GENERALIZED PROCRUSTES ANALYSIS WITH ITERATIVE WEIGHTING TO ACHIEVE RESISTANCE

Peter Verboon en K. Ruben Gabriel

Department of Data Theory
University of Leiden

GENERALIZED PROCRUSTES ANALYSIS WITH ITERATIVE WEIGHTING TO ACHIEVE RESISTANCE

Peter Verboon & K. Ruben Gabriel

Abstract

The paper studies the generalized Procrustes problem with the extension of variable weights. A distinction is made between configurations and configuration classes, depending on the type of data used. The main interest is in configuration classes. Two different types of weighting are introduced: the Huber and Tukey weights. The objective is to use the weights to yield a resistant procedure for the generalized Procrustes problem. An algorithm is described and some attention is paid to the starting values in order to avoid non-optimal solutions. An example and simulation results are presented which show that the weighted approach outperforms least squares when there are outliers in the data.

Key words : generalized Procrustes analysis, resistance, outliers, Huber and Tukey weights , configuration classes

Introduction

For the Procrustes problem in its most basic form, the objective is to rotate a set of points towards another set. Solutions to the basic problem were already given in the fifties and sixties (Green, 1952; Cliff, 1966; Schönemann, 1966). Since then all kinds of generalizations have been studied by many people. For instance, the problem has been extended with the estimation of scaling factors and translations; generalizations to more sets with different dimensionalities have also been studied (Gower, 1975; Peay, 1988).

In this paper we study another extension. Consider data matrices $\mathbf{X}_k$ (k=1,...,p) with rows $\mathbf{x}_{ik}'$ (i=1,...,n) representing n points, or objects, and elements x_{ijk} (j=1,...,m) representing dimensions or variables, respectively. Our interest is in studying the similarity between the $\mathbf{X}_k$ by applying a generalized Procrustes analysis (GPA) in order to find an average or centroid (sometimes called *consensus*) configuration and in seeing which points do not fit such a centroid. This paper extends the generalized Procrustes problem to the use of iterative weights. These weights are not known beforehand and thus are part of the estimation procedure. Our objective is to develop a procedure which is resistant to aberrant points (outliers) in one or more sets. The weights are used to downweight the outliers in order to minimize their disturbing effects upon the overall solution. This approach is related to robust estimation procedures in regression analysis (e.g. Holland & Welsch, 1977) and is also used in the ordinary Procrustes problem (Verboon & Heiser, 1992). It is expected that such a resistant procedure will give a better understanding of the similarity between the data matrices when outliers are present than is provided by ordinary least squares.

GPA can be applied for different types of data. We distinguish between data consisting of *configurations* and of *configuration classes*. A configuration $\mathbf{X}$ is defined as a set of n points in an m-dimensional space and its *form* as the collection of the lengths of its

vectors and the angles between them (Goodall, 1991). For a configuration **X** with rows $\mathbf{x}_1,...,\mathbf{x}_n$, the form is given by the matrix $\mathbf{B} = \mathbf{XX}'$, whose elements are $(i,e = 1,...,n)$,

$$b_{ie} = \mathbf{x}_i'\mathbf{x}_e,$$

so that

$$\sqrt{b_{ii}} = \| \mathbf{x}_i \|$$

and

$$\cos^{-1}\left(\frac{b_{ie}}{\sqrt{(b_{ii}b_{ee})}}\right) = \text{angle between } \mathbf{x}_i \text{ and } \mathbf{x}_e.$$

The information in the form matrix **B** thus is about the magnitudes and positions of the objects relative to one another but not about their common *orientation* in space. For a given set of objects, there may be several differently oriented configurations with the same form. A collection of all configurations with the same form, but with different orientations, is an equivalence class. For form **B**, this configuration class is

$$\mathbf{C} = \{ \mathbf{X} \mid \mathbf{XX}' = \mathbf{B} \}.$$

This is readily seen to be closed under the set of orthogonal rotations.

Configurations are usually represented in a space of which the dimensions are clearly interpretable and meaningful. Examples are the ratings by p judges for n products on m attributes, or the readings at a number of localities of the magnetic North and the true North. In these examples both the attributes and the geographical coordinates are meaningful for the purpose of the analysis. It follows that both form and orientation are important aspects of configurations and so, when GPA is applied to configurations, the estimated rotation angles are parameters of interest. Together with the fit of the model, the angles between judges can be used as an indication of similarity (consensus) between judges.

For configuration classes the dimensions of the space in which the points are represented, have no clear meaning or are not interesting for the purposes of the analysis. Examples are the solutions of multidimensional scaling (MDS) or principal components analysis (PCA), (see Sibson, 1978; Krzanowski, 1988). In the comparison of different MDS or PCA solutions, we are only interested in the position of the points *relative to each other* and thus are concerned only with form, i.e. with a configuration class. If we want to study the similarity of different solutions, we first apply a rotation to eliminate all variation in orientation, as that is not of interest. This means that when we apply GPA to configuration classes, the obtained rotation angles (orientations) can be seen as "nuisance parameters".

The distinction between configurations and configuration classes leads to different types of outliers. For configurations an outlier is simply an element x_{ijk} that is very different from the other x_{ijg}'s (all $g \neq k$); for instance, going back to our example, one judge is not able to perceive the attribute "sweetness" of a particular product. For configuration classes, on the other hand, an outlier is defined in terms of the angles between a vector $\mathbf{x}_{ik}$ and the other vectors $\mathbf{x}_{ek}$ (all $e \neq i$), as compared to the angles between vector $\mathbf{x}_{ig}$ and the other vectors $\mathbf{x}_{eg}$ (all $e \neq i$), for all $g \neq k$.

Thus, an outlier is either a coordinate that is extreme within a set (configuration) or it is a vector that has quite a different position in a set compared to its corresponding vectors in the other sets (configuration class).

In this paper we concentrate on configuration classes. It is shown how weights can be assigned to the points in a configuration class in order to obtain resistance to outliers. A simulation study is presented, which shows that resistant weighting can be very useful when there are outliers in the data.

The unweighted generalized Procrustes problem

Consider the problem of comparing configuration classes (or configurations) $\mathbf{X}_k$ (k=1,...,p). These matrices are first orthogonally rotated towards some centroid configuration $\mathbf{Z}$ to obtain maximum agreement in the least squares sense. As Commandeur (1991) has pointed out, this is equivalent to minimizing the total sums of squares of the differences between all pairs of configurations. For simplicity the method is restricted to orthogonal rotations and reflections, so that all $\mathbf{X}_k$'s are assumed to have been centered (approximately) on the origin and scaled similarly, that is, if we would estimate central dilation (scaling) factors, these factors would approximately be equal to one. The minimization then is that of loss function

$$\sigma(\mathbf{Z};\mathbf{T}_1,...,\mathbf{T}_p) = \sum_{k=1}^{p} \sum_{i=1}^{n} \sum_{j=1}^{m} (z_{ij} - \mathbf{x}_{ik}'\mathbf{t}_{jk})^2, \tag{1}$$

where $\mathbf{x}_{ik}$ is the ith row of $\mathbf{X}_k$ and $\mathbf{t}_{jk}$ the jth column of a rotation matrix $\mathbf{T}_k$ (m x m), which has the restriction $\mathbf{T}_k'\mathbf{T}_k = \mathbf{T}_k\mathbf{T}_k' = \mathbf{I}$. In this problem the matrices $\mathbf{X}_k$ (n x m) are known and fixed and $\mathbf{Z}$ (n x m) is the centroid configuration, which is the least squares estimate

$$\mathbf{Z} = p^{-1} \sum_{k=1}^{p} \mathbf{X}_k\mathbf{T}_k. \tag{2}$$

A solution for this problem was given by Kristof and Wingersky (1971). The extension with scaling parameters and translations was solved by Gower (1975) and some computational aspects of this solution were improved by Ten Berge (1977). These solutions are iterative and may be started with some initial $\mathbf{Z}$, towards which all configurations are rotated, yielding $\mathbf{T}_k$'s, which are called Procrustes estimates. Next $\mathbf{Z}$ is updated by (2) and the process continues until convergence.

Weighted generalized Procrustes methods

From (1), we may define residual elements, which can be gathered in matrices $\mathbf{R}_k$ (n x m), defined as:

$$\mathbf{R}_k = \mathbf{Z} - \mathbf{X}_k\mathbf{T}_k, \tag{3}$$

so that

$$\sigma(\mathbf{Z},\mathbf{T}) = \sum_{k=1}^{p} \text{trace } (\mathbf{R}_k\mathbf{R}_k') \tag{4}$$

In the present discussion we concentrate on assigning weights to entire rows of the residual matrix $\mathbf{R}_k$ i.e., vectorwise weighting. This means that we assign a weight to the distance between a point (vector) in a data matrix and the corresponding point (vector) in the centroid. In scalar notation the loss function becomes

$$\sigma(\mathbf{Z},\mathbf{T},\mathbf{V}) = \sum_{k=1}^{p} \sum_{i=1}^{n} v_{ik} \sum_{j=1}^{m} (z_{ij} - \mathbf{x}_{ik}'\mathbf{t}_{jk})^2, \tag{5}$$

where $\mathbf{V} = \{v_{ik}\}$ is the n x p matrix of weights. Note that if all weights are chosen equal to one, (5) becomes equal to the least squares criterion given in (1).

The algorithm to minimize the ordinary unweighted GPA of (1) is iterative; in the weighted GPA of (5) that we propose, the weights v_{ik} depend on the residuals. As the residuals depend on the weights, the procedure to solve it needs an additional weighting step in the iteration process.

To be more specific about our algorithm with vectorwise weighting, one may distinguish three steps: the T-step, the Z-step and the V-step. Each of these steps solves for one set of parameters, i.e. $\mathbf{T}_k$ (k=1,...,p), $\mathbf{Z}$ or $\mathbf{V}$, while the others are considered as fixed. The T-step actually solves p (the number of configuration classes) independent rotation

problems. The weights do not complicate this step, since one can always write - switching to matrix notation - the loss function for this step as

$$\sigma(\mathbf{Z}^*,\mathbf{T}) = \text{tr}\ \Sigma_k(\mathbf{Z}^* - \mathbf{X}^*_k\mathbf{T}_k)'(\mathbf{Z}^* - \mathbf{X}^*_k\mathbf{T}_k), \tag{6}$$

where $\mathbf{Z}^* = \mathbf{V}_k^{1/2}\mathbf{Z}$ and $\mathbf{X}^*_k = \mathbf{V}_k^{1/2}\mathbf{X}_k$, with $\mathbf{V}_k$ being the diagonal n x n matrix of the kth column of $\mathbf{V}$, and that is the classical Procrustes problem. In the Z-step the centroid configuration is computed as the weighted least squares minimizer

$$\mathbf{Z} = (\sum_{k=1}^{p} \mathbf{V}_k)^{-1} \sum_{k=1}^{p} \mathbf{V}_k\mathbf{X}_k\mathbf{T}_k. \tag{7}$$

Using these solutions for $\mathbf{T}$ and $\mathbf{Z}$ the V-step computes the Euclidean distances ($\mathbb{d}_{ik}$) between each point in $\mathbf{X}_k$ and its corresponding point in the centroid,

$$\mathbb{d}_{ik} = (\sum_{j=1}^{m} (z_{ij} - \mathbf{x}_{ik}'\mathbf{t}_{jk})^2)^{1/2}, \tag{8}$$

and uses these distances to find new weights. The iteratively reweighted GPA procedure cycles through the three steps until convergence. It is an immediate generalization of the algorithm proposed in Verboon and Heiser (1992) which solves the ordinary Procrustes problem with only two configurations; that paper also contains a proof on the convergence of the algorithm.

Instead of the vectorwise approach, in which the rows of the residual matrix are weighted, weights w_{ijk} can also be assigned to the elements of $\mathbf{R}_k$, the matrix of residuals. The loss function written in scalar notation, becomes

$$\sigma(\mathbf{Z},\mathbf{T},\mathbf{W}) = \sum_{k=1}^{p}\sum_{i=1}^{n}\sum_{j=1}^{m} w_{ijk}(z_{ij} - \mathbf{x}_{ik}'\mathbf{t}_{jk})^2, \tag{9}$$

where the vector $\mathbf{x}_{ik}'$ is the ith row of $\mathbf{X}_k$ and $\mathbf{t}_{jk}$ the jth column of $\mathbf{T}_k$.

The two types of weighting, elementwise (9) and vectorwise (6), correspond with the distinction between configurations and configuration classes. Elementwise weighting is meaningful for configurations, but not for configuration classes, in which the dimensions are not meaningfully defined and may be rotated without loss of interpretability. Vectorwise weighting, as introduced here, is specially suited for configuration classes, since the vectors $\mathbf{x}_i'$ are considered as units of analysis, and outliers are aberrant vectors, hence, weights are assigned vectorwise. The vectorwise weighting approach may not yield satisfactory results for configurations, because either whole points may be downweighted by it, although only one coordinate is outlying, or it may fail to detect an outlier in which the outlying co-ordinate is masked by the other "good" coordinates. For instance, a judge who is not able to assess sweetness may well be able to assess other qualities of a product. The vectorwise approach would either see this judge as a bad assessor (outlier) or it would ignore his lack of sweetness assessment. In the elementwise approach, however, one may be able to discover that a particular judge rates sweetness differently from the other judges, while is ratings on other attributes is in line with the ratings of the other judges.

Weighting to achieve resistance

The objective in this study is to minimize (5), where the weights are defined so as to yield some resistant properties of the procedure. We choose the weights as some decreasing function of the vectorwise residuals or Euclidean distances between the points in the configuration classes and the corresponding point in the centroid.
One choice of weight function is Huber's (Huber, 1981),

$$v_{ik} = \begin{cases} 1 & \text{if} \quad \mathbb{d}_{ik} < c \\ \dfrac{c}{\mathbb{d}_{ik}} & \text{if} \quad \mathbb{d}_{ik} \geq c. \end{cases} \tag{10}$$

The constant c is called the *tuning constant*. The algorithm for minimizing (5) solves a weighted GPA problem in the T and Z steps and updates the weights via (10) in the V step. This choice of a function optimizes the squared Euclidean distances if a distance found in the previous step (denoted as $\mathbb{d}_{ik}$) is smaller than c. Otherwise it obtains weights as the reciprocal of the previous distances, which results in downweighting large distances.

Another type of weight function, which is known to yield better results when the contamination in the data is more severe, is based on the biweight or Tukey function (Mosteller & Tukey, 1977), computed as

$$v_{ik} = \begin{cases} (1 - (\dfrac{\mathbb{d}_{ik}}{c})^2)^2 & \text{if} \quad \mathbb{d}_{ik} < c \\ 0 & \text{if} \quad \mathbb{d}_{ik} \geq c. \end{cases} \tag{11}$$

These weight functions are well-known in the context of robust estimation procedures. Both allow some tolerance for large residuals (distances), because these large values contribute relatively little to the loss. This property will therefore diminish the influence of outliers on the solution.

In the present context we chose the tuning constant as $\frac{2}{3}\sigma$ for Huber and $1\frac{3}{4}\sigma$ for the biweight, with σ being a spread measure of the least squares residuals defined as

$$\sigma = \text{median}\,(\mathbb{D}) + 4\ \text{MAD}\,(\mathbb{D}), \tag{12}$$

where $\mathbb{D}$ is the n x p matrix with the (positive valued) least squares residuals $\mathbb{d}_{ik}$ (i=1,...,n; k=1,...,p) and MAD the median absolute deviation, a resistant measure of spread. Similar choices of the tuning constant have proved to yield satisfactory results in other applications (e.g. Verboon, 1993).

Initialization of the algorithm

Given the above algorithm, one must find adequate starting values for the parameters to reduce the probability of ending at a local minimum of the objective loss function and to speed up the iteration process. It seems natural to start with finding a good estimate for the centroid $\mathbf{Z}$. To this end, we propose to compute all inner product (form) matrices $\mathbf{B}_k=\mathbf{X}_k\mathbf{X}_k'$. As already shown in the introduction, the diagonal elements of these matrices represent the square lengths of the vectors in the configurations, and the off-diagonal elements are related to the angles between the vectors. The next step is to compute $\mathbf{B}_0$ as the elementwise *medians* of the $\mathbf{B}_k$'s. So, (nxn) medians are computed, each time over p elements. After having found $\mathbf{B}_0$, a decomposition

$$\mathbf{B}_0 = \mathbf{Z}_0\mathbf{Z}_0', \tag{13}$$

is required so that the $\mathbf{Z}_0$ (nxm) can be used as the initial estimate of $\mathbf{Z}$. This decomposition is obtained from the eigenvectors $\mathbf{u}_1,...,\mathbf{u}_m$ and eigenvalues $\beta_1,...,\beta_m$ of $\mathbf{B}_0$ by defining

$$\mathbf{Z}_0= (\mathbf{u}_1\sqrt{\beta_1},..., \mathbf{u}_m\sqrt{\beta_m}), \tag{14}$$

where $\sqrt{\beta_j}$ is set to zero whenever $\beta_j < 0$ (j=1,...,m).

This way of initializing has two advantages: first, the matrices $\mathbf{B}_k$ are invariant over rotations of the matrices $\mathbf{X}_k$. Secondly, by using the median to find the elements in $\mathbf{B}_0$, we exclude the influence of vector outliers on the initial estimate, because outliers in the $\mathbf{X}_k$'s will cause deviant elements in the $\mathbf{B}_k$. From the way $\mathbf{B}_0$ has been defined, it is not necessarily true that $\mathbf{B}_0$ is positive definite. However, following Theorem 14.4.2 of Mardia *et al.* (1979), the approximation in (14) is the best possible, even if some of the eigenvalues are negative. This property justifies our use of $\mathbf{Z}_0$ as an initial estimate.

Example

The well known car model data (Donoho *et al.*, 1985) are reanalyzed here by separating the models according to their origin into those from America, from Europe and from Japan. For the cars of each origin k (=1,2,3), the 6-by-6 matrix of correlations $\mathbf{B}_k$ was obtained for the following six car variables: Gallons per mile (G), number of cylinders (C), displacement (D), horsepower (H), weight (W), and acceleration (A). These variables play the role of objects in the GPA analyses. The last object was analyzed with a negative sign to make it correlate positively with the other objects, as it is easier to visualize positive than negative correlations, the former being displayed as acute, the latter as obtuse angles.

Each correlation matrix was decomposed as $\mathbf{B}_k=\mathbf{X}_k\mathbf{X}_k'$ along its principal axes and the configurations of the six row vectors of $\mathbf{X}_1$, $\mathbf{X}_2$, and $\mathbf{X}_3$ were subjected to GPA by standard least squares (method LSQ) as well as by iteratively reweighted least squares with Huber weights (method HUB) and with Tukey biweights (method BIW).

Each GPA resulted in mutually rotated configurations $\mathbf{X}_1\mathbf{T}_1$, $\mathbf{X}_2\mathbf{T}_2$, and $\mathbf{X}_3\mathbf{T}_3$ as well as a centroid configuration $\mathbf{Z}$. Their first two dimensions are displayed in Figures 1, 2, and 3 for the results of the GPA's by methods LSQ, HUB, and BIW, respectively. Points are labeled only for objects A and C for which they do not cluster closely around the consensus vector. All these figures show that the correlations between the four objects G, D, H, and W were very high (very acute angles), whereas the number of cylinders C and minus acceleration A were less highly correlated with these four objects (less acute angles) and almost uncorrelated with each other (an angle of close to 90^{o}). The angular separation of the last two objects from the others was not the same for all sources (countries). The C vector was closer to the other vectors for American cars than for Japanese and European cars, indicating that the number of cylinders is more closely

connected to the features of US cars than to the features of cars of other sources. The A vector was more widely separated from other vectors for European than for other cars, which points to acceleration as being less correlated with the other features of European cars than with those of US and Japanese cars. It appears that US car makers are more prone to adjust the number of cylinders to the type of car, and that for European cars acceleration is not closely related to other characteristics of the cars.

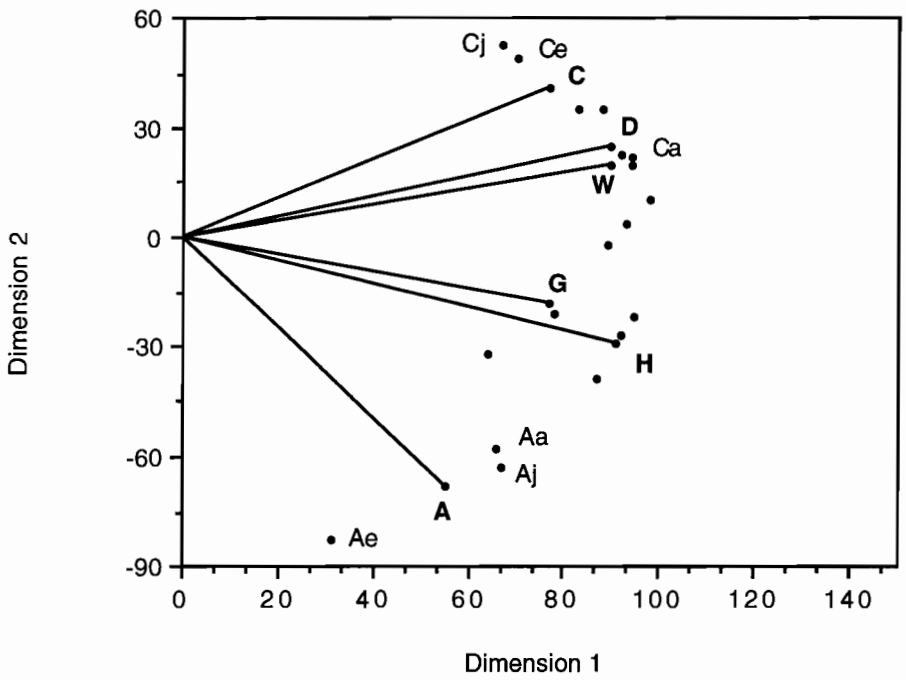

Figure 1. Results of GPA of Car Data with LSQ Method. The Six Variables Are Represented by Points for Each of the Three Groups (j,e,a) and the Centroid Configuration by Lines. Lower Case a,e,j in Label Indicate American, European, and Japanese, Respectively.

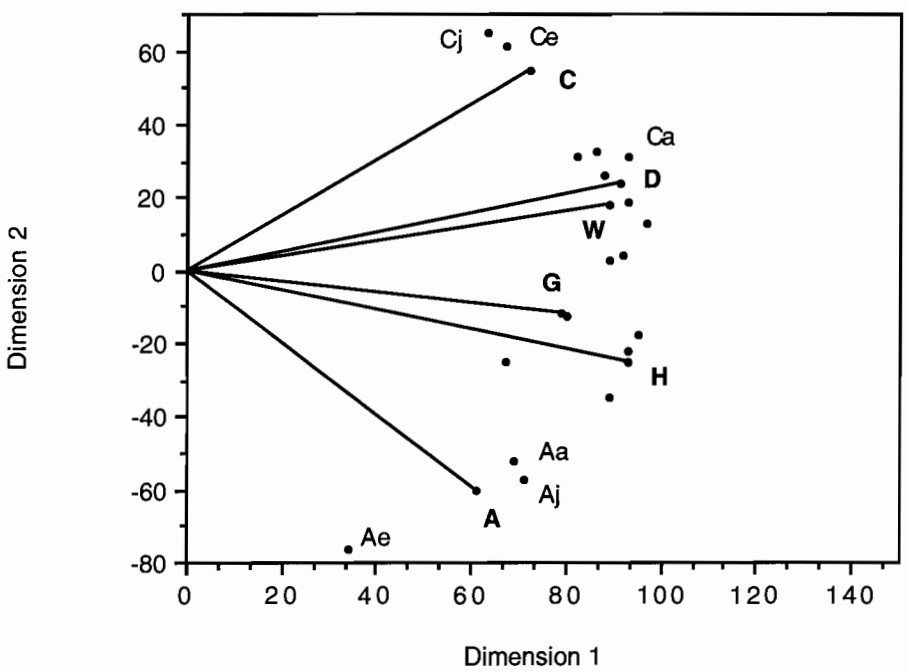

Figure 2. Results of GPA of Car Data with HUB Method. The Six Variables Are Represented by Points for Each of the Three Groups (j,e,a) and the Centroid Configuration by Lines. Lower Case a,e,j in Label Indicate American, European, and Japanese, Respectively.

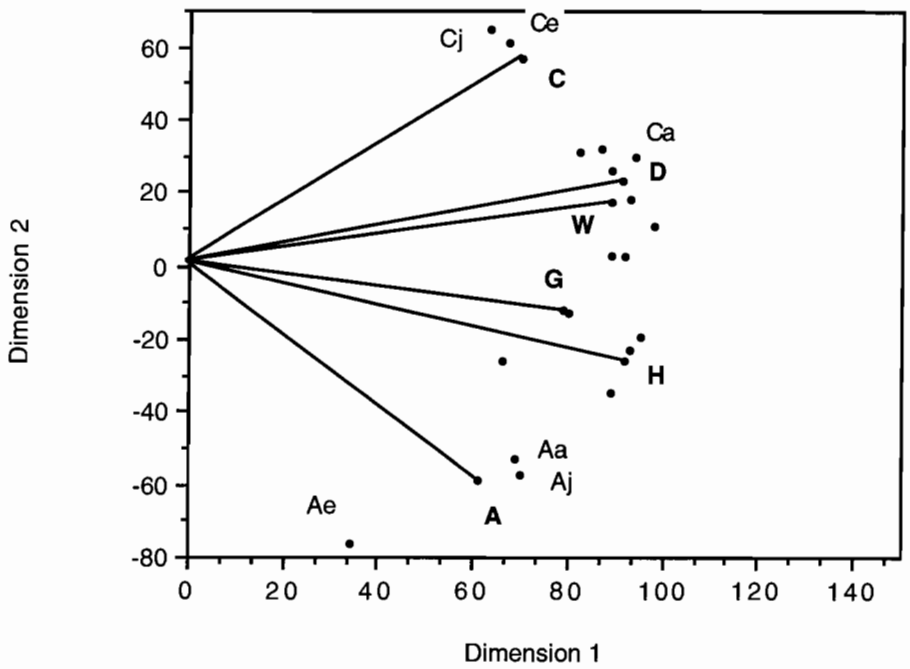

Figure 3. Results of GPA of Car Data with BIW Method. The Six Variables Are Represented by Points for Each of the Three Groups (j,e,a) and the Centroid Configuration by Lines. Lower Case a,e,j in Label Indicate American, European, and Japanese, Respectively.

The figures all show the centroid vector for each object to be in an average position relative to the three origins' vectors. This is a direct result of the Z-step of the GPA algorithm. For four of the objects this average is much the same on the three figures, and is evidently independent of the method of GPA used. Not so for the objects C and A, in which only the LSQ centroid vectors are really at centroid positions. For each of these objects the vector of one of the origins is well away from the other two vectors which are close together - the US vector outlying for object C and the European vector for object A. For these objects the centroid vectors obtained by methods HUB and even more so by method BIW, are very close to the two neighbouring vectors, being hardly

influenced at all by the one outlying vector. This illustrates reduction of the influence of outliers by resistant methods, especially by using biweights.

The resistant methods reduce the influence of outliers by downweighting them relative to other vectors. This is reflected in the weights each method has assigned to the various origins' objects, as shown in Table 1. Both resistant methods are seen to have assigned fairly similar weights to all vectors except C for American cars and A for European cars, which have lower weights. This is especially striking for method BIW in which these vectors receive zero weights. That method censors outliers more severely than method HUB's use of Huber weights. The weights evidently flag outliers effectively, and thus are a useful tool for data analysis, in addition to being instrumental in obtaining a resistant centroid.

INSERT TABLE 1

It is also of interest to compare the centroid configurations obtained by the three methods of GPA. To this purpose the three centroids were mutually rotated by GPA in their six-dimensional space, using biweights to highlight possible differences, and a common centroid obtained. Figure 4 presents the projections of these rotated centroids onto the first two principal axes of the common centroid. The overlap of the three methods' centroid vectors is very striking (for D, W, G, and H all centroids are plotted onto each other, and are therefore enclosed by circles in the figure), with two exceptions, those of method LSQ's centroid vectors for objects C and A. This brings home the influence of the outliers on ordinary least squares as compared to resistant methods.

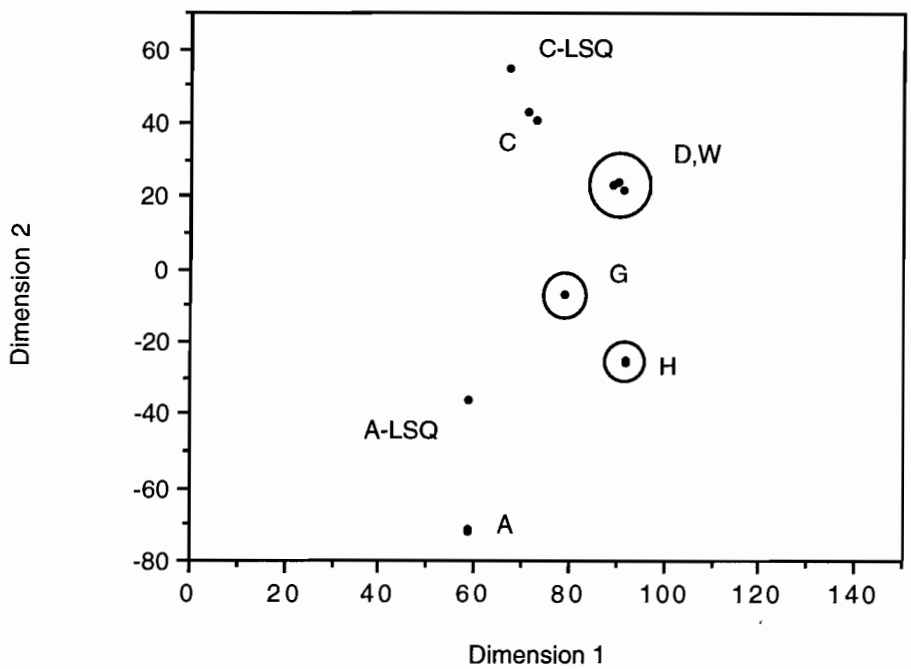

Figure 4. Results of GPA with BIW Method of Centroids Obtained from GPA's of Car Data. Two Atypical Points Are Labeled Individually.

Application of ordinary and iteratively reweighted least squares GPA to other examples has generally given similar results. We have noted, however, that potential outliers have little influence even with the least squares method, unless the number of outliers is relatively large compared to the total number of "good" points, meaning that there are many good points in a configuration or there are many configurations.

As only three configuration classes were compared in this example, it may seem a little bit far fetched to use the term outlier, just because one object differs from the other two. However, the example merely illustrates that weighted GPA can be used to highlight certain deviances from a general pattern. In this context the term outlier does not imply 'wrong observation', but that there is a point which deviates in some way.

A Simulation Study

To study the behaviour of the iterative weighting procedures systematically a simulation experiment has been carried out. In this experiment we have studied the effects of outliers on the form of a configuration class and we have examined the possible resistant properties of the approaches that use Huber weights or biweights.

Procedure

The simulation study used a four factor design for sets of configurations, with ten replicate sets t (t =1,...,10) at each factor combination. Each set was generated from an original configuration by rotation, random perturbation and outlier contamination. The factors of the design were

(1) the size n (n = 10, 20) of the configurations in the set,

(2) the number of rotated configurations p (p = 4, 8, 12) in the set,

(3) the error variance as a proportion ε (ε = 0.05, 0.20) of the variance of the elements of the rotated configurations in the set,

(4) the outliers as a percentage α (α = 0%, 5%, 10%, 15%, 20%) of the vectors in the set.

For each factor combination and each replication, each GPA method M (M = LSQ, HUB, BIW) was applied to the set of rotated configurations to obtain a centroid configuration $\mathbf{Z}_M^{(\alpha,\varepsilon,p,n,t)}$. The deviation of the form of this centroid from the form of the original configuration $\mathbf{Z}_t$ of that factor combination and replication was then measured as

$$L_M^{(\alpha,\varepsilon,p,n,t)} = \min_{\mathbf{R}} 1/n \, \| \mathbf{Z}_M^{(\alpha,\varepsilon,p,n,t)} - \mathbf{Z}_t \mathbf{R} \|^2, \quad \mathbf{R'R} = \mathbf{RR'} = \mathbf{I}, \tag{15}$$

where **R** is obtained by the ordinary Procrustes method. A small value of L_M indicates that GPA method M has closely recovered the form of the original configuration. In addition, for methods HUB and BIW the final weights were recorded separately for outlying vectors and normal vectors to see if the weights discriminate between these two types of vectors.

Details of the simulation study are as follows. For each size *n* original configurations were created by random drawing of *n* vectors from a store of some 500 two-element vectors whose coordinates had been generated independently from the uniform distribution with mean 0 and variance 1. For *number of configurations* p=4, the original configuration was subjected to rotations by angles (0, $4\pi/6$, $7\pi/6$, $11\pi/6$); for p=8, by angles (0, $\pi/6$, $3\pi/6$, $4\pi/6$,$7\pi/6$, $8\pi/6$, $10\pi/6$, $11\pi/6$); and for p=12, by angles (0, $\pi/6$, $2\pi/6$, ..., $11\pi/6$). For the *random error* of proportion ε, a uniform variable of mean 0 and variance ε times the variance of the configurations' elements was added to each coordinate. The percentage α *outlier contamination* was obtained by randomly choosing α percent of the vectors and multiplying them by -1.5. The minus sign turned these points into outliers with respect to form and the multiplication by a factor larger than one ensured that the outliers would not be too close to the origin, where their influence would have been negligible.

Results

The factorial design with the L-statistic taken as the dependent variable is analyzed by an analysis of variance. Per loss function the different sources with all second order interactions are examined to find out how the factors contribute to the variance of L_M. The analyses of variances of these 60 (2x3x2x5) averaged L_M values are given in the Tables 2, 3, and 4 for the three GPA methods.

INSERT TABLES 2,3,4

Each one shows that by far the strongest effect on the closeness of GPA fitting is that of the number of configurations compared. The average L_M is much larger for p=4 than for p=8 and 12, and somewhat larger for p=8 than for p=12. This trend with number of configurations has been confirmed by other simulations, and may be analogous to the law of increasing precision of estimates from larger samples.

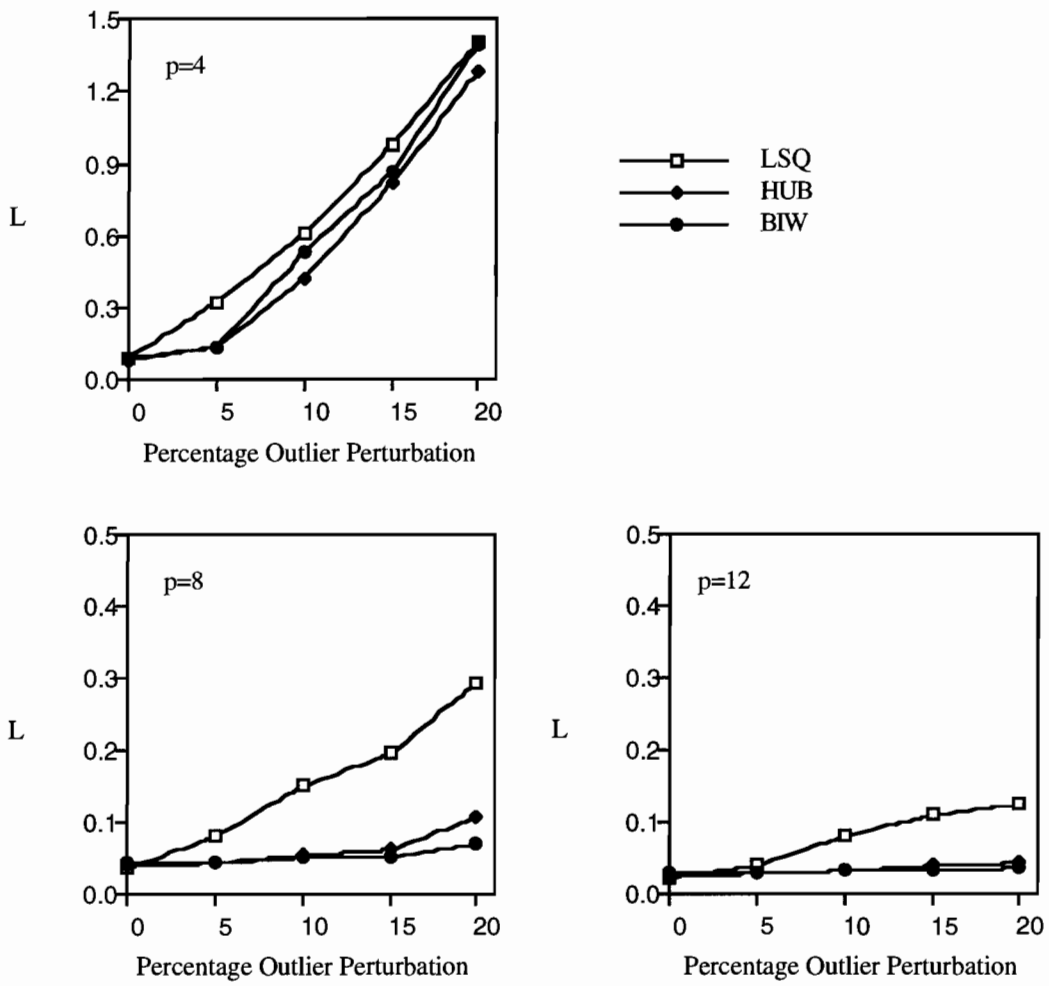

Figure 5. Effects of Sample Size and Outlier Perturbation on L. Each L Is Averaged over the Ten Replications, the Two Random Error Levels, and the Two Sizes.

The next largest effects, according to the analyses of variance, are those of the outlier contamination percentage α and its interaction $\alpha * p$ with the number of configurations. Average L evidently increases with increasing contamination, but does so most markedly when there are few configurations. This is evident from the left panels of Figure 5 in which L, averaged over n and ε, is seen to increase rapidly with α when p=4, but much more slowly when p=8 and 12.

These comments apply to all three methods of GPA, but the last one applies differently to the different methods. For LSQ the average L increases for all p's, though less rapidly for the larger number of configurations. By contrast, for BIW the increase of average L with p is quite small for p=8 and 12. This reflects the relative robustness of the different methods: LSQ is not robust for any p, HUB and BIW are very robust at p=12, whereas at p=8 HUB seems to be somewhat more robust than BIW.

The analyses of variance further suggest that there is some effect of random error ε on the fit of the GPA centroids, but that this does not interact with the other factors. They also show that size of configuration n has some effect on the fits by LSQ but none on the fit by HUB and BIW, that is, for LSQ the average L is a little larger for n=20 than for n=10, especially when p is small. As a result, for small p the robust methods, and especially HUB have a clear advantage only when n is larger.

It should be noted that in the last situation, small p and large n, HUB is found to be the most resistant whereas for larger p BIW was equally, if not more, resistant.

The variable weights can be inspected to see if smaller weights are assigned to the residuals corresponding to the outliers than to the residuals corresponding to the other points. In Figure 6 the weights assigned to the outliers are shown for all levels of p and α, averaged over sample size and random error levels. The figure also shows the weights that are assigned to the good points (averaged over all conditions, except outlier perturbations).

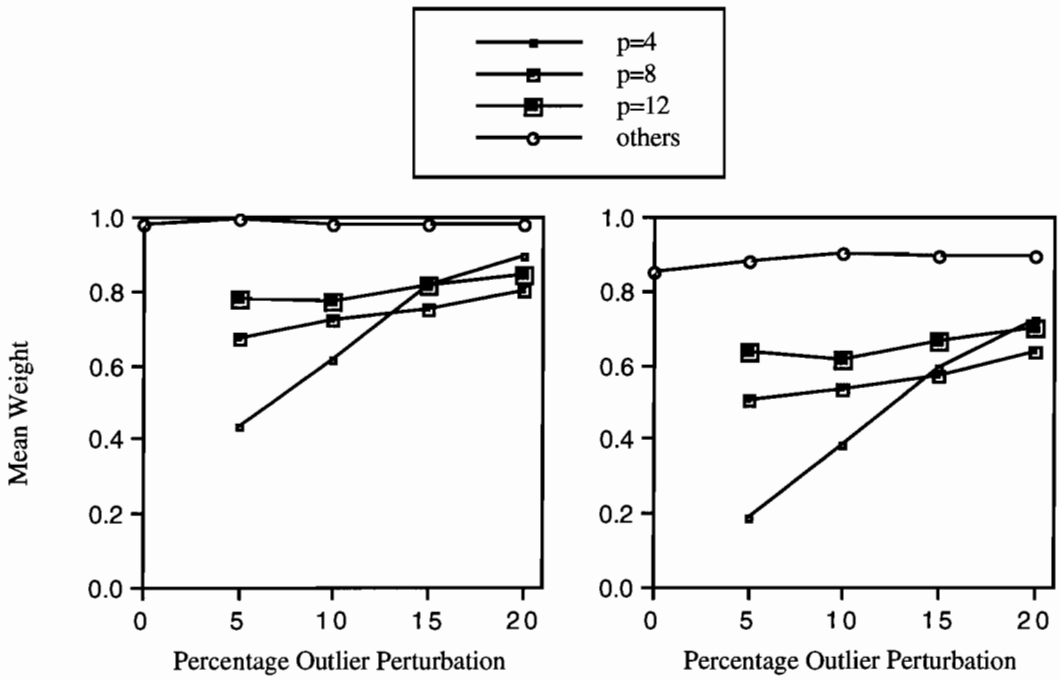

Figure 6. Weights Versus Proportion Outliers for HUB(left) and BIW(right). The Weights Are Averaged over n and ε for the Outliers, and over n, ε, and p for the Others.

It is clear that the weights distinguish quite well between outliers and other points. With increasing outlier perturbation the weights distinguish less well between outliers and other points, but this effect is particularly strong in the conditions with four sets. For p=4 the outliers are very well downweighted, while the other points obtain weights close to 1.0 for HUB, and close to 0.9 for BIW; for p=8 or p=12 the distinction between outliers and other points is much smaller but still clearly visible. The weights in BIW are generally smaller than those in HUB, but the patterns for both functions look quite similar.

Discussion and Conclusions

The proposed methods of adding weights to GPA seem to perform well in the presence of outliers. Not only could the influence of the outliers upon the estimate of the centroid be decreased, but the points which were outlying could be easily detected by inspecting the weights. The biweight GPA proved to be slightly more resistant than the GPA with Huber weights, except when the number of configurations compared was quite small.
The example revealed outliers by inspection of the weights, which would have been hard to find in the least squares context. The outliers did not have much impact upon the solution, but the small change in the centroid and the identification of the outliers was informative.
The simulation results indicate that up to 20% outliers in the data can reasonably well be handled by resistant approaches. Number of sets and amount of outlier perturbation appear to have much influence on the results. The number of objects and the level of random error seems to have only a moderate effect on the loss.

The algorithms using the Tukey weights sometimes converged to local minima. When we used random starts, that is the first centroid was randomly chosen, the final solutions were sometimes slightly different. Using the trimmed centroid as the starting value, the loss was never larger than in the random situation. This indicates that the median centroid start will be a reasonable defence against local minima, although we could never be sure of attaining the real global minimum.

Form a computational point of view it is not difficult to compute optimal dilation factors and translations in weighted GPA. In both the vectorwise and elementwise approach these parameters can be added in (5) and (9), respectively. Setting the partial derivatives of the functions with respect to one of these parameters equal to zero, immediately yields

the required result. However, as was also found in Verboon and Heiser (1992), these parameters tend to interfere with the weights, which could make the results of the analyses difficult to interpret. For this reason we have not exploited the idea of adding dilations factors and translations any further, and we have concentrated on the weights to achieve resistance.

In the initialization step of the algorithm negative eigenvalues are set to zero. Theoretically, a problem may arise when the eigenvalues are large and negative. If this might occur an alternative way to find the intial $\mathbf{Z}_0$ is to compute it from the median matrix, i.e. the matrix minimizing the sum of the absolute values $\Sigma_k \mid \mathbf{B}_0 - \mathbf{B}_k \mid$

References

Cliff, N. (1966). Orthogonal rotation to congruence. *Psychometrika, 31*, 33-42.

Commandeur, J.J.F. (1991). *Matching Configurations*. Leiden: DSWO-Press.

Donoho, A., Donoho, D.L., Gasko, M. & Olson, C.W. (1985). *MACSPIN Graphical Data Analysis Software*. Austin, Texas: D^2 Software Inc.

Goodall, C. (1991). Procrustes methods in the statistical analysis of shape. *Journal of the Royal Statistical Society, B, 53,* 285-339.

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

Green, B.F. (1952). The orthogonal approximation of an oblique structure in factor analysis. *Psychometrika, 17*, 429-440.

Huber, P.J. (1981). *Robust Statistics*. New York: Wiley.

Holland, P. W. & Welsch, R. E. (1977). Robust regression using iteratively reweighted least squares. *Comm. in Statistics, A6*, 813-827.

Kristof, W. & Wingersky, B. (1971). Generalizations of the orthogonal Procrustes rotation procedure to more than two matrices. *Proceedings of the 79th Annual Convention of the American Psychological Association, 6*, 89-90.

Krzanowski, W.J. (1988). *Principles of multivariate analysis*. London: Academic Press.

Mardia, K.V., Kent, J.T. & Bibby, J.M. (1979). *Multivariate analysis*. London; Academic Press Ltd.

Mosteller, F. & Tukey, J.W. (1977). *Data analysis and regression*. Massachus.: Addison-Wesley.

Peay, E.R. (1988). Multidimensional rotation and scaling of configurations to optimal agreement. *Psychometrika, 53*, 199-208.

Schönemann, P.H. (1966). A generalized solution of the orthogonal Procrustes problem. *Psychometrika, 31*, 1-10.

Sibson, R. (1978). Studies in the robustness of multidimensional scaling: Procrustes statistics. *Journal of the Royal Statistical Society, B, 40*, 234-238.

Ten Berge, J.M.F. (1977). Orthogonal Procrustes rotation for two or more matrices. *Psychometrika, 42*, 267-276.

Verboon, P. (1993). Robust nonlinear regression analysis. *British Journal of Mathematical and Statistical Psychology, 46*, 77-94.

Verboon, P. & Heiser, W.J. (1992). Resistant orthogonal Procrustes analysis. *Journal of Classification, 9*, 237-256.

Table 1. Weights Obtained From GPA Of Car Models.

	Huber (TC = .23)			Biweight (TC = .61)		
Objects	USA	EUR	JAP	USA	EUR	JAP
G	1	1	1	0.79	0.99	0.81
C	0.62	1	1	0.00	0.99	0.99
D	1	1	1	0.82	0.97	0.91
H	1	1	1	0.96	0.88	0.97
W	1	1	1	0.95	0.92	0.95
A	1	0.69	1	0.99	0.00	0.99

Table 2. Analysis of Variance For LSQ.

Source	df	SSQ	MS	F-ratio	Prob.
α	4	2.364	0.591	199.330	< .001
ε	1	0.013	0.013	4.432	0.044
α*ε	4	0.010	0.002	0.841	0.510
p	2	4.326	2.163	729.700	< .001
α*p	8	2.212	0.276	93.261	< .001
ε*p	2	0.003	0.001	0.429	0.655
n	1	0.019	0.019	6.249	0.018
α*n	4	0.042	0.011	3.548	0.017
ε*n	1	< .001	< .001	0.036	0.851
p*n	2	0.023	0.012	3.901	0.031
Error	30	0.089	0.003		
Total	59	9.100			

Table 3. Analysis of Variance For HUB.

Source	df	SSQ	MS	F-ratio	Prob.
α	4	1.522	0.380	79.814	< .001
ε	1	0.023	0.023	4.803	0.036
α*ε	4	0.027	0.007	1.408	0.255
p	2	3.300	1.650	346.110	< .001
α*p	8	2.495	0.312	65.439	< .001
ε*p	2	0.003	0.001	0.290	0.751
n	1	< .001	< .001	< .001	0.995
α*n	4	0.006	0.001	0.312	0.868
ε*n	1	0.006	0.006	1.217	0.279
p*n	2	< .001	< .001	0.003	0.997
Error	30	0.143	0.005		
Total	59	7.524			

Table 4. Analysis of Variance For BIW.

Source	df	SSQ	MS	F-ratio	Prob.
α	4	1.651	0.413	33.733	< .001
ε	1	0.012	0.012	1.021	0.320
α*ε	4	0.096	0.024	1.957	0.127
p	2	4.174	2.087	170.600	< .001
α*p	8	3.065	0.383	31.319	< .001
ε*p	2	0.001	< .001	0.032	0.968
n	1	0.012	0.012	1.018	0.321
α*n	4	0.038	0.010	0.785	0.544
ε*n	1	< .001	< .001	< .001	0.991
p*n	2	0.025	0.012	1.018	0.373
Error	30	0.367	0.012		
Total	59	9.441			