



Л е к ц и я № 6

Дискриминантный анализ

Разработал профессор П. М. Васильев
Кафедра фармакологии и биоинформатики

Для студентов, обучающихся по направлению 06.03.01 «Биология»
профили Биохимия, Генетика
при изучении дисциплины «Цифровые технологии в биологии»

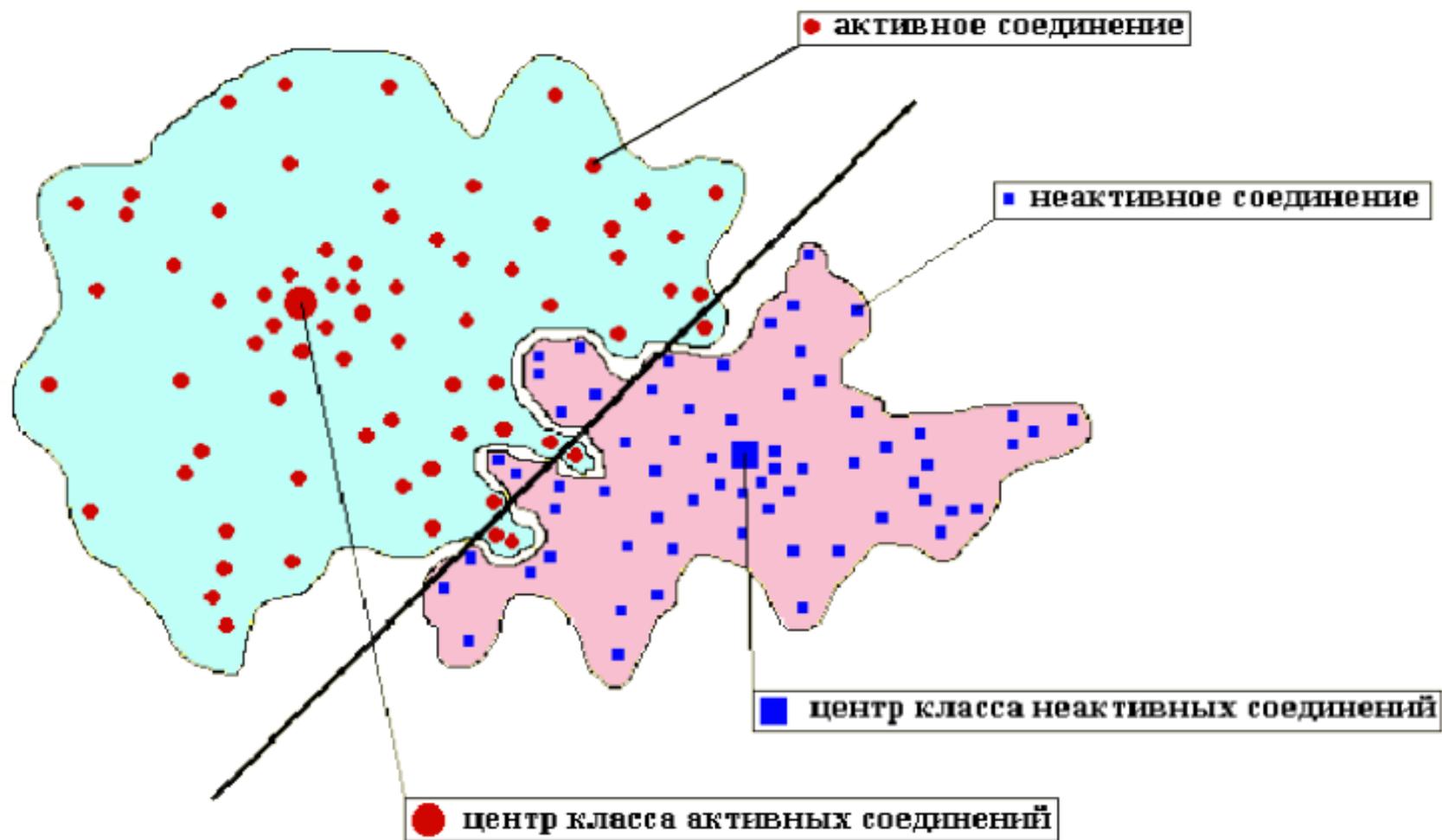
П л а н л е к ц и и

- Теория распознавания образов
- Что такое дискриминантный анализ
- Необходимые условия
- Линейный дискриминантный анализ
- Дискриминантные функции
- Пошаговый дискриминантный анализ

Теория распознавание образов

Разрабатывает принципы построения систем, предназначенных для определения принадлежности данного объекта к одному из заранее выделенных классов объектов.

Задача классификации



Что такое дискриминантный анализ

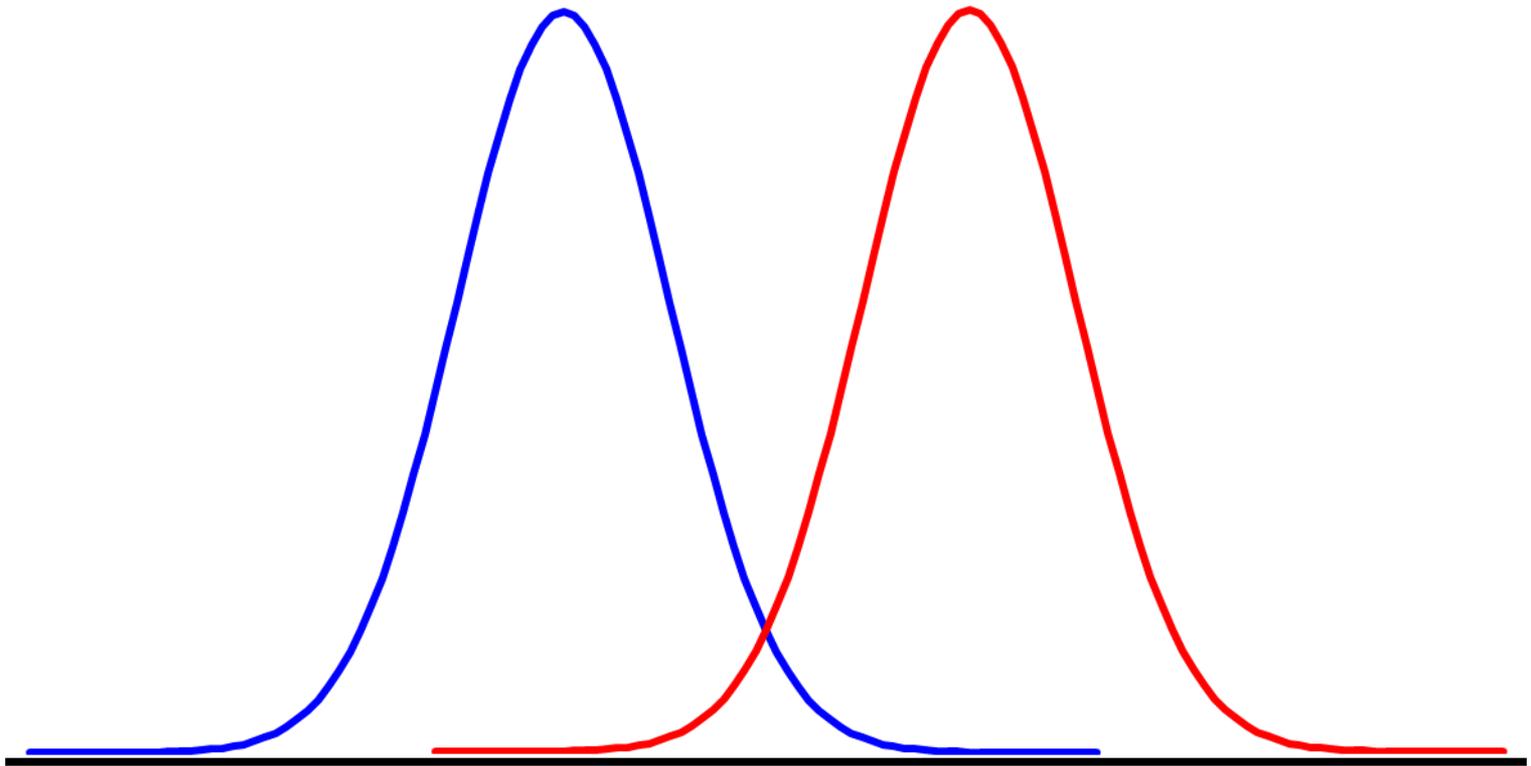
Метод многомерного статистического анализа для решения задач распознавания образов при обучении с учителем, который позволяет предсказывать принадлежность объектов к нескольким группам.

Необходимые условия

- Зависимая переменная должна быть дискретной (бинарной)
- Независимые переменные должны быть непрерывными (соблюдается не всегда)

Вероятностная трактовка

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$



Дискриминантные функции

$$g_1 = a_{01} + a_{11}x_1 + a_{21}x_2 + \dots + a_{m1}x_m$$

$$g_2 = a_{02} + a_{21}x_1 + a_{22}x_2 + \dots + a_{2m}x_m$$

**Объект C принадлежит классу 1,
если $g_1 \geq g_2$**

Линейный дискриминантный анализ

Пример исходных данных

Code	Y3	Y2	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13
LT-47	high	h	-8.5	-6.5	-7.9	-7.1	-8.2	-8.7	-8.3	-6.2	-6.7	-7.5	-8.5	-6.4	-6.3
LT-51	high	h	-8.4	-6.7	-7.4	-6.6	-8.1	-8.6	-8.3	-6.1	-6.5	-7.4	-8.4	-6.8	-6.1
LT-52	high	h	-8.6	-7.1	-8.2	-7.0	-9.4	-9.2	-8.6	-6.4	-6.8	-8.2	-8.6	-7.5	-6.4
L-91	high	h	-8.6	-6.9	-6.9	-8.3	-9.2	-8.7	-8.9	-6.1	-6.4	-7.5	-8.2	-7.0	-6.3
LT-13	high	h	-8.6	-7.2	-8.2	-7.2	-9.0	-8.9	-8.7	-6.3	-6.5	-7.5	-8.2	-6.4	-5.8
LT-23	high	h	-8.5	-6.8	-7.0	-7.7	-8.8	-8.2	-8.4	-6.1	-6.4	-7.6	-8.1	-7.0	-6.0
LT-48	high	h	-8.5	-6.4	-7.4	-6.8	-8.4	-8.6	-8.2	-6.2	-6.6	-7.6	-8.0	-6.6	-6.1
LT-46	high	h	-8.2	-6.2	-7.8	-7.8	-7.9	-8.3	-7.9	-6.0	-6.4	-7.4	-8.2	-6.2	-6.0
H-16	high	h	-8.3	-6.5	-7.1	-6.6	-8.0	-8.1	-8.1	-5.7	-6.1	-7.3	-8.0	-6.6	-6.4
H-88	high	h	-8.6	-7.0	-7.3	-8.0	-9.3	-8.4	-8.0	-6.1	-6.4	-7.2	-7.9	-6.8	-6.3
LT-22	high	h	-8.5	-6.9	-7.4	-8.3	-9.0	-8.3	-8.4	-6.0	-6.3	-7.5	-8.2	-6.8	-6.5
LT-49	high	h	-9.0	-6.6	-7.0	-6.8	-8.4	-8.5	-8.4	-6.0	-6.3	-7.6	-8.0	-6.3	-6.3
L-87	moderate	nh	-8.5	-7.0	-7.5	-7.8	-8.6	-8.5	-8.9	-6.2	-6.4	-7.3	-8.2	-6.9	-6.3
LT-55	moderate	nh	-8.6	-7.0	-7.5	-7.3	-8.5	-8.3	-9.1	-6.1	-6.4	-7.8	-8.1	-6.7	-6.0
LT-9	moderate	nh	-8.6	-6.8	-7.4	-8.0	-8.4	-8.2	-8.4	-6.1	-6.4	-6.9	-7.7	-6.3	-6.0
L-20	moderate	nh	-8.4	-6.6	-7.1	-7.3	-8.2	-8.3	-8.1	-6.0	-6.3	-7.4	-8.1	-6.6	-6.1
L-31	moderate	nh	-8.4	-6.9	-7.4	-7.8	-8.0	-8.4	-8.1	-6.2	-6.6	-7.4	-8.0	-6.4	-6.6
L-86	moderate	nh	-8.3	-6.8	-6.8	-8.5	-8.9	-8.2	-8.6	-6.1	-6.2	-7.2	-8.0	-6.9	-6.2
LT-21	moderate	nh	-8.6	-6.8	-7.1	-8.2	-8.8	-8.2	-8.1	-6.0	-6.2	-7.3	-8.0	-6.9	-6.1
LTG-2	moderate	nh	-7.8	-6.8	-6.9	-7.3	-7.9	-8.0	-7.4	-6.2	-6.5	-8.1	-8.4	-6.9	-6.8
L-19	moderate	nh	-8.3	-6.6	-7.1	-6.7	-8.3	-8.1	-7.9	-6.2	-6.4	-7.6	-7.9	-6.6	-6.1
LT-24	moderate	nh	-8.8	-7.1	-6.8	-7.9	-8.8	-8.5	-8.2	-5.9	-6.3	-7.4	-7.9	-6.5	-6.4
LT-33	moderate	nh	-8.0	-6.1	-7.2	-7.0	-7.9	-7.9	-7.7	-5.8	-5.9	-7.1	-8.0	-6.1	-5.5
L-17	moderate	nh	-8.0	-6.6	-7.4	-8.3	-8.1	-7.9	-7.8	-6.1	-6.3	-7.4	-8.0	-6.5	-6.0
LT-34	moderate	nh	-8.2	-6.3	-7.6	-6.5	-8.3	-8.1	-7.9	-6.0	-6.3	-7.1	-8.1	-6.3	-6.1
LT-36	moderate	nh	-8.2	-6.3	-7.1	-6.2	-7.8	-8.1	-7.7	-5.8	-6.0	-7.0	-8.1	-6.7	-5.9
LT-50	moderate	nh	-8.7	-6.6	-7.4	-6.8	-8.3	-8.5	-7.9	-6.2	-6.4	-7.5	-7.9	-5.9	-6.4
LT-5	moderate	nh	-8.5	-6.5	-7.1	-8.0	-8.1	-7.9	-8.0	-5.8	-6.2	-7.0	-7.5	-6.1	-5.6
H-18	moderate	nh	-8.6	-6.7	-6.8	-6.6	-8.2	-8.0	-8.2	-6.1	-6.3	-7.4	-7.9	-6.3	-6.4
H-69	low	nh	-8.2	-6.5	-7.1	-6.5	-8.0	-8.1	-8.1	-5.9	-6.2	-7.4	-7.8	-6.3	-6.4
LT-30	low	nh	-7.6	-6.1	-6.5	-7.2	-7.8	-7.2	-7.5	-5.6	-6.0	-7.0	-7.5	-6.0	-6.1
LT-31	low	nh	-7.8	-6.1	-6.1	-5.9	-7.8	-7.2	-7.3	-5.6	-5.9	-7.1	-7.6	-6.1	-5.7
LT-35	low	nh	-8.1	-6.2	-7.0	-6.5	-8.5	-8.2	-8.0	-6.0	-6.3	-7.0	-7.8	-6.6	-5.9
LT-39	low	nh	-8.2	-6.2	-6.7	-6.3	-8.4	-8.1	-8.2	-5.7	-6.1	-7.3	-7.6	-6.2	-5.8
LT-29	low	nh	-7.3	-5.9	-6.3	-6.8	-7.7	-6.8	-7.2	-5.6	-5.8	-7.1	-7.2	-6.0	-5.8
LT-32	low	nh	-7.7	-6.2	-6.0	-6.7	-7.6	-7.3	-7.6	-5.3	-5.7	-7.0	-7.4	-6.1	-5.8
LT-38	low	nh	-9.3	-6.4	-7.1	-6.0	-8.2	-8.2	-8.1	-5.9	-6.2	-7.2	-7.6	-6.3	-5.8
LT-37	low	nh	-9.3	-6.3	-7.0	-6.2	-8.3	-8.1	-8.1	-5.9	-6.2	-7.1	-7.7	-6.5	-5.7
H-42	low	nh	-7.4	-6.0	-6.2	-6.4	-7.5	-7.3	-7.0	-5.4	-5.6	-7.0	-7.3	-6.3	-5.8
LT-53	low	nh	-7.5	-6.1	-6.3	-6.8	-7.4	-7.4	-7.2	-5.6	-5.9	-7.0	-7.4	-6.3	-5.9
H-29	low	nh	-7.1	-5.5	-6.2	-7.6	-7.1	-7.1	-6.7	-5.3	-5.6	-6.6	-6.8	-5.9	-5.5

Линейный дискриминантный анализ

STATISTICA - Docking TDAz_L17

File Edit View Insert Format Statistics Data Mining Graphs Tools Data Window Help

Resume... Ctrl+R Add to MS Word

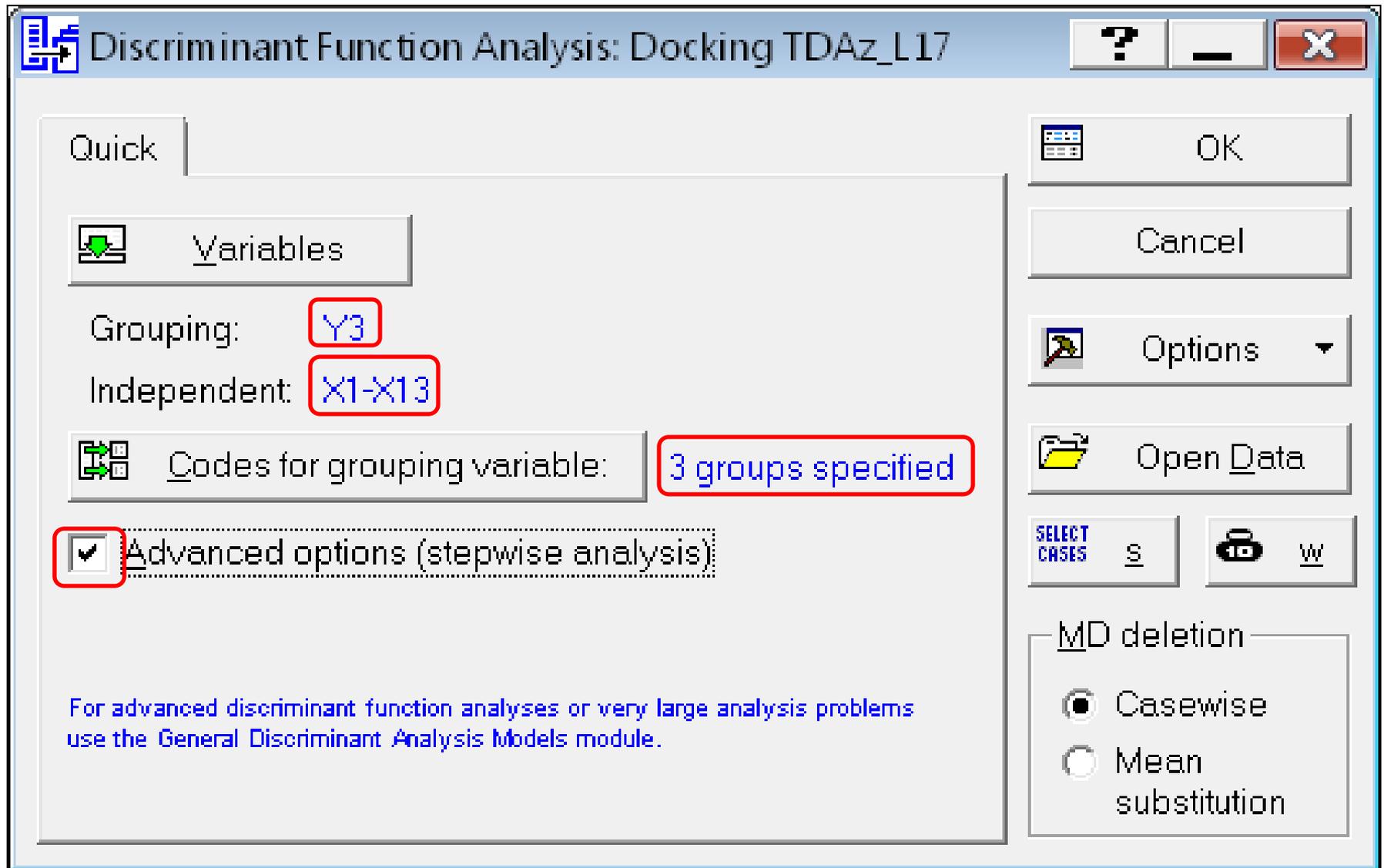
Basic Statistics/Tables
 Multiple Regression
 ANOVA
 Nonparametrics
 Distribution Fitting
 Advanced Linear/Nonlinear Models
Multivariate Exploratory Techniques
 Industrial Statistics & Six Sigma
 Power Analysis
 Automated Neural Networks
 PLS, PCA, Multivariate/Batch SPC
 Variance Estimation and Precision (VEPAC)
 Statistics of Block Data
 STATISTICA Visual Basic
 Batch (ByGroup) Analysis
 Probability Calculator

Cluster Analysis
 Factor Analysis
 Principal Components & Classification Analysis
 Canonical Analysis
 Reliability/Item Analysis
 Classification Trees
 Correspondence Analysis
 Multidimensional Scaling
Discriminant Analysis
 General Discriminant Analysis Models

Data: Docking TDAz_L17* (15)

	LDA																
	1																
	Y3																
LT-47	high																
LT-51	high																
LT-52	high																
L-91	high																
LT-13	high																
LT-23	high																
LT-48	high																
LT-46	high																
H-16	high																
H-88	high																
LT-22	high																
LT-49	high																
L-87	moderate	nh	-8.5	-7.0	-7.5	-7.8	-8.6	-8.5	-8.9	-6.2	-6.4	-7.3	-8.2	-6.9	-6.3		
LT-55	moderate	nh	-8.6	-7.0	-7.5	-7.3	-8.5	-8.3	-9.1	-6.0	-6.4	-7.8	-8.1	-6.7	-6.0		
LT-9	moderate	nh	-8.6	-6.8	-7.4	-8.0	-8.4	-8.2	-8.4	-6.1	-6.4	-6.9	-7.7	-6.3	-6.0		
L-20	moderate	nh	-8.4	-6.6	-7.1	-7.3	-8.2	-8.3	-8.1	-5.9	-6.3	-7.4	-8.1	-6.6	-6.1		
L-31	moderate	nh	-8.4	-6.9	-7.4	-7.8	-8.0	-8.4	-8.1	-6.2	-6.5	-7.4	-8.0	-6.4	-6.6		
L-86	moderate	nh	-8.3	-6.8	-6.8	-8.5	-8.9	-8.2	-8.6	-6.1	-6.2	-7.2	-8.0	-6.9	-6.2		
LT-21	moderate	nh	-8.6	-6.8	-7.1	-8.2	-8.8	-8.2	-8.1	-6.0	-6.2	-7.3	-8.0	-6.9	-6.1		
LTG-2	moderate	nh	-7.8	-6.8	-6.9	-7.3	-7.9	-8.0	-7.4	-6.2	-6.5	-8.1	-8.4	-6.9	-6.8		
L-19	moderate	nh	-8.3	-6.6	-7.1	-6.7	-8.3	-8.1	-7.9	-6.2	-6.4	-7.6	-7.9	-6.6	-6.1		
LT-24	moderate	nh	-8.8	-7.1	-6.8	-7.9	-8.8	-8.5	-8.2	-5.9	-6.3	-7.4	-7.9	-6.5	-6.4		
LT-33	moderate	nh	-8.0	-6.1	-7.2	-7.0	-7.9	-7.9	-7.7	-5.8	-5.9	-7.1	-8.0	-6.1	-5.5		

Линейный дискриминантный анализ



Линейный дискриминантный анализ

Model Definition: Docking TDAz_L17

Variables: X1-X13

Quick **Advanced** Descriptives

Method: Forward stepwise

Tolerance: .010

Stepwise options:

E to enter: 1.00

F to remove: 0.00

Number of steps: 13

Display results: Summary only

OK
Cancel
Options

Линейный дискриминантный анализ

The screenshot shows the 'Discriminant Function Analysis Results: Docking TDAz_L17' dialog box in SPSS. The main text area displays the results of a stepwise analysis:

```
Stepwise Analysis - Step 8 (Final Step)  
Number of variables in the model: 8  
Last variable entered: X2          F (2,31) = .8441100 p < .4396  
Wilks' Lambda: .1308371 approx. F (16,62) = 6.837880 p < .0000
```

The 'Wilks' Lambda' line is highlighted with a red box. Below the text area, there are three tabs: 'Quick', 'Advanced', and 'Classification'. The 'Quick' tab is active, showing a 'Summary' section with two buttons: 'Variables in the model' (highlighted with a red box) and 'Variables not in the model'. On the right side of the dialog, there are buttons for 'Summary', 'Cancel', 'Options', and 'By Group'.

Линейный дискриминантный анализ

Workbook1* - Discriminant Function Analysis Summary (Docking TDAz_L17)

Discriminant Function Analysis Summary (Docking TDAz_L17)
Step 8, N of vars in model: 8; Grouping: Y3 (3 grps)
Wilks' Lambda: .13084 approx. F (16,62)=6.8379 p< .0000

N=41	Wilks' Lambda	Partial Lambda	F-remove (2,31)	p-level	Toler.	1-Toler. (R-Sqr.)
X11	0.174515	0.749720	5.174378	0.011506	0.388729	0.611271
X4	0.187292	0.698573	6.688083	0.003848	0.658850	0.341150
X12	0.137398	0.952252	0.777211	0.468437	0.502204	0.497796
X6	0.135811	0.963377	0.589238	0.560843	0.295146	0.704854
X8	0.181885	0.719341	6.047507	0.006060	0.122478	0.877522
X9	0.164291	0.796372	3.963258	0.029329	0.124629	0.875371
X5	0.156260	0.837302	3.011847	0.063778	0.295545	0.704455
X2	0.151834	0.861713	2.487429	0.099568	0.378543	0.621457

Discriminant Function Analysis Summary (Docking TDAz_L17)

Линейный дискриминантный анализ

Discriminant Function Analysis Results: Docking TDAz_L17

Stepwise Analysis - Step 8 (Final Step)

Number of variables in the model: 8
Last variable entered: X2 $F(2,31) = .8441100$ $p < .4396$
Wilks' Lambda: .1308371 approx. $F(16,62) = 6.837880$ $p < .0000$

Quick | Advanced | **Classification**

Classification functions
Use selection conditions to classify selected cases only **SELECT CASES** Select

Classification matrix

Classification of cases

Squared Mahalanobis distances

Posterior probabilities

Save scores

A priori classification probabilities

- Proportional to group sizes
- Same for all groups
- User defined

Score to save for each case

- Save classification for case
- Save distance for case
- Save posterior probability for case

Max. number of cases in a single results spreadsheet: 100000

Summary

Cancel

Options

By Group

Линейный дискриминантный анализ

Classification Matrix (Docking TDAz_L17)
Rows: Observed classifications
Columns: Predicted classifications

Group	Percent Correct	high p=.29268	moderate p=.41463	low p=.29268
high	91.66666	11	1	0
moderate	88.23529	2	15	0
low	91.66666	0	1	11
Total	90.24390	13	17	11

Линейный дискриминантный анализ

Model Definition: Docking TDAz_L17

Variables: X1-X13

Quick Advanced Descriptives

Method: Forward stepwise

Tolerance: .010

Stepwise options:

E to enter: 2.00

F to remove: 0.00

Number of steps: 13

Display results: Summary only

OK

Cancel

Options

To be continued ...

