Title
A GPS-based Analysis of Household Travel Behavior

Permalink
https://escholarship.org/uc/item/3hq1f5nb

Authors
Golledge, Reginald G
Zhou, Jack

Publication Date
1999-08-01
A GPS-based Analysis of Household Travel Behavior
by
Jack Zhou
and
Reginald Golledge
Department of Geography
and
Research Unit on Spatial Cognition and Choice
UCSB

Presented at WRSA Annual Meeting, Kauai, Hawaii
August 1999

Acknowledgement: We are grateful to DOT for providing us with a CD-ROM of the Lexington Study and to UCTC Grant #DTRS99-G009
Abstract

While characteristics of daily travel behavior have been determined from analyses of the reconstructed household travel behavior recorded in travel diaries, such reconstructions are subject to criticisms that people lie or falsely recall information about destinations, times of travel, trip purpose, trip destination and other critical characteristics, such as under-reporting of short trips and the number of stops in a trip chain. In 1997 the Department of Transportation carried out a one week study in Lexington, Kentucky in which the cars of 100 households were equipped with GPS and in-car computers. Every stop was logged by the GPS receiver and the purpose of the stop was recorded at that time on an in-car computer. The final report of the study gave descriptions of travel behavior but performed little analysis on the data so collected. Using a CD-ROM data record of all the transactions provided by DOT, we propose to examine questions such as: To what extent were the travel behaviors recorded on each day highly correlated? Are there recurring cyclic patterns that show similar patterns of repetitive behavior on different days of the week? How well can a model calibrated for activity behavior on one day predict behavior on other days? How well does this data support panel and/or diary studies? To what extent do discrete choice models of travel behavior fit the data for different days with approximately the same parameters? To what extent did travelers use optimal travel routes (e.g. shortest path)? What proportion of different trip purposes were undertaken entirely on suburban streets and what proportion relied on different quantities of freeway travel?

Knowing the answers to these questions provides the final pieces to the puzzle of repetitiveness in household travel behavior. It is of considerable importance to the calibration of predictive models of behavior and to successfully implementing Advanced Transportation Management and Information Systems (ATMIS).

Introduction

Goals

A major goal of Transportation Science is to enhance the effectiveness of ITS services. Recently a number of traffic management centers have been developed capable of using multimedia investigation techniques to analyze volumes and patterns of traffic flow. One area to which attention is constantly turning is that related to traffic dynamics. While there are a large number of means for observing traffic, research on traffic flow theory has still lacked that easily accessible and highly accurate geocoded database as a support for building and testing destination choice, traffic flow, and cyclical effects in traffic dynamics. Part of the successful implementation of traffic management centers depends on the adequacy and accuracy of their information services. The more we know about the actual behavior of people not just in the aggregate, but in terms of the many repetitive individual patterns that are likely to be pursued by households living in different parts of the city, the more it is likely that the effectiveness of traffic management services will be enhanced. It is anticipated that our examination of the dynamics of travel behavior as revealed in the GPS generated tracking of destination choices,
will further our knowledge of several still unresearched questions on repetition behavior, path selection, multi-stop trips, cyclic travel patterns, and the relevance of single models for predicting daily travel behavior.

Objectives

Our objectives are to complement and evaluate knowledge gained from traditional diary and survey methods, by adding the effects of real-time on-site data collection using GPS and in-car computer data entry. We also will try to recreate patterns of stops on single and multi-stop trips and predict what routes may have been taken using correctional network based trip allocation models.

Behavioral Geographers are concerned with where people carry out various activities: where they live, work, shop, use other services or find their recreation, and how people come to their travel-related decisions. Traditionally, considerable attention has been given to the study of locational and spatial choice aspects of human activities. It is relatively easy to study the locational attributes attached to people’s activities occurring in a spatial context in GIS. But what is usually overlooked is the time label related to activities, which typically refers to a clock time or calendar time.

Typically, the study of human activities may be approached from the perspective of a spatial-temporal context. One innovative and well-known view is Hagerstrand’s time geography. This approach shows that people’s activities are described as trajectories or paths defined within a bounded region of time and space, from when and where an individual comes into being (birth) to the point when and where he or she ceases to live. As a constrained environment, time and space are inseparable from the intricacies of human behavior. However, using time and space as an approach is contingent upon the other constraints that operate on individuals. For example in the study of the relationship between time and people’s daily travel behavior, at the individual level, it is nearly impossible to extract useful information about the influence of time on people’s choice behavior, about the observable realism of movements made by individuals. Human decisions about travel seem to be random, non—optimal, and to a certain extent, non-periodic, in a temporal sense. Any appearance of intervening opportunities may shatter the seemingly reasonable behavior pattern into pieces. Real world human behavior is different from the scaled-down controlled environment intentionally formed in a lab. Human behaviors are subject to various physical, psychological, social, cultural, and other conditions.

However, It is possible for us to delve into human behavior patterns along a time axis at an aggregate level. Individual variations in the perception, conception, and measurement of space and time, variations in capability constraints, coupling constraints and authority constraints that are encountered by individuals in an environment are sometimes accounted for by averaging. Concerning the temporal aspects, activities of various types are related to the frequency and regularity with which a particular social group chooses to participate in a specified activity. The possible form that such regularities might take are to a certain extent determined by the regional characteristics of the study area, age composition of selected social groups, or the local social-
cultural environment. Typically, trips for different purposes, such as work trips, social and recreational trips, or shopping trips, tend to exhibit quite contrasting time distributions.

Here, we would like to focus our study on variations of regularity of different kinds of trips across a week. This time scale, although relatively coarser than hour times used by other behavioral researchers, is sufficient to reveal the underlying regional travel behavioral pattern and satisfy our needs of trying to identify the change of trip frequency during the period of a week.

Research Area and Data Collection Method

**Research Area:** The study area of this research is located at Lexington, central Kentucky. This area includes two counties — Fayette and Jessamine counties - which encompass an area of approximately 461 square miles with a total population of approximately 350,000. Travel data are collected using an automatic device that collected real-time self-reported travel-related information along with automatically recording GPS position information of the vehicles in use. This new data collection method, compared to the approach of recall-interview or travel diary that has been used a lot in the past, has the advantage of accuracy. When data is entered in real-time, the chance of omitting very short trips or rounding trips times are greatly reduced.

The participants for this travel survey were recruited using a sample plan based on demographic factors. In addition to gender, the sampling objectives were satisfied with the following categories:

- Age 18 - 24 with no children
- Age 18 - 24 with children.
- Age 25 - 49 with no children
- Age 25 - 49 with children
- Age 50 - 64 with or without children.
- Age 65+ with or without children.

During the recruitment process, efforts were made to assure some degree of geographic distribution among the participants within Fayette and Jessamine county planning areas. The adjustment was achieved by altering the recruiting telephone calling patterns based on the postal zip code of households.

**Research Plan:** Much research over the last two decades has paid attention to the study of locational and spatial choice aspects of human activities. The development of Geographical Information Systems (GIS) has provided an ability to produce geocoded spatial databases and has consequently assisted in the determination of relevant variables for destination choice. The temporal sequencing of activities has also received increasing attention. Access to the GPS/in-car computer data record from the Lexington study provides an opportunity to combine both real-time data input with GPS locational accuracy.
Typically the study of human activities can be approached from the perspective of a spatial-temporal context. Following the fundamental ideas expressed in Hagerstrand’s “Time Geography” (1970, 1976) and those of his followers such as Lentorp (1978), and more recently the works of Miller (1991, 1999), a spatio-temporal approach to the examination of people’s activities should be able to focus on the correlation between daily trajectories. Various diary surveys have concluded that on a daily basis many activities are non-periodic, and destination choices are difficult to rationalize. We plan in this study to use the GPS-based data from Lexington to look in detail at the actual time paths of individuals, to determine to what extent time-paths of individuals or groups can be best fit by simple decision making models based on criteria such as shortest path and least time. We also plan to explicitly examine trip chains and discover what regularity exists both in the periodic occurrence of similar chains and the sequences in which particular activities are performed.

BACKGROUND

The past decades have seen a paradigm shift in transportation planning from the construction of new infrastructure (supply) to the more effective management of travel (demand). Part of the reason for this shift was the recognition that building new highways was only a temporary measure to relieve movement problems such as congestion. The shift to travel demand management as a significant traffic control strategy has consolidated in the last decade. Both the spatial and the temporal dimensions of travel behavior are being examined and increasing emphasis has been placed on flex-time working hours, telecommuting, and an increasing concern with the in-car dynamic reception and use of Advanced Transportation Management and Information Systems (ATMIS). These measures are designed to facilitate movement through existing systems by: a.) reducing travel demand through the suppression and selective elimination of trips; b.) targeting single occupant vehicles at peak period commuting times, and reducing traffic volume on key links in the period of peak vehicle flow; c.) reducing driver frustration and stress along with affecting traffic flow by providing timely in-car, en-route or pre-travel information about hazards such as congestion, construction, or accidents; and d.) examining in more detail the activity patterns of individuals and households to more completely understand features such as the allocation of resources (e.g. vehicles) among household members, the timing of household activities, and the significance of multipurpose and multi-stop trips (trip chaining) in the episodic activity behavior of household members. In particular, the latter trend has attempted to treat travel behavior in more realistic terms: this has required a search both for new data that is being produced either by survey and panel research, by travel simulators, by new types of travel demand models (see Mahmassani, Hatcher, and Caplice, 1996; Stopher, 1996a; Jones 1990; van Aerde, Hellinga, Baker, and Rakha, 1996), and in the future, by GPS based in-car tracking systems.

Apart from working within the idea of constrained activity spaces, this research began emphasizing behavioral changes or behavioral dynamics represented by human decision making and choice behavior when confronted with changes in the travel environment. These changes could vary from the process of switching between driving alone and carpooling to work, to the more real-time adjustment of changing destinations, changing routes, substituting destinations, changing the time scale at which activities are undertaken, deleting and delaying activities as a
response to information about changing travel environments, or delaying departure times. Other travel behavior features that have come under investigation include trip chaining, scheduling of activities over a time span rather than restricting them to a single time, substituting out-of-home for in-home activities (such as might be the case with two-person-working households who begin dining out, instead of eating at home in the evenings), and an emphasis on household members’ life cycle stages (Mahmassani and Herman, 1990; Stopher and Lee-Gosselin, 1997; Gärling and Hirtle, 1990; Janelle, Klinkenberg, and Goodchild, 1998).

One of the characteristics of the activity approach is that it extends interest in what is going on beyond the physical nature of the trip itself Individual cognitions (e.g., cognitive maps), person-to-person and among-household relationships, and other coupling phenomena (e.g., working out with a friend, sharing rides to a transit terminal) as well as phenomena such as variability in path selection criteria for different trip purposes, all come into play in the attempt to understand the reasons behind movement. This detailed personal and small group knowledge is often extremely difficult to obtain and to code and process (Janelle et al. 1998; Goodchild, 1998). Nevertheless, there is substantial evidence that the benefits associated with adopting the methods, the ideas, and the concerns of this general behavioral approach appear to more than compensate for those difficulties, particularly by giving increased knowledge of the nature and structure of decision and choice processes of travelers. In the spirit of these new research directions, we propose the analysis of in-car GPS and real time computer data entry procedures as represented in a novel data collection projects (The Lexington Project), with a view to searching for new insights about household travel and the evaluation of the GPS based methodology.

**Driver Simulators and Computational Process Models**

As the potential for ATMIS to influence driver behavior and traffic conditions in a network has become more obvious, a number of research efforts have focused on examining the impacts of real-time traffic condition information on dynamic driver behavior. Because few ATMIS systems have been implemented in the real world, much research has concentrated on using computer-based interactive simulation rather than working with real-time field studies. Driver simulation procedures differ from the more classic revealed preference studies in that, while the former require individuals to answer hypothetical questions about technologies they have yet to experience, in a driver simulator subjects are given the opportunity to experience different traffic or information scenarios such that their decision and choice processes are revealed by the consequent actions they take. Those examining driver behavior under ATMIS conditions in recent years include Bonsall and Parry (1991), Ayland and Bright (1991), Koutsopoulos, Lotan, and Yang (1994), Chen and Mahmassani (1993), Vaughn, Abdel-Aty, Kitamura, and Jovanis (1993), Adler, Recker and McNally (1992a, b, and 1993; and Hu, Rothery, and Mahmassani (1992).

Adler et al. (1992a, b, and 1993) have developed an interactive computer-based simulator named FASTCARS. Its purpose is to gather data for estimating and calibrating predictive models of driver behavior under conditions of real-time information. It was written in Turbo Pascal, and
designed to run on a 386-series PC running at least 33 megahertz and equipped with VGA graphics and a voice adapter. The authors claim that it is not a pure driving simulator, but rather simulates real-time travel decision-making conditions. It presents a series of possible environments tied to the experience that a subject has with the basic environment and studies temporal and spatial factors such as perceptions of speed and volume, time lapse, network familiarity, information acquisition, and travel goal specification and evaluation. The simulator encompasses the entire driving process from pre-trip planning to arrival at the destination. During the trip, players are required to make a range of choices, including specification of goals, rerouting where necessary, changing lanes, and making decisions as to whether or not to use specific information technologies, such as in-car guidance or advisory signage. Pre-trip planning involves selection of departure time and initial route choice. In addition, travel objectives for each trip must be specified. During post-trip debriefing, subjects evaluate their success in meeting their pre-trip goals.

Researchers at MIT have also developed an interactive simulator to facilitate data collection and calibration of a route choice model. Their simulator uses fuzzy set theory, fuzzy control, and approximate reasoning (Koutsopoulos, Polydoropoulos, and Ben-Akiva, 1995). The model is loosely based on a previous study of the dynamics of driver behavior under conditions of provision of real-time information (IGOR: Interactive Guidance On Routes, Bonsall and Parry, 1991). IGOR simulates en-route travel through a network and emulates an in-vehicle navigation system to provide players with real-time route guidance so that drivers’ compliance with guidance advice can be evaluated. The quality of advice is manipulated, and the relationship between advice quality and advice acceptance was determined. The Koutsopoulos et al. (1994) model enhances the use of the interface, allows for modeling different operating conditions, improves the information provision capabilities available to the driver, and accounts for the driving task. A graphic display shows a car moving through a network and information is presented through a roadside display/broadcasting system as well as a graphical in-vehicle information window.

Chen and Mahmassani (1993) also developed a simulator that integrates a traffic simulator program and offers the capability for multiple driver participants (DYNASMART). This models pre-trip planning, en-route travel, and post-trip evaluation. Pre-trip planning involves selection of a departure time and a path. At the selected departure time, players see a display of the network with expected travel time for specific routes. An option is provided to allow departure on time or to delay the trip. The initial route is selected at the time of departure. The explicit purpose behind building this simulator was to examine the behavioral processes underlying commuter decisions on route diversions including en-route and day-to-day departure time and route choices as influenced by the provision of real-time traffic information. Three components are visually displayed by this simulator in en-route conditions: a network illustrator, a legend window for explaining color codings, and a real-time message display. Real-time updates of vehicle location are provided and where turn decisions have to be made, this information set is analyzed to determine whether the current route will be continued or an alternate path selected. The emphasis in the Texas simulator is on investigating day-to-day adjustments. It allows manipulation of departure times, offers a capability for real-time interaction with and among multiple driver participants within a traffic network, and considers both system performance as influenced by driver response to real-time traffic information and driver behavior as influenced by real-time
traffic information. The Texas simulator actually simulates traffic conditions, for its engine is a traffic flow simulator and ATIS information generator that displays information consistent with the processes actually taking place in the simulated traffic system. This dynamic approach allows the researcher to investigate day-to-day evolution of individual decisions under different information strategies. Thus driver learning behavior is allowed; this provides a longer-term dimension to simulation of driver behavior than is possible with other models.

Using driver simulators then provides data that acts as a basis for the development of user response models that in turn influence simulation assignment tools and their evaluation of network performance under conditions of real-time information provision (Mahmassani, 1996). Although currently state-of-the-art in terms of the types of information that can be obtained from driver responses to changing real-time traffic conditions, the various simulation experiments are not intended to totally replace actual field demonstrations and tests, but to provide information on what conceivably may happen as different traffic condition and information flow parameters are manipulated. Solving such issues are critical to the further development of IVHS technologies.

Adler et al. (1993) and Adler (1997) suggest that real-world implementation of ATMIS will involve multiple media formats which are both auditory and visual as well as varied message contents, route guidance and traffic condition information, and information display formats. Some information may be most suited to roadside posting and be passively available to all drivers even without appropriately modified in-car vehicle guidance systems. HAR, for example, would be available to most drivers but would require active acquisition via advisory radio signals. This in turn requires a deliberate effort by drivers to tune to the advisory radio station. Yet other information will be available only to a subset of drivers who will pay extra for this service but also have to make an active decision as to when to acquire it. The cost of an in-vehicle guidance system for general consumption is still being explored and simulation experiments are still being undertaken to determine drivers’ willingness to both acquire and use ATMIS of different complexity. The Adler, Recker and McNally implementation of FASTCARS (Freeway and Arterial Street Traffic Conflict Arousal and Resolution Simulator) was specifically designed to address this question.

Multiobjective travel planning has also been examined using computational process modeling (CPM). For example, Gärling et al. (1994) and Kwan (1995) have presented related models (SCHEDULER and GISICAS) that model the scheduled trip behavior of household members over specific time periods (e.g., 24 hours). Unfortunately, their models have not fully explored the problems of using an priori scheduled set of activities, which throughout the day are influenced by changing traffic conditions such as congestion and travel delays, to explore how the initial priorities set on different trip purposes can result in adjustments such as rerouting, activity deletion, activity delay, destination substitution, and activity rescheduling.

Simulation models such as those described above still need to be evaluated against real world activities. Few have been able to do this well because of the lack of precision in available panel or diary based data sets, where criticisms of the errors induced by self report delays and recall are often mentioned as a prime limiting factor. For example, perhaps the most important part of CPM is the Sequencer module that fits together trip purpose, activity duration, activity
salience, and feasible destinations. We expect that analysis of the GPS tracks of travel and destination selection in the Lexington data will contribute to obtaining information critical to the design of such trip sequencing modules.

**Choice Rules and Strategies**

Mahmassani (1996) has suggested that the behavior of repetitive travelers is guided by simple heuristic strategies and a limited set of mental choice rules. As such it has been necessary to depart somewhat from the formal utility maximizing paradigm, replacing this rigid constraint with the more flexible one that travelers act in a boundedly rational manner by searching for an “acceptable” outcome. In previous work, (Mahmassani and Chang, 1985, 1987) and in work by Supernak (1992) evidence is gathered from psychological and behavioral decision theory literature to show that boundedly rational processes were more realistic and reliable predictors of travel behavior than were predictions made on the basis of utility maximization. Mahmassani (1996) suggests that the first strategy concerns the willingness of a commuter to change their latest choice of route, departure time, or both, and that these choices are made conditional on each other.

A second set of influences concern the degree to which a potential traveler is familiar with route and traffic conditions, and the extent to which exogenous information is likely to be accepted. Important concepts include the traveler’s preferred arrival time at a destination which are known to be dependent on attitudes towards risk as well as conditions in the workplace itself (e.g., traffic volume, congestion, parking availability). The boundedly rational character of the decision process is operationalized via a satisficing rule. This specifies that the user does not change departure time if the schedule delay (on day I) is within a user-specified indifference band or tolerance band. The limits of this tolerance band represent the earliest acceptable arrival times. Information about trip delay and experienced congestion, influence the location of the upper and lower limits of the indifference band.

Route selection requires another set of rules. As Gärling and Hirtle (1990) have indicated, many trips may be regarded and modeled as a set of choices in which local decisions relating to route switching or selection of route segments to complete a trip chain, may result in behaviors that are other than globally optimal. For example, a traveler’s schedule on a particular day may involve a trip to the workplace, a trip from workplace to another destination (e.g., lunch), a return to work trip, a trip to recreation or social activities from work, a trip from recreation/social to shopping and a trip from shopping to home in the evening. On different segments of such a trip’ criteria for path selection may change from minimizing time or distance to maximizing aesthetics, maximizing use of freeway segments, or restricting travel to arterial or local streets such that signalization at intersections is minimized. The result may be a trip that is perfectly acceptable and satisfactory for the traveler but which significantly exceeds traditional time, cost or distance minimization rules. We anticipate that we should be able to do data mining on the Lexington material to investigate this hypothesis.
In general it is assumed that the purpose of providing exogenous information to travelers en route is to help them optimize route selection and overall to optimize system performance. But, what if this is not relevant for most habitual or repetitive trips? What then is the purpose of ATMIS? Experimentation with different choice rules has only recently begun (e.g., Mahmassani, 1996). There appears to be considerable potential, both in the use of driver simulators and the use of Geographical Information Systems (GIS) to explore how different decision rules, when implemented, impact features such as departure time and path selection (see Golledge and Egenhofer, 1998, for a selection of articles of these problems). As the traffic system evolves throughout the day, travel plans developed prior to initial departure, may have to be modified or altered. These changes invariably are related to the type, amount, time and reliability of information that can be accessed by the potential traveler.

En-route decisions and Information Processing

Included among the temporal and spatial factors that are involved in en-route driver decision making and information processing are those such as perception of speed, perception of traffic volume, perception of time lapse associated with completing segments of the designated route, familiarity with the network through which travel takes place, the ease and rate of information acquisition about the driving environment, travel goal specification and destination choice, evaluation of the priority of the goals associated with a specific trip, knowledge of landmarks and waypoints on and off a particular route, and perception of and familiarity with areas through which a route passes (including perceived safety, perceived areas to be traveled, and perceived feasibility for destination substitution). These factors appear to be elemental to the decision-making processes of trip making, goal specification, route choice, information search, and reaction to possible diversion. They all lend themselves to specification within experimental conditions and will be manipulated to produce participant reactions.

The information needed to allow the previously specified decision-making processes to work include knowledge of the location of the destination at which a goal can be achieved, plus information about a feasible set of possible alternative destination sites; information that will influence route segment selection and the overall configuration of the most satisfactory route for achieving goals; information about congestion and its applicability to different lanes along which traffic flows; and the availability of information technologies to provide input to each of these decision making strategies. These information technologies can be embedded in ATMIS, signalization, and radio broadcasts. Both general network information and individualized knowledge structures can be combined to provide the basic en-route scenarios that allow an individual to either conform to or change route selection criteria during a trip, or provides the opportunity for driver preferences for handling both network and traffic conditions to be expressed in the ongoing decision-making process. Since only observed trip purposes and destinations are recorded in the Lexington data, there will be little direct contribution to this segment of travel behavior modeling, but we will see if new insights may be generated that would impact variable selection for these purposes.
Over the years the California PATH institution sponsored a number of projects and a
number of investigators who have conducted research on activity patterns and travel behavior
(Recker, 1995; Recker, Root, and McNally, 1983; Recker, McNally, and Root, 1985; McNally,
1997, 1998; McNally and Recker, 1987; Wang, 1996). Currently there is little active work on
tests of the efficacy of tracking vehicles and collecting in-car data as part of ongoing research on
activity patterns, although some studies have been proposed at UC Santa Barbara (Fohi, Curtin,
Goodchild, and Church, 1996), which relate to the use of GPS-based traveler information
systems. Other research includes that of Farrell, Barth, Galijan, and Sinko (1998) but the latter is
more concerned with GPS as an aid to INS than to providing data on recurring activity patterns.

Methodology

It is possible to delve into human behavior patterns along a time axis at both an individual
and an aggregate level. Individual variations in the perception, cognition, and measurement of
space and time, and variations in capability constraints, coupling constraints, and authority
constraints, may be averaged to produce an aggregate constraints picture when individuals are
 grouped in different ways. Activity patterns are related to the frequency and regularity with
which particular social groups choose to participate in those activities. A possible form that such
regularities might take is to a certain extent determined by the study area itself, demographics of
the local population, and the sociocultural environment. For different groups it appears that quite
different patterns of work, social, recreational, and shopping trips occur (Janelle et al 1998).

In this study we plan to focus on variations or regularity of different kinds of trips across a
seven day period. This time scale can reveal the underlying local patterns and help to identify
changes of trip frequency and episodic intervals of trip making over the weekly period.

Our methodology is a pioneer method of data collection compared to the diary recall-
interview approached often used in data collected on travel behavior. It is suggested that the GPS
and real-time in-car computer recording of travel purposes will have a considerable advantage in
terms of accuracy and should not suffer from misrecall, deliberate or other forms of
misstatements. In particular, one of the major problems in many travel diary studies is the
omittance of very short trips or the rounding of trip times to selected time intervals. GPS
recording can provide more accurate locational information and can tie it more precisely into
specific times.

Since time can effect people’s travel choices in different ways, the first step in our
methodology is to create an approximately complete set of activity choices in order to evaluate
the potential influence of time on people’s actions. This provides the basic temporal framework
of the sample’s activity spaces. This will allow us to make comparisons say between the
temporal pattern and activity structure on Thursday as compared to Friday, or Sunday as opposed
to Wednesday. The specific classes of activity in the Lexington data included trips to pick up
passengers, work trips, return home trips, shopping, religion, drop off passengers, work related
business, trips to school, college, or university, eating out, social or recreational, medical or
dental, and other errands.
Comparing all possible combinations of two days in a week with respect to household travel behavior, may or may not yield significant differences between the two patterns (e.g. maybe Monday and Friday events will be correlated, while Sunday and Wednesday will not). Using multiple criterion measures such as the K-group multivariate analysis of variance (MANOVA) and discriminate analysis however can help determine if differences do exist.

Based on the days of the week the next step is to reformat the collected trip counts from the Lexington travel survey data into seven daily groups (Sunday through Saturday). Trip counts for each day of the week will then be further divided based on different trip purposes or trip types as dependent variables. Using the K-group MANOVA capability we can compare 13 different variables simultaneously for each of the seven working days with a null hypotheses of no significant difference between the patterns produced on any given day. In other words, the null hypothesis suggests there is no significant difference between people’s activities across the seven days of the week. While a number of travel diary studies have given empirical evidence that trip making varies significantly on different days (e.g. Hanson and Huff, 1982, 1983, 1985, 1988; Timmermans and Golledge, 1990), to our knowledge no statistical testing has been put forward that explicitly measures the degree of coincidence of trips across such a seven day period. We anticipate we will reject the null hypothesis and support the general hypothesis that people’s activities differ over the set of 13 trip variables and the seven week days.

Following this we will use discriminate analysis for describing major differences among the seven day groups previously used in the multiple analysis of variance. Using seven week days and 13 trip purposes the number of possible discriminate functions is six. The coefficients for each of the trip variables on the six discriminate functions will then be examined. The structure matrix will also be examined to show the correlation between each discriminate function and each of the original trip purposes. Generally it is assumed that greatest stability of the function variable correlations found in the structure matrix exist in small or medium sized samples, especially when there are high or fairly high interrelations among the variables. Also the correlations give a direct indication of which variables are most closely aligned with the unobserved trait which the canonical variables (discriminate function) represents. We will also determine variable redundancy. A useful device for determining directional differences among the groups when two or more discriminate functions exist is in to graph them in the discriminate plane. In the horizontal direction, correspondents of the first discriminate function and lateral separation among the groups indicates how much they have been differentiated on the basis of this function. The vertical dimension corresponds to the second discriminate function and tells which groups are being differentiated in a way unrelated to the way that they were separated on the first discriminate function. We will use this graphical display method to search for other regularities among variables and groups. This should help us decide which trip purposes are related to other trip purposes on which particular day or days of the week. This explores the basic nature of multi-purpose or multi-stop trips, as well as giving insights on trip sequencing.

Finally, post-hoc procedures using Hotelling’s T2 and univariate T tests will be undertaken. This is done to find out which groups and which variables are responsible for the overall association found in the discriminate analysis. Hotelling’s multivariate tests (T2) and univariate T-tests determine which pair or groups differ significantly on the set of trip variables. This
should highlight which days of the week are most similar to each other and which trip purposes are most frequently repeated on those days to create the patterns of similarity.

A second problem will be to compare the trips made by individuals to those suggested by various optimization network flow models. At this stage we simply plan to use path selection models found in TRANSCAD and the expanded ArcInfo Network module. Basically the study area has been geocoded and the network exists as a node-link database. Using ArcInfo functionalities we can determine distance and times from specific origins to given destinations in the geocoded network, and compare them with the actual trips and temporal intervals between origin and destination GPS-based recordings of locations. Although many conceptual criticisms have been made of standard network flow optimization models, there have been very few opportunities to compare the specific routes taken by real-world travelers with those predicted by network models.

To do model comparisons, we plan to construct origin destination matrices for both actual and different model predictions and then, using an innovative Quadratic Assignment Program (QAP) that is designed to determine which model most accurately reproduces the original data, we should be able to evaluate the accuracy of each of the different model types. The QAP method (Hubert, Golledge, and Costanzo, 1982; Hubert and Golledge, 1981, 1982; Hubert and Schultz, 1976) takes as its critical operating factors an original distance, similarities, or correlation matrix calculated on the real data, and compares via a cross-product procedure the degree of coincidence with a model-predicted database. This is done by calculating an initial cross-product statistic, then randomly commuting the rows and columns of the model matrix to create a distribution of possible outcomes. The procedure is distribution free and generates its own reference distribution. The proportion of randomly commuted matrices that give better or worse fits than the original model then define the significance level of the measured association.

We may also use Factor Analysis to examine the relations among the set of random variables observed or counted or measured, which are associated with trips. In this case, we view the trips recorded in the Lexington area as a trip group. Obviously, each trip made by a household will be an ‘individual entity’ within this group. The random variables of the set to be analyzed associated with trips may consist of any attributes on which the individual trips of the group differ. With transportation problems, the random variables we are most concerned with are probably trip time, trip length, and trip purpose. The contributing factors could be trip types, day of week on which the trip happened, and so on.

Travel behavior is typically related to the movement of individuals from place to place, with the emphasis being upon the actual quantitative changes that take place and the nature of the localities concerned. One important aspect of travel trips as representations of people’s movements in a two dimensional space is the physical direction of the movements. Once information about the trips made by people has been recorded, inherently the change of direction of the trips are available to us. These kinds of data that get direction measurement involved are usually called circular data. When dealing with problems arising with circular data, we typically use some relatively unusual statistics.
For example, we know that one of inherent attributes about circular data is its periodicity. In a circular world, simple arithmetic operations like addition do not work. Suppose we have one direction, say 200 degrees from north; if we add 360 degrees to it, we are not getting 560, but will return to the original direction. This unusual form of addition can clearly cause problems, especially in the case where we have the individual trip orientation information of a study area and intend to collectively determine the local main traffic flow direction (the average of trip orientation). A solution for this could be taking the vector means rather than using the traditional addition operation. For the Lexington travel data, the trip information has been collected within a period of a week. Depending on the trip types to which each trip belongs and day of week on which the trip happened, trips may be long-distance or short-distance. By following trips and breaking the trips at those main turning points, we may divide each trip into a set of vector segments and then compute the mean orientation for the single trip. A very good representation of the final mean trip orientation data would be a circular histogram. Hopefully, we can show a trip flow direction circular histogram for each day of the week and dynamically link them using the Java media player. The final result would be a vivid demonstration of the variation of traffic flow direction change in the Lexington area across a week.

A Case Study: Variations in Travel by Days of the Week

Analytical Method used and Results

K-group Multivariate Analysis of Variance and Discriminant Analysis

Time might affect people’s travel choice in more then one way. An approximately complete set of activity choice may be formed to evaluate the potential influence of time on people’s activity choice. The result derived will help us understand the temporal nature of activity spaces with respect to the overall mix of activities in which people participate. For instance, if we were comparing people’s activities on the two days of a week, say, Thursday and Friday, we would obtain a more detailed and informative breakdown of the differential effects of time on people’s travel behavior if they were split into its detailed types: pick up passenger, return home, work place, etc. Comparing two days of a week on total trip counts made by a household might yield no significant difference; however, the multiple criterion measures mentioned above may make such a difference.

Based on days of the week we reformatted the collected trip counts from the Lexington Travel Survey Data into groups — Monday, Tuesday Wednesday, Thursday Friday, Saturday and Sunday. Trips counts for each day of the week were further divided based on different trip purposes or trip types as dependent variables. Table one shows the trip types used. Because of the limited number of trips recorded for trip types -9, 13, 14, 16, 17 and 18, they are not used as dependent variables in this study. With the K group MANOVA capability provided by SPSS, we compare the 7 groups on the remaining 13 dependent variables simultaneously. Our null hypothesis for MANOVA analysis is: $H_0: \mu_1=\mu_2=\mu_3=\mu_4=\mu_5=\mu_6=\mu_7$

(Population mean vectors are equal. Namely, there is no difference on people’s travel activities across the week.)
Table 1. Activity types

<table>
<thead>
<tr>
<th></th>
<th>Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Pick Up Passenger</td>
</tr>
<tr>
<td>2</td>
<td>Drop Off Passenger</td>
</tr>
<tr>
<td>3</td>
<td>Work Place</td>
</tr>
<tr>
<td>4</td>
<td>Work-Related Business</td>
</tr>
<tr>
<td>5</td>
<td>School, College, University</td>
</tr>
<tr>
<td>6</td>
<td>Shopping</td>
</tr>
<tr>
<td>7</td>
<td>Other Errands</td>
</tr>
<tr>
<td>8</td>
<td>Eat Out</td>
</tr>
<tr>
<td>9</td>
<td>Social or Recreational</td>
</tr>
<tr>
<td>10</td>
<td>Medical or Dental</td>
</tr>
<tr>
<td>11</td>
<td>Return Home</td>
</tr>
<tr>
<td>12</td>
<td>Religious Activities</td>
</tr>
<tr>
<td>13</td>
<td>Volunteer Work</td>
</tr>
<tr>
<td>14</td>
<td>Community Meetings</td>
</tr>
<tr>
<td>15</td>
<td>Other</td>
</tr>
<tr>
<td>16</td>
<td>To Day Care or Preschool</td>
</tr>
<tr>
<td>17</td>
<td>Go Along For The Ride</td>
</tr>
<tr>
<td>18</td>
<td>Work or School</td>
</tr>
</tbody>
</table>

Table 2 gives the multivariate F’s from the SPSS MANOVA run on the problem. Using 0.05 as the criterion for rejection, the significance of F indicates we should reject the null hypothesis and conclude that people’s activities differs overall on the set of 13 trip variables.

Table 2. Multivariate F’s from SPSS MANOVA

<table>
<thead>
<tr>
<th>Test Name</th>
<th>Value</th>
<th>Approx. F</th>
<th>Hypoth. DF</th>
<th>Error DF</th>
<th>Sig. of F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pillais</td>
<td>.15292</td>
<td>1.58530</td>
<td>78.00</td>
<td>4728.00</td>
<td>.001</td>
</tr>
<tr>
<td>Hotellings</td>
<td>.16251</td>
<td>1.62784</td>
<td>78.00</td>
<td>4688.00</td>
<td>.000</td>
</tr>
<tr>
<td>Wilks</td>
<td>.85421</td>
<td>1.60697</td>
<td>78.00</td>
<td>4323.27</td>
<td>.001</td>
</tr>
<tr>
<td>Roys</td>
<td>.08744</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Next, we use Discriminant Analysis for describing major differences among the seven day-of-week groups in MANOVA. As we have k=7 groups and p=13 dependent variables, then the number of possible discriminant functions is the minimum of p and (k-1), which is 6. After the test procedure is performed for determining how many of the discriminant functions, only the first discriminant function remains. The coefficients for each of the trip variables of the six discriminant functions are listed as follows:
### Table 3: Standardized Canonical Discriminant Function Coefficients

<table>
<thead>
<tr>
<th>Activity</th>
<th>Function</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pick-up passenger</td>
<td>PURP1</td>
<td>.114</td>
<td>.389</td>
<td>-.288</td>
<td>.049</td>
<td>-.227</td>
<td>-.392</td>
</tr>
<tr>
<td>Medical or Dental</td>
<td>PURP10</td>
<td>.307</td>
<td>.060</td>
<td>.083</td>
<td>.264</td>
<td>.466</td>
<td>.562</td>
</tr>
<tr>
<td>Return Home</td>
<td>PURP11</td>
<td>.007</td>
<td>-.095</td>
<td>-.485</td>
<td>-.816</td>
<td>-.391</td>
<td>.652</td>
</tr>
<tr>
<td>Religious Activities</td>
<td>PURP12</td>
<td>-.530</td>
<td>.699</td>
<td>.150</td>
<td>.164</td>
<td>.476</td>
<td>-.051</td>
</tr>
<tr>
<td>Other</td>
<td>PURP15</td>
<td>.140</td>
<td>-.249</td>
<td>-.187</td>
<td>.113</td>
<td>-.169</td>
<td>.274</td>
</tr>
<tr>
<td>Drop off Passenger</td>
<td>PURP2</td>
<td>.130</td>
<td>.317</td>
<td>.336</td>
<td>-.086</td>
<td>.052</td>
<td>.161</td>
</tr>
<tr>
<td>Workplace</td>
<td>PURP3</td>
<td>.595</td>
<td>-.098</td>
<td>.444</td>
<td>-.228</td>
<td>.325</td>
<td>-.348</td>
</tr>
<tr>
<td>Work-related Business</td>
<td>PURP4</td>
<td>.283</td>
<td>.415</td>
<td>-.108</td>
<td>.655</td>
<td>-.375</td>
<td>.033</td>
</tr>
<tr>
<td>School/College</td>
<td>PURP5</td>
<td>.259</td>
<td>.224</td>
<td>.037</td>
<td>.053</td>
<td>.003</td>
<td>-.070</td>
</tr>
<tr>
<td>Shopping</td>
<td>PURP6</td>
<td>-.057</td>
<td>.080</td>
<td>-.013</td>
<td>.177</td>
<td>.231</td>
<td>-.237</td>
</tr>
<tr>
<td>Other Errands</td>
<td>PURP7</td>
<td>.114</td>
<td>-.372</td>
<td>-.520</td>
<td>.268</td>
<td>.544</td>
<td>-.293</td>
</tr>
<tr>
<td>Eat Out</td>
<td>PURP8</td>
<td>.168</td>
<td>-.096</td>
<td>.313</td>
<td>.119</td>
<td>-.058</td>
<td>.032</td>
</tr>
<tr>
<td>Social/Recreational</td>
<td>PURP9</td>
<td>-.072</td>
<td>-.230</td>
<td>.459</td>
<td>.304</td>
<td>-.166</td>
<td>.130</td>
</tr>
</tbody>
</table>

Another useful table in discriminant analysis of MANOVA is the Structure Matrix, which shows the correlation between each discriminate function and each of the original variable (in this case, trip with specific purpose, e.g. PURP1).

### Table 4: Structure Matrix

<table>
<thead>
<tr>
<th>Activity</th>
<th>Function</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Workplace</td>
<td>PURP3</td>
<td>.635(*)</td>
<td>.515</td>
<td>.318</td>
<td>-.371</td>
<td>.184</td>
<td>-.214</td>
</tr>
<tr>
<td>School/College</td>
<td>PURP5</td>
<td>.367(*)</td>
<td>.246</td>
<td>-.128</td>
<td>-.117</td>
<td>.024</td>
<td>.018</td>
</tr>
<tr>
<td>Religious Activities</td>
<td>PURP12</td>
<td>-.343</td>
<td>.614(*)</td>
<td>.054</td>
<td>-.084</td>
<td>.360</td>
<td>.197</td>
</tr>
<tr>
<td>Pick-up passenger</td>
<td>PURP1</td>
<td>.320</td>
<td>.374(*)</td>
<td>-.254</td>
<td>-.103</td>
<td>-.165</td>
<td>-.366</td>
</tr>
<tr>
<td>Drop off Passenger</td>
<td>PURP2</td>
<td>.249</td>
<td>.262(*)</td>
<td>.121</td>
<td>-.175</td>
<td>.110</td>
<td>.113</td>
</tr>
<tr>
<td>Other Errands</td>
<td>PURP7</td>
<td>.221</td>
<td>-.208</td>
<td>-.531(*)</td>
<td>.148</td>
<td>.449</td>
<td>-.130</td>
</tr>
<tr>
<td>Social/Recreational</td>
<td>PURP9</td>
<td>-.062</td>
<td>-.217</td>
<td>.374(*)</td>
<td>.216</td>
<td>-.188</td>
<td>.154</td>
</tr>
<tr>
<td>Eat Out</td>
<td>PURP8</td>
<td>.207</td>
<td>-.058</td>
<td>.251(*)</td>
<td>.064</td>
<td>.014</td>
<td>.071</td>
</tr>
<tr>
<td>Work-related Business</td>
<td>PURP4</td>
<td>.344</td>
<td>.389</td>
<td>-.168</td>
<td>.518(*)</td>
<td>-.377</td>
<td>.110</td>
</tr>
<tr>
<td>Return Home</td>
<td>PURP11</td>
<td>.247</td>
<td>.252</td>
<td>-.323</td>
<td>-.465(*)</td>
<td>-.010</td>
<td>.413</td>
</tr>
<tr>
<td>Shopping</td>
<td>PURP6</td>
<td>.010</td>
<td>.010</td>
<td>-.205</td>
<td>.094</td>
<td>.309(*)</td>
<td>-.042</td>
</tr>
<tr>
<td>Medical or Dental</td>
<td>PURP10</td>
<td>.308</td>
<td>.055</td>
<td>-.007</td>
<td>.247</td>
<td>.478</td>
<td>.573(*)</td>
</tr>
<tr>
<td>Other</td>
<td>PURP15</td>
<td>.137</td>
<td>-.111</td>
<td>-.204</td>
<td>.108</td>
<td>-.069</td>
<td>.402(*)</td>
</tr>
</tbody>
</table>

There are two methods for interpreting the discriminant functions:

1. Examine the standardized coefficients--- these are obtained by multiplying the raw coefficient for each variable by the standard deviation for that variable.
2. Examine the discriminant function-variable correlation (structure matrix).
Generally, it is assumed that greater stability of the correlation exists in small or-medium sized samples, especially when there are high or fairly high interrelations among the variables. Also the correlation gives a direct indication of which variables are most closely aligned with the unobserved trait which the canonical variate (discriminant function) represents. Usually, we use the correlation for substantive interpretation of the discriminant functions, but use the coefficients to determine which of the variables are redundant given that others are in the set.

When there are two or more discriminant functions, then a useful device for determining directional differences among the groups is to graph them in the discriminant plane. The horizontal direction corresponds to the first discriminant function and thus lateral separation among the groups indicates how much they have been distinguished on this function. The vertical dimension corresponds to the second discriminant function and thus vertical separation tells us which groups are being distinguished in a way unrelated to the way they were separated on the first discriminant function. (Figure 1 shows the positions of Groups for Lexington’s travel survey data in the discriminant plane defined by Function 1 & 2 and function 1 & 3).

For interpreting the first discriminant function, as mentioned earlier, we use both the standardized coefficients and the discriminant function-variable correlation. Examining Table 4 for the first discriminant function, we see that it is primarily the two variables--trip with PURP3 (work place) and trip with PURP5 (school, college, university) that defines the function, with PURP12 and PURP4 secondarily involved. Since the correlation for PURP12 is negative, this means that the groups that have higher PURP12 trip counts scored lower on the first discriminant.

Now, evaluating the standardized coefficients to determine which of the variables are redundant given others in the set, we see that purp12 and purp3 are not redundant (coefficients of -0.53 and 0.595 separately), but that purp1, purp6 and purp9 are redundant since their coefficients are close to zero. Combined with the information from the coefficients and discriminant function-variable correlation, we can say that the first discriminant function is characterized as work—school—religious activity dominant. The three kinds of trips maximized the difference of people’s
activities among the days of a week. And on the other hand, we know that return home, shopping and social recreation trips show not much variation within the period of one week. Note, from the group centroids, it is on weekend people that live in Lexington area go to religious activities and don’t go to work place or study place (school, college and university), which is obvious. We may also infer from the relative large positive coefficient of purp10 that medical or dental trips are typically made during weekdays but not on weekend.

Post Hoc Procedures

Hotelling T2’s and Univariate t tests

After a significant overall multivariate result one would like to know which groups and which variables were responsible for that association, i.e., a more detailed breakdown. Pairwise multivariate tests (T2’s) and univariate t test are used to determine which pair of groups differs significantly on the set of trip variables. Here, we use a loose criterion, say, each at the .05 level for univariate t tests, to determine which kind of trips are contributing the significant pairwise difference of people’s travel pattern on two different days of the week. Some interesting results are showed as follows:

Pairwise activity comparison based on groups:

Thursday - Friday Not Significant
purp1 pickup passenger Significant

Tuesday - Thursday Not Significant

Monday - Thursday Not Significant
purp4 work-related business Significant

Thursday - Saturday Significant
purp1 pickup passenger Significant
purp3 work place Significant
purp4 work-related business Significant
purp5 school, college, university Significant
purp11 return home Significant

Saturday - Sunday Significant
purp7 other errands Significant
purpl2 religious activities. Significant

Wednesday - Sunday Significant
purp3 work place Significant
purp10 medical or dental Significant
purpl2 religious activities Significant
Tuesday - Saturday Significant
purp3 work place. Significant

Monday - Tuesday Not Significant

Wednesday - Saturday Not Significant
purp3 work place Significant
purp10 medical or dental Significant

Friday - Saturday Significant
purp3 work place Significant
purp7 other errands Significant
purpl2 religious activity Significant

The major difference between activities on weekdays and weekends is whether to make work-trips. Work trips usually go from home to a specific place then back on a regular basis, and usually they occur on five days per week.

It is worth noticing that people’s activity-choice on Saturday and Sunday also show contrasting difference. Saturday time typically is devoted to errands compared to that of Sunday, while Sunday’s religious activity can be viewed in terms of routines. It is more or less performed on a weekly basis. Thus it is easy for us to understand why the two points that represents Saturday and Sunday on the discriminant plane are separated widely.

Ongoing Research

Following this case study, we found that some of the household data in the Lexington study was incomplete and thus we have to reduce the household set from 100 to about 80. This is still a significant number. There also appears to be some problems with the digitized basemap; a preliminary search through the CD-ROM obtained from DOT shows that only one of the two counties was thoroughly geocoded. Thus, we will have to obtain an equivalent mapping of the second county so that members of the activities of sample residing in that area can be fully taken into consideration. It appears that we may also have to obtain land-use maps and geocode the locations of specific urban functions such as shopping centers, educational institutions, parks and recreational areas, churches, and so on. This preliminary session on evaluating the cleanliness of the dataset and transforming it to one that we can use will be followed by a MANOVA based examination of the activity patterns, thus determining the nature of activity patterns for specific purposes on different days of the week. A second project will be to evaluate standard network models that define optimal routes for origin destination pairs and trip chains and then compare them with the trips actually made by the sample members. This will be followed by an evaluation (using Quadratic Assignment Procedures) of which of the models most closely reproduce the actual dataset. A third project is to examine the daily and weekly temporal sequencing of trips to define repetitive cycles. Two-dimensional (Spatial) Spectral Analysis will be used. A fourth project will be to examine directionality in trip making using circular statistics. We will examine the directional relation between different trip types, as well as looking for different trip patterns for a given type on different weekdays.
Ultimately, we will focus on two areas; (1) adding a final touch to the considerable volume of data on destination choice activity behavior as reconstructed in the past from whole or partial week diaries; (2) evaluating which of the set of prevailing optimizing network flow models most closely represent the actual behavior of sample members. This will provide us some hard information on the degree to which travelers are spatially rational or whether other criteria appear to be more important in choosing routes for single and multi-purpose trips.

References


Fohi, P., Curtin, K.M., Goodchild, M.F., and Church, R.L. (1996) A non-planar, lane-based navigable data model for ITS. In M.J. Kraak and M. Molenaar (Eds.), *Proceedings, Seventh International Symposium on Spatial Data Handling* (pp. 7B.1-7B.29). Delft, August 12-16


Institute of Transportation Engineers (1989) A toolbox for alleviating traffic congestion. Institute of Transportation Engineers, Washington D.C.


Road Transportation Research (1988) Route guidance and in-car communication systems. Road Transport Research, Organization for Economic Coordination and Development, Paris, France.


