

Orthogonal regression for three-part compositional data via linear model with type-II constraints

E. FIŠEROVÁ¹ and K. HRON¹

¹Department of Mathematical Analysis and Applications of Mathematics - Palacký University Olomouc, Czech Republic, fiserova@inf.upol.cz

Orthogonal regression is a proper tool for fitting two-dimensional data points when errors occur in both the variables. This type of modelling technique is also called the total least squares (TLS) in the statistical literature. In its simplest form it attempts to fit a line that explains the set of n two-dimensional data points in such a way that the sum of squared distances from data points to the estimated line is minimal. Orthogonal regression is invariant under the orthogonal rotation of coordinates and thus it is convenient for regression analysis of three-part compositional data, performed after isometric logratio transformation.

The difficulty or even impossibility of deeper statistical analysis (confidence regions, hypotheses testing) using the standard solution for orthogonal regression based on maximum-likelihood method can be overcome by calibration line technique based on linear statistical models, namely linear models with type-II constraints (constraints involve in addition to the unknown model's parameters the other unobservable ones). The main advantage of the linear model approach is its validity for finite samples in contrast to the standard techniques. It means we can determine exact variances and covariances of estimated line's coefficients (for the standard technique we have only asymptotic variances and covariances). Further, under assumption of normality, we can make any standard statistical inference, e.g., construct confidence regions and bounds and test hypotheses that is then also easy to interpret on the simplex sample space. Consequently, we can apply various standard approaches to checking the model and its assumptions for adequacy and validity, e.g. coefficient of determination, residuals analysis or normality tests.

The only restrictive condition that must be fulfilled in order to ensure a meaningful analysis of compositional data, concretely invariance of the regression line's parameters in the sample space of compositions, the simplex, with respect to the choice of the orthonormal basis for the ilr transformation, concerns the covariance structure of variables that needs to be very simple (homoscedastic) in this case. Moreover, from the theory of linear statistical models it follows that estimation by linear models (least squares method) and the orthogonal regression give the same results under this condition.

The aim of the contribution is to present an iterative algorithm for estimating the regression line via linear models with type-II constraints and some statistical inference, together with the corresponding interpretation for compositional data. The theoretical results will be applied to real-world example.