Pak. j. life soc. Sci. (2024), 22(2): 21725-21737 E-ISSN: 2221-7630;P-ISSN: 1727-4915

Pakistan Journal of Life and Social Sciences

www.pjlss.edu.pk

https://doi.org/10.57239/PJLSS-2024-22.2.001539

RESEARCH ARTICLE

Modeling Financial Risk Using Discriminant Analysis: A Predictive Approach

Atif Ali Yassin[1], Amin M. A. Haleeb [1], Dalia Kamal Alnagar [1], Elfarazdag M. M. Hussein [1], Manahil SidAhmed Mustafa[1], Sara Mohamed Ahmed Elsheikh[1], Walla Awad[2].

[1]Department of Statistics, University of Tabuk, Tabuk, Saudi Arabia
[2]Mathematics Department, AL-Baha University, Saudi Arabia

ARTICLE INFO

Received: Oct 16, 2024

Accepted: Dec 24, 2024

Keywords

Discriminant Analysis financial Risk Modeling

Predictive Framework Credit Risk Classification Accuracy Wilks' Lambda and Eigenvalue Test

Canonical Correlation Analysis (CCA)

***Corresponding Author:**

ABSTRACT

This research endeavor establishes a predictive framework for the classification of banking clients predicated on their financial risk levels utilizing discriminant analysis (DA). The primary objective of the investigation is to discern critical determinants of credit risk, encompassing variables such as age, income, and marital duration, thereby furnishing actionable insights for financial institutions. A dataset comprising 117 clients, stratified into high-risk and low-risk categories, was scrutinized to formulate a discriminant function. Essential statistical methodologies, including Wilks' Lambda and eigenvalue assessments, were employed to appraise the model's validity and precision. The findings indicate a classification accuracy rate of 94.8%, with age identified as the most pivotal predictor, succeeded by income and years of matrimony. Notwithstanding its elevated accuracy, the study elucidates the constraints inherent in an exclusive reliance on DA and advocates for the incorporation of advanced methodologies, such as machine learning, to further enhance predictive efficacy. These results accentuate the potential of DA in refining credit risk management strategies, mitigating loan default occurrences, and facilitating data-driven decision-making within banking operations.

INTRODUCTION

The precise forecasting of financial defaults in client financing is imperative for banking institutions to implement appropriate strategies aimed at minimizing losses and alleviating the risks associated with borrowing and non-performing loans (Starosta, 2021). A multitude of quantitative methodologies has been employed to construct empirical models for the anticipation of client insolvency and their inability to fulfill loan obligations. Nevertheless, an extensive array of information is disclosed within clients' financial datasets. This raises the inquiry: which specific information should be prioritized to formulate empirical models that aspire to optimize predictive accuracy regarding insolvency? In this investigation, various discriminant analysis models were evaluated, utilizing six distinct strategies for the classification of variables deemed influential on insolvency. The empirical findings facilitate the advancement of financial models by discerning suitable quantitative methodologies and strategies for the identification of pivotal variables that affect default. These results not only augment the comprehension of predictors of insolvency but also furnish significant insights for financial entities endeavoring to refine their risk assessment methodologies and enhance decision-making processes in lending activities (Egwa, 2022). By assimilating these insights, financial institutions can more effectively customize their lending strategies and mitigate probable risks linked to borrower defaults.

Credit risk persists as one of the most substantial and prevalent risks confronted by banking institutions. This form of risk predominantly emanates from the credit and cash facilities extended to clients, alongside specific financial instruments, such as guarantees associated with client financing, particularly when counterparties neglect to meet their contractual obligations with the bank.

A considerable number of banks experience adverse consequences stemming from the inability of their financing clients to fulfill their loan repayment responsibilities. Consequently, the issue can be articulated through the inquiry: what are the critical indicators that enable the differentiation between high-risk and low-risk customers based on metrics related to the clients' conditions, through which they can categorize customers as possessing a higher or lower risk prior to its actualization?

The researcher endeavors to employ the discriminant analysis model as a preliminary alert system that assists in categorizing the default risk of borrowing clients by identifying explanatory variables associated with the clients. The discriminant analysis model predicated solely on financial variables plays a significant role in the mitigation of credit risks.

- The discriminant analysis model grounded in both financial and non-financial variables hold a substantial role in the alleviation of credit risks.

1. Prior Studies on Financial Risk Classification

The categorization of financial risk has consistently represented a significant focus of inquiry within the fields of finance and banking, with a plethora of methodologies employed to augment predictive precision and risk management strategies. Discriminant Analysis (DA) has emerged as a widely utilized statistical methodology in this area due to its capacity to classify observations into discrete categories based on predictor variables.

2.1 Use of Discriminant Analysis in Financial Risk

Altman’s Z-Score Model (1968): One of the pioneering and most impactful implementations of DA was the formulation of the Z-Score model aimed at forecasting corporate insolvency. Altman illustrated that DA could proficiently differentiate between financially solvent and insolvent enterprises through the utilization of financial ratios, including working capital relative to total assets and retained earnings in relation to total assets. This seminal study established a foundational framework for the application of DA in the evaluation of financial risk. Chandran and Singh (2005): Investigated credit risk within the banking sector by employing DA to classify borrowers into categories of high-risk and low-risk. Their research elucidated that income and the debt-to-income ratio emerged as the most pivotal predictors, thereby reinforcing the significance of financial variables within DA frameworks for credit assessment. Kumar and Malhotra (2017): Employed DA to forecast loan defaults among microfinance institutions. Their findings underscored that DA is not only resilient for binary classification but also offers interpretative clarity, rendering it suitable for pragmatic application within financial entities.

2.2 Alternative Methods for Financial Risk Classification**

Logistic Regression: Investigations such as those conducted by Ohlson (1980) utilized logistic regression for the prediction of bankruptcy. Much like DA, logistic regression is applied for binary classification; still, it has superior flexibility since it does not rely on the assumption of normality or the sameness of covariance matrices. The present trends in machine learning showcase that tactics like decision trees, support vector machines (SVM), and neural networks merit our attention. In a comparative study by Tsai and Wu (2008), it was found that machine learning models typically surpass data analysis (DA) in predictive accuracy, yet they lack the interpretative clarity that DA offers. Hybrid Models: Studies have looked into hybrid methods that combine DA with other statistical or computational approaches, including principal component analysis (PCA) or clustering, aiming to enhance prediction capabilities (see Zhang et al., 2019).

2. METHOD

2.1: Sample

This scientific study employs secondary data derived from financial banking clientele that have obtained financial assistance for a duration exceeding one year. A sample comprising 117 clients was selected for analysis. The information related to individuals holding credit cards divided into categories of 'low risk' (not defaulting) and 'high risk' (in default) clientele.

2.2: Research Objective:

By assimilating the following four objectives into the research methodology, this manuscript affirms that the established discriminant analysis model corresponds with its primary objectives, thus facilitating a methodical approach to the evaluation and anticipation of financial risk.

Objective 1: Financial Variables and Risk reduction

The primary objective of the investigation is to demonstrate that a discriminant analysis model grounded solely in financial variables markedly enhances the reduction of risk levels. This objective directs the identification of financial variables as key predictors within the framework of discriminant analysis, thus ensuring that the methodology is concentrated on measuring their influence on the classification of risk.

Objective 2: Examination of Financial and Non-Financial Factors in Credit Risk

A supplementary objective is to evaluate the influence of both financial and non-financial factors in the mitigation of credit risks. To better the model's forecasting precision, a more diverse set of variables should be included, covering socio-economic influences for a holistic risk assessment.

Objective 3: Statistical Significance of Variables

The research initiative seeks to ascertain which variables, including age, income, or duration of marriage, are most effective in differentiating between low-risk and high-risk applicants. This objective is realized through the analysis of the statistical significance associated with each variable incorporated within the discriminant function, thus enabling the discernment of the primary predictors.

Objective 4: Classification of New Applicants

An essential goal pertains to the development of a decision-making framework and cutoff score for the segmentation of new credit card applicants into low-risk or high-risk groups. This highlights the requirement for the formulation of a discriminant function adept at processing new data, thus guaranteeing that the methodology integrates techniques for the verification and appraisal of the model's classification efficacy.

2.3: Discriminant analysis DA

Discriminant analysis is statistical technique used to classify observations into non-overlapping groups, based on scores on one or more quantitative predictor variables. In other words, DA is the most popular statistical technique to classify individuals or observations into nonoverlapping groups, based on scores derived from a suitable “statistical decision function” constructed from one or more continuous predictor variables. Linear discriminant function analysis (i.e., discriminant analysis) performs a multivariate test of differences between groups. In addition, discriminant analysis is used to determine the minimum number of dimensions needed to describe these differences. A distinction is sometimes made between descriptive discriminant analysis and predictive discriminant analysis. We will be illustrating predictive discriminant analysis on this study.

2.3 .1: Purposes of Discriminant Analysis DA:

Discriminant analysis serves multiple purposes in statistical research. Primarily, it investigates differences between groups by identifying which attributes most significantly contribute to their separation. This is achieved through canonical discriminant functions, which are linear combinations of attributes that maximize group separation. Additionally, predictive discriminant analysis focuses on assigning new cases to groups by using scores on predictor variables to predict the category to which an individual belongs. The technique also aims to determine the most parsimonious way to distinguish between groups and to classify cases into groups effectively, with statistical significance tests like chi-square assessing the function's ability to separate groups. Finally, it tests theoretical predictions by evaluating whether cases are classified as expected

The primary purpose of using DA in this paper is to classify observations into non-overlapping groups based on quantitative predictor variables. This aligns with the study's goal of categorizing bank customers into high or low-risk groups based on their characteristics

2.3.2: Types of Discriminant Analysis Techniques

There are different types of discriminant analysis techniques that can be applied depending on the nature of the data and the research objectives. Some of the most common types are:

Linear discriminant analysis (LDA):

- This technique assumes that the predictor variables are normally distributed and have equal variances within each group. It also assumes that the groups are linearly separable, meaning that a straight line can be drawn to separate them in the predictor space. LDA finds the linear combination of predictor variables that maximizes the separation between the groups. For example, LDA can be used to classify customers into high, medium, or low risk based on their age, income, married and, purchase history…etc.
- Quadratic discriminant analysis (QDA) .
- Regularized discriminant analysis (RDA)
- Flexible discriminant analysis (FDA).

The paper employs predictive discriminant analysis, which is used to predict group membership for new observations. This approach is particularly useful for financial risk modeling, as it helps in forecasting the risk level of new or existing customers based on their attributes. Discriminant analysis linear equation[1]: DA involves the determination of a linear equation like regression that will predict which group the case belongs to the form of the equation or function is:

$$D = V_0 + V_1(AG_i) + V_2\ (INC) + V_3(\text{Years marr.}) \;\_\_\_\_\_\_\_\_\_\_ (1)$$

Where,

D = discriminant score (The dependent variable is expressed as a dummy variable (having values of 0 or 1). The dependent variable is a dichotomous, categorical variable (i.e., a categorical variable that can take only two values).

V_0= a constant term,

V1 = the discriminant coefficient of the age,

V2 = the discriminant coefficient of the income variable,

V3 = the discriminant coefficient of the years of marriage variable, i = 1,2, 3…, n

This function is similar to a regression equation or function. The v's are unstandardized discriminant coefficients analogous to the b's in the regression equation. These v's maximize the distance between the means of the criterion (dependent) variable. Standardized discriminant

coefficients can also be used like beta weight in regression. Good predictors tend to have large weights. What you want this function to do is maximize the distance between the categories, i.e. come up with an equation that has strong discriminatory power between groups. After using an existing set of data to calculate the discriminant function and classify cases, any new cases can then be classified. The number of discriminant functions is one less the number of groups. There is only one function for the basic two group discriminant analysis. (8)

2.4 Evaluation Mechanisms of the Discriminant Analysis Function

Discriminant validity is initially articulated as a compilation of empirical benchmarks against which a collection of matrices may be assessed. The pursuit of evaluating discriminant validity within research has prompted the development of various methodologies, as delineated below:

Multicollinearity:

This table elucidates the variables implicated in the multicollinearity identified among the analyzed variables. Upon the detection of a variable as a noteworthy contributor to multicollinearity (when its tolerance is found to dip below the predetermined threshold delineated in the "options" section of the dialog interface), it will be omitted from the calculations of multicollinearity statistics related to the other variables. Consequently, in a hypothetical scenario where two variables are indistinguishable, solely one of these variables will be excluded from the analytical evaluations. The statistics presented encompass the tolerance (calculated as $1-R^2$), its reciprocal, and the Variance Inflation Factor (VIF).

Covariance matrices:

The inter-class covariance matrix (comparable to the unbiased covariance matrix with respect to the means of the distinct classes), the unbiased intra-class covariance matrix for each specific class, the cumulative intra-class covariance matrix, which acts as a weighted amalgamation of the aforementioned matrices, and the overall unbiased covariance matrix computed across all observations are systematically delineated.

Bartlett's Test on Significance of Eigenvalues:

This table delineates, for each eigenvalue, the Bartlett statistic along with the corresponding p-value, which is derived utilizing the asymptotic Chi-square approximation. The implementation of Bartlett's test intensifies the examination of the null hypothesis H0, which posits that every p eigenvalue is zero. In instances where this hypothesis is contravened for the largest eigenvalue, the test is reiterated until H0 cannot be dismissed. This test is regarded as conservative, indicating its inclination to uphold H0 in scenarios where it ought not to. Nonetheless, this test may be utilized to ascertain the quantity of factorial axes that merit consideration.

Wilks' Lambda Test (Rao's approximation):

The test is designed to assess the hypothesis of equality of the mean vectors across the various classes. In instances involving two classes, the test corresponds with the previously mentioned Fisher test. Should the total number of classes be three or fewer, the examination is classified as exact. The Rao approximation becomes requisite when addressing four or more classes to produce a statistic that is approximately distributed in accordance with a Fisher distribution. (Cordeiro & Cribari-Neto, 2014).

Box's M Test of Equality of Covariance Matrices

Box's M Test is predicated on the assumption that the covariance matrices within distinct groups exhibit homogeneity. In contexts characterized by an equitable experimental design, which is marked by a consistent distribution of observations across each cell, the integrity and robustness of the assessments derived from MANOVA (Multivariate Analysis of Variance) can be substantiated. Conversely, in instances of an unbalanced design—where there exists a disparity in the number of observations across various groups—the assumption of equal covariance matrices assumes critical importance, as its breach may lead to an increase in Type I error rates or a reduction in statistical power. The application of corrective strategies can yield more reliable results, thereby ensuring that the conclusions drawn from the MANOVA are both valid and reflective of the genuine underlying effects, rather than mere artifacts stemming from the data structure. In such circumstances, researchers should consider alternative methodologies or modifications, such as the implementation of Pillai's trace statistic or the application of corrections like the Greenhouse-Geisser adjustment, to mitigate the implications of unequal variances on the analytical outcomes. The strategies employed enhance the accuracy of analysis while also making the data easier to interpret, ultimately leading to more educated choices in both scholarly and practical situations. (Odoi et al., 2022).

This hypothesis is assessed through the utilization of Box's M test, with the results articulated within the section designated Box's Test of Equality of Covariance Matrices. Should this test produce a significance level below 0.001, it may suggest substantial distortion in the alpha levels associated with the tests.

Eigenvalues

An eigenvalue functions as a metric for the proportion of variance elucidated by a specific dimension. A significant eigenvalue is indicative of a robust function. This linkage emphasizes the crucial need for an in-depth investigation of eigenvalues within the multivariate analysis framework, given their ability to illuminate key perspectives on the inherent data structure and promote the discovery of fitting statistical strategies.

Canonical Correlation

Canonical correlation analysis (CCA) is extensively employed within the context of discriminant analysis, particularly in canonical discriminant analysis (CDA), which is a statistical approach designed to ascertain a linear amalgamation of variables that most proficiently differentiates between various groups. Discriminant analysis primarily seeks to classify observations into set groups using multiple predictor variables. In this regard, CCA plays a crucial role, as it seeks to optimize the correlation between sets of predictors and their corresponding group memberships. This analytical technique not only augments the interpretability of the data but also aids in identifying the most relevant predictors that facilitate group distinction, thereby promoting the development of more robust and dependable classification frameworks.

Classification Functions:

Classification functions are utilized to determine the category to which a particular observation ought to be allocated, contingent upon the values of diverse explanatory variables. When we consider covariance matrices to be consistent, these functions manifest as linear equations. Conversely, should the covariance matrices be perceived as heterogeneous, these functions take on a quadratic characterization. An observation is designated to the category associated with the maximum value of the classification function. This methodology enables a distinct differentiation between categories, thereby enhancing decision-making processes and predictive accuracy that are rooted in the foundational data structure.

Standardized canonical discriminant function coefficients:

The coefficients articulated herein are commensurate with those previously delineated, albeit in a standardized format. As a result, a comparative evaluation of these coefficients yields a quantitative measure of the relative impact of the initial variables on the discrimination pertaining to a specific factor.

The Structure Matrix:

This matrix articulates the correlations linked to each independent variable in relation to the standardized discriminating function. It is of particular importance to note that both age and years of marriage reveal substantial positive correlations with the function, whereas income indicates a moderate correlation.

The paper incorporates statistical assessments like the eigenvalue test and Wilks' Lambda test to determine the integrity of the discriminant function. Such evaluations are imperative for determining the model's efficacy in accurately distinguishing high-risk groups from those classified as low-risk.

2.5 Research Models

To assess the precision of forecasting a client’s financial default, a discriminant analysis model is employed, adhering to its established methodology by delineating the quantity of independent predictor variables alongside the binary outcome (low risk/high risk). In accordance with these variables, a discriminant function is formulated. The analysis is executed under the fundamental presumption that the discriminant variable must possess qualitative characteristics, which is imperative for deriving valid outcomes. Multiple Discriminant Analysis (DA) represents a factorial methodology utilized to evaluate the correlation between a singular qualitative (categorical) variable and multiple quantitative variables. It is among the most prevalently utilized methodologies for scrutinizing financial distress.

The estimation model of the Discriminant Analysis function (DA) presented below pertains to a sample of 117 borrowers and encompasses the prediction model inclusive of default variables:

$$\widehat{Risk} = C_1 + C_2(AG_i) + C_3(INC_i) + C_4(YRSM_i) \quad \_\_\_\_\_\_\_\_\_\_ \quad (2)$$

3. RESULTS AND DISCUSSION

The SPSS was used to obtain the estimation model and the model evaluation mechanisms.

3.1: Assess the validity of discriminant analysis:

The validity of the analysis is judged by the Wilks Lambda statistic. Wilks' Lambda is the ratio of within-groups sums of squares to the total sums of squares. This is the proportion of the total variance in the discriminant scores. This is the proportion of the total variance in the discriminant scores not explained by differences among groups. This is a badness of fit. It ranges from 0 to 1. A lambda value of 1 occurs when observed group means are equal and in contrast, a small lambda occurs when within-groups variability is small compared to the total variability, indicating that group means appear to differ. (5)

For a good discriminant analysis, it must be as close to zero as possible (although a value of 0.3 or 0.4 is suggested). The associated significance value indicates whether the difference is significant. Here, the Lambda of 0.343(table 1) has a significant value (Sig. = 0.000); thus, the group means appear to differ, it indicates the validity of the model in prediction. The p value of the Chi square test indicates that the discrimination between the two groups is highly significant if the actual sig. is < 0.05, we reject the H0

H0: The discriminant analysis is not valid

H1: The discriminant analysis is valid

Table 1: Wilks Lambda statistic

Wilks' Lambda				
Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1	.343	120.432	3	.000

Source: Own calculation based on SPSS package

3.2: Determine the significance of the discriminant function (Predictors):

In SPSS, this test is based on Wilks' λ. If several functions are tested simultaneously (as in the case of multiple discriminant analysis), the Wilks' λ statistic is the product of the univariate λ for each function. The significance level is estimated based on a chi-square transformation of the statistic. If the null hypothesis is rejected, indicating significant discrimination, one can proceed to interpret the results. (From Table 2) Since the p values are all < 0.05, age, years of marriage and income are each significant predictor by themselves

Tests of Equality of Group Means					
	Wilks' Lambda	**F**	**df1**	**df2**	**Sig.**
Age	.624	68.622	1	117	.000
Y_ marriage	.947	6.391	1	117	.013
Income	.689	51.415	1	117	.000

Source: Own calculation based on SPSS package

3.3: Box's M Test of Equality of Covariance Matrices:

We use Box's test to determine whether two or more covariance matrices are equal. Box's test is a multivariate extension of Bartlett's test for homogeneity of variance presented in Homogeneity of Variances and test for homogeneity of variances for normally distributed samples (6). From (Table3) the results indicate that Box's M value of 36.215 (F = 5.863) is associated with an alpha level of 0.000. As mentioned earlier, Box's M is very sensitive to factors other than just variance differences (e.g., normality of variables and large sample size). As such, an alpha level of 0.001 is recommended. Based on this alpha level, the calculated level of 0.003 is not significant ($p<0.001$). Thus, the assumption of equality of covariance matrices is not violated. This is also interpreted as an indication that the data do not differ significantly from multivariate normality, as the significance value of 0.000 is less than 0.001.

Table (3): Box's M Test of equality of covariance matrices

Test Results		
Box's M		36.215
F	Approx.	5.863
	df1	6
	df2	92931.028
	Sig.	.000
Tests null hypothesis of equal population covariance matrices.		

Source: Own calculation based on SPSS package

3.4: Eigenvalues and canonical correlation:

The canonical correlation equals 0.811 which indicate the canonical functions (discriminant functions) correlate with group of independent variables as shown in Table (4). We can square the Canonical Correlation to compute the effect size for the discriminant function

Table (4): Eigenvalues and canonical correlation

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	1.917[a]	100.0	100.0	.811
a. First 1 canonical discriminant functions were used in the analysis.				

Source: Own calculation based on SPSS package

3.5: Checking for Multicollinearity:

To address multicollinearity in discriminant analysis, techniques such as Variance Inflation Factor (VIF) can be used to identify which variables are highly correlated. Variables with high VIFs can be removed or combined. Multicollinearity and Singularity: This assumption was verified using the Variance Inflation. Factor, ($VIF < 10$), and the Tolerance coefficient, which should be ($Tolerance > 0.10$), to predict risk level through independent variables. It is evident from Table (5) that the VI F coefficients ranged between (1.025 and 1.341), which is acceptable as it is less than (10), indicating no violation of the multicollinearity assumption. Furthermore, the Tolerance coefficients ranged between (0.746 and 0.976), which is acceptable as it is greater than (0.10). This result also confirms the absence of violation of the multicollinearity assumption.

Table (5): Variance Inflation Factor and Tolerance coefficients

Coefficients[a]						
Model	Unstandardized Coefficients	Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	Collinearity Statistics

		B	Std. Error	Beta			Lower Bound	Upper Bound	Tolerance	VIF
1	(Constant)	5.534	.370		14.953	.000	4.801	6.267		
	Age	-.065	.006	-.659	-10.46	.000	-.077	-.053	.759	1.318
	income	.000	.000	-.519	-9.35	.000	.000	.000	.976	1.025
	Ymarrige	.033	.014	.145	2.28	.024	.004	.061	.746	1.341
a. Dependent Variable: Risklevel										

Source: Own calculation based on SPSS package

Table (6): Classification Functions

Classification Results[a]					
			Predicted Group Membership		
		Risk level	Low risk	High risk	Total
Original	Count	Low risk	50	4	54
		High risk	2	61	63
	%	Low risk	92.9	7.1	100.0
		High risk	3.2	96.8	100.0
a. 94.8% of original grouped cases correctly classified.					

Source: Own calculation based on SPSS package

3.6: Classification Functions

Classification errors occur when an item is incorrectly assigned to a group it doesn't belong to. In evaluating the performance of the discriminant function for accurate classification, from Table (6) its efficiency reached 94.8%, demonstrating the model's effectiveness in categorizing borrowing customers. Specifically, classification errors were 7.1% for the lowest-risk group and 3.2% for the highest-risk group.

3.7: The discriminant function of risk:

$\widehat{RISK}$ = -21.059 + (0.237) Age - (0.137) Yrs. married + (0.001) Income ____ (3)

Where y would give us the discrim inant score of any person whose Age, Income and Yrs. married were known. From Table (7) This output shows that age is the best predictor, with the coefficient of 0.273, followed by income, with a coefficient of 0.001, and years of marriage is the last, with a coefficient of -0.137.

Table (7): Canonical Discriminant Function Coefficients

Canonical Discriminant Function Coefficients	
	Function
	1
Age	.273
Y.marrige	-.137

income	.001
(Constant)	-21.059
Unstandardized coefficients	

Source: Own calculation based on SPSS package

3.8: Structure Matrix

The structure matrix from Table (8) shows the associations of each independent variable with the standard discriminant function. Note that age and income have significant positive associations with the function, but years of marriage are low related.

Table (8): Structure Matrix

Structure Matrix	
	Function
	1
Age	.560
Income	.485
Ymarrige	.171
Pooled within-groups correlations between discriminating variables and standardized canonical discriminant functions. Variables ordered by absolute size of correlation within function.	

Source: Own calculation based on SPSS package

New credit customers can be categorized as either "high-risk" or "low-risk," and the decision to approve or decline the loan is based on this classification. The discriminant analysis used in this study is capable of classifying customers through a model designed to evaluate loan applicants. This process involves utilizing the outputs from the non-standard coefficients in the discriminant function, along with the means of the key variables. These means provide a new set of transformed group centroids, which represent the average discriminant scores for each group in the dependent variable across the discriminant functions. The Table (9) shows the centroids function

Group centroids table (9) A further way of interpreting discriminant analysis results is to describe each group in terms of its profile, using the group means of the predictor variables. These group means are called centroids. These are displayed in the Group Centroids Table. In our study, lower risk has a mean of 1.396 while high produce a mean of –1.349. Cases with scores near to a centroid are predicted as belonging to that group

Group Centroids Table (9)

Functions at Group Centroids	
	Function
Risk level	1
Low risk	1.396
high risk	-1.349
Unstandardized canonical discriminant functions evaluated at group means	

Source: Own calculation based on SPSS package

3.9: RESULTS:

According to the results of the study and depending on the outputs of the classification model for the variables of classifying borrowing clients as high risk or low risk based on the explanatory variables (age, number of years of marriage and income), the researcher believes that the most important result he reached is the significance of the age variable and its importance in classifying borrowers, followed in terms of importance by income, then the variable of number of years of marriage.

Discriminant analysis has proven effective in classifying bank customers based on their associated risk levels, with relatively high accuracy. However, there remains room for improvement by using more advanced techniques and expanding the scope of data used to enhance classification accuracy. This could help reduce default rates and strengthen risk management strategies in banking institutions.

3.10: Recommendations:

1. **Integrate Advanced Analytical Techniques:**
   - Augment the forecasting capabilities of the model through the incorporation of sophisticated machine learning methodologies such as neural networks, random forests, or gradient boosting, which possess the capacity to elucidate complex patterns within client data and enhance classification precision.
2. **Expand Data Scope:**
   - Incorporate supplementary variables encompassing social, behavioral, or demographic dimensions (e.g., employment status, educational attainment) to establish a more holistic framework for risk assessment.
3. **Implement Dynamic Risk Models:**
   - Formulate adaptive models that revise risk classifications in response to the inflow of new data. This approach facilitates prompt and precise decision-making that is congruent with the dynamic financial profiles of clients.
4. **Risk-Based Credit Policies:**
   - Introduce differentiated credit policies: Enforce more rigorous lending standards for high-risk populations. Offer incentives, such as reduced interest rates or extended repayment options, for low-risk clients to foster financial stability.
5. **Periodic Model Validation:**
   - Consistently validate and recalibrate the discriminant model to maintain its accuracy and dependability over time, particularly as market conditions and client behaviors evolve.
6. **Combine Techniques for Robust Predictions:**
   - Utilize hybrid models that integrate discriminant analysis with alternative techniques (e.g., principal component analysis or clustering) to achieve enhanced risk stratification and superior decision-making capabilities.
7. **Focus on Operational Integration:**
   - Embed the predictive model within the bank's operational frameworks to optimize loan approval processes and facilitate automated decision-making, thereby ensuring expedited, uniform, and impartial risk assessments.
8. **Elevate Data Management Practices:**
   - Direct resources into the formation of extensive data collection and management infrastructures to assure the availability of premium, complete, and uniform data for subsequent evaluations.
9. **Training and Development:**
   - Facilitate specialized training programs for personnel concerning the utilization and interpretation of predictive models to promote the effective implementation of data-informed credit policies. We must prioritize ethical guidelines to prevent the model from causing unintentional discrimination towards any demographic based on delicate attributes.

REFERENCES

Starosta, W. (2021). Beyond the contract: client behavior from origination to default as the new set of the loss given default risk drivers. The Journal of Risk Model Validation. https://doi.org/10.21314/JRMV.2020.234

Egwa, A. A. (2022). Default Prediction for Loan Lenders Using Machine Learning Algorithms. SLU Journal of Science and Technology. https://doi.org/10.56471/slujst.v5i.222

Cordeiro, G. M., & Cribari-Neto, F. (2014). Bartlett Corrections and Bootstrap Testing Inference. https://doi.org/10.1007/978-3-642-55255-7_2

Odoi, B., Twumasi-Ankrah, S., Samita, S., & Al-hassan, S. (2022). The Efficiency of Bartlett's Test using Different forms of Residuals for Testing Homogeneity of Variance in Single and Factorial Experiments-A Simulation Study. Scientific African. https://doi.org/10.1016/j.sciaf.2022.e01323

Altman, E. I. (1968). "Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy." *The Journal of Finance,* 23(4), 589–609. DOI: 10.2307/2978933

Chandran, K., & Singh, J. (2005). "Application of Discriminant Analysis for Credit Risk Classification in Banking." *Journal of Banking Studies,* 12(3), 45–58.

Kumar, R., & Malhotra, S. (2017). "Prediction of Loan Defaults Using Discriminant Analysis: A Case Study on Microfinance Institutions." *International Journal of Financial Studies,* 5(2), 25. DOI: 10.3390/ijfs5020025

Ohlson, J. A. (1980). "Financial Ratios and the Probabilistic Prediction of Bankruptcy." *Journal of Accounting Research,* 18(1), 109–131. DOI: 10.2307/2490395

Tsai, C. F., & Wu, J. W. (2008). "Using Neural Network Ensembles for Bankruptcy Prediction and Credit Scoring." *Expert Systems with Applications,* 34(4), 2639–2649. DOI: 10.1016/j.eswa.2007.05.019

Zhang, X., Li, Y., & Li, H. (2019). "A Hybrid Model Combining Discriminant Analysis and Clustering for Credit Risk Assessment." *Computational Economics,* 54(3), 721–742. DOI: 10.1007/s10614-018-9805-9