Variance-based importance measures for high-dimensional linear model via Johnson indices: Insights and comparisons

LAURA CLOUVEL
EDF R&D, PERICLES Department, Saclay, France

BERTRAND IOOSS, VINCENT CHABRIDON
EDF R&D, PRISME Department, Chatou & SINCLAIR AI Lab., Saclay, France

In regression analysis, importance measures are effective tools for feature selection and model interpretation, allowing for the ranking of the most influential regressors. In particular, *variance-based importance measures* (VIMs) are a prominent topic in both fields of statistics and global sensitivity analysis. This is due to their accessible interpretation as variance shares of the explained variable. As proposed in [1], this work focuses on a linear regression model between an explained real-valued output random variable Y and d explanatory input random variables $\boldsymbol{X} = (X_1, \ldots, X_d)$: $Y = \boldsymbol{X}\boldsymbol{\beta} + \varepsilon$ with $\boldsymbol{\beta} \in \mathbb{R}^d$ is an unknown vector of coefficients and ε is a centered Gaussian random error. It addresses some of the practical challenges that arise when the component of $\boldsymbol{X}$ are dependent inputs and the input dimensionality d is large. Specifically, the goal is to discuss the formulation and interpretation of Johnson indices [2, 3], which have empirically demonstrated their value both in high-dimensional contexts and their ability to approximate the not so well-known LMG indices [4].

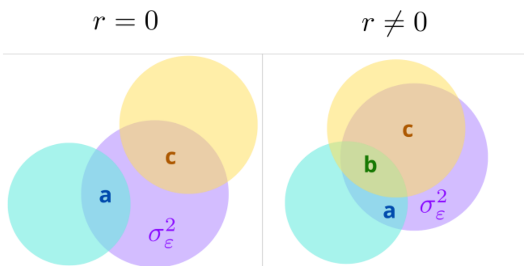
We start by providing some theoretical elements and interpretations to define the context in which Johnson indices can be used in comparison to LMG and PMVD indices [5]. In the literature of linear regression analysis, VIMs are built from the decomposition of the *coefficient of determination* R^2 which quantifies the percentage of output variability explained by the model. A VIM associated with a regressor is thus defined as its proportional contribution to R^2, accounting for both its direct effect (correlation with Y) and combined effects with other variables [6]. Various R^2 decomposition strategies have thus been proposed, leading to different interpretations. The choice of the R^2 decomposition suitable for defining the VIM can then be established based on *desirability criteria*:

- (C_1) *Proper decomposition*: the sum of all shares should be equal to the R^2;
- (C_2) *Nonnegativity*: all shares should be nonnegative;
- (C_3) *Exclusion*: if $\beta_j = 0$, then the share of X_j should be zero;
- (C_4) *Inclusion*: if $\beta_j \neq 0$, then the share of X_j should be nonzero;
- (C_5) *Grouping*: all shares should tend to equate for highly correlated inputs.

The first four criteria were defined by Gromping [7], while the last one relates to regularization techniques [8]. Criteria (C_1) and (C_2) are essential for interpreting VIMs as a percentage of R^2. Criterion (C_4) is also fundamental to highlight inputs with direct influence. However, (C_5) contradicts the exclusion property (C_3). If the interpretation is focused on the direct influence of the inputs on the model output, then (C_3) is appropriate; if the correlations among data can carry necessary information for the interpretation, (C_5) is relevant instead. In this context, the LMG and Johnson indices favor the (C_5) criterion whereas the PMVD indices (C_3). In fact, both methods aim to decompose the R^2, but they differ in how they average the marginal contribution of each variable across all the permutations. LMG uses an arithmetic average while PMVD weights these contributions based on the proportion of variance attributable to each variable.

To better understand and illustrate the concept of *multicollinearity*, we also use Venn diagrams on a two-input regression model ($d = 2$), see Fig. 1. The Venn diagrams are formed by three circles associated with the variances of Y (in purple), X_1 (σ_1 in blue) and X_2 (σ_2 in yellow), by two overlapping area measuring the *additional explanatory power* of X_1 (a) and of X_2 (c), and by the area representing the *combined effect* of the inputs on the model $Y(\mathbf{X})$ (b). We prove, in particular, (with a different demonstration from the one of [9] which relies on geometrical arguments) that there is an equivalence between the LMG and the standardized Johnson indices in the case of a two-input model [1].

Finally, we apply these indices to the well-known dataset of the `R` package `AmesHousing`, which contains 79 features describing house sale prices in Ames, USA [1]. The computational cost of the LMG and PMVD indices is exponential with the number of input variables. It appears impossible to calculate them for the entire set of input variables. This example shows that there are cases where it is impossible to determine the LMG and PMVD indices, and where it is necessary to use approximate methods to conduct sensitivity analyses. In this case, we calculate the Johnson and the well-known SRC^2 indices [1] for the set of 34 quantitative variables. We then determined the 10 most influential variables and we determine all the VIMs for these 10 variables (see Fig. 2).



$\mathbf{a} = \beta_1^2\sigma_1^2(1-r^2)$
$\mathbf{b} = \beta_1^2\sigma_1^2 r^2 + 2\beta_1\beta_2\sigma_1\sigma_2 r + \beta_2^2\sigma_2^2 r^2$
$\mathbf{c} = \beta_2^2\sigma_2^2(1-r^2)$

	$r = 0$	$r \neq 0$	$\mathcal{C}_1$	$\mathcal{C}_2$	$\mathcal{C}_3$	$\mathcal{C}_4$	$\mathcal{C}_5$
$\mathrm{LMG}_1 =$ $\mathrm{Johnson}_1$	$\frac{a}{a+c+\sigma_\varepsilon^2}$	$\frac{a+b/2}{a+b+c+\sigma_\varepsilon^2}$	yes	yes	no if $r \neq 0$	yes	yes
PMVD_1	$\frac{a}{a+c+\sigma_\varepsilon^2}$	$\frac{a[1+b/(a+c)]}{a+b+c+\sigma_\varepsilon^2}$	yes	yes	yes	yes	no
R^2	$\frac{a+c}{a+c+\sigma_\varepsilon^2}$	$\frac{a+b+c}{a+b+c+\sigma_\varepsilon^2}$					

Figure 1: Interpretations of VIMs: Venn diagrams and desirability criteria.

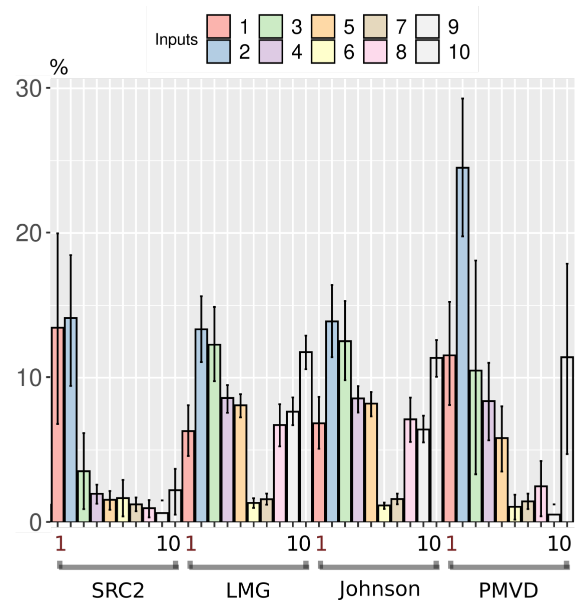


Figure 2: Results on the Ames housing dataset.

References:

[1] L. Clouvel et al. "An overview of variance-based importance measures in the linear regression context: comparative analyses and numerical tests". In: *Preprint* (2024). URL: `https://hal.science/hal-04102053`.

[2] J.W. Johnson. "A heuristic method for estimating the relative weight of predictor variables in multiple regression". In: *Multivariate Behavioral Research* 35 (2000), pp. 1–19.

[3] A. Genizi. "Decomposition of R^2 in multiple regression with correlated regressors". In: *Statistica Sinica* (1993), pp. 407–420.

[4] R. H. Lindeman, P. F. Merenda, and R. Z. Gold. *Introduction to bivariate and multivariate analysis*. Glenview, IL: Scott Foresman and Company, 1980.

[5] B. Feldman. "Relative Importance and Value". In: *SSRN Electronic Journal* (Mar. 2005). DOI: `10.2139/ssrn.2255827`.

[6] J.W. Johnson and J.M. LeBreton. "History and use of relative importance indices in organizational research". In: *Organizational Research Methods* 7 (2004), pp. 238–257.

[7] U. Grömping. "Estimators of relative importance in linear regression based on variance decomposition". In: *The American Statistician* 61.2 (2007).

[8] H. Zou and T. Hastie. "Regularization and Variable Selection via the Elastic Net". In: *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 67.2 (2005), pp. 301–320.

[9] D.R. Thomas et al. "On Johnson's (2000) relative weigths method for assessing variable importance: A reanalysis". In: *Multivariate Behavioral Research* 49 (2014), pp. 329–338.

[Laura Clouvel, EDF R&D, Saclay France; laura.clouvel@edf.fr]
[Bertrand Iooss, EDF R&D, Chatou France; bertrand.iooss@edf.fr]
[Vincent Chabridon, EDF R&D, Chatou France; vincent.chabridon@edf.fr]