Classification of Aedes Adults Mosquitoes in Two Distinct Groups Based on Fisher Linear Discriminant Analysis and FZOARO Techniques

Friday Zinzendoff Okwonu [1,2*], Hamady Dieng [3], Abdul Rahman Othman [2] and Ooi Seow Hui [3]

1. School of Distance Education, Univesiti Sains Malaysia,11800,Pulua,Pinang,Malaysia
2. Department of Mathematics and Computer Science, Delta State University, P.M.B. 1, Abraka, Nigeria
3. School of Biological Science, Universiti Sains Malaysia,11800,Pulau Pinang, Malaysia

* E-mail of the corresponding author: okwonu4real@yahoo.com

Abstract

This paper describes the breeding, feeding and measurement of Aedes mosquitoes based on body size (wing length). Due to similarity in body size measurements, we were constrained on gender recognition. To reveal the gender identity of these mosquitoes, Fisher linear discriminant analysis and FZOARO classification models were considered suitable for prediction and classification. We randomly selected 15 mosquitoes from each groups and categorize the body size as small and large and applied the classification procedures. Both classification techniques perform similar. The numerical simulation reveals that 86.67% were classified as male for group one and 80% were correctly classified as female in group two.

Keywords: Fisher linear discriminant analysis; FZOARO; Classification.

1. Introduction

Linear discriminant analysis (LDA) is one of the most widely used dimension reduction technique due to its simplicity and effectiveness. Fisher (Fisher R. A., 1936) introduced his linear discriminant analysis approach to analyze the iris data set. Fisher linear discriminant analysis (FLDA) is a conventional learning model based statistical classifier designed to allocate an unknown observation vector to one of two multivariate Gaussian populations with different mean vectors and common covariance matrix (Sarunas R., & Duin, R. P. W. 1998). Classical Fisher linear discriminant analysis was originally proposed for two groups but it has however been applied and extended to more than two groups (Barker, M., & Rayens, W., 2003; Deirdre, T., Gerard, D., & Thomas, B. M. 2011; Rao, C. R., 1948; Reynes, S. de Souza, 2006). FLDA finds application in real word lower dimensional data set (Belhumeur, P. N., Hespanha, J. P., & Kriengman, D. J. 1997; Cevikalp, H., Neamtu, M., Wilkes, M., & Barkana, A. 2005; Daniel, L. S., & Weng, J. 1996; Jian, Y., Lei ,Z., Jing-yu, Y., & Zhang, D. 2011; Liu, C. J., & Wechsler, H. 2000; Ye, J. et al 2004; Yu, H., &Yang, J. 2001). Thus, it performs poorly if the data set has more dimension than the sample size (Alok ,S., & Kuldip, K. P. 2012; Marcel, K., Hans -Gunter, M., Hans, D. & Dietrich, K. 2002; Zhao, Z., &Tommy, W. S. C. 2012). FZOARO classification approach is a modified version of FLDA that incorporate compensate constant into the discriminant coefficient. It has comparable classification performance with FLDA. The optimality of these techniques lies on classical assumptions of normality and equal variance covariance matrix and the size of the data dimension(Alvin, C. R. 2002; Joseph, F. H., Jr, 1998; Maurice, M. T., & Lohnes, R. P. 1988). In this paper our focus is to apply these techniques to classify and predict laboratory breeded Aedes mosquitoes using their body size (wing length) as the predictor variables. Based on information obtained from body size measurement, the predictor variables are categorized as large and small, this information are applied to learn and validate the model to obtain gender classification and prediction.

This paper is organized as follows. Fisher linear discriminant analysis is presented in section two. Section three contains FZOARO classification model. Breeding procedure and body size measurement is described in section four. Simulations and conclusions are presented in sections five and six respectively.

2. Fisher Linear Discriminant Analysis

Fisher suggested transforming multivariate observations to univariate observations such that the univariate observations derived from each population is maximally separated. The separation of these univariate observations can be examined by their mean difference(Richard, A. J., & Dean W. W. 1998; Richard, A. J., & Dean, W. W. 2007). Fisher classification rule maximizes the variation between samples variability to within samples variability (Alvin, C. R. 2002; Fukunaga, K., 1991; Johnson, R. A., & Dean,W. W. 1998; Maurice, M. T., & Lohnes, R. P. 1988; Peter, A. L. 1975; William, R. D.,& Goldstein, M. 1984).

Consider classifying an observation vector $\mathbf{x}$ into one of two populations say $\pi_i : N_p(\mu_i, \Sigma), (i = 1, 2)$, the population mean vectors and covariance matrix are denoted as μ, Σ respectively. Since the population mean vectors and covariance matrices are unknown, the sample equivalent is applied in this paper. We define the sample mean vectors, sample covariance matrices and pooled sample covariance matrix as follows;

$$\bar{\mathbf{x}}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \mathbf{x}_{ij},$$

$$S_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \left(\mathbf{x}_{ij} - \bar{\mathbf{x}}_i\right)\left(\mathbf{x}_{ij} - \bar{\mathbf{x}}_i\right)',$$

$$S_{pooled} = \frac{\sum_{i=1}^{g=2} (n_i - 1) S_i}{\sum_{i=1}^{g=2} n_i - g}.$$

Where $\mathbf{X}$ is the independent predictor multivariate variable, $\mathbf{x}_{ij}$ denotes the jth training sample in group i, g denote number of groups, $n_i = \frac{N}{g}, \left(N = \sum_{i=1}^{g} n_i\right)$ is the number of equal size training samples in each group i and $\bar{\mathbf{x}}_i$ is the within group means. Using the definitions of the above parameters, Fisher linear discriminant analysis(Fisher R. A., 1936) can be started as follows.

$$\Omega = C'X = \left(\overline{\mathbf{x}}_1 - \overline{\mathbf{x}}_2\right)' S_{pooled}^{-1} X, \tag{1}$$

$$\overline{\Omega} = (\overline{\mathbf{x}}_1 + \overline{\mathbf{x}}_2) C / 2, \tag{2}$$

Where Ω denote discriminant score, $C = \left(\overline{\mathbf{x}}_1 - \overline{\mathbf{x}}_2\right)' S_{pooled}^{-1}$ is the discriminant coefficient, and $\overline{\Omega}$ is the discriminant mean. The classification rule based on equations (1) and (2) can be described as follows

assign $\mathbf{x}_1$ to population one π_1 if

$$\Omega \geq \overline{\Omega},$$

otherwise assign $\mathbf{x}_1$ to population two π_2 if

$$\Omega < \overline{\Omega},.$$

Fisher maintained that his techniques most adhere strictly to equal variance covariance matrix of the two normal populations. FLDA has been extended to more than two groups(Fisher R. A., 1936; Rao C. R., 1948).

3. FZOARO Classifical Model

We propose a comparable classification and dimension reduction technique called FZOARO classification model(Okwonu, F. Z., & Othman, A. R. 2012). The proposed technique is developed based on the following assumptions; the sample size is greater than the sample dimension, secondly, the variance covariance matrices are equal and are obtained from multivariate normal distribution. Since the population parameters are unknown, the sample equivalent is applied. Using the definitions of sample mean vectors, sample covariance matrices and pooled sample covariance matrix define in section two; the proposed approach is stated mathematically as follows

$$Z_{emonu} = \eta_{\omega} \mathbf{X} = (\mathbf{w}' \delta_{orogu}) \mathbf{X}, \tag{3}$$

where

$$\eta_{\omega} = \mathbf{w}' + \delta_{orogu},$$

$$\mathbf{w} = \left(\overline{\mathbf{x}}_{emu1} - \overline{\mathbf{x}}_{emu2}\right)' S_{pooled}^{-1},$$

$$\delta_{orogu} = \left(\frac{\sum S_{pooled}^{-1}}{\sqrt{\chi_p^2 \alpha = h}} \right),$$

$$h = \frac{3n}{4n},$$

$$\bar{Z}_{emonu/orogu} = \left(\frac{(\bar{\mathbf{x}}_{emu1} + \bar{\mathbf{x}}_{emu2})\eta_{\omega}}{2} \right). \tag{4}$$

where Z_{emonu} denote the discriminant score, $\mathbf{w}$ is the discriminant coefficient, δ_{orogu} internally generated constant that enhances the robustness of this approach and $\bar{Z}_{emonu/orogu}$ is the discriminant mean. The classification procedure is performed by comparing the discriminant score with the discriminan mean. That is,

Classify $\mathbf{x}_{obs1}$ to population one ϖ_1 if

Z_{emonu} is greater than or equal to $\bar{Z}_{emonu/orogu}$

or classify $\mathbf{x}_{obs1}$ to population two ϖ_2 if

Z_{emonu} is less than $\bar{Z}_{emonu/orogu}$

This approach is relatively computationally complex than FLDA because equation (3) comprises of a constant which also requires Chi-square value. This constant is considered to have dual functions one as a stabilizer and two as control factor.

4. Breeding, Feeding and Body Size Measurement Process

This section describes the breeding, feeding and measurement of Aedes adult mosquitoes based on their wing length. The larvae are categorized into two distinct groups or populations $\Pi_i, (i = 1, 2)$. Population one Π_1 contains 100 larvae and population two Π_2 doubles population one. Group one, 100 newly hatched larvae (100 L1) were feed with 3gram larvae food in 100ml dechlorinated water throughout the developmental stage for group one. Group two, 200 newly hatched larvae (200 L1) were feed with 1.5 gram larvae food in 100ml dechlorinated water throughout the developmental stage for group two. The variation in feeding pattern is due to population size. The volume of water in rearing tray is about 800ML to one liter respectively. Tables 1 and 2 describe the feeding regiment and water renewal pattern during developmental stage(s), see tables 1 and 2 below.

4.1 Determination of Body Size

Wing length was used as a proxy for body size. Because both wings on an individual mosquito has been shown to be of statistically similar size (Gleiser, R. M., Gorla, D. E. & Schelotto, G. 2000; Jonathan, M. C., & Tiffany, M. K. 2003), consequently, here, one wing, either the left or the right, from each individual adult mosquito was used to measure its length. Mosquitoes were adhered to a glass slide and wings measured with an ocular micrometer in a binocular microscope (Microscope name) from the apical notch to the axillary margin, excluding the wing fringe. 15 adult Aedes mosquitoes were randomly selected in each group and the wing length measured. In each group, we categorize the body sizes as large $\mathbf{x}_1$ and small $\mathbf{x}_2$. Our focus is to obtain the number of male and female in each group based on their body size measurement. The predicament we are confronted with is that each group size measurement is similar. Based on this gender difficulty, we apply multivariate statistical classification techniques to classify the two groups using the body size measurements. Based on previous discussions in sections two and three, Fisher linear discriminant analysis (FLDA) and FZOARO classification techniques are better options for this purpose.

5. Simulation

Numerical simulation is performed to classify and predict breeded Ades mosquitoes into male and female using Fisher linear discriminant analysis and FZOARO classification techniques. Group one contains 100 unclassified Aedes mosquitoes and group two contains 200 unclassified mosquitoes. In each group we randomly selected 15 adult mosquitoes and the wing length was measured to determine the body size. The preferred predictor variables are $\mathbf{x}_1$ for large body size and $\mathbf{x}_2$ for small body size. We are constrained to identify male and female. Based on these data set we apply Fisher's linear discriminant analysis and FZOARO classification models. Both models were applied and the classification and prediction results are presented in table 3. Table 3 reveals that 13 male were correctly classified to belong to group one and two male were misclassified to group two. On the other hand, three female were wrongly classified to group one and 12 female were correctly classified to group two. Table 3 also contains the percentage in bracket of correct classifications for each group. Table 4 indicate the hit-ratio which explain the percentage of correct classifications, both techniques performs exactly with 83.33% correct classification and 16.67% misclassifications. To confirm the performance of these procedures, figure 1 show that 2 male mosquitoes were wrongly classified as female mosquitoes and one of the male mosquitoes is seen as an outlier. Figure 1 also affirms that 3 female mosquitoes were wrongly classified as male and 12 were correctly classified. Figure 2 illustrates scatter plot of data set obtained from our experiment. It reveals the separation between male and female mosquitoes. Tables 3 and 4 and figure 1 affirms correct prediction of the mosquitoes into male and female categories. This work reveals that information obtained from body size measurement (wing length) of the mosquitoes can be used to classify and predict the gender of Aedes breeded mosquitoes.

6. Conclusions

In this paper, we applied Fisher linear discriminant analysis and FZOARO classification models. As an application, the above mentioned techniques were used to classify laboratory breeded Aedes mosquitoes based on their wing length measurement. Numerical simulation reveals that 86.67% were correctly classified as male and

80.00% as female in each group. The numerical experiment reveals that mosquito wing length can be used to predict and classify mosquitoes as male or female.

References

Alok, S., & Kuldip, K. P. (2012). A two-stage linear discriminant analysis for face recognition. *Pattern recognition letters, 33*, 1157-1162.

Alvin, C. R. (2002). *Methods of Multivariate Analysis(3rd ed)*: A John Wiley & Sons, Inc.

Barker, M., & Rayens, W. (2003). Partial least squares for discrimation. *Journal of Chemometrics, 17*, 166-173.

Belhumeur, P. N., Hespanha, J. P., & Kriengman, D. J. (1997). Eigenfaces vs Fisherfaces:recognition using class specific linear projection. *IEEE Transactions on pattern analysis and machine intelligence, 19*(7), 711-720.

Cevikalp, H., Neamtu, M., Wilkes, M., & Barkana, A. (2005). Discriminantive common vectors for face regonition. *IEEE Transactions on pattern analysis and machine intelligence, 27*(1), 4-13.

Daniel, L. S., & John, W. (1996). Using discriminant eigenfeatures for image retrieval. *IEEE Transactions on pattern analysis and machine intelligence, 18*(8), 831-836.

Deirdre, T., Gerard, D., & Thomas, B. M. (2011). Semi-Supervised linear discriminant analysis. *Journal of Chemometrics, 25*(12), 624-630.

Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics, 7*, 179 - 188.

Fukunaga, K. (1991). *Introduction to statistical pattern recognition* (2nd ed.): New York :Academic.

Gleiser, R. M.,Gorla, D. E., & Schelotto, G. (2000). Population dynamics of Aedes albifasciatus (Diptera: Culicidae) South of Mar Chiquita Lake, Central Argentina. *J. Med. Entomol, 37*, 21–26.

Jian, Y., Lei, Z., Jing-yu, Y., & Zhang, D. (2011). From classifiers to discriminators:A nearest neigbor rule induced discriminant analysis. *Pattern recognition, 44*, 1387-1402.

Johnson, R. A., & Wichen, D. W. (1998). *Applied Multivariate Statistical Analysis* (4th ed.): Prentice Hall:New York.

Jonathan, M. C., & Tiffany, M. K. (2003). Drought-induced mosquito outbreaks in wetlands. *Ecology Letters,, 6*, 1017–1024.

Joseph, F. H., Jr., Ronald, L. T., & William, C. B. (1998). *Multivariate Data Analysis* (5th ed.): Prentice Hall, Upper Saddle River, New Jersey.

Liu, C. J., & Wechsler, H. (2000). Robust coding schemes for indexing and retrieval from large face databases. *IEEE Transactions on image processing, 9*(1), 132-137.

Marcel, K., Hans -Gunter, M., Hans, D., & Dietrich, K. (2002). Robustness of linear discriminant analysis in automatic speech recognition. *ICPR'02 proceedings of the 16th international conference on pattern recogntion(ICPR'02),IEEE computer society, 3*, 1-4.

Maurice, M. T., & Lohnes, R. P. (1988). *Multivariate Analysis* (2nd ed.): Macmillan Publishing Company, New York.

Okwonu, F. Z., & Othman, A. R. (2012). A Model Classification Technique for Linear Discriminant Analysis for Two Groups. *IJCSI , 9*(3),125-128.

Peter, A. L. (1975). *Discriminant Analysis*: Hafner Press.

Rao, C. R. (1948). The ultilization of multiple measurements in problems of biological classification. *Journal of royal statist. soc.,, 10*(B), 159-193.

Reynes, C., Sabrina, de S, Robert, S., Giles, F., & Bernard, V. (2006). Selection of discriminant wavelenght intervals in NIR spectrometry with genetic algorithms. *Journal of Chemometrics, 20*, 136-145.

Richard, A. J., & Dean, W. W. (1998). *Applied Multivariate Statistical Analysis* (4th ed.): Prentice HaLL International Editions.

Richard, A. J., & Dean, W. W. (2007). *Applied multivariate statistical analysis* (6th ed.): Pearson Prentice Hall, Upper Saddle River.

Sarunas, R., & Duin, R. P. W. (1998). Expected classification error of the Fisher linear classifier with Pseudo-inverse covariance matrix. *Pattern recognition letter, 19*, 385-392.

William, R. D., & Goldstein, M. (1984). *Multivariate Analysis Methods and Applications*: John Wiley & Sons.

Ye, J., Janardan, R., Park, C., & Park, H. (2004). An optimization criterion for generalized discriminant analysis on undersampled problems. *IEEE Transactions on pattern analysis and machine intelligence, 26*(8), 982-994.

Yu, H., & Yang, J. (2001). A direct LDA algorithm for high dimensional data with application to face recognition. *Pattern recognition, 34*(10), 2067-2070.

Zhao, Z., & Tommy, W. S. C. (2012). Robust linearly optimized analysis. *Neurocomputing, 79*, 140-157.

Table 1. Feeding Regiment of Aedes Adult Mosquitoes in Two Distinct Size Classes

GROUPS	DAY1(D1)	DAY 3(D3)	DAY5(D5)	DAY6(D6)
100L 1/3 gram	3 ml	6ml	6ml	6ml
200 L 1/1.5 gram	3 ml	3 ml	3 ml	0

Table 2. Water Renewal Pattern before Feeding

Groups	Renewal 1	Renewal 2
100 L 1/3 gram	Before adding D3 food	Before adding D6 food
200 L 1/1.5 gram	Before adding D3 food	Before adding D6 food

Day 1 (D1) = Day of the appearance of larvae following egg flooding

Day 3 (D3) = 3rd day after the appearance of L1 (newly hatched larvae)

Day 5 (D5) = 5th day after the appearance of L1 (newly hatched larvae)

Day 6 (D6) = 6th day after the appearance of L1 (newly hatched larvae)

Table 3. Confusion Matrix For Aedes Body Size Measurement

Actual group membership	Predicted group membership	
	Group one (Male)	Group two (Female)
Group one (Male)	13 (86.67%)	2 (13.33%)
Group two(Female)	3 (20.00%)	12 (80.00%)

Table 4. Hit-Ratio/Methods

Methods	CFISHER	FZOARO
Observations Correctly Classified (%)	83.33*(16.67)*	83.33(*16.67*)

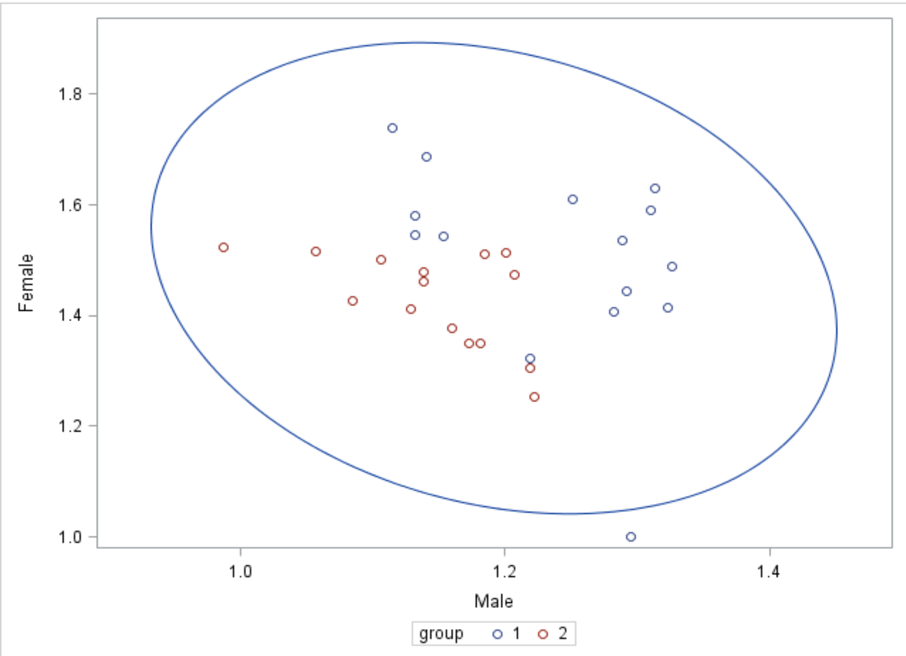

Figure 1. 97.5% Ellipse of Aedes Body Size Measurement

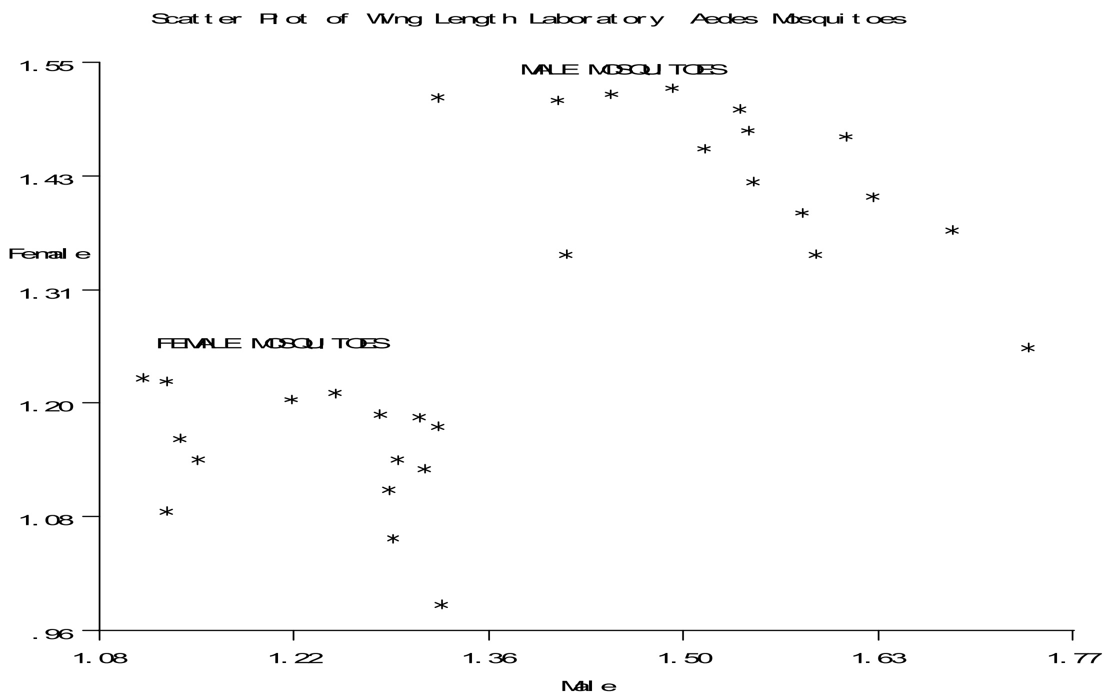

Figure 2. Scatter Plot of Wing Length Measurements of Laboratory Breeded Aedes Mosquitoes