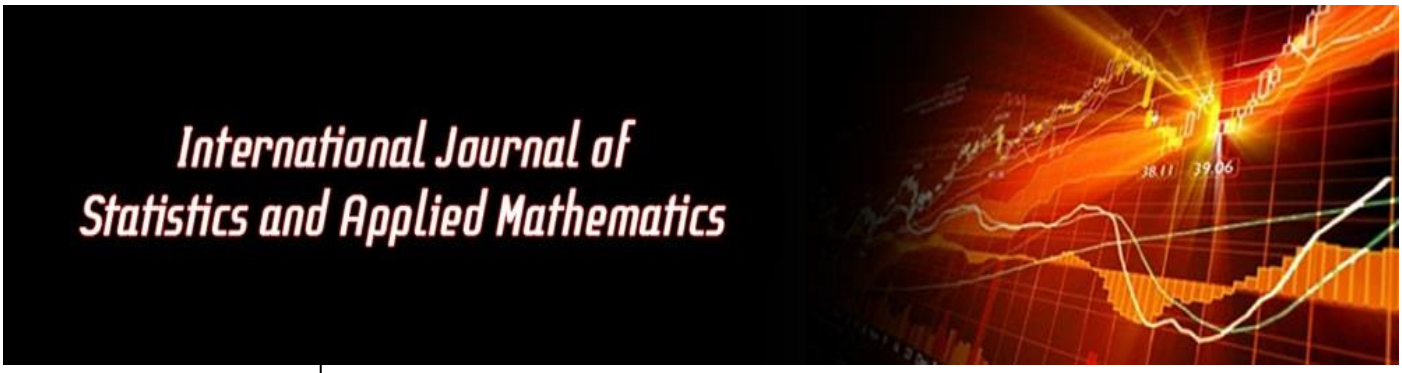
International Journal of Statistics and Applied Mathematics 2024; SP-9(2): 155-162



ISSN: 2456-1452
Maths 2024; SP-9(2): 155-162

www.mathsjournal.com
Received: 02-01-2024
Accepted: 03-02-2024

Rajarathinam A
Department of Statistics, Manonmaniam Sundaranar University, Tirunelveli, Tamil Nadu, India

Corresponding Author:
Rajarathinam A
Department of Statistics, Manonmaniam Sundaranar University, Tirunelveli, Tamil Nadu, India

Fisher's discriminant modeling for human platelet parameters

Rajarathinam A

DOI: https://doi.org/10.22271/maths.2024.v9.i2Sc.1697

Abstract
This study employed multivariate modeling techniques to explore the relationship between platelet parameters and demographic factors. Assumptions such as normality, multicollinearity, and covariance matrix equality were rigorously assessed. Data normalization using the Box-Cox method improved normality, facilitating more robust analyses. Multivariate analysis of Variance revealed significant variations in platelet parameters across demographic groups. Furthermore, linear discriminant analysis demonstrated the capacity to classify individuals based on platelet parameters, particularly concerning gender. The findings underscore the importance of platelet parameters in understanding population characteristics and their potential implications for medical research and clinical practice.

Keywords: Platelet parameters, Box-cox, MANOVA, Wilks' Lambda, and Fisher Discriminant analysis.

1. Introduction
Platelets, crucial blood components, play a fundamental role in various physiological processes, including clotting and immune response. Understanding platelet parameters and their associations with demographic factors is essential for medical research and clinical practice. Platelet parameters, such as distribution width, volume, large cell ratio, and count, reflect platelet morphology and function, providing valuable insights into an individual's health status. Discriminant analysis, a multivariate technique, has been widely employed to explore relationships between platelet parameters and demographic variables. This method allows for classifying individuals based on observed characteristics, facilitating the identification of factors influencing platelet biology.
Discriminant analysis (DA) is one of the methods used in multivariate analysis, along with the dependency method. It is used in cases where the independent variable is matrix data and the dependent variable is non-metric data. DA undertakes the same task as multiple linear regression by predicting an outcome, and it is used to build a predictive model of group membership based on the observed characteristics of each case.

The main objectives of the study are
- To identify linear combinations of platelet parameters that effectively discriminate between demographic groups.
- To identify most significant platelet parameters contribute to the discrimination between demographic categories.
- To assess the statistical significance of the discriminant functions and ascertain whether significant differences exist among demographic groups based on platelet parameters.

2. Review of literature
Durrant and Kaban (2010) [7] utilized a novel approach incorporating random projections with Fisher's Linear Discriminant (FLD) classifier for classification tasks. Unlike previous methods focusing solely on preserving pairwise distances under projection, the authors' approach emphasizes leveraging the inherent class structure within the data.

Ramayah *et al*. (2010) [19] present a step-by-step example, making it easier for readers to comprehend the intricacies of discriminant analysis. They carefully explain the necessary assumptions and procedures involved in discriminant analysis, including data preparation, model estimation, and interpretation of results.

Nainggolan *et al*. (2018) [15] used discriminant analysis and classified Hypertension women aged 27 to 54 years living in the village in the central district of Bogor. The results of the multivariate discriminant analysis showed that the level of Vo2 max is the only distinction maker in the incidence of hypertension.

ALKubaisi *et al*. (2019) [1] used discriminant analysis with three criteria to test the developed model, producing excellent projecting precision. The discriminant function properly assessed and classified about 67% of the cases in the analysis. Also, the analysis produced two discriminant functions: the first explained 77% and the second explained 23% of the Variance.

Dibal and Abraham (2020) [5] applied Fisher's linear Discriminant Analysis (FLDF) to health data on diabetic patients from the University of Port Harcourt Teaching Hospital, Rivers, Nigeria. He created a predictive discriminant model that classifies patients into one of two groups (Diabetic and Non-Diabetic). Fisher's linear discriminant function correctly classifies 65.4% of the total observation.

Ndako *et al*. (2020) [16] investigated if hematological measurements could differentiate between typhoid-positive and -negative pediatric patients. Using Fisher's Linear Discriminant Method, 200 patients were analyzed. A discriminant score threshold of -0.0067 was established, with patients above classified as harmful and below as positive. Classification efficacy was assessed using retribution estimate and leaving-one-out approaches, indicating a 75.8% and 74.7% prevalence for typhoid-positive patients, respectively. These findings suggest a high prevalence of typhoid fever among pediatric patients, emphasizing the need for improved point-of-care diagnostics with robust positive predictive value.

Explored Discriminant Function Analysis (DFA) to evaluate the effectiveness of Indigenous health-and-wellness programs, particularly in the Eeyou Istchee territory, Canada. By analyzing various health parameters, DFA models were developed to discriminate between individuals with and without Type 2 Diabetes Mellitus (T2DM). The models exhibited high specificity (~97%) in classifying non-T2DM individuals. This research underscores the potential of DFA in point-of-contact evaluations for monitoring and assessing health interventions in rural and remote Indigenous communities, providing valuable insights for T2DM management and prevention strategies among the James Bay Cree population.

Ding *et al*. (2023) [6] introduced the Sparse Variables Selection Exponential Local Fisher Discriminant Analysis (SELFDA) model to address shortcomings in fault classification using Local Fisher Discriminant Analysis (LFDA). By automatically identifying key faulty variables through the minor absolute shrinkage and selection operator, SELFDA enhances fault diagnosis performance and model interpretability. It overcomes the Small Sample Size (SSS) problem by employing a matrix exponential strategy, ensuring full-rank within-class scatter matrices. This approach, tested on the Tennessee Eastman process and a real-world diesel working process, outperforms existing methods, demonstrating its effectiveness in practical industrial applications.

Rahamneh *et al*. (2023) [18] utilized discriminant analysis to distinguish between two types of Bowel and Esophageal cancer in Jordan, identifying significant variables such as sex, weight, and Platelets Count P.C. The correct classification rates for the first and second groups were 62.8% and 77% respectively, with misclassification rates of 37.2% and 23%. The proper classification ratio was 71.6%, with a false classification ratio of 28.4%. The method effectively identified vital independent variables for diagnosing both cancer types, with correct classification probabilities of 66.4% and 77.6% for the first and second groups, respectively.

Henry *et al*. (2023) [10] investigated the spectral differences of tobacco leaves under macronutrient deficiencies. They employed information entropy and spectral derivatives methods to identify the most effective wavelengths for discrimination. Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) algorithms were utilized to reduce data dimensionality and classify the symptoms. The study's findings revealed that the overall accuracy for classifying young, intermediate, and mature plants was 92%, 82%, and 75%, respectively. The results also indicated that nitrogen, sulfur, and magnesium deficiencies significantly impacted the classification accuracy. In contrast, deficiencies in phosphorus and potassium had minimal effect on the classification outcomes.

3. Materials and Methods

3.1 Materials

The dataset under investigation is collected through Mendeley Data (https://data.mendeley.com/datasets/5t8dr6d73f/1). The dataset comprises comprehensive information on platelet parameters and red cell distribution width (RDW) for 1883 samples. It includes measurements obtained from individuals across various demographics, encompassing platelet distribution width, mean platelet volume, platelet large cell ratio, plateletcrit, total platelet count, and RDW values. Each entry in the dataset provides specific values for these parameters, allowing for a detailed analysis of their distributions, variability, and potential interrelationships.

3.2 Methods

3.1.1 Box-Cox method

The Box-Cox method is used in statistics and econometrics to transform non-normal data into approximately normal data. It is named after statisticians George Box and Sir David Cox and was introduced in 1964. Let $y = (y_1, y_2, ..., y_n)'$ be the date on which the Box-Cox transformation is applied. Box and Cox (1964) defined their transformation as

$$y_i^{(\lambda)} = \begin{cases} \lambda^{-1}(y_i^{\lambda} - 1) & if \ \lambda \neq 0 \\ \log(y_i) & if \ \lambda = 0 \end{cases} \quad (1)$$

Such that for unknown λ

$$y^{(\lambda)} = X\beta + \varepsilon \quad (2)$$

Where $y^{(\lambda)}$ is the λ transformed data, X is the design matrix (possible covariates of interest), β is the set of parameters associated with the λ transformed data, and $\varepsilon = (\varepsilon_1, \varepsilon_2, ..., \varepsilon_n)$ is the error term. Since the aim of Equation (1) is that

$$y^{(\lambda)} \square N(X\beta, \sigma^2 I_n) \quad (3)$$

$\varepsilon \sim N(0, \sigma^2)$ The transformation in Equation (1) is only valid for $y_i > 0, i = 1, 2, ..., n$ modifications to be made when negative observations are present (Vélez *et al*., (2015)).

3.2.1 Multivariate Analysis of Variance

A MANOVA technique (Johnson & Wichern, (1998)) [12] is employed to test the significance of variation among all the five parameters considered simultaneously. The MANOVA model for comparing the population means vectors is as follows:

$$Y_{ij} = \mu + V_i + E_{ij} \quad (4)$$

Where, E_{ij} is a vector of random error distributed as $N_P(0, \Sigma)$. Here, the parameter vector μ is the overall mean and V_i represents the model's status in (4); each component of the observation vector Y_{ij} satisfies the univariate model, and the variance-covariance matrix Σ is the same for all populations.

3.2.2 Variance Inflation Factor

The variance inflation factor is used to measure how much the Variance of the estimated regression coefficient is inflated if the independent variables are correlated. VIF is calculated as

$$VIF = \frac{1}{1 - R^2} \quad (5)$$

Where the tolerance is simply the inverse of the VIF; the lower the tolerance, the more likely the multicollinearity among the variables. The value of VIF=1 indicates that the independent variables are not correlated. If the value of VIF is 1 < VIF < 5, it specifies that the variables are moderately correlated. If the VIF value is above 5, there will be multicollinearity among the predictors in the regression model (Goldstein, (1993) [9] and Shrestha, (2020) [21]. Another one is the scatterplot graphical method, which signifies the linear relationship between pairs of independent variables. It is essential to look for scatterplots that indicate a linear relationship between pairs of independent variables. The correlation coefficient is calculated using the formula:

$$r = \frac{n(\sum XY) - (\sum x)(\sum y)}{\sqrt{\left[n\sum X^2 - (\sum)^2\right]\left[n\sum Y^2 - (\sum Y)^2\right]}} \quad (6)$$

Where, r is the correlation coefficient, n is the number of observations, X represents the first variable in the context, and Y is the second variable in the context. If the correlation coefficient value is higher with the pairwise variables, it indicates the possibility of collinearity.

3.2.3 Box's-M test

The Test for homogeneity of covariance matrices, introduced in 1949, examines the covariance matrices derived from multivariate normal data considering one or more classification factors. This test assesses the similarity between the separate covariance matrices by comparing the product of their log determinants to the log determinant of the combined covariance matrix, similar to a likelihood ratio test. The test statistic employs a chi-square approximation.

3.2.4 Wilk's lambda

In discriminant analysis, Wilk's lambda is utilized to assess the contribution of each level of an independent variable to the model. This scale ranges from 0 to 1, where a value of 0 indicates complete discrimination, while a value of 1 signifies no discrimination. To test the impact of each independent variable, it is successively included and excluded from the model, generating a Λ statistic. The significance of the change in Λ is evaluated using an F-test; if the computed F-value exceeds the critical value, the variable is retained in the model (Onwukwe, (2014) [17]. Thus, a non-significant Wilks' lambda value is always preferred.

$$Wilks\ lamda(\Lambda) = \frac{|W|}{|B+W|} \quad (7)$$

B is the between-groups matrix, and W is the within-group matrix. The Eigenvalue can be explained as the ratio of the between-groups sum of squares to the within-group sum of squares (McGarigal *et al*., (2000) [14].

3.2.5 Multiple Discriminant Analysis

Multiple Discriminant Analysis (MDA) is an extension of discriminant analysis; it shares ideas and techniques with various analyses of Variance (MANOVA). MDA aims to classify cases into three or more categories using continuous or dummy categorical variables as predictors (Cramer, (2003) [4], Jang *et al*., (2015) [11]. The term DA refers to numerous types of analyses. DA is the most popular statistical technique to classify individuals or observations into non-overlapping groups based on scores derived from a suitable "statistical decision function" constructed from one or more continuous predictor variables. While investigating the differences between the groups or categories, the necessary step is to identify the attributes with the most contributions to maximum reparability between known groups or categories to classify a given observation into one of the groups. For that purpose, DA successively identifies the linear combination of attributes known as canonical discriminant functions (equations) that contribute maximally to group separation. Predictive DA addresses the question of how to assign new cases to groups.

The form of the Equation or function is:

$$D_i = \alpha_0 + \alpha_1 X_{i1} + \alpha_2 X_{i2} + \alpha_3 X_{i3} + ... + \alpha_j X_{ik}$$

Where D is an independent variable and D_i is the value of discriminant score from the i[th] category (i=1,2,…n), α_j is the discriminant coefficient of j[th] attributes (j=0,1,2,…,k), and X_{ik} is the k[th] independent variable of the i[th] category. This function is similar to a regression equation or function. The $\alpha's$ re are unstandardized discriminant coefficients analogous to the ones in the regression equation. These $\alpha's$ maximize the distance between the means of the dependent variable, and the standardized discriminant coefficients can also be used, like beta

weight in regression.

4. Results and Discussion

The results presented in Table 1 reveal a population's (N=1883) age and platelet-related parameters. With an average age of 37.73 years and a standard deviation of 13.68, there's notable age diversity. Platelet distribution width (mean: 11.65, SD: 1.51) shows moderate variability, while mean platelet volume (mean: 10.28, SD: 0.72) exhibits less. Platelet large cell ratio (mean: 27.00, SD: 5.95) suggests significant diversity, unlike Plateletcrit (mean: 0.27, SD: 0.04), indicating consistency. Total platelet count (mean: 264.39, SD: 46.10) varies notably. Skewness and kurtosis imply generally symmetrical distributions, barring slight negative skewness for platelet large cell ratio and high kurtosis for plateletcrit. These insights are crucial for medical research and clinical evaluations.

Table 1: Characteristics of summary statistics for platelet parameters

	Mean	Std. Deviation	Variance	Skewness	Kurtosis
Age	37.73	13.68	187.01	0.67	-0.22
Platelet distribution width	11.65	1.51	2.27	0.64	-0.10
Mean platelet volume	10.28	0.72	0.52	0.46	-0.44
Platelet large cell ratio	27.00	5.95	35.39	0.44	-0.54
Plateletcrit	0.27	0.04	0.00	-0.04	-0.70
Total platelet count	264.39	46.10	2124.96	0.06	-0.62

The Kolmogorov-Smirnov and Shapiro-Wilk tests were conducted (Table 2) to assess the normality of the distributions for platelet parameters before and after applying the Box-Cox transformation. Before the transformation, all variables exhibited statistically significant deviations from normality ($p<0.05$), with varying degrees of skewness. However, after applying the Box-Cox transformation, there was an improvement in the normality of the distributions for most variables, as indicated by non-significant p-values ($p>0.05$) in both tests. Specifically, platelet distribution width, mean platelet volume, large cell ratio, and plateletcrit showed notable improvements in normality post-transformation. Total platelet count also exhibited improved normality, although to a lesser extent. These results suggest that the Box-Cox transformation effectively normalized the distributions of platelet parameters, making them more suitable for subsequent statistical analyses that assume normality, such as parametric tests.

Table 2: Characteristics of normality test for platelet parameters

Before Box-Cox						
Variables	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	DF	Sig.	Statistic	DF	Sig.
Platelet distribution width	0.08	1883	0.00	0.96	1883	0.08
Mean platelet volume	0.08	1883	0.00	0.97	1883	0.08
Platelet large cell ratio	0.06	1883	0.00	0.97	1883	0.06
Plateletcrit	0.06	1883	0.00	0.98	1883	0.06
Total platelet count	0.03	1883	0.00	0.99	1883	0.03

In Figure 1, the histogram black curve shows the Gaussian distribution, while the histogram shows the distribution of 1882 platelet cells of different parameters. The top bars in the histogram match nicely with the Gaussian distribution; therefore, after the Box-Cox method, the dataset was perfectly normally distributed. The points in the histogram plot form a bell-shaped line since the dataset's quantiles nearly match the dataset's quantiles, which would theoretically be the customarily distributed dataset.

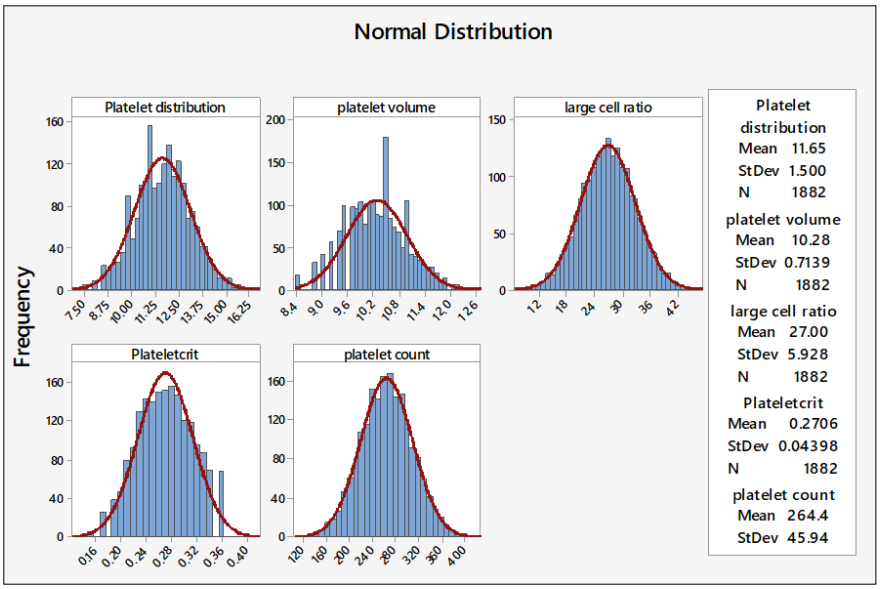


Fig 1: Normality plot for platelet parameters

Table 3: Characteristics of multivariate Analysis of Variance for platelet parameters

Statistics	Value	Error DF	Sig.
Pillai's Trace	0.09	1872	0.00
Wilks' Lambda	0.91	1872	0.00
Hotelling's Trace	0.10	1872	0.00
Roy's Largest Root	0.10	1872	0.00

Various multivariate tests, such as Pillai's trace, Wilks' lambda, Hotelling's trace, and Roy's most significant root tests, were utilized to assess the collective variation of all five platelet parameters across the gender groups. The outcomes of these tests are presented in the Table 3. These MANOVA statistics provide insights into the multivariate effects of the analysis. Pillai's Trace, Wilks' Lambda, Hotelling's Trace, and Roy's Largest Root are all measures of the significance of the overall model. In this case, the extremely low p-value (0.00) indicates that the model has a significant overall effect. The values of these statistics (ranging from 0.09 to 0.10) suggest the proportion of Variance in the dependent variables explained by the independent variables in the model.

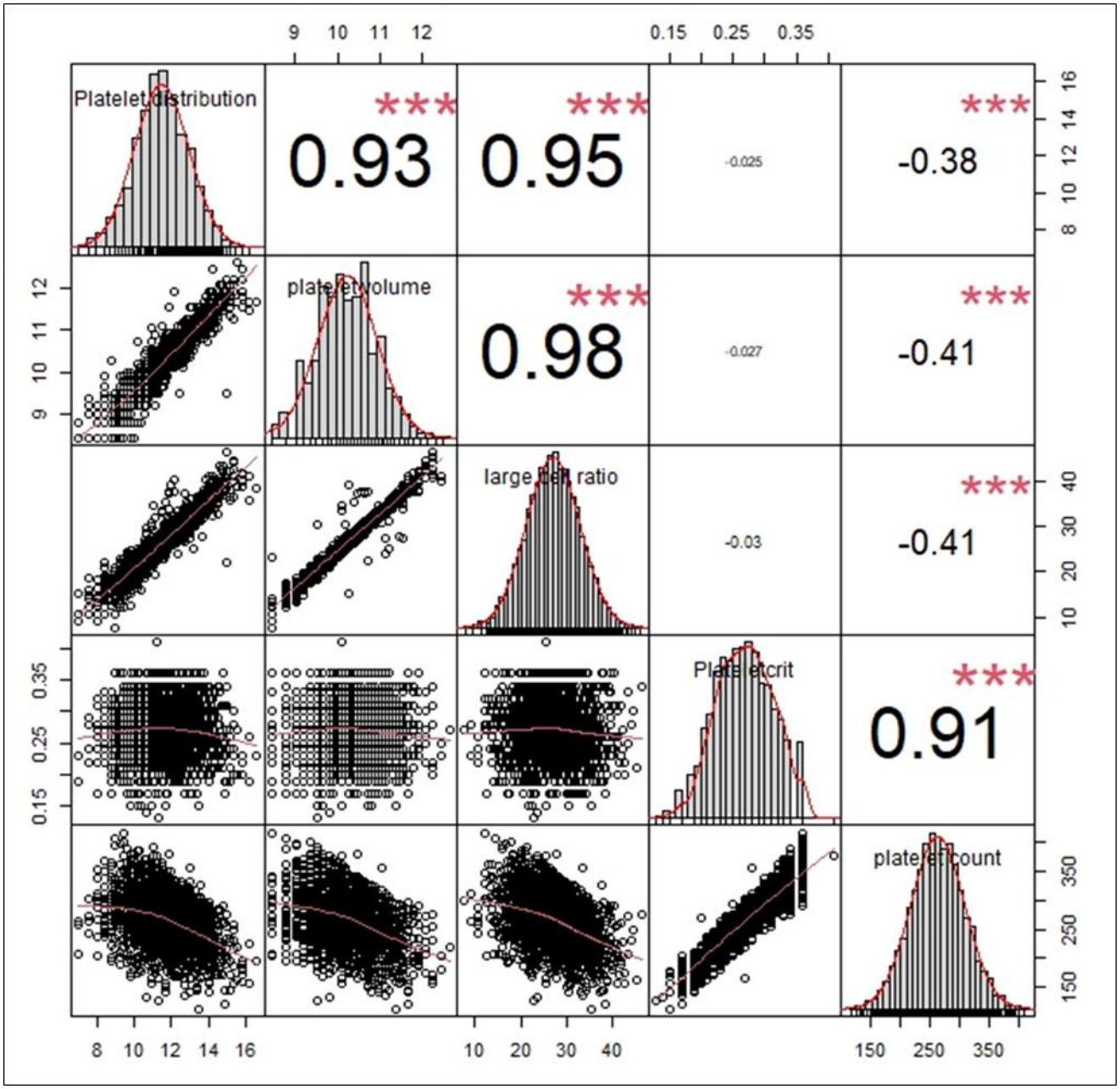


Fig 2: Correlation matrix for platelet parameters association

The Pearson correlation between the study variables' platelet parameters has been calculated. It is depicted in Figure 3.2. The upper triangular matrix shows the Pearson correlation and its significance level (as stars). Each significance level is associated with a symbol: p-values 0.001 (***), 0.01 (**), and 0.05 (*).

The result reveals that the Platelet large cell ratio is highly positively correlated (0.98) with platelet volume and platelet distribution significance ($p<0.01$). Plateletcrit has a positive correlation with platelet total count. The platelet total count negatively correlated with platelet volume and platelet distribution. Figure 2 provides evidence of no strong correlation among the independent variables; hence, multicollinearity doesn't occur in this problem.

Table 4: Characteristics of multicollinearity for platelet parameters

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
	B	Std.Err	Beta			Tolerance	VIF
(Constant)	-1.874	0.629		-2.979	0.003		
Platelet distribution width	-0.136	0.022	-0.422	-6.223	0.000	0.106	4.462
Mean platelet volume	0.277	0.080	0.408	3.461	0.001	0.035	2.593
Platelet large cell ratio	0.010	0.010	0.129	1.025	0.306	0.031	3.429
Plateletcrit	2.707	1.558	0.236	1.738	0.082	0.026	3.907
Total platelet count	0.000	0.002	-0.012	-0.082	0.935	0.022	4.445

These coefficients in Table 4 represent the relationship between the model's independent variables (platelet parameters) and the dependent variable. The standardized coefficients (Beta) indicate the strength and direction of the relationship, while the t-values and significance levels (Sig.) indicate the statistical significance of each coefficient. Collinearity statistics such as Tolerance and VIF assess multicollinearity among the independent variables. In this model, platelet distribution width and mean platelet volume show significant negative and positive relationships with the dependent variable. However, platelet large cell ratio, plateletcrit, and total platelet count do not show statistically significant relationships. Furthermore, all variables exhibit acceptable levels of multicollinearity because the Variance inflation factor values are below 5.

Table 4: Characteristics of Box's M method

Box's M	88.441
Approx.	5.878
DF1	15
DF2	8662237.732
Sig.	0.210

Box's M test statistics reported in Table 4 is a diagnostic test used to assess the equality of covariance matrices across groups in multivariate analysis of Variance (MANOVA). The test evaluates whether the assumption of homogeneity of covariance matrices (homoscedasticity) is violated. In this case, the p-value (Sig.) of 0.210 suggests no significant violation of this assumption, indicating that the covariance matrices are approximately equal across groups. Therefore, the assumption of homogeneity of covariance matrices is met, and the MANOVA results can be interpreted reliably.

Table 5: Characteristics of Wilk's Lambda test statistics

Test of Function(s)	Wilks' Lambda	Chi-square	DF	Sig.
Function 1	0.908	182.114	5	0.000

In Table 5 results, "Function 1" refers to the specific function being tested. The Wilks' Lambda value of 0.908 indicates the proportion of Variance in the dependent variables not accounted for by the independent variables. The associated Chi-square statistic of 182.114 and the degrees of freedom (DF) of 5 results in a highly significant p-value (Sig.) of 0.000, suggesting that the overall model or the specific function being tested significantly affects the dependent variables.

Table 6: Characteristics of eigenvalues for the first function

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
Function 1	0.102	100	100	0.304

In Table 6, the eigenvalue of 0.102 indicates the amount of variance explained by first discriminant function. Since the percentage of variance is 100.0%, function one accounts for the entire variance in the data. The cumulative % also reflects this, as it reaches 100.0%. The canonical correlation of 0.304 represents the correlation between the observed and canonical variables derived from the function.

Table 7: Characteristics of canonical discriminant function coefficients

	Platelet distribution width	Mean platelet volume	Platelet large cell ratio	Plateletcrit	Total platelet count
Function 1	-1.434	1.531	0.411	0.526	0.259

These coefficients presented in Table7 indicated the weights assigned to each variable in the canonical discriminant function. The coefficients signify the magnitude and direction of the relationship between each predictor variable (platelet parameters) and the discriminant function. Positive coefficients suggest a positive association with the function (mean platelet volume, platelet large cell ratio, Plateletcrit, and total platelet count). In contrast, negative coefficients (Platelet distribution width) imply a negative association. The values reflect the relative importance of each variable in discriminating between groups or explaining the variability in the data.

Table 8: Characteristics of fisher linear discriminant function

Variables	Gender	
	F	M
Platelet distribution width	10.834	11.467
Mean platelet volume	401.953	400.522
Platelet large cell ratio	-38.343	-38.389
Plateletcrit	-352.782	-361.242
Total platelet count	3.599	3.595
(Constant)	-1639.676	-1627.220

Fisher linear discriminant results presented in the Table 8 represents the relationship between each platelet parameter and gender in a linear discriminant analysis. Positive coefficients indicate an increase in the value of the platelet parameter associated with the specified gender, while negative coefficients indicate a decrease. The constant term represents the intercept of the linear discriminant function for each gender group. The equations representing the relationship between each platelet parameter and gender in the linear discriminant analysis are as follows:

Female (F):

$$Y_F = -1639.676 + 10.834_{\text{Platelet distribution width}} + 401.953_{\text{Mean platelet volume}} - 38.343_{\text{Platelet large cell ratio}} - 352.782_{\text{Plateletcrit}} + 3.599_{\text{Total platelet count}}$$

Male (M):

$$Y_M = -1627.220 + 11.467_{\text{Platelet distribution width}} + 400.522_{\text{Mean platelet volume}} - 38.389_{\text{Platelet large cell ratio}} - 361.242_{\text{Plateletcrit}} + 3.595_{\text{Total platelet count}}$$

In these equations, Y_F and Y_M represent the discriminant scores for females and males based on the given platelet parameters.

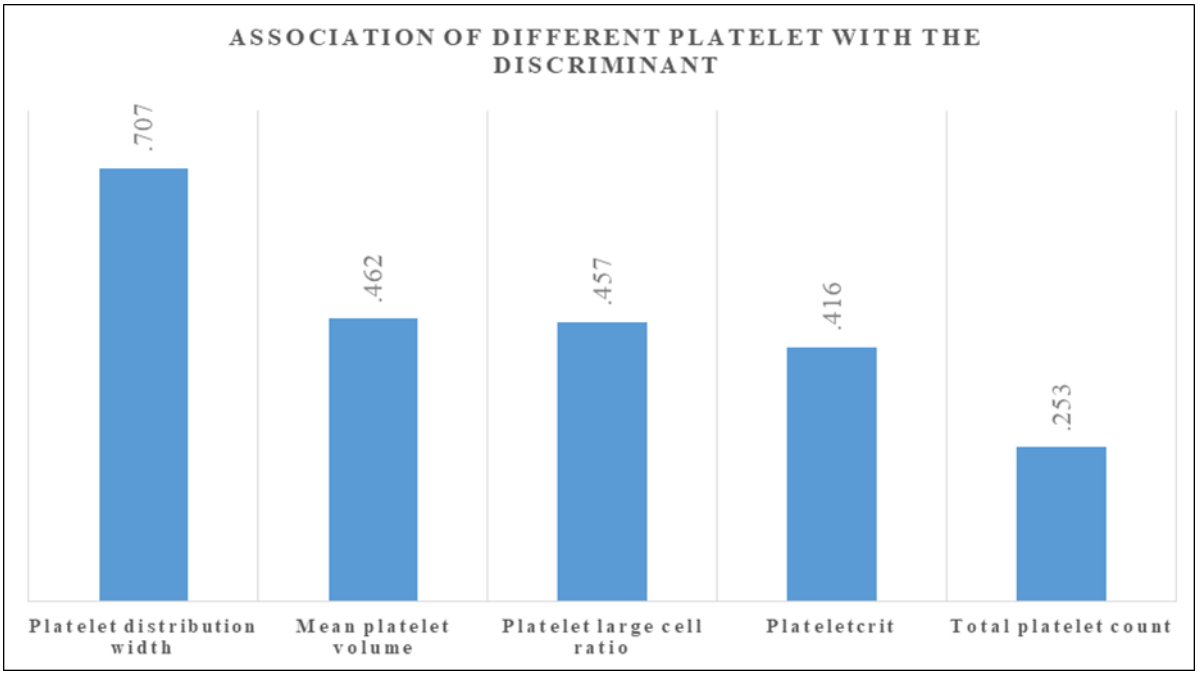


Fig 3: Association between the different platelet parameters with the discriminant function

The coefficients of each platelet parameter in the linear discriminant weights are depicted in the Figure 3. Platelet distribution width exhibits a notably high positive correlation with gender. Additionally, all platelet parameters demonstrate positive correlations with gender. Consequently, these four platelet parameters are significantly influenced by the gender.

5. Conclusion

In this study, the analysis of the platelet parameter dataset reveals valuable insights into population characteristics and their associations with platelet biology. The dataset represents a diverse population and includes comprehensive information on platelet parameters and red cell distribution width. Descriptive statistics highlight variability and distribution characteristics, while normality tests indicate the effectiveness of transformation methods. Multivariate analysis confirms significant overall effects of platelet parameters, and correlation analysis identifies potential interdependencies among variables. Fisher Linear discriminant analysis underscores the influence of platelet parameters on gender classification, with acceptable levels of multicollinearity ensuring reliability. Moreover, tests for covariance matrix equality and canonical analysis provide further insights. These findings contribute to a better understanding of platelet biology's importance in medical research and clinical practice. The Platelet parameters exhibit significant associations with population characteristics and gender, which are crucial for medical research and clinical applications.

6. References

1. AlKubaisi M, Aziz WA, George S, Al-Tarawneh K. Multivariate discriminant analysis managing staff appraisal case study. Acad Strategic Manag J. 2019;18(5):1-12.
2. Box GE. A general distribution theory for a class of likelihood criteria. Biometrika. 1949;36(3/4):317-346.
3. Box GE, Cox DR. An analysis of transformations. J R Stat Soc Ser B Stat Methodol. 1964;26(2):211-243.
4. Cramer D. Advanced quantitative data analysis. McGraw-Hill Education (UK); c2003.
5. Dibal PN, Abraham AC. On applying linear discriminant function to evaluate data on diabetic patients at the University of Port Harcourt Teaching Hospital, Rivers, Nigeria. Am J Theor Appl Stat. 2020;9(3):53-56.
6. Ding Z, Xu Y, Zhong K. Exponential Local Fisher Discriminant Analysis with Sparse Variables Selection: A Novel Fault Diagnosis Scheme for Industry Application. Machines. 2023;11(12):1066.
7. Durrant RJ, Kabán A. Compressed Fisher linear discriminant analysis: Classification of randomly projected data. In: Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining. 2010:1119-1128.
8. Fisher RA. The use of multiple measurements in taxonomic problems. Ann Eugen. 1936;7(2):179-188.
9. Goldstein R. Conditioning Diagnostics: Collinearity and Weak Data in Regression. J Oper Res Soc. 1993;35(1):85-86.
10. Henry J, Veazie P, Furman M, Vann M, Whipker B. Spectral discrimination of micronutrient deficiencies in greenhouse-grown flue-cured tobacco. Plants. 2023;12(2):280.
11. Jang DH, Anderson-Cook CM, Kim Y. Graphical methods for the sensitivity analysis in discriminant analysis. Commun Stat Appl Methods. 2015;22(5):475-485.
12. Johnson RA, Wichern DW. Applied Multivariate Statistical Analysis. Biometrics. 1998;54(3):1203.
13. Liberda EN, Zuk AM, Martin ID, Tsuji LJ. Fisher's linear discriminant function analysis and its potential utility as a tool for the assessment of health-and-wellness programs in indigenous communities. Int J Environ Res Public Health. 2020;17(21):7894.
14. McGarigal K, Stafford S, Cushman S. Discriminant analysis. In: Multivariate statistics for wildlife and ecology research. 2000:129-187.
15. Nainggolan O, Tjandrarini DH, Indrawati L. Discriminant Analysis to Predict the Hypertension in Women Aged 25–54 Years. Glob J Health Sci. 2018;10(10):93.
16. Ndako JA, Olisa JA, Ifeanyichukwu IC, Okolie CE, Ojo

SK, Jegede SL. Predictive evaluation of pediatric patients based on their typhoid fever status using linear discriminant model. Med Hypotheses. 2020;144:110264.
17. Onwukwe CE, Ogbonna EN, Ayeni A. Application of Wilks' Lambda and Hotelling's T2 with MANOVA on Drug Addiction and Drug Abuse Data. Adv Life Sci Technol. 2014;24:53-59.
18. Rahamneh AAA, Jresat SS, Zubaidi F, Al-Hawary SIS. Using the Linear Discriminant Analysis Method to Classify Types of Bowels and Esophageal cancer in Jordan. Inf Sci Lett. 2023;12(3):1299-1305.
19. Ramayah T, Ahmad NH, Halim HA, May-Chiun SRMZ. Discriminant analysis: An illustrated example. Afr J Bus Manag. 2010;4(9).
20. Selvajothi P, Packiaraj R. Dataset of platelet parameters and red cell distribution width. Mendeley Data. 2021. doi:10.17632/5t8dr6d73f.1
21. Shrestha N. Detecting Multicollinearity in Regression Analysis. Am J Appl Math Stat. 2020;8(2):39-42. DOI: 10.12691/jams-8-2-1
22. Vélez JI, Correa JC, Marmolejo-Ramos F. A new approach to the Box-Cox transformation. Front Appl Math Stat. 2015;1:12. DOI: 10.3389/fams.2015.00012
23. Young DS. Handbook of regression methods. Chapman and Hall/CRC; c2018.