

**Sample Weighting and Expansion  
Part III: Average Saturday and Sunday Weights**

**California Household Travel Survey 2012/13  
for the  
San Francisco Bay Area**

Planning Section  
Metropolitan Transportation Commission  
101 Eighth Street  
Oakland, California 94607

October 2013

**Table of Contents**

I. Introduction ..... 1  
II. Exploring Census and Survey Characteristics ..... 2  
III. Validation and Evaluation of Weighting Methods..... 3  
IV. Next Steps..... 5

**List of Tables and Figures**

Tables in Body of Report

- Table 1. CHTS 2012/13 Bay Area Sample Households by Day of Week
- Table 2. Raking Levels and Corresponding Appendix Tables
- Table 3. Range of Average Weekend Sample Weights by Weighting Model

Appendices

- Appendix A. Exploring Census and Survey Characteristics
- Appendix B. Household Level Model Validation: Models #4, #4c1, #5, #5c1

## **I. INTRODUCTION**

This working paper is the third in a series for documenting procedures and results of the year 2012/2013 California Household Travel Survey conducted in the San Francisco Bay Area (CHTS12/13). The purpose of this working paper, *Sample Weighting and Expansion: Part III: Average Weekend Weights*, is to describe procedures for weighting and expanding CHTS12/13 household and person files for the “Saturday sample” and “Sunday sample” (that is, persons providing travel data on the assigned weekend day).

Four sets of weights are envisioned for this study:

- 1) Average Daily weights (for the combined samples);
- 2) Average Weekday weights (for the Monday through Friday samples);
- 3) Average Saturday weights (for the Saturday sample); and
- 4) Average Sunday weights (for the Sunday sample).

Working papers such as this report tend to be a “work in progress” and may be updated to incorporate other improvements, clarifications and analyses. Please check with MTC to obtain the most current version of this and other working papers.

Background on “what is weighting” and “what is raking” and previous household travel surveys is covered in Part I of this three part report on sample weighting and expansion. It is suggested that the Part I report is reviewed before Part III (Saturday and Sunday weights).

The Bay Area portion of the California Household Travel Survey (2012/13) was a one-day travel/activity data from 9,719 households. Data was collected between February 1, 2012 and January 31, 2013. Of the 9,719 sample households, 8,086 provide weekday travel data; 717 provide Saturday data; and 916 households provide Sunday data.

**Table 1**  
**CHTS 2012/13 Bay Area Sample Households by Day of Week**

Day of Week	Sample Households	% of Total
Monday	775	8.0%
Tuesday	2,149	22.1%
Wednesday	2,124	21.9%
Thursday	2,160	22.2%
Friday	878	9.0%
Saturday	717	7.4%
Sunday	916	9.4%
TOTAL	9,719	100.0%
Weekday Total	8,086	83.2%
Weekend Total	1,633	16.8%
Tuesday-Thursday Total	6,433	66.2%

The proposed raking schemes for the Bay Area CHTS2012/13 SATURDAY and SUNDAY samples have six raking levels:

- 1) County (9) by Minority Status of Householder (2);
- 2) County (9) by Age of Householder (5);
- 3) County (9) by Vehicles in Household (4);
- 4) County (9) by Workers in Household (4);
- 5) County (9) by Tenure (2); and
- 6) County (9) by Household Size (5).

The small sample size of the Saturday and Sunday sample would suggest that any geographic expansion below county level, e.g., PUMA-level, would be very difficult, sparse, and complicated due to excessive need to collapse geographies to suitable levels.

## **II. EXPLORING CENSUS AND SURVEY CHARACTERISTICS**

The purpose of this section is to compare census “marginal control total” data to the corresponding patterns from the CHTS 2012/13 Saturday and Sunday samples in the Bay Area. This will highlight the critical biases in the survey that are correctable using appropriate weighting schemes.

Detailed data tables included in Appendix A (Saturday) and Appendix B (Sunday) are reported in this section. As appropriate, census data sources (Census 2010 “short form” data versus American Community Survey 2007/11) are cited.

More detailed analyses of the “combined sample” relative to Census and ACS data is included in the Part I report. This Part II report focuses on the raking levels that compare the WEEKDAY sample to the same census/ACS data.

Table 2 summarizes the raking level, and corresponding appendix table, used in developing the various raking/weight Saturday and Sunday models.

**Table 2**  
**Raking Levels and Corresponding Appendix Tables**

Table	Raking Level Description
A.8, B.8	County (9) by Minority Status (2)
A.4, B.4	County (9) by Age of Householder (5)
A.7, B.7	County (9) by Vehicles in Household (4)
A.6, B.6	County (9) by Workers in Household (4)
A.3, B.3	County (9) by Tenure (2)
A.2, B.2	County (9) by Household Size (5)

The ordering of the raking levels is not important, except for the final, last raking level. The “last rake” will show the best fit, comparing the modeled, expanded households to the “marginal control totals.” For previous and current MTC weighting approaches, the focus is on obtaining accurate estimates of households by geography by household size, in order to ensure the best approximation of total household population.

### **III. VALIDATION AND EVALUATION OF WEIGHTING METHODS**

Four weighting methods are examined in this section. Detailed data tables are included as Appendices C.

The study consultant developed a set of weights, which are denoted as “Model #0” weights in this overall study. MTC staff developed multiple sets of raking models/weights, denoted as Model #1, Model #1c, Model #2, Model #2c1, Model #2c2, Model #3, Model #3c1 and Model #3c2. The “c” stands for “constrained”.

The Model #1 and #2 series are for the “combined sample” also known as the “average daily sample.” The Model #3 series of weights are for the “weekday sample”.

Model #4 and #4c1 are for the Saturday weights.

Model #5 and #5c1 are for the Sunday weights.

Model #4 and #5 weights were constrained using floors, but no “ceilings” or caps on the weights. One set of constrained weights were produced for the Saturday and Sunday samples:

- 1) Model #4c1 (Saturday), constraining the weights to a 10.0 floor; and
- 2) Model #5c2 (Sunday), constraining the weights to a 10.0 floor.

**Table 3**  
**Range of Average Weekend Sample Weights by Weighting Model**

	Median	Mean	Minimum	Maximum
Model #4	1404.66	3637.41	0.002932	56,983.5
Model #4c1	1404.66	3638.02	10.00	56,983.6
Model #5	1552.06	2847.19	3.17E-08	35,455.0
Model #5c1	1552.05	2847.47	10.00	35,455.0

The maximum weights in the Saturday and Sunday samples are magnitudes higher than the “combined sample” and “weekday sample” households. This may be a concern when analyzing the “regional” travel characteristics on “average Saturday” and “average

Sunday” travel, so the data use may want to test the sensitivity of the characteristic being analyzed (e.g., weekend transit shares, VMT per household), by omitting the high outliers (say, weights greater than 25,000) from the analysis.

Appendix C summarizes the household level validation of Models #3c2 (weekday), #4c1 (Saturday) and #5c2 (Sunday), at the regional and county levels. Regional level results are provided in Table C.1.1 through C.1.6.

Results are also shown at the county-level (Table C.2.1) and PUMA-level (Table C.2.2). The PUMA-level tabulation indicates that the Saturday and Sunday samples should basically be analyzed and reported only at the county or regional level. Again, the sample size is too small to support sub-county level analyses.

The remaining tables in Appendix C (Tables C.3 through C.8) show the Saturday, Sunday and Weekday expanded/weighted households by county, by various socio-economic stratifications. The general findings are that the Saturday and Sunday samples are probably to support even sub-regional analyses. Travel characteristics are probably best presented at the regional totals, only.

What is not included in this analysis of weekend weights is the weighted/expanded characteristics of the household population.

The last set of information provided are the “person correction factors” used to correct for the under-estimation of the very large households in the sample survey. For the “combined sample” and “weekday sample”, person correction factors were calculated at the PUMA-level. For the weekend weights, regional level person correction factors were prepared:

- |  |            |
|--|------------|
| a) Saturday, Model #4, PCF for Households, 4+:   | 1.07078431 |
| b) Saturday, Model #4c1, PCF for Households, 4+: | 1.07078975 |
| c) Sunday, Model #5, PCF for Households, 5+:     | 1.06721082 |
| d) Sunday, Model #5c1, PCF for Households, 5+:   | 1.06721791 |

The Saturday “person correction factors” are applied for all persons-in-households of household size four-or-more. This is needed since there are no Saturday households of 5+ persons in San Francisco County. So, the correction factors are applied to four-person and five-or-more-person households.

Some Saturday and Sunday results are included in the Part II report on weekday weights. Part II, Appendix C, Tables C.12 through C.16, examines average (mean) characteristics by market segment: average household size, average workers per household, average students per household, and the average age of persons in household. The rates are shown for the consultant-provided “Model #0”, Saturday (Model #4c1), Sunday (Model #5c1) and the three weekday models (#3, #3c1, #3c2).

#### **IV. NEXT STEPS**

The recommendation is that Model #4c1 (Saturday) and Model #5c1 (Sunday) weights are the final MTC weights on the Bay Area WEEKEND sample households in the CHTS 2012/13 database. This is for the “Saturday” sample (N=717 sample households) and the “Sunday” sample (N=916 sample households).

It is envisioned that the “combined sample weights” will be used when estimating auto ownership level models, and other analyses focusing on the demographics of the household. For purposes of estimating aggregate “average weekday” travel characteristics, only data from the “weekday sample households” would be used.

It is envisioned that the Saturday and Sunday samples are analyzed at the regional level for purposes of estimating aggregate regional “average Saturday” and “average Sunday” characteristics. These values will be useful when annualizing estimates of transit usage, VMT, bicycle travel, travel by tour purpose, travel by means of transportation, etc.

Future steps in the analysis of the CHTS 2012/13 travel survey for the Bay Area include detailed processing of the unlinked trip records to produce a linked trip, tour and sub-tour files. The product of this trip linking/chaining process will be both traditional linked trip files, as used in trip-based travel demand models; and tour-based travel files, for supporting the current and future generation of MTC travel behavior models.

Procedures to impute missing values will be documented in separate technical reports. Other “data cleaning” notes will be included in MTC staff notes and technical documentation.

New weighting/raking methods have also been developed for the entire statewide CHTS 2012/13 databases. Technical reports documenting these methods will be produced. In addition, the recommended household and person-level weights will be extracted and provided to CHTS 2012/13 data users. Appropriate metadata will be developed to assist the data user.

Further research on raking methods will be undertaken as time permits. Options may include simplifying some of the three-dimensional raking schemes (e.g., omitting in tenure in the PUMA by tenure by household size) to analyze the impacts on extreme weights, and raking model closure.

Procedures to impute missing trips and tours may be required, and will probably be included in future technical reports.