

RECENT PHD THESES

Faculté Universitaire des Sciences Agronomiques (FUSAGx)

Planchon Viviane. *Outliers detection in mixtures of dissymmetric distributions for data sets with spatial constraints*. (2007) – Promotor: Prof. Jean-Jacques Claustrioux.

In the case of soil chemical analyses, frequency distributions for some elements show a dissymmetrical aspect, with a very marked spread to the right or to the left. A high frequency of extreme values is also observed and a possible mixture of several distributions, due to the presence of various soil types within a single geographical unit, is encountered. Then, for the outliers detection and the establishment of detection limits, an original outliers detection procedure has been developed; it allows estimating extreme quantiles above and under which observations are considered as outliers. The estimation of these detection limits is based on the right and the left of the distribution tails. A first estimation is realised for each elementary geographical unit to determine an appropriate truncation level. Then, a spatial classification allows creating adjoining homogeneous groups of geographical units to estimate robust limit values based on an optimal number of observations.

Université de Liège (ULg)

Magis David. *Influence, information and item response theory in discrete data analysis*. (4/5/2007) – Promotor: Prof. Gentiane Haesbroeck.

The main purpose of this thesis is to consider usual statistical tests for discrete data and to present some recent developments around them. Contents can be divided into three parts.

In the first part we consider the general issue of misclassification and its impact on usual test results. A suggested diagnostic examination of the misclassification process leads to simple and direct investigation tools to determine whether conclusions are very sensitive to classification errors. An additional probabilistic approach is presented, in order to refine the discussion in terms of the risk of getting contradictory conclusions whenever misclassified data occur.

In the second part we propose a general approach to deal with the issue of multiple sub-testing procedures. In particular, when the null hypothesis is rejected, we show that usual re-applications of the test to selected parts of the data can provide non-consistency problems. The method we discuss is based on the concept of decisive subsets, set as the smallest number of categories being sufficient to reject the null hypothesis, whatever the counts of the remaining categories. In this framework, we present an iterative step-by-step detection process based on successive interval building and category count comparison. Several examples highlight the gain our method can bring with respect to classical approaches.

The third and last part is consecrated to the framework of item response theory, a field of psychometrics. After a short introduction to that topic, we propose first two enhanced iterative estimators of proficiency. Several theoretical properties and simulation results indicate that these methods ameliorate the usual Bayesian estimators in terms of bias, among others. Furthermore, we propose to study the link between response pattern misfit and subject's variability (the latter as individual latent trait). More precisely, we present "maximum likelihood"-based joint estimators of subject's parameters (ability and variability); several simulations suggest that enhanced estimators also have major gain (with respect to classical ones), mainly in terms of estimator's bias.