



Contribution ID: 28

Type: **not specified**

Variance Inflation Factor (VIF) Decomposition to investigate multicollinearity in experimental designs

As a reference frame, balanced factorial designs are used in this presentation, because these designs are orthogonal for all linear models, they can be used for. Orthogonality means that the experimental factors are mutually orthogonal (angles of 90°) and as such are independent and not correlated. As a consequence, the parameters of the fitted linear models are also independent, leading to straightforward interpretations of the parameters and increasing knowledge about the factors of influence on the system under study. If the primary interest is an as accurate as possible estimation of the predicted response, orthogonality is not a prerequisite for optimal efficiency expressed as goodness of fit. Nevertheless, balanced factorial designs are D- and G-optimal, providing both optimal variance properties for parameter and predicted response.

Multicollinearity in experimental designs arises when designs are unbalanced or when additional constraints are imposed on the experimental factors (fi designs for mixture systems). It is generally understood that least squares estimation can lead to misleading results when the independent variables are correlated. The multicollinearity becomes harmful, when estimation or hypothesis testing is more affected by the multicollinearity among the regressor variables than by the relationship between the dependent and the regressor variables. Multicollinearity can imply incorrect signs and large variances of parameter estimates, unreliable test statistics and variable selection criteria.

Most of the time the multicollinearity structure of a design is presented by the correlation matrix (heatmap), calculated from the extended design matrix. Although in principle the correlation matrix is showing the full multicollinearity structure, the structure is difficult to interpret. This talk proposes to investigate the multicollinearity by calculating the VIF's for each parameter to be estimated, before the experiment is carried out. Therefore, the unknown error variance of the dependent variable is put equal to 1. Next, the design matrix (X) is centered and scaled to unit length (W), to eliminate the effect of different coding on the calculations. As a result, the correlation matrix of the design can be calculated as $W'W$. Since the VIF's are equal to the head diagonal of the $(W'W)^{-1}$ matrix, a SVD on the W matrix ($W=ULV'$) results in the VIF's to be equal to the diagonal of $V'L^{-2}V$. Or on a per parameter basis each VIF_i can be decomposed into a sum of components, each associated with one and only one singular value.

With p the number of columns in X and W and thus the number of parameters to estimate, i the i th parameter, j the j th singular value. This decomposition shows how much one or more small singular values will inflate the parameter estimates. The corresponding eigenvectors of these small singular values are defining multicollinearity constraints as orthogonal linear combinations in the original centred and scaled experimental factors (W). As a consequence, the experimental factors that are most important in defining the multicollinearity can be investigated by evaluating the Principal Components of the $W'W$ matrix with the smallest corresponding eigenvalues.

Some examples of balanced and unbalanced factorial designs, as well as mixture designs are explored to demonstrate the theory.

Special/ Invited session

Classification

Both methodology and application

Keywords

Variance Inflation Factors, Design of Experiments, Multicollinearity

Primary author: Prof. SCHREVEENS, Eddie (KU Leuven, Belgium)

Presenter: Prof. SCHREVEENS, Eddie (KU Leuven, Belgium)

Track Classification: Design of Experiments