

Análise de Arquétipos como um método de imputação para dados faltantes

Pórtia Piscitelli Cavalcanti ¹, Matthew Frederick Parry ², Bryan Frederick John Manly ³ Carlos Tadeu dos Santos Dias ⁴

Resumo: *Esse trabalho tem por objetivo avaliar o uso de arquétipos na imputação de valores em um conjunto de dados multivariados sobre o golfinho de Hector, endêmico da Nova Zelândia, com dados faltantes (missing data) simulados. O conjunto de dados consiste de seis medidas corporais de 59 esqueletos de golfinhos das Ilhas Norte e Sul. Foram simulados dados faltantes completamente ao acaso (missing completely at random, MCAR) da seguinte forma: nove observações foram removidas aleatoriamente, das quais foram removidos valores das variáveis resposta também aleatoriamente, considerando que o número de valores removidos segue a distribuição de probabilidade Binomial truncada em zero. Assim, a Análise de Arquétipos (AA) foi aplicada nos dados completos padronizados e, a partir dos três arquétipos encontrados, foram imputados os valores faltantes. Os resultados da utilização da AA como um método de imputação para dados faltantes do tipo MCAR foram promissores e as medidas corporais dos golfinhos foram imputadas com sucesso via arquétipos.*

Palavras-chave: arquétipos; dados imputados; estatística multivariada.

¹Department of Mathematics & Statistics - University of Otago/ Departamento de Ciências Exatas - ESALQ/USP. e-mail: portia@usp.br

²Department of Mathematics & Statistics - University of Otago. e-mail: mparry@maths.otago.ac.nz

³Manly-Biostatistics Limited. e-mail: BryanManly@xtra.co.nz

⁴Departamento de Ciências Exatas - ESALQ/USP/ Departamento de Ciências do Solo - CCA/UFC. e-mail: ctsdias@usp.br