THE USE OF DIVERSE SAMPLING PLANS FOR THE COLLECTION OF TRANSPORTATION DATA

Paul Rackow, Tri-State Transportation Committee

I INTRODUCTION

The problem of improving transportation systems for the growing number of urbanized regions throughout the United States has been receiving increased attention in recent years. Great impetus to Urban Area Transportation Studies was given through the passage of the Federal Highway Act of 1962. This legislation provided for the use of federal-aid highway money in the planning of integrated region-wide transportation systems. A region is analogous to the definition of a Standard Metropolitan Statistical Area (SMSA) as used by the Bureau of the Census in the 1960 Census of Population.

In August 1961 an informal arrangement between the Governors of New York, New Jersey, and Connecticut led to the formation of the Tri-State Transportation Committee. The three State Governments in conjunction with various Federal Agencies mutually agreed upon appropriate financing arrangements. Membership on the Committee included not only representatives from each of the three State Governments but also included participants from the U. S. Bureau of Public Roads, the U. S. Housing and Home Finance Agency, the Federal Aviation Agency and the New York City Planning Commission. Local cooperating committees representing cities and counties within the Tri-State Region are important adjuncts to the Committee. Following the structuring of the Organization, the next step was the completion of a Prospectus for the Committee which would outline the projects to be undertaken in order to fulfill its mission. In its own words the purpose of a transportation plan is to provide for "the expeditious movement of persons and goods throughout the region which is essential for the continued economic growth of the 17 million people who presently reside within it."

As might be expected, the first step in the process of transportation plan development is the large scale collection of relevant data. Previous urban area studies have shown the existence of an intimate functional relationship between trip generation and population, socio-economic, vehicle ownership and land use characteristics of a region. The projects assigned to the Data Collection Section were the design and field work associated with a number of travel inventories. These inventories are to determine the current functional relationships among the variables mentioned above for use in the projection of trips by mode and purpose in future periods. Hence while the survey results themselves are answering enumerative questions, the use to which the data will be used in transportation planning involves an analytic study. Thus, the survey is both enumerative and analytic.

II SURVEY SAMPLING PLANS

The Travel Inventories comprise three areas of Data Collection:

1. Home Interview Survey
2. Truck-Taxi Survey
3. External Cordon Line Survey

The aim of the Travel Inventories is to describe the totality of trips made by persons and vehicles via all possible modes of travel on an average weekday within a defined Cordon Area. The Cordon Area is defined as that portion of the total Tri-State Region which contains now or is expected to contain by 1980 most of the urban development. The line encircling this area is the frame within which we wish to estimate the totality of trips. Each of the three inventories provide a segment of the total trips made within the Cordon. The Home Interview Survey provides those trips originating at the living places within the Cordon by all modes of travel except trucks and taxis. These trips are estimated through the Truck-Taxi Survey by sampling from Motor Vehicle Bureau files. The remaining segment of trips made within the Cordon Area originate outside the Cordon. The External Cordon Line Survey samples trips made by automobiles, trucks, and taxis crossing the Cordon that are destined inside or outside the line. Those trips found to be garaged within the Cordon Area are duplicated by results from trips made either by the Home Interview or Truck-Taxi Survey. These must be eliminated before the results from the three surveys can be added together.

This paper does not propose to discuss the survey methods associated with the External Cordon Line Survey. That survey was the responsibility of each of the three concerned State Highway Departments. Basically, the sampling plan involved the collection of at least 10% of the trips made across the Cordon Line via all roadways with an average of 2,000 vehicles or more per day. For safety and other practical reasons, it was felt that a probability mechanism was not applicable to the type of roads concerned. It is the opinion of the author that a modification of a probability sampling technique described in a paper he and others presented to the American Statistical Association in 1960 may be applicable even here.[1]

[1] Leslie Kish, Warren Lovejoy and Paul Rackow "<u>A Multi-State Probability Sample for Traffic Surveys</u>" Presented at the 1960 Annual Meeting of the American Statistical Association.

A. Sources Available for Sample Selection

There are a number of sources available for selection of the Home Interview Survey in the New York-New Jersey-Connecticut Tri-State Region. These range from various field listing techniques to the use of Sanborne or other maps, Reverse Telephone Directories, or the records of the various Utility Companies that cover the region. Typically, Urban Area Studies have used the results of a Land Use Inventory which collects data on the development of the land including a detailed listing of all living places within the area as a means for a selection of a sample of households. However, the situation at Tri-State required that both the Land Use Inventory and the two Travel Inventories be conducted almost simultaneously.

In the case of the Truck-Taxi Survey, the source for selection was the records of Truck and Taxi Registrations within the Cordon Area of the three concerned States. Each State presented unique problems of sampling due to both the form in which their records were kept and their willingness to cooperate in processing the data.

B. Requirements of the Sample Design

The sample design for the two travel inventories had to be economical and practical in order to meet the necessary requirements for speedy field work and reasonable workloads, conditions of work and overall cost.

There were a number of specific requirements which had to be met by the Survey Design:

(1) Sample estimates had to be produced for a great variety of characteristics. These included such things as trips between geographic areas, trips by mode of travel, purpose of trip or land use category. In additions, such items as car ownership rates and incomes distributions for various household types by varying population density groups in the region was necessary as well. These estimates were for trips made on an average weekday over the period of the survey.

(2) The sampling frame included all housing units and other special dwelling places within the defined Cordon Area for the Home Interview Survey and all trucks and taxis registered with the three State Motor Vehicle Departments within the Cordon Area for the Truck-Taxi Survey.

(3) The sample was a probability sample requiring that every housing unit (or special dwelling place) or truck-taxi registration have a known probability of selection achieved by randomized selection procedures.

(4) Built into the sample design were procedures whose purpose was to measure the sampling errors associated with the various sample estimates to be made.

(5) The statistical reliability required in the survey specified that with the expected overall sample of 450,000 trips, a sample estimate of 1% or 4,500 trips, would have a coefficient of variation of the order of 2.5% or less.

The design had to be tailored to the practical problems inherent in the actual conduct of a Home Interview or Truck-Taxi Survey, and these considerations necessitated the use of the following procedures:

(6) The survey budgets permitted the use of some 250 interviewers, control clerks, and other administrative personnel to collect the necessary data over a relatively short time period. The Home Interview Sample was randomly assigned to 80 travel dates while the Truck-Taxi sample was spread over 91 dates.

(7) The survey designs included a 1% probability sample of housing units and other special dwelling places for the Home Interview Survey and a 3% probability sample of appropriate truck and taxi registrations.

(8) To improve coverage of different origin and destination patterns and mode of travel usage, both samples are randomly spread over time in such a way as to insure geographic dispersion for each weekday assignment.

(9) Average workload per interviewer ranged from about eight home interview assignments in the City of New York per weekday to a lesser amount in the more suburban areas. Assignments usually averaged 7 sample units for the Truck-Taxi Survey.

C. Field Survey Procedures

The Home Interview Survey was designed to obtain weekday trip information and other socio-economic data from a sample of persons residing in housing units and other special dwelling places. As a means to this end, the frame for sampling purposes was defined as living places located within the Cordon Area. Sample units selected were randomly allocated to unique weekday travel dates over an 80 day period. The field interviewer was assigned to obtain the necessary data from each person 5 years of age and older residing at the living place for the randomly chosen travel date only. This was accomplished by visiting the residence of the day following the travel date after a letter announcing the intended visit had been sent a week earlier. Up to three days after the travel date assignment was allowed for completion of the interview.

Similarly, the Truck-Taxi Survey was designed to obtain weekday trip and other data from a sample of trucks and taxis registered within the Cordon Area.

The sampling frame comprised trucks and taxis registered with each of the three State Motor Vehicle Departments. After selection of the sample and random allocation to 91 travel dates, field interviewers were required to complete their questionnaires by oral interview of truck or taxi owners, dispatchers, drivers, etc. within three days after the assigned travel date. As in the Home Interview Survey, letters to prospective respondents were sent well in advance to insure a maximum of cooperation.

D. Types of Sampling Plans

1. Home Interview Survey

A number of sampling frames were used as sources for selection of a one-percent probability sample of living places in New York-New Jersey-Connecticut Cordon Area.

New York City

In cooperation with another public agency which had already completed the necessary preparatory work a one-percent clustered area probability sample was selected. The sampling frames were defined by two strata:

a. The civilian, non-institutional population living in housing units and other special dwelling places in existence according to the 1960 Census of Population and Housing.

b. Housing units built since the 1960 Census up to and including February 28, 1963 as represented by occupancy certificates obtained from the New York City Department of Buildings

Blocks, as defined by the Bureau of the Census, are assigned measures of size approximately equal to one-eighth the number of housing units found in the 1960 Census. These measures of size are always whole numbers. A block with a given measure of size is said to contain that many clusters of housing units which, hopefully, average approximately eight housing units.

Some of the defined Census blocks received no independent measures of size. A block with less than six housing units recorded in the 1960 Census was arbitrarily joined to an adjacent block that contained six or more housing units. Such linked blocks were treated as single pseudo-blocks.

Certain blocks were excluded in whole or in part. Census blocks wholly excluded are those within military areas and blocks along the waterfront which yield a largely transient population depending on the number of boats docked. Blocks excluded in part are those portions containing institutional population as determined by the Census. However, housing units for institutional staff members were not excluded.

Non-institutionalized persons living in group quarters were converted into measures of size on the basis of twenty-five people per measure. These measures were added to that for housing units to obtain a total measure of size for the block.

After all conversions to measures of size were completed, geographic area sequence was randomized as follows: First, a random arrangement of boroughs and within each borough a random ordering of health districts was defined. Finally, within health districts, a random sequence of health areas was determined. As a rule, a health area contains more than five census tracts. Upon completion of this randomization procedure, every 100th measure of size was selected following a random start. In the borough of Richmond, every 50th measure was selected and later, after definition of the interviewing cluster, a random half per cluster was taken so that each cluster averaged four housing units.

It is felt that the randomization procedure used reduced to insignificance the likihood that final ordering of the areas contained any periodicity in the occurrence of characteristics of the population that would be a multiple of the sampling interval. Hence, this systematic sample of measures of size attached to randomly ordered areas may be considered as a simple random sample of measures of size.

In New York City, whenever a new residential structure is constructed or an old structure is converted or modified, occupancy permits must be secured from the Department of Building before such residential quarters may be occupied. These permits are a source for determining new construction since some base period such as the April 1960 Census.

Data on occupancy permits issued since the 1960 Census up to an including February 1963 were secured and included physical location(address and apartment number) and the number of residential units involved.

The structures involved were arranged in tax block sequence within each borough. Measures of size, approximately equal to one-eighth the number of units covered by occupancy permits, were assigned to each tax block, and every 100th measure was selected after a random start.

If one or more of the structures on the census block were included in the occupancy permit frame, whether or not they were selected from that frame, the census block was considered to contain all housing units outside the occupancy permit structures. Thus, if a sample hit were made in the selection of measures from stratum A (1960 Census) turns out to be in a structure which is a member of the occupancy permit frame, that measure is given a "no units" designation and is not covered in the field interviewing. Also, "no units" designations can result for measures selected from stratum A but which were demolished since the 1960 Census with no subsequent construction of residential quarters until after the interviewing period.

Those Census blocks containing selected measures of size had to be listed in detail by field personnel. After listing all housing units within the block, they were grouped into segments equal to the number of measures previously assigned to the block. A great effort was made to define the measures with a maximum of physical contiguity to insure unique identification of the segment assignment. Then preceding in a logical, pre-established sequence, the appropriate selected measure was determined.

Definition of the sample from the frame of occupancy permits was relatively simple. No comparison had to be made to exclude census block population. The unique physical limits of the selected measure had to be determined.

A modification of the sample design described above was introduced to reduce sampling error in that portion of the sample containing hotel and motel living places. This was done because of the Census Bureau definition of housing unit and our desire to sample persons living in transient hotel and motel rooms.

Another sampling frame was defined consisting of hotels and motels in New York City having 100 or more rooms assigned to transient guests. Measures of size equal to approximately one-tenth the number of transient guest rooms were assigned to each hotel and motel in the frame. After geographic stratification of the frame, every 200th measure was systematically selected in two independent replications. A precise definition of transient status was given to each interviewer and all assignments were allocated to travel dates at random over a time period. Interviewers were instructed to interview only in rooms within the assigned measure containing transient guests. The introduction of this new sampling frame had an effect on the major sample described previously. The effect was as follows:

(1) All previous sample measures that fell in hotels or motels included in the transient sampling frame are to be treated specially. Only rooms within the defined measures which contain permanent guests were to be included in the sample and interviewed. Others were to be considered out-of-scope.

(2) All other previous sample measures that fell in hotels or motels which are not members of the transient sampling frame were to be interviewed in entirety as a catch-all measure. That is, all persons living in rooms defined by the measure are to be interviewed whether they are transient or permanent guests.

In summary, there are then three types of sample measures defined for hotels and motels in New York City:

(1) Transient Measures
(2) Permanent Measures
(3) Catch-all Measures

Outside New York City

The necessity for a speedy, inexpensive sample selection scheme required the use of the records of the various Electric Utility Companies as a sampling frame for most of the remainder of the Cordon Area outside New York City. Although initial costs were insignificant (all utilities cooperated as a public service with no charge to Tri-State) and the sample was selected quickly so that processing for field assignment could proceed rapidly, other methodological problems arose which led to considerable expenditure of time and money.

Utility company records are broadly split into two groups which are of interest to us in sampling living places:

(1) Residential rate customers usually representing individual housing units in the Census definitional sense.

(2) Commercial or General Service rate customers representing multi-unit housing units in the Census definitional sense.

A residential account or meter may represent two or more housing units (living places) in some cases but generally are in bi-unique correspondence to one housing unit.

However, commercial rate accounts range from representing an office in a commercial building with no housing units to a motel representing 50 separate family units. Commercial or General Service rate accounts do not include a source for sampling large housing complexes such as Public Housing or Private cooperative Housing which is supplied electricity through demand meters (master-metering). However, many of the utility companies did supply us with information about non-public large housing complexes supplied electricity in this manner. Further assistance in this problem was obtained through the cooperation of the various County Planning Boards that comprise the Tri-State Region. Information about large public housing complexes with demand meters was secured from the various Public Housing Authorities.

Each utility company was asked to select one in every one hundred residential rate accounts following a random start. Records of each company are in meter-reading cycle order which is, in effect, a geographic stratification of the universe. A meter cycle lasts either 21 or 42 days. In addition, a 100% listing of all commercial or general service accounts were obtained as well. The sample frame included both active and inactive meters; hence, vacant units as well as occupied ones are reflected in the sample. However, since each sample was selected from utility company records at an instant in time, there is no reflection of housing unit growth from the time of selection to the end of the field interviewing period.

The commercial rate accounts although numbering about 225,000 represent much less than 5% of the final residential units sample. Nevertheless, to insure complete coverage of the living place frame, a sample from this portion is necessary though costly. It was decided to stratify this frame into three groups based on the number of living places likely to be associated with an account.

(1) Accounts likely to have no housing units or other special dwelling places.

(2) Accounts likely to have from one to twenty-five housing units or other special dwelling places.

(3) Accounts likely to have more than twenty-five housing units or other special dwelling places.

Group (1) was sampled at a rate of one in every one hundred; group (2) at one in every ten; and group (3) was covered one hundred percent. All accounts selected from groups (1) and (2) were assigned to field personnel whose job was to determine the number of housing units or special dwelling places associated with each account. These results were then sampled at a second stage to provide an overall probability of 1 in 100. That is,

Group (1) 1/100 = 1/100 X 1/1

Group (2) 1/100 = 1/10 X 1/10

Accounts in group (3), such as hotels, motels, rooming houses, hospital staff quarters and school dormitories, were sampled at a rate of 1 in 100 at the 1st stage to obtain a 1% sample. Information on the number of housing units and/or other special dwelling places was obtained for each in-scope account by mail or telephone contact. After this data is obtained the 1% sample is selected systematically following a random start.

Data on the number of housing units associated with structures supplied electricity through demand meters (master-metering) was obtained through the cooperation of the various utility companies, county planning boards or public housing authorities. Again, a 1% systematic random sample was selected following a random start.

Some towns within the Cordon Area outside New York City were not covered for one reason or another by sampling the records of the various utility companies. These enclaves were sampled by means of a block field listing procedure analogous to the area probability design in New York City. Estimates of the number of housing units in existence by block were available to us through previous work of the Tri-State Land Use Inventory. With this data as a frame, a 1% random sample was selected systematically in five independent replications of 0.2% each. All unduplicated blocks in which the sample hits fell were then again field listed in detail with great care taken to physically identify each individual housing unit. Each sample hit was then uniquely associated with its appropriate physically defined housing unit. Additional sample units or a reduced nu mber was a function of the accuracy of the original block estimates of the number of housing units. The replication feature of the survey design reduced the field listing work as well as providing a simple procedure for calculating sampling errors of the estimated to be produced.

Military Sample

All branches of the Armed Services (Army, Navy, Marines, Coast Guard and Air Force) were contacted and asked to provide data on the number of family housing units and barracks living quarters (rooms, cots, beds, etc.) available for officers and enlisted men within each base in the Tri-State Cordon Area. A systematic random sample of 1% of these living places were selected and assigned for interviewing. Additional sample units were interviewed when original estimates were found to be low in relation to actual numbers of living places. Conversely, a reduction in the original assignment occurred when estimates from the military overstated the true number.

2. Truck-Taxi Survey

The sample frames used for selection of the three percent probability sample of trucks and taxis were the records of the Motor Vehicle Departments in New York, New Jersey and Connecticut. The records are kept in a different manner by each of the three states ranging from micro-film on cards to IBM punch cards to 1401 computer tape files. The sampling plans were tailored to fit each of these three record forms without sacrificing the requirements for the proper selection of a probability sample. The basic design incorporated the selection of three independent subsamples of one percent each with varying degrees of natural geographic stratification. Replication of the sample in the form of independent subsamples has many statistical advantages the primary one being the ease of calculating the standard errors of the survey. Each subsample was selected from a random start followed by systematic selection of every one-hundreth license plate thereafter. This procedure was accomplished in New York and Connecticut with selection taking place in a serpentine fashion through each District office or tax town thereby providing initial natural geographic stratification. In New Jersey, an initial ten percent sample was selected by associating a random number with the last digit of the in-scope truck or taxi registration. After geographic stratification of this initial sample by county was effected, three independent subsamples of one percent each was selected by a systematic random sample of 30% of the initial sample.

A separate selection of three one percent samples of in-scope Post Office Department and Military Base vehicles was effected manually after lists of appropriate license plate numbers were received. Stratification by municipality and base respectively was accomplished before selection.

E. Audits Applied to the Data

Immediately after selection of the home interview sample, various checks were applied to the sample to insure its reasonableness when compared to published sources. Outside New York City, for each municipality, town or other political subdivision within the defined Cordon Area, building permit data was obtained from appropriate state agencies covering the time period from the 1960 Census up to as close to the survey periods as was available. Building permit data includes information on the number of housing units that are expected to be built in the political subdivision. While, in some cases, the housing unit may not in fact be built by the contractor, this is rare. Usually a unit is built within three to six months after a building permit is taken out by the contractor. All building permit data was available up to at least the end of December 1962. This is six months before the beginning of the field survey period for the Home Interview Survey. Building permit data added to the 1960 Census of housing figures for each political subdivision gave us reasonable estimates of the number of housing units in existence as of the beginning of the survey period. The one percent sample after expansion by the inverse of the sampling fraction was compared to these independent estimates. Any large discrepancies were checked in detail to account for the difference.

In New York City, a different type of audit was conducted after the completion of the sample survey. The primary sample unit in New York City was a uniquely defined cluster of about eight housing units. To check the accuracy of interviewers in carrying out instructions to interview all units within their assigned cluster, an independent assessment of the number of units within a sample of clusters was made. A simple random sample of clusters was selected after stratification by health districts and week of work assignment. The reason for health district stratification was the apriori knowledge of varying difficulty in carrying out assignments in the different parts of the city. Week of assignment was considered an important mode of stratification since, apriori, it was felt that quality interviewing takes some time to achieve. Hence, after these two modes of stratification, a simple random sample of clusters was taken; one of 20% for the first four weeks of assignment and 10% for the remaining weeks. The results of the audit indicated an overall understatement of about 1% of the housing units that should have been interviewed but were missed. This varied considerably by geographic area in the city ranging from a 0% to 4% understatement.

III SURVEY RESULTS

The objectives of the survey consists of making many estimates of trip production for various geographic couplets by mode of travel and purpose of trip. This data is related to the use of the land at the couplet as well as car ownership rates and income category at households generating trips. Estimates are projected for an average weekday throughout the survey period although analysis of peak and non-peak hour periods will also be ascertained.

A. Sample Estimates

Any simple estimate from the survey such as some aggregate of a characteristic may be estimated simply in the form X" = Kx where K is the inverse of the sampling rate and x is the weighted count (including adjustment for non-response) in the sample with the specified characteristic. An analogous estimate for a ratio would be of the form X^1 = = p where p is some proportion made up of two characteristics x and y. In the usual case, x may represent the number of trips made by y household or p is a statistic of trips per household.

B. Sampling Errors

Two types of error arise in sample surveys: sampling errors and non-sampling errors. Non-sampling errors come primarily from errors of response in collecting data, errors contributed during the processing of the results and any bias in the sample due to non-response. Errors of response were kept to a minimum by careful training of interviewers who were supervised by people experienced in home interview techniques or familiar with the trucking industry. Processing errors were kept within bounds through the use of quality control methods. The possibility of bias due to non-response of a small proportion of the sample assignments remains. The method for dealing with non-response is discussed in the next section under non-sampling errors.

Sampling errors of the survey estimates arise from the fact that the characteristics as mirrored by the sample do not exactly coincide with the characteristics that would emerge from an equal complete coverage of the entire frame.

For computing the sampling errors in New York City we use a model in which the entire selection for the Health District within a borough consists of two random and independent halves. Actually, by this method of collapsed strata[2] we create the two computing units. Since this disregards further stratification actually accomplished, this results in slight over-estimation of the variance.

The two halves for the jth Health District may be represented as follows in estimating the proportion for some characteristic:

$$p_j = \frac{x_j}{y_j} = \frac{x_{j1} + x_{j2}}{y_{j1} + y_{j2}}$$

For the entire city of 30 Health Districts the simmilar estimator is the ratio of the sums of the Health Districts.

$$p = \frac{\bar{x}}{\bar{y}} = \frac{x}{y} = \frac{\sum_{j=1}^{30}(x_{j1} + x_{j2})}{\sum_{j=1}^{30}(y_{j1} + y_{j2})}$$

The "relvariance" (the square of the coefficient of variation) of p can be estimated by:

$$C_p^2 \doteq C_{\bar{x}}^2 + C_{\bar{y}}^2 - 2\,C_{\bar{x}\bar{y}}$$

$$C_{\bar{x}}^2 = \frac{1}{x^2}\sum_{j=1}^{30}(x_{j1} - x_{j2})^2$$

$$C_{\bar{y}}^2 = \frac{1}{y^2}\sum_{j=1}^{30}(y_{j1} - y_{j2})^2$$

$$C_{\bar{x}\bar{y}} = \frac{1}{xy}\sum_{j=1}^{30}(x_{j1} - x_{j2})(y_{j1} - y_{j2})$$

Thus from the sum of 30 Health District contrasts for the City's estimates the relvariance can be computed with 30 degrees of freedom equal to the number of Health Districts within N. Y. C. These computations are made for a large number of characteristics which are then plotted and average values, subject to smaller variations, are used for estimating standard errors.

The relvariance for a simple estimate (X" = Kx) is just $C_{\bar{x}}^2$ defined above.

[2] Nathan Keyfitz, "Estimates of Sampling Variance Where Two Units are Selected from Each Stratum", Journal of the AMerican Statistical Association, 52, (1957), Pp. 503 - 510.

Outside New York City the method of random group estimation of variance will be used. The entire sample is divided into t random groups of K units each. An estimation of the variance[3] may be expressed as

$$S_K^2 = \sum_{g=1}^{T} \frac{(x_g - \bar{x}')^2}{K(t-1)}$$

where x_g = the value of the characteristic in the gth group

$\bar{x}'$ = the mean per group of the t group totals

For both estimates inside or outside New York, sampling errors of various size cells for a representative group of important characteristics can be plotted on a graph. Sampling errors for other estimates not plotted can be read off the graph.

The design of the Truck-Taxi Survey as a replicative sample in three subsamples provides a simple procedure for producing estimates and calculating sampling errors.

Simple inflation estimates are of the form $X' = Kx$ in which K is the inverse of the sampling fraction and is the value of the characteristic for the three subsamples added together. Ratio estimates of the form $f = \frac{x}{y}$ where x is the value of some characteristic in the sample and y is the value of some other characteristic in the sample, both for the three subsamples combined.

The variance of the two types of estimates may be expressed as follows:

$$\text{Var. } X' = \left(1 - \frac{R}{K}\right)\frac{K^2}{R(R-1)} \sum_{i=1}^{M} \sum_{j=1}^{R} \left(x_{ij} - \bar{x}_{i.}\right)^2$$

$$\text{Var. } f = \left(1 - \frac{R}{K}\right)\frac{R}{y^2}\frac{1}{R-1} \sum_{i=1}^{M} \sum_{j=1}^{R} \left[\left(x_{ij} - \bar{x}_{i.}\right) - f\left(y_{ij} - \bar{y}_{i.}\right)\right]^2$$

where $R = 3$ = the number of independent subsamples

K = sample inflation factor

x_{ij} = value of characteristic x in the jth subsample and ith stratum

y_{ij} = value of characteristic in the jth subsample and ith stratum

$\bar{x}_{i.}$ = mean value for characteristic in the ith stratum

$\bar{y}_{i.}$ = mean value for characteristic in the ith stratum

C. Non-Sampling Errors

There is no consistent or unbiased way of adjusting for non-response in surveys of a human population. The magnitude of the biases resulting from any subjective procedure is not known. Hence, great efforts were taken to achieve small non-response rates within the limits of the budget. The effort resulted in a non-response rate of about 10% for the Home Interview Survey and less than that proportion for the Truck-Taxi Survey. These rates are computed on the base of in-scope housing units or licensed trucks and taxis for which trips are possible.

Criteria for representing Home Interview non-responding sample units by factoring completed samples include:

(1) Geographic area on a Census Tract grouping level in New York City and a Municipality grouping outside of New York City.

(2) Composite of structure types and housing density grouping.

The strata so defined are felt to be appropriate with respect to the characacteristics being measured by the survey.

The similar criteria for the Truck-Taxi Survey sample include:

(1) Geographic area on a county group level for the truck or taxi base of operations.

(2) Body type (truck) or vehicle type (taxi).

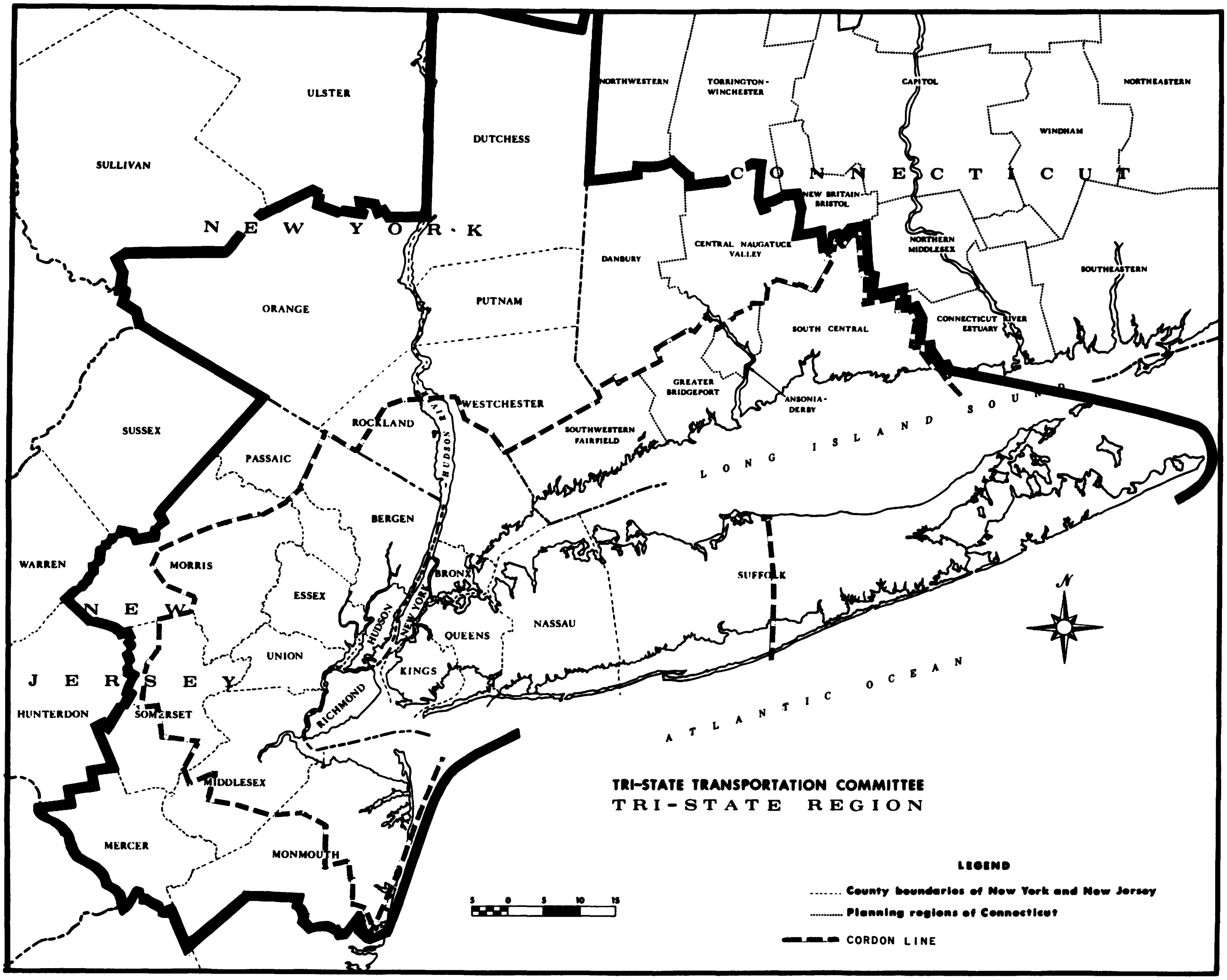
Each of the three independent subsamples are to be treated separately throughout.

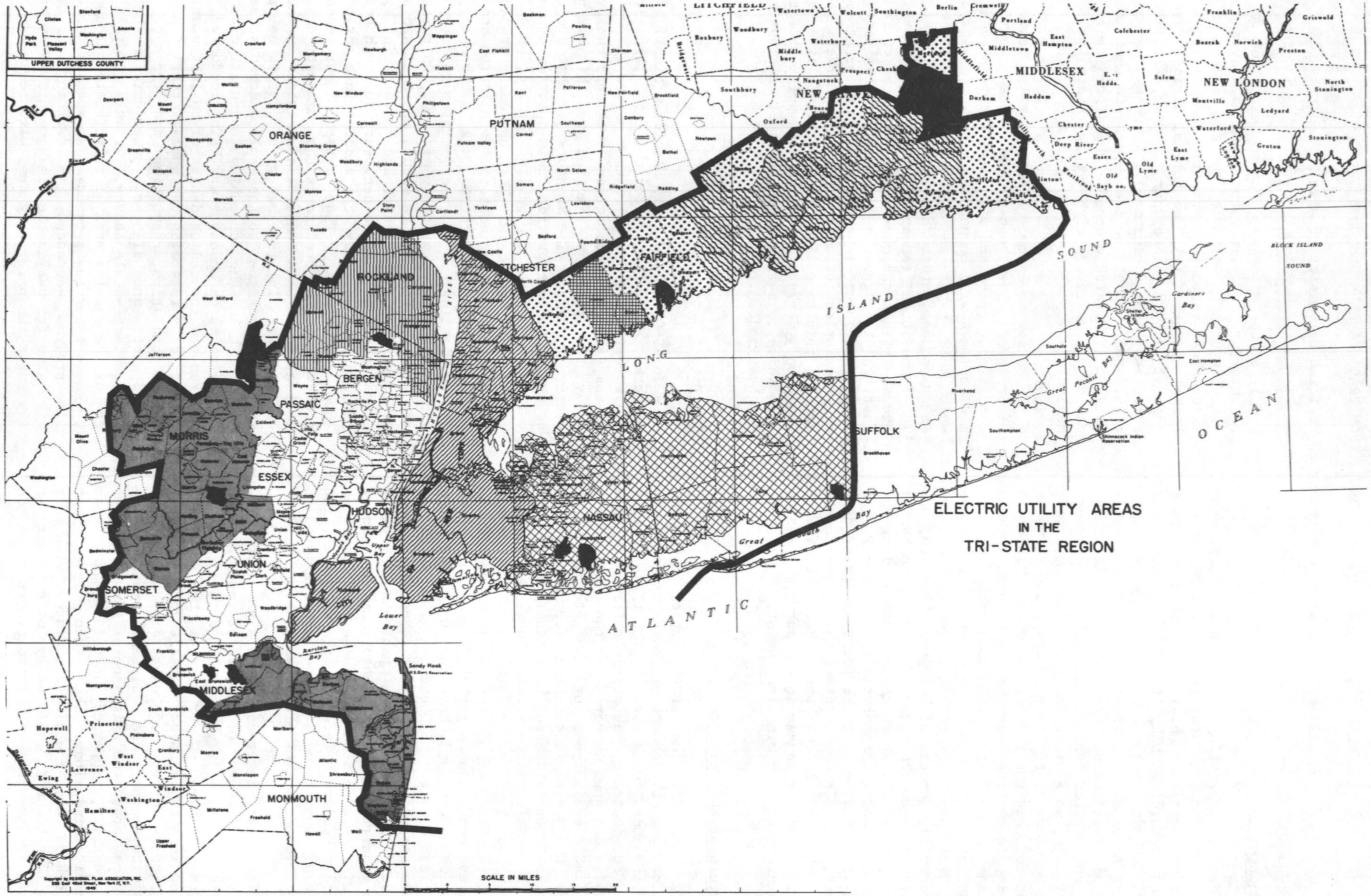
[3] M. H. Hansen, W. N. Hurwitz, and W. G. Madow, "Sample Survey Methods and Theory," Volume I, Pp. 440 - 444.

For each of the strata defined above, the ratio Interviewed Households (Trucks, Taxis) Plus Non-Interviewed Households (Trucks, Taxis) Interviewed Households (Trucks, Taxis) is computed.

These ratios are applied to all data for interviewed households in the corresponding strata, except in groups where the ratio exceeds 2. In such cases, the process of combining strata is continued until the ratio becomes 2 or less.

APPENDIX





ELECTRIC UTILITY AREAS IN THE TRI-STATE REGION

]