

## IMPROVING THE CLASSICAL USES OF REGRESSION MODELLING

Thierry Foucart, University de Poitiers, France

Regression and correlation are among the most often employed statistical methods and then are taught in all the courses of statistics and on different levels. We are interested here in applied statistics. Teaching regression and correlation from this point of view requires the teacher to call users to be careful to the way of performing regression. In fact, since statistical softwares are easier to use and more powerful than ten years ago, many procedures of regression which were difficult to apply because of computations are now often available. That is the case with simulation methods, with explanatory variables selection, with influent cases detection and with collinearity effects measure. These methods bring up the lack of precision of estimates which is given sometimes by the rough method of minimum squared errors, and improve them by leading to another model, other explanatory variables systems or a biased estimator for example. So users have to be able to criticise the usual minimum squared errors estimator and the results which it gives, and eventually choose the appropriate method to amend their model and thus obtain better results. Many examples can be obtained here with simulation methods: we generate many samples of data with the same statistical parameters in the aim of comparing different regression coefficients systems. Introducing instability of estimates, Cook's distance or simulating predictions is very easy from these data. It is convenient to present the difficulties which can be encountered and then the way of eliminating them if possible. We present detailed examples with discussion about:

Detecting influent cases: jackknife's methods, Cook's distance;

Selecting explanatory variables: partial correlation coefficients, stepwise methods, minimisation of the unbiased residual variance estimate, Mallow's Cp;

Analysing collinearity effects: instability of estimates, principal component regression, ridge regression. In some cases simple figures make easy to understand why the rough method does not give satisfying results. A numerical example is developed in each case, on real or simulated data. Data and examples are selected from well-known reference books or created by simulation. All the computations are performed by an

available software running on DOS, which contains most of statistical methods, particularly in regression and the programs which are related to previously published papers about symmetric definite positive matrices.