4-2002

Automatic Passenger Counter Evaluation: Implications for National Transit Database Reporting

Thomas J. Kimpel
Portland State University

James G. Strathman
Portland State University

David Griffin
TriMet

Steve Callas
TriMet

Rick Gerhart
TriMet

Automatic Passenger Counter Evaluation:
Implications for National Transit Database Reporting

Thomas J. Kimpel
James G. Strathman

Center for Urban Studies
College of Urban and Public Affairs
Portland State University
Portland, OR 97207
(503) 725-4020

David Griffin
Steve Callas
Richard L. Gerhart

Tri-Met
4012 SE 17th Ave.
Portland, OR 97202
(503) 239-3000

FINAL REPORT

April 2002

Abstract

This paper presents findings from a study assessing the accuracy and precision of automatic passenger counter (APC) technology at Tri-County Metropolitan Transportation District of Oregon (Tri-Met). Video surveillance cameras, rather than ride checkers, were used to establish reference values for determining APC accuracy and precision. Analysis of data collection, processing and reporting methods at Tri-Met indicates that APC data, along with a properly designed sampling plan, can be used for internal monthly ridership reporting and annual National Transit Database (NTD) reporting. Presently, NTD sampling plans require that bus trips be randomly selected prior to manual data collection efforts. The sampling methodology developed for this analysis allows APC data to be matched with a random selection of bus trips following automated data collection.

I. Introduction

Automatic passenger counter (APC) technology is utilized by many transit properties throughout North America (Attanucci & Vozzolo, 1983; Hodges, 1985; Boyle, 1998). One of the main benefits of APC technology is that it allows data to be collected at reasonable cost to the agency relative to manual data collection efforts (Casey et. al., 1996; Boyle, 1998). Early research pertaining to APC technology focused almost exclusively on the issue of accuracy relative to manual passenger counts (Attanucci and Vozzolo, 1983). In general, the studies found that APCs were more accurate at recording boarding activity than alighting activity and that there was a tendency for APCs to undercount passenger activity. Although discrepancies between APC counts and manual counts were found to exist, the differences were often not statistically significant (Multisystems Inc., 1982; Strathman & Hopper, 1991).

One of the conclusions from Transit Cooperative Research Report (TCRP) Synthesis 29 is that active management of the system is necessary to ensure the continued success of APC technology (Boyle, 1998). In particular, investments in time and effort are needed in the areas of data collection, processing, storage, and retrieval as well as the refinement of analytical techniques. While the actual uses of APC data vary according to the specific needs of each agency, the use of APC technology for collecting system-level ridership information is relatively rare (Boyle, 1998).

The objectives of the present study are to (1) assess the accuracy and precision of passenger count data at the stop level, and (2) to develop sampling plans that can be used for internal monthly ridership reporting and annual NTD reporting.

Tri-Met has been a strong proponent of APCs since the agency first started experimenting with the technology in 1983. At present, the Tri-Met bus fleet is approximately 61% APC-equipped. Many transit properties rotate APC-equipped vehicles throughout the system in order to ensure that bus trips are sampled a minimum number of times during a given time period. Because of widespread deployment of APCs at Tri-Met, the agency rarely needs to assign APC-equipped vehicles to specific bus trips for data collection purposes.

This study differs from most previous studies in that APC counts were compared to counts derived from video cameras, rather than ride checkers. The video camera information is considered ground truth and used as the basis for determining APC accuracy. Only one other study was found to have used video cameras to assess APC accuracy. The study was part of a Transit-IDEA project that looked at errors associated with pressure sensitive mats for various boarding scenarios on one bus over a four hour time period (Greneker et. al., 1996). Each vehicle analyzed in the present analysis was equipped with three digital video cameras– two pointed towards each door and one down the aisle to maximize surveillance capabilities. Only "clean" APC observations validated during post-processing were utilized in the study. Approximately 35% of APC data at Tri-Met is rejected because of data quality considerations. The post-processing screening criteria address (1) the compatibility of each of the four infrared beam counts in relation to each other and (2) whether the difference between boardings and alightings divided by total boardings and alightings is less than 10% at the block level for the assignment day. Overall, 2,921 stop-level observations were analyzed, at least partially representing 57 vehicle blocks, 134 trips, and 21 bus routes.

II. APC Precision and Accuracy

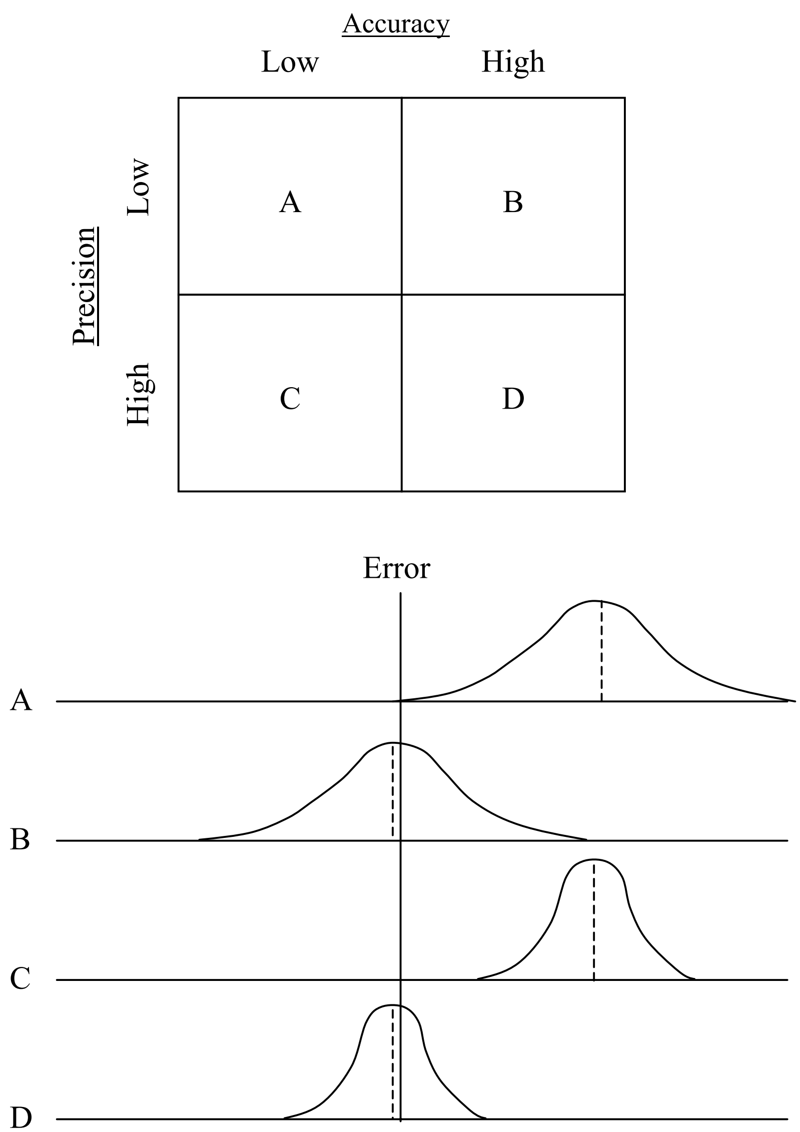
Two factors that need to be discussed in greater detail concern APC accuracy and precision. APC accuracy relates to the systematic over or undercounting of passengers relative to their true values. Estimates of passenger activity that systematically differ from ground truth are inaccurate, or biased. If the systematic bias is known, then correction factors can be applied to APC ridership counts to control for the amount of bias.

APC precision concerns the distribution of error between the measured and true value of passenger activity. Precision is always stated with respect to a specific level of confidence. For example, a transit agency may require that monthly ridership figures have a precision level of +/- 5% at the 95% level of confidence. The agency wants to be 95% certain that APC-generated ridership values fall within +/- 5% of their true value. Unlike accuracy, a correction factor cannot be applied to ridership estimates in cases of low precision.

The maintained assumption for determining APC accuracy using data collected by ride checkers is that the manual counts are error free. Previous work established the accuracy of APCs relative to manual counts, but the assumption of zero measurement error in manual counts is suspect (Hobbs & Poirier, 1983; Boyle, 1998; Strathman et. al., 2001). An important distinction between automatic and manual data collection techniques is that error associated with manual data collection efforts is likely be random (Boyle, 1998) whereas error associated with APCs is likely to be both systematic and random. In the first case, error results from human factors associated with the collection of passenger activity information on survey trip sheets and the transferring of information to daily record sheets. In the second case, error results from a number of factors including mechanical problems, environmental factors, passenger behavior, and post-processing of APC data. The use of video cameras to establish reference ridership is likely to be more precise given that data transcribers can replay video in cases where ambiguity in passenger activity exists. Furthermore, establishing reference ridership using video transcribers (few persons) is likely to result in less measurement error relative to ride checkers (many persons).

Figure 2.1 shows the relationship between accuracy and precision. Example A in the top part of the figure represents a case where passenger activity estimates have low accuracy and low precision. As can be seen in the lower part of the figure for Example A, the mean value of the sample distribution is far from the true value indicating low accuracy due to systematic over counting. The spread of the distribution is also relatively wide, indicating low precision (a high degree of measurement error). At the opposite extreme is Example D where passenger activity estimates have high accuracy and high precision. The sample mean is much closer to the true value and the sample distribution is also much narrower.

Figure 2.1: Diagram of Accuracy and Precision



The treatment of bus operators is an important factor to consider when discussing APC accuracy. The question arises as to whether to count bus operators as boarding and alighting passengers when establishing reference ridership values using video cameras. APCs are accurate to the extent that they count boarding and alighting activity correctly, regardless of whether the activity is associated with passengers or operators. In theory, APCs should always count slightly high relative to ground truth because of the activity of bus operators. In order to adequately test APC accuracy, bus operators were included in boarding and alighting estimates derived from video. In contrast, bus operators were not included in passenger load estimates generated from video in order to remain consistent with NTD reporting practices.

An important issue with respect to accuracy concerns post-processing of APC data. Tri-Met uses a post-processing algorithm that balances boardings and alightings on a per-vehicle block basis such that the number of boardings and alightings are equal at the end of the service day. Passenger loads are estimated from the "massaged" boarding and alighting data in a manner that ensures that loads are zero when vehicles return to the garage at the end of the service day. Presently, the algorithm does not allow for negative passenger loads at individual bus stops. The consequence of not allowing for negative passenger loads is that errors introduced at early points in the service day tend to propagate over time, resulting in inflated load estimates. Preliminary analysis of the passenger load information derived from the post-processing algorithm indicated that loads needed to be calculated in a more direct manner. For the present analysis, a new passenger load variable was developed based on the following criteria: (1) raw boarding and alighting data (rather than the massaged data) were used, (2) negative passenger loads were allowed and (3) loads were set to zero at locations with layovers greater than 5 minutes. This procedure produced APC load estimates that were more consistent with camera load estimates than those obtained from the post-processing algorithm.

Analysis of APC accuracy was undertaken with means tests and a regression model. Analysis of APC precision was undertaken with a reverse regression model. Three different samples were used in the analysis, including the full sample and two smaller samples differentiated by bus fleet. The justification for splitting the full sample according to bus fleet was to ascertain whether significant differences in APC accuracy and precision exist depending on bus type.

Table 2.1 shows the variables used in the analysis. Passenger activity variables generated from camera observations include boardings (ON), alightings (OFF), and loads (LOAD).

Table 2.1: Description of Variables

Variable	*Description*
ON	Boardings (Camera)
ONS	Boardings (APC)
OFF	Alightings (Camera)
OFFS	Alightings (APC)
LOAD	Load (Camera)
LOADS	Load (APC)
DON	Boarding difference (ONS - ON)
DOFF	Alighting difference (OFFS- OFF)
DLOAD	Load difference (LOADS – LOAD)
GIL	Bus type (1=Gillig, 0=NFI)

Passenger activity variables generated from APC observations include boardings (ONS), alightings (OFFS), and loads (LOADS). Variables representing the difference between APC and camera observations are boarding difference (DON), alighting difference (DOFF), and load difference (DLOAD). GIL is a dummy variable for Gillig bus type, New Flyer Industries (NFI) bus type otherwise. NFIs are low-floor buses. Because all buses analyzed in the study use similar APC technology, the main distinction between Gillig and NFI bus types is not due to differences in APC equipment but to bus configuration. Bus configuration may affect APC accuracy due to differences in sensor location, the behavior of passengers near doorways (e.g., stairwell vs. low floor), passenger bunching effects due to different doorway widths, and door exiting behavior (e.g., front vs. rear door).

Descriptive statistics for selected variables are presented in Table 2.2. Descriptive statistics for all of the variables used in the analysis are presented in Appendix A. The percentage difference between APC and camera counts is shown in the right hand column of Table 2.2. In the full sample, the difference between APC and camera counts (DON) is small, with APCs showing a tendency to overestimate boardings by 0.88%. For alightings, the difference between APC and camera counts (DOFF) is slightly larger, with APCs overestimating alightings by 1.83%. The percentage difference between APC and camera loads (DLOAD) in the full sample is 6.08%. When differentiated by bus type, APCs on Gilligs are shown to overestimate boardings and alightings by a noticeable margin (4.24% and 5.37% respectively) whereas NFIs tend to underestimate boardings and alightings by a small margin (-1.06% and -0.38% respectively). The difference between APC and camera loads (DLOAD) is considerably higher for Gilligs at 9.82% compared to NFIs at 3.94%. The reason for the greater disparity between APC and camera loads for Gilligs compared to NFIs is because the difference between boardings and alightings is much greater for Gilligs.

Table 2.2: Descriptive Statistics

Full Sample							
Variable	*N*	*Mean*	*St. Dev.*	*Var.*	*Min.*	*Max.*	*Diff. (%)*
DON	2,921	0.0130	0.5206	0.2711	-7	7	0.88
DOFF	2,921	0.0243	0.5219	0.2724	-9	8	1.83
DLOAD	2,921	0.8545	2.9381	8.6326	-9	13	6.08

Gillig Sample							
Variable	*N*	*Mean*	*St. Dev.*	*Var.*	*Min.*	*Max.*	*Diff. (%)*
DON	1,155	0.0580	0.4436	0.1968	-2	4	4.24
DOFF	1,155	0.0693	0.5558	0.3089	-4	8	5.37
DLOAD	1,155	1.2745	3.1532	9.9428	-9	12	9.82

NFI Sample							
Variable	*N*	*Mean*	*St. Dev.*	*Var.*	*Min.*	*Max.*	*Diff. (%)*
DON	1,766	-0.0164	0.5635	0.3176	-7	7	-1.06
DOFF	1,766	-0.0051	0.4964	0.2464	-9	7	-0.38
DLOAD	1,766	0.57984	2.755	7.5899	-6	13	3.94

APC Accuracy

Although differences are shown to exist between APC and video camera counts, the question arises as to whether these differences are statistically significant. Means tests were run for the passenger activity variables expressed as differences between APC and camera counts. APC measurements are deemed accurate if the 95% confidence interval (p <0.05) encompasses the value zero. The confidence intervals were calculated using Equation 1.

$$C.I. = \Delta \overline{X} \ {+/-}\ 1.96 * \left(\sigma / \sqrt{N} \right) \qquad \text{[Equation 1]}$$

Table 2.3 shows the results of the means tests for the three samples. The results indicate that boardings in the full sample and boardings and alightings in the NFI sample are measured with statistical accuracy. APCs are shown to systematically over count boardings on Gilligs and passenger loads irrespective of bus type. The finding of no statistically significant difference between APC and camera counts for boarding passengers in the full sample is consistent with previous research by Strathman and Hopper (1991). The finding of a significant difference between APC and camera counts for alighting passengers in the full sample runs contrary to the results reported in their study.

Table 2.3: APC Accuracy Means Tests Results

Variable	*Full Sample: 95% C.I.*		*Gillig Sample: 95% C.I.*		*NFI Sample: 95% C.I.*	
DON	-0.0059	0.0319	0.0324	0.0836	-0.0427	0.0099
DOFF	0.0054	0.0432	0.0372	0.1013	-0.0283	0.0181
DLOAD	0.7480	0.9611	1.0927	1.4564	0.4514	0.7083

Results from the means tests indicate that calibration factors may need to be applied in certain instances. The National Institute of Standards and Technology recommends that calibration factors be applied to values derived from a particular instrument when systematic errors are known to exist (NIST, 1998). This is particularly relevant in cases where an instrument cannot be calibrated or where an instrument has reached its maximum precision. Within the transit industry, OC Transpo in Ottawa, Ontario, Canada applies a correction factor to address the systematic undercounting of alightings at the vehicle block level (Attanucci & Vozzolo, 1983). Correction factors were calculated using the following equation (Equation 2).

$$C.F. = 1 - \left(\left(\overline{X}_{APC} - \overline{X}_{CAM} \right) / \overline{X}_{CAM} \right) \qquad \text{[Equation 2]}$$

Based on the results of the analysis for the full sample, a correction factor of 0.9426 should be applied to APC estimates of passenger loads at Tri-Met. The analysis indicates that no correction factor is warranted for boardings.

Analysis of the difference between APC and camera counts for the three passenger activity variables against Gillig bus type shows that there are significant differences in accuracy between Gilligs and NFIs (Table 2.4). The individual coefficients represent the magnitude of the difference between the two bus types. The results show that APC estimates of boardings, alightings and loads on Gilligs are significantly higher compared to NFIs.

Table 2.4: Differences in Passenger Counts by Bus Type

DV	*IV*	*Coef.*	*Std. Err.*	*T-Ratio*	*N*
DON	GIL	0.0744	0.0197	3.7860	2,921
DOFF	GIL	0.0744	0.0197	3.7740	2,921
DLOAD	GIL	0.6946	0.1105	6.2880	2,921

An additional series of means tests were run to test the effect of the cutoff value used to screen APC data at Tri-Met. During post-processing, APC readings are flagged as "good" if the ratio of the difference between boardings and alightings to total boardings and alightings is less than 10% on a per-vehicle block basis (PDIFF). Because the data used in the present analysis have a cutoff value of 5% or less (out of 57 blocks represented), a full test from 10% was not possible. Difference in means tests were run on samples with PDIFF decremented by one additional percentage point each time. Results of the analysis are presented in Table 2.5

Table 2.5: Analysis of APC Screening Criteria

		Boarding Difference (DON)		*Alighting Difference (DOFF)*	
Cutoff Value	*N*	*95% C.I.*	*95% C.I.*	*95% C.I.*	*95% C.I.*
PDIFF = < 5%	2,921	-0.0059	0.0319	0.0054	0.0432
PDIFF = < 4%	2,705	-0.0028	0.0368	0.0038	0.0442
PDIFF = < 3%	2,550	-0.0050	0.0348	0.0011	0.0437
PDIFF = < 2%	2,100	-0.0072	0.0377	-0.0059	0.0440
PDIFF = < 1%	1,298	-0.0031	0.0539	-0.0033	0.0650

The results show that the cutoff value used to validate APCs does not improve accuracy in any appreciable manner until the 2% threshold is reached. At this point, there is no statistically significant difference between APC and camera alightings.

APC Precision

A test of APC precision was also applied to the passenger activity variables. Reverse regression is commonly used to test for the relative effects of measurement error (Wonnacott & Wonnacott, 1970). The first step was to estimate regressions for each of the three passenger activity variables using APC values as the dependent variables and camera value as the independent variables (Equation 3). A second series of regression were run using camera values as the dependent variables and APC values as the independent variables (Equation 4). The inverse of the beta coefficients from the first regressions were used with the beta coefficients from the second regressions in T-tests (Equation 5).

$$Y_{APC} = \alpha + \beta_{CAM} X_{CAM} + \varepsilon \qquad \text{[Equation 3]}$$

$$Y_{CAM} = \alpha + \beta_{APC} X_{APC} + \varepsilon \qquad \text{[Equation 4]}$$

$$TEST_{\alpha=0.05} = \beta_{APC} - \left(1/\beta_{CAM}\right) = 0$$ [Equation 5]

T-tests were used to determine whether measurement error associated with APC was significantly different from zero at the 95% level of confidence. The formula for determining the amount of estimation bias attributed to APC estimates is given in Equation 6.

$$BIAS = \left(\left(1/\beta_{CAM}\right) - \beta_{APC}\right) / \left(1/\beta_{CAM}\right)$$ [Equation 6]

The results of the reverse regressions are presented in Table 2.6. The T-tests show that the amount of measurement error attributed to APCs is statistically significant for each of the passenger activity variables in all three samples. The relative magnitude of estimation bias for APC boardings is 5.5% in the full sample, 4.2% in the Gillig sample, and 6.3% in the NFI sample. Boardings are shown to have less measurement error than alightings in the Gillig sample. In contrast, alightings are shown to have less measurement error than boardings in the NFI sample. The greatest amount of measurement error tends to be associated with passenger loads, particularly on Gilligs.

Table 2.6: Measurement Error Analysis

Full Sample						
DV	*IV*	β(APC)	1/β(CAM)	*Bias (%)*	*Significant*	*N*
ON	ONS	0.9358	0.9907	5.54	Yes	2,921
OFF	OFFS	0.9553	1.0238	6.60	Yes	2,921
LOAD	LOADS	0.9031	0.9903	8.81	Yes	2,921

Gillig Sample						
DV	*IV*	β(APC)	1/β(CAM)	*Bias (%)*	*Significant*	*N*
ON	ONS	0.9263	0.9664	4.15	Yes	1,155
OFF	OFFS	0.8905	0.9713	8.32	Yes	1,155
LOAD	LOADS	0.8684	0.9961	12.83	Yes	1,155

NFI Sample						
DV	*IV*	β(APC)	1/β(CAM)	*Bias (%)*	*Significant*	*N*
ON	ONS	0.9409	1.0039	6.28	Yes	1,766
OFF	OFFS	0.9910	1.0479	5.42	Yes	1,766
LOAD	LOADS	0.9160	0.9830	6.82	Yes	1,766

Discussion of APC Accuracy and Precision Results

The question of whether APC calibration factors are necessary at Tri-Met for both monthly ridership reporting and NTD reporting deserves further discussion because (1) not all APC-equipped bus fleets are represented in the study sample, (2) some of these other fleets are equipped with a different generation of APC technology, and (3) the proportion of NFIs to Gilligs in the analysis differs from that of the Tri-Met system. The reason why certain APC-equipped bus fleets were not included in the analysis is because they lacked video cameras.

Table 2.7. shows the breakdown of APCs by bus fleet at Tri-Met. Percentages are shown in relation to the total fleet (N=700) and the portion of the fleet that is APC-equipped (N=426).

Table 2.7: Differences in APC Coverage by Bus Fleet

Fleet	*APCs*	*% Fleet- Total*	*% Fleet- APC only*
Gillig	65	9.29%	15.26%
NFI	200	28.57%	46.95%
Other	161	23.00%	37.79%
Total	426	60.86%	100.00%

The breakdown of observations in the study sample is 39.5% Gilligs and 60.5% NFIs. Although Gilligs are shown to be measured with error in the partial sample, the full sample shows that APCs are statistically accurate when both bus types are analyzed together. Since the proportion of Gilligs in the study sample is greater than that for the system, it can be ascertained that APC boardings do not need to be calibrated when limited to these two fleets. The only caveat is that there is missing information on the accuracy of APCs on fleets not included in the study. Whether or not APCs are accurate on these fleets is compounded by the fact that many use an earlier generation of APC technology.

APC accuracy is partially contingent on the precision of the reference data. If reference ridership values are measured without error, then one is more likely to find that APC estimates of passenger activity are inaccurate. The results of a previous Tri-Met study (Strathman et. al., 2001) comparing APC-generated load estimates with those from ride checkers found that the relative amount of measurement error-related estimation bias attributed to APCs was 26.84%. In contrast, the amount of estimation bias attributed to APCs using reference ridership obtained from video for the full sample is 8.8%. By basing APC accuracy analysis on data that is less prone to measurement error, a more stringent test of APC accuracy is provided in the present analysis.

The results of the APC accuracy analysis show that APC boardings are measured with statistical accuracy and that APC-based load estimates require a correction factor of 0.9426 passenger (from Equation 2). Because the present study is based on an analysis of only two bus fleets, the question arises as to whether it is appropriate to extrapolate the results of the analysis to the Tri-Met system. We feel that the omission of certain bus fleets from the study is a less serious problem than that of establishing APC accuracy using reference ridership of questionable precision. APC accuracy will likely improve over time as Tri-Met gradually phases out many of these earlier generation, APC-equipped bus fleets.

Quarterly Route Performance Reports

In addition to monthly ridership reports and annual NTD reports, Tri-Met also prepares quarterly route performance reports. The passenger activity values presented in the route performance reports are derived from bus trips with working APCs and are not drawn from a statistical sample. For the route by trip level summary, Tri-Met flags the precision of boardings and load factors by calculating a confidence interval statistic shown in Equation 7. The cutoff value is +/-

20% at the 80% level of confidence. Passenger activity estimates are suspect if the value is greater than 0.2.

$$FLAG = 1.282 * \sqrt{\left(\sigma^2_{ONS} / N\right)} \Big/ \overline{X}_{ONS} \qquad \text{[Equation 7]}$$

No data quality statistics are presented for any of the other summary levels including route, route by time of day, and route by hour although they would yield important information about the precision of the estimates. Closer inspection of the data presented in the route performance reports reveals that many of the estimates are based on a limited number of observations and thus do not provide a sufficient number of degrees of freedom for the formula presented above. For samples with less than 40 APC observations, the appropriate critical T-value should be used. Tri-Met might reexamine whether the APC precision cutoff value is set too low for agency applications given that a considerable number of APC-equipped vehicles have been brought online over the past several years. With respect to accuracy, the finding that APC accuracy varies according to bus type has implications for quarterly route performance reporting. The finding is particularly relevant in instances where certain bus routes operate Gilligs almost exclusively.

III. Sampling Plans for Internal Monthly Ridership Reporting and NTD Reporting

The third component of the study consists of developing sampling methodologies for internal monthly system-level ridership reporting and annual NTD reporting. NTD reporting at Tri-Met is based upon the manual data collection procedures outlined in Federal Transit Administration Circular No. C 2710.1A (Urban Mass Transit Administration, 1988). Until recently, the finance department estimated monthly ridership using a revenue-based model. With the gradual proliferation of special fare programs using various fare media, boarding estimates using revenue-based models were becoming increasingly suspect. In September of 2001, Tri-Met started using APC estimates of passenger activity for monthly ridership reporting.

Previously, Strathman and Hopper (1991) showed that data collection for NTD reporting using APCs necessitated a different sampling methodology because APCs are tied to vehicle blocks rather than bus trips. A multistage cluster sample was proposed to identify the minimum number of randomly selected vehicle blocks needed to meet NTD reporting requirements. The cluster sampling approach was proposed because APC deployment was limited at the time. Due to widespread deployment of APCs at Tri-Met, it is now possible to develop a trip-based sampling plan for both NTD and monthly ridership reporting.

NTD reporting uses random sampling procedures to generate annual estimates of passenger activity with a precision of +/- 10% at the 95% level of confidence. The necessary sample size depends upon (1) the desired level of precision, (2) the level of statistical confidence, (3) the inherent variation in passenger activity, and (4) frequency of data collection. We recommend using a precision of +/- 5% with a confidence level of 95% in accordance with standard statistical practice. The target number of trips (N) is defined in Equation 8. The two quantities necessary to calculate minimum sample size are the mean and standard deviation of boardings per hour. The two parameter estimates are determined on a per month basis by day type.

$$N = \left(1.96 * \sigma_{ONS} / 0.05 * \overline{X}_{ONS}\right)^2 \qquad \text{[Equation 8]}$$

Table 3.1 shows the minimum sample size needed for ridership reporting based upon the most variable month of service for each day type. The results show that a minimum sample size of 817 trips is needed for ridership reporting. For purposes of simplicity, we estimate monthly ridership by sampling 1,000 trips from the schedule.

Table 3.1: Sample Size Determination Based Upon Most Variable Month of Service

Day Type	Year /Month	Mean ONS/Hour	Std. Dev. ONS/Hour	Sample Size
Weekday	2001/08	40.42	29.48	817
Saturday	2001/07	37.83	25.71	710
Sunday	2001/08	37.43	25.65	721

A Monte Carlo technique, characterized by repeated random samples, was used to generate a distribution of passenger boardings with minimal sampling error. Increasing sample size is a valid technique for reducing sampling error under the assumption of a normal population distribution. If the population distribution is not normal, then increasing sample size does little to reduce sampling error. Data generated from APCs are not normally distributed because certain trips are likely to be over or underrepresented in the sample. This is because trips are tied to vehicle blocks which may or may not be equipped with APCs. However, we know from the central limit theorem that repeated samples from a non-normal population will produce a mean value for a given variable similar to that from a normal population distribution.

For monthly ridership reporting, 300 samples of 1,000 trips were selected randomly from the schedule and used to generate distributions of passenger activity. An algorithm was developed that matched each sampled bus trip with valid APC data. To control for any changes in scheduled service during the month, 250 trips were sampled each week. Thus, the maximum number of times a single trip could be sampled in a particular month was four. Each instance of a particular trip in the schedule was assigned a random number to ensure that a trip occurring in any given week had an equal probability of being selected. In the event that a successful match was not found at any time during the month, then the closest available adjacent trip in the sequence of trips was used as a substitute. The closest available adjacent trip was determined by taking the absolute value of the difference between adjacent trip sequence numbers, then selecting the trip with the lowest absolute value difference. If the absolute value difference between adjacent trip numbers was the same, then the trip with the lowest random number was selected as the substitute trip. On average, our method allowed us to directly match 940 trips with valid APC data on the first attempt, with 51 additional trips selected through trip substitution.

Because our method is based upon repeated sampling, the variable of interest is the grand mean of the estimated number of boardings over 300 replications for each day type. The first step was to estimate the mean number of boardings per hour for each replication of the sample (Equation

9). The equation is simply the ratio of total boardings to total actual revenue time for the 1,000 sampled trips. To arrive at the estimated number of boardings per month for a single replication of the sample, the average number of boardings per hour was multiplied by total hours of scheduled service for a given month (Equation 10). The last step involved estimating the grand mean over 300 replications (Equation 11).

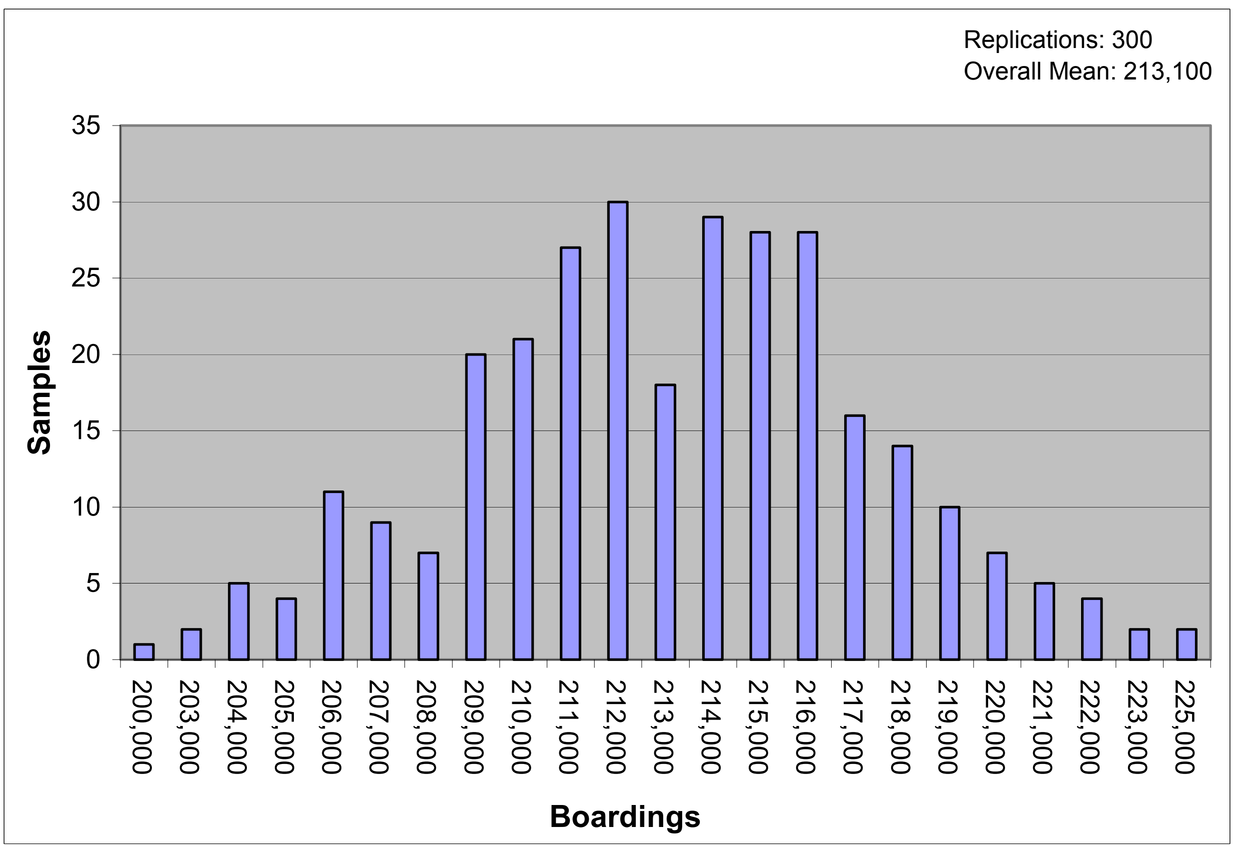
$$\overline{ONS}_{HOUR} = \sum_{i=1}^{1000} ONS_i \Big/ \sum_{i=1}^{1000} REV\ HR_i \qquad \text{[Equation 9]}$$

$$ONS_{MONTH} = \overline{ONS}_{HOUR} * SCHED\ HR_{MONTH} \qquad \text{[Equation 10]}$$

$$\overline{\overline{ONS}}_{MONTH} = \frac{1}{300} * \sum_{j=1}^{300} ONS_{MONTH} \qquad \text{[Equation 11]}$$

Figure 3.2 shows the results of the Monte Carlo simulation for estimating weekday boardings for September, 2001. Figures for Saturday and Sunday boardings are presented in Appendix B.

Figure 3.2: Sampling Results– Weekday Boardings (September, 2001)



The average number of boardings over 300 replications is 213,100 passengers. The range between the lowest and highest estimated number of boardings on weekdays is 25,000 persons.

The figure shows how the use of repeated sampling techniques can produce results that approximate a normal distribution.

Table 3.3 compares monthly ridership estimates from the revenue-based model with those derived from APCs using the repeated sampling technique presented in this report. The data show that estimation of monthly ridership using APCs leads to an average increase in the number of weekday boardings of 1.7% over the revenue-based model, a decrease in Saturday boardings of 2.1% and an increase in Sunday boardings of 6.2% over a 15-month period. Although the two methodologies used to estimate ridership are vastly different, with one based on financial information and the other on archived bus operations data, the results are remarkably similar.

Table 3.3. Comparison of Boardings Estimates Using APCs and a Revenue-Based Model

	Weekdays			*Saturdays*			*Sundays*		
Date	*APC*	*Revenue*	*% Diff*	*APC*	*Revenue*	*% Diff*	*APC*	*Revenue*	*% Diff*
2000/06	208,800	207,000	0.9%	108,600	113,300	-4.1%	71,100	55,000	29.3%
2000/07	203,200	191,600	6.1%	106,200	108,900	-2.5%	70,400	65,900	6.8%
2000/08	199,700	195,400	2.2%	105,300	110,000	-4.3%	72,500	60,200	20.4%
2000/09	209,300	198,300	5.5%	105,300	101,600	3.6%	70,600	64,400	9.6%
2000/10	215,900	209,900	2.9%	109,000	111,900	-2.6%	70,400	66,300	6.2%
2000/11	210,200	212,600	-1.1%	103,300	109,900	-6.0%	66,400	70,000	-5.1%
2000/12	194,700	199,400	-2.4%	102,500	106,700	-3.9%	68,200	62,600	8.9%
2001/01	207,000	200,500	3.2%	102,400	103,100	-0.7%	64,500	69,300	-6.9%
2001/02	211,400	215,400	-1.9%	103,400	111,100	-6.9%	68,500	71,100	-3.7%
2001/03	207,600	210,300	-1.3%	105,400	105,400	0.0%	67,500	69,600	-3.0%
2001/04	212,800	212,800	0.0%	105,600	103,900	1.6%	66,900	67,200	-0.4%
2001/05	218,900	215,400	1.6%	111,700	108,000	3.4%	73,800	67,200	9.8%
2001/06	213,800	214,600	-0.4%	112,400	115,700	-2.9%	73,100	67,600	8.1%
2001/07	208,000	196,900	5.6%	106,700	112,700	-5.3%	72,700	69,500	4.6%
2001/08	203,300	192,800	5.4%	107,700	107,100	0.6%	73,900	63,100	17.1%
Mean	208,307	204,860	1.7%	106,367	108,620	-2.1%	70,033	65,933	6.2%

For annual NTD reporting, the sampling procedure was modified somewhat because repeated random sampling is not necessary. This is because internal monthly ridership reporting at Tri-Met requires more frequent sampling at a higher precision than NTD reporting. For NTD reporting, we randomly selected 100 trips per month from the schedule for a total sample of 1,200 observations. The values of interest for NTD reporting are unlinked passenger trips (boardings) and passenger miles of service for the average weekday, Saturday and Sunday, as well as annual totals. The first step involved calculating the mean number of boardings per revenue hour for the 1,200 sampled trips (Equation 12). This value was then multiplied by the average hours of scheduled service for each day type to arrive at the mean number of boardings for each day type (Equation 13). Total annual boardings were determined by multiplying the average number of boardings for each day type by the number of respective service days in a given year, then summing over all day types (Equation 14).

$$\overline{ONS}_{REV\ HR} = \sum_{i=1}^{1200} ONS_i \Big/ \sum_{i=1}^{1200} REV\ HR_i \qquad \text{[Equation 12]}$$

$$\overline{ONS}_{DAY\ TYPE} = \overline{ONS}_{REV\ HR} * \overline{SCHED\ HR}_{DAY\ TYPE}$$ [Equation 13]

$$ONS_{YEAR} = \sum_{DAY\ TYPE}^{3} \left(\overline{ONS}_{DAY\ TYPE} * DAYS_{DAY\ TYPE} \right)$$ [Equation 14]

The formulas for determining average and annual passenger miles of service are similar to the ones presented above except that passenger miles are substituted for boardings (Equations 15-17).

$$\overline{P\ MILES}_{REV\ HR} = \sum_{i=1}^{1200} P\ MILES_i \Big/ \sum_{i=1}^{1200} REV\ HR_i$$ [Equation 15]

$$\overline{P\ MILES}_{DAY\ TYPE} = \overline{P\ MILES}_{REV\ HR} * \overline{SCHED\ HR}_{DAY\ TYPE}$$ [Equation 16]

$$P\ MILES_{YEAR} = \sum_{DAY\ TYPE}^{3} \left(\overline{P\ MILES}_{DAY\ TYPE} * DAYS_{DAY\ TYPE} \right)$$ [Equation 17]

The numerator in Equation 15 is determined by multiplying the passenger load by the distance between stops for the sampled trips and then summing over all trips. The load calibration factor determined previously should be applied at this stage.

A comparison of ridership estimates for NTD reporting using both manual and APC data collection methods is presented in Table 3.4. To test the sampling methodology, we generated 2001 APC estimates based upon a random sample of 100 trips per month for the last 6 months of the year (596 total trips).

Table 3.4 Comparison of NTD Estimates by Estimation Method

Year (Method)/Measure	*1999 (NTD)*	*2000 (NTD)*	*2001 (APC)*
Unlinked Passenger Trips (in 000's)	58,926.1	61,818.8	63,236.3
Passenger Miles (in 000's)	206,844.1	207,760.5	208,030.8
Avg. Trip Length (miles)	3.51	3.36	3.29

The data show that the number of unlinked passenger trips and passenger miles of service are consistent with trends from previous years.

IV. Conclusions

The present study addresses the accuracy and precision of APC data at Tri-Met. APC-based sampling plans for internal monthly ridership reporting and annual NTD reporting are presented. The sampling plan for monthly ridership reporting is based on a repeated sampling technique that addresses the problem of sampling from a non-normal population. The sampling plans for both NTD reporting and monthly ridership reporting are based on a random selection of bus trips from the schedule that are matched with valid APC data using an innovative search routine. The difference in means tests comparing APC counts to camera counts showed that APC estimates of boardings are accurate at the system level. When the sample was differentiated by bus type, it was found that APCs on NFIs count boardings accurately but that APCs tended to overestimate boardings on Gilligs by a statistically significant margin. This finding is particularly relevant for quarterly route performance reporting where many bus routes operate Gilligs almost exclusively. The study also found that APCs overestimate passenger loads by a statistically significant margin, irrespective of bus type. A correction factor was developed for passenger load estimates.

Although measurement error was shown to exist with APCs, the problem is relatively minor compared to error associated with manual data collection techniques. Likewise, the problem of not having all APC-equipped bus fleets represented in the study sample due to lack of video surveillance capabilities is offset by the problem of basing accuracy tests on manually collected data of questionable precision. The older, APC-equipped bus fleets are gradually being phased out as new vehicle purchases are made at Tri-Met. The APC accuracy analysis showed that the criteria used to screen APC data at Tri-Met during post-processing is sufficient. Although we could not test for the effects of moving from a 10% accuracy cutoff value to values greater than 5%, we found that dropping the accuracy cutoff value below 5% produced negligible improvements in accuracy.

The tendency for the load balancing algorithm to count high due to the propagation of errors by not allowing negative passenger loads should be corrected. It is recommended that the new algorithm zero out loads at locations with layovers greater than 5 minutes. Ideally, the new algorithm would balance boardings and alightings at the vehicle block level first, then at the trip level so as to minimize the propagation of errors over time. Examination of the data presented in the quarterly route performance reports identified a problem where appropriate T-values were not being used when constructing confidence intervals for cases with less than 40 observations. It is recommended that the data quality flag be applied to estimates of passenger activity at each summary level, not just the route by trip summary level. Furthermore, Tri-Met might look at the possibility of tightening up the criteria used to flag suspect passenger activity estimates given the large number of APC-equipped buses deployed throughout the system.

One of the main benefits of widespread APC deployment at Tri-Met is that sufficient data are generated allowing for ex post facto sampling of scheduled bus trips for ridership reporting purposes. Because the agency no longer needs to assign APC-equipped vehicles to specific bus trips for data collection purposes, an important communication problem between operations personnel and garage managers over vehicle assignment has been effectively eliminated. A critical mass has been reached with respect to the number of APCs deployed throughout the system allowing the agency to take advantage of the technology in new and innovative ways. Tri-Met is presently able to use a consistent data source (APCs) for all of the major ridership reporting tasks within the agency. The agency will realize considerable cost savings as ridership reporting is largely automated through database queries on archived APC data and also because of a diminished need for sending ride checkers out in the field to collect passenger activity information.

Bibliography

Attanucci, J. & Vozzolo, D. (1983). Assessment of operational effectiveness, accuracy, and costs of automatic passenger counters. Transportation Research Record, 947, 15-23.

Boyle, D. B. (1998). Passenger counting technologies and procedures. TCRP Synthesis of Transit Practice 29. Washington, DC: Transportation Research Board.

Casey, R. F., Labell, L. N., Holmstrom, R., LoVecchio, J. A., Schweiger, C. L., & Sheehan, T. (1996). Advanced Public Transportation Systems: The State of the Art Update '96 (Report No. FTA-MA-26-7007-96-1). Washington, DC: Federal Transit Administration.

Greneker, E., Johnson, B., Murphy, K., & Rausch, E. (1996). Improved passenger counter and classifier system for transit operations. Transit-IDEA Program Project Final Report. Washington, DC: Transportation Research Board.

Hobbs, V. J. & Poirier, P. J. (1983). Automated Passenger Counting Systems: A Synopsis of Working Group Meeting (Report No. DOT-TSC-UMTA-83-23). Transportation Systems Center, Cambridge: MA.

Hodges, C. C. (1985). Automatic Passenger Counter Systems: The State of the Practice (Report No. DOT-1-87-36). Washington, DC: Urban Mass Transportation Administration.

Koffman, J. (August, 1990). Automatic Passenger Counting Data: Better Schedules Improve On-Time Performance. Paper presented at the Fifth Workshop on Computer-Aided Scheduling of Public Transport, Montreal.

Multisystems, Inc. (1982). An Assessment of Automatic Passenger Counters. (Report No. DOT-I-82-43). Washington, DC: Urban Mass Transit Administration.

National Institute of Standards and Technology (1998). Handbook 44: Specifications, Tolerances, and Other Technical Requirements for Weighing and Measuring Devices. Washington, DC: GPO.

Strathman, J. G., & Hopper, J. R. (1991). Evaluation of automatic passenger counters: Validation, sampling and statistical inference. Transportation Research Record, 1308, 69-77.

Strathman, J.G., Kimpel, T. J., Dueker, K. J., Gerhart, R., Turner, K. Griffin, D., & Callas, S. (2001). Bus transit operations control: Review and an experiment involving Tri-Met's automated bus dispatch system. Journal of Public Transportation, 4(1), 1-26.

Urban Mass Transit Administration (1988). Sampling Procedures for Obtaining Fixed Route Bus Operating Data Required Under the Section 15 Reporting System (Circular No. UMTA C 2710.1A). Washington, DC: Author.

Wonnacott, R. J., & Wonnacott, T. H. (1970). Econometrics. John Wiley and Sons: New York.

Appendix A: Descriptive Statistics

Full Sample

Name	*N*	*Mean*	*St. Dev.*	*Var.*	*Min.*	*Max.*
ON	2,921	1.4783	2.1283	4.5297	0	23
ONS	2,921	1.4913	2.2103	4.8856	0	24
OFF	2,921	1.3252	2.0004	4.0017	0	24
OFFS	2,921	1.3495	2.0238	4.0959	0	26
LOAD	2,921	14.0350	9.3561	87.5380	0	53
LOADS	2,921	14.8890	9.8933	97.8780	-5	53
DON	2,921	0.0130	0.5206	0.2711	-7	7
DOFF	2,921	0.0243	0.5219	0.2724	-9	8
DLOAD	2,921	0.8545	2.9381	8.6326	-9	13
GIL	2,921	0.3954	0.4890	0.2391	0	1

Gillig Sample

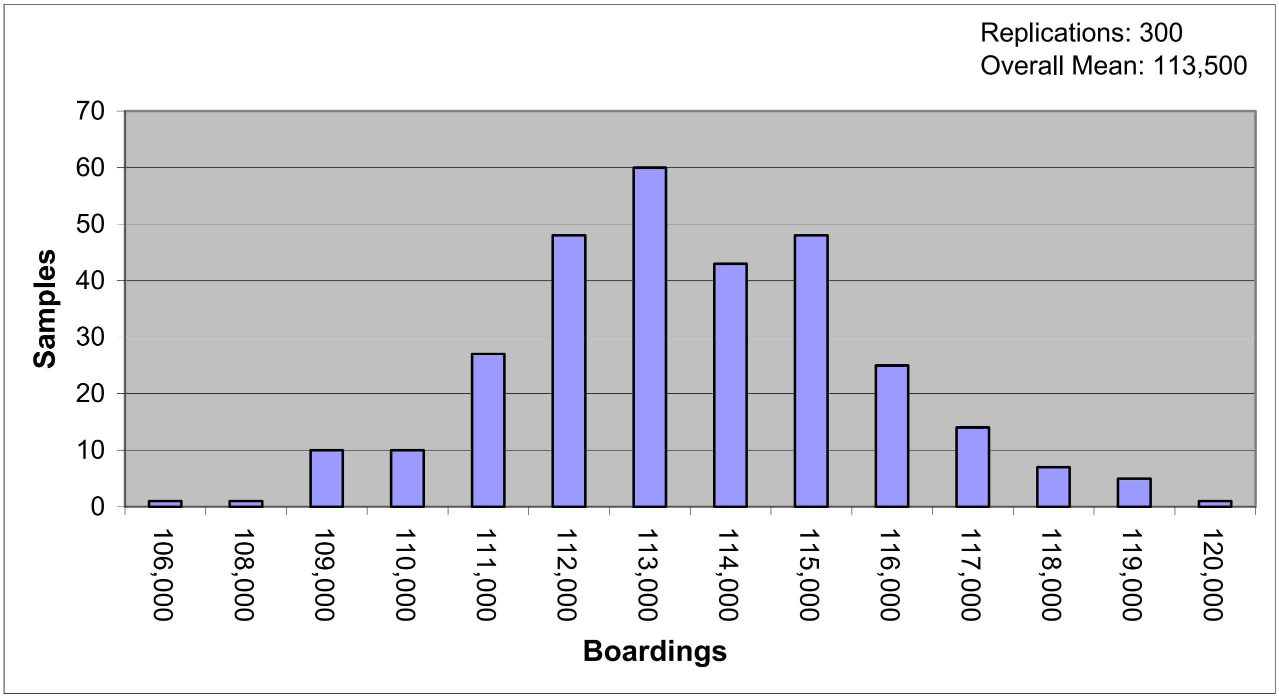
Name	*N*	*Mean*	*St. Dev.*	*Var.*	*Min.*	*Max.*
ON	1,155	1.3688	2.0341	4.1377	0	16
ONS	1,155	1.4268	2.1500	4.6227	0	19
OFF	1,155	1.2892	1.7842	3.1832	0	16
OFFS	1,155	1.3584	1.9185	3.6808	0	16
LOAD	1,155	12.9780	8.1880	67.0440	0	40
LOADS	1,155	14.2530	8.8039	77.5080	-5	41
DON	1,155	0.0580	0.4436	0.1968	-2	4
DOFF	1,155	0.0693	0.5558	0.3089	-4	8
DLOAD	1,155	1.2745	3.1532	9.9428	-9	12

NFI Sample

Name	*N*	*Mean*	*St. Dev.*	*Var.*	*Min.*	*Max.*
ON	1,766	1.5498	2.1853	4.7757	0	23
ONS	1,766	1.5334	2.2485	5.0558	0	24
OFF	1,766	1.3488	2.1302	4.5378	0	24
OFFS	1,766	1.3437	2.0904	4.3696	0	26
LOAD	1,766	14.7250	9.9889	99.7790	0	53
LOADS	1,766	15.3050	10.5270	110.8100	-5	53
DON	1,766	-0.0164	0.5635	0.3176	-7	7
DOFF	1,766	-0.0051	0.4964	0.2464	-9	7
DLOAD	1,766	0.57984	2.755	7.5899	-6	13

Appendix B: Sampling Results for Monthly Ridership Reporting

Sampling Results– Saturday Boardings (September, 2001)



Sampling Results– Sunday Boardings (September, 2001)