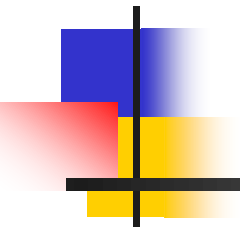

Análise Discriminante

Análise Discriminante

- TÉCNICA DE ANÁLISE MULTIVARIADA QUE ENVOLVE UMA RELAÇÃO DE DEPENDÊNCIA ENTRE VARIÁVEIS, SENDO A VAR. DEPENDENTE CATEGÓRICA E AS INDEPENDENTES EM NÍVEL PELO MENOS INTERVALAR.
- PRODUZ COMBINAÇÕES LINEARES DAS VARIÁVEIS INDEP. (FUNÇÃO DISCRIMINANTE) QUE MELHOR DISCRIMINAM OS GRUPOS ESTABELECIDOS PELA VARIÁVEL DEPENDENTE. ASSIM, SÃO DEFINIDAS AS REGRAS DE CLASSIFICAÇÃO DOS ELEMENTOS EM CADA GRUPO.

Exemplos de Aplicação

- ECONOMIA: FORAM USADAS 6 VAR. FINANC. P/ AJUDAR EMPRESAS DE SEGURO A IDENTIF. DENTRE SEUS CLIENTES A PREDISP. À INSOLVÊNCIA.
- MARKETING: MEDIDAS DE VARIÁVEIS SÓCIO-DEMOGR. E DE ATITUDES FORAM USADAS PARA DEFINIR O PERFIL DE CADA SEGM. DE MERCADO
- CIÊNCIAS SOCIAIS: COM A IDENTIF. DE VARIÁVEIS DISCRIMINADORAS DE BAIRROS AGREGADOS EM GRUPOS INTERN. HOMOG. OBTEVE -SE UM MODELO PARA ORIENTAR AÇÕES PREVENTIVAS EM CADA LOCAL CONFORME A SUA CLASSIFIC. EM ALGUM DOS GRUPOS
- EDUCAÇÃO: MEDIDAS DE 5 VAR. DE AVALIAÇÃO NO 2O GRAU FORAM USADAS COMO PREVISORAS DO DESEMP. DE ALUNOS NA FACULDADE

Exemplos de Aplicação - Perguntas

- ECONOMIA: COMO OS BONS CLIENTES DE SEGURO SE DIFEREM DOS POTENCIALMENTE INSOLVENTES?
- MARKETING:COMO OS SEGMENTOS DE MERCADO DE UMA CATEGORIA DE PRODUTOS DIFEREM ENTRE SI?
- CIÊNCIAS SOCIAIS: COMO OS GPOS DE BAIRROS SE DIFERENCIAM?
- EDUCAÇÃO: COMO OS ALUNOS COM PREVISÃO FAVORÁVEL DE DESEMPENHO NA FACULDADE SE DIFEREM DOS QUE TÊM PREVISÃO DESFAVORÁVEL?

Objetivo

- OBTER UMA FUNÇÃO DISCRIMINATÓRIA POR MEIO DE COMBINAÇÕES LINEARES DAS VARIÁVEIS INDEPEND., A PARTIR DA QUAL SEJA POSSÍVEL CLASSIFICAR OS ELEMENTOS EM CADA UMA DAS CATEGORIAS DA VARIÁVEL DEPENDENTE.

Questões respondidas pela Análise Discriminante

- **QUAIS AS VARIÁVEIS INDEPENDENTES QUE MELHOR DISCRIMINAM OS GRUPOS? (OU SEJA, QUE FAZEM COM QUE A VARIABILIDADE ENTRE OS GRUPOS SEJA MAIOR QUE A VARIABILIDADE DENTRO DOS GRUPOS)**
- **AS MÉDIAS DE CADA VARIÁVEL INDEPENDENTE EM CADA GRUPO SÃO ES TATISTICAMENTE DIFERENTES?**
- **QUAL É O PERFIL DE CADA GRUPO?**
- **QUAL É O GRAU DE EFICIÊNCIA DO MODELO DE CLASSIFICAÇÃO?**

O que deve ser aprendido nesta aula

- **DIFERENCIAR ANÁLISE DISCRIMINANTE DAS OUTRAS TÉCNICAS MULTIVARIADAS**
- **IDENTIFICAR SITUAÇÕES FAVORÁVEIS PARA O SEU USO**
- **IDENTIFICAR CUIDADOS NECESSÁRIOS COM OS DADOS ANTES DOS CÁLCULOS DA TÉCNICA**
- **IDENTIFICAR VARIÁVEIS EFETIVAMENTE DISCRIMINANTES DE GRUPOS PREVIAMENTE DEFINIDOS**
- **IDENTIFICAR PERFIL DE CADA GRUPO**

Etapas da Análise

- SELEÇÃO DAS VAR. DEPEND. E INDEP.
- VERIFICAÇÃO DAS HIPÓTESES DA TÉCNICA
- ESTIMAÇÃO DAS FUNÇÕES DISCRIMINANTES
- ANÁLISE DAS ESTATÍSTICAS
- IDENTIFICAÇÃO DAS VARIÁVEIS INDEP. COM MAIOR PODER DISCRIMINATÓRIO
- ANÁLISE DA MATRIZ DE CLASSIFIC.
- VALIDAÇÃO DOS RESULTADOS

Exemplo

- **OBJETIVO: ANALISAR SE AS VARIÁVEIS IDADE E PONTOS OBTIDOS PARA CLASSE SOCIOECONÔMICA SÃO BOAS DISCRIMINADORAS DE 2 GRUPOS DE PESSOAS: FIÉIS E NÃO FIÉIS À MARCA X DO PRODUTO HIDRATANTE PARA O CORPO**

PARTE DO BANCO DE DADOS

FIDELIDADE		IDADE	CLASSE
FIEL	1	35	22
FIEL	1	40	20
FIEL	1	40	25
FIEL	1	37	28
INFIEL	2	38	28
INFIEL	2	39	30
INFIEL	2	37	33
INFIEL	2	32	30

Exemplo: Questões

- COMO OS CONSUMIDORES FIÉIS À MARCA X SE DISTINGUEM DOS NÃO FIÉIS?
- QUAL A VARIÁVEL INDEPENDENTE QUE MELHOR DISCRIMINA OS 2 GRUPOS?
- AS MÉDIAS DE CADA VARIÁVEL INDEPENDENTE EM CADA GRUPO SÃO ES TATISTICAMENTE DIFERENTES?
- QUAL É O PERFIL DE CADA GRUPO?
- QUAL É O GRAU DE EFICIÊNCIA DO MODELO DE CLASSIFICAÇÃO?
- PREVISÃO: UMA PESSOA COM 35 ANOS E 30 PONTOS SERÁ FIÉL À MARCA X?

Exemplo: Output

FUNÇÃO DISCRIMINANTE:

D = - 0,0441178 * IDADE + 0,3406238 * CLASSE - 7,5534548

MÉDIAS DOS GRUPOS

FIDELIDADE	IDADE	CLASSE	ESCORE
FIEL	38	23,75	- 1,14012
INFIEL	36,5	30,25	1,14012
TOTAL	37,25	27	

ESCORES DISCRIMINANTES

FUNÇÃO DISCRIMINANTE APLICADA A CADA PESSOA

EXEMPLO:

1O CASO:

D = - 0,0441178 * 35 + 0,3406238 * 22 - 7,5534548 = -1,6039

Exemplo: Output

CASOS	GPO REAL	ESCORES	GPO PREVISTO
1	1	-1,6039	1
2	1	-2,5057	1
3	1	-0,8026	1
4	1	0,3517	2
5	2	0,3075	2
6	2	0,9447	2
7	2	2,0548	2
8	2	1,2535	2

RESULTADOS DA CLASSIFICAÇÃO

		GPO PREVISTO	
		1	2
GPO REAL	1	21	7
	2	0	22

% CLASSIFICAÇÕES CORRETAS: 86,0 %

Exemplo: Output

REGRA DE DECISÃO PARA CLASSIF. DOS ELEMENTOS NOS GRUPOS:

VALOR CRÍTICO : EM GERAL , DC É A MÉDIA ENTRE OS ESCORES OBTIDOS NAS MÉDIAS DOS GRUPOS

NESTE EXEMPLO:

DC = (- 1,14012 + 1,14012) / 2 = 0

DECISÃO:

SE ESCORE INDIV. > DC --> GRUPO 2

SE ESCORE INDIV. < DC --> GRUPO 1

OUTRA REGRA : CÁLCULO DE PROBAB.

UMA PESSOA COM 35 ANOS E 30 PONTOS.....

HIPÓTESES DA ANÁLISE DISCRIMINANTE

- **NORMALIDADE DAS VARIÁVEIS INDEPENDENTES**
- **LINEARIDADE DAS RELAÇÕES**
- **SEM PROBLEMAS DE MULTICOLINEARIDADE**
- **VARIÂNCIAS IGUAIS NOS GRUPOS**
- **SEM PROBLEMAS DE OUTLIERS**

ESTATÍSTICAS DA ANÁLISE DISCRIMINANTE

- **CENTRÓIDE**
- **FUNÇÃO DISCRIMINANTE**
- **PESOS E CARGAS DISCRIMINANTES**
- **WILKS'LAMBDA OU ESTATÍSTICA U**
- **CORRELAÇÕES ENTRE AS VARIÁVEIS INDEPENDENTES**

ESTATÍSTICAS DA ANÁLISE DISCRIMINANTE

- **ESTATÍSTICA BOX'S M**
- **EIGENVALUE**
- **CORRELAÇÃO CANÔNICA**
- **COEFIC. DE CLASSIFIC. DE FISHER**
- **MATRIZ DE CLASSIFICAÇÃO**
- **GRÁFICOS**

CENTRÓIDE

- **MÉDIA DE CADA GRUPO OBTIDA A PARTIR DOS ESCORES DISCRIMINANTES DENTRO DE CADA GRUPO**
- **2 GRUPOS : 2 CENTRÓIDES**
- **3 GRUPOS : 3 CENTRÓIDES**

FUNÇÃO DISCRIMINANTE

- **COMBINAÇÃO LINEAR DE VARIÁVEIS INDEPENDENTES SELECIONADAS POR SEU PODER DISCRIMINATÓRIO NA ALOCAÇÃO DE ELEMENTOS A GRUPOS**

- **X GRUPOS : X - 1 FUNÇÕES DISCRIMINANTES**

- **EX: 3 GRUPOS - 2 FUNÇÕES DISCRIMINANTES ; 1A FUNÇÃO SEPARA UM GRUPO DOS OUTROS DOIS E A 2A SEPARA OS DOIS GRUPOS RESTANTES**

PESOS DISCRIMINANTES

- **PESOS DISCRIMINANTES (COEFIC. DA FUNÇÃO DISCRIMINANTE PADRONIZADOS) CONTRIBUIÇÃO RELATIVA DE CADA VARIÁVEL INDEPENDENTE PARA A FUNÇÃO DISCRIMINANTE (SINAL DESCONSIDERADO)**

- **ANÁLISE PREJUDICADA SE HOUVER MULTICOLINEARIDADE**

CARGAS DISCRIMINANTES

- **CORRELAÇÃO ENTRE FUNÇÃO DISCR. E VAR. INDEP. : CORRELAÇÃO ENTRE OS ESCORES DISCRIMINANTES E CADA VARIÁVEL INDEPENDENTE**

- **A HIERARQUIA DO PODER DISCRIMINANTE DAS VARIÁVEIS É A MESMA OBTIDA PELA ESTATÍSTICA WILKS'LAMBDA**

WILKS'LAMBDA (U)

$$U = \frac{SQ_{dentro\ dos\ grupos}}{SQ_{total}}$$

$$U = \frac{SQ_{total} - SQ_{entre\ grupos}}{SQ_{total}}$$

$$U = 1 - \frac{SQ_{entre\ grupos}}{SQ_{total}}$$

WILKS'LAMBDA (U)

- **U PRÓXIMO DE 1**

PARA DETERM. VAR. INDEPEND. AS MÉDIAS SÃO = NOS GRUPOS

A V. INDEP. NÃO DISCRIMINA

- **U PRÓXIMO DE 0**

AS MÉDIAS PARECEM SER DIFERENTES

A V. INDEP . DISCRIMINA

- **TESTE DE SIGNIFICÂNCIA**
- **H_0: AS MÉDIAS SÃO = NOS GPOS**

- **TESTE F: COMPARAR FOBS COM FCRÍTICO**

SE FOBS > FCRÍT REJ. H_0

OU ANALISAR N. S. SE N.S.OBS < N. S.CRÍT REJ. H_0

CORRELAÇÕES ENTRE VARIÁVEIS INDEPENDENTES

- **ESPERA-SE QUE AS VARIÁVEIS INDEP. NÃO SEJAM ALTAMENTE CORRELACIONADAS ENTRE SI**

- **NO PROGRAMA STEPWISE SERÁ PRIORIZADA A INCLUSÃO DAS VARIÁVEIS COM ALTO PODER DISCRIMINATÓRIO E QUE SEJAM MENOS CORRELACIONADAS ENTRE SI**

ESTATÍSTICA BOX'S M

- **TESTE DA IGUALDADE DAS VARIÂNCIAS E COVARIÂNCIAS NOS GRUPOS**

  SE N.S.OBS < N. S.CRÍT REJ. H_0

- **A ESTATÍSTICA BOX'S M PODE SER SENSÍVEL AO TAMANHO DA AMOSTRA E AO NÃO ATENDIMENTO DA HIPÓTESE DE DISTRIBUIÇÃO NORMAL**

EIGENVALUE

$$EIGENVALUE = \frac{SQ_{entre\ os\ grupos}}{SQ_{dentro\ dos\ grupos}}$$

- **EIGENVALUE ALTO IMPLICA BOAS FUNÇÕES DE DISCRIMINAÇÃO**

CORRELAÇÃO CANÔNICA

$$CORR.CANON. = \sqrt{\frac{SQ_{entre\ os\ grupos}}{SQ_{total}}}$$

- **MEDE GRAU DE ASSOC. ENTRE ESCORES DISCRIM. E GRUPOS**

$$U + CORR.CANON.^2 = 1$$

CORREL. ENTRE FUNÇÃO DISCR. E VAR. INDEPEND.

- **CORRELAÇÃO ENTRE OS ESCORES DISCRIMINANTES E CADA VARIÁVEL INDEPENDENTE**

- **CORRELAÇÕES ALTAS**

A VARIÁVEL INDEPENDENTE DISCRIMINA

- **CORRELAÇÕES BAIXAS**

A VARIÁVEL INDEPENDENTE NÃO DISCRIMINA

- **A HIERARQUIA DO PODER DISCRIMINANTE DAS VARIÁVEIS É A MESMA OBTIDA PELA ESTATÍSTICA WILKS'LAMBDA**

COEFICIENTE DE CLASSIFICAÇÃO DE FISHER

- **PODEM SER USADOS PARA CLASSIFICAÇÃO**
- **HÁ UM CONJUNTO DE COEFICIENTES PARA CADA GRUPO**
- **CADA CASO SERÁ CLASSIFICADO NO GRUPO ONDE O ESCORE DISCRIM. FOR MAIOR**

MATRIZ DE CLASSIFICAÇÃO

		GRUPO PREVISTO	
		1	2
G. REAL	1	N1	N2
	2	N3	N4

- **% CLASSIF. CORRETAS =**

$$\frac{(N1+N4)}{(N1+N2+N3+N4)}.100$$

MÉTODOS DE ANÁLISE DISCRIMINANTE

- **TODAS AS VARIÁVEIS INDEPENDENTES INCLUÍDAS**

- **STEPWISE: SELEÇÃO DAS VARIÁVEIS INDEPENDENTES COM MAIOR PODER DE DISCRIMINAÇÃO E NÃO CORRELACIONADAS ENTRE SI ;**

1o PASSO - SELECIONAR A VARIÁVEL QUE MAIS DISCRIMINA,

2o PASSO - ESCOLHER VARIÁVEL QUE MAIS CONTRIBUI PARA MELHORAR O PODER DISCRIMINATÓRIO DA FUNÇÃO EM COMBINAÇÃO COM A 1A ESCOLHIDA ETC ;

REMOVER VARIÁVEIS JÁ INCLUÍDAS SE SUA INFORMAÇÃO SOBRE AS DIFERENÇAS DOS GRUPOS ESTIVER NA COMBINAÇÃO DE VARIÁVEIS DEPOIS INCLUÍDAS.

MÉTODO STEPWISE DE ANÁLISE DISCRIMINANTE

- **WILKS' LAMBDA**

- **MAHALANOBIS DISTANCE**

$$D^2_{ab} = (n-g)\sum_{i=1}^{p}\sum_{j=1}^{p} w_{ij}^{*}(\overline{X}_{ia} - \overline{X}_{ib})(\overline{X}_{ja} - \overline{X}_{jb})$$

onde

p número variáveis independentes

g número de grupos

$\overline{X}_{ia}$ média i – ésima variável no grupo a

w_{ij}^{*} elemento do inverso da matriz de covariância dentro dos grupos

TAMANHO DA AMOSTRA

- **NÚMERO MÍNIMO DE OBSERVAÇÕES POR VARIÁVEL INDEPENDENTE : 5**
- **NÚMERO RECOMENDADO: 20 OBSERVAÇÕES POR VARIÁVEL**
- **NÚMERO DE OBSERVAÇÕES POR GRUPO : O MENOR GRUPO DEVE EXCEDER O NÚMERO DE VARIÁVEIS INDEPENDENTES**
- **CADA GRUPO : NO MÍNIMO 20 OBSERVAÇÕES**

DETERMINAÇÃO DO ESCORE DE CORTE

- **GRUPOS DE MESMO TAMANHO**

$$Z_{CE} = \frac{Z_A + Z_B}{2},$$

onde Z_A = centróide grupo A

Z_B = centróide grupo B

DETERMINAÇÃO DO ESCORE DE CORTE

- **GRUPOS DE TAMANHO DIFERENTE**

$$Z_{CE} = \frac{N_B Z_A + N_A Z_B}{N_A + N_B},$$

onde Z_A = centróide grupo A

Z_B = centróide grupo B

N_A = tamanho grupo A

N_B = tamanho grupo B

TEST t PARA DETERMINAÇÃO NÍVEL DE SIGNIFICÂNCIA PRECISÃO DA CLASSIFICAÇÃO

- **GRUPOS DE MESMO TAMANHO**

$$t = \frac{p - 0{,}5}{\sqrt{\frac{0{,}5(1{,}0 - 0{,}5)}{N}}} \text{ , onde}$$

$p = \text{proporção classificações corretas}$

$N = \text{tamanho da amostra}$

PRECISÃO DA CLASSIFICAÇÃO PELO CRITÉRIO DO ACASO

- **GRUPOS DE MESMO TAMANHO**

$$C = \frac{1}{\text{número grupos}}$$

- **GRUPOS DE TAMANHO DIFERENTE**

$$C_{PRO} = p^2 + (1-p)^2, \text{onde}$$

$$p = \text{proporção elementos grupo 1}$$

$$1-p = \text{proporção elementos grupo 2}$$

PRECISÃO DA CLASSIFICAÇÃO PELO CRITÉRIO DO ACASO

- **CRITÉRIO SUGERIDO PARA MÍNIMO ACEITÁVEL PARA A PROPORÇÃO DE CLASSIFICAÇÕES CORRETAS : PELO MENOS 25% A MAIS DO QUE O OBTIDO AO ACASO (GRUPOS DE MESMO TAMANHO)**

TESTE χ^2 DO PODER DISCRIMINATÓRIO DA MATRIZ DE CLASSIFICAÇÃO COMPARADA COM O MODELO DO ACASO

$$\text{Press's } Q = \frac{\left[N - (nk)\right]^2}{N(k-1)}$$

onde

$N = \text{tamanho amostra}$

$n = \text{número classificações corretas}$

$k = \text{número de grupos}$

1 g.l.

TESTE χ^2 DO PODER DISCRIMINATÓRIO DA MATRIZ DE CLASSIFICAÇÃO COMPARADA COM O MODELO DO ACASO

$$\text{Press's } Q = \frac{\left[N - (nk)\right]^2}{N(k-1)}$$

onde

$N = \text{tamanho amostra}$

$n = \text{número classificações corretas}$

$k = \text{número de grupos}$

1 g.l.

TESTE χ^2 DO PODER DISCRIMINATÓRIO DA MATRIZ DE CLASSIFICAÇÃO COMPARADA COM O MODELO DO ACASO

$$\text{Press's } Q = \frac{\left[N - (nk)\right]^2}{N(k-1)}$$

onde

$N = \text{tamanho amostra}$

$n = \text{número classificações corretas}$

$k = \text{número de grupos}$

1 g.l.

EXEMPLO

		Mean	Std. Dev
IRISTYPE			
1	PETALLEN	42,8	4,537
	PETALWID	13,33	1,941
	SEPALLEN	59,55	5,033
	SEPALWID	27,8	3,096
2	PETALLEN	55,52	5,519
	PETALWID	20,26	2,747
	SEPALLEN	65,88	6,359
	SEPALWID	29,74	3,225
Total	PETALLEN	49,22	8,136
	PETALWID	16,83	4,214
	SEPALLEN	62,75	6,538
	SEPALWID	28,78	3,294

EXEMPLO

Pooled Within-Groups Matrices - Correlation				
	PETALLEN	PETALWID	SEPALLEN	SEPALWID
PETALLEN	1,000	0,484	0,813	0,457
PETALWID	0,484	1,000	0,364	0,574
SEPALLEN	0,813	0,364	1,000	0,473
SEPALWID	0,457	0,574	0,473	1,000

Tests of Equality of Group Means			
	Wilks' Lambda	F	Sig.
PETALLEN	0,382	156,685	1,155E-18
PETALWID	0,316	209,666	1,155E-18
SEPALLEN	0,763	30,073	3,306E-07
SEPALWID	0,912	9,357	0,0028714

EXEMPLO

Box's M		35,87693554
F	Approx.	3,427830601
	df1	10
	df2	44940,07607
	Sig.	0,000166153
Tests null hypothesis of equal population covariance matrices		

	Entered	**Wilks' Lambda**	
		Statistic	
Step			Sig.
1	PETALWID	0,316	1,15486E-18
2	SEPALWID	0,279	0
3	PETALLEN	0,234	7,4041E-30
4	SEPALLEN	0,217	2,5455E-30
At each step, the variable that minimizes the overall Wilks' Lambda is entered.			

EXEMPLO

Eigenvalues				
Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	3,6023244	100	100	0,88471379

Structure Matrix		
	Function	
	1	
PETALWID	0,775	
PETALLEN	0,670	
SEPALLEN	0,293	
SEPALWID	0,164	
Pooled within-groups correlations between discriminating variables and standardized canonical discriminant functions		

EXEMPLO

Standardized Canonical Discriminant Function Coefficients	
	Function
	1
PETALLEN	0,939
PETALWID	0,779
SEPALLEN	-0,537
SEPALWID	-0,460

Canonical Discriminant Function Coefficients	
	Function
	1
PETALLEN	0,186
PETALWID	0,327
SEPALLEN	-0,094
SEPALWID	-0,145
(Constant)	-4,596
Unstandardized coefficients	

EXEMPLO

Functions at Group Centroids	
	Function
IRISTYPE	1
1	-1,8978
2	1,8598
Unstandardized canonical discriminant	
functions evaluated at group means	

Classification Function Coefficients		
	IRISTYPE	
	1	2
PETALLEN	-0,195	0,503
PETALWID	-0,031	1,199
SEPALLEN	1,528	1,176
SEPALWID	1,625	1,079
(Constant)	-64,404	-81,602
Fisher's linear discriminant functions		

EXEMPLO

Classification Results					
			Predicted Group Members		Total
		IRISTYPE	1	2	
Original	Count	1	47	2	49
		2	1	49	50
	%	1	95,918	4,082	100
		2	2	98	100
	97,0% of original grouped cases correctly classified.				