

ESTIMATORS OF THE OVERALL MISCLASSIFICATION PROBABILITY IN DISCRIMINANT ANALYSIS

Altemir da Silva BRAGA¹
Daniel Furtado FERREIRA¹
Patrícia de Siqueira RAMOS¹

- **ABSTRACT:** Discriminant analysis is one of the multivariate statistics techniques which idea consists in classifying new individuals in one of several populations known *a priori*. Thus, several estimators for the parametric overall misclassification probability (OMP) were proposed, using jackknifing methods and whose performance was assessed through Monte Carlo simulation. In the present work, the performance of OMP₁, OMP₂, OMP₃, OMP₄, OMP₅ and OMP₆ estimators was compared for two homoscedastic multivariate normal populations, considering the same costs of misclassification and *priori* probabilities. The first one is Lachenbruch & Mickey's method (1968), based on *Jackknife* methods, the second one was derived from Lachenbruch & Mickey's method (1968), using a common variance estimator into the function which estimates OMP. Third and fourth methods were proposed in the present work, in which Lachenbruch & Mickey's method (1968) was been modified, associating Fisher's linear function with *Jackknife* methodology. Fifth and sixth methods were derived using the same previous reasoning, setting the linear combination vector Γ_1 of the variates and applying the *Jackknife* for the constant of the Fisher's linear combination. The performance was assessed through bias and quadratic mean square estimator. Thus, the mean vector from population π_1 was set to $\mathbf{0}$ ($\mu_1 = \mathbf{0}$). The approximate search of μ_2 from population π_2 , for a settled value of the Mahalanobis distance Δ^2 , was accomplished by trial and error. For population π_1 , the sampling sizes were $n_1 = 10, 50, 100$ and for π_2 , $n_2 = 10, 50, 100$ that were factorially combined with $p = 2$ and 10 variates and correlation coefficient $\rho=0, \rho=0.5$ and $\rho=0.9$. The estimators OMP₅ and OMP₆ underestimated OMP, whereas OMP₁, OMP₂, OMP₃ and OMP₄ overestimated it. The OMP₃, OMP₄, OMP₅ and OMP₆ estimators were more efficient than OMP₁ one, originally proposed by Lachenbruch & Mickey (1968). The OMP₃ estimator with heterogeneous variance estimators was considered optimum, due to the smallest positive bias.
- **KEYWORDS:** Multivariate; estimator; classifying; simulation; *Jackknife*.

¹Departamento de Ciências Exatas - DEX, Universidade Federal de Lavras -- UFLA, Caixa Postal 3037, CEP: 37200000, Lavras, MG, Brasil. E-mail: altemirbraga@hotmail.com / danielff@ufla.br