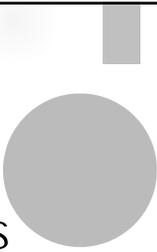


Discriminant Function Analysis

LECTURE 05



Objectives

At the end of this series of lectures, you should be able to:

- ▶ Define terms.
- ▶ Describe DFA.
- ▶ Describe the use and limitations of DFA.
- ▶ Explain the assumptions of DFA.
- ▶ Interpret the results of DFA.
- ▶ Perform DFA using R



Review

- ▶ Logistic regression very similar in goals to a DFA.
 - ▶ Difference is the type of data accepted as independent variables.
 - ▶ Assumptions of the logistic regression are easier met than for DFA.
 - ▶ Power is somewhat better for a DFA than logistic regression.
 - ▶ Sample sizes required for logistic regression are much larger.
- ▶ DFA is a classification technique and should only be casually used for prediction.
- ▶ Logistic regression is a predictive technique and should only be used casually as a classification technique.



Review

- ▶ There are similarities multiple regression and DFA.
 - ▶ Use ratio/interval values to predict a "dependent variable."
 - ▶ Variate form
 - ▶ "Dependent variable" ratio/interval in multiple regression, nominal in DFA.
 - ▶ Multiple regression is a predictive technique and DFA is a classification technique.

Overview

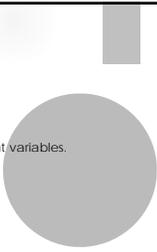
- ▶ Intent -- Statistical
 - ▶ Find a set of axes that maximize the amount of variation among the groups relative to the amount of variation within the groups. (Sounds like cluster analysis)
 - ▶ Like PCA, DFA reduces the number of dimensions you need to consider.
 - ▶ Like PCA, DFA is an eigenvector procedure

Overview

- ▶ Intent -- Uses
 - ▶ Different than noted in the book
 - ▶ Classify samples into existing groups or classes
 - ▶ Establish the importance of independent variables in the classification process.
 - ▶ (Use the results to predict the group membership of an individual that was not included in the initial analysis.)

Overview

- ▶ Data requirements
 - ▶ There are, strictly speaking, no dependent or independent variables. However:
 - ▶ Independent variable (Predictor) – Ratio/Interval
 - ▶ Dependent variable (Groups) – Nominal



Overview

- ▶ Data requirement
 - ▶ The number of variables should be at least 2 less than the number of cases included in the analyses.
 - ▶ Group sizes should be about equal.
 - ▶ At least two real groups in the dependent variable.



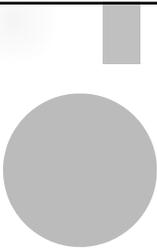
Overview

- ▶ Rationale
 - ▶ Discriminant analysis refers to a variety of techniques.
 - ▶ The most common result of these analyses is a linear equation that can be used to distinguish between the sample groups.
 - ▶ The technique we are going to use, is often referred to as canonical variates analysis or discriminant function analysis.
 - ▶ An exploratory technique, some variations are used in hypothesis testing but we will concentrate on an exploratory variation.



Overview

- ▶ Rationale
 - ▶ Demonstrated in class



Assumptions

- ▶ Random sampling
- ▶ Independent observations
- ▶ Equal variance within groups for each predictor variable (Homoscedasticity).
- ▶ Correlations between variables are the same for all groups.
- ▶ Linearity
- ▶ Multivariate normality (for hypothesis testing)
- ▶ (No multicollinearity)



Issues

- ▶ Issues
 - ▶ Sensitive to outliers
 - ▶ Determination of actual values (circular reasoning)