

Trabalho MATLAB

Cintia Izumi Shinoda

I Introdução

Este trabalho tem como objetivo realizar uma Análise de Componentes Principais (PCA) utilizando MATLAB como ferramenta a fim de demonstrar o domínio desta.

O *dataset* analisado é referente ao consumo de 9 tipos de proteínas em 25 países europeus (WEBER, 1973).

Foram utilizados os conhecimentos adquiridos na disciplina Análise Multivariada da Especialização em Métodos Matemáticos Aplicados.

A análise PCA permitirá destacar os principais componentes (redução da complexidade dos dados) que explicam a maior parte da variação no consumo de proteínas entre os países, oferecendo *insights* sobre hábitos alimentares e agrupamentos de países com padrões de consumo de proteína semelhantes.

2 PCA – Análise de Componentes Principais

2.1 Importação dos dados

25x9 [table](#)

	RedMeat	WhiteMeat	Eggs	Milk	Fish	Cereals	Starch	Nuts	Fr_Veg
Albania	10.1	1.4	0.5	8.9	0.2	42.3	0.6	5.5	1.7
Austria	8.9	14	4.3	19.9	2.1	28	3.6	1.3	4.3
Belgium	13.5	9.3	4.1	17.5	4.5	26.6	5.7	2.1	4
Bulgaria	7.8	6	1.6	8.3	1.2	56.7	1.1	3.7	4.2
Czechoslovakia	9.7	11.4	2.8	12.5	2	34.3	5	1.1	4
Denmark	10.6	10.8	3.7	25	9.9	21.9	4.8	0.7	2.4
E Germany	8.4	11.6	3.7	11.1	5.4	24.6	6.5	0.8	3.6
Finland	9.5	4.9	2.7	33.7	5.8	26.3	5.1	1	1.4
France	18	9.9	3.3	19.5	5.7	28.1	4.8	2.4	6.5
Greece	10.2	3	2.8	17.6	5.9	41.7	2.2	7.8	6.5
Hungary	5.3	12.4	2.9	9.7	0.3	40.1	4	5.4	4.2
Ireland	13.9	10	4.7	25.8	2.2	24	6.2	1.6	2.9
Italy	9	5.1	2.9	13.7	3.4	36.8	2.1	4.3	6.7
Netherlands	9.5	13.6	3.6	23.4	2.5	22.4	4.2	1.8	3.7
Norway	9.4	4.7	2.7	23.3	9.7	23	4.6	1.6	2.7
Poland	6.9	10.2	2.7	19.3	3	36.1	5.9	2	6.6
Portugal	6.2	3.7	1.1	4.9	14.2	27	5.9	4.7	7.9
Romania	6.2	6.3	1.5	11.1	1	49.6	3.1	5.3	2.8
Spain	7.1	3.4	3.1	8.6	7	29.2	5.7	5.9	7.2
Sweden	9.9	7.8	3.5	24.7	7.5	19.5	3.7	1.4	2
Switzerland	13.1	10.1	3.1	23.8	2.3	25.6	2.8	2.4	4.9
UK	17.4	5.7	4.7	20.6	4.3	24.3	4.7	3.4	3.3
USSR	9.3	4.6	2.1	16.6	3	43.6	6.4	3.4	2.9
W Germany	11.4	12.5	4.1	18.8	3.4	18.6	5.2	1.5	3.8
Yugoslavia	4.4	5	1.2	9.5	0.6	55.9	3	5.7	3.2

Figura 1: Tabela com os dados importados.

2.2 Matriz de Covariância

A matriz de covariância é utilizada para encontrar as direções de maior variância dos dados.

Matriz de Covariâncias									
11.2029	1.8918	2.1906	11.9609	0.6942	-18.3622	0.7407	-2.3225	-0.4481	
1.8918	13.6462	2.5614	7.3884	-2.9413	-16.7761	1.8941	-4.6576	-0.4086	
2.1906	2.5614	1.2491	4.5704	0.2494	-8.7385	0.8259	-1.2423	-0.0918	
11.9609	7.3884	4.5704	50.4869	3.3335	-46.2218	2.5824	-8.7630	-5.2342	
0.6942	-2.9413	0.2494	3.3335	11.5772	-19.5759	2.2454	-0.9942	1.6335	
-18.3622	-16.7761	-8.7385	-46.2218	-19.5759	120.4459	-9.5634	14.1868	0.9215	
0.7407	1.8941	0.8259	2.5824	2.2454	-9.5634	2.6702	-1.5390	0.2488	
-2.3225	-4.6576	-1.2423	-8.7630	-0.9942	14.1868	-1.5390	3.9429	1.3431	
-0.4481	-0.4086	-0.0918	-5.2342	1.6335	0.9215	0.2488	1.3431	3.2541	

Figura 2: Tabela com a matriz de covariâncias.

2.3 Matriz de Correlação

A matriz de correlação, é a matriz de covariância normalizada.

Matriz de Correlações

1.0000	0.1530	0.5856	0.5029	0.0610	-0.4999	0.1354	-0.3494	-0.0742
0.1530	1.0000	0.6204	0.2815	-0.2340	-0.4138	0.3138	-0.6350	-0.0613
0.5856	0.6204	1.0000	0.5755	0.0656	-0.7124	0.4522	-0.5598	-0.0455
0.5029	0.2815	0.5755	1.0000	0.1379	-0.5927	0.2224	-0.6211	-0.4084
0.0610	-0.2340	0.0656	0.1379	1.0000	-0.5242	0.4039	-0.1472	0.2661
-0.4999	-0.4138	-0.7124	-0.5927	-0.5242	1.0000	-0.5333	0.6510	0.0465
0.1354	0.3138	0.4522	0.2224	0.4039	-0.5333	1.0000	-0.4743	0.0844
-0.3494	-0.6350	-0.5598	-0.6211	-0.1472	0.6510	-0.4743	1.0000	0.3750
-0.0742	-0.0613	-0.0455	-0.4084	0.2661	0.0465	0.0844	0.3750	1.0000

Figura 3: Tabela com a matriz de correlações.

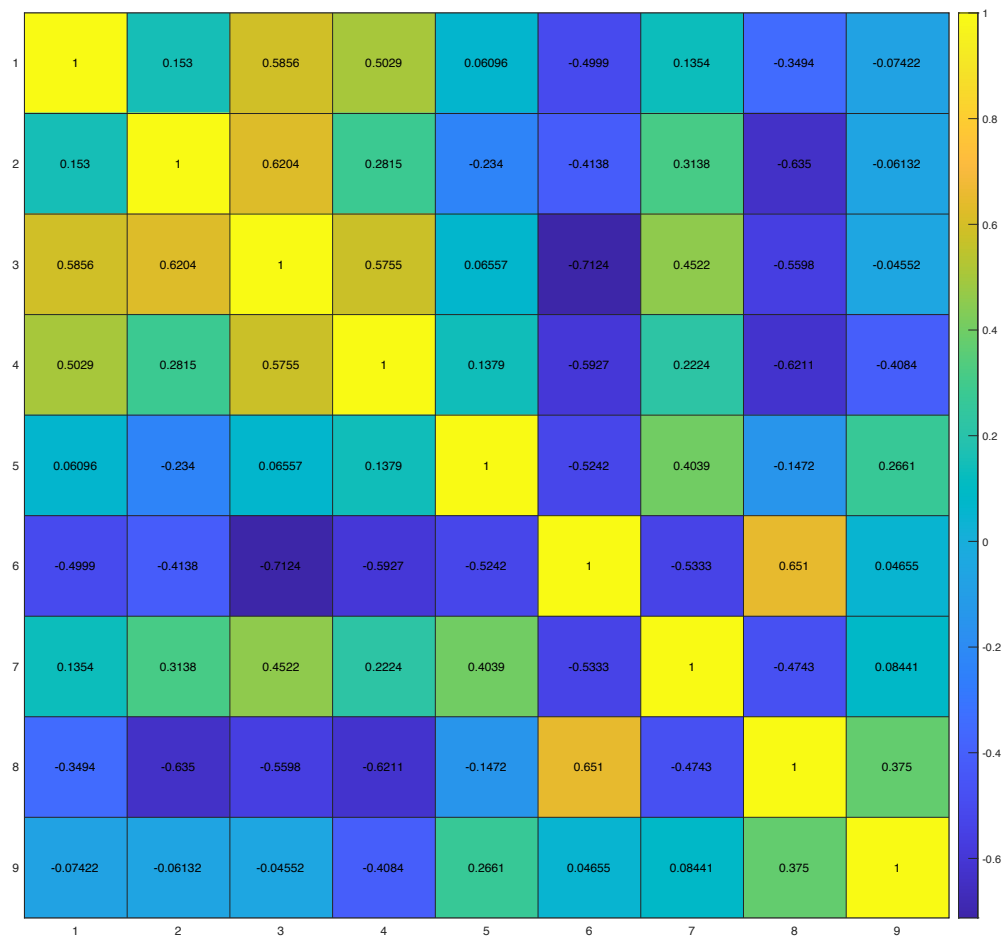


Figura 4: Gráfico de calor para a matriz de correlação.

2.4 PCA

Componentes principais:

-0.2903	-0.0951	-0.4041	0.5846	0.3872	-0.4404	-0.0257	-0.1218
-0.3132	-0.2639	0.6117	0.0584	-0.2927	-0.1637	-0.1644	-0.3114
-0.4132	-0.0742	0.1232	0.3223	0.0757	0.5315	-0.2390	0.5988
-0.3852	-0.1539	-0.3958	-0.1363	-0.2995	0.3205	0.6192	-0.1408
-0.1165	0.6978	-0.2385	-0.1648	-0.2889	-0.1572	-0.2739	0.1896
0.4343	-0.2753	0.0642	-0.1025	0.1808	-0.1569	0.2873	0.5117
-0.2796	0.3638	0.3086	-0.3045	0.7010	0.1175	0.2449	-0.1483
0.4400	0.0769	-0.1164	0.2701	0.1231	0.5740	-0.2029	-0.4321
0.1568	0.4372	0.3478	0.5736	-0.2151	-0.0382	0.5226	0.0565

Variância explicada:

75.5071
12.4314
6.3770
3.7651
1.0638
0.4627
0.3307
0.0623

Figura 5: Porcentagem de variância explicada por cada componente.

Autovalores:

1.8390
0.3028
0.1553
0.0917
0.0259
0.0113
0.0081
0.0015

Figura 6: Variância explicada por cada componente.

3 Referências

SCOTT, J. G. Protein data set. Disponível em:

<https://raw.githubusercontent.com/jgscott/STA380/master/data/protein.csv>. Acesso em: 30 nov. 2024.