

# Practical Assessment, Research & Evaluation

A peer-reviewed electronic journal.

Copyright is retained by the first or sole author, who grants right of first publication to *Practical Assessment, Research & Evaluation*. Permission is granted to distribute this article for nonprofit, educational purposes if it is copied in its entirety and the journal is credited. PARE has the right to authorize third party reproduction of this article in print, electronic and database forms.

Volume 7, Number 16, April, 2001

ISSN=1531-7714

## Profile Analysis: Multidimensional Scaling Approach

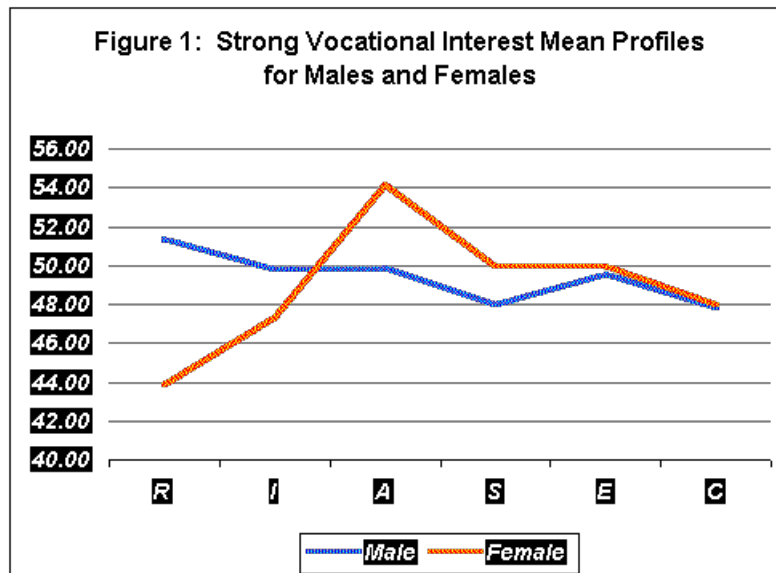
Cody S. Ding

Arizona State University

A great many current investigations, either in psychology or in education, deals with profiles of test scores. The terminology "profile" has been widely used in education settings to indicate a student's performance on a set of test scores such as reading, math, writing, and critical thinking skills. It is not uncommon that students receive test report with their score profiles, representing the strength and weakness in their performance on different tests. Due to this common practice in education, the profile analysis was sometimes considered by education practitioners as simply depicting test scores. In addition, a variety of exploratory techniques have been used to identify profile patterns in a set of data. These methods includes various cluster analytic approaches (e.g., Konald, Glutting, McDermott, Krush, & Watkins, 1999), configural frequency analysis (e.g., Stanton & Reynolds, 2001), and model profile analysis (e.g., Moses & Pritchard, 1995), which is a hybrid of cluster analysis and Cattell's (1967) Q-factor analysis. With the wide use of profile analysis, researchers and practitioners need to be aware of issues on the topic and how it should be dealt with. The goal of this paper is, thus, to introduce the problem and how it can be dealt with appropriately and provide an example of how profile analysis can be done, especially based on multidimensional scaling.

### How does profile analysis work?

The term "profile" comes from the practice in applied work in which scores on a test battery are plotted in terms of graph or profile. Figure 1 shows an example of profiles for males and females on six variables of Strong Vocational Interest (Strong, 1955): Realistic, Investigative, Artistic, Social, Enterprising, and conventional.



As shown in Figure 1, the profile provides three types of information for any person or group: level, dispersion, and the shape. The profile level is defined as an unweighted average of the scores in the profile, that is, the mean score over the six interest variables. In Figure 1, we would obtain level by adding scores on the six interest variables and dividing by 6 for males and females, respectively. It seems that the level for males (level = 49.41) is higher than that for females (level = 48.91).

The profile dispersion is defined as how much each score in the profile deviates from the mean. A measure of the dispersion is the standard deviation of scores for each person or group. In Figure 1, we would compute dispersion for males and females by subtracting the scores on each of the six interest variables from each group's level. Whereas it is possible to make a direct interpretation of level, it is difficult to do so for dispersion because the profile dispersions for people in general depend on the correlations among variables in the profile. According to Nunnally (1978), the most sensible way to interpret the dispersion of scores is to compare the dispersions of scores for two people or groups. In

Figure 1, profile for females has a much larger dispersion (dispersion = 3.43) than that for males (dispersion = 1.33) on the six interest variables. Thus, it may be concluded that females' interests are more variable than those of males.

The last information provided by profile is the profile shape. The profile shape is defined as the "ups" and "downs" in the profile and can be determined by the rank-order of scores. The high and low points in the profile indicate the salient characteristics of the person or group who resembles the profile. It should be noted that the actual appearance of a particular profile depends on the way variables are listed. Since it is arbitrary which variable is listed in which position, the physical appearance of the profile can be arbitrarily changed without impacting level, dispersion, or shape. Thus in Figure 1, females have the highest score on Artistic and the lowest score on Realistic, whereas males have the highest score on Realistic and lowest score on Conventional.

The profiles in Figure 1 are the simplest way of profile analysis in that all that involved is to plot a set of scores of the variables for any given person or group, as we often do for educational testing data. Profile analysis is, however, a generic term for all methods concerning groupings of persons. There are two sets of research problems underlying the use of profile analysis. One set of problems in profile analysis is that group memberships are known in advance of the analysis. In this case, the purpose of the profile analysis is to distinguish the groups from one another on the basis of variables in the data matrix. One such a problem, for example, could be start with groups of males and females in a population on vocational interests, as we have done in Figure 1. The purpose of analysis would be to distinguish these two groups in terms of variables used to measure the traits. This type of problems relates to multivariate analysis of variance (MANOVA) or discriminate analysis, which is concerned with a priori groupings of people. Thus in profile analysis via MANOVA, one simultaneously test hypotheses of a priori group differences on a set of variables; on conversely, one test hypotheses, in discriminate analysis, of differentiation of a priori groups with a set of variables and forms linear combinations of those variables that will most effectively differentiate in that regard.

The second type of problems in profile analysis occurs when group membership of people are not stated in advance of the analysis; the purpose of analysis is, thus, to classify people according to their scores on set of variables. Accordingly, while profile analysis via MANOVA or discriminate analysis is focused on the hypothesis testing about the extent to which a priori groups hang together, the clustering profiles is focused on discovering groups of people that hang together. However, it is important to note that the problem of test for significance of difference between groups on a set of variables considered simultaneously should not be confused with the major problem of profile analysis that we are attempting to deal with here, namely, classify people via proper methods. This is the topic that we now turn to.

### **Profile analysis via Multidimensional Scaling approach (PAMS) model**

As we mentioned above, there is a variety of exploratory techniques that have been used to identify profile patterns. This paper outlines an exploratory multidimensional scaling based approach to identifying the major profile patterns in the data, called Profile Analysis via Multidimensional Scaling (PAMS) that was originated by Davison (Davison, 1994). The Multidimensional Scaling is not a new technique but the profile interpretation of dimension is new. In comparison with other approaches of profile analysis (e.g., cluster analysis and configural frequency analysis), there are four major distinct features of the PAMS model. First, the model includes the Q-factor model (Cattell, 1967) as a special case. Furthermore, unlike techniques based on the Q-factor approach, the PAMS model can readily be applied to samples of any size. In fact, in PAMS analysis, the sample size consists of the number of variables, and to use that as N in the usual formulas forces one to consider random samples of variables in a domain.

Second, the PAMS model is designed for the study of latent "person." If one has theories concerning "types" among people, as opposed to "factors" among variables in factor analysis, one should use the PAMS model, which studies clusters of people and each cluster is a hypothetical "prototypical person." The prototypical person is defined in terms of "his/her" complete set of responses to the variables involved. Third, cluster analysis is very similar to factor analysis and it is designed to classify variables into "clusters" and it has been used to describe people in terms of discrete groupings. On the other hand, the PAMS model describes people in terms of continuous person profile indices that specify to what extent people are mixtures of the various types. That is, people are assumed to vary around each of several "prototypical" persons. Fourth, the person profile indices can be used either as predictors or criterion variables in subsequent analysis to study the relationships between individual profile patterns and other variables such as treatment outcomes. There is no such counterpart in other profile analysis technique procedures as far as the author knows.

I now turn to a description of the PAMS model and the analysis based on that model.

A full PAMS analysis involves two parts: (1) estimation of latent or prototypical profiles in a population, and (2) estimation of person profile index. Each person profile index quantifies the degree of correspondence between the observed profile of that individual and one of the latent or prototypical profiles. In this paper I will focus on the first step, estimation of latent or prototypical profiles in a population.

In classical MDS representations of structure, each data point is represented in a Euclidean space of continuous dimensions, that is:

$$m_{sv} = \sum_k w_{sk} x_{vk} + c_s + e_{vs} \quad (1)$$

Where  $m_{sv}$  is an observed score of subject  $s$  on variable  $v$ ,  $c_s$  is a level parameter, and  $e_{vs}$  is an error term representing measurement error and systematic deviations from the model. Each term in the sum on the right side of Equation 1 refers to a latent profile pattern  $K$  and this latent profile pattern corresponds to a multidimensional scaling dimension. The  $x_{vk}$ , the scale value parameter, equals the scores of variables in latent profile  $K$ . The  $w_{sk}$ , person profile index, is a measure of profile match that indexes the degree of match between the observed profile of subject  $s$  and the latent profile  $K$ . In an analysis based on the model of equation (1), the goal is to estimate the number of dimensions  $K$ , the scale value parameters,  $x_{vk}$ , along each dimension, the person profile index  $w_{sk}$ , and a measure of fit.

To estimate those parameters, one needs a set of assumptions and restrictions that uniquely identify the parameters (See Davison, 1996). One of the important assumptions is ipsative, which states that the mean of the scores in each latent profile equals zero, that is,  $\sum_v x_{vk} = 0.0$  for all  $K$ . Consequently, latent profiles will reproduce observed profile patterns (i.e., scatter plus shape), but not the level of the observed profiles which is reproduced by the level parameters. Under those assumptions, one can perform the parameter estimation based on the Squared Euclidean distance that satisfies the fundamental assumptions of common MDS analyses. One can, hence, estimate the scale value parameters in the PAMS model (i.e., Equation 1) by computing the squared Euclidean distance matrix for all possible pairs of variables and submitting that matrix to an MDS analysis that can be found in major statistical packages. The MDS analysis should produce one dimension for each latent profile  $K$ , along which the scale value,  $x_{vk}$ , for variable  $v$  is our estimate of the score for that variable in latent profile. The number of dimensions or latent profiles may be determined based on theory, previous research, or statistical methods. One such method is to use badness-of-fit index of .05 or less. One can choose a given dimensional solution based on which dimensional solution produces the smallest badness of fit index by the MDS analysis.

The PAMS model leads to the interpretation of MDS dimensions as latent profiles. Each MDS dimension  $K$  represents a prototype individual, that is, each MDS dimension represents a group of individuals with the similar characteristics. How does one interpret the level parameter  $c_s$ , scale value  $x_{vk}$ , and person profile index  $w_{sk}$ ? The parameter  $c_s$  equals the mean score in row; that is, it indexes the overall elevation of subject's profile, and it is therefore called level parameter. As mentioned earlier, Q-factor model (Cattell, 1967) is a special case of the PAMS model in which the observed data are standardized across subjects and therefore this level parameter,  $c_s$ , drops from the model because it equals zero for every subject  $s$ .

The set of scale values  $x_{vk}$  along each dimension represents a profile of scores for that prototypical person  $k$  on measure  $v$ . They indicate deviations from the profile level. The person profile index  $w_{sk}$  indicates how well a person resembles the prototypical or latent profile. The product of the person profile index,  $w_{sk}$ , and the scale values,  $x_{vk}$ , forms the profile patterns that provide addition information that is beyond and above that provided by the profile level (i.e., the average score). In contrast, in the profile analysis of priori defined group of individuals via analysis of variance, the information provided by the analysis is only based on profile level  $c_s$  of equation 1, as we illustrated in Figure 1.

## An example

To illustrate the profile analysis via Multidimensional Scaling (PAMS) approach, I used the same six variables of Strong Vocational Interest as in Figure 1 to identify "types" of people who have similar configurations of interests. Analysis was run for the total sample of 344 subjects since the emphasis here was on the results for the total sample. The scale value parameters in the model were estimated using the most common, nonmetric MDS analyses found in SAS (or other most major computer packages). The 2 dimensional solution was obtained using a non-metric scaling MDS procedure in SAS. For an illustrative purpose, below I included data for 10 subjects, the brief descriptions of the variables (the full discussion of the original terms can be found in Holland's 1963 paper), and the SAS program codes to perform the analysis.

| Obs | RTHEME | ITHEME | ATHEME | STHEME | ETHEME | CTHEME |
|-----|--------|--------|--------|--------|--------|--------|
| 1   | 45     | 39     | 38     | 56     | 51     |        |
| 2   | 47     | 51     | 56     | 53     | 48     | 47     |
| 3   | 65     | 67     | 51     | 39     | 54     | 55     |
| 4   | 45     | 54     | 43     | 45     | 45     | 45     |
| 5   | 52     | 47     | 51     | 49     | 49     | 54     |
| 6   | 64     | 52     | 53     | 32     | 58     | 45     |
| 7   | 54     | 57     | 29     | 45     | 56     | 62     |
| 8   | 52     | 57     | 39     | 57     | 59     | 54     |
| 9   | 60     | 57     | 55     | 64     | 65     | 54     |
| 10  | 49     | 54     | 57     | 47     | 48     | 48     |

RTHEM = Realistic, indicating people who enjoy activities requiring physical strength, aggressive action, motor coordination and skill.

ITHEME = Investigative, indicating task-oriented people who generally prefer to think through rather than act out problems.

ATHEME = Artistic, indicating people who prefer indirect relations with others, that is, dealing with environmental

problems through self expression in artistic media.

STHEME = Social, indicating people who prefer teaching role, which may reflect a desire for attention and socialization in a structured setting.

ETHEME = Enterprising, indicating people who prefer to use their verbal skills in situations which provide opportunities for dominating, selling, or leading others.

CTHEME = Conventional, indicating people who prefer structured verbal and numerical activities, and subordinate role.

```
OPTIONS nocenter nodate;

* Read in the data;
DATA one;
INFILE 'a:interest';

* Create dissimilarity matrix between variables;
PROC transpose DATA=one OUT=out ;
VAR rtheme--ctheme;
RUN;

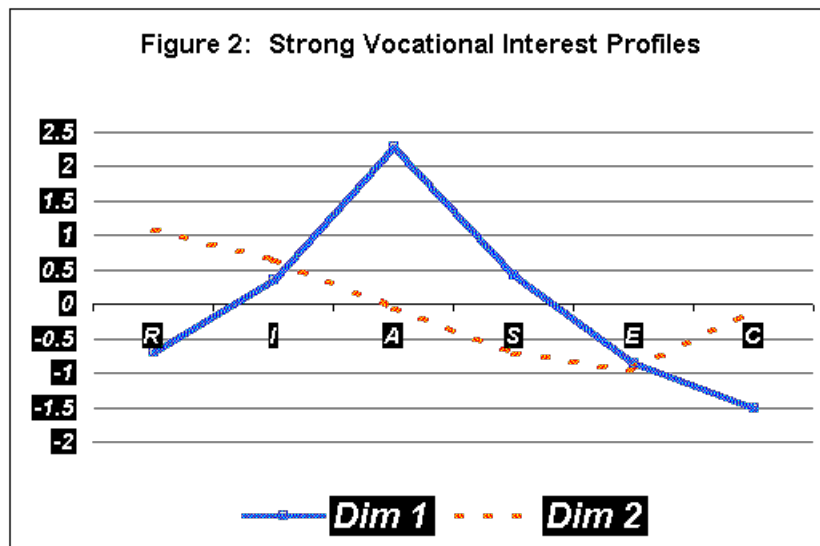
%INCLUDE 'C:\program files\SAS institute\sas\v8\STAT\SAMPLE\xmacro.SAS';
%INCLUDE 'C:\program files\SAS institute\sas\v8\STAT\SAMPLE\distnew.SAS';
%INCLUDE 'C:\program files\SAS institute\sas\v8\STAT\SAMPLE\stdize.SAS';
%DISTANCE ( DATA=out,METHOD=EUCLID, id=_name_, OUT=two );
RUN;

* Perform MDS analysis.
PROC MDS
DATA=two COEF=i DEC=2 DIM=2 FORMULA=1 LEVEL=ordinal FIT=2
CONDITION=matrix SHAPE=square NOP PFINAL OUT=coord;
RUN;
```

Table 1 shows the estimated scale values for these two dimensions and these scale values were plotted in Figure 2 as the latent profiles of the six interest variables.

Table 1: Estimated 2-Dimensional Scale Values of the Six Interest Variables

|        | Dim 1 | Dim 2 |
|--------|-------|-------|
| RTHEME | -0.70 | 1.09  |
| ITHEME | 0.36  | 0.65  |
| ATHEME | 2.28  | -0.04 |
| STHEME | 0.42  | -0.69 |
| ETHEME | -0.85 | -0.94 |
| CTHEME | -1.51 | -0.07 |



As I mentioned above, the actual appearance of a particular profile depends on the way variables are listed and since it is arbitrary which variable is listed in which position, the physical appearance of the profile can be arbitrarily changed without impacting profile level and profile patterns. Because of this reason, some researchers call high points in a latent profile a profile and low points in the latent profile a mirror image of the profile (Davison, 1994).

As shown in Figure 2, Dimension 1 has a high score on Artistic on the positive end and Convention/Enterprising on the negative end. This dimension looks like Prediger's (1982) Data vs Ideas Dimension. According to theory, the Data profile should be marked by the Conventional/Enterprising scales and the Ideas profile should be marked by the Artistic/Investigative scales, although in our profile the Investigative scale does not appear to be that salient as theory would lead one to expect. Thus, those whose interest profile resembles the Artistic vs. Convention profile shapes tend to be either more interested in Artistic than in Convention or more in Convention than in Artistic.

On the other hand, Dimension 2 has a high score on Realistic on the positive end and high scores on Social and Enterprising at the negative end. This dimension represents, according to theory, People vs. Things latent profiles (Prediger, 1982). People whose vocational interest profile resembles this profile shape tend to either involve in more realistic-type interest than social-type interest or involve in more social-type interest than realistic-type interest. That is, a subject resembles People profile would rather work with people than things and a subject resembles Things profile would rather work with things than people. Taken together, Dimensions 1 and 2 constitute a spatial representation of the major profile patterns in the data matrix as recovered by MDS.

## Conclusion

In this paper I examined several different approaches of profile analysis. Particularly, I illustrated profile analysis by simply depicting a set of test scores and that by PAMS model. Proper choice of a method for a specific investigation requires knowledge of the assumptions, limitations, and information utilized in the methods of measuring profile. It is important not to confuse the clustering of person in the PAMS approach with the group comparison on a set of scores in multivariate analysis of variance approach (MANOVA), as we did in the first example. It is also important to distinguish analyzing variables as in factor analysis from analyzing people as we did in here. In this latter approach, one is concerned with identifying prototypical persons in a population rather than identifying constructs from a set of variables. The PAMS model has advantages of being easily applied to samples of any size, classifying person on a continuum scale, and using person profile index for further hypothesis studies, but there are some caveats that need to be noted. First, the determination of number of dimensions or latent profiles is based on interpretability, reproducibility, and the fit statistics. Second, the interpretation of significance of the scale values is somewhat arbitrary. There is no objective criteria for decision making regarding which scale values are salient. Some researchers (e.g., Kim and Davison, 2001) used bootstrapping method to estimate standard error of scale values and statistical significance of these scale values was used for making decisions. Third, it is not well known as to what degree the latent profiles recovered by MDS solutions can be generalized across populations. Based on the limited research so far, it seems that different latent profiles may be recovered in two different populations such as in female population and male population, although the same latent profile solutions are expected to emerge on different data sets in the same population. More researches are needed in these areas.

## Note

A second dataset and corresponding SAS control cards can be found at <http://pareonline.net/htm/v7n16/profileanalysis.htm>

## References

- Cattell, R. B. (1967). The three basic factor analysis research designs: Their interrelations and derivatives. In D. N. Jackson & S. Messick (Eds.), *Problems in human assessment*. New York: McGraw-Hill.
- Davison, M. L. (1994). Multidimensional scaling models of personality responding. In S. Strack & M. Lorr (Eds.), *Differentiating normal and abnormal personality*. New York: Springer.
- Davison, M. L. (1996). Addendum to "Multidimensional scaling and factor models of test and item responses." Unpublished report, Department of Educational Psychology, University of Minnesota.
- Davison, M. L., Blake, R., & Sackett, S. (August, 1997). Relationships between personality and interest profiles: A canonical regression approach. Paper presented at the annual meeting of the American Psychological Association, Washington D.C.
- Holland, J. L. (1963). Explorations of a theory of vocational choice and achievement: II. A four-year prediction study. *Psychological Reports*, 12, 547-594.
- Kim, S. & Davison, M. L. (2001). Bootstrapping for estimating standard errors in multidimensional scaling. Unpublished paper presented to the American Educational Research Association, Seattle, WA.
- Konold, T. R., Glutting, J.J., McDermott, P.A., Kush, J. C., & Watkins, M.M. (1999). Structure and diagnostic benefits of a normative subtest taxonomy developed from the WIS-II standardization sample. *Journal of School Psychology*, 37, 29-48.
- Moses, J. A., Jr. & Pritchard, D. A. (1995). Modal profiles for the Wechsler Adult Intelligence Scale-Revised. *Archives of Clinical Neuropsychology*, 11, 61-68.
- Nunnally, J. C. (1978). *Psychometric theory* (2<sup>nd</sup> ed). New York: McGraw-Hill.
- Prediger, D. J. (1982). Dimensions underlying Holland's hexagon: Missing link between interests and occupations. *Journal of Vocational Behavior*, 21, 259-287.
- Stanton, H. C. & Reynolds, C. R. (2001). Configural frequency analysis as a method of determining Wechsler profile types. *School Psychology Quarterly*, 15, 434-448.
- Strong, E. K. Jr. (1955). *Vocational interests 18 years after college*. Minneapolis, MN: University of Minnesota Press.
- Tabachnick, B. G. & Fidell, L. S. (1989). *Using multivariate statistics* (2<sup>nd</sup> ed.). New York: HarperCollins Publisher.

### Author Notes

Correspondence relating to this article can be addressed to Cody Ding, 301 Payne Hall, Arizona State University, Tempe, AZ 85287, or via email at [cody.ding@asu.edu](mailto:cody.ding@asu.edu). The author would like to express his acknowledgments and appreciation to Mark Davison who provided a book and the articles from which I have drawn many of my ideas and arguments relating to this paper.

**Descriptors:** Profile analysis; scaling

**Citation:** Ding, Cody S. (2001). Profile analysis: multidimensional scaling approach. *Practical Assessment, Research & Evaluation*, 7(16). Available online: <http://PAREonline.net/getvn.asp?v=7&n=16>.