11-13-2007

# Estimation and Prediction of Average Vehicle Occupancies using Traffic Accident Records

Kaiyu Liu
*Florida International University*

FLORIDA INTERNATIONAL UNIVERSITY

Miami, Florida

ESTIMATION AND PREDICTION OF AVERAGE VEHICLE OCCUPANCIES

USING TRAFFIC ACCIDENT RECORDS

A dissertation submitted in partial fulfillment of the

requirements for the degree of

DOCTOR OF PHILOSOPHY

in

CIVIL ENGINEERING

by

Kaiyu Liu

2007

To: Interim Dean Amir Mirmiran
    College of Engineering and Computing

This dissertation, written by Kaiyu Liu, and entitled Estimation and Prediction of Average Vehicle Occupancies using Traffic Accident Records, having been approved in respect to style and intellectual content, is referred to you for judgment.

We have read this dissertation and recommend that it be approved.

_____
L. David Shen


_____
Fang Zhao


_____
Zhenmin Chen


_____
Albert Gan, Major Professor


Date of Defense: November 13, 2007

The dissertation of Kaiyu Liu is approved.


_____
Interim Dean Amir Mirmiran
College of Engineering and Computing


_____
Dean George Walker
University Graduate School


Florida International University, 2007

DEDICATION

I dedicate this dissertation to my dear wife, Jun Zheng, for her love, understanding, and constant support; to my lovely daughter, Tianyao Liu; to my late father, Zhaoyuan Liu; my mother, Zhenying Yang; my elder sister Zili Liu and her husband Huichun Jiang; and to my lovely nephew, Shenfeng Jiang, for their love and encouragement.

ABSTRACT OF THE DISSERTATION

ESTIMATION AND PREDICTION OF AVERAGE VEHICLE OCCUPANCIES

USING TRAFFIC ACCIDENT RECORDS

by

Kaiyu Liu

Florida International University, 2007

Miami, Florida

Professor Albert Gan, Major Professor

As congestion management strategies begin to put more emphasis on person trips than vehicle trips, the need for vehicle occupancy data has become more critical. The traditional methods of collecting these data include the roadside windshield method and the carousel method. These methods are labor-intensive and expensive. An alternative to these traditional methods is to make use of the vehicle occupancy information in traffic accident records. This method is cost effective and may provide better spatial and temporal coverage than the traditional methods. However, this method is subject to potential biases resulting from under- and over-involvement of certain population sectors and certain types of accidents in traffic accident records.

In this dissertation, three such potential biases, i.e., accident severity, driver's age, and driver's gender, were investigated and the corresponding bias factors were developed as needed. The results show that although multi-occupant vehicles are involved in higher percentages of severe accidents than are single-occupant vehicles, multi-occupant vehicles in the whole accident vehicle population were not overrepresented in the accident database. On the other hand, a significant difference was found between the

distributions of the ages and genders of drivers involved in accidents and those of the general driving population.

An information system that incorporates adjustments for the potential biases was developed to estimate the average vehicle occupancies (AVOs) for different types of roadways on the Florida state roadway system. A reasonableness check of the results from the system shows AVO estimates that are highly consistent with expectations. In addition, comparisons of AVOs from accident data with the field estimates show that the two data sources produce relatively consistent results.

While accident records can be used to obtain the historical AVO trends and field data can be used to estimate the current AVOs, no known methods have been developed to project future AVOs. Four regression models for the purpose of predicting weekday AVOs on different levels of geographic areas and roadway types were developed as part of this dissertation. The models show that such socioeconomic factors as income, vehicle ownership, and employment have a significant impact on AVOs.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1. Background

Faced with a need to increase the efficiency of transportation systems, transportation agencies are making a greater effort to monitor travel trends and measure the impacts of plans, policies, and procedures at the regional and local levels. The Intermodal Surface Transportation Efficiency Act (ISTEA) of 1991 recognizes the importance of these activities by mandating development of six management systems, including a congestion management system (CMS). To measure system performance and use systematic procedures to assess alternative congestion mitigation strategies, an increasingly recognized key element is vehicle occupancy data.

Traditionally, vehicle occupancy rates are used to convert person trips to vehicle trips in the four-step travel demand forecasting model and to determine the required parking spaces for fixed-seat facilities. The ISTEA expanded this traditional role of vehicle occupancy rates by emphasizing person movements rather than vehicle movements. Today, these data are used to formulate transportation strategies, to analyze air quality and energy efficiency measures, and to assess such programs as ride-sharing and HOV-lane implementation. Occupancy data are also used to compute person delays and person-miles traveled and to set policies for managed lanes (Gan *et al*., 2005).

With the increasing need for vehicle occupancy data comes the need to study the methods used to collect these data. Unlike counting vehicles, which can be automatically recorded when vehicles run over pneumatic road tubes, counting the number of persons

in a vehicle in the field remains largely the task of human observers. In the face of budget reductions, agencies must find better ways and define acceptable practices for collecting vehicle occupancy data that not only meet the needed accuracy, but also the limits of a restricted budget (Gan *et al*., 2005).

Existing vehicle occupancy data collecting methods can be categorized into five primary methods: the roadside windshield method, the carousel method, the photographic and video surveillance method, the survey method, and the accident data extraction method (Gan *et al*., 2005). The roadside windshield method has been the most widely used method for collecting vehicle occupancy data. It involves stationing observer(s) along the roadside to perform physical counts of vehicles and occupants. Record keeping may be carried out using paper and pencil, or by using an electronic counter. For safety reasons, the observers must stand in a protected area, such as behind a guardrail, on an overpass, or at a reasonable distance from the traffic stream, which may limit the observer's view.

The carousel method positions observers in vehicles traveling on a highway to collect vehicle occupancy data on the passing vehicles in the same direction. During data collection, the observer vehicle drives slightly slower than the general traffic, resulting in the continuous flow of traffic by the observation vehicle. This method provides additional time for the observer to make a more accurate occupancy count. However, its use is limited to only multilane highways.

The photographic and video surveillance method uses video cameras to capture passing vehicles and then uses a human observer to view the captured videos to extract occupancy data. The extra cost from equipment purchase and setup may be offset by

time-savings from sampling or fast-forwarding to skip large vehicle gaps. Although a few researchers have attempted to use computer vision techniques to automatically recognize the number of occupants in a vehicle from captured images or videos, an operational system does not currently exist.

Several survey methods have been developed to collect information directly from travelers. They include telephone interviews, travel diaries, parking lot surveys, employer surveys, household surveys, and mail-out surveys. Most of these methods require contact with residents to obtain occupancy information. Once the surveys are returned, the data have to be manually entered into a computer.

The accident data extraction method extracts vehicle occupancy estimates from police accident records for a particular road segment, corridor, or metropolitan area for specified time periods. This method relies on police reports concerning vehicles involved in traffic accidents and is only applicable when police reports record the total number of occupants in each vehicle. As of 1992, accidents forms in 16 states, including Florida, had specific fields for vehicle occupancy data, making it possible for these states to use these data to derive vehicle occupancy rates.

1.2. Problem Statement

The more traditional methods for vehicle occupancy data collection, including surveys, roadside windshield, and carousel, as described above, are labor-intensive and expensive. These procedures also require an advanced survey design to ensure unbiased portrayal of the sampled population. Also, different methods have different limitations. The roadside windshield method provides limited time to view and count passengers in a

vehicle, and it is especially difficult to count small children or backseat occupants in general.   In addition, the observer's safety, the infeasibility of collecting data on highways with inadequate shoulder areas, and the infeasibility of collecting data during the night-time or near night-time hours, are other limitations of this method.   The carousel observation method requires a continuously moving traffic stream and, as such, can only be performed on multilane highways.  It may also obstruct other traffic and create unsafe conditions because the survey vehicle travels slower than the surrounding traffic.  The photographic and video surveillance methods today still require the use of a human observer to view the photos or videos, and the quality of these photos and videos may prevent the observer from obtaining an accurate occupancy count.

Compared to other traditional collection methods, the accident data extraction method offers the greatest promise of readily available, low-cost vehicle occupancy data (Gan *et al*., 2007).   It does not require any data collection effort nor any data post-processing, the latter of which is tedious and time consuming. It provides good regional samples, and the new data are regularly updated.  This method can cover all time periods, including night hours when field data collection cannot be performed.  However, it is known that the average occupancy rates computed from accident records could be skewed by over- or under-involvement of drivers of a certain age group such as the over-representation of younger and older drivers in traffic accidents.   It has also been hypothesized that having more occupants in a vehicle may increase the likelihood of an accident occurrence due to increased interference from passengers.

A very limited number of studies can be found in the literature that examined this method and its issues as related to the extraction of vehicle occupancy information.  No

comprehensive analysis of vehicle occupancy rates based on accident records has been conducted to assist users in estimating the average vehicle occupancy (AVO) and select the associated geographic, temporal, and vehicle coverage for specific applications. On the other hand, while both accident data and field data can be used to study historical trends of occupancy rates and to collect and estimate current vehicle occupancy rates, only one existing study has attempted to develop models to predict occupancy rates as a function of socioeconomic variables.

## 1.3. Objectives and Scope

The goal of this dissertation is to develop procedures and models that make use of accident data to estimate and forecast vehicle occupancy rates. These procedures and models will attempt to minimize potential biases associated with the accident data extraction method and improve the accuracy of the estimated vehicle occupancy rates. The specific objectives of this dissertation are to:

1) Study the feasibility of accident records as a potential source of vehicle occupancy data and analyze the factors affecting average vehicle occupancies.

2) Investigate potential biases associated with the accident vehicle occupancy data and recommend acceptable and detailed procedures for estimating average vehicle occupancies using traffic accident data.

3) Develop tools to facilitate the collection, processing, and analysis of vehicle occupancy rates using accident data.

4) Develop vehicle occupancy prediction models as a function of local socio-economic data.

1.4. Dissertation Organization

This dissertation consists of the following eight chapters:

1) Chapter 1 introduces the background of this dissertation research, describes the problem to be solved, and sets forth the goal and objectives to be achieved.

2) Chapter 2 contains a review of the literature. A review of vehicle occupancy data collection method coverage, sampling procedures, and the advantages and disadvantages of all existing methods is first presented. The feasibility of and procedures for making use of accident data to estimate vehicle occupancy rates are also reviewed, and the biases derived from accident data and the adjustment factors for the estimation procedures are then discussed.

3) Chapter 3 presents the results from both parametric and nonparametric statistical tests using the vehicle occupancy data from accident records. The test results identify the contributing factors affecting AVOs in the state of Florida.

4) Chapter 4 focuses on the adjustment factors to account for potential biases stemming from using accident records to estimate AVOs.

5) Chapter 5 describes a user-friendly AVO information system developed to analyze and generate AVO estimates at the district, county, and corridor levels and to monitor statewide, regional, and site-specific vehicle occupancy trends.

6) Chapter 6 evaluates the AVOs extracted from accident records based on both reasonableness checks and comparisons with vehicle occupancy field data.

7) Chapter 7 presents the method and process of developing weekday AVO prediction models that are represented as a function of socioeconomic information. Evaluation of the model accuracy using the test data set is also discussed.

8) In Chapter 8 the major research results from each chapter are summarized, conclusions are drawn, and issues for future research and system improvements are recommended.

CHAPTER 2


LITERATURE REVIEW


While many studies have included reviews of the procedures and issues related to vehicle occupancy data collection methods, only a few studies have been carried out using the accident data extraction method. This section provides a review of the literature on issues of vehicle occupancy rate estimation and forecast, sampling process, and accident data extraction procedures and strategies.

2.1. Vehicle Occupancy Data Collection Design Issues

Various studies have been conducted to collect vehicle occupancy data for different purposes. For example, Rutherford *et al*. (1990) reviewed the methods used by agencies for short-term and long-term monitoring of high-occupancy vehicle (HOV) lane violations. Levine and Wachs (1994) presented a methodology for conducting regional and corridor-level vehicle occupancy surveys. Heidtman *et al*. (1997) compared various vehicle occupancy data collection methodologies based on field studies and cost consideration. Most of these studies included the following conditions during data collection:

- Observation locations and days were chosen arbitrarily, rather than sampled randomly, to represent vehicle occupancy accurately for a particular purpose.

- Locations with higher traffic volume are more likely to be chosen to the exclusion of lower volume locations.

- Sites near central business districts (CBDs) or major trip attraction areas where congestion is assumed to be highest were also chosen frequently.

- Many agencies appear to acquire average vehicle occupancy (AVO) estimates using the simple mean calculation procedure. Vehicle occupancy observations taken under these conditions will create biased estimates, and, in turn, unreliable and inaccurate findings will be derived.

The Heidtman *et al*. (1997) study also provides a chronological summary of studies conducted in terms of study objectives, study methodology, and use of temporal and spatial coverages. The study concluded that:

- The roadside windshield method was the most commonly used method.

- There was a trend to save data entry time by switching recording from manual count boards to laptop computers or electronic counter boards.

- Measurements were usually taken during either the a.m. or p.m. peak period.

- Observers generally counted in short periods rather than continuously counting for one hour.

- Observations were typically taken in the middle of the week (Tuesdays, Wednesdays, and Thursdays) to avoid unusual conditions of recreational travel prevailing on Mondays, Fridays, and weekends.

The existing literature on vehicle occupancy collection methods shows that almost all studies have focused on the following design and implementation issues and considerations: geographic and temporal coverage, locations, vehicle types and percentages, and sampling methods.

## 2.1.1. Geographic Coverage

Geographic units used to evaluate vehicle occupancy may include the entire state, air quality non-attainment areas, urban areas, subareas, regions, corridors, activity centers, functional classes, and external cordon lines (Heidtman *et al*., 1997; Zakaria, 1991). Depending on the specific study objectives, one or more of these units may be appropriate. The traditional emphasis has been on estimating vehicle occupancy at either the site-specific (e.g., factory entrance and critical roadway section) or area-wide (e.g., metropolitan and county) level. The more recent focus, however, has been on the intermediate level of geographic coverage, which is referred to as the corridor level (e.g., at cutlines and Central Business District cordon lines).

## 2.1.2. Temporal Coverage

Temporal variability in AVO is a common issue for all data collection methods. Many existing studies show significant variations in vehicle occupancy rates by time-of-day, day-of-week, month-of-year, and season-of-year (Benke and Sjoberg, 1977; Kuzmyak, 1981; Heidtman *et al*., 1997; Reed *et al*., 1998; Levine and Wachs, 1998; Liu and Desai, 1998).

*Time-of-Day*: These studies show that the AVO during the a.m. peak period tends to be lower and heavily dominated by home-based work trips, primarily with single-occupant vehicles. As the day progresses, an increasing variety of other trip purposes occur that are likely to shift vehicle occupancy upward. Hence, during the p.m. peak period, vehicle occupancy may be higher than that during the a.m. peak period due to a multitude of different trip purposes, many of which involve more than one person.

*Day-of-Week*:  Vehicle occupancy rates vary on different days of the same week, with Saturday typically having higher rates.  However, studies did not find a consistent day-of-week pattern for variations in AVO.  In general, it is a common practice to exclude Monday, Friday, and weekend counts because these days are assumed to contain atypical and non-recurrent trips.  Measurements were typically taken on Tuesday, Wednesday, and Thursday only, as these days are considered most representative of average weekday travel behavior and commute conditions (MARC, 2002).

*Month/Season-of-Year*:  Vehicle occupancy rates have been found to vary with month and season.  It is, thus, important that observations be taken throughout the year (and not restricted to any one season) in order to properly represent AVOs (Heidtman *et al.*, 1997).

*Temporal Trends*:  The following temporal trends in AVO have been observed in various studies (Roach and Lester, 1978; URS Corporation, 1997):

- Weekday a.m. peak AVO is normally lower than weekday midday and p.m. peak AVO.

- Weekend AVO is normally higher than weekday AVO.

- Off-peak AVO is normally higher than either a.m. or p.m. peak AVO.

- Summer AVO is higher than winter AVO for northern cities.

Based on data from Florida, Liu and Desai (1998) made the following general recommendations with regard to temporal coverage:

- The data collection period should ideally be within the period of interest (e.g., peak hours, etc.).

11

- Counts of one to two hours should produce data with sufficient accuracy and precision for most purposes.

- Tuesdays and Wednesdays are normally adequate for data collection, and the best days for collecting data are Thursdays.

- Mondays and Fridays should be avoided.

- Time of day chosen for data collection is important.

- As a rule of thumb, mid-morning to mid-afternoon counts are adequate for most purposes. However, if the 5-6 p.m. period were chosen for performance monitoring, adjustments would be necessary to derive the AVO for the day.

2.1.3. Facility Types and Trip Purposes

Different types of roadways typically have different occupancy levels. Roadways of the higher functional hierarchy will typically be expected to have a lower AVO. Vehicles traveling on freeways have lower occupancy rates than vehicles traveling on arterials and collectors (Barton-Aschman Associates, 1989). Kuzmyak (1981), in his nationwide personal transportation study, compared the AVO rates based on different trip purposes. The result indicated that average vehicle occupancy rate was 1.87, ranging from a low of 1.32 persons per vehicle mile for travel related to earning a living to a high of 2.44 persons per vehicle mile for social and recreational travel.

Freeways and major arterials are heavily used for home-based work trips and commercial trips, particularly in the a.m. and p.m. peak periods. These two trip purposes tend to have lower AVO. Conversely, minor arterials, collectors, and local streets are

utilized more extensively for trips involving home-based shopping, home-based school, and social/recreational purposes, resulting in higher AVO.

The data collection methods for freeway segments with high occupancy vehicle (HOV) lanes have seen improvement. Rutherford *et al.* (1990) reviewed different technologies used by state agencies for monitoring HOV facilities. Pavlidis *et al.* (1999) studied wave band and computer vision methods to automatically count vehicle occupants in HOV lanes. However, the team could not achieve the accuracy and reliability needed for real-world automated enforcement of HOV lanes.

## 2.1.4. Locations

AVO can differ from one location to another. Benke (1977) found that significant differences occur in occupancy rates at different subareas of a corridor, probably relating to trip length, origin/destination density, resident demographics, etc. This also means that the spatial variations of AVO are related to the distribution of household types and work places. A research project in Phoenix examined factors that have influences on AVO. It was found that the lowest vehicle occupancy rates occur in the core area of the region and that the highest occupancy rates occur in the outlying suburban areas (Barton-Aschman Associates, Inc., 1989).

Site selection for traffic data collection is dependent on the purposes involved and the overall expectations of the study (Liu and Desai, 1998). To ensure that a representative sample of the population will be selected, data collection locations should be sampled randomly from all possible highway segments for area-wide studies. If an initial study has been performed, the entire random selection procedure should be

repeated again for subsequent studies. The same sampled locations should not be chosen unless they are randomly selected. In contrast to randomly selected locations for area-wide studies, locations are often judgmentally selected due to the nature of corridor/site-specific studies.

2.1.5. Vehicle Types

The vehicle types included during data collection are determined by the purpose for which the data are to be collected. Different study purposes may utilize different criteria for interpreting AVO. Different vehicle types have different occupancy rates. To compare vehicle occupancy between household survey data and roadside observation data, Grush and Gross (1995) studied the Nationwide Personal Transportation Survey (NPTS) survey occupancy data and the National Highway Traffic Safety Administration (NHTSA) study data. It was found that the highest average occupancy rate occurred in vans, followed by passenger cars, and then pick-up trucks. Pamanikabud (2000) also analyzed each type (private passenger car, private pick-up, and public taxi) of vehicle's morning occupancy rate entering the city center of Bangkok, Thailand. From this morning peak period analysis, the lowest occupancy rate appeared in passenger cars. Pick-ups' occupancy rate was in the middle of all three types. Taxis had the highest rate of 2.009. To collect public transportation service data, Demery (2002) listed peak-period vehicle occupancy statistics (passengers per vehicle and passengers per meter of vehicle length) for U.S. and Canadian rapid bus and rapid rail services.

In most vehicle occupancy studies, only data from passenger vehicles or light vehicles (private passenger automobiles, pick-ups, vans, recreational vehicles, and

motorcycles) are usually counted. Buses are typically excluded or counted separately because it is difficult to count all of the occupants using the roadside windshield method and carousel method. Trucks are generally excluded because they are used mainly for goods movement and have little to do with human mobility. Note that the definition of vehicle type remains an issue in vehicle occupancy data collection. There is no consistency in definitions across different regions or different transportation models.

2.1.6. Vehicle Percentages

Instead of recording data from all vehicles traversing a link during a selected time period, a subset of vehicles that is assumed to reflect the same characteristics as the target population is generally observed. Ferlis (1981) recommended a systematic short-count procedure, in which observations are made for a fixed interval in each hour of the day. This procedure can be used to produce relatively accurate daily estimates while conserving manpower resources. The following three basic types of short-count procedures were suggested by Ferlis (1981):

- Position one or more observers to count all vehicles that pass by during a fixed interval within each hour (e.g., count for 45 minutes and rest for 15 minutes, thus representing a 75% systematic sample).

- Position one observer to count vehicles that pass by on each lane during a fixed interval within each hour (e.g., count each of three lanes during successive 15-minute periods and rest for 15 minutes within each hour, thus representing a 25% systematic sample).

- Position one or more observers to systematically observe two or more locations concurrently by counting all vehicles passing a particular location during the same time interval within each hour (e.g., count vehicles at one location from 7:00 to 7:15, 8:00 to 8:15, etc., and at another location from 7:30 to 7:45, 8:30 to 8:45, etc., thus representing a 25% systematic sample).

## 2.1.7. Sampling Issues

The sampling process that is conventionally used in vehicle data collection methods includes simple collection and stratified collection. Simple collection involves the selection of sample units at random from the entire population so that each sample unit has an equal probability of being selected. Stratified collection, on the other hand, begins by dividing the entire sample frame into mutually exclusive strata, then selecting sample units from each stratum using simple collection.

## 2.1.7.1. Simple Collection

The highly referenced sampling procedures were developed by Ferlis (1979 and 1981). Ferlis's research showed that a reliable occupancy estimate using statistical survey techniques of stratification and sampling could be obtained at a lower cost than traditional approaches that used less frequent but longer data collection periods. Another report (FHWA, 1980) complements Ferlis's work. Field tests of these methods in Atlanta (Fisher *et al*., 1980) indicated that the sampling techniques were very useful in obtaining reliable occupancy estimates and that the methods were cost-effective.

The sample size estimation formula developed by Ferlis (1981) is as follows:

$$N = \left( \frac{Z \times \sigma}{T} \right)^2 \qquad\qquad\qquad (2\text{-}1)$$

where

$N$ = the number of link-days to be sampled,

$Z$ = the normal variant for the specific level of confidence,

$T$ = the desired tolerance, and

$\sigma$ = the composite standard deviation of AVO.

The basic sampling unit is a "link-day," representing an estimate of the survey measures made for a particular link on a particular day for a specified time period. The composite standard deviation of AVO was based on the following formulas:

Area-wide level:

$$\sigma = \left( \sigma_L^{\,2} + \sigma_S^{\,2} + \sigma_W^{\,2} \right)^{1/2}$$

Site-specific level:

$$\sigma = \left( \sigma_D^{\,2} + \sigma_S^{\,2} + \sigma_W^{\,2} \right)^{1/2}$$

where

$\sigma_L$ = the standard variation of AVO across link-days within a season,

$\sigma_D$ = the standard variation of AVO from day to day within a season,

$\sigma_S$ = the standard variation of AVO from season to season, and

$\sigma_W$ = the standard variation of AVO across time periods within a day.

Additional sources of variation have been suggested in other studies. Ulberg and McCormack (1988) examined some potential sources of error and concluded that observer counting error ($\sigma_O$) should be included. Accordingly, Ferlis's formulas for composite standard deviation were modified to:

Area-wide level:

$$\sigma = \left(\sigma_L{}^2 + \sigma_S{}^2 + \sigma_W{}^2 + \sigma_O{}^2\right)^{1/2}$$

Site-specific level:

$$\sigma = \left(\sigma_D{}^2 + \sigma_S{}^2 + \sigma_W{}^2 + \sigma_O{}^2\right)^{1/2}$$

Depending on the sampling procedure, some of these sources of variance may not apply. The link-day variation, $\sigma_L$, reflects the fact that different AVO measurements can vary by measured locations as well as by different days within a season. For an area-wide study, $\sigma_L$ accounts for a much greater variability than the other terms. The daily variation term, $\sigma_D$, should be included for corridor and site-specific studies because it can vary substantially from one day to the next. The seasonal variation, $\sigma_S$, should be included if the estimated occupancy is intended to represent more than one season, and the data collection is, therefore, extended throughout this period. Similarly, the within-day variation, $\sigma_W$, should be included only if a short-count method is used.

A study was conducted in 1997 by URS Corporation for the Florida Department of Transportation to collect vehicle occupancy data from 21 individual sites throughout the state of Florida. Accordingly, due to the time periods during which the occupancy data for private vehicles were collected via the use of short-counts, variations for day of week and month of year could be established. Four seasonal sites were observed twice each month over a one-year period to determine the monthly and seasonal variation in vehicle occupancy.

Table 2-1 shows the results computed from local data for different sources of variation, along with the suggested values from studies by Ferlis (1981) and Ulberg and McCormack (1988). The table shows that the standard deviations observed in the data

from the URS study are generally larger than those suggested by Ferlis. Note that the values for the within-day term, $\sigma_W$, are based on an assumed systematic sampling rate of 25%.

Table 2-1: Sources of Variation in Average Occupancy

| Geographic Scope | Standard Variation | Ferlis (1981) | Ulberg and McCormack (1988) | Data from URS Study (1997) |
|---|---|---|---|---|
| Area-wide | $\sigma_L$ | 0.063 | | 0.076 |
| | $\sigma_S$ | 0.015 | | 0.068 |
| | $\sigma_W$ | 0.017 | | 0.008 |
| | $\sigma_O$ | | 0.006 | |
| Site-specific or Corridor | $\sigma_D$ | 0.015 | | 0.028 |
| | $\sigma_S$ | 0.015 | | 0.068 |
| | $\sigma_W$ | 0.017 | | 0.008 |
| | $\sigma_O$ | | 0.006 | |

2.1.7.2. Stratified Collection

Agencies often place emphasis on collecting separate AVO estimates for various subsets of the area-wide transportation network (e.g., separate estimates for individual counties, freeways and arterials, and freeways by HOV and mixed-flow lanes). In this case, the collection population is stratified for separate sampling (Ferlis, 1981; Heidtman *et al* 1997).

In contrast to the simple estimating method, the composite standard deviation is estimated for each stratum as:

Area-wide Level: $$\sigma_h = \left( \sigma_{Lh}^{2} + \sigma_{Sh}^{2} + \sigma_{Wh}^{2} + \sigma_{Oh}^{2} \right)^{1/2}$$

Corridor Level: $$\sigma_h = \left( \sigma_{Dh}^{2} + \sigma_{Sh}^{2} + \sigma_{Wh}^{2} + \sigma_{Oh}^{2} \right)^{1/2}$$

where

$\sigma_h$ = composite standard deviation of AVO in stratum *h,*

19

$\sigma_{Lh}$ = standard deviation of AVO across link-days within a season in stratum $h$,

$\sigma_{Dh}$ = standard deviation of AVO across days within a season in stratum $h$,

$\sigma_{Wh}$ = standard deviation of AVO across time periods within a day in stratum $h$,

$\sigma_{Sh}$ = standard deviation of AVO across seasons in stratum $h$, and

$\sigma_{Oh}$ = standard deviation of AVO due to observer error in stratum $h$.

The sample size of link-days required to estimate stratum AVO within a desired tolerance can be computed as:

$$N_h = \left( \frac{Z \times \sigma_h}{T_h} \right)^2 \qquad\qquad (2\text{-}2)$$

where

$N_h$ = number of link-days required for stratum $h$,

$T_h$ = desired tolerance for stratum $h$, and

$\sigma_h$ = composite standard deviation of AVO in stratum $h$.

Instead of determining the sample size to be collected, traffic accident records are used to determine if the database contains a sufficient number of records to make the required estimation. The sampling unit is the vehicle(s) involving in an accident. In Heidtman's report, the above formula was used to calculate the sample size for Chicago accident data vehicle occupancy analysis.

## 2.2. AVO Calculation

Heidtman *et al*. (1997) listed all AVO calculating methods, which include those for field observational methods and the mail-out questionnaire method. The most common method of deriving an AVO from accident data is to use a method similar to the

observational method's way of deriving AVO. This method simply divides the total number of occupants by the total number of vehicles for a specified geographic area and the time period. This calculation assumes that the traffic volumes observed at each location is proportional to the actual traffic flow at the corresponding location. The simple estimate of AVO and its corresponding composite standard deviation can be computed as follows:

$$AVO = \frac{\sum_i P_i}{\sum_i V_i} \qquad (2\text{-}3)$$

$$\sigma = \left[ \left( \frac{N}{N-1} \right) \frac{\sum_i (P_i - AVO \times V_i)^2}{\left( \sum_i V_i \right)^2} \right]^{1/2} \qquad (2\text{-}4)$$

where $P_i$ is number of persons counted in session $i$ and $V_i$ is number of vehicles counted in session $i$. These formulas can also be applied to data collected for each stratum of a stratified data collection to compute the AVO and its composite standard deviation $\sigma_h$. This method is not applicable given the fact that accident data are likely to reflect a biased sample of the population. To accommodate these biases, some adjustments and filtering of the data, as well as a weighted means procedure, are required to remove biases inherent in these types of data.

## 2.3. Accident Data Extraction Method

The number of passengers in a vehicle involved in an accident is routinely recorded in many states by police officers at the site of the accident. Because of the

accessibility and low cost associated with obtaining accident data, several states have shown interest in extracting vehicle occupancy estimates from police accident reports (Ahuja and Hanscom, 1996; Asante *et al.*, 1996; Gaulin, 1991). Past studies on this method have focused on estimation procedures and the potential biases existent in this method.

2.3.1. Estimate Procedures

A Connecticut DOT report (Gaulin, 1991) provides what has served as the basic method of calculating vehicle occupancy from traffic accident data. In Gaulin's report, estimates from roadside windshield surveys, which were conducted in 1982 and 1984, were compared to estimates obtained from accident data. The results indicated a relative difference in estimates of less than 2.4% over comparable time periods. Several factors that can influence accident estimates were identified in the report as part of the verification process for using this new accident data procedure. For example, the impact of male versus female drivers, good versus bad weather, and alcohol/drug-related accidents were considered. In addition, sample size requirements for different tolerance levels were discussed. This report illustrates how this procedure can be used for assessing the viability and performance of facility-specific projects. It also shows that it is feasible to calculate vehicle occupancy rates from accident data.

To provide system-level average automobile occupancy (AAO) data, the New York State Department of Transportation (NYSDOT) built upon Gaulin's procedure and tailored the algorithm (Asante *et al.*, 1996). In the NYSDOT procedure, accident data were classified and cross-classified by year, time-of-day interval, day of week, month,

and county. New York's 62 counties were classified into three urban-size groups for AAO analysis: 1) large urban counties with populations over 1 million; 2) large urban counties with county populations under 1 million but more than 100,000; and 3) rural counties with populations under 100,000. The ANOVA test indicated that most of the variations in the means of accident-based automobile occupancy observations could be explained by two main effects: the day-of-week groups and time-of-day intervals. Two other effects are the seasonal groups for months and population-size groups for counties.

Because the AAOs calculated from accident data were significantly higher than multiple-station roadside-observed AAOs, a simple linear regression model was developed to adjust the accident-based AAO estimate, $AAO_{(accident)}$, so that it would be comparable to roadside observer-based AAOs. The adjusted AAO, $AAO_{(adjusted)}$, from the fitted regression model was (Asante *et al.*,1996):

$$AAO_{(adjusted)} = 0.30 + 0.69 \times AAO_{(accident)} \qquad (2\text{-}5)$$

Based on the statistical analyses reported, NYSDOT decided that accident-based AAO was adequate for purposes of network-level, area-wide CMS analysis processes. The NYSDOT analysis of accident-based AAO shows that these data accurately reflect the distribution of automobile occupancy during various periods of the day. Calculated occupancy rates were found to be the lowest during the morning peak and highest in the evening period. The weekday rates were lower than weekend rates. These findings were consistently observed across all counties and urban-area-size groups and were supported by other studies and data sources (Heidtman *et al.*, 1997; Barton-Aschman Associates, 1989; FHWA, 1980).

2.3.2. Bias Issues and Adjustments

A major problem associated with the use of accident data for vehicle occupancy has been the potential biases resulting from the over-representation of younger drivers, alcohol- and drug-related accidents, and fatal accidents.

2.3.2.1. Age Bias

Heidtman *et al.* (1997) studied this problem by using Chicago region accident data (years 1993 and 1994). It was found in this study that occupancy rates vary by the age of the driver for the Chicago area. AVO is clearly highest for the youngest drivers, dropping sharply by the age of 20, with slight increases during childbearing years and retirement (Figure 2-1). To adjust for this bias, a weighted mean procedure using the census data was used in this evaluation. To obtain the overall AVO estimate, the weighted mean procedure would weight each age group's AVO in proportion to its age group's contribution to the overall driving population, as defined by the census data.

Although there were substantial differences in the AVO estimates for various times of the week, the Chicago accident data set did not show significant differences between the unadjusted and census-weighted estimates. This lack of difference may be an anomaly peculiar to this one data set. Over-representation of vehicles driven by younger drivers may have been offset by the under-representation of vehicles driven by older drivers.

2.3.2.2. Vehicle At-Fault Bias

Heidtman *et al.* (1997) also investigated the impact of another potential biasing factor: unsafe drivers, such as those that are fatigued or under the influence of drugs or

alcohol. The AVO for vehicles "more likely at-fault" was found to be significantly lower than that of the additional vehicles involved in the accident. However, the vehicle at-fault for the accident was judged personally by the reporting officer, and this information does not reflect the court's opinion of the vehicle at-fault or responsible for the accident.
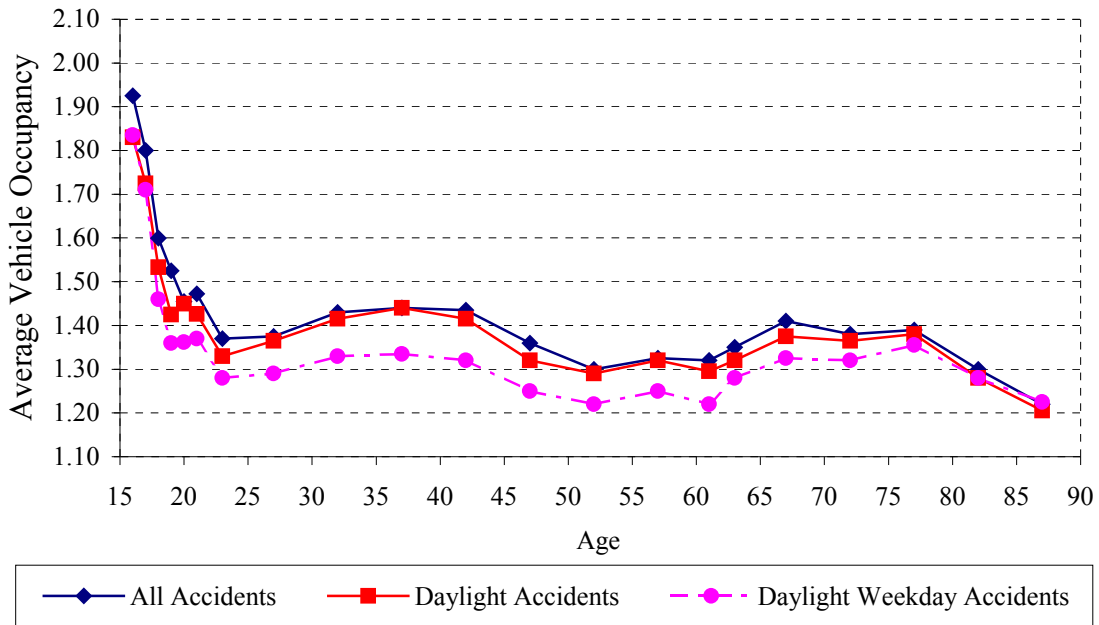


Figure 2-1: Average Vehicle Occupancy by Age for Chicago Area

2.3.2.3. Accident Severity Bias

The Asante *et al*. study (1996) has shown that accident records significantly overestimate vehicle occupancy. One primary explanation is that high-occupancy vehicles are more likely to be involved in severe accidents and are, thus, more likely to be reported in the accident database. This relationship arises from the fact that an increase in the number of people in the vehicle increases the likelihood of injury and fatality because the number of occupants exposed to the risk of being injured or killed increases

with higher occupancies, while in vehicles with lower occupancies, unoccupied seat space provides protection from intrusion. A study by Chang and Mannering (1998) resulted in similar findings. In their study, a database containing all reported traffic accidents occurring on principal arterials, state highways, and interstate highways in King County (in Washington State) during 1994 compared the average vehicle occupancies for accidents with different severity levels. It was found that multi-occupant vehicles were more likely to be involved in severe accidents.

Based on studies that examined variables that significantly influence accident severity probabilities (Lui *et al*., 1988; Shankar *et al*., 1996; O'Donnell and Connor, 1996), Chang and Mannering also developed a nested logit model to estimate the probability of different occupancies and the probability of different accident severities. Estimated results from the model were then compared with actual observations for the same roadways to determine the appropriate adjustment factors to be applied in the model.

## 2.4. Forecast Issues in Vehicle Occupancy Data Collection

The literature on studies of vehicle occupancy data forecasting has been very limited in recent years. The Federal Highway Administration (1972) sponsored a study to review four methods used in developing vehicle occupancy models: average factor method, curve fitting, cross classification, and regression analysis. Examples of studies that have used these methods were listed in this report. Forbord (1966) used regression techniques to relate vehicle occupancy to socio-economic variables using data from Twin

Cities, Minnesota. The following vehicle occupancy models were developed to predict the auto occupancy rates using socioeconomic data:

Work trip auto occupancy:

$$AVO = [1.411 - 0.202 \times 10^{-4}\,(Income)] \times [0.972 + 5.878 \\ \times 10^{-4}\,(Employment\_Density)] \qquad (2\text{-}6)$$

Other trip auto occupancy:

$$AVO = 1.75 - 0.16 \times 10^{-4}\,(Income) \qquad (2\text{-}7)$$

It was also pointed out that there were similarities and relationships between the development of a modal split model and the development of an auto occupancy model. The characteristics or variables considered in the development of both models are the same and include the characteristics of the trip (purpose and length), the trip maker (auto ownership, income, residential density), land use type (residential, manufacturing, commercial, etc.), employment density, and the transportation system (travel time and parking costs).

To forecast the potential use of highway facilities, Ulberg (1994) attempted to incorporate the best of previous transportation models with new information on psychological and demographic determinants of mode choice into a model that could forecast the vehicle occupancy for a specific highway facility. The research began with Tybout and Hauser's (1981) "integrated model of consumer choice" as a basis. Four survey data sets were then analyzed. However, the results and findings in this study were related mode choice models only. Also, the results concerning the impact of socioeconomic variables on mode choices were mixed.

2.5. Summary

In this chapter, a comprehensive literature search and review has been performed to investigate and assess vehicle occupancy data collection practices. The major findings of the review are summarized as follows:

- Study Coverage: Most vehicle occupancy data collection studies were one-time, special studies undertaken as part of a large project, typically the development of a travel demand model system. Regular monitoring of vehicle occupancies on a predefined periodic basis is much less common.

- Accident Extraction Method: Collection of vehicle occupancy data has been primarily conducted by using traditional collection methods such as windshield surveys and mail-out trip diaries or postcards. Because of the readily availability and low cost of accident data, the accident extraction method holds great potential as a rich and economical source of vehicle occupancy data. Past studies on this method have focused on the estimation procedures and potential biases associated with this method. The feasibility of this method was verified, and some states have already used this method to collect vehicle occupancy data for their transportation planning projects. However, comprehensive analysis of variables that affect vehicle occupancy estimates from accident data has been very limited, and a system tool that can ease data collection, processing, and analysis has been conspicuously absent.

- Geographic Coverage: Geographic units used to evaluate vehicle occupancy may include the entire state, air quality non-attainment areas, urban areas, subareas, regions, corridors, activity centers, functional classes, and external cordon lines.

Depending on the specific study objectives, one or more of these units may be appropriate.

- Variations: Temporal variability in AVO is a common issue across all data collection methods. Existing studies show significant variations in vehicle occupancy rates by time-of-day, day-of-week, and season-of-year. Different types of roads typically have different occupancy levels. For example, roadways of the higher functional hierarchy will typically be expected to have lower AVO. It is, thus, important to sample all roadway types to generate a representative estimate of regional vehicle occupancy. AVOs can also differ from one location to another. The spatial variations of AVO are related to the distribution of household types and work places.

- Sampling Procedure and AVO Calculation: Most research and studies have focused on sampling procedures and AVO calculation methods used in different vehicle occupancy data collection methods. These procedures and formulas can be used to obtain reliable and accurate vehicle occupancy data at the regional, corridor, and site levels. Although these procedures and formulas were provided over 20 years ago, the recommended stratified sampling techniques and weighted average procedures have seldom been used. Many states appear to use the simple mean calculation procedure.

- AVO Prediction Model: Few studies on AVO prediction models have been carried out in recent years. Only one existing study attempted to develop models to predict occupancy rates as a function of socioeconomic variables.

CHAPTER 3


FACTORS AFFECTING AVO: EVIDENCE FROM ACCIDENT DATA


3.1. Introduction

The number of passengers in a vehicle involved in an accident is routinely recorded in many states by police officers at the site of the accident. The use of this data source for vehicle occupancy data offers greater benefits over the traditional field data collection methods. This chapter describes an effort to apply the vehicle occupancy data in Florida's accident records to examine the impacts the following factors may have on vehicle occupancy: accident year, accident month, accident day of week, accident time-of-day, county location, roadway type, accident severity, driver's age group, driver's gender, driver's race, and weather conditions.


3.2. Data Preparation

A total of 16 years (1990–2005) of accident records for Florida's state roadway system were used in the analysis presented in this chapter. The data were obtained from the Florida Department of Transportation (FDOT) and integrated into a single database. Only passenger vehicles (i.e., cars, SUVs, pick-up trucks, and vans) were included in the analysis. The AVO for a specific category was calculated as follows:

$$AVO_j = \frac{\sum_i P_{ij}}{V_j} \qquad (3\text{-}1)$$

where $P_{ij}$ is the number of persons in vehicle $i$ in category $j$ and $V$ is the total number of vehicles in category $j$.

One issue in the data preparation process was to determine the data intervals for some of the factors considered. The use of too few intervals can mask the differences in AVO. On the other hand, the use of too many intervals will result in sample sizes being too small to achieve the desired confidence in the AVO estimates. Based on these considerations, the original data intervals for some of the variables considered were re-classified for this analysis as follows:

- The original data include the specific hour during which the accident occurred. The hours were re-classified into the following intervals: 1) 6:00 a.m. – 6:59 a.m., morning period; 2) 7:00 a.m. – 8:59 a.m., morning peak period; 3) 9:00 a.m. – 3:59 p.m., off-peak period; 4) 4:00 p.m. – 5:59 p.m., afternoon peak period; 5) 6:00 p.m. – 10:59 p.m., evening period; and 6) 11:00 p.m. – 5:59 a.m., late night and early morning. The purpose of the 6:00 a.m. – 6:59 a.m. morning period was to evaluate the transition from the late-night period to the morning peak.

- Florida's counties were classified into three urban-size groups: 1) large urban counties with county populations over 500,000; 2) medium-size counties with county populations under 500,000 and over 100,000; and 3) rural counties with populations under 100,000.

- The original data included the specific age of each driver, which was re-classified into one of the following nine groups: 16-19, 20-24, 25-29, 30-39, 40-49, 50-59, 60-69, 70-79, and $\geq 80$.

- Roadway functional classifications were re-classified from their original classes in FDOT's Roadway Characteristics Inventory (RCI) to the following four major groups: collectors, minor arterial, major arterial, and freeways.

3.3. Statistical Analysis Methods

Both parametric analyses of variance (ANOVA) test and non-parametric analyses of variance test were used to determine if there are significant differences in AVO among the different categories of independent variables (i.e., the factors being examined for their impact on the AVO). The general parametric ANOVA model used for the AVO variance analysis is defined as follows:

$$Y_{ij} = \mu_i + \varepsilon_{ij} \tag{3-2}$$

where

$Y_{ij}$ = the number of occupants in an observed vehicle $j$ for group $i$,

$\mu_i$ = the group mean, and

$\varepsilon_{ij}$ = the random error associated with individual observations.

Parametric tests require the assumptions of stabilized variance and normally distributed data, which usually affect the $F$-test and $t$-test. In this dissertation study, nonparametric tests were used essentially to check the conclusions of the parametric ANOVA tests.

The Kruskal–Wallis $H$-test (McClave and Benson, 1988) is often viewed as the nonparametric equivalent of the parametric one-way analysis of variance (one-way ANOVA). The advantage of the Kruskal–Wallis $H$-test is that no assumptions about the nature of the sampled populations are needed. The $H$-test statistic is given by:

$$H = \frac{12}{n(n+1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 3(n+1) \tag{3-3}$$

where

$n_i$ = the number of measurements in the $i$th sample,

$n$ = the total sample size ($n=n_1+n_2+\ldots+n_k$), and

$R_i$ = the rank sum of the $i$th sample.

Another nonparametric test called the Friedman $F_r$ test (McClave and Benson, 1988) was also used to test for significant AVO differences across different factor groups, such as time-of-day interval. The Friedman $F_r$ test is defined as follows:

$$F_r = \frac{12}{bp(p+1)} \sum_{i=1}^{p} R_i^2 - 3b(p+1) \qquad (3\text{-}4)$$

where

$b$ = the number of blocks (i.e., counties),

$p$ = the number of treatments (i.e., time-of-day intervals), and

$R_i$ = the rank sum of the $i$th treatment, where the rank of each measurement is computed relative to its position within its own block.

The ANOVA test results are summarized in Table 3-1, which includes the factor results and contrast results. The top section of Table 3-1 lists the 12 factors with ANOVA results for each. The $F$ value associated with each factor is used to test the null hypothesis that the treatment (or level or group) means are equal. $Pr > F$ gives the level of significance associated with the $F$ value. The $F$-test results for all 12 factors shown at this section indicate that these factors significantly affect AVO data at the 95% significance level.

The contrast section listed in the bottom of Table 3-1 shows ANOVA results for selected treatment levels. Contrasts compare treatment means to determine significant differences among the levels. For example, given that there were significant differences

in AVO data among the time-of-day intervals, contrasts were used to test if the a.m. peak

AVO data are significantly different from the p.m. peak AVO data.

Table 3-1:  Results of ANOVA Test

| Factors | DF | Type I SS | Mean Square | F Value | Pr > F |
|---------|-----|-----------|-------------|---------|--------|
| Severity | 2 | 1513.879176 | 756.939588 | 1167.75 | <.0001 |
| Year | 15 | 8199.115310 | 546.608091 | 659.73 | <.0001 |
| Weekday_Group | 3 | 2446.706441 | 815.568814 | 1258.20 | <.0001 |
| Month_Group | 3 | 26.786878 | 8.928959 | 13.77 | <.0001 |
| Hour_Group | 6 | 821.791828 | 136.965305 | 211.30 | <.0001 |
| County_Group | 2 | 333.004620 | 166.502310 | 256.87 | <.0001 |
| Functional Class | 10 | 328.088703 | 32.808870 | 50.62 | <.0001 |
| Driver Age_Group | 8 | 959.839016 | 119.979877 | 185.10 | <.0001 |
| Driver Race | 3 | 510.087734 | 170.029245 | 262.31 | <.0001 |
| Driver Gender | 1 | 100.919664 | 100.919664 | 155.69 | <.0001 |
| Weekday | 6 | 65.164458 | 10.860743 | 16.76 | <.0001 |
| Weather | 4 | 13.655780 | 3.413945 | 5.27 | 0.0003 |

| Contrast | DF | Contrast SS | Mean Square | F Value | Pr > F |
|----------|-----|-------------|-------------|---------|--------|
| Fall vs Summer | 1 | 32.11088607 | 32.11088607 | 47.66 | <.0001 |
| Fall vs Winter | 1 | 16.64394871 | 16.64394871 | 24.70 | <.0001 |
| Spring vs Winter | 1 | 0.08336151 | 0.08336151 | 0.12 | 0.7250 |
| Summer vs Winter | 1 | 5.79238464 | 5.79238464 | 8.60 | 0.0034 |
| Morning Period vs AM Peak | 1 | 5.6275180 | 5.6275180 | 8.68 | 0.0032 |
| AM Peak vs PM Peak | 1 | 270.7720362 | 270.7720362 | 417.73 | <.0001 |
| PM Peak vs Early Evening | 1 | 12.8135720 | 12.8135720 | 19.77 | <.0001 |
| PM Peak vs Late Evening | 1 | 87.6609210 | 87.6609210 | 135.24 | <.0001 |
| Mon. vs Tue.&Thr. | 1 | 3.5400978 | 3.5400978 | 5.33 | 0.0210 |
| Tue.&Th.r vs Fri. | 1 | 106.9256005 | 106.9256005 | 160.98 | <.0001 |
| Fri. vs Weekends | 1 | 938.3923881 | 938.3923881 | 1412.76 | <.0001 |
| Thr. vs Fri. | 1 | 46.1432708 | 46.1432708 | 69.48 | <.0001 |
| Fri. vs Sat. | 1 | 565.2579437 | 565.2579437 | 851.19 | <.0001 |
| Sat. vs Sun. | 1 | 34.1376941 | 34.1376941 | 51.41 | <.0001 |
| Mon. vs Tue. | 1 | 3.5921517 | 3.5921517 | 5.41 | 0.0200 |
| Tue. vs Wed. | 1 | 0.6033771 | 0.6033771 | 0.91 | 0.3405 |
| Wed. vs Thu. | 1 | 6.7420803 | 6.7420803 | 10.15 | 0.0014 |
| Mon. vs Fri. | 1 | 43.1139781 | 43.1139781 | 64.91 | <.0001 |

The Kruskal–Wallis *H*-test results are also summarized in Table 3-2, which

includes the factor test results.  The chi-square value associated with each factor is used

to test the null hypothesis that the treatment (or level or group) means are equal.  *Pr >*

chi-square gives the level of significance associated with the chi-square value. Table 3-3 lists the Friedman $F_r$ test output in the Statistical Package for the Social Sciences (SPSS) statistical software. The Row Mean Scores Differ value is the same as the Friedman's chi-square value. The detailed discussions of all factor test results are described in the next section.

Table 3-2: Results of Kruskal–Wallis $H$-Test

| Factors | DF | Chi-Square Value | $Pr$ > Chi-Square |
|---|---|---|---|
| Severity | 2 | 3122.1043 | <.0001 |
| Year | 15 | 1802.5280 | <.0001 |
| Weekday_Group | 3 | 3995.1749 | <.0001 |
| Month_Group | 3 | 44.0005 | <.0001 |
| Hour_Group | 6 | 2195.6960 | <.0001 |
| County_Group | 2 | 1636.5245 | <.0001 |
| Functional Class | 10 | 1074.2189 | <.0001 |
| Driver Age_Group | 8 | 1559.6965 | <.0001 |
| Driver Race | 3 | 4437.6859 | <.0001 |
| Driver Gender | 1 | 110.2614 | <.0001 |
| Weekday | 6 | 4051.6560 | <.0001 |
| Weather | 4 | 437.6859 | <.0001 |

Table 3-3: Results of Friedman $F_r$ Test

Summary Statistics for Hour_Group by AvgOfOccupants

Controlling for COUNTY

Cochran-Mantel-Haenszel Statistics (Based on Rank Scores)

| Statistic | Alternative Hypothesis | DF | Value | Prob |
|---|---|---|---|---|
| 1 | Nonzero Correlation | 1 | 11.4307 | 0.0007 |
| 2 | Row Mean Scores Differ | 6 | 190.4363 | <.0001 |

Total Sample Size = 468

35

3.4. Identification of Factors Affecting AVO

3.4.1. Accident Year

Figures 3-1(a), 3-1(b), and 3-1(c) show the AVO trends by year (from 1990 to 2005) based on all accidents, weekday accidents, and weekend accidents, respectively. It can be seen from all three figures that AVOs have continued to drop over the period. The AVOs remained relatively stable during the second half of the 1990s, but started to drop again from the year 2001 onward. The ANOVA test results ($F$ value=659.73, $Pr>F < 0.0001$) indicated that there were significant differences in AVOs from the different years, which means that the decrease in AVO over the years has been significant overall. The chi-square statistics (chi-square = 1802.528, $Pr>$chi-square $< 0.0001$) for the Kruskal–Wallis $H$-test confirmed this conclusion. Figures 3-1(b) and 3-1(c) also show that the weekend AVOs are consistently higher than the weekday AVOs.
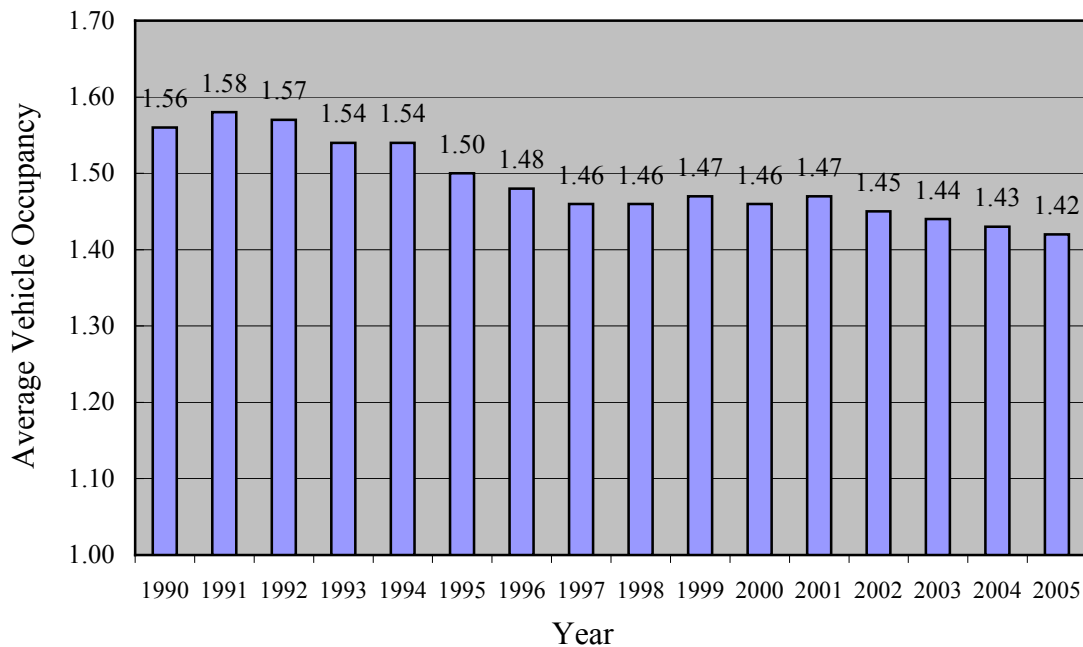


Figure 3-1(a): Daily AVO Variations by Year

36
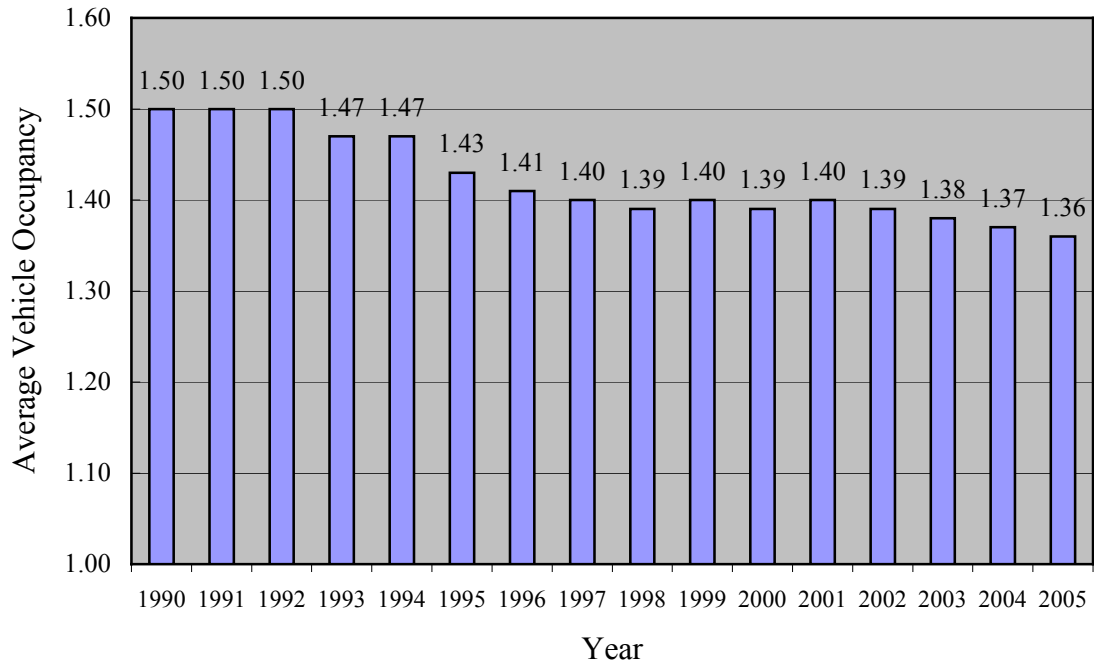
Figure 3-1(b): Weekday AVO Variations by Year
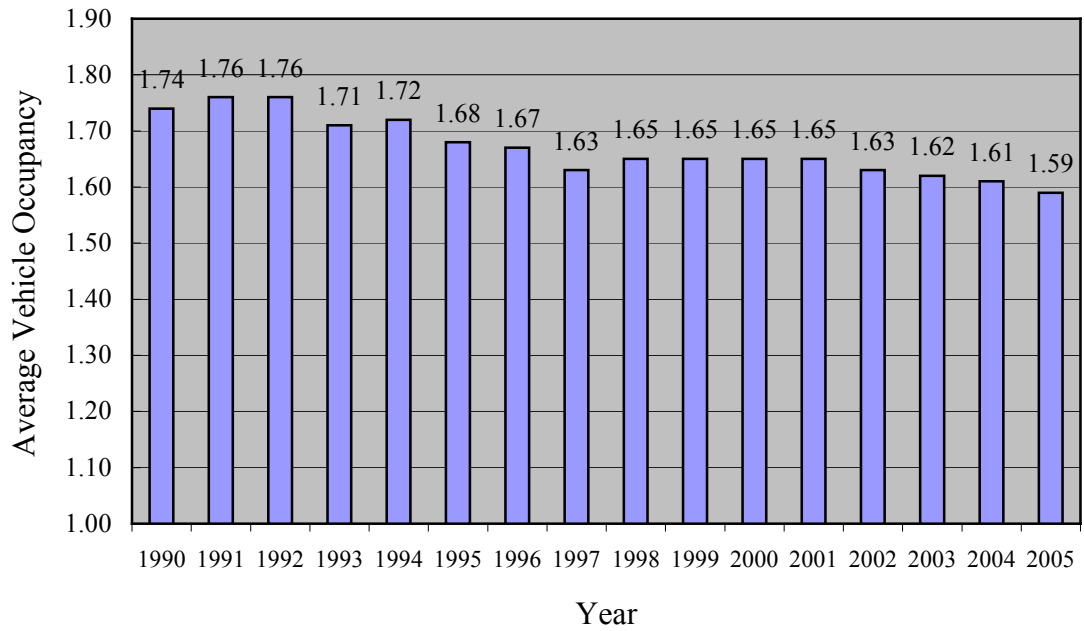


Figure 3-1(c): Weekend AVO Variations by Year

## 3.4.2. Accident Month

Figure 3-2 shows the AVO trend by accident month based on the 2005 statewide accident data. The figure shows that the month of July experienced the highest AVO, while the months of August, September, and November experienced the lowest AVOs. Two other months that also experienced a higher AVO are March, which experienced traffic from spring breakers, and December, which experienced traffic from holiday shoppers and vacationers. Both the nonparametric and ANOVA tests indicate significant differences in AVO among the months of the year. The pair test result ($F$ value=0.121, $Pr>F$ = 0.7250) in Table 3-1 shows that there is no significant difference between the winter months (December through March) and the spring months (April through June).
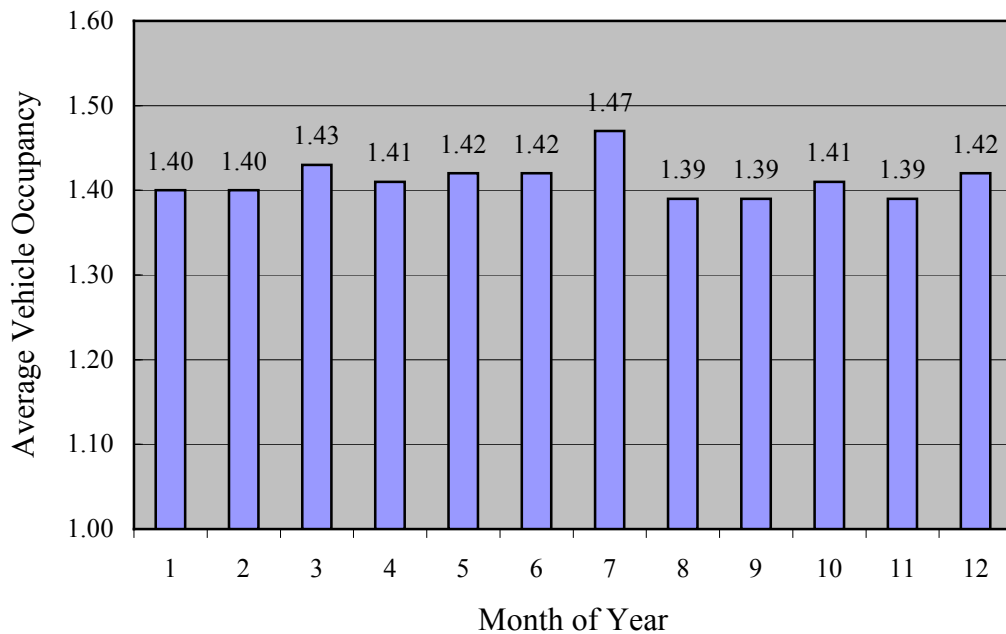


Figure 3-2: AVO Variations by Month

3.4.3. Accident Day-of-Week

Figure 3-3 shows the AVO trend by the accident day-of-week, also based on the year 2005 statewide accident data. It is clear from the distribution that the weekday AVOs tend to be significantly lower than the weekend AVOs. The figure also shows that Sundays experienced the highest AVO, followed by Saturdays and Fridays. While Mondays and Fridays are generally excluded from data collection that aims to obtain data that are representative of typical weekdays, the distribution from Figure 3-3 shows that, for the purposes of vehicle occupancy data collection, Mondays may not be that different from Tuesdays, Wednesdays, and Thursdays, which are typically days used for data collection. The AVO on Fridays is normally higher than that on other weekdays, but remains significantly lower than the AVO for weekends. The ANOVA pair test result ($F$ value=0.91, $Pr>F = 0.3405$) indicates that AVOs on Tuesdays and Wednesdays are similar and are generally lower than other days of the week.
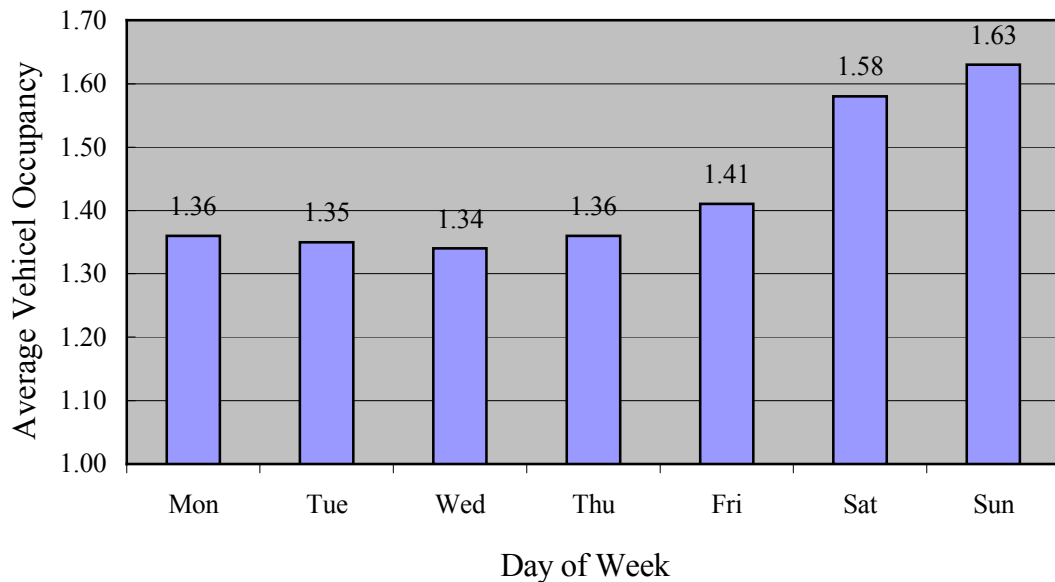


Figure 3-3: AVO Variations by Day-of-Week

3.4.4. Accident Hour

Figure 3-4 shows that the AVOs are generally low in the early morning, increase after the morning peak period, remain level during the middle of the day, and reach a high point in the evening. The figure also shows that more vehicles involved home-based work trips during the a.m. peak than during the p.m. peak, and that vehicles have the highest occupancy on average during the evening hours. The ANOVA test result confirmed significant differences in AVOs among the time-of-day intervals at a 95% significance level. The nonparametric Friedman $F_r$ test was also used to test for significant AVO differences across time-of-day intervals. Each county was treated as a block, AVOs for time-of-day intervals were ranked within each block, and rank sums of each of the seven levels (time-of-day intervals) were computed. The Friedman $F_r$ test result (Row Mean Scores Differ value = 190.4363 and $p$-value <0.0001) in Table 3-3 also confirms the distributions of the AVO for hour groups are significantly different.
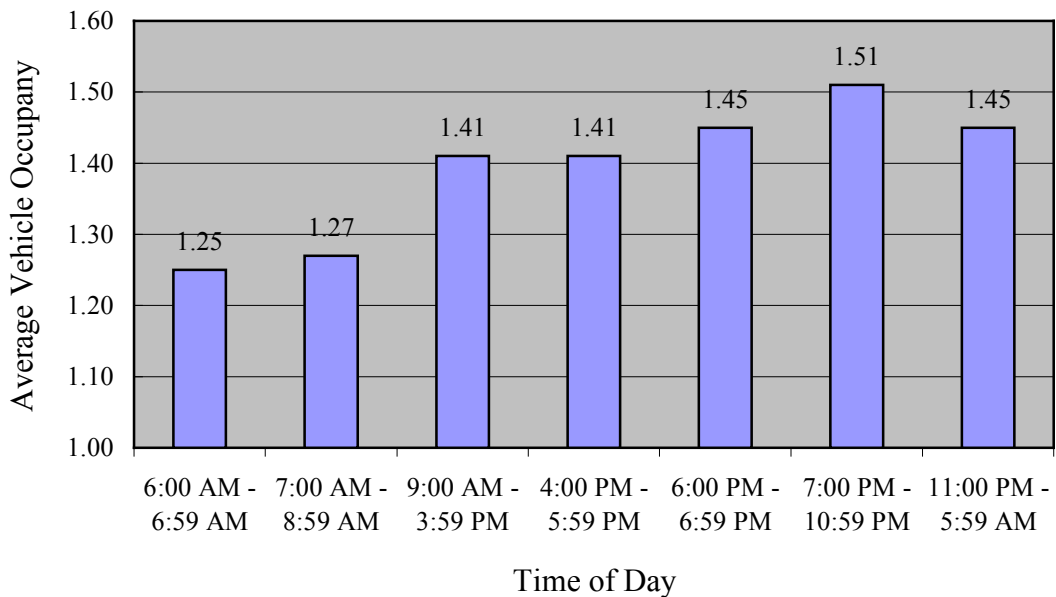


Figure 3-4: AVO Variations by Hours

3.4.5. County

In this analysis, all 67 counties in Florida were classified into three urban-size groups: 1) large urban counties with county populations over 500,000; 2) medium-size counties with county populations under 500,000 and over 100,000; and 3) rural counties with populations under 100,000.  Figure 3-5 shows that the highest AVOs are present in small rural counties, and that the lowest AVOs are present in large urban counties, with medium-size urban counties ranked in the middle.  The ANOVA test result ($F$ value for this factor is 256.87, $Pr>F < 0.0001$) in Table 3-1 and the Kruskal–Wallis $H$-test result (chi-square = 1636.5245, $Pr>$chi-square $< 0.0001$) in Table 3-2 both showed that the differences of AVOs for large urban counties, medium urban counties, and rural counties are significant.
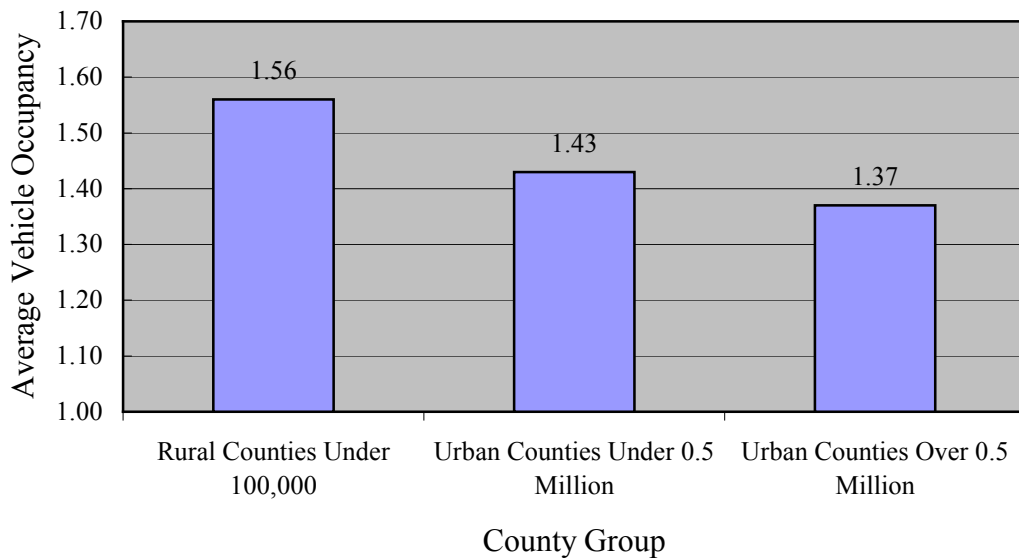


Figure 3-5:  AVO Variations by County Group

3.4.6. Roadway Type

Figure 3-6 shows the AVOs for four roadway functional classifications for rural and urban areas separately. In general, urban roadways with higher functional classifications experienced lower AVO, as these roadways serve a higher proportion of single-occupant work trips. Interestingly, the trend is generally reversed in the case of rural roadways, as rural roadways with higher functional classifications in Florida serve a high proportion of tourists and visitors who tend to travel in groups. The ANOVA test result ($F$ value for this factor is 50.62, $Pr>F < 0.0001$) in Table 3-1 and the Kruskal–Wallis $H$-test result (chi-square $= 1074.2189$, $Pr>$chi-square $< 0.0001$) in Table 3-2 confirmed that the differences among the AVOs are significant.
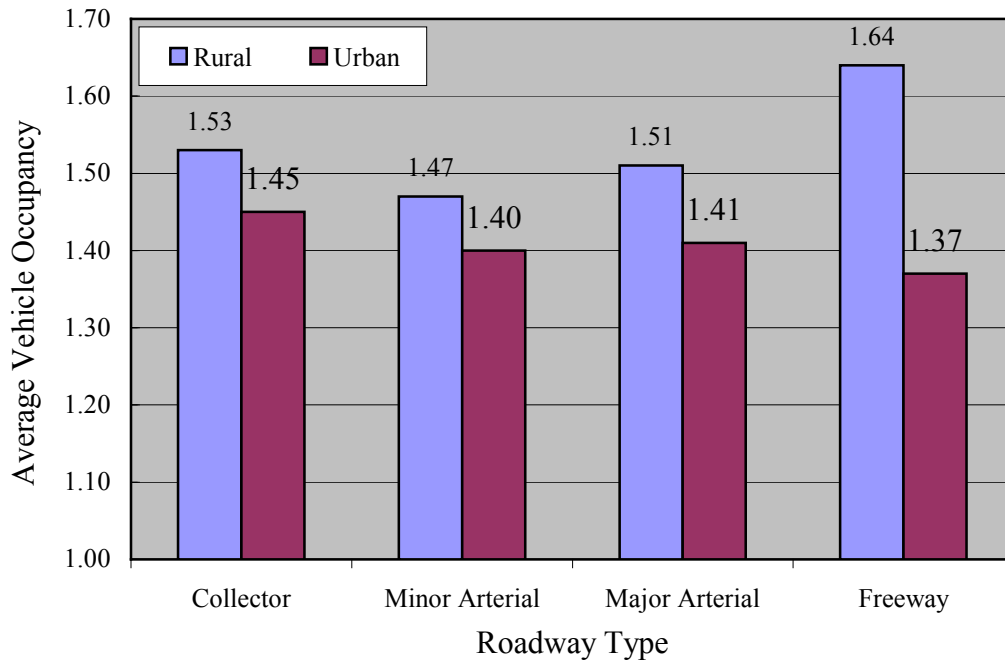


Figure 3-6: AVO Variations by Roadway Type

## 3.4.7. Accident Severity

Each accident can be classified into one of three categories: property damage only (PDO), injury, and fatality. Figure 3-7 shows that fatality accident vehicles have the highest AVOs by a significant margin and that PDO accident vehicles have the lowest vehicle occupancy rates. The statistical tests also indicated that these differences are significant. These results were expected because multiple-occupant vehicles have a higher likelihood of injury and fatality than single-occupant vehicles in a given accident. The results from both the ANOVA test ($F$ value for this factor is 1167.75, $Pr>F <$ 0.0001) in Table 3-1 and the Kruskal–Wallis $H$-test (chi-square = 3122.1043, $Pr$>chi-square < 0.0001) in Table 3-2 confirmed that the differences among the AVOs are significant.
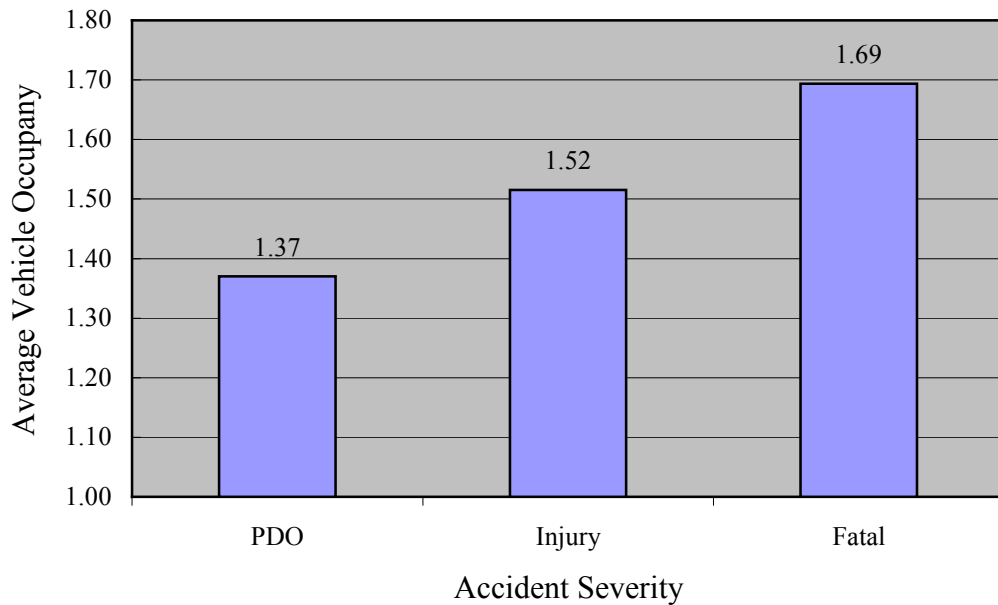


Figure 3-7:  AVO Variations by Accident Severity

3.4.8. Driver's Age

Figure 3-8 shows that the AVOs vary between different age groups. In general, the younger driver groups tend to have higher AVOs than the older driver groups, with the youngest driver group (16-19 years old) having the highest AVO. This is followed by the 30-39 and 25-29 driver age groups, respectively, which is likely due to a higher percentage of active parents with children in these age groups. The AVO dropped significantly for drivers in the 50-59 age group (i.e., the "empty nest" group) and then increased again in the 60-69 and 70-79 retired age groups. The ANOVA test result ($F$ value for this factor is 185.10, $Pr>F < 0.0001$) and the Kruskal–Wallis $H$-test result (chi-square = 1559.6965, $Pr>$chi-square $< 0.0001$) confirmed that the AVOs for different driver age groups are significantly different.
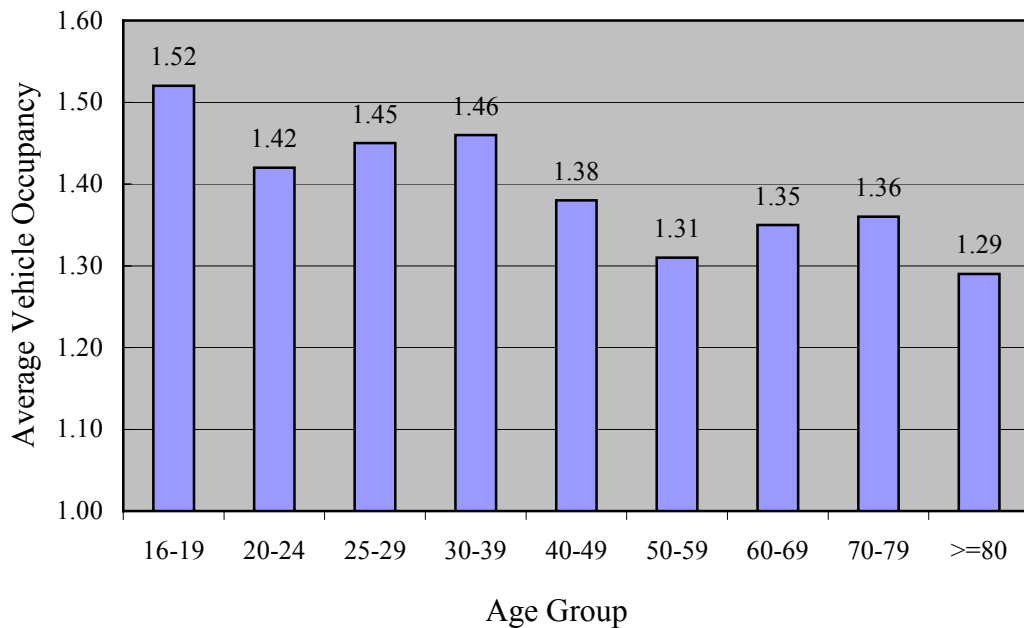


Figure 3-8: AVO Variations by Driver's Age

3.4.9. Driver's Gender

Figure 3-9 shows that female drivers tend to have more passengers in their vehicles than male drivers. This observation was consistent throughout the different days of a week. One explanation for this travel behavior is that female drivers are more likely to be the ones to take their children to and from schools, daycare centers, doctor offices, etc. The ANOVA test result in Table 3-1 and the Kruskal–Wallis $H$-test result (chi-square = 110.2614, $Pr$>chi-square < 0.0001) confirmed that AVOs for vehicles driven by female drivers are significantly higher than AVOs for vehicles driven by male drivers.
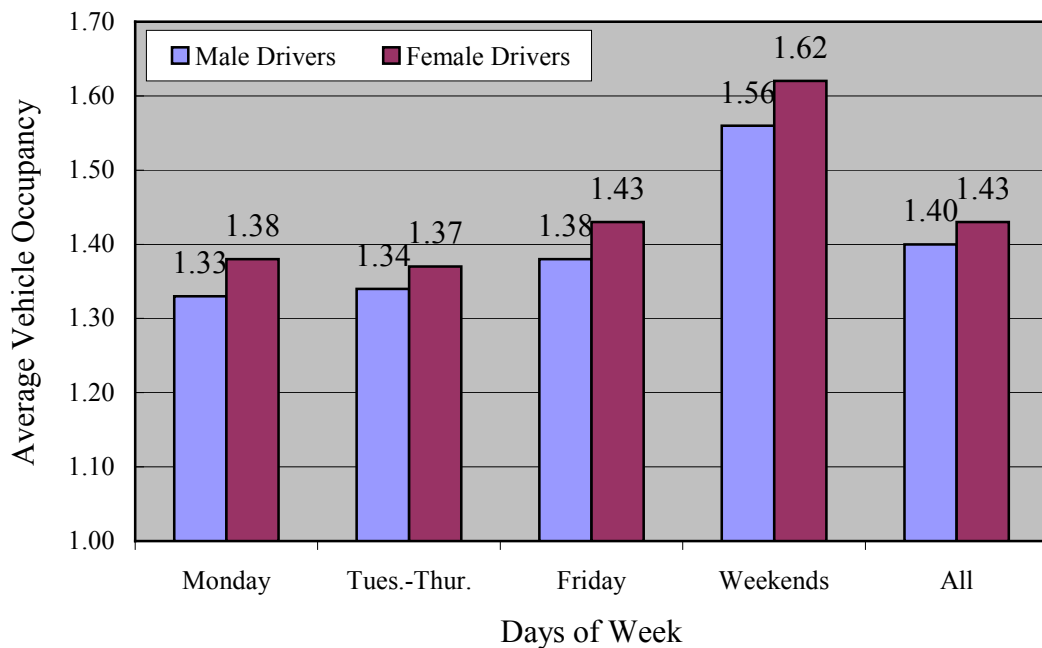


Figure 3-9: AVO Variations by Driver's Gender

3.4.10. Driver's Race

Figure 3-10 shows the AVOs based on the driver's race. It shows that white drivers are more likely to be in a single-occupant vehicle than others, while blacks and Hispanics are more likely to carpool. This travel behavior is likely a direct result of the

generally better economic well-being of the white population, which contributes to a higher vehicle ownership level and, thus, lowers their AVO. The ANOVA test showed that the difference is significant ($Pr>F < 0.0001$) at the 95% confidence level.
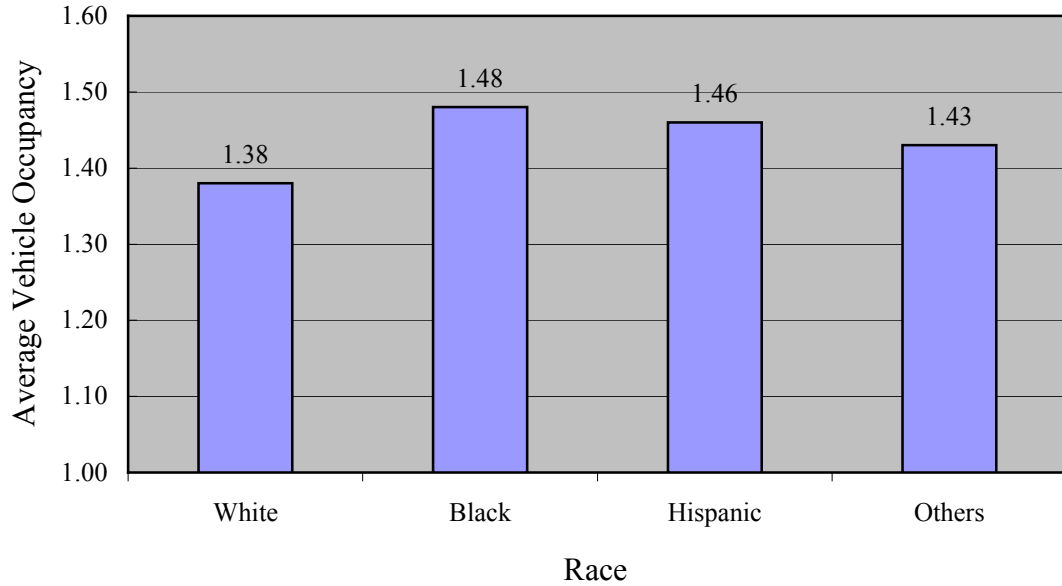


Figure 3-10: AVO Variations by Driver's Race

## 3.4.11. Weather Conditions

Figure 3-11 shows the AVO trend by weather conditions. This figure shows that the vehicles involved in accidents that occurred during rain had a slightly higher AVO than vehicles involved in accidents during clear or cloudy conditions. The Kruskal–Wallis *H*-test result (chi-square = 437.6859, *Pr*>chi-square < 0.0001) indicated that the AVO difference between different weather conditions is significant. Also, these vehicles occupy about 10% of the total vehicle population.
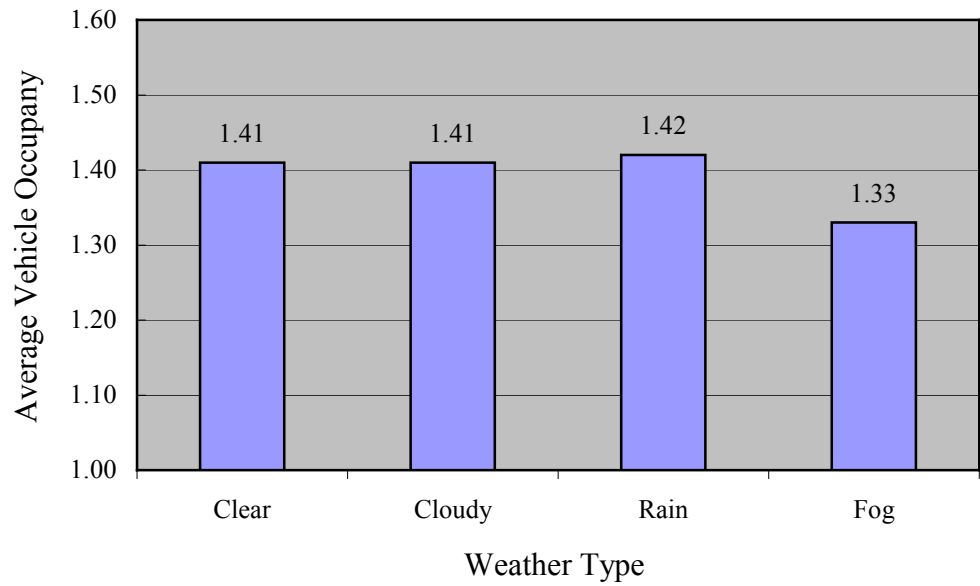
Figure 3-11: AVO Variations by Weather Type

CHAPTER 4

ADJUSTMENTS FOR POTENTIAL BIASES IN ACCIDENT DATA

4.1. Introduction

As mentioned in Chapter 1, a major problem associated with the use of accident data for vehicle occupancy has been the potential biases resulting from over-representation of younger drivers, alcohol- and drug-related accidents, and fatal accidents (Heidtman *et al.*, 1997; Ahuja and Hanscom, 1996; Chang and Mannering, 1998). The premise is that these types of accidents tend to involve vehicles with higher occupancies, thus causing the AVO to be overestimated. This chapter examines three potential factors, namely, accident severity, driver's age, and driver's gender; develops the corresponding adjustment factors as needed; and discusses other potential factors that may affect the estimates of AVO from accident data. While other potential factors such as driver's race and weather conditions are likely to also contribute to similar biases, adjustment for these factors cannot currently be performed due to the lack of exposure data needed to derive the adjustment factors.

4.2. Data Preparation

The Vehicle Occupancy Collection Methods research (Phases I and II) were funded by the Research and Development Center of the Florida Department of Transportation (FDOT). In the Phase I study, the different methods available to estimate vehicle occupancy were reviewed and summarized, and the procedural guidelines based on a statistical sampling technique were established for the manual counting methods

(Gan *et al*., 2005).  The Phase II study demonstrates the entire process of determining the sample size, selecting the survey locations, performing the actual data collection, and applying rigorous statistical methods to analyze the field collected data and accident vehicle data (Gan *et al*., 2007).   The field vehicle occupancy data collected during the Phase II study are used to examine the potential bias factors in this chapter.  These data are also used in this study to compare with those extracted from traffic accident records in Chapter 6.  In addition to the field collection data, the census demographic data were also used to examine the potential bias factors associated with a driver's age, race, and gender.

4.2.1. Field Data Collection Locations

To establish a statistically sound baseline from which future vehicle occupancy evaluations can be performed, an initial area-wide study was conducted in Phase II. Miami-Dade County was selected as the study area for field data collection.  For this study, the target collection population was divided into strata by roadway facility, including surface streets, freeways, and toll facilities.

Based on the standard deviation values derived from the data collected by URS (1997), as well as the sampling scheme developed by Ferlis (1981) that was discussed in Chapter 2, to attain a 95% confidence level within a precision level of ±0.03 for the area-wide study, a minimum sample size of 44 was required for the area-wide study.  To facilitate the arrangement of data collection tasks, the sample size was increased to 48. The 48 observation locations were further divided into 24 sites on surface streets, 12 sites

on freeways, and 12 sites at toll facilities. The detailed addresses of these locations can be seen in Tables 4-1, 4-2, and 4-3.

Table 4-1: Initial Selected Observation Roadway Segments for Surface Streets

| Index | 2000 Census Tract ID | Survey Roadway | Intersecting Roadways |
|-------|----------------------|----------------|-----------------------|
| 1 | 116 | NW 72nd Ave | NW 36th St and NW 31st St |
| 2 | 34 | NW 116th Way | NW South River Dr and NW 100th Rd |
| 3 | 53 | Douglas Rd Et | Langley Rd and Alibaba Ave |
| 4 | 238 | S Dixie Hwy | SW 24th Ave and SW 22nd Ave |
| 5 | 317 | SW 184th St | SW 132nd Ave and SW 127th Ave |
| 6 | 45 | NW 57th Ave | NW 165th Te and NW 159th St |
| 7 | 309 | SW 152nd St | SW 152nd Ave and SW 148th Ct |
| 8 | 52 | NW 151st St | NW 27th Ave and NW 24th Ave |
| 9 | 247 | SW 37th Ave | SW 28th St and Bird Rd |
| 10 | 288 | SW 88th St | SW 79th Ave and SW 77th Ave |
| 11 | 307 | SW 112th St | SW 97th Ave and SW 94th Ave |
| 12 | 192 | NW 37th Ave | NW 3rd St and W Flagler St |
| 13 | 314 | Old Cutler Rd | SW 157th Te and SW 168th St |
| 14 | 97 | NW 17th Ave | NW 103rd St and NW 95th St |
| 15 | 220 | Granada Bd. | SW 8th St and Venetia Te. |
| 16 | 91 | NW 103rd St | NW 35th Ave and NW 32nd Ave |
| 17 | 241 | SW 82nd Ave | SW 24th St and SW 28th St |
| 18 | 356 | S Dixie Hwy | SW 380th St and SW 384th St |
| 19 | 283 | SW 127th Ave | SW 76th St and SW 80th St |
| 20 | 22 | NE 190th St | NE 29th Ave and Country Club Dr |
| 21 | 76 | NE 2nd Ave | NE 117th St and NE 114th St |
| 22 | 8 | NW 215th St | Florida TP and NW 7th Ave |
| 23 | 134 | N Miami Ave | NE 57th St and NE 54th St |
| 24 | 269 | SW 117th Ave | SW 64th St and SW 72nd St |

Table 4-2: Selected Observational Roadway Segments for Freeways

| Index | Survey Roadway | Crossing Roadways |
|---|---|---|
| 1 | Julia Tuttle CY | Biscayne Blvd. and N Bay Rd |
| 2 | SR 826 EX | W 63rd St and W 52nd St |
| 3 | SR 826 EX | NW 17th St and NW 15th St |
| 4 | SR 826 EX | NW 48th St and NW 43rd St |
| 5 | SR 836 EX | NW 36th Ave and NW 31st Ave. |
| 6 | SR 826 EX | NW 12th St and NW 7th St |
| 7 | SR 826 EX | SW 75th St and SW 85th St |
| 8 | SR 836 EX | NW 57th Ave and Perimeter Rd. |
| 9 | SR 826 EX | NW 41st St and NW 36th St |
| 10 | SR 836 EX | NW 111th Pl. and NW 107th Ave(Ramps) |
| 11 | SR 826 EX | NW 35th Te and NW 34th St (Ramps) |
| 12 | I-95 | NW 5th Ct and SR 836 EX |

Table 4-3: Selected Observation Roadway Segments for Toll Facilities

| Index | Survey Roadway | Crossing Roadways |
|---|---|---|
| 1 | SR 821 EX | SW 154th St and SW 180th St |
| 2 | SR 821 ET | SW 134th St and Louis St |
| 3 | SR 821 HY | SW 24th St and SW 35th Te |
| 4 | SR 874 EX | SW 56th St and SW 66th St |
| 5 | Florida TP | NW 183rd St and NW 170th St |
| 6 | Florida TP | NW 202nd St and W 194th St |
| 7 | SR 821 ET | SW 106th Ave and SW 112th Te |
| 8 | SR 874 EX | SW 117th St and SW 120th St |
| 9 | Florida TP | NW 194th St and NW 186th St |
| 10 | Opa Locka EX | NW 135th St and NW128th St |
| 11 | Florida TP | NW 186th Te and NW 183rd St |
| 12 | Florida TP | Near NW 27th Ave |

## 4.2.2. Field Data Collection Methods

It was initially determined that the roadside windshield method would be used for the entire data collection effort.  However, it was later determined that this method would have presented safety concerns were the survey crew stationed on high-speed freeways, and so the carousel method was used on some freeways.  It was also determined that, where toll facilities are concerned, the observers would have the option of locating themselves near toll plazas; thus, the windshield method could be used in these instances.

To conserve manpower resources, a systematic short-count procedure was applied.  Observations were made of multiple lanes at a time for a fixed interval during each hour.  This involved scheduled shifts from direction to direction.  One observer for each observation session was preliminarily assigned.  To examine the feasibility of this arrangement, a field data collection test was conducted on one of the selected segments for surface streets.  The heavy traffic during peak periods not only placed an intensive workload on the observer, but also increased the chances of the observer recording data incorrectly.  The observers worked in two-person teams, and each observer collected data for one travel direction.  Observations were made of multiple lanes in each direction for 20- or 40-minute intervals during each hour.  A 20-minute interval was reserved for reviewing the counts taken and for rest.  Observations were made from 7:00 a.m. to 6:00 p.m. Data were collected only for passenger vehicles (including pick-up trucks).  Data were collected on Tuesdays, Wednesdays, and Thursdays only.

4.2.3. Field Data Collection Tool

  To facilitate field data collection, a Pocket PC application tool was developed. The tool attempts to eliminate the need for manual data post-processing by allowing the user to make use of the touch-screen interface on a Pocket PC to record the number of occupants and time for vehicles. This Pocket PC application tool is a typical Windows installation designed to run on the Microsoft Windows CE operating system. The system can work with any Pocket PC that has a minimum viewing screen of at least 3.5".

  Figure 4-1 shows the data entry screen. The screen provides buttons to enter up to seven occupants in a vehicle. In the case when a vehicle has more than seven occupants, which is extremely rare for passenger cars, the user may record seven occupants for the current vehicle and add the difference to the total occupants of the next vehicle.
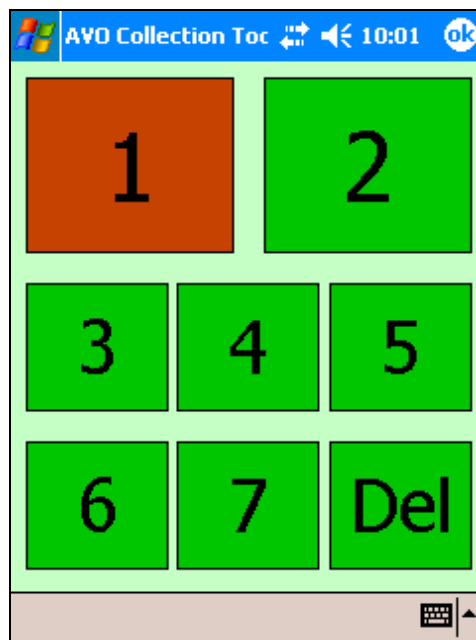


Figure 4-1: Data Entry Screen

4.2.4. Other Data Sets

The 2000 census demographic data, which were categorized or cross-categorized into age, race, and gender groups, were used to identify and adjust potential biases. Only the population with an age greater than 15 is used in this analysis. The years 2000 and 2005 Miami-Dade County accident data are also used in this chapter.

4.3. Examination of Potential Bias Factors

4.3.1. Accident Severity

The accident data from Miami-Dade County during the year 2005 were used to test the potential bias in AVO estimation due to accident severity. Miami-Dade County was used due to the availability of field vehicle occupancy data, which were collected as part of this study. Table 4-4 lists the distribution of accident severity by vehicle occupancy. Of the 54,026 vehicles involved in accidents, the data show that multi-occupant vehicles have a tendency to produce more severe accidents than single-occupant vehicles. For example, injury and fatal accidents account for over 55% and 0.56%, respectively, of all multi-occupant vehicle accidents. In contrast, the total accident count (single- and multi-occupant accidents) shows that, overall, only about 46% of all accidents involved injury and only 0.41% of all accidents involved fatality. These statistics support the argument that vehicle occupancy affects accident severity. The data show that the AVOs of vehicles involved in accidents involving fatalities and injuries are significantly higher than those of vehicles involved in PDO accidents. There are two possible reasons for this relationship, the first of which is fairly obvious: an increase in the number of people in a vehicle increases the likelihood of injury and fatality because

the number of occupants exposed to the risk of being injured or killed increases with higher occupancies. Also, in vehicles with lower occupancies, unoccupied seat space provides protection from intrusion.

Table 4-4:  Distribution of Accident Severity by Vehicle Occupancy

| No. of Occupants | PDO | Injury | Fatality |
|---|---|---|---|
| 1 occupant | 23,379 (55.87%) | 18,315 (43.77%) | 148 (0.35%) |
| 2 occupants | 3,400 (43.40%) | 4,390 (56.04%) | 44 (0.56%) |
| 3 occupants | 1,167 (43.03%) | 1,527 (56.31%) | 18 (0.66%) |
| 3+ occupants | 711 (43.46%) | 912 (55.75%) | 13 (0.79%) |
| All | 28,657 (53.04%) | 25,146 (46.54%) | 223 (0.41%) |

Awareness of the relationship between occupancy and accident severity is the key to being able to use occupancy data from accident reports to estimate occupancy rates. As mentioned, some past studies have shown that accident records significantly overestimate vehicle occupancy (Asante *et al*., 1996).  This occurs if the accident data over-represent multi-occupant vehicles.  However, a problem also arises if the field vehicle data used for comparison with the accident vehicle data under-represent multi-occupant vehicles because observers in the field have unclear sight and/or limited time to count the number of occupants in moving vehicles.  To examine this problem, Miami-Dade County 2006 vehicle occupancy field data were compared with vehicle occupancy data derived from Miami-Dade County 2005 accident vehicles.

Table 4-5 shows a comparison of vehicle occupancy distribution between accident data and field data.  It can be seen from the table that, for a single occupant/driver, the vehicle percentages for both data sets are very close.  This suggests that single-occupant vehicles did not under-represent nor over-represent the vehicle population in accident

data. In other words, although multi-occupant (two and more than two occupants) vehicles were involved in higher percentages in severe accidents (injury and fatality) than were single-occupant vehicles (see Table 4-4), the number of multi-occupant vehicles involved in accidents in general occupied almost the same percentage as that in the field data set. This suggests that multi-occupant vehicles are not over-represented in the accident database. However, high-occupancy (three and more than three) vehicles do appear with higher frequency in accident data than in field data. Obviously, there are other potential explanations for this difference, one of which might be that observers using the windshield method missed children and infants, who were difficult to see.

Table 4-5: Vehicle Occupancy Distributions of Accident and Field Data

| Roadway Type | Number of Occupants | Accident Data (%) | Field Data (%) |
|---|---|---|---|
| Urban Arterials | 1 | 80.62 | 80.52 |
| | 2 | 13.32 | 16.57 |
| | 3 | 4.08 | 1.25 |
| | 3+ | 1.98 | 1.65 |
| Freeways and Expressways | 1 | 84.56 | 83.70 |
| | 2 | 11.0 | 14.55 |
| | 3 | 3.21 | 1.21 |
| | 3+ | 1.23 | 0.54 |
| All | 1 | 81.75 | 81.74 |
| | 2 | 12.57 | 15.80 |
| | 3 | 3.83 | 1.23 |
| | 3+ | 1.85 | 1.23 |

4.3.2. Driver's Age

To detect potential bias resulting from driver's age, the year 2000 Miami-Dade County accident data were used to determine if the age distribution of drivers in the

56

accident database was different from the age distribution for the entire driving population in Miami-Dade County. This can be analyzed by comparison with the 2000 census data for the same area. In this analysis, the driving population is defined as persons 16 years of age or older. A comparison of the distribution of ages obtained from both accident and census data was conducted and is shown in Figure 4-2. This figure indicates a significant difference between the ages of drivers involved in accidents and the ages of the general driving population in Miami-Dade County. Specifically, younger drivers were more likely to be involved in accidents than older drivers. Figure 4-3 shows that the AVOs vary significantly with the ages of drivers in the Miami-Dade County area. The figure shows that the AVOs for the youngest drivers are the highest and continue to drop over the age 20-50 groups. The AVOs are lower for drivers over 50 years old than they are for other groups.
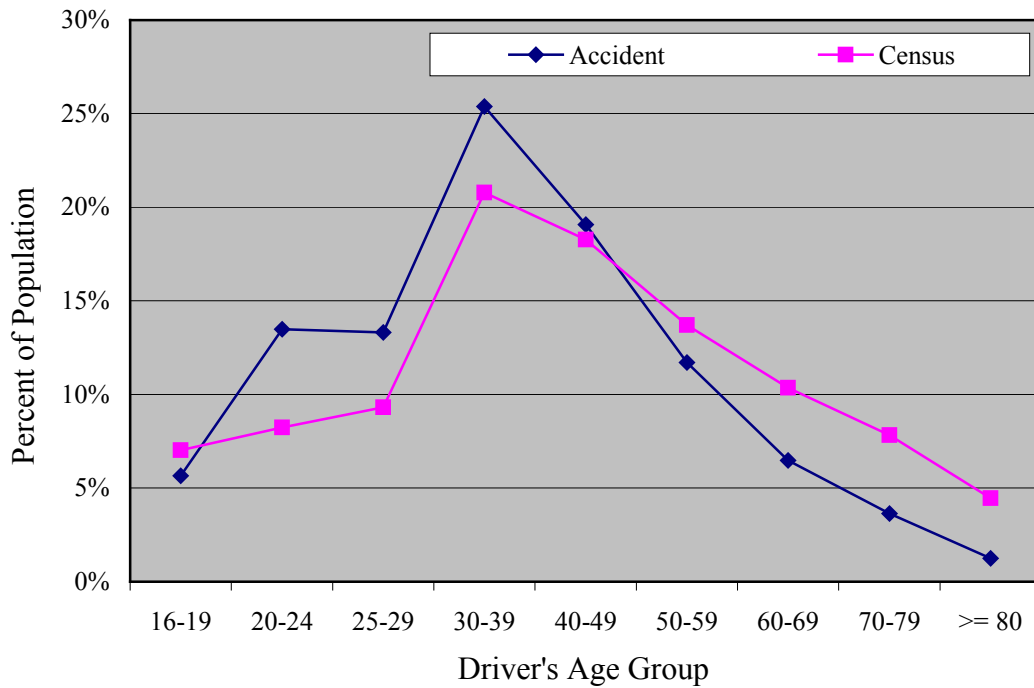


Figure 4-2: Distributions of Driver's Age for Miami-Dade County
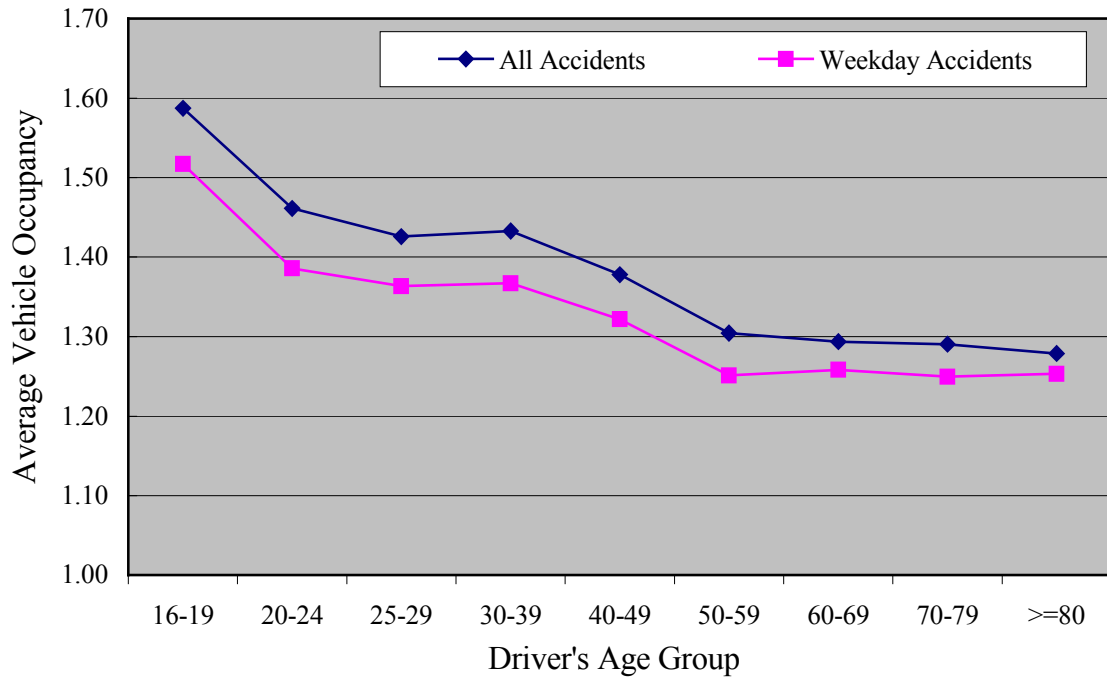
Figure 4-3:  Average Vehicle Occupancy by Driver's Age for Miami-Dade County

### 4.3.3. Driver's Gender

Figure 4-4, reproduced from Figure 3-8 of Subsection 3.4.9, shows that female drivers tend to have more occupants in their vehicles than male drivers and that this observation is consistent throughout the different days of the week.

Table 4-6 lists year 2000 accident vehicle frequency comparisons between male drivers and female drivers in the state of Florida.  On the whole, male drivers get in accidents more frequently than female drivers, especially during weekends.  Based on 2000 Census demographic data, the male population percentage in the state of Florida is 48.21%, and the female population percentage is 51.79% (see Figure 4-5).  Hence, if the vehicles driven by male drivers are over-represented in the whole accident vehicle population, the overall AVO value estimated from the accident records would have a bias.
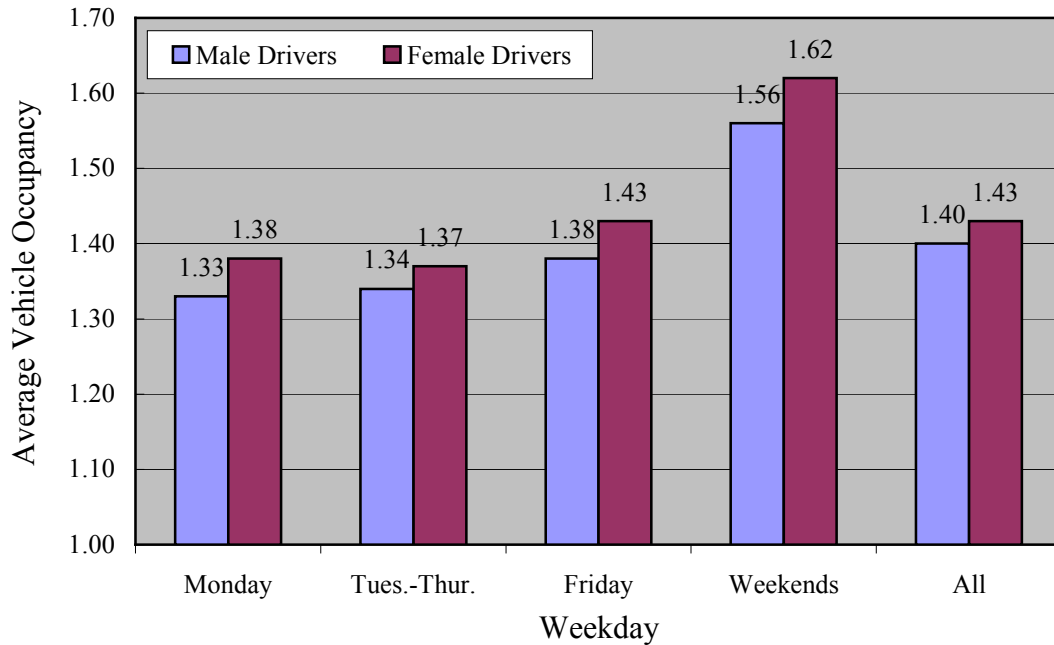
Figure 4-4:  AVO Comparisons between Male and Female Drivers

Table 4-6:  Vehicle Frequency Comparisons Between Male and Female Drivers

| Weekday Type | Male Driver | | Female Driver | |
|---|---|---|---|---|
| | Number | Percentage | Number | Percentage |
| Monday | 19,472 | 58.43% | 13,856 | 41.57% |
| Tuesday through Thursday | 58,467 | 57.83% | 42,637 | 42.17% |
| Friday | 24,633 | 59.44% | 16,811 | 40.56% |
| Weekends | 36,268 | 63.54% | 20,813 | 36.46% |
| All | 138,840 | 59.60% | 94,117 | 40.40% |

4.3.4. Other Potential Bias Factors

Other potential factors, including the driver's race and weather conditions, have been shown to have an impact on AVO.  Over-representation of one of these subgroups in traffic accidents may distort the AVO estimates from accident data.  However, the lack of detailed racial data and detailed weather information for different sub-groups prevents the adjustment factors from being developed.
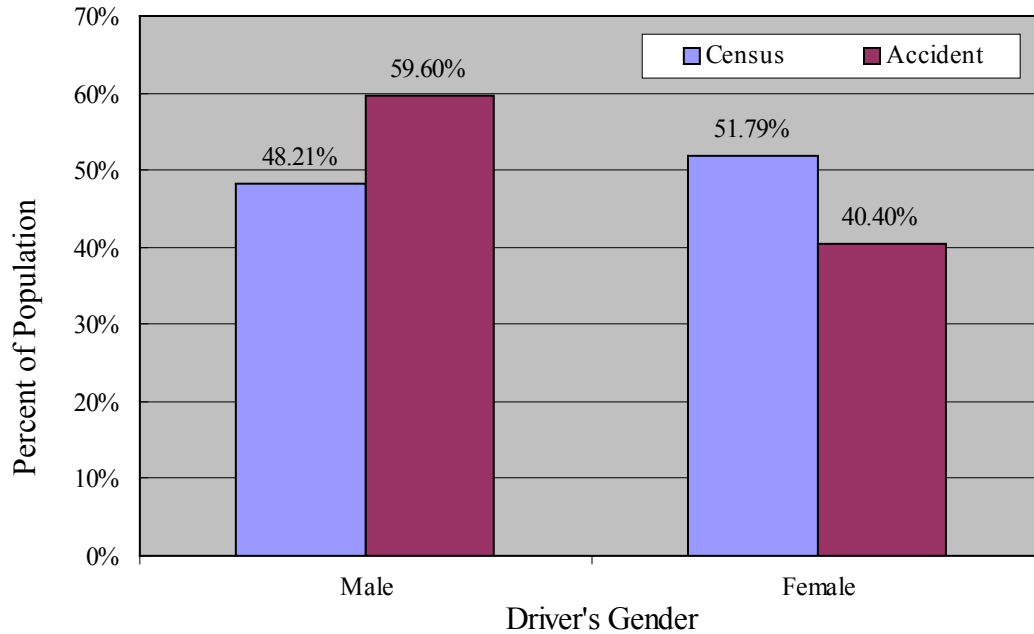
Figure 4-5: Distributions of Driver's Gender from Accident and Census Data

## 4.4. Development of Bias Adjustment Factors

While it is important to obtain adjusted AVOs to correct for potential biases for such applications as converting person trips to vehicle trips in the traditional demand forecasting process, it may not be as important for such adjustments to be made for the purpose of monitoring AVO trends in a study area. This is because, for the latter, it is the relative values of AVOs, rather than absolute AVOs, that are important. This assumes that the bias factors do not change over time or the range of variables considered.

### 4.4.1. Improved Weighting Mean Method

Heidtman *et al.* (1997) used a weighted mean procedure using the census data to adjust for the driver's age bias in the Chicago area. This procedure would weight each age group's AVO in proportion to its age group's contribution to the overall driving

population.  During this weighting procedure, the census population data for the Chicago area are grouped and calculated for each age group.  The ratio of each age group's population to the whole driving population is then obtained as a weighting coefficient.  To obtain the overall AVO estimate, the weighted mean procedure will weight each age group's AVO by using that group's weighting coefficient.  Hence, the sum of each age group's weighted AVO is the overall adjusted AVO.  This procedure can be represented using the following formulas:

$$AVO_{adj} = \sum_i \left(W_i \times AVO_i\right) \tag{4-1}$$

$$W_i = \frac{P_i}{\sum_k P_k} \tag{4-2}$$

where

$AVO_{adj}$ = the overall adjusted AVO,

$W_i$ = the weighting coefficient for driver age group $i$,

$AVO_i$ = the unweighted AVO for driver age group $i$, and

$P_i$ = the driving population for driver age group $i$.

The assumptions for this procedure include: 1) that each age group has a uniform percentage of vehicle driving activity and 2) that each age group has a uniform accident rate in terms of vehicle miles driven and driving population.  However, the driving activity for each age group's drivers was different (U.S. DOT, 2001).  Based on U.S. DOT 2001 national household travel survey result, Table 4-7 lists the travel minutes and miles differences based on adult demographic characteristics for age and gender.

Overall, for all adults, including nondrivers and those who may not have driven on a given day, 55 minutes are spent behind the wheel driving 29 miles a day.  Usually

men drive further (38 vs. 21 miles) and longer (67 vs. 44 minutes). Teenagers drive the least on an average day (about 25 minutes and 12 miles). Adults over the age of 65 also drive less on an average day compared to other adult age groups (20 to 64 years).

Table 4-7: Minutes Spent and Miles Driven Daily by Age and Gender Groups

|  | Minutes | Miles |
|---|---|---|
| Male | 67.3 | 37.6 |
| Female | 43.8 | 21.2 |
| 15-19 years | 24.6 | 12.2 |
| 20-24 years | 51.7 | 28.9 |
| 25-54 years | 64.1 | 35.0 |
| 55-65 years | 57.7 | 29.7 |
| 65 years and older | 39.3 | 17.0 |
| All | 55.1 | 29.1 |

To reflect these travel characteristic differences on driver's ages and genders, equation 4-2 is revised to the following formula:

$$W_i = \frac{P_i M_i}{\sum_k P_k M_k} \qquad (4\text{-}3)$$

where

$P_i$ = the driving population for driver age group $i$, and

$M_i$ = the average miles driving by age group $i$.

This improved weighting method will be used not only for the driver's age bias adjustment, but also for the driver's gender bias adjustment, in which the AVOs are grouped by driver's gender instead of driver's age.

4.4.2. Bias Adjustment Factors for Driver's Age

Because of the drivers' age bias in the accident data, estimates of AVO derived from accident data should be adjusted to minimize the effect of this bias on AVO calculations. During this weighting procedure, the census population data for each county in Florida are grouped and calculated for each age group. Because the Florida 2001 household travel survey data are not available, the 2001 national household travel survey data are used in this research. Table 4-8 lists each age group's weighting coefficient, the unadjusted AVO, and the weighted AVO for each age group for Miami-Dade County. This table also lists the final overall adjusted AVO, which is 1.39. (See Appendix A.1 for driver's age group weighting coefficients for all counties.)

Table 4-8: Weighting Coefficients for Age Groups for Miami-Dade County

| Age Group | Weighting Coefficient ($W_i$) (a) | Unadjusted $AVO_i$ (b) | Weighted $AVO_i$ (a) x (b) |
|---|---|---|---|
| 16-19 | 0.0294 | 1.591 | 0.0468 |
| 20-24 | 0.0817 | 1.465 | 0.1197 |
| 25-29 | 0.1120 | 1.430 | 0.1602 |
| 30-39 | 0.2586 | 1.439 | 0.3597 |
| 40-49 | 0.2195 | 1.384 | 0.3039 |
| 50-59 | 0.1526 | 1.308 | 0.1996 |
| 60-69 | 0.0832 | 1.294 | 0.1076 |
| 70-79 | 0.0457 | 1.291 | 0.0590 |
| ≥80 | 0.0261 | 1.283 | 0.0334 |
| Overall | 1.0 | 1.408 | 1.390 |

Table 4-9 lists the adjusted AVOs for some counties based on the obtained weighting coefficients. (See Appendix A.2 for the AVOs adjusted by age group weighting coefficients for all counties.) Although these differences appear small, most

AVO values are adjusted lower. The paired-samples $T$ test indicates that the differences between unadjusted AVOs and census weighted AVOs are significant ($T = 12.439$, $P$ value $< 0.0001$) at the 95% confidence level.

Table 4-9 also lists the driver's age adjustment factor ($F_{age}$) for each county, which was calculated as follows:

$$F_{age} = \frac{AVO_{adj}}{AVO_{unadj}} \qquad (4\text{-}4)$$

Table 4-9: AVOs and Age Adjustment Factors for Some Counties

| County No. | County Name | Unadjusted AVO ($AVO_{unadj}$) | AVO Adjusted by Age ($AVO_{adj}$) | Age Adjustment Factor ($F_{age}$) |
|---|---|---|---|---|
| 3 | Collier | 1.567 | 1.539 | 0.982 |
| 12 | Lee | 1.517 | 1.482 | 0.977 |
| 13 | Manatee | 1.475 | 1.450 | 0.983 |
| 54 | Jefferson | 1.664 | 1.695 | 1.018 |
| 56 | Liberty | 1.648 | 1.563 | 0.948 |
| 86 | Broward | 1.446 | 1.424 | 0.985 |
| 87 | Miami-Dade | 1.408 | 1.390 | 0.987 |
| 89 | Martin | 1.518 | 1.478 | 0.974 |
| 93 | Palm Beach | 1.443 | 1.426 | 0.988 |
| 94 | St. Lucie | 1.555 | 1.520 | 0.977 |

Assuming that the age composition will not change from year to year, this average adjustment factor can be used in lieu of the weighting coefficients described above. Both should lead to the same adjustment if the composition of different subgroups does not change. To obtain the AVO adjusted for driver's age bias, simply multiply the unadjusted AVO by the corresponding $F_{age}$ for the respective county. In most cases, the

$F_{age}$ is a value below 1.0, which reduces the AVO to adjust for the overall over-representation of subgroups with higher AVOs in traffic accidents.

4.4.3. Bias Adjustment Factors for Driver's Gender

Because the AVOs between male drivers and female drivers are significantly different, the improved census weighting mean method is also used to minimize the effect of the bias derived from the over-representation of male drivers or female drivers in the accident vehicle data. During this weighting procedure, the census population data for each county in Florida are grouped and calculated for both male and female groups. Table 4-10 lists weighting coefficients, the unadjusted AVO, and the weighted AVO for each gender group for Miami-Dade County. This table also lists the final overall adjusted AVO, which is 1.403. (See Appendix A.3 for driver's gender weighting coefficients for all counties.)

Table 4-10: Weighting Coefficients for Gender Groups for Miami-Dade County

| Gender Group | Weighting Coefficient ($W_i$) (a) | Unadjusted AVO$_i$ (b) | Weighted AVO$_i$ (a) x (b) |
|---|---|---|---|
| Male | 0.6171 | 1.386 | 0.8553 |
| Female | 0.3829 | 1.430 | 0.5475 |
| Overall | 1.0 | 1.408 | 1.403 |

Table 4-11 lists the adjusted AVOs for some counties based on the obtained weighting coefficients. (See Appendix A.4 for the AVOs adjusted by gender weighting coefficients for all counties.) Overall, the adjustment scale is very small, perhaps because the accident vehicle population already reflects the travel characteristic differences for male and female drivers. The paired-samples $T$ test result ($T = 8.751$, $P$ value $< 0.0001$)

indicates that the differences between unadjusted AVOs and census weighted AVOs in some counties are significant at the 95% confidence level.

Table 4-11: AVOs and Gender Adjustment Factors for Some Counties

| County No. | County Name | Unadjusted AVO ($AVO_{unadj}$) | AVO Adjusted by Gender ($AVO_{adj}$) | Gender Adjustment Factor ($F_{gender}$) |
|---|---|---|---|---|
| 3 | Collier | 1.567 | 1.561 | 0.996 |
| 12 | Lee | 1.517 | 1.512 | 0.997 |
| 13 | Manatee | 1.475 | 1.468 | 0.995 |
| 54 | Jefferson | 1.664 | 1.664 | 1.000 |
| 56 | Liberty | 1.648 | 1.588 | 0.994 |
| 86 | Broward | 1.446 | 1.439 | 0.995 |
| 87 | Miami-Dade | 1.408 | 1.403 | 0.997 |
| 89 | Martin | 1.518 | 1.505 | 0.992 |
| 93 | Palm Beach | 1.443 | 1.436 | 0.995 |
| 94 | St. Lucie | 1.555 | 1.546 | 0.994 |

Table 4-11 also lists the driver's gender adjustment factor ($F_{gender}$) for these counties, which was calculated as follows:

$$F_{gender} = \frac{AVO_{adj}}{AVO_{unadj}} \tag{4-5}$$

Assuming that the gender composition will not change from year to year, this adjustment factor can be used to obtain the AVO adjusted for driver's gender bias. In most cases, the $F_{gender}$ is a value below 1.0, which reduces the AVO to adjust for the overall over-representation of the female group with higher AVOs in traffic accidents.

4.4.4. Composite Adjustment for Driver's Age and Gender

The age adjustment factor $F_{age}$ and the gender adjustment factor $F_{gender}$ obtained in Subsections 4.4.2 and 4.4.3 can be used to adjust the overall AVO values separately. To avoid both potential driver's age and gender biases and reflect the travel characteristic differences in terms of both driver's age and gender, the following formula is used to estimate composite adjusted AVOs:

$$AVO_{adj} = AVO_{unadj} \times F_{age} \times F_{gender} \qquad (4\text{-}6)$$

Based on the above formula, Table 4-12 lists the final composite adjusted AVO values for some counties. Assuming that the gender and age compositions will not change from year to year, both adjustment factors can be used in this formula to obtain the composite adjusted AVO values simply and quickly.

Table 4-12: AVOs and Composite Adjusted AVOs for Some Counties

| County Name | Unadjusted AVO ($AVO_{unadj}$) | Age Adjustment Factor ($F_{age}$) | Gender Adjustment Factor ($F_{gender}$) | Composite Adjusted AVO ($AVO_{adj}$) |
|---|---|---|---|---|
| Collier | 1.567 | 0.982 | 0.996 | 1.533 |
| Lee | 1.517 | 0.977 | 0.997 | 1.477 |
| Manatee | 1.475 | 0.983 | 0.995 | 1.443 |
| Jefferson | 1.664 | 1.018 | 1.000 | 1.674 |
| Liberty | 1.648 | 0.948 | 0.994 | 1.505 |
| Broward | 1.446 | 0.985 | 0.995 | 1.417 |
| Miami-Dade | 1.408 | 0.987 | 0.997 | 1.385 |
| Martin | 1.518 | 0.974 | 0.992 | 1.466 |
| Palm Beach | 1.443 | 0.988 | 0.995 | 1.419 |
| St. Lucie | 1.555 | 0.977 | 0.994 | 1.510 |

CHAPTER 5

AVO INFORMATION SYSTEM

5.1. Introduction

This chapter describes a user-friendly system that makes use of Florida accident records to derive AVO estimates. The system, called the Florida Accident Vehicle Occupancy Rate Information Estimator, or FAVORITE, can estimate AVOs for selected roadway segments, corridors, or counties for specific time periods. Because the system makes use of a comprehensive statewide database, it can potentially be a highly cost-effective means for monitoring statewide, county, and possibly site-specific vehicle occupancy trends. The current system not only provides rich data analysis functions, but also incorporates bias adjustment factors and sampling issues.

5.2. System Implementation

The information system was developed using a combination of Visual Basic, Microsoft Access, and the ESRI MapObjects Developer Library. The current database includes the 1990-2005 accident data for the entire Florida state roadway system. The accident data include the number of passengers of each accident vehicle. In addition, the database also includes a number of variables that can be used for analysis, including district, county, hour of day, day of week, month of year, vehicle type, facility type, area type, and accident severity. These variables allow accident records to be filtered to include only certain types of records for specific analysis needs.

5.2.1. System Functionalities

The major functionalities of the system include:

- Applying filters for time of day, day of week, month of year, type of vehicle, type of road, type of area, and accident severity. These filters allow AVOs to be analyzed for their sensitivity with respect to location, roadway, time, vehicle, etc.

- Defining a new variable by grouping the available categories of a variable.

- Calculating and displaying AVOs in a cross table of two variables, for example, the AVO for different days of a week for different FDOT districts. The total AVOs are also provided for all rows and columns of the cross table.

- Calculating and displaying AVOs on line, bar, or pie charts.

- Calculating and displaying AVOs on a GIS map at the district, county, and corridor levels.

- Exporting tables and charts to Excel.

5.2.2. Incorporation of Bias Adjustment Factors

Because of the drivers' age bias and drivers' gender bias found in the accident data, any estimates of AVO derived from accident data should be adjusted to minimize the effect of both biases on AVO calculations. To adjust for these biases, a census weighting procedure as discussed in Chapter 4 was implemented in FAVORITE. The user may choose whether or not to apply bias adjustment factors. As noted in Section 4.3, it is not necessary that the adjustment factor be applied for AVO monitoring for trends in

one area, as it is the relative values of AVOs, rather than the absolute AVOs, that are important in this case.

## 5.2.3. Incorporation of Minimum Sample Size

A second major design issue in the development of the AVO information system was to consider the minimum number of accident records needed to provide an AVO estimate that will meet the desired accuracy. Unlike other vehicle occupancy data collection methods, the sampling plan for the accident method is not used to determine the number of sample points to be collected. Instead, it is used to determine if the existing accident database contains a sufficient number of accident records to make the required estimation within a specified level of precision. The sampling unit is the vehicle(s) involved in an accident. Based on the defined population of interest in terms of geographic scope and temporal coverage, the number of vehicles required to estimate AVO within a specified precision level at a certain confidence level can be computed for each stratum by using the following formula (Ferlis, 1981):

$$N_h = \left( \frac{Z_{1-\alpha/2} \times \sigma_h}{T_h} \right)^2 \tag{5-1}$$

where

$N_h$ = the number of vehicles required for stratum $h$,

$Z_{1-\alpha/2}$ = the upper $1-\alpha/2$ percentile of the standard normal distribution,

$\sigma_h$ = the composite standard deviation for stratum $h$, and

$T_h$ = the desired tolerance for stratum $h$.

The standard deviation estimates of vehicle occupancy for different strata of accidents can be obtained from existing studies. To use the data, the population of

interest in terms of geographic coverage and temporal conditions has to be clearly defined. Within each selected subpopulation, the variability in occupancy rates will dictate the number of vehicles needed, for which estimates can be obtained to achieve a specified precision level. For example, the required number of vehicles needed to derive an AVO estimate within 0.1, with 95% confidence, is:

$$N_h = \left( \frac{1.96 \times \sigma_h}{0.1} \right)^2 = 384.16 \times \sigma_h^2$$

Table 5-1 provides estimates of the standard deviation, $\sigma_h$, of passenger vehicle occupancy for several strata, as well as the number of reported vehicles that will be needed to estimate AVO. The results calculated are based on three years' (1999, 2000, and 2001) total number of accidents occurring in Miami-Dade county, are restricted to drivers 16 years and older, and are restricted to passenger vehicles with fewer than seven occupants.

Table 5-1: Required Number of Accident Vehicles for Miami-Dade County

|  | Actual # of Vehicles in Database | Standard Deviation ($\sigma_h$) | Minimum Required # of Vehicles |
|---|---|---|---|
| All accidents | 154,152 | 0.810 | 253 |
| Weekday | 115,549 | 0.733 | 207 |
| Weekend | 38,603 | 0.982 | 371 |
| Weekday a.m. rush | 14,252 | 0.584 | 132 |
| Weekday midday | 45,980 | 0.689 | 182 |
| Weekday p.m. rush | 19,791 | 0.747 | 215 |
| Weekend a.m. rush | 1,757 | 0.722 | 201 |
| Weekend midday | 14,823 | 0.941 | 341 |
| Weekend p.m. rush | 4,667 | 1.047 | 422 |

Notes:
   Midday: 9:00 a.m.–4:00 p.m.; a.m. rush: 7:00 a.m.–9:00 a.m.; and p.m. rush: 4: 00 p.m.–6:00 p.m.
   Required number of vehicles involved in accidents reflects ±0.1 precision level.

5.3. Input Specifications

Figure 5-1 shows the main input interface of the system. The screen allows the user to make the following three major selections:

1) Accident data years: The user can select any number of years of accident data to include in his/her analysis.

2) Analysis locations: Location selection can be made at three different levels: district, county, and corridor.

3) Variable filters: The system includes filters for accident severity, hour of day, month of year, day of week, vehicle type, area type, and roadway type. This allows the user to include only a subset of accident records in the AVO estimates. For example, the user may be interested in only the occupancy from passenger cars. In this case, only "Automobile" should be selected. The user can further constrain the accident record set to, for example, weekday traffic, by not selecting the "Sat" and "Sun" categories.

The program also allows the user to group the different categories of a variable. A re-categorized variable can be saved as a new variable that can then be used just like any other existing variable. This feature is particularly useful because the existing categories of a variable may not suit a particular analysis need.

5.3.1. Accident Data Years

The FAVORITE system includes the complete 1990–2005 accident records for Florida's state roadway system. The user can select any number of years of data within

this timeframe to include in the analysis using the "From" and "To" dropdown lists on the main screen.
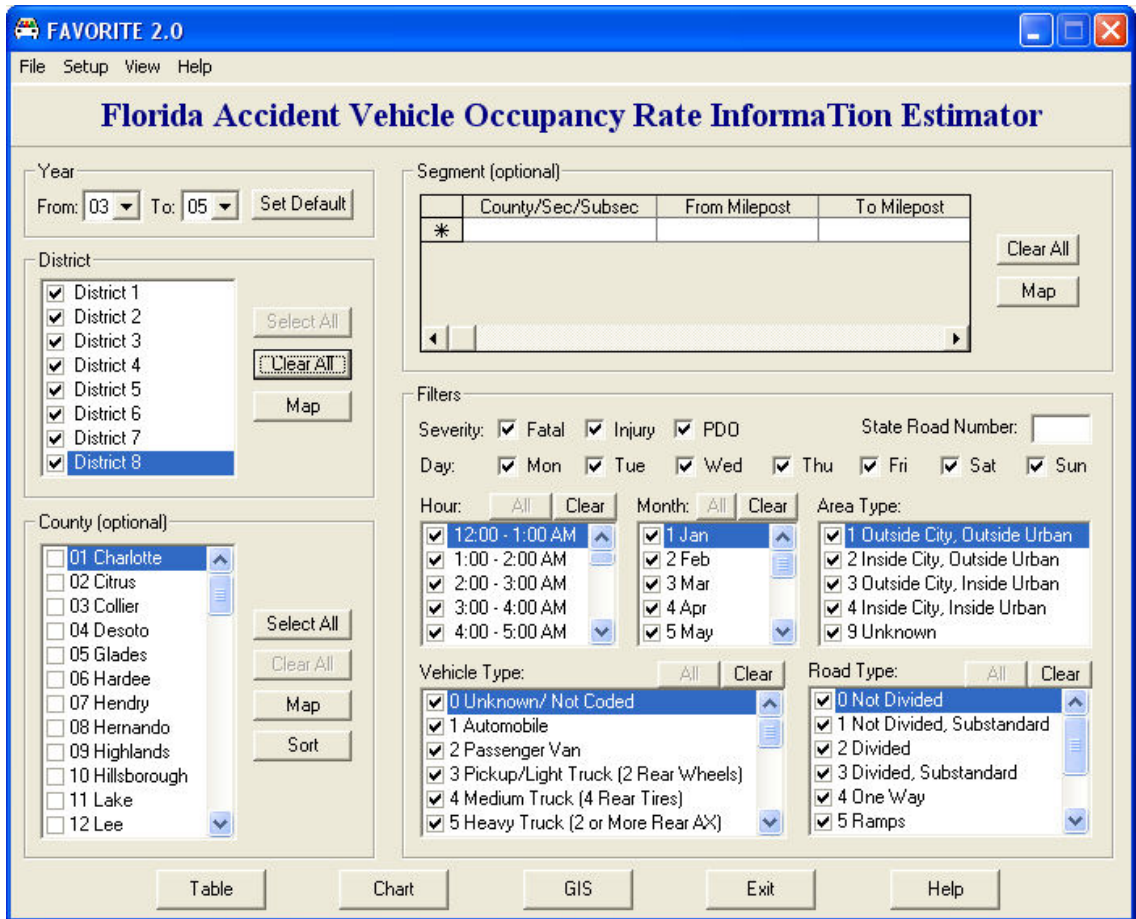


Figure 5-1: FAVORITE Main Screen

5.3.2. Location Selection

Location selection can be made at three different levels: district, county, and corridor. More than one item in each level can be selected. For example, one can select Districts 4 and 6 to cover the Southeast Florida counties. When a district is selected, all of the counties in it will be listed under the "County" list box. The selection of counties is optional. If no counties are selected, the selection is considered to have been made at

the district level only, such that all counties in a selected district will be included. When a county is selected, only data for that county are included.

The selection of districts or counties is made by checking the appropriate checkboxes. Alternatively, one can click the "Map" button beside the "District" or "County" list box to select by map. Selection by map is convenient when the user wants to select a contiguous area such as selecting two adjacent counties. Figure 5-2 shows a map screen in which Miami-Dade and Broward counties are selected.
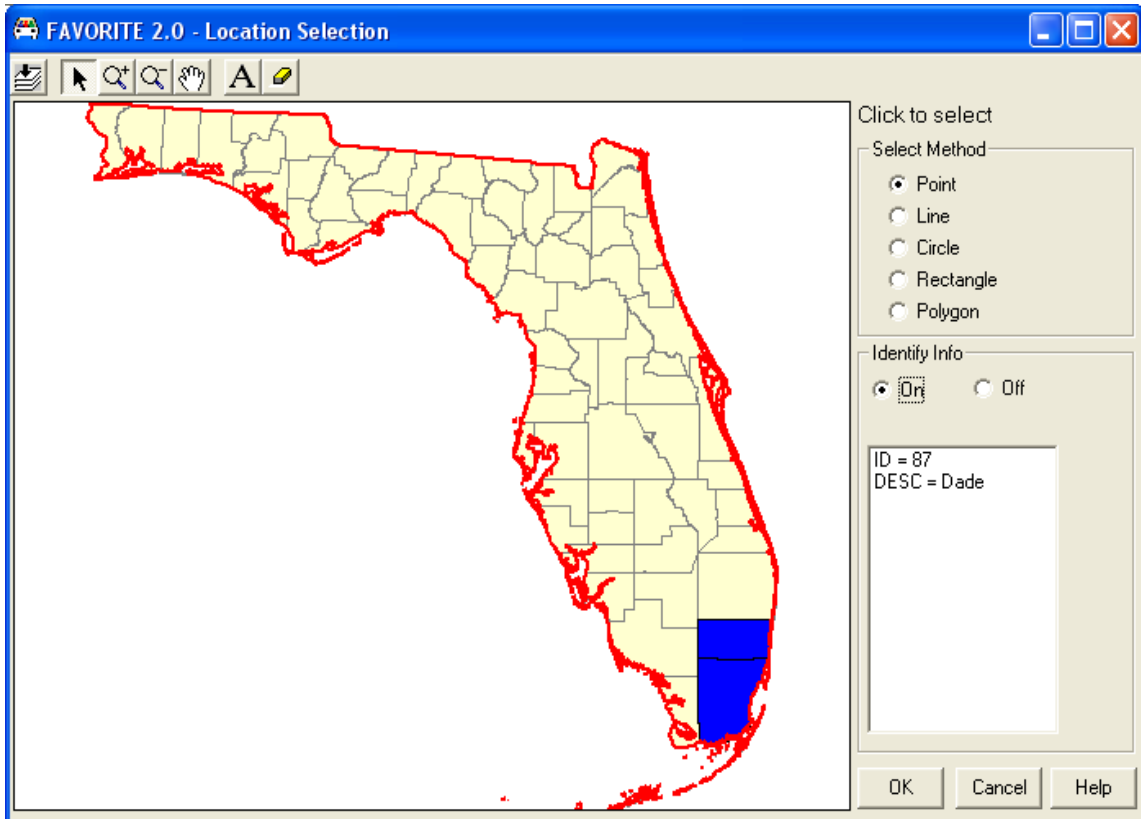


Figure 5-2: Select Locations by Map

To select a specific corridor, the user can either enter the standard roadway ID, which is county/section/subsection, and the beginning and ending mileposts, or the user

can click the adjacent "Map" button to select a particular roadway by pointing and clicking on the map.

5.4. System Output

Once the input specifications described in the preceding section are completed, the user can click on the "Table," "Chart," or "GIS" button at the bottom of the input screen to start computing, analyzing, and displaying AVOs.

5.4.1. Table and Chart Display

The "Table" and "Chart" output options are similar, and they share the same input interface.  Figure 5-3 shows a table that is cross-classified by district and day of week.  As Figure 5-4 shows, the "Chart" option simply displays the same AVOs, but on a chart.  For both tables and charts, the user is given the following display options:

- Select any one or two variables from the two dropdown menus at the top of the screen to classify AVOs.

- Display occupancy rates, total number of vehicles, or total number of occupants by clicking one of the three corresponding radio buttons at the top of the screen.

- Select whether or not to show the overall AVOs across each row and each column. For example, in Figure 5-3, the overall AVO for District 1 is 1.49 (shown under the "Total" column) and the overall AVO for Monday is 1.37.  The table also gives an overall AVO of 1.44, which is shown in the bottom-right cell**.**

- The user can choose whether or not to display AVOs that do not meet the minimum sample sizes, which are computed by the system based on the sample

size formula (see Section 5.2.3) for each user-selected cross-class. Thus, if a user chooses to require the minimum sample size, the corresponding cell will be displayed only if the number of vehicles (note: not the number of accidents) involved in the accidents used to calculate the AVO meets a minimal sample size.

- Select Adjust by census weighting factors to avoid age and gender biases found in the accident vehicle data.
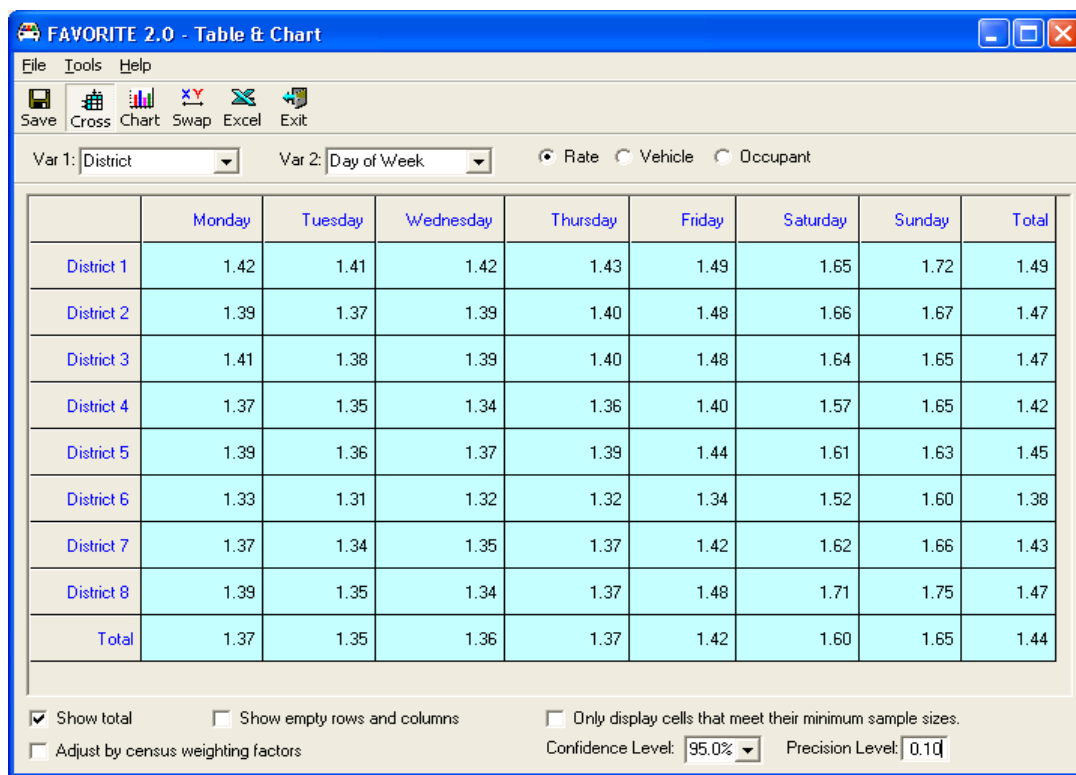


Figure 5-3:  AVOs Cross-Classified Table by FDOT District and Day of Week

The following functions are available from the table and chart screens:

- The user is allowed to select up to two variables by which to cross classify the AVO estimates.  By default, AVO estimates are displayed when the table is first displayed.

76

- The corresponding cross tables for the number of vehicles and number of occupants, respectively, can be displayed by clicking the appropriate radio buttons.

- The user can export the current view to Excel.

- The user can swap the rows and columns of the cross table.

- Different options may be checked or unchecked to customize and trim the output result, such as showing total AVO values, enabling the bias adjustment feature, and considering minimum sample size.
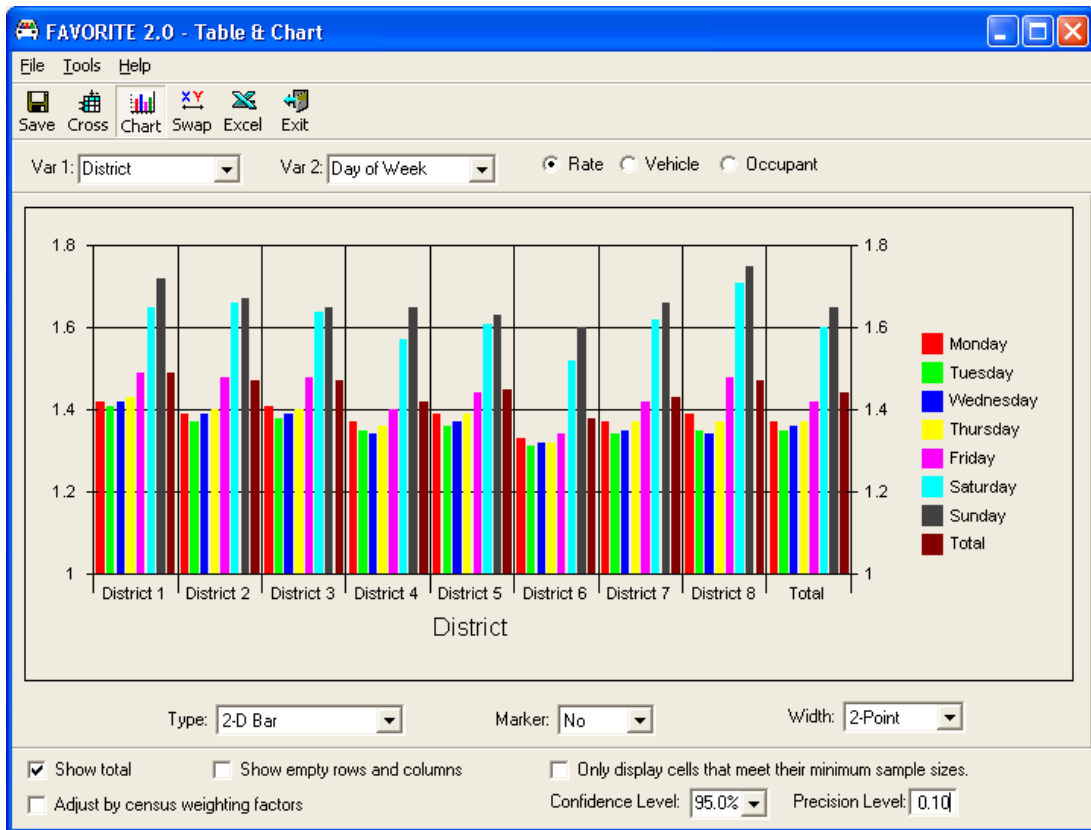


Figure 5-4: AVOs Cross-Classified Chart by FDOT District and Day of Week

5.4.2. GIS Display

   While tables and charts are able to display AVOs that are cross classified by variables, they cannot show the AVOs spatially. FAVORITE provides a geographic information system (GIS) interface that can display AVOs by district, county, and segment. Figure 5-5 shows the main interface of the GIS display. The left side of the screen allows the user to make various selections, while the right side of the screen displays a map view of the state. The top-left corner lists the GIS layers and the corresponding colors used for display. Below this layer box is a "Variable" dropdown box that allows the user to specify the variable options to include in the calculation and analysis of AVOs. Once a variable is selected, all options will be displayed in the box below it. The user can select the options to include by checking the appropriate checkboxes.

   Once a variable option is selected, the six specification boxes below the option list box will become active. The first of these boxes allows the user to select a theme to display by county, district, or segment. Figure 5-5 shows an example of AVOs displayed by county. Figure 5-6 shows another example of AVOs displayed by segments.

   The "Class" dropdown list box allows the user to select the number of classes for the theme. The default number of classes is seven. The first class is always assigned to 0.00 to 0.99, which normally includes features that have no accidents. The last class includes any number above a certain threshold. All classes between these two boundary classes are divided based on the increment specified to the right. By default, the increment is 0.1.

The "Style" option allows the user to specify whether to show themes by color, line width, or a combination of both. The default option displays by different colors. Obviously, the line width option applies only to the segment theme. The "Color" option allows the user to select a color scheme. By default, random colors are used. The user may choose to use the "Blue" and "Red" color schemes, which display features from gradual light to dark colors to indicate low to high AVOs. The GIS display is automatically refreshed as soon as any one of these specifications is changed.
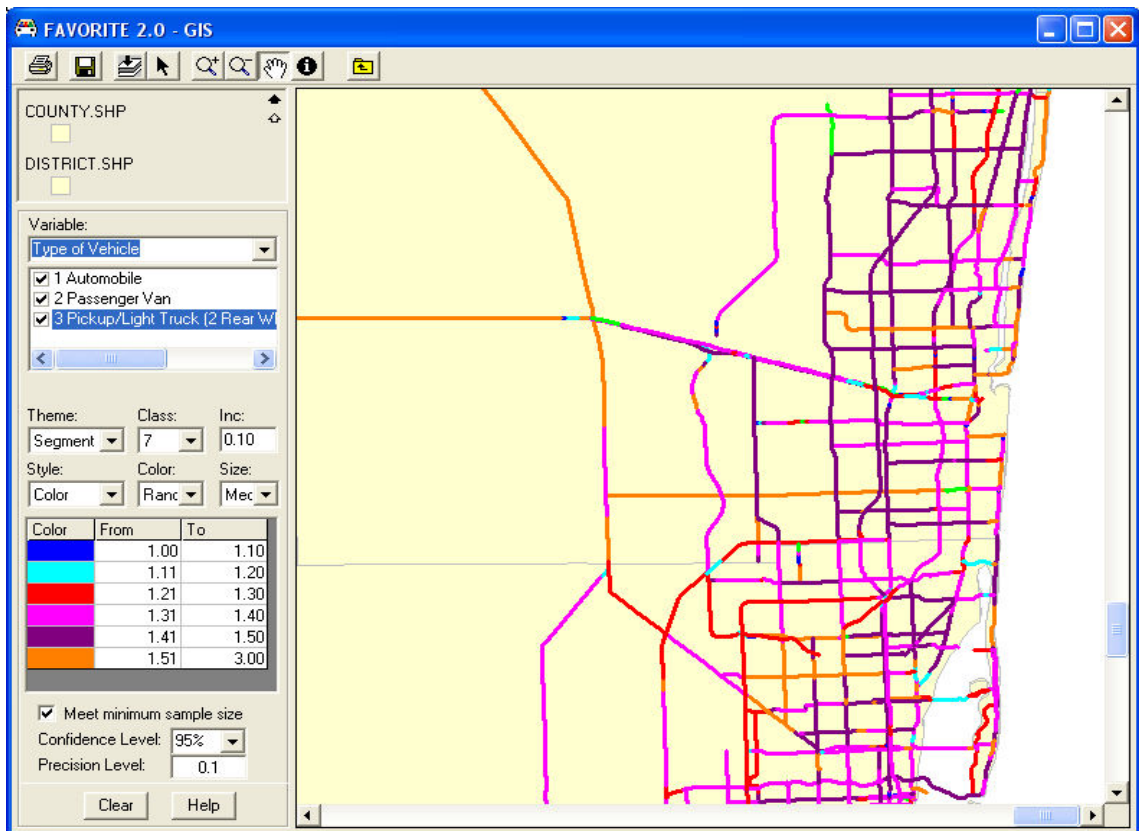


Figure 5-5: Automobile AVOs Displayed by County
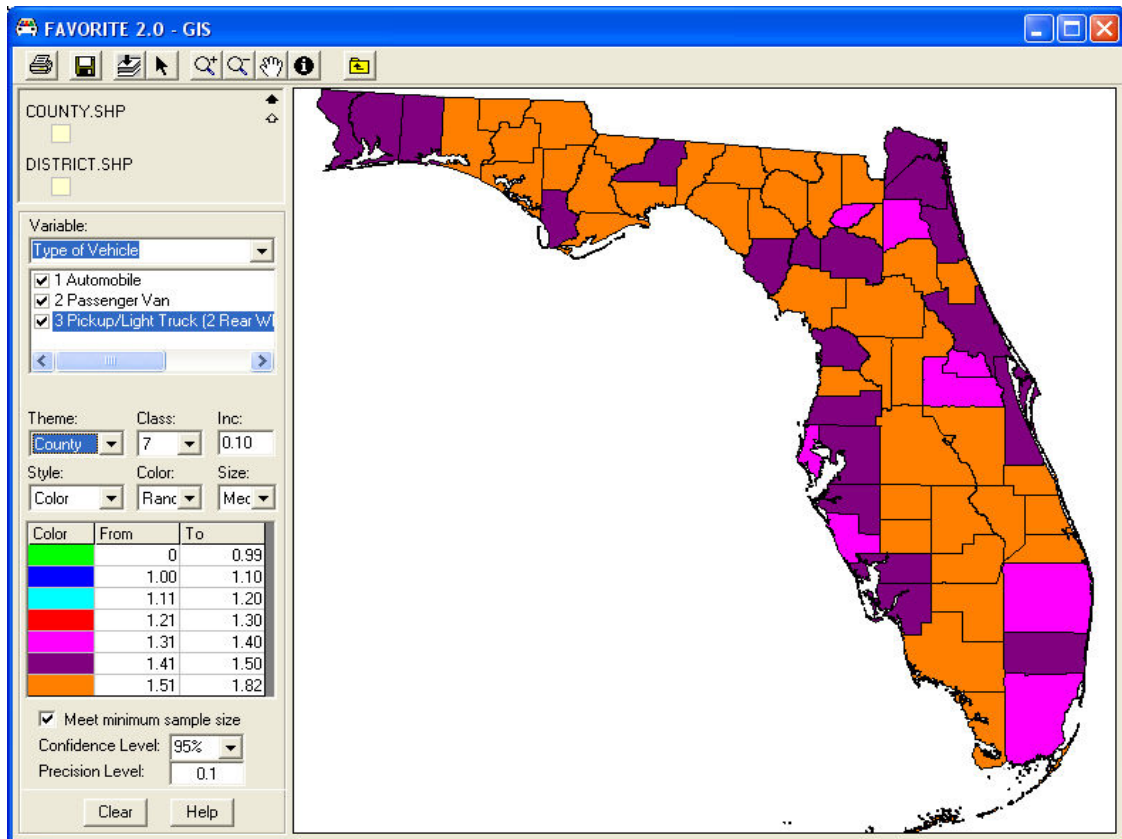
Figure 5-6: Automobile AVOs Displayed by Segments

A number of tool buttons are available on the GIS screen. These buttons can be used to operate map functions, such as zoom in, zoom out, print map, and identify segment information. Figure 5-7 shows the GIS map print layout screen which allows users to enter a title or a footnote, copy the print layout to the clipboard, or print the map to the default printer.

## 5.5. Variable Re-Categorization

FAVORITE allows the user to re-categorize the options of each variable. This function is very useful for AVO category analysis. Users can add re-categorized variables based on the original raw variables. Figure 5-8 shows a newly re-categorized

variable, "Season," that is based on the original accident month. If the user is not familiar with the code numbers, he or she should click the cell for which codes are to be entered, and then press the F2 function key. This will invoke a new window that lists all of the codes and option names for the selected variable.
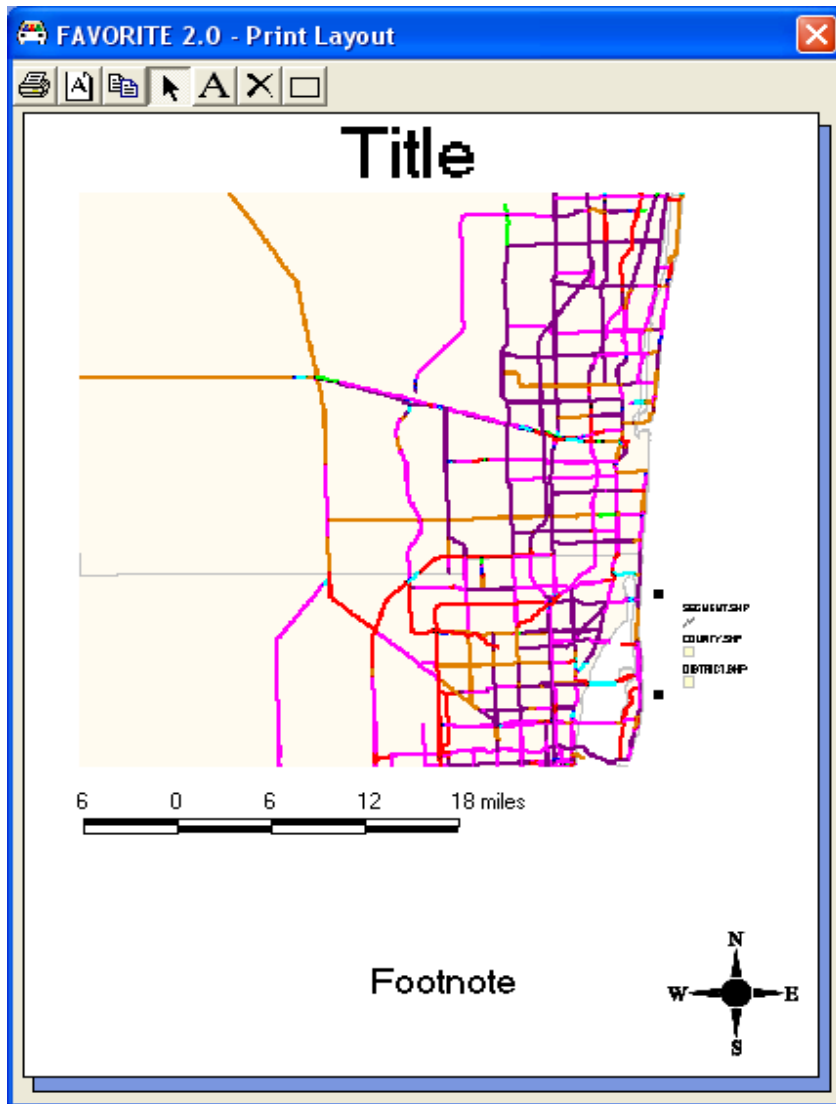


Figure 5-7: GIS Map Print Layout

Figure 5-8: Screen for Adding a Category

CHAPTER 6

EVALUATION OF AVO ESTIMATES FROM ACCIDENT DATA

6.1. Introduction

In this chapter, the accuracy of AVO estimates from accident data are evaluated based on two approaches. The first is used to determine if the AVO trends from the accident data are as expected. The second is used to compare the AVOs from accident data with those collected from the field at the same locations. A description of the field data source is already given in Chapter 4.

6.2. Reasonableness Checks

Reasonableness checks involve examining the AVO estimates to determine if they match the expected trends.

6.2.1. AVO Trends by Year and Day of Week

Figure 6-1 shows that, as expected, weekend AVOs for passenger vehicles are higher than weekday AVOs, with Sunday having the highest AVOs, followed by Saturdays and Fridays. Furthermore, Fridays tend to have slightly higher AVOs than the other weekdays, which tend to have similar AVOs. These observations were consistent over each of the four years of data shown.
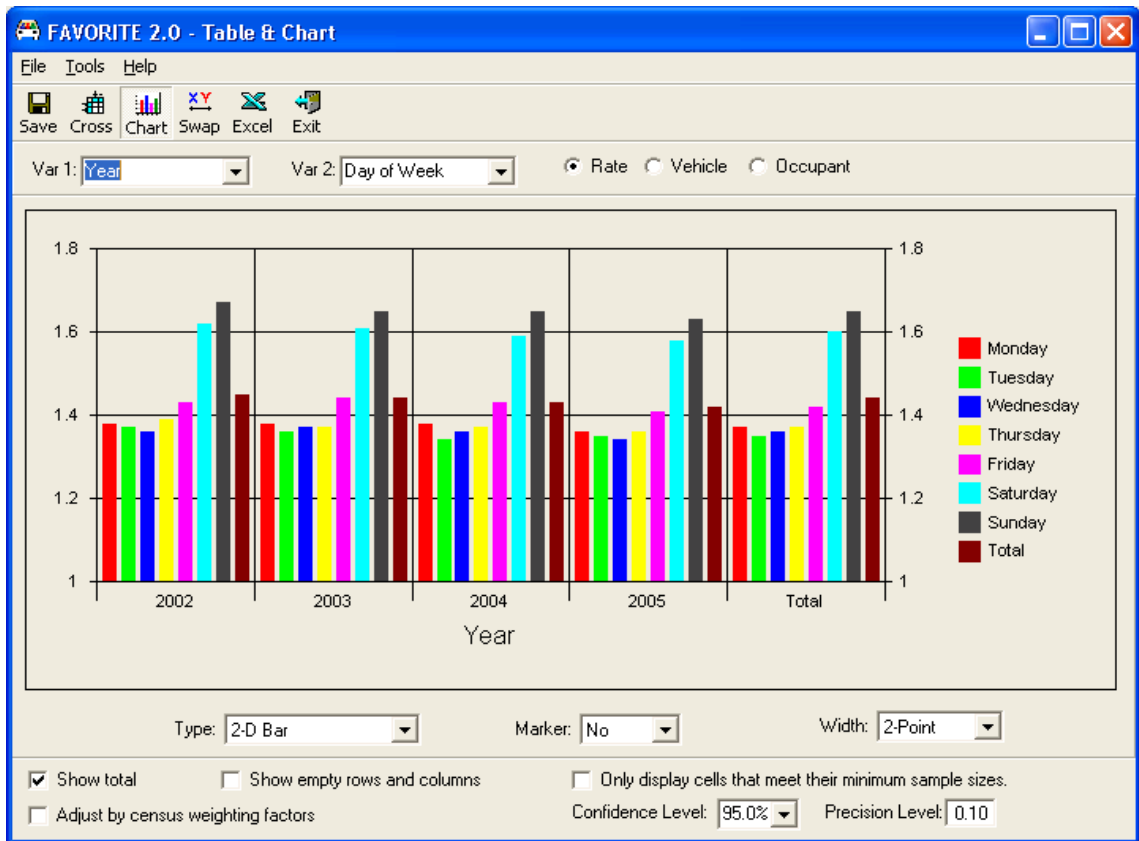
Figure 6-1:  AVO Trend by Year and Day of Week

## 6.2.2. AVO Trends by Month

Figure 6-2 shows that, as expected, the month of July tends to have the highest AVO, followed by the month of March, which coincides with Spring Break traffic in Florida, and the month of December, which coincides with Christmas holiday traffic.
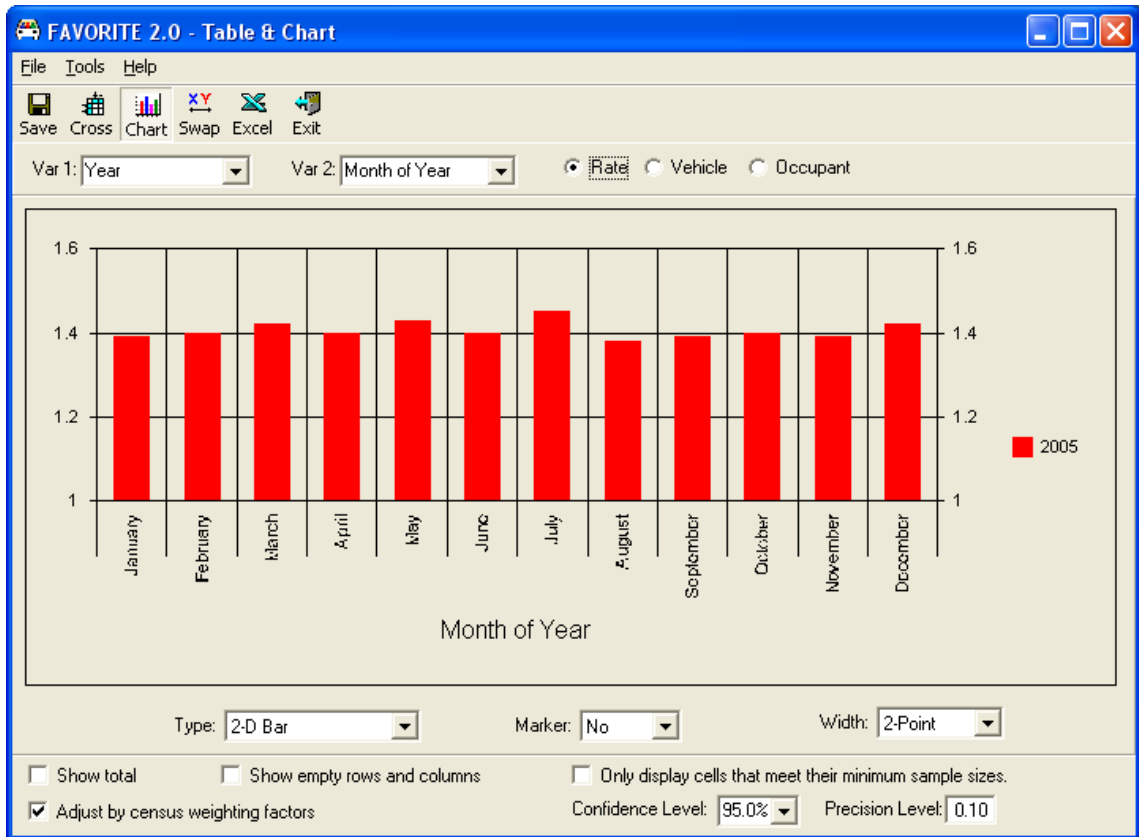
Figure 6-2:  AVO Trend by Month

6.2.3. AVO Trends by Hours

Figure 6-3 shows that, as expected, AVOs during the morning peak hours are lower than those of the afternoon peak hours and that daytime AVOs are lower than night-time AVOs.

6.2.4. AVO Trends by Area Type

Figure 6-4 shows a GIS thematic map of AVO distribution by county.  The darker the color, the higher the AVO.  The map shows that, as expected, rural counties generally have higher AVOs than do urbanized counties.
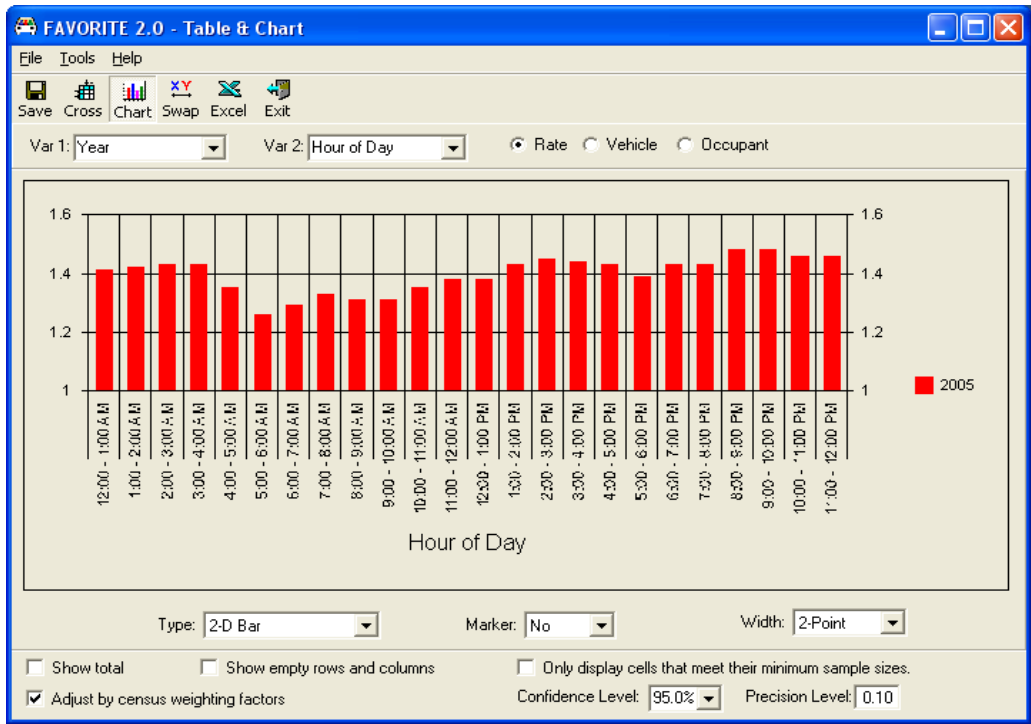
Figure 6-3: AVO Trend by Time of Day
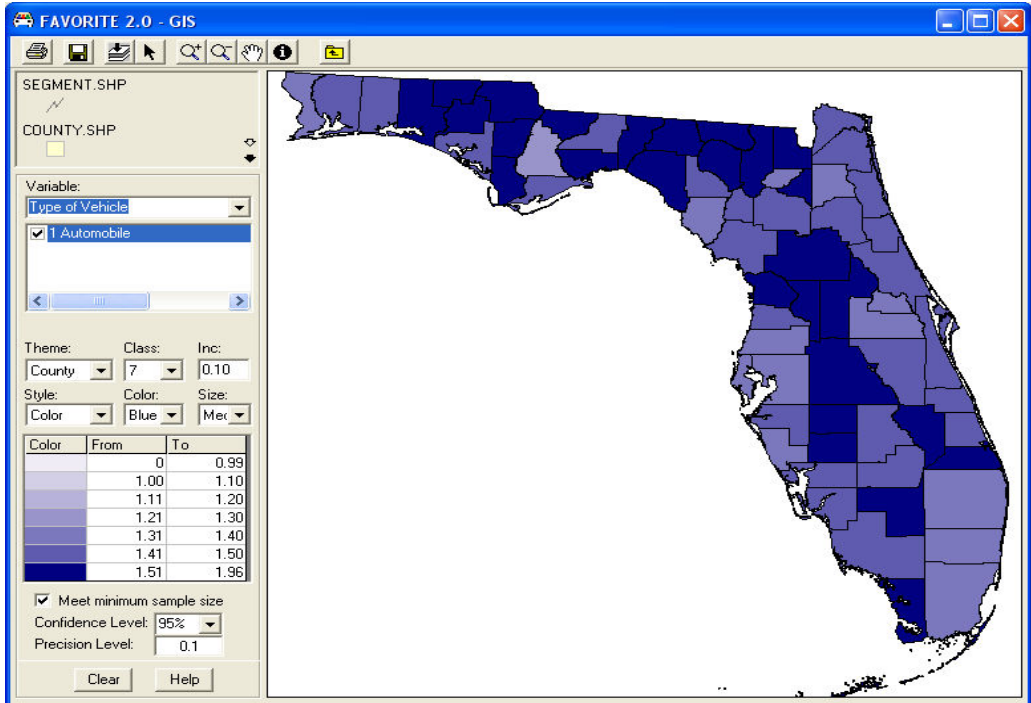


Figure 6-4: AVO Trend by County

6.2.5. AVO Trends by Vehicle Type

Figure 6-5 shows the AVOs for different types of vehicles. The AVOs obtained appear reasonable compared to the expected AVOs for different vehicle types. Some observations include:

- AVOs of passenger vans are higher than those of passenger vehicles.

- Buses have a significantly higher AVO than any other vehicle types.

- Bicycles have the lowest AVOs.

    These observations were consistent throughout the four years of data used.
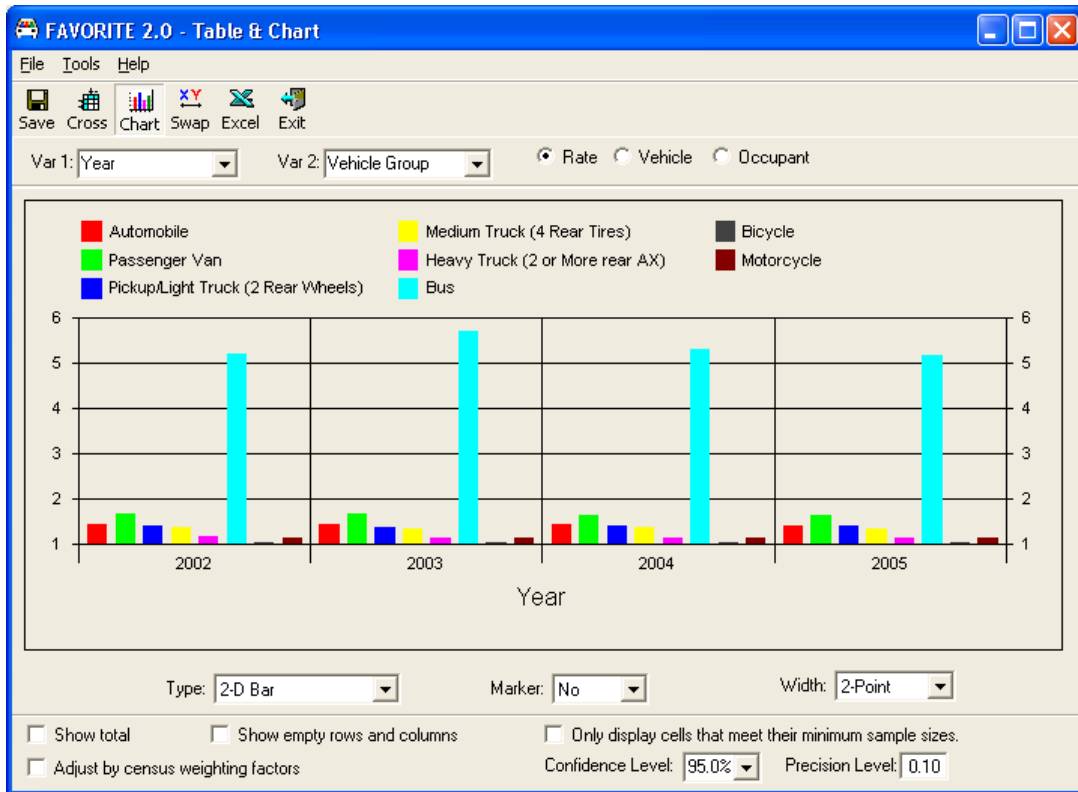


Figure 6-5: AVO Trend by Vehicle Type

6.3. Comparisons with Field Data at the County Level

To compare the AVOs obtained by the two data sources, the Miami-Dade County year 2005 accident data were tailored to match the field data collection time (7:00 a.m. to 6:00 p.m.) and days of the week (Tuesday through Thursday) of a 2006 field collection study. Given that vehicle occupancy rates vary by time of day, critical time periods during the aforementioned 7:00 a.m. to 6:00 p.m. time span were selected for separate analysis. The time periods of interest included the a.m. peak period (7:00 to 9:00 a.m.), midday period (11:00 a.m. to 1:00 p.m.), p.m. peak period (4:00 to 6:00 p.m.), and off-peak period (9:00 to 11:00 a.m. and 1:00 to 4:00 p.m.). Also, the locations of interest were classified into three facility types: surface streets, toll facilities, and freeways.

Figure 6-6 shows that the AVOs from accident records are consistently slightly higher for all time periods. Figure 6-7 shows that the estimates are very close for freeway facilities, but are slightly higher for those from accident records for the surface street and toll facility locations. These results are expected because small children are difficult to see using the windshield method. On the other hand, in the carousel method that was used to collect freeway data, because the observer vehicle moves in parallel with the observed vehicles, the observers had more time to observe the occupants, including younger children, hence the closer results for the freeway AVOs.

6.4. Comparisons with Field Data at the Site-Specific Level

To compare the field and accident AVOs gathered at the same survey points, accident data from one mile upstream and one mile downstream of the survey point were included. Figure 6-8 shows that a positive relationship exists between the two data

sources. However, the AVO estimates from the accident data tend to be higher than the field estimates. Again, this is likely due to infants and small children being missed by observers using the windshield method.
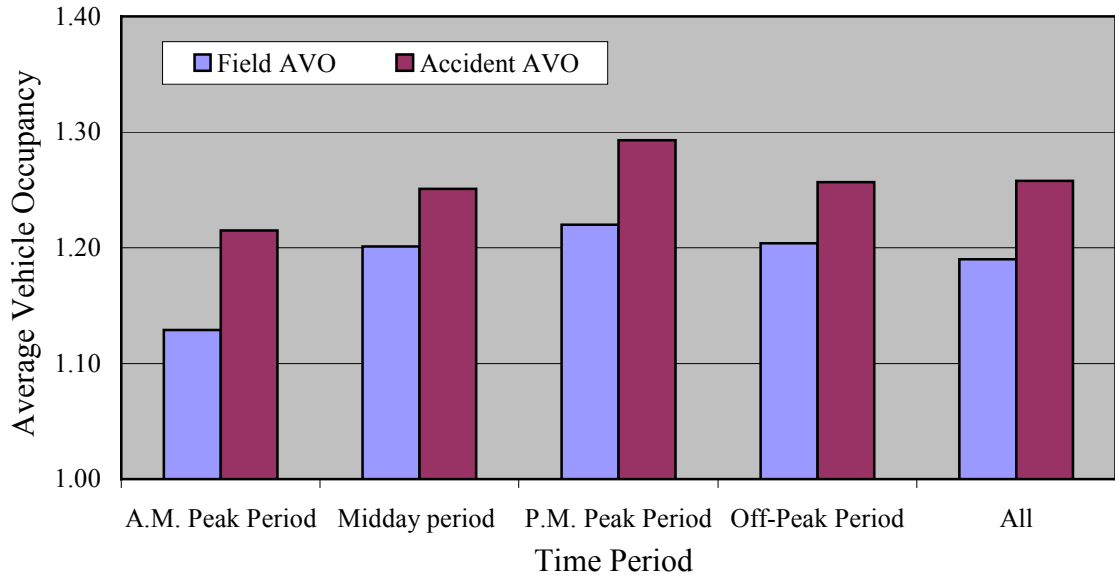


Figure 6-6: Comparisons of Countywide AVOs for Different Time Periods
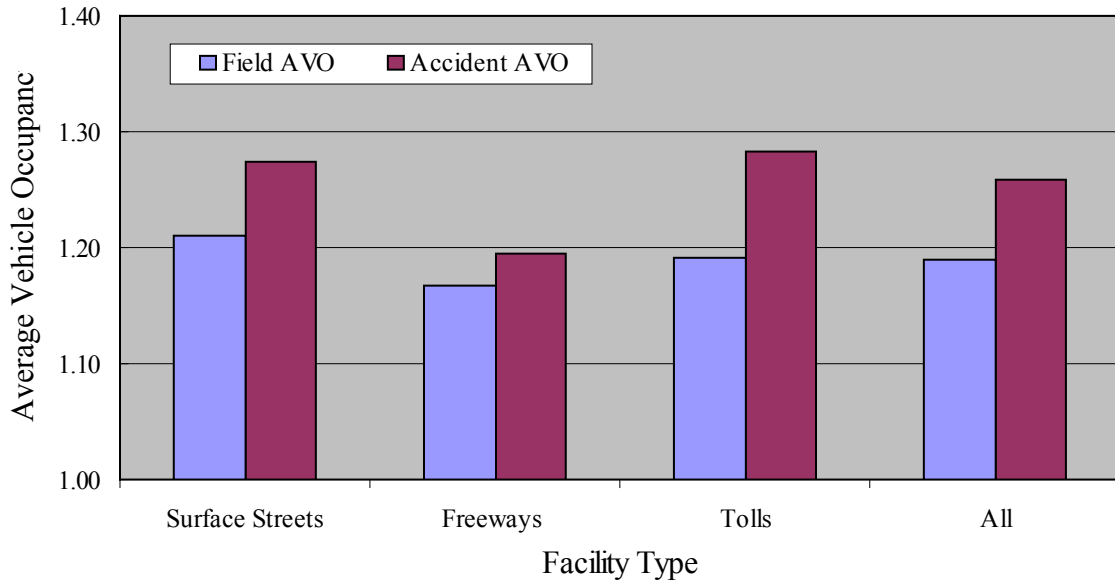


Figure 6-7: Comparisons of Countywide AVOs for Different Facility Types
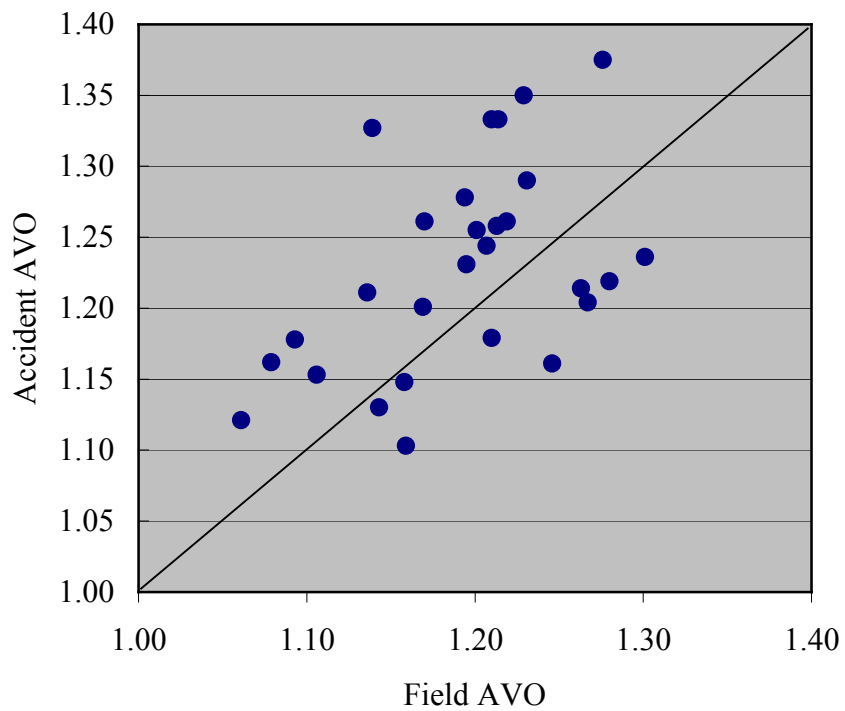
Figure 6-8:  Comparisons of AVOs from the Field and from Accidents

CHAPTER 7

VEHICLE OCCUPANCY PREDICTION MODELS

7.1. Introduction

While historical occupancy information can be obtained from accident records, and the current occupancy information can be collected using the field collection method, there have been no known methods developed to predict future occupancy rates. This chapter describes an effort to explore the feasibility of developing prediction models that can be used to estimate the expected vehicle occupancy rates given the future socioeconomic conditions of the study areas.

7.2. Data Sources

The Census 2000 demographic data from Summary Tape File #1 (STF #1) and Summary Tape File #3 (STF #3) were the main source of data for this study. The demographic data include age, gender, race and ethnic origin, household income level, vehicle ownership, families, and housing units, etc. The data by age are generally limited to persons of working age, defined as 16 years and older.

The 2002 Info USA employment data were also used in this study. The employment data were categorized into industry, service, and commerce types. The 2002 employment data were the closest to the 2000 Census that were available for this study. Both census demographic and employment data were aggregated into the census blockgroup GIS layer for model regression analysis. Appendix B.1 lists the variable attributes, together with their definitions, from the two data sources. Except for

MEDIAN_AGE, AV_HH_SIZE, and AVG_HH_INC, all other variables were converted to density (i.e., divided by the total area) in the analysis. Also three employment variables and some household variables were converted to percentage (i.e., divided by the total number of employees, or divided by the total household units) during modeling procedures.

In addition, the variables listed in Table 7-1 were calculated from two or more of the attributes from the 2000 Census demographic data. These composite variables were also used to investigate the impacts on the AVOs in the analysis.

Table 7-1: Composite Variables

| Variable Name | Description |
|---|---|
| YOUNG_AGE | Density of population with age in the range of 0–14 years |
| YOUNG_AGE_PER | Percentage of population with age in the range of 0–14 years |
| OLD_AGE | Density of population with age equal to or greater than 70 |
| OLD_AGE_PER | Percentage of population with age equal to or greater than 70 |
| TOTAL_VEH | Density of household vehicles |
| HH_GT_1VEH | Density of household units with two or more vehicles |
| HH_GT_1VEH_PER | Percentage of household units with two or more vehicles |
| HH_GT_2VEH | Density of household units with three or more vehicles |
| HH_GT_2VEH_PER | Percentage of household units with three or more vehicles |
| PERPVEH | Average persons per vehicle |

The vehicle occupancy data were only derived from years 1999 and 2000 accident data to maintain compatibility with the census year. These accident data were filtered by vehicle type and weekday type, leaving only passenger vehicles involved in weekday accidents.

7.3. Weekday AVO Models at Area Level

AVO can differ from one location to another. The statistical analysis in Chapter 3 indicated that the AVOs for different counties were significantly different from each other. To model the relationship between an area's AVO and the associated socioeconomic characteristics, the AVO for the area was calculated from vehicles involved in accidents that occurred inside the area. The census blockgroup data were aggregated at the area level. In this study, only two area levels are analyzed: county level and census blockgroup level.

1) *County level*: The AVOs for all 67 counties in the state of Florida were calculated from vehicles involved in accidents, and the census blockgroup data were aggregated into the county level.

2) *Census blockgroup level*: All vehicles involved in accidents that were located inside each census blockgroup were counted. A 0.25-mile buffer size was applied to each census blockgroup to extend the influence of the areas to include vehicles involved in accidents that occurred on the roadway between two or more census blockgroups. Also, the AVO for each blockgroup was calculated to model the relationship between the AVO and socioeconomic variables at the census blockgroup level.

7.3.1. AVO Model at County Level

Tables 7-2a and 7-2b list the SPSS output for this model. The stepwise method in the SPSS statistical analysis package selected two predictors: HH_GT_1VEH (i.e., household density with more than one vehicle) and TT_30_59M (i.e., worker density

with travel time to work in the range of 30–59 minutes).  However, the variance inflation factor (VIF) value for each of these two predictors is 19.125.  This is obviously much larger than four, indicating that there is a high correlation between the two predictors. Hence, TT_30_59M was deselected in the final model refinement procedure because it contributes a very small part to $R^2$ (0.059).

Table 7-2a:  AVO Weekday County Model Summary

| Model | $R$ | $R$ Square | Adjusted $R$ Square | Std. Error of the Estimate | Change Statistics | | | | Durbin-Watson |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | $F$ Change | df1 | df2 | Sig. $F$ Change | |
| 1 | 0.663 | 0.439 | 0.430 | 0.05970 | 53.660 | 1 | 65 | 0.000 | 2.429 |

Table 7-2b:  AVO Weekday County Model Coefficients

| Model | Unstandardized Coefficients | | Standardized Coefficients | $t$ | Sig. | Collinearity Statistics | |
|---|---|---|---|---|---|---|---|
| | $B$ | Std. Error | Beta | | | Tolerance | VIF |
| (Constant) | 1.556 | 0.007 | | 210.341 | 0.000 | | |
| LOG_VEH_GT_1VEH | -0.043 | 0.006 | -0.672 | -7.325 | 0.000 | 1.000 | 1.000 |

The final weekday AVO model at the county level ($R^2$/adjusted $R^2$ = 0.439/0.430) has the following form:

$$AVO = 1.667 - 0.040 \times \ln(HH\_GT\_1VEH) \qquad (7\text{-}1)$$

In this model, the negative sign of the coefficient of predictor HH_GT_1VEH is also meaningful.  It indicates that a higher household density with more than one vehicle will decrease AVO.   Figure 7-1 shows the fit of this model for the given data.

7.3.2. AVO Model at Census Blockgroup Level

For this modeling, Miami-Dade County census blockgroup data and vehicle occupancy data were used. The stepwise method involved the selection of many more

variables: HH_0_VEH, HH_2_VEH, VACANT_HU, MEDIAN_AGE, AV_HH_SIZE, BLACK, AVG_HH_INC, IND_EMP, and COM_EMP. Tables 7-3a and 7-3b list the SPSS output for this model. They indicate that the AVO in each census blockgroup is related to average household size, average household income, industry and commercial employee density, vacant household unit density, and household density without a vehicle and with two vehicles.



$$AVO = 1.667 - 0.040 \times ln\,(HH\_GT\_1VEH)$$
$$R2 = 0.439$$
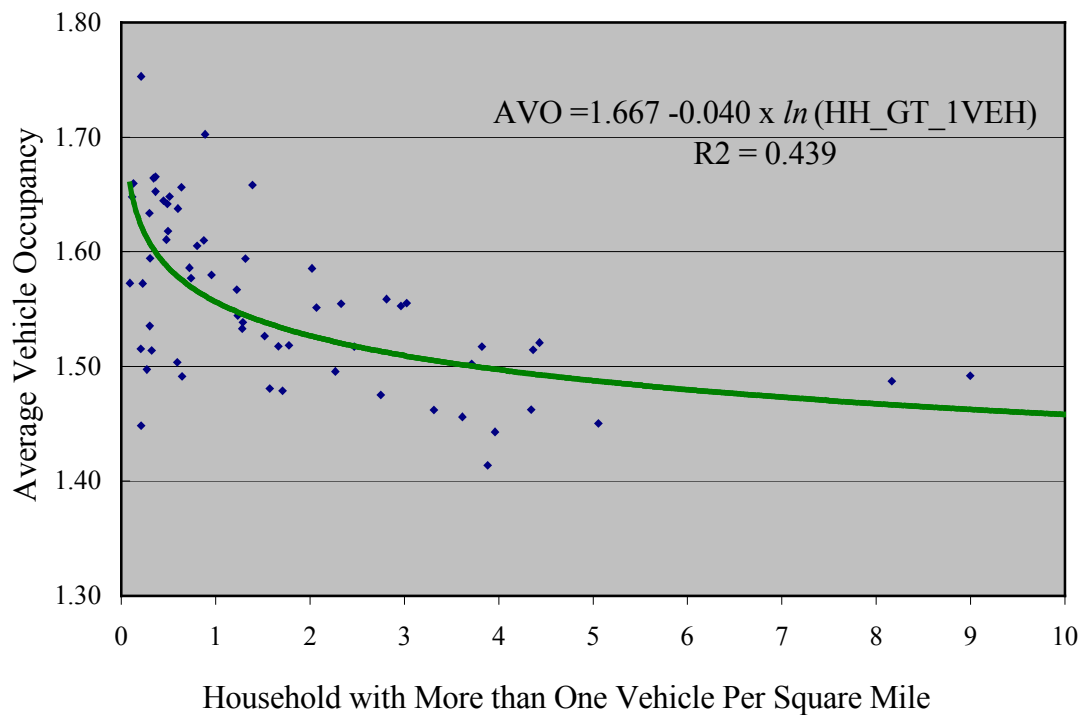
Figure 7-1: Weekday County AVO Model

In this model, the signs of the coefficients are as expected. For instance, the negative sign for AVG_HH_INC indicates a higher average household income with lower average vehicle occupancy because of a higher possibility of multiple vehicle ownership. The positive sign for HH_0_VEH indicates that the higher density of a

95

household with no vehicles will increase the possibility of carpooling, which will result in a higher occupancy rate. However, the $R^2$ in this model is low, at 0.088.

Table 7-3a: Miami-Dade County AVO Weekday Model Summary

| Model | $R$ | $R$ Square | Adjusted $R$ Square | Std. Error of the Estimate | Change Statistics | | | | Durbin-Watson |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | $F$ Change | df1 | df2 | Sig. $F$ Change | |
| 1 | 0.297 | 0.088 | 0.080 | 0.07745 | 11.050 | 8 | 913 | 0.000 | 1.370 |

Table 7-3b: Miami-Dade County AVO Weekday Model Coefficients

| Model | Unstandardized Coefficients | | Standardized Coefficients | $t$ | Sig. | Collinearity Statistics | |
|---|---|---|---|---|---|---|---|
| | $B$ | Std. Error | Beta | | | Tolerance | VIF |
| (Constant) | 1.342 | 0.020 | | 66.284 | 0.000 | | |
| IND_EMP | -0.004 | 0.001 | -0.105 | -3.206 | 0.001 | 0.923 | 1.084 |
| COM_EMP | 0.003 | 0.001 | 0.096 | 2.766 | 0.006 | 0.834 | 1.200 |
| AVG_HH_INC | -2.88E-007 | 0.000 | -0.079 | -2.227 | 0.026 | 0.793 | 1.261 |
| AV_HH_SIZE | 0.018 | 0.004 | 0.147 | 4.035 | 0.000 | 0.751 | 1.331 |
| MEDIAN_AGE | -0.001 | 0.000 | -0.074 | -2.146 | 0.032 | 0.837 | 1.194 |
| VACANT_HU | 0.006 | 0.001 | 0.155 | 4.052 | 0.000 | 0.686 | 1.458 |
| HH_0_VEH | 0.004 | 0.001 | 0.116 | 2.902 | 0.004 | 0.628 | 1.592 |
| HH_2_VEH | -0.012 | 0.003 | -0.169 | -4.728 | 0.000 | 0.782 | 1.278 |

7.4. AVO Model at Corridor Level

Different roadway types serve different people for different trip purposes. The federal roadway functional classification system defines the five urban roadway types as given in Table 7-4. In this study, these roadway types were regrouped into two major groups for analysis: urban freeways, consisting of those with class codes 11 and 12, and urban arterials, consisting of those with class codes 14 and 16. Codes 17 and 19 are not modeled in this study.

Table 7-4:  Urban Roadway Types

| Functional Classification Code | Definition |
|---|---|
| 11 | Principal Arterial - Interstate |
| 12 | Principal Arterial - Other Freeways and Expressways |
| 14 | Principal Arterial - Other |
| 16 | Minor Arterial |
| 17 | Collector |
| 19 | Local |

For urban arterials, all urban counties are also classified into two categories: large urban counties with populations over 500,000 and medium-size urban counties with populations over 100,000 but less than or equal to 500,000.  Table 7-5 lists all large urban counties in the state of Florida.

Table 7-5:  Large Urban Counties

| County Number | County Name | Total Population |
|---|---|---|
| 87 | Miami-Dade | 2,253,362 |
| 86 | Broward | 1,623,018 |
| 93 | Palm Beach | 1,131,184 |
| 10 | Hillsborough | 998,948 |
| 15 | Pinellas | 921,482 |
| 75 | Orange | 896,344 |
| 72 | Duval | 778,879 |

7.4.1. Data Preparation

The data preparation for the AVO model at the corridor level includes four considerations: roadway segmentation strategy, AVO calculation for roadway segments, buffer size definition, and predictor selection rules.

7.4.1.1. Segmentation

In Florida, traffic and geometric information for the state roadway system is stored in the Roadway Characteristics Inventory (RCI) system, which includes more than 200 roadway features (e.g., functional classification, annual average daily traffic , number of lanes, median width, etc.).  Three basic fields (i.e., RoadwayID, BeginMilepost, and EndMilepost), were used to identify segments with special geometric characteristics.  To model the relationship between the AVO in each corridor and the socioeconomic characteristics along the corridors, each state roadway is divided into segments based on the roadway functional class type, defined in the field FUNCLASS.  Each of these segments has a unique functional class, and all have different lengths.

In the preliminary experiments, to achieve the best possible modeling results, each urban arterial was divided into 1-, 2- and 4-mile segments.  Also, each of these segments had a unique roadway functional class.  However, this fixed-length segmentation method did not provide good results.  It was found that models based on this fixed-length segmentation method generally had much smaller $R^2$ values than those models that were based only on roadway function class types.  Hence, only the roadway function class type is used to split up the urban arterials.

7.4.1.2. Aggregation

The key fields in the accident database are Roadway ID and Milepost. Both field values for each accident record are referred to as this accident location spatially. Each roadway segment has three fields to represent unique segment: Roadway ID, Begin Milepost, and End Milepost. Vehicles involved in accidents located inside each segment are aggregated, and the corresponding vehicle occupants and vehicle numbers are counted. Table 7-6 shows a sample of how the AVO for each segment was calculated.

Table 7-6: Segmentation and Aggregation Results

| Roadway ID | Begin Milepost | End Milepost | FUNCLASS | Occupants | Vehicles | AVO |
|---|---|---|---|---|---|---|
| 88010000 | 0 | 10.13 | 14 | 879 | 633 | 1.39 |
| 88010000 | 17.131 | 17.681 | 14 | 22 | 14 | 1.57 |
| 88010000 | 18.581 | 20.981 | 14 | 111 | 84 | 1.32 |
| 88060000 | 21.82 | 32.142 | 14 | 825 | 583 | 1.42 |
| 88060000 | 32.142 | 33.592 | 16 | 70 | 50 | 1.40 |
| 88060001 | 0 | 1.364 | 14 | 104 | 79 | 1.32 |

7.4.1.3. Buffer Sizes

The GIS buffer analysis technique has been widely used in transportation modeling procedures to estimate traffic data, such as annual average daily traffic (Zhao and Chung, 1999). This technique, combined with the regression method, is also used in the analysis of weekday AVO models for the corridor level. During the compilation of the socioeconomic data around each roadway segment, the GIS buffer function was used to create buffers around roadway segments to compile data for regression analysis. Buffer sizes of 0.5, 0.75, 1, 1.25, 1.50, and 2 mile(s) were used.

7.4.1.4. Predictor Selection

In the preliminary procedures, all independent variables and the dependent variable AVO for each buffer data set were loaded into the SPSS system. By using a linear regression stepwise method, all significant predictors for each buffer size data set were selected based on the following rules:

1)  Each predictor must be significant at the 0.05 confidence level.

2)  Variance inflation factors (VIF) must be less than four to avoid a high correlation between independent variables.

7.4.2. AVO Model for Arterials in Large Urban Counties

Large urban counties, in which there were a total of 277 urban arterial segments, possess different socioeconomic characteristics than medium-size urban counties. Along these arterial segments, different buffers with different buffer sizes were created for model selection. Table 7-7 lists the initial results of the preliminary procedures. [Note: All independent variables are explained by density (per acre) except average variables, such as AVG_HH_INC.] In this table, a buffer size of 1.50 miles gives the highest $R^2$ value. For all of these initial models, three predictors were chosen: INC75_100K (i.e., household density with household income from \$75,000 to \$99,999), IND_EMP (i.e., industry employee density), and AVG_HH_INC (i.e., average household income). In other words, these predictors are more representative than others.

The next steps involve making curve estimations for selected predictors, as well as filtering out some predictors that contribute a very small part of $R^2$. Also, abnormal predictors were deleted from the models. Table 7-8 lists the final refined models for

different buffer sizes. At 1.5 miles, the model reaches the highest $R^2$ value (0.357) and

adjusted $R^2$ value (0.351).


Table 7-7: Model Predictors Selected by Stepwise Procedure

| Buffer Size (miles) | Model Predictors | $R^2$/Adjusted $R^2$ |
|---|---|---|
| 0.50 | INC75_100K, IND_EMP, PERPVEH, AVG_HH_INC | 0.299/0.289 |
| 0.75 | INC75_100K, IND_EMP, AVG_HH_INC, OTHER_ENG, WORKATHOME | 0.335/0.320 |
| 1.00 | INC75_100K, IND_EMP, AVG_HH_INC, OTHER_RACE | 0.340/0.327 |
| 1.25 | INC75_100K, IND_EMP, AVG_HH_INC, TT_GT_60M, OTHER_ENG | 0.343/0.331 |
| 1.50 | INC75_100K, IND_EMP, AVG_HH_INC, TT_GT_60M, OTHER_ENG, MENTL_DISB | 0.360/0.346 |
| 2.00 | INC75_100K, IND_EMP, AVG_HH_INC, MENTL_DISB, COM_EMP | 0.335/0.326 |


Table 7-8: Models for Different Buffer Sizes

| Buffer Size (miles) | Model | $R^2$/Adjusted $R^2$ |
|---|---|---|
| 0.50 | AVO = 1.363-0.039×ln(INC75_100K)-0.032×ln(IND_EMP) -1.66$E$-006 ×AVG_HH_INC | 0.293/0.285 |
| 0.75 | AVO = 1.362-0.042×ln(INC75_100K)-0.032×ln(IND_EMP) -1.79$E$-006 ×AVG_HH_INC | 0.316/0.309 |
| 1.00 | AVO = 1.361-0.045×ln(INC75_100K)-0.032×ln(IND_EMP) -1.87$E$-006 ×AVG_HH_INC | 0.336/0.328 |
| 1.25 | AVO = 1.354-0.047×ln(INC75_100K)-0.031×ln(IND_EMP) -1.81$E$-006 ×AVG_HH_INC | 0.351/0.344 |
| 1.50 | AVO = 1.349-0.049×ln(INC75_100K)-0.030×ln(IND_EMP) -1.75$E$-006 ×AVG_HH_INC | 0.357/0.351 |
| 2.00 | AVO = 1.348-0.047×ln(INC75_100K)-0.031×ln(IND_EMP) -1.68$E$-006 ×AVG_HH_INC | 0.350/0.344 |

During the buffer analysis process, different buffer sizes were applied for urban major arterial and minor arterial subgroups. When mixed arterial buffer sizes (1.5 miles for major arterials and 1.0 mile for minor arterials) were applied to all urban county arterials, the $R^2$ (0.360) for the final model did not improve significantly compared to a uniform buffer size for all urban arterials. Hence, the following model created using a buffer size of 1.50 miles was selected:

$$\text{AVO} = 1.349 - 0.049 \times \ln(\text{INC75\_100K}) - 0.03 \times \ln(\text{IND\_EMP}) - 1.75E\text{-006} \times \text{AVG\_HH\_INC} \tag{7-2}$$

Tables 7-9a and 9-9b give the SPSS output for this model, and Figures 7-2, 7-3, and 7-4 plot the scatter charts for each of the three predictors, respectively.

Table 7-9a: Summary for AVO Model for Arterials in Large Urban Counties

| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate | Durbin-Watson |
|---|---|---|---|---|---|
| 1 | 0.597 | 0.357 | 0.351 | 0.08027 | 1.906 |

Table 7-9b: Coefficients for AVO Model for Arterials in Large Urban Counties

| Model | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. | Collinearity Statistics | |
|---|---|---|---|---|---|---|---|
| | B | Std. Error | Beta | | | Tolerance | VIF |
| (Constant) | 1.342 | 0.020 | | 66.284 | 0.000 | | |
| LOG_INC75_100K | -0.049 | 0.008 | -0.368 | -6.067 | 0.000 | 0.639 | 1.564 |
| LOG_IND_EMP | -0.030 | 0.007 | -0.289 | -4.584 | 0.000 | 0.591 | 1.692 |
| AVG_HH_INC | -1.75E-006 | 0.000 | -0.212 | -3.824 | 0.000 | 0.763 | 1.311 |

Figure 7-2:  AVO vs. Industry Employees
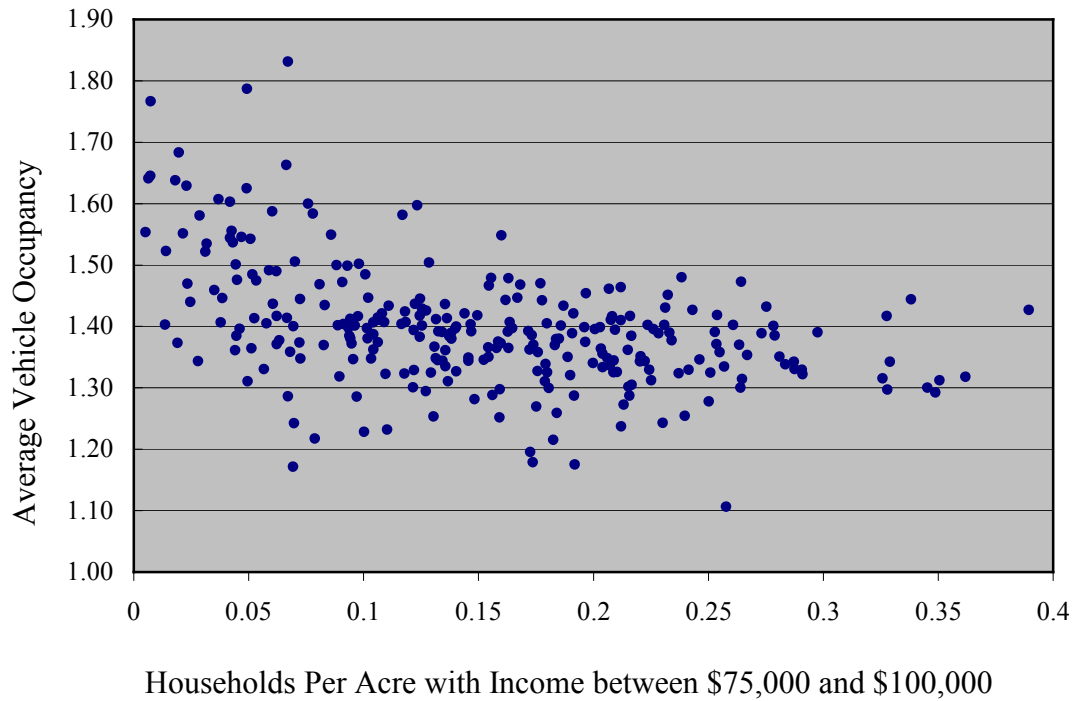


Figure 7-3:  AVO vs. Average House Income Between $75,000 and $100,000

Figure 7-4: AVO vs. Average Household Income

### 7.4.3. AVO Model for Arterials in Medium-Size Urban Counties

Counties with populations under 500,000 and above 100,000 were grouped into the medium-size urban county group. The urban arterials in these counties were aggregated to model the relationship between average vehicle occupancy and socioeconomic characteristics. The same buffering processes used earlier for large urban county arterial models were applied. For different buffer sizes, each model includes some common predictors: HH_2_VEH (i.e., household density with two vehicles), VACANT_HU (i.e., vacant household density), HH_W6_PERS (i.e., household density with six family members), and IND_EMP (i.e., industry employee density). Some other abnormal predictors, such as SPAN_ONLY (i.e., density of population speaking only

Spanish), were also deselected from the final model. At a buffer size of 1.0 mile, the $R^2$ value and adjusted $R^2$ value reach 0.182 and 0.169, respectively. The final model is given as follows:

$$AVO = 1.461 + 3.332 \times HH\_W6\_PERS - 0.376 \times HH\_2\_VEH - 0.131 \times IND\_EMP$$
$$+ 0.151 \times VACANT\_HU + 0.101 \times COM\_EMP \qquad (7\text{-}2)$$

The SPSS output for this model is listed in Tables 7-10a and 7-10b. In this model, the signs of the coefficients for all predictors are significant. For instance, the negative sign for IND_EMP indicates that a higher industry employee density will decrease the AVO due to more work-based trips.

Table 7-10a: Medium-Size County Arterial AVO Weekday Model Summary

| Model | $R$ | $R$ Square | Adjusted $R$ Square | Std. Error of the Estimate | Durbin-Watson |
|---|---|---|---|---|---|
| 1 | 0.427[a] | 0.182 | 0.169 | 0.08941 | 1.671 |

Table 7-10b: Medium-Size County Arterial AVO Weekday Model Coefficients

| Model | Unstandardized Coefficients | | Standardized Coefficients | $t$ | Sig. | Collinearity Statistics | |
|---|---|---|---|---|---|---|---|
| | $B$ | Std. Error | Beta | | | Tolerance | VIF |
| (Constant) | 1.461 | 0.011 | | 134.249 | 0.000 | | |
| HH_2_VEH | -0.376 | 0.051 | -0.611 | -7.427 | 0.000 | 0.381 | 2.621 |
| VACANT_HU | 0.151 | 0.053 | 0.169 | 2.857 | 0.005 | 0.735 | 1.361 |
| HH_W6_PERS | 3.332 | 0.739 | 0.354 | 4.508 | 0.000 | 0.418 | 2.391 |
| IND_EMP | -0.131 | 0.036 | -0.287 | -3.600 | 0.000 | 0.406 | 2.464 |
| COM_EMP | 0.101 | 0.035 | 0.257 | 2.932 | 0.004 | 0.337 | 2.970 |

7.4.4. AVO Model for Urban Freeways

Because freeways are fully access-controlled, the AVOs for freeways reflect vehicle occupancy over a much larger area. The common predictor selected for this

facility by the stepwise method is HH_2_VEH (i.e., household density with two vehicles), and, at a buffer size of 2 miles, the initial model has the highest $R^2$ value. This initial model selected has two predictors: HH_2_VEH and CARPOOLED (i.e., carpooling population density). However, the VIF for each predictor is 5.505, which is greater than four, and, because predictor CARPOOLED contributes a very small part to $R^2$, it was deleted. Tables 7-11a and 7-12b list the SPSS output for the final model, given as follows:

$$AVO = 1.29 - 0.119 \times \ln(HH\_2\_VEH) \qquad (7\text{-}3)$$

This model has an $R^2$ value of 0.386 and an adjusted $R^2$ value of 0.379. Figure 7-5 shows the scattered plot of this model.

Table 7-11a:  Freeway Weekday AVO Model Summary

| Model | $R$ | $R$ Square | Adjusted $R$ Square | Std. Error of the Estimate | Durbin-Watson |
|---|---|---|---|---|---|
| 1 | 0.622 | 0.386 | 0.379 | 0.12384 | 2.001 |

Table 7-11b:  Freeway Weekday AVO Model Coefficients

| Model | Unstandardized Coefficients | | Standardized Coefficients | $t$ | Sig. | Collinearity Statistics | |
|---|---|---|---|---|---|---|---|
| | $B$ | Std. Error | Beta | | | Tolerance | VIF |
| (Constant) | 1.290 | 0.022 | | 57.692 | 0.000 | | |
| LOG_HH_2_VEH | -0.119 | 0.016 | -0.622 | -7.315 | 0.000 | 1.000 | 1.000 |
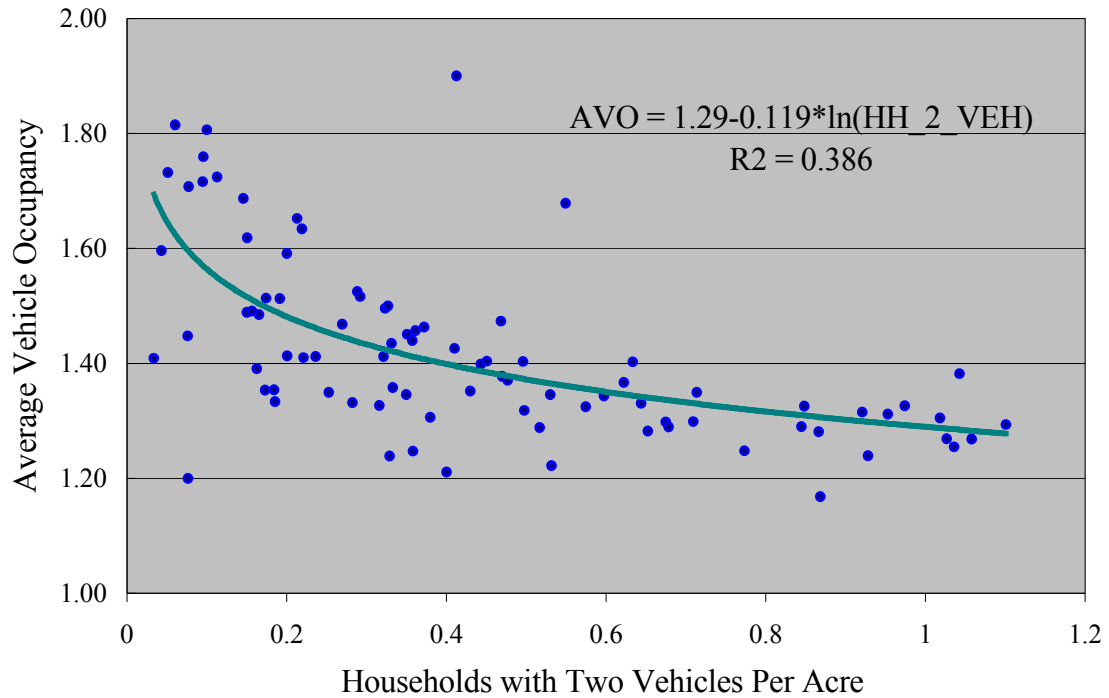
Figure 7-5:  Freeway Weekday AVO Model

7.5. Model Validation

A test of the developed models' predictive capability involves using the models to predict the AVOs for counties, large and medium size urban county arterials, and freeways and then comparing the predicted AVO values with the actual AVO values derived from the accident records.  At the beginning of the analysis, each data set for county area, urban arterials, and freeways is divided into two groups, one used for model calibration, and the other for model validation.  Tables 7-12 and 7-13 summarize the test results and the accumulative percent errors for four models including the countywide weekday AVO prediction model, urban freeway weekday AVO prediction model, large-size urban county arterial weekday AVO prediction model, and medium-size urban county arterial weekday AVO prediction model.  These four models are represented as

models A, B, C, and D separately in both tables.  The mean square of error (MSE) is the sum of the squared differences between actual AVO values and predicted AVO values divided by the sample size minus the number of model parameters.  The total error is the error in percentage for the entire testing data set, which is under 1% for all models.  Table 7-13 gives the cumulative percent errors for the four models.  For instance, for Model A (AVO Model at County Level), 57.14% of the testing counties have an error less than 5%.  The maximum error for each model is also given in the last row of Table 7-13.

Table 7-12:  Model Testing Result Summary

| Model | Model A | Model B | Model C | Model D |
|---|---|---|---|---|
| R-square | 0.439 | 0.386 | 0.357 | 0.182 |
| Adj. R-square | 0.43 | 0.379 | 0.351 | 0.169 |
| MSE | 0.004 | 0.015 | 0.006 | 0.008 |
| Total Error (%) | -0.06 | 0.08 | 0.07 | 0.08 |

Note: Total Error=(Sum of Predict AVO – Sum of AVO)/Sum of AVO * 100

Table 7-13:  Model Testing Result Summary

| Error Range (%) | Cumulative Percentage of Testing Samples | | | |
|---|---|---|---|---|
| | Model A | Model B | Model C | Model D |
| Sample Size | 7 | 9 | 28 | 32 |
| 5 | 57.14 | 55.55 | 64.29 | 62.50 |
| 10 | 71.43 | 77.77 | 78.57 | 75.00 |
| 15 | 100.00 | 77.77 | 82.14 | 84.38 |
| 20 | 100.00 | 88.88 | 82.14 | 96.88 |
| 25 | 100.00 | 88.88 | 100.00 | 100.00 |
| 30 | 100.00 | 100.00 | 100.00 | 100.00 |
| Max Error (%) | 13.80 | 23.00 | 21.24 | 20.58 |

7.6. Summary of Final Models

　　All final weekday AVO models recommended are listed in Table 7-14. These models include those for countywide weekday AVO, large urban county arterial weekday AVO, medium-size urban county arterial weekday AVO, and urban freeway weekday AVO. The model for census blockgroup level weekday AVO is not included as one of the recommended models in the table because of its low $R^2$ value.

Table 7-14: Final Models and Buffer Sizes

| Type | Model Form | $R^2$ / Adjusted $R^2$ | Buffer Size (miles) |
|---|---|---|---|
| County Area Level | AVO = 1.667 - 0.04 × ln(HH_GT_1VEH) | 0.439/0.430 | N/A |
| Large-size Urban County Arterial | AVO = 1.349-0.049×ln(INC75_100K)- 0.03×ln(IND_EMP) -1.75$E$-006 ×AVG_HH_INC | 0.357/0.351 | 1.5 |
| Medium-size Urban County Arterial | AVO = 1.461+3.332 × HH_W6_PERS - 0.376 × HH_2_VEH - 0.131 × IND_EMP +0.151 × VACANT_HU + 0.101 × COM_EMP | 0.182/0.169 | 1.0 |
| Urban Freeway | AVO = 1.29 - 0.119 × ln(HH_2_VEH) | 0.386/0.379 | 2.0 |

　　The $R^2$ values for these models are not high. Because accident vehicle occupancy data do not explicitly represent the context in which accident vehicle trips were made, these data do not deal explicitly with choices people make about how many automobiles to own, where to live, or where to work. Particularly for smaller areas, the vehicles involved in accidents inside these areas may not have much to do with the socioeconomic factors specific to these areas. This is also the reason that the AVO model based on census blockgroup level has a small $R^2$ value. For large areas, however, such as those at

the county level, higher $R^2$ values were achieved because large-scale trends in travel patterns can often be accounted for by changing demographics.

CHAPTER 8

CONCLUSIONS AND RECOMMENDATIONS

This dissertation study has accomplished the following major research objectives: (1) performed an extensive literature review of vehicle occupancy data collection methods, including the accident data extraction method and the feasibility of accident data as a potential source of vehicle occupancy data; (2) performed statistical analysis of vehicle occupancy rates using the accident data with respect to their geographic, temporal, and vehicle coverage design; (3) investigated and identified potential biases associated with the accident vehicle occupancy data as well as a improved procedure used to minimize the existing biases; (4) developed tools to facilitate the collection, processing, analysis, and estimating of vehicle occupancy; (5) used the multiple linear regression analysis method to develop vehicle occupancy forecasting models as a function of local socio-economic data.

The accomplished research tasks and major conclusions are stated in detail below. In addition, recommendations for future studies are also discussed.

8.1. Summary of Research Results and Conclusions

The results and conclusions from each task described above are summarized as follows:

(1) Literature Review

A comprehensive literature search and review has been performed to investigate and assess vehicle occupancy data collection practices. The main review tasks included:

1) the study of vehicle occupancy data collection method coverage, advantages, and disadvantages of all existing methods; 2) the identification of factors that affect vehicle occupancy data collection; 3) the feasibility of making use of accident data to estimate the vehicle occupancy rates, as well as the feasibility of estimation procedures associated with the accident extraction method; 4) an examination of the biases existent in the accident vehicle population, as well as corresponding adjustment methods; 5) a description of the issues associated with the AVO forecasting models. Through the literature review, the feasibility of deriving vehicle occupancy rates from accident data was confirmed, and the research goal and objectives of this project were further determined and explained.

(2) Factors Affecting Vehicle Occupancies

Parametric and nonparametric statistical tests were applied to examine the contributing factors affecting AVOs in the state of Florida. The following factors were analyzed: year, month, day-of-week, time periods, county, driver's age, driver's gender, driver's race, accident severity, facility type, and weather conditions. Both parametric and non-parametric analyses of variance (ANOVA) were applied, and it was found that AVOs were affected by all of the factors examined to various degrees. Despite continued efforts over the past decades to encourage people to carpool, the data show that the overall AVO in Florida has continued to decline over the years, from a high of 1.58 occupants per vehicle in 1990 down to 1.42 occupants per vehicle in 2005, or a 10% decrease. This study also found that a.m. peaks tend to have the lowest AVO in a day and that weekend AVOs were significantly higher than weekday AVOs. In addition,

younger drivers were found to have higher AVOs than older drivers, female drivers tend to have more occupants in their vehicles than male drivers, black and Hispanic Americans tend to have higher AVOs than white Americans, and rural AVOs are significantly higher than urban AVOs.

(3) Potential Biases and Adjustment Factors

As a potential source of vehicle occupancy data, accident data has met with some doubt due to its possible biases. Based on findings from past studies, three possible biases were analyzed and discussed in detail: accident severity, driver's age, and driver's gender. By comparing accident vehicle data with field data, however, analysis results showed that although multi-occupant (two and more than two) vehicles occupy higher percentages of severe accidents (injury and fatality) than do single-occupant vehicles, multi-occupant vehicles in the whole accident vehicle population were not over-represented in the accident database. However, a driver's age bias was found in the accident data. By comparing the year 2000 Miami-Dade County accident data and the 2000 census demographic data, it was found that there is a significant difference between the distribution of the ages of drivers involved in accidents and the ages of the general driving population in the county. A driver's gender bias was also identified. An improved census weighting method was used to adjust for this bias in the AVO estimates. Other potential bias factors may include driver's race and weather conditions. However, adjustment factors for these potential biases cannot currently be developed because of the lack of driving exposure data associated with their population subgroups.

(4) FAVORITE Information System

FAVORITE is a user-friendly information system capable of estimating average vehicle occupancies from multiple years of accident records for Florida's state roadway system.  As a powerful integrated information system, the potential biases resulting from accident severity and driver's age were examined and the corresponding adjustment factors for driver's age were developed and implemented in the system.   Also incorporated in the system was consideration given to the minimum required number of accident records necessary to meet the desired accuracy in the AVO estimates.

The current version of the system includes the 1990-2005 accident data that are cross classified by year, month, day, and hours, making it possible to perform temporal monitoring. The database also includes district, county, roadway section, and area types that can be used for spatial comparisons and monitoring.  Comparisons can also be made by vehicle type, facility type, and accident severity.  Because the system makes use of a comprehensive, statewide database, it can potentially be a highly cost-effective means of monitoring statewide, regional, and site-specific vehicle occupancy trends.

(5) Evaluation of AVO Estimates from Accident Data

A reasonableness check of the results from the FAVORITE system shows AVO estimates that are consistent with expectations.  In addition, comparisons of AVOs from accident data with field estimates show that the two data sources produce results that are generally consistent.   The comparisons also show that the AVO estimates from accident data tend to be slightly higher, most likely because accident records are able to include infants and small children who are often difficult to see during field observation.

(6) AVO Prediction Models

Regression models for the purpose of predicting weekday AVOs were developed in this study using census demographic data integrated with 2002 InfoUSA employment data. Four regression models were developed: a weekday county area level model, a weekday large urban county arterial level model, a medium urban county arterial level model, and a weekday urban freeway model. Essentially, these models expressed the average occupancy rate for arterials and areas as a function of several statistically significant socioeconomic variables.

For the weekday county area level AVO model, the $R^2$ value is 0.452, which is considered acceptable. During the model development, a weekday model for the census blockgroup level was also developed. However, the $R^2$ value for this model is relatively low. Hence, this model was not recommended. The $R^2$ values for the three recommended arterial models are 0.357 for the large urban county arterial level model, 0.386 for the urban freeway model, and 0.182 for the medium-size urban county arterial level model. These models show that such socioeconomic factors as income, vehicle ownership, and employment play a part in vehicle occupancy forecasting models.

Although the AVO prediction models developed were statistically significant, their $R^2$ values are not high. The accident vehicle occupancy data do not explicitly represent the context in which accident vehicle trips were made. As more accident vehicle trip information and driver and passenger information become available, an update of the model parameters and predictors may be conducted to produce more accurate predictions. Consequently, one of the main future studies could involve

developing a survey data method to obtain the needed data and a procedure for automatically integrating these survey data into the current accident database.

8.2. Recommendations for Future Development

Although this study has achieved the proposed goal and objectives, there remains a need for further study. Some major points for the future research are summarized below.

(1) Potential Biases and Adjustment Factors

One major objective of this study is to identify potential biases and develop adjustment factors. The driver's age bias and driver's gender bias, identified in this study, can be adjusted by using the improved census weighting method, or an adjustment factor. If more detailed travel information for the state of Florida, or even for each county, is available, the adjustment factors need to be recalculated to achieve greater accuracy.

While it is not necessary that AVOs derived from accident records be adjusted for potential biases for continuous AVO monitoring, further research may address other potential biases, such as weather conditions and driver's races, and possibly develop the corresponding adjustment factors for the benefit of applications that require absolute AVO estimates.

(2) FAVORITE Information System

Although FAVORITE is already freely available through web download, it is recommended that a web version be developed using ArcGIS Server as the software platform to take advantage of the instant accessibility of web applications. Additional

functionalities can also be developed, for example, the ability to automatically upload new accident records as they become available.

(3) Parameter Adjustments of the Developed Prediction Models

The quality of the AVO prediction models developed in this study still needs to be improved. The accident vehicle occupancy data do not explicitly represent the context in which accident vehicle trips were made. As more accident vehicle trip, driver, and passenger information become available, an update of the model parameters and predictors may also be conducted to produce more accurate predictions. Consequently, one of the main future studies could involve developing a survey data structure to obtain appropriate information, as well as a procedure for automatically integrating these survey data into the current accident database.

REFERENCES

Ahuja, V., and E. Hanscom, "Feasibility Study to Estimate the Average Vehicle Occupancy by Utilizing a Traffic Accident Database," Presented at the Annual Meeting of the Transportation Research Board, Washington, DC, January 1996.

Asante, S. A., L. H. Adams, J. J. Shufon, and J. P. McClean, "Estimating Average Automobile Occupancy from Accident Data in New York State," *Transportation Research Record 1553*, DC, pp. 115-123, 1996.

Barton-Aschman Associates, Inc., *Vehicle Occupancy Determinators*, Final Report Prepared for Arizona Department of Transportation, August 1989.

Benke, R. J. and R. A. Sjoberg, "Auto Occupancy Patameter Variations," *Traffic Engineering*, pp. 21-27, 1977.

Chang, L. Y. and F. L. Mannering, "Predicting Vehicle Occupancies from Accident Data: An Accident Severity Approach," *Transportation Research Record 1635*, Transportation Research Board, Washington, DC, pp. 93-104, 1998.

Demery, L. W., "Peak-period Vehicle Occupancy Statistics for U.S. and Canadian Rapid Bus and Rapid Rail Services," http://www.publictransit.us, October 25, 2002.

Federal Highway Administration, *1977 Nationwide Personal Transportation Study*, 1980.

Federal Highway Administration, *Estimating Auto Occupancy: A review of Methodology*, September 1972.

Ferlis, R. A., "Field Data Collection and Sampling Procedures for Measuring Regional Vehicle Classification and Occupancy," *Transportation Research Record 701*, pp. 1-6, 1979.

Ferlis, R. A., *Guide for Estimating Urban Vehicle Classification and Occupancy*, Final Report Prepared for FHWA Office of Highway Planning, March 1981.

Fisher, R. F., G. J. Williams, and J. P. Boyd, "Atlanta Vehicle Occupancy Monitoring," *Transportation Research Record 779*, pp. 27-32, 1980.

Forbord, R. J., *Twin Cities Modal Split Model*, Minnesota Highway Department, January 1966.

Gan, A., K. Liu, and R. Jung, *Vehicle Occupancy Data Collection Methods (Phase II)*, Final Report Prepared for the Transportation Statistics Office, Florida Department of Transportation, August 2007.

Gan, A., R. Jung, K. Liu, X. Li, and D. Sandoval, *Vehicle Occupancy Data Collection Methods*, Final Report Prepared for the Transportation Statistics Office, Florida Department of Transportation, February 2005.

Gaulin, R., *A Procedure to Calculate Vehicle Occupancy Rates from Traffic Accident Data*, Final Report Prepared for the Connecticut Department of Transportation, Bureau of Planning, April 1991.

Grush, W. and B. Gross, *Vehicle Occupancy: A Comparison of Household Survey Data and Roadside Observation Data*, FHWA Office of Highway Information Management, September 1995.

Heidtman, K., B. Skarpness, and C. Tornow, *Improved Vehicle Occupancy Data Collection Methods*, FHWA Office of Highway Information Management, April 1997.

Kuzmyak, J. R., *Vehicle Occupancy*, 1981.

Levine, N. and M. Wachs, "Factors Affecting Vehicle Occupancy Measurement," *Transportation Research, Part A, Policy and Practice*, Vol. 32A, No. 3, pp.215-229, 1998.

Levine, N. and M. Wachs, *Methodology for Vehicle Occupancy Measurement: Part I, Operations Manual for Measuring Vehicle Occupancy – Instructions and Procedures*, October 1994.

Levine, N. and M. Wachs, *Methodology for Vehicle Occupancy Measurement: Part II, Technical Issues in Vehicle Occupancy Measurement*, October 1994.

Liu, Q. X. and H. Desai, "What do We Know about Vehicle Occupancy? A Statewide Pilot Study in Florida," Compendium of Technical Papers, ITE, 1998.

Lui, K. J., D. McGee, P. Rhodes, and D. Pollock, "An Application of a Conditional Logistic Regression to Study the Effects of Safety Belts, Principal Impact Points, and Car Weights on Drivers' Fatalities," *Journal of Safety Research*, 19, pp. 197-203, 1988.

McClave, J. T., and P. G. Benson, *Statistics for Business and Economics*, 4th ed., Dellen and MacMiallan, 1988.

Mid-America Regional Council (MARC), *2002 Vehicle Occupancy Study for the Kansas City Metropolitan Area*, Final Report, 2002.

O'Donnell, C. and Connor, D., "Predicting the Severity of Motor Vehicle Accident Injuries Using Models of Ordered Multiple Choice," *Accident Analysis and Prevention*, 1996, 28(6), pp.739-753.

Pamanikabud, P., "Morning Vehicle Occupancy Rate Entering the City Center," *Urban Transportation and Environment*, pp. 899-903, 2000.

Pavlidis, I., B. Fritz, P. Symosek, N. P. Papanikolopoulos, V. Morellas, and R. Sfarzo, "Automatic Passenger Counting in the High Occupancy Vehicle (HOV) Lanes," Proceedings of the 9th ITS America Meeting, 1999.

Reed, G. L., H. Desai, and L. B. Krueger, "Determination of Vehicle Occupancy for Small and Medium Sized Areas," Sixth National Conference on Transportation Planning for Small and Medium-Sized Communities, Spokane, Washington, 1998.

Roach, W. T. and M. Lester, *Techniques for Monitoring Auto Occupancy and Seattle Area Research Results*, Final Report Prepared for FHWA Offices of Research and Development, September 1978.

Rutherford, G. S., R. K. Kinchen, and L. N. Jacobson, "Agency Practice for Monitoring Violations of Highway-Occupancy-Vehicle Facilities," *Transportation Research Record 1280*, pp.141-147, 1990.

Shankar, V. N., Mannering, F. L. and Barfield, W., "Statistical Analysis of Accident Severity on Rural Freeways," *Accident Analysis and Prevention*, 28(3), pp. 391-401, 1996.

Tybout, A. M. and Hauser, J. R., "A Marketing Audit Using a Conceptual Model of Consumer Behavior: Application and Evaluation," *Journal of Marketing*, 1981, 45(3), 82-101.

U.S. Department of Transportation, *2001 Nationwide Household Travel Survey*, http://www.bts.gov/publications/highlights_of_the_2001_national_household_travel_survey/, 2001.

Ulberg C. G., *Vehicle Occupancy Forecasting*, Final Report Submitted for Washington State Department of Transportation, February 1994.

Ulberg, C. G. and E. McCormack, "Accuracy and Other Factors Affecting a Continuous Vehicle Occupancy Monitoring Program," *Transportation Research Record 1206*, Transportation Research Board, Washington, DC, pp. 35-47, 1988.

URS Consultants, Inc., *Vehicle Occupancy Data Collection and Analysis*, Final Report Prepared for the Florida Department of Transportation, Systems Planning Office, December 1997.

Zakaria, T., "Cordon Line Highway Survey for the Delaware Valley Region," *Transportation Research Record 1305*, pp. 93-102, 1991.

Zhao, F. and S. Chung, "Estimation of Annual Average Daily Traffic in a Florida County Using GIS and Regression," *Transportation Research Record 1660*, pp. 32-34, 1999.

APPENDICES

## A.1. Weighting Coefficients for Age Groups

Table A-1: Weighting Coefficients for Different Age Groups

| County Name | 16-19 | 20-24 | 25-29 | 30-39 | 40-49 | 50-59 | 60-69 | 70-79 | >=80 |
|---|---|---|---|---|---|---|---|---|---|
| Charlotte | 0.0186 | 0.0385 | 0.0535 | 0.1498 | 0.1813 | 0.1864 | 0.1765 | 0.1307 | 0.0645 |
| Citrus | 0.0205 | 0.0389 | 0.0547 | 0.1524 | 0.1873 | 0.1924 | 0.1735 | 0.1223 | 0.0580 |
| Collier | 0.0228 | 0.0597 | 0.0825 | 0.1980 | 0.1916 | 0.1730 | 0.1412 | 0.0918 | 0.0394 |
| DeSoto | 0.0335 | 0.1024 | 0.1213 | 0.2029 | 0.1802 | 0.1456 | 0.1108 | 0.0714 | 0.0317 |
| Glades | 0.0256 | 0.0688 | 0.1008 | 0.2075 | 0.1981 | 0.1727 | 0.1339 | 0.0673 | 0.0253 |
| Hardee | 0.0372 | 0.1028 | 0.1247 | 0.2299 | 0.1964 | 0.1445 | 0.0852 | 0.0572 | 0.0222 |
| Hendry | 0.0433 | 0.1279 | 0.1235 | 0.2399 | 0.1946 | 0.1391 | 0.0765 | 0.0384 | 0.0169 |
| Hernando | 0.0226 | 0.0483 | 0.0608 | 0.1652 | 0.1922 | 0.1797 | 0.1520 | 0.1252 | 0.0540 |
| Highlands | 0.0253 | 0.0568 | 0.0687 | 0.1580 | 0.1742 | 0.1601 | 0.1582 | 0.1354 | 0.0633 |
| Hillsborough | 0.0289 | 0.0840 | 0.1129 | 0.2500 | 0.2320 | 0.1552 | 0.0721 | 0.0422 | 0.0225 |
| Lake | 0.0223 | 0.0513 | 0.0736 | 0.1952 | 0.1996 | 0.1671 | 0.1434 | 0.1005 | 0.0471 |
| Lee | 0.0219 | 0.0548 | 0.0769 | 0.1911 | 0.2013 | 0.1772 | 0.1362 | 0.0955 | 0.0452 |
| Manatee | 0.0231 | 0.0585 | 0.0810 | 0.1981 | 0.2080 | 0.1670 | 0.1198 | 0.0934 | 0.0511 |
| Pasco | 0.0226 | 0.0511 | 0.0734 | 0.2005 | 0.2031 | 0.1681 | 0.1226 | 0.1007 | 0.0578 |
| Pinellas | 0.0214 | 0.0560 | 0.0829 | 0.2108 | 0.2284 | 0.1721 | 0.1005 | 0.0764 | 0.0515 |
| Polk | 0.0297 | 0.0741 | 0.0947 | 0.2140 | 0.2125 | 0.1655 | 0.1064 | 0.0695 | 0.0335 |
| Sarasota | 0.0186 | 0.0437 | 0.0624 | 0.1697 | 0.2023 | 0.1805 | 0.1410 | 0.1148 | 0.0670 |
| Sumter | 0.0191 | 0.0530 | 0.0786 | 0.1796 | 0.1743 | 0.1760 | 0.1921 | 0.0951 | 0.0322 |
| Alachua | 0.0470 | 0.2029 | 0.1205 | 0.1915 | 0.1979 | 0.1335 | 0.0561 | 0.0322 | 0.0184 |
| Baker | 0.0356 | 0.0862 | 0.1059 | 0.2469 | 0.2415 | 0.1627 | 0.0759 | 0.0310 | 0.0143 |
| Bradford | 0.0277 | 0.0835 | 0.1066 | 0.2390 | 0.2371 | 0.1596 | 0.0808 | 0.0432 | 0.0225 |
| Columbia | 0.0324 | 0.0798 | 0.0927 | 0.2153 | 0.2354 | 0.1758 | 0.0950 | 0.0507 | 0.0229 |
| Dixie | 0.0276 | 0.0705 | 0.0891 | 0.2017 | 0.2164 | 0.1864 | 0.1241 | 0.0618 | 0.0222 |
| Gilchrist | 0.0355 | 0.1390 | 0.0822 | 0.1932 | 0.2145 | 0.1687 | 0.0955 | 0.0485 | 0.0230 |
| Hamilton | 0.0282 | 0.1016 | 0.1120 | 0.2431 | 0.2295 | 0.1538 | 0.0774 | 0.0346 | 0.0198 |
| Lafayette | 0.0286 | 0.0947 | 0.1345 | 0.2555 | 0.1969 | 0.1459 | 0.0842 | 0.0393 | 0.0204 |
| Levy | 0.0291 | 0.0586 | 0.0798 | 0.1988 | 0.2218 | 0.1941 | 0.1221 | 0.0671 | 0.0288 |
| Madison | 0.0341 | 0.0838 | 0.1093 | 0.2211 | 0.2195 | 0.1597 | 0.0917 | 0.0502 | 0.0307 |
| Marion | 0.0259 | 0.0553 | 0.0756 | 0.1921 | 0.2068 | 0.1720 | 0.1349 | 0.0963 | 0.0410 |
| Suwannee | 0.0329 | 0.0724 | 0.0861 | 0.1957 | 0.2203 | 0.1875 | 0.1126 | 0.0603 | 0.0321 |
| Taylor | 0.0304 | 0.0711 | 0.1026 | 0.2160 | 0.2317 | 0.1820 | 0.0942 | 0.0488 | 0.0231 |
| Union | 0.0264 | 0.0710 | 0.1157 | 0.3000 | 0.2599 | 0.1396 | 0.0550 | 0.0228 | 0.0095 |
| Bay | 0.0288 | 0.0760 | 0.0993 | 0.2332 | 0.2367 | 0.1678 | 0.0901 | 0.0479 | 0.0201 |
| Calhoun | 0.0279 | 0.0826 | 0.1193 | 0.2384 | 0.2193 | 0.1536 | 0.0840 | 0.0462 | 0.0288 |
| Escambia | 0.0369 | 0.1014 | 0.1069 | 0.2218 | 0.2209 | 0.1582 | 0.0827 | 0.0470 | 0.0242 |
| Franklin | 0.0207 | 0.0653 | 0.0998 | 0.2249 | 0.2229 | 0.1836 | 0.1068 | 0.0488 | 0.0273 |
| Gadsden | 0.0345 | 0.0820 | 0.1034 | 0.2262 | 0.2439 | 0.1630 | 0.0819 | 0.0414 | 0.0237 |
| Gulf | 0.0249 | 0.0582 | 0.0927 | 0.2223 | 0.2350 | 0.1781 | 0.1088 | 0.0552 | 0.0249 |
| Holmes | 0.0291 | 0.0777 | 0.1050 | 0.2333 | 0.2067 | 0.1732 | 0.0973 | 0.0480 | 0.0297 |
| Jackson | 0.0317 | 0.0831 | 0.1029 | 0.2259 | 0.2271 | 0.1660 | 0.0872 | 0.0472 | 0.0289 |
| Jefferson | 0.0285 | 0.0707 | 0.0880 | 0.2136 | 0.2516 | 0.1884 | 0.0831 | 0.0493 | 0.0268 |
| Leon | 0.0469 | 0.1823 | 0.1203 | 0.2035 | 0.2115 | 0.1407 | 0.0511 | 0.0276 | 0.0160 |

Table A-1:  Weighting Coefficients for Different Age Groups (continued)

| County Name | 16-19 | 20-24 | 25-29 | 30-39 | 40-49 | 50-59 | 60-69 | 70-79 | >=80 |
|---|---|---|---|---|---|---|---|---|---|
| Liberty | 0.0258 | 0.0833 | 0.1375 | 0.2711 | 0.2236 | 0.1385 | 0.0722 | 0.0343 | 0.0137 |
| Okaloosa | 0.0299 | 0.0874 | 0.1048 | 0.2414 | 0.2326 | 0.1572 | 0.0887 | 0.0416 | 0.0166 |
| Santa Rosa | 0.0300 | 0.0610 | 0.0906 | 0.2516 | 0.2562 | 0.1697 | 0.0872 | 0.0388 | 0.0149 |
| Wakulla | 0.0311 | 0.0624 | 0.0917 | 0.2480 | 0.2624 | 0.1743 | 0.0798 | 0.0359 | 0.0145 |
| Walton | 0.0250 | 0.0607 | 0.0896 | 0.2147 | 0.2348 | 0.1868 | 0.1107 | 0.0552 | 0.0226 |
| Washington | 0.0278 | 0.0684 | 0.0921 | 0.2257 | 0.2206 | 0.1833 | 0.0997 | 0.0509 | 0.0315 |
| Brevard | 0.0260 | 0.0575 | 0.0736 | 0.2158 | 0.2368 | 0.1702 | 0.1134 | 0.0725 | 0.0342 |
| Clay | 0.0330 | 0.0675 | 0.0909 | 0.2490 | 0.2559 | 0.1781 | 0.0746 | 0.0342 | 0.0169 |
| Duval | 0.0299 | 0.0872 | 0.1181 | 0.2543 | 0.2368 | 0.1517 | 0.0653 | 0.0372 | 0.0196 |
| Flagler | 0.0212 | 0.0401 | 0.0548 | 0.1602 | 0.2013 | 0.1984 | 0.1741 | 0.1119 | 0.0381 |
| Nassau | 0.0275 | 0.0627 | 0.0869 | 0.2274 | 0.2461 | 0.1912 | 0.0966 | 0.0442 | 0.0173 |
| Orange | 0.0303 | 0.0990 | 0.1252 | 0.2632 | 0.2279 | 0.1391 | 0.0628 | 0.0346 | 0.0179 |
| Putnam | 0.0302 | 0.0669 | 0.0797 | 0.1917 | 0.2228 | 0.1881 | 0.1208 | 0.0712 | 0.0286 |
| Seminole | 0.0283 | 0.0738 | 0.1006 | 0.2494 | 0.2547 | 0.1695 | 0.0684 | 0.0368 | 0.0185 |
| St. Johns | 0.0271 | 0.0578 | 0.0742 | 0.2132 | 0.2596 | 0.1899 | 0.0934 | 0.0579 | 0.0269 |
| Volusia | 0.0269 | 0.0709 | 0.0791 | 0.1969 | 0.2190 | 0.1714 | 0.1123 | 0.0799 | 0.0434 |
| Broward | 0.0246 | 0.0645 | 0.0998 | 0.2533 | 0.2364 | 0.1537 | 0.0741 | 0.0533 | 0.0402 |
| Miami-Dade | 0.0294 | 0.0817 | 0.1120 | 0.2499 | 0.2195 | 0.1526 | 0.0832 | 0.0457 | 0.0261 |
| Indian River | 0.0245 | 0.0513 | 0.0673 | 0.1777 | 0.2064 | 0.1675 | 0.1336 | 0.1125 | 0.0591 |
| Martin | 0.0205 | 0.0469 | 0.0613 | 0.1824 | 0.2154 | 0.1791 | 0.1326 | 0.1062 | 0.0556 |
| Monroe | 0.0178 | 0.0525 | 0.0813 | 0.2189 | 0.2628 | 0.2060 | 0.0924 | 0.0477 | 0.0208 |
| Okeechobee | 0.0346 | 0.0853 | 0.1015 | 0.2178 | 0.2049 | 0.1596 | 0.1076 | 0.0631 | 0.0257 |
| Osceola | 0.0312 | 0.0846 | 0.1105 | 0.2495 | 0.2294 | 0.1579 | 0.0768 | 0.0402 | 0.0201 |
| Palm Beach | 0.0235 | 0.0584 | 0.0835 | 0.2192 | 0.2204 | 0.1577 | 0.0998 | 0.0850 | 0.0524 |
| St. Lucie | 0.0261 | 0.0579 | 0.0767 | 0.2064 | 0.2182 | 0.1643 | 0.1217 | 0.0888 | 0.0399 |

## A.2. AVOs Adjusted by Driver's Age and Age Adjustment Factors

Table A-2: AVOs Adjusted by Driver's Age and Age Adjustment Factors

| County No. | County Name | Unadjusted AVO ($AVO_{unadj}$) | AVO Adjusted by Age ($AVO_{adj}$) | Age Adjustment Factor ($F_{age}$) |
|---|---|---|---|---|
| 1 | Charlotte | 1.526 | 1.507 | 0.987 |
| 2 | Citrus | 1.481 | 1.455 | 0.983 |
| 3 | Collier | 1.567 | 1.539 | 0.982 |
| 4 | DeSoto | 1.653 | 1.638 | 0.991 |
| 5 | Glades | 1.660 | 1.654 | 0.997 |
| 6 | Hardee | 1.642 | 1.621 | 0.988 |
| 7 | Hendry | 1.535 | 1.540 | 1.003 |
| 8 | Hernando | 1.559 | 1.543 | 0.990 |
| 9 | Highlands | 1.577 | 1.545 | 0.980 |
| 10 | Hillsborough | 1.487 | 1.465 | 0.985 |
| 11 | Lake | 1.517 | 1.501 | 0.989 |
| 12 | Lee | 1.517 | 1.482 | 0.977 |
| 13 | Manatee | 1.475 | 1.450 | 0.983 |
| 14 | Pasco | 1.521 | 1.498 | 0.985 |
| 15 | Pinellas | 1.444 | 1.418 | 0.982 |
| 16 | Polk | 1.553 | 1.529 | 0.985 |
| 17 | Sarasota | 1.462 | 1.434 | 0.981 |
| 18 | Sumter | 1.702 | 1.684 | 0.989 |
| 26 | Alachua | 1.462 | 1.456 | 0.996 |
| 27 | Baker | 1.656 | 1.672 | 1.009 |
| 28 | Bradford | 1.533 | 1.512 | 0.986 |
| 29 | Columbia | 1.610 | 1.598 | 0.993 |
| 30 | Dixie | 1.572 | 1.566 | 0.996 |
| 31 | Gilchrist | 1.504 | 1.452 | 0.966 |
| 32 | Hamilton | 1.753 | 1.783 | 1.017 |
| 33 | Lafayette | 1.448 | 1.361 | 0.940 |
| 34 | Levy | 1.594 | 1.574 | 0.987 |
| 35 | Madison | 1.645 | 1.626 | 0.989 |
| 36 | Marion | 1.586 | 1.567 | 0.988 |
| 37 | Suwannee | 1.586 | 1.577 | 0.994 |
| 38 | Taylor | 1.515 | 1.504 | 0.992 |
| 39 | Union | 1.492 | 1.477 | 0.990 |
| 46 | Bay | 1.551 | 1.502 | 0.968 |
| 47 | Calhoun | 1.514 | 1.510 | 0.998 |
| 48 | Escambia | 1.414 | 1.400 | 0.990 |
| 49 | Franklin | 1.573 | 1.566 | 0.996 |
| 50 | Gadsden | 1.594 | 1.569 | 0.984 |
| 51 | Gulf | 1.497 | 1.417 | 0.946 |

Table A-2:  AVOs Adjusted by Driver's Age and Age Adjustment Factors (continued)

| County No. | County Name | Unadjusted AVO ($AVO_{unadj}$) | AVO Adjusted by Age ($AVO_{adj}$) | Age Adjustment Factor ($F_{age}$) |
|---|---|---|---|---|
| 52 | Holmes | 1.648 | 1.634 | 0.991 |
| 53 | Jackson | 1.605 | 1.604 | 0.999 |
| 54 | Jefferson | 1.664 | 1.695 | 1.018 |
| 55 | Leon | 1.450 | 1.436 | 0.990 |
| 56 | Liberty | 1.648 | 1.563 | 0.948 |
| 57 | Okaloosa | 1.496 | 1.478 | 0.988 |
| 58 | Santa Rosa | 1.479 | 1.465 | 0.991 |
| 59 | Wakulla | 1.611 | 1.594 | 0.990 |
| 60 | Walton | 1.618 | 1.605 | 0.992 |
| 61 | Washington | 1.638 | 1.647 | 1.006 |
| 70 | Brevard | 1.456 | 1.435 | 0.986 |
| 71 | Clay | 1.514 | 1.489 | 0.983 |
| 72 | Duval | 1.492 | 1.467 | 0.983 |
| 73 | Flagler | 1.580 | 1.525 | 0.965 |
| 74 | Nassau | 1.545 | 1.525 | 0.987 |
| 75 | Orange | 1.453 | 1.442 | 0.992 |
| 76 | Putnam | 1.538 | 1.499 | 0.974 |
| 77 | Seminole | 1.354 | 1.336 | 0.987 |
| 78 | St. Johns | 1.555 | 1.530 | 0.984 |
| 79 | Volusia | 1.502 | 1.474 | 0.981 |
| 86 | Broward | 1.446 | 1.424 | 0.985 |
| 87 | Miami-Dade | 1.408 | 1.390 | 0.987 |
| 88 | Indian River | 1.519 | 1.492 | 0.983 |
| 89 | Martin | 1.518 | 1.478 | 0.974 |
| 90 | Monroe | 1.634 | 1.608 | 0.984 |
| 91 | Okeechobee | 1.665 | 1.651 | 0.991 |
| 92 | Osceola | 1.658 | 1.664 | 1.003 |
| 93 | Palm Beach | 1.443 | 1.426 | 0.988 |
| 94 | St. Lucie | 1.555 | 1.520 | 0.977 |

## A.3. Weighting Coefficients for Gender Groups

Table A-3: Weighting Coefficients for Driver's Gender Groups

| County No. | County Name | Male | Female |
|---|---|---|---|
| 1 | Charlotte | 0.6148 | 0.3852 |
| 2 | Citrus | 0.6153 | 0.3847 |
| 3 | Collier | 0.6385 | 0.3615 |
| 4 | DeSoto | 0.7032 | 0.2968 |
| 5 | Glades | 0.6882 | 0.3118 |
| 6 | Hardee | 0.6854 | 0.3146 |
| 7 | Hendry | 0.7003 | 0.2997 |
| 8 | Hernando | 0.6101 | 0.3899 |
| 9 | Highlands | 0.6230 | 0.3770 |
| 10 | Hillsborough | 0.6238 | 0.3762 |
| 11 | Lake | 0.6193 | 0.3807 |
| 12 | Lee | 0.6244 | 0.3756 |
| 13 | Manatee | 0.6180 | 0.3820 |
| 14 | Pasco | 0.6141 | 0.3859 |
| 15 | Pinellas | 0.6106 | 0.3894 |
| 16 | Polk | 0.6248 | 0.3752 |
| 17 | Sarasota | 0.6091 | 0.3909 |
| 18 | Sumter | 0.6689 | 0.3311 |
| 26 | Alachua | 0.6242 | 0.3758 |
| 27 | Baker | 0.6662 | 0.3338 |
| 28 | Bradford | 0.6995 | 0.3005 |
| 29 | Columbia | 0.6432 | 0.3568 |
| 30 | Dixie | 0.6737 | 0.3263 |
| 31 | Gilchrist | 0.6711 | 0.3289 |
| 32 | Hamilton | 0.7190 | 0.2810 |
| 33 | Lafayette | 0.7352 | 0.2648 |
| 34 | Levy | 0.6186 | 0.3814 |
| 35 | Madison | 0.6572 | 0.3428 |
| 36 | Marion | 0.6167 | 0.3833 |
| 37 | Suwannee | 0.6239 | 0.3761 |
| 38 | Taylor | 0.6495 | 0.3505 |
| 39 | Union | 0.7856 | 0.2144 |
| 46 | Bay | 0.6307 | 0.3693 |
| 47 | Calhoun | 0.6797 | 0.3203 |
| 48 | Escambia | 0.6334 | 0.3666 |
| 49 | Franklin | 0.7049 | 0.2951 |
| 50 | Gadsden | 0.6075 | 0.3925 |
| 51 | Gulf | 0.6738 | 0.3262 |

Table A-3:  Weighting Coefficients for Driver's Gender Groups (continued)

| County No. | County Name | Male | Female |
|---|---|---|---|
| 52 | Holmes | 0.6692 | 0.3308 |
| 53 | Jackson | 0.6651 | 0.3349 |
| 54 | Jefferson | 0.6511 | 0.3489 |
| 55 | Leon | 0.6109 | 0.3891 |
| 56 | Liberty | 0.7340 | 0.2660 |
| 57 | Okaloosa | 0.6436 | 0.3564 |
| 58 | Santa Rosa | 0.6363 | 0.3637 |
| 59 | Wakulla | 0.6554 | 0.3446 |
| 60 | Walton | 0.6503 | 0.3497 |
| 61 | Washington | 0.6518 | 0.3482 |
| 70 | Brevard | 0.6253 | 0.3747 |
| 71 | Clay | 0.6268 | 0.3732 |
| 72 | Duval | 0.6188 | 0.3812 |
| 73 | Flagler | 0.6154 | 0.3846 |
| 74 | Nassau | 0.6289 | 0.3711 |
| 75 | Orange | 0.6304 | 0.3696 |
| 76 | Putnam | 0.6266 | 0.3734 |
| 77 | Seminole | 0.6238 | 0.3762 |
| 78 | St. Johns | 0.6208 | 0.3792 |
| 79 | Volusia | 0.6211 | 0.3789 |
| 86 | Broward | 0.6160 | 0.3840 |
| 87 | Miami-Dade | 0.6171 | 0.3829 |
| 88 | Indian River | 0.6192 | 0.3808 |
| 89 | Martin | 0.6267 | 0.3733 |
| 90 | Monroe | 0.6707 | 0.3293 |
| 91 | Okeechobee | 0.6749 | 0.3251 |
| 92 | Osceola | 0.6272 | 0.3728 |
| 93 | Palm Beach | 0.6177 | 0.3823 |
| 94 | St. Lucie | 0.6235 | 0.3765 |

## A.4. AVOs Adjusted by Gender and Gender Adjustment Factors

Table A-4: AVOs Adjusted by Gender and Gender Adjustment Factors

| County No. | County Name | Unadjusted AVO ($AVO_{unadj}$) | AVO Adjusted by Gender ($AVO_{adj}$) | Gender Adjustment Factor ($F_{gender}$) |
|---|---|---|---|---|
| 1 | Charlotte | 1.526 | 1.520 | 0.996 |
| 2 | Citrus | 1.481 | 1.476 | 0.997 |
| 3 | Collier | 1.567 | 1.561 | 0.996 |
| 4 | DeSoto | 1.653 | 1.630 | 0.986 |
| 5 | Glades | 1.660 | 1.662 | 1.001 |
| 6 | Hardee | 1.642 | 1.624 | 0.989 |
| 7 | Hendry | 1.535 | 1.522 | 0.991 |
| 8 | Hernando | 1.559 | 1.552 | 0.996 |
| 9 | Highlands | 1.577 | 1.575 | 0.999 |
| 10 | Hillsborough | 1.487 | 1.477 | 0.993 |
| 11 | Lake | 1.517 | 1.507 | 0.993 |
| 12 | Lee | 1.517 | 1.512 | 0.997 |
| 13 | Manatee | 1.475 | 1.468 | 0.995 |
| 14 | Pasco | 1.521 | 1.517 | 0.997 |
| 15 | Pinellas | 1.444 | 1.440 | 0.997 |
| 16 | Polk | 1.553 | 1.540 | 0.992 |
| 17 | Sarasota | 1.462 | 1.459 | 0.998 |
| 18 | Sumter | 1.702 | 1.693 | 0.994 |
| 26 | Alachua | 1.462 | 1.455 | 0.995 |
| 27 | Baker | 1.656 | 1.654 | 0.999 |
| 28 | Bradford | 1.533 | 1.508 | 0.984 |
| 29 | Columbia | 1.610 | 1.593 | 0.989 |
| 30 | Dixie | 1.572 | 1.568 | 0.997 |
| 31 | Gilchrist | 1.504 | 1.478 | 0.983 |
| 32 | Hamilton | 1.753 | 1.726 | 0.985 |
| 33 | Lafayette | 1.448 | 1.434 | 0.990 |
| 34 | Levy | 1.594 | 1.590 | 0.997 |
| 35 | Madison | 1.645 | 1.630 | 0.991 |
| 36 | Marion | 1.586 | 1.573 | 0.992 |
| 37 | Suwannee | 1.586 | 1.572 | 0.991 |
| 38 | Taylor | 1.515 | 1.495 | 0.986 |
| 39 | Union | 1.492 | 1.453 | 0.974 |
| 46 | Bay | 1.551 | 1.542 | 0.994 |
| 47 | Calhoun | 1.514 | 1.514 | 1.000 |
| 48 | Escambia | 1.414 | 1.404 | 0.993 |
| 49 | Franklin | 1.573 | 1.573 | 1.000 |
| 50 | Gadsden | 1.594 | 1.584 | 0.993 |

Table A-4: AVOs Adjusted by Gender and Gender Adjustment Factors (continued)

| County No. | County Name | Unadjusted AVO (AVO$_{unadj}$) | AVO Adjusted by Gender (AVO$_{adj}$) | Gender Adjustment Factor ($F_{gender}$) |
|---|---|---|---|---|
| 51 | Gulf | 1.497 | 1.454 | 0.971 |
| 52 | Holmes | 1.648 | 1.619 | 0.982 |
| 53 | Jackson | 1.605 | 1.581 | 0.985 |
| 54 | Jefferson | 1.664 | 1.664 | 1.000 |
| 55 | Leon | 1.450 | 1.441 | 0.994 |
| 56 | Liberty | 1.648 | 1.588 | 0.964 |
| 57 | Okaloosa | 1.496 | 1.490 | 0.996 |
| 58 | Santa Rosa | 1.479 | 1.474 | 0.997 |
| 59 | Wakulla | 1.611 | 1.606 | 0.997 |
| 60 | Walton | 1.618 | 1.614 | 0.998 |
| 61 | Washington | 1.638 | 1.612 | 0.984 |
| 70 | Brevard | 1.456 | 1.449 | 0.995 |
| 71 | Clay | 1.514 | 1.507 | 0.995 |
| 72 | Duval | 1.492 | 1.481 | 0.992 |
| 73 | Flagler | 1.580 | 1.574 | 0.996 |
| 74 | Nassau | 1.545 | 1.539 | 0.996 |
| 75 | Orange | 1.453 | 1.445 | 0.994 |
| 76 | Putnam | 1.538 | 1.530 | 0.995 |
| 77 | Seminole | 1.354 | 1.349 | 0.997 |
| 78 | St. Johns | 1.555 | 1.547 | 0.995 |
| 79 | Volusia | 1.502 | 1.495 | 0.995 |
| 86 | Broward | 1.446 | 1.439 | 0.995 |
| 87 | Miami-Dade | 1.408 | 1.403 | 0.997 |
| 88 | Indian River | 1.519 | 1.512 | 0.996 |
| 89 | Martin | 1.518 | 1.505 | 0.992 |
| 90 | Monroe | 1.634 | 1.625 | 0.995 |
| 91 | Okeechobee | 1.665 | 1.657 | 0.995 |
| 92 | Osceola | 1.658 | 1.640 | 0.989 |
| 93 | Palm Beach | 1.443 | 1.436 | 0.995 |
| 94 | St. Lucie | 1.555 | 1.546 | 0.994 |

B.1. Socioeconomic Attributes at Census Blockgroup Level

Table B-1:  Socioeconomic Attributes at Census Blockgroup Level

| Attribute Name | Description |
| --- | --- |
| TOTAL_POP | Total population |
| WHITE | Total White population |
| BLACK | Total Black or African-American population |
| INDI_ALASK | Total American Indian and Alaska Native population |
| ASIAN | Total Asian population |
| PACIFIC | Total Native Hawaiian and Other Pacific Islander population |
| OTHER_RACE | Total population of some other race |
| MIXED_RACE | Population of two or more races |
| HISPANIC | Total population of Hispanic or Latino origin |
| MALE | Total male population |
| FEMALE | Total female population |
| AGE_0_4 | Number of persons between 0 and 4 years old |
| AGE_5_9 | Number of persons between 5 and 9 years old |
| AGE_10_14 | Number of persons between 10 and 14 years old |
| AGE_15_17 | Number of persons between 15 and 17 years old |
| AGE_18_19 | Number of persons between 18 and 19 years old |
| AGE_20_24 | Number of persons between 20 and 24 years old |
| AGE_25_29 | Number of persons between 25 and 29 years old |
| AGE_30_39 | Number of persons between 30 and 39 years old |
| AGE_40_49 | Number of persons between 40 and 49 years old |
| AGE_50_59 | Number of persons between 50 and 59 years old |
| AGE_60_69 | Number of persons between 60 and 69 years old |
| AGE_70_79 | Number of persons between 70 and 79 years old |
| AGE80_OVER | Number of persons 80 or above years old |
| MEDIAN_AGE | Median age |
| HOUSE_UNIT | Total number of housing units |
| OCPIED_HU | Number of occupied housing units |
| VACENT_HU | Number of vacant housing units |
| OWNER_HU | Number of housing units occupied by owners |
| RENTER_HU | Number of housing units occupied by renters |
| TOTAL_HH | Total number of households |
| HH_W1_PERS | Households with one person |
| HH_W2_PERS | Households with two persons |
| HH_W3_PERS | Households with three persons |
| HH_W4_PERS | Households with four persons |
| HH_W5_PERS | Households with five persons |
| HH_W6_PERS | Households with six persons |
| HH_W7_PERS | Households with seven or more persons |
| AV_HH_SIZE | Average household size |

Table B-1:  Socioeconomic Attributes at Census Blockgroup Level (continued)

| Attribute Name | Description |
|---|---|
| TOTAL_EMP | Total employees (source: InfoUSA employment data) |
| IND_EMP | Total industry employees (source: InfoUSA employment data) |
| COM_EMP | Total commercial employees (source: InfoUSA employment data) |
| SER_EMP | Total service employees (source: InfoUSA employment data) |
| FORGN_BORN | Foreign-born population |
| DRIVEALONE | Drive alone to work for workers 16 years or over |
| CARPOOLED | Carpooled to work for workers 16 years or over |
| TRANSIT | Take transit to work for workers 16 years or over |
| OTHR_MODES | Use other modes to work for workers 16 years or over |
| WORKATHOME | Work at home for workers 16 years or over |
| TT_LT_15M | Travel time to work less than 14 min for workers 16 years or over |
| TT_15_29M | Travel time to work is 15-29 min for workers 16 years or over |
| TT_30_59M | Travel time to work is 30-59 min for workers 16 years or over |
| TT_GT_60M | Travel time to work is 1 hour or more for workers 16 years or over |
| TTT_LT_30M | Transit travel time to work is less than 30 min for workers 16 years or over |
| TTT_30_44M | Transit travel time to work is 30-44 min for workers 16 years or over |
| TTT_45_59M | Transit travel time to work is 45-59 min for workers 16 years or over |
| TTT_GT_59M | Transit travel time to work is 1 hour or more for workers 16 years or over |
| TOTAL_DISB | Total population 5 years and over with disabilities |
| SENSO_DISB | Total population 5 years and over with sensory disabilities |
| PHY_DISB | Total population 5 years and over with physical disabilities |
| MENTL_DISB | Total population 5 years and over with mental disabilities |
| EMPLY_DISB | Total population 5 years and over with employment disabilities |
| OTHER_DISB | Total population 5 years and over with other disabilities |
| EMPLOYED | Employed population 16 years or over |
| UNEMPLOYED | Unemployed population 16 years or over |
| INC_<10K | Household income less than $10,000 |
| INC_10<20K | Household income from $10,000 to $19,999 |
| INC_20<30K | Household income from $20,000 to $29,999 |
| INC_30<40K | Household income from $30,000 to $39,999 |
| INC_40<50K | Household income from $40,000 to $49,999 |
| INC_50<60K | Household income from $50,000 to $59,999 |
| INC_60<75K | Household income from $60,000 to $74,999 |
| INC75<100K | Household income from $75,000 to $99,999 |
| INC_>100K | Household income $100,000 or more |

Table B-1: Socioeconomic Attributes at Census Blockgroup Level (continued)

| Attribute Name | Description |
|---|---|
| AVG_HH_INC | Average household income |
| HH_0_VEH | Housing units with no vehicle |
| HH_1_VEH | Housing units with one vehicle |
| HH_2_VEH | Housing units with two vehicles |
| HH_3_VEH | Housing units with three vehicles |
| HH_>3_VEH | Housing units with three or more vehicles |

VITA

KAIYU LIU

Sep. 1968      Born, Hunan Province, P. R. China

EDUCATION

Jul. 1991      B.S. in Computer Science and Application
               Shenyang Institute of Aeronautical Industry, Liaoning, P. R. China
Apr. 1994      M.S. in Computer Science
               Beijing University of Aeronautics and Astronautics, Beijing, P. R. China
Dec. 2005      Doctoral Candidate in Civil Engineering
               Florida International University, Miami, FL

EMPLOYMENT

1994 - 1997    Software Engineer and Designer
               China Software Test Center, Beijing, P. R. China
1997 - 1999    Director, Department of Technology
               China Computer Users Weekly, Beijing, P. R. China
1999 - 2001    Visiting Scholar
               School of Computer Science, Florida International University, Miami, FL
2001 - 2007    Senior Software Engineer
               Lehman Center for Transportation Research
               Florida International University, Miami, FL

AFFILIATIONS

Member, Institute of Transportation Engineers (ITE)

PUBLICATIONS AND PRESENTATIONS

1.  Gan, Albert, Haifeng Wang, Kaiyu Liu, Fabian Cevallos, and Ike Ubaka, "An Automated System for Collection and Analysis of Transit Stop Inventories," Accepted for Presentation at the 10th International Conference on Application of Advanced Technologies in Transportation, ASCE, Athens, Greece, May 27-31, 2008.

2.  Gan, Albert, Haifeng Wang, and Kaiyu Liu, "Improving Management and Applications of Highway Safety Improvement Projects through a Statewide Web-based Information System," Accepted for Presentation at *ITE 2008 Technical Conference and Exhibit: Making a Difference in Transportation Safety,* Miami, March 30 - April 2, 2008.

3.  Gan, Albert, Kaiyu Liu, and L. David Shen, "Development of a Prototype Information System for Estimating Average Vehicle Occupancies from Traffic Accident Records," Accepted for Presentation at *the 87th Annual Meeting of the*

*Transportation Research Board* and accepted for Publication in the *Journal of the Transportation Research Board*, National Research Council, Washington, DC, 2008.

4. Liu, Kaiyu and Albert Gan, "Vehicle Occupancy Trends in Florida: Evidence from Traffic Accident Records," Accepted for Presentation at *the 87th Annual Meeting of the Transportation Research Board* and Pending Final Publication Decision in the *Journal of the Transportation Research Board*, National Research Council, Washington, DC, 2008.

5. Gan, Albert, Haifeng Wang, and Kaiyu Liu, "A Web-Based Information System for the Recording and Analysis of Highway Safety Improvement Projects," Accepted for Presentation at *the 87th Annual Meeting of the Transportation Research Board*, National Research Council, Washington, DC, 2008.

6. Gan, Albert, Kaiyu Liu, and Rax Jung, "Occupancy Data Collection Methods (Phase II)," Final Research Report Prepared for Florida DOT, August 2007.

7. Bian, Jie, Peng Zhu, Albert Gan, and Kaiyu Liu, "Analysis of Traffic Crashes on Five Freeways in Miami-Dade County," Final Research Report Submitted to ITS Division, District 6 Florida DOT, April 2007.

8. Liu, Kaiyu and Albert Gan, "A Robust Dynamic Segmentation Tool for Highway Safety Analysis," *Proceedings of the 9th International Conference on Application of Advanced Technologies in Transportation*, ASCE, Chicago, IL, August 2006.

9. Gan, Albert, Kaiyu Liu, and Ike Ubaka, "Florida Transit Geographic Information System (FTGIS)," *Proceedings of the 2005 Conference on GIS in Transit, National Center for Transit Research (NCTR)*, Tampa, FL, November 2005.

10. Gan, Albert, Rax Jung, and Kaiyu Liu, "Occupancy Data Collection Methods," Final Research Report Prepared for Florida DOT, February 2005.

11. Gan, Albert and Kaiyu Liu, "A Database System for the Maintenance and Application of Crash Reduction Factors for Highway Safety Improvements," *Proceedings of the ASCE International Conference on Application of Advanced Technologies in Transportation*, pp. 109-114, Beijing, China, May 2004.