Method Selection and Application for Variable Relative Importance Evaluation

Authors: Zhu Xun, Gu Xin, Gu Xin

Date: 2022-07-26T00:00:00+00:00

Abstract

Against the backdrop of an explosion in high-dimensional data, psychological research urgently requires effective methods for assessing variable relative importance. The crux of relative importance assessment lies in selecting appropriate evaluation metrics and statistical inference methods. Among the wide variety of evaluation metrics for relative importance, dominance analysis and relative weights are recommended as key indicators. Statistical inference methods for relative importance are suited to different contexts: Bootstrap sampling is a commonly employed method for inferring the importance of single variables and differences in importance between two variables, while Bayesian testing is an emerging method for evaluating the ordering of importance among multiple variables. Beyond linear regression models, research on relative importance has expanded to include Logistic regression models, structural equation models, and multilevel models, among others, though the applicable data types remain relatively limited. Although relative importance assessment has been widely applied in psychological empirical research, problems of inappropriate metric interpretation and method selection persist. To this end, the assessment process for variable relative importance is illustrated through specific examples.

Full Text

Methods for Evaluating Predictors' Relative Importance: Selection and Application

ZHU Xun, GU Xin
(Department of Educational Psychology, East China Normal University, Shanghai 200062, China)

Abstract

Evaluating predictors' relative importance becomes increasingly critical in psychological research amid the explosion of high-dimensional data. The key to relative importance assessment lies in selecting appropriate measures and statistical inference methods. Among the numerous available indices, dominance analysis and relative weight analysis are particularly recommended. Different statistical inference methods apply to different contexts: Bootstrap sampling is commonly used to infer the importance of single variables and differences between pairs of variables, while Bayesian tests represent a novel approach for evaluating the importance orderings of three or more variables. Beyond linear regression models, relative importance research has expanded to logistic regression models, structural equation models, and multilevel models, though the applicable data types remain somewhat limited. Relative importance assessment has been widely applied in psychological empirical research, yet issues persist regarding inappropriate interpretation of indices and method selection. To address this, we illustrate the evaluation process for variable relative importance using concrete examples.

Keywords: relative importance, dominance analysis, relative weight, Bootstrap, Bayes factor

Against the backdrop of contemporary psychological research focusing on high-dimensional big data, studies increasingly involve more variables, making the assessment of variable importance ever more crucial. For instance, numerous factors at individual, family, school, and societal levels may influence students' school bullying behaviors. Simply put, variable importance refers to its contribution in predicting and explaining the dependent variable. Relative importance assessment helps researchers explore, verify, and refine theories—for example, determining whether family or school factors exert greater influence on bullying behavior, and ranking the importance of variables such as parent-child interaction, parenting styles, and parental emotion regulation within family factors. Relative importance assessment is particularly vital for the effective utilization of variables. First, exploring variable relative importance helps decision-makers identify priorities. Although factors like teacher-student relationships, teachers' attitudes toward bullying, school culture, and class size all affect bullying behavior, administrators with limited resources and time must choose more cost- and time-efficient interventions—for instance, deciding whether developing curricula to raise student awareness of bullying or strengthening teachers' skills in handling bullying incidents is more urgent. Additionally, assessing variable relative importance helps identify individuals requiring intervention. If individual self-esteem and body image are determined to be important factors in bullying behavior, at-risk student groups can be identified early for proactive intervention to prevent or reduce bullying incidents. Therefore, research on methods and applications of variable relative importance holds significant practical value.

Research on variable relative importance originated from regression models.

Darlington's (1968) article in *Psychological Bulletin* was the earliest to discuss relative importance measures, including correlation coefficients, partial correlation coefficients, and standardized regression coefficients, though all were subsequently shown to have notable limitations. Later, psychological researchers proposed dominance analysis (Budescu, 1993; Azen & Budescu, 2003) and relative weights (Johnson, 2000) as relative importance measures, providing a general definition: the proportional contribution of each independent variable to R^2 (the coefficient of determination in regression models) when considering both individual and partial effects (Johnson & LeBreton, 2004). These measures differ in meaning, calculation method, and applicable contexts.

Many empirical studies applying relative importance methods only report estimated importance values and derived rankings while ignoring hypothesis testing results. However, hypothesis testing is essential for statistical inference. After calculating relative importance indices, researchers should ideally use hypothesis testing to infer the significance of variable relative importance and their differences. For example, whether one independent variable's importance is significantly greater than another's, or whether sufficient data evidence supports a particular importance ordering. Researchers typically use bootstrap sampling to calculate standard errors and construct confidence intervals for index estimates (Azen & Budescu, 2003; Tonidandel et al., 2009). Meanwhile, Bayesian factor-based relative importance assessment methods have emerged, offering the advantage of testing relative importance orderings among three or more independent variables while quantifying evidence supporting the importance ordering (Gu, 2021).

In recent years, relative importance research has expanded to other models, such as multivariate regression models (Azen & Budescu, 2006; LeBreton & Tonidandel, 2008), logistic regression models (Azen & Traxel, 2009; Tonidandel & LeBreton, 2010), multilevel models (Luo & Azen, 2013; Rights & Sterba, 2019), and structural equation models (Gu, 2022). Researchers have also developed numerous statistical software packages for calculating variable relative importance indices and inferring their significance, such as the R package `dominanceanalysis` for conducting dominance analysis and the R package `yhat` for implementing both dominance analysis and relative weight methods. Due to their intuitive interpretability and comprehensive model and software support, these methods have gained widespread application in psychology (Casillas et al., 2012; Richardson et al., 2021; Modersitzki et al., 2021).

With numerous relative importance assessment measures available and different statistical inference methods and models applicable to different situations, relevant review literature remains scarce domestically. This paper aims to assist psychological researchers in selecting appropriate variable importance assessment measures and inference methods through discussion and comparison of relative importance assessment methods, while providing direction for future methodological research.

2 Measures of Variable Relative Importance

This section introduces several commonly used relative importance measures within the regression framework. The regression model can be expressed as:

$$\mathcal{Y} = \alpha + \beta_{1X}1 + \cdots + \beta_{JX}J + \epsilon,$$

where $\mathcal{Y}$ is the dependent variable of length N (number of participants), α is the intercept, $X_1, \ldots, X_J$ are J independent variables of length N, $\beta_1, \ldots, \beta_J$ are the corresponding regression coefficients, and $\epsilon \sim N(0, \sigma^2)$ is a residual vector of length N with variance σ^2.

For illustration, consider a simple example. As shown in Figure 1 [Figure 1: see original paper], we establish a regression model with exam scores as the dependent variable and subject ability, study duration, and anxiety level as independent variables:

$$= \alpha + \beta_1 \text{Subject Ability} + \beta_2 \text{Study Duration} + \beta_3 \text{Anxiety Level} + \epsilon_i, \quad i = 1, 2, \ldots, N,$$

assuming the correlation matrix among variables is as shown in Table 1 . Exam scores have a strong correlation with subject ability, study duration affects exam scores, and student anxiety correlates with both subject ability and study duration. We will analyze the relative importance of subject ability, study duration, and anxiety level in this regression model.

2.1 Correlation Coefficients and Regression Coefficients

The correlation coefficient r_j between dependent variable $\mathcal{Y}$ and each independent variable X_j is a fundamental importance index, representing the correlation between an independent variable and the dependent variable and measuring the variable's independent influence and direct predictive power. However, the correlation coefficient ignores other independent variables in the model and fails to consider mutual influences among variables. For example, the correlation between anxiety level and exam scores is $r_3 = 0$, yet anxiety level relates to both subject ability and study duration, thereby also affecting exam scores.

The standardized regression coefficient $\tilde{\beta}_j$ is the most common traditional index used in psychological research to compare independent variable importance. It represents the standardized partial effect of X_j on dependent variable $\mathcal{Y}$ while holding other independent variables constant, interpreted as the expected change in $\mathcal{Y}$ when X_j increases by one standard deviation. However, standardized regression coefficients have drawbacks: their estimation depends on other variables in the model, and a zero regression coefficient does not indicate complete lack of importance—its importance may be masked by other variables. For instance, based on the correlation matrix in Table 1, the standardized regression coefficients for subject ability, study duration, and anxiety level are $\tilde{\beta}_1 = 0.83$, $\tilde{\beta}_2 = 0.00$, and $\tilde{\beta}_3 = -0.17$, respectively, but this does not mean study duration is unimportant for exam scores.

Since both correlation coefficients and standardized regression coefficients can be negative, their squares r_j^2 are generally used as importance indices. When independent variables are uncorrelated, correlation coefficients equal standardized regression coefficients. However, when independent variables are correlated, the two may differ. Correlation coefficients represent each independent variable's individual contribution, whereas standardized regression coefficients represent each variable's unique contribution after controlling for other variables—that is, the variable's unique contribution to the dependent variable.

The product $r_j\tilde{\beta}_j$ is another importance measure. Its advantage over r_j^2 is that it simultaneously considers both individual and unique contributions, and the sum of all independent variables' $r_j\tilde{\beta}_j$ equals R^2, enabling decomposition of R^2—that is, decomposition of explained variance in the dependent variable. However, the product index combines the disadvantages of both correlation and standardized regression coefficients. When either r_j or $\tilde{\beta}_j$ is zero, $r_j\tilde{\beta}_j = 0$. This means that under the product index, an independent variable significantly correlated with the dependent variable but with a zero standardized regression coefficient is deemed unimportant; similarly, a variable uncorrelated with the dependent variable but valuable for explanation or prediction also receives zero importance. In the example, the product indices for study duration and anxiety level are $r_2\tilde{\beta}_2 = r_3\tilde{\beta}_3 = 0$. Additionally, correlation and regression coefficients may have opposite signs, making the product index negative, which complicates interpretation (Johnson & LeBreton, 2004).

Psychological research often involves variable groups, such as socioeconomic status (including income, occupation, education level, etc.) and personality traits (including openness, agreeableness, conscientiousness, etc.). Researchers may wish to assess the joint contribution of variable groups in explaining and predicting the dependent variable—that is, the importance of variable groups. Another disadvantage of correlation coefficients, standardized regression coefficients, and their product indices is their inability to measure variable group importance. The sum or average of correlation or standardized regression coefficients for a group of variables cannot be interpreted as the group's joint effect. For example, subject ability and anxiety level have standardized regression coefficients of 0.83 and -0.17, respectively, but their joint effect is not 0.66. The sum of product indices for a variable group may be negative or smaller than individual product indices, making interpretation difficult. Furthermore, categorical independent variables in regression models, such as occupation and education level, must be converted into sets of dummy variables, and the importance of categorical variables cannot be measured or compared using standardized regression coefficients.

2.2 R^2 Increment

The R^2 increment ΔR^2 in regression models is an intuitive variable importance measure, reflecting the increase in explained variance of the dependent variable after an independent variable enters the model—that is, the variable's unique contribution. The R^2 increment when variable X_j enters last can be expressed as:

$$\Delta R^2(X_j) = R^2(\mathcal{Y}) - R^2(X_{-j}),$$

where $R^2(\mathcal{Y})$ is the total model R^2 and $R^2(X_{-j})$ represents the model R^2 without independent variable X_j. The ΔR^2 method can measure variable group importance; for example, the joint contribution of independent variables X_1, X_2, X_3 can be assessed by their simultaneous entry $\Delta R^2(X_{1X_{2X}}3) = R^2(\mathcal{Y}) - R^2(X_{-123})$, where $R^2(X_{-123})$ denotes the model R^2 without X_1, X_2, X_3. However, similar to standardized regression coefficients, a zero ΔR^2 does not indicate complete absence of effect—the variable's contribution to explained variance may be masked by variables entering earlier. Moreover, the sum of ΔR^2 values when all independent variables enter last does not equal $R^2(\mathcal{Y})$ and thus cannot represent decomposition of R^2. For the exam score regression model, based on the correlation matrix in Table 1, we can calculate $R^2(\mathcal{Y}) = 0.66$, with $\Delta R^2(X_1) = 0.51$ for subject ability, $\Delta R^2(X_2) = 0.00$ for study duration, and $\Delta R^2(X_3) = 0.03$ for anxiety level.

$\Delta R^2(X_j)$ only considers R^2 change when X_j enters last, whereas independent variables can enter in different orders. The average ΔR^2 across different entry orders, denoted $\overline{\Delta R^2}(X_j)$, is considered a more appropriate importance measure (Kruskal, 1987). For example, with four independent variables X_1, X_2, X_3, X_4, there are $4! = 24$ possible entry orders, with Figure 2 [Figure 2: see original paper] presenting some variable entry sequences. For entry order (a), X_3's R^2 increment is $R^2(X_{1X_{2X}}3) - R^2(X_{1X}2)$; for entry order (c), X_3's R^2 increment is $R^2(X_{2X}3) - R^2(X_2)$, where $R^2(\cdot)$ represents the model R^2 with corresponding independent variables. Averaging X_3's R^2 increments across all entry orders yields $\overline{\Delta R^2}(X_3)$. Notably, the sum of average R^2 increments across all independent variables equals the overall model $R^2(\mathcal{Y})$, thus assessing each variable's proportional contribution to explained variance in the dependent variable. For the exam score regression model, $\overline{\Delta R^2}(\text{Subject Ability}) = 0.57$, $\overline{\Delta R^2}(\text{Study Duration}) = 0.08$, and $\overline{\Delta R^2}(\text{Anxiety Level}) = 0.01$. This method incorporates different entry orders to comprehensively measure variable importance and can decompose model R^2. However, it incurs substantial computational costs when many independent variables exist, requiring consideration of $J!$ entry orders for J variables—over one million scenarios when $J = 10$.

2.3 Shapley Value

In fact, the R^2 increment produced by an independent variable depends only on the subset of variables it joins, not on the order of other variables before or after its entry. For example, in Figure 2, X_3's R^2 increment is identical under orders (a) and (b), both being $R^2(X_{1X_{2X}}3) - R^2(X_{1X}2)$. Therefore, only variable subsets need consideration; these subsets can be empty or contain one or more independent variables. Figure 3 [Figure 3: see original paper] illustrates the subsets that X_3 can enter. Subsequently, a weighted average index of R^2 increments across different subsets is calculated, using weights because probabilities of entering each subset may differ.

Shapley value decomposition is based on this principle. Proposed by mathematician Lloyd Shapley in the context of cooperative game theory (Nandlall & Millard, 2019), it measures a player's contribution to a cooperative game's outcome score. A player j's unique contribution to a participation situation (i.e., the existing players when they join) is the score change after player j joins; the Shapley value is the weighted sum of the player's unique contributions across all participation situations, with the sum of all players' Shapley values equaling the total game score when all players participate. Combining linear regression models with the Shapley value method treats independent variables as game players, the game score as R^2, and the player's unique contribution as the R^2 increment when a variable joins a subset.

Variable X_j's Shapley value is denoted $S(X_j)$. For example, with four independent variables X_1, X_2, X_3, X_4, Figure 3 presents the subsets that X_3 might join. X_3's R^2 increment when entering subset (A) is $R^2(X_{1X_{3X}}4) - R^2(X_{1X}4)$, with subset (A) size 2 and total variables 4, giving X_3's entry probability as $2!(4-2-1)! = 1/12$. X_3's R^2 increment when entering subset (C) is $R^2(X_{2X}3) - R^2(X_2)$, with subset (C) size 1, yielding entry probability $1!(4-1-1)! = 1/12$. The weighted average of X_3's R^2 increments across all possible subsets is $S(X_3)$. For the exam score regression model, $S(\text{Subject Ability}) = 0.57$, $S(\text{Study Duration}) = 0.08$, and $S(\text{Anxiety Level}) = 0.01$. Shapley values also decompose R^2, and the subset approach improves computational efficiency. For J independent variables, this method requires considering $2^J - 1$ subsets—just over one thousand scenarios when $J = 10$.

2.4 Dominance Analysis

Another method considering variable entry into subsets is dominance analysis (Azen & Budescu, 2003), which uses pairwise variable comparisons and defines three dominance patterns: complete dominance, conditional dominance, and general dominance. Complete dominance compares two independent variables' R^2 increments across all common variable subsets they can join. For four independent variables X_1, X_2, X_3, X_4, the common variable sets that X_1 and X_2 can join are the empty set $\{\}$, $\{X_3\}$, $\{X_4\}$, and $\{X_{3X}4\}$. If adding X_1 yields greater R^2 increments than adding X_2 across all these subsets, then X_1 completely dominates X_2.

Complete dominance is strict; when multiple independent variables exist in a model, complete dominance often cannot be identified. For example, X_1 and X_2 might show $R^2(X_1) > R^2(X_2)$ when added to the empty set, but $(R^2(X_{1X}3) - R^2(X_3)) < (R^2(X_{2X}3) - R^2(X_3))$ when added to $\{X_3\}$. Therefore, Azen and Budescu (2003) introduced two more lenient indices: conditional dominance and general dominance. Conditional dominance compares the average R^2 increments of two independent variables when added to subsets of equal size. If X_1's average R^2 increment exceeds X_2's for every subset size, then X_1 conditionally dominates X_2. Conditional dominance may also be unidentifiable. General dominance compares the overall average of two variables' conditional dominance indices; if X_1's grand mean R^2 increment exceeds X_2's, then X_1 generally dominates X_2. Complete dominance necessarily implies conditional dominance, which in turn implies general dominance. Variable X_j's general dominance index is denoted $d(X_j)$.

Using the exam score regression model as an example, Table 2 presents the dominance analysis results, where k represents the size of the variable subset entered. The R^2 increment when a variable enters the empty set equals the square of its correlation with the dependent variable; for example, X_1's R^2 increment when entering the empty set is 0.64 (see the $k = 0$ average row in Table 2). X_1's R^2 increment when entering $\{X_2\}$ is $R^2(X_{1X}2) - R^2(X_2) = 0.48$, and when entering $\{X_3\}$ is $R^2(X_{1X}3) - R^2(X_3) = 0.67$, giving X_1's average R^2 increment across subsets of size $k = 1$ as $(0.48 + 0.67)/2 = 0.57$ (see the $k = 1$ average row). X_1's general dominance index is the average of all conditional dominance indices: $(0.64 + 0.57 + 0.51)/3 = 0.57$ (see the overall average row). Table 2 shows that for explaining and predicting exam scores, subject ability completely dominates both study duration and anxiety level. No complete or conditional dominance exists between study duration and anxiety level, but comparing general dominance indices reveals that study duration generally dominates anxiety level.

The general dominance index also possesses the R^2 decomposition property: the sum of all independent variables' general dominance indices equals the model $R^2(\mathcal{Y})$. The general dominance index assesses each variable's contribution to explaining variance in the dependent variable, while complete and conditional dominance provide additional information about importance patterns. Moreover, dominance analysis can obtain restricted relative importance by fixing certain variables in the model and examining the relative importance of remaining variables.

Although the average R^2 increment across different orders, Shapley value, and general dominance analysis measure different concepts, these three methods yield identical variable importance metrics:

$$\overline{\Delta R^2}(X_j) = S(X_j) = d(X_j),$$

and share the same properties: (1) measuring the average contribution of independent variables to explained variance in the dependent variable; (2) decomposing R^2 by allocating variance contributions to each independent variable; and (3) enabling measurement of variable group importance, where a group's joint importance equals the sum of its constituent variables' importance—for example, in general dominance analysis, the joint importance of X_1 and X_2 is $d(X_1) + d(X_2)$.

2.5 Commonality Analysis

Another method for partitioning explained variance is commonality analysis, which differs from previous methods by not assuming that explained variance can be independently allocated to each independent variable, instead recognizing that common variance explanation exists among multiple variables. Commonality analysis partitions explained variance into two components: a unique effect, representing variance explainable only by a particular independent variable, and a common effect, representing variance jointly explained by multiple variables. As shown in Figure 4 [Figure 4: see original paper], ellipses represent each variable's variance, with shaded areas representing variance in $\mathcal{Y}$ explained by X_1 and X_2—that is, their joint contribution, which divides into X_1's unique contribution U_1, X_2's unique contribution U_2, and the common contribution C_{12} of X_1 and X_2. X_1's individual contribution is expressed as $U_1 + C_{12}$. An independent variable's unique effect U is the R^2 increment when that variable enters last, namely the previously mentioned ΔR^2. In Figure 4, the three effects are respectively:

$$U_1 = R^2(X_{1X}2) - R^2(X_2),$$
$$U_2 = R^2(X_{1X}2) - R^2(X_1),$$
$$C_{12} = R^2(X_{1X}2) - (U_1 + U_2) = R^2(X_1) + R^2(X_2) - R^2(X_{1X}2).$$

Unique and common effects are collectively termed commonality coefficients (CC). With multiple independent variables, commonality coefficient formulas can be derived through polynomial expansion. Table 3 lists the commonality analysis results for the exam score regression model, where the sum of all commonality coefficients equals $R^2(\mathcal{Y})$. Variance explainable solely by subject ability accounts for 76% of total variance, while variance jointly explained by subject ability and study duration accounts for 24% of total variance, with these two components explaining the vast majority of variance, indicating that subject ability is most important for explaining or predicting exam scores. Additionally, study duration's unique effect is 0 but it has a substantial common effect with subject ability; anxiety level's unique effect is non-zero, and its common effect with subject ability is negative. Table 3 shows that commonality analysis can produce negative importance estimates, which Ozdemir (2015) explains as indicating the presence of suppressor variables.

Commonality analysis can explain relationships among variables, detect suppressor variables, and comprehensively assess independent variables' contributions to dependent variable variance by considering both unique and common effects (Ozdemir, 2015). Potential issues include difficulty interpreting higher-order and negative commonality effects.

2.6 Relative Weight Analysis

Relative weight analysis assesses variable relative importance from a different perspective. When independent variables are mutually independent, the sum of their squared standardized regression coefficients equals R^2. Therefore, the relative weight method seeks a set of orthogonal variables maximally correlated with the original independent variables to serve as approximations. If original independent variables are not highly correlated, the squared standardized regression coefficients of orthogonal variables can approximate the original variables' relative importance indices. However, if original independent variables are highly correlated, orthogonal variables cannot adequately approximate them, and their squared standardized regression coefficients cannot serve as relative importance indices.

To address this issue, the relative weight method regresses original independent variables onto orthogonal variables (Johnson & LeBreton, 2004). Let X_j be original independent variables and Z_k be orthogonal variables. $\tilde{\beta}_{Z_k}$ is the standardized regression coefficient of dependent variable $\mathcal{Y}$ regressed on orthogonal variable Z_k. Since Z_k are uncorrelated, $\tilde{\beta}^2_{Z_k}$ represents the proportion of variance in $\mathcal{Y}$ explained by Z_k. λ_{jk} is the standardized regression coefficient of X_j regressed on Z_k. Since Z_k are uncorrelated, λ_{jk} equals the correlation between X_j and Z_k, and λ^2_{jk} represents the proportion of variance in X_j explained by Z_k. Moreover, because original independent variables can be expressed as linear combinations of orthogonal variables, orthogonal variables can completely explain original variables' variance, meaning $\sum_k \lambda^2_{jk} = 1$, and due to symmetry of the standardized regression coefficient matrix from original to orthogonal variables, $\sum_j \lambda^2_{jk} = 1$.

Therefore, λ^2_{jk} summed over k yields the proportion of variance in $\mathcal{Y}$ explained by X_j. For three independent variables X_1, X_2, X_3, X_1's relative weight is:

$$w_1 = \lambda^2_{11} + \lambda^2_{12} + \lambda^2_{13}.$$

For the exam score regression model, $w(\text{Subject Ability}) = 0.57$, $w(\text{Study Duration}) = 0.08$, and $w(\text{Anxiety Level}) = 0.01$.

Relative weight analysis can be summarized in four steps: (a) create orthogonal approximations of original independent variables; (b) obtain coefficients between original and orthogonal variables; (c) obtain coefficients between orthogonal and dependent variables; and (d) combine the two sets of coefficients.

The relative weight method can represent R^2 decomposition, measure variable group importance (with w_1+w_2 interpretable as the joint importance of X_1 and X_2), and most importantly, offers high computational efficiency. A potential criticism is that results may differ depending on the orthogonalization procedure employed. Additionally, Thomas et al. (2014) noted that the method's allocation of variance explained by orthogonal variables back to original independent variables remains correlation-based.

2.7 Comparison and Selection of Importance Indices

Table 4 summarizes the properties of various variable relative importance indices. Most relative importance indices are non-negative, while indices related to correlation and standardized regression coefficients cannot compare variable groups. Indices related to R^2 increment can be interpreted as contributions to explained variance in the dependent variable and can compare variable groups. Indices including $r_j\tilde{\beta}_j$, CC, $\overline{\Delta R^2}(X_j)$, $S(X_j)$, $d(X_j)$, and w_j can all decompose R^2. However, both the product index $r_j\tilde{\beta}_j$ and commonality analysis coefficients (CC) suffer from interpretational difficulties.

The average R^2 increment across different orders, Shapley value, and general dominance index yield identical assessment results. Among these, dominance analysis offers intuitive interpretation of variable importance meanings and defines different importance patterns, thus holding particular advantage. The average R^2 increment across orders, Shapley value, dominance analysis, and commonality analysis all require substantial computation. Relative weight analysis provides results nearly identical to general dominance analysis and can be viewed as its approximation, with the advantage of higher computational efficiency.

Note: r_j^2 is the square of the correlation coefficient; $\tilde{\beta}_j^2$ is the square of the standardized regression coefficient; $r_j\tilde{\beta}_j$ is the product of correlation and standardized regression coefficients; $\Delta R^2(X_j)$ is the R^2 increment; CC is the commonality coefficient; $\overline{\Delta R^2}(X_j)$ is the average R^2 increment; $S(X_j)$ is the Shapley value; $d(X_j)$ is the dominance analysis index; w_j is the relative weight.

Based on the above discussion and practical research questions, we offer the following guidance for selecting relative importance indices: First, if researchers focus only on independent variables' individual contributions, correlation coefficients can serve as importance indices; if focusing only on unique contributions, standardized regression coefficients can be used. However, neither index can compare variable groups or categorical variable importance. Second, if researchers wish to explain variable relationships through both unique and common contributions, commonality analysis is appropriate. Third, if researchers focus on independent variables' contributions to explained variance in the dependent variable, dominance analysis and relative weights are suitable importance measures that simultaneously consider individual and unique contributions, measuring each variable's proportional contribution to R^2. Based on Table 4's comparison, we particularly recommend dominance analysis and relative weight indices. The average R^2 increment and Shapley value produce identical results to dominance analysis, but dominance analysis better emphasizes psychologically meaningful importance interpretation. Specifically, we recommend dominance analysis when research focuses on importance patterns and interpretations, and relative weight analysis when assessing relative importance among numerous independent variables. Finally, note that all relative importance indices can be calculated directly from the variable correlation matrix without requiring raw data samples (Gu, 2021).

3 Statistical Inference for Variable Relative Importance

This section discusses statistical inference methods for relative importance. Using the exam score regression model from Section 2 as an example, we illustrate frequentist and Bayesian inference processes for different relative importance indices. We generated multivariate normal data with sample size $N = 200$, zero means, and covariance matrix as shown in Table 1. For illustration, we generated only one dataset, constraining sample means and covariance matrix to equal population values exactly.

3.1 Frequentist Inference

Traditional relative importance indices such as correlation coefficients and standardized regression coefficients follow t distributions. Using standardized regression coefficients as an example, testing the null hypothesis $H_0 : \tilde{\beta}_j = 0$ yields significance tests for standardized coefficients. Regressing on the simulated sample, subject ability yields $\tilde{\beta}_1 = 0.83$, $P < 0.001$; study duration yields $\tilde{\beta}_2 = 0$, $P = 1$; and anxiety level yields $\tilde{\beta}_3 = -0.17$, $P < 0.001$, indicating that subject ability and anxiety level have significant importance for exam scores when using standardized regression coefficients as importance indices. To test whether two independent variables' standardized regression coefficients differ significantly, Wald tests can assess coefficient differences via $H_0 : \tilde{\beta}_1 - \tilde{\beta}_2 = 0$. However, for inferences involving three or more variables, multiple Wald tests suffer from low statistical power even when feasible (Braeken et al., 2015).

The sampling distributions of general dominance analysis and relative weight indices are generally unknown, but bootstrap sampling can construct approximations. Using general dominance analysis as an example, for a sample of size N, we resample S bootstrap samples, computing general dominance indices within each bootstrap sample to obtain the sampling distribution. Figure 5 [Figure 5: see original paper] shows the bootstrap sampling distribution of general dominance indices from the simulated sample.

From the sampling distribution, we can calculate standard errors, confidence intervals, and other statistics for dominance indices or their differences. Bootstrap provides percentile confidence intervals and BCa (bias-corrected and accelerated) confidence intervals, with 95% confidence interval lower and upper bounds being the 2.5th and 97.5th percentiles of bootstrap samples, respectively. BCa confidence intervals adjust percentile intervals using bias-correction and acceleration for greater precision. Wald statistics can be constructed from estimates and standard errors to calculate significance. Since general dominance indices are always non-negative, significance tests for these indices are one-sided. In the simulated sample, subject ability's general dominance index d_1 is estimated at 0.57 with standard error 0.04, bootstrap percentile confidence interval [0.49, 0.65], BCa confidence interval [0.49, 0.65], and Wald statistic 14.36 with $P < 0.001$, indicating that subject ability's general dominance index is significantly different from zero. The difference between subject ability and study duration general dominance indices, $d_1 - d_2$, is estimated at 0.49 with standard error 0.06, bootstrap percentile confidence interval [0.37, 0.59], BCa confidence interval [0.37, 0.60], and Wald statistic 8.79 with $P < 0.001$, showing that subject ability's general dominance index is significantly greater than study duration's—that is, subject ability is more important than study duration for explaining or predicting exam scores. Wald tests and bootstrap methods can only infer relative importance orderings between two independent variables; comparisons involving three or more variables still require multiple testing.

The frequency with which estimated importance index values maintain their ordering across bootstrap samples is called reproducibility of the relative importance ordering. Using general dominance indices as an example, the simulated sample yields estimates $\hat{d}_1 = 0.57$, $\hat{d}_2 = 0.08$, $\hat{d}_3 = 0.01$, with importance ordering $d_1 > d_2 > d_3$. After 1,000 bootstrap samples with general dominance indices computed for each, 980 samples satisfy $d_1 > d_2 > d_3$, giving a reproducibility of 98% for the importance ordering.

3.2 Bayesian Inference

Psychological theories concerning variable relative importance can be expressed as ordering hypotheses. For example, $H_1 : d_1 > d_2 > d_3$ indicates that when explaining or predicting exam scores, the three independent variables' importance under the general dominance index ordering from highest to lowest is subject ability, study duration, and anxiety level. Bayesian tests can directly evaluate ordering hypotheses, with the Bayes factor being the core metric. Compared to traditional methods, Bayesian tests offer several advantages (Hoijtink et al., 2019): they can accept null hypotheses (which significance tests cannot), quantify evidence supporting hypotheses, simultaneously test two or more possible ordering hypotheses without multiple testing corrections, and update results as data accumulate.

The Bayes factor is the ratio of marginal likelihoods of data under two hypotheses. Since null or ordering hypotheses are nested within unconstrained alternative hypotheses, Bayes factors for null versus alternative or ordering versus alternative hypotheses simplify to ratios of posterior to prior densities or probabilities for the tested indices. This simplified Bayes factor expression is known as the Savage-Dickey density ratio or probability ratio (Mulder et al., 2022). Prior densities or probabilities can be calculated directly from specified prior distributions for indices (e.g., the density of a normal prior at index value 0 gives the prior density). Posterior densities or probabilities can be estimated via posterior sampling (e.g., the proportion of posterior samples conforming to the ordering hypothesis gives the posterior probability). Bayes factor calculations for null or ordering hypotheses can be performed using the R package `bain`; see Gu et al. (2018) and Gu (2021) for theoretical details.

Criteria for interpreting Bayes factor evidence are shown in Table 5 (Hu et al., 2018). For comparisons among three or more hypotheses, Bayes factors can be converted to posterior model probabilities.

Using standardized regression coefficients as an example, Bayesian factors can test not only the significance of individual variable importance and differences between two variables' importance but also more informative ordering hypotheses such as $\tilde{\beta}_1 > \tilde{\beta}_2 > \tilde{\beta}_3$ (Gu et al., 2014). In the simulated sample, compared to the alternative hypothesis H_u, $BF_{1u} = 8.17 > 3$ for $H_1 : \tilde{\beta}_2 = 0$ indicates acceptance of H_1, showing study duration's standardized regression coefficient is not significant; $BF_{2u} = 0.00 < 1$ for $H_2 : \tilde{\beta}_2 > \tilde{\beta}_3$ indicates rejection of H_2, showing subject ability and study duration's standardized regression coefficients differ significantly; $BF_{3u} = 4.14 > 3$ for $H_3 : \tilde{\beta}_1 > \tilde{\beta}_2 > \tilde{\beta}_3$ indicates acceptance of H_3, establishing the importance ordering under standardized regression coefficients as subject ability, anxiety level, and study duration.

General dominance indices can likewise be tested using Bayesian methods. For example, if we believe subject ability is unquestionably most important for explaining or predicting exam scores, but are uncertain whether study duration is more important than anxiety level or whether they have equal importance, we might consider two competing hypotheses:

$$H_4 : d_1 > d_2 > d_3$$
$$H_5 : d_1 > d_2 = d_3$$

Calculations yield $BF_{4u} = 8.63$, $BF_{5u} = 0.92$, and $BF_{45} = BF_{4u}/BF_{5u} = 9.44$, indicating that hypothesis H_4 receives substantially more data support than H_5.

Comparing frequentist and Bayesian statistical inference reveals that bootstrap sampling is conceptually simple and suitable for inferring single variable importance and differences between two variables' importance across various indices. However, when comparing three or more variables, bootstrap methods can only provide reproducibility of importance orderings without offering statistical tests. In contrast, Bayesian methods can test any hypothesis about importance orderings, providing a direct and effective approach for relative importance assessment. Nevertheless, Bayesian methods involve more complex statistical principles and computations, and their application in relative importance research remains less common than bootstrap methods.

4 Model Applications of Variable Relative Importance

Extending relative importance assessment based on standardized regression coefficients to multivariate regression models, logistic regression models, and structural equation models is straightforward. Once standardized coefficients and their standard errors are estimated in each model, variable importance can be ranked and inferred using processes similar to those for standardized coefficient estimation, ranking, and inference in regression models; we therefore omit further discussion here. Below we focus on applications of dominance analysis and relative weight methods across various models.

The key to extending dominance analysis to other models is selecting an appropriate variance-explained effect size R^2 (Azen & Budescu, 2006). R^2 relates to model fit and should satisfy: (a) boundedness: R^2 between 0 and 1, where 0 indicates complete lack of fit and 1 indicates perfect fit; (b) linear invariance: R^2 remains unchanged under nonsingular variable transformations; (c) monotonicity: adding an independent variable to the model should not decrease R^2; and (d) interpretability: R^2 should be intuitively understandable. Once R^2 is determined, an independent variable's unique contribution to a specific model can be obtained through its R^2 increment when added to the model. Azen and Budescu (2006) used a multivariate association measure based on canonical correlation as multivariate R^2 to extend general dominance analysis to multivariate regression models. Logistic regression models apply to binary dependent variables; Azen and Traxel (2009) used a likelihood ratio-based R^2 for dominance analysis of independent variables. Luo and Azen (2013) applied dominance analysis to multilevel models. For a comprehensive framework of R^2 effect size measurement in multilevel models, see Rights and Sterba (2019); researchers can select appropriate R^2 based on levels of interest to assess relative importance of independent variables at different levels. Structural equation models can analyze relationships among latent variables; Gu (2022) extended dominance analysis to structural equation models by calculating R^2 for latent variable regression models through model-implied correlation matrices to obtain dominance analysis indices.

The key to extending relative weight analysis to other models is designing orthogonalization procedures and appropriately estimating coefficients. For multivariate linear regression models, directly estimating coefficients fails to account for the intrinsic correlation among dependent variables; LeBreton and Tonidandel (2008) therefore created orthogonal approximations of dependent variables as an intermediate step between orthogonal variables and dependent variables. For logistic regression models, the least squares framework no longer applies; Tonidandel and LeBreton (2010) used standardized logistic regression coefficients as coefficients between orthogonal variables and the dependent variable.

Across these models, relative weight analysis shows consistency with dominance analysis. Relative weight methods have also been extended to regression models containing interaction terms, quadratic terms, or other higher-order terms to assess the relative importance of these effects (Tonidandel & LeBreton, 2011).

Relative importance applications have expanded from linear regression models to binary logistic regression models, multivariate regression models, multilevel models, and latent variable regression models. However, systematic applications to the two major model classes of generalized linear mixed models and structural equation models remain unseen. For example, mediation models are widely used in psychological research; common methods for assessing different mediators' importance involve comparing specific indirect effects, which can be measured by the product of coefficients along each mediator's path (Preacher & Hayes, 2008; Preacher & Kelley, 2011). After calculating specific indirect effects, bootstrap methods can construct percentile and bias-corrected confidence intervals for contrasts between two mediators' specific indirect effects; if the confidence interval excludes 0, the two specific indirect effects differ significantly (Preacher & Hayes, 2008; Fang et al., 2014). Comparison and inference of multiple mediation effects can be implemented in structural equation modeling software such as Mplus and the R package `lavaan`. However, in relative importance assessment, multiple mediators' indirect effect indices share the same problem as standardized coefficient indices in regression models—namely, one mediator's importance may be masked by other mediators. Meanwhile, the recommended dominance analysis and relative weight methods for assessing independent variable importance in regression models have not been extended to multiple mediation models, possibly because multiple mediation models involve more complex paths than regression models, making R^2 interpretation, calculation, and decomposition more difficult. Furthermore, existing model application studies have not addressed categorical independent variables, and the robustness of relative importance assessment when applied to categorical data remains to be examined.

5 Applications of Variable Relative Importance in Psychological Research

To promote practical application of relative importance methods, researchers have developed various software packages across multiple platforms including SPSS, Stata, SAS, and R for calculating relative importance indices and conducting statistical inference (Kraha et al., 2012; Luchman, 2021; Groemping & Matthias, 2021; Mulder et al., 2021). Among these, R packages provide comprehensive and flexible analytical tools: the `dominanceanalysis` package can perform dominance analysis for linear regression models (single and multiple dependent variables), generalized linear models, and multilevel linear models while calculating bootstrap reproducibility; the `yhat` package implements both dominance analysis and relative weight methods in linear regression models and calculates bootstrap confidence intervals (Nimon & Oswald, 2013); the `bain` package can compute Bayes factors for null and ordering hypotheses of any importance index under general statistical models using relative importance index estimates and covariance matrices, enabling statistical inference for multi-variable importance orderings (Gu et al., 2018).

With the development of relative importance methods and software implementation, relative importance assessment has been widely applied in psychological research. Many researchers use dominance analysis and relative weight methods to compare variables' relative importance from the perspective of proportional contribution to dependent variable variance (Casillas et al., 2012; Hakanen et al., 2021; Richardson et al., 2021; Modersitzki et al., 2021). For example, Casillas et al. (2012) used dominance analysis to compare the importance of psychosocial and behavioral factors (such as motivation, self-regulation, and social control) versus prior academic achievement for high school students' academic performance. Modersitzki et al. (2021) used relative weights to compare the relative contributions of sociodemographic variables and personality trait variables (including Big Five and Dark Triad traits) in predicting psychological outcomes during the COVID-19 pandemic. Due to its computational efficiency advantages, the relative weight method is more suitable for models with many independent variables and has been used in numerous meta-analyses to assess relative importance (Young et al., 2018; MacCann et al., 2020). For instance, Young et al. (2018) used relative weights in a meta-analysis to determine different personality traits' relative importance in predicting employee engagement.

Domestic psychological research using relative importance methods has also grown, primarily focusing on mental health, educational psychology, and organizational research (Zhou & Long, 2014; Song et al., 2015; Zhang & Feng, 2018; Xu & Li, 2019; Zhang et al., 2021; Hu & Yuan, 2021). For example, Zhang and Feng (2018) used dominance analysis to compare the relative contributions of impulsivity dimensions to academic procrastination. Xu and Li (2019) used relative weights in a meta-analysis to compare different leadership styles' explanatory power for employee engagement. Hu and Yuan (2021) used Shapley values to compare the importance of family versus school inputs for primary and secondary education outcomes.

Beyond descriptive assessment of relative importance, some psychological studies have focused on statistical inference (Vaessen & Blomert, 2010; Dalal et al., 2012; Ainsworth & Oldfield, 2019). For example, Vaessen and Blomert (2010) used confidence intervals of dominance analysis indices for statistical inference, comparing contributions of phonological awareness and rapid automatized naming to reading fluency across different grades. Ainsworth and Oldfield (2019) used relative weight analysis to compare various factors promoting teacher resilience in predicting job satisfaction, burnout, and well-being, using confidence intervals to infer significance.

Analysis of applied literature reveals several issues in relative importance applications: (1) Inappropriate index selection and interpretation—importance indices have different interpretations and should be selected based on specific research questions. (2) Most applied models are linear regression, possibly because methodological development for relative importance assessment in other models has concentrated mainly in the past decade, with method and software development for multilevel models and structural equation models still incomplete. (3) Importance assessment reporting is primarily descriptive, especially multi-variable rankings, neglecting statistical inference information (Braun et al., 2019). (4) Lack of review literature guidance—despite many new developments and models in relative importance over the past two decades, psychological literature lacks relevant review articles to guide researchers.

6 Practical Example of Variable Relative Importance

This section illustrates practical application of general dominance analysis using a study on children's viewing of the *Sesame Street* television program (Stevens, 2012). Researchers sought to explore factors influencing children's number knowledge. The study data are available in the "sesamesim" dataset in the R package `bain`, including 240 children aged 34 to 69 months. The dependent variable is Postnumb, children's number knowledge test scores one year after watching *Sesame Street*. Independent variables are pre-viewing number knowledge score Prenumb, age Age, and Peabody mental age score Peabody. Their relationship can be modeled with the linear regression:

$$\text{Postnumb} = \alpha + \beta_1 \text{Prenumb} + \beta_2 \text{Age} + \beta_3 \text{Peabody} + \epsilon.$$

Table 6 shows the correlation matrix among dependent and independent variables and estimated standardized regression coefficients. Based on Table 6's correlation matrix, we calculate each independent variable's general dominance index and obtain standard errors, bootstrap percentile and BCa confidence intervals, and Wald test results through bootstrap sampling, as shown in Table 7 . Wald test results indicate that all three independent variables have significant importance in predicting the dependent variable.

Researchers can construct and test specific independent variable importance orderings based on psychological theory while exploring and evaluating all possible importance orderings. Here, to explore variable importance ordering, we evaluate the following hypotheses:

$$H_1 : \theta_{\text{Prenumb}} > \theta_{\text{Age}} > \theta_{\text{Peabody}}$$
$$H_2 : \theta_{\text{Prenumb}} > \theta_{\text{Peabody}} > \theta_{\text{Age}}$$
$$H_3 : \theta_{\text{Age}} > \theta_{\text{Prenumb}} > \theta_{\text{Peabody}}$$
$$H_4 : \theta_{\text{Age}} > \theta_{\text{Peabody}} > \theta_{\text{Prenumb}}$$
$$H_5 : \theta_{\text{Peabody}} > \theta_{\text{Prenumb}} > \theta_{\text{Age}}$$
$$H_6 : \theta_{\text{Peabody}} > \theta_{\text{Age}} > \theta_{\text{Prenumb}}$$

where θ represents importance indices. Using standardized regression coefficients and general dominance indices as examples, Table 8 presents Bayesian inference results for these hypotheses, including Bayes factors for each hypothesis. For standardized regression coefficients, compared to the alternative hypothesis H_u, $BF_{1u} = 0.44$ indicates data neither supports nor opposes H_1, while $BF_{2u} = 5.77$ shows H_2 receives data support. For general dominance indices, $BF_{2u} = 5.54$ indicates data support H_2; all other hypotheses have Bayes factors less than 1/3, indicating greater support for the alternative hypothesis H_u. Among all considered hypotheses, H_2 has the largest Bayes factor BF_{2u}, indicating it receives the most data support. Bayesian tests for both standardized regression coefficients and general dominance indices yield the same result: pre-viewing number knowledge is most important for predicting post-viewing number knowledge test scores, followed by Peabody mental age, with chronological age being least important.

7 Summary and Outlook

Variable relative importance research holds significant meaning for psychology, education, and other social sciences. Although current research practice often uses traditional indices like standardized regression coefficients to compare variable importance, these indices cannot simultaneously consider independent variables' direct effects and their effects after controlling for other variables when independent variables are correlated. The average R^2 increment across different orders, Shapley value, dominance analysis, and relative weights provide unified assessment indices for variable relative importance, yielding nearly identical results. Dominance analysis and relative weight indices simultaneously consider independent variables' individual and unique contributions, decompose R^2, and can measure importance of categorical variables and variable groups, making them recommended measures. More specifically, dominance analysis is recommended when research focuses on importance patterns and interpretations, while relative weight analysis is recommended when many independent variables are involved. Since sampling distributions of dominance analysis and relative weight indices are often unknown, bootstrap confidence intervals are the recommended statistical inference tool. When comparing three or more independent variables' relative importance, Bayesian factor-based statistical inference methods are recommended for assessing multi-variable importance orderings.

Several areas remain to be explored in variable relative importance research. First, model applications of dominance analysis and relative weight methods remain limited, lacking systematic application to generalized linear mixed models and structural equation models. As previously discussed, when comparing different mediators' importance in multiple mediation models, current methods based on product-of-coefficients specific indirect effects may underestimate some mediators' importance. Mediation effects can also be measured using R^2, understood as variance in the dependent variable explained jointly by independent and mediator variables (Lachowicz et al., 2018). Extending dominance analysis and relative weight methods to assess relative importance of mediators represents a future research direction.

Second, although dominance analysis and relative weight analysis are currently the most recommended relative importance assessment methods, each has limitations. Dominance analysis requires computing R^2 increments for all possible subsets, making it computationally inefficient with many variables. Potential improvements include grouping variables for dimensionality reduction and parallel computing based on general dominance analysis's R^2 decomposition property. Additionally, R^2 selection and calculation can be difficult in complex models. For example, multilevel models have many different R^2 measures (Rights & Sterba, 2019), requiring researchers to select appropriate R^2 for dominance analysis. Methodological research on dominance analysis should provide specific application guidelines for such models. Relative weight analysis is widely used due to its similar results to general dominance analysis and high computational efficiency, but its method for allocating variance explained by orthogonal variables back to original independent variables remains correlation-based, potentially leading to inappropriate variance partitioning. This issue must be resolved to extend relative weight methods to broader model classes.

Third, the robustness of variable importance index estimation and inference across different data types in practical analysis requires investigation. For example, psychological research often uses Likert scales, which may assume normally distributed scores to meet regression model requirements. How violations of normality assumptions affect relative importance assessment warrants study. Additionally, psychological research frequently involves categorical variables; whether relative importance assessment of categorical independent variables (e.g., education level) is affected by the number of categories (e.g., graduate, undergraduate, college, high school, junior high or below versus college or above, high school or below) is also worth exploring. Therefore, robustness simulation studies of relative importance assessment represent a future research direction. Finally, integrated analysis software for relative importance assessment that simultaneously performs index calculation and statistical inference is lacking. Future methodological research should therefore focus more on model applications, data suitability, and software development.

Note: Figure translations are in progress. See original paper for figures.