



UNIVERSIDADE DE SÃO PAULO
FACULDADE DE ECONOMIA, ADMINISTRAÇÃO E CONTABILIDADE
DEPARTAMENTO DE ADMINISTRAÇÃO



RAD1509 –Estatística Aplicada à Administração II

Lista 05

Análise de Agrupamentos



(1)

A tabela a seguir apresenta o resultado de um questionário aplicado em uma amostra de 10 alunos para saber o grau de satisfação destes sobre o curso de administração. Os quesitos avaliados foram: conhecimento adquirido, avaliação do corpo docente e mercado de trabalho. As notas são atribuídas em uma escala quantitativa de 0 a 10 para cada questão.

Preencha os valores do grau de satisfação do aluno A com os três últimos dígitos do teu número USP.

Aluno	Conhecimento	Docentes	Mercado
A			
B	9	7	6
C	8	7	8
D	2	3	2
E	6	5	4
F	6	7	8
G	4	2	3
H	3	2	4
I	6	4	5
J	3	1	3

Com base nesta base de dados e por meio da aplicação da técnica de análise de conglomerados hierárquicos com a distância quadrática euclidiana com a utilização do método da ligação individual (*Single Linkage*), identifique o número de conglomerados e interprete os *outputs* gerados.

(1.a) Utilize as variáveis originais (ou seja, não padronize os dados). Considere a distância euclidiana ao quadrado e obtenha a matriz de dissimilaridades.

(1.b) Obtenha o dendrograma correspondente à aplicação do método hierárquico descrito acima.

(1.c) Indique o número de agrupamentos adequado nesta análise e quais alunos são integrantes de cada grupo.

(1.d) Obtenha, para cada grupo, as médias do grau de satisfação em cada uma das três dimensões, conhecimento, Docentes e Mercado.

(1.e) Interprete os resultados obtidos. Ou seja, caracterize cada grupo obtido, utilizando os resultados dos itens anteriores.

(2) Considere a base de dados “*SupermercadosBrasileiros.csv*”, disponibilizada no STOA.

Sejam as variáveis relativas a faturamento bruto em 2005 (R\$), área de vendas (m²) e número de funcionários de 30 grupos supermercadistas do Brasil:

(2.a)

Aplique a técnica de análise de conglomerados hierárquicos, utilizando a distância quadrática euclidiana e o método *Nearest Neighbor*. Antes, padronize as variáveis pelo método Z scores. Analise todas as saídas do processamento (*hierarchical cluster* e *K-means*) e ofereça recomendações quanto ao agrupamento das empresas, justificando sua decisão.

(2.b)

Refaça o item (a), porém agora com o método *Between Groups* e distância quadrática euclidiana.

(2.c)

Refaça o item (a), porém agora sem os três maiores *players*

(3) Os dados estão no arquivo “*Siderurgia.csv*”, disponibilizada no STOA.

Sendo um analista de mercado, você está interessado em segmentar um conjunto de empresas que atuam no setor de siderurgia e metalurgia, com base nas informações financeiras publicadas. Para tanto foram selecionadas 25 empresas que atuam neste segmento de acordo com os dados divulgados pela revista Exame – Melhores e Maiores do ano de 2005, sendo escolhidos quatro indicadores para o estudo:

Fatur: Faturamento expresso em milhões de dólares;

Rent: Rentabilidade do Patrimônio Líquido (%);

Endiv: Endividamento geral da empresa (%);

Empreg: Número de empregados.

(3.a) Faça uma análise de agrupamento utilizando o método hierárquico *Between Groups*:

- Padronize as variáveis pelo Score-z
- Considere a distância euclidiana quadrática
- Grave duas soluções: uma solução com 2 agrupamentos e outra com 3 agrupamentos.
- Obtenha o Dendrograma

(3.b) Faça uma análise de agrupamento utilizando o método hierárquico de Ward

Considere os mesmos itens do exercício (3.a)

(3.c) Faça uma análise de agrupamento utilizando o método não-hierárquico K-means.

- Antes de utilizar o método, padronize as variáveis.
- Escolha o número de Clusters igual a dois
- Grave os integrantes dos agrupamentos
- Observe os centros dos clusters finais

(3.d) Repita o exercício 3.c escolhendo agora o número de Clusters igual a três.

(3.e) Compare todas as análises realizadas. Qual delas você indicaria como melhor para a segmentação?