www.statsoft.com

- Products
- Solutions
- Buy
- Trials
- Support

What can we help you find? Search

TextbookGeneral Discriminant Analysis

Looking for info about statistics?
We wrote the book on it.
And you can read it for free!

Elementary Concepts
Statistics Glossary
Basic Statistics
ANOVA / MANOVA
Association Rules
Boosting Trees
Canonical Analysis
CHAID Analysis
C & R Trees
Classification Trees
Cluster Analysis
Correspondence Analysis
Data Mining Techniques
Discriminant Analysis
Distribution Fitting
Experimental Design
Factor Analysis
General Discrim. Analysis
General Linear Models
Generalized Additive Mod.
Generalized Linear Mod.
General Regression Mod.
Graphical Techniques
Ind.Components Analysis
Linear Regression
Log-Linear Analysis
MARSplines
Machine Learning
Multidimensional Scaling
Neural Networks
Nonlinear Estimation
Nonparametric Statistics
Partial Least Squares
Power Analysis
Process Analysis
Quality Control Charts
Reliability / Item Analysis
SEPATH (Structural eq.)
Survival Analysis
Text Mining
Time Series / Forecasting

General Discriminant Analysis (GDA

Introductory Overview

General Discriminant Analysis (GDA) is called a "general" discriminant analysis because it applies the methods of the general linear model (see also General Linear Models (GLM)) to the discriminant function analysis problem. A general overview of discriminant function analysis, and the traditional methods for fitting linear models with categorical dependent variables and continuous predictors, is provided in the context of Discriminant Analysis. In *GDA*, the discriminant function analysis problem is "recast" as a general multivariate linear model, where the dependent variables of interest are (dummy-) coded vectors that reflect the group membership of each case. The remainder of the analysis is then performed as described in the context of General Regression Models (GRM), with a few additional features noted below.

To index

Advantages of GDA

Specifying models for predictor variables and predictor effects. One advantage of applying the general linear model to the discriminant analysis problem is that you can specify complex models for the set of predictor variables. For example, you can specify for a set of continuous predictor variables, a polynomial regression model, response surface model, factorial regression, or mixture surface regression (without an intercept). Thus, you could analyze a constrained mixture experiment (where the predictor variable values must sum to a constant), where the dependent variable of interest is categorical in nature. In fact, GDA does not impose any particular restrictions on the type of predictor variable (categorical or continuous) that can be used, or the models that can be specified. However, when using categorical predictor variables, caution should be used (see "A note of caution for models with categorical predictors, and other advanced techniques" below).

Stepwise and best-subset analyses. In addition to the traditional stepwise analyses for single continuous predictors provided in Discriminant Analysis, *General Discriminant Analysis* makes

Variance Components
Statistical Advisor
Distribution Tables
References Cited
Send Comments
Business Solutions
Free Resources
About Textbook

available the options for stepwise and best-subset analyses provided in General Regression Models (GRM). Specifically, you can request stepwise and best-subset selection of predictors or sets of predictors (in multiple-degree of freedom effects, involving categorical predictors), based on the *F-to-enter* and *p-to-enter* statistics (associated with the multivariate Wilks' *Lambda* test statistic). In addition, when a cross-validation sample is specified, best-subset selection can also be based on the misclassification rates for the cross-validation sample; in other words, after estimating the discriminant functions for a given set of predictors, the misclassification rates for the cross-validation sample are computed, and the model (subset of predictors) that yields the lowest misclassification rate for the cross-validation sample is chosen. This is a powerful technique for choosing models that may yield good predictive validity, while avoiding overfitting of the data (see also *Neural Networks*).

Desirability profiling of posterior classification probabilities. Another unique option of *General Discriminant Analysis (GDA)* is the inclusion of Response/desirability profiler options. These options are described in some detail in the context of Experimental Design (DOE). In short, the predicted response values for each dependent variable are computed, and those values can be combined into a single desirability score. A graphical summary can then be produced to show the "behavior" of the predicted responses and the desirability score over the ranges of values for the predictor variables. In *GDA*, you can profile both simple predicted values (like in *General Regression Models*) for the coded dependent variables (i.e., dummy-coded categories of the categorical dependent variable), and you can also profile posterior prediction probabilities. This unique latter option allows you to evaluate how different values for the predictor variables affect the predicted classification of cases, and is particularly useful when interpreting the results for complex models that involve categorical and continuous predictors and their interactions.

A note of caution for models with categorical predictors, and other advanced techniques. General Discriminant Analysis provides functionality that makes this technique a general tool for classification and data mining. However, most -- if not all -- textbook treatments of discriminant function analysis are limited to simple and stepwise analyses with single degree of freedom continuous predictors. No "experience" (in the literature) exists regarding issues of robustness and effectiveness of these techniques, when they are generalized in the manner provided in this very powerful analysis. The use of best-subset methods, in particular when used in conjunction with categorical predictors or when using the misclassification rates in a cross-validation sample for choosing the best subset of predictors, should be considered a heuristic search method, rather than a statistical analysis technique.

The use of categorical predictor variables. The use of categorical predictor variables or effects in a discriminant function analysis model may be (statistically) questionable. For example, you can use *GDA* to analyze a 2 by 2 frequency table, by specifying one variable in the 2 by 2 table as the dependent variable, and the other as the predictor. Clearly, the (ab)use of *GDA* in this manner would be silly (although, interestingly, in most cases you will get results that are generally compatible with those you would get by computing a simple Chi*). On the other hand, if you only consider the parameter estimates computed by GDA as the least squares solution to a set of linear (prediction) equations, then the use of categorical predictors in GDA is fully justified; moreover, it is not uncommon in applied research to be confronted with a mixture of continuous and categorical predictors (e.g., income or age which are continuous, along with occupational status, which is categorical) for predicting a categorical dependent variable. In those cases, it can be very instructive to consider specific models involving the categorical predictors, and possibly interactions between categorical and continuous predictors for classifying observations. However, to reiterate, the use of categorical predictor variables in discriminant function analysis is not widely documented, and you should proceed cautiously before accepting the results of statistical significance tests, and before drawing final* *conclusions from your analyses. Also remember that there are alternative methods available to perform similar analyses, namely, the* multinomial logit *models available in* Generalized Linear Models (GLZ)*, and the methods for analyzing multi-way frequency tables in* Log-Linear*. -square test for the 2 by 2 table*

To index

Copyright © 2018 TIBCO Software Inc All Rights Reserved. | Privacy Policy