AN ANALYSIS OF RESIDENTIAL TRIP GENERATION

IN CAPE TOWN

by

ERIC IVAN PAS B.Sc.(Eng.)(Cape Town)

A thesis submitted in partial fulfilment of the requirements for the Master of Science degree in the Faculty of Engineering at the University of Cape Town.

Department of Civil Engineering,
University of Cape Town.

September, 1974.

DECLARATION

I, ERIC IVAN PAS, hereby declare that this thesis is my own work (except where assistance is acknowledged), and that it has not been submitted for a degree at another University.

Signed by candidate

September, 1974.

TO

My Late Parents,

BLUME and ARON

for their unselfish efforts to provide me with a sound education.

TABLE OF CONTENTS

CHAPTER I - Introduction

CHAPTER II - The Analysis of Residential Trip Generation

CHAPTER III - Data Collection

Page

SYNOPSIS

Some of the factors contributing to the growth of a 'transportation problem' in Cape Town are outlined, and the main trends are illustrated. A brief introduction to the Urban Transportation Planning Process is presented, and the importance of the trip generation sub-model within the conventional traffic forecasting model is highlighted.

A review of past and current practice in the analysis of residential trip generation is presented. The least-squares and category analysis techniques are compared, and the dummy variable method is briefly described. The unit of analysis to be used in the development of residential trip generation models is discussed. The independent variables frequently incorporated in models of residential trip-making are examined, and their effects on home-based trip generation are analysed. Some of the problems associated with the development of regression models are examined.

The planning and execution of a home questionnaire survey of households in the Cape Town area is described. The survey was very successful - the response rate being approximately 80 per cent, and therefore a non-response survey was not considered necessary.

The results of the survey, which was completed by 1 690 households, are presented. The purpose, mode, and time of day characteristics of the 10 641 home-based trips reported by the 4 571 respondents are examined.

In the analysis of the survey data, particular attention was given to the question of the appropriate unit of analysis. To this end, the survey information was coded in a manner to facilitate the development of models at the personal, household, and zonal levels of analysis, and six preliminary trip generation equations were derived.

The analysis of variance technique was employed to examine the effect of aggregating trip-making data from one level of analysis to another. These investigations revealed that if the effects of aggregation are taken into account, the personal model is best able to explain the between-person variation in trip-making. However, even the personal model only accounted for 15 per cent of this variation.

The personal equation was found to be marginally superior to the household model when used for estimating zonal trip generation in Cape Town. In addition, the personal equation was found to produce better estimates than the household model, when used for predicting household trip generation rates in four foreign urban areas.

The distribution of the residual error term was investigated for the personal and household equations, and it was found that the personal model was better able to satisfy the least-squares assumption of constant error variance.

Recommendations are made for future research in the analysis of residential trip generation. In particular, it is suggested that the personal level of analysis be further investigated. The need for a comprehensive land-use/transportation study in the Cape Metropolitan Area is emphasized, and urgent attention needs to be given to the 'latent demand' for travel on the Cape Flats.

ACKNOWLEDGEMENTS

I wish to express my gratitute to:-

Associate Professor, Dr. W.H. King (Department of Civil Engineering, University of Cape Town), under whose supervision this research was conducted.

Neil Aplin (Country Roads Board, Victoria, Australia; formerly of the Department of Civil Engineering, University of Cape Town), for all his valuable advice and encouragement during the course of this research project.

Associate Professor A.D.W. Sparks (Department of Civil Engineering, University of Cape Town), for having interested me in analytical transport planning.

Messrs. R. del Mistro, P. Tomlinson and K. Sturgeon (Cape Town City Engineer's Department), for their helpful co-operation in providing information at various stages in this work.

Mrs. Ensor and Mrs. Swann (Computer Centre, University of Cape Town), for their assistance with the punching of the sample addresses and survey data.

Mr. F.J.P. Roux (formerly of the Department of Civil Engineering, University of Cape Town), for his assistance with the translation of the letter and survey forms.

Miss I. Williams (formerly of the Department of Civil Engineering, University of Cape Town), for typing the survey forms.

Messrs. A. Abrahamson, J.S. Braude, J.P. Gibberd, A.B. Jordaan, B.G. Lacey, C.T. Turvey and P. Uys, for successfully carrying out the Cape Town home questionnaire survey.

Messrs. J. Ntanjana and A. Siko (Department of Civil Engineering, University of Cape Town), for their assistance with clerical work during the home questionnaire survey.

The eight young ladies, who assisted with the coding of the survey data.

Dr. A.H. Money and Mr. M. Green (Department of Mathematical Statistics, University of Cape Town), for their advice on statistical problems.

Colin Goldberg and Barry Lichtman (Department of Computer Science and Computer Centre, University of Cape Town), for advice on computing problems.

Miss Bridget Spalding (Department of Civil Engineering, University of Cape Town), for her accurate typing of this thesis.

Mr. K.A.C. Martin, for his assistance with the preparation of the diagrams.

My mother-in-law, Dr. P.L. Back, for proof-reading this manuscript.

Mr. Charles Basson, for his proficient reproduction of the manuscript.

My wife, Michele, for the patience, encouragement, understanding and practical assistance, rendered at all times during the preparation of this thesis. These efforts have been very much appreciated.

CHAPTER I

INTRODUCTION

1.1 Preliminary Remarks

This thesis is concerned with the analysis of residential trip generation. Specifically, an analysis of residential trip generation in Cape Town, is presented.

Trip generation may be considered as the estimation of the number of trips produced by, and attracted to, an areal unit as a function of the socio-economic, locational and land-use characteristics of that areal unit. This study is about trips produced by, and attracted to, residential areas.

Trip generation is the first phase of the traffic forecasting model. This model is part of the urban transportation planning process, which is part of the overall planning of cities and regions.

Analytical transport planning is a relatively new science which has developed as a result of an increasing awareness by engineers, planners, government authorities and others, of the growing importance of urban transportation.

1.2 The Urban Transportation Problem in Cape Town

The increasing importance of urban transportation stems basically from the rapid growth of urban populations[1]*. This growth has been brought about as a result of the combined effects of the normal population growth and migration of people from rural to urban areas. This latter phenomenon is referred to as urbanization.

In keeping with the trend in other countries of the developed world, the urban population of South Africa has been growing rapidly, relative to the rural population of the country. This trend is well illustrated by the population figures plotted in Figure 1.1[2]. The growth of the population in the Cape and Wynberg magisterial districts is shown in Figure 1.2[3].

Another major factor contributing to the increasing concern about urban transportation is the almost dramatic increase in the motor car

---

*Refers to the notes at the end of the chapter.

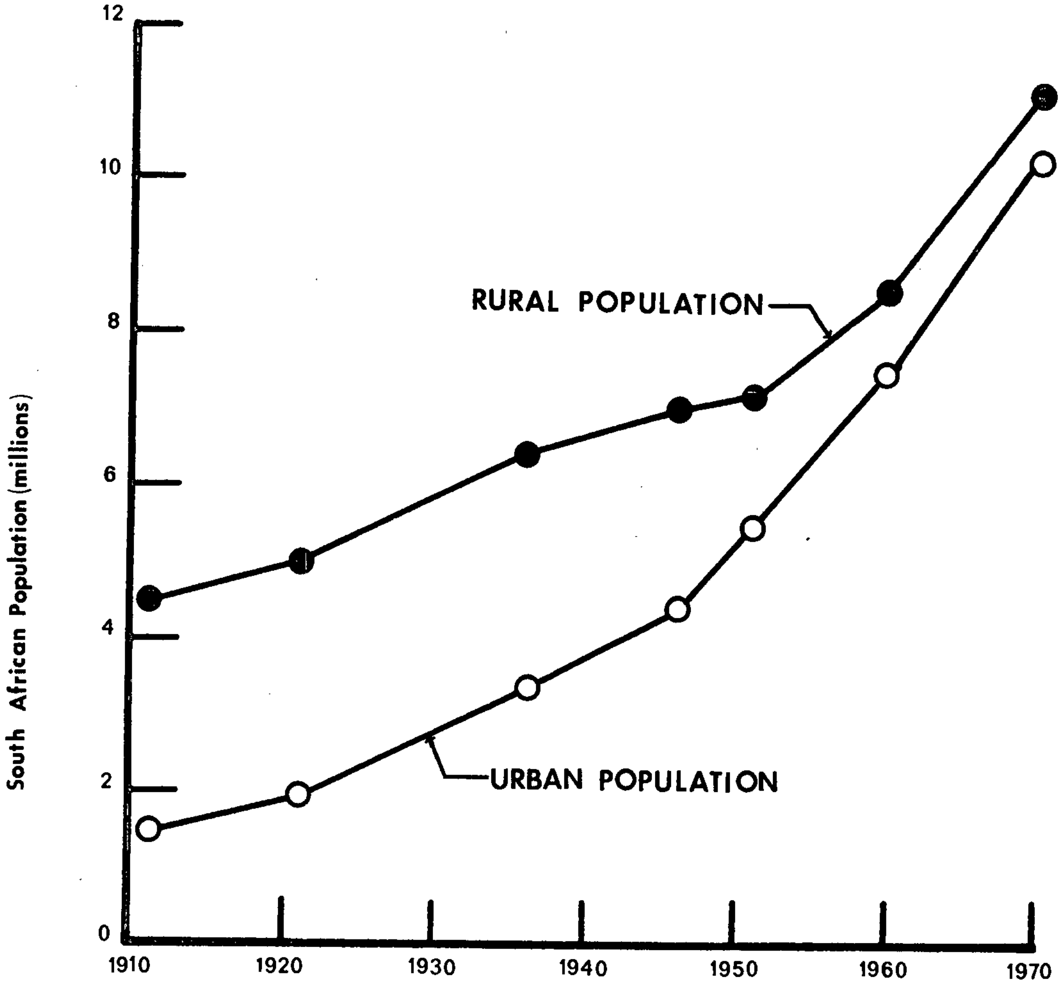


FIG 1.1 Urbanization In South Africa

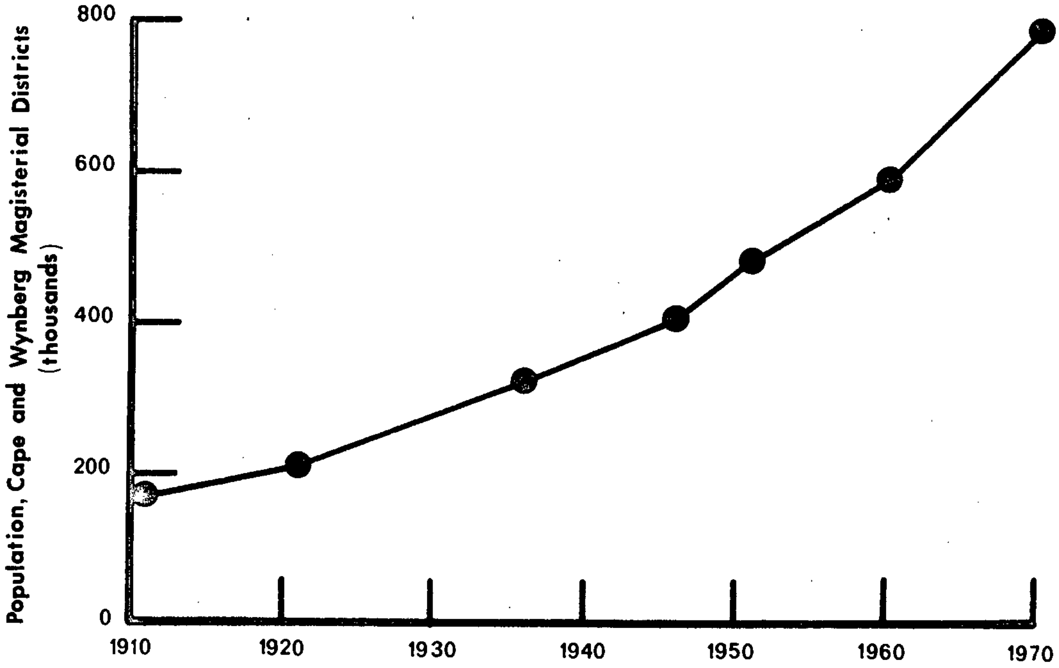


FIG. 1.2 Population Growth In Cape Town

population of urban areas. Figure 1.3 illustrates the growth of the Cape Town (CA) and Bellville (CY) motor car registration figures between 1934 and 1970[4].

Consideration of the increase in the number of cars per head of population leads to a better understanding of the problem. Rees[5] points out that the number of motor vehicles per thousand people in Cape Town increased from 164 to 345 between 1962 and 1970, for the white population group. Comparable figures for the coloured population group are 7 and 23 cars per thousand people.

Accompanying this rapid increase in car ownership has been a significant decrease in the use of public transport. The information shown graphically in Figure 1.4 illustrates the declining patronage of the local buses[6].

The fact that there is a concentration of trips in the peak periods, and that the majority of work trips are oriented to the C.B.D.[7], further complicates the matter. In fact, Floor[8] has described the urban passenger transportation problem as, ".... the overlapping in time and space of too many movements."

The growth of large metropolitan areas is another factor to be considered. These areas fall under the jurisdiction of a number of local authorities. In the absence of a metropolitan planning authority, there is a lack of co-ordinated planning in these areas. A new Cape Metropolitan Planning body is being formed at present[9], but up to now there has been no co-ordinated planning at this level.

In Cape Town, as in the other urban areas of South Africa, the problem is further complicated by the effects of Government policy. Many of the lower income members of the population have been forced to live on the periphery of the cities and thus many, who would otherwise have been able to walk to work, are forced to make long and expensive journeys to get to work.

Public transport, which is essential to the vast majority of these people, cannot serve the peripheral areas adequately. In addition, public transport vehicles are used inefficiently because of the separate seating arrangements.

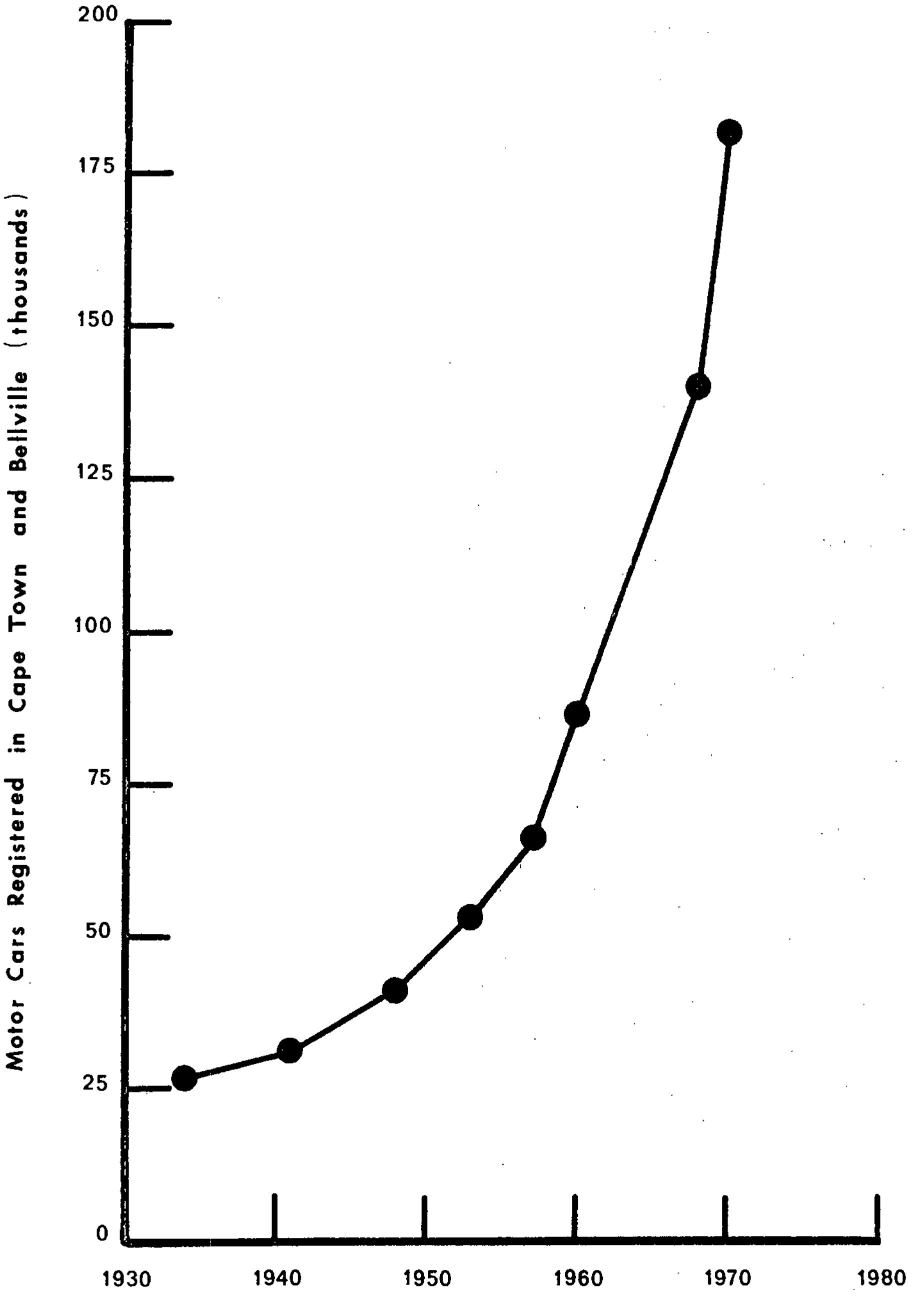


FIG. 1.3 Increasing Car Ownership In Cape Town

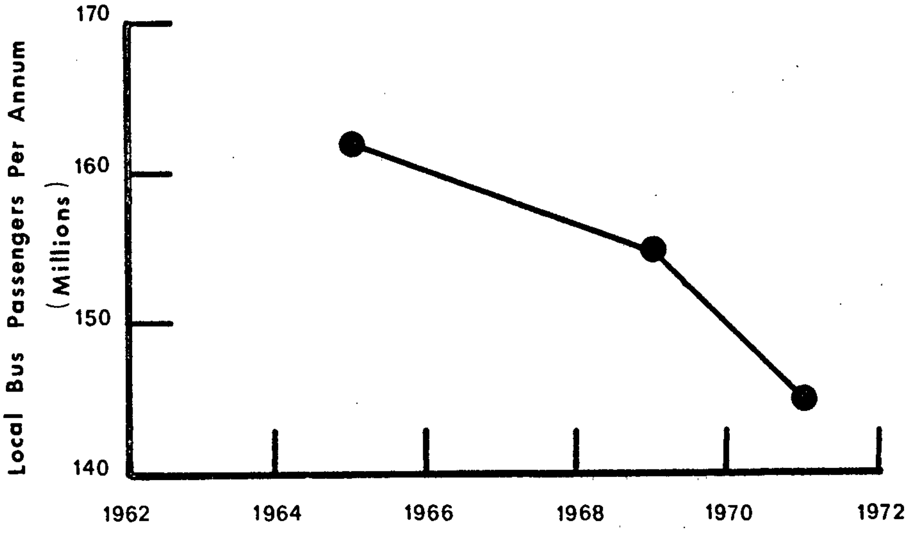


FIG. 1.4 Decreasing Bus Usage In Cape Town

The trends illustrated by Figures 1.2, 1.3 and 1.4 are typical of urban areas in developed countries throughout the Western world. These trends have led to what is often called the 'urban transportation problem' and sometimes called (perhaps exaggeratedly) the 'urban transportation crisis'. While most of us may not consider the situation to be of crisis proportions, many people acknowledge that a problem does exist.

1.3 The History of Urban Transportation Planning

Over the years, different approaches to tackling the urban transportation problem have emerged. According to Oi and Shuldiner[1], the historical development of scientific research in the area of urban transportation may be divided into three periods of time:-

(i) Prior to 1944.
(ii) Between 1944 and 1951/1952.
(iii) From 1952 onwards.

Although the boundaries between these periods are obviously fairly vague, they do indicate notable changes in the approach taken in urban transportation planning and research.

In the period prior to 1944, traffic engineers were dominant. They were essentially concerned with technological design problems such as street capacity, traffic signals, etc.

In 1944 the first origin and destination study was conducted by the Bureau of Public Roads in the United States of America. This was the first time that recognition was given to the idea that there is a need to collect information regarding travel activity within urban areas. During the seven years following the first origin-destination study, a large amount of travel data was collected in a number of such studies.

The period since 1952 is characterized by numerous analytical studies of the data collected in the early, and subsequent, travel surveys. During this period, the field of urban transportation research and planning has been infiltrated by sociologists, economists, city planners, geographers and others.

This mixing of disciplines has led the transportation engineer and his colleagues to analyses of travel demands and behaviour patterns. In addition, the advent of the computer age has encouraged the application of sophisticated and powerful statistical tools. This has enabled us to develop a greater understanding of urban travel behaviour.

1.4 Transportation Planning in Cape Town

At the time when the present study was initiated, the city of Cape Town had not really expanded its transportation planning and research beyond the 1944 type of O-D study.

In 1956 an O-D study was carried out by the City Engineer's Department of the Cape Town City Council. At that point in time, very few O-D surveys had been undertaken outside the U.S.A. and Cape Town's effort is to be commended. However, O-D surveys are fairly limited, and the City Engineer's Department in Cape Town has tended to rest on its laurels since 1956.

A very extensive and expensive urban freeway system has been planned in Cape Town, based on the information gathered in the 1956 O-D survey and subsequent traffic counts. The Shand Committee[10], appointed in 1962, was only concerned with road and traffic planning for the foreshore and central area of Cape Town.

When the field-work for this study was virtually complete, the City Engineer's Department in Cape Town began a survey of travel habits in the Cape Metropolitan Area[11]. This survey dealt only with the first trip undertaken by all respondents on a particular day (Thursday, 25th October, 1973). The information was requested on a mailed questionnaire, and reply-paid envelopes were provided.

Approximately 20 000 householders throughout the Metropolitan area were sampled for this survey. As would be expected from a survey of this nature, a fairly low response rate was obtained. It is likely that the replies were obtained from a rather biased sample of the households in the survey area.

1.5 The Conventional Urban Transportation Planning Process

Since the middle of the 1950's, a philosophy and methodology has been developed, by transportation planners and other research workers, for analysing travel behaviour in urban areas as a basis on which to plan and design future transportation systems.

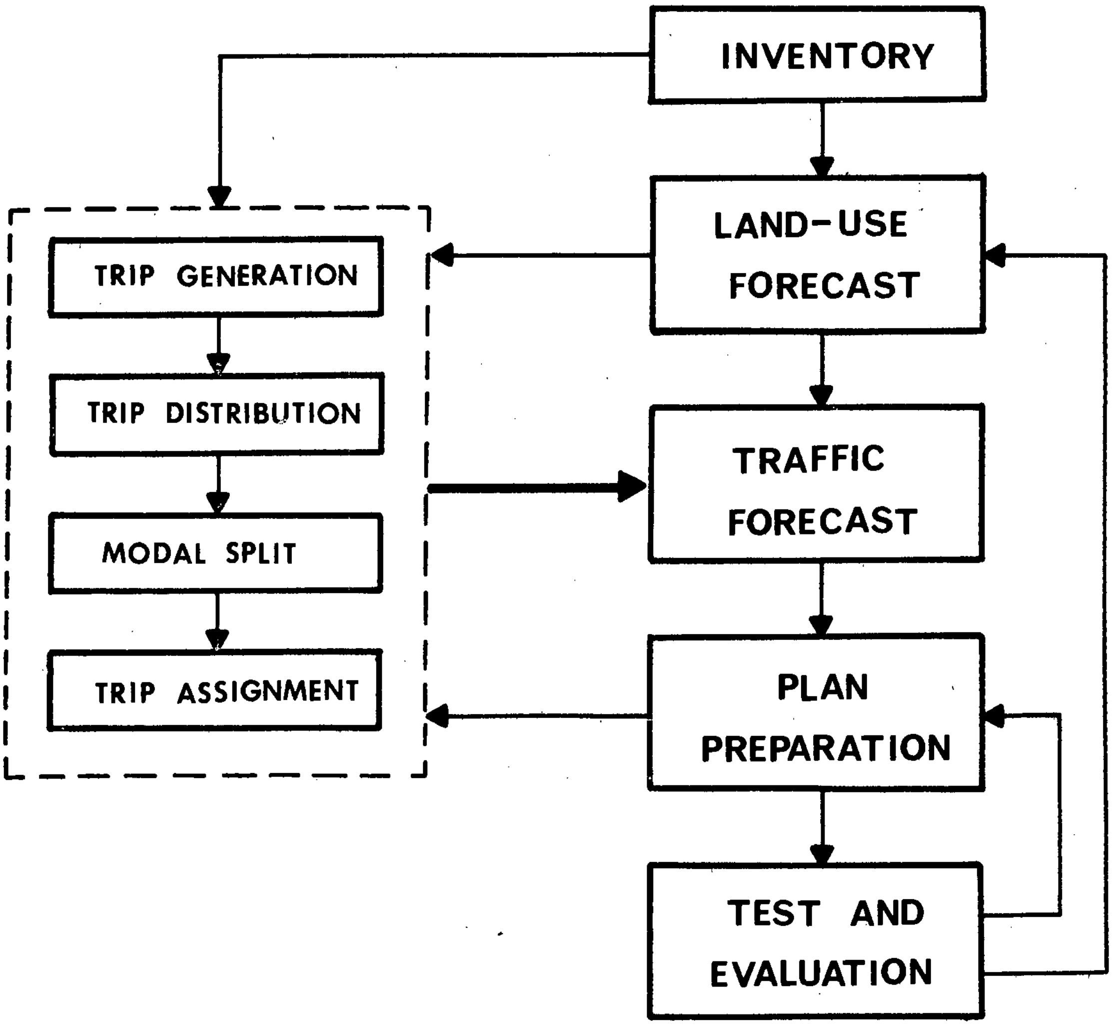


FIG. 1.5 The Conventional Transportation Planning Process

The literature to-day abounds with criticisms of what may be termed the 'conventional urban transportation planning process'. While some modifications may be urgently needed, the transportation planning process that has evolved provides a reasonable basis for the planning and design of transportation facilities required at some future date.

The basic premise of the conventional transportation study is that the demand for travel is repetitive and predictable[12]. Figure 1.5 illustrates the conventional urban transportation planning process. The traffic forecasting model consists of four sub-models as shown on the left of Figure 1.5.

Although this thesis is concerned specifically with residential trip generation in Cape Town, it is necessary that we examine the transportation planning process and the traffic forecasting model in order to appreciate the role of the trip generation sub-model.

The first step in the conventional planning process is to prepare inventories of population, land use, employment, transportation facilities and traffic patterns in the survey area. The information collected is used for the development of the four sub-models of the traffic forecasting phase.

The data collected in the base year inventories is also used for preparing a land-use forecast for the survey area. This land-use plan is input to the traffic forecasting model and a traffic forecast is produced. At first, the existing transportation system is assumed for the design year and its adequacy is tested. Subsequently, alternative transportation plans are prepared, and each is tested and evaluated. Finally, a decision is made as to the most suitable land-use and transportation plan for the survey area.

Two important aspects are not shown in Figure 1.5, nor have they been mentioned above. The reason being that, although they are important, they have not generally been treated formally as part of the transportation planning process.

Firstly, the goals and objectives for the study area should be defined before any comprehensive study is commenced, since these are vital to the evaluation phase of the planning process.

Secondly, the implementation of the proposed transportation plans should be incorporated as a formal step in the planning process. Together with the implementation, there should also be a continuing study to up-date the rather long-range forecasts made during the initial study. These follow-up studies would also provide valuable information needed for evaluating the adequacy of existing planning techniques.

At this stage we can review the specific function of each of the four traffic forecasting sub-models.

1.5.1 Trip Generation

As we have already mentioned, trip generation analysis aims at developing a model which is capable of estimating the traffic flow to and from a given piece of land. This is achieved by relating the number of trips to various socio-economic, locational and land-use characteristics of the piece of land.

Trip generation models may be developed for different trip purposes (e.g. work, shopping, pleasure), as well as for different time periods (e.g. morning peak, afternoon peak), and for different modes of transport (e.g. car-driver, car passenger, bus passenger). The basic concept remains the same, whatever the type of trip we are studying.

In trip generation analysis we relate a dependent variable (e.g. trips per household per day) to a number of independent variables (e.g. number of cars per household, number of people per household, household income). The technique commonly used to develop these relationships is known as multiple linear least-squares regression.

Lately, a second basic approach to trip generation model development has been introduced. This technique is usually referred to as a category analysis.

1.5.2 Trip Distribution

The survey area is usually divided into a number of zones (sometimes called traffic analysis zones), and the trip generation models are used to estimate the number of trip ends in each zone.

The function of the trip distribution model is to use these trip end estimates as input, together with other relevant information, and to produce a matrix of inter-zonal trips. In other words, the trip distribution model distributes the trips, estimated by the trip generation model, between the various zones of the study area.

It is important to notice that current trip distribution techniques are primarily designed to produce the relative rate of attraction between zones, and the actual volumes of inter-zonal movement are obtained only when these rates are applied to the trip generation estimates[13].

We see that the trip distribution model (and therefore the rest of the traffic forecasting process) is completely dependent on the output from the trip generation model.

1.5.3 Modal Split

Modal split may be introduced either before or after trip distribution. The function of the modal split model is to estimate what proportion of the trips will be made by each mode of transport.

If several different modes of transport are available for making a particular trip, then certain factors will influence the choice of mode. Modal split analysis attempts to take account of these factors and to produce a model capable of explaining the distribution of trips between the different modes of transport. The most common technique used for structuring this model is multiple least-squares regression.

1.5.4 Trip Assignment

The output from the above three models may be considered as a set of origin-destination tables, one for each mode of transport. The trip assignment model makes use of these tables as input, together with a description of the transportation networks in the study area.

In travelling from one zone to another, by a particular mode of transport, there may be more than one route that can be used by the traveller. The trip assignment model attempts to estimate what proportion of the trips will use each of the available routes. As output from this model, we obtain estimates of the traffic volumes on each link of the transportation networks of the study area.

1.5.5 The Traffic Forecasting Model

Each of the four sub-models of the traffic forecasting process is calibrated for base-year conditions. The traffic forecasting model is then used, together with a land-use forecast for the study area, to predict future travel patterns.

During the last five years, a number of researchers have suggested the replacement of the four sub-models by a single model[14, 15, 16]. While some of the arguments put forward for the proposed single model may be valid, the author feels that similar results could be achieved by proper use of the conventional four-step traffic forecasting model.

The outline given above of the traffic forecasting process illustrates the importance of the trip generation model. Since the other three sub-models merely distribute the trips predicted by the generation phase, the volume of trips estimated on any link of the transportation network is completely dependent on the output of the trip generation model.

In spite of this fact, much more research effort has been devoted to the development of the trip distribution, modal split, and trip assignment models, than has been devoted to the study of trip generation. One of the reasons for this neglect is that trip generation is a relatively simple concept. However, we should ensure that the complexities of the distribution, modal split, and assignment problems, do not overshadow the importance of the relatively simpler concept of trip generation[13].

1.6 Summary

This introductory chapter has outlined the circumstances which have motivated this study. Briefly, these may be summarized as follows:-

(i) The recognition of the growth of a transportation problem in the Cape Town Area.

(ii) The realization that insufficient use has been made in Cape Town of the tools of the relatively new science of analytical transport planning.

(iii) The recognition that relatively little research effort has been devoted to the very important trip generation phase of the transportation planning process.

Ideally, a comprehensive land-use and transportation study should have been undertaken for the whole of the Cape Metropolitan Area. However, since very limited resources were available to the author, two alternatives were considered. Either, a comprehensive study could have been attempted on a rather superficial level; or an in-depth study could have been made of a particular aspect. It was decided that a detailed investigation was preferable, since a comprehensive study would have involved too many generalizations and assumptions.

As a result of the relative neglect of trip generation analysis as a field of research, the author decided to carry out a study of residential trip generation in Cape Town.

NOTES - CHAPTER I

1. Oi, W. and Shuldiner, P.W., An Analysis of Urban Travel Demands (Northwestern University Press, 1962).

2. Data extracted from report number 02-05-01 of the South African Department of Statistics.

3. Data extracted from report number 02-02-01 of the South African Bureau of Statistics.

4. Data kindly supplied by the South African Department of Statistics.

5. Rees, Brian J., Group Areas Legislation - Its Impact on Urban Structure and Movement Patterns with Special Reference to Cape Town. (Paper presented at the Fifth Quinquennial Convention of the South African Institution of Civil Engineers, Johannesburg, August 1973).

6. Financial Mail, October 27th, 1972.

7. Central Business District.

8. Floor, B.C., The Urban Transportation Problem in South Africa. (Transportation Research Centre, Stellenbosch, 1968).

9. The Argus, March 9th, 1974.

10. Report of the Committee of Consultants on Road, Traffic and Town Planning in Foreshore and Central City Area of Cape Town.

11. The Cape Times, October 25th, 1973.

12. Worral, R.D., Trip Generation. (Notes for a Short Course in Urban Transportation Planning at the University of Stellenbosch, January, 1972).

13. Douglas, A.A. and Lewis, R.J., "Trip Generation Techniques - 1. Introduction", Traffic Engineering and Control, November 1972.

14. Domencich, T.A., Kraft, G. and Vallette, J-P., "Estimation of Urban Passenger Travel Behaviour: An Economic Demand Model", Highway Research Record, Number 238, 1968.

15. Brand, Daniel, "Theory and Method in Land Use and Travel Forecasting", Highway Research Record, Number 422, 1973.

16. Manheim, Marvin L., "Practical Implications of some Fundamental Properties of Travel-Demand Models", Highway Research Record, Number 422, 1973.

CHAPTER II

THE ANALYSIS OF RESIDENTIAL TRIP GENERATION

2.1 The Rationale of Trip Generation Analysis

2.1.1 Introduction

The rationale of trip generation analysis has been summarized as follows[1]:-

(a) Travel is an aspect of derived demand.

(b) The intensity of travel to or from a given piece of land is a function of the land-use activity.

(c) The intensity of travel may be estimated independently of the transportation service provided, and independently of the set of opportunities available.

(d) The relationship between the trip rates and the characteristics of the areal unit (to which these rates apply), may be assumed to remain stable over time.

Each of these aspects is discussed in detail in the next four sub-sections of this chapter.

2.1.2 The Derived Demand for Travel

If we consider a typical week-day, it is obvious that very few trips are made purely for the pleasure that they themselves provide. This means that a trip is usually part of some other activity such as working, shopping or attending school, and most transportation studies use these activities for classifying trips according to the purpose for which they were made. Because travel is a derived demand, we expect to be able to express variations in travel behaviour in terms of socio-economic, locational, and land-use variables.

Oi and Shuldiner[2], when developing a theory of consumer behaviour in urban travel, suggest that we may divide trips into two broad categories. Trips made for the purpose of earning a living may be thought of as 'production-oriented', while leisure trips (for example), are 'consumption-oriented'.

Some trips cannot easily be classified in this way, but the difference between production and consumption-oriented trips is important. A consumption-oriented trip usually involves expenditure over-and-above the time and money cost of the trip itself, while the price of a work trip is usually relatively small compared with the income earned. As a result, we would expect the price of travel to exert little effect on the number of work trips generated by a household.

2.1.3 Effect of Land-Use Activity

It is important to realize that a given piece of land will generate a different number of trips in a given time period (e.g. hour or day), depending on the land-use activity. In fact, not only will the intensity of the traffic generated be different, but the composition of the traffic will also be a function of the land-use activity.

Consider a given piece of land being used for the following three alternative purposes:

(a) Thirty single dwelling units.

(b) Three blocks of flats, each with twenty-five units.

(c) A large shopping centre.

A brief reflection on the quantity (and type) of traffic produced by these alternatives should be enough to convince the reader that land-use activity has a vital effect on the intensity of travel to or from a given piece of land. The need for comprehensive land-use and transportation planning is clearly illustrated by this example.

2.1.4 Effect of Transportation Service and Available Opportunities

In most of the urban transportation studies carried out to date it has been implicitly assumed, in cases where it was not explicitly stated, that the intensity of travel may be estimated independently of the transportation service provided and the set of opportunities available. In some of the latest studies attempts have been made to account for the different levels of service provided to various parts of the study area by the existing and planned transportation systems.

The results of the London Transportation Study[3] do not show level of service to be a significant explanatory variable in the analysis of trip generation. However, intuitively one might postulate some relationship between the number of trips generated by a piece of land and the transport service provided to that land. Blunden[4] has said that, "..... traffic is the joint consequence of land-use activity and transport capability. One or other on its own will not produce traffic".

Domencich[5] et al use an extreme example to illustrate this point. A woman living on a relatively isolated island would probably make very few shopping trips to the mainland because of the time, cost, and general difficulty of making the trips. Each trip would be well planned and executed. In contrast, consider a woman living close to a large shopping district. She may well make a large number of trips, each of which would not be carefully planned. If, on a particular visit to the shops she forgets an item, the cost of making a further trip is relatively minor.

Although the example is extreme, it does show that the location of a household, relative to the transport network and the activities available in the study area, could have an effect on the trips made by the members of the household.

2.1.5 Temporal Stability of Trip Generation Models

If one attempts to predict future travel patterns on the basis of current relationships, then one is forced to assume that the present relationships will remain stable with time. The only alternative is to study the same phenomenon at a number of points in time and to incorporate the time effect in the calibration of the model.

The latter approach is obviously more suitable, although very costly and time consuming. In addition, it is very difficult to account for future unforeseen technological developments; yet these developments may be extremely important factors influencing future urban travel patterns.

The author feels that, to a large extent, the temporal stability of a trip generation equation depends on the proportion of the variation, in the dependent variable, which is accounted for by variations in the independent variables included in the model. However, the problem of possible future changes in transport technology and policy still exists, since the regression analysis cannot easily take these factors into account.

2.1.6 Summary

In sub-sections 2.1.2 to 2.1.5 we have dealt with the rationale of trip generation under the four headings mentioned by Worral[1]. However, we must point out that Worral himself questions the logic of assuming trip generation to be independent of the level of transport service provided.

Trip generation equations developed in different study areas have been found to be significantly different from each other. It is possible that the assumption, that trip generation is independent of the transport service, is partly responsible for some of these differences.

The temporal stability of such equations can certainly be questioned, since if they are inconsistent from one area to the next, are they likely to remain stable with time? However, these differences may also be the result of other factors as we will point out later in this chapter.

It may well be asked whether we can justify trying to quantify a phenomenon which is a result of the habits and attitudes of a number of individuals. However, as we have already pointed out, certain factors motivate and influence the generation of trips. "These motivations cause urban travel to become repetitious and therefore subject to predetermination by formulae and laws similar to that of physical phenomena"[6].

2.2 Methods of Analysis

2.2.1 Introduction

Two basic approaches have been used in the analysis of residential trip generation. The earlier transportation studies relied exclusively on the technique of multiple least-squares regression, while some of the latest studies have used the technique known as category analysis.

Both of these approaches are standard statistical methods of analysis, but a fairly detailed discussion is warranted, since much confusion has arisen in the literature on trip generation analysis.

2.2.2 Multiple Least-Squares Regression

The concept of regression is a simple one. In a general sense, regression assumes that a variable y is related to other variables $x_1, x_2, \ldots\ldots x_k$. The variable y is the one under study and is known as the dependent variable. The variables x_1, x_2, ... x_k are known as the independent variables. Since they seemingly explain variations in the value of y, the independent variables are sometimes termed 'explanatory variables'.

Basically, we have observations of different values of y and corresponding values of x_1, x_2, ... x_k; and regression analysis enables us to fit the 'best' surface to the observed data. The criterion normally used for determining the best surface is the 'principle of least-squares'.

It is extremely important to note that the least-squares approach only enables us to fit the 'best' surface once the shape of the surface has been chosen.

The relationship between y and x_1, x_2, ... x_k is usually assumed to be linear and hence we talk of linear least-squares regression. Usually more than one independent variable is included in the trip generation regression model and we talk of multiple linear least-squares regression. This model is of the form:

$$y = a_o + a_1 x_1 + a_2 x_2 + \ldots\ldots a_k x_k + E \qquad (2.1)$$

where y is the dependent (or response variable)

x_1 ..... x_k are the independent (or explanatory variables)

a_o ..... a_k are the model parameters

E is usually called the error (or disturbance) term.

The error term indicates that the model does not necessarily take account of all the factors that influence the value of y. In addition, this term represents the effect of random deviations.

Using the principle of least-squares, the model parameters may be estimated to obtain the equation

$$y = b_o + b_1 x_1 + b_2 x_2 + \ldots\ldots + b_k x_k \qquad (2.2)$$

where b_o ..... b_k are the least-squares estimators of the unknown model parameters a_o ..... a_k. These estimators are chosen in such a way as to minimize the squared sum of the deviations from the estimated regression surface.

The least-squares principle is outlined in Appendix 1. The reader who is unfamiliar with the technique, or the associated terminology, is referred to this appendix, since much of the later discussion relies on a knowledge of the least-squares concept.

A number of important assumptions are implicitly made by the analyst when using the least-squares technique[7]. These are:

(a) The mean and co-variance of the error (or disturbance) terms are zero.

(b) The variance of the error terms is constant and their distribution is normal.

(c) The independent variables are not correlated with one another.

In addition, we must note that the least-squares model estimates the mean value of the dependent variables for given values of the independent variables. Hence, measurement errors in the independent variables are not accounted for by the regression model[8].

Of the assumptions listed above, the one which is most commonly violated is the condition of constant error variance[7]. If the error variance is dependent on the independent variables, then the data is said to be heteroscedastic. Heteroscedasticity can lead to an overstatement of the accuracy of the regression model, and hence its consequences can be serious[9].

The effects of the independent variables in the regression model are assumed to be additive, and therefore the independent variables are assumed to be uncorrelated with each other. The separate effects of two or more inter-correlated variables are difficult to isolate, and the condition of multi-collinearity is said to exist. This condition is sometimes reflected by regression coefficients having a magnitude or sign contrary to logical expectation.

It is important to realize that there are a number of statistical tests available to the analyst when he examines a multiple least-squares regression equation. These tests are discussed in Appendix 1. Besides being able to evaluate the overall significance of the regression equation, one can test the significance of individual independent variables, and estimate the predictive accuracy of the regression equation.

2.2.3 Category Analysis

The method known as category (or cross-classification) analysis, has been applied to trip generation analysis mainly due to the increasing dissatisfaction with the results obtained using the least-squares regression technique.

The poor results are not entirely due to the method itself, but largely arise from mis-use of the technique. Nevertheless, the category analysis approach has become popular lately, especially in the United Kingdom.

It is important to appreciate that category analysis is a proper form of regression analysis[10]. This will become clear when one considers the method, as outlined below, in relation to the basic concept of regression as outlined at the beginning of section 2.2.2. Therefore, least-squares and category analysis are two alternative approaches to regression analysis, and the title of Winsten's paper[10], "Regression Analysis versus Category Analysis", is rather misleading.

The basis of the cross-classification method is that we determine the average value of the dependent variable for defined categories of the independent variables. The categories are defined by constructing a multi-dimensional matrix in which each dimension represents one of the independent variables. All the independent variables are stratified into a number of discrete class intervals and every observation is assigned to a cell of the matrix. The mean value of the dependent variable is found for all the observations falling into each cell.

As an example, suppose we were studying the effect of sex and height on the mass of a person. We would take a number of observations on these three variables and we could set up the following two-dimensional (since we have only two explanatory variables) matrix:

Height (Metres)	Sex: Male	Sex: Female
< 1,52		
1,52 - 1,60		
1,61 - 1,70		
1,71 - 1,80		
> 1,80		

Note:
The average mass of the people in each category is recorded in the relevant cell.

Table 2.1

Notice that in the above example we have stratified the one variable (sex) into two discrete classes, while the other variable (height) was stratified into five discrete classes. The number of independent variables to be used in setting up the matrix, and the number of classes into which each of these variables is stratified, are chosen by the analyst.

However, increasing the number of independent variables to be considered, or increasing the number of classes into which each of these variables is stratified, increases the demands made on the data collected. In order to use the average value of the dependent variable for all observations falling into a particular cell, we need to have a reasonable number of observations in that cell.

If, in the above example, we decided to take account of the age of the people studied, in addition to the other two independent variables, and we stratified age into four discrete classes, we would increase the number of cells in the matrix from ten to forty. This would mean collecting far more data than would be necessary if we only considered the original two independent variables.

The first application of the category analysis technique in trip generation analysis was in the Puget Sound Regional Transportation Study[11]. In this original application, the method was termed rank-classification; but it has subsequently been called category, or cross-classification analysis.

In the Puget Sound Study the model was developed and applied at the zonal level. The usual data was collected at the household level and the zonal averages for each of the variables were computed. Each zone was then allocated to a cell in the multi-dimensional matrix. The average trip rate was then computed for each cell in the matrix.

The category analysis method can be applied at the household, zonal, or even personal level of analysis; as can the least-squares regression procedure. The question of the appropriate unit of analysis is discussed in section 2.3.

The basic assumption of the category analysis technique, when applied to the analysis of residential trip generation, is that the average trip rates observed during the base-year study will remain stable with time.

It is very important to note that the use of the category analysis method frees the analyst of many of the assumptions inherent in the least-squares regression technique. In particular, the shape of the response surface does not have to be defined, and no assumptions regarding the distribution of the error terms need be made. We also notice that non-quantifiable variables (e.g. sex) are easily incorporated in the category analysis approach.

The category analysis technique has been developed to a considerable degree in the United Kingdom. Wooton and Pick [12] have developed a method for allocating households to the various categories. A system of distribution functions (e.g. Gamma, Poisson) is used for this task. In their work, Wooton and Pick studied three independent variables:

(a) Household Structure - based on the number of employed and unemployed people in the household.

(b) Income.

(c) Car Ownership.

The household structure and income variables were stratified into six classes each, while the car ownership variable was stratified into three classes. This means that the resultant three-dimensional matrix had 108 cells. Assuming that the data were uniformly distributed, more than 1 000 home interviews would be required to obtain only ten observations in each cell of the matrix. If one further takes into account the mode and purpose of the trips, one realizes that the data requirements of the category analysis method are rather excessive - especially if we wish to consider more than three variables.

Douglas and Lewis [13] point out that many important categories could be poorly represented if a conventional random sampling technique is used. In some study areas, multi-car owning households may not be very common in the base year survey data, but will very likely feature rather prominently in the design year.

The category analysis technique is conceptually simple, but it is deficient as a model building technique. This is because there is no means for testing the statistical significance of the various explanatory variables thought to influence the dependent variable. The variables to be used in setting up the multi-dimensional matrix must be chosen by the analyst a priori, and we have already pointed out the problem of sample size which arises if many explanatory variables are considered.

2.2.4 The Dummy Variable Technique

Two of the advantages of the cross-classification method over the least-squares regression approach are:

(a) The shape of the regression surface does not have to be defined in the cross-classification technique.

(b) Non-quantifiable variables are easily handled by the category analysis method.

However, both these advantages can be overcome, to a large extent, by incorporating the dummy variable technique in a least-squares regression analysis.

The dummy variable idea is not new, but it was not formally documented until 1957[14]. In this method, a value of one is assigned to the dummy variable corresponding to the class within which the observed value of the independent variable falls, while the other dummy variables in the set are assigned the value of zero.

The technique has been used essentially for incorporating, in a least-squares regression analysis, the effects of variables that are not conventionally measured on a numerical scale, e.g. race, sex, religion, etc. Suppose we are studying (as before) the effect of sex and height on the mass of a person. The variable 'sex' can be considered as two dummy variables, S_1 and S_2. For a male, S_1 would be one, while S_2 would be zero; and vice versa. In general, presence in a class is indicated by unity, and absence by zero.

Although the variable 'height' is continuous, we could incorporate it also as a set of dummy variable classes. It is important to notice that if we treat all the explanatory variables as dummy variables, then the least-squares regression technique will be very similar to the category analysis method[15]. In fact, if we have only one independent variable, then the least-squares regression equation (with dummy variables) will give the same results as the category analysis approach.

The introduction of dummy variables creates a problem which is easily solved, but not always appreciated. The utilization of all the classes of any one dummy variable set will render the normal equations (see Appendix 1) indeterminate. Basically this results from the fact that the dummy variables of any set are linearly related.

Considering the example above, of the two dummy variables classes (S_1 and S_2) used to represent the variable 'sex', we see that for all observations $S_1 + S_2 = 1$ (assuming that all the people studied were either male or female). The effect of this linear relationship on the normal equations is investigated in Appendix 2.

Two approaches have been suggested by Heathington and Isibor[16] for overcoming the indeterminancy which arises if all the dummy variables of a particular set are included in a least-squares regression analysis. Both the approaches involve the addition of an extra constraint to the solution of the normal equations, and both involve modification of the usual interpretation of the partial regression coefficients. The easier method is to delete one of the dummy variables from each set, and the partial regression coefficients will then express variations relative to the excluded class, in each set.

We have already mentioned that the use of dummy variables can overcome the problem of specifying the shape of the response surface. If we stratify a quantifiable variable (e.g. family size) into a number of dummy variable classes, and use these variables in a multiple linear regression analysis, we are in fact approximating the relationship between the independent and dependent variable by means of a series of straight lines.

McCarthy[17] has suggested that the use of dummy variables in trip generation analysis be more fully investigated.

2.3 The Unit of Analysis

2.3.1 Introduction

The data used in the analysis of residential trip generation is generally collected at the household level, but it is usually aggregated to the zonal level before the analysis is undertaken. The aggregation of household data to the zonal level, prior to analysis, has been suggested as the major reason for the fact that trip generation models have been found to differ significantly from one study to the next[17,18,19,20,21,22].

The least-squares regression technique and the category analysis method can both be used either with aggregated or raw data. However, in the sub-sections which follow, we will mainly use the least-squares method in discussing the difference between the alternative units of analysis, since the majority of transportation studies have used this method for structuring the trip generation models they have developed.

2.3.2 Zonal Analysis of Residential Trip Generation

The primary basis for utilizing data aggregated to the zonal level in analysing residential trip generation, has been the assumption that geographical proximity results in a similarity of households with respect to trip-making and socio-economic characteristics.

Another factor to be considered is the fact that the output required from the trip generation phase is a set of zonal trip productions and attractions. Of course, this does not mean that the analysis need be carried out at the zonal level, but possibly this was not always appreciated.

In addition, analysis at the zonal level means that the regression program needs to deal with far fewer observations, since we have only one set of observations per zone. It should be remembered that when the earlier transportation studies were carried out, the speed and capacity of the computers available may well have been a limiting factor.

In developing zonal regression models, we have the choice of either using zonal total or rate variables. The difference between these two variables is best understood by considering a zone in which we have determined the number of trips, households, cars, etc. Let these quantities be denoted by $s_1, s_2, \ldots\ldots, s_i, \ldots\ldots, s_k$. In addition, consider a set of units (e.g. streets, households, etc.) into which the items above can be grouped. Let the number of each of these units, in the zone, be denoted by $t_1, t_2, \ldots\ldots, t_j, \ldots\ldots, t_m$.

A zonal <u>total</u> variable is defined as s_i per zone, e.g. trips per zone.

A zonal <u>rate</u> variable is defined as s_i/t_j per zone, e.g. trips per household per zone.

Douglas and Lewis[9] provide the following sample zonal regression equations:-

$$\text{Personal trips per zone} = 3{,}02X_3 \quad (R^2 = 0{,}98) \qquad (2.3)$$

$$\text{Personal trips per household per zone} = 3{,}05x_3 \quad (R^2 = 0{,}42) \qquad (2.4)$$

where X_3 = persons over 5 years per zone
and x_3 = persons over 5 years per household per zone.

Zonal total and zonal rate models are similar, in that both seek to explain the variation in trip-making between zones. No account can be taken of the variation in trip-making within the zones, since the least-squares algorithm has only zonal values, either total or rate, to analyse.

Obviously, all zonal total variables must reflect the effect of zone size. Douglas and Lewis[9] have found strong indications of a linear correlation between the standard error of the residuals and the number of households per zone, from a number of models developed using zonal total variables. This means that a zonal total model will probably violate the least-squares assumption of constant error variance, and the use of zonal rate variables is recommended in preference to zonal total variables.

Kassof and Deutschman[22] also recommend the use of rate variables in preference to total variables, since the rate variables are less strictly tied to the particular geographic system of data aggregation.

The use of rate variables in a least-squares analysis does not guarantee that the assumption of constant error variance will not be violated; however, the assumption is more likely to be satisfied when rate, rather than total, variables are used.

2.3.3 Household Analysis of Residential Trip Generation

In this approach each household is considered as a separate observation. This means that the variations in household characteristics and trip-making are fully incorporated into the model, and hence the model must attempt to explain the variation in trip-making between households, as against explaining the variation in trip-making between zones - which is what the zonal models attempt to do. Examples of the variables used in a household analysis are trips per household, persons per household, etc.

2.3.4 Zonal Analysis versus Household Analysis

In a survey carried out in the United Kingdom in 1967, O'Sullivan[23] found that all the trip generation models that had been developed up to that time were established at the zonal level - in other words aggregation had preceded analysis.

O'Sullivan comments on his survey as follows, "I have no knowledge of the basis of the zonal delineation of those studies, but the variations in the constant and coefficients and the signs of these for equations explaining similar dependent variables in different cities might be accounted for to some extent by differences in zonal size and shape". He goes on to say that the above comments certainly strengthen the case for household trip generation models.

Other researchers[17,18,19,20,21,22] have also noted a disparity in the signs and coefficients of aggregated variables used in the analysis of trip generation.

We may well ask why so much confidence was placed in zonal analysis in the past. The answer to this question is simple. A mis-use of the coefficient of determination led the analysts to the fallacious conclusion that they had developed sound regression equations.

If least-squares regression is carried out at the zonal level, extremely high coefficients of determination can be obtained. These coefficients can be as high as 0,95 when the analysis is carried in terms of zonal total variables. Having obtained such high values for the coefficients of determination, the analysts were satisfied that the regression equations could satisfactorily be used for predicting future trip ends, since they (seemingly) explained a very large percentage of the variation in trip-making observed during the base-year study.

A very clear example of this misconception is provided by the Bombay Traffic and Transportation Study[6] report which states, "The high statistical correlations obtained justified the use of these equations in future projections". This conclusion was reached in connection with trip generation equations developed at the zonal level, for which the coefficient of multiple correlation was found to be of the order of 0,93.

However, the crucial point is that a least-squares regression equation merely endeavours to explain the variation in the dependent variable for the observations analysed. If these observations are at a zonal level, the regression equation can only take account of the variations between zones, and not within zones.

Obviously, if all the zones are very nearly homogeneous, then the within-zone variation will be small when compared with the between-zone variation. However, to obtain zones which are very nearly homogeneous one would have to use very small zones, and therefore the amount of information to be collected increases tremendously, since one needs to collect enough data in each zone for it to be representative of the zone.

The collection of larger amounts of information increases the cost of the home interview survey and may place excessive demands on computer facilities. Notice that, in the limit, as we consider smaller and smaller traffic analysis zones, we are left with the individual households.

Fleet and Robertson[21] have investigated the degree of homogeneity within the zones, by comparing the standard deviation of a particular variable (car ownership) within each of 270 traffic analysis zones, to the standard deviation of this variable throughout the study area. In eighteen per cent of the zones, the standard deviation of the car ownership variable was found to be as large as, or larger than, the standard deviation of the same variable for the whole study area. Hence a considerable degree of variation exists within the zones.

McCarthy[17] has extended this idea and has derived a 'critical relative homogeneity index'. Any zone having a relative homogeneity index greater than this critical value is considered to be heterogeneous. After analysing data gathered in an origin-destination survey in Raleigh, North Carolina, McCarthy[17] concluded that a large enough number of zones showed a significant degree of non-homogeneity, and hence the assumption of zonal homogeneity can be generally refuted.

If most of the variation in trip-making is within the zones, then we lose much information by aggregating the household data to the zonal level. Fleet and Robertson[21] have used a one-way analysis of variance program in order to partition the total variation into that attributable to difference between groups, and that occurring within each group.

The total variation was expressed in terms of the sum of the squared deviations about a mean value. Mathematically, this is the same as the total sum of squares described in Appendix 1, and can be expressed as

$$\sum_{j=1}^{g} \sum_{i=1}^{n_j} (X_{ij} - \bar{X})^2 \qquad (2.5)$$

where X_{ij} = i^{th} observation on some variable (X), taken in the j^{th} group,

$\bar{X}$ = overall mean of variable X,

n_j = the total number of observations in the j^{th} group,

g = number of groups.

When the data are grouped, the total sum of squares can be divided into two components. These two components are called the 'between sum of squares' and the 'within sum of squares'. Mathematically, these components may be expressed as follows:-

$$\text{Between Sum of Squares} = \sum_{j=1}^{g} n_j (\bar{X}_j - \bar{X})^2 \qquad (2.6)$$

$$\text{Within Sum of Squares} = \sum_{j=1}^{g} \sum_{i=1}^{n_j} (X_{ij} - \bar{X}_j)^2 \qquad (2.7)$$

where $\bar{X}_j$ is the mean of variable X for observations in the jth group.

These two components can be interpreted as measures of the between and within group variations, respectively.

Fleet and Robertson[21] found that the within sum of squares was more than 79 per cent of the total sum of squares, when the data from 5 255 households was aggregated to 247 zones. This means that a regression analysis at the zonal level would attempt to explain only about 20 per cent of the between household variation in trip-making.

Least-squares regression equations at the household level have mainly been developed in research work and not in operational transportation studies. However, relatively low coefficients of determination (R^2), say between 0,30 and 0,40; have been obtained at the household level. Although this is much lower than the R^2 values found at the zonal level (as high as 0,95), we can now appreciate the reason for this occurrence.

In order to make a valid comparison we must use the household equation to estimate zonal trip ends. Fleet and Robertson[21] carried out such an analysis and their results are reported in Table 4.7, where they are compared with the results obtained in this study. It will be seen that when the household equation is applied at the zonal level, the household and zonal equations produce comparable results in terms of the coefficient of determination and the standard error of estimate.

When one considers that the output required from the trip generation phase of the traffic forecasting process is an estimate of zonal trip ends, the discussion above demonstrates that the analysis should be carried out at the household level, and the household equation can subsequently be used to estimate zonal trip ends. In other words, aggregation should take place after analysis.

Since the household equation has to explain a much larger variation in the dependent variable, it is very reasonable to believe that the household equation is more likely to remain stable with time.

In spite of the above quantative discussion, probably the most compelling argument for analysing trip generation at the household level is the fact that the household is a much more logical unit of trip-making than the zone. Walker[24] summarized the question of household trip generation analysis when he wrote, ".... the household is the generator of trip productions; the characteristics of its members determine the types and amounts of trips produced".

A further problem is often encountered if one analyses trip generation at the zonal level, in that the simple correlation between aggregated variables has been found to be generally higher than the simple correlation between the same variables before aggregation[7].

In Chapter IV of this thesis the author will investigate the effect of aggregating the household data, collected in Cape Town, to the zonal level.

2.3.5 Analysis at the Personal Level

At this point in time not much consideration has been given to the possibility of analysing residential trip generation at the personal level. However, the author feels that since one usually studies the number of individual trips made, perhaps it is worthwhile giving some consideration to the analysis of trip generation at the personal level.

Oi and Shuldiner[2] feel that "...... some trips are clearly initiated by individual household members without regard to the rest of the household. For other trips, such as shopping for groceries or earning a livlihood, the travel decisions probably represent some collective decision. Social-recreation and business trips are probably the outcome of joint decisions. Thus, although the household appears to be the pertinent unit of analysis, cognizance must be taken of the size and composition of its members".

McCarthy[17] has reported that a study by the Bureau of Public Roads[25] indicates that, "..... although the individual is, in essence, the basic trip-making unit, the magnitude of the unexplained variation that exists in the individual's travel behaviour makes trip generation analysis impractical at this level".

We have already pointed out that analysis at the household level appears to leave a large unexplained variation, when compared with analysis at the zonal level, although this can be shown to be untrue. Apparently the same fallacy is to be found in the above statement.

Probably, the coefficients of determination at the personal level will be found to be relatively low; however, a full investigation is warranted before the personal level of analysis can be dismissed, or otherwise. In Chapter IV the author presents a detailed investigation to determine whether the personal level of analysis is worthy of consideration in future trip generation studies.

2.4 Variables used in the Analysis of Residential Trip Generation

2.4.1 Introduction

We have mentioned that the variables employed in the analysis of trip generation may be divided into three groups, as follows:-

(a) Socio-economic.
(b) Locational.
(d) Land-Use.

In this section of the report we propose to discuss some of the explanatory variables that have been used in previous studies of residential trip generation. Since the household socio-economic characteristics have dominated the scene, the reader will find that the following discussion is mainly devoted to this type of variable.

2.4.2 Car Ownership

One of the most important factors influencing the trip generation of a household is the number of cars that members of the household have available for their use. Even if the car is not owned by the members of the household, but is available for their use, it will influence the trip generation rate of the household.

The private motor car opens up many opportunities not generally available to people confined to using public transport only,[26] particularly for non-work trips, and for trips not directed towards the C.B.D. This is because public transport, where it is available, is mainly geared to serve work trips directed to the C.B.D. For this reason we expect car-driving households to generate more trips than non-car-driving households.

Oi and Shuldiner[2] mention that, " ..... car ownership is the one variable which exhibits the closest association with reported trip generation rates". Although the above comment summarizes the results of analyses at the zonal level, it is likely that similar conclusions will be reached at the household level.

A significant point which must be raised at this stage is the question of whether the relationship between trip generation and car ownership can logically be expected to be linear. All other things being equal, do we expect the acquisition of a second car by a one-car-owning family, to have the same effect on household trip generation, as the acquisition of a first car by a non-car-owning family?

Results reported by Shuldiner[27] indicate that the intensity of vehicular use, in terms of trips per vehicle, is lower for multi-car-owning households than for single-car-owning households. In other words, the first car has the greatest impact on trip generation rates, and each successive car is driven less intensively than the first one. On the other hand, Stowers and Kanwit[18] found that car ownership was linearly related to the number of trips made per household per day.

From the above discussion it seems necessary to investigate the linearity of the car ownership variable, before using it in a multiple linear least-squares regression analysis.

2.4.3 Household Size

The number of trips made by a household is the result of both communal and individual decisions. Because some trips are the result of communal decisions and needs, we would expect the number of trips per household to increase at a slower rate than the increase in the household size. In addition, larger households typically contain higher proportions of children, which also indicates a non-linear relationship between household trip generation and household size.

As is the case with car ownership, some studies have reported household size to be non-linearly related to household trip generation[2], whereas other studies indicate a fairly linear relationship between these variables[18].

Evidently, one needs to examine the linearity assumption for the household size variable, before using it in a multiple linear least-squares regression equation.

2.4.4 Income

The total income of all household members has sometimes been used as an explanatory variable in the analysis of residential trip generation. If we think of travel as an economic good, and we consider the importance of income on consumer demands for nearly all goods, we soon realize that income is likely to be fairly highly correlated with household trip generation.

However, it is important to consider whether it is the household income itself which affects the trip generation rate, or whether the income of the household is found to be fairly highly correlated to trip generation because of its association with car ownership.

It has been found[2] that when car ownership and household size are held constant, the partial correlation between income and trips per household is rather low, although still significant. In other words, most of the effect of income on household trip generation is accounted for by the car ownership and household size variables.

2.4.5 Occupation

In a broad sense, occupation (particularly that of the household head), reflects the level of family income and car ownership, as well as providing an indication of a household's social status.

The problem with occupation, as with income, is that it is fairly highly correlated with car ownership; and it is difficult to isolate the separate effects of the different explanatory variables.

2.4.6 C.B.D. Distance

We have already mentioned that the location of a household within the study area is likely to affect the trip-making behaviour of that household. C.B.D. distance is one variable which has commonly been used to represent the household's location.

In most of the large cities, at least until fairly recently, the central business district was the focal point of economic and social activity. In virtually all studies, trips per household have been found to increase steadily as a function of distance from the C.B.D. However, the inclusion of C.B.D. distance as an explanatory variable, cannot really be justified theoretically. In fact, one is tempted to feel that the reason for the observed correlation between C.B.D. distance and trips per household, is that C.B.D. distance is positively correlated to other variables, particularly car ownership.

If one thinks of the structure of the classical American city, with a high density core consisting of the lower income groups, then one can see why Oi and Shuldiner[2] report that, using the Detroit home interview data, they found that 64 per cent of all families owning no cars resided within 4,8 km of the C.B.D. Only 19 per cent of all multi-car-owning families were found to be living within the same distance of the C.B.D.

Hence it would seem that the observed relationship between C.B.D. distance and trips per household is merely due to the association between C.B.D. distance and other explanatory variables. Stowers and Kanwit[18] found C.B.D. distance to be insignificant in explaining variations in household trip generation rates, when considered simultaneously with other variables (such as family size, automobiles owned, income, etc.).

2.4.7 Residential Density

It is intuitively plausible that residential (or population) density might be an important explanatory factor in the analysis of residential trip generation. One would expect higher vehicular trip rates in low density areas since fewer needs can be satisfied by means of pedestrian trips in such areas.

In spite of this logical explanation, it has been found[2] that, " .... careful analysis fails to isolate density as a significant explanatory variable when the effects of vehicle ownership and family size are taken into consideration". Once again, the inter-correlation among possible explanatory variables causes some problems.

2.4.8 Accessibility

Accessibility to public transport could increase the trip generation rate of zero- and one-car households in particular, while the improvement of the road system would be likely to increase the trip generation rate of car-owning households, particularly multi-car-owning households.

The need to take account of the transport supply when analysing trip generation has been recognized for some time; however, the quantification of an accessibility index presents some serious problems. In the London Traffic Study[3], Freeman, Fox and Partners, used the following index as a measure of public transport accessibility for home-based trips:[28]

$$\text{Bus Accessibility Index} = \sum_i \sqrt{N_{ij}} \Big/ \sqrt{A_j}$$

where N_{ij} = the off-peak frequency of buses on route i and passing through zone j,

A_j = the area of zone j.

A train accessibility index was also included as a locational variable in the analysis of trip generation in London, but was found to have no apparent influence. However, Wooton and Pick[12] have reported that, "..... where the bus index was high it caused trip generation rates to increase slightly, and there was also an apparent transfer of allegiance from rail to buses. The effect of transfer from car to bus is also evident, but probably less than from rail. In any event, these indices suggest that the effects of changes in levels of public transport service are of less importance than the household characteristics".

Apparently the use of accessibility indices in trip generation analysis warrants further research, although their effect appears to be only secondary when compared with household socio-economic characteristics.

2.4.9 Other Explanatory Variables

Other explanatory variables have been used in some studies of residential trip generation. Shuldiner[27], for example, reports the use of so-called 'social area indices'. These indices have been shown to exhibit a strong association with certain aspects of social behaviour, and since urban travel is a manifestation of urban social behaviour, we see that it is reasonable to consider the use of these indices in the analysis of residential trip generation. One of the three indices, the degree of urbanization, was found to be significantly correlated to household trip generation[27].

We have discussed the use of household size as an explanatory variable in the analysis of residential trip generation, but have also pointed out that the characteristics of the members of the household will affect the trip-making behaviour of the household. Wooton and Pick[12] used six household structure types, based on the number of employed and non-employed adults in the household, in their category analysis of the London Traffic Survey data.

The age distribution of the members of a household is also likely to affect the trip generation rate of the household. For this reason a variable called 'stage in the family life cycle' has been investigated[18], but was found to be relatively insignificant when the effect of other variables was accounted for. However, it is likely to be useful to include the number of adults and the number of children separately, instead of the total household size.

2.5 Modelling Trip Generation

2.5.1 Introduction

"Essentially, a model is a representation of reality. It is usually a simplified and generalized statement of what seem to be the most important characteristics of a real-world situation; it is an abstraction from reality which is used to gain conceptual clarity - to reduce the variety and complexity of the real world to a level we can understand and clearly specify"[29].

The extract above, from a book on the use of models in planning, clearly indicates the benefits to be derived from the use of models. The use of mathematical models in transportion planning is relatively new. However, if we consider the size, breadth, cost and complexity of urban transportation to-day; the need for mathematical models, to assist the transportation planner in making important decisions, becomes obvious. As we have pointed out before, the advent of the computer age has greatly assisted in the development of models for transportation planning.

2.5.2 Classes of Models

Lowry[30] has identified three different model classes, namely:

(a) Descriptive.
(b) Predictive.
(c) Planning.

Descriptive models seek merely to replicate the relevant features of an existing phenomenon without trying to account for any cause and effect process. In other words, it may suffice in a descriptive model, to notice that x and y are co-variant. For example, we may find that the equation $y = 2x$ fits our observations very closely.

However, when the aim is to predict the value of y at some time in the future, we require a model which specifies a casual relationship. This means, using the above example, that the model predicts that a one-unit change in x will cause the value of y to change by two units. In this situation, we can predict the future value of y if we know the future value of x.

Hence the predictive model must be built within a logical framework. In other words, the explanatory (or independent) variables used in a predictive model need to be logical causes of changes in the dependent variable.

Planning models are a class whose technology is not far developed, according to Lowry[30]. A planning model not only predicts future events based on certain assumptions, but also evaluates the outcome in terms of the planner's goals. One such type of model is known as linear programming.

From the above discussion, we see that we are interested in predictive models. The most important feature of this type of model, is that it must specify a casual relationship, and for this reason the question of causal effects is discussed in the next section.

2.5.3 Causal Effects

It is extremely important to realize that a regression analysis cannot take into account the question of whether an independent variable has a causal effect on the dependent variable. The regression techniques (least-squares and category analysis), merely assist the researcher to specify a relationship between the dependent variable and the chosen independent variables. In other words, regression techniques are ideally suited for the development of descriptive models.

However, we need not despair. Regression techniques can be used for developing predictive models, as long as the analyst is aware that the onus is on him to choose carefully the explanatory variables employed in the model, and to examine thoroughly the resulting regression model.

In the case of least-squares regression, the analyst must study with care the coefficients (and their signs) in the regression equation. In the case of category analysis the researcher is more limited when examining the results of his calculations, but he can still investigate whether the independent variables appear to have the expected effect on the dependent variable.

The problem of causality is clearly illustrated by the old story of the man who studied the number of storks and the number of new-born babies in a particular town over a period of time. He found that the two variables were highly correlated, and used his results to substantiate the argument that storks bring babies.

The fact that relationships can be found, but not explained, by a regression analysis, is one of the major weaknesses of the technique. Sophisticated computer programs have been developed for multiple least-squares regression analysis, which can result in impressive looking, but illogical equations being developed. However, regression analysis, in the hands of a critical analyst, can be used to good effect.

2.5.4 Ease of Forecasting the Independent Variables

The aim of expressing present day trip generation in terms of various explanatory variables, is to be able to forecast future trip generation rates from forecasted values of the independent variables. Obviously there is no point in expressing trip generation in terms of variables which are more difficult to forecast than the dependent variable itself.

Hence, one of the major requirements of the variables used in a regression analysis, is that they can be forecast fairly easily and reasonably accurately. It is necessary to remember, in this connection, that a regression analysis does not take account of errors in the present, or future, estimates of the independent variables.

2.5.5 Model Building

One of the biggest disadvantages of category analysis, when compared with least-squares regression analysis, is that we have no means of testing the significance of the individual explanatory variables included in the model. Therefore, we cannot build up a model, using only significant explanatory variables, with the category analysis approach; since all the explanatory variables selected by the analyst are automatically included in the model. The following discussion will illustrate the advantage of the least-squares approach, as far as model building is concerned.

Suppose we wish to establish a multiple regression equation for a particular dependent variable (y) in terms of a number of independent variables (x_1, x_2 .... x_k). On the one hand, we would want to include as many of the x's as possible in the model, so that reliable predictions can be made. On the other hand, because of the difficulty, and cost, involved in forecasting future values of a large number of independent variables, we should like to include as few x's as possible in the regression.

A compromise has to be found between these extremes, and this process is usually termed 'selecting the best regression equation'. No unique statistical method is available, and inevitably personal judgement will be a necessary part of any selection procedure.

Draper and Smith[8] discuss a number of statistical procedures which have been developed to aid the analyst in selecting the best regression equation. These methods are:

(a) All possible regressions.
(b) Backward elimination.
(c) Forward selection.
(d) Stepwise regression.
(e) Stagewise regression.

It suffices to report here that Draper and Smith conclude that, as a practical regression method, the stepwise procedure is the most suitable. However, in a theoretical sense, the all-regressions procedure is best, in that all possible combinations of the independent variables are investigated.

A number of sophisticated computer programs for stepwise regression are available. In the stepwise procedure, the regression equation is built up by successively adding explanatory variables. The variable added at any step is the one which produces the greatest reduction in the residual (or unexplained) sum of squares. At every step the significance of each variable is checked, and only variables which are significant are added to the regression set; while variables already in the regression equation, which become insignificant, are removed.

The stepwise regression procedure will be discussed in more detail when the stepwise regression program used in this study is outlined (see 4.3.2).

2.5.6 The Problem of Multi-Collinearity

Multi-collinearity, "..... is the name given to the general problem which arises when some or all of the explanatory variables in a relation are so highly correlated one with another that it becomes very difficult, if not impossible, to disentangle their separate influences and obtain a reasonably precise estimate of their relative effects"[31].

The fact that the least-squares regression procedure breaks down completely in the case of perfect correlation between two or more variables, has already been mentioned. In addition, the author feels that the problem of multi-collinearity should not be forgotten when using the category analysis technique.

The practical interpretation of multi-collinearity is that two or more independent variables will be trying to explain largely the same effect. The individual effects of such explanatory variables on the dependent variable will probably be obscured if all enter a regression equation.

Unfortunately, there are no generally accepted rules for deciding when the multi-collinearity is severe enough to warrant action, or what action to take, in cases where it is considered to be bad enough.

Douglas and Lewis[15], point out that it has been suggested that if the correlation between two independent variables exceeds their individual correlations with the dependent variable, then one of the independent variables should be omitted from further consideration. If one of the variables is easier to forecast, or more meaningful, then the other variable should be omitted.

Fortunately, if two variables are highly correlated then it is unlikely that both will enter the regression equation, if a stepwise procedure is used.

It is relevant to note here that car ownership and household income have sometimes been included as explanatory variables in models for estimating trip generation from the home. There is obviously a fairly high correlation between these two variables, for present day conditions, and the inclusion of both of them in a regression model can certainly be questioned.

2.6 Summary

We have examined the rationale of trip generation analysis, and have seen that the most important factor is that travel is an aspect of derived demand. In addition, it is necessary to remember that the assumption, that trip generation is independent of the available transport service and opportunities, is a possible weakness of the rationale.

The two basic methods of analysis were discussed, and the advantages and disadvantages of each were outlined. We have seen that the category analysis approach is an extremely simple concept, but is deficient as a model-building technique, and also makes large demands on the data collection phase.

On the other hand, the least-squares regression technique, while requiring that certain assumptions be met, is a powerful model-building tool when used intelligently; and we have seen that the dummy variable technique can be used to overcome some of the difficulties involved in the least-squares method.

It should be emphasized that the dissatisfaction that arose with the least-squares regression technique led to the introduction of the category analysis procedure. However, the poor results were not due to the least-squares method itself, but because of its mis-use. Douglas[20] points out that instead of rejecting zonal least-squares regression in favour of household least-squares regression, the transportation planners rejected the least-squares approach in favour of the category analysis technique.

The question of the most suitable unit to be used in the analysis of residential trip generation has been discussed in some detail. The zonal level of analysis will not be used in future studies. Not much has been reported about the personal level of analysis, but a reasonable amount of information is becoming available about studies at the household level. Most researchers suggest that the household is the best unit to use in the analysis and model-building phase.

We have pointed out the need for trip generation models to be developed within a logical framework, if they are to be used for forecasting future trip generation rates. For this reason, we have had a brief look at some of the variables that have been used in residential trip generation analysis. In addition, we have pointed out that some of the commonly used explanatory variables may well be non-linearly related to the dependent variable.

In order to investigate some of the questions raised in this chapter, and to gather some of the data needed for transportation planning in Cape Town, the author decided to carry out an analysis of residential trip generation in Cape Town. The collection of the necessary data is discussed in Chapter III, while in Chapter IV the analysis and model-building phase is reported.

NOTES - CHAPTER II

1. Worral, R.D., Trip Generation (Notes for a Short Course in Urban Transportation Planning at the University of Stellenbosch, January 1972).

2. Oi, W. and Shuldiner, P.W., An Analysis of Urban Travel Demands (Northwestern University Press, 1962).

3. Reported by H.J. Wooton and G.W. Pick - see Note 12.

4. Blunden, W.R., "Introduction to Traffic Science - 1. Basic Concepts of Interaction", Traffic Engineering and Control, February 1967.

5. Domencich, T.A., Kraft, G. and Vallette, J-P., "Estimation of Urban Passenger Travel Behaviour: An Economic Demand Model", Highway Research Record, Number 238, 1968.

6. Bombay Traffic and Transportation Study, Wilbur Smith and Associates, 1963.

7. Douglas, A.A. and Lewis, R.J., "Trip Generation Techniques - 1. Introduction", Traffic Engineering and Control, November 1970.

8. Draper, N.R. and Smith, H., Applied Regression Analysis, Wiley, 1966.

9. Douglas, A.A. and Lewis, R.J., "Trip Generation Techniques - 2. Zonal Least-Squares Regression Analysis", Traffic Engineering and Control, December 1970.

10. Winsten, C.B., "Regression Analysis versus Category Analysis", P.T.R.C. Seminar, September 1967.

11. Walker, J.R., "Rank Classification: A Procedure for Determining Future Trip Ends", Highway Research Record, Number 240, 1968.

12. Wooton, H.J. and Pick, G.W., "A Model for Trips Generated by Households", Journal of Transport Economics and Policy, 1967.

13. Douglas, A.A. and Lewis, R.J., "Trip Generation Techniques - 4. Category Analysis and Summary of Trip Generation Techniques", Traffic Engineering and Control, February 1971.

14. Suits, D.B., "Use of Dummy Variables in Regression Equations", Journal of the American Statistical Association, December 1957.

15. Douglas, A.A. and Lewis, R.J., "Trip Generation Techniques - 3. Household Least-Squares Regression Analysis", Traffic Engineering and Control, January 1971.

16. Heathington, K.W. and Isibor, E., "The Use of Dummy Variables in Trip Generation Analysis", Transportation Research, Vol. 6, 1972.

17. McCarthy, G.M., "Multiple Regression Analysis of Household Trip Generation - A Critique", Highway Research Record, Number 297, 1969.

18. Stowers, J.R. and Kanwit, E.L., "The Use of Behavioural Surveys in Forecasting Transportation Requirements", Highway Research Record, Number 106, 1966.

19. Brant, A.E. and Low, D.E., "Cost-Saving Techniques for Collection and Analysis of Origin-Destination Survey Data", Highway Research Record, Number 205, 1967.

20. Douglas, A.A., "Home-Based Trip End Models - A Comparison between Category Analysis and Regression Analysis Procedures", Transportation, Vol. 2, 1973.

21. Fleet, C.R. and Robertson, S.R., "Trip Generation in the Transportation Planning Process", Highway Research Record, Number 240, 1968.

22. Kassoff, H. and Deutschman, H.D., "Trip Generation: A Critical Appraisal", Highway Research Record, Number 297, 1969.

23. O'Sullivan, P., "Review (and List) of Home Trip Estimating Equations Derived from Traffic Surveys in U.K.", P.T.R.C. Seminar, September 1967.

24. Walker, J.R., "Social Status of Head of Household and Trip Generation from Home", *Highway Research Record*, Number 114, 1965.

25. *Guideline for Trip Generation Analysis*, U.S. Bureau of Public Roads, 1967.

26. Such a person is generally termed a 'transit captive'.

27. Shuldiner, P.W., "Trip Generation and the Home", *Highway Research Board Bulletin*, Number 347, 1962.

28. A home-based trip is one which begins or ends at the home of the person making the trip.

29. Lee, Colin, *Models in Planning - An Introduction to the Use of Quantitative Models in Planning*, Pergamon Press, 1973.

30. Lowry, I.S., "A Short Course in Model Design", *Journal of the American Institute of Planners*, May 1965.

31. Johnston, J., *Econometric Methods*, McGraw-Hill, 1963.

CHAPTER III

DATA COLLECTION

3.1 Introduction

In order to carry out an analysis of residential trip generation in Cape Town, information had to be gathered about the trip-making and socio-economic characteristics of households[1] in the Cape Town area. This chapter describes:

(i) The selection of the home questionnaire technique as the method of data collection.

(ii) The definition of the study area.

(iii) The selection of the sample households.

(iv) The design of the survey forms and the pre-interview letter.

(v) The pre-test survey.

(vi) The Cape Town home questionnaire survey.

3.2 Methods of Data Collection

3.2.1 Introduction

Together with the development of a philosophy and methodology for the planning of urban transportation systems, a number of techniques have been developed for the collection of the basic data required by the various models used in the urban transportation planning process.

One of the most expensive and time-consuming aspects of the data collection phase of a land-use/transportation study is the collection of socio-economic and trip-making information for a sample of households in the survey area, but this data is essential for a study of residential trip generation. In addition, information regarding the household's location may be needed, but this can be obtained from sources external to the sample households.

3.2.2 The Home Interview Survey

The technique commonly used for the collection of the household information is known as the home interview survey.[2] In this method, interviewers visit a sample of households in the survey area within two days of the day for which travel information is required. Usually each sample household receives an introductory letter before the visit of the interviewer. The relevant information for each sample household is gathered by interviewing those members of the household present at the time of the interview. Generally, trips made by all residents five years of age or older are recorded. Each interviewer is expected to complete eight interviews in an eight hour day.

3.2.3 Modifications to the Home Interview Survey

In most of the transportation studies carried out in the U.S.A. the above technique was applied with very little or no modification.

One modification which has been used is called the pre-interview technique.[2] In this method, the interviewers visit the sample households before, and after, the survey day. The aim of the pre-survey day visit is to obtain more complete and accurate travel information.

On the first visit the interviewer supplies trip cards on which the respondents are asked to record the trips they make on the survey day. The interviewer arranges to return at a mutually acceptable time within three days of the survey day.

On the second visit the interviewer studies the trip cards completed by members of the household and transfers this information, together with additional household and personal information, to the standard interview form.

Another modification, known as the intensive interviewing technique[2], has been tried with some success. In this methods all persons over the age of 16 years are interviewed, either when the interviewer first visits the household (after the survey day); or at additional call-backs, either personally or by telephone.

Other techniques have been used with some degree of success, these are telephone and post-card surveys[2]. In both these cases, the comprehensiveness of the information obtained is limited by the method.

3.2.4 Newspaper Survey

Consideration was given to the use of one or more newspapers for the distribution of a questionnaire. The advantages of such an approach are that a very large number of questionnaires can be distributed with very little effort and at a reasonable cost, assuming the co-operation of the newspapers could be obtained.

The main reason that this approach was not pursued was the realization that such a survey would produce a very low response rate, and that replies would be received from a rather biased sample of the households in the area.

3.2.5 Home Questionnaire Survey

A method based on the home interview technique has been successfully used in three urban areas in Britain[3]. This technique has been termed the 'home questionnaire survey'. The method was first employed by the Christchurch Regional Planning Authority in 1959. The essential difference between this technique and the usual home interview procedure, is that in the home questionnaire survey the number of verbal questions is kept to a minimum.

As with the home interview survey, the first step in this procedure is the selection of the sample households. Following this, an introductory letter is mailed to each sample household about a week prior to the survey. An interviewer visits the sample households and delivers the questionnaire forms to the members of the household by obtaining the co-operation of at least one responsible member of the household. After explaining the requirements of the survey, and asking a few questions (so as to get the respondents started on the questionnaire), the interviewer requests the members of the household to record all the trips they make on a specific day following the delivery of the forms. The interviewer arranges to collect the completed forms at a mutually acceptable time within two days of the specified survey day.

When collecting the forms the interviewer has the opportunity of checking them to see if they have been correctly completed. If not, he is able to assist in the correct completion of the questionnaire. A rate of fifty interviews per interviewer per four-day working week has been suggested as an average rate[3]. The two days prior to the survey day are used for distributing the questionnaires, while the two days following the survey day are used for collecting the completed forms.

3.2.6 Supply of Manpower

In the search for a suitable method of data collection, an important factor was the lack of resources available to the author for this project.

Fortunately, seven final year civil engineering students were made available to the author for assistance with this study. The work was to be done as part of the final year thesis project. Because of the nature of the undergraduate thesis project, each student would be able to collect data for about five weeks.

As a result, a method of data collection was needed which would obtain as much information as possible in a period of approximately five weeks.

3.2.7 Day-to-Day Variation in Trip-Making.

Another factor which was considered is the day-to-day variation in trip-making of a given household.[4] Since not much is known about this variation, we realized that a very large amount of data would have to be collected if this variation were to be taken into account. Therefore we decided to choose an assumed typical weekday and to collect trip-making information relevant to only this day.

3.2.8 Choice of Method

Having decided that the trip-making data would be collected for a particular weekday, and remembering that a fixed period of time would be available for the field work, we had to choose a method which would enable the maximum amount of data to be collected within these constraints.

Using the home interview method, each interviewer would be able to interview approximately eighty households during the available five week period. Using the pre-interview technique, this figure could perhaps have been improved to approximately one hundred and fifty interviews per interviewer.

However, using the home questionnaire method, each interviewer could carry out approximately 250 interviews during the five week period. Hence this method was selected for the collection of the trip-making and socio-economic data needed for the analysis of residential trip generation in

Cape Town.

After much consideration, Thursday was chosen as the survey day to be used in this study. Monday, Tuesday and Wednesday would be available for distributing the questionnaires, and Friday and Saturday could be used for the collection of the completed forms.

3.3 Definition of the Study Area

We have already mentioned that one of the factors contributing to the urban transportation problem is the growth of metropolitan areas encompassing a number of local authorities.

Ideally, this study should have covered the whole of the Cape Metropolitan Area. However, because of the limited resources available to the author, the survey had to be restricted to a reasonably small area, containing a relatively small population.

An important factor which was considered in the definition of the study area, is that the generally lower socio-economic coloured and black population groups live in areas poorly served by public transport. Since these people are mostly dependent on public transport, the author believes that there is a 'latent demand'[5] for travel amongst these people. Further, as we have already pointed out, the movement patterns of the coloured and black population groups are much influenced by government policy.

As a result of the above considerations, it was decided that this study be limited to the white population group. However, we realize that as increasing standards of living are attained by the coloured and black people, an enormous transportation problem will develop in a fairly confined area on the Cape Flats. For this reason, the author hopes to carry out a study of transportation needs on the Cape Flats in the near future.

The area chosen for the present study incorporates part of the Cape Town municipal area, the Pinelands municipality, and part of the Divisional Council area in Constantia. A map of the study area is shown in Figure 3.1.

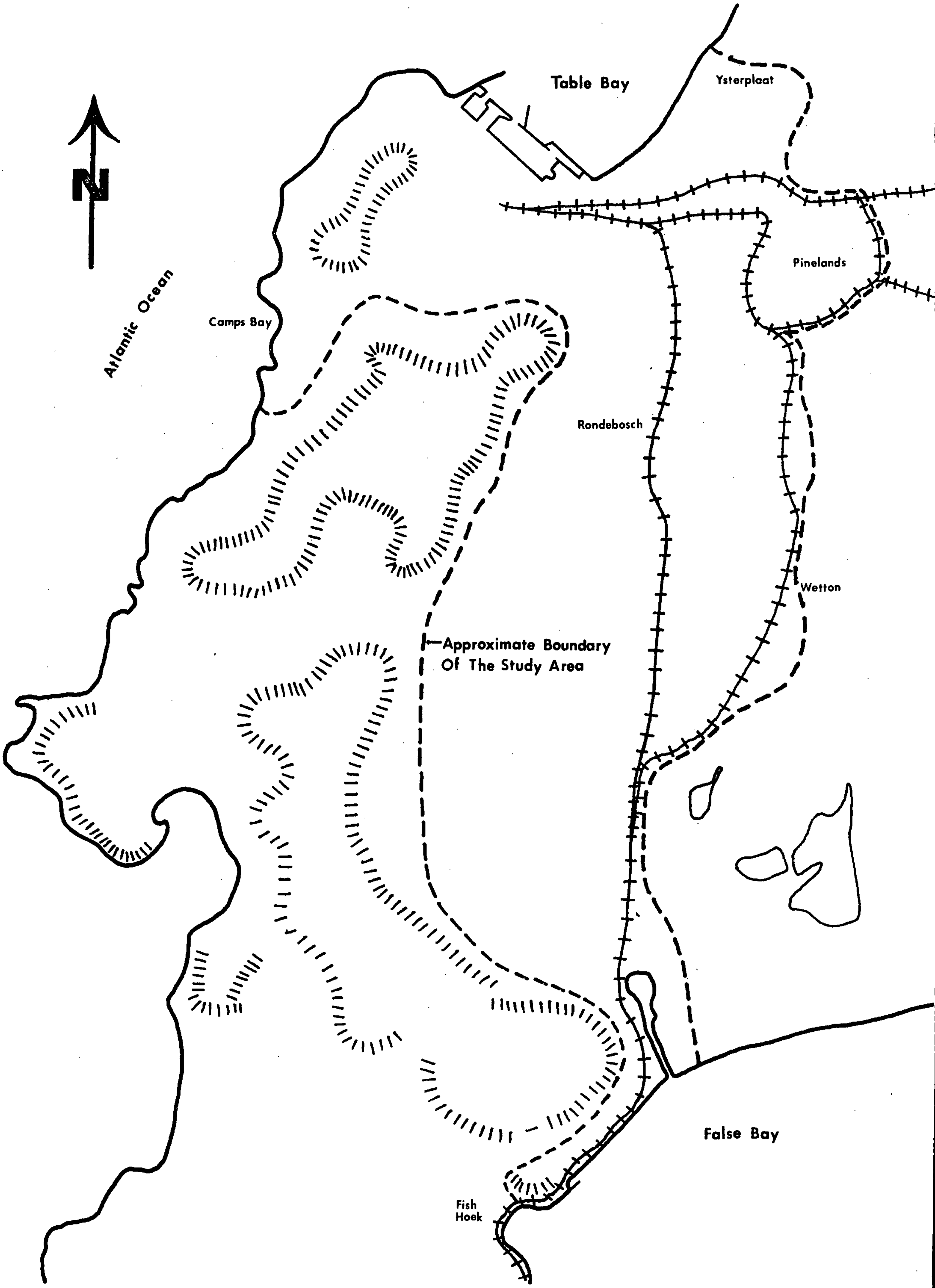


FIG. 3.1 The Study Area

3.4 Selection of Sample Addresses

3.4.1 Introduction

Initially we intended to use the record of addresses kept by the City Treasurer's Department (Cape Town) for the mailing of electricity bills, as the sampling frame[6]. After much effort to obtain this information, we were informed that a decision taken by the City Council, some years ago, forbids the use of these records for any purpose other than the mailing of electricity bills to consumers.

In addition to supplying a list of addresses from which the sample addresses could be selected, we had hoped that the electricity consumers' list would provide us with the electricity consumption of the sample households. This could be used as a measure of the socio-economic standing of each household. Intuitively, household income and electricity consumption should be fairly highly correlated.

The use of the local telephone directory as a sampling frame was considered unsuitable, since this would have introduced a bias into the sample, as a significant proportion of lower socio-economic households do not possess a telephone.

The Cape Times Peninsula Directory[7] was selected as the most suitable source of the information required for the selection of the sample households. In addition to the Cape Times Directory, a number of publications of the Department of Statistics[8] were used. These publications, based on the 1970 census of population, supplied the author with the number of houses, flats and people in each of thirty seven geographical areas within the survey area.

3.4.2 Sampling Techniques

The need for choosing a representative sample, as opposed to interviewing all the households in the survey area, is fairly obvious. A number of different sampling methods are available, and each is suitable in certain situations. The following sampling techniques are commonly used[9]:-

(i) Random Sampling.
(ii) Stratified Sampling
(iii) Systematic Sampling

(iv) Proportional Sampling.
(v) Cluster Sampling.

The use of more than one of the above techniques, in a sequential sampling process, is known as multi-stage sampling. This latter method was employed in selecting the households to be surveyed in this study.

3.4.3 Sample Size

In designing the method of sample selection, we were aiming at a sample which would be a representative cross-section of the socio-economic and locational characteristics of the households in the study area. In order to achieve this aim, we first clustered the households in the survey area into the thirty seven geographical areas used in the Department of Statistics' publications. We then stratified the households in each of these areas into two groups - houses and flats. The idea was to select a proportional sample of houses and flats (considered separately) from each of these thirty seven suburbs.

Table 3.1 (see page 56) illustrates the calculation of the number of households to be sampled in each suburb.

In order to calculate the total sample size $(N_H + N_F)$, we first estimated the total number of households in the survey area (in 1973) by assuming a 2% annual growth rate[10], i.e.

$$T_{1973} = T_{1970} \left(1 + \frac{2}{100}\right)^3$$

where T_{1973} = total number of households in the survey area in 1973,

$T_{1970} = T_H + T_F$ = total number of households in the survey area in 1970.

Now $T_{1970} = 70\ 445$

$\therefore T_{1973} = 74\ 760$

The following sample sizes, to be adopted in a home-interview study, are given by Bruton[11]. (Table 3.2)

Table 3.1

Suburb	No. of Houses	No. of Flats	Proportion of Houses	Proportion of Flats	No. of Sample Houses	No. of Sample Flats
1	H_1	F_1	P_{H1}	P_{F1}	N_{H1}	N_{F1}
2	H_2	F_2	P_{H2}	P_{F2}	N_{H2}	N_{F2}
3	H_3	F_3	P_{H3}	P_{F3}	N_{H3}	N_{F3}
.						
.						
i	H_i	F_i	P_{Hi}	P_{Fi}	N_{Hi}	N_{Fi}
.						
.						
.						
37	H_{37}	F_{37}	P_{H37}	P_{F37}	N_{H37}	N_{F37}
	$\sum_{i=1}^{37} H_i = T_H$	$\sum_{i=1}^{37} F_i = T_F$	$\sum_{i=1}^{37} P_{Hi} = 1$	$\sum_{i=1}^{37} P_{Fi} = 1$	$\sum_{i=1}^{37} N_{Hi} = N_H$	$\sum_{i=1}^{37} N_{Fi} = N_F$

where H_i = number of houses in suburb i (1970)

F_i = number of flats in suburb i (1970)

P_{Hi} = proportion of houses in suburb i (1970)

$= H_i/T_H$

P_{Fi} = proportion of flats in suburb i (1970)

$= F_i/T_F$

T_H = total number of houses in the survey area (1970)

T_F = total number of flats in the survey area (1970)

N_{Hi} = number of sample houses in suburb i (1973)

$= P_{Hi} \times N_H$

N_{Fi} = number of sample flats in suburb i (1973)

$= P_{Fi} \times N_F$

N_H = total number of sample houses in survey area (1973)

N_F = total number of sample flats in survey area (1973)

Population of Study Area	Recommended (Dwelling Units)	Minimum (Dwelling Units)
Under 50 000	1 in 5	1 in 10
50 000 - 150 000	1 in 8	1 in 20
150 000 - 300 000	1 in 10	1 in 35
300 000 - 500 000	1 in 15	1 in 50
500 000 - 1 000 000	1 in 20	1 in 70
More than 1 000 000	1 in 25	1 in 100

Table 3.2

The population of the survey area in 1973 was approximately a quarter million people. Using the minimum sampling fraction suggested in Table 3.2, we estimated the minimum number of households to be interviewed as $1/35 \times 74760 = 2\ 140$.

3.4.4 Selection of Sample Houses

The first step in the selection of the sample houses was to choose a set of streets for each suburb. The use of cluster sampling at this stage was influenced by two factors. Firstly, the layout of the Cape Times Directory, where house addresses are listed street-wise. Secondly, it was felt that the use of cluster sampling at this stage of the sampling process would make the fieldwork easier for the interviewers, since less travelling would be required than would be the case had individual houses been selected at random.

It was estimated that, on the average, approximately five houses should be surveyed in each selected street[12]. Hence the number of streets to be selected for suburb i was estimated as

$$S_i = \frac{N_{Hi}}{5}$$

where S_i = number of sample streets in suburb i.

For each suburb the required number of streets was selected, using a multi-stage sampling process[13], from the alphabetical list of street names provided by the Cape Times Directory.

The next step was the selection of the sample houses in each of these streets. Using the Cape Times Directory, we determined the total number of houses in the streets chosen for each suburb. This enabled us to determine a sampling fraction for houses in suburb i as

$$\frac{N_{Hi}}{T_{Hi}}$$

where T_{Hi} = total number of houses in the streets selected in suburb i.

Systematic sampling was used to select the sample houses from the chosen streets.

3.4.5 Selection of Sample Flats

The selection of the sample flats followed a slightly different procedure to that used for the houses, since the Cape Times Directory provides only an alphabetical list of residential flats in the area covered by the directory. No list of the occupants, or the numbers of the units in each of these blocks of flats, is provided. Although we did not require a list of the occupants, we did require the total number of flats in each block, as well as the individual unit numbers.

The first step was the selection of the sample blocks of flats. This, as with the first step in the selection of the sample houses, was a cluster sampling process. It was estimated that, on the average, we should survey approximately eight flats per block selected[14]. Hence the number of blocks of flats to be selected in suburb i was estimated as

$$B_i = \frac{N_{Fi}}{8}$$

where B_i = number of survey blocks of flats in suburb i.

One hundred and twenty blocks of flats were chosen, using a multi-stage sampling process[15], and each block was visited by the author so as to obtain a list of the individual flat numbers in each block. This information was used for selecting a systematic sub-sample from each block of flats. The sampling proportion for the selection of this sub-sample was calculated for suburb i as

$$\frac{N_{Fi}}{T_{Fi}}$$

where T_{Fi} = total number of flats in the sample blocks selected in suburb i.

3.4.6 Selected Sample Size

Using the techniques outlined above, a sample of 1 235 houses and 927 flats was selected. The total number of sample households was 2 162, representing approximately 3% of the households in the survey area.

3.5 Design of Survey Forms and Pre-Interview Letter

3.5.1 Introduction

The design of the survey forms and 'Dear Householder' letter is extremely important in a study of the kind reported here. If not enough attention is given to this task, it is likely that a fairly low response rate will be obtained. In addition, the replies that are received would provide misleading information if there were any ambiguities in the questionnaire.

An additional problem associated with surveys in a bilingual country, is the need to prepare all the survey documents (letter, forms, etc.) in both official languages. It is important to ensure that the translations are undertaken by someone who is fully bilingual, so as to obtain two sets of documents which are as similar as possible. For this reason the author obtained assistance with translations needed in this study[16].

3.5.2 The Pre-Interview Letter

The 'Dear Householder' letter, which is sent to the sample households before they are visited by the interviewer, is the first point of contact between the sample households and the survey staff. As such, it has a very important role to play in creating a favourable impression.

Goode and Hatt[17] mention a number of aspects which should be covered by the pre-interview correspondence. Firstly, it should mention the auspices under whom the study is being conducted. Secondly, the aims of the survey should be clearly indicated, while the importance of the response of each respondent should be stressed. The anonymity of the respondents should be guaranteed. Although all the above topics should be adequately covered, the introductory letter must be kept as brief as possible.

Particular care was taken in the design of the introductory letter used in the Cape Town home questionnaire survey. Copies of the English and

Afrikaans versions of the letter, sent to all sample households, can be found in Appendix 3.

Considerable attention was given to the envelope in which the pre-interview correspondence was mailed to the survey households. It was decided that the envelopes be addressed to 'The Householder', since very recent information on the name of the occupants of houses was not available. In addition, it was realized that the use of a name would decrease the respondent's feeling of anonymity.

There is a problem, however, in addressing a letter to 'The Householder'. Many commercial and charitable organizations send mail addressed to 'The Householder'. As a result, there is a general tendency for householders to ignore such mail.

In an attempt to prevent our introductory letter receiving the same fate, we decided to use envelopes bearing the name of the University of Cape Town in both English and Afrikaans. In addition, a special rubber stamp was purchased so as to enable us to mark each envelope with the words "Transportation Research Project" in both English and Afrikaans.

In a further effort to ensure that respondents would pay attention to the pre-interview letter, it was decided that the necessary three cent stamp would be affixed to each envelope, instead of using the franking facility provided by the post office for bulk postage.

Some consideration was given to addressing each envelope by hand, or having each address typed on the envelope. However, it was felt that since various lists of addresses would be needed during the course of the survey, it would be wise to have the sample household addresses punched on computer cards.

A computer program was written by the author in order to print the address labels, as well as for sorting the addresses into various groups as required. An example of the envelope and address label is shown in Appendix 3.

3.5.3 The Survey Forms

The basic requirements of a questionnaire are that the questions should be unambiguous and as simple as possible. The term 'questionnaire' usually refers to a device for securing answers to questions using a form which the respondent completes himself[18]. In this study, some of the answers were recorded by the interviewer; however, we will sometimes use the term questionnaire when referring to the survey forms.

When carrying out a survey of some magnitude, the temptation always arises to collect more information than is really necessary for the study at hand. The reasoning being that while the survey is being carried out, not much extra cost will be incurred by the collection of some additional information. On the other hand, one must ensure that all the relevant information is collected. We must admit that some additional information was collected in this survey; however, it is hoped that use will be made of this data in the near future.

Before designing a questionnaire, one needs to consider the information required. Basically, we needed the household's socio-economic characteristics, as well as details of the trips made by members of the household.

The socio-economic characteristics of the sample households were obtained using a questionnaire entitled 'Household Information - Form A' (see Appendix 3). This form determined the number of people in each of three age groups, the number of cars available for use by the members of the household[19], the number of school children, etc. No information regarding income was requested as it was felt that such a question would considerably lower the response rate. However, the occupation of each of the employed people was requested, and it was anticipated that the occupation information would be used instead of income data. In addition, we have already noted that car ownership and income are highly correlated.

Each member of the sample households, five years of age or older, received one copy of Form 'B' and the Trip Record Form (see Appendix 3). These two forms were stapled together in the home questionnaire survey. Form 'B' was designed to obtain information relevant to each person for whom trip information was requested. The aim was to be able to carry out an analysis of trip generation at the personal level, in addition to the household and zonal analyses.

In a home questionnaire survey the most difficult information to collect accurately and completely is the trip-making data. For this reason, great care was taken by the author when designing the layout and contents of the trip record form. Since a typewriter could not give the variation in size and spacing of lettering desired, each letter was individually set by the author using a special lettering set[20].

The purpose categories to be used on the trip record form was the major problem encountered in the design of this form. From a brief look at the literature, it was obvious that in most transportation studies too many purpose categories had been specified at the survey stage. At the analysis stage some of these categories had to be aggregated[21].

In this study it was decided to use six different trip purpose categories:

(i) Work.
(ii) Shopping.
(iii) Return Home.
(iv) Education.
(v) Pleasure (social - recreation).
(vi) Other.

It was anticipated that a fairly large number of the trips in the 'other' category would be non-home-based and therefore not of vital interest to this study.

Since we only intended analysing the trips made by motorized forms of transport, the six mode categories were easily identified as follows:

(i) Car driver.
(ii) Car passenger.
(iii) Bus.
(iv) Train.
(v) Taxi.
(vi) Motor cycle.

In order to be able to analyse separately the trips made during the peak and off-peak periods, three alternative time periods were presented to the respondents on the trip record form. These were:

(i) 7.00 a.m. - 9.00 a.m.
(ii) 4.30 p.m. - 6.30 p.m.
(iii) Rest of Day.

Some people, such as doctors, taxi-drivers, service-men, etc., make many non-home-based trips during a typical week-day. The intention was to ask such people to record the full details of all their home-based trips, but only to record the total number of the numerous non-home-based trips they made on the survey day.

All the survey forms were carefully translated, and Afrikaans copies can be found in Appendix 3.

3.6 Pre-Test Survey

However much care the researcher may take when designing the content and layout of a questionnaire, it is still possible for ambiguities to exist. For this reason, as well as for checking on the planned survey procedure, the use of a pre-test survey has been strongly recommended[22].

Since the aim of the pre-test is to detect any possible flaws in the questionnaire or the survey procedure, the pre-test must be carried out early enough (relative to the proposed start of the actual survey), so that any necessary modifications can be made in good time. The proposed starting date for the home questionnaire survey was Monday, September 10th, 1973. The pre-test survey was carried out between Tuesday, August 14th and Saturday, August 18th.

For the pre-test survey a sample of sixteen households was chosen, eight houses and eight flats. These households were subjectively picked from two streets and a block of flats. All the households selected were in the Gardens and Oranjezicht areas, but they were chosen so as to cover the range of socio-economic characteristics expected in the full-scale survey. This was essential, since we wanted to examine what problems would arise with the different types of people we expected to interview in the home questionnaire survey.

The introductory letters were mailed to the pre-test survey sample households on Thursday, August 10th. On Tuesday, August 14th the author visited all sixteen households. One of the selected houses was found to be vacant - this was established after consultation with the people living next door. Contact was made with a responsible member in each of the other fifteen households.

All the people contacted were most co-operative, and completed questionnaires were collected from all fifteen households. The good response obtained during the pre-test survey was most encouraging. The author feels that this was largely due to the fact that most people contacted had either read the pre-survey letter, or had been informed of it by some other member of the household. The fact that the survey was part of a research project assisted in the attainment of a good response.

When collecting the completed questionnaires the author noted that most of the respondents had found the forms easy and quick to complete and very few had encountered any major difficulties in completing the forms. The replies were examined and it appeared that the questionnaires had been satisfactorily completed.

The author was very satisfied with the pre-test survey, both from the point of view of the response obtained, as well as from the fact that the survey forms appeared to be suitable for use in the actual survey, after one or two minor corrections were made.

One factor which was disturbing, was the fact that all the people in the pre-test survey spoke English and this meant that the Afrikaans forms were not used at all during the pre-test survey, and had therefore not been put to the test. However, it was not possible to carry out another pre-test survey, and we had to rely on the quality of the translation as far as the Afrikaans questionnaires were concerned.

3.7 <u>The Cape Town Home Questionnaire Survey</u>

3.7.1 <u>Introduction</u>

The home questionnaire survey was carried out between Monday, September 10th and Saturday, October 27th, 1973. Of the seven Thursdays which occurred between these dates, six were used for the recording of trip-making by residents of the survey area. The seventh, (Thursday, September 27th) was not used, since both the school and university vacations occurred this week, and it was felt that trip-making information collected during this week would not correspond to normal weekday behaviour.

Hence the trip-making data used in this study applies to the following Thursdays:

13th September, 1973
20th September, 1973
4th October, 1973
11th October, 1973
18th October, 1973
25th October, 1973

Thursday, October 11th, was the day following Kruger day[23], but it was assumed that this would not have much effect on the trip-making behaviour of the respondents.

In an attempt to improve the response obtained from the sample households, the author tried to obtain some publicity for the project. Contact was made with The Cape Times and Die Burger, and the author was interviewed by reporters representing these newspapers. A report about the project was published in The Cape Times on the Saturday prior to the commencement of the survey, but apparently no report appeared in Die Burger. For a copy of the published report see Appendix 3.

3.7.2 Administration of the Survey

At the beginning of the home questionnaire survey, a set of instructions was issued to each of the seven interviewers. A copy of these instructions is to be found in Appendix 3.

Each interviewer was allocated a list of addresses which he was to visit during the survey, but he was free to choose the addresses to be surveyed each week. A master copy of all the sample addresses was kept by the author.

Fairly close contact was maintained throughout the survey between the author and the team of interviewers. A meeting was held each Monday and Thursday morning during the survey. At these meetings problems pertaining to the survey were discussed, and suggestions for improvement were made.

The regular Monday meeting was also used for distributing to each interviewer the forms he required for the coming week. Each trip record form was stamped with the date for which trips were to be recorded (see Appendix 3), and it was therefore necessary for the interviewers to collect fresh supplies for each week of the survey.

On each Thursday, two weeks before a survey day, the interviewers had to supply the author with a list of addresses to be surveyed. This was necessary in order that the address labels could be attached to the envelopes, and the relevant introductory letters posted on the Tuesday evening preceding each survey week. There was a lot of clerical work involved in the preparation of the introductory letters and survey forms, and assistance was obtained within the Department of Civil Engineering at the University of Cape Town.

At the Monday morning meeting, each interviewer reported to the author his previous week's work. The master copy of addresses kept by the author was updated each week, and reflected the progress of the survey. The interviewers used a field schedule (see Appendix 3) to record the outcome of each visit made to the sample households. The following code was used for recording the progress of the survey.

C - Completed Interview
V - Vacant
U - Unable to make contact
R - Refused
N - Non-existent[24].

3.7.3 Modifications Made During the Survey

During the pre-test survey, and for the first week of the home questionnaire survey, the interviewers completed a blank trip record form in the presence of the person contacted in each household. This was done in order to explain the completion of this form. However, at the second Monday morning meeting, the interviewers pointed out that it was unnecessarily time-consuming to complete an example at each household, and it was suggested that copies of a standard example sheet be supplied to the interviewers. This suggestion was accepted, and a copy of the form is to be found in Appendix 3.

A second modification made during the survey was in connection with the recording of occupations. The interviewers correctly pointed out that in the case of a retired person it was useful to record the previous occupation, as well as the fact that the person was retired, since a wide variety of retired people can be found. This modification was brought into force during the second week of the home questionnaire survey.

3.7.4 Survey Response

The sample chosen for this survey consisted of 1 235 houses and 927 flats, giving a total sample size of 2 162 households. Table 3.3 illustrates the break-down of this sample.

Total Sample		Vacant		Non-Existent		Refused		Unable to Contact		Spoilt		Completed	
H	F	H	F	H	F	H	F	H	F	H	F	H	F
1235	927	50	67	26	25	72	74	57	66	16	19	1014	676
2162		117		51		146		123		35		1690	

Table 3.3

The sampling failure rate may be defined as

$$\frac{\text{Vacant + Non-Existent}}{\text{Original Sample}} \times 100$$

The interview failure rate may be defined as

$$\frac{\text{Refused + Unable to Contact + Spoilt}}{\text{Original Sample} - (\text{Vacant + Non-Existent})} \times 100$$

The overall failure rate may be defined as

$$\left(1 - \frac{\text{Completed Interviews}}{\text{Original Sample}}\right) \times 100$$

The sampling, interview, and overall failure rates were computed for houses and flats separately, as well as for all households. These results are shown in Table 3.4.

Description	Houses	Flats	All Households
Sampling Failure Rate	6,2	9,9	7,8
Interview Failure Rate	12,5	19,0	15,3
Overall Failure Rate	17,8	27,1	21,8

Table 3.4

Since all three failure rates were found to be higher for flats than for houses, we decided to test whether these differences were statistically significant. A test for differences in proportions[25] was used, and the computations are shown in Appendix 4. The statistical tests showed that, at a 1% level of significance, all three failure rates were significantly higher for flats than for houses.

The higher failure rates for flats was anticipated, since there is a higher turnover of occupants in flats and hence they are more likely to be vacant. In addition, we realized that it would be more difficult to make contact with flat-dwellers. For this very reason the ratio of sample houses to flats (1,33) was chosen slightly lower than the ratio of houses to flats (1,41) in the survey area in 1970. However, the ratio of completed interviews (1,50) shows that the actual difference in failure rates was higher than had been anticipated.

The figures shown in Table 3.3 indicate that the respondents were generally extremely co-operative, only 146 of the households contacted refused to assist with the survey. This represents only 6,8% of the sample households, or 7,8% of the households actually contacted. The interviewers felt that to a large extent the low rate of refusals was due to the fact that the survey was part of a research project.

In view of the fact that the overall failure rate was only just over 20%, it was decided that there was no need to carry out a non-response survey. In other words, we could reasonably assume that the households who completed the questionnaire were as representative of all the households in the study area, as were the households originally sampled.

The more trips taken by a household, the greater is the likelihood that no one will be at home. Similarly, households with few residents will be less likely to have at least one person at home. Therefore, there is reason to suppose that the population of 'unable to contact' households differs significantly from the overall population from which data were collected. If after three visits, the interviewers were unable to make contact, they reported this to the author but substitute households were not interviewed. The possible effect of the 'unable to contact' households was ignored since they comprised only 5,7% of the total sample.

3.8 Critique of the Cape Town Home Questionnaire Survey

In order to ascertain the opinion of the interviewers as regards the survey, a questionnaire was prepared by the author and distributed at the conclusion of the survey. A copy of this questionnaire is shown in Appendix 3.

All seven interviewers felt that the average of fifty households to be interviewed each week was a reasonable task. It should be noted that a five-day working week was used for this survey, as opposed to the four-day week mentioned by Taylor[3].

Only two of the interviewers felt that the newspaper publicity had any effect on the co-operation (or lack of co-operation) received from the sample households, while five felt that the newspaper publicity had no effect at all. It would seem therefore that not much benefit was obtained from the report that appeared in the Cape Times. This does not necessarily mean that the aid of newspapers should not be sought in future surveys of this kind.

The interviewers were unanimous in the opinion that the pre-interview letter was a necessary part of the survey; however, three of them felt that it was not necessary to have personally contacted the respondents prior to the survey day. This opinion, on the pre-interview technique, endorses the conclusion drawn by Memmott[2].

All the interviewers agreed that the standard trip record example form was a success. The use of this document in future home questionnaire surveys is recommended, not only as a time-saver, but also as a means of ensuring that the respondents have a document to which they can refer if they have any trouble in completing the trip record form.

The seven interviewers felt that the design of the survey forms was 'good' or 'very good'. One comment that was made by two of the interviewers, referred to the choice of the purpose categories. They felt that the category 'business trips' should have been included on the trip record form. Consideration was given to the inclusion of this category when the trip record form was designed; however, it was decided that this category comprised only a relatively small proportion of home-based trips.

The major ambiguity in the survey forms was only noticed when about one-third of the survey was completed. In the General Information Section on Form 'B', question three reads as follows:-

"3. If employed, please state (a) Occupation ....
(b) Industry or Business................."

Many respondents thought that in part (b) they were merely required to state whether they were employed in industry or in business. However, the intention was to find out the type of industry or business in which the respondent was employed, e.g. 'clothing industry'. This problem was pointed out to the interviewers, and they subsequently checked the replies to this question fairly carefully when they collected the completed forms.

The other major short-coming in the questionnaire used in this survey is linked to the above problem. The aim of the questions on occupation was to be able to form some assessment of the socio-economic standing of the respondents. However, ".....the occupation of a person is merely the kind of work performed by that person, regardless of the industry in which this work is performed and of the status of employment of the individual"[26].

If one wishes to determine occupational status, one needs to find out not only the type of work performed but also the position of the individual in the hierarchy. Unfortunately, the author only started looking at various occupational classifications after the survey was completed, and then only was the difference between occupation and occupational status fully appreciated.

It should be pointed out that question number four of the General Information Section of Form 'B' did request the respondent to record whether he was self-employed or not. Together with the respondent's occupation, this would give some idea of the occupational status. However, a fairly large number of the respondents did not answer this question, and for some unknown reason the interviewers did not pick this up when they collected the completed questionnaires.

In spite of the above-mentioned shortcomings, the Cape Town Home Questionnaire Survey was very successful. The very low percentage of spoilt replies (2,0%), and the very high overall success rate (78,2%), confirms this statement.

Consideration was given to sending thank-you letters to the households that responded to the survey; however, the cost of this task proved prohibitive. A thank-you letter was written to the editor of the Cape Times, and was published in that newspaper. A copy of this letter is to be found in Appendix 3.

NOTES - CHAPTER III

1. For the purposes of this study, a household is defined as a person or group of persons living at a single address, except where the premises are institutions, boarding houses, hotels, etc.

2. Memmott, F.W., "Home Interview Survey and Data Collection Procedures", Highway Research Board, Number 41, 1963.

3. Taylor, M.A., "Travel Surveys by Home-Questionnaire", Traffic Engineering and Control, July 1965.

4. Sullivan, S.W., "Variations in Personal Travel Habits by Day of Week", Highway Research Record, Number 41, 1963.

5. Hoel, L.A., Perle, E.D., Kansky, K.J., Kuehn, A.A., Roszner, E.S. and Nesbitt, H.P., Latent Demand for Urban Transportation, (Transportation Research Institute, Carnegie-Mellon University, Pittsburgh, 1969).

6. The sampling frame is the document from which the sample is drawn. See Methods of Social Research by M. Stacey (Pergamon Press, 1969).

7. Cape Times Peninsula Directory with Street Maps, 1973.

8. The following Department of Statistics News Releases were used: P.11 No. 3, P.11 No. 5, P.11 No. 6, P.11 No. 33, P.11 No. 36, P.11 No. 58.

9. Slonim, M.J., "Sampling in a Nutshell", Journal of American Statistics Association, June 1957.

10. This is based on the growth rate indicated in Figure 1.2.

11. Bruton, M.J., Introduction to Transportation Planning, Hutchinson Technical Education, 1970.

12. From a random sample of streets, selected from the Cape Times Directory, it was found that the average number of houses per street (in the survey area) was approximately ten. We

decided to survey, on the average, every alternate house in the selected streets. On this basis, the ratio of five houses per street was obtained.

13. The S_i survey streets for suburb i were selected as follows: Twenty six of the letters (A through Z) of a scrabble set were used to sample S_i letters at random. For each chosen letter, a street in suburb i, beginning with that letter, was selected from the alphabetical list of street names in the directory. The latter part of this selection procedure was not quite random sampling, but it is felt that this should not have affected the chosen sample of streets.

14. From a random sample of blocks it was found that, on average, each block contained approximately fifteen units. We decided to survey, on the average, every alternate flat in each selected block. On this basis, the ratio of eight flats per block was determined.

15. The sampling procedure used for the selection of blocks of flats was very similar to that used for selecting the survey streets - see note 13.

16. The assistance received in this connection has been gratefully acknowledged at the beginning of the thesis.

17. Goode, W.J. and Hatt, P.K., Methods in Social Research, McGraw-Hill, 1952.

18. Moser, C.A. and Kalton, G., Survey Methods in Social Research, Heinemann Educational Books, 1971.

19. In other words, not only cars owned by the members of the household, but also cars normally available for use by one or more members of the household.

20. This rather laborious task was completed by the author using the LETTER-PRESS range of lettering.

21. For example, a home interview survey in Toronto reported all trips in ten trip purpose categories, but at the analysis stage only four

different purpose categories were used.

22. Parten, M., Surveys, Polls and Samples: Practical Procedures, Harper and Brothers, 1950.

23. Kruger Day is a public holiday in the Republic of South Africa.

24. Includes any selected address which was found to fall outside the scope of the survey, as well as households which had been demolished.

25. Spiegel, M.R., Theory and Problems of Statistics, (Schaum's Outline Series, McGraw-Hill, 1961).

26. General Register Office, Classification of Occupations, (HMSO, 1960).

CHAPTER IV

ANALYSIS OF THE SURVEY DATA

4.1 Preliminaries

4.1.1 Introduction

Acceptable replies were received from a total of 1 690 households. Trip-making information was obtained about 4 571 people, five years of age or older, residing in these households. Since a large quantity of data had been collected, and it was intended to carry out fairly sophisticated statistical analyses, the aid of a digital computer was essential for the analysis of the data collected in the home questionnaire survey.

4.1.2 Coding the Survey Data

The first step was to transfer the information from the questionnaires to standard computer coding sheets. A number of students were employed by the author, during the summer vacation, to assist with this work. Approximately 275 man-hours were required to complete the task.

The survey information was coded in such a way that the study of trip generation could subsequently be carried out at the personal, household and zonal levels of analysis.

A number of points should be raised at this stage. Firstly, the trips were 'linked' manually during the coding process. The concept of a linked trip was first introduced by the Chicago Area Transportation Study, and refers to a trip made for essentially one purpose, but by more than one mode of transport. Figure 4.1 illustrates the linked trip concept.

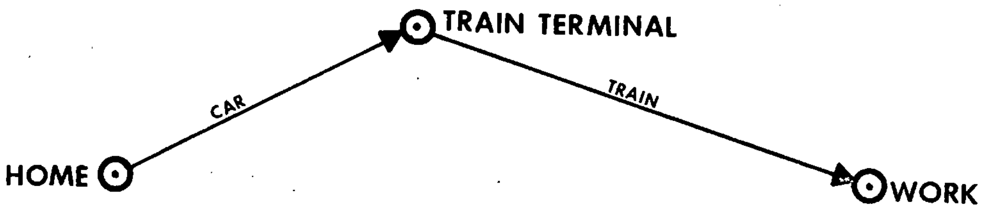


FIG. 4.1 The Linked Trip Concept

Before the linked trip idea was introduced, the above movement would have been considered as two trips. First, a home-based trip by car 'to catch the train', and then a non-home-based trip by train 'to get to work'. Using the linked trip concept, the above journey is defined as a home-based trip 'to get to work', by the major mode of transport used in making the trip.

It should also be pointed out that the occupational data was not coded, because the author felt that the occupational information collected in the home questionnaire survey was lacking in certain respects (see section 3.8). In addition, a suitable occupational classification system could not be found.

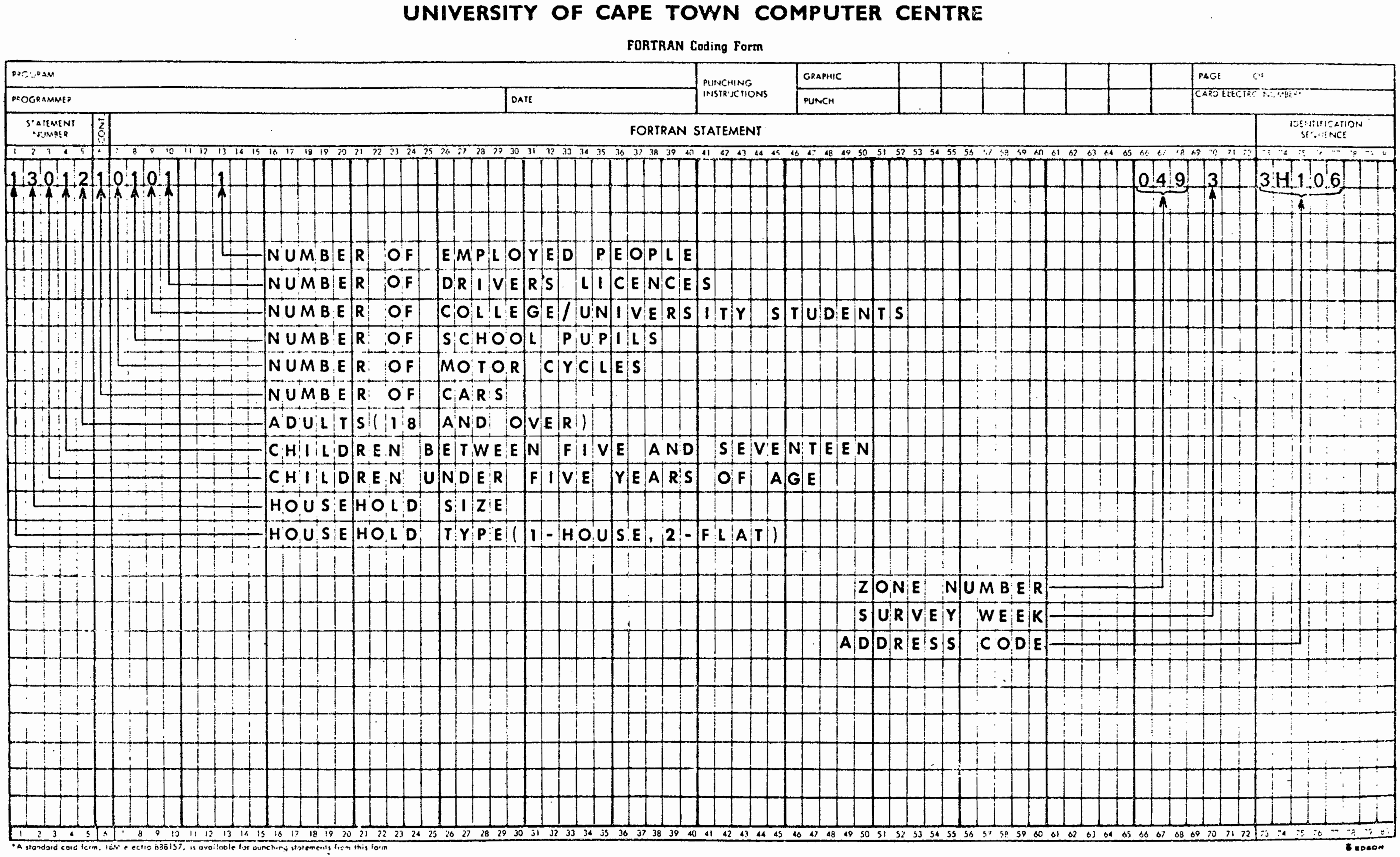
Instead of boring the reader with a description of the coding, the author has prepared Figures 4.2 and 4.3, which clearly illustrate the layout of the information on the coding sheets.

The zone numbers referred to in Figure 4.2 are those used by Aplin[1], while the zones to which these numbers refer are those used by the Advance Planning Section of the City Engineer's Department[2]. The decision to use these zones was influenced by two major considerations. Firstly, the Cape Town City Engineer's Department is collecting information relevant to these zones, and it seemed sensible to use the same zone boundaries so that the zonal information collected in this study could be used in future projects. Secondly, these zones had been delineated so as to be approximately homogeneous (with respect to socio-economic and land-use characteristics); and (supposedly) homogeneous zones were needed in order to analyse the data at the zonal level.

4.1.3 Storing the Data on the Computer

The information was transferred from the coding sheets to computer cards. The punching and verification of the cards was kindly undertaken by the punch operators of the Computer Centre at the University of Cape Town (U.C.T.)[3].

A total of 6 261 computer cards were used to store all the information gathered in the home-questionnaire survey. For each household, one card was punched containing the household's socio-economic characteristics, zone number, week of survey and the address code[4]. In addition, one card was used for storing the personal and trip-making data for each of the 4 571 respondents. Each card containing a household's information was



UNIVERSITY OF CAPE TOWN COMPUTER CENTRE

FORTRAN Coding Form

PROGRAM

PROGRAMMER

DATE

PUNCHING INSTRUCTIONS

GRAPHIC

PUNCH

PAGE OF

CARD ELECTRO NUMBER*

STATEMENT NUMBER | CONT | FORTRAN STATEMENT | IDENTIFICATION SEQUENCE

```
1301210101  1                                                    049 3  3H106
```

NUMBER OF EMPLOYED PEOPLE

NUMBER OF DRIVER'S LICENCES

NUMBER OF COLLEGE/UNIVERSITY STUDENTS

NUMBER OF SCHOOL PUPILS

NUMBER OF MOTOR CYCLES

NUMBER OF CARS

ADULTS(18 AND OVER)

CHILDREN BETWEEN FIVE AND SEVENTEEN

CHILDREN UNDER FIVE YEARS OF AGE

HOUSEHOLD SIZE

HOUSEHOLD TYPE(1-HOUSE, 2-FLAT)

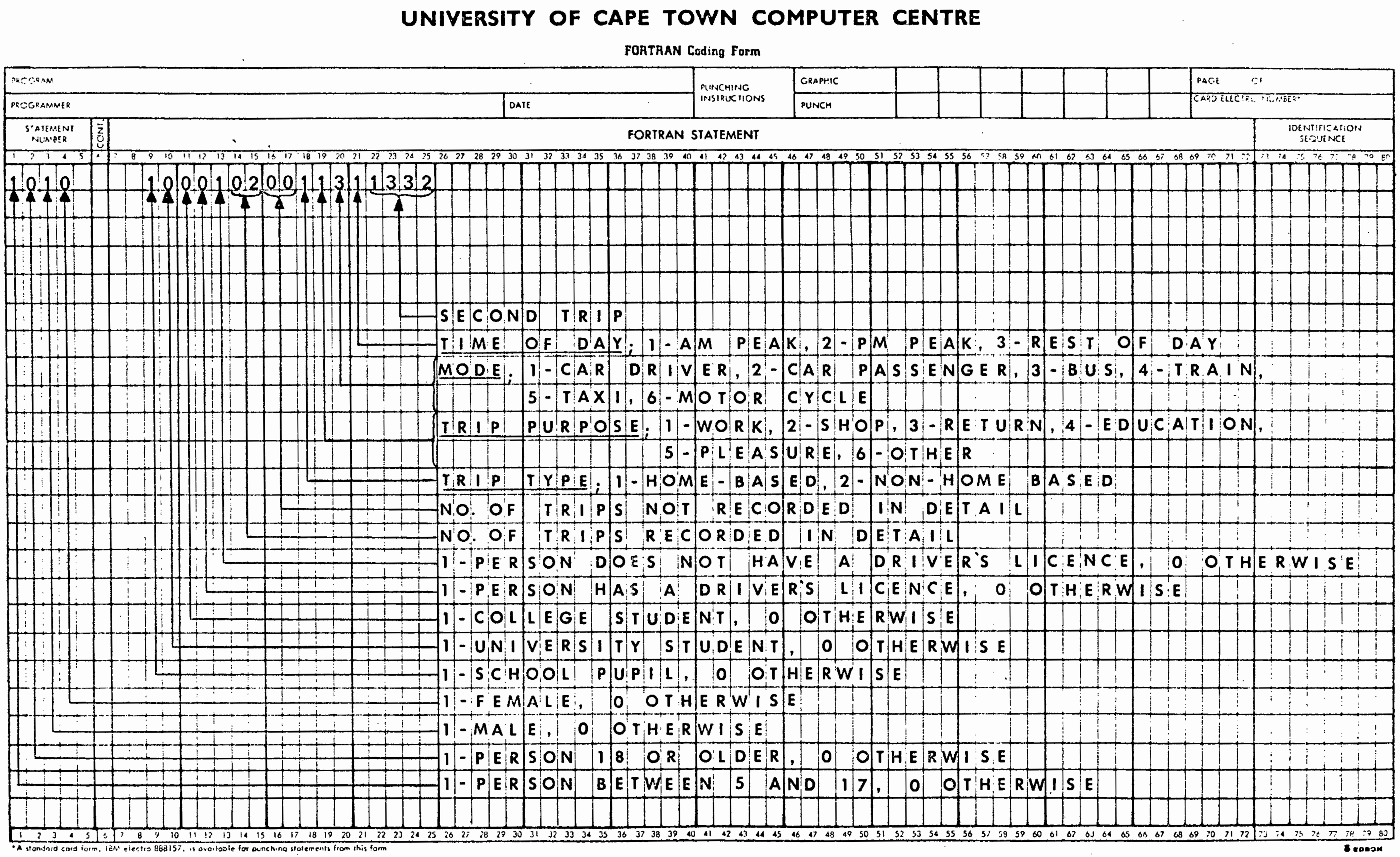
ZONE NUMBER

SURVEY WEEK

ADDRESS CODE

*A standard card form, IBM electro 888157, is available for punching statements from this form

FIG. 4.2 Coding- Household Data



UNIVERSITY OF CAPE TOWN COMPUTER CENTRE

FORTRAN Coding Form

PROGRAM

PROGRAMMER

DATE

PUNCHING INSTRUCTIONS

GRAPHIC

PUNCH

PAGE OF

CARD ELECTRO NUMBER*

STATEMENT NUMBER

CONT.

FORTRAN STATEMENT

IDENTIFICATION SEQUENCE

```
1010    10001020011311332
```

SECOND TRIP

TIME OF DAY; 1-AM PEAK, 2-PM PEAK, 3-REST OF DAY

MODE, 1-CAR DRIVER, 2-CAR PASSENGER, 3-BUS, 4-TRAIN,
5-TAXI, 6-MOTOR CYCLE

TRIP PURPOSE, 1-WORK, 2-SHOP, 3-RETURN, 4-EDUCATION,
5-PLEASURE, 6-OTHER

TRIP TYPE, 1-HOME-BASED, 2-NON-HOME BASED

NO. OF TRIPS NOT RECORDED IN DETAIL

NO. OF TRIPS RECORDED IN DETAIL

1-PERSON DOES NOT HAVE A DRIVER'S LICENCE, 0 OTHERWISE

1-PERSON HAS A DRIVER'S LICENCE, 0 OTHERWISE

1-COLLEGE STUDENT, 0 OTHERWISE

1-UNIVERSITY STUDENT, 0 OTHERWISE

1-SCHOOL PUPIL, 0 OTHERWISE

1-FEMALE, 0 OTHERWISE

1-MALE, 0 OTHERWISE

1-PERSON 18 OR OLDER, 0 OTHERWISE

1-PERSON BETWEEN 5 AND 17, 0 OTHERWISE

*A standard card form, IBM electro 888157, is available for punching statements from this form

FIG. 4.3 Coding-Personal And Trip-Making Data followed by the cards for all the members of that household. Figure 4.4 illustrates the arrangement of the data cards.

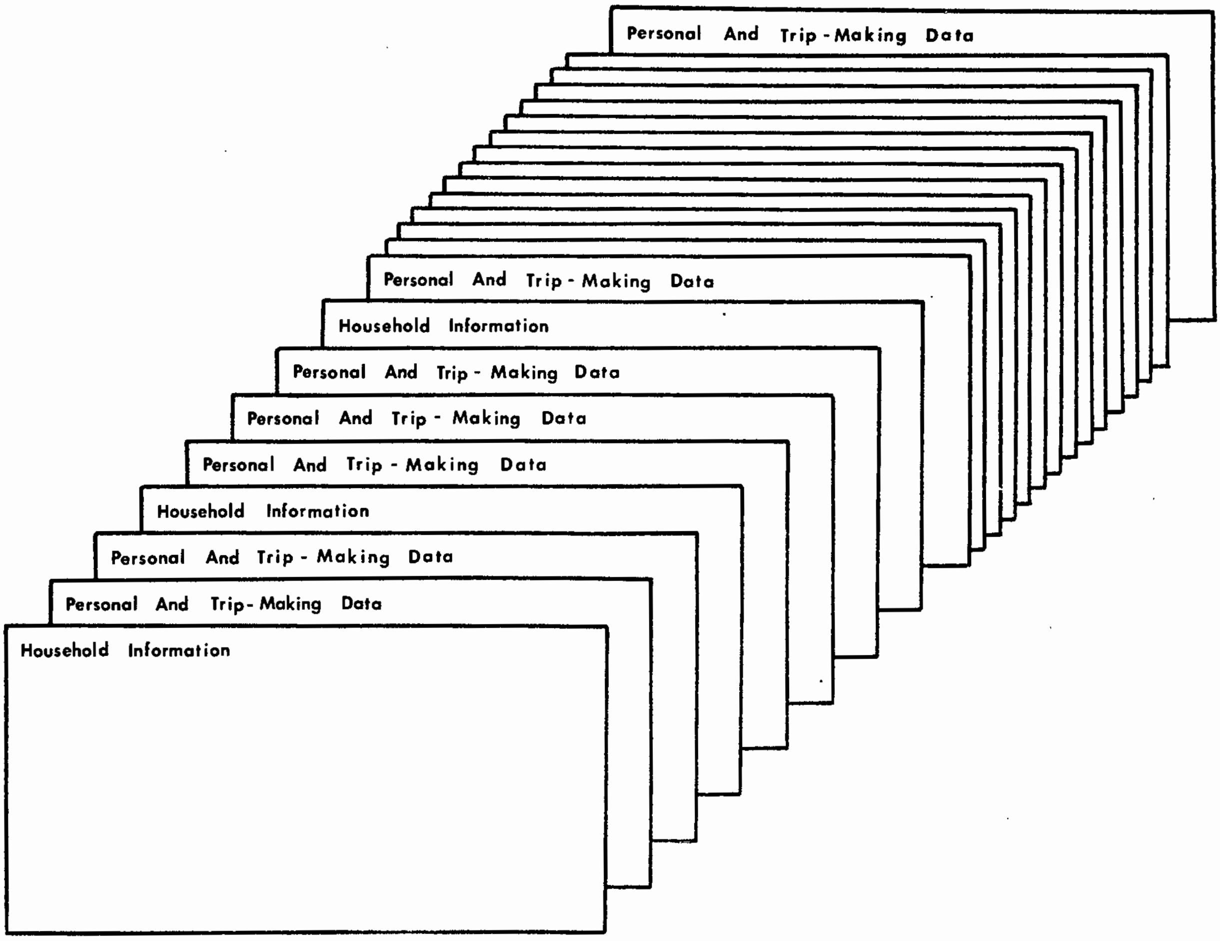


FIG. 4.4 Contents Of File Surveydata

The information recorded on the data cards was stored in a datafile[5], called SURVEYDATA, on the UNIVAC 1106 computer at U.C.T. A computer program was written by the author in order to:

(i) check for inconsistencies in the data;

(ii) create additional datafiles from the original file.

A number of coding and punching mistakes were located with the aid of this program, and the file SURVEYDATA was up-dated to eliminate these mistakes. The original card deck was not up-dated, but a back-up copy of SURVEYDATA was stored on the computer for security reasons.

Once all the obvious coding and punching mistakes had been corrected, the computer program carried out its second phase - the creation of new datafiles from the basic information stored in SURVEYDATA. Three additional datafiles were created by the program, as follows:-

(a) RTG 1

(b) RTG 2

(c) RTG 3

The file RTG 1 contained 1 690 records, one for each household. Each record contained not only the socio-economic characteristics, zone number, survey week and address code, but also the number of trips made by members of the household, in each of sixteen categories. There were fifteen categories for home-based trips and one for non-home-based trips. The fifteen home-based categories covered six different trip purposes, six different transport modes, and three time periods (see section 3.5.3).

The file RTG 2 contained 4 571 records, one for each person in the 1 690 households. These records contained the personal and trip-making information needed in order to carry out an analysis of residential trip generation at the personal level. The trip-making information was recorded for the same sixteen categories mentioned above. Each record also contained the socio-economic characteristics of the household to which the person belonged.

The file RTG 3 contained 12 333 records, one for each trip recorded in detail[6] by the respondents. Each record in this file consisted of four digits, by means of which the trip's characteristics were described (see Figure 4.3).

The data in file RTG 1 was aggregated to the zonal level, so that trip generation analyses could be carried out in terms of both zonal total and rate variables[7]. The zonal total data was stored in a file called RTG 4, while the zonal rate data was stored in file RTG 5. The formation of the various data files is illustrated in Figure 4.5

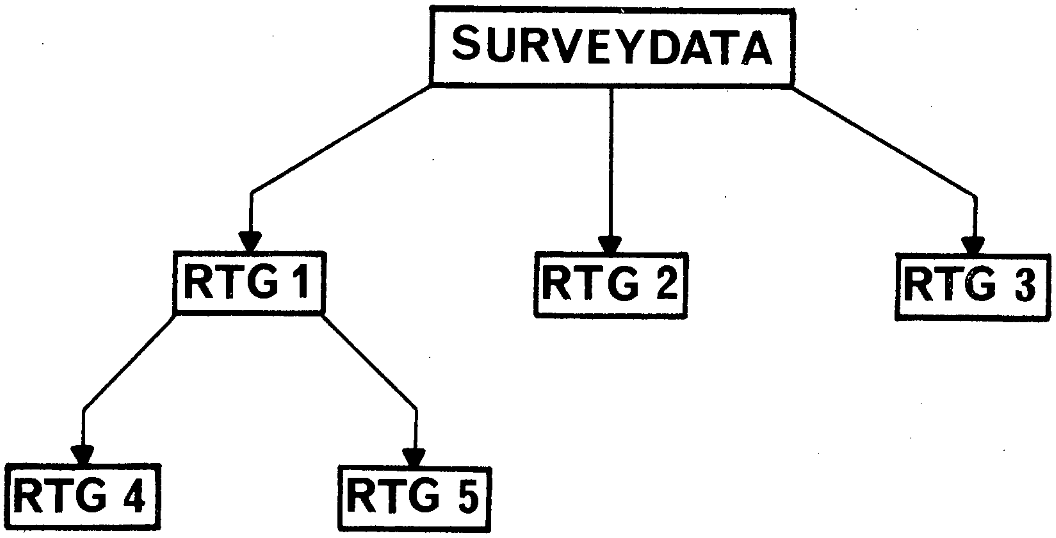


FIG. 4.5 Formation Of Datafiles

4.2 Home-based Trips by Purpose, Mode and Time of Day

4.2.1 Introduction

A computer program was written by the author to analyse the trip data contained in file RTG 3. The function of the program was to separate the home-based trips from the non-home-based trips, and to analyse the home-based trips with respect to their purpose, mode and time of day characteristics.

A total of 10 641 home-based trips were reported by the 4 571 people who completed the questionnaire. Thus, on the average, each person made approximately 2,33 home-based trips per day. In addition, each person made an average of 0,55 non-home-based trips per day.

4.2.2 Purpose

The distribution of the home-based trips by purpose is shown below in Figure 4.6.

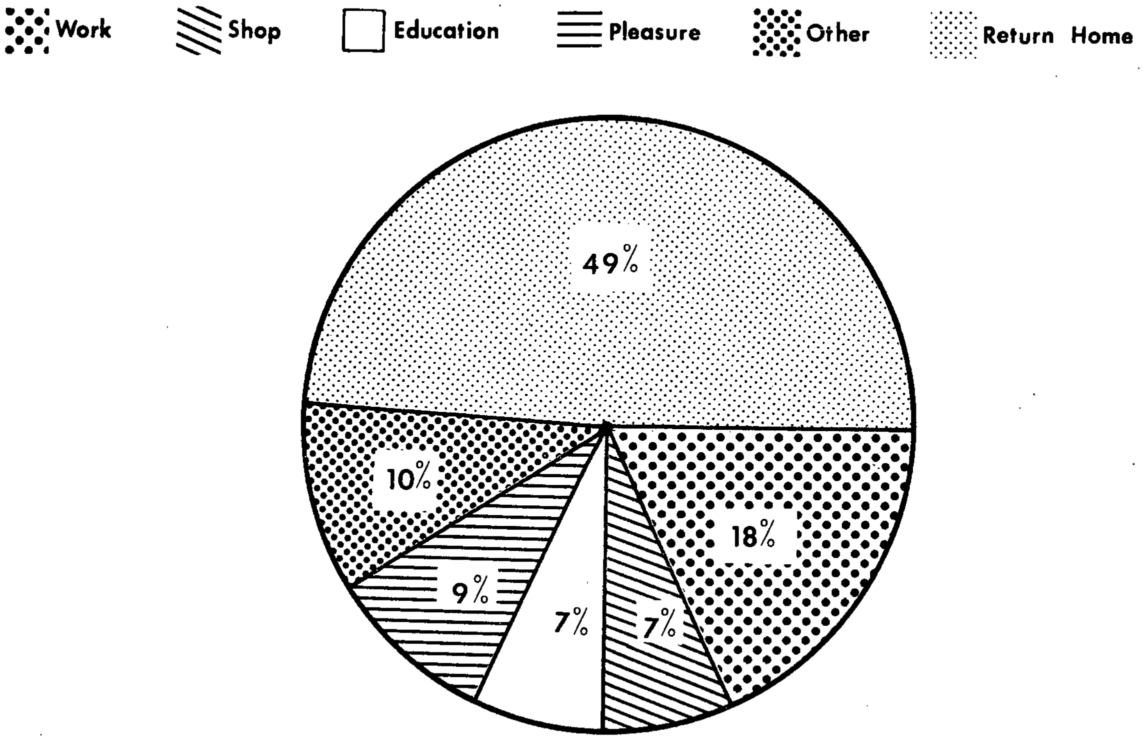


FIG. 4.6 Distribution Of Home-Based Trips By Purpose

Studying the results shown in Figure 4.6, we see that it is reasonable to conclude that the number of purpose categories used was satisfactory, in that only 10 per cent of the trips are in the 'other' category.

The purpose distribution of home-based trips was compared with the results obtained in other study areas. This comparison is presented in Table 4.1. Some of the definitions used differed from one study area to the next; however, the major differences are noted in the table. Allowing for these differences, one sees a fairly good comparison between the results obtained in Cape Town and the other study areas.

Urban Area	Year of Survey	Percentage of Home-Based Trips[1]				
		Work[2]	Shopping	Social-Recreation	School	Other
Detroit	1953	41,6	13,9	20,1	6,3	18,1
Greater London	1962	56,1	9,3	15,2	5,6	13,8
Kansas City, Missouri	1957	33,4	17,2	22,7	6,0	19,7
Washington D.C.	1955	43,1	14,2	12,5	9,4	19,8
Cape Town	1973	35,0	14,0	17,0	14,0[3]	20,0

1. Except for Cape Town, these figures refer to trips to and from the stated purpose. The Cape Town figures refer to the percentage of 'from-home' trips for each purpose.
2. Except for Cape Town, work trips include 'employer's business' trips.
3. This figure applies to trips made to all institutions of learning including colleges and universities.

Table 4.1

4.2.3 Mode

The modal split of home-based personal trips was analysed for each of the different trip purposes, and the results of this analysis are shown in Figure 4.7. We see that 80 per cent of all home-based trips were made by car - the person travelling either as the driver or as a passenger. Using the results shown in Figure 4.7(a), we can estimate the average car occupancy (for home-based trips) as 1,51[8].

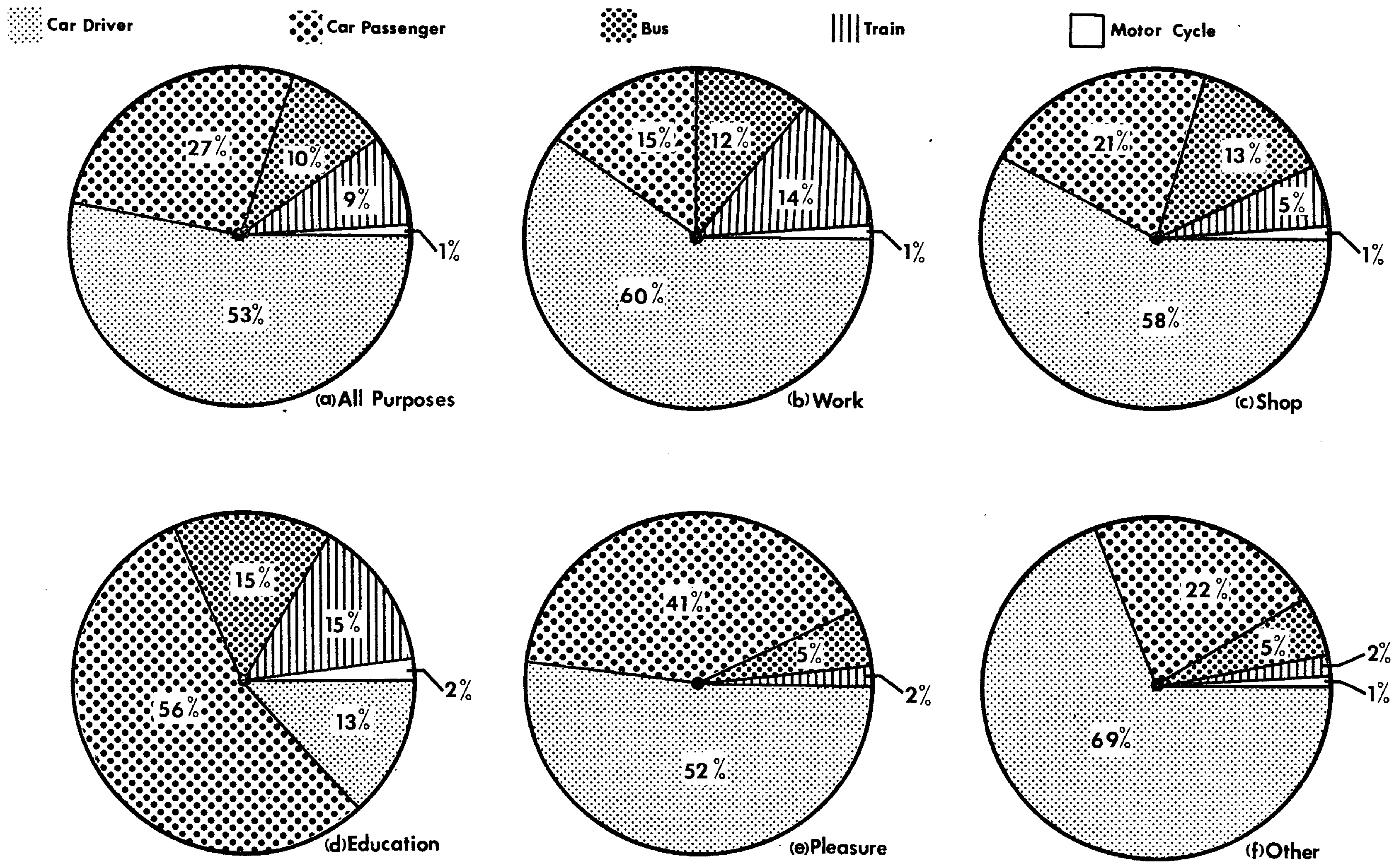


FIG. 4.7 Distribution Of Home-Based Trips By Mode, For Different Purposes

The low level of public transport patronage is to be expected if one considers that, on the average, there were only 1,75 people (over the age of eighteen) per car. It will be noticed that trips by taxi do not appear in Figure 4.7, since the number of such trips was negligible.

The home-to-work journey reveals a very high percentage of car driver trips, while the proportion of car passenger trips for this purpose was fairly low. Based on the data in Figure 4.7(b), we can estimate the average car occupancy for the home-to-work trip as 1,25. This compares very well with the results reported by Aplin[1]. Public transport trips accounted for 26 per cent of home-to-work trips, compared with the 19 per cent of all home-based trips made by public transport.

Figure 4.7(c) illustrates the modes of transport used for home-based shopping trips. Once again, a very high proportion of trips by motor car is seen. The average car occupancy for home-based shopping trips can be estimated as 1,36, while the proportion of public transport trips is seen to be only 18 per cent.

'Education trips' cover all movements from home to places of learning, including schools, colleges and universities. For this reason, 13 per cent of education trips are made by car drivers. A very high percentage (56 per cent) of all home-based education trips are made as car passengers.

However, the modal split for education trips is probably fairly severely influenced by the fact that no account was taken of walking trips, or trips made by pedal cycle. Since these modes of transport probably account for a significant proportion of education trips, we decided to investigate this further.

The total number of students and scholars in the sample was 1 145. Assuming that 90 per cent of these students and scholars attend school, college or university on any one day, we can estimate that our sample would make approximately 1 030 (0,90 × 1 145) education trips on any one day. The total number of home-based education trips recorded was 782, and hence we see that approximately 250 education trips were probably made either by foot or by pedal cycle. On this basis, approximately 42 per cent of education trips (by all modes of transport) were made as car passengers. Is this some indication of an affluent society?

It must be pointed out that while coding the survey data, the author noticed that a fair proportion of scholars were given a lift to school but appeared to walk home, since no return trip was recorded. As would be expected, a fairly large proportion of education trips are made by public transport (30 per cent).

As far as pleasure (social-recreation) trips are concerned, we see that, as we would expect, a very high percentage (93 per cent) are made either as car driver or as car passenger. This clearly shows the advantage of a car, in that the people having a private vehicle are able to participate in more of the activities the city has to offer.

The high percentage of 'other' trips made as car driver is to be expected, since this category includes the purpose commonly termed 'serve passenger'; in other words, trips made for the purpose of transporting someone else.

4.2.4 Time of Day

The distribution of trips at different times of the day was investigated for different trip purposes. These results are illustrated in Figure 4.8. Fifty per cent of all home-based trips are made during the morning and afternoon peak periods. In other words, fifty per cent of all home-based trips are made during four hours of the twenty four hour day. This clearly illustrates the magnitude of the so-called 'peak period problem'.

As would be expected, the vast majority of home-to-work trips are made during the morning peak period, whilst most of the home-based shopping trips are made during the off-peak period. However, we notice that 28 per cent of home-based shopping trips are made during the morning and afternoon peak periods. Perhaps some of these trips could be made at other times of the day?

The results for education trips are as expected, and they show that to a large extent these trips overlap with the morning journey-to-work. The fact that schools, colleges and universities have a starting time which is similar to that in commerce and industry, enables a number of children to get lifts to school; but it also places additional loads on the public transport system during the morning peak period.

As we would expect, a large proportion of pleasure trips are made

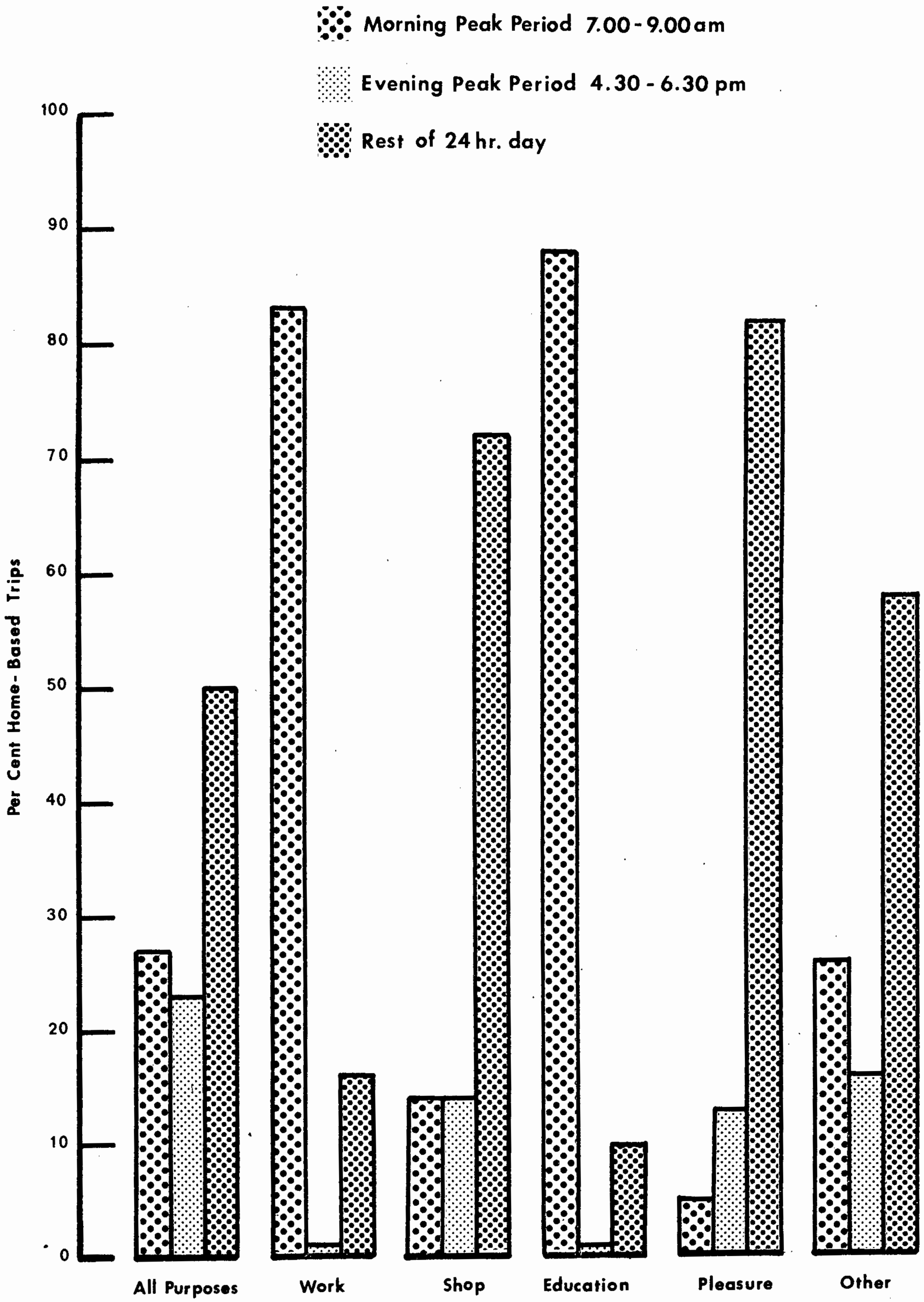


FIG. 4.8 Distribution Of Home-Based Trips By Time Of Day, For Different Purposes

during 'the rest of the day', but a possibly significant percentage (13 per cent) are made during the afternoon peak period.

We have discussed the peak period problem in terms of the percentage of home-based trips (for different purposes) made during the three different time periods analysed in this study. It is very interesting to see the peak period problem in terms of the usage of alternative modes of transport.

Figure 4.9 shows the percentage of home-based trips by each mode in each time period. These results clearly indicate that the peak period problem is more pronounced in the case of public transport, especially the train, than in the case of the private motor car.

The time distribution of trips by motor-cycle and taxi is not shown in Figure 4.9, since trips by these modes form a very small proportion of all home-based trips made by the study area population.

4.3 Development of Preliminary Equations

4.3.1 Introduction

In order to study the effects of analysing the data at different levels of analysis, a number of preliminary equations were developed using the home questionnaire survey information.

An extremely interesting feature of the work reported here is the fact that the personal level of analysis was not automatically dismissed from further consideration, as has happened in the past. Hence, preliminary trip generation equations were developed at the personal, household, and zonal levels of analysis.

The zonal models were developed in terms of both total and rate variables. In addition, two preliminary household equations had to be developed in order that the required comparison could be made.

4.3.2 Least-Squares Computer Program (BMD02R)

The computer program, used in the development of the least-squares regression equations reported in this thesis, was selected from the

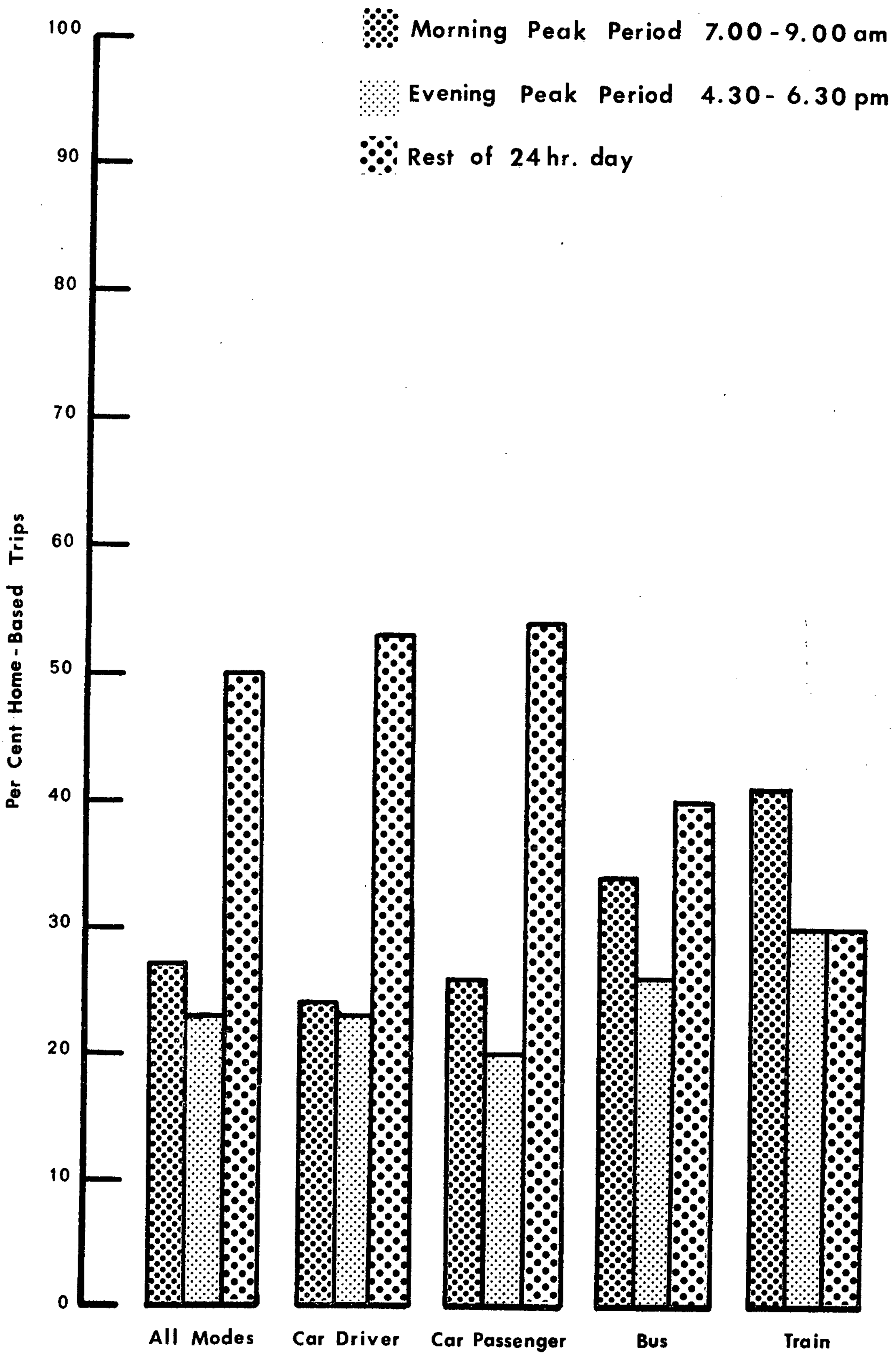


FIG 4.9 Distribution Of Home-Based Trips By Time Of Day, For Different Modes

BioMeDical (BMD) Computer Programs[9]. This package was developed at the University of California in 1964, mainly for use in medical research projects. The programs in this package are very powerful and versatile, and enable the user to undertake a large variety of statistical analyses.

We have already mentioned (section 2.5.5) that a stepwise regression procedure has been suggested as being the most suitable for selecting the 'best' regression equation when one has a number of possible explanatory variables. The BMD02R program computes a sequence of multiple linear regression equations in a stepwise manner, and one variable is added or deleted at each step of the process.

The criterion used for entering or deleting a variable is the contribution made by that variable to the explained sum of squares, given the effect of the variables already in the regression. The significance of the reduction in the unexplained sum of squares, brought about by the inclusion of a particular variable, is measured by the so-called partial F-test.

At this point it should be noted that a full description of the BMD02R computer program is given in Appendix 5. The reader who is unfamiliar with the stepwise regression concept in general, and the BMD02R program in particular, is referred to this Appendix. The basic features of the program are briefly discussed below.

The BMD02R program provides the user with the ability to force variables into the regression equation even when their F-values are very low. In addition, one can delete a variable from the analysis, although it would normally have entered on the strength of its contribution to the regression.

This feature is particularly useful when one is dealing with dummy variables, since if one has n dummy variable classes in a set, then either $n - 1$ of them must enter the regression, or none at all.

The simple correlation matrix, which can be obtained prior to performing the regression, is extremely useful for detecting cases of multicollinearity. The analyst should study this matrix very carefully before developing any final regression models.

The BMDO2R program was found to be generally very satisfactory, except for what appears to be an error in the user manual. This manual points out that one must specify the F-level to be used for inclusion and deletion of variables in the stepwise process. If these values are not specified, they are assumed to be 0,01 and 0,005 respectively. One is led to believe that the level of significance is required, but in fact it is the F-value which is needed. This point is discussed in detail in Appendix 5.

4.3.3 The Preliminary Personal Level Equation

Since no comparable analysis could be found in the literature, the author had no basis for selecting the explanatory variables to be used in this investigation. For this reason, the preliminary regression analysis at the personal level was carried out by allowing a large number of explanatory variables to enter the regression equation in a stepwise procedure.

Although the analysis was carried out at the personal level, for each person[10], not only was the personal information used, but also the socio-economic characteristics of the household in which the person resided. Hence the variables considered in this analysis can be divided into two groups:

(a) Personal Variables.
(b) Household Variables.

After two steps of the stepwise regression procedure, the following equation was obtained:

$$\text{HBTPD} = 2{,}50 - 1{,}25\ \text{NOLICE} + 0{,}29\ \text{CARS} \qquad (4.1)$$

$$(R^2 = 0{,}153; \quad S_E = 1{,}75)$$

where HBTPD = number of home based trips per person per day,
NOLICE = 1, if the person does not possess a driver's licence,
0, if the person possesses a driver's licence.
CARS = car ownership of the household in which the person lives.

In other words, the variable NOLICE is a dummy variable at the personal level, but CARS is a household characteristic used here at the personal level[11]. It is important to notice that the variable LICENC, which is the complement of NOLICE, was deleted from the regression analysis for the reason given in section 2.2.4.

After seven variables were included in the regression equation, the coefficient of determination (R^2) had only increased to 0,196, while the standard error of estimate (S_E) had only decreased to 1,71. At the same time, a number of relatively highly inter-correlated variables had entered the regression equation. For these reasons, equation 4.1 was selected as the personal level equation to be used in the preliminary investigations.

4.3.4 The Preliminary Household Level Equations

Two preliminary household relationships were derived, the first having the number of home-based trips per household per day as the dependent variable, while the second had the number of home-based trips per person per household per day as the dependent variable.

The second equation may be thought of as a household rate model[12]. This relationship was developed in order to compare the household and personal level analyses, while the first model was derived for comparing the household and zonal level analyses.

The first preliminary household model was as follows:-

$$\mathrm{HBTHD} = -0{,}84 + 1{,}53\ \mathrm{FAMSIZ} + 1{,}75\ \mathrm{CARS} + 0{,}43\ \mathrm{EMPLOY} \qquad (4.2)$$

$$(R^2 = 0{,}437; \quad S_E = 3{,}93)$$

where HPTHD = number of home-based trips per household per day,
FAMSIZ = number of people per household,
CARS = number of cars per household,
EMPLOY = number of employed people per household.

Only the above three explanatory variables were allowed to enter the regression equation, this choice being based on an examination of past studies at the household level.

As far as the household rate equation was concerned, we were not able to use the guidance of previous research findings in order to select the explanatory variables to be utilized in this preliminary analysis. Hence a number of explanatory variables were free to enter the regression equation. After five steps of the stepwise process the following equation was obtained:-

$$\begin{aligned} HBTPHD = {} & 0{,}92 + 0{,}63\ CARSP + 1{,}45\ UNIVERP + 0{,}94\ DRIVLIP \\ & + 0{,}63\ EMPLOYP + 1{,}24\ SKOLARP \\ & (R^2 = 0{,}188; \quad S_E = 1{,}32) \end{aligned} \tag{4.3}$$

where HBTPHD = number of home-based trips per person per household per day,

CARSP = number of cars per person per household,

UNIVERP = number of college/university students per person per household,

DRIVLIP = number of driver's licences per person per household,

EMPLOYP = number of employed people per person per household,

SKOLARP = number of school children per person per household.

The main point to note in connection with the above equation is the relatively low value of the coefficient of determination, even with the inclusion of five explanatory variables. Equation 4.3 was used only for preliminary comparisons, and therefore no account was taken of the fairly high inter-correlation between some of the explanatory variables.

4.3.5 The Preliminary Zonal Level Equations

Three preliminary zonal level models were developed, one in terms of zonal total variables, and two in terms of zonal rate variables.

The first preliminary zonal rate equation was as follows:-

$$\begin{aligned} HBTHZD = {} & 0{,}64 + 1{,}48\ FAMSIZHZ + 2{,}60\ CARSHZ - 1{,}72\ EMPLOYHZ \\ & (R^2 = 0{,}570; \quad S_E = 1{,}56) \end{aligned} \tag{4.4}$$

where HBTHZD = number of home-based trips per household per zone per day,

FAMSIZHZ = number of people per household per zone,

CARSHZ = number of cars per household per zone,

EMPLOYHZ = number of employed people per household per zone.

The coefficient of the variable EMPLOYHZ in the above equation has a negative sign, which is contrary to what one would have expected. This is probably due to the fact that the correlation between FAMSIZHZ and EMPLOYHZ was 0,60; which is relatively high. However, since equation 4.4 is to be used in the preliminary investigations only, the illogical sign of the coefficient need not worry us at this stage.

An equation was derived for the number of home-based trips per person per zone per day. This equation was as follows:-

$$\text{HBTPZD} = 1{,}82 + 2{,}52\ \text{DRIVLIPZ} - 1{,}97\ \text{EMPLOYPZ} \qquad (4.5)$$
$$(R^2 = 0{,}431; \quad S_E = 0{,}45)$$

where HBTPZD = number of home-based trips per person per zone per day,
DRIVLIPZ = number of driver's licences per person per zone,
EMPLOYPZ = number of employed people per person per zone.

As with equation 4.4, there is a coefficient with an illogical sign, but this will be ignored during the preliminary investigations.

The preliminary zonal total equation was as follows:-

$$\text{HBTZD} = -1{,}95 + 1{,}22\ \text{FAMSIZZ} + 3{,}24\ \text{CARSZ} - 0{,}94\ \text{EMPLOYZ} \qquad (4.6)$$
$$(R^2 = 0{,}971; \quad S_E = 34{,}4)$$

where HBTZD = number of home-based trips per zone per day,
FAMSIZZ = number of people per zone,
CARSZ = number of cars per zone,
EMPLOYZ = number of employed people per zone.

Notice the very high coefficient of determination for the above model, and the illogical sign.

4.3.6 Summary - Preliminary Equations

At this stage it is worth summarizing the statistics which apply to the preliminary equations reported above. The table below provides the relevant information.

Level of Analysis	No. of Observations	R^2	S_E	$\bar{Y}$	$\frac{S_E}{\bar{Y}} \times 100$
Zone (Total)	49	0,97	34,4	217,1	15,9
Zone (Rate)[1]	49	0,57	1,56	6,73	23,0
Zone (Rate)[2]	49	0,43	0,45	2,26	19,9
Household (Total)	1 690	0,44	3,93	6,30	62,4
Household (Rate)	1 690	0,19	1,32	2,29	57,6
Person	4 571	0,15	1,75	2,33	75,1

1 - dependent variable, trips per household per zone per day,
2 - dependent variable, trips per person per zone per day.

Table 4.2

On the basis of these results it would appear that the zonal models are far better (statistically) than the equations derived at the other levels of analysis. However, the effects of data aggregation must be investigated before comparing the results reported above.

4.4 The Effects of Data Aggregation

4.4.1 Introduction

In section 2.3.4 a method was outlined for studying the consequences of aggregating data from one level to another. In this section of the report we will examine the effects of aggregating trip-making information from the household to the zonal level, as well as from the personal to the household and zonal levels.

4.4.2 Effects of Aggregation - Household to Zonal Level

The variable, home-based trips per household per day, was considered in order to study the effects of aggregating trip-making information from 1 690 households to 49 zones. The results of the one-way analysis of variance are presented in Table 4.3 below.

Total Sum of Squares (TSS) = 46 340
Between Sum of Squares (BSS) = 11 205
Within Sum of Squares (WSS) = 35 135
BSS as a percentage of TSS = 24,2 %

Table 4.3

The significance of these results is that 75,8 per cent of the variation in home-based trips per day, observed between different households, is to be found within the zones, when household trip-making data is aggregated to the zonal level. Therefore, only 24,2 per cent of the variation remains to be explained by a zonal equation. Fleet and Robertson[13] found, when aggregating data from 5 255 households to 247 zones, that 79,6 per cent of the heterogeneity was within the zones.

4.4.3 Effects of Aggregation - Personal to Household Level

This investigation proved most interesting since no comparable study could be found in the literature. The variable home-based trips per person per day was used in order to examine the effect of aggregating the trip-making data of 4 571 people to 1 690 households. The results of the one-way analysis of variance are shown below.

Total Sum of Squares (TSS)	=	16 498
Between Sum of Squares (BSS)	=	9 037
Within Sum of Squares (WSS)	=	7 461
BSS as a percentage of TSS	=	54,8 %

Table 4.4

Thus 45,2 per cent of the variation in home-based trips per person per day is to be found within households, with only 54,8 per cent of the variation remaining to be explained by a regression analysis at the household level. Therefore, although there is some degree of homogeneity within households, a significant proportion of the variation between individuals is obscured when the personal trip-making data is aggregated to the household level.

4.4.4 Effects of Aggregation - Personal to Zonal Level

As with the above analysis, no comparable study could be found, and therefore the results reported in Table 4.5 are extremely interesting. The variable used in this analysis was home-based trips per person per day.

Total Sum of Squares (TSS)	=	16 498
Between Sum of Squares (BSS)	=	1 866
Within Sum of Squares (WSS)	=	14 632
BSS as a percentage of TSS	=	11,3 %

Table 4.5

It is obvious that only an extremely small proportion of the variation between individuals is retained when personal trip-making data is aggregated to the zonal level, and a zonal regression analysis would attempt to explain only 11,3 per cent of the overall variation in trip-making.

4.4.5 Effects of Aggregation - Summary

The effect of aggregating the individual's trip-making information to other units of analysis is summarized in Table 4.6.

Unit	No. of Observations	BSS	WSS	$\frac{BSS}{TSS} \times 100$	$\frac{WSS}{TSS} \times 100$
Person	4 571	16 498	0	100,0	0,0
Household	1 690	9 037	7 461	54,8	45,2
Zone	49	1 866	14 632	11,3	88,7

Table 4.6

The results above clearly show the tremendous loss of information which takes place when personal trip-making data is aggregated to other (higher) levels of analysis.

4.5 Estimating Zonal Home-Based Trip Generation Using Household and Personal Level Equations

4.5.1 Introduction

As output from the trip generation phase of the traffic forecasting process, estimates of zonal trip generation are required. This means that irrespective of the unit of analysis used in the development of a trip generation equation, the model must be capable of estimating the number of zonal trip ends. Therefore the author investigated the ability of the preliminary personal and household level equations to estimate the number of home-based trip ends per day in each of the 49 zones within the study area.

4.5.2 Applying the Household Equation at the Zonal Level

The preliminary household equation (4.2) was used for predicting the number of home-based trips per day for each of the 49 analysis zones. For each zone, the average value of the three explanatory variables was inserted in equation 4.2 in order to estimate the average number of home-based trips per household per day, for that zone. The average rate for each zone was multiplied by the number of households in the zone to obtain an estimate of home-based trips per day for the zone.

A linear least-squares regression analysis was carried out between the observed and predicted zonal home-based trips per day. The results of this analysis were as follows:-

$$y = 11{,}9 + 0{,}95\ x \tag{4.7}$$

$$(r^2 = 0{,}960; \quad S_E = 38{,}2)$$

where y = zonal home-based trips per day estimated by the preliminary household equation,

x = observed zonal home-based trips per day.

The above results reveal a very high linear correlation between x and y. It is important to notice that although the regression line does not pass through the origin, the constant term in the equation is less than six per cent of the mean value of y. In addition, the least-squares regression line is inclined at very nearly 45 degrees to the horizontal. A graphical presentation of the above results is to be found in Figure 4.10.

The ability of a household level least-squares regression equation to predict zonal trip generation is clearly illustrated by the above results. These results may be compared to those reported by Fleet and Robertson[13], as shown below in Table 4.7.

Level of Analysis	No. of Observations	R^2	S_E	$\bar{Y}$	$\frac{S_E}{\bar{Y}} \times 100$
Zone[1] (Total Variables)	143	0,95	296,1	1679	17,6
Zone[2] (Total Variables)	49	0,97	34,4	217,1	15,9
Household[1]	5 255	0,36	3,89	5,20	74,9
Household[2]	1 690	0,44	3,93	6,30	62,4
Household at Zonal Level[1]		0,94[4]	-[3]	-[3]	19,4
Household at Zonal Level[2]		0,96[4]	38,2	217,2	17,6

1 - Results reported by Fleet and Robertson
2 - Results obtained in this study
3 - These results were not reported
4 - Strictly speaking, these are simple coefficients of determination, not multiple coefficients as suggested by the column heading.

Table 4.7

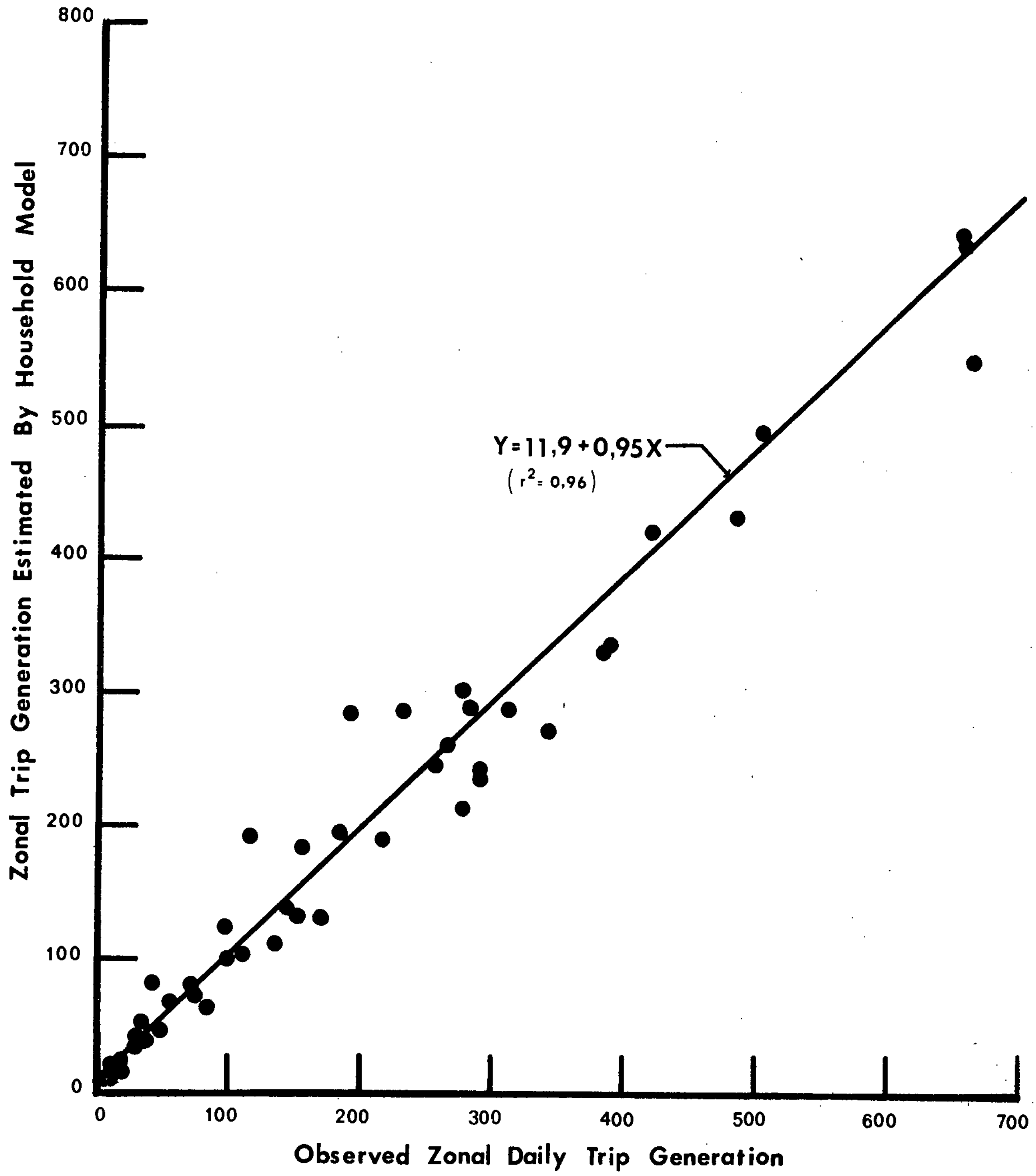


FIG. 4.10 Estimating Zonal Trip Generation Using A Household Level Model

4.5.3 Applying the Personal Equation at the Zonal Level

The preliminary personal equation is repeated here for convenience only.

$$\text{HBTPD} = 2{,}50 - 1{,}25\ \text{NOLICE} + 0{,}29\ \text{CARS} \qquad (4.1)$$

To apply this equation at the zonal level, the author inserted for the variable NOLICE the proportion of people (five years of age and older) in the zone not having a driver's licence, and for the variable CARS he inserted the average number of cars per household in the zone. The result was then multiplied by the number of people in the zone.

A least-squares linear regression analysis was carried out between the observed and predicted zonal trip ends, and the following results were obtained:-

$$y = 1{,}49 + 0{,}96x \qquad (4.8)$$
$$(r^2 = 0{,}968; \quad S_E = 34{,}4)$$

where y = zonal home-based trips per day predicted by the personal level equation,

x = observed zonal home-based trips per day.

We notice that the line defined by equation (4.8) is inclined at very nearly 45 degrees to the horizontal and virtually passes through the origin. In addition, the coefficient of determination is very high. Thus the personal level equation can provide an extremely good estimate of zonal trip ends. These results are depicted in Figure 4.11.

4.5.4 Summary

Table 4.8 summarizes the results of applying the preliminary household and personal level equations at the zonal level.

Description	r^2	S_E	$\frac{S_E}{\bar{Y}} \times 100$
Household Equation at Zonal Level	0,96	38,2	17,6
Personal Equation at Zonal Level	0,97	34,4	16,4

Table 4.8

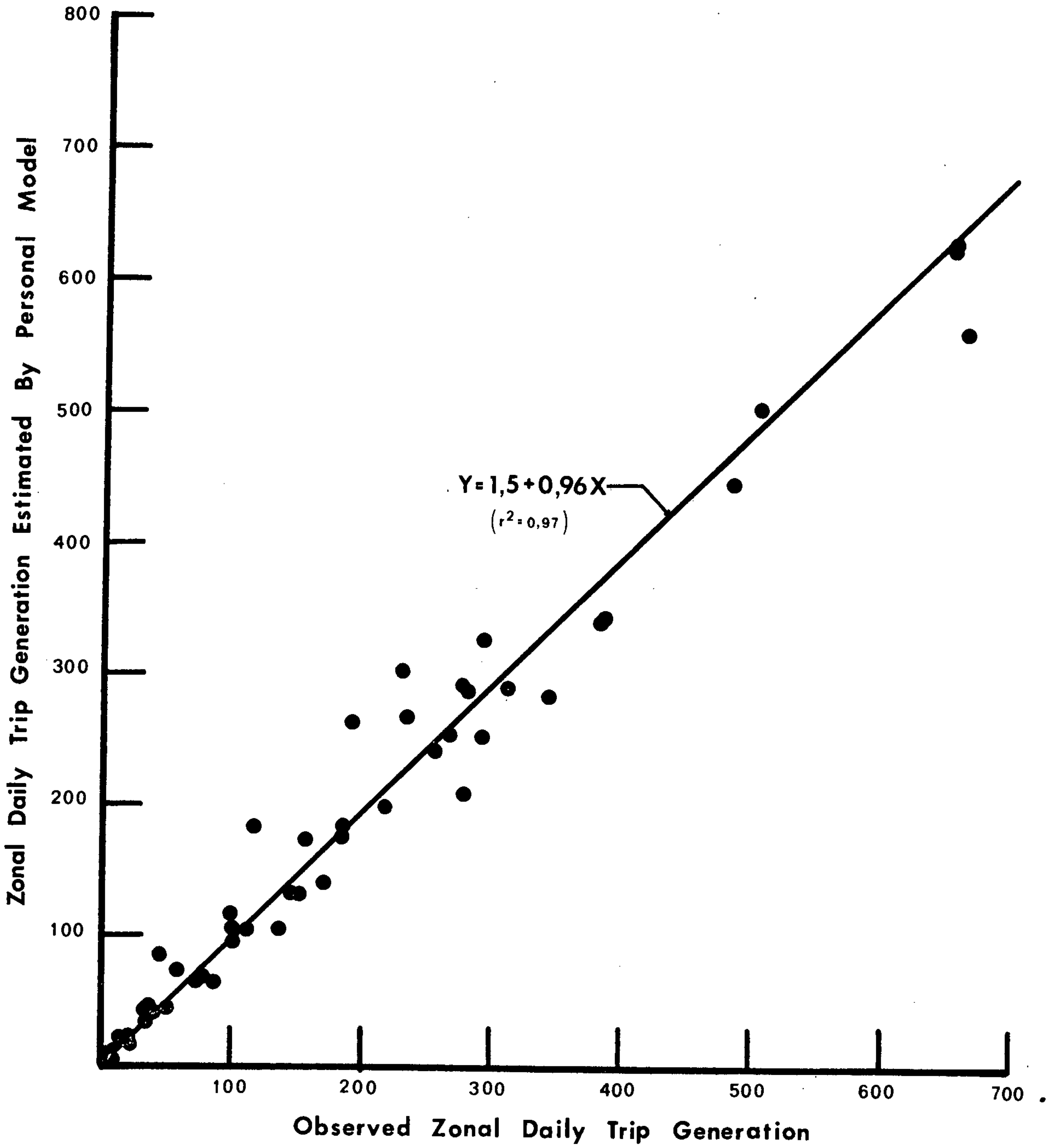


FIG. 4.11 Estimating Zonal Trip Generation Using A Personal Level Model

Hence the personal equation is marginally better than the household equation for predicting zonal home-based trip generation.

4.6 Comparative Analyses of the Preliminary Trip Generation Equations

4.6.1 Introduction

Various preliminary trip generation equations have been developed, and the effects of data aggregation have been investigated. The ability of the personal and household level equations to predict zonal trip generation has been confirmed. However, we must now investigate what proportion of the overall variation in trip-making is explained by the various preliminary models.

4.6.2 Comparing the Preliminary Household and Zonal Equations

The relevant preliminary household and zonal equations had the following coefficients of determination (R^2):-

Household (Total)	0,44	(Equation 4.2)
Zonal (Rate)	0,57	(Equation 4.4)

We have seen that approximately 75 per cent of the variation in trip generation is hidden within the zones, when household data is aggregated to the zonal level. This means that whereas the household equation explains approximately 44 per cent of the variation in trip-making between households, the zonal equation accounts for only about 14 per cent (24,2 × 0,57) of this variation.

In fact, noticing that the preliminary household equation explains 44 per cent of the variation in trip-making between households, and realizing that only 24,2 per cent of this variation is considered in a zonal analysis, we see that even a zonal model having a coefficient of determination equal to 1,00 would explain less of the between-household variation, than the household equation.

This clearly shows the futility of attempting to obtain a close-fitting equation at the zonal level, by including a large number of explanatory variables.

4.6.3 Comparing the Preliminary Personal and Household Equations

The preliminary personal trip generation equation had an R^2 of 0,153, while the preliminary household (rate) equation had an R^2 of 0,188. On this basis it would seem that the household model is marginally better (statistically) than the personal model.

However, when the individual's trip-making data is aggregated to the household level, the between sum of squares is only approximately 55 per cent of the total sum of squares. Thus the household equation explains only about 10 per cent (0,55 × 18,8) of the observed variation in trip-making between individiuals, compared with the 15 per cent explained by the preliminary personal equation.

4.6.4 Comparing the Preliminary Personal and Zonal Equations

Looking at the statistics reported earlier in this thesis for the personal equation and the second zonal rate equation (4.5), we would conclude that the zonal rate model is more likely to remain stable with time, since it appears to explain a larger proportion of the variation in trip generation between people. In addition, the zonal rate model has a lower percentage standard error of estimate.

However, we have already seen that when personal trip-making information is aggregated to the zonal level, we obscure almost 89 per cent of the between-person variation in trip generation. Thus the preliminary personal level equation explains 15 per cent of the variation in trip-making between individuals, while the relevant zonal rate equation explains only approximately 5 per cent (0,113 × 43,1) of this variation.

4.6.5 Summary

The following tables are presented to summarize the results reported in this section. Table 4.9 indicates the percentage of the between-person variation in trip generation explained by the various relevant preliminary models.

Level	R^2	Percentage Explained[1] (%)
Zone	0,431	4,9
Household	0,188	10,3
Person	0,153	15,3

1 - Refers to the between-person variation in trip-making

Table 4.9

On the other hand, taking into account the effects of aggregating data from the household to the zonal level, we have the following results:-

Level	R^2	Percentage Explained[1] (%)
Zone	0,570	13,8
Household	0,437	43,7

1 - Refers to the between-household variation in trip-making

Table 4.10

These results clearly illustrate that although one can obtain a larger coefficient of determination at a higher level of aggregation, yet the actual percentage of the variation which is explained is very much reduced at this higher level of aggregation.

4.7 Some Further Studies at the Household Level

4.7.1 Introduction

In this section of the thesis the author intends to report some of the household results not yet reported, and to examine the household level trip generation model in more detail.

The models derived thus far do not necessarily comply with the various assumptions of a least-squares regression analysis. In addition, we have not given any consideration to the actual form of the equations. Hence the author would not recommend them for use in forecasting future trip generation - at least, not before further study.

4.7.2 Household Socio-Economic Characteristics

Two of the household socio-economic characteristics used in the preliminary household level models will be studied in greater detail in this section; they are car ownership and household size. Figures 4.12 and 4.13 show frequency distributions for these two variables. Notice that both distributions are fairly highly skewed, as would be expected.

Figure 4.12 is particularly interesting, since it reveals that only 382 households, out of the total sample of 1 690 households (23 per cent), did not have a car available. We also see that 560 households, approximately one third of the sample, had two or more cars.

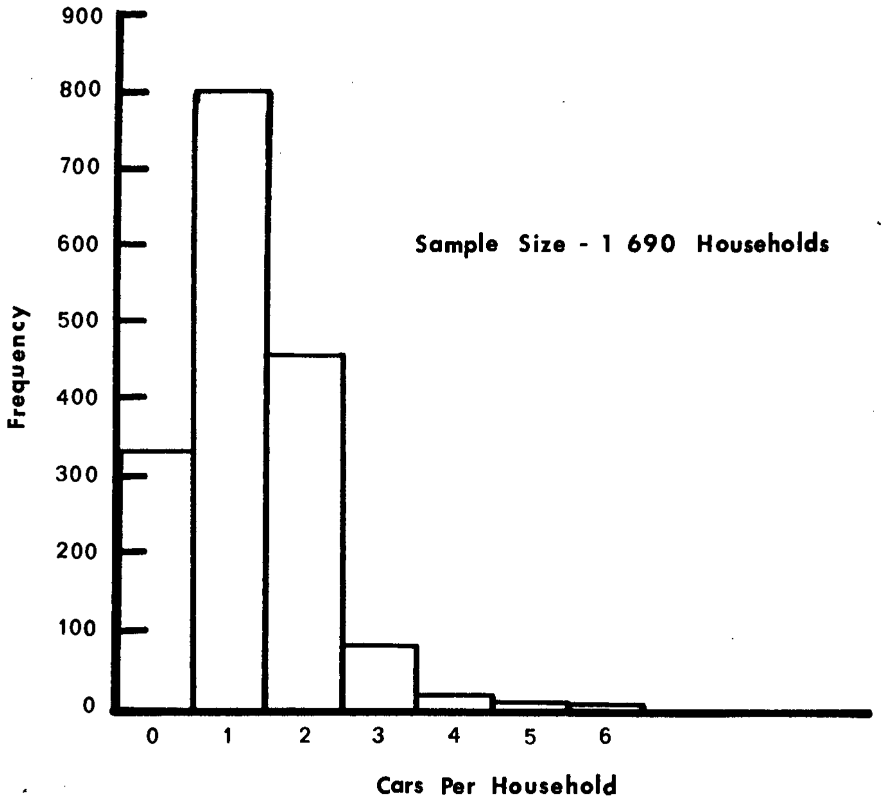


FIG. 4.12 Frequency Distribution, Household Car Ownership

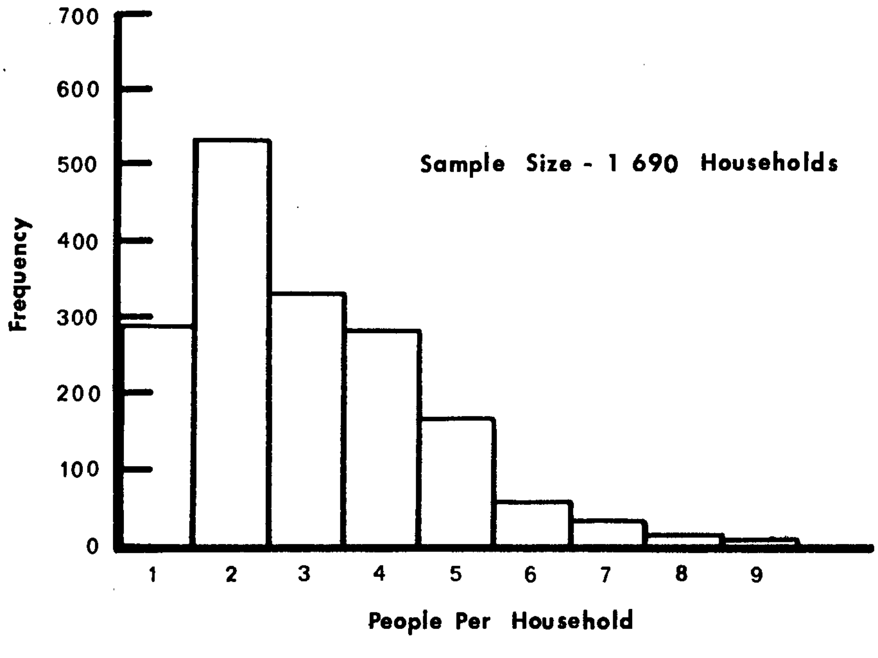


FIG. 4.13 Frequency Distribution, Household Size

The mean and standard deviation were computed for each of the household socio-economic characteristics. These results are tabulated below.

Variable	Description	Mean	Std. Deviation
FAMSIZ	People per household	2,93	1,55
UNDER5	Children under five years of age per household	0,22	0,55
5TO17	Children between five and seventeen per household	0,60	1,03
OVER18	Adults (eighteen and over) per household	2,11	0,83
CARS	Number of cars per household	1,21	0,88
M/C	Number of motor cycles per household	0,04	0,23
SKOLAR	Number of school pupils per household	0,58	1,00
UNIVER	Number of college/university students per household	0,10	0,38
DRIVLI	Number of driver's licences per household	1,44	0,91
EMPLOY	Number of employed people per household	1,23	0,85

Table 4.11

From the results shown in Table 4.11 one can estimate the number of cars per 1 000 people as

$$\frac{1{,}21}{2{,}93} \times 1\ 000 = 413.$$

This is not very far below the saturation level postulated by Tanner[14] for car ownership in Great Britain. It would seem that car ownership amongst the survey area population will not increase very much more, and will probably increase at a decreasing rate until a saturation level is reached.

4.7.3 Further Investigation of the Household Level Trip Generation Model

The preliminary household total equation (4.2) was derived by allowing the three variables FAMSIZ, CARS and EMPLOY to enter the regression in a stepwise manner. The variables were all allowed to enter the equation irrespective of their contribution to reducing the unexplained sum of squares. In other words, they were forced into the model. The following table summarizes the stepwise development of the equation:-

Step No.	Equation	$100R^2$	$\frac{S_E}{\bar{Y}} \times 100$
1	HBTHD = 0,44 + 2,00FAMSIZ	35,1	66,9
2	HBTHD = - 0,66 + 1,60FAMSIZ + 1,86CARS	43,4	62,6
3	HBTHD = - 0,84 + 1,53FAMSIZ + 1,75CARS + 0,43EMPLOY (21,8) (14,4) (3,4)	43,8	62,4

Note: The t-value is shown below each independent variable at step No. 3

Table 4.12

Looking at the above table we see that the inclusion of the variable EMPLOY, once FAMSIZ and CARS have already been included, does not seem to make much difference to the model's statistics. However, such arbitrary decisions are not acceptable, and the effect of the variable EMPLOY was evaluated statistically (See Appendix 5). The analysis showed that this variable made a significant contribution to the regression sum of squares, at a 1 per cent level of significance.

In addition to the above test, the author investigated the simple correlations between the variables in the above equation. The table below contains data extracted from the correlation matrix provided by the BMDO2R program.

		(1)	(2)	(3)	(4)
HBTHD	(1)	1,00	0,593	0,490	0,369
FAMSIZ	(2)	-	1,00	0,378	0,417
CARS	(3)	-	-	1,00	0,377
EMPLOY	(4)	-	-	-	1,00

Table 4.13

From the above table it can be seen that there is a fairly high correlation between FAMSIZ and EMPLOY. This is larger than the correlation between the dependent variable (HBTHD) and EMPLOY. Accepting the criterion suggested by Douglas and Lewis[15] (see section 2.5.6), we would not include both FAMSIZ and EMPLOY in the same regression equation, although both are significant in explaining the observed variation in the dependent variable.

As a result of the above considerations it was decided to study only FAMSIZ and CARS in more detail. This decision was also influenced to a large extent by the findings of Oi and Shuldiner[16].

The multiple linear regression equation used in the following investigations is that shown as step 2 in Table 4.12, i.e.

$$\text{HBTHD} = -0{,}66 + 1{,}60\text{FAMSIZ} + 1{,}86\text{CARS} \qquad (4.9)$$
$$(R^2 = 0{,}434; \quad S_E = 3{,}94)$$

The overall significance of this regression equation was evaluated by comparing the F-ratio of 647 to $F_{0,01;\ 2;\ 1\ 687} = 4{,}6$. (This test is described in Appendix 1). Hence, the equation was found to be highly significant at a 1 per cent level of significance.

The next aspect to be investigated was the form of the relationship between the independent variables and the dependent variable. As a first step, the author cross-classified the 1 690 households into 20 groups, and the mean number of home-based trips per household per day was computed for each group. The results of this category analysis are shown in Table 4.14.

		FAMSIZ				
		1	2	3	4	5+
CARS	0	1,39 (141)	2,30 (102)	3,40 (52)	6,00 (11)	6,65 (26)
	1	2,94 (135)	4,71 (309)	5,78 (145)	7,59 (117)	10,24 (92)
	2	2,88 (8)	5,36 (114)	7,52 (114)	10,34 (112)	12,48 (109)
	3+	4,00 (1)	5,00 (6)	8,29 (17)	11,39 (39)	15,97 (40)

The figures in parentheses indicate the number of households in each category.

Table 4.14

These results are depicted graphically in Figures 4.14 and 4.15. In order to plot these graphs, the mean number of people in households having five or more members, and the mean number of cars in households possessing three or more cars, had to be calculated. The data shown in Figures 4.12

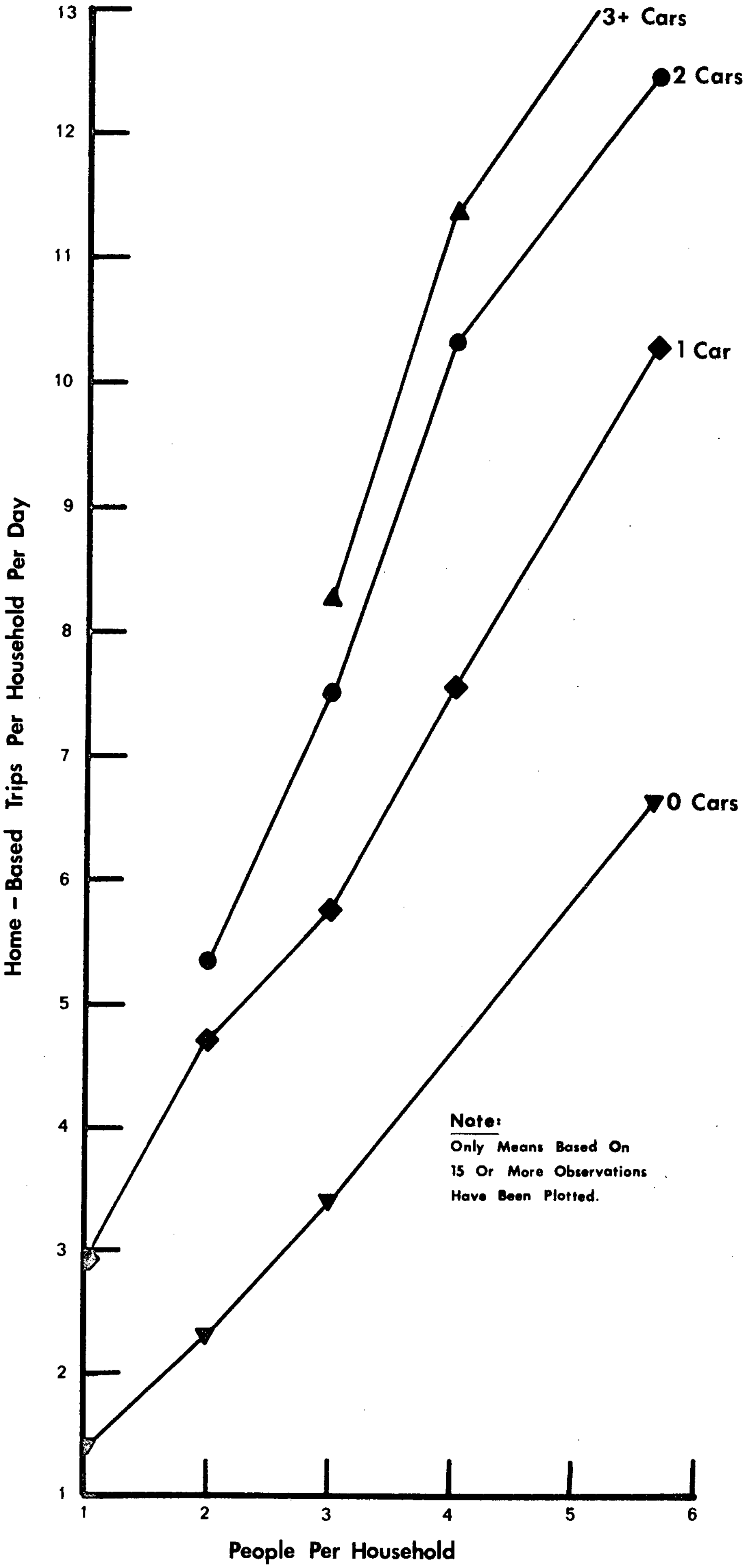


FIG. 4.14 Effect Of Household Size For Given Car Ownership

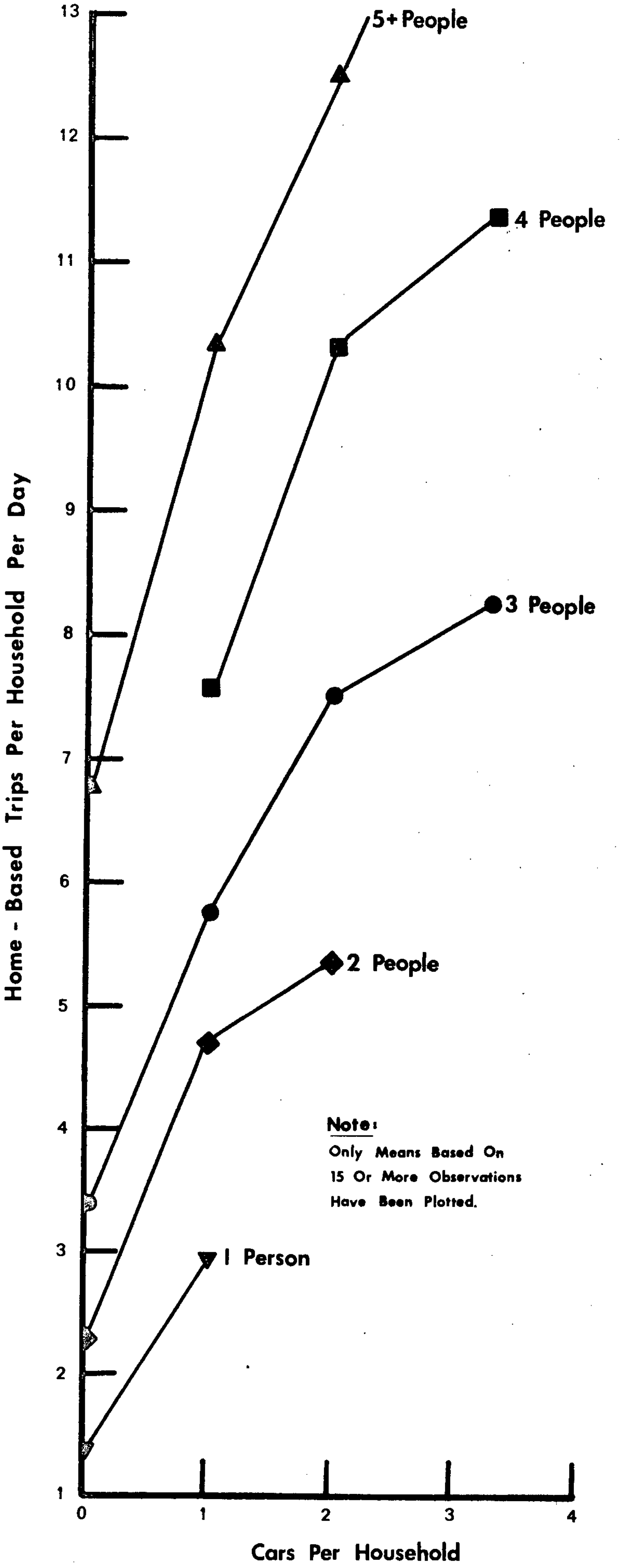


FIG. 4.15 Effect Of Car Ownership For Given Household Size

and 4.13 was used to calculate the following:-

Mean number of people in households having 5 or more members = 5,64
Mean number of cars in households having 3 or more cars = 3,30

Studying the results shown in Figure 4.14 we see that persons per household appears to be approximately linearly related to home-based trips per household per day, allowing for variations in car ownership.

Also, for a given household size, the acquisition of the first car has a greater impact on home-based trips per household per day, than the addition of a second or third car. This effect is particularly noticeable for small household sizes; but it is very important, since it indicates that the relationship between home-based trips per household per day and car ownership could be significantly non-linear, allowing for variations in household size.

The effect of car ownership on trip-making, for different household sizes, was studied next. The relevant category analysis results are shown in Figure 4.15. We again notice that the non-linear effect of car ownership on trip generation is more evident for smaller households than for the larger ones.

Before further examination of the effect of household size and car ownership on household trip generation, the Cape Town data was compared with the Modesto (California) data reported by Oi and Shuldiner. These comparisons are illustrated in Figures 4.16 and 4.17. We see that the Cape Town and Modesto data reveal the same trends, although the Cape Town trip generation rates tend to be slightly higher than the Modesto results, for the large values of the independent variables.

The next step in the household analysis, was the use of two dummy variable sets to replace the variables FAMSIZ and CARS. These two variables were stratified into a number of classes, corresponding to the classes used in the category analysis reported earlier. The following dummy variables were used:-

F1 = 1 when FAMSIZ = 1, 0 otherwise
F2 = 1 when FAMSIZ = 2, 0 otherwise
F3 = 1 when FAMSIZ = 3, 0 otherwise
F4 = 1 when FAMSIZ = 4, 0 otherwise

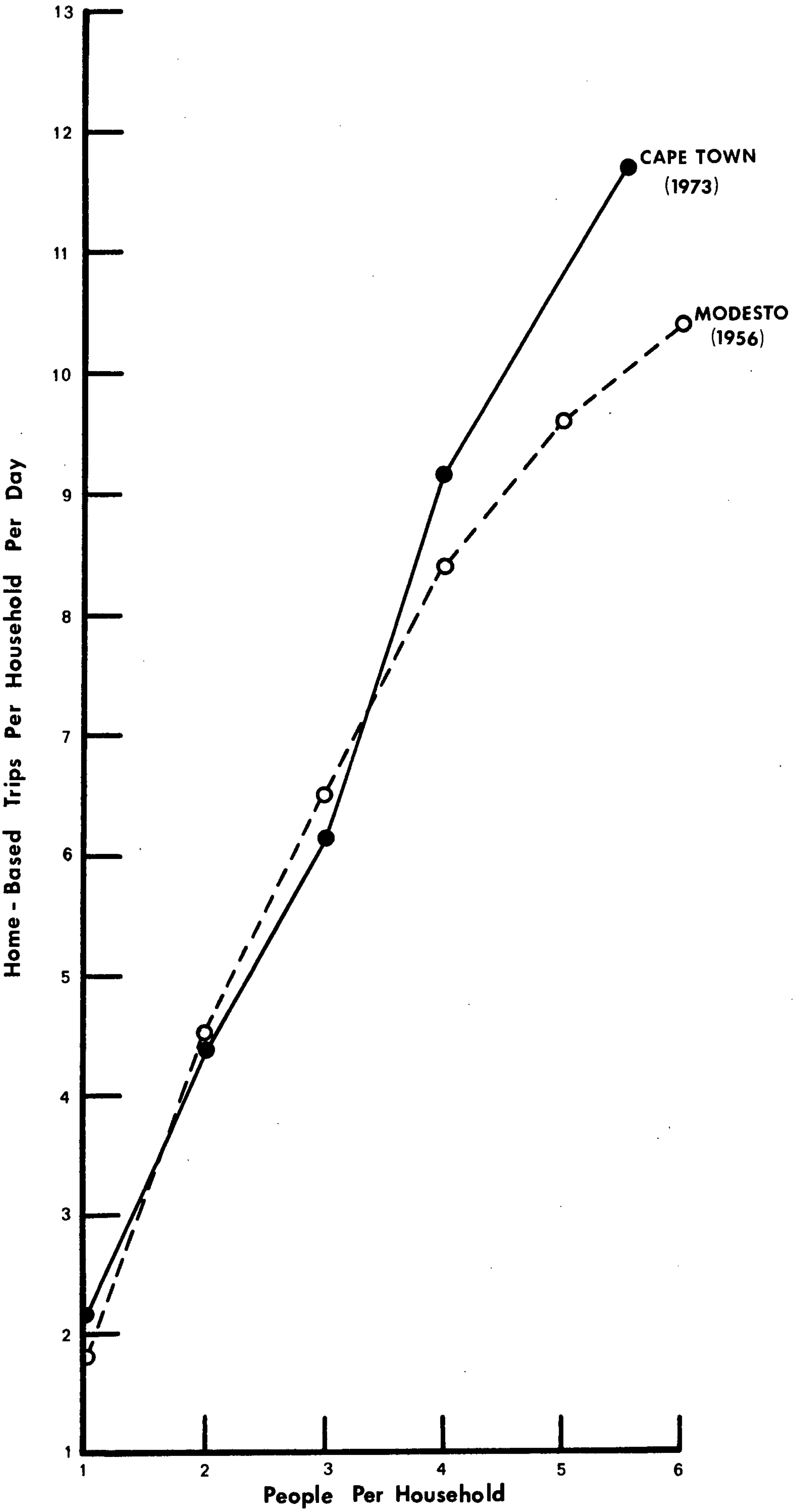


FIG. 4.16 Effect Of Household Size On Household Trip Generation

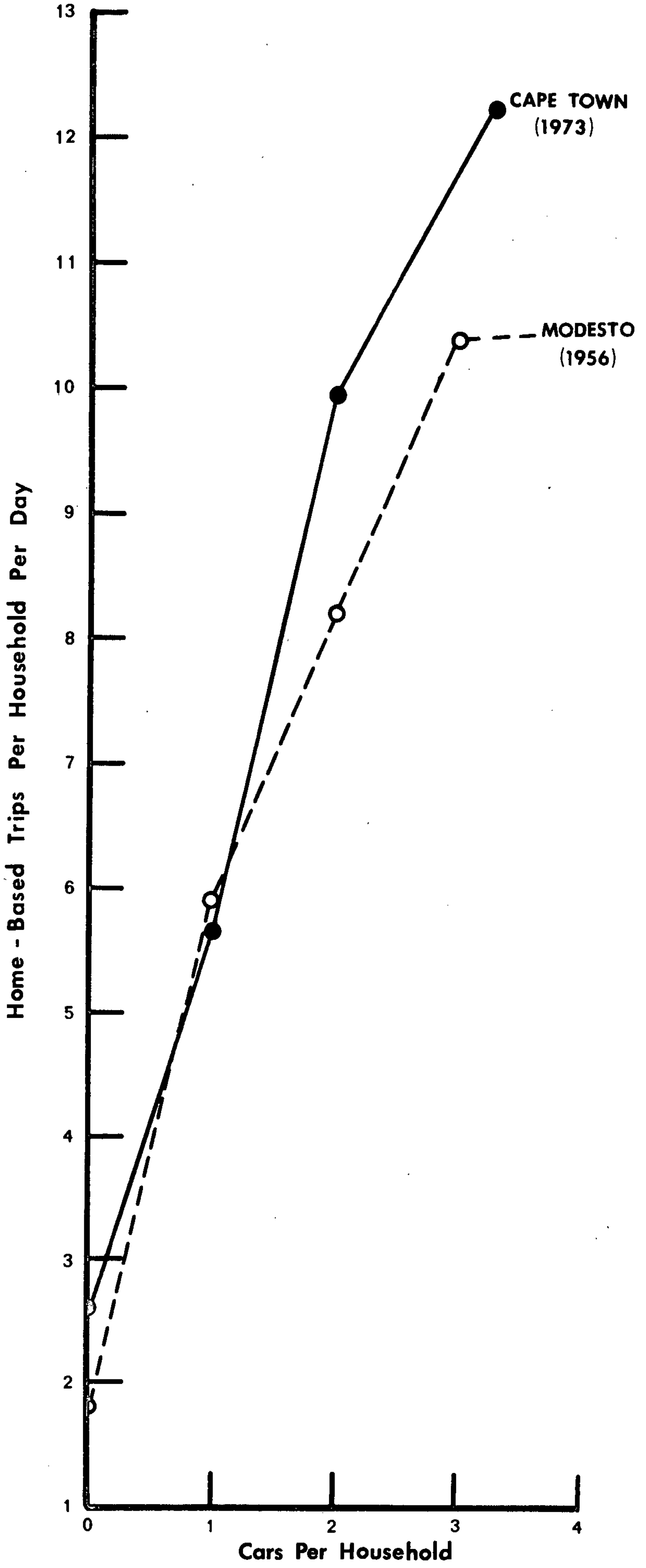


FIG. 4.17 Effect Of Car Ownership On Household Trip Generation

F5 = 1 when FAMSIZ ≥ 5, 0 otherwise
C1 = 1 when CARS $= 0$, 0 otherwise
C2 = 1 when CARS $= 1$, 0 otherwise
C3 = 1 when CARS $= 2$, 0 otherwise
C4 = 1 when CARS ≥ 3, 0 otherwise.

The dummy variables F5 and C4 were not allowed to enter the regression, and the following equation was obtained:-

$$\begin{aligned} \text{HBTHD} = {} & 14{,}45 - 7{,}46F1 - 6{,}23F2 - 4{,}94F3 \\ & - 2{,}60F4 - 5{,}96C1 - 3{,}82C2 - 2{,}12C3 \qquad (4.10) \\ & (R^2 = 0{,}435; \quad S_E = 3{,}95) \end{aligned}$$

The F-ratio for this model was 185, which is far greater than $F_{0,01;\ 7;\ 1\,682} = 2{,}64$. Hence the regression equation is highly significant at a 1 per cent level of significance.

Comparing the R^2 and S_E values for this model, with those reported for equation 4.9, reveals that there is virtually no difference between the two models, as far as explanatory or predictive power is concerned.

The interpretation of equation 4.10 is fairly simple. As we have mentioned earlier, the coefficient of a dummy variable must be interpreted relative to the excluded class (of the set to which that dummy variable belongs). For example, car ownership being equal, a two-member household will, on the average, make 6,23 home-based trips per day less than a household having five or more members (since F5 was the excluded class). Similar interpretations can be applied to the other coefficients in the equation.

The results obtained using the least-squares regression approach (with and without dummy variables), and the results of the category analysis, are compared in Figures 4.18 and 4.19.

Very little difference is seen between the two alternative least-squares models, except for large values of the independent variables. Hence it would appear that the introduction of the dummy variables did not affect the model significantly.

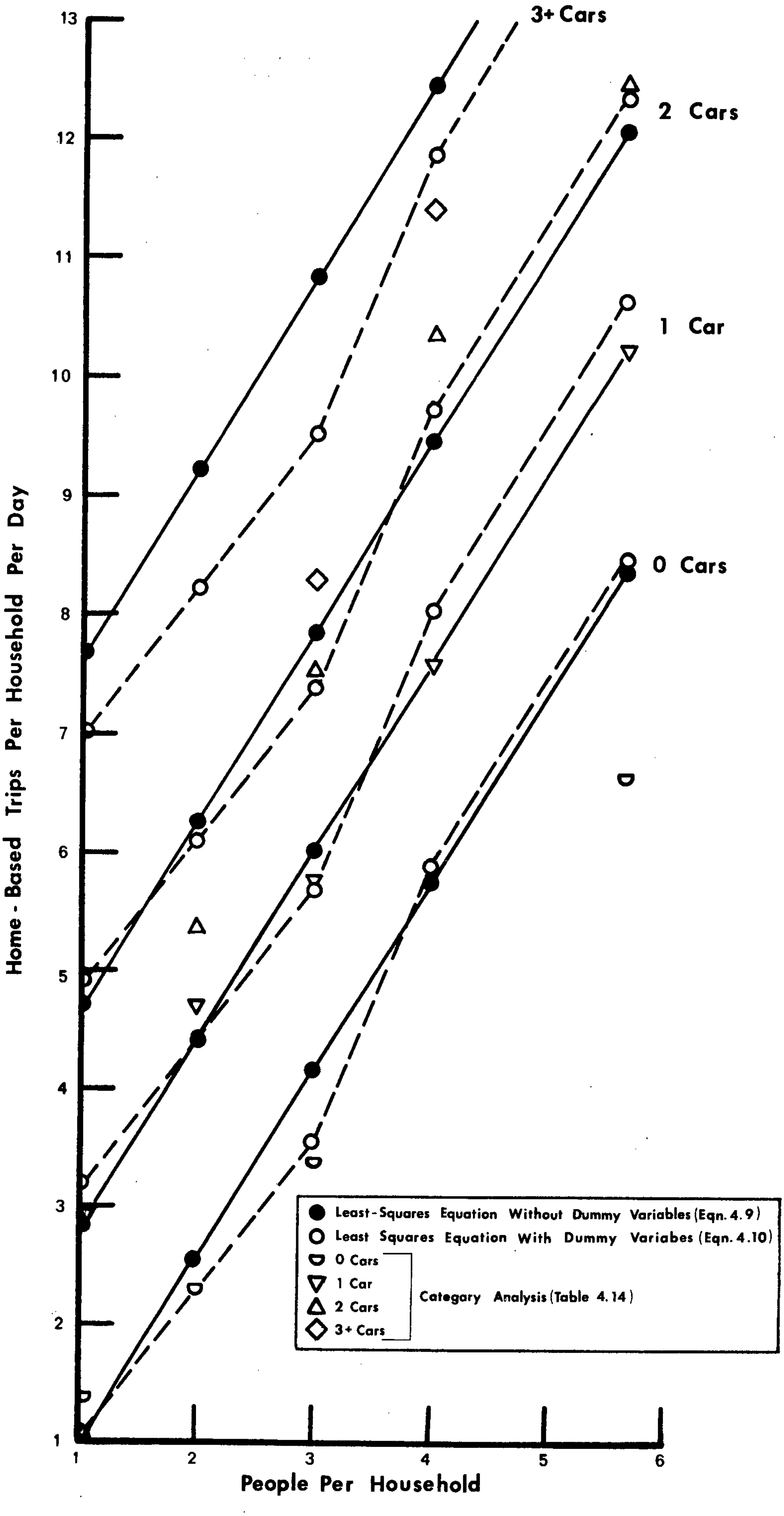


FIG. 4.18 Household Least-Squares And Category Analysis Results

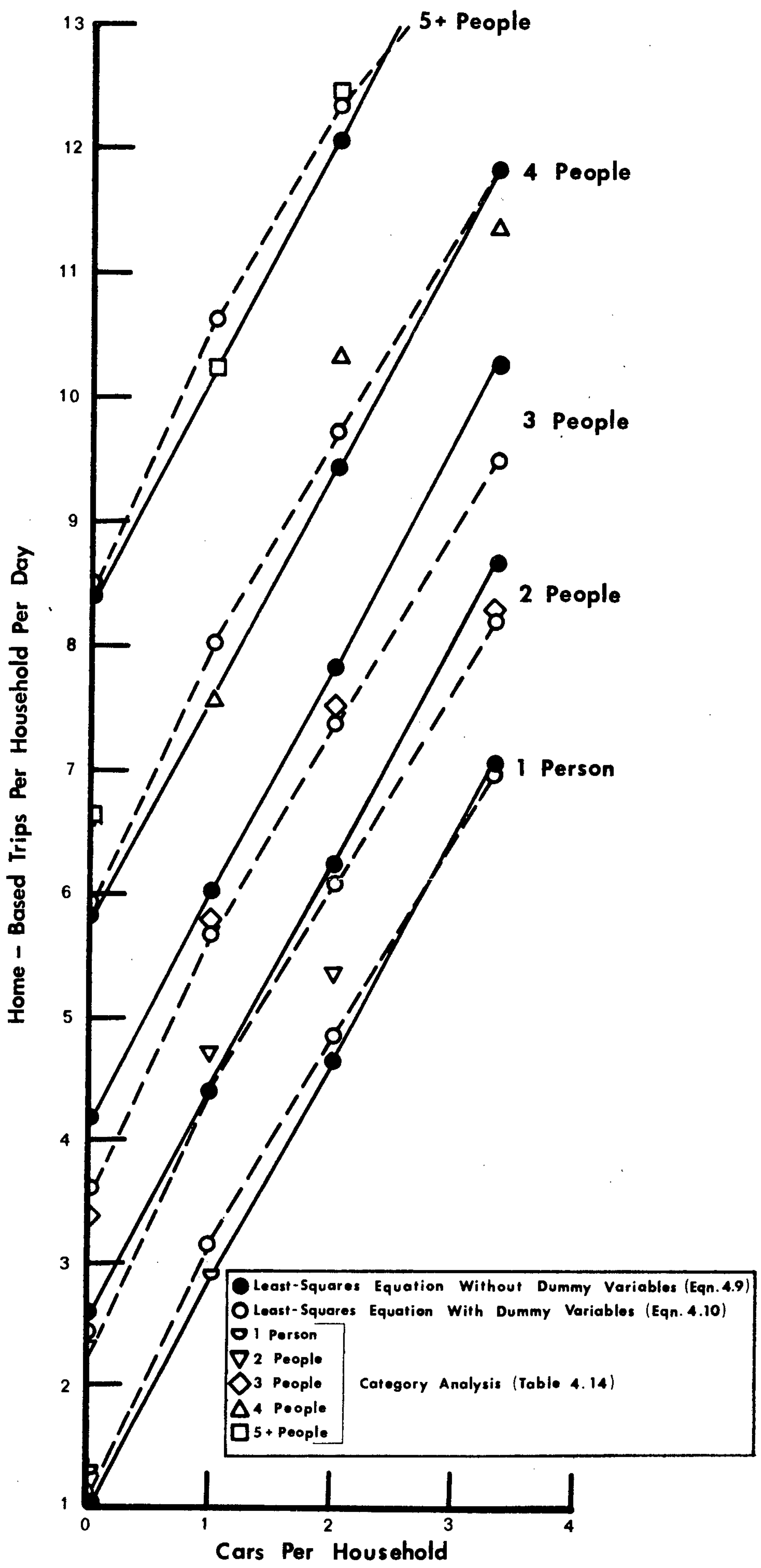


FIG. 4.19 Household Least-Squares And Category Analysis Results

However, Figures 4.18 and 4.19 both reveal a noticeable difference between the cross-classification results and the least-squares regression results. The major differences occur for large values of the independent variables, while in the intermediate range of the independent variables there is a good comparison between the results obtained using the different techniques.

As a result of the above considerations, it was decided to investigate equation (4.9) in more detail. Figure 4.20 shows a frequency chart for the residual error of this model. From this figure it can be concluded that the residuals approximately satisfy the zero mean and normality assumptions.

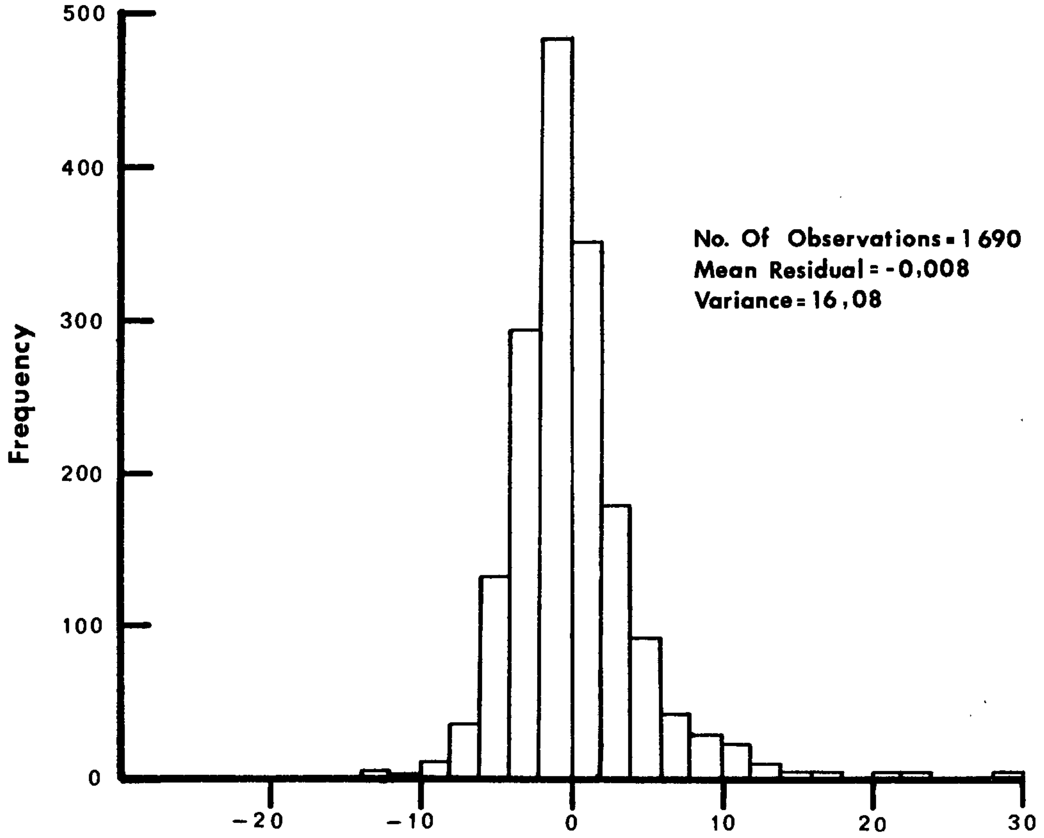
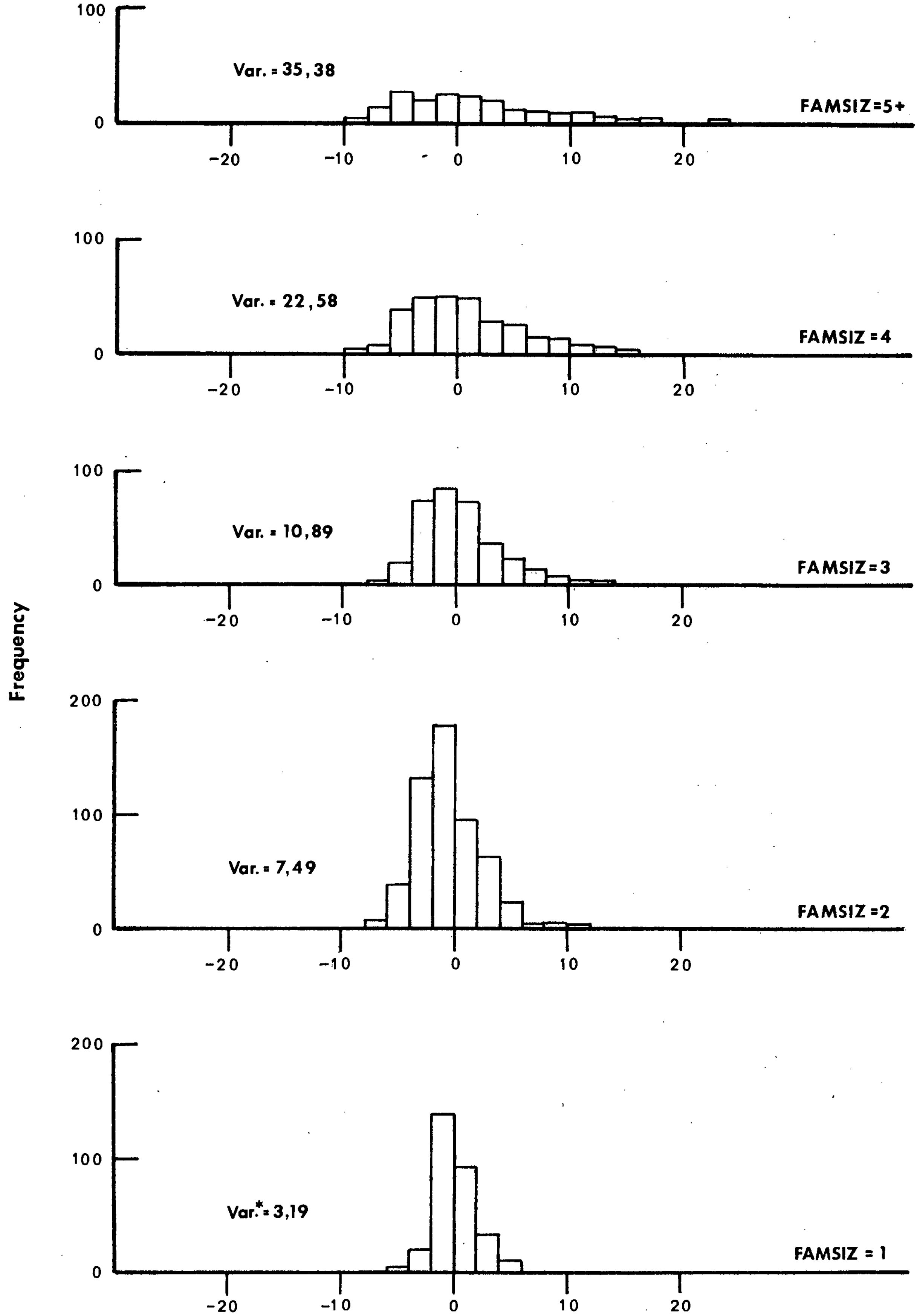


FIG. 4.20 Residual Error Term - Household Model



* Variance

FIG. 4.21 Residual Error Term – Household Model

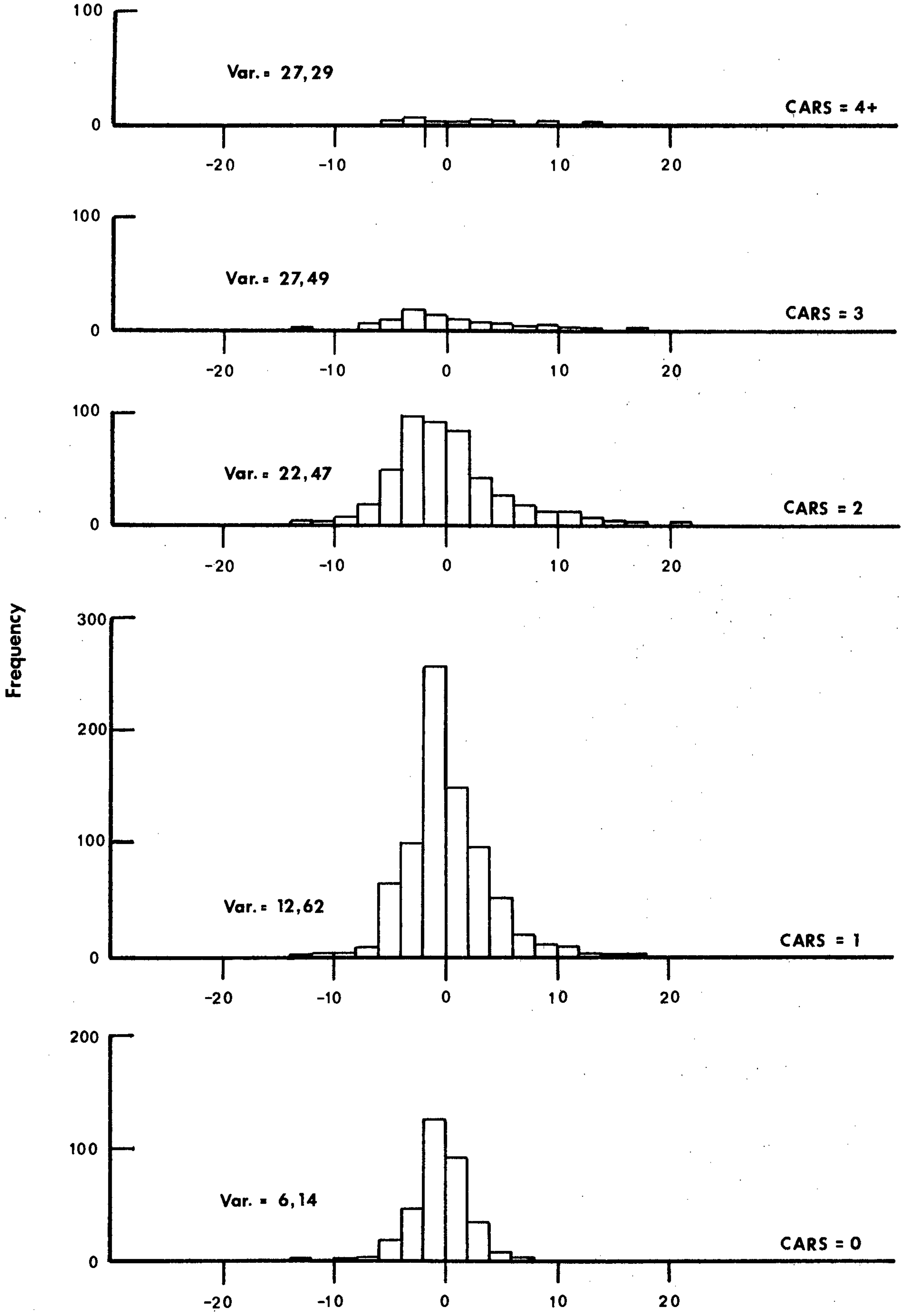


FIG. 4.22 Residual Error Term – Household Model

The residuals, plotted for different values of the two independent variables, are shown in Figures 4.21 and 4.22. From these figures the author concluded that the constant error variance assumption is not met and the condition of heteroscedasticity exists.

As a result of the above investigations, equation (4.9) was accepted as a satisfactory model of home-based trip generation, with the reservation that the constant error variance assumption is not satisfied. However, Douglas and Lewis[17], as we have already pointed out, indicate that the accuracy of the model will be overstated if the condition of heteroscedasticity exists, but there is nothing intrinsically wrong with the model itself.

The two explanatory variables in the household model, and their effects on residential trip generation, have already been discussed in detail (see section 2.4), and hence the model has been developed within a logical framework.

The interpretation of the model is straightforward. If household size increases by one unit, while car ownership remains constant, then home-based trips per household per day will increase by 1,60.

Of greater interest, is the effect of increasing car ownership, given that the household size remains constant. In this case, a one unit increase in car ownership, will cause an increase of 1,86 in the dependent variable. The average household car ownership for the survey area was found to be 1,21, or 413 cars per 1 000 people. Assuming this increases to a saturation level of about 520 cars per 1 000 people, the average household car ownership would increase to 1,52 (if the average household size remains constant). This would increase the average number of home-based trips per household per day by

$$1{,}86 \times (1{,}52 - 1{,}21) = 0{,}58 \text{ trips.}$$

In other words, the average number of home-based trips per household per day would increase from 6,29 to 6,87. This is not a large change, but since the car ownership level amongst the study area population is already high, we would expect only a small increase in trip generation as a result of increasing car ownership.

Another feature of equation (4.9) is that the constant term is only approximately 10 per cent of the mean of the dependent variable, and this is indicative of a stable equation.

Although other explanatory variables probably deserve further consideration, it is felt that such research falls outside the limited scope of this thesis. However, the necessary data has been collected and the author intends carrying out further studies at the household level in the future.

4.8 Further Studies at the Personal Level

4.8.1 Introduction

The development of the preliminary trip generation equation at the personal level has been discussed and it has been shown that this model can be successfully used for predicting zonal trip ends. We have, however, not had a close look at the form of the equation, nor have we examined the characteristics of the individual people who responded to the survey. These topics will be covered in this section.

4.8.2 Personal Characteristics

The 4 571 people who completed the questionnaire were studied with regard to their age, sex, and the possession (or otherwise) of a driver's licence. The following table summarizes the results of this investigation.

Age Group	Male	Female
Between 5 and 17 years of age.	527	481
Eighteen years of age and over.	1 636	1 927
Eighteen and over, having a driver's licence.	1 433	1 017
Eighteen and over, not having a driver's licence.	203	910

Table 4.15

Looking at the results above one notices that the proportion of males (eighteen years and over), having a driver's licence, appears to be much larger than the comparable proportion for females. In fact, almost 88 per cent of eligible males have a driver's licence; while the comparable figure for females is 53 per cent.

The author decided to test whether this difference was significant. The test for differences in proportions was used (see Appendix 6), and it was concluded that the proportion of eligible males having a driver's licence is significantly greater than the comparable proportion for females, at a 1 per cent level of significance.

Since we would expect the possession (or otherwise) of a driver's licence to have a notable effect on modal choice, it is not surprising that Aplin[1] found that the proportion of females in a zone was a significant explanatory variable in a zonal modal split model.

4.8.3 Further Investigation of the Personal Trip Generation Model

The preliminary personal level trip generation equation, reported in section 4.3.3, was as follows:-

$$\text{HBTPD} = 2{,}50 - 1{,}25\text{NOLICE} + 0{,}29\text{CARS} \qquad (4.1)$$

$$(R^2 = 0{,}153; \quad S_E = 1{,}75)$$

The variables in equation (4.1) were the first to enter during the stepwise process. Both the variables are extremely significant in explaining variations in the dependent variable. The partial F-values for NOLICE and CARS respectively were 529,7 and 97,5; while $F_{0,01;\ 1;\ 4\,568} = 6{,}6$.

The two explanatory variables in equation (4.1) could be expected to be fairly highly correlated with each other. From the printout of the BMD02R program, the correlation between the explanatory variables was found to be - 0,287, while their individual correlations with the dependent variable were - 0,367 and 0,234 respectively.

The criterion mentioned earlier in regard to multi-collinearity is not quite satisfied. However, it is not very seriously violated, and it was decided to retain both the above explanatory variables in our further investigations.

The next step was to carry out a category analysis at the personal level, in terms of the above explanatory variables. Table 4.16 shows the results of this analysis.

		CARS				
		0	1	2	3	4
NOLICE	1	1,19 (564)	1,56 (870)	1,85 (549)	2,07 (98)	3,42 (12)
	0	1,73 (100)	2,87 (1096)	3,11 (963)	3,23 (228)	3,75 (57)

The figures in parentheses indicate the number of people in each category.

Table 4.16

In order to have a better picture of the above results Figure 4.23 was drawn, and only means based on fifteen or more observations were plotted. Also in Figure 4.23, the two straight, parallel lines defined by equation (4.1) have been plotted.

The two analyses give very comparable results for the sub-sample of people not having a driver's licence. In the case of people possessing a driver's licence the two methods yield fairly similar results, except for people belonging to households having no cars.

Looking at equation (4.1), the reason for this large discrepancy can easily be seen. The model specifies a large difference in trip-making between people having a driver's licence and those not having a driver's licence. Obviously, the potential benefit of a driver's licence can only be realized by those people having possible access to a car. Therefore, the least-squares equation overstates the trip generation of licensed drivers belonging to zero-car households.

The distribution of the residual error terms was investigated, to see whether the personal model satisfied the assumptions of the least-squares technique better than the household model.

The relevant histograms and statistics are shown in Figures 4.24 and 4.25. From these figures it can be seen that the least-squares assumptions (with respect to the residuals), are fairly well satisfied in the case of the personal trip generation model.

As in the case of the household model, other variables need to be examined at the personal level, but these investigations fall beyond the scope of this thesis. The author hopes to continue this work in the future.

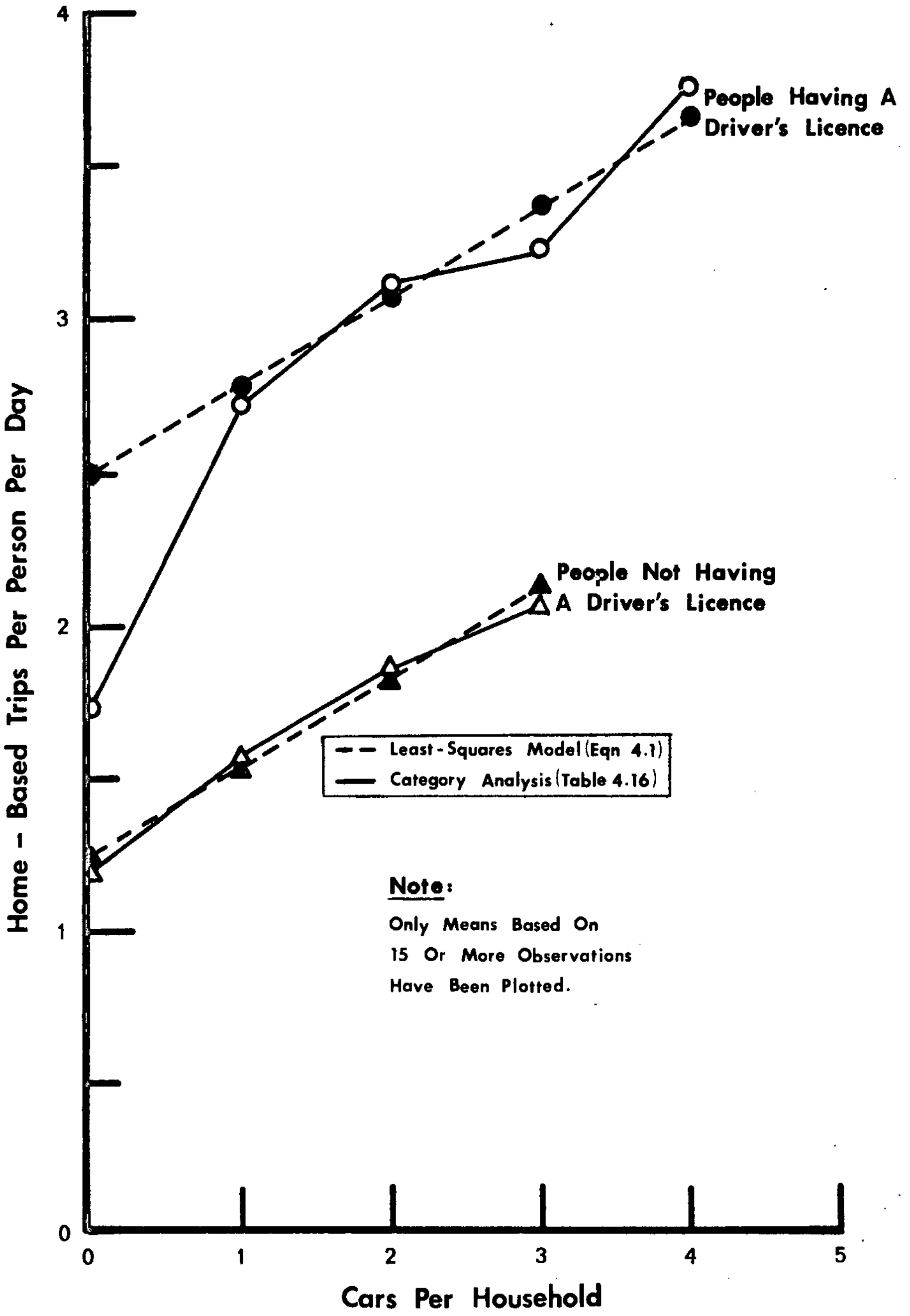


FIG. 4.23 Personal Least-Squares And Category Analysis Results

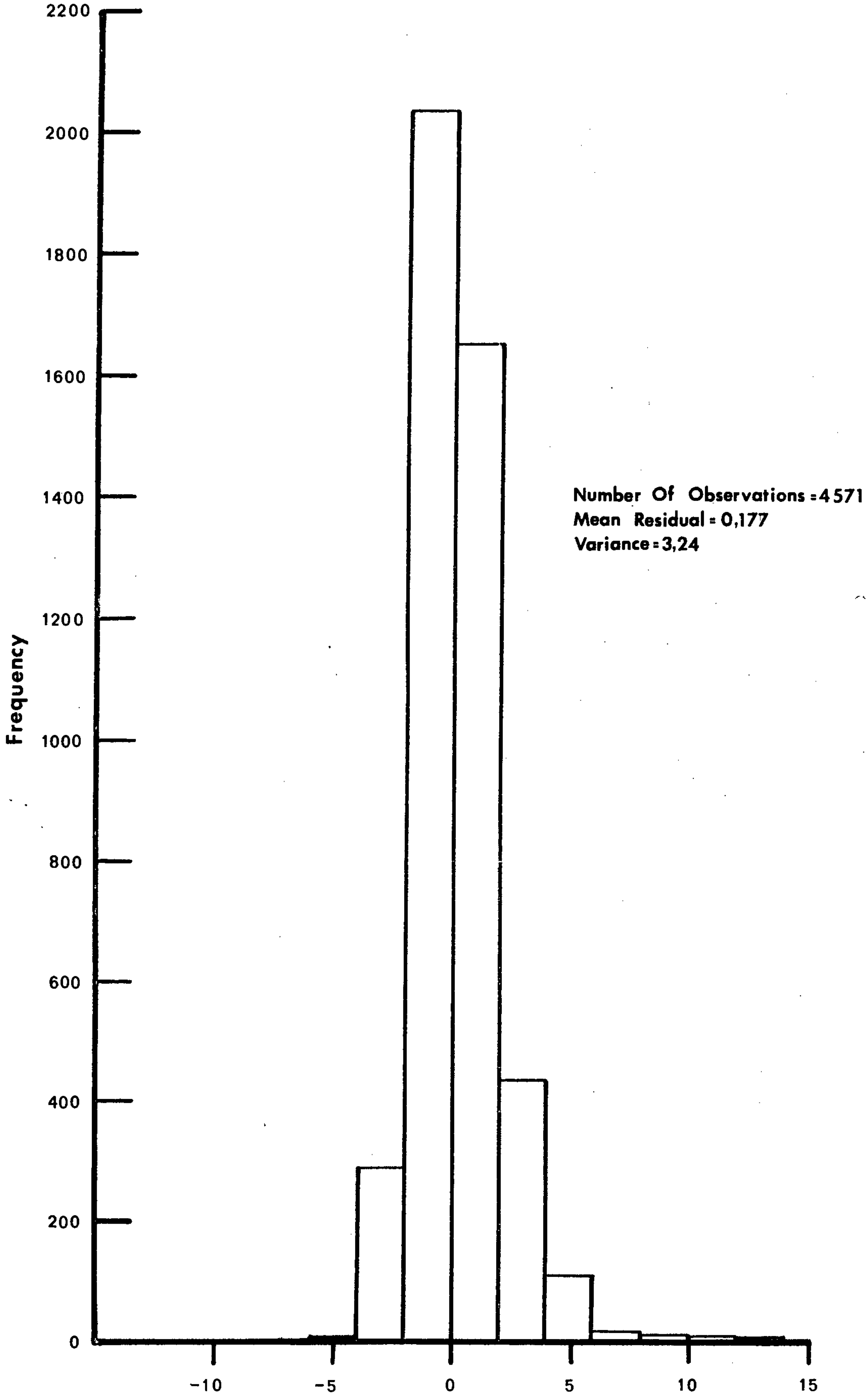


FIG. 4.24 Residual Error Term – Personal Model

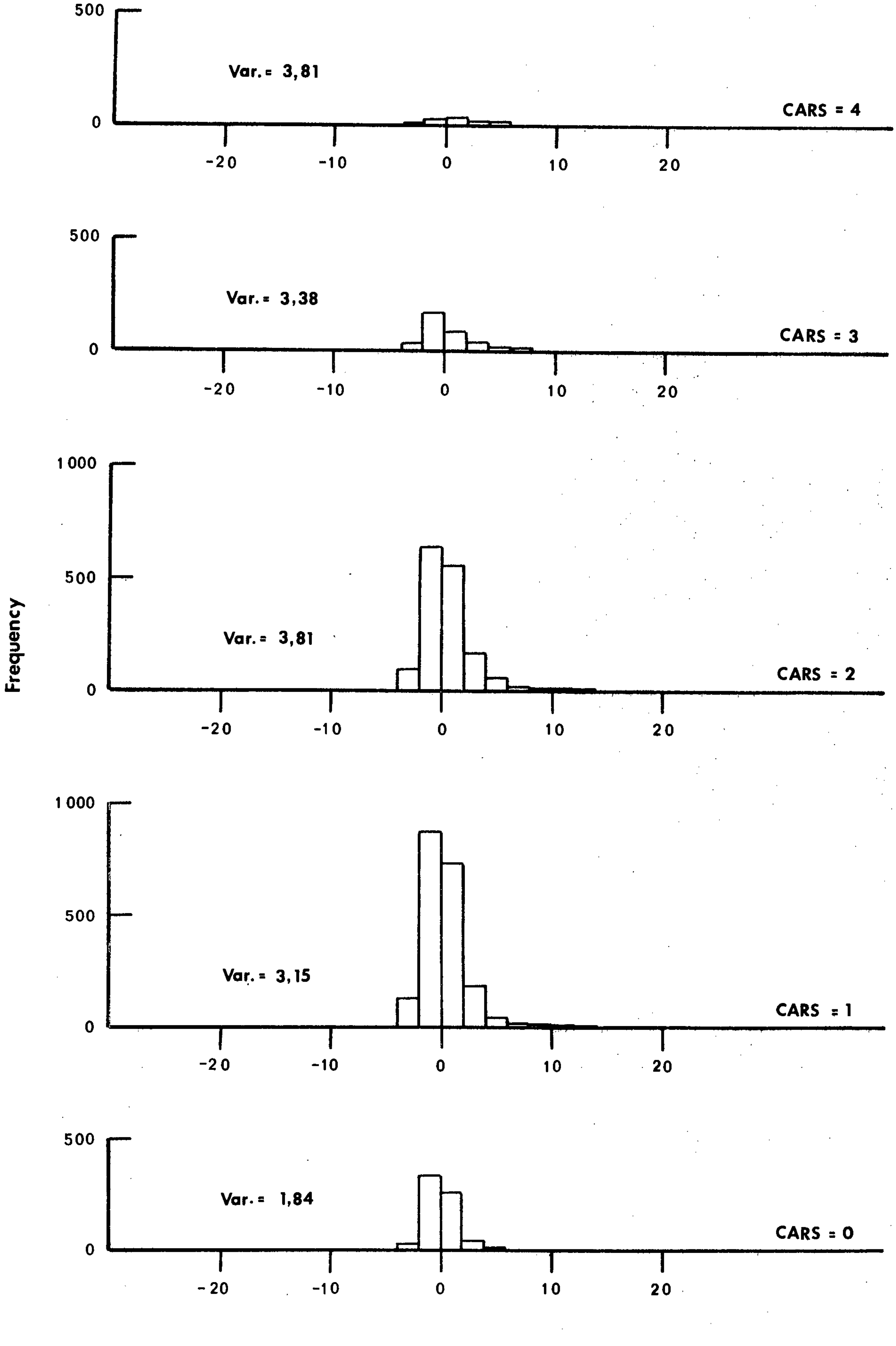


FIG. 4.25 Residual Error Term – Personal Model

4.9 Comparison with other Study Data

We have already seen a reasonable comparison between some of the Cape Town and Modesto results (Figures 4.16 and 4.17). In this section some of the Cape Town results are compared with those found in four other study areas. The results from the other areas were presented in the London Traffic Survey report[18].

In addition to comparing the results obtained in the various studies, the Cape Town household and personal models were used to estimate the average number of home-based trips per household per day, for the other study areas. These predictions, and the other comparisons, are tabulated in Table 4.17.

In order to use the personal trip generation model to predict household trip generation in the other study areas, two assumptions had to be made, because of the limited data available to the author. Firstly, it was assumed that the ratio

$$\frac{\text{average number of driver's licences per household}}{\text{average number of cars per household}}$$

was the same for all the study areas. Secondly, the ratio

$$\frac{\text{average number of children under 5 per household}}{\text{average number of people per household}}$$

was assumed to be the same for all the study areas. The author is of the opinion that both these assumptions are fairly reasonable ones.

From Table 4.17 it can be seen that the level of car ownership in the Cape Town study area exceeds the level in the other areas. However, we must remember that the other studies were carried out many years before the Cape Town study. It is also evident, from the results reported in Table 4.17, that the average household size does not vary much from one area to the next.

The most interesting results reported in Table 4.17 are the predicted home-based trips per household per day for the different study areas. We see that the personal trip generation model is able to make better estimates of the average household trip generation rates for the other study areas, than the household trip generation model. However, even the personal model can only produce estimates that are generally 25 to 30 per cent higher than

Study Area	Year of Study	Persons per Household	Persons per Car	Cars per Household	Trips per Person	Home-Based Trips per Household[1]	Home-Based Trips per Household[2]	Home-Based Trips per Household[3]
Detroit, Michigan	1953	3,31	3,51	0,94	1,77	4,67	6,38	6,05
Greater London	1962	2,87	6,81	0,42	1,54	3,38[4]	4,71	4,27
Kansas City, Missouri	1957	3,07	3,26	0,95	2,18	5,14	6,02	5,77
Washington D.C.	1955	3,02	3,75	0,81	1,67	4,23	5,68	5,30
Cape Town	1973	2,93	2,42	1,21	2,88[5]	6,29	-	-

1 - as reported in the survey,
2 - predicted by the household trip generation model (Equation 4.9),
3 - predicted by the personal trip generation model (Equation 4.1),
4 - unlinked trips,
5 - trips per person five years of age or older.

Table 4.17

the reported household trip generation rates. These differences could possibly be the result of peculiar local factors.

NOTES - CHAPTER IV

1. Aplin, W.N., Modal Split Analysis for the Journey to Work, Thesis submitted in partial fulfilment of the requirements for the M.Sc.(Eng.) Degree, Department of Civil Engineering, University of Cape Town, 1974.

2. Information kindly supplied by Messrs. K. Sturgeon and P. Tomlinson of the Advance Planning Section, City Engineer's Department, Municipality of Cape Town.

3. The assistance provided by Mrs. Ensor and Mrs. Swann is acknowledged at the beginning of this thesis.

4. Each sample household was allocated a code, by means of which the author could identify the interviewer, determine whether the dwelling was a house or flat, and determine the household's address. These codes were recorded on the forms by the interviewers when they collected the completed questionnaires from the respondents.

5. 'Datafile' is the term used for a collection of records stored on a mass-storage device (e.g. magnetic tape or disk) on the computer.

6. In section 3.5.3 we mentioned that people such as doctors, taxi-drivers, service-men, etc. were not expected to record the details (purpose, mode and time) of their numerous non-home-based trips.

7. Zonal total and rate variables were defined in section 2.3.2.

8. These estimates are based on the assumption that the car passenger trips made by respondents in cars belonging to other households (not in the survey), are equal (in number) to the car passenger trips made by other people (not in the sample) in cars belonging to the sample households.

9. Dixon, W.J. (ed.), Biomedical Computer Programs, University of California Press, 1968.

10. 'Person' here obviously means people five years of age or older, since we did not collect personal and trip-making data for people younger than five.

11. Similarly, zonal characteristics (such as residential density) have been used in household level trip generation analyses.

12. A household rate variable is analogous to a zonal rate variable (see section 2.3.2). An example of a household rate variable is the number of cars per person per household.

13. Fleet, C.R. and Robertson, S.R., "Trip Generation in the Transportation Planning Process", Highway Research Record, Number 240, 1968.

14. Reported by R. Lane, T.J. Powell and P. Prestwood Smith, Analytical Transport Planning, Duckworth, 1971.

15. Douglas, A.A. and Lewis R.J., "Trip Generation Techniques -3. Household Least-Squares Regression Analysis", Traffic Engineering and Control, January 1971.

16. Oi, W. and Shuldiner, P.W., (An Analysis of Urban Travel Demands, Northwestern University Press, 1962) found car ownership and household size to be the dominant variables in household trip generation.

17. Douglas, A.A. and Lewis, R.J., "Trip Generation Techniques -1. Introduction", Traffic Engineering and Control, November 1970.

18. London Traffic Survey, Freeman, Fox and Partners, 1964.

CHAPTER V

CONCLUSIONS AND RECOMMENDATIONS

5.1 Conclusions

A summary of the main conclusions which can be drawn from the research reported in this thesis is presented below:-

(i) Trends in Cape Town reflect the growth of a 'transportation problem' in the metropolitan area.

(ii) Urban transportation planning and research in Cape Town have not really evolved beyond the early type of origin-destination study. A marked exception is the recent work carried out by Aplin[1].

(iii) Trip generation is an extremely important phase of the traffic forecasting process, but has had comparatively little research effort devoted to it.

The following conclusions are based specifically on the Cape Town home questionnaire survey and the results of that study.

(iv) Fifty interviews (per interviewer) per week is a reasonable task when the home questionnaire survey technique is used, particularly if a five day working week is utilized.

(v) The citizens of Cape Town are amenable to a home questionnaire-type survey[2].

(vi) The response rate in flats was found to be significantly lower than in houses.

(vii) Car ownership amongst the survey population has already reached a very high level of approximately 413 cars per 1 000 people.

(viii) A significantly larger proportion of eligible males (than females) possess a driver's licence.

(ix) The purpose distribution of home-based trips in Cape Town was found to be similar to that reported by other studies.

(x) As a result of the high car ownership level there is a very low degree of patronage of public transport facilities, and 80 per cent of home-based trips are made by car.

(xi) A very large proportion of home-to-work trips are made by car, either as a car driver (60 per cent) or car passenger (15 per cent).

(xii) Trips from home to school, colleges and universities, largely overlap (timewise) with the morning journey-to-work. This enables many scholars to obtain a lift to school, but it also increases the peak period load on public transport facilities.

(xiii) The peak period problem is clearly illustrated by the fact that 50 per cent of all home-based trips are made during four hours of the twenty four hour day.

(xiv) The peak period problem is far more pronounced in the case of public transport than the private motor car.

(xv) Superficially, one would conclude that the zonal level of analysis yields least-squares models better able to explain observed variations in trip-making.

(xvi) Both households and zones exhibit a significant degree of heterogeneity, and much information is lost when personal trip-making data is aggregated to the zonal level.

(xvii) Thorough investigation revealed that household and personal trip generation models are better able to explain observed variations in trip-making, when the effects of data aggregation are accounted for.

(xviii) The personal model explains the largest proportion of the between-person variation in trip-making. Therefore, this model is most likely to remain stable with time. However, even the personal model only explains 15 per cent of the between-person variation in trip-making.

(xix) Both personal and household trip generation models are able to estimate zonal trip generation just as successfully as a zonal model. The personal model was found to be marginally superior in this respect.

(xx) The least-squares assumption of constant error variance is better satisfied by a personal model of trip-making, than by household and zonal models.

(xxi) Household size and car ownership appear to be non-linearly related to home-based trips per household per day, but the assumption of a linear regression surface is still reasonably satisfactory.

(xxii) The personal trip generation model was found to be more successful than the household model when used to estimate the average household trip generation in four foreign study areas.

(xxiii) The trip generation models reported here do not take account of all the factors which could possibly influence future travel volumes[3]. However, the equations can be expected to remain reasonably stable with time, unless large changes in transport policy or costs take place in the future.

5.2 Recommendations

Some of the recommendations made below are not based on the survey and subsequent analyses reported here, but result from the author's readings during the course of preparing this thesis. However, all the recommendations are associated with issues raised at various stages of the report.

(i) A comprehensive land-use/transportation study of the Cape Metropolitan Area is needed, and future planning must be done at this level. To this end, the establishment of a Metropolitan Planning body[4] is seen as a constructive step.

(ii) The Group Areas Act, as it affects the supply of and demand for urban transportation, must be investigated. In this connection, particular attention must be given to the transportation needs of the black people living on the Cape Flats, where a latent demand[5] for travel probably exists. In addition, the effect on travel patterns of increasing standards of living and car ownership must be carefully evaluated.

(iii) The introduction in Cape Town of bus priority and bus lane schemes[6] should be investigated.

(iv) The provision of parking space at the suburban railway stations in Cape Town must be examined, so that the 'park-and-ride' and 'kiss-and-ride'[7] ideas can be encouraged.

(v) The use of the home questionnaire survey technique, as distinct from a home interview survey, should be considered in future studies of the kind reported here.

(vi) In future research into residential trip generation, the analysts should consider the development of models at both the personal and household levels. This will mean collecting, coding and storing the data in such a way that personal trip-making information is available for analysis, but can subsequently be aggregated to the household level. Thus the personal level of analysis will be further investigated.

(vii) The possibility of using smaller sample sizes, when the analysis is carried out at the personal level, should be examined.

(viii) As a model-building technique, the least-squares method is far superior to the category analysis procedure. However, the use of a cross-classification table for reporting trip generation rates is recommended. The author is of the opinion that a cross-classification table will mean far more to a town or regional planner than a least-squares regression equation.

(ix) The category analysis method should be developed until it can be used as a model-building technique. Possibly the excessive data requirements of this method can be reduced in the future.

(x) The category analysis and least-squares techniques should not be seen as mutually exclusive, since both have a role to play in our attempts to simulate and understand the phenomenon of trip generation.

(xi) The effect on residential trip generation (and travel patterns in general), of such factors as accessibility, cost, road pricing[8], and future communications technology[9], must be examined. In this connection, the question of the difference between trip generation and travel demand[10] should be further clarified.

The author hopes that the research reported here represents a small, but worthwhile, contribution to the state-of-the-art of urban transportation planning, particularly in Cape Town. The large number of recommendations above indicates the magnitude of the scope for future research in this field.

NOTES - CHAPTER V

1. Aplin, W.N., Modal Split Analysis for the Journey to Work, Thesis submitted in partial fulfilment of the requirements for the M.Sc.(Eng.) Degree, Department of Civil Engineering, University of Cape Town, 1974.

2. This is in contrast to the results of the mailed questionnaire survey carried out by the Cape Town City Engineer's Department in October 1973. The response rate in this survey varied very much from area to area, particularly since all population groups were represented in the sample, thereby covering a very wide range of socio-economic characteristics. The response rate in the mailed questionnaire survey, varied (on a zonal basis) between 0 and 30 per cent, resulting in an average response rate of approximately 20 per cent.

3. The effect of such factors as accessibility, cost, and future communications technology, has not been accounted for (see recommendation number (xi)). In particular, the 'oil crisis' (which was touched off by last year's war in the Middle East), reminded one very clearly that petrol shortages (whether real, or due to a cut-back in supply), are a factor to be considered. It is likely that the price of petrol will rise fairly substantially, and continually, in the future.

4. The formation of a Cape Metropolitan Planning Body was announced in The Argus, March 9th, 1974.

5. Hoel, L.A., Perle, E.D., Kansky, K.J., Kuehn, A.A., Roszner, E.S. and Nesbitt, H.P., "Latent Demand for Urban Transportation", Transportation Research Institute, Carnegie-Mellon University, Pittsburgh, 1969.

6. See (for example), Louis de Waal, Bus Lanes and Bus Priorities (Paper presented at the Fifth Quinquennial Convention of the South African Institution of Civil Engineers, Johannesburg. August, 1973).

7. See (for example), Joseph Barnett, "Express Bus Mass Transit", Transportation Engineering Journal, Proceedings of the A.S.C.E., May 1970.

8. Roth, G., Paying for Roads - The Economics of Traffic Congestion, Penguin Special, 1967.

9. Harkness, Richard C., "Communications Innovations, Urban Form and Travel Demand: Some Hypotheses and a Bibliography", Transportation, July 1973.

10. In discussing a paper by Fleet and Robertson (Highway Research Record, Number 240, 1968), R.D. Worral points out that one must not confuse trip generation and travel demand. Although both deal with trip frequency, Worral feels that they are different concepts having different objectives, output and principles. Worral suggests that we think of trip generation as an inelastic subset of travel demand. Therefore, travel demand depends on cost, accessibility, etc., but trip generation is independent of these factors.

The difference between these two concepts is therefore important, and the possibility of replacing the trip generation model (as the first phase of the traffic forecasting process), by a model of travel demand, must be considered. This would make it possible to account for changes in travel costs when forecasting future travel volumes. In addition, we would be able to have a meaningful feedback loop from the trip assignment model, thus taking into account the effect of congestion on traffic volumes.

APPENDIX 1

LEAST-SQUARES REGRESSION ANALYSIS

The least-squares regression concept can best be understood by considering the two-variable situation shown in Figure A1.1.

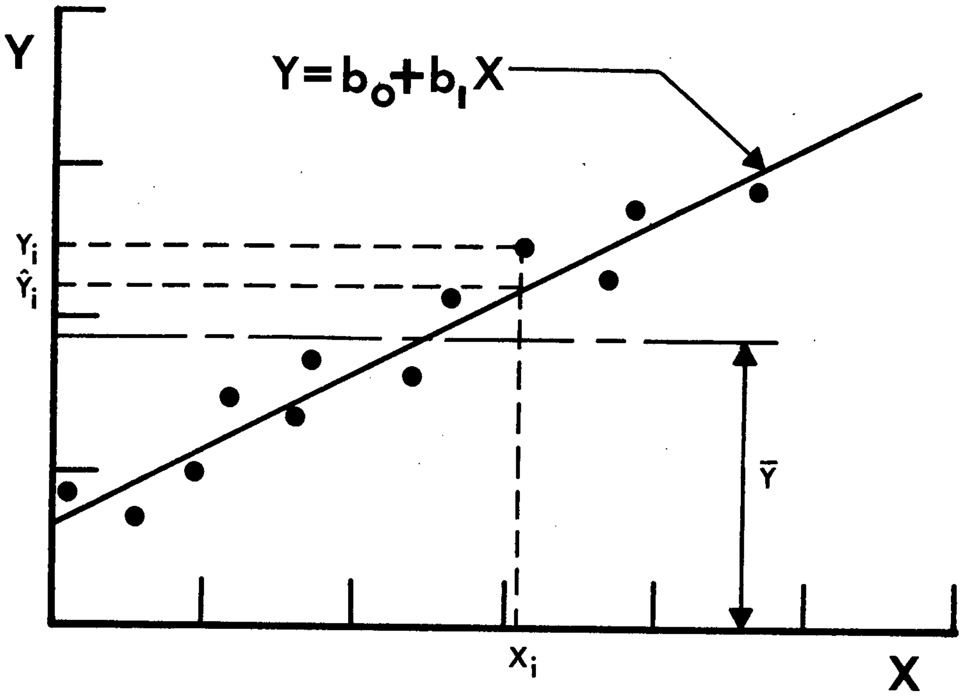


FIG. A1.1 The Least-Squares Concept

where Y_i is the observed value of Y for a particular value of X,

$\hat{Y}_i$ is the value of Y predicted by the regression line for the same value of X,

$\bar{Y}$ is the mean value of Y for all the observations,

and $e = Y_i - \hat{Y}_i$ is called the residual (or deviation).

The least-squares technique enables one to find the values of b_0 and b_1 such that

$$\sum_{i=1}^{n} e_i^2 \text{ is a minimum}$$

where n is the number of observations.

Thus we are trying to minimize

$$\sum_{i=1}^{n} (Y_i - b_o - b_1 X_i)^2$$

Differentiating this expression with respect to b_o and b_1, and equating the results to zero, we obtain

$$-2\,\Sigma(Y_i - b_o - b_1 X_i) = 0 \qquad (A1.1)$$

and

$$-2\,\Sigma\, X_i\,(Y_i - b_o - b_1 X_i) = 0 \qquad (A1.2)$$

Re-arranging the terms in equations (A1.1) and (A1.2) we obtain equations (A1.3) and (A1.4), which are known as the normal equations.

$$\Sigma\, Y_i = nb_o + b_1\, \Sigma\, X_i \qquad (A1.3)$$

$$\Sigma\, X_i\, Y_i = b_o\, \Sigma\, X_i + b_1\, \Sigma\, X_i^2 \qquad (A1.4)$$

Solving equations (A1.3) and (A1.4) yields the value of b_o and b_1.

The above concept may be extended to the case where we have more than one independent variable. There will be k + 1 normal equations for k independent variables, since k + 1 coefficients must be estimated.

Referring again to Figure A1.1, we see that the observed values of Y are dispersed about their mean $(\bar{Y})$. The regression line attempts to explain this dispersion in terms of the observed variations in the value of X. A measure of the success of the regression line in explaining the variations in Y, in terms of the variations in X, is given by

$$r^2 = \frac{\sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^{n} (Y_i - \bar{Y})^2} \qquad (A1.5)$$

where r^2 is the coefficient of determination, while the square root of r^2 is the coefficient of correlation (r).

The numerator of the right-hand-side of equation (A1.5) is known as the 'explained sum of squares' and the denominator is called the 'total sum of squares'. The 'unexplained sum of squares' is given by

$$\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$$

We see that the unexplained sum of squares is identically equal to the sum of squares of the residual, and for this reason it is often referred to as the 'residual sum of squares'. The explained sum of squares is often called the 'regression sum of squares'.

The aim of the least-squares technique is to minimize the unexplained (or residual) sum of squares. Minimizing the residual sum of squares is equivalent to maximizing the explained (or regression) sum of squares since

Total Sum of Squares = Explained Sum of Squares + Unexplained Sum of Squares,

$$\text{or} \quad \sum_{i=1}^{n}(Y_i - \bar{Y})^2 = \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2 + \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 \qquad \text{(A1.6)}$$

If the regression line passed through all the observations then the unexplained sum of squares would be zero, the explained sum of squares would equal the total sum of squares, and the coefficient of determination would be unity. However, if all the observations lay on the line $Y = \bar{Y}$, then the value of r^2 would be zero.

The magnitude of r^2 is often expressed as a percentage, and in this form it expresses the percentage of the total variation of the dependent variable which is explained by variations in the independent variable.

The above ideas may be extended to the case where we consider more than one independent variable. When dealing with two or more independent variables, we speak of the coefficients of multiple determination (R^2) and multiple correlation (R). The coefficients $b_1, \ldots\ldots, b_k$ in a multiple least-squares regression equation are called partial regression coefficients, since they indicate the effect that a change in a particular independent variable has on the dependent variable, when the other independent variables are held constant.

In addition to the coefficient of multiple determination, there are a number of other statistics which are available to the analyst when examining a multiple least-squares regression equation. The standard error of estimate is computed as

$$S_E' = \sqrt{\frac{\sum_{i=1}^{n} (Y_i - \bar{Y}_i)^2}{n - k - 1}} \qquad \text{(A1.7)}$$

where k is the number of independent variables in the regression equation.

The numerator under the square root in equation (A1.7) is the unexplained sum of squares; hence the larger the unexplained sum of squares, the larger the standard error of estimate. In other words, the standard error of estimate is a measure of the deviation of the observed values about the estimated regression surface, and it is therefore a measure of the accuracy of predictions made using the regression equation.

The denominator under the square root in equation (A1.7) is the residual degrees of freedom. If there are k independent variables in the multiple least-squares regression equation, then we have to estimate $k + 1$ parameters (b_o, b_1, ..., b_k), and the residual degrees of freedom will be $n - k - 1$.

The standard error of estimate is often expressed as a percentage of the mean value of the dependent variable. This is a more meaningful form since it is dimensionless, and having been standardized, it can be used for making comparisons.

The overall significance of the regression equation can be evaluated by means of the F-ratio. This ratio is usually contained in the analysis of variance table, which is standard output from a least-squares regression computer program. An example of an analysis of variance table is shown below[1].

Source	D.F.(1)	S.S.(2)	M.S.(3)	F-Ratio(4)
Regression	1(5)	16495,31	16495,31	132,0
Residual	9	1124,33	124,93	
Total	10	17619,64		

(1) Degrees of Freedom

(2) Sum of Squares

(3) Mean Square = Sum of Squares/Degrees of Freedom

(4) The F-ratio is calculated as M.S.(Regression)/M.S.(Residual)

(5) Only one explanatory variable.

TABLE A1.1

The overall significance of the regression may be evaluated by comparing the F-ratio to the F-statistic having (1,9) degrees of freedom. The null and alternate hypotheses we would construct are as follows:-

H_0: The regression equation explains a significant proportion of the variation in the dependent variable.

H_1: The regression equation does not explain a significant proportion of the variation in the dependent variable.

Now $F_{0,01;\ 1;\ 9} = 10,56$ is far less than the computed F-ratio; hence we would reject the null hypothesis and accept the alternate hypothesis, at a one per cent level of significance.

The above test is exactly the same as testing the significance of the sample coefficient of correlation.

It can be shown[2] that the partial regression coefficients divided by their standard deviation have a t-distribution with $n - k - 1$ degrees of freedom. Thus we are able to test the significance of each explanatory variable, by finding out whether its partial regression coefficient is significantly different from zero or not. This test is used in a stepwise multiple least-squares regression analysis (see Appendix 5).

NOTES - APPENDIX 1

1. The table was extracted from, Johnston, J, Econometric Methods, McGraw-Hill, 1963.

2. Yamane, T., Statistics: An Introductory Analysis, Harper and Row, 1969.

APPENDIX 2

EFFECT OF PERFECT LINEAR CORRELATION BETWEEN TWO EXPLANATORY VARIABLES

If two or more of the explanatory variables are linearly related, then the least-squares technique breaks down completely. The dummy variable classes of a particular set are linearly related, and therefore if all are introduced into a multiple least-squares regression analysis, the algorithm will break down.

The author decided to study the effect of including both dummy variables of a two-variable set, and the investigation is presented below for the interested reader.

We will consider the following model:-

$$Y = a_0 + a_1 X_1 + a_2 X_2 + a_3 X_3 + E \qquad \text{(A2.1)}$$

where X_1 and X_2 are dummy variables belonging to the same set

i.e. $$X_1 + X_2 = 1 \text{ (for all observations)} \qquad \text{(A2.2)}$$

Since $X_1 = 1$ when $X_2 = 0$, and vice versa, we see that for all observations

$$X_1 \cdot X_2 = 0 \qquad \text{(A2.3)}$$

The normal equations which will be used to find the least-squares estimators (b_0, b_1, b_2, b_3) for the variables in the above model are as follows:-

$$b_0 n + b_1 \Sigma X_1 + b_2 \Sigma X_2 + b_3 \Sigma X_3 = \Sigma Y \qquad \text{(A2.4)}$$

$$b_0 \Sigma X_1 + b_1 \Sigma X_1^2 + b_2 \Sigma X_1 X_2 + b_3 \Sigma X_1 X_3 = \Sigma X_1 Y \qquad \text{(A2.5)}$$

$$b_0 \Sigma X_2 + b_1 \Sigma X_2 X_1 + b_2 \Sigma X_2^2 + b_3 \Sigma X_2 X_3 = \Sigma X_2 Y \qquad \text{(A2.6)}$$

$$b_0 \Sigma X_3 + b_1 \Sigma X_3 X_1 + b_2 \Sigma X_3 X_2 + b_3 \Sigma X_3^2 = \Sigma X_3 Y \qquad \text{(A2.7)}$$

Using the relationship given in equation (A2.3), the normal equations become

$$b_0 n + b_1 \Sigma X_1 + b_2 \Sigma X_2 + b_3 \Sigma X_3 = \Sigma Y \quad (A2.8)$$

$$b_0 \Sigma X_1 + b_1 \Sigma X_1^2 + 0 + b_3 \Sigma X_1 X_3 = \Sigma X_1 Y \quad (A2.9)$$

$$b_0 \Sigma X_2 + 0 + b_2 \Sigma X_2^2 + b_3 \Sigma X_2 X_3 = \Sigma X_2 Y \quad (A2.10)$$

$$b_0 \Sigma X_3 + b_1 \Sigma X_3 X_1 + b_2 \Sigma X_3 X_2 + b_3 \Sigma X_3^2 = \Sigma X_3 Y \quad (A2.11)$$

Adding (A2.9) and (A2.10) we get

$$b_0(\Sigma X_1 + \Sigma X_2) + b_1 \Sigma X_1^2 + b_2 \Sigma X_2^2 + b_3(\Sigma X_1 X_3 + \Sigma X_2 X_3) = \Sigma X_1 Y + \Sigma X_2 Y \quad (A2.12)$$

Now $\Sigma X_1 + \Sigma X_2 = n$, since $X_1 + X_2 = 1$ for all observations, and also $\Sigma X_1^2 = \Sigma X_1$ and $\Sigma X_2^2 = \Sigma X_2$ (since $X_1 = 0$ or 1, and $X_2 = 0$ or 1)

Hence (A2.12) becomes

$$b_0 n + b_1 \Sigma X_1 + b_2 \Sigma X_2 + b_3(\Sigma X_1 X_3 + \Sigma X_2 X_3) = \Sigma X_1 Y + \Sigma X_2 Y \quad (A2.13)$$

Because of the relationship between X_1 and X_2 we have that

$$\begin{array}{ll} & X_1 X_3 = 0 \ \underline{\text{and}} \ X_2 X_3 = X_3 \\ \text{or} & \\ & X_1 X_3 = X_3 \ \underline{\text{and}} \ X_2 X_3 = 0 \end{array} \quad (A2.14)$$

Let the number of observations for which $X_1 = 1$ be p, and let the number of observations for which $X_2 = 1$ be q (i.e. $p + q = n$); hence we have that

$$\begin{aligned} \Sigma X_1 X_3 &= p/n \ \Sigma X_3 \\ \Sigma X_2 X_3 &= q/n \ \Sigma X_3 \end{aligned} \quad (A2.15)$$

Similarly it can be shown that

$$\begin{aligned} \Sigma X_1 Y &= p/n \ \Sigma Y \\ \Sigma X_2 Y &= q/n \ \Sigma Y \end{aligned} \quad (A2.16)$$

Substituting (A2.15) and (A2.16) in (A3.13) we obtain

$$b_0 n + b_1 \Sigma X_1 + b_2 \Sigma X_2 + b_3 \Sigma X_3 = \Sigma Y \qquad (A2.17)$$

However, equation (A2.17) is identically equal to equation (A2.4), and since we obtained (A2.17) by adding (A2.5) and (A2.6), we see that the set of normal equations only contains three linearly independent equations, although there are four coefficients to be estimated.

<u>Note</u>: The above analysis was not extended to the general case, but the example studied above clearly shows that the least-squares procedure breaks down if two of the explanatory variables are linearly related.

Appendix 3

Survey Documents

UNIVERSITY OF CAPE TOWN

DEPARTMENT OF CIVIL ENGINEERING

URBAN TRANSPORTATION RESEARCH PROJECT

Dear Sir/Madam,

Survey of Trips Made by Residents of the Cape Town Area

The purpose of this survey is to provide detailed and accurate information about the trips which people living in the Cape Town area make by car, bus, motor cycle, train and taxi. Information such as this is not at present available.

In order to ensure that the information collected is as accurate as possible, more than 1500 households are being interviewed over a period of seven weeks. The families to be interviewed have been selected on a random basis and your family is one of those chosen. The co-operation of the families chosen is most important to ensure the success of the survey and your help will be highly valued.

Early next week an interviewer will be calling at your home. He will deliver a number of survey forms to you and your family. These forms will request information about your family and the trips they make on the THURSDAY immediately following the visit of the interviewer. The forms will take only five minutes to complete but the information obtained from this survey will be invaluable to those responsible for planning urban transportation systems both locally and elsewhere.

Your name, address and other indentifying particulars are not included on the questionnaire. We guarantee that all information will be treated in the strictest confidence. We thank you for your help in this survey.

Yours faithfully,

E.I. PAS, B.Sc.(Eng.) (Cape Town)

VIR AFRIKAANS SIEN
AANGEHEGDE BLADSY

UNIVERSITEIT VAN KAAPSTAD

DEPARTEMENT SIVIELE INGENIEURSWESE

STEDELIKE VERVOERWESE NAVORSINGSPROJEK

Geagte Meneer/Mevrou,

Opname van ritte onderneem deur bewoners van die Gebied Kaapstad

Die doel van hierdie opname is om uitvoerige en akkurate inligting in te win omtrent ritte wat bewoners van die gebied Kaapstad per motor, bus, motorfiets, trein en huurmotor onderneem. Inligting van hierdie aard is op die oomblik nie beskikbaar nie.

Om te verseker dat die bepaalde inligting so akkuraat as moontlik is, sal daar oor 'n tydperk van sewe weke met ongeveer 1500 huisgesinne onderhoude gevoer word. Die betrokke huisgesinne, waaronder ook u gesin, is na willekeur gekies. Die samewerking van die gekose gesinne is baie belangrik om 'n suksesvolle opname te voltooi en u hulp sal hoog op prys gestel word.

'n Ondervraer sal u vroeg aanstaande week tuis besoek en 'n aantal vraelyste aan u en u gesin uitdeel. Hierdie vraelyste verlang inligting i.v.m. die ritte wat lede van u gesin sal onderneem gedurende die eerste DONDERDAG wat volg op die besoek van die ondervraer. Die vraelyste sal net vyf minute neem om to voltooi, maar die inligting sal van onskatbare waarde wees vir die toekomstige beplanning van stedelike vervoermetodes.

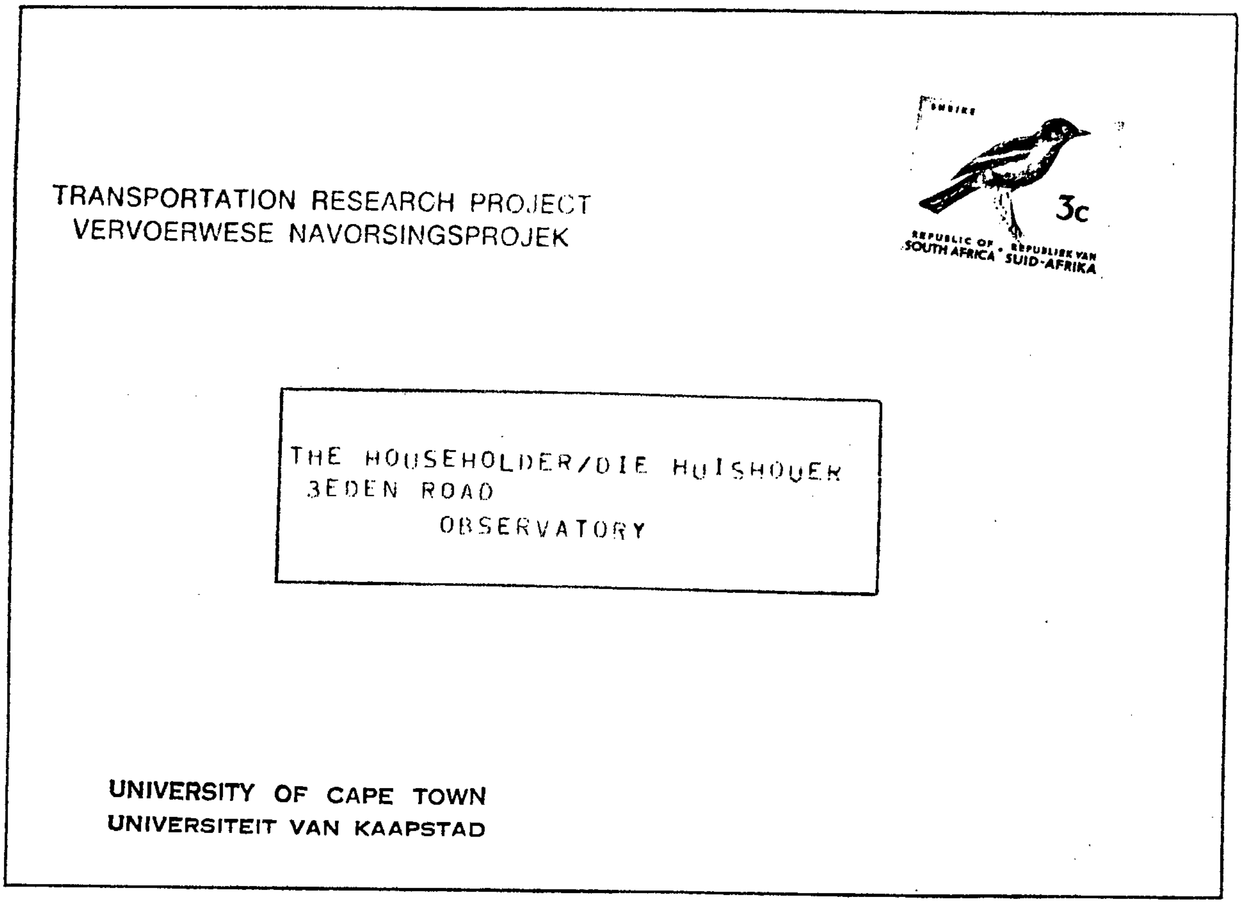
Ons vestig u aandag daarop dat u naam, adres en enige ander persoonlike besonderhede nie op die vraelys verskyn nie. Ons verseker u dat alle inligting as streng vertroulik beskou sal word.

By voorbaat dank vir u hulp met hierdie projek.

Die uwe,

E.I. Pas, B.Sc.(Ing) (Kaapstad)

SEE ATTACHED SHEET
FOR ENGLISH



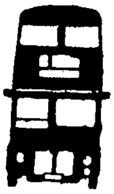
TRANSPORTATION RESEARCH PROJECT
VERVOERWESE NAVORSINGSPROJEK

THE HOUSEHOLDER/DIE HUISHOUER
3EDEN ROAD
OBSERVATORY

UNIVERSITY OF CAPE TOWN
UNIVERSITEIT VAN KAAPSTAD

UNIVERSITY OF CAPE TOWN

DEPARTMENT OF CIVIL ENGINEERING



URBAN TRANSPORTATION RESEARCH PROJECT

HOUSEHOLD INFORMATION - FORM 'A'

Date: Interviewer: Sample No:

Note: Please exclude domestic servants when answering these questions. All information will be treated as Strictly Confidential.

1. Is this dwelling a house ☐ or a flat ☐ ?
2. How many people reside at this address? ☐
3. Please complete the tables below:

How Many People?		
Under 5 years	5 - 17 years	18 and over

How Many Vehicles?	
Cars and Light Vans	Motor Cycles

4. How many school children reside at this address? ☐
5. How many college/university students reside at this address? ☐
6. How many licenced drivers reside at this address? ☐
7. What is the occupation of the head of the household? ....

....

8. How many people at this address are employed (excluding housewives)? ☐
9. Please state the occupations of all the people employed (other than the household head).

....

....

....

....

....

....

UNIVERSITY OF CAPE TOWN

DEPARTMENT OF CIVIL ENGINEERING

URBAN TRANSPORTATION RESEARCH PROJECT

TRIP RECORD - FORM 'B'

Date: Interviewer: Sample No:

1. Please complete one copy of this form for each member of the household 5 years of age or older.
2. Please complete the General Information section below and the attached Trip Record.
3. Remember that the interviewer will be calling to collect the completed forms. If you require any assistance the interviewer can provide it when he calls.

GENERAL INFORMATION

Please fill in the following. This information is about the PERSON whose trips are recorded overleaf. (Please tick the appropriate square.)

1. AGE GROUP: 5 - 17 years ☐ 18 years and over ☐
2. MALE ☐ FEMALE ☐
3. If employed, please state (a) Occupation ....
   (b) Industry or Business ....
4. Are you self-employed? Yes ☐ No ☐
5. Are you a student? At School ☐ University ☐ or College ☐
6. Do you hold a driving licence? Yes ☐ No ☐

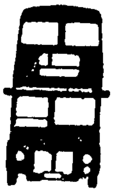
PLEASE RECORD ALL THE TRIPS YOU MADE ON THURSDAY [illegible]

TRIP NUMBER ↓	WHERE DID YOU GO ?		WHY DID YOU GO ?						HOW DID YOU GO ?						AT WHAT TIME ?		
	IN THE 'FROM' COLUMN PLEASE RECORD THE AREA FROM WHICH THE TRIP STARTED (e.g. 'MOWBRAY' or 'CAPE TOWN'). IF THE TRIP BEGAN AT YOUR HOME, PLEASE RECORD 'HOME'.	IN THE 'TO' COLUMN PLEASE RECORD THE AREA IN WHICH THE TRIP ENDED (e.g. 'OTTERY' or 'NEWLANDS'). IF THE TRIP ENDED AT YOUR HOME, PLEASE RECORD 'HOME'.	TO GET TO WORK	TO DO SHOPPING	TO RETURN HOME	TO GET TO SCHOOL UNIVERSITY COLLEGE	FOR PLEASURE (e.g. SPORT, CINEMA VISIT FRIENDS. EAT OUT, etc.)	FOR ANY OTHER REASON	DROVE A CAR/LIGHT VAN	PASSENGER IN CAR/LIGHT VAN	BY BUS	BY TRAIN	BY TAXI	BY MOTOR CYCLE	BETWEEN 7 a.m. AND 9 a.m.	BETWEEN 4.30 p.m. AND 6.30 p.m.	REST OF THE DAY
	FROM	TO															
1																	
2																	
3																	
4																	
5																	
6																	
7																	
8																	
9																	
10																	
11																	
12																	
13																	
14																	
15																	
16																	
17																	
18																	
19																	
20																	

A/14.

UNIVERSITEIT VAN KAAPSTAD

DEPARTEMENT SIVIELE INGENIEURSWESE



STEDELIKE VERVOERWESE NAVORSINGSPROJEK

HUISHOUDELIKE INLIGTING - VORM 'A'

Datum:

Ondervraer:

Monster Nr:

Let Wel: Moet asseblief nie bediendes insluit wanneer u hierdier vrae beantwoord nie.

Alle inligting sal as streng konfidensieel beskou word.

1. Is hierdie woonplek 'n huis ☐ of 'n woonstel ☐ ?
2. Hoeveel mense is in hierdie woonplek woonagtig ? ☐
3. Voltooi asseblief die volgende tabelle:

Hoeveel van die bewoners is:		
Minder as 5 jaar oud	5 - 17 jaar oud	Meer as 18 jaar oud

Hoeveel Voertuie is daar?	
Motors en Ligte Vervoerwaens	Motorfietse

4. Hoeveel skoolkinders woon hier? ☐
5. Hoeveel Kollege/Universiteitstudente woon hier? ☐
6. Hoeveel van die mense hier woonagtig besit rybewyse? ☐
7. Gee asseblief die beroep van die gesinshoof ....

....

8. Hoeveel van die bewoners van hierdie woonplek is werknemers (moet nie huisvrouens insluit nie)? ☐
9. Gee asseblief die beroepe van al die werknemers hier woonagtig (die gesinshoof uitgeslote).

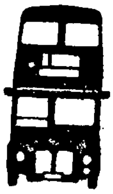
....

....

....

UNIVERSITEIT VAN KAAPSTAD

DEPARTEMENT SIVIELE INGENIEURSWESE



STEDELIKE VERVOERWESE NAVORSINGSPROJEK

REISOPNAME - VORM 'B'

Datum: Ondervraer: Monster Nr:

Let Wel: ALLE INLIGTING SAL AS STRENG KONFIDENSIEEL BESKOU WORD

1. Voltooi asseblief een afskrif van hierdie vorm vir elke lid van die huishouding wat meer as 5 jaar oud is.

2. Voltooi ook die Algemene Inligting-afdeling hieronder en die Reisopname wat aangeheg is.

3. Onthou dat die ondervraer u weer sal besoek om die voltooide vorms te kom haal. As u hulp nodig het, kan die ondervraer dit verskaf wanneer hy u weer besoek.

Algemene Inligting

Voltooi asseblief die volgende. Die inligting is in verband met die PERSOON wie se ritte op die aangehegde vorm beskryf word. (Merk asseblief net die betrokke blokkie.)

1. Ouderdoms Groep: 5 - 17 jaar ☐ Meer as 18 jaar ☐

2. Manlik ☐ Vroulik ☐

3. As u werk, vermeld asseblief (a) Beroep ....

(b) Nywerheid ....

4. Is u u eie werkgewer? Ja ☐ Nee ☐

5. Is u 'n student? Op skool ☐ Universiteit ☐ of Kollege ☐

6. Besit u 'n rybewys? Ja ☐ Nee ☐

VERSKAF ASSEBLIEF ALLE REISE WAT OP DONDERDAG 16 OKTOBER 1975 DEUR U ONDERNEEM IS.

REIS NOMMER ⟵	WAARHEEN HET U GEGAAN ?		HOEKOM HET U GEGAAN ?						HOE HET U GEGAAN ?						TUSSEN WATTER TYE ?		
	GEE IN DIE 'VAN' KOLOM DIE OMGEWING WAARIN DIE REIS BEGIN HET (bv. 'NUWELAND' of 'KAAPSTAD'). AS DIE REIS BY U TUISTE BEGIN HET, SKRYF 'TUISTE'.	GEE IN DIE 'NA' KOLOM DIE OMGEWING WAARIN U DIE REIS BEËINDIG HET (bv. 'MOWBRAY' of 'BELLVILLE') AS U DIE REIS BY U TUISTE BEËINDIG HET, SKRYF 'TUISTE'.	OM BY DIE WERK TE KOM	OM INKOPIES TE DOEN	HUIS TOE	OM SKOOL UNIVERSITEIT KOLLEGE TOE TE GAAN	VIR PLESIER (bv. SPORT, BIOSKOOP, VRIENDE BESOEK, UIT EET, ens.)	VIR ENIGE ANDER REDE	'n MOTOR of LIGTE VERVOERWA BESTUUR	'n PASSASIER IN MOTOR of LIGTE VERVOERWA	PER BUS	PER TREIN	PER HUURMOTOR	PER MOTORFIETS	TUSSEN 7 vm. EN 9 vm.	TUSSEN 4.30 nm. EN 6.30 nm.	RES VAN DIE DAG
	VAN	NA															
1																	
2																	
3																	
4																	
5																	
6																	
7																	
8																	
9																	
10																	
11																	
12																	
13																	
14																	
15																	
16																	
17																	
18																	
19																	
20																	

A 17.

UCT to study City transport

Cape Times Municipal Reporter

THE Department of Civil Engineering at the University of Cape Town has launched a computer-aided research project into transport demands generated in Cape Town's suburbs.

The survey is part of a comprehensive project directed by a lecturer in the department, Mr E. I. Pas, into transport demands in Cape Town. The present survey, in which civil engineering students will participate as part of their final-year curriculum, starts on Monday and will cover the transport demands of between 1 500 and 2 000 households in Cape Town and Pinelands municipalities and in the Cape Divisional Council.

Householders will be interviewed for information relating to the number of trips they make, the modes of transport they use, the purposes for which trips are made and the periods when they are made.

Field work for the project will take seven weeks and the results are expected to become available in April next year. According to Mr Pas, the project will contribute data necessary for transport planning in the City. Major cities oversea and in South Africa already had such data, but he believed the City Council did not have much data on transport demands beyond a few traffic counts.

He said the project was confined at this stage to White areas in Cape Town, but warned that studies into transport demands in the non-White areas would become necessary as there was a latent demand there which could result in a "phenomenal growth in a confined area".

The Cape Times

8.9.73

City survey of transport demands

From Mr E. I. PAS (Department of Civil Engineering, University of Cape Town):

ON SEPTEMBER 8 your Municipal Reporter reported that the above Department had launched a research project into transport demands generated in Cape Town's suburbs. As reported, this survey is part of a comprehensive research project into transport demands in Cape Town.

Since September 10 a team of seven interviewers (final-year students in the Department of Civil Engineering) have interviewed approximately 2 000 households in the Cape Town and Pinelands municipalities, as well as in the adjoining Divisional Council areas.

Few refusals

The response we obtained from the survey households was most gratifying. Approximately only six per cent of the chosen households refused to assist us in the survey. This we consider to be a very low refusal rate.

We would like to thank those people who gave up their time to assist us in this survey. Preliminary examination of the completed questionnaires indicates that the majority were correctly completed.

We hope to have the results of the survey available by April 1974.

The Cape Times

6.11.73

Instructions to Interviewers:

1. This project is being carried out under the auspices of the University of Cape Town. Please do not do (or say) anything which could be detrimental to the University's name.

2. People will generally be co-operative as long as you are reasonable in your approach.

3. The initial interview must be carried out with a responsible member of the survey household.

4. The best way to introduce yourself is to say, "Good evening (afternoon), I am from the Transportation Research Project". At this stage many respondents will recall the introductory letter, or possibly the newspaper report.

5. If the respondent is aware of the survey and its aims, find out whether he/she minds answering a few questions. Proceed to fill in Form 'A' if the respondent is willing. After you have completed most of Form 'A', explain to the respondent how Form 'B' is to be completed for each member of the household five years of age or older (see 8 below).

6. You will find that even when the person you are interviewing has read the covering letter (or has been informed of its contents by other members of the household), it will be useful to outline briefly some of the salient points from the introductory letter.

7. In a few cases you will find that a more detailed explanation is required by the respondent. If this is so, you must be as helpful as possible.

8. When you are explaining the requirements of Form 'B', the following points should be emphasized:-

   (i) One copy of this form is to be completed for each member of the household five years of age or older.

   (ii) If a particular member of the household does not make any motorized trips on the survey day this should be clearly indicated on the relevant trip record form, but the 'General Information' section should still be completed.

(iii) Only motorized trips made on the survey day should be recorded. The survey day is the coming Thursday.

(iv) A journey from (say) home to Cape Town and back home is to be recorded as two trips.

(v) A movement made by more than one motorized form of transport should not be recorded as one trip. For example, a person might catch a bus to the nearest railway station (or drive there by car), and catch the train to work. Such a movement should be recorded as two separate trips. The first purpose would be "For some other reason", while the second is "To get to work".

(vi) Doctors, salesmen, repair-men, etc., make many non-home-based trips during a typical weekday. In such cases the respondent need record in detail (purpose, mode, time of day) only those trips which began or ended at his home, and those trips not made for business purposes. At the bottom left-hand corner of the trip record form the respondent should estimate the number of trips made which are not recorded in detail.

9. After having emphasized the above points, complete a trip record form as an example[1], again emphasizing the important aspects. It is advisable to do this even if it appears to you that the respondent fully understands the instructions.

10. Remind the respondent that you will be calling to collect the completed forms after the survey day, and that you will be able to assist if any problems are experienced in completing the questionnaire.

11. Arrange a time (if possible) at which you will collect the completed forms. This should be on the Friday or Saturday following the survey day. Remember to observe people's religious practices where necessary.

12. If you establish that a particular house or flat is vacant, please record this on your field schedule.

13. The following codes should be used on the field schedule:-

C - completed interview

V - vacant

U - unable to contact

R - refused

N - non-existent.

This information should be transferred to your list of addresses.

14. The field schedule has space available for four visits to each household. If after four visits you have not been able to contact a responsible member of the household, record a 'U' on your field schedule. Do not substitute households under any circumstances.

15. When collecting the forms check to see whether they have been correctly completed.

Notes:

1. After the first week of the survey the standard trip record example sheet was introduced, and the interviewers did not complete an example sheet in the presence of the respondent.

No.	Address	1st Call	2nd Call	3rd Call	4th Call	Collect	Code	Remark

PLEASE RECORD ALL THE TRIPS YOU MADE ON THURSDAY

TRIP NUMBER	WHERE DID YOU GO ?		WHY DID YOU GO ?						HOW DID YOU GO ?						AT WHAT TIME ?		
	IN THE 'FROM' COLUMN PLEASE RECORD THE AREA FROM WHICH THE TRIP STARTED (e.g. 'MOWBRAY' or 'CAPE TOWN'). IF THE TRIP BEGAN AT YOUR HOME, PLEASE RECORD 'HOME'.	IN THE 'TO' COLUMN PLEASE RECORD THE AREA IN WHICH THE TRIP ENDED (e.g. 'OTTERY' or 'NEWLANDS'). IF THE TRIP ENDED AT YOUR HOME, PLEASE RECORD 'HOME'.	TO GET TO WORK	TO DO SHOPPING	TO RETURN HOME	TO GET TO SCHOOL UNIVERSITY COLLEGE	FOR PLEASURE (e.g. SPORT, CINEMA, VISIT FRIENDS, EAT OUT, etc.)	FOR ANY OTHER REASON	DROVE A CAR/LIGHT VAN	PASSENGER IN CAR/LIGHT VAN	BY BUS	BY TRAIN	BY TAXI	BY MOTOR CYCLE	BETWEEN 7 a.m. AND 9 a.m.	BETWEEN 4.30 p.m. AND 6.30 p.m.	REST OF THE DAY
	FROM	TO															
1	Home	Claremont Station						✓			✓				✓		
2	Claremont Station	Cape Town	✓									✓			✓		
3	Cape Town	Claremont Station						✓				✓				✓	
4	Claremont Station	Home			✓						✓					✓	
5	Home	Rosebank					✓		✓								✓
6	Rosebank	Home			✓				✓								✓
7																	
8																	
9																	
10																	
11																	
12																	
13																	
14																	
15																	
16																	
17																	
18																	
19																	
20																	

A23

PROJECT: Analysis of Residential Trip Generation in Cape Town.
Questionnaire to be completed by interviewers.

1. Do you consider that the average of 50 households to be interviewed per week was a reasonable task?

   Definitely Yes Yes No Definitely Not

2. Do you consider that the newspaper publicity had any effect on the co-operation (or otherwise) received from sample households?

   Definitely Yes Yes No Definitely Not

3. Do you feel that it was necessary to have a pre-interview letter sent to the sample household?

   Definitely Yes Yes No Definitely Not

4. Would you consider the co-operation of the sample households to have been

   Very good Good Bad Very Bad

5. Do you feel that the basic technique used for collecting the data was

   Very Successful Successful Unsuccessful Very Unsuccessful

6. Do you feel that it was necessary to contact people before the survey day?

   Definitely Yes Yes No Definitely Not

7. Do you think that the standard trip record example form was

   Very Successful Successful Unsuccessful Very Unsuccessful

8. Do you consider the design of the survey forms to have been

   Very Good Good Poor Very Poor

9. What percentage of your sample households were Afrikaans speaking?

10. What percentage of your sample households could speak neither English nor Afrikaans?

11. Any comments?

APPENDIX 4

Test for Differences in Proportions - Failure Rates in Houses and Flats

Let P_1 and P_2 be the sample proportions obtained from two large samples of size N_1 and N_2, drawn from populations having proportions p_1 and p_2 respectively. The null hypothesis is that there is no difference between the population parameters, i.e. $p_1 = p_2$

It can be shown that the sampling distribution of the differences in proportions is approximately normal with mean $\mu_{P_1-P_2} = 0$, and standard deviation

$$\sigma_{P_1-P_2} = \sqrt{pq\left(\frac{1}{N_1} + \frac{1}{N_2}\right)}$$

where $$p = \frac{N_1P_1 + N_2P_2}{N_1 + N_2}$$

and $$q = 1 - p$$

Therefore the standardized variable

$$Z = \frac{P_1 - P_2 - 0}{\sigma_{P_1-P_2}} = \frac{P_1 - P_2}{\sigma_{P_1-P_2}}$$

can be used to test the observed differences in proportions at an appropriate level of significance.

In the following analysis '1' refers to houses and '2' refers to flats.

(i) Testing the Overall Failure Rate

$N_1 = 1\ 236$ $\qquad P_1 = 0{,}166$

$N_2 = 928$ $\qquad P_2 = 0{,}250$

$$p = \frac{1\ 236 \times 0{,}166 + 928 \times 0{,}250}{1\ 236 + 928} = 0{,}202$$

$$\therefore \quad q = 1 - 0{,}202 = 0{,}798.$$

Now $\sigma_{P_1-P_2} = \sqrt{(0{,}202)(0{,}798)(\frac{1}{1\ 236} + \frac{1}{928})}$

$= 0{,}0174.$

$\therefore \quad Z = \frac{P_1 - P_2 - 0}{0{,}017} = 4{,}83$

$H_o : p_1 = p_2$

$H_1 : p_1 < p_2$ (i.e. a one-tailed test)

Now $Z_{0{,}01} = 2{,}33$

Therefore reject H_o and conclude that the overall failure rate in flats is higher than in houses, at a 1 per cent level of significance.

(ii) Testing the Interview Failure Rate

$N_1 = 1\ 160 \qquad P_1 = 0{,}111$

$N_2 = 836 \qquad P_2 = 0{,}167$

$\therefore p = 0{,}128,\quad q = 0{,}872$ and $\sigma_{P_1-P_2} = 0{,}0152$

$\therefore Z = 3{,}7$

$H_o : p_1 = p_2$

$H_1 : p_1 < p_2$

Now $Z_{0{,}01} = 2{,}33$

Therefore reject H_o and conclude that, at a 1 per cent level of significance, the interview failure rate for flats is higher than for houses.

(iii) Testing the Refusal Rate

$N_1 = 1\ 103 \qquad P_1 = 0{,}065$

$N_2 = 770 \qquad P_2 = 0{,}096$

$\therefore p = 0{,}078,\quad q = 0{,}922$ and $\sigma_{P_1-P_2} = 0{,}0126$

$\therefore Z = 2{,}45$

$H_o : p_1 = p_2$

$H_1 : p_1 < p_2$

Now $Z_{0,01} = 2,33$

Therefore reject H_o at a 1 per cent level of significance, and conclude that the refusal rate for flats is higher than for houses.

APPENDIX 5

Stepwise Least-Squares Regression Program (BMD02R)

In section 4.3.2 a brief outline of the BMD02R program is presented. The aim here is to provide a more detailed description of this program, and to illustrate the stepwise regression procedure.

The BMD02R program computes a series of multiple linear regression equations in a stepwise procedure. At each stage, a variable is added to (or deleted from) the regression equation. The variable added is the one which makes the greatest reduction in the regression sum of squares, and equivalently it is the variable which, if it were added, would have the highest F-value. Variables can be forced into the regression equation, while non-forced variables are automatically removed when their F-values become too low. The user is able to select a regression equation with a zero intercept; in other words, the regression surface can be forced through the origin. A further feature of the program is that the analyst can perform a number of transformations (e.g., square, square root, logarithm) on the variables before the regression analysis.

The F-value referred to above must not be confused with the F-ratio defined in Appendix 1. The F-ratio refers to the whole regression equation, while the F-value applies to a particular variable. The F-statistic is used for testing the significance of both the F-ratio and F-value.

The F-value can be calculated as follows:-

$$\text{F-value} = \frac{\text{S.S.(Regression)}_2 - \text{S.S.(Regression)}_1}{\text{M.S.(Residuals)}_2} \qquad \text{(A5.1)}$$

where $\text{S.S.(Regression)}_2$ = regression sum of squares <u>after</u> the variable under consideration is included in the equation.

$\text{S.S.(Regression)}_1$ = regression sum of squares <u>before</u> the variable under consideration is included in the equation.

M.S.(Residuals)_2 = mean square of the residuals <u>after</u> the inclusion of the variable being considered.

This F-value can be compared to the F-statistic having (1, n - k - 1) degrees of freedom.

Alternatively, the F-value is computed as

$$F_i = (b_i/S_i)^2 \tag{A5.2}$$

where F_i = F-value of the i^{th} independent variable,

b_i = partial regression coefficient of the i^{th} independent variable,

S_i = standard deviation of the coefficient of the i^{th} independent variable (i.e. the standard deviation of b_i).

From equation (A5.2) it can be seen that the F-value is the square of the t-value defined in Appendix 1. Since $F_{\alpha,1,n-k-1} = (t_{\frac{\alpha}{2},n-k-1})^2$, we see that testing the significance of a partial regression coefficient is equivalent to testing whether the F-value is significant.

However, the interpretation is slightly different in the two cases. A significant F-value indicates that the variable will make a significant reduction in the unexplained sum of squares if included in the equation. On the other hand, a significant t-value shows that the partial regression coefficient is significantly different from zero.

Let us have a look at how these tests are used in the stepwise regression procedure. The stepwise development of the household (total) equation is shown in Table 4.12. At the third step of the process the following equation was obtained:-

$$\underset{}{HBTHD} = -0{,}84 + \underset{(21{,}8)}{1{,}53FAMSIZ} + \underset{(14{,}4)}{1{,}75CARS} + \underset{(3{,}35)}{0{,}43EMPLOY} \tag{4.2}$$

where HBTHD = number of home-based trips per household per day,
FAMSIZ = number of people per household,
CARS = number of cars per household,
EMPLOY = number of employed people per household.

Below each independent variable in the equation above, the t-value is shown. We can set up the following hypotheses:-

H_0 : coefficient of EMPLOY = 0
H_1 : coefficient of EMPLOY $\neq$ 0

Now $t_{0{,}005;1\,686} = 2{,}58$.

Therefore, H_0 is rejected and we conclude that the coefficient of the variable EMPLOY is significantly different from zero at a 1 per cent of significance.

Prior to the inclusion of the variable EMPLOY the regression sum of squares was 20 112, while after inclusion of this variable the regression sum of squares increased to 20 286. After EMPLOY was included in the equation the mean square of the sum of the residuals was found to be 15,46. Therefore the F-value can be computed as

$$\frac{20\ 286 - 20\ 112}{15{,}46} = 11{,}2.$$

The following hypotheses are set up:-

H_0 : EMPLOY does not produce a significant increase in the regression sum of squares.

H_1 : EMPLOY does produce a significant increase in the regression sum of squares.

Now $F_{0{,}01;1;1\ 686} = 6{,}6$

Therefore, reject H_0 and conclude that, at a 1 per cent level of significance, EMPLOY does produce a significant increase in the regression sum of squares when included in the equation after FAMSIZ and CARS have entered the regression model.

As has been mentioned above, the F- and t-tests are equivalent, and in practice only one is used in a stepwise regression program. The BMD02R program uses the F-test for evaluating whether a variable is significant or not.

The BMD user manual[1] mentions that the analyst must specify the 'F-level for inclusion' and the 'F-level for deletion'. The default values for these two parameters are given as 0,01 and 0,005 respectively. One immediately interprets these as being the levels of significance (i.e. 1 % and 0,5 %) for inclusion and deletion of non-forced variables.

However, on first using the program the author found that variables with low F-values (say 0,5) entered the regression equation when the F-levels for inclusion and deletion were not specified (i.e. the default values applied).

At a 1 per cent level of significance and for a large number of residual degrees of freedom (such as we had, because of the large number of observations), the F-statistic has a value of 6,6. Thus variables having an F-value of 0,5 should not have entered the equation.

After further investigation, the author came to the conclusion that one actually had to specify an F-value (not level of significance) as the criterion for inclusion and deletion of variables.

The output from the BMD02R program includes:

(1) At each step:

- (a) Multiple coefficient of correlation (R).
- (b) Standard error of estimate (S_E).
- (c) Analysis-of-variance table.
- (d) For each variable in the equation:
  - Partial regression coefficient
  - Standard deviation
  - F-to-remove (i.e. F-value).
- (e) For each variable not in the equation:
  - Partial correlation coefficient
  - F-to-enter (i.e. F-value).

(2) Optional output prior to performing the regression:

- (a) Mean and standard deviation of each variable.
- (b) Covariance matrix.
- (c) Correlation matrix.

(3) Optional output after performing the regression:

- (a) List of residuals.
- (b) Plots of residuals against specified variables.
- (c) Summary table - summarizes the stepwise development of the regression equation.

The correlation matrix is particularly useful to the analyst for investigating the correlation between pairs of explanatory variables. It also enables the user to study the correlation between the dependent variable and individual independent variables.

The possibility of obtaining plots of the residuals against specified variables appears to be a useful feature, since it would assist the analyst in identifying particular trends in the residual error term.

However, if one has a large number of observations, the resulting plots become very crowded and trends cannot be easily identified. This is particularly the case when one is dealing with non-continuous variables such as the number of people (or cars) per household.

In general, the BMD02R is a very powerful and versatile program, well suited for the development of trip generation models.

Notes:

1. Dixon, W.J., (ed.), Biomedical Computer Programs, University of California Press, 1968.

APPENDIX 6

Test for Differences in Proportions - Males and Females having a Driver's Licence

The test for difference in proportions is described in detail in Appendix 4, and only the relevant calculations are shown here.

Subscript '1' refers to males, and '2' refers to females.

$N_1 = 1\ 433 \qquad P_1 = 0{,}876$

$N_2 = 1\ 017 \qquad P_2 = 0{,}528$

$\therefore \quad p = 0{,}686$

$\quad\;\; q = 0{,}312$

$\therefore \; \sigma_{P_1 - P_2} = 0{,}016$

and $Z = 21{,}75$

$H_0 : p_1 = p_2$

$H_1 : p_1 > p_2$ (i.e. one-tailed test)

Now $Z_{0,01} = 2{,}33$

Therefore, reject H_0 and conclude that, at a 1 per cent level of significance, the proportion of eligible males (i.e. those eighteen and over) having a driver's licence, is higher than the comparable proportion for females.