COMPARATIVE APPROACH TO PLANNING ANALYSIS
- A MODEL FOR CBD FLOOR SPACE PREDICTION -

by

HOW-YIN LEUNG
B. Sc. Arch. Eng.
Taiwan Provincial Cheng-Kung University, 1966

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE

in the School
of
Community and Regional Planning

We accept this thesis as conforming to the
required standard

THE UNIVERSITY OF BRITISH COLUMBIA
May, 1970

School of Community and Regional Planning

The University of British Columbia
Vancouver 8, Canada

Date April 30th, 1970.

ABSTRACT

Due to the introduction of modern computing technology to planning, it is anticipated that the techniques for quantitative analysis in planning will be changed radically; the expanding use of model and the increasing importance of inter-city data collection may have an effect in redefining the process of planning itself.

In this thesis, an effort has been made to investigate the current problems of building and using models in planning so that a methodology, combining several multivariate statistical methods and the modern computation algorithms, is developed for planning analysis and prediction from the comparative point of view. It is hypothesized that simple prediction models can be improved by the use of these comparative techniques. As a testing mechanism of the proposed methodology, a simple regression model for CBD floor space prediction is devised, and the result shows that the comparative analysis is effective for better prediction.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure Page

ACKNOWLEDGEMENT

I wish to express my sincere gratitude to Dr. N. d. Cherukupalle, my thesis advisor, for her constructive criticism and suggestions during the preparation of this thesis. As well, I happily acknowledge my grateful appreciation to Dr. H. C. Davis and Professor P. Roer for their particular interest in my study, and to Miss Sheung-Ling Chan, my colleague, for smoothing my English.

I am also indebted to the planning agencies who promptly answered my questionnaires and provided me invaluable information for this study.

COMPARATIVE APPROACH TO PLANNING ANALYSIS

- A MODEL FOR CBD FLOOR SPACE PREDICTION -

CHAPTER I

INTRODUCTION

1.1 Nature and Purpose of the Study

1.2 Organization of the Thesis

CHAPTER I

INTRODUCTION

1.1 Nature and Purpose of the Study

Since the entry of social scientists into city planning began in the 1930's, the theory and practice of city planning have been transforming gradually from the exclusive master plan approach which tries to portray the future physical condition of the city, to the so-called goal-oriented approach[1] which develops programmes to allocate limited resources in order to achieve the goals of the community, and concerns more about the implications of people living in the environment. In order to make policies achieving the desired goals, or even before formulating these planning goals, intensive understanding of the existing conditions of the area concerned is indispensable. Furthermore, such analytical work for the goal-oriented planners will be more complicated and more important than for the master-plan-oriented planners, because in providing a basis for decision making they have to formulate and evaluate alternative programmes not only according to the tangible or physical effects but also to the intangible or human consequences which require the insight into human behavior upon the change of environment. This necessitates the use of sophisticated tools in planning.

---

[1] Herbert J. Gans, People and Plans, Essays on Urban Problems and Solutions (New York: Basic Books, Inc., 1968), Chapter 6, pp. 78-83.

No matter which way the planners go, planning is, as Britton Harris puts it, essentially oriented to the future so that the planner should "devise policies which can influence the development in desired directions, by means and at costs which are acceptable to the community as a whole".[2] The first and important step to achieving this goal, therefore, is to comprehend the present conditions so as to predict the probable future development and identify the desirable alternative patterns and directions of development.[3]

Nowadays, as urban problems become more and more complicated, the planners and other experts, mainly the social scientists, are increasing their awareness of the implications and consequences of interaction of many human activities. In the process of planning analysis, planners very often come across a situation where the complexity and uncertainty of human interactions are unlikely to be predictable by means of simple techniques employed in the past. Fortunately, by the late 1950's another transformation of planning theory and practice has been taking place, influencing mainly the field of planning techniques. "Planning techniques have been most highly refined in the model building and computer technology applied since 1955 by the massive transportation studies in Chicago, Detroit, Philadelphia, Pittsburgh, and other large cities".[4]

---

[2]Britton Harris, "New Tools for Planning," Journal of the American Institute of Planners, May 1965, pp. 90-95.

[3]Ibid.

[4]I. William Goodman, ed., Principles and Practice of Urban Planning (International City Managers' Association, 1968), p. 27.

The transportation planners did introduce several innovations into city planning. Now, city planners are able to collect large masses of data after bringing the newly available computer into the profession. Under the inspiration and influence of individual transportation studies, they formulate a number of smaller planning schemes rather than a single one.[5] Perhaps the most important influence is that the recently developed simulation models for planning decision making have lightened an effective way towards solving the complex urban problems.

Obviously, the techniques for planning analysis and prediction are undertaking revolutionary advancement. No doubt the computer takes an influential role for the innovations. Harris, recognizing the importance of computer in the society as well as in the planning profession, puts it this way:

> It is no exaggeration to say that, with regard to human affairs in general, the invention of the computer ranks with such major technological innovations as fire and the wheel, or at least the steam engine . . .
>
> Computer is no more and no less than a tool in the planning process. It is, however, a tool of such revolutionary new potential that it may have an effect in redefining the process of planning itself.[6]

The computer makes it possible to obtain, process, store and retrieve masses of information for planning; it analyzes information and

---

[5]Gans, Op. cit., p. 66.

[6]Britton Harris, "Computer and Urban Planning", Socio-Economic Planning Science, vol. 1, 1968, pp. 223-230.

makes prediction effectively; what is more, it serves as a catalyst and as an ingredient in the process of model-building to find solutions to planning problems. By means of high speed electronic computer, many a cumbersome statistical calculation, that could not extensively be manipulated before, can be handled without difficulty and applied to planning easily today. New techniques are devised, and many mathematical models are operated for the purpose of decision-making. Therefore, in the past few years significant new needs and capabilities have quickened that trend to the point of a methodological revolution.[7]

Although the development of planning techniques has attained dramatic achievements since last decade, it is as yet in its infancy; it has high potentiality but it needs further modifications. It is the fact that, except some of the leading agencies who brought the computer and simulation models into practice, the planning agencies in general have not yet completely accepted these innovations in the planning process, because of some inherent difficulties which will be discussed in the next chapter. Nevertheless, opportunities are widely opened in every established or new direction from which new models can be built and systems of techniques can be devised.

Combining the modern computation algorithms and some of the multivariate statistical methods, this thesis attempts to develop, from a comparative point of view, a system of techniques to penetrate into the

---

[7]Goodman, op. cit., p. 277.

urban problems, so that a simple model can be derived for the purpose of planning prediction. Moreover, as a testing mechanism of the proposed system of techniques, a model is devised to predict the floor space of the central business district (CBD) of a city.

1.2 Organization of the Thesis

Since the study in this thesis involves the using of multivariate statistical methods which may be beyond those who have little knowledge in statistics, the thesis is so organized that its main body, consisting of five chapters, largely concerns with the evolution, concepts, procedures, results, and implications of the study as well as the proposed model, with a minimum application of statistics and mathematics. The detailed statistical interpretations of various techniques involved constitute a series of appendices for those who are interested in further exploration of the subject matters.

After this introductory remarks on the nature of the study, the next chapter will bear upon the evolution of the comparative approach and the formulation of the general hypothesis, beginning with a brief review of the existing planning techniques and their inherent problems. The third chapter discusses the concepts and methodology of the proposed comparative approach. After these, an empirical study aiming at developing a simple model to test the effectiveness of this approach is formulated. The fourth chapter deals with the results of the study while the last chapter examines and evaluates the applicability of the prediction model as well as the comparative approach that will be introduced in this thesis.

CHAPTER II

ORGANIZATION OF THE STUDY - A REVIEW OF THE EXISTING TECHNIQUES FOR PLANNING ANALYSIS AND PREDICTION

CHAPTER II

ORGANIZATION OF THE STUDY - A REVIEW OF THE EXISTING TECHNIQUES FOR PLANNING ANALYSIS AND PREDICTION

2.1 Descriptive Approach Versus Analytical Approach

It is obvious from a glimpse of some introductory texts that the outlined fundamental areas of studies for the preparation of policy plans are scarcely the same. For example, Goodman singles out the studies of community facilities whereas Chapin emphasizes the studies of activities system.[1] Deliberately, taking consideration of all possible alternatives, the many basic studies can be broadly classified into five areas as the following:

1. Population studies, such as population estimation and forecasting, population characteristics studies, etc.
2. Economic studies, such as economic base study, input-output study, income expenditure analysis, employment study, etc.
3. Transportation and support system studies, such as traffic count, O-D survey, travel behaviour study, community utilities study, etc.
4. Social studies, such as activities study, social attitude survey, etc.
5. Land use studies, such as existing land use survey, housing

---

[1]See Goodman, op. cit., and F. Stuart Chapin, Jr., Urban Land Use Planning (Urbana: University of Illinois Press, 1965).

market analysis, community facilities study, land use forecasting, etc.

Two different approaches have been employed for these basic studies. The first is descriptive approach which is in the nature of an investigation of the specific problem with little consideration of its causalities characterizes the traditional techniques for planning analysis. The recently developed analytical approach, on the other hand, aims at exploring and explaining the causal factors and the relationships thereof for the implicit urban problems.

To say that the traditional planning techniques are descriptive is not to say that they are definitely of little value for explanation and prediction. In fact, some of the techniques, such as the multiplier methods used in economic studies, capture some of the cause-effect relationships[2] which can further be applied to prediction. Nevertheless, descriptive techniques provide indispensable information for analytical studies. Another advantage of using the descriptive techniques for prediction is that the operational simplicity is challenging. However, due to the lack of a rigorous theoretical framework, explanation and prediction attained from the descriptive approach can hardly be desirable.

---

[2]As an illustration, take $E_s = (k/(1-k))\ E_b$ in which E_s and E_b are employments in service and basic industries respectively, the change of basic employment (cause) partially explains the change of service employment (effect) in this formula. Another example is the population projection by extrapolation of a time series. Although change of time yields change of population, this is not the sufficient condition for the population change.

As urban problems became more complex, with the growing awareness of the fact that the forecast of urban growth involves the consideration of a great number of influencing factors, such as the population trends, socio-economic characteristics of the area, economic factors, government policies, traffic conditions and accessibilities, and so on, there is a strong feeling that:

> Metropolitan planning is often handicapped by attempts to superimpose outdated technologies and methodologies upon old world planning concepts, . . . [for] traditional urban growth theories provide highly imperfect analytical tools for the preparation of operationally useful forecasts of land utilization.[3]

Due to this necessity and to the impetus of modern technology, planning techniques changed radically during the past decade. This is manifested in the prevalence of the analytical approach over the traditional descriptive approach which has withdrawn to such a position for providing basic information for further analysis only.

2.2 The Role of Models in the Planning Process

The most important and effective tool for planning analysis as well as for prediction is the use of models.

A model is merely a simplified representation of the real world; it may be in the form of a small-scale physical object, a diagram or

---

[3]Center for Real Estate and Urban Economics (CREUE), Jobs, People and Land: Bay Area Simulation Study, Special Report No. 6 (Berkeley: Univ. of California, 1968), p. 1.

mathematical statements.[4] Wagner explains the importance of a model in this way:

> Constructing a model helps you put the complexities and possible uncertainties attending a decision-making problem into a logical framework amenable to comprehensive analysis. . . . [It] is a vehicle for arriving at a well-structured view of reality.[5]

This is not all of what a model can provide, however. Reflecting the usefulness of models in planning, planning agencies have reported that there are three major purposes of using a model:

1. To improve the rationality of the planning decisions.
2. To analyze and evaluate the policy alternatives.
3. To forecast and analyze urban growth.[6]

Above all, the greatest advantage seems to be its educational benefit. "In order to build even the most simple model it is necessary to think very clearly about the phenomena under investigation, and to understand them even more thoroughly than in most traditional descriptive or anecdotal approaches."[7]

---

[4]N. Paul Loomba, Linear Programming, An Introductory Analysis (New York: McGraw-Hill, 1964), p. 14.

[5]Harvey M. Wagner, Principles of Operations Research, With Applications to Managerial Decisions (Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1969), p. 10.

[6]G. C. Hemmens, "Planning Agency Experience with Urban Development Models and Data Processing", Journal of the American Institute of Planners, Sept. 1968, pp. 323-327.

[7]P. Cowan, J. Ireland, and D. Fine, "Approaches to Urban Model-Building", Regional Studies, vol. 1, Dec. 1967, pp. 163-172.

Therefore, using a model in planning would not only provide the desirable outcome for decision-making, but also impose the rigor required of planners to explore the implications of urban problems, as Lowry points out:

> The process of model building is educational. The participants invariably find their perceptions sharpened, their horizons expanded, their professional skills augmented.[8]

2.3 Existing Models Applied to Planning

According to the purposes of model-building, there are four kinds of models which are now in use in urban planning:[9]

1. Allocation models distribute established totals of population, employment, or land use. For example, the Employment Location Submodel developed as a component of the BASS Model (Bay Area Simulation Study) allocating the total forecasted employment into various categories,[10] and Bourne's probabilistic model based on the theory of Markov chains forecasting land occupancy in a central city[11] are typical deterministic and probabilistic allocation

---

[8] Ira S. Lowry, "A Short Course in Model Design", Journal of the American Institute of Planners, May 1965, pp. 158-165.

[9] Hemmens, op. cit.

[10] CREUE, op. cit., Chapter 3, pp. 95-178.

[11] L. S. Bourne, Forecasting Land Occupancy-Changes Through Markovian Probability matricies, A Central City Example, Research Report No. 14, Urban development study (Toronto: University of Toronto, 1967).

models respectively.

2. Policy impact models measure the effect of changes in some public facility systems on the pattern of land development or of population or employment distribution. For example, the agencies of the Detroit Regional Transportation and Land Use Study have developed Facilities Model to measure the effect of new land use and activity distribution on the sewer and water system.[12]

3. Activity estimation models measure the amount and location of an activity, such as retail shopping or recreation, which results from a given land use or population distribution. The typical example is the gravity model in transportation study which estimates the amount of traffic generated between different zones with given populations.[13]

4. Aggregate projection models estimate future levels of employment or population for the area as a whole, and in all cases are intended to provide inputs to allocation models used by the same agency. The Employment and Population Submodel used in BASS providing inputs of aggregated population and employment to another allocation submodel is a typical example.[14]

---

[12]Highway Research Board, Urban Development Models, Special Report 97 (Washington, D.C., 1968), p. 257.

[13]See, for examples, R. J. Bouchard and C. E. Pyers, "Use of Gravity Model for Describing Urban Travel" Traffic Research Record No. 88, 1965, pp. 1-43, and R. L. Smith, "Gravity Model Theory Applied to a Small City Using a Small Sample of O-D Data" Traffic Research Record, No. 88, 1965, pp. 85-115.

[14]CREUE, op. cit., Chapter 2, pp. 29-94.

The combination of the four kinds of models described above provides a total picture of methodology contemporarily applied to comprehensive planning. The aggregate models forecast the future population and employment levels which are distributed to various categories by allocation models. Activity models estimate the various activities generated as a result of allocations. Finally, the policy impact models examine and evaluate the impacts of the planning decisions upon the area concerned.

It has been mentioned in chapter I that the pioneers in model-building related to urban planning were the transportation engineers who undertook their studies by means of these revolutionary techniques. Urban development models, the other main stream of models in planning, were mostly devised under the inspiration and influence of transportation models. Thus, transportation and land use constitute two broad areas of model-building in city planning to-day.[15]

Studying and forecasting the travel behavior in an urban area, the transportation models usually employ multiple regression techniques or least square fitting to estimate the parameters based on empirical observations. Similarly regression is also a widely used tool for building land use models. As one of the possible routes to model-building, the regression models and the like require careful and successive

---

[15] A. M. Voorhees, "The Nature and Use of Models in City Planning", Journal of the American Institute of Planners, May 1959, pp. 57-60.

empirical investigations.[16] On the other hand, some successful land use models were developed in terms of linear programming to optimize an objective function,[17] of simultaneous equations to express the relationships of factors shaping land use patterns,[18] or of stochastic process to give outcomes in probability distributions,[19]--the second possible route of model building which is "a deductive one requiring the careful construction of theory in advance."[20]

2.4 Problems of Using Models in Planning

Some inherent difficulties hinder the wider spread of using this new methodology in the planning profession. These difficulties may be external as well as internal.

The external difficulties which have been experienced by planning agencies are mainly the results of insufficiency of computer facilities

---

[16] B. Harris, "Conference Summary and Recommendations" in Highway Research Board, Urban Development Models, Special Report 97 (Washington, D.C., 1968), pp. 3-17.

[17] For example, refer to the Penn-Jersey Regional Growth Model. See: N. A. Irwin, "Review of Existing Land Use Forecasting Techniques", Highway Research Record, No. 88, 1965, pp. 182-216; and B. Harris, Linear Programming and the Projection of Land Uses, Penn-Jersey Transp. Study, PJ Paper No. 20, Nov. 1962.

[18] For example, see: I. S. Lowry, A Model of Metropolis, Memorandum RM-4035-RC (Santa Monica, Calif.: The RAND Corp., 1964).

[19] This kind of model is referred to as probabilistic model which predicts a whole distribution of outcomes in terms of probabilities as against the deterministic model which yields prediction of a single outcome. For example, see L. S. Bourne, op. cit.

[20] Harris, see fn. 16.

and the related operational problems such as inadequate experience in computer programming, difficulties in collecting relevant data, and discrepancy in communication between the model-builders and the model-users.

Internal difficulties in model-using reflects the theory and attitude of current model-building process. Therefore it merits a more detailed discussion below.

Most models are built on the basis of a particular problem so that they are unlikely to be transferable to or freely used by other agencies. This is the general weakness of a model using the regression technique based on local observations. The parameters are determined upon the observations of a particular case, and even the predictor-criterion relationships devised in this manner may not also be applied competently to the other cases. The situation will be worse if the model involves some sort of judgment rather than purely mathematical manipulation. The CATS Model (Chicago Area Transportation Study), for example, is absolutely not transferable and even the results of the forecast are probably not reproducible unless done by the same study team.[21]

It also recognizes that:

> The validation and testing of models will require not only statistical tests, but also their application in urban areas over time and in different cities.[22]

---

[21] Irwin, op. cit., p. 187.

[22] Harris, op. cit., p. 12.

Judging from the above criterion, most existing models fail to be universally applicable. Under such circumstances, the planning agencies could either build their own models and accept the new methodology, or completely discard this new technology and keep on using the "classical" techniques.

Further, the regression models which were widely used in the studies of travel behavior and urban development seemingly have less rigorous theoretical frameworks to assure correct predictions. The argument is that most multiple regression models are derived from the so-called stepwise regression method which selects the "best" regression on the trial-and-error basis with empirical data instead of digging into the theoretical implication of the problem. The general existence of multicollinearity in this kind of models even makes the situation worse.[23] Thus, their lack of transferability is obvious.

Another inherent problem concerning the use of models is the complexity of the models themselves. Although complexity implies increased approximation of the real world situation, the difficulty in operation, manipulation, and the increased measurement error possibly affect its superiority over a model in simpler form. Some models are devised in a chain of several submodels such that the output of the preceding submodel becomes the input of its successors. It has been found

---

[23]For more detailed account for the multicollinearity, see S. L. Chan, Forthcoming Multicollinearity in Transportation Models, (Forthcoming Master's Thesis, School of Planning, UBC, 1970).

that operation in chains always compounds the errors of the separate sub-models and leads to rapid deterioration of prediction.[24]

Above all, the models consisting of a number of predictors which are most desirable to simulate the reality encounter additional difficulty when it comes to forecasting. Whether it be a transportation model or land use model, and whether the model takes the form of regression, linear programming or simultaneous equations, the forecasting of a criterion must be a result of a series of exogenous forecastings of all the predictors involved. The work concerned is cumbersome and highly undesirable.

In summary, the complex model is inadvisable for the purpose of planning prediction in a sense that it has operational difficulties and leads to an undesirable deterioration of prediction. Thus, one should prefer the simple model.

2.5 Simple Model Versus Complex Model

It is not easy to distinguish whether a model is simple or complex since the scale of complexity of a model is subjective rather than objective, continuous rather than discrete. However, for the convenience of this study, a simple model should satisfy the following conditions:

1. There is only one predictor in the model.

---

[24] William Alonso, "Predicting Best with Imperfect Data", Journal of the American Institute of Planners, July, 1968, pp. 248-255.

2. If the model exists in a group of several submodels, the simple model should be independent of or parallel to other models of this group.

3. The operation of the model should not involve any complicated mathematical procedure. It should be a linear model or at most linear in a simple logarithm transformation.

Alonso distinguishes two types of error in a model: the measurement error which arises from inaccuracy in assessing a magnitude and can be accumulated during the mathematical procedures, and the specification error which arises from a misunderstanding or purposeful simplification in the model.[25] The simple model predicts the criterion with less measurement error that results solely from the inaccuracy of the input data, but as the complexity of the model increases, the compounding of measurement error increases. Because of the oversimplicity of the simple model, its specification error is large, but decreases as the complexity of the model increases. The relationships between the errors of a model and its complexity are shown in Figure 1.[26]

It is clear that a simple model has smaller measurement error but larger specification error compared with a complex one, and that making the simple model complicated reverses the situation entirely. Analogizing model building to gold mine digging and analyzing the situation by means of the theory of games, Bartos advises that "it is better, in general, to

---

[25] Ibid., p. 248.

[26] Ibid., p. 251.

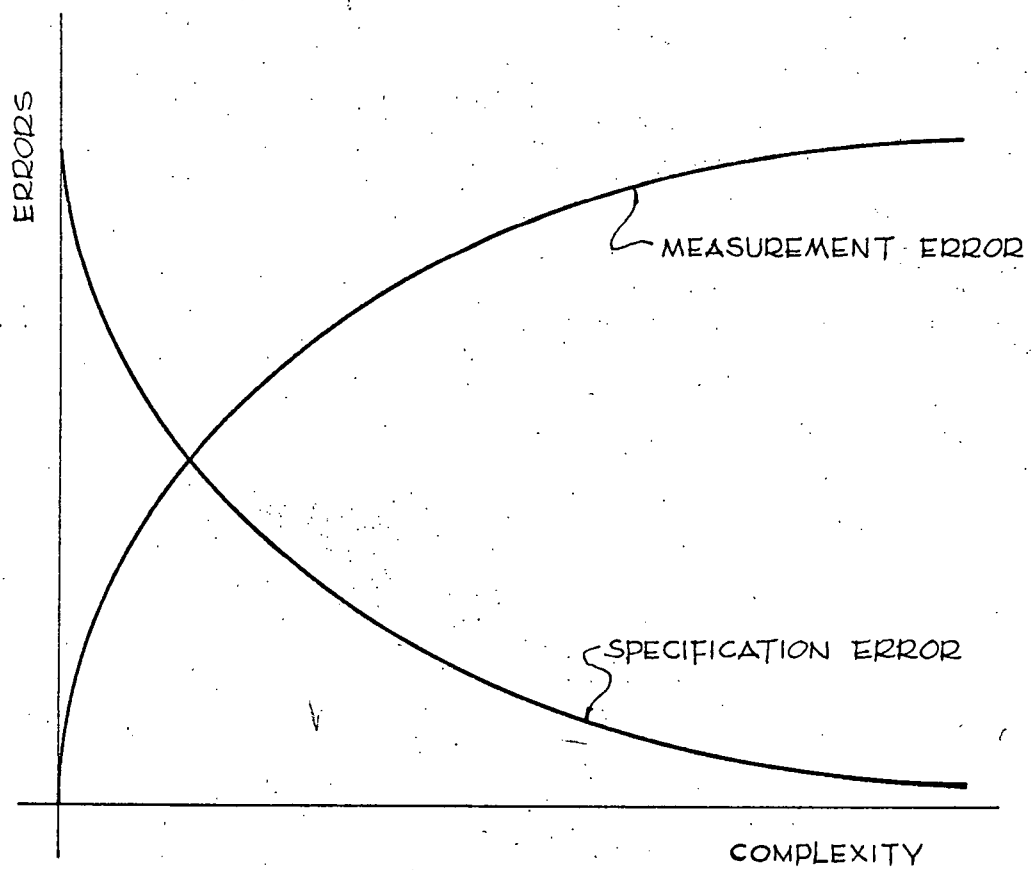


FIG. 1 RELATIONSHIP BETWEEN ERRORS AND COMPLEXITY OF A MODEL

explore many models in a preliminary fashion than to explore one model in depth" and that "to explore many simple models is predicated on the assumption that making a simple model more complex is not worth the effort if the simple model shows no promise to start with."[27]

With the very high degree of generality achieved from discarding sufficient detail, a simple model might become so general as to be useless when applied to the real world.[28] Therefore, when it comes to make simulation more realistic, we cannot beat complex model.

It is a dilemma to choose an exact type of model in the process of model building. The advantages of a simple model are the simplicity of operation and the small measurement error in prediction even with imperfect data. The notorious weakness, on the other hand, is the poor theoretical framework in describing and explaining the urban phenomena.

In attempting to solve this problem, the following study aims at devising a model for prediction in its simplest form, in which the specification errors are decreased indirectly from introducing a system of techniques.

---

[27] Otomar J. Bartos, Simple Models of Group Behavior (New York: Columbus University Press, 1967), p. 319.

[28] Cowan et al., op. cit.

2.6 The Evolution of the Comparative Approach[29]

Supposing a simple regression model, $Y = a + b X$, predicts the CBD floor space (Y) from the total population of a city (X). Although it is obvious that the amount of CBD floor space required is a direct function of the city population, it could be argued that the model is too general to be realistic. Such factors as the geographical location and economic situation of the city, the socio-economic characteristics of the population, etc., exhibit substantial influences in shaping the CBD. These cause-effect relationships can never be adequately explained by the above model.

Applying Alonso's interpretation of model errors, the hypothetical specification error of this simple model can be shown in Figure 2.

Imagine that there is a system of submodels of whatever level of complexity parallel to and completely independent of this simple model such that the output of the former does not affect the measurement error of, nor becomes the input directly to the latter. The scale of complexity of this system of submodels, if drawn in the graph, is perpendicular to that of the simple model. Thus, the composite specification errors of these models as a whole, supposing they can be connected in such way successfully, is a surface in the three dimensional space shown in Figure 3 below:

---

[29] For definition, see section 2.8.

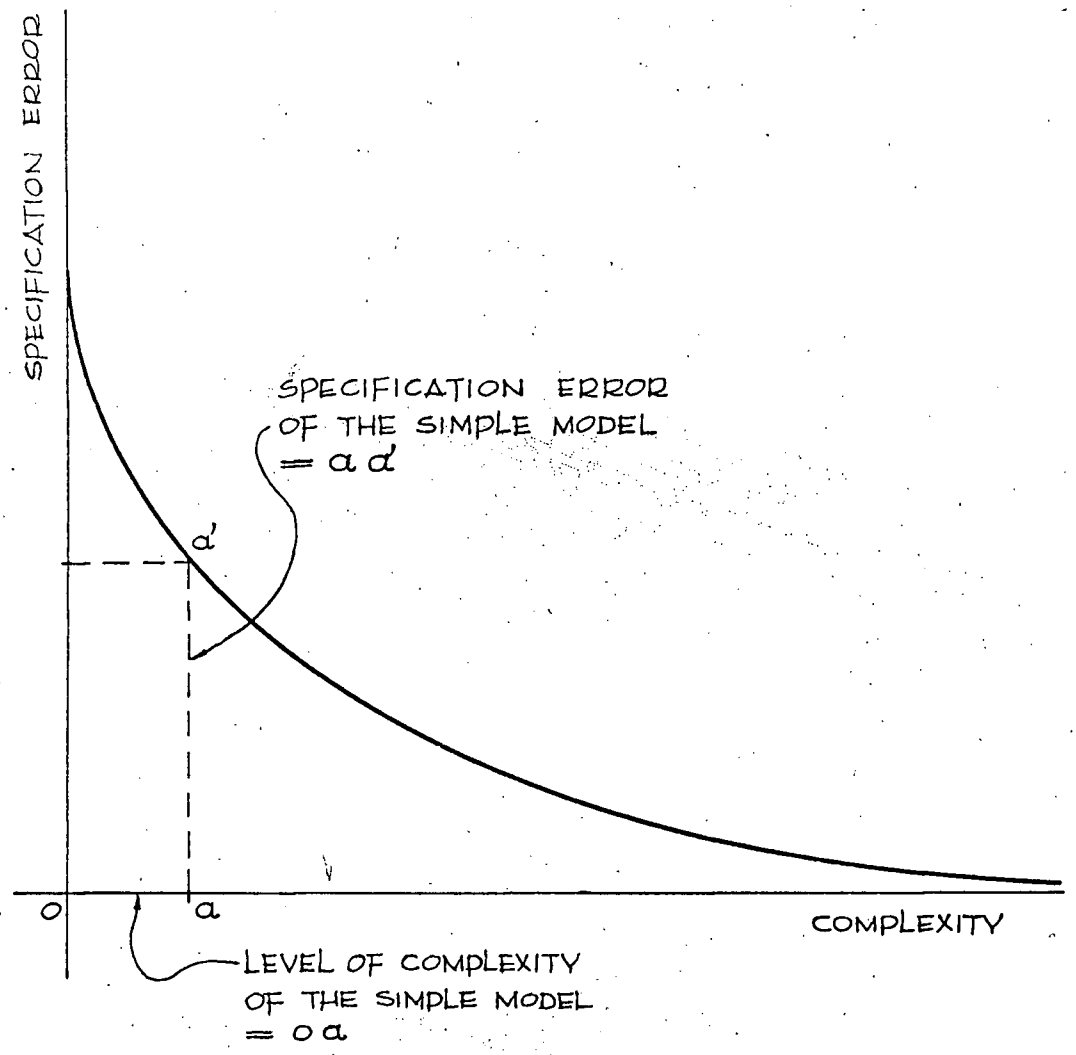


FIG. 2 SPECIFICATION ERROR OF A SIMPLE MODEL (HYPOTHETICAL)

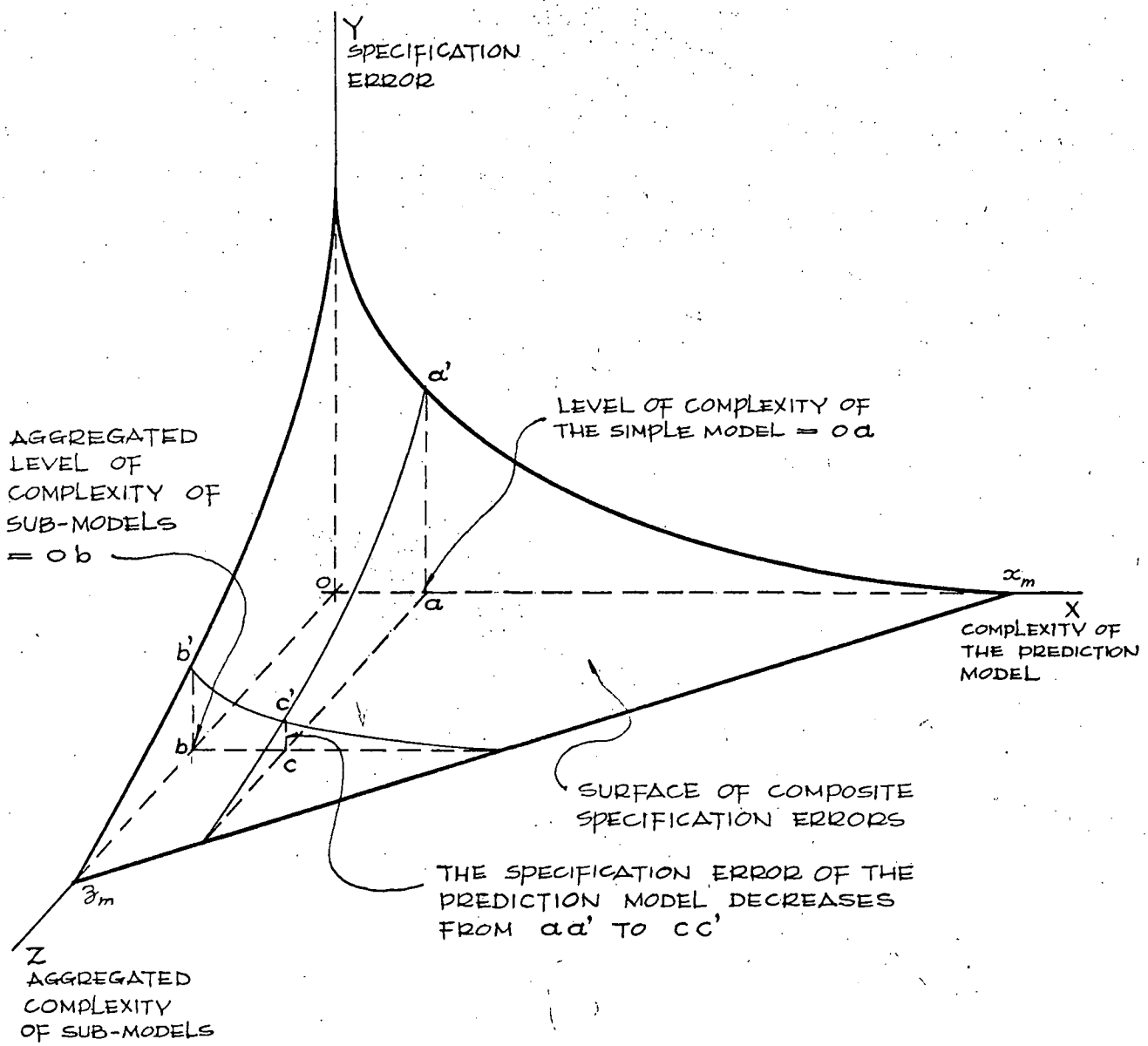


FIG. 3 COMPOSITE SPECIFICATION ERROR OF A GROUP OF MODELS

In Figure 3, the line $X_m Z_m$ represents the optimal combinations of models on which the specification errors are zero. The surface bounded by the three curves is the specification errors resulting from any combination of the complexities of the models. Thus, the specification error, aa', of the simple model will decrease substantially to cc' if a set of parallel submodels with aggregated complexity operates simultaneously. In other words, the ignored causal factors in the simple model are recaptured from another way, yet all the merits of this simple model are preserved.

The remaining problem is whether this system of submodels exists. Harold M. Mayer suggests that:

> By modern methods, such as factor analysis, it is possible to classify cities and metropolitan areas by the extent of their similarities and differences with respect to a large number of variables. . . . [It] would be of assistance in determining the extent to which successful planning solutions to the problems of a given city or metropolitan area might be transferable to other cities or metropolitan areas which are similar with respect to relevant characteristics. By such methods, the extent of similarity or difference by any or all combinations of characteristics may be measured; thus, multivariate classifications of cities and areas of geographis significance could be useful to the planner.[30]

Inspired by these remarkable statements, encouraged by the rapid development of the modern computation algorithms and also awakened by the planners' negligence of this virgin land, a system of techniques are

---

[30] Harold M. Mayer, "Urban Geography and City and Metropolitan Planning", Urban Research and Policy Planning, L. F. Schnore and H. Fagin, ed. (Beverly Hills, Calif.: Sage Publications, Inc., 1967) p. 229.

developed in this thesis such that the urban problems can be analysed from a comparative point of view.

2.7 The General Hypothesis

It is hypothesized that:

> Comparative approach is an effective way to analyze urban problems, and simple planning prediction models can be improved by the use of comparative techniques.

2.8 Definitions

Some important terminologies applied above and in the following chapters are defined below:

1. Model: A simplified representation of some subjects of inquiry, such as objects, events, processes, systems. It may be physical, analog, mathematical, or a combination of these, differing from the degree of abstraction imposed, as below:[31]

Scale of abstraction

least abstract		most abstract
Physical models (mostly for description)	Analog models (for example, colored map, graph, diagrams, etc.)	Mathematical Models (mostly for explanation & prediction)

2. Technique: A systematic method based on logical or mathematical

---

[31] C. W. Churchman, R. L. Ackoff, and E. L. Arnoff, Introduction to Operations Research, Chapter 7 (New York: John Wiley and Sons, 1957) p. 151.

relationships to describe, analyze or forecast some specific processes or problems.

3. Simulation: Simulation is the technique centering upon the construction of a type of machine called a simulator, whose functioning is intended more or less directly to resemble the behavior of a specific existing or potentially existing operational system. A simulator usually consists of a digital computer, plus a program of instructions, and data.[32]

4. Prediction, forecasting and projection: In sense the three terms are synonymous, meaning the process of purposely estimating the future based on the knowledge of past and/or present conditions. In this study, "prediction" means the estimation of the future as well as the existing conditions by means of some techniques or models; "forecasting" means the prediction of the future only; "projection" means the forecasting of the future under the condition that the present situations remain unchanged.

5. Regression analysis: The statistical process of describing by an equation the relationship between a dependent variable (criterion) and one or more independent variables (predictors) so that the equation so derived describes the relationship, as represented by a number of observations, with a minimum of error.[33] Simple

---

[32] James R. Jackson, "Simulation as Experimental Mathematics", Symposium on Simulation Models: Methodology and Applications in the Behavioral Sciences, A. C. Hoggatt and F. E. Balderston, ed. (South-Western Publishing Co., 1963), p. 245.

[33] Irwin, op. cit., p. 184.

regression consists of one independent variable only whereas multiple regression consists of more than one independent variable.

6. Factor analysis: A multivariate statistical technique to reduce and orthogonalize the dimensionality in correlated systems of measurements so that a smaller number of independent factors can be attained for further analysis.
7. Discriminant analysis: A procedure for estimating the positions of lines or linear functions that best separate classes or groups in order to test whether these classes or groups are significantly discriminant, if so, the relative importance of the factors contributing to the discrimination is detected.
8. Classification: Given the "a priori" groups, it is the procedure for estimating the probabilities of an individual belonging to those groups in order to assign the individual into one of the groups properly.[34]
9. Grouping: Given a population, it is a procedure to classify the population into groups according to certain characteristics.[35]
10. Comparative approach: It is the manner for acquiring analytical findings by means of comparing the specific characteristics of a subject with those of others so that generalization and deviations can be detected, described, explained or even predicted.

---

[34]For detailed descriptions of these multivariate statistical techniques, see standard text, for example, William W. Cooley and Paul R. Lohnes, Multivariate Procedures for the Behavioral Sciences (New York: John Wiley and Sons, Inc., 1962).

[35]For detail, see Appendix A.4.

2.9 Summary

Reviewing the situation of building and using models in planning, it appears that the prospects are promising. However, the major problems that may hinder future development could be either that the model is too complex to be operated easily without much error, or that it is too simple to capture sufficient causalities to build up a rigorous theoretical framework which is one of the necessary conditions to make the model of universal application. Comparative Approach to planning analysis and prediction is thus introduced, trying to resolve this dilemma.

CHAPTER III

THE COMPARATIVE APPROACH TO PLANNING ANALYSIS AND PREDICTION

CHAPTER III

THE COMPARATIVE APPROACH TO PLANNING ANALYSIS AND PREDICTION

3.1 Concepts of the Comparative Approach

To study urban phenomena using a comparative approach is not new to some urban researchers. Nelson studied and classified the American cities according to their service functions;[1] Moser and Scott classified British towns according to their social and economic characteristics;[2] as for more specific study, it was Reynolds who employed some sorts of specialization quotients to compare ninety American CBD's.[3] However, the comparative approach is badly neglected by planners. Bogue complains that:

> [There] is the scarcity of "comparative" population and urban studies from which one can ascertain whether or not general trends for all metropolitan areas are also characteristic of most metropolitan areas individually. . . . [The deficiencies] characterize most areas of social research where "groups of people", rather than "persons" are the units in terms of which the data should be collected and analyze.[4]

---

[1]For detail, see H. J. Nelson, "A Service Classification of American Cities", Economic Geography, Vol. 31, 1955, pp. 189-210.

[2]For detail, see C. A. Moser, and W. Scott, British Towns: A Statistical Study of their Social and Economic Differences (London: Oliver and Boyd, Ltd., 1961).

[3]For detail, see R. B. Reynolds, "Retail Specialization of CBD's" Journal of the American Institute of Planners, No. 1960, pp. 313-316.

[4]D. J. Bogue and D. L. Harris, Comparative Population and Urban Research Via Multiple Regression and Covariance Analysis (Scripps Foundation for Research in Population Problems, Miami Univ. Press, 1954), p. 1.

It also appears that "procedures for making comparative studies of groups of people must constitute an important part of social science methodology".[5]

A comparative study of groups of people may attain additional advantage that the mathematical models which involve the consideration of human values and which are derived from the usual approaches can hardly obtain. It is because human values and choices have to be quantified for use in these models, and it is precisely at this point that the models are most severely criticized on the grounds that such things cannot be adequately quantified to make them worthwhile.[6] Instead of explaining that the attitude of a group of people changes because its influencing factors change simultaneously, as most models do, the comparative approach emphasizes the fact that the attitude of this group of people differs from that of others because they are actually different in some other characteristics, and that if the discrepencies of these characteristics decrease, the likelihood of having the similar attitude increases. Thus, it is interested in discovering the facts causing the change of attitudes among groups of people and in deducing generalities applicable to such groups. Predictions made on the basis of these generalities need not precisely quantify the human factors which are considered as collective efforts affecting the characteristics used for comparison. For example,

---

[5] Ibid.

[6] John Dakin, "Models and Computers in Planning", Plan Canada, The Town Planning Institute of Canada, Vol. 6, No. 1, July, 1965, pp. 11-35.

in order to predict the buying power of a group of people with increase in average income, it is necessary to consider such factors as the socio-economic characteristics, the population composition and characteristics, the propensity to save with increase in income, and so forth, then to construct logical and mathematical relationships between buying power and all the above factors, tangible as well as intangible. However, from the comparative point of view, if a second group of people which is similar to the first group in many related characteristics did increase their buying power upon the similar increment of income, it can be predicted that the first group of people will do so under the similar circumstances.

Two levels of studies have the possibility of employing the comparative approach. At the local or city-wide level, it is concerned with comparing the characteristics of small groups of people within a community or a city whereas at the inter-city level, it deals with variations among cities or metropolitan areas as a whole. Both levels of studies are based on the premise that there are variations from the grand average of the groups concerned and also from the average for various groups.

It is clear that the objective of this approach is to explain the differences among groups of people or among several areas so that generalized patterns can be derived for prediction. Furthermore, it is postulated that from the comparative point of view, people as well as urban areas can be classified into groups according to stated characteristics.

The most important concepts to compare several subjects are similarity and comparability. The former answers the conceptual question of "what it is to be compared". The latter answers the technical question of "how and whether it can be compared".

In a sense, to compare something means to examine their similarities. The term "similarity" has the following aspects:

1. Relative aspect

   It is meaningless to say that A is similar to B without recognizing in what ways they are similar. Therefore similarity is a term relative to a set of predetermined criteria which characterizes the group thus formed.

   The fact that several subjects are similar in respect to a specific set of characteristics does not imply that they are also similar in other ways. However, the members of a similar group do not necessarily have to exist at the same time. For example, Sakyamuni and Jesus could be in the same group because they were great religious philosophers and had tremendous influences on their descendents, the discrepancy of time does not affect their similarities at all.

2. Indirect aspect

   Similarity can not be measured directly, nor is there any way to express similarity properly.

Comparing A's annual income of $12,000 with B's $11,500, we may say that their income levels are similar. As a matter of fact, we are not measuring their similarity, but dissimilarity. That the difference, $500, is within our acceptance limit of dissimilarity enables us to assert that similarity does exist between them. It is a common sense that the lesser the difference between two subjects, the more likely the similarity between them. In other words, to minimize dissimilarity yields the same result as to maximize similarity.

3. Subjective aspect

It is subjective more than objective to say that two subjects are similar. Since we do not have an efficient way to measure similarity directly, we should determine acceptance level of dissimilarity[7] to perform paired comparison. The determination of this acceptance level varies according to the nature and purpose of the analysis, and to the analyst himself. The difference of income level, say $10,000, for instance, may be thought stringent for the purpose of classifying millionaires into groups, but ill-defined as it is applied to the low income groups. The analyst's point of view is another important factor affecting the determination of acceptance level as long as a workable and definite method for this purpose has not been developed.

---

[7]Acceptance level of dissimilarity is defined as the maximum degree of dissimilarity within which similarity between subjects is asserted.

4. Representative aspect

Once two or more subjects have been classified as a similar group according to some characteristics, the members can be represented by some common features which characterize the whole group. The mean(s), or the centroid,[8] is generally used to represent the characteristics of the whole group as well as the individual members. The mid-point of class intervals for statistical analysis is an example. However, it will imply some losses of information when it comes to represent the individual scores by a unique centroid, and the homogeneity of the group will also deteriorate as these losses of information increase. The degree of homogeneity, as is measured indirectly from the amount of loss of information, and the degree of similarity, as is measured indirectly from the degree of dissimilarity, constitute the two essential properties of a group with similar characteristics. In order to get a group representative of all its members, the members of this group must be similar, that is, the differences between every two members must be as small as possible, yet the group must be homogeneous, that is, the total loss of information must be minimized.

These four characteristics of similarity, namely, relative, indirect, subjective, and representative aspects are fundamental concepts to form groupings for further studies in the Comparative Approach.

---

[8]Centroid is a vector of means of the measuring variables.

On the other hand, the concept of comparability firmly relates to measurement which is "the assignment of numerals to things so as to represent facts and conventions about them".[9] To compare some subjects, their characteristics must be quantified by means of scales of measurement which can be norminal, ordinal, interval or ratio. However, in order to make things comparable, two conditions must be satisfied:

1. Consistency of scale

   Although all the characteristics concerned for comparison may not be in terms of the same scale, a specific characteristic should be expressed consistently in the same scale using the same unit, if any.

2. Orthogonalization for multivariate comparison

   As in most cases, the comparison involves consideration of more than one characteristic. These quantified characteristics must be orthogonal to each other because of the following reasons:

   a) The orthogonalization of quantified characteristics (or dimensionalities) assures minimum intercorrelation among them which otherwise will deteriorate the rational framework for comparison.

---

[9] S. S. Stevens, "On the Theory of Scale of Measurement", Philosophy of Science, A. Danto, ed. (New York: The World Publishing Co., Meridian Books, 1967), pp. 141-149.

b) In case of measuring similarity by means of distance (or dissimilarity) in a multi-dimensional score space, orthogonalization is a necessary mathematical condition.

In summary, the Comparative Approach to planning analysis and prediction is built on the premise that variations exist among subjects. These variations can be detected, analyzed, and applied to prediction, based on the ground rules of similarity and comparability.

3.2 Techniques Applicable to the Comparative Approach

Two important steps have to be dealt with in the Comparative Approach. The first is to form groupings based on the concepts of similarity and comparability. The second is, if the groups have already been formed, to examine whether they are significantly different and to detect the salient characteristics so that a generalized pattern among these groups can be derived. For accomplishing these, five established multivariate statistical methods can be applied:

1. Factor analysis

Factor analysis serves as a "black box" to change the structure of the characteristics or variables used for comparison to another structure of factors with the following properties:

a) The factors obtained are independent of each other;[10] the

[10]Factors can purposefully be oblique to each other in some particular cases which are out of the scope of this study. For detail, see

intercorrelation between every two factors approaches to zero so that they have been orthogonalized.

b) The scale of measurement for the factors is unique. Every factor score of the subject is expressed in terms of the standard deviation of that factor, with mean zero and standard deviation one.

c) The number of factors attained is usually less than the number of original variables concerned.

Because of the first two properties, characteristics of the subjects become comparable, and because of the last property, a great number of variables becomes manageable in terms of fewer factors.

2. Hierarchical grouping

The standardized and orthogonalized factors scores of the subjects in question are compared so as to form groupings based on the criteria that the members of the specific groups must be similar and homogeneous in terms of these factors.

Several grouping algorithms have been formulated. The one that is most applicable to this study forms a group of two members[11] by

---

H. H. Harmon, Modern Factor Analysis (Chicago: University of Chicago Press, 1960).

[11]The two members in a grouping procedure may be an individual, another group, or both.

measuring the distance between them and at the same time optimizing an objective function which may be of maximizing the similarity or of minimizing the loss of information.[12] Further, a non-parametric grouping technique has been developed to form groups of individuals measured in norminal or ordinal scale.[13]

The subjects for grouping can be cities, metropolitan areas, sub-areas of a community, groups of people, or whatever individual entities, depending on the nature and purpose of the study.

3. Multivariate analysis of variance

This technique enables one to detect two kinds of variations after groupings have been formed. The first is the variations from the grand mean of the groups, the second is the variations from the mean for various groups. Statistical procedures are provided to test the significance of these variations.[14]

The multivariate analysis of variance is usually one of the components of a complete grouping procedure mentioned above or of a discriminant procedure introduced below.

---

[12]For detail of grouping algorithms, see Appendix A.4.

[13]The non-parametric grouping and discrimination are well developed by Kendall, see M. G. Kendall, "Discrimination and Classification". In P. R. Krishnaiah (ed.), Multivariate Analysis, Proceedings of an International Symposium held in Dayton, Ohio, 1965 (New York: Academic Press, 1966), pp. 165-185.

[14]Cooley and Lohnes outlined several tests applicable to the situation. The most useful one may be the Wilks' lambda criterion. For

4. Discriminant analysis

If the "a priori" groups are given, the discriminant analysis aims at determining a set of linear functions to give optimal separation of the groups. At the same time, significant test of the discrimination can be performed on the basis of these linear functions. The most important feature of this technique is that, by examining and comparing the correlations of the factors with these functions, the relative importance of the factors contributing to the discrimination of the groups can emerge.

By means of discriminant analysis, insights of the urban structure and problem can be attained.

5. Classification

The purpose of classification technique is to assign an individual subject to one of the "a priori" groups. Several methods, have been derived. The most desirable one involves computing the probabilities of the individual assigning to various groups.[15] The highest probability means the most likelihood that the individual will be the member of a certain group. After this assignment, generalities of the individual's characteristics can be

---

detail, see W. W. Cooley and P. R. Lohnes, op. cit., pp. 60-115, or T. W. Anderson, An Introduction to Multivariate Statistical Analysis (New York: John Wiley and Sons, 1958), Chapter 10.

[15]For parametric methods, see Cooley and Lohnes, op. cit., Chapter 7. For non-parametric methods, see M. G. Kendall, op. cit.

attained and future trends can be derived with reference to the other members of the group.

Each of the above multivariate statistical methods can be used individually, or as one of the components of a system of techniques for comparative study, depending on the nature of each case.

3.3 Methodology--An Analog Model for the Comparative Approach

The objective of the comparative approach is twofold: (1) to study and analyze the characteristics of a subject in depth, and (2) to derive generalized behavioral patterns which serve as a means for prediction.

Synthesizing the applicable multivariate statistical techniques and keeping the above objectives in mind, an Analog Model for comparative approach is derived as in Figure 4.

As shown in Figure 4, the methodology for the comparative approach consists of two models:

1. Comparative Analysis Model

   This model composes of four submodels, namely, factor analysis, grouping, discriminant and classification submodels.

   Aggregated data are put in to the model. If the data are comparable, and if the subjects have already been grouped, the discriminant submodel is employed directly and analytical results

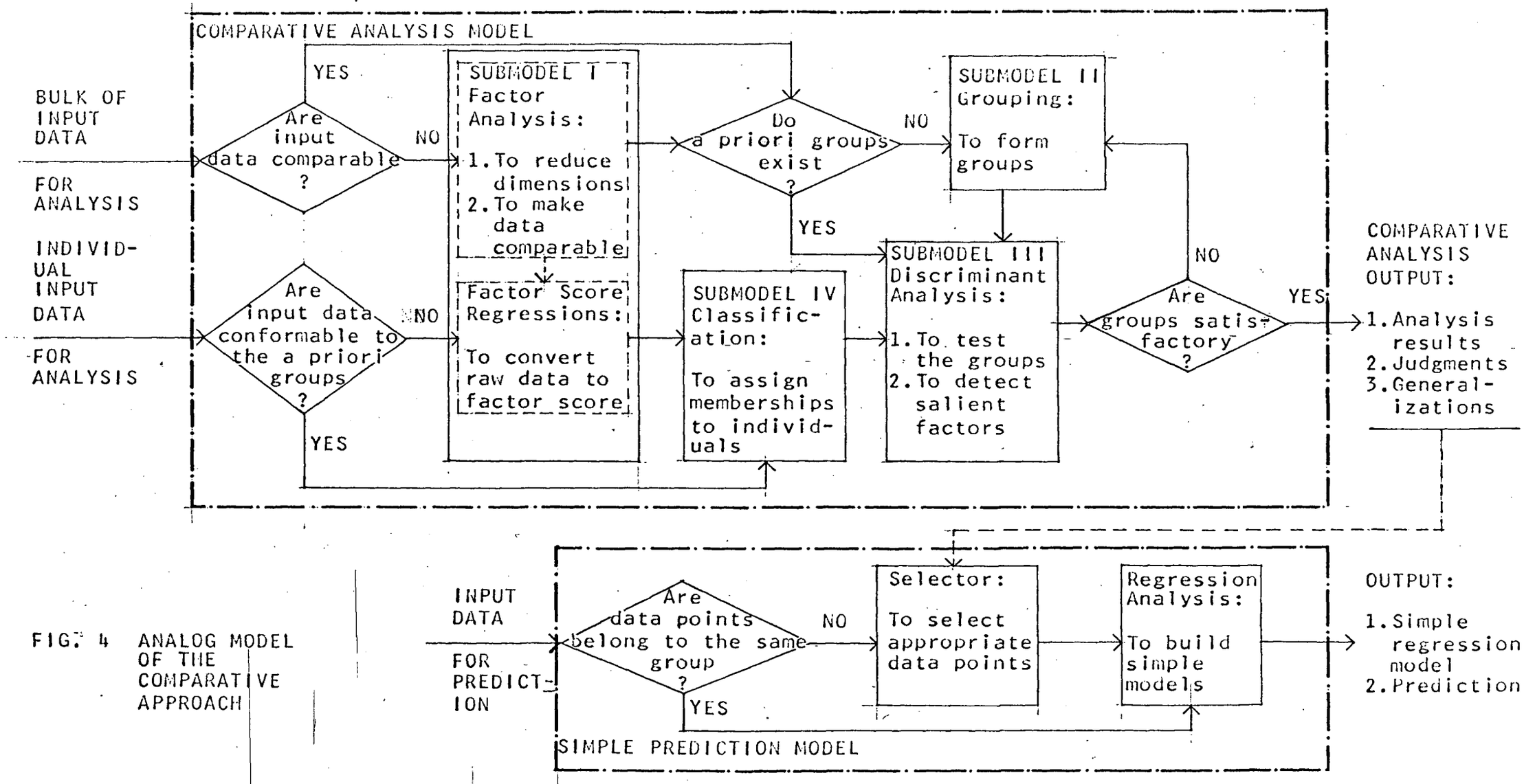


FIG. 4 ANALOG MODEL OF THE COMPARATIVE APPROACH

will be attained after those groups have successfully been discriminated. If, on the other hand, the data are not comparable, factor analysis submodel will serve as a means to standardize and orthogonalize the data, and to provide factor score regression equations to convert the raw data of a single subject for classification. In case the subjects have not been grouped, factor scores are put in to the grouping submodel and the resultant groupings are tested by the discriminant submodel.

Furthermore, a new subject can also be put in to this model in order to assign it an appropriate membership by means of classification submodel and its affiliations. (Refer to figure 4).

2. Simple Prediction Model

The best statistical technique for deriving this model is simple regression. A suitable independent variable (predictor) which should obviously correlate to the dependent variable (criterion) is chosen at first. The data points (observations) should be so selected that they are all members of the same group derived from the Comparative Analysis Model mentioned above.

There is no direct connection between the Comparative Analysis Model and the Simple Prediction Model. However, the former, which is responsible to most of the specification errors, does provide useful information for the latter, which is responsible to all the measurement errors and some of the specification

errors, to select the appropriate data to build a more efficient prediction model.

Finally, the applicability of the above methodology will be tested in the following way. An usual simple regression model is built to predict the CBD floor space from the population of North American cities. Since some of the factors influencing the CBD growth pattern have not been considered, and some of the variances of the criterion remain unexplained, comparative models are employed to overcome these weaknesses and to derive an "improved" simple regression model. If the original model has been improved, the general hypothesis stated in chapter 2 is verified.

3.4 Empirical Study--The Prediction of CBD Floor Space and the Alternative Hypothesis

Different approaches to study and predict the CBD growth patterns have been suggested by some researchers. Weiss, in her excellent research paper, summarizes that the methodological approaches for estimating CBD space requirements can be divided into five groups, namely, population, purchasing power, business establishments, employment, and daytime population.[16] She also remarks that "a review of recent CBD forecasts reveals that population as a basis for projection is still the most widely used approach".[17] In addition, researchers as well as

---

[16]Shirley F. Weiss, The Central Business District in Transition, Methodological Approaches to CBD Analysis and Forecasting Future Space Requirements, City and Regional Planning Studies, Research Paper No. 1 (Chapel Hill: University of North Carolina, 1957).

[17]Ibid., p. 27.

planning agencies have employed regression techniques for the CBD studies and forecasting. For example, Berry correlated CBD variables in terms of simple regression models,[18] Rotoff studies the relationship between population and business floor areas,[19] and the staff of the Technical Planning Board predicted the retail and office areas for future Downtown Vancouver by comparing those of some American cities.[20]

In comparative analysis, CBD data must be collected from individual cities that have been chosen for study. However, it is found that "so far no uniform method of delimiting the district has been used, [and] that for each city the limits of the CBD have been largely a matter of local agreement".[21] In order to lessen the discrepancy of delimiting CBD's, the study has to be narrowed to that of the floor spaces which are functionally and desirably located at the CBD's. Therefore, the term "CBD floor space" is defined here as the following:

> CBD floor space means the gross floor space used for central business functions inclusively. It is obtained "by subtracting, from total floor space in the CBD, the floor space devoted to non-central

---

[18]For detail, see Brian J. L. Berry, "The Retail Component of the Urban Model", Journal of the American Institute of Planners, May, 1965, pp. 150-155.

[19]For detail, see B. M. Rotoff, "The CBD and Its Umland", Plan Canada, The Town Planning Institute of Canada, vol. 10, no. 2, 1969, pp. 16-23.

[20]For detail, see Technical Planning Board, Downtown Vancouver 1955-1976, City of Vancouver Twenty-Year Development Plan, Aug. 1956, pp. 32-35.

[21]Raymond E. Murphy and J. E. Vance, Jr., "Delimiting the CBD" Economic Geography, vol. 30, July 1954, pp. 189-222.

business uses: residence, public and organizational functions, industry, wholesaling and vacancy.[22]

Thus, a simple regression equation is derived for forty-five North American cities,[23] to predict the CBD floor space from the total population of the incorporated cities.[24] This equation is:

$$\log Y = -1.5772 + 1.0279 \log X$$

where Y = CBD floor space in mil. sq. ft.

X = incorporated city population in 1,000 persons

and $R = 0.90$

$R^2 = 0.81$

S.E. of $Y = 0.24$

Although the coefficient of correlation is so high that might be considered as a desirable model for prediction, it is still dissatisfactory because of the following reasons:

1. Due to the sampling error, the true coefficient of correlation of the population is 99% sure to fall within the range of 0.95 and

---

[22]Raymond E. Murphy, and J. E. Vance, Jr., "A Comparative Study of Nine Central Business Districts", Economic Geography, Oct. 1954, pp.

[23]A sampling of 100 North American cities was draw, questionnaires were sent in order to collect appropriate CBD data. However, by the time of this analysis, only 45 data points were applicable. For detail, see Appendices A.1 and A.2.

[24]The size of CBD depends closely upon the incorporated city population. For detail, see R. E. Murphy, and J. E. Vance, Jr., op. cit., pp. 324-326.

0.77, approximately.[25] In other words, the true correlation between population and the CBD floor space is probably as low as 0.77, and correspondingly only 59% of the total variances has been accounted for.

2. Many influencing factors have not been considered.

3. Refer to the scatter diagram in Figure 5, in a certain population range, some of the data points exhibit tremendous differences of CBD floor space. These differences can never be explained by this model.

Therefore, the original simple model will be revised indirectly by means of the Comparative Approach.

An alternative hypothesis is thus formulated:

> Cities can be classified into groups with similar demographic, social, economic, and geographic characteristics as a basis for better prediction of CBD floor space.

It is asserted that if the alternative hypothesis is verified, so is the general hypothesis.

3.5 Research Design

A flow diagram designed for the remaining research is shown in Figure 6.

---

[25] See Table XIII in Appendix A.6.

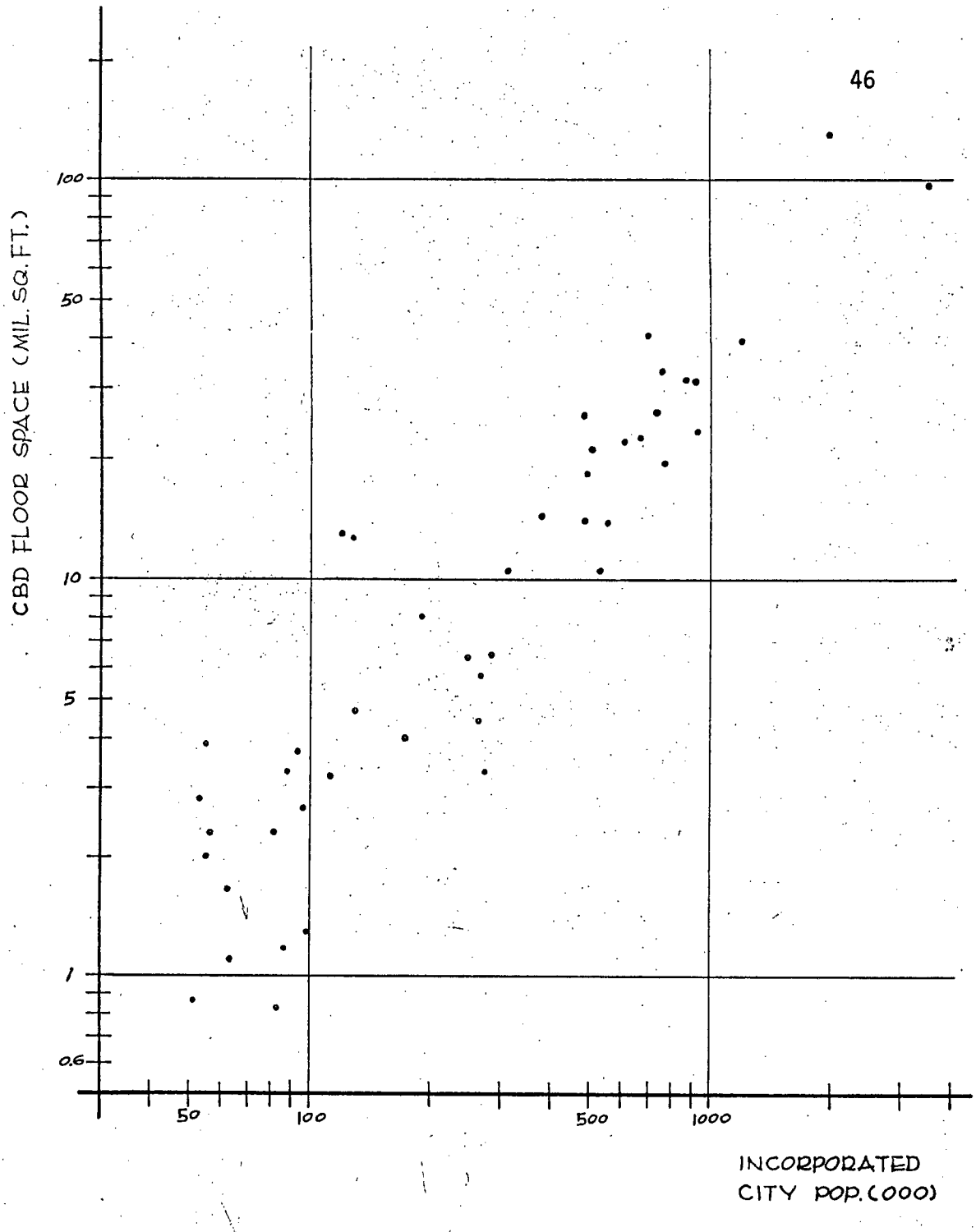


FIG. 5 SCATTER DIAGRAM OF 45 NORTH AMERICAN CITIES

FOR ACTUAL FIGURES SEE APPENDIX A.2

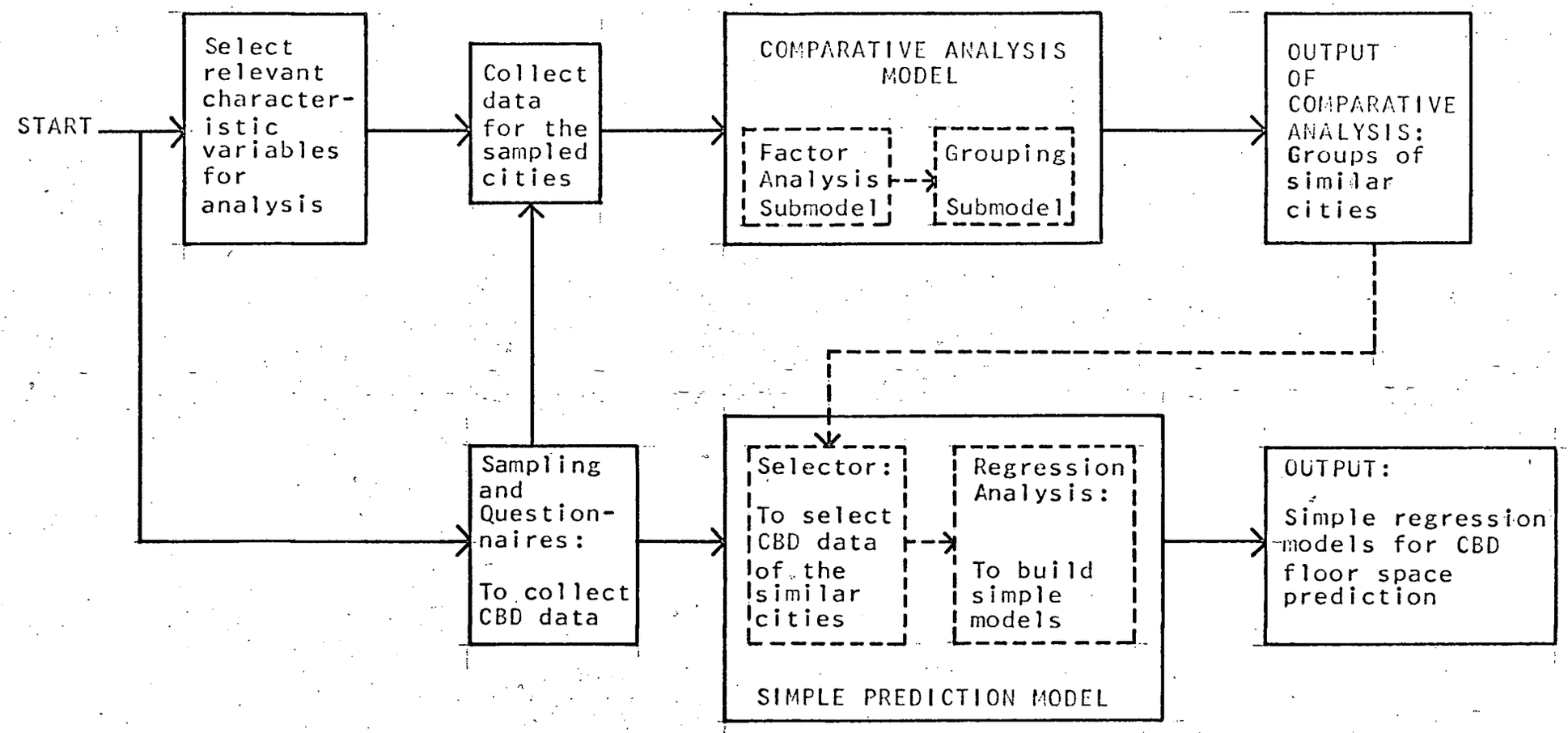


FIG. 6 FLOW DIAGRAM OF THE RESEARCH DESIGN

Similar to the Analog Model presented in section 3.3 above, the process will begin with selecting relevant variables and collecting appropriate data for the chosen cities. The data are put in to the Comparative Analysis Model which consists of a factor analysis submodel and a grouping submodel modified with variance test. The output which will contain several groupings of cities with similar characteristics serves as a reference for selecting appropriate data points to build a simple regression model.

3.6 Summary

The concepts of similarity and comparability constitute the theoretical framework of the proposed Comparative Approach to planning analysis and prediction. For the purposes of getting insights into urban phenomena and for prediction, methodology in terms of an Analog Model is presented, which composes of a Comparative Analysis Model and a Simple Prediction Model parallel to and independent of each other. In order to test the general hypothesis that the comparative approach is effective for planning analysis as well as for prediction, the following study aims at devising a better simple regression model to predict the CBD floor space. The results will be described in the following chapter.

CHAPTER IV

APPLICATION OF THE COMPARATIVE APPROACH--A SIMPLE MODEL FOR CBD FLOOR SPACE PREDICTION

CHAPTER IV

APPLICATION OF THE COMPARATIVE APPROACH--A SIMPLE MODEL FOR CBD FLOOR SPACE PREDICTION

4.1 Characteristic Variables

A sample of 100 North American cities with population of 50,000 and above has been drawn. (Refer to Appendix A.1) The next step of the study is to select relevant characteristics or variables influencing the amount of CBD floor space in a city. The criteria are:

1. the selected variables should be representative of the characteristics related to CBD growth,
2. the data of these variables must be comparable across cities.

Thus, five groups of characteristics, consisting of a total of 30 variables, are selected as follows:

A. Population characteristics

1. Total incorporated city population (TICP)
2. Percentage of population aged under 5 (PA5-)
3. Percentage of population aged over 65 (PA65+)
4. City population density (CPD)
5. Percentage of population non-white (PNW)
6. Percentage of population change from 1950-1960 (PC)

B. Socio-economic characteristics

7. Percentage of occupied dwellings owner occupied (HOWNER)
8. Median value of owner occupied dwelling (MVOOD)
9. Number of persons per household (HDSIZE)
10. Percentage of occupied dwellings with one car (D1CAR)
11. Percentage of occupied dwellings with two or more cars (D2+CAR)
12. Percentage of occupied dwellings with home food freezer (DHFF)
13. Percentage of occupied dwellings with T.V. (DTV)
14. Median family income (MFI)
15. Percentage of families of income level under $3,000 (FI3T-)
16. Percentage of families of income level over $10,000 (FI10T+)

C. Business Characteristics

17. Retail sales per capita (RSPC)
18. Wholesale trade per capita (WTPC)
19. All service receipts per capita (SRPC)

D. Labour force characteristics

20. Percentage of population employed labour force (EMPT)
21. Percentage of employed labour force in retailing and wholesaling (LF-RW)
22. Percentage of employed labour force in white collar occupations[1] (LF-WC)

---

[1]White collar occupations include professionals, managers, officials, clerical and sales workers.

23. Percentage of employed labour force in manufacturing (LF-M)
24. Percentage of employed labour force in communication and transportation (LF-CT)
25. Percentage of employed labour force in construction (LF-C)

E. Geographical characteristics

26. Member of a CMA[2] or SMSA[3] (MEMBER)
27. Central city of a CMA of SMSA (CCITY)
28. Located on railroad network (RAIL)
29. Having port facilities (PORT)
30. Served by commercial airline (COMAIR)

It has been mentioned above that the CBD floor space depends very closely on the incorporated city population. Therefore, the variables of total population and population change are very crucial in this study. Moreover, proportions of infant and aged populations influence the consumption patterns; population density affects accessibility to CBD; non-white population affects the purchasing power as a result of different cultural background and living habit.

---

[2]CMA, Central Metropolitan area, means groups of urban communities in Canada which are in close economic, geographic and social relationship.

[3]SMSA, Standard Metropolitan Statistical area. By definition, every city of over 50,000 in U.S. is included in an SMSA.

The socio-economic characteristics directly or indirectly influence purchasing power of the population in a city. Home-ownership and income level reflect the economic condition; possession of cars, home food freezer and television indicates the living standard; household size reinforces the estimation of the purchasing power of the household unit.

Business characteristics affect directly the amount of business spaces. Further, the sale volumes of retail, wholesale and service in the economy partly reflect its economic affluence and partly indicate the level of consumption of the local population.

Labour force characteristics depict the relative importance of various economic activities in terms of distributions of employments, thus reveal the functional diversification of the city and influence the CBD land use patterns.

Finally, geographical characteristics indicate the locational importance of a city within the region. Whether it is the central city of a CMA or a SMSA dictates whether it has strong functional influence over the surroundings, therefore affects the space requirements of its CBD. Similarly, the availabilities of rails, port facilities and commercial airlines indicate the prosperity of central business activities.

Since the data for these characteristic variables have to be collected from various census sources of the United States and Canada, discrepancies are inevitable. In order to make data comparable across

the 100 sampled cities, most of them have been converted to either percentages or ratios. In addition, the qualitative variables, such as the geographical characteristics, are measured in two discrete numbers 1 and 0 denoting YES and NO respectively.

4.2 Factor Analysis of the Characteristics of 100 Sampled Cities

A detailed statistical interpretation of factor analysis can be found in Appendix A.3.

After analyzing the variance-covariance structure of the above 30 variables, 11 factors are extracted to account for approximately 85% of the original variances. These factors are described in detail below:[4]

1. Factor I is referred to as Socio-economic Index (SEIND)

   This factor positively correlates with such variables as value of owner-occupied dwellings, households with two or more cars, median family income, family income more than $10,000, retail and service sales per capita, and employment in white-collar occupations. On the other hand, it negatively correlates with such variables as family income less than $3,000 and central city of a CMA or SMSA. Therefore cities with high positive scores on Factor I denote a high socio-economic condition of their citizens.

---

[4]Note that the mean and standard deviation of the scores for any factor are equal to 0 and 1 respectively. The following discussions concerning correlations between factors and variables are only good for those with absolute values greater than 0.3. For details see Table X in Appendix A.3.

2. Factor II is referred to as Business Status Index (BSIND)

   Since this factor negatively correlates with retail sales, wholesale trade and service receipts per capita, and also with employments in retail and wholesale, a city having negative score on the factor shows that it is of above average business status.

3. Factor III is referred to as Living Standard Index (LSIND)

   It positively correlates with such variables as home-ownership, value of dwelling, households with two or more cars, households with food freezer, employments in retail and wholesale, and member of a CMA or SMSA. It negatively correlates with city density. Thus, high positive score means high standard of living in the city.

4. Factor IV is referred to as Household Size Index (HSIND)

   This factor positively correlates with such variables as population aged 5 or less, population change, household size, and employments in construction. It negatively correlates with population aged 65 and over, member and/or central city of a CMA or SMSA. A positive score on this factor indicates that the city is above average in household size.

5. Factor V is referred to as Functional Diversification Index (FDIND)

   This factor positively correlates with the variables of employment distributions in retail and wholesale, white collar

occupation, communication and transportation, and construction, as well as with some geographical variables such as central city of a CMA or SMSA, and availability of commercial airlines. It negatively correlates with employment in manufacturing.

This factor provides an indicator of functional diversification of a city. High positive score means that the relative importances of retailing, wholesaling, manufacturing and other functions are likely to be equal. As this score approaches to zero, it shows that the economy in question relies more on some particular activities, among which manufacturing is the most probably one.[5] Further, a high negative score on this factor shows that manufacturing dominates the economy.[6]

6. Factor VI is referred to as Population Index (PIND)

This factor positively loads on such variables as population, density, and commercial airlines, but negatively loads on retail sales per capita. A high positive score represents a large population.

Moreover, that it negatively correlates with retail sales per capita reveals that the larger the city in terms of population

---

[5]Manufacturing is the only variable that has high negative loading on this factor.

[6]For example, Detroit, a predominantly manufacturing city, has score of -1.11 on this factor.

size, the smaller the purchasing power of the citizens living there.

7. Factor VII is referred to as Population Change Index (PCIND)

   This factor only has negative correlations with such variables as population change, employment level, labour force in construction, and commercial airlines. Thus, high negative score on this factor means that the population change of the city during the last decade was considerable.

8. Factors VIII, IX, X are referred to as Geographical Factors (GF-I, II, & III)

   These factors associate with geographical characteristics of the cities, but they are not easily defined.

9. Factor XI is referred to as Poverty Index (POVIND)

   The meaning of this factor is obscure. Since it negatively correlates with such variables as non-white population, income less than $3,000, and member of a CMA or SMSA, but positively correlates with car-ownership, it may be interpreted as a measure of the poverty of the city. The higher the score, the lesser the poverty.

Structures of the first six factors are shown in Figure 7. With these 11 orthogonalized and standardized factor scores as criteria for

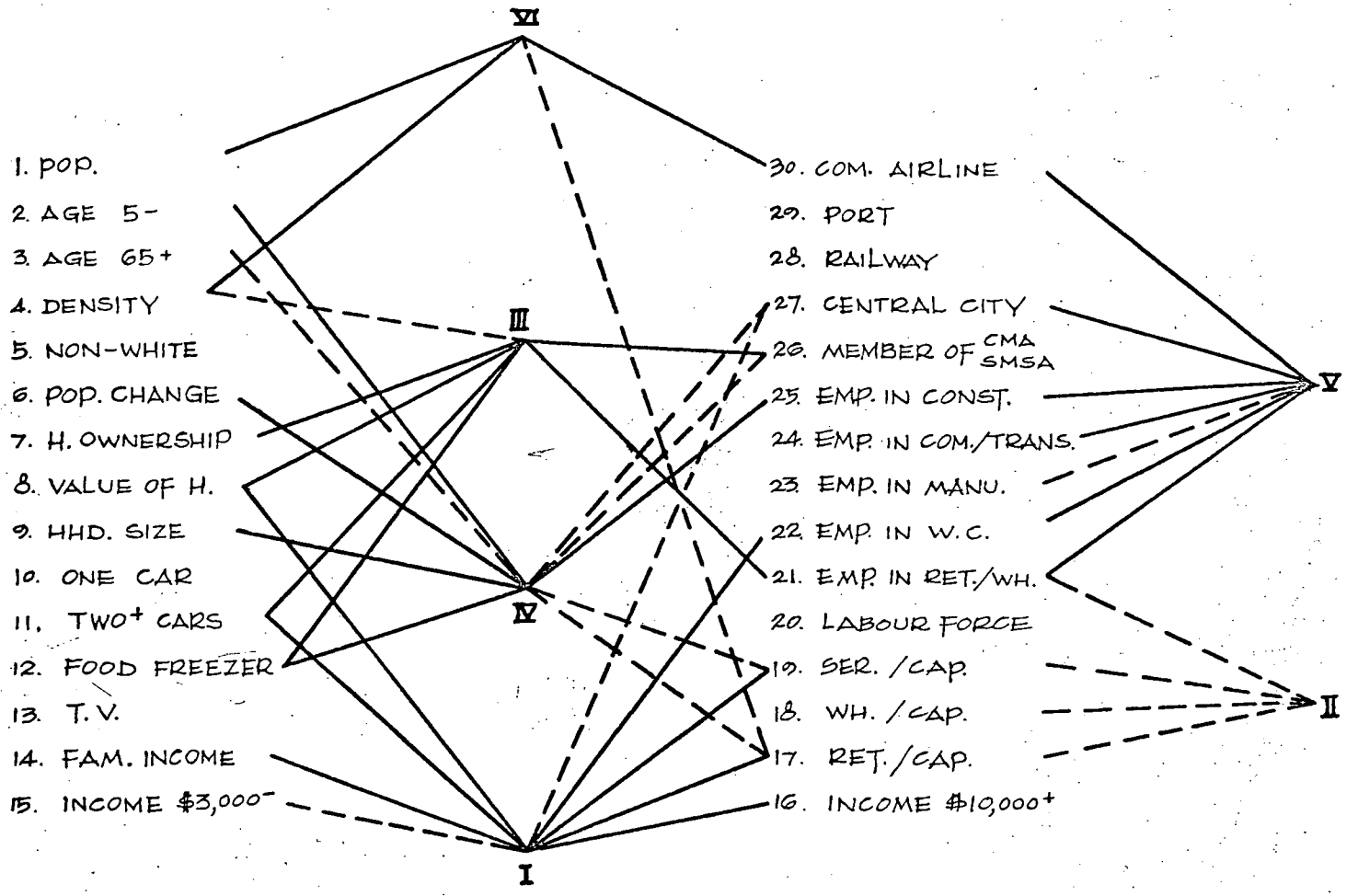


FIG. 7 STRUCTURE OF FACTORS I TO VI % TRACE: 69

——— +IVE CORRELATION OF 0.30+

- - - - - -IVE CORRELATION OF 0.30+

I SOCIO-ECONOMIC INDEX

II BUSINESS STATUS INDEX

III LIVING STANDARD INDEX

IV HOUSEHOLD SIZE INDEX

V FUNCTIONAL DIVERSIFICATION INDEX

VI POPULATION INDEX

comparison, the sampled cities are grouped, and the results are described in the following section.

4.3 Grouping of Cities with Maximum Similarities

In the factor space, the similarities of a pair of cities can be measured in terms of distance-scaling.[7] Groupings can thus be formed based on the criterion that the similarity between the two cities is maximum or that the characteristic homogeneity of the group is optimal. In this study, the former is employed because of the following reasons:

1. The purpose of grouping in this study is not to classify several homogeneous groups for analyzing their characteristics, but to select some similar cities for improving the original prediction model.

2. CBD data are available for 45 cities only. It is recognized that these data will not be sufficient for a number of homogeneous groups of nearly equal size to undertake regression analysis at the same time.[8]

3. Grouping based on the criterion of maximum similarity provides a large group for which regression analysis can be carried out with

---

[7] B. J. L. Berry, "Grouping and Regionalization: An Approach to the Problem of Using Multivariate Analysis", in W. L. Garrison and D. F. Marble (eds.) Quantitative Geography (Northwestern University Studies in Geography, No. 13), pp. 219-251, or see Appendix A.4.

[8] Discussion and comparison of different grouping techniques can be found in Appendix A.4.

this limited stock of CBD data.

A computer programme is written and run. One principal group of maximum similarity with 36 city members is obtained. The characteristics of this group are summarized in Table I.

With reference to Table I, the characteristics of this group represented by its centroid indicate that the 36 cities as a whole are very likely of average quality. Among the 11 factors taken for comparison, none of them is significantly different from its grand mean that is equal to zero, although they do exhibit some degrees of difference. These cities did not have rapid change in population in the last decade; they are seemingly middle-size cities; their economic structures emphasize slightly on particular kinds of activities, for example, manufacturing, so that their functional diversification indices are comparatively low; the average household size of these cities is fairly small while the overall living standard is above average. Finally, it is perhaps difficult to generalize that their business status and socio-economic condition are below average, but more than 60% of them do have low scores on these factors.

In this group of 36 cities, only 17 of them have been supplied with adequate CBD data for building a revised regression model, they are listed in Table II below:

TABLE I

CHARACTERISTICS OF A GROUP OF 36 SIMILAR CITIES

Factors	Centroid	Max. Value	Min. Value	No. & % of +ive Values	No. & % of -ive Values
I. SEIND	-0.166	1.34	-1.33	14 (39%)	22 (61%)
II. BSIND*	0.089	1.16	-1.71	24 (67%)	12 (33%)
III. LSIND	0.027	1.17	-1.29	22 (61%)	14 (39%)
IV. HSIND	-0.613	0.34	-1.40	6 (17%)	30 (83%)
V. FDIND	-0.186	1.50	-1.24	12 (33%)	24 (67%)
VI. PIND	0.200	1.97	-0.87	21 (58%)	15 (42%)
VII. PCIND*	0.244	0.85	-0.56	27 (75%)	9 (25%)
VIII. GF-I	0.126	0.85	-0.35	22 (61%)	14 (39%)
IX. GF-II	0.231	0.95	-0.67	26 (72%)	10 (28%)
X. GF-III	0.021	1.01	-1.39	21 (58%)	15 (42%)
XI. POVIND	0.035	1.07	-1.32	20 (56%)	16 (44%)

*Negative score denotes above average. Note that the grand mean for all factors are equal to zero.

TABLE II

17 CITIES OF SIMILAR CHARACTERISTICS

No.	Code	Name	No.	Code	Name
1	104	Dubuqe, Iowa	10	610	Cleveland, Ohio
2	208	Springfield, Mo.	11	614	St. Louis, Mo.
3	310	Chattanooga, Tenn.	12	615	Milwaukee, Wis.
4	405	Springfield, Mass.	13	618	Pittsburgh, Pa.
5	508	Minneapolis, Minn.	14	619	Seattle, Wash.
6	510	Denver, Colo.	15	620	Buffalo, N.Y.
7	605	Cincinnati, Ohio	16	621	Philadelphia, Pa.
8	608	Boston, Mass.	17	814	Vancouver, B.C.
9	609	Baltimore, Md.			

Before undertaking regression analysis, it is interesting to compare this group of 17 similar cities with the original group of 45 so that the effort of "re-grouping" or "selecting" on an arbitrary group can be revealed. The comparison is shown in Table III below:

TABLE III

COMPARISON OF STANDARDIZED* FACTOR SCORES
BEFORE (N=45) AND AFTER (N=17) GROUPING

Factors	Number & Percentage of cities with Positive Scores N=45	N=17	Number & Percentage of cities with Negative Scores N=45	N=17
I. SEIND	18 (40%)	7 (41%)	27 (60%)	10 (59%)
II. BSIND	25 (56%)	7 (41%)	20 (44%)	10 (59%)
III. LSIND	21 (47%)	9 (53%)	24 (53%)	8 (47%)
IV. HSIND	22 (49%)	3 (18%)**	23 (51%)	14 (82%)**
V. FDIND	21 (47%)	3 (18%)**	24 (53%)	14 (82%)**
VI. PIND	22 (49%)	13 (76%)**	23 (51%)	4 (24%)**
VII. PCIND	23 (51%)	13 (76%)**	22 (49%)	4 (24%)**

Factors VIII to XI are omitted

*Factor scores are standardized on the basis of 45 cities

**Significant at 0.05 level of two-tail test

Table III reveals that the process of "re-grouping" changes the characteristic structure of the group dramatically. For the first three factors, although the changes of positive-negative proportions are not significant, yet they have inversed the relative importances. For the last four factors taken into consideration, the changes are significant

at 0.05 level.[9]

Further, the scores on selected factors and variables of the 45 cities are standardized and plotted on graphs so as to compare the distributions of cities before and after grouping (see Figures 8, 9, 10, 11, and 12). It is concluded that, by deleting those dissimilar cities, the grouping process has actually changed the characteristic structure of the original arbitrary group. A comparison of these graphs also indicates that the CBD floor space correlates highly with the incorporated city population; this correlation will be increased if dissimilar cities are removed.

4.4 Revised Simple Regression Model

The last step of this study is to derive another simple regression model from these 17 similar cities and test whether this model is in fact better than the original one.

This revised model takes the form of:

$$\log Y = -1.7507 + 1.1191 \log X$$

where Y = CBD floor space in mil. sq. ft.

X = Incorporated city population in 1,000 persons

and $R = 0.96$

$R^2 = 0.92$

S.E. of $Y = 0.13$

---

[9]The probability of occurring these changes are less than or equal to 5%.

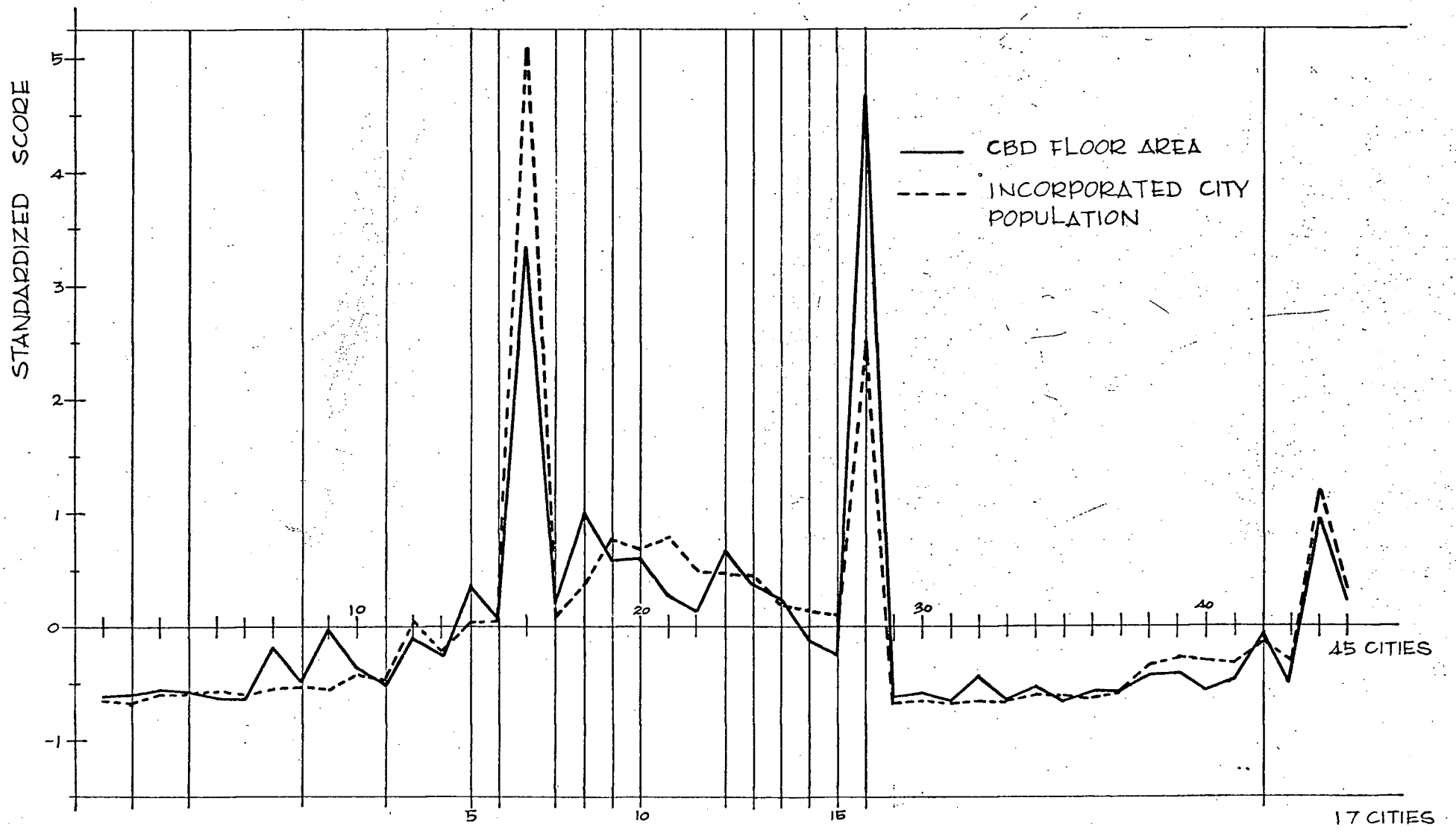


FIG. 8 BEFORE - AFTER COMPARISON

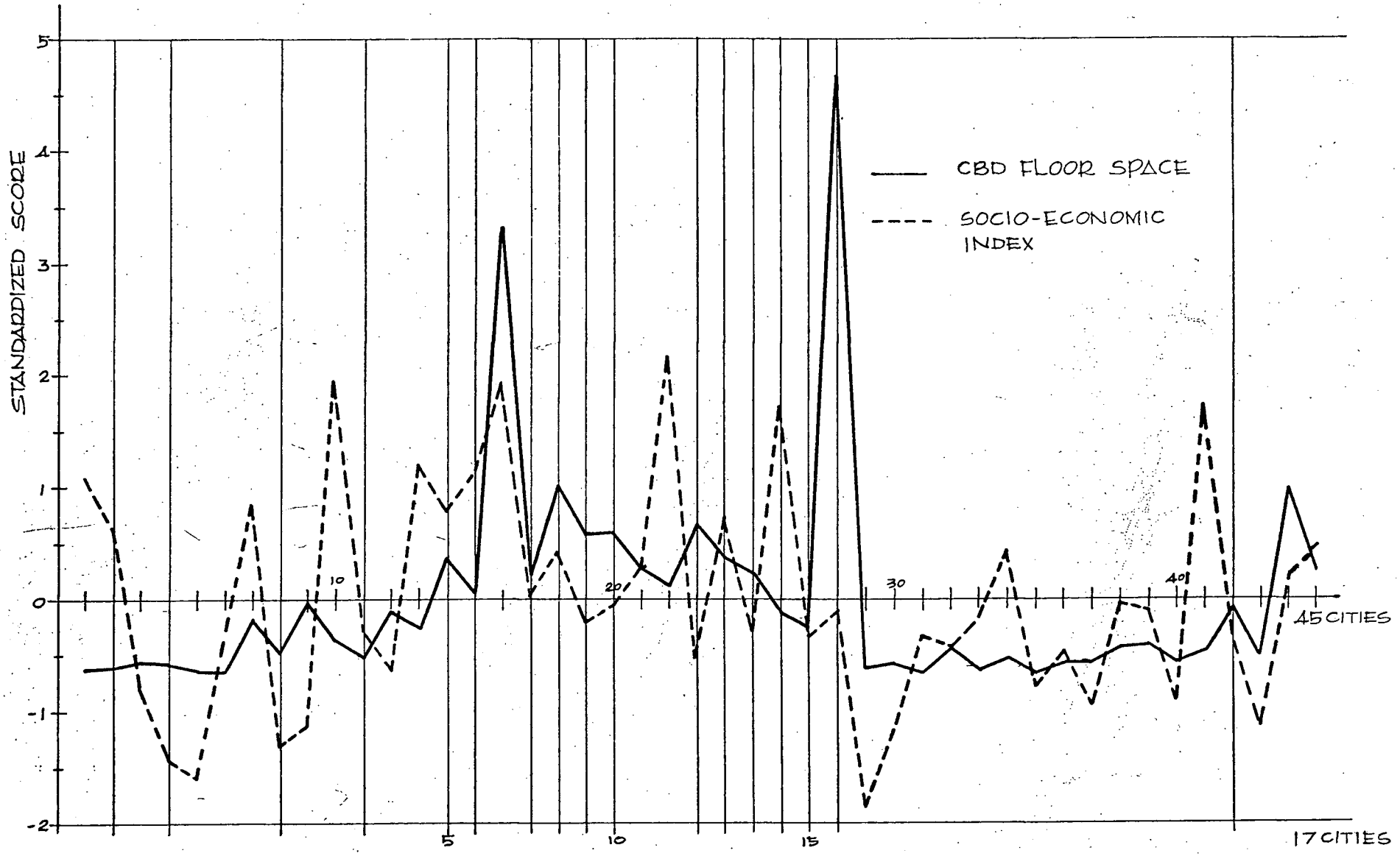


FIG. 9 BEFORE-AFTER COMPARISON

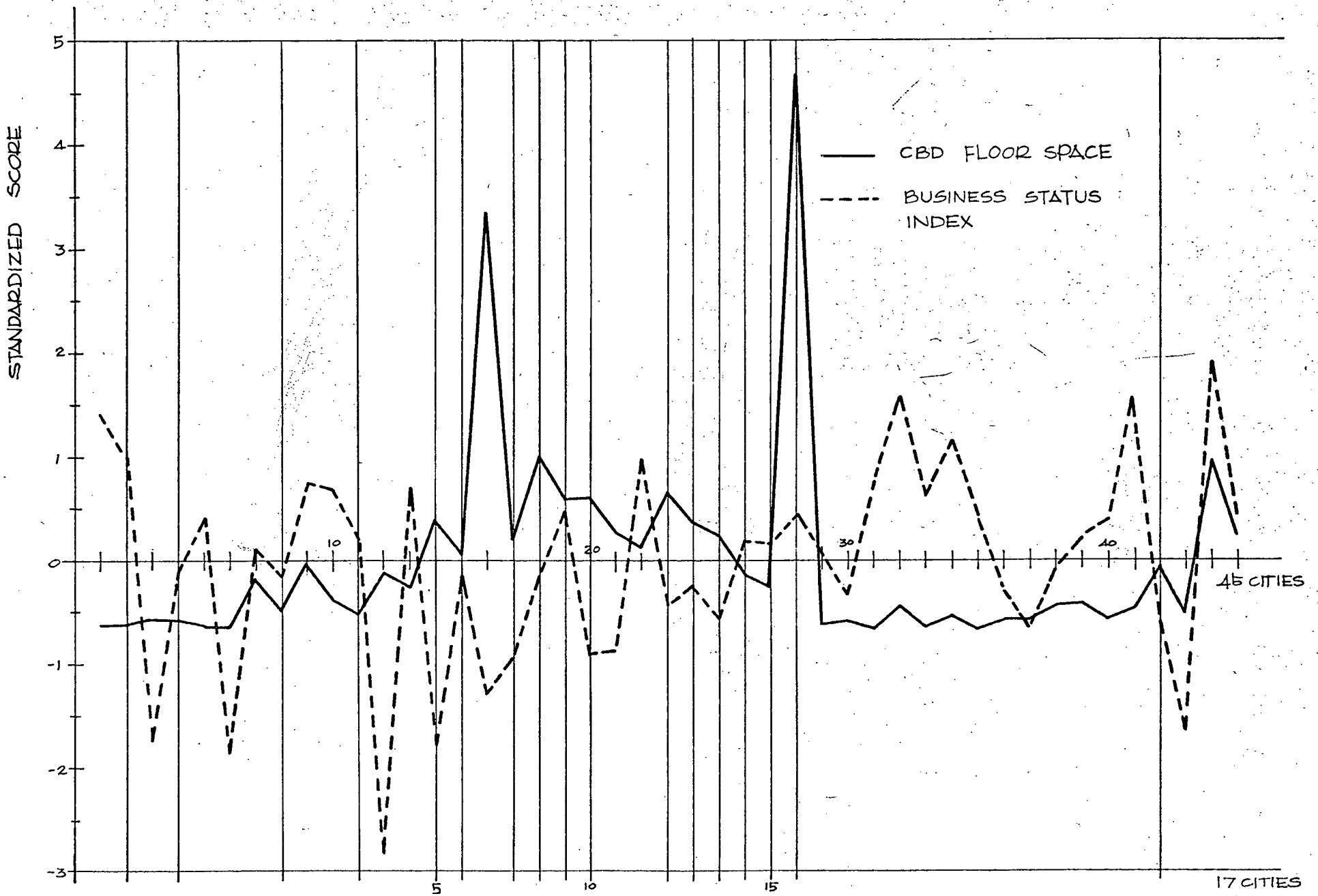


FIG. 10 BEFORE - AFTER COMPARISON

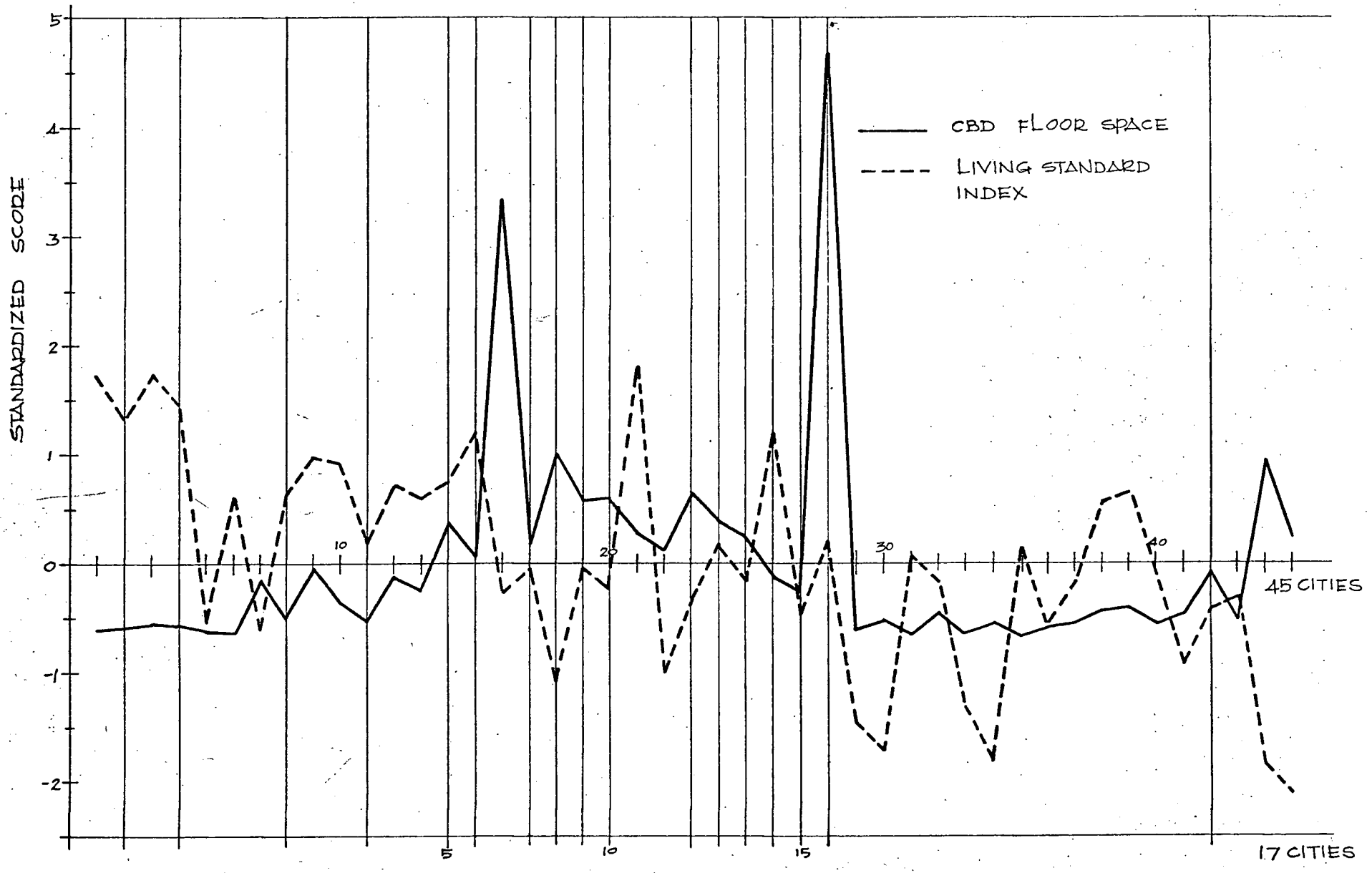


FIG. 11 BEFORE - AFTER COMPARISON

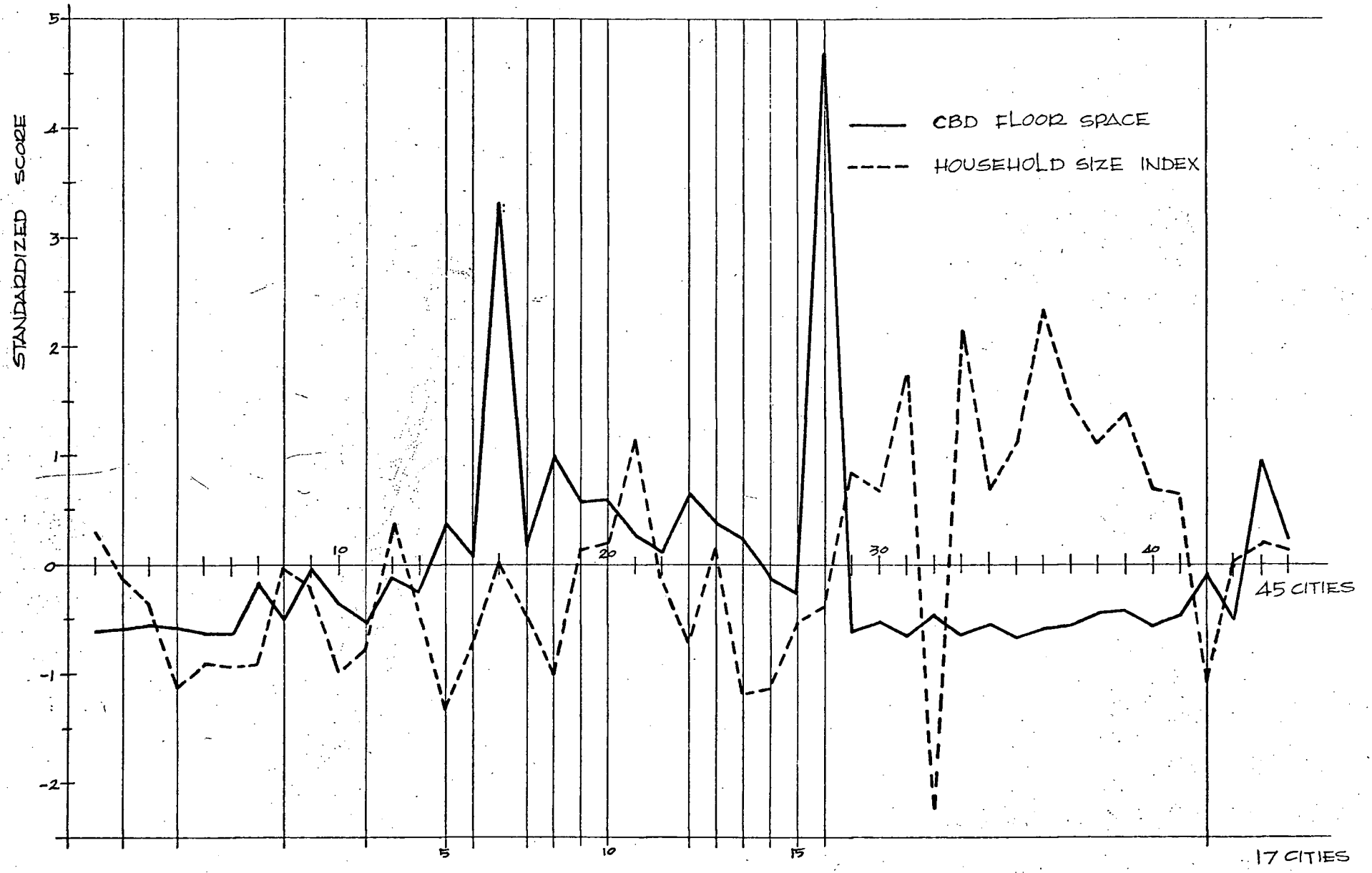


FIG. 12 BEFORE - AFTER COMPARISON

The two regression models are compared below:

TABLE IV

COMPARISON OF THE TWO REGRESSION MODELS

General equation: $\log Y = a + b \log X$

	R	R^2	S.E. of Y
1. Original equation a = -1.5772, b = 1.0279, N = 45	0.90	0.81	0.24
2. Revised equation a = -1.7507, b = 1.1191, N = 17	0.96	0.92	0.13
Difference	0.06	0.11	0.11
Percentage improvement	6.67%	13.6%	45.8%

From the above table, it can be seen that the standard error of prediction and the coefficient of determination of the regression model has been improved substantially by means of comparative analysis. The improvement, or the difference, of coefficient of correlation, however, is not as obvious, therefore additional statistical test is required.

Fisher's z-transformation is applied.[10] It is found that the change of coefficient of correlation is not significant at the conventional 0.05 level for a one-tail test. Notwithstanding this, the probability of obtaining as large or larger difference is 0.059, about 6 times out of 100 a Type I error is committed if the null hypothesis

[10]For details, see Appendix A.5.

that there is no difference between the two correlations is rejected. Therefore, the judgement is suspended.[11] The resolution can be attained by either undertaking replication of the experiment or judging it from other considerations.

Finally, it is asserted that the revised model does exhibit improvements. The following arguments support this conclusion:

1. Since the coefficient of correlation of the original regression is so high (0.90) that it is not possible to gain significant improvement. In other words, the revised equation with such coefficient of correlation (0.96) has probably reached the maximum level that any regression model based on empirical data can attain.

2. Comparing the two equations, it is found that the percentage changes of the coefficient of determination and the standard error of estimates are 13.6% and 45.8% respectively. These improvements can be considered significant.

3. Two graphs showing the relationships of observed and estimated

---

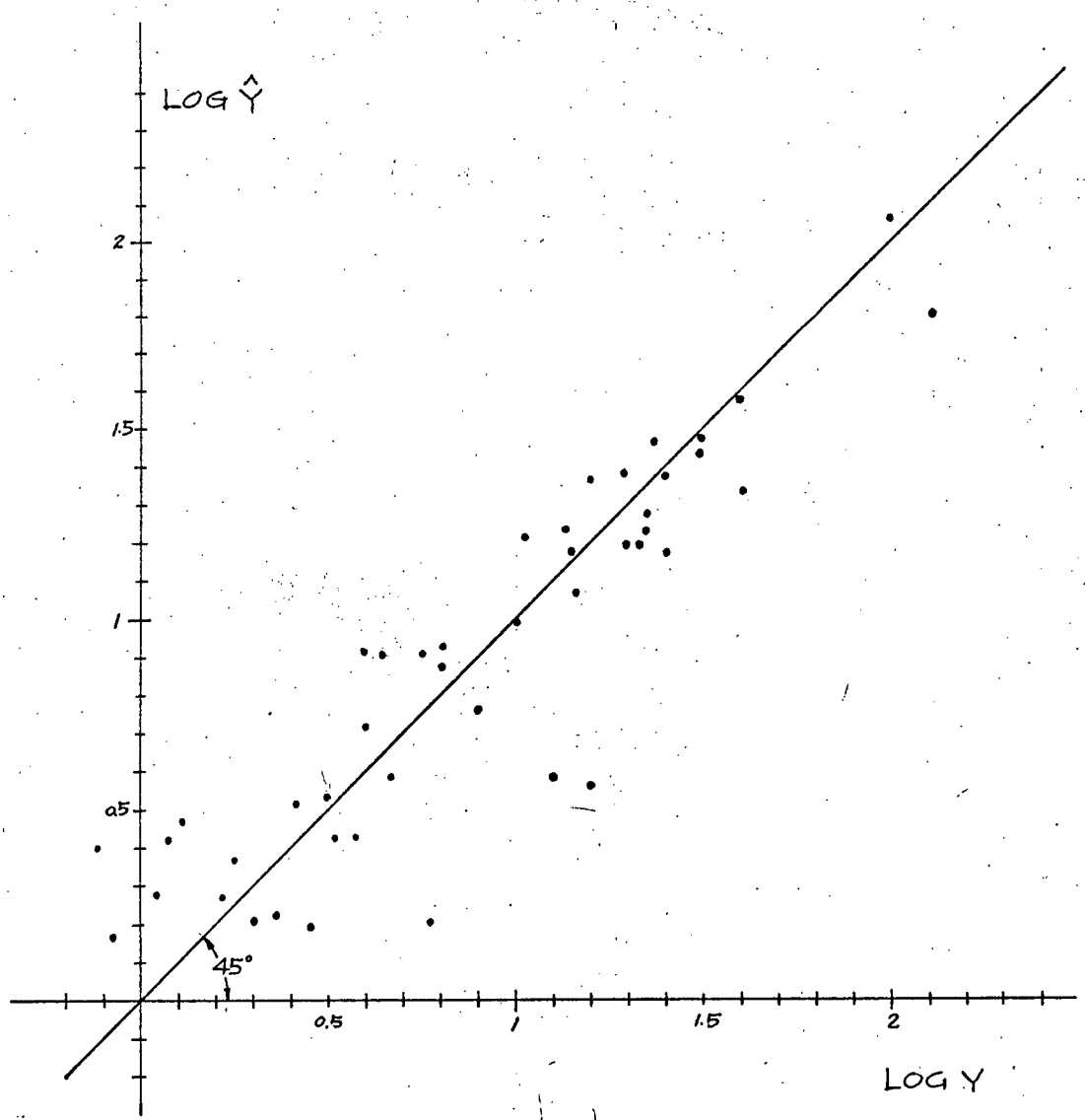
[11] "Instead of confining ourselves to a two-choice decision--rejection or acceptance--we might allow a third possibility, that of suspended judgement, which usually calls for a replication of the experiment. For example, if the deviation is significant at the 0.01 level or better we might reject H_0; if the deviation is smaller than the boundary of the critical region at the 0.10 level, we might accept H_0. Between the two levels, 0.10 and 0.01, we might suspend judgement". See Guilford, J. P., Fundamental Statistics in Psychology and Education, (New York: McGraw-Hill, Inc., 1965), p. 207.

values for the two regression equations are constructed for investigation. The scattering of points in Figure 13 indicates that quite a few predictions are not satisfactory due to their locations away from the 45-degree line, a line representing perfect prediction. Some estimations by the original model yield errors as high as or even higher than 0.55 mil. sq. ft. Figure 14 is a similar graph for the revised model. The points are located very close to the perfect prediction line. In fact, the greatest error of estimation is 0.28 mil. sq. ft., that is, the precision of estimation has been doubled.

Consequently, it is concluded that the simple model for CBD floor space prediction has been improved after the dissimilar cities have been removed.

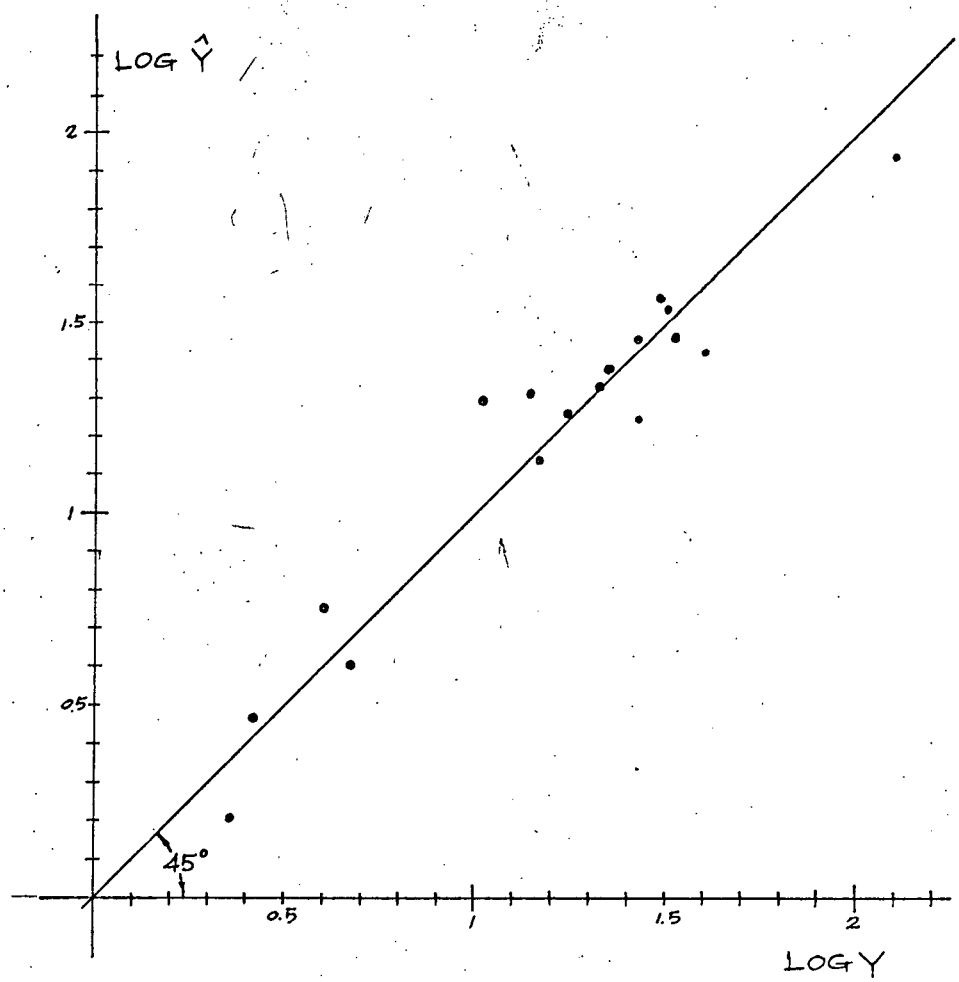
4.5 Modification of the Simple Model and the Concept of Maximum and Minimum Requirements

The above model predicts the possible trends of growth of CBD floor space for all cities provided that they are the members of the same group. Classify this group of cities into several sub-groups according to their population ranks, and single out the maximum and minimum CBD floor space for each sub-group as in the following table, two additional regression models can be derived.



Y CBD FLOOR SPACE
Ŷ ESTIMATED Y FROM LOG Y = -1.5772 + 1.0279 LOG X

FIG. 13 RESIDUAL OF PREDICTION BEFORE GROUPING 45 CITIES



Y CBD FLOOR SPACE

Ŷ ESTIMATED Y FROM $\log Y = -1.7507 + 1.1191 \log X$

FIG. 14 RESIDUAL OF PREDICTION AFTER GROUPING 17 CITIES

TABLE V

MAXIMUM AND MINIMUM CBD FLOOR SPACES FOR DIFFERENT POPULATION RANKS

Population (1,000 per.)	Maximum (mil. sq. ft.)	Minimum (mil. sq. ft.)
50- 80	2.291	-
80-150	4.730	2.635
150-250	-	4.000
250-350	-	-
350-450	14.600	-
450-550	26.000	10.498
550-650	-	13.910
650-750	41.000	26.000
750-850	-	-
850-950	-	31.000
950+	129.000	-

A regression equation, which exhausts all the maximum values of CBD floor space for different population ranks, and which gives the upper limit of a prediction based on the past experiences of many other similar cities, is called a maximum requirement regression. It takes the form of:

$$\log Y = -1.7507 + 1.1191 \log X$$

where $R^2 = 0.99$, and S.E. of $Y = 0.07$

Similarily, a minimum requirement regression can also be defined. The equation is:

$$\log Y = -1.7832 + 1.0807 \log X$$

where $R^2 = 0.98$, and S.E. of Y = 0.10

The three equations are shown in Figure 15. It is interesting to note that as population increases, so does CBD floor space, and that the range of upper and lower limits of prediction also increases. In other words, a large city not only can maintain a much-greater-than-normal CBD but also can be serviced with merely a much-smaller-than-normal CBD. The requirement flexibility for a larger city is greater than that for a smaller city.

Explanation of the slow increase of minimum requirements may raise another proposition of diminishing space requirements for increasing population. Take a simple example, the space requirement for two persons is always lesser than the doubling of that for one. On the other hand, the fact that large city probably is also central city with substantial influences over the surroundings explains the fast increase of maximum requirements against city size.

In practice, these three regression equations perform different functions, and together they generate a total picture for CBD floor space predictions. After forecasting the future level of population, the most probably requirement can be projected, accompanied by corresponding maximum and minimum requirements reflecting the impact of changing development policy. In forecasting process by using these models, it is not necessary to forecast the characteristic variables beforehand.

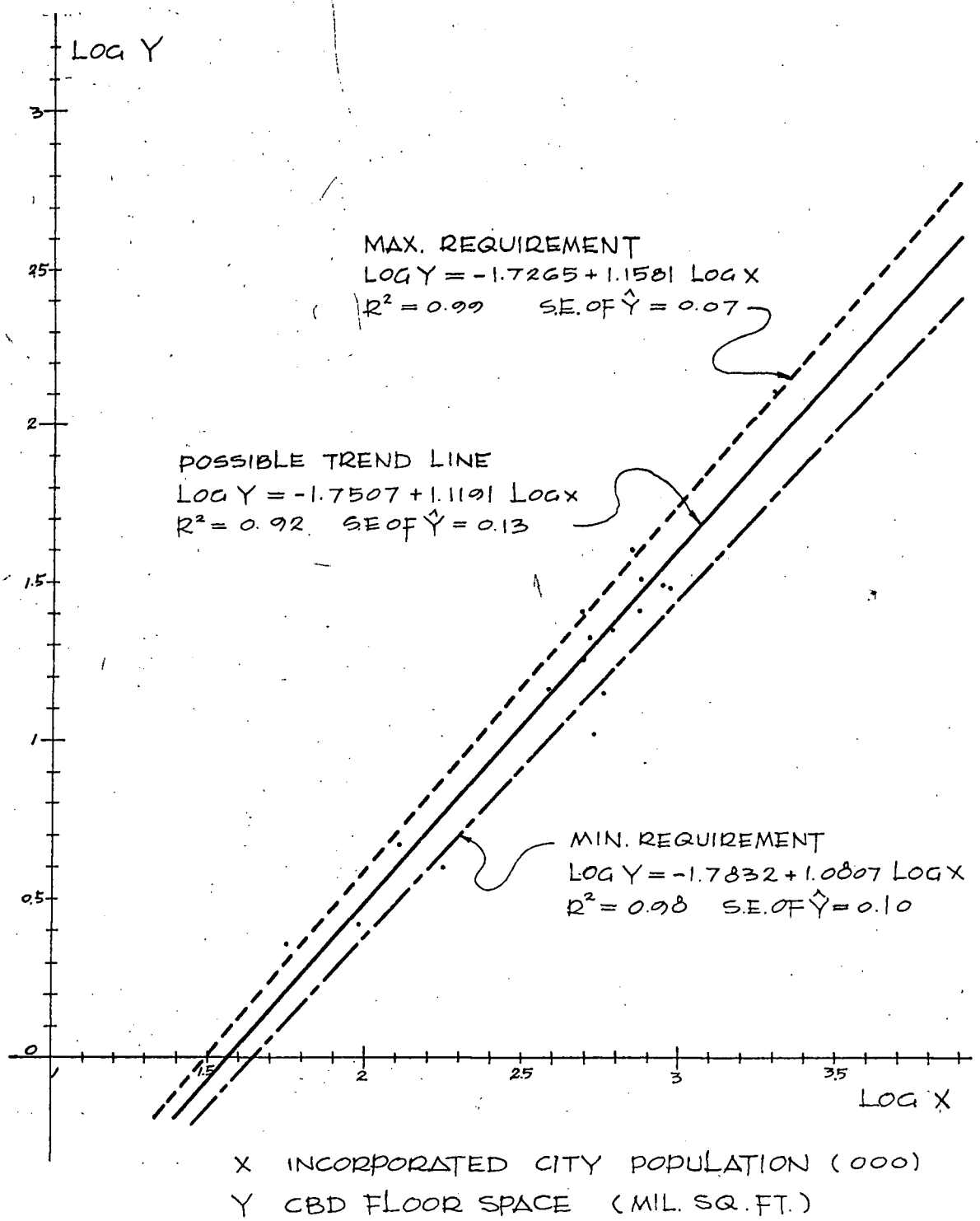


FIG. 15 REVISED REGRESSION MODEL FOR CBD FLOOR SPACE PREDICTION

4.6 Summary

An improved simple regression model for CBD floor space prediction has been derived indirectly by means of comparative analysis. The procedure comprises of factor analysis which makes characteristic data comparable, and grouping technique which provides information to select appropriate data points among the original observations. The whole idea stems from the hypothesis that cities can be classified into groups with similar characteristics as a basis for better prediction.

In addition, among the similar cities maximum and minimum requirement regression models can also be derived to provide upper and lower limits of a prediction.

CHAPTER V

VERIFICATION AND IMPLICATIONS OF THE COMPARATIVE APPROACH--A CONCLUSION

CHAPTER V

VERIFICATION AND IMPLICATIONS OF THE COMPARATIVE APPROACH--A CONCLUSION

5.1 Verification of the Hypotheses

It has been verified that the simple regression model derived in Chapter IV provides improved predictions over the original one. This improvement has the following implications:

1. The precision of prediction has been improved. The measurement errors of the model remain unchanged since it is still in its originally simple form. Therefore, the increase of precision entirely attributes to the decrease of specification errors.

2. The large variations of CBD floor spaces that can not be explained by the original model can now be identified as the consequences of dissimilar characteristics among cities. These variations automatically decrease as the dissimilar cities are excluded from consideration, and the remaining variations are those within the group which exist as the composite result of slightly and acceptably different characteristics. For illustration, see Figure 16 below. The great variation between A and B can not be explained by ordinary simple model. However, if groups I and II are formed, it is clear that difference between A and B is the result of significant dissimilarity between them, and that the small

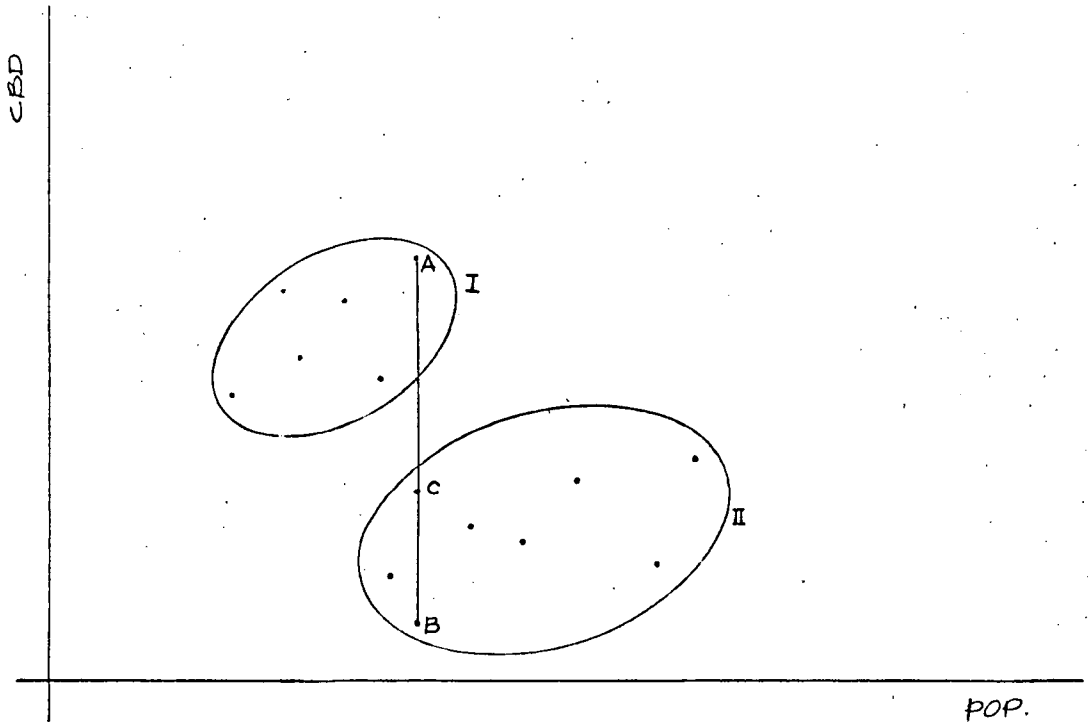


FIG. 16 DEVIATIONS BETWEEN AND WITHIN GROUPS (HYPOTHETICAL)

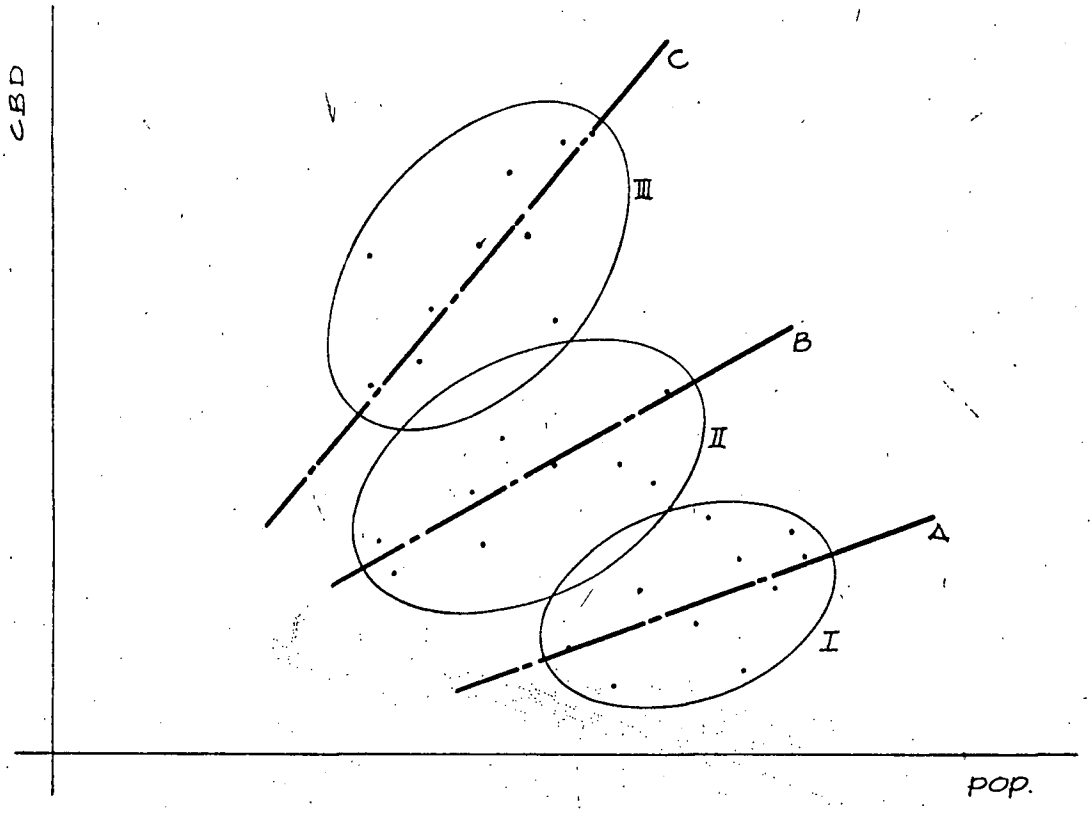


FIG. 17 HIERARCHICAL GROWTH PATTERN (HYPOTHETICAL)

difference between B and C is acceptable as a deviation within the group.

3. The improved simple model can be used for prediction with considerable precision for all cities that are members of the same group from which the model is derived. Thus, the model serves as a "law" governing a particular growth pattern among the group members. It is therefore transferable within the group.

4. If several groups are formed at the same time and appropriate simple prediction models are constructed for each of these groups, the combination of these simple models provides a hierarchical growth pattern and depicts the impact of significant differences of characteristics among groups. In Figure 17 above, for example, simple regression models A, B and C represent the CBD growth patterns for distinct groups I, II and III respectively. In terms of growth rates, a hierarchical pattern can be derived, that is, from A to B to C. Model A predicts CBD floor space from population assuming that the characteristics[1] of the city remain unchanged or with only a slight change. If it is foreseen that the characteristics of a city will change significantly in the future, its growth pattern will move from A to B or C according to its likelihood of being a member of group II or III.[2] This

---

[1]Note that these characteristics are mostly measured in terms of ratios or indices, for the convenience of comparison.

[2]This can be done by comparing the characteristics of all the groups, or by applying classification techniques.

hierarchical growth pattern permits one to forecast the future trends upon dynamic changes of characteristics.

Considering the above implications, it is clear that the result of Comparative Analysis makes possible a simple model to predict better in four ways: it re-captures the ignored causalities indirectly; it explains the large variations among observations; it makes the model effectively transferable; and it provides for dynamic changes. Hence, the alternative hypothesis stating that cities can be classified into groups with similar characteristics as a basis for better prediction, and the general hypothesis stating that Comparative Approach is an effective way to analyze urban problems are justified.

5.2 Applications and Limitations of the Comparative Approach

The applicabilities of the Comparative Approach are twofold: for analysis and for prediction.

1. For analysis

If the interest is in exploring the differences and their causal factors of certain criteria among several subjects, the application of the Comparative Approach is beneficial. Several areas of study that can make use of this approach are suggested in the following:

a) In urban renewal program, it is possible to characterize some areas, by means of factor analysis and discriminant analysis,

so as to range priorities of areas for slum clearance.

b) In traffic accident study, intersections and segments of roadways can be grouped according to frequency of accidents occurred, discriminant analysis will disclose the important factors influencing the amount of traffic accidents.

c) In traffic modal choice study, individuals can be grouped according to their preference of transportation modes for a certain purpose, say journey to work. Socio-economic as well as modal characteristics can be analyzed among these groups so that important factors that determine choices can be separated.

d) As for the recreation study of a region, several sub-regions of different recreational potentiality can be formed for further studies. The procedure is firstly to select and compare the physical features and relevant characteristics of small areas, and secondly to group the nearest neighbours which not only similar in stated characteristics but also contiguous to each other.

e) In inter-city utility study, such as water consumption, cities with similar level of water consumption can be grouped. Therefore, relationships with the influencing factors such as population level, socio-economic characteristics, degree of specialization of industry, geographic

features of the city can be generated quantitatively.

f) In land use studies, important characteristics associated with different land use patterns can be detected by means of the Comparative Approach at the inter-city level.

The first four areas of study suggested above focus on the local problems whereas the remainders compare urban phenomena across cities.

2. For prediction

Prediction is a by-product of the Comparative Analysis. The result of the Comparative Analysis Model[3] may or may not be applicable to construct a simple regression model, depending on the grouping structures obtained. The reason is that in order to construct a regression model, at least three members in a group is necessary.[4] Further, for the sake of predicting with better confidence, a lower limit of six members in a group is recommended.[5] Therefore, the capability of building a simple model from the Comparative Approach is a function of (a) the number

---

[3]Refer to Figure 4, Analog Model of the Comparative Approach, in Chapter III.

[4]Sample size of less than three is meaningless for regression analysis because: with one point infinitive number of regression lines can be drawn, and two points a regression of perfect "fit" can be constructed but reveals nothing about the association between dependent and independent variables.

[5]See Appendix A.6.

of subjects in the group, and (b) the availability of appropriate data. It is further realized that:

a) At the local level, to build such prediction models is probably not realistic, if the subjects concerned are sub-areas or zones instead of people in the city, because there is no guarantee to get enough members for each group to build regression models.

b) At the inter-city level, the city in question has this possibility of being the single member of a group after comparison. In such case regression analysis is inapplicable.

For the above reasons, prediction using simple regression model should not be the final objective of the Comparative Analysis, but an additional outcome attainable under favourable circumstances. Therefore it is necessary to emphasize that the optimal objective of the Comparative Approach to planning analysis is to study and clearly identify the relevant factors to lay a foundation for more complete and generalized scientific explanation of the basic forces influencing certain urban phenomena in question. Such superior explanation obtained from considering a number of similar cases of the subject matter would undeniably lead to better prediction by means of simple regression or other techniques. However, the inability of ensuring this analytical result to be further used for prediction limits its practical value.

Despite the conceptual weakness regarding its practical value, a further limitation is data difficulty. Two implications of data problem are considered here.

1. For Comparative Analysis at a local level, data must be collected and aggregated according to the smallest basic unit under consideration, for example, a person, a small group of people, or a small area. This may not be a serious problem if an intra-urban data bank has been intensively developed. However, for the inter-city level, an inter-urban data bank containing important statistics of many other urban areas is a pre-requisite.

2. In the process of Comparative Analysis, the selection of relevant characteristic variables is crucial because it affects the outcome of the analysis considerably. For this reason, usually as many variables are extracted as possible and then are converted into a manageable number of factors. Thus the large number of variables concerned aggravates the data problem.

Fortunately, data processing techniques are fast developing as the result of computerization. Moreover, the awareness of the importance of establishing intensive data bank on the inter-city basis is apparent.[6] It can be foreseen that data difficulty will be substantially alleviated in the near future so that urban researchers can undertake studies more on the comparative basis to gain insights into urban problems.

---

[6]A conference on urban development models held in 1967 came up with a recognition that "desirably, with respect to selected data items, an

5.3 Summary and Conclusion

Comparative Approach to planning analysis is proposed in this thesis. Although such approach has been employed occasionally in planning, a systematic and complete exploration can scarcely be found in the planning literature. The general negligence of this approach is probably attributed to the more sophisticated techniques involved, and the extensive data requirements. The modern computer technique offers a solution to these problems by providing "canned" computer programmes for profound and cumbersome computation, and by enabling the setting up of an intensive inter-urban data bank. Under such circumstance, planning analysis on a comparative basis can be achieved.

The Comparative Approach is introduced in this thesis on the assumption that it is not only one of the effective methods to comprehend in depth the specific urban problem but an indirect means permitting improved predictions through the use of a simple model. Thus the current problems of building and using models in planning are objectively discussed and hypotheses are boldly formulated. The Approach is further presented in three ways. Firstly, the concepts of similarity and comparability are discussed so as to provide a theoretical framework for practical undertaking. Secondly, multivariate statistical techniques applicable to this Approach are presented in non-mathematical manner.

---

effort might be made to achieve uniformity not only over time, but also between cities". B. Harris, "Conference Summary and Recommendation", in Urban Development Models, Special Report 97 (Highway Research Board, 1968), p. 12.

Finally, these techniques are brought together to form a systematic methodology in terms of an Analog Model consisting of a Comparative Analysis Model and a Simple Prediction Model.

The effectiveness of the proposed approach is tested by experimenting a simple model for prediction. After introducing the comparative analysis, if prediction is significantly improved, the validity of this approach to planning prediction as well as analysis can be ascertained.

After a pilot study of predicting CBD floor space, it is found that the Approach is effective for better prediction, as well as for explanation. However, it is also found that, because of the particular relationship[7] between the Comparative Analysis Model and the Simple Prediction Model, the benefit gained from better prediction can not always be attained for all cases. Unless this obstacle is overcome, the Approach for the purpose of prediction is of little practical value. Nevertheless, its application to analysis is almost endless, especially in this computerized society.

Finally, several areas of study that can possibly be undertaken on the comparative basis are suggested for further research. It is hoped that by means of continuous effort in this direction, the weakness of the Approach can be detected and overcome. Hopefully, this initial investigation will pave the way for this methodology to develop from its infancy to maturity.

---

[7]Refer to the Analog Model in Chapter III, and to Section 5.2 above.

BIBLIOGRAPHY

BIBLIOGRAPHY

General References

Ackoff, R. and B. Harris. "Planning, Operations Research, and Metropolitan Systems". Conference Proceedings, American Institute of Planners (1964), pp. 92-96.

Ahmad, Q. Indian Cities: Characteristics and Correlates, Research Paper No. 102. Department of Geography. Chicago: Univ. of Chicago Press, 1965.

Allpass, J. "Urban Structure and the Composition of City Centers". Ekistics (Feb., 1964), pp. 86-95.

Alonso, W. "Predicting Best with Imperfect Data". Journal of the American Institute of Planners (July, 1968), pp. 248-255.

Bartholomew, H. Land Uses in American Cities. Harvard City Planning Studies, XV. Cambridge: Harvard University Press, 1955.

Bachi, R. "Standard Distance Measures and Related Methods for Spatial Analysis". The Regional Science Association, Vol. 10, pp. 83-132.

Berry, B. J. L. and D. F. Marble. Spatial Analysis, a Reader in Statistical Geography. Englewood Cliffs, N. J.: Prentice-Hall, Inc., 1968.

———. "A Critique of Contemporary Planning for Business Centers". Land Economic (Nov., 1959), pp. 306-312.

———. "Grouping and Regionalization: An Approach to the Problem of Using Multivariate Analysis". Quantitative Geography, ed. W. L. Garrison and D. F. Marble. Northwestern University Studies in Geography, No. 13, pp. 219-251.

———. Commercial Structure and Commercial Blight, Research Paper No. 85, Department of Geography. Chicago: University of Chicago Press, 1963.

———. "A Note Concerning Methods of Classification". Annals, Association of American Geographers, Vol. 48 (1958), pp. 300-303.

Berry, B. J. L. "Multivariate Socio-Economic Regionalization: A Pilot Study in Central Canada". Regional Statistical Studies. Edited by T. Rymes, and S. Ostry. Toronto: University of Toronto Press, 1966.

———. "The Retail Component of the Urban Model". Journal of the American Institute of Planners (May, 1965), pp. 150-151.

Blumenfeld, Hans. "Are Land Use Patterns Predictable?" Journal of the American Institute of Planners (May, 1959), pp. 61-66.

Bogue, D. J. and D. L. Harris. Comparative Population and Urban Research Via Multiple Regression and Covariance Analysis. Scripps Foundation for Research in Population Problems, No. 8. Miami: Miami University Press, 1965.

Bouchard, R. J. and C. E. Pyers. "Use of Gravity Model for Describing Urban Travel: An Analysis and Critique". Highway Research Record, No. 88, 1965, pp. 1-43.

Bourne, L. S. "A Spatial Allocation Land Use Conversion Model of Urban Growth". Journal of Regional Science, Vol. 9, No. 2 (1969).

———. "Forecasting Land Occupancy - Changes Through Markevian Probability Matrices, A Central City Example. Research Report No. 14, Urban Development Study, Univ. of Toronto, 1967.

Center for Real Estate and Urban Economics. Jobs, People and Land: Bay Area Simulation Study. Special Report No. 6. Berkeley: University of California Press, 1968.

Chan, S. L. Multicollinearity in Transportation Models. Forthcoming Master's Thesis, School of Community and Regional Planning, University of British Columbia, May, 1970.

Chapin, F. S., Jr. Urban Land Use Planning. Second edition. Urbana: University of Illinois Press, 1965.

———. "A Model for Simulating Residential Development". Journal of the American Institute of Planners (May, 1956), pp. 120-125.

Clark, P. J. and F. C. Evens. "Distance to Nearest Neighbor as a Measure of Apatial Relationships in Populations". Ecology, Vol. 35 (1954), pp. 445-453.

Cohen, H. "Planners Need an Improved Census". Journal of the American Institute of Planners (Dec., 1955), pp. 33-42.

Cowan, P.; J. Ireland, and D. Fine. "Approach to Urban Model-Building". Regional Studies, Vol. 1 (Dec., 1967), pp. 163-172.

Dacey, M. F. "A Note on the Derivation of Nearest-Neighbor Distances". Journal of Regional Science, Vol. 2 (1960), pp. 81-87.

Dakin, J. "Models and Computers in Planning". Plan Canada, Vol. 6, No. 1 (1965), pp. 11-35.

Danto, A. (ed.) Philosophy of Science. Meridian Books. New York: The World Publishing Co., 1960.

Dominion Bureau of Statistics. The 10th Census of Canada, 1961. Ottawa: Queen's Printer, 1963.

Donnelly, T. C.; F. S. Chapin, Jr., and S. F. Weiss. A Probabilistic Model for Residential Growth. An Urban Studies Research Monograph. Chapel Hill: University of North Carolina, 1964.

Duncan, O. D. and A. J. Reiss, Jr. Social Characteristics of Urban and Rural Communities, 1950. New York: John Wiley and Sons, Inc., 1956.

Forrester, J. W. Urban Dynamics. Cambridge, Mass: The MIT Press, 1969.

Fricker, U. J. Regional Land Use Allocation Models and their Application to Planning. Unpublished Master's Thesis, School of Community and Regional Planning, University of British Columbia, May, 1969.

Gans, Herbert. People and Plans, Essays on Urban Problems and Solutions. New York: Basic Books, Inc., 1968.

Goddard, J. "Multivariate Analysis of Office Location Patterns In the City Center: A London Example". Regional Studies, Vol. 2 (1968), pp. 69-85.

Goode, W. J. and P. K. Hatt. Methods in Social Research. New York: McGraw-Hill Book Co., 1952.

Goodman, W. J. and E. C. Freund (ed.). Principles and Practice of Urban Planning. Municipal Management Series. International City Managers' Association, 1968.

Harris, B. "The Use of Theory in the Simulation of Urban Phenomena". Journal of the American Institute of Planners (Sept., 1966), pp. 258-272.

———. "Computer and Urban Planning". Socio-Economic Planning Science, Vol. 1 (1968), pp. 223-230.

———. "Urban Development Models: New Tools for Planning". Journal of the American Institute of Planners (May, 1965), pp. 90-95.

Harris, B. "Plan or Projection: An Examination of the Use of Models in Planning". Journal of the American Institute of Planners (Nov., 1960), pp. 265-272.

———. Linear Programming and the Projection of Land Uses. Penn-Jersey Transportation Study, PJ Paper No. 20, Nov. 1962.

Hemmens, G. C. "Planning Agency Experience with Urban Development Models and Data Processing". Journal of the American Institute of Planners (Sept., 1968), pp. 323-327.

Irwin, N. A. "Review of Existing Land-Use Forecasting Techniques". Highway Research Record, No. 88 (1965), pp. 182-216.

King, L. J. Statistical Analysis in Geography. Englewood Cliffs, N. J.: Prentice-Hall, Inc., 1969.

———. "Cross-Sectional Analysis of Canadian Urban Dimensions: 1951 and 1961". Canadian Geographer, Vol. 10 (1966), pp. 205-224.

Lowery, I. S. "A Short Course in Model Design". Journal of the American Institute of Planners (May, 1965), pp. 158-166.

———. A Model of Metropolis. Memorandum RM-4035-RC. Santa Monica, Calif.: The RAND Corp., 1964.

Miernyk, W. H. The Elements of Input-Output Analysis. New York: Random House, 1965.

Moser, C. A. and W. Scott. British Towns: A Statistical Study of their Social and Economic Differences. London: Oliver and Boyd Ltd., 1961.

Murphy, R. E. and J. E. Vance, Jr. "Delimiting the CBD". Economic Geography (July, 1954), pp. 189-222.

——— —— ———. "A Comparative Study of Nine CBD's". Economic Geography (Oct., 1954), pp. 301-336.

Nelson, H. J. "A Service Classification of American Cities". Economic Geography, Vol. 31 (1955), pp. 189-210.

Price, D. O. "Factor Analysis in the Study of Metropolitan Centers". Social Force, Vol. 20 (1942), pp. 449-455.

Rannells, J. "Approach to Analysis". Journal of the American Institute of Planners, Special Issue in CBD Studies (Feb. 1961), pp. 17-25.

Reynolds, R. B. "Retail Specialization of CBD's". Journal of the American Institute of Planners (Nov., 1960), pp. 313-316.

Rotoff, B. M. "The CBD and It's Umland, A Study of the Relationship between Population and Business Floor Areas in Manitoba Service Centers". Plan Canada, Vol. 10, No. 2 (1969), pp. 16-23.

Schnore, L. F. and H. Fagin (ed.). Urban Research and Police Planning. Urban Affairs Annual Reviews, Vol. 1. Beverly Hills, Calif.: Sage Publication, Inc., 1967.

Smith, R. H. T. "Method and Purpose in Functional Town Classification". Annals, American Association of Geographers, Vol. 55 (1965), pp. 539-548.

Smith, R. L. "Gravity Model Theory Applied to a Small City Using a Small Sample of Origin-Destination Data". Highway Research Record, No. 88 (1965), pp. 85-115.

Stone, R. "A Comparison of the Economic Structure of Regions Based on the Concept of Distance". Journal of Regional Science, Vol. 2 (1960), pp. 1-20.

Taaffe, E. J.; R. L. Morrill, and P. R. Gould. "Transport Expansion in Underdeveloped Countries: A Comparative Analysis". Geographical Review, Vol. 53 (1963), pp. 503-529.

Tiebout, C. M. The Community Economic Base Study. Supplementary Paper No. 6 Published by the Committee for Economic Development, New York, 1962.

U. S. Bureau of the Census. The 18th Decennial Census of the United States, 1960. Washington, D. C.: Government Printing Office, 1963.

———. County and City Data Book, 1962, A Statistical Abstract Supplement. Washington, D. C.: Government Printing Office, 1962.

Vancouver. Technical Planning Board, Downtown Vancouver 1955-1976. City of Vancouver Twenty-Year Development Plan, Aug. 1956.

Veldman, D. J. Fortran Programming for the Behavior Sciences. New York: Holt, Rinehart and Winston, 1967.

Voorhees, A. M. "The Nature and Use of Models in City Planning". Journal of the American Institute of Planners (May, 1959), pp. 57-60.

———. "Applications of Model Techniques in Metropolitan Planning". Proceedings, the American Institute of Planners (1964), pp. 110-119.

Weiss, S. F. The CBD in Transition: Methodological Approach to CBD Analysis and Forecasting Future Space Requirements. City and Regional Studies, Research Paper No. 1. Chapel Hill: University of North Carolina, 1957.

Statistical and Mathematical Reference

Anderson, T. W. An Introduction to Multivariate Statistical Analysis. New York: John Wiley and Sons, Inc., 1958.

Attneave, F. "Dimensions of Similarity". American Journal of Psychology, Vol. 63 (1950), pp. 516-556.

Bailey, N. T. J. The Elements of Stochastic Process. New York: John Wiley and Sons, Inc., 1964.

Bartos, O. J. Simple Models of Group Behavior. New York: Columbia University Press, 1967.

Casetti, E. "Analysis of Spatial Association by Trigonometric Polynomials". The Canadian Geographer, Vol. 10 (1966), pp. 199-204.

Churchman, C. W.; R. L. Ackoff and E. L. Arnoff. Introduction to Operations Research. New York: John Wiley and Sons, Inc., 1956.

Cooley, W. W. and P. R. Lohnes. Multivariate Procedures for the Behavioral Sciences. New York: John Wiley and Sons, Inc., 1962.

Cox, D. R. "Note on Grouping". Journal of American Statistical Association, Vol. 52 (1957), pp. 543-547.

David, F. D. Tables of the Correlation Coefficient. Cambridge, England: The University Press, 1938.

Davies, O. L. Statistical Methods in Research and Production. London: Oliver and Boyd, 1957.

Davis, P. J. The Mathematics of Matrices, A First Book of Matrix Theory and Linear Algebra. New York: Blaisdell Pub. Co., 1965.

Draper, N. and H. Smith. Applied Regression Analysis, New York: John Wiley and Sons, Inc., 1966.

Fisher, R. A. Statistical Methods for Research Workers. London: Oliver and Boyd, 1954.

Fisher, R. A. "The Use of Multiple Measurements in Taxonomic Problems". Annals, Eugenics, Vol. 7 (1936), pp. 179-188.

Fisher, W. D. "On Grouping for Maximum Homogeneity". Journal of American Statistical Association, Vol. 53 (1958), pp. 789-798.

Freund, J. E. Modern Elementary Statistics. Third edition. Englewood Cliffs, N. J.: Prentice-Hall, Inc., 1967.

Guilford, J. P. Fundamental Statistics in Psychology and Education. Fourth edition. New York: McGraw-Hill, Co., 1965.

Harmon, H. H. Modern Factor Analysis. Chicago: University of Chicago Press, 1960.

Hoggatt, A. C. and F. E. Balderston (ed.). Symposium on Simulation Models: Methodology and Applications to the Behavioral Sciences. South-Western Publishing Co., 1963.

Kemeny, J. G.; J. L. Snell and G. L. Thompson. Introduction to Finite Mathematics. Second edition. Englewood Cliffs, N. J.: Prentice-Hall, Inc., 1966.

Kendall, M. G. "Discrimination and Classification". Multivariate Analysis. Krishnaiah, P. R. (ed.). Proceedings of an International Symposum held in Dayton, Ohio, 1965. New York: Academic Press, 1966, pp. 165-185.

Lawley, D. N. and A. E. Maxwell. Factor Analysis as a Statistical Method. London: Butterworths, 1963.

Loomba, N. P. Linear Programming, An Introductory Analysis. New York: McGraw-Hill, 1964.

MacRae, D. "Direct Factor Analysis of Sociometric Data". Sociometry, Vol. 23 (1960), pp. 360-371.

Maxwell, A. E. "Calculating Maximum Likelihood Factor Loadings". Journal of the Royal Statistical Society, 127, Part 2 (1964), pp. 238-241.

Mood, A. M. and F. A. Graybill. Introduction to the Theory of Statistics. Second edition. New York: McGraw-Hill, 1963.

Moroney, M. J. Facts from Figures. Pelican Books, Harmondsworth: Penguin Books Ltd., 1965.

Rae, C. R. Advanced Statistical Methods in Biometric Research. New York: John Wiley and Sons, Inc., 1952.

Reichman, W. J. Use and Abuse of Statistics. Pelican Books. Harmondsworth: Penguin Books Ltd., 1961.

Rohlf, F. J. and R. R. Sokal. Statistical Tables. San Francisco: W. H. Freeman and Co., 1969.

Rulon, P. J. and D. V. Tiedeman, et al. The Profile Problems: A Methodological Study of the Interpretation of Multiple Test Scores. Cambridge, Mass: Educational Research Corp., 1954.

———. "Distinctions Between Discriminant and Regression Analysis and a Geometric Interpretation of the Discriminant Functions". Harvard Educational Review, Vol. 21 (1951), pp. 80-90.

Siegel, S. Nonparametric Statistics for the Behavioral Sciences. New York: McGraw-Hill, 1956.

Sokal, R. R. and P. H. A. Sneath. Principles of Numerical Taxonomy. San Francisco: W. H. Freeman and Co., 1963.

Stuart, A. Basic Ideas of Scientific Sampling. Griffin's Statistical Monographs and Courses, No. 4. London: Charles Griffin and Co. Ltd. 1968.

Torgerson, W. S. "Multidimensional Scaling of Similarity". Psychometrika, Vol. 30 (1965), pp. 379-393.

Ward, J. H., Jr. "Hierarchical Grouping to Optimize an Objective Function". Journal of American Statistical Association, Vol. 58 (1963), pp. 236-244.

Whittle, P. Prediction and Regulation by Linear Least-Square Methods. Princeton, N. J.: D. Van Nostrand Co. Inc., 1963.

Williams, E. J. Regression Analysis. New York: John Wiley and Sons, Inc., 1959.

Wagner, H. M. Principles of Operations Research, With Applications to Managerial Decisions. Englewood Cliffs, N. J.: Prentice-Hall, Inc., 1969.

APPENDICES

APPENDIX A

STATISTICAL INTERPRETATIONS

A.1 Stratified Random Samples

A.2 Questionnaires

A.3 Factor Analysis

A.4 Distance-scaling of Similarities and Grouping Techniques

A.5 Statistical Test of the Regression Models

A.6 Relationships Between Sample Correlation and Population Correlation Against Sample Size

APPENDIX A

STATISTICAL INTERPRETATIONS

A.1 Stratified Random Samples

The 1960 U.S. Census and the 1961 Canadian Census provide that the numbers of cities having population over 50,000 are 300 and 29 respectively, which total 329. The reason of using 50,000 as starting point for sampling is twofold. Firstly, the CBD of a city with population less than 50,000 is not well formed. Secondly, statistics for some important socio-economic characteristics such as retail sales volume, percentage of dwellings having food freezer, etc., are not available for small cities.

Due to the constraints of time and other resources, all of the 329 cities can not be included in the study so that 100 stratified random samples[1] are drawn.

The advantage of stratified random sampling over simple random sampling is that, when it is used properly, the former increases precision. That is, it decreases sampling error.

---

[1]Stratified random sampling is one of the sampling techniques. The procedure begins with grouping the population into a number of subpopulations (each called a stratum). Then, a simple random sample is drawn from each of these strata, according to its corresponding sampling fraction.

Two rules must be followed in stratified random sampling to maximize precision:[2]

1. Strata should be so constructed that their averages are as different as possible, and their variances within the strata are as small as possible.

2. Large sampling fractions should be used in the more variable strata, or more precisely, the sampling fraction in each stratum should be proportional to the square root of the variance in that stratum.

According to the above rules, and making use of Figure 5, the following strata and sampling fractions are derived:

TABLE VI

STRATIFIED RANDOM SAMPLING

Code	Pop. Range	Strata Size	Prob.	Sample Size
American Cities				
100	50,000- 75,000	120	1/10	12
200	75,000-100,000	60	1/6	10
300	100,000-150,000	49	1/5	10
400	150,000-250,000	30	1/4	8
500	250,000-500,000	30	1/3	10
600	500,000+	21	1	21
Sub-total		300	(23.6%)	71
Canadian Cities				
700	50,000- 75,000	12	1	12
800	75,000+	17	1	17
Sub-total		29	(100%)	29
Grand Total		329	(30.4%)	100

[2]A. Stuart, Basic Ideas of Scientific Sampling, Griffins Statistical Monographs and Courses (London: Charles Griffin & Co. Ltd., 1968), pp. 44-66.

TABLE VII

STRATIFIED RANDOM SAMPLES OF 100 NORTH AMERICAN CITIES

A. UNITED STATES

No.	Code	City	No.	Code	City
		100--50,000-75,000			
1	101	High Point, N.C.	7	107	Ogden, Utah
2	102	Independence, Mo.	8	108	Lima, Ohio
3	103	Joliet, Ill.	9	109	Asheville, N.C.
4	104	Dubuque, Iowa	10	110	Terre Haute, Ind.
5	105	Green Bay, Wis.	11	111	White Plains, N.Y.
6	106	Waukegan, Ill.	12	112	Chester, Pa.
		200--75,000-100,000			
13	201	Binghamton, N.Y.	18	206	Royal Oak, Mich.
14	202	Clifton, N.J.	19	207	Decatur, Ill.
15	203	Sioux City, Iowa	20	208	Springfield, Mo.
16	204	Racine, Wis.	21	209	Fall River, Mass.
17	205	Roanoke, Va.	22	210	Orlando, Fla.
		300--100,000-150,000			
23	301	Tacoma, Wash	28	306	Savannah, Ga.
24	302	Scranton, Pa.	29	307	Albany, N.Y.
25	303	Utica, N.Y.	30	308	Trenton, N.J.
26	304	Cambridge, Mass.	31	309	Kansas City, Kans.
27	305	Evansville, Ind.	32	310	Chattanooga, Tenn.
		400--150,000-250,000			
33	401	Youngstown, Ohio	37	405	Springfield, Mass.
34	402	Sacramento, Calif.	38	406	Worcester, Mass.
35	403	New Haven, Conn.	39	407	San Jose, Calif.
36	404	Albuquerque, N.Mex.	40	408	Charlotte, N.C.
		500--250,000-500,000			
41	501	Portland, Oreg.	46	506	Atlanta, Ga.
42	502	Newark, N.J.	47	507	St. Paul, Minn.
43	503	Miami, Fla.	48	508	Minneapolis, Minn.
44	504	Birmingham, Ala.	49	509	Memphis, Tenn.
45	505	Indianapolis, Ind.	50	510	Denver, Colo.
		600--500,000+			
51	601	Chicago, Ill.	56	606	New Orleans, La.
52	602	San Diego, Calif.	57	607	New York, N.Y.
53	603	San Antonio, Texas	58	608	Boston, Mass.
54	604	Los Angeles, Calif.	59	609	Baltimore, Md.
55	605	Cincinnati, Ohio	60	610	Cleveland, Ohio

(continued)

TABLE VII (Continued)

STRATIFIED RANDOM SAMPLES OF 100 NORTH AMERICAN CITIES

A. UNITED STATES

No.	Code	City	No.	Code	City
		600--500,000+			
61	611	Detroit, Mich.	67	617	Dallas, Texas
62	612	Houston, Texas	68	618	Pittsburgh, Pa.
63	613	Washington, D.C.	69	619	Seattle, Wash.
64	614	St. Louis, Mo.	70	620	Buffalo, N.Y.
65	615	Milwaukee, Wis.	71	621	Philadelphia, Pa.
66	616	San Francisco, Calif.			

B. CANADA

No.	Code	City	No.	Code	City
		700--50,000-75,000			
72	701	Saint John, N.B.	78	707	St-Michel, Que.
73	702	Kingston, Ont.	79	708	Kitchener, Ont.
74	703	Hull, Que.	80	709	Oshawa, Ont.
75	704	Sarnia, Ont.	81	710	Sherbrooke, Que.
76	705	Victoria, B.C.	82	711	Trois-Rivieres, Que.
77	706	Brantford, Ont.	83	712	St. John's, Nfld.
		800--75,000+			
84	801	Halifax, N.S.	93	810	Quebec, Que.
85	802	St. Catharines, Ont.	94	811	Edmonton, Alta.
86	803	Saskatoon, Sask.	95	812	Hamilton, Ont.
87	804	Sudbury, Ont.	96	813	Ottawa, Ont.
88	805	Verdun, Que.	97	814	Vancouver, B.C.
89	806	London, Ont.	98	815	Winnipeg, Man.
90	807	Regina, Sask.	99	816	Montreal, Que.
91	808	Windsor, Ont.	100	817	Toronto, Ont.
92	809	Calgary, Alta.			

A.2 Questionnaires

In order to acquire adequate CBD data for analysis, questionnaires have been sent to the planning departments of the sampled cities. The result is described below:

October 27, 1969

Director

City Planning Department

Dear Sir:

I am a student of the School of Community and Regional Planning, University of British Columbia, Canada, and am doing my master's thesis on developing a technique to predict the future land use of a central business district.

In order to carry out my research project, a bulk of CBD data, which is unfortunately insufficiently provided at the university library, is indispensable. Therefore, it would greatly be appreciated if you could kindly fill in the accompanying short questionnaire and return it to the following address as soon as possible. Without your assistance, hardly can I accomplish my research.

Thank you very much.

Yours very truly,

(Sgnd.) H. Y. Leung

How Yin Leung

Mailing address:

How Yin Leung
c/o School of Community and Regional Planning
University of British Columbia
Vancouver 8, B. C.
Canada

Code _____

A. Please indicate the CBD gross floor area of your city in 1960 according to various land use categories:

Please note: - The breakdown of each land use category can be left blank if such data are not available.
- If 1960 data are not available, those within 1960±3 years are also applicable.

Year _____

	Floor area in sq. ft.
Total Retail-Type Use	
Department stores	
General and variety stores	
Amusement, drinking and eating	
Hotels, motels	
Service station	
Total Office-Type Use	
Finance, real estate, insurance	
General business offices	
Professional services	
Total Wholesaling	
With stock	
Without stock	
Total Industrial	
Manufacturing	
Processing	
Total Transportation	
Off-street parking	
Railroad terminals	
Bus depots	
Total Public and Quasi-Public	
Governmental buildings	
Organizations	
Schools, hospitals, and other institutional use	

(to be cont'd on next page)

	Floor area in sq. ft.
Total Residential (All types)	
Total Other Uses (Please specify, if any)	
GRAND TOTAL	

B. Please describe briefly the method of delimiting CBD, that is currently used in your department:

- Thank You -

TABLE VIII

AGGREGATED CBD DATA OF 45 SAMPLED CITIES

Code	Name	Pop. (X)	log X	CBD F.S. (Y) (mil. sq. ft.)	log Y
102	Independence, Mo.	62.33	1.795	1.685	0.227
104	Dubuque, Iowa*	56.61	1.753	2.291	0.360
203	Sioux City, Iowa	89.16	1.950	3.343	0.524
208	Springfield, Mo.*	95.87	1.982	2.635	0.421
209	Fall River, Mass.	99.94	2.000	1.286	0.109
210	Orlando, Fla.	88.14	1.945	1.169	0.068
307	Albany, N.Y.	129.73	2.113	12.900	1.111
309	Kansas City, Kans.	121.90	2.086	16.000	1.204
310	Chattanooga, Tenn.*	130.01	2.114	4.730	0.675
402	Sacramento, Calif.	191.67	2.283	8.005	0.903
405	Springfield, Mass.*	174.46	2.242	4.000	0.602
506	Atlanta, Ga.	487.46	2.688	14.000	1.146
507	St. Paul, Minn.	313.41	2.496	10.400	1.017
508	Minneapolis, Minn.*	482.87	2.684	26.000	1.415
510	Denver, Colo.*	493.89	2.694	18.200	1.260
601	Chicago, Ill.	3,550.40	3.550	97.000	1.987
605	Cincinnati, Ohio*	502.55	2.701	21.287	1.328
608	Boston, Mass.*	697.20	2.843	41.000	1.613
609	Baltimore, Md.*	939.02	2.973	31.000	1.491
610	Cleveland, Ohio*	876.05	2.943	31.300	1.496
612	Houston, Texas	938.22	2.972	23.200	1.365
613	Washington, D.C.	763.96	2.883	19.700	1.294
614	St. Louis, Mo.*	750.03	2.875	33.000	1.407
615	Milwaukee, Wis.*	741.32	2.870	26.000	1.204
618	Pittsburgh, Pa.*	604.33	2.781	22.587	1.354
619	Seattle, Wash.*	557.09	2.746	13.910	1.143
620	Buffalo, N.Y.*	532.76	2.727	10.498	1.021
621	Philadelphia, Pa.*	2,002.51	3.302	129.000	2.111
701	Saint John, N.B.	55.15	1.742	2.000	0.301
702	Kingston, Ont.	53.53	1.729	2.800	0.447
704	Sarnia, Ont.	50.98	1.707	0.873	-0.059
705	Victoria, B.C.	54.94	1.740	5.888	0.770
712	St. John's, Nfld.	63.63	1.804	1.100	0.041
801	Halifax, N.S.	92.51	1.966	3.700	0.568
802	St. Catharines, Ont.	84.47	1.927	0.825	-0.836
804	Sudbury, Ont.	80.12	1.904	2.300	0.251
807	Regina, Sask.	112.14	2.050	3.193	0.504
809	Calgary, Alta.	249.64	2.397	6.351	0.803
811	Edmonton, Alta.	281.03	2.449	6.457	0.810
812	Hamilton, Ont.	273.99	2.438	3.142	0.497
813	Ottawa, Ont.	268.21	2.585	5.742	0.759
814	Vancouver, B.C.*	384.52	2.585	14.600	1.164
815	Winnipeg, Man.	265.43	2.424	4.400	0.644
816	Montreal, Que.	1,191.06	3.076	39.600	1.598
817	Toronto, Ont.	672.41	2.828	22.392	1.350

*Members of a group with similar characteristics.

Number of cities with CBD data available before sending questionnaires:		18
Total no. of questionnaires sent:	85	
Total no. of questionnaires returned:	46	
Percentage of response:	54%	
Total no. of usable questionnaires		27
Total no. of cities with adequate CBD data for study:		45

For original questionnaire see pages 99-101 inclusive. The CBD data of these 45 cities are shown in Table VIII (page 102).

A.3 Factor Analysis

Mathematics of Factor Analysis

Factor Analysis has become the generic term for a variety of procedures developed for the purpose of analyzing the inter-correlations within a set of variables.[3] In nature, it can be broadly divided into two categories:[4]

1. Principal-component method

It is a procedure for breaking down a covariance or correlation matrix into a set of orthogonal components or axes equal in number to the number of variates concerned. These correspond to the

---

[3] W. W. Cooley and P. R. Lohnes, op. cit., p. 151.

[4] D. N. Lawley and A. E. Maxwell, Factor Analysis as a Statistical Method (London: Butterworths, 1963), pp. 1-4.

latent roots and latent vectors of the matrix. No assumption of random variates has to be made for this method. In mathematical terms the concept of the principal-component method can be expressed as:

$$x_i = \sum_{r=1}^{p} w_{ir} z_r$$

where $i = 1,2,3,\ldots,p$

$r = 1,2,3,\ldots,p$

p = no. of variates

The score on i variate is viewed as a linear function of p factors, where w_{ir} and z_r are factor loading and factor score for the rth factor.

2. Factor analysis

It is a procedure to account for, or explain, the matrix of covariance by a minimum, or at least a small number of hypothetical variates or 'factors', as the following:

$$x_i = \sum_{r=1}^{k} l_{ir} f_r + e_i$$

where $i = 1,2,3,\ldots,p$, no. of variates

$r = 1,2,3,\ldots,k$, no. of factors

l_{ir} = factor loading of ith variate on r^{th} factor

f_r = factor score of rth factor

and $k < p$

The score, x_i, follows a multivariate normal distribution, and is

a linear function of k common factors and a residual, e_i, representing sources of variation affecting only the variate x_i.

In this study, the principal-component method is employed.

If the intercorrelation matrix of the original p variates is R, the solution of its characteristic equation, $|R - \lambda I| = 0$, gives the required latent roots $\lambda_{i,i=1,2,\ldots,p}$, and the corresponding vectors $v_{i,i=1,2,\ldots,p}$. The latent roots have the following property:

$$\sum_{i=1}^{p} \lambda_i = \sum_{j=1}^{p} r_{jj}$$

This is, the sum of the roots is equal to the trace of R, the total variance of the original variates to be accounted for.

Thus, factor loading matrix A can be calculated by the following formula:

$$A = V \Lambda^{\frac{1}{2}}$$

where V is the matrix of the latent vectors which has been normalized, i.e., $\sum_{i=1}^{p} v_i^2 = 1.0$, and $\Lambda^{\frac{1}{2}}$ is the diagonal matrix with $\lambda_i^{\frac{1}{2}}$ as the diagonal elements.

The number of components is exactly equal to the number of variates, and therefore in the factor loading matrix A, the sum of the squared loadings for a given low is equal to 1, i.e.:

$$\sum_{j=1}^{p} a_{ij}^{2} = 1.0$$

If only k factors are selected for further analysis, the following relationship should hold:

$$\sum_{j=1}^{k} a_{ij}^{2} = h^{2}$$

where h^2 is the communality of variate i when k factors are extracted. The remaining concern is the problem of how many factors to extract for further analysis. Usually, those factors with corresponding latent roots greater than 1 are used. However, in order to preserve higher degree of communality, in this study the factors extracted should account for at least 85% of the original variances.

Furthermore, the factor scores can be obtained by the following matrix multiplication:

$$F = Z\ A\ L_{dg}$$

where F is n by p factor score matrix of n subjects for p factors,

Z is n by p original standardized score matrix of n subjects for p variates,

A is p by p factor loading matrix of p variates and p factors

L_{dg} is diagonal matrix of p order with $1/\lambda_i$ on the principal diagonal.

Statistical Interpretation and Findings

A "canned" programme called FACTO is provided by the UBC Computing Center to perform factor analysis. In order to obtain a desirable percentage of trace in the factor space yet a less number of factors, two separate FACTO programmes have been run, using the same input data. The trial FACTO output, with 20 latent roots and 97% of trace, shows that the first eleven latent roots, accounting for 85% of the total variance, are capable of providing a clear factor structure. Then, a second FACTO is run in such manner that only eleven factors are extracted.

Table IX shows the intercorrelations of the 30 variables in an upper triangular correlation matrix. The highest inter-correlation is 0.896, median family income (14) with percentage of income over $10,000 (16). The second highest is -0.850, median family income (14) with percentage of income under $3,000 (15). These are undeniable truths. Except these, no more inter-correlation is higher than 0.80.

From this correlation matrix factor loadings are derived. Table X shows the relationships between each factor and the variables. The meanings of these factors are interpreted in Chapter IV.

For the convenience of further interpretation, the 100 sampled cities are arbitrarily classified into three groups as below:

Category	Population range
Small cities	50,000- 75,000
Medium cities	75,000-500,000
Large cities	500,000+

Table IX Intercorrelation of the 30 Variables

Variables*	(Means)	(Standard Deviations)	1	2	3	4	5
1. TICP	208.05(1000per)	513.86	1.0000	-0.1095	0.0097	0.3730	0.1730
2. PA5-	11.08(%)	1.58		1.0000	-0.7261	-0.1963	-0.1086
3. PA65+	10.08(%)	2.40			1.0000	0.0490	-0.1660
4. CPD	5976.08(per/sq.mi)	4156.26				1.0000	0.1817
5. PNW	11.12(%)	11.18					1.0000
6. PC	22.47(%)	35.89					
7. HOWNER	55.74(%)	13.48					
8. MVOOD	12453.37($)	3485.56					
9. HDSIZE	3.18(per)	0.26					
10.D1CAR	57.38(%)	5.67					
11.D2+CAR	17.32(%)	6.96					
12.DHFF	10.90(%)	4.66					
13.DTV	89.31(%)	5.77					
14.MFI	5871.31($)	879.38					
15.FI3T-	16.92(%)	6.10					
16.FI10T+	14.67(%)	6.67					
17.RSPC	1841.85($/per)	570.35					
18.WTPC	2991.58($/per)	2341.47					
19.SRPC	296.11($/per)	137.91					
20.EMPT	80.41(%)	384.00					
21.LF-RW	19.53(%)	3.02					
22.LF-WC	44.39(%)	6.65					
23.LF-M	28.48(%)	10.68					
24.LF-CT	8.08(%)	2.76					
25.LF-C	5.09(%)	1.47					
26.MEMBER	0.98	0.14					
27.CCITY	0.73	0.44					
28.RAIL	0.96	0.20					
29.PORT	0.28	0.44					
30.COMAIR	0.43	0.50					

*For complete variable names, see Chapter IV.

Table IX Intercorrelation of the 30 Variables (Cont'd)

Variables	6	7	8	9	10	11	12	13
1. TICP	-0.0856	-0.2271	0.1065	-0.1195	-0.3503	-0.1320	-0.1449	-0.0520
2. PA5-	0.5817	0.2644	-0.0857	0.7409	0.2670	0.2205	0.5079	0.0325
3. PA65+	-0.5743	-0.0703	-0.1110	-0.7347	0.0319	-0.3410	-0.3052	-0.1065
4. CPD	-0.1928	-0.5348	0.1725	0.0193	-0.4128	-0.4860	-0.4946	0.0642
5. PNW	-0.1641	-0.2335	-0.2134	-0.0527	-0.6205	0.0639	-0.2955	-0.2180
6. PC	1.0000	0.1950	0.1581	0.3822	0.1506	0.3872	0.4356	0.0634
7. HOWNER		1.0000	-0.3014	-0.0066	0.4803	0.6153	0.6753	0.2154
8. MVOOD			1.0000	-0.0103	0.1350	0.0813	0.0081	0.1556
9. HDSIZE				1.0000	0.1412	-0.0806	0.1146	0.0824
10. D1CAR					1.0000	0.0126	0.3776	0.3078
11. D2+CAR						1.0000	0.6173	0.1234
12. DHFF							1.0000	0.0692
13. DTV								1.0000

Table IX Intercorrelation of the 30 Variables (Cont'd)

Variables	14	15	16	17	18	19	20	21
1.TICP	0.0165	0.0176	0.0597	-0.1283	0.2158	0.4142	0.0001	-0.0169
2.PA5-	0.0461	-0.2231	-0.1547	-0.3562	-0.2772	-0.4751	0.1638	-0.0224
3.PA65+	-0.1062	0.1394	-0.0669	0.2470	0.0806	0.2338	-0.1116	0.2735
4.CPD	0.0146	-0.0876	-0.0018	-0.1950	0.0003	0.2147	-0.0574	-0.4405
5.PNW	-0.3614	0.6082	-0.0814	0.0482	0.2567	0.2821	-0.0755	-0.0535
6.PC	0.1309	-0.2009	0.0814	0.0063	-0.1044	-0.1001	0.2778	-0.0130
7.HOWNER	0.2705	-0.2087	0.1322	0.0179	-0.2079	-0.3598	0.0607	0.4320
8.MVOOD	0.6859	-0.5997	0.7142	0.4427	0.2812	0.4287	0.1237	-0.2794
9.HDSIZE	-0.1206	-0.0968	-0.2595	-0.3938	-0.2777	-0.4762	0.0054	-0.2433
10.D1CAR	0.3603	-0.5482	0.0835	0.0451	-0.3150	-0.3899	-0.0453	0.0943
11.D2CAR	0.4533	-0.1936	0.5165	0.2726	0.1148	0.0264	0.2259	0.2584
12.DHFF	0.3550	-0.3411	0.2189	0.0618	-0.0852	-0.1982	0.1109	0.4015
13.DTV	0.3743	-0.4259	0.2301	-0.0361	-0.0965	-0.2164	-0.0016	-0.1008
14.MFI	1.0000	-0.8500*	0.8955*	0.3453	0.1098	0.1141	0.1361	-0.1476
15.FI3T-		1.0000	-0.5675	-0.0285	0.0646	0.1117	-0.0981	0.2849
16.FI10T+			1.0000	0.5843	0.3107	0.3757	0.1038	-0.0731
17.RSPC				1.0000	0.3352	0.5509	0.0815	0.1982
18.WTPC					1.0000	0.6805	-0.0206	0.2981
19.SRPC						1.0000	0.0618	0.1131
20.EMPT							1.0000	-0.0608
21.LF-RW								1.0000

*Very high correlation (r > 0.8000)

Table IX Intercorrelation of the 30 Variables (Cont'd)

Variables	22	23	24	25	26	27	28	29	30
1. TICP	0.0276	-0.0556	0.0189	-0.0428	0.0379	0.1574	0.0578	0.2530	0.2904
2. PA5-	-0.0823	0.0839	0.1174	0.3384	-0.2652	-0.1335	-0.1071	-0.0285	-0.1419
3. PA65+	0.0901	-0.1257	0.0501	-0.2912	0.1856	0.4122	0.2776	0.0784	0.0902
4. CPD	-0.1738	0.1382	-0.1380	-0.1919	-0.1123	-0.1314	-0.1463	0.2019	0.1326
5. PNW	-0.3329	-0.1012	-0.1122	0.0372	0.1232	0.0917	0.2007	-0.0062	0.2008
6. PC	0.1530	-0.0636	-0.0078	0.6749	-0.2993	-0.2836	-0.2933	-0.1191	0.0097
7. HOWNER	0.2102	0.0684	0.1247	0.0863	0.2694	-0.0014	-0.1248	-0.2978	-0.1591
8. MVOOD	0.4755	-0.1303	-0.1526	0.0596	-0.0450	-0.3508	-0.2065	0.1084	-0.0022
9.HDSIZE	-0.2350	0.1656	0.0400	0.1878	-0.3286	-0.2244	-0.1918	-0.0366	-0.1307
10.D1CAR	0.0712	0.2499	0.0260	-0.1475	-0.0460	-0.2048	-0.0642	-0.1282	-0.2868
11.D2+CAR	0.4596	-0.1075	-0.0174	0.3045	0.2651	-0.1010	-0.2618	-0.2783	-0.0764
12.DHFF	0.3915	-0.2316	0.2653	0.3254	0.0036	-0.0755	0.0341	-0.1583	-0.0867
13.DTV	-0.0128	0.2751	-0.0360	-0.1392	0.2504	-0.1522	-0.2683	0.0556	-0.1102
14.MFI	0.5353	0.1115	-0.1678	-0.1051	0.1659	-0.4059	-0.3381	-0.0885	-0.1138
15.FI3T-	-0.3519	-0.2386	0.0743	0.1087	0.0257	0.3969	0.3154	-0.0037	0.1246
16.FI10T+	0.5961	-0.0665	-0.2403	-0.0296	0.2330	-0.3525	-0.2666	-0.1482	-0.0647
17.RSPC	0.4852	-0.3920	-0.1274	0.1296	0.1930	-0.1082	0.1705	-0.2284	-0.1049
18.WTPC	0.2838	-0.3325	0.0561	0.0502	0.1177	0.1213	0.0357	-0.0131	0.2719
19.SRPC	0.3769	-0.4417	-0.0654	0.1055	0.1336	0.1201	0.1529	0.1262	0.3569
20.EMPT	0.0876	0.0060	-0.1088	0.2306	0.0173	0.0640	0.0223	0.1840	0.1260
21.LF-RW	0.3379	-0.4656	0.3915	0.1930	0.1270	0.4733	0.1679	-0.0618	0.0924
22.LF-WC	1.0000	-0.6188	0.2678	0.2693	0.0746	-0.0027	-0.1822	-0.1475	0.1419
23.LF-M		1.0000	-0.5035	-0.4719	0.0308	-0.3430	-0.2939	0.0238	-0.3512
24.LF-CT			1.0000	0.1829	-0.1336	0.2330	0.0706	0.0535	0.3145
25.LF-C				1.0000	-0.2196	0.0618	-0.0139	-0.0212	0.2760
26.MEMBER					1.0000	0.2383	0.1786	0.0378	-0.0005
27.CCITY						1.0000	0.3406	0.2967	0.2969
28.RAIL							1.0000	0.1216	0.1783
29.PORT								1.0000	0.1276
30.COMAIR									1.0000

TABLE X

FACTOR LOADINGS ON 30 VARIABLES AFTER ROTATION

VARIABLES	FACTORS I SEIND	II BSIND	III LSIND	IV HSIND	V FDIND	VI DIND	VII PCIND	VIII GF-I	IX GF-II	X GF-III	XI POVIND
1. TICP						76					
2. PA5-				86							
3. PA65+				-90							
4. CPD			-68			37					
5. PNW											-78
6. PC				55			-62				
7. HOWNER			86								
8. MVOOD	82		-32								
9. HDSIZE				87							
10. D1CAR											82
11. D2+CAR	38		75								
12. DHFF			73	33							
13. DTV									84		
14. MFI	90										
15. FI3T-	-71										-54
16. FI10T+	90										
17. RSPC	51	-39		-34		-45					
18. WTPC		-84									
19. SRPC	34	-65		-35							
20. EMPT							-73			-30	
21. LF-RW		-38	62		42						
22. LF-WC	63				56						
23. LF-M					-81						
24. LF-CT					82						
25. LF-C				30	39		-69				
26. MEMBER			36	-32					50		-37
27. CCITY	-43			-32	31					-53	
28. RAIL								90			
29. PORT										-82	
30. COMAIR					43	56	-39				

Figures have been converted to percentages and only those larger than 30 are included

It is of interest to compare the characteristics of these groups of cities. Three graphs are so constructed that all these cities are plotted against two factors at a time, and the percentages of distributions are also provided.

Figure 18 shows the distributions of cities on the factor space composed of Socio-economic Index (I) and Living Standard Index (III). It can be seen that 67% of North American cities, including Vancouver and Montreal, are more or less of low socio-economic status. Again, 60% of them reveal that the living standard is below average. The concentration of Canadian Cities in the 3rd quadrant means that, in comparison with the U.S. cities, they exhibit lower socio-economic status and living standard. For the small cities, only 16.7% have positive socio-economic scores, and as much as 70.8% are of low living standard. The situation for medium sized cities are better. 51% of them show high living standard and 33.9% show sound socio-economic status. Finally, 47.8% of the large cities have high scores on factor I and 73.9% have scores on factor III less than zero. This means that the large cities usually have low living standard but good socio-economic status.

Figure 19 shows factor scores of all the sampled cities on Factor II, Business Index, and Factor V, Functional Diversification Index. It reveals that Canadian cities are more functionally diversified than the U.S. cities. In terms of business status, large cities are better than the smaller ones because the percentages of having high negative scores on Factor II for the large, medium and small cities are 65.2,

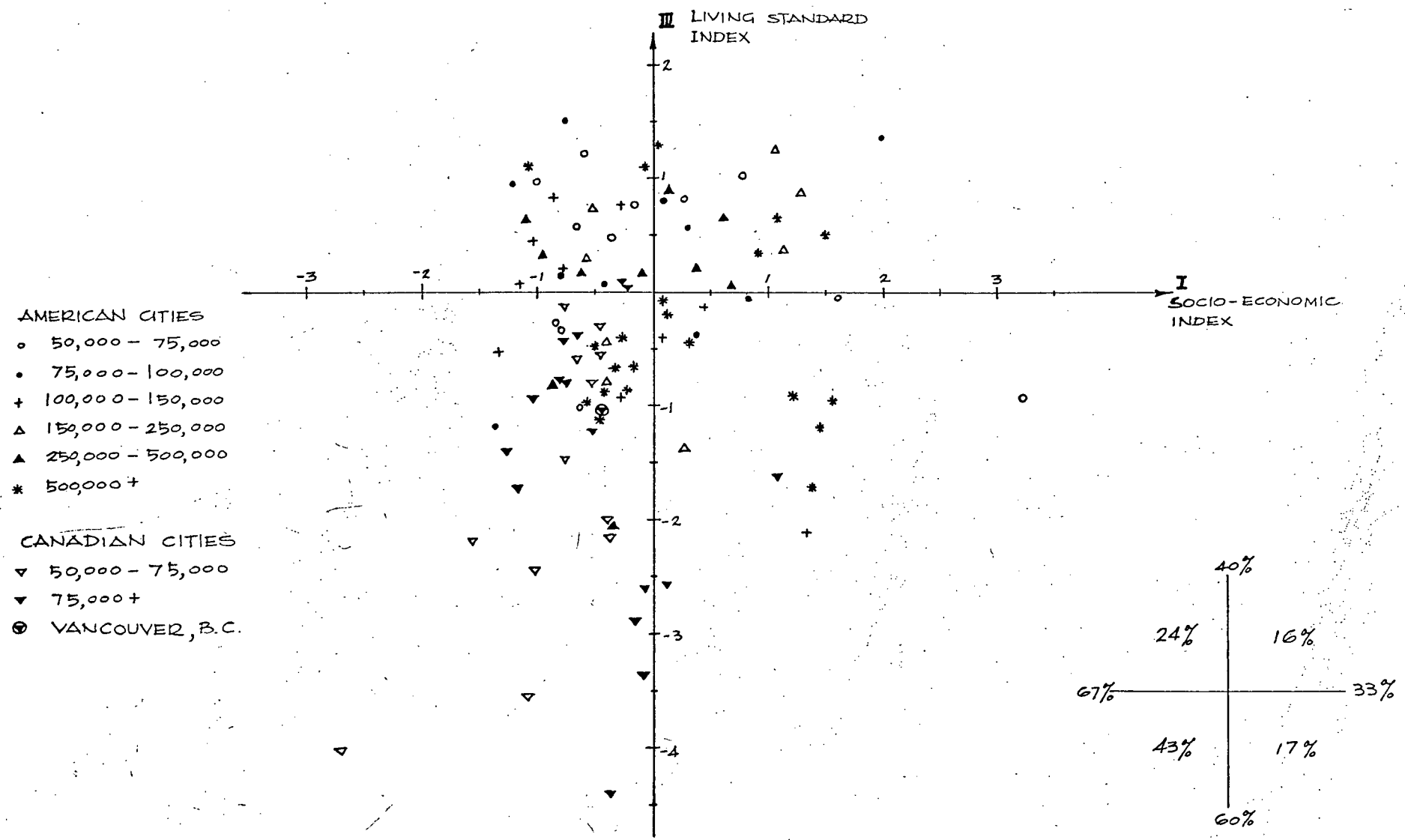


FIG. 18 FACTOR SCORES OF 100 NORTH AMERICAN CITIES ON FACTORS I & III

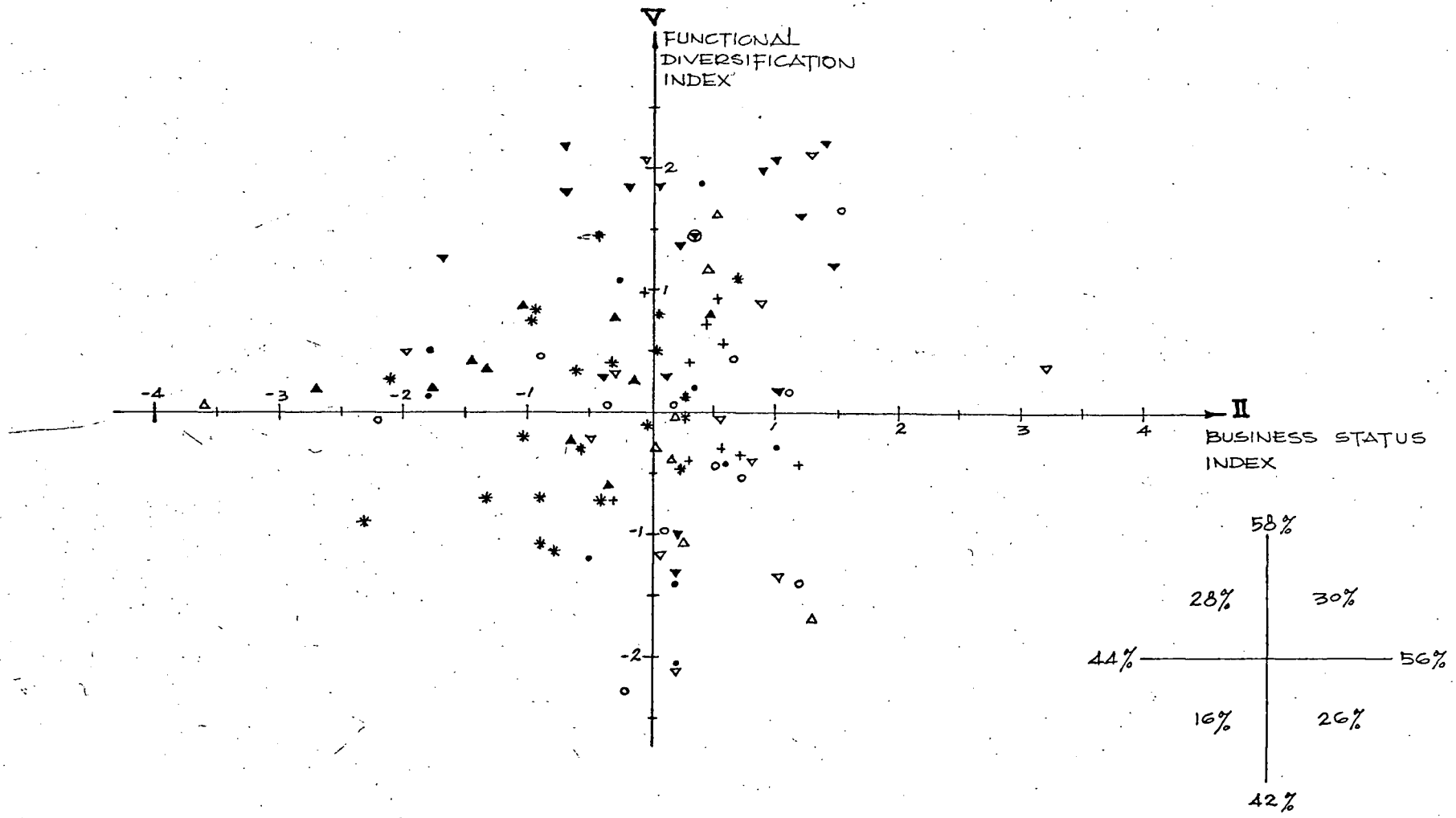


FIG. 19 FACTOR SCORES OF 100 N. AMER. CITIES ON FACTORS II & V

39.6, and 33.3 respectively. That Vancouver, Toronto, and Montreal are in the first quadrant means that they are functionally diversified but they do not have very sound business condition.

Finally, Figure 20 shows the distributions of sampled cities against population and household size indices. Obviously, the large cities have high scores on Factor VI, the Population Index. Because of the much higher proportion of Canadian cities lying in the upper half of the graph, it is concluded that the average household size of Canadian cities is much larger than that of the U.S. cities. However, Vancouver is an exception because it shows negative score on Household Size Index.

Summarized statistics are given in Table XI.

A.4 Distance-scaling of Similarities and Grouping Techniques

As it has been mentioned in Chapter III, similarity of two subjects is usually measured indirectly from the dissimilarity between them. In a two dimensional Cartesian coordinates, two subjects, A and B, measured by two variates X and Y, can be represented as two points $A(X_a,Y_a)$ and $B(X_b,Y_b)$. Obviously, their dissimilarity is the distance between point A and point B. As in the following graph, the distance is calculated by Pythagorous theorem.

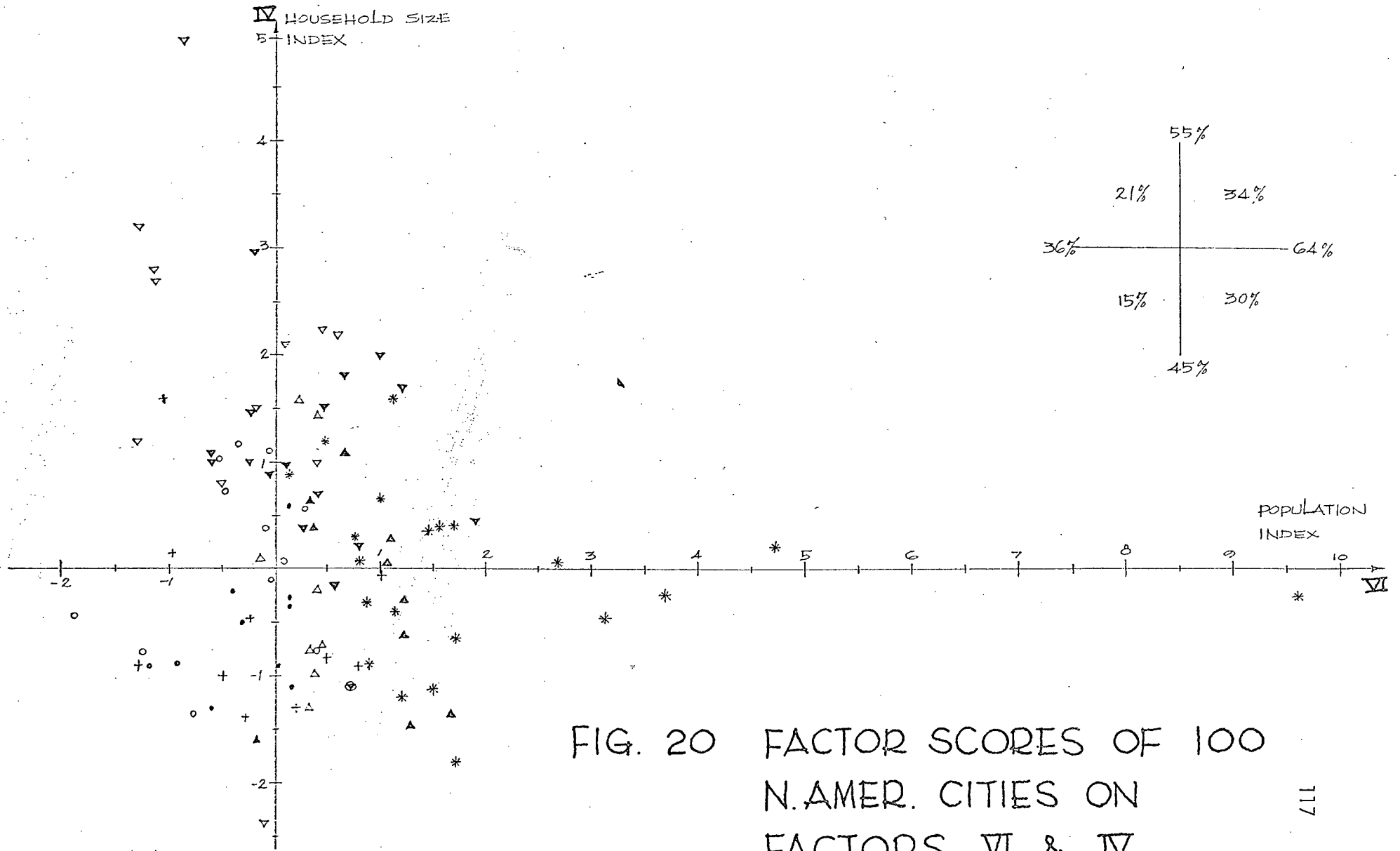


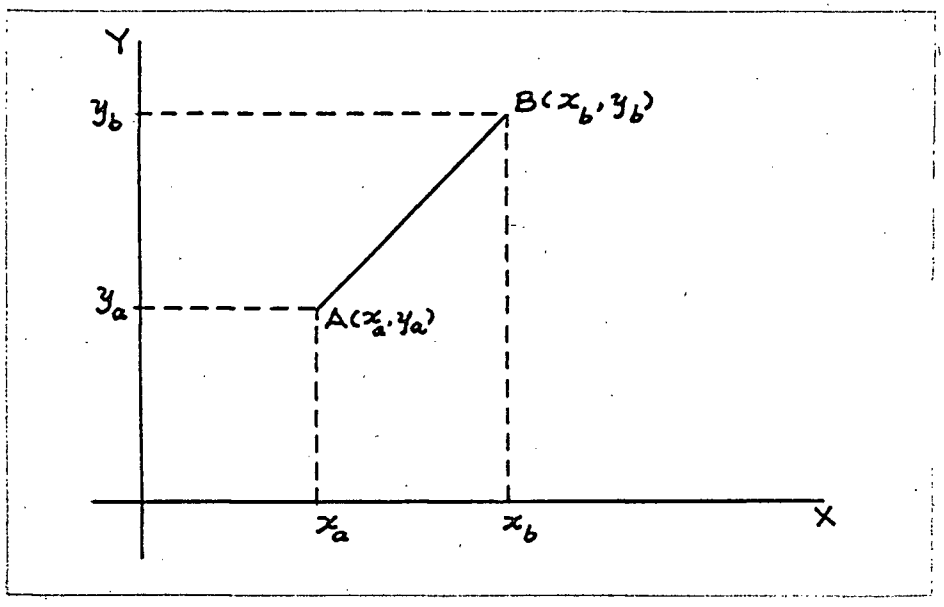
FIG. 20 FACTOR SCORES OF 100 N.AMER. CITIES ON FACTORS VI & IV

TABLE XI

SUMMARIZED STATISTICS FOR FIGURES 18, 19, & 20

LOCATION	SMALL CITIES	MEDIUM CITIES	LARGE CITIES	OVERALL
A. Refer to Fig. 18, abscissa = I, ordinate = III				
Quadrant - 1	8.3	18.9	17.4	16.0
2	5.0	32.1	8.7	24.0
3	15.0	34.0	43.5	43.0
4	2.0	15.0	30.4	17.0
Half-Right	16.7	33.9	47.8	33.0
Left	83.3	66.1	52.2	67.0
Upper	29.2	51.0	26.1	40.0
Lower	70.8	49.0	73.9	60.0
B. Refer to Fig. 19, abscissa = II, ordinate = V				
Quadrant - 1	29.2	32.1	26.1	30.0
2	20.8	32.1	26.1	28.0
3	12.5	7.5	39.1	16.0
4	37.5	28.3	8.7	26.0
Half-Right	66.7	60.4	34.8	56.0
Left	33.3	39.6	65.2	44.0
Upper	50.0	64.2	52.2	58.0
Lower	50.0	35.8	47.8	42.0
C. Refer to Fig. 20, abscissa = VI, ordinate = IV				
Quadrant - 1	25.0	28.3	56.5	34.0
2	50.0	17.0	0.0	21.0
3	20.8	18.9	0.0	15.0
4	4.2	35.8	43.5	30.0
Half-Right	29.2	64.1	100.0	64.0
Left	70.8	35.9	0.0	36.0
Upper	75.0	45.3	56.5	55.0
Lower	25.0	54.7	43.5	45.0

**Figures are percentages of distribution in the specific locations in the corresponding figure, i.e., the first quadrant, the upper half of the graph, etc.



$$D_{AB} = ((X_b - X_a)^2 + (Y_b - Y_a)^2)^{\frac{1}{2}}$$

If there are p variates measuring the subjects with which distances are calculated, the formula is generalized below:

$$D_{AB} = (\sum_{i=1}^{p} (F_{Ai} - F_{Bi})^2)^{\frac{1}{2}}$$

where F_{Ai} is the score of subject A on variate or factor i, and i run from 1 to p.

The "distance measures provide objective quantitative estimates of the similarity of sampled individuals"[5] for multivariate analysis. Berry considers this procedure as distance-scaling of similarities.[6]

---

[5]B. J. L. Berry, "A Note Concerning Methods of Classification", Annals, American Association of Geographers, Vol. 48, 1958, pp. 300-303.

[6]_____, "Grouping and Regionalization: An Approach to the problem of Using Multivariate Analysis", Quantitative Geography (Northwestern University, No. 13), pp. 219-251.

An important feature of distance-scaling in an p-dimensional space is that the variates must be orthogonal, that is, perpendicular, to each other. Furthermore, Mahalanobis D^2 statistics should be used in case that for every observation several values of each variate are taken.[7]

Given an n by p score matrix of n sampled subjects on p factors, there are two alternatives of grouping algorithms based on the principal of distance measure of similarities:[8]

1. Maximum similarity grouping

   Consider n points in the p-dimensional space, the procedure starts with calculating an n by n distance matrix (more precisely, an n by n-1 upper or lower triangular matrix). Link the pair with minimum distance, then calculate their centroid. The next step begins with n-1 points, actually n-2 points plus 1 centroid, and link the closest pair again. The procedure is repeated until finally only one group is obtained.

---

[7]For detail, see C. R. Rao, Advanced Statistical Methods in Biometric Research (New York: John Wiley and Sons, 1952).

[8]Berry suggests a third alternative, gravity grouping, which is of less importance and thus is omitted here. For detail, see B. J. L. Berry, "A Synthesis of Formal and Functional Regions Using a General Field Theory of Spatial Behavior", in Berry and Marble (eds.), Spatial Analysis, A Reader in Statistical Geography (Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1968), pp. 419-428, particularly on p. 424.

2. Optimum homogeneity grouping

Suppose that two or more points are linked, the group can be represented by its centroid, a vector of p means. However, this implies some losses of information. Ward[9] suggests a "value-reflecting" number, or error sum of squares, to measure the loss of information after grouping. That is:

$$ESS = \sum_{i=1}^{g} \sum_{j=1}^{p} (F_{ij} - C_j)^2$$

where F_{ij} is factor score of subject i in the group with g members on variate j, C_j is the corresponding element of the centroid.

In order to obtain an optimum homogeneity group, the above objective function has to be minimized at each step of grouping. The procedure is exactly the same as the first method, except that an matrix of error sum of squares instead of a distance matrix is employed.

Comparing the two grouping techniques, several points are worth mentioning:

1. After investigating the nature and procedure of each of the algorithms, it is not difficult to find that the optimum homogeneity

---

[9] J. H. Ward, Jr., "Hierarchical Grouping to Optimize an Objective Function", Journal of American Statistical Association, March 1963, pp. 236-244.

grouping tends to maintain groups of similar size whereas the maximum similarity grouping tends to form at least one group of large size. As an illustration, suppose that A, B, C and D are subjects measured in two-factor space as in Figure 21 below, A and B is grouped at the first step and their resultant centroid is at E.

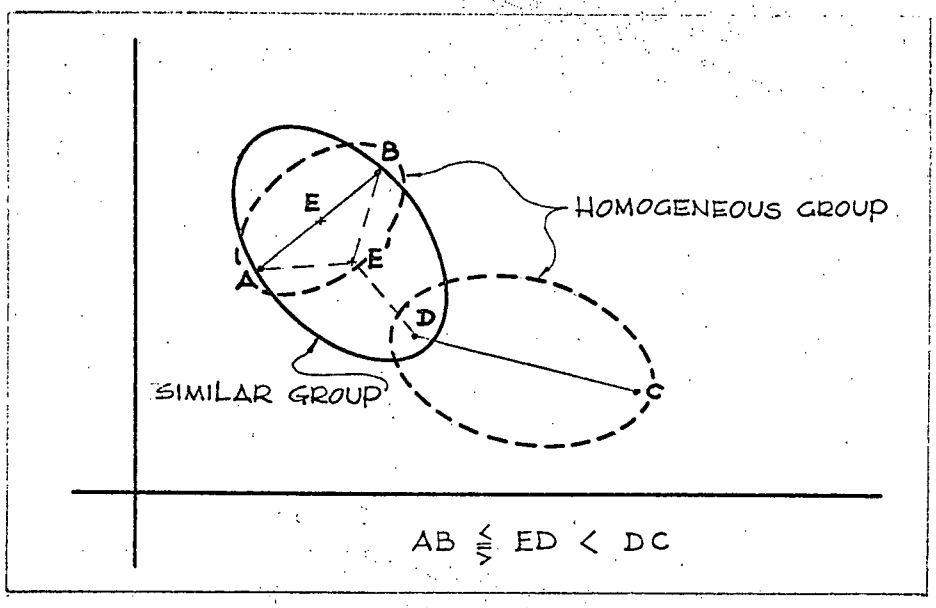


FIGURE 21

HYPOTHETICAL GROUPING FROM DIFFERENT CRITERIA

Obviously ED is smaller than CD, thus D is added to form a large group containing A, B, and D, based on the criterion of maximum similarity. However, based on the criterion of optimum homogeneity, if the summation of AE', BE' and DE' is greater than CD where E' is the centroid of this large group, a group of C and D will be formed, against that of A and B.

2. The maximum similarity grouping creates more distinct groups, that is to say, the overlapping areas among groups are smaller, than the optimum homogeneity grouping does.

3. The group members created by optimum homogeneity grouping are not necessarily of optimum similarity. On the contrary, the members in a maximum similarity group exhibit less homogeneity and more loss of information.

4. The algorithm of maximum similarity grouping is the simplest and the most straightforward among all the other grouping techniques.

5. Since both techniques have merits and weaknesses, the selection of an appropriate one depends on the nature and purpose of the analysis.

The criterion of maximum similarity is employed in this study.[10]

In addition to the distance measures of similarity (squared), two testing criteria have also been programmed to the IBM system 360/67, they are:

1. Squared average intergroup distance to check the degree of dissimilarity resulted from grouping. The formula is:

$$D^2 = \sum_{i=1}^{g} \sum_{k=i+1}^{g} \sum_{j=1}^{p} (F_{ij} - F_{kj})^2 / {}_gC_2$$

[10]For reasons, see Chapter IV.

where g = no. of subjects in the group

p = no. of variates (factors)

F_{ij} = score of subject i on factor j

${}_gC_2$ = no. of combinations of g subjects taken 2 at a time, that is, $g! / (g-2)!\ 2!$

The maximum value of D^2 can be obtained at the final step of grouping at which only one group is remained. The ratio D^2/ Max.D^2 will give the relative degree of dissimilarity at each step. If the ratio exceeds a certain acceptance level, the grouping should be terminated at one step before.

2. Squared Bachi's standard distance[11] which is closely analogous to Ward's error sum of squares is used to measure the degree of heterogeneity (or average loss of information) at each step of grouping. Bachi's squared standard distance takes the form of:

$$d^2 = \sum_{i=1}^{g} \sum_{j=1}^{p} (F_{ij}-C_j)^2/g$$

where F_{ij} is the factor score of subject i on factor j, and C_j is the corresponding element of the centroid vector. Similarly, an acceptance ratio of d^2/Max.d^2 is predetermined to maintain a desirable degree of homogeneity in the group.

---

[11]For detailed account for standard distance, see R. Bachi, "Statistical Analysis of Geographical Series", in Berry and Marble, Spatial Analysis, A Reader in Statistical Geography (Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1968), pp. 101-109, and L. King, Statistical Analysis in Geography (Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1968), pp. 92-93.

Using an 100 by 11 factor matrix as input data, a grouping programme has been run.[12] The maximum squared average intragroup distance and the maximum squared Bachi's standard distance are:

$\text{Max.}D^2 = 21.87$ $\qquad$ $\text{Max.}d^2 = 10.89$

This means that, if all the sampled cities are treated as one group, the average degree of dissimilarity is 21.87 and the average loss of information is 10.89. In order to get a stringent group with high degree of similarity as well as homogeneity among all members, the acceptance levels for both criteria are set as one-third of the maximum values. Thus:

Upper limit of $D^2 = 7.29$ $\qquad$ Upper limit of $d^2 = 3.63$

Any value of D^2 of d^2 at a step exceeding the corresponding acceptance level will result in cutting off the grouping process. At the same time, profiles showing the change of D^2 and d^2 at each grouping step are constructed for the main group which absorbs all the sampled cities at the final step. In Figure 22, it is easily seen that each of the profiles shows a series of terraces; steadily flattened at the first 30 steps and dramatically steepened at the last 10 steps. No doubt should the cutting point of the grouping process be somewhere at the edge of the next terrace between the 30th and the 90th steps. The acceptance levels for D^2 and d^2 indicate that such action should be done at the 53rd step so that 47 groups are left at this stage.

---

[12]For this grouping programme, see Appendix B.3.

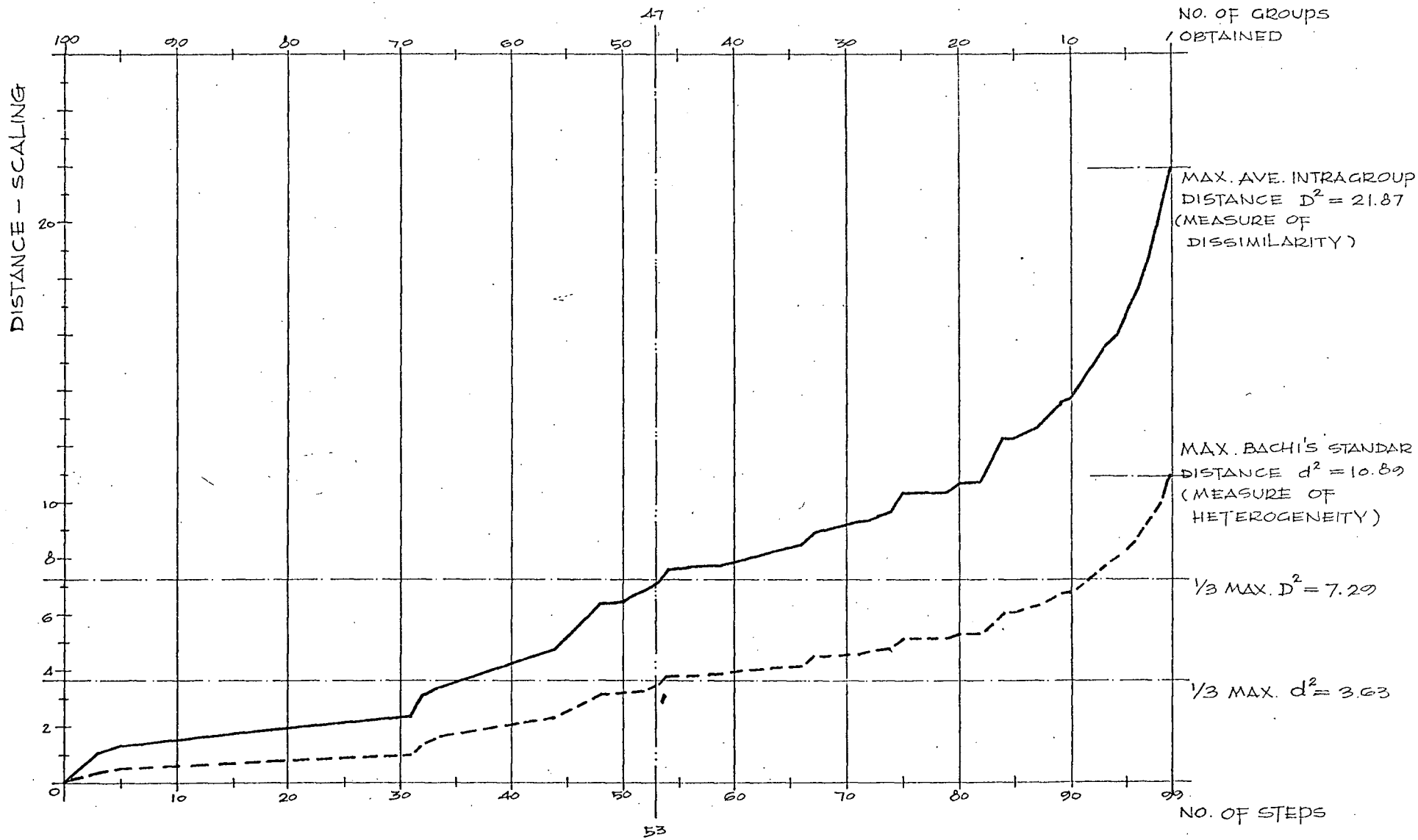


FIG. 22 PROFILES OF D^2 AND d^2

As a counter check of the cutting action, another profile showing the rate of change of linkage (i.e., distance between centroids) also indicates that, except at the very beginning of the grouping process, the first greater-than-average rate of change of linkage occurs immediately after the 53rd step. (See Figure 23).

The linkage tree showing the entire grouping process is presented in Figure 24, and the members of the 47 groups are listed in Table XII below.

A.5 Statistical Test of the Regression Models

In order to test the difference of coefficients of correlation between two regression models, Fisher's z-transformation is applied as following:

Original model - $R_1 = 0.90$ $N_1 = 45$

Revised model - $R_2 = 0.96$ $N_2 = 17$

$$Z_1 = 1.15 \log ((1+R_1)/(1-R_1))$$
$$= 1.47$$

$$Z_2 = 1.15 \log ((1+R_2)/(1-R_2))$$
$$= 1.95$$

$$Z = (Z_2-Z_1)/((1/(N_1-3)) + (1/(N_2-3)))^{\frac{1}{2}}$$
$$= 1.56$$

Z - Probability = 0.059 > 0.050

It is not significant at the conventional 0.05 level for a one

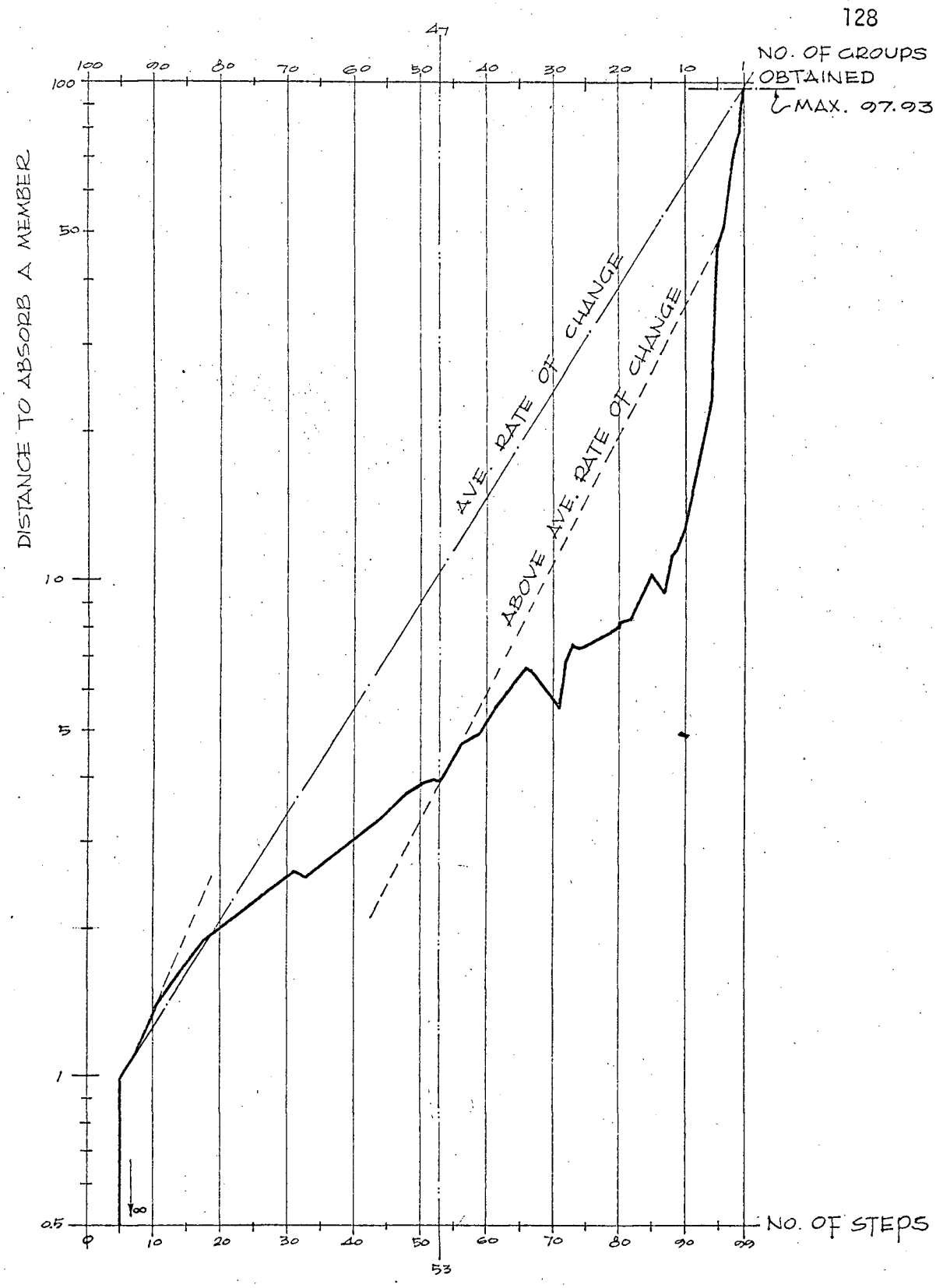


FIG. 23 RATE OF CHANGE OF LINKAGE AT EACH STEP OF GROUPING

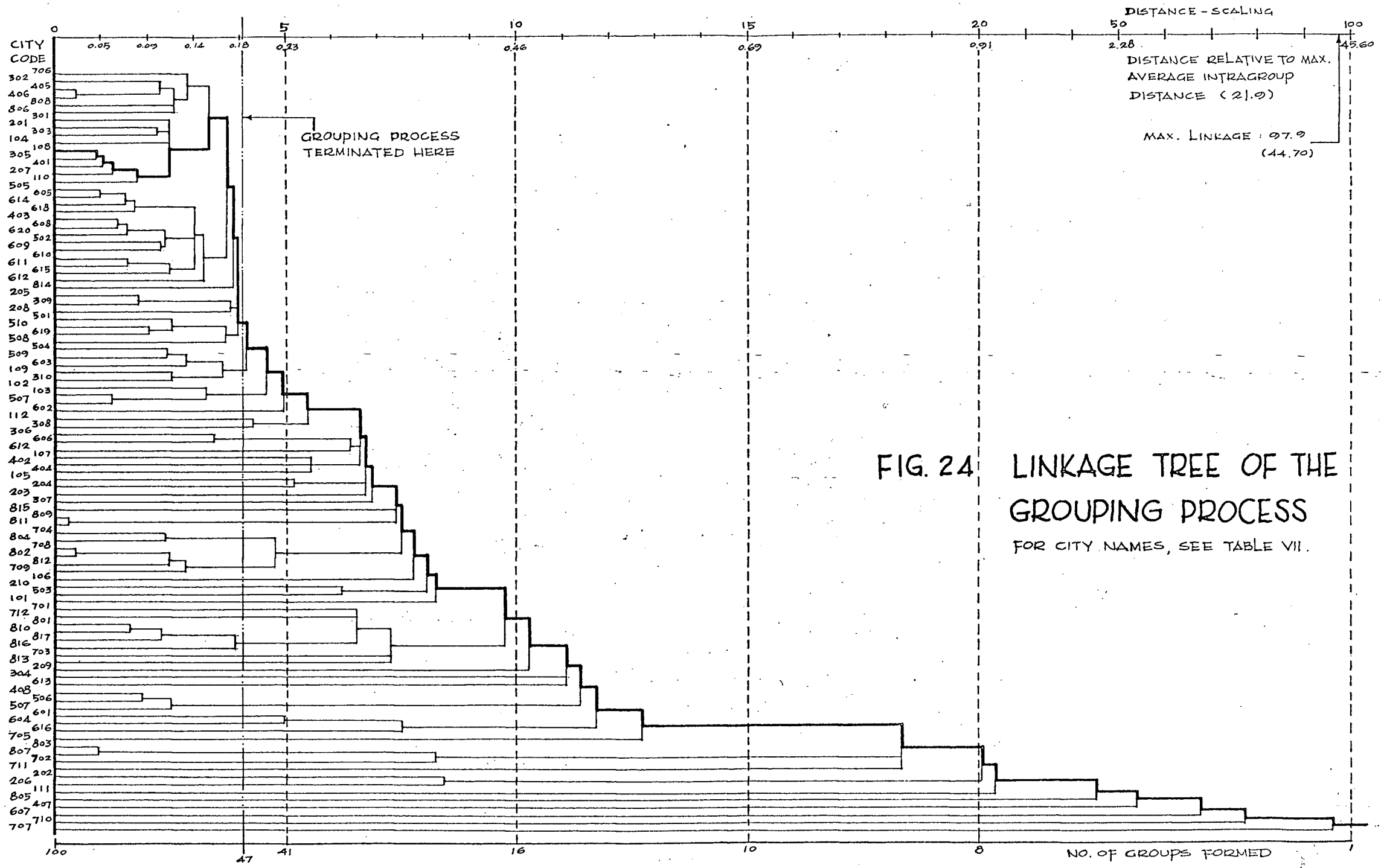


FIG. 24 LINKAGE TREE OF THE GROUPING PROCESS

FOR CITY NAMES, SEE TABLE VII.

TABLE XII

MEMBERS OF THE 47 GROUPS

Group	
1	706, 302, 405, 406, 808, 806, 301, 201, 303, 104, 108, 305, 401, 207, 110, 505, 605, 614, 618, 403, 608, 620, 502, 609, 610, 611, 615, 612, 814, 205, 309, 208, 501, 510, 619, 508 (36 members)
2	504, 509, 603, 109, 301 (5 members)
3	102, 103, 507 (3 members)
4	602 (1 member)
5	112 (1 member)
6	308 (1 member)
7	306, 606 (2 members)
8	612 (1 member)
9	107 (1 member)
10	402 (1 member)
11	404 (1 member)
12	105 (1 member)
13	204 (1 member)
14	203 (1 member)
15	307 (1 member)
16	815 (1 member)
17	809, 811 (2 members)
18	704, 804 (2 members)
19	708, 802, 812, 709 (4 members)
20	106 (1 member)
21	210 (1 member)
22	503 (1 member)
23	101 (1 member)
24	701 (1 member)
25	712 (1 member)
26	801, 810, 817, 816 (4 members)
27	703 (1 member)
28	813 (1 member)
29	209 (1 member)
30	304 (1 member)
31	613 (1 member)
32	408, 506, 507 (3 members)
33	601 (1 member)
34	604 (1 member)
35	616 (1 member)
37	803, 807 (2 members)
38	702 (1 member)
39	711 (1 member)
40	202 (1 member)
41	206 (1 member)
42	111 (1 member)
43	805 (1 member)
44	407 (1 member)
45	607 (1 member)
46	710 (1 member)
47	707 (1 member)

For city names, refer to Table VII.

tail test. However, since the probability is very close to the critical value, the judgement is suspended. (Refer to Section 4.4).

A.6 Relationships Between Sample Correlation and Population Correlation Against Sample Size

Because of the sampling error, the correlation obtained from a set of sampled observations may be different to the true population correlation. Confident belts for deriving upper and lower limits of estimated true population correlation have been devised.[13] The following table summarizes the intervals of a sampled correlation of 0.90 for different confident levels and for different sample sizes:

TABLE XIII

LOWER AND UPPER LIMITS OF ESTIMATED TRUE POPULATION CORRELATION FROM SAMPLE CORRELATION OF 0.90 AT 90%, 95% AND 99% CONFIDENT LEVELS

Sample	Confident Levels					
	90%		95%		99%	
Size	Upper	Lower	Upper	Lower	Upper	Lower
N = 3	-0.55	+1.00	-0.76	+1.00	-0.96	+1.00
N = 4	+0.00	+0.98	-0.27	+0.99	-0.73	+1.00
N = 5	+0.29	+0.97	+0.95	+0.99	-0.40	+1.00
N = 6	+0.46	+0.97	+0.30	+0.98	-0.07	+0.99
N = 10	+0.67	+0.96	+0.60	+0.97	+0.43	+0.98
N = 45	+0.88	+0.94	+0.83	+0.95	+0.77	+0.95

Source: F. N. David, Table of the Correlation Coefficient (Cambridge: The Univ. Press, 1938), Charts I, II and IV.

[13] These confident belts have been provided by F. N. David, in Tables of the Correlation Coefficient (Cambridge, England: The University Press, 1938).

Correlation obtained from a sample less than three subjects is meaningless. Further, from the above table, a sample with greater than six subjects provides acceptable estimates of true population correlation coefficient, so that this is considered the minimum sample size for regression analysis in this study.

APPENDIX B

DATA AND COMPUTER PROGRAMME

B.1 Raw Data of 100 North American Cities

B.2 Factor Scores of 100 North American Cities

B.3 Maximum Similarity Grouping Programme

APPENDIX B.1

RAW DATA OF 100 NORTH AMERICAN CITIES ON 30 CHARACTERISTIC VARIABLES

(I5,9X,(8F7.0))

VARIABLES NAMES:
IDENTIFICATION TICP PA5- PA65+ CPD PNW PC HOWNER MOVOOD
HDSIZE D1CAR D2+CAR DHFF DTV MFI FI3T- FI10T+
RSPC WTPC SRPC EMPT LF-RW LF-WC LF-M LF-CT
LF-C MEMBER CCITY RAIL PORT COMAIR

```
10101          62.06  11.9   6.3    2048.  18.1   55.3   57.9   9000.
3.4    52.1    25.1   11.7   86.2   5117.  20.7   11.1
1572.  2920.   276.   44.4   15.9   33.1   49.4   4.06
4.72   1.      0.     1.     0.     0.
10201          62.33  11.9   8.3    4517.  1.2    68.6   73.8   12600.
3.1    62.6    25.7   21.2   92.7   6535.  10.9   16.0
1571.  391.6   146.7  39.0   20.9   46.7   30.6   9.48
6.18   1.      0.     1.     0.     0.
10301          66.78  11.0   10.1   4703.  7.0    29.4   63.4   14300.
3.2    62.7    17.5   10.5   91.5   6958.  11.2   21.6
2256.  1470.   241.8  37.9   17.9   45.7   33.2   11.05
5.14   1.      0.     1.     0.     1.
10401          56.61  12.3   11.5   4162.  0.3    14.    63.6   13400.
3.3    62.1    15.2   15.7   83.6   6373.  12.1   16.2
1533.  1372.   156.5  37.1   21.1   43.6   33.3   6.19
4.05   1.      1.     1.     0.     0.
10501          62.89  12.9   9.2    3743.  0.9    19.3   65.0   11800.
3.4    67.7    15.7   16.0   95.4   5981.  12.2   11.7
1927.  3493.   305.5  35.43  24.0   43.9   28.2   11.42
4.48   1.      1.0    1.0    1.0    0.0
10601          55.72  12.8   7.1    5020.  8.4    43.1   57.2   17000.
3.3    66.3    19.5   13.5   93.5   7254.  8.5    23.5
2306.  1890.   286.   41.5   16.0   45.3   36.4   6.06
3.84   1.0     0.0    1.0    0.0    0.0
```

10701		70.2	12.8	8.6	3714.	3.5	22.9	65.5	12100.
3.3	53.4	30.2	17.5	90.2	6145.	14.0	15.7		
1885.	1681.	195.1	36.5	19.0	50.2	14.3	12.9		
5.86	1.	1.	1.	0.	0.				
10801		51.04	12.1	11.0	6149.	10.5	1.6	62.4	9900.
3.2	62.1	16.5	11.4	88.4	5637.	18.6	11.2		
1973.	1501.	232.5	37.2	19.7	42.3	35.5	7.85		
4.14	1.	1.	1.	0.	0.				
10901		60.19	9.6	11.4	2508.	19.0	13.6	58.4	9401.
3.1	55.1	17.8	9.7	84.0	4648.	30.0	10.7		
2510.	2020.	384.	39.2	22.8	46.5	18.1	6.78		
5.09	1.	1.	1.	0.	0.				
11001		72.5	10.1	13.7	2935.	5.8	12.9	70.1	6900.
2.9	59.8	17.	10.8	89.0	5292.	21.9	9.4		
2440.	1920.	229.	37.22	23.3	43.3	25.9	9.26		
4.45	1.	1.	1.	0.	0.				
11101		50.49	8.6	11.	5371.	11.9	16.1	41.1	25500.
3.	59.3	19.9	10.7	92.4	8012.	9.1	37.3		
4084.	7875.	594.	45.6	18.8	55.8	16.1	5.65		
5.22	1.	0.	1.	0.	0.				
11201		63.66	11.0	8.8	14468.	33.4	-3.6	50.9	8900.
3.3	50.2	9.3	5.7	92.1	5343.	22.1	10.3		
1439.	674.	148.9	36.22	16.7	30.7	41.9	4.33		
3.98	1.	0.	1.	0.	0.				
20101		75.94	9.5	13.6	6967.	1.7	-5.9	48.1	13800.
3.0	59.4	15.4	7.8	91.5	6251.	12.6	17.0		
2064.	1467.	264.	40.5	19.	46.	36.3	6.18		
4.24	1.	1.	1.	0.	0.				
20201		82.08	9.6	8.6	7016.	0.2	27.2	65.9	18000.
3.2	62.	25.	6.9	97.	7271.	7.3	22.6		
1198.	4820.	245.	42.4	16.5	46.5	50.3	6.33		
4.23	1.	0.	0.	0.	0.				
20301		89.16	11.4	11.5	1805.	2.1	6.2	66.2	9900.
3.1	60.8	17.9	11.2	91.4	5812.	15.9	13.5		
1673.	7725.	229.	38.3	27.2	47.4	22.4	8.23		
5.58	1.	1.	1.	0.	0.				

20401		89.14	12.4	9.7	7959.	5.4	25.2	65.0	13800.
3.3	64.2	16.7	14.6	95.3	6758.	10.2	18.2		
1512.	1022.	160.	39.0	21.1	37.7	44.7	4.02		
3.47	1.	1.	1.	1.	0.				
20501		97.11	9.9	9.9	3735.	17.	5.6	63.9	10300.
3.2	56.8	16.4	9.9	86.6	5103.	23.8	10.9		
1694.	2235.	307.	38.7	21.8	46.7	18.7	15.42		
5.32	1.	1.	1.	0.	1.				
20601		80.61	13.1	5.5	6718.	0.2	71.9	80.6	14200.
3.5	58.3	35.5	14.8	97.3	8184.	6.2	30.7		
1740.	1131.	179.	36.7	18.4	59.1	39.5	5.75		
4.73	1.	0.	0.	0.	0.				
20701		78.	10.9	11.2	3960.	7.7	17.7	69.1	11900.
3.	61.2	18.8	12.9	88.7	6054.	15.5	14.7		
2067.	2650.	232.	38.2	20.7	47.4	28.9	9.8		
4.36	1.	1.	1.	0.	0.				
20801		95.87	9.7	12.2	2763.	2.6	43.7	65.4	9100.
2.9	58.7	20.1	14.9	88.6	4955.	24.	9.3		
1862.	2832.	281.	38.3	25.3	46.8	20.3	11.18		
5.9	1.	1.	1.	0.	1.				
20901		99.94	9.2	12.0	2948.	0.6	-10.7	30.9	10600.
3.0	56.8	11.0	2.0	94.6	4970.	20.	6.6		
1325.	1034.	132.	41.6	15.	28.1	52.2	3.85		
4.09	1.	1.	1.	1.	0.				
21001		88.14	10.1	11.9	4177.	23.5	68.3	57.5	13900.
2.9	54.	23.2	8.2	85.1	5037.	25.5	14.8		
3253.	4620.	567.	37.7	23.	49.9	12.4	5.72		
9.94	1.	1.	1.	0.	1.				
30101		147.98	10.4	12.2	3115.	5.3	3.	67.5	9400.
2.9	55.8	20.2	16.6	89.4	5993.	17.3	13.9		
1800.	2480.	245.	34.9	20.7	45.1	23.2	8.66		
6.15	1.	1.	1.	1.	1.				
30201		111.44	8.8	12.6	4405.	0.7	-11.2	52.	9000.
3.1	60.9	9.7	5.	93.7	5005.	21.1	9.3		
1460.	1945.	176.5	36.5	20.6	42.0	33.7	10.38		
3.21	1.	1.	1.	0.	1.				
30301		100.41	9.9	13.9	5906.	3.2	-1.1	57.5	14000.
3.0	60.1	10.9	7.1	90.6	5873.	15.7	13.9		
1494.	1633.	278.	37.6	18.5	46.2	30.4	6.9		
3.19	1.	1.	1.	1.	0.				

30401		107.72	8.6	11.7	17098.	6.3	-10.8	22.5	13800.
2.8	51.9	8.6	2.5	81.9	5923.	15.3	17.4		
2209.	4560.	492.	42.9	13.3	50.6	26.1	5.4		
2.6	1.	0.	1.	0.	0.				
30501		141.54	10.9	10.6	4423.	6.6	10.	66.6	8600.
3.1	58.	18.	9.6	87.1	5299.	21.6	10.8		
1551.	2005.	221.	35.4	22.1	42.9	30.6	6.95		
4.58	1.	1.	1.	0.	0.				
30601		149.25	12.9	6.9	3596.	35.7	24.7	46.8	10100.
3.4	49.1	19.2	10.5	84.1	4761.	28.7	10.7		
1276.	1952.	162.	33.9	20.6	41.5	20.7	10.7		
6.92	1.	1.	1.	1.	0.				
30701		129.73	9.5	13.0	6828.	8.5	-3.9	39.6	14800.
2.9	54.2	8.3	4.8	89.8	5778.	17.5	15.9		
1940.	4075.	378.	41.3	18.3	51.3	14.3	9.15		
5.96	1.	1.	1.	1.	0.				
30801		114.12	9.1	11.5	15428.	22.6	-10.8	59.	8600.
3.2	53.9	14.2	5.7	91.1	5840.	16.8	16.5		
1729.	1751.	281.5	40.3	15.1	37.1	33.7	4.58		
5.66	1.	1.	1.	0.	1.				
30901		130.01	10.6	9.2	3542.	33.2	-0.8	47.3	8300.
3.2	49.5	14.2	10.0	84.8	4438.	31.5	9.5		
2055.	3620.	300.	37.4	18.0	36.3	28.3	6.6		
4.32	1.	1.	1.	0.	0.				
31001		121.9	11.7	10.9	3002.	23.2	-5.9	64.4	8200.
3.1	59.1	17.4	11.1	89.1	5583.	18.1	10.2		
1187.	3600.	178.5	39.1	19.2	38.1	26.6	14.6		
4.22	1.	1.	1.	0.	1.				
40101		166.69	10.8	10.9	5021.	19.1	-1.0	66.8	11300.
3.4	59.3	18.0	9.2	92.2	5749.	16.2	12.9		
1473.	1660.	218.	35.2	18.4	35.5	40.1	5.93		
5.08	1.	1.	1.	0.	0.				
40201		191.67	9.8	11.0	4250.	12.7	39.3	57.7	13700.
2.8	50.7	25.8	14.4	86.6	6943.	11.9	23.8		
2419.	3220.	437.5	41.5	18.8	54.6	12.9	8.78		
7.28	1.	1.	1.	0.	1.				
40301		152.05	9.8	11.7	8494.	14.9	-7.5	33.8	16200.
2.9	53.1	13.8	5.3	87.7	5864.	16.6	15.4		
1528.	3230.	368.2	41.4	16.3	42.4	31.5	7.94		
3.97	1.	1.	1.	1.	1.				

40401		201.19	14.1	4.9	3580.	2.9	107.8	69.1	13800.
3.5	55.3	33.7	19.9	90.8	6621.	13.6	21.8		
1804.	1660.	263.	35.5	20.9	59.1	9.4	7.61		
8.88	1.	1.	1.	0.	1.				
40501		174.46	11.1	11.8	5271.	7.7	7.4	50.2	12100.
3.1	61.6	12.2	4.8	90.4	5994.	14.6	13.2		
1629.	1950.	268.	38.5	19.8	44.7	33.9	6.12		
4.17	1.	1.	1.	0.	1.				
40601		186.59	9.4	13.6	5043.	1.2	-8.3	44.5	12900.
3.1	59.6	11.1	5.0	91.5	5804.	15.4	13.9		
1668.	2242.	238.	38.9	18.6	44.4	37.9	5.74		
3.76	1.	1.	1.	0.	1.				
40701		204.20	13.5	7.6	3747.	3.3	114.3	63.5	16300.
3.2	55.0	31.7	15.7	89.2	6949.	11.5	20.8		
2253.	2470.	369.	3605.	17.9	49.7	29.0	5.38		
8.24	1.	1.	1.	1.	1.				
40801		201.56	12.4	5.8	3111.	28.0	50.4	53.3	12300.
3.4	48.3	26.7	9.5	87.2	5592.	21.0	16.2		
1951.	13320.	472.	41.9	23.2	46.7	21.5	8.68		
5.68	1.	1.	1.	0.	1.				
50101		372.68	8.7	14.2	5546.	5.6	-0.3	61.5	10800.
2.7	55.2	19.6	18.3	85.0	6335.	15.1	16.9		
1845.	8050.	453.	40.3	23.9	51.3	18.8	11.0		
4.94	1.	1.	1.	1.	1.				
50201		405.22	10.9	9.0	17170.	34.4	-7.6	22.5	13500.
3.1	47.9	8.9	3.5	90.6	5454.	18.9	12.1		
1158.	4610.	419.	40.0	16.0	31.	36.2	6.91		
3.72	1.	1.	1.	1.	1.				
50301		291.69	8.2	12.7	8529.	22.6	17.0	42.3	13100.
2.6	52.5	15.0	5.0	80.3	4450.	29.7	9.2		
2245.	3810.	702.	43.3	22.3	40.2	12.5	8.8		
5.64	1.	1.	1.	1.	1.				
50401		340.89	11.3	8.6	4576.	39.7	4.6	53.25	9700.
3.3	49.2	19.2	9.9	85.3	4947.	27.0	10.4		
1664.	3360.	248.	36.6	21.4	42.1	24.8	8.32		
4.87	1.	1.	1.	0.	1.				
50501		476.26	12.2	9.9	6689.	20.7	11.5	55.4	10800.
3.1	57.7	17.1	9.8	89.9	6106.	15.6	16.2		
1924.	4860.	353.	40.75	19.4	44.1	29.6	7.26		
4.69	1.	1.	1.	0.	1.				

50601		487.46	10.9	7.9	3802.	38.3	47.1	45.5	12000.
3.2	47.8	20.1	10.3	87.8	5029.	26.6	15.8		
2084.	8950.	564.	40.4	21.8	44.1	17.9	8.67		
5.69	1.	1.	1.	0.	1.				
50701		313.41	11.1	11.9	6004.	3.0	0.7	60.3	13800.
3.1	60.0	16.8	16.5	90.6	6543.	11.8	17.7		
1560.	3890.	282.	40.3	19.0	49.3	25.2	10.5		
5.0	1.	0.	1.	0.	1.				
50801		482.87	9.5	13.4	8546.	3.2	-7.4	52.6	13700.
2.8	56.6	16.3	14.2	87.7	6401.	13.9	17.3		
1782.	8170.	712.5	43.9	22.0	50.6	24.4	7.97		
4.3	1.	1.	1.	0.	1.				
50901		497.52	12.3	7.8	3881.	37.1	25.6	55.3	10300.
3.4	50.9	17.8	13.0	85.4	4915.	27.0	11.2		
1512.	6750.	297.	37.4	23.1	43.2	20.6	9.0		
5.92	1.	1.	1.	0.	1.				
51001		493.89	10.8	10.8	6956.	7.1	18.8	53.4	13200.
2.9	50.9	26.6	14.7	86.9	6361.	15.1	18.8		
1735.	5450.	440.	39.8	21.8	52.6	18.0	8.31		
5.35	1.	1.	1.	0.	1.				
60101		3550.4	10.7	9.8	15836.	23.6	-1.9	34.4	18000.
3.0	51.6	8.5	6.5	88.9	6738.	13.6	21.3		
1586.	4930.	623.	42.3	17.9	41.7	33.5	8.7		
3.8	1.	1.	1.	1.	1.				
60201		573.22	11.1	7.6	2979.	7.8	71.4	52.6	16700.
3.0	56.7	25.7	14.3	89.0	6614.	14.4	20.9		
1386.	1130.	308.	31.4	18.6	51.9	25.2	4.95		
6.7	1.	1.	1.	1.	1.				
60301		587.72	13.5	7.1	3662.	7.4	43.9	63.7	9000.
3.6	55.8	23.	14.9	84.1	4691.	27.9	9.6		
1235.	1619.5	216.	31.6	24.0	45.5	11.3	6.56		
7.35	1.	1.	1.	0.	1.				
60401		2479.02	10.1	10.2	5451.	16.8	25.8	46.2	17300.
2.8	49.2	28.5	9.6	87.4	6896.	14.4	25.1		
1801.	3815.	618.	41.	18.8	50.3	26.8	6.0		
4.65	1.	1.	1.	1.	1.				
60501		502.55	11.2	11.7	6501.	21.8	-0.3	40.4	15100.
3.0	49.6	14.0	7.3	88.5	5701.	19.6	15.8		
1591.	6275.	425.	37.8	18.0	42.0	29.4	7.05		
5.0	1.	1.	1.	0.	1.				

60601		627.53	11.4	8.6	3157.	37.4	10.0	37.5	16000.
3.2	47.6	12.2	11.7	86.0	4807.	27.8	12.9		
1276.	3350.	294.	35.6	22.4	44.1	13.6	12.1		
5.9	1.	1.	1.	1.	1.				
60701		7781.9	88.8	10.5	24697.	14.7	-1.4	21.8	17000.
2.9	38.3	4.1	2.5	88.1	6091.	15.2	18.5		
1348.	6800.	903.	42.5	18.9	48.1	26.3	8.68		
3.75	1.	1.	1.	1.	1.				
60801		697.2	9.5	12.3	14586.	9.8	-13.0	27.2	13500.
2.9	45.5	6.8	2.4	86.6	5747.	16.7	13.6		
1778.	6210.	380.	41.1	18.6	43.8	24.4	7.91		
4.31	1.	1.	1.	1.	1.				
60901		939.02	10.9	9.0	11886.	35.0	-1.1	54.4	9000.
3.3	48.9	10.1	6.0	91.0	5659.	18.6	15.0		
1402.	2860.	315.5	38.6	18.5	40.3	28.0	8.15		
5.0	1.	1.	1.	1.	1.				
61001		876.05	11.5	9.9	10789.	28.9	-4.2	44.8	13900.
3.2	54.8	13.6	6.6	90.6	5935.	17.2	13.0		
1459.	5110.	443.	38.6	16.3	32.9	40.8	7.31		
3.7	1.	1.	1.	1.	1.				
61101		1670.14	10.3	9.5	11964.	29.2	-9.7	58.2	12000.
3.2	54.5	17.7	7.1	92.2	6069.	19.0	17.8		
1379.	4080.	508.	36.7	18.3	40.2	37.4	5.88		
3.3	1.	1.	1.	1.	1.				
61201		938.22	13.0	5.6	2860.	23.2	57.4	60.3	10900.
3.3	52.7	28.3	19.3	87.6	5902.	18.8	17.5		
1722.	4360.	381.8	38.7	22.6	47.0	19.6	9.46		
9.4	1.	1.	1.	1.	1.				
61301		763.96	10.2	9.1	12442.	54.8	-4.8	29.7	15400.
2.9	43.7	9.0	7.6	83.5	5993.	17.3	21.7		
1855.	2690.	564.	44.8	14.0	50.0	6.2	8.5		
4.97	1.	1.	1.	1.	1.				
61401		750.03	11.0	12.3	12296.	28.8	-12.5	38.1	12000.
2.9	52.8	8.0	6.4	84.8	5355.	21.7	10.8		
1424.	5480.	467.	39.2	17.3	37.6	31.2	8.6		
3.47	1.	1.	1.	0.	1.				

61501		741.32	11.8	9.6	8137.	8.9	16.3	48.2	15100.
3.2	60.2	11.6	11.5	91.7	6664.	11.4	16.7		
1452.	3235.	304.	40.75	18.4	40.1	40.6	6.65		
3.81	1.	1.	1.	1.	1.				
61601		740.32	7.9	12.6	15553.	18.4	-4.5	35.0	17300.
2.4	46.8	11.1	6.7	77.4	6717.	13.5	22.6		
1990.	7140.	815.	44.7	20.5	52.4	16.4	10.35		
4.23	1.	1.	1.	1.	1.				
61701		679.68	11.8	7.0	2428.	19.3	56.4	59.7	11300.
3.2	52.0	29.8	12.7	89.0	5976.	18.4	18.9		
1895.	7950.	512.	42.2	23.1	50.3	20.4	8.13		
6.94	1.	1.	1.	0.	1.				
61801		604.33	9.6	13.0	11171.	0.3	-10.7	48.9	11000.
3.1	51.1	9.6	5.5	90.4	5605.	18.4	14.3		
1621.	4970.	565.	36.8	20.4	43.6	26.1	8.11		
4.96	1.	1.	1.	0.	1.				
61901		557.09	9.3	12.0	6295.	8.4	19.1	57.5	13500.
2.7	53.0	20.5	16.9	84.1	6942.	11.8	22.9		
1811.	5090.	396.	41.3	21.1	54.7	24.1	8.79		
4.7	1.	1.	1.	1.	1.				
62001		532.76	10.3	11.6	13522.	13.8	-8.2	44.3	11700.
3.1	56.7	9.4	5.6	92.4	5713.	17.3	13.1		
1317.	3955.	271.	37.2	18.7	39.3	35.9	8.24		
4.04	1.	1.	1.	1.	1.				
62101		2002.5	110.0	10.4	15743.	26.7	-3.3	62.0	8700.
3.2	48.1	7.9	6.7	91.5	5782.	17.1	14.2		
1243.	3050.	351.	39.4	18.9	40.8	33.2	6.73		
4.22	1.	1.	1.	1.	1.				
70101		55.15	12.7	8.71	4055.	2.54	8.6	9.38	10346.
3.7	54.5	2.53	2.17	90.3	4138.	26.4	1.95		
1441.	1621.	194.2	34.6	20.9	40.8	19.9	14.67		
5.24	1.	1.	1.	1.	1.				
70201		53.53	11.3	8.8	4225.	4.85	60.0	40.5	14190.
3.5	66.9	7.91	9.64	89.95	5144.	14.61	4.81		
1378.	718.	206.1	36.3	12.89	37.6	22.7	5.68		
5.15	0.	0.	1.	0.	1.				

70301		56.93	15.6	4.74	8626.	0.35	30.9	30.3	12980.
4.2	60.1	2.55	8.5	94.	4836.	15.68	3.2		
792.	295.	144.2	34.6	13.7	40.6	18.3	7.1		
9.18	1.	0.	1.	0.	0.				
70401		50.98	15.1	6.08	2870.	3.14	46.9	69.25	11812.
3.7	73.	10.37	13.71	92.7	5804.	11.1	6.18		
1170.	498.	151.8	33.8	13.39	40.0	38.35	8.72		
8.7	1.	0.	1.	1.	1.				
70501		54.94	10.2	20.39	7578.	6.37	7.0	48.1	10197.
2.8	52.1	5.32	8.28	76.3	4826.	16.28	2.87		
2170.	1085.	386.	22.75	22.4	60.0	17.6	12.0		
6.4	1.	1.	1.	1.	1.				
70601		55.2	12.7	9.79	3206.	4.0	50.3	62.7	10663.
3.4	66.3	8.42	9.43	92.1	4832.	15.1	3.22		
1145.	557.	160.	35.9	14.15	37.9	14.65	5.55		
3.54	1.	0.	1.	0.	0.				
70701		55.98	18.8	3.75	18536.	0.72	431.2	10.55	13828.
4.2	56.6	1.3	8.44	94.	4454.	16.05	1.97		
735.	309.	64.6	32.7	14.21	26.8	36.6	10.94		
15.3	0.	0.	0.	0.	0.				
70801		74.49	13.35	6.9	4194.	2.96	66.0	62.2	12985.
3.6	66.8	9.95	12.75	90.25	5318.	9.87	4.05		
1300.	1010.	163.8	40.25	15.67	37.7	45.4	5.0		
6.99	1.	1.	1.	0.	0.				
70901		62.42	13.6	4.82	2936.	4.03	50.2	71.2	12770.
3.6	71.5	9.55	12.3	92.5	5348.	8.26	4.16		
1385.	457.	148.2	36.4	12.31	30.8	54.6	5.72		
3.96	1.	0.	1.	1.	0.				
71001		66.55	15.2	6.45	3003.	0.38	31.7	20.6	12983.
4.0	58.4	2.69	6.08	9.02	4239.	24.42	2.6		
1180.	1123.	176.5	31.4	15.8	38.8	31.1	6.22		
6.7	0.	0.	1.	0.	0.				
71101		53.48	13.1	5.99	7140.	0.3	16.1	14.45	13789.
4.1	50.5	1.48	3.91	92.	4611.	21.7	4.34		
1089.	920.	176.5	31.6	14.45	39.2	35.5	9.46		
5.92	0.	0.	1.	1.	0.				

71201		63.63	14.15	6.6	12881.	1.75	20.4	36.5	15316.
4.6	51.75	7.36	4.64	91.4	4669.	22.35	4.57		
1495.	2090.	194.	31.4	23.	51.5	9.0	15.0		
5.	1.	1.	1.	1.	1.				
80101		92.51	9.44	7.47	13446.	4.7	8.1	34.9	16947.
3.9	53.	5.68	4.55	89.8	4930.	17.21	5.23		
1495.	1850.	276.	42.2	16.15	43.8	9.24	11.26		
3.08	1.	1.	1.	1.	1.				
80201		84.47	11.85	7.69	3301.	2.13	122.4	67.4	11312.
3.6	69.4	11.07	11.45	92.4	5186.	10.9	4.03		
1101.	429.	163.2	33.6	13.75	37.7	44.75	7.05		
5.64	1.	0.	1.	1.	0.				
80301		95.53	13.6	9.01	3256.	1.99	79.3	63.6	11752.
3.5	66.4	7.4	24.6	84.95	5033.	13.41	4.38		
1164.	2182.	210.5	34.0	22.18	48.	11.19	12.6		
7.39	0.	0.	1.	0.	1.				
80401		80.12	15.0	3.5	1722.	1.48	88.9	42.75	14492.
4.0	68.1	5.54	11.02	92.9	5482.	7.84	4.55		
1400.	1017.	194.6	34.7	14.75	32.	13.68	6.95		
5.4	1.	1.	1.	0.	1.				
80501		78.32	11.52	7.55	23877.	0.77	1.2	4.54	16030.
3.5	48.4	1.59	3.23	95.4	5131.	9.55	4.02		
844.	210.	110.	37.9	15.5	43.1	35.7	16.5		
3.65	0.	0.	0.	0.	0.				
80601		169.57	12.4	8.72	2572.	3.35	77.9	57.5	13071.
3.4	64.6	11.12	11.13	93.	5348.	11.55	5.11		
1405.	1250.	212.	37.52	17.2	45.5	27.21	8.45		
4.84	1.	1.	1.	0.	1.				
80701		112.14	14.23	7.4	4360.	2.23	57.2	60.4	12190.
3.6	66.5	9.67	19.6	90.	5433.	11.11	5.42		
1180.	2150.	238.	38.4	23.13	53.25	10.7	13.02		
7.21	0.	0.	1.	0.	1.				
80801		114.37	11.32	10.8	8859.	4.11	-4.7	54.8	9365.
3.4	57.9	6.2	6.2	94.	4797.	17.65	3.69		
1090.	874.	241.	34.1	14.35	37.9	36.4	7.95		
5.64	1.	1.	1.	0.	1.				

80901		249.64	14.03	7.38	3448.	2.96	93.4	53.1	15252.
3.4	65.25	11.59	18.15	87.	5616.	10.69	7.06		
1320.	2215.	254.	36.3	21.5	50.75	12.25	11.5		
8.38	1.	1.	1.	0.	1.				
81001		171.98	10.47	7.4	19130.	0.35	4.9	6.99	15248.
3.8	44.2	2.17	5.31	92.	4638.	21.0	4.58		
1380.	2175.	268.	37.2	16.25	45.4	16.59	9.38		
5.64	1.	1.	1.	1.	1.				
81101		281.03	13.5	6.37	4869.	2.78	76.	55.5	14692.
3.6	64.5	10.32	22.75	86.4	5400.	12.41	5.96		
1271.	1824.	257.	36.9	21.	48.1	13.88	11.55		
8.18	1.	1.	1.	0.	1.				
81201		273.99	11.68	8.75	5754.	3.44	31.5	60.6	13402.
3.6	62.75	8.25	8.3	91.6	5198.	12.78	3.98		
1145.	1150.	203.	36.9	14.75	35.43	44.4	6.34		
6.44	1.	1.	1.	1.	0.				
81301		268.21	11.29	7.95	6306.	4.09	32.7	40.	17655.
3.7	62.8	9.38	11.12	93.	6228.	9.8	11.58		
1287.	1055.	244.	39.0	12.25	58.	8.41	6.99		
5.18	1.	1.	1.	0.	1.				
81401		384.52	8.32	13.85	8874.	6.91	11.5	55.	13783.
3.1	53.6	9.05	6.34	84.1	5366.	14.32	6.87		
1360.	3620.	367.	36.95	19.56	45.4	19.7	13.01		
5.28	1.	1.	1.	1.	1.				
81501		265.43	10.38	10.95	10803.	2.45	12.6	51.1	12220.
3.4	52.25	6.88	9.77	88.25	5010.	15.61	4.88		
1510.	11420.	301.	40.25	20.0	43.	21.3	14.05		
5.91	1.	1.	1.	0.	1.				
81601		1191.06	11.73	10.61	25256.	1.24	16.6	42.5	15750.
3.5	44.2	2.01	5.03	90.	4934.	16.98	5.19		
1183.	2022.	318.	38.3	14.85	40.2	33.15	11.15		
6.36	1.	1.	1.	1.	1.				
81701		672.41	10.4	11.0	19234.	4.83	-0.5	25.0	17523.
3.7	46.7	6.85	3.24	86.4	4972.	19.51	5.71		
1705.	2790.	543.	43.6	16.25	41.9	18.85	9.14		
6.52	1.	1.	1.	1.	1.				

APPENDIX B.2

INPUT DATA FOR GROUPING PRCGRAMME - FACTOR SCORES OF 100 NORTH AMERICAN CITIES

(I3,3X,11F6.0)

FACTOR NAMES:

ID	SEIND	BSIND	LSIND	HSIND	FDIND	PIND	PCIND	GF-I	GF-II	GF-III	POVIND
101	-0.673	-0.221	0.593	1.122	-2.285	-0.052	-0.168	0.163	-0.458	0.987	-0.938
102	0.595	1.105	1.240	0.571	0.159	0.208	-0.149	0.079	-0.069	0.817	0.589
103	0.829	0.664	-0.276	-0.076	0.441	0.400	-0.257	0.887	0.935	1.695	0.662
104	0.262	0.734	0.819	0.082	-0.524	0.077	0.773	0.436	-1.142	-0.399	0.959
105	-0.169	-0.902	0.719	1.013	0.451	-0.508	0.972	0.369	0.731	-1.908	1.504
106	1.603	0.099	-0.036	1.108	-0.958	-0.346	0.509	1.118	0.285	0.698	0.579
107	0.762	1.509	1.013	0.714	1.652	-0.475	0.498	-0.101	-0.323	-0.399	-0.882
108	-0.390	0.515	0.476	-0.115	-0.421	-0.341	0.634	0.315	-0.290	-0.036	0.245
109	-0.814	-0.346	0.550	-0.785	0.052	-1.502	0.079	-0.139	-1.124	-0.020	-0.937
110	-0.987	0.186	0.964	-1.352	0.053	-0.778	0.217	-0.244	-0.367	0.181	0.448
111	3.219	-2.207	-0.929	-0.433	-0.054	-1.888	0.252	0.851	-0.077	0.546	0.008
112	-0.617	1.195	-1.009	0.391	-1.375	-0.087	0.649	0.174	0.475	1.026	-1.632
201	0.386	0.600	-0.365	-1.280	-0.438	-0.577	0.256	-0.127	-0.080	-0.131	0.615
202	0.831	-0.496	-0.052	-0.276	-1.236	0.116	0.298	-4.187	1.280	0.533	0.603
203	-0.775	-1.759	1.256	-0.218	0.137	-0.408	0.699	-0.584	0.101	-0.455	1.054
204	0.311	0.188	0.570	0.601	-1.446	0.132	0.629	0.307	0.557	-1.761	1.035
205	-0.801	0.443	0.149	-0.342	1.875	0.142	0.175	0.156	0.457	0.749	-0.447
206	1.986	1.013	1.356	0.941	-0.284	0.026	0.230	-4.128	0.575	0.228	-0.650
207	0.077	0.352	0.808	-0.500	0.163	-0.305	0.530	0.049	0.294	0.021	0.405
208	-1.226	-0.273	0.954	-1.113	1.083	0.184	-0.742	-0.302	0.133	0.716	0.591
209	-1.358	0.230	-1.183	-0.863	-2.073	-0.928	-0.264	-0.264	0.800	-1.438	0.700
210	-0.422	-1.850	0.072	-0.906	0.541	-1.166	-2.376	0.051	-0.210	0.996	-0.815
301	-0.266	0.579	0.775	-0.837	0.568	0.509	-0.587	0.130	0.200	-0.718	0.213
302	-1.334	0.336	-0.516	-1.327	0.419	0.194	0.191	-0.059	1.236	0.687	0.988
303	0.099	0.567	-0.414	-1.010	-0.304	-0.505	0.543	-0.174	-0.134	-1.657	0.794
304	1.338	0.726	-2.210	-1.357	-0.387	-0.288	0.696	-0.124	-1.520	0.688	-0.730

305	-0.848	0.298	0.838	-0.482	-0.425	-0.250	0.401	-0.374	-0.573	-0.116	0.022
306	-0.797	0.447	0.188	1.558	0.724	-1.067	0.351	-0.169	-0.544	-1.770	-2.180
307	0.437	-0.050	-1.269	-0.886	0.991	-1.293	0.086	-0.363	-0.175	-1.732	-0.104
308	-0.289	1.190	-0.943	-0.964	-0.443	0.804	-0.622	0.263	1.046	1.257	-0.994
309	-1.149	-0.317	0.063	0.158	-0.733	-0.976	0.551	0.260	-0.650	0.105	-1.837
310	-1.026	0.519	0.453	-0.052	0.933	1.037	0.556	0.580	1.008	0.922	-0.058
401	-0.572	0.263	0.320	0.086	-1.065	-0.116	0.350	0.225	0.548	0.134	-0.074
402	1.232	0.461	0.392	-0.972	1.191	0.369	-1.143	0.141	-0.318	0.805	-0.839
403	0.266	0.208	-1.361	-0.750	-0.001	0.335	-0.145	0.227	0.340	-0.961	-0.257
404	1.064	0.541	1.294	1.659	1.634	0.229	-1.457	0.071	0.080	0.211	-0.774
405	-0.398	0.023	-0.442	-0.721	-0.291	0.455	-0.215	0.410	0.621	0.547	0.861
406	-0.398	0.168	-0.772	-1.336	-0.404	0.310	-0.202	0.210	0.678	0.538	1.025
407	1.276	1.341	0.898	-0.186	-1.674	0.406	-6.719	0.734	-0.172	-2.770	-0.202
408	-0.509	-3.596	0.746	1.466	0.078	0.414	0.103	-0.194	0.603	0.280	-1.203
501	0.133	-1.045	0.923	-1.465	0.888	1.293	0.134	-0.282	-0.491	-0.879	0.629
502	-0.345	-0.355	-2.046	0.352	-0.597	1.109	0.266	0.296	1.021	-0.939	-1.382
503	-0.863	-1.316	-0.805	-1.616	0.385	-0.182	-0.954	-0.071	-0.815	-0.656	-0.765
504	-0.968	-0.141	0.344	0.416	0.268	0.380	0.239	0.329	0.058	0.584	-1.887
505	-0.108	-0.651	0.177	0.038	-0.210	1.078	0.042	0.601	0.495	0.633	-0.146
506	-0.659	-2.704	0.175	0.663	0.208	0.331	-0.231	0.161	0.344	0.563	-1.658
507	0.689	0.483	0.060	-0.273	0.800	1.230	0.236	0.617	0.279	1.304	0.774
508	0.379	-1.784	0.213	-1.359	0.163	1.660	0.172	0.045	-0.330	0.377	0.822
509	-1.103	-1.454	0.653	1.117	0.440	0.668	0.230	0.318	0.084	0.416	-1.399
510	0.619	-0.301	0.689	-0.617	0.769	1.246	-0.180	-0.255	-0.518	0.241	-0.495
601	1.215	-1.334	-0.910	0.224	-0.731	4.737	1.079	0.420	-0.790	-1.614	-0.373
602	0.896	0.263	0.380	0.301	-0.012	0.757	-1.258	0.079	-0.136	-1.077	-0.149
603	-1.082	0.057	1.102	1.205	0.829	0.498	-0.527	0.108	-0.706	0.039	-0.680
604	1.492	-0.931	0.525	-0.460	-0.732	3.121	-0.107	-0.115	-1.157	-1.445	-0.864
605	-0.163	-1.021	-0.654	-0.338	-0.216	0.874	-0.105	0.282	0.252	0.410	-0.585
606	-0.488	-0.474	-0.460	0.918	1.471	0.124	0.301	0.336	-0.008	-1.380	-1.614
607	1.564	-2.319	-0.974	-0.254	-0.938	9.598	1.885	-0.846	-3.038	-2.628	-0.983
608	0.109	-0.339	-1.810	-0.964	0.404	0.899	0.130	-0.383	-0.070	-1.246	-0.519
609	-0.345	0.275	-0.663	0.379	0.148	1.470	0.281	0.192	0.794	-0.802	-1.480
610	-0.241	-0.974	-0.851	0.438	-1.077	1.698	0.355	0.676	0.870	-0.865	-0.399
611	0.091	-0.783	-0.061	0.054	-1.111	2.695	0.639	0.215	0.466	-0.970	-0.834
612	0.016	-0.960	1.290	1.600	0.747	1.150	-0.945	0.191	-0.071	-1.066	-1.106

613	1.395	0.720	-1.723	0.076	1.168	0.807	0.192	0.900	-0.508	-0.798	-2.982
614	-0.590	-0.557	-0.989	-0.621	-0.324	1.717	0.467	0.519	-0.145	0.646	-0.575
615	0.324	-0.387	-0.408	0.400	-0.732	1.570	0.218	0.705	0.651	-1.131	0.890
616	1.455	-0.938	-1.184	-1.827	0.812	1.736	0.185	-0.261	-1.543	-1.026	-0.744
617	-0.059	-2.159	1.201	0.699	0.265	1.030	-0.574	-0.202	0.180	0.389	-0.847
618	-0.406	-0.684	-0.830	-1.190	0.351	1.190	-0.096	-0.280	0.282	0.372	0.215
619	1.069	0.010	0.670	-1.115	0.517	1.509	-0.067	-0.130	-0.822	-0.951	0.043
620	-0.447	-0.010	-1.111	-0.438	-0.095	1.152	0.224	0.008	1.038	-0.909	0.196
621	-0.262	0.250	-0.384	-0.201	-0.446	3.190	0.545	-0.162	0.237	-0.943	-0.998
701	-1.560	-0.082	-2.177	1.210	2.071	-1.304	0.232	0.221	0.993	-1.706	0.172
702	-1.045	-0.479	-2.454	1.039	-0.247	0.412	-1.344	0.184	-1.837	2.599	3.199
703	-0.361	0.932	-2.189	3.209	0.961	-1.282	-0.185	0.890	0.689	0.030	0.122
704	-0.446	0.536	-0.548	2.259	-0.046	0.434	-1.135	1.644	1.513	0.255	1.748
705	-0.511	1.272	-0.822	-2.424	2.112	-0.108	-0.804	-0.302	-1.732	-0.906	0.709
706	-0.649	0.838	-0.596	0.811	-0.430	-0.560	0.222	0.966	0.190	0.828	1.018
707	-2.719	-2.016	-4.017	4.991	0.531	-0.876	-5.611	-5.613	-0.934	1.875	2.931
708	-0.772	0.054	-0.132	1.525	-1.170	-0.176	-0.601	0.772	0.373	0.061	1.440
709	-0.452	0.214	-0.316	2.126	-2.181	0.064	0.220	1.457	0.875	-0.382	1.779
710	-0.794	1.013	-1.479	2.239	-1.376	0.597	0.337	0.503	-12.68	1.068	1.770
711	-1.056	-0.266	-3.544	2.694	0.442	-1.158	0.073	-1.118	-2.125	-0.607	1.608
712	-0.327	0.389	-2.018	2.824	3.392	-1.177	1.088	0.035	1.119	-2.312	-0.590
801	0.114	0.890	-2.560	1.011	2.003	-0.583	0.673	0.426	0.950	-1.467	-0.452
802	-0.767	0.213	-0.459	1.518	-1.297	-0.207	-0.878	0.509	0.770	-0.191	1.808
803	-1.181	-0.681	-0.540	1.708	1.826	1.242	-0.882	-0.080	-2.896	2.249	3.656
804	-0.539	-0.412	-1.245	2.960	0.312	-0.213	-0.537	2.097	1.499	0.293	1.492
805	-0.374	1.410	-4.416	0.967	2.199	0.104	0.975	-5.589	-1.520	0.697	1.244
806	-0.648	0.120	-0.375	0.701	0.397	0.384	-0.714	0.812	1.062	0.620	1.382
807	-0.817	-0.747	-0.798	2.006	2.218	1.046	-0.516	-0.309	-2.206	1.953	3.396
808	-1.260	1.044	-1.236	-0.147	0.166	0.576	-0.522	0.460	1.380	0.801	0.493
809	-0.216	-0.214	0.006	1.532	1.850	0.462	-0.995	0.884	0.119	0.271	1.368
810	-0.059	1.027	-3.347	1.081	2.094	-0.630	-0.007	-0.221	0.799	-1.795	-1.116
811	-0.262	0.063	0.088	1.848	1.856	0.666	-0.648	1.115	0.007	0.256	1.144
812	-0.791	0.225	-0.770	1.045	-0.982	-0.252	-0.294	0.361	0.657	-1.451	1.097
813	1.080	1.231	-1.620	0.989	1.664	-0.052	-0.194	1.108	0.760	0.193	0.381
814	-0.425	0.350	-1.067	-1.090	1.438	0.747	-0.201	-0.223	0.053	-0.821	0.513
815	-1.030	-1.682	-0.939	0.247	1.522	0.873	0.331	-0.193	0.936	0.425	0.694
816	-0.075	1.462	-2.596	0.454	1.269	1.905	-0.222	-0.415	0.403	-1.350	-0.609
817	0.142	0.238	-2.876	0.394	1.376	0.265	-0.361	0.004	0.027	-1.535	-0.772

END

APPENDIX B.3

MAXIMUM SIMILARITY GROUPING PROGRAMME

```
C PROGRAMME GROUPING
C
C GROUPING CRITERION: MINIMUM SQUARED DISTANCE
C TESTING CRITERIA: INCREMENT TO POOLED WITHIN-GROUPS DEVIATION MATRIX
C                   AND F-RATIO TO TEST THE HYPOTHESIS THAT C1=C2=...=CG
C                   OF G GROUPS (IF KF = 0)
C                   AND/OR AVERAGE INTRAGROUP DISTANCES AND BACHI'S
C                   STANDARD DISTANCES  (IF KD = 0)
C
C PARAMETER CONTROL CARD:
C     COL 1-5 NS = # OF OBSERVATIONS (MAX = 100)
C     COL 6-7 NV = # OF VARIABLES (MAX = 15)
C     COL 8-12 LEVEL = LEVEL TO TERMINATE GROUPING; THAT IS THE NUMBER
C                      OF GROUPS REQUIRED AT THE FINAL STAGE OF GROUPING
C                      LEVEL SHOULD NOT GREATER THAN NS OR LESS THAN 1
C     COL 13  KW = 0 ROW VECTOR TO WEIGH THE VARIATES IS READ IN
C                    ACCORDING TO SPECIFIC FORMAT,OTHERWISE SET KW = 1
C     COL 14  KM = 0 MATRIX OF GROUP MEMBERSHIPS AT THE FINAL STAGE OF
C                    GROUPING IS REQUIRED, OTHERWISE SET KM = 1
C     COL 15 KC = 0 LOWER TRIANGULAR CONTIGUITY MATRIX (EXCLUDING
C                   ELEMENTS ON PRINCIPAL DIAGONAL) ARRANGED LOW BY LOW
C                   ACCORDING TO FORMAT (50F1.0/(50F1.0)) IS READ IN
C                   OTHERWISE SET KC = 1
C     COL 16 KS = 0 RAW DATA HAVE TO BE STANDARDIZED
C                   OTHERWISE SET KS = 1
C     COL 17 KF = 0 F-RATIO IS CALCULATED, OTHERWISE SET KF = 1
C     COL 18 KD = 0 BACHI'S SQUARED STANDARD DISTANCE IS CALCULATED,
C                   OTHERWISE SET KD = 1
C
C CARDS SET UP:
C     #1    $SIGNON CARD
C     #2    PASSWARD
C     #3    $RUN (NAME OF PROGRAM) (DEFINING LOGICAL UNITS)
```

```
C       #4     TITLE CARD OF MAX 80 CHARACTERS
C       #5     PARAMETER CONTROL CARD
C       #6     DATA FORMAT CARD NOT TO EXCEED 80 COLUMNS
C       #7  ) DATA CARDS OF NS BY NV MATRIX ARRANGED LOW-WISE ACCORDING TO
C       #7+ ) THE SPECIFIC FORMAT, NOTE THAT IDENTIFICATIONS (FORMAT I4)
C             SHOULD BE PUT BEFORE EACH SET OF DATA
C       OPTIONAL: IF KW = 0, THE SUBSEQUENT CARDS ARE:
C       (1) F-FORMAT CARD OF THE ROW VECTOR TO WEIGH THE VARIATES (NOT TO
C           EXCEED 80 CHARACTERS)
C       (2) CARDS CONTAINING THIS VECTOR
C       OPTIONAL: IF KC = 0, THE SUBSEQUENT CARDS ARE:
C       DATA CARDS CONTAINING THE CONTIGUITY MATRIX,FOR EXAMPLE, THE 1 ST
C       CARD CONTAINS (2,1), THE 2 ND CARD CONTAINS (3,1),(3,2), ETC.
C       NOTE THAT 0'S DENOTE ADJACENT AREAS, 1'S OTHERWISE
C       #N-1   $ENDFILE
C       #N     $SIG
C
C THE PROGRAMME READS RAW DATA OR STANDARDIZED DATA FROM UNIT 4
C AND OTHERS FROM UNIT 5, AND WRITES ALL OUTPUT ON UNIT 6
C
C THE OUTPUT GIVES:
C (1) GROUP FORMED AT EACH STAGE AND THE MINIMUM DISTANCE
C (2) CENTROIDS OF THE GROUPS
C (3) INCREMENT, RATIO AND RATE OF CHANGE OF THE INCREMENT OF POOLED
C       WITHIN-GROUPS DEVIATION MATRIX (IF KF = 0)
C (4) F-RATIO AND CORRESPONDING DEGREES OF FREEDOM (IF KF = 0)
C (5) AVERAGE INTRAGROUP DISTANCES AND BACHI'S SQUARED STANDARD
C       DISTANCES (IF KD=0)
C
C SUBROUTINES REQUIRED IN THIS PROGRAMME ARE DET AND STANDA
C
      DIMENSION GPDIST(100),GPSTDD(100)
      DIMENSION F(100,15),SC(100,15),NG(100,100),WD(15,15),TD(15,15),
     1DD(100,100),C(100,15),GPM(100),IFMT(20),KTITLE(20),ID(100),
     2IFMT2(20),WEIGHT(15)
      EQUIVALENCE (GPM(1),ID(1))
      EQUIVALENCE (KTITLE(1),IFMT(1),IFMT2(1))
      EQUIVALENCE (TD(1),WD(1)),(GPM(1),ID(1))
      READ(5,1) KTITLE
```

```
      WRITE(6,15) KTITLE
1     FORMAT (20A4)
      READ(5,2) NS,NV,LEVEL,KW,KM,KC,KS,KF,KD,IFMT
2     FORMAT (I5,I2,I5,I1,I1,I1,I1,I1,I1,/20A4)
      WRITE(6,6) NS,NV,LEVEL
6     FORMAT ('NUMBER OF OBSERVATIONS TO BE GROUPED='I5/'NUMBER OF VARIA
     1BLES ='I2/'LEVEL TO TERMINATE GROUPING ='I5)
      DO 7 I=1,NS
7     READ(4,IFMT) ID(I),(F(I,J),J=1,NV)
      IF(KW) 20,8,20
8     READ(5,1) IFMT2
      READ(5,IFMT2) (WEIGHT(I),I=1,NV)
      WRITE(6,10)
10    FORMAT ('THE ROW VECTOR TO WEIGH THE VARIATES IS:')
      WRITE(6,IFMT2) (WEIGHT(I),I=1,NV)
15    FORMAT (//20A4)
20    IF(KC) 30,21,30
21    DO 24 I=2,NS
      MC=I-1
      DO 22 J=1,MC
22    READ(5,23) DD(I,J)
23    FORMAT (50F1.0/(50F1.0))
24    CONTINUE
30    IF(KS) 40,31,40
31    CALL STANDA(F,NS,NV)
C
C INITIATE THE CENTROIDS OF GROUPS
40    NGP=NS
      DO 55 I=1,NS
      DO 50 J=1,NV
      SC(I,J)=F(I,J)
50    C(I,J)=F(I,J)
55    CONTINUE
      DO 57 I=1,NS
      DO 56 J=1,NS
56    NG(I,J)=0
```

```
 57   NG(I,I)=I
      WRITE(6,60) NGP
 60   FORMAT (//I5,'  GROUPS BEFORE GROUPING IS EXECUTED')
      WRITE(6,61)
 61   FORMAT ('THE CENTROIDS OF THESE GROUPS ARE:'/'GROUP  ID
     1            CENTROIDS')
      DO 63 I=1,NS
      WRITE(6,64) I, ID(I), (SC(I,J),J=1,NV)
 63   CONTINUE
 64   FORMAT (/I5,2X,I4,4X,10F10.5/(15X,10F10.5))
      IF(KF.GT.0) GO TO 97
C
C COMPUTE THE TOTAL SAMPLE DEVIATION MATRIX
      XORIGI=0.0
      DO 90 I=1,NV
      DO 89 J=1,NV
      TD(I,J)=0.0
      DO 88 NN=1,NS
 88   TD(I,J)=TD(I,J)+F(NN,I)*F(NN,J)
 89   CONTINUE
 90   CONTINUE
      DETWDA=0.0
      CALL DET(TD,NV,DETTD)
      WRITE(6,91) DETWDA,DETTD
 91   FORMAT ('INITIAL DETERMINANTS OF POOLED WITHIN-GROUPS DEVIATION MA
     1TRIX AND TOTAL SAMPLE DEVIATION MATRIX ARE:'/10X,E20.10,'  AND  'E
     220.10)
C
C COMPUTE (NS-1)*(NS-1) UPPER TRIANGULAR DISTANCE MATRIX
 97   IF(LEVEL.GE.NGP) STOP
      IF(KW) 98,100,98
 98   DO 99 I=1,NV
 99   WEIGHT(I)=1.0
 100  II=NS-1
      DO 105 I=1,II
      KK=I+1
      DO 104 K=KK,NS
      DD(I,K)=0.0
      DO 103 J=1,NV
```

```
103   DD(I,K)=DD(I,K)+((F(I,J)-F(K,J))**2)*WEIGHT(J)
104   CONTINUE
105   CONTINUE
      WRITE(6,106)
106   FORMAT ('INITIAL UPPER TRIANGULAR DISTANCE MATRIX'/'GROUP (I)
     1                                GROUP (I+1 TO NS)')
      DO 107 I=1,II
      KK=I+1
107   WRITE(6,108) I,(DD(I,K),K=KK,NS)
108   FORMAT (I5,5X,10F10.3/(10X,10F10.3))
      IF(KD.GT.0) GO TO 110
C INITIATE GROUPING DISTANCES
      DO 109 I=1,NS
      GPDIST(I)=0.0
109   GPSTDD(I)=0.0
C
C SEARCH FOR MINIMUM DISTANCE IN THE D-MATRIX
C AND COMPUTE THE CENTROID OF THE NEW GROUP
      TEMPX=0.0
110   DO 200 I=1,NS
200   GPM(I)=1.0
300   X=10.0**10
      DO 305 I=1,II
      IF(GPM(I).EQ.0.0) GO TO 305
      KK=I+1
      DO 304 K=KK,NS
      IF(NG(K,K).EQ.0) GO TO 304
      IF(KC.GT.0) GO TO 301
      IF(DD(K,I)) 304,301,304
301   IF(DD(I,K)-X) 303,304,304
303   X=DD(I,K)
      LR=I
      LC=K
304   CONTINUE
305   CONTINUE
      TEMPLR=GPM(LR)
      TEMPLC=GPM(LC)
      IF(GPM(LC).GT.1.0) GO TO 306
      NG(LR,LC)=LC
```

```
      NG(LC,LC)=0
      GPM(LR)=GPM(LR)+1.0
      GPM(LC)=0.0
      GO TO 309
  306 DO 308 NGM=1,NS
      IF(NG(LC,NGM)-1) 308,307,307
  307 NG(LR,NGM)=NGM
      NG(LC,NGM)=0
      GPM(LR)=GPM(LR)+1.0
  308 CONTINUE
      GPM(LC)=0.0
  309 DO 310 J=1,NV
      SC(LR,J)=SC(LR,J)+SC(LC,J)
      C(LR,J)=SC(LR,J)/GPM(LR)
  310 CONTINUE
      NGP=NGP-1
      IF(KF.GT.0) GO TO 399
C
C COMPUTE THE WD-MATRIX
      DO 312 I=1,NV
      DO 311 J=1,NV
  311 WD(I,J)=0.0
  312 CONTINUE
      DO 321 K=1,NS
      IF(GPM(K).LE.1.0) GO TO 321
      DO 320 I=1,NV
      DO 319 J=1,NV
      DO 318 NGM=1,NS
      IF(NG(K,NGM)-1) 318,317,317
  317 WD(I,J)=WD(I,J)+(F(NGM,I)-C(K,I))*(F(NGM,J)-C(K,J))
  318 CONTINUE
  319 CONTINUE
  320 CONTINUE
  321 CONTINUE
C
C PRINT GROUP MEMBERSHIP
  399 WRITE(6,400)
  400 FORMAT (/'*****************************************************
     1*****************************************************$')
      WRITE(6,401) NGP,LR,LC
  401 FORMAT (//I5,'  GROUPS AFTER COMBINING GROUPS 'I5,'  AND  'I5)
      WRITE(6,4011) X
```

```
 4011 FORMAT ('THE DISTANCE BETWEEN THESE TWO GROUPS IS:'F20.4)
      DX=X-TEMPX
      TEMPX=X
      WRITE(6,4012) DX
 4012 FORMAT ('INCREMENTAL DISTANCE FOR GROUPING = 'F20.4)
      WRITE(6,402) LR,GPM(LR)
 402  FORMAT ('GROUP 'I5,' CONTAINS 'F5.0,' SUBJECTS, THEY ARE:')
      DO 405 NGM=1,NS
      IF(NG(LR,NGM)-1) 405,403,403
 403  WRITE(6,404) NGM
 404  FORMAT (55X,I3)
 405  CONTINUE
      WRITE(6,406) LR
 406  FORMAT ('CENTROID OF GROUP 'I5,' IS:')
      WRITE(6,407) (C(LR,J),J=1,NV)
 407  FORMAT (5X,10F10.5/(5X,10F10.5))
      IF(KF.GT.0) GO TO 4301
C
C COMPUTE F-RATIO AND INCREMENT OF WD-MATRIX
      DETWDB=0.0
      CALL DET(WD,NV,DETWDB)
      WRITE(6,4071) DETWDB
 4071 FORMAT ('THE DETERMINANT OF POOLED WITHIN-GROUPS DEVIATION MATRIX
     1IS:'E20.10)
      XINCRE=DETWDB-DETWDA
      IF(DETWDA-0.0) 409,408,409
 408  WRITE(6,4081)
 4081 FORMAT ('RATIO OF INCREMENT INVALID')
      RINCRE=0.0
      GO TO 411
 409  RINCRE=XINCRE/DETWDA
 411  WRITE(6,410) XINCRE, RINCRE
 410  FORMAT ('INCREMENT OF THE DETERMINANT OF POOLED WITHIN-GROUPS MATR
     1IX IS 'E20.10/'RATIO OF INCREMENT TO TOTAL WITHIN-GROUPS DEVIATION
     2 MATRIX IS 'E20.10)
      IF(XORIGI.EQ.0.0) GO TO 4102
      ROFCH=(XINCRE-XORIGI)/XORIGI
      WRITE(6,4101) ROFCH
 4101 FORMAT ('RATE OF CHANGE OF THE INCREMENT OF WITHIN-GROUPS DEVIATIO
```

```
     1N MATRIX IS:'/10X,E20.10)
4102 XORIGI=XINCRE
     XLAMB=ABS(DETWDB/DETTD)
     EM=NV
     EK=NGP
     EN=NS
     IF(EK.EQ.1) GO TO 425
     IF(EM-2.0) 412,412,413
412  DFN=2.0*(EK-1.0)
     DFD=2.0*(EN-EK-1.0)
     Y=SQRT(XLAMB)
     IF(Y.EQ.0.0) GO TO 416
     FRATIO=(1.0-Y)*DFD/(Y*DFN)
     GO TO 419
413  S=SQRT(((EM**2)*((EK-1.0)**2)-4.0)/((EM**2)+((EK-1.0)**2)-5.0))
     IF(S) 414,414,415
414  DFN=EK-1.0
     DFD=EN-EK
     FRATIO=((1.0-XLAMB)/XLAMB)*(DFD/DFN)
     GO TO 419
415  Y=XLAMB**(1.0/S)
     IF(Y.EQ.0.0) GO TO 416
     XM1=(EN-1.0)-((EK+EM)/2.0)
     XL=-((EM*(EK-1.0))-2.0)/4.0
     R1=(EM*(EK-1.0))/2.0
     DFN=2.0*R1
     DFD=(XM1*S)+(2.0*XL)
     FRATIO=((1.0-Y)/Y)*(DFD/DFN)
     IF(FRATIO.LT.0.0) WRITE(6,418)
418  FORMAT ('-- F-RATIO INVALID --')
     GO TO 419
416  WRITE(6,417)
417  FORMAT ('THE F-RATIO SHOWS THAT THE CENTROIDS OF GROUPS ARE NOT
    1EQUAL')
     GO TO 430
419  WRITE(6,420) FRATIO,DFN,DFD
420  FORMAT ('F-RATIO ='E20.10,' WITH DEGREES OF FREEDOM: DFN='F10.1,'
    1DFD='F10.1)
     GO TO 430
```

```
 425  WRITE(6,426)
 426  FORMAT ('THE F-RATIO SHOWS THAT THE CENTROIDS OF GROUPS ARE EQUAL'
     1)
 430  DETWDA=DETWDB
 4301 IF(KD.GT.0) GO TO 4302
C COMPUTE INTRAGROUP DISTANCES
      SDD=0.0
      XJOINT=0.0
      KNS=NS-1
      DO 350 I=1,KNS
      KK=I+1
      IF(NG(LR,I).LT.1) GO TO 350
      DO 345 K=KK,NS
      IF(NG(LR,K).LT.1) GO TO 345
      DO 340 J=1,NV
 340  SDD=SDD+((F(I,J)-F(K,J))**2)*WEIGHT(J)
      XJOINT=XJOINT+1.0
 345  CONTINUE
 350  CONTINUE
C COMPUTE AND PRINT THE AVERAGE INTRAGROUP DISTANCE AND ITS INCREMENT
      AVED=SDD/XJOINT
      WRITE(6,351) SDD
 351  FORMAT ('SUM OF THE INTRAGROUP DISTANCES ='F15.4)
      WRITE(6,352) AVED
 352  FORMAT ('AVERAGE INTRAGROUP DISTANCE ='F15.4)
      IF(TEMPLR-1.0) 3522,3522,3520
 3520 IF(TEMPLC-1.0) 3523,3523,3521
 3521 XINCRD=AVED-(GPDIST(LR)*TEMPLR+GPDIST(LC)*TEMPLC)/(TEMPLR+TEMPLC)
      GO TO 3524
 3522 XINCRD=AVED-GPDIST(LC)
      GO TO 3524
 3523 XINCRD=AVED-GPDIST(LR)
 3524 GPDIST(LR)=AVED
      WRITE(6,353) XINCRD
 353  FORMAT ('INCREMENTAL AVERAGE INTRAGROUP DISTANCE ='F15.4)
C
C COMPUTE AND PRINT SQUARED STANDARD DISTANCE OF THE NEW GROUP
      STDD=0.0
      DO 380 I=1,NS
      IF(NG(LR,I).LT.1) GO TO 380
      DO 375 J=1,NV
 375  STDD=STDD+((F(I,J)-C(LR,J))**2)*WEIGHT(J)
```

```
 380  CONTINUE
      STDD=STDD/GPM(LR)
      IF(TEMPLR-1.0) 3803,3803,3801
 3801 IF(TEMPLC-1.0) 3804,3804,3802
 3802 XINCRS=STDD-(GPSTDD(LR)*TEMPLR+GPSTDD(LC)*TEMPLC)/(TEMPLR+TEMPLC)
      GO TO 3805
 3803 XINCRS=STDD-GPSTDD(LC)
      GO TO 3805
 3804 XINCRS=STDD-GPSTDD(LR)
 3805 GPSTDD(LR)=STDD
      WRITE(6,381) STDD
 381  FORMAT (/'SQUARED STANDARD DISTANCE ='F15.4)
      WRITE(6,382) XINCRS
 382  FORMAT ('INCREMENTAL SQUARED STANDARD DISTANCE ='F15.4)
C
C MODIFY THE DISTANCE MATRIX
 4302 X=10.0**10
      LRR=LR-1
      IF(LR.EQ.1) GO TO 4501
      DO 450 I=1,LRR
      IF(DD(I,LR).EQ.X) GO TO 450
      DD(I,LR)=0.0
      DO 449 J=1,NV
 449  DD(I,LR)=DD(I,LR)+((C(I,J)-C(LR,J))**2)*WEIGHT(J)
 450  CONTINUE
 4501 KK=LR+1
      DO 452 K=KK,NS
      IF(DD(LR,K).EQ.X) GO TO 452
      DD(LR,K)=0.0
      DO 451 J=1,NV
 451  DD(LR,K)=DD(LR,K)+((C(K,J)-C(LR,J))**2)*WEIGHT(J)
 452  CONTINUE
      LCC=LC-1
      DO 453 I=1,LCC
 453  DD(I,LC)=10.0**10
      IF(LC.EQ.NS) GO TO 455
      KK=LC+1
      DO 454 K=KK,NS
 454  DD(LC,K)=10.0**10
```

```
C MODIFY THE CONTIGUITY MATRIX
C
      IF(KC.GT.0) GO TO 455
      IF(LR.EQ.1) GO TO 4542
      DO 4541 J=1,LRR
      DD(LC,J)=DD(LC,J)*DD(LR,J)
 4541 DD(LR,J)=DD(LC,J)
 4542 KK=LC+1
      IF(LC.EQ.NS) GO TO 4545
      DO 4543 I=KK,NS
      DD(I,LR)=DD(I,LR)*DD(I,LC)
 4543 DD(I,LC)=DD(I,LR)
 4545 LRR=LR+1
      IF(LRR.EQ.LCC) GO TO 455
      DO 4544 J=LRR,LCC
      DD(LC,J)=DD(LC,J)*DD(J,LR)
 4544 DD(J,LR)=DD(LC,J)
 455  IF(LEVEL-NGP) 460,500,500
 460  IF(NGP-1) 500,500,300
 500  IF(KM) 600,501,600
 501  DO 505 I=1,NS
      IF(GPM(I)) 505,505,502
 502  WRITE(6,503)
      WRITE(6,504) I,(NG(I,J),J=1,NS)
 503  FORMAT ('MATRIX OF GROUP MEMBERSHIPS OF THE FINAL STAGE OF GROUPIN
     1G IS:'/'GROUP                                MEMBERS')
 504  FORMAT (/I5,5X,25I4/(10X,25I4))
 505  CONTINUE
 600  STOP
      END
```

```
C
C SUBROUTINE DET
C A IS ORIGINAL MATRIX FROM THE MAIN PROGRAM
C N IS THE ORDER OF A-MATRIX NOT TO EXCEED 50
C
      SUBROUTINE DET(A,N,DETERM)
      DIMENSION IPIVOT(15),A(15,15),INDEX(15,2),PIVOT(15)
      EQUIVALENCE (IROW,JROW),(ICOLUM,JCOLUM),(AMAX,T,SWAP)
C
C INITIALIZATION
      DETERM=1.0
 15   DO 20 J=1,N
 20   IPIVOT(J)=0
 30   DO 550 I=1,N
C SEARCH FOR PIVOT ELEMENT
 40   AMAX=0.0
      DO 105 J=1,N
      IF(IPIVOT(J)-1) 60,105,60
 60   DO 100 K=1,N
 70   IF(IPIVOT(K)-1) 80,100,740
 80   IF(ABS(AMAX)-ABS(A(J,K))) 85,100,100
 85   IROW=J
 90   ICOLUM=K
 95   AMAX=A(J,K)
 100  CONTINUE
 105  CONTINUE
 110  IPIVOT(ICOLUM)=IPIVOT(ICOLUM)+1
C INTERCHANGE ROWS TO PUT PIVOT ELEMENT ON DIAGONAL
 130  IF(IROW-ICOLUM) 140,260,140
 140  DETERM=-DETERM
 150  DO 200 L=1,N
 160  SWAP=A(IROW,L)
 170  A(IROW,L)=A(ICOLUM,L)
 200  A(ICOLUM,L)=SWAP
 260  INDEX(I,1)=IROW
```

```
 270   INDEX(I,2)=ICOLUM
 310   PIVOT(I)=A(ICOLUM,ICOLUM)
 320   DETERM=DETERM*PIVOT(I)
C
C DIVIDE PIVOT ROW BY PIVOT ELEMENT
 330   A(ICOLUM,ICOLUM)=1.0
 340   DO 350 L=I,N
 350   A(ICOLUM,L)=A(ICOLUM,L)/PIVOT(I)
C
C REDUCE NON-PIVOTROWS
 380   DO 550 LL=1,N
 390   IF(LL-ICOLUM) 400,550,400
 400   T=A(LL,ICOLUM)
 420   A(LL,ICOLUM)=0.0
 430   DO 450 L=1,N
 450   A(LL,L)=A(LL,L)-A(ICOLUM,L)*T
 550   CONTINUE
C
C INTERCHANGE COLUMNS
 600   DO 710 I=1,N
 610   L=N+1-I
 620   IF(INDEX(L,1)-INDEX(L,2)) 630,710,630
 630   JROW=INDEX(L,2)
 650   DO 705 K=1,N
 660   SWAP=A(K,JROW)
 670   A(K,JROW)=A(K,JCOLUM)
 700   A(K,JCOLUM)=SWAP
 705   CONTINUE
 710   CONTINUE
 740   RETURN
       END
```

```
C
C SUBROUTINE STANDA TO STANDARDIZE THE RAW DATA
C A IS THE MATRIX TO BE STANDARDIZED, M IS # OF SUBJECTS, N IS # OF
C VARIABLES
C
      SUBROUTINE STANDA(A,M,N)
      DIMENSION A(100,15)
      DO 25 J=1,N
      XM=M
      SA=0.0
      DO 10 I=1,M
 10   SA=SA+A(I,J)
      XMEAN=SA/XM
      SSA=0.0
      DO 15 I=1,M
 15   SSA=SSA+(A(I,J)-XMEAN)**2
      STD=SQRT(SSA/(XM-1.0))
      DO 20 I=1,M
 20   A(I,J)=(A(I,J)-XMEAN)/STD
 25   CONTINUE
      RETURN
      END

END
```