

Projeto Final para Disciplina Fundamentos da Matemática para Computação

16/05/2018

Erick Fernandes da Cruz - Universidade Federal do ABC - Santo André - SP

Ferramenta Escolhida e Arquivos Submetidos

Para realização do projeto foi utilizada a ferramenta Scilab que apresenta ferramentas adequadas para realização das atividades propostas.

A submissão contém 4 arquivos sem contar o presente relatório, sendo eles:

O arquivo neighborhood.m que apresenta o grafo de vizinhança entre os países da pesquisa.

A tabela drinksEdit.csv que apresenta os dados da pesquisa editados para um formato adequado para extração de suas informações pelo Scilab.

O script SegmentacaoEmClusters.sci que realiza a segmentação da matriz de vizinhança entre países em quatro clusters.

E por fim o script CalcularEstatisticasEXChapeu.sci que extrai a estatística da matriz inicial e da matriz relacionada a cada um dos clusters.

Para reprodução do projeto é necessário executar primeiramente o script SegmentacaoEmClusters.sci e em seguida CalcularEstatisticasEXChapeu.sci.

Clusters Gerados

Para separação dos países em cluster foi utilizada a técnica de decomposição espectral, utilizando como base um grafo contendo a vizinhança entre estes países. Para realização do presente projeto inicialmente foram gerados dois clusters, e a partir desses dois clusters foram gerados mais dois clusters de cada um, totalizando um total de quatro clusters. A seguir está a lista de países pertencentes a cada um desses clusters.

- Cluster A:

Argentina, Belize, Bolivia, Brazil, Canada, Chile, Colombia, Costa Rica, Cuba, Ecuador, El Salvador, Fiji, Guatemala, Guyana, Honduras, Ireland, Marshall Islands, Micronesia, Nauru, Nicaragua, Panama, Paraguay, Peru, St. Kitts & Nevis, Samoa, Sao Tome & Principe, Suriname, Tuvalu, United Kingdom, USA, Uruguay e Venezuela.

- Cluster B:

Antigua & Barbuda, Bahrain, Barbados, Cabo Verde, Jamaica, Kiribati, Niue, St. Vincent & the Grenadines, Sri Lanka, Trinidad & Tobago e Vanuatu.

- Cluster C:

Por conter 138 países, não irei especificar cada um deles, basta considerar que são os países que não estão nos Clusters A, B e D.

- Cluster D:

Dominica, Dominican Republic, Grenada, Haiti, Iceland, Japan, Malta, Micronesia, St. Lucia, Seychelles, Singapore e Tonga.

Dados Estatísticos

Foram selecionados quatro tipos de métricas estatísticas para auxiliar na análise sobre os dados de cada um dos conjuntos estudados: Mediana, média, variância e desvio padrão. Os resultados obtidos estão expostos nas tabelas a seguir separadas pelos seus respectivos conjuntos.

Todos os Países				
	Cerveja	Destilados	Vinho	Total de Litros de Álcool
Mediana	76,00	56,00	8,00	4,20
Média	106,16	80,99	49,45	4,72
Variância	10229,93	7794,12	6351,71	14,24
Desvio Padrão	101,14	88,28	79,70	3,77

Cluster A				
	Cerveja	Destilados	Vinho	Total de Litros de Álcool
Mediana	154,00	99,00	8,50	5,55
Média	152,78	96,94	48,78	5,37
Variância	7989,21	4091,35	5200,50	7,96
Desvio Padrão	89,38	63,96	72,11	2,82

Cluster B				
	Cerveja	Destilados	Vinho	Total de Litros de Álcool
Mediana	102,00	104,00	9,00	4,00
Média	97,82	113,64	13,64	4,04
Variância	4451,56	4625,45	200,25	5,23
Desvio Padrão	66,72	68,01	14,15	2,29

Cluster C				
	Cerveja	Destilados	Vinho	Total de Litros de Álcool
Mediana	46,00	28,00	7,00	2,95
Média	95,17	67,37	53,62	4,52
Variância	11096,87	7017,74	7466,98	16,65
Desvio Padrão	105,34	83,77	86,41	4,08

Cluster D				
	Cerveja	Destilados	Vinho	Total de Litros de Álcool
Mediana	113,00	124,00	22,00	6,40
Média	115,83	165,25	36,17	5,82
Variância	5781,42	21026,75	1336,15	10,45
Desvio Padrão	76,04	145,01	36,55	3,23


Além dessas medidas estatísticas, foi realizada uma solução aproximada do sistema linear para a matriz completa e para cada um dos clusters separadamente. Como estamos resolvendo um sistema sobredeterminado, estamos interessados em uma solução aproximada. A solução é obtida utilizando a seguinte fórmula $X = (A^T * A)^{-1} * A^T * b$. As soluções destes problemas estão exibidas na tabela abaixo separadas por cada uma das bebidas.

	Cerveja	Destilados	Vinho
Total	24,05	17,50	12,14
Cluster A	28,93	17,12	9,31
Cluster B	24,80	28,09	3,02
Cluster C	23,35	15,37	14,33
Cluster D	17,39	32,17	5,08

Interpretação dos Resultados

Inicialmente vamos considerar os dados obtidos pela resolução do sistema linear. É possível visualizar que os resultados têm relação com quantidade de álcool puro que cada uma das bebidas possuem e que ao tentarmos visualizar o conjunto de todos os países, não temos uma análise muito adequada visto a diferença que os diferentes clusters de países apresentaram.

Se considerarmos o Cluster D por exemplo a taxa de álcool em destilados é quase o dobro que na cerveja, já no cluster A podemos visualizar exatamente o oposto. O cluster que mais se aproxima da análise do conjunto completo é o Cluster C, mas isso também se deve



ao tamanho do conjunto ser muito grande, o que pode camuflar certas peculiaridades regionais. No geral podemos ver que a taxa de álcool puro no vinho foi a menor em todos os clusters, sendo que no cluster C a taxa se mostrou bem semelhante a encontrada em destilados.

Agora considerando os dados estatísticos obtidos é possível visualizar que todas as regiões apresentam valores de consumo muito diferentes entre os países de um mesmo cluster, ou seja, mesmo que as taxas de álcool nas bebidas apresentem algumas diferenças regionais (como vimos anteriormente), o consumo destas bebidas não apresenta uma boa aproximação, como pode ser visto pela variância e desvio padrão de todos os conjuntos de dados.