Profile Analysis of Multivariate Data: A Brief Introduction to the `profileR` Package

Okan Bulut[*1] and Christopher David Desjardins[†2]

[1]University of Alberta
[2]St. Michael's College

Abstract

Profile analysis is a multivariate statistical technique, which is the equivalent of multivariate analysis of variance (MANOVA) for repeated measures. This technique is widely used by researchers in education, psychology, and medicine for the non-orthogonal decomposition of observed scores into level and pattern effects. A suite of procedures for decomposing observed scores into level and pattern effects and statistical techniques utilizing these effects exists for the R programming language in the **profileR** package (Bulut & Desjardins, 2018). 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 showcases several of these methods, illustrating their applications with various data sets included with the package. The `profileR` package is geared towards researchers in the social sciences and medicine, with limited familiarity with R, and aims to lower the entry to using these methods for this audience.

1 Introduction

A profile of test scores may arise from a single test consisting of multiple subtests, repeated assessments in a classroom, parallel or alternate forms of a test, or a psychological instrument(s) that measures multiple latent domains. For example, a test taker's scores on a cognitive test battery (e.g., Visual Processing, Quantitative Knowledge, Fluid Reasoning, and Short-Term Memory) can be considered a profile. Another profile might be test scores on the SAT, a required exam for college admissions in the United States. Each test taker receives three subscores for the math section of the SAT: Heart of Algebra, Problem Solving and Data Analysis, and Passport to Advanced Math. These subscores can be used to create a profile of math skills. A profile can consist of not only of an individual's test scores but can also be aggregate or average test scores for a sample or for a group. For example, profiles of male and female students on the SAT can demonstrate potential college readiness differences between the two groups.

The suite of techniques known as profile analysis exists to help researchers and practitioners identify, quantify, and interpret the extent to which individual test takers or groups of test takers show a distinct profile(s). Profile analysis is utilized for both inter- and intra-individual interpretations of test scores and allows the quantification of the degree of similarity between observed and expected test score profiles. Specifically, it can be used to quantify the amount of variability associated with the level and pattern effects. To examine the similarities/differences among test scores, profile analysis focuses on the decomposition, analysis, and quantification of the elevation, variation, and configuration of test scores (Cronbach & Gleser, 1953; Stanton & Reynolds, 2000). Throughout this manuscript, we will use *test scores* and *variables* interchangeably depending on the context when describing profiles.

---

[*]bulut@ualberta.ca (Corresponding author)
[†]cdesjardins@smcvt.edu

Several statistical frameworks either include or resemble profile analysis. The most common form of profile analysis is repeated measures MANOVA where researchers investigate whether multiple measurements of a set of variables from a single sample change over time. A more complex form of profile analysis relies on a combination of multidimensional scaling and factor analysis (M. Davison & Kuang, 2000; M. Davison, Kuang, & Kim, 1999; M. L. Davison, Gasser, & Ding, 1996). Cluster analysis is a related but distinct technique. Cluster analysis involves the classification of individuals into groups based on the similarity of their profile. In contrast, the methods in the **profileR** package focus on testing and understanding the different components of an individual's or group's profile and not on classification.

The **profileR** package implements the profile analytic methods described, and developed, in Bulut (2013), Bulut, Davison, and Rodriguez (2017), M. L. Davison (1994), M. L. Davison and Davenport (2002), and M. L. Davison, Kim, and Close (2009) as well as various unpublished techniques. The suite of functions available in the **profileR** package have been selected as they are the most commonly used techniques in the context of profile analysis. In addition, we have implemented new methods from this field into **profileR** that should of interest to applied researchers working with profiles. Our package aims to be a one-stop shop for researchers utilizing profile analysis by consolidating these methods into a single R package and providing a consistent and unified framework.

2   Principal methods available in profileR

2.1   Profile plots

A profile analysis typically begins with a graphical examination of the relative behavior of all the test scores or variables in a profile. A profile plot can be created by plotting the sample means for each variable by individuals, for a single group, or across multiple groups. To have a clear and meaningful graphical summary of the variables in a profile, all of the variables must have the same units of measurement. For instance, if test scores from a mathematics test have a range of 0 to 100 and test scores from a reading test have a range of 0 to 50, then test scores from both tests must be rescaled to have the same range or transformed into a z-score. In a typical profile plot, the average scores for each group or individual scores for each person are plotted against the measured domains. Below we present the code to create the profile plot shown in Figure 1.

```
data <- data.frame(Domain = factor(rep(c("Reading", "Math", "Science", "Social"),  2),
                                    levels = c("Reading", "Math", "Science", "Social")),
                 Student = factor(c(1, 1, 1, 1, 2, 2, 2, 2)),
                 Score = c(35, 50, 40, 50, 70, 80, 70, 85))

ggplot(data, aes(x = Domain, y = Score, group = Student, shape = Student)) +
  geom_line() + geom_point(size = 4, fill = "white") +
  theme(panel.background = element_rect(fill = 'white', colour = 'black'),
        panel.grid.minor = element_blank(), panel.grid.major = element_blank(),
        axis.text = element_text(size = 14),
        axis.title = element_text(size = 15, face = "bold")) +
  scale_y_continuous(name = "Score", limits = c(30, 90),breaks = seq(30, 90, 10)) +
  geom_text(data = data, aes(x = Domain, y = Score, label = Score),
            size = 5, vjust = -1.5)
```

Figure 1 presents 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 to assess whether the profiles are parallel (i.e., whether the scores are similar to each other). In Figure 1, the lines appear to be parallel across the four content domains. The `profileplot` function in **profileR** allows users to quickly create a profile plot and additional functions are available in **profileR** that can be used to formally test whether the profiles are parallel. These procedures will be described in more detail in the following
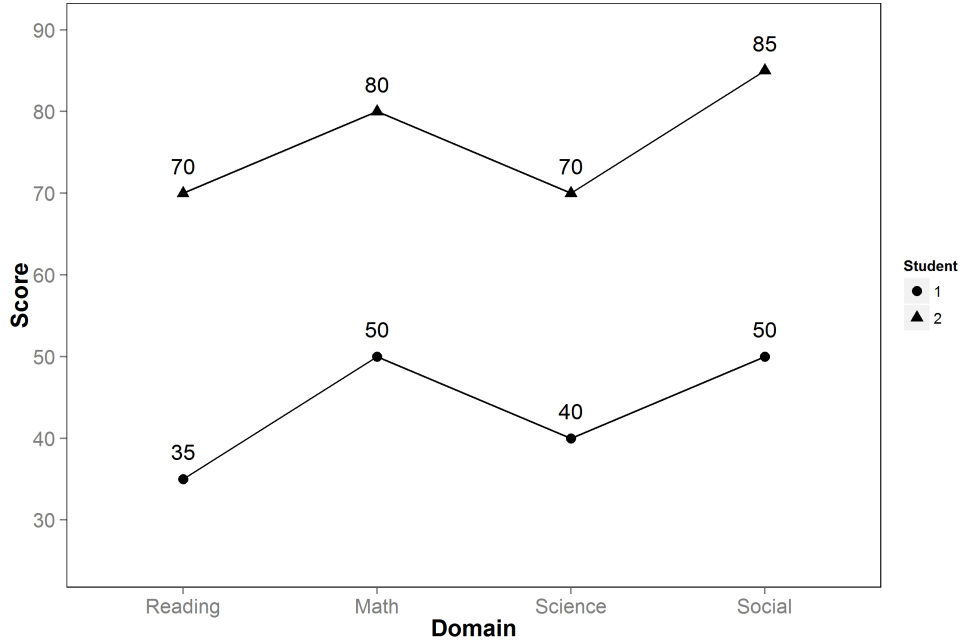
sections.

Figure 1: A profile plot with two persons and four content domains.

2.2 Profile analysis for one sample

Often a researcher is interested in testing the equivalence of the means of the variables obtained from a sample of individuals. For instance, a researcher can investigate whether students' scores on a mathematics assessment based on subdomains (e.g., algebra, geometry, and arithmetics) are similar. If students perform differently across the subdomains, the researcher can test whether this variability is significant. To test the equivalence of the means of the variables, separate univariate t-tests could be used for comparing every possible pair of variables. Alternatively, one can conduct an overall test of difference between the variables that minimizes the chance of committing a type I error and adjusts for the correlations among the variables.

Profile analysis for one sample can be performed using Hotelling's T^2 which tests the null hypothesis that the univariate means of the variables in a profile are equivalent. For this test, hypothesis testing proceeds as follows:

$$H_0 : \boldsymbol{\mu} = \boldsymbol{\mu_0} \text{ against } H_a : \boldsymbol{\mu} \neq \boldsymbol{\mu_0} \tag{1}$$

where $\boldsymbol{\mu}$ is the vector of observed means from p variables, $(\boldsymbol{\mu} = [\mu_1, \mu_2, \ldots, \mu_p])$ and $\boldsymbol{\mu_0}$ is the vector of hypothesized means for p variables $(\boldsymbol{\mu_0} = [\mu_1^0, \mu_2^0, \ldots, \mu_p^0])$. Equation 1 is equivalent to testing the null hypothesis that all observed means are equal to their hypothesized means:

$$H_0 : \frac{\mu_1}{\mu_1^0} = \frac{\mu_2}{\mu_2^0} = \cdots = \frac{\mu_p}{\mu_p^0} = 1 \tag{2}$$

against the alternative hypothesis that at least one of the observed means is different from its hypothesized mean:

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

Furthermore, instead of testing whether the ratios of the observed means over their hypothesized means are equivalent to one, profile analysis can also 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 is often tested:

$$\text{H}_0 : \frac{\mu_1}{\mu_1^0} = \frac{\mu_2}{\mu_2^0} = \cdots = \frac{\mu_p}{\mu_p^0}. \tag{4}$$

A profile analysis for one sample, as outlined above, can be easily carried out with the `paos` function in **profileR**.

2.3 Assessing parallelism, equality, and flatness in a profile across group or time

Data collection involving multiple variables is not typically collected on a single group or at one time point but on multiple groups or multiple occasions. In such situations, there is often interest in assessing various features of a profile across the groups or time points:

1. Are the profiles parallel across groups or time points?

2. Do the profiles have equal levels across the groups or time points?

3. Are the profiles flat across the groups or time points?

Parallelism is the main test of interest in a profile analysis by group because the test of parallelism examines whether each segment of a profile is identical. A segment refers to the difference in the values of the same variables across multiple time points or the difference between multiple variables in a single time point. Therefore, the segment is simply the slope of the line between the means of two adjacent variables. Parallelism is assessed using a one-way MANOVA. If the null hypothesis of parallelism is rejected, there is a significant interaction between group membership and the variables or group membership and the time points (e.g., if a test is repeatedly administered). In other words, the amount of increase or decrease between successive measurements of the variable is different for at least one of the groups.

Given that the profiles are parallel one typically tests equality of the levels, which examines whether the profiles coincide (i.e., there are no group differences). This test is used for determining whether at least one group scored higher than other groups, on average, across all the variables 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, the resulting procedure becomes a univariate test and a between-groups main effect in ANOVA is performed. The test simply measures the relative contributions of between-group and within-group variations to the total sum of squared errors. Based on this test, if the group levels are significantly different from one another, then the null hypothesis of equal levels is rejected. That is, at least one group performs significantly higher or lower than the other groups based on the average of p variables.

Flatness is a measure of the extent to which the profiles are flat within any group (i.e., there are no differences in the average values of the variables or the average value of a single variable measured across multiple time points), given that the profiles are parallel. The null hypothesis of flatness is that the segments are 0. This is the analog to the profile analysis for one sample except for multiple groups or repeated measurements. Like profile analysis for one sample, a 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:

$$T^2 = N(\bar{\mathbf{X}}_{\mathbf{tot}})^\top (\mathbf{S}_{\mathbf{wg}})^{-1} (\bar{\mathbf{X}}_{\mathbf{tot}}) \tag{5}$$

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:

$$\lambda = \frac{1}{1 + T^2} \tag{6}$$

It should be noted that if parallelism is rejected, the other two tests (equal levels and flatness) are not necessary. In that case, flatness may be assessed within each group, and various within- and between-group contrasts may be analyzed. The `pbg` function (i.e., profile by group) implements three tests that correspond to testing whether the profiles are parallel, have equal levels, and are flat across two groups or time points. An example using `pbg` is presented below.

2.4 Profile analysis via multidimensional scaling (PAMS)

An exploratory latent variable technique to identify patterns in profiles is profile analysis via multidimensional scaling (PAMS) (M. L. Davison, 1994). This approach examines the major profile patterns by estimating latent profiles, unlike the previous methods which rely solely on observed scores. The PAMS model can be more advantageous than other profile analytic methods because:

1. It can be used with any sample size.

2. Unlike the common factor model, 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.

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 model, latent person profiles and person profile indices corresponding to the latent person 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. An 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:

$$m_{ij} = \sum_{k=1}^{K} w_{ik} x_{jk} + c_i + e_{ij}, \tag{7}$$

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 the 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 (i.e., the average latent score). 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 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 7, PAMS requires several assumptions and restrictions (M. L. 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 a 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 (M. L. Davison, 1996; Ding, 2001).

The `pams` function, available in the **profileR** package, 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.

2.5 Criterion-related profile analysis

When performing a profile analysis, there may be explicit interest in an external criterion or dependent variable and understanding the proportion of variability in the criterion that can be explained by either the level or the pattern effect. This type of profile analysis is referred to as criterion-related profile analysis (M. L. Davison & Davenport, 2002) and is implemented in **profileR** with the `cpa` function.

Consider the `interest` data set from a hypothetical interest inventory. This data set contains scores on cognitive and personality measures and interest in various professions. Imagine we are interested in predicting someone's interest in being doctor given information about their cognitive abilities, specifically their mathematical and verbal abilities. We might be interested in understanding if it is just having high cognitive ability or a certain pattern that piques someone's interest in being a doctor. We can answer this question using a subset of the `interest` data set, which we will save as `doctor.interest`. The first few rows of the `doctor.interest` data set are presented in Table 1 and can be created using the following code:

```
doctor.interest <- interest[, c(4:9, 28)]
head(doctor.interest)
```

Table 1: Hypothetical interest inventory. (*Note*: vocab = vocabulary test, reading = reading comprehension, sentcomp = sentence completion, mathmtcs = mathematics, geometry = geometry, analyrea = analytical reasoning, and doctor corresponds to interest in the doctor profession.)

vocab	reading	sentcomp	mathmtcs	geometry	analyrea	doctor
1.67	1.67	1.46	0.90	0.49	1.65	1.95
-1.19	-1.43	-1.19	-0.17	0.28	0.23	-0.50
1.56	1.11	2.25	1.40	1.85	2.04	0.69
-0.59	-0.03	0.12	0.05	-0.48	-0.17	-0.07
0.15	0.98	-0.05	0.20	-0.68	0.58	0.35
1.21	1.04	0.75	-0.01	-0.05	-0.37	1.04

If we wanted to predict individual p's, interest in being a doctor, we may consider the following regression model:

$$C_p = \beta_0 + \beta_1 X_{Vp} + \beta_2 X_{Rp} + \beta_3 X_{Sp} + \beta_4 X_{Mp} + \beta_5 X_{Gp} + \beta_4 X_{Ap} \tag{8}$$

where the βs correspond to the intercept and slopes associated with our measures of cognition and $X_V \dots X_A$ correspond to the cognitive variables shown in Table 1 (i.e., vocab, $\dots$, analyrea). We can re-express this equation in terms of the *pattern* and *level* effects, as follows:

$$C_p = \beta_0 + \sum_{v=1}^{V} (\beta_v - \bar{\beta})(X_{vp} - X_{\bar{p}}) + \sum_{v=1}^{V} \bar{\beta} X_{\bar{p}} \tag{9}$$

where V corresponds to the number of cognitive measures (e.g., six in this example). The first sum corresponds to the pattern of a test score profile, this vector is defined as the difference between person p's score on a test (X_{vp} for vp = vocab, $\dots$, analyrea) and their overall mean across the tests ($X_{\bar{p}} = \frac{1}{v} \sum_{v=1}^{V} X_{vp}$). The pattern effect is multiplied by the difference between the estimated parameter associated with a particular domain, β_v, and the mean of these parameter estimates across the domains, $\bar{\beta}$. The second summation term is strictly a function of person p's level, $X_{\bar{p}}$. Further, M. L. Davison and Davenport (2002) show that this equation can be re-expressed as:

$$C_p = \beta_0 + (V/k)Cov_{pc} + V\bar{\beta} X_{\bar{p}} \tag{10}$$

where $Cov_{pc} = (1/V) \sum_{v=1}^{V} (X_{vp} - X_{\bar{p}})(k\beta_v - k\bar{\beta})$ is the covariance between the pattern effect and the scores in the criterion-pattern vector, $\mathbf{x}_c = k(\beta_v - \bar{\beta})$, where k is some scalar greater than 0 and the other variables were described above.

Rewriting the multiple regression of the criterion variable onto the cognitive measures 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 in the example below.

Criterion-related profile analysis in the **profileR** package is performed using the `cpa` function.

2.6   Profile reliability

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

Examining an individual's strengths and weaknesses from pattern scores has brought some uncertainties due to lack of evidence for reliability and validity of test scores (Watkins, Glutting, & Youngstrom, 2005). Profile reliability is conceived of as a multivariate analog to the traditional definition of univariate reliability. 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 test scores. This approach is an extension of canonical test reliability originally proposed by Conger and Lipshitz (1973).

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

$$\rho_p = \frac{(\mathbf{T}_i - \bar{\mathbf{T}}_i)^\top \mathbf{A}(\mathbf{T}_i - \bar{\mathbf{T}}_i)}{(\mathbf{X}_i - \bar{\mathbf{X}}_i)^\top \mathbf{A}(\mathbf{X}_i - \bar{\mathbf{X}}_i)}, \tag{11}$$

where $(\mathbf{T}_i - \bar{\mathbf{T}}_i)$ 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 & 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. The total score variation (T) of a test score profile can be expressed as the sum of the variances associated with these effects for V tests:

$$T = B + W, \tag{12}$$

where the total variation (T) is the sum of two orthogonal components: between-person variation (B) and within-person variation (W) (M. L. Davison et al., 2009). 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; M. L. 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 $B = V * \sigma^2_{\bar{X}_p}$, 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 $B_T = V * \sigma^2_{\bar{T}_p}$; where $\sigma^2_{\bar{T}_p}$ 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:

$$\rho_B = \frac{B_T}{B}. \tag{13}$$

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

$$W = \sum_{v=1}^{V} \left[\frac{1}{P} \sum_{p=1}^{P} (X_{vp} - \bar{X}_p)^2 \right], \tag{14}$$

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:

$$W_T = \sum_{v=1}^{V} \left[\frac{1}{P} \sum_{p=1}^{P} (T_{vp} - \bar{T}_p)^2 \right], \tag{15}$$

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:

$$\rho_B = \frac{W_T}{W}. \tag{16}$$

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 scores 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 computed as follows:

$$\hat{\rho}_W = \frac{\sum_{p=1}^{P}\left(\sum_{v=1}^{V}(X_{vp}-\bar{X}_p)(X_{vp'}-\bar{X}_{p'})\right)}{\sum_{p=1}^{P}(\sum_{v=1}^{V}(X_{vp}-\bar{X}_p)^2}, \tag{17}$$

and

$$\hat{\rho}_B = \frac{\hat{\sigma}(\bar{X}_p\bar{X}_{p'})}{\sqrt{\hat{\sigma}^2(\bar{X}_p)\hat{\sigma}^2(\bar{X}_{p'})}} = \frac{\bar{\sigma}(\bar{X}_p\bar{X}_{p'})}{\hat{\sigma}^2(\bar{X}_p)}. \tag{18}$$

Please note that the $'$ refers to a parallel form of a test in Equations 17 and 18, not the transpose of a matrix. 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)).

The `pr` function uses subscores from two parallel test forms and computes profile reliability coefficients.

3 Examples

A typical use of **profileR** to conduct a profile analysis may include using one or some of the aforementioned functions to quantify variability associated with a particular effect, to test a profile hypothesis, to create a profile plot, or to extract information for further analyses. The demonstration of the following functions are intended to help researchers understand how they can use **profileR** to conduct a particular technique within the profile analytic framework.

3.1 Profile analysis of `nutrient` data with Hotelling's T^2

Although it is not a testing-related data set, we begin by looking at the `nutrient` data set to demonstrate the use of a one-sample profile analysis. This example shows an application of the **profileR** package outside of a testing framework. The `nutrient` data set comes 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. In this example, we analyze the `nutrient` data set to test the two hypotheses: 1) that the ratios of the five nutrients were equal to 1 and 2) to test if ratios of the nutrients were equivalent. In other words, did the 737 women, on average, have the same amount of calcium, iron, protein, vitamin A and vitamin C intakes.

Because the variables in the data are not on the same scale, when we call the `paos` function we will pass the argument `scale=TRUE` to standardize the variables. The following output shows the results of two hypothesis test:

```
> 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 for each nutrient, while the second test rejects the null hypothesis that the ratio of the mean intake over the recommended intake is the same for each nutrient (i.e., that the difference is 0). Note that failing to reject the first null hypothesis renders the second hypothesis test meaningless. We conclude that the average intake for at least one of these five nutrients is different from the rest. In order to know which nutrient(s) differ, post hoc comparisons of the variables can be conducted.

3.2 Testing parallelism, equal levels, and flatness in the `spouse` data

A researcher might consider profile analysis by group if he or she wants to understand the interaction between a grouping factor and a set of test scores from multiple tests in a single time point or tests scores from the same test administered repeatedly. Once parallelism (i.e., group and test score interaction) is rejected, the other two tests (i.e., equal levels and flatness) may help the researchers understand the reason behind group differences.

In this example, we use the `spouse` data set that contains four rating scale items where 60 spouses (30 husbands and 30 wives) rated each other. The `pbg` function is used to test whether the profiles of husbands and wives are parallel, have equal levels, and are flat.

```
> mod <- pbg(spouse[,1:4], spouse[,5], original.names = FALSE, profile.plot = TRUE)
> 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 average scores on the four items for the husbands and wives are printed by default and because we set `profile.plot = TRUE` then a profile plot is created (see Figure 2). In Figure 2, the lines appear to be largely parallel across the four items although the direction of the mean difference in the first question is different from the other three questions (i.e., husbands provided higher rates than wives). Overall, the means of the four 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, with equal levels, or flat. The output shows that assuming $\alpha = .05$, the first and second hypotheses were not rejected; however the third hypothesis was rejected. The results suggest that the profiles of husbands and wives are parallel and have equal levels, but are not flat.
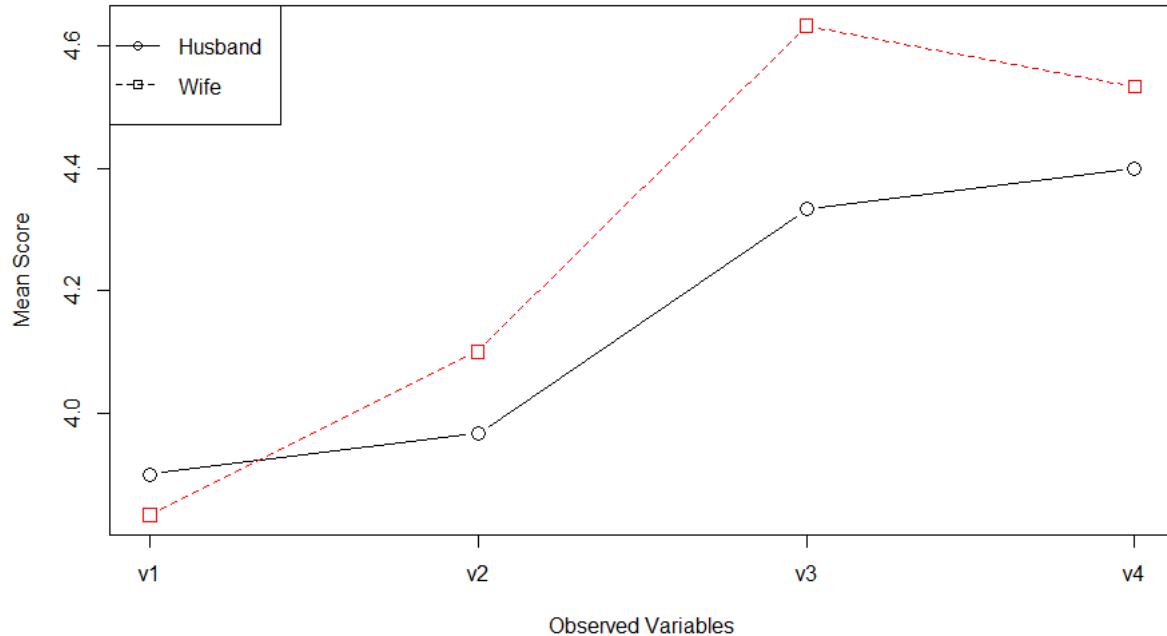
10

Figure 2: Profile plot of the `spouse` dataset.

```
> summary(mod)
Call:
pbg(data = spouse[, 1:4], group = spouse[, 5], original.names = FALSE,
    profile.plot = TRUE)

Hypothesis Tests:
$`Ho: Profiles are parallel`
  Multivariate.Test Statistic Approx.F num.df den.df    p.value
1            Wilks 0.8785726 2.579917      3     56 0.06255945
2           Pillai 0.1214274 2.579917      3     56 0.06255945
3  Hotelling-Lawley 0.1382099 2.579917      3     56 0.06255945
4              Roy 0.1382099 2.579917      3     56 0.06255945

$`Ho: Profiles have equal levels`
          Df Sum Sq Mean Sq F value Pr(>F)
group      1  0.234  0.2344   1.533  0.221
Residuals 58  8.869  0.1529

$`Ho: Profiles are flat`
        F df1 df2      p-value
1 24.82071   3  57 0.0001554491
```

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

The `doctor.interest` dataset that we created above will be used to demonstrate the use of the `cpa` function. The `cpa` function calls the `glm` function and returns information about the level and pattern effects. For the `doctor.interest` data, we want to understand whether the level and pattern of scores on a cognitive assessment can predict someone's interest in being a doctor. A typical call to `cpa` looks like:

11

```
> mod <- cpa(doctor ~ ., data = doctor.interest)
```

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

```
> mod
Call:
cpa(formula = doctor ~ ., data = doctor.interest)


Coefficients

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

Coefficients:
(Intercept)         vocab       reading      sentcomp      mathmtcs      geometry      analyrea
   -0.04789       0.08834       0.05677       0.21794       0.18475      -0.14591      -0.03267

Degrees of Freedom: 249 Total (i.e. Null);   243 Residual
Null Deviance:      276
Residual Deviance: 241.4  AIC: 716.7
```

This output is similar to output of the `glm` function, and corresponds to the estimates of the parameters defined in Equation 8. The proportion of variability associated with the model, the profile, and level can be extracted from the `mod` object.

```
> mod$r2
                 R2
Full Model 0.125320
Pattern    0.030635
Level      0.096605
```

To test the proportion of variability in the criterion variable explained by the various components of the model, a simple call with the `anova` function can be used. The `anova` function gives the following output that shows a summary of the statistical tests:

```
> anova(mod)
Call:
cpa(formula = doctor ~ ., data = doctor.interest)


Analysis of Variance Table


                 df1 df2   F value        Pr(>F)
R2.full = 0        6 243   5.802624 1.145210e-05
R2.pat = 0         5 243   1.535914 1.792564e-01
R2.lvl = 0         1 243  25.985292 6.919813e-07
R2.full = R2.lvl   5 243   1.595478 1.619787e-01
R2.full = R2.pat   1 243  26.304860 5.959596e-07
```

We see in this example that three of the four hypothesis tests were rejected. We reject the null hypothesis that the proportion of variability in interest in begin a doctor explained by our full model is zero (`R2.full = 0`); we fail to reject that the proportion of variability in interest in the doctor profession explained by the pattern effect is significantly different than zero (`R2.pat = 0`); we reject that the level effect explains no variability in someone's interest in the doctor profession (`R2.lvl = 0`); we fail to reject that the proportion of variability explained by the model is equal to the proportion of variability explained by the level effect (`R2.full = R2.lvl`) and finally, we reject that the proportion of variability explained by the model is equal to the proportion of variability explained by the pattern effect (`R2.full = R2.pat`). In this example, we can conclude that that it is the overall level of aptitude and not the pattern of the scores on the cognitive test that are associated with someone's interested in the profession of medicine.

Additional information can be extracted from the `mod` object by calling the `summary` function. This includes the proportion of variability explained by the different effects, the level component, and the pattern component. This information is omitted here because the pattern component would be a large matrix (250 rows by 6 columns). Other useful information from the criterion-related profile analysis is stored in `mod` and the contents of this object can be examined by running `str(mod)`.

Finally, a cross-validation technique is also available for criterion-related profile analysis using the `pcv` function. The `pcv` cross-validation technique randomly splits the data into two equally sized samples and then estimates the level and pattern effects in each sample.

3.4 An example of profile analysis via multidimensional scaling

In PAMS, the researcher's main interest is the variation in the latent profiles of examinees, instead of the examinees' observed profiles. 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 `PS` data include three types of profile patterns: Linearly increasing, inverted V, and linearly decreasing. To preview the `PS` data:

```
> data("PS, package = "profileR"")
> 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 set is a person ID variable, and the remaining columns are scores from the neurotic, psychotic, and character disorder scales respectively. Assuming that there are two latent dimensions that can be extracted from the four scores in the `PS` data set, the `pams` function 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 each of the two latent dimensions that we hypothesized, 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 the two latent dimensions extracted from the data.

3.5 Profile reliability using the `EEGS` data set

We will use the `EEGS` data set to demonstrate the concept of profile reliability and how it works using the `pr` function. The `EEGS` data set includes three subtests: quantitative 1 (Q1), quantitative 2 (Q2), and verbal (V). For each subtest, there are two subscores that come from two parallel forms. The first six rows of the data set can be viewed using the `head` function:

```
> 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
```

To compute between-person and within-person subscore reliability of the three subscores in the `EEGS` data set, the `pr` function can be used as:

```
> result <- pr(EEGS[, c(1, 3, 5)], EEGS[, c(2, 4, 6)])
> 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. Calling `plot(result)` would return a scatter plot of the level values from the two parallel forms.

4 Conclusions and future direction

The **profileR** package has been developed to provide researchers and partitioners in the social sciences and medicine that are interested in using profile analysis with a suite of easily accessible and commonly employed techniques for the R programming language without needing to worry about the underlying statistical model. Currently there is no specialized software for profile analysis, although popular matrix languages — such as `PROC IML` for the SAS and the Matrix Command Language in SPSS — can be used for programming a profile analysis procedure. Because profile analysis is a multivariate statistical technique, any software program capable of running multivariate analyses could be used. However, such an approach would require additional, intermediate steps of data analysis. For instance, the `pams` function in the **profileR** package performs profile analysis and multidimensional scaling together, whereas the same procedure would require creating a dissimilarity matrix between variables using a matrix computation procedure and analyzing this matrix using either the `PROC MDS` function in SAS, the `mds` function in R, or the `ALSCAL` function in SPSS. Similarly, the `cpa` and `pbg` functions are capable of testing multiple hypotheses simultaneously, without requiring the user to implement separate analyses.

In addition to the functions described above, the **profileR** package includes functions for a within-person factor model to derive a score profile (M. L. Davison et al., 2009) (using the **lavaan** package Rosseel (2012))

and experimental support for moderated profile analysis (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 the testing of these effects separately), Bayesian profile analysis, and the addition of increased graphical capabilities using the **DiagrammeR** package (Sveidqvist et al., 2015). We welcome all contributions to the **profileR** package and would like to encourage the R community and practitioners in this field to submit pull 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.* (Unpublished doctoral dissertation). University of Minnesota.

Bulut, O., Davison, M. L., & Rodriguez, M. C. (2017). Estimating between-person and within-person subscore reliability with profile analysis. *Multivariate Behavioral Research*, *52*(1), 86-104. doi: 10.1080/00273171.2016.1253452

Bulut, O., & Desjardins, C. D. (2018). profiler: Profile analysis of multivariate data in r [Computer software manual]. Retrieved from `https://CRAN.R-project.org/package=profileR` (R package version 0.3-5)

Conger, A. J., & Lipshitz, R. (1973). Measures of reliability for profiles and test batteries. *Psychometrika*, *38*(3), 411–427.

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

Cronbach, L. J., & Gleser, G. C. (1953). Assessing similarity between profiles. *Psychological Bulletin*, *50*(6), 456.

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

Davison, M., 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). Washington, DC: American Psychological Association.

Davison, M. L. (1994). Multidimensional scaling models of personality responding. In S. Strack & M. Lorr (Eds.), *Differentiating normal and abnormal personality.* Springer Publishing Co.

Davison, M. L. (1996). *Addendum to multidimensional scaling and factor models of test and item responses.* (Unblished Report, Department of Educational Psychology, University of Minnesota)

Davison, M. L., & Davenport, E. C. (2002). Identifying criterion-related patterns of predictor scores using multiple regression. *Psychological Methods*, *7*(4), 468–484.

Davison, M. L., 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, M. L., Kim, S. K., & Close, C. (2009). Factor analytic modeling of within person variation in score profiles. *Multivariate Behavioral Research*, *44*, 668–687.

Ding, C. (2001). Profile analysis: Multidimensional scaling approach. *Practical Assessment, Research & Evaluation*, *7*(16).

Kuder, M. W., G. F.and Richardson. (1937). The theory of the estimation of test reliability. *Psychometrika*, *2*(3), 151–160. Retrieved from `http://dx.doi.org/10.1007/BF02288391` doi: 10.1007/BF02288391

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36. Retrieved from `http://www.jstatsoft.org/v48/i02/`

Stanton, H. C., & Reynolds, C. R. (2000). Configural frequency analysis as a method of determining wechsler profile types. *School Psychology Quarterly*, *15*(4), 434–448.

Sveidqvist, K., Bostock, M., Pettitt, C., Daines, M., Kashcha, A., & Iannone, R. (2015). Diagrammer: Create graph diagrams and flowcharts using R [Computer software manual]. Retrieved from `https://github.com/rich-iannone/DiagrammeR` (R package version 0.8.2)

Watkins, M. W., Glutting, J. J., & Youngstrom, E. A. (2005). Issues in subtest profile analysis. In F. DP & H. PL (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues.* (2nd ed., pp. 251 – 268). New York: Guilford Press.