FACTOR ANALYSIS AND MULTIDIMENSIONAL SCALING

Vishwanath Mantha

Department for Electrical and Computer Engineering
Mississippi State University, Mississippi State, MS 39762
mantha@isip.msstate.edu

ABSTRACT

Factor Analysis (FA) is a multivariate statistical technique that is often used to create new variables that summarize all of the information that might be available in the original variables. It is used to study relationships that might exist among the measured variables in a data set. Similar to Principal Components Analysis (PCA), it is a variable-directed technique. One basic objective of FA is to determine whether the response variables exhibit patterns or relationships with each other. While PCA produces an orthogonal transformation of the variables and does not depend on an underlying model, FA does depend on a reasonable statistical model. Multidimensional scaling (MDS) is a mathematical technique that allows us to map the distances between points in a high dimensional space into a lower dimensional space. It is most useful when we can map into a two-dimensional space as this will help us visually confirm the different class groupings. The basic principle is to reduce distances between points in a two dimensional space. In this paper, we will cover the basic mathematical foundations of these two techniques and perform a comparative analysis using a test example.

1. INTRODUCTION

Multivariate data occur in almost all branches of science[1]. Whenever more than one attribute is measured for each experimental unit, the data is termed as “multivariate”. The statistical methods used to analyze such data are called multivariate statistical techniques. These are very useful for researchers who strive to make sense of large, complicated and complex data sets.One fundamental distinction between multivariate methods is that some are classified as “variable directed techniques” while others are classified as “individual directed techniques”. The former are primarily concerned with relationships that might exist among the response variables being measured while the latter are primarily concerned with relationships that might exist among the experimental units and/or individuals being measured. In this paper, we shall explore two such techniques, namely, Factor Analysis and Multidimensional scaling. The mathematical framework of these shall be covered in detail and will be applied to a test example.

2. FACTOR ANALYSIS

2.1. FA vs. PCA

Factor Analysis (FA) is a variable-directed multivariate statistical technique[2]. Principal Components Analysis (PCA) produces an orthogonal transformation of the variables and does not depend on an underlying model. FA, however does depend on a reasonable statistical model. FA is more concerned with explaining the covariance and/or correlation structure among the measured variables, whereas PCA is more concerned with explaining the variability in the variables.

2.2. Objectives of FA

Another basic purpose of factor analysis is to derive, create, or develop a new set of uncorrelated variables, called *underlying factors* or *underlying characteristics*, with the hope that these new variables will give a better understanding of the data being analyzed. These new variables can then be used in future analyses of data.

Whenever a large number of variables ar measured on each experimental unit, the variables are often related to each other in many different ways. The FA model assumes there is a smaller set of uncorrelated

variables that, in some sense, drives or controls the value of the variables that are actually being measured.

2.3. The FA model

Suppose one observes a *p*-variate response vector $\bar{x}$ from a population that has mean $\bar{\mu}$ and variance-covariance matrix Σ. The general FA model assumes there are *m* underlying factors (certainly we want $m < p$) denoted by $f_1, f_2, \ldots, f_m$ such that

$$x_j = \begin{array}{c} \mu_j + \lambda_{j1} \cdot f_1 + \lambda_{j2} \cdot f_2 + \ldots \\ \ldots + \lambda_{jm} \cdot f_m + \eta_j \end{array} \quad (1)$$

for $j = 1, 2, \ldots, p$.

2.3.1. Assumptions

In the preceding model we assume that

- the f_k's are independently and identically distributed with mean 0 and variance 1 for $k = 1, 2, \ldots, m$;
- the η_j's are independently distributed with mean 0 and variance ψ_j for $j = 1, 2, \ldots, p$; and
- f_k and η_j have independent distributions for all combinations of k and j, $k = 1, 2, \ldots m$ and $j = 1, 2, \ldots p$.

Also, it is almost always assumed that $\mu_j = 0$ and $var(x_j) = 1$ for every *j*. This can always be the case if one simply standardizes the measured variables before beginning a factor analysis. This is the default in almost all statistical software packages.

With the above assumptions, the new FA model becomes

$$x_j = \begin{array}{c} \lambda_{j1} \cdot f_1 + \lambda_{j2} \cdot f_2 + \ldots \\ \ldots + \lambda_{jm} \cdot f_m + \eta_j \end{array} \quad (2)$$

where the x's have been centered about their means.

2.3.2. Matrix form

In matrix form the above model becomes

$$\bar{x} = \bar{\Lambda} \cdot \bar{f} + \bar{\eta} \quad (3)$$

where $\bar{x}$ has been centered and

$$\bar{x} = [x_1, x_2, \ldots, x_p]' \quad (4)$$

$$\bar{f} = [f_1, f_2, \ldots, f_p]' \quad (5)$$

$$\bar{\eta} = [\eta_1, \eta_2, \ldots, \eta_p]' \quad (6)$$

$$\bar{\Lambda} = \begin{bmatrix} \lambda_{11} & \lambda_{12} & \ldots & \lambda_{1m} \\ \lambda_{21} & \lambda_{22} & \ldots & \lambda_{2m} \\ \ldots & \ldots & \ldots & \ldots \\ \lambda_{p1} & \lambda_{p2} & \lambda_{p3} & \lambda_{pm} \end{bmatrix} \quad (7)$$

In matrix form the FA model assumptions become

- $\bar{f} \sim (0, \boldsymbol{I})$
- $\bar{\eta} \sim (0, \bar{\psi})$ where $\bar{\psi} = diag(\psi_1, \psi_2, \ldots, \psi_p)$, and
- $\bar{f}$ and $\bar{\eta}$ are independent.

2.4. Factor Analysis Equations

Note that equation [3] implies that

$$\Sigma = Cov(\bar{x}) \quad (8)$$

$$= Cov(\bar{\Lambda} \cdot \bar{f} + \bar{\eta}) \quad (9)$$

$$= \bar{\Lambda} \cdot Cov(\bar{f}) \cdot \bar{\Lambda}' + \bar{\psi} \quad (10)$$

$$= \bar{\Lambda} \cdot I \cdot \bar{\Lambda}' + \bar{\psi} \quad (11)$$

$$= \bar{\Lambda} \cdot \bar{\Lambda}' + \bar{\psi} \quad (12)$$

Thus, to determine if $\bar{f}$, $\bar{\Lambda}$ and $\bar{\eta}$ exist that satisfy equation [3], we instead try to find $\bar{\Lambda}$ and $\bar{\psi}$ that satisfy equation [12].

The relationships described in equation [12] are called the *factor analysis equations.*

2.5. Non uniqueness of the factors

If $m > 1$, the factor loading matrix is not unique. That is, if $\bar{\Lambda}$ and $\bar{\psi}$ exist that satisfy equation [12], i.e

$$\Sigma = \bar{\Lambda} \cdot \bar{\Lambda}' + \bar{\psi} \quad (13)$$

$$= \bar{\Lambda} \cdot \bar{\boldsymbol{T}} \cdot \bar{\boldsymbol{T}}' \cdot \bar{\Lambda}' + \bar{\psi} \quad (14)$$

for every orthogonal matrix $\bar{T}$ .It then follows that

$$\Sigma = (\bar{\Lambda} \cdot \bar{T}) \cdot (\bar{\Lambda} \cdot \bar{T})' + \bar{\psi} \quad (15)$$

Thus, if $\bar{\Lambda}$ is a loading matrix, then $\bar{\Lambda} \cdot \bar{T}$ is also a loading matrix for every orthogonal matrix $\bar{T}$.

Hence the factor loading matrix is not unique. This leads to a new concept called *factor rotation.*

2.6. Solving the factor analysis equations

In this section, we consider solutions to the factor analysis equations. In practice, factor analysis procedures are almost always applied to the *Z* scores and to the correlation matrix $\bar{P}$. This is what we shall also assume here. Note the same techniques that will be discussed here can also be applied to the variance-covariance matrix also.

To determine whether a set of *m* underlying factors exist, one determines if $\bar{\Lambda}$ and $\bar{\psi}$ exist such that

$$\bar{P} = \bar{\Lambda} \cdot \bar{\Lambda}' + \bar{\psi} \quad (16)$$

The number of unknown quantities in $\bar{\Lambda}$ and $\bar{\psi}$ is $pm + p = p(m+1)$. The number of known quantities in $\bar{P}$ is $p(p+1)/2$ (since $\bar{P}$ is symmetric. Thus the FA equations give rise to $p(p+1)/2$ equations in $p(m+1)$ unknowns that must be solved.

If $p(m+1) > p(p+1)/2$ or, equivalently, if $m > (p-1)/2$, then there are more unknowns than there are equations, and no unique solution exists.

2.6.1. Choosing the number of factors

Before we actually begin to solve the FA equations, we should try to estimate or guess at the value of *m.* That is, guess how many underlying characteristics or factors there really are that are driving the values of the variables actually being measured.

One fairly good method that can be used to make an initial guess as to the number of underlying factors is to begin with a PCA, and determine how many principal components would be required to account for the variability. Then, we use this as the number of factors that are required. An initial guess as to the number of factors will not always agree with a final determination. Still, we have to start somewhere, and this may be as good a place to start as any.

There are also other objective methods for choosing the number of underlying factors. One of the more common methods is based on a maximum likelihood approach, which is based on the data having a multivariate normal distribution. One advantage of this approach is that it allows us to compute a likelihood ratio test (LRT) statistic that can be used to test the adequacy of the chosen number of factors[3].

Many statistical software packages use a statistic called the "Akaike's information criterion" (AIC)[4]. Supposedly the number of factors that provides the minimum value for AIC is considered to be the best choice.

2.6.2. Computer Solutions of the FA equations

Many different methods have been proposed for solving the FA equations. These include

- principal factoring with or without iteration,
- Rao's canonical factoring,
- alpha factoring
- image factoring
- maximum likelihood,
- unweighted Least-Squares factor analysis, and
- Harris factoring.

The first method is the most popular and will be discussed here.

The principal factoring method initially requires suitable estimates of the communalities, or, equivalently, estimates of the specific variances, $\psi_1, \psi_2, \ldots, \psi_p$.

Next since

$$\bar{P} = \bar{\Lambda} \cdot \bar{\Lambda}' + \bar{\psi} \quad (17)$$

must be satisfied, we must have

$$\bar{\Lambda} \cdot \bar{\Lambda}' = \bar{P} - \bar{\psi} \quad (18)$$

To obtain a unique solution for $\bar{\Lambda}$, we can force $\bar{\Lambda} \cdot \bar{\Lambda}'$ to be a diagonal matrix. Let $\bar{\Lambda} \cdot \bar{\Lambda}' = \bar{D}$.

Then

$$\bar{\Lambda} \cdot \bar{\Lambda}' = \bar{P} - \bar{\psi} \quad (19)$$

implies

$$\bar{\Lambda} \cdot \bar{\Lambda}' \cdot \Lambda = (\bar{\boldsymbol{P}} - \bar{\psi}) \cdot \Lambda \quad (20)$$

which implies

$$\Lambda \cdot \bar{\boldsymbol{D}} = (\bar{\boldsymbol{P}} - \bar{\psi}) \cdot \bar{\Lambda} \quad (21)$$

which by looking at the columns of both sides of this matrix equation, implies that

$$[d_1 \bar{\lambda_1}] = [(\bar{\boldsymbol{P}} - \bar{\psi}) \cdot \bar{\lambda}_1] \quad (22)$$

Similar equations ca written for $d_2, \ldots, d_m$. This implies

$$(\bar{\boldsymbol{P}} - \bar{\psi}) \cdot \bar{\lambda}_k = d_k \cdot \bar{\lambda}_k \quad (23)$$

for $k = 1, 2, \ldots, m$.

The only way the above equations can be true is if the diagonal elements of $\bar{\boldsymbol{D}}$ are eigenvalues f $\bar{\boldsymbol{P}} - \bar{\psi}$ and if the columns of $\bar{\Lambda}$ are their corresponding eigenvectors.

Any subset of *m* eigenvectors and eigenvalues would solve the FA equations, but the vectors corresponding to the *m* largest eigenvalues are chosen since the elements of $\bar{\boldsymbol{D}}$ are communalities and can be expressed as

$$d_k = \sum_{j=1}^{p} \lambda_{jk}^2, \quad (24)$$

for $k = 1, 2, \ldots, m$.

By choosing vectors corresponding to the *m* largest eigenvalues, we can maximize the communalities, which, in turn, tends to maximize the λ_{jk}'s (the factor loadings). Hence these vectors should correspond to the most important factors.

The above method can also be performed in an iterative fashion. It begins in the same way as described above. However, after an initial solution is found, the communalities corresponding to this solution are determined. Then these communalities are used as an initial guess, and the procedure starts all over again. This cycle is repeated until either all estimates converge or a nonsensical result is obtained. These possibilities appear to be equally likely when working with real data sets. Note that each iteration of this process actually produces a solution to the FA equations.

2.7. Factor Rotation

As stated earlier, when a set of factors is derived, they are not always easy to interpret. We may want to rotate the factors in hope of finding a set that is easy to interpret.

Many different orthogonal rotation algorithms have been developed. Orthogonal rotation procedures keep the factors uncorrelated whenever we start with a set of uncorrelated factors. Some orthogonal rotation methods that have been proposed are Quartimax, Varimax, Transvarimax, Equamax, and Parsimax. The most popular of these is the Varimax rotation procedure, which is described below.

Suppose

$$\bar{\boldsymbol{B}} = \bar{\Lambda} \cdot \bar{\boldsymbol{T}} \quad (25)$$

where $\bar{\boldsymbol{T}}$ is an orthogonal matrix. The sum of the variances of the squared loadings within each column of the factor matrix is a measure of simple structure. The Varimax criteria to maximize would then be

$$V^{\circ} = \sum_{q=1}^{m} \left(\frac{\left[\sum_{j=1}^{p} b_{jq}^4 - \left(\sum_{j=1}^{p} b_{jq}^2 \right)^2 \right] / p}{p} \right) \quad (26)$$

Note that the quantity within the larger parentheses in this expression is the variance of the squared loadings within the *q*th column of $\bar{\boldsymbol{B}}$. Since the squared loadings are all between 0 and 1, trying to maximize the variance of the squared loadings within a column is somewhat equivalent to trying to spread out the squared loadings within a column across the columns. The orthogonal matrix $\bar{\boldsymbol{T}}$ that produces a maximum to this sum of column variances results in the Varimax rotation of the factor loading matrix $\bar{\Lambda}$.

Some research has been done into rotation methods that are not orthogonal but oblique. The results obtained are not encouraging and hence these methods are not discussed here.

2.8. Factor Scores

Factor Analysis is frequently used to reduce a large number of responses to a smaller set of uncorrelated

variables. If this new set of variables is going to be used in subsequent statistical analyses, it is necessary to assign a score or value for each of the new variables for each experimental unit in the data set.

Evaluation of factor scores is not simple because the model for each individual is

$$\bar{x} = \Lambda \cdot \bar{f} + \bar{\eta} \tag{27}$$

with $\bar{\eta}$ unknown and $\bar{\Lambda}$ estimated. Thus for a given observation vector $\bar{x}$, it is not actually possible to determine $\bar{f}$ explicitly, although things should be improved if we consider only nontrivial factors.Two formal methods have been proposed for estimating $\bar{f}$ for a given individual. One of these is known as *Bartlett's method* and the other is known as *Thompson's method*. These are describe in the following sections.

2.8.1. Bartlett's method

After solving the FA model, we have

$$\bar{x} = \bar{\Lambda} \cdot \bar{f} + \bar{\eta} \tag{28}$$

where $\bar{\eta} \sim (0, \bar{\psi})$. Bartlett suggested that the next step be to find the $\bar{f}$ that minimizes

$$(\bar{z}_r - \bar{\Lambda} \cdot \bar{f}) \cdot \bar{\psi}^{-1} \cdot (z_r - \bar{\Lambda} \cdot \bar{f}) \tag{29}$$

where $\bar{z}_r$ is the standardized data vector for the *r*th individual. For a given $\bar{z}_r$, the preceding equation is minimized when

$$\bar{f}_r = (\bar{\Lambda}' \cdot \bar{\psi}^{-1} \cdot \bar{\Lambda})^{-1} \cdot \bar{\Lambda} \cdot \bar{\psi}^{-1} \cdot z_r \tag{30}$$

Then $\bar{f}_r$ is taken as the vector of estimated factor scores for the *r*th individual, $r = 1, 2, \ldots, N$.

2.8.2. Thompson's method

Thompson noted that for normally distributed data the joint distribution of $\bar{x}$ (standardized) and $\bar{f}$ is

$$\begin{bmatrix} \bar{z} \\ \bar{f} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \bar{P} & \bar{\Lambda} \\ \bar{\Lambda}' & I \end{bmatrix} \right) \tag{31}$$

This implies that the conditional expectation of $\bar{f}$ given that $\bar{z} = \bar{z}$ is

$$E[\bar{f} / \bar{z} = \bar{z}^{\circ}] = \bar{\Lambda}' \cdot \bar{P}^{-1} \cdot \bar{z}^{\circ} \tag{32}$$

Therefore, Thompson's method estimates the vector of factor scores for the *r*th individual as

$$\bar{f}_r = \bar{\Lambda}' \cdot \bar{P}^{-1} \cdot \bar{z}_r \tag{33}$$

There are also a couple of ad hoc methods for scoring factors. Even though both of these are ad hoc procedures, they hold some advantages over the formal procedures described. One advantage they hold over the formal procedures suggested above is that it is easy to explain what each means and it is very clear how each can be interpreted. These methods assume that only orthogonal rotations have been used.

When scoring a selected factor, all you really need is some variable that is

- highly correlated with the selected factor
- uncorrelated with other factors

Any procedure that produces factor scores having these two properties should be acceptable.

One way to score factors that would generally have the two properties listed earlier would be to take an adjusted average of all of the variables that have high correlation with that factor. The word adjusted is used here to mean that variables that have high positive correlations with the factor would be added, whereas those that have high negative correlations would be subtracted.

A second ad hoc way to score factors is to take the variable that has the highest correlation with the factor and use its value for the score of the factor.

3. MULTIDIMENSIONAL SCALING

Multidimensional Scaling (MDS) is a mathematical technique that allows us to map the distances between points in a high dimensional space into a lower dimensional space[6]. It is most useful when we can map distances from a high dimensional space into a two-dimensional space. In this case, the data points can be plotted in a two-dimensional space, and we can examine the plot to see which points tend to fall close to one another. Consequently, multidimensional scaling can be used as another technique to use when we want to cluster observations into groups.

To apply multidimensional scaling to a set of data points (possibly a set of group means), we must first calculate the distances between all pairs points. One of the most reasonable distance measures to use is to standardize the data, and then use the standardized ruler distance formula. A few common distance measures are described here:

•Ruler Distance: One simple measure of dissimilarity is the standard Euclidean distance, also called the *ruler distance.* This is the distance between two observations if we were able to plot the two observations in the p-dimensional sample space and measure the distance between them using a ruler. The formula that calculates the Euclidean distance between two points $\boldsymbol{x}_r$ and $\boldsymbol{x}_s$ is

$$\boldsymbol{d}_{rs} = [(\boldsymbol{x}_r - \boldsymbol{x}_s)' \cdot (\boldsymbol{x}_r - \boldsymbol{x}_s)]^{1/2} \quad (34)$$

•Standardized ruler distance: Another possibility for measuring the distance between a pair of points is to first standardize all of the variables and then compute the Euclidean distance between points using their standardized $\boldsymbol{Z}$ scores. For most situations, this is probably the best choice for measuring dissimilarities. The formula that calculates this is

$$\boldsymbol{d}_{rs} = [(\boldsymbol{z}_r - \boldsymbol{z}_s)' \cdot (\boldsymbol{z}_r - \boldsymbol{z}_s)]^{1/2} \quad (35)$$

•Mahalanobis distance: A third possibility is to compute a Mahalanobis-type distance between points. This would require estimates of within cluster variance-covariance matrices, after which these matrices would be pooled across clusters. The formula that calculates a Mahalanobis type distance is

$$\boldsymbol{d}_{rs} = [(\boldsymbol{x}_r - \boldsymbol{x}_s)' \cdot \Sigma^{-1} \cdot (\boldsymbol{x}_r - \boldsymbol{x}_s)]^{1/2} \quad (36)$$

• Note that Σ has to be replaced by a suitable estimate.

In the next section, mathematical formulas are given that reduce distances between points to a two-dimensional space. Note that these arguments can be extended to cases where we are attempting to reduce a p-dimensional data set to q dimensions where q is any number less than p. Because of a desire for simplicity, only the case where $q = 2$ is considered here.

3.1. MDS Mathematical Formulation

Suppose that $\boldsymbol{D}_{rs}$ represents the actual distance between the r^{th} and the s^{th} point in a p-dimensional sample space for $r = 1, 2, \ldots, N$ and $m = 1, 2, \ldots, N$ where N is the total number of data points

Let $\boldsymbol{z}_r$ and $\boldsymbol{z}_s$ be the standardized scores (measurements) corresponding to the r^{th} and s^{th} data points

Standardized ruler distance between these two points is

$$\boldsymbol{D}_{rs} = [(\boldsymbol{z}_r - \boldsymbol{z}_s)' \cdot (\boldsymbol{z}_r - \boldsymbol{z}_s)]^{1/2} \quad (37)$$

These distances can be ordered from smallest to largest. Let $\boldsymbol{D}_{r_1 s_1}$ be the distance between the two closest points, $\boldsymbol{D}_{r_2 s_2}$ be the distance between the next two closest points, ……, and $\boldsymbol{D}_{r_{N(N-1)/2} s_{N(N-1)/2}}$ be the distance between the farthest points. Note that the number of distinct pair of points is equal to $N(N-1)/2$, so there are $N(N-1)/2$ possible pairwise distances

Next consider plotting N points in a two-dimensional space and let the distance between the r^{th} and the s^{th} points be denoted by $\boldsymbol{d}_{rs}$. In this two-dimensional space, let $\boldsymbol{d}_{r_1 s_1}$ be the distance between the two closest points, $\boldsymbol{d}_{r_2 s_2}$ be the distance between the next two closest points, ……, and $\boldsymbol{d}_{r_{N(N-1)/2} s_{N(N-1)/2}}$ be the distance between the two farthest points. Multidimensional scaling tries to locate N points in a two-dimensional space so that the distances between the pairs of points in this space match as closely as possible with the true ordered distances between the observed points, namely,

$$\boldsymbol{D}_{r_1 s_1} < \boldsymbol{D}_{r_2 s_2} < \ldots < \boldsymbol{D}_{r_{N(N-1)/2} s_{NN-1/2}} \quad (38)$$

To assess the quality of the fit, it is customary to plot the actual differences between pairs of points against their modeled distances[7]. If the plot of these pairs of distances reveals a monotonically increasing

trend, then we can deduce that the two-dimensional plot accurately illustrates the closeness of the pair of points.

MDS programs attempt to locate the observed data points in a reduced dimensional space so that

$$E = \frac{\left[\sum_{r=1}^{N}\sum_{s=1}^{r-1}(\boldsymbol{D}_{rs}-\boldsymbol{d}_{rs})^2/\boldsymbol{D}_{rs}\right]}{\left[\sum_{r=1}^{N}\sum_{s=1}^{r-1}\boldsymbol{D}_{rs}\right]} \quad (39)$$

is minimized where $\boldsymbol{D}_{rs}$ is the distance between the r^{th} and the s^{th} observations and $\boldsymbol{d}_{rs}$ is the distance between those same two points in the reduced space.

The quality of this data embedding is measured by means of the stress function E. This formulation also gives rise to an alternate definition of MDS:

> •MDS takes a set of dissimilarities and returns a set of points in a two-dimensional space such that the distances between these points are directly proportional to the corresponding dissimilarities.

The minimization problem discussed above is sensitive to local minima and care should be taken to avoid this[6].

4. COMPARISON

4.1. FA vs. PCA

FA tries to explain the covariance and/or correlation structure whereas PCA tries to account for the variability structure. FA and PCA both perform transformations on the correlation matrix but FA assumes an underlying model unlike PCA.

4.2. FA vs. MDS

FA and MDS are both useful for clustering purposes[8]. When we have actual data points, it is always preferable to use FA. In cases where the distances between points is available rather than the actual data points, MDS is more useful for clustering.

5. EXAMPLE

We shall illustrate the applications of FA and MDS by way of a small example. The data for the example is shown in Table [1]. Nine subjects are asked six questions and their responses constitute the data points. Our objective is to determine which two subjects' answers are the closest - which two subjects are of a similar nature. This is a classic case of the need for clustering[9] and we shall perform the same using both FA and MDS. We use the software package SAS[10] to perform these analyses and study the detailed output.

5.1. FA Solution

The first step would be to calculate the correlation matrix from the given data. Once we have this, eigenvalue computation is done on the correlation matrix[11]. The next step would involve choosing the appropriate number of factors. For this example, hypothesis testing allows only two factors to be used. Then principal factoring with iteration is used to solve the FA equations.The values of the two factors obtained for each subject are given in Table [2]. The plot obtained is also shown in Figure [1].

	Q1	Q2	Q3	Q4	Q5	Q6
SUBJECT1	7	2	3	4	5	6
SUBJECT2	6	3	2	1	3	2
SUBJECT3	3	6	7	3	6	3
SUBJECT4	2	2	2	5	3	4
SUBJECT5	3	4	2	4	2	3
SUBJECT6	6	3	4	2	3	2
SUBJECT7	1	2	3	7	2	2
SUBJECT8	3	3	2	3	4	3
SUBJECT9	2	1	1	6	2	5

Table 1. Example Data

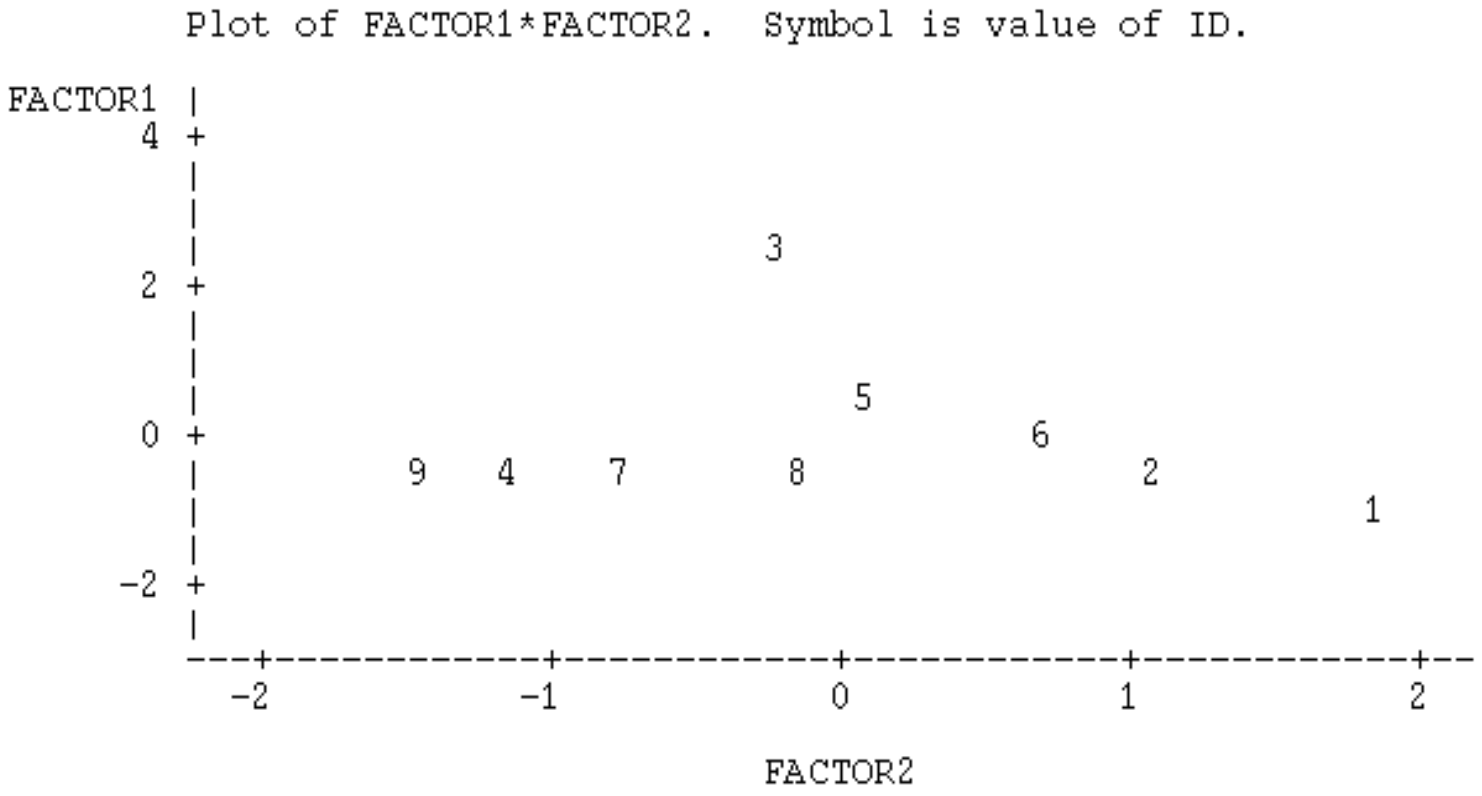

Figure 1. FA plot for the example described in section 5.

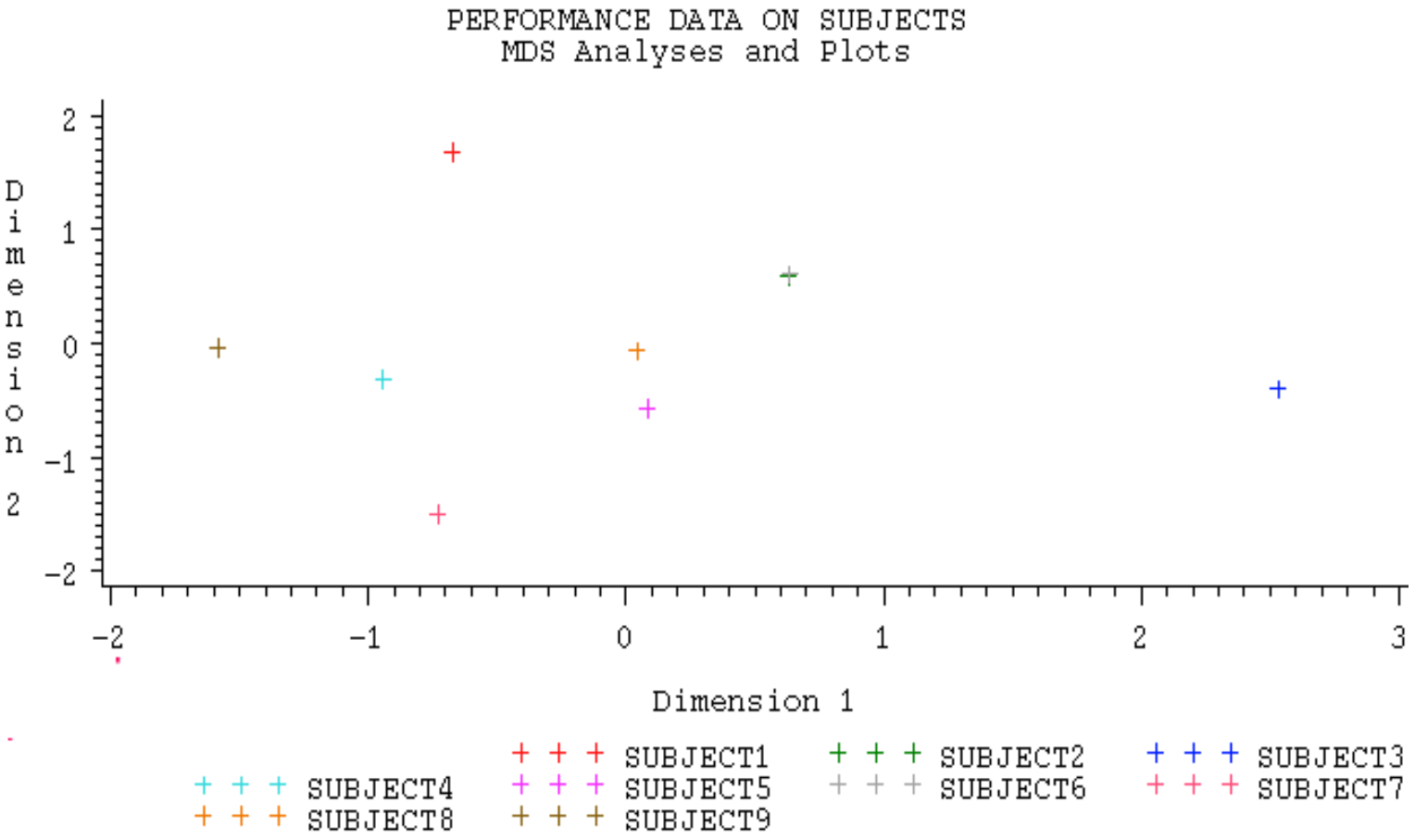

Figure 2. MDS plot for the example described in section 5.

	FACTOR 1	FACTOR 2
SUBJECT 1	-0.869	1.881
SUBJECT 2	-0.443	1.094
SUBJECT 3	2.439	-0.235
SUBJECT 4	-0.294	-1.181
SUBJECT 5	0.604	0.111
SUBJECT 6	0.168	0.730
SUBJECT 7	-0.604	-0.746
SUBJECT 8	-0.313	-0.184
SUBJECT 9	-0.686	-1.468

Table 2. Factor values for each subject

	DIM 1	DIM 2
SUBJECT 1	-0.671	1.688
SUBJECT 2	0.630	0.595
SUBJECT 3	2.529	-0.401
SUBJECT 4	-0.942	-0.313
SUBJECT 5	0.084	-0.568
SUBJECT 6	0.635	0.606
SUBJECT 7	-0.727	-1.499
SUBJECT 8	0.045	-0.068
SUBJECT 9	-1.585	-0.036

Table 3. New coordinates of the two-dimensional space

From the table and the plot, it can be inferred that subjects 4 and 9 are the closest.

5.2. MDS solution

The first step here would be to calculate the distance matrix using the standardized ruler distance metric. The next step would be do an iterative search to fit the data in a two-dimensional space so that the distances are more or less preserved. The metrics used for these are Gradient and Monotone convergence measures. The new coordinates of the two-dimensional space for each subject are shown in Table [3]. The plot of the data in the reduced two-dimensional space is shown in Figure [2]. According to the MDS solution, subjects 8 and 5 appear to be the most similar.

5.3. Comments

For the above example, the actual data points are available and hence the FA solution is more reliable as it has a more rigorous mathematical framework. Note that the MDS solution is also not far from the correct solution. Subjects 8 and 5 are also close to one another according to the FA plot also.

6. SUMMARY

This paper presents an overview of two multivariate statistical techniques, FA and MDS. FA is a useful technique for determining a set of underlying new variables, called *factors*, that appear to be driving the measurements that are being made with the original variables. MDS is a technique that is useful to map distances into a lower dimensional space and can aid in visualization of the data and clustering. Both these techniques were applied to a test example and their outputs were studied.

Relevant topics that can be further researched include oblique factor rotation, hypothesis testing techniques for choosing the number of factors and iterative search strategies in MDS.

7. REFERENCES

[1] D.E. Johnson, *Applied Multivariate Methods for Data Analysts*, Duxbury Press, Pacific Grove,California,1998.

[2] C.E. Spearman, “General Intelligence Objectively Determined and Measured”, *American Journal of Psychology*, vol. 15, pp. 201-293, 1904.

[3] K.G. Joreskog, and D. Sorbom, *Advances in Factor Analysis and Structural Equation Models*, Abt Books, Cambridge, MA, 1979.

[4] H. Akaike, “A New Look at the Statistical Identification Model”, I*EEE transactions on Automatic Control*, vol. 19, pp. 716-723, 1974.

[5] H. F. Kaiser, “The Varimax Criterion for Analytic Rotation in Factor Analysis”, *Psychometrika*, vol 123, pp. 187-200, 1958.

[6] F.W. Young, and R.M. Hamer, *Multidimensional scaling: History, Theory, and Applications*, Lawrence Erlbaum Associates, Hillsdale, NJ, 1987.

[7] H. Klock and J.M. Buhmann, Data Visualization by Multidimensional Scaling: A Deterministic Annealing Approach, *EMMCVPR’97*, Venice, Italy, 1997.

[8] L. Kaufman, and P.J. Rouseeuw, *Finding Groups in Data: An introduction to Cluster Analysis*, John Wiley, New York, NY, 1990.

[9] K. Fukunga, *Introduction to Statistical Pattern Recognition*, Macmillan, San Diego, CA, 1990.

[10] “Using the SAS system”, available at g*opher://gopher.vt.edu:10010/02/5/126*, July 1993.

[11] R.J. Schalkoff, *Pattern Recognition: Statistical, Structural and Neural Approaches*, John Wiley & Sons Inc., New York 1992.