



RAD5053 – Análise de Dados Multivariados I

Prof. Dr. Evandro Marcos Saidel Ribeiro

Lista 03 – Análise Fatorial

Exercícios do CAPÍTULO 5 do livro “Análise de Dados Multivariados; Lattin, Carroll e Green”

5.1 Arquivo 01_PSYCH_TEST.XLSX

Para fins de ilustração no capítulo, usamos um subconjunto de dados de teste psicológico coletados por Holzinger e Swineford (1939). De fato, Holzinger e Swineford coletaram dados de 145 crianças da 7ª e 8ª séries em um total de 26 diferentes testes psicológicos. Um conjunto maior de nove testes (incluindo os cinco utilizados no capítulo) é apresentado a seguir:

- X_1 Percepção visual
- X_2 Cubos
- X_3 Losangos
- X_4 Compreensão de parágrafos
- X_5 Complementação de sentenças
- X_6 Significado de palavras
- X_7 Adição
- X_8 Contagem de pontos
- X_9 Letras maiúsculas retas e curvas

A matriz de correlação (mostrada na Tabela 5.17) está disponível no arquivo *PSYCH_TEST*. Analise esses dados utilizando a análise fatorial. Quantos fatores há? Como você os interpretaria? Como esses resultados diferem daqueles baseados nos cinco testes apresentados no capítulo?

Tabela 5.17

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9
X_1	1								
X_2	0,318	1							
X_3	0,436	0,419	1						
X_4	0,335	0,234	0,323	1					
X_5	0,304	0,157	0,283	0,722	1				
X_6	0,326	0,195	0,350	0,714	0,685	1			
X_7	0,116	0,057	0,056	0,203	0,246	0,17	1		
X_8	0,314	0,145	0,229	0,095	0,181	0,113	0,585	1	
X_9	0,489	0,239	0,361	0,309	0,345	0,28	0,408	0,512	1

5.2 Arquivo 02_RTE_CEREAL.XLSX

Os dados brutos sobre os cereais prontos para comer coletados por Roberts e Lattin (usados como exemplo de problema no capítulo) estão disponíveis no arquivo *RTE_CEREAL*. O arquivo contém 27 variáveis definidas como segue: a primeira coluna contém a identidade do objeto, a segunda contém o número de identidade do cereal avaliado e as 25 colunas restantes contêm as avaliações de cada um dos 25 atributos. O número de identidade do cereal e a lista ordenada de atributos são dados na Tabela 5.8.

- a. Conduza sua própria análise fatorial dos dados do *RTE_CEREAL*. Tente extrair e rotar cinco fatores, determinando se isso faz alguma diferença em sua interpretação dos dados.

Tabela 5.8 Lista de 25 atributos de cereais RTE e 12 marcas de cereais RTE

Cereais	Atributos	
1. All Bran	Satisfaz	Família
2. Cerola Muesli	Natural	Calorias
3. Just Right	Fibra	Simples
4. Kellogg's Corn Flakes	Doce	Crocante
5. Komplete	Fácil	Regular
6. NutriGrain	Sal	Açúcar
7. Purina Muesli	Gratificante	Fruta
8. Rice Bubbles	Energia	Processo
9. Special K	Divertido	Qualidade
10. Sustain	Crianças	Prazer
11. Vitabrit	Encharcado	Chato
12. Weetbix	Econômico	Nutritivo
	Saúde	

5.3 Arquivo 03_SOFT_DRINKS.XLSX

Sessenta estudantes classificaram 10 marcas de refrigerantes (Coca-Cola, Diet Pepsi, Dr. Pepper, Mt. Dew, Pepsi, Royal Crown, 7Up, Sprite, Diet 7Up, Tab) em relação a quatro atributos (calorias, doçura, saciar a sede e popularidade com os outros) em dois momentos diferentes durante o semestre (setembro e novembro). As variáveis no conjunto de dados são definidas como segue:

- X_1 Calorias (setembro)
- X_2 Calorias (novembro)
- X_3 Doçura (setembro)
- X_4 Doçura (novembro)
- X_5 Sacia a sede (setembro)
- X_6 Sacia a sede (novembro)
- X_7 Popularidade (setembro)
- X_8 Popularidade (novembro)

A matriz de correlação é dada na Tabela 5.18 (e no arquivo *SOFT_DRINKS*).

Analise os dados usando a análise fatorial. Quantos fatores há? Como você os interpretaria?

Tabela 5.18 Matriz de correlação para os dados de *SOFT_DRINKS*

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
X_1	1							
X_2	0,886	1						
X_3	0,649	0,597	1					
X_4	0,588	0,621	0,649	1				
X_5	0,067	0,034	-0,08	-0,136	1			
X_6	0,054	0,076	-0,075	-0,092	0,542	1		
X_7	0,037	0,029	-0,018	-0,054	0,446	0,274	1	
X_8	0,075	0,102	0,089	0,069	0,225	0,267	0,730	1

5.4 Arquivo 04_SIX_VARIABLES.XLSX

Imagine que você foi contratado como assistente de pesquisa por um professor universitário que está fazendo uma análise empírica. Esse professor possui um conjunto de dados com seis variáveis correlacionadas ($n = 100$) e ele solicitou a você que conduzisse uma análise fatorial desses dados. Você recebeu as seguintes informações:

- As variáveis são simplesmente denominadas X_1 a X_6 . O professor quer que você seja informado somente pelos padrões de associação que observa nos dados (e não pelos nomes das variáveis).
 - De acordo com o professor, a maioria das pesquisas empíricas publicadas em sua área é baseada no modelo de fator comum.
 - O professor pretende usar os dados em uma análise subsequente. Portanto, é importante que ele possa interpretar facilmente os resultados que você apresentar a ele.
- A matriz de correlação é apresentada na Tabela 5.19. Os dados brutos estão no arquivo *SIX_VARIABLES*. O professor solicitou-lhe que forneça a ele as respostas às seguintes questões:
- a. Quantos fatores você extrairia dessas seis variáveis? Explique claramente as razões por trás de sua decisão.
 - b. Quanta informação do conjunto original das seis variáveis é explicada por esses fatores?
 - c. Explique claramente (mas de modo sucinto) a relação entre os fatores escolhidos e as variáveis originais.
 - d. Usando sua solução fatorial proposta, como você descreveria as diferenças entre as duas primeiras observações na amostra?

Tabela 5.19 Matriz de correlação para os dados *SIX_VARIABLES*

	X_1	X_2	X_3	X_4	X_5	X_6
X_1	1,000					
X_2	0,849	1,000				
X_3	0,462	0,442	1,000			
X_4	0,416	0,439	0,909	1,000		
X_5	0,409	0,360	0,499	0,501	1,000	
X_6	0,455	0,334	0,478	0,459	0,862	1,000

5.5 Arquivo 05_LUXURY_CAR.XLSX

A matriz de correlação para os dados coletados por Roberts (1984) descrita no início do capítulo é apresentada na Tabela 5.20 (e está disponível no arquivo *LUXURY_CAR*):

- X_1 Luxo e conforto
- X_2 Estilo e modelo
- X_3 Confiabilidade
- X_4 Economia de combustível
- X_5 Segurança
- X_6 Custos de manutenção
- X_7 Qualidade
- X_8 Durabilidade e valor de revenda
- X_9 Desempenho na estrada

Tabela 5.20 Matriz de correlação para os dados *LUXURY_CAR*

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9
X_1	1,000								
X_2	0,591	1,000							
X_3	0,356	0,350	1,000						
X_4	-0,098	0,072	0,38	1,000					
X_5	0,573	0,408	0,382	0,062	1,000				
X_6	0,156	0,232	0,517	0,424	0,303	1,000			
X_7	0,400	0,414	0,611	0,320	0,401	0,479	1,000		
X_8	0,282	0,375	0,512	0,346	0,308	0,463	0,605	1,000	
X_9	0,519	0,484	0,467	0,167	0,455	0,311	0,574	0,557	1,000

Analise os dados de Roberts usando o modelo de fator comum. Extraia e faça a rotação de três fatores comuns e descreva suas descobertas (isto é, forneça uma interpretação concisa e perceptiva dos resultados de sua análise). Como os seus resultados diferem dos resultados de dois fatores apresentados no começo do capítulo?

5.6 Arquivos 06_FOOD_RESEARCH_A.XLSX e 06_FOOD_RESEARCH_B.XLSX

Uma estudante de doutorado em Pesquisa de Alimento possui um conjunto de dados com 10 variáveis. Ela está preocupada com o fato de 10 variáveis ser um número elevado demais para uma análise subsequente que necessita realizar. Ela pretende reduzir de alguma forma seu conjunto de dados – talvez omitindo algumas variáveis – para não mais que cinco variáveis, mas sem prejuízo extremo das informações originais contidas nos dados.

- a. Como você faria para ajudar essa estudante a resolver seu problema? Que abordagem você recomendaria? Qual é a sua solução? Lembre-se de que o resultado final deve ser um conjunto de dados que a estudante possa usar em análise subsequente. Os dados brutos estão disponíveis no arquivo *FOOD_RESEARCH_A*.
- b. A estudante tem a sorte de ter separado um conjunto de dados para servir como amostra de teste com o mesmo número de observações. Como você usaria esses dados para aumentar a confiança da estudante na abordagem proposta no item “a”? Os dados do teste estão disponíveis no arquivo *FOOD_RESEARCH_B*.

5.7 Arquivo 07_MBA_CAR.XLSX

Em janeiro de 1998, 303 estudantes de MBA foram entrevistados a respeito de suas avaliações e preferências sobre 10 diferentes automóveis. Os automóveis, listados em ordem de apresentação na pesquisa, foram BMW 328i, Ford Explorer, Infiniti J30, Jeep Grand Cherokee, Lexus ES300, Chrysler Town & Country, Mercedes C280, Saab 9000, Porsche Boxster e Volvo V90. Cada estudante classificou todos os 10 carros. Para os fins deste exercício, um carro foi selecionado aleatoriamente por cada um dos estudantes, resultando em um tamanho de amostra de 303 avaliações.

Os estudantes classificaram cada carro em relação a 16 atributos. As primeiras oito questões pediam que os estudantes avaliassem o grau em que cada uma das seguintes palavras descrevia um determinado carro (em que 5 = "Extremamente descritiva" e 1 = "Não descreve de maneira alguma"): excitante, confiável, luxuoso, próprio para o ar livre, poderoso, estiloso, confortável e vigoroso. As oito questões seguintes solicitavam que os estudantes classificassem seu nível de concordância com cada uma das seguintes afirmações sobre um determinado carro (em que 5 = "Concordo totalmente" e 1 = "Discordo totalmente"):

- "Este carro é gostoso de dirigir".
- "Este carro é seguro".
- "Este carro tem um grande desempenho".
- "Este carro é para a família".
- "Este carro é versátil".
- "Este carro é esportivo".
- "Este carro é um carro de alto padrão".
- "Este carro é prático".

Os dados brutos estão disponíveis no arquivo *MBA_CAR*. Há 18 variáveis no arquivo, definidos como segue:

Identidade do estudante (disfarçado)
Identidade do carro:

1 = BMW 328i	7 = Mercedes C280
2 = Ford Explorer	8 = Saab 9000
3 = Infiniti J30	9 = Porsche Boxster
4 = Jeep Grand Cherokee	10 = Volvo V90
5 = Lexus ES300	X_1 Excitante
6 = Chrysler Town & Country	X_2 Confiável
X_3 Luxuoso	X_{10} Seguro
X_4 Próprio para o ar livre	X_{11} Desempenho
X_5 Poderoso	X_{12} Família
X_6 Estiloso	X_{13} Versátil
X_7 Confortável	X_{14} Esportivo
X_8 Vigoroso	X_{15} Status
X_9 Gostoso	X_{16} Prático

- a. Realize uma análise de fator comum do conjunto de dados. Quantos fatores você reteria? Como você os interpretaria?
- b. Salve os escores fatoriais e esquematize os escores fatoriais médios para cada um dos 10 carros avaliados pelos estudantes. O que o gráfico diz sobre as semelhanças e as diferenças dos 10 modelos de carro?

5.8 Arquivo 08_EMOTIONS.XLSX

Westbrook e Oliver (1991) investigaram os tipos de respostas emocionais para a experiência de consumo. Eles realizaram um estudo de uma amostra de proprietários de carros recém-adquiridos. Um total de 125 questionários foi preenchido com respeito às reações emocionais dos respondentes em relação aos seus automóveis. Os autores usaram a medida DES-II de Izard (1997), que contém 10 subescalas que representam a frequência com a qual os sujeitos experimentam cada uma das 10 emoções relacionadas a seguir:

X ₁	Interesse	X ₅	Raiva	X ₉	Vergonha
X ₂	Alegria	X ₆	Desgosto	X ₁₀	Culpa
X ₃	Surpresa	X ₇	Desprezo		
X ₄	Tristeza	X ₈	Medo		

A matriz de correlação, reproduzida na Tabela 5.21, está disponível no arquivo *EMOTIONS*. Em sua opinião, qual é a dimensionalidade do espaço psicológico que contém os vários padrões de resposta emocional? Como você interpretaria a(s) dimensão(ões) que encontrar?

Tabela 5.21 Matriz de correlação para os dados *EMOTIONS*

	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈	X ₉	X ₁₀
X ₁	1									
X ₂	0,20	1								
X ₃	0,08	0,30	1							
X ₄	0,13	-0,30	0,37	1						
X ₅	0,27	-0,22	0,28	0,80	1					
X ₆	0,22	-0,23	0,39	0,84	0,85	1				
X ₇	0,22	-0,20	0,45	0,76	0,82	0,92	1			
X ₈	0,33	-0,07	0,46	0,67	0,72	0,80	0,83	1		
X ₉	0,31	-0,08	0,40	0,55	0,60	0,67	0,77	0,76	1	
X ₁₀	0,25	-0,21	0,48	0,74	0,68	0,78	0,78	0,7	0,70	1