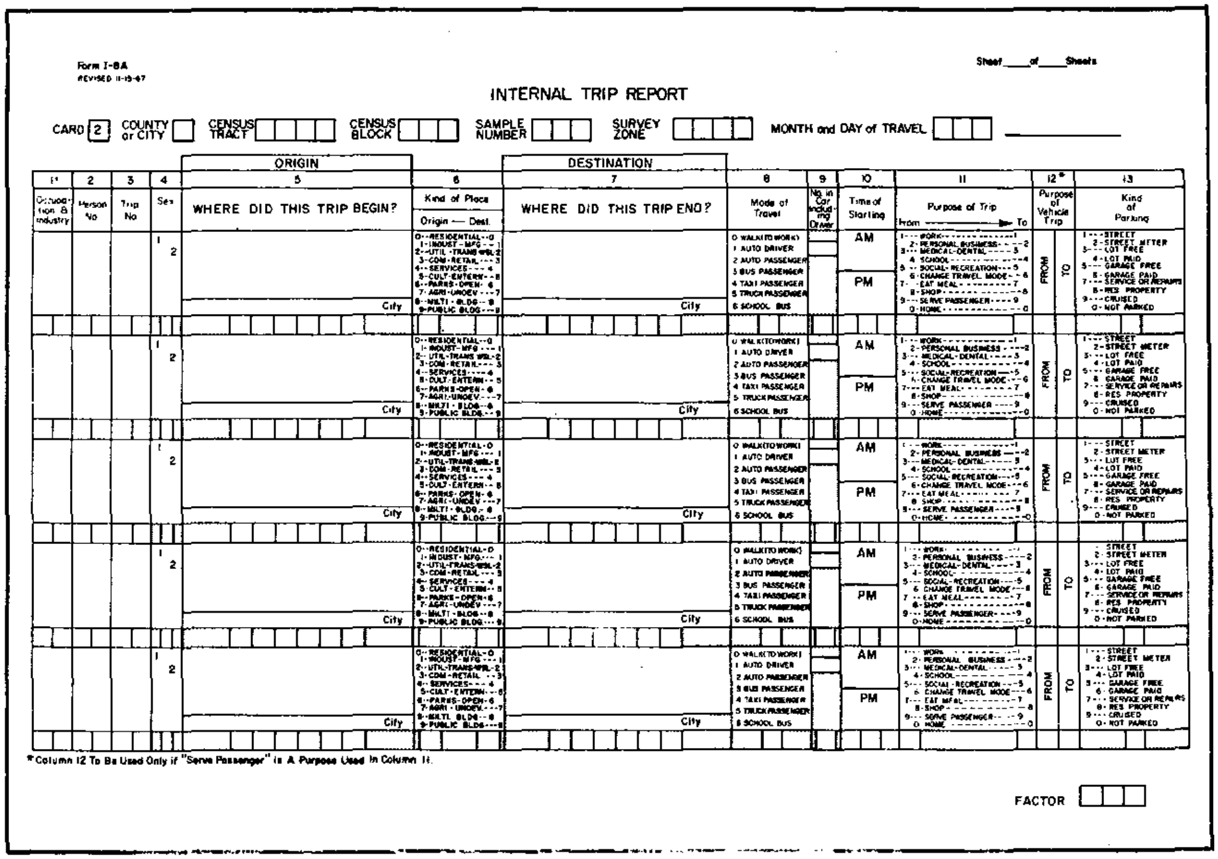
Form I-8A
INTERNAL TRIP REPORT
Sheet ___ of ___ Sheets
CARD 2 COUNTY or CITY CENSUS TRACT CENSUS BLOCK SAMPLE NUMBER SURVEY ZONE MONTH and DAY of TRAVEL
ORIGIN
DESTINATION
WHERE DID THIS TRIP BEGIN?
WHERE DID THIS TRIP END?
Kind of Place
Mode of Travel
Time of Starting
Purpose of Trip
Purpose of Vehicle Trip
Kind of Parking
AM
PM
City
*Column 12 To Be Used Only if "Serve Passenger" is A Purpose Used in Column 11.
FACTOR

FIGURE I-2: PHOTO REDUCTION OF THE QUESTIONNAIRE USED TO COLLECT THE TRIP DATA

1) Selection of three zones with apparently homogeneous socioeconomic characteristics should minimize the "within" and "between" variances. It is reasonable to expect, therefore, the variability within most other zones in urban transportation studies would generally be greater than or equal to that observed for the three selected zones. Hence, the analysis might be more easily interpreted as representing a boundary condition.

2) Adjacent zones exhibiting apparently homogeneous socioeconomic characteristics present the opportunity for convenient aggregation into a single zone for additional analyses.

The area selected for the 100 percent dwelling unit survey is located in the north central portion of San Antonio. It consists of three adjacent survey zones located between two north-south arterials (i.e., Blanco Road and West Avenue) as shown in Figures I-3 and I-4. The three zones were delineated so that the commercial developments along West Avenue and Blanco Road were not included in the selected survey zones.

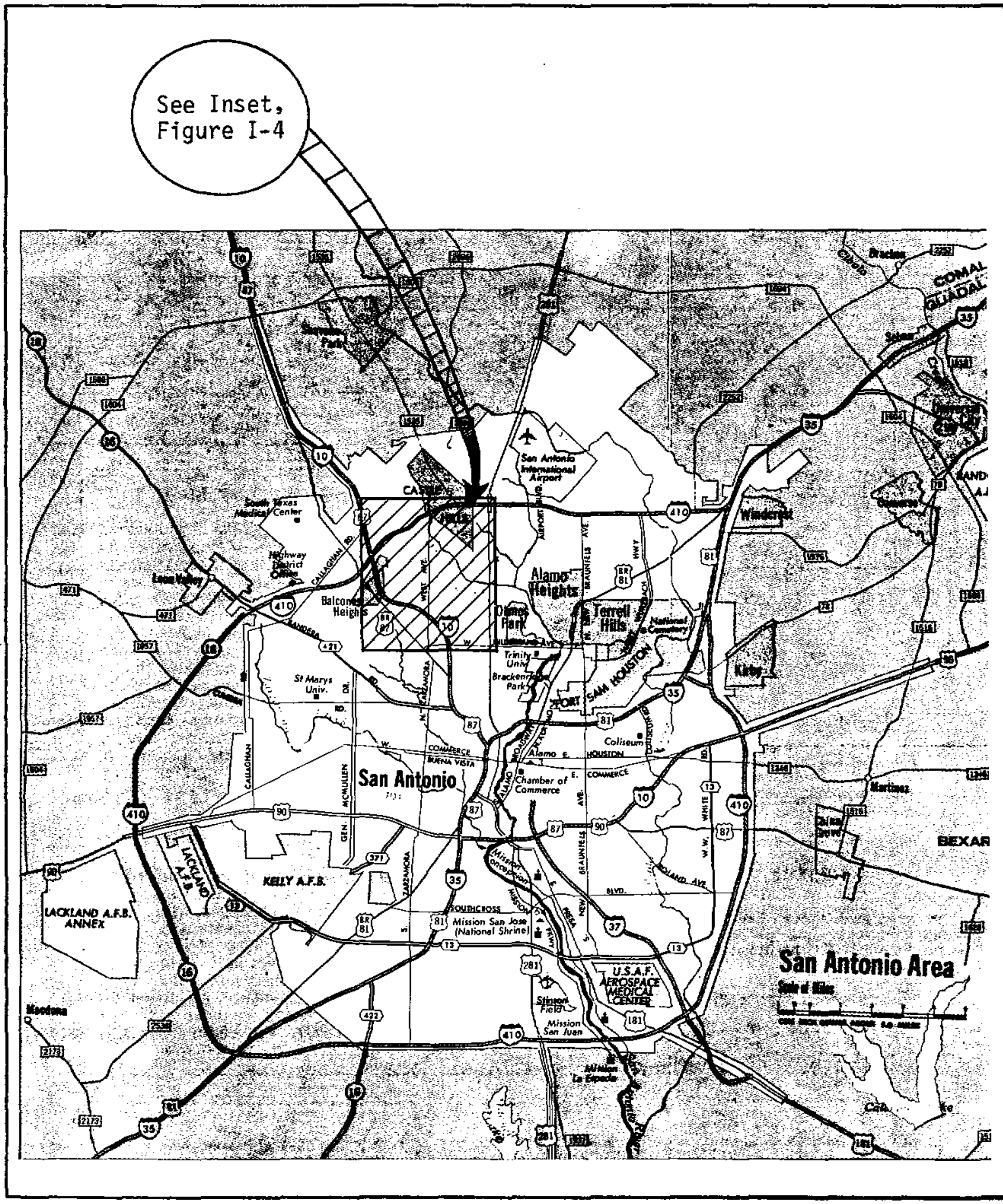

FIGURE I-3: MAP OF SAN ANTONIO AREA

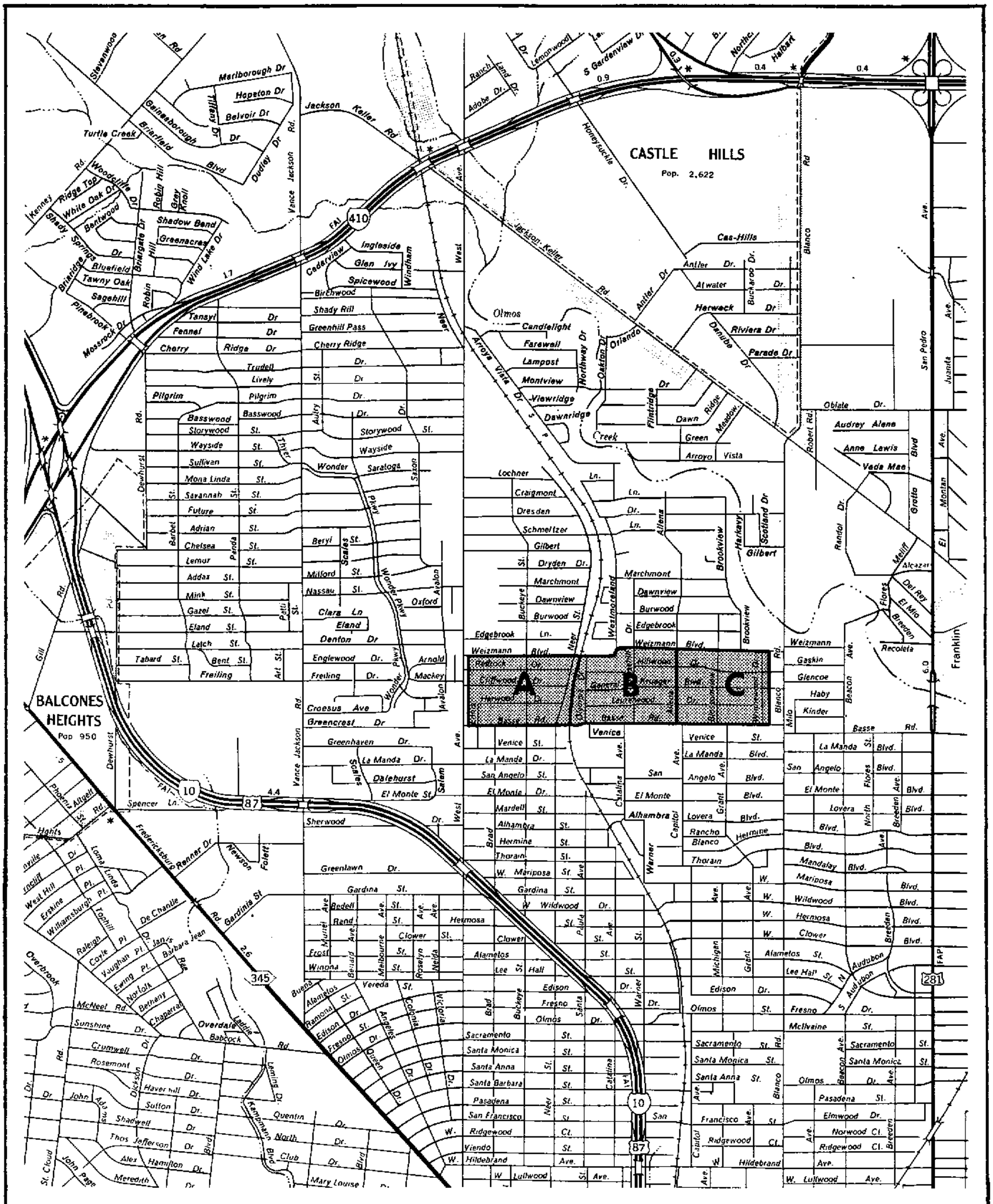

FIGURE I-4: INSET FROM FIGURE I-3 ILLUSTRATING THE LOCATION OF THE THREE SELECTED ZONES

The general appearance of the area typifies a lower-middle class neighborhood containing only single-family dwelling units. The dwellings are typical of mid- and late-1950 construction and most have single-car attached garages--some of which have been converted into living space. The vast majority of the dwellings and home sites are well maintained. Inspection of the area prior to interviewing indicated that the residents have a reasonable degree of personal mobility, as evidenced by the number of automobiles parked in driveways and at curbside. The number of boats, camper trailers, etc., suggests that the family incomes are sufficient for most to engage in a variety of recreational and other activities of their choosing. Figure I-5 illustrates the typical housing and local streets comprising the survey area. The residential density of the three survey zones is about 6,000 persons per square mile, including a small park, but excluding the commercial development located along the adjacent arterial streets.

Completeness of Data

The complete census effort was thwarted somewhat by the higher percentage of interview refusals in contrast to that normally encountered in origin-destination surveys conducted in Texas. The San Antonio vicinity had experienced considerable commercial marketing survey activity and, as a consequence, residents were less eager to respond than other cities in Texas. Furthermore, at the time of the survey, the community was plagued with a terrorist which created a tense and apprehensive atmosphere. This undoubtedly contributed to the refusal rate for the three zones reaching 14 percent which is three to five times the refusal rate previously encountered by O-D surveys in Texas.

	Zone A	Zone B	Zone C	Combined
Occupied Dwelling Units	108	201	185	494
Surveyed Dwelling Units	96	164	164	424
Refusals and No Contacts	12	37	21	70
Vacant Dwelling Units	7	4	5	16

TABLE I-1: COMPLETENESS OF SURVEY

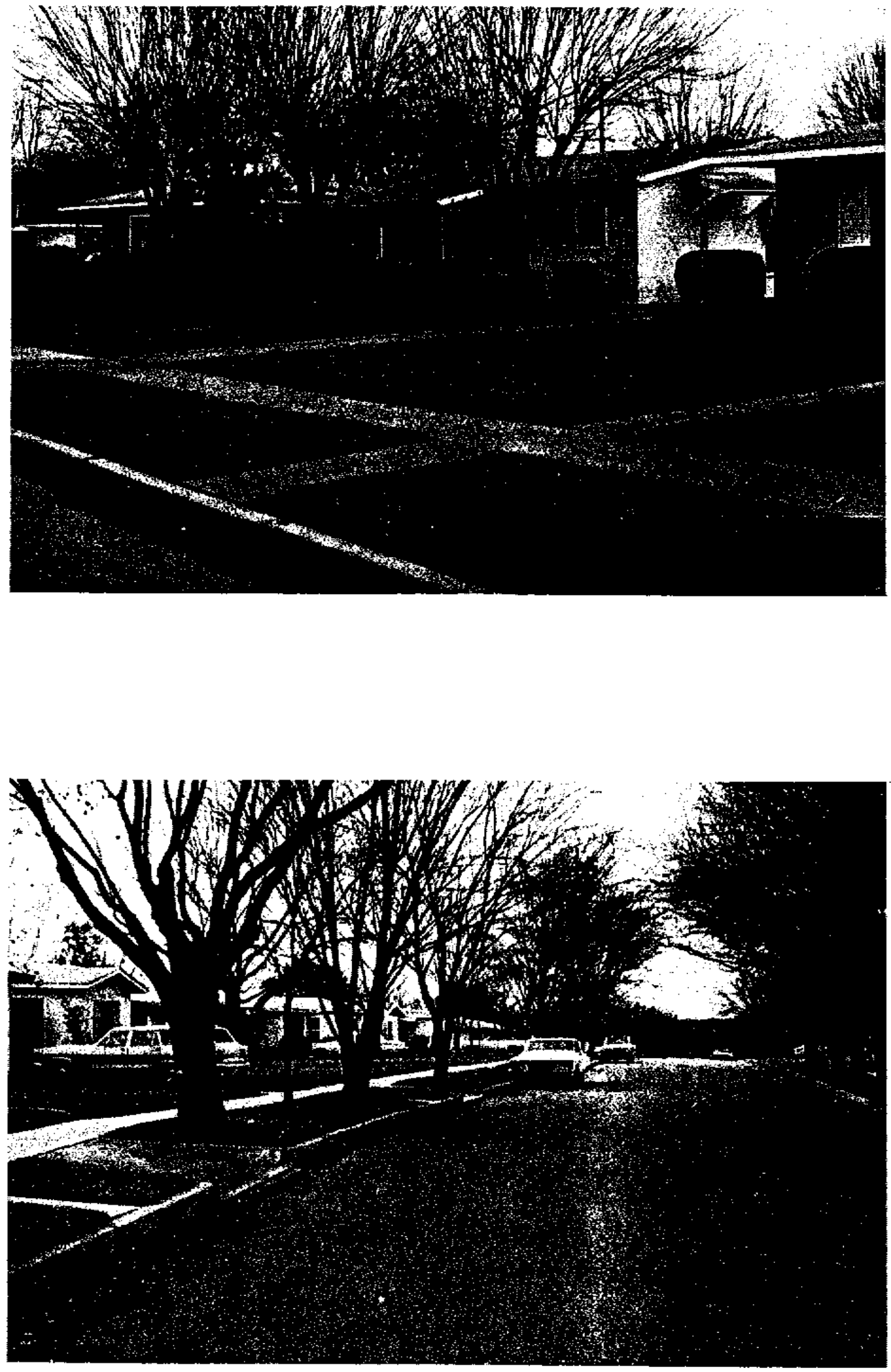

FIGURE I-5: PHOTOGRAPHS ILLUSTRATING THE TYPICAL HOUSING AND LOCAL STREETS

It is believed that the completed interviews are sufficient to establish a set of population data for the 100 percent survey area whereby the accuracy of sample data may be evaluated. Thus, for the purposes of analysis, the data collected in the 424 dwelling unit interviews are considered to be the population data (or 100 percent data) for the area. In other words, the analysis is based on the data collected from the 424 occupied dwelling units in which interviews were completed.

Data Summary

The following sections of this chapter summarize the data collected and identify the character of the area residents:

Socioeconomic Characteristics

The general dwelling unit data collected in the survey provide information on the income and employment characteristics of the area, the residents in each household, and the number of vehicles; a summary of the basic dwelling unit (except income) is given in Table I-2.

Despite the homogeneity in the socioeconomic image of the area, a fairly large variance can be observed in the dwelling unit summarized in Table I-2. However, the coefficients of variation are lowest for the fundamental household attributes, such as family size, automobile ownership, persons employed, etc. The coefficients of variation are relatively consistent among each of the three zones which substantiates that a degree of uniformity exists, as expected.

The annual household income distribution for the area is summarized in Table I-3. Residents of all but one of the 424 dwelling unit interviews responded to the income question which asked for an identification of the income range into which the household fell. The median income for the combined area, as well as Zones B and C, is in the \$6,000 to \$6,999 range. The median income for Zone A is between \$7,000 and \$7,999. The exact mean income, of course, cannot be computed from the grouped data; however, the midpoint of each range* may be used to approximate the mean. Using the midpoints of each range indicates a mean annual household income of approximately \$7,400.

---

*For the income range of "\$25,000 and above" the upper end of the range was assumed to be \$30,000. Since only two of the 424 dwelling units were in this range, it is doubtful that this assumption would have a significant effect on the computed median.

Dwelling Unit Attribute	TOTAL Zone A	TOTAL Zone B	TOTAL Zone C	TOTAL Combined	MEAN PER DWELLING UNIT Zone A	MEAN PER DWELLING UNIT Zone B	MEAN PER DWELLING UNIT Zone C	MEAN PER DWELLING UNIT Combined	STANDARD DEVIATION Zone A	STANDARD DEVIATION Zone B	STANDARD DEVIATION Zone C	STANDARD DEVIATION Combined	COEFFICIENT OF VARIATION Zone A	COEFFICIENT OF VARIATION Zone B	COEFFICIENT OF VARIATION Zone C	COEFFICIENT OF VARIATION Combined
Number of residents	324	490	463	1277	3.38	2.99	2.82	3.01	1.58	1.62	1.49	1.57	0.47	0.54	0.53	0.52
Number of residents 5 years of age or older	289	448	425	1162	3.01	2.73	2.59	2.74	1.51	1.44	1.30	1.41	0.50	0.53	0.50	0.51
Length of residence (years)	844	1462	1549	3855	8.79	8.91	9.45	9.09	7.30	6.79	6.94	6.95	0.83	0.76	0.73	0.76
Number of autos owned	152	246	234	632	1.58	1.50	1.43	1.49	0.84	0.72	0.79	0.78	0.53	0.48	0.55	0.52
Number of autos borrowed	7	10	12	29	0.07	0.06	0.07	0.07	0.26	0.24	0.26	0.25	3.58	3.94	3.57	3.69
Number of trucks available	17	20	20	57	0.18	0.12	0.12	0.13	0.44	0.38	0.36	0.39	2.46	3.12	2.98	2.88
Total number of vehicles available	176	276	266	718	1.83	1.68	1.62	1.69	0.89	0.78	0.85	0.83	0.49	0.46	0.52	0.49
Number of licensed drivers	185	316	292	793	1.93	1.93	1.78	1.87	0.93	0.84	0.84	0.86	0.48	0.44	0.47	0.46
Total number of students	96	115	107	318	1.00	0.70	0.65	0.75	1.27	1.18	1.08	1.17	1.27	1.69	1.65	1.56
Number of elementary students	44	53	55	152	0.46	0.32	0.34	0.36	0.92	0.74	0.70	0.77	2.00	2.30	2.10	2.15
Number of junior high students	19	20	17	56	0.20	0.12	0.10	0.13	0.45	0.35	0.33	0.37	2.27	2.84	3.14	2.77
Number of senior high students	23	31	27	81	0.24	0.19	0.16	0.19	0.52	0.49	0.47	0.49	2.16	2.59	2.88	2.56
Number of college students	10	11	8	29	0.10	0.07	0.05	0.07	0.31	0.27	0.22	0.26	2.95	4.09	4.43	3.83
Number of persons employed	120	209	198	527	1.25	1.27	1.21	1.24	0.73	0.77	0.73	0.74	0.58	0.60	0.60	0.60
Number of persons working on day of survey	111	187	179	477	1.16	1.14	1.09	1.13	0.76	0.78	0.73	0.76	0.66	0.69	0.67	0.67
Number of persons making trips	259	378	366	1003	2.70	2.30	2.23	2.37	1.62	1.55	1.42	1.53	0.60	0.67	0.64	0.65
Number of persons making no trips	26	67	59	152	0.27	0.41	0.36	0.36	0.61	0.66	0.64	0.64	2.24	1.62	1.77	1.79

TABLE I-2: SUMMARY OF DWELLING UNIT CHARACTERISTICS

Annual Household Income	Zone A		Zone B		Zone C		Combined	
	No.	%	No.	%	No.	%	No.	%
Under $3,000	2	2.1	7	4.3	9	5.5	18	4.2
$3,000 - 3,999	4	4.2	8	4.9	13	7.9	25	5.9
$4,000 - 4,999	7	7.3	18	11.0	16	9.8	41	9.7
$5,000 - 5,999	11	11.4	25	15.2	23	14.0	59	13.9
$6,000 - 6,999	15	15.2	25	15.2	32	19.5	72	17.0
$7,000 - 7,999	18	18.8	22	13.4	20	12.2	60	14.2
$8,000 - 8,999	13	13.5	25	15.2	15	9.2	53	12.5
$9,000 - 9,999	8	8.3	11	6.7	14	8.5	33	7.8
$10,000 - 12,499	14	14.6	12	7.3	17	10.4	43	10.1
$12,500 - 14,999	4	4.2	6	3.7	1	0.6	11	2.6
$15,000 - 24,999	0	0	4	2.5	2	1.2	6	1.4
Over $25,000	0	0	1	0.6	1	0.6	2	0.5
No Response	0	0	0	0	1	0.6	1	0.2
TOTAL	96	100.0	164	100.0	164	100.0	424	100.0

TABLE I-3: DISTRIBUTION OF DWELLING UNITS
BY ANNUAL HOUSEHOLD INCOME

About ten percent of those interviewed indicated an annual household income of less than $4,000. Based upon the home interview observations and the dwelling unit data, it is estimated that a major portion of these households are families in which the head of the household is retired. Survey responses indicated that 70 percent of these 43 dwelling units had no employed residents. Of the 43 dwelling units with incomes under $4,000, 40 percent had only one resident and 47 percent had two residents. Only three of these dwelling units reported student residents (all of which were college students).

About 15 percent of those interviewed indicated an annual household income of $10,000 or above. Of these households, 70 percent reported two or more employed residents.

Of the 424 interviews, 13 percent reported that there were no employed residents and 54 percent reported only one employed resident. Two employed residents were reported in 39 percent of the dwelling units while only four percent reported three or more employed.

The average length of residence reported by the heads of the households was approximately nine years. The average number of residents per dwelling is approximately three. However, only 22 percent of the dwelling units actually had three residents; 47 percent reported less than three residents, and 31 percent reported more than three.

Of the 1,277 residents reported living in the 424 dwelling units, 91 percent were five years of age or older. Only 57 of the dwelling units (13 percent) reported one child under five years of age; only 28 reported two or more.

Students (kindergarten through college) composed 25 percent of the total residents (318 out of 1,277). As indicated by the following figures, no students lived in a majority of the dwelling units:

Number of Students	Number of Dwelling Units	Percent of Total Dwelling Units
No students	256	60
1 student	84	20
2 students	42	10
3 or more students	42	10

Of the 56 households reporting no employed residents, none indicated having any students in high school or below; however, three of these dwelling units did report one college student resident each.

One or more autos were owned by residents in 95 percent of the dwelling units. Company-owned or borrowed vehicles were principally garaged at 29 of the dwelling units, and 12 percent of the 424 dwelling units used trucks (pickups) for personal trips. The degree of vehicular mobility is indicated by the following:

Number of Vehicles Available	Number of Dwelling Units	Percent of Total Dwelling Units
No vehicle available	22	5
1 or more	402	95
2 or more	246	58
3 or more	60	12

One or more licensed drivers resided in 95 percent of the dwelling units, and 72 percent had two or more. A total of 237 (56 percent) of the dwelling units reported having two or more licensed drivers and two or more vehicles.

Travel Characteristics

The travel characteristics of the residents of the area are not unusual. Indeed, they are typical of such residential areas in Texas.

As expected, the personal vehicle is the dominant means of transportation. Of the 4,134 person trips inventoried (excluding walk trips), their mode of travel was as follows:

Mode	Number of Trips	Percent of Total Trips
auto driver trips	2802	67.8%
auto passenger trips	1265	30.6%
bus passenger trips	47	1.1%
taxi passenger trips	0	0
truck passenger trips	5	0.1%
school bus passenger trips	15	0.4%

In addition, 26 walk-to-work trips were reported. This is not surprising, since the area is bordered on both the east and west by neighborhood shopping developments which include three supermarkets.

The inventoried trips made by residents in each zone in the combined area are summarized in Table I-4 by trip purpose. It should be noted that walk trips are excluded from this table and that the passenger trips include auto passenger, bus passenger, truck passenger, and school bus passenger trips. The mean, standard deviation, and coefficient of variation of the trips per dwelling unit are also summarized in Table I-4.

Comparison of Tables I-2 and I-4 indicates the coefficients of variation for the various trip categories are noticeably larger than those pertaining to the dwelling unit characteristics. Passenger trip productions in every instance show larger coefficients of variation than either personal-vehicle trips or all person-trips. Yet, combining passenger trips with personal-vehicle trips to obtain all person-trips, in most cases, results in a larger coefficient of variation than for personal-vehicle trips alone.

The combined area is used in discussion of the distribution of dwelling units by trips produced for the various trip purposes, in order to take advantage of the largest possible data base and to minimize the effect of a few unusual observations.

TRIP CATEGORY	TOTAL				MEAN TRIPS PER DWELLING UNIT				STANDARD DEVIATION				COEFFICIENT OF VARIATION			
	Zone A	Zone B	Zone C	Combined	Zone A	Zone B	Zone C	Combined	Zone A	Zone B	Zone C	Combined	Zone A	Zone B	Zone C	Combined
Internal, home-based work																
Automobile trip productions	173	261	247	681	1.80	1.59	1.51	1.61	1.71	1.71	1.43	1.61	0.95	1.08	0.95	1.00
Passenger trip productions	32	77	60	169	0.33	0.47	0.37	0.40	0.78	1.18	0.75	0.94	2.33	2.51	2.05	2.37
Person trip productions	205	338	307	850	2.14	2.06	1.87	2.00	1.70	2.01	1.39	1.72	0.80	0.97	0.74	0.86
Internal, home-based nonwork																
Automobile trip productions	390	503	516	1409	4.06	3.07	3.15	3.32	3.43	3.30	3.11	3.27	0.84	1.08	0.99	0.99
Passenger trip productions	265	339	357	961	2.76	2.07	2.18	2.27	3.81	3.22	3.14	3.34	1.38	1.56	1.44	1.47
Person trip productions	655	842	873	2370	6.82	5.13	5.32	5.59	6.47	5.82	5.45	5.86	0.95	1.13	1.02	1.05
Internal, nonhome-based																
Automobile trip productions	146	272	267	685	1.52	1.66	1.63	1.62	2.06	2.44	2.34	2.32	1.36	1.47	1.44	1.43
Passenger trip productions	67	64	62	193	0.70	0.39	0.38	0.46	2.38	1.10	1.47	1.61	3.40	2.83	3.89	3.53
Person trip productions	213	336	329	878	2.22	2.05	2.01	2.07	3.79	2.96	3.25	3.27	1.71	1.44	1.62	1.58
Internal, home-based																
Automobile trip productions	563	764	763	2090	5.86	4.66	4.65	4.93	3.91	3.76	3.50	3.72	0.67	0.81	0.75	0.76
Passenger trip productions	297	416	417	1130	3.09	2.54	2.54	2.67	3.96	3.65	3.20	3.56	1.28	1.44	1.26	1.33
Person trip productions	860	1180	1180	3220	8.96	7.20	7.20	7.59	6.81	6.57	5.63	6.31	0.76	0.91	0.78	0.83
All internal																
Automobile trip productions	709	1036	1030	2775	7.39	6.32	6.28	6.54	5.15	5.37	5.03	5.20	0.70	0.85	0.80	0.79
Passenger trip productions	364	480	479	1323	3.79	2.93	2.92	3.12	5.65	4.35	3.79	4.49	1.49	1.48	1.30	1.44
Person trip productions	1073	1516	1509	4098	11.18	9.24	9.20	9.67	9.47	8.58	7.53	8.43	0.85	0.93	0.82	0.87
External																
Automobile trip productions	12	7	8	27	0.13	0.04	0.05	0.06	0.46	0.26	0.29	0.33	3.72	6.00	5.92	5.14
Passenger trip productions	2	3	4	9	0.02	0.02	0.02	0.02	0.20	0.17	0.22	0.20	9.80	9.52	9.03	9.39
Person trip productions	14	10	12	36	0.15	0.06	0.07	0.08	0.50	0.33	0.45	0.42	3.44	5.36	6.16	4.96
All																
Automobile trip productions	721	1043	1038	2802	7.51	6.36	6.33	6.61	5.15	5.36	5.06	5.21	0.69	0.84	0.80	0.79
Passenger trip productions	366	483	483	1332	3.81	2.95	2.95	3.14	5.64	4.34	3.82	4.49	1.48	1.47	1.30	1.43
Person trip productions	1087	1526	1521	4134	11.32	9.30	9.27	9.75	9.46	8.56	7.57	8.43	0.84	0.92	0.82	0.86

TABLE I-4: SUMMARY OF TRIPS PER DWELLING UNIT BY PURPOSE

Table I-5 presents distribution of dwelling units in the combined area by trips produced for the following eight trip purposes:

Home-based person trips - HB(P)
Home-based work person trips - HBW(P)
Home-based nonwork person trips - HBNW(P)
Nonhome-based person trips - NHB(P)
Home-based auto-driver trips - HB(AD)
Home-based work auto-driver trips - HBW(AD)
Home-based nonwork auto-driver trips - HBNW(AD)
Nonhome-based auto-driver trips - NHB(AD)

Number of Trips Per DU	Number of DUs by Trip Purpose							
	Person Trips				Auto Driver Trips			
	HB	HBW	HBNW	NHB	HB	HBW	HBNW	NHB
0	28	102	81	183	43	151	102	188
1- 2	69	212	94	129	99	189	133	145
3- 4	87	93	75	54	112	71	76	50
5- 6	52	12	44	28	69	9	57	24
7- 8	51	2	37	11	45	3	24	6
9-10	45	1	23	9	23	0	15	6
11-12	26	1	23	2	14	1	10	2
13-14	15	1	10	4	11	0	4	3
15-16	15	0	13	1	6	0	2	0
17-18	11	0	9	0	0	0	1	0
19-20	4	0	2	1	2	0	0	0
21-22	5	0	3	0	0	0	0	0
23-24	4	0	4	0	0	0	0	0
25-26	2	0	2	1	0	0	0	0
27-28	6	0	2	1	0	0	0	0
29-30	0	0	0	0	0	0	0	0
31-32	3	0	1	0	0	0	0	0
33-34	1	0	1	0	0	0	0	0
TOTALS	424	424	424	424	424	424	424	424

TABLE I-5: DISTRIBUTION OF DWELLING UNITS IN THE COMBINED AREA BY TRIP PRODUCTIVITY FOR VARIOUS TYPES OF TRIPS

As can be seen in Table I-5, the distribution is skewed for each trip purpose. The range in trip generation is substantial; for example, 28 households produced no home-based person trips on the survey day, whereas one dwelling produced a total of 34 home-based person trips.

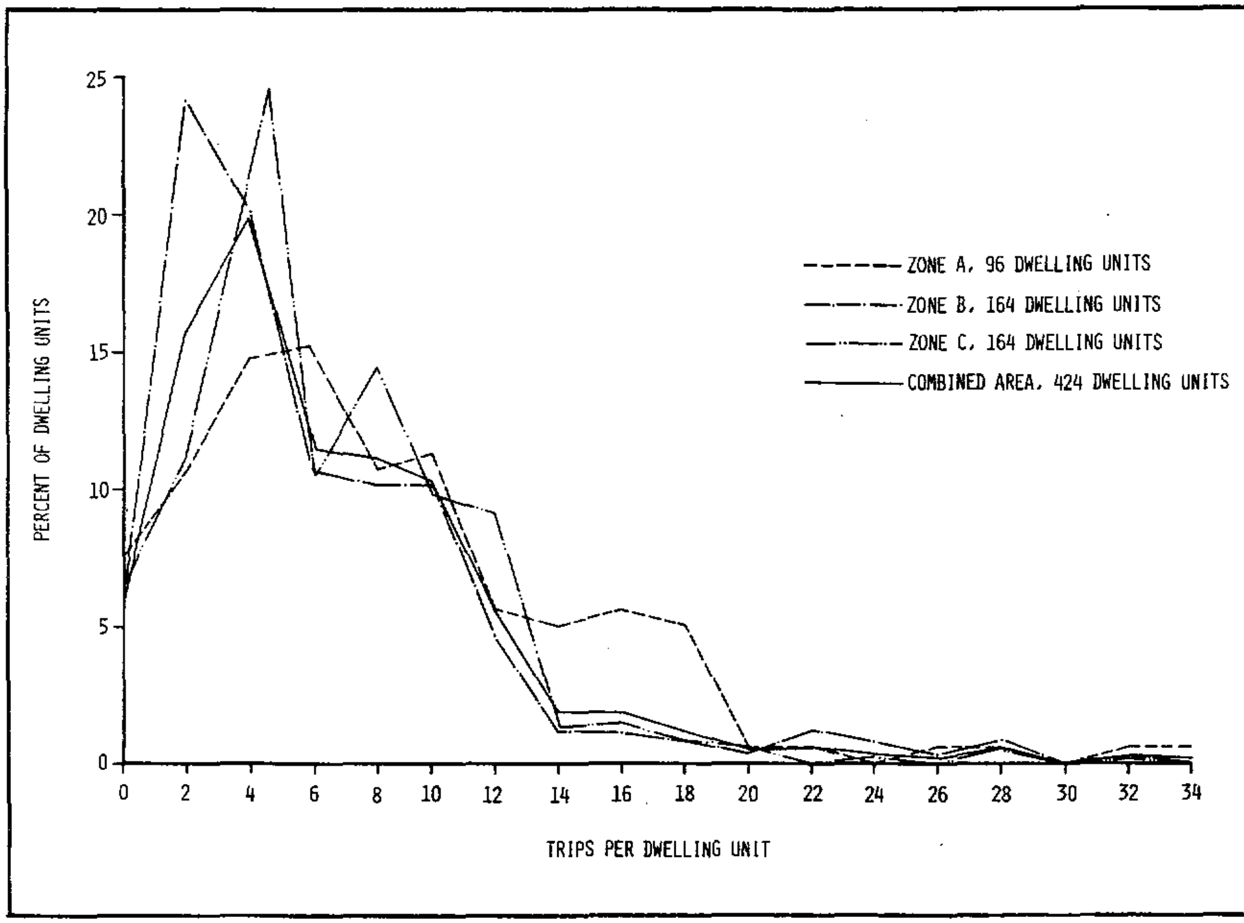

FIGURE I-6: DISTRIBUTIONS OF DWELLING UNITS BY HOME-BASED PERSON TRIP PRODUCTIONS

The home-based work trips are somewhat unusual, in that 24 percent of the dwelling units indicated no home-base work person trips. It should be recalled, however, that 15 percent of the households reported no employed residents. Approximately six percent of the dwelling units reported one or more employed residents but had no one going to work on the day of the survey. The remaining three percent of the dwelling units had trips to work but which fell in the nonhome-based category or had only walk-to-work trips which are not included in the analysis.

The distributions of dwelling units by home-based person trip productions for the individual zones, as well as the combined area, are illustrated in Figure I-6. Much of the dispersion may be attributed to variations in persons per dwelling unit.

In the 100 percent survey area, the number of residents ranged from 53 dwellings with one resident to one dwelling reporting twelve residents.

The number of trips per person (five years of age or older) are summarized in Table I-6. There were only seven persons recorded as making trips reported as unknown; these trips were excluded in computing the means and standard deviations. It is interesting to note that, with four exceptions, the coefficients of variation for the trips per person are generally larger than the trips per dwelling unit (as previously summarized in Table I-4). These exceptions are: internal home-based person trips, total internal person trips, total external person trips, and total (both internal and external) person trips.

The distributions of persons in the combined area by trip productivity for various trip categories are summarized in Table I-7. Like the distributions of dwelling units by trip productivity, these distributions are skewed. It is interesting to observe that the mode for six of the eight distributions occurs at zero while the mode for the remaining two (i.e., HB(P) and HBNW(P)) occurs at one to two trips. The distributions for the individual zones demonstrate the same general characteristics as those for the combined area. For example, the distributions of individuals by home-based person trip productions for each zone and the combined area are shown graphically in Figure I-7.

TRIP CATEGORY	TOTAL				MEAN TRIPS PER PERSON				STANDARD DEVIATION				COEFFICIENT OF VARIATION			
	Zone A	Zone B	Zone C	Combined	Zone A	Zone B	Zone C	Combined	Zone A	Zone B	Zone C	Combined	Zone A	Zone B	Zone C	Combined
Internal, Home-based work																
Automobile trip productions	173	261	247	681	0.61	0.59	0.58	0.59	0.96	0.96	0.91	0.95	1.59	1.64	1.57	1.60
Passenger trip productions	32	77	60	169	0.11	0.17	0.14	0.15	0.43	0.60	0.50	0.52	3.82	3.46	3.53	3.59
Person trip productions	205	338	307	850	0.72	0.76	0.72	0.74	0.99	1.05	0.96	1.00	1.37	1.38	1.33	1.36
Internal, home-based nonwork																
Automobile trip productions	390	503	516	1409	1.37	1.13	1.21	1.22	2.20	1.94	1.97	2.02	1.61	1.72	1.62	1.66
Passenger trip productions	265	339	357	961	0.93	0.76	0.84	0.83	1.36	1.29	1.32	1.32	1.47	1.69	1.57	1.59
Person trip productions	655	842	873	2370	2.30	1.89	2.05	2.05	2.23	2.09	2.06	2.12	0.97	1.10	1.00	1.03
Internal, nonhome-based																
Automobile trip productions	146	272	267	685	0.51	0.61	0.63	0.59	1.28	1.39	1.43	1.38	2.50	2.27	2.28	2.32
Passenger trip productions	67	64	62	193	0.24	0.14	0.15	0.17	0.77	0.55	0.86	0.73	3.26	3.86	5.92	4.39
Person trip productions	213	336	329	878	0.75	0.76	0.77	0.76	1.42	1.48	1.64	1.53	1.90	1.96	2.12	2.01
Internal, home-based																
Automobile trip productions	563	764	763	2090	1.98	1.72	1.80	1.81	2.37	2.15	2.17	2.21	1.20	1.25	1.21	1.22
Passenger trip productions	297	416	417	1130	1.04	0.93	0.98	0.98	1.41	1.39	1.40	1.40	1.35	1.49	1.43	1.43
Person trip productions	860	1180	1180	3220	3.02	2.65	2.78	2.79	2.11	2.05	2.04	2.06	0.70	0.77	0.74	0.74
All internal																
Automobile trip productions	709	1036	1030	2775	2.49	2.33	2.42	2.40	3.19	3.07	3.08	3.10	1.28	1.32	1.27	1.29
Passenger trip productions	364	480	479	1323	1.28	1.08	1.13	1.15	1.85	1.65	1.78	1.75	1.45	1.53	1.57	1.53
Person trip productions	1073	1516	1509	4098	3.76	3.41	3.55	3.55	2.99	2.96	2.98	2.98	0.79	0.87	0.84	0.84
External																
Automobile trip productions	12	7	8	27	0.04	0.02	0.02	0.02	0.27	0.16	0.18	0.20	6.52	9.94	9.59	8.57
Passenger trip productions	2	3	4	9	0.01	0.01	0.01	0.01	0.12	0.11	0.14	0.12	16.85	15.69	14.54	15.54
Person trip productions	14	10	12	36	0.05	0.02	0.03	0.03	0.30	0.19	0.23	0.23	6.07	8.38	8.00	7.49
All																
Automobile trip productions	721	1043	1038	2802	2.53	2.34	2.44	2.43	3.20	3.07	3.08	3.11	1.27	1.31	1.26	1.28
Passenger trip productions	366	483	483	1332	1.28	1.09	1.14	1.15	1.85	1.65	1.77	1.75	1.44	1.52	1.56	1.52
Person trip productions	1087	1526	1521	4134	3.81	3.43	3.58	3.58	2.98	2.95	2.98	2.97	0.78	0.86	0.83	0.83

TABLE I-6: SUMMARY OF TRIPS PER PERSON BY TYPE OF TRIP

Number of Trips Per Person	Number of Persons by Trip Purpose							
	Person Trips				Auto Driver Trips			
	HB	HBW	HBNW	NHB	HB	HBW	HBNW	NHB
0	167	702	369	777	543	792	719	864
1- 2	538	433	481	266	321	345	245	206
3- 4	312	17	196	78	188	16	113	54
5- 6	89	3	69	23	64	2	47	22
7- 8	34	0	26	2	28	0	21	1
9-10	10	0	9	5	9	0	8	5
11-12	5	0	5	3	2	0	2	3
13-14	0	0	0	0	0	0	0	0
15-16	0	0	0	1	0	0	0	0
SUBTOTALS	1155	1155	1155	1155	1155	1155	1155	1155
Persons under 5	115	115	115	115	115	115	115	115
Persons with trips unknown	7	7	7	7	7	7	7	7
TOTALS	1277	1277	1277	1277	1277	1277	1277	1277

TABLE I-7: DISTRIBUTION OF PERSONS IN THE COMBINED AREA BY TRIP PRODUCTIVITY FOR VARIOUS TRIP PURPOSES

Auto occupancy provides still further insight into the travel characteristics of the study area. It should be noted that in recording the auto occupancy, persons under five years of age were included. The auto occupancy distributions of auto-driver trips in the combined area for various trip purposes are summarized in Table I-8. As can be seen from this table, a majority of the vehicular trips for each trip purpose had only one occupant (i.e., the driver). Over 96 percent of the home-based work auto driver trips had only one occupant.

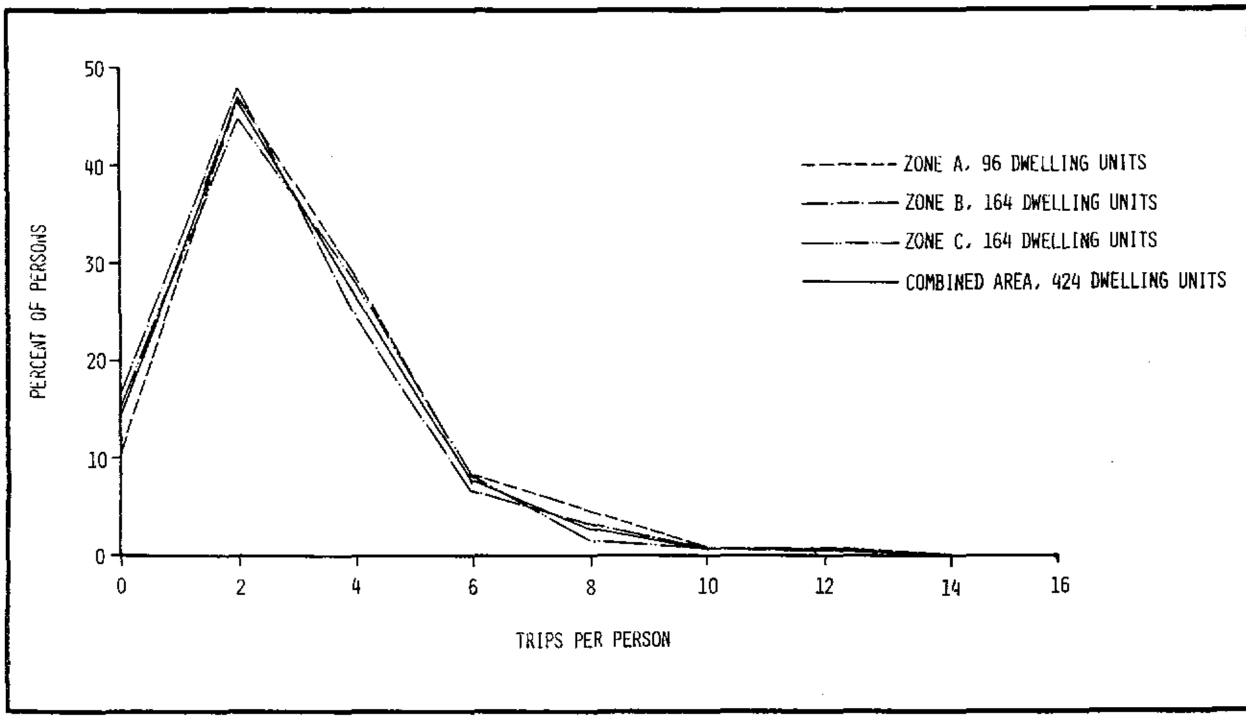

FIGURE I-7: DISTRIBUTION OF PERSONS BY TRIP PRODUCTIVITY

NUMBER OF OCCUPANTS	TRIP PURPOSE							
	HB		HBW		HBNW		NHB	
	NUMBER	PERCENT	NUMBER	PERCENT	NUMBER	PERCENT	NUMBER	PERCENT
1	1435	68.66	656	96.33	799	55.29	423	61.75
2	423	20.24	20	2.93	403	28.60	164	23.94
3	134	6.41	3	0.44	131	9.30	44	6.42
4	59	2.82	0	0.00	59	4.19	27	3.94
5	22	1.05	0	0.00	22	1.56	14	2.05
6	7	0.34	1	0.15	6	0.43	5	0.73
7	4	0.19	0	0.00	4	0.28	3	0.44
8	4	0.19	0	0.00	4	0.28	1	0.15
9	2	0.10	1	0.15	1	0.07	4	0.58
TOTALS	2090	100.00	681	100.00	1409	100.00	685	100.00
Avg. Trips/ Occupant	1.51		1.06		1.72		1.69	

TABLE I-8: DISTRIBUTION OF AUTO DRIVER TRIPS BY AUTO OCCUPANCY FOR VARIOUS TRIP PURPOSES

TECHNICAL APPROACH

The purpose of the study was to evaluate the accuracy of home interview sample survey data in estimating travel for urban transportation studies. External trips were excluded since they are normally estimated from an external survey. The analysis included two types of trips: person trips and auto-driver trips. Each of these types of trips were subdivided into the three trip purposes generally used in transportation studies in Texas; these are: home-based work trips, home-based non-work trips, and nonhome-based trips. For purposes of analysis, all home-based trips (i.e., home-based work plus home-based nonwork) were also used. In essence, the analysis was performed on four trip purposes for each of the two types of trips, or a total of eight trip categories, as follows:

Trip Category	Abbreviation
Home-based person trips	HB(P)
Home-based auto driver trips	HB(AD)
Home-based work person trips	HBW(P)
Home-based work auto driver trips	HBW(AD)
Home-based nonwork person trips	HBNW(P)
Home-based nonwork auto driver trips	HBNW(AD)
Nonhome-based person trips	NHB(P)
Nonhome-based auto driver trips	NHB(AD)

Theoretical Constructs

Sampling theory provides the basic framework for much of the subsequent analysis and inference. While population data provide opportunities for substantial empirical observations, their conformance to statistical theory both amplifies and generalizes the observations, thereby enhancing their value. The following describes the sampling theory utilized in subsequent analyses.

Sample Size and Expected Error

In a home interview survey, the unit of observation is, of course, the dwelling unit. The zone, therefore, represents a finite population having a finite variance σ^2 and mean μ for each trip category. Given that (for a given trip category) the distribution of the estimated mean trips per dwelling unit from repeated random samples is normally distributed,* the theoretical calculation of the sample size required for a desired level of accuracy is given by [11]:

$$n = \frac{\left(\frac{t\sigma}{d_1}\right)^2}{1 + \frac{1}{N}\left(\frac{t\sigma}{d_1}\right)} \qquad \text{(Equation 1)}$$

or

$$n = \frac{\left(\frac{Nt\sigma}{d_2}\right)^2}{1 + \frac{1}{N}\left(\frac{Nt\sigma}{d_2}\right)^2} \qquad \text{(Equation 2)}$$

n = the required sample size.

N = the number of dwelling units in the zone sampled.

t = the value from the normal table which corresponds to the desired level of confidence (e.g., 1.96 for 95 percent confidence).

σ^2 = the variance of distribution of dwelling units by trip productivity for the given trip category.

d_1 = the desired level of accuracy (or half the magnitude of the expected error range) in estimating the mean trips per dwelling unit for the zone. For example, at the 95 percent confidence level, n would be the sample size for which there is a 95 percent probability of drawing a random sample which would yield an estimate of the mean in the range ($\mu \pm d_1$).

d_2 = the desired level of accuracy (or half the magnitude of the expected error range) in estimating the total number of trips for the zone. For example, at the 95 percent confidence level, n would be the sample size for which there is a 95 percent probability of drawing a random sample which would yield an estimate of the total trips in the range ($N\mu \pm d_2$).

---

* The normality of the distribution of the estimated means will be subsequently demonstrated.

For purposes of analysis, it is convenient to restate equations 1 and 2 as follows:

$$d_1 = \frac{t\sigma}{\sqrt{\frac{Nn}{N - n}}} = \frac{t\sigma}{\sqrt{N}} \sqrt{\frac{1-P}{P}} \qquad \text{(Equation 3)}$$

$$d_2 = \frac{Nt\sigma}{\sqrt{\frac{Nn}{N - n}}} = (t\sigma \sqrt{N}) \sqrt{\frac{1-P}{P}} \qquad \text{(Equation 4)}$$

Note that d_2 is simply equal to Nd_1 and P is equal to the percent sample expressed as a fraction.

Variance of Estimate

As previously noted, the above relationships assume that the distribution of the estimated means ($\bar{X}$) from repeated random samples will be normally distributed. This, of course, suggests that the distribution of the estimates of the total trips for the zones (Y, where $Y = N\bar{X}$) from repeated random samples will also be normally distributed. From equations 3 and 4 it can be seen that the estimated variances for these distributions are given by:

$$S_{\bar{X}}^2 = \frac{\sigma^2}{\left(\frac{Nn}{N - n}\right)} \qquad \text{(Equation 5)}$$

and

$$S_Y^2 = \frac{N^2\sigma^2}{\left(\frac{Nn}{N - n}\right)} \qquad \text{(Equation 6)}$$

where, $S_{\bar{X}}^2$ = the variance of estimates relative to the mean trips per dwelling unit ($\bar{X}$).

S_Y^2 = the variance of estimates relative to the total trips for the zone (Y).

Note that S_Y^2 is simply equal to $N^2 S_{\bar{X}}^2$. In essence, equations 3 and 4 simplify to:

$$d_1 = tS_{\bar{X}} \quad \text{and} \quad d_2 = tS_Y$$

Estimation of Population Variance

It is important to note that the preceding equations assume that the population variance (σ^2) is known. In practice, however, the population variance is seldom known. While there are a number of ways of estimating population variances for sample size determinations, the method often employed in urban transportation studies involves drawing an initial sample from which an estimate of the population variance is obtained and the sample size requirements computed. Unfortunately, little is known relative to the distribution of estimates of the population variance from repeated random samples. It may be expected, however, that the distribution would approach a normal distribution as the sample size increases.

The 100 percent data provide unique opportunities: to draw a large number of random samples of a given sample size; to test the normality of the distribution of estimates of the population variance, and to estimate the variance of this distribution. This empirical information provides substantial insight into the accuracy of sample data in estimating the population variance for a zone and provides a basis for various statistical inferences relative to this accuracy.

Sampling Rate and Percent Error

For purposes of this analysis, it is more convenient to express the relationship in Equations 3 and 4 in terms of sampling rate (i.e., percent sample) and percent error range rather than sample size and error range. The percent error relative to the mean is, of course, equal to the percent error relative to the total trips, as illustrated by the following:

$$E = \frac{d_1}{\mu} = \frac{d_2}{N\mu} = \text{percent error expressed as a fraction}$$

The sampling rate may be defined as:

$$P = \frac{n}{N} = \text{percent sample expressed as a fraction}$$

Therefore, substituting into either equation 3 or 4, the following may be obtained:

$$E = \left(\frac{t\sigma}{\mu}\right)\left(\frac{1}{\sqrt{N}}\right)\sqrt{\frac{1 - P}{P}} \qquad \text{(Equation 7)}$$

Since the coefficient of variation is equal to the standard deviation divided by the mean, the relationship between the sampling rate and the expected percent error may be simply stated as follows:

$$E = \frac{tc}{\sqrt{N}} \sqrt{\frac{1 - P}{P}} \qquad \text{(Equation 8)}$$

Where, E = expected percent error range expressed as a fraction (i.e., for ± 25 percent error E = 0.25)

P = percent sample expressed as a fraction (i.e., for a 5 percent sample, P = 0.05)

C = coefficient of variation

N = number of occupied dwelling units

t = the value from the normal table which corresponds to the desired level of confidence

Since the values of C and N are known for the eight trip categories in each of the three zones and the combined area, the above formula may be used to quantify the relationship between the sampling rate and the expected percent error for the zones being studied.

Analysis Approach

The first step in the study of the 100 percent data was the development of a thorough description of the area based on the data. The description of the area (presented in the Introduction) provides insight into the general character of the area, both in terms of its socioeconomic characteristics and its travel characteristics.

For convenience, the following analyses of the 100 percent data have been grouped into three areas:

- general statistical analyses
- disaggregate analyses
- aggregate analyses

The designation of disaggregate and aggregate analysis areas provides a convenient means of grouping analyses results which are generally pertinent to either disaggregate or aggregate trip generation analysis.

GENERAL STATISTICAL ANALYSES

The general statistical analyses consist of those analyses directed toward the confirmation of certain conditions or assumptions which provide a foundation for subsequent analyses. These analyses were specifically directed toward:

- testing the homogeneity of the zones with regard to their trip generation characteristics.
- testing the assumption regarding normality of the distribution of the estimates of the mean trips per dwelling unit from repeated random samples.
- empirical validation of the applicability of statistical formula for sample size requirements.
- testing the normality of the distribution of the estimates of the population variance from the repeated random samples.

The following describes the results of these analyses.

Homogeneity of Zones

Although the three zones appeared homogeneous, their basic home-based trip generation characteristics were subjected to statistical tests to substantiate their homogeneity. The hypothesis, that the mean number of trips per dwelling unit was the same for each of the three zones, was tested for various trip purposes using a one-way analysis of variance at a confidence level of 0.95. In each case, the test indicated that there were no statistically significant differences between the means for the individual zones.

The hypothesis,that the variance of the number of trips per dwelling unit was the same for each of the zones, was tested for various trip purposes using Bartlett's test at the 0.95 confidence level. In all but one case (i.e., the home-based work auto-driver trips), the test indicated that there were no statistically significant differences between the variances for the three zones. In the case of the home-based work auto-driver trips, the test results indicated that the hypothesis of equal variances could have been accepted at the 90 percent confidence level. It is important to realize that Bartlett's test is sensitive to the assumption of normality of the parent frequency distribution and tends to give too many rejections when applied to skewed distributions. Since the parent frequency distributions for each of the eight trip categories are highly skewed to the right, it is not surprising to have encountered one rejection at the 0.95 confidence level. The hypothesis would not be rejected for any of the eight trip categories at a confidence level of 0.95.

Based on these tests, it was concluded that the three zones were homogeneous and that they may be combined for the purposes of analysis to observe the effect of zone size.

Distribution of Estimated Means from Sampling

Application of statistical formulas for sample size requirements assumes that the estimates of the means are normally distributed.(11) Therefore, standard statistical tests were applied to test the hypothesis of the normality of the estimates of the mean number of trips at the several sampling levels.

The Central Limit Theorem may be stated as follows:

> If a population has a finite variance of σ^2 and mean, then the distribution of the sample mean approaches the normal distribution with variance σ^2/η and mean η as the sample size η increases. (12)

The theorem, however, only asserts that the distribution will approach a normal distribution. It does not suggest that for any given sample size the distribution will adequately approximate a normal distribution such that the normality assumption for the formulas is reasonably satisfied. The availability of the 100 percent data provides a basis for the empirical verification of the normality of the distribution of the estimated means at various sampling rates.

Test for Normality

Sets of random samples at various sampling rates were drawn from each zone and the combined area; 1,000 random samples were drawn at the sampling rates of 2, 5, 6.7, and 10 percent; and, for the sampling rates of 12.5, 15, 20, 30, 40, 50, 60, 70, 80, and 90 percent, two hundred random samples were drawn. The distribution of the estimates of the mean number of trips per dwelling unit for each of the eight trip categories (HB, HBW, HBNW, NHB for both person trips and auto-driver trips) was obtained at each sampling rate. The Kolmogorov-Smirnov and Chi-Square Goodness-of-Fit tests (at a level of significance of $\alpha = 0.05$) were used to test whether the distributions were significantly different from a normal distribution.

The following points should be borne in mind in interpreting the results of these tests:

- The sample size required for the distribution of the estimates of the means to be approximately normally distributed is in part dependent upon the skewness of the parent frequency distribution. Therefore, for a given population, a larger sample size will be required for a trip purpose which has a parent frequency distribution which is more skewed than the parent frequency distribution for some other trip purpose. For example, since the parent frequency distribution for home-based work person trips is substantially more skewed than the home-based person trips in the combined area (see Table I-5), it may be expected that, for the distribution of the estimated means for home-based work person trips to be approximately normally distributed, a larger sample size may be required.
- Since the Central Limit Theorem is an asymptomatic property, the number of observations in the sample, not the sampling rate, is the determining factor. Hence, a higher sampling rate may be required for a small population than for a large population when both have about the same degree of skewness.

The required sampling rate is, therefore, dependent upon both the zone size and the skewness of the parent frequency distribution (i.e., the distribution of the dwelling units in that zone by their trip productivity for the given trip purpose).

For the combined area, the tests indicated that sampling rates of from two to five percent were sufficient for the distributions of the estimated means to approximate normal distributions. For the six home-based trip categories for the individual zones, the tests indicated that sampling rates of from five to ten percent were sufficient for the assumption of normality. Sampling rates of from 6.7 to 30 percent were found necessary to accept the normality hypothesis for the nonhome-based trip categories.

Variations in the sample size requirements for the different trip categories for each zone, and the combined area, are largely attributable to the variations in the skewness of the parent frequency distributions.

The variation in required sampling rates between the combined area and the individual zones is only partially attributable to variations in the skewness of the parent frequency distributions, since only relatively minor variations in skewness were observed. The variation between the individual zones and the combined area, therefore, is attributable to the use of various levels of sampling rates rather than sample sizes with different size populations. For example, a ten percent sample in Zone A consists of roughly ten dwelling units, whereas a ten percent sample in the combined area consists of 42 dwelling units.

Empirical Validation of Sampling Formulas

The Kolmogorov-Smirnov and Chi-Square tests indicate that the assumption of normal distribution of trips per person and trips per dwelling is not accepted at the smaller sampling levels; therefore, the question remains as to whether these differences are, in reality, sufficient to limit the practical application of the formulas at these rates. The random samples drawn for the normality tests were subsequently used to empirically demonstrate the practical applicability of the formulas.

Using Equation 3 and the population standard deviations, the expected 95 percent probability limits were computed for various trip purposes and zones for two and five percent samples. The number of random samples having an error was within the expected error range and was counted. The results of these tests are summarized in Table III-1; 92.9 percent to 97.2 percent of the samples were within the expected error ranges, for 95 percent confidence. Since the level of confidence assumes an infinite number of samples, it is felt that the range of 92.9 to 97.2 is reasonable for a thousand samples.

Similar tests were performed at the 80 percent and 95 percent levels of confidence for the entire range of sampling rates using the home-based person trips for the combined area. The results of these tests are summarized in Table III-2; 92 percent to 96 percent were within the expected error ranges, for 95 percent confidence. At 80 percent confidence level, 76 percent to 83 percent were within the expected error ranges. It is, therefore, concluded that the distributions are not - in a practical sense - sufficiently different from normal to limit the practical application of the statistical formulas at these lower sampling rates.

Trip Categories	Percent of Samples Within Expected Error Range: 2 Percent Samples, Zone A	Zone B	Zone C	5 Percent Samples, Zone A	Zone B	Zone C
Home-based Person Trips	95.2	94.4	96.2	95.0	96.6	94.2
Home-based Auto Driver Trips	95.4	95.2	94.2	95.4	95.6	94.2
Home-based Work Person Trips	95.4	94.6	96.6	93.0	94.8	94.8
Home-based Work Auto Driver Trips	96.3	96.2	97.2	95.2	95.0	93.6
Home-based Nonwork Person Trips	94.2	94.8	95.8	95.1	96.3	95.0
Home-based Nonwork Auto Driver Trips	94.7	95.4	95.4	95.6	96.0	95.4
Nonhome-based Person Trips	94.8	94.8	95.6	92.9	96.2	97.2
Nonhome-based Auto Driver Trips	94.3	95.6	94.0	94.8	95.3	95.8

TABLE III-1: PERCENTAGE OF SAMPLES WITHIN EXPECTED ERROR RANGES FOR 95 PERCENT PROBABILITY LIMITS

Percent Sampling Rate	Percent of Samples Within Expected Error Range: 80 Percent Confidence	95 Percent Confidence
2	79.0	96.0
5	77.0	95.4
6	78.0	95.7
10	79.2	94.4
12	75.5	95.5
15	83.0	96.5
20	81.0	96.5
30	82.5	92.0
40	78.0	95.0
50	79.5	93.0
60	81.5	95.5
70	81.0	95.0
80	83.5	95.0
90	76.0	95.0

TABLE III-2: PERCENTAGE OF SAMPLES WITHIN EXPECTED ERROR RANGES AT 80 AND 95 PERCENT PROBABILITY LEVEL: HOME-BASED PERSON TRIPS FOR COMBINED AREA

Distribution of Estimated Population Variance

It may be expected that the distribution of estimates of the population variance in repeated sampling will approach a normal distribution as the sample size is increased. To empirically verify this, 1,000 samples were drawn from the combined area, using sampling rates of 2.5, five, and ten percent. The distribution of estimates of the population variance for each of the eight trip categories was obtained at each sampling rate. The Kolmogorov-Smirnov test was used to determine whether the distributions were significantly different from a normal distribution. The results of these tests indicated that the hypothesis of normality of the distributions of estimates at the 12.5 percent sampling rate could be rejected at a 99 percent confidence level. At the five and ten percent sampling rates the hypothesis could not be rejected at the 80 percent confidence level. Hence, it is concluded that the distributions of estimates of the population variance for each of the eight trip categories will closely approximate a normal distribution for sample sizes of 21 and above (i.e., sampling rates of five percent and above for the combined area).

ANALYSIS OF DISAGGREGATE DATA

The analysis of disaggregate data focuses on the estimation of certain travel characteristics (such as the mean trips per dwelling unit and the variance of trips per dwelling unit) which describe the distribution of dwelling units by trip productivity (i.e., the parent frequency distribution from which samples are drawn). Specifically, the analysis focuses on:

- accuracy in the estimation of the mean trips per dwelling unit from sample data.
- the variance of estimates of significant zonal travel characteristics from sampling.

The results of this analysis provide substantial insight into the accuracy of home-interview data used in disaggregate trip generation analysis.

Accuracy in Estimation of the Mean from Sample Data

As will be recalled, the general statistical analysis demonstrated the practical applicability of the theoretical relationships between sample size and expected error and between sampling rate and expected percent error. The unique position of possessing the population data allows the direct application of these formulas to quantitatively study these relationships for the three zones and the combined area. For convenience, the formulas expressing the relationship between sample rates and expected percent error ranges (i.e., Equations 7 and 9 from section entitled, "Theoretical Constructs") were extensively used in this analysis. The use of these formulas offers two salient advantages:

- Practitioners are generally more accustomed to thinking in terms of percent samples and percent error.
- The quantitative results are applicable to the estimation of the mean trip per dwelling unit as well as to the estimation of the total trips for the zone.

The formulas expressing the relationship between sample sizes and expected error ranges (i.e., Equations 1 and 3 from the section entitled "Theoretical Constructs") were also used extensively to provide added perspective in the accuracy of sample data. The use of these formulas emphasizes the severity of the percent errors observed.

Sampling Rate Versus Expected Percent Error Range

For each of the eight trip categories and each of the three zones and the combined area, the required sampling rates at the 95 percent probability level were computed for error ranges from ±5 percent to ±95 percent. Similar computations were performed for the required sampling rates at the 80 percent probability level. The results of these computations are illustrated graphically in Figures IV-1 through IV-4.

A number of general observations become apparent from careful inspection and comparison of the relationships depicted graphically in these figures between sampling rates and their associated error ranges for the various trip categories and zones. These include:

- Rather large sampling rates are required to estimate the mean trips per dwelling unit for a zone (or the combined area) within reasonable ranges of error even at the 80 percent probability level.
- Increasing the sampling rate decreases the error ranges at a decreasing rate. For example, increasing the sampling rate from, say, five to ten percent results in a larger reduction in the error range than increasing the sampling rate from, say, 55 to 60 percent.
- As expected, increasing the zone size within a homogeneous area tends to reduce the error ranges associated with any sampling rate.
- The error ranges for the estimation of auto driver trips are generally less than those for the estimation of person trips. (Home-based nonwork trips in Zone A and nonhome-based trips in Zone B produced the only exceptions.)
- When estimating either person or auto driver trips, the four trip purposes would generally be ranked based on the percent error (smallest to largest) as follows:

  home-based trips
  home-based work trips
  home-based nonwork trips
  nonhome-based trips

It must be remembered that the last two observations relate to the percent error rather than the absolute error. Although the expected percent error in estimating the mean number of home-based work trips per dwelling unit is larger than for all home-based trips, the magnitude of the error ranges is smaller for home-based work trips than for all home-based trips.

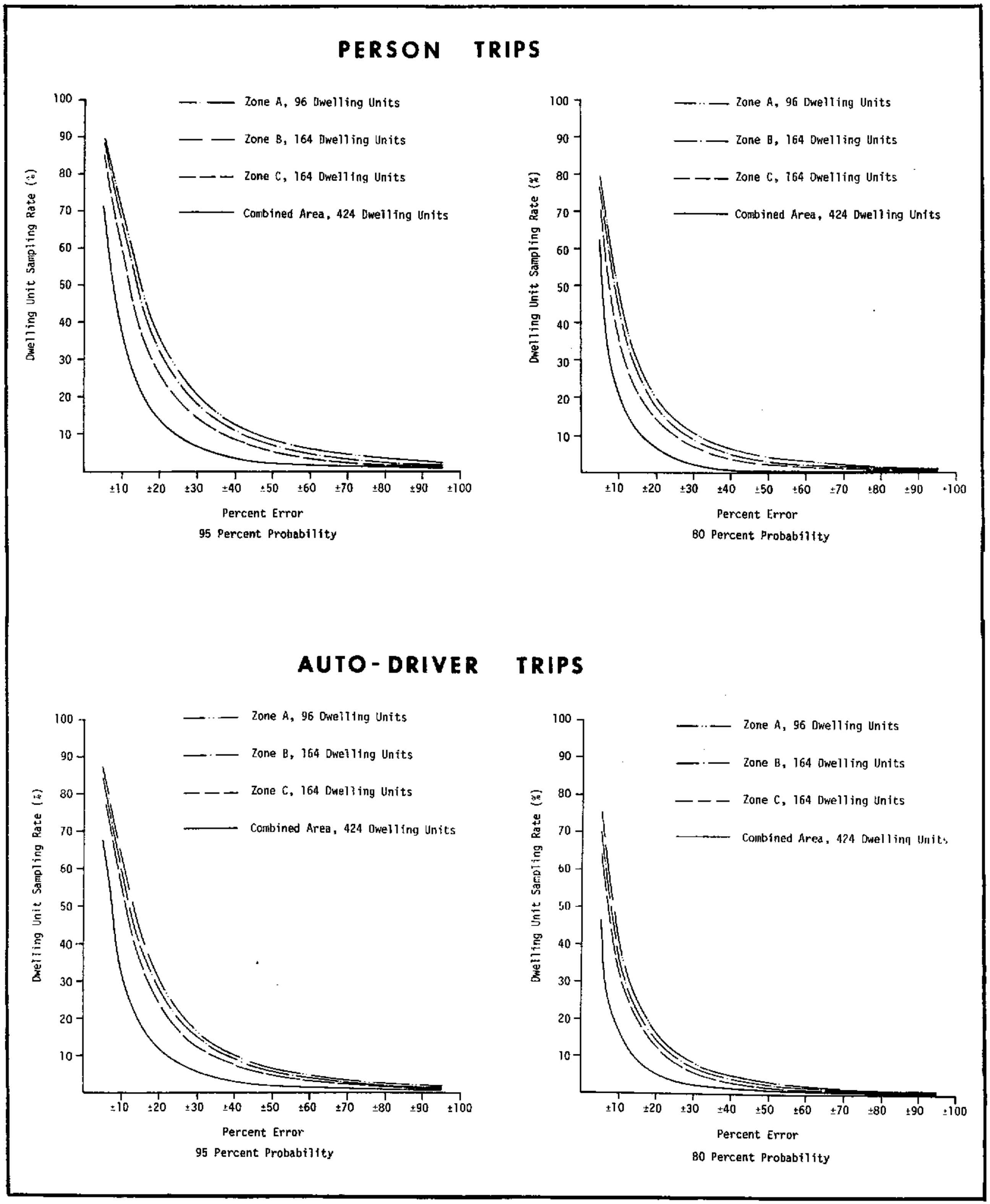

FIGURE IV-1: MARGIN OF ERROR VERSUS SAMPLING RATE FOR HOME-BASED TRIPS

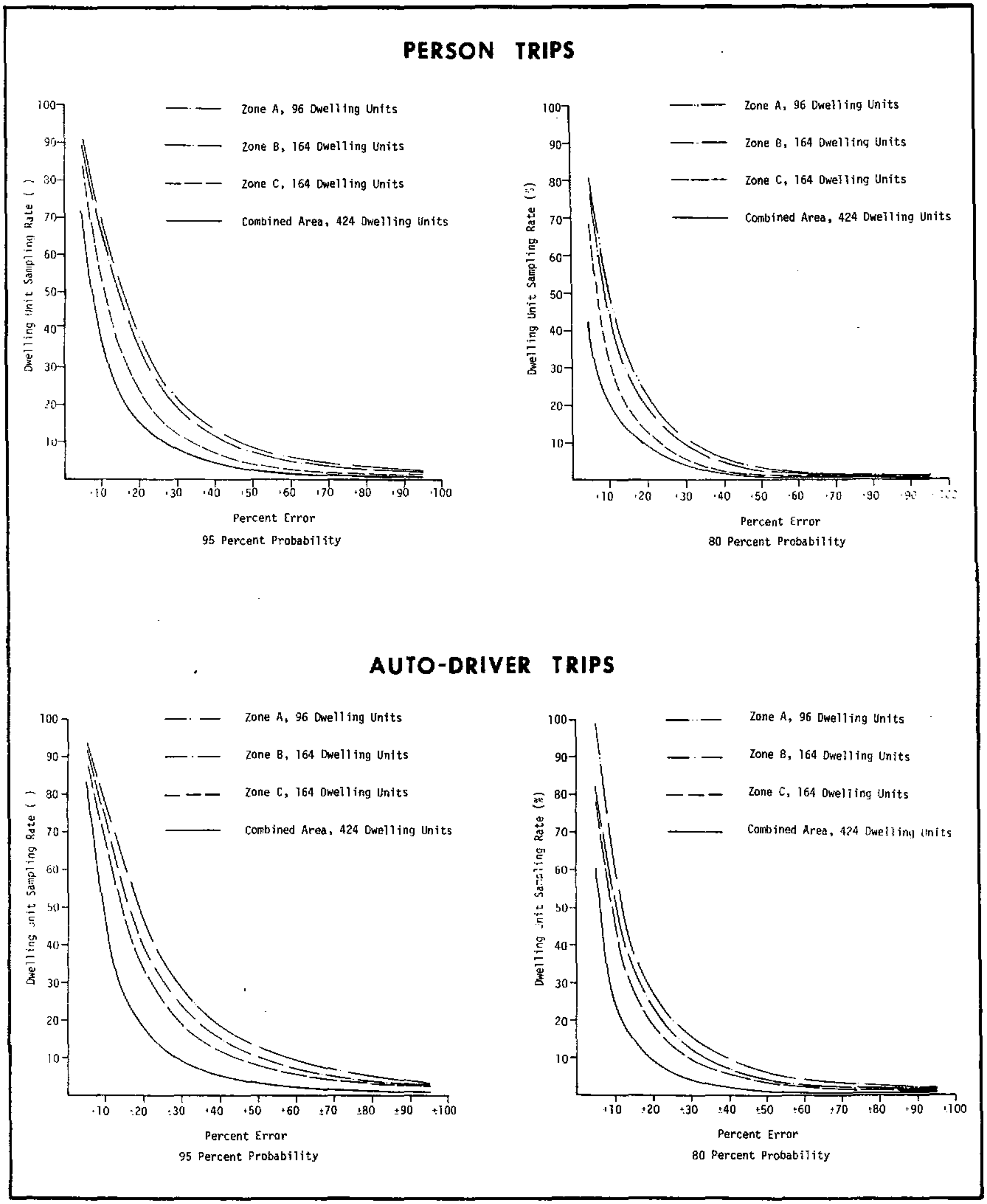

FIGURE IV-2: MARGIN OF ERROR VERSUS SAMPLING RATE FOR HOME-BASED WORK TRIPS

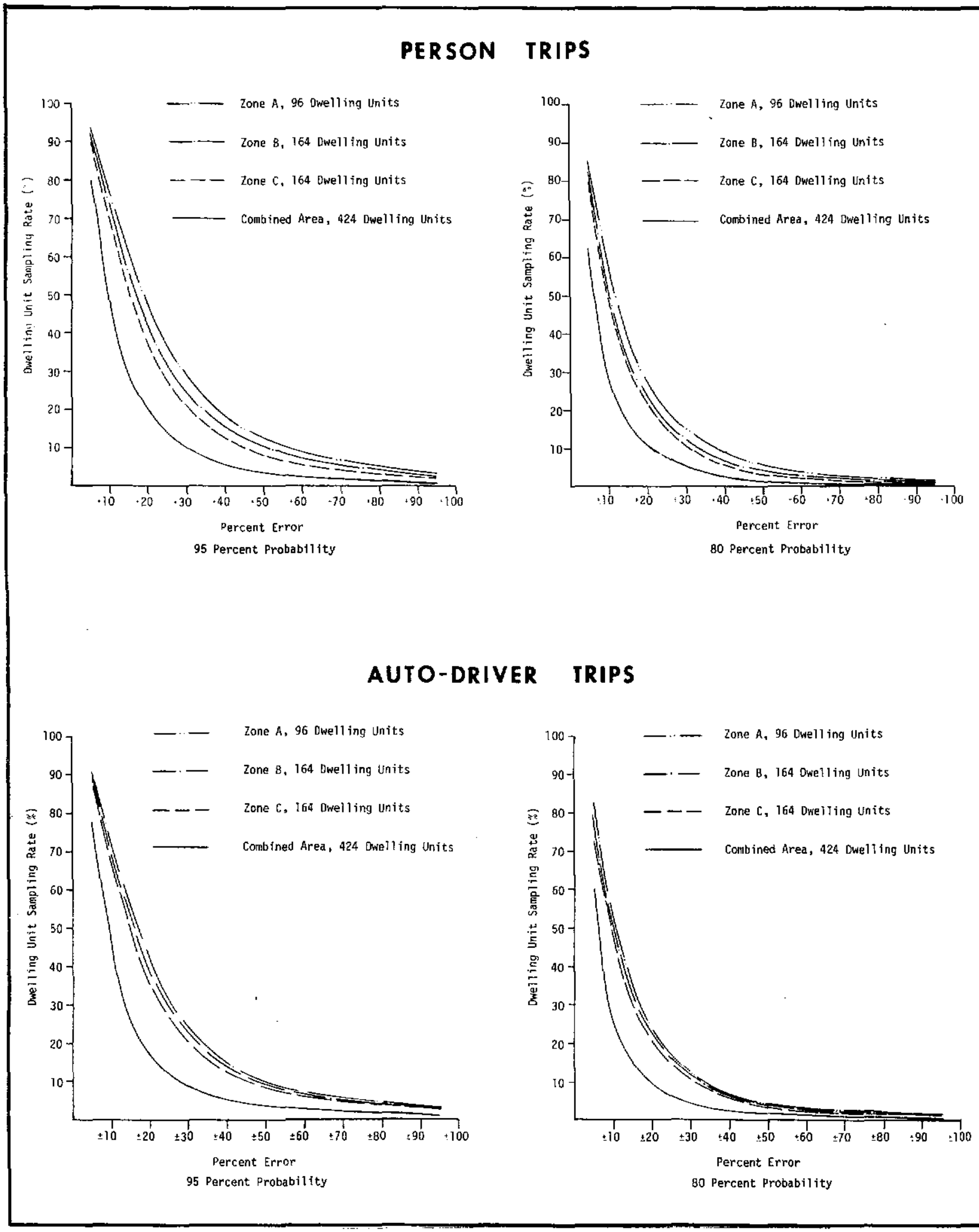

FIGURE IV-3: MARGIN OF ERROR VERSUS SAMPLING RATE FOR HOME-BASED NON-WORK TRIPS

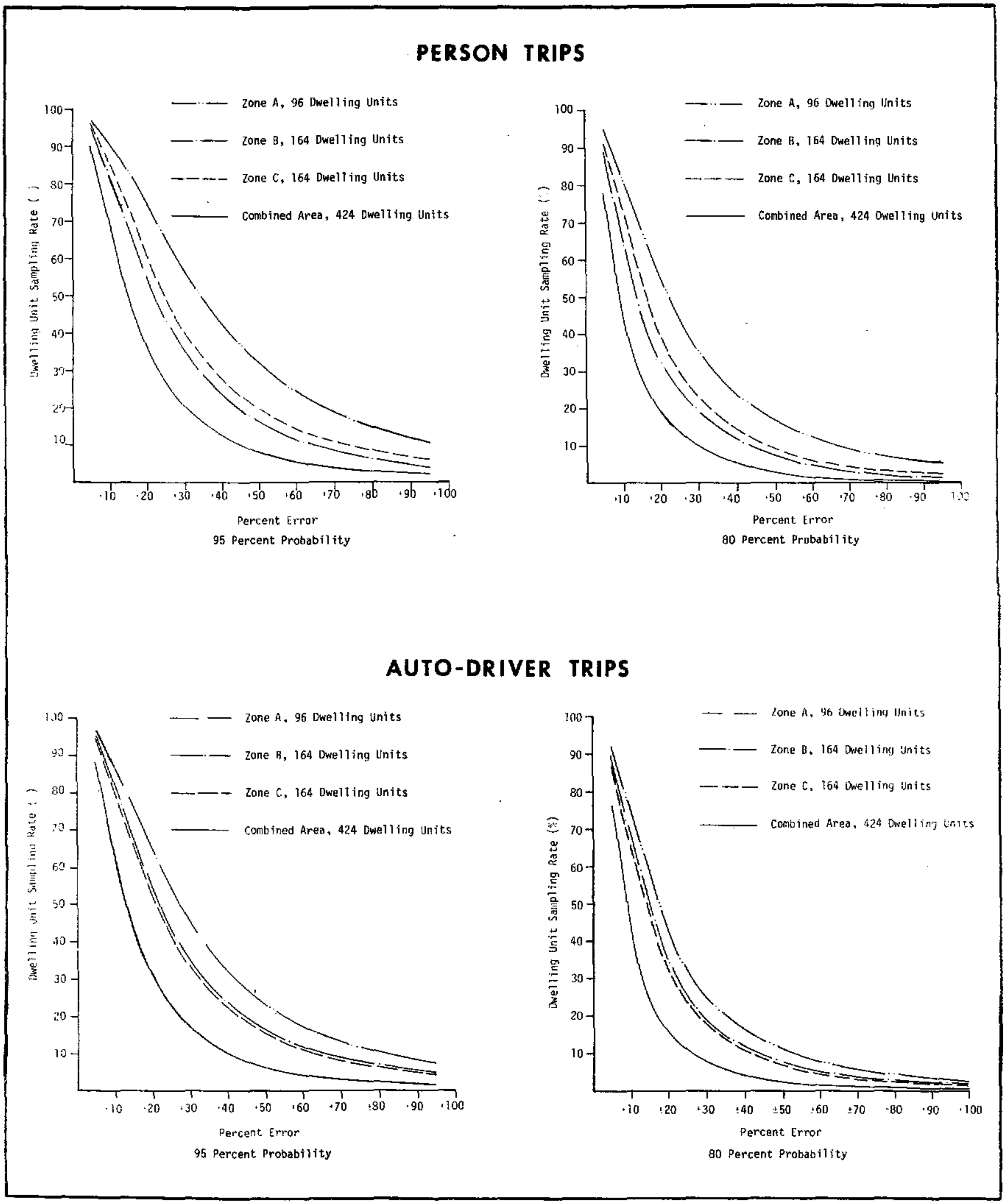

FIGURE IV-4: MARGIN OF ERROR VERSUS SAMPLING RATE FOR NONHOME-BASED TRIPS

Sample Size Versus Expected Error Range

Applying the formulas expressing the relationship between sample size and expected error range (i.e., Equations 1 and 3) provides useful perspective in the accuracy of sample data in estimating the mean trips per dwelling unit. These computations were limited to the eight trip categories for the combined area. The expected error ranges at the 95 percent probability level were computed for sample sizes from ten dwelling units to 250 dwelling units (i.e., sampling rates from approximately 2.4 percent to approximately 60 percent). The estimates of the means at the extremes of the error ranges (i.e., $\pm$ d) were then computed. Similar computations were performed at the 80 percent probability level. The results of these computations are illustrated in Figures IV-5 and IV-6.

As in the analysis relative to percent sample versus percent error, a number of general observations become apparent from careful inspection and comparison of the results depicted graphically in these figures. These include:

- Even at the 80 percent probability level, large sample sizes are required to estimate the mean trips per dwelling unit for the combined area within a reasonable expected error range for any of the trip categories.
- Again, increasing the sample size decreases the expected error range at a decreasing rate.
- For any trip purpose, the magnitude of the error ranges for the estimation of auto driver trips is generally less than that for the estimation of person trips. (It is interesting to note that a similar observation was made in terms of the percent error ranges, thereby suggesting that, at any given sampling level, a higher degree of accuracy may be generally expected in the estimation of the mean auto driver trips per dwelling unit than in the estimation of the mean person trips per dwelling unit.)
- When estimating the mean for either person or auto driver trips, the four trip purposes would generally be ranked, based on the magnitude of the error ranges (smallest to largest) as follows:

  home-based work trips
  nonhome-based trips
  home-based nonwork trips
  home-based trips

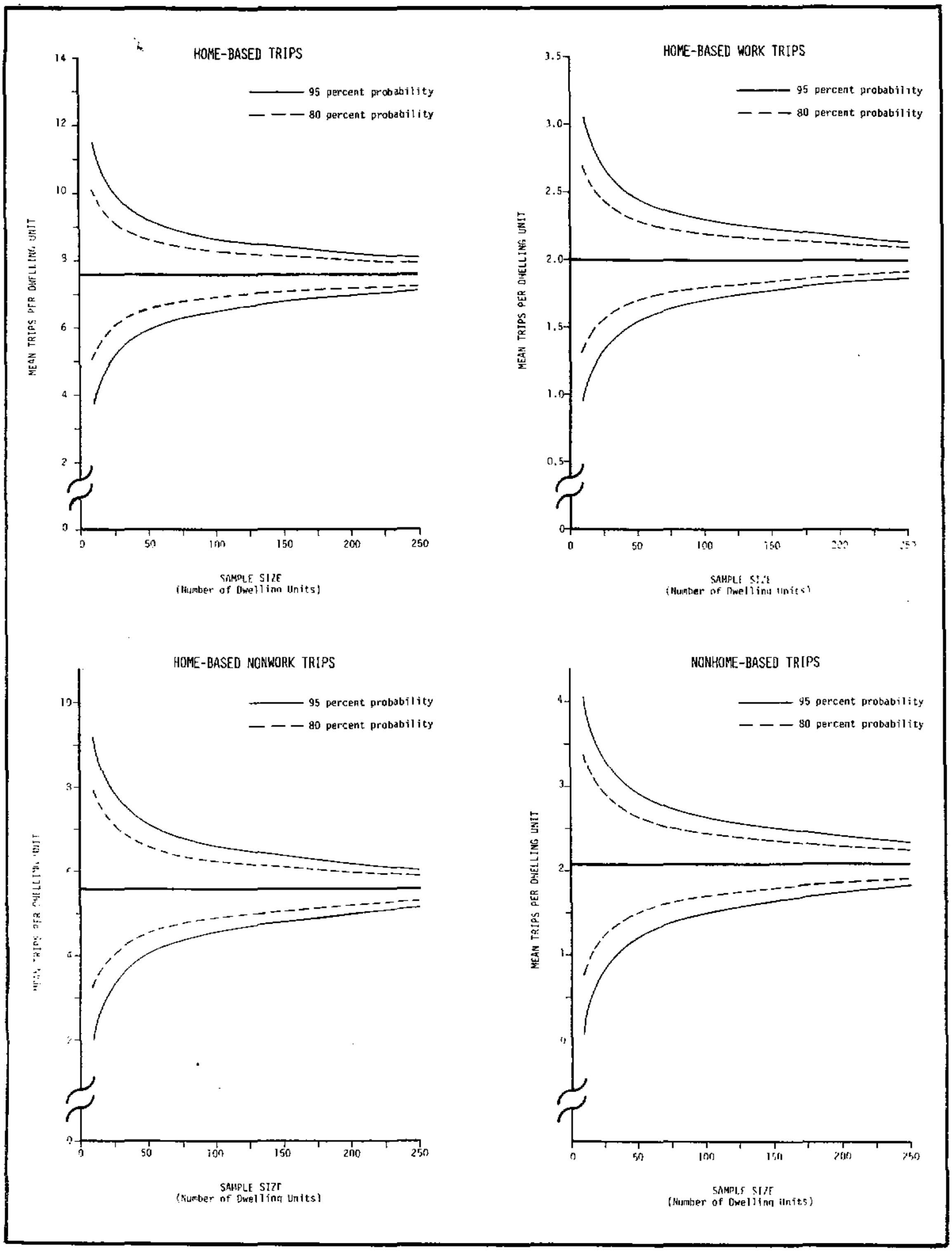

FIGURE IV-5: EXPECTED ERROR RANGES IN ESTIMATING THE MEAN PERSON TRIPS PER DWELLING UNIT FOR THE COMBINED AREA BY SAMPLE SIZE

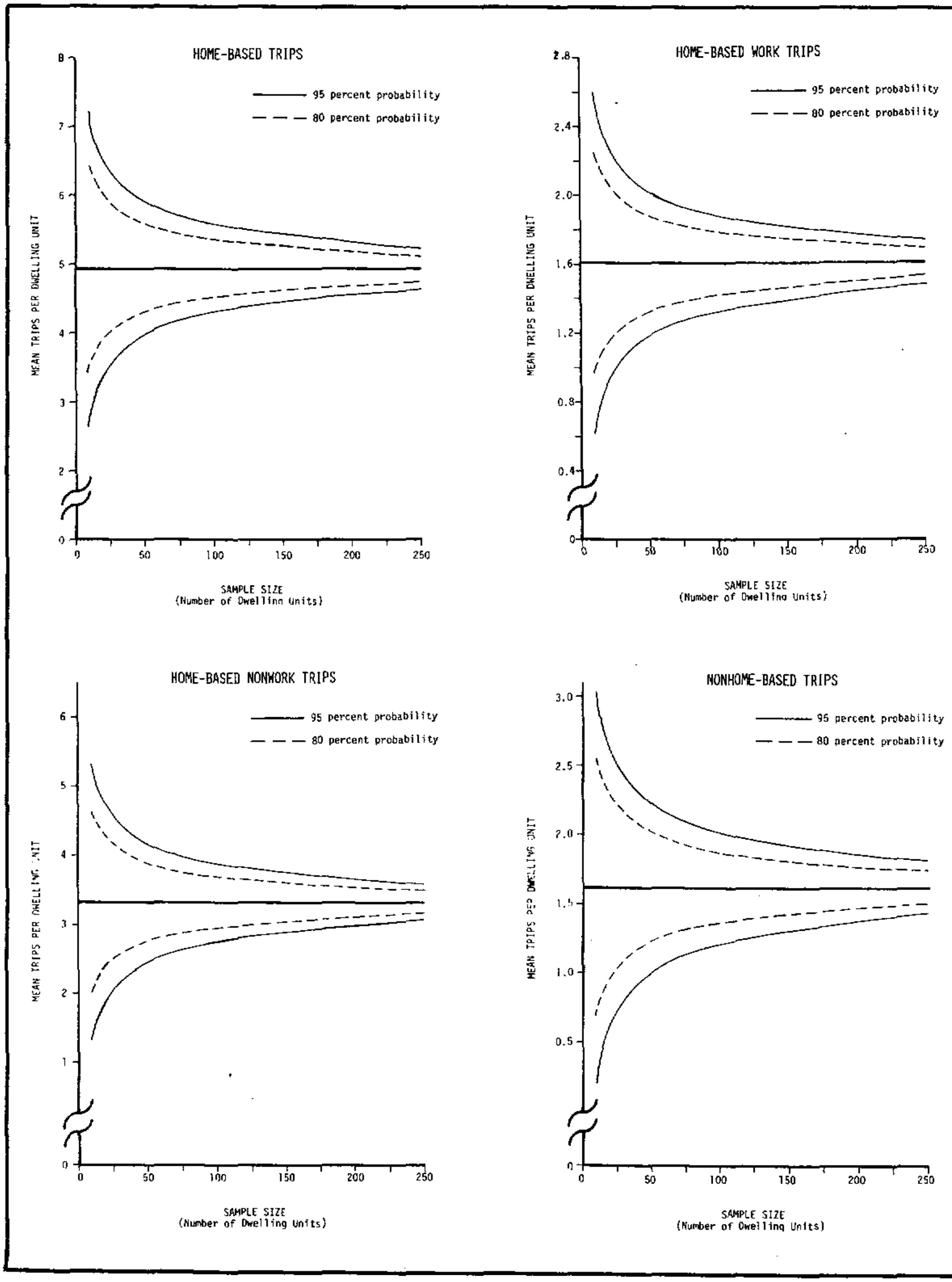

FIGURE IV-6: EXPECTED ERROR RANGES IN ESTIMATING THE MEAN AUTO DRIVER TRIPS PER DWELLING UNIT FOR THE COMBINED AREA BY SAMPLE SIZE

Again, it must be remembered that the last observation relates to the magnitude of the error range rather than the percent error range. The estimation of the mean home-based trips per dwelling unit has the smallest expected percent error range of the four trip purposes; however, the magnitude of this error range (in terms of trips per dwelling unit) is the largest of the four trip purposes.

Traditional Sampling Rates

The traditional sampling rates normally used in the home interview surveys are (13):

Population of area	Sample Size	Nominal sampling rate
under 50,000	1 in 5 dwelling units	20%
50,000 to 150,000	1 in 8 dwelling units	12.5%
150,000 to 300,000	1 in 10 dwelling units	10%
300,000 to 500,000	1 in 15 dwelling units	6.7%
500,000 to 1,000,000	1 in 20 dwelling units	5%
over 1,000,000	1 in 25 dwelling units	4%

The home interview origin-destination surveys performed in conjunction with the urban transportation studies in Texas have normally used either a nominal sampling rate of five percent (large urban areas) or 12.5 percent (small urban areas). Therefore, there was an interest in further evaluation at these two sampling rates. The expected percent error ranges at the 80 and 95 percent probability level for five percent and 12.5 percent sampling rates for the eight trip categories are given in Tables IV-1 and IV-2.

To demonstrate the severity of the magnitude of the error ranges, the estimates of the mean at the extremes of the error ranges for the various trip purposes were also computed for both the five and 12.5 percent sampling rates at both the 80 and 95 percent probability levels and are summarized in Tables IV-3 and IV-4.

These tables suggest that if repeated samples were drawn from Zone A, for example, using a nominal five percent sampling rate, the mean number of home-based person trips per dwelling unit may be expected to be estimated within a $\pm$64.9 percent error range (i.e., between 3.1 and 14.8 trips per dwelling unit) by 95 percent of the randomly drawn samples. In other words, there is a 95 percent probability of drawing a five percent sample from Zone A which will estimate that the mean number of home-based person trips per dwelling unit is within $\pm$64.9 percent of the true mean (i.e., between 3.1 and 14.8 trips per dwelling unit). This, of course, was demonstrated in the previous section, whereby the results of approximately 95 per-

NOMINAL 5 PERCENT SAMPLING RATE				
Trip Category	Expected Percent Error Range			
	Zone A	Zone B	Zone C	Combined Area
Home-based Person Trips	±64.9%	±61.7%	±52.9%	±34.7%
Home-based Auto Driver Trips	±56.9%	±54.6%	±50.8%	±31.5%
Home-based Work Person Trips	±68.0%	±65.9%	±50.2%	±35.8%
Home-based Work Auto Driver Trips	±81.0%	±72.6%	±64.2%	±41.8%
Home-based Nonwork Person Trips	±80.9%	±76.6%	±69.3%	±43.7%
Home-based Nonwork Auto Driver Trips	±72.1%	±72.7%	±66.8%	±41.2%
Nonhome-based Person Trips	±145.8%	±97.6%	±109.5%	±65.8%
Nonhome-based Auto Driver Trips	±115.6%	±99.4%	±97.1%	±59.9%

NOMINAL 12.5 PERCENT SAMPLING RATE				
Trip Category	Expected Percent Error Range			
	Zone A	Zone B	Zone C	Combined Area
Home-based Person Trips	±40.2%	±37.0%	±31.7%	±20.9%
Home-based Auto Driver Trips	±35.3%	±32.7%	±30.5%	±19.0%
Home-based Work Person Trips	±42.1%	±39.5%	±30.1%	±21.6%
Home-based Work Auto Driver Trips	±50.2%	±43.5%	±38.5%	±25.2%
Home-based Nonwork Person Trips	±50.2%	±45.9%	±41.5%	±26.4%
Home-based Nonwork Auto Driver Trips	±44.7%	±43.5%	±40.0%	±24.8%
Nonhome-based Person Trips	±90.4%	±58.5%	±65.6%	±39.8%
Nonhome-based Auto Driver Trips	±71.7%	±59.6%	±58.2%	±36.2%

TABLE IV-1: EXPECTED PERCENT ERROR RANGES AT THE 95 PERCENT PROBABILITY LEVEL FOR FIVE AND 12.5 PERCENT SAMPLING RATES

NOMINAL 5 PERCENT SAMPLING RATE				
Trip Category	Expected Percent Error Range			
	Zone A	Zone B	Zone C	Combined Area
Home-based Person Trips	±43.3%	±39.6%	±34.0%	±22.5%
Home-based Auto Driver Trips	±38.1%	±35.3%	±32.6%	±20.6%
Home-based Work Person Trips	±45.6%	±42.2%	±32.2%	±23.3%
Home-based Work Auto Driver Trips	±54.1%	±47.0%	±41.4%	±27.1%
Home-based Nonwork Person Trips	±54.1%	±49.2%	±44.4%	±28.5%
Home-based Nonwork Auto Driver Trips	±47.8%	±47.0%	±43.1%	±26.8%
Nonhome-based Person Trips	±97.3%	±62.7%	±70.6%	±42.8%
Nonhome-based Auto Driver Trips	±77.4%	±64.0%	±62.7%	±38.8%

NOMINAL 12.5 PERCENT SAMPLING RATE				
Trip Category	Expected Percent Error Range			
	Zone A	Zone B	Zone C	Combined Area
Home-based Person Trips	±26.3%	±24.1%	±20.6%	±13.7%
Home-based Auto Driver Trips	±23.2%	±21.4%	±19.8%	±12.5%
Home-based Work Person Trips	±27.6%	±25.6%	±19.6%	±14.2%
Home-based Work Auto Driver Trips	±32.8%	±28.5%	±25.1%	±16.5%
Home-based Nonwork Person Trips	±32.8%	±29.9%	±27.0%	±17.3%
Home-based Nonwork Auto Driver Trips	±29.0%	±28.5%	±26.2%	±16.3%
Nonhome-based Person Trips	±59.1%	±38.1%	±42.8%	±25.0%
Nonhome-based Auto Driver Trips	±47.0%	±38.9%	±38.1%	±23.5%

TABLE IV-2: EXPECTED PERCENT ERROR RANGES AT THE 80 PERCENT PROBABILITY LEVEL FOR FIVE AND 12.5 PERCENT SAMPLING RATES

Nominal 5 Percent Sampling Rate

	Zone A	Zone B	Zone C	Combined Area
Home-based Person Trips	3.1 to 14.8	2.8 to 11.6	3.4 to 11.0	5.0 to 10.2
Home-based Auto Driver Trips	2.5 to 9.2	2.1 to 7.2	2.3 to 7.0	3.4 to 6.5
Home-based Work Person Trips	0.7 to 3.6	0.7 to 3.4	0.9 to 2.8	1.3 to 2.7
Home-based Work Auto Driver Trips	0.3 to 3.3	0.4 to 2.7	0.5 to 2.5	0.9 to 2.3
Home-based Nonwork Person Trips	1.3 to 12.3	1.2 to 9.1	1.6 to 9.0	3.2 to 8.0
Home-based Nonwork Auto Driver Trips	1.1 to 7.0	0.8 to 5.3	1.1 to 5.3	2.0 to 4.7
Nonhome-based Person Trips	0.0 to 5.5	0.1 to 4.1	0.0 to 4.2	0.7 to 3.4
Nonhome-based Auto Driver Trips	0.0 to 3.3	0.0 to 3.3	0.1 to 3.2	0.7 to 2.6

Nominal 12.5 Percent Sampling Rate

	Zone A	Zone B	Zone C	Combined Area
Home-based Person Trips	5.4 to 12.6	4.5 to 9.9	4.9 to 9.5	6.0 to 9.2
Home-based Auto Driver Trips	3.8 to 7.9	3.1 to 6.2	3.2 to 6.1	4.0 to 5.9
Home-based Work Person Trips	1.2 to 3.0	1.3 to 2.9	1.3 to 2.4	1.6 to 2.4
Home-based Work Auto Driver Trips	0.9 to 2.7	0.9 to 2.3	0.9 to 2.1	1.2 to 2.0
Home-based Nonwork Person Trips	3.4 to 10.2	2.8 to 7.5	3.1 to 7.5	4.1 to 7.1
Home-based Nonwork Auto Driver Trips	2.3 to 5.9	1.7 to 4.4	1.9 to 4.4	2.5 to 4.1
Nonhome-based Person Trips	0.2 to 4.2	0.9 to 3.3	0.7 to 3.3	1.3 to 2.9
Nonhome-based Auto Driver Trips	0.4 to 2.6	0.7 to 2.7	0.7 to 2.6	1.0 to 2.2

TABLE IV-3: SUMMARY OF ERROR RANGES IN TERMS OF THE MEAN TRIPS PER DWELLING UNIT AT 95 PERCENT PROBABILITY LEVEL

Nominal 5 Percent Sampling Rate

	Zone A	Zone B	Zone C	Combined Area
Home-based Person Trips	5.1 to 12.9	4.3 to 10.1	4.8 to 9.6	5.9 to 9.3
Home-based Auto Driver Trips	3.7 to 8.1	3.1 to 6.3	3.2 to 6.2	3.9 to 5.9
Home-based Work Person Trips	1.1 to 3.1	1.2 to 3.0	1.3 to 2.5	1.5 to 2.5
Home-based Work Auto Driver Trips	0.8 to 2.8	0.9 to 2.3	0.9 to 2.1	1.2 to 2.1
Home-based Nonwork Person Trips	3.1 to 10.5	2.6 to 7.6	2.9 to 7.7	4.0 to 7.2
Home-based Nonwork Auto Driver Trips	2.2 to 6.0	1.7 to 4.5	1.9 to 4.5	2.4 to 4.2
Nonhome-based Person Trips	0.0 to 4.4	0.8 to 3.4	0.6 to 3.4	1.2 to 3.0
Nonhome-based Auto Driver Trips	0.3 to 2.7	0.6 to 2.8	0.6 to 2.6	1.0 to 2.3

Nominal 12.5 Percent Sampling Rate

	Zone A	Zone B	Zone C	Combined Area
Home-based Person Trips	6.6 to 11.4	5.5 to 8.9	5.7 to 8.7	6.6 to 8.6
Home-based Auto Driver Trips	4.5 to 7.3	3.7 to 5.7	3.8 to 5.6	4.3 to 5.5
Home-based Work Person Trips	1.5 to 2.7	1.6 to 2.6	1.5 to 2.3	1.7 to 2.3
Home-based Work Auto Driver Trips	1.2 to 2.4	1.1 to 2.1	1.1 to 1.9	1.4 to 1.9
Home-based Nonwork Person Trips	4.6 to 9.0	3.6 to 6.6	3.9 to 6.7	4.6 to 6.6
Home-based Nonwork Auto Driver Trips	2.9 to 5.3	2.2 to 4.0	2.4 to 4.0	2.8 to 3.9
Nonhome-based Person Trips	0.9 to 3.5	1.3 to 2.9	1.1 to 2.9	1.5 to 2.6
Nonhome-based Auto Driver Trips	0.8 to 2.2	1.1 to 2.3	1.0 to 2.2	1.2 to 2.0

TABLE IV-4: SUMMARY OF ERROR RANGES IN TERMS OF THE MEAN TRIPS PER DWELLING UNIT AT 80 PERCENT PROBABILITY LEVEL cent of the sample did lie within the expected error range.

Increasing the sampling rate, of course, will reduce the error ranges. Comparing the error ranges for a nominal 12.5 percent sample with those for a nominal five percent sample, it can be seen that by increasing the sampling rate by a factor of 2.5, the error ranges are only reduced by a factor of roughly 0.6. Even with the 12.5 percent sample, the error ranges remain disturbingly large—especiall in view of the costs associated with such a sampling rate.

Comparing the error ranges in Tables IV-1 and IV-3 with those in Tables IV-2 and IV-4, it may be observed that the expected error ranges at the 80 percent probability level are approximately 65 percent the magnitude of those at the 95 percent probability level. Thus, there is an 80 percent probability of drawing a five percent sample from Zone A, for example, which will estimate the mean home-based person trips within ±42.4 percent of the true mean, as compared with the ±64.9 percent at the 95 percent probability level. However, it must be remembered that the 80 percent probability level implies that, on the average, one out of every five samples may be expected to yield an estimate of the mean which lies outside the expected error range.

Using a 50 percent probability level would reduce the error ranges still further, such that, for a nominal five percent sample in Zone A, there would be a 50-50 chance of drawing a sample which would estimate the mean number of home-based person trips per dwelling unit for Zone A within approximately ±22 percent of the true mean (i.e., between 7.0 and 10.9 trips per dwelling unit) and the mean number of nonhome-based person trips per dwelling unit within approximately ±50 percent of the true mean (i.e., between 1.1 and 3.3 trips per dwelling unit). It must be emphasized that, while lowering the probability level reduces the expected error ranges, it does not improve the accuracy of estimates from sample data. Quite the contrary, lowering the probability level, and thereby reducing the expected error range, simply increases the probability of drawing a sample which will yield an estimate of the mean outside the specified expected error range.

In view of the magnitude of these error ranges and the homogeneity and the nonunique character of the area, it would seem reasonable to expect that an experienced urban transportation study analyst could, through a careful inspection of the area, estimate the means within the error ranges that may be expected from conventional sampling rates. Such procedures utilizing experienced analysts would be considerably less expensive than the traditional home interview survey, and at the same time would provide a level of accuracy that should be comparable to that of a home interview survey.

It is to be emphasized, however, that this does not suggest that the home interview surveys performed in the past have been unnecessary. To the contrary, these surveys provide the extensive travel data upon which the experienced analyst must be able to draw.

It must also be emphasized that it does not suggest that home interview surveys should be completely abandoned. It does, however, suggest a new direction for home interview surveys. Instead of a general survey of the entire urban area, special surveys, involving a limited number of observations, might be employed which would be specifically directed at monitoring and updating previously established trip generation rates and investigating areas which exhibit unique or unusual characteristics.

Significant Parameters

From Equation 8 it may be observed that the two significant parameters determining the relationship between the sampling rate and the expected percent error range (at a given probability level) are the zone size and the coefficient of variation of the distribution of dwelling units by trip productivity (for the specified trip category). Similarly, from Equation 3 it may be observed that the two significant parameters determining the relationship between sample size and the expected error range are the zone size and the standard deviation of the distribution of dwelling units by trip productivity. The following focuses attention on the sensitivity of the relationships (i.e., Equations 3 and 8) to these parameters and their implications regarding the delineation of zones.

Coefficient of Variation

For a given zone, the differences between the percent error ranges for the different trip categories are attributable only to variations in the population means and standard deviations; or, more specifically, to the differences in the coefficients of variation. The sensitivity of the percent error range to changes in the coefficient of variation may be observed by comparison of the coefficients of variation in Table IV-5 and the percent error ranges in Tables IV-1 and IV-2. For example, in Zone A, the coefficients of variation range from 0.67 to 1.71, while the associated percent error ranges at the 95 percent probability level vary from ± 57 percent to ± 146 percent. Since the other terms in the formula for percent error (Equation 8) are constants for any given zone, it can be seen that the percent error varies directly with the coefficient of variation. The coefficient of variation is, therefore, an extremely significant parameter in the relationship between the sampling rate and the percent error range.

Since the percent error range will vary directly with the coefficient of variation, one of the objects in zonal delineation should be the minimization of the coefficient of variation. This might be attempted by defining zones such that the dwelling units within each zone exhibit, as nearly as possible, the same socioeconomic characteristics. In other words, delineate zones so that each zone represents as homogeneous a group of dwelling units as is practicably possible.

Standard Deviation

For a given zone, the differences in the magnitudes of the error ranges (i.e., the variable d, in Equation 3) are, of course, attributable only to differences in the population standard deviation. The sensitivity of the magnitude of the error range to the standard deviation may be observed by comparison of the standard deviations in Table IV-5 and the magnitude of the error ranges in Tables IV-3 and IV-4. For example, in Zone A, the standard deviations vary from 0.67 to 1.71, while the associated percent error ranges at the 95 percent probability level vary from ± 57 percent to ± 146 percent. Since the other parameters in the formula for percent error (Equation 8) are held constant for any given zone, it can be seen that the percent error varies directly with the standard deviation.

Zone Size

The number of occupied dwelling units in the zone is the other significant parameter in both the relationship between the sampling rate and the percent error range (Equation 8) and the relationship between the sampling rate and the magnitude of the error range (Equation 3). In this regard, the difference between the error ranges for the individual zones and the error ranges for the combined area might be noted. For each trip category, (Table IV-5) the coefficient of variation and the standard deviation for the combined area consistently lie within the range of the coefficients of variation and the standard deviations for the individual zones. Therefore, the significant differences in the error ranges between the individual survey zone and the combined area are largely attributable to the effect of zone size.

As can be seen from Equations 3 and 8, if the coefficient of variation and standard deviation remain constant, the percent error range (Equation 8) and the magnitude of the error range (Equation 3) will vary inversely with the square root of the number of occupied dwelling units in the zone. In other words, assuming a constant coefficient of variation and standard deviation, the larger the number of occupied dwelling units in a zone, the smaller the percent error range (as well as

COEFFICIENTS OF VARIATION

Trip Category	Zone A	Zone B	Zone C	Combined Area
Home-based Person Trips	0.76	0.91	0.78	0.83
Home-based Auto Driver Trips	0.67	0.81	0.75	0.76
Home-based Work Person Trips	0.80	0.97	0.74	0.86
Home-based Work Auto Driver Trips	0.95	1.08	0.95	1.00
Home-based Nonwork Person Trips	0.95	1.13	1.02	1.05
Home-based Nonwork Auto Driver Trips	0.84	1.08	0.99	0.99
Nonhome-based Person Trips	1.71	1.44	1.62	1.58
Nonhome-based Auto Driver Trips	1.36	1.47	1.44	1.43

STANDARD DEVIATIONS

Trip Category	Zone A	Zone B	Zone C	Combined Area
Home-based Person Trips	6.81	6.57	5.63	6.31
Home-based Auto Driver Trips	3.91	3.76	3.50	3.72
Home-based Work Person Trips	1.70	2.01	1.39	1.72
Home-based Work Auto Driver Trips	1.71	1.71	1.43	1.61
Home-based Nonwork Person Trips	6.47	5.82	5.45	5.86
Home-based Nonwork Auto Driver Trips	3.43	3.30	3.11	3.27
Nonhome-based Person Trips	3.79	2.96	3.25	3.27
Nonhome-based Auto Driver Trips	2.06	2.44	2.34	2.32

ZONE SIZES

Zone A = 96 occupied dwelling units

Zone B = 164 occupied dwelling units

Zone C = 164 occupied dwelling units

Combined Area = 424 occupied dwelling units

TABLE IV-5: SUMMARY OF THE SIGNIFICANT PARAMETERS FROM THE 100 PERCENT DATA the magnitude of the error range) for a given sampling rate.

It is important to note that, in these observations, the sampling rate is assumed to remain constant; thus, the sample size increases as the zone size increases. In other words, assuming the sampling rate remains constant, if the zone size is increased by a factor of 4, then the sample size will also be increased by a factor of 4 and the expected error range (both magnitude and percent) relative to the mean will be reduced by one-half.

Since the expected error relative to the mean varies inversely with the square root of the number of dwelling units, another objective in zonal delineation should be to the maximize "zone size" assignment network. Unfortunately, this objective generally conflicts with the objectives relative to the minimization of the coefficient of variation and standard deviation since these parameters are a function of the variance which will increase with "zone size." It may be assumed that, when increasing the zone size, the coeffieicnt of variation will vary directly with the standard deviation. Then, the objective of zonal delineation should be to minimize the function $[C/\sqrt{N}]$; that is to minimize the ratio of the coefficient of variation to the square root of the zone size. However, the square root of the number of observations generally increases faster than the increase in the variance.

Variance of Estimates of Population Parameters

The discussion of the accuracy of sample data in estimating the population mean is predicted on the fact that the distribution of the estimates of the population mean from repeated random samples will approximate a normal distribution. Under the normality assumption, the variance of the estimates of the mean at the various sampling levels provided the bases for the statistical inferences relative to the expected error ranges at those sampling levels. It is worthwhile, therefore, to specifically examine this variance of estimate in order to provide further perspective in the accuracy of sample data in estimating the population mean.

Since sample data are generally used to estimate the population variance, it is worthwhile to review the accuracy of sample data in the estimation of this population parameter. The results of repeated random samples drawn from the 100 percent data provide a useful insight into the distribution and variance of these estimates at various sampling levels. Unfortunately, sampling theory does not provide the same useful theoretical constructs for this analysis as were available relative to the estimation of the mean. The statistical inferences relative to the estimation of the population variance will, therefore, be limited to those sampling levels at which statistical tests could not reject the assumption of the normality of the distribution of these estimates. At these sampling levels, the estimated variance of these population parameters (i.e., the variance observed from repeated random samples) may be used to estimate the expected error range at a given level of confidence.

Population Mean

The formulas for the expected error ranges in estimating the mean trips per dwelling unit (as discussed in the section entitled "Theoretical Constructs") not only assume that the distribution of the estimates of the mean will approximate a normal distribution but that the variance of this distribution may be estimated by:

$$S^2 = \frac{\sigma^2}{\frac{Nn}{N - n}}$$

Where

S^2 = the expected variance of the estimates of the mean from repeated random samples.

σ^2 = the population variance (i.e., the variance between dwelling units in trip productivity).

N = the number of occupied dwelling units in the zone.

n = the sample size.

It is worthwhile, therefore, to review the variance of the estimates of the population mean, both in terms of the theoretical variance (computed using the above formula) and the estimated variance, from a large number of random samples drawn from the 100 percent data. Both the theoretical and observed variances for sample sizes up to 300 dwelling units in the combined area are displayed graphically in Figures IV-7 and IV-8 for the eight trip categories. It must be emphasized that the "observed" points in these figures are indeed estimates, since the true variance for a given sample size may be obtained only by drawing all possible samples of that size from the 100 percent data. The close correlation between the theoretical and observed data in these figures again demonstrates the applicability of the formulas used extensively in the preceding discussion of the accuracy of survey data in estimating the mean.

Careful inspection of these figures leads to some interesting observations regarding the cost effectiveness sample size relative to the variance of estimates. First of all, increasing sample size, of course, decreases the variance of estimates but at a decreasing rate. That is, increasing the sample size from 20 to 25 dwelling units results in a greater reduction in the variance of estimates than increasing the

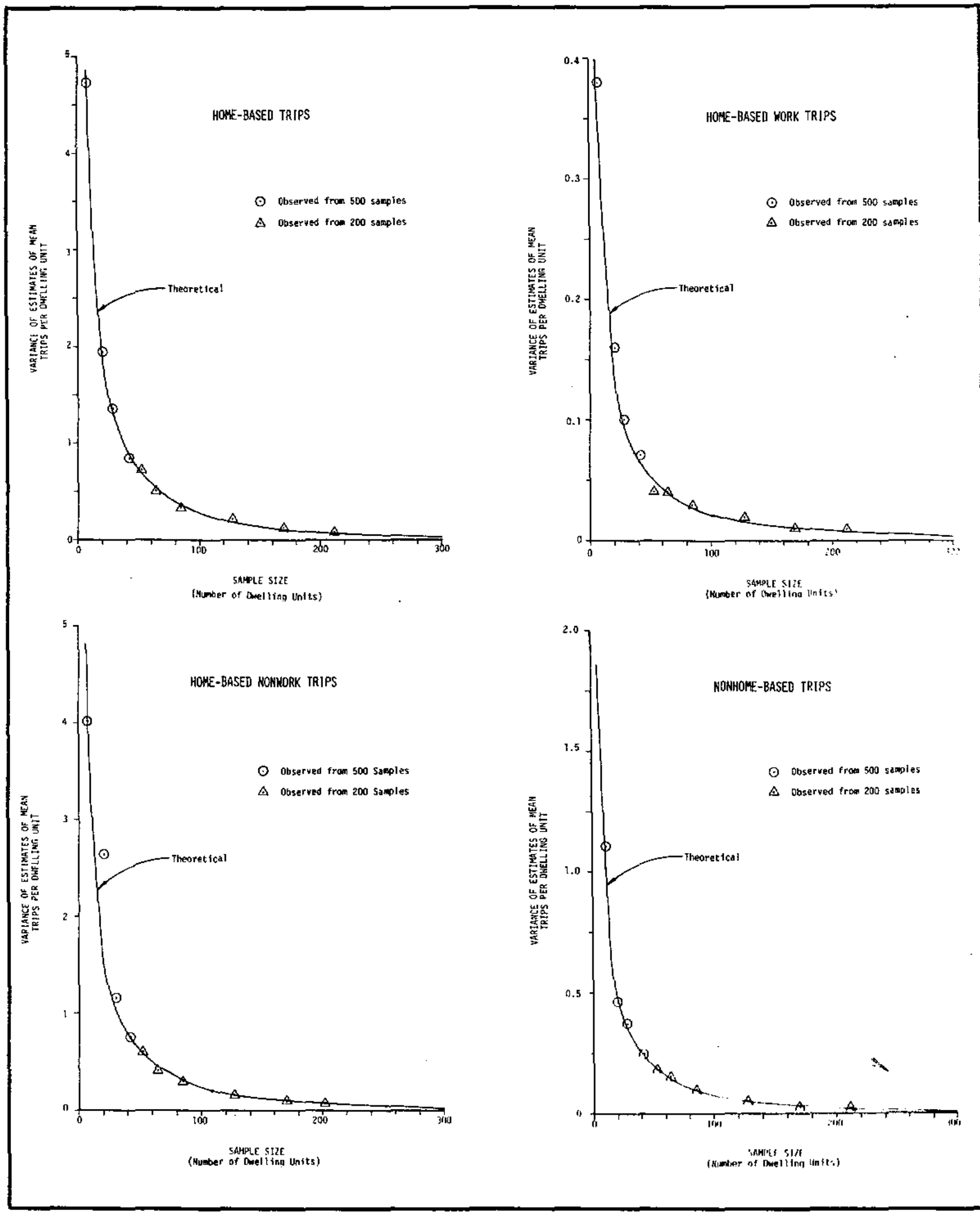

FIGURE IV-7: VARIANCE OF ESTIMATES OF THE MEAN PERSON TRIPS PER DWELLING UNIT BY SAMPLE SIZE FOR THE COMBINED AREA

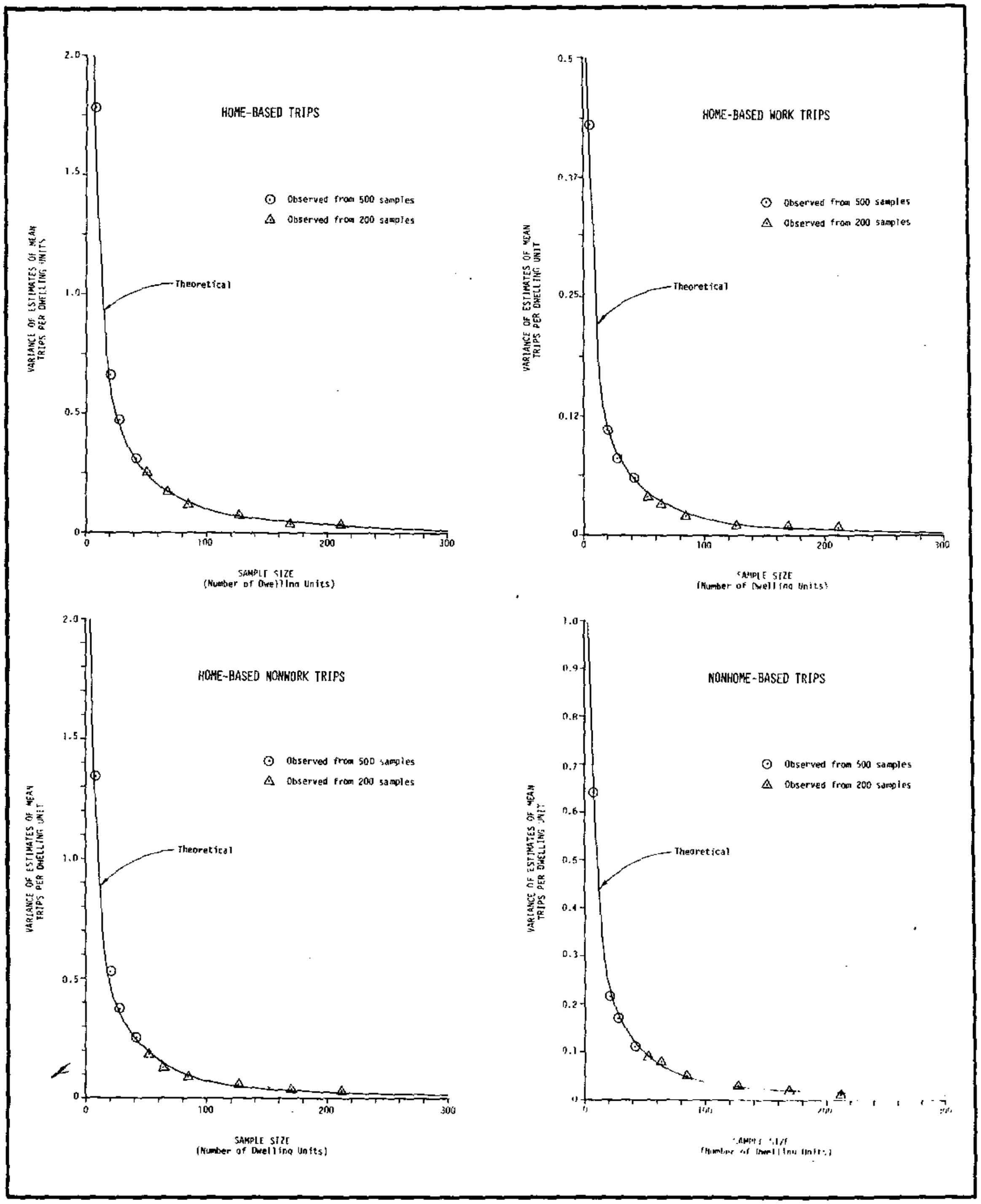

FIGURE IV-8: VARIANCE OF ESTIMATES OF THE MEAN AUTO DRIVER TRIPS PER DWELLING UNIT BY SAMPLE SIZE FOR THE COMBINED AREA sample size from say 40 dwelling units to 45 dwelling units. If the cost per home interview is assumed to be a constant beyond some small threshold sample size, then the cost of each additional observation (i.e., home interview) yields a smaller return on investment (in terms of the amount of reduction realized in the variance of estimates) than the preceding observation. It is also interesting to note that, in each instance, the curve becomes almost a straight line with a small slope beyond the "knee" of the curve. Further, regardless of the trip category, this tendency of the curve to "level off" occurs at, or about, a sample size of 100 to 120 dwelling units. It may, therefore, be argued that, regardless of the trip category, the reduction in the variance of estimates obtained by increasing the sample size beyond approximately 100 to 120 dwelling units is probably not sufficient to offset the increase in costs.

Population Variance

When attempting to determine the sample size required to adequately estimate the trips in an urban area, an estimate of the population variance is required. This estimate of the population variance is generally obtained from a small random sample. Through the use of repeated random samples drawn from the 100 percent data, it was shown that the distribution of the estimates of the population variance approximates a normal distribution. Assuming this distribution is normal, the observed variance of these estimates may be used to define intervals for the 80 and 95 percent confidence levels, whereby the probability of drawing a sample which will yield an estimate of the population variance within these intervals is 80 and 95 percent, respectively. These intervals (or expected error ranges) for the estimation of the population variance are computed (for given sample size), using the formula

$$\overline{X} \pm t\sqrt{S^2}$$

Where

$\overline{X}$ = the average estimate of the population variance.

S^2 = the observed variance of the estimates of the population variance from the samples drawn from the 100 percent data.

t = the value from the normal table corresponding to the desired level of confidence (e.g., 1.96 for 95 percent confidence).

It must be emphasized that the observed variance of the estimates is only an estimate of the true variance of estimate, since the true variance of estimate may only be computed by drawing all possible samples from the data. Nevertheless, the observed variance is felt to be an accurate estimate since it is based on 1,000 random samples. The intervals for 80 and 95 percent confidence levels were computed for the eight trip categories using the combined area for sampling rates of five and ten percent. The results of these computations are shown in Table IV-6. As can be seen, the expected error ranges in the estimation of the population variance, even at the 80 percent confidence level, are disturbingly large.

Since these estimates of the population variance may be used to estimate the sample size requirements for the combined area, the end points of the intervals were used to compute the estimated required sampling rates for estimating the mean trip per dwelling unit within ± 10 percent. The results of these computations are displayed in Table IV-7. These results suggest that, for example, there is a 95 percent probability of drawing a five percent sample from the combined area, which, when used to compute the sample size requirements for estimating the mean home-based work person trips within ± 10 percent, would yield a required sampling rate between 0 and 85.5 percent. Similarly, there is an 80 percent probability of drawing a five percent sample from the combined area which would result in an estimate of the required sampling rate (for estimating the mean home-based work person trips within ± 10 percent) in the range from a 37 percent sample to an 82.1 percent sample. From this table, it can be seen that the expected error range in the estimation of the required sampling rate from sample data is very large.

Implications Regarding Trip Generation Analysis

The analysis, thus far, has focused largely on the relationships between sampling rates and the expected error in the estimation of the mean trips per dwelling unit. It is appropriate at this point to review some of the implications of this analysis relative to disaggregate trip generation analysis.

The statistical relationships between sampling rate and percent error are also applicable to the estimation of trip generation rates by cross-classification. Reference to population has been generally used in regard to the number of dwelling units within a specified geographical area (i.e., within the zonal boundaries). Cross-classification procedures can be employed to group occupied dwelling units by socioeconomic characteristics rather than grouping zones or aggregated data. In other words, a cross-classification scheme would define what might be termed "socioeconomic zones" (i.e., populations possessing certain common socioeconomic characteristics). In applying Equation 5 to cross-classifications, therefore, the parameter N would refer to the total number of occupied dwelling units within an

Trip Purpose	No. Dwelling Units Sampled	No. Samples	Estimated Population Variance				
			Mean	95 Percent Confidence		80 Percent Confidence	
				Low	High	Low	High
HB(P)	21	1,000	38.76	3.90	73.62	16.00	61.52
	42	1,000	40.12	15.43	64.81	24.00	56.24
HB(AD)	21	1,000	13.80	3.01	24.59	6.75	20.85
	42	1,000	13.92	6.52	21.32	9.09	18.75
HBW(P)	21	1,000	2.83	-0.51	6.17	0.65	5.01
	42	1,000	2.91	0.50	5.32	1.33	4.49
HBW(AD)	21	1,000	2.53	0.01	5.05	0.89	4.17
	42	1,000	2.56	0.82	4.30	1.43	3.69
HBNW(P)	21	1,000	33.84	1.42	66.26	12.67	55.01
	42	1,000	34.68	12.40	56.96	20.13	49.23
HBNW(AD)	21	1,000	10.76	1.74	19.78	4.87	16.65
	42	1,000	10.74	4.50	16.98	6.66	14.82
NHB(P)	21	1,000	10.24	-8.36	28.84	-1.91	22.39
	42	1,000	10.68	-2.53	23.89	2.05	19.31
NHB(AD)	21	1,000	5.38	-1.28	12.04	1.03	9.73
	42	1,000	5.38	1.03	9.73	2.54	8.22

TABLE IV-6: EXPECTED ERROR RANGES IN ESTIMATING THE POPULATION VARIANCE

Trip Purpose	No. Dwelling Units Sampled	Percent Sampling Rates for ±10 Percent Error			
		95 Percent Confidence		80 Percent Confidence	
		Low	High	Low	High
HB(P)	21	78.1	98.6	93.6	98.3
	42	93.4	98.3	95.8	98.1
HB(AD)	21	73.3	95.8	86.1	95.0
	42	85.4	95.3	89.2	94.6
HBW(P)	21	0.0	85.8	37.0	82.1
	42	31.4	83.0	54.7	80.4
HBW(AD)	21	1.4	82.1	44.8	79.2
	42	42.9	79.7	56.6	77.1
HBNW(P)	21	56.4	98.6	92.2	98.1
	42	92.0	98.1	94.8	97.9
HBNW(AD)	21	61.3	94.8	81.6	93.9
	42	80.4	94.1	85.8	93.2
NHB(P)	21	0.0	96.5	0.0	95.5
	42	0.0	95.8	65.1	94.8
NHB(AD)	21	0.0	91.7	48.3	89.9
	42	48.6	89.9	69.8	88.2

TABLE IV-7: EXPECTED ERROR RANGES IN ESTIMATING THE REQUIRED SAMPLING RATES urban area which satisfy a given set of cross-classification critera.

By the judicious selection of cross-classification criteria, it is possible to define considerably larger and more homogeneous populations than is normally possible with geographical zones. In doing so, it may be expected that the ratio of the coefficient of variation to the square root of the population size would be substantially smaller than the ratios for geographical zones. This, of course, suggests that, for a given sampling rate, substantially smaller percent error ranges might be expected to be associated with the trip rates for homogeneous socioeconomic populations than with the geographic population (or zone).

These relationships are also applicable to regression analysis of disaggregate data (i.e., regression analysis using the dwelling unit as the unit of observation). In this instance, the populations are essentially defined by the independent variables. For example, in performing a simple regression analysis of disaggregate data with income level as the independent variable, the collection of points corresponding to a given income level essentially represent a sample of what has been referred to as a "socioeconomic" population. The relationships between sampling rate and expected error give an indication of how accurately the sample may be expected to represent the population of dwelling units having a given income level.

With judicious selection of the independent variables, the ratio of the coefficient of variation to the square root of the population size can be minimized more efficiently than for geographical zones. This suggests that disaggregate regression equations on cross-classification ratios should provide statistically more efficient estimates of the zonal trip ends than the expansion per dwelling unit trip data to zone totals and the use of aggregate models.

ANALYSIS OF AGGREGATE ZONAL DATA

The analysis of aggregate zonal data focuses on the estimation of total zonal travel by trip purpose. Specifically, the analysis deals with:

- accuracy in the estimation of the total trips for the zone by trip purpose from sample data.
- the variance of estimates of zonal trips by trip purpose from sampling.
- the implications of these results relative to the use of the RMS error in trip generation analysis

The results of this analysis provide substantial insight into the accuracy of home interview data used in aggregate trip generation analysis.

Since the relationships between sampling rate and expected percent error are applicable to the estimation of the total trips per zone as well as the mean trips per dwelling unit, the information in this chapter concerning the expected percent error relative to the estimation of the total trips is essentially identical to that in the preceding chapter relative to the mean trips per dwelling units. This information has been repeated for the sake of completeness as well as for the convenience of the reader.

Accuracy in Estimation of the Trips Per Zone

As will be recalled, the general statistical analysis demonstrated the practical applicability of the theoretical relationships between sample size and expected error and between sampling rate and expected percent error. The unique position of having the population data available allows the direct application of these formulas to quantitatively study these relationships for the three zones and the combined area. For convenience, the formulas expressing the relationship between sample rates and expected percent error ranges (i.e., Equations 7 and 8 from section entitled "Theoretical Constructs") were extensively used in this analysis. The use of these formulas offers two salient advantages:

- Practitioners are generally more accustomed to thinking in terms of percent samples and percent error.
- The quantitative results are applicable to the estimation of the total trips for the zone.

The formulas expressing the relationship between sample sizes and expected error ranges relative to the total trips (i.e., Equations 2 and 4 from the section entitled "Theoretical Constructs") were also used extensively to provide added perspective in the accuracy of sample data. The use of these formulas emphasizes the severity of the percent errors observed.

Sampling Rate Versus Percent Error Range

For each of the eight trip categories and each of the three zones and the combined area, the required sampling rates at the 95 percent probability level were computed for error ranges from ±5 percent to ±95 percent. Similar computations were performed for the required sampling rates at the 80 percent probability level. The results of these computations are illustrated graphically in Figures V-1 through V-4.

A number of general observations become apparent from a careful inspection and comparison of the relationships depicted graphically in these figures between sampling rates and their associated error ranges for the various trip categories and zones. These include:

- Rather large sampling rates are required to estimate the total trips by trip purpose for a zone (or the combined area) within reasonable ranges of error even at the 80 percent probability level.
- Increasing the sampling rate decreases the error ranges at a decreasing rate. For example, increasing the sampling rate from, say, five to ten percent results in a larger reduction in the error range than increasing the sampling rate from, say, 55 to 60 percent.
- As expected, increasing of zone size within a homogeneous area tends to reduce the error ranges associated with any sampling rate.
- The error ranges for the estimation of auto driver trips are generally less than those for the estimation of person trips. (Home-based nonwork trips in Zone A and nonhome-based trips in Zone B produced the only exceptions.)
- When estimating either person or auto driver trips, the four trip purposes would generally be ranked, based on the percent error (smallest to largest), as follows:

  home-based trips
  home-based work trips
  home-based nonwork trips
  nonhome-based trips

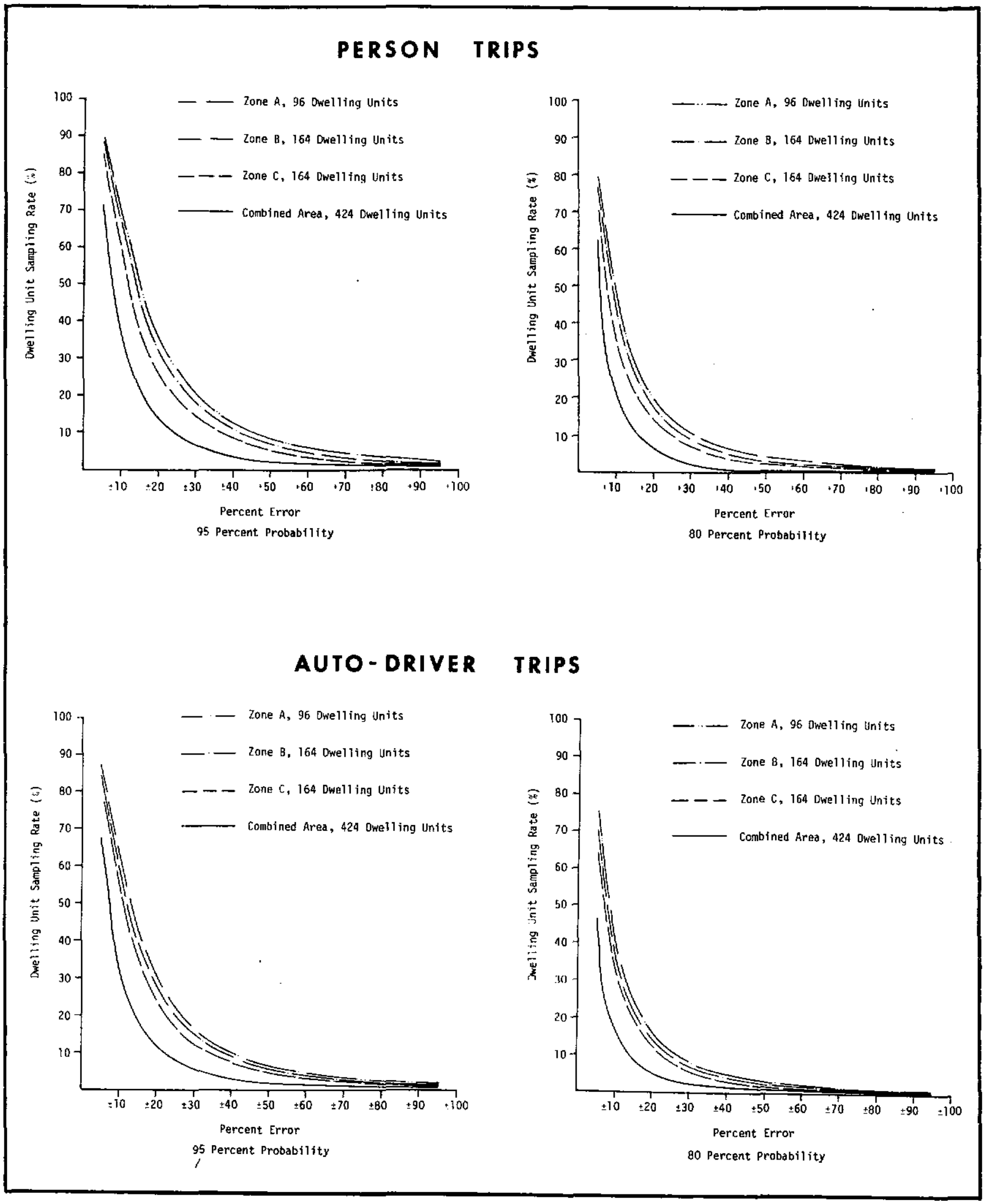

FIGURE V-1: MARGIN OF ERROR VERSUS SAMPLING RATE
FOR HOME-BASED TRIPS

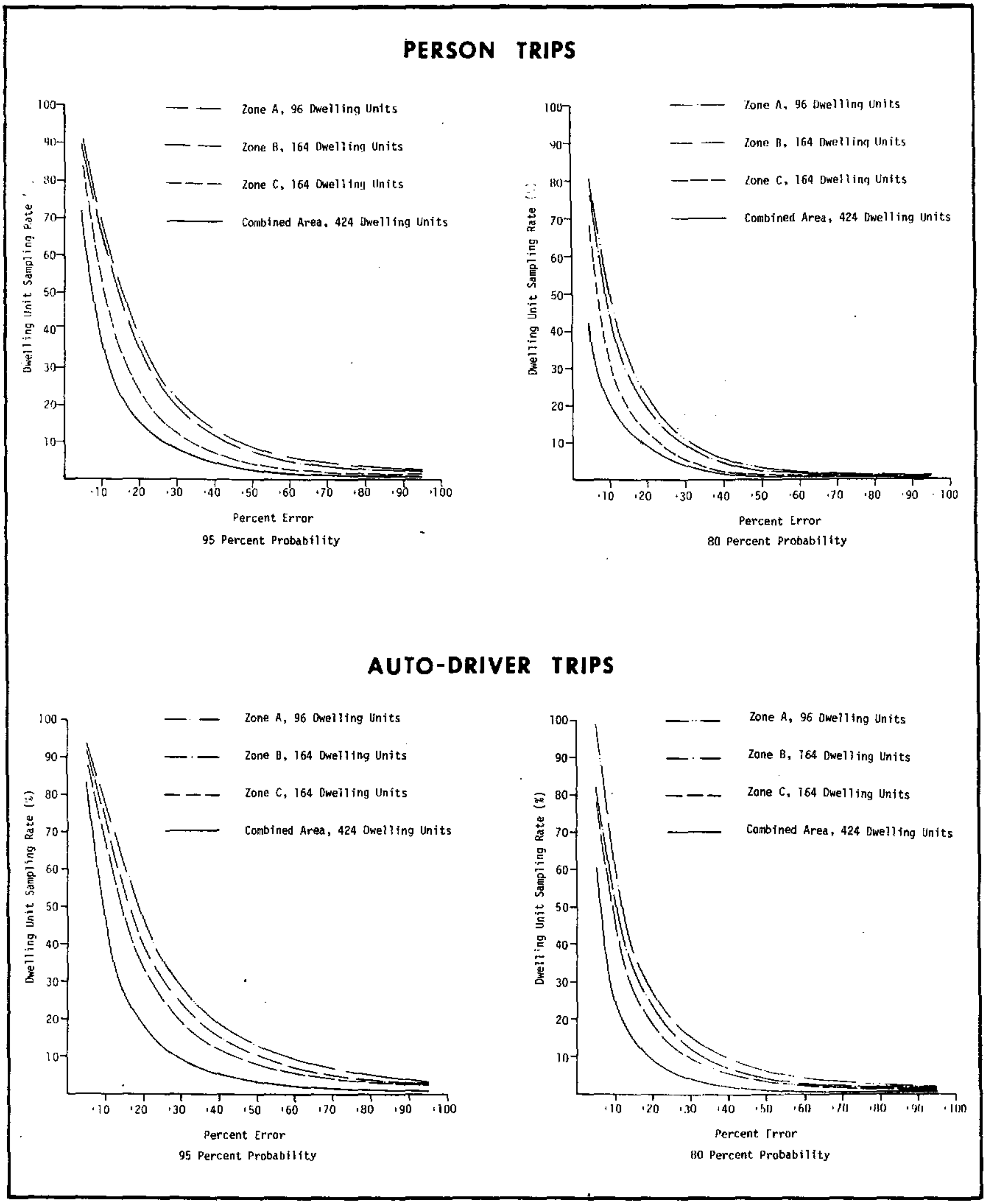

FIGURE V-2: MARGIN OF ERROR VERSUS SAMPLING RATE FOR HOME-BASED WORK TRIPS

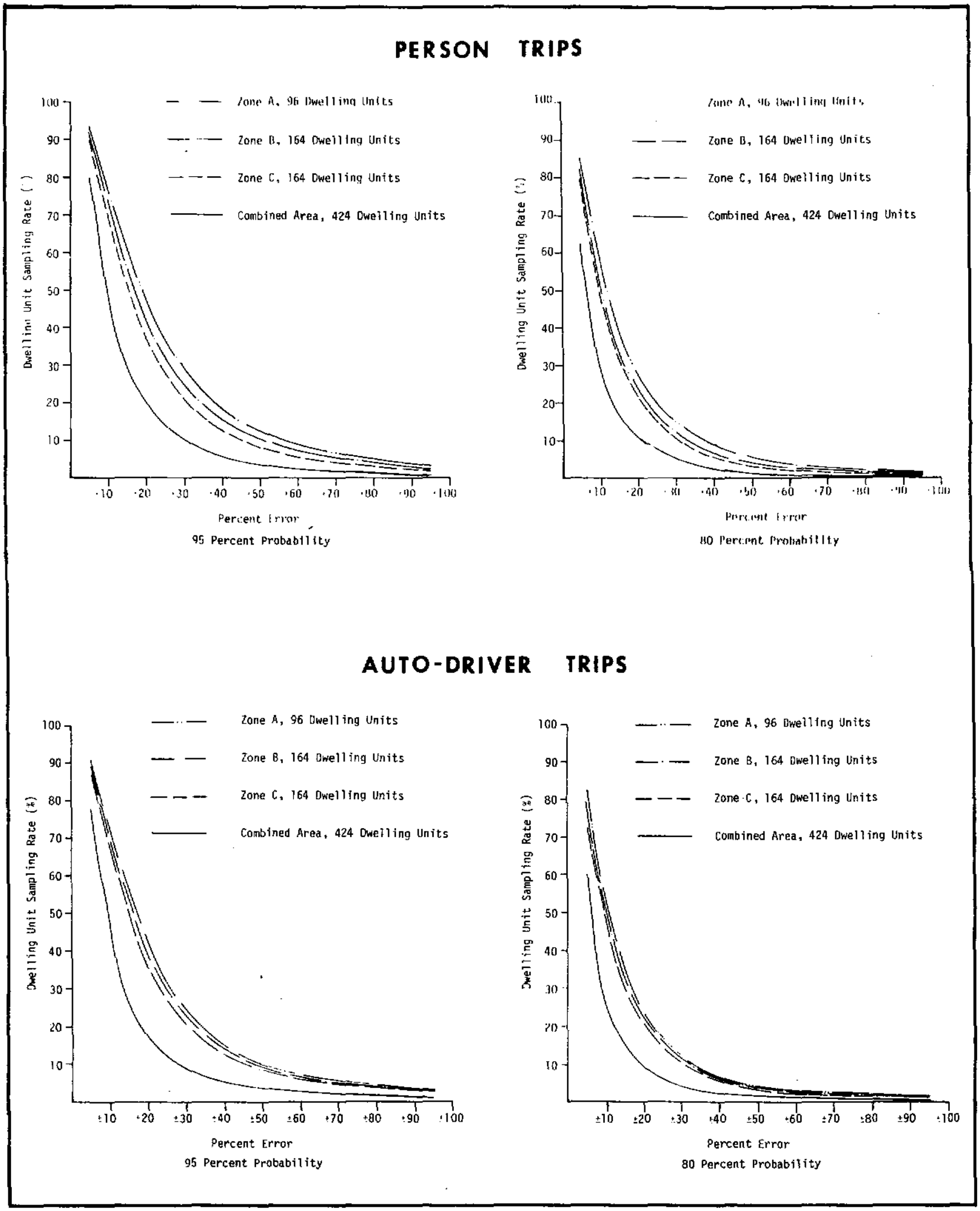

FIGURE V-3: MARGIN OF ERROR VERSUS SAMPLING RATE FOR FOR HOME-BASED NON-WORK TRIPS

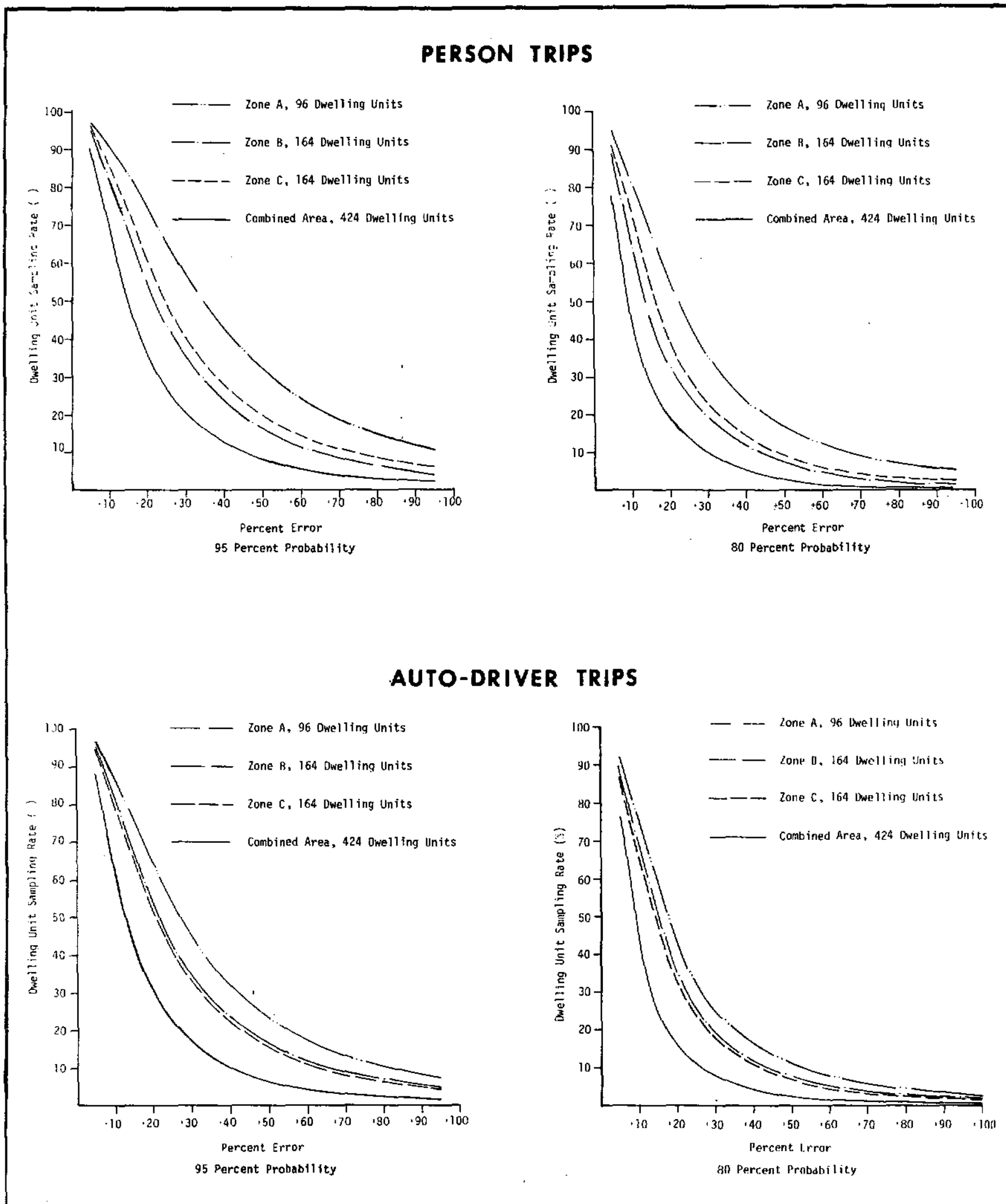

FIGURE V-4: MARGIN OF ERROR VERSUS SAMPLING RATE FOR NONHOME-BASED TRIPS

It must be remembered that the last two observations relate to the percent error rather than the absolute error. For example, while the expected percent error in estimating the number of home-based work trips is larger than for total home-based trips, the magnitude of the error range is smaller for home-based work trips than for total home-based trips.

<u>Sample Size Versus Expected Error Range</u>

Applying the formulas expressing the relationship between sample size and expected error range (i.e., Equations 2 and 4) provides useful perspective in the accuracy of sample data in estimating the trips per zone. These computations were limited to the eight trip categories for the combined area. The expected error ranges at the 95 percent probability level were computed for sample sizes from 10 dwelling units to 250 dwelling units (i.e., sampling rates from approximately 2.4 percent to approximately 60 percent). The estimates of the trips at the extremes of the error ranges (i.e., the actual of trips $\pm d$) were then computed. Similar computations are illustrated in Figures V-5 and V-6.

As in the analysis relative to percent sample versus percent error, the following observations become apparent from careful inspection and comparison of the results:

- Even at the 80 percent probability level, large sample sizes are required to estimate the trips for the combined area within a reasonable expected error range for any of the trip categories.
- Increasing the sample size decreases the expected error range at a decreasing rate.
- For any trip purposes, the magnitude of the error ranges for the estimation of auto driver trips is generally less than that for the estimation of person trips. A similar observation was made in terms of the percent error ranges; this indicates that, at any given sampling level, a higher degree of accuracy may be generally expected in the estimation of the auto driver trips than in the estimation of the person trips.
- When estimating either person or auto driver trips, the four trip purposes would generally be ranked, based on the magnitude of the error ranges (smallest to largest), as follows:

  home-based work trips
  nonhome-based trips
  home-based nonwork trips
  home-based trips

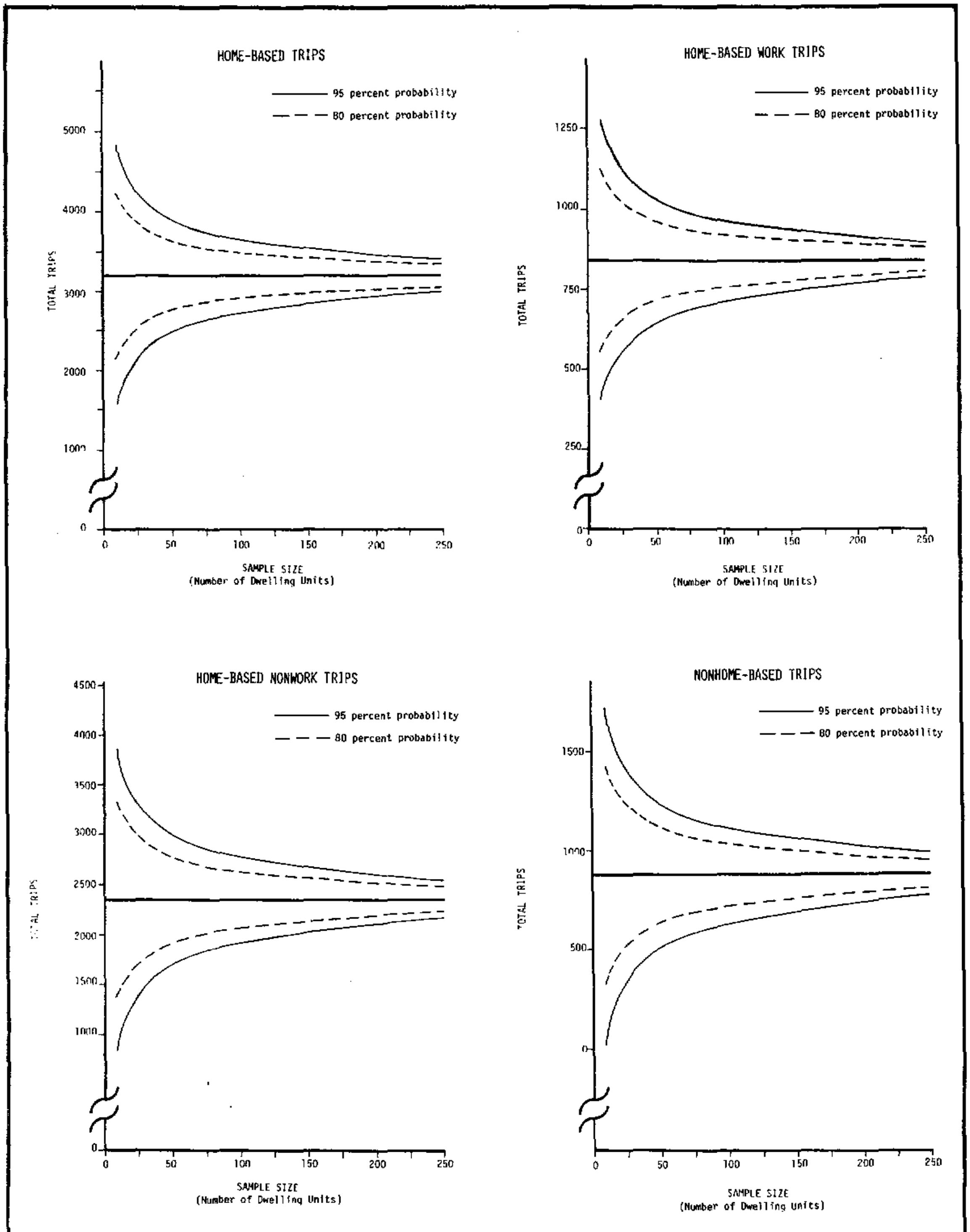

FIGURE V-5: EXPECTED ERROR RANGES IN ESTIMATING THE PERSON TRIPS FOR THE COMBINED AREA BY SAMPLE SIZE

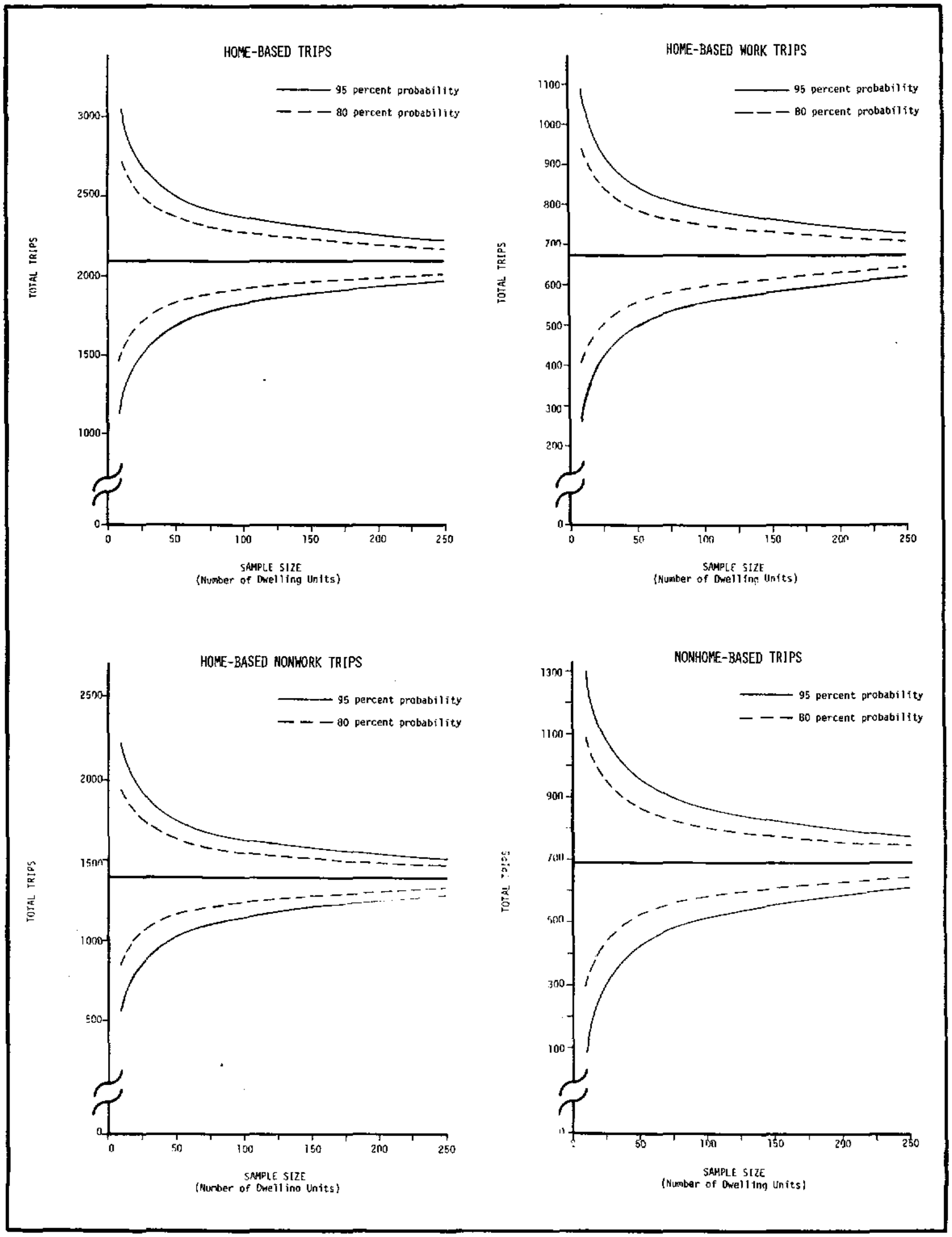

FIGURE V-6: EXPECTED ERROR RANGES IN ESTIMATING THE AUTO DRIVER TRIPS FOR THE COMBINED AREA BY SAMPLE SIZE

Again, it must be remembered that the last observation relates to the magnitude of the error range rather than the percent error range. Although the estimation of the home-based trips has the smallest expected percent error range of the four trip purposes, the magnitude of this error range (in terms of the number of trips) is the largest of the four trip purposes.

Traditional Sampling Rates

The traditional sampling rates normally used in the home interview surveys are[13]:

Population of area	Sample Size	Nominal sampling rate
under 50,000	1 in 5 dwelling units	20%
50,000 to 150,000	1 in 8 dwelling units	12.5%
150,000 to 300,000	1 in 10 dwelling units	10%
300,000 to 500,000	1 in 15 dwelling units	6.7%
500,000 to 1,000,000	1 in 20 dwelling units	5%
over 1,000,000	1 in 25 dwelling units	4%

The home interview origin-destination surveys performed in conjunction with the urban transportation studies in Texas have normally used either a nominal sampling rate of five percent (large urban areas) or 12.5 percent (small urban areas). Therefore, there was interest in further evaluation at these two sampling rates. The expected percent error ranges at the 80 and 95 percent probability level for five percent and 12.5 percent sampling rates for the eight trip categories are given in Tables V-1 and V-2. To demonstrate the severity of the magnitude of the error ranges, the estimates of the trips at the extremes of the error ranges for the various trip purposes were also computed for the five and 12.5 percent sampling rates at both the 80 and 95 percent probability levels and are summarized in Tables V-3 and V-4.

These tables indicate that if repeated samples were drawn from Zone A, for example, using a nominal five percent sampling rate, the number of home-based person trips may be expected to be estimated within a ±64.9 percent error range (i.e., between 298 and 1421 trips) by 95 percent of the samples. In other words, there is a 95 percent probability of drawing a five percent sample from Zone A which will estimate the number of home-based person trips is within ±64.9 percent of the true number (i.e., between 298 and 1421 trips). This, of course, was demonstrated in the general statistical analysis, whereby the results of approximately 95 percent of the sample did lie within the expected error range.

Increasing the sampling rate will, of course, reduce the error ranges. Com-

NOMINAL 5 PERCENT SAMPLING RATE				
Trip Category	Expected Percent Error Range			
	Zone A	Zone B	Zone C	Combined Area
Home-based Person Trips	±64.9%	±61.7%	±52.9%	±34.7%
Home-based Auto Driver Trips	±56.9%	±54.6%	±50.8%	±31.5%
Home-based Work Person Trips	±68.0%	±65.9%	±50.2%	±35.8%
Home-based Work Auto Driver Trips	±81.0%	±72.6%	±64.2%	±41.8%
Home-based Nonwork Person Trips	±80.9%	±76.6%	±69.3%	±43.7%
Home-based Nonwork Auto Driver Trips	±72.1%	±72.7%	±66.8%	±41.2%
Nonhome-based Person Trips	±145.8%	±97.6%	±109.5%	±65.8%
Nonhome-based Auto Driver Trips	±115.6%	±99.4%	±97.1%	±59.9%

NOMINAL 12.5 PERCENT SAMPLING RATE				
Trip Category	Expected Percent Error Range			
	Zone A	Zone B	Zone C	Combined Area
Home-based Person Trips	±40.2%	±37.0%	±31.7%	±20.9%
Home-based Auto Driver Trips	±35.3%	±32.7%	±30.5%	±19.0%
Home-based Work Person Trips	±42.1%	±39.5%	±30.1%	±21.6%
Home-based Work Auto Driver Trips	±50.2%	±43.5%	±38.5%	±25.2%
Home-based Nonwork Person Trips	±50.2%	±45.9%	±41.5%	±26.4%
Home-based Nonwork Auto Driver Trips	±44.7%	±43.5%	±40.0%	±24.8%
Nonhome-based Person Trips	±90.4%	±58.5%	±65.6%	±39.8%
Nonhome-based Auto Driver Trips	±71.7%	±59.6%	±58.2%	±36.2%

TABLE V-1: EXPECTED PERCENT ERROR RANGES AT THE 95 PERCENT PROBABILITY LEVEL FOR 5 AND 12.5 PERCENT SAMPLING RATES

NOMINAL 5 PERCENT SAMPLING RATE				
Trip Category	Expected Percent Error Range			
	Zone A	Zone B	Zone C	Combined Area
Home-based Person Trips	±43.3%	±39.6%	±34.0%	±22.5%
Home-based Auto Driver Trips	±38.1%	±35.3%	±32.6%	±20.6%
Home-based Work Person Trips	±45.6%	±42.2%	±32.2%	±23.3%
Home-based Work Auto Driver Trips	±54.1%	±47.0%	±41.4%	±27.1%
Home-based Nonwork Person Trips	±54.1%	±49.2%	±44.4%	±28.5%
Home-based Nonwork Auto Driver Trips	±47.8%	±47.0%	±43.1%	±26.8%
Nonhome-based Person Trips	±97.3%	±62.7%	±70.6%	±42.8%
Nonhome-based Auto Driver Trips	±77.4%	±64.0%	±62.7%	±38.8%

NOMINAL 12.5 PERCENT SAMPLING RATE				
Trip Category	Expected Percent Error Range			
	Zone A	Zone B	Zone C	Combined Area
Home-based Person Trips	±26.3%	±24.1%	±20.6%	±13.7%
Home-based Auto Driver Trips	±23.2%	±21.4%	±19.8%	±12.5%
Home-based Work Person Trips	±27.6%	±25.6%	±19.6%	±14.2%
Home-based Work Auto Driver Trips	±32.8%	±28.5%	±25.1%	±16.5%
Home-based Nonwork Person Trips	±32.8%	±29.9%	±27.0%	±17.3%
Home-based Nonwork Auto Driver Trips	±29.0%	±28.5%	±26.2%	±16.3%
Nonhome-based Person Trips	±59.1%	±38.1%	±42.8%	±25.0%
Nonhome-based Auto Driver Trips	±47.0%	±38.9%	±38.1%	±23.5%

TABLE V-2: EXPECTED PERCENT ERROR RANGES AT THE 80 PERCENT PROBABILITY LEVEL FOR 5 AND 12.5 PERCENT SAMPLING RATES

Nominal 5 Percent Sampling Rate

	Zone A	Zone B	Zone C	Combined Area
Home-based Person Trips	298 to 1421	459 to 1902	558 to 1804	2120 to 4325
Home-based Auto Driver Trips	240 to 883	344 to 1181	377 to 1148	1442 to 2756
Home-based Work Person Trips	67 to 346	115 to 558	148 to 459	551 to 1145
Home-based Work Auto Driver Trips	29 to 317	66 to 443	82 to 410	382 to 975
Home-based Nonwork Person Trips	125 to 1181	197 to 1476	262 to 1476	1357 to 3392
Home-based Nonwork Auto Driver Trips	106 to 672	131 to 869	180 to 869	848 to 1993
Nonhome-based Person Trips	0 to 528	16 to 672	0 to 689	297 to 1442
Nonhome-based Auto Driver Trips	0 to 317	0 to 541	16 to 552	297 to 1102

Nominal 12.5 Percent Sampling Rate

	Zone A	Zone B	Zone C	Combined Area
Home-based Person Trips	518 to 1210	738 to 1624	804 to 1558	2544 to 3901
Home-based Auto Driver Trips	365 to 758	508 to 1017	525 to 1000	1696 to 2502
Home-based Work Person Trips	115 to 228	213 to 476	213 to 394	678 to 1018
Home-based Work Auto Driver Trips	86 to 259	148 to 377	148 to 344	509 to 848
Home-based Nonwork Person Trips	326 to 979	508 to 1230	508 to 1230	1738 to 3010
Home-based Nonwork Auto Driver Trips	221 to 566	279 to 722	312 to 722	1060 to 1738
Nonhome-based Person Trips	19 to 403	148 to 541	115 to 541	551 to 1230
Nonhome-based Auto Driver Trips	38 to 250	115 to 443	115 to 426	424 to 933

TABLE V-3: SUMMARY OF ERROR RANGES IN TRIPS PER ZONE AT THE 95 PERCENT PROBABILITY LEVEL

Nominal 5 Percent Sampling Rate

	Zone A	Zone B	Zone C	Combined Area
Home-based Person Trips	490 to 1238	705 to 1656	787 to 1574	2502 to 3943
Home-based Auto Driver Trips	355 to 778	508 to 1033	529 to 1017	1654 to 2502
Home-based Work Person Trips	106 to 298	197 to 492	213 to 410	636 to 1060
Home-based Work Auto Driver Trips	77 to 269	148 to 377	148 to 344	509 to 890
Home-based Nonwork Person Trips	298 to 1008	426 to 1246	476 to 1263	1696 to 3053
Home-based Nonwork Auto Driver Trips	211 to 576	279 to 738	312 to 738	1018 to 1781
Nonhome-based Person Trips	0 to 422	131 to 558	98 to 558	509 to 1272
Nonhome-based Auto Driver Trips	29 to 259	98 to 459	98 to 426	424 to 975

Nominal 12.5 Percent Sampling Rate

	Zone A	Zone B	Zone C	Combined Area
Home-based Person Trips	634 to 1094	902 to 1460	935 to 1427	2798 to 3646
Home-based Auto Driver Trips	432 to 701	607 to 935	623 to 918	1823 to 2332
Home-based Work Person Trips	144 to 259	262 to 426	246 to 377	721 to 975
Home-based Work Auto Driver Trips	115 to 230	180 to 344	180 to 312	594 to 806
Home-based Nonwork Person Trips	442 to 864	590 to 1082	640 to 1099	1950 to 2798
Home-based Nonwork Auto Driver Trips	278 to 509	361 to 656	394 to 656	1187 to 1654
Nonhome-based Person Trips	86 to 336	213 to 476	180 to 476	636 to 1102
Nonhome-based Auto Driver Trips	77 to 211	180 to 377	164 to 361	509 to 848

TABLE V-4: SUMMARY OF ERROR RANGES IN TRIPS PER ZONE AT THE 80 PERCENT PROBABILITY LEVEL

COEFFICIENTS OF VARIATION

Trip Category	Zone A	Zone B	Zone C	Combined Area
Home-based Person Trips	0.76	0.91	0.78	0.83
Home-based Auto Driver Trips	0.67	0.81	0.75	0.76
Home-based Work Person Trips	0.80	0.97	0.74	0.86
Home-based Work Auto Driver Trips	0.95	1.08	0.95	1.00
Home-based Nonwork Person Trips	0.95	1.13	1.02	1.05
Home-based Nonwork Auto Driver Trips	0.84	1.08	0.99	0.99
Nonhome-based Person Trips	1.71	1.44	1.62	1.58
Nonhome-based Auto Driver Trips	1.36	1.47	1.44	1.43

STANDARD DEVIATIONS

Trip Category	Zone A	Zone B	Zone C	Combined Area
Home-based Person Trips	6.81	6.57	5.63	6.31
Home-based Auto Driver Trips	3.91	3.76	3.50	3.72
Home-based Work Person Trips	1.70	2.01	1.39	1.72
Home-based Work Auto Driver Trips	1.71	1.71	1.43	1.61
Home-based Nonwork Person Trips	6.47	5.82	5.45	5.86
Home-based Nonwork Auto Driver Trips	3.43	3.30	3.11	3.27
Nonhome-based Person Trips	3.79	2.96	3.25	3.27
Nonhome-based Auto Driver Trips	2.06	2.44	2.34	2.32

ZONE SIZES

Zone A = 96 occupied dwelling units

Zone B = 164 occupied dwelling units

Zone C = 164 occupied dwelling units

Combined Area = 424 occupied dwelling units

TABLE V-5: SUMMARY OF THE SIGNIFICANT PARAMETERS FROM THE 100 PERCENT DATA paring these error ranges for a nominal 12.5 percent sample with those for a nominal five percent sample, it can be seen that increasing the sampling rate by a factor of 2.5 results in the error ranges being reduced by a factor of roughly 0.6 Even with the 12.5 percent sample, the error ranges are very large.

Comparison of the error ranges in Tables V-1 and V-3 with those in Tables V-2 and V-4 shows that the expected error ranges at the 80 percent probability level are approximately 65 percent the magnitude of those at the 95 percent probability level. For example, there is an 80 percent probability of drawing a five percent sample from Zone A which will estimate that the home-based person trips are within $\pm$42.4 percent of the true number of trips, as compared with the $\pm$64.9 percent at the 95 percent probability level. However, it must be remembered that 80 percent probability level implies that, on the average, one out of every five samples may be expected to yield an estimate of the trips which lies outside the expected error range. At a 50 percent probability level, a nominal five percent sample from Zone A would yield a 50-50 chance of drawing a sample which would estimate the total number of home-based person trips in that zone within approximately $\pm$22 percent of the true number (i.e., between 671 and 1049), and, the number of non-home-based person trips would be estimated within approximately $\pm$50 percent of the true mean (i.e., between 107 and 320 trips).

It must be emphasized that, while lowering the probability level reduces the expected error ranges, it does not improve the accuracy of estimates from sample data. It simply increases the probability of drawing a sample which will yield an estimate outside the specified expected error range.

In view of the magnitude of these error ranges and the homogeneity and non-unique character of the area, it would seem reasonable to expect that an experienced urban transportation study analyst, through a careful inspection of an area, and using available data, could estimate the number of trips within such limits of accuracy.

This does not imply that the home interview surveys performed in the past have not been necessary. It is these surveys that provide the extensive travel data upon which the experienced analyst must be able to draw.

It must also be emphasized that this does not suggest that home interview surveys should be completely abandoned. However, it does indicate a new direction for home interview surveys. Instead of a general survey of the entire urban area, a limited number of home interviews might be used in order to monitor travel characteristics and update previous trip generation rates as well as to investigate areas which exhibit unique or unusual characteristics.

Significant Parameters

From Equation 8 it may be observed that the two significant parameters in the relationship between the sampling rate and the expected percent error range (at a given probability level) are the zone size (i.e., the number of occupied dwelling units in the zone) and the coefficient of variation. Similarly, from Equation 4 it may be observed that the two significant parameters determining the relationship between sample size and the expected error range (at a given probability level) are the zone size and the standard deviation of the distribution of dwelling units. The values of the parameters are again summarized in Table V-5 for convenience.

Coefficient of Variation

For a given zone, the differences between the percent error ranges for the different trip categories are attributable only to variations in the population means and standard deviations or, more specifically, in the coefficients of variation. Since the other terms in the formula for percent error (Equation 8) are constants for any given zone, it can be seen that the percent error varies directly with the coefficient of variation; therefore, it is a significant parameter in the relationship between the sampling rate and the percent error range. The sensitivity of the percent error range to changes in the coefficient of variation may be observed by comparison of the coefficients of variation in Table V-5 and the present error ranges in Tables V-1 and V-2. For the combined area for example, the coefficients of variation range from 0.76 to 1.58, while the associated percent error ranges at the 95 percent probability level vary from ± 31.5 percent to ± 65.8 percent.

Since the percent error range will vary directly with the coefficient of variation, one of the objectives in zonal delineation should be the minimization of the coefficient of variation. This might be attempted by defining zones such that the dwelling units within a zone exhibit, as nearly as possible, the same socioeconomic characteristics. In other words, zones should be defined so that each zone represents as homogeneous a group of dwelling units as is practicably possible.

Standard Deviation

For a given zone, the differences in the magnitudes of the error ranges (i.e., the variable d, in Equation 4) are attributable only to variations in the population standard deviation since the other terms in the formula are constants for any given zone. The sensitivity of the magnitude of the error range to the standard deviation may be observed by comparison of the standard deviations in Table V-5 and the magnitude of the error ranges in Tables V-3 and V-4. For example, the standard deviations for the combined area vary from 1.72 to 6.31 while the associated error ranges (for a five percent sample at the 95 percent probability level) vary from ±297 trips to ±1102 trips (i.e., from a range of 551 - 1145 trips to a range of 2120 - 4325 trips).

Since the magnitude of the error range will vary directly with the standard deviation, one of the objectives in zonal delineation should be the minimization of the standard deviation. As in the case of the coefficients of variation, this might be attempted by defining the zones such that the dwelling units within a zone exhibit similar socioeconomic characteristics.

Zone Size

The significant differences in the error ranges between the individual survey zone and the combined area are largely attributable to the effect of zone size. If the coefficient of variation and standard deviation are constant, the percent error range (Equation 8) will vary inversely with the square root of the zone size, and the magnitude of the error range (Equation 4) will vary directly with the square root of the zone size. In other words, assuming a constant coefficient of variation and standard deviation, the larger the zone size (number of occupied dwelling units), the smaller the percent error range for a given sampling rate, but the larger the magnitude of the error range.

It is important to note that in making these comments, the sampling rate is assumed to remain constant; thus, the sample size increases as the zone size increases. In other words, assuming the sampling rate remains constant and the standard deviation and coefficient of variation remain constant, if the zone size is increased by a factor of 4, the sample size will also be increased by a factor of 4, while the expected percent error range will be reduced by one-half, and the magnitude of the expected error range of the number of trips in the zone will double.

Since increasing the zone size, *ceteris paribus* will increase the expected magnitude of the error range, while at the same time reducing expected percent error, the objective of zonal delineation, relative to zone size, must be directed toward reducing the expected percent error rather than the magnitude of the error in estimating the total number of trips (by any trips purpose) using aggregate zonal data. Increasing the zone size will result in an increase in the variance; however, it will increase at a slower rate than the increase in the square root of the zone size. Hence, the general objective:

> should be to minimize the ratio of the coefficient of variation to the square root of the zone size (i.e., minimize $[C/\sqrt{N}]$) within the constraints imposed by the traffic assignment network.

As a practical matter, this suggests that zones should be defined to consist of as large a group of dwelling units exhibiting reasonably consistent socioeconomic characteristics as is possible, while remaining within the constraints of the traffic assignment network.

The level of detail of the traffic assignment network and the density of urban development generally provides an upper bound or limit on zone size. Consider, for example, a network representing corridors of movement corresponding to the freeways and major arterials within an urban area. If the arterials outside the CBD intersect at roughly one-mile spacings in developed areas, then the general limit on zone size within developed areas outside the CBD would be roughly one square mile. A residential area developed at approximately two dwelling units per acre (including streets, alleys, etc.) would result in a one square mile zone which consists of about 1300 dwelling units (or about three times the size of the combined area studied).

Variance of Estimates of Trips per Zone

The entire discussion of the accuracy of sample data in estimating the trips per zone is predicated on the fact that the distribution of the estimates of the total trips (as well as the mean trips per dwelling unit) from repeated random sample will approximate a normal distribution. Under the normality assumption, the variance of the estimates of the total trips at the various sampling levels provided the bases for the statistical inferences relative to the expected error ranges at those sampling levels.

The formulas for the expected error ranges in estimating the trips per zone (as discussed in the section entitled "Theoretical Constructs") not only assume that the distribution of the estimates will approximate a normal distribution, but that the variance of this distribution may be estimated by:

$$S^2 = \frac{N^2 \sigma^2}{\frac{Nn}{N - n}}$$

Where

S^2 = the expected variance of the estimates of the zonal trip ends from repeated random samples.

σ^2 = the population variance (i.e., the variance between dwelling units in trip productivity).

N = the number of occupied dwelling units in the zone.

n = the sample size.

The theoretical and observed variances for sample sizes up to 300 dwelling units in the combined area are displayed graphically in Figures V-7 and V-8 for the eight trip categories. It is to be noted that the "observed" points in these figures are estimates, since the true variance for a given sample size may be obtained only by drawing all possible samples of that size from the 100 percent data. The close correlation between the theoretical and observed data in these figures again demonstrates the applicability of the formulas used extensively in the preceding discussion of the accuracy of survey data in estimating the trips per zone.

Careful inspection of these figures indicate that increasing sample size decreases the variance of estimates but at a decreasing rate. That is, increasing the sample size from, say, 20 to 25 dwelling units results in a greater reduction in the variance of estimates than increasing the sample size from, say, 40 dwelling units to 45 dwelling units. If the cost per home interview is assumed to be relatively constant regardless of the sample size, then the cost of each additional observation (i.e., home interview) yields a smaller return on investment (in terms of the amount of reduction realized in the variance of estimates) than the preceding observation. It may be observed that, in each instance, the curve becomes almost a straight line with a small slope beyond the "knee" of the curve. Regardless of the trip category, this tendency of the curve to "level off" occurs at, or about, a sample size of 100 to 120 dwelling units. Increasing from a sample size of 80 to 120 dwelling units results in a greater reduction in variance than doubling the sample size from 120 to 240 dwelling units. It may, therefore, be argued that, for all trip categories, the reduction in the variance of estimates obtained by increasing the sample size beyond approximately 100 to 120 dwelling units is probably not sufficient to offset the increase in costs.

The variance of estimates by sample size (both theoretical and observed) of the home-based person trips for each of the three zones and the combined area is displayed graphically in Figure V-9. These curves tend to "level off" (i.e., reach an almost constant rate of decline) at, or about, sample sizes of 25 to 30 dwelling units from Zone A, 40 to 50 dwelling units for Zones B and C, and 100 to 120 dwelling units for the combined area, respectively. It is not a coincidence that these sample sizes all represent ranges of sampling rates of roughly 25 to 30 percent as may be demonstrated from the theoretical constructs. The formula for the variance of estimates of zonal trip ends may be stated in terms of sampling rate rather than sample size as follows:

$$S^2 = \sigma^2 N \frac{1 - P}{P}$$

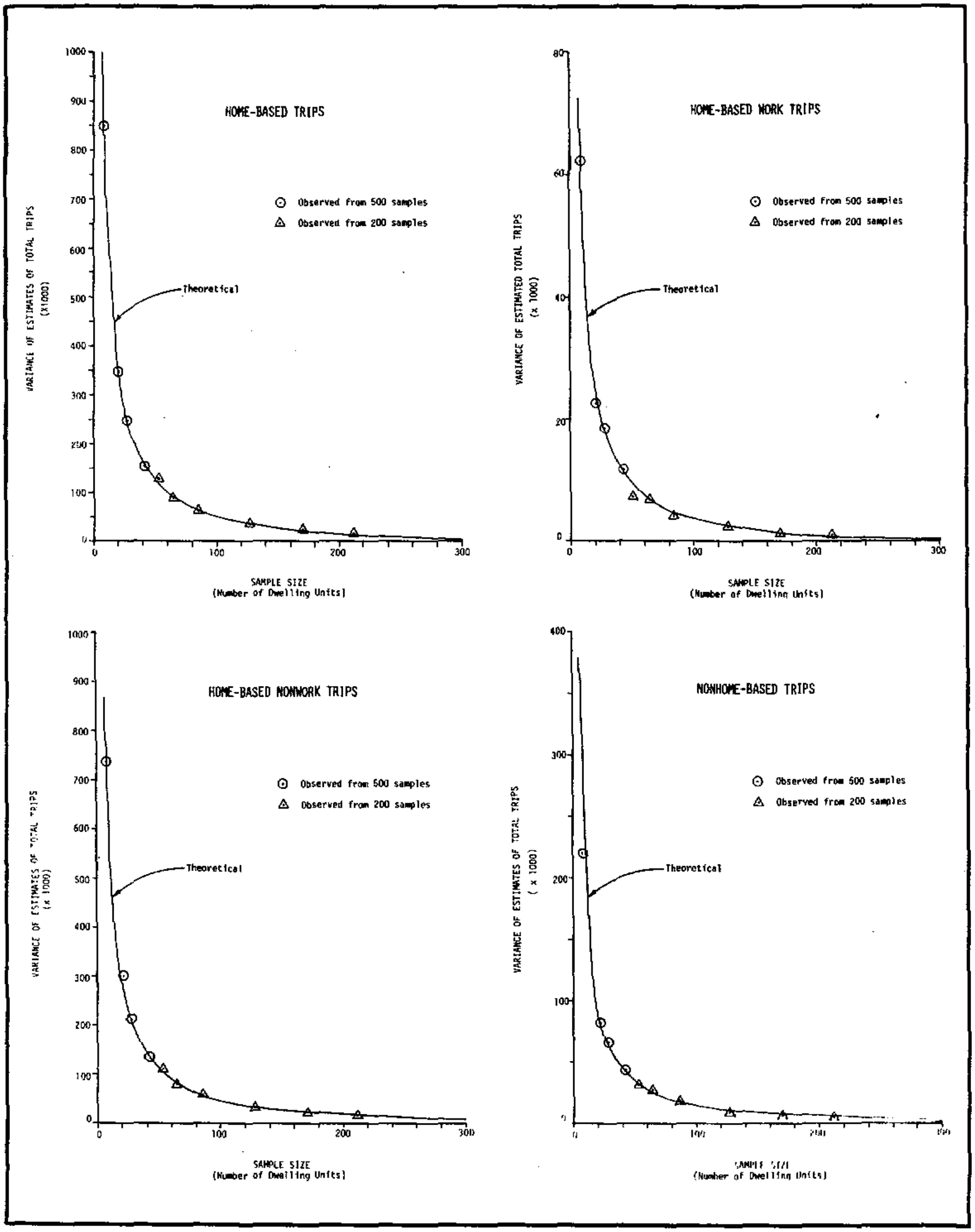

FIGURE V-7: VARIANCE OF ESTIMATES OF THE PERSON TRIPS BY SAMPLE SIZE FOR THE COMBINED AREA

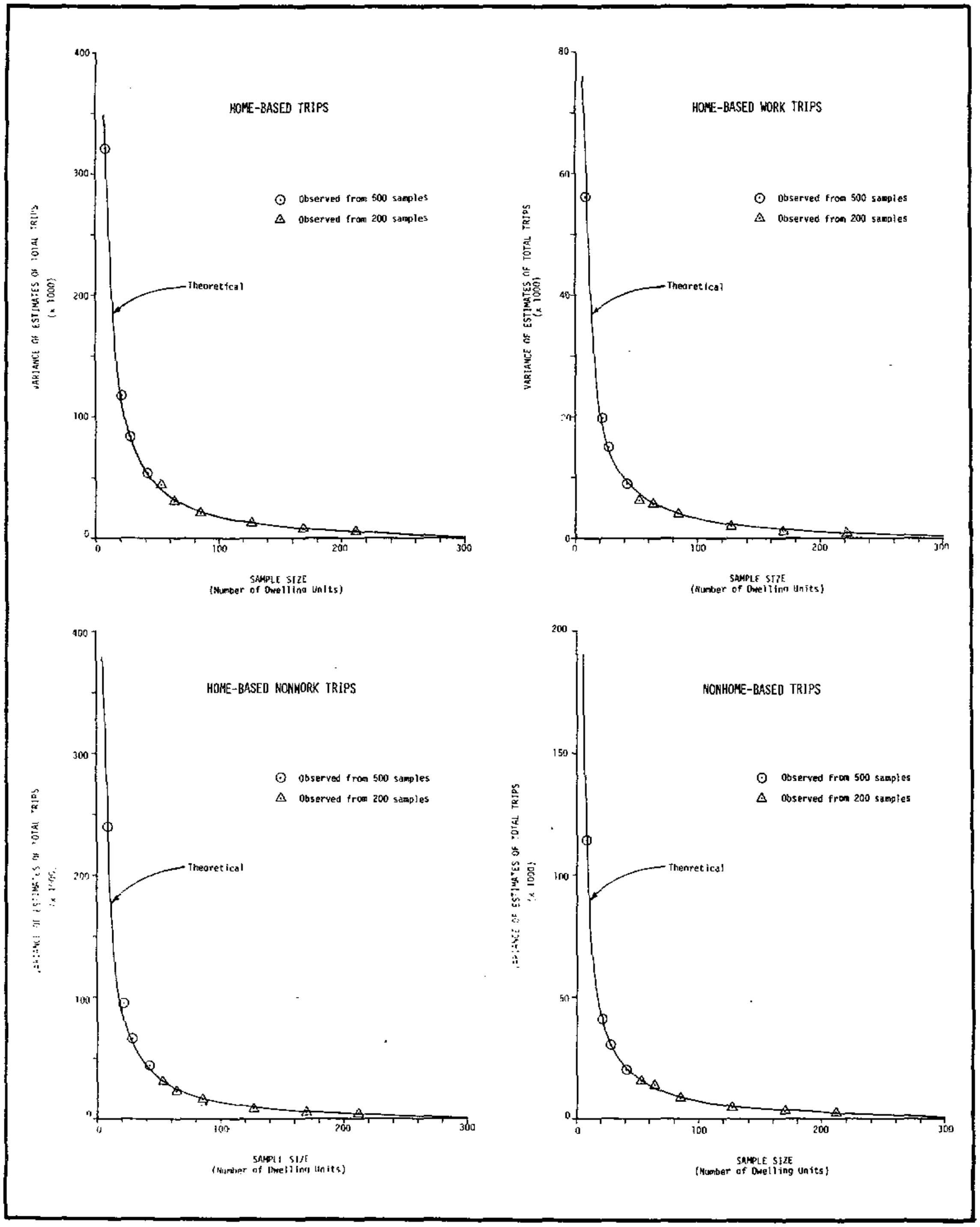

FIGURE V-8: VARIANCE OF ESTIMATES OF THE AUTO DRIVER TRIPS BY SAMPLE SIZE FOR THE COMBINED AREA

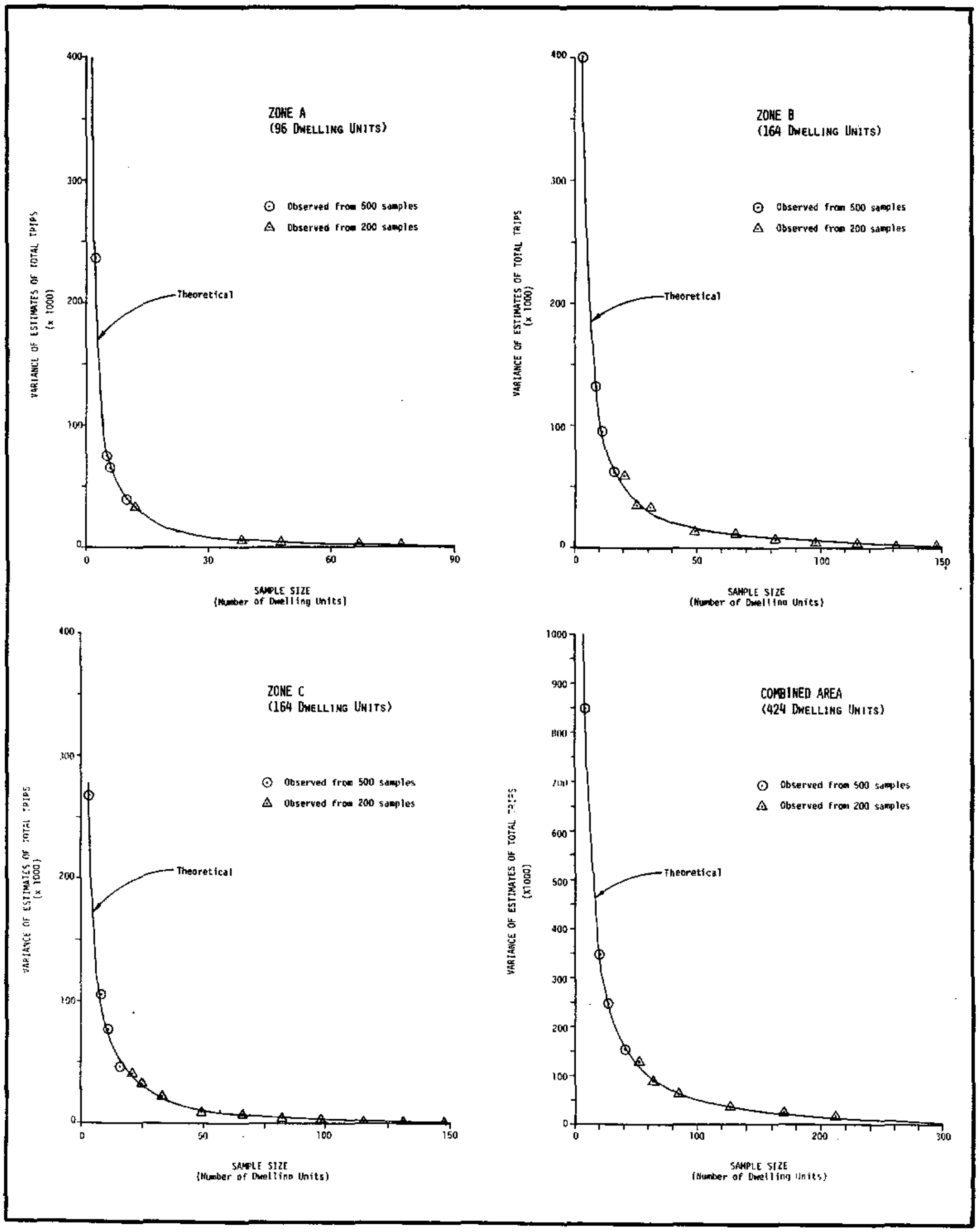

FIGURE V-9: VARIANCE OF ESTIMATES OF THE HOME-BASED PERSON TRIPS BY SAMPLE SIZE

Where

S^2 = the expected variance of the estimates of the zonal trip ends from repeated random samples.

σ^2 = the population variance (i.e., the variance between dwelling units in trip productivity).

N = the number of occupied dwelling units in the zone.

P = the percent sample expressed as a fraction

Since the value of N and σ^2 will remain constant for any given zone and trip category, the variance of estimates will vary directly with the function f(P) defined as follows:

$$f(P) = \frac{1 - P}{P} \qquad (0 < P \leq 1.0)$$

This relationship is depicted graphically in Figure V-10. Since the variance of estimates varies directly with f(P), this figure shows that beyond a sampling rate of roughly 30 percent, substantial increases in sampling rate would yield a very small reduction in the variance of estimates. Even increasing the sampling rate for roughly ten to 30 percent would yield only a modest reduction in the variance of estimates.

The analysis of the variance of estimates of the population variance indicate that the estimates of the population variance from traditionally used sampling roles are not of sufficient accuracy to yield reliable estimates of sample size requirements for estimating zonal aggregate trips.

Implications Regarding Trip Generation Analysis

It generally has been presumed that the major portion, if not all, of the difference between the observed number of trip ends in a zone and the number estimated from a regression model or trip generation rate was due to inadequacies in the regression model or cross-classification scheme. Technically stated, the independent variables did not account for all the variations in the dependent variable. The analysis of trip generation conducted under Study 2-8-63-60 (Research Report 60-12) found that the differences in trip generation rates in different zones could not be explained by socioeconomic-demographic variables. Interpretation of the 100 percent survey analyses suggests that much of the difference between "observed" and estimated trip ends in zonal aggregate totals is due to sampling error in the number of observed trips. The analyses demonstrated that distribution of sampling error is normally distributed with a mean of zero.

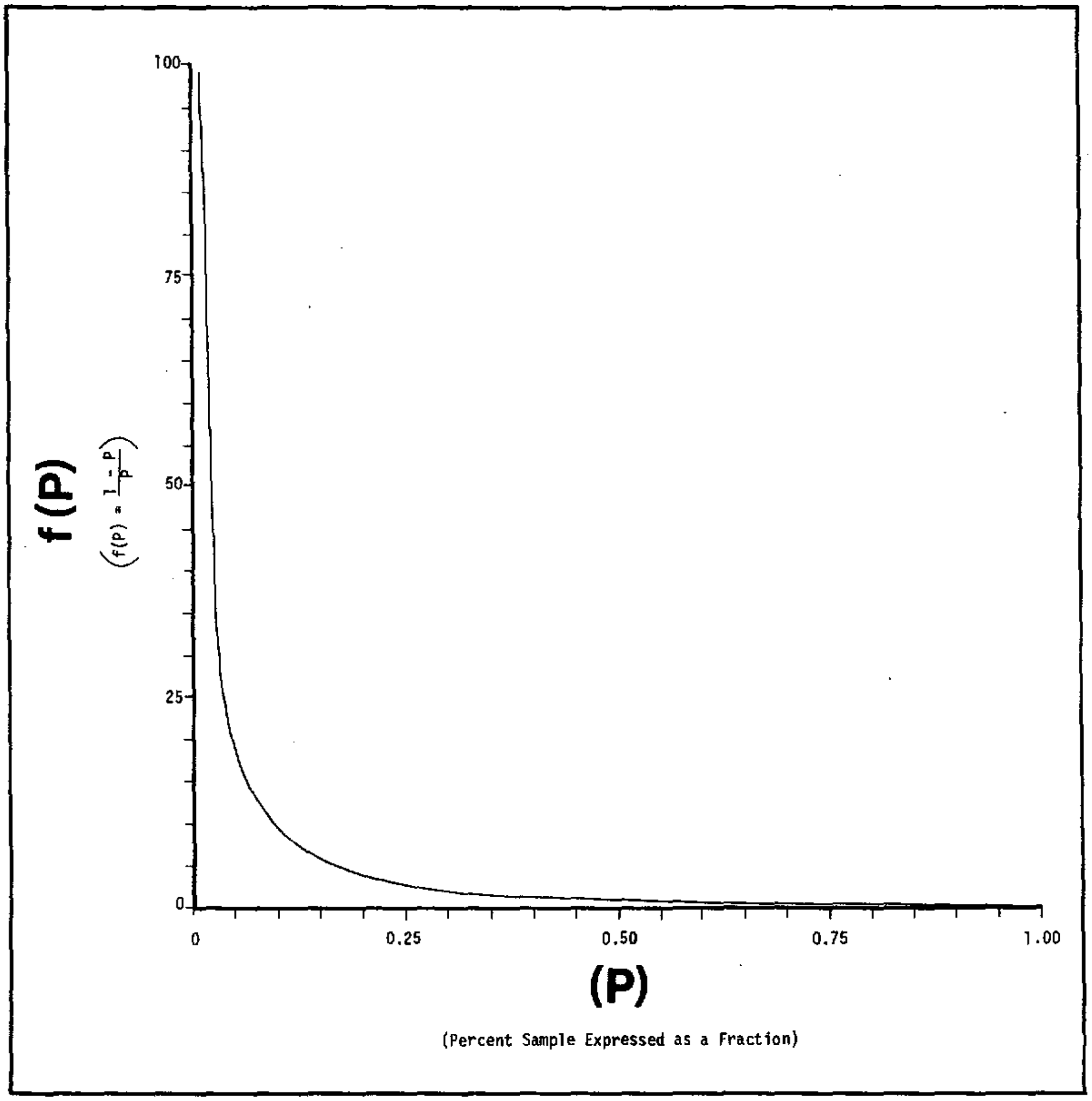
100
75
50
25
0
f(P)
$\left(f(P) = \frac{1-P}{P}\right)$
0
0.25
0.50
0.75
1.00
(P)
(Percent Sample Expressed as a Fraction)

FIGURE V-10

As a result, it may be concluded that, whereas the total number of trip ends in individual zones may be over or underestimated by the home interview O-D Survey, such errors are compensating when several zones having similar characteristics are involved. This suggests that the use of an estimating equation (regression model or cross-classification rates) will provide a better estimate of the total number of trip ends by zone than the O-D Survey directly.

The 100 percent survey analyses indicate that there is substantial "sampling noise" in the dependent variable (number of trips in a zone). The various measures of "goodness of fit" should interpret, in light of the precision with which O-D Survey data measures, the total number of trips in any zone.

The RMS error is commonly used in transportation studies and can be employed as a measure of the dispersion of the estimated number of trip ends from the observed number of trips in the several zones. However, the dispersion of the observed total number of trip ends from the actual trip ends can be very large at the feasible and generally used sampling rates. Hence, the "RMS error" of the sample from actual may be larger than the RMS error of estimated trips (using either regression or cross-classification analysis) from samples. When comparing the RMS errors for two or more sets of estimated trips, there is no guarantee that the set having the smallest calculated RMS error (estimated to observed) would produce the smallest true RMS error (estimated to actual). This would simply imply that the calculation and the use of RMS error for evaluating the adequacy of a set of estimated trips are of highly questionable value.

The results from the five and ten percent random samples drawn from each of the zones and the combined area may be used to demonstrate the potential problem with the use of the RMS error for evaluating the adequacy of a set of estimated trips. To avoid any potential confusion over terminology, the following definitions are presented:

Actual Trips - the number of home-based person trips for each zone based on the 100 percent data

Observed Trips - the number of observed home-based person trips for a zone at a given sampling rate was computed as follows:

$$\text{Observed Trips} = \bar{X} - t\sqrt{S^2}$$

Where

X = the actual number of home-based person trips for the zone.

S^2 = the variance of estimates of the home-based person trips for the zone observed from repeated random samples at the given sampling rate.

t = the number from the normal distribution table corresponding to the desired level of confidence.

Estimated Trips - the number of estimated home-based person trips for a zone at a given sampling rate was computed as follows:

$$\text{Estimated Trips} = N(\mu - t\sqrt{S^2})$$

Where

N = the number of occupied dwelling units in the zone.

μ = the mean home-based person trips per dwelling unit for the combined area.

S^2 = the variance of estimates of the mean home-based person trips for the combined area observed from repeated random samples at the given sampling rate.

t = the number from the normal distribution table corresponding to the desired level of confidence.

In this terminology then, "observed trips" correspond to those computed from the direct expansion of survey data, while the "estimated trips" correspond to those which would result from the use of a cross-classification based on the sample data collected from the combined area.

Using the variances observed from 1000 randomly drawn five percent samples from each zone and the combined area, the observed trips and the estimated trips were computed for each of the three zones for various levels of confidence. At each level of confidence the RMS error was computed for each of the following:

- Observed Trips Versus Actual Trips
- Estimated Trips Versus Actual Trips
- Estimated Trips Versus Observed Trips

The results of these computations are displayed in Figure V-11. Similar computations were performed using the variances observed from the 1000 randomly drawn ten percent samples drawn from each zone and the combined area and the results displayed in Figure V-12. The curves represent the maximum RMS error which may be expected at a given level of confidence. For example, using the five percent sampling rate, one may be 80 percent confident that the RMS error for the "Estimated Trips" versus the "Actual Trips" will be less than or equal to approximately 256 (note that 80 percent confidence corresponds to approximately 1.28σ). Referring to a normal table, it can be seen that one standard deviation (i.e., 1σ) corresponds to 68.3 percent confidence and two standard deviations (i.e., 2σ) correspond to 95.5 percent confidence.

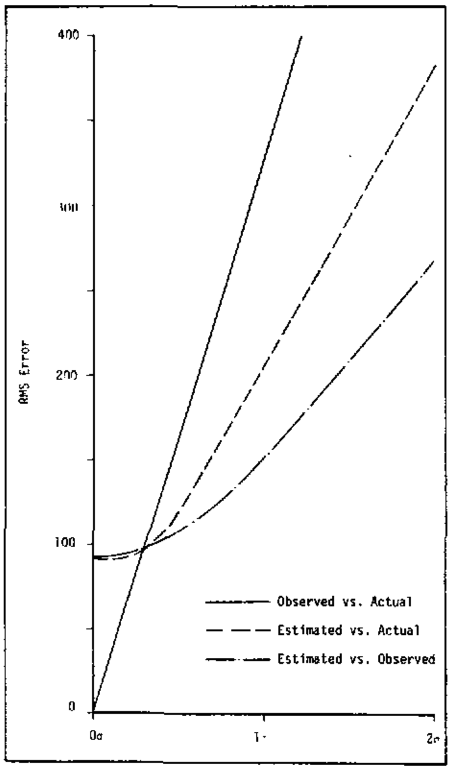

FIGURE V-11: RMS ERROR CURVES FOR FIVE PERCENT SAMPLES

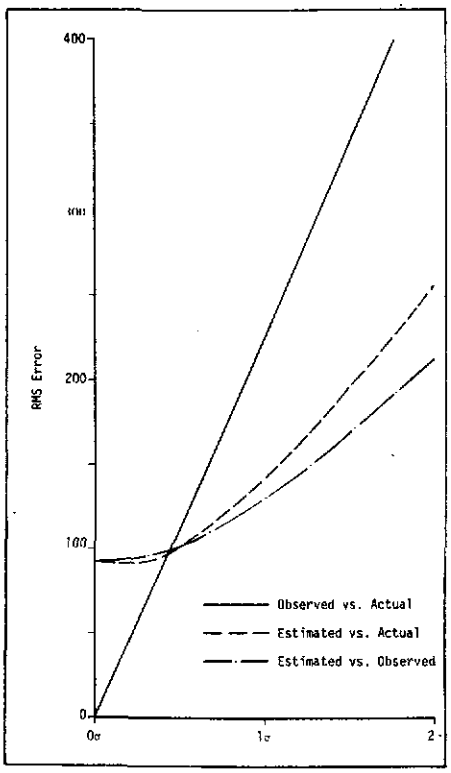

FIGURE V-12: RMS ERROR CURVES FOR TEN PERCENT SAMPLES

The general observations may be made from the curves in Figure V-11 and V-12 include:

- At the five percent sampling level, it may be observed that beyond approximately 0.35 (i.e., approximately 27 percent confidence) the curve for the RMS error of the "Observed Trips" versus the "Actual Trips" is consistently the highest and the curve for the RMS error of the "Estimated Trips" versus the "Observed Trips" is consistently the lowest. This suggests that, when using a five percent sampling rate in the combined area, the probability is approximately 73 percent that the RMS error of the observed trips versus the actual trips would exceed the RMS error for the estimated versus observed trips (the RMS error normally employed in transportation studies). Likewise, there is a 73 percent probability that the RMS error for the estimated versus actual trips will exceed the RMS error of the estimated versus observed trips. The vertical separation of the curves suggests that, when using a five percent sampling rate, the RMS error of the estimated versus observed trips would generally be substantially less than the RMS error of the observed versus actual trips.

- Comparing Figures V-11 and V-12, it may be observed that increasing the sampling rate would generally reduce the expected RMS error. Further, the curve for the RMS error of the "Estimated Trips" versus "Actual Trips" decreases faster with an increase in the sampling rate than the other curves. The curve for the RMS error of the "Estimated Trips" versus the "Observed Trips" appears the least responsive to increases in sampling rates. This suggests that at higher sampling rates (say above 40 percent) the RMS error of the estimated versus actual trips might generally be expected to be less than the RMS error of the estimated versus observed trips. It must be recognized, however, that this would only occur at sampling rates well beyond those normally used in urban transportation studies.

To nondimensionalize this analysis, the following ratios were computed for both the five and ten percent sampling rates:

- Ratio A = $\dfrac{\text{RMS error of "Observed Trips" versus "Actual Trips"}}{\text{RMS error of "Estimated Trips" versus "Observed Trips"}}$
- Ratio B = $\dfrac{\text{RMS error of "Observed Trips" versus "Actual Trips"}}{\text{RMS error of "Estimated Trips" versus "Actual Trips"}}$
- Ratio C = $\dfrac{\text{RMS error of "Estimated Trips" versus "Actual Trips"}}{\text{RMS error of "Estimated Trips" versus "Observed Trips"}}$

The results of these computations are illustrated graphically in Figures V-13 and V-14. Ratio C is, of course, simply the quotient of Ratio A divided by Ratio B. Comparing these figures, it can be seen that Ratio C is approaching a constant ratio of one as the sampling rate increases. Even at the ten percent sampling level, it may be observed that Ratio C is generally larger than one. This suggests that, for traditional sampling rates, the RMS error of the estimated versus observed trips may be generally expected to underestimate the RMS error of the estimated versus actual trips.

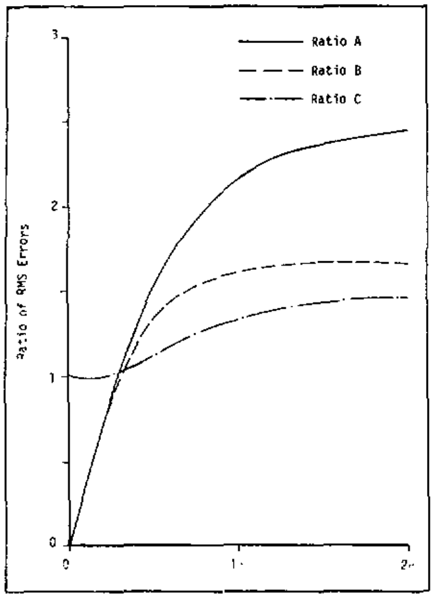

FIGURE V-13: RATIOS OF RMS ERRORS FOR FIVE PERCENT SAMPLES

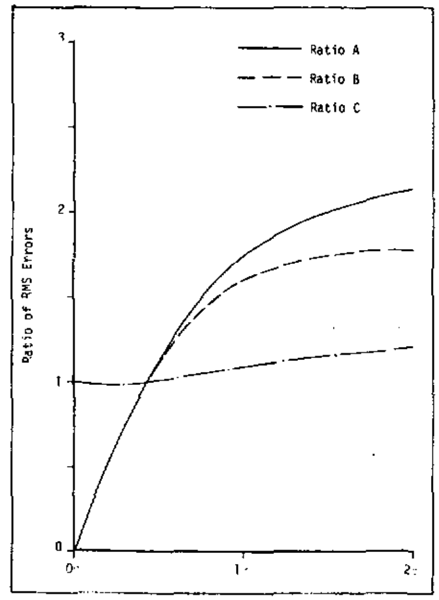

FIGURE V-14: RATIOS OF RMS ERRORS FOR TEN PERCENT SAMPLES

INTERPRETATION AND RECOMMENDATIONS

The analyses presented in this report have focused on the accuracy of home interview data in estimating the magnitude of disaggregate and aggregate zonal travel characteristics normally used in trip generation analyses. The results not only provide a new perspective for reviewing the study results which utilize the traditional O-D survey approach, but suggest a new study approach which would result in a significant cost reduction in the urban transportation study process.

While traditional approaches to trip generation analysis vary, all approaches utilize observed travel data (i.e., zonal travel data computed from home interview data obtained with a given zone) either directly or indirectly. Approaches at the zonal aggregate analysis utilize observed zonal travel data directly in the development of estimating equations (either regression models or cross-classification rates) and in measuring the adequacy of the estimating equations (i.e., comparisons of the estimated versus observed trips per zone).

Disaggregate trip generation approaches essentially discard the artificial zonal boundaries and utilize dwelling unit data directly in the development of estimating equations (either regression models or cross-classification rates). While these approaches do not use observed zonal travel data directly in the development of the estimating equations, they generally use the observed zonal travel data in measuring the adequacy of the estimating equations (i.e., comparisons of the estimated versus observed trips per zone).

The results of the analyses presented in this report demonstrate that, for small- to medium-sized zones, the observed zonal travel data computed from a traditional home interview survey (i.e., a survey employing traditional sampling rates) are subject to an extremely large variance of estimates. In other words, the magnitude of the sampling error which may be expected in computing observed zonal travel data from traditional survey data is disturbingly large at any reasonable level of probability. The results of these analyses also demonstrated the practical applicability of formulas which describe the relationships between sampling rates and expected percent error ranges and between sample sizes and the magnitude of the expected error ranges. From these relationships the following general observations may be made:

- When increasing the percent sample within a zone, the expected percent error range in computing either the observed trips for the zone or observed mean trips per dwelling unit for the zone will decrease at a decreasing rate. It may be generally stated that the expected percent error range for a zone (at a given level of probability) will vary directly with the function f(P) defined as follows:

$$f(P) = \sqrt{\frac{1 - P}{P}}$$

Where

p = the percent sample expressed as a fraction.

The analysis also suggests that beyond a sampling rate of approximately 30 percent, substantial increases in the sampling rate will result in relatively small changes in the expected percent error ranges. The analyses indicate that the increase in sampling rate required to obtain a reasonable level of accuracy for observed zonal travel data (at a reasonable level of confidence) is not feasible.

- At a given sampling rate, the expected percent error range (in computing either the observed trips for the zone or the observed mean trips per dwelling unit for the zone) will vary directly with the ratio of the coefficient of variations (C) of the distribution of dwelling units within the zone by trip productivity and the square root of the number of occupied dwelling units in the zone $(\sqrt{N})$. This suggests that the objective of zonal delineation should be the minimization of the ratio $C/\sqrt{N}$. The zone size, which would be required to achieve a reasonable level of accuracy for observed zonal travel data, would require a zone size which would be incompatible with most traffic assignment networks.

Interpretation of the 100 percent survey analysis suggests that much of the difference between "observed" and estimated trip ends is due to sampling error in the number of observed trips. The analyses indicate that distribution of sampling error is normally distributed with a mean of zero. As a result, it may be concluded that: whereas the total number of trip ends in individual zones may be over or underestimated by the home interview O-D Survey, such errors are compensating when several zones having similar characteristics are involved. This suggests that the use of an estimating equation (regression model of cross-classification rates) will provide a better estimate of the total number of trip ends by zone than the O-D Survey directly.

For zones of conventional size, an experienced urban transportation study analyst could, through inspection of the area, estimate the zonal trip ends with at least as good accuracy as that obtained from expanded home interview data using traditional sampling rates. In view of the accuracy of the home interview survey approach, it would appear that a synthetic study approach (a much more refined approach than the simple inspection of an area) may be utilized to obtain an even higher level of accuracy at an enormous reduction in cost.

Such a synthetic approach might utilize small, special purpose, home interview surveys and selected areas to update trip generation rates obtained from previous surveys and/or to study specific areas exhibiting unique or unusual characteristics. The cost of these limited special purpose surveys would be substantially less than the cost of performing the traditional home interview survey for the entire urban area.

REFERENCES

1. V.G. Stover and R.R. Roberts
"Analysis of Sample Size Requirements," Research Report No. 3
Rock Hill Area Transportation Study, December 1967

2. Arthur B. Sosslan and Glen E. Brukke
"Appraisal of Sample Size Based on Phoenix O-D Survey Data," Bulletin No. 253
Highway Research Board, 1960

3. M. D. Harmelink, G. C. Harper, and H. M. Edwards
"Trip Generation and Attraction Characteristics in Small Cities," Report No. RR121
Department of Highway, Ontario, November 1966

4. Edward J. Kannel
"A Disaggregate Analysis of Urban Travel Behavior," Final Report Project C-36-69I
Joint Highway Research Project, Purdue University, July 1972

5. Harold Kassoff and Harold D. Dentchman
"Trip Generation: A Critical Appraisal," Record No. 297
Highway Research Board, 1969

6. Gerald M. McCarthy
"Multiple-Regression Analysis of Household Trip Generation - A Critique," Record No. 297
Highway Research Board, 1969

7. Creighton, Hamburg, Planning Consultants
"Data Requirements for Metropolitan Transportation Planning", NCHRP Report No. 120
National Cooperative Highway Research Program, Highway Research Board, 1971

8. W. Y. Oi and Paul W. Shuldiner
An Analysis of Urban Travel Demands
Northwestern University Press, 1962

9. Paul W. Shuldiner
"Trip Generation and the Home," Bulletin No. 347
Highway Research Board, 1962

10. Vergil G. Stover and J. D. Benson
"Traffic Projection and Assignment"
Texas Transportation Institute, Research Report 119-3F, September 1971

11. William G. Cochran
Sampling Techniques
John Wiler and Sons, 1953

12. Bernard Ostle
Statistics in Research
Iowa University Press, 1964

13. "Manual of Procedures for Home Interview Traffic Study," Bureau of Public Roads, Department of Commerce, Revised Edition 1954 (Reprinted 1957).

14. O. K. Anderson
"Statistical Evaluation of Origin-Desgination Data" Thesis, Bureau of Highway Traffic, Yale University, May 1951