A review on variance-based importance measures in the linear regression context

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

A review on variance-based importance measures in the linear regression context

Laura Clouvel[1], Bertrand Iooss[2], Vincent Chabridon[2], Marouane Il Idrissi[2], and Frédérique Robin[1]

[1]EDF R&D, PERICLES Department, Saclay, France
[2]EDF R&D, PRISME Department, Chatou, France & SINCLAIR AI Lab., Saclay, France

Abstract
The identification of causal effects and influential variables related to some phenomena of interest is one of the fundamental issues in many socio-environmental studies. 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. The practical relevancy of such tools is highlighted through their empirical study on simulated data, as well as public datasets. Other test cases, as well the use of the VIMS in a classification context (via the logistic linear regression model), are also presented in the supplementary materials.

Keywords
Multicollinearity; Proportional values; Relative weight analysis; Sensitivity analysis; Variance decomposition

1 Introduction

The identification of causal effects and influential variables related to some phenomena of interest is one of the fundamental issues in many socio-environmental studies (Razavi et al., 2020). 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 (Kruskal, 1987; Grömping, 2015). 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 (Genizi, 1993; Budescu, 1993; Johnson & LeBreton, 2004; Bi, 2012; Iooss et al., 2022). From a practical viewpoint, they are essential in data analysis and post-hoc interpretation of learned models (Darlington & Hayes, 2017; Molnar et al., 2020; Lepore et al., 2022). 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 (Saltelli et al., 2000; Da Veiga et al., 2021; Antoniadis et al., 2021).

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., Helton et al. (2006); Iooss & Lemaître (2015); Wei et al. (2015); Borgonovo & Plischke (2016)). However, the ignorance of GSA or poor understanding on the part of practitioners can lead to flaws while conducting their interpretation of data or models (Saltelli et al., 2020). As an example, one can cite the controverse created by the work of Sovacool et al. (2020) concluding that, contrarily to renewable energy, larger nuclear attachments of a country do not tend to associate with lower carbon emissions. Based on multiple regression analyses on datasets (carbon emissions, renewable electricity production fraction and nuclear electricity production fraction from 123 countries), this

work has been shown to present many statistical strong biases and errors (Wagner, 2021; Perez, 2022). In addition, 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., Johnson & LeBreton (2004); Grömping (2015)).

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 Achen (1982) 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 Johnson & LeBreton (2004), 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* (Budescu, 1993).

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 Clouvel (2019) and also studied in Il Idrissi et al. (2021)) 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 (Lindeman et al., 1980) and the PMVD indices (Feldman, 2005) 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 (Johnson, 2000) (see also Genizi (1993)) based on relative weight allocations are highlighted.

While this work is not intended as an exhaustive review (we refer to Grömping (2015) 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 Grömping (2015); Wallard (2015, 2019); Blanchard (2023)). Finally, using the implementation of these VIMs in several `R` packages (especially in the `'sensitivity'` package (Iooss et al., 2023)), 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 (see Appendix A).

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 (the supplementary material 1 contains additional test cases). 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 2. 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)
$\boldsymbol{x} := (x_1, \cdots, x_d)$	Vector of deterministic variables (bold, italic and lowercase)
$\boldsymbol{X} := (X_1, \cdots, X_d)$	Random vector (bold, italic and uppercase)
$\mathrm{x}_j^{(i)}$	i-th observation of the variable x_j
$\mathbf{x}^{(i)}$	i-th observation of the vector $\boldsymbol{x}$
$\mathbf{X}^n := \left(\mathrm{x}_1^{(i)}, \ldots, \mathrm{x}_d^{(i)}\right)_{i=1,\ldots,n}$	n-observation matrix (bold and capital letter)
$\widehat{\beta}$	Estimator of the parameter β (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 $\boldsymbol{X} = (X_1, \ldots, X_d)$ (inputs or regressors), denoted by:

$$(\mathbf{X}^n, \mathbf{y}^n) = \left(\mathrm{x}_1^{(i)}, \ldots, \mathrm{x}_d^{(i)}, y^{(i)}\right)_{i=1,\ldots,n}. \tag{1}$$

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:*

$$\mathrm{E}[X_j] = 0 \textit{ for } j = 1, \ldots, d, \textit{ and } \mathrm{E}[Y] = 0.$$

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

$$Y = \boldsymbol{X}\beta + \varepsilon, \tag{2}$$

where $\beta = (\beta_1, \ldots, \beta_d)^\top \in \mathbb{R}^d$ is an unknown vector of coefficients, and ε is a random error assumed to be Gaussian and centered, i.e., $\varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2)$, and such that $\mathrm{E}\left[\varepsilon | \boldsymbol{X}\right] = 0$. Specifically, for each observation $\mathbf{x}^{(i)}$ of $\boldsymbol{X}$ and $y^{(i)}$ of Y, the previous relationship can be written as $y^{(i)} = \mathbf{x}^{(i)}\beta + \varepsilon^{(i)}$, where for all $i = 1, \ldots, n$, the $\varepsilon^{(i)}$s are independent and identically distributed (i.i.d.) according to the same law as ε. We thus deduce that:

$$\mathrm{E}\left[Y | \boldsymbol{X} = \left(\mathrm{x}_1^{(i)}, \ldots, \mathrm{x}_d^{(i)}\right)\right] = \mathbf{x}^{(i)}\beta, \quad \text{for } i = 1, \ldots, n.$$

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., Christensen (1990)) can be used to estimate the vector of parameters β by using the unbiased maximum likelihood estimator given by:

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

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:

$$R^2 = R^2_{Y(X)} := 1 - \frac{\mathrm{E}\left[\mathrm{VAR}(Y|X))\right]}{\mathrm{VAR}(Y)} = \frac{\mathrm{VAR}(\mathrm{E}[Y|\boldsymbol{X}])}{\mathrm{VAR}(Y)}. \tag{4}$$

Provided a consistent estimator $\widehat{\beta}$ of β, 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:

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

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* $R^2_{\mathrm{adj}} = 1 - \left|1 - R^2\right|\left|\dfrac{n-1}{n-(1+d)}\right|$ *(see Karch (2020) 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 (Marrel et al., 2008; Fekhari et al., 2023).*

3 Variance-based importance measures

Importance measures in regression models (Darlington & Hayes, 2017) 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 (Saltelli et al., 2000; Da Veiga et al., 2021). 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 (Kurowicka & Cooke, 2006)), 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 $\mathrm{VAR}(Y)$.

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:

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

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:

$$\mathrm{VAR}(\mathrm{E}[Y|\boldsymbol{X}]) = \beta^\top \Sigma_{XX}\beta, \quad \mathrm{E}[\mathrm{VAR}(Y|\boldsymbol{X})] = \sigma^2_\varepsilon, \tag{6}$$

where $\Sigma_{XX} = \left(\mathrm{COV}(X_i, X_j)\right)_{1\le i,j\le d}$ 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 (Johnson & LeBreton, 2004; Grömping, 2007). 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 Grömping (2007), 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 (Zou & Hastie, 2005; Wallard, 2019):

- (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:

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

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

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

where σ_Y and σ_j are the standard deviations associated with Y and the input X_j, respectively. Hence, the *squared SRC* β_j^{*2} can then be used as a VIM (Grömping, 2006; Antoniadis et al., 2021). 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:

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

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

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

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* (Belsley et al., 1980) 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" (McCullagh & Nelder, 1989)). 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., Deng et al. (2015)).

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* $\mathrm{COV}(X_1, X_2) = r\sigma_1\sigma_2$*, one has:*

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

Similarly, we can easily determine the squared CCs:

$$r^2_{Y,X_1} = \frac{(b_1 + rb_2)^2}{b_1^2 + 2b_1b_2r + b_2^2 + \sigma_\varepsilon^2} \quad \text{and} \quad r^2_{Y,X_2} = \frac{(b_2 + rb_1)^2}{b_1^2 + 2b_1b_2r + b_2^2 + \sigma_\varepsilon^2} \ . \tag{9}$$

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) (Fox & Monette, 1992; Johnson & LeBreton, 2004) defined as:

$$\text{VIF}_j = \frac{1}{1 - R^2_{X_j(\boldsymbol{X}_{-j})}} \ , \tag{10}$$

where $\boldsymbol{X}_{-j}$ is the vector of all the inputs except X_j, and where $R^2_{X_j(\boldsymbol{X}_{-j})}$ represents the R^2 from the linear regression where X_j is considered as the output, and by taking $\boldsymbol{X}_{-j}$ 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 (James et al., 2014).

Example: two-input regression model (Section 4.1, continued).
From Eq. (10)*, one simply has* $\text{VIF}_1 = \text{VIF}_2 = \dfrac{1}{1 - r^2}$ *with* $r \neq \pm 1$.

Remark 3. *The* generalized variance inflation factor *(GVIF) has been proposed by Fox & Monette (1992) 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 $\boldsymbol{X}_{-j}$. It is in that spirit that the *partial correlation coefficient* (PCC) has been introduced (Saltelli et al., 2000). It is defined as:

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

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

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:*

$$r^2_{(Y,X_1)|X_2} = \frac{R^2_{Y(X_1,X_2)} - r^2_{Y,X_2}}{1 - r^2_{Y,X_2}} \ ,$$

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

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

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 (Saltelli et al., 2000; Helton et al., 2006; Iooss & Lemaître, 2015), 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 $\boldsymbol{X}_{-j}$ 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 (Johnson & LeBreton, 2004). SPCC is defined as the proportion of the output variance explained by X_j after removing the "information brought" by $\boldsymbol{X}_{-j}$ (as a difference of explained variance). It is formally given by the CC (noted $r_{Y,(X_j|\boldsymbol{X}_{-j})}$) between Y and the residuals of the regression of X_j on $\boldsymbol{X}_{-j}$. The *squared SPCC* is intrinsically linked to the R^2 since it can be written as:

$$r^2_{Y,(X_j|\boldsymbol{X}_{-j})} = R^2_{Y(\boldsymbol{X})} - R^2_{Y(\boldsymbol{X}_{-j})} \ . \tag{12}$$

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:*

$$r^2_{Y,(X_1|X_2)} = R^2_{Y(X_1,X_2)} - r^2_{Y,X_2} \ .$$

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

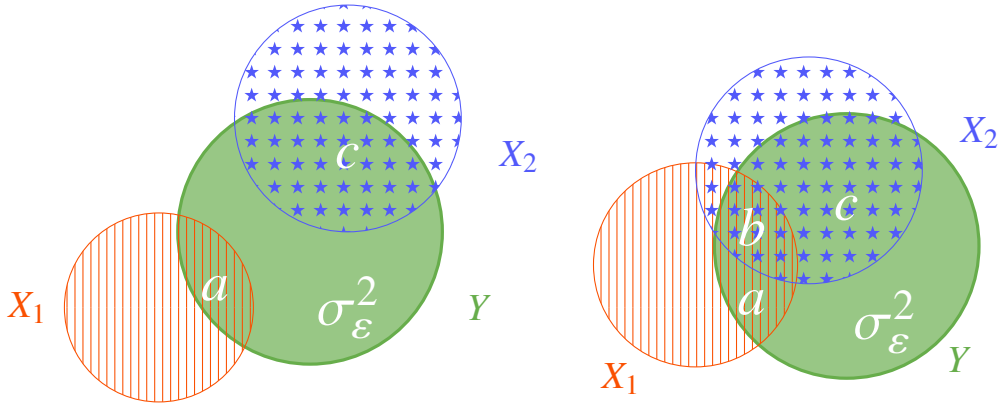
$$r^2_{Y,(X_1|X_2)} = \frac{b_1^2(1-r^2)}{b_1^2 + 2b_1b_2r + b_2^2 + \sigma_\varepsilon^2} \quad \text{and} \quad r^2_{Y,(X_2|X_1)} = \frac{b_2^2(1-r^2)}{b_1^2 + 2b_1b_2r + b_2^2 + \sigma_\varepsilon^2} \ . \tag{13}$$

In Genizi (1993), 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 (Clouvel, 2019; Il Idrissi et al., 2021).



(a) $r = 0, b = 0$ **(b)** $r \neq 0, b \neq 0$

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 $a + b + c + \sigma^2_\varepsilon$ with σ^2_ε 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:

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

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:

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

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:

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

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 :

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

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

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

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* (Budescu, 1993). 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" (Lindeman et al., 1980)). These VIMs have been studied extensively (see, e.g., Budescu (1993); Grömping (2006)). 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 $\{1, \ldots, d\}$ and $\boldsymbol{X}_u = (X_j : j \in u)$ 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 $Y(\boldsymbol{X}_u)$ by the increase in R^2 that results from adding that predictive variable to the regression model:

$$\mathrm{LMG}_j = \frac{1}{d!} \sum_{\substack{\pi \in \text{permutations} \\ \text{of } \{1,\ldots,d\}}} r^2_{Y,(X_j|\boldsymbol{X}_\pi)} \tag{15}$$

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

$$\mathrm{LMG}_j = \frac{1}{d} \sum_{i=0}^{d-1} \sum_{\substack{u \subseteq -\{j\} \\ |u|=i}} \binom{d-1}{i}^{-1} r^2_{Y,(X_j|\boldsymbol{X}_u)} = \frac{1}{d} \sum_{u \subseteq -\{j\}} \binom{d-1}{|u|}^{-1} r^2_{Y,(X_j|\boldsymbol{X}_u)} \tag{16}$$

with $\binom{n}{k} = \dfrac{n!}{(n-k)!\,k!}$ and $r^2_{Y,(X_j|\boldsymbol{X}_u)} = R^2_{Y(\boldsymbol{X}_{u\cup\{j\}})} - R^2_{Y(\boldsymbol{X}_u)}$.

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., Kruskal (1987); Genizi (1993)). 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:*

$$LMG_1 = \frac{1}{2}\left(R^2_{Y(X_1,X_2)} - R^2_{Y(X_2)} + R^2_{Y(X_1)}\right) \ , \ LMG_2 = \frac{1}{2}\left(R^2_{Y(X_1,X_2)} - R^2_{Y(X_1)} + R^2_{Y(X_2)}\right),$$

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

$$LMG_1 = \frac{b_1^2 + b_1 b_2 r + \dfrac{r^2}{2}(b_2^2 - b_1^2)}{b_1^2 + 2b_1 b_2 r + b_2^2 + \sigma_\varepsilon^2} \ , \ LMG_2 = \frac{b_2^2 + b_1 b_2 r + \dfrac{r^2}{2}(b_1^2 - b_2^2)}{b_1^2 + 2b_1 b_2 r + b_2^2 + \sigma_\varepsilon^2}. \tag{17}$$

This result is also given in Grömping (2007).

Going back to the Venn diagram illustration, (see Fig. 1), one has (Il Idrissi et al., 2021):

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

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), LMG_1 and 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 Feldman (2005) and Grömping (2007), 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, Feldman (2005) 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 Grömping (2007, 2015). The PMVD indices are defined as follows:

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

where:

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

Example: two-input regression model (Section 4.1, continued).
Eq. (18) *becomes (Grömping, 2007):*

$$PMVD_1 = \frac{b_1^2 + b_1^2/(b_1^2+b_2^2)2b_1b_2r}{b_1^2 + 2b_1b_2r + b_2^2 + \sigma_\varepsilon^2} \ , \ PMVD_2 = \frac{b_2^2 + b_2^2/(b_1^2+b_2^2)2b_1b_2r}{b_1^2 + 2b_1b_2r + b_2^2 + \sigma_\varepsilon^2}.$$

Moreover, one can notice that:

- *If we put $|r| = 1$, we obtain $PMVD_1 = b_1^2/(b_1^2+b_2^2)$ and $PMVD_2 = b_2^2/(b_1^2+b_2^2)$. 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:*

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

Going back to the Venn diagram analogy (see Fig. 1), one has (see also Hérin et al. (2022b)):

$$\begin{aligned} \text{PMVD}_1 &= a[1 + b/(a+c)]/(a+b+c+\sigma_\varepsilon^2), \\ \text{PMVD}_2 &= c[1 + b/(a+c)]/(a+b+c+\sigma_\varepsilon^2). \end{aligned}$$

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:

$$\text{PMVD}_1 = a/(a+c) \ , \ \text{PMVD}_2 = c/(a+c) \ .$$

and one can notice that the PMVD does not depend on b anymore. While this behavior is known when dealing with two inputs, Grömping (2007) 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 Feldman (2005) and Grömping (2007), 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 and in the supplementary materials.

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 Il Idrissi et al. (2021), 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 2 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 Weber (1988) and Feldman (2005). 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 (Shapley, 1953). 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* (Feldman, 2000), 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 σ_ε^2 is equal to 0 (Saltelli et al., 2000; Helton et al., 2006). 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^2 is actually equal to the *first-order Sobol' index*, which is defined outside of the realm of linear models (Sobol', 1993). 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 (Owen, 2014), where chosen value function is the closed Sobol' index of a subset of player, leading to the *Shapley effects*. Several authors (Song et al., 2016; Benoumechiara & Elie-Dit-Cosaque, 2019; Iooss & Prieur, 2019; Plischke et al., 2021) proposed and studied several dedicated estimation algorithms for nonlinear models. Analytical formulas have also been exhibited for linear models with Gaussian inputs (Owen & Prieur, 2017), and can be efficiently computed by finely tuned algorithms (Broto et al., 2019). 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 (Broto et al., 2020; Bénard et al., 2022).

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, Hérin et al. (2022a) 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* (Johnson, 2000) 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 Nimon & Oswald (2013) and Grömping (2015)). As an example, one can mention the work of Genizi (1993) 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 (Johnson & LeBreton, 2004; Clouvel, 2019; Clouvel et al., 2019), 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 $\boldsymbol{X} = (X_1, \ldots, X_d)$).

6.1 Johnson indices

In the case of the Johnson indices (Johnson, 1966, 2000))[1], 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 Johnson (1966) , it consists in finding $\mathbf{Z}^n$ and $\mathbf{W} \in \mathbb{R}^{d\times d}$ such as:

$$\left\{\begin{array}{rcl} \mathbf{X}^n & = & \mathbf{Z}^n\mathbf{W} \\ (\mathbf{Z}^n)^\top\mathbf{Z}^n & = & \mathbf{I} \\ \mathbf{Z}^n & = & \underset{\mathbf{\Pi}^n}{\arg\min}\ \mathrm{Tr}\,(\mathbf{X}^n - \mathbf{\Pi}^n)^\top(\mathbf{X}^n - \mathbf{\Pi}^n) \end{array}\right. \tag{19}$$

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

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

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

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

$$\mathbf{X}^n = \mathbf{P}^n\mathbf{\Delta}\mathbf{Q}^\top, \tag{21}$$

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 $\delta_1 \geq \ldots \geq \delta_d > 0$. In that sense, the new set of uncorrelated variables $z_1, \ldots, z_d$ is maximally correlated with the original set of correlated variables $x_1, \ldots, x_d$ (the columns of $\mathbf{X}^n$).

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

$$\Sigma_{XX} = W^2. \tag{22}$$

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 $\boldsymbol{\alpha} \in \mathbb{R}^d$:

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

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

$$\mathrm{VAR}\,\mathrm{E}(Y|\boldsymbol{Z}) = \sum_{j=1}^{d}\alpha_j^2. \tag{23}$$

[1]Note, that there are two different authors with the same name. Johnson (2000) suggested determining the matrix $\mathbf{W}$ as the weights of the regression of $\mathbf{X}^n$ on $\mathbf{Z}^n$ contrary to Johnson (1966) which regressed $\mathbf{Z}^n$ on $\mathbf{X}^n$.

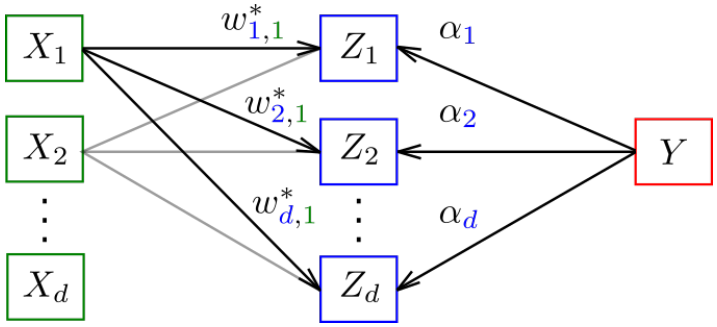


Figure 2: Representation of the Johnson relative weight calculation associated with the input X_1.

The α_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, Johnson (2000) 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:*

$$\widehat{\boldsymbol{\alpha}} = \mathbf{W}\widehat{\boldsymbol{\beta}}. \tag{24}$$

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

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

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}:

$$\widehat{w}^*_{ij} = \frac{\widehat{w}_{ij}}{\sqrt{\sum_k \widehat{w}^2_{kj}}} = \widehat{r}_{Z_i,X_j}, \tag{25}$$

and thus for all j:

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

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 $\widehat{\alpha}_i^2$ of variance in Y accounted for by Z_i by the proportion $(\widehat{w}^*_{ij})^2$ of each Z_i accounted for by X_j. The Johnson index associated with the input X_j can thus be expressed as:

$$J_j = \sigma_Y^{-2} \sum_{i=1}^{d} \alpha_i^2 w^{*2}_{ij}. \tag{27}$$

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 $\widehat{\Sigma}_{Y,X}$ and $\widehat{\Sigma}_{X,X}$ (Grömping, 2006):

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

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

$$\widehat{R}^2 = \widehat{\sigma}_Y^{-2} \widehat{\beta}^t \widehat{\Sigma}_{X,X} \widehat{\beta}. \tag{28}$$

Using Eqs.(22) and (24), Eq. (28) thus gives the decomposition[2]:

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

In the paper of Johnson (2000), 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:

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

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

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

Finally, it is important to note that the standardization of the inputs is equivalent to directly calculate the matrix W^* and α^* thanks to the multivariate correlation matrices R_{XX} and R_{XY}, as in the initial paper of Johnson (1966). The eigen decomposition of the correlation matrix R_{XX} gives:

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

The matrix $\mathbf{W}^*$ is then given (similarly to the Eq. (20)) by[3]:

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

and the vector α^* is determined thanks to the relation:

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

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

$$\boldsymbol{\alpha}^* = \mathbf{W}^* \beta^*, \tag{30}$$

with β^* 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 $\mathbf{W}^*$ is distributed over the Johnson indices. The similar behavior between the LMG and Johnson indices has been confirmed in Thomas et al. (2014) who show their strict equality in the two-dimensional case (see also the proof in Appendix B).

[2] By construction the vector α and β are associated with the same quadratic minimization problem of the function $S(\beta) = \|\mathbf{y}^n - X^n\beta\|^2 = \left\|\mathbf{y}^n - P^n Q^t Q \Delta Q^t \beta\right\|^2 = \|\mathbf{y}^n - \mathbf{Z}^n W \beta\|^2 = \|\mathbf{y}^n - \mathbf{Z}^n \alpha\|^2 = S(\alpha)$.

[3] As a reminder, in Johnson (1966), $\mathbf{W}^* = \mathbf{Q}^* \boldsymbol{\Delta}^{*^{-1}} \mathbf{Q}^{*^\top}$ because $\mathbf{Z}^n$ is regressed on $\mathbf{X}^n$.

7 Applications on toy functions and public datasets

In this section, the different VM previously introduced (VIF, SRC^2, PCC^2, $SPCC^2$, 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 in the supplementary material 1) 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	$\check{g}$	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	Suppl. material 1	5	111	qt/ql	`airquality` dataframe
Boston housing	7.2	12	506	qt/ql	`BostonHousing2` dataframe (`mlbench` package)
Car prices (r)	Suppl. material 1	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^2 of the input X_j (and so on for the other metrics).

Input	VIF	SRC^2	PCC^2	$SPCC^2$	LMG	Johnson	PMVD
X_1	1.02	4.57	98.7	4.49	3.0	3.01	4.56
X_2	1.06	8.66	99.3	8.20	16.7	16.61	10.47
X_3	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.

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^2, $SPCC^2$, LMG, Johnson, and PMVD manage to capture the large influence of X_3, while PCC^2 only measures the linearity of the inputs w.r.t. the output. Figure 4 provides 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 ($R^2 = 0.999$).

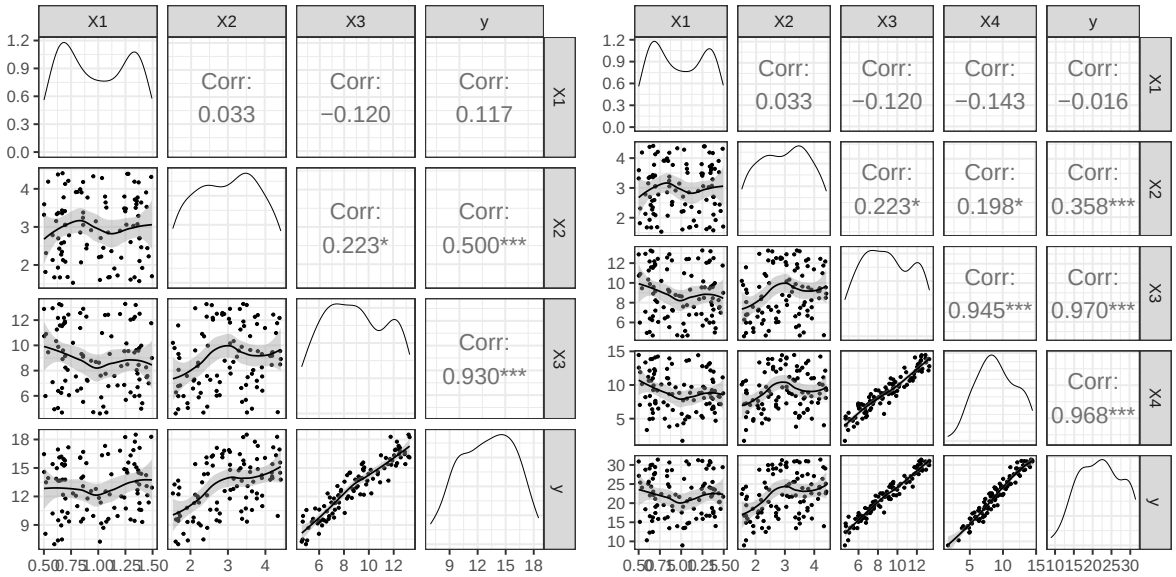


Figure 3: Data pairs plot for the independent linear regression case (left) and multicollinear case (right).

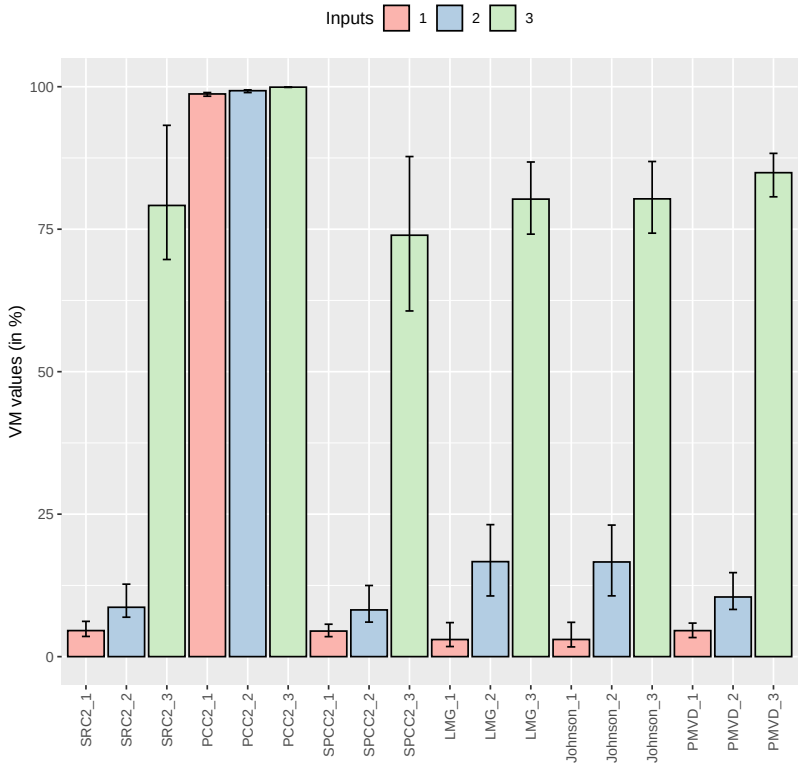


Figure 4: Estimates (with bootstrap) of the VM in the independent linear regression case.

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^2 corroborate the previous results and enable to identify the collinearity (as their sum are far from R^2). As for PCC^2, it only points out the linear input-output relationships. Finally, $SPCC^2$ does 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.

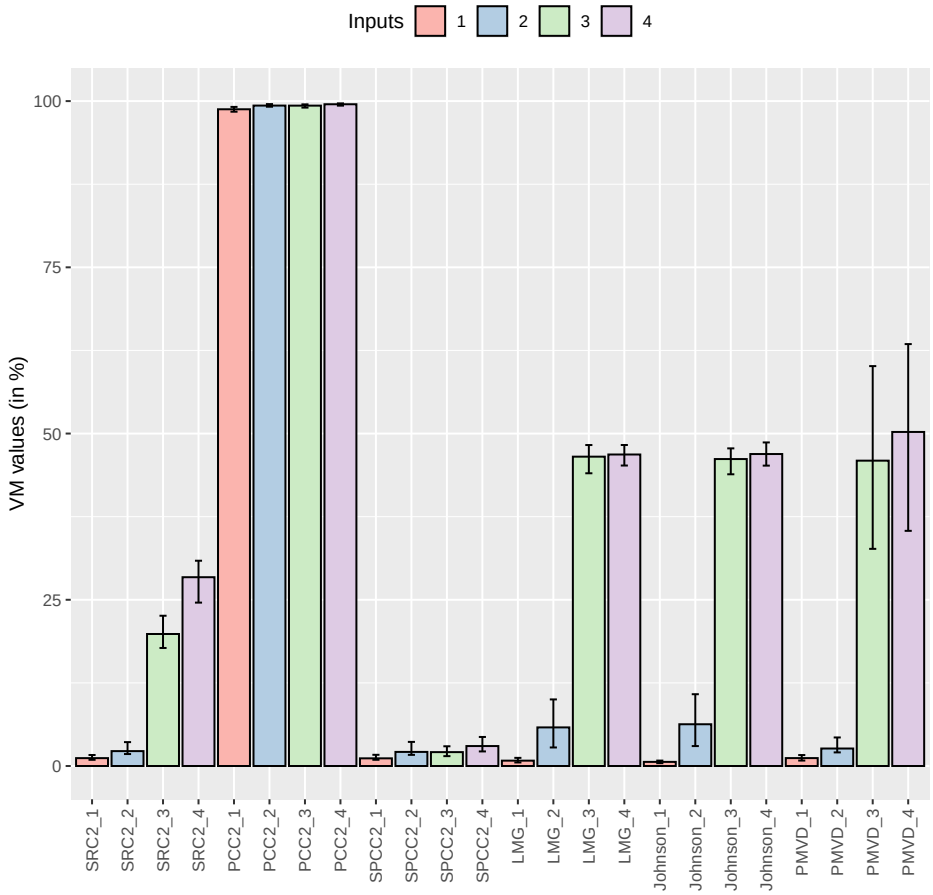


Figure 5: Estimates (with bootstrap) of the VM in the multicollinear case.

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.

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.

Input	VIF	SRC^2	PCC^2	$SPCC^2$	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.

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 SRC^2, PCC^2, $SPCC^2$ and PMVD only measure the effect of X_1 (seen as a pure linear relationship with Y). Finally, the VIM LMG, Johnson and PMVD emphasize two different interesting behaviors. LMG and Johnson allocate a part of contribution to both X_1 and X_2 while PMVD only assigns the 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 (Iooss & Prieur, 2019; Hérin et al., 2022a), 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 dataset: The 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 6. 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) and $Q^2 = 0.721$. VM are given in Table 8. 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.

8 Conclusion

In this work, various methods have been proposed to assess the relative importance of predictors/inputs in the linear regression model (the supplementary material 2 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., Iooss et al. (2022)), as such guides have been shown to be useful in the GSA community (Iooss & Lemaître, 2015).

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.,

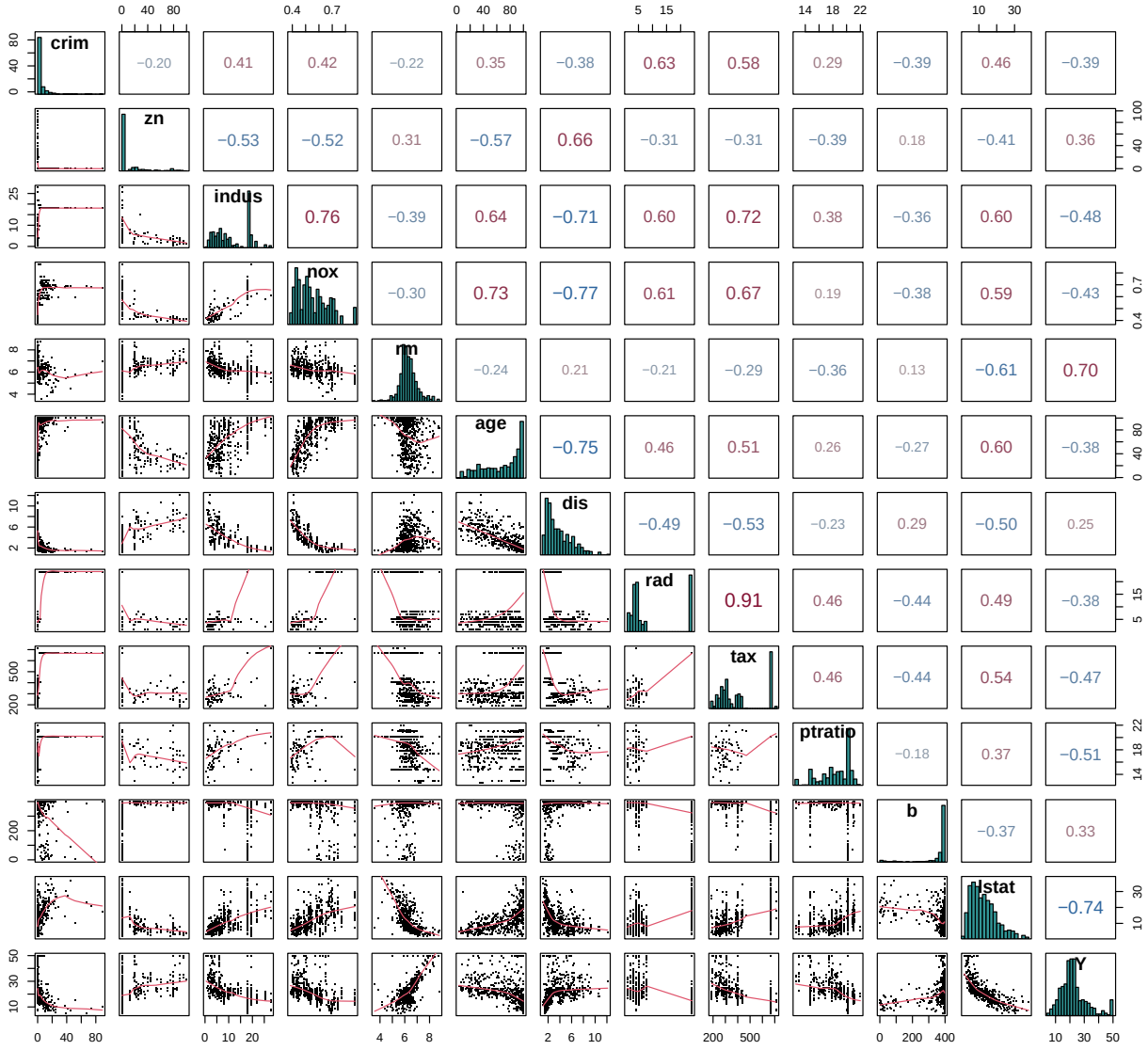


Figure 6: Data pairs plot for the Boston housing dataset.

Input	n°	VIF	SRC^2	PCC^2	$SPCC^2$	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
ptratio	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 8: VIF and VM (in %) for the Boston housing data.

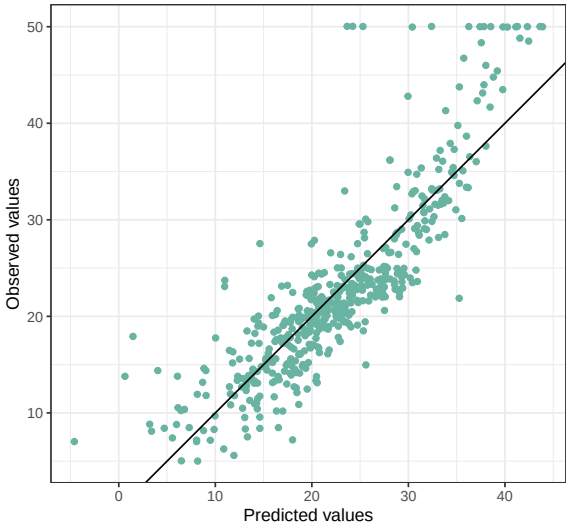


Figure 7: Linear model prediction vs. observation data for the Boston housing data.

correlation with other inputs). A distinction was made between the VM based on a single regression analysis (SRC^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) 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 9 (inspired from Grömping (2015)) synthesizes the desirability criteria (described in detail in Section 3.1) that these three VIM satisfy.

VIM	(C_1)	(C_2)	(C_3)	(C_4)	(C_5)
LMG	x	x		x	x
PMVD	x	x	x	x	
Johnson	x	x		x	x

Table 9: 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 (Grömping, 2015; Zhao & Hastie, 2021).

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.

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, Hérin et al. (2022a) 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 Iooss &

Clouvel (2023)).

A Computational details for reproducibility

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

`https://gitlab.com/LauraClouvel/toydata/.`

Several `R` packages have been used and are briefly described below.

The `sensitivity` package (Iooss et al., 2023). This package[4] 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 multiple-analysis ones (see Section 5) are provided by this package, via the functions: `src()` (for SRC^2), `pcc()` (for PCC^2 and $SPCC^2$), `lmg()` (for LMG), `pmvd()` (for PMVD) and `johnson()` (for Johnson indices). The correlation ratio (see the Supplementary Material 2) is computed using the `correlRatio()` function.

The `car` package. This package provides the VIF and GVIF metrics (`vif()` function) for multicolinearity detection (see Section 4.2).

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.

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

The equivalence between the LMG and the standardized Johnson indices in dimension two is proved with a different demonstration from the one of Thomas et al. (2014) which relies on geometrical arguments.

Proposition 1. *If $d = 2$, the LMG and the standardized Johnson indices are equal:*

$$J_j^{*2} = LMG_j \ \ for \ \ j = 1, 2.$$

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:

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

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:

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

[4]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.

Knowing that, with standardized variables:

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

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):

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

□

References

Achen, C. H. (1982). *Interpreting and using regression*, volume 29. Sage.

Antoniadis, A., Lambert-Lacroix, S., & Poggi, J.-M. (2021). Random forests for global sensitivity analysis: A selective review. *Reliability Engineering & System Safety*, 206(107312).

Belsley, D., Kuh, E., & Welsch, R. (1980). *Regression diagnostics: Identifying influential data and sources of collinearity*. John Wiley & Sons, Inc.

Benoumechiara, N. & Elie-Dit-Cosaque, K. (2019). Shapley effects for sensitivity analysis with dependent inputs: bootstrap and kriging-based algorithms. *ESAIM: Proceedings and Surveys*, 65:266–293.

Bi, J. (2012). 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.

Blanchard, J.-B. (2023). Sensitivity analysis with correlated inputs: focus on the linear case. *International Journal for Uncertainty Quantification*.

Bénard, C., Biau, G., Veiga, S. D., & Scornet, E. (2022). SHAFF: Fast and consistent SHApley eFfect estimates via random Forests. In *Proceedings of the 25th International Conference on Artificial Intelligence and Statistics*, Virtual.

Borgonovo, E. & Plischke, E. (2016). Sensitivity analysis: A review of recent advances. *European Journal of Operational Research*, 248:869–887.

Broto, B., Bachoc, F., & Depecker, M. (2020). Variance reduction for estimation of Shapley effects and adaptation to unknown input distribution. *SIAM/ASA Journal on Uncertainty Quantification*, 8:693–716.

Broto, B., Bachoc, F., Depecker, M., & Martinez, J.-M. (2019). Sensitivity indices for independent groups of variables. *Mathematics and Computers in Simulation*, 163:19–31.

Budescu, D. (1993). Dominance analysis: A new approach to the problem of relative importance of predictors in multiple regression. *Psychological Bulletin*, 114:542–551.

Christensen, R. (1990). *Linear models for multivariate, time series and spatial data*. Springer-Verlag.

Clouvel, L. (2019). *Uncertainty quantification of the fast flux calculation for a PWR vessel*. Thèse de l'Université Paris-Saclay.

Clouvel, L., Mosca, P., Martinez, J., & Delipei, G. (2019). Shapley and Johnson values for sensitivity analysis of PWR power distribution in fast flux calculation. In *M&C 2019*, Portland, USA.

Da Veiga, S., Gamboa, F., Iooss, B., & Prieur, C. (2021). *Basics and Trends in Sensitivity Analysis. Theory and Practice in R*. SIAM.

Darlington, R. & Hayes, A. (2017). *Regression analysis and linear models*. The Guilford Press.

Deng, X., Yin, L., Peng, S., & Ding, M. (2015). An iterative algorithm for solving ill-conditioned linear least squares problems. *Geodesy and Geodynamics*, 6(6):453–459.

Fekhari, E., Iooss, B., Muré, J., Pronzato, L., & Rendas, J. (2023). Model predictivity assessment: incremental test-set selection and accuracy evaluation. In Salvati, N., Perna, C., Marchetti, S., & Chambers, R., editors, *Studies in Theoretical and Applied Statistics, SIS 2021, Pisa, Italy, June 21-25*, pages 315–347. Springer.

Feldman, B. (2000). The proportional value of a cooperative game. In *Econometric Society World Congress 2000 Contributed papers*, number 1140. Econometric Society.

Feldman, B. (2005). Relative importance and value. *SSRN Electronic Journal*.

Fox, J. & Monette, G. (1992). Generalized Collinearity Diagnostics. *Journal of the American Statistical Association*, 87(417):178–183.

Genizi, A. (1993). Decomposition of R^2 in multiple regression with correlated regressors. *Statistica Sinica*, pages 407–420.

Grömping, U. (2006). Relative importance for linear regression in R: the Package relaimpo. *Journal of Statistical Software*, 17:1–27.

Grömping, U. (2007). Estimators of relative importance in linear regression based on variance decomposition. *The American Statistician*, 61(2).

Grömping, U. (2015). Variable importance in regression models. *WIREs Comput Stat*, 7(137-152).

Helton, J., Johnson, J., Salaberry, C., & Storlie, C. (2006). Survey of sampling-based methods for uncertainty and sensitivity analysis. *Reliability Engineering & System Safety*, 91:1175–1209.

Hérin, M., Il Idrissi, M., Chabridon, V., & Iooss, B. (2022a). Proportional marginal effects for global sensitivity analysis. *Preprint, arXiv:2210.13065*.

Hérin, M., Il Idrissi, M., Chabridon, V., & Iooss, B. (2022b). 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.

Il Idrissi, M., Chabridon, V., & Iooss, B. (2021). Developments and applications of Shapley effects to reliability-oriented sensitivity analysis with correlated inputs. *Environmental Modelling & Software*, 143(105115).

Il Idrissi, M., Iooss, B., & Chabridon, V. (2021). Mesures dimportance 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.

Iooss, B., Chabridon, V., & Thouvenot, V. (2022). 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.

Iooss, B. & Clouvel, L. (2023). 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*.

Iooss, B., Da Veiga, S., Janon, A., & Pujol, G. (2023). *sensitivity: Global Sensitivity Analysis of Model Outputs*. R package version 1.29.0.

Iooss, B. & Lemaître, P. (2015). A review on global sensitivity analysis methods. In Meloni, C. & Dellino, G., editors, *Uncertainty management in Simulation-Optimization of Complex Systems: Algorithms and Applications*, pages 101–122. Springer.

Iooss, B. & Prieur, C. (2019). Shapley effects for sensitivity analysis with dependent inputs: comparisons with Sobol' indices, numerical estimation and applications. *International Journal for Uncertainty Quantification*, 9:493–514,.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2014). *An introduction to statistical learning: With applications in R, 7th edition*. Springer.

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

Johnson, J. & LeBreton, J. (2004). History and use of relative importance indices in organizational research. *Organizational Research Methods*, 7:238–257.

Johnson, R. (1966). The minimal transformation to orthonormality. *Psychometrika*, 31:61–66.

Karch, J. (2020). Improving on Adjusted R-Squared. *Collabra: Psychology*, 6(1):1–11.

Kruskal, W. (1987). Relative importance by averaging over orderings. *The American Statistician*, 41:6–10.

Kurowicka, D. & Cooke, R. (2006). *Uncertainty analysis with high dimensional dependence modelling*. Wiley.

Lepore, A., Palumbo, B., & Poggi, J.-M., editors (2022). *Interpretability for Industry 4.0: Statistical and Machine Learning Approaches*. Springer.

Lindeman, R. H., Merenda, P. F., & Gold, R. Z. (1980). *Introduction to bivariate and multivariate analysis*. Scott Foresman and Company, Glenview, IL.

Marrel, A., Iooss, B., Van Dorpe, F., & Volkova, E. (2008). An efficient methodology for modeling complex computer codes with Gaussian processes. *Computational Statistics and Data Analysis*, 52:4731–4744.

McCullagh, P. & Nelder, J. (1989). *Generalized linear models*. Chapman & Hall.

Molnar, C., Casalicchio, G., & Bischl, B. (2020). Interpretable machine learning - A brief history, state-of-the-art and challenges. In *PKDD/ECML Workshops 2020*, pages 417–431.

Nimon, K. & Oswald, F. (2013). Understanding the results of multiple linear regression: Beyond standardized regression coefficients. *Organizational Research Methods*, 16:650–674.

Owen, A. (2014). Sobol' indices and Shapley value. *SIAM/ASA Journal on Uncertainty Quantification*, 2:245–251.

Owen, A. & Prieur, C. (2017). On Shapley value for measuring importance of dependent inputs. *SIAM/ASA Journal on Uncertainty Quantification*, 5:986–1002.

Perez, D. (2022). An attempt of reproduction of sovacool et al.'s differences in carbon emissions reduction. *EPJ Nuclear Science & Technology*, 8:24.

Plischke, E., Rabitti, G., & Borgonovo, E. (2021). Computing Shapley effects for sensitivity analysis. *SIAM/ASA Journal on Uncertainty Quantification*, 9:14111437.

Razavi, S., Jakeman, A., Saltelli, A., Prieur, C., Iooss, B., Borgonovo, E., Plischke, E., Lo Piano, S., Iwanaga, T., Becker, W., Tarantola, S., Guillaume, J., Jakeman, J., Gupta, H., Melillo, N., Rabiti, G., Chabridon, V., Duan, Q., Sun, X., Smith, S., Sheikholeslami, R., Hosseini, N., Asadzadeh, M., Puy, A., Kucherenko, S., & Maier, H. (2020). The future of sensitivity analysis: An essential discipline for systems modelling and policy making. *Environmental Modelling and Software*, 137(104954).

Saltelli, A., Bammer, G., Bruno, I., Charters, E., Di Fiore, M., Didier, E., Espeland, W., Kay, J., Lo Piano, S., Mayo, D., Jr, R., Portaluri, T., Porter, T., Puy, A., Rafols, I., Ravetz, J., Reinert, E., Sarewitz, D., Stark, P., & Vineis, P. (2020). Five ways to ensure that models serve society: a manifesto. *Nature*, 582:482–484.

Saltelli, A., Chan, K., & Scott, E., editors (2000). *Sensitivity analysis*. Wiley Series in Probability and Statistics. Wiley.

Shapley, L. (1953). A value for n-persons game. In Kuhn, H. & Tucker, A., editors, *Contributions to the theory of games II, Annals of mathematic studies*. Princeton University Press, Princeton, NJ.

Sobol', I. (1993). Sensitivity estimates for non linear mathematical models. *Mathematical Modelling and Computational Experiments*, 1:407–414.

Song, E., Nelson, B., & Staum, J. (2016). Shapley effects for global sensitivity analysis: Theory and computation. *SIAM/ASA Journal on Uncertainty Quantification*, 4:1060–1083.

Sovacool, B., Schmid, P., Stirling, A., Walter, G., & MacKerron, G. (2020). Differences in carbon emissions reduction between countries pursuing renewable electricity versus nuclear power. *Nature Energy*, 5:928.

Thomas, D., Zumbo, B., Kwan, E., & Schweitzer, L. (2014). On johnson's (2000) relative weigths method for assessing variable importance: A reanalysis. *Multivariate Behavioral Research*, 49:329–338.

Wagner, F. (2021). CO2 emissions of nuclear power and renewable energies: a statistical analysis of european and global

data. *The European Physical Journal Plus*, 136:562.

Wallard, H. (2015). 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.

Wallard, H. (2019). Grouping property and decomposition of explained variance in linear regression. In Skiadas, C. & Bozeman, J., editors, *Data Analysis and Applications 1: Clustering and Regression, Modeling-estimating, Forecasting and Data Mining*, pages 73–89. Wiley.

Weber, R. J. (1988). Probabilistic values for games. In *The Shapley Value*, pages 101–120. Cambridge University Press, 1st edition.

Wei, P., Lu, Z., & Song, J. (2015). Variable importance analysis: a comprehensive review. *Reliability Engineering & System Safety*, 142:399–432.

Zhao, K. & Hastie, T. (2021). Causal interpretations of black-box models. *Journal of Business & Economic Statistics*, 39(1):272–281.

Zou, H. & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 67(2):301–320.

Supplementary Material 1 - "Other test cases" - of the paper "A review on variance-based importance measures in the linear regression context"

Laura Clouvel[1], Bertrand Iooss[2], Vincent Chabridon[2], Marouane Il Idrissi[2], and Frédérique Robin[1]

[1]EDF R&D, PERICLES Department, Saclay, France
[2]EDF R&D, PRISME Department, Chatou, France & SINCLAIR AI Lab., Saclay, France

1 Public dataset 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 1, 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.

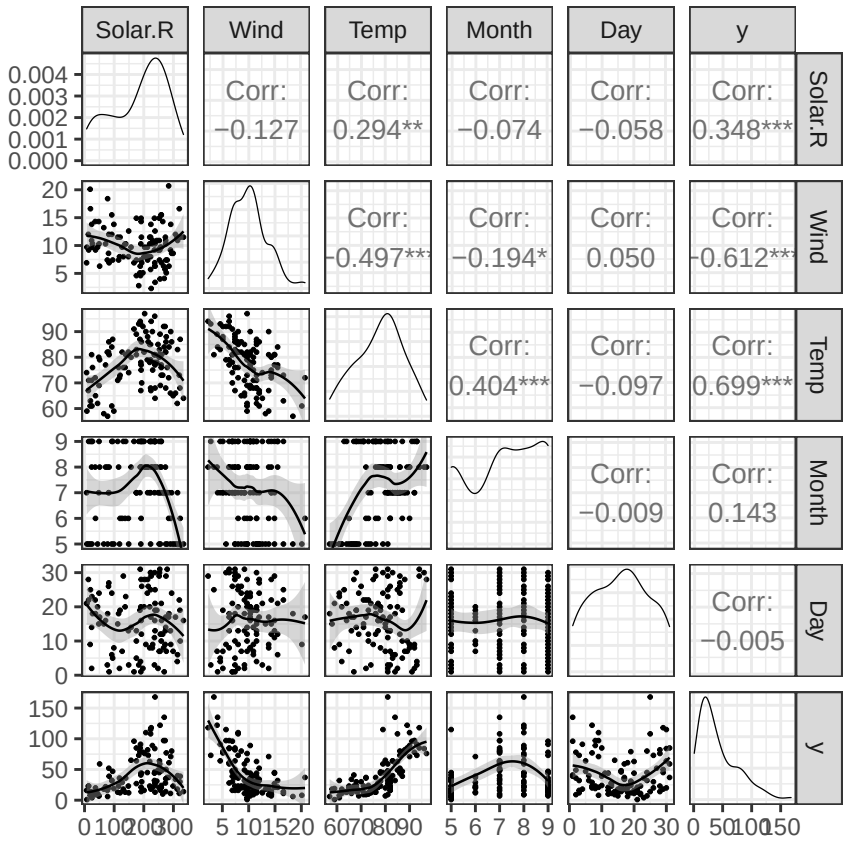


Figure 1: Data pairs plot for the air quality dataset.

The linear regression between the output and the inputs gives $R^2 = 0.625$ (see also Fig. 2) and $Q^2 = 0.582$. Table 1 and Figure 3 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.

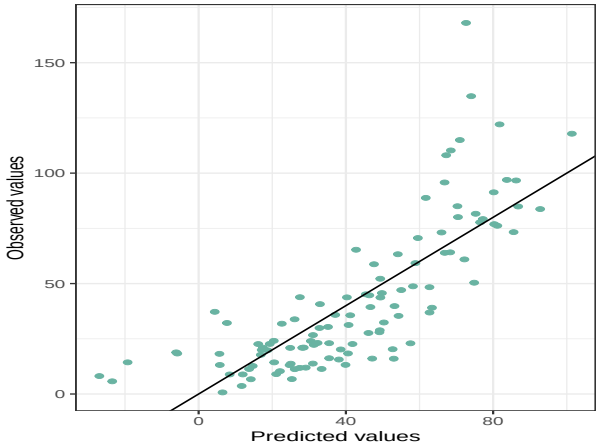


Figure 2: Linear model prediction vs. observation data for the air quality data.

Input	n°	VIF	SRC^2	PCC^2	$SPCC^2$	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 1: VIF and VM (in %) for the air quality data.

2 Public dataset 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: 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. 4) and $Q^2 = 0.911$. VM are given in Table 2 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.

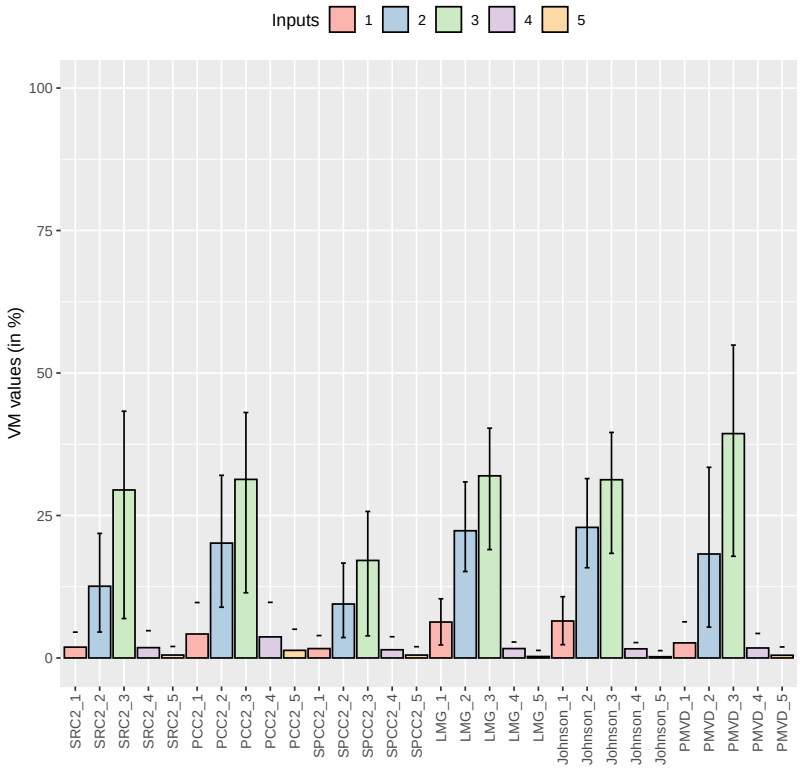


Figure 3: Estimates (with bootstrap) of the VM for the air quality dataset.

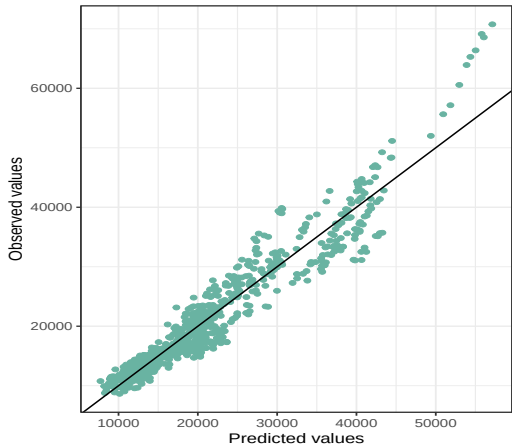


Figure 4: Linear model prediction vs. observation data for the cars data.

Input	n°	VIF	SRC^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 2: VIF and VM (in %) for the cars data. The last column gives the sense of variation of inputs with significantly influence (LMG> 1).

Supplementary Material 2 - "Classification case" - of the paper "A review on variance-based importance measures in the linear regression context"

Laura Clouvel[1], Bertrand Iooss[2], Vincent Chabridon[2], Marouane Il Idrissi[2], and Frédérique Robin[1]

[1]EDF R&D, PERICLES Department, Saclay, France
[2]EDF R&D, PRISME Department, Chatou, France & SINCLAIR AI Lab., Saclay, France

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. 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 (McCullagh & Nelder, 1989) 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:

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

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 (McCullagh & Nelder, 1989).

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:

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

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

Remark 1. *The logistic regression parameters (i.e. $\beta_i, i = 0, \ldots, d$ 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. Raguet & Marrel (0047); Marrel & Chabridon (2021)) 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., Zheng & Agresti (2000) for a review). A popular one is the following (Guisan & Zimmerman, 2000):

$$R^2 = 1 - \frac{D}{D_0} \tag{3}$$

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 (Tonidandel & LeBreton, 2010) but their study is beyond the scope of this report.

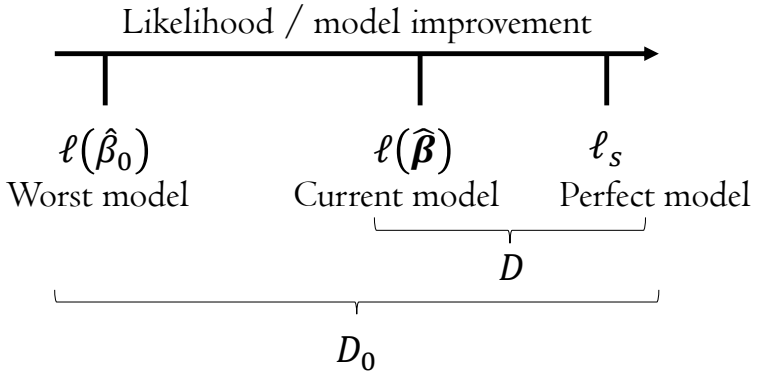


Figure 1: Illustration of deviance and null deviance for GLM validation (inspired from García-Portugués (2021)).

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 obser-

vations:

$$\varepsilon = \frac{\mathrm{FP} + \mathrm{FN}}{n} \ . \tag{4}$$

- 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:

$$\tau = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \ . \tag{5}$$

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 (Saporta, 1990):

$$\mathrm{CR}_j = \eta^2_{X_j|Y} = \frac{\mathrm{VAR}\,\mathrm{E}(X_j|Y)}{\mathrm{VAR}\,X_j} \tag{6}$$

where one can recognize a first-order Sobol' index (Sobol', 1993) formula. CR is also equivalent to the coefficient of determination (R^2) of the linear regression explaining the quantitative variable by the qualitative one (Saporta, 1990).

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

$$\widehat{\eta}^2_{X_j|Y} = \frac{n_0 n_1}{n} \frac{\left(\overline{X}_{j,0} - \overline{X}_{j,1}\right)^2}{\sum_{i=1}^{n} \left(X_j^{(i)} - \overline{X}_j\right)^2} \tag{7}$$

where n_0 and $\overline{X}_{j,0}$ (resp. n_1 and $\overline{X}_{j,1}$) 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 Menard (2004), Tonidandel & LeBreton (2010) 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

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

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

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

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

$$A_{logit} = (\mathbf{Z}^t\mathbf{Z})^{-1}\mathbf{Z}^t\boldsymbol{G}_{(\boldsymbol{p})} = \mathbf{Z}^t\boldsymbol{G}_{(\boldsymbol{p})} = (\alpha_{logit,j})_{1 \le j \le d}, \tag{10}$$

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

$$J_{logit,i} = R^2 \sum_{j=1}^{d} \alpha^{*2}_{logit,j} \pi^{*2}_{ij}. \tag{11}$$

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	$\breve{g}$	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:

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

with $k \in \mathbb{R}$ and $X_i \sim \mathcal{N}(0, 1)$ $i = 1, \ldots, d$. 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).

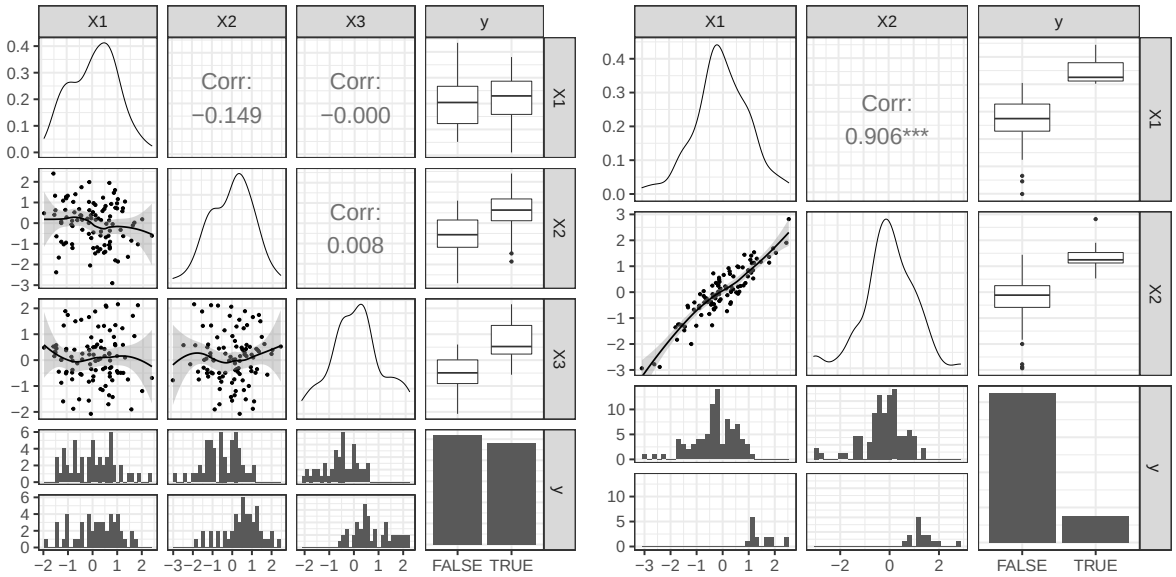


Figure 2: Data pairs plot for the linear classification case (left) and the dummy-correlated-variable classification case (right). The upper panel provides the CC of each variable pair; the diagonal panel gives the kernel density estimation of the marginals; the lower panel gives scatter plots and fitted 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^2 (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.

Input	VIF	CR	SRC^2	PCC^2	$SPCC^2$	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.

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

$$Y = \mathbb{1}_{X_1+\eta\geq 1} \tag{13}$$

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^2	PCC^2	$SPCC^2$	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.

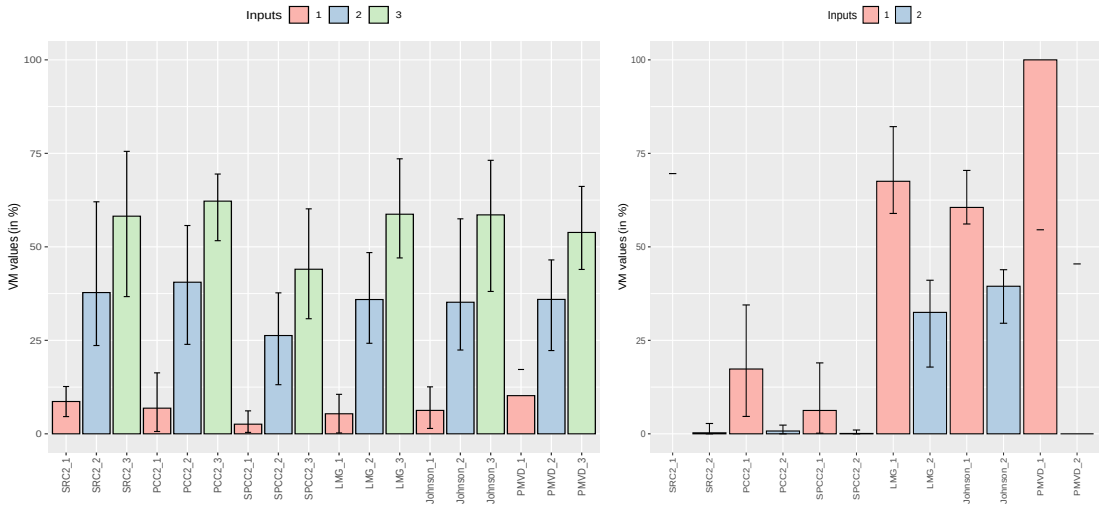


Figure 3: Estimates (with bootstrap) of the VM in the linear classification case (left) and in the *dummy-correlated-variable* classification case (right). SRC2_j corresponds to the SRC^2 of the input X_j (and so on for the other VM).

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^2	PCC^2	$SPCC^2$	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

García-Portugués, E. (2021). *Notes for Predictive Modeling*. `https://bookdown.org/egarpor/PM-UC3M/`. Version 5.9.0.

Guisan, A. & Zimmerman, N. E. (2000). Predictive habitat distribution models in ecology. *Ecological Modelling*, 135:147–186.

Marrel, A. & Chabridon, V. (2021). Statistical developments for target and conditional sensitivity analysis: Application on safety studies for nuclear reactor. *Reliability Engineering & System Safety*, 214:107711.

McCullagh, P. & Nelder, J. (1989). *Generalized linear models*. Chapman & Hall.

Menard, S. (2004). Six approaches to calculating standardized logistic regression coefficients. *The American Statistician*, 58(3):218–223.

Raguet, H. & Marrel, A. (2018, URL https://arxiv.org/abs/1801.10047). Target and conditional sensitivity analysis with emphasis on dependence measures. *Working paper*.

Saporta, G. (1990). *Probabilités, analyse de données et statistique*. éditions Technip.

Sobol', I. (1993). Sensitivity estimates for non linear mathematical models. *Mathematical Modelling and Computational Experiments*, 1:407–414.

Tonidandel, S. & LeBreton, J. (2010). Determining the relative importance of predictors in logistic regression: An extension of relative weight analysis. *Organizational Research Methods*, 13:767–781.

Zheng, B. & Agresti, A. (2000). Summarizing the predictive power of a generalized linear model. *Statistics in Medicine*, 19:1771–1781.