2. SURVEY METHODS AND RESULTS APPLICATION

This chapter summarizes the 2003 WSF South Sound Travel Survey design process, methodology, administration and response statistics. The survey serves as an update to the 1999 WSF Travel Survey for five routes at the south end of the system (Point Defiance-Tahlequah, Fauntleroy-Vashon, Southworth-Vashon, Seattle-Vashon passenger-only, and Fauntleroy-Southworth). The context of the survey data and subsequent analysis is framed in terms of response rates, usable survey records, survey expansion, response precision estimates, and other applications of the results.

2.1 SURVEY ADMINISTRATION AND CODING

This section summarizes the sampling plan, administration, and coding of the WSF travel survey. Readers seeking additional information regarding these topics are referred to the *WSF South Sound Ferry Routes Travel Survey: Technical Report of Methods* dated February 2004 and prepared by NuStats for Washington State Ferries as part of this study.

Survey Periods and Sampling Plan

The survey sampling plan called for administering a travel survey to a sample of weekday during October 2003. The sampling plan was developed to obtain travel information from users during the weekday PM peak period and the remaining non-peak PM hours of the day (PM non-peak period) for each of the five routes. For the PM peak period, survey questionnaires were offered to all persons age 12 or older on every vessel sailing departing between 3 and 7 PM on each route was surveyed.[1] PM non-peak period riders were surveyed on the same day as the PM peak period survey from a sample of approximately one-half of the vessel sailings that occurred during the non-peak PM hours. When appropriately expanded, the two weekday survey periods combined represent the travel patterns for the PM half day ridership.

Survey Administration

The 2003 WSF Travel Survey was administered as planned over two days: Tuesday, October 14 for Point Defiance-Tahlequah and Wednesday, October 15 for the remaining south sound routes. Boarding and alighting counts were completed by vessel at each terminal in order to produce ridership control totals, especially in the case where one vessel serves three terminals (Fauntleroy, Vashon and Southworth).

[1] Age determination was somewhat at the discretion of the survey workers; the cutoff age used in 1999 was 15 years. A few exceptions to the PM peak period definition were made to include vessel sailings that were within the 3-7 PM window in 1999 but had been subsequently rescheduled to be just outside this period.

2.2 SURVEY RESPONSE TRENDS

Approximately 4,270 questionnaires were distributed system-wide to weekday passengers on the sampled vessel sailings, and just over 2,000 of these were returned and tallied.[2] Computerized scanning of survey questionnaires was employed for all survey responses except for the reported addresses for the rider's trip origin, trip destination, and home location, which were manually entered for subsequent latitude/longitude geocoding. After data processing and an iterative, thorough quality review, there were 1,818 general usable survey records for the PM peak and PM non-peak survey periods, and 1,809 trip table usable records.

Usable Survey Records

For consistency with the 1999 Travel Survey, survey records deemed usable for tabulation and analysis were divided into two categories: ***general usable*** and ***trip table usable***. A general usable record is defined as survey data for one respondent that is judged complete and valid for analysis purposes. To meet the general usable criteria, a survey record must include:

- Indication of the trip purpose;
- Indication (or observation) of the boarding method as in-vehicle or walk-on, and if the latter, further indication of the access mode to the departure ferry terminal and the egress mode from the arrival ferry terminal; ***and***
- Geocodable address information for the respondent's trip origin and destination that are either:
  - Sufficiently dissimilar in geography to be consistent with a one-way trip on the ferry crossing surveyed; ***or***
  - Geographically similar but with a reported trip purpose of sightseeing, indicating that the respondent made a continuous round-trip, beginning and ending at the same ferry terminal.

Because the survey questionnaire was designed to gather data about a one-way trip, it was necessary to exclude those records for the small number of respondents who erroneously provided information about a round-trip from the general usable category. One exception was made for respondents who indicated that they were making a continuous round-trip with the sole purpose of "sightseeing", effectively boarding and alighting at the same ferry terminal. For these cases, the ferry component of a respondent's one-way trip and round-trip are indistinguishable.

A further restriction was applied to yield survey records categorized as trip table usable. This involved excluding the nine general usable survey records that reported round-trip travel with the trip purpose of "sightseeing".

---

[2] Approximately 3% of these were returned by mail via a business reply mail permit.

Response and Usable Rates

Table 2-1 presents the survey response statistics and usable rates for both sets of usable records by survey period. Response rates are measured by dividing the number usable records by category into the relevant total number of surveys distributed. Usable rates are calculated by dividing the number of usable records (by category) into the relevant number of surveys that were returned and coded.

In most cases regarding queries and tabulations of general ferry use, it is appropriate to use the set of general usable records. For the two survey periods, when expanded using the supplied expansion factors, it is this set of data that gives the expected ridership by the PM peak and PM non-peak periods. There are a few applications, such as mapping one-way trip origins and destinations, where it may be appropriate to use the slightly reduced subset of trip table usable records that excludes the sightseeing continuous round-trip travel patterns.

Table 2-1
Survey Distribution, Response and Usable Record Statistics

Survey Category	*October 2003 South Sound Travel Survey*		
	Total Survey (PM ½ Day)	*Weekday PM Peak*	*Weekday PM Non-Peak*
Surveys Distributed	4,270	3,173	1,097
Surveys Returned and Coded	2,021	1,575	446
Less: Missing Required Data Fields	(90)	(78)	(12)
Trip O & D too Similar / Incomplete / Inconsistent	(113)	(79)	(34)
Usable Records for General Tabulation	1,818	1,418	400
Less: Usable w/ Trip Purpose Sightseeing & Similar O & D	(9)	(3)	(6)
Usable Records for Trip Table Analysis (One-Way Trip Origins and Destinations)	1,809	1,415	394
Response Rates (% of Distributed Surveys)			
General Usable Response Rate	42.6%	44.7%	36.5%
Trip Table Usable Response Rate	42.4%	44.6%	35.9%
Usable Rates (% of Returned & Coded Surveys)			
General Usable Record Rate	90.0%	90.0%	89.7%
Trip Table Usable Record Rate	89.5%	89.8%	88.3%

Expansion Factors

Basic tabulations of the general usable survey records provide results that represent those who participated in the survey. However, for each route, this group represents a subset of the population of all riders within the survey period on the survey day. In order to present the survey results as representative of the survey period population of riders, it is necessary to expand the usable numbers to reflect the actual ridership that occurred by various subcategories on the respective survey days. Thus, expansion factors are calculated to adjust the usable responses back up to the actual survey period ridership — as if all ferry

riders had actually completed the survey and all completed surveys were perfectly usable. Expanding the usable survey responses requires accurate ridership counts by route, vessel sailing and direction, and boarding method to provide the correct expansion targets.

Weekday Survey Period Expansion Issues

The weekday PM peak and PM non-peak survey results have been expanded to be representative of the actual ridership during these periods using boarding and/or alighting counts by vessel sailing. Boarding counts (and alighting counts where necessary for vessels serving multiple destinations) were conducted as part of the fielding process during the survey periods, with separate counts by boarding method (in-vehicle and walk-on.) In general, boarding counts prove to be an accurate source for ridership data, though in some cases, it was difficult to count all vehicle passengers, particularly for vehicles arriving at the last possible moment and boarding without stopping. Similarly, there were cases on the larger routes where the rush of walk-on passenger volumes just prior to departure may have exceeded the ability of survey workers to accurately count boardings.

> ***"The weekday PM peak and PM non-peak survey results have been expanded to be representative of the actual ridership during these periods…."***

PM Peak Period Expansion Factors

The PM peak period was defined as the four hours between 3 and 7 PM, and surveys were typically distributed to every rider age 12 and older on all routes for vessel sailings scheduled to begin within this interval. Therefore, the PM peak period expansion factors were calculated by dividing the number of usable responses for each combination of route, direction and boarding mode into its associated survey period ridership.

Overall, the 1,418 general usable PM peak survey records correspond to an expanded weekday PM peak survey period ridership base of 4,383 ferry users, which corresponds to an average expansion factor of 3.09.

PM Non-Peak Period Expansion Factors

The PM non-peak period generally represents the non-peak PM hours of the day — those PM hours outside of the 3-7 PM peak window. Although questionnaires were distributed to every rider age 12 and older on all routes for the vessel sailings surveyed, only about one-half of the PM non-peak vessel sailings were surveyed, and ridership counts were only conducted for surveyed sailings. As such, the expansion factors for the PM non-peak period generally included a response factor component and a boarding factor component. The response factors expand the usable survey records to the actual ridership on all of the sampled sailings by route, direction and boarding mode. The boarding factor then expands the number of sampled sailings by route to the total number of vessel sailings in the PM non-peak period.

This method assumes that the vessel sailings excluded from the sample, have, on average, the same ridership characteristics and overall usage as those included in the survey. More

accurate survey expansion results could have been obtained for this period by conducting ridership counts for all of the PM non-peak sailings, or better yet, to have both counted and surveyed all of the PM non-peak sailings. These alternatives were not undertaken due to budgetary constraints.

Overall, the 400 usable survey records obtained correspond to an expanded weekday PM non-peak survey period ridership base estimated at 2,528 ferry users, including the estimates for the non-sampled sailings. This results in an average expansion factor, inclusive of both the response and boarding factors, of 6.32.

2.3 SURVEY PRECISION ESTIMATES

Given the survey response trends, this section sheds light on the accuracy or precision by which the survey sample results are representative of the true underlying population of ferry riders. It is divided into sections covering a technical discussion of sample precision and typical precision levels achieved by route.

Technical Discussion

Precision Levels, Confidence Intervals, and Population Parameters

It is useful to define several terms that are key to this discussion. The first is precision, which refers to close the survey sample represents the characteristics of the underlying population. The population in this case is the universe of all ferry travelers during the defined survey period on the day in which the survey was conducted. While it is likely that the sample is also representative of a broader population of ferry riders beyond the survey period and day, especially since many riders are regular users and the survey month of May is the month most representative of annual ridership, precision levels can only be estimated in reference to the daily survey period population from which survey respondents were obtained. In addition, one must consider the confidence level of the precision estimates. A confidence interval of 95% implies that with repeated sampling of survey riders, 95 out of 100 samples will give sample estimates of population parameters that are within the specified precision range. In other words, the precision of a sample estimate of some underlying population characteristic or parameter is a function of the population itself, the prescribed confidence level, and the size of the sample. A tightening of the desired precision level and/or increasing the confidence interval will require a larger sample.

The WSF 2003 South Sound Travel Survey covered multiple data collection objectives represented by the different questions or question components, each of which may require a different sample size to achieve a desired precision level at an acceptable confidence interval. For example, some questions serve to estimate population proportions — the percentage distribution of ferry users across a set of categories (e.g., distributions of trip purpose, type of ticket purchased, transit improvements desired, etc.), and some serve to estimate population means or averages (e.g., minutes waited for ferry, number of one-way rides in the last week, amount paid for parking, people in vehicle, etc.) Furthermore, some questions seeking population means which are later transformed to category ranges, such as respondent age, then require population proportion precision estimates. It is important to

consider the results for which precision levels are desired, as the formulas for estimating the sample precision are different when estimating population proportions than for population means.

Absolute versus Relative Precision Estimates

Finally, there are two types of precision levels to consider: absolute precision and relative precision. Consider hypothetical survey results for trip purpose, where 20% of respondents indicated a trip purpose of work/school. Absolute precision levels consider a range about the sample estimate measured in the units of the parameter being estimated. Thus an absolute precision level of ±5%, at a 95% confidence level, would mean that with 95% certainty in this sample, the true share of work/school trip purposes lies between 15% and 25%. Conversely, relative precision levels measure the percentage deviation about the sample estimate. Thus a relative precision level of ±5%, at a 95% confidence level, implies that with 95% certainty in this sample, the true share of work/school trip purposes lies between 19% and 21%. In this case, a relative precision level of ±5% would require a larger sample than an absolute precision level, since it provides an overall lower margin of error.

Route Level Precision Estimates

Table 2-2 presents the effective absolute precision levels for population proportion estimates by route for the PM peak period. These precision values apply to simple proportion distribution tabulations (e.g., trip purpose) for each route across all boarding methods in both directions. In general, the determination of the precision level achieved for a survey sample estimate representing the survey ridership population really requires an application-specific calculation. For instance, the results of a particular cross-tabulation of two survey questions or a single question by boarding method each represent unique subsets of the ridership population and sampled survey data. Given a statistical sample and the type of estimate it represents (population proportion or population mean), an absolute and/or relative precision range can then be uniquely estimated. Since all cases are different, it is not possible to provide all of the hundreds of combinations of precision levels for the various routes, directions, boarding modes, survey periods, etc.

Note that since results are tabulated for travel between Southworth and Seattle as if a unique route served these terminals, responses for riders making the transfer at Vashon were extracted from either the Seattle-Vashon passenger-only or the Southworth-Vashon responses, and precision estimates were reported separately. This has the effect of reducing the precision of the results obtained on the Seattle-Vashon and Southworth-Vashon routes, but the results, including those for "Seattle-Southworth" travel, more closely match how the data is analyzed.

Table 2-2
Survey Distribution, Response and Precision Statistics — Weekday PM Peak Period

Route		PM Peak Survey Period Ridership	PM Peak Surveys Distributed	PM Peak Returned Surveys	PM Peak Usable Surveys[1]	Usable Sample Share of PM Peak Ridership	Sample Req'd for Pop. Proportion ± 5% Absolute Precision Range[2]	Expected Absolute Precision Range for Estimating Pop. Proportions
Pt. Defiance–Tahlequah		806	686	427	351	44%	260	± 3.9%
Fauntleroy–Vashon		1,721	2,111[3]	429	388	23%	314	± 4.4%
Fauntleroy–Southworth		1,090		390	377	35%	284	± 4.1%
Southworth–Vashon	*Conventional Trips*	*294*		*81*	*69*	*23%*	*167*	*± 10.3%*
	Transfers to Seattle-Vashon	*75*		*15*	*10*	*13%*	*N/A*	*N/A*
	Total Trips	369		96	79	21%	188	± 9.8%
Seattle–Vashon PO	*Conventional Trips*	*267*		*153*	*149*	*56%*	*158*	*± 5.3%*
	Transfers to Southworth-Vashon	*129*		*80*	*74*	*57%*	*N/A*	*N/A*
	Total Trips	396	376	233	223	56%	195	± 4.3%
Seattle–Southworth (*Combined Transfers*)*		204	N/A	95	84	41%	133	± 8.2%
Overall South Sound Routes*		4,382	3,173	1,575	1,418	32%	353	± 2.1%

Footnotes

[1] Returned surveys meeting completeness criteria: "good" origin & destination information, plus data for trip purpose, boarding method, access & egress modes.

[2] Absolute Precision estimated were adjusted by the finite population correction factor on all routes.

[3] Totals for three routes due to shared vessels (vessel departures from each terminal may serve two destinations or routes).

* Measures travel of riders on both the Seattle-Vashon & Southworth Vashon routes via a transfer on Vashon. PM peak period ridership values are estimates, and South Sound totals may include some double counting of travelers between Seattle and Southworth.

Comments

• Sample sizes required for absolute or relative precision when estimating population means (averages) will vary with the range of responses to each question.

On five of the six routes, the absolute precision level achieved for the PM peak period survey was less than ±5%, though Seattle-Vashon slightly exceeded ±5%after separating out the responses for those transferring to/from the Southworth-Vashon route. The precision level for Southworth-Vashon was less robust, as this route has both a small ridership population and is short in duration, the latter contributing to a low response rate. This problem is somewhat exacerbated by removing the responses for those transferring on Vashon to the passenger-only ferry in order to examine this travel separately. This "seventh route" exhibited an absolute precision level of ±8.1%.

In general, the precision of a single day survey can only be improved by increasing the response rate or selecting a survey day with a higher level of expected ridership. Sampling on more than one day to improve precision levels creates another set of statistical problems due to sampling with replacement (sampling some riders more than once.)

Sample Share of Survey Population

Table 2-2 also presents the general usable response sample share of the total PM peak survey period ridership (survey population) for each route. For all South Sound routes as a whole, the usable survey sample — which was restricted to riders age 12 and older — represents 32% of the total ridership during the 3-7 PM peak period survey window. Put another way, 32% of all PM peak period riders completed a valid, usable questionnaire. This ridership-based response rate is lower than the rate shown in Table 2-1, which is based on the number of surveys successfully distributed.

Regardless of measurement techniques, the sample shares or response rates realized for the WSF 2003 South Sound Travel Survey, especially during the PM peak period, are above average for voluntary travel surveys.

2.4 WSF GEOGRAPHIC ZONE SYSTEM

This section describes the geographic information system (GIS) elements of the 1999 WSF Travel Survey, including the geocoding of addresses to latitude-longitude (X-Y) coordinates and the revisions made to the existing WSF Transportation Analysis Zone (TAZ) system.

Geocoding Process and GIS Application

Home, origin, and destination addresses reported by survey respondents were geocoded to latitude-longitude (X-Y) coordinates in the ArcView GIS software application. Utilizing street coverage files derived from 1997 U.S. Census Bureau TIGER Line files[3], addresses were geocoded through a combination of batch (automated) and interactive (one address at a time) processes. Geocoding quality control measures included extensive visual checks for location accuracy, comparison of the geocoded zip code with the zip code given by the respondent, and comparison of the given direction of travel with that indicated by the geocoded origin and destination.

WSF TAZ System and District Schemes

Transportation Analysis Zone (TAZ) System

The Transportation Analysis Zone (TAZ) system used for this study is identical to that for the WSF 1999 Travel Survey, which, in turn, is an adaptation of the existing Puget Sound Regional Council (PSRC) TAZ system. Specifically, the PSRC TAZ boundaries have been employed for counties within the PSRC region (King, Pierce, Snohomish, and Kitsap Counties). Zones external to the PSRC TAZ system in the eight outlying counties (Clallam, Jefferson, Island, Mason, San Juan, Skagit, Thurston, and Whatcom) have been sized and developed to reflect concentrations of activities (e.g., population) within each county. External TAZs were created to represent areas in Washington State but outside of the study area, other states in the U.S., and certain parts in British Columbia, Canada.

Geocoded home, origin, and destination addresses were assigned to the WSF TAZ system utilizing a spatial join in ArcView GIS. This is an automated GIS function which locates each geocoded address on the TAZ map and assigns the appropriate TAZ number to that record. For those addresses located outside of the study area, external TAZs were assigned manually.

[3] The street coverage files were indexed in ArcView 3.2 software to State Plane 1983 Washington North.

System and Route-Specific District Schemes

TAZs were aggregated into route-specific district schemes for purposes of analyzing and presenting survey results. The district schemes for the routes surveyed are identical to those used in 1999, which were sized with the intention of providing more detail closer to the route's ferry terminals. TAZ-to-district equivalency tables for each route are not included with this document, but can be found in Appendix D of the 1999 report.