THE JOURNEY TO WORK
A SINGULAR BASIS FOR TRAVEL PATTERN SURVEYS

SEPTEMBER 1967

NO. 26

by

G. A. SHUNK
W.L. GRECCO
V.L. ANDERSON

Joint Highway Research Project

PURDUE UNIVERSITY
LAFAYETTE INDIANA

Technical Paper

THE JOURNEY TO WORK;
A SINGULAR BASIS FOR TRAVEL PATTERN SURVEYS

To: G. A. Leonards, Director
Joint Highway Research Project

From: H. L. Michael, Associate Director
Joint Highway Research Project

September 27, 1967

File No.: 3-7-2

Project No.: C-36-69B

Attached is a Technical Paper entitled "The Journey to Work; A Singular Basis for Travel Pattern Surveys".

It has been authored by Messrs. G. R. Shunk, W. L. Grecco, and V. L. Anderson of our staff. The paper is from research currently being conducted under the HPR, Part I, project "Major Aspects of the Urban Transportation Planning Process". The material contained in this paper has not previously been reported to the Board but will also be contained in a Progress Report on this research to be presented at an early date.

The paper has been offered to the Highway Research Board for presentation at its 1968 Annual Meeting. It is presented to the Board for approval of publication if it is accepted by the HRB for approval. It will also be forwarded to the Highway Commission and the BPR for their review and approval of publication.

Respectfully submitted,

Harold L. Michael

Harold L. Michael
Associate Director

HLM:mf

Attachment

Copy: F. L. Ashbaucher
W. L. Dolch
W. H. Goetz
W. L. Grecco
G. K. Hallock
M. E. Harr
R. H. Harrell
J. A. Havers
V. E. Harvey
J. F. McLaughlin
F. B. Mendenhall
R. D. Miles
J. C. Oppenlander
C. F. Scholer
M. B. Scott
W. T. Spencer
H. R. J. Walsh
K. B. Woods
E. J. Yoder

Technical Paper

THE JOURNEY TO WORK:

A SINGULAR BASIS FOR TRAVEL PATTERN SURVEYS

by

G. A. Shunk, Graduate Instructor

W. L. Grecco, Research Engineer

and

V. L. Anderson, Professor of Statistics

Joint Highway Research Project
File No.: 3-7-2
Project No.: C-36-69B

Prepared as part of an Investigation

Conducted by

Joint Highway Research Project
Engineering Experiment Station
Purdue University

in cooperation with the

Indiana State Highway Commission

and the

U. S. Department of Transportation
Federal Highway Administration
Bureau of Public Roads

Lafayette, Indiana

September 27, 1967

THE JOURNEY TO WORK; A SINGULAR BASIS FOR TRAVEL PATTERN SURVEYS

INFORMATIVE ABSTRACT

This research involved a study of the feasibility of using the patterns of work trips alone to represent the patterns of travel for all purposes in an urban area. Further, the feasibility of using peak hour travel patterns to represent those of the entire day was investigated. The objective was to develop an approach to travel surveys which would satisfactorily reproduce the results of a conventional home interview survey.

Using data from a 1964 comprehensive transportation survey in Indianapolis, Indiana, an analysis of variance was run to determine the effect of the commonly defined factors, mode, purpose, and time, on trip volume and average trip length. Based on the extremely high significance of all main effects and interactions, a second variance analysis was run to determine the effect of more specific purpose, time and mode factors on the traffic assigned to the freeway and arterial links of the highway system. The significance of all main effects and a mode-purpose interaction were the basis for regression models accounting for mode, the peak hour, and the work purpose. A high degree of the variation in total trips on all major street system links was explained by multiple linear regression equations based on link volumes for the work purpose. Using the same regression approach, high degrees of explanation were achieved for total day, all purpose trips using all peak hour trips, and for all peak hour trips using total day work trips.

THE JOURNEY TO WORK; A SINGULAR BASIS FOR TRAVEL PATTERN SURVEYS

INTRODUCTION

The important role that the journey to work assumes in urban area travel is readily acknowledged by transportation planners. The significance of peak hour work travel would be corroborated by any commuter who has ever braved the "rush hour". But the possibilities of such significance have apparently escaped exploration. It is conceivable that, due to this great influence, it might be possible to employ a survey of work trips alone as a basis for developing the pattern of all travel in an urban area. It is also possible that the pattern of peak hour work travel is adequate to develop a highway system capable of serving trips of all purposes for the entire day. The study of these two proposals is the subject of the research reported here.

The work trip is the most stable and ritualistic component of urban travel. It occurs between two of the most readily predictable land uses. It is less influenced by separation than almost any other purpose trip. Work trip generation rates are characteristically stable and predictable. In general, no other function concentrates more people per net unit of area than the work process. Because of this property, there is a related concentration of traffic at the workplace. Traffic intensity on facilities serving centralized work places or very large single workplaces can extend for considerable distances.

The important role that work oriented travel assumes in the total picture of urban area transportation should be apparent from the above discussion. Travel in nearly every city is dominated by the journey to work. Only in cities whose function is somehow extraordinary can instances be seen where work travel is less than the principal purpose of travel. The characteristics of work travel as a sufficient representative of the total urban travel pattern are the essential elements of predictability. Based on these attributes, it is proposed that work travel would be valid and sufficient means for prediction of total urban travel patterns. It is hypothesized that, for use in the urban transportation planning process, the pattern of travel developed using work oriented trips alone presents a sufficient representation of the major street system used by all urban travel. It is further proposed that this implication is nearly as valid for home-based work trips as for all work trips.

The feasibility of representing the distribution pattern of trips of all purposes by the distribution pattern of work purpose trips alone will be tested. These travel patterns will be examined over the basic system of major streets and highways. Zone-to-zone travel patterns cannot be used as the basis for testing, however, since they are strictly dominated by the functions of respective zones. Since the desired results of the prediction procedure concern the major street system, the links of the network representing this system will be used as the basis for comparison. Travel of the respective purpose groups will be assigned to the network, and a test of the reproducability will be made on a link-by-link basis. In this manner the zone centroid influence is virtually eliminated in favor of testing the conditions on the major street network.

It has been implied that the peak hour is an important factor in urban transportation, and work travel is a dominant aspect of the peak hour. It is then apparent that system design based on peak hour volumes may be feasible. The principal advantage of this approach is that the result is a functional basis for design. The methods by which ADT volumes are obtained in current planning practice contain considerable inherent error, attributable to natural variation and the sampling procedure. Further compounding this by introducing a factor to obtain a design hour volume seems unreasonable. By the time factoring has been accomplished, the care of a very expensive survey might well have been wasted.

The peak hour is the one most consistent and significant point of stress of the transportation system. It would seem only too obvious to deal directly with the maximum loading condition rather than to factor to it. The most significant argument raised against peak hour oriented design is that peaks occur at different times, at different locations, and for different purposes. Such a situation can only be examined empirically. To follow this argument, assume a system based on a single peak hour, say for work. Beyond this system heavy non work volumes, whenever their peak, could be accounted for by specialized surveys or analyses. The residential end is served by the work oriented system. Regarding the time shifts of peak volumes, as long as the maximum volume has been defined relative to the system, the location of its occurrence is immaterial. The trips involved are still full trips, from origin to destination, and will, sometime during the peak hour, contribute to the volume on the link. This situation, of course, is contingent upon a peaking definition such as used here, i.e. maximum trips on the entire system.

The hypothesis proposed for test in this second instance involves the reproducability of total day loadings by peak hour loadings. The reproducability of peak hour loadings using a survey of work trips alone might also be feasible. These will be tested using link-by-link comparison of the respective conditions.

The hypotheses proposed are directed at elimination of the home-interview survey technique and replacement of it by a special survey of work purpose trips only. Such a survey could take any of several forms. One of these, useful principally for an updating function, involves current proposals to obtain workplace information of employed persons as a portion of the decenial census. The possibilities of such an application may be further examined after evaluation of research such as that proposed here.

One of the main reasons that interviewing has traditionally been conducted at the dwelling unit is the requirement to obtain trips of all purposes by all members of the household. Quite naturally, if the interest in trip making were confined to a single purpose of travel, the interview place should, if possible, be oriented to the destination of trips for that purpose. This is the concept of the destination place interview.

To temporarily diverge somewhat, the concept of a destination place interview can be seen on reflection to be of quite apparent utility. The destination place of any trip purpose group is in every case at least as densely attractive as the residential end of the trip. Only social trips (i.e. human interaction) can be considered to generate travel at a rate per net unit of area as low as the home. In all other situations the very concept of activity or service provision on a production line basis implies high density trip confluence.

What traveler has not at some time or another been caught in a traffic jam in the central business area or near an industrial plant at shift change or near the site of a major sporting event? At all these locations a service or activity is offered which attracts persons from all residential locations to one relatively small area. Such activity can exist only if major interest can be focused upon it and transportation service provided for it. Very rarely if ever do traffic problems occur at the location of residence; probably the only case immediately apparent concerns high density residential areas, which of course fall into a destination place type situation. The object of highway planning should be provision of service at location of maximum stress, i.e. the points of traffic concentration. But for the exigencies of predictability and multipurpose considerations, the household interview approach is somewhat less attractive than the destination place survey. Destination place surveys could also, depending on the procedure, obtain measures of non-home trips. If the basic system indicated by these trips is developed, it will function for other travel as well, providing the exceptions are taken into account.

The applicability of the destination-place interview concept to work oriented travel points to surveys at the place of employment. Depending on whether or not non-home oriented trips are necessary, such a survey would take one of two forms. A survey of home-based travel would be confined to examination of employee records, coding the address of each worker's residence. Such a procedure eliminates all response error due to interviewing. Should it be desirable to obtain information on non-home trips, an alternate procedure would be to utilize either a questionnaire or an interview procedure. Interviewing would of course yield more and better information, since direct contact is made, but the balance in difficulty and cost would have to be weighed against utility.

In summary, revised survey procedures will be faster, easier, less expensive, more complete, and more reliable than the current household interview approach. Only their feasibility remains to be verified.

DATA PREPARATION

The data used in this study were obtained from the Indianapolis Regional Transportation and Development Study (IRTADS) through one of its sponsors, the Indiana State Highway Commission. Use of 1964 data from a study area the size of Indianapolis should imply the generality of any results to any city of similar character. The specific information selected from the IRTADS data file was the travel data from the home interview survey, as coded and punched on 'number 2' cards. No use was made of either the truck-taxi or the external survey data since the principal objective of this work was the elimination of the home interview survey. The inventory of the 1964 street network as punched in standard (BELMN) format was also obtained. Transit network information was not used because the proposed analysis was principally highway oriented. Transit trips were included in one phase of the analysis only because no network information was required.

The complete file of home-interview travel data was pre-processed to put it in a suitable form for subsequent manipulation. All trips with an origin or destination outside the study area were eliminated because of the lack of information on the external terminus and the difference in motivation and character between those and the wholly internal trips. This deletion eliminated 2416 trips from the basic file. The remaining trip cards were grouped according to home orientation. Home-based trips were placed in one file and sorted on zone of residence. Non-home based trips were placed in a second file and sorted on zone of origin. These files were the input for the travel data processing programs for the total day condition.

In order to select a representative peak period the combined home- and non-home-based files, with externals deleted, were processed by the program PEAKS.[1] PEAKS scanned the standard trip survey cards for trip purpose, mode, and times of start and arrival. The times were recorded separately according to the mode and purpose of the trip. PEAKS then computed the number of trips entering and leaving the transportation system, by mode and purpose and in tenth of an hour increments throughout the day. PEAKS then aggregated the incremental periods into successive one hour blocks and produced the number of trips in progress, by mode and purpose, for contiguous one hour periods, successively by tenth of an hour increments through the day. The hours during which trips in progress by each mode and purpose reached their maximum were then available. PEAKS also aggregated modes and purposes into total tables. The peak hour selected for use in the current analysis was that for all auto driver trips. This was the condition considered to place the most stress on the highway system. A single hour was selected because the objective in the peak hour phase of this study was to obtain a single hour volume on which to base design recommendations. The peak hour selected for subsequent study was from 4:24 to 5:24 p.m.

Processing of the travel and network data involved extensive use of the computer program package disseminated by the Bureau of Public Roads for use in operational transportation planning studies. The IRTADS data had originally been coded according to formats required by the system, since IRTADS also used these programs. Several decisions regarding format of the data for the final analysis had to be made prior to the initiation of bulk processing. It was decided that trips of three specific purpose groups would be obtained in addition to the all-purpose group. The selected purposes were home-based work, home-based shop, and non-home-based work. Non-home-based trips could reasonably be classified by either the to or from purpose, since neither is at the place of residence. For the current study a non-home-based work trip is one having a work purpose at either end of the trip.

The modes selected for analysis were auto driver and highway person, reflecting the highway orientation. The former represented vehicle trips and could be indicative of traffic volume. Highway person trips included auto driver trips as well as passengers in automobiles, trucks, and taxis. Use of a person trip orientation in revised survey procedures would of course permit development of modal split relations. No transit or school bus trips were included because of the lack of knowledge of a network for either group. Two periods, total day and the single afternoon peak hour, defined the time conditions to be considered.

Further processing of the basic data in preparation for the analysis used the BELMN program package. The shorter of two phases using BELMN involved processing of the street network. The 1964 street inventory had been coded and punched on cards by IRTADS. Program PR-6 processed the cards containing among other information the length of, speed on, and node number at the terminals of every link in the street system. Program PR-1 scanned the network description for link direction and travel time and the terminal node numbers and proceeded to build "trees" from each zone to all other zones in the study area. These trees are the link-by-link description of the path taken in moving from one zone to another. For the current study the minimum travel-time path was chosen. Travel-time includes terminal time at both ends of the trip. PR-1 prepared a binary tape describing the minimum time path trees, link-by-link, for all zones in the study area. Program PR-130 further processed the binary tree information by summing the time to traverse the links in each tree, producing the accumulated time to move between each zone pair on the minimum time path, known as a "skimmed tree."

The major utilization of the BELMN programs was concerned with processing the travel data. The necessary pre-processing of the trip cards was described previously. Input to program PR-133 was in two phases, home-based and non-home-based trip cards. This separation was necessary because of the requirements of PR-133. The procedures were essentially identical, and the results were combined at a later stage. Output were "trip tables," cumulative zone-to-zone movements by purpose and mode. Because of the nature of PR-133, all final trips tables are complementary; i.e. they do not overlap. Program PR-152 was used to merge certain tables in an additive manner in order to obtain the purpose combinations specified for analysis.

The entire trip processing and assignment procedure was executed for both the total day and the peak hour situations. The traffic assignment process, utilizing program PR-2 assigned to each link of the minimum time path tree the zone-to-zone movements given in the trip table. This process accumulated the trip volumes on each link for all zone-to-zone movements. No attempt was made to apply capacity restraint to the loaded networks, since the differences in absolute volumes would have yielded inconsistent results from such a procedure. The objective of the research was to match the control loading condition; that having been accomplished, restraint procedures would be applied to the synthesized loadings. There were sixteen separate network loadings, made up by the 12 specified purpose situations and 4 totals. The loaded networks were summarized by program PR-124 to obtain a more readily processable output format and a listing of the loading on each link. The output from PR-124 was processed by a data reduction program which summarized for each link the loading under each of the sixteen specified conditions.

Since the principal influence of the hypothesis was intended to be over the major street system, it was necessary to select the links in this group. In order to provide an objective basis for the selection, the functional street classifications developed by IRTADS were used to group the links. The IRTADS system was composed of five groups: local, collector, arterial, expressway, and freeway. Because of the small number in their groups, expressway and freeway links were combined under the latter title. All local links, centroid and external node connectors, were removed because of domination of travel on them by the zone represented by the centroid. Links connecting between different street classes were considered collectors. The link groups at this point represent a modification of the IRTADS system and are referred to as the MOD-1 system.

Examination of the volume distributions in the respective groups indicated certain exigencies for which modifications were appropriate. The link volumes for the auto driver, total day, all purpose condition were chosen as a criterion since they were the best available representation of actual traffic volumes. Based on the estimated standard error of the group, all links with volumes less than 140 were deleted.

This was because the true volumes on these links in the average situation might be reasonably considered not different from zero. Links that were the only connection between the system and local links were deleted for the same reason as the local links. Links previously classified arterial or freeway, but having volumes less than 1000 were merged with collectors. Collectors with volumes greater than 5000 but less than 12,000 were merged with arterials; those with volumes over 12,000 were merged with freeways.

The rationale for these modifications was based on the fact that no capacity restraint was used in the assignment process. As a result trips were assigned to the absolute minimum time path without consideration for the capacity of the links used. Such a situation would explain the failure of links to carry volumes commensurate with their functional classification. In order to correct for the situation, links having arterial level volumes were defined as arterials etc.; this was the reasoning behind volume considerations when reorganizing the groups. The volume criteria for each group were established by a generalization capacity analysis of the respective street classes.

The MOD-2 system used in the final analyses was as follows:

Freeways	218 links
Arterials	529 links
Collectors	1793 links

ANALYSES

The analytical procedures employed to test the proposed hypotheses fell into two distinct phases. The first was directed toward establishing a basis for consideration of the second. The second phase was directly concerned with testing the principal hypotheses regarding work and peak hour travel.

Phase one undertook examination of the hypothesis that the several factors of trip purpose, means of travel, and time of trip do significantly influence the character of person movement in an urban area. The variables chosen for examination were travel volume and length of trip. Travel volume was defined as number of trips made, where each survey card represents a trip. Trip length was the time required to complete a given trip on the minimum time path from the zone of origin to the destination zone. The purpose of a trip was that indicated on the survey card at the point of destination. Purpose was considered in six groups:

1. Work
2. Shopping
3. Social-Recreational/Eat Meal
4. Personal Business/Medical-Dental
5. School
6. Other

Mode of travel was defined in three groups:

1. Auto Driver
2. Non-transit passenger
3. Transit passenger

The definition of transit includes school busses as well as other bus vehicles; there is no other form of transit in Indianapolis. The non-transit passenger group includes passengers in private automobiles, taxis, and trucks. Time was defined in 24 one hour groups. The mean of the start and arrive times reported for the trip maker was used to place the trip in its time group.

The basic data source was the IRTADS home inverview survey file consisting of 76,396 records, each describing one trip wholly within the area made by a resident of a household selected for interview. The sampling unit for the survey was the household. The household[9] had been selected in a systematic manner from public utility records and represented approximately five percent of the dwelling units in the study area. The appropriate skimmed tree time was appended to the individual record of each trip by the program LENGTH,[1] being made an additional permanent part of each trip record.

The nature of the hypothesis to be tested was appropriate for investigation by the analysis of variance technique (ANOVA). This statistical procedure involves classification of each observation of a variable according to the conditions of several factors, the object of the investigation being to determine the extent to which the factors affect the observed variable. Thus an observation occurring under a particular set of conditions would be grouped only with observations which occurred under similar circumstances. In the type of study undertaken here, termed a complete factorial, there are the same number of such groups, or cells, as there are combinations of possible conditions (levels) of the factors considered. This procedure isolates the quantitative effect that each factor has on the variable analyzed, but also permits evaluation of effects occurring due to factors acting in combination (interaction).

In order to test the significance of the effects due to factors and interactions, the ANOVA uses an estimate of experimental error, i.e.

natural variability (not due to the factors analyzed) to be expected in the occurrence of the variable. One means of obtaining such an estimate in experimentation is to replicate or repeat at least a portion of the experiment, since variability in observations made under identical conditions can be attributed to experimental error. For the present investigation it was decided to select four random subsamples from the basic trip file. These four complete subsamples provided the necessary estimate of experimental error. In order to simplify the sample selection procedure, the observation selected for testing was the mean trip length value over all trips in each cell for each subsample.

The equation representing the analyses, commonly called the analysis of variance model was:

$$X_{ijkl} = \mu + P_i + M_j + T_k + PM_{ij} + PT_{ik} + MT_{jk} + PMT_{ijk} + \epsilon_{(ijk)l}$$

where: X_{ijkl} represents trip volume or trip length, depending on the analysis, for the i^{th} purpose, by the j^{th} mode, in time period k, for the l^{th} subsample;

μ is the respective overall mean;

P_i is the effect of the i^{th} purpose, i = 1....6;

M_j is the effect of the j^{th} mode, j = 1.......3;

T_k is the effect of the k^{th} time period (hour), k = 1.....24;

PM_{ij} is the effect of the purpose-mode interaction;

MT_{jk} is the effect of the time-mode interaction;

PT_{ik} is the effect of the purpose-time interaction;

PMT_{ijk} is the effect of the purpose-mode-time interaction;

$\epsilon_{(ijk)l}$ is the experimental error;

l is the number of the subsample l = 1.....4.

It will be noted that all effects are fixed, i.e. they are not random samples from an infinite population of such values. The inference permitted can, therefore, only be considered applicable for those levels of the respective factors included in this analysis.

A theoretical consideration at this point involved the inference space of the results. The original objective was to imply validity not only for the city of Indianapolis, but for the nation as a whole. Such an implication is valid if the trip data used is considered a randomly selected single cluster sample from a nationwide population of trips.

The random subsamples of the basic systematic sample may be considered random samples of trips in Indianapolis, When four subsamples of 10,000 each were drawn from the original 76,396 trips that represented a five percent sample, each subsample was effectively a sample of less than one in one hundred and fifty and was considered drawn from an infinite population. Under these circumstances infinite theory was closely approximated, and no finite population correction was necessary.

Selection and processing of the four samples was accomplished by the program SAMPLR[(1)]. Input to SAMPLR was four sets of unique, sorted random numbers developed by the program RANDOM[(1)], and the sorted trip card file, augmented with trip lengths. The four files of random numbers were stacked on magnetic tape for use by SAMPLR.

SAMPLR read the random numbers and, based on each, selected the data occuring in the designated location of the trip card file. The records selected were tabulated by purpose, mode, and time of trip. Each record used was deleted from the input trip file, and those remaining were written out to await selection of the next sample by the subsequent pass of SAMPLR. The sample selection process was repeated four times. These data were punched on cards in preparation for their analysis.

The complete factorial analysis of variance computations were executed by program BIMD-2V[2]. The analyses are summarized below.

Volume - all trips

factor	d.f.	SS.	MS.	F
1	2	390,696.9	195,348.4	*
2	23	492,524.0	21,414.1	*
3	5	467,009.9	93,402.0	*
12	46	227,956.2	4,955.6	*
13	10	585,531.9	58,553.2	*
23	115	1,057,252.5	9,193.5	*
123	230	752,835.5	3,273.2	*
ε	1296	28,157.0	21.7	---
Total	1727	4,001,963.9	---	---

* F ratios are not shown because of the obvious significance of every factor and interaction.

(1 = Mode; 2 = Time; 3 = Purpose)

Average length - all trips

factor	d.f.	SS.	MS.	F
1	2	4388.9	2194.4	*
2	23	15708.4	683.0	*
3	5	10766.2	2153.2	*
12	46	5261.0	114.4	*
13	10	1731.7	173.2	*
23	115	9045.9	78.7	*
123	230	10649.1	46.3	*
ε	1296	29177.6	22.5	--
Total	1727	86728.9	--	--

*F ratios are not shown because of the obvious significance of every factor and interaction.

All main effects and interactions are significant ($\alpha = 0.01$). The high significance of the main effects had been expected. It implied that the volume and length of trips observed in Indianapolis differed significantly with variation in the purpose of trip, the mode of travel, and the time of observation. The high significance of the interactions was not anticipated. It implied, for example, that the relationships between volume or trip length and the single factors (e.g. purpose) were inconsistent if any other factor was not held constant. The results of this analysis emphasized the fact that the factors being examined in regard to travel pattern development were very worthy of consideration. They also indicated that further analyses would have to account for the interactions.

Phase two of the analysis involved testing the principal hypotheses, concerned with the use of work and peak hour trips to represent total daily travel. The objective was to determine the degree to which trips of a single purpose or a particular time period could be expected to reproduce the pattern of all travel in an urban area and define the transportation system used thereby. Travel volume on individual links of the highway network was the decision variable selected; the form that the variable took depended on the analysis performed. Reflection on the objective of the research pointed up the necessity of retaining the all-purpose loading as the control condition, against which the hypothesized revisions would be tested. The nature of the situation, with the variable to be predicted containing the variable used to predict, indicated that a regression approach would be most appropriate.

The extent of the regression analysis required was investigated by a modification and extension of the analysis of variance performed in phase one. The objectives of this second ANOVA were to determine which factors should be included in the regression models and what different models were necessary. This analysis was designed to test the effect on individual link volumes of change in the factors purpose, mode, and time.

Definitions of the factors and variable for this analysis were modified from those applied in the first investigation. In this analysis purpose was considered at three levels: home-based work, non-home based work, and non-work. This reflected a split of the previous P_1 (work) and combination of P_2 to P_5. Mode was included at only two levels, transit trips having been deleted. The time levels were redefined as peak hour, one particular hour, against non peak hour, the remaining 23 hours combined. The observed variable was relative assigned traffic volume. This variable was obtained by assigning trips (variable in the first analysis) to the links of the highway network and dividing each resulting link volume by the link-trip total over all links for its particular factor level combination. This manipulation eliminated between cell differences attributable only to differences in absolute total volumes of trips observed for respective purposes. The effect of the absolute totals had been examined in the first analysis; the second analysis was to examine the degree to which selected observed effects extended to the highway system. The resulting variable, termed link-relative-importance (LRI), was indicative of the status the particular link assumed regarding movement of traffic in the area.

If there was no significant difference found due to purpose, it could be reasoned that each link was as important for moving work trips as for moving other trips. Lack of significance due to mode would imply that passenger travel is distributed on the system in the same manner as vehicle travel. And no significance attributed to time would infer that peak hour traffic uses the same links as non peak movement. Should any main effects not be considered significant (0.25), regression analysis of that situation would not be necessary. If the main effects were found significant it could be reasoned that sufficient difference occurred between purposes, modes, or time periods for these factors to be considered in the regression analysis.

Of particular interest in this ANOVA was whether the significance of interactions carried through from the first analysis. An interaction implies that the results of varying one factor under the constant level of another factor might not match the results of identical variations of the first factor under different conditions of the second factor.

Thus, a significantly different relationship might be found between volume and purpose for auto driver trips than for passenger trips. Interaction significance would imply a need for different regression models at each level combination of the interacting factors. This analysis would yield a rational basis for the form of the regression equations and contribute to the understanding of underlying relationships.

Consideration of two requirements of the analysis of variance was necessary. The ANOVA procedure bases its tests of significance on properties of the normal distribution and requires that the experimental error within the classification groups or cells be normally and independently distributed. Tests of this condition utilizing the Komolgorov-Smirnov (K-S) test for goodness of fit, indicated that the raw LRI values were not normally distributed. It was found, however, that after a square root transformation of the raw data was carried out, the K-S test showed significant (5% level) departures from normality in only a very few cases for the arterial and freeway classes of the MOD-2 highway system. The collector class was discarded from further consideration in the ANOVA examination because the observed departures from normality could not be considered insignificant. The use of the square root transformation has a basis in theory; the data were merely a traffic occurence or frequency distribution on the highway system, known to be distributed in a Poisson manner. The square root is the characteristic transformation to a normal distribution for a Poisson distributed variable.

The other ANOVA assumption tested concerned homogeneity of variance between cells. The common test for this condition is that attributed to Bartlett[3]. The square root transformed data were processed by two computer routines which yielded the chi-square values to be tested. It was apparent that cell variances of the design were quite non-homogeneous. Box (5) has considered the variance problem and indicates that the robust nature of the ANOVA is capable of withstanding quite a degree of hetero-scedasticity.

In spite of the lack of variance homogeneity and the minor variations from normality, it was decided to continue with the ANOVA as proposed. The analysis was run separately for the two street classes with no attempt being made to examine between class effects. This decision was based principally on the variation in the number of observations between the classes. The ANOVA models took the form:

$$L_{ijkl} = \mu + P_i + M_j + T_k + PM_{ij} + PT_{jk} + MT_{jk} + PMT_{ijk} + \epsilon_{(ijk)l}$$

where: L_{ijkl} is the link-relative-importance for the i^{th} purpose, by the j^{th} mode, in time period k, for the l^{th} observation.

P_i is the effect of the i^{th} purpose, $i = 1....3$;

M_j is the effect of the j^{th} mode, $j = 1....2$;

T_k is the effect of the k^{th} time period, $k = 1....2$;

$\epsilon_{(ijk)l}$ is the experimental error;

l represents the link considered, $l = i....218$, for freeways, $l = 1....529$, for arterials.

The interactions are similar to those defined for the first ANOVA. computations for this analysis were performed by program BIMD-2V(2).

The analyses are summarized below.

Analysis of Variance

Freeway LRI

Factor	D.F.	S.S.	M.S.	F
1	1	28.37	28.37	3.31*
2	1	17.67	17.67	2.06*
3	2	511.07	225.54	26.7*
12	1	2.31	2.31	0.27
13	2	3.23	1.61	0.19
23	2	0.41	0.20	0.23
123	2	15.90	7.95	0.93
6	2604	22318.74	8.57	----
Total	2615	22897.68	----	----

*Significant at $\alpha = 0.25$

Analysis of Variance

Arterial LRI

Factor	D.F.	S.S.	M.S.	F
1	1	39.28	39.28	4.95*
2	1	253.22	253.22	31.8*
3	2	1783.42	891.71	112.2*
12	1	3.84	3.84	0.48
13	2	8.01	4.00	0.50
23	2	67.98	33.99	4.38*
123	6336	50380.69	7.95	----
Total	6347	52554.68	----	----

*Significant at $\alpha = 0.25$

(1 = Mode; 2 = Time; 3 = Purpose)

The significance level chosen for testing the F ratios was 0.25. This choice was based on the fact that probability of "type II" or "β" error (accepting a false hypothesis) was of importance. Increasing the "α", or probability of type I error (rejecting a true hypothesis), to the level of 0.25 reduces the probability of β error. The low β error was considered necessary because the objective of the test was to determine which effects were not significant and could thereby be eliminated from consideration in model development. All significance tests were made using an F ratio with only the error mean square, since the model was composed completely of fixed effects.

The tests on the freeway links indicated significant results due to the main effects of time, purpose, and mode. No effect on freeway LRI was noted due to interaction. It can be concluded that LRI does vary between the peak and non-peak periods, due to change in consideration of the work or the non-work purpose, and due to travel mode. The implications are, that, for freeways, separate models describing peak and non-peak traffic would yield better results than a single model. Further, there is sufficient effect due to the work purpose and mode that models describing travel must include recognition of the factors. The extension of these results is valid and consistent only over the factors and levels considered here.

The tests in the ANOVA for arterials indicated the same effects observed for freeways as well as a significant mode-purpose interaction. This additional effect may reflect the change in orientation of traffic from movement to land service as street class decreases. The variability in the influence on volume exerted by work purposes can not be considered the same for all modes, and conversely, as was the case for freeways. This implies a need for more models to account for interaction.

The hypothesis for the ANOVA of relative volumes was rejected for the purpose, mode, and time main effects on freeway links and for these effects as well as the mode-purpose interaction on arterial links. The remaining effects could not be rejected at $\alpha = 0.25$. The meaning of these results must be tempered by the failure of the data to satisfy the criterion of homoscedasticity. Reflection on the trends observed gives cause for contemplation on the results which might have occurred had the collector class exhibited normality.

The models for the regression analysis were developed in accordance with the results of the variance analyses, and included factors representing purpose, mode, and time. The definitions of the variables and factors for regression were further modified from those used previously. The dependent regression variable (Y), in accordance with the control condition selected, was the number of trips for all purposes that were assigned to the individual links of the highway network. This represented a combination of the three purpose levels tested in the second ANOVA. The independent regression variables (X's) were similarly assigned volumes, but represented trips for specific purposes: home-based work, non-home-based work, and home-based shop. The first two were identical to classifications in the second ANOVA; the third was an additional factor included because of general interest and availability of the data. The shop level was not included separately in the purpose factor of the second ANOVA because the objective at that point was to define the effect of work relative to all other purposes combined. Levels of the mode factor were auto driver, identical to M_1 in both previous analyses, and highway person, a combination of the M_1 and M_2 levels of the second ANOVA. Time was treated in a similar manner: peak hour corresponded to T_1 and total day was the combined T_1 and T_2 levels. The definition of the regression factors closely approached the definitions of the original principal hypotheses. The only variation occurred in the second level of purpose; P_2 was defined as non-home based rather than all work because of build-up approach. Adding non-home-based work to an equation including home-based work yielded the desired effect of total work. The factor definitions for the second ANOVA and the regression analysis are listed below with the corresponding nomenclature for the principal hypotheses.

	HYPOTHESES	ANOVA	REGRESSION
PURPOSE:	HBW	HBW	HBW
	All Work	NHBW	NHBW
	All Purposes	NW	All Purposes
MODE:	Driver	Driver	Driver
	Person	Passenger	Person
TIME:	Peak Hour	Peak Hour	Peak Hour
	Total Day	Non-Peak Hour	Total Day

(H:HOME, B:BASED, W:WORK, N:NON)

The factor level combinations for the regression variables are listed below.

INDEPENDENT VARIABLES

1. Home-based work, Auto driver, Total day
2. Non-home-based work, Auto driver, Total day
3. Home-based work, Highway person, Total day
4. Non-home-based work, Highway person, Total day
5. Home-based work, Auto driver, Peak hour
6. Non-home-based work, Auto driver, Peak hour
7. Home-based work, Highway person, Peak hour
8. Non-home-based work, Highway person, Peak hour

9. Home-based shop, Auto driver, Total day
10. Home-based shop, Highway person, Total day
11. Home-based shop, Auto driver, Peak hour
12. Home-based shop, Highway person, Peak hour

DEPENDENT VARIABLES

1. All purpose trips, Auto driver, Total day
2. All purpose trips, Highway person, Total day
3. All purpose trips, Auto driver, Peak hour
4. All purpose trips, Highway person, Peak hour

Separate models were developed for each combination of the levels of the mode and time factors. The ANOVA results indicated that the use of each additional level of the purpose factor would increase the variation explained. Further, each mode-time combination would yield different levels of predictability, each of which was consistent within time, but not necessarily within mode. The regression equations represent the relations within the condition groups or cells of the ANOVA.

It should be emphasized that this analysis was not oriented to developing predictive relationships, but rather to determining the degree to which variation in the all-purpose group was explained by variation in specific purpose groups. It is not inferred that the equations shown are applicable generally, but rather that variation explained (R^2) may be universal, and that the respective expansion ratios (slopes) are typical.

The first regression analysis examined the simple linear relation between peak hour volume for all purposes and total day volume for all purposes. The analysis considered both modes and treated the three street classes separately.

TABLE 1 - SIMPLE CORRELATION

X: TOTAL PEAK, Y: TOTAL DAY; AUTO DRIVER

(F)	$Y = 824 + 5.446X$	$R^2 = 0.969$
(A)	$Y = 899 + 5.150X$	$R^2 = 0.905$
(C)	$Y = 314 + 5.307X$	$R^2 = 0.875$

X: TOTAL PEAK, Y: TOTAL DAY; HIGHWAY PERSON

(F)	$Y = 1549 + 5.304X$	$R^2 = 0.961$
(A)	$Y = 1410 + 5.090X$	$R^2 = 0.895$
(C)	$Y = 493 + 5.383X$	$R^2 = 0.857$

The predictions (R^2) vary between modes by a maximum of 0.018; the difference increases with decreasing street class importance. The prediction of total traffic based on peak hour traffic is apparently quite reliable, link-by-link, throughout each street class.

The multiple linear regression analysis used a build-up technique to synthesize models in order to examine the feasibility of the development approach employed. The build-up procedure adds independent variables to the regression equation in a stepwise manner, adding at each step the one variable which will cause the greatest decrease in the error sum of squares of the analysis. There is concomitant increase in R^2, the variation explained, with decrease in the error sum of squares. The procedure yields, step by step, the best equation possible, including those variables which most significantly affect the dependent variable.

Subsequently additional multiple linear regression equations were developed by the same stepwise procedure, but including at each step the variable previously selected according to the logical progression of the proposed survey procedures. Each additional step, or variable, represented a further extension of the modified survey implied by the first variable entered. It was thereby shown just how much increase in explained variation and decrease in standard error of estimate could be implied by progressive extensions of the modified survey. The results of this 'selected' run, compared to the previous 'free' run verified the validity of the ANOVA oriented development approach.

The computations for this analysis were performed by program BIMD-2R[2]. Four sets of five equations were developed. Each set had the same all-purpose dependent variable; the sets differed according to the particular time-mode combination of the dependent variable. Three equations of each set represented the three separate street class groups: freeways (F), arterials (A), and collectors (C). The other equations represented combinations of these groups: freeways and arterials (FA), and freeways, arterials, and collectors (FAC). Results of the analysis of the three separate groups are given in tables 2 and 3. Only the R^2 and standard error of estimate values are show, and these are arrayed according to street class, mode, and purpose. As one proceeds in decreasing order of survey complexity (to the right) and decreasing importance of street class (down), the R^2 of the equations decrease. Similar movement in the tables shows the standard error of estimate to increase. Moving down columns does not imply results for cumulative street groups; moving horizontally does imply results cumulative for purposes. Thus, the cell in the upper left corner is for a three purpose, vehicle trip survey to predict total day traffic on the freeway system. The cell in the lower

TABLE 2 - R^2 FOR SEPARATE STREET CLASSES

		WORK and SHOP		WORK		H-B WORK	
		driver	person	driver	person	driver	person
Freeways	peak	.9959	.9949	.9926	.9925	.9869	.9859
	day	.9943	.9912	.9816	.9702	.9730	.9616
Arterials	peak	.9845	.9751	.9625	.9448	.9157	.9013
	day	.9759	.9602	.9104	.8692	.8613	.8160
Collectors	peak	.9744	.9610	.9374	.9149	.8546	.8305
	day	.9631	.9388	.8866	.8229	.8108	.7566

TABLE 3 - STANDARD ERROR OF ESTIMATE FOR SEPARATE STREET CLASSES

		WORK and SHOP		WORK		H-B WORK	
Freeways	peak	57	93	76	112	101	154
	day	368	658	660	1210	799	1370
Arterials	peak	59	105	92	155	137	208
	day	397	711	765	1287	951	1525
Collectors	peak	37	65	57	96	87	136
	day	250	475	437	808	565	947

right corner is for a home-based work, person trip survey to predict peak hour volumes on the collector system. In summary, the R^2 values decrease with decreasing complexity of survey and with decreasing levels of time and street class. The standard errors of estimate decrease with decreasing survey complexity and increase with change in time group and street class.

The R^2 values describe the percent variability explained by the factors in the respective regression equation. Increase in R^2 implies increase in variation explained; respective increases in R^2 should be balanced against additional survey costs necessary to attain these increases. The standard error of estimate implies the width of a confidence band on each prediction. Adding one standard error to a predicted volume assures 0.84 probability that the true volume is not greater than the result; two standard errors implies 0.97 probability of enclosure.

The results of the two class group combinations are given in tables 4 and 5. The tables show the three most significant variables in solution for each time-mode situation. These variables are arrayed according to both the manner in which they entered solution freely and the order that was selected, based on the survey procedure proposed. The R^2 and stepwise increase in R^2 are shown along with the standard error at each step. Also shown are the intercepts and regression coefficients for the selected models.

It was apparent that inclusion of all three street groups (FAC) in the analysis was very beneficial. The R^2 values were higher and the standard error values lower under all conditions in this situation than for any single street class other than freeways. The R^2 values for the

Table 4 - Regression Analysis of Freeway and Arterial Classes (FA)

		FREE VARIABLES				SELECTED VARIABLES							
		Name	R^2	R^2 increase	Standard error	Name	R^2	R^2 increase	Standard error	Intercept	X_1 coefficient	X_2 coefficient	X_3 coefficient
Peak Driver	X_1	HBW-D	.9561	--	129	HBW-D	.9561	--	129	147	1.131	--	--
	X_2	NHBW-D	.9799	.0238	88	NHBW-D	.9799	.0238	88	62	1.116	1.222	--
	X_3	HBS-D	.9911	.0112	58	HBS-D	.9911	.0112	58	36	1.064	1.140	1.248
Peak Person	X_1	HBW-P	.9519	--	195	HBW-P	.9519	--	195	231	1.167	--	--
	X_2	NHBW-P	.9737	.0217	145	NHBW-P	.9737	.0217	145	111	1.151	1.401	--
	X_3	HBS-D	.9869	.0132	102	HBS-D	.9869	.0132	102	67	1.092	1.321	1.955
Day Driver	X_1	HBW-D	.9266	--	919	HBW-D	.9266	--	919	1183	1.382	--	--
	X_2	HBS-D	.9659	.0393	627	NHBW-D	.9524	.0257	741	715	1.220	1.611	--
	X_3	NHBW-D	.9868	.0209	391	HBS-D	.9868	.0344	391	314	1.125	1.456	1.513
Day Person	X_1	HBW-P	.9043	--	1492	[illegible]-P	.9043	--	1492	2047	1.465	--	--
	X_2	HBS-D	.9588	--	979	NHBW-P	.9305	.0262	1273	1391	1.275	2.057	--
	X_3	NHBW-P	.9793	.0205	695	HBS-D	.9793	.0488	695	689	1.158	1.827	2.555

(H:Home; N:Non; B:Based; W:Work; S:Shop; D:Driver; P:Person)

Table 5 - Regression Analysis of Major Street System (FAC)

		FREE VARIABLES				SELECTED VARIABLES							
		Name	R^2	R^2 increase	Standard error	Name	R^2	R^2 increase	Standard error	Intercept	X_1 coefficient	X_2 coefficient	X_3 coefficient
Peak Driver	X_1	HBW-D	.9510	---	105	HBW-D	.9510	---	105	77	1.185	---	---
	X_2	NHBW-D	.9791	.0281	69	NHBW-D	.9791	---	69	31	1.134	1.271	---
	X_3	HBS-D	.9911	.0120	45	HBS-D	.9911	.0120	45	14	1.070	1.209	1.331
Peak Person	X_1	HBW-P	.9442	---	161	HBW-P	.9442	---	161	126	1.227	---	---
	X_2	NHBW-P	.9719	.0277	114	NHBW-P	.9719	.0277	114	59	1.172	1.486	---
	X_3	HBS-D	.9866	.0147	79	HBS-D	.9866	.0147	79	31	1.099	1.422	2.106
Day Driver	X_1	HBW-D	.9307	---	717	HBW-D	.9307	---	717	620	1.486	---	---
	X_2	HBS-D	.9642	.0335	516	NHBW-D	.9573	.0266	564	336	1.290	1.587	---
	X_3	NHBW-D	.9875	.0233	304	HBS-D	.9875	.0233	304	123	1.140	1.489	1.615
Day Person	X_1	HBW-P	.9076	---	1188	HBW-P	.9076	---	1188	1083	1.601	---	---
	X_2	HBS-D	.9565	.0489	815	NHBW-P	.9337	.0261	1007	682	1.379	1.936	---
	X_3	NHBW-P	.9792	.0266	565	HBS-D	.9792	.0455	565	294	1.183	1.851]	2.820

(H:Home; N:Non; B:Based; W:Work; S:Shop; D:Driver; P:Person)

FAC group were always close to those for the FA group, but the FAC standard errors were consistently lower. Thus, by including all streets in the analysis, very little prediction is sacrificed, and the precision of the results is improved. Further discussion will center on the FAC situation.

Of particular interest was the order of entrance and consistency of variables used in the free analysis. The home-based variable for the respective mode-time condition always entered first. The next entry was home-based shop for total day prediction and non-home-based work for the peak-hour situation, both regardless of mode. Home-based shop in person trip models, for both peak hour and total day, entered as a vehicle rather than a person trip variable. This implies that a license plate survey of shoppers would be an adequate technique if supplementary travel data were considered necessary. The change from peak hour to total day and from vehicle to person trips decreased the R^2.

The implications of the results in table 5 are important to the design of survey procedures. Based on the increase in R^2 and the decrease in the standard error of estimate, the order of survey complexity follows directly. The same progression follows for both peak and total day prediction. The time periods sampled in each survey correspond to the time period being predicted. A home-based work driver trip survey requires employer records of those employees driving to work, drivers being defined by supervisor tabulation. A home-based work person trips survey requires tabulation of all employee records and a tabulation regarding mode. An all work trip (home- and non-home based) survey requires interview of drivers or all employees, depending on the mode of interest. Extension

TABLE 6 - MULTIPLE CORRELATION SUMMARY - PEAKS by TOTAL DAY

	VEHICLE			
	Name	R^2	R^2 incr.	S.E.
Freeway	EW	.9732	---	144
	NEW	.9764	.0032	135
Arterial	EW	.8761	---	165
	NEW	.9104	.0344	141
Collector	EW	.8110	---	99
	NEW	.8643	.0533	84
Freeways and Arterials	EW	.9314	---	162
	NEW	.9486	.0172	140
Network	EW	.9341	---	122
	NEW	.9320	.0279	104

to include shopping trips requires a license plate survey at shopping districts in addition to the procedure selected in the manner outlined above.

One further analysis concerned prediction of peak hour travel (a direct design hour) based on a survey of total day work trips. The R^2 values for this analysis are presented in table 6. It was apparent that such a procedure may well be feasible.

Decisions on the form of any revised survey procedure are best made according to costs and feasibility. This research has provided the study director with alternative procedures for replacing the costly home interview survey. It must be his decision, in light of the community conditions, to select a feasible alternative which will provide valid travel patterns at the least possible cost.

REFERENCES CITED

1. Computer Programs Developed by the Authors in Connection with the Research.

2. Computer Programs Available From the Health Sciences Computing Facility, The University of California at Los Angeles.

3. Ostle, B.; *Statistics in Research*; The Iowa State University Press; Ames; 1963.

4. Hicks, C. R.; *Fundamental Concepts in the Design of Experiments*; Holt, Rinehart and Winston; New York; 1964.

5. Box, G. E. P.; "Effects of Inequality of Variance in the One-Way Classification," Annals of Mathematical Statistics; June, 1954; pp. 290-302.

6. Cherniak, N.; "Critique of Home-Interview Type O-D Surveys in Urban Areas;" Highway Research Board, Bulletin 253; Washington, D. C.; 1960.

7. Lapin, H. S.; *Structuring the Journey to Work*; University of Pennsylvania Press; Philadephia; 1964.

8. *Traffic Assignment Manual*; U. S. Bureau of Public Roads, Office of Planning, Urban Planning Division; Washington, D. C.; 1964.

9. *Calibrating and Testing a Gravity Model for Any Size Urban Area*; U. S. Bureau of Public Roads, Office of Planning, Urban Planning Division; Washington, D. C.; 1965.