Plano para a Apresentação da Análise Fatorial de Correspondência Simples e Múltiplas

A. Observar, conservar as observações e resumir

B. Comparar e comunicar os resultados das comparações

C. Vantagens e incovenientes do modo analógico e do modo digital de comunicação de uma mensagem

D. Representação gráfica da informação contida em uma Tabela de Contingência

E. Critérios a serem respeitados na busca de um novo referencial de representação

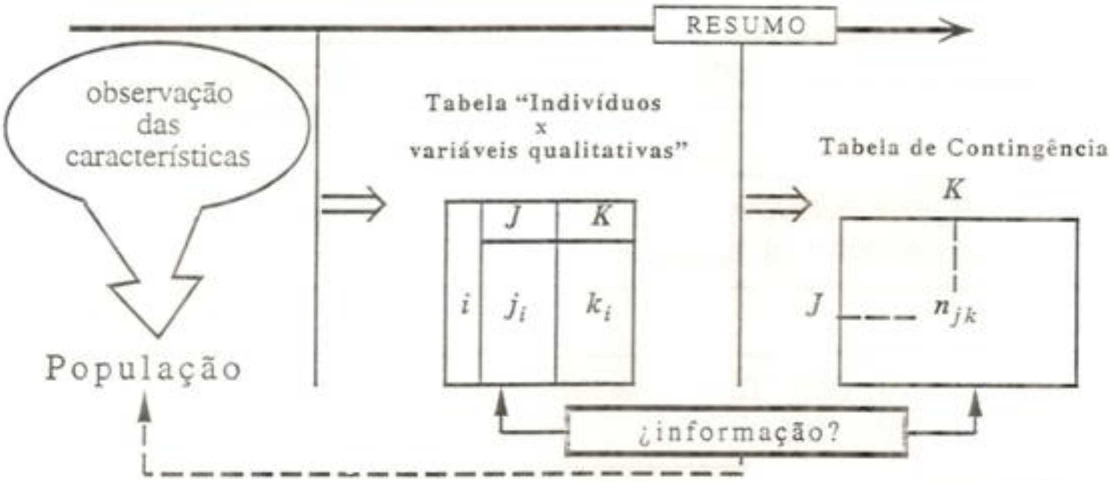
G. Representação fatorial da informação contida numa Tabela de Contingência

H. Exemplo de uma aplicação da Análise Fatorial de Correspondências Simples

I. Generalização da Análise Fatorial de Correspondências Simples: Análise Fatorial de Correspondências Múltiplas

A. Observar, conservar as observações e resumir

1. Observar para comparar



☞ A observação sistemática de um aspecto da realidade social tem por objetivo comparar os elementos observados.

☞ Ler a informação disponível em uma Tabela «indivíduos x variáveis qualitativas» implica :

■ Comparar todos os indivíduos entre si, a fim de se avaliar o grau de semelhança que existe entre eles;

■ Avaliar o nível de associação existente entre as características observadas.

A. Observar, conservar as observações e resumir

2. Resumir → perder informações

observações ⇒

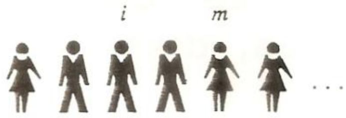


Tabela
"Ind. x var. qualitativas"

i \ K	X	Y
1	x_1	y_1
	⋮	⋮
i	x_i	y_i
	⋮	⋮
m	x_m	y_m
	⋮	⋮
n	x_J	y_K

conservação ⇒

¿como comparar o indivíduo **i** e o indivíduo **m** ?

x_i = código y_m = código

$x_i \overset{=}{\neq} x_m$

$x_i \overset{\leq}{\geq} x_m$

$\cancel{x_i + x_m} = ?$

$\boxed{n_{jk} = n_{jk} + 1}$

Tabela de Contingência

	Y_1	...	Y_k	...	Y_K
X_1	n_{11}	...	n_{1k}	...	n_{1K}
	...	...	...	...	...
X_j	n_{j1}	...	n_{jk}	...	n_{jK}
	...	...	...	...	...
X_J	n_{J1}	...	n_{Jk}	...	n_{JK}

resumo ⇒

☞ A tabela de contingência contém menos informações que a tabela «indivíduos x variáveis qualitativas».

☞ Conhecendo-se a Tabela de Contingência não se pode deduzir a Tabela «indivíduos x variáveis qualitativas».

☞ Na Tabela de contingência se perde informações sobre os indivíduos.

3. Avaliação das comparações : similaridades, dissimilaridades e distáncias

Se define matematicamente a semelhança ou a similaridade de dois indivíduos, i e j , mediante uma função s_{ij} dosvalores observados que figuram nas linhas i e j da Tabela «indivíduos x variáveis qualitativas».

Notas :

- Define-se habitualmente similaridade por meio de uma relação simétrica ($s_{ij} = s_{ji}$).

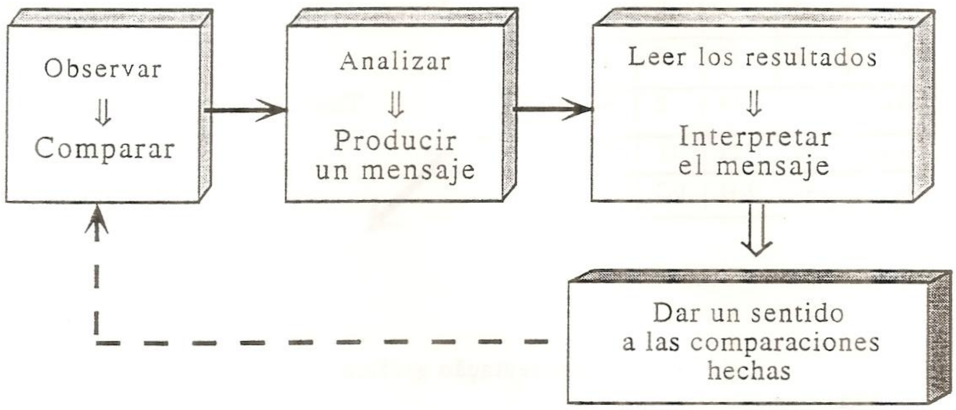
- A maioria das medidas de similaridade sao construídas de tal maneira que seus valores estejam compreendidas entre 0 e 1 : $0 < s_{ij} < 1$

- As medidas de similaridade da forma $0 < s_{ij} < 1$ permitem definir as medidas de dissimilaridade $d_{ij} = 1 - s_{ij}$, que também sao simétricas e também compreendidas entre 0 e 1.

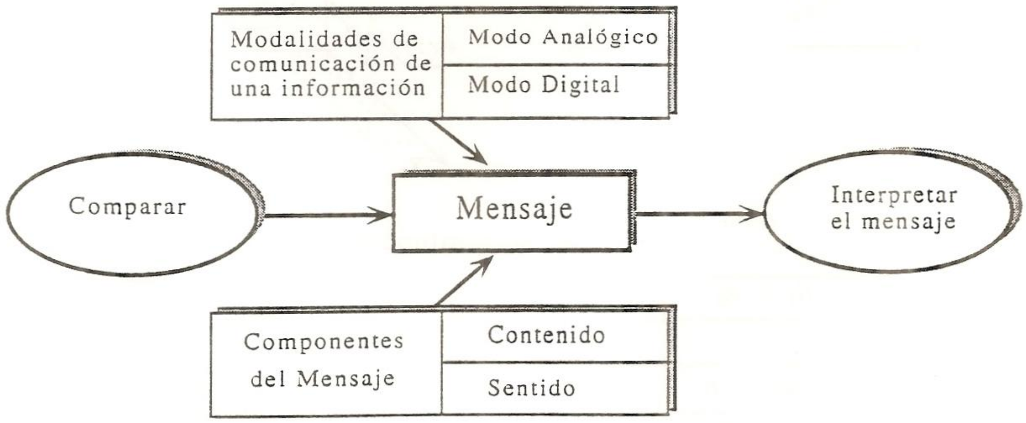
B. Comparar e comunicar os resultados das comparações

1. O tratamento estatístico produz uma mensagem

Estabelecer a similaridade ou dissimilaridade entre dois indivíduos (ou a distância em um triplo de indivíduos) da Tabela de dados implica na produção de uma mensagem que explicita as informações elementares contidas nesta Tabela de Dados.



2. Componentes e modalidades da mensagem



C. Vantagens e inconvenientes do modo analógico e do modo digital de comunicação de uma mensagem

1. Exemplo : comunicação da informação contida uma Tabela de Contingência

Tabela de Contingência

	Y_1	Y_2	Y_3	margen	p_j
X_1	5	2	2	9	0,13
X_2	8	12	7	27	0,39
X_3	17	15	2	34	0,49
margen	30	29	11	70	
p_k	0,43	0,41	0,2		

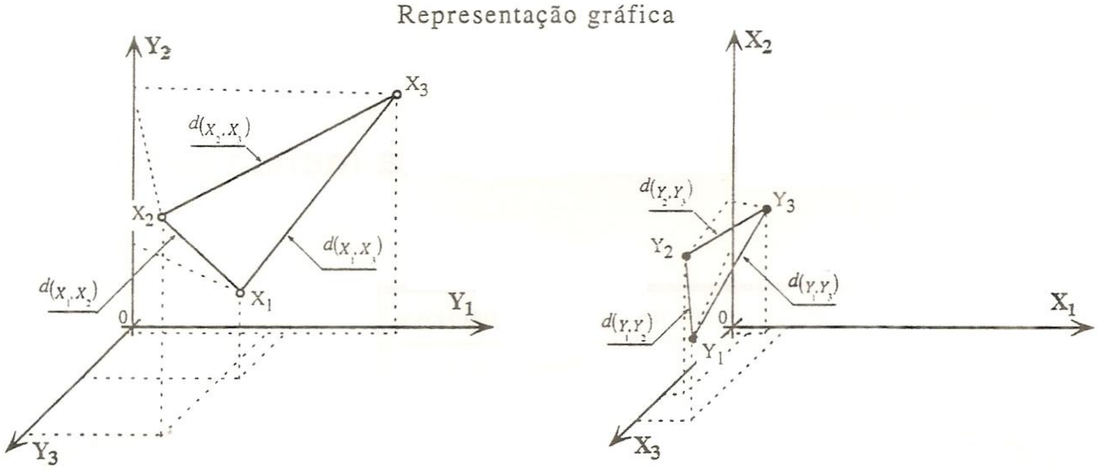
Tabela de perfis ponderados em linha

X_1	0.56	0.22	0,22
X_2	0.30	0,44	0,26
X_3	0.50	0,44	0,06

Tabela de perfis ponderados em columna

Y_1	Y_2	Y_3
0.17	0.07	0,16
0.27	0,41	0,64
0.57	0,52	0,18

Representação gráfica



Evaluação das similitudes

	X_1	X_2	X_3
X_1	0		
X_2	0,285	0	
X_3	0,293	0.32	0

	Y_1	Y_2	Y_3
Y_1	0		
Y_2	0,135	0	
Y_3	0,661	0,459	0

C. Vantagens e inconvenientes do modo analógico e do modo digital de comunicação de uma mensagem

2. Notas sobre o índice de similaridade e a utilidade do conceito de distância

☞ Os coeficientes de dissimilaridade $\mathbf{d}_{ij}$ que não satisfazem a relação de desigualdade triangular:

❶ Nao permitem criar um espaço euclidiano capaz de representar os resultados da comparação de todos os indivíduos.

❷ Apenas permitem «figurar» ou representar as semelhanças entre os indivíduos através de uma nuvem de pontos sobre uma imagem plana qualquer.

Para representar adequadamente elementos de uma tabela de dados torna-se necessário :

a) *definir um espaço de representação gráfica capaz de transmitir o sentido da relação observada entre os elementos comparados;*

b) *utilizar como critério de comparação uma expressão da distância entre dois pontos que seja apropriada ao espaço de representação adotado.*

☞ Mas como representar a comparação entre os indivíduos de uma Tabela quando o número de atributos é muito grande?

☞ Os métodos estatísticos de análise de grandes tabelas (quadros numéricos) produzem mensagens analógicas e digitais que permitem compreender o sentido da informação contida.

D. Representação gráfica da informação contida numa Tabela de Contingência

1. A Tabela de Contingência e a Tabela de Frequências relativas associada

TABELA «T» TABELA DE CONTINGÊNCIA

VARIABLE "L" \ VARIABLE "C"	1	...	j	...	J	
1	n_{11}	...	n_{1j}	...	n_{1J}	$n_{1.}$
$\vdots$	...	...	$\vdots$	...	$\vdots$	$\vdots$
i	n_{i1}	...	n_{ij}	...	n_{iJ}	$n_{i.}$
$\vdots$	...	...	$\vdots$	...	$\vdots$	$\vdots$
I	n_{I1}	...	n_{Ij}	...	n_{IJ}	$n_{I.}$
	$n_{.1}$	...	$n_{.j}$	...	$n_{.J}$	$n_{..}$

- n_{ij} : termo geral

- $n_{i.}$: é o i-ésimo elemento da margem direita de T

$$n_{i.} = \sum_{j=1}^{J} n_{ij} \quad \forall i \in I = (1, 2, \ldots, I)$$

- $n_{.j}$: é o j-ésimo elemento da margem inferior de T

$$n_{.j} = \sum_{i=1}^{I} n_{ij} \quad \forall j \in J = (1, 2, \ldots, J)$$

- $n_{..}$: é a soma de todas as células de T

$$n_{..} = \sum_{i=1}^{I} \sum_{j=1}^{J} n_{ij} = \sum_{i=1}^{I} n_{i.} = \sum_{j=1}^{J} n_{.j}$$

D. Representação gráfica da informação contida numa Tabela de Contingência

1. A Tabela de Contingência e a Tabela de Frequências Relativas associada (continuação)

Tabela T (3 x 3)

Variable "L"	Variable "C" C1	C2	C3	
L1	n_{11}	n_{12}	n_{13}	$n_{1.}$
L2	n_{21}	n_{22}	n_{23}	$n_{2.}$
L3	n_{31}	n_{32}	n_{33}	$n_{3.}$
	$n_{.1}$	$n_{.2}$	$n_{.3}$	$n_{..}$

Tabela F (3 x 3)

Variable "L"	Variable "C" C1	C2	C3	
L1	f_{11}	f_{12}	f_{13}	$f_{1.}$
L2	f_{21}	f_{22}	f_{23}	$f_{2.}$
L3	f_{31}	f_{32}	f_{33}	$f_{3.}$
	$f_{.1}$	$f_{.2}$	$f_{.3}$	$f_{..}$

- termo geral : $f_{ij} = \dfrac{n_{ij}}{n_{..}}$

- elementos marginais:

$$f_{i.} = \frac{n_{i.}}{n_{..}} = \sum_{j=1}^{J} \frac{n_{ij}}{n_{..}} = \sum_{i=1}^{J} f_{ij} \qquad \forall i \in I = (1, 2, \ldots, I)$$

$$f_{.j} = \frac{n_{.j}}{n_{..}} = \sum_{i=1}^{I} \frac{n_{ij}}{n_{..}} = \sum_{i=1}^{I} f_{ij} \qquad \forall j \in J = (1, 2, \ldots, J)$$

- A soma na margem é evidentemente igual a 1 , uma vez que a tabela T foi dividida pelo valor de n..

$$f_{..} = \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{n_{ij}}{n_{..}} = \frac{1}{n_{..}} \sum_{i=1}^{I} \sum_{j=1}^{J} n_{ij} = \frac{n_{..}}{n_{..}} = 1$$

D. Representação gráfica da informação contida numa Tabela de Contingência

Apresentação de um exemplo numérico :

Mercado anual de «centrais telefônicas»

Um pesquisador se propõe a distinguir :

☞ os países que satisfizeram às necessidades de seus mercados internos através da importação de centrais telefônicas, daqueles países que optaram por desenvolver a produção local;

☞ os países que desenvolveram a produção de centrais para atender às necessidades de seus mercados internos, daqueles países que orientaram sua produção para a exportação.

Tabela «M.A.C.T. (3 x 3)»

	Venta de nuevas Centrales Telefónicas en cada país			
Producción de Centrales telefónicas en cada país	a	b	c	Producción total
A	75	25	54	154
B	56	78	189	323
C	89	132	202	423
Total Ventas:	220	235	445	900

☞ As «questões» que o pesquisador se propõe a resolver não comportam uma formulação em termos de causalidade.

Para estabelecer uma relação de causalidade entre a produção e o consumo dos bens referidos, torna-se necessário programar uma observação do mercado :

❶ que se repita no tempo (observação longitudinal);

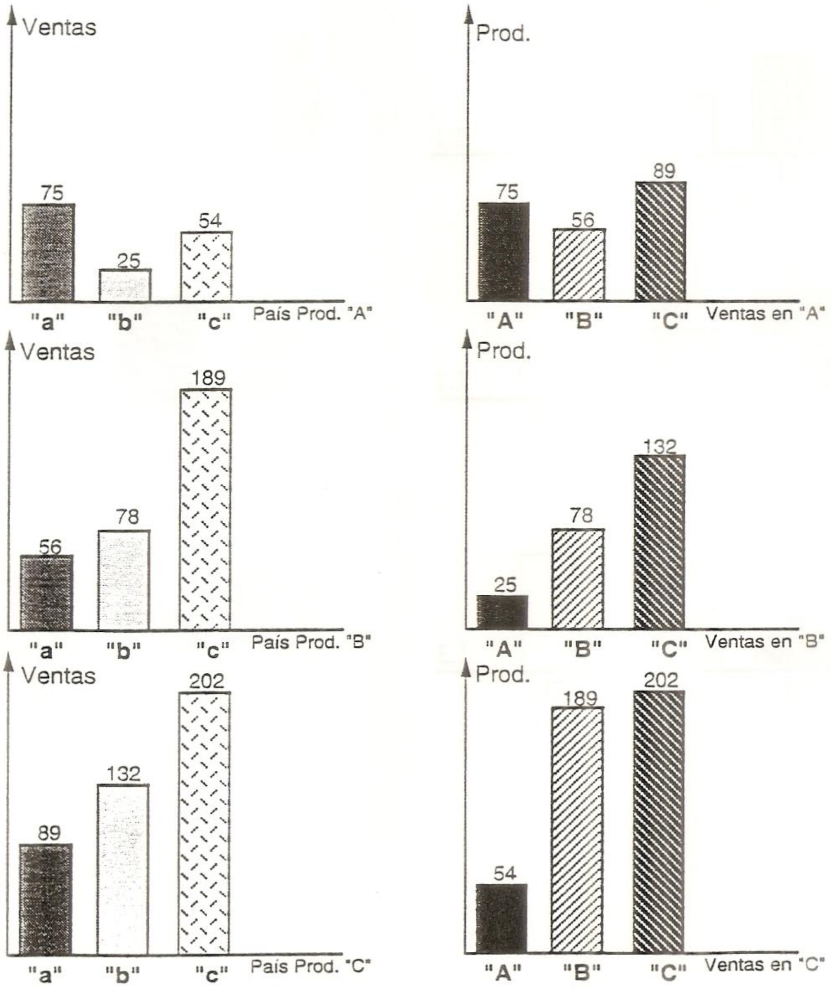
❷ que tome em conta as estratégias de produção das firmas e a evolução das exigências técnicas dos utilizadores finais das centrais telefônicas.

D. Representação gráfica da informação contida numa Tabela de Contingência

2. Exemplo numérico : (continuação)

1. Primeira representação gráfica :

Deve-se construir o «diagrama de barras» das distribuições observadas de cada linha e deve se proceder da mesma forma com as colunas.

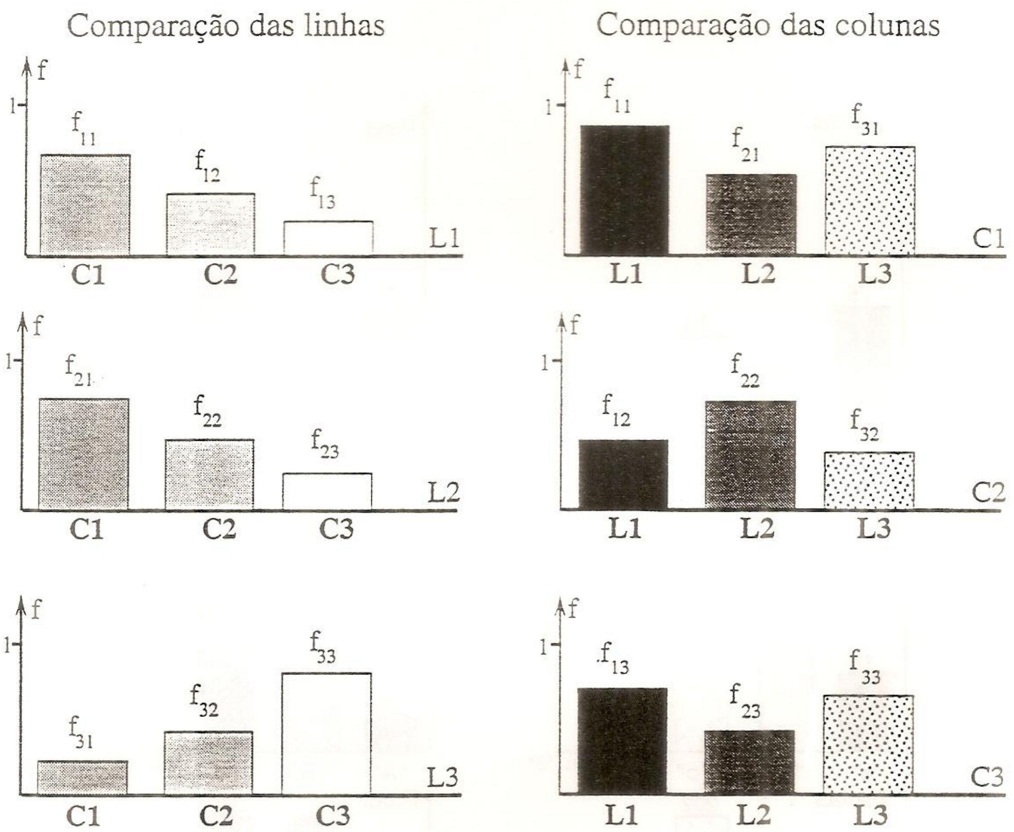


D. Representação gráfica da informação contida numa Tabela de Contingência

2. Exemplo numérico : (continuação)

1. Segunda representação gráfica :

Se representarmos a Tabela F de Frequências Relativas, obteremos um gráfico equivalente

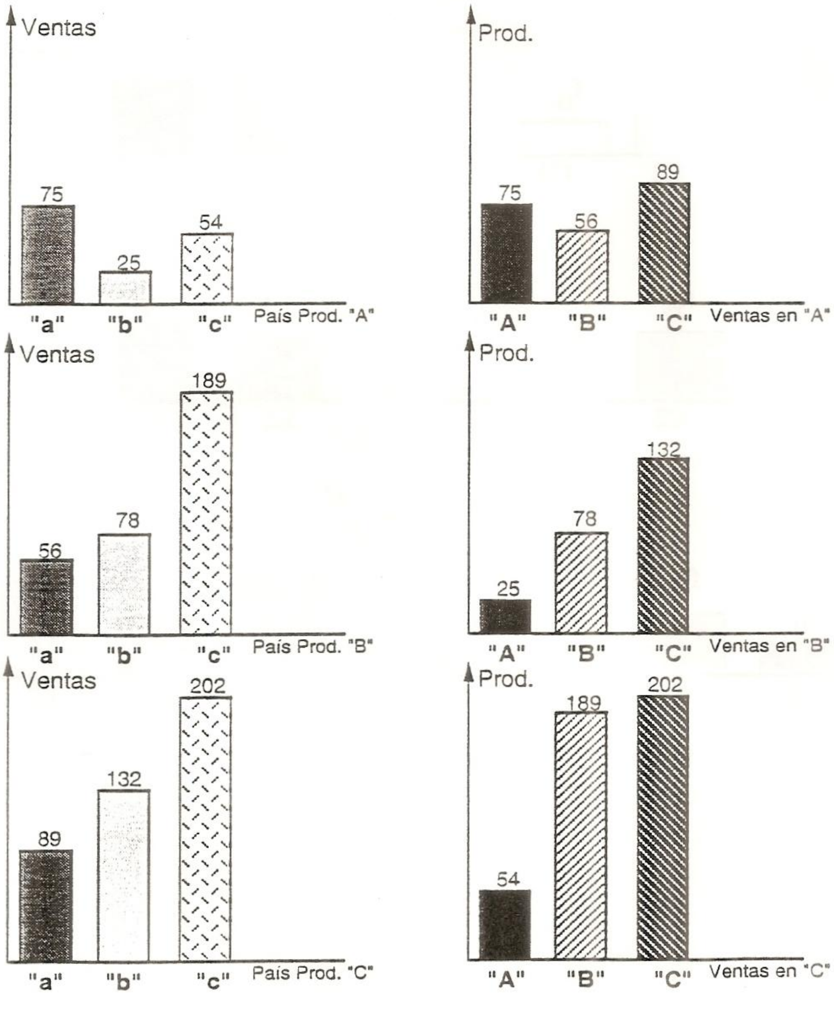


D. Representação gráfica da informação contida numa Tabela de Contingência

2. Exemplo numérico : (continuação)

1. Primeira representação gráfica :

Deve-se construir o «diagrama de barras» das distribuições observadas de cada linha e deve se proceder da mesma forma com as colunas.



D. Representação gráfica da informação contida numa Tabela de Contingência

2. Exemplo numérico : (continuação)

- **Indice de similaridade entre elementos :**

Completamos a representação gráfica por meio de um «índice» que resuma o grau de similaridade existente entre elementos da tabela F.

- Definição do «índice de similaridade» entre linhas de uma tabela F(I x J) correspondente à primeira representação gráfica:

$$s_{(i,i')} = \sum_{j=1}^{J} \left(f_{ij} - f_{i'j} \right) = \sum_{j=1}^{J} \left(\frac{n_{ij}}{n_{..}} - \frac{n_{i'j}}{n_{..}} \right) \quad \forall\, i, i'$$

- Definição do «índice de similaridade» entre colunas de uma tabela F(I x J) correspondente à primeira representação gráfica:

$$s_{(j,j')} = \sum_{i=1}^{I} \left(f_{ij} - f_{ij'} \right) = \sum_{i=1}^{I} \left(\frac{n_{ij}}{n_{..}} - \frac{n_{ij'}}{n_{..}} \right) \quad \forall\, j, j'$$

D. Representação gráfica da informação contida numa Tabela de Contingência

2. Exemplo numérico : (continuação)

3. Discordância entre os modos de representação da informação das tabelas T ou F

Para representar todas as relações entre elementos da tabela F, devemos produzir tantas «mensagens» (analógicas - digitais) quantos pares de elementos - linha e de elementos - coluna contenha a tabela.

Este modo de comunicar a informação não transmite a forma da informação contida em F, mas tão somente seus conteúdos. (Cada conteúdo separadamente).

Assim a interpretação da informação transmitida, (a verificação do seu sentido), fica subordinada à «leitura» subjetiva dos conteúdos da mesma.

Ocorre portanto perda do sentido da mensagem produzida a partir daquilo que foi observado.

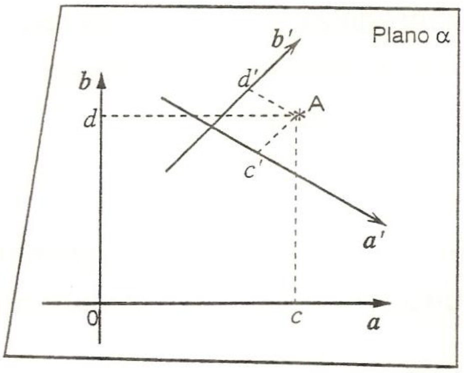
Assim sendo, torna-se necessário definir outras maneiras analógicas e digitais de comunicar a informação de uma tabela T ou de uma tabela F.

D. Representação gráfica da informação contida numa Tabela de Contingência

3. Representação de uma Tabela de Contingência no espaço euclidiano

1. Espaços de representação :

Sabemos colocar um ponto no espaço α comum que nos é habitual...



O plano α é um espaço de representação de duas dimensões (comprimento x largura).

Todo referencial composto por um par de eixos graduados permite que se coloque neste plano qualquer ponto do espaço.

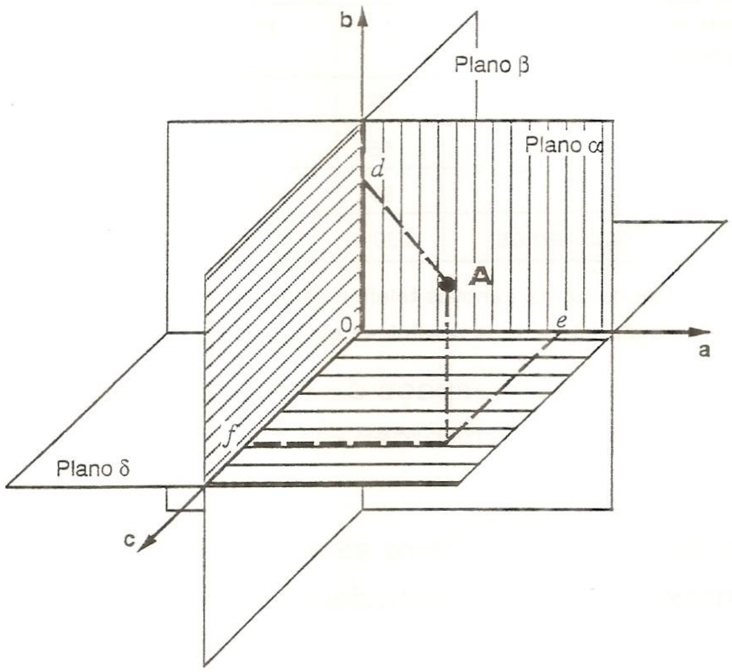
D. Representação gráfica da informação contida numa Tabela de Contingência

3. Representação de uma Tabela de Contingência no espaço euclidiano

1. Espaços de representação : (continuação)

Em um espaço de três dimensoes...

Mesmas considerações.



Estes tipos de espaços de representação sao chamados espaços euclidianos.

Correspondem a nossa percepção habitual do espaço.

As propriedades formais destes espaços podem ser generalizadas para espaços de maiores dimensões.

Porém, nao podemos desenhá-los...

D. Representação gráfica da informação contida numa Tabela de Contingência

3. Representação de uma Tabela de Contingência no espaço euclidiano

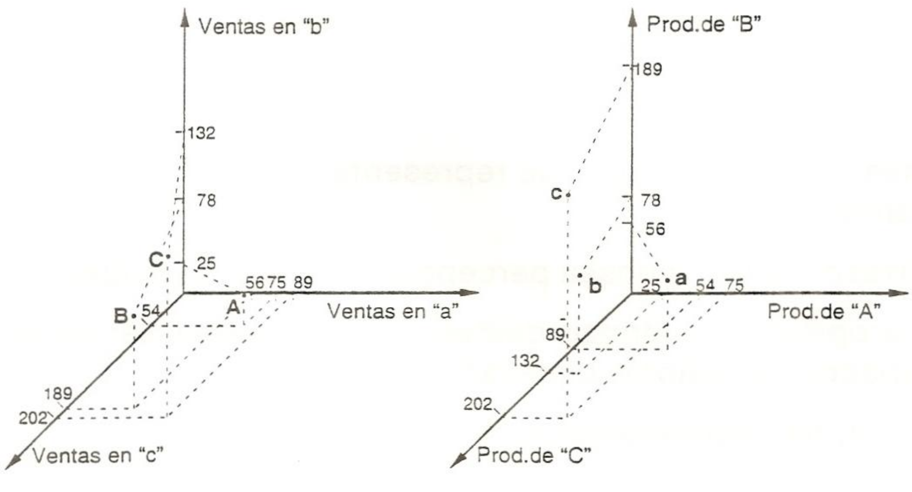
2. Modo analógico de representação da Tabela de Contigência em um espaço euclidiano :

Tabela «M.A.C.T.(3 x 3)»

	Venta de nuevas Centrales Telefónicas en cada país			
Producción de Centrales telefónicas en cada país	a	b	c	Producción total
A	75	25	54	154
B	56	78	189	323
C	89	132	202	423
Total Ventas:	220	235	445	900

Trad.: Vendas de novas centrais telefinicas em cada país.
Produção de centrais telefinicas em cada país.

- Traçamos em um espaço de representação os três pontos-linha a partir das « coordenadas» dos mesmos no referencial das três colunas.
- Procedemos de maneira similar para a representação dos pontos-coluna no referencial das linhas.



Escala: 1 cm = 50 unidades de obs.

D. Representação gráfica da informação contida numa Tabela de Contingência

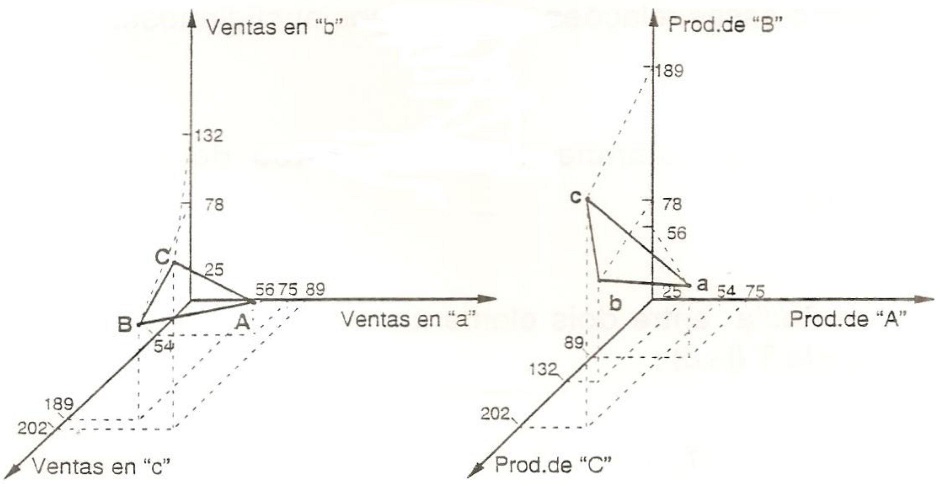
3. Representação de uma Tabela de Contingência no espaço euclidiano

2. Modo analógico de representação da Tabela de Contigência em um espaço euclidiano :

(continuação)

Expressamos assim a comparação entre os elementos da Tabela em termos de distância entre pontos nos espaços euclidianos.

As distáncias relativas dos pontos entre si representam a forma da informação contida na Tabela de Contingência.



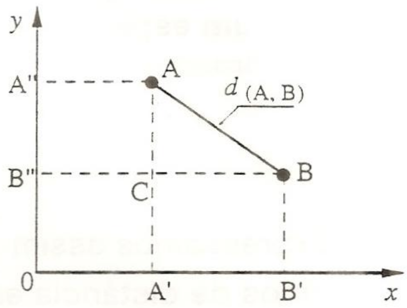
Escala: 1 cm = 50 unidades de obs.

D. Representação gráfica da informação contida numa Tabela de Contingência

3. Representação de uma Tabela de Contingência no espaço euclidiano

3. Modo digital de resumir a informação da tabela T :

Calculemos a distáncia entre dois pontos (Teorema de Pitágoras).



A distância entre A e B

$$d_{(A,B)} = \sqrt{(B' - A')^2 + (B'' - A'')^2}$$

Generalizando para o caso de espaços de três dimensões

$$d_{(A,B)} = \sqrt{(B' - A')^2 + (B'' - A'')^2 + (B''' - A''')^2}$$

Aplicando essas relações nos espaços euclidianos...

Definição da distância entre elementos de uma Tabela de Contingência...

- Distância entre dois elementos linha de uma tabela T (I x J) :

$$d_{(i,i')} = \sqrt{\sum_{j=1}^{J} \left(n_{ij} - n_{i'j}\right)^2} \quad \forall\, i, i'$$

- Distância entre dois elementos coluna de uma tabela T (I x J) :

$$d_{(j,j')} = \sqrt{\sum_{i=1}^{I} \left(n_{ij} - n_{ij'}\right)^2} \quad \forall\, j, j'$$

D. Representação gráfica da informação contida numa Tabela de Contingência

3. Representação de uma Tabela de Contingência no espaço euclidiano

4. Correspondência entre os dois modos de representação da informação da tabela T

Este resumo... expressa realmente a mesma informação que o gráfico , porém de uma forma diferente...?

Propriedades :

❶ Similaridade ou diferença entre dois elementos I e J :

A distáncia entre dois pontos de um espaço euclidiano é nula ou positiva:

$$d_{(i,i')} \geq 0 \quad \forall i,i' \text{ e } d_{(j,j')} \geq 0 \quad \forall j,j'$$

❷ Similaridade entre dois elementos **i** e **j** :

$$d_{(i,i')} = 0 \iff i = i' \text{ e } d_{(j,j')} = 0 \iff j = j'$$

❸ Similaridade ou diferença entre elementos adotada qualquer ordem de comparação...

A distáncia do ponto i ao ponto j é a mesma entre a distáncia do ponto j ao ao ponto i :

$$d_{(i,i')} = d_{(i',i)} \quad \forall i,i' \text{ e } d_{(j,j')} = d_{(j',j)} \quad \forall j,j'$$

❹ Desigualdade triangular:

$$d_{(i,i')} \leq d_{(j,k)} + d_{(i',k)} \quad \forall i,i',k \text{ e } d_{(j,j')} \leq d_{(j,k)} + d_{(j',m)} \quad \forall j,j',m$$

Este «índice de comparação» possui as quatro propriedades que determinam uma distância entre dois pontos de um espaço euclidiano.

Este índice «avalia» corretamente estas distâncias e traduz o modo analógico de comparação utilizado para «ler» a informação da tabela.

D. Representação gráfica da informação contida numa Tabela de Contingência

3. Representação de uma Tabela de Contingência no espaço euclidiano

5. Distorção da informação contida na tabela T :

■ Comparando os elementos linha (coluna) da tabela T, conseguimos estabelecer relações entre elementos comparáveis...?

Os valores das células da tabela T serão tão maiores quanto for a população observada.

As diferenças observadas em duas linhas (coluna) conduzem a uma apreciação da diferença entre estes elementos que é ampliada pelo efeito do tamanho da população observada.

■ Em nosso exemplo numérico :

➥ (se o volume do mercado é de 900 centrais (n.. = 900)

$$\begin{aligned} d^2_{(\text{“}A\text{”},\text{“}B\text{”})} &= (75-56)^2 + (25-78)^2 + (54-189)^2 \\ &= 21\,395 \quad \Rightarrow \quad d_{(\text{“}A\text{”},\text{“}B\text{”})} = 146{,}27 \end{aligned}$$

➥ (se o volume do mercado é de 90 centrais (n.. = 90)

$$\begin{aligned} d^2_{(\text{“}A\text{”},\text{“}B\text{”})} &= (8-6)^2 + (3-8)^2 + (5-19)^2 \\ &= 225 \quad \Rightarrow \quad d_{(\text{“}A\text{”},\text{“}B\text{”})} = 15 \end{aligned}$$

ê necessário neutralizar o efeito amplificador das comparações produzido pelo tamanho da população observada.

D. Representação gráfica da informação contida numa Tabela de Contingência

3. Representação de uma Tabela de Contingência no espaço euclidiano

6. Representação da tabela F(3 x 3) em um espaço euclidiano :

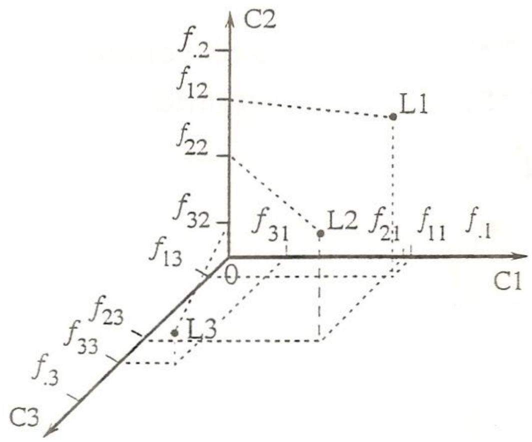
- Para neutralizar o efeito do tamanho → proporção de cada co-ocorrência observada na população total...

- Se associa o tabela T a tabela F :

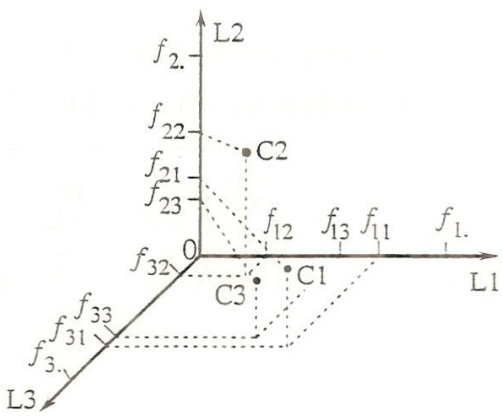
$$F\ (I \times J) = (1/n_{..})\ .\ T\ (I \times J)$$

- Se representa a informação contida na tabela F :

Representação dos pontos-linha no referencial das colunas



Representação dos pontos-coluna no referencial das linhas



D. Representação gráfica da informação contida numa Tabela de Contingência

3. Representação de uma Tabela de Contingência no espaço euclidiano

7. Definição do modo digital de resumir a informação da tabela F :

■ A expressão geral de uma distáncia no caso da comparação de dois elementos-linha de uma tabela F(I x J) é a seguinte :

$$d_{(i,i')} = \sqrt{\sum_{j=1}^{J} \left(\frac{n_{ij}}{n_{..}} - \frac{n_{i'j}}{n_{..}} \right)^2} = \sqrt{\sum_{j=1}^{J} \left(f_{ij} - f_{i'j} \right)^2} \quad \forall\, i, i'$$

■ A expressão geral de uma distáncia no caso da comparação de dois elementos-coluna de uma tabela F(I x J) é a seguinte :

$$d_{(j,j')} = \sqrt{\sum_{i=1}^{I} \left(\frac{n_{ij}}{n_{..}} - \frac{n_{ij'}}{n_{..}} \right)^2} = \sqrt{\sum_{i=1}^{I} \left(f_{ij} - f_{ij'} \right)^2} \quad \forall\, j, j'$$

D. Representação gráfica da informação contida numa Tabela de Contingência

3. Representação de uma Tabela de Contingência em um espaço euclidiano

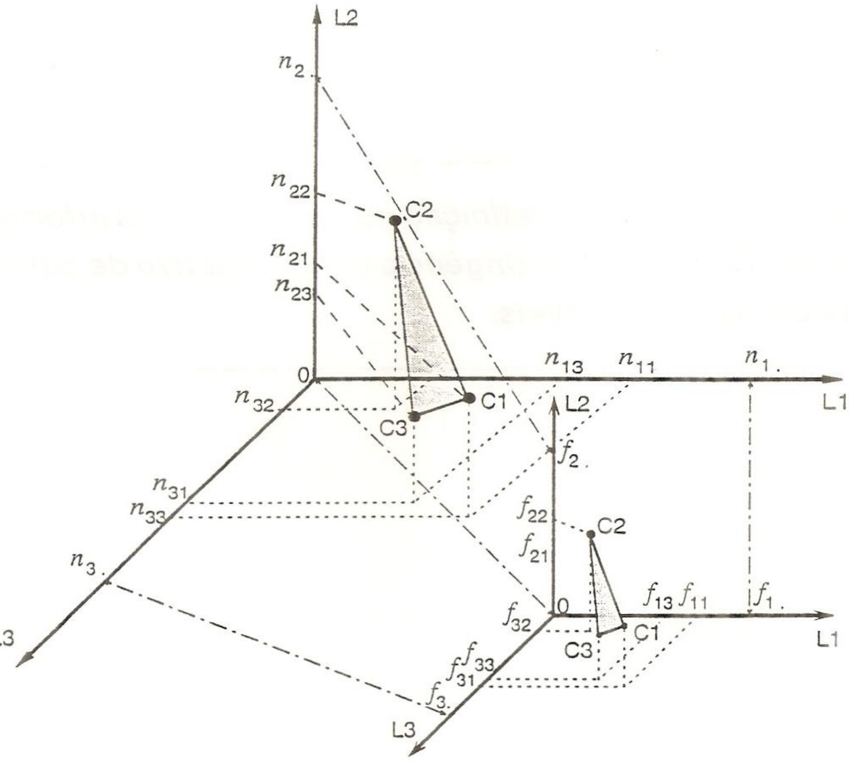
8. A Tabela F «conserva» a informação da tabela T

- Se neutraliza assim o «efeito tamanho» da tabela T. Porque ?...

- As representações da informação da tabela F trazem a mesma «mensagem» sobre o que foi observado do que a representação proposta pela tabela T.

- Porque ?...

➥ O espaço de representação da tabela T é homotético do espaço de representação da tabela F. (Eixos proporcionais).



O espaço de representação da tabela F é a fotocópia reduzida (com uma escala 1/n..) do espaço de representação da tabela T.

3. Representação de uma Tabela de Contingência no espaço euclidiano

9. Distorção da informação da Tabela F :

☞ O tamanho da tabela observada não é o único fator de possíveis distorções da leitura da informação.

☞ As linhas e as colunas de uma tabela F . . . são comparáveis ?...

Devemos melhorar a definição dos conteúdos da informação de uma Tabela de Contingência com o objetivo de comparar elementos comparáveis.

D. Representação gráfica da informação contida numa Tabela de Contingência

4. Representação da informação em termos de comparação de perfis

1. Definição dos perfis dos elementos da tabela F :

Um perfil em linha expressa as proporções de indivíduos que apresentam cada característica da variável na coluna, dentro do subgrupo da população que apresenta a característica correspondente numa linha.

- Perfis em linha de uma tabela T :

$$f_{i|Cj} = \frac{n_{ij}}{n_{i.}} \quad \forall i \in I; \ \forall j \in J$$

e

$$\sum_{j=1}^{J} f_{i|Cj} = \sum_{j=1}^{J} \frac{n_{ij}}{n_{i.}} = \frac{1}{n_{i.}} \sum_{j=1}^{J} n_{ij} = \frac{n_{i.}}{n_{i.}} = 1$$

- Perfis em coluna de uma tabela T :

$$f_{j|Li} = \frac{n_{ij}}{n_{.j}} \quad \forall i \in I; \ \forall j \in J$$

e

$$\sum_{i=1}^{I} f_{j|Li} = \sum_{i=1}^{I} \frac{n_{ij}}{n_{.j}} = \frac{1}{n_{.j}} \sum_{j=1}^{J} n_{ij} = \frac{n_{.j}}{n_{.j}} = 1$$

D. Representação gráfica da informação contida numa Tabela de Contingência

4. Representação da informação em termos de comparação de perfis

1. Definição dos perfis dos elementos da tabela F (continuação) :

- Perfis em linha de uma tabela F :

$$f_{i|Cj} = \frac{f_{ij}}{f_{i.}} \quad \forall i \in I; \ \forall j \in J$$

e $$\sum_{j=1}^{J} f_{i|Cj} = \sum_{j=1}^{J} \frac{f_{ij}}{f_{i.}} = \frac{1}{f_{i.}} \sum_{j=1}^{J} f_{ij} = \frac{f_{i.}}{f_{i.}} = 1$$

- Perfis em coluna de uma tabela F :

$$f_{j|Li} = \frac{f_{ij}}{f_{.j}} \quad \forall i \in I; \ \forall j \in J$$

e $$\sum_{i=1}^{I} f_{j|Li} = \sum_{i=1}^{I} \frac{f_{ij}}{f_{.j}} = \frac{1}{f_{.j}} \sum_{j=1}^{J} f_{ij} = \frac{f_{.j}}{f_{.j}} = 1$$

➤ Um perfil linha (coluna) expressa a distribuição de frequências condicionais da subpopulação em linha (coluna) para cada um dos caracteres em coluna (linha).

➤ $f_{i|Cj}$: é a proporção da i-ésima subpopulação que apresenta a modalidade j da variável «C».

D. Representação gráfica da informação contida numa Tabela de Contingência

4. Representação da informação em termos de comparação de perfis

1. Definição dos perfis de elementos da Tabela F (continuação)

- Se associa o tabela T, ou o tabela F, duas tabelas de perfis.

Perfis em linha da tabela T (3 x 3)

		Variable "C"			
		C1	C2	C3	
Variable "L"	L1	$f_{1\mid C1}$	$f_{1\mid C2}$	$f_{1\mid C3}$	1
	L2	$f_{2\mid C1}$	$f_{2\mid C2}$	$f_{2\mid C3}$	1
	L3	$f_{3\mid C1}$	$f_{3\mid C2}$	$f_{3\mid C3}$	1
		p_{C1}	p_{C2}	p_{C3}	1

Perfis em coluna da tabela T (3 x 3)

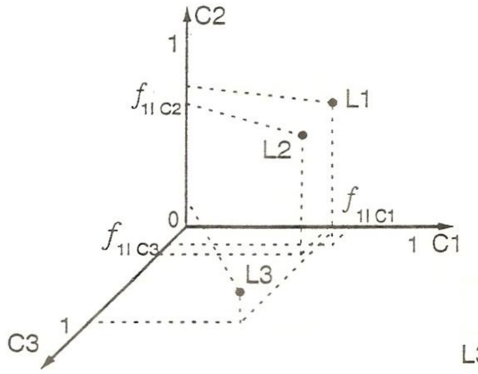
		Variable "C"			
		C1	C2	C3	
Variable "L"	L1	$f_{1\mid L1}$	$f_{2\mid L1}$	$f_{3\mid L1}$	p_{L1}
	L2	$f_{1\mid L2}$	$f_{2\mid L2}$	$f_{3\mid L2}$	p_{L2}
	L3	$f_{1\mid L3}$	$f_{2\mid L3}$	$f_{3\mid L3}$	p_{L3}
		1	1	1	1

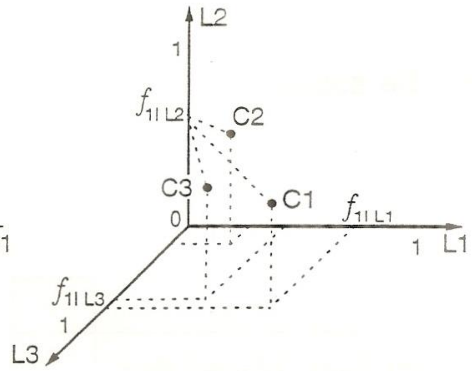
- Para a comunicação da informação contida nessas tabelas, se utilizam as representações analógicas e digitais que já foram vistas.

D. Representação gráfica da informação contida numa Tabela de Contingência

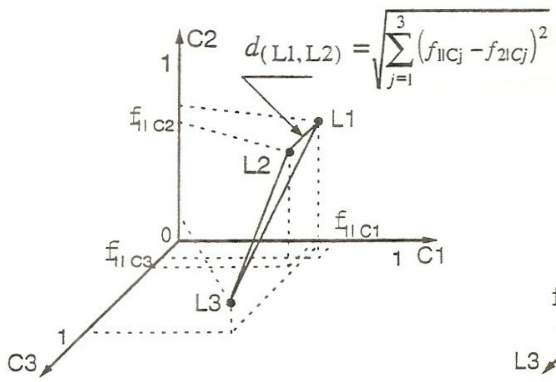
4. Representação da informação em termos de comparação de perfis

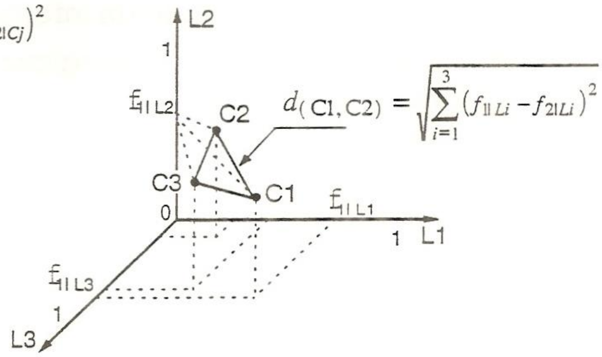
2. Representação gráfica (analógica) da informação das tabelas de perfis :





➤ Podemos materializar sobre os gráficos as «distáncias» relativas entre os pontos-perfis :





D. Representação gráfica da informação contida numa Tabela de Contingência

4. Representação da informação em termos de comparação de perfis

3. Representação digital da informação das tabelas de perfis :

■ Avaliação da distância entre dois perfis-linha, associadas a uma tabela F(I x J) :

$$d_{(i,i')} = \sqrt{\sum_{j=1}^{J} \left(\frac{f_{ij}}{f_{i.}} - \frac{f_{i'j}}{f_{i'.}} \right)^2} \quad \forall i, i'$$

■ Avaliação da distância entre dois perfis-coluna associados a uma tabela F(I x J) :

$$d_{(j,j')} = \sqrt{\sum_{i=1}^{I} \left(\frac{f_{ij}}{f_{.j}} - \frac{f_{ij'}}{f_{.j'}} \right)^2} \quad \forall j, j' \qquad (1)$$

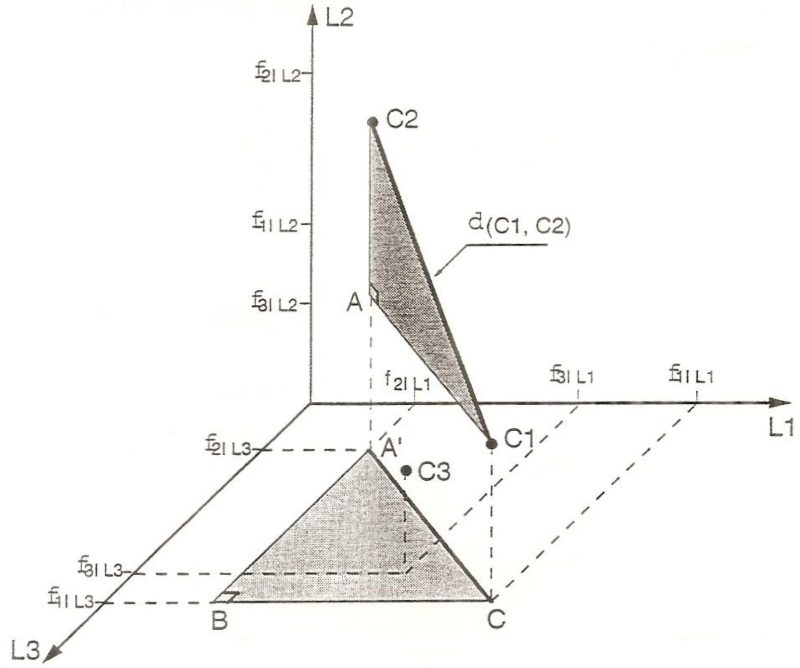
O índice $d_{(i,i')}$ ou $d_{(j,j')}$ avalia a «distância» que caracteriza a semelhança entre dois perfis no espaço euclidiano de representação dos mesmos

D. Representação gráfica da informação contida numa Tabela de Contingência

4. Representação da informação em termos de comparação de perfis

4. Representação gráfica da informação contida em uma tabela de Contingência

Neste espaço, a distância entre dois pontos-coluna $d_{(j\,,\,j')}$ representa o grau de semelhança entre dois elementos.



$$d^2_{(C1,\,C2)} = \sum_{i=1}^{3}\left(f_{1\,|\,Li} - f_{2\,|\,Li}\right)^2 = \sum_{i=1}^{3}\left(\frac{f_{i1}}{f_{.1}} - \frac{f_{i2}}{f_{.2}}\right)^2$$

De modo que :

$$d_{(C1,\,C2)} = \sqrt{\sum_{i=1}^{3}\left(f_{1\,|\,Li} - f_{2\,|\,Li}\right)^2} = \sqrt{\sum_{i=1}^{3}\left(\frac{f_{i1}}{f_{.1}} - \frac{f_{i2}}{f_{.2}}\right)^2}$$

A distância entre as colunas C_1 e C_2 é medida pela mesma expressão que [1].

D. Representação gráfica da informação contida numa Tabela de Contingência

4. Representação da informação em termos de comparação de perfis

5. Relação entre a informação das tabelas de perfis e a informação da tabela F :

Tabla F
Espacio de Puntos-Columna

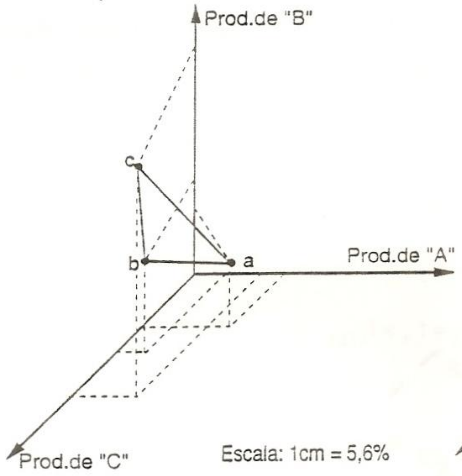


Tabla de Perfiles Asociada a la Tabla F
Espacio de Puntos-Columna

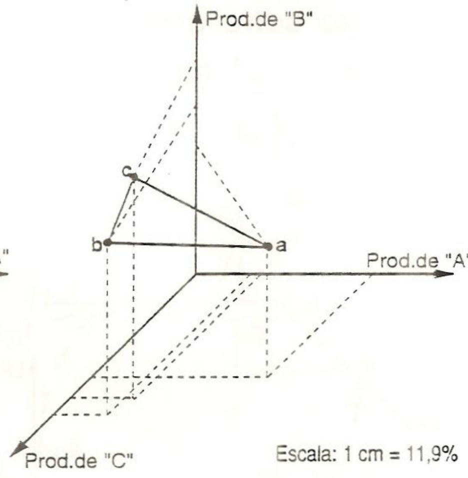


Tabla F
Espacio de Puntos-Línea

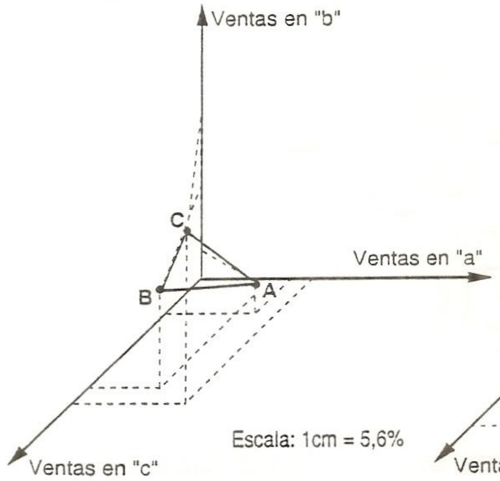
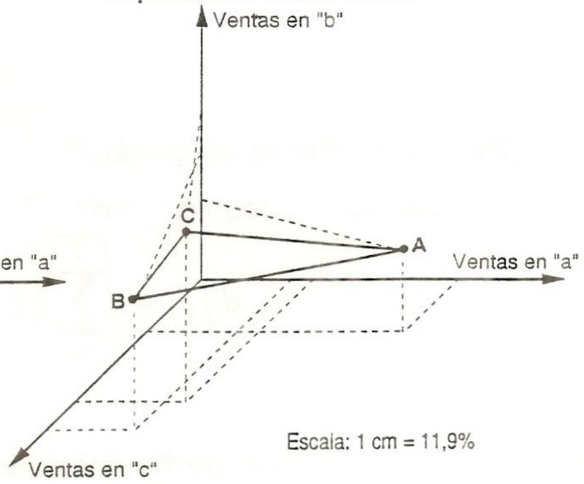


Tabla de Perfiles Asociada a la Tabla F
Espacio de Puntos-Línea



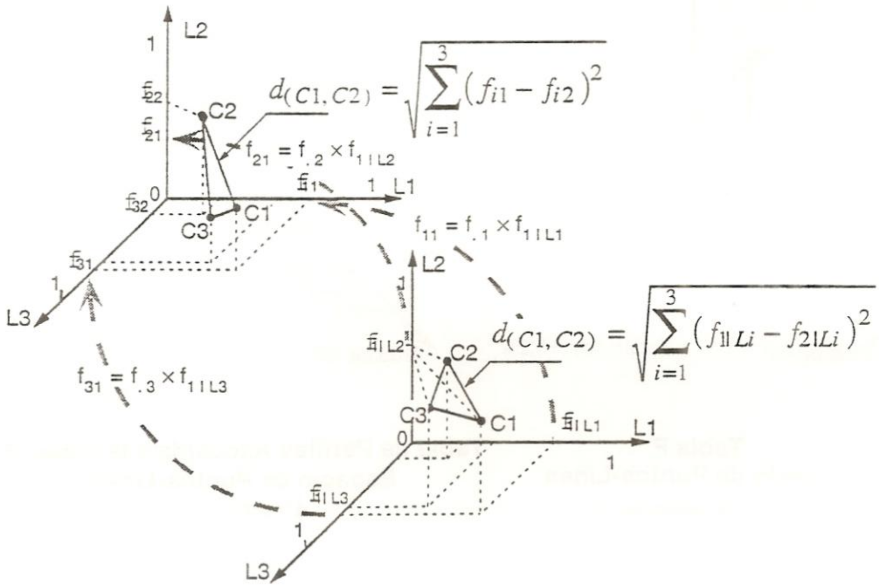
➤ Que transformação sofreram os espaços de representação dos elementos da tabela F...?

D. Representação gráfica da informação contida numa Tabela de Contingência

4. Representação da informação em termos de comparação de perfis

5. Relação entre a informação das tabelas de perfis e a informação da tabela F : (continuação)

Cada eixo sofre uma dilatação diferente. Cada eixo sofre uma «dilatação» proporcional ao «peso» dos elementos na tabela F.



A avaliação digital da distância também se transforma :

- Distância entre duas linhas da tabela F :

$$d_{(i,i')} = \sqrt{\sum_{j=1}^{J}\left(f_{ij} - f_{i'j}\right)^2}$$

- Distância entre dois perfis-linhas associados o Tabela F :

$$d_{(i,i')} = \sqrt{\sum_{j=1}^{J}\left(\frac{f_{ij}}{f_{i.}} - \frac{f_{i'j}}{f_{i'.}}\right)^2} = \frac{1}{f_{i.}f_{i'.}}\sqrt{\sum_{j=1}^{J}\left(f_{i'.}f_{ij} - f_{i.}f_{i'j}\right)^2}$$

D. Representação gráfica da informação contida numa Tabela de Contingência

4. Representação da informação em termos de comparação de perfis

6. Distorção da informação trazida pelas tabelas de Perfis

- A correção da informação, mediante a comparação dos perfis, ainda é insuficiente...

Assim, torna - se necessário comparar os perfis ponderados.

D. Representação gráfica da informação contida numa Tabela de Contingência

5. Representação da informação em termos de comparação ponderada de perfis

1. Comparação ponderada de perfis :

- Expressao da distáncia entre dois perfis-linha :

$$d_{(i,i')}\sqrt{\sum_{j=1}^{J}\left(\frac{f_{ij}}{f_{i.}}-\frac{f_{i'j}}{f_{i'.}}\right)^2} \qquad \forall\, i, i'$$

☞ A «contribuição do atributo j para a comparação de dois perfis - linha» = valor de cada termo da soma.

- Expressao da distáncia entre dois perfis-coluna :

$$d_{(j,j')}\sqrt{\sum_{i=1}^{I}\left(\frac{f_{ij}}{f_{.j}}-\frac{f_{ij'}}{f_{.j'}}\right)^2} \qquad \forall\, j, j'$$

☞ A «contribuição do atributo i para a comparação de dois perfis-coluna» = valor de cada termo da soma.

Importância de cada elemento da tabela F = peso do elemento.

Se considera a contribuição de cada atributo para a comparação de dois elementos... em valor relativo ao peso do atributo.

Uma diferença entre dois perfis - linha para uma modalidade J na coluna, contribui tanto mais para a comparação das linhas quanto mais raro for o atributo J.

e necessário ponderar as contribuições para cada modalidade j (ou i) para a comparação de duas linhas (colunas) com o valor inverso do peso da coluna (linha).

D. Representação gráfica da informação contida numa Tabela de Contingência

5. Representação da informação em termos de comparação ponderada de perfis

2. Avaliação da comparação de perfis : a distância de χ^2 (ou do Qui²)

- Distância do Qui² entre dois perfis-linha i e i' :

$$d_{(i,i')}\sqrt{\sum_{j=1}^{J}\frac{1}{f_{.j}}\left(\frac{f_{ij}}{f_{i.}}-\frac{f_{i'j}}{f_{i'.}}\right)^2}=\sqrt{\sum_{j=1}^{J}\frac{1}{p_j}\left(\frac{f_{ij}}{f_{i.}}-\frac{f_{i'j}}{f_{i'.}}\right)^2} \qquad \forall\, i,i' \in I$$

- Distância do Qui² entre dois perfis-coluna j e j' :

$$d_{(j,j')}\sqrt{\sum_{i=1}^{I}\frac{1}{f_{i.}}\left(\frac{f_{ij}}{f_{.j}}-\frac{f_{ij'}}{f_{.j'}}\right)^2}=\sqrt{\sum_{i=1}^{I}\frac{1}{p_i}\left(\frac{f_{ij}}{f_{.j}}-\frac{f_{ij'}}{f_{.j'}}\right)^2} \qquad \forall\, j,j' \in J$$

⇨ «Distância do Qui²» ou « distância de Benzécri»

⇨ Trata-se de uma distância euclidiana

A «distância do Qui²» neutraliza todas as distorções na representação da informação da Tabela de Contingência

D. Representação gráfica da informação contida numa Tabela de Contingência

5. Representação da informação em termos de comparação ponderada de perfis

3. Espaço de representação com a distância do Qui^2 :

Se associa às Tabelas de perfis as tabelas de Perfis Ponderados.

- cada perfil-linha é dividido por $\sqrt{p_j}$.

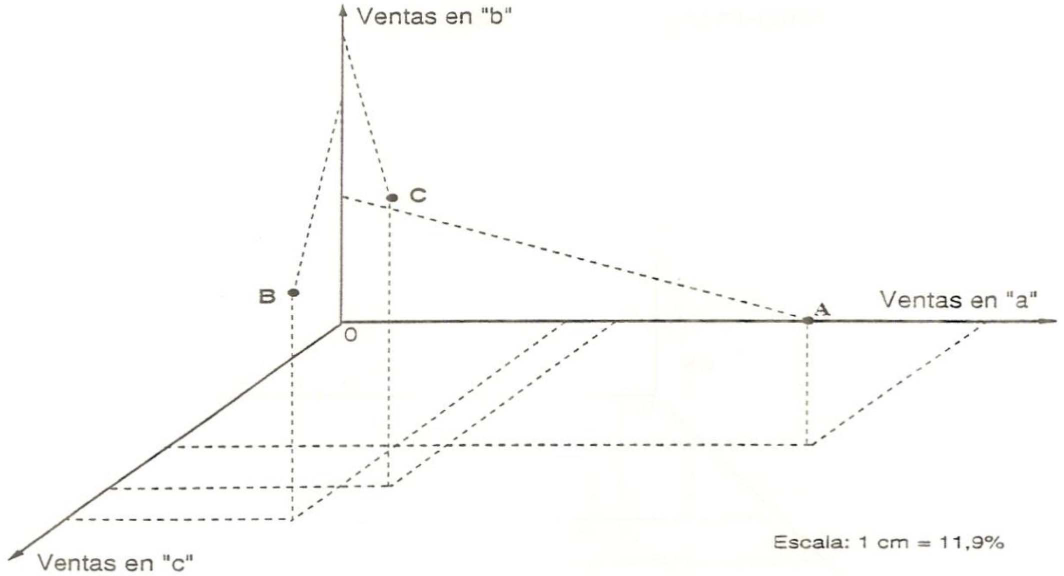
- cada perfil-coluna é dividido por $\sqrt{p_i}$.

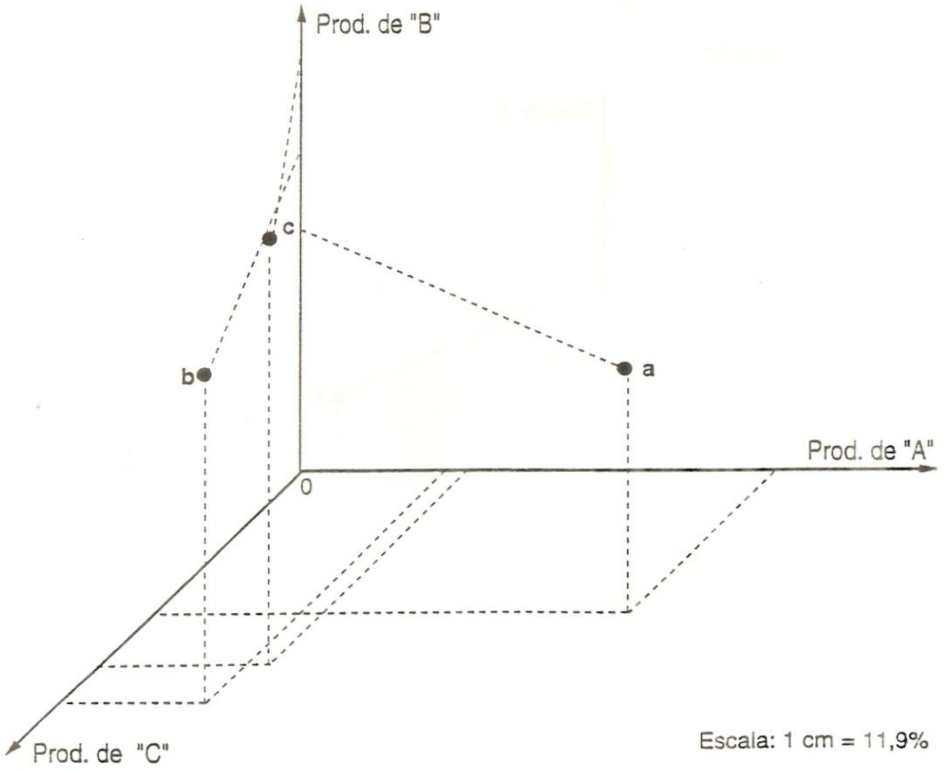
As Tabelas de Perfis-Ponderados, em linha ou em coluna, são as coordenadas de representação dos elementos da tabela M.A.C.T.(3 x 3) num espaço euclidiano

D. Representação gráfica da informação contida numa Tabela de Contingência

3. Distância do Qui^2 (exemplo numérico)

Representação gráfica da Tabela de Perfis - ponderados associados à tabela M.A.C.T.(3 x 3)

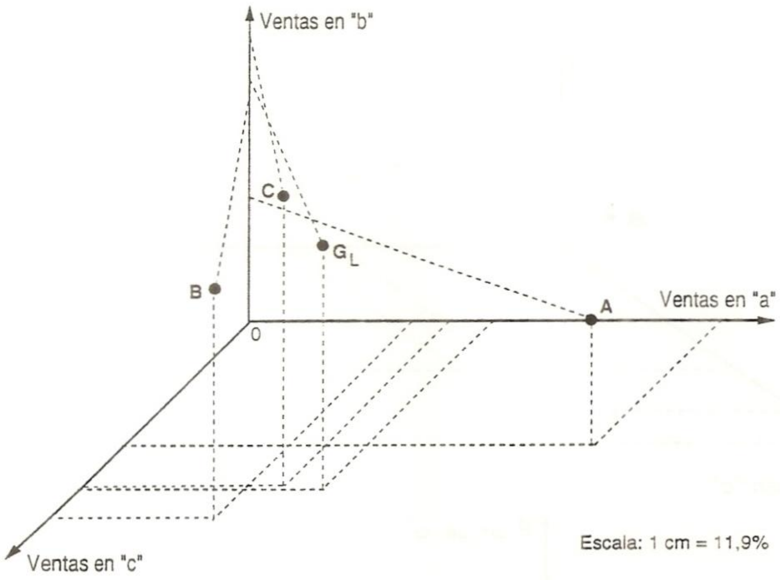


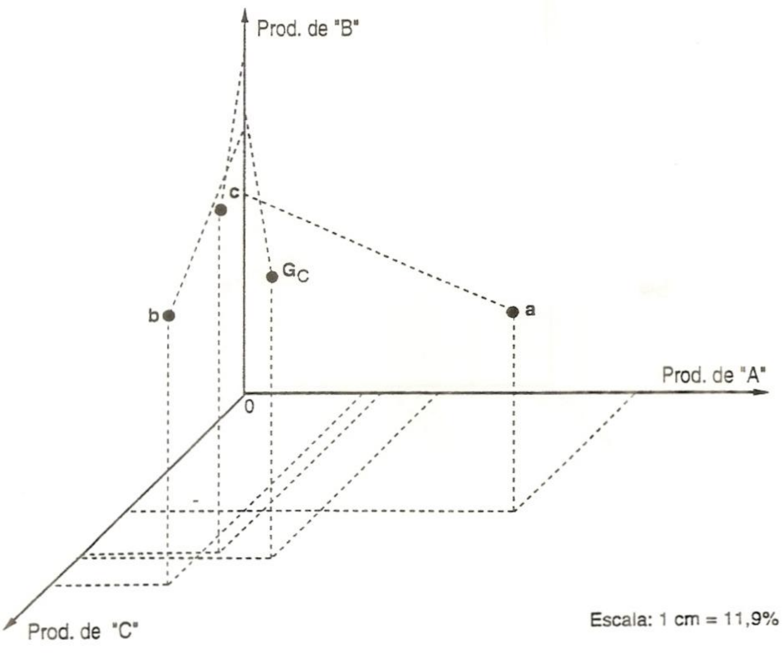


D. Representação gráfica da informação contida numa Tabela de Contingência

3. Distância do Qui² (exemplo numérico)

Perfis ponderados na margem ⇒ representação do ponto-margem na linha e do ponto-margem na coluna, respectivamente.





5. Representação da informação em termos de comparação ponderada de perfis

4. Correspondência entre os dois modos de representação da informação das tabelas de perfis - ponderados

- Se pode demonstrar facilmente que o modo digital de avaliação da comparação de dois perfis-ponderados é compatível com o modo analógico de representação dos conteúdos da informação das tabelas de perfis -ponderados.

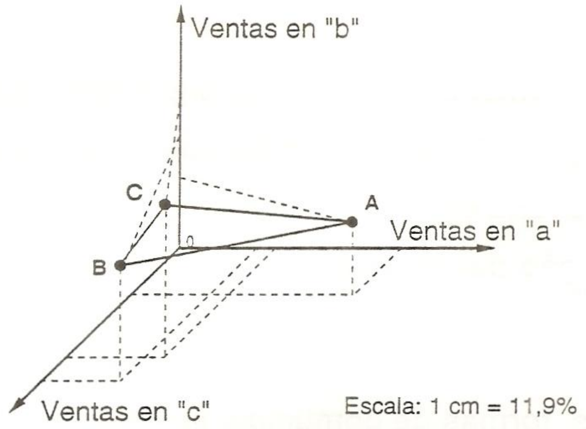
- As duas formas de comunicação da informação se apóiam em uma comparação de perfis que neutralizam todas as distorções dos conteúdos da informação.

D. Representação gráfica da informação contida numa Tabela de Contingência

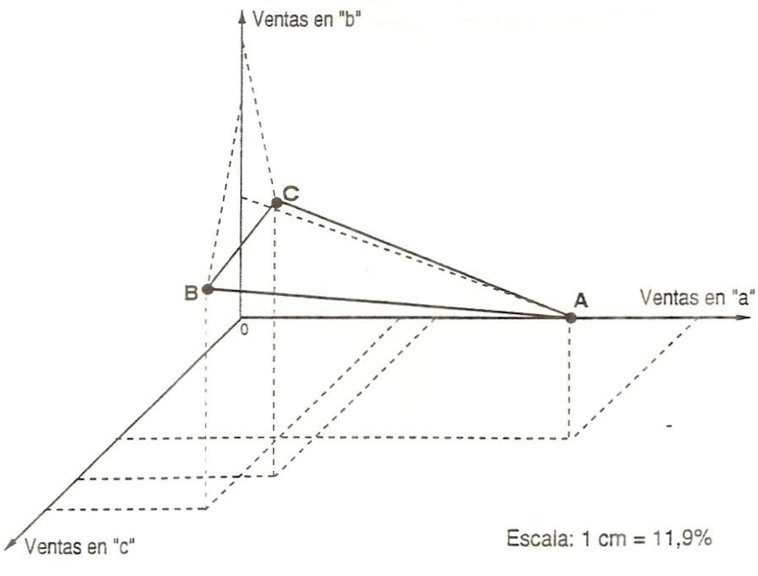
5. Representação da informação em termos de comparação ponderada de perfis

5. Comparação : Tabela F e tabela de Perfis-Ponderados.

Pontos - perfis em linha da Tabela F



Pontos - perfis ponderados em linha



D. Representação gráfica da informação contida numa Tabela de Contingência

5. Representação da informação em termos de comparação ponderada de perfis

5. Comparação : Tabela F e tabela de Perfis-Ponderados (continuação)

- Que deformação sofreram os espaços de representação dos pontos-perfis na nova metáfora da informação da tabela de contingência ...?

- A informação da Tabela F se apresenta agora sob a forma da «posição relativa dos pontos, dotados de massa, em um espaço euclidiano».

- Podemos fazer essas representações fazendo figurar a massa associada a cada ponto.

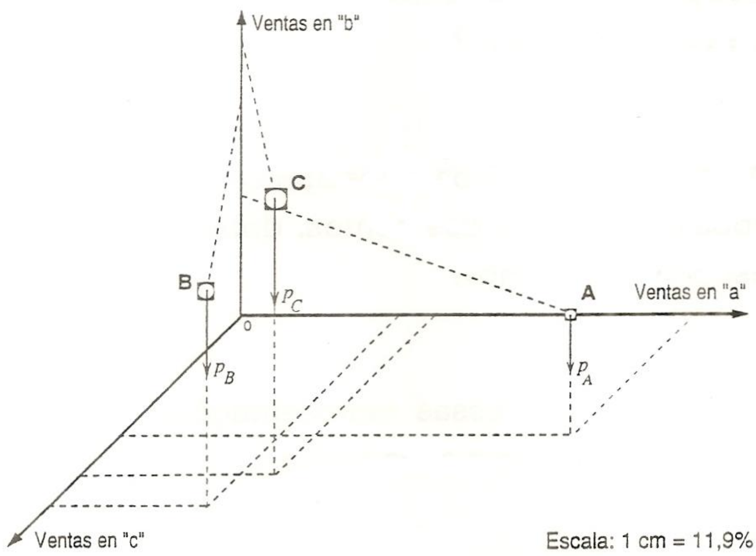
Construímos assim uma nova metáfora analógica da informação da Tabela F.

D. Representação gráfica da informação contida numa Tabela de Contingência

5. Representação da informação em termos de comparação ponderada de perfis

5. Comparação : Tabela F e tabela de Perfis-Ponderados (continuação)

Pontos Perfis - coluna dotados de peso



A forma da informação trazida por esta nova metáfora gráfica não pode ser lida em termos de posição relativa de pontos - perfis uma vez que eles estão dotados de massa.

D. Representação gráfica da informação contida numa Tabela de Contingência

5. Representação da informação em termos de comparação ponderada de perfis

6. Propriedades da distância do Qui ^{2} : a equivalência distribucional

- A distância do Qui^2 satisfaz a propriedade chamada de «equivalência distribucional».

Em espaços dotados da distáncia do Qui^2 , se duas linhas (coluna) são proporcionais (têm o mesmo perfil) se pode substituir essas linhas (coluna) por uma linha (coluna) igual à soma das duas linhas (colunas) proporcionais sem que isto modifique as distâncias entre as colunas (linhas).

- Verificação com um exemplo numérico :

Tabela T

	C1	C2	C3	
L1	10	9	5	24
L2	6	3	3	12
L3	8	2	4	14
	24	14	12	**50**

Tabela de Perfis associada a Tabela T

Perfis em Linha

	C1	C2	C3	
L1	0,42	0,38	0,20	1
L2	0,5	0,25	0,25	1
L3	0,57	0,14	0,29	1
	0,48	0,28	0,24	1

Perfis em Coluna

	C1	C2	C3	
L1	0,42	0,64	0,42	0,48
L2	0,25	0,22	0,25	0,24
L3	0,33	0,14	0,33	0,28
	1	1	1	1

D. Representação gráfica da informação contida numa Tabela de Contingência

5. Representação da informação em termos de comparação ponderada de perfis
6. Propriedades da distância do Qui ^{2} : a equivalência distribucional

Representação Gráfica

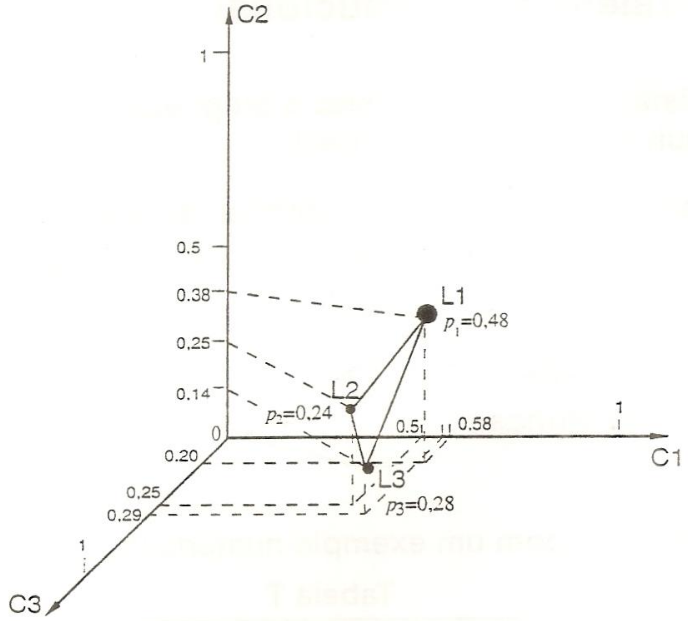
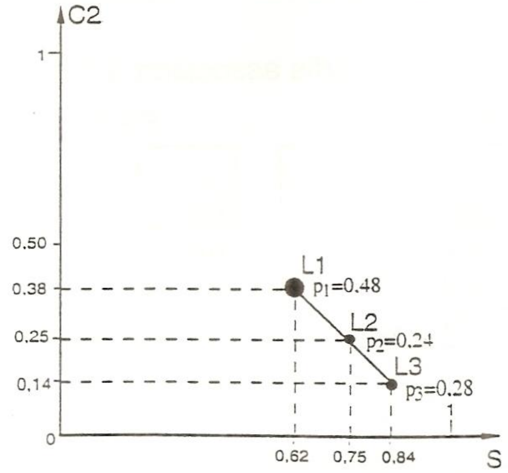


Tabela T'(3 x 2)

	S	C2	
L1	15	9	24
L2	9	3	12
L3	12	2	14
	36	14	50

Tabela de Perfis em linha de T

	S	C2	
L1	0.62	0.38	1
L2	0.75	0.25	1
L3	0.86	0.14	1
	0.72	0.28	50



D. Representação gráfica da informação contida numa Tabela de Contingência

5. Representação da informação em termos de comparação ponderada de perfis

6. Resumo :

Definimos assim um modo analógico e digital apropriado para comunicar a informação trazida por uma Tabela de Contingência, mas...

- Como devemos «ler» nesses espaços de representação a forma da informação da Tabela F ... ?

outra pergunta ...

- Como avaliar globalmente a informação trazida por uma Tabela F...?

D. Representação gráfica da informação contida numa Tabela de Contingência

6. Representação da informação em termos de inércia de uma nuvem de pontos.

1. Inércia de uma nuvem de pontos-perfis munidos de peso

- Inércia de uma nuvem de pontos N(I) com respeito a um ponto m qualquer → dispersão da nuvem de pontos em torno deste ponto.

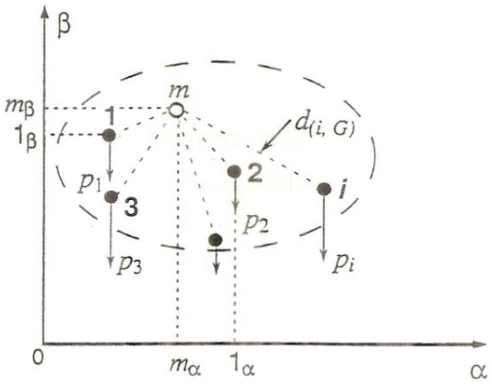
- Como avaliar a dispersão da nuvem de pontos em torno de um ponto m...?

Seja N(I) :

- um conjunto de pontos - perfis : $I = \{1, 2, 3, ..., i\}$

- munido de pesos, $p_i = \{p_1, p_2, p_3, \dots, p_i\}$,

Através de suas coordenadas sobre os eixos α e β, estes pontos são localizados em um espaço euclidiano.



A inércia (dispersão) da nuvem de pontos N(I) com respeito a m (um ponto qualquer do mesmo espaço)

é assim definida :

$$I_m = \sum_{i=1}^{I} p_i d^2(i,m) \quad \forall i \in I$$

A inércia (dispersão) da nuvem de pontos com respeito a m é igual à soma, para todos os pontos, do produto do quadrado da distância de cada ponto a m pelo peso associado a cada ponto.

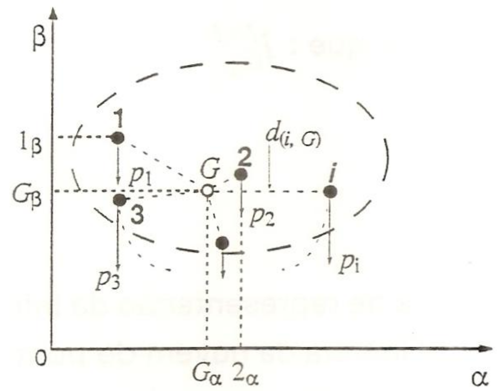
D. Representação gráfica da informação contida numa Tabela de Contingência

6. Representação da informação em termos de inércia de uma nuvem de pontos.

1. Inércia de uma nuvem de pontos-perfis munidos de peso (continuação)

A «contribuição do ponto i « para a inércia com respeito a m da nuvem de pontos N(I) é igual ao termo correspondente a i na soma que mede a dispersão da nuvem N(I) com respeito a m.

$$\mathrm{Contr}\, I_m(i) = p_i \times d^2(i,m) \quad \forall i \in I$$



Os pontos da nuvem N(I) estão em equilíbrio em torno de G se a rotação do sistema de pontos em torno de G é nula.

D. Representação gráfica da informação contida numa Tabela de Contingência

6. Representação da informação em termos de inércia de uma nuvem de pontos.

2. Inércia dos pontos-perfis linha ponderados com respeito aos pontos G_L e G_C :

$$I_{G_L}^{N(I)} = \sum_{i=1}^{I} \text{Contr}\, G_L(i)$$

$$I_{G_C}^{N(J)} = \sum_{j=1}^{J} \text{Contr}\, G_C(j)$$

e se pode demonstrar que : $I_{G_L}^{N(I)} = I_{G_C}^{N(J)}$

Em todos os espaços de representação da informação das Tabelas de Perfis Ponderados, a inércia da nuvem de pontos N(I) com respeito a G_L é a mesma que a inércia da nuvem de pontos N(J) com respeito a G_C.

Que significa esta dispersão comum a essas duas nuvens de pontos - perfis...?

D. Representação gráfica da informação contida numa Tabela de Contingência

6. Representação da informação em termos de inércia de uma nuvem de pontos.

3. Desvio da situação de independência : associação de duas variáveis qualitativas.

Tabela de independência ou tabela teórica :

Termo geral : $f_{ij}^{*} = f_{i.} \times f_{.j}$

com : $\sum_{i=1}^{I} f_{i.}^{*} = \sum_{i=1}^{I} f_{i.} = 1$ e $\sum_{j=1}^{J} f_{.j}^{*} = \sum_{j=1}^{J} f_{.j} = 1$

As distribuições da tabela teórica sao proporcionais às suas distribuições marginais:

$$\sum_{j=1}^{J} f_{ij}^{*} = f_{i.} \sum_{j=1}^{J} f_{.j} = f_{i.} \quad \forall i \in I$$

$$\sum_{i=1}^{I} f_{ij}^{*} = f_{.j} \sum_{i=1}^{I} f_{i.} = f_{.j} \quad \forall j \in J$$

Para medir a associação entre duas variáveis qualitativas deve - se avaliar a importância entre o afastamento entre aquilo que foi observado e a situação de independência.

D. Representação gráfica da informação contida numa Tabela de Contingência

6. Representação da informação em termos de inércia de uma nuvem de pontos.

3. Desvio da situação de independência : (continuação)

- Tabela de desvio da situação de independência

Termo geral : $e_{ij} = f_{ij} - f_{ij}^{*} \quad \forall i \in I; \forall j \in J$

Problemas na interpretação desta tabela...

- Tabela de desvios ponderados :

Termo geral : $e_{ij}^{*} = \dfrac{\left(f_{ij} - f_{ij}^{*}\right)^2}{f_{ij}^{*}} \quad \forall\, i \in I; \quad \forall\, j \in J$

- Interesse desta tabela :

A soma dos valores desta tabela é igual ao coeficiente do φ^2

$$\varphi^2 = \sum_{i=1}^{I} \sum_{j=1}^{J} e_{ij}^{*} = \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{\left(f_{ij} - f_{ij}^{*}\right)^2}{f_{ij}^{*}} \quad \forall\, i \in I; \quad \forall\, j \in J$$

O coeficiente do φ^2 pode ser expresso com os elementos da tabela F (I ,J) :

$$\varphi^2 = \sum_{i=1}^{I} \sum_{j=1}^{J} e_{ij}^{*} = \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{\left(f_{ij} - f_{i.} \times f_{.j}\right)^2}{f_{i.} \times f_{.j}} \quad \forall\, i \in I; \quad \forall j \in J$$

D. Representação gráfica da informação contida numa Tabela de Contingência

6. Representação da informação em termos de inércia de uma nuvem de pontos.

4. A inércia das nuvens N(I) e N(J), com respeito ao ponto G correspondente, mede o grau de associação das variáveis qualitativas da tabela F (I x J).

Demonstração : (Ver documento anexo) :

$$I_{G_j}^{N(J)} = \sum_j p_j \, d^2\left(j, G_j\right)$$

$$d^2_{\left(j,G_j\right)} = \sum_i \left(\frac{f_{ij}}{f_{.j}\sqrt{f_{i.}}} - \sqrt{f_{i.}} \right)^2$$

$$= \sum_i \left(\frac{f_{ij} - f_{.j} f_{i.}}{f_{.j}\sqrt{f_{i.}}} \right)^2 \quad = \sum_i \frac{\left(f_{ij} - f_{.j} f_{i.}\right)^2}{f_{.j}^2 f_{i.}}$$

$$I_{G_j}^{N(J)} = \sum_j f_{.j} \sum_i \frac{\left(f_{ij} - f_{.j} f_{i.}\right)^2}{f_{.j}^2 f_{i.}}$$

$$I_{G_j}^{N(J)} = \sum_j \sum_i \frac{\left(f_{ij} - f_{.j} f_{i.}\right)^2}{f_{.j} f_{i.}} \quad = \varphi^2 = \frac{\chi^2}{n}$$

6. Representação da informação em termos de inércia de uma nuvem de pontos.

5. Os pontos perfis - marginais G destes espaços são os Centros de Gravidade dos sistemas de massa

- Cálculo da média ponderada das coordenadas dos pontos-perfis sobre cada eixo do espaço...

• Considerando os valores absolutos das massas :

$p_{*_{G_L}} = n_{1.} + n_{2.} + \cdots + n_{i.} = n$ e $p_{*_{G_C}} = n_{.1} + n_{.2} + \cdots + n_{.j} = n$

• Coordenadas dos pontos-perfis coluna sobre o eixo i :

$$\left(\frac{f_{i1}}{f_{i.}\sqrt{f_{.1}}} ; \frac{f_{i2}}{f_{i.}\sqrt{f_{.2}}} ; \ldots ; \frac{f_{ij}}{f_{i.}\sqrt{f_{.j}}} \right)$$

• Média ponderada dessas coordenadas :

$$\bar{x} = \frac{1}{n}\sum_{j=1}^{J} n_{.j}\left(\frac{f_{ij}}{f_{.j}\sqrt{f_{i.}}}\right) = \frac{1}{n}\sum_{j=1}^{J} n_{.j}\sum_{j=1}^{J}\frac{f_{ij}}{f_{.j}\sqrt{f_{i.}}}$$

$$= \frac{1}{\sqrt{f_{i.}}}\sum_{j=1}^{J}\frac{f_{ij}}{f_{.j}} = \frac{1}{\sqrt{f_{i.}}}\frac{\sum_{j=1}^{J} f_{ij}}{\sum_{j=1}^{J} f_{.j}} = \frac{1}{\sqrt{f_{i.}}}\frac{f_{i.}}{1} = \sqrt{f_{i.}}$$

A média ponderada das coordenadas dos pontos de N (J) sobre o eixo i é igual à coordenada de G sobre o eixo i .

D. Representação gráfica da informação contida numa Tabela de Contingência

6. Representação da informação em termos de inércia de uma nuvem de pontos.

6. Leitura da informação de uma tabela F(I x J) associada a Tabela T(I x J)

Os pontos perfis-linha constituem um sistema de massas em equilíbrio em torno de G.

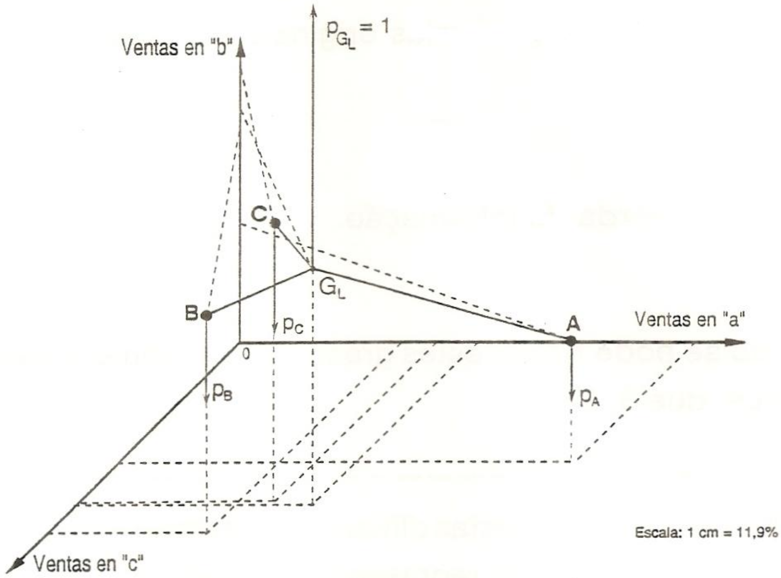
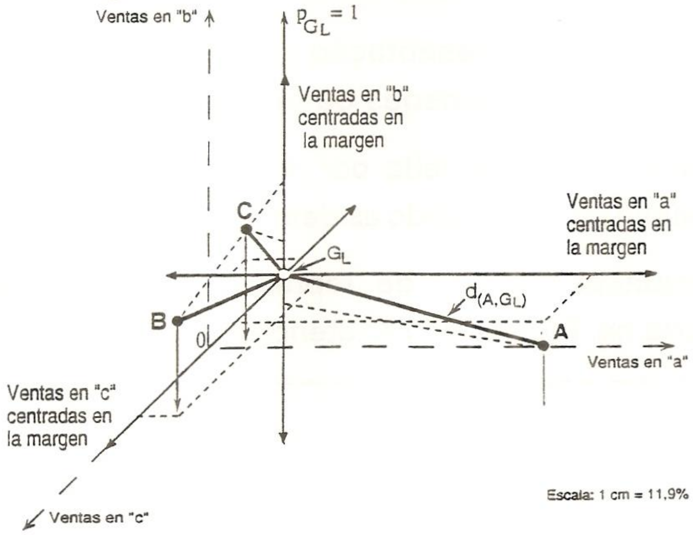


Imagem do mesmo espaço centrado em G



D. Representação gráfica da informação contida numa Tabela de Contingência

6. Representação da informação em termos de inércia de uma nuvem de pontos.

7. Resumo

☞ A operação que consiste em «centrar» a nuvem de pontos-perfis nao modifica as distáncias originais de cada ponto-perfil com respeito a G.

☞ Nao há perda de informação.

☞ Nao se pode traçar estes gráficos se as dimensões do espaço sao maiores que 3.

Para se contornar estas dificuldades torna-se necessário produzir uma nova representação desses espaços de maneira :

- que essa representação possa ser feita quaisquer sejam as dimensoes da tabela T(J x K)
- que possa ser feita por meio de gráficos de duas dimensões evitando as deformações de perspectiva;

Os espaços fatoriais de representação da informação contida na Tabela T(J x K) atendem a estas exigências

E. Critérios a serem respeitados na busca de um novo referencial de representação

1. Forma dos espaços dos pontos - perfis ponderados e independência das variáveis qualitativas.

- Toda tabela observada apresenta uma certa associação entre as características observadas.

- O coeficiente $\varphi^2 > 0$ e a inércia $I_{GL} = I_{GC} > 0$

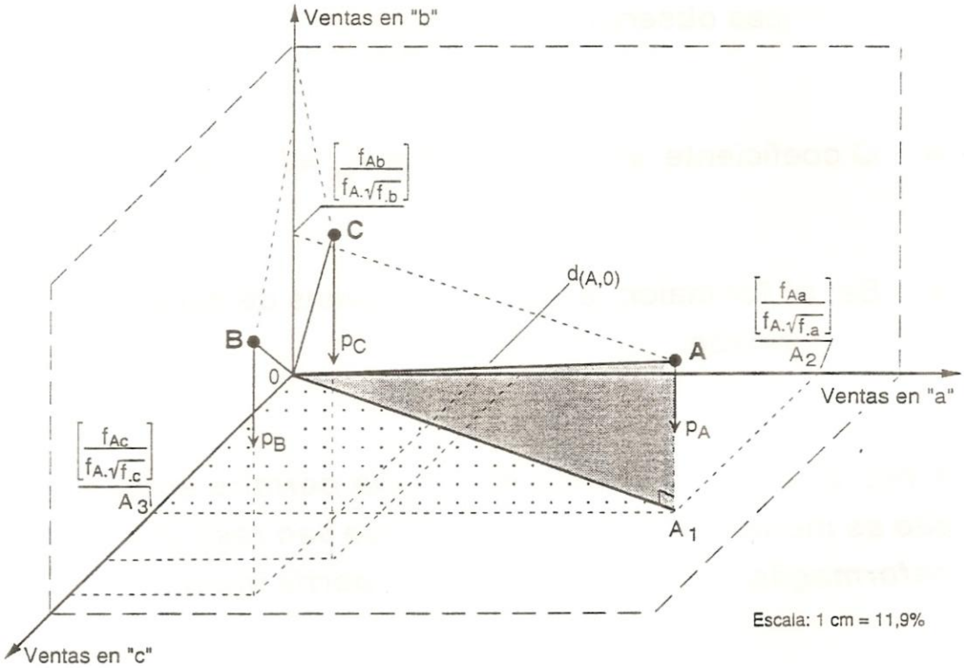
- Se φ^2 for maior, a forma das nuvens de pontos N(I) e N(J) se afasta da esferidade.

A representação gráfica da inércia permite estudar quais são as modalidades da Tabela T que sao responsáveis pela deformação da nuvem de pontos-perfis munidos de peso.

Para compreender como se pode estudar a deformação de uma nuvem de pontos-perfis munidos de peso, é necessário verificar como se pode decompor a inércia de uma nuvem de pontos-perfis...

2. Decomposição da inércia das nuvens de pontos N(I) e N(J) com respeito o origem

Representação gráfica dos pontos-perfis linha de uma tabela T em um espaço euclidiano, com a distância do Qui²



Esse gráfico permite calcular a inércia dos pontos da nuvem N(I) com respeito a origem.

Os triángulos retángulos sombreados $(A, \hat{A}_1, O)$ y $(A_1, \hat{A}_3, O)$ permitem mostrar que,

$$d^2_{(A,0)} = \frac{f^2_{Aa}}{f^2_{A}.f_{.a}} + \frac{f^2_{Ab}}{f^2_{A}.f_{.b}} + \frac{f^2_{Ac}}{f^2_{A}.f_{.c}}$$

2. Decomposição da inércia das nuvens de pontos N(I) e N(J) com respeito o origem

■ Definição da inércia ponto-perfil em linha com respeito o origem :

$$I_0 = \sum_{i=1}^{I} p_i \, d^2_{(i,0)} \tag{1}$$

Desenvolvendo esta expressão :

$$\begin{aligned} I_0 = {} & p_A \left[\frac{f^2_{Aa}}{f^2_{A.} f_{.a}} + \frac{f^2_{Ab}}{f^2_{A.} f_{.b}} + \frac{f^2_{Ac}}{f^2_{A.} f_{.c}} \right] + \\ & p_B \left[\frac{f^2_{Ba}}{f^2_{B.} f_{.a}} + \frac{f^2_{Bb}}{f^2_{B.} f_{.b}} + \frac{f^2_{Bc}}{f^2_{B.} f_{.c}} \right] + \\ & p_C \left[\frac{f^2_{Ca}}{f^2_{C.} f_{.a}} + \frac{f^2_{Cb}}{f^2_{C.} f_{.b}} + \frac{f^2_{Cc}}{f^2_{C.} f_{.c}} \right] \end{aligned}$$

e reordenando os termos ...

$$\begin{aligned} I_0 = {} & \left[p_A \frac{f^2_{Aa}}{f^2_{A.} f_{.a}} + p_B \frac{f^2_{Ba}}{f^2_{B.} f_{.a}} + p_C \frac{f^2_{Ca}}{f^2_{C.} f_{.a}} \right] + \\ & \left[p_A \frac{f^2_{Ab}}{f^2_{A.} f_{.b}} + p_B \frac{f^2_{Bb}}{f^2_{B.} f_{.b}} + p_C \frac{f^2_{Cb}}{f^2_{C.} f_{.b}} \right] + \\ & \left[p_A \frac{f^2_{Ac}}{f^2_{A.} f_{.c}} + p_B \frac{f^2_{Bc}}{f^2_{B.} f_{.c}} + p_C \frac{f^2_{Cc}}{f^2_{C.} f_{.c}} \right] \end{aligned} \tag{2}$$

E. Critérios a serem respeitados na busca de um novo referencial de representação

2. Decomposição da inércia das nuvens de pontos N(I) e N(J) com respeito o origem (continuação)

■ Em (2), os termos entre colchetes constituem a inércia projetada ao longo do eixo a, do eixo b e do eixo c.

■ Inércia projetada = «inércia ao longo de um eixo» :

- A inércia de N(I) com respeito o origem :

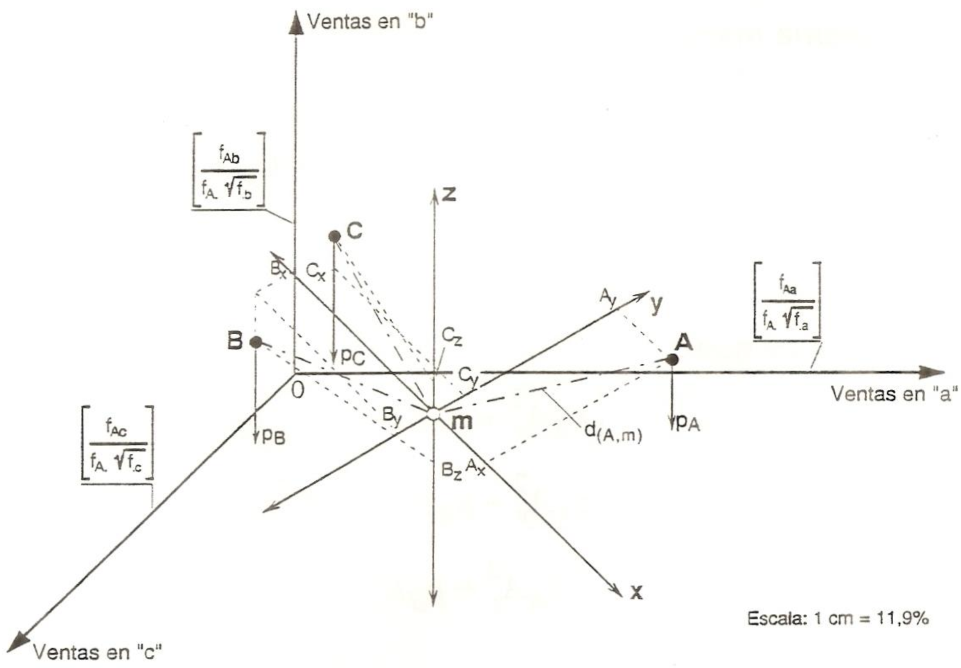
$$I_O^{N(I)} = I_{Largo_a}^{N(I)} + I_{Largo_b}^{N(I)} + I_{Largo_c}^{N(I)}$$

- A inércia de N(J) com respeito o origem :

$$I_O^{N(J)} = I_{Largo_A}^{N(J)} + I_{Largo_B}^{N(J)} + I_{Largo_C}^{N(J)}$$

3. Inércia das nuvens de pontos perfis N(I) e N(J) com respeito a um ponto m qualquer do espaço.

Por m passam três eixos ortogonais (x , y , z) ...



A inércia de N(I) com respeito a m :

$$I_m = \sum_{i=1}^{I} p_i \, d^2_{(i,m)} = p_A \, d^2_{(A,m)} + p_B \, d^2_{(B,m)} + p_C \, d^2_{(C,m)}$$

No entanto a distância do ponto A ao ponto m é :

$$d^2_{(A,m)} = A_x^2 + A_y^2 + A_z^2 \quad \textbf{com} \quad A_z^2 = 0$$

E. Critérios a serem respeitados na busca de um novo referencial de representação

3. Inércia das nuvens de pontos perfis N(I) e N(J) com respeito a um ponto m qualquer do espaço (continuação).

da mesma maneira :

$$d^2_{(B,m)} = B_x^2 + B_y^2 + B_z^2 \quad y \quad d^2_{(C,m)} = C_x^2 + C_y^2 + C_z^2$$

concluímos que :

$$I_m^{N(I)} = \left(p_A A_x^2 + p_B B_x^2 + p_C C_x^2\right) + \left(p_A A_y^2 + p_B B_y^2 + p_C C_y^2\right) + \left(p_A A_z^2 + p_B B_z^2 + p_C C_z^2\right)$$

A inércia de N(I) com respeito ao ponto m é igual a soma da inércia da nuvem com respeito a m «ao longo de um eixo x» ... mais a inércia da nuvem com respeito a m «ao longo de um eixo y» perpendicular a x ... mais a inércia da nuvem com respeito a m «ao longo de um eixo z» perpendicular a x e a y.

$$I_m^{N(I)} = I^{N(I)}_{//x^m} + I^{N(I)}_{//y^m \perp x^m} + I^{N(I)}_{//z^m \perp y^m \perp x^m}$$

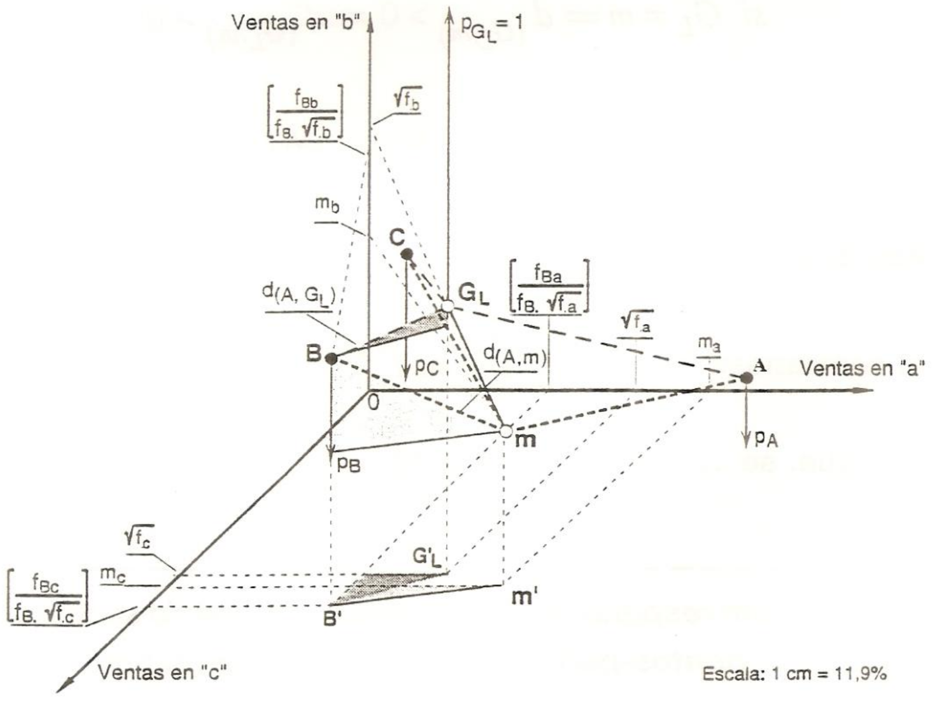
E. Critérios a serem respeitados na busca de um novo referencial de representação

4. Importância da descrição da forma das nuvens de pontos centradas em G.

Teorema de Huygens :

A inércia de uma nuvem de pontos N(I) com respeito a um ponto m do espaço pode ser descomposta na soma da inércia da nuvem de pontos N(I) com respeito ao Centro de Gravidade G_L mais a inércia do Centro de Gravidade, munido de massa total da nuvem de pontos, com respeito ao ponto m.

O gráfico permite compreender como se demonstra este teorema.



(demostração ... ver documento anexo)

E. Critérios a serem respeitados na busca de um novo referencial de representação

5. O ponto G de uma nuvem de pontos-perfis é um ponto característico do espaço

Pelo teorema de Huygens :

$$I_m^{N(I)} = I_{G_L}^{N(I)} + M_{tot}\, d^2_{(G_L,m)}$$

Dado que as massas de cada ponto são positivas, a soma das massas é > 0, de modo que : $M_{tot} > 0$.

As distáncias entre dois pontos são sempre > 0 de modo que :

$$si\ \ G_L \equiv m \Rightarrow d_{(G_L,m)} = 0$$

$$si\ \ G_L \neq m \Rightarrow d_{(G_L,m)} > 0 \Rightarrow d^2_{(G_L,m)} > 0$$

e

$$si\ \ G_L \neq m \Rightarrow I_m^{N(I)} > I_{G_L}^{N(I)}$$

dado que, se ... $d_{(G_L,m)} > 0 \Rightarrow M_{tot}\, d^2_{(G_L,m)} > 0$

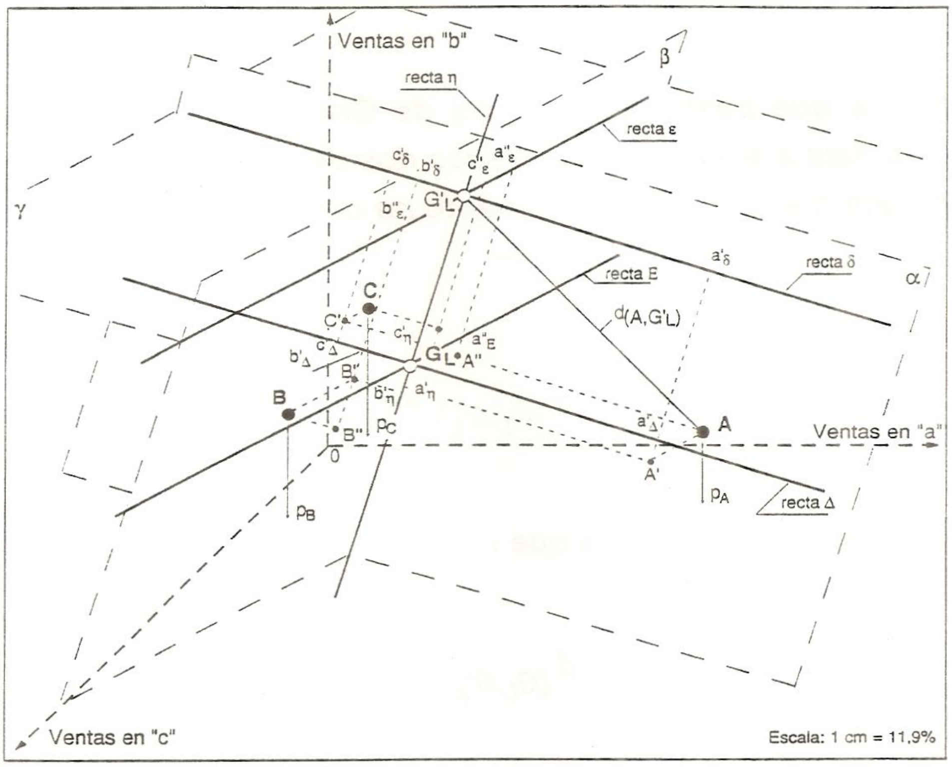
em contraposição, $si\ \ G_L \equiv m \Rightarrow I_m^{N(I)} = I_{G_L}^{N(I)}$

dado que, se ... $d_{(G_L,m)} = 0 \Rightarrow M_{tot}\, d^2_{(G_L,m)} = 0$

A inércia, com respeito ao Centro de Gravidade do sistema de massas dos pontos-perfis constitue um valor mínimo.

Importáncia desta conclusão...

6. Redução de uma nuvem de pontos-perfis ao Centro de Gravidade



A partir desse gráfico, é possível demostrar que :

$$I_m^{N(I)} = I_{G_L}^{N(I)} + M_{tot}\, d^2_{(G_L,m)}$$

De modo que a inércia de N(I) com respeito a G'_L pode ser assim formulada :

$$I_{G'_L}^{N(I)} = I_{\substack{G_L \\ //\Delta}}^{N(I)} + I_{\substack{G_L \\ //E\perp\Delta}}^{N(I)} + I_{\substack{G_L \\ //E\perp\Delta\perp\eta}}^{N(I)} + M_{tot}\, d^2_{(G_L,G'_L)}$$

6. Redução de uma nuvem de pontos-perfis ao Centro de Gravidade (continuação)

A reta que passa pelo Centro de Gravidade é a reta que minimiza a soma dos quadrados dos desvios dos pontos da nuvem medidos ortogonalmente na direção da reta.

Além disso :

$$si \;\; G'_L \neq G_L \Rightarrow d_{(G_L, G'_L)} > 0 \Rightarrow I_{G'_L}^{N(I)} > I_{G_L}^{N(I)}$$

Mas também se verifica que :

$$si \;\; G'_L \equiv G_L \Rightarrow d_{(G_L, G'_L)} = 0 \Rightarrow I_{G'_L}^{N(I)} = I_{G_L}^{N(I)}$$

Todas estas propriedades podem ser generalizadas no caso de espaços de mais de três dimensoes.

F. Critérios a serem respeitados na busca de um novo referencial de representação

1. Principais características do novo referencial

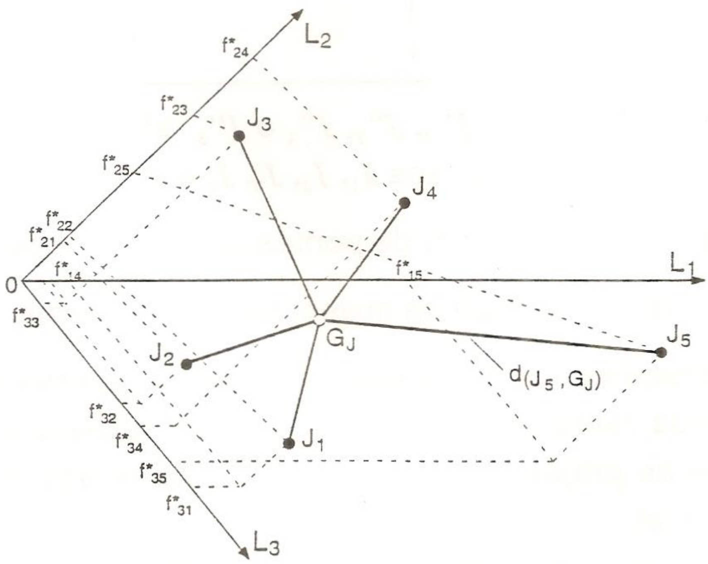
O novo referencial deve ser :

- centrado em G_L Y em G_C , respectivamente... porque...?
- um sistema de eixos ortogonais por construção, porque...?
- como fazer para construir um sistema de eixos ortogonais...?

Admitindo-se que sabemos definir esse sistema de eixos...

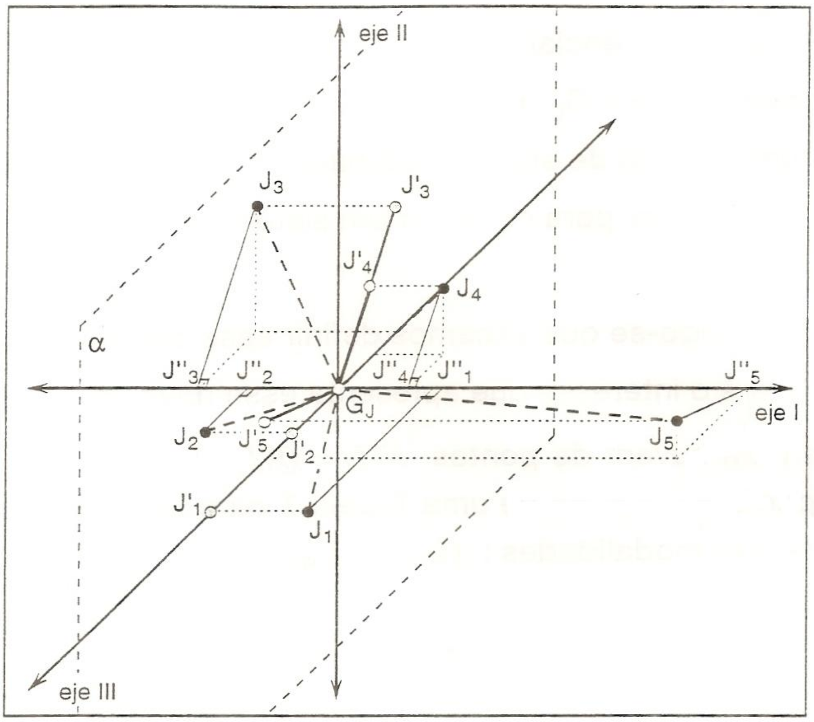
Qual é o interesse que apresenta esse novo referencial...?

Seja uma nuvem de pontos $N(J) = Q\{J_1, J_2, J_3, J_4, J_5\}$ de cinco pontos-perfis em coluna de uma Tabela T na qual a variável em linha apresenta três modalidades : $\{L_1, L_2, L_3\}$.



Nesse espaço, criamos um novo referencial centrado e ortogonal (por construção)...

2. Referencial ortogonal e centrado em G_J



Os pontos-perfis J''_1, J''_2, J''_3, J''_4 e J''_5 são determinados pela projeção ortogonal dos pontos J_1, J_2, J_3, J_4 e J_5 sobre o eixo I.

Qual é a inércia da nuvem de pontos-perfis ao longo do eixo I...?

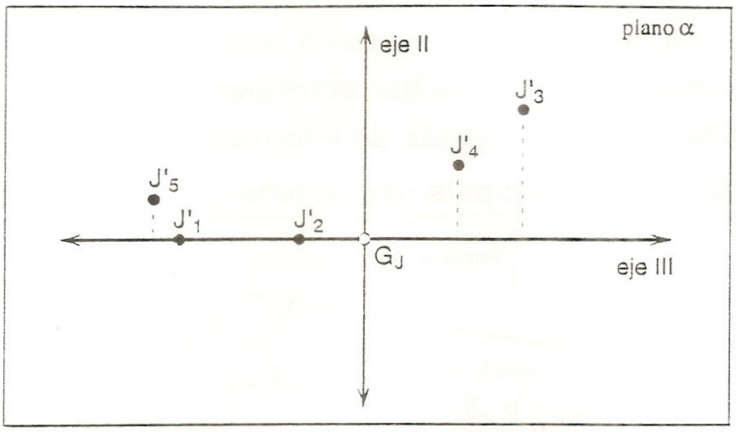
Qual é a inércia residual da nuvem de pontos-perfis...?

Para representar as distâncias de todos os pontos com respeito a G_J, podemos representar – em todos os sub-espaços de duas dimensões – as projeções da nuvem de pontos sobre os eixos do novo referencial.

Primeiro sub-espaço de uma dimensao

F. Critérios a serem respeitados na busca de um novo referencial de representação

3. Segundo sub-espaço de duas dimensões



A ortogonalidade dos eixos permite apresentar a inércia da nuvem com respeito a G em uma sucessao de gráficos dos sub-espaços de uma ou duas dimensões.

O novo tipo de referencial apresenta três vantagens importantes :

❶ Permite a análise completa da forma da nuvem de pontos-perfis;

❷ Permite a representação da forma da nuvem de pontos-perfis qualquer que sejam as dimensões das nuvens de pontos.

❸ Permite produzir a representação objetiva dessas nuvens de pontos, quer dizer decidir «estáveis», independentes do ponto de vista do analista.

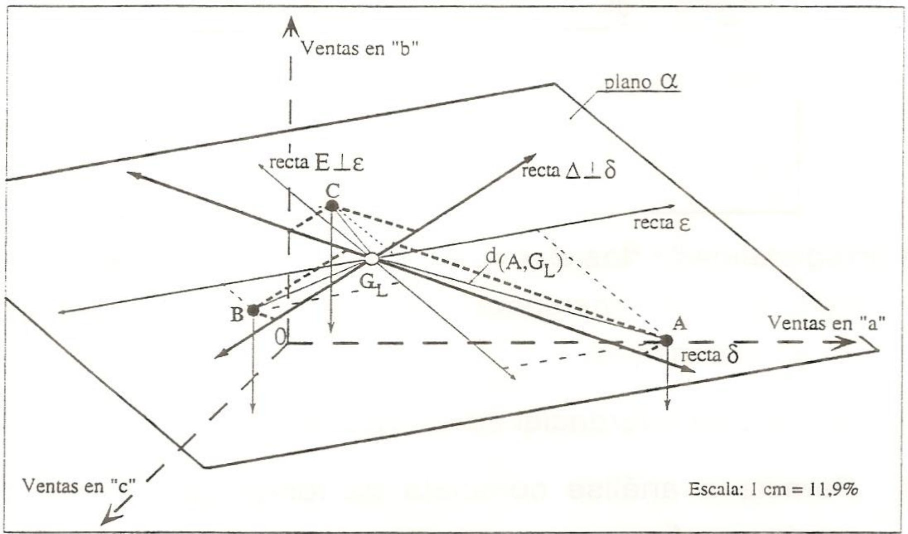
ê possível associar a essas representações gráficas os coeficientes de controle que constituem o componente digital complementária desse modo analógico de comunicar o sentido da informação trazida pela Tabela de Contingência.

F. Critérios a serem respeitados na busca de um novo referencial de representação

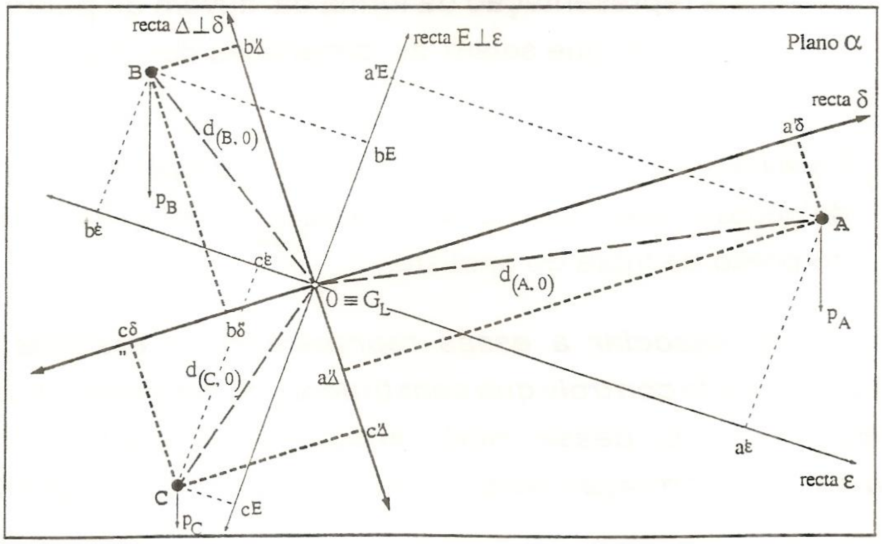
2. Como construir uma «boa» representação da informação da Tabela T...?

Por um ponto do espaço passam uma infinidade de sistemas de eixos ortogonais... Como escolher entre eles o «bom» sistema que sirva de referencial de representação de informação da Tabela T...?

Direções principais das nuvens de pontos-perfis



Plano α em um gráfico de duas dimensões ...



F. Critérios a serem respeitados na busca de um novo referencial de representação

3. Como escolher entre esses sistemas de eixos igualmente possíveis...?

Qual deles indica a «melhor» representação da dispersão da nuvem de pontos em torno de G...?

■ Critério de seleção...

O menor referencial é aquele para o qual as projeções dos pontos-perfis sobre um eixo sejam as maiores possíveis , de tal modo que os desiíos da nuvem de pontos-perfis — medidos ortogonalmente em relação ao eixo— sejam os menores possíveis.

Com esse critério podemos escolher o melhor referencial entre os dois possíveis...

Se queremos obter, por projeção ortogonal, a melhor imagem possível de uma nuvem de pontos-perfis é necessário projetar os pontos de uma nuvem sobre uma reta que maximize a inêrcia projetada da nuvem ao longo dela.

Ou então , é necessário projetar a nuvem ao longo de uma reta em relação a qual, a inércia residual ortogonal da nuvem de pontos-perfis seja mínima.

■ A reta que responde a essa exigência é chamada «direção principal de deformação» da nuvem de pontos-perfis.

4. Processo de construção de um novo referencial

A definição do novo referencial se faz em quatro etapas :

1) PRIMEIRA ETAPA : se busca a primeira direção principal de deformação da nuvem de pontos-perfis N(I) e N(J). Se determina, por exemplo, reta p_1.

2) SEGUNDA ETAPA : se define o sub-espaço de J-1 dimensões, ortogonal à primeira direção principal escolhida (se define a inércia residual ortogonal da nuvem de pontos),

3) TERCEIRA ETAPA : se busca nesse sub-espaço, a segunda direção principal de deformação e se determina a reta p_2.

4) QUARTA ETAPA : se repetem as etapas 1, 2 e 3. até que não haja mais inércia residual ortogonal a projetar em um último sub-espaço ortogonal.

A inércia I_G das nuvens de pontos- perfis N(I) e N(J) será assim decomposta em :

$$I_G = I^{N(I)}_{//\overset{\rho}{\delta_1}} + I^{N(I)}_{//\overset{\rho}{\delta_2}} + \ldots + I^{N(I)}_{//\overset{\rho}{\delta_I}} = I^{N(J)}_{//\overset{\rho}{\delta_1}} + I^{N(J)}_{//\overset{\rho}{\delta_2}} + \ldots + I^{N(J)}_{//\overset{\rho}{\delta_J}}$$

dado que : $I_G = I^{N(I)}_{GL} = I^{N(J)}_{GC}$, sendo...

$$I^{N(I)}_{//\overset{\rho}{\delta_1}} \geq I^{N(I)}_{//\overset{\rho}{\delta_2}} \geq \ldots \geq I^{N(I)}_{//\overset{\rho}{\delta_I}} \quad y \quad I^{N(J)}_{//\overset{\rho}{\delta_1}} \geq I^{N(J)}_{//\overset{\rho}{\delta_2}} \geq \ldots \geq I^{N(J)}_{//\overset{\rho}{\delta_J}}$$

por definição as «direções principais de deformação» das nuvens de pontos-perfis.

Quantas direções principais será necessário extrair para decompor completamente a inércia de uma nuvem de pontos perfis...?

F. Critérios a serem respeitados na busca de um novo referencial de representação

4. Processo de construção do novo referencial (continuação) :

Para realizar a «melhor» representação possível da inércia I_G de uma nuvem de pontos-perfis é necessário e suficiente :

❶ determinar as p direções principais de deformação da nuvem,

❷ construir as p retas que passam por G e que são colineares com essas direções principais,

❸ projetar os pontos ortogonalmente a essas retas, construindo assim os p eixos fatorais,

❹ representar sucessivamente os planos fatoriais que resultam da decomposição da inércia I_G da nuvem de pontos-perfis...

F. Critérios a serem respeitados na busca de um novo referencial de representação

5. Como se calculam as retas que constitem o novo referencial...?

■ As direções principais de uma nuvem de pontos-perfis N(I) se determinam por meio de um processo de cálculo chamado «diagonalização de uma matriz de inércia».

Este processo é um problema clássico do cálculo numérico pelo qual se determinam os «auto-valores e os auto-vetores» associados às matrizes de inércia.

■ Não levamos em conta, aqui, a forma como se realiza a diagonalização das matrizes de inércia associada às nuvens de pontos N(I) e N(J). (ver documento anexo).

Desse processo de cálculo, resultam os auto-vetores u_α de $V_{(J, J)}$ e w_a de $B_{(I, I)}$, os quais :

1) estão associados aos mesmos auto-valores $0 < \lambda_\alpha$ 1..

2) estão ligados pelas relações seguintes :

$$\underset{[J\times 1]}{u_\alpha} = \frac{1}{\sqrt{\lambda_\alpha}} \underset{[J\times I]}{K'} \times \underset{[I\times 1]}{w_\alpha} \qquad \text{y} \qquad \underset{[I\times 1]}{w_\alpha} = \frac{1}{\sqrt{\lambda_\alpha}} \underset{[I\times J]}{K} \times \underset{[J\times 1]}{u_\alpha}$$

Para determinar o novo referencial de cada um dos espaços de representação dos pontos-linha e dos pontos-coluna, se segue uma estratégia econômica de cálculo que consiste em trabalhar com a matriz de inércia correspondente ao espaço que contenha menos pontos. Os vetores próprios do outro espaço sao definidos por meio das relações anteriores.

Como se constróem os eixos fatoriais, por projeção ortogonal dos pontos sobre as direções principais de deformação das nuvens de pontos-perfis?.