CLUSTER SAMPLING METHODS FOR MONITORING ROUTE-LEVEL TRANSIT RIDERSHIP

by

MARK BUCCIARELLI

B.S. Systems Engineering, Case Western Reserve University (1987)

Submitted to the Department of Civil Engineering in Partial Fulfillment of the Requirements for the Degree of

MASTER OF SCIENCE

in

OPERATIONS RESEARCH

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

July 1991

©Massachusetts Institute of Technology 1991

Signature of Author _____

Operations Research Center

July 29, 1991

Certified by _____

Nigel H. M. Wilson, Professor of Civil Engineering

Thesis Supervisor

Accepted by _____

Thomas Magnanti, Operations Research Center

Co-Director

Accepted by _____

Ole S. Madsen, Departmental Committee on Graduate Students

Chairman

CLUSTER SAMPLING METHODS FOR MONITORING ROUTE-LEVEL TRANSIT RIDERSHIP

by

MARK BUCCIARELLI

Abstract

The past 20 years of federal subsidies have improved transit infrastructure and stabilized ridership, but these subsidies have replaced farebox revenue as the primary revenue source and transit market share has decreased. In 1984 the Urban Mass Transportation Administration defined a new federal policy regarding public transit emphasizing that private industry be encouraged to compete for the provision of public transit services. Such a policy requires a transit agency to focus on service planning and monitoring as opposed to a service operation. The costs of monitoring a private contractor should be included in any cost/benefit analysis of whether or not to contract out service.

This research presented investigated the cost of monitoring transit service at the route level using five years of ridership data from the MBTA Commuter Boat service. Five sampling plans were compared based on two measures of efficiency: required sample size, and a generalized cost function. The five sampling plans investigated were: simple random sampling of departures, proportional random sampling of departures stratified by time-of-day, simple random sampling of days, systematic random sampling of days, and simple random sampling of days with post-stratification by time-of-day.

Parameters of variation were found to be very stable from year to year, despite rapid service growth and large seasonal fluctuations. In addition, some parameters of variation were found to be consistent with those from other transit services. The within-day variance was found to dominate the between-day variance, leading to the result that simple random theory can be used to compute the required sample size and cluster sampling can then be used to obtain the sample.

Systematic cluster sampling and proportional stratified sampling were found to require an average of 25% less observations than simple random sampling to reach the same precision. Using a cost function which distinguished between the cost of making an observation and the cost of getting to the observation location, a comparison of these two plans resulted in systematic cluster sampling clearly being more cost efficient.

Thesis Supervisor: Dr. Nigel H. M. Wilson
Title: Professor of Civil Engineering
School of Engineering, MIT

Acknowledgements

I would like to acknowledge the good fortune I had to be given the two parents I have. Without their support, encouragement, and love, this thesis would not have been possible. I would also like to thank my brother Louis whose intuitive love has been a bedrock of inspiration.

Special thanks to my research advisor Nigel H.M. Wilson for his guidance and patience throughout the duration of this project. His monitoring of my thesis performance required a large systematic sample (he met with me regularly) from which he produced an efficient estimate of the contents of this thesis (and wasted little time in our meetings).

The effectiveness and efficiency of the faculty and staff at the Operations Research Center must be acknowledged. My initial impression of the ORC never changed much; here is a place where people are both extremely creative *and* organized--something I had never run into before!

I would like to acknowledge what good company my classmates here have been. Although they might not appreciate good puns (okay, puns) all the time, they never threw anything at me. From them I got a measure how accepting people can be.

Finally, some stanzas from a poem called "Don't Quit" which is on Laura's wall, and which seems appropriate to this thesis.

Life is queer with its twists and turns,
As everyone of us sometimes learns,
And many a failure turns about,
When he might have won had he stuck it out;
Don't give up though the pace seems slow,
You may succeed with another blow.

Success is failure turned inside out,
The silver tint of the clouds of doubt,
And you never can tell how close you are,
It may be near when it seems so far;
So stick to the fight when you're hardest hit,
It's when things seem worse,
that you must not quit.

Contents

List of Tables

List of Figures

Chapter 1. Introduction

In recent years, the use of private companies to provide mass transit service has increased. This trend was formalized in 1984, when the Urban Mass Transportation Administration (UMTA) announced its privatization initiative which encouraged transit agencies to use private firms. While private sector involvement had been mentioned in the 1968 enabling legislation, this marked an important change in the federal attitude towards transit service provision. The executive branch of the federal government, through UMTA, was expressing disappointment with the results of 20 years of capital financing and 10 years of operating cost subsidies at the federal level. UMTA's new approach has been to use federal influence to encourage competition for the provision of public transit service.

Certainly, some type of change was needed. During the 10 year period from 1975 to 1985, the nation-wide supply of annual transit vehicle miles increased 28%, from 2,176 x 10^6 to 2,790.7 x 10^6. However, productivity decreased 7.5%, from an average of 3.35 passengers boarding* per vehicle mile in 1975, to 3.10 passengers in 1985[†]. During this same period, operating costs per passenger boarding increased 47%, from $0.97 to $1.43, after inflation. Due to fare increases, the revenue per passenger boarding posted a modest increase of 11¢, from $0.56 to $0.67. However, governmental operating assistance per passenger boarding increased 105%, from $0.39 in 1975 to $0.80 in 1985[††] (American Public Transit Association 1990). Although federal aid stabilized fares and ridership, it replaced farebox revenue as the primary revenue source, while transit market share continued to decrease.

---

* Here, unlinked passenger boardings. A transfer is counted as a separate boarding. One passenger who takes a bus to get on the subway is counted as two unlinked passenger boardings.

[†] The passenger boardings statistics are not entirely reliable. Before 1980, the collection procedure had not been formalized by UMTA. Before 1984, they did not include demand responsive transit, rural or small system ridership.

[††] All dollar amounts are adjusted for inflation.

The financial decline during this ten year period was caused in large part by inefficiencies which have existed for many years. For example, since the 1920's, when the automobile replaced mass transit as the preferred mode for social, excursion, recreational, and family travel, the demand for transit has been primarily for work-to-home and home-to-work trips. This demand was focused and intensified by the establishment of the 8 hour work day and 5 day work week in 1938 (Jones 1985). Transit systems must have sufficient capital and labor resources on hand every weekday to handle the largest load. In addition, the cost of labor is increased by contractual work-rules, which traditionally have not allowed part-time labor (although this has changed in the past ten years). Thus, operators work at a higher pay rate to cover both peaks, and there is overstaffing during the midday period. Yet the flat-rate fare structure of the 1900's is still the norm; it costs a commuter no more to travel during a peak period than in an off-peak period. Buses and trains are filled to capacity during peak periods, and under-utilized in off-peak periods.

By encouraging competition, UMTA hopes that some of these structural inefficiencies will change naturally. The rational is the standard free-market argument: a market in which firms compete freely will produce goods at a lower cost to consumers, and will react more quickly to changing consumer demands. In the context of the current mass transit structure, the goods are transportation options available to citizens and are provided largely by governmentally-controlled public monopolies. Increasing competition by allowing private firms to provide service will provide citizens with better service for the same cost or the same service at a lower cost.

However clear this ideal sounds, its implementation is not so clear. Few private firms have the capital resources necessary to provide large scale mass transit. The large capital requirements, combined with short contract durations, 31 months on average (Teal 1988), and the historic lack of profitability of transit service makes banks leery about lending money to would-be private service providers.

10

One way around this difficulty is for the transit agency to purchase the capital equipment, and hire a private company to operate this equipment. This approach encourages competition as it allows private firms with less capital resources to enter the bidding process. It is interesting to note that this approach was first tried in Boston in the 1890's, when this city financed, built, and owned the nation's first subway, which was operated by a private company (Jones 1985). In 1984 this approach was commonly used for privatization, with 83% of the vehicles used by private firms to provide contracted fixed-route service owned by public agencies (Teal 1988).

A major concern of any agency contracting out to a private company is that the level of service be maintained or improved. If costs decrease, but service deteriorates at the same time, savings might be illusory or short-lived. In order to maintain control over the level of service, an agency will retain responsibility for scheduling, fare levels, and safety requirements. Of the three sets of fixed route services at the Massachusetts Bay Transportation Authority (MBTA) that are contracted out (commuter rail, commuter boat and bus), the MBTA controls all three factors for each service. Control of these factors leaves the private operator with one competitive edge: labor costs.

Notwithstanding these constraints on private contractors, in 1984 private operators provided a significant amount of public transit service. In 1988, a survey of 468 fixed-route operators (reporting 1983-84 data) showed that 84 (18%) operated some service through private contractors. These privately contracted services were generally less than 50 vehicles in size. Of the 1037 total operators of all services in the survey (fixed route, demand responsive, commuter, and weekend services), 421 were operated by private companies. Of these 421, only 17 were larger than 50 vehicles. These 421 services comprised 5.1% of nation-wide transit operating expenses and 8.6% of the nation-wide revenue vehicle miles of bus service produced (Teal 1988).

1.1 The Transit Agency as a Monitor

When a transit agency contracts out a service to a private operator, its role with respect to that service changes from that of a service provider to that of a service planner and monitor. First and foremost, it must monitor the physical condition of the vehicles used to provide the service to guarantee the safety of its customers. Second, it should monitor the on-time performance to ensure that satisfactory schedule adherence is achieved. Third, it should monitor route-level ridership so that schedules can be adjusted to provide an appropriate level of service. In a more general sense it should make sure that high quality service is provided consistent with the contract.

Additional monitoring responsibilities may be created by specific incentive and/or penalty clauses in the service contract. For example, the agency might fine the operator for failing to meet an on-time performance goal, and reward the operator for ridership above and beyond some base monthly level. Such contractual arrangements are intended to give the contractor a financial interest in the success of the service.

The costs associated with monitoring a private contractor should be included in any cost/benefit analysis of whether or not to contract out service. It is not a fair comparison if the contract amount is used as the cost of the private operation if the agency is spending additional resources to monitor the operator. How much these extra-contractual costs amount to can be difficult to establish. A management team in Los Angeles made 28 different cost comparisons while trying to evaluate the savings of a bus privatization program, with the resulting savings ranging from 5 to 60 percent (DOT 1986).

However, there is one aspect of measuring the cost of monitoring which is relatively straightforward to analyze objectively. The cost of monitoring ridership on a contracted route or set of routes is directly related to the sampling plan used. At one extreme, an agency can collect and process data for every trip provided by the operator. At the other extreme, the agency can use ridership reports from the operator with no checking other than against the reported ticket revenues. If the operator is responsible for collecting ticket revenues, the latter approach would not detect potential fraud by the operator. In the

wide range of alternatives between these two extremes, an agency can implement various sampling methodologies. Which methodology is chosen can have significant impacts on the cost of monitoring and hence on the overall cost-effectiveness of service contracting.

1.2 Thesis Statement

The thesis of this research is that the appropriate use of cluster sampling can significantly reduce the costs of monitoring ridership for a transit agency. To investigate this thesis, cluster sampling will be compared with ran. m sampling and stratified sampling using ridership data from the MBTA Commuter Boat service. This data set includes the number of passengers travelling on every trip which departed between June 1984 and May 1990. Ratio-to-revenue estimation was not used because departure-level revenue records were not available.

Cluster sampling, as its name implies, means selecting departures for the sample in a cluster. In this research, a cluster is defined to be a day, and all trips departing on the selected day are included in the sample. However, a cluster does not have to be a day. A cluster could also be the set of all trips a crew operates in a shift, or all departures on a given route in the am peak. The two criteria which define a good cluster choice are: (1) it is inexpensive to collect a cluster of observations, and (2) there is a minimum of variation between clusters.

The advantage to cluster sampling is a lower cost of data collection. Though more data may need to be collected using this method, because it is collected in clusters, it is less expensive to collect. For example, the administrative overhead of selecting two days, and scheduling people to measure all departures on these two days, say there are 50, is likely to be much less than scheduling people to measure 40 departures that are randomly selected from the 500 trips which depart over a four week period.

In this research, the empirical precision of each plan is compared with the theoretical precision obtained from statistical theory. The performance of the methods for

13

small samples sizes and deviations, if any, from the theoretical precision is noted. Then, the plans are compared with each other, both by sample size and by using a generalized cost function. The cost function decomposes the cost of sampling into two components: (1) the field cost of physically taking the measurement, and (2) the overhead costs associated with getting a person out to the count location to make the measurement. This analysis is performed for each of the five years in the data set.

The rest of this thesis is as follows. Chapter 2 presents a literature review which covers the topic of transit performance from the perspective of government, academia, and privatization efforts. A historical review of research regarding the application of statistical methodology to measuring transit ridership is presented, along with some of the more recent case studies. Chapter 3 presents the statistical theory for the simple random sampling, stratified sampling, and cluster sampling methodologies investigated in this research. Chapter 4 presents a history of the transit service that produced the data set, describes the experiments, and presents the results of these experiments. Chapter 5 provides a summary of the results of the research and presents suggestions for future extensions of this work.

Chapter 2. Research Context

This chapter reviews literature pertaining to transit performance and describes a consensus of governmental, academic, and private concerns regarding the need for quality transit performance monitoring. A review of literature pertaining to the application of statistical methodology to transit monitoring is presented. Two recent case studies of the data collection efforts of large transit systems are presented in detail, with particular attention paid to data collection costs. Although an extensive search was done for papers detailing the costs of monitoring private operators, there is little published research on this topic.

2.1 A Consensus Regarding Monitoring Transit Performance

From a review of literature regarding transit performance, it is clear that there is a consensus opinion. This consensus was perhaps most articulately expressed in Miller and Kirby (1984).

> *"Improved resource management is now a common goal for all public transport*
> *systems; large and small, old and new. Improved procedures for performance*
> *monitoring are the building blocks for accomplishing this goal, and deserve*
> *greatly increased attention in the immediate future."*

This consensus reaches across governmental, academic, and private concerns.

2.1.1 Governmental Perspective

Congress first expressed its interest in transit performance data in 1974, when it mandated that within three years, the Secretary of Transportation create a uniform system of reporting for all transit operations receiving federal funds. This reporting requirement is referred to as Section 15, after the section amended to the Urban Mass Transportation Act of 1964 in 1974. Congress intended that the

(Uniform Reporting System) shall be designed to assist in meeting the needs of individual public mass transportation systems, Federal, State and local governments, and the public for information on which to base planning for public transportation services, and shall contain information appropriate to assist in the making of public sector investment decisions at all levels of government. (UMTA, 1977).

This Uniform Reporting System was issued in 1977, and the first report, containing 1978-79 data, was released in 1981.

This standardized system of reports has allowed Congress to create a performance based allocation of transit subsidies. Previous formula allocation had been based solely on population and population density. The Federal Public Transportation Act of 1982, Title III of the Surface Transportation Assistance Act of 1982 altered the formula grant program to include transit service data. This new formulation was intended to provide "financial incentives to communities for increases in transit productivity" (Congress, 1982). The measures added to population and population density were revenue vehicle miles, route miles, and passenger miles.

However, this mechanism for encouraging improved transit performance has remained static. No new performance measures have been introduced since 1982, and the relative importance of the measures, as measured by the percentage of total fund allocation, has not changed significantly since 1982. In addition, the only measure relating to service utilization, passenger miles, has remained the least important component of the formula, decreasing from 7% of the total disbursement in 1982 (Fielding 1987) to 5.8% in 1991 (Congress 1990). Table 2.1 presents the various measures and their weights.

Measure	Weight in 1982	Weight in 1991
revenue vehicle miles	45%	50.3%
population and population density	36%	31.2%
route miles	11%	12.7%
passenger miles	7%	5.8%

Table 2.1 Weights of Elements in Federal Formula

2.1.2. Academic Perspective

The pioneering academic work on transit performance indicators was done by Fielding & Glauthier in 1976 (Talley 1986). They created a system of performance indicators which distinguished between the effectiveness and the efficiency of service provision. The availability of Section 15 data spurred some researchers to extract as many performance indicators as possible from these data: Holec, Schwager, and Fandialan (1980) created 47, and Fielding identified 48 distinct performance measures. Fielding used statistical analysis to partition this set of 48 into the following seven groups: cost efficiency, service utilization, revenue generation, labor efficiency, vehicle efficiency, maintenance efficiency, and safety. He identified one "marker indicator" for each of these seven categories (Fielding 1987).

Lee (1989) developed a hierarchical set of performance indicators which can be used for internal evaluation or peer comparison, both at the agency level or the route level. The indicators are modular and hierarchical, i.e. the indicators at the coarsest level of detail are some function of the indicators at the next highest level of detail. There are three levels in this hierarchy. The indicators at the first level are:

deficit / vehicle hour = (cost / vehicle hour) - (revenue / vehicle hour),

cost / passenger = (cost / vehicle hour) / (passengers / vehicle hour), and

17

deficit / passenger = (cost / passenger) - (revenue / passenger).

The indicators are designed to be used with published data and using Section 15 data, Lee graphically demonstrates the power of his modular approach. Operating ratios are compared by plotting operating cost / vehicle hour on the x-axis and revenue / vehicle hour on the y-axis. Cost versus utilization for the agencies are compared by plotting revenue / passenger on the x-axis and the passengers / vehicle hour on the y-axis. (Note that revenue = cost * utilization.) By comparing these two graphs, conclusions about relative system characteristics can be made. For example, New York and Golden Gate transit agencies have operating ratios close to 50%, and are similar in terms of cost and revenues. Thus in the first graph, they appear next to each other. However, Golden Gate charges an average fare of $1.40, whereas New York charges $0.45 on average. Golden Gate carries around 25 passengers per vehicle hour and New York 85. Thus these two are very far apart on the second graph.

Bates and Lawrence (1986) look at the issue of transit performance from a marketing perspective. They present a historical perspective of marketing in mass transit, identify weaknesses of the system, and make recommendations for improvement. Three weaknesses identified are as follows. The basic product, i.e. the CBD core served by radial routes, has remained unchanged while steadily losing business for 20 years. The one-price fare philosophy does not take advantage of different market segments. "Different people use transit at different times for different reasons. Value of the trip to the tripmaker, ability of the tripmaker to pay for the trip, and considerations of benefit to cost vary widely". In fact, the one-fare structure creates inequity, as setting fares to cover average costs subsidizes the first passenger and overcharges the last.

The final weakness identified by Bates and Lawrence is the difficulty in performing effective evaluation of the results of marketing plans once implemented. The effectiveness of any marketing strategy can be measured in changes in ridership. However, aggregate totals are not fine-grained enough to reveal any improvements, even if the marketing

strategy was successful. For example, if a marketing strategy is aimed at promoting off-peak bus service as an inexpensive, convenient, and environmentally conscious means for suburban homemakers to make shopping trips, then system-wide monthly totals of passenger boardings are not sufficient for accurate evaluation. They conclude,

Given these difficulties, an appropriate strategy for transit marketers is to promote overall service and performance evaluation. To do so requires strong market research on existing as well as potential markets and services provided; and formulation of specific objectives not only for the marketing function but for the entire transit operation.

2.1.3. Privatization Perspective

It is safe to assume that the private sector in general is more concerned with performance objectives (albeit narrow ones related to profitability), the monitoring of service, and evaluation of whether these objectives are achieved. However, how this assumption plays out in situations where private companies are managed by public agencies to provide transportation services is not so clear. Leick and Hedemann (1987) document the privatization of a 25 vehicle bus system in Johnson County, Kansas. This county had experienced changes common to much of suburbia. In the 1960's, the land use patterns changed from rural to residential. In the 1970's, employment in the county grew 91%. The number of residents working within the county grew 122%, and the number of workers commuting into the county increased 130%. In 1984, the county contained 289,000 people in roughly 500 square miles, giving an average population density of 578 people per square mile.

While the pattern of demand for transportation changed dramatically, the pattern of supply did not; consequently, public transit did not capture this new market. In 1980, the system route structure, i.e. radial service to the Kansas City CBD, had not changed significantly from the private commuter routes that were in operation prior to the formation

of the Kansas City Area Transportation Authority (KCATA) in 1970. Of the people who used transit in 1980, only 8% used it for intra-county commutes, and 60% used it for commuting to the CBD.

In 1982, the county commissioners decided to contract service to private operators. Although the county saved $17,000 in the first year, there were many difficulties that remained unresolved over the next four years. For example, Kansas City, threatened by the county's withdrawal from KCATA, imposed a $500 annual fee for every bus entering the city and reduced the allowed bus stops from eight to two. Another difficulty was that the transit vehicles were not maintained properly and had deteriorated to an unacceptable condition after four years of use. Due to these factors and a decline in the regional economy, ridership declined 32% from 1981 through 1984. This trend was not countered with any marketing of the service, as the majority of the commissioners were "philosophically opposed" to such activity. The management of the system was also a problem, with the responsibility for system management changing every year; in 1983 and 1984, no one person was responsible for managing the system. As a result of all these difficulties, in 1984, two of the five county commissioners expressed interest in discontinuing public transportation in the county altogether.

In 1984, the commissioners decided to restructure the system in an attempt to save the system. The county produced a request for proposals (RFP) which divided the transit product into three services: express service to the CBD, intra-county service, and demand responsive service. A timed transfer system was designed for immediate implementation, and bidders were invited to bid on these components separately or in combinations. The RFP required authorization for the county to survey passengers, and included performance measures, such as 90% on-time performance and close monitoring of safety and maintenance. Seven private operators submitted bids on the services, while KCATA did not formally bid, having determined that it could not account for the vehicle fleet separately and would require 28 exemptions from contractual terms.

In 1986, the restructured system and the private operation proved successful. In the first four weeks of service, ridership increased 40%. The new system was extremely flexible, implementing a total of 14 route and schedule changes in the first quarter of operation. Summarizing the causes of this dramatic turnaround, the authors identify seven elements necessary for successful contracting to the private sector:

1. constant commitment by elected officials to public transportation;
2. smooth transition from public to private operation;
3. long term contracts;
4. full-time management and monitoring by the public jurisdiction as well as the private operator;
5. performance standards;
6. marketing; and
7. public acceptance.

Clearly, effective monitoring is an integral component of a successful privatization program. While political conflicts and bureaucratic inertia are likely to dominate the enterprise over the short term, the long term success will depend on how well the service is monitored and evaluated. To quote the authors,

As in any jurisdiction, when public transit service is contracted to a private operator, the official body is still responsible for managing the system. Staff must be kept in place to monitor the contract's provisions, to analyze the system, to monitor ridership patterns and farebox revenues, and to make timely suggestions for route changes.

2.2 A Framework for the Cost of Monitoring Ridership

Before discussing the specific case studies of transit monitoring, it is useful to develop a general framework for the cost of monitoring transit ridership. The cost of monitoring ridership depends on two components: the transit system, and the sampling method. The transit system is the physical network of routes, frequency of trips on each route, route types (e.g. feeder, shuttle, or diffuse), and the data collection equipment

21

installed. The sampling method is the sampling methodology used to collect information about ridership characteristics, e.g. simple random, stratified, cluster, or some combination of these three.

The cost of monitoring ridership characteristics depends first on the degree of precision desired. For example, a low cost statistical estimate of the total annual system-wide boardings can be computed by measuring the passengers boarding on one randomly selected bus trip, and multiplying this number by the total number of bus trips provided during the year. However, the variation possible using this estimation procedure is very large, and the confidence in the precision of such an estimate is small. In statistical terminology, the variance of this estimator is large.

Another major component of the cost of monitoring depends on what quantity is being estimated. The more aggregate the quantity, the less expensive to estimate with precision. For example, estimating total system-wide boardings to a specified precision is less expensive than estimating annual boardings on each route with the same precision. Estimating something a marketer would be interested in, such as the annual change in working women boardings on each route with similar precision would be even costlier.

There are many other factors which contribute to the cost of monitoring ridership characteristics. Transit system characteristics such as available data collection technology, fleet size, service region size, route type, and labor costs affect the cost of monitoring. For example, a system which uses ride checks will be more expensive to monitor than a system which can use point checks because most routes are of the feeder type.

Finally, the patterns of variation in the quantity being estimated contribute to the cost of monitoring. Such patterns can occur over time and over space. For example, if the quantity of interest is passenger boardings, one pattern over time is the daily peaks in boardings which correspond to home-to-work trips, and another is the similarity of day-to-day total boardings within a season. A spatial pattern might be similarity between boardings on radial routes that run through the same neighborhood. If the sampling

methodology is properly matched to these patterns, then the cost of monitoring is minimized.

The interactions between these and other factors in the cost of monitoring are represented in Figure 2.1. Note that a "measurement" refers to one observation of the quantity of interest, e.g. the total boardings on one bus run. When deciding on a sampling plan for a particular transit system, the patterns of variation in the quantity of interest are rarely well known. Thus, a best guess of these patterns must be constructed from previous studies of the system of interest, peer systems, or simply common sense combined with experience. This best guess, the chosen sampling plan, and the desired precision define how many measurements are needed. The cost of making these measurements depends on the sampling plan, and the cost of the data collection resources available.

2.3 Statistical Methodology for Transit Systems

The amount of literature dealing with statistical methods applied to mass transit has been, until recently, very sparse. Prior to 1977, there were only two reports of note. The 1974 American Transit Association Bus Scheduling Manual suggested frequencies for ride and point checks, but did not present any statistical rationalization of the methodology. The 1976 Manual of Traffic Engineering Studies contains an appendix on determining the statistical precision of an estimate, but this is done only for a simple random sample. No explicit guidelines are presented for determining the frequency and/or magnitude of data collection. Some of the more prominent papers published on this topic since 1977 are reviewed below.
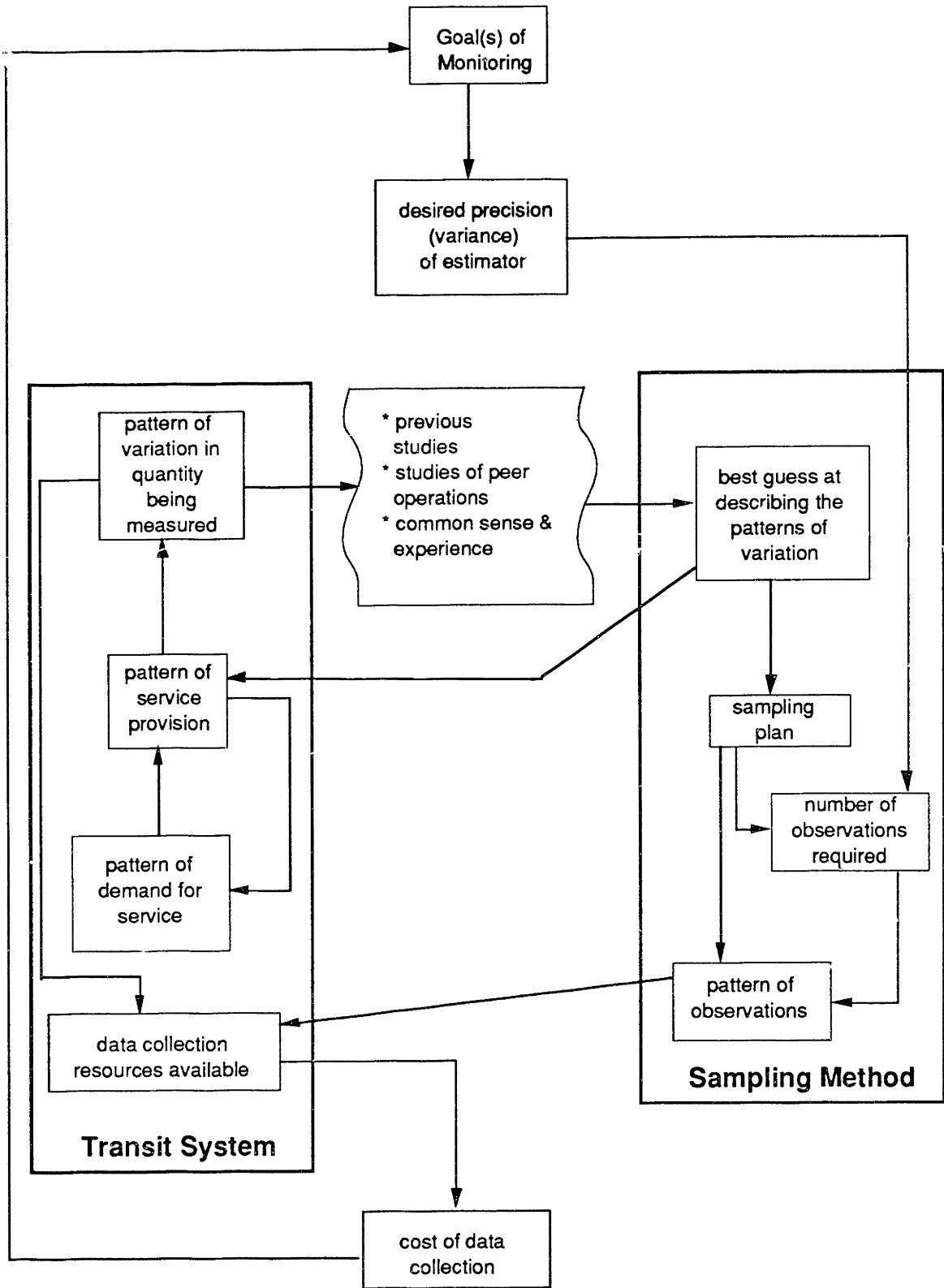
Figure 2.1 Cost of Monitoring Ridership.

2.3.1 Wells Memorandum

In 1977, the Wells Research Company published a study detailing alternative sampling methods for the collection of bus passenger data. This study was the basis for UMTA's circular 2710.1 (UMTA 1978), which suggested a methodology for producing an estimate of system-wide annual unlinked passenger boardings and passenger-miles to the required precision of ± 10 percent of the true value at a 95% confidence level. This method was based on the two-stage cluster sampling method described in Hansen, Hurwitz, and Madow (1956). In the first stage, a set of days are randomly selected, and in the second stage, one-way trips are randomly selected from within the days selected in the first stage. Using conservative estimates of inter-day coefficient of variation of 0.1 and an intra-day coefficient of variation of 1.0, this method requires a sample of 183 days, with 3 departures selected from each day to produce an estimate which is within 9.1% of the true value 95 times out of 100. Because of the large number of days, 183 out of a possible 365, a systematic sample (i.e roughly every third day) of clusters was considered equivalent to a random sample.

2.3.2 Bus Transit Monitoring Manual

In 1979, UMTA began a study to improve the use of statistical methodology for the measurement of transit performance. One product of this study was the Bus Transit Monitoring Manual produced by Attanucci, Burns, and Wilson. This manual proposes a three-phase method for monitoring route-level ridership characteristics. During the first phase, or baseline phase, a large data collection effort is performed to establish a good description of the patterns of variation in ridership characteristics on a route level. Such measures as total boardings, passenger miles, revenue, running time, and schedule adherence are suggested for collection. By measuring the relationships between the variation in these measures at the route, day of week and time of day level, a "snapshot" of the system is taken.

25

Another important aspect of the baseline phase is the identification of conversion factors, or ratio measures. These measures allow the estimation of an expensive-to-collect data item from an inexpensive-to-collect data item. For example, for many systems it is relatively inexpensive to collect a precise measure of total passenger boardings, and expensive to collect a precise measure of total passenger miles. If the total passenger boardings is known precisely, then it will be less expensive to estimate the average passenger miles per passenger boarding than to estimate total passenger miles. A sample is taken to estimate this ratio, which is then multiplied by the precise measure of total passenger boardings. Another common version of this technique is to use the ratio average passenger boarding per dollar revenue and an accurate total of annual revenue.

In the second phase, or monitoring phase, the data collection requirements are much less. Here, monitoring is performed in order to satisfy internal planning and external reporting (e.g. Section 15) requirements. Utilizing the information obtained in the baseline phase, sampling methods can be designed to match the patterns of variability in the system. In addition, the monitoring phase tracks key measures which are designed to indicate any significant changes which require a reformulation of the sampling methodology used in the monitoring phase. For example, service modifications may result in the average passenger miles per dollar revenue changing to such an extent that it is necessary to establish new ratio measures.

The third and final phase, called the follow-up phase, is entered when such an indicator measure changes significantly, or when service modifications are substantial. Obvious examples (in addition to the situation mentioned in the previous paragraph) are adding or cancelling a route, or changing the fare structure. This phase is a focussed version of the baseline phase, where instead of investigating the entire system, the data collection focuses on the component of the system which has changed.

A revised and shortened version of the Bus Transit Monitoring Manual was published in 1985 (TCDM 1985).

2.3.3 A 1984 Survey of Section 15 Collection Techniques

Smith (1984) produced a survey of Section 15 data collection methods used by the nations large transit systems. Systems with 100 or more peak-hour buses were investigated with one aim being to identify "the range of techniques used by large transit properties to collect Section 15 passenger data". Fifty-eight systems were surveyed in total. Results pertaining to the type of sampling plan used, the number of checkers employed, and the percentage of checker staff time devoted to ride checks were summarized for all 58 properties. Case studies of six systems were presented in detail.

The survey revealed that there is a large amount of diversity in how the transit systems produce the Section 15 passenger data. Only 60.3% (35 systems) used the method suggested by UMTA, outlined above in Section 2.3.1. Another 19% (11) obtained a sample from a large set of ride checks. Generally, this set of ride checks includes one element for each daily one-way trip operated. A total of 15.5% (9) use a ratio method, where the estimate for total passenger miles is obtained by estimating average passenger miles per passenger, and multiplying by the total annual passengers. Finally, 5.2% (3) have extensive ride check programs allowing accurate estimates at the route level. The distribution of sampling plans is summarized below in Table 2.1.

Sampling Method	Frequency
Two-stage cluster	60.3% (35)
Sample from extensive ride checks	19.0% (11)
Ratio	15.5% (9)
Extensive random sampling (route level)	5.2% (3)
Total:	100% (58)

Table 2.2 Distribution of Sampling Methods in 58 Large Transit Systems.

Smith's survey also identified the large extent to which large transit systems utilize ride checks. Of the 35 systems using Well's method, 23 integrated their extensive ride

checks into this random sampling procedure. Adding this to the 11 systems which randomly sample only from their ride checks, makes a total of 34 (58%) which perform extensive ride checks. Most systems had between 1 and 2 full-time checkers per 100 peak buses, with only 17% having less than 1 and the remaining 41% having more than 2. Only 10% of the systems don't spend any time on ride checks.

2.3.4 UMTA's Alternative Methodology

In 1985, UMTA released circular 2710.4 (UMTA, 1985) which outlined an alternative methodology for collecting and reporting total annual system-wide passenger boardings and passenger miles. This methodology used the ratio measures of passenger boardings and passenger miles to dollar of revenue collected. This method is more efficient because there is much less bus trip-to-bus trip variation in passenger miles per dollar revenue than in passengers miles per bus trip, and so a smaller sample size is required to estimate this ratio measure necessary to meet the precision requirement of ±10% at 95% confidence level. Hence, you need a smaller sample, 208 trips per year, to obtain the same precision.

In an effort to minimize the impacts of seasonal variation on the sample selection, UMTA suggests randomly selecting 4 bus trips every week. In essence, this is a stratified random sample, where each week is a different stratum. Each trip operated during the week should have the same probability of being selected. For a transit agency to use this method, it must create a complete list of all trips departing during the week of interest, and then randomly select four of them. This list of trips must be accurate, and thus must incorporate all recent schedule changes.

2.3.5. Recent Research

Most of the recent research on the use of statistics in transit monitoring is along the same lines as outlined above, i.e. describing how Section 15 data is collected. Apparently,

either very little work is being done on the type of monitoring required for route level performance monitoring and/or marketing, or if it is being done, it is not being published. One exception to this is the paper by Furth, Killough, and Ruprecht (1988), reviewed at the end of this section.

Two good examples detailing the Section 15 data collection efforts at specific transit agencies are Foote and Hancox (1989), and Ferguson (1987). Foote and Hancox (1989) detail the data collection methods used at the Chicago Transit Authority (CTA). CTA is responsible for a rail and bus system which together serves an area of 250 square miles. The bus system alone operates roughly 9 million trips a year. The focus of this study is the cost associated with collecting Section 15 data, particularly the costs faced by a large transit agency. The requirement to select days randomly within each week is criticized, and an analysis of variance is presented which shows how insignificant the variation due to the day of week is. The researchers suggest that sampling in clusters would be less expensive, perhaps one 2 - 3 week period four times each year.

Ferguson (1987) reports an alternate sampling methodology used at Orange County Transit, California. The author describes how data produced from collection procedures already in place are used to produce Section 15 measures. This method uses ratio estimation from extensive ride checks and driver check sheets, the latter being an independent source for passenger boardings. Neither data source is sampled randomly. The authors justify this lack of randomness by presenting an analysis of variation and arguing that the stratification scheme used matches the pattern of variation. In addition, the sample of driver check sheets is very large, roughly 40% of all trips.

Furth, Killough, and Ruprecht (1987) review five cluster sampling methodologies and compare these techniques using data from the Southern California Rapid Transit District (SCRTD, Los Angeles) and Pittsburgh, Pennsylvania. The five methods are: (1) simple random sampling of clusters, (2) stratified ratio-to-cluster size sampling, (3) stratified cluster sampling with probability of cluster selection proportional to cluster size,

(4) ratio-to-revenue cluster sampling, and (5) cluster sampling with stratification after sample selection (ex post facto stratification).

The Los Angeles data set is a fare check data set for February, 1987, weekdays only. For the purposes of the study, a cluster is defined as a half bus run where a bus run is the set of trips completed by a bus operator during a day. While the cluster sizes that result are not described, the average cluster size per stratum is shown, ranging from a minimum of 2.4 trips to a maximum of 10.7 trips. As these averages are over many clusters (each stratum has over 130 clusters), it is possible that some clusters are much larger than 10.7 trips.

The comparison of methods is based solely on sample size, the issue of cost of data collection (other than that associated with the sample size) is not explicitly considered in such comparisons. The most efficient cluster sampling technique was found to be direct stratification of clusters by average vehicle utilization per cluster, using 8 strata based on total boardings. Each cluster was put into a stratum by estimating the total boardings of each trip contained in the cluster from previous data. A ninth stratum was created for clusters which contained trips with no previous ridership information. Direct stratification of clusters required a sample of 34 clusters for a total of 139 trips. The ratio-to-cluster size method required 53 clusters, which contained 228 trips. Stratified cluster sampling with cluster selection proportional to cluster size required 60 clusters and 382 trips.

The Los Angeles analysis was done on a system-wide basis. Comparing the cost of stratified (4 strata) cluster sampling to stratified random sampling, the authors show that cluster sampling requires 2.2 more trips in the sample to achieve the same precision. However, they state that the cost of cluster sampling is less than the cost of random sampling by a factor of 3 or 4. Thus, cluster sampling is estimated to cost one-half to two-thirds as much as simple random sampling.

The Pittsburgh data set was used to analyze the efficiency of cluster sampling at the route-level. The data set consisted of a series of extensive ride checks that were done in the

30

spring of 1984. In this analysis, a cluster was defined to be the set of trips in a driver run which covered the same line. Most of the clusters used in the analysis contained 4 or 5 trips; those clusters with less than four trips were discarded as the authors were interested in larger clusters. The mean cluster size was 5 trips, the minimum size was 4 and the maximum size was 16. The standard deviation of cluster size was 1.3 trips. The ratios boardings per maximum load and passenger miles per maximum load were computed as well as the mean boardings.

The performance of cluster sampling was first analyzed at the route level. The data set contained 18 routes with 8 or more clusters of size 4 or greater. The average number of clusters per route for these 18 routes was 40, with a minimum of 10 and a maximum of 97 clusters per route. Comparing cluster sampling with simple random sampling at this level of detail, on average, cluster sampling required 10% to 32% more bus trips in the sample to attain the same precision. However, the variation of this factor was quite large from one route to another. Estimating mean boardings with cluster sampling required from 12% to 320% more bus trips to achieve the same precision as simple random sampling.

The performance of cluster sampling was also investigated at the line/direction/time-period (L/D/TP) level of detail. In the Pittsburgh data set, there were 69 L/D/TP combinations. The average number of clusters per L/D/TP combination was 13.6, with a minimum of 8, and a maximum of 33. The average cluster size was 2.3 trips, with a minimum of 2, a maximum of 8, and a standard deviation of 0.414. The time periods used (with number of L/D/TP's in parenthesis) were: early am (4), am peak (5), base (25), pm peak (3), evening (7), Saturday (15), and Sunday (10). Averaging over all L/D/TP combinations, cluster sampling required the same sample size as simple random sampling. The L/D/TP's which contained larger clusters, the Saturday and Sunday all day periods, required 1.19 and 1.10 times as many trips in the sample. One Saturday L/D/TP required 1.75 times as many trips, which was the largest of all of these factors. Based on these results, the authors conclude that cluster sampling of weekday L/D/TP groups can be

considered as simple random sampling. They suggest increasing weekend all-day periods cluster sampling by 20% over the simple random sample sizes.

Chapter 3. Sampling Methodology

This chapter presents the statistical theory which underlies statements of precision about estimates of ridership. Two basic assumptions are described, and the theory of simple random sampling, stratified sampling, and cluster sampling is presented.

3.1. Terminology

There are some statistical terms which are used frequently in the following discussion and warrant definition. When sampling, there is always a "population" of "elements" (e.g. bus trips, passengers, buses, rail cars, etc), each element having a value for some characteristic of interest (e.g. average passengers per bus trip, average income of passengers, fuel consumption rate of bus fleet, car cleanliness, etc.). As it is too expensive to measure this characteristic for every element of the population, a subset of elements is selected, and an "estimate" of the true population value is made from this subset. The subset of elements is called a "sample", and the way this sample is selected is the "sampling plan" or "sampling methodology".

The formula for computing the estimate from the sample is the "estimator". For each particular sample of elements the value of the estimator is the estimate for that sample. If there are N elements in the population, and n elements in the sample, then there are $_NC_n$ possible values for the estimator. If all these estimates were known, a "frequency histogram" could be plotted, with values of the estimator on the x–axis, and the number of times this estimate occurs on the y–axis. If the values on the y–axis are normalized by the total number of estimates, $_NC_n$, the histogram is the "probability distribution" of the estimator. If instead the "cumulative distribution" was plotted, the x–axis would be the same, and the y axis would show the number of estimates that were less than or equal to the corresponding estimate value. Thus, the estimator is a "random variable", with a "mean" and "standard deviation" of its own.

If an estimator is "unbiased" then its "expected value" is equal to the true population value. The expected value of an estimator is equal to the sum of the product of each possible value of the estimator and the probability of getting this value. Another description of unbiasedness is that if all possible samples were generated, and an estimate computed for each sample, then if the average of all these estimates was equal to the true population value, then this estimator would be unbiased.

Finally, the "efficiency" of an estimator is a quantity used to compare it with other estimators. If one estimator is more efficient than another, then for the same sample size, the more efficient estimator will produce an estimate of higher precision. The more efficient estimator can also produce an estimate of the same precision from a smaller sample. The efficiency of an estimator is directly related to its variance; the smaller the variance of an estimator, the more efficient.

3.2 Assumptions.

There are two assumptions which are used frequently in the theory of sampling. The first assumption is that population elements are selected independently in the sample. The second assumption is that the characteristic being measured is normally distributed (Cochran 1977).

3.2.1 Independence

The assumption of independence states that the elements in the sample are selected independently of each other. This is different from the more restrictive requirement that the values of the elements in the population are independent of each other. It is possible to have elements selected independently from a population which contains elements whose values are dependent upon each other.

For example, say that the population consists of all departures on two bus routes, and the characteristic of interest is the average boardings per departure. One bus line

34

second bus route is dependent to some extent on the ridership on the first bus route. The probability of there being X riders on this second bus is equal to the probability that there will be Y people transferring from the first bus times the probability that there will be X - Y passengers who don't transfer. Depending on how large the ratio Y/X is, the probability distribution describing the ridership on the second bus is more or less dependent on the ridership on the first bus.

Here, the elements of the population are the departures, and the population is defined over some period of time, e.g. a week, a month, a year, etc. If the elements in the population are selected randomly to create a sample, then they are considered independent. By selecting elements randomly, the correlation that exists between measurements of ridership made on different trips on the same day is negated.

3.2.2 Normality

By definition, an estimate from a sample does not include all elements of the population in the calculation. The precision of an estimate depends on the sampling plan, the number of elements in the sample, and the number of elements in the population. Calculation of the precision of a given estimator is based on the assumption that the estimator is normally distributed. This assumption is based on the Central Limit Theorem, which may be stated as:

if r is the sum of n independent identically distributed random variables, then as

n approaches infinity, the cumulative probability distribution for r approaches

the Gaussian distribution. (Drake, 1967)

This theorem is extremely powerful. For example, when estimating the mean or the total of n measurements, it is immaterial what type of probability distribution is generating the characteristics you are measuring. As long as the sample contains enough elements, i.e. n is "large enough", the estimator will be normally distributed. When the estimator is normally distributed, you can compute the precision of any given estimate with confidence.

For many distributions, n need only be as large as 30 to apply the Gaussian distribution with little error (Kvanli, 1986). See Figure 3.1 for a graphic illustration of this theorem.

The speed with which the total, or equivalently, the average, of n observations approaches the normal distribution depends on the symmetry of the probability distribution which describes the variability among the population elements. The central portion of the probability distribution for an average will become Gaussian more quickly than will the tails of the distribution. These observations are illustrated in Figure 3.1 below.

3.2.2.1 Normal Theory

Given that the measure of interest is normally distributed, the theory for calculating the precision of an estimate made from a simple random sample is well known. To present this theory, we need to label the elements in the sample, define the sample size, and the mean and variance of a sample. Let $\{y_1, \ldots, y_n\} \equiv$ a sample of n elements, each y_i normally distributed as $N(\mu, \sigma^2)$. Then the estimator for the population mean is the sample mean,

$$\overline{y} = \frac{\sum_{i=1}^{n} y_i}{n} ,$$

and the estimator for the population variance is the sample variance,

$$s^2(y) = \frac{\sum_{i=1}^{n} (y_i - \overline{y})^2}{n - 1} .$$
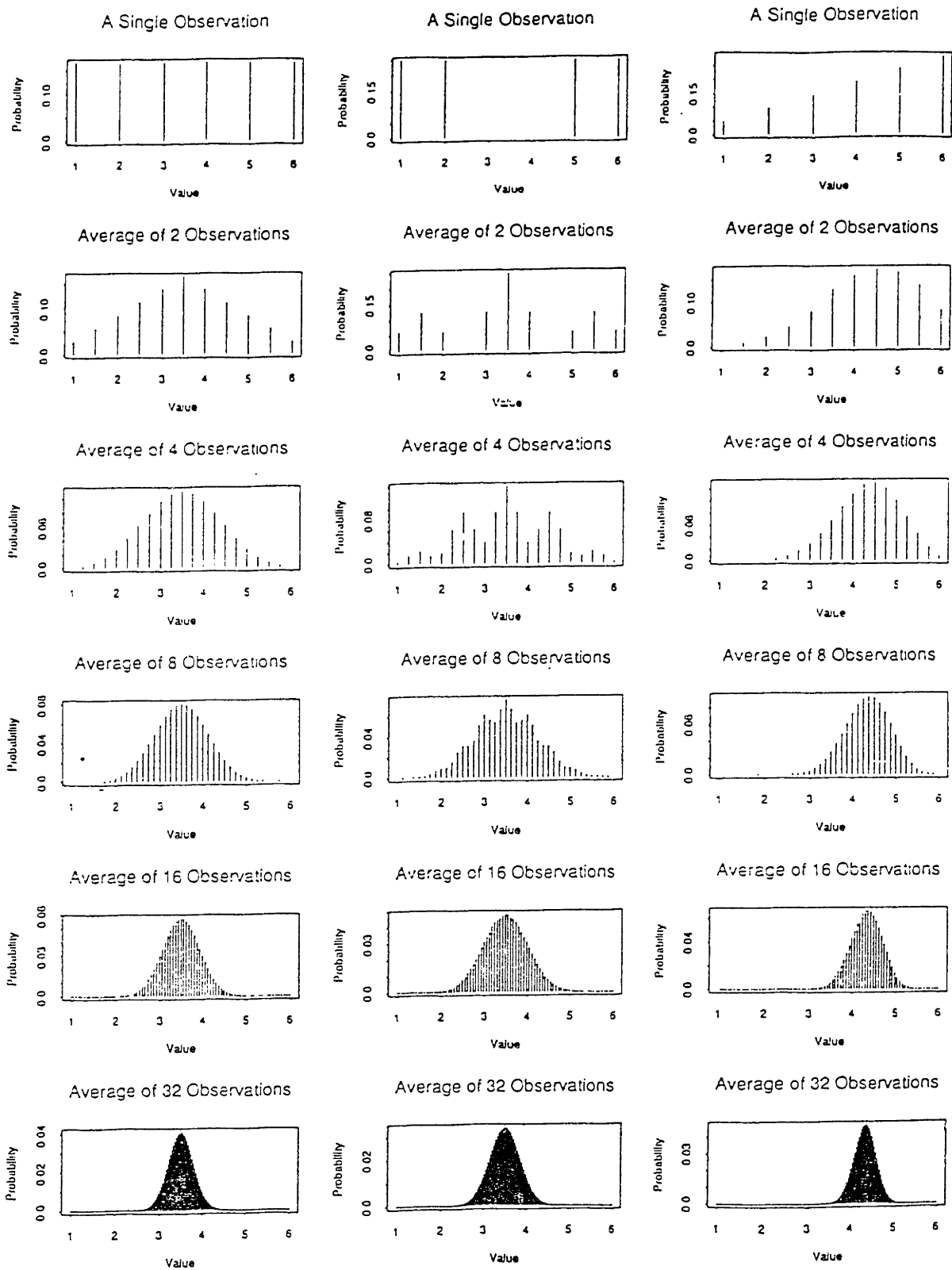
Figure 3.1 Averages of Some Distributions.

Location

There are a number of ways the precision of an estimator for the mean can be expressed. A common method is to calculate a "confidence interval". This interval is defined by lower and upper bounds which are equidistant from the estimate of the population mean. The width of the interval is determined by the desired level of confidence. If this level of confidence is 95%, then normal theory states that if a large number of samples are taken then the true population mean will fall within 95% of these intervals.

If we know that an estimator is described by a normal distribution, we calculate a confidence interval using the unit normal distribution. A number randomly generated from a unit normal distribution, $N(0,1)$, has a 95% chance of being contained in the range [-1.96, 1.96]. If a random variable y is distributed as $N(\mu, \sigma^2)$, then the variable

$$z = \frac{y - \mu}{\sigma}$$

is distributed as $N(0,1)$ and

$$P\{-1.96 < z < 1.96\} = 0.95 .$$

If the sample size is "large enough" that the sample mean is normally distributed, the variable

$$z_1 = \frac{\bar{y} - \mu}{\sigma / \sqrt{n}}$$

is distributed as the unit normal distribution. Note that the computation of a specific value of this variable requires knowledge of the population standard deviation, σ. If the variance of the population is unknown, a more common situation, then the population standard

deviation in the above expression is replaced by the sample estimator for the population standard deviation. The variable

$$z_2 = \frac{\overline{y} - \mu}{s / \sqrt{n}} = \frac{\overline{y} - \mu}{s(\overline{y})}$$

is not distributed as the unit normal. Instead, it has a Student's t distribution with $n - 1$ degrees of freedom. This distribution incorporates the additional variability in z_2 which is due to having to estimate the population standard deviation. Note that for $n > 30$, the Student's t distribution is very close to the unit normal distribution, and the two can be considered equivalent for all practical purposes (Kvanli, 1986).

If the sample size is less than 30, to compute a confidence interval, a table of Student t values must be used instead of a table for the unit normal distribution. The values in such a table depend on two factors: the sample size n, and the confidence level, $(1 - \alpha)$, $\alpha \in [0,1]$. Each value in a t-table can be labeled $t_{n-1}^{\alpha/2}$. For an estimate made from a sample of less than 30,

$$P\left(-t_{n-1}^{\alpha/2} < z_2 < t_{n-1}^{\alpha/2} \right) = 1 - \alpha .$$

Substituting for z_2 and after some algebraic manipulation, we get

$$P\left(\overline{y} - t_{n-1}^{\alpha/2} s(\overline{y}) < \mu < \overline{y} + t_{n-1}^{\alpha/2} s(\overline{y}) \right) = 1 - \alpha .$$

Thus, the true population mean will fall within this interval $(1 - \alpha)$ % of the time over many trials.

The precision as specified by UMTA for Section 15 service consumed reports is in another format. UMTA specifies that the estimates for total system-wide annual passenger

boardings and total system-wide passenger miles be within ±10% of their true population

values at a 95% confidence level. This definition is slightly different from the definition

used in statistical theory. A statistician would require that the true population value be

within ±10% of the sample estimate at a 95% confidence level. Since the difference

between these two definitions of precision is insignificant, the latter will be used hereafter.

Using this latter definition, UMTA requires that

$$\frac{t_{n-1}^{.025} \times s(\overline{y})}{\overline{y}} = 0.1$$

As a final note, consider the situation that faces an agency trying to estimate total

system boardings to this required level of accuracy. How do they decide what the sample

size n should be? As

$$s(\overline{y}) = \frac{s(y)}{\sqrt{n}},$$

we have

$$n = \left(\frac{t_{n-1}^{\alpha/2} \times s(\overline{y})}{0.1 \times \overline{y}}\right)^2.$$

However, this requires knowledge of $s(y_i)$, which may not be available. If it is available,

then it is not likely to be very precise. For if agency staff can describe the variability

among population elements precisely, then it is also likely that they know the elements'

values just as well. Finally, a value of $t_{n-1}^{\alpha/2}$ is required to compute n. But this value

depends on n, which makes the above equation impossible to solve exactly for small n.

Generally, what is done is that a $t_{n-1}^{\alpha/2}$ is selected that is a conservative estimate of

the precision. As stated previously, for n > 30, this value can be replaced by a value from

the unit normal distribution, and does not depend on n. Thus, the lack of knowledge of

s (y) is more significant in terms of the errors in computing the precision of the estimator

for the population mean. Although after the sample is taken s (y) is recomputed from the

fresh data, this sample estimate for the standard deviation of the population is far more

sensitive to departures from normality then is the estimate for the mean.

Dispersion

Confidence intervals for the estimator of the population variance are similar to those

for the mean. Whereas before the estimator of the mean was a normal variate, here the

random variable

$$z_3 = \frac{(n - 1) \times s^2(y)}{\sigma^2}$$

is distributed as a Chi–squared distribution with $n - 1$ degrees of freedom. Like the

Student's t distribution, the Chi–squared distribution has tables of values, each value

labeled by the degrees of freedom and the confidence level. Note that the Chi–squared is

not symmetric as is the Student's t distribution, and thus we need to look up two values for

a given confidence interval. These values are the lower and upper $100 (\alpha/2)$ percentiles

of the $(\chi^2)_{n-1}$ distribution, i.e.

$$P\left\{ (\chi^2)_{n-1}^{1-\alpha/2} < z_3 < (\chi^2)_{n-1}^{\alpha/2} \right\} = (1 - \alpha) \; .$$

Substituting for z_3, and after some algebraic manipulations, we get

$$P\left\{ \frac{(n - 1) s^2(y)}{(\chi^2)_{n-1}^{\alpha/2}} < \sigma^2 < \frac{(n - 1) s^2(y)}{(\chi^2)_{n-1}^{1-\alpha/2}} \right\} = (1 - \alpha) \; .$$

This is the $(1 - \alpha)$ % confidence interval for the population variance, given a sample

estimator for this variance.

Alternatively, this relationship between the sample estimator for the population variance, $s^2(y)$, and the true value of the population variance, σ^2, can be expressed as

$$P\left\{\frac{(\chi^2)_{n-1}^{1-\alpha/2}\sigma^2}{(n-1)} < s^2(y) < \frac{(\chi^2)_{n-1}^{\alpha/2}\sigma^2}{(n-1)}\right\} = (1 - \alpha) \quad .$$

This expression is useful when you know the true population variance, and want to test how precise some estimator of this value is.

3.3 Simple Random Sampling of Elements

In order to illustrate the various sampling methodologies investigated in this research, I will make use of a population of shapes. This population is comprised of three different types of shapes, each type having more or less of a characteristic called "roundness". The main purpose of these illustrations is to demonstrate graphically ways of grouping the elements of the population in order to achieve increased efficiency of sampling. For simple random sampling, there is no grouping involved. The figure for this sampling methodology is one group, the population itself.
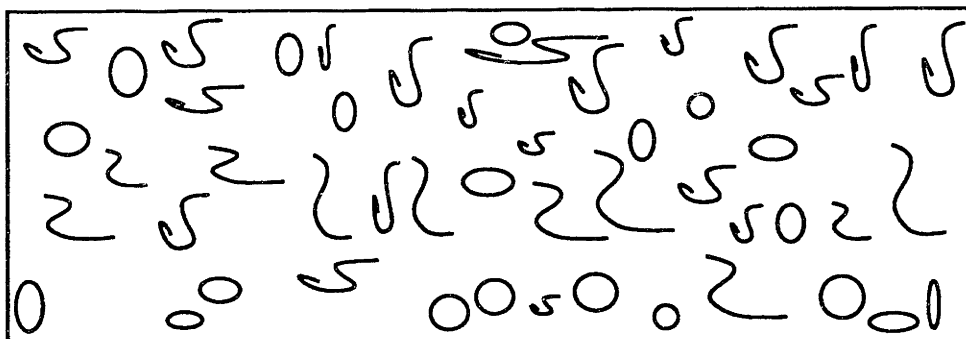

Figure 3.2 A Population of Shapes.

Given an infinitely large population the expressions for $\overline{Y}$ and $s^2(y)$ from normal theory are the estimates for the population mean and population variance from a sample of

size n. Both estimates are unbiased, i.e. if all $_NC_n$ distinct samples are generated, then the average of all the estimates computed will be equal to the appropriate population value (Cochran, 1977).

The above sample estimates of the population mean and variance are valid for sampling without replacement. When sampling without replacement, once an element is drawn, it cannot be drawn again, and each remaining element has an equal probability of being selected. When sampling with replacement, at each draw every element of the population has an equal probability of being drawn. For an infinite population, there is no distinction between these two approaches, no matter how many you draw, there is still an infinity of elements left, and each element has the same probability of being selected as it did on the first draw.

The sample estimate of the population variance is different if the population being sampled does not have an infinity of elements. If N is the number of elements in the population, then the sample estimate of the population variance is

$$s^2(y) = \frac{\sum_{i=1}^{n} (y_i - \overline{y})^2}{n} \cdot \frac{N - n}{N} \ .$$

Another quantity of interest is the variance of the sample estimate for the population mean. This is used to compute the precision of the estimate for the population mean. The less variance the mean estimator has, the more precise it is. Given a sample of n elements from a population of N elements, the estimate for the mean is itself a random variable, taking a different value for each distinct sample. Assuming n is "large enough" and applying the Central Limit Theorem , the sample estimate of the population mean is normally distributed. The variance of this distribution,

$$s^2(\bar{y}) \;=\; \frac{1}{n}s^2(y) \;=\; \frac{1}{n} \frac{\displaystyle\sum_{i=1}^{n} (y_i - \bar{y})^2}{n-1} \frac{N-n}{N} \;,$$

and the standard deviation is

$$s(\bar{y}) \;=\; \sqrt{s^2(\bar{y})} \;=\; \frac{s(y)}{\sqrt{n}} \;.$$

As before, the confidence interval for the population mean is

$$P\left\{ \bar{y} - t_{n-1}^{\alpha/2}\, s(\bar{y}) < \mu < \bar{y} + t_{n-1}^{\alpha/2}\, s(\bar{y}) \right\} \;=\; 1 - \alpha \;.$$

If the sampling fraction n/N is small, the finite population factor is negligible. Cochran suggests that this factor can be ignored if the sampling fraction does not exceed 5% and even 10% in some cases. If the sampling fraction is small, the size of the population does not enter the equation for the variance of the estimate for the population mean. This leads to the somewhat counter-intuitive observation that an estimate of a population mean from a sample of 500 is roughly of the same precision for two populations, one of size 200,000 and one of size 10,000 (Cochran, 1977).

3.4 Stratified Sampling of Elements

Stratification of the population into groups can significantly improve the efficiency of sampling, resulting in either a smaller sample size to produce an estimate of equal precision, or an estimate of higher precision for the same sample size. A group can be defined by any characteristic that seems appropriate, e.g. all trips departing during the same period of time. Whether or not a characteristic is appropriate depends on the variation within groups defined using the characteristic. To increase the efficiency, the variation between elements within the groups must be less than the variation between the elements when they are considered as one big group (i.e. the population).

Returning to the population of shapes, from looking at the total population it is easy to see that the way to stratify the population so as to achieve the greatest increase in efficiency is to group the elements such that each element falls in one of three basic shape types: a hook, a squiggle, or an oval as illustrated in Figure 3.3 below.
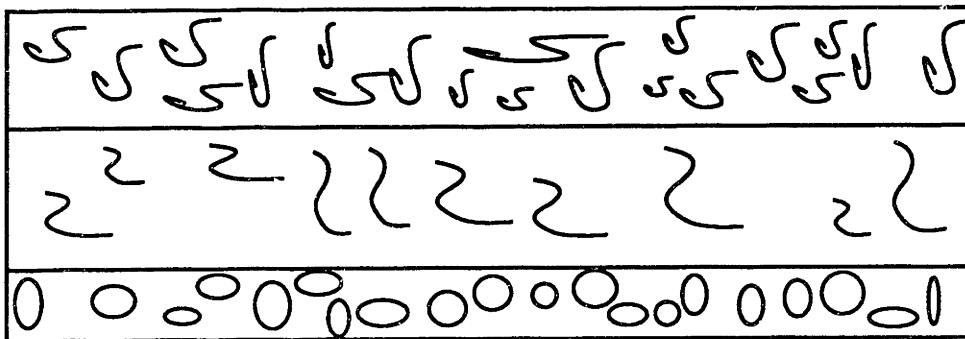
Figure 3.3 A Stratified Population of Shapes.

From this figure it should be clear that it is possible to construct a more efficient estimator for a stratified population than for an unstratified population. First, obtain an independent sample from each stratum. Second, compute the estimate for the population mean by weighing each sample stratum mean by the ratio of population stratum size to total population size.

More formally, stratify the population of N elements into ST groups or strata which are mutually exclusive and collectively exhaustive. Each group h contains N_h elements, labeled $y_{h1}, \ldots, y_{hN_h}$. A total sample of n elements is taken, with n_h elements selected from stratum h. The estimate of the population mean from this stratified sample is

$$\overline{y}_{STRAT} = \frac{\sum_{h=1}^{ST} N_h \overline{y}_h}{N} = \sum_{h=1}^{ST} W_h \overline{y}_h \quad ,$$

where W_h is the weight factor and is equal to N_h/N.

45

If each stratum mean $\bar{y}_h$ is unbiased, then this estimate of the population mean is also unbiased (Cochran, 1977). If the samples are drawn independently from the strata, the sample estimator for the variance of $\bar{y}_{STRAT}$ is

$$s^2(\bar{y}_{STRAT}) = \sum_{h=1}^{ST} w_h^2 \ s^2(\bar{y}_h)$$

,

where

$$s^2(\bar{y}_h) = \frac{1}{n_h - 1} \sum_{i=1}^{n_h} (y_{hi} - \bar{y}_h)^2 \ .$$

For situations where $\bar{y}_{STRAT}$ is normally distributed and $s^2(\bar{y}_{STRAT})$ is well determined (i.e. "large enough" n), then the confidence interval for the stratified estimator of the population mean is

$$P\left\{ \ \bar{y} - t_{n-1}^{\alpha/2} \ s(\bar{y}_{STRAT}) < \mu < \bar{y} + t_{n-1}^{\alpha/2} \ s(\bar{y}_{STRAT}) \ \right\} = 1 - \alpha \ .$$

Note that this confidence interval is exactly the same as the one for simple random sampling, except that the variance of the estimator is different. This observation is consistent for all sampling plans.

Assuming that a statistician does not have reliable a priori information regarding the intra–strata variances, there is a tradeoff between the number of strata the population is partitioned into, ST, and the ability to obtain a good estimate of the variance of each stratum, $s^2(\bar{y}_{STRAT})$, and hence a good estimate of the precision of the mean estimator. The more groups the population is partitioned into, the more precise the estimate of the population mean will be. However, the more groups there are, the fewer samples there will be per strata, and the less "well determined" the strata variances.

In general, the distribution of $s^2(\bar{y}_{STRAT})$ is too complex to permit use of the Student's t table for small stratum samples size although one can approximate the "effective" degrees of freedom for $s(\bar{y}_{STRAT})$ when the n_h are small. Satterthwaite (Miller, 1986) presents an estimate for the effective degrees of freedom,

$$n_e = \frac{\left(\sum_{h=1}^{N_s} g_h s_h^2 \right)^2}{\sum_{h=1}^{N_s} \frac{g_h^2 s_h^4}{n_h - 1}} ,$$

where

$$g_h = \frac{N_h(N_h - n_h)}{n_h} .$$

Note, however, that this estimate assumes that the y_{hi} are normally distributed. If the distribution of y_{hi} has a positive kurtosis, then the sample variance is underestimated, the effective degrees of freedom are overestimated, and the estimated confidence level is overstated (Miller, 1986).

3.5 Cluster Sampling

While stratification of the population improves the efficiency of estimation, clustering the sample can reduce the cost of collecting data. For example, if a sample is collected by selecting elements sequentially in time, the cost of collecting the sample may be lower than if the elements are selected randomly in time. To reduce the potential loss of efficiency due to selecting elements in a non-random manner, the population is clustered into groups which have little variation between groups. If the population is partitioned into

such groups or clusters, then the loss of precision due to sampling entire clusters can be small relative to the savings in collecting the data.

For example, it is often reasonable to assume that the ridership on a bus line varies little between weekdays. Assuming this is true, a sample which contains all trips departing on one weekday will produce an estimate that is no less efficient than an estimate computed from a sample of the same size randomly selected from all departures over a week. (In fact, if ridership varies significantly over different time-of-day periods, the sample of all trips departing on a weekday is likely to be more efficient.) It is a much simpler and less expensive proposition to measure the ridership on all trips on the same day instead of some random scattering of trips across all departures in a month.

The presentation of cluster sampling methods in this paper is not exhaustive. Notably absent is any mention of the class of cluster sampling where the clusters are chosen with probability proportional to their size. As the cluster units in this study are days and all days contain roughly the same number of elements (departures), equal probability sampling of clusters is sufficient.

3.5.1 Clusters of Equal Size

To use this sampling plan, the population must be partitioned into clusters of equal size. Returning to the population of shapes, each cluster should have the same number of elements, and roughly the same proportion of each shape type. Thus, there is little variation between the clusters. One such partitioning is illustrated below.
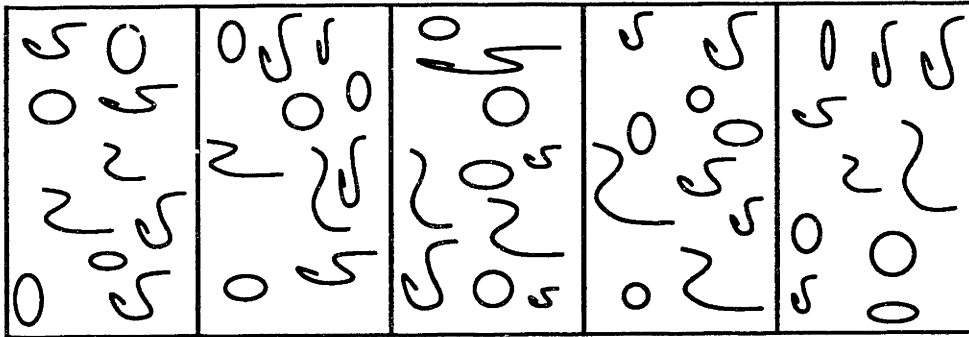
Figure 3.4 A Clustered Population of Shapes.

A random sample of cl clusters is selected from a population of CL clusters with each cluster containing EL elements. Thus the total sample size $N = CL * EL$.

Mean per Cluster Estimator.

One estimator for the population total is generated by computing the sample mean per cluster and multiplying by the total number of clusters in the population. This estimator is in effect the same as a simple random sample of population elements with an element redefined to be a cluster. If the value of element j in cluster i is y_{ij}, then the total for cluster i is

$$y_i = \sum_{j=1}^{EL} y_{ij} \quad ,$$

and the estimator of the population total is

$$Y_{CLUST1} = \frac{CL}{cl} \sum_{i=1}^{cl} y_i$$

with variance

$$s^2(Y_{CLUST1}) = \frac{CL - cl}{CL} \frac{CL^2}{cl} \frac{1}{cl - 1} \sum_{i=1}^{cl} (y_i - \overline{Y}_{CLUST1})^2 \quad .$$

49

The estimator for the population mean per cluster is equal to the estimator for the population total divided by the number of clusters in the population, or

$$\overline{Y}_{CLUST1} = \frac{Y_{CLUST1}}{CL} \quad .$$

The estimator for the population mean per element is equal to the estimator for the population total divided by the total number of elements in the population, or

$$\overline{\overline{Y}}_{CLUST1} = \frac{Y_{CLUST1}}{N} \quad ,$$

with variance

$$s^2\left(\overline{\overline{Y}}_{CLUST1}\right) = \frac{s^2\left(Y_{CLUST1}\right)}{N^2} \quad .$$

Mean per Element Estimator.

Another estimator that can be constructed from a clustered sample uses the mean per element instead of the mean per cluster. The mean per element estimator,

$$\overline{\overline{Y}}_{CLUST2} = \frac{1}{cl\ EL} \sum_{i=1}^{cl} \sum_{j=1}^{EL} Y_{ij}$$

is an unbiased estimate of the population mean (Cochran, 1977). The variance of this estimator can be expressed in terms of the intracluster correlation coefficient, defined as

$$\rho = \frac{E\left(Y_{ij} - \overline{\overline{Y}}_{CLUST2}\right)\left(Y_{ik} - \overline{\overline{Y}}_{CLUST2}\right)}{E\left(Y_{ij} - \overline{\overline{Y}}_{CLUST2}\right)^2}$$

$$= \frac{2 \sum_{i=1}^{cl} \sum_{j=1}^{EL} \sum_{k=j+1}^{EL} \left(Y_{ij} - \overline{\overline{Y}}_{CLUST2}\right)\left(Y_{ik} - \overline{\overline{Y}}_{CLUST2}\right)}{\left(EL - 1\right)\left(cl \cdot EL - 1\right)s^2(y)}$$

where

$$s^2(y) = \frac{\sum_{i=1}^{cl} \sum_{j=1}^{EL} \left(y_{ij} - \overline{\overline{y}}_{CLUST2}\right)^2}{cl \cdot EL - 1}.$$

The variance of this estimator is

$$s^2\left(\overline{\overline{y}}_{CLUST2}\right) = \frac{CL - cl}{CL} \frac{1}{cl} \frac{cl \cdot EL - 1}{EL^2(cl - 1)} s^2(y)\left(1 + (EL - 1)\rho\right).$$

Note that except for the last term in parenthesis, this expression is the same as that for the estimator from a simple random sample of EL·cl elements. So if the factor

$$\left(1 + (EL - 1)\rho\right)$$

is less than 1, the variance of $\overline{\overline{y}}_{CLUST2}$ is less than that of a simple random sample of the same number of elements, and thus more efficient. This factor is less than one for $\rho < 0$. This factor is called the *deff* (*design effect*) factor, as it represents the change in precision due to changing the design of the sampling plan from simple random sampling (Cochran 1977).

3.5.2 Stratified Clusters.

The techniques of clustering and stratification can be combined to improve the cost efficiency of sampling. Such a plan comes in two basic varieties. The first method is first to stratify the population, as illustrated in Figure 3.3, and then break each stratum into clusters. If the clusters are selected from each stratum independently, then the cluster estimators for each strata can be combined using the theory for a stratified estimator from Section 3.4 to create population estimators.

Another method of combining clustering and stratification is first to break the population into clusters, randomly select clusters, and then stratify the clusters that have been selected. The stratum estimators are combined to create a population estimator. This second method is useful when the groups most suited to be clusters contain multiple strata.

Such is the case in this study, where the natural cluster unit is a day and the natural strata are the time periods during the day.

Furth (1987) presents an estimator for this second method of grouping the population. This estimator is more general than is needed here; he allows each cluster to have a different number of strata, and each stratum within a cluster to have a different number of elements from the same stratum in a different cluster. In this application, a day cluster will always have the same number of time-period strata. However, because of variation in the boat schedule over time, there are times where the same stratum will contain different numbers of elements in different clusters.

Returning one last time to the population of shapes, the partitioning of the population required by this sampling plan is illustrated in Figure 3.5.

Figure 3.5 A Clustered and Stratified Population of Shapes

Adopting Furth's terminology, a day is called a "supercluster" (a column in Figure 3.5), a time period is a stratum (a row in Figure 3.5), and a day-time-period is a cluster (one of the 15 boxes in Figure 3.5).

On to the notation. There are SCL superclusters in the population, scl of which are randomly selected for the sample. Each supercluster has ST strata within it. Thus there are CL clusters per stratum for the population, and cl clusters per stratum for the sample.

The number of trips in stratum h cluster i is EL_{hi}. The value for element j of stratum h cluster i is y_{hij}. The total number of elements from stratum h that are in the sample is

$$el_h = \sum_{i=1}^{cl} EL_{hi}$$

and the estimator for the mean per element in stratum h is

$$\bar{y}_h = \frac{\sum_{i=1}^{cl} \sum_{j=1}^{EL_{hi}} y_{hij}}{el_h} .$$

If the number of elements in stratum h population-wide is

$$EL_h = \sum_{i=1}^{CL} EL_{hi}$$

then the estimator for the population total is

$$Y_{CLUST4} = \sum_{h=1}^{ST} \bar{y}_h EL_h .$$

As the selection of clusters is not random, i.e. all clusters within a supercluster are selected, the variance estimator for the system total must include terms which capture the covariance between clusters that lie in the same supercluster. The following derivation omits terms that represent second order interaction effects (or higher). This estimator for the variance of Y_{CLUST4} is

$$s^2(Y_{CLUST4}) = \sum_{h=1}^{ST} EL_h^2 s^2(\bar{y}_h) + 2\sum_{h=1}^{ST} \sum_{h'=h+1}^{ST} EL_h EL_{h'} \cdot Cov(\bar{y}_h, \bar{y}_{h'}) .$$

The first term in this expression represents the component of the variation within clusters (intracluster variance), and the second term the variation between clusters but within superclusters (intercluster and intrasupercluster variance). Starting with the intracluster

53

variance, we know that the variance of the stratum mean estimator is equal to the variance of the estimator of the stratum total divided by the total elements in the stratum squared, or

$$s^2(\overline{y}_h) = \frac{s^2(Y_h)}{EL_h^2} \, .$$

If y_{hi} is the total of all elements in stratum h cluster i, i.e.

$$y_{hi} = \sum_{j=1}^{EL_{hi}} y_{hij} \, ,$$

then using the theory of ratio to size estimator we can state that

$$s^2(Y_h) = \frac{CL^2}{cl} \frac{CL - cl}{CL} \frac{1}{cl-1} \sum_{i=1}^{cl} (y_{hi} - EL_{hi}\overline{y}_h)^2 \, .$$

Putting this all together, we get

$$s^2(\text{intracluster}) = \frac{CL^2}{cl} \frac{CL - cl}{CL} \frac{1}{cl-1} \sum_{h=1}^{ST} \frac{1}{EL_h} \sum_{i=1}^{cl} (y_{hi} - EL_{hi}\overline{y}_h)^2 \, .$$

To compute the intercluster component of the variation, we use the relation that

$$Cov(\overline{y}_h, \overline{y}_{h'}) = \frac{1}{el_h el_{h'}} Cov\left[\sum_{i=1}^{cl} y_{hi} \, , \, \sum_{i=1}^{cl} y_{h'i}\right] \, .$$

Using the identity

$$Cov\left(\sum^i A_i \, , \, \sum^j B_j\right) = \sum^i \sum^j Cov(A_i \, , \, B_j)$$

we get

$$Cov\left(\overline{y}_h , \overline{y}_{h'}\right) = \frac{1}{el_h el_{h'}}\sum_{i=1}^{cl} \sum_{j=1}^{cl} Cov\left(y_{hi}, y_{h'j}\right) .$$

Because the superclusters are selected independently, the covariance between two clusters is non-zero only for those clusters which are contained within a common supercluster. Given that the two clusters h and h' are in the same supercluster, the covariance between these two is

$$s_{hh'}^2 = \frac{\sum_{k=1}^{cl}\left(y_{kh} - EL_{kh}\overline{y}_h\right)\left(y_{kh'} - EL_{kh'}\overline{y}_{h'}\right)}{cl - 1} .$$

Substituting for $Cov(y_{hi}, y_{h'i})$ in the previous equation, we get

$$Cov\left(\overline{y}_h , \overline{y}_{h'}\right) = \frac{1}{el_h el_{h'}}scl_{hh'} \cdot s_{hh'}^2 ,$$

where $scl_{hh'}$ is the number of superclusters in the sample that contain both a cluster h and a cluster h'. Putting this all together, we get

$$s^2(\text{intercluster}) = 2\sum_{h=1}^{ST} \sum_{h'=h+1}^{ST} \frac{EL_h EL_h}{el_h el_{h'}}scl_{hh'} \cdot s_{hh'}^2 .$$

Chapter 4 The Experiments

This chapter provides a brief history of the transit operation which produced the ridership data set used in this study and presents the experimental design and the results of these experiments. The efficiency of five sampling plans is compared, first on the basis of sample size, and then on the basis of a parameterized cost function. Simple random sampling of departures is used as a base for comparisons, against which proportional stratified sampling by time-of-day (TOD), and cluster sampling are compared. Three different implementations of cluster sampling are examined.

The first cluster sampling plan selects day-clusters randomly from all days (excepting holidays) on which service was provided during the year. The relative efficiency of this method depends on how the average boardings per departure varies from day-to-day in comparison to the variation between all departures over the year. The second cluster sampling plan selects days on a systematic basis in an attempt to exploit any systematic patterns in ridership due to seasonal and/or growth factors. If there are regular patterns in the average boardings per departure on a daily basis, then this method will be more efficient than simple random selection of days. The third cluster sampling method randomly selects day-clusters as in method one, stratifies the departures by TOD, and computes a stratified estimate for average boardings per departure. If the benefits of stratification by TOD outweigh the within day correlation between time periods, then this method will be more efficient that simple random selection of clusters.

Based on the population parameters found in this study and reported in others, eight typical bus routes are defined and the relative cost-efficiency of proportional stratified sampling is compared with that of systematic selection of clusters. A cost function is developed to define the cost of data collection in terms of the frequency of service, the length of the route, the daily hours of service, the coefficient of variation of average

boardings per trip per day and the coefficient of variation of average boardings per departure within each day.

4.1 The MBTA Commuter Boat Service

4.1.1 Service History

The Massachusetts Bay Transportation Authority (MBTA) operates a commuter boat service between Hingham and Boston. The trip covers a distance of roughly 10 miles in 35 minutes, with an average speed of 15 knots*. The Commuter Boat Service began in 1984 as part of the effort to mitigate delays anticipated from reconstruction of the Southeast Expressway. As a result of a competitive bidding process, two operators, Boston Harbor Commuter Service and Massachusetts Bay Commuter Services were chosen as joint contractors. MBCS had to give up its unsubsidized ferry service to George's Island in order to receive the subsidy payment. BHCS, a new operator in the harbor as of the spring of 1984, purchased two high speed vessels originally designed to service oil rigs in the Gulf of Mexico.

The contracts are put out for rebidding every three years. However, because the capital equipment required to provide the service is expensive and uncommon, there has not been much competition. The service requires boats which can carry at least 150 people and cruise at speeds of up to 24 knots. The cost of such a boat is at least $500,000. In the six year history of the service, the only competition occurred in the bidding for fiscal year 1989, when there were three companies competing for two spots. During the service history, the rebidding process could be characterized as more of a renegotiating period with the current contractors. Only four different operators have ever provided the service and one of these was terminated after one month. MBCS, one of the original contractors,

---

* One knot is a nautical mile per hour. A nautical mile is 1.11 miles.

eventually went out of business. BHCS has provided service since its inception and was the sole provider for a time after MBCS went out of business. Massachusetts Bay Lines (MBL) joined BHCS four years ago, and continues to provide service today.

Originally, MBCS provided 18 one-way trips daily and BHCS provided 10. Currently, BHCS provides 22 one way trips daily and MBL 16. In the interim period, the boat schedule has changed 28 times although many of these changes were minor, for example in 16 changes the total departures changed by no more than two trips. A number of the changes were made in response to growth and seasonal variation in ridership, especially during the first three years of service.

Aside from the initial phases of the service, the subsidy structure has remained the same. When the service began in March 1984, each operator was paid $1.50 for every seat during rush hour trips regardless of the ridership. During Phase I, which ran from March to June 1984, the service levels, schedules and ridership patterns were monitored closely, and as a result, the subsidy structure and schedules were modified for Phase II. During this second phase, which ran from June 1984 to November 1985, the subsidy formula was slightly more complicated; on all trips, a direct subsidy of $2.70 was paid for all empty seats, as long as at least 20 passengers boarded the boat. A vessel carrying 100 passengers was considered full and received no subsidy. In addition, the commonwealth paid $150 per day to keep a second BHCS boat as a contingency vessel. The current subsidy structure has been in place since the end of Phase II. The operator is penalized if on-time weekly performance is not at least 90%. On-time means the trip arrives or departs no more than 5 minutes after the scheduled time. In addition, the total late time for trips operated during a day should not exceed 15 minutes. The contract also includes financial incentives for the operators to increase ridership: for every passenger over a minimum daily ridership, the operator retains an average of $1.38. This incentive can be significant, for example, on every day of service in June 1990, one operator carried at least 500 people over the minimum level.

Considering the primitive state of marketing in mass transit, it is no surprise that in 1984 there was no marketing done for the service. The majority of riders who first used the service found out about it through word of mouth. Once a person made it to the docks, they were not informed about waiting room facilities. In the first user survey, "comfort of waiting/boarding areas" received the lowest rating: the facilities existed, but the organization was poor. For a time, the Boston waiting area was closed during rush hour! The first batch of schedules, printed in February of 1984, was exhausted quickly, and most people were hard pressed to get their hands on one. The operators were likely to give out information pertaining only to the trips they ran, and not the other company's trips.

Given the troubled beginning, the service has done extraordinarily well, with ridership increases exceeding all predictions. The second user survey, done in 1985 by the MIT course 1.102, had 35% of all rider comments express that the service was "great, outstanding, or must continue". Another 14% wanted more frequent service. Most people said that they would not switch back to their previous mode after the expressway work was complete. An interesting result from this survey was that most people did not start using the boat because of construction delays. Most thought that if more people knew about the service, patronage would grow considerably, as indeed it has as discussed in the following section.

4.1.2 Seasonal Trends and Growth Rate

When considering seasonal trends and growth rate of ridership, it is important to note the distinction between these trends in terms of total boardings and in terms of average boardings per trip. Total boardings provides a clear measure of overall seasonal and growth-related patterns. The scheduling of departure times and frequencies of departures acts as a filter through which the total boardings pattern is passed to produce the patterns in average boardings per trip. The difference between these two measures is illustrated below in Figures 4.1 and 4.2.
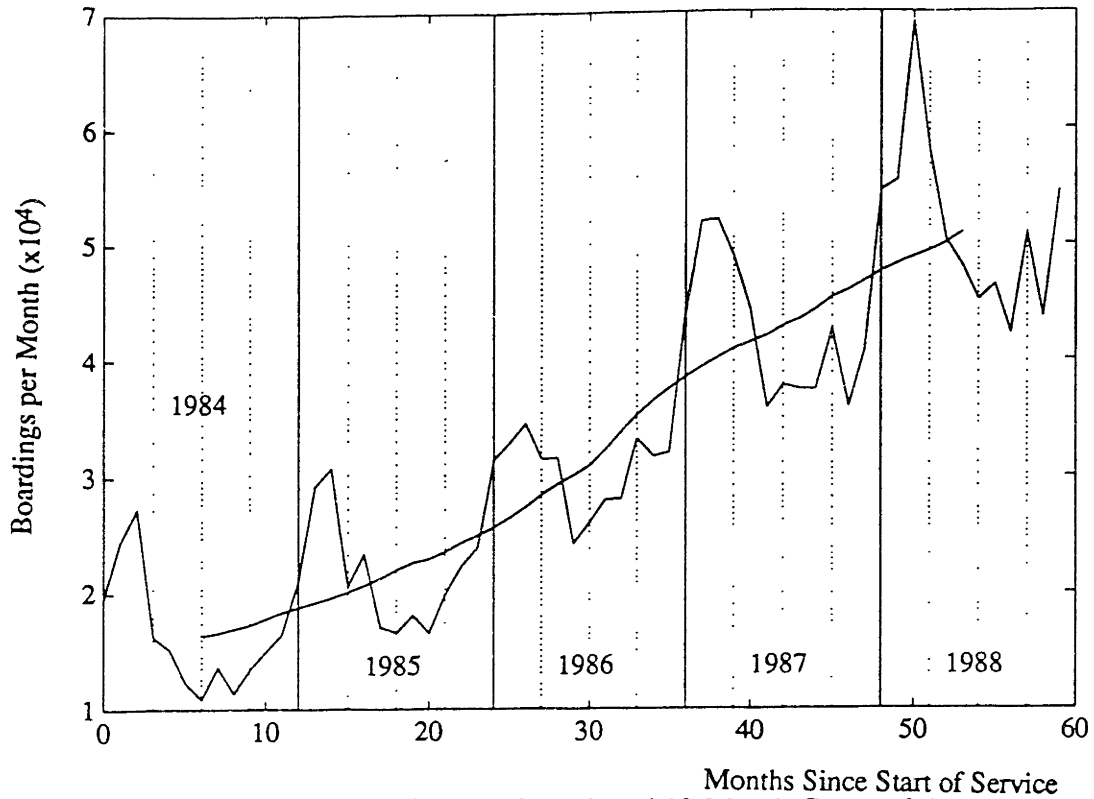
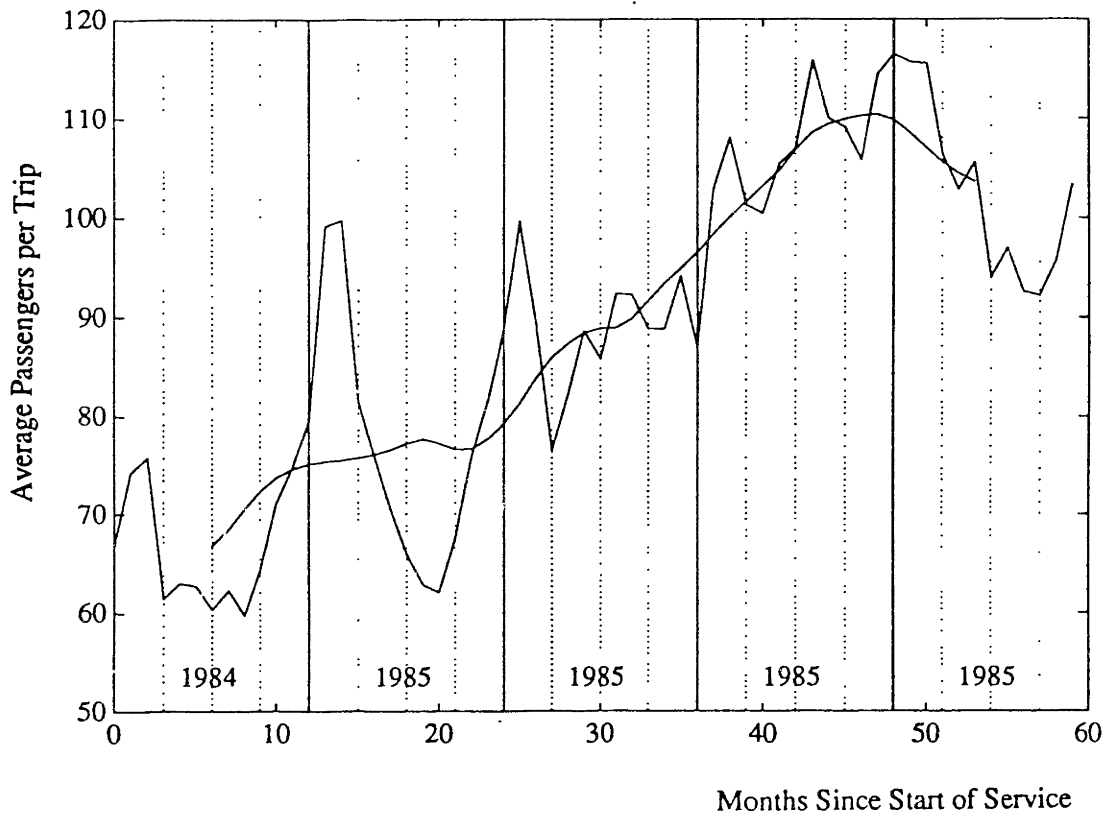Figure 4.1 Total Boardings per Month and 12-Month Centered Average

Figure 4.2 Average Boardings per Departure per Month and 12-Month Centered Average

60

In contrast to the continuous growth in total passenger boardings per month, the growth trend for average boardings per departure stops in the fourth quarter of boat-year (BY) 1987*, and declines thereafter. In addition, the regular seasonal pattern apparent in the plot of total boardings is not as clear in the plot of average boardings, particularly in BY1986 and BY 1987. Finally, the variation in total boardings about its 12-month centered average grows as time passes, whereas the variation of average boardings decreases.

These differences can be explained by examining how the number of departures per day varies over this five year period. Appendix A contains plots showing all the changes in the number of departures per day for each of the five years. The drop-off in growth in average boardings per trip in BY1988 is due mainly to the increase in the number of daily departures. In the winter and spring of BY1987, a total of 17 trips departed day. In the summer of BY 1988, the number of departures increased to 22 in June, then to 26 in July, August, and September, and decreased to 24 for the remainder of BY1988. Departures per day thus increased more than the supply of passengers, and while the total boardings per month increased modestly, the average boardings per trip decreased.

The perturbations of the seasonal patterns are likewise due to schedule changes. In both BY1986 and BY1987, the number of trips was reduced in the fall, which effectively offset the decreases evident for these two quarters in total boardings. This level of departures per day remained the same for the winter of BY1986 and decreased an additional 2 trips in BY1987, which accounts for the average boardings per trip being above the centered average for these two winter seasons.

_____

*   A boat-year starts in June and continues through to May of the following calendar year. Thus, the first quarter (summer) of a BY is contains June, July, and August, the second quarter (fall) contains September, October, and November, the third quarter (winter) contains December, January, and February, and the fourth quarter (spring) contains March, April, and May.

Figure 4.2 can be considered a measure of how effectively the schedules were adjusted to correspond to ridership patterns over the years. Going to an extreme, if the supply of transportation exactly matched the variation in demand, then Figure 4.2 would be a horizontal line, intercepting the y-axis at the desired level of average monthly vehicle utilization. In this context, the decrease in variation about the 12-month centered average shown in Figure 4.2 for BY1986 and BY1987 indicates that the scheduling best matched the systematic factors in these two years. BY1985 is clearly the worst in this respect, having constant 15 departures per day for the entire period. BY1988 shows less responsiveness to the seasonal patterns, but capped the annual passenger growth at roughly 110 passengers per trip. Given that the average capacity of a departure in BY1988 was 165 passengers, this level corresponds to an average vehicle utilization rate of 67%.

4.2 The Sampling Plans Investigated

Five of the sampling methods described in Chapter 3 were investigated in this research. These five are: simple random sampling, proportional sampling stratified by time-of-day (TOD), simple random sampling of day-clusters, systematic sampling of day-clusters, and simple random sampling of day-clusters which are then stratified by time-of-day. Simple random sampling is the base case against which the other plans are compared. Proportional stratified sampling exploits the variation in average boardings per trip between different times of day, e.g. the a.m. peak period versus the mid-day period. Simple random cluster sampling investigates the consequences of sampling trips by selecting all trips departing during a day. Pre-stratification of clusters is aimed an exploiting the systematic patterns of seasonality and growth, and post-stratification by TOD attempts to gain the efficiency of proportional stratified sampling while maintaining the cost benefits of cluster sampling.

Each plan was implemented on the five years of ridership data, one year at a time. The word population as used in the descriptions below refers to one ooat-year of ridership

data, which includes all days of service (only weekdays) minus holidays. Appendix B lists the days left out for each year. Below is a description of how each plan was implemented.

4.2.1. Simple Random Sampling (without replacement)

Simple random sampling makes no attempt to exploit any systematic patterns that exist in average boardings per departure. Neither the pattern between departures within a day nor those between months within a year are utilized. All departures over the year are included in one group, from which a random sample is selected as follows:

(1) Store the ridership of each departure during the year in a list. Index this list from 1 to N, where N is the total number of departures.

(2) Generate a random number rv between 1 and N inclusive, where each number between 1 and N has an equal probability of being generated.

(3) Check to see if this number has been used before. If

yes, discard and generate another.

no, select corresponding departure and add the rv to a list of numbers already used.

(4) Repeat until n departures have been selected.

4.2.2. Proportional Stratified Sampling

Proportional stratified sampling attempts to exploit the difference in average boardings per departure that exist over different hours within a day. In this investigation, four within-day strata were used. Based on departure time, each departure during the year is grouped into one of these four strata, and a random sample is selected from each stratum. The stratum sample size is proportional to the total number of trips the stratum contains.

Two different approaches were used to define the strata. The first method partitioned the day into four contiguous time-of-day (TOD) periods: one for the a.m. peak period, one for the mid-day period, one for the p.m. peak period, and one for the evening

63

period. The second method partitioned departures based on the yearly average boardings per departure. Each method utilized a time-of-day profile for each year which plots the average yearly boardings per trip as a function of time of day, with one value for every 15 minute interval. These profiles are contained in Appendix C.

The first method of defining the strata was implemented by looking at these profiles and visually partitioning the day into four contiguous blocks of time. Using simple heuristics such as the a.m. peak period should end when the a.m. rush hour has tapered off, the TOD groups in Table 4.1 and in Figure 4.3 were defined.

h	stratum h	depart during
0	a.m. peak	6:00 to 8:30
1	mid day	8:31 to 15:15
2	p.m. peak	15:16 to 18:30
3	evening	18:31 to 21:00

Table 4.3 Contiguous Time of Day Strata
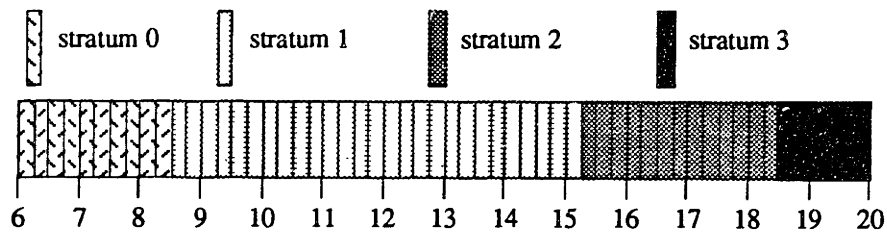
Figure 4.3 Contiguous Time of Day Strata.

The second method of defining the strata was that of Dalenius and Hodges (Cochran 1977) to obtain a quick approximation to the optimum strata definition. This method groups the average boardings per trip on the daily profile by their magnitude, and does not necessarily produce contiguous blocks of time. This method was applied

64

separately to each year, the results of which were combined to produce the strata illustrated

below in Figure 4.4.
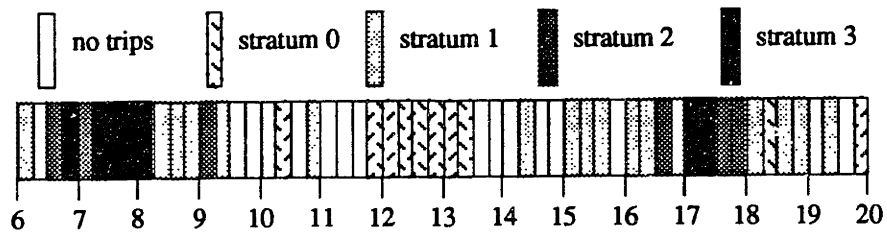
Figure 4.4 Non-contiguous Strata

The effectiveness of these two groupings can be compared by computing the

coefficient of variation between and within the TOD strata. The grouping with the smaller

$COV_W(TOD)$'s will be the more efficient strategy. As Table 4.2 below shows, the second

stratification scheme is more effective, achieving an average decrease in $COV_W(TOD)$ of

26%.

	Method 1		Method 2	
Year	COV_W	COV_B	COV_W	COV_B
1984	.451	.449	.359	.559
1985	.378	.400	.284	.457
1986	.346	.304	.270	.461
1987	.360	.396	.286	.432
1988	.379	.385	.317	.457

Table 4.2 A Comparison of Two Stratification Schemes.

The drawback to using this second scheme of stratification is that in order to

implement it, the operator must have a good idea of the average boardings per departure. If

an operator has some means of ascertaining these data, then the non-contiguous strata will

produce a more efficient stratification scheme. However, as this information may not be

readily available, the contiguous TOD strata are easier to define and implement. It is assumed that most transit monitors would be more likely to use the contiguous TOD strata, and this is the sampling plan used in the cost comparison.

The proportional stratified sampling plan was implemented as follows.

(1) Group each departure into one of the strata defined in Figure 4.3 .

(2) Count the number of elements in each stratum h, N_h, and the number of elements in the population, N.

(3) Compute the stratum h sample sizes, n_h using the formula

$$n_h = \text{round}\left[\left(\frac{N_h}{N}\right) n\right].$$

Here, the operator round rounds the floating point number within the brackets to the nearest integer. It is possible that the n_h do not add up to n. If their sum is greater (less) than n, reduce the n_h that was rounded up the most (least).

(4) Obtain a simple random sample (without replacement) of n_h from each stratum h.

4.2.3. Simple Random Cluster Sampling

Simple random cluster sampling is aimed at exploiting the cost efficiency of collecting samples concentrated in the minimum number of days instead of randomly scattered over the year. Implementation is the same as for simple random sampling, with departures replaced by days. This plan was implemented as follows:

(1) Group each departure into a day cluster with all departures on the same day going into the same cluster.

(2) Count the total number of clusters, CL, in the population.

(3) Obtain a simple random sample of cl clusters. For comparison with the other methods, the sample size

```
n = round[cl * (avg. number of departures per day)]
```

66

is used.

(4) Divide the total passenger boardings by the total number of departures to estimate average boardings per departure.

4.2.4 Systematic Cluster Sampling

Pre-stratification of clusters by season is aimed at improving efficiency by exploiting any systematic patterns of seasonal or growth trends. The year is divided into contiguous groups of days of equal size, and one day is randomly selected from each of these groups. By spreading the sample evenly over time, any systematic patterns over the months or between seasons will be better represented. If these patterns are significant, then systematic cluster sampling will be more efficient.

(1), and (2) as above in Section 4.2.3.

(4) For each sample size cl, partition the days in the population into cl groups, each of size

$$group\_size = (int) (CL / cl).$$

(The (int) operator returns the integer part of its argument.) As $group\_size$ must be an integer, it is possible to have more days in the year than cl * $group\_size$. Any such remaining days were included in the last group.

(5) Randomly select a day in each group.

4.2.5 Post-Stratification of Clusters by TOD

Here, a simple random sample of clusters is obtained as in section 4.2.3, and then the departures on these days are stratified into the TOD groups defined in section 4.2.2.. The variance of this estimator depends on two factors: the between-day variance of each TOD group, and the within-day covariance of the TOD groups. The between-day variance of the average daily boardings per departure for each TOD group should be lower than the between-day variance of the average daily boardings per departure for all TOD groups

taken together, as trips within the same strata have similar characteristics. If the reduction in variation due to this stratification is greater than the increase in variation due to the correlation between departures leaving on the same day, then this sampling method will improve efficiency.

Somewhat surprisingly, in three out of the five boat-years investigated, the post-stratification of clusters was less efficient than simple random selection of clusters. In BY1984, the simple random selection of clusters was significantly more efficient than post-stratification, while in BY1987, the post-stratification was more efficient than the simple random selection. The variances of these two estimators are presented below in Table 4.3 for each boat-year.

	1984	1985	1986	1987	1988
Post stratification	12.20	13.88	8.34	8.75	12.65
Simple random	8.43	13.81	8.89	10.19	11.03

Table 4.3 Variance of Post-Stratification of Clusters and Simple Random Estimators

The contribution of the covariance term to the total estimator variance varied widely over the five years, reaching a high of 61% in BY1986 and a low of 33% in BY1985. The average proportion over the five years was 50%. This lack of stability was an aggregate result of a similar lack of stability in the covariances between TOD periods within each year. Table 4.4 below presents the correlation coefficients between TOD groups for each year. All values, except one, are greater than zero, indicating that when ridership goes up (down) in one time-period, the increase (decrease) occurs in all TOD groups. The two time periods with the largest correlation are a.m. peak and p.m. peak, which is no surprise as most people who take the boat to work also take the boat home. However, even this most

obvious correlation has an exception: in BY1986, these two periods have a correlation of 0.004.

This very low value is due to two effects unique to this year. First, until the end of August, the p.m. peak ridership was much higher (e.g. 50 riders per day per p.m. peak) than its average, while the a.m. peak was equal to or slightly lower than its average. Second, in March 1987 one trip was added to the p.m. peak while no change was made to the a.m. peak. Thus, for the three spring months, the p.m. peak ridership was lower than its average while the a.m. peak ridership was higher. These two sections of the year resulted in enough negative terms in the covariance calculation to produce the low value that appears below.

TODs	BY1984	BY1985	BY1986	BY1987	BY1988
a.m. / mid	.21	.57	.04	.07	.21
a.m. / p.m.	.74	.83	.004	.63	.78
a.m. / eve	.38	.60	.26	.14	.37
mid / p.m.	.42	.80	.49	.36	.68
mid / eve	.90	.54	.47	.40	.48
p.m. /eve	.53	.55	.03	-.04	.50

Table 4.4 Correlation Coefficients of Time-of-Day Groups for the Commuter Boat

The other rows in this table show similar instability. For example, the correlation coefficients for p.m. peak /evening TOD groups is near 50 for three years, and near 0 for the other two. It is difficult to isolate the causes of such instability, or even provide a rationalization in some cases. Due to the lack of constant parameters, the complexity of assigning cause, and the general lack of any efficiency gain over simple random selection of clusters, this method was not pursued further.

4.3 The Stability of Parameters of Variation

The relative efficiency of these sampling plans depends on how the average boardings per departure varies over time. The efficiency of simple random selection of departures depends on the variation in boardings per departures between all departures over the year. Stratified sampling of departures improves efficiency if the variation within each stratum is less than this population variance. Cluster sampling of days will improve efficiency if the variation between daily mean of boardings per departure is small relative to the population variance. These measures of variation are represented by the coefficient of variation between all departures within the population, COV(pop), the coefficient of variation between the daily average of boardings per departure, $COV_B(day)$, and the average coefficient of variation between departures within each stratum, $COV_W(TOD)$. Table 4.4 below shows these population parameters for each of the five years as well as the average coefficient of variation between departures within each day, $COV_W(day)$, and the coefficient of variation between the stratum averages of boardings per departure, $COV_B(TOD)$. (The contiguous TOD strata used here.)

Referring to this table we see that stratification by time-of-day will increase efficiency over simple random sampling, as the $COV_W(TOD)$ is consistently lower that the COV(pop). In addition, the low values for the between day coefficients of variation give us hope that using cluster sampling will not result in much lower efficiency relative to simple random sampling.

Parameter	1984	1985	1986	1987	1988
COV(pop)	.59	.51	.42	.49	.50
$COV_W(TOD)$	.45	.38	.35	.36	.38
$COV_B(TOD)$	.45	.40	.30	.40	.39
$COV_W(day)$	.61	.50	.42	.49	.50
$COV_B(day)$	.12	.18	.10	.10	.11
avg. trips / day	11.7	13.9	16.9	19.8	24.1

Table 4.5 Population Parameters for the MBTA Commuter Boat.

Considering the many differences between the years in the data set, the stability of these parameters is truly remarkable. The service experienced large growth, multiple schedule changes, and five different New England winters, yet the parameters show very little difference from year to year. The COV(pop) is right around .50 for all years except 1984 and 1986. The higher value in 1984 is not surprising because this is the first year of service, and both operators and passengers can be expected to exhibit more variation. In 1986, there generally seems to be less variation than in other years, possibly due to more effective scheduling. Excluding 1984, the following parameters are very similar:

- $COV_W(TOD)$ ranges from .27 to .32, with a mean of .29

- $COV_W(day)$ mirrors the similarity in COV(pop)

- $COV_B(TOD)$ ranges from .43 to .46, with a mean value of .45.

The $COV_B(day)$ exhibits behavior which is different from the other parameters in that its values are very close in every year except BY1985. Excluding this year but including BY1984, the values range from .10 to .12, with an average value of .11. The reasons for this different behavior are: (1) the major component of the population variation is the variation between departures within any day, and (2) the major components of the variation between days within the year are the systematic factors of seasonality and growth. Thus, although BY1984 exhibits higher variation between departures within each day, higher variation between departures within each stratum, and higher variation between average boardings per stratum, the scheduling of departures per day during this year better matched the systematic trends over the year, and the $COV_B(day)$ is lower than in BY1985, where no schedule changes were made to match the systematic trends.

Given this parameter stability over the years, it is interesting to compare them with parameters obtained from other systems. Note that because some of these other parameters are aggregated at the route level, and others are based on small samples, any comparisons made can only be at a rudimentary level. However, due to the special nature of the service

71

studied in the research, such comparisons are useful checks for any outlier parameter values.

The most aggregate measure used to compare the variation in average boardings per departure is the population coefficient of variation, COV(pop). This measure pools the variation due to time-of-day, seasonal changes in weather, service growth, and other factors into one measure for a route. The largest value for this parameter I found published was 1.24 for a circulator route in Denver (Smith, 1984), and the smallest was .30 for a route in Albany, New York (Smith, 1984).

The comparison of this parameter is more meaningful if peer routes are compared. If similar routes are compared across different systems, stronger conclusions can be drawn as to the stability of the patterns in variation across systems. At least three characteristics have been used to define peer routes: (1) the average vehicle utilization, expressed as either average peak load, or average boardings per departure, (2) the time-of-day during which the route operates, (3) the route type, e.g. express, local, or circulator.

Smith (1984) stratifies data from the Denver Section 15 report by route type, and computes the COV(pop) for each route type. These parameter are presented below in Table 4.6. Note that by stratifying by route type, he is implicitly stratifying by vehicle utilization, with the smaller local and circulator routes having the lowest utilization, the inter-city and larger local routes in the middle, and the express service having the highest utilization. Looking at this table, we see a high negative correlation between vehicle utilization and the COV(pop): the higher the utilization, the lower the COV(pop). The correlation coefficient between these two parameters is -.92 for this Denver data.

Route Type	avg boardings	COV(pop)	sample size
Local (Longmont)	4.2	1.07	12.
Circulator (Denver)	11.4	1.24	34
Local (Boulder)	21.1	.83	15
Inter-city	28.3	.88	16
Local (Denver)	40.9	.63	260
Express	45.6	.40	37
All Routes	35.8	.723	381

Table 4.6 Parameter Values for Denver Routes, by Route Type.

Ferguson (1987), in an analysis of variance of passenger trips per bus by bus-line for 35 different bus routes in Orange County, California found an average coefficient of variation of .45, with two-thirds falling between .32 and .60. These estimates are computed from very small samples: an average of 5.7 observations per route over a year. The correlation between mean boardings per route and the COV(pop) for these routes is -0.65 for the twelve routes which had seven or more observations, and -0.98 for the four routes which had ten or more observations. Although this latter covariance is similar to that from the Denver data, the COVs did not vary as widely for the Orange County data set. For routes which had less than an average of 36 boardings per trip, the COV(pop) was .49. For those routes which had an average of 36 or more boardings per trip, the COV(pop) was .40.

Given that the MBTA Commuter Boat is an express service with a capacity utilization rate of 62%, its COV(pop) value of 0.50 is somewhat higher than either Denver or Orange County. The express route in Denver had a value of .40, and the highly utilized routes in Orange County had an average value of .40. A likely explanation for this variation is higher levels of variation between seasons and between time-of-day periods. Another interpretation is that the levels of variation are the same as in other systems, but that the boat schedule does not adjust to these changing levels of demand as well as the bus

schedules do. Some such inflexibility is built into the service, as the capacity of a boat is roughly three times that of a bus, and thus the supply of transportation cannot be adjusted as finely using boats.

The other common dimension along which routes are compared is the time-of-day they operate in. This comparison is often between subdivisions of a route, as many routes operate all day. The Transit Data Collection Manual (Wilson, 1986) provides default coefficients of variation based on a two-way classification of routes by TOD and average peak load. These default parameters were culled from extensive data sets from San Francisco MUNI, Chicago CTA and RTA, and Pittsburgh PAT. The groups for which significant differences were found, and the default parameters for these groups are presented below in Table 4.7. The corresponding parameters from the MBTA commuter boat appear parenthetically in bold. As the boat capacity is 150 passengers and a bus' capacity is 60, the two systems are compared in terms of the average vehicle utilization, i.e. average passengers / maximum capacity.

Comparing the parameters from the TDCM to those from the Commuter Boat, we see that the boat has slightly more variation in the peak period, much more variation in the off-peak period, and much less variation in the evening period. The similarity of peak period COVs suggests there is stability between systems for peak period routes. The larger variability in the off-peak period suggests that much of the additional variation in the Commuter Boat discussed above is due to the off-peak period. Finally, the smaller variation in the evening stratum is due to the short duration of evening service for the Commuter Boat, which ends at 8:00pm. Most other evening periods run until at least midnight, and thus contain much more variation.

TOD	All Peak Loads		
	< 35	≥ 35	
		35 - 55	> 55
Peak	.42	.35 (.40)	
Off-Peak	.45	.40 (.62)	.35
Evening	.73 (.36)		
Owl	.80		
Sat., 7 a.m. - 6 p.m.	.45		
Sat., 6 p.m. - 1 a.m.	.73		
Sun., 7 a.m. - 1a.m.	.73		

Table 4.7 Route Level Population Parameters from TDCM and Commuter Boat (in bold).

Systems can also be compared in terms of the variation between TOD groups. This measure indicates how large the differences are between the average boardings per departure for each stratum. Values for this parameter are presented below in Table 4.8. Although you might expect that the value of the parameter for the Commuter Boat would be higher, given the larger vehicle capacity, this expectation is not born out by the values. It is difficult to draw any conclusions from these data other than to identify the range for the values of this parameter.

Transit System	$COV_B(TOD)$
MBTA Commuter Boat	.38
Orange County, CA, system-wide	.32
Denver, CO, system-wide	.24
Ottawa, system-wide	.44
Ottawa, route 95, two-week period	.33

Table 4.8 Coefficient of Variation Between TOD Croups for Five Transit Systems

One conclusion which can be drawn from this analysis is that there is a strong correlation between the vehicle utilization of a route and the coefficient of variation in boardings per departure for the route. The relationship is that as the vehicle utilization

increases, the coefficient of variation decreases. This relationship was already demonstrated for the Commuter Boat in Section 4.2.2, when it was shown how stratifying explicitly by average vehicle utilization instead of TOD periods improved the efficiency of the proportional stratified sampling plan. If a sampling plan is to involve stratification, then the first choice for stratification should be average vehicle utilization. If this is not possible, then the routes should be stratified by some characteristic which is thought to be closely correlated with vehicle utilization, such as route type or TOD.

A second result of this analysis is that routes that have high levels of vehicle utilization have very similar coefficient of variation values across systems. San Francisco, Chicago, Pittsburgh, Denver, Orange County, and the Commuter Boat all had COV(pop) values very close to .40 for high utilization trips. Given the wide range of environmental conditions surrounding the services that produced these parameters, this result is remarkable.

One final remark on the stability of the COV(pop) parameter. To a large extent, this parameter reflects how much variation occurs within a day. Referring to Table 4.5, we see that the Commuter Boat COV(pop) remained constant between years which had large differences in the $COV_B(day)$, e.g. BY1985 to BY1987. While the average daily boardings per departure varied more in BY1985 because the schedule did not adapt to seasonal and growth factors, the COV(pop) remained stable. Notice also that the average variation between departures on a day also remained stable. This implies that the variation between departures within the day dominates the variation due to seasonal and growth factors in determining the COV(pop). This characteristic is investigated from a different perspective in the following section.

4.3.1 Simple Random Theory, Cluster Cost

The fact that the within day variation is so predominant in determining the population coefficient of variation suggests a very simple model for estimating the between

day variance. Consider a situation where the only information available is an estimate of the COV(pop), the pooled variation of all trips over the entire year. If an estimate of the inter-day COV is needed, a very crude model would be to consider every trip that departs over the year to be independently and identically distributed (iid) with covariance equal to COV(pop). Using this model,

$$COV_B(day) = COV(pop) / \sqrt{trips / day} \, .$$

Considering how far from reality this model is, with all the variation in boardings per trip between different TOD strata and different months, it is startling to find that for three of the five years of the boat service, the estimate from this crude model is very close to the true value. So close that at a 95% confidence level, the hypothesis that the COV computed from this model is equal to the true value cannot be rejected. Table 4.9 below presents the true values, the estimates from this model, and the summary statistics of the hypothesis test.

Year	True Value	Estimate	95% Test	p-value
1984	.12	.17	reject	.0000
1985	.18	.14	reject	.0000
1986	.10	.10	cannot reject	.7872
1987	.10	.11	cannot reject	.3174
1988	.11	.10	cannot reject	.3174

Table 4.9 Summary Statistics for Test of iid Hypothesis

Similar results have appeared in other research. Hsu (1983) investigates a data set from Pittsburgh which consists of two six week periods, one week per month, of full ride

checks for nine routes and five time periods. This same hypothesis was tested, and in eight of the 10 time periods, the hypothesis could not be rejected at the 95% confidence level. Furth (1988) found that for weekday Line / Direction / Time Period strata, selecting day-clusters did no worse than simple random sampling.

The condition for when this simple model fits can be determined by considering the theoretical relationship between the population variance, the between-day variance, and the within-day variance. Let

$$N = \text{number of clusters (e.g. days) in population}$$

$$M = \text{number of elements (e.g. trips) per cluster}$$

$$\sigma^2_{Wi} = \text{variance between elements in cluster } i$$

$$\sigma^2_W = \text{avg. within-day variance} = \frac{1}{N-1} \sum_{i=1}^{N} \sigma^2_{Wi}$$

$$\overline{x}_i = \text{avg. for cluster } i$$

$$\overline{\overline{x}} = \text{population average per element}$$

$$\sigma^2_C = \text{variance between cluster means} = \frac{1}{N-1} \sum_{i=1}^{N} \left(\overline{x}_i - \overline{\overline{x}} \right)^2$$

$$\sigma^2_B = \text{variance between elements based on cluster means} = M \, \sigma^2_C$$

$$\sigma^2 = \text{population variance between elements}$$

Then, we have

$$\sigma^2 = \frac{(N-1)\sigma^2_B + N(M-1)\sigma^2_W}{NM-1}$$

Assuming $N > 50$, we can approximate this equality with the expression

$$\sigma^2 = \frac{\sigma^2_B + (M-1)\sigma^2_W}{M} \quad .$$

(Cochran, 1977). Replacing σ^2_B with $M\sigma^2_C$ and σ^2_C with the value computed from the simple model, i.e. σ^2 / M, we get

$$\sigma^2 = \sigma^2_W \qquad \text{and} \qquad \frac{\sigma^2_W}{\sigma^2_B} = M$$

Thus, for any population in which the ratio of the within cluster coefficient of variation to the between cluster coefficient of variation is greater than or equal to the square root of the number of elements in a cluster, then obtaining a simple random sample of clusters will produce an estimate that is at least as precise as that obtained from a simple random sample of M departures. For such populations, a sampling plan can be constructed that has the best of all possible worlds: the simplest sampling theory (simple random sampling) can be used to compute the required sample size, and the lowest cost sampling plan (cluster sampling) can be used to obtain the sample.

This model was tested considering each TOD stratum as the population, and using all trips in a given stratum which depart on the same day as the cluster. The true between day variance as well as the ratio of the variance from the model to the true variance is presented in Table 4.10 for each TOD group for each year. This analysis produced different results for each of the strata.

Stratum	BY1984	BY1985	BY1986	BY1987	BY1988	Average
a.m. pk	.113 / 1.91	.110 / 1.61	.083 / 1.57	.070 / 1.57	.090 / 2.00	.093 / 1.40
midday	.616 / .73	.669 / .63	.487 / .63	.395 / .71	.500 / .72	.533 / .62
p.m. pk	.194 / 1.17	.156 / 1.03	.125 / 1.03	.094 / 1.22	.129 / 1.52	.140 / 1.11
evening	.481 / 1.02	.286 / .98	.302 / .98	.228 / 1.10	.222 / 1.20	.304 / 1.04

Table 4.10 Between Day COVs and Ratio to Model Estimate for TOD Clusters.

For the a.m. peak stratum, the model overestimated the between day variance. This reflects an even greater dominance of the within-day variation over the between-day variation for the a.m. peak stratum. This result suggests that there is very little day-to-day variation in the a.m. peak, due to the large number of work trips during this period. The midday stratum shows more between day variability, as the model underestimates the

between-day variation. This is due to the dominance of irregular non-work trips during the midday. In the p.m. peak the model slightly overestimates the between-day variation because of the mix of work and non-work trips served during this period. Finally, the model fits the evening stratum well, a result due largely to the small number of daily departures (2 or less) during this period.

This analysis shows that the iid model works well for peak period time-of-day clusters as well as day clusters. The sample size required by simple random sampling will produce a more precise estimate if collected by cluster sampling. This is not true for the off-peak periods, where the cluster estimator will be less precise than that from simple random sampling. In this case, a cluster is not an entire day of trips, but all trips during a time-of-day group (e.g. a.m. peak) that depart on the same day.

4.4. N-Efficiency

Given that the entire population is known in this study, we can compute the theoretical efficiency of each estimator which is a function of the variance of the estimator. Using the formulas presented in Chapter 3, and replacing the sample estimates of population, within-stratum, and between-cluster coefficients of variation with the true population values, we can compute the variance of each estimator for each sample size.

To compute the theoretical efficiency of simple random sampling, we need the number of trips which departed over the year, the average boardings per trip for the year, and the standard deviation of boardings per trip. For stratified sampling, we need these three measures for each stratum. For cluster sampling, we need the the number of days in the year, the total number of departures during the year, average total boardings per day, and the standard deviation of the total boardings per day. The formulas for the theoretical variance of each estimator are repeated below.

Simple Random Sampling

$$\sigma(\overline{y}_R) = \sqrt{\frac{N - n}{N} \frac{1}{n}} \, \sigma(y)$$

Stratified Sampling

$$\sigma(\overline{y}_S) = \frac{1}{N} \sqrt{\sum_{h=1}^{4} N_h^2 \frac{N_h - n_h}{N_h} \frac{\sigma^2(y_h)}{n_h}}$$

Cluster Sampling

$$\sigma(\overline{y}_C) = \sqrt{\frac{CL - cl}{CL} \frac{1}{cl}} \, \sigma_C$$

4.4.1. A Sample Calculation

As an example of calculating the theoretical efficiencies of these three plans, consider the data for 1984, using a sample size of 120 departures. In this case, the standard deviation of the simple random sampling estimator will be

$$\sigma(\overline{y}_R) = \sqrt{\frac{2912 - 120}{2912} \frac{1}{120}} \; 39.84 \; = \; 3.56$$

Using a Student's t-table, we find that $t_{119}^{.025}$, the t-value for a sample size of 120 at the 95% confidence level, is equal to 1.98, and thus the maximum error is 1.98 * 3.56 = 7.05 boardings per departure. As the population average boardings per departure is 67.18, a sample size of 120 trips will give a precision of 7.05 / 67.18 = 10.5% at the 95% confidence level.

To compute the standard deviation of the proportional stratified sampling estimator, first compute the strata sample sizes. Each stratum sample size is proportional to the total number of trips it contains relative to the total number of trips in the population. Thus the sample size for the a.m. peak period is (1045 / 2912) * 120 = 40. Similarly, the sample sizes for the mid day, p.m. peak, and evening strata are 29, 40, and

11 respectively. Using these sample sizes and the appropriate population parameters, the standard deviation is

$$\sigma(\bar{y}_s) = \frac{1}{2912} \left[\begin{array}{l} \left(971^2 \ \frac{971 - 40}{971} \ \frac{1}{40} \ 35.76^2\right) \\ + \left(711^2 \ \frac{711 - 29}{711} \ \frac{1}{29} \ 28.91^2\right) \\ + \left(979^2 \ \frac{979 - 40}{979} \ \frac{1}{40} \ 36.67^2\right) \\ + \left(251^2 \ \frac{251 - 11}{251} \ \frac{1}{11} \ 14.64^2\right) \end{array} \right]^{1/2}$$

$$= 2.97,$$

with a maximum error of 5.89 boardings per departures which corresponds to a precision of 8.8%.

To compute the standard deviation of the cluster sampling estimator, we need to compute the number of days required to give at least a sample size of 120 departures. As the average number of departures per day is 11.7, a sample of 10 days should give a number of departures very close to 120. Using this sample size, the standard deviation of the cluster sampling estimator is

$$\sigma(\bar{y}_c) = \sqrt{\frac{248 - 10}{248} \ \frac{1}{10} \ 8.43} = 2.61 .$$

with a maximum error of 5.90 boardings per departure and a precision of 8.8%.

4.4.2. Empirical Analysis

The experiments conducted in this study were run to test how well each sampling plan performed empirically. The empirical performance of a plan can be tested by repeatedly drawing random samples from the population. The empirical precision is computed as follows:

1. For random sample i, compute the estimate y_i.

2. Take the difference between this estimate for the population mean, and the true

population mean, $\overline{Y}$.

3. Label this quantity d_i.

4. Find the value D, such that

$$I\{\overline{Y} \in [y_i - D, y_i - D]\} = 0.95 \times \text{ITMAX} ,$$

where ITMAX is the total number of samples generated, $I\{\text{true}\} = 1$ and $I\{\text{false}\} = 0$.
If one of the ITMAX estimates is randomly selected, say y_j, then

$$P\{y_j - D \leq \overline{Y} \leq y_j + D\} = 0.95 .$$

This quantity D defines the empirical 95% confidence interval, which can then be compared
with the theoretical quantity $t_{N-1}^{.025}\sigma(\overline{Y})$.

As these experiments aim at determining the difference between the theoretical

performance, and the empirical performance, one question that is important to answer is:

what is the effect of generating a limited number of samples since it is impossible to

generate all the possible samples. For example, the total number of ways a sample of 100

can be selected from a population of 2000 is more than 1×10^{170}. If all the possible

samples are not generated, this introduces error in the empirical precision computation.

For simple random sampling, it is possible to calculate the error due to limited

iterations. Define the sample size, in departures, to be N. ITMAX is the number of times a

sample of size N is generated. Let the true population variance of the boardings per

departure be σ^2. Then, the "true" (assuming normality of the mean estimator) standard

deviation of the mean estimator for simple random sampling is

$$\frac{1}{\sqrt{N}} \, \sigma$$

Now, say that we want the error due to limited iterations to be no greater that 5%. Let s^2 be the sample variance of the ITMAX means. From statistical theory, we know that the quantity

$$\frac{\text{ITMAX} - 1}{\sigma^2 \, / \, N} \, s^2 \sim \chi^2$$

(is distributed as chi-squared) (Kvanli, et al. 1986). We also know that when the sample size ITMAX is larger than 30, the quantity

$$\sqrt{2\chi^2} \sim N\left(\sqrt{2(N-1)} - 1 \,,\, 1\right) \text{ (Kendall 1961)}.$$

As ITMAX will definitely be larger than 30 in this situation, we will consider only this case. Thus, we have

$$P\left\{ \sqrt{2(\text{ITMAX}-1)-1} - 1.96 \;\leq\; \sqrt{2\chi^2} \;\leq\; \sqrt{2(\text{ITMAX}-1)-1} + 1.96 \right\} \;=\; 0.95$$

.

Using these two relationships, after some algebra we can state that

$$P\left\{ \begin{array}{l} \frac{1}{2} \dfrac{\sigma^2}{N(\text{ITMAX}-1)} \left(\sqrt{2(\text{ITMAX}-1)-1} - 1.96\right)^2 \;\leq\; s^2 \\[2mm] \leq \frac{1}{2} \dfrac{\sigma^2}{N(\text{ITMAX}-1)} \left(\sqrt{2(\text{ITMAX}-1)-1} + 1.96\right)^2 \end{array} \right\} \;=\; 0.95$$

If many samples of ITMAX estimates are generated, 95% of the many sample variances should be within this interval. Thus, at a confidence level of 95%, the error due to limited iterations will be no more that 5% if

$$\frac{\left| \frac{1}{\sqrt{2}} \dfrac{\sigma}{\sqrt{N(\text{ITMAX}-1)}} \left(\sqrt{2(\text{ITMAX}-1)-1} - 1.96\right) - \frac{1}{\sqrt{N}} \, \sigma \right|}{\frac{1}{\sqrt{N}} \, \sigma} \;=\; 0.05$$

.

Using the approximation

$$\sqrt{2(\text{ITMAX}-1)-1} \approx \sqrt{2(\text{ITMAX}-1)}$$

this gives a value of 770 for ITMAX, independent of the sample size N.

This number of iterations was also used for the proportional sampling. For simple random cluster sampling 5,000 iterations were executed for each sample size. For systematic cluster sampling, all possible samples were generated.

4.4.3 Theory vs. Empirical

For both random sampling and proportional stratified sampling, the empirical precision was very close to the theoretical precision. The only difference was that the empirical precision was slightly higher than that predicted by the theory for sample sizes smaller than 40. This happened because the actual distribution of the estimate approached normality faster that the corresponding Student's t-distribution.

4.4.4 A Comparison of Sampling Plans

Figures 4.5 - 4.9 show the efficiency of the sampling plans for each year. As discussed in the analysis of the population parameters in Section 4.3, proportional stratified sampling of departures does better than simple random sampling of departures in every year. Also expected is the similarity in efficiency between simple random sampling of clusters and simple random sampling of departures for BY1986, BY1987, and BY1988. In BY1984, where the ratio of $COV_W(\text{day}) / COV_B(\text{day})$ is greater than the square root of the number of trips per day, cluster sampling does better than simple random sampling. In BY1985, where this ratio is less than square root of the number of trips per day, cluster sampling does worse.

It is apparent from these figures that a systematic selection of clusters improves the efficiency of cluster sampling in four out of the five years. Referring to Figure 4.2, BY1985 has the largest and most regular seasonal pattern, and thus it is no surprise that the improvement due to a systematic selection of clusters is the greatest of all the years. Note that in the years in which a simple random selection of clusters is least efficient, a systematic selection of clusters shows the greatest improvement. This is because the years in which simple random selection of clusters does poorly have large between day variation, which is mainly due to seasonal and growth factors. Thus a systematic selection of clusters which are spread out evenly over the year provides a more representative sample of the entire year.

The efficiency of each plan relative to simple random sampling is summarized in Table 4.11 below. The sample sizes are those required by each plan to produce an estimate of average annual boardings per departure that is within ±10% of the true population value at a 95% confidence level.

Year	N_R	N_S	N_C	N_{SC}	N_S / N_R	N_{SC}/N_R
1984	132	95	84(7)	60(5)	.72	.45
1985	102	70	168(12)	84(6)	.69	.82
1986	70	60	68(4)	68(4)	.86	.97
1987	95	67	80(4)	60(3)	.71	.63
1988	99	73	120(5)	72(3)	.74	.73

Table 4.11 Sample Sizes Required to Reach ±10% Precision at 95% Confidence Level

Reviewing the figures in this table, it is probable that systematic cluster sampling will be the most cost-efficient method since it generally has the same n-efficiency as proportional stratified sampling. In fact, the average performance of systematic cluster sampling is better than proportional stratified sampling over the five years; the average

systematic cluster sampling improvement in efficiency over simple random sampling is 28% compared with 25.6% for proportional stratified sampling.
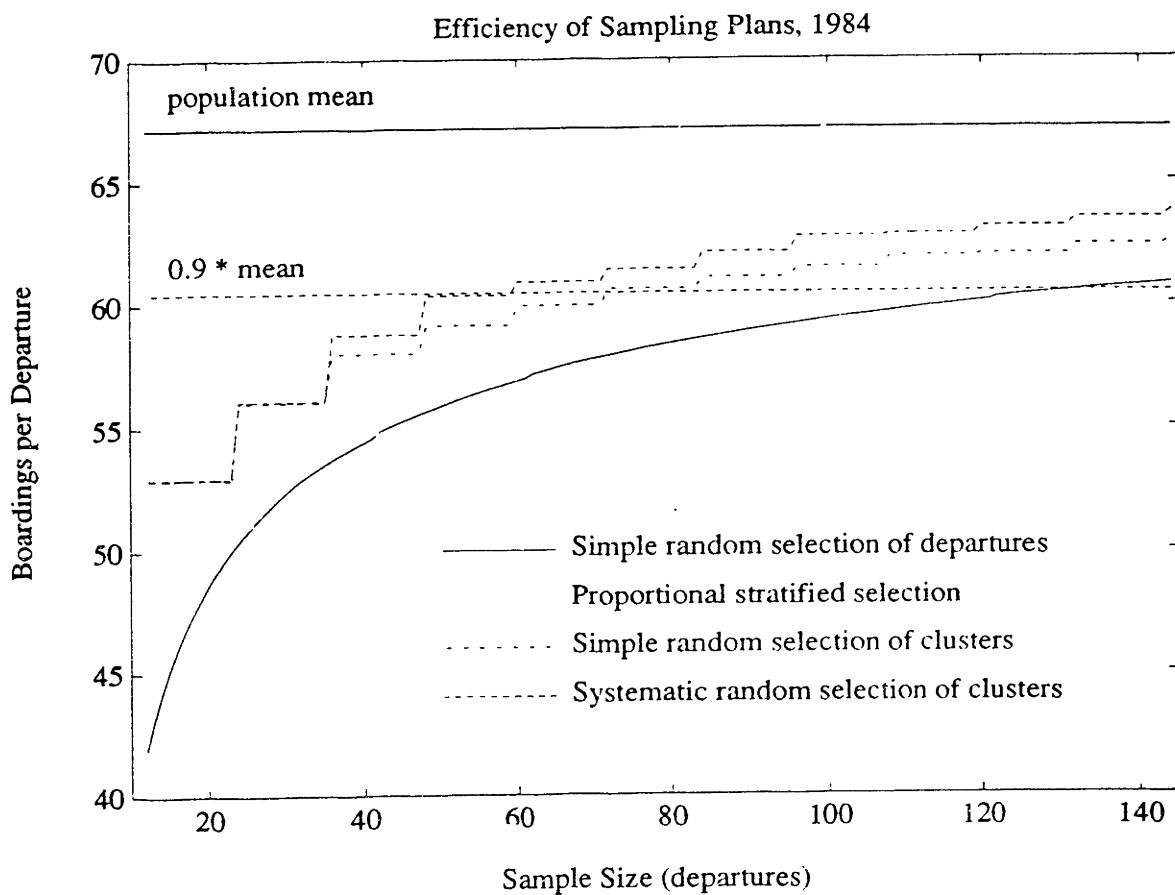
Efficiency of Sampling Plans, 1984

Figure 4.5 Relative Efficiency of Plans, 1984

Efficiency of Sampling Plans, 1985

population mean

0.9 * mean

—————— Simple random selection of departures

. Proportional stratified selection

- - - - - - Simple random selection of clusters

- - - - - - - - Systematic random selection of clusters
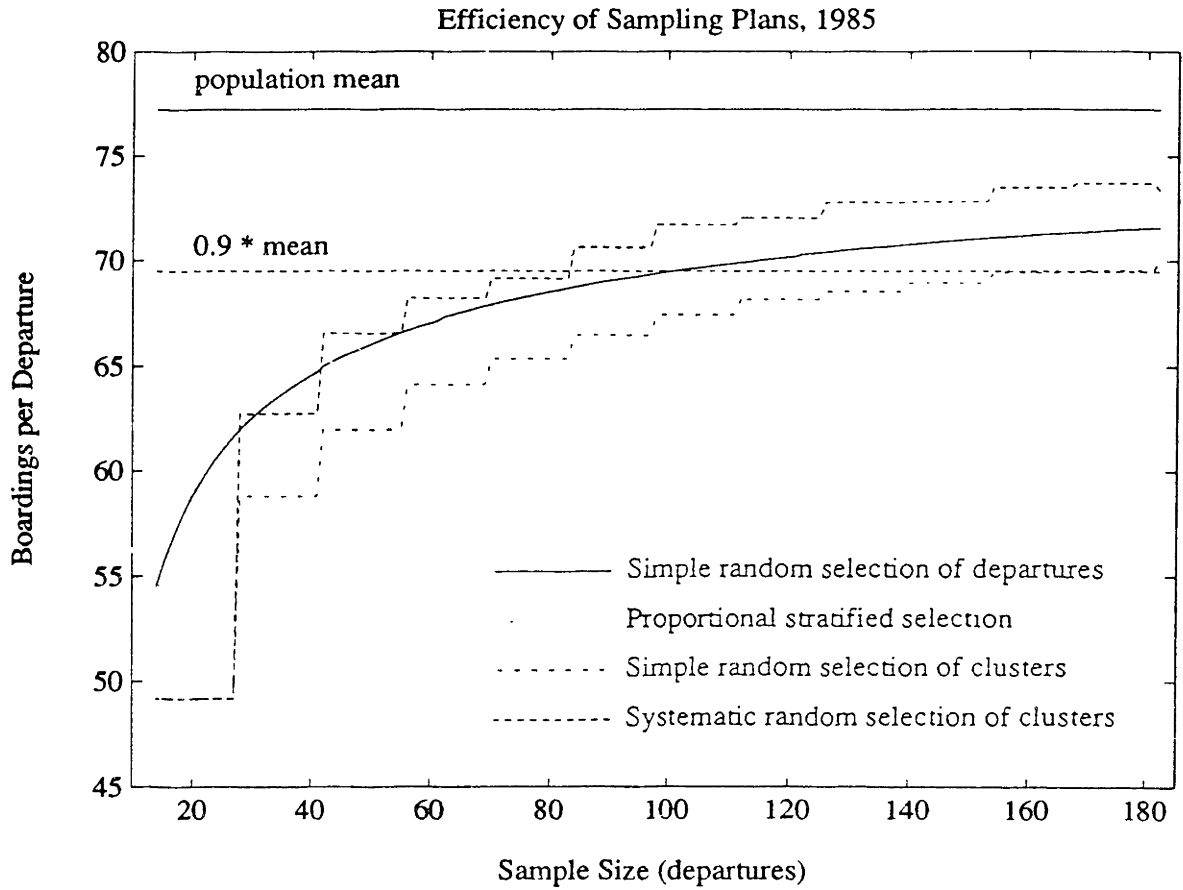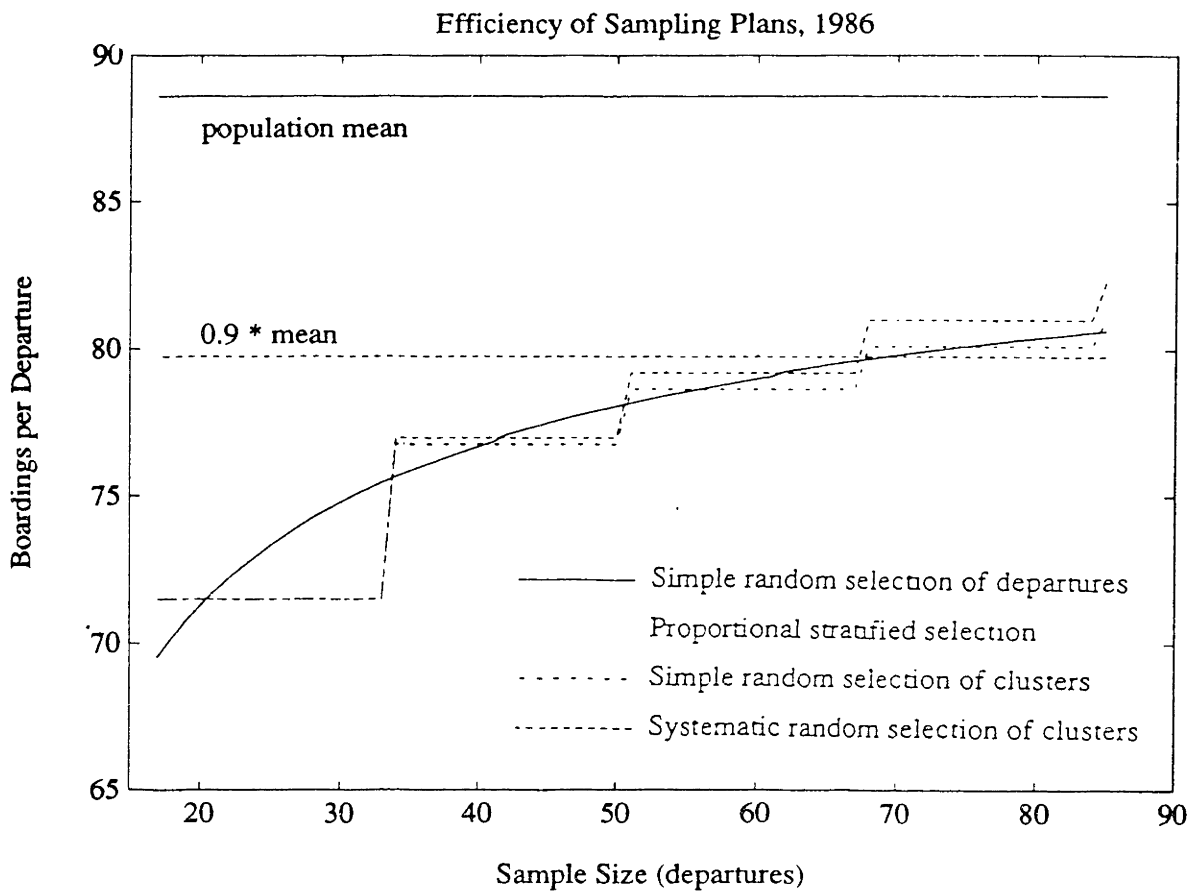
Boardings per Departure

Sample Size (departures)

Figure 4.6 Relative Efficiency of Plans, 1985

Figure 4.7 Relative Efficiency of Plans, 1986

Figure 4.8 Relative Efficiency of Plans, 1987

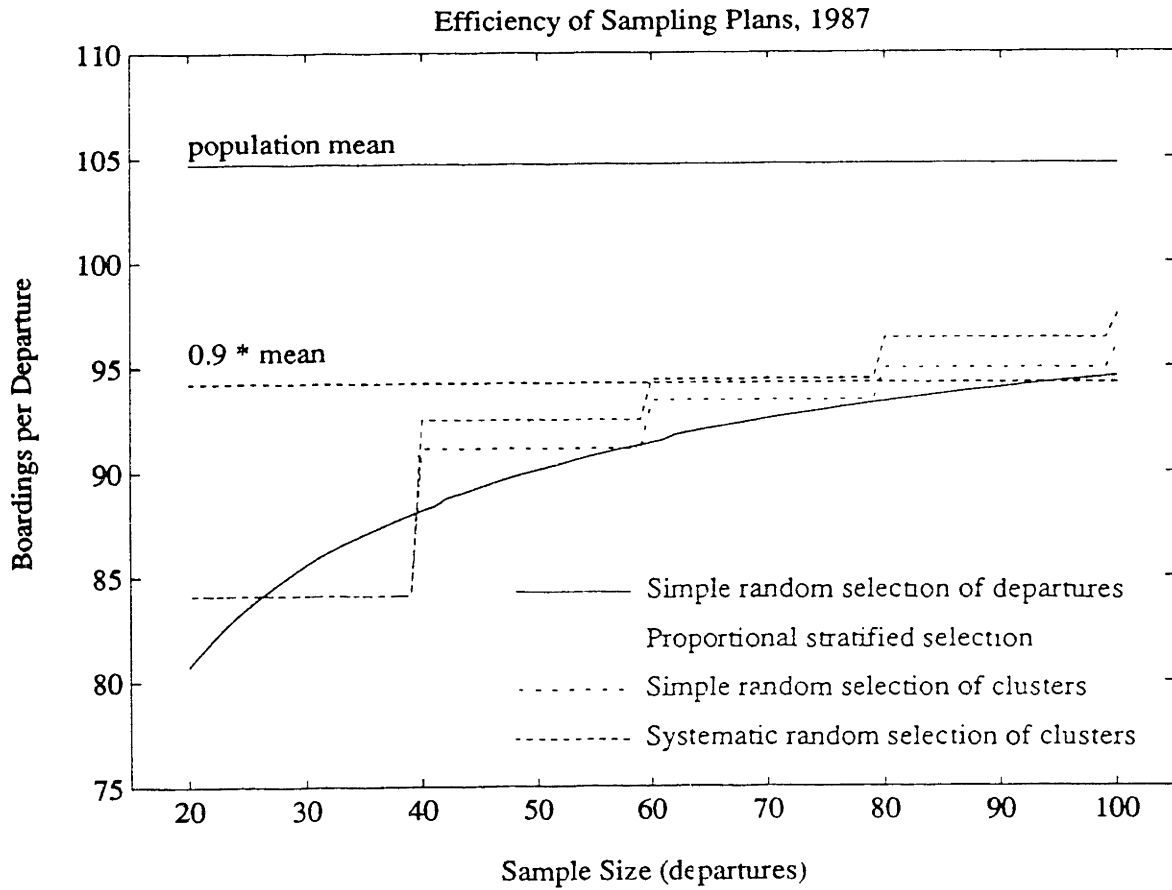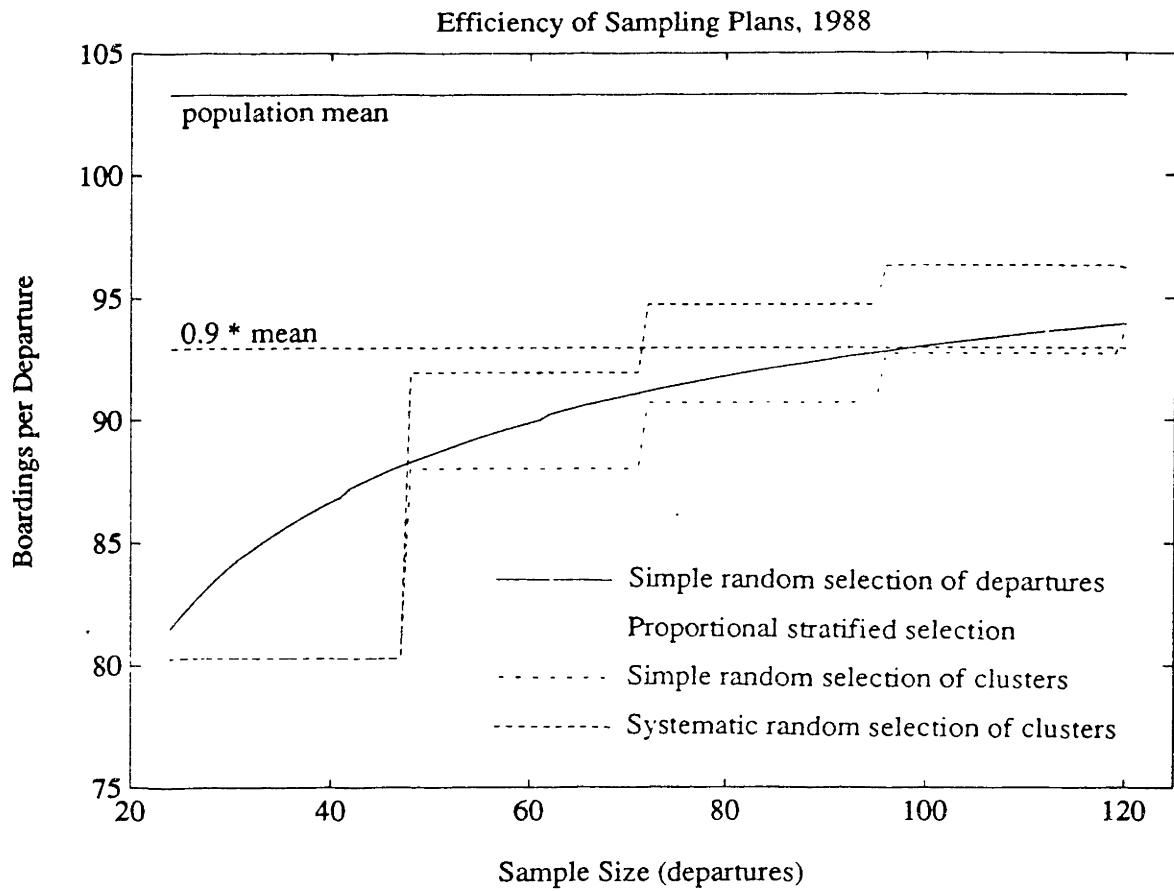## Efficiency of Sampling Plans, 1988

Figure 4.9 Relative Efficiency of Plans, 1988

4.5 Cost Efficiency

Comparing the efficiency of cluster sampling to the efficiency of simple random sampling or proportional stratified sampling on the basis of sample size alone is not a complete comparison. The key advantage of cluster sampling is that the sample is collected in clusters, and thus can be significantly less expensive to schedule and to collect. This advantage is not reflected in a comparison based on sample size alone. To obtain a proportional stratified sample, the entire population must be partitioned into the appropriate strata, and then a random sample must be drawn from each strata independently. This results in observations that are scattered randomly over time. This is expensive to schedule and collect, especially for large transit agencies.

In an attempt to make a more complete comparison of the sampling methods used here, a cost function is defined which breaks the cost of sampling into two components. The first component is the overhead cost associated with getting a collector to the location where the measurement is to be made. The second component is the cost of actually making the observation, once at the proper place and time.

The two plans compared using this cost function are proportional stratified sampling and cluster sampling with pre-stratification of clusters. Each was compared assuming that an estimate of average annual boardings per departure was required for the route at a precision of ±10% at the 95% confidence level.

4.5.1 Cost Function

More formally, let

h = mean headway of service (hours),

L = one-way running time of route (hours),

OV = overhead cost (hours),

F = field cost of collecting data ($ / hour),

H = daily service duration (hours),

92

W = working hours per day for one checker,

T = number of trips offered in a day,

C_S = the cost of collecting a proportional stratified sample of N_S departures,

C_C = the cost of collecting a random sample of N_C departures using cluster sampling.

Point Checks

Consider the situation where the measurement can be made through a point check. If a point check can be made in t_1 hours, and ignoring the low-probability event that two (or more) randomly selected departures happen to fall sequentially, the cost of collecting a random sample of N departures is equal to the product of the sum of the overhead cost of sending a person to the count location plus the time required to make the count, the cost per hour of a checker, and the number of departures, or

$$C_S = (OV + t_1) * F * N_S.$$

The cost of collecting a clustered sample is different from the cost for a proportional stratified sample in two ways. First, every time a checker is assigned, she is assigned for a full W or H hours, whichever is smaller. Each checker observes as many trips as she can in one day of work. Thus, the number of checkers required for a day of service is equal to the total service duration divided by the difference between the working hours available with one checker and the time required to get the checker to and from the observation location. Thus,

$$\text{\# checkers needed per day} = \lceil H / (W - OV) \rceil.$$

(NOTE: The operator $\lceil x \rceil$ returns the smallest integer that is greater than x.)

The last term required to compute the cost of data collection using a clustered sampling plan is the number of days required. As each day has T trips, and N_C trips are required in the sample, the

$$\text{\# of days required in sample} = \lceil N_C / T \rceil.$$

Putting this all together, we get that the cost of collecting a clustered sample of N_r trips is

$$C_C = \lceil N_C / T \rceil * \lceil H / (W - OV) \rceil * W * F$$

Ride Checks

We can construct a similar cost function for routes for which point checks are not sufficient and ride checks must be used. Here, the time to make one observation, t_1, is equal to the one-way travel time for the route, L. In this case, the expression for the cost of a proportional stratified sample of departures is

$$C_S = (OV + L) * F * N_S$$

The integer constraint for the number of days can be relaxed for cluster ride check sampling. Nothing in this data analysis indicates that a systematic selection of trips from TOD groups over different days of a week will reduce the efficiency of the sampling plan. This formulation is also more realistic; for example, it is unlikely that a transit agency will assign 20 checkers to cover an entire day when they can assign 4 to cover 5 consecutive weekdays. The number of checker-days needed is simple the number of trips to be sampled divided by the number of trips observed per checker-day. Assuming that every observation requires the equivalent of two one-way travel times, as if the operator was interested in measuring one direction for a particular route, and was not interested in the off-peak-direction trips, the cost of collecting a simple random sample of clusters is

$$C_C = \lceil N_C / \lfloor (W - OV) / 2L \rfloor \rceil * W * F$$

where $\lfloor \rfloor$ returns the largest integer smaller than its argument.

4.5.2 A Cost Comparison for Various Route Types

In this section the cost of data collection for proportional stratified sampling and cluster sampling with pre-stratification of clusters by season was compared for a variety of route types. Eight route types were considered, each having parameters COV(pop), h, and H, at one of two possible values. The parameter for departure headway, h, was either 0.17 (10 minute headways) for high frequency routes, or 1.00 (1 hour headways) for low frequency routes. The parameter for the daily service duration, H, was either 18 hours for all day service, or 6 hours for peak or express service. The parameter for the population coefficient of variation, COV(pop), was either 1.00 for routes with a high degree of variation or 0.40 for routes with a low amount of variation.

Based on the results in Table 4.11, the required sample size for both proportional stratified sampling and systematic cluster sampling were assumed to be 75% of that required for a simple random sampling.

All of the routes are compared using checkers that work eight hours a day, and all routes were assumed to be of equal length, i.e. 30 minutes one-way running time. The time to make a point check, t1, was set equal to 15 minutes. In reality, the value for these last two parameters will be subject to variability as trip durations vary depending on congestion levels, whether or not the vehicle breaks down, gets in an accident, or otherwise gets delayed. In addition, vehicles may depart or arrive off schedule, and checkers will also not always make it to the correct place by the scheduled time. The effect of all this uncertainty is that given the required sample size, in order to assure that enough departures are observed, the route operator should overschedule observations by a small percentage.

Table 4.12 below presents the results of this cost comparison for both point checks and ride checks in terms of a break-even overhead cost. This is the value of overhead time per observation at which the systematic selection of clusters is as cost efficient as the proportional stratified selection of departures. If the actual overhead cost for an

95

observation is greater than this break-even value, then systematic cluster sampling is more cost efficient.

route	COV	h	H	point checks	ride checks
1	1.00	.17	18	0	.65
2	1.00	.17	6	0	.65
3	1.00	1.0	18	1.11	.65
4	1.00	1.0	6	1.08	.65
5	.40	.17	18	.25	.67
6	.40	.17	6	.08	.67
7	.40	1.0	18	1.25	.67
8	.40	1.0	6	1.08	.67

Table 4.12 Break-Even Overhead Costs for Cluster Sampling

The obvious conclusions from this table are that cluster sampling will always be preferable for point checks on high frequency routes, and is likely to be strongly competitive for all ride check data collection. For the point check cost comparison, the route parameter which dominates the break-even overhead cost is the mean headway. Cluster sampling does much better for the routes which have higher service frequencies. This makes sense, because the number of checkers required to provide a day's worth of departures is independent of the mean headway. Thus, the more departures that occur during a day, the better the cluster sampling method will perform.

A secondary factor in the difference between break-even overhead costs for the point checks is the "lumpiness" of the cluster sampling sample sizes. The cluster sample size can only be increased one day at a time, and thus is lumpy as opposed to the stratified sample size which can be increased one departure at a time. Usually, the cluster sample

contains more departures than are required by the precision specification. For example, route 5 has a COV of 0.40 which requires a sample size of 64 departures to meet ±10% precision at 95% confidence level. However, this route has a high frequency of service and operates 18 hours a day, and thus one day contains 105 departures. The smallest sample size possible using cluster sampling is 105, which produces an estimate that is within ±6.8% of the true population mean at the 95% confidence level.

The break-even overhead costs for the ride checks are all close to 40 minutes. This is a low overhead time for a real-world route: 20 minutes to get to the ride check location, and 20 minutes to return from the end of the ride check. Considering that every cluster sampling observation was penalized 30 minutes more than a proportional stratified observation, this is strong evidence that cluster sampling will be more cost efficient for most types of routes.

Consider a more general cost comparison which includes factors that are difficult to quantify such as time required to select and schedule each random observation. In this more general context, cluster sampling is clearly more efficient. To generate a proportional stratified sample, all departures during the year must be put into one of the four strata. Then the total number of departures in each group must be counted, and the proportional sample sizes computed. Then a random sample must be drawn from each of the strata. For a route with many departures and schedule changes over the year, this procedure can be tedious.

The cluster sampling procedure is clearly much easier, at least for point checks. Simply count the number of regular-service days in the population, and compute the average number of departures per day. Then, determine how many days are required in the sample by dividing the sample size required by the average number of departures per day, and rounding up to the next integer. Divide the year into this many groups of contiguous days, and randomly select one day from each group. The scheduling of data collectors is

97

also easier, as all trips during the day are to be observed, as opposed to departures randomly scattered across days within each TOD group.

The critical assumption underlying this cost comparison, and the conclusion that cluster sampling is more cost efficient than proportional stratified sampling, is that systematic cluster sampling and stratified sampling will result in the same reduction in required sample size over simple random sampling. Although not explicitly stated, it is also assumed that both systematic cluster sampling and stratified sampling will be more efficient that simple random sampling. These assumptions are reasonable given the results of this investigation as well and the findings of Hsu and Furth alluded to previously.

This investigation has produced strong evidence that systematic cluster sampling of days is more cost efficient than both simple random sampling and proportional stratified sampling for most route types. The commuter boat data set had large differences in seasonal and growth factors across the five years, yet the performance of the systematic cluster estimator was very stable. Based on the population parameter values found in this data set and presented in other research, the above cost comparison suggests that cluster sampling will always be preferable for point checks on high frequency routes, and is likely to be strongly competitive for all ride check data collection. Considering the other hard-to-quantify costs associated with selecting and scheduling the random observations, systematic cluster sampling should be the best choice of sampling plan for most routes.

Chapter 5. Summary of Results and Research Extensions

5.1 Summary

There is a consensus among governmental, academic, and private players that a primary objective of the public transit industry should be to improve productivity. The past 20 years of federal subsidies have improved transit infrastructure and stabilized ridership, but these subsidies have replaced farebox revenue as the primary revenue source and transit market share has decreased. During the period from 1975 to 1985, the average national transit operating cost per passenger boarding increased 47%, from $0.97 to $1.43, (after inflation), while the revenue per passenger boarding increased by 11 cents, from $0.56 to $0.67, and governmental operating assistance per passenger boarding increased 105%, from $0.39 to $0.80. Although governmental intervention has saved this critical industry, there are still many structural inefficiencies that are causing productivity to decline despite the public support.

In 1984, the executive branch of the federal government, through UMTA, defined a new federal policy regarding public transit emphasizing that private industry be encouraged to compete for the provision of public transit services. By encouraging public transit agencies to contract out services to private operators, it is hoped that the private vs. private and private vs. public competition will cause inefficiencies to dissipate naturally

Transit Agency as Planner and Monitor

When a transit agency contracts with a private operator to provide service, its role becomes that of a service planner and monitor. The safety of the customers must be assured, service quality maintained or improved, and service adjusted to match changing patterns of demand. In addition, case studies of successful privatization efforts indicate that one key to success is the use of financial incentives and penalties for performance. When such clauses are included in a contract, the transit operator is responsible for

determining whether or not such performance goals are met. The costs of monitoring a private contractor should be included in any cost/benefit analysis of whether or not to contract out service.

The research presented in this paper investigated the cost of monitoring transit service at the route level. Five years of complete ridership data from the MBTA Commuter Boat service were used to analyze the patterns of variability in average boardings per departure over time. The population parameters describing this variability for this service were compared with those from other routes in other transit systems, and some striking similarities were found. Five sampling plans were compared based on two measures of efficiency: required sample size, and a generalized cost function. The five sampling plans investigated were: simple random sampling of departures, proportional random sampling of departures stratified by time-of-day, simple random sampling of days, systematic random sampling of days, and simple random sampling of days with post-stratification by time-of-day period.

The Stability of Parameters of Variation

The parameters describing the variation within each year of the MBTA Commuter Boat data set exhibit remarkable stability, considering the many differences between the years. The service experienced rapid growth over the five years, and had many schedule changes, with an average of 12 departures in BY1984, and an average of 24 departures in BY1988. In addition, the seasonal ridership variations due to climate and weather were as various as only New England's can be. Despite these many differences the variation between all boardings in each years was stable, as well as the variation between time-of-day groups and the between-day variation.

The close relationship between the scheduling and the patterns of variation in average boardings per departure was noted. This relationship was particularly evident in the between-day variation, which was seen to depend on how well the scheduling adjusted

to the systematic patterns of seasonality and growth. Another manifestation of this relationship is the dominance of the within-day variation over the between-day variation in contributing to the total population variation. The heavy peaking characteristic of transit service, combined with policy headways and minimum levels of service result in this predominance of within-day variation.

The parameters found in the Commuter Boat data were compared with those published in studies of other systems. There was found to be a high correlation between average vehicle utilization and the coefficient of variation, with the more heavily utilized routes having a lower coefficient of variation. The coefficient of variation for the most heavily utilized routes was found to be very close for all the studies: San Francisco, Chicago, Pittsburgh, Denver, Orange County, and the Commuter Boat all had COV values very close to .40 for high utilization trips. Given the wide range of environmental conditions surrounding the services that produced these parameters, this result is remarkable.

Simple Random Theory, Cluster Sampling Cost

A result of the predominance of the within-day variation in determining the population coefficient of variation is the applicability of a very simple model for estimating the between day variance. If all departures over a year are considered to be independently and identically distributed (iid) with coefficient of variation equal to COV, then the between-day coefficient of variation is equal to $COV / \sqrt{(\# \text{ trips per day})}$. For the Commuter Boat Data, the the hypothesis that the COV computed from this model is equal to the true value cannot be rejected at a 95% confidence level for the three most recent years of service. Thus, the simplest sampling theory (simple random sampling of departures) can be used to compute the required sample size, and the lowest cost sampling plan (cluster sampling) can then be used to obtain the sample.

It is noted that this result has appeared in other research. Hsu (1983) found that a this same hypothesis could not be rejected at the 95% confidence level for ride check data from Pittsburgh. In his research, the clusters were defined to be route/time-of-day groups, instead of the entire day of departures on a route. In addition, a 1987 analysis of data from Pittsburgh found this same result, and concluded that "cluster sampling for L/D/TP level statistics can be considered to be just as good as simple random sampling for weekday time periods" (Furth, 1987). This model was tested for the Commuter Boat time-of-day groupings, and was found to produce sample sizes which, when collected in clusters, gave estimates of greater precision for the a.m. and p.m. peak periods.

Generalized Cost Function

Comparing the efficiency of cluster sampling with the efficiency of simple random sampling or proportional stratified sampling on the basis of sample size alone is not a complete comparison. The key advantage of cluster sampling is that the sample is collected in clusters, and thus can be significantly less expensive to schedule and to collect. This advantage is not reflected in a comparison based on sample size alone. In order to make a more equitable comparison, a cost function was defined which breaks the cost of sampling into two components. The first component is the overhead cost associated with getting a collector to the location where the measurement is to be made. The second component is the cost of making the observation, once at the proper place and time.

Based on the results from the Commuter Boat population, both proportional stratified sampling and systematic cluster sampling are assumed to require 25% fewer departures in a sample than what is required by simple random sampling. Based on this assumption, these two sampling plans are compared for eight different hypothetical routes using this cost function, for both point checks and ride checks. Each route has a unique combination of three parameter values: mean headway, daily hours of service, and population coefficient of variation. The mean headway is either 10 minutes or 1 hour, the

102

daily hours of service is either 18 hours or 6 hours, and the population coefficient of variation is either 1.00 or 0.40. Each route was assumed to have the same one-way running time, 30 minutes.

The cost comparison results were presented in terms of a break-even overhead cost. This is the cost, in hours, for a checker to travel between work and the observation location for each observation obtained. If the actual overhead cost is greater than this break-even amount, then the systematic selection of clusters is more cost efficient than proportional stratified sampling. For the point check cost comparison, the route parameter which dominates the break-even overhead cost is the mean headway. For the short headway routes, systematic clusters sampling was dominant, while for the long headway routes, the break-even overhead cost was a little over one hour.

For ride checks, the break-even overhead cost was roughly 40 minutes for all the routes. A cluster observation was modeled in the cost function as requiring 1.00 hour to collect (the round-trip running time), and a proportional stratified observation was modeled as requiring 30 minutes to collect. In addition, the cost function used in this research did not consider cost factors that are difficult to quantify such as time required to select and schedule each random observation. Thus, this result was interpreted as strong evidence that systematic cluster sampling is more cost efficient that both simple random sampling and proportional stratified sampling.

5.2 Research Extensions

The most difficult aspect of monitoring transit ridership data is that the precision of any sampling plan must be computed from estimates of the parameters of variation made from the sample. This is a far different situation from that in this paper, where all departures in the population were know. There are two possible approaches to this problem. One is to continue to compare population parameters between systems, and to seek out any consistency and to find the causes of this consistency. The results from this
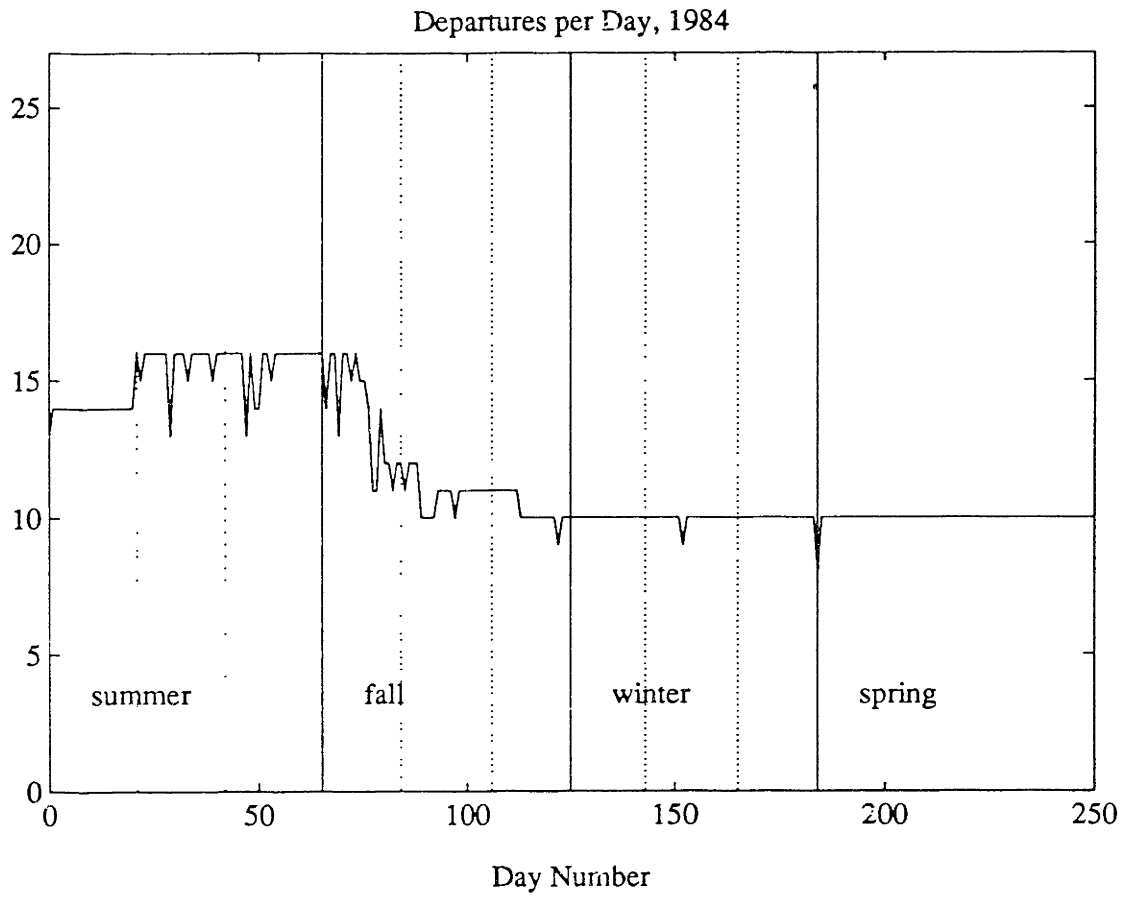
103

research suggest that there is considerable consistency, which is likely due to scheduling similarities. The other approach is to take complete populations such as this one, and empirically compute the frequency distributions of sample variance estimates for different sample sizes to see how these estimators behave.

The relationship between scheduling and the parameters of variation found in this research suggest that these parameters of variation are a type of performance indicator. For example, if the between-day variation is higher for one route than another over the same year, but their overall population variation is the same, then one route is not adjusting as well to seasonal patterns. As Automatic Passenger Counters become more prevalent in transit systems, the use of such performance indicators will be a reasonable way to condense the huge amounts of data into route-level scheduling performance indicators.
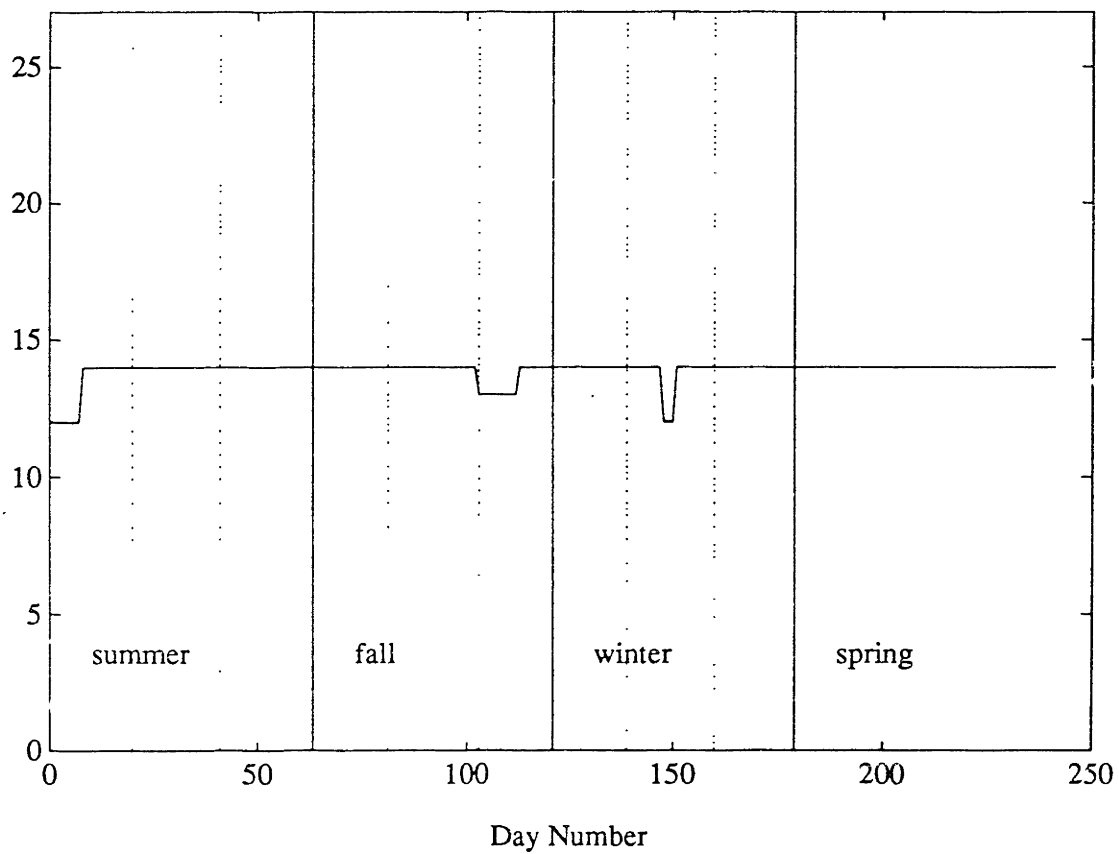
Finally, there is the issue of market research of transit customers. The technique of cluster sampling can be applied in this domain just as well as in the ridership domain. This is an area with much room for growth, given the primitive state of marketing in public transit and the need to improve productivity.
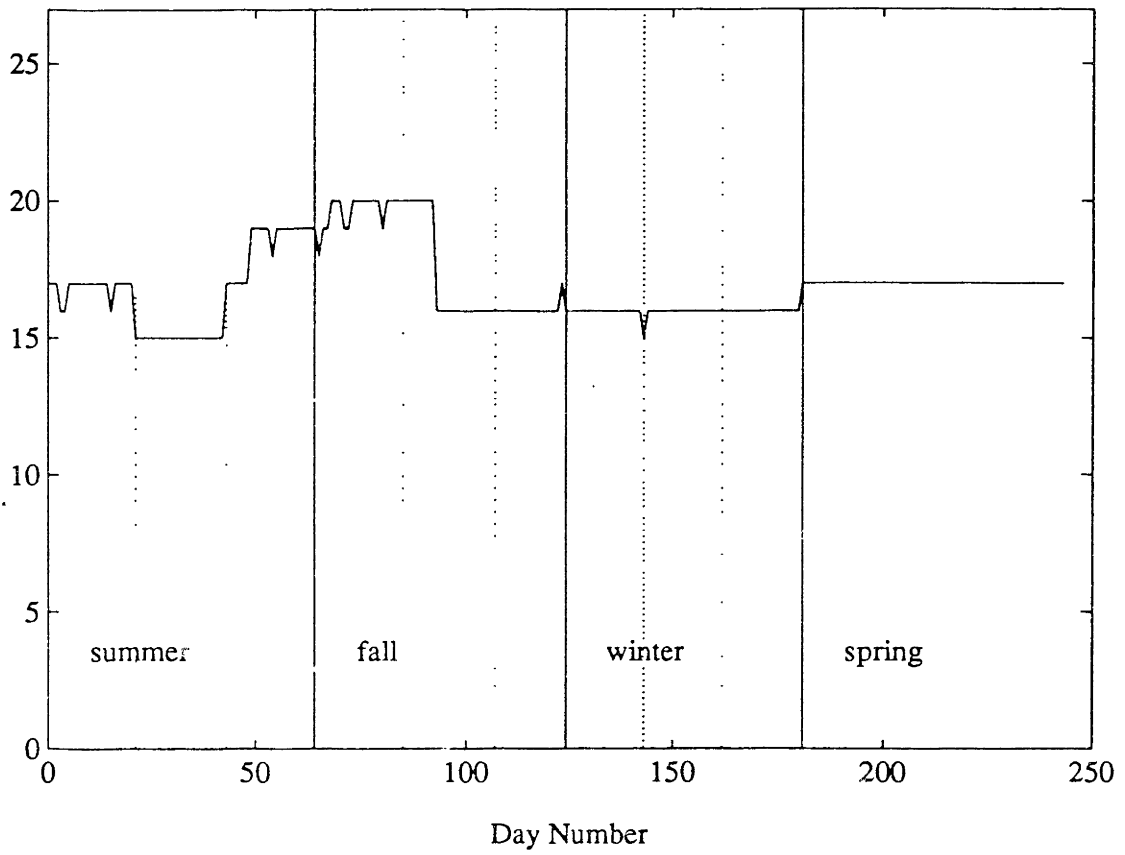
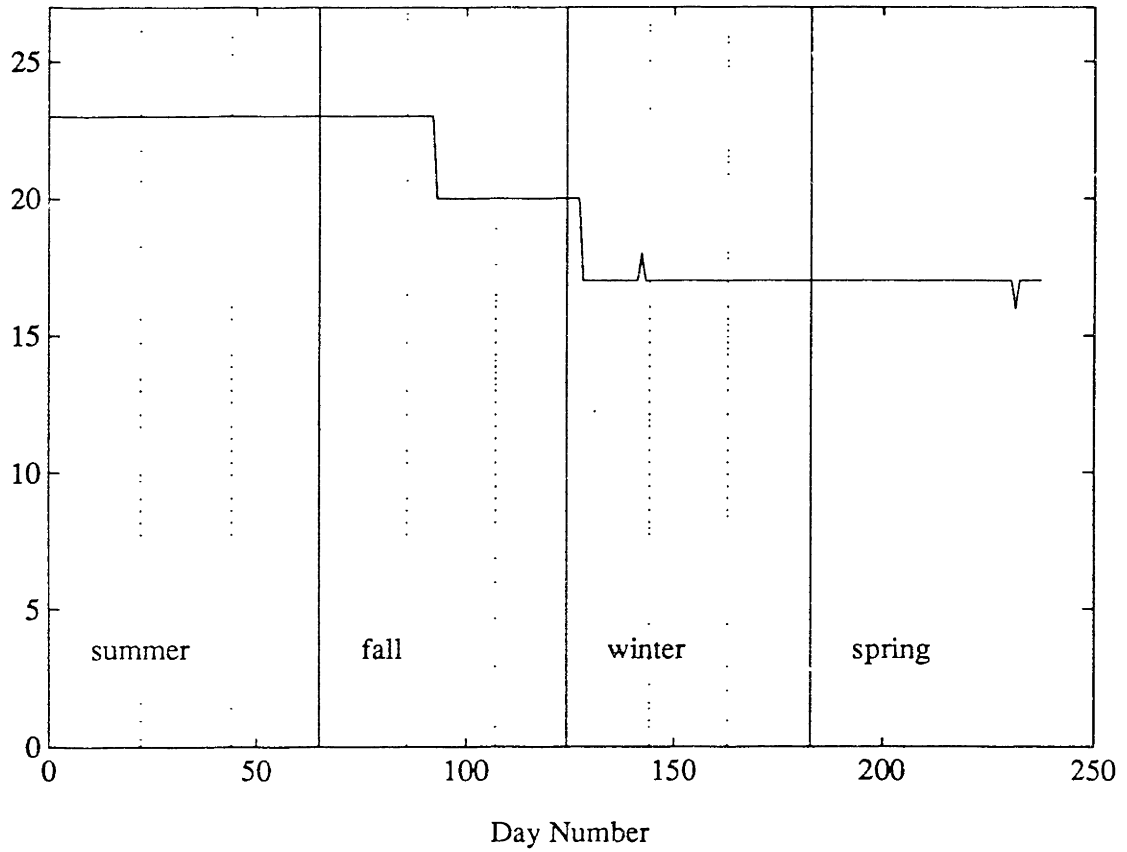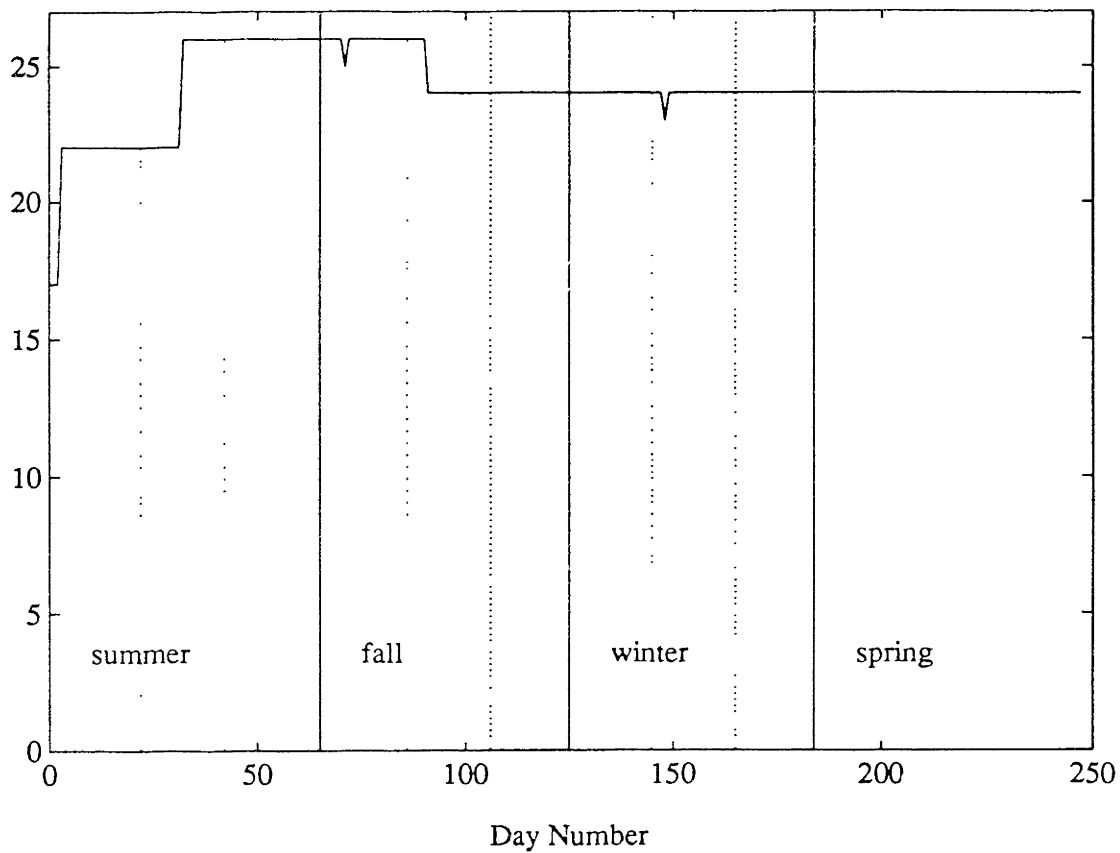# A. Yearly Patterns of Departures per Day

Departures per Day, 1984

Departures per Day, 1985

Day Number

summer   fall   winter   spring

Departures per Day, 1986

| summer | fall | winter | spring |

Day Number

Departures per Day, 1987

Day Number

Departures per Day, 1988

B. Holidays

The days listed below were not included in BY1984 as they were either holidays on which no service was operated, or holidays on which a light schedule was operated, or other days close to holidays on which light shedules were operated or abnormally low ridership occured.

July 4	Wed	no trips, Independence Day
Sept 3	Mon	no trips, Labor Day
Oct 8	Mon	no trips, Columbus Day
Nov 12	Mon	no trips, Veterans Day
Nov 22	Thr	no trips, Thanksgiving
Nov 23	Fri	light ridership, day after Thanksgiving
Dec 24	Mon	light ridership, day before Christmas
Dec 25	Tue	no trips, Christmas
Dec 31	Mon	light ridership, day before New Years
Jan 1	Tue	no trips, New Years Day
Feb 18	Mon	no trips, Presidents' Day
Apr 15	Mon	no trips, Patriots' Day
May 27	Mon	no trips, Memorial Day

The days listed below were not included in BY1985 as they were either holidays on which no service was operated, or holidays on which a light schedule was operated, or other days close to holidays on which light shedules were operated or abnormally low ridership occured.

July 4	Thr	no trips, Independence Day
July 5	Fri	light ridership, day after Independence Day
Oct 14	Mon	no trips, Columbus Day
Nov 11	Mon	light schedule, Veterans Day
Nov 28	Thr	no trips, Thanksgiving
Nov 29	Fri	light ridership, day after Thanksgiving
Dec 24	Tue	light ridership, day before Christmas
Dec 25	Wed	no trips, Christmas
Dec 26	Thr	light ridership, day after Christmas
Dec 31	Tue	light ridership, day before New Years
Jan 1	Wed	no trips, New Years Day
Jan 20	Mon	light schedule, Martin Luther King's Day
Feb 17	Mon	no trips, Presidents' Day
Apr 21	Mon	light schedule, Patriots' Day
May 26	Mon	no trips, Memorial Day

The days listed below were not included in BY1986 as they were either holidays on which no service was operated, or holidays on which a light schedule was operated, or other days close to holidays on which light shedules were operated or abnormally low ridership occured.

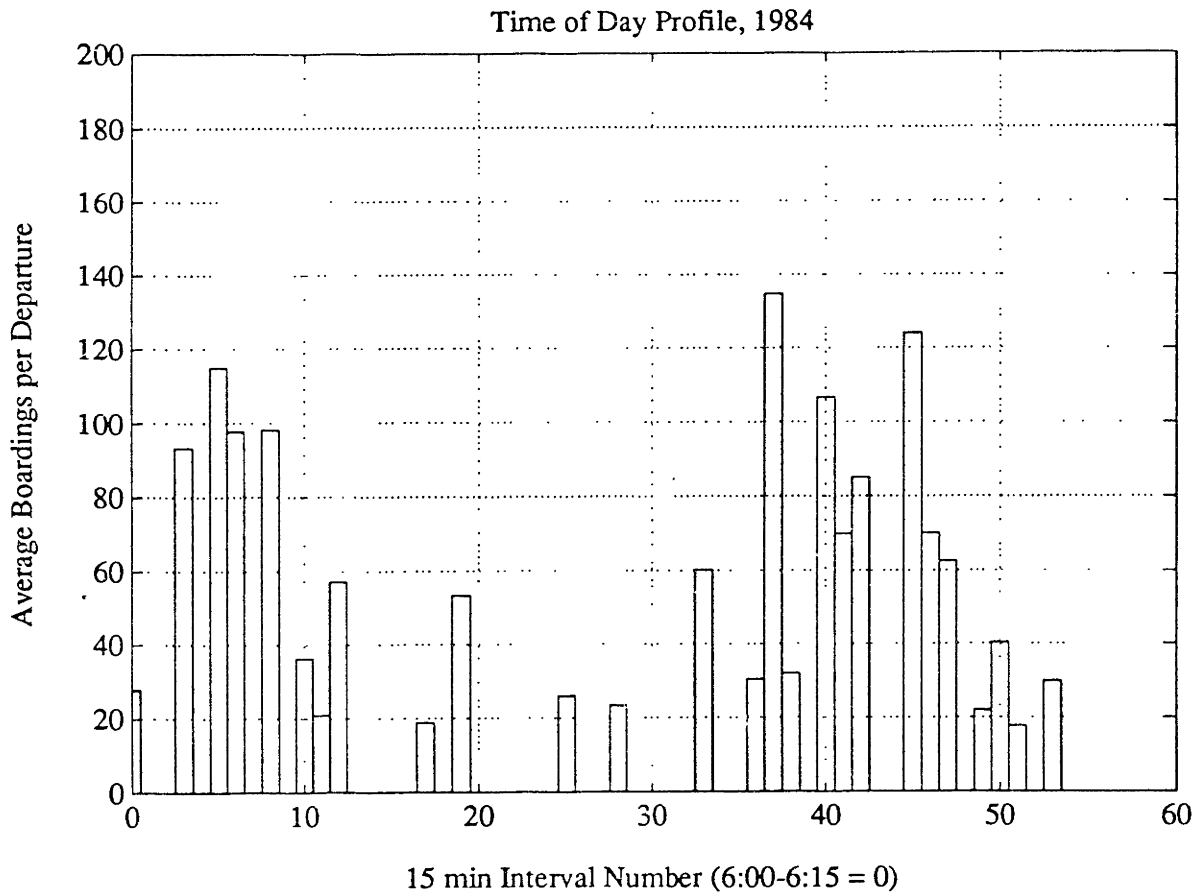|		
July 4	Fri	no trips, Independence Day
Sept 1	Mon	no trips, Labor Day
Oct 13	Mon	no trips, Columbus Day
Nov 11	Tue	no trips, Veterans Day
Nov 27	Thr	no trips, Thanksgiving
Nov 28	Fri	light ridership, day after Thanksgiving
Dec 24	Wed	light ridership, day before Christmas
Dec 25	Thr	no trips, Christmas
Dec 26	Fri	light ridership, day after Christmas
Jan 1	Thr	no trips, New Years Day
Jan 2	Fri	light schedule, day after New Years Day
Jan 19	Mon	no trips, Martin Luther King's Day
Apr 20	Mon	light schedule, Patriots' Day
May 25	Mon	no trips, Memorial Day

The days listed below were not included in BY1987 as they were either holidays on which no service was operated, or holidays on which a light schedule was operated, or other days close to holidays on which light shedules were operated or abnormally low ridership occured.

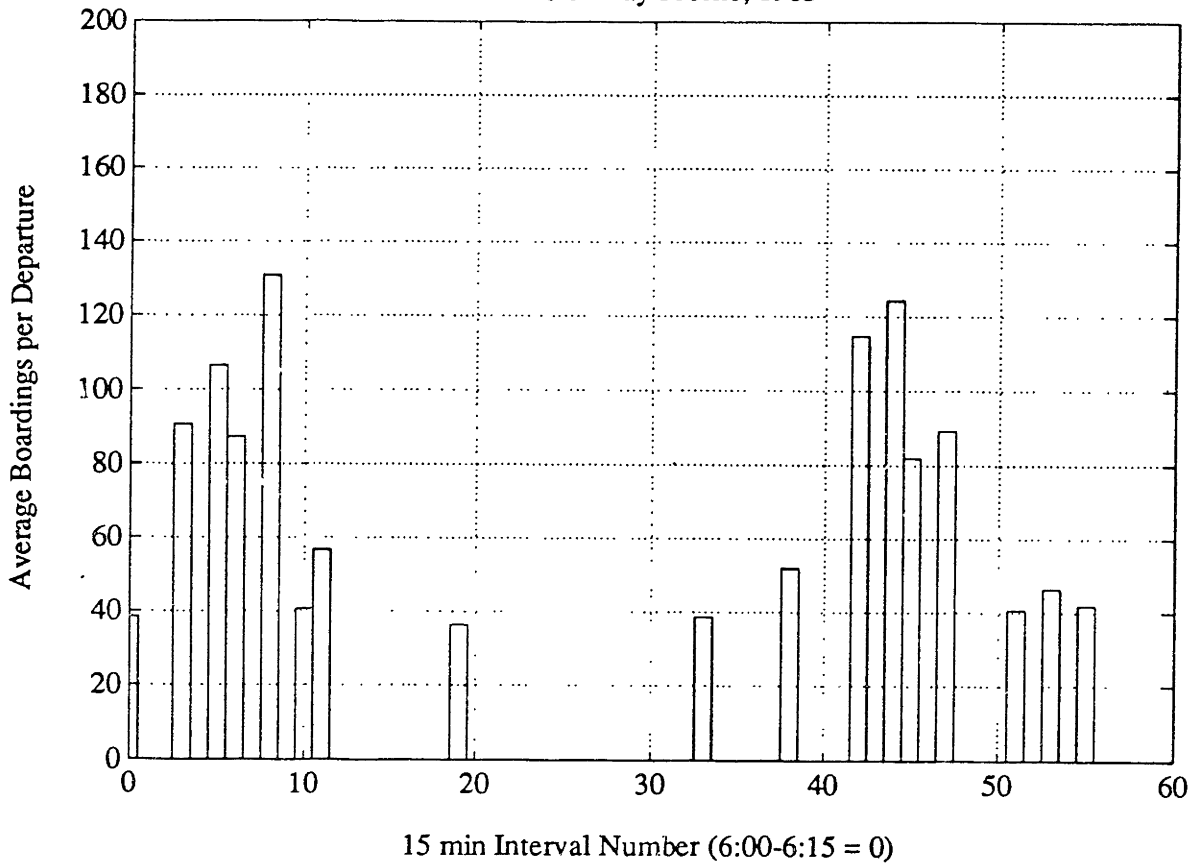| July 3 | Fri | light schedule, Friday before Independence Day |
Sept 7	Mon	no trips, Labor Day
Oct 13	Mon	no trips, Columbus Day
Nov 11	Wed	light schedule, Veterans Day
Nov 12	Thr	light schedule, day after Veterans Day
Nov 26	Thr	no trips, Thanksgiving
Nov 27	Fri	light ridership, day after Thanksgiving
Dec 24	Thr	light ridership, day before Christmas
Dec 25	Fri	no trips, Christmas
Dec 31	Thr	light ridership, day before New Years Day
Jan 1	Fri	no trips, New Years Day
Jan 18	Mon	light ridership, Martin Luther King's Day
Feb 15	Mon	no trips, Presidents' Day
Apr 18 - 30		missing 2 full weeks (10 days) of ridership data
May 30	Mon	no trips, Memorial Day

The days listed below were not included in BY1988 as they were either holidays on which no service was operated, or holidays on which a light schedule was operated, or other days close to holidays on which light shedules were operated or abnormally low ridership occured.

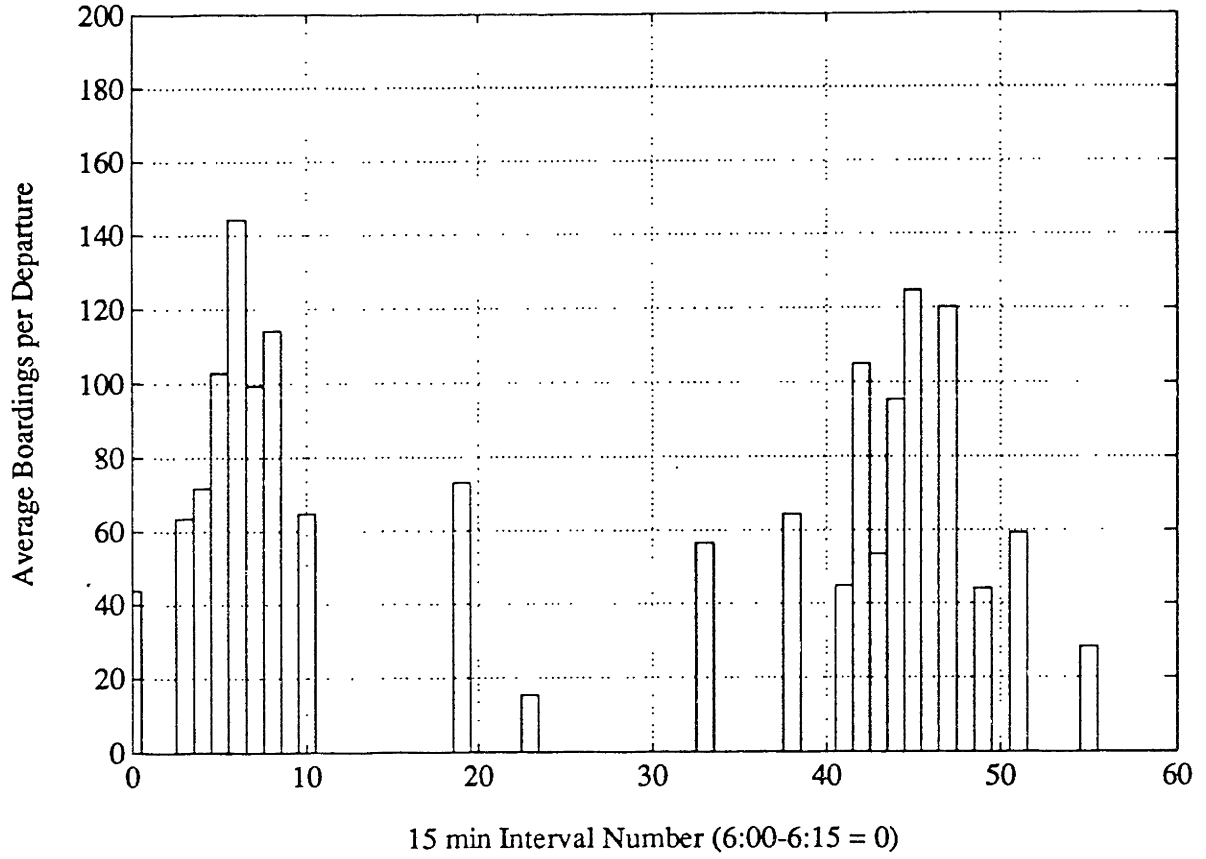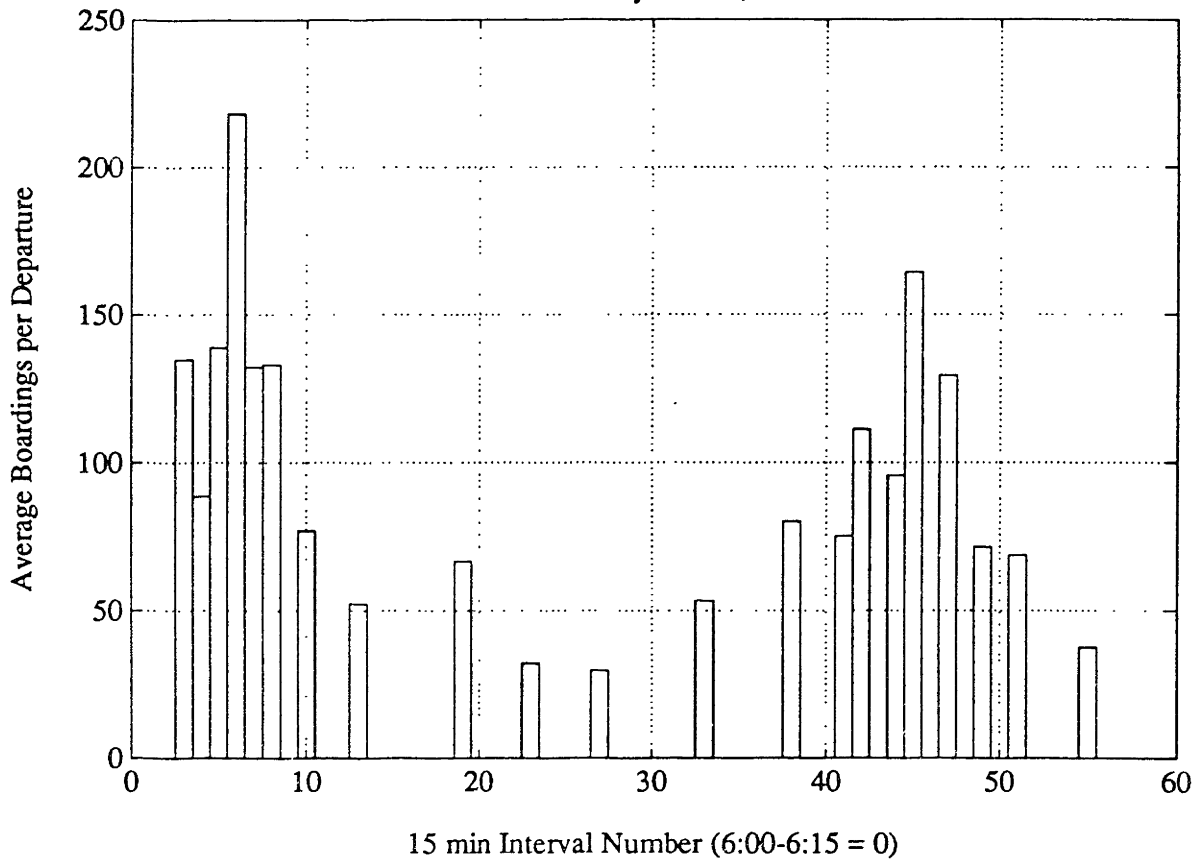| July 4 | Mon | no trips, Independence Day |
Sept 5	Mon	no trips, Labor Day
Oct 10	Mon	no trips, Columbus Day
Nov 11	Fri	no trips, Veterans Day
Nov 24	Thr	no trips, Thanksgiving
Nov 25	Fri	light ridership, day after Thanksgiving
Dec 23	Fri	light ridership, day before Christmas
Dec 26	Mon	no trips, Christmas
Jan 2	Mon	no trips, day after New Years Day
Jan 16	Mon	light ridership, Martin Luther King's Day
Feb 20	Mon	no trips, Presidents' Day
Apr 17	Mon	light ridership, Patriots' Day
May 29	Mon	no trips, Memorial Day

C. Yearly Time of Day Profiles
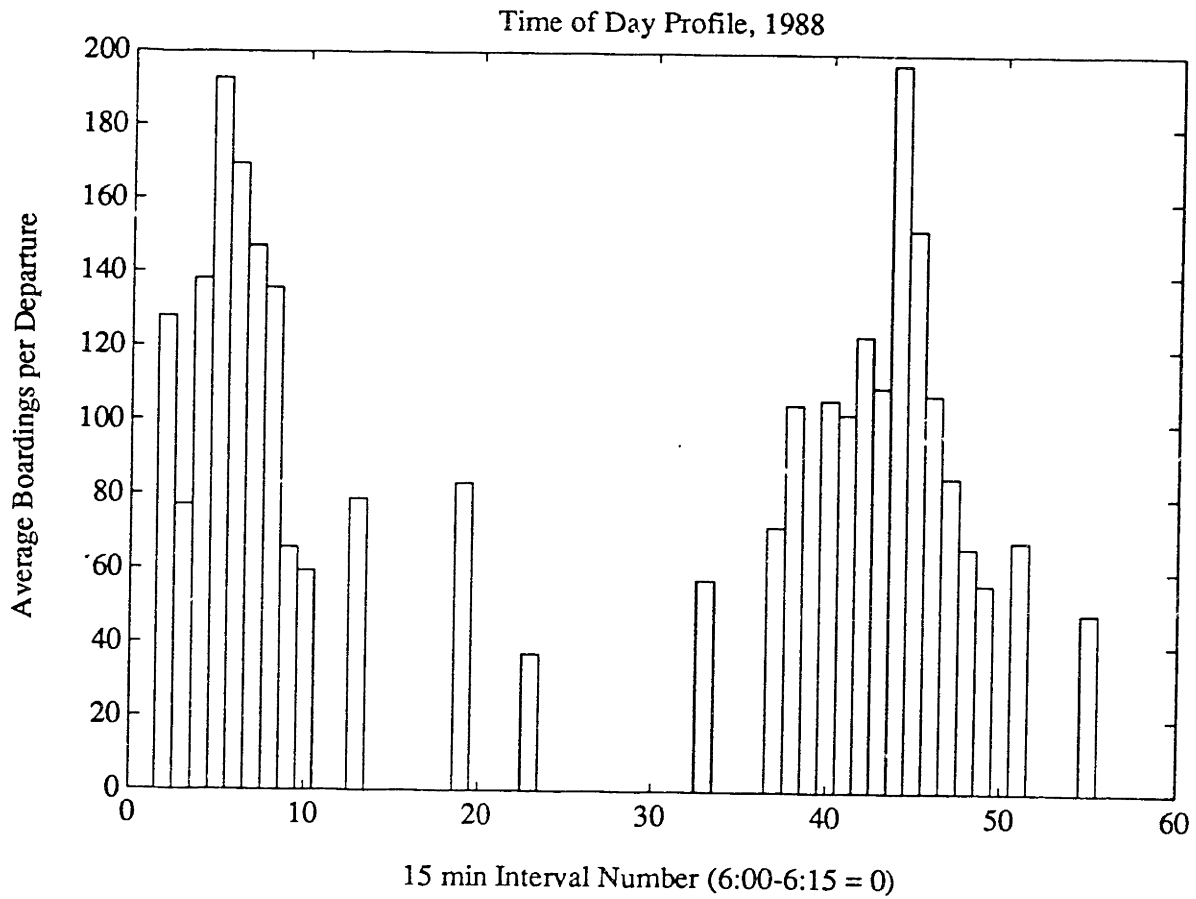
Time of Day Profile, 1984

Average Boardings per Departure vs. 15 min Interval Number (6:00-6:15 = 0)

Time of Day Profile, 1985

15 min Interval Number (6:00-6:15 = 0)

Average Boardings per Departure

Time of Day Profile, 1986

Time of Day Profile, 1987

Time of Day Profile, 1988

Average Boardings per Departure vs. *15 min Interval Number (6:00-6:15 = 0)*

D. Reference List

American Public Transit Association. 1990 Transit Fact Book. Washington, D.C.: American Public Transit Association, 1990.

Bates, John W., J. Dewel Lawrence. "Transit Marketing: A Strategic Approach." Transportation Quarterly 40, no. 4 (October 1986): 549-558.

Canada, Ottawa-Carleton Regional Transit Commission, Systems Planning Department. Report Documentation For Users: Automatic Passenger Counting System. Ottawa: Ottawa-Carleton Regional Transit Commission, June 1988.

Cochran, William G. Sampling Techniques. New York: John Wiley & Sons, Inc., 1977.

Congress, 1982. see The Federal Public Transportation Act of 1982.

Congress, 1990. see The Federal Mass Transportation Act.

Drake, Alvin W. Fundamentals of Applied Probability Theory. New York: McGraw-Hill Book Company, 1966; A McGraw-Hill Classic Textbook Reissue, 1988.

DOT, 1986. see U.S., Department of Transportation. Private Sector Involvement in Urban Transportation: Case Studies.

Ferguson, Erik. Total Annual Passenger Miles: Ratio Estimation with a Sample From Extensive Ride Checks. Draft Report, Service Planning Section, Orange County Transit District, Garden Grove, California, 1987.

The Federal Mass Transportation Act. U.S. Code Annotated. Title 49 Appendix, secs. 1607a, 1607a-1, 1617 (1990).

The Federal Public Transportation Act of 1982. U.S. Code, P.L. 97-424, Legislative History (1982).

Fielding, Gordon J. Managing Public Transit Strategically. San Francisco: Jossey-Bass Publishers, 1987.

Foote, Peter J., William A. Hancox. Producing Section 15 Service-Consumed Data: A Dynamic Challenge for Large Transit Authorities. Preprint of paper 880585 from the 68th Annual Meeting of the Transportation Research Board, 1989.

Furth, Peter G., Keith L. Killough, Gary F. Ruprecht. Cluster Sampling Techniques for Estimating Transit Patronage. Submitted to the Transportation Research Board for presentation in the 1988 Annual Meeting and for subsequent publication.

Hsu, Poh Ser, Nigel H.M. Wilson. Analysis of Transit Data. Draft Final Report, Center for Transportation Studies, Massachusetts Institute of Technology, 1983.

Jones, David W. Jr. Urban Transit Policy. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1985.

Kendall, Maurice G. The Advanced Theory of Statistics. New York: Hafner Publishing Company, 1961.

Kvanli, Alan H., C. Stephen Guynes, Robert J. Pavur. Introduction to Business Statistics. St. Paul: West Publishing Company, 1986.

Lee, Douglass. "Transit Cost and Performance Measurement." Transport Reviews 9, no. 2 (1989): 147-170.

Leick, Janet D., Janice M. Hedemann. "Contracting to Private Transit Operators." Transportation Quarterly 41, no. 3 (July 1987): 427-438.

Miller, Gerald K., Ronald F. Kirby. "A Structured Approach to Monitoring and Evaluating Public Transportation Services." Transportation Quarterly 38, no. 1 (January 1984): 23-46.

Miller, Rupert G. Jr., Beyond ANOVA, Basics of Applied Statistics. New York: John Wiley & Sons, 1986.

Smith, Robert L. Jr. "Improving Section 15 Data Collection Techniques." Transportation Research Record XXXX, 1984.

TCDM, 1985. see U.S., Department of Transportation, Urban Mass Transportation Administration, Office of Planning Assistance.

Teal, Roger F. "Public Transit Service Contracting: A Status Report." Transportation Quarterly 42, no. 2 (April 1988): 207-222.

UMTA, 1977. see U.S., Department of Transportation, Urban Mass Transportation Administration, Office of Transit Management. Urban Mass Transportation Industry Uniform System of Accounts and Records and Reporting System.

UMTA, 1978. see Department of Transportation, Urban Mass Transportation Administration, Office of Transit Management. UMTA C 2710.1.

UMTA, 1985. see Department of Transportation, Urban Mass Transportation Administration, Office of Transit Management. UMTA C 2710.4.

U.S., Department of Transportation, Urban Mass Transportation Administration, Office of Transit Management. Urban Mass Transportation Industry Uniform System of Accounts and Records and Reporting System. UMTA-IT-06-0094-77-1. Springfield, Virginia: National Technical Information Service, January 1977.

U.S., Department of Transportation, Urban Mass Transportation Administration, Office of Transit Management. UMTA C 2710.1 Sampling Procedures for Obtaining Fixed Route Bus Operating Data Required Under the Section 15 Reporting System. Washington D.C.: Urban Mass Transportation Administration, February 22, 1978.

U.S., Department of Transportation, Urban Mass Transportation Administration, Office of Transit Management. UMTA C 2710.4 Revenue Based Sampling Procedures for Obtaining Fixed Route Bus Operating Data Required Under the Section 15 Reporting System. Washington D.C.: Urban Mass Transportation Administration, May 10, 1985.

U.S., Department of Transportation, Urban Mass Transportation Administration, Office of Transit Management. Transit Data Collection Design Manual. Washington, D.C.: U.S. Department of Transportation, DOT-1-85-38, June 1985.

U.S., Department of Transportation. Private Sector Involvement in Urban Transportation: Case Studies. Washington, D.C.: U.S. Department of Transportation, December 1986.