International Journal of Physics and Applications

E-ISSN: 2664-7583
P-ISSN: 2664-7575
IJOS 2025; 7(1): 16-19

www.physicsjournal.in
Received: 09-11-2024
Accepted: 14-12-2024

Ravi Kumar Verma
Assistant Professor, Model College, Suggabathan, Godda Sido Kanhu Murmu University, Dumka, Jharkhand, India

Corresponding Author:
Ravi Kumar Verma
Assistant Professor, Model College, Suggabathan, Godda Sido Kanhu Murmu University, Dumka, Jharkhand, India

A study on the applications of discriminant analysis

Ravi Kumar Verma

DOI: https://doi.org/10.33545/26647575.2025.v7.i1a.121

Abstract

One of the challenging tasks facing a researcher is the data analysis section where the researcher needs to identify the correct analysis technique and interpret the output that he gets. The analysis wise is very simple, just by the click of a mouse the analysis can be done. The more demanding part is the interpretation of the output that the researcher gets. Many researchers are very familiar and well exposed to the regression analysis technique whereby the dependent variable is a continuous variable.

But what happens if the dependent variable is a nominal variable? Then the researcher has 2 choices: either to use a discriminant analysis or a logistic regression. Discriminant analysis is used when the data are normally distributed whereas the logistic regression is used when the data are not normally distributed. The purpose of this paper is to help novice researchers as well as seasoned researchers on how best the output from the SPSS can be interpreted and presented in standard table forms. Each application represents a real practical case of applying the discriminant analysis technique, finally a general conclusion about discriminant analysis.

Keywords: SPSS, regression analysis, data, researcher, dependent variable, discriminant analysis

Introduction

Discriminant Analysis technique was afforded by Fisher in 1936 as a statistical approach for discriminating new data into two separate groups. It is a statistical method that is a useful technique for the researchers to investigate the differences between two or more sets of variables regarding to different variables simultaneously. The technique is also studied the relationship among categorical variable and a set of variables that is not related to each other. It offers to the researcher how they can separate between two groups by using linear relationship function between the variables to determine which variables are important to do the separation process and how to separate the unknown objects into one set of the groups. It is a classical discrimination method that used to separate a single categorical variable by using different attributes. Also, it appoints the observations to one of the predetermined set of data taking into account the knowledge of the different-attributes. It is frequently used to establish a predictive and descriptive model of group separation depending on the observed predictor variables and on classifying each observation into one of the groups. This technique was used by Fisher for the first time (1936), a researcher who was looking for solving a problems in physical anthropology and biology fields, this method also was used in the field of social science by (Tastuoka and Tiedeman, 1954) [18]. Also in the field of political scientists this approach was discovered useful in investigating the behavior of the citizen voting, it is greatly used in the domain of pattern recognition that is interested specifically with images, the purpose of the pattern recognition is to make the processes carried out by humans is automated (John Wiley and Sans, Hoboken), some interested domains in that this approach has been implemented in the personal placement testing, roll call analysis of legislatures, psychological testing of children, the effects of medical treatments, predicting voting behavior and any more domains. The most important things in this method are two or more groups exist on different variables, discriminate analysis also help in analyzing the differences between groups.

Concept of Discriminant Analysis

Discriminant or discriminant function analysis is a parametric technique to determine which weightings of quantitative variables or predictors best discriminate between 2 or more than 2 groups of cases and do so better than chance (Cramer, 2003) [19].

The analysis creates a discriminant function which is a linear combination of the weightings and scores on these variables. The maximum number of functions is either the number of predictors or the number of groups minus one, whichever of these two values is the smaller.

$$Zjk = a + W_1X_{1k} + W_2X_{2k} +. + W_nX_{nk}$$

Where:
Z_{jk} = Discriminant Z score of discriminant function *j* for object *k.*
a = Intercept.
W_i = Discriminant coefficient for the Independent variable *i.*
X_j = Independent variable *i* for object *k.*
Again, caution must be taken to be clear that sometimes the focus of the analysis is not to predict but to explain the relationship, as such, equations are not normally written when the measures used are not objective measurements.

Types of discriminant analysis
Linear Discriminant Analysis (LDA)

- **Definition:** LDA is the most commonly used form of discriminant analysis, suitable when the predictor variables are continuous and follow a normal distribution. It assumes that each group has identical variance-covariance matrices.
- **Purpose:** LDA maximizes the separation between two or more groups by finding linear combinations of predictor variables that best distinguish them.
- **Application:** LDA is often used in image recognition, for instance, to distinguish between images of faces or objects.

Quadratic Discriminant Analysis (QDA)

- **Definition:** QDA is similar to LDA but allows for distinct variance-covariance matrices for each group, making it more flexible in cases where groups do not have identical covariances.
- **Purpose:** QDA provides better classification accuracy when groups differ in variance-covariance structures.
- **Application:** QDA is suitable for complex classification tasks where groups have distinct distributions, such as differentiating between types of tumors in medical diagnostics.

Canonical Discriminant Analysis (CDA)

- **Definition:** CDA, also known as canonical correlation analysis, finds the linear combinations (canonical variables) that best separate multiple groups. CDA is used to analyse relationships between multiple predictor and response variables.
- **Purpose:** CDA is particularly useful in exploratory data analysis and for reducing the dimensionality of data.
- **Application:** CDA can be applied in educational research to determine which factors best separate students by academic performance.

Regularized Discriminant Analysis (RDA)

- **Definition:** RDA is a modified version of LDA and QDA that introduces regularization to reduce over fitting. It combines features of both LDA and QDA, balancing the two approaches based on the dataset.
- **Purpose:** RDA is valuable when the dataset has many predictor variables and a small sample size, making it a robust choice for high-dimensional data.
- **Application:** RDA is widely used in genomic studies to classify gene expressions, where datasets often have a large number of variables.

Flexible Discriminant Analysis (FDA)

- **Definition:** FDA is a non-linear extension of discriminant analysis, using non-parametric regression techniques to classify observations into groups.
- **Purpose:** FDA is effective for datasets with complex, non-linear relationships that linear methods cannot capture.
- **Application:** FDA is applicable in fields like ecology, where relationships between species and environmental factors may be non-linear.

Reasons why discriminant analysis is better than logistics regression

Discriminant function analysis is very similar to logistic regression, and both can be used to answer the same research questions. Logistic regression does not have as many assumptions and restrictions as discriminant analysis. However, when discriminant analysis' assumptions are met, it is more powerful than logistic regression. Unlike logistic regression, discriminant analysis can be used with small sample sizes. It has been shown that when sample sizes are equal, and homogeneity of variance/covariance holds, discriminant analysis is more accurate.

Methods in discriminant analysis
Estimating Group Means and Covariances

In discriminant analysis, the mean vector and covariance matrix for each group are estimated from the data. LDA assumes equal covariances across groups, while QDA allows for distinct covariances.

Discriminant Functions

- Discriminant functions are linear or quadratic equations derived from predictor variables. Each function is optimized to maximize the distance between group means and minimize the variance within groups.
- For example, in LDA, a single linear discriminant function is created for two groups, while multiple functions can be generated for multi-group classification.

Eigenvalues and Canonical Variables

In canonical discriminant analysis, eigenvalues represent the variance explained by each canonical variable, indicating its contribution to separating the groups. Canonical variables are derived from the linear combination of predictor variables, offering insights into the relationships between variables.

Classification Rule and Decision Boundary

A classification rule is established to assign observations to groups based on their discriminant scores. Decision boundaries are then drawn to indicate which region of the data space each observation falls into. In LDA, decision boundaries are linear, while QDA creates quadratic boundaries.

Cross-Validation and Performance Evaluation

To ensure accuracy, discriminant analysis often employs cross-validation techniques, such as k-fold cross-validation, to evaluate classification performance. Metrics such as classification accuracy, sensitivity, and specificity are used to assess how well the model predicts group membership.

Evaluation criteria for discriminant analysis

When results of a discriminant analysis are obtained, there are three basic questions to ask:

a) Which independent variables are good discriminators?
b) How well do these independent variables discriminate among the two groups?
c) What decision rule should be used for classifying individuals? More complete answers to these questions require a synopsis of the theoretical derivation of the discriminant function. The other steps to look for are;

1. Deriving the Discriminant Function, and
2. Determining the Effect of Independent Variables

Advantages and limitations

Advantages:

- **Predictive Accuracy:** Discriminant analysis is highly effective for classification and prediction tasks.
- **Interpretability:** The discriminant functions provide interpretable models, showing which variables contribute most to group separation.
- **Versatility:** Discriminant analysis supports linear, quadratic, and non-linear relationships, allowing flexibility across different datasets.

Limitations

- **Assumptions of Normality:** LDA assumes normal distribution of predictor variables, which may not hold in all datasets.
- **Sensitivity to Outliers:** Outliers can significantly impact the model's accuracy, particularly in LDA.
- **Complexity with High-Dimensional Data:** Large numbers of predictor variables can complicate the model, potentially leading to overfitting in QDA or LDA without regularization.

Applications of discriminant analysis

Marketing and Customer Segmentation

- **Objective:** A company wants to classify customers based on purchasing behavior to design targeted marketing strategies.
- **Method:** Using LDA, the company analyzes variables like purchase frequency, total spending, and product preferences to group customers into categories (e.g., high-value, moderate-value, low-value).
- **Outcome:** By distinguishing customer groups, the company can create tailored marketing campaigns for each segment.

Medical Diagnostics

- **Objective:** A healthcare provider seeks to classify patients based on medical test results to predict disease outcomes.
- **Method:** QDA is used to account for different variances in patient groups, allowing for more accurate classification between disease types.
- **Outcome:** Discriminant analysis helps predict patient diagnoses, improving the accuracy of early detection and treatment strategies.

Financial Risk Assessment

- **Objective:** A bank wants to classify loan applicants based on the likelihood of default.
- **Method:** LDA is applied to historical data, including income, credit score, and debt-to-income ratio, to classify applicants into low-risk and high-risk categories.
- **Outcome:** By classifying borrowers, the bank can better manage loan approval processes and mitigate financial risk.

Ecology and Species Classification

- **Objective:** Researchers aim to classify different plant species based on environmental variables like soil type, temperature, and rainfall.
- **Method:** Flexible Discriminant Analysis (FDA) captures non-linear relationships between species and environmental factors.
- **Outcome:** The model helps ecologists understand species distribution, informing conservation efforts and environmental management.

Educational Research

- **Objective:** A school district seeks to classify students by performance levels based on academic, social, and behavioral metrics.
- **Method:** Canonical Discriminant Analysis (CDA) helps identify the factors that best distinguish high-performing and low-performing students.
- **Outcome:** This information can guide targeted interventions for students needing additional support, improving educational outcomes.

Conclusion

Discriminant analysis is a statistical technique which allows the researchers to study the differences between two or more groups of objects with respect to several variables simultaneously. The applications discussed showed that discriminant analysis can be effectively used to classify observations into two or more known set of data depending on the existence of one or more quantitative variables.

Discriminant analysis is a valuable tool for classifying observations into groups based on predictor variables. Through techniques like LDA, QDA, CDA, and FDA, discriminant analysis enables researchers and analysts to distinguish groups, predict outcomes, and identify variables that drive differences. Widely used in fields such as finance, healthcare, and ecology, discriminant analysis continues to provide insightful solutions for classification and prediction tasks.

References

1. Alexakos CE. Predictive efficiency of two multivariate statistical techniques in comparison with clinical predictions. J Educ Psychol. 1966;57:297-306.
2. Anderson GJ, Walberg HJ, Welch WW. Curriculum effects on the social climate of learning: A new representation of discriminant functions. Am Educ Res J. 1969;6:315-328.
3. Afifi AA, Sacks ST, Liu VY, Weil MH, Shubin H. Accumulative prognostic index for patients with barbiturate, glutethimide and meprobamate intoxication. N Engl J Med. 1971;285:1497-1502. DOI:10.1056/NEJM197112302852702
4. Albert A. Discriminant analysis based on multivariate response curves: A descriptive approach to dynamic allocation. Stat Med. 1983;2:95-106. DOI:10.1002/sim.4780020111
5. Albert JM, Kshirsagar AM. The reduced-rank growth

curve model for discriminant analysis of longitudinal data. Aust J Stat. 1993;35:345-357. DOI:10.1111/j.1467-842X.1993.tb01342.x
6. Azen SP, Afifi AA. Two models of assessing prognosis on the basis of successive observations. Math Biosci. 1972;14:169-176. DOI:10.1016/0025-5564(72)90016-8
7. Bagui SC, Mehra KL. Classification of multiple observations using multi-stage nearest rank neighbour rule. J Stat Plan Inference. 1999;76:163-183. DOI:10.1016/S0378-3758(98)00137-2
8. Chastian K. Prediction of success in audio-lingual and cognitive classes. Lang Learn. 1969;19:27-39.
9. Cochran WG. On the performance of the linear discriminant function. Technometrics. 1974;6:179-190.
10. Eisenbeis RA, Avery RB. Discriminant Analysis and Classification Procedures: Theory and Applications. Lexington, Mass.: D.C. Heath & Co.; 1972.
11. Gnanadesikan R. Methods for Statistical Data Analysis of Multivariate Observations. London: Wiley; 1977.
12. Joy OM, Tollefson JO. On the financial applications of discriminant analysis. J Financ Quant Anal. 1975;10:723-739.
13. Lachenbruch PA. An almost unbiased method of obtaining confidence intervals for the probability of misclassification in discriminant analysis. Biometrics. 1967;23:639-645.
14. Malkovich JF, Afifi AA. On tests for multivariate normality. J Am Stat Assoc. 1973;68:176-179.
15. Marriott FHC. The Interpretation of Multiple Observations. London: Academic Press; 1974.
16. Mosteller F, Wallace DL. Inference in the authorship problem. J Am Stat Assoc. 1963;58:275-309.
17. Saupe JL. Factorial-design multiple discriminant analysis: A description and an illustration. Am Educ Res J. 1965;2:175-184.
18. Tatsuoka MM, Tiedeman DV. Chapter iv: Discriminant analysis. Review of Educational Research. 1954 Dec;24(5):402-420.
19. Cramer D. Advanced quantitative data analysis. McGraw-Hill Education (UK); 2003 Jul 1.