Profile Analysis of Multivariate Data in R: An Introduction to the profileR Package

Christopher David Desjardins
University of Iceland

Okan Bulut
University of Alberta

Abstract

Profile analysis is a multivariate data analysis technique that is the equivalent of a repeated measures extension of MANOVA. Profile analysis is mainly concerned with test scores; more specifically with profiles of test scores obtained from an educational or psychological assessment. A test score profile shows differences in subscores on tests that are commonly administered in medical, psychological, and educational studies to rank participants of a study on some construct. Practitioners are typically interested in quantifying both an individuals' overall performance on the subtests (i.e., the *level*) and the variation between the scores on the subtests (i.e., the *pattern*). A suite of profile analytic procedures for decomposing observed scores into both level and pattern effects exists for the R programming language in the **profileR** package (Bulut and Desjardins 2015). This package includes routines to perform criterion-related profile analysis, profile analysis via multidimensional scaling, moderated profile analysis, profile analysis by group, and a within-person factor model to derive score profiles. This article describes several of these methods, illustrating their application with data sets included within the package, as well as describing the future direction for the **profileR** package.

Keywords: Profile analysis, univariate, multivariate, criterion-related patterns, psychometrics, R, **profileR**.

1. Introduction

A profile is a score vector that includes a person's scores from a test consisting of multiple subtests, subscales, or strands, measuring reading or mathematics abilities or some other construct. Features in a test score profile, such as elevation, dispersion, and shape, bring valuable pieces of information about the person, and collectively summarize average, variance, and rank order of test scores within a profile (Watkins, Glutting, and Youngstrom 2005). A test score profile conveys information about a person's strengths and weaknesses in certain
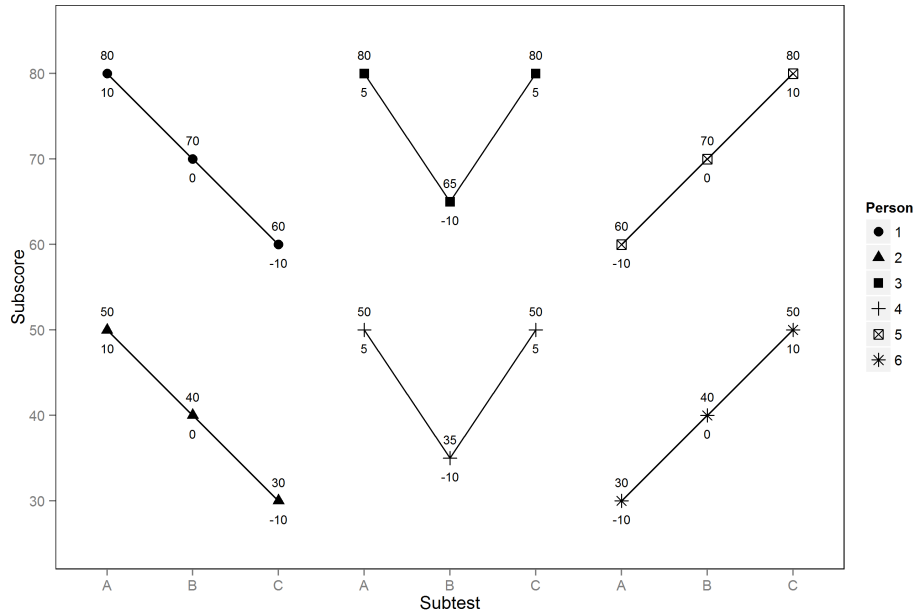
Figure 1: Score profiles of six individuals

domains, skills, or abilities as measured by a battery of tests. In education, test score profiles can be used to identify academic strengths and areas malleable to growth and can be used to guide the development of instructional interventions or student placement. In psychological settings, a test score profile can be used to identify which traits, or ability factors, manifest themselves to some degree in the pattern of a person's clinical profile on a latent construct.

The mean score of a score vector of tests (or subtests) indicates the *level* of a score profile and can be interpreted as a measure of the overall strength of the profile. If the level is subtracted from each score in a score profile, a vector of deviations containing the *pattern* in the score profile is obtained (Davison, Kim, and Close 2009). Variation in profile levels of all examinees can be interpreted as the variation among the examinees' average profile score (i.e., between-person variation). Variation in the profile pattern of an examinee can be interpreted as within-person variation in the score profile (Davison *et al.* 2009). Figure 1 shows est score profiles of six individuals on three domains of a hypothetical test battery. For each person, the three scores in the person's profile vector are shown above the profile while each person's deviation scores are shown below the profile. Individual differences in the profile level can be seen by comparing within an column for the three profiles. Comparing within rows and across columns, the variation in the deviation scores (i.e., the pattern effect) is visible. The first two test score profiles display a linearly decreasing pattern; the second two display a V shape pattern; and the last two display a linearly increasing pattern.

2. Profile analysis theory

Profile analysis involves the analysis and quantification of the elevation, variation, and configuration of multiple test scores to identify individuals among subgroups (Cronbach and Gleser 1953; Stanton and Reynolds 2000). These test scores could arise from a single test composed

of multiple subtests, repeated assessments in a classroom, parallel or alternate forms of a test, or a psychological instrument(s) measuring multiple latent domains. Henceforth, we will refer to these subtest and multiple tests generically as "variables". The most common form of profile analysis relies on a combination of multidimensional scaling and factor analytic techniques which extract patterns in a score profile in order to average latent scores and the variance of the profile scores (Davison, Gasser, and Ding 1996; Davison and Kuang 2000; Davison, Kuang, and Kim 1999). The utility and validity of profile analysis for analyzing patterns of individual test takers' scores have been discussed extensively in the literature (e.g., Davison *et al.* (1996) and Watkins *et al.* (2005)).

Profile analysis can be used by researchers or practitioners to identify whether two or more groups of test takers have significantly distinct or similar score profiles. It can be used to quantify the amount of variability associated with the level and pattern effects. It serves as an aid in the analysis and interpretation of individual patterns. It may be conducted across or within an individual. Furthermore, profile analysis allows for both inter- and intra-individual interpretations of test scores from an individual and the quantification of the degree of similarities with observed test score profiles.

There are several statistical frameworks that implement profile analysis. For example, cluster analysis involves the partitioning of individuals into meaningful classes where both the number of classes and the composition of the classes can be evaluated (Kaufman and Rousseeuw 2008). This method primarily focuses on the identification of profile patterns and the classification of individuals based on their observed score profiles. Cluster analysis may be performed in R using the `kmeans` function provided in the **stats** package. If data are hierarchical, then the `hclust` function, also provided in the **stats** package, should be considered. In addition, other classification frameworks (e.g., classification trees) could be considered and can be fit using the **rpart** package (Therneau, Atkinson, and Ripley 2015).

The **profileR** package implements several profile analytic methods described in Bulut (2013), Davison (1994), Davison and Davenport (2002), and Davison *et al.* (2009) as well as other published and unpublished techniques. The methods in the **profileR** package differ from the existing classification methods in R by focusing on the quantification of variability associated with the different components of a profile without the need for classification of individuals into classes or groups.

The **profileR** package is in version 0.3 on CRAN and has been in active development for 3 years. The **profileR** package has long been a part of the CRAN Psychometric Task View. Development of the package occurs at `https://github.com/cddesja/profileR` and all are encouraged to contribute pull requests and file feature requests on the project's GitHub site.

2.1. Assumptions

There are several statistical assumptions in profile analysis that are similar to the assumptions in MANOVA. First, the test scores should have a multivariate normal distribution. This assumption usually holds if there are more subjects in the smallest cell of the data than the number of variables and if the sample sizes are equal across the variables. This assumption can be assessed by checking skewness and kurtosis of the variables. If a variable is not normally distributed marginally, then transformations (e.g., log or square-root transformations) may be considered. In addition, univariate and multivariate outliers should be addressed prior to performing profile analysis as they may produce either a Type I or Type II error and

give no indication as to which type of error may have occurred in the analysis. Fortunately, there are a multitude of packages available in R, such as the **MVN** (Korkmaz, Goksuluk, and Zararsiz 2014), **mvoutlier** (Filzmoser and Gschwandtner 2015), and **mvnormtest** (Jarek 2012) packages, that can readily and robustly assess this assumption.

The second assumption is homogeneity of the variance-covariance matrices of the test scores. If the sample sizes are equal, this assumption is usually not an issue (French, Poulsen, and Yu 2015). However, if sample sizes are unequal, then a test for homogeneity (e.g., Box's M test) may be performed. Box's M is implemented in the **biotools** package (da Silva 2015) in R and could be used. However, care must be taken as Box's M is highly susceptible to non-normality.

The final assumption is that the variables are linearly related. When the variables are normally distributed and the sample size is large enough, this assumption is typically not an issue. However, in case of any evidence against normality or unequal sample sizes, this assumption should be assessed by constructing joint distribution plots of the test scores.

2.2. Profile plots

A profile plot is a graphical exploratory profile analytic technique for examining the relative behavior of all variables in a multivariate data set. Profile plots can be created by plotting the sample means for each variable for each individual, for a single group, or across multiple groups. An individual profile plot examines the behavior of all variables but only on a subset of the data (e.g., looking at all the test scores only for female examinees).

To have a clear and meaningful graphical summary of the variables in the data, all of the variables must have the same units of measurement prior to plotting. For instance, if weight-related variables are measured in different units (e.g., grams, kilograms, and pounds), they must be placed on the same scale of measurement or transformed into a standardized score (e.g., z scores). In a typical profile plot, the sample means for each group or scores for each person are plotted against the measured domains. Figure 2 shows test score profiles of two students on four content domains (reading, math, science, and social skills) on a general aptitude test. One of the principal purposes for creating a profile plot is exploratory in order to assess whether the profiles are parallel. In Figure 2, the lines appear to be parallel across the four content domains. Profile analysis can be used to test whether the lines are indeed parallel. The details of this procedure are explained in the following section.

2.3. Profile analysis with Hotelling's T^2 test

We can formally test the null hypothesis that the univariate mean of the test scores in a profile are equivalent using Hotelling's T^2 test. For this test, hypothesis testing proceeds as follows:

$$\text{H}_0 : \mu = \mu_0 \text{ against } \text{H}_a : \mu \neq \mu_a \tag{1}$$

which is equivalent to testing the following null hypothesis:

$$\text{H}_0 : \frac{\mu_1}{\mu_1^0} = \frac{\mu_2}{\mu_2^0} = \cdots = \frac{\mu_p}{\mu_p^0} = 1 \tag{2}$$
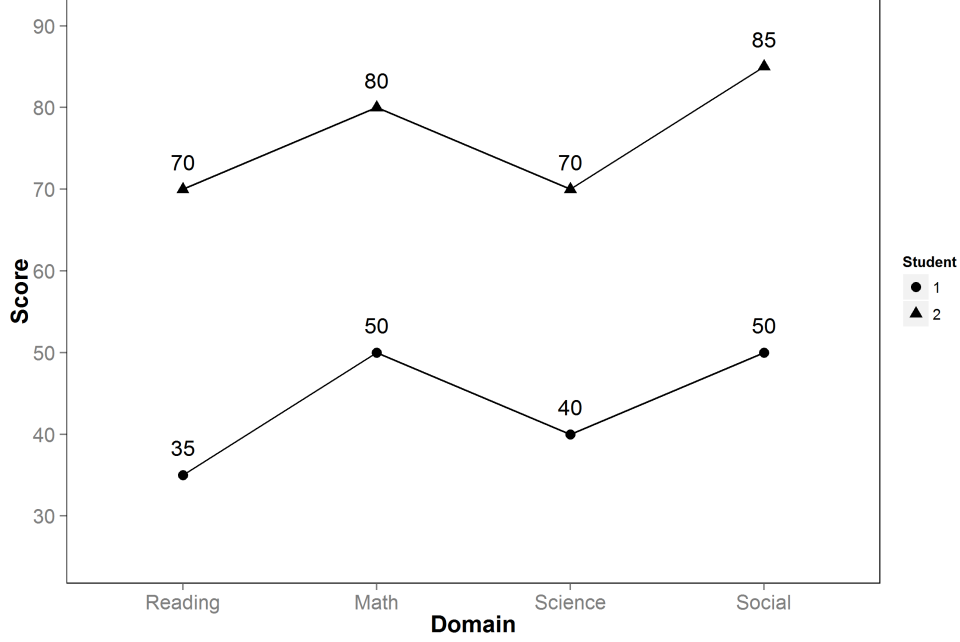
Figure 2: A profile plot with two persons and four content domains

against the alternative that at least one of these ratios is not equal to 1:

$$H_a : \frac{\mu_j}{\mu_j^0} \neq 1 \text{ for at least one } j \in \{1, 2, \ldots, p\} \tag{3}$$

Furthermore, instead of testing whether the ratios of the means over their hypothesized means are equivalent to one, profile analysis can be used for testing the null hypothesis that all of these ratios are equivalent. After rejecting the null hypothesis in Equation 2, a second null hypothesis can be tested as:

$$H_0 : \frac{\mu_1}{\mu_1^0} = \frac{\mu_2}{\mu_2^0} = \cdots = \frac{\mu_p}{\mu_p^0}. \tag{4}$$

A typical test of the profile using Hotelling's T^2 would proceed as follows:

Step 1: The differences between successive ratios are computed. For this computation, the ratio of the $j + 1^{\text{th}}$ variable over its hypothesized mean is calculated and then this ratio is subtracted from the ratio of jth variable over its hypothesized mean:

$$D_{ij} = \frac{X_{ij+1}}{\mu_{j+1}^0} - \frac{X_{ij}}{\mu_j^0}, \tag{5}$$

where D_{ij} represents this difference for observation i. It should be noted that testing the null hypothesis that all of the ratios are equal to one another is equivalent to testing the null hypothesis that all the mean differences are equal to 0:

$$H_0 : \mu_D = 0. \tag{6}$$

Step 2: Hotelling's T^2 test is then applied to test the null hypothesis that the mean of the differences is equal to 0.

To motivate the use of the Hotelling's T^2, consider the square of the t-statistic for testing a hypothesis regarding a univariate mean. Under the null hypothesis, the t statistic has a distribution with $n-1$ degrees of freedom:

$$t = \frac{(\bar{x} - \mu_0)}{s/\sqrt{n}}. \tag{7}$$

Now consider squaring this test statistic as shown below:

$$t^2 = \frac{(\bar{x} - \mu_0)^2}{s^2/n} = n(\bar{x} - \mu_0)\left(\frac{1}{s^2}\right)(\bar{x} - \mu_0) \sim F_{1,n-1} \tag{8}$$

If a t-distributed random variable with $n-1$ degrees of freedom is squared, a random variable with an $F_{1,n-1}$ distribution is obtained. The null hypothesis of H_0 is rejected if $t^2 > F_{1,n-1,\alpha}$, where α typically is 0.05.

In the expression for Hotelling's T^2, the difference between the sample mean and the μ_0 in the univariate t-test is replaced with the difference between the sample mean vector, $\bar{x}$, and the hypothesized mean vector, $\boldsymbol{\mu_0}$. The sample variance is replaced by the inverse of the sample variance-covariance matrix $\mathbf{S}$, yielding the matrix form of the equation:

$$T^2 = n(\bar{x} - \boldsymbol{\mu_0})'\mathbf{S}^{-1}(\bar{x} - \boldsymbol{\mu_0}) \tag{9}$$

For large n, T^2 is approximately chi-square distributed with p degrees of freedom. If the sample variance-covariance matrix $\mathbf{S}$ is replaced by the population variance-covariance matrix $\boldsymbol{\Sigma}$, then Hotelling's T^2 becomes

$$T^2 = n(\bar{x} - \boldsymbol{\mu_0})'\boldsymbol{\Sigma}^{-1}(\bar{x} - \boldsymbol{\mu_0}). \tag{10}$$

This test is exactly chi-square distributed with p degrees of freedom. For small samples, the chi-square approximation for T^2 does not take into account variation due to estimating the sample variance-covariance matrix $\mathbf{S}$. Therefore, better results can be obtained from the transformation of the Hotelling's T^2 statistic to an F statistic as shown below:

$$F = \frac{n-p}{p(n-1)}T^2 \sim F_{p,n-p}. \tag{11}$$

Under the null hypothesis, $\mathrm{H}_0 : \mu = \mu_0$, this will have a F distribution with p and $n-p$ degrees of freedom. The null hypothesis, H_0, is rejected if the test statistic F is greater than the critical value from the F-table with p and $n-p$ of freedom, evaluated at level α as follows:

$$F > F_{p,n-p,\alpha}. \tag{12}$$

As for other statistical tests, it is necessary to consider the assumptions under which the Hotelling's T^2 test is carried out, and to assess which of those assumptions may be satisfied for the data. There are four assumptions underlying Hotelling's T^2 test:

1. The data from population i is sampled from a population with mean vector $\boldsymbol{\mu_i}$. This assumption essentially means that there are no subpopulations with different population mean vectors.

2. The data from both populations have common variance-covariance matrix, $\boldsymbol{\Sigma}$. This assumption may be assessed using Bartlett's Test by testing the null hypothesis that $\boldsymbol{\Sigma_1}$ from population 1 is equal to $\boldsymbol{\Sigma_2}$ from population 2 against the general alternative that these variance-covariance matrices are not equal.

3. The subjects from both populations are independently sampled. It should be noted that this assumption does not require the variables to be independent of one another. The results of Hotelling's T^2 test are not generally robust to violations of this assumption.

4. Both populations should have a multivariate normal distribution. According to the Central Limit Theorem, the sample mean vectors tend to be approximately multivariate normally distributed regardless of the distribution of the original variables as the sample size becomes larger. Thus, Hotelling's T^2 test may not be adequately sensitive to violations of this assumption.

Again, many of these assumptions can be readily assessed in R.

2.4. Profile analysis by groups

In profile analysis, multivariate data are typically collected across multiple groups, time points, variables, etc. A profile consisting of multivariate data can be examined as lines in a profile plot (as described earlier) representing a set of variables across time points or groups. Profile analysis then allows us to answer three basic questions that arise from this plot:

1. Are the groups parallel between time points or variables?

2. Are the groups at equal levels across time points or variables?

3. Do the profiles exhibit flatness across time points or variables?

If the answer to any of these questions is no (i.e., that a specific null hypothesis is rejected), then there is a significant effect. The type of effect depends on which of these null hypotheses is rejected.

Parallelism

In profile analysis, parallelism is usually the main test of interest because the test examines whether each segment is the same across all individuals or groups. A segment in this context is simply the difference in the response between time points or variables. Therefore, the segment is equivalent to the slope of the line between two points on the x-axis. Parallelism is assessed using a multivariate test which compares the multiple segments of the profile.

A one-way MANOVA can be used to test whether there is evidence for significant non-parallelism between the groups. The within-group variance comes from subtracting the segment matrix for each individual in the group from the group mean. The between-group variance is obtained by subtracting each group mean segment matrix from the grand mean

segment matrix. If the null hypothesis of parallelism is rejected, there is a significant interaction between group membership and the measured variables or group membership and the time points (e.g. if a test is repeatedly administered).

Equal levels (coincident profiles)

The test of equal levels across the profiles is the most straightforward test in profile analysis. Given that the profiles are parallel, an equal level test examines whether the profiles coincide (i.e., there are no group differences). The test is used for determining whether one group scores higher on average across all the variables (e.g., domains or subtests) or time points. To evaluate this, the grand mean of all time points or variables is calculated for each group. Since all of the time points or variables are collapsed into a group mean, this becomes a univariate test. This is equivalent to a between-groups main effect in ANOVA. Mathematically, the test simply measures the relative contributions of between-group and within-group variations to the total sum of squared errors. This should look familiar, as it is the basis behind simple ANOVA models. For i groups measured on j variables:

$$\sum_i \sum_j (X_{ij} - \bar{X}_{tot})^2 = np \sum_j (\bar{X}_j - \bar{X}_{tot})^2 + p \sum_i \sum_j (X_{ij} - \bar{X}_j)^2. \tag{13}$$

where $\bar{X}_{tot}$ is the grand mean, n is the sample size of each group, and p is the number of groups. Based on this test, if the group levels are significantly different from one another, then the null hypothesis of equal levels is rejected.

Flatness

Flatness evaluates the extent to which the profiles are flat (i.e., no differences in mean response within any group) given the profiles are parallel. The flatness null hypothesis is that the segments are 0 (i.e., the slope of each line segment is zero and the profile is flat). This is evaluated independently for each group, making this a within-subjects test. If the line is not flat (i.e., any of the segments vary significantly from 0) then there is a within groups main effect of the time point (if the data are from repeated assessments) or a measured variable (e.g., multiple subtests or a test measuring multiple domains). Hotelling's T^2 test can be used to test the difference of the zero-matrix and the segmented data for each group as follows:

$$T^2 = N(\bar{\mathbf{X}}_{\mathbf{tot}})'(\mathbf{S}_{\mathbf{wg}})^{-1}(\bar{\mathbf{X}}_{\mathbf{tot}}) \tag{14}$$

where $\bar{\mathbf{X}}_{\mathbf{tot}}$ is the grand mean, N is the number of segments, and $\mathbf{S}_{\mathbf{wg}}$ is the within-group variance-covariance matrix. Wilk's λ can then be calculated from the T^2 statistic using the following equation:

$$\lambda = \frac{1}{1 + T^2} \tag{15}$$

It should be noted that if parallelism is rejected, the other two tests (equal levels and flatness) are meaningless. In that case, flatness may be assessed within each group, and various within- and between-group contrasts may be analyzed.

2.5. Criterion-related profile analysis

Davison and Davenport (2002) extended their previous work in profile analysis by defining pattern and level components as linear components of a score profile and considered whether the profiles had any theoretical or applied validity with respect to an external, dichotomous criterion. They termed this approach as criterion-related profile analysis.

Consider the test score data from a personality inventory presented in Table 1 (available in **profileR** via IPMMc dataset).

Table 1: Inventory of Personality and Mood Manifestation.

A	H	S	B	C
75	60	50	50	1
60	75	45	55	1
60	60	55	45	1
50	50	75	60	0
45	55	60	75	0
55	45	60	60	0

Note. A = anxiety, H = hypochrondriasis, S = schizophrenia, B = bipolar disorder, and C corresponds to the criterion, 1 = neurotic and 0 = psychotic. Table adapted fromDavison and Davenport (2002).

If we wanted to predict whether an individual's personality profile is related to whether they are neurotic or psychotic we may consider the following regression model:

$$C_P = \beta_0 + \beta_1 X_{PA} + \beta_2 X_{PH} + \beta_3 X_{PS} + \beta_4 X_{PB} \tag{16}$$

where the βs correspond to the intercept and the regression weights and $X_{PA} \dots X_{PB}$ correspond to the predictor variables in Table 1. We can re-express this equation in terms of the *pattern* and *level* effects.

$$C_P = \beta_0 + \sum_V (\beta_V - \beta_.)(X_{PV} - X_{p.}) + \sum_V \beta_. X_{P.} \tag{17}$$

where β_0 again corresponds to the intercept, the first sum corresponds to a linear combination of the elements in person p's pattern vector which is defined as the difference between the domain scores (i.e., A, H, S, and B) and the overall mean score for person p, $(X_{PV} - X_{p.})$ (the pattern effect), multiplied by the difference between the regression weight associated with a domain from the personality inventory, β_V, and the mean regression weight, $\beta_.$, and the second summation term is a function of person p's level, $X_{p.}$, defined as $1/V \sum_V X_{PV}$, where V is the number of measured domains (i.e., 4 in this example), multiplied by the mean regression weight. Further, Davison and Davenport (2002) show that this equation can be expressed as:

$$C_P = \beta_0 + (V/k)Cov_{pc} + V\beta_. X_{p.} \tag{18}$$

where $Cov_{pc} = (1/V) \sum_V (X_{PV} - X_{P.})(k\beta_V - k\beta_.)$ is the covariance between the pattern effect and the scores in the criterion-pattern vector, $\mathbf{x}_c = k(\beta_V - \beta_.)$, where k is some scalar greater than 0 and β_V and $\beta_.$ were defined above.

Rewriting the multiple regression of the criterion variable onto the test scores in this fashion, allows the calculation of the proportion of variability associated with the level and pattern effects, respectively, and allows for the testing of several hypotheses which are described later.

2.6. Profile analysis via multidimensional scaling (PAMS)

One of the exploratory techniques that has been used to identify patterns in test score profiles is profile analysis via multidimensional scaling (PAMS) (Davison 1994). This approach examines the major profile patterns by estimating latent profiles. The PAMS model is more advantageous than other profile analytic methods because of the following reasons: 1) it can be used with any sample size; 2) unlike typical factor analysis, it takes the person component into consideration when estimating latent profiles; 3) it creates continuous person profile indices, rather than discrete categories, that represent the extent that people possess a mixture of the various measured abilities/constructs/domains; and 4) the latent person profile indices can be used either as predictors or criterion variables to describe the relationships between individual profile patterns and other variables such as treatment outcomes (Ding 2001).

In the PAMS analysis, latent person profiles and person profile indices corresponding to these profiles are estimated simultaneously. Each person profile index represents the degree of similarity between a person's observed profile and each of their estimated latent profiles. A Euclidean space is created to indicate the distance between each data point and the continuous dimensions (i.e., latent profile indices). The PAMS model can be written as:

$$m_{ij} = \sum_{k=1}^{K} w_{ik} x_{jk} + c_i + e_{ij}, \tag{19}$$

where m_{ij} is an observed score of person i on variable j $(j = 1, \ldots, J)$, w_{ik} is the person profile index that indicates the distance between the observed profile of person i and the estimated latent profile k, x_{jk} is the scale value parameter that equals the scores of variables in latent profile k $(k = 1, \ldots, K)$, c_i is the profile level (i.e., the average of J scores for person i), and e_{ij} is an error term representing measurement error and systematic deviations from the model. As described above, w_{ik}, x_{jk}, c_i, and e_{ji} define a latent profile pattern k that corresponds to a multidimensional scaling dimension (Ding 2001). PAMS first finds the number of latent profile dimensions (K), then estimates the scale value parameters (x_{jk}) for each dimension, and finally computes the person profile index (w_{ik}), and a measure of overall fit for the model.

The scale value parameters (x_{jk}) represent a latent profile k with j scores. These parameters indicate deviations from the profile level. The person profile index w_{ik} indicates how closely a person reflects the latent profile k defined by the x_{jk} values. The product of x_{jk} and w_{ik} leads to the profile patterns of persons that can be described as additional information above and beyond what the profile level (c_i) explains for each person. Unlike a typical ANOVA model, the analysis of a group of persons on a set of variables via profile analysis provides information that is not only based on the level but is also based on the pattern.

In order to estimate the parameters defined in Equation 19, PAMS requires several assump-

tions and restrictions (Davison 1996). First, the mean of the scores in each latent profile should equal zero (i.e., $\bar{x}_{jk} = 0$ for each k). Therefore, latent profiles can reproduce observed profile patterns, but not the level of the observed profiles which is reproduced by the level parameters. Second, the number of dimensions or latent profiles is user-defined based on theory, previous research, or statistical methods such as choosing the solution with either a badness-of-fit index of .05 or less or the smallest badness of fit index (Ding 2001).

Under these assumptions, the parameters in the PAMS model can be estimated by computing the squared Euclidean distance matrix for all possible pairs of variables and one can then use this distance matrix to perform multidimensional scaling analysis (which is available in many major statistical programs). For instance, the `cmdscale` function in R can analyze such data. This analysis produces one dimension for each latent profile k, in which the scale value, x_{jk}, for variable j is the estimate of the score for that variable in latent profile (Davison 1996; Ding 2001).

2.7. Profile reliability

The aim of a test battery is to obtain reliable subscores that can be used to evaluate examinees' skills for diagnostic, classification, or selection purposes. The higher the reliability of the subscores, the more precisely examinees are able to be evaluated based on their test scores. If the purpose of an assessment is to use each subscore to compare individuals, then traditional reliability indices such as coefficient alpha (Cronbach 1951) and Kuder-Richardson 20 (Kuder and Richardson 1937) can be used.

Examining an individual's strengths and weaknesses from pattern scores has brought some uncertainties due to lack of evidence for subscore reliability and validity (Watkins *et al.* 2005). Profile reliability is conceived of as a multivariate analog to the traditional definition of univariate reliability as the ratio of the variance of true scores to the variance of observed scores. In order to estimate the precision of unique patterns of test score profiles, Bulut (2013) proposed an approach for estimating the reliability of individual differences in test score profiles based upon the total variation, the variation among individuals, and the variation among the subscores. This approach is mainly an extension of canonical test reliability originally proposed by Conger and Lipshitz (1973).

Conger and Lipshitz (1973) defined the observed difference vector as $\overrightarrow{X}_i - \overrightarrow{X}_{i.}$; where $\overrightarrow{X}_i$ is the vector of scores for person i and $\overrightarrow{X}_{i.}$ is the average of person i's scores. Using the observed and true difference vectors, canonical reliability can be written for any distance function as:

$$\rho_p = \frac{(\overrightarrow{T}_i - \overrightarrow{T}_{i.})^{\mathrm{T}} \mathbf{A} (\overrightarrow{T}_i - \overrightarrow{T}_{i.})}{(\overrightarrow{X}_i - \overrightarrow{X}_{i.})^{\mathrm{T}} \mathbf{A} (\overrightarrow{X}_i - \overrightarrow{X}_{i.})}, \tag{20}$$

where $(\overrightarrow{T}_i - \overrightarrow{T}_{i.})$ is the true difference vector and $\mathbf{A}$ is a square matrix used for weighting the reliability. The square matrix $\mathbf{A}$ can either be a correlation matrix of the subscores on the test or an identity matrix if weighting is not desired (Conger and Lipshitz 1973).

Using the canonical reliability framework defined by Conger and Lipshitz (1973), Bulut (2013) presented a profile reliability approach that made use of both the vector of difference scores and the vector of the average score in a test score profile. For a given person, the level

represents the mean of the subscores in the profile which can be expressed as

$$\bar{X}_p = \frac{1}{v} \sum_v X_{pv}, \tag{21}$$

where X_{pv} is the score of person p ($p = 1, \ldots, P$) on domain v ($v = 1, \ldots, V$) and $\bar{X}_p$ is the average scores from V domains for person p. When Equation 21 is applied to each person in the data, a level vector that consists of P level scores is obtained.

The pattern of a test score profile is a vector of the score differences (i.e., ipsatized scores) between each score and the mean of the scores for a given person. A pattern vector of a test score profile can be shown as $(X_{p1} - \bar{X}_p), \ldots, (X_{pv} - \bar{X}_p)$; where X_{pv} is the score of person p on test v and $\bar{X}_p$ is the mean of the subscores that person p obtained from V tests. Using the level and pattern components defined above, the total score variation (T) of a test score profile can be defined as the sum of the variances for the V tests:

$$T = \sum_{v=1}^{V} \sigma_v^2; \tag{22}$$

where the total variation (T) is the sum of two orthogonal components: between-person variation (B) and within-person variation (W) (Davison *et al.* 2009). The B corresponds to the level effect and W is corresponds to the pattern effect. Essentially, B is the between-person variation due to individual differences in profile level; and W is the within-person variation due to individual differences in profile patterns (Bulut 2013; Davison *et al.* 2009).

To define reliability based on between-person variation and within-person variation, the relationship between observed and true test scores in the classical test theory (CTT) framework can be used (Bulut 2013). In CTT, reliability is defined as the proportion of observed score variation that is attributable to true scores. That is, the ratio of true score variation to observed score variation leads to a reliability index ranging from 0 to 1 where higher values indicate higher reliability.

If the total observed score variation is defined as the sum of the variances for V subtests, then the observed total level variance becomes $B = V * \sigma_{\bar{X}_p}^2$, where B is the observed total level variance and V is the number of domains, subtests, variables, etc. Similarly, the true total level variance based on the true level value becomes $B_T = V * \sigma_{\bar{T}_p}^2$; where $\sigma_{\bar{T}_p}^2$ is the variance of true level scores and B_T is the total true level variance. Using the observed and true level variances, between-person reliability (ρ_B) can be defined as the ratio of true level variation to observed level variation:

$$\rho_B = \frac{B_T}{B}. \tag{23}$$

Within-person reliability can be defined in a similar fashion. In a test score profile, the total observed pattern variance is:

$$W = \sum_{v=1}^{V} \left[\frac{1}{P} \sum_{p=1}^{P} (X_{pv} - \bar{X}_p)^2 \right], \tag{24}$$

where W represents the total observed within-person variation due to individual differences in the subscores. Similarly, the total true pattern variance can be shown as:

$$W_T = \sum_{v=1}^{V} \left[\frac{1}{P} \sum_{p=1}^{P} (T_{pv} - \bar{T}_p)^2 \right], \tag{25}$$

where W_T is the total true within-person variation in the test score profile. By using the same approach with the ratio of observed and true scores, within-person reliability can be defined as the ratio of true pattern variation to observed pattern variation as follows:

$$\rho_B = \frac{W_T}{W}. \tag{26}$$

The within-person reliability coefficient can be interpreted as the proportion of variation in observed profile patterns that can be attributed to true pattern variation in the test score profile. Within-person reliability can also be interpreted as a weighted average of the within-person reliability for each subtest and as a weighted average of the person profile reliabilities.

As explained above, within-person and between-person reliability coefficients are based on the relationship between true and observed subscores in a test score profile. It is assumed in CTT that true scores are unknown and observed scores are the approximations of true scores. When there are parallel forms of a test, the covariance of the two forms provides an estimate of the true score variation. If true scores from two tests are perfectly correlated (i.e., congeneric) and equally reliable, then the correlation between the observed scores provides an estimate of the proportion of true score variance to observed score variance.

After obtaining an estimate of the covariance between every possible pair of parallel tests v and v', the true score variation becomes equal to the average of all possible covariances because the tests v and v' are assumed to have equal variances. Based on the test scores from parallel test forms, within-person and between-person reliability coefficients can be formulated as follows (Bulut 2013):

$$\hat{\rho}_W = \frac{\sum_{p=1}^{P} \left(\sum_{v=1}^{V} (X_{pv} - \bar{X}_p)(X_{pv'} - \bar{X}_{p'}) \right)}{\sum_{p=1}^{P} (\sum_{v=1}^{V} (X_{pv} - \bar{X}_p)^2)}, \text{and} \tag{27}$$

$$\hat{\rho}_B = \frac{\hat{\sigma}(\bar{X}_p \bar{X}_{p'})}{\sqrt{\hat{\sigma}^2(\bar{X}_p)\hat{\sigma}^2(\bar{X}_{p'})}} = \frac{\bar{\sigma}(\bar{X}_p \bar{X}_{p'})}{\hat{\sigma}^2(\bar{X}_p)}. \tag{28}$$

It should be noted that the within-person reliability coefficient is a weighted average of within-person reliability coefficients from all subtests. Without averaging over the subtests, within-person reliability coefficients can also be computed to evaluate reliability for each subtest (for more details, see Bulut (2013)).

3. Principal functions in the profileR package

The following is a brief description of the principal functions in the **profileR** package that implements the aforementioned analyses and tests.

3.1. Profile analysis for one sample using Hotelling's T^2

The `paos` function implements profile analysis for a single sample using Hotelling's T^2 test and tests the two hypotheses described earlier. A call to the `paos` function proceeds as follows:

```
paos(data, scale = TRUE)
```

where `data` represents a data frame with persons (rows) with multiple variables (columns). The argument of `scale` asks whether variables should be standardized (i.e., divided by their respective standard deviations). This is a necessary step if the variables are not on the same scale. The function returns a summary table that prints the results of the two hypothesis tests.

3.2. Testing parallel, coincidental, and level profiles by groups

The `pbg` function (i.e., profile by group) implements three tests that correspond to testing whether the profiles are parallel, coincidental, and level across two groups defined by the grouping variable. The `pbg` function accepts the following arguments:

```
pbg(x, y, original.names=FALSE, profile.plot=FALSE, ...)
```

where `x` is a data frame with multiple subscores and `y` is a grouping variable which can be either a character or numerical variable or a factor. The `original.names` asks whether the original column names in the data frame should be used. If `profile.plot = TRUE`, then a profile plot of scores for the groups is drawn.

The `pbg` function returns a summary of the means of the observed variables (e.g. subscores on the tests or domains) by the grouping variable, a correlation table among the variables by the grouping variable, and the results of F-tests for testing parallel, coincidental, and level profiles across two groups.

3.3. Criterion-related profile analysis

Criterion-related profile analysis in the **profileR** package is performed using the `cpa` function. The `cpa` function utilizes the same formula syntax as the familiar `lm` and `glm` functions but utilizes slightly different arguments. The full default call to `cpa` is:

```
cpa(formula, data, k = 100, na.action = "na.fail", family = "gaussian",
    weights = NULL)
```

where `formula` is an object of class "formula" (i.e., of the format `Y ~ X1 + X2`); `data` is a data frame, list, or matrix containing the variables; `k` is the scalar constant which must be greater than 0; `na.action` is how the missing data should be handled; `family` can be any family defined in the **stats** package; and finally `weights`, an optional vector of weights to be used in the regression.

The `cpa` function returns an object of S3 class `critpat` which has the following generic functions defined: `print`, `summary`, `plot`, and `anova`.

A cross-validation technique for criterion-related profile analysis, described in Davison and Davenport (2002), can be conducted using the `pcv` function which takes the following arguments:

```
pcv(formula, data, seed = NULL, na.action = "na.fail", family = "gaussian",
        weights = NULL)
```

where `formula` is an object of class "formula" ; `data` is a data frame, list, or matrix containing the variables; `na.action` is how the missing data should be handled; `family` can be any family defined in the **stats** package; and finally `weights`, an optional vector of weights to be used in the regression. In addition, a seed, for reproducibility, may be set using the `seed` argument.

3.4. Profile analysis via multidimensional scaling

The `pams` function implements profile analysis via multidimensional scaling as described by Davison, Davenport, and Bielinski (1995) and Davenport, Ding, and Davison (1995). The function computes similarity/dissimilarity indices based on Euclidean distances between the scores provided in the data, and then extracts dimensional coordinates for each score using multidimensional scaling. A weight matrix, level parameters, and fit measures are computed for each subject in the data. The `pams` function takes the following arguments:

pams(x,dim)

where `x` is a data frame in which rows represent individuals and columns represent variables, and `dim` is the desired number of dimensions to be extracted from the data. The `pams` function returns a matrix that provides prototypical profiles of dimensions extracted from the data and a weight matrix that includes the subject correspondence weights for all dimensions, level parameters, and the subject fit measure which is the proportion of variance in the subject's actual profiles accounted for by the prototypical profiles.

3.5. Profile reliability

The `pr` function uses subscores from two parallel test forms and computes profile reliability coefficients as described in Bulut (2013). A typical call to `pr` involves:

pr(form1,form2)

where `form1` and `form2` are matrices or data frames that include two or more subscores for each examinee. Rows represent individuals and columns represent subscores. Both forms should have the same individuals and subscores in identical order. Using the parallel test forms or multiple administration of the same test form, within-person and between-person reliability coefficients are computed. Within-person reliability is an indicator of variability between the subscores of examinees and between-person reliability is an indicator of the average subscore variation among all examinees.

4. Examples

4.1. Profile analysis of nutrient data with Hotelling's T^2

In this example, we analyze the `nutrient` data to test the two hypotheses that the ratio of the variables are equal to 1 and that they are all equivalent, where failure to reject the

former renders the second hypothesis unnecessary. The `nutrient` data come from a study of women's nutrition commissioned by the United States Department of Agriculture (USDA) in 1985. Nutrient intake was measured on a random sample of 737 women aged 25-50 years. Five nutrients were measured: calcium, iron, protein, vitamin A and vitamin C. The first six rows of the `nutrient` data are shown below.

```
> data(nutrient)
> head(nutrient)

  calcium    iron protein      a       c
1  522.29 10.188  42.561 349.13  54.141
2  343.32  4.113  67.793 266.99  24.839
3  858.26 13.741  59.933 667.90 155.455
4  575.98 13.245  42.215 792.23 224.688
5 1927.50 18.919 111.316 740.27  80.961
6  607.58  6.800  45.785 165.68  13.050
```

Using the `nutrient` data, we want to test whether the ratios of these variables are equal to 1 or each other in the sample. Because the variables in the data are not on the same scale, `scale=TRUE` should be used to get the variables standardized. The following output shows the results of two hypothesis tests:

```
> paos(nutrient, scale=TRUE)

Profile Analysis for One Sample with Hotelling's T-Square:


                                          T-Squared         F df1 df2 p-value
Ho: Ratios of the means over Mu0=1         1392.347 276.9559   5 732       0
Ho: All of the ratios are equal to each other 1278.073 318.2159   4 733       0
```

The results indicate that both hypotheses should be rejected for the nutrient data. The first test rejects the null hypothesis that the ratio of the mean intake over the recommended intake is equal to 1 each nutrients. The second test rejects the null hypothesis that the ratio of the mean intake over the recommended intake is the same for each nutrients (i.e., that the difference is 0).

4.2. Testing parallel, coincidental, and level profiles in the Spouse data

The `spouse` data set includes four rating scale items where 60 spouses (30 husbands and 30 wives) rate each other. The `pbg` function can be used to test whether the profiles of husbands and wives are parallel, coincidental, or level.

```
> #Read spouse data from Professor Richard Jonhson's website
> spouse <- read.table(file="http://www.stat.wisc.edu/~rich/JWMULT02dat/T6-14.DAT")
> head(spouse)
```

```
   V1 V2 V3 V4      V5
1  2  3  5  5  Husband
2  5  5  4  4  Husband
3  4  5  5  5  Husband
4  4  3  4  4  Husband
5  3  3  5  5  Husband
6  3  3  4  5  Husband
```

The first four columns are the items and the last column in the data is the group variable. We will run `pbg` on this data set and store it in an object called `mod`.

```
> mod <- pbg(spouse[,1:4],spouse[,5],labels=FALSE,profile.plot=TRUE)
> print(mod)

Data Summary:
     Husband      Wife
v1 3.900000 3.833333
v2 3.966667 4.100000
v3 4.333333 4.633333
v4 4.400000 4.533333
```

The `print` function displays the average scores on the items for the husband and wife groups. It can be seen that the means of the rating scale items are quite similar between the husbands and wives. The `summary` function prints the findings of the three hypotheses corresponding to whether the profiles are parallel, coincidental, or level. The output shows that assuming $\alpha = .05$, the first and third hypotheses were not rejected; however the second hypothesis was rejected. The results suggest that the profiles of husbands and wives are parallel and level, but not coincidental.

```
> summary(mod)

Call:
pbg(x = spouse[, 1:4], y = spouse[, 5], profile.plot = TRUE,
    labels = FALSE)

Hypothesis Tests:
                                   F df1 df2       p-value
Ho: Profiles are parallel    8.016171   3  56 0.0625594505
Ho: Profiles are coincidental 1.532770  1  58 0.2206853266
Ho: Profiles are level      24.820709   3  57 0.0001554491
```

The `pbg` function returns a profile plot of the husbands' and wives' profiles because `profile.plot = TRUE`. In Figure 3, the lines appear to be parallel across the four items although the direction of the mean difference in the first question is different from the other three questions.

In the following example, we create a pseudo data set where two groups (i.e., males and females) have different profiles on three continuous variables. For both groups, the variables have a multivariate normal distribution and the correlations among the variables are fixed to
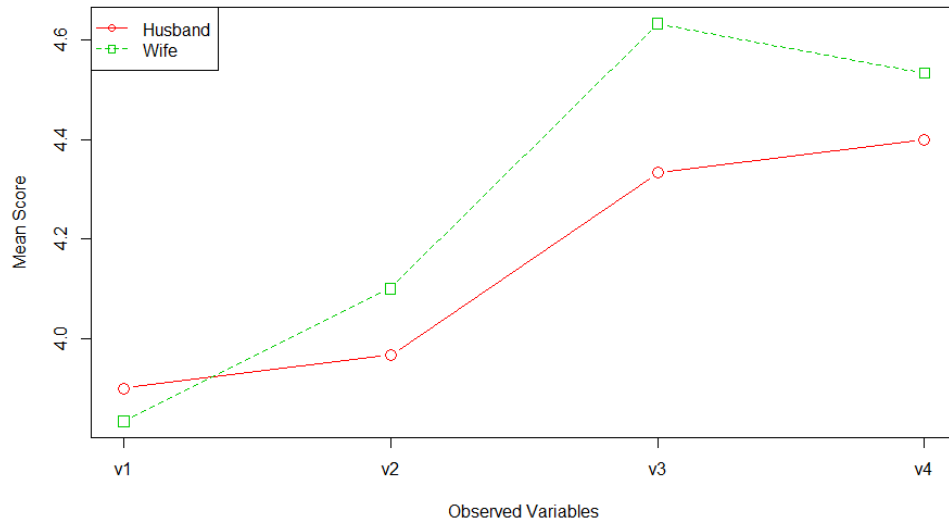
Figure 3: Profile plot of husbands and wives on the four rating scale items

$\rho = .3$. Although the variables have the same variance ($\sigma^2 = 1$) for both groups, the means of the variables for the male group are different from the female group. We use the `mvrnorm` function from the **MASS** package to simulate pseudo data for this example.

```
library(MASS)
Sigma <- matrix(c(1,0.3,0.3,
                  0.3,1,0.3,
                  0.3,0.3,1),3,3)

set.seed(2015)
male <- mvrnorm(n = 500, rep(0, 3), Sigma)
female <- mvrnorm(n = 500, c(0.1, 0.2, 0.3), Sigma)
data <- as.data.frame(rbind(male, female))
gender <- c(rep("Male",500), rep("Female",500))
data <- cbind(data,gender)
```

The resulting data set has 1000 rows, three columns of variables for the profiles, and an additional column of the group variable (i.e., gender). The first three rows of the simulated data set are:

```
> head(data,3)

          V1          V2          V3 gender
1  0.7956448   1.2150803   1.3751826   Male
2  0.6166696   0.9726083  -0.4316263   Male
3  1.7023255  -0.1173691   0.7960128   Male
```

If we use the `pbg` function to assess the profiles of the male and female groups, we obtain the following results:

```
> mod2 <- pbg(data[,1:3],data[,4],labels=FALSE,profile.plot=TRUE)
> print(mod2)

Data Summary:
        Female        Male
v1 0.05149379 -0.04385813
v2 0.18629128  0.07732983
v3 0.29538716  0.03379842

> summary(mod2)

Call:
pbg(x = data[, 1:3], y = data[, 4], profile.plot = TRUE, labels = FALSE)

Hypothesis Tests:
                                   F df1 df2      p-value
Ho: Profiles are parallel     6.680799   2 997 3.593899e-02
Ho: Profiles are coincidental 11.654924  1 998 6.660775e-04
Ho: Profiles are level        20.608530  2 998 3.757353e-05
```

The results show that all three null hypotheses were rejected in this example. Figure 4 shows the difference in the profiles of male and female groups based on the three variables and it is clearly visible that the profiles are indeed not parallel.
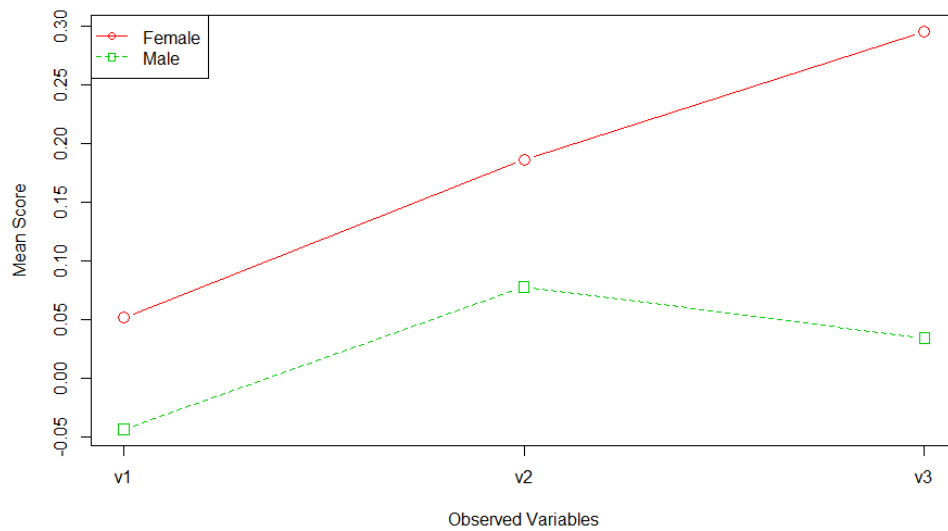

Figure 4: Profile plot of male and females from the pseudo data set

4.3. Criterion-related profile analysis and cross-validation using the IPMMc data set

A typical call to the `cpa` function using the data in Table 1, available in the IPMMc data set, would involve:

```
data(IPMMc)
mod <- cpa(R ~ A + H + S + B, data = IPMMc)
```

The `cpa` function will print the following by default:

```
print(mod)
```

```
> mod
Call:
cpa(formula = R ~ A + H + S + B, data = IPMMc)


Coefficients


Call:  glm(formula = formula, family = family, data = data, na.action = na.action)

Coefficients:
(Intercept)            A            H            S            B
   0.500000     0.009231     0.023077    -0.009231    -0.023077

Degrees of Freedom: 5 Total (i.e., Null);  1 Residual
Null Deviance:            1.5
Residual Deviance: 0.04615          AIC: -0.1779
```

This output should look familiar to users of `lm` and `glm`. This output corresponds to estimates of the parameters defined in Equation 16 and measures of model fit. Hypothesis testing can then proceed by a call to the `anova` function.

```
anova(mod)
```

```
> anova(mod)
Call:
cpa(formula = R ~ A + H + S + B, data = IPMMc)


Analysis of Variance Table


                df1 df2     F value     Pr(>F)
R2.full = 0       4   1   7.87500e+00 0.2604188
R2.pat = 0        3   1   1.05000e+01 0.2221903
R2.lvl = 0        1   1   0.00000e+00 1.0000000
R2.full = R2.lvl  3   1   1.05000e+01 0.2221903
R2.full = R2.pat  1   1  -7.21645e-15 1.0000000
```

This output corresponds to testing that the proportion of variability in the criterion variable explained by the full model is zero; that the proportion of variability in the criterion variable explained by the pattern effect is zero; that the proportion of variability in the criterion variable explained by the level effect is zero; that the proportion of variability in the criterion variable explained by by the full model is equal to the proportion explained by the level effect; and finally that the proportion of variability in the criterion variable explained by the full model is equal to the proportion explained by the pattern effect. These hypothesis tests are summarized in Table 2.

We see in the example that none of the hypothesis tests were rejected and we can conclude that the proportion of variability in the criterion variable that is explained by the full model is not significantly different than 0. Do note that the strange F-values and small degrees of freedom are a function of the sample size of the data set.

Table 2: Hypothesis tested reported from `anova(mod)`

R output	Null hypothesis test
R2.full = 0	$H_0 : R^2_{Full} = 0$
R2.pat = 0	$H_0 : R^2_{Pattern} = 0$
R2.lvl = 0	$H_0 : R^2_{Level} = 0$
R2.full = R2.lvl	$H_0 : R^2_{Full} = R^2_{Level}$
R2.full = R2.pat	$H_0 : R^2_{Full} = R^2_{Pattern}$

Finally, the proportion of variability in the criterion variable explained by the full model, pattern effect, and level effect can be extracted using the `summary` function. The `summary` function also provides the level and the pattern components.

summary(mod}

```
> summary(mod)
Call:
cpa(formula = R ~ A + H + S + B, data = IPMMc)

Relability
              R2
Full Model 0.969231
Pattern    0.969231
Level      0.000000

 Level Component
     1     2     3     4     5     6
58.75 58.75 55.00 58.75 58.75 55.00

 Pattern Component
       A     H     S     B
1  16.25  1.25  -8.75  -8.75
```

```
2   1.25  16.25 -13.75  -3.75
3   5.00   5.00   0.00 -10.00
4  -8.75  -8.75  16.25   1.25
5 -13.75  -3.75   1.25  16.25
6   0.00 -10.00   5.00   5.00
```

Other useful information from the criterion-related profile analysis is stored in the `critpat` object and can be easily extracted. The contents of this object can be examined by running `str(mod)`.

To demonstrate the cross-validation technique, we'll generate a data set sufficiently large enough to do cross-validation, that is very loosely based on the `IPMMc` data set.

```
set.seed(13251)
dv <- rbinom(n = 100, size = 1, prob = mean(IPMMc$R))
A <- rnorm(n = 100, mean = mean(IPMMc$A))
H <- rnorm(n = 100, mean = mean(IPMMc$H))
S <- rnorm(n = 100, mean = mean(IPMMc$S))
B <- rnorm(n = 100, mean = mean(IPMMc$B))
dat0 <- data.frame(dv, A, H, S, B)
```

To perform cross-validation, we call `pcv` and save the output in `cv.model` (which is also a `critpat` object):

```
cv.model <- pcv(dv ~ A + H + S + B, data = dat0)
```

As with `cpa`, `pcv` uses the same syntax as a call to `glm` does. Regression coefficients can be obtained through printing.

```
print(cv.model)

Call:
pcv(formula = dv ~ A + H + S + B, data = dat0)

Coefficients
$`Random Sample 1`
           A            H            S            B
 0.0514905676  0.0515755913 -0.0285983736  0.0005443284

$`Random Sample 2`
           A            H            S            B
-0.087656366 -0.013529506 -0.155730560 -0.007347574
```

It may be useful to set the seed when doing cross-validation and this can be set through the `seed` argument.

```
cv.model <- pcv(dv ~ A + H + S + B, data = dat0, seed = 4135)
```

The proportion of variability associated with each effect can be obtained using `summary` and hypothesis tests (described in Table 2) can be obtained using `anova` for each sample.

```
summary(cv.model)

Call:
pcv(formula = dv ~ A + H + S + B, data = dat0, seed = 4135)

Relability
$`R2.full = 0`
Random Sample 1 Random Sample 2
       0.015496         0.005036

$`R2.pat = 0`
Random Sample 1 Random Sample 2
       0.000088         0.001244

$`R2.lvl = 0`
Random Sample 1 Random Sample 2
       0.015489         0.004141

$`R2.full = R2.lvl`
Random Sample 1 Random Sample 2
       0.000088         0.001244

$`R2.full = R2.pat`
Random Sample 1 Random Sample 2
       0.015489         0.004141

anova(cv.model)

Call:
pcv(formula = dv ~ A + H + S + B, data = dat0, seed = 4135)

Analysis of Variance Table

$`R2.full = 0`
                     R2 df1 df2  F value    Pr(>F)
Random Sample 1 0.015496   1  47 0.739768 0.394102
Random Sample 2 0.005036   1  47 0.237895 0.627998

$`R2.pat = 0`
                     R2 df1 df2  F value    Pr(>F)
Random Sample 1 0.000088   1  48 0.004206 0.948561
Random Sample 2 0.001244   1  48 0.059785 0.807878

$`R2.lvl = 0`
```

```
                         R2 df1 df2  F value   Pr(>F)
Random Sample 1 0.015489   1  48 0.755175 0.389167
Random Sample 2 0.004141   1  48 0.199594 0.657058


$`R2.full = R2.lvl`
                      R2  F value df1 df2  Pr(>F)
Random Sample 1 0.000088 0.000327   1  48 0.98564
Random Sample 2 0.001244 0.043183   1  48 0.83626


$`R2.full = R2.pat`
                     R2 df1 df2  F value   Pr(>F)
Random Sample 1 0.015489   1  48 0.751236 0.390397
Random Sample 2 0.004141   1  48 0.182944 0.670768
```

As with `cpa`, additional information can be examined and extracted using `str(cv.model)`.

4.4. An example of profile analysis via multidimensional scaling

To demonstrate the use of the `pams` function, we will use a data set which contains score profiles of six respondents to a hypothetical personality scale. Each score profile consists of three scores from the neurotic, psychotic, and character disorder scales. The data include three types of profile patterns: Linearly increasing, inverted V, and linearly decreasing. To preview the `PS` data:

```
> data(PS)
> head(PS)
  Person Neu Psy CD
1      1  60  70 80
2      2  30  40 50
3      3  65  80 65
4      4  35  50 35
5      5  80  70 60
6      6  50  40 30
```

The first column in the `PS` data is a person ID variable, and the remaining columns are scores from the neurotic, psychotic, and character disorder scales. Assuming that there are two latent dimensions that can be extracted from the PS data, `pams` is called by:

```
> result <- pams(PS[,2:4], dim=2)
> print(result)
$weights.matrix
     weight1 weight2 level R.sq
[1,]     1.5 0.00000    70    1
[2,]     1.5 0.00000    40    1
[3,]     0.0 2.12132    70    1
[4,]     0.0 2.12132    40    1
```

```
[5,]     -1.5 0.00000    70    1
[6,]     -1.5 0.00000    40    1

$dimensional.configuration
     Dimension1 Dimension2
Neu  -6.666667  -2.357023
Psy   0.000000   4.714045
CD    6.666667  -2.357023
```

The first part of the output shows a weight matrix that includes the subject correspondence weights for all dimensions, level parameters, and the subject fit measure which is the proportion of variance in the subject's actual profiles accounted for by the prototypical profiles. The second part of the output shows a matrix that provides prototypical profiles of two latent dimensions extracted from the data.

4.5. Profile reliability using the EEGS data set

The `pr` function can be used to examine profile reliability of the subscores in the `EEGS` data set. The `EEGS` data set includes three subtests for examinees: quantitative 1 (Q1), quantitative 2 (Q2), and verbal (V). For each subtest, there are two subscores that come from parallel forms. To activate the EEGS data set and preview the first six rows of the data.

```
> data(EEGS)
> head(EEGS)
     Form1_Q1 Form2_Q1 Form1_Q2 Form2_Q2 Form1_V Form2_V
[1,]        2        0        2        0       0       0
[2,]        4        9        0        0       3       4
[3,]        4        3        8        6      27      27
[4,]        2        6        0        0      26      29
[5,]        7        4        3        2       8       6
[6,]       18       16        1        3      14      15
```

The returning output shows two columns for each subtest. These columns represent subscores from parallel forms from each subtest. To compute between-person and within-person subscore reliability of the three subscores in EEGS, the `pr` function can be called as:

```
result <- pr(EEGS[,c(1,3,5)],EEGS[,c(2,4,6)])
```

The `print` function returns the following output:

```
> print(result)
Subscore Reliability Estimates:

        Estimate
Level   0.9245548
Pattern 0.9338338
Overall 0.9308374
```

In the output, level refers to between-person reliability, pattern refers to within-person reliability, and overall refers to the total profile reliability which is the weighted average of between-person and within-person reliability coefficients. In this example, the three subtests of EEGS indicate high levels of between-person and within-person reliability. `plot(result)` returns a scatter plot of the level values from the two parallel forms each with three subtests.

5. Discussion

The **profileR** package has been developed to provide researchers and partitioners interested in profile analysis with an intuitive syntax and framework for the R programming language. In addition to the functions described above, functions have been written use a within-person factor model to derive a score profile (Davison *et al.* 2009) (using the engine from the **lavaan** package (Rosseel 2012)) and experimental support for moderated profile analysis has recently been added (based on unpublished work by Davison and Davenport). In future versions of **profileR**, we intend to expand the package to allow for multi-level profile analysis (i.e., creating score profiles for test takers and schools while controlling for the inherent nesting of students within classrooms and testing these effects seperately), Bayesian profile analysis, and the addition of increased graphical capabilities using the **DiagrammeR** package (Sveidqvist, Bostock, Pettitt, Daines, Kashcha, and Iannone 2015). We welcome all contributions to the **profileR** package and would like to encourage the community to open feature requests on the GitHub website to help steer the direction of **profileR**.

References

Bulut O (2013). *Between-person and Within-person Subscore Reliability: Comparison of Unidimensional and Multidimensional IRT Models*. Ph.D. thesis, University of Minnesota.

Bulut O, Desjardins CD (2015). *profileR: Profile Analysis of Multivariate Data in R*. R package version 0.3.

Conger AJ, Lipshitz R (1973). "Measures of reliability for profiles and test batteries." *Psychometrika*, **38**, 411–427.

Cronbach LJ (1951). "Coefficient alpha and the internal structure of tests." *Psychometrika*, **16**(3), 297–334.

Cronbach LJ, Gleser G (1953). "Assessing similarity between profiles." *Psychological Bulletin*, (50), 456–473.

da Silva AR (2015). *biotools: Tools for Biometry and Applied Statistics in Agricultural Science*. R package version 2.1, URL http://CRAN.R-project.org/package=biotools.

Davenport EC, Ding S, Davison ML (1995). *PAMS: SAS Template*. Unpublished report, Department of Educational Psychology, University of Minnesota.

Davison ML (1994). "Multidimensional scaling models of personality responding." In S Strack, M Lorr (eds.), *Differentiating normal and abnormal personality*. Springer, New York.

Davison ML (1996). *Addendum to "Multidimensional scaling and factor models of test and item responses.".* Unpublished report, Department of Educational Psychology, University of Minnesota.

Davison ML, Davenport EC (2002). "Identifying criterion-related patterns of predictor scores using multiple regression." *Psychological Methods*, **7**(4), 468–484.

Davison ML, Davenport EC, Bielinski J (1995). *PAMS: SAS Template.* Unpublished report, Department of Educational Psychology, University of Minnesota.

Davison ML, Gasser M, Ding S (1996). "Identifying major profile patterns in a population: An exploratory study of WAIS and GATB patterns." *Psychological Assessment*, (8), 26–31.

Davison ML, Kim SK, Close C (2009). "Factor analytic modeling of within person variation in score profiles." *Multivariate Behavioral Research*, (44), 668–687.

Davison ML, Kuang H (2000). "Profile patterns: Research and professional interpretation." *School Psychology Quarterly*, (15), 457–464.

Davison ML, Kuang H, Kim S (1999). "The structure of ability profile patterns: A multidimensional scaling perspective on the structure of intellect." In P Ackerman, P Kyllonen, R Roberts (eds.), *Learning and individual differences: Process, trait, and content determinants*, pp. 187–207. American Psychological Association, Washington, DC.

Ding CS (2001). "Profile analysis: multidimensional scaling approach." *Practical Assessment, Research & Evaluation*, **7**(16). Retrieved August 31, 2015, URL http://PAREonline.net/getvn.asp?v=7&n=16.

Filzmoser P, Gschwandtner M (2015). *mvoutlier: Multivariate outlier detection based on robust methods.* R package version 2.0.6, URL http://CRAN.R-project.org/package=mvoutlier.

French A, Poulsen J, Yu A (2015). *Multivariate analysis of variance (MANOVA).* Retrieved August 12, 2015, URL http://userwww.sfsu.edu/~efc/classes/biol710/manova/manovanew.html.

Jarek S (2012). *mvnormtest: Normality test for multivariate variables.* R package version 0.1-9, URL http://CRAN.R-project.org/package=mvnormtest.

Kaufman L, Rousseeuw PJ (2008). *Introduction*, pp. 1–67. John Wiley and Sons, Inc. ISBN 9780470316801. doi:10.1002/9780470316801.ch1. URL http://dx.doi.org/10.1002/9780470316801.ch1.

Korkmaz S, Goksuluk D, Zararsiz G (2014). "MVN: An R Package for Assessing Multivariate Normality." *The R Journal*, **6**(2), 151–162. URL http://journal.r-project.org/archive/2014-2/korkmaz-goksuluk-zararsiz.pdf.

Kuder G, Richardson M (1937). "The theory of the estimation of test reliability." *Psychometrika*, **2**(3), 151–160. ISSN 0033-3123. doi:10.1007/BF02288391. URL http://dx.doi.org/10.1007/BF02288391.

Rosseel Y (2012). "lavaan: An R Package for Structural Equation Modeling." *Journal of Statistical Software*, **48**(2), 1–36. URL http://www.jstatsoft.org/v48/i02/.

Stanton HC, Reynolds CR (2000). "Configural frequency analysis as a method of determining Wechsler profile types." *School Psychology Quarterly*, (15), 434–448.

Sveidqvist K, Bostock M, Pettitt C, Daines M, Kashcha A, Iannone R (2015). *DiagrammeR: Create Graph Diagrams and Flowcharts Using R.* R package version 0.8, URL https://github.com/rich-iannone/DiagrammeR.

Therneau T, Atkinson B, Ripley B (2015). *rpart: Recursive Partitioning and Regression Trees.* R package version 4.1-10, URL http://CRAN.R-project.org/package=rpart.

Watkins M, Glutting J, Youngstrom E (2005). "Issues in subtest profile analysis." In D Flanagan, P Harrison (eds.), *Contemporary intellectual assessment: Theories, tests, and issues*, 2nd edition, pp. 251–268. Guilford Press, New York.

Affiliation:

Christopher David Desjardins
Science Institute, Taeknigardur
University of Iceland
Dunhaga 5
107 Reykjavik, Iceland
E-mail: cddesjardins@gmail.com

Okan Bulut
University of Alberta
6-110A Education Centre North
11210 87 Ave NW
Edmonton, AB T6G 2G5 Canada
E-mail: bulut@ualberta.ca
URL: https://www.ualberta.ca/~bulut/