TECHNICAL REPORT STANDARD TITLE PAGE

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

4. Title and Subtitle

ACCURACY OF TRAVEL PATTERN ESTIMATES FROM THE HOME INTERVIEW SURVEY

5. Report Date
March 1974

6. Performing Organization Code

7. Author(s)

J. D. Benson, David F. Pearson, and Vergil G. Stover

8. Performing Organization Report No.
Research Report 167-8

9. Performing Organization Name and Address

Texas Transportation Institute
Texas A&M University
College Station, Texas 77843

10. Work Unit No.

11. Contract or Grant No.
Research Study 2-10-71-167

12. Sponsoring Agency Name and Address

Texas Highway Department
11th & Brazos
Austin, Texas 78701

13. Type of Report and Period Covered
Interim- September 1972
March 1974

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 patterns. The study is primarily 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 demonstrates the general conformance of observed travel characteristics with expected characteristics from urban travel theory. The 100 percent data were used as a data base from which sets of repeated random samples were drawn at various sampling rates. Comparison of the results from the sets of random samples with the actual population data demonstrates the levels of accuracy which may be expected in estimating zonal interactions, interchange volumes, and trip length frequency data. In addition, the entire San Antonio home interview survey was used as the data base to determine the sample size needed to adequately estimate the mean trip length for the urban area.

17. Key Words

Origin-Destimation Surveys, Urban Transportation Studies, Transportation Planning, Zonal Interactions, Zonal Interchange Volumes, Trip Length Frequency.

18. Distribution Statement

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	73	

Form DOT F 1700.7 (8-69)

ACCURACY OF TRAVEL PATTERN ESTIMATES FROM THE HOME INTERVIEW SURVEY

by

J. D. Benson
Assistant Research Planner

David F. Pearson
Research Assistant

and

Vergil G. Stover
Study Supervisor

Research Report 167-8

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

March 1974

TABLE OF CONTENTS

ABSTRACT

This report presents the results of a study of the accuracy of home interview survey data in estimating zonal travel patterns. The study is primarily 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 demonstrates the general conformance of observed travel characteristics with expected characteristics from urban travel theory. The 100 percent data were used as a data base from which sets of repeated random samples were drawn at various sampling rates. Comparison of the results from the sets of random samples with the actual population data demonstrates the levels of accuracy which may be expected in estimating zonal interactions, interchange volumes, and trip length frequency data. In addition, the entire San Antonio home interview survey was used as the data base to determine the sample size needed to adequately estimate the mean trip length for the urban area.

Keywords: Origin-Destination Surveys, Urban Transportation Studies, Transportation Planning, Zonal Interactions, Zonal Interchange Volumes, Trip Length Frequency.

SUMMARY

The purpose of this study was to investigate the accuracy of home interview origin-destination surveys in estimating zonal travel patterns. 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 analyses performed focused on zonal interactions, interchange volumes, and trip length frequency characteristics. Analyses of the basic data from the three zones demonstrate the general conformance of observed travel pattern characteristics with expected characteristics from urban travel theory. The 100 percent data were used as the data base from which sets of repeated random samples were drawn at various sampling rates. Analyses of these sets of samples indicated that very large variances of estimates of zonal travel pattern characteristics may be expected when using traditional sampling rates. This indicates that a relatively low level of accuracy would be associated with zonal travel pattern estimates.

Using the entire San Antonio home interview survey data as a data base, it was demonstrated that a reasonable level of accuracy may be attained in estimating the mean trip length for the urban area using a relatively small random sample from the urban area.

IMPLEMENTATION STATEMENT

The results of this study demonstrate that extremely large variances may be expected in estimating travel patterns at the zonal level. At the area-wide level, however, it was demonstrated that a reasonable estimate of the mean trip length for the urban area may be obtained from a comparatively small random sample of the urban area. A similar study using the San Antonio 100 percent data was directed toward the accuracy of zonal trip end estimates from the home interview survey (the results of the trip end study were reported in Research Report 167-7).

Based largely on the findings of this study and the trip end study, the Texas Highway Department has abandoned 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 already has resulted in a net savings to THD in excess of $1,000,000.

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, trip distribution models, mode split, etc.

The basic assumption underlying the bulk of the research and modeling of travel characteristics has been that the O-D survey provides a reliable measure of zonal trip ends and travel patterns. The purpose of this 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, the data collection costs would be completely out of reason. 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 zonal trip ends and travel patterns might be studied.

The results of the analyses associated with the accuracy of trip end estimates were reported in Research Report 167-7 entitled "Accuracy of Trip End Estimates from the Home Interview Survey." This report presents the results of the analyses regarding the accuracy of travel pattern estimates from the home interview survey. The analyses of travel patterns reported herein involve interaction analysis, interchange volume analysis, and trip length frequency analysis.

The Data Base

San Antonio, Texas, was selected as the site for the 100 percent data collection, primarily because it was a major metropolitan area in which an origin-destination study was to be performed. 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 comparison with the normal 5 percent survey.

The area selected for the 100 percent dwelling unit survey is located in the north central portion of San Antonio. The basis for the selection of the zones was their apparent homogeneity and nonunique characteristics. 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 to 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 of the area 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. The residential density of the three survey zones is about 7,500 persons per square mile, including a small park but excluding the commercial development located along the adjacent arterial streets.

The same data were collected in the three zones as were to be collected in the home interviews in the San Antonio - Bexar County Urban Transportation Study. A conscientious effort was made to collect the needed data from every occupied dwelling unit in the designated area. However, the San Antonio area has been subjected to numerous market surveys; furthermore, the community was plagued with a terrorist rapist which created a tense and apprehensive atmosphere. This undoubtedly contributed to a higher refusal rate than previously encountered by O-D surveys in Texas. The completed interviews, however, are sufficient to establish a set of population data whereby the accuracy of sample data may be evaluated. For the purpose of this analysis, the zones will be defined to consist of the following number of occupied dwelling units:

- Zone A = 96 dwelling units
- Zone B = 164 dwelling units
- Zone C = 164 dwelling units
- Combined Area = 424 dwelling units

Socioeconomic Characteristics

The general dwelling unit data collected in the survey provide an insight into the character of the area. The median annual household income (1969) for the area was in the $6,000 to $6,999 range and the mean annual household income was estimated to be approximately $7,400. Approximately 85% reported an annual household income in the $4,000 to $12,499 range, while 10% of the households reported incomes of less than $4,000, and 5% reported incomes in excess of $12,500.

Other general dwelling unit data collected in the survey are summarized in Table I-1. It is interesting to note that, despite the homogeneity in the socioeconomic image of the area, a fairly large variance can be observed in the dwelling unit data summarized in Table I-1. 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.

Travel Characteristics

Travel characteristics are typical of residential areas in Texas. Of the 4,134 person trips inventoried (excluding walk trips), 98.4% were made by private auto (68.7% were auto-driver trips and 30.6% were auto passenger trips).

Table I-2 summarizes the trip production characteristics observed for the three zones as well as the combined area. The observed travel characteristics from the 100% data such as interactions, interchange volumes, and trip length frequency are presented in the subsequent sections dealing with the analysis of these respective characteristics.

TABLE I-1: SUMMARY OF DWELLING UNIT CHARACTERISTICS

	TOTAL				MEAN				STANDARD DEVIATION				COEFFICIENT OF VARIATION			
Dwelling Unit Attribute	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
Number of residents	324	490	463	1277	3.38	2.99	2.82	3.01	1.58	1.67	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

Summarized in this table are the dwelling unit data collected in the three 100 percent survey zones.

TABLE I-2: TRIP PRODUCTION SUMMARY BY PURPOSE

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	50	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

The data presented summarize the observed zonal trip production by purpose for each of the three 100 percent data zones and the combined area. Also summarized are the characteristics of the distribution of dwelling units by trip productivity for each of the three zones and the combined area.

INTERACTION ANALYSIS

An interaction is defined as a production zone and an attraction zone which interchange one or more trips. Early trip distribution techniques, such as the FRATAR method, relied heavily on the observed interactions from origin-destination data. In urban transportation studies involving primarily small to medium-sized zones, it is reasonable to expect that a substantial portion of the interactions would not be detected by the survey. In such instances, the number of interactions would be underestimated. Since the total number of trips in the study area are estimated accurately, the interchange volume between zone pairs having trips would be overestimated.

The gravity model theory, on the other hand, essentially assumes that for any given production zone there is some probability of an interaction with every nonzero attraction zone. Using the gravity model theory, an "expected" number of trips will be computed for each nonzero production and attraction zone pair. In urban transportation studies involving primarily small- to medium-sized zones, the number of zone pairs often exceeds the number of trips being distributed and the gravity model tends to "spread" the trips between virtually all possible zone pairs such that a large portion of the zone pairs interchange only a fraction of a trip.

The same situation often is encountered at the zonal level when the number of nonzero attraction zones exceeds the production volume for a given zone. (In other words, there are more attraction zones competing for the trips produced by a zone than there are trips to be distributed.) Various rounding procedures have been instituted in the gravity model programs which, of course, reduce the number of interactions. Although the rounding procedures significantly improve the gravity model results, the resulting trip matrices may still tend to overestimate the number of interactions which might be expected on any given day and, thereby, tend to underestimate many of the higher volume interchanges between zone pairs.

The Texas Trip Distribution Model, developed by TTI in cooperation with the Texas Highway Department, handles this problem by limiting the number of zone pairs which are eligible to interchange trips. The approach taken in this model is to limit the number of attraction zones with which a given production zone is eligible to interact. In using the model, the analyst is asked to provide an estimate of the maximum number of attraction zones with which a zone of a given production volume may be expected to interact. In essence, the interaction constraint serves as a maximum bounding condition in the model. The subset of eligible attraction zones for a given production is selected on the basis of accessibility.

With the development of this modeling capability, interest has naturally begun to focus on the number of attraction zones with which a given production zone may be expected to interact. Indeed, zonal interaction is an element of travel pattern which is always estimated (either directly or indirectly) in the modeling process. The 100% data collected from the three zones in San Antonio has, therefore, been used to analyze the interaction characteristics of each of these three zones as well as the combined area and to investigate the characteristics of sampled interaction data.

Analysis of Basic Interaction Data

The population data for three zones in San Antonio provides a unique opportunity to study the basic interaction characteristics of these zones as well as the combined area (i.e., treating the three zones as a single zone).

Interactions and Trips

The relationship between interactions and trip productivity may be described as a probabilistic relationship. The probability that a new trip will produce an interaction is a conditional probability based on the number of interactions produced by the preceding trips. Thus, it may be stated that, as the trip production volume and interactions for a zone increase, there is a decreasing probabilit that an additional trip will produce an interaction. This suggests that, as the number of trips produced by a zone increases, the expected number of interactions will increase at a decreasing rate and will asymptotically approach the bounding condition on this relationship (i.e., the maximum number of interactions is obviousl the number of nonzero attraction zones, which is a bounding condition). This also suggests that, as the trip production volume from a zone increases, the average trips per interaction may also be expected to increase.

The relationship between trip production volume and interaction is dependent upon the zonal structure superimposed on the urban area and the distribution of attractions of various separations for the various production zones (or, indeed, the accessibility distribution for each production zone). In essence, the trip production and interaction relationship may be expected to vary from city to city, from zonal structure to zonal structure within a given city, and from trip purpose to trip purpose within a given city and given zonal structure.

A summary of interactions and trips by zone and trip purpose from the 100% data area is given in Table II-1. This table also summarizes the average trip per interaction (i.e., the trips produced by the zone divided by the number of interactions) and the interactions as a percent of the 1,569 zones in the network.

Since the three 100% survey zones are adjacent zones, it may be expected that they would have similar accessibility characteristics; therefore, careful inspection of Table II-1 leads to a number of interesting observations relative to their general conformance to expected characteristics. First of all, it may be generally observed that for any given trip purpose, the number of interactions increases as the trip productions increase. Second, for any given trip category, the average trips per interaction for the individual zones remained approximately equal but shows a substantial increase for the combined area. This suggests that the rate of increase in interactions over the range of production volumes represented by the individual zones would appear to be almost constant; however, the average number of trips per interaction apparently increases at a decreasing rate as the production volume approached that represented by the combined area. This does not contradict the previous hypothesis that, as a trip production volume increases, the interactions may be expected to increase at a decreasing rate. It simply suggests that, over the volume ranges represented by the individual zones, the decrease in this rate of increase is probably very small.

Comparison of Zones B and C for any given trip category provides another interesting observation. Both zones had the same or similar production volumes for any given trip category, which resulted in approximately the same number of inter-

TABLE II-1: SUMMARY OF INTERACTIONS AND TRIPS

Trip Purpose and Zone	Trips Produced	Number of Interactions	Trips Per Interaction	Interactions as a Percent of Attraction Zones
Home-based Person Trips				
Zone A	860	167	5.1	10.7
Zone B	1180	234	5.0	14.9
Zone C	1180	219	5.4	14.0
Combined Area	3220	399	8.0	25.6
Home-based Auto Driver Trips				
Zone A	563	159	3.5	10.1
Zone B	764	212	3.6	13.5
Zone C	763	205	3.7	13.1
Combined Area	2090	372	5.6	23.8
Home-based Work Auto-Driver Trips				
Zone A	173	70	2.5	4.5
Zone B	261	109	2.4	6.9
Zone C	247	99	2.5	6.3
Combined Area	681	216	3.1	13.9
Home-based Nonwork Auto Driver Trips				
Zone A	390	107	3.6	6.8
Zone B	503	133	3.8	8.5
Zone C	516	142	3.6	9.1
Combined Area	1409	253	5.5	16.3

Summarized are the observed trips and observed interactions by trip purpose for the three 100 percent data zones. Also summarized are the average trips per interaction and the percent of the attraction zones with which the zone interacted.

actions. It may be noted, however, that even when these two zones produced identical trip volumes, the resulting number of interactions was subject to some variation. Since the relationship between trips and interactions is a probabilistic relation, such a variation is not unexpected.

Differences between trip purposes provide still more interesting observations. In comparing home-base person trips with the home-base auto driver trips, it may be observed that very different production volumes resulted in a very similar number of interactions. This is expected, since person trips include both auto driver and auto passenger trips. Person trips consistently resulted in a few more interactions than did the auto driver trips. This suggests that some of the persons interviewed either traveled with auto drivers from other zones or were bus passengers. Comparison of home-based work auto driver trips with home-based nonwork auto driver trips indicates that the work trips exhibit a more dispersive travel pattern. For example, the 261 work trips produced by Zone B interacted with 109 attraction zones, whereas the 503 nonwork trips produced interacted with 133 attraction zones. This reinforces the hypothesis that interactions are more than just a function of volume; they are also a function of the zonal accessibility for a given trip category.

The limited number of observations from these three zones, of course, does not provide a sufficient base to prove the hypothesis advanced in relation to zonal interaction characteristics. However, their general conformance to the expected characteristics certainly strengthens the hypothesis advanced in relation to zonal interactions.

Spatial Characteristics

The population data for the three zones in San Antonio also provide a unique opportunity to observe the spatial character of interactions. The spatial separation between zones and urban transportation studies is normally measured in terms of either network travel time, or network travel time plus a terminal time. This analysis used the network travel time as the measure of spatial separation, since this is the measure normally used by the Texas Highway Department in its urban transportation studies.

Using the existing network for the San Antonio study area, trees were built and skimmed to obtain the needed spatial separation data (rounded to the nearest one minute). For convenience, the spatial separations for the combined area were estimated by simply averaging the spatial separations measured for the three individual zones.

The spatial separations of attraction zones relative to the three production zones varied from 1 to 64 minutes. The average spatial characteristics of attraction zones, interactions, and trips by purpose and production zone are summarized in Table II-2. The average separation for attraction zones was computed by simply dividing the sum of the spatial separations from the production zone to each zone in the network by the number of zones in the network. This measure, of course, does not vary from trip purpose to trip purpose.

The average separation for interactions was computed by summing the spatial separation for each attraction zone which interacted with the given production zone, and dividing this sum by the number of interactions. This provides a rough measure of the spatial dispersiveness of interactions. The average trip length was computed by summing the spatial separation associated with each trip, and dividing by the number of trips.

The data presented in Table II-2 provides a number of interesting observations relative to the spatial character of interactions. As might be expected, the average separations for all attraction zones are consistently larger than the average separations for interactions. This simply reflects the impedance to interzonal travel due to spatial separations between zones. Comparison of the average separation for interactions with the average trip length indicates that (except for home-based work auto driver trips) the average separation for interactions consistently exceeds the average trip length. This again reflects the impedance to travel due to spatial separation and the resulting tendency for generally larger interchange volumes at the shorter separations. In the case of the home-based work auto driver trips, however, the average trip length slightly exceeds the average separation for interactions. This was largely

TABLE II-3: SPATIAL DISTRIBUTIONS OF INTERACTIONS AND ZONES

HOME-BASED PERSON

SPATIAL SEPARATION INTERVAL (network minutes)	NUMBER OF ATTRACTION ZONES				OBSERVED INTERACTIONS				PERCENT OF ATTRACTION ZONES RECEIVING TRIPS			
	Zone A	Zone B	Zone C	Combined	Zone A	Zone B	Zone C	Combined	Zone A	Zone B	Zone C	Combined
1 - 5	34	34	37	33	21	24	23	24	61.8	70.6	62.2	72.7
6 - 10	178	150	141	155	51	62	60	106	28.7	41.3	42.6	68.4
11 - 15	427	421	395	420	61	97	81	165	14.3	23.0	20.5	39.3
16 - 20	306	302	303	299	15	31	32	63	4.9	10.3	10.6	21.1
21 - 25	241	261	280	260	15	14	16	30	6.2	5.4	5.7	11.5
26 - 30	147	155	148	155	2	6	5	9	1.4	3.9	3.4	5.8
31 - 35	105	104	97	105	1	0	2	1	1.0	0	2.1	1.0
36 - 40	86	90	98	90	1	0	0	1	1.2	0	0	1.1
41 & above	45	52	70	52	0	0	0	0	0	0	0	0
All	1569	1569	1569	1567	167	234	219	399	10.6	14.9	14.0	25.5

HOME-BASED AUTO DRIVER

SPATIAL SEPARATION INTERVAL (network minutes)	NUMBER OF ATTRACTION ZONES				OBSERVED INTERACTIONS				PERCENT OF ATTRACTION ZONES RECEIVING TRIPS			
	Zone A	Zone B	Zone C	Combined	Zone A	Zone B	Zone C	Combined	Zone A	Zone B	Zone C	Combined
1 - 5	34	34	37	33	21	23	23	24	61.8	67.6	62.2	72.7
6 - 10	178	150	141	155	50	57	57	102	28.1	38.0	40.4	65.8
11 - 15	427	421	395	420	56	87	72	150	13.1	20.7	18.2	35.7
16 - 20	306	302	303	299	14	28	31	58	4.6	9.3	10.2	19.4
21 - 25	241	261	280	260	14	11	16	27	5.8	4.2	5.7	10.4
26 - 30	147	155	148	155	2	6	4	9	1.4	3.9	2.7	5.8
31 - 35	105	104	97	105	1	0	2	1	1.0	0	2.1	1.0
36 - 40	86	90	98	90	1	0	0	1	1.2	0	0	1.1
41 & above	45	52	70	52	0	0	0	0	0	0	0	0
All	1569	1569	1569	1567	159	212	205	372	10.1	13.5	13.1	23.7

HOME-BASED WORK AUTO DRIVER

SPATIAL SEPARATION INTERVAL (network minutes)	NUMBER OF ATTRACTION ZONES				OBSERVED INTERACTIONS				PERCENT OF ATTRACTION ZONES RECEIVING TRIPS			
	Zone A	Zone B	Zone C	Combined	Zone A	Zone B	Zone C	Combined	Zone A	Zone B	Zone C	Combined
1 - 5	34	34	37	33	5	4	12	16	14.7	11.8	32.4	48.5
6 - 10	178	150	141	155	16	22	17	45	9.0	14.7	12.1	29.0
11 - 15	427	421	395	420	30	49	40	91	7.0	11.6	10.1	21.7
16 - 20	306	302	303	299	6	20	15	38	2.0	6.6	5.0	12.7
21 - 25	241	261	280	260	9	8	10	16	3.7	3.1	3.6	6.2
26 - 30	147	155	148	155	2	6	4	9	1.4	3.9	2.7	5.8
31 - 35	105	104	97	105	1	0	1	0	1.0	0	1.0	0
36 - 40	86	90	98	90	1	0	0	1	0	0	0	0
41 & above	45	52	70	52	0	0	0	0	0	0	0	0
All	1569	1569	1569	1567	70	109	99	216	4.5	6.9	6.3	13.8

HOME-BASED NONWORK AUTO DRIVER

SPATIAL SEPARATION INTERVAL (network minutes)	NUMBER OF ATTRACTION ZONES				OBSERVED INTERACTIONS				PERCENT OF ATTRACTION ZONES RECEIVING TRIPS			
	Zone A	Zone B	Zone C	Combined	Zone A	Zone B	Zone C	Combined	Zone A	Zone B	Zone C	Combined
1 - 5	34	34	37	33	21	22	22	24	61.8	64.7	59.5	72.7
6 - 10	178	150	141	155	40	45	50	90	22.5	30.0	35.5	58.1
11 - 15	427	421	395	420	32	48	41	93	7.5	11.4	10.4	22.1
16 - 20	306	302	303	299	8	11	18	27	2.6	3.6	5.9	9.0
21 - 25	241	261	280	260	6	5	9	16	2.5	1.9	3.2	6.2
26 - 30	147	155	148	155	0	2	1	2	0	1.3	0.7	1.3
31 - 35	105	104	97	105	0	0	1	1	0	0	1.0	1.0
36 - 40	86	90	98	90	0	0	0	0	0	0	0	0
41 & above	45	52	70	52	0	0	0	0	0	0	0	0
All	1569	1569	1569	1567	107	133	142	253	6.8	8.4	9.1	16.1

Summarized are the distributions of attraction zones and interactions by spatial separation for the three 100 percent data zones and the combined area. The impedance to travel due to spatial separation is reflected in the distributions of interactions and in the percent of the attraction zones in each separation interval which received trips.

TABLE II-4: SPATIAL DISTRIBUTIONS OF INTERACTIONS AND TRIPS

HOME-BASED PERSON

SPATIAL SEPARATION INTERVAL (network minutes)	OBSERVED TRIPS				OBSERVED INTERACTIONS				AVERAGE TRIPS PER INTERACTION			
	Zone A	Zone B	Zone C	Combined	Zone A	Zone B	Zone C	Combined	Zone A	Zone B	Zone C	Combined
1 - 5	304	320	368	893	21	24	23	24	14.5	13.5	16.0	37.2
6 - 10	239	398	356	1103	51	62	60	106	4.7	6.4	5.9	10.4
11 - 15	208	305	256	746	61	97	81	165	3.4	3.1	3.2	4.5
16 - 20	49	75	83	217	15	31	32	63	3.3	2.4	2.6	3.4
21 - 25	53	53	88	193	15	14	16	30	3.5	3.8	5.5	6.4
26 - 30	3	29	23	64	2	6	5	9	1.5	4.8	4.6	7.1
31 - 35	2	0	6	2	1	0	2	1	2.0	0	3.0	2.0
36 - 40	2	0	0	2	1	0	0	1	2.0	0	0	2.0
41 & above	0	0	0	0	0	0	0	0	0	0	0	0
All	860	1180	1180	3220	167	234	219	399	5.1	5.0	5.4	8.0

HOME-BASED AUTO DRIVER

SPATIAL SEPARATION INTERVAL (network minutes	OBSERVED TRIPS				OBSERVED INTERACTIONS				AVERAGE TRIPS PER INTERACTION			
	Zone A	Zone B	Zone C	Combined	Zone A	Zone B	Zone C	Combined	Zone A	Zone B	Zone C	Combined
1 - 5	176	193	234	562	21	23	23	24	8.4	8.4	10.2	23.4
6 - 10	164	238	206	656	50	57	57	102	3.3	4.2	3.6	6.4
11 - 15	137	209	171	499	56	87	72	150	2.4	2.4	2.4	3.3
16 - 20	33	59	63	165	14	28	31	58	2.4	2.1	2.0	2.8
21 - 25	46	38	70	152	14	11	16	27	3.3	3.4	4.4	5.6
26 - 30	3	27	13	52	2	6	4	9	1.5	4.5	3.2	5.8
31 - 35	2	0	6	2	1	0	2	1	2.0	0	3.0	2.0
36 - 40	2	0	0	2	1	0	0	1	2.0	0	0	2.0
41 & above	0	0	0	0	0	0	0	0	0	0	0	0
All	563	764	763	2090	159	212	205	372	3.5	3.6	3.7	5.6

HOME-BASED WORK AUTO DRIVER

SPATIAL SEPARATION INTERVAL (network minutes)	OBSERVED TRIPS				OBSERVED INTERACTIONS				AVERAGE TRIPS PER INTERACTION			
	Zone A	Zone B	Zone C	Combined	Zone A	Zone B	Zone C	Combined	Zone A	Zone B	Zone C	Combined
1 - 5	11	7	28	45	5	4	12	16	2.2	1.8	2.3	2.8
6 - 10	34	51	39	124	16	22	17	45	2.1	2.3	2.3	2.8
11 - 15	72	111	82	257	30	49	40	91	2.4	2.3	2.1	2.8
16 - 20	11	39	34	94	6	20	15	38	1.8	1.9	2.3	2.5
21 - 25	38	28	51	114	9	8	10	16	4.2	3.5	5.1	7.1
26 - 30	3	25	9	45	2	6	4	9	1.5	4.2	2.3	5.0
31 - 35	2	0	4	0	1	0	1	0	2.0	0	4.0	0
36 - 40	2	0	0	2	1	0	0	1	2.0	0	0	2.0
41 & above	0	0	0	0	0	0	0	0	0	0	0	0
All	173	261	247	681	70	109	99	216	2.5	2.4	2.5	3.2

HOME-BASED NONWORK AUTO DRIVER

SPATIAL SEPARATION INTERVAL (network minutes)	OBSERVED TRIPS				OBSERVED INTERACTIONS				AVERAGE TRIPS PER INTERACTION			
	Zone A	Zone B	Zone C	Combined	Zone A	Zone B	Zone C	Combined	Zone A	Zone B	Zone C	Combined
1 - 5	165	186	206	517	21	22	22	24	7.9	8.5	9.4	21.5
6 - 10	130	187	167	532	40	45	50	90	3.3	4.2	3.3	5.9
11 - 15	65	98	89	242	32	48	41	93	2.0	2.0	2.2	2.6
16 - 20	22	20	29	71	8	11	18	27	2.8	1.8	1.6	2.6
21 - 25	8	10	19	38	6	5	9	16	1.3	2.0	2.1	2.4
26 - 30	0	2	4	7	0	2	1	2	0	1.0	4.0	1.5
31 - 35	0	0	2	2	0	0	1	1	0	0	2.0	2.0
36 - 40	0	0	0	0	0	0	0	0	0	0	0	0
41 & above	0	0	0	0	0	0	0	0	0	0	0	0
All	390	503	516	1409	107	133	142	253	3.6	3.8	3.6	5.6

Summarized are the distributions of interactions and trips by spatial separation for the 100 percent data area. The impedance to travel due to spatial separation is apparent in these data.

Most of the zones in the CBD lie in the separation interval 11-15 minutes, which accounts for a substantial portion of the observed interaction and trips in this interval. Two major regional shopping centers lie in a 6-8-minute interval relative to the three zones. Military installations, major employment centers for the San Antonio area, lie in the following intervals:

- 16-20 minutes -- Fort Sam Houston.
- 21-25 minutes -- Kelly Air Force Base.
- 26-30 minutes -- Randolph Air Force Base and Lackland Air Force Base.

In addition, there are several community shopping centers and substantial strip development in both the 1- to 5-minute and 6- to 10-minute intervals relative to the three zones. While spatial separation exerts a tremendous influence on travel patterns, it is well recognized in trip distribution theory that spatial separation alone is not sufficient to explain this observed travel pattern. This is probably most apparent if one considers the home-based work auto driver trips in the 21- to 25-minute interval. Sixty-eight of the 114 observed trips in the 21- to 25-minute interval were attracted to a single zone (i.e., Kelly AFB). The impact of this is most apparent in the average number of trips per interaction which is larger for the 21- to 25-minute interval than for any other interval.

In summary, the population data from the three zones in San Antonio demonstrate the impedance to interzonal travel due to spatial separation between zones. This impedance is not only reflected in the trip length frequency, but is also reflected in the spatial distribution of interactions. As generally accepted trip distribution theory would suggest, the spatial distribution of interactions is not simply a function of spatial separation, but is more precisely a function of spatial separation and the distribution of attractions. Such a conclusion is certainly consistent with the data observed.

Because interactions essentially ignore the magnitude of the interchange volume involved (so long as the volume is greater than or equal to one), the measures of the spatial characteristics of interactions will generally differ from the measures of the spatial characteristic of trips. A comparison of the spatial characteristics of interactions with trip production volume makes the salient difference in the spatial characteristic of trips and interactions apparent. As the production volume for a zone becomes sufficiently large such that the number of interactions approaches the number of nonzero attraction zones, the average separation for interactions will approach the average separation of attraction zones. The average trip length, on the other hand, will not necessarily change as production volume increases.

Interaction Characteristics of Dwelling Units

Cross-classification of dwelling units by trip productivity and the number of attraction zones with which the dwelling unit interacted gives an indication of the dispersiveness of travel at the dwelling unit level. The cross-classification of the dwelling units in the combined area by trip productivity and dwelling unit interactions is presented in Table II-5. The tendency toward "paired" trips generally observed in O-D surveys is, of course, apparent in the table. The following two observations, relative to the analysis of interactions, may be made from this table and similar cross-classifications of dwelling units for other trip purposes:

1. As the trip productivity of a dwelling unit increases, the expected number of interactions for that dwelling unit increases but at a decreasing rate. This tendency may be observed in the average number of interactions per trip for the various production volumes.
2. The basic relationships between dwelling unit interactions and zonal interactions are:
   - The sum of the dwelling unit interactions will be greater than the number of zonal interactions (except in the rare special case where no two dwelling units in a zone interact with a common attraction zone). For example, the sum of the dwelling unit interactions for the home-based person trips in the combined area is 1326 which is substantially greater than the 399 interactions observed for the zone.
   - The set of interactions for the zone is equal to the union of the sets of dwelling unit interactions.

These basic relationships are important considerations relative to the application of sampling theory and will be discussed in greater detail under the section entitled "Interactions and Sampling Theory."

TABLE II-5: DWELLING UNIT INTERACTIONS

TRIP PRODUCTION	DWELLING UNIT INTERACTIONS												ROW TOTALS	TRIPS	AVERAGE NUMBER OF INTERACTIONS PER TRIP
	0	1	2	3	4	5	6	7	8	9	10	11			
0	28												28	0	-
1 - 2		52	17										69	136	0.63
3 - 4		11	53	20	3								87	344	0.55
5 - 6		1	23	18	7	3							52	310	0.46
7 - 8			9	17	17	6	2						51	403	0.44
9 - 10			2	8	15	13	6	1					45	446	0.44
11 - 12			2	3	10	6	3		2				26	309	0.38
13 - 14				2	1	5	3	3		1			15	206	0.40
15 - 16		1			2	6	3	1	2				15	239	0.33
17 - 18					1	3	3	1	1			2	11	196	0.38
19 - 20							1		1	2			4	78	0.41
21 - 25					1	1	1	3	1		1	1	9	204	0.32
26 - 30						1	4		1	1		1	8	220	0.26
31 - 35					1	1		1	1				4	129	0.19
Column Totals	28	65	106	68	58	45	26	10	9	4	1	4	424	3220	-

The table is a cross classification of the 424 dwelling units in the 100 percent data area by the number of home-based person trips produced by the dwelling units versus the number of attraction zones with which those trips interacted. These data give an indication of the dispersiveness of travel at the dwelling unit level.

Interactions and Sampling

The sampling theory generally associated with the home-interview O-D survey was reviewed with regard to its application to interactions and the expected impact of its application on interaction estimation. The population data from the three zones provide a unique data base for analysis of the estimation of interactions from sample data. Sets of random samples were drawn from the 100% data set using various sampling rates; the characteristics of sampled interaction data were then compared with population data. The relationship between sampled trips and sampled interactions was also analyzed.

Sets of Random Samples

A computer program was developed to perform the desired sampling and to summarize the results. A uniform random number generator was used to draw sets of 100 random samples from each zone and from the combined area for sampling rates of 5, 12.5, 25, 50 and 90%. The sampling process was constrained such that no dwelling unit could appear more than once in any given sample (i.e., sampling without replacement.) In order to verify the adequacy of these sets of 100 random samples in representing the results from repeated random samples, sets of 500 random samples were drawn from Zone B and the combined area for sampling rates of 5 and 50%; a set of 1,000 5-percent random samples compared with those results obtained using 100 random samples. This comparison indicated that the 100 random samples did not produce significantly different results from the larger sets of samples; therefore, the sets of 100 random samples were accepted as a data base for the analysis of interactions and sampling.

The observed means and standard deviations of the distributions of observed interactions per sample for the various sets of random samples are summarized in Table II-6. The average percent of the total interactions (for a given zone and trip purpose) which was detected in sampling is also summarized in this table.

Distribution of Observed Interactions from Sampling

As may be recalled from the analysis of trip end estimates from the 100% data (1), it was demonstrated that the distribution of observed trips

TABLE II-6: INTERACTIONS OBSERVED IN SAMPLED DATA

HOME-BASED PERSON

NOMINAL SAMPLING RATE (percent)	MEAN NUMBER OF INTERACTIONS DETECTED PER SAMPLE				STANDARD DEVIATION OF THE INTERACTIONS DETECTED PER SAMPLE				MEAN PERCENT OF TOTAL INTERACTIONS DETECTED PER SAMPLE			
	Zone A	Zone B	Zone C	Combined	Zone A	Zone B	Zone C	Combined	Zone A	Zone B	Zone C	Combined
5	14.9	22.5	22.8	50.8	3.9	5.6	4.3	6.3	8.9	9.6	10.4	12.7
12.5	33.0	48.8	48.7	102.7	5.6	6.6	6.4	9.0	19.8	20.9	22.2	25.6
25	58.7	85.2	81.2	169.3	6.2	7.7	6.8	9.3	35.2	36.4	37.1	42.2
50	100.6	144.9	133.1	267.9	5.8	8.1	8.3	9.3	60.2	61.9	60.8	66.8
90	155.0	217.1	204.3	379.5	3.3	4.5	4.0	5.2	92.8	92.8	93.3	94.6

HOME-BASED AUTO DRIVER

NOMINAL SAMPLING RATE (percent)	MEAN NUMBER OF INTERACTIONS DETECTED PER SAMPLE				STANDARD DEVIATION OF THE INTERACTIONS DETECTED PER SAMPLE				MEAN PERCENT OF TOTAL INTERACTIONS DETECTED PER SAMPLE			
	Zone A	Zone B	Zone C	Combined	Zone A	Zone B	Zone C	Combined	Zone A	Zone B	Zone C	Combined
5	13.9	20.1	20.6	46.4	3.7	5.2	4.6	5.8	8.7	9.5	10.1	12.4
12.5	31.0	43.7	44.1	94.8	5.2	6.1	6.5	8.6	19.5	20.6	21.5	25.4
25	55.4	76.3	74.5	156.0	5.7	6.8	6.7	8.6	34.8	36.0	36.3	41.7
50	94.5	130.5	123.0	248.7	5.4	7.3	8.0	8.7	59.4	61.6	60.0	66.5
90	147.3	196.5	190.7	354.1	3.1	4.2	4.0	4.7	92.6	92.7	93.0	94.7

HOME-BASED WORK AUTO DRIVER

NOMINAL SAMPLING RATE (percent)	MEAN NUMBER OF INTERACTIONS DETECTED PER SAMPLE				STANDARD DEVIATION OF THE INTERACTIONS DETECTED PER SAMPLE				MEAN PERCENT OF TOTAL INTERACTIONS DETECTED PER SAMPLE			
	Zone A	Zone B	Zone C	Combined	Zone A	Zone B	Zone C	Combined	Zone A	Zone B	Zone C	Combined
5	4.3	7.4	6.5	16.8	2.1	2.4	2.2	3.6	6.1	6.8	6.6	7.7
12.5	10.6	18.0	16.3	40.3	3.3	3.5	2.7	4.1	15.1	16.5	16.5	18.5
25	20.5	32.4	29.3	73.0	3.9	4.0	3.8	6.2	29.3	29.7	29.6	33.5
50	37.9	61.4	54.9	130.4	3.8	4.2	4.5	5.3	54.1	56.3	55.5	59.8
90	64.3	99.9	90.6	202.9	2.3	2.5	2.3	3.8	91.9	91.7	91.5	93.1

HOME-BASED NONWORK AUTO DRIVER

NOMINAL SAMPLING RATE (percent)	MEAN NUMBER OF INTERACTIONS DETECTED PER SAMPLE				STANDARD DEVIATION OF THE INTERACTIONS DETECTED PER SAMPLE				MEAN PERCENT OF TOTAL INTERACTIONS DETECTED PER SAMPLE			
	Zone A	Zone B	Zone C	Combined	Zone A	Zone B	Zone C	Combined	Zone A	Zone B	Zone C	Combined
5	10.1	13.1	14.8	31.8	2.9	4.8	4.4	5.5	9.4	9.9	10.4	12.5
12.5	21.9	27.3	30.7	63.0	4.1	5.6	6.0	7.8	20.5	20.5	21.6	24.7
25	38.2	48.0	52.1	104.0	4.6	6.2	6.1	8.2	35.7	36.1	36.7	40.8
50	64.7	80.8	84.3	165.8	5.2	7.1	6.6	8.6	60.5	60.8	59.4	65.0
90	99.3	122.5	132.2	240.7	2.8	4.2	3.7	4.4	92.8	92.1	93.1	94.4

Sets of 100 random samples at various sampling rates were drawn from each of the three 100% data zones and the combined area. This table summarizes the means and standard deviations of the distributions of interactions observed from repeated random samples. The mean percents of the interactions detected per sample are also shown.

approaches a normal distribution as a sample size increases. Since interactions are a function of trips, it would seem reasonable to expect that the distribution of observed interactions from repeated random samples might likewise be expected to approach a normal distribution as a sample size increases. The frequency distribution of the home-based person interaction from the 500 5% random samples in the combined area is shown in Figure II-1. As expected, this distribution approximates a normal distribution.

Using the sets of random samples at the various sampling rates drawn from each zone, the distribution of the observed interactions from repeated random samples for each of the four trip categories (i.e., home-based person trips, home-based auto driver trips, home-based work auto driver trips, and home-based nonwork auto driver trips) was obtained for each of the five sampling rates (i.e., 5, 12.5, 25, 50, and 90%). The Kolmogorov-Smirnov test was used to determine whether the distributions were significantly different from a normal distribution. The results of these tests indicate that the normality of the distributions of estimates for the 5-percent sampling rate for Zones B and C and the combined area could not be rejected at the 80% confidence level. In essence, these results suggest that the distribution of observed interactions from repeated random sampling, using a sample size of eight dwelling units and above, may generally be expected to closely approximate a normal distribution.

Interactions and Sampling Theory

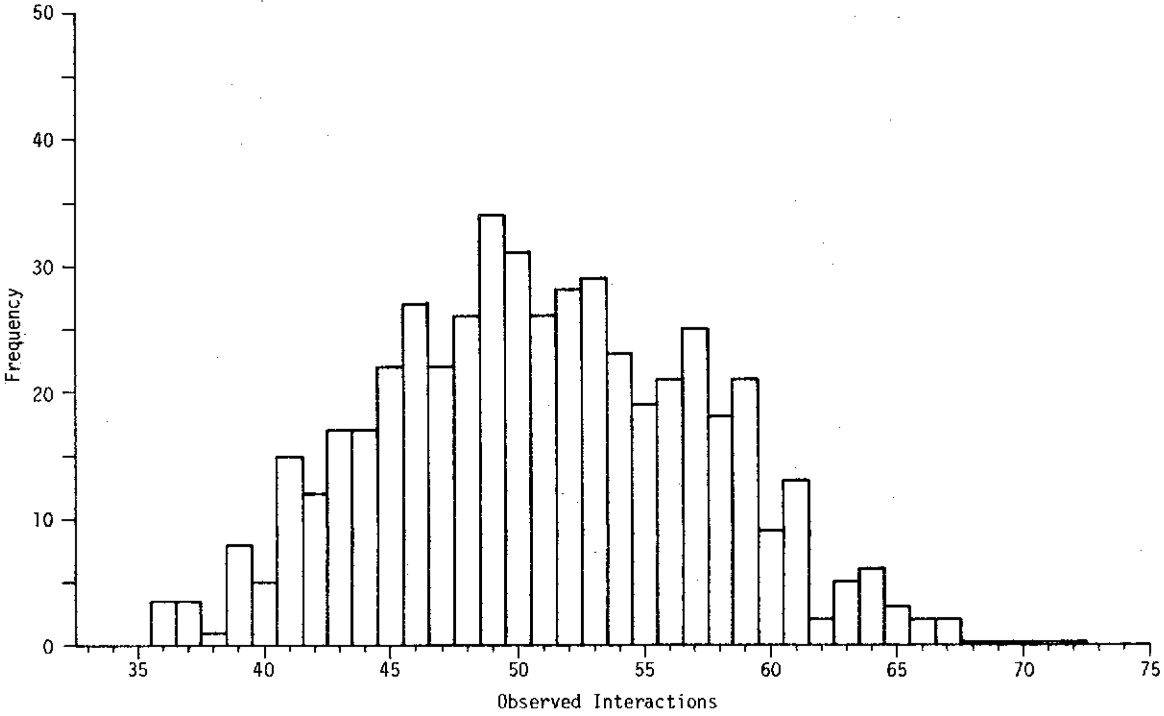
The sampling theory generally associated with home interview O-D surveys is not applicable to the estimation of interactions because the basic assumptions are not met by interactions. This can perhaps best be demonstrated mathematically by comparing the basic relationship involved in the estimation of the total trips produced by a zone versus the relationship associated with interactions. Stated mathematically, the basic relationship between the total trips produced by a zone and the trips produced by the dwelling unit (the unit of observation in home interview surveys) is:

$$P_x = t_1 + t_2 + t_3 + \ldots + t_n = \sum_{j=1}^{n} t_j$$

where:

P_x = the total trips produced by zone x

FIGURE II-1: DISTRIBUTION OF SAMPLED INTERACTIONS



Five hundred random 5% dwelling units samples were drawn from the combined area. The distribution of the 500 samples by the number of home-based person trip interactions actually observed in each sample approximates a normal distribution.

t_j = the trips produced by dwelling unit j in zone x

N = the number of dwelling units in zone x

The important aspects of this relationship which are necessary assumptions for the application of the sampling theory generally associated with home interview surveys are:

- the relationship is additive (i.e., the whole is equal to the sum of its parts)
- the observations (t_j for j = 1, ..., N) are independent.

The basic relationship between interactions and dwelling units, on the other hand, may be expressed mathematically using set theory* as follows:

$$I_x = i_1 \cup i_2 \cup i_3 \cup \ldots \cup i_N$$

where:

I_x = the set of interactions for zone x

i_j = the set of interactions associated with dwelling unit j in zone x.

N = the number of dwelling units in zone x.

Although the observations are independent, it is generally not an additive relationship as is a necessary assumption for the sampling theory usually associated with the home interview O-D surveys.

It is interesting to note, however, that an additive relationship between interactions and dwelling units may be expressed as follows:

$$I_x = k_1 + k_2 + k_3 + \ldots + k_N$$

where:

I_x = the number of interactions for zone x.

k_j = the number of new interactions produced by dwelling unit j in zone x (i.e., the number of interactions produced by dwelling unit j which was not previously produced by dwelling units 1 through j-1).

N = the number of dwelling units in zone x.

---

*The notation "U" is a mathematical operation indicating the union of sets.

Although an additive relationship, the observations are not independent. Indeed, each observation is a function of the preceding observations. Except for the special case where no two dwelling units interact with the same zone, the following is true:

$$I_x < \sum_{j=1}^{N} i_j$$

where:

I_x = the number of interactions for zone x.

i_j = the number of interactions associated with dwelling unit j.

N = the number

This, of course, suggests that the expansion of observed interactions from sample data may be expected to overestimate the number of interactions for the zone. The general tendency to overestimate interactions when expanding sample data may be observed from the results of the sets of random samples previously summarized in Table II-6. The mean percent of the total interactions detected per sample consistently exceeds the sampling rate by a substantial margin. Since the distribution of observed interactions for the combined area approximates a normal distribution for sampling rates of 5% and 12.5%, the observed means and standard deviations were used to estimate the expected error ranges at the 80% and 95% confidence levels which would result from the expansion of sample data. The results of these calculations (summarized in Table II-7) suggest, for example, that if repeated random 5% samples were drawn from the combined area and observed interactions were expanded to estimate the number of interactions associated with the home-based person trips, then approximately 95% of these samples would be expected to overestimate the true number of interactions by 94% to 220%. Similarly, approximately 80% of the samples would overestimate the true number of home-based person trip interactions by 116% and 198%. On the average, the repeated random 5% samples from the combined area would be expected to overestimate the number of home-based person trip interactions

TABLE II-7: EXPECTED PERCENT ERROR IN ESTIMATION OF INTERACTIONS BY EXPANSION OF SAMPLE DATA

5 PERCENT SAMPLING RATE

Trip Purpose	Actual Interactions	Average Expected Percent Error	Expected % Error Ranges	
			95% Confidence	80% Confidence
Home-Based Person Trips	399	+157	+94 to +220	+116 to +198
Home-Based Auto Driver Trips	372	+152	+90 to +214	+112 to +192
Home-Based Work Auto Driver Trips	216	+ 57	- 9 to +170	+ 14 to +100
Home-Based Nonwork Auto Driver Trips	253	+154	+68 to +240	+ 98 to +210

12.5 PERCENT SAMPLING RATE

Trip Purpose	Actual Interactions	Average Expected Percent Error	Expected % Error Ranges	
			95% Confidence	80% Confidence
Home-Based Person Trips	399	+105	+71 to +141	+83 to +129
Home-Based Auto Driver Trips	372	+104	+68 to +140	+80 to +128
Home-Based Work Auto Driver Trips	216	+ 49	+20 to + 79	+30 to + 69
Home-Based Nonwork Auto Driver Trips	253	+ 99	+51 to +148	+68 to +131

The data presented demonstrates the extremely high percent error which would be expected if the number of interactions for the combined area were estimated by the usual direct expansion of the observations from sample data using traditional sampling rates.

by 157% (i.e., the average sample would be expected to produce an estimate of the number of home-based person trip interactions, which approximately two-and-one-half time the actual number of interactions.

With the exception of home-based work trip interactions, the 5% sampling rate resulted in an average expected percent error of roughly 150% for each of the three trip purposes; the 12.5% samples resulted in an average percent error of roughly 100% for the same three trip purposes. In other words, with the exception of the home-based work trip interactions, the 5% sampling rate produced estimates which were, on the average, about two-and-one-half times the actual; the average 12.5% sample would be expected to produce an estimate which is roughly double the actual number of interactions. There is a 99% probability of over estimating home-based person, home-based auto driver, and home-based nonwork auto driver trips at a 5% sampling rate.

The average expected percent error for home-based work trip interactions were 57% and 49% for the 5% and 12.5% sampling rates respectively. The expected percent error rante for the home-based work auto driver trips at the 5% sampling level (at the 95% confidence level) has a lower bound which indicates a slight underestimate of the actual number of interactions. Nevertheless, there is about a 95% probability of overestimating home-based work trip interactions a the 5% sampling level.

The empirical data presented confirm the hypothesis that expanded sample data may generally be expected to overestimate the interactions for a zone. Indeed, the data suggest that the magnitude of this overestimate would be substantial (i.e., generally exceeding 100% error).

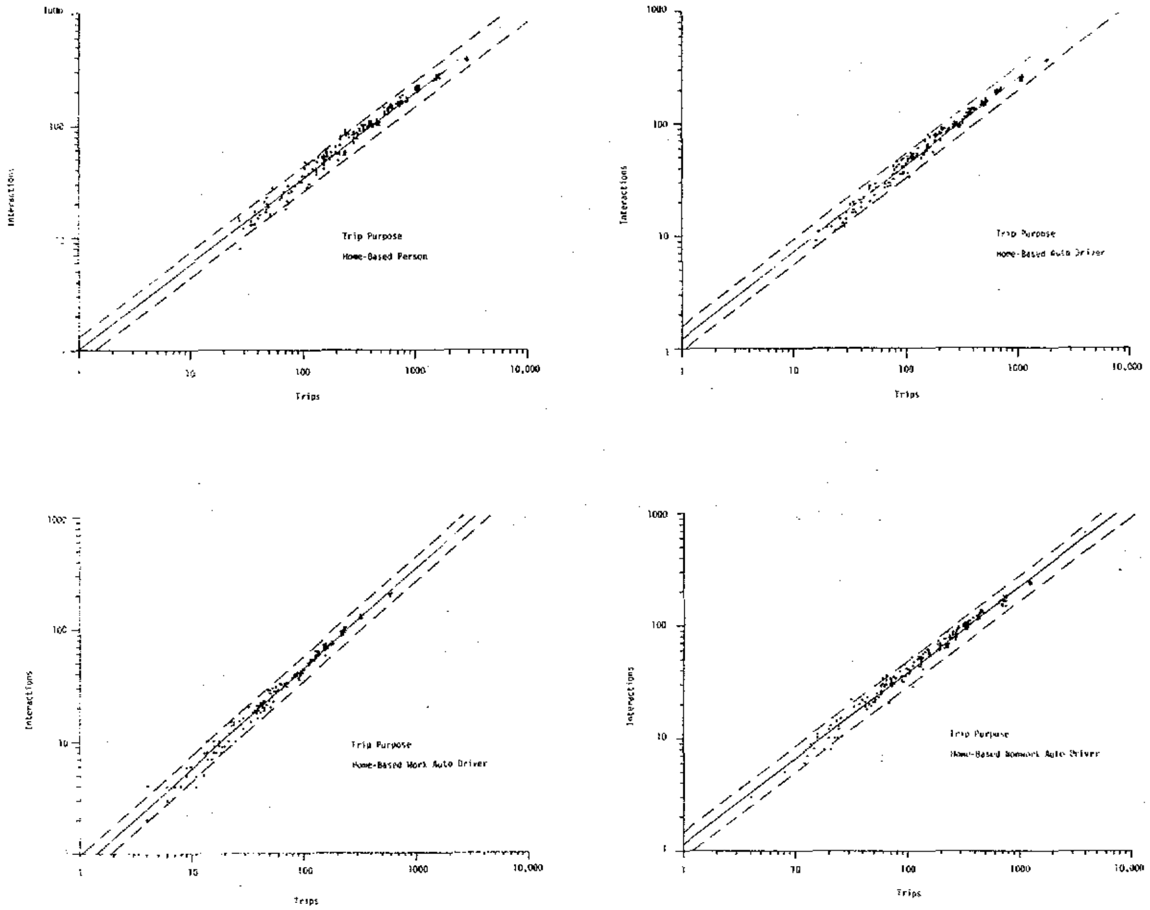
The magnitude of the overestimate is a function of zone size (as measured by trip productivity or number of dwelling units). As the zone size increases, the percent of the interactions detected at a given sampling rate may be expected to increase. This may be observed for Table II-6 by comparing the mean percent of total interactions detected per sample for the individual zones versus the combined area. In this example the total number of zones in the network remains roughly constant (i.e., 1,569 zones for the individual zones versus 1,567 zones for the combined area). In essence, this suggests that, within a given zonal structure, the magnitude of the expected error at a given sampling rate may generally be expected to increase as the size of the production zone increases. Similarly, if a new zonal structure were superimposed on the urban area which is substantially larger than the zone sizes throughout the area, then the magnitude of the percent error at a given sampling rate would likewise be expected to generally increase. In short, this simply suggests that, as the number of attraction zones is reduced by increasing zone sizes, the percent of interactions detected using a given sampling rate would be expected to increase, thereby increasing the expected percent error which results from the direct expansion of sampled interactions.

Interactions and Trips

Since interactions cannot be estimated by traditional expansion of sample data, the emphasis in the investigation of the relationship between observed trips and observed interactions was to determine if there is a relationship which might be used to estimate the number of interactions from survey data. To provide a reasonable size data set for graphical analysis, the first ten random samples drawn from each of the four zones for each of the five sampling rates were selected for analysis. This approach provided 200 data points for the analysis of each trip purpose (each point corresponds to the number of trips observed versus the number of interactions observed in a given sample for the specified trip purpose). The set of 200 data points for a given trip purpose also represents a range of sample sizes varying from 5 dwelling units (i.e. a nominal 5% sampling rate in zone A) to 382 dwelling units (i.e., a nominal 90% sampling rate in the Combined Area).

Since the data exhibited a general exponential tendency, log-log plots of the data were prepared and are exhibited in Figure II-2. As can be seen from these plots, the data points might be reasonably approximated by a straight line, thereby suggesting an exponential relationship. The regression line illustrated on each plot was determined by a least squares curve fit of the form.

FIGURE II-2: OBSERVED TRIPS AND INTERACTIONS



The data points represent the observed trip production volumes and interactions (unexpanded) from repeated random samples at various sampling rates and from the 100 percent data. These data suggest an exponential relation exists between trips and interactions.

$$\log(I) = a + b(\log(P))$$

or

$$I = AP^b$$

where:

I = number of interactions
P = production volume
b = regression coefficient
a = regression intercept
A = constant equal to 10^a

A problem in the use of such an estimating equation is that there is a critical value, P_c, such that if P is greater than P_c then the number of interactions will exceed the maximum number of nonzero attraction zones. This problem may be avoided by stating the estimating equation as follows:

$$I = \begin{cases} AP^b & \text{for } P < P_c \\ N & \text{for } P \geq P_c \end{cases}$$

where:

I = number of interactions
A = constant determined from regression
P = production volume
b = coefficient of regression
N = number of nonzero attraction zones
P_c = the critical production volume such that $P_c^b = N$

The dashed parallel lines on either side of the regression line represent the ±25% band relative to the number of interactions estimated by the regression line for a given production volume (note, this band is not the confidence limit of the regression equation). The significance of these plots is:

- A relationship exists between interactions and trips which might be used in estimating interactions from sample data.
- The relationship is such that as the production volume increases, the expected number of interactions increases at a decreasing rate.
- The relationship varies somewhat by trip purpose.

It appears from the data observed that this approach gives at least a reasonable estimate of the average expected number of interactions that might be associated with a zone of a given production volume. It must be remembered that it is the average expected number which is being estimated and the actual number is, of course, subject to some variance. Referring to Figure II-2, it may be observed that the data points were generally within a ± 25% band of the regression line. This analysis, of course, utilized data from repeated random samples from the three zones and the combined area representing sampling rates from 5% to 90%. Such data would obviously not be available to the analyst in the course of a normal O-D survey. Nevertheless, it would seem that such an approach might be employed using observed trips and observed interactions for a given trip purpose. A line of best fit may be determined and used to calculate an estimate of the number of interactions for the estimated production volumes. The analyst should, of course, carefully review the reasonableness of the results.

The user of the Texas Trip Distribution Model is, of course, not directly interested in estimating the number of interactions for a zone but in estimating a reasonable maximum number of interactions (i.e. a reasonable upper bound) for a zone of a given production volume. Thus, the results from the proposed approach would not be used directly but would be a useful guide in estimating a bounding condition for a given trip purpose in a given urban area.

Conclusions

The following summarizes the conclusions from both the analysis of the basic interaction data and the analysis of interactions and sampling:

- A general relationship exists between trips and interactions such as the production volume for a zone increases, the expected number of interactions will increase at a decreasing rate until the bounding condition (i.e., the maximum number of possible attraction zones) is reached. This relationship may vary from urban area to urban area, from zonal structure to zonal structure within a given urban area, and from trip purpose to trip purpose for a given urban area and structure.

- The impedance to travel due to spatial separation is reflected in the spatial distribution of interactions.
- As the trip productivity of a dwelling unit increases, the expected number of dwelling unit interactions (i.e., the number of attraction zones with which the dwelling unit is expected to interact) increases but at a decreasing rate.
- The relationship between dwelling unit interactions and zonal interactions may be stated mathematically as follows:

$$I_x = i_1 \cup i_2 \cup i_3 \cup \ldots \cup i_N$$

where

I_x = the set of interactions for zone x

i_j = the set of interactions associated with dwelling unit j in zone x.

N = the number of dwelling units in zone x.

- Except for the special case where no two dwelling units interact with the same zone, the following is true:

$$I_x < \sum_{j=1}^{N} i_j$$

where:

I_x = the number of interactions for zone x.

i_j = the number of dwelling unit interactions associated with dwelling unit j in zone x.

N = the number of dwelling units in zone x.

- The sampling theory generally associated with home interview O-D surveys is not applicable to the estimation of interactions. The expansion of observed interactions for sample data employing traditional sampling rates resulted in very large overestimates of the actual number of interactions for the zones studied.
- The analysis of observed trips and observed interactions indicated that there is an exponential relationship between trips and interactions. Such a relationship might be estimated from sample data and subsequently used to estimate the expected number of interactions for given production volumes.

INTERCHANGE VOLUME ANALYSIS

The production-attraction interchange volume, T_{ij}, is the number of trips produced by zone i which are attracted to zone j and are the principal output of a trip distribution model. Since an interaction is a zone pair having a nonzero interchange volume, the estimation of interactions is simply the estimation of the number of nonzero interchange volumes associated with a production zone. The number of trips at a given spatial separation (a basic element in trip length frequency) is essentially the sum of the interchange volumes for all the zone pairs having a given separation.

As with the interaction analysis, the interchange volume analysis was divided into two phases: The analysis of the basic interchange volume data and, the accuracy of interchange volumes estimated from sample data.

Analysis of Basic Interchange Volume Data

The analysis of the basic interchange volume data is an analysis of the population data (i.e., the 100% sample) for the three zones in San Antonio. The primary objective of such an analysis is to review the actual travel patterns observed with regard to interchange volumes. Since interactions are nonzero interchange volumes, this analysis obviously overlaps with the analysis of the basic interaction data presented in the previous section. This overlap is not only natural but desirable,in that it supplements and provides further perspective into this aspect of travel patterns.

Interchange Volumes and Production Volumes

Since trip distribution models generally focus on the estimation of interchange volumes between zone pairs, it is worthwhile to briefly review trip distribution theory with regard to the relationship between interchange volumes and production volumes. The general form of the estimating equation associated with the Gravity Model, the Intervening Opportunities Model, the Competing Opportunities Model, and the Texas Trip Distribution Model is simply:

$$T_{ij} = P_i \, p_{ij}$$

where:

T_{ij} = the expected number of trips produced by zone i which are attracted to zone j (i.e. the interchange volume between zone pair ij).

P_i = the production volume for zone i

p_{ij} = the probability of trips produced by zone i being attracted to zone j

(Note: $\sum_{j=1}^{N} p_{ij} = 1$

where N = the number of attraction zones).

These models basically differ in their approach toward the estimation of p_{ij}. Using this general form, the average interchange volume for zone i is:

$$\overline{T}_i = \frac{\sum_{j=1}^{N} T_{ij}}{N} = \frac{P_i}{N}$$

where:

$\overline{T}_i$ = *the average interchange volume for zone i*

T_{ij} = the expected interchange volume for zone pair ij.

P_i = the production volume for zone i.

N = the number of attraction zones.

The average interchange volume is, of course, directly proportional to the production volume and inversely proportional to the number of attraction zones. Any two zones in a given network having the same production volume will have the same average interchange volume. Obviously, such a measure gives very little information relative to the travel pattern for a zone. A better measure, while still a relatively gross measure, is the average nonzero interchange volume which may be defined as follows:

$$\overline{T}_{Ii} = \frac{\sum_{j=1}^{N} T_{ij}}{I_i} = \frac{P_i}{I_i}$$

where:

$\overline{T}_{Ii}$ = the average nonzero interchange volume for zone i (i.e. the average trips per interaction).

I_i = the number of nonzero interchange volumes for zone i (i.e. the number of interactions).

As was discussed in the section of this research report entitled "Interaction Analysis," the expected number of interactions for a zone is a function of the production volume for the zone. This may be expected to vary from urban area to urban area, with different zonal structures for a given urban area, and from trip purpose to trip purpose within a given urban area. Nevertheless, this relationship may be expected to exhibit the following general characteristic: as the production volume for a zone increases, the number of interactions may be expected to increase at a decreasing rate and approach the bounding condition (i.e. the number of possible attraction zones). Therefore, as the production volume for a zone increases, the average nonzero interchange volume (average trips per interaction) would be expected to increase. This relationship is dependent upon the zonal structure superimposed upon the urban area and the distribution of attractions at various separations for the various production zones (or, indeed, the accessibility distribtuion for each production zone).

A summary of the average interchange volumes and the average nonzero interchange volumes by zone and trip purpose from the 100 percent data area is given in Table III-1, together with the number of nonzero interchange volumes (i.e. the number of interactions) and the percent of interchange volumes which are nonzero. Since the three zones being studied are adjacent zones, they have similar accessibility characteristics. As indicated by the data presented in Table III-1, the average interchange volume increases as the production volume increases, since the average interchange volume is directly proportional to the production volume. For each trip category, the average nonzero interchange volume was nearly equal for the individual zones; however, it is substantially larger in the conbined area. Recalling from the interaction analysis, the relationship between production volume and the number of interactions was that the expected number of interactions increases at a decreasing rate as the production volume increases. A corollary to this is that as the production volume increases, the average nonzero interchange volume will increase at an increasing rate. This may be observed by comparing the average nonzero interchange volume for the individual zones with that for the combined area. However, this relationship is not apparent from the comparison of the individual zones since their average nonzero interchange volumes are roughly constant. This does not contradict the hypothesized relationship, but suggests that the increase in the average nonzero interchange volume is very small over the volume range represented by the individual zones and is subject to some variation.

Comparison of different zones for any given trip category provides another observation. Zones B and C (for example) have similar production volumes for each trip category, which result in approximately the same number of interactions. However, the number of interactions and, therefore, the average nonzero interchange volumes were subject to some variations since the relationship between zonal production volumes and interchange volumes is probabilistic.

The frequency distributions of interchange volumes (including interchange volumes of zero) by volume groups and the frequency distributions of interactions (i.e. interchanges with nonzero volumes) by volume groups are summarized in Table III-2. The average interchange volumes and the average nonzero interchange volumes are, of course, the means of these distributions.

TABLE III-1: AVERAGE INTERCHANGE VOLUMES

Trip Purpose and Zone	Production Volume	Average Interchange Volume	Percent with Nonzero Interchange Volumes*	Number of Nonzero Interchange Volumes**	Average Nonzero Interchange Volume***
Home-based Person Trips					
Zone A	860	0.55	10.7	167	5.1
Zone B	1180	0.75	14.9	234	5.0
Zone C	1180	0.75	14.0	219	5.4
Combined Area	3220	2.05	25.6	399	8.0
Home-based Auto Driver Trips					
Zone A	563	0.36	10.1	159	3.5
Zone B	764	0.49	13.5	212	3.6
Zone C	763	0.49	13.1	205	3.7
Combined Area	2090	1.33	23.8	372	5.6
Home-based Work Auto Driver Trips					
Zone A	173	0.11	4.5	70	2.5
Zone B	261	0.17	6.9	109	2.4
Zone C	247	0.16	6.3	99	2.5
Combined Area	681	0.43	13.9	216	3.1
Home-based Nonwork Auto Driver Trips					
Zone A	390	0.25	6.8	107	3.6
Zone B	503	0.32	8.5	133	3.8
Zone C	516	0.33	9.1	142	3.6
Combined Area	1409	0.90	16.3	253	5.5

* Interactions as a percent of attraction zones.
** Number of interactions.
*** Average trips per interaction.

Summarized are the production volumes, the average interchange volumes, and the average nonzero interchange volumes observed from the three 100 percent data zones and the combined area. The relationship between production volumes and both the interchange volumes and nonzero interchange volumes may be observed in these data.

As may be observed from this table, about 75% to 95% of the interchange volumes were zero for the three zones and the combined area. In other words, the zones interacted with only about 5% to 25% of the possible attraction zones. Only two interchange volumes exceeded 150 trips and both were associated with the home based person trips for the combined area. None of the interchange volumes associated with the individual zones exceeded 100 trips. The only interchange volumes associated with individual zones which exceeded 50 trips were associated with the home based person trips. Indeed, about 85% to 99% of the nonzero interchange volumes involve 10 or fewer trips; and, a majority of these interchange volumes involve fewer than 5 trips.

These data indicate that, in studies of large urban areas using small- to medium-size zones, most of the production zones will interact with fewer than half of the possible attraction zones. This suggests that, in general, more than half of the interchange volumes may be expected to be zero. In studies of smaller urban areas involving small- to medium-sized zones, the proportion of nonzero interchange volumes may be expected to increase. In urban transportation studies using small- to medium-size zones (regardless of the size of the urban area), it may be expected that the majority of the nonzero interchange volumes will involve 10 or fewer trips.

This, of course, has some obvious implications relative to the estimation of interchange volumes from sample data. For example, at a 12.5% sampling rate (i.e., an expansion factor of 8), the *minimum nonzero* interchange volume that may be estimated is 8 trips. Similarly, at a 5% sampling rate (i.e., an expansion factor of 20), the minimum nonzero interchange volume that may be estimated is 20 trips. This means that the small interchange volumes detected would be overestimated. As will be recalled from the interaction analysis, however, a large portion of the interchanges would not be detected using traditional sampling rates and, therefore, would be underestimated. The accuracy of the estimation of interchange volumes from sample data will be discussed in greater detail in the subsequent section entitled "Interchange Volumes and Sampling."

TABLE III-2: FREQUENCY DISTRIBUTIONS OF INTERCHANGE VOLUMES

FREQUENCY DISTRIBUTIONS OF INTERCHANGES AND TRIPS BY INTERCHANGE VOLUME GROUPS

Trip Purpose and Zone	Production Volume	Interchange Volume Groups												
		0	1 to 10		11 to 20		21 to 50		51 to 100		101 to 150		151 to 182	
		Pct. of Att. Zones	Pct. of Att. Zones	Pct. of Trips	Pct. of Att. Zones	Pct. of Trips	Pct. of Att. Zones	Pct. of Trips	Pct. of Att. Zones	Pct. of Trips	Pct. of Att. Zones	Pct. of Trips	Pct. of Att. Zones	Pct. of Trips
Home-based Person Trips														
Zone A	860	89.35	9.82	56.40	0.32	7.56	0.32	15.46	0.19	20.58	0.0	0.0	0.0	0.0
Zone B	1180	85.09	13.70	53.22	0.45	9.41	0.57	21.95	0.19	15.42	0.0	0.0	0.0	0.0
Zone C	1180	86.04	12.56	48.56	0.64	11.44	0.57	24.92	0.19	15.08	0.0	0.0	0.0	0.0
Combined Area	3220	74.54	22.08	37.92	1.72	11.46	0.83	13.11	0.57	19.56	0.13	7.24	0.13	10.71
Home-based Auto Driver Trips														
Zone A	563	89.87	9.56	66.61	0.25	10.30	0.32	23.09	0.0	0.0	0.0	0.0	0.0	0.0
Zone B	764	86.49	12.75	67.02	0.38	11.91	0.38	21.07	0.0	0.0	0.0	0.0	0.0	0.0
Zone C	763	86.93	12.11	59.11	0.64	20.31	0.32	20.58	0.0	0.0	0.0	0.0	0.0	0.0
Combined Area	2090	76.26	21.57	47.99	0.83	8.90	0.96	21.63	0.32	12.39	0.06	9.09	0.0	0.0
Home-based Work Auto Driver Trips														
Zone A	173	95.54	4.40	87.86	0.0	0.0	0.06	12.14	0.0	0.0	0.0	0.0	0.0	0.0
Zone B	261	93.05	6.88	94.64	0.07	5.36	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Zone C	247	93.69	6.25	86.64	0.0	0.0	0.06	13.36	0.0	0.0	0.0	0.0	0.0	0.0
Combined Area	681	86.22	13.53	84.43	0.19	5.58	0.0	0.0	0.06	9.99	0.0	0.0	0.0	0.0
Home-based Nonwork Auto Driver Trips														
Zone A	390	93.18	6.37	62.05	0.19	11.03	0.26	26.92	0.0	0.0	0.0	0.0	0.0	0.0
Zone B	503	91.52	7.84	57.65	0.32	15.31	0.32	27.04	0.0	0.0	0.0	0.0	0.0	0.0
Zone C	516	90.95	8.29	56.40	0.57	25.19	0.19	18.41	0.0	0.0	0.0	0.0	0.0	0.0
Combined Area	1409	83.85	14.49	40.74	0.58	10.08	0.83	27.53	0.19	13.84	0.06	7.81	0.0	0.0

FREQUENCY DISTRIBUTIONS OF INTERACTIONS AND TRIPS BY INTERCHANGE VOLUME GROUPS

Trip Purpose and Zone	Production Volume	Inter-Actions	Interchange Volume Groups											
			1 to 10		11 to 20		21 to 50		51 to 100		101 to 150		151 to 182	
			Pct. of Int.*	Pct. of Trips	Pct. of Int.	Pct. of Trips	Pct. of Int.	Pct. of Trips	Pct. of Int.	Pct. of Trips	Pct. of Int.	Pct. of Trips	Pct. of Int.	Pct. of Trips
Home-based Person Trips														
Zone A	860	167	92.22	56.40	2.99	7.56	2.99	15.46	1.80	20.58	0.0	0.0	0.0	0.0
Zone B	1180	234	91.88	53.22	2.99	9.41	3.85	21.95	1.28	15.42	0.0	0.0	0.0	0.0
Zone C	1180	219	89.95	48.56	4.57	11.44	4.11	24.92	1.37	15.08	0.0	0.0	0.0	0.0
Combined Area	3220	399	86.72	37.92	6.77	11.46	3.26	13.11	2.25	19.56	0.50	7.24	0.50	10.71
Home-based Auto Driver Trips														
Zone A	563	159	94.34	66.61	2.52	10.30	3.14	23.09	0.0	0.0	0.0	0.0	0.0	0.0
Zone B	764	212	94.34	67.02	2.83	11.91	2.83	21.07	0.0	0.0	0.0	0.0	0.0	0.0
Zone C	763	205	92.69	59.11	4.88	20.31	2.44	20.58	0.0	0.0	0.0	0.0	0.0	0.0
Combined Area	2090	372	90.86	47.99	3.49	8.90	4.03	21.63	1.34	12.39	0.28	9.09	0.0	0.0
Home-based Work Auto Driver Trips														
Zone A	173	70	98.57	87.86	0.0	0.0	1.43	12.14	0.0	0.0	0.0	0.0	0.0	0.0
Zone B	261	109	99.08	94.64	0.92	5.36	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Zone C	247	99	98.99	86.64	0.0	0.0	1.01	13.36	0.0	0.0	0.0	0.0	0.0	0.0
Combined Area	681	216	98.15	84.43	1.39	5.58	0.0	0.0	0.46	9.99	0.0	0.0	0.0	0.0
Home-based Nonwork Auto Driver Trips														
Zone A	390	107	93.46	62.05	2.80	11.03	3.74	26.92	0.0	0.0	0.0	0.0	0.0	0.0
Zone B	503	133	92.48	57.65	3.76	15.31	3.76	27.04	0.0	0.0	0.0	0.0	0.0	0.0
Zone C	516	142	91.55	56.40	6.34	25.19	2.11	18.41	0.0	0.0	0.0	0.0	0.0	0.0
Combined Area	1409	253	89.72	40.74	3.56	10.08	5.14	27.53	1.18	13.84	0.4	7.81	0.0	0.0

* Percent of Interactions

The data presented is based on the observed interchange volumes from the San Antonio 100 percent data. There are 1569 interchanges (i.e. attraction zones) associated with the individual zones and 1,567 interchanges associated with the Combined Area. The frequency distributions of interactions are the distributions of the nonzero interchange volumes.

Spatial Characteristics

As in the analysis of the spatial characteristics of interactions, this analysis used the network travel time as a measure of spatial separation. For convenience, the spatial separation data were rounded to the nearest one minute. The spatial separations for the combined area were estimated by averaging the spatial separations measured for the three individual zones.

Since interactions are zone pairs with a nonzero interchange volume, the analysis of the spatial characteristics of interchange volumes also overlaps the previous analysis of the spatial characteristics of interactions. In order to minimize duplication, the reader is simply referenced to this previous analysis relative to interactions.

Table III-3 summarizes the spatial frequency distributions of interchange volumes and trips, together with the average nonzero interchange volume observed for each spatial separation interval. The data presented in this table was previously presented in Tables II-3 and II-4, but is repeated here for convenience.

The impedance to travel due to spatial separation is evidenced by the percent of the interchange volumes greater than zero within each separation interval, the number of trips within the various separation intervals, and the average nonzero interchange volumes within the various separation intervals. As trip distribution theory would suggest, the variations observed at the various separations are due to the distribution of attractions by spatial separation. For example, the variation observed in the 21- to 25-minute separation interval relative to the work trips is largely attributable to a major employment center (i.e., Kelley Air Force Base), which exerted a tremendous influence on the zones studied. Nevertheless, as trip distribution theory would suggest, the average nonzero interchange volume generally tended to decrease as the spatial separation increased.

Interchange Volumes and Sampling

The population data from the three zones provide a unique data base for an analysis of the accuracy of the percentage volumes estimated from small samples. Using the 100% survey data base, sets of random samples

TABLE III-3: SPATIAL CHARACTERISTICS OF INTERCHANGE VOLUMES AND TRIPS

HOME-BASED PERSON

SPATIAL SEPARATION INTERVAL (network minutes)	NUMBER OF INTERCHANGE VOLUMES				PERCENT OF INTERCHANGE VOLUMES GREATER THAN ZERO				NUMBER OF TRIPS				AVERAGE NONZERO INTERCHANGE VOLUME			
	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
1 - 5	34	34	37	33	61.8	70.6	62.2	72.7	304	320	368	893	14.5	13.5	16.0	37.2
6 - 10	178	150	141	155	28.7	41.3	42.6	68.4	239	398	356	1103	4.7	6.4	5.9	10.4
11 - 15	427	421	395	420	14.3	23.0	20.5	39.3	208	305	256	746	3.4	3.1	3.2	4.5
16 - 20	306	302	303	299	4.9	10.3	10.6	21.1	49	75	83	217	3.3	2.4	2.6	3.4
21 - 25	241	261	280	260	6.2	5.4	5.7	11.5	53	53	88	193	3.5	3.8	5.5	6.4
26 - 30	147	155	148	155	1.4	3.9	3.4	5.8	3	29	23	64	1.5	4.8	4.6	7.1
31 - 35	105	104	97	105	1.0	0	2.1	1.0	2	0	6	2	2.0	0	3.0	2.0
36 - 40	86	90	98	90	1.2	0	0	1.1	2	0	0	2	2.0	0	0	2.0
41 & above	45	52	70	52	0	0	0	0	0	0	0	0	0	0	0	0
All	1569	1569	1569	1567	10.6	14.9	14.0	25.5	860	1180	1180	3220	5.1	5.0	5.4	8.0

HOME-BASED AUTO DRIVER

SPATIAL SEPARATION INTERVAL (network minutes)	NUMBER OF INTERCHANGE VOLUMES				PERCENT OF INTERCHANGE VOLUMES GREATER THAN ZERO				NUMBER OF TRIPS				AVERAGE NONZERO INTERCHANGE VOLUME			
	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
1 - 5	34	34	37	33	61.8	67.6	62.2	72.7	176	193	234	562	8.4	8.4	10.2	23.4
6 - 10	178	150	141	155	28.1	38.0	40.4	65.8	164	238	206	656	3.3	4.2	3.6	6.4
11 - 15	427	421	395	420	13.1	20.7	18.2	35.7	137	209	171	499	2.4	2.4	2.4	3.3
16 - 20	306	302	303	299	4.6	9.3	10.2	19.4	33	59	63	165	2.4	2.1	2.0	2.8
21 - 25	241	261	280	260	5.8	4.2	5.7	10.4	46	38	70	152	3.3	3.4	4.4	5.6
26 - 30	147	155	148	155	1.4	3.9	2.7	5.9	3	27	13	52	1.5	4.5	3.2	5.8
31 - 35	105	104	97	105	1.0	0	2.1	1.0	2	0	6	2	2.0	0	3.0	2.0
36 - 40	86	90	98	90	1.2	0	0	1.1	2	0	0	2	2.0	0	0	2.0
41 & above	45	52	70	52	0	0	0	0	0	0	0	0	0	0	0	0
All	1569	1569	1569	1567	10.1	13.5	13.1	23.7	563	764	763	2090	3.5	3.6	3.7	5.6

HOME-BASED WORK AUTO DRIVER

SPATIAL SEPARATION INTERVAL (network minutes)	NUMBER OF INTERCHANGE VOLUMES				PERCENT OF INTERCHANGE VOLUMES GREATER THAN ZERO				NUMBER OF TRIPS				AVERAGE NONZERO INTERCHANGE VOLUME			
	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
1 - 5	34	34	37	33	14.7	11.8	32.4	48.5	11	7	28	45	2.2	1.8	2.3	2.8
5 - 10	178	150	141	155	9.0	14.7	12.1	29.0	34	51	39	124	2.1	2.3	2.3	2.8
11 - 15	427	421	395	420	7.0	11.6	10.1	21.7	72	111	82	257	2.4	2.3	2.1	2.8
16 - 20	306	302	303	299	2.0	6.6	5.0	12.7	11	39	34	94	1.8	1.9	2.3	2.5
21 - 25	241	261	280	260	3.7	3.1	3.6	6.2	38	28	51	114	4.2	3.5	5.1	7.1
26 - 30	147	155	148	155	1.4	3.9	2.7	5.8	3	25	9	45	1.5	4.2	2.3	5.0
31 - 35	105	104	97	105	1.0	0	1.0	0	2	0	4	0	2.0	0	4.0	0
36 - 40	86	90	98	90	0	0	0	0	2	0	0	2	2.0	0	0	2.0
41 & above	45	52	70	52	0	0	0	0	0	0	0	0	0	0	0	0
All	1569	1569	1569	1567	4.5	6.9	6.3	13.8	173	261	247	681	2.5	2.4	2.5	3.2

HOME-BASED NONWORK AUTO DRIVER

SPATIAL SEPARATION INTERVAL (network minutes)	NUMBER OF INTERCHANGE VOLUMES				PERCENT OF INTERCHANGE VOLUMES GREATER THAN ZERO				NUMBER OF TRIPS				AVERAGE NONZERO INTERCHANGE VOLUME			
	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
1 - 5	34	34	37	33	61.8	64.7	59.5	72.7	165	186	206	517	7.9	8.5	9.4	21.5
6 - 10	178	150	141	155	22.5	30.0	35.5	58.1	130	187	167	532	3.3	4.2	3.3	5.9
11 - 15	427	421	395	420	7.5	11.4	10.4	22.1	65	98	89	242	2.0	2.0	2.2	2.6
16 - 20	306	302	303	299	2.6	1.6	5.9	9.0	22	20	29	71	2.8	1.8	1.6	2.6
21 - 25	241	261	280	260	2.5	1.9	3.2	6.2	8	10	19	38	1.3	2.0	2.1	2.4
26 - 30	147	155	148	155	0	1.3	0.7	1.3	0	2	4	7	0	1.0	4.0	3.5
31 - 35	105	104	97	105	0	0	1.0	1.0	0	0	2	2	0	0	2.0	2.0
36 - 40	86	90	98	90	0	0	0	0	0	0	0	0	0	0	0	0
41 & above	45	52	70	52	0	0	0	0	0	0	0	0	0	0	0	0
All	1569	1569	1569	1567	6.8	8.4	9.1	16.1	390	503	516	1409	3.6	3.8	3.6	5.6

Summarized are the distributions of interchange volumes, nonzero interchange volumes, and trips by spatial separation for the three 100 percent data zones and the combined area. Again, the impedance to travel due to spatial separation is apparent in these data.

were drawn using various sampling rates. The results of these random samples were used to study the characteristics of sampled interchange volume data.

Interchange Volumes and Sampling Theory

The sampling theory generally associated with home interview O-D surveys is applicable to the estimation of interchange volumes. The basic relationship between interchange volume and the trips produced by a dwelling unit may be stated mathematically as follows:

$$T_{ij} = t_{ij1} + t_{ij2} + \dots + t_{ijN} = \sum_{k=1}^{N} t_{ijk}$$

where:

T_{ij} = the number of trips produced by zone i that are attracted to zone j.

t_{ijk} = the number of trips produced by dwelling unit k in zone i that are attracted to zone j.

N = the number of dwelling units in zone i.

Since the relationship is additive, the traditional sampling theory is applicable to interchange volumes (i.e. each specific zone pair). A distinct parent frequency distribution is associated with each interchange. The parent frequency distribution associated with a given interchange is the distribution of dwelling units in the production zone by a number of trips each dwelling unit contributes to the given interchange volume. The mean of the parent frequency distribution is the average trips per dwelling unit from the production zone which were attracted to the given attraction zone.

If it may be assumed that the distribution of estimates of the mean of the parent frequency distribution from repeated random samples would approximate a normal distribution, then, assertions could be made relative to the expected accuracy. The Central Limit Theorem indicates that:

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

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 of estimates will adequately approximate a normal distribution such that normality may be assumed. The theorem also assumes random sampling with replacement, while in O-D surveys, the sampling is obviously without replacement. However, the theorem is not extremely sensitive to the sampling with replacement assumption, except in cases where the parent frequency distribution is highly skewed such that the number of dwelling units actually contributing to a given interchange volume is very small. For example, if only one dwelling unit from a zone contributes trips to a given interchange volume (regardless of the size of the interchange volume), it is impossible for the distribution of the estimated means from repeated random samples, when sampling without replacement, to approximate a normal distribution, regardless of the sample size. This is especially a problem in dealing with small interchange volumes, since only a very small portion of the dwelling units in a zone can contribute to the interchange volume (i.e. the number of dwelling units contributing trips to a given interchange volume will always be less than or equal to the interchange volume). It appears, however, that in dealing with an interchange volume having as few as say four to six contributing dwelling units (i.e. dwelling units contributing one or more trips to the given interchange volume), the estimated means from repeated random samples would reasonably approximate a normal distribution above some sample size. Unfortunately, as will be subsequently shown, the required sample size for such an assumption would probably far exceed the traditional sampling rates used in urban transportation studies.

Application of Sampling Theory

In estimating zonal trip ends, the proportion of dwelling units in a zone which produce no trips for the given trip purpose would be expected to be a very small portion of the dwelling units in that zone. In contrast, the number of dwelling units not contributing trips to a given interchange volume may comprise a larger portion of the total dwelling units in a zone. In studies utilizing small- to medium-size zones, it would generally be anticipated that a majority of the dwelling units in the zone would not contribute trips to any specific interchange volume. In essence, this suggests that, although the parent frequency distributions associated with the estimation of zonal trip ends would generally be expected to be a substantially skewed distribution, the parent frequency distribution associated with the estimation of a given interchange volume would be expected to be still much more highly skewed than those associated with zonal trip ends. This suggests that larger sample sizes would generally be required for the assumption of normality relative to interchange volumes than the sample sizes required for the estimation of zonal trip ends. Indeed, in some instances in dealing with interchange volumes, normality cannot be assumed regardless of the sample size.

Another problem in the application of sampling theory relative to the estimation of zonal interchange volumes lies in the sheer magnitude of the problem. Since an application of sampling theory applies to the estimation of each interchange, then for each zone pair and trip purpose, there is a distinct parent frequency distribution. This suggests, for example, that when a 5% random sample is drawn from one of the zones, it represents drawing a random 5% sample from each of 1,569 distinct parent frequency distributions for a given trip purpose. If the 5% sample is being used to estimate the zonal interchange volumes for four trip purposes, then the sample essentially represents a 5% random sample from 6,267 distinct parent frequency distributions.

Since the expected error is a function of parent frequency distribution and sample size, analysis of the expected error for the three zones and the combined area for four trip purposes would essentially involve some 25,096 distinct parent frequency distributions. Of these 25,096 parent frequency distributions, 22,000 are associated with interchange volumes of zero and are, therefore, identical parent frequency distributions. Interchange volumes of zero will, of course, always be estimated exactly and will require no further analysis.

The remaining 3,096 parent frequency distributions, however, involve one or more trips. To individually analyze each of the 3,096 nonzero interchange volumes would be a formidable task with a great deal of needless redundancy. Therefore, the analysis generally focused on the nonzero interchange volumes less than or equal to ten trips. For interchange volumes of greater than ten trips, the analysis focused on four selected interchanges which reasonably represent the range of observed interchange volumes.

Sampling and Interchange Volumes of Zero

As previously stated, the interchange volumes of zero will always be estimated exactly. In urban studies involving a large number of small- to medium-size zones, a large portion of the interchange volumes will, of course, be zero. Indeed, in many studies the number of zone pairs between which trips may be interchanged substantially exceeds the number of trips in the urban area. When estimating the number of interactions (i.e., number of nonzero interchange volumes) which may be expected for a zone, one is in essence estimating the number of zero interchange volumes which might be associated with that zone. Therefore, the preceding section entitled "Interaction Analysis" addresses the problems associated with the estimation of zero interchange volumes.

Sampling and Interchange Volumes of One

The interchange volume of one is the simplest case to address relative to the accuracy of estimation from the home interview survey. The parent frequency distribution associated with an interchange volume of one would, of course, involve only one contributing dwelling unit, with the remainder of the dwelling units in the zone contributing zero trips to that given interchange volume. As was previously discussed, the distribution of the estimated means from repeated random samples for such a parent frequency distribution (regardless of the interchange volume involved) will not reasonably approximate a normal distribution regardless of the sample size involved.

Unless the sample size exceeds a sampling rate of 50%, the resulting percent error will either be negative 100% (i.e., an estimated interchange volume of zero), or will be greater than or equal to 100% (since an effective sampling rate of 50% or less implies an expansion factor of greater than or equal to 2). In short, an interchange volume of one will either be estimated as an interchange volume of zero or as an interchange volume equal to the expansion factor associated with the sampling rate.

Unfortunately, in an urban study such as San Antonio, involving a large number of small- to medium-size zones, the interchange volumes of one probably account for a significant portion of both the nonzero interchange volumes and the trips produced by the zone. It would seem reasonable to expect that the portion of these interchange volumes detected by a sample survey would be roughly equivalent to the sampling rate (i.e., with a sampling rate of 10%, roughly 10% of the interchange volumes of one would be detected as nonzero interchange volumes). With traditional sampling rates of 5% to 12.5%, only roughly 5% to 12.5% of these interchange volumes would probably be detected in the home interview survey and, those detected would be overestimated by a factor of roughly 8 to 20. In short, the expected error of estimation associated with interchange volumes of one using traditional sampling rates are, to say the least, substantial.

Sampling and Interchange Volumes of From 1 to 10 Trips

Interchange volumes involving one to ten trips may be addressed as a group relative to the expected accuracy of estimation from sample data. It can be demonstrated (assuming the zone size of 100 or more occupied dwelling units and sampling without replacement) that sampling rates substantially greater than those traditionally associated with home interview surveys are required to reasonably assume normality in the distribution of the interchange volumes. As a result, no reasonable statistical assertions may be made relative to the expected accuracy of estimation for home interview surveys. Nevertheless, the interchange volumes of from 1 to 10 trips are too important to be ignored, especially in studies involving a large number of small- to medium-size zones. As will be recalled from Table III-2, these interchange volumes accounted for 85% to 99% of the observed nonzero interchange volumes, depending on the zone and trip purpose These interchange volumes also accounted for 35% to 95% of the observed trips (again varying from zone to zone and trip purpose to trip purpose).

Since it is not possible to make any reasonable statistical inferences relative to the expected accuracy of estimation of interchange volumes of from 1 to 10 trips using traditional sampling rates, empirical data must be relied on for observations relative to the expected accuracy. The data used for this analysis were the one hundred 5% random samples and the one hundred 12.5% random samples drawn from the combined area. Tables III-4 and III-5 summarize the sets of random samples relative to the estimation of the interchange volumes from 1 to 10. The data presented are the home-based person trip interchange volumes. The data for other trip purposes did not significantly differ from data presented in the tables; the parent frequency distributions for 347 interchange volumes presented in these tables are representative of the parent frequency distributions which might be encountered in dealing with these small interchange volumes for a zone of this size.

For the 5% samples summarized in Table III-4, there is roughly 80% to 95% probability of estimating a zero interchange volume for an interchange volume of 1 to 10 trips. This suggests that on the average, roughly 92% of these interchanges would remain undetected by a sample survey using a nominal sampling rate of 5%. The percent error of estimate associated with these undetected interchange volumes is -100%.

Since the use of a nominal 5% sampling rate implies an expansion factor of roughly 20 (i.e., roughly twice the maximum interchange volume being considered), the percent error associated with detected interchange volumes would range from roughly 100% to roughly 1,900% (i.e., the resulting estimates for detected interchange volumes would be roughly 2 to 20 times as large as the actual interchange volume). In essence, at a 5% sampling rate there is a 100% probability that the error of estimate associated with any interchange volume of 1 to 10 trips will be ±100% or greater. As may be observed from Table III-4, only 28% of the interchange volumes detected were estimated to involve 20 trips, while about 82% resulted in estimates of 40 or more trips. In short, the expected accuracy in estimating any one of these interchange volumes using a 5% sampling rate is very poor.

TABLE III-4: Frequency Distribution of Estimates of Small Interchange Volumes Using a 5 Percent Sampling Rate

Actual Interchange Volume	Number of Interchanges	Estimated Interchange Volumes*											Average Estimate
		0	20	40	61	81	101	121	141	162	182	202	
1	61	95.3	4.7										0.94
2	122	94.9	0.4	4.7									1.96
3	29	92.7	2.5	2.8	2.0								2.85
4	39	91.4	1.2	5.3	0.2	1.9							4.02
5	20	89.7	3.6	3.2	0.4	1.9	1.2						4.99
6	25	88.9	2.2	4.6	0.4	1.9	0.1	1.9					6.4
7	15	86.4	2.3	7.0	0.5	1.5	1.0	0.8	0.5				7.43
8	14	85.2	4.5	5.9	1.6	1.7	0.0	0.9	0.1	1.0			7.97
9	10	81.7	5.6	5.4	1.1	3.4	0.1	1.8	0.7	0.2	0.0		10.30
10	12	82.8	2.5	9.1	1.2	2.6	0.7	0.8	0.0	0.0	0.0	0.3	9.20

* Expansion factor = 20.2

The cross classification of estimated versus actual interchange volumes observed demonstrates the low level of accuracy which might be expected in estimating such interchange volumes from 5 percent sample data.

TABLE III-5: FREQUENCY DISTRIBUTION OF ESTIMATES OF SMALL INTERCHANGE VOLUMES USING A 12.5 PERCENT SAMPLING RATE

Actual Interchange Volume	Number of Interchanges	Percent of Observations by Estimated Interchange Volumes*											Average Estimate
		0	8	16	24	32	40	48	56	64	72	80	
1	61	87.7	12.3										0.98
2	122	86.7	1.1	12.2									
3	29	81.6	5.9	6.9	5.6								
4	39	80.1	2.4	11.5	0.6	5.4							
5	20	76.5	6.6	7.4	1.9	3.8	3.8						
6	25	75.5	3.8	9.6	1.3	5.6	0.6	3.7					
7	15	70.1	4.9	12.2	2.9	3.7	2.9	2.3	1.0				
8	14	64.3	5.8	16.6	4.4	4.9	0.5	2.2	0.2	1.1			
9	10	63.2	7.6	11.5	4.3	6.5	1.2	2.9	2.0	0.5	0.3		
10	12	60.1	5.2	18.7	2.8	6.4	1.9	3.3	0.3	0.6	0.0	0.7	9.57

* Expansion factor = 8.0

The cross-classification of estimated versus actual interchange volumes observed from one hundred random 5-percent samples demonstrate the low level of accuracy which might be expected in estimating such interchange volumes from 5 percent sample data.

The distributions of the observed estimates associated with the various interchange volumes are also summarized in Table III-4. As expected, the average estimate is generally very close to the interchange volume being estimated. This simply suggests that the mean of the distribution of estimates from reapeated random samples associated with a given interchange volume would generally be expected to closely approximate the interchange volume being estimated, but the expected error associated with any given estimate remains substantial.

Table III-5 summarizes the results from one hundred 12.5% random samples. Increasing the sample size, of course, increases the probability of detecting these interchanges and reduces the magnitude of the expected error associated with the estimates (i.e., increasing the sample size reduces the variance of estimates from repeated random samples). As can be seen from this table, however, the accuracy of estimates associated with any given interchange remains very poor.

Increasing the zone size, while holding the sampling rate constant, would have very little effect in the expected accuracy associated with interchange volumes of from 1 to 10 trips. The increasing of zone sized, however, would be expected to reduce both the proportion of the nonzero interchange volumes which lie in this range and the proportion of the trips accounted for by these interchange volumes (see Table III-3).

Sampling and Interchange Volumes Greater Than 10

The proportion of the nonzero interchange volumes involving more than 10 trips, of course, may be expected to vary by zone size and trip purpose. In the case of the home-based person trips in the in the Combined Area, 13% of the nonzero interchange volumes involved more than 10 trips and accounted for 62% of the trips produced (see Table III-3). In contrast, the interchange volumes greater than 10 accounted for only 1% fo the nonzero interchange volumes associated with the home-based work auto driver trips from Zone B and accounted for only 5% of the trips produced.

For purposes of analysis, four interchanges were selected for detailed analysis which were reasonably representative of the range of observed interchange volumes greater than 10. These interchanges were selected from those associated with the home-based person trips for the Combined Area and represent interchange volumes of 16, 51, 96 and 182 trips. Table III-6 summarizes the parent frequency distributions associated with each of the four selected interchanges. As expected, each of the parent frequency distributions are highly skewed distributions.

Using the sets of random samples drawn from the combined area, the distribution of the estimated interchange volumes from repeated random samples for each of the four selected interchanges were obtained for sampling rates of 5, 12.5%, 25%, and 50%. The Kolmogorov-Smirnov test was used to determine if the distributions were significantly different from a normal distribution. The results of these tests indicated that the normality of the distributions associated with the interchange volumes of 96 and 182 trips could not be rejected at the 80% confidence level for any of the four sampling rates. The results of these tests indicated, however, that the normality of the distributions associated with the interchange volumes of 16 and 51 trips could be rejected at the 80% confidence levels for the 25% and 50% sampling rates.

Based upon the normality assumptions and the mean and variance of the parent frequency, sample size requirements may be computed for various desired levels of accuracy. The following summarizes the sample rate that would be required for the four selected interchanges for error tolerances of ±50% and ±100% at the 95% confidence level:

	Sampling Rate Required at 95% Confidence Level			
	Interchange Volume			
Error Tolerance	16	51	96	182
±50%	64%	69%	26%	23%
±100%	31%	36%	8%	7%

These figures suggest, for example, that a sampling rate of 64% would be required to estimate the interchange volume of 16 within ±50% at the 95% confidence level. They further suggest that sampling rates substantially greater than 12.5% would probably be required for any interchange volume of 50 or fewer trips in order to be 95% confident that the resulting estimate is within ±100% of the actual interchange volume. To be 95% confident that a larger interchange

TABLE III-6: PARENT FREQUENCY DISTRIBUTIONS

Trips Produced	Number of Dwelling Units I. V.* of 16	I. V. of 51	I. V. of 96	I. V. of 182
0	415	408	381	366
1	2	5	2	7
2	7	7	34	20
3		0	2	7
4		2	5	13
5		0		5
6		0		5
7		0		1
8		1		
9-15		0		
16		1		

* Abbreviation for interchange volume.

The above table illustrates the parent frequency distribution for the four selected interchanges. Shown for each interchange volume is the distribution of the 424 dwelling units by the number of trips which they contributed to the given interchange volume.

volume of, say, 100 to 180 trips is estimated within ±50% would also require, based on these estimates, sampling rates substantially greater than 12.5%.

Table III-7 summarizes the results of one hundred 5% random samples and one hundred 12.5% random samples relative to their accuracy in estimating the interchange volumes for the four selected interchanges. These observations suggest that the expected level of accuracy in using traditional sampling rates to estimate interchange volumes of from 11 to 50 trips remains very poor. Unfortunately, in most studies, the interchange volumes of 1 to 50 trips account for a large majority of the nonzero interchange volumes and of the trips in the urban area. These observations further suggest that, using traditional sampling rates, moderate levels of accuracy may be expected in estimating interchange volumes of from roughly 100 to 200 trips.

Conclusions

The following summarizes the conclusions from the interchange volume analysis:

- The characteristics of the interchange volumes associated with the three zones and the combined area generally conform with the expected characteristics based on trip distribution theory.
- In urban transportation studies involving a large number of small- to medium-size zones such as San Antonio, a large majority of the interchange volumes would generally be expected to be zero. Of the nonzero interchange volumes, a large majority would generally be expected to fall in the volume range of from 1 to 10 trips.
- Using traditional sampling rates, a large majority of the small interchange volumes of 1 to 10 trips, may generally be expected to remain undetected by the survey (i.e., an error of estimate of -100%). Those detected by the survey may generally be expected to be substantially overestimated (i.e., an error estimate generally exceeding +200%). In studies utilizing a large number of small- to medium-size zones, these small interchange volumes not only account for a majority of the nonzero interchange volumes but account for a substantial portion of the trips.

TABLE III-7: ACCURACY OF ESTIMATES FOR FOUR SELECTED INTERCHANGES

5 Percent Sampling Rate

Error Range	Percent of Samples Within Error Range Interchange Volume 16	Interchange Volume 51	Interchange Volume 96	Interchange Volume 182
± 25%	5	28	28	36
± 50%	5	28	49	60
± 75%	5	38	78	70
± 99%	5	43	82	86
±100%	66	83	97	92

12.5 Percent Sampling Rate

Error Range	Percent of Samples Within Error Range Interchange Volume 16	Interchange Volume 51	Interchange Volume 96	Interchange Volume 182
± 25%	27	15	56	52
± 50%	46	39	89	85
± 75%	46	64	95	92
± 99%	46	69	99	98
±100%	93	78	99	99

The table summarizes the results of one hundred 5 percent random samples and one hundred 12.5 percent random samples drawn from the Combined Area relative to their accuracy in estimating the interchange volumes for the four selected interchanges.

- Sampling rates of 25% and above would generally be required to be 95% confident that estimates of nonzero interchange volumes of 50 or less trips would be estimated within ±100%. In studies utilizing a large number of small- to medium-size zones, interchange volumes of 50 and fewer trips will generally account for a large majority of the trips in the urban area.

- Interchange volumes of 100 to 200 trips will generally require sampling rates of greater than 5% to be 95% confident that the estimates of these interchange volumes are within ±100 percent. Empirical data from repeated random samples indicate that when using a 5% sampling rate to estimate these interchange volumes, the error associated with roughly 40% to 50% of these estimates will exceed ±50%. Using a 12.5% sampling rate, roughly 80% to 90% of the estimates will be within ±50 of the actual volume. In essence, it was felt that traditional sampling rates would result in only moderate levels of accuracy relative to the estimation of interchange volumes of 100 to 200 trips.

TRIP LENGTH FREQUENCY DISTRIBUTION ANALYSIS

A trip length frequency distribution (TLFD) is defined as the distribution of trips by spatial separation. As in the preceding analyses, the measure of spatial separation used in this analysis is the network travel time.

The evaluation of travel patterns as employed in the State of Texas involves the use of a trip length frequency distribution directly in the trip distribution process. This is generally the distribution for the entire urban area which is obtained from the data collected in home-interview surveys. Basically, the distribution for the entire urban area may be thought of as the aggregation of the zonal distributions as estimated from the data collected in each zone. For this reason, the accuracy of sample data in the estimation of trip length frequency distributions is analyzed from two standpoints:

- On the basis of accuracy of sample data in predicting the zonal trip length frequency distribution. This is done using the 100% data for the combined area zone; and
- On the basis of the accuracy of a much smaller sample size in predicting the trip length frequency distribution for the entire urban area. This is done using the data collected in the O-D survey for the San Antonio Transportation Study.

Both of the above analyses also considered the mean trip length as estimated from sample data. Before proceeding, it will be helpful to first consider sampling theory and its relationship with trip length frequency distributions.

Sampling Theory and TLFDs

Sampling theory associated with home interview O-D surveys is applicable to the estimation of the number of trips in a given separation interval. The basic relationship between a given separation interval and the trips produced by a dwelling unit may be stated as follows:

$$T_{KS} = t_{1KS} + t_{2KS} + t_{3KS} + \ldots + t_{NKS}$$

where:

T_{KS} = total trips produced by zone K whose travel time falls within the separation intervals.

t_{iKS} = total trips produced by dwelling unit i located in zone K whose travel time falls within the separation interval S.

N = number of dwelling units in zone K.

Since the relation is additive and the individual observations are independent, sampling theory is applicable. The subsequent analyses are directed toward determining the accuracy of sampling at the zone level and the question as to the sample size required to achieve an acceptable estimate of the trip length frequency distribution for the entire urban area.

Analysis Using 100% Survey Data

Analysis of the trip length frequency distribution and mean trip length was performed using the combined area only. A computer program was developed to facilitate the analysis with the following functions being performed:

1. Dwelling units were selected randomly for analysis; 100 samples were drawn at sampling rates of 5%, 12.5%, and 25%;
2. Expected and observed trip length frequency distributions were computed;
3. Chi-square and Kolmogorov-Smirnov (K-S) tests for goodness of fit were performed in comparing the expected and observed distributions;
4. Test results for all samples drawn were tabulated;
5. For each sample set, the mean, percent error, and standard deviation were computed; and
6. For all the samples drawn, frequency distributions of the errors were developed and normality tests performed using the Chi-Square and Kolmogorov-Smirnov (K-S) goodness of fit test for the following:

- trips per dwelling unit;
- total travel time per dwelling unit; and
- the mean trip length.

The trip length frequency distributions (TLFD) generated from the sample data were quite different from the actual frequency distributions of the 100% survey data. The results of the statistical tests performed on the 100 samples drawn at 5%, 12.5%, and 25% sampling rates are presented in Table IV-1.

The hypothesis that the sample TLFD is equal to the TLFD from the 100% data, is rejected by both the Chi-Square and K-S tests a high percentage of the time. The differences between the percentage of the samples rejected by the Chi-Square test and the K-S test stem from the fact that the Chi-Square test is very sensitive to differences in the tails of distribution Since trip length frequency distributions are skewed with rather long tails, one would expect to find more samples rejected by the Chi-Square test. For this reason, the K-S test is generally considered to be the better of the two tests for analyses of these data.

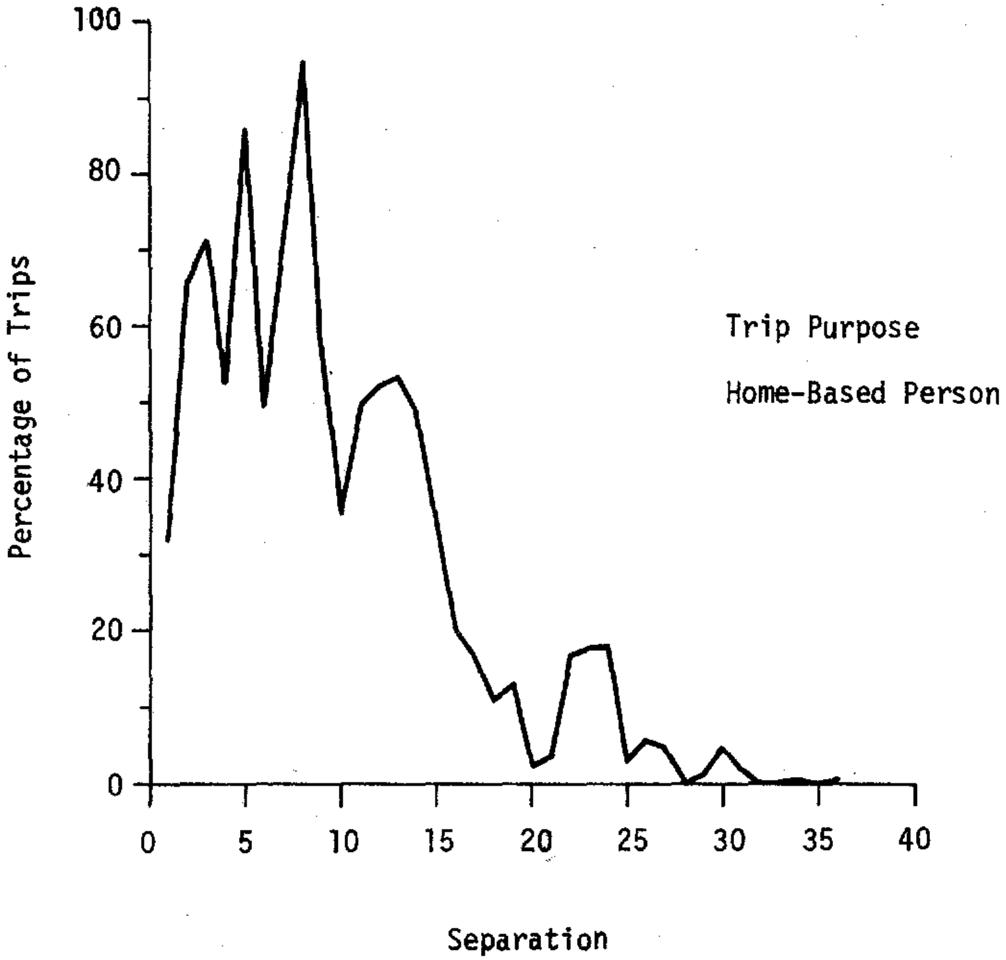
It was concluded that the observed trip length frequency distribution from the 100% data was not satisfactorily estimated from sample data. There should be no inference with respect to the use of sample data in estimating the trip length frequency distribution for the entire urban areas. This stems from the possibility of errors on the zonal level having a canceling effect when aggregated to estimate the distribution for the urban area.

TABLE IV-1: RESULTS OF STATISTICAL TESTS

Trip Purpose	HBP			HBAD			HBWAD			HBNWAD		
Sampling Rate	5%	12.5%	25%	5%	12.5%	25%	5%	12.5%	25%	5%	12.5%	25%
Chi-Square	99%	99%	99%	91%	89%	73%	60%	91%	73%	83%	79%	72%
Kolmogorov-Smirnov	75%	75%	62%	47%	38%	26%	36%	33%	30%	41%	42%	18%

The observed trip length frequency distributions (TLFD) from 100 random samples taken at sampling rates of 5%, 12.5%, and 25% were compared to the population TLFD by using the Chi-Square and Kolmogorov-Smirnov goodness-of-fit tests. Using a confidence level of 80% and the null hypothesis that the sample TLFD was the same as the population TLFD, the percentage of the samples drawn where the null hypothesis was rejected was computed.

FIGURE IV-1: HOME-BASED PERSON TRIP LENGTH FREQUENCY DISTRIBUTION



The figure shows the trip length frequency distribution for HBP trip productions for the combined area. It represents the population distribution from which samples were drawn. As will be observed, the distribution is quite jagged and illustrates the difficulty in predicting the trip length frequency distribution for a given zone from sample data.

With expansion factors for the subsamples being calculated for individual zones, the survey trips from Zone A may be expanded by a different factor than those from Zone B and the effective sampling rate will vary between zones. The expansion factors are computed in the following manner:

$$DUF_i = \frac{N_i - [Ni/Si\ (Vi)]}{S_i - (V_i + R_i)}$$

N_i = Total number dwelling units in Zone i

S_i = Number of dwelling units selected to be interviewed in Zone i

V_i = Number of vacant dwelling units encountered in the attempt to interview S_i dwelling units in Zone i

R_i = Number of refusals, no contacts, etc. encountered in the attempt to interview S_i dwelling units in Zone i

DUF_i = Dwelling unit factor for Zone i

The trip factors are then computed using the following:

$$TF_i = DUF_i \frac{P_i}{P_i - U_i}$$

DUF_i = Dwelling unit factor for Zone i

P_i = Total number of persons 5 years and older in the interviewed dwelling units in Zone i

U_i = Total number of persons 5 years and older making trips that were unknown to the person being interviewed dwelling units in Zone i

Since the effective sampling rate varied between survey zones, the dwelling units interviewed represented varying numbers of other dwelling units. Thus, if the sample were selected at random from among the interviewed dwelling units, two courses of action would be open. One would be to weigh each dwelling unit equally with no expansion. Thus, each would have an equal probability of being selected, but the total urban area would not be represented.

The other approach would be to expand the trips from the subsampled dwelling units using each zones respective trip factor. This would imply that each dwelling unit sampled in the 5% O-D survey would represent a cluster of dwelling units, each having the same trips. The number of dwelling units represented in each cluster would vary between zones as the trip factors varied. If a random sample of the sampled dwelling units from the 5% O-D survey were drawn, the number of dwelling units represented in the urban area could be large and would also vary between samples. Thus, each dwelling unit in the 5% O-D survey would have an equal chance of being included in a given subsample, but each dwelling unit in the urban area would not have an equal chance.

A procedure which would stimulate sampling from the entire urban area is illustrated by the following example of 5 zones with a total of 400 DU's.

Zone No.	No. D. U. Interviewed	D. U. Factor	Est. No. D. U.	Cumulative Total
1	5	20.0	100	100
2	2	20.0	40	140
3	20	5.0	100	240
4	10	10.0	100	340
5	3	20.0	60	400

1. Randomly select a number between 1 and 400; for this example, say, the first number picked is 150;
2. Looking into our above table under the cumulative column, the number 150 lies in Zone 3;
3. Then the first dwelling unit selected will be one of the twenty interviewed in survey Zone 3;
4. Randomly select a number between 1 and 20, for this example, say, that number is 5;
5. Then the first dwelling unit in the sample will be the 5th dwelling unit surveyed in Zone 3; and
6. Repeat steps 1 through 5 until a total of 20 dwelling units have been selected for analysis.

Mean Trip Length

If the distribution of the estimates of the mean trip length are normally distributed, the sample size needed to estimate the mean trip length with a given margin of error and confidence level can be determined, using the technique outlined in Appendix B. Consequently, an appropriate analysis was conducted to determine if such was the case.

The Chi-Square and Kolmogorov-Smirnov (K-S) tests were used to test the normality of the distribution of errors using 100 samples randomly drawn at 5%, 12.5%, and 25% sampling rates. Results of the statistical test on the distribution of errors of the mean trip length are summarized in Table IV-2. The null hypothesis could not be rejected for any of the sampling rates or trip purposes by the K-S test at the 80% confidence level.

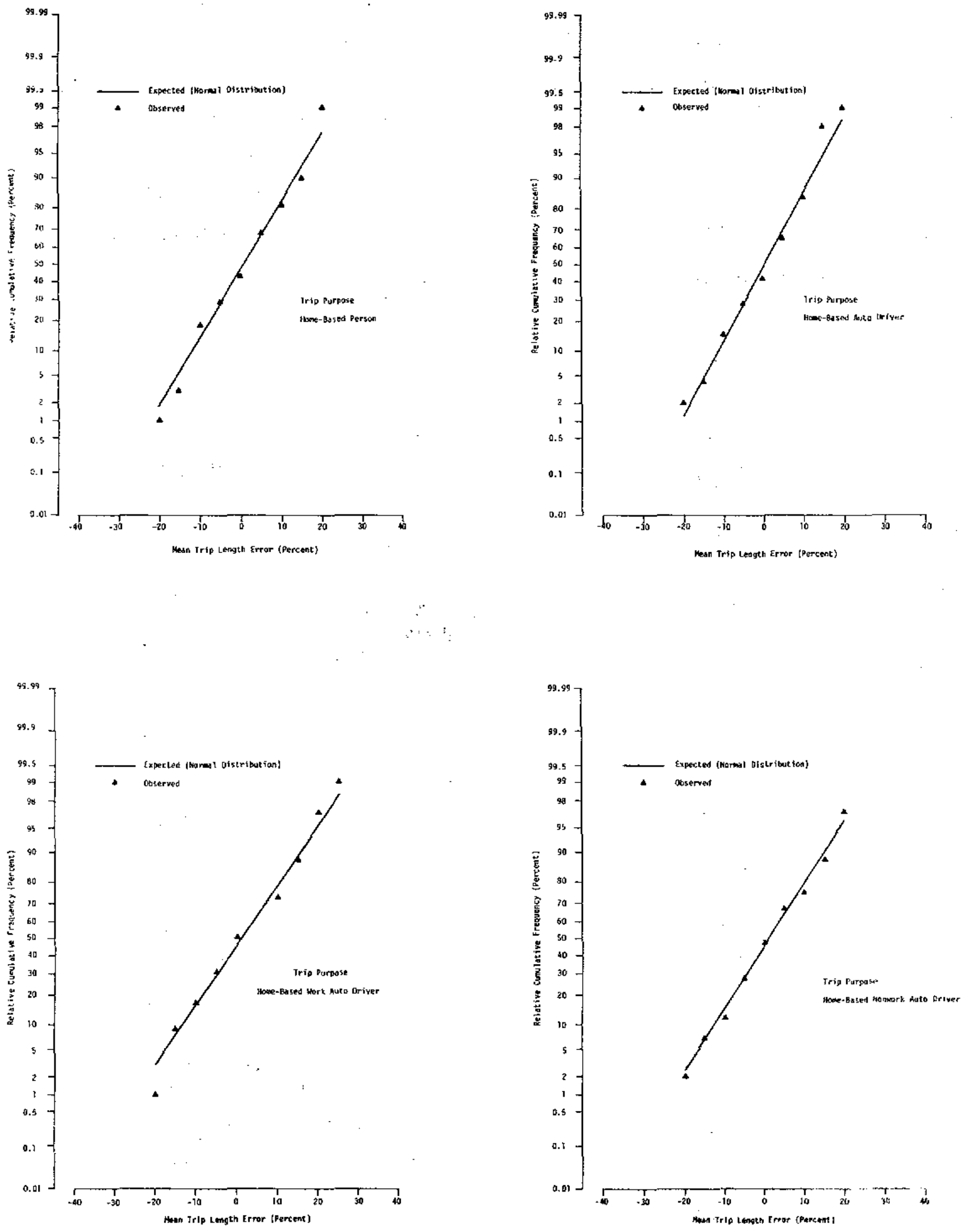
TABLE IV-2: NORMALITY TESTS ON ERROR FREQUENCY

Trip Purpose	HBP			HBAD			HBWAD			HBNWAD		
Sampling Rate	5%	12.5%	25%	5%	12.5%	25%	5%	12.5%	25%	5%	12.5%	25%
Rejected by Chi-Square	No	No	No	No	No	No	No	No	No	No	No	No
Rejected by Kolmogorov-Smirnov	No	No	No	No	No	No	No	No	No	No	No	No

Verification of the normality of the frequency distribution of errors in the mean trip length estimates was accomplished by drawing 100 random samples at sampling rates of 5%, 12.5%, and 25%. The resulting frequency distribution of errors was computed for each trip purpose and sampling rate and compared with the expected distribution based on a normal distribution. The comparison was made using the Chi-Square and Kolmogorov-Smirnov goodness-of-fit tests. Using a confidence level of 80% and the null hypothesis that the frequency distribution of errors was normal, the tests were applied. The above table summarizes the results of those applications for each trip purpose and sampling rate.

Probability plots of the relative cumulative frequency distribution of errors are shown on Figure IV-2. Two plots are shown for the sampling rate of

FIGURE IV-2: PROBABILITY PLOTS FOR 100% DATA



The above figures show the probability plots of the relative cumulative frequency distribution of mean trip length errors as obtained from 100 samples drawn at the sampling rate of 5%. Shown for each trip purpose are the observed values and the expected values (i.e. solid line) based on a normal distribution. As may be seen, the observed values closely approximate the expected values, thus indicating the distribution of errors is normal.

5%; one is of the average observed values from the 100 samples and the other is the expected values (based on a normal distribution). The normality of the observed values is indicated by how well those points fit the straight line formed by the expected values. As can be seen, the observed values fit the straight line formed by the expected values very well. The slope of the line in Figure IV-2 is a measure of the variance in the distribution.

Since the hypothesis that the errors are normally distributed can be accepted, the formulas presented in Appendix A were applied to evaluate the sample size required to estimate the mean trip length. The relationship between percent error and the percent sampling rate for the various trip purposes analyzed is shown in Figure IV-3. These curves indicate that a relatively good estimate of the mean trip length may be obtained at the smaller sampling rates.

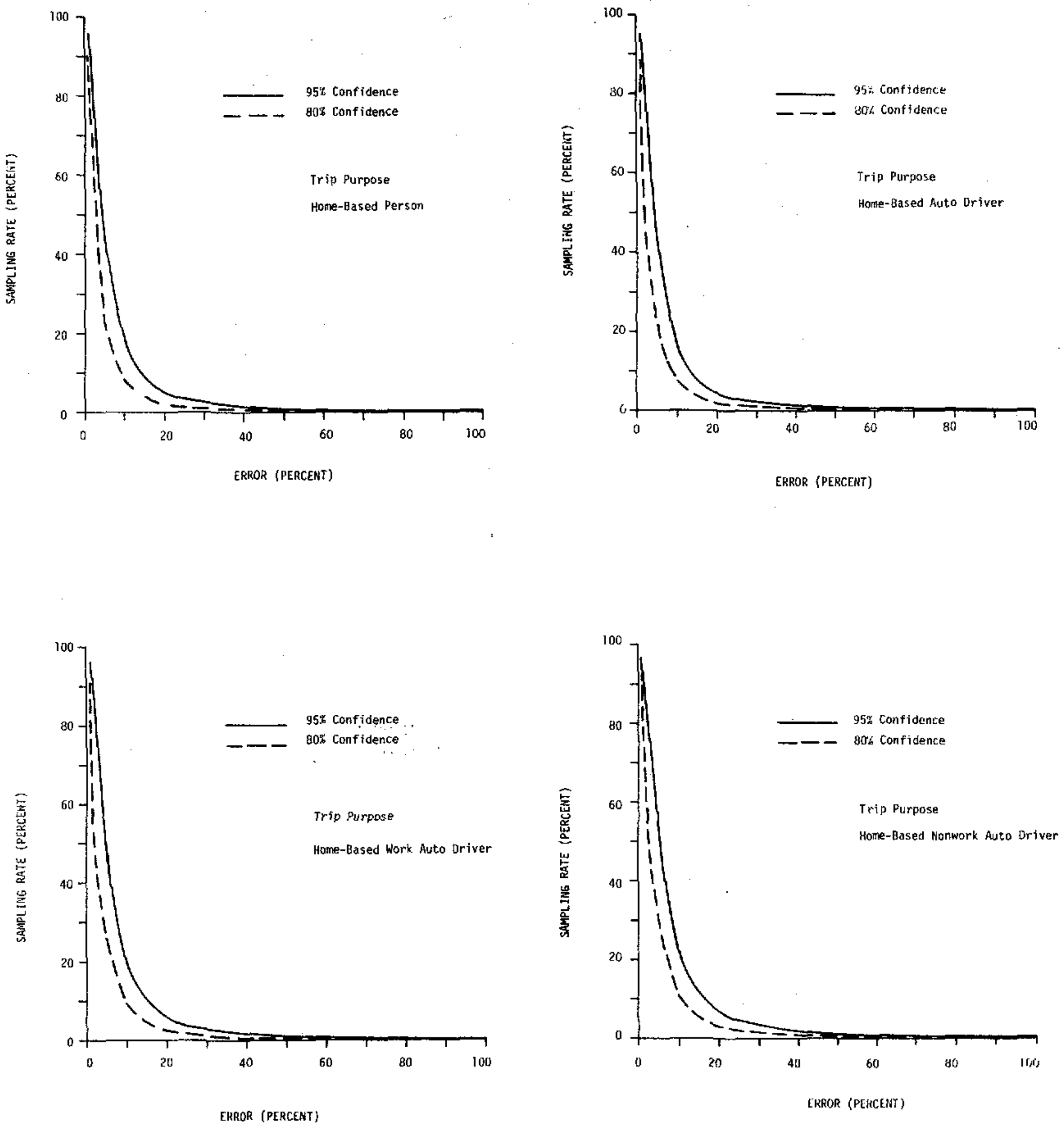
Summary

In summarizing, it was found that samples provided, at best poor estimates of the trip length frequency at the zonal level. However, no inference should be drawn as to the accuracy of the estimate of the trip length frequency distribution for the entire urban area. The probable reason for the poor estimates lies in the population distribution from which the samples were drawn, an example of which is shown in Figure IV-1. As can be seen, the distribution is somewhat jagged, as might be expected, considering the zone contained only 424 dwelling units. With regard to the estimation of the mean trip length, the analysis performed indicated that reasonable estimates may be obtained using sample data. This supports the application of a similar analogy regarding the mean trip length for the entire urban area.

Analysis of San Antonio O-D Study Data

Further analysis of sample size necessary to predict the trip length frequency distribution and length was performed using the 5% survey data collected as part of the San Antonio-Bexar County Urban Transportation Study. The analysis was performed using the same trip purposes (HBP, HBAD, HBNWAD) as used in the analysis of the 100% survey data. However, the subsamples were selected using a different procedure than used with the 100% data since these samples were drawn from a sample data set. The ideal sampling procedure for

FIGURE IV-3: SAMPLING REQUIREMENTS FOR 100% DATA



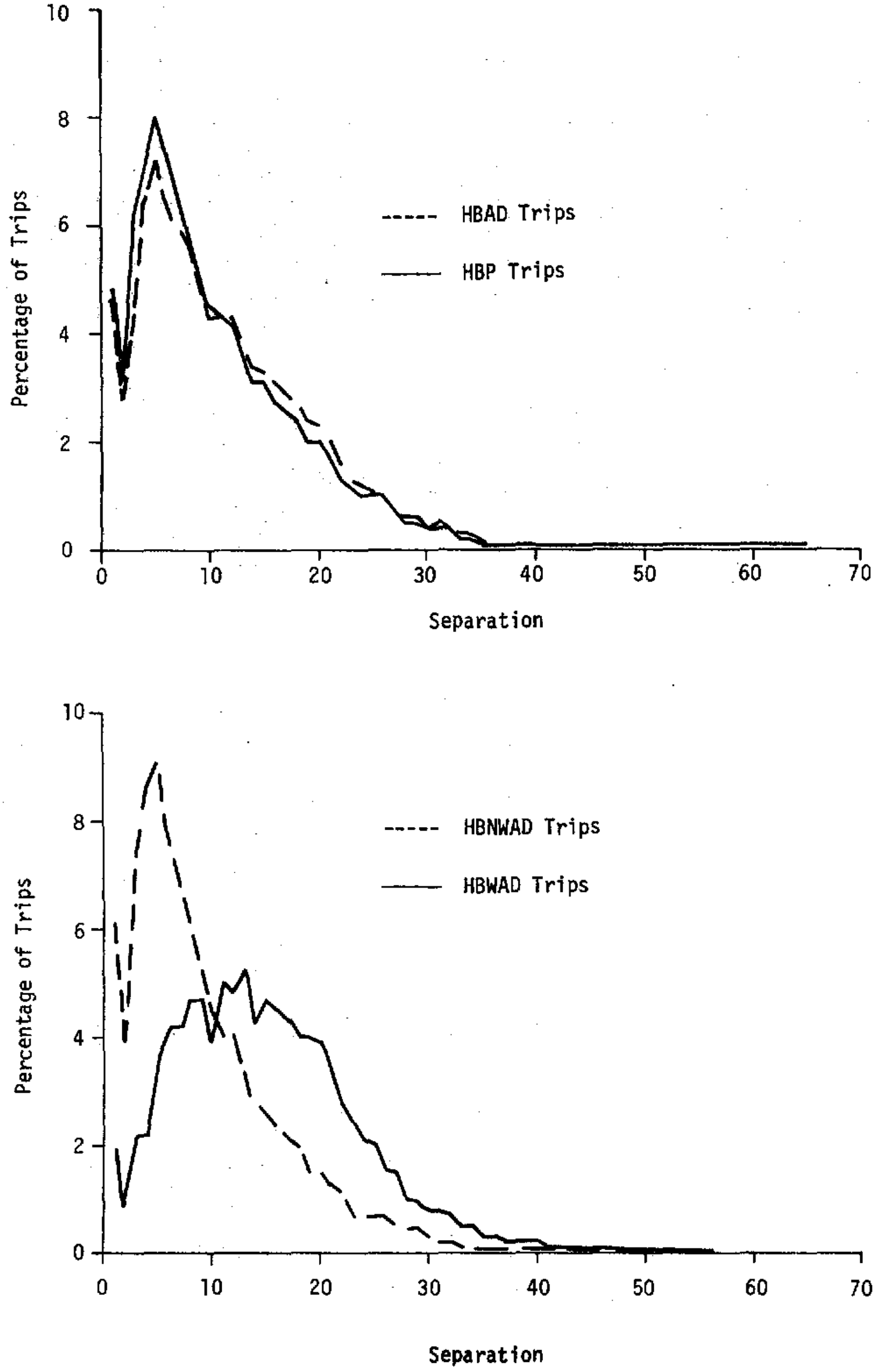
These figures show the sampling rates required for a given level of accuracy at confidence levels of 80% and 95% in estimating mean trip length at the zonal level. The curves are based on the error distribution being normal (as previously shown) and are computed using the relationships illustrated in Appendix A.

such an analysis would be random sampling without replacement with each dwelling unit in the urban area (not just those from the home interview survey) equally likely. To attempt to simulate this condition, to the degree possible having only sample data available, a two stage random selection procedure was used in drawing a dwelling unit observation for a sample:

- Stage 1: Initially each dwelling unit in the urban area (not just those in the survey data) was assigned a unique number and the zone containing the dwelling unit identified. The dwelling units for a sample were randomly selected (without replacement) from the list of dwelling units in the urban area and the desired number of observations from each zone tabulated.
- Stage 2: Having tabulated the desired number of observations from a given zone for a given sample, the observations used were randomly selected from the available dwelling unit observations in the zone.

Three sets of 100 random samples were drawn representing sample sizes of 400, 800 and 2,000 dwelling units. Each sample drawn was used to estimate the trip length frequency distribution for each of the four trip purposes. For each trip purpose, the "observed" trip length frequency distribution from the complete O-D survey (illustrated in Figure IV-4) using the Chi-squared and Kolmogorov-Smirnov goodness of fit tests at the 80% confidence level. The results of the tests are summarized in Table IV-3. Of the 1,200 distributions tested, the Chi-squared test indicated statistically significant differences in all but 15 of the distributions tested. This unusually high rejection rate is probably largely attributable to the sensitivity of the test to minor variations in the tails of the distributions. The Kolmogorov-Smirnov (K-S) test is felt to be a more powerful test since it is less sensitive to such minor tail variations. Nevertheless, the tests indicated statistically significant differences in 643 of the 1,200 distributions tested. This was probably largely attributable to the use of an 80% confidence level which is somewhat low for such tests. Substantially fewer distributions would have been rejected (i.e., indicated statistically

FIGURE IV-4: POPULATION FREQUENCY DISTRIBUTIONS



The above figures show the trip length frequency distribution for each trip purpose as observed from the O-D survey conducted during the San Antonio Transportation Study.

TABLE IV-3: STATISTICAL TEST RESULTS

	Trip Purpose											
	HBP			HBAD			HBWAD			HBNWAD		
No. Dwelling Units Sampled	400	800	2,000	400	800	2,000	400	800	2,000	400	800	2,000
Percentage of Samples Rejected by Chi^2	100	100	100	99	100	100	99	99	100	98	99	91
Percentage of Samples Rejected by K-S	80	78	76	50	48	42	47	44	38	52	48	40

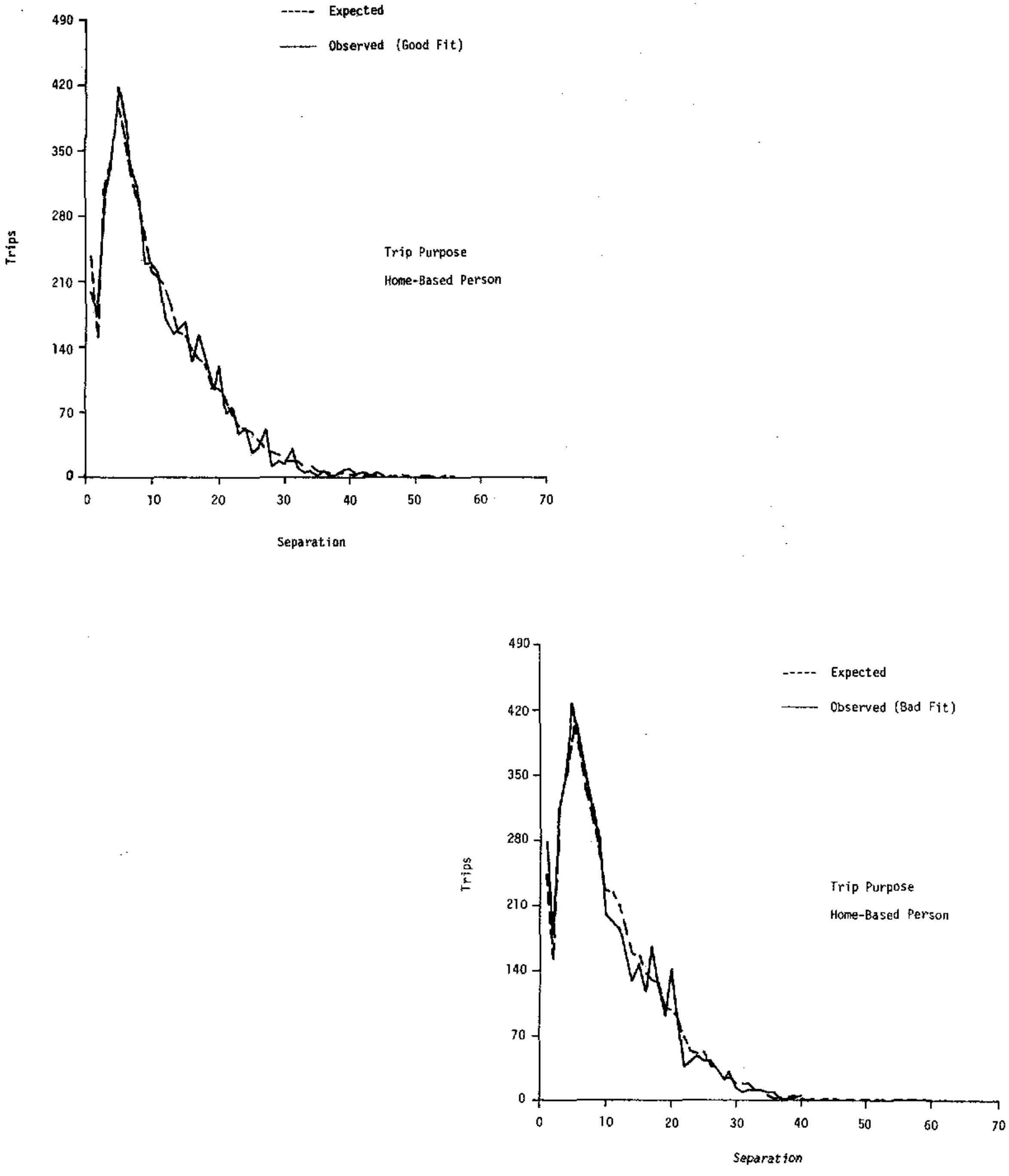
The observed trip length frequency distributions (TLFD) from 100 random samples of 400, 800, and 2,000 dwelling units were compared to the TLFD found in the San Antonio 5% O-D survey by using the Chi-Square and Kolmogorov-Smirnov goodness of fit tests. Using a confidence level of 80% and the null hypothesis that the sample TLFD was the same as the population TLFD, the percentage of samples drawn where the null hypothesis was rejected was computed.

significant differences) with confidence levels of 90% and 95%. Figure IV-5 illustrates two of the distributions: one of which was "accepted" and one was "rejected" by the K-S test. In the case of the rejected distribution, it is interesting to note that much of the significant differences would be eliminated by an analyst by simple smoothing the distribution.

Mean Trip Length

Recognizing that a smoothing technique can eliminate much of the irregularities that might be encountered from the use of very small samples, a much more critical concern is the accuracy of estimates of the mean trip length. The relationship between sample size and the expected variance of estimates from sampling theory assumes that the distribution of estimates from repeated random samples are normally distributed. Having estimates of the mean trip lengths (by purpose) from the sets of 100 random samples drawn at each of the sample sizes studies (i.e., 400, 800 and 2,000 dwelling units), provided an opportunity to test the normality assumption at these sampling levels. Testing the normality of these using the Kolmogorov-

FIGURE IV-5: SAMPLE ESTIMATES OF TRIP LENGTH FREQUENCY

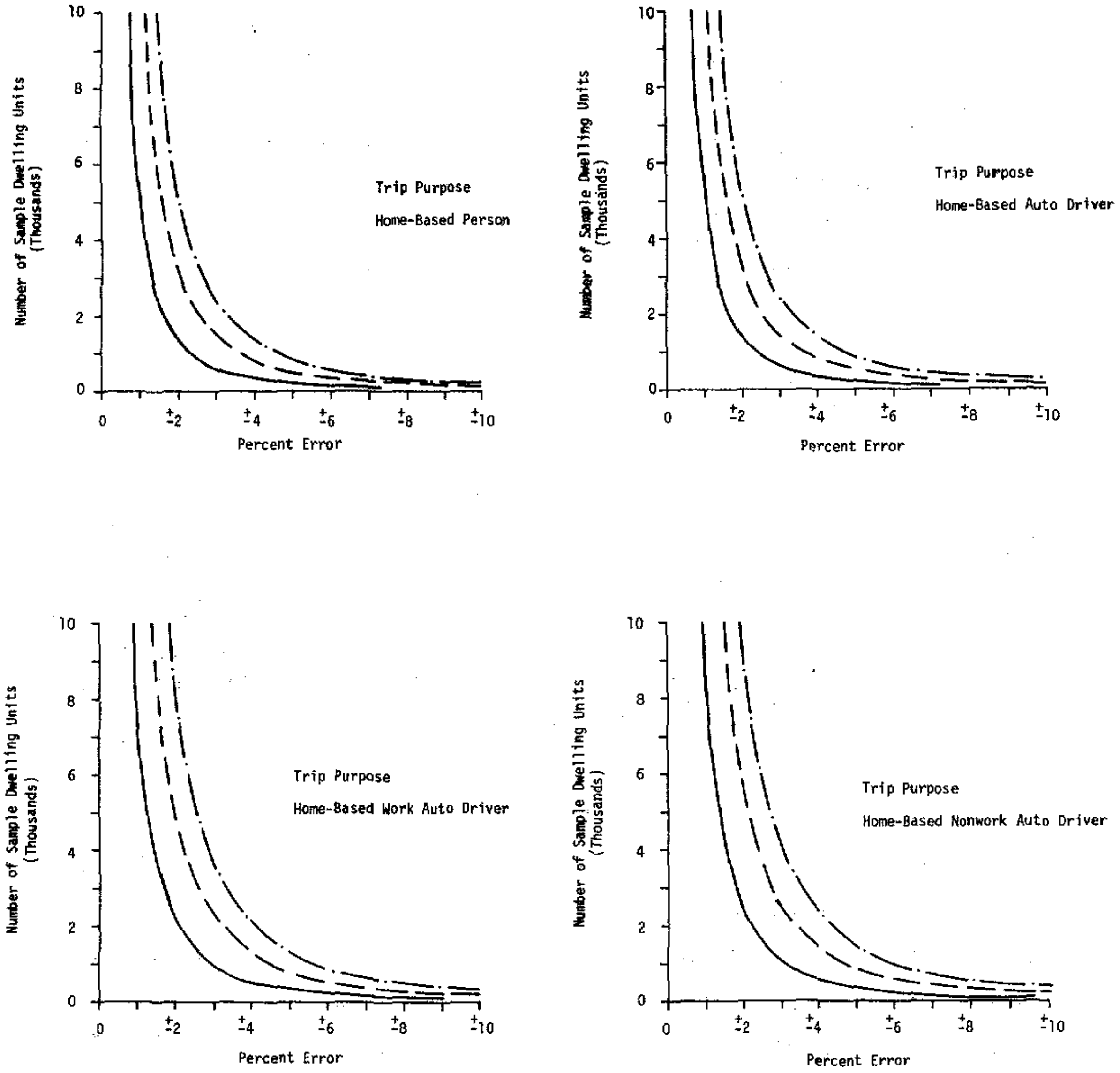


The above figures are representative plots of observed trip length frequency distributions versus the expected distributions as obtained from samples of 800 dwelling units. The top figure shows a sample where the K-S goodness of fit test indicated a good fit and the bottom figure shows another sample where the K-S test indicated a bad fit. Both plots are for home-based person trips.

Smirnov test at the 90% confidence level did not indicate statistically significant differences. The relationship between sampling rate and the expected error, which assumes the normality of the distribution of estimates, as described in Appendix A. The sampling theory used also assumed random samples of trips rather than dwelling units, however, to assume a random sample of 400 or more dwelling units represents a random sample of trips is not a serious violation of the basic assumptions of the statistical theory used. The relationship between sample size and the expected percent error ranges for the four trip purposes is illustrated in Figure IV-6. These relationships were compared with the results observed from the sets of random samples drawn. These comparisons indicated the validity of the application of the relationships. For example, 95% of the samples provided estimates within the expected error ranges indicated for the 95% confidence level.

Reviewing the results from the sets of random samples indicated that sample sizes of 400 dwelling units and above generally achieved acceptable levels of accuracy relative to the estimate of the mean trip lengths by purpose. Further, much of the irregularities in the trip length frequency distribution that might be encountered in using such samples can be overcome either by a smoothing technique or a modeling technique which uses the estimated mean trip length to estimate the trip length frequency distribution. Indeed, estimating trip length frequency distributions from small sample data using such approaches would be expected to produce sufficiently reliable estimates for modeling the urban travel pattern. An important implication of these findings is that, at sample sizes of 400 to 600 dwelling units, it may be practicable to consider to periodically draw such samples to monitor the trends in mean trip length over time.

FIGURE IV-6: SAMPLE SIZE REQUIREMENTS



The above figures show the sample size (in number of dwelling units) required for estimating the mean trip length for an urban area within a given level of accuracy at confidence levels of 80%, 95%, and 99%. The curves are computed based on the relationships shown in Appendix A. As will be noted, a high level of accuracy may be achieved with a relatively small sample.