Appendix to
Semantical Communication in Anti-predator Alarm Calls

A Significance Test for the Generalized Mahalanobis-squared Distance

Russell O. Davis
Division of Reproductive Biology and Medicine
Department of Obstetrics and Gynecology
School of Medicine
University of California
Davis, California 95616

The Mahalanobis-squared distance (D^2), commonly calculated in multivariate-discriminant analysis, is the distance between pairwise comparisons of the multivariate means of populations. As such, it provides a quantitative estimate of the relative separation of groups. However, such distances may not be significant, i.e., the hypothesized groups may all belong to the same population. Therefore, it is desirable to develop a significance test for the generalized Mahalanobis distance.

By assumption, the estimates of all p variates in all k groups are independent and normally distributed with mean μi and covariance matrix Σ. The covariance matrix is assumed to be equal for all groups. Thus,

$$Y_{ij} \sim N(\mu_i, \Sigma), \qquad [1]$$

where, i=1,...,k=the group and j=1,...,N_i=case in the group.

Therefore, the multivariate mean of the ith group is distributed as:

$$Y_i \sim N(\mu_i, 1/N_i \Sigma). \qquad [2]$$

The difference between the means of two groups is distributed as:

$$Y_i - Y_j \sim N(\mu_i - \mu_j, (1/N_i + 1/N_j)\Sigma) \qquad [3]$$

Therefore, Yi is independent of Yj, and:

$$W = ((N_i N_j)/(N_i + N_j)) (Y_i - Y_j) \sim N(\mu_i - \mu_j, \Sigma). \qquad [4]$$

If $\mu_i = \mu_j$ (the null hypothesis), then:

$$W \sim N(0, \Sigma), \qquad [5]$$

so that:

$$W'\Sigma^{-1}W \sim X^2(p). \qquad [6]$$

Since the covariance matrix must be estimated,

$$W'\Sigma^{-1}W \sim F(p, \xi) \qquad [7]$$

where, $\xi = N_i - k$.

Since ξ is typically large, it approximately holds that:

$$W'\Sigma^{-1}W \sim X^2(p) \qquad [8]$$

Therefore,

$$W'\Sigma^{-1}W = ((N_i N_j)/(N_i + N_j))(Y_i - Y_j)'\Sigma^{-1}(Y_i - Y_j) \qquad [9]$$

$$= ((N_i N_j)/(N_i + N_j))D^2, \qquad [10]$$

where D^2 is the Mahalanobis-squared distance. Thus, the critical value of D^2 for p=0.05 is:

$$D^2(\text{crit}) = X^2 0.05(p)(N_i + N_j)/(N_i N_j), \qquad [11]$$

where (p) is the number of variates used in the analysis.

An example of such a calculation follows:

if N_1=10, N_2=25, and 24 variates are used, then $X^2_{0.05(24)}$=36.42 and the critical value for D^2 is:

$$(35/250)(36.42)=5.10. \qquad [12]$$

D^2 values that equal or exceed 5.10, and whose population sizes are 10 and 25, indicate that the group means are significantly different at p=0.05.