

MBA  
USP  
ESALQ

# Unsupervised Machine Learning: Análise de Correspondências Simples e Múltiplas

Rafael de Freitas Souza

A LEGO minifigure of a man with curly hair and a beard, wearing a backpack and holding a map, standing in a field. The minifigure is smiling and has a backpack on its back. It is holding a map that shows a compass rose and some terrain lines. The background is a blurred field with some dry grass. A semi-transparent dark grey banner is at the bottom of the image.

# Introdução

# Introdução às Análises de Correspondências Simples e Múltiplas

As Análises de Correspondências, sejam elas simples ou múltiplas, são técnicas adequadas para se trabalhar com dados que se manifestam de forma qualitativa, não possuindo a intenção de se fazer previsões para observações não presentes na amostra.

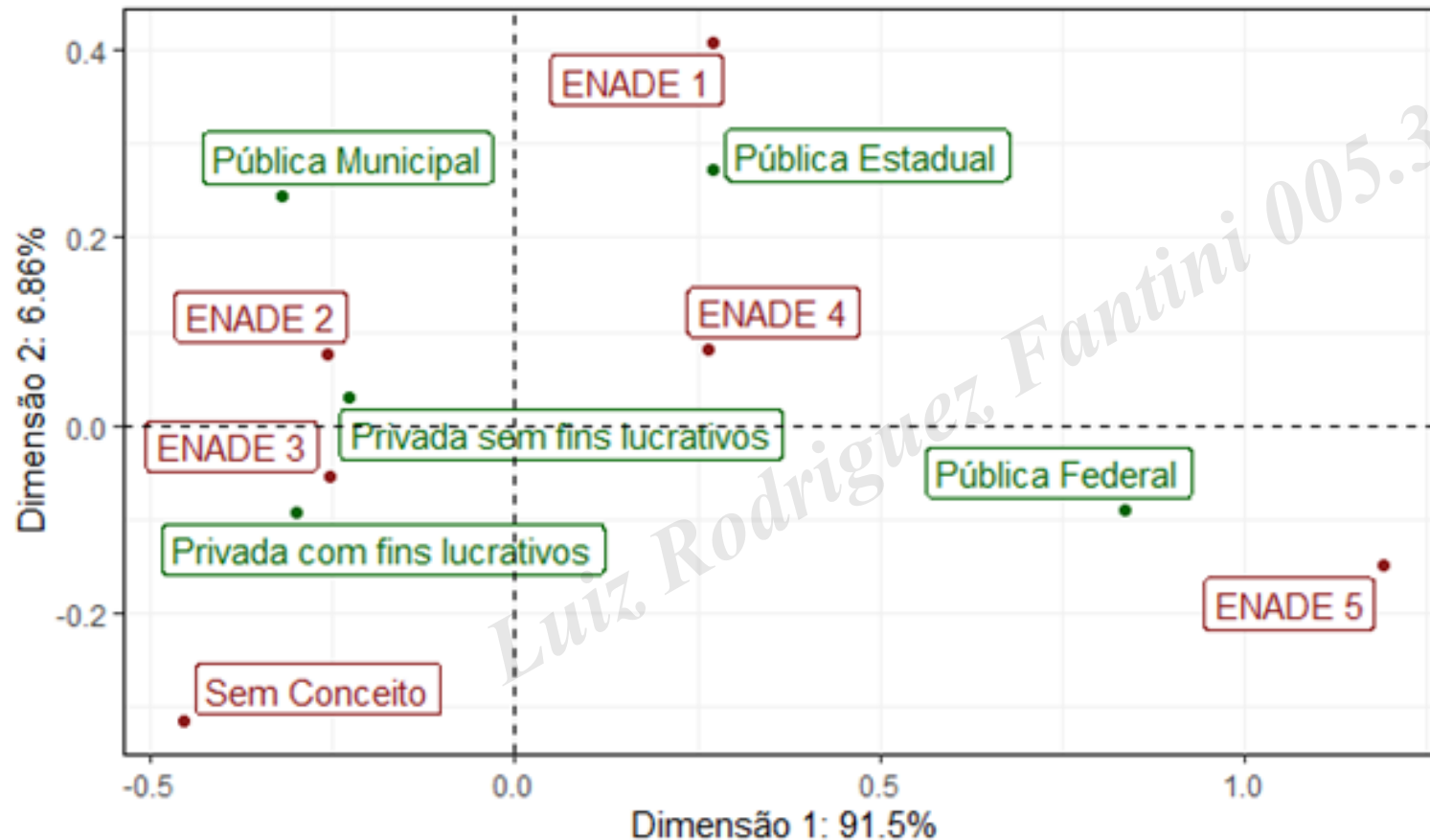
Quando o número de variáveis de interesse for igual a 2, utilizaremos as Análises de Correspondências Simples (ANACOR); por outro lado, quando o número de variáveis de interesse for maior do que 2, utilizaremos as Análises de Correspondências Múltiplas (ACM).

# Ideia e Objetivos das Técnicas

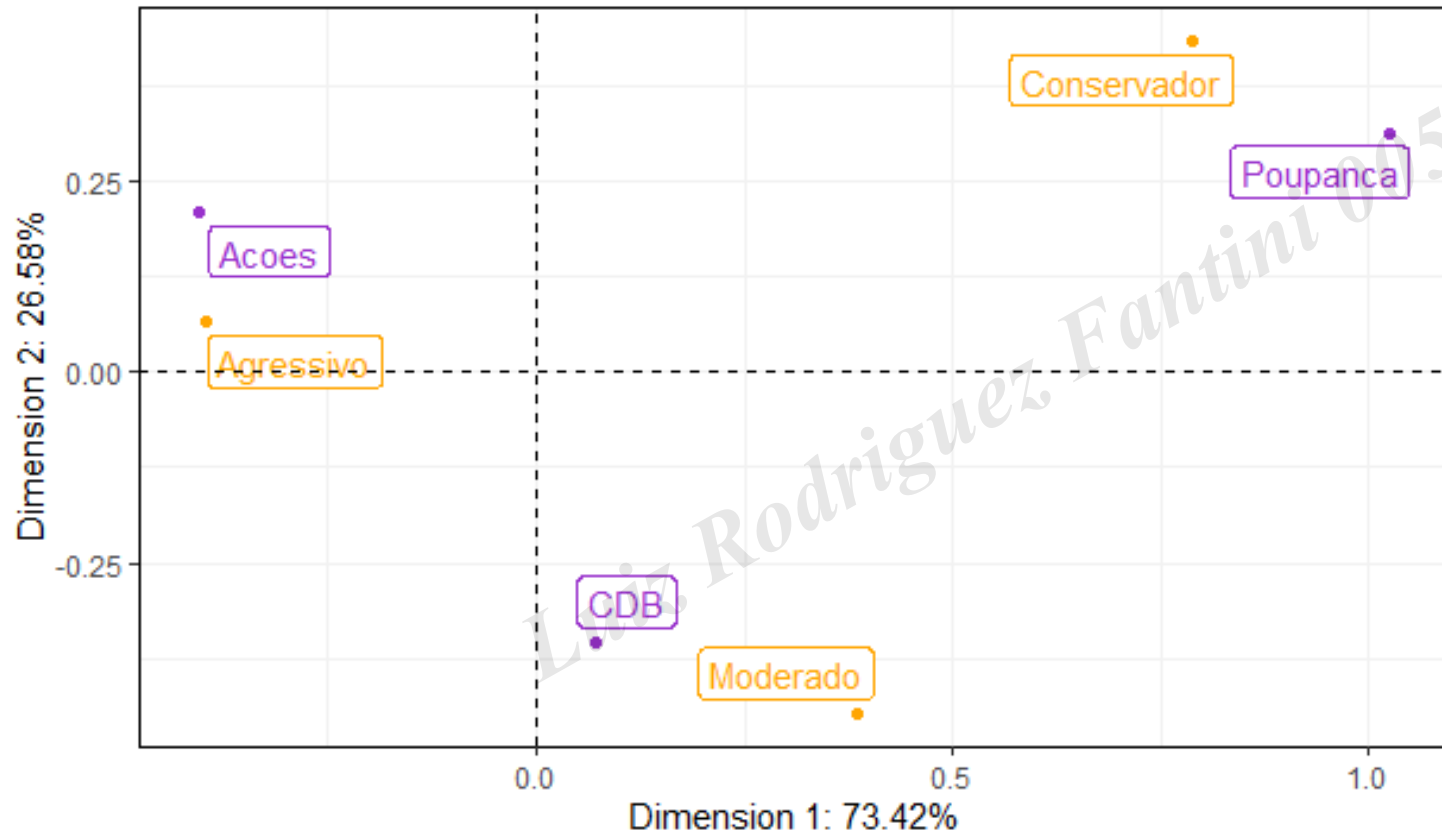
A ideia, portanto, é a de se estudar as relações de interdependência em razão das associações entre as categorias das variáveis de interesse.

O grande objetivo das Análises de Correspondências é o estabelecimento de um mapa perceptual. O mapa perceptual é uma espécie de gráfico que utilizará coordenadas que representarão as linhas e as colunas de uma tabela de contingências.

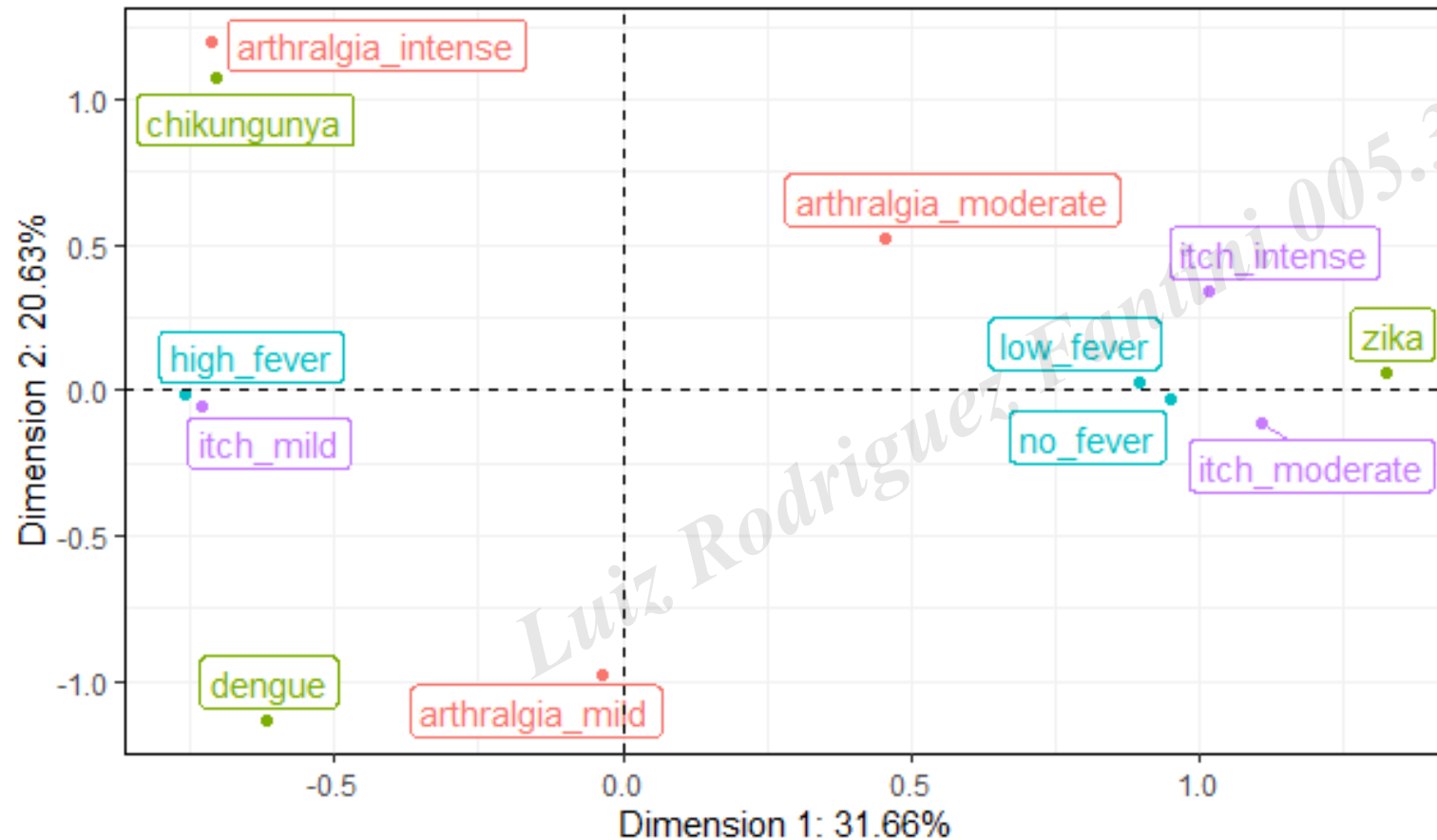
# Exemplo de um Mapa Perceptual de uma ANACOR



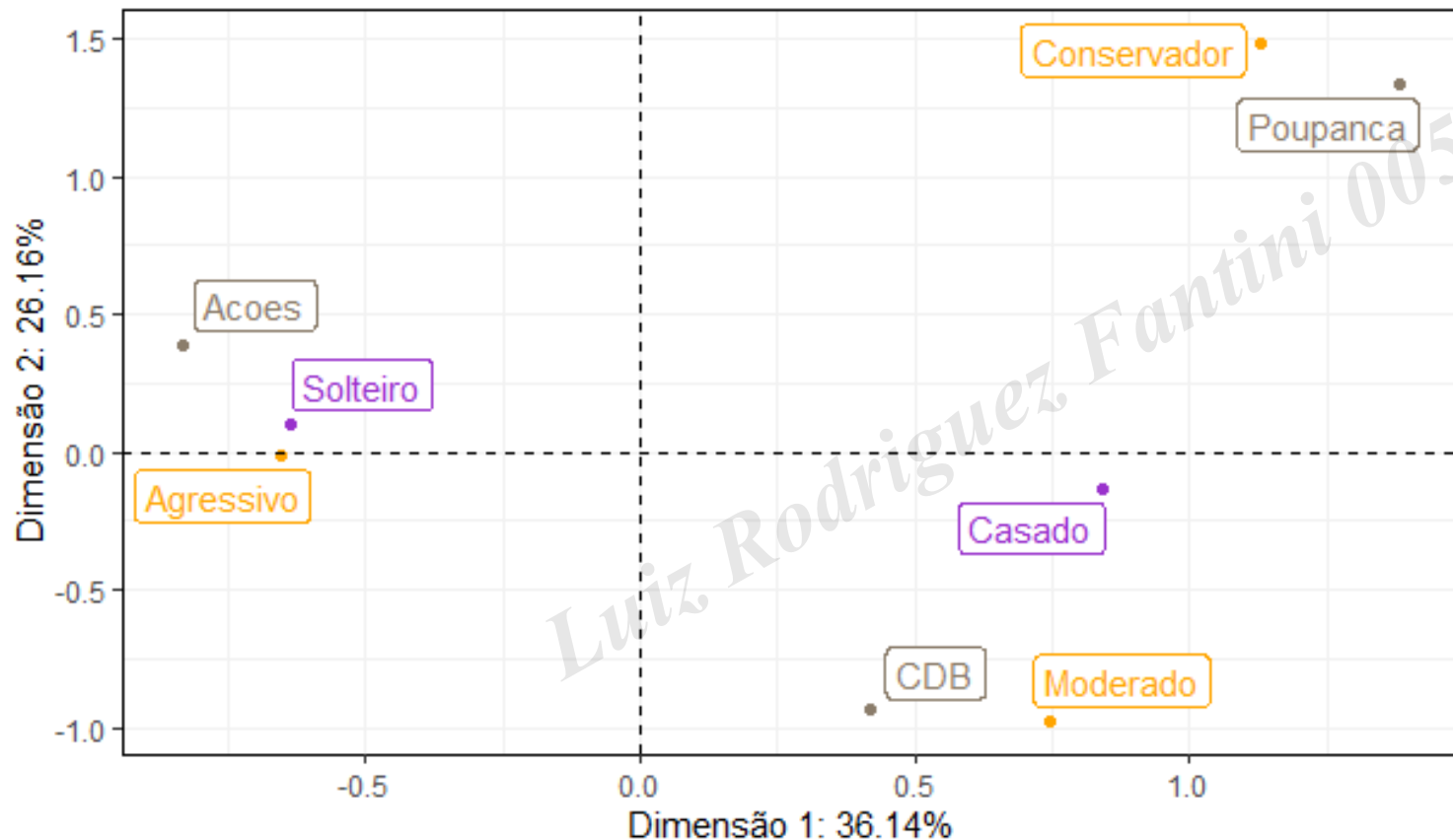
# Exemplo de um Mapa Perceptual de uma ANACOR



# Exemplo de um Mapa Perceptual de uma ACM



# Exemplo de um Mapa Perceptual de uma ACM







005.374.619-81

## Análise de Correspondência Simples (ANACOR)





# Análise de Correspondência Simples (ANACOR)

PRIMEIRO PASSO

A Construção de uma Tabela de Frequências Observadas

# *A construção de uma Tabela de Contingências de Valores Observados:*

Considerando-se, portanto, duas variáveis categóricas, em que a primeira possui  $I$  categorias, e a segunda possui  $J$  categorias, devemos estabelecer uma tabela de contingências  $\mathbf{X}_0$ .

A matriz  $\mathbf{X}_0$  conterá as frequências absolutas observadas das categorias  $I$  e  $J$ , em que cada célula  $ij$  possui certa quantidade  $n_{ij}$  ( $i = 1, 2, 3, \dots, I$  e  $j = 1, 2, 3, \dots, J$ ) de observações.

# A Tabela de Contingências de Valores Observados

O total de observações  $N$  que farão parte do estudo é ser dada por  $N = \sum_{i=1}^I \sum_{j=1}^J n_{ij}$ . Assim, um exemplo teórico de uma tabela de contingências, seria:

	1	2	...	$J$	Total
1	$n_{11}$	$n_{12}$	...	$n_{1J}$	$\sum l_1$
2	$n_{21}$	$n_{22}$		$n_{2J}$	$\sum l_1$
$\vdots$	$\vdots$	$\vdots$		...	$\vdots$
$I$	$n_{I1}$	$n_{I2}$		$n_{IJ}$	$\sum l_i$
Total	$\sum c_1$	$\sum c_2$	...	$\sum c_J$	$N$

*Construída a Tabela de Contingências, A Matriz  $\mathbf{X}_0$ , será:*

$$\mathbf{X}_0 = \begin{bmatrix} n_{11} & n_{12} & \cdots & n_{1J} \\ n_{21} & n_{22} & \cdots & n_{2J} \\ \vdots & \vdots & \ddots & \vdots \\ n_{I1} & n_{I2} & \cdots & n_{IJ} \end{bmatrix}$$

# Análise de Correspondência Simples (ANACOR)

## SEGUNDO PASSO

# O Estudo das Associações entre as Categorias das Variáveis de Interesse



# *O estudo das associações entre as categorias das variáveis:*

Para o estudo das associações entre as categorias das variáveis, utilizaremos dois instrumentos: o Teste  $\chi^2$  e a análise dos resíduos padronizados ajustados.

O Teste  $\chi^2$  estudará se as associações entre as categorias das variáveis se associam, ou não, de forma aleatória; enquanto a análise dos resíduos padronizados ajustados revelará os padrões característicos de cada categoria de uma variável segundo o excesso ou falta de ocorrências de sua combinação com cada categoria de outra variável.

Logo, deveremos estabelecer uma matriz de valores esperados e, a seguir, uma matriz de resíduos – isso bastará para o estabelecimento do Teste  $\chi^2$ .

No caso da análise dos resíduos padronizados ajustados, deveremos utilizar a matriz de resíduos para definirmos uma matriz de resíduos padronizados. Após isso, conseguiremos estabelecer uma matriz de resíduos padronizados ajustados.

# Os Valores Esperados:

Nós sabemos que o somatório entre os valores  $\sum c_1 + \sum c_2 + \dots + \sum c_J = \sum l_1 + \sum l_2 + \dots + \sum l_I = N$ . Então as frequências esperadas, serão:

	<b>1</b>	<b>2</b>	...	<b>J</b>
<b>1</b>	$\left( \frac{\sum c_1 \times \sum l_1}{N} \right)$	$\left( \frac{\sum c_2 \times \sum l_1}{N} \right)$	...	$\left( \frac{\sum c_J \times \sum l_1}{N} \right)$
<b>2</b>	$\left( \frac{\sum c_1 \times \sum l_2}{N} \right)$	$\left( \frac{\sum c_2 \times \sum l_2}{N} \right)$		$\left( \frac{\sum c_J \times \sum l_2}{N} \right)$
$\vdots$	$\vdots$	$\vdots$		
<b>I</b>	$\left( \frac{\sum c_1 \times \sum l_I}{N} \right)$	$\left( \frac{\sum c_2 \times \sum l_I}{N} \right)$		$\left( \frac{\sum c_J \times \sum l_I}{N} \right)$



# A Matriz $\mathbf{X_E}$ de Valores Esperados:

$$\mathbf{X_E} = \begin{bmatrix} \left( \frac{\sum c_1 \times \sum l_1}{N} \right) & \left( \frac{\sum c_2 \times \sum l_1}{N} \right) & \dots & \left( \frac{\sum c_J \times \sum l_1}{N} \right) \\ \left( \frac{\sum c_1 \times \sum l_2}{N} \right) & \left( \frac{\sum c_2 \times \sum l_2}{N} \right) & \dots & \left( \frac{\sum c_J \times \sum l_2}{N} \right) \\ \vdots & \vdots & \ddots & \vdots \\ \left( \frac{\sum c_1 \times \sum l_I}{N} \right) & \left( \frac{\sum c_2 \times \sum l_I}{N} \right) & \dots & \left( \frac{\sum c_J \times \sum l_I}{N} \right) \end{bmatrix}$$

# A Matriz **R** dos Resíduos

Entenderemos por resíduo, a diferença entre os valores observados e os valores esperados. Assim sendo, a matriz **R** é dada por:

$$\mathbf{R} = \begin{bmatrix} n_{11} - \left( \frac{\sum c_1 \times \sum l_1}{N} \right) & n_{12} - \left( \frac{\sum c_2 \times \sum l_1}{N} \right) & \cdots & n_{1J} - \left( \frac{\sum c_J \times \sum l_1}{N} \right) \\ n_{21} - \left( \frac{\sum c_1 \times \sum l_2}{N} \right) & n_{22} - \left( \frac{\sum c_2 \times \sum l_2}{N} \right) & \cdots & n_{2J} - \left( \frac{\sum c_J \times \sum l_2}{N} \right) \\ \vdots & \vdots & \ddots & \vdots \\ n_{I1} - \left( \frac{\sum c_1 \times \sum l_I}{N} \right) & n_{I2} - \left( \frac{\sum c_2 \times \sum l_I}{N} \right) & \cdots & n_{IJ} - \left( \frac{\sum c_J \times \sum l_I}{N} \right) \end{bmatrix}$$

## O Teste $\chi^2$ :

$$\chi^2 = \sum_{i=1}^I \sum_{j=1}^J \frac{\left[ n_{ij} - \left( \frac{\sum c_j \times \sum l_i}{N} \right) \right]^2}{\left( \frac{\sum c_j \times \sum l_i}{N} \right)}, \text{ com } (I - 1) \times (J - 1) \text{ graus de liberdade.}$$

São hipóteses do Teste  $\chi^2$ :

$H_0$ : as duas variáveis categóricas se associam de forma aleatória;

$H_1$ : as duas variáveis categóricas não se associam de forma aleatória.

# Os Resíduos Padronizados:

Os resíduos padronizados são dados por:  $r_{padronizado_{ij}} = \frac{n_{ij} - ne_{ij}}{\sqrt{ne_{ij}}}$ . Então a matriz  $\mathbf{R}_{padronizados}$  é descrita por:

$$\mathbf{R}_{padronizados} = \begin{bmatrix} \frac{n_{11} - \left(\frac{\sum c_1 \cdot \sum l_1}{N}\right)}{\sqrt{\left(\frac{\sum c_1 \cdot \sum l_1}{N}\right)}} & \frac{n_{12} - \left(\frac{\sum c_2 \cdot \sum l_1}{N}\right)}{\sqrt{\left(\frac{\sum c_2 \cdot \sum l_1}{N}\right)}} & \dots & \frac{n_{1J} - \left(\frac{\sum c_J \cdot \sum l_1}{N}\right)}{\sqrt{\left(\frac{\sum c_J \cdot \sum l_1}{N}\right)}} \\ \frac{n_{21} - \left(\frac{\sum c_1 \cdot \sum l_2}{N}\right)}{\sqrt{\left(\frac{\sum c_1 \cdot \sum l_2}{N}\right)}} & \frac{n_{22} - \left(\frac{\sum c_2 \cdot \sum l_2}{N}\right)}{\sqrt{\left(\frac{\sum c_2 \cdot \sum l_2}{N}\right)}} & \dots & \frac{n_{2J} - \left(\frac{\sum c_J \cdot \sum l_2}{N}\right)}{\sqrt{\left(\frac{\sum c_J \cdot \sum l_2}{N}\right)}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{n_{I1} - \left(\frac{\sum c_1 \cdot \sum l_I}{N}\right)}{\sqrt{\left(\frac{\sum c_1 \cdot \sum l_I}{N}\right)}} & \frac{n_{I2} - \left(\frac{\sum c_2 \cdot \sum l_I}{N}\right)}{\sqrt{\left(\frac{\sum c_2 \cdot \sum l_I}{N}\right)}} & \dots & \frac{n_{IJ} - \left(\frac{\sum c_J \cdot \sum l_I}{N}\right)}{\sqrt{\left(\frac{\sum c_J \cdot \sum l_I}{N}\right)}} \end{bmatrix}$$

# Os Resíduos Padronizados Ajustados:

Então, os resíduos padronizados ajustados podem ser calculados da seguinte maneira:  $r_{padronizado\ ajustado_{ij}} = \frac{r_{padronizado_{ij}}}{\sqrt{\left(1 - \frac{\sum c_j}{N}\right) \cdot \left(1 - \frac{\sum l_i}{N}\right)}}$ . Logo, a matriz

$R_{padronizados\ ajustados}$  será dada por:

$$R_{padronizados\ ajustados} = \begin{bmatrix} \frac{r_{padronizado_{11}}}{\sqrt{\left(1 - \frac{\sum c_1}{N}\right) \cdot \left(1 - \frac{\sum l_1}{N}\right)}} & \frac{r_{padronizado_{12}}}{\sqrt{\left(1 - \frac{\sum c_2}{N}\right) \cdot \left(1 - \frac{\sum l_1}{N}\right)}} & \dots & \frac{r_{padronizado_{1J}}}{\sqrt{\left(1 - \frac{\sum c_J}{N}\right) \cdot \left(1 - \frac{\sum l_1}{N}\right)}} \\ \frac{r_{padronizado_{21}}}{\sqrt{\left(1 - \frac{\sum c_1}{N}\right) \cdot \left(1 - \frac{\sum l_2}{N}\right)}} & \frac{r_{padronizado_{22}}}{\sqrt{\left(1 - \frac{\sum c_2}{N}\right) \cdot \left(1 - \frac{\sum l_2}{N}\right)}} & \dots & \frac{r_{padronizado_{2J}}}{\sqrt{\left(1 - \frac{\sum c_J}{N}\right) \cdot \left(1 - \frac{\sum l_2}{N}\right)}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{r_{padronizado_{I1}}}{\sqrt{\left(1 - \frac{\sum c_1}{N}\right) \cdot \left(1 - \frac{\sum l_I}{N}\right)}} & \frac{r_{padronizado_{I2}}}{\sqrt{\left(1 - \frac{\sum c_2}{N}\right) \cdot \left(1 - \frac{\sum l_I}{N}\right)}} & \dots & \frac{r_{padronizado_{IJ}}}{\sqrt{\left(1 - \frac{\sum c_J}{N}\right) \cdot \left(1 - \frac{\sum l_I}{N}\right)}} \end{bmatrix}$$



The background of the slide features a blurred image of a financial market display. It includes various stock indices such as OMXC25, OMXRGI, and OMXI8, along with their respective values and percentage changes. A line chart is also visible, showing a fluctuating trend. The overall color scheme is dominated by blue and red, typical of financial data visualizations.

# Análise de Correspondência Simples (ANACOR)

TERCEIRO PASSO

A Decomposição da Inércia Principal Total

# A Inércia Principal Total $I_T$ :

O Teste  $\chi^2$  consegue apontar se associação entre as categorias de dadas variáveis se dá, ou não, de forma aleatória.

Porém, o Teste  $\chi^2$  aumenta à medida que a amostra aumenta. Então, no lugar de decompor o  $\chi^2$  para realizarmos a análise de correspondência, o padronizaremos em razão do tamanho amostra para, então, decompô-lo.

A inércia principal total é dada por:  $I_T = \frac{\chi^2}{N}$ .

A decomposição inercial para a elaboração da análise de correspondência perpassa pela extração dos *eigenvalues* de uma matriz  $\mathbf{W}$ , dada por  $\mathbf{W} = \mathbf{A}'\mathbf{A}$ , em que  $\mathbf{A} = \mathbf{D}_l^{-\frac{1}{2}} \cdot (\mathbf{P} - lc') \cdot \mathbf{D}_c^{-\frac{1}{2}}$ .

# A Determinação dos Eigenvalues para a Decomposição Inercial:

O método de decomposição dos autovalores, tradicionalmente, é o método de Eckart-Young, em que quantidade  $m$  de *eigenvalues* é dada por  $m = \min(I - 1, J - 1)$ .

O passo inicial para a determinação dos autovalores é a construção de uma matriz **P** de frequências relativas observadas.

	1	2	...	$J$	Total
1	$\frac{n_{11}}{N}$	$\frac{n_{12}}{N}$	...	$\frac{n_{1J}}{N}$	$\frac{\sum l_1}{N}$
2	$\frac{n_{21}}{N}$	$\frac{n_{22}}{N}$		$\frac{n_{2J}}{N}$	$\frac{\sum l_2}{N}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
$I$	$\frac{n_{I1}}{N}$	$\frac{n_{I2}}{N}$		$\frac{n_{IJ}}{N}$	$\frac{\sum l_I}{N}$
Total	$\frac{\sum c_1}{N}$	$\frac{\sum c_2}{N}$	...	$\frac{\sum c_J}{N}$	$N$



# *A Matriz **P** de frequências relativas observadas*

$$\mathbf{P} = \frac{1}{N} \cdot \mathbf{X}_O = \begin{pmatrix} \frac{n_{11}}{N} & \frac{n_{12}}{N} & \dots & \frac{n_{1J}}{N} \\ \frac{n_{21}}{N} & \frac{n_{22}}{N} & \dots & \frac{n_{2J}}{N} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{n_{I1}}{N} & \frac{n_{I2}}{N} & \dots & \frac{n_{IJ}}{N} \end{pmatrix}$$

# *Massas das Linhas e Massas das Colunas:*

Calculadas as frequências relativas observadas, presentes na matriz **P**, podemos entender o conceito de massa (*profiles*) como medidas influência de determinada categoria em relação às demais. Nós precisaremos dessas massas para prosseguirmos para a decomposição dos *eigenvalues*.

A seguir, vamos estudar como calcular as *column profiles* e as *row profiles*.

# Column Profiles

	1	2	...	<i>J</i>	Massa Média
1	$\left(\frac{n_{11}}{\sum c_1}\right)$	$\left(\frac{n_{12}}{\sum c_2}\right)$	...	$\left(\frac{n_{1J}}{\sum c_J}\right)$	$\frac{\sum l_1}{N}$
2	$\left(\frac{n_{21}}{\sum c_1}\right)$	$\left(\frac{n_{22}}{\sum c_2}\right)$		$\left(\frac{n_{2J}}{\sum c_J}\right)$	$\frac{\sum l_2}{N}$
⋮	⋮	⋮		⋮	⋮
<i>I</i>	$\left(\frac{n_{I1}}{\sum c_1}\right)$	$\left(\frac{n_{I2}}{\sum c_2}\right)$		$\left(\frac{n_{IJ}}{\sum c_J}\right)$	$\frac{\sum l_I}{N}$
Total	1,000	1,000		1,000	

# Row Profiles

	1	2	...	<i>J</i>	Massa Média
1	$\left(\frac{n_{11}}{\sum l_1}\right)$	$\left(\frac{n_{12}}{\sum l_1}\right)$	...	$\left(\frac{n_{1J}}{\sum l_1}\right)$	1,000
2	$\left(\frac{n_{21}}{\sum l_2}\right)$	$\left(\frac{n_{22}}{\sum l_2}\right)$		$\left(\frac{n_{2J}}{\sum l_2}\right)$	1,000
⋮	⋮	⋮		⋮	⋮
<i>I</i>	$\left(\frac{n_{I1}}{\sum l_I}\right)$	$\left(\frac{n_{I2}}{\sum l_I}\right)$		$\left(\frac{n_{IJ}}{\sum l_I}\right)$	1,000
Total	$\frac{\sum c_I}{N}$	$\frac{\sum c_2}{N}$		$\frac{\sum c_J}{N}$	

## *Definições das matrizes $\mathbf{D}_l$ e $\mathbf{D}_c$ :*

$$\mathbf{D}_l = \begin{bmatrix} \frac{\sum l_1}{N} & 0 & \dots & 0 \\ 0 & \frac{\sum l_2}{N} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{\sum l_I}{N} \end{bmatrix}$$

$$\mathbf{D}_c = \begin{bmatrix} \frac{\sum c_1}{N} & 0 & \dots & 0 \\ 0 & \frac{\sum c_2}{N} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{\sum c_J}{N} \end{bmatrix}$$

## *Definição da matriz $lc'$ :*

Sendo  $\mathbf{C}$ , a matriz que contém as massas das *column profiles* de cada célula considerada; e  $\mathbf{L}$ , a matriz que contém as massas das *row profiles* de cada célula considerada, a matriz  $lc$  é definida por:

$$lc' = \mathbf{C} \otimes \mathbf{L}$$

## Definição da Matriz **W**:

Construídas as matrizes  $\mathbf{D}_l$ ,  $\mathbf{P}$ ,  $lc'$  e  $\mathbf{D}_c$ , podemos definir a matriz  $\mathbf{A}$ , como:

$$\mathbf{A} = \mathbf{D}_l^{-\frac{1}{2}} \cdot (\mathbf{P} - lc') \cdot \mathbf{D}_c^{-\frac{1}{2}}$$

E, após isso, podemos definir a matriz  $\mathbf{W}$ , dada por  $\mathbf{W} = \mathbf{A}'\mathbf{A}$ , e dela extrair os *eigenvalues*, cuja soma é igual à  $I_T$ .

Luiz Rodriguez

Lucas

\$99.40



# A Decomposição do Valor Singular de $\mathbf{A}$ :

A decomposição do valor singular de  $\mathbf{A}$  é necessária para o cálculo de seus respectivos *eigenvectors* que serão necessários para os cálculos das coordenadas a respeito de cada categoria em nosso mapa perceptual. Chamaremos os vetores dos *eigenvectors* de:

$$\mathbf{U} = \begin{pmatrix} \mathbf{u}_1 \\ \vdots \\ \mathbf{u}_I \end{pmatrix}$$

$$\mathbf{V} = \begin{pmatrix} \mathbf{v}_1 \\ \vdots \\ \mathbf{v}_I \end{pmatrix}$$

# O Cálculo das Coordenadas do Mapa Perceptual:

Variável em linha na tabela de contingências:

- Coordenadas da primeira dimensão (abcissas):

$$\mathbf{X}_l = \begin{pmatrix} \mathbf{x}_{l1} \\ \vdots \\ \mathbf{x}_{li} \end{pmatrix} = \lambda_1 \cdot \mathbf{D}_l^{-1/2} \cdot \mathbf{u}_1$$

- Coordenadas da segunda dimensão (ordenadas):

$$\mathbf{Y}_l = \begin{pmatrix} \mathbf{y}_{l1} \\ \vdots \\ \mathbf{y}_{li} \end{pmatrix} = \lambda_2 \cdot \mathbf{D}_l^{-1/2} \cdot \mathbf{u}_2$$

- Coordenadas da  $k$ -ésima dimensão:

$$\mathbf{Z}_r = \begin{pmatrix} \mathbf{z}_{l1} \\ \vdots \\ \mathbf{z}_{li} \end{pmatrix} = \lambda_k \cdot \mathbf{D}_l^{-1/2} \cdot \mathbf{u}_k$$

# O Cálculo das Coordenadas do Mapa Perceptual:

Variável em coluna na tabela de contingências:

- Coordenadas da primeira dimensão (abscissas):

$$\mathbf{X}_c = \begin{pmatrix} \mathbf{x}_{c1} \\ \vdots \\ \mathbf{x}_{cJ} \end{pmatrix} = \lambda_1 \cdot \mathbf{D}_c^{-1/2} \cdot \mathbf{v}_1$$

- Coordenadas da segunda dimensão (ordenadas):

$$\mathbf{Y}_c = \begin{pmatrix} \mathbf{y}_{c1} \\ \vdots \\ \mathbf{y}_{cJ} \end{pmatrix} = \lambda_2 \cdot \mathbf{D}_c^{-1/2} \cdot \mathbf{v}_2$$

- Coordenadas da  $k$ -ésima dimensão:

$$\mathbf{Z}_c = \begin{pmatrix} \mathbf{z}_{c1} \\ \vdots \\ \mathbf{z}_{cJ} \end{pmatrix} = \lambda_k \cdot \mathbf{D}_c^{-1/2} \cdot \mathbf{v}_k$$



005.374.619-81

## Análise de Correspondência Múltipla (ACM)

# A Construção da Matriz Binária $Z$ :

Imagine uma base de dados com  $N$  observações, com  $Q$  variáveis ( $Q > 2$ ), e que cada variável  $q$  ( $q = 1, 2, \dots, Q$ ) possua  $J_q$  categorias. Logo, o número total de categorias em uma ACM é:

$$J = \sum_{q=1}^Q J_q$$

	1	2	...	$Q$
1	Categoria 1	Categoria 4	...	Categoria 2
2	Categoria 2	Categoria 1		Categoria 1
3	Categoria 1	Categoria 3		Categoria 1
4	Categoria 3	Categoria 2		Categoria 2
⋮	⋮	⋮		⋮
$N$	Categoria 2	Categoria 4		Categoria 2
Número de categorias $J_q$	3	4	...	2

# A Matriz Binária Z:

Obs.	Variável 1			Variável 2				...	Variável Q	
	Cat. 1	Cat. 2	Cat. 3	Cat. 1	Cat. 2	Cat. 3	Cat. 4		Cat. 1	Cat. 2
1	1	0	0	0	0	0	1	...	0	1
2	0	1	0	1	0	0	0		1	0
3	1	0	0	0	0	1	0		1	0
4	0	0	1	0	1	0	0		0	1
⋮	⋮								⋮	
N	0	1	0	0	0	0	0		0	1

# A Inércia Principal Total na ACM

$$I_T = \frac{\sum_{q=1}^Q (J_q - 1)}{Q} = \frac{J - Q}{Q}$$



# A Matriz de Burt **B**:

Considerando-se a matriz binária  $\mathbf{Z} = [\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_Q]$ , podemos definir a matriz de Burt da seguinte maneira:

$$\mathbf{B} = \mathbf{Z}'\mathbf{Z}$$
$$\mathbf{B} = \begin{pmatrix} \mathbf{z}'_1 \cdot \mathbf{z}_1 & \mathbf{z}'_1 \cdot \mathbf{z}_2 & \cdots & \mathbf{z}'_1 \cdot \mathbf{z}_K \\ \mathbf{z}'_2 \cdot \mathbf{z}_1 & \mathbf{z}'_2 \cdot \mathbf{z}_2 & \cdots & \mathbf{z}'_2 \cdot \mathbf{z}_K \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{z}'_K \cdot \mathbf{z}_1 & \mathbf{z}'_K \cdot \mathbf{z}_2 & \cdots & \mathbf{z}'_K \cdot \mathbf{z}_K \end{pmatrix}_{J \times J}$$



# *O Cálculo das Coordenadas na ACM:*

As coordenadas geradas pela matriz binária **Z** são chamadas de coordenadas-padrão; Já as coordenadas geradas pela matriz **B** são chamadas de coordenadas principais, cuja relação é dada por:

$$(coordenadas\ principais_{dim.k})_B = \lambda_k \cdot (coordenadas - padrão_{dim.k})_Z$$



**USP**

**MUITO OBRIGADO!**

**Rafael de Freitas Souza**