Advances in Natural Science *ISSN 1715-7862* *Vol.2, No.1, 2009*
Canadian Research & Development Center of Sciences and Cultures *30/06/2009*
Http://www.cscanada.org Http://www.cscanada.net
E-mail: ans@cscanada.org; caooc@hotmail.com

Discriminant Analysis of Animal Species Odor's Response

Hezlin Aryani Abd Rahman[1]

Haslinda Ab. Malek[2]

Muniroh Mohd Fadzil[3]

Kamaruzaman Jusoff[4]

Abstract: The basis of our study is to identify the discriminating groups that are present in the observations as well as looking into the details of the classification of the observation that forms each group. The observations were obtained as a secondary data from a clinical experiment done by Wuensch, K. L in 1992 in his research paper, to identify the effects on the response of the fostered house mice towards species odor. The subjects used are only from the house mice of the species Mus. The nursing mothers selected were only from three species, which are house-mouse (Mus), deer mouse (Peromyscus) or rat (Rattus). The method used in this study is the discriminant analysis techniques. This study established the discriminant functions based on three groups of cross-forested nursing mothers in identifying the effects of response of the subjects towards the species odor. For new predicted membership, it is found that the largest group is group 3 which is the rat (Rattus) group. The resubstitution of the error rate is 30.6% and the cross validation error rate is 38.9%. Thus, because of the new observation was allocated to group of rat, it shows that the linear discriminant function obtained has been justified with the Discriminant Function Coefficient which showed that Rat-V is the predictor that is most heavily weighted on the first discriminant function. Mainly, this study can provide a platform and guidelines for other researchers to understand the classification characteristics of fostered animal species in response to species odor. Other than that, it will open other opportunities for other researchers to study discriminating factors of other species for the same objectives.

Key words: Disriminant analysis; Odor; Animal species; Classification; Response

[1] Faculty Computer Science & Mathematics, Universiti Teknologi MARA, Shah Alam, 40450 Selangor. Malaysia. E-mail: hezli921@salam.uitm.edu.my
[2] Faculty of Computer Science & Mathematics, Universiti Teknologi MARA, Shah Alam, 40450, Selangor. Malaysia
[3] Faculty of Computer Science & Mathematics, Universiti Teknologi MARA, Shah Alam, 40450, Selangor. Malaysia
[4] TropAIR, Faculty of Forestry, Universiti Putra Malaysia, Serdang 43400, Selangor. Malaysia
*Received 15 May 2009; accepted 10 June 2009

1. INTRODUCTION

This study is an extension from the research paper "Fostering house mice onto rats and deer mice: Effects on response to species odors. (*Animal Learning and Behavior*, 1992, *20*, 253-258)" by Wuensch. In his research paper, the researcher used multivariate repeated measures ANOVA as the method. The basis of our study is to identify the discriminating groups that are present in the observation as well as looking into the details of the classification of the observations that forms each group. The observations were obtained from a clinical experiment to identify the effects on the response of the fostered house mice towards species odor. Wild-strain house mice were, at birth, cross-fostered onto house-mouse (*Mus*), deer mouse (*Peromyscus*) or rat (*Rattus*) nursing mothers. Ten days after weaning, each subject was tested in an apparatus that allowed it to enter tunnels scented with clean pine shavings or with shavings bearing the scent of *Mus*, *Peromyscus*, or *Rattus*. The number of visits and the length of time spent for each tunnel by each subjects were recorded for the purpose of the study (Wuensch, 1992). *Mus or Mus Musculus or house-mouse is a* small, scaly-tailed mouse with a distinct notch in the cutting surface of upper incisors with short hair. The ears are moderately large with pale brown color and naked with tail brownish with black tip with no distinct bicolor, but paler on underside, ears, feet drab or buffy, and the tips of toes white (Hovanic et. al., 2008; **The Mammals of Texas**, 1994a; Wikipedia, 2009; Animal Diversity Web, 2009a). *Mus* are generally considered both territorial and colonial when living commensally with humans. However, territoriality is not as pronounced in wild conditions (**The Mammals of Texas**, 1997a; Animal Diversity Web, 2009a). *Peromyscus or Peromyscus Maniculatus or deer mouse is described as a* small, white-footed mouse with sharply bicolor tail, white beneath and dark above with ears usually shorter than hind foot, prominent and leaflike. They also have upperparts bright fulvous or brownish, intermixed with dusky and the underparts and feet white (**The Mammals of Texas**, 1994b; Wikipedia, 2009; Animal Diversity Web, 2009b). Roof rats are in close association with man. This species are black in color with a lighter colored belly in any combination of black, white, grey, and agouti. The skull and nasal bones are relatively narrow (**The Mammals of Texas**, 1994c; Wikipedia, 2009; Animal Diversity Web, 2009c). Discriminant analysis is a technique for classifying and separating individuals into different groups (dependent variables) based on the set quantitative independent random variables. Discriminant analysis involves deriving the linear combination of predictor variables (called Discriminant functions) that will discriminate best between the given groups (Johnson, 1998; Hair, et. al., 1987). The terms *Fisher's linear discriminant* and *Linear Discriminant Analysis* are often used interchangeably, although Fisher's original article *The Use of Multiple Measures in Taxonomic Problems* (1936) actually describes a slightly different discriminant, which does not make some of the assumptions of LDA such as normally distributed classes or equal class covariances (Johnson, 1998; Hair et. al., 1987).

2. MATERIALS AND METHODS

The study analyzed the cases into groups using a discriminate prediction equation between four groups of independent constructs and three group of dependent construct. The independent are labeled as clean-V, Mus-V, Pero-V and Rat-V and a dependent construct are nursing group.

2.1 Assumptions for Discriminant Analysis

The followings assumptions are required in this stud, namely (a) Discriminant function analysis is computationally very similar to MANOVA, and all assumptions for *MANOVA* apply, (b) It is assumed that the data (for the variables) represent a sample from a multivariate normal distribution, (c) It is

assumed that the variance/covariance matrices of variables are homogeneous across groups. Other than that the multivariate Box *M* test for homogeneity of variance/covariance is particularly sensitive to deviations from multivariate normality, and should be taken, (d) It is the major threat to the validity of significance tests occurs when the means for variables across groups are correlated with the variances (or standard deviations). If there is large variability in a group with particularly high means on some variables, then those high means are not reliable. However, the overall significance tests are based on pooled variances, that is, the average variance across all groups. Thus, the significance tests of the relatively larger means (with the large variances) would be based on the relatively smaller pooled variances. 5. Unequal sample sizes are acceptable. The sample size of the smallest group needs to exceed the number of predictor variables. As a "rule of thumb", the smallest sample size should be at least 20 for a few (4 or 5) predictors. The maximum number of independent variables is n - 2, where n is the sample size. While this low sample size may work, it is not encouraged, and generally it is best to have 4 or 5 times as many observations and independent variables (Johnson, 1998; Hair et. al., 1987; Manly, 1994).

2.2 Data Description

The study involves both the qualitative and quantitative variables. The quantitative data variables are X_1 = Clean-V (Clean), X_2 = Mus-V (Mus-scented), X_3 = Pero-V (Peromyscus-scented) and X_4 = Rat-V (Rattus-scented). Meanwhile, the qualitative data variables comprised of the Nurs group, particularly 1 = Mus reared, 2 = Peromyscus reared and 3 = Rattus reared with a sample size *(N)* 36 observations.

3. RESULTS & DISCUSSIONS

Table 1 below shows the normality test result for each variable. Result shows that all variable are normality distributed (p-value<0.05) except for variable Rat-V.

Table 1. Normality Test of the three group predictors for three groups

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	Df	Sig.	Statistic	Df	Sig.
Clean-V	.145	36	.053	.920	36	.013
Mus-V	.125	36	.166	.911	36	.007
Pero-V	.147	36	.047	.940	36	.049
Rat-V	.196	36	.001	.845	36	.000
a. Lilliefors Significance Correction						

From the chi square plot in Figure 1, it shows that most of the data points fall along the diagonal line. Thus, we can conclude that the distribution is multivariate normal. However, there are presents of three outliers. From the plot in Figure 2, it can be seen that the data from the discriminating patterns. Thus, proceed with discriminant analysis. Equality of variance-covariance matrices can be access through Box's M test. Homogeneity of variance-covariance matrices resulting in insignificant of this test means that the variance-covariance is equal. On the other hand, if the test is significant, it indicates that the variance-covariance matrices not equal. Test statistics shows a p-value = 0.495 which indicates that the

variance-covariance matrices are equal. The eigenvalue is the ratio of the between-groups to the within groups sum of squares.

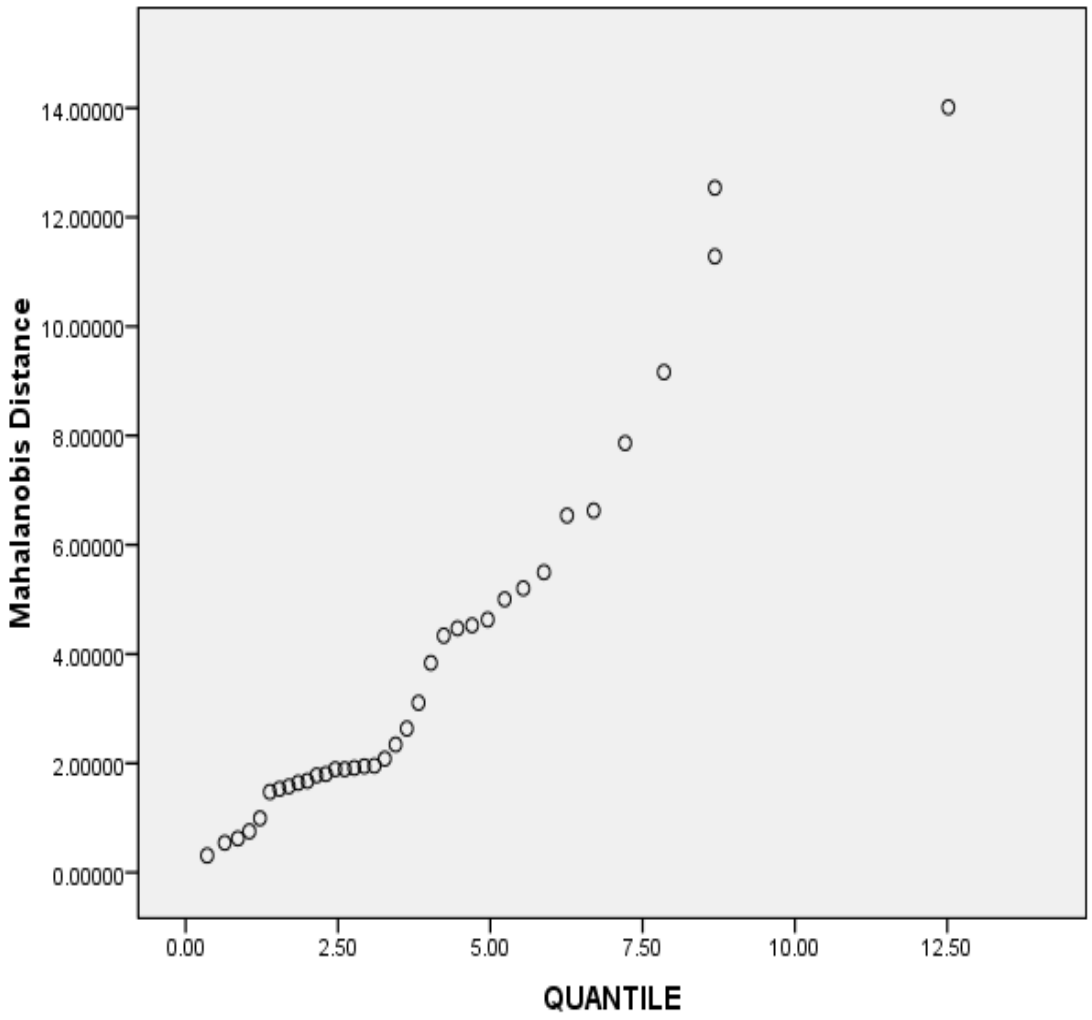

Figure 1. Chi square plot

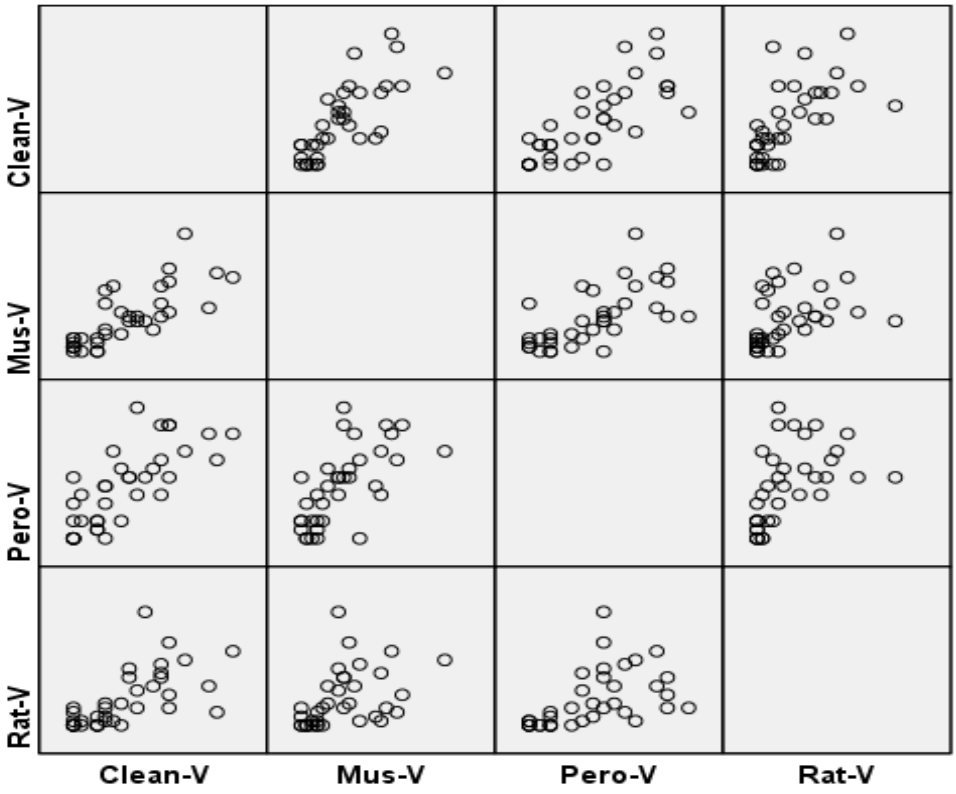

Figure 2. Scatter plot of nurse group

From Table 1, it can be observed that the eigenvalue for the first discriminant function is higher (1.461) compare the second and third discriminant functions. The first discriminant functions explain 93.7 of the total variance and the remaining for the second discriminant function show 2.7 of total variance. A canonical correlation analysis then can be used to investigate the relationship between the two groups. The canonical correlation is the square root of the ratio of the between groups to the total sum of square. The first discriminant function for canonical correlation shows the largest (0.771) compare second discriminant functions. The first discriminant function is significant as the p-value = 0.00 <0.05. It can be concluded that the means for the independent variables between the groups are not the same accept for the second discriminant function (p-value = 0.743 > 0.05). The first discriminant function provided the maximum or the best separation between the groups. The second discriminant function will provide the next best separation between which is unrelated or orthogonal to the first discriminate function and so on. We had more than one groups, so we have more than one discriminant function. Results from Canonical Discriminant Function Coefficient showed that Rat-V is the predictor that is most heavily weighted on the first discriminant function (1.039), followed by Clean-V (0.57). Pero-V (1.293) is the predictor that is heavily weighted on the second discriminant function, followed by Rat-V (0.262). For the first discriminant function, the mean for the Rat is higher comparing others group (1.621). For the second discriminant function, Mus group is higher comparing other groups (0.252). The model of discriminant analysis is $Z = a + W_1$(Mus reared) + W_2(Peromyscus reared) + W_3(Rattus reared).Based on the Classification Function Coefficient, the following Fisher's linear discriminant functions was obtained as follows:

Mus reared: $\hat{d}_1(x_0) = -2.028 - 0.089x_1 + 0.067x_2 + 0.284x_3 + 0.081x_4$

Peromyscus reared: $\hat{d}_2(x_0) = -1.701 + 0.038x_1 + 0.068x_2 + 0.159x_3 - 0.042x4$

Rattus reared: $\hat{d}_3(x_0) = -5.781 + 0.018x_1 + 0.063x_2 + 0.104x_3 + 0.604x_4$

where, C = constant, X_1 = Clean-V, X_2 = Mus-V, X_3 = Pero-V, X_4 = Rat-V

The resubstitution error rate is 30.6% (100-69.4%) and the cross validation error rate is 38.9% (100-61.1%). Both of these error rates are considered a reasonable since it is not too far from 20%. Thus, the discriminant functions to classify new observations can be used. Discriminant functions are considered good classification function if their error rate is low. The separation of the three group means is fully explained in the two dimensional "discriminant space". The group means and the scatter of the individual observations in the discriminant coordinate system are shown in Figure 3. The separation between the groups is clear. From the discriminant functions obtained, we can classify new observations based on certain values of X_1, X_2, X_3 and X_4.

For example; Let x_1 = 16, x_2 = 17, x_3 = 14, x_4= 11. Since $\hat{d}_3(x_0) = 3.678$, is the largest discriminant score, thus we allocate the new observation to group 3. For new predicted membership, it is found that the largest group is group 3 which is the rat (*Rattus*) group. Thus, because of the new observation is allocated to group of rat, it shows that the linear discriminant function obtained is justified with the Discriminant Function Coefficient which showed that Rat-V is the predictor that is most heavily weighted on the first discriminant function (1.039).

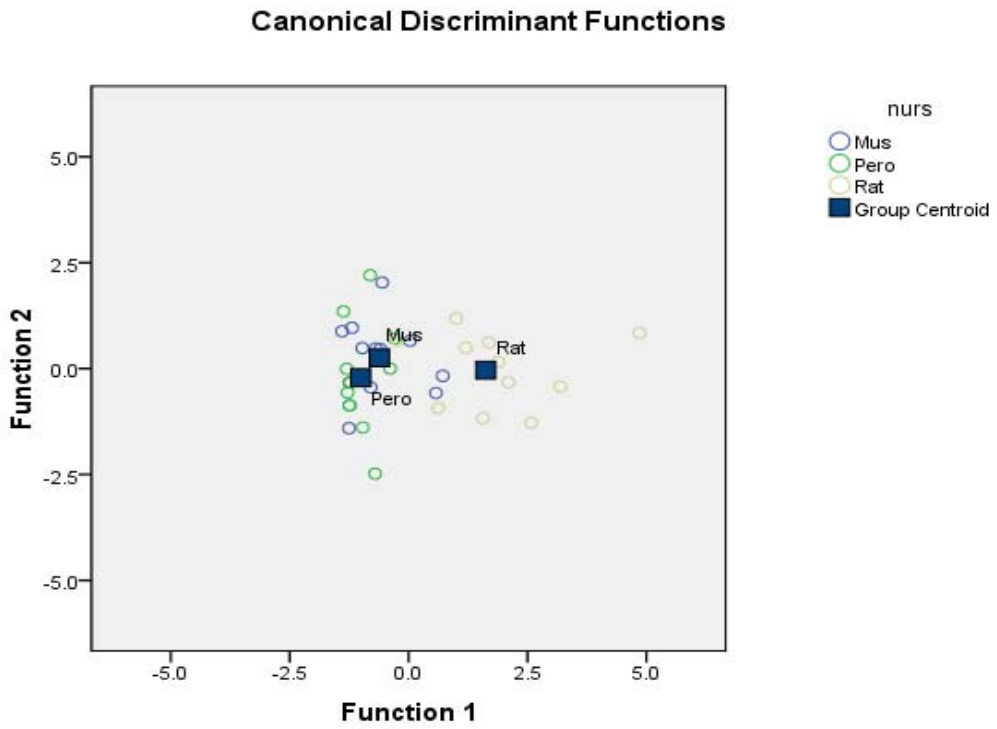

Figure 3. Samples in discriminant Analysis

4. CONCLUSION

This study establishes the discriminant functions based on 3 groups of cross-forested nursing mothers in identifying the effects of response of the subjects towards the species odor. The data used for this study is secondary data. After conducting the analysis for the test of normality, the distribution shows to be a multivariate normal. Box M shows that the variance-covariance matrices are equal. As the conclusion, we can proceed to do the discriminant analysis and derive the linear discriminant functions. The resubstitution of the error rate is 30.6% and the cross validation error rate is 38.9%. The discriminant function obtained from this analysis can be considered reasonable since the error rates are not very far from 20%. For new predicted membership, it is found that the largest group is group 3 which is the rat (*Rattus*) group. Thus the new observation is allocated to group 3 with all of the analyses done with success. We can say that all of our objectives for this study have been achieved. Thus, because of the new observation is allocated to group of rat, it shows that the linear discriminant function obtained is justified with the Discriminant Function Coefficient which showed that Rat-V is the predictor that is most heavily weighted on the first discriminant function. After conducting the analyses, it is recommended that a larger sample size should be utilized in order to further enhance the difference between subject groups. The method using Discriminant Analysis can be further elaborated in the use of different set of data in different field of interest.

REFERENCES

Animal Diversity Web (2009, June). Mus Musculus, University of Michigan Museum of Zoology, Retrieved April 14, 2009 from *http://animaldiversity.ummz.umich.edu/site/accounts/information/Mus_musculus.html*

Animal Diversity Web (2009, June). Peromyscus Maniculatus, University of Michigan Museum of Zoology, Retrieved April 14, 2009 from *http://animaldiversity.ummz.umich.edu/site/accounts/information/Peromyscus_maniculatus.html*

Animal Diversity Web (2009, June). Rattus, University of Michigan Museum of Zoology, Retrieved April 14, 2009 from *http://animaldiversity.ummz.umich.edu/site/accounts/information/Rattus_rattus.html*

Balakrishnam, S., Ganapathiraju, A., Picone, J. (1999). *Linear Discriminant Analysis For Signal Processing Problems*. Institute for Signal and Information Processing, Mississippi State University.

Chan, Y.H. (2005). *Biostatistics 303: Discriminant Analysis* [Electronic Version]. Singapore Med J 46(2) : 54

Hovanic, C., Collman, S.J., Pehling D., with additional information provided by King County Health Department and Snohomish Health District, Rats & Mice [Electronic Version] (Updated links April 21, 2008), Retrieved June, 23 2009 from the Online Library of Washington State University.

Johnson D.E. (1998). *Applied Multivariate Methods for Data Aohnalysis*. 1st Edition, A division of International Thomson Publishing.

Jospeh F. Hair, Jr, Rolph E. Anderson, Ronald L.Tatham. (1987). *Multivariate Data Analysis*. 2nd Edition, Macmillan Publishing Company.

Manly, B.F.J. (1994). *Multivariate Statistical Methods.* 2nd Edition, Chapman and Hall.

The Mammals of Texas Online Edition (1994). House Mouse, Retrieved April, 14 2009 from *http://www.nsrl.ttu.edu/tmot1/mus_musc.htm*

The Mammals of Texas Online Edition (1994). Deer Mouse, Retrieved April, 14 2009 from *http://www.nsrl.ttu.edu/tmot1/peromani.htm*

The Mammals of Texas Online Edition (1994). Roof Rat, Retrieved April, 14 2009 from *http://www.nsrl.ttu.edu/tmot1/rattratt.htm*

Wikipedia(2009, June). Mouse. Retrieved April 14, 2009 from *http://en.wikipedia.org/wiki/Mouse*