Current Trends on Biostatistics & Biometrics

DOI: 10.32474/CTBB.2020.03.000162

ISSN: 2644-1381

Research Article

A Three-Way Discriminant Analysis of Ovulation Interval of Selected Women

Matthew Chukwuma Michael[1]*, Yoro Rume Elize[2] and Obukata Victor[3]

[1]*Department of Mathematics and Statistics, Delta State Polytechnic, Ogwashi, Uku*

[2,3]*Department of Computer Science, School of Applied Sciences, Delta State Polytechnic, Ogwashi, Uku*

***Corresponding author:** Matthew Chukwuma Michael, Department of Mathematics and Statistics, Delta State Polytechnic, Ogwashi-Uku, Delta State, Nigeria

Received: September 15, 2020 **Published:** September 22, 2020

Abstract

This paper aimed at discriminating between women that ovulate in shorter time than the expected twenty-eight (28) days, those that ovulate in the expected twenty-eight (28) days and those that ovulate in longer than the universally expected twenty-eight (28) days. A total of two hundred (200) women in their reproductive age interval were selected for the study. Questionnaires were used to get the relevant information relating to their ovulation interval. The factors affecting ovulation considered in the study include Age, Height, Weight, Work Time (stress), Menstrual Duration, Number of Conceptions, Number of Births and Exposure to Sun. The three-way linear discriminant function was formulated for the data. Using the formulated functions, the women were classified and it was observed that the probability of misclassification into short ovulation and normal ovulation when a woman is actually in long ovulation interval is 0.9863; the probability of misclassifying a woman into short or long ovulation interval when she is actually experiencing normal ovulation is 0.3; the probability of misclassifying a woman into long or normal ovulation when she is actually experiencing short ovulation is 1.0 and the total probability of misclassification of the discriminant function is 0.715.

Keywords: Three-way; discriminant analysis; ovulation; misclassification; multivariate; probability.

Introduction

A good knowledge of ovulation interval of women is imperative in preconception gender selection. Different beliefs and methods for procreation of off springs with desired sex have been adopted but recent development in research has shown that proper determination of ovulation interval has so much to contribute to the efficiency of sex determination. Discriminant Analysis has been applied to the post-mortem discrimination of Felis Catus by their sexes using skull measurements [1]. The level of fluctuation of ovulation interval in women is high and dependent on many variables. While some ovulate normally (every twenty-eight days), some experience short ovulation (not more than twenty-four days) and the other group experience long ovulation (greater than thirty-two days). It is the objective of this paper to classify the women in the study population into their respective ovulation experience.

Methodology

The data for this study were collected from 200 selected women in Ika North-East and Ika South Local Government Areas, Delta State. The women were mainly health workers and teachers in primary and secondary schools in the selected area. Data were obtained on their age, height, weight, work time (hrs/day), menstrual duration, number of conceptions, number of births, exposure to sun and ovulation interval. The women were classified according to the length of their stated interval and the a priori probabilities for the groups were obtained. Discriminant analysis is concerned with the problem of discrimination between two or more groups and assigning a new observation into a group with low probability of misclassification [2-7]. Anderson [8] developed a method for discrimination and classification which shows that for known a priori probabilities and misclassification costs, the optimum rule is based upon the likelihood ratio of all pairs of multivariate normal populations f_i (X) and f_j (X). Then the ratio of the i[th] to the j[th] density is

$$\frac{f_i(X)}{f_j(X)} = \frac{\exp\left\{-\frac{1}{2}(X-\mu_i)'\Sigma^{-1}(X-\mu_i)\right\}}{\exp\left\{-\frac{1}{2}(X-\mu_j)'\Sigma^{-1}(X-\mu_j)\right\}}; \forall i \neq j \quad (1)$$

$$\frac{f_i(X)}{f_j(X)} = \exp\left\{-\frac{1}{2}\left[(X-\mu_i)'\Sigma^{-1}(X-\mu_i) - (X-\mu_j)'\Sigma^{-1}(X-\mu_j)\right]\right\} \quad (2)$$

The region of classification into population, π_i is the set of X^'s for which Equation 2 is greater than k (k suitably chosen). That is,

$$\exp\left\{-\frac{1}{2}\left[(X-\mu_i)'\Sigma^{-1}(X-\mu_i) - (X-\mu_j)'\Sigma^{-1}(X-\mu_j)\right]\right\} > k$$

Thus,

$$-\frac{1}{2}\left[(X-\mu_i)'\Sigma^{-1}(X-\mu_i) - (X-\mu_j)'\Sigma^{-1}(X-\mu_j)\right] > \log_e k$$

$$Y_{ij} = X'\Sigma^{-1}(\mu_i - \mu_j) - \frac{1}{2}(\mu_i + \mu_j)'\Sigma^{-1}(\mu_i - \mu_j) \quad (3)$$

The population parameters, μ_i, μ_j and Σ may be estimated with their respective sample estimators X ̅_i, X ̅_j and S where

$$S = \frac{\sum_{i=1}^{k}(n_i - 1)S_i}{\sum_{i=1}^{k} n_i - k} \quad (4)$$

is the pooled variance-covariance matrix for k groups.

For the common case of unknown parameters, the discriminant function is

$$\hat{Y}_{ij} = X'S^{-1}(\bar{X}_i - \bar{X}_j) - \frac{1}{2}(\bar{X}_i + \bar{X}_j)'S^{-1}(\bar{X}_i - \bar{X}_j) \forall i \neq j \quad (5)$$

The discriminant function, Y ij, for every i≠j will discriminate between two completely specified groups (populations), π_i and π_j. The classification rule now is: assign X to π_i if

$$\hat{Y}_{ij} > \log_e k \forall i \neq j \quad (6)$$

But if a priori probabilities $f_i(X)$ and $f_j(X)$ are known then k is given by

$$k = \frac{f_j(X)c(i/j)}{f_i(X)c(j/i)} \quad (7)$$

as defined in Equation 1. If the two populations are equally likely, that is, $f_i(X) = f_j(X)$; and also the misclassification costs being equal, that is $C(i/j) = C(j/i)$, this leads to k being equal to 1. Then, $\log_e k = \log_e 1 = 0$. The best classification with known and equal a priori probabilities and misclassification cost is: assign observation with measurement X to population π_i if $\hat{Y}_{ij} > 0 \forall i \neq j$; assign to π_j if otherwise. With the relations $Y_{ij} = -Y_{ij}$ and $Y_{1i} - Y_{1i} = Y_{ij}, k$, k populations or groups will require k-1 linearly independent discriminant function(s) to obtain the best regions of classification.

The discriminant procedure is evaluated with the aid of confusion matrix. The Apparent Error Rate (AER), a measure of the tendency that individual items are wrongly classified, is appropriately determined as the proportion of misclassified items. Let

$$P_i = P(\pi_{i1}', \pi_{i2}', \ldots, \pi_{ip}' / \pi_i) = \frac{\sum_{j=1}^{p} w_{i'j}}{n_i}; \forall i \neq i' \quad (8)$$

where P_i is the probability of misclassifying an item that truly belongs to the i^th group into any of i^' groups.

For the case of three groups

$$P_1 = P(\pi_2, \pi_3 / \pi_1) = \frac{W_{21} + W_{31}}{n_1} \quad (9)$$

$$P_2 = P(\pi_1, \pi_3 / \pi_2) = \frac{W_{12} + W_{32}}{n_2} \quad (10)$$

$$P_3 = P(\pi_1, \pi_2 / \pi_3) = \frac{W_{13} + W_{23}}{n_3} \quad (11)$$

The total probability of misclassification denoted by P ̅ is given by

$$\bar{P} = \frac{W_{12} + W_{13} + W_{23} + W_{21} + W_{31} + W_{32}}{n_{\cdots}}$$

where w_{ij} is the number of items misclassified into i while they belong to j; $i \neq j; C_{ij}$ $i = j$ are correctly classified items.

Data Analysis

x_i;i=1,2,...,9, represent Age in years, Height in meters, Weight in Kg, Work Time in hours/day, Menstrual Duration in days, Number of Conceptions, Number of births, Exposure to Sun in hours/day and x_9 is the ovulation interval in days. The average Ovulation Interval, x ̅_9, is 16.06 for the 47 women experiencing short interval, 28.89 for the 80 women experiencing normal interval and 48. 95 for the 73 women experiencing long interval. The group variance-covariance matrices and the pooled variance covariance matrix were obtained. The two linearly independent discriminant functions, Y_12 and Y_13 was obtained

$$Y_{12} = -0.0080094x_1 - 0.119718x_2 - 0.009589x_3 + 0.033052x_4 + 0.003926x_5 - 0.067772x_6 + 0.032326x_7 + 0.006512x_8 + 0.383174$$

$$Y_{13} = -0.014784x_1 - 0.109057x_2 - 0.008104x_3 + 0.019137x_4 + 0.046557x_5 - 0.075191x_6 + 0.204162x_7 + 0.066034x_8 + 0.406651$$

The Classification Rules: assign an individual with measurement X of unknown origin to π_1 if $Y_{12}>0.531879$ and $Y_{13}>0.440312$ from Equation 6; assign the individual to π_2 if $Y_{21}>-0.531879$ and $Y_{32}>0.0915672$. $P_3=0.9863$, $P_2=0.3$ and $P_1=1$. The misclassification probabilities show that 986 out of every 100 of the women experiencing long ovulation were misclassified as experiencing short or normal ovulation; 300 out of 1000 women experiencing normal ovulation were misclassified as experiencing short or long ovulation; all the women experiencing short ovulation were misclassified as experiencing long or normal ovulation and the total probability of misclassification is 0.715 showing that 715 of the women were misclassified by the classification/discriminant function.

Conclusion

From the results of data analysis, the following conclusions may be reached: the discriminant function is associated with high probability of misclassification; there may be important variables excluded in the study and the selected women may be experiencing fluctuating ovulation intervals such that tracking a woman's interval would require studying her experiences independently over a long period of time.

References

1. Du Toit SH, Aarde RJ (1980) Sex Determination of the Feral Cat (Felis Catus) using Multivariate Analysis. South African Journal of Wildlife Research 10(2): 82-87.
2. Onyeagu SI (2003) A First Course Multivariate Analysis. Mega Concept.
3. Ogum GE (2002) Introduction to Methods of Multivariate Analysis. Aba Nigeria Afri Tower Limited.
4. Johnson RA (1992) Applied Multivariate Statistical Analysis. New York Englewood Cliffs, USA.
5. Cooley WW (1962) Multivariate Procedure for the Behavioural Sciences. Wiley and Sons Inc New York, USA.
6. Lachnbruch PA (1975) Discriminant Analysis. Hafner Press New York, USA.
7. Welch BL (1939) Note on Discriminant Functions. Biometrika 31(2): 218-220.
8. Anderson TW (1958) An Introduction to Multivariate Analysis. John Wiley and Sons Inc New York, USA.

DOI: 10.32474/CTBB.2020.03.000162