

Teste Qui-quadrado de independência

Teste de Independência

Objetivo: Verificar se existe independência entre duas variáveis medidas nas mesmas unidades experimentais.

Exemplo 1: Uma grande empresa de comunicação no Brasil fez um levantamento com 1300 usuários de seus recursos midiáticos, para verificar se a preferência por um determinado canal de informação para se interar de notícias é independente do nível de instrução do indivíduo. Os resultados obtidos foram:

	Tipo de mídia				
Grau de instrução	Internet	TV	Rede Social	Outras	Total
Fundamental	10	27	5	8	50
Médio	90	73	125	162	450
Superior	200	130	220	250	800
Total	300	230	350	420	1300

Vamos calcular *proporções segundo os totais das colunas* (poderiam também ser calculadas pelos totais das linhas). Temos a seguinte tabela:

	Tipo de mídia				
Grau de instrução	Internet	TV	Rede Social	Outras	Total
Fundamental	3,33%	11,74%	1,90%	1,43%	3,85%
Médio	30,00%	31,74%	38,57%	35,71%	34,62%
Superior	66,67%	56,52%	59,52%	62,86%	61,54%
Total	100,00%	100,00%	100,00%	100,00%	100,00%

⇒ O que representam as porcentagens na colunas?

Distribuição de grau de instrução por tipo de mídia.

⇒ Independentemente da preferência por um tipo de mídia:

3,85% dos usuários têm ensino fundamental,

34,62% têm ensino médio e

61,54% têm ensino superior.

Sob independência entre grau de instrução e preferência por um tipo de mídia, o número esperado de usuários que têm:

- **Fundam. e preferem Internet** é igual a $300 \times 0,0385 = \mathbf{11,54}$ ($=300 \times 50/1300$),
- **Médio e preferem Internet** é $300 \times 0,3462 = \mathbf{103,85}$ ($=300 \times 450/1300$),
- **Superior e preferem Internet** é $300 \times 0,6154 = \mathbf{184,62}$ ($=300 \times 800/1300$).

	Tipo de mídia				
Grau de instrução	Internet	TV	Rede Social	Outras	Total
Fundamental	10 11,54 (3,85%)	27 8,85 (3,85%)	5 13,46 (3,85%)	8 16,15 (3,85%)	50 (3,85%)
Médio	90 103,85 (34,62)%	73 79,62 (34,62%)	125 121,15 (34,62%)	162 145,38 (34,62%)	450 (34,62%)
Superior	200 184,62 (61,54%)	130 141,54 (61,54%)	220 215,38 (61,54%)	250 258,46 (61,54%)	800 (61,54%)
Total	300	230	350	420	1300

As diferenças entre os valores observados e os esperados não são muito pequenas. Preferência por um tipo de mídia e grau de instrução parecem não ser *independentes*.

Teste de Independência – Metodologia

Em geral, os dados referem-se a mensurações de duas características (A e B) feitas em n unidades experimentais, que são apresentadas conforme a seguinte tabela:

$A \setminus B$	B_1	B_2	...	B_s	Total
A_1	O_{11}	O_{12}	...	O_{1s}	$O_{1.}$
A_2	O_{21}	O_{22}	...	O_{2s}	$O_{2.}$
...
A_r	O_{r1}	O_{r2}	...	O_{rs}	$O_{r.}$
Total	$O_{.1}$	$O_{.2}$...	$O_{.s}$	n

Hipóteses a serem testadas – **Teste de independência:**

H_0 : A e B são variáveis independentes

H_1 : As variáveis A e B não são independentes

→ Quantas observações devemos esperar em cada casela, se A e B forem independentes?

$$E_{ij} = \frac{O_{i.} \times O_{.j}}{n}$$

Distância entre os valores observados e os valores esperados sob a suposição de independência:

$$\chi^2 = \sum_{i=1}^s \sum_{j=1}^r \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

Supondo H_0 verdadeira,

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^s \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \sim \chi_q^2$$

aproximadamente, sendo $q = (r - 1) \times (s - 1)$ o número de **graus de liberdade**.

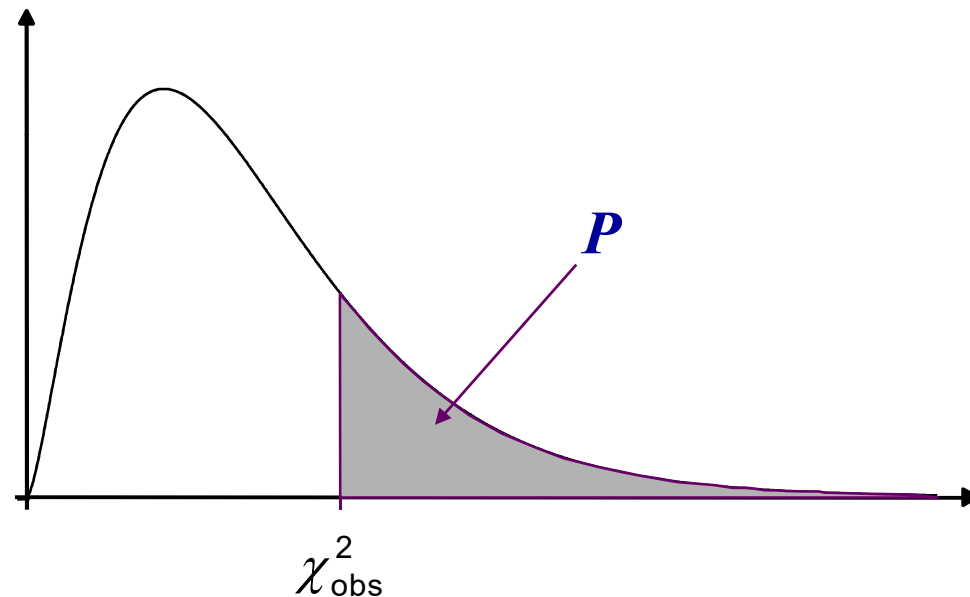
Regra de decisão:

Pode ser baseada no **valor-p** (nível descritivo), neste caso

$$P = P(\chi_q^2 \geq \chi_{obs}^2),$$

em que χ_{obs}^2 é o valor calculado, a partir dos dados, usando a expressão apresentada para χ^2 .

Graficamente:



Se, para α fixado, obtemos **valor-p** $\leq \alpha$, **rejeitamos a hipótese H_0 de independência.**

Exemplo 1 (continuação): Estudo da independência entre preferência por um tipo de mídia e grau de instrução. 1300 usuários foram entrevistados ao acaso.

Hipóteses: H_0 : As variáveis preferência por um tipo de mídia e grau de instrução são independentes.

H_1 : Existe dependência entre as variáveis.

Tabela de valores observados

	Tipo de mídia				
Grau de instrução	Internet	TV	Rede social	Outras	Total
Fundamental	10	27	5	8	50
Médio	90	73	125	162	450
Superior	200	130	220	250	800
Total	300	230	350	420	1300

→ Exemplo do cálculo dos valores esperados sob H_0 (independência):

- Número esperado de usuários que têm fundamental e preferem internet:

$$E_{11} = \frac{300 \times 50}{1300} = 11,54$$

Tabela de valores observados e esperados (entre parênteses)

	TIPO DE MÍDIA				
GRAU DE INSTRUÇÃO	Internet	TV	Rede social	Outras	Total
Fundamental	10 (11,54)	27 (8,85)	5 (13,46)	8 (16,15)	50
Médio	90 (103,85)	73 (79,62)	125 (121,15)	162 (145,38)	450
Superior	200 (184,62)	130 (141,54)	220 (215,38)	250 (258,46)	800

Médio e prefere TV:

$$E_{22} = \frac{230 \times 450}{1300} = 79,62$$

Superior e prefere outras mídias:

$$E_{34} = \frac{420 \times 800}{1300} = 258,46$$

Lembre-se:

$$E_{ij} = \frac{O_{i.} \times O_{.j}}{n_{..}}$$

Cálculo da estatística de qui-quadrado:

	Tipo de Mídia				
Grau de instrução	Internet	TV	Rede social	Outras	Total
Fundamental	10 (11,54)	27 (8,85)	5 (13,46)	8 (16,15)	50
Médio	90 (103,85)	73 (79,62)	125 (121,15)	162 (145,38)	450
Superior	200 (184,62)	130 (141,54)	220 (215,38)	250 (258,46)	800
Total	300	230	350	420	1300

$$\begin{aligned}
 \chi_{obs}^2 &= \frac{(10-11,54)^2}{11,54} + \frac{(27-8,85)^2}{8,85} + \frac{(5-13,46)^2}{13,46} + \frac{(8-16,15)^2}{16,15} \\
 &+ \frac{(90-103,85)^2}{103,85} + \frac{(73-79,62)^2}{79,62} + \frac{(125-121,15)^2}{121,15} + \frac{(162-145,38)^2}{145,38} \\
 &+ \frac{(200-184,62)^2}{184,62} + \frac{(130-141,54)^2}{141,54} + \frac{(220-215,38)^2}{215,38} + \frac{(250-258,46)^2}{258,46} \\
 &= 0,21 + 37,25 + 5,32 + 4,12 + 1,85 + 0,55 + 0,12 + 1,90 + 1,28 + 0,94 + 0,10 + 0,28 \\
 &= 53,91.
 \end{aligned}$$

Determinação do número de graus de liberdade:

- Categorias de Grau de instrução: $s = 3$
 - Categorias de Tipo de mídia: $r = 4$
- $\Rightarrow q = (r - 1) \times (s - 1) = 3 \times 2 = 6$

O nível descritivo (valor p): $P = P(\chi_6^2 \geq 53,910) < 0,0001$

Supondo $\alpha = 0,05$, temos valor-p $< \alpha$.

Assim, temos evidências para rejeitar a independência entre as variáveis grau de instrução e preferência por tipo de mídia para informação, ao nível de 5% de significância, i.é, a preferência por uma mídia depende do grau de instrução do usuário.

Análise de Correspondência Simples

A Análise de Correspondência Simples (ACS) é a análise aplicada a uma tabela de contingência de dupla entrada, isto é, com duas variáveis categorizadas, em que a variável alocada nas linhas da tabela tem r possíveis categorias e a variável alocada nas colunas, s categorias. Assim, a matriz de dados tem dimensão $r \times s$.

Exemplo 2: suponha que uma pesquisa foi realizada com 400 pessoas, em quatro países (A, B, C, D), e a cada uma delas foi perguntada a preferência por um de 3 filmes de cinema (X, Y, Z). Os dados estão na Tabela 1.

Tabela 1: Frequência da preferência por filmes

No país A todas as 120 pessoas consultadas preferiram o filme X, enquanto que no país D, dentre as 80 pessoas, as preferências foram de 40 para o filme X e 40 para o filme Z.

	X	Y	Z	Total
A	120	0	0	120
B	20	10	70	100
C	0	50	50	100
D	40	0	40	80
Total	180	60	160	400

Como os números de pessoas consultadas em cada país não são iguais, devemos calcular as frequências relativas dadas em proporções ou porcentagens para podermos fazer comparações entre os países. A Tabela 2 mostra as proporções correspondentes à Tabela 1.

Tabela 2: Proporções a preferência por filmes

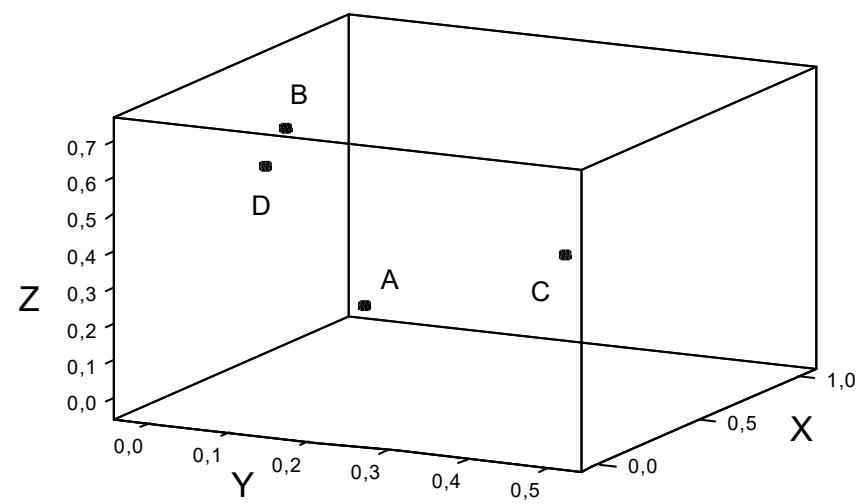
	X	Y	Z	Total	Massa
A	1,0	0,0	0,0	1,0	0,30
B	0,2	0,1	0,7	1,0	0,25
C	0,0	0,5	0,5	1,0	0,25
D	0,5	0,0	0,5	1,0	0,20
Total	0,45	0,15	0,40	1,0	1,00

Por essa tabela tem-se que 20% das pessoas consultadas no país B preferem o filme X enquanto que 50% das pessoas do país C preferem o filme Y. As linhas da Tabela 2 são chamadas perfis linha (perfis dos países), isto é, o perfil do país A é $(1,0 \quad 0,0 \quad 0,0)$. Notamos que os países têm perfis diferentes. O perfil correspondente aos totais das colunas da Tabela 1 é chamado perfil linha médio e no caso do Exemplo 2 é igual a $(0,45 \quad 0,15 \quad 0,40)$. Isto significa que, de todas as 400 pessoas consultadas, 45% preferiram o filme X, 15% o filme Y e 40% o filme Z.

A Tabela 2 foi calculada tomando como base os totais das linhas da Tabela 1, uma vez que as amostras foram tomadas em cada país. Outra possibilidade é calcular as frequências relativas tomando como base os totais das colunas e nesse caso teríamos os perfis coluna (perfis dos filmes) definidos analogamente aos perfis linha. A Análise de Correspondência pode ser executada tanto usando os perfis linha como os perfis coluna.

Os dados da Tabela 2 podem ser representados em um espaço de 3 dimensões, isto é, cada linha da tabela tem as coordenadas nos 3 eixos correspondentes aos 3 filmes. A representação está apresentada na Figura 1.

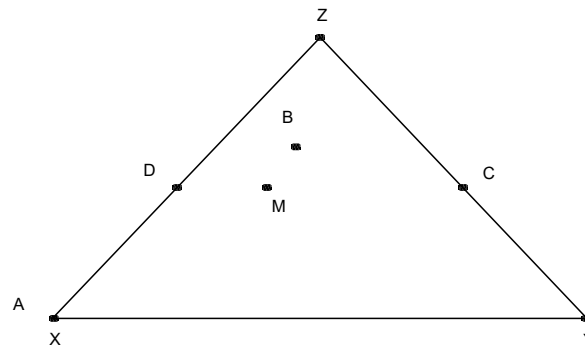
**Figura 1: Representação dos pontos da Tabela 2
no diagrama de dispersão**



Todos os pontos da tabela estão dentro da área do cubo definida pelas coordenadas $(1, 0, 0)$, $(0, 1, 0)$ e $(0, 0, 1)$, limitada pelo fato de que a soma das coordenadas é igual a 1. Na verdade a representação pode ser feita de forma exata em um espaço de dimensão 2, ou seja, no triângulo definido pelos pontos $(1, 0, 0)$, $(0, 1, 0)$ e $(0, 0, 1)$. Esse triângulo é denominado simplex regular de 3 vértices.

A representação no simplex é dada na Figura 2.

Figura 2: Representação dos pontos da Tabela 2 no simplex



No simplex, os vértices representam as categorias de uma variável (filmes) – são as coordenadas padrão. Os pontos representados no simplex são as categorias da outra variável (países) – são as coordenadas principais. Essa representação será vista adiante nos mapas assimétricos da Análise de Correspondência. O perfil tende a estar mais próximo do vértice (filme) para o qual ele tem o maior valor.

Uma tabela de contingência com r linhas e s colunas pode ser exatamente representada em um plano com dimensão igual ao mínimo entre $r-1$ e $s-1$.

A massa de uma linha é a proporção de observações nessa linha com relação ao número total de observações. A última coluna da Tabela 2 exibe as massas das linhas no Exemplo 2.

A medida qui-quadrado (χ^2) é bastante conhecida e utilizada em testes de associação entre variáveis categorizadas.

A inércia é definida como

$$I_n = \frac{\chi^2}{n}$$

A Análise de Correspondência é uma representação da tabela de contingência em algum espaço de dimensão menor do que a da representação exata, analogamente à Análise de Componentes Principais. A diferença é que na Análise de Componentes Principais a distância utilizada na representação é a distância euclidiana enquanto que na Análise de Correspondência é a distância qui-quadrado.

Exemplo 3: considere que uma pesquisa foi feita com 1500 pessoas e a cada uma delas foi perguntada a idade, distribuída em 4 faixas etárias, de 18 a 25 anos, de 26 a 40 anos, de 41 a 60 anos, e acima de 60 anos, e a preferência por um de 5 produtos (A, B, C, D, E). A tabela de contingência está exibida na Tabela 3.

Tabela 3: Frequência da preferência por produtos

Produto	18 - 25	26 - 40	41 - 60	+ 60	Total
A	110	122	68	24	324
B	102	110	66	35	313
C	98	98	60	48	304
D	61	82	82	87	312
E	31	42	82	92	247
Total	402	454	358	286	1500

Tabela 4: Perfis de linhas da preferência por produtos

Produto	18 - 25	26 - 40	41 - 60	+ 60	Total	Massa
A	0,340	0,376	0,210	0,074	1	0,216
B	0,326	0,351	0,211	0,112	1	0,209
C	0,322	0,322	0,198	0,158	1	0,202
D	0,195	0,263	0,263	0,279	1	0,208
E	0,126	0,170	0,332	0,372	1	0,165
Total	0,268	0,303	0,239	0,190	1	

Tabela 5: Perfis de colunas da preferência por produtos

Produto	18 - 25	26 - 40	41 - 60	+ 60	Total
A	0,273	0,269	0,190	0,084	0,216
B	0,254	0,242	0,184	0,122	0,209
C	0,244	0,216	0,168	0,168	0,202
D	0,152	0,181	0,229	0,304	0,208
E	0,077	0,092	0,229	0,322	0,165
Total	1	1	1	1	1
Massa	0,268	0,303	0,239	0,190	

Figura 3: Mapa assimétrico dos produtos no espaço das faixas etárias

Faixas etárias - coordenadas padrão
Produtos - coordenadas principais

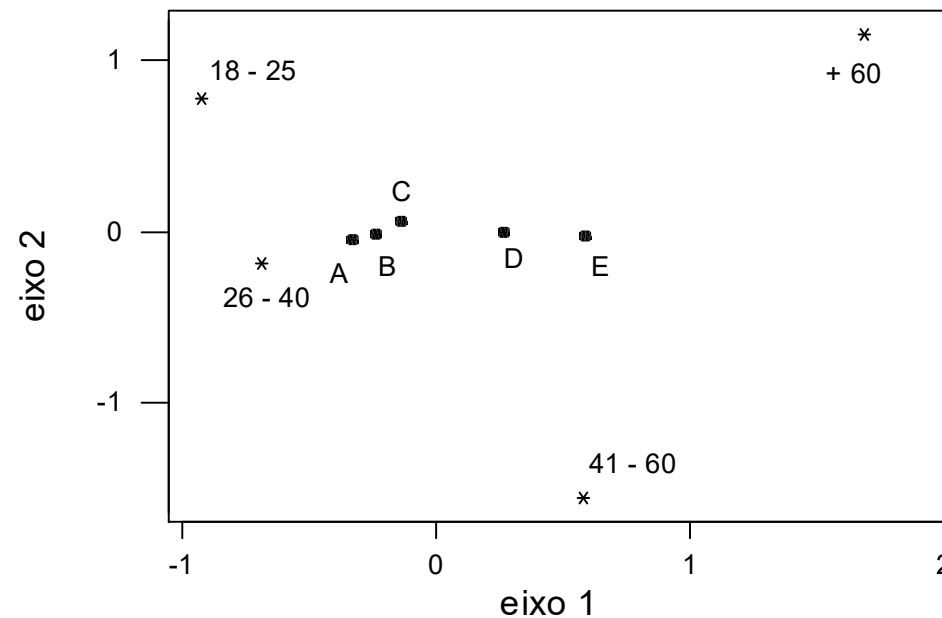


Figura 4: Mapa assimétrico das faixas etárias no espaço dos produtos

Produtos - coordenadas padrão

Faixas etárias - coordenadas principais

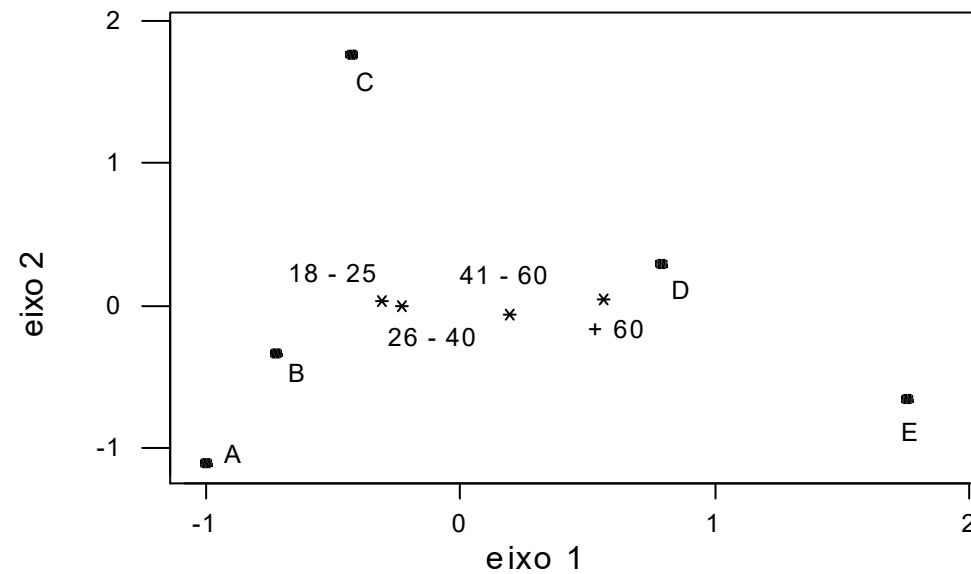
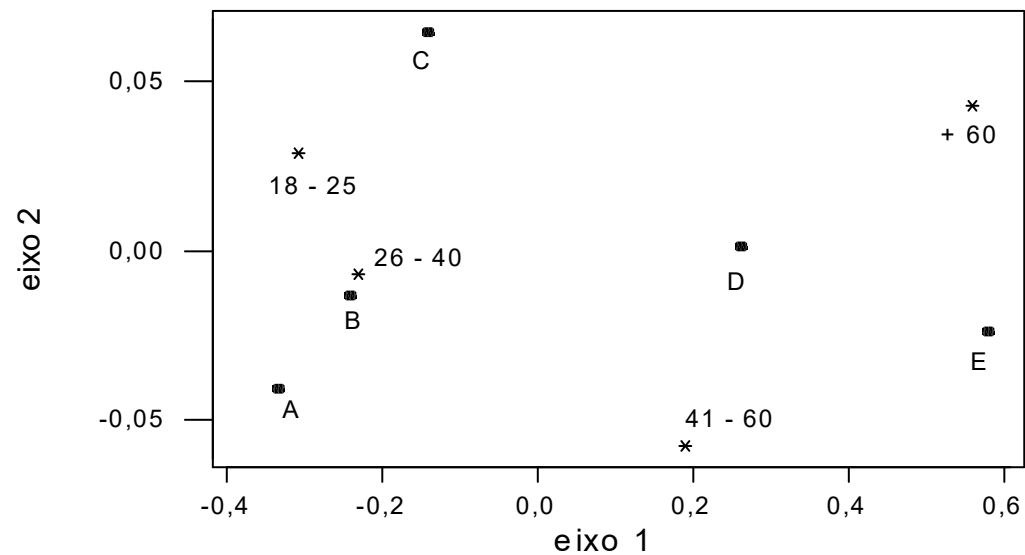


Figura 5: Mapa simétrico de produtos e faixas etárias

Faixas etárias e Produtos - coordenadas principais

Distâncias linha-linha e coluna-coluna \approx distância qui-quadrado



Alguns comentários sobre estes gráficos são importantes. Nos mapas assimétricos linhas e colunas estão exibidos no mesmo espaço euclidiano. Nos simétricos, as representações de linhas e colunas são sobrepostos em dois espaços diferentes e a interpretação das distâncias entre linhas e colunas deve ser feita com cautela. Nos assimétricos é possível interpretar essas distâncias e por isso é importante que os eixos tenham a mesma escala. Nos simétricos, interpretamos somente as direções.

Nos mapas assimétricos os vértices são empregados para dar nomes aos eixos em que os pontos são interpretados. A distância entre os pontos são aproximações da distância qui-quadrado, a proximidade entre os pontos nos dá ideia da similaridade entre linhas, entre colunas e da associação entre linhas e colunas.

Observando-se a projeção dos pontos no primeiro eixo da Figura 5, pode-se verificar que os produtos A, B e C estão mais associados com os mais jovens e os produtos D e E com os mais idosos. Se tivéssemos que escolher produtos para relacionar com as faixas de idade a projeção no eixo 1 nos levaria a escolher o produto A para a faixa de 18 a 25 anos, os produtos B e C para a faixa de 26 a 40 anos, o produto D para a faixa de 41 a 60 anos e o produto E para as pessoas com mais de 60 anos. Outra possibilidade seria escolher os produtos A, B e C para as pessoas com até 40 anos e os produtos D e E para pessoas acima dessa idade. Neste exemplo, o número de categorias de ambas as variáveis é pequeno, e essas associações podem ser visualizadas diretamente das tabelas de proporções. Em casos com variáveis com número maior de categorias já não é tão simples entender as associações existentes e a Análise de Correspondência é uma ferramenta muito útil.

A Tabela 6 mostra a decomposição da inércia por linhas e eixos do Exemplo 3, enquanto que a Tabela 7 mostra o mesmo para colunas.

Tabela 6: Decomposição da inércia por linhas da preferência por produtos

Produto	Eixo 1	Eixo 2	Eixo 3	Total
A	0,0239	0,0004	0,0000	0,0243
B	0,0120	0,0000	0,0000	0,0120
C	0,0040	0,0009	0,0001	0,0050
D	0,0143	0,0000	0,0004	0,0147
E	0,0559	0,0001	0,0001	0,0561
Total	0,1101	0,0014	0,0006	0,1121

Tabela 7: Decomposição da inércia por colunas da preferência por produtos

Faixa etária	Eixo 1	Eixo 2	Eixo 3	Total
18 - 25	0,0255	0,0002	0,0002	0,0259
26 – 40	0,0161	0,0000	0,0003	0,0164
41 – 60	0,0089	0,0008	0,0001	0,0098
+ 60	0,0596	0,0004	0,0000	0,0600
Total	0,1101	0,0014	0,0006	0,1121

A Tabela 8 mostra as inércias principais dos 3 eixos e a porcentagem explicada por cada um deles no Exemplo 3. O primeiro eixo explica 98,2% da inércia total e os 2 primeiros eixos, ou seja, a representação em 2 dimensões, 99,5%.

Tabela 8: Inércias principais da preferência por produtos

Eixo	Inércia principal	% explicada
1	0,1101	98,22
2	0,0014	1,25
3	0,0006	0,53
Total	0,1121	100

A contribuição das linhas e colunas para as inércias principais no Exemplo 3 é dada nas Tabelas 9 e 10, respectivamente.

Tabela 9: Inércias principais explicadas pelas linhas da preferência por produtos

Produto	Eixo 1	Eixo 2	Eixo 3	Total
A	21,71	28,57	0,00	21,68
B	10,90	0,00	0,00	10,71
C	3,63	64,29	16,67	4,46
D	12,99	0,00	66,66	13,11
E	50,77	7,14	16,67	50,04
Total	100	100	100	100

Tabela 10: Inércias principais explicadas pelas colunas da preferência por produtos

Faixa etária	Eixo 1	Eixo 2	Eixo 3	Total
18 - 25	23,16	14,29	33,33	23,11
26 - 40	14,62	0,00	50,00	14,63
41 - 60	8,09	57,14	16,67	8,74
+ 60	54,13	28,57	0,00	53,52
Total	100	100	100	100

A correlação ao quadrado das linhas com eixos no Exemplo 3, é dada na Tabela 11 e das colunas na Tabela 12.

Tabela 11: Contribuição dos eixos principais para linhas da preferência por produtos

Produto	Eixo 1	Eixo 2	Eixo 3	Total
A	0,9835	0,0165	0,0000	1
B	1,0000	0,0000	0,0000	1
C	0,8000	0,1800	0,0200	1
D	0,9728	0,0000	0,0272	1
E	0,9964	0,0018	0,0018	1

Tabela 12: Contribuição dos eixos principais para colunas da preferência por produtos

Faixa etária	Eixo 1	Eixo 2	Eixo 3	Total
18 - 25	0,9846	0,0077	0,0077	1
26 – 40	0,9817	0,0000	0,0183	1
41 – 60	0,9082	0,0816	0,0102	1
+ 60	0,9933	0,0067	0,0000	1

A qualidade de cada linha (coluna) na representação de 2 dimensões é dada pela soma das contribuições na Tabela 11 (Tabela 12) das parcelas correspondentes aos 2 primeiros eixos. Essas qualidades, para o Exemplo 3, são apresentadas nas Tabelas 13 e 14, respectivamente, para linhas e colunas.

Tabela 13: Qualidade das linhas em 2 dimensões da preferência por produtos

Produto	Qualidade
A	1,0000
B	1,0000
C	0,9800
D	0,9728
E	0,9982

Tabela 14: Qualidade das colunas em 2 dimensões da preferência por produtos

Faixa etária	Qualidade
18 - 25	0,9923
26 – 40	0,9817
41 – 60	0,9898
+ 60	1,0000

As correlações entre linhas (colunas) e eixos são a raiz quadrada dos valores da Tabela 11 (Tabela 12), com o sinal da coordenada correspondente e são apresentadas nas Tabelas 15 e 16.

Tabela 15: Correlações das linhas e eixos da preferência por produtos

Produto	Eixo 1	Eixo 2	Eixo 3
A	-0,992	-0,128	0,000
B	-1,000	-0,000	0,000
C	-0,894	0,424	0,141
D	0,986	0,000	-0,165
E	0,998	-0,042	0,042

Tabela 16: Correlações de colunas e eixos da preferência por produtos

Faixa etária	Eixo 1	Eixo 2	Eixo 3
18 - 25	-0,992	0,088	-0,088
26 – 40	-0,991	-0,000	0,135
41 – 60	0,953	-0,286	-0,101
+ 60	0,997	0,082	0,000

Exemplo: Saúde

Foi perguntado a um grupo de pessoas:

Como você considera sua saúde?

Categorias de respostas:

MB = muito boa

BO = boa

RE = regular

RU = ruim

MR = muito ruim

Saúde *versus* Idade

	MB	BO	RE	RU	MR	Total
16-24	243	789	167	18	6	1.223
25-34	220	809	164	35	6	1.234
35-44	147	658	181	41	8	1.035
45-54	90	469	236	50	16	861
55-64	53	414	306	106	30	909
65-74	44	267	284	98	20	713
75+	20	136	157	66	17	396
Total	817	3.542	1.495	414	103	6.371

Saúde *versus* Idade

Principal inertias (eigenvalues):

dim	value	%	cum%	scree plot
1	0.136603	97.3	97.3	*****
2	0.002090	1.5	98.7	
3	0.001292	0.9	99.7	
4	0.000474	0.3	100.0	

Total:	0.140458	100.0		

Saúde *versus* Idade

Rows :

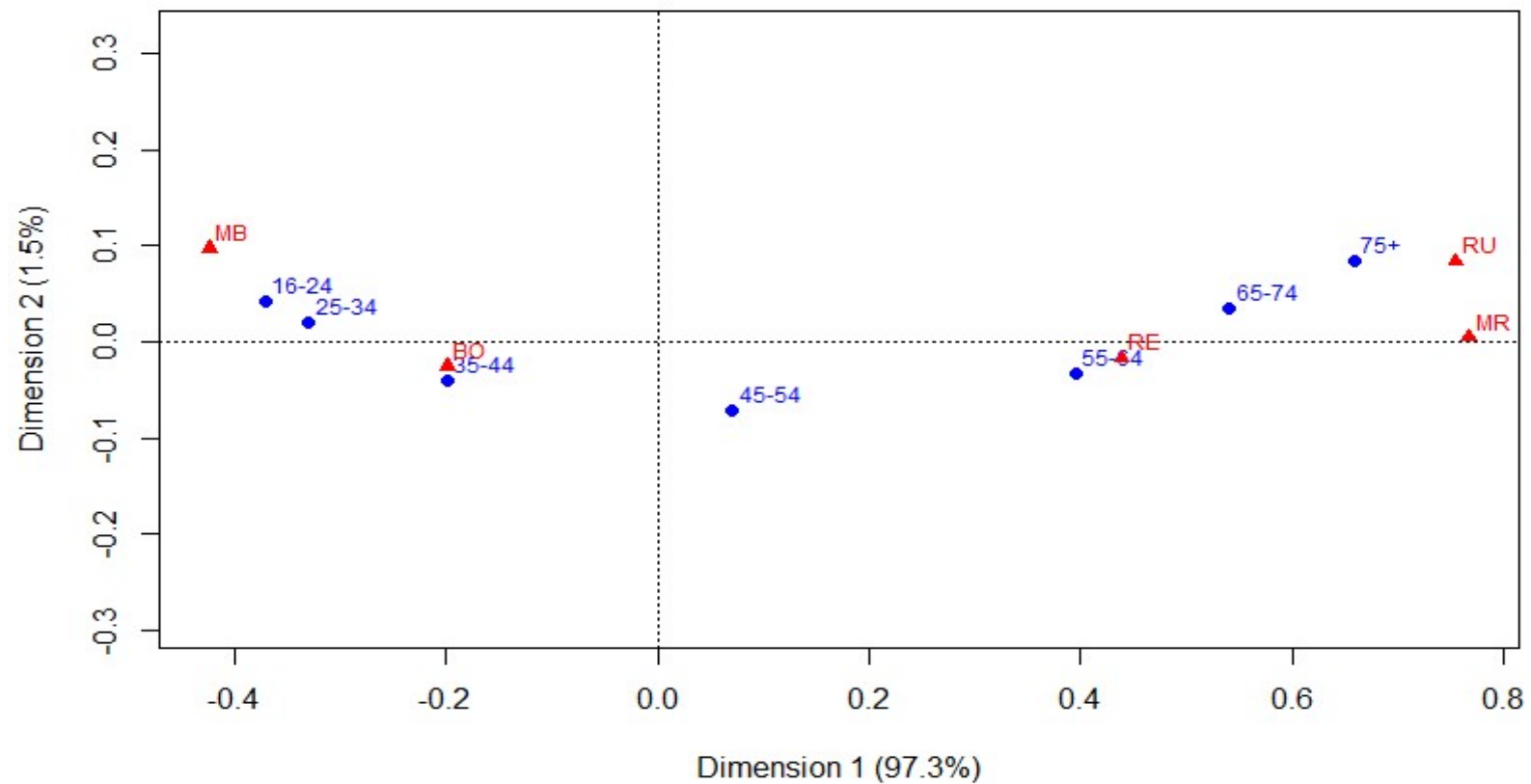
	name	mass	qlt	inr	k=1	cor	ctr	k=2	cor	ctr
1	16-24	192	991	192	-371	978	193	42	13	164
2	25-34	194	992	152	-330	989	154	20	3	35
3	35-44	162	971	49	-199	932	47	-41	39	129
4	45-54	135	815	12	71	408	5	-71	407	325
5	55-64	143	990	162	396	983	163	-33	7	75
6	65-74	112	987	237	541	983	239	34	4	63
7	75+	62	994	196	658	978	197	84	16	208

Columns :

	name	mass	qlt	inr	k=1	cor	ctr	k=2	cor	ctr
1	MB	128	994	173	-423	944	168	97	50	579
2	BO	556	994	159	-198	978	160	-25	16	171
3	RE	235	989	326	439	987	331	-17	1	31
4	RU	65	988	270	755	976	271	84	12	219
5	MR	16	949	71	767	949	70	5	0	0

Saúde *versus* Idade

Mais jovens - boa saúde, mais velhos - saúde ruim



Saúde *versus* Sexo

	MB	BO	RE	RU	MR	Total
M	448	1.789	636	177	39	3.089
F	369	1.753	859	237	64	3.282
Total	817	3.542	1.495	414	103	6.371

Saúde *versus* Idade e Sexo - T.combinadas

	MB	BO	RE	RU	MR	Total
M16-24	145	402	84	5	3	639
M25-34	112	414	74	13	2	615
M35-44	80	331	82	24	4	521
M45-54	54	231	102	22	6	415
M55-64	30	219	119	53	12	433
M65-74	18	125	110	35	4	292
M75+	9	67	65	25	8	174
Total	448	1.789	636	177	39	3.089

Saúde *versus* Idade e Sexo - T.combinadas

	MB	BO	RE	RU	MR	Total
F16-24	98	387	83	13	3	584
F25-34	108	395	90	22	4	619
F35-44	67	327	99	17	4	514
F45-54	36	238	134	28	10	446
F55-64	23	195	187	53	18	476
F65-74	26	142	174	63	16	421
F75+	11	69	92	41	9	222
Total	369	1.753	859	237	64	3.282

Saúde *versus* Idade e sexo

Tabelas combinadas

Principal inertias (eigenvalues):

dim	value	%	cum%	scree plot
1	0.141665	94.5	94.5	*****
2	0.003924	2.6	97.1	*
3	0.003123	2.1	99.2	*
4	0.001264	0.8	100.0	

Total: 0.149976 100.0

Columns:

	name	mass	qlt	inr	k=1	cor	ctr	k=2	cor	ctr
1	MB	128	985	177	-432	900	169	-133	85	577
2	BO	556	985	158	-202	957	161	35	28	169
3	RE	235	977	327	451	974	337	25	3	38
4	RU	65	965	262	755	944	262	-111	20	202
5	MR	16	901	76	796	896	72	-56	4	13

Saúde *versus* Idade e sexo

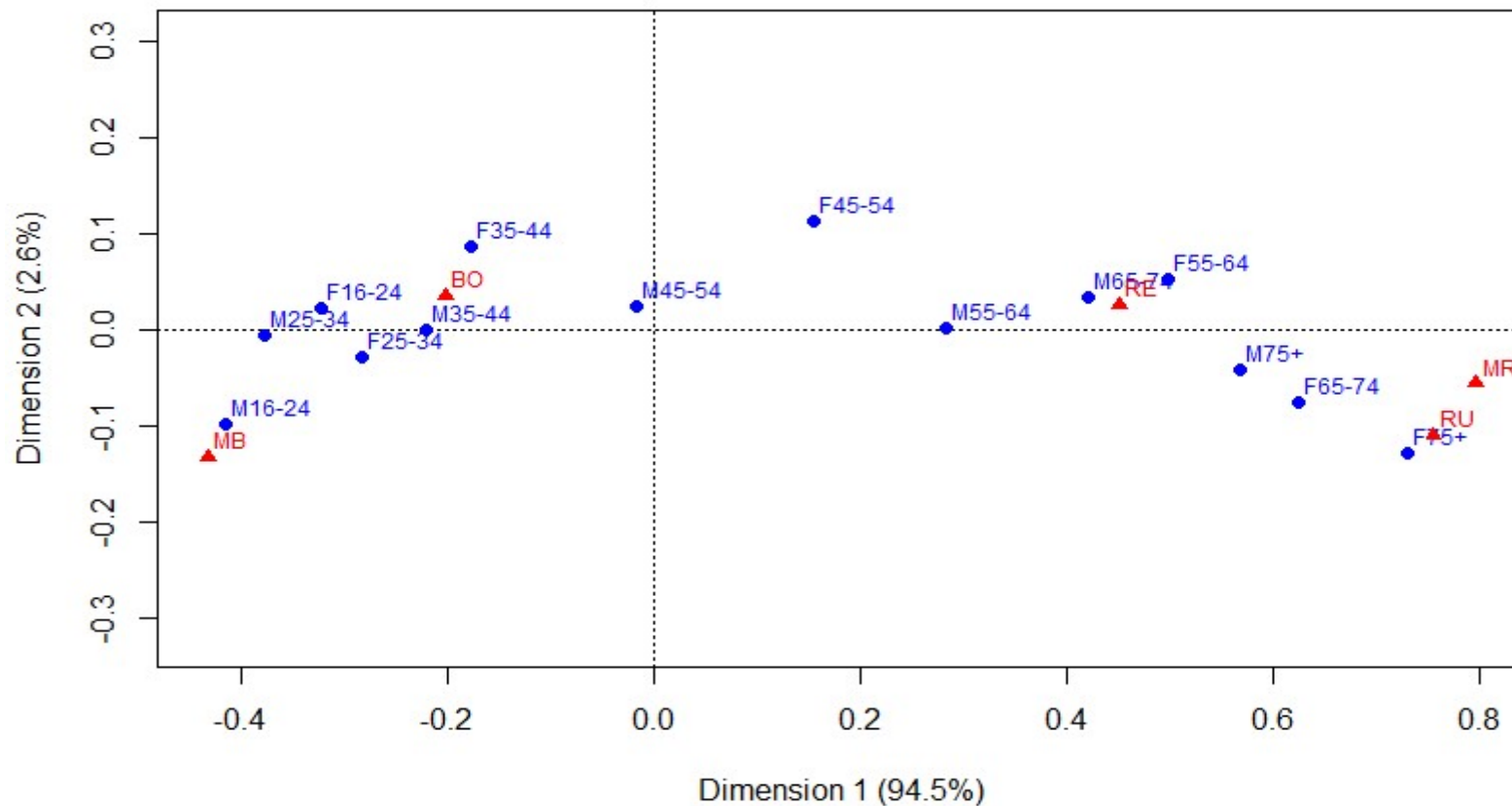
Tabelas combinadas

Rows :

	name	mass	qlt	inr	k=1	cor	ctr	k=2	cor	ctr
1	M1624	100	962	127	-415	910	122	-100	52	253
2	M2534	97	995	92	-377	994	97	-7	0	1
3	M3544	82	937	28	-220	937	28	0	0	0
4	M4554	65	284	1	-17	96	0	24	188	9
5	M5564	68	818	44	282	818	38	2	0	0
6	M6574	46	903	60	420	897	57	34	6	13
7	M75+	27	977	60	567	972	62	-41	5	12
8	F1624	92	998	64	-323	993	67	23	5	12
9	F2534	97	992	53	-284	982	55	-28	10	20
10	F3544	81	998	21	-178	806	18	87	192	154
11	F4554	70	928	18	154	602	12	113	326	229
12	F5564	75	978	128	499	967	131	52	10	51
13	F6574	66	997	175	624	982	182	-76	15	98
14	F75+	35	994	128	729	964	131	-129	30	147

Saúde *versus* Idade e Sexo - T. combinadas

Mais jovens - boa saúde, mais velhos - saúde ruim
Masculino sempre à esquerda do Feminino



Saúde *versus* Idade e Sexo

	MB	BO	RE	RU	MR	Total
16-24	243	789	167	18	6	1.223
25-34	220	809	164	35	6	1.234
35-44	147	658	181	41	8	1.035
45-54	90	469	236	50	16	861
55-64	53	414	306	106	30	909
65-74	44	267	284	98	20	713
75+	20	136	157	66	17	396
M	448	1.789	636	177	39	3.089
F	369	1.753	859	237	64	3.282

Saúde *versus* Idade e sexo

Tabelas empilhadas

Principal inertias (eigenvalues):

dim	value	%	cum%	scree plot
1	0.072031	97.1	97.1	*****
2	0.001215	1.6	98.8	
3	0.000667	0.9	99.7	
4	0.000259	0.3	100.0	

Total:	0.074171	100.0		

Inércia total Idade = 0,140458

Inércia total Sexo = 0,007884

Média = 0,074171

Saúde *versus* Idade e sexo

Tabelas empilhadas

Rows :

	name	mass	qlt	inr	k=1	cor	ctr	k=2	cor	ctr
1	16-24	96	987	182	-371	979	184	-32	7	82
2	25-34	97	995	144	-330	991	147	-20	3	31
3	35-44	81	963	47	-199	930	45	38	34	95
4	45-54	68	889	11	72	420	5	76	469	322
5	55-64	71	988	153	396	984	155	25	4	37
6	65-74	56	986	224	540	981	227	-37	5	62
7	75+	31	995	186	657	975	186	-95	20	230
8	M	242	992	27	-89	949	27	-19	43	73
9	F	258	992	26	84	949	25	18	43	68

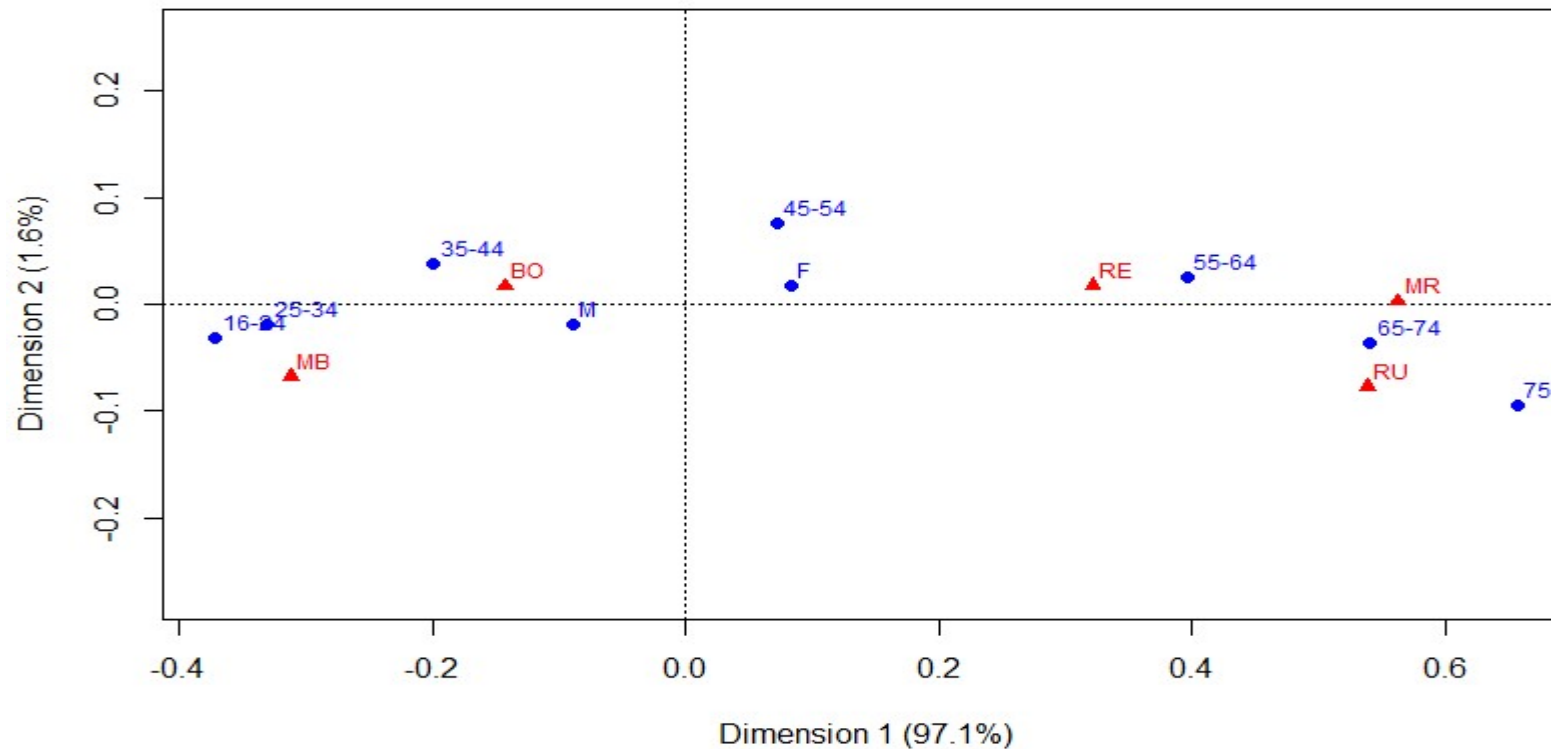
Columns :

	name	mass	qlt	inr	k=1	cor	ctr	k=2	cor	ctr
1	MB	128	990	178	-312	946	173	-68	45	486
2	BO	556	992	157	-143	978	158	17	14	135
3	RE	235	990	331	321	987	337	18	3	60
4	RU	65	990	262	538	970	261	-77	20	320
5	MR	16	951	73	563	951	71	3	0	0

Saúde *versus* Idade e Sexo - T. empilhadas

Mais jovens e Masculino - boa saúde

Mais velhos e Feminino - saúde ruim



ACM: questões com mesmas respostas

Tabela indicadora

			Q1				Q2				Q3			
Q1	Q2	Q3	R1	R2	R3	R4	R1	R2	R3	R4	R1	R2	R3	R4
1	3	2	1	0	0	0	0	0	1	0	0	1	0	0
2	3	3	0	1	0	0	0	0	1	0	0	0	1	0
4	3	2	0	0	0	1	0	0	1	0	0	1	0	0
4	4	4	0	0	0	1	0	0	0	1	0	0	0	1
3	2	2	0	0	1	0	0	1	0	0	0	1	0	0

ACM: questões com mesmas respostas

Tabela de Burt

	R1	R2	R3	R4	R1	R2	R3	R4	R1	R2	R3	R4
R1	2501	0	0	0	172	1107	1131	91	355	1710	345	91
R2	0	476	0	0	7	129	335	5	16	261	181	18
R3	0	0	79	0	1	6	72	0	1	17	61	0
R4	0	0	0	362	1	57	108	196	7	96	55	204
R1	172	7	1	1	181	0	0	0	127	48	4	2
R2	1107	129	6	57	0	1299	0	0	219	997	61	22
R3	1131	335	72	108	0	0	1646	0	24	989	573	60
R4	91	5	0	196	0	0	0	292	9	50	4	229
R1	355	16	1	7	127	219	24	9	379	0	0	0
R2	1710	261	17	96	48	997	989	50	0	2084	0	0
R3	345	181	61	55	4	61	573	4	0	0	642	0
R4	91	18	0	204	2	22	60	229	0	0	0	313

ACM: questões com mesmas respostas

Z = Tabela indicadora

B = Tabela de Burt

$$B = Z^T Z$$

Q = número de questões

J = $\sum J_q$ = número total de categorias

$$Inercia(Z) = \frac{1}{Q} \sum_q Inercia(Z_q) = \frac{1}{Q} \sum_q (J_q - 1) = \frac{J - Q}{Q}$$

inércias principais (Burt) = quadrado (inércias principais(indicadora))

Como as inércias são menores do que 1, as de Burt são menores do que as da matriz indicadora.