



**HAL**  
open science

# Variance-based importance measures in the linear regression context: Review, new insights and applications

Laura Clouvel, Bertrand Iooss, Vincent Chabridon, Marouane El Idrissi, Frédérique Robin

## ► To cite this version:

Laura Clouvel, Bertrand Iooss, Vincent Chabridon, Marouane El Idrissi, Frédérique Robin. Variance-based importance measures in the linear regression context: Review, new insights and applications. 2023. hal-04102053v1

**HAL Id: hal-04102053**

**<https://hal.science/hal-04102053v1>**

Preprint submitted on 22 May 2023 (v1), last revised 5 Jul 2024 (v3)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Variance-based importance measures in the linear regression context: Review, new insights and applications

Laura Clouvel, Bertrand Iooss, Vincent Chabridon,  
Marouane El Idrissi and Frédérique Robin (EDF R&D)

## Abstract

In the context of regression analysis, *importance measures* are effective tools to perform feature selection or to interpret a model by ranking the most influential regressors. In particular, *variance-based importance measures* (VIMs) are prominent in the field of statistics, but also in the most recent field of global sensitivity analysis, due to their accessible interpretation as variance shares of the explained variable. By focusing on the linear regression model, this work aims at revisiting the overview of the most well-founded methods (some of them being rather old and sometimes, misunderstood), while clarifying their respective positioning, conditions of use, intrinsic capabilities, and interpretation. Some challenges are discussed, such as the case of dependent inputs and the case of a high input dimension. A particular set of VIMs, called the *Johnson indices*, has shown the potential to solve both previously mentioned challenges. Their equivalence with other known VIMs, the *Lindeman-Merenda-Gold* (LMG) indices, is proved in the case of two inputs. The practical relevancy of such tools is highlighted through their empirical study on simulated data, as well as public datasets. An application to classification tasks (logistic regression) is also presented in the supplementary material.

**Keywords:** Linear model, Multicollinearity, Proportional values, Relative weight analysis, Sensitivity analysis, Shapley effects, Variance decomposition

## 1 Introduction

In the context of regression analysis, *importance measures* are relevant tools to either perform an insightful feature selection or to interpret a model by allowing to rank the explanatory variables (also called “inputs”) with respect to (w.r.t.) their influence [41, 25]. Indeed, numerous methods allow for quantifying the relative importance of inputs involved in a model used to predict a specific explained variable of interest (also called the “output”). In particular, *variance-based importance measures* (VIMs) are popular due to their intrinsic interpretation as shares of the output’s variance [22, 10, 38, 5, 31]. From a practical viewpoint, they are essential in data analysis and post-hoc interpretation of learned models [16, 47, 43]. Moreover, their properties have motivated their use in the recent field of global sensitivity analysis (GSA) of model outputs, where their versatility and ease of estimation provide undeniable practical benefits [52, 15, 2].

In GSA, VIMs derived from a linear regression analysis constitute, most of the time, the basic elements of any preliminary study according to various methodological reviews (see, e.g., [26, 34, 59, 7]). However, it appears (to the best of the authors’ knowledge) that a large part of the GSA literature dedicated to VIM in linear models omitted some crucial aspects that appeared in the historical development of these importance measures in the statistical research community. As an example, one can mention the *desirable criteria* that an importance measure should verify to be well defined (see, e.g., [38, 25]).

The present work aims at revisiting some well-founded VIMs for linear regression from the statistical literature. Among them, some can be old but still poorly known and underused in practice. A particular focus is put on their properties, conditions of use, and subsequent interpretation. In addition, a discussion about the need to have a clear definition for *relative importance* (sometimes called “relative weight” or “relative contribution” in the literature) from the user viewpoint is proposed. In particular, the VIMs are associated with the definition of *dispersion importance* introduced by [1] and linked to the influence of the inputs on the output variance. In the context of linear regression model, the *coefficient of determination*  $R^2$ , which quantifies the percentage of output variability captured by the linear regression model, can be a key metric to build VIMs. Thus, in accordance with [38], the following definition of what a VIM associated with a specific regressor is: “*the proportionate contribution each variable makes to  $R^2$  (ratio of explained variance to the total response variance), considering both its direct effect (i.e., its correlation with the response) and its effect when combined with the other variables in the model.*” In this way, we reintroduce the approach of the *general dominance analysis* which consists in defining an  $R^2$  decomposition by exhibiting hierarchy among regressors regarding some *dominance criteria* [10].

In this context, a first difficulty arises: how to meaningfully allocate shares of  $R^2$  between statistically dependent inputs? In this paper, we refer to the concept of *multicollinearity* to deal with it. It proposes an intuitive representation of the multicollinearity based on the Venn diagrams (inspired from [13] and also studied in [29]) for a two-input regression model case. Furthermore, it presents classic metrics to deal with multicollinearity and shows that the latter cannot be directly used for an  $R^2$  decomposition. Finally, it justifies the use of more complex VIMs to separate the individual effects of each variable on the output variable. In the literature, it exists various ways to partition the  $R^2$ . The LMG indices [44] and the PMVD indices [20] appears to be the most interesting VIMs but basic desirability criteria need to be defined to differentiate them and to determine their conditions of use.

A second difficulty arises: how to estimate VIMs in high dimensions knowing that the computational complexity of the LMG and PMVD indices is exponential with the number of inputs. Due to their ability to solve the last two issues, the Johnson indices [37] (see also [22]) based on relative weight allocations are highlighted. Their equivalence with the LMG indices is shown for the two-input regression model case.

While this work is not intended as an exhaustive review (we refer to [25] for that purpose), and aside from the theoretical results related to the Johnson indices, the novelty of this paper is threefold. First, we emphasize the links between VIMs in the statistical literature and the field of GSA. Second, it points out recent works on the presented VIMs which pave the way towards more meaningful and theoretically sound interpretation of linear models, especially in the context of highly correlated inputs. For that purpose, ill-defined and non-robust proposed importance measures from the literature are omitted from this work (e.g., first/last methods, pratt, CAR scores, wefila, studied in [25, 56, 57, 6]). Finally, using the implementation of these VIMs in several R packages (especially in the ‘*sensitivity*’ package), their numerical behavior on simulated and public datasets are highlighted. From these empirical studies, practical recommendations are derived. Every R script is also made available, from reproducibility purposes.

The structure of the paper is as follows. Section 2 reminds some basics about the multivariate linear regression model. Section 3 develops standard VIMs based on variance decomposition obtained with independent inputs. Section 4 introduces the effects and issues multicollinearity can bear on variance decomposition. Then, Section 5 develops several VIMs adapted to correlated inputs, obtained from allocation rules, while Section 6 presents the Johnson indices. Section 7 applies all the studied metrics on several simulated or public datasets. Finally, Section 8 provides a synthesis and draws some prospects regarding the current remaining challenges. Generalizations of these VIMs for classification tasks (i.e., logistic regression) are provided in the supplementary material. Table 1 (resp. Table 2) provides a table of acronyms (resp. notations) used all along the paper.

Table 1: Main acronyms.

CC	Correlation Coefficient (Pearson coefficient)
CI	Confidence Interval
LMG	Lindeman-Merenda-Gold indices (Shapley effects for linear models)
PCC	Partial Correlation Coefficient
PMVD	Proportional Marginal Variance Decomposition
RWA	Relative Weight Allocation (non-normalized Johnson indices)
GSA	Global Sensitivity Analysis
SPCC	Semi-Partial Correlation Coefficient
SRC	Standardized Regression Coefficient
SVD	Singular Value Decomposition
VIF	Variance Inflation Factor
VIM	Variance-based Importance Measure
VM	Variance-based Metrics

Table 2: Main notations.

$x_j$	$j$ -th deterministic variable (lowercase and italic)
$X_j$	$j$ -th random variable (uppercase and italic)
$\boldsymbol{x} := (x_1, \dots, x_d)$	Vector of deterministic variables (bold, italic and lowercase)
$\boldsymbol{X} := (X_1, \dots, X_d)$	Random vector (bold, italic and uppercase)
$x_j^{(i)}$	$i$ -th observation of the variable $x_j$
$\boldsymbol{x}^{(i)}$	$i$ -th observation of the vector $\boldsymbol{x}$
$\boldsymbol{X}^n := \left( x_1^{(i)}, \dots, x_d^{(i)} \right)_{i=1, \dots, n}$	$n$ -observation matrix (bold and capital letter)
$\widehat{\beta}$	Estimator of the parameter $\beta$ (circumflex accent)

## 2 Basics of multivariate linear regression

In this section, the multivariate linear regression framework is recalled. Consider an experimental design with  $n$  observations of an explained (real-valued) random variable  $Y$  (output) and of  $d$  explanatory random variables  $\boldsymbol{X} = (X_1, \dots, X_d)$  (inputs or regressors), denoted by:

$$(\boldsymbol{X}^n, \boldsymbol{y}^n) = \left( x_1^{(i)}, \dots, x_d^{(i)}, y^{(i)} \right)_{i=1, \dots, n}. \quad (1)$$

For simplicity and without any loss of generality, we use the following usual assumption.

**Assumption 1** (Centered inputs and output). *All inputs and the output are centered such that:*

$$\mathbb{E}[X_j] = 0 \text{ for } j = 1, \dots, d, \text{ and } \mathbb{E}[Y] = 0.$$

The relationship between the random inputs  $\boldsymbol{X}$  and the random output  $Y$  is modeled as being linear such that:

$$Y = \boldsymbol{X}\boldsymbol{\beta} + \varepsilon, \quad (2)$$

where  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_d)^\top \in \mathbb{R}^d$  is an unknown vector of coefficients, and  $\varepsilon$  is a random error assumed to be Gaussian and centered, i.e.,  $\varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2)$ , and such that  $\mathbb{E}[\varepsilon|\boldsymbol{X}] = 0$ . Specifically, for each observation  $\boldsymbol{x}^{(i)}$  of

$\mathbf{X}$  and  $y^{(i)}$  of  $Y$ , the previous relationship can be written as  $y^{(i)} = \mathbf{x}^{(i)}\beta + \varepsilon^{(i)}$ , where for all  $i = 1, \dots, n$ , the  $\varepsilon^{(i)}$ s are independent and identically distributed (i.i.d.) according to the same law as  $\varepsilon$ . We thus deduce that:

$$\mathbb{E} \left[ Y | \mathbf{X} = \left( x_1^{(i)}, \dots, x_d^{(i)} \right) \right] = \mathbf{x}^{(i)}\beta, \quad \text{for } i = 1, \dots, n.$$

If the sample size is large enough (i.e.,  $n \gg d$ ), and  $(\mathbf{X}^n)^\top \mathbf{X}^n$  is a positive-definite matrix, the Ordinary Least-Squares method (see, e.g., [12]) can be used to estimate the vector of parameters  $\beta$  by using the unbiased maximum likelihood estimator given by:

$$\widehat{\beta} = ((\mathbf{X}^n)^\top \mathbf{X}^n)^{-1} (\mathbf{X}^n)^\top \mathbf{y}^n. \quad (3)$$

Statistical techniques then allow for checking whether the use of a linear model is licit or not. An important goodness-of-fit metric is the *coefficient of determination*  $R^2$  which quantifies the percentage of output variability captured by the linear regression model. Its theoretical value is given by:

$$R^2 = R_{Y(X)}^2 := 1 - \frac{\mathbb{E} [\text{VAR}(Y|X)]}{\text{VAR}(Y)} = \frac{\text{VAR}(\mathbb{E}[Y|\mathbf{X}])}{\text{VAR}(Y)}. \quad (4)$$

Provided a consistent estimator  $\widehat{\beta}$  of  $\beta$ , one can build a plug-in consistent estimator of  $R^2$  based on the design matrix described by Eq. (1), leading to the following formula:

$$\widehat{R}^2 = \frac{\sum_{i=1}^n \left( \widehat{y}^{(i)} - \bar{y} \right)^2}{\sum_{i=1}^n \left( y^{(i)} - \bar{y} \right)^2}, \quad \text{where } \bar{y} = \frac{1}{n} \sum_{i=1}^n y^{(i)} \quad \text{and} \quad \widehat{y}^{(i)} = \mathbf{x}^{(i)} \widehat{\beta}.$$

**Remark 1.** *If the sample size is close to the number of inputs, there is a risk of overfitting. In that sense, an adjusted coefficient, such as  $R_{\text{adj}}^2 = 1 - |1 - R^2| \left| \frac{n-1}{n-(1+d)} \right|$  (see [40] for an overview of the various formulations of adjusted coefficients) can be used in order to penalize this dimension drawback. Moreover, cross-validation techniques can also be used to validate the regression model so as to avoid overfitting. It mainly consists in computing a predictivity coefficient  $Q^2$  based on a validation sample extracted from the learning sample using dedicated techniques [45] [18].*

### 3 Variance-based importance measures

Importance measures in regression models [16] broadly consist in quantifying the relative importance of the inputs to the output. In the field of GSA, importance measures are usually called *sensitivity indices*, and many different metrics (e.g., the variance, the entropy, a dependence measure or a dissimilarity measure between an input and the output) have been proposed to define them mathematically [52] [15]. A first approach consists in quantifying the amount of input uncertainty that creates dispersion in the output, the “dispersion” being traditionally quantified by the variance. Hence, the importance of an input can be naturally understood as the amount of uncertainty (i.e., in terms of variance) it brings to the system.

Besides GSA, and more generally in statistics, variance decomposition plays a central role in practical studies (e.g., in uncertainty analysis as illustrated in [42]), where it has been deemed to be an appropriate measure of information for a long time. In a nutshell, a *variance-based importance measure* (VIM) aims at quantifying the contribution of each input  $X_i$  to the variance of the output  $Y$ , denoted by  $\text{VAR}(Y)$ .

### 3.1 The variance decomposition

In the context of a multivariate linear regression model, the VIMs are based on the variance decomposition given by the law of total variance:

$$\text{VAR}(Y) = \underbrace{\text{VAR}(E[Y|\mathbf{X}])}_{\text{explained variance}} + \underbrace{E[\text{VAR}(Y|\mathbf{X})]}_{\text{residual variance}}, \quad (5)$$

which is valid for any real-valued random variable  $Y$ . The first term is usually called the *explained variance*, while the second term is usually interpreted as the *residual variance* which can be due to unaccounted inputs in the modelling, or to measurement errors. In particular, with Eq. (2), this decomposition gives:

$$\text{VAR}(E[Y|\mathbf{X}]) = \beta^\top \Sigma_{XX} \beta, \quad E[\text{VAR}(Y|\mathbf{X})] = \sigma_\varepsilon^2, \quad (6)$$

where  $\Sigma_{XX} = (\text{COV}(X_i, X_j))_{1 \leq i, j \leq d}$  is the covariance matrix of the inputs. Finally, one can notice the direct link between the explained variance and the theoretical definition of the  $R^2$  coefficient in Eq. (4), which is nothing more than a percentage of the total variance explained by the inputs.

### 3.2 Criteria for $R^2$ decomposition

As seen above, the  $R^2$  is directly linked to the notion of explained variance and to the variance decomposition (Eqs. (4) and (5)). Thus, historical developments of IMs in the literature of linear regression analysis naturally focused on partitioning the  $R^2$  among the  $d$  inputs [38, 24]. Many decomposition types have been proposed, leading to various  $R^2$  partitioning strategies (leading to various meanings). To sum up such a large panel of strategies, several authors defined some *desirability criteria* (i.e., properties) of what a “relevant decomposition” should be. For instance, according to [24], four basic desirability criteria can be sought after for a VIM resulting from an  $R^2$  decomposition:

- $(C_1)$  *Proper decomposition*: the sum of all shares should be equal to the total variance (or to the  $R^2$  itself in the case of normalized shares);
- $(C_2)$  *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.

Criteria  $(C_1)$  and  $(C_2)$  constitute the fundamental properties that VIMs should verify as they allow for a proper interpretation as a percentage of  $R^2$ . In addition, the criterion  $(C_4)$  seems fundamental to highlight inputs with direct influence in the model. The criterion  $(C_3)$  is also relevant in terms of its interpretation.

For the sake of completeness, one can mention an additional criterion that is sometimes mentioned in the literature, but more related to regularization-based techniques [61, 57]:

- $(C_5)$  *Grouping*: shares tend to equate for highly correlated inputs.

However, as it will be shown in Sections 5 and 7 for the VIMs that are considered in this paper, the grouping property  $(C_5)$  can be contradictory to the exclusion property  $(C_3)$ . Thus, the choice of a specific VIM should depend on the case of study and on the desired criteria. If the interpretation is focused on the direct influence of the inputs on the model output, then the exclusion property  $(C_3)$  seems to be appropriate; if the correlations among data can carry necessary information for the interpretation (as sometimes in GSA), it can be useful to consider the  $(C_5)$  property instead.

### 3.3 Regression coefficients and Pearson correlation for independent inputs

Provided that the inputs are independent, the law of total variance in Eq. (5) becomes:

$$\text{VAR}(Y) = \sum_{j=1}^d \beta_j^2 \sigma_j^2 + \sigma_\varepsilon^2,$$

and naturally allows to partition the output variance with respect to any input  $X_j$ , with  $j = 1, \dots, d$ , by means of a *standardized regression coefficient* (SRC)  $\beta_j^*$  defined as:

$$\beta_j^* = \beta_j \frac{\sigma_j}{\sigma_Y}.$$

where  $\sigma_Y$  and  $\sigma_j$  are the standard deviations associated with  $Y$  and the input  $X_j$ , respectively. Hence, the *squared SRC*  $\beta_j^{*2}$  can then be used as a VIM [23, 2]. It can be understood as the share of variance explained by each input  $X_j$ , since:

$$R^2 = \sum_{j=1}^d \beta_j^{*2}.$$

One can see that the squared SRC respect the four desirability criteria ( $C_1$ ), ( $C_2$ ), ( $C_3$ ) and ( $C_4$ ) mentioned previously. Moreover, one can notice that the SRC is strongly connected to the input-output *Pearson correlation coefficient* (CC), denoted by  $r_{Y,X_j}$ , which allows to measure the linear correlation between an input  $X_j$  and the output  $Y$ :

$$r_{Y,X_j} = \frac{\text{COV}(Y, X_j)}{\sigma_Y \sigma_j}.$$

In fact, for independent inputs, both quantities are equal, i.e.,  $r_{Y,X_j} = \beta_j^*$  and thus, one obtains:

$$R^2 = \sum_{j=1}^d r_{Y,X_j}^2. \quad (7)$$

**Remark 2.** As a reminder, if the input  $X_j$  admits a perfect linear relationship with the output  $Y$ ,  $r_{Y,X_j}$  is equal to 1 or  $-1$ . If  $X_j$  and  $Y$  are independent,  $r_{Y,X_j}$  is equal to 0. However, a  $r_{Y,X_j}$  equals to 0 does not imply that  $X_j$  and  $Y$  are independent as the dependency between  $X_j$  and  $Y$  might be nonlinear.

As soon as inputs are not independent anymore, the squared SRC is no longer an admissible VIM, since it does not take the contribution due to the covariance between the inputs of Eq. (6) into account. Thus, the VIM desirability criterion ( $C_1$ ) is not respected anymore. The following sections are dedicated to study alternatives which can be used when the independence is no more ensured.

## 4 Dealing with multicollinearity

In a regression setting, *multicollinearity* occurs whenever two or more inputs exhibit a statistically significant linear dependence. It generalizes the notion of *collinearity* [3] to encompass a linear link between more than two variables. Two variables  $X_1$  and  $X_2$  are said to be *perfectly collinear* if and only if the CC  $r_{X_1,X_2}$  is equal to 1 or  $-1$  (see Remark 2). Similarly, there is a *perfect multicollinearity* when there are two or more inputs

perfectly collinear. In practice, a perfect linear correlation almost never occurs and we speak of multicollinearity when there are several correlated variables with each other.

Several drawbacks can arise due to a high degree of multicollinearity. For instance, the least-square estimates of the linear coefficients can be impacted (this consequence is sometimes known as the “aliasing effect” [46]). Even if the matrix  $(\mathbf{X}^n)^\top \mathbf{X}^n$  appearing in Eq. (3) is theoretically invertible, a computer algorithm may be unsuccessful or inaccurate enough to obtain a precise approximation of the inverse matrix due to ill-conditioning. Several methods exist to circumvent this phenomenon, such as regularization (see, e.g., [17]).

Another issue can occur during the estimation the impact of an input variable on the output  $Y$ . The greater the multicollinearity effect, the more difficult it is to separate the individual effects of each variable on the output variable. Therefore, this section focuses on this difficulty by investigating several classic metrics proposed in the literature to deal with multicollinear inputs. However, as it will be shown, these metrics do not rely on the  $R^2$  decomposition and are, consequently, not able to separate the individual effects of each input variable on the output variable. In the following, these metrics will be called *variance-based metrics* (VM) so as to distinguish them from the VIM.

#### 4.1 An illustrative example: a two-input regression model

This subsection aims at providing a first simple example which will be used throughout the paper for illustration purposes of several metrics (and the corresponding properties).

**Example: two-input regression model.** Consider the linear regression model of the Eq. (2) (for  $d = 2$ ) of the output  $Y$  modeled by two inputs  $X_1$  and  $X_2$ . For the sake of simplicity, let us introduce the following notations:

$$b_1 := \beta_1 \sigma_1, \quad b_2 := \beta_2 \sigma_2, \quad \text{and } r := r_{X_1, X_2}.$$

From Eqs. (2) and (4), recalling that  $\text{COV}(X_1, X_2) = r \sigma_1 \sigma_2$ , one has:

$$R^2 = \frac{b_1^2 + 2b_1 b_2 r + b_2^2}{b_1^2 + 2b_1 b_2 r + b_2^2 + \sigma_\varepsilon^2}. \quad (8)$$

Similarly, we can easily determine the squared CCs:

$$r_{Y, X_1}^2 = \frac{(b_1 + r b_2)^2}{b_1^2 + 2b_1 b_2 r + b_2^2 + \sigma_\varepsilon^2} \quad \text{and} \quad r_{Y, X_2}^2 = \frac{(b_2 + r b_1)^2}{b_1^2 + 2b_1 b_2 r + b_2^2 + \sigma_\varepsilon^2}. \quad (9)$$

Both Eqs. (9) and (8) highlight the fact that, when the inputs are correlated (i.e.,  $r \neq 0$ ), the squared CCs do not satisfy the  $R^2$  decomposition as in the case of independent inputs given by Eq. (7). Therefore, squared CCs do not satisfy the criterion ( $C_1$ ). Moreover, assuming that  $r = 1$ , both squared CCs will be the same, even if either  $b_1$  or  $b_2$  are set to zero, which makes the criterion ( $C_3$ ) not fulfilled.

#### 4.2 Variance inflation factor

A standard and well-known metric of multicollinearity is the *variance inflation factor* (VIF) [21, 38] defined as:

$$\text{VIF}_j = \frac{1}{1 - R_{X_j}^2(\mathbf{X}_{-j})}, \quad (10)$$



where  $\mathbf{X}_{-j}$  is the vector of all the inputs except  $X_j$ , and where  $R_{X_j(\mathbf{X}_{-j})}^2$  represents the  $R^2$  from the linear regression where  $X_j$  is considered as the output, and by taking  $\mathbf{X}_{-j}$  as inputs. The smallest value of VIF is 1 and corresponds to the absence of high collinearity. A standard rule of thumb is that a VIF value exceeding 5 or 10 indicates a substantial amount of collinearity [36].

**Example: two-input regression model (Section 4.1 continued).**

From Eq. (10), one simply has  $\text{VIF}_1 = \text{VIF}_2 = \frac{1}{1-r^2}$  with  $r \neq \pm 1$ .

**Remark 3.** The generalized variance inflation factor (GVIF) has been proposed by [27] in order to provide a similar measure of multicollinearity as the VIF in the case of categorical inputs. The GVIF also works if one desires to group polynomial terms related to the same input.

### 4.3 The partial correlation coefficient

It can also be interesting to quantify the degree of association between the output  $Y$  and an input  $X_j$  by cancelling the effect of other inputs, gathered in  $\mathbf{X}_{-j}$ . It is in that spirit that the *partial correlation coefficient* (PCC) has been introduced [52]. It is defined as:

$$r_{(Y,X_j)|\mathbf{X}_{-j}} = r_{\varepsilon_{Y|\mathbf{X}_{-j}}, \varepsilon_{X_j|\mathbf{X}_{-j}}}, \quad (11)$$

where  $\varepsilon_{Y|\mathbf{X}_{-j}}$  (respectively  $\varepsilon_{X_j|\mathbf{X}_{-j}}$ ) represents the random error in the linear regression model of  $Y$  (respectively  $X_j$ ) with respect to  $\mathbf{X}_{-j}$ . In other words, the PCC measures the residual information of  $X_j$  on  $Y$  which is not explained by the variables  $\mathbf{X}_{-j}$ .

**Example: two-input regression model (Section 4.1 continued).**

Eq. (11) can be written as a function of the coefficient of determination and the CC such as:

$$r_{(Y,X_1)|X_2}^2 = \frac{R_{Y(X_1,X_2)}^2 - r_{Y,X_2}^2}{1 - r_{Y,X_2}^2},$$

and using Eqs. (8) and (9):

$$r_{(Y,X_1)|X_2}^2 = \frac{b_1^2(1-r^2)}{b_1^2(1-r^2) + \sigma_\varepsilon^2} \quad \text{and} \quad r_{(Y,X_2)|X_1}^2 = \frac{b_2^2(1-r^2)}{b_2^2(1-r^2) + \sigma_\varepsilon^2}.$$

Note that the squared PCC is equal to 1 if the model is perfectly linear (i.e., if  $\sigma_\varepsilon^2 = 0$  with  $b_j \neq 0$ ) and equal to zero if  $X_1$  and  $X_2$  are perfectly correlated. Thus, even if it respects the *exclusion* criterion ( $C_3$ ), it does not respect the *inclusion* criterion ( $C_4$ ). Finally, even if in the GSA literature [52, 26, 34], the PCC has been proposed as a substitute for the SRC in the case of dependent inputs, it does not respect the fundamental desirability criterion ( $C_1$ ), i.e., the proper  $R^2$  decomposition.

### 4.4 The semi-partial correlation coefficient

Instead of controlling the potential linear effects of  $\mathbf{X}_{-j}$  with  $X_j$ , as done with the PCC, the *semi-partial correlation coefficient* (SPCC) quantifies the additional explanatory power of a variable  $X_j$  on the variance of  $Y$  [38]. SPCC is defined as the proportion of the output variance explained by  $X_j$  after removing the “information

brought” by  $X_{-j}$  (as a difference of explained variance). It is formally given by the CC (noted  $r_{Y,(X_j|X_{-j})}$ ) between  $Y$  and the residuals of the regression of  $X_j$  on  $X_{-j}$ . The *squared SPCC* is intrinsically linked to the  $R^2$  since it can be written as:

$$r_{Y,(X_j|X_{-j})}^2 = R_{Y(X)}^2 - R_{Y(X_{-j})}^2. \quad (12)$$

In the case of independent inputs, the SPCC is equal to the usual CC.

**Example: two-input regression model (Section 4.1 continued).**

Eq. (12) can be written as a function of the coefficient of determination and the CC such as:

$$r_{Y,(X_1|X_2)}^2 = R_{Y(X_1,X_2)}^2 - r_{Y,X_2}^2.$$

and using Eqs. (8) and (9):

$$r_{Y,(X_1|X_2)}^2 = \frac{b_1^2(1-r^2)}{b_1^2 + 2b_1b_2r + b_2^2 + \sigma_\varepsilon^2} \quad \text{and} \quad r_{Y,(X_2|X_1)}^2 = \frac{b_2^2(1-r^2)}{b_1^2 + 2b_1b_2r + b_2^2 + \sigma_\varepsilon^2}. \quad (13)$$

In [22], the squared SPCC is called the *marginal reduction* due to  $X_j$ . This comes from the fact that it quantifies the *loss of  $R^2$*  induced by removing  $X_j$  from the linear model.

On the other hand, as illustrated in the above example, the SPCC can become artificially small in situations of highly correlated inputs, which subsequently renders any importance ranking task quite difficult. Moreover, one can notice that the squared SPCC is not an admissible VIM, since it does not respect the fundamental VIM desirability criterion ( $C_1$ ).

## 4.5 Illustration with the Venn diagrams

In order to provide an intuitive understanding of the multicollinearity, we propose to use the Venn diagrams in the case of a standard linear regression with two inputs [13, 29].

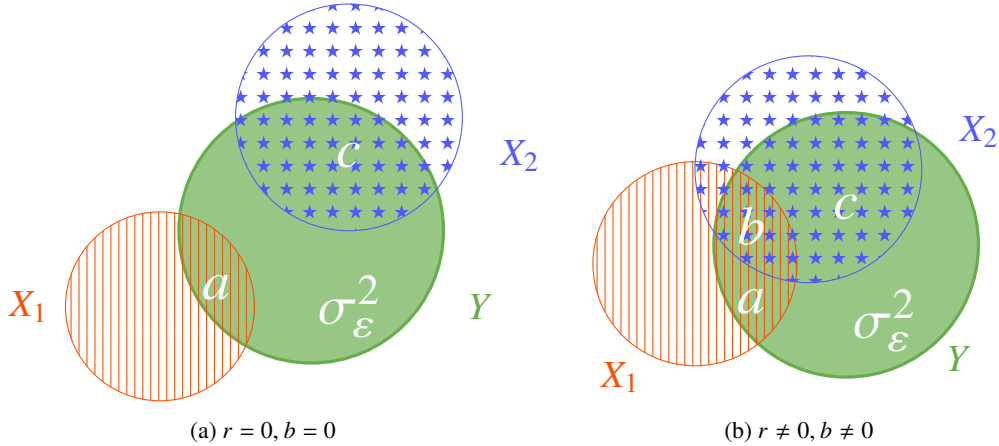


Figure 1: Illustration of the multicollinearity effects with an output  $Y$  and two inputs  $X_1$  and  $X_2$ .

Figure 1 is to be understood as follows: in both sub-figures, the total variance of  $Y$  is represented as the green area by  $a + b + c + \sigma_\varepsilon^2$  with  $\sigma_\varepsilon^2$  the unexplained share of variance (i.e., the model error). The orange area represents the variance of  $X_1$ , while the blue area represents the variance of  $X_2$ . The area  $a$  (resp.  $c$ ) represents the additional explanatory power of the variable  $X_1$  (resp.  $X_2$ ) on the regression model  $Y(X)$  (defined by Eq. (2)) given by the nominator of the squared SPCC (Eqs. (13)). We thus can write that:

$$\begin{aligned} a &= b_1^2(1 - r^2), \\ c &= b_2^2(1 - r^2), \\ b &= b_1^2 r^2 + 2b_1 b_2 r + b_2^2 r^2. \end{aligned} \quad (14)$$

In the Figure 1a the variables  $X_1$  and  $X_2$  are independent; the orange and the blue areas do not overlap ( $r = 0, b = 0$ ). In this case, the squared CCs (Eq. (9)) and the squared SPCCs (Eq. (13)) are equal:

$$r_{Y,(X_1|X_2)}^2 = r_{Y,X_1}^2 = \frac{a}{a + c + \sigma_\varepsilon^2} \quad \text{and} \quad r_{Y,(X_2|X_1)}^2 = r_{Y,X_2}^2 = \frac{c}{a + c + \sigma_\varepsilon^2} \quad (\text{in Figure 1a}).$$

The orange area  $a$  and the blue area  $c$  finally represent the proportion of the variance in  $Y$  respectively explained by  $X_1$  and  $X_2$ , and allow sharing the determination coefficient  $R^2$ :

$$R^2 = \frac{a + c}{a + c + \sigma_\varepsilon^2} \quad (\text{in Figure 1a}).$$

In the Figure 1b the variables  $X_1$  and  $X_2$  are correlated; the orange and the blue areas overlap ( $r \neq 0, b \neq 0$ ). The proportion of the variance in  $Y$  explained by  $X_1$  (resp.  $X_2$ ) is now equal to  $a + b$  (resp.  $c + b$ ). The squared CCs (Eq. (9)) are thus equal to :

$$r_{Y,X_1}^2 = \frac{a + b}{a + b + c + \sigma_\varepsilon^2} \quad \text{and} \quad r_{Y,X_2}^2 = \frac{c + b}{a + b + c + \sigma_\varepsilon^2},$$

and the squared SPCCs (Eq. (13)) are equal to :

$$r_{Y,(X_1|X_2)}^2 = \frac{a}{a + b + c + \sigma_\varepsilon^2} \quad \text{and} \quad r_{Y,(X_2|X_1)}^2 = \frac{c}{a + b + c + \sigma_\varepsilon^2}.$$

We can notice that the sum of the squared CCs or the sum of the squared SPCCs is not equal to the determination coefficient  $R^2$ :

$$R^2 = \frac{a + b + c}{a + b + c + \sigma_\varepsilon^2}.$$

Therefore, the squared CCs and the squared SPCCs cannot be used as VIM knowing that they do not meet the criterion  $C_1$  for the  $R^2$  decomposition.

Finally, the notion of multicollinearity can be appreciated as follows: the higher (in absolute) the Pearson CC  $r$  between two variables are, the higher the overlap area  $b$  is. Similarly, the areas  $a$  and  $c$  are getting smaller (see Eq. (14)) and the squared CCs and the squared SPCCs do not meet the criterion  $C_1$ . That is why the presence of multicollinearity effect makes the  $R^2$  decomposition difficult and it is necessary to use more complex VIMs than the classic VMs previously presented to separate the individual effects of each variable on the output variable.

## 5 Importance measures from allocation rules

In the context of statistical learning, an interesting approach is to define  $R^2$  decomposition by exhibiting hierarchy between the inputs among predictors (i.e., inputs) regarding some *dominance criteria* [10]. This is known as the *general dominance analysis*. This approach is analogous to the definition of allocations in the field of cooperative game theory. In this section, the most interesting developments around this idea are exhibited.

### 5.1 Lindeman-Merenda-Gold indices

One particular VIM arising from general dominance analysis is the LMG indices (acronym coming from the initials of the authors' names, i.e., "Lindeman-Merenda-Gold" [44]). These VIMs have been studied extensively (see, [10, 23]). They are based on the averaging sequential sums of squares over all orderings of inputs.

Formally, let  $u$  denote a subset of indices in the set of all subsets of  $\{1, \dots, d\}$  and  $\mathbf{X}_u = (X_j : j \in u)$  represents a subset of inputs. Dominance analysis is based on the measure of the elementary contribution of any given variable  $X_j$  to a given subset model  $Y(\mathbf{X}_u)$  by the increase in  $R^2$  that results from adding that predictive variable to the regression model:

$$\text{LMG}_j = \frac{1}{d!} \sum_{\substack{\pi \in \text{permutations} \\ \text{of } \{1, \dots, d\}}} r_{Y, (X_j | \mathbf{X}_\pi)}^2 \quad (15)$$

where the squared SPCC  $r_{Y, (X_j | \mathbf{X}_\pi)}^2 = R_{Y(\mathbf{X}_{v \cup \{j\}})}^2 - R_{Y(\mathbf{X}_v)}^2$  is to be understood with  $v$  being the indices preceding  $j$  in the order  $\pi$ . An equivalent formula is given by:

$$\text{LMG}_j = \frac{1}{d} \sum_{i=0}^{d-1} \sum_{\substack{u \subseteq -\{j\} \\ |u|=i}} \binom{d-1}{i}^{-1} r_{Y, (X_j | \mathbf{X}_u)}^2 = \frac{1}{d} \sum_{u \subseteq -\{j\}} \binom{d-1}{|u|}^{-1} r_{Y, (X_j | \mathbf{X}_u)}^2 \quad (16)$$

with  $\binom{n}{k} = \frac{n!}{(n-k)!k!}$  and  $r_{Y, (X_j | \mathbf{X}_u)}^2 = R_{Y(\mathbf{X}_{u \cup \{j\}})}^2 - R_{Y(\mathbf{X}_u)}^2$ .

In Eq. (16) (resp. Eq. (15)), this averaging process over all combinations (resp. permutations) is carried out in the absence of order between the inputs. This VIM has been extensively studied in the literature (see, e.g., [41] and [22]). The main drawback in regards to its broad utilization in practice is its exponential complexity (i.e., one needs to perform  $2^d - 1$  different linear regressions to compute the summands in Eq. (16)), which can be challenging even for moderate size  $d$ .

**Example: two-input regression model (Section 4.1, continued).**

Eq. (15) becomes:

$$\text{LMG}_1 = \frac{1}{2} \left( R_{Y(X_1, X_2)}^2 - R_{Y(X_2)}^2 + R_{Y(X_1)}^2 \right), \quad \text{LMG}_2 = \frac{1}{2} \left( R_{Y(X_1, X_2)}^2 - R_{Y(X_1)}^2 + R_{Y(X_2)}^2 \right),$$

and using Eqs. (8) and (9):

$$\text{LMG}_1 = \frac{b_1^2 + b_1 b_2 r + \frac{r^2}{2} (b_2^2 - b_1^2)}{b_1^2 + 2b_1 b_2 r + b_2^2 + \sigma_\varepsilon^2}, \quad \text{LMG}_2 = \frac{b_2^2 + b_1 b_2 r + \frac{r^2}{2} (b_1^2 - b_2^2)}{b_1^2 + 2b_1 b_2 r + b_2^2 + \sigma_\varepsilon^2}. \quad (17)$$

This result is also given in [24].

Going back to the Venn diagram illustration, (see Fig. 1), one has [29]:

$$\begin{aligned} \text{LMG}_1 &= (a + b/2)/(a + b + c + \sigma_\varepsilon^2), \\ \text{LMG}_2 &= (c + b/2)/(a + b + c + \sigma_\varepsilon^2). \end{aligned}$$

Focusing on the numerators, one can notice that the LMG redistributes  $b$  equally between  $X_1$  and  $X_2$  (each variable gets half of the variance due to their correlation).

Note also what happens in the two following particular cases:

- If  $|r|$  tends to 1 (i.e., the inputs are collinear),  $\text{LMG}_1$  and  $\text{LMG}_2$  tends to be both equal to 0.5. The grouping property (criterion  $(C_5)$  of Section 3.2) is respected.
- If one input is not in the model, for example  $X_2$  (then  $\beta_2 = 0$  and  $b_2 = c = 0$ ), its LMG cannot be zero as long as it is correlated with  $X_1$ . The exclusion property  $(C_3)$  (see, Section 3.2) is thus not respected.

In conclusion, the LMG index respects the fundamental VIM desirability criteria  $(C_1)$  ( $R^2$  decomposition) and  $(C_2)$  (positivity). Moreover, as stated by [20] and [24], it respects criteria  $(C_4)$  (inclusion) and  $(C_5)$  (grouping), but not  $(C_3)$  (exclusion).

## 5.2 The proportional marginal variance decomposition

By analogy with the LMG indices, [20] proposed the *proportional marginal variance decomposition* (PMVD). They also make use of sequential sum of squares, but differ from the LMG on the averaging process over the different orderings of inputs. These indices have been extensively studied in [24, 25]. The PMVD indices are defined as follows:

$$\text{PMVD}_j = \sum_{\substack{\pi \in \text{permutations} \\ \text{of } \{1, \dots, d\}}} \frac{L(\pi)}{\sum_{\pi} L(\pi)} r_{Y, (X_j | \mathbf{X}_\pi)}^2, \quad (18)$$

where:

$$L(\pi) = \prod_{i=1}^{d-1} \left[ r_{Y, (\mathbf{X}_{\pi_{i+1}, \dots, \pi_d} | \mathbf{X}_{\pi_1, \dots, \pi_i})}^2 \right]^{-1}.$$

**Example: two-input regression model (Section 4.1, continued).**

Eq. (18) becomes [24]:

$$\text{PMVD}_1 = \frac{b_1^2 + b_1^2/(b_1^2 + b_2^2)2b_1b_2r}{b_1^2 + 2b_1b_2r + b_2^2 + \sigma_\varepsilon^2}, \quad \text{PMVD}_2 = \frac{b_2^2 + b_2^2/(b_1^2 + b_2^2)2b_1b_2r}{b_1^2 + 2b_1b_2r + b_2^2 + \sigma_\varepsilon^2}.$$

Moreover, one can notice that:

- If we put  $|r| = 1$ , we obtain  $\text{PMVD}_1 = b_1^2/(b_1^2 + b_2^2)$  and  $\text{PMVD}_2 = b_2^2/(b_1^2 + b_2^2)$ . These two values can be strongly different in cases of large differences between  $b_1$  and  $b_2$ . This shows that the grouping property (criterion  $(C_5)$  of Section 3.2) is not respected.
- If one input is not in the model, for example,  $X_2$ , then  $\beta_2 = 0$  and  $b_2 = c = 0$ , and subsequently, its PMVD is equal to zero. Therefore, the exclusion property  $(C_3)$  is respected.

- If  $\sigma_\varepsilon^2 = 0$ , the equations simplify to:

$$PMVD_1 = \frac{b_1^2}{b_1^2 + b_2^2}, \quad PMVD_2 = \frac{b_2^2}{b_1^2 + b_2^2}.$$

Going back to the Venn diagram analogy (see Fig. 1), one has (see also [28]):

$$\begin{aligned} PMVD_1 &= a[1 + b/(a + c)]/(a + b + c + \sigma_\varepsilon^2), \\ PMVD_2 &= c[1 + b/(a + c)]/(a + b + c + \sigma_\varepsilon^2). \end{aligned}$$

In this case, the share  $b$  due to the correlation between inputs is not equally shared, as for the LMG indices, but rather “proportionally” shared with respect to the magnitude of the shares  $a$  and  $c$ . In the particular case where  $\sigma_\varepsilon^2 = 0$ , the above equations simplify to:

$$PMVD_1 = a/(a + c), \quad PMVD_2 = c/(a + c).$$

and one can notice that the PMVD does not depend on  $b$  anymore. While this behavior is known when dealing with two inputs, [24] shows that it does not generalize to situations with more inputs.

In conclusion, the PMVD respects the fundamental VIM desirability criteria ( $C_1$ ) ( $R^2$  decomposition) and ( $C_2$ ) (positivity). Moreover, as stated by [20] and [24], it respects criteria ( $C_3$ ) (exclusion) and ( $C_4$ ) (inclusion).

### 5.3 Synthesis

Table 3 synthesizes the analytical expressions of the discussed VIMs based on the illustration of Figure 1. The equations for  $CC^2$ ,  $PCC^2$  and  $SPCC^2$  are displayed as functions of  $a$ ,  $b$ ,  $c$  (using Eqs. (14)). As seen in this case,  $CC^2$ ,  $PCC^2$  and  $SPCC^2$  are not admissible VIMs because they do not sum to  $R^2$ . Contrarily, LMG and PMVD are admissible. Moreover, LMG does not respect the exclusion property but respects the inclusion property, while the PMVD respects both. The behavior of these indices is illustrated and studied on more general examples and on real datasets in Section 7.

Input	$CC^2$	$PCC^2$	$SPCC^2$	LMG	PMVD
$X_1$	$\frac{a + b}{a + b + c + \sigma_\varepsilon^2}$	$\frac{a}{a + \sigma_\varepsilon^2}$	$\frac{a}{a + b + c + \sigma_\varepsilon^2}$	$\frac{a + \frac{1}{2}b}{a + b + c + \sigma_\varepsilon^2}$	$\frac{a + \frac{a}{a+c}b}{a + b + c + \sigma_\varepsilon^2}$
$X_2$	$\frac{c + b}{a + b + c + \sigma_\varepsilon^2}$	$\frac{c}{c + \sigma_\varepsilon^2}$	$\frac{c}{a + b + c + \sigma_\varepsilon^2}$	$\frac{c + \frac{1}{2}b}{a + b + c + \sigma_\varepsilon^2}$	$\frac{c + \frac{c}{a+c}b}{a + b + c + \sigma_\varepsilon^2}$

Table 3: Different VM associated with the decomposition of  $R^2 = (a + b + c)/(a + b + c + \sigma_\varepsilon^2)$ .

**Remark 4.** In [30], estimation schemes for LMG and PMVD have been proposed for both linear and logistic models and applied to regression and classification tasks, respectively. The supplementary material presents results for logistic regression.

## 5.4 Links with game theory and global sensitivity analysis

The two above-presented VIMs (LMG and PMVD) are inherently linked with cooperative game theory. The sequential approach (i.e., the formulations using permutations) is related to the notion of *random order allocation* by [58] and [20]. These allocations (also called solution concepts) allow decomposing a quantity (in this case, the  $R^2$ ) by means of quantifying the “value” of each player using a value function (here, the square SPCC). Through this lens, the LMG indices are none other than the so-called Shapley values of the cooperative game [53]. It is well known that this value is egalitarian in its redistribution (i.e., the behavior of splitting  $b$  is half in the Venn diagram analogy actually hold in higher dimensions). On the other hand, the PMVD is analogous to the *proportional values* [19], allowing for a proportional redistribution.

In GSA, when dealing with a linear numerical model, the only difference with the present study is the fact that  $\sigma_\varepsilon^2$  is equal to 0 [52, 26]. GSA actually encompasses the definition of VIMs of more-general models (i.e., not necessarily linear). For instance, whenever the inputs are assumed to be independent, the SRC<sup>2</sup> is actually equal to the *first-order Sobol’ index*, which is defined outside of the realm of linear models [54]. Additionally, provided that the error is null, the  $R^2$  can be directly comparable to the closed Sobol’ indices, which need not be restricted to linear models to be defined.

The use of the Shapley values to define sensitivity indices for variance-based GSA has been recently introduced [49], where chosen value function is the closed Sobol’ index of a subset of player, leading to the *Shapley effects*. Several authors [55, 4, 35, 51] proposed and studied several dedicated estimation algorithms for nonlinear models. Analytical formulas have also been exhibited for linear models with Gaussian inputs [50], and can be efficiently computed by finely tuned algorithms [9]. However, these algorithms require the knowledge and the ability to draw randomly from the joint density of the inputs. Thus, one needs to know how to model the dependence structure (i.e., the copula) between the inputs. Typically, such a condition is not met in common statistical learning (or machine learning) practice and the estimation of such VIMs often appears to be difficult (either because of its complexity, to the input dimension or since one only has limited data). Recently, given-data algorithms have been proposed to leverage this issue [8, 11].

It has also been noticed that, theoretically, the Shapley effects can grant *exogenous inputs* (i.e., which are not explicitly included in the structural equations of the model) some importance, especially when these inputs are correlated to *endogenous inputs* (i.e., effectively present in the model). Inspired by the PMVD, [27] proposed to use the proportional values instead of the Shapley values. It lead to novel GSA indices, called *proportional marginal effects* (PME). These indices allow the detection of exogenous inputs, despite the correlation, in a nonlinear setting (this is analogous to the exclusion criterion).

## 6 Dealing with high-dimensional inputs via the relative weight allocations

When  $d$  is large (e.g., several tens), interesting VIM, coming from a singular value decomposition (SVD), in order to transform the correlated inputs into uncorrelated variables and an appropriate reweighing process, have allowed some authors to propose the so-called *relative weight measures* [37] and called later *Johnson’s relative weights* or *Johnson indices*. Note that this approach has been proposed earlier by several authors (see associated references in [48] and [25]). As an example, one can mention the work of [22] which led to the so-called *Genizi’s approach*. All in all, these VIM based on a preliminary transformation of inputs are known to be adapted to large input dimension as well as providing similar results to those obtained via LMG indices [38, 13, 14], at a highly reduced computational cost. In this section, we explain how to get the Johnson indices.

The Johnson indices are part of a package of methods based on a *Relative Weights Allocations* (RWA). The

idea of these methods is to find an orthogonal matrix  $\mathbf{Z}^n$  of the space generated by the column vectors of  $\mathbf{X}^n$  (the  $n$ -size sample of the  $d$  inputs  $\mathbf{X} = (X_1, \dots, X_d)$ ).

## 6.1 Johnson indices

In the case of the Johnson indices ([39] and [37])<sup>1</sup>, the matrix  $\mathbf{X}^n \in \mathbb{R}^{n \times d}$  of the design of experiments is transformed in an orthogonal matrix  $\mathbf{Z}^n \in \mathbb{R}^{n \times d}$  in the least square sense. Figure 2 summarizes the global approach of the Johnson indices. For [39], it consists in finding  $\mathbf{Z}^n$  and  $\mathbf{W} \in \mathbb{R}^{d \times d}$  such as:

$$\begin{cases} \mathbf{X}^n = \mathbf{Z}^n \mathbf{W} \\ (\mathbf{Z}^n)^\top \mathbf{Z}^n = \mathbf{I} \\ \mathbf{Z}^n = \arg \min_{\mathbf{\Pi}^n} \text{Tr}(\mathbf{X}^n - \mathbf{\Pi}^n)^\top (\mathbf{X}^n - \mathbf{\Pi}^n) \end{cases} \quad (19)$$

where  $\mathbf{I} \in \mathbb{R}^{d \times d}$  is the identity matrix.

Johnson shows that the solution of Eq. (19) is:

$$\mathbf{Z}^n = \mathbf{P}^n \mathbf{Q}^\top \text{ and } \mathbf{W} = \mathbf{Q} \mathbf{\Delta} \mathbf{Q}^\top. \quad (20)$$

$\mathbf{P}^n \in \mathbb{R}^{n \times d}$  and  $\mathbf{Q} \in \mathbb{R}^{d \times d}$  are the matrices defined by the singular value decomposition:

$$\mathbf{X}^n = \mathbf{P}^n \mathbf{\Delta} \mathbf{Q}^\top, \quad (21)$$

which contains respectively the eigenvectors of  $\mathbf{X}^n \mathbf{X}^{n\top}$  and  $\mathbf{X}^{n\top} \mathbf{X}^n$ , and  $\mathbf{\Delta} \in \mathbb{R}^{d \times d}$  a diagonal matrix which itself contains the singular values of  $\mathbf{X}^n$  such as the singular values  $\delta_1 \geq \dots \geq \delta_d > 0$ . In that sense, the new set of uncorrelated variables  $z_1, \dots, z_d$  is maximally correlated with the original set of correlated variables  $x_1, \dots, x_d$  (the columns of  $\mathbf{X}^n$ ).

**Remark 5.** Note also that Eq. (20) gives:

$$\Sigma_{XX} = \mathbf{W}^2. \quad (22)$$

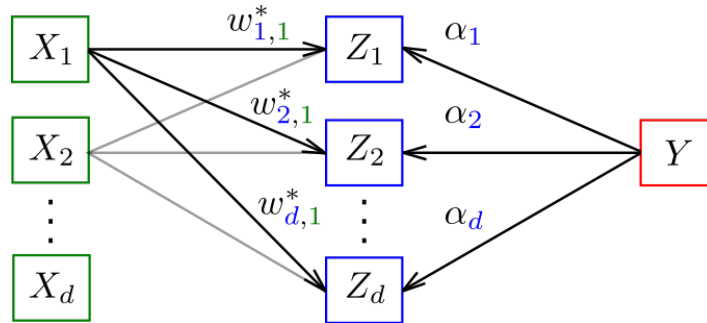


Figure 2: Representation of the Johnson relative weight calculation associated with the input  $X_1$ .

A first least square regression of  $\mathbf{y}^n$  ( $n$ -size sample of the variable  $Y \in \mathbb{R}$ ) on  $\mathbf{Z}^n$  allows determining the vector  $\boldsymbol{\alpha} \in \mathbb{R}^d$ :

$$\hat{\boldsymbol{\alpha}} = ((\mathbf{Z}^n)^\top \mathbf{Z}^n)^{-1} (\mathbf{Z}^n)^\top \mathbf{y}^n = (\mathbf{Z}^n)^\top \mathbf{y}^n = (\hat{\alpha}_j)_{1 \leq j \leq d}.$$

<sup>1</sup>Note, that there are two different authors with the same name. [37] suggested determining the matrix  $\mathbf{W}$  as the weights of the regression of  $\mathbf{X}^n$  on  $\mathbf{Z}^n$  contrary to [39] which regressed  $\mathbf{Z}^n$  on  $\mathbf{X}^n$ .



Because the new transformed predictors  $z_j$  are uncorrelated to one another, the predictable variance of  $Y$  can be decomposed such as:

$$\text{VARE}(Y|\mathbf{Z}) = \sum_{j=1}^d \alpha_j^2. \quad (23)$$

The  $\alpha_j^2$ 's are considered to be close approximations to the relative weights of the original set of correlated variables  $x_1, \dots, x_d$ , but they do not give close representations, particularly if some original variables are highly correlated. To take into account the correlation effects, [37] thus suggests computing the regression coefficients of  $\mathbf{X}^n$  on  $\mathbf{Z}^n$ .

**Remark 6.** Using Eqs. (3), Eq. (19) and (22), note also that:

$$\hat{\alpha} = \mathbf{W}\hat{\beta}. \quad (24)$$

The  $d$  linear combinations between  $\mathbf{X}^n$  and  $\mathbf{Z}^n$  therefore allows determining the matrix of the *weights*  $\mathbf{W}$ :

$$\hat{\mathbf{W}} = (\mathbf{Z}^n)^\top \mathbf{X}^n = (\hat{w}_{ij})_{1 \leq i, j \leq d}.$$

Using Eq. (20), it can be shown that the standardized matrix  $\hat{\mathbf{W}}^*$  is composed of an estimation of the CCs  $r_{Z_i, X_j}$ :

$$\hat{w}_{ij}^* = \frac{\hat{w}_{ij}}{\sqrt{\sum_k \hat{w}_{kj}^2}} = \hat{r}_{Z_i, X_j}, \quad (25)$$

and thus for all  $j$ :

$$\sum_{i=1}^d (\hat{w}_{ij}^*)^2 = 1. \quad (26)$$

$w_{ij}^*$  therefore represents the proportion of variance in  $Z_i$  accounted by  $X_j$ .

Finally, the proportionate contribution of  $X_j$  to  $Y$  can then be estimated by multiplying the proportion  $\hat{\alpha}_i^2$  of variance in  $Y$  accounted for by  $Z_i$  by the proportion  $(\hat{w}_{ij}^*)^2$  of each  $Z_i$  accounted for by  $X_j$ . The Johnson index associated with the input  $X_j$  can thus be expressed as:

$$J_j = \sigma_Y^{-2} \sum_{i=1}^d \alpha_i^2 w_{ij}^{*2}. \quad (27)$$

## 6.2 Standardized Johnson indices for the variance decomposition

As discussed in the previous section 3.1, the  $R^2$  decomposition ( $C_1$ ) (linked to the variance decomposition) is a fundamental property that VIM must have.

As presented in the Section 2, the  $R^2$  can be decomposed and estimated thanks to the covariance matrices  $\hat{\Sigma}_{Y,X}$  and  $\hat{\Sigma}_{X,X}$  (see [23]):

$$\hat{R}^2 = \hat{\sigma}_Y^{-2} \hat{\Sigma}_{Y,X} \hat{\Sigma}_{X,X}^{-1} \hat{\Sigma}_{X,Y},$$

and using  $\hat{\Sigma}_{X,Y} = \hat{\Sigma}_{X,X} \hat{\beta}$ , the latter equation gives:

$$\hat{R}^2 = \hat{\sigma}_Y^{-2} \hat{\beta}' \hat{\Sigma}_{X,X} \hat{\beta}. \quad (28)$$

Using Eqs. (22) and (24), Eq. (28) thus gives the decomposition<sup>2</sup>:

$$\widehat{R}^2 = \widehat{\sigma}_Y^{-2} \widehat{\alpha}' \widehat{\alpha}.$$

In the paper of [37], it is quickly said that the input samples are *expressed in standard score form*. Using Eqs. (25) and (22), the standardization of the predictors implied that  $w_{ij}^* = w_{ij}$  and by the symmetry of  $W$ ,  $\sum_{i=1}^d w_{ij}^{*2} = \sum_{j=1}^d w_{ij}^{*2} = 1$ . The sum of the  $d$  relative weights  $\sum_{i=1}^d \alpha_i^2 w_{ij}^{*2}$  thus forms the variance decomposition of Eq. (23) and finally:

$$\sum_{j=1}^d J_j = \sigma_Y^{-2} \alpha' \alpha.$$

With Eq. (6.2), the standardization of the inputs thus gives:

$$\widehat{R}^2 = \sum_{j=1}^d \widehat{J}_j. \quad (29)$$

Finally, it is important to note that the standardization of the inputs is equivalent to directly calculate the matrix  $W^*$  and  $\alpha^*$  thanks to the multivariate correlation matrices  $R_{XX}$  and  $R_{XY}$ , as in the initial paper of [39]. The eigen decomposition of the correlation matrix  $R_{XX}$  gives:

$$R_{XX} = \mathbf{Q}^* \Lambda^{*2} \mathbf{Q}^{*\top}.$$

The matrix  $W^*$  is then given (similarly to the Eq. (20)) by<sup>3</sup>:

$$W^* = \mathbf{Q}^* \Lambda^* \mathbf{Q}^{*\top},$$

and the vector  $\alpha^*$  is determined thanks to the relation:

$$\alpha^* = W^{*-1} R_{XY}.$$

**Remark 7.** As previously with the Eq. (24), we can also write:

$$\alpha^* = W^* \beta^*, \quad (30)$$

with  $\beta^*$  the standardized coefficient presented in Section 3.3

The *standardized Johnson index* associated with the input  $X_j$  is directly given by:

$$J_j^{*2} = \sum_{i=1}^d \alpha_i^{*2} w_{ij}^{*2}.$$

The standardized Johnson index respects the fundamental VIM desirability criteria ( $C_2$ ) (positivity) and ( $C_1$ ) ( $R^2$  decomposition). Moreover, as LMG, it respects criteria ( $C_4$ ) (inclusion) and ( $C_5$ ) (grouping), but not ( $C_3$ ) (exclusion). Indeed, Eq. (27) intuitively shows that the correlation structure of the inputs carried by  $W^*$  is distributed over the Johnson indices. The similar behavior between the LMG and Johnson indices is confirmed in the next section which shows their strict equality in the two-dimensional case.

<sup>2</sup>By construction the vector  $\alpha$  and  $\beta$  are associated with the same quadratic minimization problem of the function  $S(\beta) = \|\mathbf{y}^n - X^n \beta\|^2 = \|\mathbf{y}^n - P^n Q' Q \Delta Q' \beta\|^2 = \|\mathbf{y}^n - Z^n W \beta\|^2 = \|\mathbf{y}^n - Z^n \alpha\|^2 = S(\alpha)$ .

<sup>3</sup>As a reminder, in [39],  $W^* = \mathbf{Q}^* \Lambda^{*-1} \mathbf{Q}^{*\top}$  because  $Z^n$  is regressed on  $X^n$ .

### 6.3 Equivalence between the LMG measures and the standardized Johnson indices for the case of two variables

In this section, we show, in a particular case, the first theoretical result (up to our knowledge) showing the equivalence between the LMG and the standardized Johnson indices.

**Proposition 1.** *If  $d = 2$  (as in Section 4.1), the LMG and the standardized Johnson indices are equal:*

$$J_j^{*2} = \text{LMG}_j \text{ for } j = 1, 2.$$

*Proof.* The correlation matrix  $R_{XX}$  is given by:

$$R_{XX} = \mathbf{W}^{*2} = \begin{pmatrix} w_{11}^{*2} + w_{12}^{*2} = 1 & w_{12}^{*2}(w_{11}^{*2} + w_{22}^{*2}) \\ w_{12}^{*2}(w_{11}^{*2} + w_{22}^{*2}) & w_{12}^{*2} + w_{22}^{*2} = 1 \end{pmatrix}.$$

The standardized Johnson index associated with the input  $X_1$  (resp.  $X_2$ ) is given according to the Eq. (6.2) by:

$$J_1^{*2} = \left[ \alpha_1^{*2} w_{11}^{*2} + \alpha_2^{*2} w_{21}^{*2} \right],$$

with  $\alpha_i^* = \beta_1^* w_{i1}^* + \beta_2^* w_{i2}^*$  for  $i \in \{1, 2\}$ . We then have:

$$J_1^{*2} = \left[ (\beta_1^* w_{11}^* + \beta_2^* w_{12}^*)^2 w_{11}^{*2} + (\beta_1^* w_{21}^* + \beta_2^* w_{22}^*)^2 w_{21}^{*2} \right]. \quad (31)$$

Because the singular values involved in Eq. (21) are positive, the diagonal elements of  $W$  are also positive. Using Eq. (26), we thus have  $w_{11}^* = w_{22}^* = \sqrt{1 - w_{12}^{*2}}$  and after several simplifications, Eq. (31) becomes:

$$J_1 = \left[ \beta_1^{*2} + 2\beta_1^* \beta_2^* w_{12}^* w_{11}^* + 2w_{11}^{*2} w_{12}^{*2} (\beta_2^{*2} - \beta_1^{*2}) \right].$$

Knowing that, with standardized variables:

$$\begin{aligned} b_1 &= \beta_1 \sigma_1 = \beta_1^* \sigma_Y, \\ b_2 &= \beta_2 \sigma_2 = \beta_2^* \sigma_Y, \\ r &= 2w_{12}^* w_{11}^*, \end{aligned}$$

we find that:

$$J_1^* = \sigma_Y^{-2} \left[ b_1^2 + b_1 b_2 r + \frac{r^2}{2} (b_2^2 - b_1^2) \right].$$

and finally with Eqs. (17):

$$J_1^* = \text{LMG}_1 \text{ (and similarly, } J_2 = \text{LMG}_2).$$

□

## 7 Applications on toy functions and public datasets

In this section, the different VM previously introduced (VIF, SRC<sup>2</sup>, PCC<sup>2</sup>, SPCC<sup>2</sup>, LMG, PMVD and Johnson) are computed and compared on several datasets. We recall that, if the inputs are not independent, only LMG, PMVD and Johnson are VIM as they respect the  $R^2$  decomposition property. All the VM estimations are associated with confidence intervals (CI) in order to capture the uncertainty due to finite sample size of the data sample. The standard bootstrap technique is used to obtain such CI at a 95%-level (typically using 100 replicas).

Table 4 provides a summary of the various datasets used in this section and their corresponding characteristics: the name and corresponding subsection, the input dimension  $d$ , the number of observations  $n$ , the information about the presence of quantitative vs. qualitative inputs (qt/ql), and the source of the dataset. The first three rows correspond to toy cases with simulated data while the remaining ones correspond to public datasets. Note that the +1 sometimes mentioned in the input dimension column refers to the fact that a dummy correlated variable is introduced (but without being explicitly part of the model).

Name	§	$d$	$n$	qt/ql	Source
Independent	7.1.1	3	100	qt	–
Multicollinear	7.1.2	4	100	qt	–
Dummy correlated	7.1.3	1 + 1	100	qt	–
Air quality	7.2	5	111	qt/ql	airquality dataframe
Boston housing	7.3.1	12	506	qt/ql	BostonHousing2 dataframe (mlbench package)
Car prices (r)	7.3.2	15	804	qt/ql	cars dataframe (caret package)

Table 4: Summary of the toy and public use cases.

### 7.1 Simulation data from linear models

#### 7.1.1 Independent inputs’ case (without noise)

We simulate a 100-size sample of  $X = (X_1, X_2, X_3)$  with  $X_1 \sim \mathcal{U}([0.5, 1.5])$ ,  $X_2 \sim \mathcal{U}([1.5, 4.5])$ ,  $X_3 \sim \mathcal{U}([4.5, 13.5])$  and we study the model:

$$Y = X_1^2 + X_2 + X_3 .$$

The data matrix plot (also known as *pairs plot*) is given in Figure 3 (left). For this figure and all the other pairs-plot figures shown in the rest of the report, the upper panel provides the CC of each variables’ pair, the diagonal panel gives the kernel density estimation (or the histogram) of each variable marginal and the lower panel gives scatter plots and fitted smoothers (with CI) of each variables’ pair.

The linear regression between the output and the inputs gives  $R^2 = 0.999$  and  $Q^2 = 0.999$ . VM are given in Table 5 (see also Fig. 4). In Fig. 4 and for all the similar figures in the rest of the paper, SRC2\_j corresponds to the SRC<sup>2</sup> of the input  $X_j$  (and so on for the other metrics).

In this first example, due to the structure of the additive model,  $X_3$  is supposed to play a major role on the output variability. As shown in Figure 3 (left), both scatter plots and CC capture this simple linear influence here. From Table 5, one can notice that VIF values are all equal to unity, which indicates the absence of collinearity, as expected. Concerning VM, SRC<sup>2</sup>, SPCC<sup>2</sup>, LMG, Johnson, and PMVD manage to capture the large influence of  $X_3$ , while PCC<sup>2</sup> only measures the linearity of the inputs w.r.t. the output. Figure 4 provides

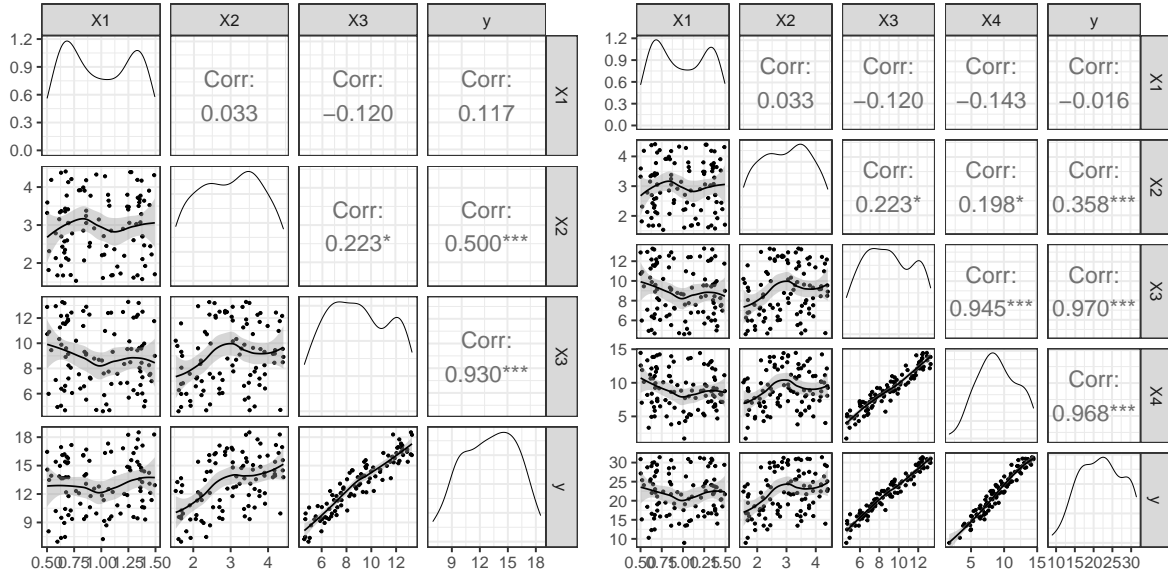


Figure 3: Data pairs plot for the independent linear regression case (left) and multicollinear case (right).

Input	VIF	SRC <sup>2</sup>	PCC <sup>2</sup>	SPCC <sup>2</sup>	LMG	Johnson	PMVD
X <sub>1</sub>	1.02	4.57	98.7	4.49	3.0	3.01	4.56
X <sub>2</sub>	1.06	8.66	99.3	8.20	16.7	16.61	10.47
X <sub>3</sub>	1.07	79.16	99.9	73.94	80.3	80.32	84.91
Sum	-	92.39	298.0	86.62	99.9	99.94	99.94

Table 5: VIF and VM (in %) for the linear model toy data.

both mean estimates together with bootstrap estimates of CI of the VM. The results associated with the LMG and the Johnson indices are strictly equivalent in this case where there is no high multicollinearity and where the inputs and the output are linearly related ( $R^2 = 0.999$ ).

### 7.1.2 Multicollinear case (without noise)

We simulate a 100-size sample of  $X = (X_1, X_2, X_3, X_4)$  with  $X_1 \sim \mathcal{U}([0.5, 1.5])$ ,  $X_2 \sim \mathcal{U}([1.5, 4.5])$ ,  $X_3 \sim \mathcal{U}([4.5, 13.5])$ ,  $X_4 = X_3 + \eta$ ,  $\eta \sim \mathcal{N}(0, 1)$  and we study the model:

$$Y = X_1^2 + X_2 + X_3 + X_4 .$$

For this example, collinearity is introduced within the model between  $X_3$  and  $X_4$ . Figure 3(right) provides the data matrix plot. One can see that the CC indicates this strong correlation between  $X_3$  and  $X_4$ . scatter plots denote the linear influence of these two inputs on the model output.

The linear regression between the output and the inputs gives  $R^2 = 1.000$  and  $Q^2 = 1.000$ . VM are given in Table 6 (see also Fig. 5). SRC<sup>2</sup> corroborate the previous results and enable to identify the collinearity (as their sum are far from  $R^2$ ). As for PCC<sup>2</sup>, it only points out the linear input-output relationships. Finally, SPCC<sup>2</sup> does

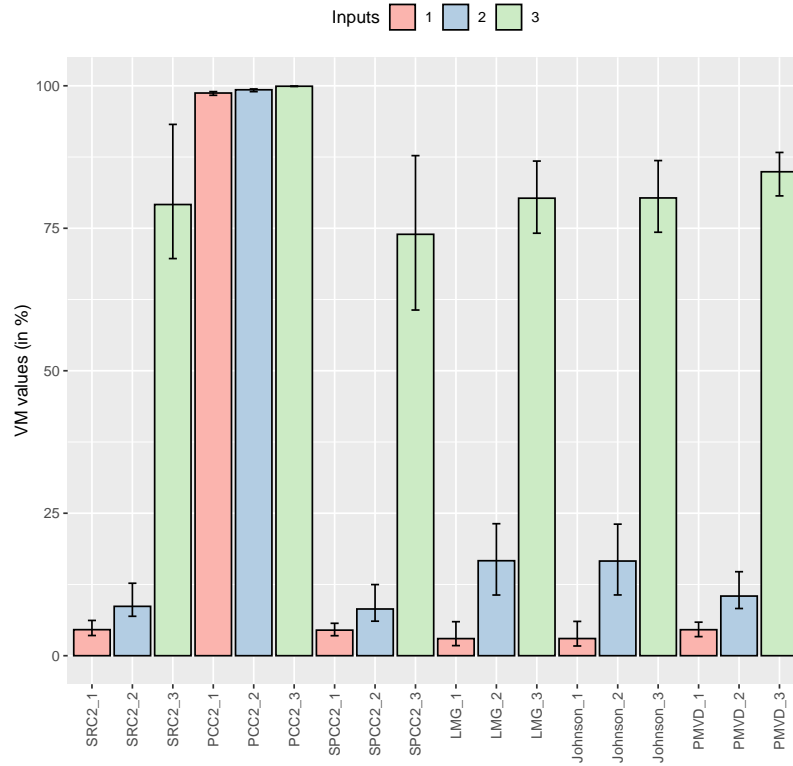


Figure 4: Estimates (with bootstrap) of the VM in the independent linear regression case.

not manage to highlight either collinearity or relative importance. One can notice that VIF values associated with  $X_3$  and  $X_4$  are above 10, which clearly indicates the collinearity between these two regressors. Here,  $SRC^2$ , LMG, Johnson and PMVD are able to capture that  $X_3$  and  $X_4$  have a similar influence.

Input	VIF	$SRC^2$	$PCC^2$	$SPCC^2$	LMG	Johnson	PMVD
$X_1$	1.03	1.19	98.8	1.16	0.81	0.62	1.19
$X_2$	1.06	2.24	99.3	2.12	5.81	6.28	2.63
$X_3$	9.51	19.85	99.3	2.09	46.52	46.16	45.92
$X_4$	9.46	28.39	99.5	3.00	46.85	46.92	50.24
Sum	-	51.67	396.9	8.36	99.99	99.99	99.99

Table 6: VIF and VM (in %) for the multicollinear case data.

Figure 5 corroborates the previous results by providing mean estimates and bootstrap estimates of CI. One can also notice the similarities between the LMG and the Johnson indices which give the same hierarchy in the VM. In this case, the PMVD, LMG and Johnson indices give similar results.

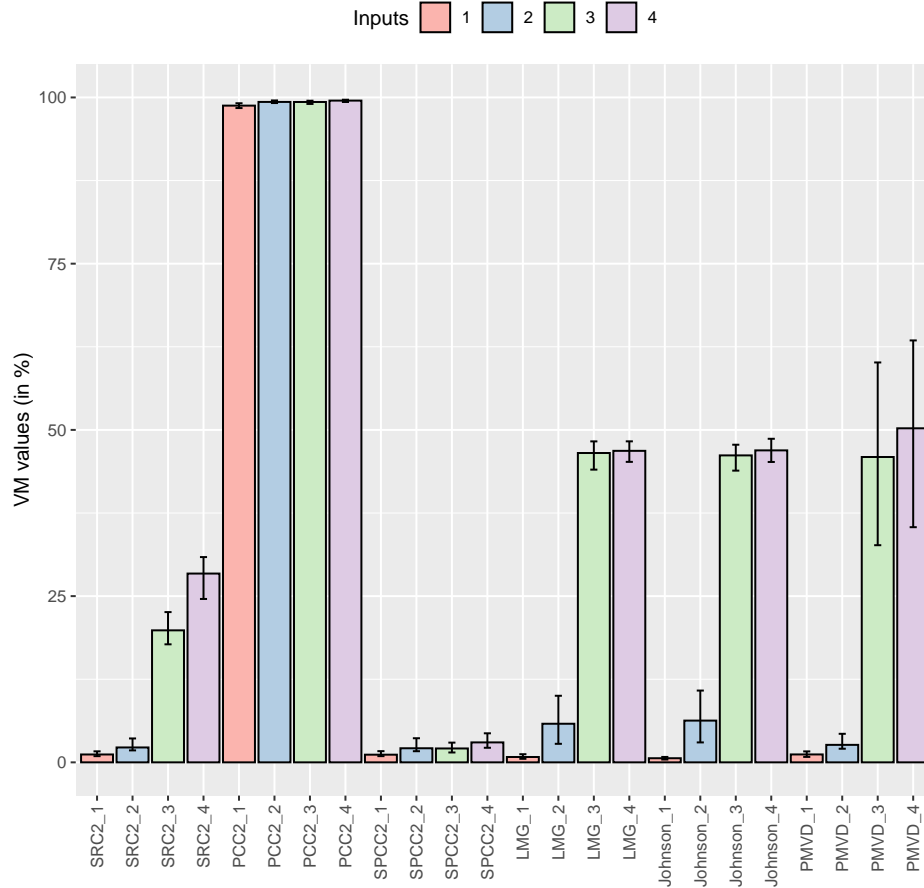


Figure 5: Estimates (with bootstrap) of the VM in the multicollinear case.

### 7.1.3 Model with a dummy (not included in the model) correlated input

We simulate a 100-size sample of  $X = (X_1, X_2)$  with  $X \sim \mathcal{N}_2 \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.9 \\ 0.9 & 1 \end{pmatrix} \right)$  and we study the model:

$$Y = X_1 + \eta ,$$

with  $\eta \sim \mathcal{N}(0, 0.01)$ . The linear regression between the output and the inputs gives  $R^2 = 0.992$  and  $Q^2 = 0.992$ . VM are given in Table 7

This case introduces collinearity by means of the variable  $X_2$  which is not directly included in the regression model, while being strongly correlated to  $X_1$ . One can see that VIF manages to catch a strong collinearity between the two inputs, while  $\text{SRC}^2$ ,  $\text{PCC}^2$ ,  $\text{SPCC}^2$  and  $\text{PMVD}$  only measure the effect of  $X_1$  (seen as a pure linear relationship with  $Y$ ). Finally, the VIM LMG, Johnson and  $\text{PMVD}$  emphasize two different interesting behaviors. LMG and Johnson allocate a part of contribution to both  $X_1$  and  $X_2$  while  $\text{PMVD}$  only assigns the

Input	VIF	SRC <sup>2</sup>	PCC <sup>2</sup>	SPCC <sup>2</sup>	LMG	Johnson	PMVD
$X_1$	6.05	102.86	95.67	17.01	58.1	58.1	99.19
$X_2$	6.05	0.04	0.84	0.01	41.1	41.1	0.04
Sum	-	102.90	96.50	17.01	99.2	99.2	99.2

Table 7: VIF and VM (in %) for the non-included correlated input model toy data.

full contribution to  $X_1$ . This highlights a fundamental difference between PMVD and LMG (recalled in Section 5): the PMVD formulation forces to get a null index for a non-included correlated input.

This test case mostly illustrates that LMG, as already pointed out for the Shapley effects [35, 27], attributes a weight to a dummy variable as soon as it is somehow correlated to another input. This behavior is also found using the Johnson indices. Such a fact is at odds with the criterion ( $C_3$ ) (namely, the exclusion one) recalled in Section 3.2.

## 7.2 Public datasets on air quality

We use the R dataframe “airquality”, which contains some measures of the air quality of New-York in 1973. There are 153 observations but only  $n = 111$  without missing data. In our analysis, we have only considered lines with non-missing data. The output is Ozone and the  $d = 5$  inputs are Solar.R, Wind, Temp, Month and Day. The matrix plot, given in Figure 6, clearly indicates that two inputs, Wind and Temp, are highly linearly correlated to the output (Temp has a positive influence and Wind a negative one). However, analyzing the relative influence and inferring collinearity with this matrix plot become more difficult as the dimension increases ( $d = 5$ ) and the patterns of the scatter plots become rather complex.

The linear regression between the output and the inputs gives  $R^2 = 0.625$  (see also Fig. 7 left) and  $Q^2 = 0.582$ . Table 8 and Figure 8 provide VM results. No strong collinearity is detected here with VIF (while a rough analysis of the matrix plot led one to believe that the correlation of  $-0.5$  between Wind and Temp, together with the correlation of  $0.4$  between Temp and Month, are potential sources of collinearity). However, significant differences appear between the PMVD and the LMG/Johnson indices (the LMG and Johnson indices are very close to each other, even with this imperfect linear model case). The PMVD indices highlight the influence of the temperature, decreasing those of the wind and the solar irradiation. This illustrates the more discriminatory power of PMVD compared to other VIM.

Input	n°	VIF	SRC <sup>2</sup>	PCC <sup>2</sup>	SPCC <sup>2</sup>	LMG	Johnson	PMVD
Solar.R	1	1.15	1.90	4.20	1.65	6.30	6.49	2.65
Wind	2	1.33	12.59	20.16	9.47	22.33	22.91	18.25
Temp	3	1.72	29.48	31.33	17.11	31.96	31.28	39.37
Month	4	1.26	1.81	3.70	1.44	1.65	1.60	1.75
Day	5	1.01	0.51	1.34	0.51	0.26	0.22	0.48
Sum		6.47	46.29	60.73	30.18	62.49	62.49	62.49

Table 8: VIF and VM (in %) for the air quality data.



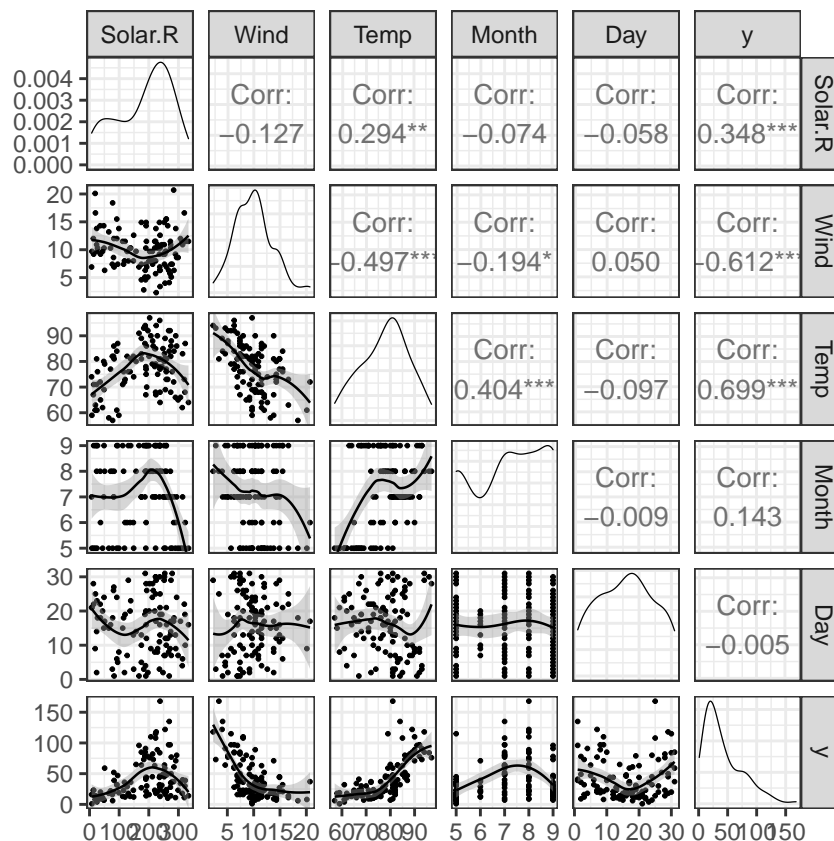


Figure 6: Data pairs plot for the air quality dataset.

### 7.3 Public datasets on economics

#### 7.3.1 Boston housing

We use the BostonHousing2 dataset of the R package `mlbench` which comes from the Boston 1970 census. There are  $n = 506$  observations, one output (`cmedv` which means median value of owner-occupied home) and  $d = 12$  inputs. The matrix plot is given in Figure 9. It shows that strong dependencies exist between inputs and that quite a complex relation links the output with the inputs.

The linear regression between the output and the inputs gives  $R^2 = 0.739$  (see also Fig. 7 right) and  $Q^2 = 0.721$ . VM are given in Table 9. Large VIF for most of the inputs show the strong multicollinearity present in these data. Therefore, large differences between  $SRC^2$  (which are no more valid in this case) and LMG/Johnson appear. Moreover, the interest of PMDV compared to LMG/Johnson is exemplified: PMVD decrease the importance values of inputs with low LMG/Johnson and strongly increase the importance values of the most influential input (`lstat`). PMVD considers that the effects of the other inputs are due to their correlation with `lstat` (which has indeed a quiet large value of VIF). Finally, the proximity between LMG and Johnson values is again highlighted, even with a moderate quality of the linear regression model.

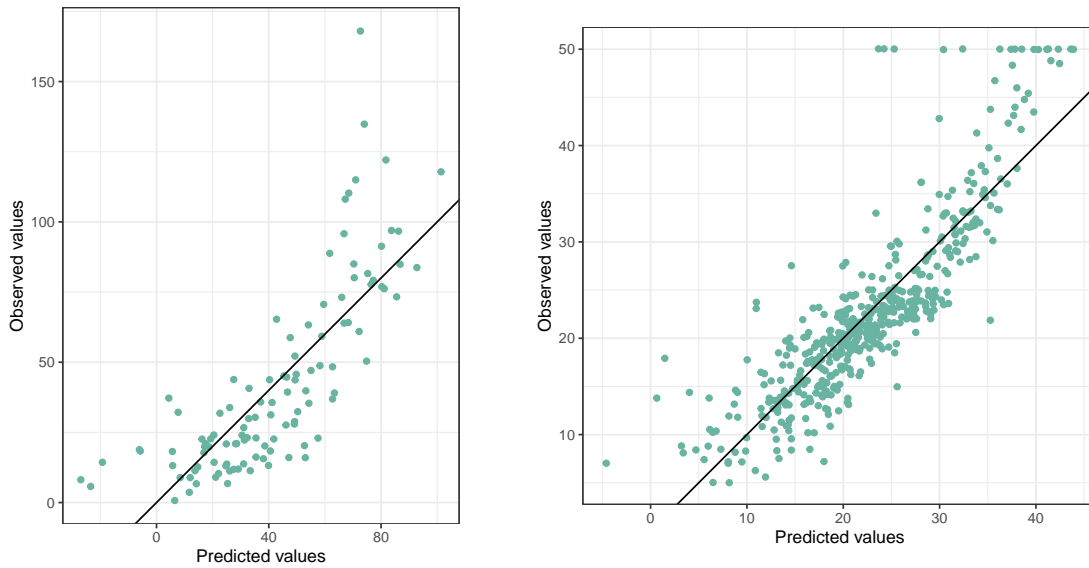


Figure 7: Linear model prediction vs. observation data for the air quality data (left) and the Boston housing data (right).

Input	n°	VIF	SRC <sup>2</sup>	PCC <sup>2</sup>	SPCC <sup>2</sup>	LMG	Johnson	PMVD
crim	1	1.79	1.09	2.28	0.61	2.79	3.29	0.72
zn	2	2.30	1.51	2.46	0.66	2.50	2.81	0.67
indus	3	3.95	0.10	0.10	0.03	3.74	3.66	0.06
nox	4	4.39	4.79	4.02	1.09	3.31	3.68	1.54
rm	5	1.93	8.59	14.57	4.45	19.01	20.59	22.71
age	6	3.09	0.01	0.09	0.00	2.20	2.70	0.00
dis	7	3.95	12.02	10.44	3.04	3.17	1.86	2.18
rad	8	7.40	9.56	4.72	1.29	2.46	2.10	0.83
tax	9	8.88	6.73	2.82	0.76	3.87	3.64	1.07
prratio	10	1.78	5.15	9.97	2.89	7.93	8.70	6.48
b	11	1.34	0.92	2.56	0.69	2.37	2.97	1.12
lstat	12	2.93	17.64	18.76	6.02	20.59	17.92	36.56
Sum		43.74	68.10	72.70	21.51	73.93	73.93	73.93

Table 9: VIF and VM (in %) for the Boston housing data.

### 7.3.2 Regression on cars prices data

We use the cars dataset of the R package `caret` which comes from Kelly Blue Book resale data (2005 model year). It contains suggested retail price (column `Price`) and various characteristics of each car. There are  $n = 804$  observations, one output (`Price` in \$) and 18 inputs. For our analysis, we keep  $d = 15$  inputs (numerical problems in linear regression with the others). One input (`Mileage`) is quantitative and the others are qualitative:

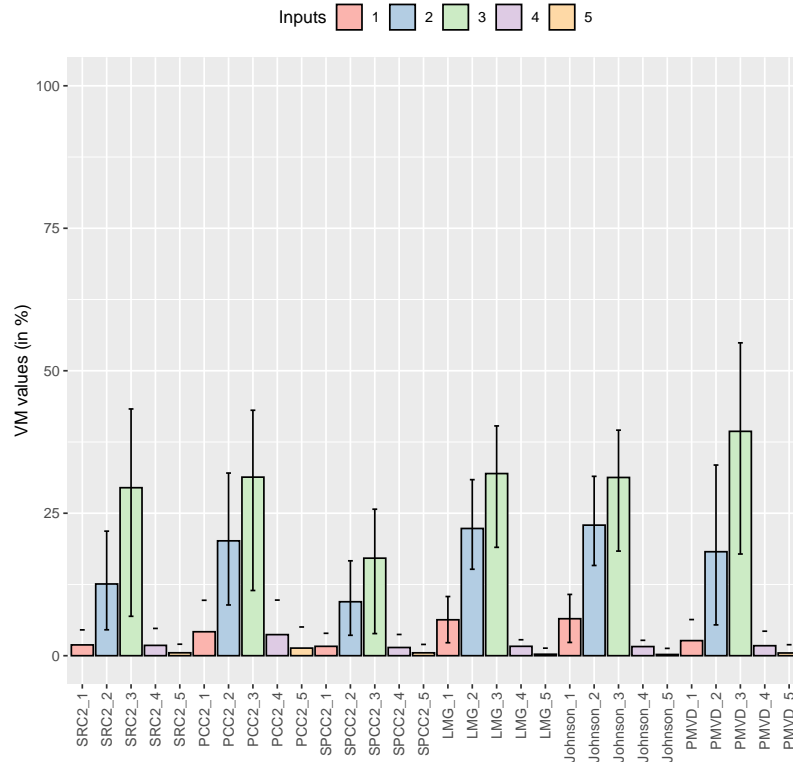


Figure 8: Estimates (with bootstrap) of the VM for the air quality dataset.

one (Cylinder) has three modes and the 13 other inputs are binary (two modes).

The linear regression between the output and the inputs gives  $R^2 = 0.915$  (see also Fig. 10) and  $Q^2 = 0.911$ . VM are given in Table 10 by adding the information on the sign of the CC (“Cor. sign”) between the price and each input in order to know the sense of variation. Large multicollinearity issues are still present in these data (large VIF values). As in the previous examples, LMG and Johnson are very close to each other, and PMVD allows for a better inputs’ influence discrimination.

## 8 Conclusion

In this work, various methods have been proposed to assess the relative importance of predictors/inputs in the linear regression model (a supplementary material of this paper shows how to extend these results to the classification context via the logistic linear regression model). Conditionally to the linear relation hypothesis, interpretations and conditions of use of the various importance measures have been developed based on the variance decomposition of the output, with a special care to the GSA context. One of the final objectives of such works is to provide a user guide for practitioners (see, e.g., [31]), as such guides have been shown to be useful in the GSA community [34].

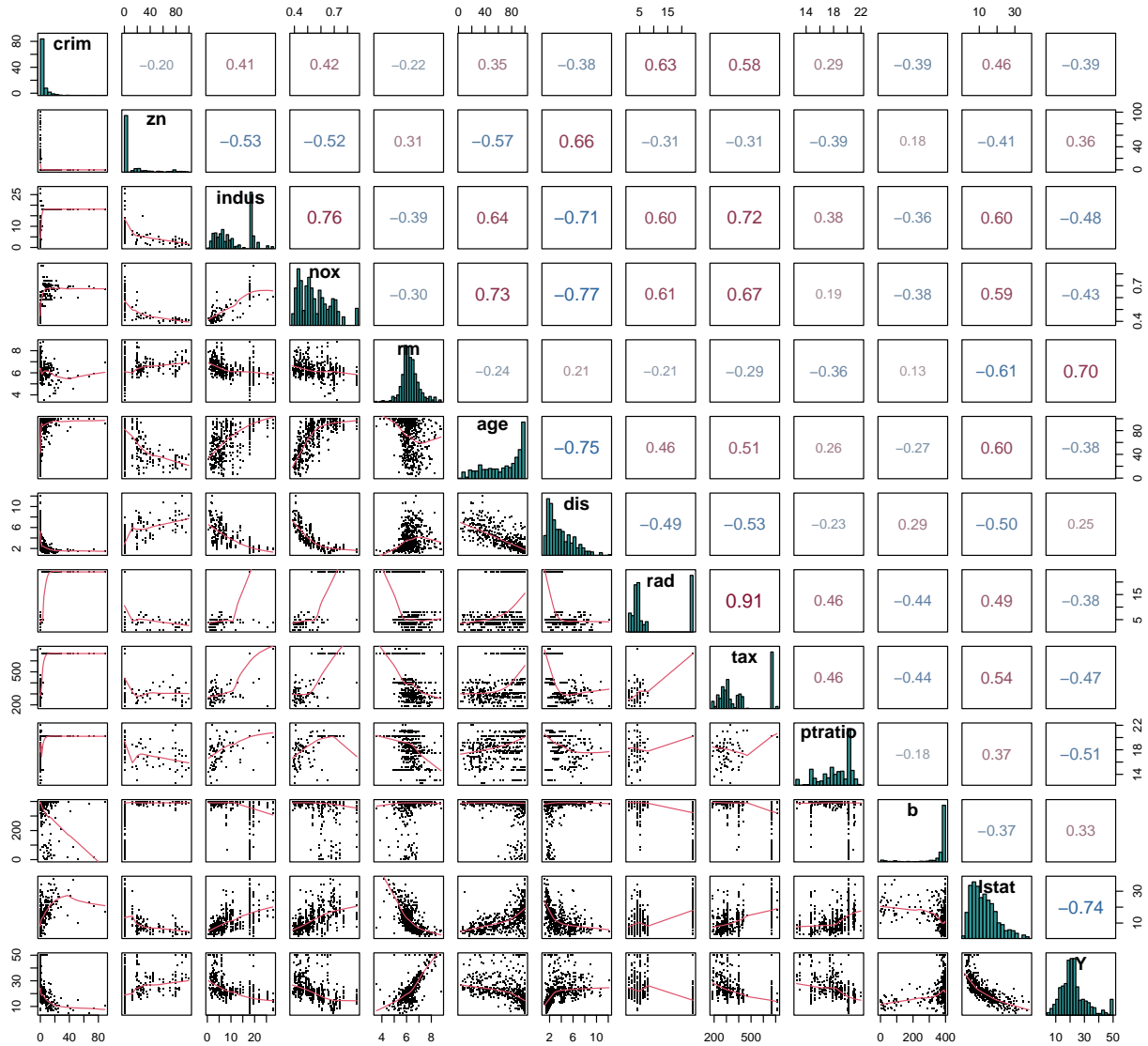


Figure 9: Data pairs plot for the Boston housing dataset.

The relative importance has been considered as the contribution each input makes to the coefficient of determination  $R^2$ , considering both its direct effect (i.e., correlation with the output) and its indirect effect (i.e., correlation with other inputs). A distinction was made between the VM based on a single regression analysis ( $SR^2$ ,  $CC^2$ ,  $PCC^2$  and  $SPCC^2$ ) and the VM requiring multiple regression analyses: the LMG corresponding to Shapley effects, the PMVD and the Johnson indices corresponding to relative weight allocations. The latter VM are VIM as they notably provide a partition of the  $R^2$  among the inputs ( $R^2$  decomposition desirability criterion)

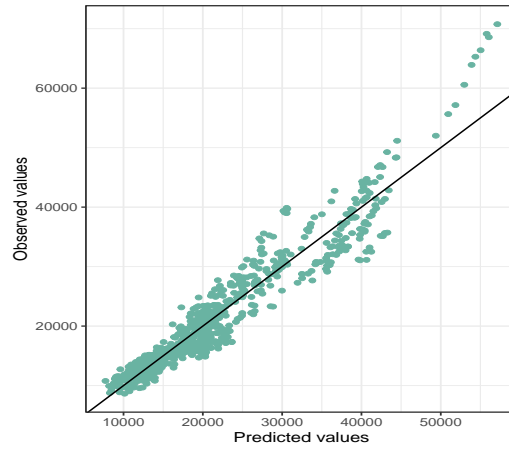


Figure 10: Linear model prediction vs. observation data for the cars data.

Input	n°	VIF	SR $C^2$	PCC $^2$	SPCC $^2$	LMG	Johnson	PMVD	Cor. sign
Mileage	1	1.01	2.33	21.42	2.31	2.25	2.24	2.15	-
Cylinder	2	2.35	26.39	56.92	11.22	21.20	21.96	25.65	+
Doors	3	4.61	1.82	4.43	0.39	1.22	1.07	0.37	+
Cruise	4	1.55	0.02	0.17	0.01	6.10	5.54	0.03	+
Sound	5	1.14	0.04	0.45	0.04	0.42	0.37	0.04	
Leather	6	1.19	0.13	1.26	0.11	1.35	1.41	0.11	+
Buick	7	2.60	0.08	0.37	0.03	0.84	0.86	0.18	+
Cadillac	8	3.33	16.39	36.70	4.92	22.40	22.58	29.57	+
Chevy	9	4.41	0.07	0.20	0.02	6.97	5.68	0.04	-
Pontiac	10	3.42	0.30	1.04	0.09	2.51	2.39	0.12	+
Saab	11	3.56	18.80	38.32	5.28	10.32	11.23	19.68	+
convertible	12	1.63	7.26	34.40	4.45	13.16	12.95	12.16	+
hatchback	13	2.45	2.86	12.10	1.17	1.70	1.93	0.76	-
sedan	14	4.51	4.83	11.19	1.07	1.08	1.30	0.65	-
Sum		37.77	81.34	218.95	31.12	91.51	91.51	91.51	

Table 10: VIF and VM (in %) for the cars data. The last column gives the sense of variation of inputs with significantly influence (LMG> 1).

in the general case of dependent inputs. By measuring the additional contribution of any input in the  $R^2$ , the LMG and Johnson indices provide a classification of the inputs which shares the contribution by taking into account the weight of the correlations among the inputs. It means that an input can have a relative importance even if it does not have a direct influence on the model output. Conversely, the PMVD indices guarantee, by construction, that an input with an estimated regression coefficient equals to zero does yield zero contributions in this relative importance measure. Table 11 (inspired from [25]) synthesizes the desirability criteria (described in detail in Section 3.1) that these three VIM satisfy.

<i>VIM</i>	( <i>C</i> <sub>1</sub> )	( <i>C</i> <sub>2</sub> )	( <i>C</i> <sub>3</sub> )	( <i>C</i> <sub>4</sub> )	( <i>C</i> <sub>5</sub> )
LMG	x	x		x	x
PMVD	x	x	x	x	
Johnson	x	x		x	x

Table 11: Adequation between the VIM and their desirability criteria.

Several datasets has been used to simulate and analyze the measured effects by these various VM. All the results confirm the theoretical properties and the intuitions, as for example the close behaviour between LMG and Johnson indices. The preference to use LMG/Johnson or PMVD thus depends on whether the user wants to consider causality effects or not [25, 60].

The main practical limitation of the LMG and the PMVD methods is the complexity of their calculation which is proportional to  $2^d$ , the number of possible subsets in a set of  $d$  inputs. It has been shown that the Johnson indices can give an excellent alternative to measure the multicollinearity effects when deriving importance measures in a regression model containing a large number (several dozens) of inputs. In this case, the LMG and the PMVD computation is practically impossible. The Johnson indices are in fact a good approximation of the LMG indices in linear regression context. Our current research works try to find similar solutions to approximate the PMVD. Moreover, in a particular case (model with two standardized inputs), a theoretical connection has been found in this paper between LMG and Johnson; further works are necessary to extend this result in more general cases.

Concerning the linear model restriction of all the metrics developed in this report, current works develop other metrics valid in the general case of nonlinear models. In particular, [27] has extended the PMVD to the nonlinear case by defining novel sensitivity indices. Inspired from cooperative game theory, these so-called proportional marginal effects (PME) are based on the proportional value allocation rule. Extension of Johnson indices to nonlinear models is also a strong remaining challenge (see, e.g., a first attempt in [32]).

## Computational details for reproducibility

The results in this paper (as well as in the Supplementary Material) were obtained using R . Both codes and datasets are available at:

<https://gitlab.com/LauraClouvel/toydata/>

In this paper, several functions of a large panel of R packages have been used: `src()` (for  $SRC^2$ ), `pcc()` (for  $PCC^2$  and  $SPCC^2$ ), `lmg()` (for LMG), `pmvd()` (for PMVD) from the `sensitivity` package; `vif()` from the `car` package; and `rwa()` from the `rwa` package. The correlation ratio (see the Supplementary Material) is computed using the `correlRatio()` function from the `sensitivity` package. Finally, all the R packages that have been used are briefly described below.

**The `sensitivity` package [33].** This package<sup>4</sup> contains a collection of functions for GSA, from factor screening, ranking to robustness analysis. Most of the functions have to be applied on a model with scalar output, but several functions support multidimensional outputs. Single-analysis metrics (see Section 4) and

<sup>4</sup>The `sensitivity` package (information: <https://cran.r-project.org/web/packages/sensitivity>, sources: <https://github.com/cran/sensitivity>) is maintained by EDF R&D (with Bertrand Iooss as the maintainer) under a GPL-2 license.

multiple-analysis ones (see Section 5) are provided by this package. This package is used to compute SRC<sup>2</sup>, PCC<sup>2</sup>, SPCC<sup>2</sup>, LMG and PMVD.

**The `car` package.** This package provides the VIF and GVIF metrics for multicollinearity detection (see Section 4.2).

**The `rwa` package.** This package provides the `rwa()` function (for “relative weight analysis”) which enables to compute the standardized Johnson indices.

**Other standard R packages.** The package `boot` is used for computing bootstrap confidence intervals for several VIMs while the package `ggplot2` is used for visualization and displaying graphics.

## References

- [1] Christopher H Achen. *Interpreting and using regression*, volume 29. Sage, 1982.
- [2] A. Antoniadis, S. Lambert-Lacroix, and J-M. Poggi. Random forests for global sensitivity analysis: A selective review. *Reliability Engineering & System Safety*, 206(107312), 2021.
- [3] D.A. Belsley, E. Kuh, and R.E. Welsch. *Regression diagnostics: Identifying influential data and sources of collinearity*. John Wiley & Sons, Inc, 1980.
- [4] N. Benoumechiara and K. Elie-Dit-Cosaque. Shapley effects for sensitivity analysis with dependent inputs: bootstrap and kriging-based algorithms. *ESAIM: Proceedings and Surveys*, 65:266–293, 2019.
- [5] J. Bi. A review of statistical methods for determination of relative importance of correlated predictors and identification of drivers of consumer liking. *Journal of Sensory Studies*, 27:87–101, 2012.
- [6] J-B. Blanchard. Sensitivity analysis with correlated inputs: focus on the linear case. *International Journal for Uncertainty Quantification*, 2023.
- [7] E. Borgonovo and E. Plischke. Sensitivity analysis: A review of recent advances. *European Journal of Operational Research*, 248:869–887, 2016.
- [8] B. Broto, F. Bachoc, and M. Depecker. Variance reduction for estimation of Shapley effects and adaptation to unknown input distribution. *SIAM/ASA Journal on Uncertainty Quantification*, 8:693–716, 2020.
- [9] B. Broto, F. Bachoc, M. Depecker, and J-M. Martinez. Sensitivity indices for independent groups of variables. *Mathematics and Computers in Simulation*, 163:19–31, 2019.
- [10] D.V. Budescu. Dominance analysis: A new approach to the problem of relative importance of predictors in multiple regression. *Psychological Bulletin*, 114:542–551, 1993.
- [11] C. Bénard, G. Biau, S. Da Veiga, and E. Scornet. SHAFF: Fast and consistent SHApley eFFect estimates via random Forests. In *Proceedings of the 25th International Conference on Artificial Intelligence and Statistics*, Virtual, March 2022.
- [12] R. Christensen. *Linear models for multivariate, time series and spatial data*. Springer-Verlag, 1990.

- [13] L. Clouvel. *Uncertainty quantification of the fast flux calculation for a PWR vessel*. Thèse de l'Université Paris-Saclay, 2019.
- [14] L. Clouvel, P. Mosca, J.M. Martinez, and G. Delipei. Shapley and Johnson values for sensitivity analysis of PWR power distribution in fast flux calculation. In *M&C 2019*, Portland, USA, August 2019.
- [15] S. Da Veiga, F. Gamboa, B. Iooss, and C. Prieur. *Basics and Trends in Sensitivity Analysis. Theory and Practice in R*. SIAM, 2021.
- [16] R.B. Darlington and A.F. Hayes. *Regression analysis and linear models*. The Guilford Press, 2017.
- [17] Xingsheng Deng, Liangbo Yin, Sichun Peng, and Meiqing Ding. An iterative algorithm for solving ill-conditioned linear least squares problems. *Geodesy and Geodynamics*, 6(6):453–459, 2015.
- [18] E. Fekhari, B. Iooss, J. Muré, L. Pronzato, and J. Rendas. Model predictivity assessment: incremental test-set selection and accuracy evaluation. In N. Salvati, C. Perna, S. Marchetti, and R. Chambers, editors, *Studies in Theoretical and Applied Statistics, SIS 2021, Pisa, Italy, June 21-25*. Springer, 2023.
- [19] B. Feldman. The proportional value of a cooperative game. In *Econometric Society World Congress 2000 Contributed papers*, number 1140. Econometric Society, 2000.
- [20] B. Feldman. Relative importance and value. *SSRN Electronic Journal*, 03 2005.
- [21] J. Fox and G. Monette. Generalized Collinearity Diagnostics. *Journal of the American Statistical Association*, 87(417):178–183, 1992.
- [22] A. Genizi. Decomposition of  $R^2$  in multiple regression with correlated regressors. *Statistica Sinica*, pages 407–420, 1993.
- [23] U. Grömping. Relative importance for linear regression in R: the Package relaimpo. *Journal of Statistical Software*, 17:1–27, 2006.
- [24] U. Grömping. Estimators of relative importance in linear regression based on variance decomposition. *The American Statistician*, 61(2), 2007.
- [25] U. Grömping. Variable importance in regression models. *WIREs Comput Stat*, 7(137-152), 2015.
- [26] J.C. Helton, J.D. Johnson, C.J. Salaberry, and C.B. Storlie. Survey of sampling-based methods for uncertainty and sensitivity analysis. *Reliability Engineering & System Safety*, 91:1175–1209, 2006.
- [27] M. Hérin, M. Il Idrissi, V. Chabridon, and B. Iooss. Proportional marginal effects for global sensitivity analysis. *Preprint, arXiv:2210.13065*, 2022.
- [28] M. Hérin, M. Il Idrissi, V. Chabridon, and B. Iooss. Proportional marginal effects for sensitivity analysis with correlated inputs. In *Proceedings of the 10th International Conference on Sensitivity Analysis of Model Output (SAMO 2022)*, Tallahassee, Florida, USA, March 2022.
- [29] M. Il Idrissi, V. Chabridon, and B. Iooss. Developments and applications of Shapley effects to reliability-oriented sensitivity analysis with correlated inputs. *Environmental Modelling & Software*, 143(105115), 2021.



- [30] M. Il Idrissi, B. Iooss, and V. Chabridon. Mesures d'importance relative par décomposition de la performance de modèles de régression. In *Actes des 52èmes Journées de Statistique de la Société Française de Statistique (SFdS), Juin 2021, Nice, France*, pages 497–502, 2021.
- [31] B. Iooss, V. Chabridon, and V. Thouvenot. Variance-based importance measures for machine learning model interpretability. In *Actes du 23ème Congrès de Maîtrise des Risques et de Sûreté de Fonctionnement ( $\lambda\mu 23$ )*, Saclay, France, october 2022.
- [32] B. Iooss and L. Clouvel. Une méthode d'approximation des effets de Shapley en grande dimension. In *Actes des 54èmes Journées de Statistique de la Société Française de Statistique (SFdS), July 2023, Brussels, Belgium*, 2023.
- [33] B. Iooss, S. Da Veiga, A. Janon, and G. Pujol. *sensitivity: Global Sensitivity Analysis of Model Outputs*, 2022. R package version 1.28.0.
- [34] B. Iooss and P. Lemaître. A review on global sensitivity analysis methods. In C. Meloni and G. Dellino, editors, *Uncertainty management in Simulation-Optimization of Complex Systems: Algorithms and Applications*, pages 101–122. Springer, 2015.
- [35] B. Iooss and C. Prieur. Shapley effects for sensitivity analysis with dependent inputs: comparisons with Sobol' indices, numerical estimation and applications. *International Journal for Uncertainty Quantification*, 9:493–514., 2019.
- [36] G. James, D. Witten, T. Hastie, and R. Tibshirani. *An introduction to statistical learning: With applications in R, 7th edition*. Springer, 2014.
- [37] J.W. Johnson. A heuristic method for estimating the relative weight of predictor variables in multiple regression. *Multivariate Behavioral Research*, 35:1–19, 2000.
- [38] J.W. Johnson and J.M. LeBreton. History and use of relative importance indices in organizational research. *Organizational Research Methods*, 7:238–257, 2004.
- [39] R. Johnson. The minimal transformation to orthonormality. *Psychometrika*, 31:61–66, 1966.
- [40] J. Karch. Improving on Adjusted R-Squared. *Collabra: Psychology*, 6(1):1–11, 2020.
- [41] W. Kruskal. Relative importance by averaging over orderings. *The American Statistician*, 41:6–10, 1987.
- [42] D. Kurowicka and R. Cooke. *Uncertainty analysis with high dimensional dependence modelling*. Wiley, 2006.
- [43] A. Lepore, B. Palumbo, and J-M. Poggi, editors. *Interpretability for Industry 4.0: Statistical and Machine Learning Approaches*. Springer, 2022.
- [44] R. H. Lindeman, P. F. Merenda, and R. Z. Gold. *Introduction to bivariate and multivariate analysis*. Scott Foresman and Company, Glenview, IL, 1980.
- [45] A. Marrel, B. Iooss, F. Van Dorpe, and E. Volkova. An efficient methodology for modeling complex computer codes with Gaussian processes. *Computational Statistics and Data Analysis*, 52:4731–4744, 2008.

- [46] P. McCullagh and J.A. Nelder. *Generalized linear models*. Chapman & Hall, 1989.
- [47] C. Molnar, G. Casalicchio, and B. Bischl. Interpretable machine learning - A brief history, state-of-the-art and challenges. In *PKDD/ECML Workshops 2020*, pages 417–431, 2020.
- [48] K.F. Nimon and F.L. Oswald. Understanding the results of multiple linear regression: Beyond standardized regression coefficients. *Organizational Research Methods*, 16:650–674, 2013.
- [49] A.B. Owen. Sobol’ indices and Shapley value. *SIAM/ASA Journal on Uncertainty Quantification*, 2:245–251, 2014.
- [50] A.B. Owen and C. Prieur. On Shapley value for measuring importance of dependent inputs. *SIAM/ASA Journal on Uncertainty Quantification*, 5:986–1002, 2017.
- [51] E. Plischke, G. Rabitti, and E. Borgonovo. Computing Shapley effects for sensitivity analysis. *SIAM/ASA Journal on Uncertainty Quantification*, 9:1411–1437, 2021.
- [52] A. Saltelli, K. Chan, and E.M. Scott, editors. *Sensitivity analysis*. Wiley Series in Probability and Statistics. Wiley, 2000.
- [53] L.S. Shapley. A value for n-persons game. In H.W. Kuhn and A.W. Tucker, editors, *Contributions to the theory of games II, Annals of mathematic studies*. Princeton University Press, Princeton, NJ, 1953.
- [54] I.M. Sobol’. Sensitivity estimates for non linear mathematical models. *Mathematical Modelling and Computational Experiments*, 1:407–414, 1993.
- [55] E. Song, B.L. Nelson, and J. Staum. Shapley effects for global sensitivity analysis: Theory and computation. *SIAM/ASA Journal on Uncertainty Quantification*, 4:1060–1083, 2016.
- [56] H. Wallard. Using explained variance allocation to analyse importance of predictors. In *Proceedings of the 16th Conference of the Applied Stochastic Models and Data Analysis*, Le Pirée, Greece, july 2015.
- [57] H. Wallard. Grouping property and decomposition of explained variance in linear regression. In C.H. Skiadas and J.R. Bozeman, editors, *Data Analysis and Applications 1: Clustering and Regression, Modeling-estimating, Forecasting and Data Mining*, pages 73–89. Wiley, 2019.
- [58] R. J. Weber. Probabilistic values for games. In *The Shapley Value*, pages 101–120. Cambridge University Press, 1st edition, October 1988.
- [59] P. Wei, Z. Lu, and J. Song. Variable importance analysis: a comprehensive review. *Reliability Engineering & System Safety*, 142:399–432, 2015.
- [60] K. Zhao and T. Hastie. Causal interpretations of black-box models. *Journal of Business & Economic Statistics*, 39(1):272–281, 2021.
- [61] Hui Zou and Trevor Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 67(2):301–320, 2005.

# Supplementary Material

## “Classification case”

### of the paper “Variance-based importance measures in the linear regression context: Review, new insights and applications”

Laura Clouvel, Bertrand Iooss, Vincent Chabridon,  
Marouane El Idrissi and Frédérique Robin (EDF R&D)

## Introduction

Variance-based Metrics (VM) and Variance-based Importance Measures (VIM) have been defined in the main paper in the classical linear regression context where the response (output) one tries to fit is a quantitative (often continuous) variable, while the predictors (inputs) can be either continuous quantitative variables or qualitative ones (but still, numerically valued). However, many practical applications deal with classification data, where the output is a categorical variable. Easy, especially In this supplementary material, by the way of the generalized linear model (GLM), we give extensions of VM and VIM to the linear logistic regression model. We deal with the case of a binary output, namely in the context of the linear logistic regression.

The structure of this supplementary material is as follows. Section 1 reminds some basics about logistic regression model. Section 2 develops the correlation ratio that is the correlation coefficient between an input and the binary output. Then, Section 3 develops the Johnson indices in the logistic regression context. Finally, Section 4 applies all the studied metrics on several simulated or public datasets. In this paper, the same acronyms and mathematical notations as those of the main paper are used.

## 1 The logistic regression model

In a classification problem, the output  $Y$  is no longer continuous (nor quantitative) but binary (e.g.  $Y \in \{0, 1\}$ ). The GLM [4] allows considering a binomial distribution for  $Y$  and to perform a linear regression on a transformed output (by a so-called “link function”). For example, if  $p = p(X) = \mathbb{P}(Y = 1|X)$ , the logistic regression model writes:

$$g(p) = \log\left(\frac{p}{1-p}\right) = X\beta. \quad (1)$$

It is usually called the “regression model on the link scale” and the link function  $g(p)$  is known as the “logit” transform. Other transforms such as the “probit” one can be used [4].

Via the “inverse logit” transform  $p = [1 + \exp(-g(p))]^{-1}$ , the model in Eq. (1) returns probability values as predictions. In practice, to predict a binary value for the output, a threshold  $s \in ]0, 1]$  has to be defined and

the following predictor is used:

$$\widehat{Y}(x^*) = \mathbb{1}_{\{\widehat{p}(x^*) \geq s\}}(x^*) \quad (2a)$$

$$\text{with } \widehat{p}(x^*) = \left[ 1 + \exp(x^* \widehat{\beta}) \right]^{-1} . \quad (2b)$$

**Remark 1.** *The logistic regression parameters (i.e.  $\beta_i, i = 0, \dots, d$  in Eq. (1)) are intrinsically interpretable, through an exponential transformation, as odds ratios. The quantity  $\exp(\beta_i)$  quantifies the marginal effect of  $X_i$  on the modeled probability  $p$ . The set of odds ratios, while providing an interpretable tool to quantify input importance in the sense of the marginal effect of a variable on the conditional probability, does not fall under the definition of an importance measure (IM) for linear models, and are thus out of the scope of this report. In the following, the focus is put on IM with respect to the linear link between the inputs and the quantity  $g(p)$ . These IMs are not directly interpretable with respect to the output of interest. IM on non-linear links between an output of interest and the inputs (see, e.g. [6, 3]) are beyond the scope this report and will be described in other works. Here, we limit ourselves to the interpretation of  $g(p)$ , being aware that IMs are not directly linked to the classes of the output (but still highly correlated).*

In order to validate the model in Eq. (1),  $R^2$  and  $Q^2$  have to be computed. Considering GLM, several metrics can be used (see, e.g. [10] for a review). A popular one is the following [2]:

$$R^2 = 1 - \frac{D}{D_0} \quad (3)$$

where  $D$  and  $D_0$  are, respectively, the *deviance* and the *null deviance*. Deviance can be seen as a generalization of the variance when the error distribution is non-Gaussian (as provided by the GLM). More precisely, the deviance is twice the difference in log-likelihood between the current model and a saturated model (i.e. a model that fits the data perfectly). As for the null deviance, it is a generalization of the total sum of squares of the linear model. Figure 1 provides an illustrative summary of how these two quantities are connected. Again, other coefficients of determination have been proposed for the logistic regression model [9] but their study is beyond the scope of this report.

The  $Q^2$  estimation is usually computed from cross-validation residuals. As this formula also involves the variance of the observations on the link scale, we compute it by dividing the variance of the linear fits (on the link scale) by  $R^2$ .

In order to validate the model in Eq. (2a), several criteria are useful:

- If one considers that the important class to be predicted (e.g. typically the one which is critical regarding safety purposes) is “TRUE” ( $Y = 1$ ) and the other class is “FALSE” ( $Y = 0$ ), the confusion matrix distinguishes:
  - the number of true positive (TP):  $Y = 1$  and  $\widehat{Y} = 1$ ;
  - the number of true negative (TN):  $Y = 0$  and  $\widehat{Y} = 0$ ;
  - the number of false positive (FP):  $Y = 0$  and  $\widehat{Y} = 1$ ;
  - the number of false negative (FN):  $Y = 1$  and  $\widehat{Y} = 0$ .
- The error rate is the number of errors (false positive and false negative) divided by the number of observations:

$$\varepsilon = \frac{\text{FP} + \text{FN}}{n} . \quad (4)$$

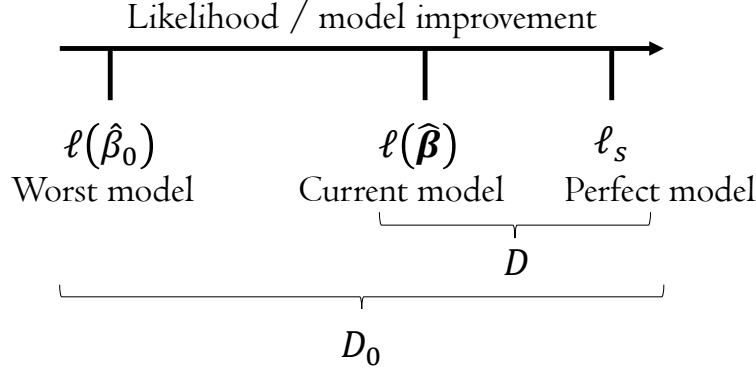


Figure 1: Illustration of deviance and null deviance for GLM validation (inspired from [11]).

- The sensitivity is related to the important class to be predicted. It is the number of good predictions in this class divided by the number of observations in this class:

$$\tau = \frac{\text{TP}}{\text{TP} + \text{FN}} . \quad (5)$$

## 2 Correlation coefficient with the binary output

In the classification context,  $Y$  is a binary variable which can be treated as a qualitative one. The analogue of CC when dealing with a qualitative  $Y$  (of any modalities) and one quantitative  $X_j$  (instead of two quantitative) variables is called the *correlation ratio* (CR). It writes [7]:

$$\text{CR}_j = \eta_{X_j|Y}^2 = \frac{\text{Var } \mathbb{E}(X_j|Y)}{\text{Var } X_j} \quad (6)$$

where one can recognize a first-order Sobol' index [8] formula. CR is also equivalent to the coefficient of determination ( $R^2$ ) of the linear regression explaining the quantitative variable by the qualitative one [7].

Returning to the binary case for  $Y$ , from the sample  $(\mathbf{X}^n, \mathbf{Y}^n)$ , it can be easily estimated by:

$$\hat{\eta}_{X_j|Y}^2 = \frac{n_0 n_1}{n} \frac{(\bar{X}_{j,0} - \bar{X}_{j,1})^2}{\sum_{i=1}^n (X_j^{(i)} - \bar{X}_j)^2} \quad (7)$$

where  $n_0$  and  $\bar{X}_{j,0}$  (resp.  $n_1$  and  $\bar{X}_{j,1}$ ) are the sample size and the empirical mean of  $X_{j,0}$  (resp.  $X_{j,1}$ ) which is the restriction of  $X_j$  to the case  $\{Y = 0\}$  (resp.  $\{Y = 1\}$ ). Let us remark that CR can also be used in a regression context (case of a quantitative variable  $Y$ ) when  $X_j$  is a qualitative variable (by exchanging the role of  $X_j$  and  $Y$  in Eqs. (6) and (7)).

### 3 Johnson indices in the logistic regression context

Following the calculation methodology of the standardized logistic regression coefficient proposed by [5], [9] suggests extending the definition of the Johnson indices to the logistic regression context. By considering the logistic regression model described by Eq. (1), the standardized logistic regression coefficient associated with the variable  $X_i$  is defined as

$$\beta_i^* = \frac{\sigma_{X_i}}{\sigma_{\text{logit}(g(p))}} \beta_i. \quad (8)$$

To define the standard deviation  $\sigma_{\text{logit}(g(p))}$ , one can use the alternative definition of  $R = (\sigma_{\text{logit}(\hat{g}(\hat{p}))})/(\sigma_{\text{logit}(g(p))})$  and thus calculate the  $\beta_i$  such as:

$$\beta_i^* = \frac{\sigma_{X_i}}{\sigma_{\text{logit}(\hat{g}(\hat{p}))}} \beta_i R. \quad (9)$$

The idea is then to apply this definition to the methodology previously defined for a classical linear regression. The matrices  $\mathbf{Z}$ ,  $\mathbf{\Pi}$  and  $A_{\text{logit}}$  are calculated in function of the variables  $X$  and  $g(p)$  standardized beforehand. In particular:

$$A_{\text{logit}} = (\mathbf{Z}^t \mathbf{Z})^{-1} \mathbf{Z}^t \mathbf{G}(p) = \mathbf{Z}^t \mathbf{G}(p) = (\alpha_{\text{logit},j})_{1 \leq j \leq d}, \quad (10)$$

and the Johnson index associated with the variable  $X_i$  in the logistic regression context is thus given by:

$$J_{\text{logit},i} = R^2 \sum_{j=1}^d \alpha_{\text{logit},j}^{*2} \pi_{ij}^{*2}. \quad (11)$$

### 4 Application cases

Classification problems deal with binary  $Y$  and Section 1 has developed the linear logistic regression model which allows modelling  $g(p) = \log \frac{p}{1-p}$  (with  $p = \mathbb{P}(Y = 1)$ ). VM of such models, fitted on the link scale, are then associated to the quantity  $g(p)$  and do not give a direct interpretation of the output on which we focus.

Table 1 provides a summary of the various datasets used in this section and their corresponding characteristics: the name and corresponding subsection, the input dimension  $d$ , the number of observations  $n$ , information about the presence of quantitative vs. qualitative inputs (qt/ql), and the source of the dataset. The first five rows correspond to toy cases with simulated data while the remaining ones correspond to public datasets. Note that the +1 sometimes mentioned in the input dimension column refers to the fact that a dummy correlated variable is introduced (but without being explicitly part of the model).

Name	$\S$	$d$	$n$	qt/ql	Source
Classif #1	4.1	3	100	qt	-
Classif #2 (dummy)	4.1	2 + 1	100	qt	-
Car prices (c)	4.2	15	804	qt/ql	cars dataframe (caret package)

Table 1: Summary of the toy and public use cases.

## 4.1 Illustration on simulation data from toy cases

We first study the three-dimensional ( $d = 3$ ) linear classification model:

$$Y = \mathbb{1}_{\sum_{i=1}^d a_i X_i \geq k} \quad (12)$$

with  $k \in \mathbb{R}$  and  $X_i \sim \mathcal{N}(0, 1)$   $i = 1, \dots, d$ . In our case, we take  $k = 0$ ,  $a = (1, 2, 3)$  and we simulate a 100-size sample of  $X$ . The matrix plot is given in Figure 2 (left).

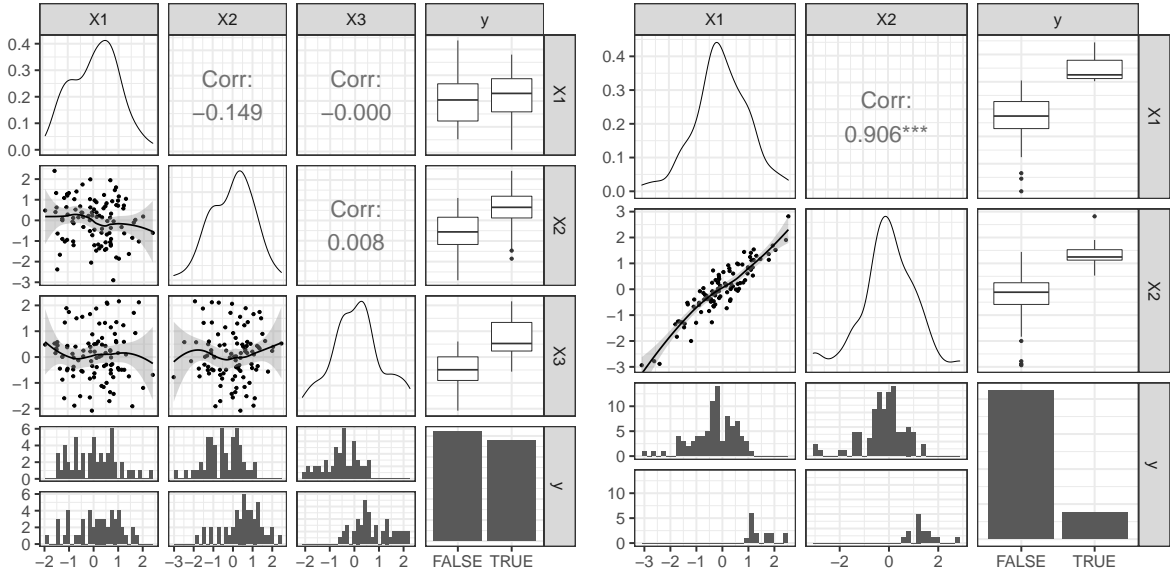


Figure 2: Data pairs plot for the linear classification case (left) and the dummy-correlated-variable classification case (right). The upper panel provides the CC of each variable pair; the diagonal panel gives the kernel density estimation of the marginals; the lower panel gives scatter plots and fitted GLM with CI. As the output variable is not continuous but binary, other representations are given in the right column and bottom line.

On the link scale, the linear regression between the output and the inputs gives  $R^2 = 1.000$  and  $Q^2 = 0.921$ . By taking the threshold  $s$  (see Eq. (2a)) at the mid-value and classical value 0.5 to distinguish the two classes, the classification error rate (Eq. (4)) is  $\varepsilon = 0$  and the classification sensitivity (Eq. (5)) is  $\tau = 1$ , which mean a perfect fit (as expected). The VMs, from the regression on the link scale, are given in Table 2 and Figure 3 (left). It shows that LMG, Johnson and PMVD provide similar results that SRC<sup>2</sup> (which is only based on the regression coefficients that give a higher weight to  $X_2$  than to  $X_1$ ). The output corresponds to a threshold exceedance that is mainly explained by  $X_3$ .  $X_1$  and  $X_2$  compete  $X_3$  only via their interaction effects (concomitant large values). Therefore, this interaction effect is shared between these inputs in the LMG/Johnson/PMVD approach, and their effect is equalized.

We now study a model with  $d = 2$  correlated inputs with one *dummy* variable (i.e. non-included in the model):

$$Y = \mathbb{1}_{X_1 + \eta \geq 1} \quad (13)$$

Input	VIF	CR	SRC <sup>2</sup>	PCC <sup>2</sup>	SPCC <sup>2</sup>	LMG	Johnson	PMVD
$X_1$	8.86	0.723	8.65	6.86	2.58	5.36	6.27	10.2
$X_2$	14.83	24.977	37.78	40.54	26.28	35.91	35.19	35.9
$X_3$	28.63	44.577	58.20	62.22	44.01	58.73	58.55	53.8
Sum	52.31	70.277	104.62	109.62	72.88	100.00	100.00	100.0

Table 2: VIF and VM (in %) for the linear classification data.

with  $\eta \sim \mathcal{N}(0, 0.01)$  and  $X \sim \mathcal{N}_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.9 \\ 0.9 & 1 \end{pmatrix}\right)$ . We simulate a 100-size sample of  $X$ . The matrix plot is given in Figure 2 (right).

On the link scale, the linear regression between the output and the inputs gives  $R^2 = 0.951$  and  $Q^2 = 0.841$ . By taking the threshold  $s = 0.5$ , we have  $\varepsilon = 0.02$  and  $\tau = 0.93$ . The VMs, from the regression on the link scale, are given in Table 3 and Figure 3 (right). PMVD allows drastically decreasing the importance measure of  $X_2$  which is only due to its correlation with  $X_1$ . One also observes the closeness between LMG and the (logistic regression-based) Johnson indices.

Input	VIF	CR	SRC <sup>2</sup>	PCC <sup>2</sup>	SPCC <sup>2</sup>	LMG	Johnson	PMVD
$X_1$	12.3	41.1	111.51	11.994	9.176	64.1	57.9	91.27
$X_2$	12.3	32.1	1.36	0.323	0.126	31.0	37.3	3.87
Sum	24.6	73.2	112.87	12.317	9.301	95.1	95.1	95.14

Table 3: VIF and VM (in %) for the non-included-input classification model toy data.

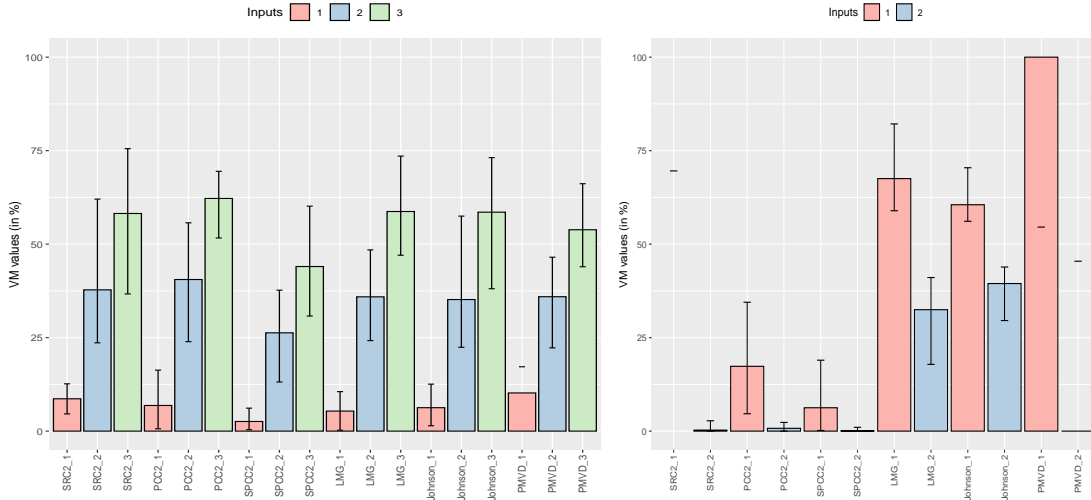


Figure 3: Estimates (with bootstrap) of the VM in the linear classification case (left) and in the *dummy-correlated-variable* classification case (right). SRC2<sub>j</sub> corresponds to the SRC<sup>2</sup> of the input  $X_j$  (and so on for the other VM).



## 4.2 Application to a public dataset: car prices data

We now use the car data for a classification exercise ( $Y$  is binary) by distinguishing the cars prices above and below a given price (\$40,000). The important class to be predicted ( $Y = 1$ ) is for the high prices. On the link scale, the linear logistic regression between the output and the inputs gives  $R^2 = 0.757$  and  $Q^2 = 0.601$ . By taking the threshold  $s = 0.2$  to distinguish the two classes, the classification error rate (Eq. (4)) is  $\varepsilon = 0.037\%$  and the classification sensitivity (Eq. (5)) is  $\tau = 1$ . The VMs, from the regression on the link scale, are given in Table 4. The difference with the regression case is that some variables (as Saab) have no more influence. The influence of the three main influential inputs (Cylinder, Cadillac and convertible) are still present.

Input	n°	VIF	CR	SRC <sup>2</sup>	PCC <sup>2</sup>	SPCC <sup>2</sup>	LMG	Johnson	PMVD	Cor. sign
Mileage	1	1.01	1.40	0.79	2.26	1.03	4.67	1.42	5.13	-
Cylinder	2	2.35	18.85	1.94	1.09	1.61	21.68	14.20	27.9	+
Doors	3	4.61	0.56	1.49	0.00	0.59	1.45	0.81	0.00	+
Cruise	4	1.55	1.72	0.21	0.00	0.04	2.30	1.06	0.00	-
Sound	5	1.14	0.26	0.00	0.00	0.02	0.31	0.13	0.02	
Leather	6	1.19	2.00	0.01	0.00	0.03	3.01	0.57	0.00	+
Buick	7	2.60	0.58	0.05	0.00	0.17	1.20	0.66	0.00	-
Cadillac	8	3.33	35.14	16.99	0.10	5.25	20.90	27.05	31.3	+
Chevy	9	4.41	1.63	0.32	0.00	0.03	3.11	1.91	0.00	-
Pontiac	10	3.42	1.20	0.44	0.00	0.30	3.34	1.15	0.00	-
Saab	11	3.56	0.87	18.91	0.00	0.42	3.10	15.44	0.15	-
convertible	12	1.63	8.78	13.47	1.69	5.10	8.90	10.56	11.2	+
hatchback	13	2.45	0.42	0.53	0.00	0.53	0.51	0.16	0.00	
sedan	14	4.51	0.01	1.79	0.00	0.63	1.27	0.62	0.00	-
Sum		37.77	73.42	56.93	5.14	15.75	75.74	75.74	75.7	

Table 4: VIF and VM (in %) for the cars classification data. The last column gives the sense of variation of inputs with significantly influence (LMG > 1).

## References

- [1] E. García-Portugués. *Notes for Predictive Modeling*. <https://bookdown.org/egarpor/PM-UC3M/>, 2021. Version 5.9.0.
- [2] A. Guisan and N. E. Zimmerman. Predictive habitat distribution models in ecology. *Ecological Modelling*, 135:147–186, 2000.
- [3] A. Marrel and V. Chabridon. Statistical developments for target and conditional sensitivity analysis: Application on safety studies for nuclear reactor. *Reliability Engineering & System Safety*, 214:107711, 2021.
- [4] P. McCullagh and J.A. Nelder. *Generalized linear models*. Chapman & Hall, 1989.
- [5] Scott Menard. Six approaches to calculating standardized logistic regression coefficients. *The American Statistician*, 58(3):218–223, 2004.
- [6] H. Raguet and A. Marrel. Target and conditional sensitivity analysis with emphasis on dependence measures. *Working paper*, 2018, URL <https://arxiv.org/abs/1801.10047>.
- [7] G. Saporta. *Probabilités, analyse de données et statistique*. éditions Technip, 1990.
- [8] I.M. Sobol'. Sensitivity estimates for non linear mathematical models. *Mathematical Modelling and Computational Experiments*, 1:407–414, 1993.
- [9] S. Tonidandel and J.M. LeBreton. Determining the relative importance of predictors in logistic regression: An extension of relative weight analysis. *Organizational Research Methods*, 13:767–781, 2010.
- [10] B. Zheng and A. Agresti. Summarizing the predictive power of a generalized linear model. *Statistics in Medicine*, 19:1771–1781, 2000.