

## ANÁLISES DISCRIMINANTES NÃO PARAMÉTRICAS APLICADAS AO ESTUDO DA DIVERSIDADE GENÉTICA BASEADO EM DADOS FENOTÍPICOS QUANTITATIVOS

Marcileia Santos SOUZA<sup>(1)</sup>, Fábio Medeiros FERREIRA<sup>(2)</sup>.

1. Mestranda pelo Programa de Pós-graduação em Agronomia Tropical, email: marcileia\_souza@hotmail.com. 2. Professor Adjunto III – UFAM/Itacoatiara, email: ferreirafmt@ufam.edu.br.

**Resumo:** Métodos multivariados de análises discriminantes visam identificar as populações nas quais um indivíduo deva pertencer, admitindo previamente, que o indivíduo compõe uma das populações avaliadas. Métodos baseados em funções discriminantes lineares têm sido usados nos estudos preditivos da diversidade no melhoramento genético, quando os dados são fenotípicos quantitativos. Entretanto, este tipo de análise pressupõe a multinormalidade das populações. Objetivou-se avaliar a efetividade das análises discriminantes não paramétricas – que não requerem probabilidade específica de distribuição – quando aplicadas a variáveis quantitativas. Dois conjuntos de dados foram utilizados: i) 83 matrizes de pupunheira, previamente alocadas em três raças primitivas, para sete variáveis do fruto; ii) 122 clones de cafeeiro, previamente alocados entre três variedades botânicas, para dez características agrônomicas. Avaliou-se os métodos não paramétricos do “vizinho médio” e dos “k-vizinhos mais próximos” sob vários cenários, conforme combinações possíveis entre técnica de análise não paramétrica  $x$  medida de distância genética  $x$  valor de  $k$   $x$  probabilidade *a priori* dos genótipos pertencerem as populações. Comparou-se a alocação dos genótipos nos diferentes cenários e com aquela obtida pelas funções discriminantes de Anderson (considerada padrão) a partir das taxas de erro aparente globais (TEA) de classificação dos indivíduos nas respectivas populações. Utilizou-se o software GENES. No geral, o método do “vizinho médio” e dos “k-vizinhos mais próximos” apresentaram TEA’s similares a análise discriminante de Anderson. As TEA’s pouco variaram em função das probabilidades definidas *a priori* para as populações. O “k-vizinhos mais próximos” exibiu TEA’s menores quando comparado ao “vizinho médio” para a população de pupunha e maiores para as de café. As medidas de distância genética que apresentaram menores TEA’s foram a distância Quadrado da distância média e a distância de Gower. Houve uma leve tendência de ao se elevar o valor de  $k$ , as TEA’s tenderem a pouco variar aumentando ou diminuindo, independente da medida de distância e probabilidades *a priori*. Conclui-se que os métodos não paramétricos foram efetivos para classificar os genótipos em suas respectivas populações; o método dos “k-vizinhos mais próximos” acertou mais em classificar os genótipos, quando comparado ao “vizinho médio”; não há diferenças expressivas de classificação entre as medidas de distância estudada.

**Palavras-chave:** divergência genética; vizinho médio; k-vizinhos mais próximos.