HOMOGENEITY ANALYSIS: EXPLORING THE DISTRIBUTION OF VARIABES AND THEIR NONLINEAR RELATIONSHIPS

Willem J. Heiser
Jacqueline J. Meulman

Department of Data Theory
University of Leiden

Homogeneity Analysis: Exploring the Distribution of Variables and Their Nonlinear Relationships

Willem J. Heiser and Jacqueline J. Meulman

Department of Data Theory,
Faculty of Social Sciences, University of Leiden,
P.O. Box 9555, 2300 RA Leiden, The Netherlands

Draft for chapter in *"Correspondence Analysis in the Social Sciences: Recent Developments and Applications"*
Edited by: Michael Greenacre, Jörg Blasius, and Walter Kristof

Contents

Homogeneity Analysis:
Exploring the Distribution of Variables and Their Nonlinear Relationships

Willem J. Heiser and Jacqueline J. Meulman

Abstract

Homogeneity analysis is the collective name for a group of exploratory techniques for nonlinear multivariate analysis. This chapter emphasizes their role in the study of reliability of measurement by presenting them in a distributional framework. A set of variables is called homogeneous if they have a single center. Loss of homogeneity is defined as a variance, which measures random deviations of the optimally transformed variables from their center. Alternating Least Squares (ALS) is briefly introduced as a flexible computational strategy that can be used for all methods presented.

Principal components analysis is discussed as a method of homogeneity analysis, by showing the connection of loss of homogeneity with the mean squared correlation of all variables with their center, and with the eigenvalues of the correlation matrix. There is also a connection with Cronbach's α, the standard measure of internal consistency that is equal to the mean of all possible split-half reliability coefficients.

Nonlinear transformations of the variables may be incorporated in loss of homogeneity to account for a greater part of the total variability. Using Huygens' Theorem, it is shown that transformation and weighting can be separated. During each ALS cycle, the center, or a linear combination of multiple centers, has to be regressed on the space of transformations. This space can be independently chosen for each variable, according to various specifications (implemented in the PRINCALS program). Homogeneity analysis in a strict sense (a method and program called HOMALS) is obtained when all variables are nominal, and is equivalent to multiple correspondence analysis. PRIMALS is a method to study the distribution of the variables before and after they have been transformed by maximizing the primary eigenvalue.

Various forms of analysis are illustrated by applying them to data from a Dutch survey on opinions towards controversial issues. A permutation test for choosing the number of eigenvalues to be maximized is proposed. For these data, clear evidence can be obtained for a bimodal distribution.

Key words and phrases: optimal scaling; quantification methods; transformation of variables; Gifi system; nonlinear multivariate analysis; principal components analysis; multiple correspondence analysis; Cronbach's α; permutation tests.

Homogeneity Analysis:
Exploring the Distribution of Variables and Their Nonlinear Relationships

Introduction

When several variables are measured on the same group of individuals, one of the first questions that needs an answer is: how to distinguish the reliable variability from the unreliable variability? As will become evident shortly, this question can be approached by phrasing it as a problem of finding out what the distribution of the variables is, a matter that is of concern independently from considerations about the distribution of individuals. The discussion starts with a key attribute of empirical distributions, their homogeneity.

Homogeneity of individuals and variables

The term homogeneity is predominantly used in statistics in connection with samples from different populations, which may – or may not – exhibit identical behavior, or display similar characteristics. If the populations are the same for the purposes of the study, they are collectively called *homogeneous*, meaning *of one kind*, and different random samples from homogeneous populations will be called homogeneous as well. A first thing to note is, that populations and subpopulations can be homogeneous in one respect and heterogeneous in another, as for example in the idealized circumstances of the familiar t-test, where it is assumed that the observations are drawn from two populations that are homogeneous in their variance and heterogeneous in their means.

Populations of individuals that are heterogeneous in their means are so common (or perhaps, believed to be common) in the social and behavioral sciences that they form the almost exclusive object of study, both in theoretical and in empirical research. To fully appreciate the truth of this remark, we have to make a slight digression to look at the distinct *aggregation levels* at which any variable can be studied.

In an operational sense, a 'variable' refers to some specified *rule* that assigns values to

individuals, like when one calls a variable 'school achievement' while in fact grade point averages of a number of students are being studied. But for purposes of analysis, individuals can be grouped into school classes, or into schools, into school districts, or into other higher-order aggregates, and the grouping defines the analysis variable. By determining the mean and the variance, and perhaps other distributional properties like the skewness or an extra bump in the tail of the distribution, we express our knowledge at the group level, and thereby characterize a certain population of students. From such a choice of aggregation level it follows, that not much more can be said about any *particular* student than that it is a member of a population with such-and-such characteristics. However, often a social or behavioral scientist is not just interested in the distribution properties at the highest aggregation level, or in a single, *a priori* chosen, grouping, because the concept of a variable is also linked with the idea of a *measuring instrument*, or a *scale*, on which the individuals have definite fixed scores, known up to some *measurement error*.

One tends to think about measurement error as an unpredictable influence that is unavoidable, albeit hopefully small, or at least not so large as to be forced to confuse for instance the ability of a person at one end of the scale with the ability of a person at the other end of the scale. It is one of the central tenets in social science methodology that it should be possible to cut back on the aggregation level by distinguishing subgroups of individuals – or subgroups of individual measurements – along the measurement scale. Such differently located subgroups constitute a translation family of populations, and that is the reason why we can say that the assumption of populations being heterogeneous in their means is so common.

The availability of a reliable measurement scale is prerequisite if we try to answer questions such as whether females are brighter than males, whether different teaching instructions are really effective or not, whether a distinction in social environment is noticeable on the scale or not, and so on. Subgroups may be defined by different experimental treatments or interventions, such as teaching instructions, or by an observable background characteristic, such as sex or age. In other situations it may be desirable to find the *right* characteristic from a number of plausible candidates. Since the distinction in background variable and response variable (similar contrasting terms are: exogenous – endogenous, or independent – dependent) is a distinction in role, not in substance,

one cannot say once and for all which variable is of one type and which of the other. Even sex and age can be response variables, as for instance in a demographic study, where regional characteristics are the background variables. Conversely, a typical response variable like endorsing a statement of opinion might serve as a background characteristic in a study of political voting.

Is it possible to reduce the uncertainty that follows from the measurement error on the response scale, or in the background characteristic, and thereby to improve the chances of convincingly demonstrating the existence of subgroups that are heterogeneous in their means? An old answer to this recurring question is to *average* over a number of *parallel* measurements, the idea being that errors on different versions of the variable will tend to cancel one another out (Spearman, 1904). When things really matter, we do not evaluate a student on a single grade, but on a grade point average. The remarkable concept here is that we consider replication across 'variables', not across individuals! Ideally, the parallel measurements are equal in the location of each individual (or of subgroups of them), and the actual data are expected to vary randomly around these equal values. The average location will tend to have smaller variance than any of the single measurements on which it is based, provided that the average is taken over a homogeneous set of variables. Because the psychometric notion of parallel measurement involves a distribution of scores across variables (in sociology the term 'indicator' denotes the same idea of using a variety of observables to pin down some theoretical construct), we can talk about homogeneous samples and populations of variables. By implication, samples of variables can be *heterogeneous* too, in which case it would not seem to be very sensible to expect improvement of reliability by averaging.

How to find out whether a set of variables is homogeneous or heterogeneous? There are a number of answers to this question, and one of them is – unsurprisingly – to do a homogeneity analysis. Although the primary aim of this type of analysis is to characterize a distribution of variables, it always *also* leads to a characterization of individuals, because homogeneity analysis relies on a common basis of comparison in which samples of individuals are maximally heterogeneous.

We have seen that the idea of a statistical distribution and related concepts can be applied to sets of individuals and sets of variables alike. It is now time to define more precisely what is meant by

'parallel measurement' or 'homogeneous sample of variables'. Indeed, various definitions are possible, and it is helpful to order them from simple to more complex, which then yields a natural ordering of the methods of homogeneity analysis.

Forms of homogeneity

Since homogeneity is a concept that describes the relationship between the elements of a distribution, our task is to specify in what sense variables can be different while still being of the same kind. A distribution is a multidimensional concept, because it has various aspects, like a centrality, a dispersion, and perhaps irregularities of shape, as noted before. A variable is a multi-dimensional concept too, because it specifies the different locations of (groups of) individuals. To keep the complexity within bounds, the following concepts and definitions are very useful:

(a) a sample of variables is called homogeneous if its elements (i.e., the variables) are equal, up to some prespecified class of information-preserving operations, and up to random deviations;

(b) information-preserving operations are operations like changing the mean, rescaling, taking the logarithm, and so on, that are considered to leave the information on the individuals, carried by the variable, unchanged;

(c) random deviations among the variables are measured through a loss function, in which they are compared with one another, or with a latent variable, which is the center of the distribution.

The simplest, and most restrictive type of information-preserving operation is the identity, which keeps any variable the same, and in this pure form the loss function can be expressed as follows:

$$\sigma^2(\mathbf{x}) = (Nm)^{-1} \sum_j \| \mathbf{h}_j - \mathbf{x} \|^2 . \qquad (1)$$

Here $\mathbf{h}_j$ is an N-vector containing the observed scores of N individuals (also called *objects*) on variable j, with $j = 1,...,m$, and the function $\| \, . \, \|$ is the Euclidean norm, which is the square root of the sum of squares of its elements. The common basis of comparison is the unknown variable $\mathbf{x}$, a vector of the same length, for which the scores of the N individuals, called *object scores*, are to be determined. In classical psychometrics, the object scores are called *true scores*, corresponding to the conceptualization of the random deviations $(\mathbf{h}_j - \mathbf{x})$ as *errors*. Geometrically, the value of $\sigma^2(\mathbf{x})$ – called *loss of homogeneity* – is the mean squared Euclidean distance between the vectors

$\mathbf{h}_1,...,\mathbf{h}_j,...,\mathbf{h}_m$ and $\mathbf{x}$. Loss of homogeneity is the multidimensional analogue of the variance; it measures the departure from $\mathbf{x}$, the multidimensional analogue of the mean.

Indeed, the minimum loss of homogeneity, denoted as $\sigma^2(*)$, is attained for $\mathbf{x} = \bar{\mathbf{h}} = m^{-1} \sum_j \mathbf{h}_j$, the mean across variables for each object (individual). Here, as elsewhere in the paper, rigorous proofs are omitted; the interested reader is referred to Nishisato (1980) or Gifi (1990) for more details. The minimal value turns out to be

$$\sigma^2(*) = (Nm)^{-1} \sum_j \| \mathbf{h}_j - \bar{\mathbf{h}} \|^2 = N^{-1} [m^{-1} \sum_j \|\mathbf{h}_j\|^2 - \|\bar{\mathbf{h}}\|^2], \tag{2}$$

up to a factor N the mean squared length of the $\mathbf{h}_j$ minus the squared length of their mean – analogous again to the familiar pocket calculator formula for the variance. It is possible to simplify $\sigma^2(*)$ still further, if it is assumed – as will be done in the remainder of the chapter – that all variables are in deviations from their own mean, and standardized as $s^2(\mathbf{h}_j) = N^{-1}\|\mathbf{h}_j\|^2 = 1$; in that case, $\sigma^2(*) = 1 - r_{..}$, where $r_{..}$ is the average correlation between all $\mathbf{h}_j$, including the self-correlations equal to 1, and all correlations r_{ij} are counted twice. Thus minimal loss of homogeneity attains its lowest value if all variables are perfectly correlated ($\sigma^2(*) = 0$), and it is maximal if the variables are altogether uncorrelated ($\sigma^2(*) = 1 - 1/m$).

Now suppose that we allow for rescaling of the variables. There can be a variety of reasons for taking this step, for instance because it is expected that the variables are different in their power to discriminate the objects and are therefore to be assigned a different weight in the analysis. In this case it is natural to modify the definition of loss of homogeneity into

$$\sigma^2(\mathbf{a},\mathbf{x}) = (Nm)^{-1} \sum_j \| a_j\mathbf{h}_j - \mathbf{x} \|^2 , \tag{3}$$

where each variable is rescaled by the scaling factor a_j. The coefficients a_j are the elements of the m-vector $\mathbf{a}$ that is included in the list of unknowns in $\sigma^2(\mathbf{a},\mathbf{x})$. Note that we do not restrict a_j in any way; if $a_j < 0$, then the scores in $\mathbf{h}_j$ change sign, a phenomenon called *reflection*; if $a_j = 0$, the jth variable drops out of the analysis. The only type of restriction that ought to be considered here consists of normalization to identify $\mathbf{a}$ and $\mathbf{x}$, e.g. $\|\mathbf{a}\|^2 = m$. Before continuing the discussion of rescaling, a number of introductory remarks on computation are in order.

Alternating least squares

Although computation is of no great concern in this chapter, the question of determining the optimal value of $\sigma^2(\mathbf{a},\mathbf{x})$ provides an excellent opportunity to introduce Alternating Least Squares (ALS), which is a very flexible computational strategy for many methods in multidimensional data analysis. Briefly, suppose we fix the a_j at some particular value $\mathbb{a}_j$ (for j=1,...,m), obtained in a previous round of computation. Then actually rescaling the variables as $\mathbb{q}_j = \mathbb{a}_j\mathbf{h}_j$ shows that our best current guess $\mathbb{x}$ of $\mathbf{x}$, which still is the average across variables, becomes equal to

$$\bar{\mathbb{q}} = m^{-1} \Sigma_j \mathbb{q}_j = m^{-1} \Sigma_j \mathbb{a}_j\mathbf{h}_j \,, \tag{4}$$

i.e., the *weighted mean* of the original variables. Because of their role in (4), the scaling factors a_j are also called *weights*. Fixing $\mathbf{x} = \bar{\mathbb{q}}$ in turn, the ALS principle assures us that the conditional minimum of $\sigma^2(\mathbf{a},\bar{\mathbb{q}})$ over $\mathbf{a}$ always improves (decreases) loss of homogeneity. The conditional minimum is attained by choosing the new scaling factors $\hat{\mathbb{a}}_j$ as

$$\hat{\mathbb{a}}_j = c(\mathbf{h}_j,\bar{\mathbb{q}}) \,/\, s^2(\mathbf{h}_j) \,, \tag{5}$$

where $c(\mathbf{h}_j,\bar{\mathbb{q}})$ denotes the covariance between $\mathbf{h}_j$ and $\bar{\mathbb{q}}$, and $s^2(\mathbf{h}_j) = N^{-1}\,\|\mathbf{h}_j\|^2$ the variance of $\mathbf{h}_j$ (which was assumed to be equal to one, but which is included in (5) to be fully explicit). Alternating between (4) and (5) yields a convergent process. Note that $\hat{\mathbb{a}}_j$ in (5) is simply the regression coefficient for the linear regression of $\bar{\mathbb{q}}$ on $\mathbf{h}_j$. In order to avoid a solution with $a_j = 0$ for all variables and, correspondingly, $\mathbf{x} = \mathbf{0}$, which trivially minimizes $\sigma^2(\mathbf{a},\mathbf{x})$, some convention on *normalization* is required; usually, one fixes either the sum of squares of the object scores or the sum of squares of the scaling factors at some prechosen value. Application of alternating least squares generally requires that the loss function can be split into independent components, and several of such decompositions will be demonstrated in the sequel.

Principal components analysis as a method of homogeneity analysis

Optimally rescaling the variables before averaging determines another center of the distribution than directly averaging them, but what is so special about it for the study of homogeneity? This question

will be answered by discussing the loss of homogeneity function in more detail.

The mean squared correlation and the eigenvalue

Frequently, the elements of $\mathbf{x}$ in (3) are scaled as z-scores. Since the latter normalization satisfies $s^2(\mathbf{x}) = 1$, substitution of the optimal weights (5) in (3) then allows us to write the loss function as

$$\sigma^2(*,\mathbf{x}) = 1 - m^{-1} \Sigma_j r^2(\mathbf{h}_j,\mathbf{x}) , \tag{6}$$

where $r(\mathbf{h}_j,\mathbf{x})$ denotes the correlation between $\mathbf{h}_j$ and $\mathbf{x}$. So under the rescaling definition of loss (3), the center of the distribution of variables, $\mathbf{x}$, is chosen in such a way that it has maximal mean squared correlation with the original variables; this center – generally different from the unweighted mean – is called the (first) *principal component*. The correlation between $\mathbf{h}_j$ and $\mathbf{x}$ is often called the *loading* of variable j on the principal component. The squared loading then is the *variance* of $\mathbf{h}_j$ that can be *accounted for* by $\mathbf{x}$ in linear regression terms, and the mean squared correlation in (6) is the average variance accounted for.

Just as one can go from (3) to (6) by substituting the conditionally optimal weights, it is possible to re-express (3) in a form that does not involve the principal component. First, we switch from normalization of the component to normalization of the weights, so that $\|\mathbf{a}\|^2 = m$, while the conditionally optimal value of $\mathbf{x}$ remains exactly as indicated in (4). Next, it is not difficult to verify a fundamental identity in Euclidean space (Gower, 1975), written in the present notation as

$$(Nm)^{-1} \Sigma_j \| a_j\mathbf{h}_j - \mathbf{x} \|^2 = (2m)^{-1}(Nm)^{-1} \Sigma_j\Sigma_l \| a_j\mathbf{h}_j - a_l\mathbf{h}_l \|^2 , \tag{7}$$

which expresses the fact that the dispersion of the rescaled variables may be formulated either with respect to the center $\mathbf{x}$, or in terms of the sum of the squared distances between all of them taken in pairs; the two formulations differ only by a factor $2m$. As a final step, using the fact that $c(a_j\mathbf{h}_j,a_l\mathbf{h}_l) = a_ja_lr(\mathbf{h}_j,\mathbf{h}_l)$, the right-hand side of (7) may be rewritten to obtain

$$\sigma^2(\mathbf{a},*) = 1 - m^{-2} \Sigma_j\Sigma_l a_ja_lr(\mathbf{h}_j,\mathbf{h}_l) . \tag{8}$$

So weights that minimize loss of homogeneity will maximize the weighted mean correlation among all pairs of variables. From (8) it becomes clear how the value $1 - r_{..}$ was obtained earlier for the

case of all weights equal to one. When optimal weights $\hat{a}_j$ and $\hat{a}_l$ are inserted in the weighted mean correlation, the quantity

$$\lambda^2 = m^{-1} \Sigma_j \Sigma_l \hat{a}_j \hat{a}_l r(\mathbf{h}_j, \mathbf{h}_l) \; . \tag{9}$$

is called the (maximum, or first) *eigenvalue* of the correlation matrix, i.e. the mxm matrix that consists of elements $r(\mathbf{h}_j,\mathbf{h}_l)$. The eigenvalue expresses the homogeneity of the variables in terms of their mutual correlations. Inserting (9) into (8), and comparing the result with (6), we reach the important conclusion that $m^{-1}\lambda^2$ is *also* equal to the mean squared correlation of the variables with their principal component.

Cronbach's α

Loss of homogeneity was expressed in (8) without reference to a center, but directly as a weighted mean of the correlations between the variables, for any set of weights satisfying the normalization constraint $\|\mathbf{a}\|^2 = m$. If all correlations are one, this weighted mean is bounded by

$$m^{-2} \Sigma_j \Sigma_l \, a_j a_l \; = (m^{-1} \Sigma_j \, a_j)(m^{-1} \Sigma_l \, a_l) = \bar{a}^2 \leq m^{-1} \|\mathbf{a}\|^2 = 1 \; , \tag{10}$$

where the inequality follows from the elementary fact that $\Sigma_j \, (a_j - \bar{a} \,)^2 \geq 0$. The average weight, $\bar{a}$, can never be larger than one, and it can only become equal to one if all weights become equal to one. In that case, $\sigma^2(\mathbf{a},*)$ attains its natural minimum of zero. When all correlations are zero, it follows from (8) that $\sigma^2(\mathbf{a},*)$ becomes $1 - m^{-1}$, for any choice of $\mathbf{a}$, where the presence of m^{-1} reflects the inclusion of $j = l$ in the summation. Because $\sigma^2(\mathbf{a},*)$ involves correlations, which are bounded by -1.0 and 1.0, and weights with a fixed sum of squares, it is natural to consider a related measure that itself has the properties of a correlation. Multiplying $\sigma^2(\mathbf{a},*)$ by the factor $m/(m - 1)$ corrects for the undesired maximum, and after reflection the quantity

$$\alpha(\mathbf{a}) = [\; 1 - m/(m - 1) \; \sigma^2(\mathbf{a},*) \;] \; / \; [\; 1 - \sigma^2(\mathbf{a},*) \;] \tag{11}$$

has the desired properties of a correlation coefficient. By substitution of (8) and some algebra on the sum of the correlations, (11) becomes

$$\alpha(\mathbf{a}) = [\ (m(m-1))^{-1}\ \Sigma_j\Sigma_{l\neq j}\ c(a_j\mathbf{h}_j, a_l\mathbf{h}_l)\ ]\ /\ s^2(m^{-1}\Sigma_j a_j\mathbf{h}_j)\ , \qquad (12)$$

i.e., the mean covariance among the weighted variables, *excluding* the variances, divided by the variance of the mean scores. In this form, or in a number of similar forms, the coefficient in (12) is best known as *Cronbach's* α. Originally, it was defined for binary variables only, but it soon became obvious that this limitation was unnecessary. Rather than deriving α from (strong) assumptions, Cronbach (1951) settled upon a definition and studied its properties. The coefficient is written here as a function $\alpha(\mathbf{a})$ to emphasize its dependence on the weights. In those days, weights were determined by a variety of methods; they were not frequently optimized, as in (9). Therefore, coefficient α was welcomed as a convenient measure to evaluate different weighting schemes or different selections of variables, most frequently test items, for computing *total scores*.

Although Cronbach's α is probably the most frequently used omnibus coefficient in applied psychometrics, it has a number of interpretations that are less known than they should be. The primary interpretation follows from (12): it is a measure of homogeneity or *internal consistency*. If one includes uncorrelated variables, the value of α drops; if one adds correlated variables, the value of α increases. Cronbach (1951) gave a number of additional properties that refer more specifically to Spearman's (1910) *split-half* approach for determining the reliability of a series of measurements, in which the battery is rescored, half the variables at a time, to get *two* estimates of the mean score. Split-half reliability is then defined as the correlation between these two sets of scores, corrected with the so-called Spearman-Brown formula, which accounts for the fact that only half of the information is used in estimating the mean score. The conventional split-half approach had been criticized because of its lack of uniqueness, due to the many different ways in which one can split m variables into two sets. However, Cronbach (1951) showed that α is the mean of all possible split-half coefficients if all splits are weighted equally. Moreover, he showed that α is the expected value of the correlation between the object scores in two independent random samples of length m, under fairly general regularity conditions on the distribution of variables (e.g., unimodality), and with the same weighting. These remarkable properties of α (and similar coefficients) relate the early psychometricians' concern with reliability to present-day statistical practice of estimating the variance of a statistic by cross-validation.

Considering α is relevant for homogeneity analysis, because homogeneity is determined by minimizing $\sigma^2(\mathbf{a},*)$ over $\mathbf{a}$, and it may be seen from (11) that $\alpha(\mathbf{a})$ increases when $\sigma^2(\mathbf{a},*)$ decreases. According to Nishisato (1980, p. 100), it was Lord (1958) who first demonstrated that maximization of $1 - \sigma^2(\mathbf{a},*)$ leads to maximization of $\alpha(\mathbf{a})$. Simple calculations with (8) and (9) show that, for optimal $\hat{\mathbf{a}}$, the value of $\alpha_{max} = \alpha(\hat{\mathbf{a}})$ can be expressed in terms of the eigenvalue as

$$\alpha_{max} = m\,(\lambda^2 - 1) / (m\text{-}1)\,\lambda^2. \tag{13}$$

The relationship between α_{max} and the eigenvalue is nonlinear and monotonically increasing. Two examples are plotted in Figure 1, for $m = 6$ and $m = 11$, to give an impression of the rather severe nonlinearity when m grows. Note that α_{max} would become negative if $m^{-1}\lambda^2 < m^{-1}$. However, it

---
Insert Figure 1 about here
---

can be shown that this condition cannot occur for the largest eigenvalue, which is the reason that each curve in Figure 1 starts with a value of m^{-1} on the horizontal axis. Clearly, especially for large m the interpretation of α_{max} and $m^{-1}\lambda^2$ must be carefully adjusted to their different ranges.

Summarizing, the result of a homogeneity analysis – as defined so far – can be given in terms of the weights or in terms of the object scores, and the measure of homogeneity can be given either as a variance, $\sigma^2(*,*)$, or as a property of the correlation matrix of the rescaled variables, $m^{-1}\lambda^2$, which is a mean, or as an upper bound to Cronbach's α, which is a correlation.

Irregularities of shape: more principal components

A uniform distribution of variables corresponds to correlations of about equal size, so that $r_{..}$ is sufficient to describe it. Non-uniform distributions can be classified by the number of regions of high density. With only one such region, the distribution is called *unimodal*. There will be a characteristic pattern of higher and lower correlations, called the *Spearman hierarchy*, and the eigenvalue will satisfactorily describe relative homogeneity, in so far as $m^{-1}\lambda^2$ is close to one.

When $m^{-1}\lambda^2$ is not close to one, it may be rewarding to look for multiple solutions. Apart from possible multimodality, there is also the possibility that some of the variables may have got small – maybe even close to zero – weights in the first principal component, so that initially they do not

really enter into the analysis. In the latter case, interest may be in finding out whether a second principal component yields larger weights for the neglected variables – an indication that the total set is distributed around a plane, rather than around a single central direction in m-space, or around a number of scattered directions. Deviations from uniformity and unimodality are instances of what was called earlier *irregularities of shape* of a distribution, and in practice they turn out to be the rule, rather than the exception.

In one way or another, the characteristics of the first weighting scheme should not reappear in the second weighting scheme (and perhaps in further ones), or at least such reoccurrences are to be avoided as much as possible. This objective can be translated into the formal analysis by requiring that the components are uncorrelated, i.e. $c(\mathbf{x}_s,\mathbf{x}_t) = 0$, where $\mathbf{x}_s$ is a particular principal component and $\mathbf{x}_t$ another one. Now, consider the minimization of

$$\sigma^2(\mathbf{A},\mathbf{X}) = (Nmp)^{-1} \Sigma_j \Sigma_s \| a_{js}\mathbf{h}_j - \mathbf{x}_s \|^2 , \qquad (14)$$

in which $s = 1,...,p$, with p the number of components sought, where $\mathbf{X}$ is an Nxp matrix of object scores, with columns $\mathbf{x}_s$, and $\mathbf{A}$ is an mxp matrix of weights $\{a_{js}\}$ for variable j with respect to principal component s (see (3), where p=1). Various algorithms exist for finding multiple principal components, but they all lead to the same result, up to a scaling factor. As before, the scale of $\mathbf{x}_s$ is chosen by convention to identify a solution – a choice that also determines the range of $\sigma^2(\mathbf{A},\mathbf{X})$, while leaving the solution essentially the same, because the coefficients in $\mathbf{A}$ get uniformly adjusted when $\mathbf{x}_s$ is scaled differently.

The optimal choice of a_{js} still is the covariance between $\mathbf{h}_j$ and $\mathbf{x}_s$ divided by the variance of $\mathbf{h}_j$, analogous to (5), and substitution in (14) with $s^2(\mathbf{h}_j) = 1$ and $s^2(\mathbf{x}_s) = 1$ shows that multiple principal components are uncorrelated variables $\mathbf{x}_s$ that optimize the function

$$\sigma^2(*,\mathbf{X}) = 1 - (mp)^{-1} \Sigma_j\Sigma_s r^2(\mathbf{h}_j,\mathbf{x}_s) , \qquad (15)$$

again one minus the mean squared correlation, but now also averaged across components. The components can be ordered by the size of their eigenvalues, which are (cf. equations (8) and (9)) equal to the mean squared $r(\mathbf{h}_j,\mathbf{x}_s)$ per component. When the squared correlation is averaged across

components, a measure called the *fit per variable* is obtained, which indicates how much each variable contributes to the analysis. This remark on the alternative decomposition of the total variance concludes the discussion of *linear* principal components analysis (PCA) as a method for studying the homogeneity of variables.

Nonlinear transformations of the variables

Linear PCA allows for linear transformations of the variables to optimize their homogeneity. All lack of homogeneity necessarily is identified as random deviation. One way to split off further systematic components from the total variability is to allow for nonlinear transformations of the variables, which can often *also* be justified on theoretical grounds, or on characteristics of the observational setting.

Scaling and nonlinearity

A lot of measurements in the social and behavioral sciences are recorded on a scale with uncertain unit of measurement. In case of a five-category Likert item, for example, it is usually quite arbitrary whether we should assign scores {1,2,3,4,5} or {–3,–1,0,1,3} to the categories {'strongly disagree', 'disagree', 'neutral', 'agree', 'strongly agree'}. Therefore, both the origin of the scale and the distance betweeen consecutive values is uncertain. Another example is time as a response variable: in studies of attention and memory one may use *reaction time* as the behavioral response, and an important aspect of emotional processes is their *duration*. Of course, one measures time in milliseconds or in minutes, but the psychological calibration of time may be different from the series of unit intervals on the dial of a clock. Therefore, it is often appropriate to use a log scale, or some other nonlinear transformation of physical time. A frequently used background variable is *age*, measured in years; here, nonlinearities arise because a lot of developmental processes have typical patterns of acceleration and deceleration. Many achievement variables first improve, then level off, and eventually deteriorate with age.

When such nonlinearies are a possibility, the uncertainty in the unit of measurement is not just a matter of measurement error, because its variability may have a systematic component. On a log

scale, for example, a distance at the lower end is longer than a distance at the upper end, compared to unit distances on the original scale. Now suppose that all observed variables are in fact different transformations of the same basic variable; then there must exist *inverse* transformations $\phi_j(\mathbf{h}_j)$ that make them equal again. Instead of a simple rescaling $a_j\mathbf{h}_j$ we thus consider a one-to-one mapping $\phi_j(\mathbf{h}_j) = \mathbf{q}_j$ of the original observations $\{h_{ij}\}$ to new *quantifications* $\{q_{ij}\}$. The mapping ϕ_j allocates to each different value of variable $\mathbf{h}_j$ a new value that can be chosen as to minimize

$$\sigma^2(\phi_1 ... \phi_m, \mathbf{x}) = (Nm)^{-1} \Sigma_j \| \phi_j(\mathbf{h}_j) - \mathbf{x} \|^2 . \qquad (16)$$

In this extended, nonlinear definition of homogeneity, the class of information preserving operations is determined by the specification of ϕ_j, which in the general case is a nonlinear mapping, and which can be chosen independently for each variable. Note that explicitly taking the logarithm is not included in formulation (16); that would simply involve the preliminary recoding $\mathbf{h}_j \Leftarrow \ln \mathbf{h}_j$ and does not form a family of transformations. However, if the homogeneity of variables can be improved by increasing the distance between smaller values of a particular variable $\mathbf{h}_j$ and by decreasing distances between larger values of the same variable, and if the class of transformations is broad enough to allow for such changes, then a plot of the transformed values $\phi_j(\mathbf{h}_j)$ against the original values $\mathbf{h}_j$ may well reveal a logarithmic type of function for ϕ_j. Thus homogeneity analysis suggests useful transformations on *a posteriori* grounds, by considering equivalent forms of the variables, and then selecting precisely those that yield a distribution with minimal dispersion.

Finding the best transformation by regression

It will be clear that the center $\mathbf{x}$ of the distribution equals the mean of the transformed variables $m^{-1} \Sigma_j \phi_j(\mathbf{h}_j)$, just as the optimal location under simple rescaling was the mean of the weighted variables (see (4)). But it is not known beforehand what transformation would be best to take. Using the ALS principle, a conditionally optimal transformation can be found for each variable separately, given the best current guess $\mathbb{x}$ of $\mathbf{x}$, and keeping the other variables constant at their current values $\mathbb{q}_l$, because then (16) can be decomposed into a constant and a variable term:

$$\sigma^2(\phi_j) = constant + (Nm)^{-1} \| \phi_j(\mathbf{h}_j) - \mathbb{x} \|^2 . \qquad (17)$$

Minimizing $\sigma^2(\phi_j)$ over all possible choices of ϕ_j is a regression problem, in which the *estimate* $\mathbb{x}$ is regressed on the space of transformations of the *data* $\mathbf{h}_j$.

In the previous section, the case $\phi_j(\mathbf{h}_j) = a_j\mathbf{h}_j$ was already considered, which corresponds to linear regression without an intercept, and which is appropriate if – but not only if – the measurements are recorded on a ratio scale. Next, the case $\phi_j(\mathbf{h}_j) = a_j + b_j\mathbf{h}_j$, appropriate if – but not only if – one deals with an interval scale, can be solved by linear regression with an intercept, and when $\phi_j(\mathbf{h}_j) = a_j + b_j\mathbf{h}_j + c_j\mathbf{h}_j^2$, formulation (17) leads to polynomial regression. Yet another possibility would be to choose piecewise polynomials, also called *splines*, which very naturally adjust themselves to nonlinear relationships, while requiring only a modest number of parameters (Winsberg and Ramsay, 1983). After all variables have been processed according to their prespecified class of transformations, yielding current values $\mathbb{q}_j$, the ALS principle tells us to continue with an updated guess $\mathbb{x}$ equal to the standardized version of $m^{-1}\,\Sigma_j\,\mathbb{q}_j$, unless this set of scores is close enough to the previous one to stop the process.

Whatever type of regression is chosen, in the majority of cases it will be possible to isolate a scaling factor from the transformed variable, so that we can write $\mathbf{q}_j = a_j\phi_j(\mathbf{h}_j)$ with $s^2(\phi_j(\mathbf{h}_j)) = 1$. Therefore, the interpretation of loss of homogeneity in (6) is still possible: the transformed variables will maximize the mean squared correlation with their principal component. Likewise, the interpretation in terms of the eigenvalue and of Cronbach's α remains valid. The next question is: what type of regression is used in the treatment of variables – like the five-category Likert item mentioned earlier – that are *non-numerical*, such as nominal and ordinal variables?

Nominal variables: intra-class regression

A nominal variable is a rule for identifying classes of individuals that are equivalent in some aspect or another. For example, the variable "religion" groups individuals through their connection with a system of religious belief, and the variable "nationality" through their connection with a nation. Typically, the individuals in one class of a nominal variable – called *category* – share at least one attribute, making them similar in a specific sense, but there is no preferred way of comparing individuals from different categories. Sometimes there might be an obvious suggestion to order or even to scale the classes, yet the analysis is supposed to ignore this information – an example being

the nominal treatment of age when nonlinear relationships between age groups and other variables are anticipated. Since so few constraints are taken into consideration, nominal treatment of a variable is the most general of a range of possibilities.

The values of a nominal variable are just labels to identify the categories, and therefore they can be replaced by any other set of values. Let the number of categories of variable j be denoted by k_j. Thus there are k_j different values in $\mathbf{h}_j$, and we are looking for k_j new values, called *category quantifications*, that minimize the sum of squared deviations in (17). This sum of squares can be decomposed into k_j separate parts, corresponding to the given grouping – by equal h_{ij}-value – of individuals into the k_j classes of variable j, and the category quantifications must be the mean $\mathbb{x}$-value (object score) of the individuals in each group. Together, the category quantifications form the *intra-class regression* of $\mathbb{x}$ onto $\mathbf{h}_j$, and in as much as they are different, they measure the extent to which the subgroups defined by variable j are heterogeneous in terms of the estimated object scores. The mean of the object scores can be freely chosen to be zero, and the sum of squares of the category quantifications, weighted by the proportion of individuals in each group, then becomes a between-group variance. When the k_j classes are well-separated in terms of the object scores, variable j is a good discriminator, and therefore the between-group variance is also called *discrimination measure* (Gifi, 1990).

As an historical remark, it was in this form – all variables nominal – that Guttman (1941) initiated the principal components analysis of categorical data, and he further demonstrated the enormous flexibility of the approach in Guttman (1946), where he introduced PCA of variables defined on *pairs of objects*, instead of the objects themselves as units, including various forms of restrictions on $\mathbf{x}$.

Ordinal variables: isotonic regression

An ordinal variable is a rule for identifying ordered classes of individuals or other objects of study. Although further subdivisions are possible, we shall take 'ordered' only to mean that all individuals within a class are equivalent (Kruskal's (1964) secondary approach to ties), and that the classes themselves are *strictly* ordered (i.e., any class is either above or below one particular other class). Because an ordinal variable carries more constraints than a nominal one, there is no

objection against having a relatively large number of classes, and even $k_j = N$ (one individual per class) is a practical possibility, especially if m is large relative to N. In the case of a Likert item, it is the semantic ordering from 'strongly disagree' to 'strongly agree' that induces a classification of individuals into ordered classes.

Intra-class regression under the additional requirement that the classes should be ordered is called *isotonic regression* (Barlow *et al.*, 1972). It minimizes the sum of squares in (17) with the specification that ϕ_j should keep the order in $\mathbf{h}_j$ unchanged (their elements are pulled in the same direction; hence the term isotonic). There are several ingenious algorithms for calculating the isotonic regression $\mathbb{q}_j$ of $\mathbb{x}$, which all lead to the same solution. What is shared with all regressions is the variance reducing property $s^2(\mathbb{q}_j) \leq s^2(\mathbb{x})$, but what is specific for regression of the isotonic kind is the fact that for two consecutive individuals i and k one always has: if $h_{ij} > h_{kj}$ *and* $x_i < x_k$, *then* $q_{ij} = q_{kj}$. If individual i has an observed value higher than k, but the ith object score is lower than the kth object score, then the best isotonic transformation will give them equal values (create a *tie*). Here it was assumed for simplicity that each individual forms one category, but with a limited number of categories the principle remains the same, except that the tie is created in the mean object scores of all individuals in the two different categories.

Multiplicity of solutions

When we are not satisfied with one single center for the distribution of variables, because we anticipate different areas of concentration, or when we are interested in a kind of residual analysis of the major deviations from the first component, the question of multiplicity of solutions arises – unlike the relatively simple situation in linear PCA. Multiplicity means that there are several ways to proceed, which can only be briefly sketched here. For a more extensive discussion of some advanced proposals in this area, the reader is referred to Bekker and De Leeuw (1988).

Multiple quantification and HOMALS

For the case of simple rescaling, it was shown in (14) how one may determine multiple principal components in the present framework. A straightforward generalization is to replace the scalar

quantities a_{js} by a multidimensional mapping $\phi_{js}(\mathbf{h}_j)$, resulting in the *multiple loss of homogeneity* function

$$\sigma^2(\Phi_1 \ldots \Phi_m, \mathbf{X}) = (Nmp)^{-1} \sum_j \sum_s \| \phi_{js}(\mathbf{h}_j) - \mathbf{x}_s \|^2 . \qquad (18)$$

By far the best known special case of (18) occurs when, for all variables, $\phi_{js}(\mathbf{h}_j)$ is specified using *multiple nominal* quantification, where repeated intra-class regressions of the different components $\mathbf{x}_s$ are called for. Multiple nominal quantification was the approach taken in Guttman's seminal 1941 paper, and the ALS method for calculating the minimal multiple loss of homogeneity forms the core technique of the Gifi system (Gifi, 1990), called HOMALS.

A convenient display of the results of a HOMALS analysis is to make a joint scatter plot (or *biplot*) of the object scores and the category quantifications for pairs of components $(\mathbf{x}_s, \mathbf{x}_t)$. Thus the joint plot contains two sets of points, one for the categories and one for the individuals. The optimal category points will be in the center of gravity of the object points that share the same category. For each variable, the k_j categories partition the total configuration of points into subconfigurations, and when the k_j centers of gravity of these subconfigurations are far apart, the variable discriminates well. Good discrimination is again expressed in a discrimination measure, which now equals the weighted mean squared distance of the category points towards the origin. The discrimination measures can be interpreted as the relative contribution of each quantified variable to the total variation of the individuals as expressed in the components $\mathbf{x}_s$. In view of (15), they are equal, dimensionwise, to squared correlations of the object scores with the quantified variables.

It can be shown (Heiser, 1981) that minimization of (18) with multiple nominal quantifications amounts to *multiple correspondence analysis*, i.e. correspondence analysis on a special type of binary table, called *indicator matrix*, which is partitioned by variable, and which codes for each of the m variables to what category each individual belongs. The results will be exactly equal if in both approaches the solution is normalized so that $s^2(\mathbf{x}_s) = 1$ and $c(\mathbf{x}_s, \mathbf{x}_t) = 0$, the standard normalization in HOMALS. A similar relationship exists with *dual scaling* of contingency tables, the Anglo-Saxon precursor of correspondence analysis (Nishisato, 1980). There is also an inverse

connection between correspondence analysis and HOMALS, in which the former is regarded as a special case of the latter, since the solution of a simple correspondence analysis can be reconstructed by appropriate renormalizations of a HOMALS with two variables (cf. De Leeuw, 1973; for still other connections, see Israëls, 1987). Unfortunately, these connections are often erroneously interpreted as equivalences, whereas – as recently emphasized by Greenacre (1991) – the methodological rationale of correspondence analysis is quite different from the rationale of homogeneity analysis as discussed here.

The indicator matrix has been the inspiration for proposals to use a so-called *fuzzy coding* of nominal variables (Van Rijckevorsel, 1987), which incorporates the estimated probability of an object being connected with any one of the categories into the analysis. For variables that carry more information than a partitioning into classes, Gifi (1990, p.169) – relying on an argument by Guttman (1959) – has explained that multiple quantification has to have a special form, called single transformation, to which we turn next.

Single transformation, but multiple weighting: PRINCALS

Suppose that we split off – as before, in the discussion of principal components – a scaling factor from the quantifications, writing $\phi_{js}(\mathbf{h}_j) = a_{js}\phi_j(\mathbf{h}_j)$, with $\|\mathbf{a}_j\|^2 = p$ for all $j = 1,...,m$ serving as identification constraints, and with $s^2(\mathbf{x}_s) = 1$ for all $s = 1,...p$. So each variable is transformed once, by $\phi_j(\,.\,)$ (hence the term single), but contributes by multiple amounts a_{js} to the components. Substituting this restriction into (18) for all variables, multiple loss of homogenity reduces to

$$\sigma^2(\phi_1...\phi_m,\mathbf{A},\mathbf{X}) = (Nmp)^{-1} \sum_j\sum_s \| a_{js}\phi_j(\mathbf{h}_j) - \mathbf{x}_s \|^2 . \qquad (19)$$

Single transformation with multiple weighting cleverly combines the idea of optimally re-expressing the variable as $\phi_j(\mathbf{h}_j)$, in the process of comparing it with the other variables, with the aim of identifying multiple areas of concentration in a multidimensional, non-uniform distribution – an aim that was presented in the present framework as the prime justification for PCA.

Computationally, all the ALS steps for the minimization of the ordinary PCA function (14) remain valid for the minimization of (19), as long as they manipulate some temporarily fixed, feasible transformation ϕ_j in the place of $\mathbf{h}_j$. Then, finding better transformations given the current

best guesses $\mathbb{A}$ and $\mathbb{X}$ of the weights and the object scores, respectively, can be based upon the decomposition

$$\sigma^2(\phi_1 \ldots \phi_m, \mathbb{A}, \mathbb{X}) = (Nmp)^{-1} \sum_j \sum_s \| a_{js} x_j - x_s \|^2 + (Nm)^{-1} \sum_j \| x_j - \phi_j(\mathbf{h}_j) \|^2 . \qquad (20)$$

Here the choice of the identification constraints $\|\mathbf{a}_j\|^2 = p$ gives the weights the geometrical interpretation of being equal, up to a factor p, to the *direction cosines* that serve to indicate, in the space of the principal components, the direction of the variable x_j defined as $x_j = p^{-1} \sum_s a_{js} x_s$, a mixture of the components x_s. The decomposition in (20) is a special application of *Huygens' Theorem*, which in general asserts that the weighted mean squared Euclidean distance of an arbitrary multidimensional point towards a number of given points equals the sum of the weighted mean squared Euclidean distance between the given points and their weighted*center of gravity* and the squared Euclidean distance between the weighted center of gravity and the arbitrary point in consideration, multiplied by the sum of the weights. To obtain (20), Huygens' Theorem has to be applied m times, for each variable separately; the given points are x_s / a_{js}, the weights are a_{js}^2 (so the sum of the weights is p), the weighted center of gravity is x_j, and the arbitrary point is $\phi_j(\mathbf{h}_j)$.

Huygens' Theorem gives the decomposition of a weighted sum of squared distances into two useful terms. The second term on the right-hand side of (20) shows that, for variable j, the component mixture x_j is regressed on the space of transformations, instead of the single component x in (17). The first term on the right-hand side of (20) shows that x_s is one of the p descriptors of the distribution of the x_j's. Another interpretation of the same two terms will be given in the next section.

PRINCALS (De Leeuw and Van Rijckevorsel, 1980; Gifi, 1985; SPSS, 1990) is a program that calculates all the quantities mentioned above, according to user-specified classes of transformations, which may be multiple nominal for some variables and single nominal, ordinal, or numerical for others. For the case that all variables are single, it can be shown that PRINCALS searches for transformations that yield a correlation matrix with maximal *sum of the first p* eigenvalues. Of course, the second and further eigenvalues (up to p) may become relatively larger than when we would have maximized only the first one.

Single transformation with analysis of residuals: generalized PRIMALS

It is often a good idea to perform a multidimensional analysis even when the optimal transformations are defined as those that maximize only the largest eigenvalue of the correlation matrix. Suppose that the largest eigenvalue is maximized by a one-dimensional HOMALS, by a one-dimensional PRINCALS, or by using similar options in a similar program. Then it is possible to proceed with an *analysis of residuals* to study the actual distribution of the transformed variables. For there are many interesting ways in which the transformed variables still may deviate from their first principal component.

More formally, assume that, starting with optimizing (20), the best $\phi_j(\mathbf{h}_j)$ are found to be $\mathcal{q}_j$. The optimal value of (20) can be written just as (8) with (9) inserted, but now with $\mathcal{q}_j$ and $\mathcal{q}_l$ in the position of $\mathbf{h}_j$ and $\mathbf{h}_l$:

$$\sigma^2(*,*) = 1 - m^{-1}\,\lambda^2 = 1 - m^{-2}\,\Sigma_j\Sigma_l\,\hat{a}_j\hat{a}_l r(\mathcal{q}_j,\mathcal{q}_l)\,. \tag{21}$$

The largest eigenvalue λ^2 of the correlation matrix with elements $r(\mathcal{q}_j,\mathcal{q}_l)$ is determined by the optimal weights $\hat{a}_j$ and $\hat{a}_l$, and the residual analysis simply consists of determining further weights and further components of the same collection of variables $\{\mathcal{q}_1,\ldots,\mathcal{q}_l\}$. The structure of the correlations *after* transformation $\{r(\mathcal{q}_j,\mathcal{q}_l)\}$ may then also be compared with the structure of the correlations *before* transformation $\{r(\mathbf{h}_j,\mathbf{h}_l)\}$. When all transformations are nominal, this approach has been called PRIMALS (because it focuses on the PRIMary eigenvalue, computed by ALS; cf. Van de Geer and Meulman, 1985). The only thing that matters, however, to support the rationale of this method as an analysis of residuals, is that the class of transformations must be general enough to be written as $a_j\phi_j(\mathbf{h}_j)$ with $s^2(\phi_j(\mathbf{h}_j)) = 1$. Otherwise, the nature of $\phi_j(\mathbf{h}_j)$ does not matter; hence it is suggested here to use the term *generalized PRIMALS* whenever the family of transformations is more general than one-to-one mappings that preserve class membership.

Determining q principal components of the quantified variables $\{\mathcal{q}_1,\ldots,\mathcal{q}_m\}$ implies optimizing the PRINCALS loss function (19) with fixed numerical variables $\mathcal{q}_j$. To distinguish this case from PRINCALS, the PRIMALS components will be called $\mathbf{z}_s$, and the PRIMALS weights will be

denoted by b_{js}. The mixture variable $\mathbf{x}_j$ becomes $\mathbf{z}_j = p^{-1} \Sigma_s b_{js}\mathbf{z}_s$, and analogously to (20), the use of Huygens' Theorem now yields the decomposition

$$\sigma^2(\mathbf{B},\mathbf{Z}) = (Nmp)^{-1} \Sigma_j\Sigma_s \| b_{js}\mathbf{z}_j - \mathbf{z}_s \|^2 + (Nm)^{-1} \Sigma_j \| \mathbb{q}_j - p^{-1} \Sigma_s b_{js}\mathbf{z}_s \|^2 . \qquad (22)$$

Here $\mathbf{z}_j$ is the linear combination of the components that is closest to $\mathbb{q}_j$. The first term on the right-hand side of this equation shows how much the unit length of $\mathbf{z}_j$ must be adjusted to get it closest to each component $\mathbf{z}_s$ separately, and this term never vanishes (in fact, it can be shown to be constant if the components are uncorrelated). The second term on the right-hand side of (22) shows the way in which $\mathbb{q}_j$ is predicted from the q components, and can be made exactly equal to zero by choosing q large enough. Because the first principal component that optimizes $\sigma^2(\mathbf{B},\mathbf{Z})$ and the eigenvalue in (21) relate to the same variables $\{\mathbb{q}_1,...,\mathbb{q}_m\}$, it must be true that the best choices are $b_{j1} = \hat{a}_j$, $b_{l1} = \hat{a}_l$, and $\mathbf{z}_1 = \mathbb{x}$, so that a vanishing second term in (22) implies, for all j,

$$(\mathbb{q}_j - p^{-1} \hat{a}_j\mathbb{x}) = p^{-1} \Sigma_{s\neq 1} b_{js}\mathbf{z}_s , \qquad (23)$$

and this shows which residuals are accounted for in generalized PRIMALS. If the components $\{\mathbf{z}_2,...,\mathbf{z}_q\}$ exhibit systematic trends, which can be studied by plotting the coefficients $\{b_{js}\}$ – either with, or without the first series $\{b_{j1}\}$ –, then there is *a posteriori* evidence of heterogeneity, even though the quantifications were found by initially assuming homogeneity.

Application of homogeneity analysis to controversial issue variables

The concepts that have been discussed in this chapter will be illustrated using an example from a survey study on opinions towards a number of controversial issues. When real observations enter the stage, statistical questions of stability and generalizability often arise. Generally speaking, the statistical evaluation of the results of a homogeneity analysis is not straightforward, because the number of parameters and the different combinations of options is usually large, and there are two types of inference to be considered: from a sample of subjects to a population of individuals, and from a sample of variables to a domain of items or tasks. Asymptotic results based on simple

multinomial sampling can be obtained (Lebart, 1976; De Leeuw, 1984), but not much is known about the quality of the approximations involved. Here, resampling techniques (Efron, 1982) and permutation tests (Edgington, 1987), which work with a minimum of assumptions, may offer a way out. We start the analysis of the example by assuming that the distribution of variables has one dominant component, and study the effect of optimizing over nonlinear transformations. Then the use of permutation tests will be demonstrated, particularly for deciding on the question of the right dimensionality.

Data description and initial analysis

The data were collected in 1974, in a Dutch survey among 575 subjects, on opinions towards a number of controversial issues (Veenhoven and Hentenaar, 1975). In Gifi (1990, chapter 13), several subsets of the survey questions have been analyzed; here we only used a subset of Likert items. Table 1 gives the description of the 11 questions that were used; the first six are statements

---

Insert Table 1 about here

---

about the *abortion* issue, and the next five are statements about the issue of *sexual freedom*. Table 1 also gives the marginal frequencies of the response categories. (The original data contain a few missing entries; because the standard option for missing data in HOMALS cannot be used easily for a study of the correlation matrix, the missing values have been replaced in the present analyses by the median response obtained from the remaining subjects, after recoding variables AB-1 and AB-2.)

There is a peculiarity in the response categories of variables AB-1 and AB-2, i.e., the first five categories indicate an increasing tolerance with respect to abortion, but category 6 denotes rejection of abortion under all circumstances. Also, categories 1 up to 5 of AB-3 and SF-1 indicate increasing intolerance, while for all other variables the order of the response categories implies decreasing intolerance. To compute a correlation matrix before transformation and its mean correlation (see Table 2), recoding of these variables is necessary. In a HOMALS analysis, the optimal quantification is supposed to take care of these anomalies automatically. From the transfor-

Insert Table 2 and Figure 2 about here

mation plots in Figure 2 it can be seen that proper recoding is achieved, apparently due to the internal consistency with the other variables. The optimal standardized quantifications (vertical axes) are plotted against the original response categories (horizontal axes); AB-4, AB-5, AB-6, SF-2, and SF-3 show a decreasing function, so that a high value on the transformed variable indicates intolerance. AB-3 and SF-1 obtain increasing functions (high values remain high, indicating intolerance), AB-1 and AB-2 obtain decreasing functions, except for the sixth category, which obtains the highest value, indicating absolute intolerance. The transformations for SF-4 and SF-5 are basically decreasing, with a small increase for the third categories.

The correlation matrix after transformation, using optimal quantification in one dimension, is given in Table 3. The largest eigenvalue of this correlation matrix equals 5.00, which is m times the eigenvalue as given by the output of programs like HOMALS and PRINCALS. The latter

Insert Table 3 about here

quantities are indicated with $m^{-1}\,\lambda^2$ in this paper, and are displayed in Table 4. Thus $5.00 = 11 \times 0.455$; this eigenvalue is equivalent to $\alpha = 0.88$. Apart from the largest eigenvalue, the subsequent eigenvalues were also computed; these are important for the study of the residuals. If the largest eigenvalue is maximized to see whether a one-dimensional structure fits the data, the components associated with the second and following eigenvalues should *not* display a structural pattern. It is possible to perform a permutation test to see whether or not this is the case for our example.

A permutation test for choosing the number of eigenvalues to be maximized

Under the null hypothesis that all variables are independent, the distribution of the various statistics that we compute can be approximated by generating a permutation distribution, which mimics independent sampling of subjects. The reader is referred to De Leeuw and Van der Burg (1986) and Ter Braak (1992) for an extensive discussion of the rationale of permutation methods.

A new feature of the present application is that we do not focus on the significance of the HOMALS eigenvalues that follow the first one, but on the residual eigenvalues of the quantification

matrix (cf. (23)). The question is, what dimensionality (p) to choose for determining the optimal transformations, and this question is answered by estimating the probability to obtain eigenvalues for the correlation matrix of the transformed variables **Q** that are as large or larger than the observed values, under random assignment of the category responses. Thus the permutation distributions of the q eigenvalues of generalized PRIMALS are employed.

The category responses of the 11 variables were independently redistributed among the subjects, with the number of permutations set equal to 1000; for each of the permuted data sets, one-dimensional homogeneity analyses were performed, with nominal treatment of all variables, and the optimal quantifications were used to obtain the permutation distributions of three eigenvalues $\lambda^2_{R(1,2,3)}$. The permutation distribution of the three eigenvalues is depicted in the form of a histogram in Figure 3. For the first eigenvalue we have $p(\lambda^2_{R(1)} \geq 4.9999) = 0.000$,

---
Insert Figure 3 about here
---

for the second eigenvalue $p(\lambda^2_{R(2)} \geq 1.4774) = 0.000$, and for the third $p(\lambda^2_{R(3)} \geq 0.8615) = 1.000$, so there is a clear structure in the first two PRIMALS components and a random residual pattern in the third. These results suggest that the right dimensionality (i.e., the number of eigenvalues to be maximized) is 2. To obtain a maximized sum of the first *two* eigenvalues, a PRINCALS analysis was done, with a single nominal transformation. It can be seen from Table 4, that the second PRINCALS eigenvalue ($11 \times 0.146 = 1.608$) is considerably larger than the mean of the

---
Insert Table 4 about here
---

permutation distribution of the second PRIMALS eigenvalue ($11 \times 0.104 = 1.146$), while the third eigenvalue ($11 \times 0.076 = 0.840$) is considerably smaller than the mean of the distribution of the third PRIMALS eigenvalue ($11 \times 0.099 = 1.085$). This fact confirms our conclusion that for the controversial issue data the maximization should be carried out over two dimensions.

Insert Figure 4 about here

Two-dimensional component loadings for the variables are displayed as vectors in Figure 4, which contains from top to bottom: PCA of the original (recoded) variables, PCA with optimization over one dimension (PRIMALS), and PCA with optimization over two dimensions (PRINCALS). As could be expected from the study of the eigenvalues, there are no major differences, only subtle ones. Compared to the PCA at the top, PRIMALS minimizes the mean squared distance of the endpoints of the vectors towards the horizontal axis. In PRINCALS, on the other hand, the endpoints have a smaller mean squared distance towards the second dimension. The PRINCALS solution shows that the SF-1 variable really belongs to the SF-subset.

Having established the bimodal character of the distribution of the variables, the abortion and sexual freedom variables were also analyzed separately. The results are given in Table 4. First of all, we note that the two mean correlations are higher (0.607 and 0.486, respectively) than the grand mean 0.414, and this remarkable result remains true when we compare the dominant PRIMALS eigenvalues (0.656 and 0.512) with the first two PRINCALS eigenvalues (0.448 and 0.146). Next, differential weighting has a slightly larger effect for the sexual freedom (0.492) than for the abortion variables (0.609). Third, the permutation results indicate that in *both* subsets of variables, the second eigenvalue is *not* larger than the one obtained by random assignment to the response categories (both p-values are 0.000). Finally, when the sum of the first two eigenvalues is maximized by PRINCALS, the second PRINCALS eigenvalue ($5 \times 0.201 = 1.004$) is not larger than the mean of the permutation distribution of the second eigenvalue of the sexual freedom variables ($5 \times 0.203 = 1.016$), and considerably smaller ($6 \times 0.126 = 0.755$) than the mean ($6 \times 0.173 = 1.039$) for the abortion variables, so it is concluded that maximization over only one dimension is justified for the two subsets separately.

Substantive interpretation

What does this analysis tell us about the controversial issues themselves? First, leniency versus strictness in sexual freedom apparently involves different beliefs and values than the abortion controversy, since they constitute two separate components that are only weakly correlated. The implication is that persons in the sample who take a tolerant stand in the abortion issue may still be either strict or lenient on sexual freedom, while persons in the sample who think abortion is never

justifiable also may be either strict or lenient on sexual freedom. Second, the statements that build up the abortion component are all about equally important, and different selections of them could be chosen as the same indicator of someone's position on this ethical dilemma. For the sexual freedom component, statement SF-1 is a bit odd, perhaps because it requires a double negation to express strictness here.

The analysis suggests using two new variables in further studies of a similar nature. To actually obtain these indicators of leniency toward sexual freedom and tolerance towards abortion, one would code the responses with the category quantifications (as given here on the vertical axis of Figure 2), and add the resulting coded variables within the two identified groups. Such indicators (equal to the object scores) will be more reliable and enable finer discrimination than each of the items alone. As an example, the object scores of the abortion component are plotted in Figure 5, in

---

Insert Figure 5 about here

---

such a way that the five subsamples generated by the five categories of statement AB-5 are clearly visible as partially overlapping distributions. Note that the regression is linear after transformation (Figure 5 is a plot of $\mathbf{x}$ versus $\mathbf{q}_5$, with the category centroids specially marked; the discrimination measure is equal to the squared correlation in this scatter plot). So the group of respondents on the right, labeled with 1, have all stated that they are fully convinced that "people who agree with abortion have little respect for life", whereas the group of respondents on the left, labeled with 5, have all declared that they completely disagree with this statement. Although the partitioning of the object scores displays a clear shift of the means (heterogeneity), it is also true that there is considerable overlap (absence of perfect discrimination). Apparently, people who are extreme in terms of "respect for life" (AB-5) need not be similarly extreme in terms of "woman's right" (AB-3), for example, and vice versa. As long as we are looking at the level of an individual variable, we have to be careful for random disturbances (be it measurement error, response bias, temporary switch of opinion, or the like). Nevertheless, on average, the abortion statements discriminate the respondents consistently from *tolerant* to *adverse*. The sexual freedom statements discriminate them

differently, on average, yielding an order from *liberal* to *authoritarian*. It could be suggested that the former is a personal and moral distinction, while the latter is a social and normative one.

Discussion

Homogeneity analysis was presented as a way to explore a distribution of variables and their nonlinear relationships. A uniform distribution occurs when the variables are equicorrelated, with uniformly zero correlation as a special case. More often, the variables have one or more areas of concentration, just as empirical univariate distributions often have one or more peaks. The centers of these areas may be determined by generalized principal components analysis, and are candidates for further practical use, since their reliability will be higher than any single variable, due to the effect of averaging out the measurement error. In fact, their reliability as measured by Cronbach's α will be maximal for new samples (of subjects *and* variables) that are similar to the one used to establish homogeneity.

What are the limits in the applicability of homogeneity analysis? Two kinds of limits should be mentioned. In the first place, certain types of nonuniform distribution are hard to describe by this form of nonlinear PCA (Heiser, 1986), for instance when the data are translation families of functions with at least one inflection point. Such cases, which require a method with more complicated basis functions than the central vectors $\mathbf{x}$ used here, are not easy to detect, except by careful examination of bivariate scatter plots and some prior knowledge of the processes involved. In the second place, homogeneity analysis – as standardly defined – focuses on the univariate and bivariate marginals, or the means, variances and covariances. If one expects the presence of higher order interactions (in the categorical case) or non-zero higher order moments (in the numerical case), special action is required to take them into account. In the categorical case, it is just a matter of redefining the variables involved into an *interaction variable* (Gifi, 1990, p. 376; Van der Lans, 1992, ch. 4), which is a variable whose categories correspond to the cells of the two- or higher-dimensional contingency table of the original variables. Thus this limit can be transgressed within the method itself.

It is also possible to present homogeneity analysis in an entirely different way, which starts from an identification of the individuals as points in m different spaces, of dimensionality $k_j - 1$, and which proceeds by finding some kind of consensus configuration in low-dimensional space that optimally represents the distances between individuals. Because this approach is extensively documented elsewhere, the present chapter has emphasized a different line of development. The reader is referred to Meulman (1986, 1992) for more information on the so-called *distance approach*. When an *a priori* partitioning of variables into K sets is available, we can profit from the nesting to study the homogeneity of the so-called *canonical variables*, which are themselves linear combinations of variables within sets. Then all possibilities of variable weighting and nonlinear transformation are open to us (Gifi, 1990, ch. 5), but an extensive discussion of these would lead too far. The reader is referred to Van der Burg *et al.* (1988) for a discussion of this situation, and the related technique called OVERALS.

Resampling techniques such as the bootstrap have been applied before in homogeneity analysis (Meulman, 1982; Greenacre, 1984; Saporta and Hatabian, 1986; Van der Burg & De Leeuw, 1988; Markus and Visser, 1992), but applications of permutation tests await further development. In our example the results were encouraging: it could definitely be decided that two homogeneous clusters of variables are to be distinguished, with slightly correlated central components, and with a joint Cronbach's α considerably higher (0.895 and 0.762, corresponding to the mean squared correlations of 0.656 and 0.512 in Table 4) than the standard PRINCALS two-dimensional homogeneity analysis that is based on uncorrelated principal components (0.877 and 0.415, corresponding to the mean squared correlations of 0.448 and 0.146 in Table 4).

Note

The programs HOMALS, PRINCALS and OVERALS, mentioned in the text and/or used in the applications, were originally developed in the Department of Data Theory of the University of Leiden, and are available in the SPSS package Categories, documented in SPSS (1990).

Acknowledgement

The authors are indebted to the editors, Shizuhiko Nishisato, Norman Verhelst, Niels Veldhuijzen, Ivo van der Lans, and Jacques Commandeur for many helpful comments on an earlier draft of this paper.

References

Barlow, R.E., Bartholomew, D.J., Bremner, J.M., and Brunk, H.D. (1972). *Statistical Inference under Order Restrictions.* New York: Wiley.

Bekker, P. and De Leeuw, J. (1988). Relations between variants of nonlinear principal component analysis. In J.L.A. Van Rijckevorsel & J. De Leeuw (Eds.), *Component and Correspondence Analysis.* New York: Wiley.

Cronbach, L.J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika,* 16, 297-334.

De Leeuw, J. (1973). *Canonical Analysis of Categorical Data.* Doctoral dissertation, University of Leiden, The Netherlands.

De Leeuw, J. (1984). *Statistical Properties of Multiple Correspondence Analysis.* Internal Report RR-84-06, Department of Data Theory, University of Leiden, The Netherlands.

De Leeuw, J. and Van der Burg, E. (1986). The permutational limit distribution of generalized canonical correlations. In E. Diday *et al.* (Eds.), *Data Analysis and Informatics, Vol.IV,* pp. 509-521. Amsterdam: North-Holland.

De Leeuw, J. and Van Rijckevorsel, J.L.A. (1980). HOMALS and PRINCALS: some generalizations of principal components analysis. In E. Diday *et al.* (Eds.), *Data Analysis and Informatics.* Amsterdam: North-Holland.

Edgington, E.S. (1987). *Randomization Tests, 2nd Ed.* New York: Dekker.

Efron, B. (1982). *The Jackknife, the Bootstrap and other Resampling Plans.* Philadelphia: SIAM.

Gifi, A. (1985). *PRINCALS.* Research Report UG-85-03. Leiden: Department of Data Theory.

Gifi, A. (1990). *Nonlinear Multivariate Analysis.* New York: Wiley.

Gower, J.C. (1975). Generalized Procrustes analysis. *Psychometrika,* 40, 33-51.

Greenacre, M.J. (1984). *Theory and Applications of Correspondence Analysis.* London: Academic Press.

Greenacre, M.J. (1991). Interpreting multiple correspondence analysis. *Applied Stochastic Models and Data Analysis,* 7, 195-210.

Guttman, L. (1941). The quantification of a class of attributes: a theory and a method of scale construction. In P. Horst (Ed.), *The prediction of Personal Adjustment.* New York: SSRC.

Guttman, L. (1946). An approach for quantifying paired comparisons and rank order. *Annals of mathematical statistics,* 17, 144-163.

Guttman, L. (1959). Metricizing rank-ordered or unordered data for a linear factor analysis. *Sankhya, A*, 21, 257-268.

Heiser, W.J. (1981). *Unfolding Analysis of Proximity Data.* Doctoral dissertation, University of Leiden, The Netherlands.

Heiser, W.J. (1986). Undesired nonlinearities in nonlinear multivariate analysis. In E. Diday *et al.* (Eds.), *Data Analysis and Informatics*, Vol IV, pp. 455-469. Amsterdam: North-Holland.

Israëls, A.Z. (1987). *Eigenvalue Techniques for Qualitative Data.* Doctoral dissertation. Leiden: DSWO Press.

Kruskal, J.B. (1964). Nonmetric multidimensional scaling: a numerical method. *Psychometrika*, 29, 115-129.

Lebart, L. (1976). The significance of eigenvalues issued from correspondence analysis of contingency tables. In J. Gordesch and P. Naeve (Eds.), *Proceedings COMPSTAT 1976.* Vienna: Physika Verlag.

Lord, F.M. (1958). Some relations between Guttman's principal components of scale analysis and other psychometric theory. *Psychometrika*, 23, 291-296.

Markus, M. Th. and Visser, R.A. (1992). Applying the bootstrap to generate confidence regions in multiple correspondence analysis; a Monte Carlo Study. In K.-H. Jöckel *et al.* (Eds.), *Bootstrapping and related techniques.* Berlin: Springer Verlag.

Meulman, J.J. (1982). *Homogeneity Analysis of Incomplete Data.* Leiden: DSWO Press.

Meulman, J.J. (1986). *A Distance Approach to Nonlinear Multivariate Analysis.* Doctoral dissertation. Leiden: DSWO Press.

Meulman, J.J. (1992). The integration of multidimensional scaling and multivariate analysis with optimal transformations. *Psychometrika*, 57, 539-565.

Nishisato, S. (1980). *Analysis of Categorical Data: Dual Scaling and its Applications.* Toronto, Canada: University of Toronto Press.

Saporta, G. and Hatabian, G. (1986). Régions de confiance en analyse factorielle. In E. Diday *et al.* (Eds.), *Data Analysis and Informatics, Vol.IV*, pp. 499-508. Amsterdam: North-Holland.

Spearman, C. (1904). "General intelligence" objectively determined and measured. *American Journal of Psychology*, 15, 201-293.

Spearman, C. (1910). Correlation calculated from faulty data. *British Journal of Psychology*, 3, 271-295.

SPSS (1990). *Categories.* Chicago: SPSS Inc.

Ter Braak, C.J.F. (1992). Permutation versus bootstrap significance tests in multiple regression and ANOVA. In K.-H. Jöckel *et al.* (Eds.), *Bootstrapping and Related Techniques*, pp. 79-86. Berlin: Springer Verlag.

Van de Geer, J.P. and Meulman, J.J. (1985). *PRIMALS.* Research Report UG-85-02. Leiden: Department of Data Theory.

Van der Burg, E. and De Leeuw, J. (1988). Use of the multinomial jackknife and bootstrap in generalized nonlinear canonical correlation analysis. *Applied Stochastic Models and Data Analysis*, 4, 159-172.

Van der Burg, E. and De Leeuw, J. (1988). Homogeneity analysis with k sets of variables: an alternating least squares method with optimal scaling features. *Psychometrika*, 53, 177-197.

Van der Lans, I.A. (1992). *Nonlinear Multivariate Analysis for Multiattribute Preference Data.* Doctoral dissertation. Leiden: DSWO Press.

Van Rijckevorsel, J.L.A. (1987). *The Application of Fuzzy Coding and Horseshoes in Multiple Correspondence Analysis.* Doctoral dissertation. Leiden: DSWO Press.

Veenhoven, R. and Hentenaar, F. (1975). *Nederlanders over Abortus.* Den Haag: Vereniging Stimezo Nederland.

Winsberg, S. and Ramsay, J.O. (1983). Monotone spline transformations for dimension reduction. *Psychometrika*, 48, 575-595.

Table 1. 6 Statements about abortion and 5 statements about sexual freedom with response categories and marginal frequencies

AB-1	A woman, 45 years of age, when menstruation fails to come, thinks menopause has started, and does not worry. Later she appears to be pregnant. She has a family with grown-up children. Until which month of pregnancy do you feel that abortion in this special case is still justified? Or in your judgement is abortion in this case not justified?
AB-2	A girl of 15 -unmarried- suspects she is pregnant. She is scared to talk about it with the family doctor or with her parents. As a result it takes much longer for her than necessary to enlist for medical aid. Until which month of pregnancy do you feel that abortion in this special case is still justified? Or in your judgement is abortion in this case not justified?
	Responses: 1 = 'justifiable until 3 months', 2 = 'until 4 months', 3 = 'until 5 months', 4 = 'until 6 months', 5 = 'after six months', 6 = 'not justifiable'.
AB-3	It is the woman's right to have abortion when she wants it .
AB-4	Medical practitioners who perform abortion are no better than murderers.
AB-5	People who agree with abortion have little respect for life.
AB-6	Abortion is justifiable under no circumstances.
	Response from 1 = 'agree completely' to 5 = 'disagree completely'.
SF-1	I don't object to children below the age of ten walking around on the beach naked.
SF-2	If sexual intercourse was separated from procreation it would soon become pure egoism.
SF-3	Parents should forbid children to have sexual play.
SF-4	Young people who have sexual intercourse before marriage do not have respect for each other.
SF-5	Parents should impress upon their children that it is better to have control over yourself and not to indulge in masturbation.
	Response from 1 = 'agree completely', to 5 = 'disagree completely'.

	Marginal frequencies					
Label	1	2	3	4	5	6
AB-1	221	48	18	8	21	259
AB-2	213	62	31	12	33	224
AB-3	178	115	36	93	153	
AB-4	41	32	77	111	314	
AB-5	114	60	69	117	215	
AB-6	43	54	62	110	306	
SF-1	130	85	56	99	205	
SF-2	84	67	85	117	222	
SF-3	124	109	101	114	127	
SF-4	49	42	56	126	302	
SF-5	124	97	92	88	174	

Table 2. Correlation matrix original variables, with mean correlation 0.414, and eigenvalues

	AB-1	AB-2	AB-3	AB-4	AB-5	AB-6	SF-1	SF-2	SF-3	SF-4	SF-5
AB-1	1.00	0.67	0.44	0.39	0.48	0.41	0.32	0.27	0.24	0.28	0.22
AB-2	0.67	1.00	0.51	0.42	0.49	0.46	0.27	0.18	0.22	0.29	0.23
AB-3	0.44	0.51	1.00	0.50	0.57	0.49	0.25	0.14	0.18	0.26	0.21
AB-4	0.39	0.42	0.50	1.00	0.71	0.68	0.27	0.27	0.28	0.36	0.30
AB-5	0.48	0.49	0.57	0.71	1.00	0.70	0.28	0.27	0.28	0.39	0.34
AB-6	0.41	0.46	0.49	0.68	0.70	1.00	0.23	0.26	0.25	0.37	0.38
SF-1	0.32	0.27	0.25	0.27	0.28	0.23	1.00	0.18	0.31	0.25	0.30
SF-2	0.27	0.18	0.14	0.27	0.27	0.26	0.18	1.00	0.36	0.39	0.38
SF-3	0.24	0.22	0.18	0.28	0.28	0.25	0.31	0.36	1.00	0.37	0.53
SF-4	0.28	0.29	0.26	0.36	0.39	0.37	0.25	0.39	0.37	1.00	0.50
SF-5	0.22	0.23	0.21	0.30	0.34	0.38	0.30	0.38	0.53	0.50	1.00
λ^2	4.657	1.512	0.059	0.797	0.642	0.600	0.503	0.445	0.321	0.295	0.270

Table 3. Correlation matrix variables after optimal quantification in one dimension, with mean correlation 0.443, and eigenvalues

	AB-1	AB-2	AB-3	AB-4	AB-5	AB-6	SF-1	SF-2	SF-3	SF-4	SF-5
AB-1	1.00	0.64	0.52	0.49	0.57	0.51	0.32	0.31	0.28	0.35	0.29
AB-2	0.64	1.00	0.58	0.52	0.59	0.58	0.27	0.24	0.25	0.36	0.29
AB-3	0.52	0.58	1.00	0.54	0.58	0.51	0.26	0.15	0.19	0.28	0.24
AB-4	0.49	0.52	0.54	1.00	0.73	0.68	0.29	0.28	0.28	0.40	0.32
AB-5	0.57	0.59	0.58	0.73	1.00	0.71	0.30	0.29	0.30	0.41	0.36
AB-6	0.51	0.58	0.51	0.68	0.71	1.00	0.27	0.28	0.25	0.40	0.40
SF-1	0.32	0.27	0.26	0.29	0.30	0.27	1.00	0.15	0.30	0.28	0.29
SF-2	0.31	0.24	0.15	0.28	0.29	0.28	0.15	1.00	0.36	0.43	0.39
SF-3	0.28	0.25	0.19	0.28	0.30	0.25	0.30	0.36	1.00	0.38	0.54
SF-4	0.35	0.36	0.28	0.40	0.41	0.40	0.28	0.43	0.38	1.00	0.51
SF-5	0.29	0.29	0.24	0.32	0.36	0.40	0.29	0.39	0.54	0.51	1.00
λ^2	5.000	1.477	0.861	0.692	0.635	0.560	0.464	0.435	0.344	0.274	0.257

Table 4. ($m^{-1} \times$) Eigenvalues compared to mean correlations ($\underline{r}$)

$\underline{r}$	All Variables 0.414			Abortion 0.607		Sexual Freedom 0.486	
	Dim-1	Dim-2	Dim-3	Dim-1	Dim-2	Dim-1	Dim-2
1	0.423	0.137	0.087	0.609	0.150	0.492	0.169
2	0.455	0.134	0.078	0.656	0.109	0.512	0.162
3	0.448	0.146	0.076	0.644	0.126	0.481	0.201
4	0.142	0.104	0.099	0.229	0.173	0.262	0.203

1 Eigenvalues principal components analysis original variables
2 Eigenvalues PRIMALS (first dimension optimal)
3 Eigenvalues PRINCALS (first and second dimension optimal)
4 Mean permutation distribution eigenvalues PRIMALS

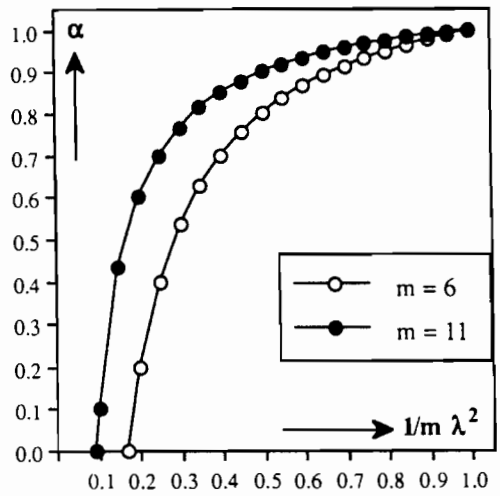

Figure 1. Relationship between Cronbach's α (vertical axis) and the eigenvalue (horizontal axis)

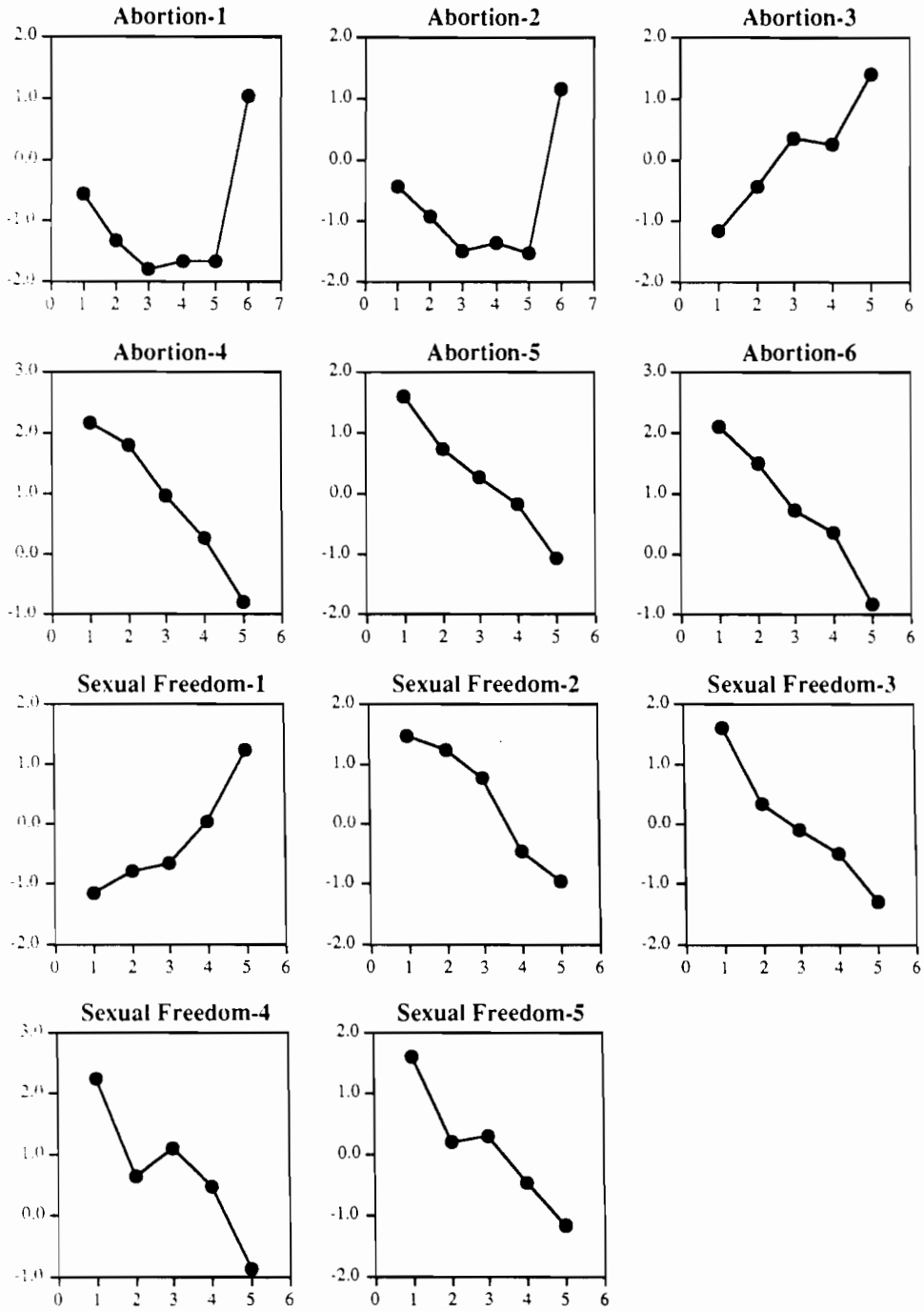

Figure 2. Transformation plots for controversial issue data. Optimal quantifications (vertical axes) versus original categories (horizontal axes)

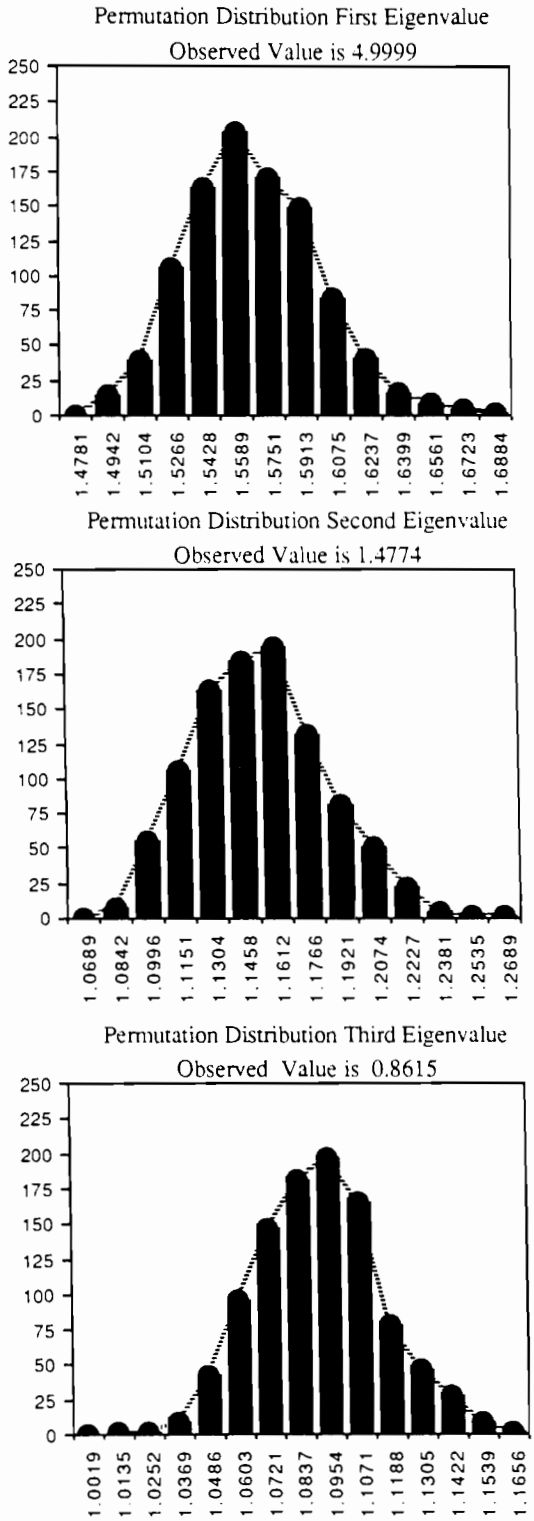

Figure 3. Permutation study of the eigenvalues of the correlation matrix after optimal quantification

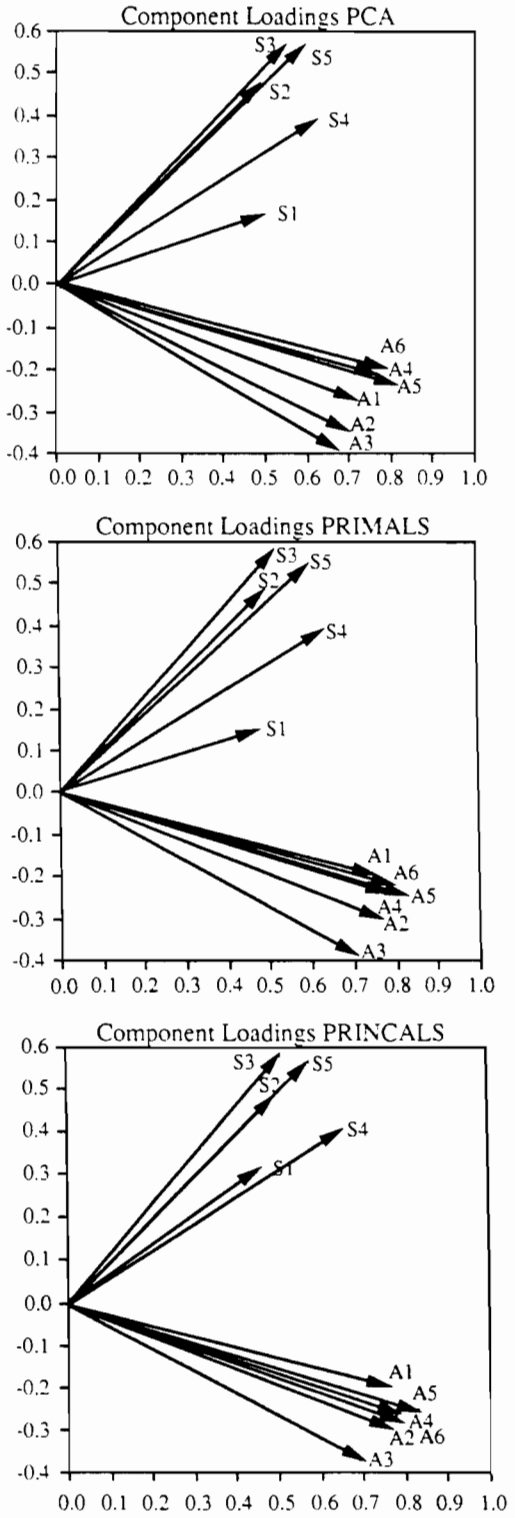

Figure 4. Three different two-dimensional representations of the variables in the analysis of the controversial issue data

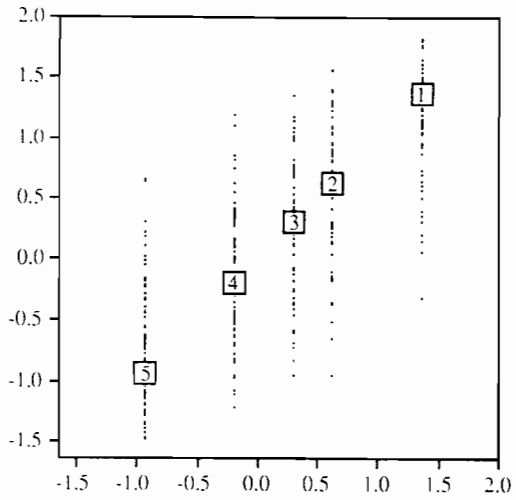

Figure 5. Object scores of the one-dimensional homogeneity analysis of the abortion statements plotted against the quantified variable AB-5