METHODOLOGICAL CONTRIBUTION TO CONTROL HETEROSCEDASTICITY IN DISCRIMINANT ANALYSIS STUDIES

R. Glèlè Kakaï[1], R. Palm[2]

[1] Faculty of Agronomic Sciences, University of Abomey-Calavi, Bénin
[2] Gembloux Agricultural University, Belgique

Email : *gleleromain@yahoo.fr*

Let $\mathbf{Y}(\boldsymbol{\mu},\boldsymbol{\Sigma})$ denote a multivariate population with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. For two groups, we consider an heteroscedastic model constituted of $\mathbf{Y}_1(\mathbf{0},\mathbf{I})$ and $\mathbf{Y}_2(\mathbf{m},\mathbf{V})$ where $\mathbf{m} = (m_1, 0, ..., 0)'$ and $\mathbf{V}$, a diagonal matrix of vector $\mathbf{v}$ of p diagonal elements so that $\mathbf{v}_{ii} = \lambda$ (>0) for $i = 1, \ldots k$ and $\mathbf{v}_{ii} = 1$ for $i = k+1, \ldots, p$ ($k \le p$). This simple model allows, by linear transformations, to extend the results of discriminant analysis studies to a large variety of real world problems. To control the heteroscedasticity of the model, a parameter Γ is considered and defined, for two covariances matrices $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$, as $\Gamma = -\sum_{i=1}^{2} \ln(|\boldsymbol{\Sigma}_i| / |\boldsymbol{\Sigma}|)$, where $\boldsymbol{\Sigma}$ is the pooled covariance matrix of the model. In the case of the populations $\mathbf{Y}_1(\mathbf{0},\mathbf{I})$ and $\mathbf{Y}_2(\mathbf{m},\mathbf{V})$ or their linear transformations, we show analytically that the parameter Γ can be expressed as a function of k and λ. Γ is considered as a measure of heteroscedasticity of the discriminant model and some attractive properties of the function $\Gamma(\lambda, k)$ are given. We discuss also about the choice of m_1 and Γ values for the sampling scheme related to Monte Carlo discriminant analysis studies. Since it is possible to compute Γ on data samples, the results of Monte Carlo studies related to discriminant analysis can be expressed as a function of the heteroscedasticity observed on the data samples.