Scaling methods: measurement of the service level of centers in Noord Brabant

Musterd, S.; Dieleman, F.M.

DOI
10.1111/j.1467-9663.1981.tb00931.x

Publication date
1981

Published in
Tijdschrift voor Economische en Sociale Geografie

SCALING METHODS: MEASUREMENT OF THE SERVICE LEVEL OF CENTERS IN NOORD-BRABANT (THE NETHERLANDS)

by

S. MUSTERD* & F. M. DIELEMAN*

Amsterdam, the Netherlands

Introduction

In the past, Dutch geographers have not been too concerned about methodological aspects of measurement. For many research questions measurement in itself is not a problem, and thus it does not command much attention. For example, when research deals with the population structure or the employment structure, it is standard practice to express the number of persons or businesses belonging to various categories in amounts or percentages; the selection of variables is also straightforward. Moreover, in such a case, measurement is on the ratio level, which permits all desired computations to be carried out.

Many Dutch geographers whose research was conducive to analysis techniques have restricted their focus to tangible phenomena that can easily be used as variables. This precludes problems of measurement, making scaling methods designed to solve problems of measurement irrelevant. Others have kept far from analysis techniques and have not applied scaling, even when the object of research warranted its use — perhaps in part because scaling is so little known in Dutch geography. In this context the continued lack of concern on the part of geographers for problems of measurement and the corresponding scaling methods is not surprising.

Before we present our arguments for more attention to scaling methods, it is pertinent to deal briefly with the concept of measurement. A study generally articulates a conceptual model which gives an overview of the characteristics (the theoretical variables) of the selected research elements. These theoretical variables receive an empirical elaboration, or operationalization, since a necessary step toward measurement is the precise definition of variables (Segers 1977). Measurement is the determination of the value of a variable by way of some observation procedure. This does not imply, as is sometimes assumed, that measurement always produces quantitative or metric information, but it does imply that the value is expressed in a symbol and that criteria determine which symbol is assigned to a research element (see e.g. Torgerson 1958, Coombs 1964, Nunnally 1967, Swanborn 1973, Maranell 1974).

In the application of scaling methods, well-defined rules are used to assign numbers (or symbols) to objects, such that the value of a certain characteristic of the objects is expressed in the numbers. The diversity of commonly used rules has led to a great variety of scaling methods. Although an extensive literature has arisen on the topic, we think that these cursory remarks suffice. References to this body of literature are included in the text below.

There is every reason for this article to reiterate the potential of scaling methods for geographic research. It may be profitable to consider using scaling methods in at least three research situations:

a. If the concepts with which the analysis is concerned are complex and abstract. Such concepts, increasingly used in Dutch geographical research, are usually hard to operationalize. They are thus often defined in such a way that only a low level of measurement (ordinal) is attained, whereas a higher level of measurement may be desirable. If this is the case, scaling

* Department of Geography, Free University, Postbus 7161, 1007 MC Amsterdam.

Received, February 1980; revised December 1980.

methods may profitably be employed. We select two examples of abstract concepts from among those frequently found in Dutch geography and planning literature to illustrate this: service level of urban centers and degree of urbanization. The discussion of the operationalization of the first concept forms the main body of this paper and will be treated in depth later.

The other example is briefly mentioned here to introduce our argument for the use of scaling methods. In the literature this concept is frequently operationalized by employing the Census Bureau's official urbanization typology of municipalities in the Netherlands (Typologie van de Nederlandse gemeenten... 1964).

Other than degree of urbanization, few variables can be found that have been related to so many other characteristics of the municipalities. Yet the degree of urbanization has been defined at a low level of measurement; it scarcely reaches the ordinal level (rank order). Due to this, other variables can only be related to degree of urbanization by means of data organized in tables. Even though this official typology (and the currently suggested improvements, see Herziening typologie... 1976) has been fruitfully employed in a large number of studies, we would like to point out that if it were possible to operationalize the concept at a higher level of measurement, the relations with other variables could be more exactly described by employing more sophisticated methods of analysis.

b. A second situation in which scaling methods are appropriate is when attitudes, opinions, behavior, etc. are the object of study. Because of their origin in psychology and sociology, scaling methods were developed to deal with such a research situation (Shepard, Romney & Nerlove 1972, Maranell 1974). The field of geography increasingly emphasizes spatial preferences and perceptions (Golledge & Rushton 1976) and topics such as geographic concept-forming by children (Dijkink & Elbers 1981) and decision-making with respect to industrial location (Townroe 1972, Hamilton 1974). Consequently, scaling methods have assumed a more important role in geographic research. This explains why these studies do emphasize the problems of measurement and employ scaling methods. Many other studies on similar topics, however, use only series of categorized scales of judgment (e.g. from very positive to very negative), which though legitimate, indicate only a superficial treatment of the problems of measurement.

c. Scaling methods may be called for when there is a demand for data material on which correlation, regression, or factor analysis may not be performed (see Voogd 1978, p. 3). If for whatever reason these methods of analysis are not applicable, even in spite of an interval level of measurement, scaling methods may still prove their merit. Such a situation occurs for instance when many objects have a value of zero on a series of variables (see Dieleman 1978, 1980), which is the case in the study of the service level of centers in Noord-Brabant, the main body of this article. As will be shown, in such a situation certain multidimensional scaling methods provide suitable alternatives for the above-mentioned analysis techniques. This is primarily because these methods (just like cluster analysis) are usually based on (dis)similarities between the objects. When data are not yet (dis)-similarities, as is the case in this study, a wide variety of coefficients is available for producing (dis)similarities based on the data.

This article does not intend to give a comprehensive review of the diverse scaling techniques with all their potentialities, limitations, and applications (well-known reviews have been compiled by Shepard 1972 and—more geographically oriented—by Golledge & Rushton 1972). Regarding the two forms of scaling analysis that are discussed in more detail in this article, only a few of the most important sources are mentioned. We chose deliberately to treat scaling methods on the basis of results of empirical research, avoiding technical explanations as much as possible. Hopefully this gives a clear picture of their potential for use while providing enough material for a critical evaluation of their application.

As mentioned before, our study of the service level of centers provides a context for the use of scaling methods, which is the central focus in this paper. We chose this context for two reasons. In the first place, even though no attitudes or preferences are being measured, this topic demonstrates that the measurement procedure is often more problematic than anticipated. In the second place, since other geographic and planning studies have measured service level (operationalized in a variety of ways), there is a basis for comparison of the results of these studies with the outcome of our application of scale analysis. Thus, the practicability of scale analysis is brought into perspective.

Scaling methods can be differentiated as unidimensional and multidimensional. From both broad categories one specific method is chosen to measure the service level. A selection is necessary because of the large number of variants of uni- and multidimensional scaling methods. Although in this way the study can only deal with a small portion of the available scaling methods, this selection illustrates their potentialities and limitations adequately. The results achieved by applying the two methods are discussed in the following part of this paper.

The service level of centers

Numerous Dutch geographic and planning studies classify the centers in a given area or in the country as a whole according to their service level (Voortman 1961/62, P.P.D. Friesland 1966, Luyten & Verberk 1968, Keuning 1971, Buursink 1971b, P.P.D. Noord-Brabant 1972, Onderzoekskring OKU 1976, Kiestra 1978, Van Heesewijk 1978). In these studies the service level of each center is generally expressed in a single number which is then used to allocate the centers to a number of classes. It is often assumed that the centers form a hierarchy of functions; based on this, service areas for the centers (or, centralities) are determined (see Buursink 1971a). Our study, however, is restricted to the service level and does not treat these other aspects.

The analyses referred to demonstrate that it is not easy to measure the service level of centers. It is clear that the service functions—retail stores, facilities for medical care, education, recreation, and cultural activities—should be included in the determination of the service level. For this reason most analyses begin with an inventory of a broad cross-section or of a specific selection (key elements) of the functions in the centers.

The problem then shifts to the determination of a value for the total number of elements of each service function to indicate the level of the services. All authors share the opinion that the functions in the various sectors are not of equal value. The availability of some functions is indicative for a high level of services, while others occur as well in centers with a low service level. Almost everyone agrees that the service level of a center cannot be expressed by the mere sum total of all elements of each service function. Therefore a procedure is usually determined to allocate values to each available service function, after which the scores for the various functions are added up per center: this total is then used to indicate the service level. On the basis of these scores, a classification of the centers is often constructed. While internally consistent, the results of various studies differ.

When the results of this procedure are evaluated, it seems on first impression that the picture conforms reasonably well to a subjective image of the service level of the centers. This can partly be explained by the fact that the researcher bases his subjective decisions regarding the allocation of values and determination of thresholds on his familiarity with the characteristics of the centers. In addition, if the area under analysis is relatively large, the service level of the centers generally differs greatly. If these differences are large enough, one can quickly find a reasonable and useful classification, no matter how the concept is operationalized. From the point of view of methodology, a number of objections may be brought forward against these arbitrary procedures. These objections can be mitigated by use of scaling methods to determine the service level of centers. Some of these objections are:

a. The measurement procedures have a subjective character and are hardly standardized: each researcher chooses his own method. The results are therefore hard to evaluate. Because the scaling methods are based on a more generally accepted measurement model, the results can be evaluated more easily. This does not mean that no subjective decisions need be made when using scaling methods. But the measurement model and the basic assumptions behind it draw explicitly on the existing body of literature and can thus be justified. One objectionable aspect of the subjective method is that it indicates the service level of a center by one number. Thus it is tacitly assumed that the concept is unidimensional; we will argue below (see point c.) that this cannot be taken for granted.

b. Criteria are not generally available for an evaluation of the extent to which the scale being constructed fits the data from which it has been deduced. With scaling methods such criteria are more adequate.

c. On the basis of the raw data, it is not easy to determine whether or not the service level of centers may be conceived of as a unidimensional phenomenon. In other words, if for example one has made an inventory of retail and educational facilities, it is difficult to check whether a certain package of retail facilities will always be found in conjunction with a specific package of educational facilities or whether these two aspects vary independently of one another. Multidimensional scaling methods can give a better indication in this respect.

d. The classification of the centers generally seems artificial. It appears to be added to data rather than derived from it. On this issue as well, scaling methods—in combination with cluster analysis—can provide some insight.

e. A frequently voiced objection is that the meaning of the scores that indicate the service level is ambiguous. The question whether the figures are of ordinal or interval level cannot be answered (see Hays 1974). For example, if three centers, A, B, and C score respectively 26, 36 and 80 points, it remains unclear whether this means only that $A < B < C$, or whether there is some meaning attached to the size of the intervals between the numbers. While this problem is not treated in the studies mentioned, the literature concerning scaling methods explicitly discusses this problem.

The listed objections are serious enough to warrant a closer look at the potentials of scaling methods for measuring service levels. Not all problems are resolved by use of scaling methods, and we do not want to imply that these methods will provide demonstrably better results than those of the previously indicated studies. Our only purpose is to illustrate the potentials and the limitations of a number of scaling methods.

The data are derived from an inventory of the situation of ca. 1976 made by the Noord-Brabant Provincial Planning Department.[1] Using these data, 42 categories of service functions were assembled (see Table 1). For 61 centers in Noord-Brabant (see Figure 1) the number of service elements was computed per function and per center. It is an advantage that the data were gathered per center; if the data are summarized for large spatial units in which different centers are located, then no further differentiation can be made between smaller and larger centers, which often have a difference in service level.

To eliminate a great variation in service level, not all centers in the province are included in this study; only the largest by population size and by total number of service elements are incorporated. Too much variation in service level tends to highlight the difference between centers at the extreme ends of the continuum. In a test situation such as this, the applicability of a scaling method can be evaluated better when used only to distinguish subgroups of a group of centers wherein the service level does not vary too extensively. Moreover, the available computer software for the multidimensional scaling analysis limited the number of centers that could be included in the study. In practice one will make a pertinent and more substantiated selection of the centers to be categorized, for example, if all centers within a province must be classified. Our selection is arbitrary; the primary reason for our inclusion

[1] We are grateful to the P.P.D. for making the data available.

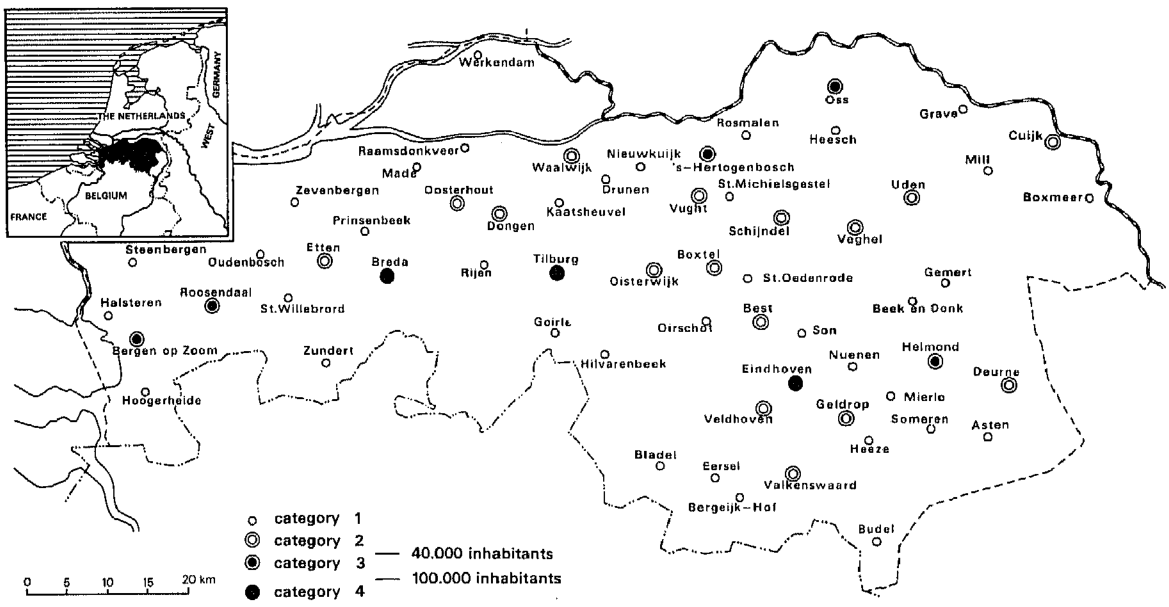


Fig. 1. The 61 centers included in the study; the subdivision in four categories is based on the selected multidimensional scaling method and on cluster analysis (compare Figure 6).

of only the *larger* centers is that more people are familiar with these and thus the results of the illustration will be more vivid.

The Guttman scale applied to service level
The following analysis sets out to measure the service level of centers by use of a unidimensional scaling method. Although unidimensional scales yield only rank order measurements, they are lucid and easy to apply. Of the set of methods available (see e.g. Torgerson 1958, Maranell 1974, Utrecht 1975) the Guttman scale is conceptually the most appropriate in this context.

The Guttman scale belongs to the set of scaling methods that uses more items (in this case, categories of service functions) to develop a single scale. The final position of the object on the scale to be constructed is not determined by its score on any one item but by the pattern of scores on a series of items. The Guttman scale is cumulative, as distinguished from differential (Thurstone) and additive (Likert) scales (see e.g. Utrecht 1975). On a cumulative scale there is a graduated distance between items where some items are more likely to show a positive score than are others. As a rule a positive score on a difficult item implies a similar score on the less difficult items; in other words, when a more difficult item is 'passed', it is assumed that the object does not 'fail' on less difficult items.

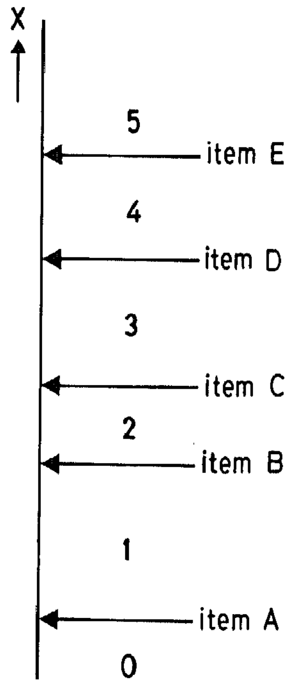


Fig. 2. The structure of the Guttman scale.

Figure 2 depicts the structure of the Guttman scale. A given property X (here, service level) which is taken as a characteristic of a population of objects (here, centers) is assumed to form a unidimensional continuum. The total of property X in each of the objects is measured by checking the individual scores of the objects on each of five items, A through E, arranged from easy to difficult. Each item must therefore relate to property X. The items are dichotomous: each object can only score positive (pass) or negative (fail) on them. An object that does not pass the easiest item A is assigned an ordinal score of zero. The model assumes that an object that does not pass item A will not pass the other items. If an object passes item C but not item D, then the ordinal score is three, because it is assumed that the easier items A and B are also passed and the more difficult item E will not be attained, etc. For a more complete treatment of the model consult for example Torgerson (1958). This measurement model is consistent with the bulk of theory regarding the service level of centers. The theory generally assumes that functional elements occur cumulatively in centers: functional elements of a lower order are always concurrent with elements of a higher order.

In practice, scoring patterns of all objects on the items may not be expected to form a perfect Guttman scale. Some objects for example will pass item E and fail to pass D, which in terms of the measurement model must be considered an error. A certain number of errors is acceptable, measured by a coefficient of reproducibility:

$$\text{Rep(roducibility)} = 1 - \frac{\text{number of errors}}{\text{number of responses (objects} \times \text{items)}}$$

If Rep is above 0.90, it is considered that the items do indeed form a valid Guttman scale. (The coefficient of reproducibility should be used with discretion; see e.g. Utrecht 1975, Nie *et al.* 1975, Lloyd 1977.)

The Guttman scale was employed to measure the service level of the 61 centers in Noord-Brabant and expresses these measurements as an ordinal score. To construct a Guttman scale, from the 42 categories of service functions (variables) it was necessary to choose some having a diverse degree of difficulty and subsequently to dichotomize them. To this end the frequency distributions of each of the 42 variables were analyzed for the 61 centers. The frequency distributions of some variables show rather obvious division points. These variables were chosen for the Guttman scale because they could be dichotomized at their division point.

It seems logical to use these very division points, because they coincide with more-or-less evident groups of centers. The final allocation of centers to these groups should be a good reflection of the empirical disparities between groups of centers. Incidentally, we have experimented with a dichotomization around the median, a method discussed in the literature. This method produced much inferior results (in terms of reproducibility), perhaps precisely because the median seldom coincides with the obvious division points in the frequency distribution.

Both the selection of suitable categories of service functions for the scale analysis and the dichotomization remain highly subjective. Moreover, these choices determine to a great extent the results of the analysis.

Figure 3 shows the results of a number of scale analyses. First two Guttman scales are presented which relate to two categories of service functions, namely, retail trade and education. Subsequently more diverse variables are combined as items in two general Guttman scales of the service level of the centers. Only four or five categories of service functions are used in each scale, because, as shown in other studies, the centers may be divided into a relatively small number of groups (see below). In addition it proved unproductive to include more items; the items are then too similar, and often only one or two centers are differentiated from the larger group.

For a Guttman scale analysis the results are indisputably good. The coefficient of reproducibility is high in all cases, which indicates that only on few occasions the centers show a pattern of scores that does not completely conform to the model. The results are also good in case other coefficients are employed to evaluate the scales (see e.g. Nie *et al.* 1975). When comparing the results of the scale analyses, some of the interesting observations concerning the service level of the 61 centers which emerge are as follows: especially in centers close to the national border, there is a disparity between the service levels with respect to retail trade and education; and centers near the largest centers have a relatively low service level. We will not elaborate on these points in this context, however.

The analysis of the service level of centers by means of a Guttman scale is valuable in a number of respects:

a. The measurement level is indisputable, the service level is recorded in an ordinal score.

b. The measurement procedure is evident and fairly easy to apply.

The conclusions regarding any ordinal score are well defined. For example, if a center scores three on the scale for the retail trade (see Figure 3), this means that the center has at least seven stores in housewares, three drugstores and five jewelry stores, but fewer than ten shoe stores and three department stores.

There is, however, no sufficient theoretical explanation why ten shoe stores count more than seven stores in housewares. But once the division points for the dichotomization are chosen, the procedure for the Guttman scale analysis determines the most suitable sequence of the items depending on the number of centers that passes each item.

c. The assumptions conditioning the use of the scale are well defined and are treated in the literature. This provides a basis for discussing the suitability of the Guttman scale to a given problem. Thus the researcher's confrontation with assumptions and subjective decisions is perhaps more solid than with a self-designed procedure.

d. The scaling method itself subdivides the centers—the centers having the same score form one group—which means that no arbitrary criteria are needed.

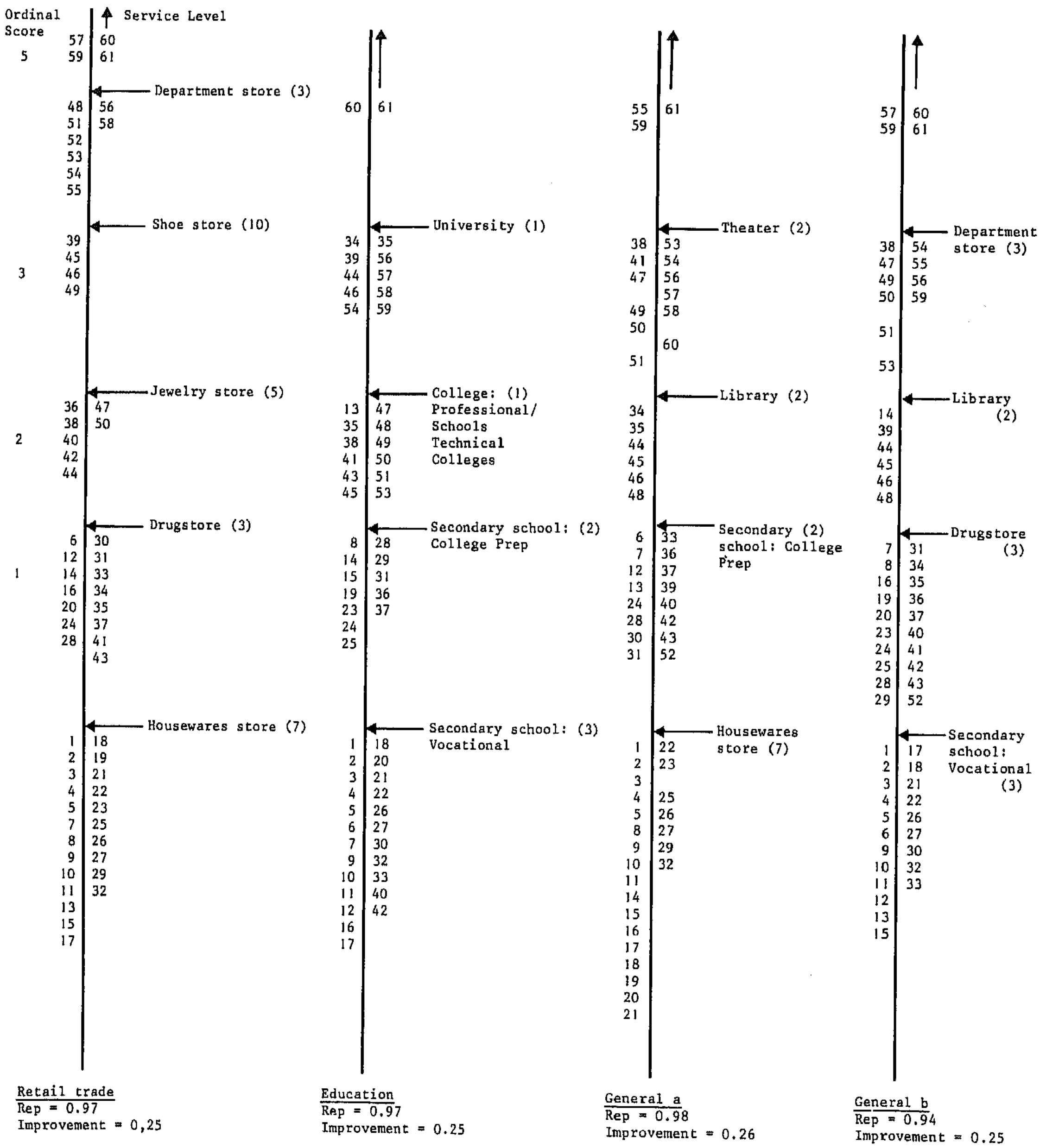
Apart from these advantages, the use of the Guttman scale to measure service levels also generates some problems:

a. The implicit assumption that service level is a unidimensional phenomenon should be investigated. The differences between the scales for retail trade and education (see Figure 3) give rise to doubts concerning its empirical validity (see Van der Meulen 1979).

b. Although the subdivision of the centers is guided by the procedure, it seems artificial; it does not become sufficiently clear whether or not the constructed groups of centers coincide with a logical grouping of the centers. The model is too limited to provide insight.

c. A number of subjective decisions are inevitable and determine to a large extent the results of the analysis. Some of these subjective decisions are intrinsic to every method, e.g. the choice of the basic variables. But when applying scale analysis, additional decisions must be made in regard to selection of service functions as items and in regard to the dichotomization. The data material is only partially exploited, and the measurement level attained is low, considering the potential of the available data.

Fig. 3. The service level of the 61 centers in Noord-Brabant: four Guttman scales.



1 Beek en Donk
2 Bergeijk-Hof
3 Prinsenbeek
4 Mierlo
5 Heesch
6 Hoogerheide
7 Nieuwkuijk
8 Mill
9 Halsteren
10 Raamsdonkveer
11 St.-Willebrord
12 Werkendam
13 Eersel
14 Bladel
15 St.-Michielsgestel
16 Grave
17 Heeze
18 Hilvarenbeek
19 Oirschot
20 Rijen
21 Son
22 Drunen
23 Steenbergen
24 Zundert
25 Someren
26 Made
27 Rosmalen
28 Budel
29 Goirle
30 St.Oedenrode
31 Zevenbergen
32 Nuenen
33 Kaatsheuvel
34 Oudenbosch
35 Gemert
36 Asten
37 Boxmeer
38 Cuijk
39 Dongen
40 Best
41 Etten
42 Oisterwijk
43 Schijndel
44 Vught
45 Deurne
46 Boxtel
47 Veghel
48 Uden
49 Veldhoven
50 Geldrop
51 Oosterhout
52 Valkenswaard
53 Waalwijk
54 Oss
55 Bergen op Zoom
56 Roosendaal
57 Helmond
58 's-Hertogenbosch
59 Breda
60 Tilburg
61 Eindhoven

The analysis of the service level of centers by means of the Guttman scale only partially accommodates some of the methodological objections that were raised in our discussion of the procedures commonly applied (see the previous section). In the following section the multidimensional scale analysis is used to bring the service level of the centers into perspective. Based on that, the advantages and disadvantages of the various measurement procedures will be discussed again.

A multidimensional scale analysis
The multidimensional scaling methods form a group of related analysis procedures that are especially suited to application in the social sciences, where data on a low (e.g. ordinal) level of measurement are frequently used.

Shepard (1972) describes how diverse schools of thought have influenced the development of these methods. They show a similarity to the better-known factor analysis. In both methods scores of objects on a relatively large number of variables (m) are reduced to scores on a much smaller number of dimensions. When determining component scores a principal component analysis can be seen as a form of metric multidimensional scale analysis.

The multidimensional scaling methods do not start, however, with a matrix of correlation coefficients, as is common for factor analysis, but with a matrix of dissimilarities. Because many (dis)similarity coefficients have been developed, this gives these methods a greater flexibility (Johnston 1976); the cluster analysis, which also departs from such a matrix, also enjoys this flexibility. Moreover, less strict assumptions are made than for factor analysis, and it is easier to work with ordinal measurements. When working with ordinal data, a form of non-metric multidimensional scale analysis must be applied.

The conceptual basis for the techniques is simple (Shepard 1972). It is assumed that for every two objects *i* and *j* of a collection of size *n*, a (dis)similarity coefficient δ_{ij} can be defined. The δ_{ij} indicates the (in)equality, association, interaction, etc.—in general, the proximity or distance—between the objects. Subsequently, a search is made for a configuration of *n* points in (Euclidian) space with as few dimensions as possible, so that it meets to the greatest extent the requirement that the distance between points, d_{ij}, be monotonically related to the (dis)similarities (Kruskal 1964a). The coordinates of the points in geometric space are the scale values. In other words, it must hold that: $d_{ij} < d_{kl}$ if $\delta_{ij} < \delta_{kl}$

This monotonicity requirement is illustrated in Figure 4. The requirement implies that if the d_{ij} and δ_{ij} values are ranked according to size, a subsequent point should never lie to the left of—or under—the previous point. Figure 4a does show and Figure 4b does not show a case of monotone relation between d_{ij} and δ_{ij}. In this last situation, one can speak of stress, because the distances (d_{ij}) derived from the scale values do not fully fit the original data. This will be dealt with later.

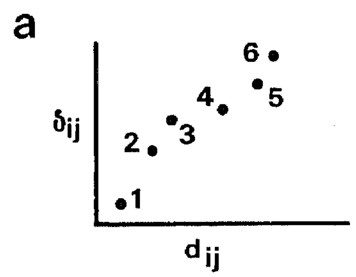


Fig. 4a. A monotone relation between d_{ij} *and* δ_{ij}.

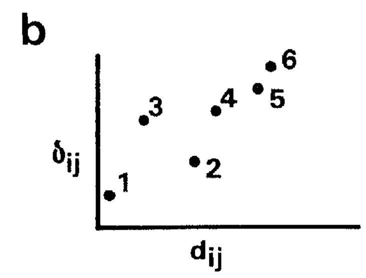


Fig. 4b. A non-monotone relation between d_{ij} *and* δ_{ij}.

To find a suitable solution to the problem concerning multidimensional scale analysis, a series of steps must be taken (see Nijkamp & Voogd 1978). This is expressed schematically in Figure 5. Starting with the data matrix, here the 42 categories of service functions in the 61 Noord-Brabant centers, a dissimilarity matrix must be computed. This study opted for the non-metric dissimilarity coefficient of Bray and Curtis (see Berkouwer 1978):

$$\delta_{ij} = \frac{\sum_{p=1}^{m} |X_{ip} - X_{jp}|}{\sum_{p=1}^{m} (X_{ip} + X_{jp})}$$

To compute the dissimilarities between the centers, varying service functions must be summed. It is therefore reasonable to choose a non-metric dissimilarity coefficient (of which the above-mentioned is an example). The coefficients of the differences between the centers do not have an absolute meaning. This concerns, incidentally, only the dissimilarities themselves. In a later phase, through a reduction in the number of dimensions (42 categories of service functions yield 2 dimensions or scales) the objects are forced into a fixed place in the space formed by the scales; then the differences between the objects are expressed on an interval scale.

Our choice of this specific dissimilarity coefficient has the advantage that the numerous double zeros—two centers that each score zero on a series of variables—play no role, and that variables with relatively high values—service functions that occur in large numbers—are not of decisive importance for the analysis. The selection of a suitable (dis)similarity coefficient is of great importance, because it determines to a large extent the results of the analysis. Deurloo (1976) reviews various options, such as introducing weighting factors or accounting for the inter-dependency of the variables. It is always advisable to experiment with different (dis)similarity coefficients in order to determine the influence of the choice of a given measure on the final result.

During the application of the multidimensional scale analysis, the (dis)similarity matrix is subsequently compared with a matrix D which expresses the distances between the objects (centers) in a space. The coordinates of the objects in this space are the scale values of the objects. The analysis is limited to geometric space and distance. The number of dimensions of the space must be determined by the researcher. This choice can be a product of the available theory, but if the theory proves to be of little help, one may start out by selecting a small number of dimensions. More dimensions may be added at a later time if and when no satisfactory solution can be found.

In our case, two dimensions were used to start with, and these proved to be sufficient. For the initial configuration of the points, the component scores of the 61 objects were utilized, a commonly followed procedure (see Golledge & Rushton 1972, Voogd 1978). The use of the principal component analysis in this context is somewhat paradoxical, because it requires input data of a higher level of measurement than is necessary for the non-

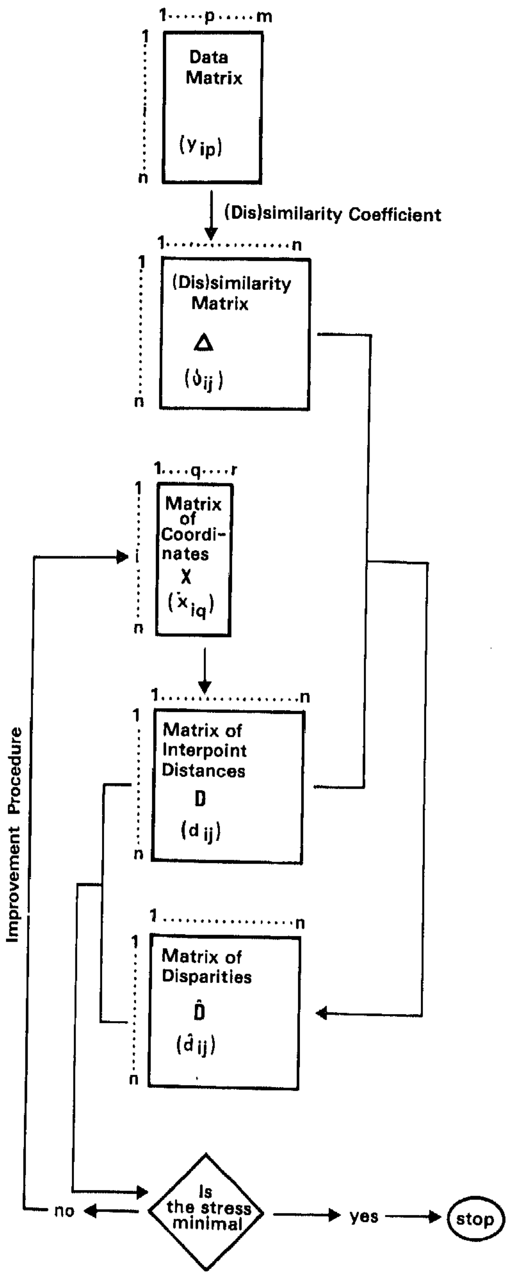


Fig. 5. A flowchart of the non-metric multidimensional scaling analysis.

metric scale analysis. A major drawback of the use of a random configuration as a starting point, however, is that a final result is not reached quickly. The matrix D can be computed from the matrix X with the aid of the geometric distance measure:

$$d_{ij} = \left[\sum_{q=1}^{r} (X_{iq} - X_{jq})^2 \right]^{1/2}$$

A comparison of matrix Δ and matrix D is made to check for stress; that is the state in which the monotonicity requirement is not fulfilled. The monotone relation may be reinstated by checking the diagram (Fig. 4b) systematically from low to high for the disturbing d_{ij} values and calculating the arithmetical average of the consecutive disturbing d_{ij} values. Every old d_{ij} value that contributes to the disturbance is then replaced by the new averages: the disparities $\hat{d}_{ij}$. Thus a matrix $\hat{D}$ is determined. The farther d_{ij} is removed from $\hat{d}_{ij}$, the greater the stress. The difference between d_{ij} and $\hat{d}_{ij}$ may be expressed by means of the normalized stress statistic of Kruskal (1964b):

$$S = \left[\frac{\sum_{i<j} (d_{ij} - \hat{d}_{ij})}{\sum_{i<j} d_{ij}^2} \right]^2$$

The choice of this stress statistic is substantiated by the tests accompanying it to judge the results as good, moderate or poor.

S is invariable when the axes are rotated and when the configuration expands and contracts. S should be as small as possible; if this is not the case, then the coordinates of the objects should be adjusted (X_{iq}). This can usually be achieved by using an iterative improvement process (see Kruskal 1964b). In our analysis the MINISSA program (Roskam & Lingoes 1975) was employed.

Kruskal (1964b) gives the following interpretation of stress. A value of S = 0.025 is very good, a value of 0.05 is good, 0.10 is moderate, and 0.20 is poor (that is, if the number of objects is not too small). If the minimal S in a given analysis is not acceptable, then an attempt may be made to find a more suitable solution by adding one or more dimensions.

The multidimensional scaling method elaborated here is only one of many possible variants. Numerous methods have been developed, all of which are called multidimensional scaling methods. For a discussion of the various procedures the handbooks may be consulted (see Golledge & Rushton 1972, Shepard, Romney & Nerlove 1972); an introductory article such as this is not the place to present a complete overview.

Actually, for some methods, not nearly all the problems have been resolved. Especially the joint space output analyses, wherein both the subjects and the objects are combined in one scale analysis, still generate unresolved difficulties (see Coombs 1964, Harman & Betak 1976, Carrol 1972). In this study only objects (here, centers) are scaled—a simple space output analysis—and the inherent problems are therefore much smaller.

Not many examples of an application of scaling methods to geographical problems can be found in the literature, due to the fact that these methods have only recently been elaborated. Some of the rare examples of the application of scaling methods that were found are the analysis of the decision-making process with respect to inter-urban migration (Golledge & Rushton 1976); the study of preferences for recreation areas (Nijkamp & Veenendaal 1978); the attempt to reveal the cognitive spatial structure of a city (Golledge, Rivizzigno & Spector 1976); Massam's (1978) application in the field of planning, where he uses multidimensional scaling as an aid to finding the best plan among a number of alternatives; and a few other studies (see Klaff 1975, Golledge & Rushton 1972, and Voogd 1977,, 1979).

The[4] results of our application of multidimensional scaling methods to study the service level of the centers in Noord-Brabant may be qualified as good in terms of stress. Based on 42 categories of service functions a two dimensional scale solution was constructed, whereby the stress S is a mere 0.04. The scores of the 61 centers on the two scales is depicted in Figure 6, which also incorporates the results of a cluster analysis. The cluster analysis combines the centers in groups but does not provide any scale values; its sole function is the identification of centers that can be considered to constitute more-or-less homogeneous groups. The grouping procedure itself was based on the dissimilarity matrix, which was also the basis of the scale analysis, and employed the method of Ward (see Deurloo 1976). The grouping was subsequently improved with the RELOC procedure from the CLUSTAN package (see Berkouwer 1978).

The combination of these two separate analyses produces a distinctive pattern. On the basis of their service level four clearly distinct groups could be identified. Eindhoven, Tilburg, and Breda—the largest cities, each with a population of more than 100,000—are grouped together to form a cluster of the highest level (cluster 4); 's-Hertogenbosch, Helmond, Roosendaal, Bergen op Zoom and Oss follow at some distance as cluster 3. Cluster 2, including centers from Waalwijk to Cuyk, shows a substantially

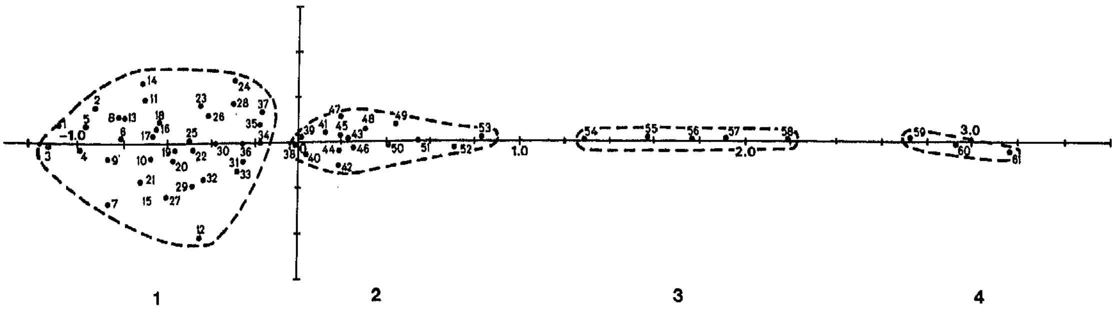


Fig. 6. A two-dimensional scaling analysis and a cluster analysis of 61 centers in Noord-Brabant, for 42 service functions. (For names of places, please consult Fig. 3 on p. 136.)

lower service level. The lower threshold of this cluster lies close to the upper threshold of cluster 1, which combines smaller centers having a population ranging from 5000 to 9500. The two latter clusters are relatively close together; it is less clear than for clusters 3 and 4 that they constitute two separate levels.

Another aspect of the service level of centers is very clearly expressed in Figure 6. Most of the variance in service level is shown on the horizontal scale. Therefore, in a case of only 61 centers, it is legitimate to interpret the service level as a unidimensional phenomenon. This is also expressed in Table 1 where the median number of elements is included for each of the 42 categories of service functions in each of the four clusters of centers. From cluster 1 to cluster 4 the median number of elements increases across the line for each of the categories. This seems to indicate the existence of a general service level whereby the increase in number of elements in one type of function accompanies an increase by growth or introduction in elements of other service functions.

There is a clear relation between the population size of the centers and the general service level. That this is decreasingly true in case of smaller centers is also clearly illustrated. Smaller centers are much more scattered around the horizontal scale, and to adequately describe the service level of these centers, different aspects must be analyzed separately. Our conclusion that more dimensions must be considered when studying smaller centers is fully consistent with the observation of Buursink (1971a) that the empirical validity of a unidimensional hierarchy concept is greatest on a regional scale.

The interpretation of the vertical dimension can be revealed through a comparison between the high and low scoring centers in cluster 1 with respect to their scores on the 42 categories of service functions. Centers with a high positive value on the vertical scale (Zundert, Bladel, St. Willebrord, etc.) also have a relatively large number of elements in one or more of the following service functions: shoes and leather goods (3), apparel (4), textiles (6), furniture (7), housewares (8), recreation articles (18), bank branch offices (22), hairdressers (26) and vocational secondary schools (MAVO and LBO) (32). Centers with a high negative score (Werkendam, Nieuwkuijk, St. Michielsgestel, Rosmalen, and others) have on the average relatively few elements in these categories. This dimension seems to place centers with an over-representation of shopping goods (luxury goods instead of convenience goods) in opposition to centers having an under-representation (note that category 32 (secondary schools) does not fit in).

A separate scale analysis of the 39 centers of cluster 1 combined with 31 others of slightly smaller population size showed the same tendency. Although we hesitate to interpret the second dimension—a profound analysis of the service level is not intended here and would require a more content-oriented analysis and presentation—it is noteworthy that among the centers with a high positive value on the second dimension there are many border towns, and that the centers with a low value include various suburban towns near large cities. Perhaps this gives an indication of the reasons why a (not substantial) variation in the value of the centers on the second dimension is encountered.

To conclude this section a comparison can be made between the results achieved with the multidimensional scale analysis and those arrived at by way of the Guttman scale (see previous section). The advantages of the multi-

Table 1. The median number of service elements in 42 categories in 61 Noord-Brabant centers as grouped in four clusters, as of 1976.

	Cluster 1 (n = 37)	Cluster 2 (n = 16)	Cluster 3 (n = 5)	Cluster 4 (n = 3)
1. Food (frequently bought)	20	38	132	410
2. Food (less frequently bought)	1	3	17	37
3. Footwear and leather wares	3	7	25	65
4. Clothing	6	16	62	114
5. Furs	0	0	3	6
6. Textiles	2	6	29	79
7. Furniture	3	8	33	86
8. Small household goods	5	13	45	114
9. Antiques, curiosa	1	5	19	45
10. Plumbing fixtures	1	1	6	7
11. Department store	0	0	2	5
12. Drugstore	2	4	11	41
13. Health aids/beauty aids	0	0	5	7
14. Books	1	3	9	37
15. Flowers and plants	2	6	18	44
16. Do-it-yourself articles, paint, wallpaper, etc.	4	9	24	56
17. Jewelry, optician	2	5	15	45
18. Recreation goods (incl. sporting goods, toys, photography articles)	4	9	36	80
19. Bicycles, motorbikes	2	4	12	40
20. Café, restaurant, snack bar	15	29	123	308
21. Hotel, amusement enterprise	1	2	11	16
22. Bank branch	3	8	17	60
23. Travel agency	0	1	2	12
24. Dry cleaner, launderette	1	3	8	35
25. Auto repair, service station	5	8	26	150
26. Hairdresser	4	7	33	130
27. Pharmacy	0	2	4	13
28. Center for the elderly, health center, etc.	2	3	6	29
29. Nursing home, hospital	0	1	2	13
30. Physician (general practitioner), dentist	6	12	34	85
31. Nursery school, primary school	10	19	55	122
32. Secondary school (vocational)	2	4	10	35
33. Secondary school (college preparatory)	0	2	5	14
34. College	0	0	4	31
35. University	0	0	0	1
36. Schools of theater, dance, ballet, music; special continuing education institutions	0	1	2	4
37. Library	1	1	2	7
38. Cultural center, community center	3	8	8	35
39. Theater	0	0	1	2
40. Cinema	0	1	4	14
41. Sports center, swimming pool	2	3	8	9
42. Post office, telecommunications	1	2	6	16

dimensional scale analysis may be summarized as follows:

a. The number of groups into which the centers are divided with respect to their service level is derived from the data; the analysis by way of the Guttman scale is not linked so directly to the data base.

b. The multidimensional scaling method provides a partial answer to the question whether or not the service level is a unidimensional phenomenon; the Guttman approach was based on the assumption of unidimensionality.

c. The original data are better utilized. In the first place, all 42 categories of service functions are considered in the analysis. Secondly, multidimensional scaling allows interpretation of the distance between the values; for example, the difference in service level between Oss (54) and 's-Hertogenbosch (58) is just about as large as that between the latter city and Eindhoven (61). The Guttman scale analysis only produced an ordinal measurement. In this regard, it should be noted that even when starting out with ordinal values, a higher level of measurement can be attained through multidimensional scaling (see Shepard, Romney & Nerlove 1972).

Besides the advantages, some limitations and disadvantages of the multidimensional scale analyses must be mentioned. One must limit the number of objects that can be included due to a restricted memory capacity of the computer and huge time consumption. The availability of many (dis)similarity coefficients provides the analysis methods with great flexibility, as has been argued earlier, but can also be considered a disadvantage: different researchers will probably make a different selection and may therefore arrive at a somewhat different final result. Also the interpretation of the various scales that result from the analyses can at times present problems similar to those encountered in factor analysis and component analysis.

The advantages of multidimensional scaling techniques outweigh the disadvantages, as demonstrated in our analysis of the service level of centers. For this type of research problem the use of multidimensional scaling methods is more advisable than the application of the Guttman scale. Nonetheless, if only larger centers are studied and the grouping of centers is not of ultimate consequence, then the application of a Guttman scale need not meet many objections; it may even be preferable because of the formal clarity and simplicity of the analysis. Even when the required assumption of unidimensionality seems unrealistic, the Guttman method can still be utilized by making parallel analyses of the different aspects of the service level, as illustrated above.

Conclusion

In the introduction to this article we proposed that in geography more attention could be given to measurement of the phenomena being studied. Scaling of variables should certainly be considered when complex concepts are employed and attitudes, behavior, etc. are included in the observation and analysis. To this purpose, it is only logical to adopt existing and developing measurement procedures from other social sciences. The measurement of the service level of centers in Noord-Brabant illustrates this. The results clearly show that the use of scaling methods to analyze such a complex concept is feasible and can produce empirically valid results.

In the second section of this article some publications from the fields of geography and planning are briefly mentioned which also deal with the measurement of service level. In these studies ranking of the service level of centers was not attempted through the scaling methods treated here but through specifically designed measurement procedures. It is not our intention to prove (or even to suggest) in this article that the use of scaling methods leads to an improved classification of centers according to their service level, even if such proof could be rendered. The sole purpose of this article is to illustrate the potentials and the limitations of the scaling methods. In doing so, it has been shown to the extent that the material allowed how the use of scaling methods can circumvent a number of methodological problems which were raised in our discussion of the measurement procedures in the literature on the service level of centres. A comparison of scaling methods with other procedures sheds more light on the implications of their use.

It should be stressed that the scaling method we have used to measure the service level of centers involves subjective decision-making. It is obvious that to use scaling procedures choices must be made on many points and that the final results are strongly influenced by such subjective choices. One of those choices which cannot be avoided in any research, not even when standardized methods are used, is the selection of the relevant variables or characteristics. When using the Guttman scale the selection of the items and the dichotomization are additional subjective moments; although these can be justified, they nonetheless remain arbitrary while the final result depends on them. The most noteworthy subjective element of the multidimensional scale analysis as used in this study is the selection of the (dis)similarity coefficient.

Scaling methods, however, have some clear advantages over other procedures. Both the Guttman scale and the multidimensional scaling methods provide more-or-less standardized statistics which can be used to determine how well the constructed scale fits the initial data. These statistics facilitate the necessary assessment of the measurement procedure before the attained results are further analyzed and interpreted. Such an assessment is much less likely if one designs one's own measurement model. Another advantage of the use of the scaling methods is that the level of measurement is unambiguous: either ordinal or interval.

The multidimensional scaling methods provided perspective on the number of dimensions involved in the concept of service level. This was important for our example, and depending on the complexity of the concepts involved, may or may not be relevant for other studies. Multivariate analysis techniques, including the multidimensional scaling methods, provide the easiest way to get a perspective on the number of dimensions contained in a problem — of course, within the context of the selected variables.

Although scaling methods do not remove all obstacles to measurement, the above analysis illustrates that the use of scaling methods to measure complex concepts is feasible and should definitely be included in more Dutch geographic research.

Acknowledgements:

The authors want to thank Mrs. N. van Weesep-Smyth for translating this article.

References:

Berkouwer, Y. (1978), CLUSTAN-IC. Utrecht: Fakulteit der Sociale Wetenschappen, Rijksuniversiteit Utrecht.

BUURSINK, J. (1971a), *Centraliteit en hierarchie; De theorie van de centrale plaatsen in enkele Nederlandse industriegebieden.* Assen: Van Gorcum.

BUURSINK, J. (1971b), De Nederlandse hierarchie der regionale centra. *Tijdschrift voor Economische en Sociale Geografie 62*, pp. 67-80.

CARROL, J. D. (1972), Individual Differences and Multidimensional Scaling. *In*: R. N. SHEPARD, A. K. ROMNEY & S. B. Nerlove, eds., *Multidimensional Scaling.* Vol. 1, pp. 105-155. New York, London: Seminar Press.

COOMBS, C. H. (1964), *A Theory of Data.* New York etc.: Wiley.

DEURLOO, M. C. (1976), Clusteranalyse van kwantitatieve gegevens, uitgewerkt voor een agrarische karakteristiek van de Nederlandse landbouwgebieden. *Tijdschrift voor Economische en Sociale Geografie* 67, pp. 213-229.

DIELEMAN, F. M. (1978), *Een analyse van spreidingspatronen van vestigingen en van werkgelegenheidsgebieden in Tilburg en Eindhoven.* Meppel: Krips Repro.

DIELEMAN, F. M. (1980), Het lokatiepatroon van het winkelapparaat in de stad: experimenten met vierkanten; *Planning, Methodiek en Toepassing*, no. 10, pp. 11-25.

DIJKINK, G. J. W. & E. ELBERS (1981), The Development of Geographic Representation in Children. *Tijdschrift voor Economische en Sociale Geografie* 72, pp. 2-16.

GOLLEDGE, R. G. & G. RUSHTON (1972), *Multidimensional Scaling: Review and Geographical Applications.* Washington, D.C.: Association of American Geographers, Technical Paper no. 10.

GOLLEDGE, R. G. & G. RUSHTON, eds. (1976), *Spatial Choice and Spatial Behavior, Geographic Essays on the Analysis of Preferences and Perceptions.* Columbus: Ohio State University Press.

GOLLEDGE, R. G., V. L. RIVIZZIGNO & A. SPECTOR (1976), Learning about a City: Analysis by Multidimensional Scaling. *In*: R.G. GOLLEDGE & G. RUSHTON, eds., *Spatial Choice and Spatial Behavior*, pp. 95-116. Columbus: Ohio State University Press.

HAMILTON, F. E. I. (1974), *Spatial Perspectives on Industrial Organization and Decision-making.* London: Wiley.

HARMAN, E. J. & J. F. BETAK (1976), Behavioral Geography, Multidimensional Scaling and the Mind. *In*: R. G. GOLLEDGE & G. RUSHTON, eds., *Spatial Choice and Spatial Behavior*, pp. 3-20. Columbus: Ohio State University Press.

HAYS, W. L. (1974), *Statistics for the Social Sciences.* London: Holt, Rinehart & Winston.

HEESEWIJK, W. C. A. VAN (1978), *Hiërarchie van kernen, Theorie en werkelijkheid.* Vught: Bureau voor ruimtelijke ordening B.V.

Herziening typologie van Nederlandse gemeenten (1976). Werkgroep Verstedelijking, Amsterdam: SISWO.

JOHNSTON, R. J. (1976), Residential Area Characteristics: Research Methods for Identifying Urban Sub-areas. *In*: D. T. HERBERT & R. J. JOHNSTON, *Social Areas in Cities.* Vol. 1, *Spatial Processes and Form*, pp. 193-235. London, etc.: Wiley.

KEUNING, H. J. (1971), Spreiding en hiërarchie van de Nederlandse verzorgingscentra op de grondslag van hun winkelapparaat. *Tijdschrift voor Economische en Sociale Geografie* 62, pp. 317-325.

KIESTRA, H. (1978), Het verzorgingsniveau van de Gelderse dorpen. *Geografisch Tijdschrift* 12, pp. 317-325.

KLAFF, V. Z. (1975), Ethnic Patterns of Residence and Socio-Economic Status as Indicators of Social Integration in Israel (Ph. D. thesis, Univ. of Wisconsin, Madison, Wis.).

KRUSKAL, J. B. (1964a), Multidimensional Scaling by Optimizing Goodness of Fit to a Nonmetric Hypothesis. *Psychometrika* 29, pp. 1-27.

KRUSKAL, J. B. (1964b), Nonmetric Multidimensional Scaling: A Numerical Method. *Psychometrika* 29, pp. 115-129.

LLOYD, R. E. (1977), Consumer Behavior after Migration: A Reassessment Process. *Economic Geography* 53, pp. 14-27.

LUYTEN, J. C. L. & V. A. M. VERBERK (1968), De verzorgingsstructuur van het Benelux-Middengebied. *Stedebouw & Volkshuisvesting* 49, pp. 393-412.

MARANELL, G. M. (1974), *Scaling: A Sourcebook for Behavioral Scientists.* Chicago: Aldine.

MASSAM, B. H. (1978), The Search for the Best Alternative Using Multiple Criteria; Singapore Transit Study. *Economic Geography* 54, pp. 245-253.

MEULEN, T. VAN DER (1979), Hierarchy of Centres; Some Notes on a Concept. *Tijdschrift voor Economische en Sociale Geografie* 70, pp. 361-365.

NIE, N. H., C. H. HULL, J. G. JENKINS, K. STEINBRENNER & D. H. BENT (1975), *Statistical Package for the Social Sciences.* New York: McGraw-Hill.

NIJKAMP, P. & W. M. VEENENDAAL (1978), *Psychometric Scaling and Preference Methods in Spatial Analysis.* Research Memorandum. Amsterdam: Economische faculteit, Free University.

NIJKAMP, P. & J. H. VOOGD (1978), *The Use of Multidimensional Scaling in Evaluation Procedures. Methodology and Application to an Industrial Location Problem.* Delft: PSC/TNO.

NUNNALLY, J. C. (1967), *Psychometric Theory*. New York: McGraw-Hill.

ONDERZOEKSKRING UTRECHT O.K.U. (1976), Hiërarchie van kernen in de provincie Utrecht. *Stedebouw en Volkshuisvesting* 57, pp. 275-286.

P. P. D. FRIESLAND (1966), *De ruimtelijke ontwikkeling van het Friese platteland*. Leeuwarden: P. P. D. Friesland.

P.P.D. NOORD-BRABANT (1972), *Verzorgingsstructuur provincie Noord-Brabant*. 's-Hertogenbosch: P.P.D. Noord-Brabant.

ROSKAM, E. E. & J. C. LINGOES (1975), Edinburgh Version of MINISSA.

SEGERS, J. W. G. (1977), *Sociologische Onderzoeksmethoden*. Assen: Van Gorcum.

SHEPARD, R. N. (1972), Introduction to Volume I. *In*: R. N. SHEPARD, A. K. ROMNEY & S. B. NERLOVE, eds., *Multidimensional Scaling*. Vol. 1, pp. 1-20. New York, London: Seminar Press.

SHEPARD, R. N. (1974), Representation of Structure in Similarity Data: Problems and Prospects. *Psychometrika* 39, pp. 373-421.

SHEPARD, R. N., A. K. ROMNEY & S. B. NERLOVE, eds., (1972), *Multidimensional Scaling*. Vol. 1, *Theory and Applications in the Behavioral Sciences*. New York, London: Seminar Press.

SWANBORN, P. G. (1973), *Variabelen en hun meting*. Meppel: Boom.

TORGERSON, W. S. (1958), *Theory and Methods of Scaling*. New York: Wiley.

TOWNROE, P. M. (1972), Some Behavioral Considerations in the Industrial Location Decision. *Regional Studies* 6, pp. 261-272.

Typologie van de Nederlandse gemeenten naar urbanisatiegraad, 31 mei 1960 (1964), Zeist: Centraal Bureau voor de Statistiek.

UTRECHT, L. C. (1975), *Schaalconstructie I. Inleiding tot de eendimensionale modellen van Thurstone, Coombs and Guttman*. Amsterdam: Psychologisch Laboratorium, Universiteit van Amsterdam.

VOOGD, J. H. (1977), *Een overzicht van meet- en schalingsmethoden ten behoeve van het kwantificeren van preferenties*. EVAPLAN, tussenrapport 1. Delft: PSC/TNO.

VOOGD, J. H. (1978), *On the Principles of Ordinal Geometric Scaling*. Research Paper 10. Delft: PSC/TNO.

VOOGD, J. H. (1979), *Ordinal Data Analysis in Urban and Regional Planning*. Paper 7th European Symposium on Urban Data Management. The Hague: publisher unknown.

VOORTMAN, A. H. (1961/62), Ontwikkeling en toepassing van een methodiek ter afleiding van patronen van verzorgende centra. *Tijdschrift voor Economische en Sociale Geografie* 52, pp. 309-318 en 53, pp. 8-14.