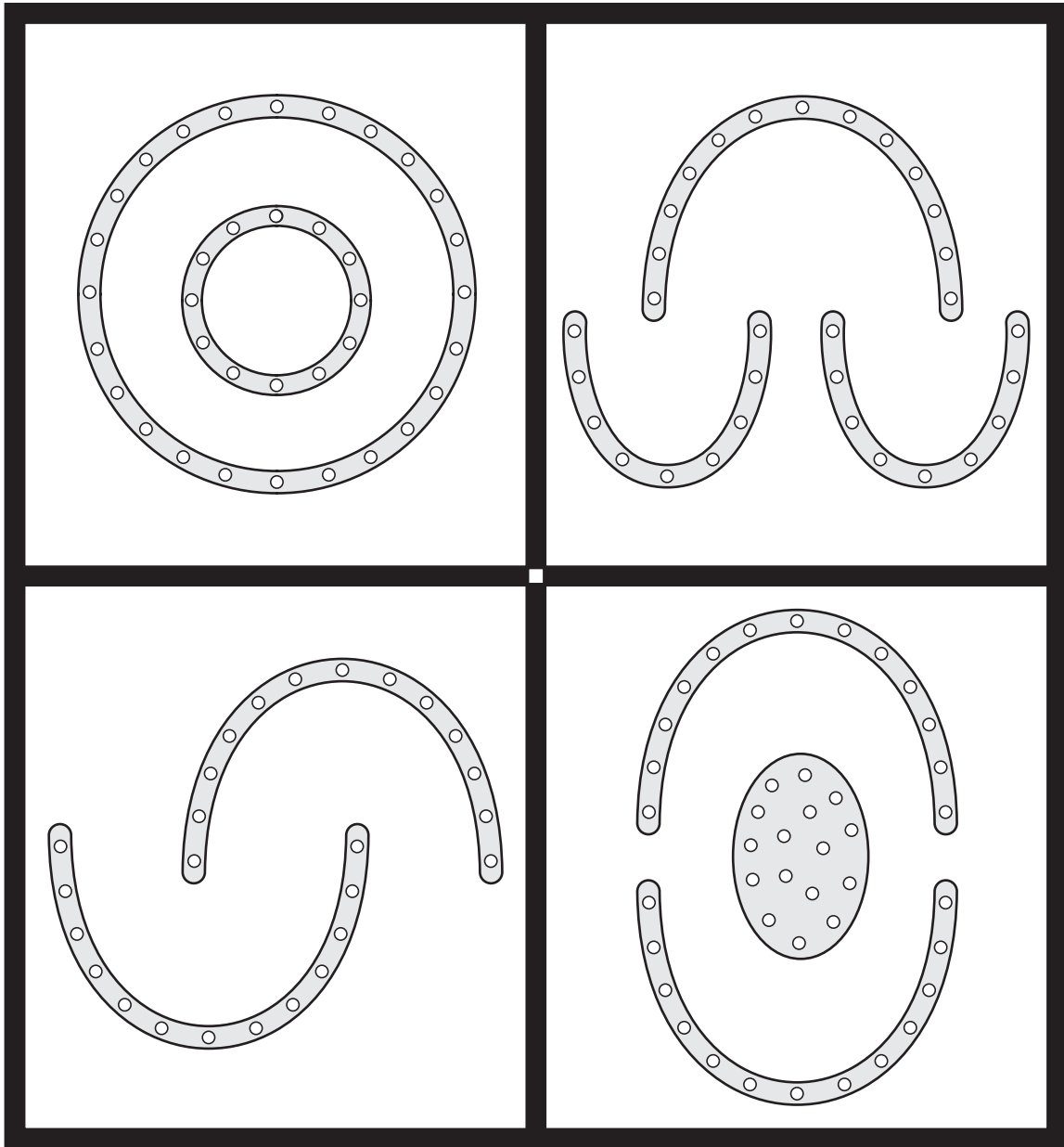




# $\rho$ -DENSITY

## Data Clustering Math



## Outlier Detection

# Generalized Distance

Prof. Aguinaldo Prandini Ricieri



## XIV

## Gedächtnisrede auf H. Minkowski.

Minkowski für jenes Maximum  $k_n$  des Minimums der quadratischen Form  $F$  auch zu einer unteren Schranke. Ein überraschendes und für die Genauigkeit der Minkowskischen Methode zeugendes Resultat ist es, daß diese untere Schranke und die früher gefundene obere Schranke asymptotisch für  $n = \infty$  ineinander fließen, so daß Minkowski die Limesgleichung

$$\lim_{n \rightarrow \infty} \frac{\log k_n}{\log n} = 1$$

aussprechen konnte.

Ch. Hermite, damals der Senior der französischen Mathematiker, hatte von Anbeginn die zahlentheoretischen Arbeiten Minkowskis mit höchstem Interesse und lebhaftester Freude verfolgt. Es ist rührend, wie rückhaltlos er die Vorzüge der Minkowskischen Methode gegenüber seinen eigenen Entwicklungen anerkennt, als Minkowski ihm die eben besprochenen Resultate mitteilt. „Au premier coup d’œil j’ai reconnu“, so schreibt Ch. Hermite in einem der an Minkowski gerichteten Briefe, „que vous avez été bien au delà de mes recherches en nous ouvrant dans le domaine arithmétique des voies toutes nouvelles.“ Und in einem zwei Jahre späteren Briefe vom November 1892 heißt es: „Je me sens rempli d’étonnement et de plaisir devant vos principes et vos résultats, ils m’ouvrent comme un monde arithmétique entièrement nouveau, où les questions fondamentales de notre science sont traitées avec un éclatant succès auquel tous les géomètres rendront hommage. Vous voulez bien, Monsieur, — et je vous en suis sincèrement reconnaissant — rapporter à mes anciennes recherches le point de départ de vos beaux travaux, mais vous les avez tant dépassées qu’elles ne gardent plus d’autre mérite que d’avoir ouvert la voie dans laquelle vous êtes entré.“

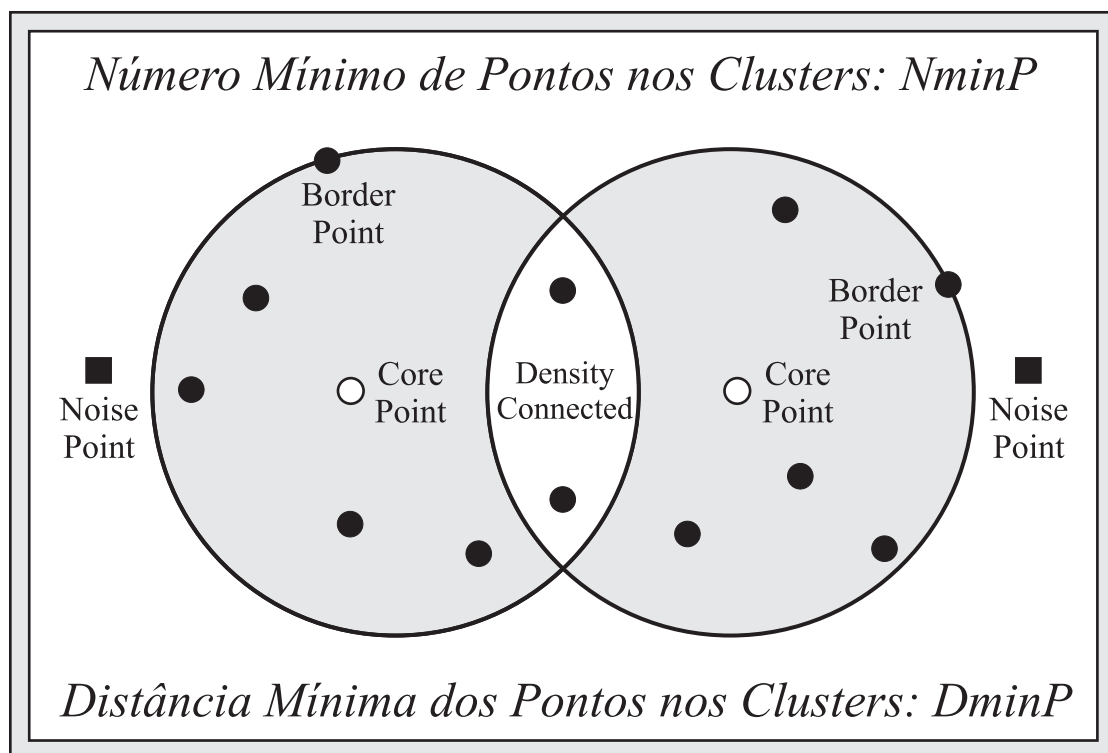
Hiernach nimmt es nicht Wunder, daß Hermites Begeisterung für die zahlentheoretischen Methoden Minkowskis keine Grenzen kannte, als die erste Lieferung seiner Geometrie der Zahlen 1896 erschien. „Je crois voir la terre promise“, so schreibt Hermite an Laugel, von dem er sich eine Übersetzung des Minkowskischen Buches zu seinem persönlichen Gebrauch anfertigen ließ. Und in der Tat, welche Fälle der verschiedenartigsten und tieflegendsten arithmetischen Wahrheiten werden in diesem Hauptwerke Minkowskis durch das geometrische Band gehalten und verknüpft! Die Theorie der Einheiten in den algebraischen Zahlkörpern, Sätze über die Ordnung einer endlichen Gruppe von homogenen linearen ganzzahligen Substitutionen und über die Zahl der Transformationen einer positiven quadratischen Form in sich, der Beweis für die Endlichkeit der Klassenanzahl von positiven quadratischen Formen mit gegebener Determinante, die Annäherung an beliebig viele reelle Größen durch rationale Zahlen mit den gleichen Nennern, die Theorie der Linearformen mit



# $\rho$ -Density

## 1.0 Clusterização por Densidade de Dados

No algoritmo de clusterização por densidade (*minimum number of points required to form a dense region*), os pontos para serem aglutinados são analisados um com todos e todos com um, tendo de medirem uma distância mínima de seus vizinhos (*euclidian, manhattan or minkowski distance, neighborhood around each data point*). O ponto do banco de dados que não pertencer a nenhum cluster, por estar muito afastado dos possíveis *core points*, como mostra o próximo esquema alegórico, será tido como um *noise point* ou também *outlier point*.



**Rótulos do Algoritmo de Clusterização  $\rho$ -Density e seu Banco de Dados**

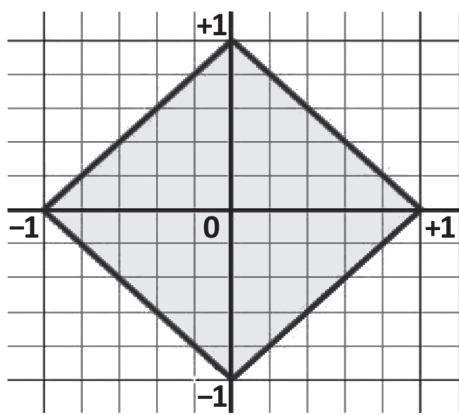
*Hierarchical Density Estimates for Data Clustering, Visualization, and Outlier Detection.*

*ACM Transactions on Knowledge, Discovery from Data (10), p. 1, 2015*

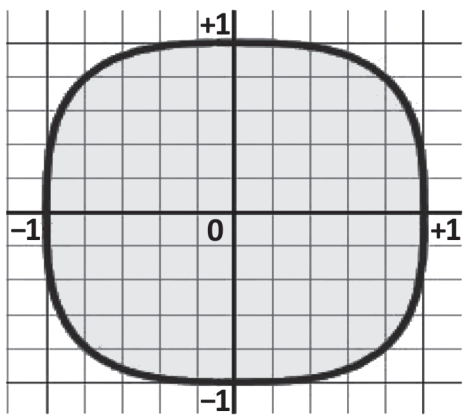
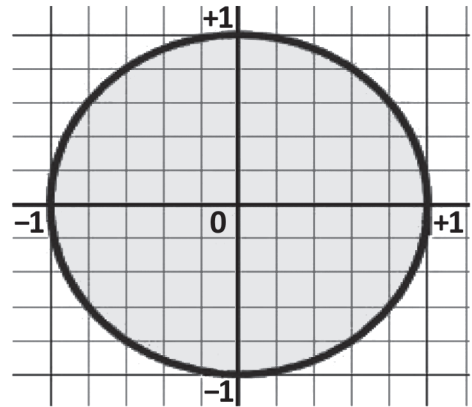
*Ricardo Campello, Davoud Moulavi, Artur Zimek, Jörg Sander*



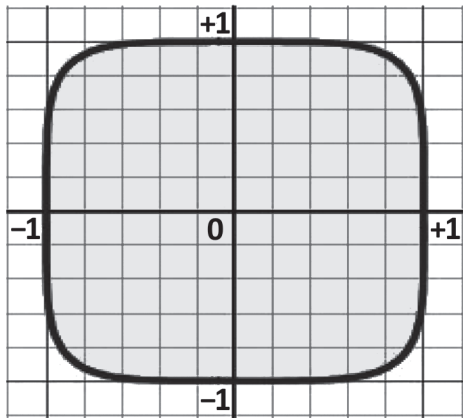
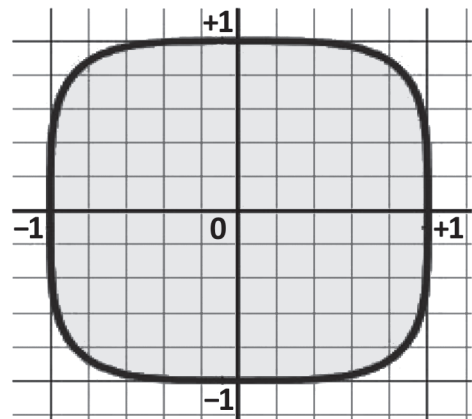
# 1.2 Morfogênese Métrica em Termos do Parâmetro m



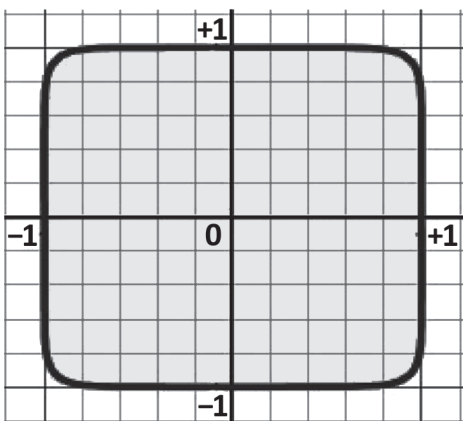
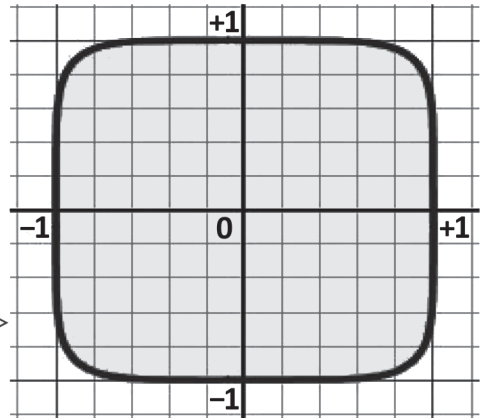
$$\begin{aligned} &\longleftarrow |x|^1 + |y|^1 = 1^1 \\ &|x|^2 + |y|^2 = 1^2 \longrightarrow \end{aligned}$$



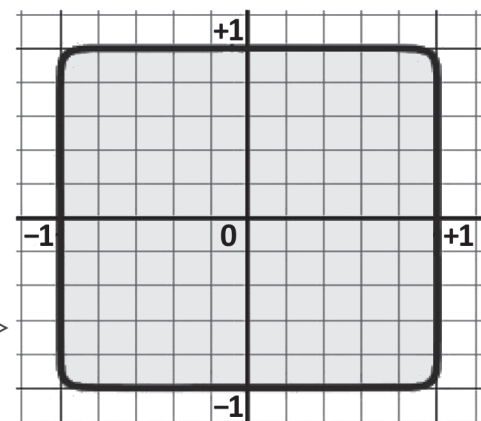
$$\begin{aligned} &\longleftarrow |x|^3 + |y|^3 = 1^3 \\ &|x|^4 + |y|^4 = 1^4 \longrightarrow \end{aligned}$$



$$\begin{aligned} &\longleftarrow |x|^5 + |y|^5 = 1^5 \\ &|x|^6 + |y|^6 = 1^6 \longrightarrow \end{aligned}$$



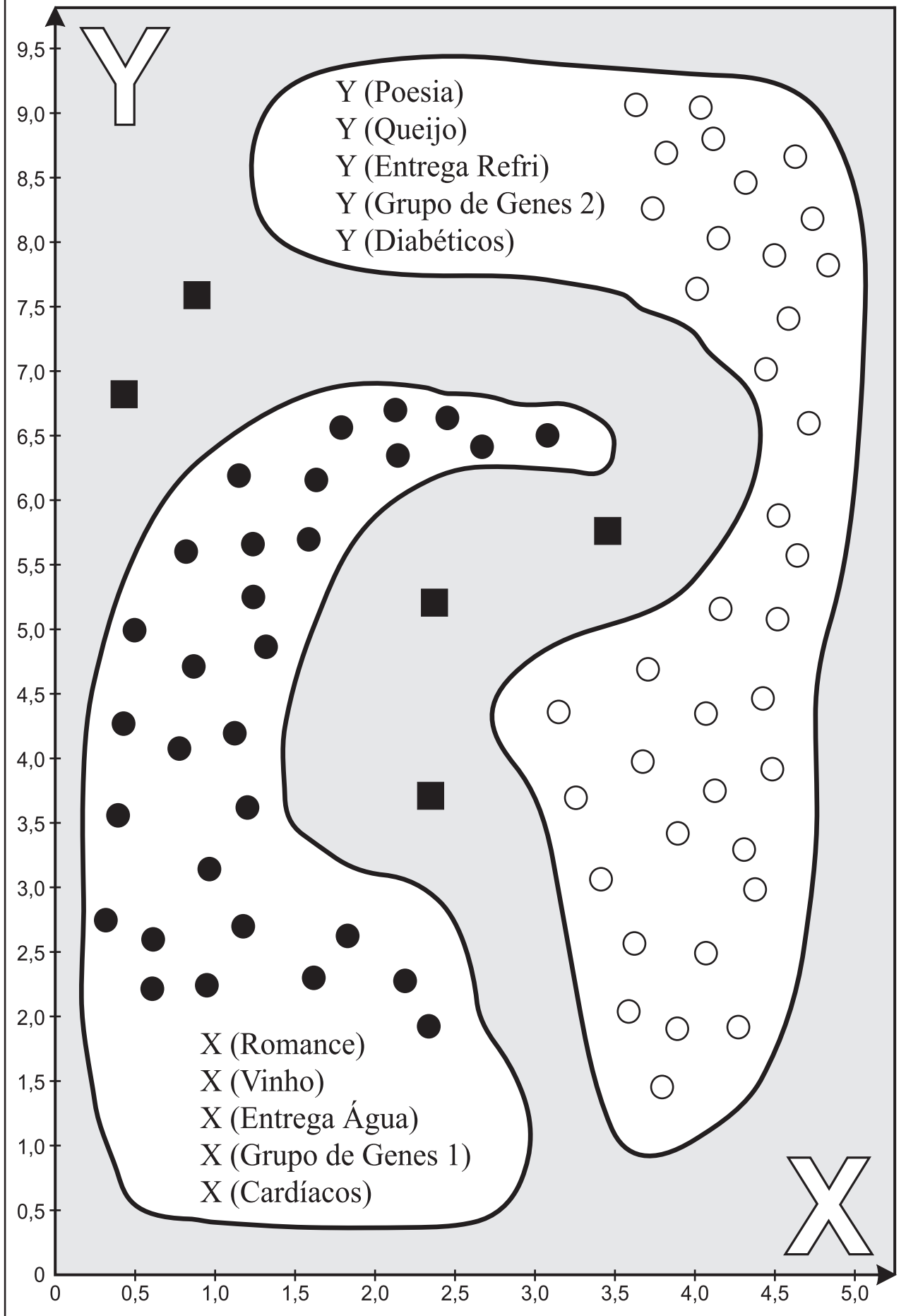
$$\begin{aligned} &\longleftarrow |x|^7 + |y|^7 = 1^7 \\ &|x|^8 + |y|^8 = 1^8 \longrightarrow \end{aligned}$$







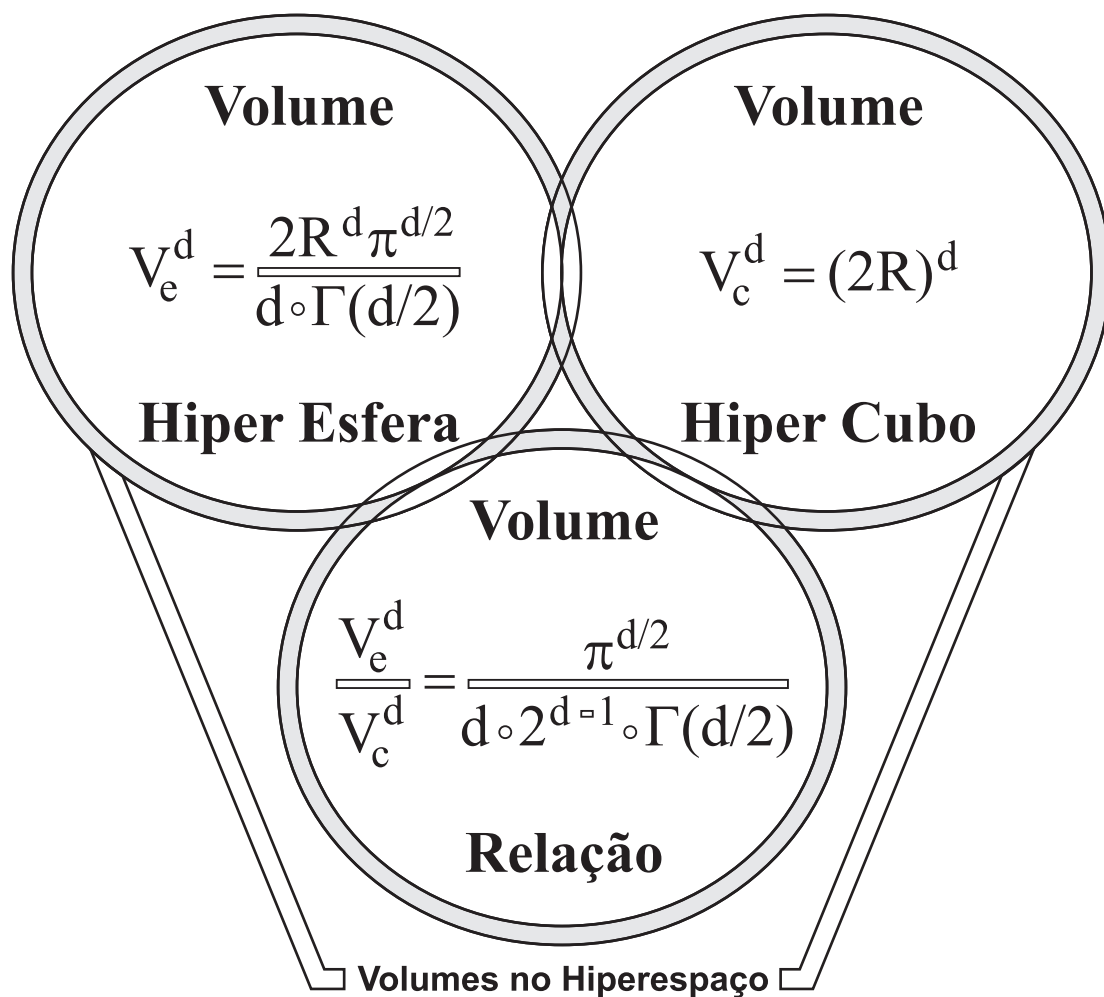
# 1.5 Exemplo da Separação de Dois Clusters





## 2.0 Maldição da Dimensionalidade dos Dados

Sendo a dimensão dos dados, relativos ao estudo de um fenômeno, o número de coordenadas intrínsecas e independentes necessárias para analisá-lo nos seus pormenores. Segundo Richard Bellman (1957) quando sua dimensionalidade aumenta, o volume do espaço de busca aumenta também de modo que o número de dados, anteriormente registrados, tornam-se “rarefeitos”, o que implica em um maior número de dados (“exponencial”) para melhor interpretar o fenômeno mirado.

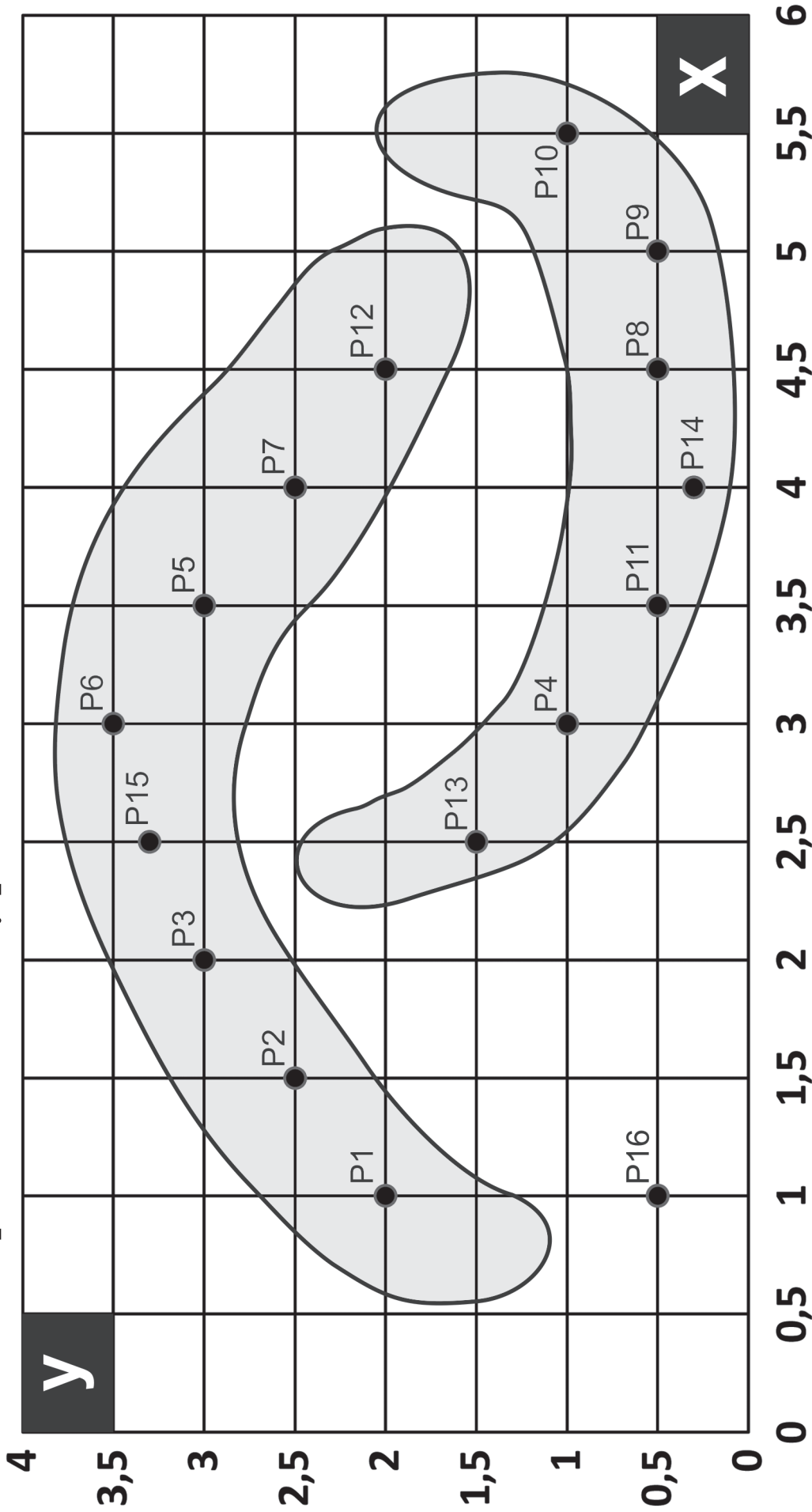


Essa “maldição computacional” pode ser melhor entendida sabendo que se a distância euclidiana, definida para muitas coordenadas aumentar ( $d \rightarrow$  grande), haverá pouca diferença entre os pares de pontos que matematizam o fenômeno.

## Estatística de Bellman $d$ -Dimensional



# 4.5 Aplicação 2-Density para $N_{min}P = 2$ e $D_{min}P = 1,0$ , Minimum

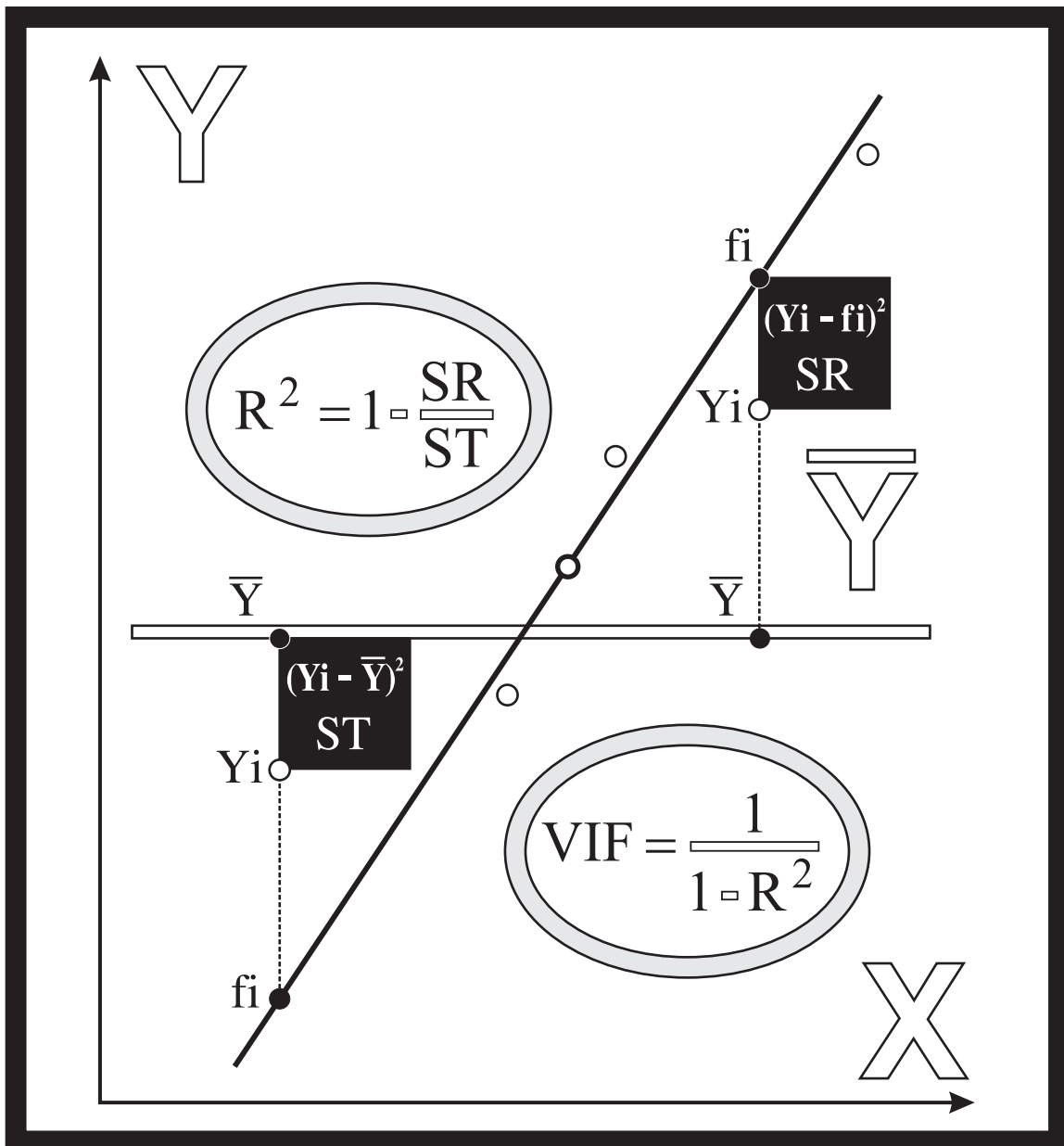


Data Clustering // Manhattan Distance // Outlier Detection. P16



# VIF-INFLATION

## Inflação da Variância



## Regressões Lineares

# Multicolinearidade de Dados

Prof. Aginaldo Prandini Ricieri





# CORRELATION AND CAUSATION

By SEWALL WRIGHT

*Senior Animal Husbandman in Animal Genetics, Bureau of Animal Industry, United States Department of Agriculture*

## PART I. METHOD OF PATH COEFFICIENTS

### INTRODUCTION

The ideal method of science is the study of the direct influence of one condition on another in experiments in which all other possible causes of variation are eliminated. Unfortunately, causes of variation often seem to be beyond control. In the biological sciences, especially, one often has to deal with a group of characteristics or conditions which are correlated because of a complex of interacting, uncontrollable, and often obscure causes. The degree of correlation between two variables can be calculated by well-known methods, but when it is found it gives merely the resultant of all connecting paths of influence.

The present paper is an attempt to present a method of measuring the direct influence along each separate path in such a system and thus of finding the degree to which variation of a given effect is determined by each particular cause. The method depends on the combination of knowledge of the degrees of correlation among the variables in a system with such knowledge as may be possessed of the causal relations. In cases in which the causal relations are uncertain the method can be used to find the logical consequences of any particular hypothesis in regard to them.

### CORRELATION

Relations between variables which can be measured quantitatively are usually expressed in terms of Galton's (4)<sup>1</sup> coefficient of correlation,

$r_{XY} = \frac{\sum X'Y'}{n\sigma_X\sigma_Y}$  (the ratio of the average product of deviations of  $X$  and  $Y$  to the product of their standard deviations), or of Pearson's (7) correlation

ratio,  $\eta_{X \cdot Y} = \frac{\sigma\left(\frac{Y - \bar{Y}}{X - \bar{X}}\right)}{\sigma_X}$  (the ratio of the standard deviation of the mean values

of  $X$  for each value of  $Y$  to the total standard deviation of  $X$ ), the standard deviation being the square root of the mean square deviation.

Use of the coefficient of correlation ( $r$ ) assumes that there is a linear relation between the two variables—that is, that a given change in one variable always involves a certain constant change in the corresponding average value of the other. The value of the coefficient can never exceed



## 4.0 Definições e Deduções Pertinentes às Regressões

$R^2$  (Coeficiente de Determinação) e VIF (Fator de Inflação da Variância) relacionam-se segundo a formatação estatística das variáveis na forma:

$$\mathbf{ST} \text{ (Soma Total)} = \sum [y_i - \bar{y}]^2$$

$$\mathbf{SR} \text{ (Soma Residual)} = \sum [y_i - f_i]^2$$

$$\mathbf{SE} \text{ (Soma Estatística)} = \sum [f_i - \bar{y}]^2$$

$$\mathbf{SZ} \text{ (Soma Zephirum)} = \sum [y_i - f_i]$$



$$ST = \sum [y_i - \bar{y}]^2 \quad \therefore \quad ST = \sum [(f_i - \bar{y}) + (y_i - f_i)]^2$$

$$ST = \sum [(f_i - \bar{y})^2 + \sum [y_i - f_i]^2 + \sum 2(f_i - \bar{y}) \cdot (y_i - f_i)]$$

$$ST = SE + SR + \cancel{SZ}^0$$

$$\frac{ST}{ST} = \frac{SE}{ST} + \frac{SR}{ST}$$

$$1 = \frac{SR}{ST} = \frac{SE}{ST} \quad \therefore \quad \frac{SE}{ST} = R^2 \quad \therefore \quad R^2 = 1 - \frac{SR}{ST}$$

$$-1 + \frac{SR}{ST} = -R^2$$

$$\frac{SR}{ST} = 1 - R^2 \quad \therefore \quad \frac{ST}{SR} = \frac{1}{1 - R^2} \quad \therefore \quad \mathbf{VIF} = \frac{1}{1 - R^2}$$

**Parâmetros Estatísticos Associados com a Soma:  $SZ \rightarrow 0$**



# 11.0 Parâmetros R<sup>2</sup> e VIF para Nova Rodada

LÇ	X1	X2	Y	f	SR	SE	ST
a	1	1	7	7,0297	0,0009	4,7102	4,8400
b	2	4	8	7,9585	0,0017	1,5413	1,4400
c	3	6	9	8,9916	0,0001	0,0434	0,0400
d	4	8	10	10,0247	0,0006	0,6801	0,6400
e	5	1	12	11,9965	0,0000	7,8204	7,8400
<b>f = 5,8923 + 1,24175 X1 - 0,10439 X2</b>				<b>0,0033</b>	<b>14,7955</b>	<b>14,8000</b>	
R <sup>2</sup> (Xi, Y) = 0,999777				VIF (Xi, Y) = 4488,602589			

COR	X1	X2	Y
X1	1,000	0,205	0,986
X2	0,205	1,000	0,042
Y	0,986	0,042	1,000

**f (X1, X2,X3, X4)**

LÇ	X1	X2	Y	f	SR	SE	ST
a	1	1	7	7,1000	0,0100	4,4100	4,8400
b	2	4	8	7,8000	0,0400	1,9600	1,4400
c	3	6	9	9,0000	0,0000	0,0400	0,0400
d	4	8	10	10,2000	0,0400	1,0000	0,6400
e	5	9	12	11,9000	0,0100	7,2900	7,8400
<b>f = 5,400 + 2,200 X1 - 0,5000 X2</b>				<b>0,1000</b>	<b>14,7000</b>	<b>14,8000</b>	
R <sup>2</sup> (Xi, Y) = 0,993243				VIF (Xi, Y) = 147,994672			

COR	X1	X2	Y
X1	1,000	0,985	0,986
X2	0,985	1,000	0,948
Y	0,986	0,948	1,000

**f (X1, X2,X3, X4)**

LÇ	X1	X2	Y	f	SR	SE	ST
a	1	1	2	2,2769	0,0767	47,9293	51,8400
b	2	4	8	6,7768	1,4962	5,8719	1,4400
c	3	6	9	10,5844	2,5103	1,9166	0,0400
d	4	8	15	14,3920	0,3697	26,9569	33,6400
e	5	1	12	11,9689	0,0010	7,6668	7,8400
<b>f = -0,8384 + 2,4230 X1 + 0,6923 X2</b>				<b>4,4538</b>	<b>90,3414</b>	<b>94,8000</b>	
R <sup>2</sup> (Xi, Y) = 0,953018				VIF (Xi, Y) = 21,284747			

COR	X1	X2	Y
X1	1,000	0,205	0,877
X2	0,205	1,000	0,600
Y	0,877	0,600	1,000

**f (X1, X2,X3, X4)**

LÇ	X1	X2	Y	f	SR	SE	ST
a	1	1	2	3,3647	1,8624	23,3801	38,4400
b	2	4	8	6,0789	3,6906	4,4991	0,0400
c	3	6	9	8,5953	0,1638	0,1563	0,6400
d	4	8	10	11,1117	1,2359	8,4780	3,2400
e	5	1	12	11,8479	0,0231	13,3072	14,4400
<b>f = 1,0461 + 2,1208 X1 + 0,1978 X2</b>				<b>6,9758</b>	<b>49,8206</b>	<b>56,8000</b>	
R <sup>2</sup> (Xi, Y) = 0,877186				VIF (Xi, Y) = 8,142394			

COR	X1	X2	Y
X1	1,000	0,205	0,923
X2	0,205	1,000	0,344
Y	0,923	0,344	1,000

**f (X1, X2,X3, X4)**

LÇ	X1	X2	Y	f	SR	SE	ST
a	1	1	7	5,9416	1,1202	18,1340	10,2400
b	2	4	8	8,6558	0,4301	2,3846	4,8400
c	3	6	9	10,9799	3,9200	0,6082	1,4400
d	4	8	15	13,3040	2,8764	9,6348	23,0400
e	5	1	12	12,1172	0,0137	3,6757	3,2400
<b>f = 4,007 + 1,5439 X1 + 0,3901 X2</b>				<b>8,3604</b>	<b>34,4372</b>	<b>42,8000</b>	
R <sup>2</sup> (Xi, Y) = 0,804663				VIF (Xi, Y) = 5,119348			

COR	X1	X2	Y
X1	1,000	0,205	0,822
X2	0,205	1,000	0,521
Y	0,822	0,521	1,000

**f (X1, X2,X3, X4)**

# $\lambda$ -RIDGE

## Multicollinearity Fixing

$$(\mathbf{X}^T \cdot \mathbf{X})^{-1} \rightarrow \nexists$$

$$\hat{y} = \mathbf{X}_1 \cdot \mathbf{b}_1 + \mathbf{X}_2 \cdot \mathbf{b}_2 + \mathbf{X}_3 \cdot \mathbf{b}_3 + \dots + \epsilon$$

$$\mathbf{b}_i = (\mathbf{X}^T \cdot \mathbf{X} + \lambda \cdot \mathbf{I})^{-1} \cdot \mathbf{X}^T \cdot \mathbf{y}$$

$$R^2 = 1 - \left( \frac{n-1}{n-1-p} \right) \cdot \left( 1 - \frac{\sum (\hat{y} - \bar{y})^2}{\sum (y_i - \bar{y})^2} \right)$$

$$VIF = \frac{1}{1 - R^2}$$

Regressão de Cume

Large Numbers of Parameters

Prof. Aguinaldo Prandini Ricieri





## 2.2 Ajuste de uma Regressão a um Banco de Dados

LÇ	Um	X1	X2	X3	X4	Y
01	1	1	4	3	9	7
02	1	5	3	2	1	4
03	1	8	2	8	3	5
04	1	5	8	5	6	6
05	1	4	3	9	6	6
06	1	8	6	2	4	5
07	1	4	1	1	8	7
08	1	6	5	6	1	4

TABELA DE DADOS COM ENTRADAS X E SAÍDAS Y AJUSTADA COM A COLUNA DE Um

$$\mathbf{b} = (\mathbf{X}^t \circ \mathbf{X})^{-1} \circ \mathbf{X}^t \circ \mathbf{Y}$$

$$\mathbf{F} = 3,571 + 0,024 \circ \mathbf{X1} - 0,042 \circ \mathbf{X2} + 0,010 \circ \mathbf{X3} + 0,405 \circ \mathbf{X4}$$

$$b_0 = + 3,571$$

$$b_1 = + 0,024$$

$$b_2 = - 0,042$$

$$b_3 = + 0,010$$

$$b_4 = + 0,405$$

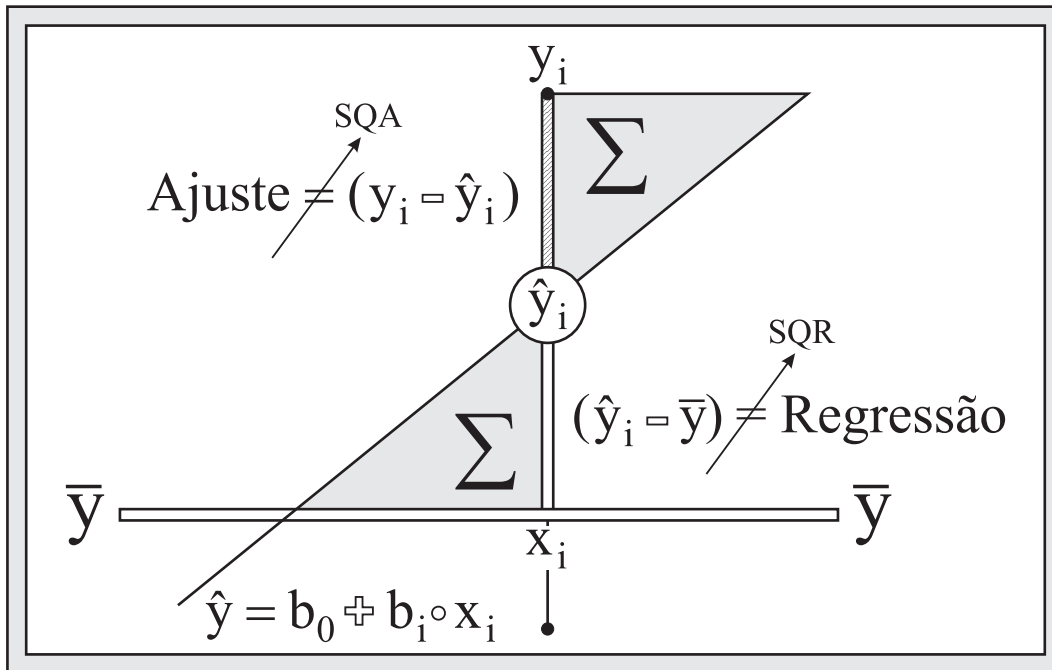
X1	X2	X3	X4	Y	F	(F-Y)^2
1	4	3	9	7	7,1050	0,0110
5	3	2	1	4	3,9920	0,0001
8	2	8	3	5	4,9750	0,0006
5	8	5	6	6	5,8420	0,0250
4	3	9	6	6	6,0630	0,0040
8	6	2	4	5	5,1560	0,0243
4	1	1	8	7	6,8750	0,0156
6	5	6	1	4	3,9740	0,0007

TABELA COM O CÁLCULO DO MSE: MINIMUM SQUARE ERROR



# 3.0 Coeficiente de Determinação R2 Ajustado

Para n dados experimentados e acomodados em p colunas de variáveis explicativas de um determinado fenômeno estatístico, define-se a qualidade da variância desses dados como R2 que varia entre os valores 0 e 1 segundo o esquema:



A soma quadrática total (SQT) das diferenças indicadas pelos rótulos SQR (Soma da regressão) e SQA (Soma do ajuste) implica na próxima formatação:

$$SQT = SQR + SQA$$

$$\sum (y_i - \bar{y})^2 = \sum (\hat{y}_i - \bar{y}) + \sum (y_i - \hat{y}_i)$$

$$R^2 = \frac{SQR}{SQT} = \frac{SQT - SQA}{SQT}$$

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

R Quadrado Primal: (R²)

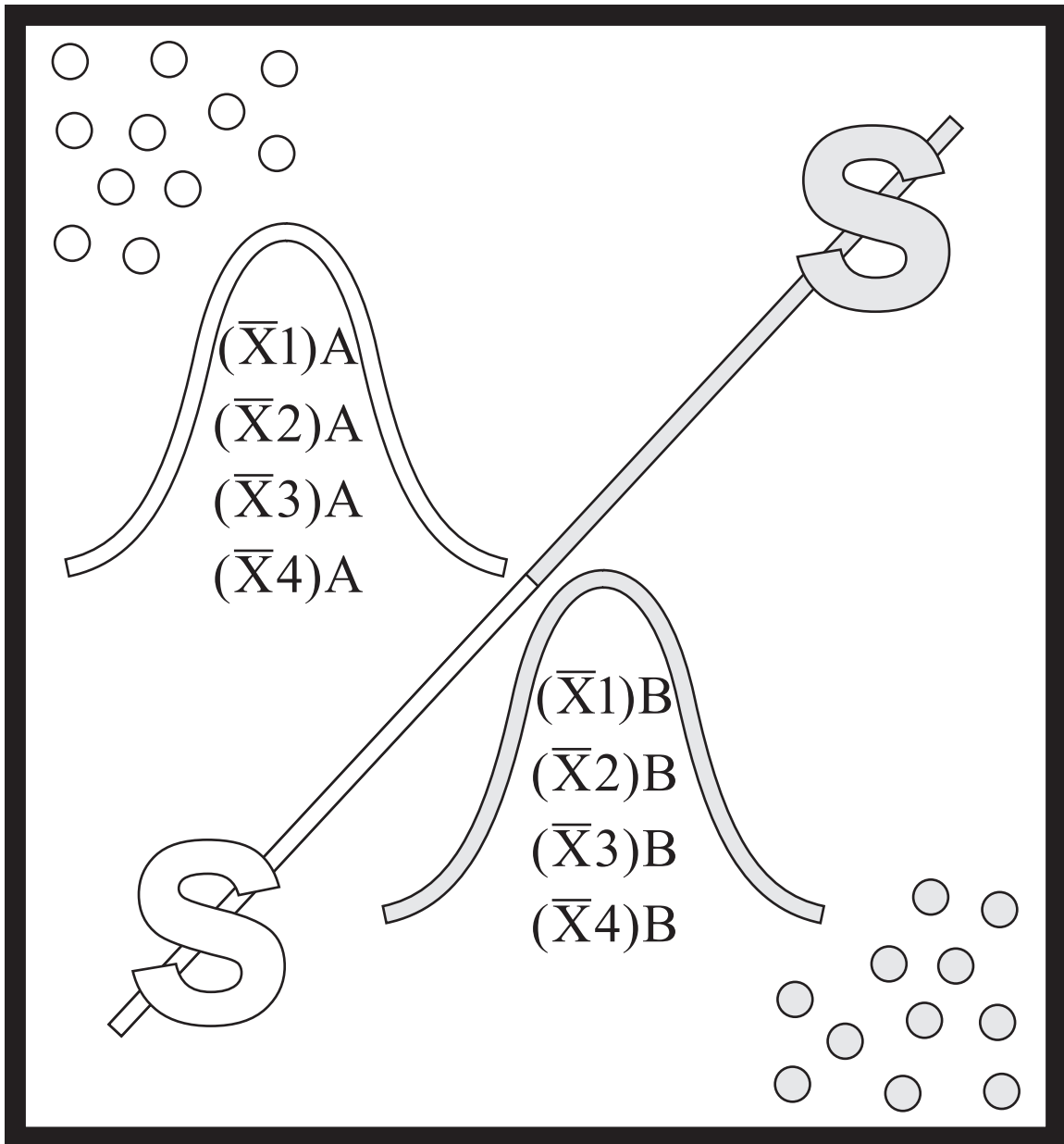
$$R2 = 1 - \left( \frac{n - 1}{n - 1 - p} \right) \cdot \left( \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \right)$$

R Quadrado Ajustado: (R2), R Deitado



# S-SEPARATOR

## Praeludium Futurum



## Transformação Linear

# Classificador Estatístico

Prof. Aguinaldo Prandini Ricieri



The consent of the family of the late Professor Benjamin Peirce has been kindly given to the publication in the *American Journal of Mathematics*, of his valuable and unique memoir on Linear Associative Algebra, of which only a small number of copies in lithograph were taken in the author's lifetime, for distribution among his friends. This publication will, it is believed, supply a want which has been long and widely felt, and bring within the reach of the general mathematical public a work which may almost be entitled to take rank as the *Principia* of the philosophical study of the laws of algebraical operation.

## *Linear Associative Algebra.*

*A Memoir read before the National Academy of Sciences in Washington, 1870.*

BY BENJAMIN PEIRCE.

WITH NOTES AND ADDENDA, BY C. S. PEIRCE, SON OF THE AUTHOR.

1. Mathematics is the science which draws necessary conclusions.

This definition of mathematics is wider than that which is ordinarily given, and by which its range is limited to quantitative research. The ordinary definition, like those of other sciences, is objective; whereas this is subjective. Recent investigations, of which quaternions is the most noteworthy instance, make it manifest that the old definition is too restricted. The sphere of mathematics is here extended, in accordance with the derivation of its name, to all demonstrative research, so as to include all knowledge strictly capable of dogmatic teaching. Mathematics is not the discoverer of laws, for it is not induction; neither is it the framer of theories, for it is not hypothesis; but it is the judge over both, and it is the arbiter to which each must refer its claims; and neither law can rule nor theory explain without the sanction of mathematics. It deduces from a law all its consequences, and develops them into the suitable form for comparison with observation, and thereby measures the strength of the argument from observation in favor of a proposed law or of a proposed form of application of a law.

Mathematics, under this definition, belongs to every enquiry, moral as well as physical. Even the rules of logic, by which it is rigidly bound, could not be deduced without its aid. The laws of argument admit of simple statement, but they must be curiously transposed before they can be applied to the living speech and verified by observation. In its pure and simple form the syllogism cannot be directly compared with all experience, or it would not have required an





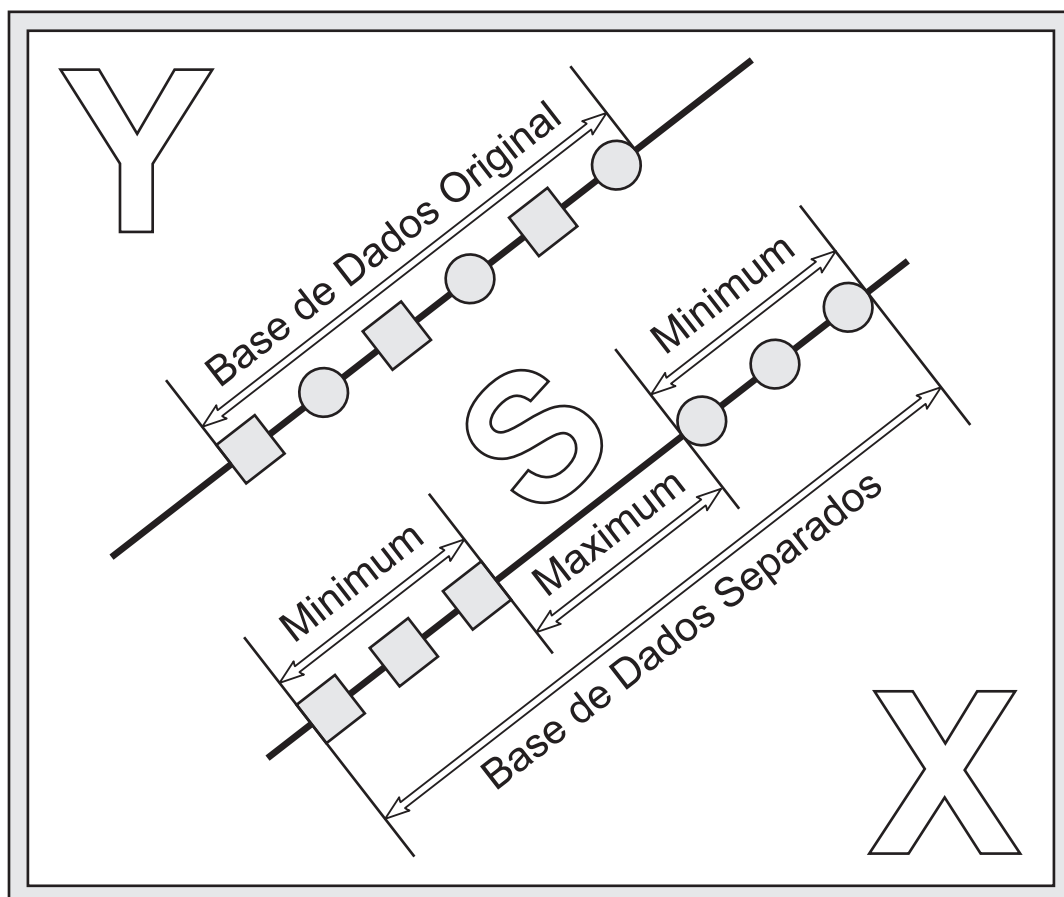
# S-Separator

## 1.0 Classificação de Classes de Dados Afins

A combinação linear de dados usada na estatística de aprendizado de máquina, permite disjuntir propriedades ou características comuns no âmbito comercial ou científico. Essa separação feita com o S-Separator de um banco de dados em duas ou mais classes  $C$ / apresenta-se, grosso modo, como o prelúdio do Big Data.

$$X1(A), X2(A), X3(B)... // X1(B), X2(B), X3(B)...$$

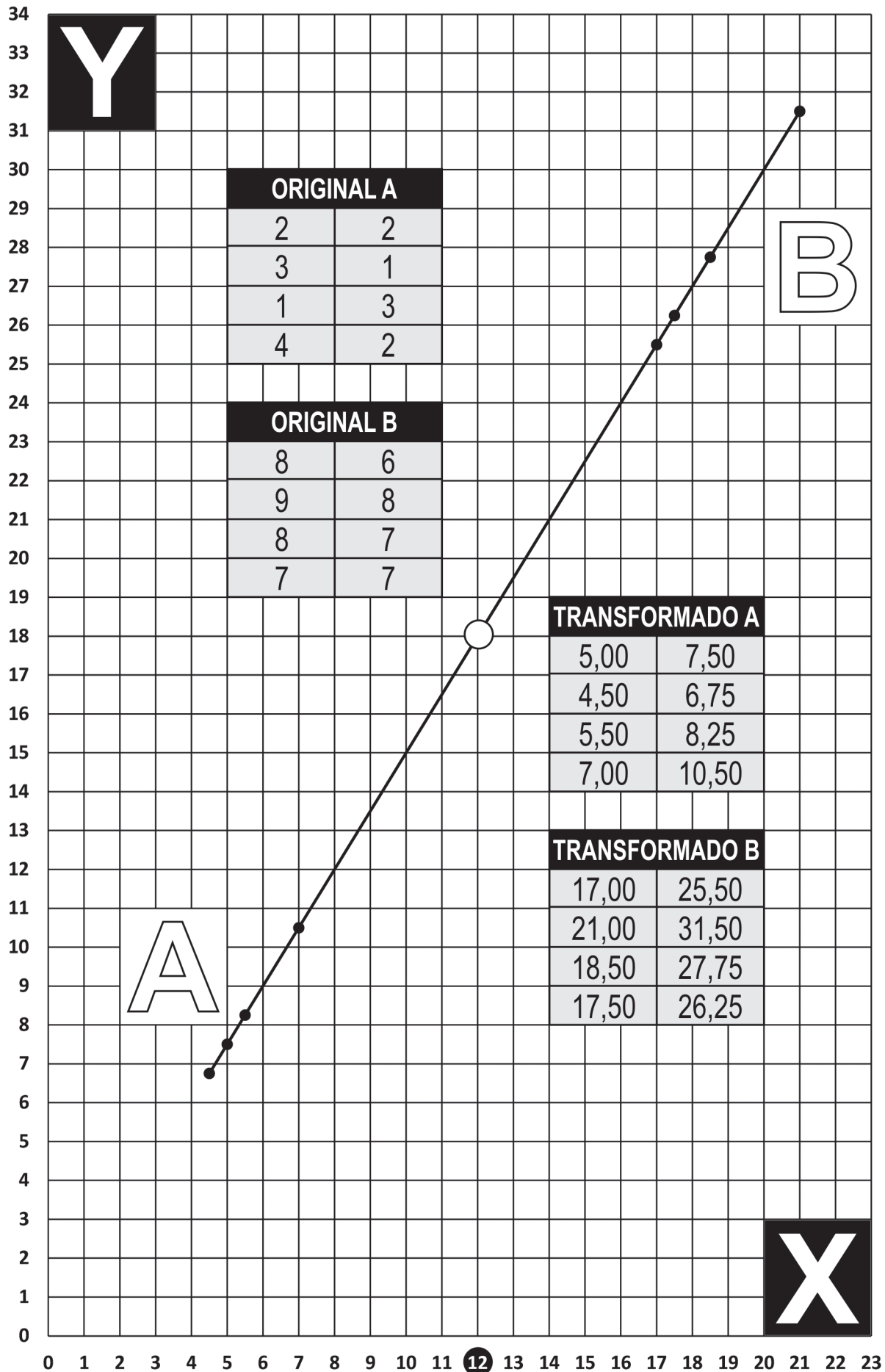
O modelo S-Separatio de aprendizagem supervisionada, com base na covariância matricial, tem por finalidade a translação dos dados iniciais para um espaço estatístico final, onde ocorre a separação das classes de dados topologicamente semelhantes, maximizando as distâncias externas entre as classes e minimizando as distâncias internas de cada uma das classes de dados como mostra o esquema:



Transformação Linear dos Dados no S-Separator

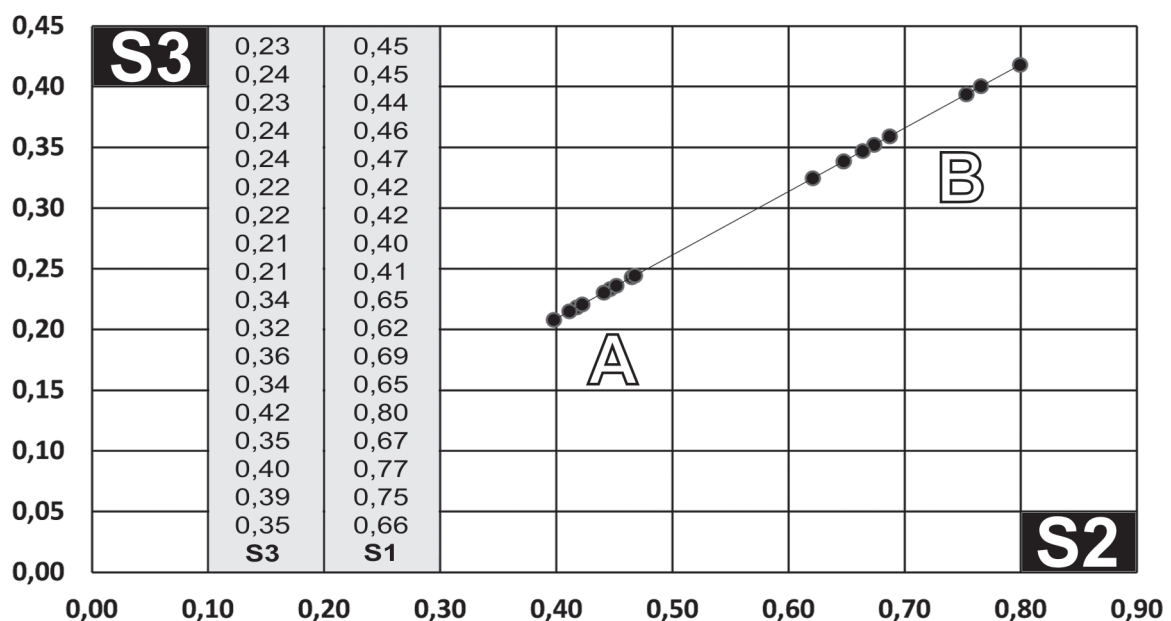
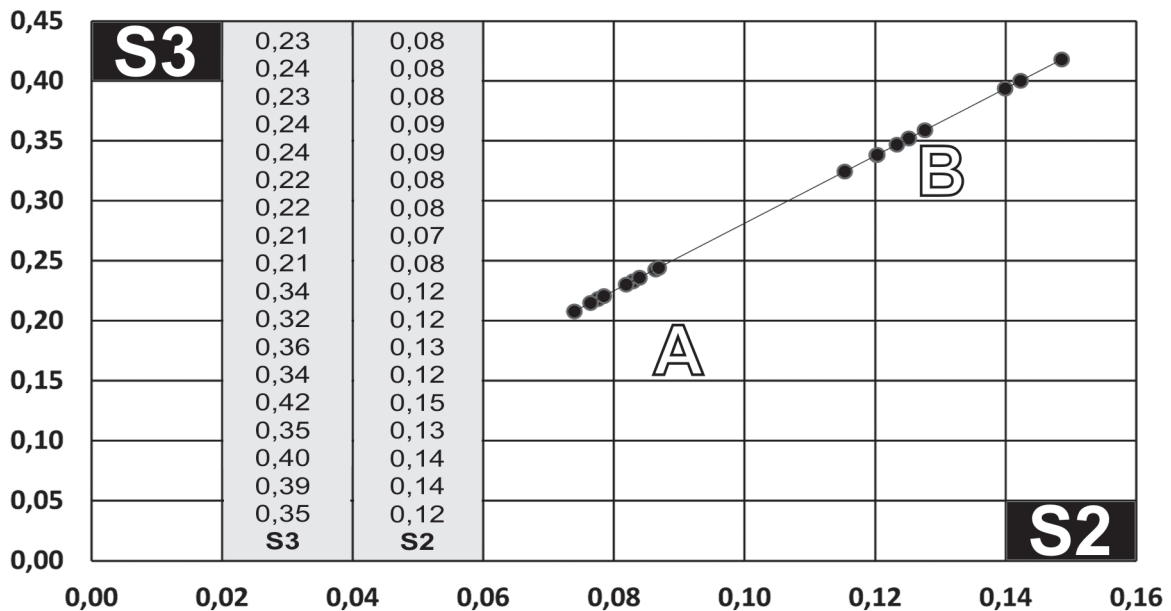
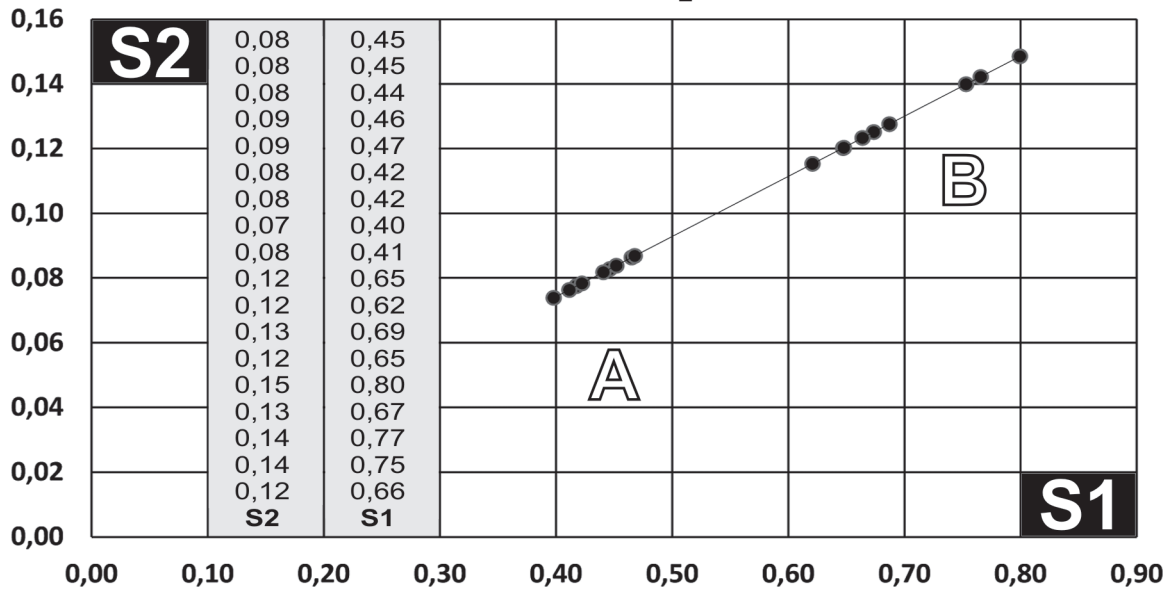


# 4.2 Transformada como Separação dos Dados





# 7.2 Transformada como Separação de Dados

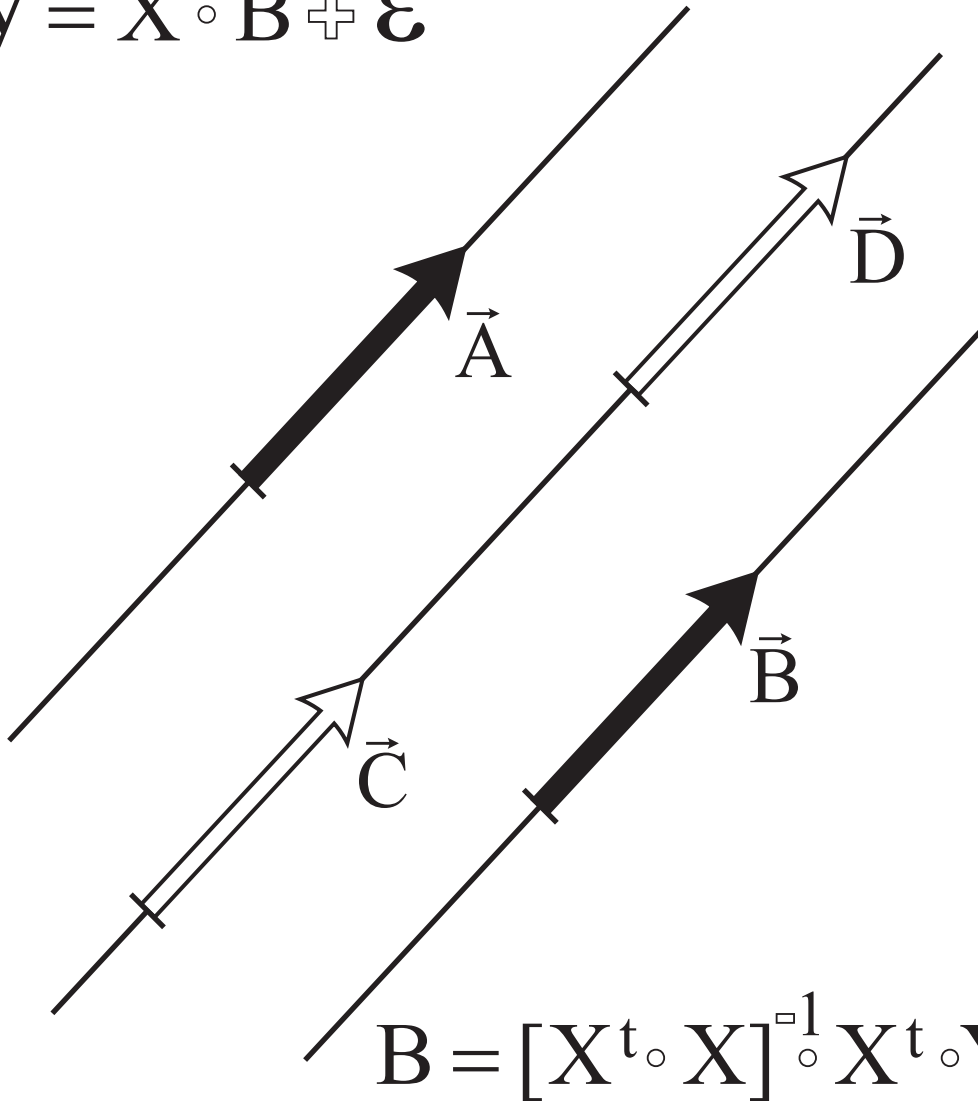




# C-COLLINEARITY

## Adjustment Function

$$\hat{y} = X \cdot B + \varepsilon$$



$$B = [X^t \cdot X]^{-1} \cdot X^t \cdot Y$$

Perfect Correlation

Incongruências dos Parâmetros  $b_i$

Prof. Aguinaldo Prandini Ricieri



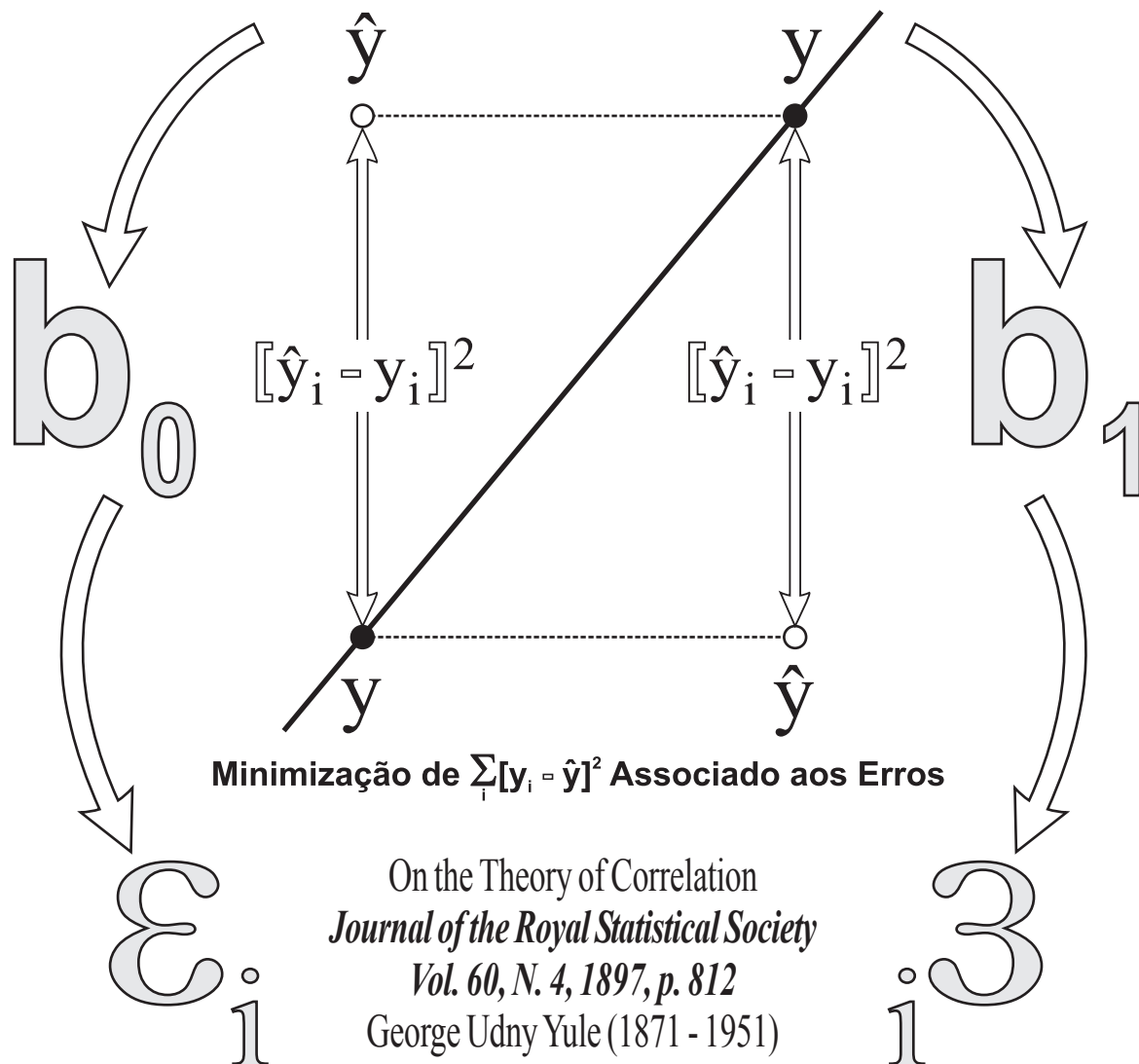


# 1.2 Variáveis Previstas e Variáveis Observadas

Um modelo de regressão de dados experimentados (termo cunhado pelo estatístico Francis Galton) projeta as variáveis estatísticas previstas (*predicted variables*  $X_1, X_2, \dots, X_n$ ) e as variáveis observadas (*observable variables*  $Y_1, Y_2, \dots, Y_n$ ) em um novo espaço vetorial ( $X'$  e  $Y'$ ). Com essa projeção estatística de ajuste procura-se a direção multidimensional nos espaços  $(X'_1, X'_2, \dots, X'_n)$ ,  $(\hat{Y}_1, \hat{Y}_2, \dots, \hat{Y}_n)$  que maximiza a variância o que possibilitará um ajuste perfeito dos dados.

Os modelos de regressão de dados usados hoje na estatística computacional tem seu início mais influente em 1897 no importante trabalho de George Udny Yule.

Como exemplo de regressão univariada ( $\hat{Y}_i = b_0 + b_1 X_i + \epsilon_i$ ) é necessário encontrar os parâmetros  $b_0$  e  $b_1$  que minimizam os erros de ajuste ( $\epsilon_i$ ) dos dados.



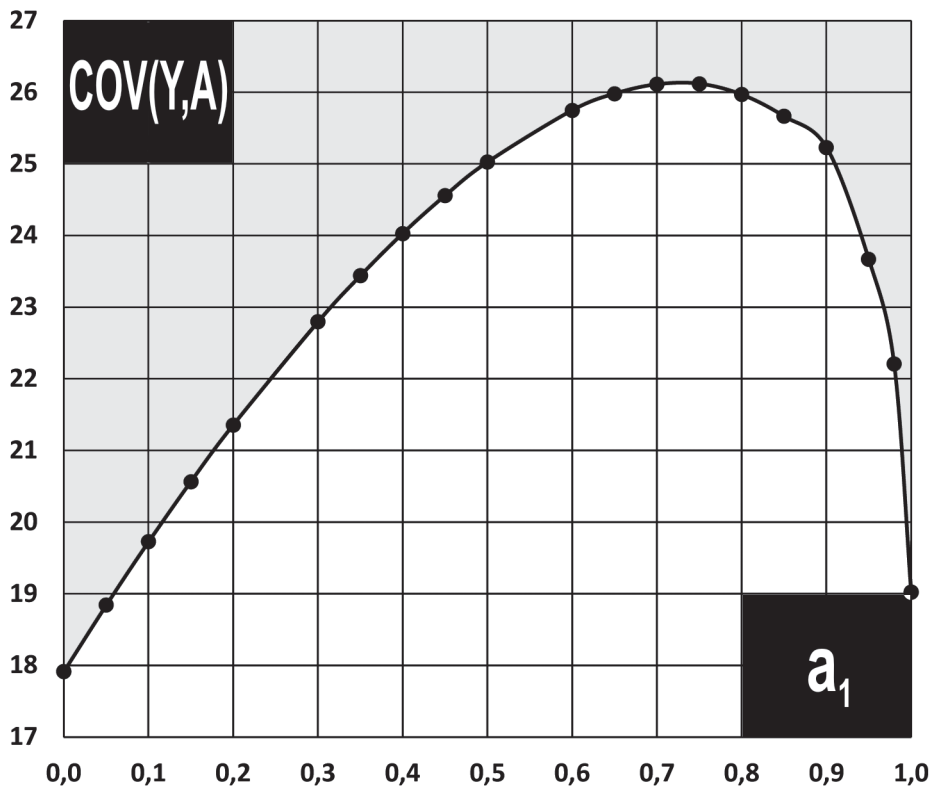


# 6.0 Coleção das Covariâncias em Termos de $a_1$ e $a_2$

**$a_1$**   
0,750

ITER	a1	a2	COV(Y, A)
01	0,000	1,000	17,917
02	0,050	0,999	18,846
03	0,100	0,995	19,730
04	0,150	0,989	20,568
05	0,200	0,980	21,360
06	0,300	0,954	22,800
07	0,350	0,937	23,443
08	0,400	0,917	24,032
09	0,450	0,893	24,563
10	0,500	0,866	25,030
11	0,600	0,800	25,750
12	0,650	0,760	25,984
13	0,700	0,714	26,115
14	0,750	0,661	26,122
15	0,800	0,600	25,972
16	0,850	0,527	25,867
17	0,900	0,436	25,832
18	0,950	0,312	23,671
19	0,980	0,199	22,213
20	1,000	0,000	19,028

**$a_2$**   
0,661



$A = a_1 \cdot x_1 + a_2 \cdot x_2$

X

$A = 0,750 \cdot x_1 + 0,661 \cdot x_2$

**A: Combinação Linear das Variáveis Predictoras  $X_1$  e  $X_2$**



# $\lambda$ -ORTOGONAL

## Transformação de Espaços

$$\begin{bmatrix} m_{11} & m_{12} & \circ\circ\circ & m_{n1} \\ m_{21} & m_{22} & \circ\circ\circ & m_{n2} \\ \vdots & \vdots & & \vdots \\ m_{n1} & m_{n2} & \circ\circ\circ & m_{nn} \end{bmatrix} \circ \begin{bmatrix} V_1 \\ V_2 \\ \vdots \\ V_n \end{bmatrix} = \begin{bmatrix} W_1 \\ W_2 \\ \vdots \\ W_n \end{bmatrix}$$

$$M, V, W, \lambda$$

$$M \circ V = W \quad \circ \quad M \circ V = \lambda \circ V$$

$$(M - \lambda I) \circ V = 0$$

$$\left( \begin{bmatrix} m_{11} & m_{12} & \circ\circ\circ & m_{n1} \\ m_{21} & m_{22} & \circ\circ\circ & m_{n2} \\ \vdots & \vdots & & \vdots \\ m_{n1} & m_{n2} & \circ\circ\circ & m_{nn} \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 & \circ\circ\circ & 0 \\ 0 & 1 & \circ\circ\circ & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \circ\circ\circ & 1 \end{bmatrix} \right) \circ \begin{bmatrix} V_1 \\ V_2 \\ \vdots \\ V_n \end{bmatrix} = 0$$

Análise Multivariada

Fenômenos Colunados

Prof. Aguinaldo Prandini Ricieri



## 2.1 Exemplificando os Conceitos Anunciados

### Exemplo Sectio Prima

$$M = \begin{bmatrix} 2 & 3 \\ 1 & 4 \end{bmatrix}$$

$$\begin{bmatrix} 2 & 3 \\ 1 & 4 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = 0$$

$$\det \begin{bmatrix} 2-\lambda & 3 \\ 1 & 4-\lambda \end{bmatrix} = 0 \quad \because \quad \begin{aligned} (2-\lambda) \cdot (4-\lambda) - 3 &= 0 \\ \lambda_1 &= +1 \text{ e } \lambda_2 = +5 \end{aligned}$$

$$\begin{bmatrix} 2 & 3 \\ 1 & 4 \end{bmatrix}^{-1} = \frac{1}{(2 \cdot 4 - 3 \cdot 1)} \cdot \begin{bmatrix} +4 & -3 \\ -1 & +2 \end{bmatrix} = \begin{bmatrix} +0,80 & -0,60 \\ -0,20 & +0,40 \end{bmatrix}$$

$$\text{tr}(M) = \sum_{i=1}^2 \lambda_i = +6$$

$$\det(M) = \prod_{i=1}^2 \lambda_i = +5$$

### Exemplo Sectio Secunda

$$M = \begin{bmatrix} 3 & 4 \\ 5 & 2 \end{bmatrix}$$

$$\begin{bmatrix} 3 & 4 \\ 5 & 2 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = 0$$

$$\det \begin{bmatrix} 3-\lambda & 4 \\ 5 & 2-\lambda \end{bmatrix} = 0 \quad \because \quad \begin{aligned} (3-\lambda) \cdot (2-\lambda) - 20 &= 0 \\ \lambda_1 &= +7 \text{ e } \lambda_2 = -2 \end{aligned}$$

$$\begin{bmatrix} 3 & 4 \\ 5 & 2 \end{bmatrix}^{-1} = \frac{1}{(3 \cdot 2 - 4 \cdot 5)} \cdot \begin{bmatrix} +2 & -4 \\ -5 & +3 \end{bmatrix} = \begin{bmatrix} -0,14 & +0,28 \\ +0,35 & -0,21 \end{bmatrix}$$

$$\text{tr}(M) = \sum_{i=1}^2 \lambda_i = +5$$

$$\det(M) = \prod_{i=1}^2 \lambda_i = -14$$

## Calculando para Matrizes 2 X 2



# 4.1 Análise da Matriz de Covariância dos Dados

## Cálculo dos Autovalores e Autovetores da Matriz M

$$M = \begin{bmatrix} 4,4 & 5,6 \\ 5,6 & 8,0 \end{bmatrix} \therefore \det[M - \lambda I] = 0 \therefore \det \begin{bmatrix} 4,4 & 5,6 \\ 5,6 & 8,0 \end{bmatrix} - \lambda \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} = 0$$

$$\det \begin{bmatrix} 4,4 - \lambda & 5,6 \\ 5,6 & 8,0 - \lambda \end{bmatrix} = 0 \Rightarrow \lambda^2 - 12,40\lambda + 3,84 = 0$$

$$M \circ V = \lambda \circ V \text{ para } \lambda_1 = 12,082$$

$$\begin{bmatrix} 4,4 & 5,6 \\ 5,6 & 8,0 \end{bmatrix} \circ \begin{bmatrix} x \\ y \end{bmatrix} = 12,082 \begin{bmatrix} x \\ y \end{bmatrix}$$

$$\begin{cases} 4,4x + 5,6y = 12,082x \\ 5,6x + 8,0y = 12,082x \end{cases} \Rightarrow \boxed{y = 1,371x} \Rightarrow \begin{bmatrix} x \\ 1,371x \end{bmatrix} \stackrel{N}{=} \begin{bmatrix} 0,580 \\ 0,795 \end{bmatrix}$$

$$M \circ V = \lambda \circ V \text{ para } \lambda_2 = 0,317$$

$$\begin{bmatrix} 4,4 & 5,6 \\ 5,6 & 8,0 \end{bmatrix} \circ \begin{bmatrix} x \\ y \end{bmatrix} = 0,317 \begin{bmatrix} x \\ y \end{bmatrix}$$

$$\begin{cases} 4,4x + 5,6y = 0,317x \\ 5,6x + 8,0y = 0,317x \end{cases} \Rightarrow \boxed{y = -1,371x} \Rightarrow \begin{bmatrix} -1,371y \\ y \end{bmatrix} \stackrel{N}{=} \begin{bmatrix} -0,795 \\ 0,580 \end{bmatrix}$$

$$V = \begin{bmatrix} 0,580 & -0,795 \\ 0,795 & 0,580 \end{bmatrix} \Rightarrow V1 = +0,580x + 0,795y \\ \Rightarrow V2 = -0,795x + 0,580y$$

Os vetores  $V1 = (0,580|0,795)$  e  $V2 = (-0,795|0,580)$  são ortogonais



# D-DIAGONAL

Autovalores (**L**) e (**V**) Autovetores

$$\begin{bmatrix} m_{11} & m_{12} & m_{13} & m_{14} & m_{15} & m_{16} \\ m_{21} & m_{22} & m_{23} & m_{24} & m_{25} & m_{26} \\ m_{31} & m_{32} & m_{33} & m_{34} & m_{35} & m_{36} \\ m_{41} & m_{42} & m_{43} & m_{44} & m_{45} & m_{46} \\ m_{51} & m_{52} & m_{53} & m_{54} & m_{55} & m_{56} \\ m_{61} & m_{62} & m_{63} & m_{64} & m_{65} & m_{66} \end{bmatrix}$$

M

$$\begin{aligned} M &= V \circ D \circ V^{-1} \\ D &= V^{-1} \circ M \circ V \end{aligned}$$

D

$$\begin{bmatrix} d_{11} & \circ & \circ & \circ & \circ & \circ \\ \circ & d_{22} & \circ & \circ & \circ & \circ \\ \circ & \circ & d_{33} & \circ & \circ & \circ \\ \circ & \circ & \circ & d_{44} & \circ & \circ \\ \circ & \circ & \circ & \circ & d_{55} & \circ \\ \circ & \circ & \circ & \circ & \circ & d_{66} \end{bmatrix}$$

Matrizes Diagonalizáveis

## Álgebra Linear

Prof. Aguinaldo Prandini Ricieri

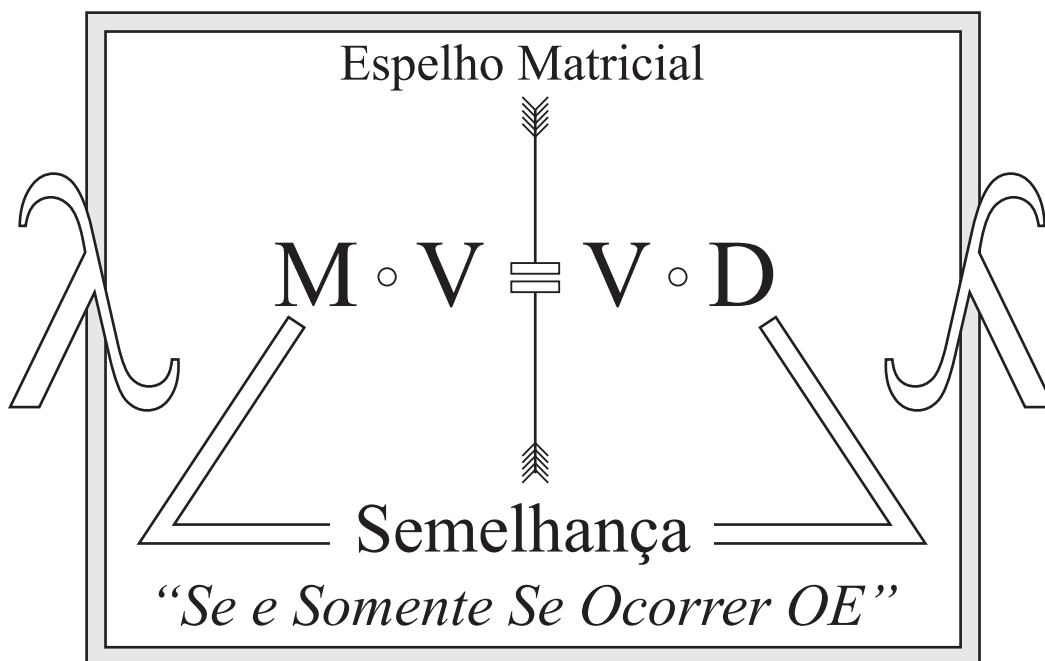




# Diagonalização

## 1.0 Semelhança Matricial Espelhada

Conhecidos os autovalores distintos e matriciados que associam os autovetores também matriciados da matriz quadrada  $M_{n \times n}$ , diz-se na álgebra linear que essa matriz é diagonalizável ou semelhante à matriz diagonal  $D$ , se e somente se, existir as duas matrizes  $V$  e  $V^{-1}$  que compõem a próxima operação espelhada (OE):



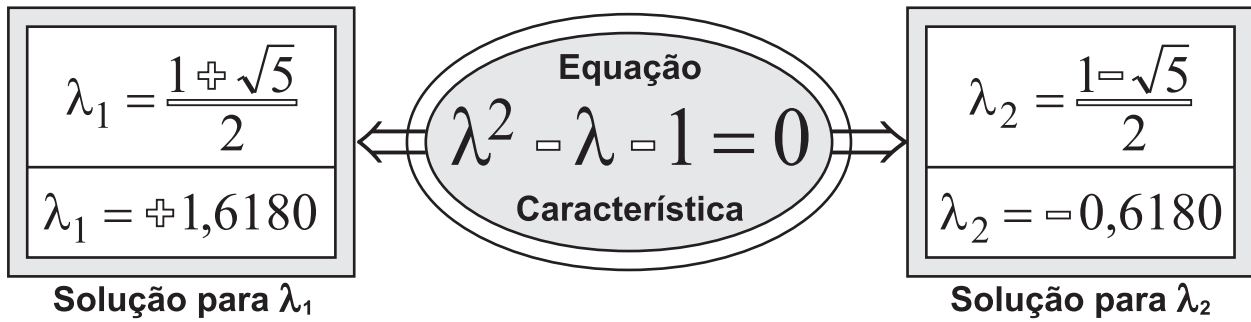
Além da importância técnica, a diagonalização de matrizes tem uma ampla gama de aplicações práticas nos diversos ramos do saber. Em física é aplicada na análise de sistemas dinâmicos lineares. Na estatística é empregada na análise de componentes principais. Em fim, à busca por autovalores e autovetores entendendo essa busca como as transformações de espaços vetoriais – tópicos que permitem a diagonalização de matrizes –, é fundamental para o desenvolvimento presente da matemática e particularmente para a estatística ou para a ciências de dados.

$$\boxed{D = V^{-1} \cdot M \cdot V} \quad \times \quad \boxed{M = V \cdot D \cdot V^{-1}}$$

**D** Associa Autovalores e **V** Associa Autovetores



## 3.1 Cálculo dos Autovalores



## 3.2 Cálculo dos Autovetores

$$\lambda_1 \circ \left( \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} - \begin{bmatrix} 1,61 & 0 \\ 0 & 1,61 \end{bmatrix} \right) \circ \begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} -1,61 & 1 \\ 1 & -1,61 \end{bmatrix} \circ \begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Rightarrow \begin{cases} X = 1,61Y \\ X = 1,61Y \end{cases}$$

$$\begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} 1,61Y \\ Y \end{bmatrix} = Y \begin{bmatrix} 1,61 \\ 1 \end{bmatrix} \Rightarrow V_1 = \begin{bmatrix} 1,61 \\ 1 \end{bmatrix}$$

$$\lambda_2 \circ \left( \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} - \begin{bmatrix} -0,61 & 0 \\ 0 & -0,61 \end{bmatrix} \right) \circ \begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} 1,61 & 1 \\ 1 & 0,61 \end{bmatrix} \circ \begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Rightarrow \begin{cases} X = -0,61Y \\ X = -0,61Y \end{cases}$$

$$\begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} -0,61Y \\ Y \end{bmatrix} = Y \begin{bmatrix} -0,61 \\ 1 \end{bmatrix} \Rightarrow V_2 = \begin{bmatrix} -0,61 \\ 1 \end{bmatrix}$$

$D = \begin{vmatrix} \frac{1 + \sqrt{5}}{2} & 0 \\ 0 & \frac{1 - \sqrt{5}}{2} \end{vmatrix}$	$V = \begin{vmatrix} \frac{1 + \sqrt{5}}{2} & \frac{1 - \sqrt{5}}{2} \\ 1 & 1 \end{vmatrix}$
--	--



# 4.0 Potenciação Sucessiva de Matrizes 2 X 2

6,00	4,00	≡	-2,00	1,00	○	4,00	0,00	○	-0,33	0,33
2,00	8,00		1,00	1,00		0,00	10,00		0,33	0,67

-8,00	10,00	○	-0,33	0,33
4,00	10,00		0,33	0,67

$$M^1 = V \circ D^1 \circ V^{-1}$$

6,00	4,00
2,00	8,00

44,00	56,00	≡	-2,00	1,00	○	16,00	0,00	○	-0,33	0,33
28,00	72,00		1,00	1,00		0,00	100,00		0,33	0,67

-32,00	100,00	○	-0,33	0,33
16,00	100,00		0,33	0,67

$$M^2 = V \circ D^2 \circ V^{-1}$$

44,00	56,00
28,00	72,00

376,00	624,00	≡	-2,00	1,00	○	64,00	0,00	○	-0,33	0,33
312,00	688,00		1,00	1,00		0,00	1000,00		0,33	0,67

-128,00	1000,00	○	-0,33	0,33
64,00	1000,00		0,33	0,67

$$M^3 = V \circ D^3 \circ V^{-1}$$

376,00	624,00
312,00	688,00

3504,00	6496,00	≡	-2,00	1,00	○	256,00	0,00	○	-0,33	0,33
3248,00	6752,00		1,00	1,00		0,00	10000,00		0,33	0,67

-512,00	10000,00	○	-0,33	0,33
256,00	10000,00		0,33	0,67

$$M^4 = V \circ D^4 \circ V^{-1}$$

3504,00	6496,00
3248,00	6752,00

3504,00	6496,00	≡	-2,00	1,00	○	1024,00	0,00	○	-0,33	0,33
3248,00	6752,00		1,00	1,00		0,00	100000,00		0,33	0,67

-2048,00	100000,00	○	-0,33	0,33
1024,00	100000,00		0,33	0,67

$$M^5 = V \circ D^5 \circ V^{-1}$$

34016,0	65984,0
32992,0	67008,0

3504,00	6496,00	≡	-2,00	1,00	○	4096,00	0,00	○	-0,33	0,33
3248,00	6752,00		1,00	1,00		0,00	1000000,00		0,33	0,67

-8192,00	1000000,00	○	-0,33	0,33
4096,00	1000000,00		0,33	0,67

$$M^6 = V \circ D^6 \circ V^{-1}$$

336064	663936
331968	668032



# $\bar{A}$ -INVERSA

## Matrizes Inversas Ampla

$$\begin{array}{c}
 I_{m \times n} \\
 \parallel \\
 A \circ \bar{A} \\
 \begin{array}{cc}
 m \times n & n \times m
 \end{array}
 \end{array}$$

$$A \circ X = B$$

$$X = \bar{A} \circ B + (I_n - A \circ \bar{A}) \circ W$$

$$\begin{array}{c}
 B_{m \times k} \\
 \parallel \\
 A \circ \bar{A} \circ B \\
 \begin{array}{ccc}
 m \times n & n \times m & m \times k
 \end{array}
 \end{array}$$

## Matrizes e Submatrizes

### Algoritmo de Searle

Prof. Aguinaldo Prandini Ricieri



## 2.0 Algoritmo da Matriz Inversa Ampla

a) Dado uma matriz  $A_{m \times n}$  de posto  $P$ , gerada por um sistema de equações, extrair dessa matriz  $A_{m \times n}$ , qualquer submatriz ( $a_n$ ) inversível de posto também  $P$ :

$$A_{m \times n} = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{vmatrix} \xleftrightarrow[\text{P}=2]{A} \begin{vmatrix} 5 & 3 & 1 \\ 7 & 1 & 3 \end{vmatrix} = A_{2 \times 3}$$

$$a_1 = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}_{2 \times 2} \quad a_2 = \begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix}_{2 \times 2} \quad a_3 = \begin{vmatrix} a_{11} & a_{13} \\ a_{21} & a_{23} \end{vmatrix}_{2 \times 2}$$

$$a_1 = \begin{vmatrix} 5 & 3 \\ 7 & 1 \end{vmatrix} \quad a_2 = \begin{vmatrix} 3 & 1 \\ 1 & 3 \end{vmatrix} \quad a_3 = \begin{vmatrix} 5 & 1 \\ 7 & 3 \end{vmatrix}$$

$$a_1^{-1} = \begin{vmatrix} -\frac{1}{16} & \frac{3}{16} \\ \frac{7}{16} & -\frac{5}{16} \end{vmatrix} \quad a_2^{-1} = \begin{vmatrix} \frac{3}{8} & -\frac{1}{8} \\ -\frac{1}{8} & \frac{3}{8} \end{vmatrix} \quad a_3^{-1} = \begin{vmatrix} \frac{3}{8} & -\frac{7}{8} \\ -\frac{1}{8} & \frac{5}{8} \end{vmatrix}$$

$$(a_1^{-1})^t = \begin{vmatrix} -\frac{1}{16} & \frac{7}{16} \\ \frac{3}{16} & -\frac{5}{16} \end{vmatrix}_{2 \times 2} \quad (a_2^{-1})^t = \begin{vmatrix} \frac{3}{8} & -\frac{1}{8} \\ -\frac{1}{8} & \frac{3}{8} \end{vmatrix}_{2 \times 2} \quad (a_3^{-1})^t = \begin{vmatrix} \frac{3}{8} & -\frac{1}{8} \\ -\frac{7}{8} & \frac{5}{8} \end{vmatrix}_{2 \times 2}$$

b) Leve as matrizes inversas transpostas para suas respectivas posições na matriz  $A_{m \times n}$ , zerando os elementos restantes de  $A_{m \times n}$  e transpondo a matriz resultante:

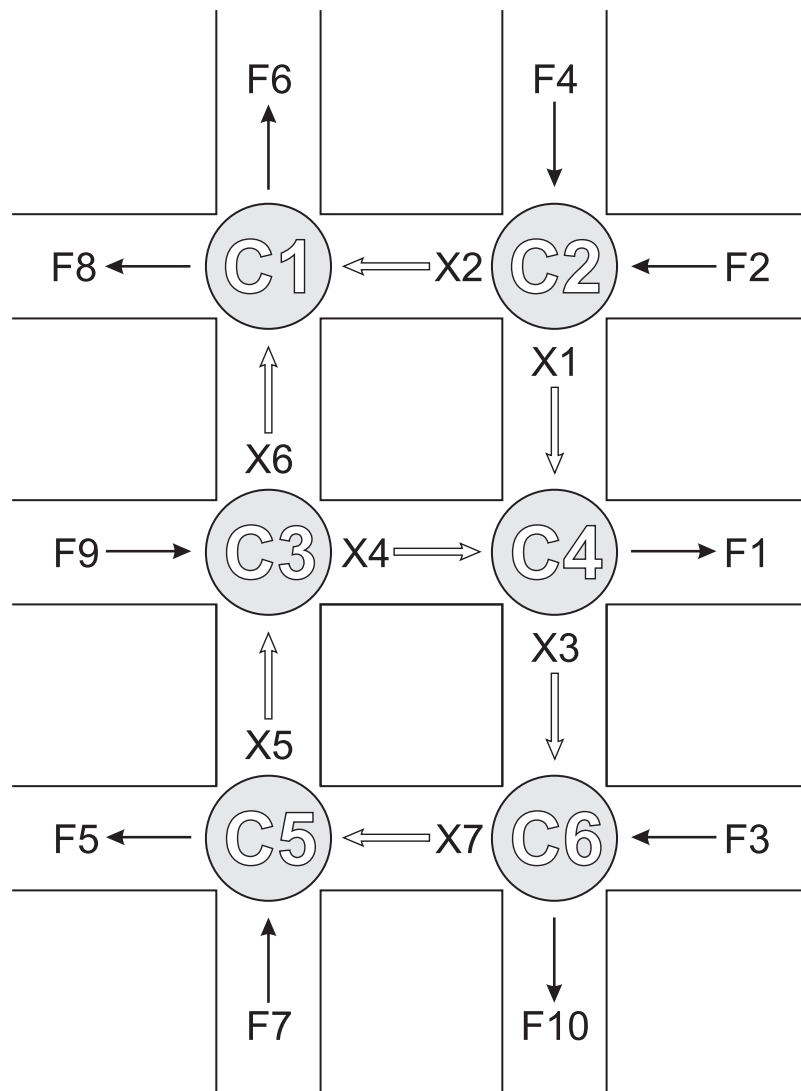
$$\begin{vmatrix} -\frac{1}{16} & \frac{3}{16} & 0 \\ \frac{7}{16} & -\frac{5}{16} & 0 \\ 0 & 0 & 0 \end{vmatrix} \quad \begin{vmatrix} 0 & \frac{3}{8} & -\frac{1}{8} \\ 0 & -\frac{1}{8} & \frac{3}{8} \\ 0 & 0 & 0 \end{vmatrix} \quad \begin{vmatrix} \frac{3}{8} & 0 & -\frac{7}{8} \\ -\frac{1}{8} & 0 & \frac{5}{8} \\ 0 & 0 & 0 \end{vmatrix}$$

$$\bar{A}_1 = \begin{vmatrix} -\frac{1}{16} & \frac{3}{16} \\ \frac{7}{16} & -\frac{5}{16} \\ 0 & 0 \end{vmatrix}_{3 \times 2} \quad \bar{A}_2 = \begin{vmatrix} 0 & 0 \\ \frac{3}{8} & -\frac{1}{8} \\ -\frac{1}{8} & \frac{3}{8} \end{vmatrix}_{3 \times 2} \quad \bar{A}_3 = \begin{vmatrix} \frac{3}{8} & -\frac{1}{8} \\ 0 & 0 \\ -\frac{7}{8} & \frac{5}{8} \end{vmatrix}_{3 \times 2}$$

$\bar{A}_1, \bar{A}_2, \bar{A}_3$ : Matrizes Inversas Amplas



## 4.2 Extensão do Modelo de Fluxo de Coisas



**Sextivium e seus Fluxos**

$$X2 + X6 = F6 + F8 \quad (C1)$$

$$X1 + X2 = F2 + F4 \quad (C2)$$

$$X5 + F9 = X4 + X6 \quad (C3)$$

$$X1 + X4 = F1 + X3 \quad (C4)$$

$$X7 + F7 = X5 + F5 \quad (C5)$$

$$X3 + F3 = F10 + X7 \quad (C6)$$

**Sistema de Equações Pertinentes ao Fluxo**





# $\theta$ -Likelihood

## Linear Regression

$$\mathbf{X} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1p} \\ 1 & x_{21} & \cdots & x_{2p} \\ \vdots & & & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{bmatrix} \quad \hat{\mathbf{y}} = \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_n \end{bmatrix} \quad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

$$\mathbf{P} = \prod_{i=1}^n f(y_i, x_i, \mathbf{b}, \sigma^2)$$

$$\ln L = -\frac{n}{2} \cdot \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} (y_i - \mathbf{b}^t \cdot \mathbf{x}_i)^2$$

$$\mathbf{b} = (\mathbf{X}^t \cdot \mathbf{X})^{-1} \cdot \mathbf{X}^t \cdot \mathbf{y}$$

Probability Distribution  
Model for Predicting

Prof. Aguinaldo Prandini Ricieri



# Verossimilhança

## 1.0 Likelihood e Algumas Particularidades

O método da verossimilhança que formata parâmetros de ajustes de modelos de previsão de fenômenos – maximizando o produto de funções de probabilidades –, permite uma relação ótima entre as variáveis independentes e dependentes.

Essa técnica estatística inicia sua modelagem com uma função de likelihood que representa a probabilidade de ocorrerem os dados registrados dado um conjunto de parâmetros otimizadores de ajuste ou de acerto. No caso de uma função linear  $\hat{y} = b_0 + b_1x_1 + b_2x_2 + \dots + \varepsilon$ , implica partir do presuposto da existência de uma distribuição probabilística  $(\sigma^2, \bar{x})$  para os erros  $|\hat{y} - y|$  residuais entre os valores reais (valores mensurados  $y$ ) e os valores previstos (valores gerados pela função  $\hat{y}$ ). Quanto aos coeficientes de ajustes  $b_0, b_1, \dots, b_p$ , esses são encontrados alterando-os de modo a maximizarem a probabilidade ( $P$ ) de ocorrência dos dados.

$$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2} \cdot \left(\frac{x-\mu}{\sigma}\right)^2} \times \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2} \cdot \left(\frac{y_i - b^t \cdot x_i}{\sigma}\right)^2}$$

Funções de Densidades de Probabilidades

$$P = \prod_{i=1}^n f(y_i, x_i, b, \sigma^2)$$

$$b_0 \text{ e } b_1$$

$$\ln L = -\frac{n}{2} \cdot \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} (y_i - b_0 - b_1 \cdot x_i)^2$$

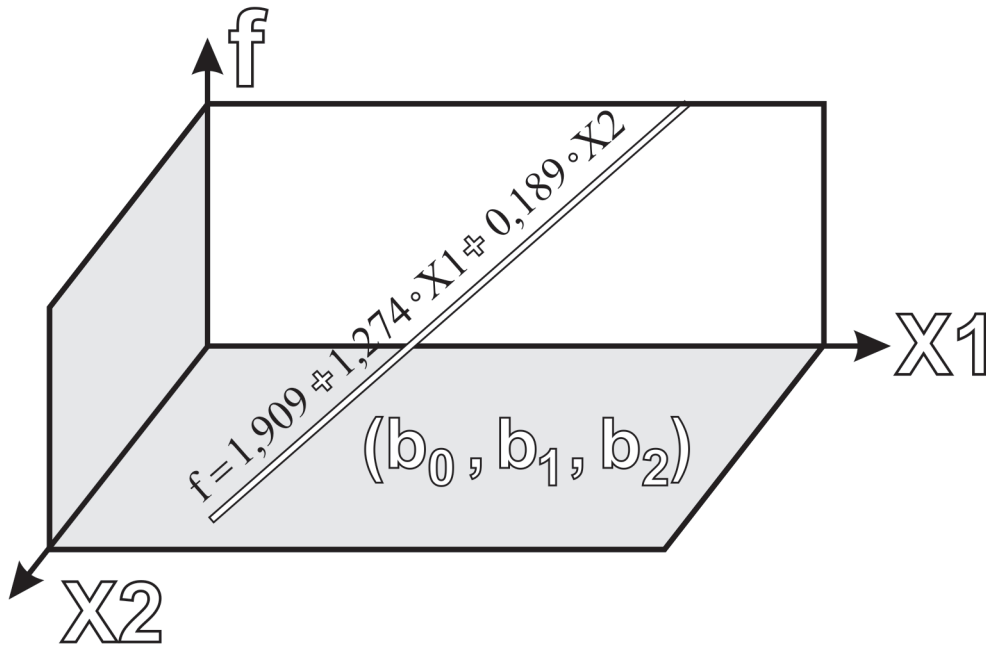
Maximizar  $\ln L$  em Termos dos Coeficientes de Ajustes  $b_0$  e  $b_1$



# 6.0 Resultados Likelihood ( $b_0, b_1, b_2$ ) que Maximizam $\ln(L)$

ITERA	b0	b1	b2	Ln(L)
1	1,5000	1,7000	1,1200	-47,3442
2	1,7322	0,1844	0,2621	-45,2365
3	1,8342	0,2731	1,4250	-32,6694
4	1,7021	1,4318	0,9750	-43,1640
5	1,9096	1,2742	0,1891	-25,8329
6	2,0354	0,9522	0,8342	-32,7880
7	0,5021	1,3245	0,9890	-40,2658
8	1,6050	0,9589	0,3241	-30,3773
9	0,9944	1,2712	0,8546	-37,9020
10	0,4522	1,8290	0,3544	-38,5620
11	1,8990	1,2240	0,1790	-26,3079
12	1,9394	1,2952	0,1984	-25,9712

**DOZE ITERAÇÕES COM TUCANOS**



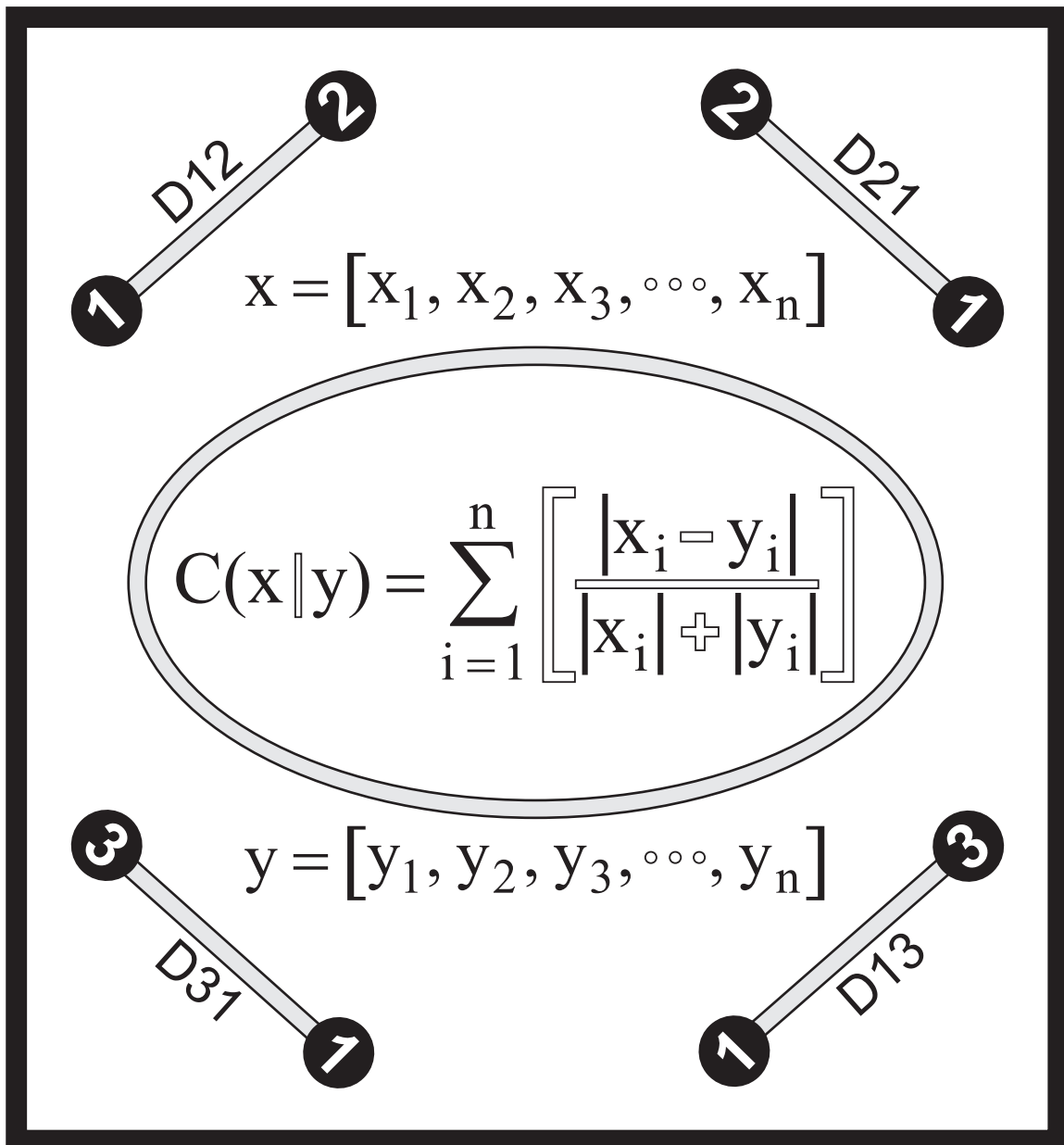
ITERA	b0	b1	b2	Ln(L)
1	1,5000	1,7000	1,1200	-47,3442
2	1,7322	0,1844	0,2621	-45,2365
4	1,7021	1,4318	0,9750	-43,1640
7	0,5021	1,3245	0,9890	-40,2658
10	0,4522	1,8290	0,3544	-38,5620
9	0,9944	1,2712	0,8546	-37,9020
6	2,0354	0,9522	0,8342	-32,7880
3	1,8342	0,2731	1,4250	-32,6694
8	1,6050	0,9589	0,3241	-30,3773
11	1,8990	1,2240	0,1790	-26,3079
12	1,9394	1,2952	0,1984	-25,9712
5	1,9096	1,2742	0,1891	-25,8329

**RESULTADOS CLASSIFICADOS MIN-MAX**



# C-CAMBERRA

## Distance for Classification



## Taxonomia Numérica

# Medida de Similaridade

Prof. Aguinaldo Prandini Ricieri



# C=CAMBERRA

## 1.0 Taxonomia Numérica e sua Distância

Trata-se de uma importante e singular medida de distância entre pares de pontos em um espaço matricial (1 x n). Proposta inicialmente em 1966 pela dupla de matemáticos australianos Willians e Lance, (Camberra é a capital da Austrália), tornou-se uma técnica indispensável na estatística de classificação de dados.

A distância de Camberra  $C$  entre os vetores de dados registrados  $x$  e  $y$ , definida em um espaço  $n$  dimensional é escrita na forma representativa e compacta:

$$\begin{array}{c}
 \mathbf{x} = [x_1, x_2, x_3, \dots, x_n] \\
 \mathbf{y} = [y_1, y_2, y_3, \dots, y_n] \\
 C(\mathbf{x}||\mathbf{y}) = \sum_{i=1}^n \left[ \frac{|x_i - y_i|}{|x_i| + |y_i|} \right]
 \end{array}$$

$C(\mathbf{x}||\mathbf{y})$ : Distância de Camberra e Pertinências

$C(\mathbf{x}||\mathbf{y})$  tem a função de quantificar o grau de similaridade (proximidade) entre  $x$  e  $y$ .

D

- 1 Computer Programs for Hierarchical Polythetic Classification, Similarity Analysis  
*Computer Journal*, 9(1), p. 60, 1966  
William T. Willians and Godfrey N. Lance

Σ

- 2 Camberra Distance on Ranked Lists  
*Procedings, Advances in Ranking*, p. 22, 2009

Giuseppe Jurman, Cesare Furlanello, Samantha Riccadonna, Roberto Visintainer



# 5.0 Estatística das 100 Intensidades de Absorções de Luz - A

i 11	i 12	i 13	i 14	i 15	SUB	i 21	i 22	i 23	i 24	i 25	SUB	i 31	i 32	i 33	i 34	i 35
400	450	500	550	600	nm	400	450	500	550	600	nm	400	450	500	550	600
70,2	28,6	48,1	28,6	67,6	01	43,2	43,2	17,6	41,9	68,9	01	15,4	89,6	44,8	60,2	21,0
71,5	19,5	42,9	22,1	65,0	02	54,0	41,9	23,0	40,5	72,9	02	16,8	44,8	44,8	65,8	14,0
71,5	27,3	45,5	23,4	68,9	03	43,2	40,5	25,7	47,3	68,9	03	18,2	89,6	46,2	60,2	22,4
66,3	24,7	45,5	19,5	76,7	04	41,9	40,5	13,5	50,0	72,9	04	16,8	77,0	43,4	70,0	22,4
76,7	24,7	39,0	22,1	70,2	05	45,9	43,2	25,7	50,0	72,9	05	14,0	81,2	46,2	65,8	19,6
70,2	22,1	40,3	20,8	71,5	06	47,3	41,9	23,0	48,6	67,5	06	19,6	82,6	43,4	71,4	22,4
78,0	18,2	45,5	28,6	75,4	07	45,9	41,9	24,3	45,9	72,9	07	16,8	44,8	44,8	65,8	22,4
78,0	28,6	45,5	32,5	75,4	08	48,6	44,6	18,9	48,6	77,0	08	15,4	84,0	43,4	65,8	18,2
67,6	26,0	39,0	26,0	71,5	09	41,9	47,3	25,7	45,9	70,2	09	15,4	91,0	46,2	70,0	14,0
70,2	23,4	46,8	31,2	68,9	10	54,0	43,2	13,5	47,3	72,9	10	18,2	89,6	47,6	64,4	21,0
75,4	20,8	46,8	19,5	72,8	11	40,5	47,3	18,9	51,3	81,0	11	19,6	81,2	47,6	64,4	18,2
71,5	26,0	40,3	20,8	76,7	12	41,9	47,3	18,9	52,7	75,6	12	18,2	89,6	43,4	68,6	16,8
66,3	23,4	39,0	27,3	72,8	13	48,6	43,2	18,9	44,6	72,9	13	19,6	44,8	47,6	71,4	16,8
67,6	27,3	44,2	32,5	72,8	14	44,6	40,5	18,9	40,5	67,5	14	15,4	81,2	44,8	65,8	15,4
72,8	28,6	49,4	19,5	70,2	15	41,9	43,2	21,6	40,5	72,9	15	15,4	78,4	44,8	72,8	19,6
66,3	27,3	44,2	32,5	71,5	16	48,6	43,2	14,9	54,0	71,6	16	21,0	85,4	42,0	71,4	21,0
65,0	15,6	46,8	19,5	72,8	17	44,6	47,3	16,2	48,6	77,0	17	14,0	77,0	42,0	65,8	19,6
67,6	16,9	41,6	23,4	65,0	18	40,5	45,9	20,3	51,3	72,9	18	18,2	81,2	42,0	65,8	15,4
65,0	24,7	49,4	20,8	65,0	19	50,0	41,9	21,6	48,6	75,6	19	18,2	84,0	42,0	68,6	18,2
74,1	22,1	41,6	26,0	75,4	20	43,2	45,9	16,2	47,3	78,3	20	21,0	82,6	44,8	64,4	16,8
78,0	15,6	46,8	28,6	72,8	21	41,9	41,9	23,0	40,5	79,7	21	19,6	81,2	46,2	64,4	19,6
72,8	24,7	41,6	22,1	76,7	22	54,0	47,3	14,9	45,9	72,9	22	16,8	82,6	44,8	71,4	16,8
70,2	18,2	41,6	27,3	72,8	23	47,3	47,3	16,2	48,6	72,9	23	16,8	79,8	47,6	65,8	16,8
65,0	16,9	45,5	31,2	68,9	24	50,0	43,2	25,7	47,3	74,3	24	15,4	81,2	47,6	65,8	19,6
75,4	18,2	48,1	29,9	71,5	25	54,0	47,3	27,0	50,0	72,9	25	21,0	88,2	44,8	67,2	19,6
70,2	16,9	44,2	24,7	72,8	26	54,0	45,9	20,3	44,6	77,0	26	19,6	84,0	42,0	58,8	14,0
76,7	23,4	46,8	22,1	75,4	27	48,6	43,2	16,2	52,7	67,5	27	15,4	81,2	43,4	67,2	21,0
72,8	16,9	42,9	23,4	78,0	28	54,0	45,9	23,0	50,0	79,7	28	15,4	44,8	47,6	58,8	21,0
65,0	23,4	45,5	22,1	65,0	29	44,6	41,9	24,3	51,3	77,0	29	18,2	78,4	44,8	63,0	19,6
70,2	19,5	44,2	23,4	78,0	30	45,9	40,5	18,9	43,2	74,3	30	21,0	79,8	42,0	67,2	16,8
72,8	23,4	44,2	28,6	75,4	31	41,9	45,9	18,9	43,2	72,9	31	21,0	78,4	46,2	72,8	16,8
76,7	16,9	40,3	22,1	70,2	32	54,0	47,3	23,0	43,2	68,9	32	16,8	86,8	43,4	60,2	15,4
68,9	23,4	45,5	22,1	68,9	33	41,9	45,9	27,0	44,6	74,3	33	19,6	77,0	47,6	65,8	15,4
75,4	23,4	48,1	26,0	65,0	34	40,5	40,5	23,0	47,3	74,3	34	16,8	84,0	46,2	65,8	15,4
70,2	26,0	49,4	31,2	68,9	35	54,0	44,6	14,9	47,3	74,3	35	14,0	84,0	43,4	70,0	22,4
76,7	20,8	48,1	20,8	66,3	36	45,9	44,6	24,3	44,6	68,9	36	14,0	89,6	42,0	65,8	15,4
65,0	22,1	41,6	19,5	71,5	37	54,0	45,9	16,2	29,7	72,9	37	19,6	84,0	42,0	68,6	19,6
76,7	15,6	39,0	49,4	74,1	38	45,9	45,9	13,5	51,3	74,3	38	15,4	84,0	46,2	70,0	18,2
72,8	19,5	39,0	27,3	66,3	39	54,0	41,9	14,9	54,0	79,7	39	15,4	85,4	42,0	67,2	15,4
71,5	24,7	46,8	20,8	78,0	40	54,0	44,6	24,3	54,0	72,9	40	21,0	82,6	43,4	60,2	21,0
67,6	28,6	44,2	26,0	72,8	41	51,3	41,9	23,0	52,7	67,5	41	14,0	79,8	46,2	61,6	19,6
65,0	28,6	40,3	20,8	74,1	42	41,9	44,6	13,5	41,9	77,0	42	15,4	86,8	47,6	65,8	16,8
45,5	24,7	42,9	28,6	65,0	43	54,0	43,2	17,6	48,6	81,0	43	21,0	77,0	42,0	60,2	15,4
72,8	24,7	41,6	49,4	71,5	44	51,3	47,3	17,6	41,9	72,9	44	16,8	84,0	43,4	71,4	18,2
75,4	18,2	42,9	28,6	74,1	45	50,0	44,6	14,9	47,3	75,6	45	18,2	89,6	44,8	72,8	16,8
68,9	23,4	39,0	20,8	70,2	46	40,5	44,6	13,5	52,7	70,2	46	16,8	79,8	43,4	60,2	15,4
78,0	16,9	44,2	26,0	65,0	47	52,7	45,9	18,9	44,6	74,3	47	19,6	44,8	46,2	67,2	22,4
78,0	24,7	40,3	23,4	67,6	48	44,6	40,5	25,7	50,0	81,0	48	19,6	85,4	43,4	68,6	19,6
68,9	24,7	49,4	49,4	67,6	49	47,3	41,9	24,3	51,3	79,7	49	18,2	81,2	46,2	60,2	19,6
72,8	18,2	45,5	27,3	76,7	50	43,2	45,9	23,0	29,7	74,3	50	16,8	85,4	46,2	65,8	18,2

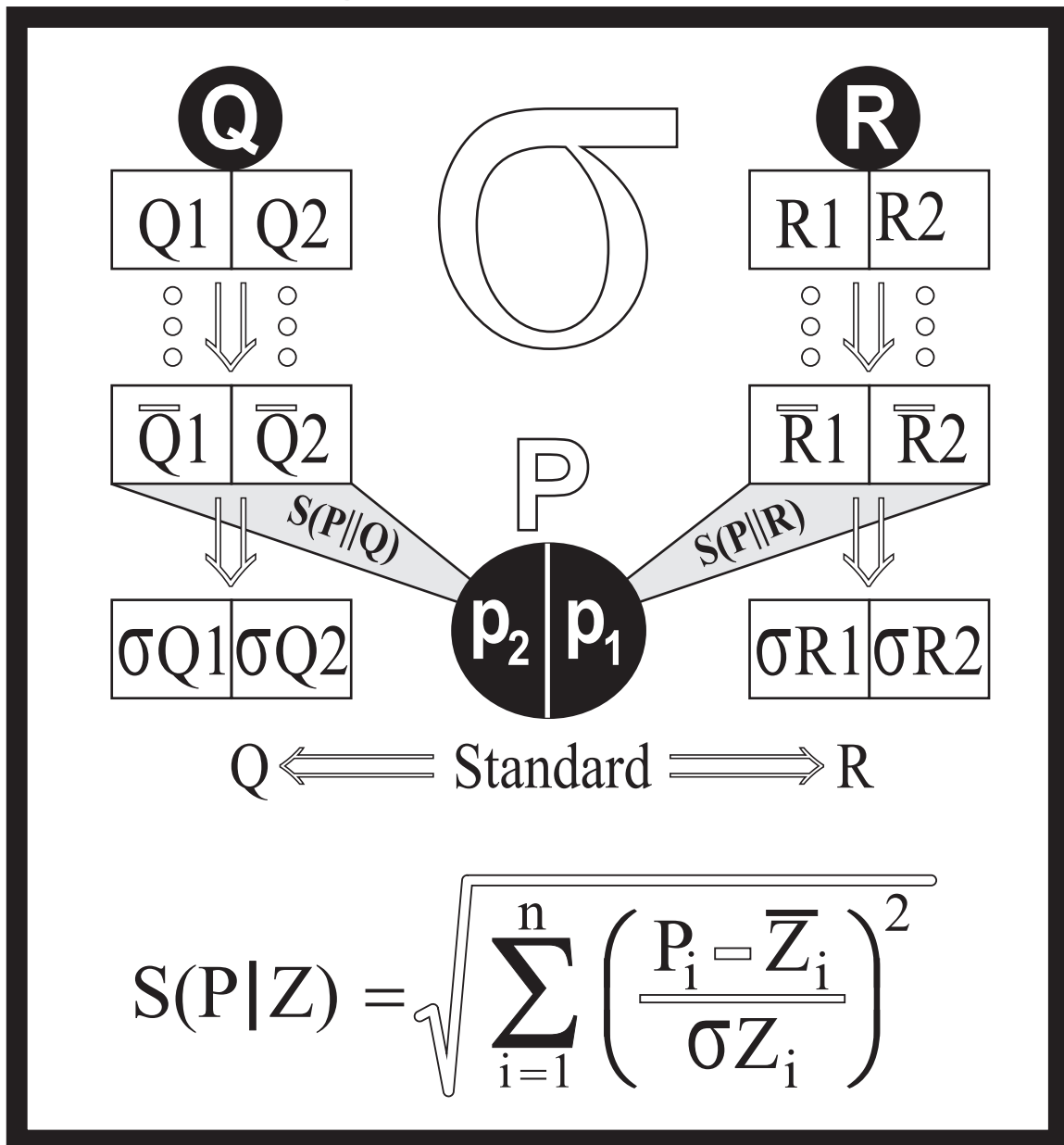
## Substâncias Químicas Clusterizadas em 3 Classes





# S-STANDARD

## Pictogram Separation



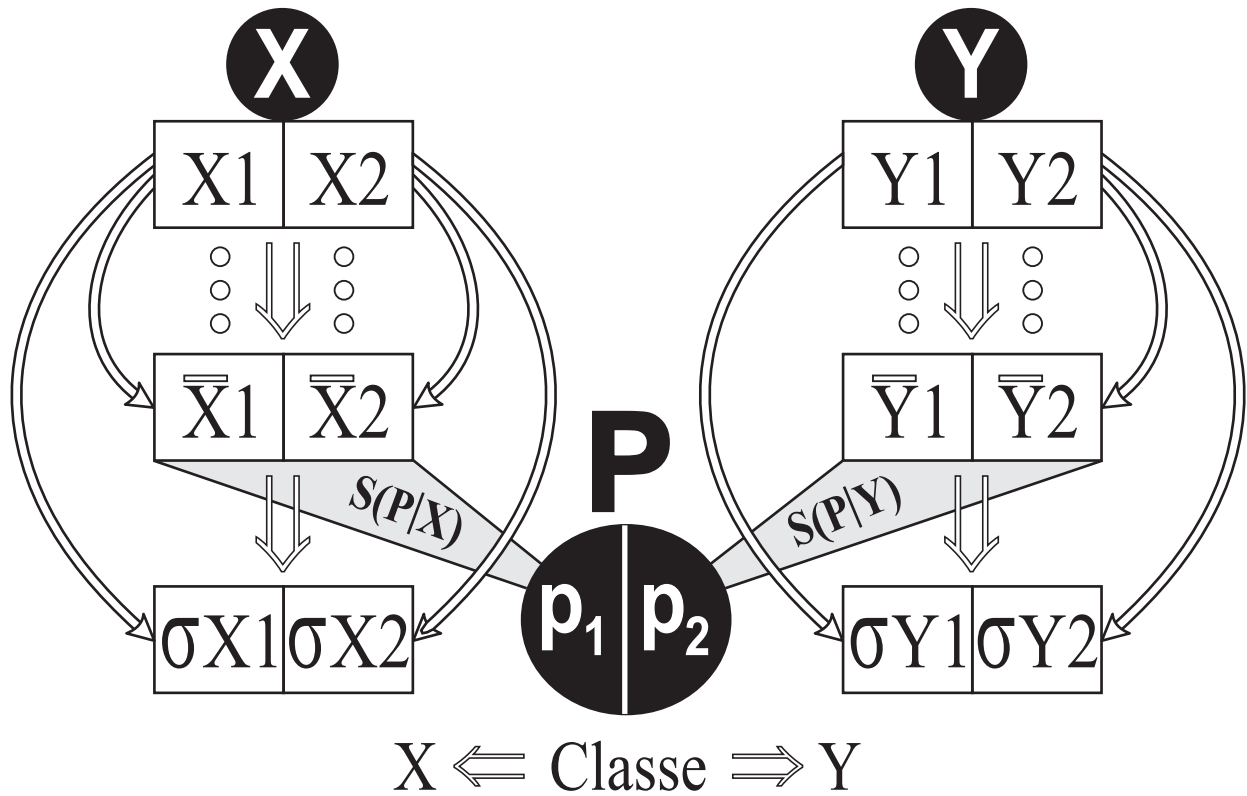
## Standard Deviation

## Agrupamento de Dados

Prof. Aguinaldo Prandini Ricieri



## 2.0 Apresentação da Distância Standard (S)



$$S(P|X) = \sqrt{\left(\frac{p_1 - \bar{X}_1}{\sigma_{X1}}\right)^2 + \left(\frac{p_2 - \bar{X}_2}{\sigma_{X2}}\right)^2}$$

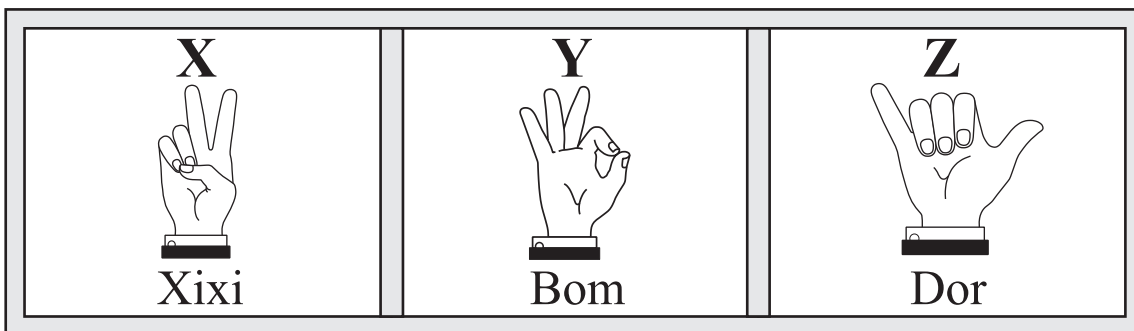
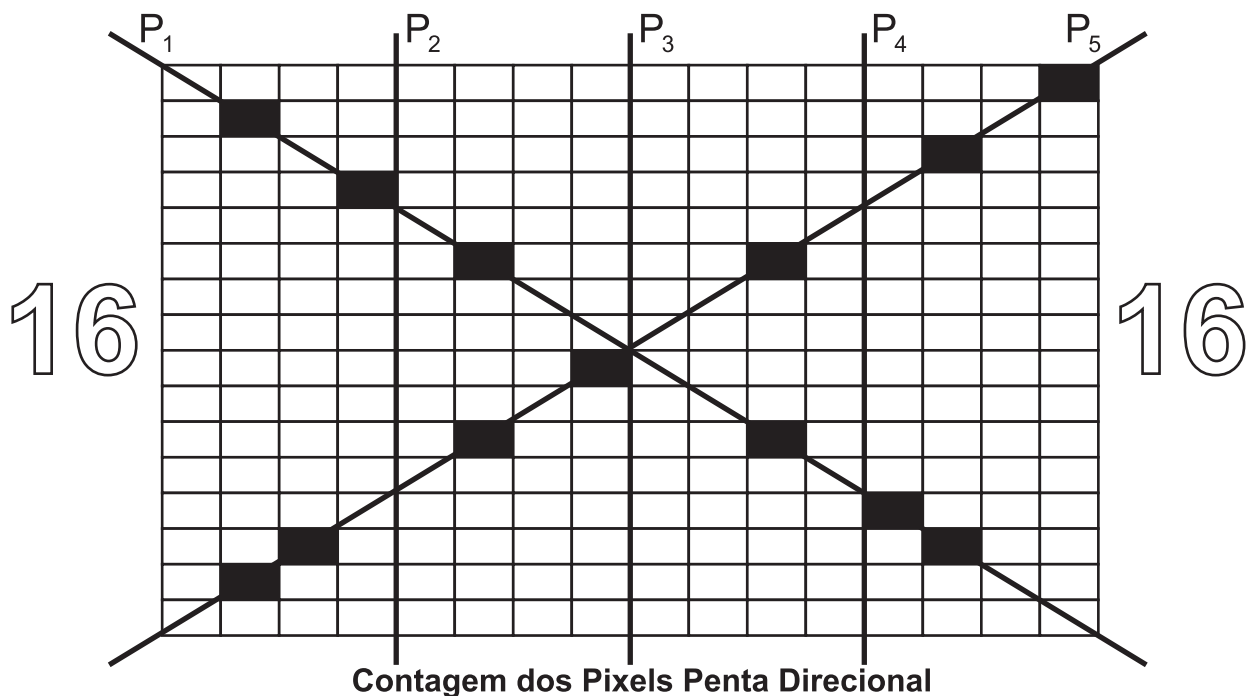
$$S(P|Y) = \sqrt{\left(\frac{p_1 - \bar{Y}_1}{\sigma_{Y1}}\right)^2 + \left(\frac{p_2 - \bar{Y}_2}{\sigma_{Y2}}\right)^2}$$

$$S(P|Z) = \sqrt{\sum_{i=1}^n \left(\frac{p_i - \bar{Z}_i}{\sigma_{Zi}}\right)^2}$$

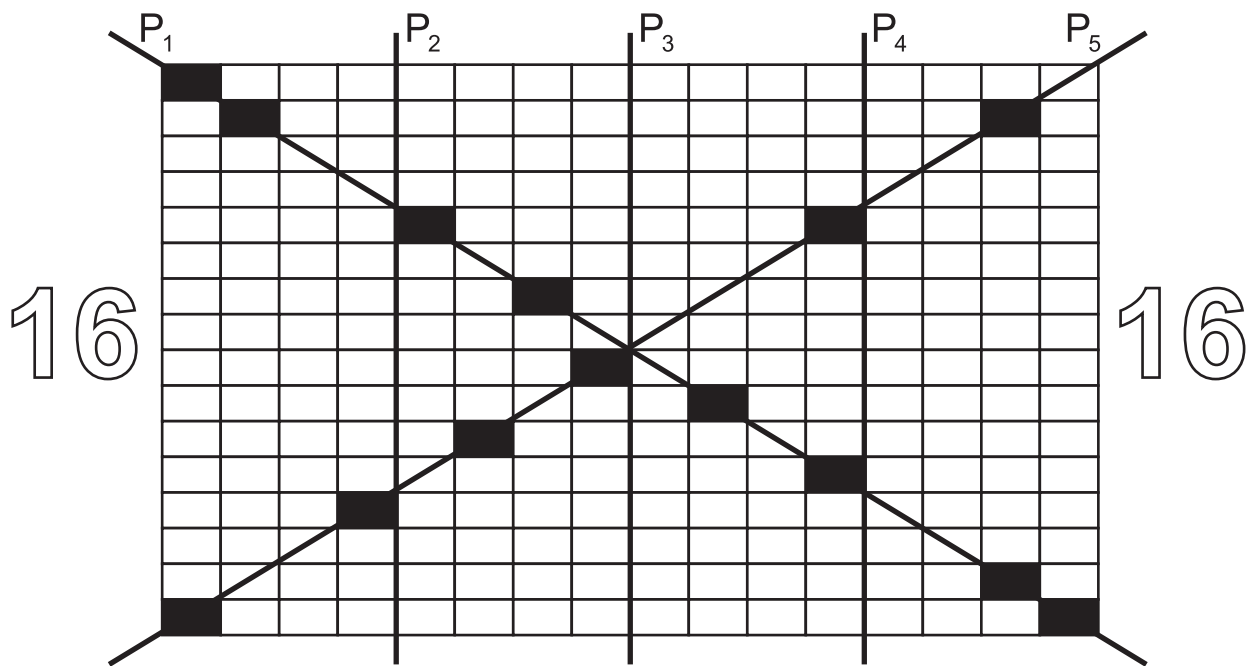
**Distância de Standard Generalizada n Dimensional**



# 3.1 Comunicação Manual da Pessoa com Deficiência



**Três Comunicados Manuais Possíveis**



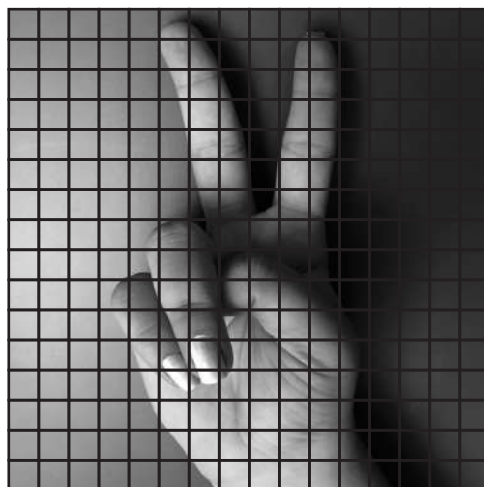
## Exemplos de Elementos de Inclusão do Deficiente



# 3.2 Coleção Lições Fotográficas das Mãos



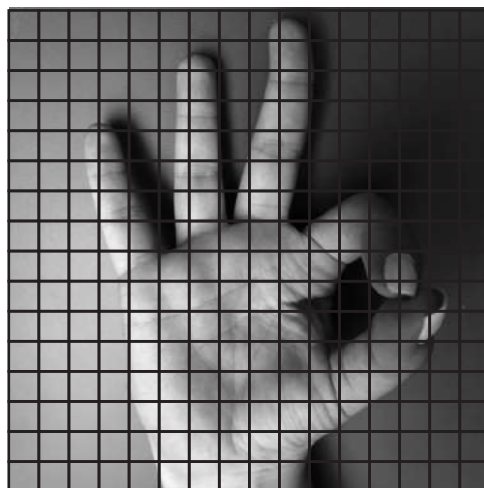
○ ○ ○  
 0  
 1  
 0  
 1  
 ○ ○ ○



Xixi



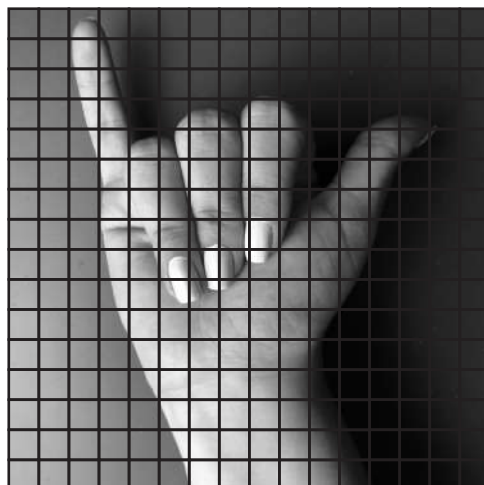
○ ○ ○  
 1  
 0  
 0  
 1  
 ○ ○ ○



Bom



○ ○ ○  
 0  
 1  
 1  
 0  
 ○ ○ ○



Dor

## Três Diferentes e Possíveis Posições das Mãos



# C-COVARIANCE

## Conjoint Analysis

$$\begin{bmatrix} A1 - \bar{A1} \\ A2 - \bar{A2} \end{bmatrix}^t \circ \begin{vmatrix} \sigma A11 & \sigma A12 \\ \sigma A21 & \sigma A22 \end{vmatrix}^{-1} \circ \begin{bmatrix} A1 - \bar{A1} \\ A2 - \bar{A2} \end{bmatrix}$$

$$\begin{bmatrix} B1 - \bar{B1} \\ B2 - \bar{B2} \end{bmatrix} \circ \begin{vmatrix} \sigma B11 & \sigma B12 \\ \sigma B21 & \sigma B22 \end{vmatrix} \circ \begin{bmatrix} B1 - \bar{B1} \\ B2 - \bar{B2} \end{bmatrix}$$

$$A11 - \bar{A1} \quad A12 - \bar{A2}$$

$$A21 - \bar{A1} \quad A22 - \bar{A2}$$

Dados Intervalares Likert

**Covariância Matricial**

Prof. Aguinaldo Prandini Ricieri



# 3.1 Bancos de Dados de Duas Classes dos Clientes do Banco

X1	X2	X3	X4	X5	X6	X7	DADOS	Y1	Y2	Y3	Y4	Y5	Y6	Y7	
3	3	5	4	3	2	4	Modelo Covariance	6	5	4	7	5	5	6	
4	2	4	5	1	2	5		8	4	7	8	5	6	6	
5	1	2	1	2	1	2		9	5	6	6	4	7	7	
1	1	2	2	1	4	1		7	3	6	8	3	7	6	
2	2	2	2	1	6	2		8	4	6	4	4	4	9	4
1	2	5	5	2	5	4		9	3	8	8	8	4	7	8
4	1	1	1	1	6	4		6	5	6	5	3	9	9	6
3	2	5	3	1	3	5		8	4	7	4	5	8	8	4
3	1	5	4	1	5	1		9	3	4	5	5	7	7	5
2	3	3	3	2	1	2		8	3	4	4	5	8	7	7
3	1	4	3	2	4	3		6	4	7	5	5	9	9	4
4	1	1	3	1	6	2		9	3	8	7	3	6	6	8
6	1	2	1	1	1	5		7	3	5	4	5	5	5	7
2	2	3	2	1	5	2		9	5	6	7	5	7	7	8
4	2	1	4	2	5	1		7	3	7	7	5	8	6	6
3	1	1	3	3	3	2		6	3	7	6	4	5	7	7
1	1	4	2	3	2	5		8	3	6	7	5	7	7	8
4	1	5	4	1	1	4		8	5	4	8	4	7	8	8
3	3	3	2	3	4	4		9	4	6	4	4	6	7	7
5	1	2	1	2	2	1		6	3	5	4	5	7	7	7
3	3	5	1	1	3	5		9	5	5	7	5	8	8	4
6	3	4	1	2	2	3		8	3	4	6	5	6	6	6
5	3	3	1	2	1	2		8	5	4	8	5	7	7	6
5	2	3	4	1	1	2		7	5	7	4	5	7	7	5
6	1	2	3	1	1	5		8	4	7	5	4	9	9	5
5	1	3	2	1	1	3		8	4	8	8	5	6	6	6
1	2	2	1	2	4	5		7	5	8	8	5	9	9	5
5	1	5	2	1	2	2		9	4	6	7	5	9	8	8
3	2	3	2	1	5	3		6	3	8	4	5	6	8	8
3	2	2	1	1	1	3		6	3	7	8	5	5	6	6
5	1	4	2	1	1	1		9	3	4	7	4	7	7	6
6	2	2	3	1	3	2		9	5	8	8	5	6	7	7
5	2	1	1	1	1	2		7	5	4	8	5	7	7	7
1	2	4	2	1	1	1		8	4	5	8	4	5	5	5
4	1	2	2	1	1	4		8	3	6	7	5	8	6	6
2	1	4	4	2	1	3		7	5	5	8	5	8	7	7
2	2	5	4	2	2	2		7	4	4	8	5	9	9	5
4	1	4	5	1	1	1		8	5	4	8	3	6	7	7
1	1	4	4	1	1	1		9	4	8	8	3	7	6	6
1	3	2	4	2	1	3		9	4	5	7	5	5	4	4
3	1	3	1	2	1	4	8	5	5	8	5	5	4	4	
1	2	4	2	2	2	5	9	4	8	6	5	9	4	4	
1	3	1	2	1	2	2	9	3	4	8	4	6	5	5	
3	2	5	2	2	1	4	8	5	4	6	5	8	7	7	
6	1	2	3	1	3	1	6	4	7	8	4	9	6	6	
5	1	5	4	1	1	4	7	3	6	8	3	9	8	8	
4	3	1	5	1	1	4	8	5	6	5	4	9	5	5	
2	2	2	4	1	1	4	8	4	4	8	5	6	8	8	
1	3	3	3	1	1	1	9	4	5	4	5	7	4	4	
3	1	3	3	3	3	2	8	3	4	8	4	5	7	7	
<b>3,3</b>	<b>1,7</b>	<b>3,1</b>	<b>2,7</b>	<b>1,5</b>	<b>2,4</b>	<b>2,9</b>	<b>MÉDIA</b>	<b>7,8</b>	<b>4,0</b>	<b>5,8</b>	<b>6,6</b>	<b>4,5</b>	<b>7,1</b>	<b>6,1</b>	





# 4.0 Alocação do Cliente C=4567 na Agência A

c1	c2	c3	c4	c5	c6	c7
6	4	4	3	2	6	5
mx1	mx2	mx3	mx4	mx5	mx6	mx7
3,30	1,74	3,06	2,66	1,50	2,36	2,86

2,70	2,26	0,94	0,34	0,50	3,64	2,14
------	------	------	------	------	------	------

**1x7**

0,47	0,24	0,08	0,07	0,21	0,09	-0,02
0,24	1,87	0,09	0,00	-0,15	0,10	-0,11
0,08	0,09	0,65	-0,17	-0,09	0,09	-0,12
0,07	0,00	-0,17	0,70	0,10	0,01	0,07
0,21	-0,15	-0,09	0,10	2,42	-0,02	-0,13
0,09	0,10	0,09	0,01	-0,02	0,41	0,02
-0,02	-0,11	-0,12	0,07	-0,13	0,02	0,56

**7x7**

**1x1**

COVARIANCE  
C(X) = 28,512  
LIKERT/AG:A

==

2,31	5,01	0,99	0,51	1,03	2,08	0,83
------	------	------	------	------	------	------

**1x7**

2,70	2,26	0,94	0,34	0,50	3,64	2,14
------	------	------	------	------	------	------

**7x1**

28,512 X 33,242

c1	c2	c3	c4	c5	c6	c7
6	4	4	3	2	6	5
my1	my2	my3	my4	my5	my6	my7
7,80	3,96	5,78	6,58	4,50	7,06	6,12

-1,80	0,04	-1,78	-3,58	-2,50	-1,06	-1,12
-------	------	-------	-------	-------	-------	-------

**1x7**

0,95	-0,06	0,04	-0,05	0,11	0,05	0,07
-0,06	1,67	0,10	-0,19	-0,21	-0,22	0,17
0,04	0,10	0,52	0,01	0,10	-0,10	0,01
-0,05	-0,19	0,01	0,49	0,16	0,10	-0,11
0,11	-0,21	0,10	0,16	2,24	0,14	0,20
0,05	-0,22	-0,10	0,10	0,14	0,63	0,09
0,07	0,17	0,01	-0,11	0,20	0,09	0,66

**7x7**

**1x1**

COVARIANCE  
C(Y) = 33,242  
LIKERT/AG:B

==

-1,99	1,24	-1,18	-2,07	-6,95	-1,41	-1,10
-------	------	-------	-------	-------	-------	-------

**1x7**

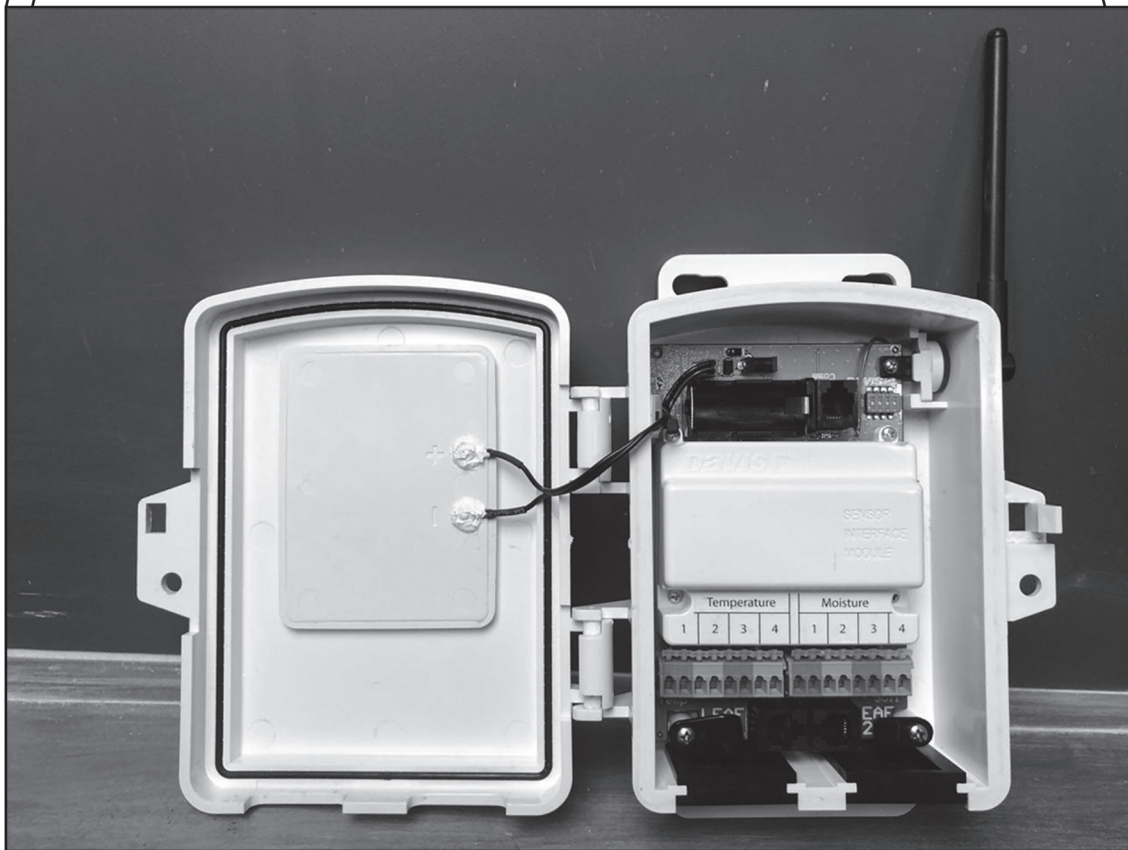
-1,80	0,04	-1,78	-3,58	-2,50	-1,06	-1,12
-------	------	-------	-------	-------	-------	-------

**7x1**

Computação do C=Covariance/Likert



# 5.0 Agricultura de Precisão e sua Monitoração



**Coletor de Dados para Fomar o Banco de Dados**



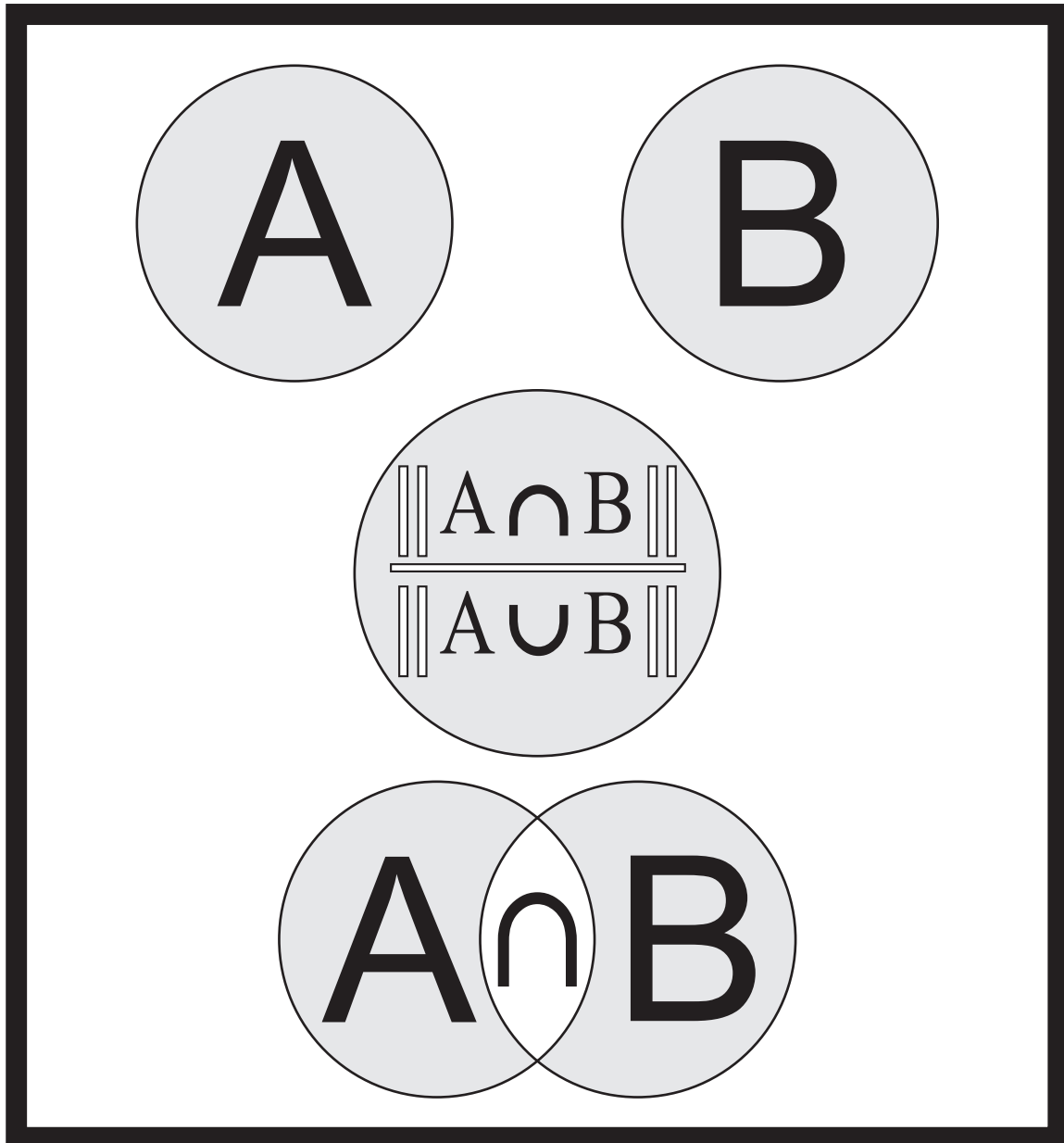
# 5.1 Banco de Dados de Duas Classes da Agrícola

X1	X2	X3	X4	X5	X6	X7	DADOS	Y1	Y2	Y3	Y4	Y5	Y6	Y7
21	36	25	55	60	11	40	Modelo Covariance	29	37	25	62	65	12	47
20	34	22	54	54	12	46		28	40	28	59	65	12	47
23	30	24	59	53	11	46		28	37	26	67	67	13	48
21	33	24	53	61	13	45		29	40	27	68	62	12	50
25	32	23	52	59	12	42		27	40	26	69	64	13	46
25	35	25	60	54	11	43		27	39	27	62	60	12	49
24	30	23	60	55	11	42		28	40	27	69	62	13	48
26	33	22	52	61	12	45		27	40	28	62	62	14	46
23	33	23	53	58	13	45		26	39	27	68	65	13	47
25	30	22	54	56	13	40		29	36	25	66	68	14	47
24	32	23	60	52	13	44		30	38	29	63	64	13	50
20	33	25	59	55	13	40		29	37	27	65	63	12	50
25	35	25	50	56	12	40		29	39	28	64	69	12	49
25	33	23	51	52	13	44		28	36	27	61	65	12	47
25	35	24	55	61	13	41		27	38	25	60	66	12	50
21	34	25	50	58	13	41		26	38	27	63	62	14	47
24	36	23	57	52	11	41		30	39	26	63	65	12	50
22	36	25	60	55	12	43		26	36	26	60	61	13	49
23	32	24	58	53	11	44		26	35	28	67	62	13	49
21	35	23	51	60	13	40		26	38	25	65	64	13	45
25	36	22	55	52	13	46		29	40	25	64	68	14	49
23	34	22	59	56	13	45		30	35	26	67	67	12	48
20	34	25	60	61	12	41		30	39	29	63	60	14	45
22	32	22	59	51	12	41		29	35	25	67	62	14	46
24	32	24	58	54	11	46		29	37	28	59	64	14	49
24	32	25	50	59	11	42		30	37	29	61	67	13	49
24	31	23	60	53	12	46		28	37	25	61	65	14	45
25	36	22	51	50	12	40		28	40	26	67	67	14	50
23	34	23	54	58	12	42		26	36	28	59	61	13	47
21	31	24	50	61	12	44		29	36	26	61	61	13	45
21	35	25	58	50	12	40		26	39	29	65	63	12	50
25	31	22	53	52	12	44		27	35	29	68	64	12	47
24	35	22	54	55	12	42		30	40	27	67	65	14	45
26	30	22	60	51	13	41		29	40	28	66	62	13	48
20	30	22	53	54	13	45		27	39	26	60	68	12	48
20	31	23	56	60	11	42		30	40	28	66	60	12	46
25	31	24	56	58	13	45		26	38	28	61	68	13	46
24	30	23	55	55	11	46		26	37	25	59	63	13	45
21	36	25	59	61	11	41		28	38	28	66	68	14	45
25	36	22	55	56	13	44		30	36	27	64	61	14	49
22	33	22	53	57	13	44	28	37	26	64	60	12	45	
21	33	22	59	57	12	42	26	37	28	62	62	14	49	
22	34	25	57	54	13	46	27	40	28	64	69	12	47	
23	32	22	51	54	12	44	27	35	27	60	60	14	46	
25	35	22	59	50	11	44	27	40	25	61	68	14	48	
24	31	22	55	54	12	40	28	36	27	59	66	12	48	
22	30	22	58	57	11	43	29	36	25	62	70	14	48	
24	35	22	50	60	12	45	26	36	26	60	61	13	46	
23	35	22	60	60	13	42	26	36	27	63	65	14	50	
26	31	25	52	60	12	41	27	39	27	60	61	13	47	
<b>23,1</b>	<b>33,1</b>	<b>23,2</b>	<b>55,4</b>	<b>55,9</b>	<b>12,1</b>	<b>42,9</b>	<b>MÉDIA</b>	<b>27,9</b>	<b>37,8</b>	<b>26,8</b>	<b>63,4</b>	<b>64,1</b>	<b>13,0</b>	<b>47,5</b>



# J-JACCARD

Measure of Similarities



Interseção Numérica

Le Coefficient de Communauté

Prof. Aguinaldo Prandini Ricieri





THE  
**NEW PHYTOLOGIST.**

VOL. XI, No. 2.

FEBRUARY, 1912.

[PUBLISHED FEBRUARY 29TH.]

THE DISTRIBUTION OF THE FLORA IN THE  
ALPINE ZONE.<sup>1</sup>

BY PAUL JACCARD,

*Professor at the Federal Polytechnic, Zürich.*



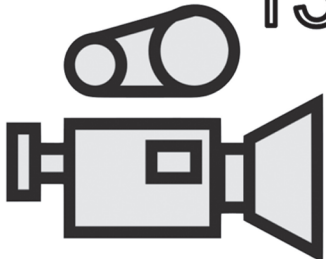






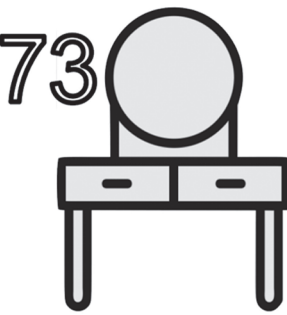

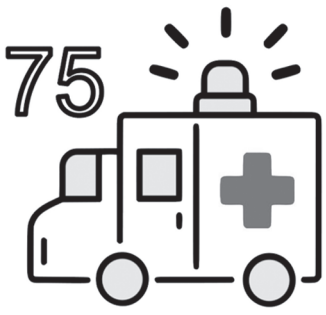

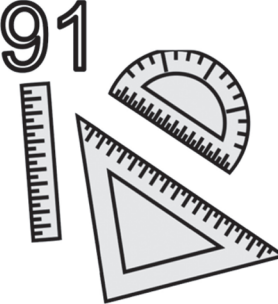

[WITH TWO TEXT-FIGS.]

**T**HE botanising hitherto done in the Alps gives us but very imperfect information on the local distribution of the alpine flora. Attention has been specially directed to rare plants, whose least localities are recorded, while the common species are often neglected. But from the standpoint of the factors which regulate distribution the common species are the most important. The rare species of the alpine flora, those which appear only in a few isolated stations, sometimes only in a single one, are usually either species with a very sporadic general distribution, or they are endemic species, or finally they may be "glacial relicts." All these categories are of great interest from the standpoint of the history of floras; their presence in the stations which they now occupy is explained not only by the ecological conditions they find in these stations, but also by historical causes, and especially by the conditions of post-glacial immigration.

<sup>1</sup> This article is translated from the French original (which appeared in the *Revue générale des Sciences*, 15th December, 1907, pp. 961—967) and published in **THE NEW PHYTOLOGIST** (by kind permission of M. Olivier, the editor of the *Revue générale*) at Professor Jaccard's request. The Editor acceded to this request the more readily, inasmuch as the statistical method employed by Professor Jaccard appears worthy of being tested over a wider range of vegetation. With suitable development, these and similar statistical methods promise to form an important means of connecting the study of floristic distribution with that of the determination and distribution of units of vegetation.—EDITOR, **NEW PHYT.**



# 4.1 Numerização de Referência para o Algoritmo J-Jaccard

<p>11</p> 	<p>12</p> 	<p>13</p> 
<p>35</p> 	<p>36</p> 	<p>37</p> 
<p>51</p> 	<p>52</p> 	<p>53</p> 
<p>73</p> 	<p>74</p> 	<p>75</p> 
<p>90</p> 	<p>91</p> 	<p>92</p> 

Exemplos das 99 Coisas de Interesse Monitoradas pela Empresa P-FaceLibro





# 4.4 Banco de Dados Anual de Fotos do Usuário Carla

11	89	89	87	26	21	89	89	90	12	12	91	34	11	89	89	94	21	89	89
12	11	89	12	12	11	31	11	11	89	21	93	12	91	43	11	26	21	96	33
11	27	21	89	12	12	89	90	21	89	85	89	11	11	45	12	21	21	89	89
89	87	89	12	21	25	40	89	93	86	12	30	12	44	11	88	89	89	89	31
25	95	89	89	95	96	11	27	11	11	43	12	12	46	27	11	44	89	36	21
88	90	89	21	45	11	89	89	21	89	89	89	36	89	29	94	12	89	89	31
27	89	89	12	41	21	11	89	12	21	89	89	21	89	89	31	25	86	90	12
12	86	87	34	11	89	89	11	89	89	11	12	89	44	98	39	21	32	11	40
29	11	30	21	33	89	21	11	30	12	11	89	89	90	89	89	11	89	27	89
35	88	29	11	89	12	89	11	11	89	21	42	11	95	11	11	11	11	89	40
21	89	33	89	89	89	92	41	89	11	21	89	37	12	45	12	11	11	85	11
26	89	43	89	40	12	12	11	21	89	21	12	89	11	98	35	37	12	38	38
31	28	21	11	28	37	12	28	89	98	27	21	93	12	11	44	88	21	85	12
11	21	21	43	89	38	12	12	92	89	11	11	89	11	42	11	11	89	21	89
95	89	89	45	89	11	12	35	89	87	12	21	11	12	96	87	11	11	44	45
11	89	21	89	21	26	12	90	11	21	32	89	32	89	29	89	89	11	89	30
11	11	89	11	89	21	21	89	89	21	94	12	40	26	89	11	98	12	98	21
12	11	91	21	89	11	12	11	38	89	11	95	89	89	39	89	32	89	89	45
21	89	89	21	94	89	21	12	89	11	11	89	11	89	89	11	11	89	12	41
12	11	89	43	12	87	89	32	89	37	85	11	12	39	27	11	11	89	11	89
89	29	11	21	12	21	89	40	12	11	12	12	21	12	89	93	43	11	11	11
11	89	12	31	30	11	89	21	89	43	87	89	89	31	89	11	21	44	12	12
41	45	12	29	97	91	89	89	11	11	11	37	11	37	11	11	28	95	89	89
96	21	11	12	11	27	11	11	89	39	40	12	12	38	12	12	34	88	11	89
37	33	11	89	89	88	89	43	94	85	27	11	38	29	21	29	89	29	38	21
87	89	12	89	11	27	11	89	11	34	11	12	89	89	89	89	28	12	41	89
11	12	89	30	40	30	12	11	96	11	89	89	89	89	12	44	11	11	89	21
89	31	89	21	89	26	97	12	94	11	37	21	89	90	89	12	11	21	26	98
89	89	32	31	97	89	11	89	45	11	45	27	30	89	86	12	11	85	11	36
89	11	89	21	89	89	21	92	84	11	21	21	89	21	90	21	36	90	34	87
89	11	11	21	89	88	45	12	12	21	89	21	11	89	93	27	32	12	44	33
90	89	11	28	89	11	34	12	90	21	89	11	34	21	12	12	40	12	29	89
11	42	35	37	11	42	40	21	11	21	90	89	89	12	12	21	45	12	11	12
40	12	12	21	12	89	41	94	30	11	89	21	34	12	87	89	89	11	12	11
89	29	32	90	89	45	21	21	38	21	21	89	11	96	11	42	29	12	30	40
89	21	93	37	11	87	31	12	89	12	11	21	12	89	89	21	89	38	21	11
39	91	89	89	21	35	89	21	26	12	12	90	21	89	21	11	37	97	21	21
21	89	89	27	91	89	11	11	89	89	89	11	89	89	21	11	89	89	97	89
12	92	89	36	11	29	11	89	42	12	89	89	11	12	89	21	42	12	11	39
11	12	11	11	45	89	89	28	11	21	21	94	21	97	11	89	21	89	42	30
12	40	42	11	89	12	12	89	11	35	11	42	11	11	89	11	90	45	12	21
89	90	89	29	26	11	11	12	89	90	94	11	11	89	11	21	98	89	41	98
87	12	89	89	11	32	89	89	11	93	12	11	11	11	96	21	11	89	89	12
35	89	90	89	12	89	12	12	89	29	42	43	89	87	12	11	89	40	25	93
89	90	12	97	11	21	21	89	90	12	89	89	85	11	89	97	92	11	11	29
11	11	12	11	11	97	89	43	12	89	89	11	27	38	12	11	21	41	89	42
89	12	89	12	11	21	89	21	89	11	21	12	11	95	89	90	89	89	89	89
38	89	89	38	21	89	89	89	89	21	21	38	11	11	89	98	89	11	92	21
89	89	89	21	21	89	29	87	11	89	11	21	98	12	11	89	11	21	39	36
11	32	94	30	11	21	89	41	89	29	12	11	21	94	44	11	12	11	11	89

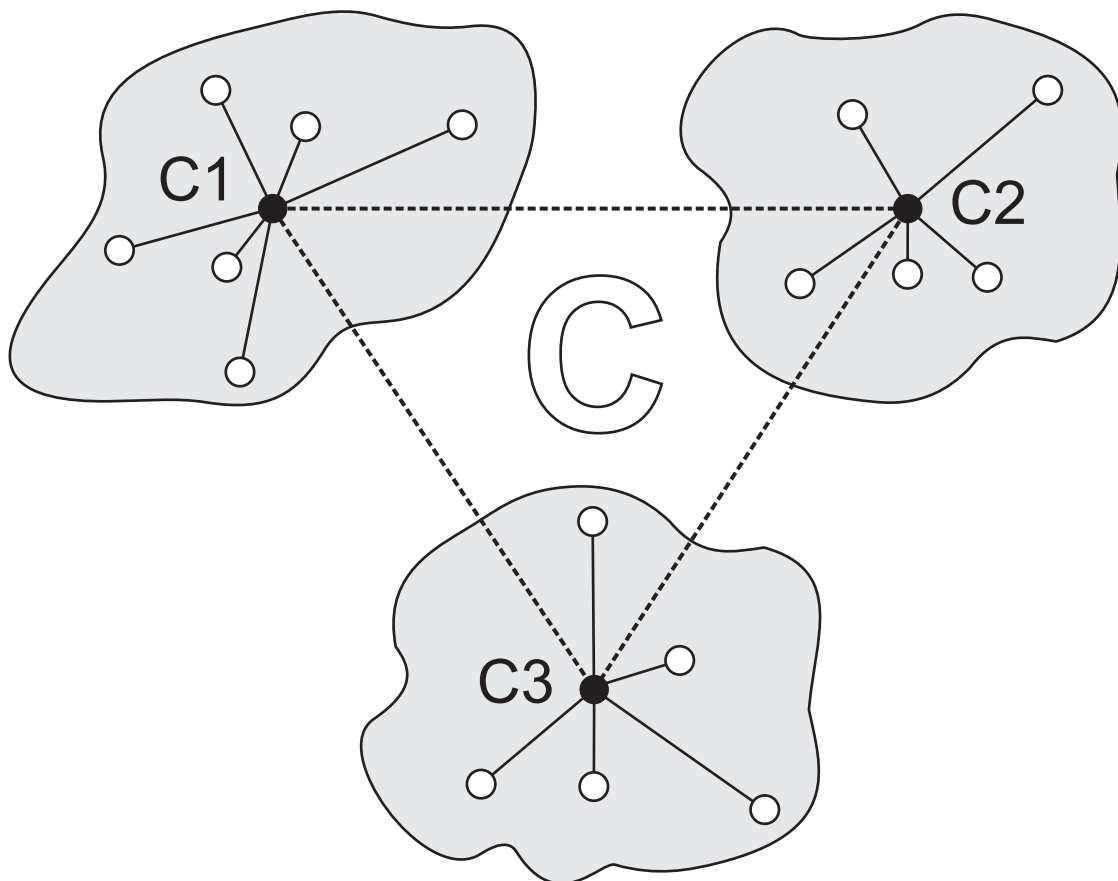
Fotografia Capturadas e Postadas pela Carla na Rede P-FaceLibro



## 7.0 Homogeneidade Versus Heterogeneidade de Dados

Os resultados dos inúmeros algoritmos de clusterização de dados – taxonomia da proximidade estatística de informações –, dependem da métrica de (dis)similaridade escolhida a qual é intimamente dependente da forma como os dados foram representados (numéricos, booleanos, categóricos... outros) e se há ou não necessidade desses dados serem normalizados.

Clusterização, propriamente dita e *ad hoc* – que é feito para um fim específico –, implica maximizar a homogeneidade dos dados dentro dos clusters (coesão interna) e maximizar a heterogeneidade entre os clusters (separação externa), o que pode ser melhor compreendido observando a próxima alegoria e legendas:



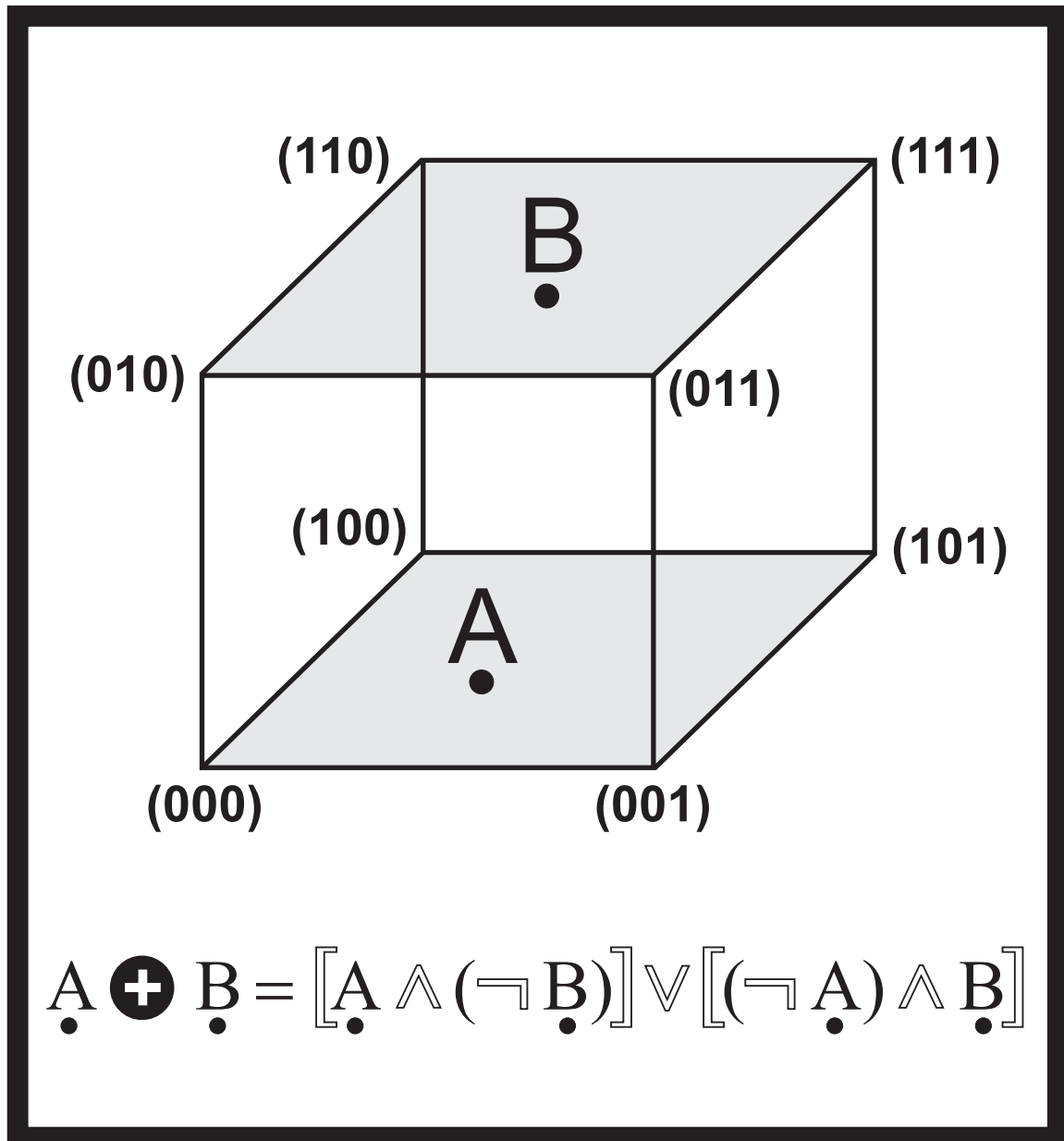
**Legenda das Representações**

- (—): Minimização das distâncias dentro dos clusters entre o omedioides (C1, C2, C3) e seus dados (0, 0, 0, ..., 0); homogeneidade dos dados.
- (----): Maximização das distâncias entre os clusters, de medioide a medioide (C1 ---- C2), (C1 ---- C3), (C2 ---- C3); heterogeneidade dos clusters.



# H-HAMMING

## Binary Distance



## Bit Statistics

# Similaridade de Conteúdos

Prof. Aguinaldo Prandini Ricieri



# 1.2 Exemplos do Cálculo da Distância de Hamming

DH (molho//milho)= 2  
 DH (molho//soja)= 14  
 DH (molho//trigo)= 13  
 DH (molho//sorgo)= 12



DH (milho//milho)= 0  
 DH (milho//soja)= 14  
 DH (milho//trigo)= 13  
 DH (milho//sorgo)= 14

DH (soje//milho)= 13  
 DH (soje//soja)= 1  
 DH (soje//trigo)= 12  
 DH (soje//sorgo)= 17



DH (soja//milho)= 14  
 DH (soja//soja)= 0  
 DH (soja//trigo)= 13  
 DH (soja//sorgo)= 18

DH (grigo//milho)= 12  
 DH (grigo//soja)= 14  
 DH (grigo//trigo)= 3  
 DH (grigo//sorgo)= 10



DH (trigo//milho)= 13  
 DH (trigo//soja)= 13  
 DH (trigo//trigo)= 0  
 DH (trigo//sorgo)= 11

DH (torgo//milho)= 13  
 DH (torgo//soja)= 17  
 DH (torgo//trigo)= 8  
 DH (torgo//sorgo)= 3



DH (sorgo//milho)= 14  
 DH (sorgo//soja)= 18  
 DH (sorgo//trigo)= 11  
 DH (sorgo//sorgo)= 0

a	01100001	n	01101110
b	01100010	o	01101111
c	01100011	p	01110000
d	01100100	q	01110001
e	01100101	r	01110010
f	01100110	s	01110011
g	01100111	t	01110100
h	01101000	u	01110101
i	01101001	v	01110110
j	01101010	w	01110111
k	01101011	x	01111000
l	01101100	y	01111001
m	01101101	z	01111010

A	01000001	N	01001110
B	01000010	O	01001111
C	01000011	P	01010000
D	01000100	Q	01010001
E	01000101	R	01010010
F	01000110	S	01010011
G	01000111	T	01010100
H	01001000	U	01010101
I	01001001	V	01010110
J	01001010	W	01010111
K	01001011	X	01011000
L	01001100	Y	01011001
M	01001101	Z	01011010



# 4.0 Comparando a Proximidade de Genes A

Droso	Droso
01010100	01000001
01000111	01000011
01000111	01000011
01000011	01000111
01000011	01000111
01000001	01010100
01000111	01000011
01010100	01000001
01000001	01010100
01000001	01010100
01010100	01000001
01000011	01000111
01010100	01000001
01000111	01000011
01010100	01000001
01000001	01010100
01000111	01000011
01000001	01010100
01000011	01000111
01010100	01000001
01000011	01000111
01000001	01010100
01000111	01000011
01010100	01000001
01000001	01010100
01000011	01000111
01010100	01000001
01000011	01000111
01000011	01000111
01000011	01000111
01000011	01000111
01000111	01000011
01000001	01010100
01000001	01010100
01000011	01000111
01000001	01010100
01000011	01000111
01000001	01010100
01010100	01000001
01010100	01000001
01000011	01000111
01000111	01000011
01000111	01000011
01000001	01010100
01010100	01000001
01010100	01000001
01000011	01000111
01000111	01000011
01000111	01000011
01000001	01010100
01000001	01010100
01000011	01000111
01000001	01010100
01000011	01000111

Drosophila



Homo	Homo
01010100	01000001
01000001	01010100
01000111	01000011
01000111	01000011
01010100	01000001
01000011	01000111
01000001	01010100
01000011	01000111
01000111	01000011
01000011	01000111
01010100	01000001
01000111	01000011
01010100	01000001
01000111	01000011
01000011	01000111
01000011	01000111
01000011	01000111
01000011	01000111
01000011	01000111
01010100	01000001
01010100	01000001
01000111	01000011
01000011	01000111
01010100	01000001
01000111	01000011
01000001	01010100
01000111	01000011
01000111	01000011
01010100	01000001
01010100	01000001
01000001	01010100
01000001	01010100
01000111	01000011
01000111	01000011
01010100	01000001
01010100	01000001
01000011	01000111
01010100	01000001
01010100	01000001
01000011	01000111
01000011	01000111

Homo sapiens





# Z-CONOVER

Dados Pareados

+	-	+	-
-	+	-	+
+	-	+	-
-			+
+	-	+	-
-	+	-	+

$$z = \frac{x - 0.5n}{0.5\sqrt{n}}$$

Teste de Hipóteses

Estatística de Sinais Algébricos

Prof. Aguinaldo Prandini Ricieri



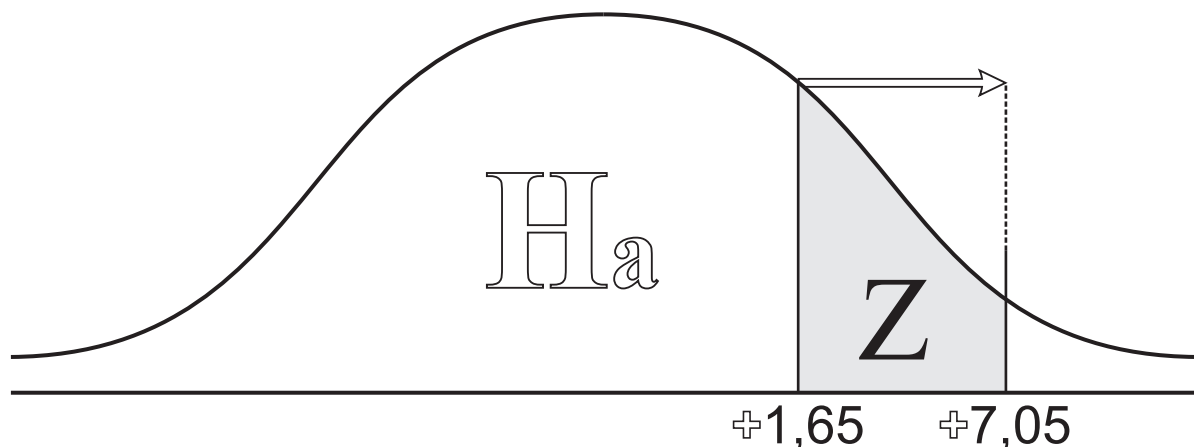


## 6.0 Cálculo da Validação das Hipóteses dos Cabritos

$$Z = \frac{\left( \text{Valor observado} \right) - \left( \text{Valor esperado} \right)}{\text{Desvio Padrão do Número de } + \text{ e de } -}$$

$$Z = \frac{75^+ - 85 \cdot \frac{1}{2}}{\sqrt{85 \cdot \frac{1}{2} \left( 1 - \frac{1}{2} \right)}} = 7,05$$

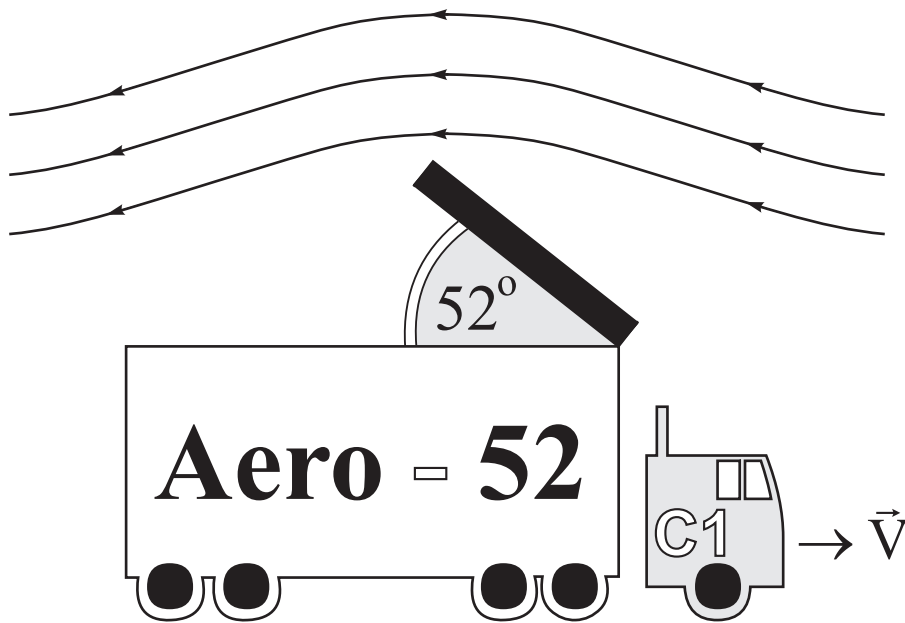
Ao nível de 0,05 (95% da área) o teste unilateral (+) rejeita a hipótese  $H_0$  pois a ração engordou os animais tratados com a ração Cabral-196: seu efeito alterou positivamente (+) a engorda dos cabritos.



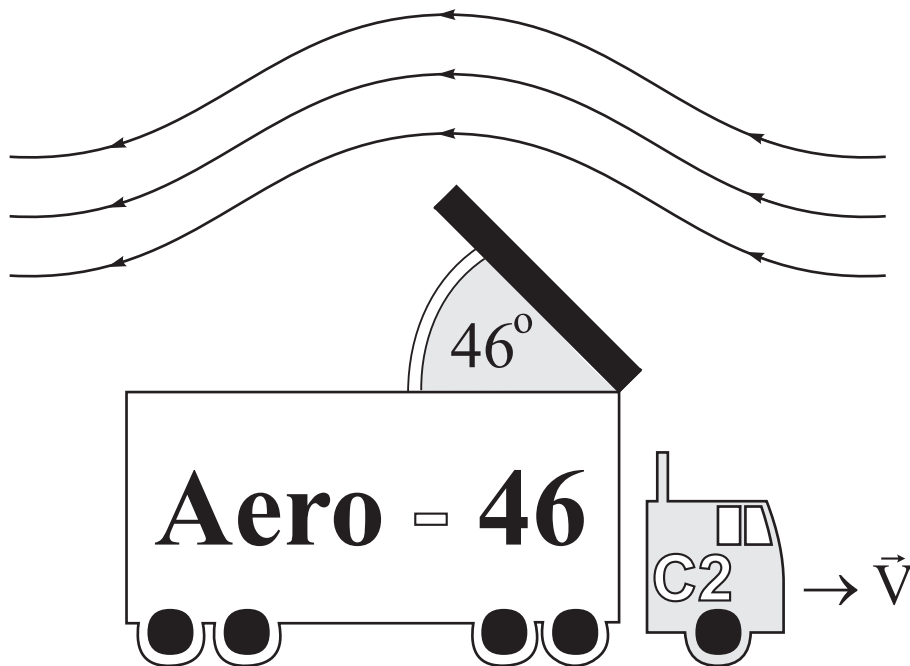
Aplicação do Teste de Conover (Cabral = 196)



## 7.0 Teste com Aerofólios Defletores de Ar Aero = 52/46



Caminhões transportando 40 toneladas de carga a velocidade de 80 km/h constante gastam certas quantidades de litros de diesel para percorrerm 360 km. Seguem os testes de 2 aerofólios de caminhões:

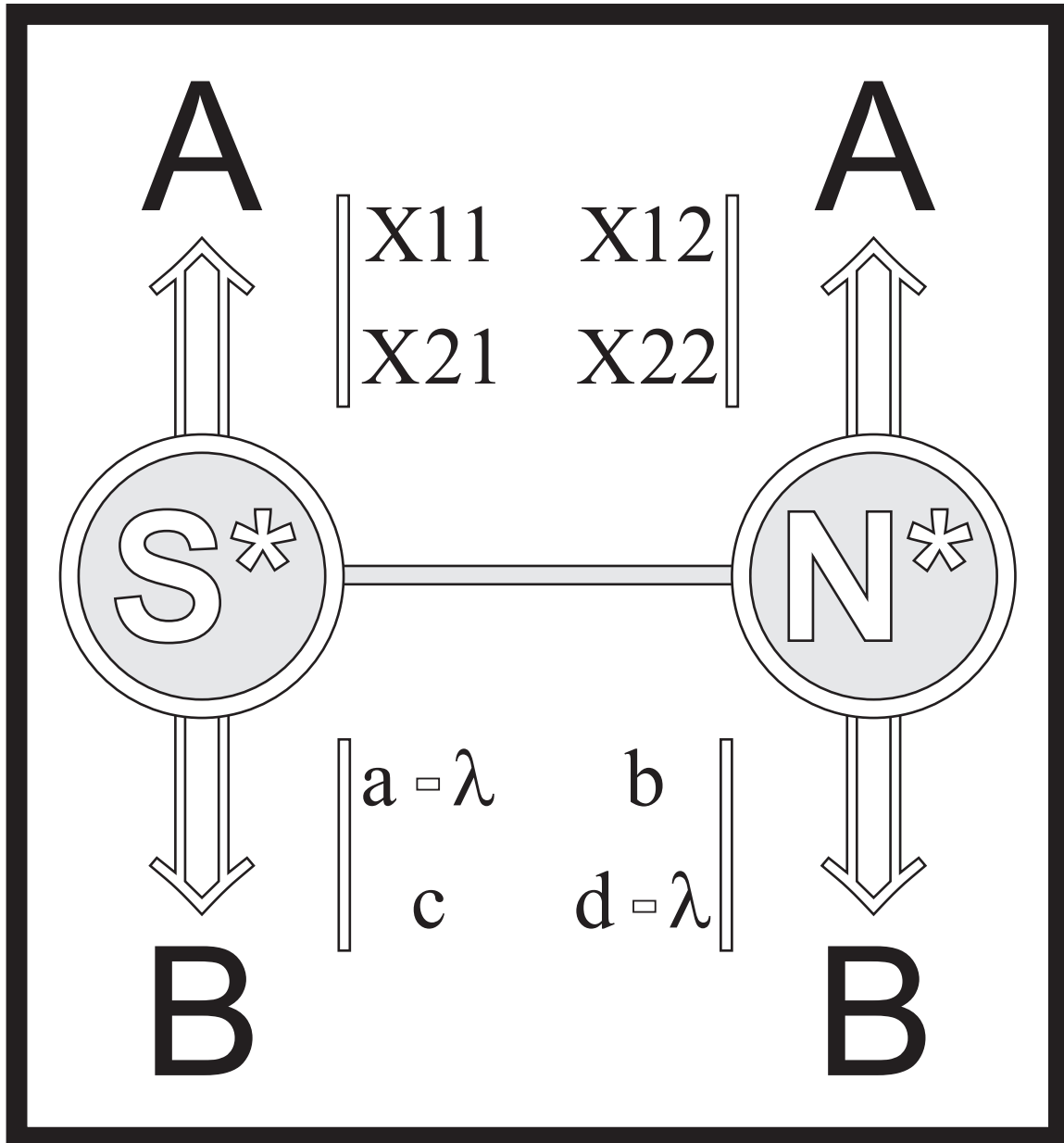


Aplicação do Teste de Conover



# D-DISJUNTOR

Desagregador de Dados



Matrizes de Covariâncias

Autovalores e Autovetores

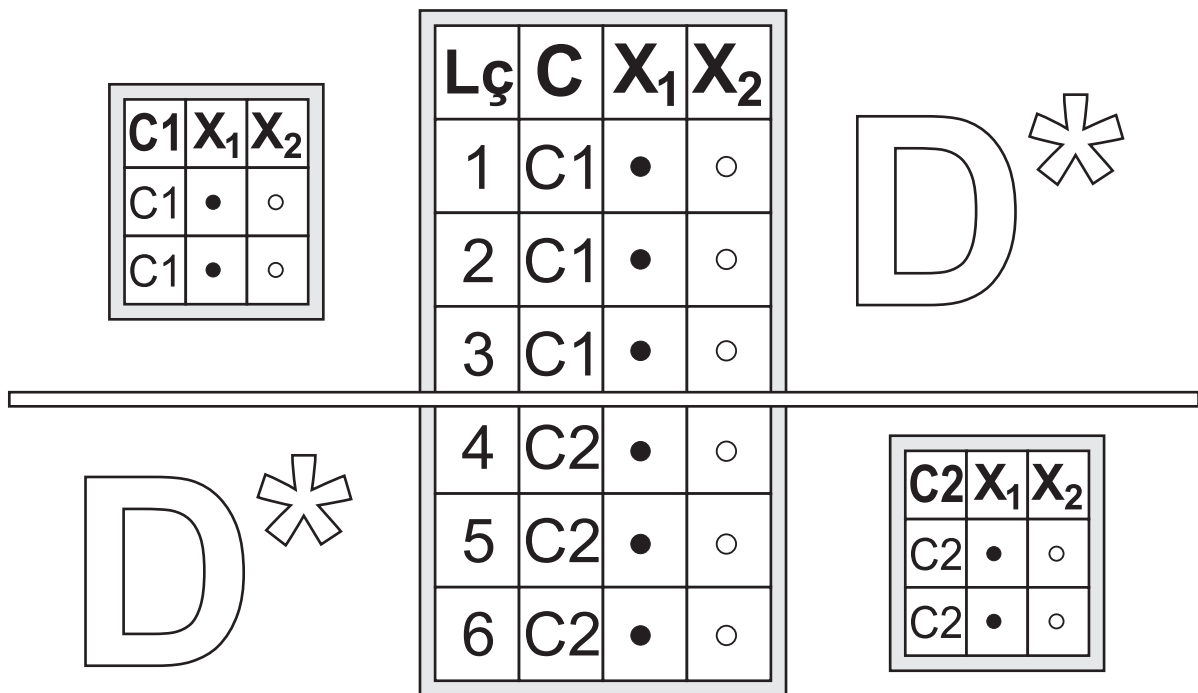
Prof. Aguinaldo Prandini Ricieri



# D=DISJUNTOR

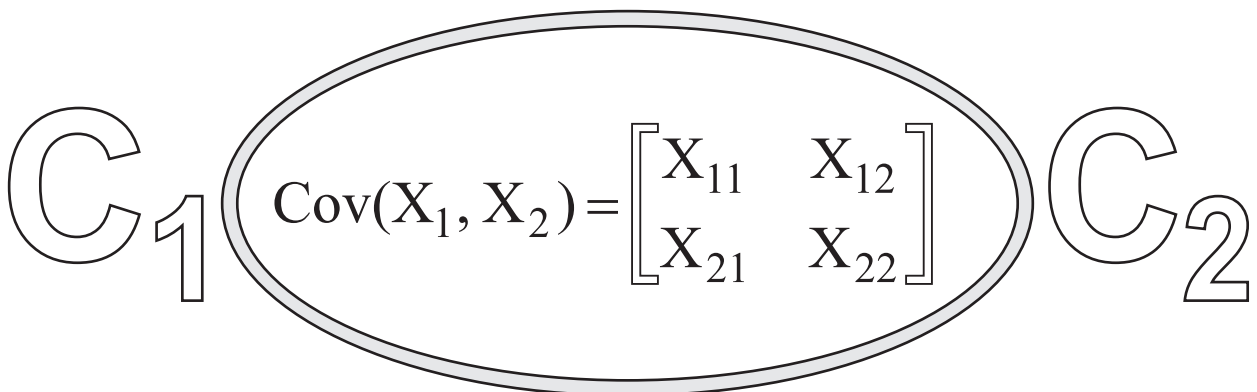
## 1.0 Classificador de Dados por Separação Variante

A finalidade desse método estatístico é disjuntar diferentes classes de dados ( $C_1, C_2$ ) associadas aos descritores ( $x_1, x_2, x_3, \dots, x_n$ ), maximizando a covariância entre as classes e minimizando a covariância dentro das classes dos dados analisados:



Separação de Classe C1 e C2

A eficiência estatística do D=Disjuntar é evidente na medida que possibilita excelente classificação de dados, que estão distribuídas nessas duas classes  $C_1$  e  $C_2$ .



Matriz de Covariância Reduzida



# 3.2 Cálculos para a Obtenção do D=Disjuntor

Vírus	P	D	X1	X2	D	$\bar{D}$	Vírus
V	01	V	11	40	41,30	41,05	V
	02	V	22	39	41,70		
	03	V	12	41	42,44		
	04	V	5	39	39,63		
	05	V	31	36	39,80		
	06	V	42	37	42,20		
B	07	B	55	39	45,82	45,55	B
	08	B	31	42	45,84		
	09	B	46	40	45,70		
	10	B	18	42	44,20		
	11	B	62	41	48,68		
	12	B	41	38	43,08		

Fronteira de Separação dos Dados: V ou B

## Cálculos dos Autovalores $\lambda_1$ e $\lambda_2$

$$\det \begin{bmatrix} (1,19 = \lambda) & 0,09 \\ 10,04 & (0,68 = \lambda) \end{bmatrix} = 0 \circ (1,19 = \lambda) \circ (0,68 = \lambda) = 0,903$$

$$\boxed{\lambda_1 = +1,919 \quad \lambda_2 = -0,049}$$

## Cálculos dos Autovetores V1 e V2

$$\begin{bmatrix} -0,729 & 0,090 \\ 10,040 & -1,235 \end{bmatrix} \circ \begin{bmatrix} x \\ y \end{bmatrix} \xrightarrow{\lambda_1} = +1,919 \circ \begin{bmatrix} x \\ y \end{bmatrix} \circ V1 = \begin{bmatrix} +0,124 \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} 1,235 & 0,090 \\ 10,040 & 0,729 \end{bmatrix} \circ \begin{bmatrix} x \\ y \end{bmatrix} \xrightarrow{\lambda_2} = -0,049 \circ \begin{bmatrix} x \\ y \end{bmatrix} \circ V2 = \begin{bmatrix} -0,072 \\ 1 \end{bmatrix}$$

$$\boxed{D = +0,124 \circ X1 + 1 \circ X2}$$



# D-DISCREPANTIA

Incomparabili Influentia

Lç	X	Y
01		
02		
03		
04	•	•

y

Lç	X	Y
01	•	•
02		
03		
04		

b<sub>0</sub>

Lç	X	Y
01		
02	•	•
03		
04		

b<sub>1</sub>

Lç	X	Y
01		
02		
03	•	•
04		

y<sup>^</sup>

Lç	X	Y
01		
02		
03		
04	•	•

Intercepto e Coeficientes  
 Distância de Pontos Estranhos

Prof. Aguinaldo Prandini Ricieri





# 3.0 Retirando (*Extraho*) e Voltando (*Reverti*) com as Lições

Lç	X	Y
01	0	1
02	1	2
03	2	4
04	3	6
05	5	8
06	7	9
07	8	8
08	9	1

$\hat{y}_0 = 3,04 \mp 0,41X$

Lç	X	Y
01	●	●
02	1	2
03	2	4
04	3	6
05	5	8
06	7	9
07	8	8
08	9	1

$\hat{y}_1 = 4,22 \mp 0,24X$

Lç	X	Y
01	0	1
02	●	●
03	2	4
04	3	6
05	5	8
06	7	9
07	8	8
08	9	1

$\hat{y}_2 = 3,66 \mp 0,33X$

Lç	X	Y
01	0	1
02	1	2
03	●	●
04	3	6
05	5	8
06	7	9
07	8	8
08	9	1

$\hat{y}_3 = 3,00 \mp 0,42X$

Lç	X	Y
01	0	1
02	1	2
03	2	4
04	●	●
05	5	8
06	7	9
07	8	8
08	9	1

$\hat{y}_4 = 2,64 \mp 0,45X$

Lç	X	Y
01	0	1
02	1	2
03	2	4
04	3	6
05	●	●
06	7	9
07	8	8
08	9	1

$\hat{y}_5 = 2,74 \mp 0,39X$

Lç	X	Y
01	0	1
02	1	2
03	2	4
04	3	6
05	5	8
06	●	●
07	8	8
08	9	1

$\hat{y}_6 = 3,20 \mp 0,29X$

Lç	X	Y
01	0	1
02	1	2
03	2	4
04	3	6
05	5	8
06	7	9
07	●	●
08	9	1

$\hat{y}_7 = 3,21 \mp 0,31X$

Lç	X	Y
01	0	1
02	1	2
03	2	4
04	3	6
05	5	8
06	7	9
07	8	8
08	●	●

$\hat{y}_8 = 1,81 \mp 0,97X$

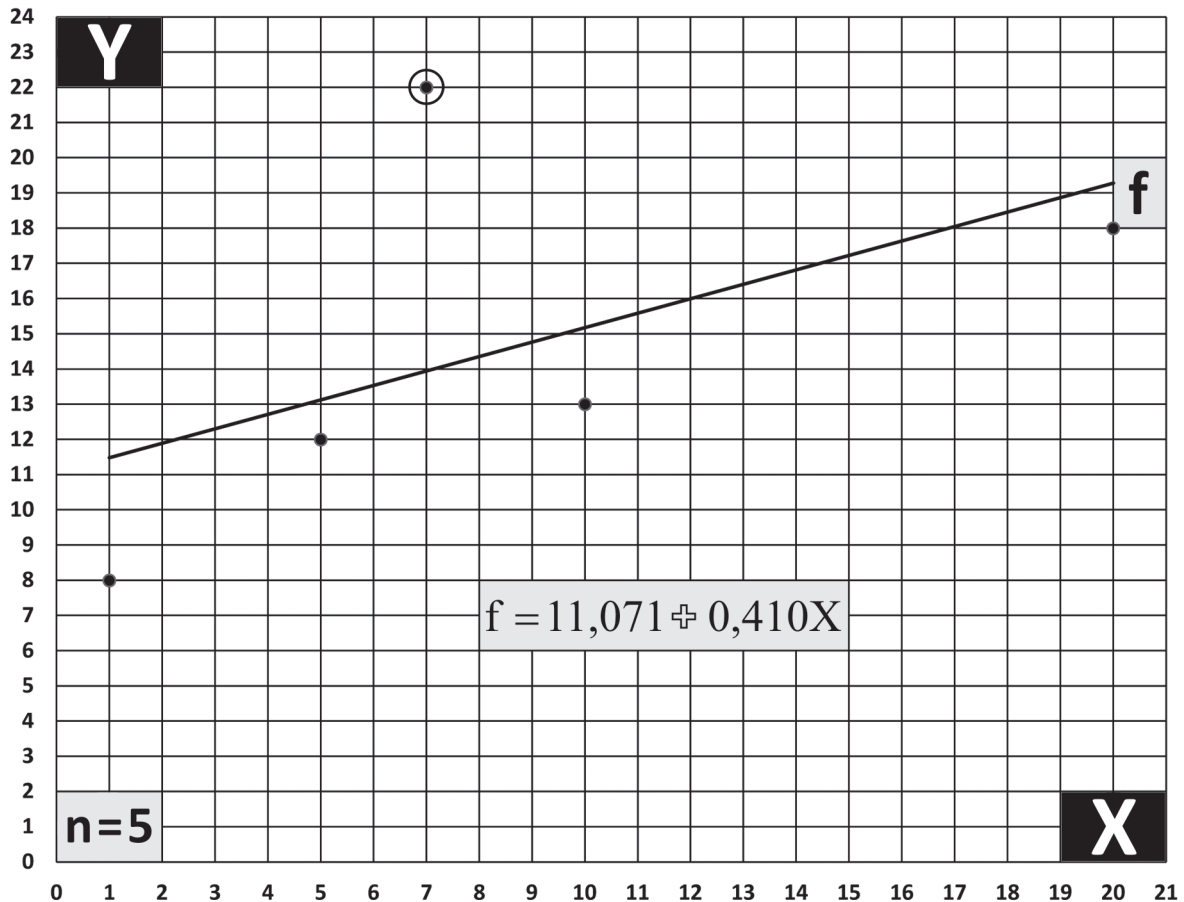
## Simulação Computacional dos Parâmetros $b_0$ e $b_1$ de $\hat{Y}$



# 6.1 Distância de Cook e Alavancagem Prima

LÇ	X	Y	f	E (i)
01	1	8	11,481	12,117
02	5	12	13,121	1,257
03	7	22	13,941	64,947
04	10	13	15,171	4,713
05	20	18	19,271	1,615
<b>MÉDIA (X) = 8,600</b>			<b>SOMA</b>	<b>84,650</b>
			<b>MSE</b>	<b>16,930</b>

ALAVANCAS		C	DISTÂNCIAS	
H1	0,481		O O K	D1
H2	0,263	D2		0,027
H3	0,212	<b>D3</b>		<b>1,035</b>
H4	0,210	D4		0,074
H5	0,833	D5		0,477
<b>PONTO DISCORDANTE</b>				



**Erro Quadrático Médio (MSE)**

$$MSE = \frac{1}{n} \cdot \left[ \sum_{i=1}^n (Y_i - f_i)^2 \right] = \frac{1}{n} \cdot \left[ \sum_{i=1}^n E(i) \right]$$

**Alavancas (Hi) e Distâncias (Di) das Lições**

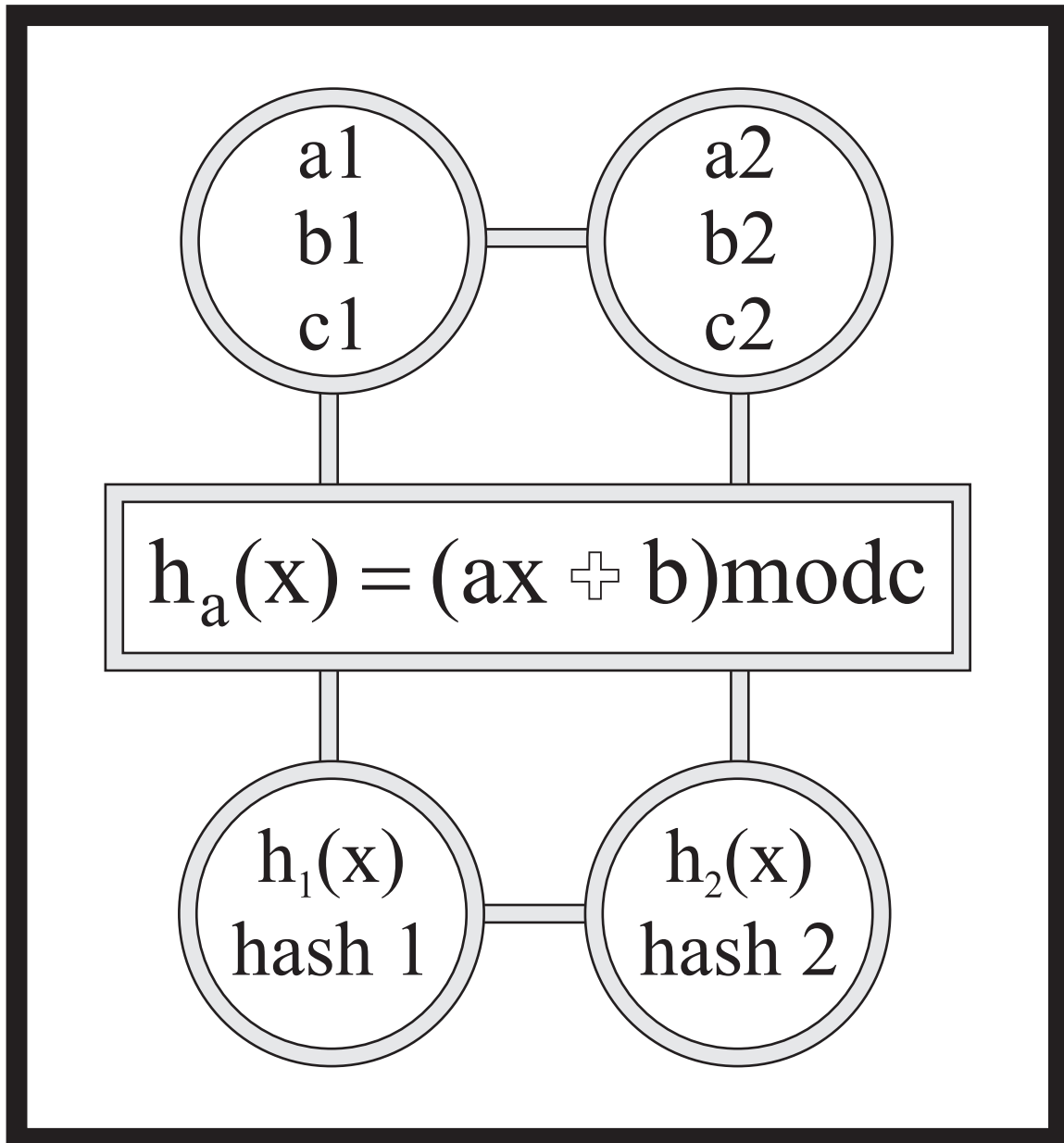
$$H_i = \frac{1}{n} + \frac{(X_i - \bar{X})^2}{\sum_{j=1}^n (X_j - \bar{X})^2} \quad \Bigg| \quad D_i = \frac{E^2(i) \cdot H_i}{MSE \cdot (1 - H_i)}$$

**A Lição (03) é Discordante do Banco de Dados**



# Ω-ALGORITMO

Flajolet - Martin



Função Hash

Cardinalidade Estatística

Prof. Aguinaldo Prandini Ricieri



# DISQUISITIONES ARITHMETICAE

## SECTIO PRIMA

DE

NVMERORVM CONGRVENTIA IN GENERE.

1. Si numerus  $a$  numerorum  $b, c$  differentiam metitur,  $b$  et  $c$  secundum  $a$  congrui dicuntur, sin minus, incongrui: ipsum  $a$  **modulum** appellatus. Uterque numerorum  $b, c$ , priori in casu alterius **residuum**, in posteriori vero *nonresiduum* vocatur.

Hae notiones de omnibus numeris integris tam positivis quam negativis \*) valent, neque

\*) Modulus manifesto semper absolute i. e. sine omni signo est sumendus.

A



# 4.0 Estimativa Estatística da Constante FM

Formatação de um Experimento Computacional Didático

$$h(x) = (ax \oplus b) \text{mod } c \quad \text{ED} = \frac{2^d}{\Omega} \quad h(x) = (5x \oplus 1) \text{mod } 11$$

$x = 1, 8, 2, 19, 7, 22, 3, 6, 7, 22$   
 $h(x) = 6, 7, 0, 8, 3, 1, 5, 9, 3, 1$   
 $B(h) = 0110, 0111, 0000, 1000, \dots, 0001$

$d = 3, \text{ ED} = 8, \Omega = 1,00$

Alguns dos Milhares de Sorteios Realizados

d	ED	FM
3	8	1,00
2	7	0,57
4	6	2,66
3	9	0,88
2	8	0,50
1	5	0,40
3	4	2,00
2	6	0,66

$d = 13$   
 $ED = 10639$   
 $\Omega = 0,769$

$d = 15$   
 $ED = 42336$   
 $\Omega = 0,773$

$d = 17$   
 $ED = 169783$   
 $\Omega = 0,772$

$d = 13$   
 $ED = 85002$   
 $\Omega = 0,770$

d	ED	FM
7	157	0,76
5	38	0,84
8	367	0,69
9	583	0,87
6	79	0,81
8	382	0,67
9	688	0,74
8	390	0,65

Varia-se os Parâmetros (a, b, c) do Hashe e as Quantidades de x do Conjunto



# α-DIVERGENCE

## Produto Interno

0,10	0,02	0,08
0,05	0,15	0,10
0,05	0,25	0,20

1

0,15	0,10	0,05
0,15	0,05	0,10
0,20	0,05	0,15

1

0,10	0,05	0,05
0,20	0,10	0,15
0,20	0,10	0,05

$$D_{\alpha}(P \parallel Q) = \frac{1}{\alpha - 1} \log \left( \sum_{i=1}^n \frac{p_i^{\alpha}}{q_i^{\alpha-1}} \right)$$
  
  

0,05	0,20	0,10
0,10	0,20	0,15
0,03	0,07	0,10

1

0,20	0,15	0,04
0,06	0,20	0,05
0,10	0,10	0,10

1

0,30	0,05	0,05
0,15	0,15	0,10
0,05	0,10	0,05

Divergência de Rényi

Bhattacharyya Distance

Prof. Aginaldo Prandini Ricieri



# Alfa - Divergência

## 1.0 Classificador de Dados Rényi

A divergência de Rényi (1960) trata-se de uma medida nas ciências estatísticas, oferecendo uma abordagem única que quantifica a diferença ou semelhança entre distribuições de probabilidades dos mais diversos fenômenos. Essa análise feita em grandes bancos de dados (condição da aplicação:  $\alpha > 0$ ,  $\alpha \neq 1$ ) engloba a distância || de Hellinger (1909) e a distância || de Kullback-Leibler (1951) :

$$H(P \parallel Q) = \frac{1}{\sqrt{2}} \sqrt{\sum_{i=1}^k (\sqrt{p_i} - \sqrt{q_i})^2} \quad KL(P \parallel Q) = \sum_{i=1}^n P_i \circ \text{Log} \frac{p_i}{q_i}$$

**Distâncias de Hellinger e de Kullback - Libler ( $p_i$ ,  $q_i$ : Probabilidades)**

Devido à sua capacidade de apontar nuances nas diferenças entre as distribuições estatísticas, a divergência alfa de Rényi como classificador de dados é hoje aplicada em diversas áreas da estatística como transmissão de sinais, internet das coisas e aprendizado de máquina, isso graças às facilidades computacionais.

A síntese matemática de Rényi é uma união entre a distância de Anil Kumar Bhattacharyya (1943) e a entropia de Claude Shannon (1948) alinhavadas com o produto interno  $\langle \vec{P}, \vec{Q} \rangle$  e as dimensões geométricas generalizadas  $D_\alpha(P, Q)$ .

P

$$D_{1/2}(P \parallel Q) = -2 \text{Log} \left( \sum_{i=1}^n \sqrt{p_i \circ q_i} \right)$$

**Expressão de Bhattacharyya ( $\alpha = 1/2$ )**

Q

*On Measures of Entropy and Information*

*Proceedings of the 4th Berkeley Symposiun, p. 547, volume 1, 1960*

*Alfréd Rényi (1921 - 1970)*





# 6.0 Classificação de Dados Estatísticos **A**

<b>A</b>	0,120	0,080
0,050	0,150	0,100
0,050	0,250	0,200

<b>P</b>	p1	p2
p3	p4	p5
p6	p7	p8

D (X || A)  
**0,1714**  
DISTÂNCIA DE RÉNYI

<b>B</b>	0,250	0,050
0,150	0,050	0,100
0,200	0,050	0,150

D (X || B)  
**0,1268**  
DISTÂNCIA DE RÉNYI

<b>C</b>	0,150	0,150
0,200	0,100	0,050
0,200	0,100	0,050

D (X || C)  
**0,1200**  
DISTÂNCIA DE RÉNYI

<b>D</b>	0,100	0,070
0,100	0,200	0,150
0,030	0,250	0,100

D (X || D)  
**0,1229**  
DISTÂNCIA DE RÉNYI

<b>E</b>	0,020	0,050
0,060	0,350	0,040
0,100	0,100	0,100

D (X || E)  
**0,4826**  
DISTÂNCIA DE RÉNYI

<b>F</b>	0,350	0,050
0,150	0,150	0,100
0,050	0,100	0,050

D (X || F)  
**0,0048**  
DISTÂNCIA DE RÉNYI

<b>G</b>	0,100	0,200
0,150	0,050	0,100
0,150	0,200	0,050

D (X || G)  
**0,1931**  
DISTÂNCIA DE RÉNYI

<b>H</b>	0,100	0,100
0,050	0,150	0,100
0,100	0,350	0,050

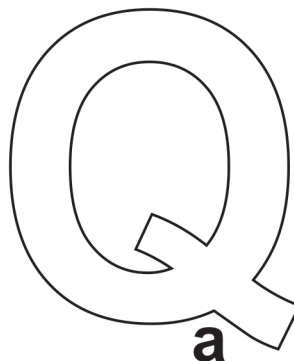
D (X || H)  
**0,1990**  
DISTÂNCIA DE RÉNYI

<b>I</b>	0,100	0,100
0,020	0,150	0,200
0,300	0,080	0,050

D (X || I)  
**0,2765**  
DISTÂNCIA DE RÉNYI



PERTENCIMENTO 1		
<b>X</b>	0,300	0,060
0,170	0,180	0,095
0,050	0,090	0,055
CLASSIFICAÇÃO DE X		

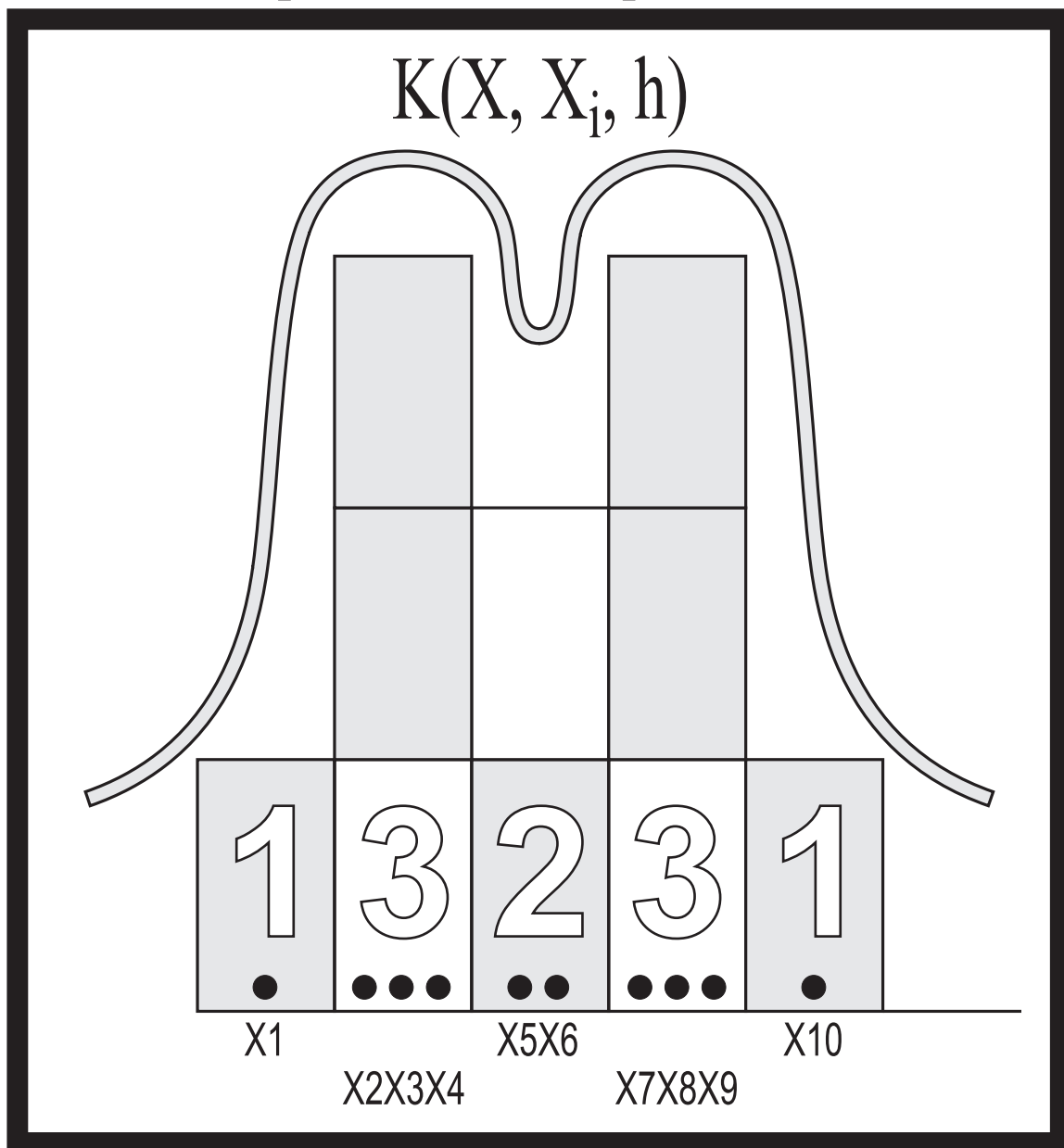


**A MATRIZ DE DADOS (X) É CLASSIFICADA NO GRUPO F: MENOR DISTÂNCIA (D)**



# C-KERNEL

## Mapas Coropléticos



Nonparametric Estimation

Densidade de Probabilidade

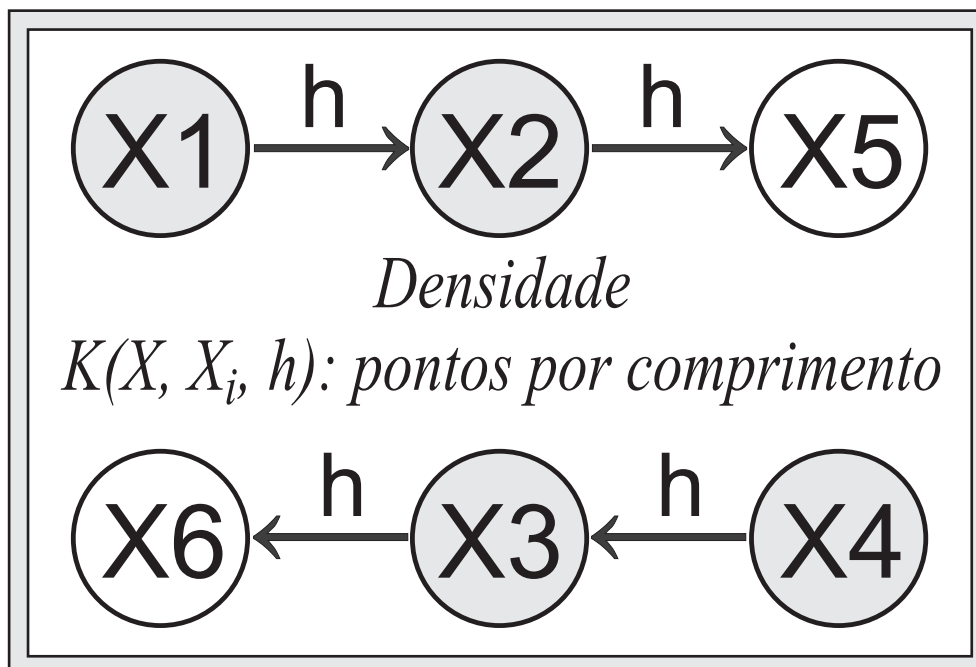
Prof. Aguinaldo Prandini Ricieri



# Estimador Kernel

## 1.0 Função Densidade de Probabilidade

O Kernel como núcleo gerenciador de dados – estatística não paramétrica – permite que se estime valores, por exemplo  $X_5$  e  $X_6$ , através de curvas de densidades de probabilidades  $f(X, X_i, h)$ , modeladas com base em amostras colhidas a certa distância – valores computacionalmente definidos de  $h$  –, que fazem o papel de núcleos estatísticos ou também pontos de ancoragem  $X_i(X_1, X_2, X_3, X_4)$ .



Núcleos  $X_i(X_1, X_2, X_3, X_4)$  para Estimativas dos Valores  $X(X_5, X_6)$

O estimador Kernel permite inferências sobre populações a partir de amostras espacialmente referenciadas quando é conhecido o adensamento dos dados.

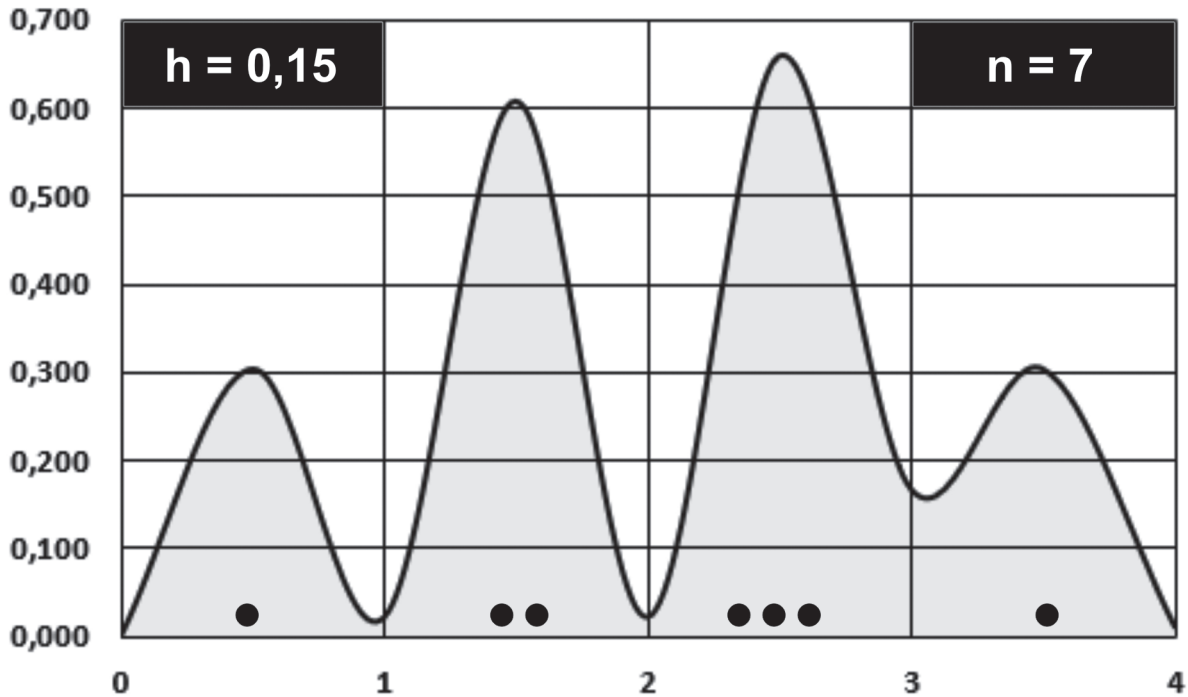
① *On Estimation of a Probability Density Function*  
*The Annals of Mathematical Statistics (33), p. 1065, 1962*  
Emanuel Parzer

② *Nonparametric Density Estimation via Diffusion Mixing*  
*University of Queensland, 2007*  
Zdravko Botev

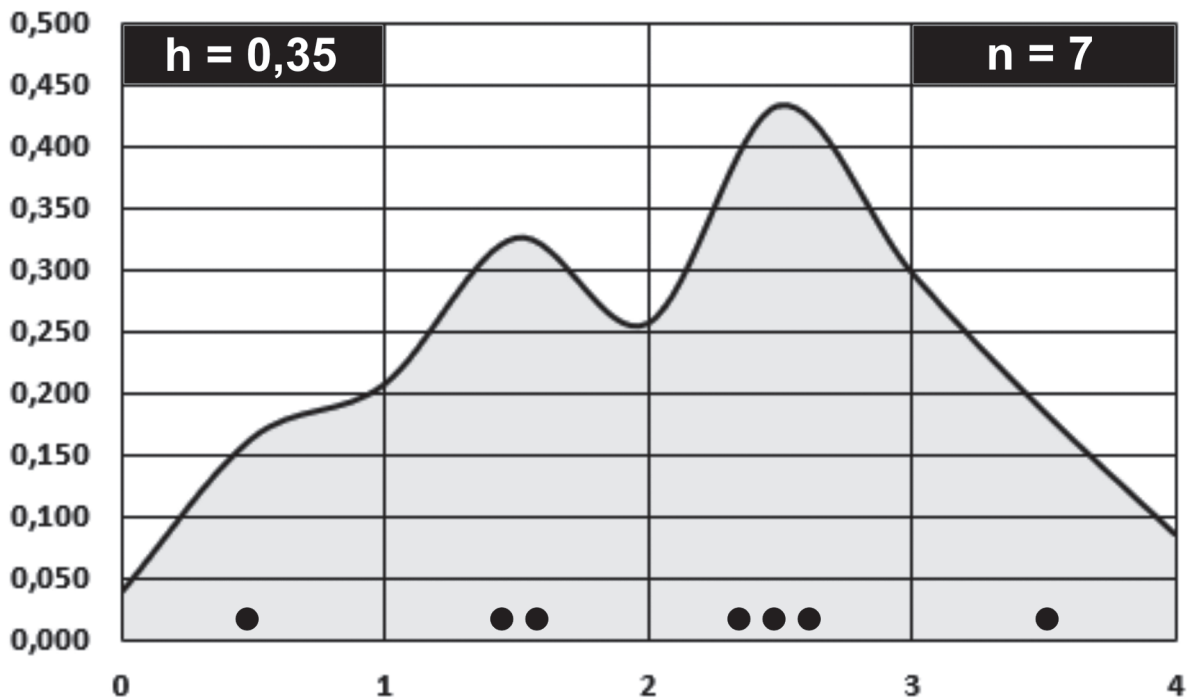


# 4.1 Função Densidade de Probabilidade Prima

$$X_i = \{0,6 \mid 1,4 \mid 1,6 \mid 2,4 \mid 2,6 \mid 2,8 \mid 3,6\}$$



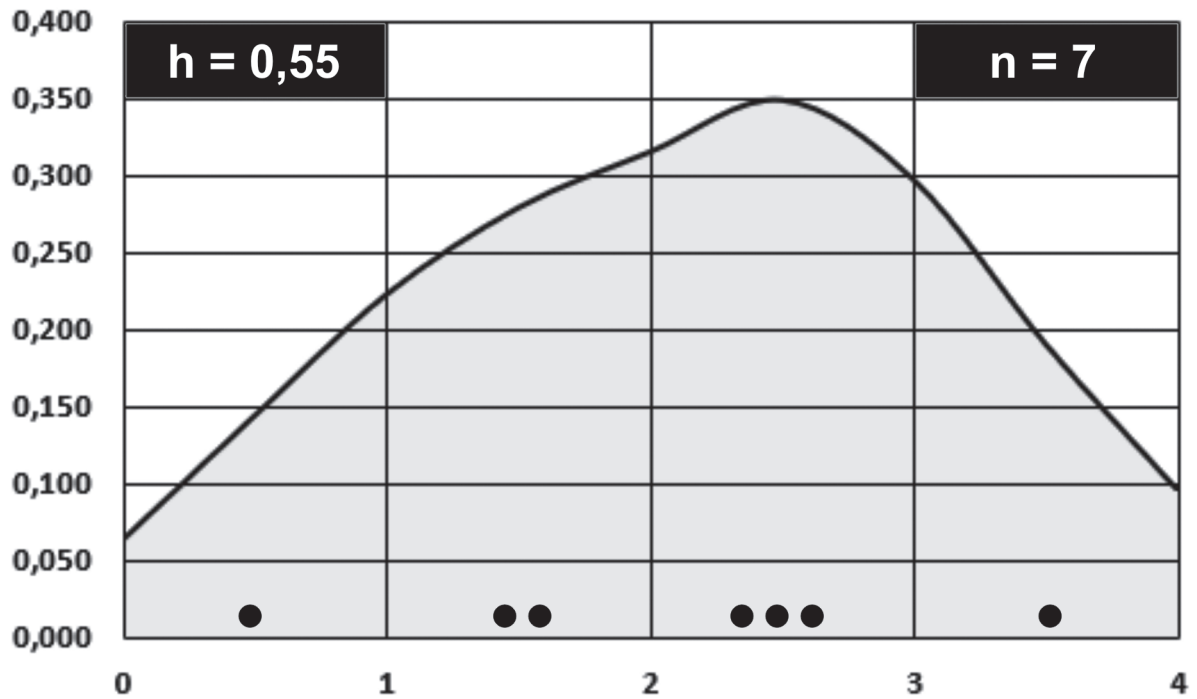
$$f(x, h) = \frac{0,0567}{h} \sum_{i=1}^7 e^{-0,5 \left( \frac{x - x_i}{h} \right)^2}$$





# 4.2 Função Densidade de Probabilidade Prima

$$X_i = \{0,6 \mid 1,4 \mid 1,6 \mid 2,4 \mid 2,6 \mid 2,8 \mid 3,6\}$$



$$f(x, h) = \frac{0,0567}{h} \sum_{i=1}^7 e^{-0,5 \left( \frac{x - X_i}{h} \right)^2}$$

