

Detecting Unusual Score Patterns in the Context of Relevant Predictors

W. Joel Schneider¹  · Feng Ji²

Received: 22 October 2022 / Revised: 4 November 2022 / Accepted: 7 November 2022 / Published online: 15 February 2023
© The Author(s), under exclusive licence to American Academy of Pediatric Neuropsychology 2023

Abstract

Neuropsychological assessment requires integrating new information with what is already known about an examinee. In constructing a case conceptualization, it can be helpful to quantify how unusual a pattern of scores is in the context of other quantitative data. This tutorial explains how two powerful statistical and psychometric concepts, conditional distributions and the Mahalanobis distance, are especially useful when used together to identify unusual score patterns after controlling for other scores. Graphical illustrations, R code, and instructions for a free and user-friendly web app are provided.

Keywords Conditional distributions · Mahalanobis distance · Psychometrics · Neuropsychological assessment

Introduction

A single test score, interpreted in isolation, means precious little in an assessment. A practitioner might interpret the same numeric score as unremarkable in one evaluation, surprisingly high in a second, and distressingly low in a third. The full meaning of any one score becomes apparent only in the context of what else is known about the individual.

Early in the interpretive process, practitioners ask whether an examinee's scores are unusual compared to scores from comparable individuals. To be sure, a complete interpretation of a score involves many considerations beyond the question of prevalence (e.g., reliability, validity, accuracy, generalizability, applicability, fairness, and cultural relevance). However, judging scores to be atypical generally prompts greater scrutiny. Perhaps a scoring error has occurred. Perhaps follow-up testing is needed. Perhaps new hypotheses need to be investigated.

Perhaps a broader range of abilities needs to be assessed. Perhaps the referral concern needs to be revisited.

In this paper, we will illustrate a flexible method for quantifying how unusual a set of scores is in the context of other scores. This method allows practitioners to answer a wide variety of questions, some of which are listed in Table 1. We describe how the application of two statistical concepts—conditional distributions and the Mahalanobis distance (Mahalanobis, 1936)—can answer important questions when used separately but are especially powerful when used in combination.

We will present these methods conceptually, graphically, and, for the mathematically intrepid, with equations. However, applying these tools in practice is virtually impossible without software. The primary calculations were performed via `unusualprofile` (Ji, 2018; Schneider & Ji, 2021), a free package developed in the R programming environment (R Core Team, 2021). The use of the `unusualprofile` package is explained in its online tutorial, which we will not duplicate here. Because most practitioners are not familiar with the R programming language, we have made the core functionality of the `unusualprofile` package available in a convenient web application here <https://w-joel-schneider.shinyapps.io/unusualprofile/>.

To avoid repeated citation, we note that in preparing this article, we have drawn heavily—and no doubt unconsciously—from many authoritative sources on psychometrics, statistical analysis, and linear algebra (e.g., Cohen et al., 2003; Furr, 2017; Crocker & Algina, 2006;

✉ W. Joel Schneider
schneider@temple.edu

¹ College of Education and Human Development,
Temple University, 1301 Cecil B. Moore Ave,
Philadelphia, PA, 19122 USA

² Department of Applied Psychology and Human Development,
University of Toronto, 252 Bloor St. West, Toronto, ON,
M5S1V6, Canada

Table 1 Questions a conditional Mahalanobis distance can answer

Questions	Examples
Given a predictor set, how unusual are the outcomes?	With a given set of cognitive ability scores, how unusual is the profile of academic scores? With a given set of cognitive ability scores, how unusual is the profile of adaptive functioning scores? With a given set of immediate memory scores, how unusual is the profile of delayed memory scores? With a given set of verbal and visual memory span tests, how unusual is the profile of verbal and visual working memory tests? With a given set of basic reading and calculation scores, how unusual is the profile of reading fluency and calculation fluency scores?
Given a set of Time 1 scores, how unusual are the Time 2 scores?	Given an examinee's broad ability scores from a previous evaluation, how unusual are the examinee's current broad ability scores?
Given one or more composite scores, how unusual are the subtest scores that compose them?	Given an examinee's general ability score, how unusual is the examinee's pattern of broad ability scores that contribute to the general ability score? Given an examinee's broad ability scores, how unusual is the pattern of the subtest scores that contribute to the calculation of those broad ability scores?

Nunnally, 1967; Raykov & Marcoulides, 2011; Eaton, 2007; Strang, 2016).

Conditional Distributions

A conditional distribution is the distribution of one or more variables when one or more other variables are held constant. If a predictor and an outcome variable can be assumed to be multivariate normal, one can use conditional distributions to answer many kinds of probability-related questions. For example, how common is it for a person with a vocabulary score of 120 to have a reading comprehension score of 90 or lower?

To answer such questions, we need to know the population means, standard deviations, and correlations of all relevant variables. Generally these parameters are approximated well from a test battery's standardization sample. In this example, the vocabulary and reading comprehension tests correlate at $\rho_{XY} = .5$ and have means of $\mu = 100$ and standard deviations of $\sigma = 15$. We also assume that people's scores are individually and identically distributed with a multivariate normal distribution so that we can omit the subscript when describing random variables (e.g., X).

Univariate Conditional Distributions

In the sections that follow, a more general set of equations for conditional means and standard deviations will be

presented. For now, we focus on the simplest conditional distribution, when a single outcome is conditioned on a single predictor. In a simple regression, the conditional mean y' is the expected value of the outcome Y when X is a specific value x :

$$y' = \mathbb{E}(Y | X = x) \quad (1)$$

The specific value of y' , can be calculated like so:

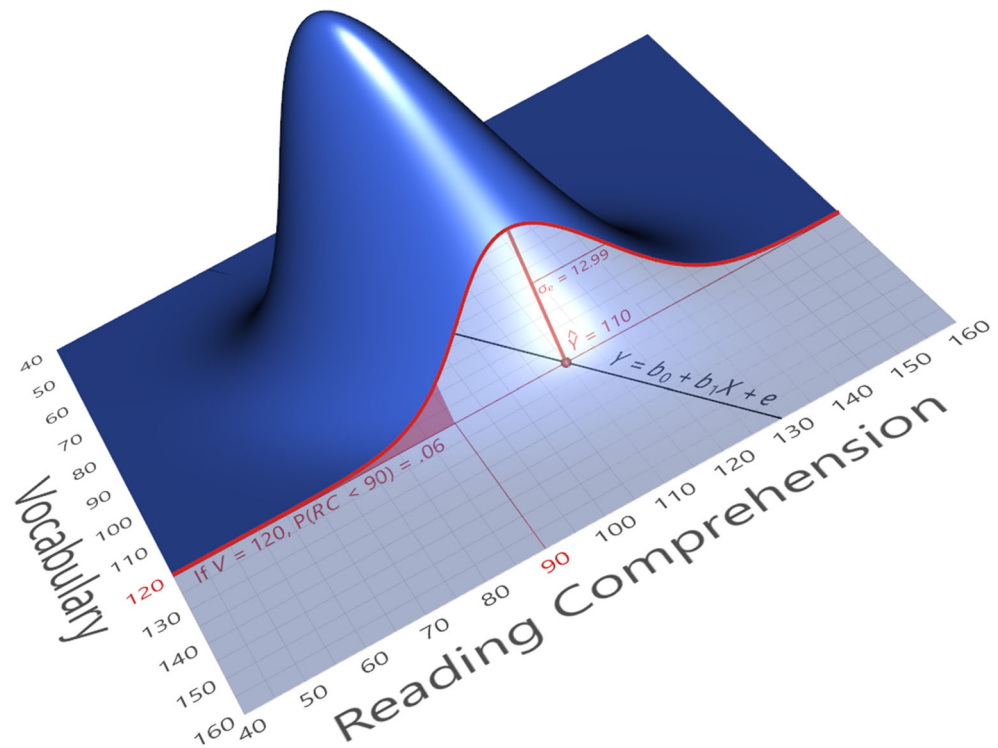
$$\begin{aligned}
 \mathbb{E}(Y | X = x) &= \sigma_Y r_{XY} \frac{x - \mu_X}{\sigma_X} + \mu_Y \\
 y' &= 15 \times .5 \frac{120 - 100}{15} + 100 \\
 y' &= 110
 \end{aligned} \quad (2)$$

In a simple regression, the conditional standard deviation, known as the *standard error of the estimate* (σ_e), can be calculated like so:

$$\begin{aligned}
 \sigma_e &= \sigma_Y \sqrt{1 - r_{XY}^2} \\
 &= 15 \sqrt{1 - .5^2} \\
 &\approx 12.99
 \end{aligned} \quad (3)$$

In Fig. 1, the unconditional bivariate normal distribution is shown in blue. The conditional distribution of reading comprehension when vocabulary is 120 is shown in red. Because reading comprehension is normally distributed, the conditional distribution is also normally distributed. Thus, we can use the standard normal cumulative distribution function $\Phi(\cdot)$ to look up the proportion of individuals

Fig. 1 When vocabulary is 120, the conditional distribution for reading comprehension (shown in red) has a conditional mean of $\hat{Y} = 110$ and a conditional standard deviation of $\sigma_e = 12.99$. Among examinees with Vocabulary = 120, about 6% have reading comprehension scores of 90 or less



with vocabulary = 120 who score 90 or less on reading comprehension. Using information from Eqs. 2 and 3:

$$P(Y \leq y' | X = x) = \Phi\left(\frac{Y - y'}{\sigma_e}\right)$$

$$P(\text{Reading Comprehension} \leq 90 | \text{Vocabulary} = 120) \approx \Phi\left(\frac{90 - 110}{12.99}\right) \approx .06 \quad (4)$$

Thus, only about 6% of individuals with vocabulary of 120 score 90 or less on reading comprehension.

Multivariate Conditional Distributions

In a multiple regression analysis with multiple predictors, one can estimate the conditional mean and conditional standard deviation of the outcome variable. When the same predictors are used on multiple outcomes (i.e., multivariate regression), one can calculate not only the conditional means and standard deviations of each outcome, but the entire conditional covariance matrix. This calculation gives us the ability to understand the conditional multivariate behavior of whole sets of variables at a time. For example, given a profile of scores at time 1, what is the likely profile at time 2? Which features of the profile at time 1 are likely to present at time 2 and which features are unstable? If two abilities differ by 20 points at time 1, what is the probability that they will differ by at least 10 points at time 2?

For interested readers, we describe how to use the standard formulas for conditional means and conditional covariances (e.g., Eaton, 2007, pp. 116–119). These formulas become especially useful when multivariate normality can be assumed. All conditional distributions of multivariate normal distributions are themselves normal (e.g., Eaton, 2007, p. 117), which allows for easy calculations of conditional probabilities.

Suppose X is a multivariate normal random variate:

$$X \sim \mathcal{N}(\mu, \Sigma)$$

If X has 2 mutually exclusive subsets $X_1 \sim \mathcal{N}(\mu_1, \Sigma_{11})$ and $X_2 \sim \mathcal{N}(\mu_2, \Sigma_{22})$, then

$$X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}\right)$$

The conditional distribution of X_1 , given that X_2 has specific values in vector x_2 , can be called $X_{1|2}$, the mean of which is:

$$\mu_{1|2} = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2) \quad (5)$$

The covariance matrix of $X_{1|2}$ is:

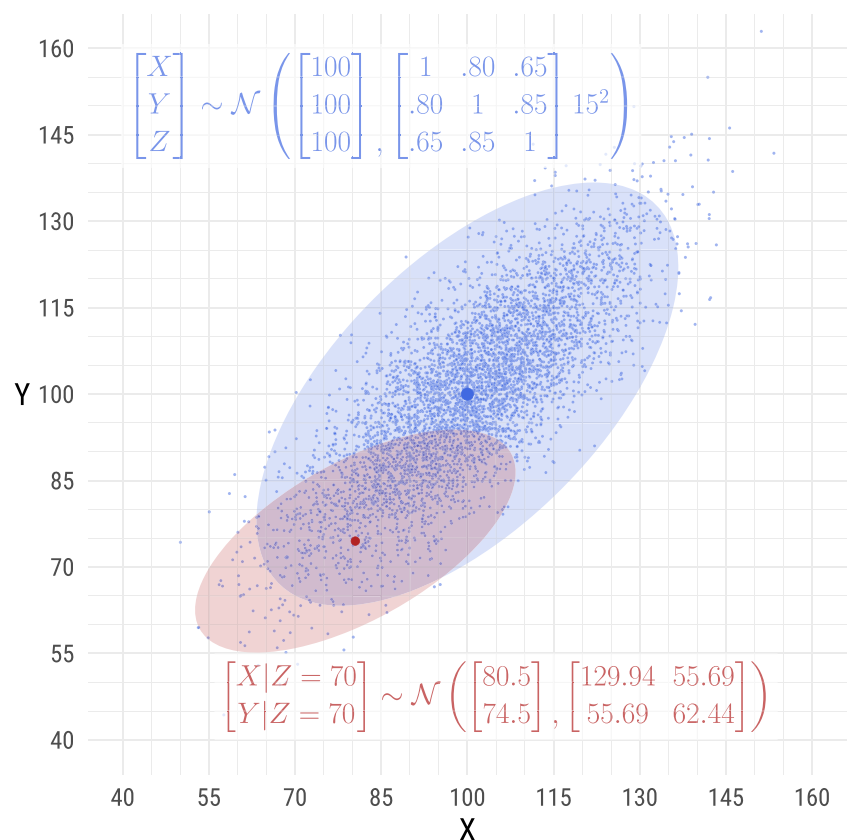
$$\Sigma_{1|2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \quad (6)$$

If X_1 is multivariate normal, the conditional distribution is also multivariate normal:

$$X_{1|2} \sim \mathcal{N}(\mu_{1|2}, \Sigma_{1|2}) \quad (7)$$

In Fig. 2, we illustrate a conditional distribution of variables X and Y when variable $Z = 70$. All variables

Fig. 2 The unconditional distribution of X and Y is shown in blue, and the conditional distribution of X and Y when $Z = 70$ is shown in red



have means of 100 and standard deviations of 15. The correlations of X , Y , and Z appear in Fig. 2.

Applying Eqs. 5 and 6, we can see that many score pairs that are ordinary in the overall (unconditional) distribution of X and Y shown in blue are outside the X and Y conditional distribution's 95% confidence ellipse shown in red. Likewise, many score pairs that are in the conditional distribution's confidence ellipse in red are outside the unconditional distribution's ellipse in blue. As will be seen later, the degree to which a conditional profile is unusual can be quantified.

Longitudinal Conditional Distributions

There is longstanding concern whether the shape of cognitive profiles is stable over time (Canivez & Watkins, 1998; Styck et al., 2019; Watkins, 2000; Watkins & Smith, 2013). We can use conditional distributions to make informed estimates about which features of a profile are likely to be stable and which are not. Using correlations from Styck et al. (2019) as our estimates of the long-term stability of the WISC-IV factor indexes (WISC-V estimates are not currently available), we can get a rough sense of how an examinee's scores are likely to change in the coming years. Because the correlations were estimated from a referred sample selected for academic difficulties, the

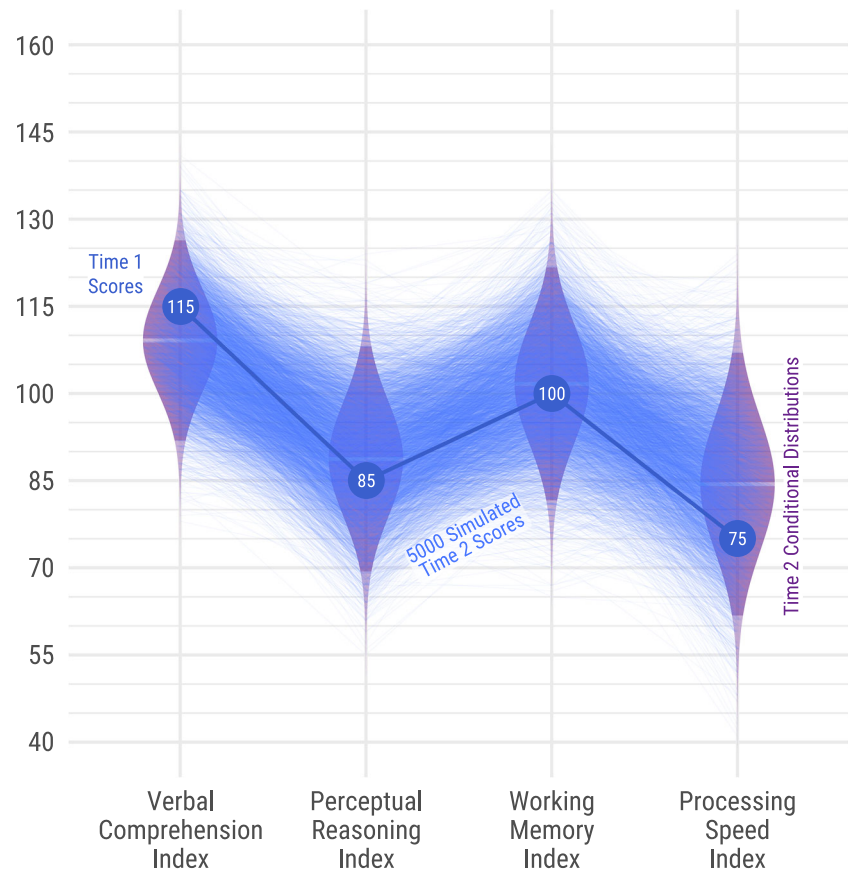
correlations were corrected for range restriction (Alexander et al., 1987).

As seen in Fig. 3, the examinee's time 1 factor index scores shown in blue are extremely uneven. Applying Eqs. 5 and 6, the conditional distributions at time 2 can be estimated.

The conditional distributions at time 2 have roughly the same shape as the time 1 profile, but the conditional distributions (shown as purple "violin plots") are wide enough that one should not expect the profile shape to be exactly the same at time 2. Because the multivariate distribution of the time 2 scores is known, one can simulate an arbitrarily large population of possible time 2 scores and observe their characteristics. This powerful technique allows for estimations of many different kinds of probabilities. The purpose of simulation is not to derive perfect probability statements, but to help guide inferences about what is likely and what is not.

In our calculations, we used the `rmvnorm` function from the `mvtnorm` package (Genz et al., 2017) to simulate multivariate normal distributions with the means and covariances of the time 2 conditional scores. In 5,000 simulated profiles, the time 2 Verbal Comprehension Index was greater than the time 2 Perceptual Reasoning Index and the Processing Speed Index in 97% of the cases. The time 2 Working Memory Index was greater than the time

Fig. 3 Time 1 WISC-IV factor index scores are shown in blue. The conditional distributions of the time 2 WISC-IV scores are shown in purple. The thin blue lines are 5,000 simulated time 2 profiles



2 Perceptual Reasoning Index in 85% of simulated cases and greater than the Processing Speed Index in 89% of cases. The probability that these differences will be large (by 10 points or more) is substantial but not high (61–89%, depending on the comparison). The probability that both VCI and WMI will be greater than both PRI and PSI is .76. The probability that both VCI and WMI will exceed both PRI and PSI by 10 points or more is .44. Thus, the time 2 profile is likely to resemble the time 1 profile, but it is easily possible that the profile will differ noticeably from the current profile.

Many other questions could have been asked besides these. In general, one can simulate a large sample of scores from a condition distribution and then calculate the frequency of any result of interest.

Mahalanobis Distances

The concept of distance might seem straightforward, but distances are not always straight, or forward for that matter. Euclidean distances measure the length of (straight) line segments in flat space with Cartesian coordinates. On curved surfaces, however, walking on straight paths is not

possible. In a city with grid-like streets, one cannot walk along diagonal paths.

For correlated psychometric scores with approximately multivariate normal distributions, it is common to plot data in flat Cartesian coordinates. In Fig. 4 bivariate normal data was simulated, and the confidence ellipse was drawn to contain 95% of the data points. Confidence ellipses are helpful to identify unusual pairs of scores.

In scatterplots with orthogonal axes (i.e., axes at right angles), the Euclidean distances between points do not have a consistent meaning. In Fig. 4, the two variables X and Y are both z -scores with means of 0 and standard deviations of 1. They correlate at .80. Points A and B are equally far from the origin, but point A is well within the confidence ellipse, and point B is outside of it. Likewise, points C and D are equally far from the origin, but C is just inside the confidence ellipse, and D is well outside of it. Thus, distances in a standard scatterplot like Fig. 4 are not comparable.

Mathematicians have devised numerous alternative coordinate systems that fit different kinds of spaces. For correlated variables, it can be productive to plot the data with oblique (skewed) axes rather than orthogonal axes (i.e., axes at right angles). Specifically, if the cosine of the angle

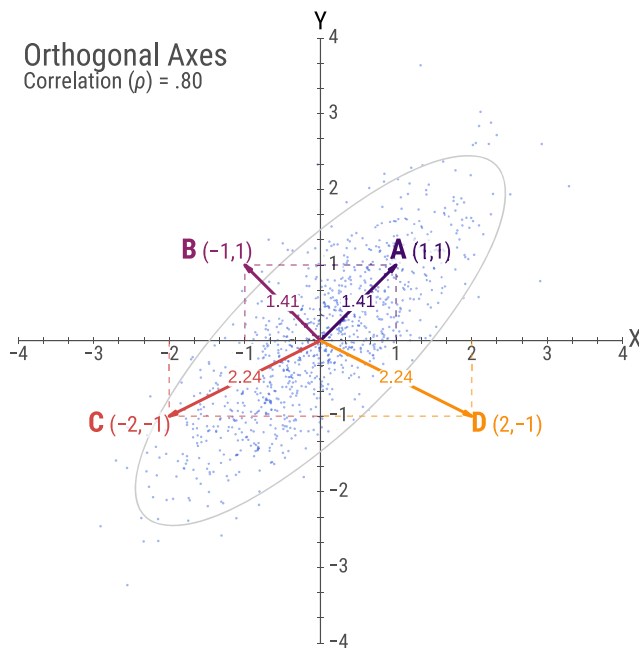


Fig. 4 When scores are plotted on orthogonal axes (i.e., on a right angle), Euclidean distances that seem similar are not comparable. Points A and B are the same distance from the origin, as are points C and D, but B and D and much more unusual than A and C

between the axes equals the correlation coefficient, the distances between each data point plotted with the new coordinates become meaningful. The same data in Fig. 4 was plotted with oblique axes in Fig. 5. Notice how the 4 points appear to have moved, and the confidence ellipse is now a perfect circle.

This idea of calculating distance with oblique axes like Fig. 5 was first articulated and explored by P. C. Mahalanobis (1936), an Indian statistician with many original contributions to the discipline of statistics (Bhattacharya, 2018). Conceptually, plotting data with oblique axes in this manner is the same as plotting data in (standardized) principal components space (Brereton, 2015). Because standardized principal components are uncorrelated z -scores, distances in principal component spaces have a consistent meaning. These Mahalanobis distances are directly comparable and can be interpreted as representing how unusual the scores are in terms of elevation and shape. In Fig. 5, the left-right dimension corresponds to the size of the composite score (i.e., profile elevation) and the up-down dimension corresponds to the standardized $Y - X$ difference score (i.e., profile shape).

In Fig. 5, Point A is now 1.05 units to the right of the origin, which happens to be point A's composite score in z -score units. The difference between A's original X and Y coordinates in Fig. 4 is 0, which is why A's vertical coordinate in Fig. 5 is also 0.

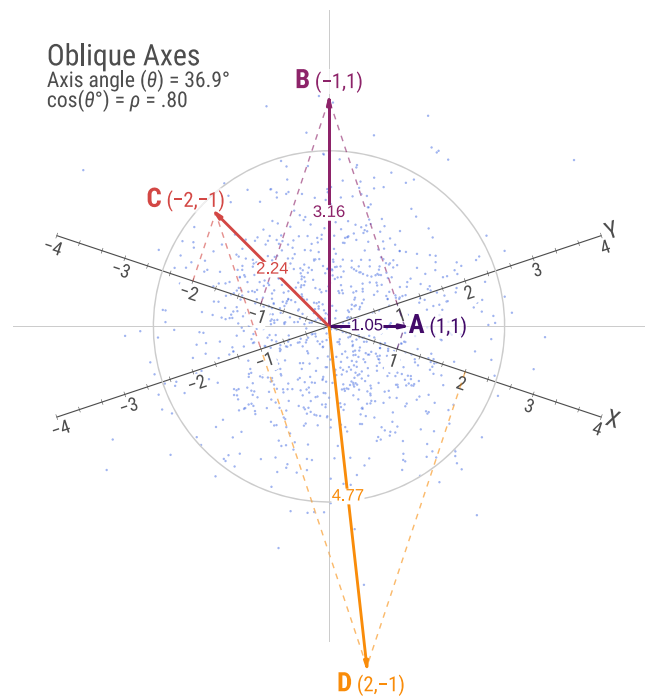


Fig. 5 When scores are plotted on oblique axes such that the cosine of the angle between the axes equals the correlation coefficient between the two variables, distances from the origin are Mahalanobis distances and have a consistent meaning

Point B is 3.16 units directly above the origin. Point B's composite score is 0, which is why it is neither to the left nor right of the origin in Fig. 5. The difference between its original Y and X coordinates is $1 - (-1) = 2$. Standardized to z -score units, this difference is 3.16, which is its placement on the vertical coordinate in Fig. 5.

If k variables are multivariate normal, the Mahalanobis distance squared is the sum of k standard normal principal component scores. Because the sum of k z -scores has a χ^2 distribution with k degrees of freedom, we can estimate how unusual a set of scores is by squaring the Mahalanobis distance and obtaining its prevalence with the χ^2 cumulative distribution function (e.g., the `pchisq` function in R or the `CHISQ.DIST` function in Excel).

In Fig. 5, there are 2 variables and thus 2 degrees of freedom. Point A is 1.05 units from the origin. In a χ^2 variate with 2 degrees of freedom, 1.05^2 is at the 42nd percentile, meaning that this combination of scores is not unusual. For Point B, 3.16^2 is at the 99.3rd percentile, meaning that this combination of scores is unusual. For Point C, 2.24^2 is at the 92nd percentile, meaning that this combination of scores is a little unusual. Point D has a Mahalanobis distance of 4.77 units, which at the 99.999th percentile for profile unusualness. This combination of scores is so rare that we ought to triple-check for scoring errors and do some follow-up testing to make sure that the ability estimates are accurate.

The Mahalanobis distance has many important applications in statistics, particularly in identifying multivariate outliers and comparing group means across many variables (Cronbach & Gleser, 1953). It has also been used to identify unusual score profiles in psychological assessment (Crawford & Allan, 1994; Huba, 1985).

As with calculating conditional distributions, calculating Mahalanobis distances requires knowing the means, standard deviations, and correlations of a set of variables. Imagine that X is a vector of k variables $\{X_1, \dots, X_k\}'$. The means of vector X are collected in vector μ (i.e., $\{\mu_1, \dots, \mu_k\}'$). Likewise, the standard deviations of vector X are collected in vector σ (i.e., $\{\sigma_1, \dots, \sigma_k\}'$). The correlations among the variables in X are collected in matrix R .

The covariance matrix Σ (a bolded capital sigma, not to be confused with the summation sign $\sum$) is calculated with diagonal standard deviation matrices like so:

$$\Sigma = \text{diag}(\sigma) R \text{diag}(\sigma) \quad (8)$$

When variables in X have specific values collected in vector x (i.e., $\{x_1, \dots, x_k\}'$), the Mahalanobis Distance (d_M) is:

$$d_M = \sqrt{(x - \mu)' \Sigma^{-1} (x - \mu)} \quad (9)$$

When X is multivariate normal, the Mahalanobis Distance squared (d_M^2) has a χ^2 distribution with k degrees of freedom. If multivariate normality can be assumed, this property of the Mahalanobis Distance can be used to quantify how unusual a profile of scores is in the population.

Conditional Mahalanobis Distances

The Mahalanobis distance can tell us if a set of scores is unusual in some way compared to population norms. However, we might want to know if an examinee's scores are unusual compared to other individuals who are similar along known dimensions. To answer such questions, we will compute conditional means and covariances and then apply a Mahalanobis distance to the conditional distribution.

We will call Mahalanobis distances applied to conditional distributions, *conditional Mahalanobis distances* (d_{CM}). In most cases, their calculation is exactly like Eq. 9, except that the means are conditional means $\mu_{1|2}$ from Eq. 5 and the covariances are conditional covariances $\Sigma_{1|2}$ from Eq. 6. When X_1 has specific values of x_1 and is conditioned on $X_2 = x_2$, the conditional Mahalanobis Distance of x_1 is

$$d_{CM} = \sqrt{(x_1 - \mu_{1|2})' \Sigma_{1|2}^{-1} (x_1 - \mu_{1|2})} \quad (10)$$

Provided that X_1 and X_2 are not collinear, and X_1 consists of k_1 multivariate normal variables, d_{CM}^2 will be χ^2 distributed with k_1 degrees of freedom. If X_2 consists of k_2

linear combinations of X_1 , $\Sigma_{1|2}$ will not be positive definite and will thus have no inverse. In this case, using Eq. 10 will be impossible. Fortunately, the conditional Mahalanobis distance can still be obtained because it is calculated from within a subspace of X_1 :

$$d_{1|2} = \sqrt{(x_1 - \mu_{1|2})' \Sigma_{11}^{-1} (x_1 - \mu_{1|2})}$$

In this case, d_{CM}^2 will be χ^2 distributed with degrees of freedom equal to the rank of $\Sigma_{1|2}$. Thus, if X_2 consists of k_2 linear combinations of X_1 , then d_{CM}^2 will be χ^2 distributed with $k_1 - k_2$ degrees of freedom.

Calculating Overall Profile Unusualness with the (Unconditional) Mahalanobis Distance

We can evaluate the unusualness of the profile as a whole, or we can evaluate the profile elevation separately from the profile shape. We will illustrate setting up code in R to make the necessary calculations. First we specify the scores, means, standard deviations. Then we use the `cond.maha` function from the `unusualprofile` package to calculate the overall Mahalanobis distance.

The result of this calculation is $d_M = 2.74$, $p = .89$, meaning that the set of scores is above average in profile unusualness (89th percentile), but is not particularly unusual overall.

One of the primary reasons that Cronbach and Gleser (1953) discouraged the use of the Mahalanobis distance with individuals is that in a profile of k variables, it gives all k principal components equal weight. Most of the time, the profile elevation, as summarized by an overall composite score, is more important than the profile shape (McDermott et al., 1990; McDermott et al., 1992; McGill et al., 2018). Additional dimensions beyond elevation may also be more important than others. For example, in cognitive data, interpretive weight is often given to broad ability scores, and individual subtest scores are given little weight until follow-up testing allows for reliable measurement with narrow ability composite scores (Flanagan et al., 2013).

With the VECI, overall elevation as measured by the composite score is the primary score to be interpreted, with the pattern of subtests scores given less attention unless the pattern is extremely unusual. After accounting for the profile elevation, there are three of the four dimensions remaining. Because these three residual dimensions come from differences in the four subtests, they are all of equal interpretive weight and can be summarized by a single number, the conditional Mahalanobis distance.

We therefore illustrate how to distinguish between the profile's elevation and the profile's shape.

```

# Load Packages
library(tidyverse)
library(unusualprofile)

# Scores
x <- c(Similarities = 9,
Vocabulary = 12,
Information = 6,
Comprehension = 11)

# Variable Names
v_names <- names(x)

# Number of Tests
k <- length(x)

# Means (setting variable names)
means <- rep(10, k) |> 'names<-'(v_names)

# Standard Deviations(setting variable names)
SDs <- rep(3, k) |> 'names<-'(v_names)

# Correlations (setting row and column names)
R <- matrix(c(1.00, 0.68, 0.65, 0.59,
0.68, 1.00, 0.71, 0.60,
0.65, 0.71, 1.00, 0.56,
0.59, 0.60, 0.56, 1.00),
nrow = k, ncol = k) |>
'dimnames<-'(list(v_names, v_names))

# Mahalanobis Distance
d_M <- cond_maha(data = x,
R = R,
v_dep = v_names,
mu = means,
sigma = SDs)

```

Evaluating Profile Elevation with a Composite Score

To quickly get a sense of the elevation of a profile, one can look at a composite score that summarizes them, if one is available. If no official composite score exists, it is easy to take the average of the scores. However, quantifying the elevation correctly requires a properly calculated composite score.

For purposes of illustration, we are going to pretend that the VECI composite did not exist and that we need to calculate a custom composite score from the four subtests. Generally these custom composites are exactly the same as the official composites, but rounding error in the reported correlations sometimes results in a score difference of a point or two. Therefore one should rely on the official composite scores when they are available.

To calculate a composite score using Eq. 11, one needs to know the subtest scores, means, standard deviations, and

correlations among the scores. Typically, an unweighted composite is used, in which case all weights $\{w_1, \dots, w_k\}'$ are set to 1. One can, of course, use a weighted composite if needed (e.g., when two tests in a larger composite score are very similar and provide redundant information).

$$\begin{aligned}
 c &= \sigma_C \frac{\mathbf{w}'(\mathbf{x} - \boldsymbol{\mu})}{\sqrt{\mathbf{w}'\boldsymbol{\Sigma}\mathbf{w}}} + \mu_C \\
 &= \sigma_C \frac{\sum_{i=1}^k w_i (x_i - \mu_i)}{\sqrt{\sum_{i=1}^k \sum_{j=1}^k w_i w_j \sigma_{ij}}} + \mu_C
 \end{aligned} \quad (11)$$

where

- c is a generic value of the composite variable C
- μ_C is the mean of composite variable C
- σ_C is the standard deviation of composite variable C
- $\mathbf{x}$ is a vector of specific values of X $\{x_1, \dots, x_k\}'$
- $\mathbf{X}$ is a vector of k variables $\{X_1, \dots, X_k\}'$
- $\boldsymbol{\mu}$ is a vector of the means of X $\{\mu_1, \dots, \mu_k\}'$
- $\boldsymbol{\sigma}$ is a vector of the standard deviations of X $\{\sigma_1, \dots, \sigma_k\}'$
- $\mathbf{w}$ is a vector of weights $\{w_1, \dots, w_k\}'$
- $\boldsymbol{\Sigma}$ is the matrix of covariances (σ_{ij}) for variables in X

In the example, we can apply Eq. 8 to calculate the covariance matrix and Eq. 11 to calculate the composite score. In R, it can be accomplished like so:

```

# Weights
w <- matrix(rep(1, k))

# Covariance Matrix
Sigma <- diag(SDs) %**% R %**% diag(SDs)

# Composite Score
composite_x <- 15 * (t(w) %**% (x - means) /
sqrt(t(w) %**% Sigma %**% w)) + 100

```

Rounded to the nearest integer, the composite score is 97, at the 42th percentile, which is in the average range.

Evaluating a Profile's Shape

In the special case of evaluating a profile of two scores, we can compute two features of the scores: the overall elevation of the two scores and the difference between the two scores.

Although difference scores are relatively easy to understand, computing difference scores for each pair of variables becomes increasingly unwieldy as the number of scores to compare rises. With k scores, $k(k - 1)/2$ difference scores are possible. Thus, with the five WISC-V factor index scores, there are ten pairs of scores to compare. When researchers need to compare the means of many groups, they perform an ANOVA instead of a series of t -tests. For similar reasons, it is desirable to evaluate how unusual

a score profile is with one multivariate distance statistic instead of a series of difference scores.

The geometry of a two-subtest composite score becomes fascinating when the scores are plotted on oblique axes as in Fig. 6. The point where the two scores intersect on oblique axes is the same place in which the standardized principal components intersect. The intersection's distance to the origin (at the variable means) is proportional to the Mahalanobis distance. The horizontal distance is equal to the composite score's deviation from the mean and is proportional to the first principal component. The vertical distance is proportional to the conditional Mahalanobis distance and to the second principal component. By the Pythagorean Theorem, the composite's deviation squared plus the conditional Mahalanobis distance squared equals the overall Mahalanobis distance squared.

There are two more additional features of Fig. 6 that are quite interesting. The angle between the composite vector and each subtest is half that of the angle between the two subtest vectors. The cosine of this composite-subtest angle equals the correlation between the composite

and the subtest. Thus, the correlation between a 2-subtest unweighted composite and its subtests can be found like so:

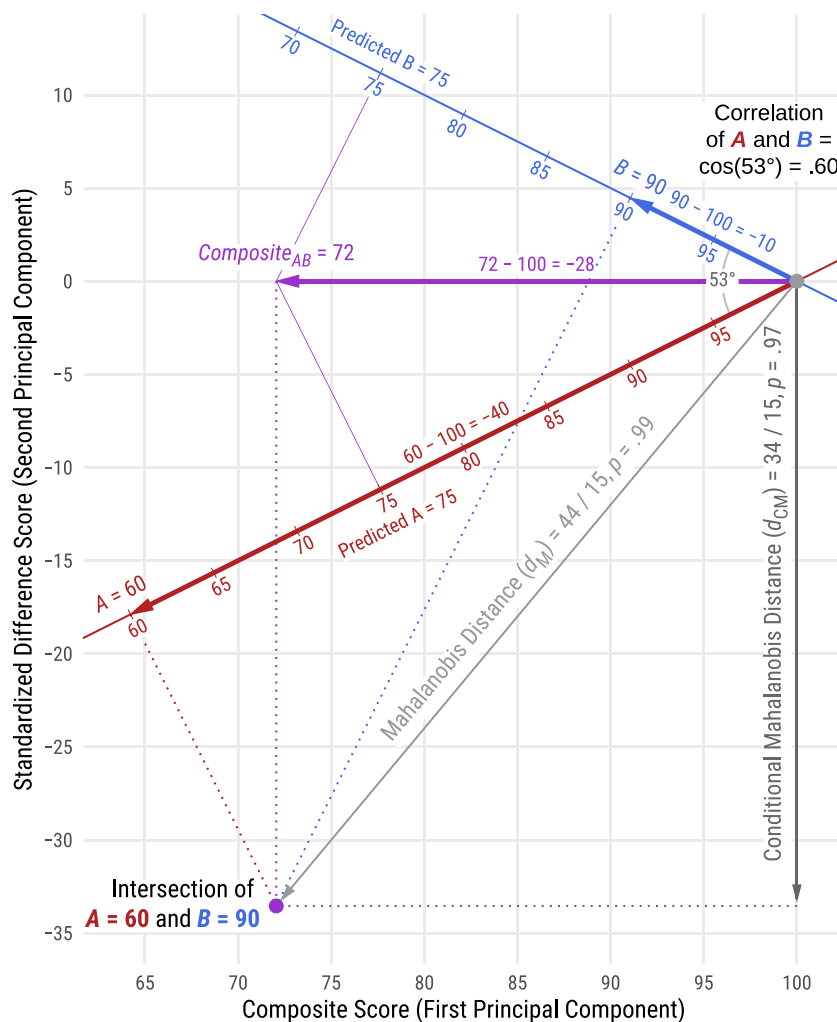
$$r_{CA} = r_{CB} = \cos\left(\cos^{-1}(r_{AB}/2)\right) \quad (12)$$

The composite vector's distance to each of the subtest axes intersects with the predicted subtest scores, given the composite score. That is, in this example, across all individuals with a composite score of 72, *A* and *B* have a conditional mean of 75. The deviations of the observed scores of *A* and *B* from their predicted values are what drives up the conditional Mahalanobis distance. If the examinee's subtest scores were exactly 75, the conditional Mahalanobis distance would be 0.

Although a thorough understanding of Fig. 6 is not necessary to use Mahalanobis distances in practice, the geometry of Mahalanobis distances reveals that, like the geometry of ordinary least squares regression, the Mahalanobis distance is not some arbitrary concept, but is tied to something fundamental about the nature of measurement.

If there are three subtests in a composite, we can plot three oblique axes in three-dimensional space such that

Fig. 6 When subtest scores are plotted with oblique axes such that the cosine of their angle equals the correlation, the length of the vector to the intersection of the subtest scores corresponds to the Mahalanobis distance. The horizontal distance to the intersection corresponds to the extremity of the composite score, and the vertical distance to the conditional Mahalanobis distance



the cosines of the three angles between the axes equals the correlations between the relevant pair of variables. The distance from the origin to the point at which the scores intersect will equal the Mahalanobis distance. A composite vector can be drawn such that the cosine of the angle between the composite vector and each of the three subtest vectors equals the composite-subtest correlation. The distance from the subtest intersection to the composite vector equals the conditional Mahalanobis distance, which summarizes the unusualness of three subtest scores after accounting for the composite score.

These geometric features hold in k dimensions, though it is hard to visualize them when k exceeds three. In Fig. 7, three z -scores form a composite score, with $X_1 = -0.3$, $X_2 = 1.8$, and $X_3 = 2.8$. The composite score that summarizes the three scores has a z -score of 1.74, meaning that the profile elevation is higher than 96 percent of score profiles. The three correlations among the tests are $r_{12} = .50$, $r_{13} = .40$, and $r_{23} = .65$. The point at which the three subtest vectors intersect is represented as a small black sphere. Mahalanobis distance of $d_M = 3.25$ is the distance from the origin to the black point of intersection. The probability associated with a $\chi^2(3) = 3.25^2$ is about .99, meaning that this profile is unusual overall. The conditional Mahalanobis distance of $d_{CM} = 2.74$ is the distance from the tip of the black composite vector to the black point of intersection. Its associated probability of .98 tells us that this 3-test composite score is unusual even after controlling for its high elevation (i.e., Composite score of 1.74).

Returning to our example with the VECI, it is possible to evaluate how unusual a profile is after accounting for one or more important dimensions. The VECI derives from an unweighted sum of the four verbal knowledge subtest scores—Vocabulary, Similarities, Information, and Comprehension. A practitioner may want to know if the pattern of subtests is unusually uneven after accounting for the VECI's elevation.

To get the `cond_maha` function to plot appropriately, we will convert all the subtest scores to the same metric as the VECI, $\mu = 100, \sigma = 15$:

```
# Names of Composite and Subtests
v_SC <- c(v_names, "VECI")

# Subtest-Composite Scores Converted Standard Scores
x_SC <- c(100 + 15 * (x - means) / SDs, composite_x) |>
'names<-`(v_SC)

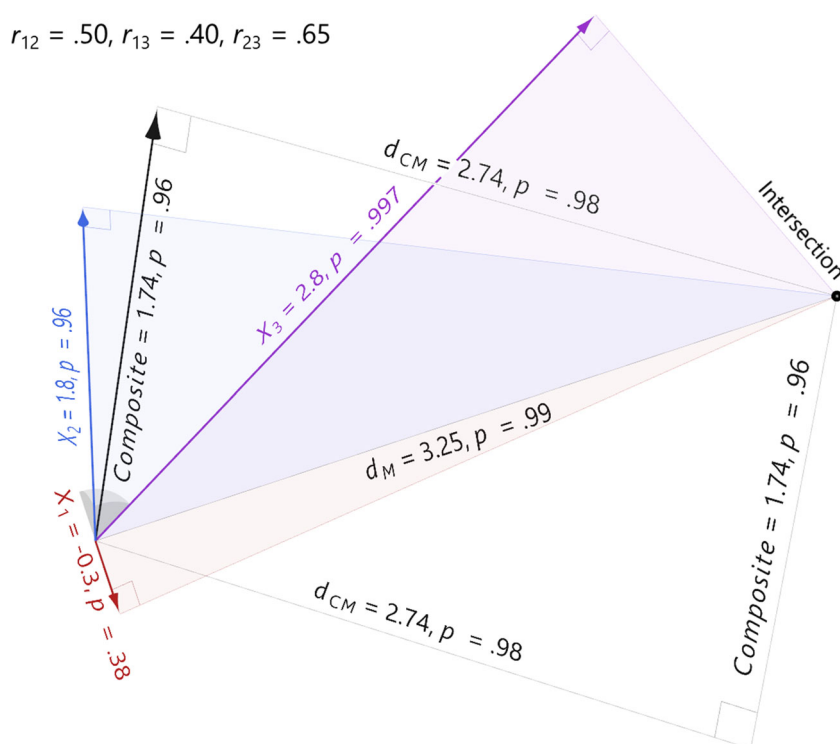
# Subtest-Composite Weights
w_SC <- cbind(diag(k), w) |>
'colnames<-`(v_SC) |>
'rownames<-`(v_names)

# Subtest-Composite Covariance
Sigma_SC <- t(w_SC) %*% (R * 15 ^ 2) %*% w_SC

# Subtest-Composite Correlation
R_SC <- cov2cor(Sigma_SC)

d_CM <- cond_maha(
  data = x_SC,
  R = R_SC,
  v_dep = v_names,
  v_ind_composites = "VECI",
  mu = 100,
  sigma = 15
)
```

Fig. 7 Three z -scores, X_1 , X_2 , and X_3 , form a composite score. The three scores are represented by vectors in three-dimensional space such that the cosine of the angles between them are the subtests' correlation coefficients. The black dot is the unique intersection of three lines at right angles to each score's vector. Representing the profile's overall unusualness, the Mahalanobis distance (d_M) is the distance from the origin to the intersection point. The d_M can be decomposed into two orthogonal distances, the profile's composite score in z -score units and the conditional Mahalanobis distance (d_{CM}), which represents the unusualness of the subtest scores after accounting for the composite score's elevation



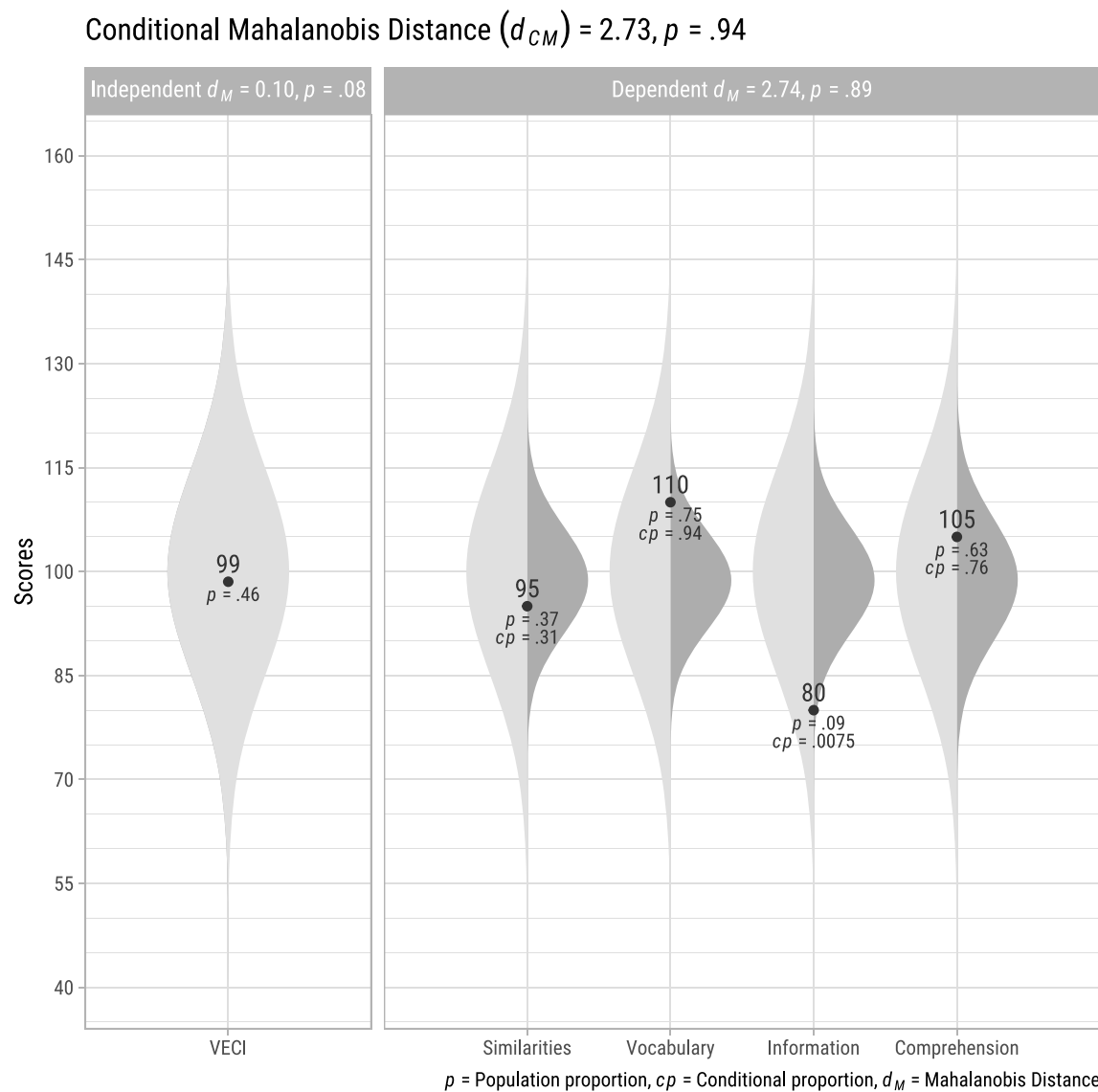


Fig. 8 After controlling for the VECI composite score, the conditional Mahalanobis distance suggests that the pattern of subtest scores is unusual, particularly the Information subtest

The results plotted in Fig. 8 suggest that after accounting for the elevation of the VECI, the subtest score have a pattern that is more unusual than 94 percent of individuals with a VECI = 97. Because this pattern is fairly unusual, it can be informative to inspect the conditional probabilities (cp) in the plot. In this case the Information subtest is unusually low, given the other scores. Possibly, this ability could be tested again with a different test designed to measure general knowledge to see if measurement error has distorted the overall estimate of verbal comprehension and knowledge.

If the entire WISC-V is given, it is not necessary to perform a separate analysis for each composite score. Instead, one can compute a single conditional Mahalanobis distance

with all the composite scores simultaneously predicting all the subtest scores. This conditional Mahalanobis distance would detect unusual subtest score fluctuations within composite groups. In addition, the pattern of factor index scores could be evaluated after controlling for the overall profile elevation (e.g., the Full Scale IQ).

Conditional Mahalanobis Distances with Variable Sets from Different Domains

To showcase the flexibility and utility of conditional Mahalanobis distances, we present two cases that involve one set of variables predicting another set of variables from a different domain.

Example 1

It is common to use cognitive ability scores to explain academic performance (Flanagan et al., 2013). In our experience, it is not always obvious what a profile of academic scores is likely to be, given a cognitive ability profile. Sometimes predictor scores that look like good explanations are revealed to be insufficient once a rigorous analysis is applied.

Suppose a bright child is referred to a neuropsychologist because of difficulty learning to read. The Woodcock-Johnson IV is given because the cognitive and academic ability tests were co-normed with thousands of cases (Schrang et al., 2014). As seen in Fig. 9, the child's WJ IV reading scores are indeed low, ranging from 74 to 87. The child's cognitive ability scores suggest that the child is

bright overall, but there is a striking weakness in auditory processing (Ga). It would appear that this weakness in auditory processing is an obvious explanation of the low reading scores, given that deficits in auditory processing are a primary cause of phonological dyslexia (Alfonso & Flanagan, 2018; Torgesen et al., 1997).

In Fig. 9, the light gray “violin plots” represent vertical normal distributions from the general population ($\mu = 100, \sigma = 15$). The auditory processing (Ga) score of 71 is at the bottom end of the distribution at the 3rd percentile. The darker gray distributions represent the conditional distributions of the academic scores after controlling for all seven cognitive ability scores. Note that some conditional distributions are pulled upward, mostly because of the high scores in Gc (Comprehension/Knowledge) and Gf (Fluid reasoning). Because of its strong relationship with

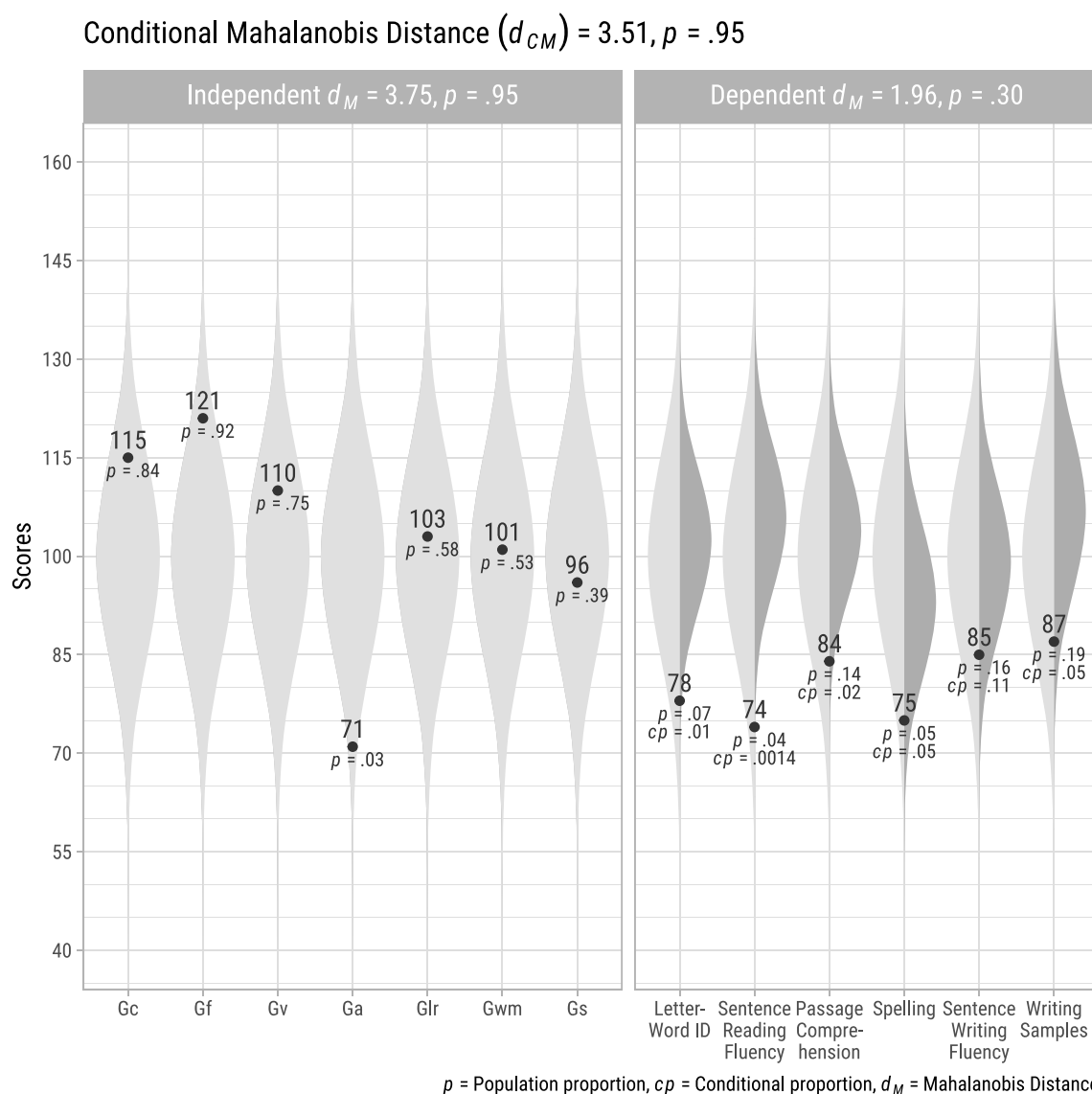


Fig. 9 Not only do these WJ IV cognitive ability scores not adequately explain the low academic scores, they deepen the mystery

auditory processing, Spelling is the only variable to have a conditional distribution below the mean at 93.1.

The Mahalanobis distance of 3.76 ($df = 7$, $p = .95$) for the independent variables (i.e., the cognitive abilities) suggests that the cognitive scores are more unusual than 95% of score profiles in the population. If needed, a separate analysis of just the cognitive profile could be conducted, but it is easy to see that the Ga score is unusually low compared to the other abilities.

The Mahalanobis distance of 1.96 ($df = 6$, $p = .30$) for the dependent variables (i.e., the academic abilities) suggests that the academic scores are not particularly unusual. The academic scores are somewhat low in elevation, but otherwise the profile is fairly flat.

The conditional Mahalanobis distance of 3.52 ($df = 6$, $p = .95$) suggests that the academic scores are more unusual than 95% of profiles from individuals with the same cognitive ability scores as the examinee. The conditional Mahalanobis distance, much like an F test in an ANOVA, does not indicate the nature of the pattern. It simply says that the pattern of score is unusual. In an ANOVA, we use post-hoc comparison tests to inspect the pattern of means, but only if the overall F test is significant. Likewise, we can inspect the conditional distributions of each academic score after we judge the conditional Mahalanobis distance to be unusual.

In Fig. 9, we can see each academic score is positioned at the lower end of the dark gray normal distributions, and

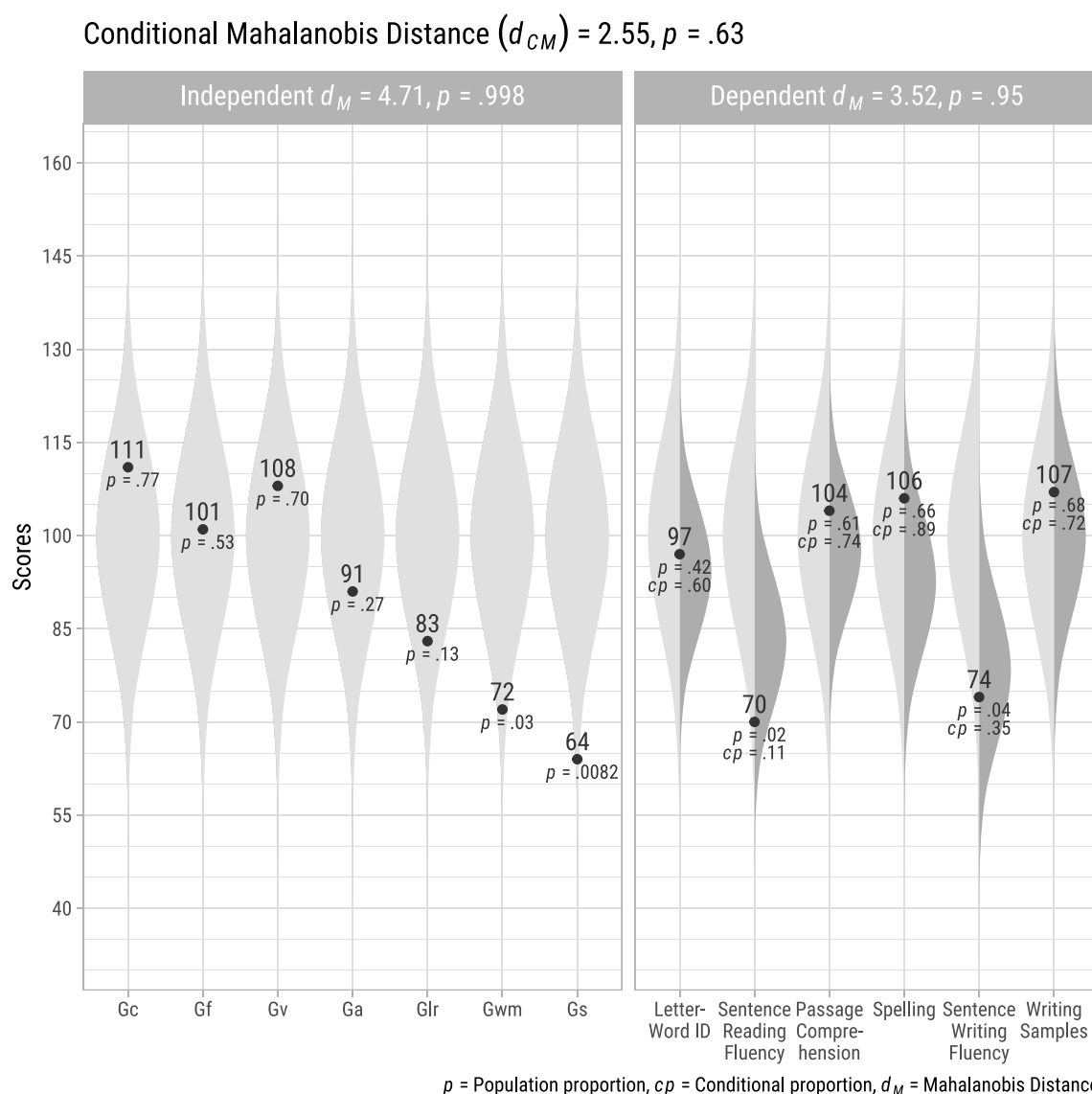


Fig. 10 These WJ IV cognitive ability scores appear to explain the pattern of academic scores

the conditional proportion (cp) is lower than the population proportion (p). Thus, each academic score is lower than expectations, given the cognitive scores.

To be sure, the low Ga score is a partial explanation of the low academic scores. Had Ga been 100 instead of 71, the predicted reading and writing scores would have been several points higher (2 to 11 points, depending on the reading score). Thus, if Ga were average, the set of cognitive ability scores would be an even poorer explanation of the observed reading and writing difficulties.

In this example, the conditional Mahalanobis distance reveals that the cognitive scores are not a sufficient explanation of the child's reading and writing difficulties. The auditory processing deficit may be a partial explanation,

but additional reasons for low academic achievement need to be found, perhaps outside the cognitive domain.

Example 2

Suppose a different child is evaluated for a different kind reading difficulty. The child seems to be able to read and understand short passages but reads slowly and becomes quickly fatigued while reading longer passages.

In Fig. 10, the cognitive scores suggest the child is bright but has low scores in working memory (Gwm) and processing speed (Gs). The child has mostly average reading and writing ability scores but has low scores in measures of reading fluency and writing fluency.

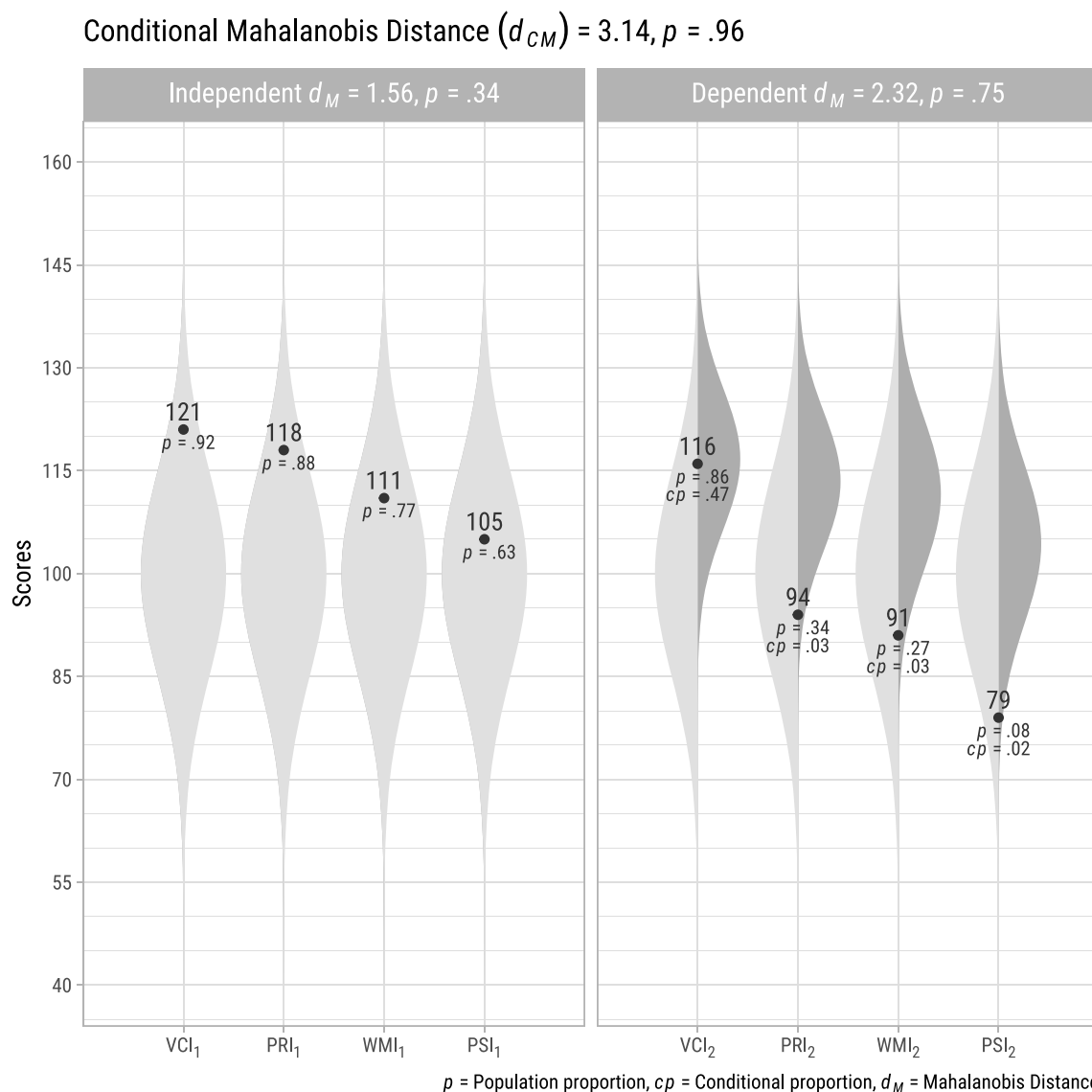


Fig. 11 The WISC-IV factor index scores at time 2 are unusually low, given the WISC-IV factor index scores at time 1

The Mahalanobis distance of 4.70 ($df = 7$, $p = .998$) for the independent variables suggests that the pattern of high and low scores in the cognitive variables is quite unusual. Likewise, the Mahalanobis distance of 3.53 ($df = 6$, $p = .95$) for the dependent variables suggests that the pattern of academic scores is also unusual.

After controlling for the cognitive ability scores, however, the academic scores are not particularly unusual. That is, the conditional Mahalanobis distance of 2.52 ($df = 6$, $p = .61$) means that compared to other individuals with the same cognitive ability scores, the academic profile is at the 61st percentile of unusualness.

An examination of where the academic scores fall in the darker gray conditional distributions suggests that academic scores are not perfectly accounted for by the cognitive scores, but they are not far from expectations. Thus, it seems reasonable that working memory and processing speed deficits are good explanations of the low scores in reading and writing fluency. Of course, the interpretive process does not stop here. Other information may help refine this explanation. For example, the Sentence Reading Fluency score is still at the 11th percentile in the conditional distribution. Perhaps there are other reasons beyond cognitive deficits that contribute to the reading fluency problem (e.g., the child is in the habit of reading aloud rather than subvocalizing the words).

Conditional Mahalanobis Distances with Longitudinal Data

Suppose a neuropsychologist is preparing to do a follow-up evaluation on an adolescent who suffered a traumatic brain injury at age 9 (about 4 years ago). While reviewing records, the neuropsychologist was surprised to see that not only had a WISC-IV been administered at age 10 after the injury, a WISC-IV had been given to the examinee at age 7 for reasons that predated the injury. The scores were clearly lower post-injury, but the neuropsychologist wants to quantify the difference.

To estimate the time 1–time 2 correlations needed to calculate the conditional Mahalanobis distance, the correlation matrices from the longitudinal study by Styck et al. (2019) was used. The sample was 296 cases from special education records in public school districts in the USA. The average retest interval in this study was 2.84 years, which is similar to the interval in this case. The correlations from this study are probably not perfect estimates of the population parameters. However, if not taken too seriously, they are likely to be accurate enough to be informative. As with the example in Fig. 3, the correlations were corrected for range restriction.

In Fig. 11, it can be seen that the WISC-IV Verbal Comprehension Index (VCI) held steady at time 2, but the

other three factor indexes were markedly lower and well below expectations. The conditional Mahalanobis suggested that the factors indexes at time 2 were more unusual than 97% of profile scores compared to individuals with the same time 1 scores as the examinee. The conditional probabilities (cp) of three of the four factor index scores at time 2 were .02 or lower. Thus, even though a Perceptual Reasoning Index (PRI) of 94 or a Working Memory Index (WMI) of 91 are average in the general population, they are unusually low for someone with this examinee's time 1 scores.

Using the Conditional Mahalanobis Distance Web Application

The unusualprofile web application is free to use here: <https://w-joel-schneider.shinyapps.io/unusualprofile/>. Click the *Download Data Template* button to save an Excel file for convenient data entry and storage. A separate file can be used to store information for each test battery a practitioner uses regularly. In the appropriate sheets, enter the examinees test scores, along with the test score means, standard deviations, and correlations. The user can specify custom composite scores to be calculated and used.

For each test battery, a user might want to compute multiple conditional Mahalanobis distances. In the *Distances* sheet, give each distance a name, indicating which role each variable has in the analysis. Enter 0 to omit the variable, 1 for dependent/outcome scores, 2 for independent/predictor scores, and 3 for composite scores that are linear combinations of the dependent/outcome scores. One may insert as many distances as needed.

Once the spreadsheet is complete, upload it to the app using the import data browse button. In the Distance Plot tab, one can select and interpret any distance that has been specified in the uploaded data template.

Conclusion

Conditional distributions are useful for quantifying the probabilities of particular outcomes. Mahalanobis distances are useful for identifying multivariate outliers. Applying Mahalanobis distances to conditional distributions is a powerful and flexible technique for identifying patterns of unusual data in the context of what else is known about examinees. When population parameters are not available or cannot be assumed to be estimated well, future work will investigate how a Bayesian framework can jointly estimate parameters and quantify unusualness at the same time so that the sampling errors for estimating population parameters can be appropriately propagated into quantifying unusualness.

It is important to note that the mere presence of unusual score patterns is not necessarily important (McDermott et al., 1990; McGill et al., 2018). The overall case conceptualization informed by theory, research, and good sense is what matters. Nevertheless, while constructing one's case conceptualization, a good estimation of a pattern's prevalence compared to similar individuals can be an aid to clear thinking.

Author contribution W. Joel Schneider had the original idea of combining conditional distributions and Mahalanobis distances and exploring the geometry of composite scores. Over several years, the authors have collaborated in exploring extensions to the original idea. Feng Ji made the initial version of the unusualprofile package as a part of his Master's Thesis project on applying conditional Mahalanobis distances to latent variable models (Ji, 2018). Dr. Schneider wrote the initial draft of the current paper, created the figures, programmed the web app, and restructured the unusualprofile package. Dr. Ji edited and refined the original draft.

Availability of data and materials All code for generating results and figures can be found at <https://doi.org/10.17605/OSF.IO/XNF7E>. The unusualprofile package is hosted at <https://github.com/wjschne/unusualprofile>. The unusualprofile package's tutorial is at https://wjschne.github.io/unusualprofile/articles/tutorial_unusualprofile.html. The unusualprofile package's web app is at <https://w-joel-schneider.shinyapps.io/unusualprofile>

Declarations

Conflict of interest In the context of illustrative examples, a few commercial tests were mentioned in the text: the WISC-IV, WISC-V, and WJ IV. W. Joel Schneider has no direct financial interest in any of these tests. However, he received a stipend from Pearson, the publisher of the Wechsler series of intelligence tests, for his one-year appointment in 2022 as external adviser to the Pearson Clinical Assessment Scientific Council. Dr. Schneider has also consulted briefly with Riverside Insights, the publisher of the Woodcock-Johnson IV. Feng Ji has no relevant conflicts of interest to disclose.

References

- Alexander, R. A., Carson, K. P., Alliger, G. M., & Carr, L. (1987). Correcting doubly truncated correlations: An improved approximation for correcting the bivariate normal correlation when truncation has occurred on both variables. *Educational and Psychological Measurement*, 47(2), 309–315. <https://doi.org/10.1177/0013164487472002>
- Alfonso, V. C., & Flanagan, D. P. (2018). *Essentials of specific learning disability identification*. New York: Wiley.
- Bhattacharya, S. (2018). Situating Professor Mahalanobis as a public intellectual in the social history of his times. *Sankhya B*, 80(S1), 237–250. <https://doi.org/10.1007/s13571-019-00203-z>
- Brereton, R. G. (2015). The Mahalanobis distance and its relationship to principal component scores: The Mahalanobis distance and PCA. *Journal of Chemometrics*, 29(3), 143–145. <https://doi.org/10.1002/cem.2692>

- Canivez, G. L., & Watkins, M. W. (1998). Long-term stability of the Wechsler Intelligence Scale for Children—Third Edition. *Psychological Assessment*, 10(3), 285–291. <https://doi.org/10.1037/1040-3590.10.3.285>
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences*. Mahwah, NJ: L. Erlbaum Associates.
- Crawford, J. R., & Allan, K. M. (1994). The Mahalanobis Distance index of WAIS-R subtest scatter: Psychometric properties in a healthy UK sample. *British Journal of Clinical Psychology*, 33(1), 65–69. <https://doi.org/10.1111/j.2044-8260.1994.tb01094.x>
- Crocker, L., & Algina, J. (2006). *Introduction to classical and modern test theory*. Mason, OH: Cengage Learning.
- Cronbach, L. J., & Gleser, G. C. (1953). Assessing similarity between profiles. *Psychological Bulletin*, 50(6), 456–473. <https://doi.org/10.1037/h0057173>
- Eaton, M. L. (2007). *Multivariate statistics: A vector space approach* (No. 53). Beachwood Ohio: Inst. of Mathematical Statistics.
- Flanagan, D. P., Ortiz, S. O., & Alfonso, V. C. (2013). *Essentials of cross-battery assessment*. Hoboken, NJ: Wiley.
- Furr, R. (2017). *Psychometrics: An introduction*, (3rd ed.). Los Angeles: SAGE.
- Genz, A., Bretz, F., Miwa, T., Mi, X., Leisch, F., Scheipl, F., & Hothorn, T. (2017). *Mvtnorm: multivariate normal and t Distributions*.
- Huba, G. J. (1985). How unusual is a profile of test scores? *Journal of Psychoeducational Assessment*, 3(4), 321–325. <https://doi.org/10.1177/073428298500300403>
- Ji, F. (2018). *Applying conditional distributions to individuals: Using latent variable models* (Unpublished master's thesis). Illinois State University.
- Mahalanobis, P. C. (1936). On the generalized distance in statistics. *Proceedings of the National Institute of Sciences of India*, 2, 49–55.
- McDermott, P. A., Fantuzzo, J. W., & Glutting, J. J. (1990). Just say no to subtest analysis: A critique on Wechsler theory and practice. *Journal of Psychoeducational Assessment*, 8(3), 290–302.
- McDermott, P. A., Fantuzzo, J. W., Glutting, J. J., Watkins, M. W., & Baggaley, A. R. (1992). Illusions of meaning in the Ipsative assessment of children's ability. *The Journal of Special Education*, 25(4), 504–526. <https://doi.org/10.1177/002246699202500407>
- McGill, R. J., Dombrowski, S. C., & Canivez, G. L. (2018). Cognitive profile analysis in school psychology: History, issues, and continued concerns. *Journal of School Psychology*, 71, 108–121. <https://doi.org/10.1016/j.jsp.2018.10.007>
- Nunnally, J. C. (1967). *Psychometric theory*. New York: McGraw-Hill.
- R Core Team (2021). *R: A language and environment for statistical computing* R Foundation for Statistical Computing, Vienna, Austria.
- Raykov, T., & Marcoulides, G. A. (2011). *Introduction to psychometric theory*. New York: Routledge. <https://doi.org/10.4324/9780203841624>
- Schneider, W. J., & Ji, F. (2021). *Unusualprofile*. AssessingPsyche.
- Schrank, F. A., McGrew, K. S., & Mather, N. (2014). *Woodcock-Johnson IV*. IL: Riverside: Rolling Meadows.
- Strang, G. (2016). *Introduction to linear algebra*, (5th ed.). Wellesley: Cambridge Press.
- Styck, K. M., Beaujean, A. A., & Watkins, M. W. (2019). Profile reliability of cognitive ability subscores in a referred sample. *Archives of Scientific Psychology*, 7(1), 119–128. <https://doi.org/10.1037/arc0000064>
- Torgesen, J. K., Wagner, R. K., Rashotte, C. A., Burgess, S., & Hecht, S. (1997). Contributions of phonological awareness and rapid automatic naming ability to the growth of word-reading skills in

- second-to fifth-grade children. *Scientific Studies of Reading*, 1(2), 161–185. https://doi.org/10.1207/s1532799xssr0102_4
- Watkins, M. W. (2000). Cognitive profile analysis: A shared professional myth. *School Psychology Quarterly*, 15(4), 465–479. <https://doi.org/10.1037/h0088802>
- Watkins, M. W., & Smith, L. G. (2013). Long-term stability of the Wechsler Intelligence Scale for Children—Fourth Edition. *Psychological Assessment*, 25(2), 477–483. <https://doi.org/10.1037/a0031653>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.