



## RAD5053 – Análise de Dados Multivariados I

Prof. Dr. Evandro Marcos Saidel Ribeiro      Lista 04 – Escalonamento Multidimensional

Exercícios do CAPÍTULO 7 do livro “Análise de Dados Multivariados; Lattin, Carroll e Green”

### 7.2 Arquivo 02\_SIMILARITY.XLSX

Responda a questão seguinte usando os dados da Tabela 7.12 (disponível no arquivo *SIMILARITY*). Os dados representam julgamentos de similaridade de um único indivíduo para todos os pares possíveis de 13 itens (representados na tabela por letras do alfabeto). Os julgamentos são classificados de 1 (menos similar) a 78 (mais similar).

Realize uma análise MDS dos dados de similaridade na tabela. Em sua opinião, o espaço perceptual de duas dimensões é suficiente para refletir adequadamente os julgamentos de similaridade observados? O mapa de MDS revela alguma estrutura especial para esses dados?

Tabela 7.12 Matriz de similaridade para o problema 7.2

	A	B	C	D	E	F	G	H	I	J	K	L	M
A													
B	53												
C	78	27											
D	47	40	44										
E	5	70	67	4									
F	42	28	39	31	50								
G	35	61	29	34	75	30							
H	8	64	22	48	14	58	55						
I	32	10	23	57	2	11	38	1					
J	21	13	3	65	52	6	18	41	12				
K	16	19	60	33	51	17	77	37	68	74			
L	72	36	54	45	59	66	43	49	76	9	71		
M	20	8	25	15	56	69	7	46	73	63	62	26	

### 7.3 Arquivo 03\_NATIONS.XLSX

Kruskal e Wish (1978) coletaram dados sobre similaridades percebidas entre 12 nações: Brasil, Congo, Cuba, Egito, França, Índia, Israel, Japão, China (continental), Rússia, Estados Unidos e Iugoslávia. As classificações médias de similaridades (quanto maior o número, mais similares as duas nações) são apresentadas na Tabela 7.13 e estão disponíveis no arquivo *NATIONS*.

**Tabela 7.13** Similaridades das nações de Kruskal e Wish (1978)

	Brasil	Congo	Cuba	Egito	França	Índia	Israel	Japão	China	Rússia	EUA	Iugoslávia
Brasil												
Congo	4,83											
Cuba	5,28	4,56										
Egito	3,44	5,00	5,17									
França	4,72	4,00	4,11	4,78								
Índia	4,50	4,83	4,00	5,83	3,44							
Israel	3,83	3,33	3,61	4,67	4,00	4,11						
Japão	3,50	3,39	2,94	3,83	4,22	4,50	4,83					
China	2,39	4,00	5,50	4,39	3,67	4,11	3,00	4,17				
Rússia	3,06	3,39	5,44	4,39	5,06	4,50	4,17	4,61	5,72			
EUA	5,39	2,39	3,17	3,33	5,94	4,28	5,94	6,06	2,56	5,00		
Iugoslávia	3,17	3,50	5,11	4,28	4,72	4,00	4,44	4,28	5,06	6,67	3,56	

Usando uma metodologia de escalonamento apropriada, analise os dados da Tabela 7.13 e responda às seguintes questões:

- Quantas dimensões você usaria para o seu mapa? Justifique sua escolha.
- Como você descreveria a qualidade de ajuste do seu resultado?
- Com base naquilo que conhece sobre as nações envolvidas nesse estudo, como você interpretaria cada dimensão em seu mapa?

#### 7.4 Arquivo 04\_CONSUMPTION.XLSX

Hoffman e Franke (1986) coletaram dados de homens e mulheres, estudantes de MBA, solicitando que cada um deles indicasse a frequência com que ele/ela adquiriu e consumiu oito diferentes marcas de refrigerantes. Esses dados foram codificados para indicar consumo frequente (1) ou consumo não frequente (0). Hoffman e Franke registraram os dados de 34 estudantes; os dados estão disponíveis no arquivo *CONSUMPTION*.

- Como você usaria esses dados para obter um mapa que representasse a similaridade relativa das oito marcas de refrigerantes? Qual método de escalonamento você utilizaria?
- Como você usaria esses dados para obter um mapa que representasse a similaridade relativa dos 34 estudantes? Qual método de escalonamento você utilizaria?

#### 7.5 Arquivo 05\_PSYCH\_60.XLSX e 05\_PSYCH\_64.XLSX

Neste capítulo, apresentamos uma análise de escalonamento multidimensional de oito revistas de psicologia usando referências cruzadas de 1964 como medidas de similaridade (esses dados estão disponíveis no arquivo *PSYCH\_64*). Dados adicionais sobre referência cruzada foram também coletados para o volume do ano 1960. Os dados são apresentados na Tabela 7.14 (cada elemento representa o número de vezes em que a revista na linha referenciou a revista na coluna) e estão disponíveis no arquivo *PSYCH\_60*.

**Tabela 7.14** Atividade de referência cruzada para oito revistas de psicologia: 1960

Referência na	Referência a							
	AJP	JASP	JAP	JCPP	JCP	JEDP	JEXP	PKA
AJP	122	4	1	23	4	2	135	1
JASP	23	303	9	11	49	4	55	7
JAP	0	28	84	2	11	6	15	13
JCPP	36	10	4	304	0	0	98	4
JCP	6	93	11	1	186	6	7	14
JEDP	6	12	11	1	7	34	24	14
JEXP	65	15	3	33	3	3	337	14
PKA	2	0	2	0	0	1	6	59

- Proponha e calcule uma medida simétrica apropriada de similaridade entre as revistas.
- Utilizando sua medida proposta de similaridade, represente as similaridades entre as revistas em espaço euclidiano bidimensional. Interprete os resultados de sua análise.
- Como você poderia determinar se a estrutura de similaridade das revistas mudou entre 1960 e 1964?

## 7.6 Arquivos 06\_BASKET.XLSX

Um pesquisador de mercado possui dados sobre o comportamento de compra de várias centenas de famílias em um período de dois anos. Os dados captam todas as idas às compras feitas pela família, o montante gasto na loja durante a visita e um indicador de aquisições por categoria de produto das diversas categorias listadas na Tabela 7.15.

**Tabela 7.15** Categorias de produtos de supermercado

Lista de categorias	Nome da variável
1. Bacon	<i>BACON</i>
2. Manteiga	<i>MANTEIGA</i>
3. Pasta dental	<i>DENTE</i>
4. Detergente para lavar roupa	<i>DETER</i>
5. Ovos	<i>OVOS</i>
6. Papel higiênico	<i>PAPEL</i>
7. Sabão em pedra	<i>SABÃO</i>
8. Analgésicos	<i>ANALG</i>
9. Toalhas de papel	<i>TOALHA</i>
10. Açúcar	<i>AÇÚCAR</i>
11. Hot dogs	<i>HTDOG</i>
12. Amaciante de roupa	<i>AMAC</i>

Para entender melhor os padrões de comportamento de compra entre as categorias, o pesquisador calculou uma matriz de “cesta de coincidências”. Cada célula  $(j, k)$  da matriz reflete o número total de vezes (em todas as visitas à loja feitas por todas as famílias) em que as categorias de produtos  $j$  e  $k$  aparecem na mesma cesta de compras; cada elemento da diagonal  $(j, j)$  reflete o número total de idas às compras em que a categoria de produto  $j$  foi adquirida. A matriz da cesta de coincidências é apresentada na Tabela 7.16 (e está disponível no arquivo *BASKET*).

O pesquisador de mercado está interessado em explorar a estrutura entre as categorias. Existe um padrão? Algumas categorias tendem a ocorrer simultaneamente na cesta mais do que outras? Qual seria um bom modo de representar essa estrutura?

Realize uma análise de escalonamento multidimensional dos dados da Tabela 7.16 e responda as seguintes questões:

- Os padrões de coincidência podem ser representados por um número pequeno de dimensões? Justifique sua resposta.
- Baseado no que conhece sobre as diferentes categorias, como você descreveria essas dimensões?
- Quais categorias parecem ser mais prováveis de serem adquiridas conjuntamente em qualquer ida às compras? Há categorias que são especialmente improváveis de serem adquiridas juntas? O que você poderia dizer sobre essa análise?

**Tabela 7.16** Matriz de coincidência de categoria para 12 categorias de produtos de supermercado

	BACON	MANT.	DENTE	DETER	OVOS	PAPEL	SABÃO	ANALG	TOALHA	AÇÚCAR	HOTDOG	AMAC
BACON	4092	1190	337	770	1361	1267	580	220	1090	580	1215	520
MANT.	1190	10449	758	1763	3063	3132	1274	498	2771	1404	1918	1229
DENTE	337	758	2853	771	846	1201	831	196	993	386	640	574
DETER	770	1763	771	6789	1653	2556	1256	403	2330	819	1255	1703
OVOS	1361	3063	846	1653	9844	2999	1263	461	2546	1370	2113	1129
PAPEL	1267	3132	1201	2556	2999	13274	1911	658	4531	1422	2304	1628
SABÃO	580	1274	831	1256	1263	1911	4510	347	1624	604	1007	802
ANALG	220	498	196	403	461	658	347	1647	607	221	367	275
TOALHA	1090	2771	993	2330	2546	4531	1624	607	11219	1178	1970	1490
AÇÚCAR	580	1404	386	819	1370	1422	604	221	1178	4224	889	535
HOT DOGS	1215	1918	640	1255	2113	2304	1007	367	1970	889	7552	908
AMAC	520	1229	574	1703	1129	1628	802	275	1490	535	908	4326

### 7.7 Arquivo 07\_STORE\_SWITCH.XLSX

Bucklin e Lattin (1992) examinaram os padrões de concorrência entre as lojas de um mercado. Eles relataram um comportamento de mudança de loja em 300 membros das famílias examinadas, que, coletivamente, fizeram 30.966 visitas a 13 lojas diferentes do mercado no período de um ano. Os autores usaram esses dados para criar uma matriz de mudança de loja, apresentada na Tabela 8.3 (Capítulo 8) e disponível no arquivo *STORE\_SWITCH*.

- Proponha e calcule uma medida apropriada de similaridade ou discrepância com base no comportamento observado de mudança de loja.
- Use o escalonamento multidimensional para criar um “mapa” do mercado baseado nesse comportamento de mudar de loja (limite-se a duas dimensões). Quão bom é o ajuste? O que o mapa diz sobre as lojas em concorrência cerrada (isto é, entre as quais há mais mudanças)?

Bucklin e Lattin também fornecem um mapa físico das localizações de cada loja do estudo. Seu mapa é reproduzido na Figura 8.3 (veja o Capítulo 8).

- O mapa físico da Figura 8.3 corresponde ao seu mapa do item “a”? O que poderia explicar as diferenças?

### 7.9 Arquivo 09\_STORE\_SHARE.XLSX

Os dados seguintes (disponíveis no arquivo *STORE\_SHARE*) são retirados de um conjunto de dados sobre um detergente líquido para lavagem de roupas. As primeiras linhas de dados são apresentadas na Tabela 7.18.

**Tabela 7.18** Número de aquisições de detergentes para lavagem de roupa por família em cada loja

S03	S04	S07	S16	S18	S21	S24	S26	S29	S36	S10	S43	S45
0	0	0	9	0	11	0	0	2	2	1	0	0
0	1	2	0	0	0	0	1	0	0	1	1	0
0	0	3	0	0	0	0	3	0	0	0	0	0
0	1	6	0	0	0	0	0	0	1	0	2	6
1	4	3	0	0	0	0	0	0	0	0	0	0
0	0	5	0	1	0	2	0	0	6	1	0	1
0	1	4	0	0	1	0	3	0	0	0	0	0
0	1	2	0	0	0	0	0	0	1	0	2	2
2	0	1	0	0	0	0	0	3	0	0	0	0

As colunas da matriz de dados representam lojas (identificadas de S03 até S45) e as linhas representam os membros das famílias estudadas. Cada célula indica o número de compras do detergente em um período fixado de tempo para cada família em cada loja. Essas famílias foram selecionadas porque têm a tendência de comprar em mais de uma loja (nenhuma família na amostra escolhe mais de 50% dos detergentes em qualquer loja).

Use um método de escalonamento multidimensional para obter um *insight* sobre a estrutura competitiva entre as lojas. Descreva seus resultados (e quaisquer pressupostos que você deva fazer).

### 7.11 Arquivo 11\_LOGISTICS.XLSX

Whitlark e Smith (2001) analisaram dados de aproximadamente 800 pessoas que lidam com cargas e expedições através do mundo. Essas pessoas avaliaram 12 companhias que fornecem comunicações, consultoria em logística e serviço de suporte de software. Tais fornecedores, bem conhecidos nos Estados Unidos (Oracle, Nokia e FedEx), foram identificados pelos nomes; os demais fornecedores foram identificados somente pelas letras do alfabeto. A Tabela 7.19 (também disponível no arquivo *LOGISTICS*) mostra a contagem da frequência de atributo da marca; os números relativamente altos para a FedEx refletem o alto nível de conhecimento da marca da companhia por todos os países no estudo.

**Tabela 7.19** Avaliações de 12 fornecedores de serviços por 800 trabalhadores que lidam com transportadoras e expedidoras de cargas

	Oracle	Nokia	FedEx	A	B	C	D	E	F	G	H	I
Inovador	155	375	476	86	30	18	25	21	190	18	408	143
Líder	157	350	675	66	21	12	23	20	307	18	549	225
Solução	109	274	550	105	25	11	33	21	305	16	467	194
Conformidade	133	318	669	110	37	16	36	26	332	25	551	191
Eficiente	151	351	748	117	40	17	34	27	355	21	613	206
Relevante	96	284	627	76	20	12	28	18	309	18	523	184
Essencial	35	91	307	30	9	2	12	9	131	10	239	121
De confiança	170	408	754	122	43	18	35	36	392	29	624	248

- Use um método apropriado de escalonamento multidimensional para obter um mapa que reflita a proximidade de associação entre os diferentes fornecedores e seus atributos. O que o mapa diz sobre quais fornecedores são percebidos como similares pelo pessoal de cargas e expedidores?
- Retire os fornecedores B, C, D, E e G dos dados (todos eles têm níveis baixos e uniformes de conhecimento de marca entre as pessoas que trabalham no setor de cargas e expedições) e faça de novo a análise. Como os resultados se alteraram (se é que houve alguma alteração)?

### 7.12 Arquivo 12\_LOCAL\_CURRENCY.XLSX

A matriz da Tabela 7.20 (também disponível no arquivo *LOCAL\_CURRENCY*) contém as correlações no retorno total trimestral (alteração no preço mais dividendo) da moeda local em 12 países diferentes.

Os dados foram obtidos da Morgan Stanley por Hunter e Coggin (1990) e cobrem o intervalo de tempo entre 1970 e 1986. Os 12 países são:

- |              |            |                        |
|--------------|------------|------------------------|
| 1. Austrália | 5. Japão   | 9. Suécia              |
| 2. Bélgica   | 6. Holanda | 10. Reino Unido        |
| 3. França    | 7. Noruega | 11. Estados Unidos     |
| 4. Itália    | 8. Espanha | 12. Alemanha Ocidental |

Suponha que essas correlações sejam um reflexo da proximidade entre dois países. Analise os dados usando o escalonamento multidimensional. Descreva os resultados de sua análise.

**Tabela 7.20** Matriz de correlação para o retorno trimestral total de 12 países

	1	2	3	4	5	6	7	8	9	10	11	12
1	1,00											
2	0,38	1,00										
3	0,38	0,69	1,00									
4	0,39	0,44	0,52	1,00								
5	0,34	0,36	0,31	0,37	1,00							
6	0,53	0,67	0,54	0,40	0,50	1,00						
7	0,32	0,44	0,35	0,21	0,07	0,34	1,00					
8	0,24	0,33	0,23	0,43	0,43	0,31	0,08	1,00				
9	0,28	0,41	0,28	0,43	0,38	0,45	0,09	0,33	1,00			
10	0,41	0,58	0,44	0,32	0,42	0,67	0,18	0,19	0,34	1,00		
11	0,57	0,50	0,45	0,33	0,48	0,69	0,30	0,28	0,42	0,60	1,00	
12	0,39	0,49	0,48	0,37	0,37	0,63	0,01	0,27	0,39	0,45	0,51	1,00

### 7.13 Arquivo 13\_INTL\_FOODS.XLSX

Hartigan (1975) registrou a porcentagem de famílias de um dado país que tinham em casa um tipo particular de alimento. Os dados estão disponíveis no arquivo *INTL\_FOODS*. A lista de códigos para os países e tipos de alimento é apresentada na Tabela 7.21.

Realize uma análise de desdobramento desses dados e descreva seus resultados. Quais países parecem ser mais similares com respeito às suas preferências alimentícias? Quais alimentos parecem ser mais similares com relação ao seu padrão de consumo nos diferentes países?

**Tabela 7.21** Códigos para os dados *INTL\_FOODS*

Países		Alimentos	
WG	Alemanha Ocidental	GC	Café moído
IT	Itália	IC	Café instantâneo
FR	França	TB	Chá ou chá em sachês
NS	Holanda	SS	Doce sem açúcar
BM	Bélgica	BP	Biscoitos em pacote
LG	Luxemburgo	SP	Sopa (em pacote)
GB	Grã-Bretanha	ST	Sopa (enlatada)
PL	Portugal	IP	Batata instantânea
AA	Áustria	FF	Peixe congelado
SD	Suíça	VF	Vegetais congelados
SW	Suécia	AF	Maçãs frescas
DK	Dinamarca	OF	Laranjas frescas
NY	Noruega	FT	Frutas enlatadas
FD	Finlândia	JS	Geleia (comprada)
SP	Espanha	CG	Dente de alho
ID	Irlanda	BR	Manteiga
		ME	Margarina
		OO	Azeite de oliva, óleo de milho
		YT	Iogurte
		CD	Pão crocante