Silva Lusitana ***16****(2): 207 - 227, 2008*
© EFN, Lisboa. Portugal

O Algoritmo BARCOR: Classificação de Cortiça para Rolhas Recorrendo a Quatro Atributos de Qualidade

Luís Soares Barreto
Professor Catedrático Jubilado
Av. do Movimento das Forças Armadas, 41-3D, 2825-372 COSTA DA CAPARICA

Sumário. O autor propõe um método de análise, denominado algoritmo BARCOR, que recorre ao TOPSIS ("Technique for Order Preference by Similarity to Ideal Solution"), regressão linear múltipla, análise de agrupamentos ("cluster analysis") e modelos de árvore para seleccionar o número mínimo de atributos da qualidade da cortiça, para o fabrico de rolhas, necessários para classificar eficientemente amostras deste material. Utiliza o algoritmo BARCOR para mostrar que é possível classificar amostras de cortiça recorrendo a um menor número de atributos, e com melhores índices de comportamento dos algoritmos de agrupamento usados. Na situação de igualdade de ponderação dos atributos, mostra bastarem as seguintes quatro características da cortiça, para este efeito: força correspondente ao binário máximo, ângulo correspondente ao binário máximo, área máxima dos poros da barriga, área máxima dos poros transversais. Ponderando os atributos, reduz este número a três: calibre, área máxima dos poros da barriga, área máxima dos poros transversais. Propõe um modelo de quatro variáveis e outro de três, para se obter, preliminar e expeditamente, um ordenamento relativo das amostras de cortiça que se queiram classificar.
Palavras-chave: Algoritmo; classificação; cortiça; qualidade; rolhas

Algorithm BARCOR: the Classification of Cork for Stoppers Using Four Quality Attributes

Abstract. The author proposes an analytical method named algorithm BARCOR that integrates TOPSIS ("Technique for Order Preference by Similarity to Ideal Solution"), regression, tree models and cluster analysis, to select the minimum number of attributes of cork quality to efficiently classify this material when used to the production of stoppers. When using attributes evenly weighted he shows that only the four attributes are sufficient. If unevenly weights of the attributes are used, this number can be reduced to three. He proposes two models with four and three attributes as explanatory variables to obtain expedite and preliminary classifications of the cork. He admits that the algorithm has potential to be applied in other similar situations.
Key words: Algorithm; classification; cork; quality; stoppers

L'Algorithme BARCOR: la Classification des Écorces pour Bouchon Utilisant Quatre Attributs de Qualité

Résumé. L'auteur propose une méthode analytique appelée algorithme BARCOR qui intègre la technique TOPSIS ("Technique for Order Preference by Similarity to Ideal Solution"), la régression, l'analyse des groupements, et modèles d'arbres pour choisir le numéro minimum

des attributs de qualité pour classifier, avec efficience, des écorces pour bouchon. L'application de l'algorithme usant des poids égaux pour les attributs permit d'utiliser seulement quatre attributs. Si les poids utilisés ne sont pas égaux il est possible d'obtenir des classifications avec l'emploi de trois attributs. L'auteur propose aussi deux modèles pour classifier de façon expéditive et préliminaire les écorces, avec quatre et trois variables indépendantes. Il admit que l'algorithme peut être appliqué à des situations similaires.
Mots clés: Algorithme; classification; écorce; qualité; bouchon

Introdução

Na competição entre rolhas de cortiça, "screwcaps" e vedantes plásticos, aquela tem em seu desfavor a maior variação das características de um material de origem biológica, comparativamente a um produto fabricado.

Neste contexto, o controlo da qualidade da cortiça ganha relevância, e factores de competitividade favorecem não só um controlo de qualidade eficiente, mas também expedito e ao menor custo possível.

Em PESTANA, LOUZADA e SARDINHA (2005a), os autores classificam rabanadas de cortiça para o fabrico de rolhas considerando só dez atributos de qualidade, obtendo um coeficiente de correlação cofenética (r) igual a 0,79 (figura 1 destes autores), para o método da análise de agrupamento ("cluster analysis") que utilizam. Quando acrescentam ao conjunto das dez variáveis o teor em suberina obtêm r=0,67 (página 61 do mesmo artigo).

Esta ocorrência levou-me a empreender a elaboração do presente estudo. Nele procurei estabelecer o menor número de atributos necessários para qualificar amostras de cortiça para o fabrico de rolhas, com o melhor desempenho dos algoritmos utilizados.

Este posicionamento analítico privilegia uma abordagem macroscópica do assunto.

O presente trabalho surge na esteira de outras contribuições, como as de PEREIRA (1998), PESTANA, LOUZADA e SARDINHA (2005a,b), e as mencionadas por estes autores.

Anotação breve sobre classificações

Esta secção responde à necessidade de proporcionar enquadramento adequado, mínimo, à discussão dos resultados da análise que exporei adiante.

O tema classificação não é trivial, mas considero desajustado alongar-me para além do expendido nesta secção. O leitor interessado numa abordagem epistemológica geral pode consultar GIL (2000). MAHNER e BUNGE, (1997: capítulo 7) tratam a classificação na perspectiva das ciências da vida.

As classificações podem ser ditas "objectivas" se qualquer pessoa, usando os mesmos dados, algoritmo e instrumentos de cálculo comparáveis, obtém o mesmo resultado. Isto é, são objectivas sobretudo por exibirem muito elevada reprodutibilidade. No entanto, métodos diferentes, ditos objectivos, de classificação, com os mesmos dados, não propiciam exactamente a mesma classificação. Não há pois classificações absolutas. Qualquer classificação é uma convenção (e.g., MAHNER e BUNGE, 1997:247).

Convém ainda anotar, por exemplo, que o mesmo método, dito objectivo, com os mesmos dados, dá resultados

diferentes se utilizarmos definições diferentes de distância, ou na situação de variáveis que representem um "custo" para a qualidade do objecto, se os valores dos atributos forem normalizados de maneira diferente.

Qualquer livro dedicado à análise multivariada recomenda a utilização de mais de um procedimento analítico, e a comparação crítica dos seus resultados (e.g., DIGBY e KEMPTON, 1987:147-149). Este artigo vem alargar o esforço realizado por PESTANA, LOUZADA e SARDINHA (2005a,b), neste âmbito, ao analisarem amostras de cortiça caracterizadas por 10 atributos de qualidade, que mediram.

Em antinomia, serão tidas como "subjectivas" as classificações em que a percepção da semelhança só recorra a representações de origem sensível (em oposição às só detectadas por algoritmos numéricos e geometricamente representáveis), onde o factor pessoal atido à perícia e experiência do classificador desempenha um papel decisivo, e por isso o resultado produzido não exibe reprodutibilidade elevada.

Outro aspecto das classificações é a sua objectividade relativa à utilização em questão. Isto passa pela selecção adequada dos dados e métodos de classificação. Por exemplo, aplicar métodos que não permitam a ponderação de atributos, em situações em que alguns destes tenham uma valorização destacada e, mesmo, decisiva na maneira como lidamos com os objectos em apreço, revela uma escolha não objectiva ou inadequada do método de classificação.

Os métodos objectivos não são todos iguais e, por isso, em cada situação devem ser criteriosamente seleccionados e os seus resultados analisados, não só no âmbito restrito da amostra analisada, mas também no contexto da sua generalizada utilização, com outros conjuntos dos mesmos objectos e com o mesmo propósito. Trivialmente, sofisticação estatística e matemática não são garante de objectividade relativa.

Particularmente, sobre os métodos da análise de agrupamento gostaria de ecoar ALDENDERFER e BLASHFIELD (1987:14-16): a) são heurísticos e não se apoiam num corpo estruturado de análise estatística, além de terem origem em diversas disciplinas, pelo que são afectados pelas suas preferências e necessidades específicas; b) não procuram uma estrutura nos dados, mas **impõem-lhe** uma estrutura, isto é, estabelecem grupos quer eles existam ou não, na realidade. Mesmo em objectos distribuídos uniformemente pelo espaço das variáveis, eles encontram grupos!

Os dados utilizados

Os dados a que recorro são os constantes no quadro 5, do interessante e recente trabalho publicado por três autores (PESTANA, LOUZADA e SARDINHA, 2005b), o que privilegia também a emergência de alguma, e desejada, comparabilidade.

Os atributos de qualidade da cortiça usados inserem-se no quadro 1, obtidos dos quadros 2 e 3 de (PESTANA *et al.*, 2005b). As variáveis do quadro 1 formam o conjunto C0.

É meu entendimento que esta utilização de informação reforça a valia do trabalho que acabo de mencionar.

Quadro 1 - Atributos da qualidade da cortiça. Fonte: PESTANA, LOUZADA e SARDINHA (2005b)

Símbolo	Descrição
x_1	Força correspondente ao binário máximo (daN)
x_2	Ângulo correspondente ao binário máximo (grau)
x_3	Calibre da cortiça (mm)
x_4	Massa volúmica da cortiça (Kg m^{-3})
x_5	Humidade (%)
x_6	Poros da barriga (análise de imagem)
x_7	Área máxima de poros de barriga (análise de imagem, mm^2)
x_8	Poros transversais (análise de imagem)
x_9	Área máxima de poros transversais (análise de imagem, mm^2)
x_{10}	Área mínima de poros transversais (análise de imagem, mm^2)

O algoritmo BARCOR

Apliquei às 38 amostras de cortiça, caracterizadas de acordo com o quadro 5 de PESTANA *et al.,* (2005b), o método TOPSIS ("Technique for Order Preference by Similarity to Ideal Solution"; YOON e HWANG, 1995:38-45), que gerou a proximidade relativa (numérica – variável y) a uma amostra virtual (arquétipo) de qualidade melhor que todas as constantes da amostra, e criada a partir delas.

O método TOPSIS, do âmbito da análise decisória mencionada, cria dois arquétipos virtuais do melhor e pior "objecto" (amostra de cortiça) e ordena entre eles os objectos gerando índices de proximidade relativa aos dois objectos virtuais. Só utilizei a proximidade ao melhor arquétipo, que pode ser aceite como um índice relativo da qualidade da cortiça. A soma das duas proximidades é igual a 1. O TOPSIS é de muito simples aplicação e localiza os objectos da colecção analisada ao longo de uma escala linear, gerando resultados inequívocos e numéricos. A criação de objectos virtuais é frequente nos métodos de decisão e classificação de atributos múltiplos, como nos algoritmos que recorrem a centróides, na análise de agrupamento. Menciono, por exemplo, o método "unweighted pair-group centroid clustering", conhecido pela sigla UPGMC (e.g., SNEATH e SOKAL, 1973:234-240).

Os valores da proximidade relativa encontrados (variável dependente) e dos dez atributos foram utilizados com a rotina "stepwise" do S-Plus, para estabelecer o "melhor" modelo linear, encontrando o inserido na caixa 2. As variáveis independentes deste modelo (x_1, x_2, x_5, x_6, x_7, x_8, x_9) serviram de fundamento para o resto da análise.

Apliquei agora o método TOPSIS, o algoritmo UPGMA ("unweighted pair-group method using arithmetic averages") ou "average method", e um algoritmo divisivo hierarquizado de agrupamento, utilizando os dez atributos, dois conjuntos de sete ([x_1, x_2, x_3, x_6, x_7, x_8, x_9], C1; [x_1, x_2, x_5, x_6, x_7, x_8, x_9], C2) e um de seis ([x_1, x_2, x_6, x_7, x_8, x_9], C3). Toda a análise de agrupamento foi feita recorrendo às rotinas específicas do R 2.00 e do S-Plus (agnes, diana, da biblioteca cluster).

Obtida a informação anterior, que criou um quadro de referência multifacetado, passei a explorar a possibilidade de reduzir o conjunto C3. Para comodidade de referência, os conjuntos de variáveis são compilados no quadro 2.

Quadro 2 - Conjuntos de variáveis. O X regista a presença da variável no conjunto

Conjuntos	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
C0	X	X	X	X	X	X	X	X	X	X
C1	X	X	X			X	X	X	X	
C2	X	X			X	X	X	X	X	
C3	X	X				X	X	X	X	
C4	X	X	X				X		X	
C5	X	X					X		X	

Para isso procurei a regressão com o menor número de variáveis que podia ajustar-se aos valores da qualidade relativa fornecidos pelo TOPSIS quando usa C3. Encontrei [x_1, x_2, x_7, x_9] (C5), passando a designar C4 o conjunto [x_1, x_2, x_3, x_7, x_9].

Submeti estes dois conjuntos a uma análise semelhante à anterior.

Todas as aplicações do método TOPSIS foram feitas no Excel. As folhas de cálculo têm características que as tornam particularmente adequadas à aplicação dos métodos da análise decisória em situação de atributos múltiplos. Os atributos (x_1, ... x_{10}) tinham todos o mesmo peso ou importância relativa.

Os resultados obtidos foram, então, avaliados, quanto à sua coerência global, sensibilidade à composição da amostra, e nível dos seguintes índices: coeficiente de agrupamento ("agglomerative coefficient"; MATHSOFT, 1999b:92), coeficiente divisivo ("divisive coefficient", MATHSOFT, 1999b:96) e coeficiente de correlação cofenética (SNEATH e SOKAL, 1973:278). Este último coeficiente foi obtido no R, usando a rotina cophenetic.

O coeficiente de agrupamento (ca) e o coeficiente divisivo (cd) avaliam o nível de estrutura grupal detectado nos dados. O coeficiente de correlação cofenética (r) mede o grau de concordância entre os valores de semelhança implícitos no dendrograma e os da matriz inicial de semelhança, sendo só aplicável aos procedimentos aglomerativos.

É importante anotar que também não há classificações absolutas das classificações proporcionadas pela análise de agrupamento. Nem todos os investigadores aceitam o r e a biblioteca cluster, comum ao S e ao R, não se lhe refere. Incluem o ca e o cd. Por isso utilizo simultaneamente ca, cd e r. Sobre as dificuldades e sérias limitações da avaliação dos resultados da aplicação dos métodos da análise de agrupamento veja-se ALDENDERFER e BLASHFIELD (1987:62-74).

O procedimento analítico anteriormente explanado é resumido e generalizado no algoritmo BARCOR, assim descrito:

1. Aplicar o método TOPSIS aos dados, depois de os submeter às transformações que se justifiquem.

2. Seleccionar a proximidade relativa ao arquétipo que for mais conveniente.

3. Usar a proximidade relativa seleccionada como variável dependente e os dados utilizados no TOPSIS como variáveis independentes e aplicar um procedimento para estabelecer o "melhor" modelo linear, como, por exemplo, a rotina "stepwise" (S-Plus; usa o Cp) ou a "step" (S-Plus e R; utiliza o AIC).

4. Utilizar as variáveis do modelo

anterior como atributos na aplicação do TOPSIS.

5. Usar a informação gerada no ponto três, e pela nova aplicação do TOPSIS como dados para algoritmos da análise de agrupamento e para estabelecer um modelo de árvore regressiva. Eventualmente, recorrer a um método de visualização (e.g., análise das componentes principais) para fins interpretativos.

6. Usar, de modo integrado, toda a informação gerada anteriormente, voltar a aplicar o ponto 3, e verificar se é possível reduzir eficientemente o número de variáveis, repetindo para isso, também, os pontos 4 e 5.

7. Interpretar os resultados obtidos e escolher a classificação que tenha o melhor desempenho à luz dos critérios de avaliação que se adoptarem

Aplicação do algoritmo BARCOR e discussão

A matriz de valores do quadro 5, dos referidos autores, foi previamente sujeita a uma "normalização". Ocorrem nelas variáveis que quanto mais elevado é o seu valor, pior é a qualidade da cortiça ("custos"). Recorremos à normalização dos "custos" e "benefícios" de acordo com o procedimento descrito em YOON e HWANG (1995:16), de uso muito generalizado no âmbito da investigação operacional. Os valores das colunas dos "benefícios" são substituídos pela razão "valor da célula/máximo da coluna". Os valores das colunas dos "custos" são substituídos pela razão "mínimo da coluna/valor da célula". Esta nova matriz serviu de entrada para os algoritmos utilizados.

A esta matriz apliquei o algoritmo TOPSIS (aplicação T10/1). Os valores da proximidade relativa ao melhor arquétipo e a identificação dos provetes exibem-se no quadro 3. A identificação dos provetes é a de PESTANA *et al.*, (2005b). A cada um é atribuído um número pelo qual passam daqui em diante a ser referenciados. Os provetes estão ordenados de acordo com os três grupos de qualidade (má, média e boa) definidos pelos mesmos autores, separados por linhas a cheio.

A matriz dos dados iniciais também pode ser normalizada multiplicando os "custos" por -1. A esta matriz também apliquei o TOPSIS (aplicação T10/0). Estes resultados também se inserem no quadro 3.

Assumindo as mesmas classes qualitativas de PESTANA *et al.* (2005b), comparando a sua classificação, em termos grupais, com a T10/0, vemos que três pares de provetes trocam de posição entre as qualidades má e média, um par entre a boa e a média. A dissemelhança é de 21%, o que atendendo à diferença dos métodos utilizados é aceitável.

O quadro 3 ilustra o que já se afirmou: os resultados da aplicação de algoritmos classificatórios, ditos objectivos, dependem destes e do tratamento pré analítico que os dados requeiram.

Submeti as proximidades da coluna T10/1 (variável dependente), e as variáveis tratadas à rotina "stepwise" do S-Plus, obtendo a saída final descrita pela eq. (1). Este modelo serve de ponto de partida para ensaiar várias alternativas de redução do número de atributos.

$$y=-0{,}1303 + 0{,}1028\, x_1 + 0{,}0801\, x_2 + 0{,}0407\, x_5 + 0{,}0874\, x_6 + 0{,}2172\, x_7 + 0{,}1406\, x_8 + 0{,}4080\, x_9 \qquad (1)$$

$(R^2_{aj}=0{,}996,\ F=1248,\ p=0,\ s=0{,}01034)$

No quadro 4, apresento os resultados da aplicação do TOPSIS aos conjuntos C0, C1, C2, C3, do método UPGMA aos

conjuntos anteriores (U/C0, U/C1, U/C2 e U/C3), e do algoritmo divisivo hierarquizado de agrupamento ao C3 (D/C3). Quanto se pretende um pequeno número de classes, neste caso três (má, média e boa) justifica-se ensaiar um algoritmo deste último tipo (e.g.,VENABLES e RIPLEY, 2002:317). Utilizei os respectivos dendrogramas (figuras 1 a 5) das cinco aplicações de algoritmos da análise de agrupamento para definir os três grupos qualitativos, certamente que com alguma subjectividade, mas esta não é totalmente evitável, em qualquer circunstância.

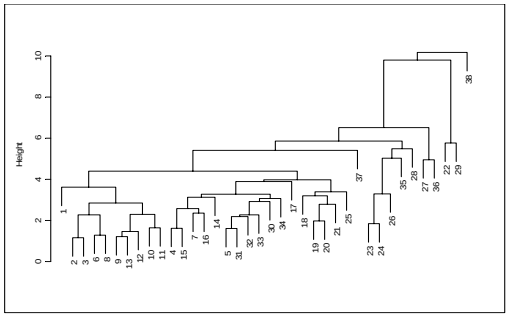


Figura 1 - Dendrograma gerado pela aplicação U/C0

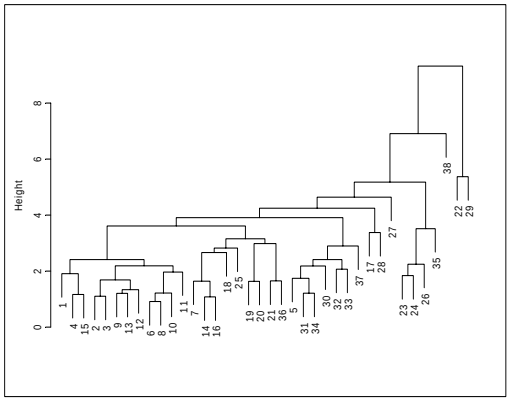


Figura 2 - Dendrograma gerado pela aplicação U/C1

No quadro 5, insiro os avaliadores do desempenho da aplicação dos algoritmos de agrupamento. Cumpri os passos 1, 2, e 3 do algoritmo BARCOR.

Por uma questão de completude da comparabilidade com PESTANA *et al.* (2005b), empreendi também a análise das componentes principais aos conjuntos C0 e C3, cujos resultados se exibem no quadro 6. Por motivos que explano adiante, não apresento os gráficos das projecções dos provetes no espaço definido pelas componentes principais.

Antes de prosseguir vale a pena avaliar os resultados já alcançados.

Vem aqui à colação o expendido na secção Anotação Breve sobre Classificações.

O quadro 4 é agora o objecto da minha atenção. Os resultados T10/1, T/C1, T/C2, T/C3, U/C0, U/C1, U/C2, U/C3, D/C3 revelam um fundo de semelhança e coerência que convém assinalar. As classificações U/C0, U/C1, U/C2 e U/C3 têm mais semelhança com as T10/1, T/C1, T/C2, T/C3 do que com os resultados da aplicação UPGMA feita por PESTANA *et al.* (2005b). Relativamente à classificação destes autores, as maiores dissemelhanças ocorrem nas classes média e boa.

As quatro aplicações do TOPSIS são muito semelhantes, ocorrendo algumas simples trocas de posição de provetes contíguos, revelando assim baixa sensibilidade ao número de atributos. É aceitável aplicar o TOPSIS recorrendo ao conjunto C3. Todos os gráficos de dendrogramas, árvores e de resíduos foram obtidos no S-Plus, exceptuando o da figura 3, obtido no R 2.

O valor de r encontrado por PESTANA *et al.* (2005b) é de 0,74, sendo o valor mínimo por mim obtido 0,85. Justifica-se a minha preferência pelas

classificações que estabeleci.

Concentro-me agora na aplicação dos métodos de agrupamento, reportando-me ao quadro 5. Descer os atributos de dez para sete e seis melhorou tanto ca como r. As classificações aglomerativas são comparáveis à divisiva, que revela um cd ligeiramente superior ao máximo ca das U/Ci.

As diferenças entre as classificações U/C3 e D/C3 estão assinaladas com um asterisco no quadro 4, coluna D/C3 e são em número de seis (16%). Os provetes 7, 21 e 36 têm qualidade má na D/C3 e média na U/C3. O provete 17 é de qualidade média no método divisivo e boa no aglomerativo. O inverso acontece com os provetes 18 e 25.

Uma conjectura digna de registo e futura exploração: provavelmente, o uso do número óptimo de atributos poderá permitir obter classificações com diferentes métodos de agrupamento exibindo níveis altos de semelhança.

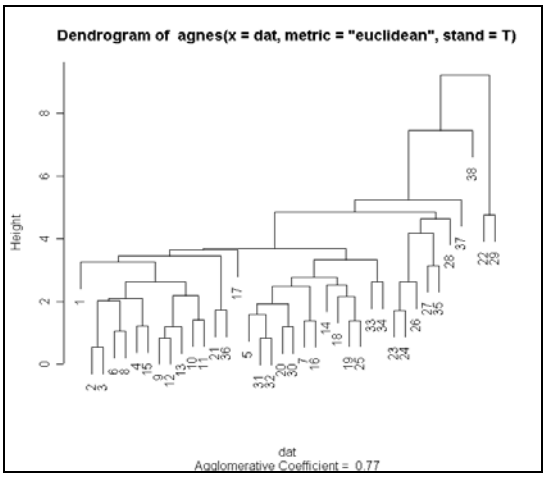


Figura 3 - Dendrograma gerado pela aplicação U/C2

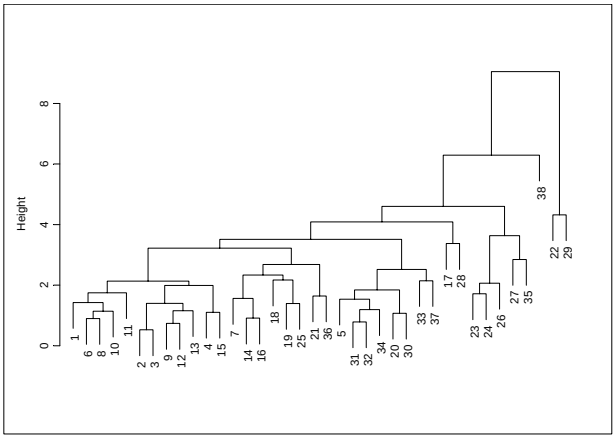


Figura 4 - Dendrograma gerado pela aplicação U/C3

Quadro 3 - Comparação da classificação de Pestana *et al.* (2005b) e os ordenamentos obtidos com o TOPSIS. Os asteriscos assinalam discrepâncias de classificação qualitativa, relativamente à obtida por Pestana *et al.* (2005b)

PESTANA *et al.*		T10/0		T10/1	
Provetes	**Número**	**Proximidade**	**Provete**	**Proximidade**	**Provete**
111B46	1	0,35300467	5	0,10294591	12
23A129	2	0,37342184	27*	0,10349049	10
23A130	3	0,43730907	4	0,11085242	11
116A33	4	0,44946552	2	0,11991388	9
36B135	5	0,51119526	10	0,12084248	13
23B131	6	0,51693133	11	0,13968393	1
23B132	7	0,52611649	9	0,14205514	3
37A210	8	0,52659503	12	0,14791679	6
320A239	9	0,52886507	1	0,14971339	2
320A312	10	0,58235631	20*	0,15254045	4
320A313	11	0,58497448	23*	0,15422573	8
320A414	12	0,59400742	13	0,17743234	37*
320A240	13	0,60143647	3	0,18111651	5
111C48	14	0,60484367	8*	0,18737105	15
116B44	15	0,62269072	21	0,18791361	31*
215A521	16	0,65660302	22	0,19689581	32*
215B523	17	0,67330812	6*	0,20524244	20
37A211	18	0,70365782	7*	0,20847031	30*
37B138	19	0,72091535	15	0,20920733	7*
116C21	20	0,73439972	18	0,214652	33*
116A32	21	0,73457294	24	0,21866969	34*
36A134	22	0,74734313	33*	0,23700578	17
36B136	23	0,75191483	26	0,25405138	16
116B45	24	0,76519723	16	0,26706377	36*
21B420	25	0,7667966	14	0,26927052	14
37B137	26	0,7688806	17	0,2799872	25
111C27	27	0,80663672	19	0,28804616	19
215A522	28	0,81082141	31	0,31593587	28
37A29	29	0,81168241	36	0,34071112	21*
21A315	30	0,82141965	34	0,3559997	27*
21A417	31	0,82661963	32	0,37500671	18
21A419	32	0,82719125	38	0,37920008	35*
21A316	33	0,8303518	30	0,40774267	38*
21A418	34	0,83333394	29	0,41145307	23*
117A125	35	0,83429605	25*	0,41157305	26*
117A126	36	0,8419685	35	0,44700608	24*
117A227	37	0,85514614	28	0,76053711	22
117A228	38	0,90553911	37	0,81826637	29*

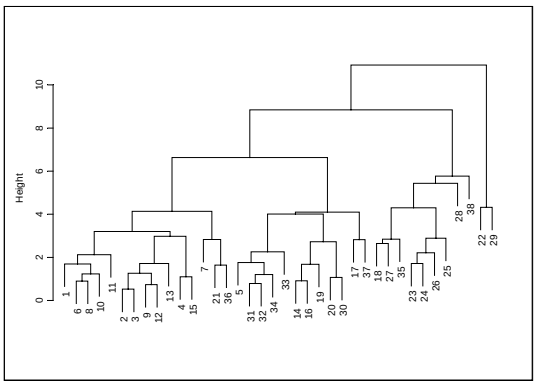


Figura 5 - Dendrograma gerado pela aplicação D/C3

Sucintamente, mas com verificações relevantes, reporto-me agora aos resultados da aplicação da análise das componentes principais (quadro 6). Destaco os seguintes aspectos:

- Verifica-se uma elevada concentração da variância total no primeiro eixo (56,5% e 60,8%).

- Anote-se o elevado valor da variância acumulada nos três primeiros eixos (80,0% e 85,6%).

- A redução do número de variáveis não alterou as características principais da aplicação da análise das componentes principais a C0. Por exemplo, nas duas análises, a área máxima dos poros de barriga (x_7) e os poros transversais (x_8) têm o maior controlo sobre a primeira componente, seguindo-se-lhe em importância os poros de barriga (x_6) e a área máxima dos poros transversais (x_9); Igualmente, o ângulo correspondente ao binário máximo (x_2) e a área máxima dos poros de barriga (x_7) controlam preponderantemente a segunda componente

Genericamente, sobre a aplicação da análise das componentes principais aos 10 atributos, podemos dizer que as variáveis de C3 (x_1, x_2, x_6, x_7, x_8, x_9) controlam preponderantemente a variância dos dois primeiros eixos, o que é coerente com a informação inserida na eq. (1), associada ao TOPSIS, e apoia o emprego deste método.

Dada a minha falta de familiarização com os provetes e rabanadas utilizados, não empreendo a análise (sempre eivada de alguma subjectividade) dos gráficos das projecções dos provetes no espaço definido pelas componentes principais, a qual, no contexto da amplitude da análise já apresentada, e a coerência global evidenciada, perde relevância.

Em PESTANA *et al.*, (2005b), a coerência do enquadramento e sustentação da classificação da qualidade dos provetes, em grande extensão, é encontrada nos resultados da classificação das rabanadas (PESTANA *et al.*, 2005a) que utiliza os mesmos procedimentos analíticos, pelo que ocorre aqui uma certa circularidade argumentativa, facto que no meu entendimento não afecta a valia e o interesse dos dois artigos destes autores.

PESTANA *et al.* (2005b), no seu quadro 11, inserem a classificação dos provetes usando o método da pontuação ponderada. Embora se trate de comparar classificações com ponderações diferentes dos atributos, assinalo que comparando a classificação proporcionada por D/C3 com esta, verifica-se uma dissemelhança de classificação de 63%.

Quadro 4 - Resumo de todas aplicações do método TOPSIS e da análise de agrupamento. A linha cheia separa as classes de qualidade (má, média, boa)

Referência		**Provetes Ordenados**										**UPGMA**			**Divisivo**	
Provetes	**Nº**	**T10/1**	**T/C1**	**T/C2**	**T/C3**	**T/C4**	**T/C5**	**U/C0**	**U/C1**	**U/C2**	**U/C3**	**U/C4**	**U/C5**	**D/C3**	**D/C4**	**D/C5**
111B46	1	12	12	10	10	1	10	1	1	1	1	1	1	1	1	1
23A129	2	10	10	12	12	12	1	2	4	2	6	11	6	6	11	6
23A130	3	11	11	11	11	10	12	3	15	3	8	6	8	8	2	8
116A33	4	9	9	9	9	11	9	6	2	6	10	10	10	10	9	10
36B135	5	13	13	13	13	9	11	8	3	8	11	8	11	11	6	2
23B131	6	1	1	3	1	13	2	9	9	4	2	2	36	2	10	3
23B132	7	3	3	1	3	3	13	13	13	15	3	3	2	3	8	9
37A210	8	6	6	6	2	4	3	12	12	9	9	9	3	9	3	12
320A239	9	2	2	2	6	2	4	10	6	12	12	12	9	12	12	4
320A312	10	4	4	8	8	6	6	11	8	13	13	13	12	13	4	15
320A313	11	8	8	4	4	8	8	4	10	10	4	4	13	4	15	11
320A414	12	37	37	5	37	5	5	15	11	11	15	15	4	15	13	36
320A240	13	5	5	37	5	30	30	7	7	21	7	7	15	7*	28	13
111C48	14	15	15	15	15	37	20	16	14	36	14	16	7	21*	7	28
116B44	15	31	31	31	31	31	37	14	16	17	16	25	14	36*	25	7
215A521	16	32	32	32	32	20	31	5	18	5	18	14	16	5	14	16
215B523	17	20	20	20	20	15	15	31	25	31	19	17	19	31	16	14
37A211	18	30	30	30	30	34	32	32	19	32	25	28	25	32	18	19
37B138	19	7	7	33	33	32	33	33	20	20	21	18	18	34	21	25
116C21	20	33	33	7	7	33	34	30	21	30	36	27	21	33	36	18
116A32	21	34	34	34	34	25	25	34	36	7	5	19	24	14	23	35
36A134	22	17	17	17	17	7	17	17	5	16	31	20	27	16	26	21
36B136	23	16	16	16	16	17	7	18	31	14	32	21	35	19	24	24
116B45	24	36	36	36	14	28	28	19	34	18	34	36	23	20	35	23
21B420	25	14	14	14	36	38	19	20	30	19	20	5	26	30	5	26
37B137	26	25	25	25	19	19	38	21	32	25	30	31	17	17*	31	5
111C27	27	19	19	19	25	36	36	25	33	33	33	34	28	37	34	20
215A522	28	28	28	28	28	16	16	37	37	34	37	30	5	18*	30	30
37A29	29	21	21	21	21	14	14	23	17	23	17	32	37	25*	32	31
21A315	30	27	27	27	27	35	27	24	28	24	28	33	33	35	33	32
21A417	31	18	18	18	18	27	35	26	27	26	23	37	38	23	37	34
21A419	32	35	35	35	35	21	21	35	23	27	24	23	20	24	19	27
21A316	33	38	38	38	38	18	18	28	24	35	26	26	30	26	20	17
21A418	34	23	26	23	26	23	23	27	26	28	27	24	31	27	27	33
117A125	35	26	23	26	23	26	26	36	35	37	35	35	32	28	17	37
117A126	36	24	24	24	24	24	24	22	38	38	38	38	34	38	38	38
117A227	37	22	22	22	22	29	29	29	22	22	22	22	22	22	22	22
117A228	38	29	29	29	29	22	22	38	29	29	29	29	29	29	29	29

Quadro 5 - Índices do desempenho dos métodos da análise de agrupamento. Aplicações a conjuntos que contém o calibre (x_3) estão assinaladas com um asterisco; na = não aplicável

	U/C0*	**U/C1***	**U/C2**	**U/C3**	**U/C4***	**U/C5**	**D/C3**	**D/C4***	**D/C5**
ca	0,71	0,77	0,77	0,80	0,80	0,85	cd=0,82	cd=0,83	cd=0,87
r	0,85	0,88	0,87	0,87	0,88	0,86	na	na	na

Quadro 6 - Análise das componentes principais (ACP) aplicada a C0 e C3. Correlações entre as variáveis e as três primeiras componentes principais. DP = desvio padrão; VT = variância total; VA = variância acumulada

ACP com 10 Variáveis				ACP com 6 Variáveis			
Variáveis	Componentes Principais			Variáveis	Componentes Principais		
	1	2	3		1	2	3
x_1	---	0,301	-0,445	x_1		-0,322	0,519
x_2		0,456	-0,506	x_2		-0,505	-0,404
x_3			0,296	x_6	0,427	-0,520	
x_4	-0,291	0,165	-0,284	x_7	0,605	0,419	-0,468
x_5		0,206	0,292	x_8	0,548	-0,242	
x_6	-0,421	0,445		x_9	0,383	0,371	0,582
x_7	-0,562	-0,471	-0,322	DP	0,366	0,199	0,123
x_8	-0,522	0,170	0,302	VT	0,608	0,179	0,068
x_9	-0,348	-0,395	0,312	VA	0,608	0,788	0,856
x_{10}	0,108		-0,278				
DP	0,386	0,207	0,137				
VT	0,565	0,163	0,071				
VA	0,565	0,728	0,800				

Concluído este ponto da situação, intercalar, prossigo com o algoritmo BARCOR, pois os resultados obtidos até agora não o desaconselham.

Seleccionei quatro variáveis de C3 (x_1, x_2, x_7, x_9) e o resultado da aplicação T/C3 (variável dependente, y) para ajustar o modelo linear seguinte:

$$y=-0,2118 + 0,2203\, x_1 + 0,1675\, x_2 + 0,3584\, x_7 + 0,4479\, x_9 \qquad (2)$$

(R^2_{aj}=0,953, F=189.5, p=0, s=0,03631)

Este ajustamento conduziu-me aos conjuntos C4 e C5, atrás descritos. Apliquei a estes conjuntos o TOPSIS (T/C4, T/C5) e os métodos de agrupamento já utilizados (U/C4, U/C5, D/C4, D/C5), como também se regista no quadro 4. Os respectivos índices de desempenho apresentam-se no quadro 5, e os dendrogramas nas figuras 6 a 9.

No quadro 7, apresento os resultados da aplicação da análise das componentes principais a C4 e C5.

Debrucemo-nos sobre os resultados obtidos, voltando ao quadro 4.

A inclusão destas novas quatro classificações não aumentou o nível de dissemelhança inter classificações de forma conspícua, mantendo-se um fundo de homogeneidade entre elas.

Os índices de desempenho do quadro 5, mostram que a estrutura grupal dos dados é mais marcada para as classificações de C5 (ca=0,85 e cd=0,87).

A classificação U/C4 mostrou-se mais eficiente que a U/C1. Iguala o valor de r (0,88) da última, com menos dois atributos, e simultaneamente maior valor de ca. As variáveis x_6 e x_8 revelam-se supérfluas. Como os calibres para rolha não têm grande amplitude de variação, as U/C4 e U/C5 evidenciam desempenho comparável quanto a r (0,88 *versus* 0,86), revelando a última o maior valor de ca (0,85 *versus* 0,80).

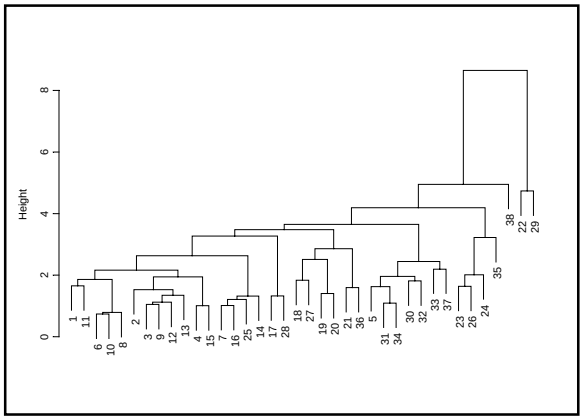


Figura 6 - Dendrograma gerado pela aplicação U/C4

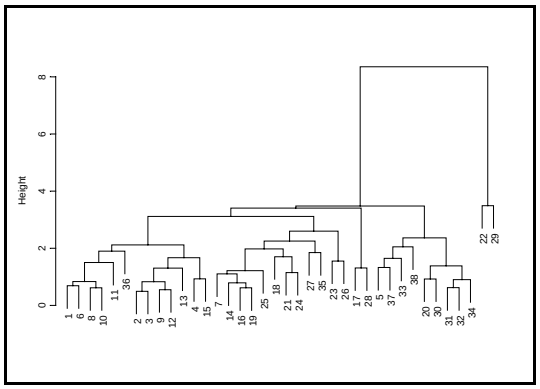


Figura 7 - Dendrograma gerado pela aplicação U/C5

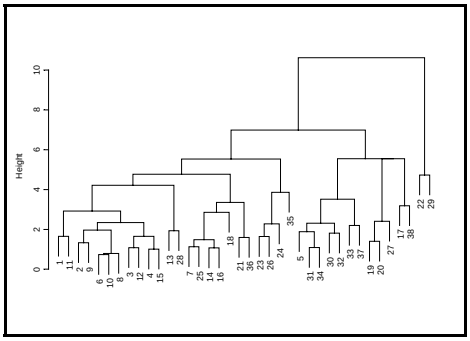


Figura 8 - Dendrograma gerado pela aplicação D/C4

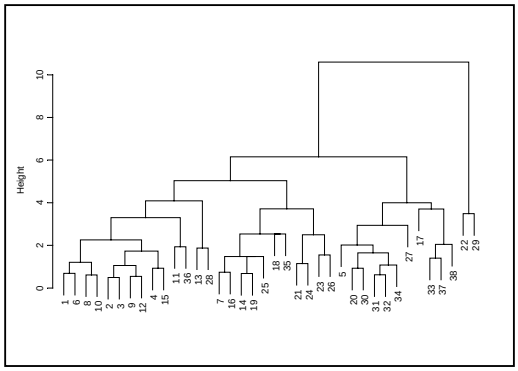


Figura 9 - Dendrograma gerado pela aplicação D/C5

Quadro 7 - Análise das componentes principais (ACP) aplicada a C4 e C5. Correlações entre as variáveis e as três primeiras componentes principais. DP = desvio padrão; VT = variância total; VA = variância acumulada

ACP Aplicada ao C4				ACP Aplicada ao C5			
Variáveis	**Componentes Principais**			**Variáveis**	**Componentes Principais**		
	1	**2**	**3**		**1**	**2**	**3**
x_1		-0,412	-0,539	x_1		0,398	-0,543
x_2		-0,906	0,238	x_2		0,915	0,192
x_3			-0,277	x_7	0,829		0,453
x_7	0,828		0,431	x_9	0,557		-0,680
x_9	0,558		-0,625	DP	0,289	0,152	0,123
DP	0,289	0,152	0,125	VT	0,634	0,174	0,114
VT	0,597	0,165	0,112	VA	0,634	0,808	0,923
VA	0,597	0,761	0,873				

A descida de seis para quatro atributos melhorou bastante as classificações divisivas. A D/C5 revela os melhores valores dos dois índices cd=0,87.

Exclusivamente pelo critério da estrutura grupal detectada, a D/C5 é a melhor classificação.

Só atendendo ao valor de r, as U/C1 e U/C4 são as melhores.

De todas as classificações aglomerativas, a U/C5 revela-se a mais equilibrada.

Relativamente ao quadro 7, o padrão do quadro 6 repete-se, mais marcadamente. A aplicação da análise das componentes principais ao C5 sustenta a aplicação do TOPSIS e confirma a eq. (3), a introduzir adiante.

Análise discriminante multivariada linear

Com o propósito de alargar a apreciação da classificação da cortiça recorrendo só a quatro variáveis, apliquei à classificação D/C5, a análise discriminante multivariada linear. Para isso recorri ao S-Plus.

Devido à extensão máxima que o artigo não pode violar, resumo a informação referente a esta análise nos seguintes termos: a) os grupos de

provetes das três qualidades mostram separação nítida, tanto pelo valor de Λ de Wilks (0,0876), como no teste T^2 de Hotelling, das diferenças das médias dos três grupos; b) As classificações preditiva (erro global=0) e de validação cruzada (erro global = 0,077) não revelam nenhum provete mal classificado.

Os resultados da análise discriminante multivariada linear confirmam a possibilidade de se classificar a cortiça, para o fabrico de rolhas, recorrendo só a quatro atributos, e não infirmam a predicabilidade da eq. (3), a apresentar na secção seguinte.

Modelo linear de predição

Usei o resultado da aplicação T/C4 (variável dependente, y) para ajustar o seguinte modelo linear:

$$y= -0{,}1281 + 0{,}1359\, x_1 + 0{,}1162\, x_2 + 0{,}3025\, x_7 + 0{,}5325\, x_9 \qquad (3)$$

$(R^2_{aj}=0{,}994,\ F=1588,\ p=0,\ s=0{,}01254)$

Na figura 10, apresento os erros percentuais das estimativas obtidas com a eq. (3). Só ocorrem seis erros superiores a 10%, em valor absoluto.

Dada a informação da caixa 4, e da figura conexa, é meu entendimento que a eq.(3) pode ser usada, preliminar e expeditamente, para se obter um ordenamento relativo das amostras de cortiça para rolha, que se queiram classificar. Fica em aberto indagar da possibilidade de a aplicar, individualmente, a cada uma das outras classes de calibre, não utilizadas na produção de rolhas (NP 298, 1993; ISSO 1216, 1990).

Uma estrutura grupal para os ordenamentos do TOPSIS

Por força da forma como se comercia a cortiça para rolha, pode existir interesse em associar, aos ordenamentos quantitativos proporcionados pelo TOPSIS, uma escala classificativa nominal (estrutura grupal). Isto pode ser feito de, pelo menos, duas formas:

A primeira. As proximidades relativas (y) que o método estabelece podem ser trivialmente transformadas numa escala 0-1. Nesta, podem-se estabelecer os limites das classes que se quiserem. Por exemplo, à classe boa correspondem os valores de $y>0{,}667$; na média teremos $0{,}333<y\leq 0{,}667$; à má confinam-se os valores de $y\leq 0{,}333$.

Outro procedimento consiste em extrair do ordenamento um modelo de árvore (MATHSOFT, 1999a: cap. 12; VENABLES e RIPLEY, 2002: cap. 9). Para as provas atinentes aos modelos de árvore ver RIPLEY (1996: cap. 7).

Recorrendo ao S-Plus (biblioteca tree), a título ilustrativo, estabeleci um modelo de árvore regressiva a partir dos dados utilizados para ajustar a eq. (3). A árvore detectada ilustra-se na figura 11. A discrepância média residual ("residual mean deviance") é igual a 0.007959 (= 0.2626 / 33).

Atente-se que para estabelecer a árvore, o algoritmo empregue só utilizou as variáveis, por ordem, x_7, x_9 e x_2. A constituição dos grupos finais insere-se no quadro 8.

No quadro 8, embora as designações sejam de escolha arbitrária, os grupos reflectem, nas suas linhas gerais, o ordenamento do quadro 4.

Na análise dos resíduos, não detectei nenhum padrão na dispersão dos resíduos. Na figura 12, os "outliers" são os provetes do grupo 3. Em baixo situam-se os provetes 23, 26 e 24; em cima, os provetes 29 e 22.

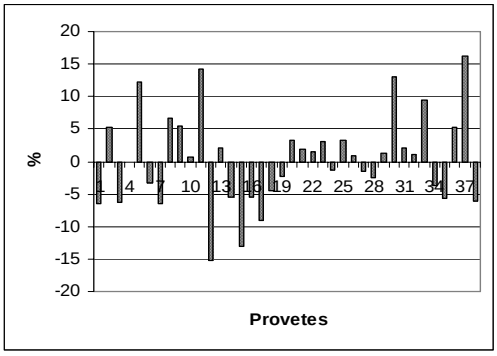
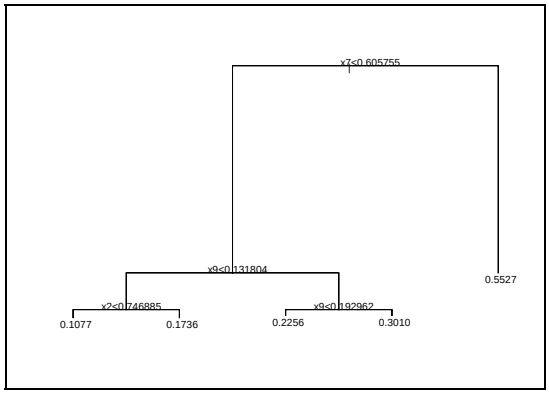


Figura 10 - Erros percentuais associados às estimativas do modelo da eq. (3)



Grupos	8	9	10	11	3

Figura 11 - Árvore extraída do ordenamento T/C4

Quadro 8 - Grupos de provetes detectados pelo modelo da árvore, no ordenamento T/C4

Grupos	Provetes Incluídos	Total	Classificação
8	1, 2, 3, 4, 6, 8, 9, 10, 11, 12, 13	11	Má inferior
9	5, 20, 25, 30, 31, 32, 33, 34, 37, 38	10	Má superior
10	7, 14, 15, 16, 17, 19, 28	7	Média inferior
11	18, 21, 27, 35, 36	5	Média superior
3	22, 23, 24, 26, 29	5	Boa

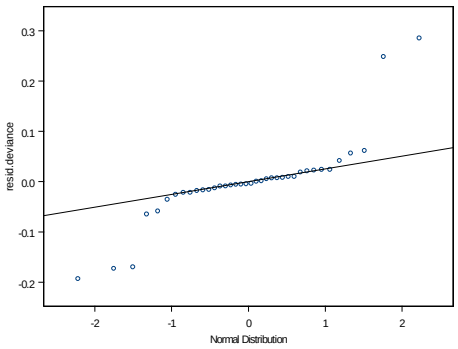


Figura 12 - Distribuição normal acumulada dos resíduos do modelo da árvore da figura 11

É possível extrair mais informação da árvore, mas para o meu propósito ilustrativo, tenho a exposta como suficiente

PEREIRA (1998) mostra a importância dominante do calibre e porosidade da cortiça, na sua valorização para produção de rolhas. Nesta perspectiva, pondero, arbitrariamente e só para propósito ilustrativo, o calibre (x_3) com 0,41, cada uma das áreas máximas dos poros (x_7 e x_9) com 0,19, e as outras variáveis com 0,03, isto é, ao calibre atribuo 41% da ponderação, à porosidade máxima 38%, e aos outros sete atributos 21%. Aplicando agora o TOPSIS e o procedimento descrito, é possível reduzir o número de variáveis relevantes a três: x_3, x_7 e x_9. De acordo com o algoritmo BARCOR, voltei a aplicar o TOPSIS só considerando estas três variáveis com os pesos, respectivamente, de 0,52, 0,24, 0,24. Voltando a ajustar um modelo de árvore detecto os grupos inseridos no quadro 9, e a árvore da figura 13, cuja discrepância média residual é igual a 0.01085 (=0.3579/33). As distribuições dos resíduos, dos dois modelos de árvore estabelecidos, inserem-se no quadro 10.

A relação entre a qualidade da cortiça (proximidade relativa ao melhor arquétipo – y) e as três variáveis supra mencionadas pode ser expressa pelo seguinte modelo:

$$y = -0{,}3155 + 0{,}4520x_3 + 0{,}8686\, x_7 + \\ + 0{,}5876\, x_9 - 0{,}7231\, x_3x_7 - 0{,}1617\, x_3x_9 - \\ - 0{,}4406\, x_7x_9 + 0{,}7091\, x_3x_7x_9 \qquad (4)$$

(R^2=0,9989, F=4062, p=0, s=0,0063)

Considerando os valores absolutos dos desvios, na predição dos valores de y com a eq. (4), ocorrem 5 desvios maiores que 10% (13%), e 4 entre 5 e 10% (10%).

Comparando as árvores das figuras 11 e 13, verifica-se que a estrutura divisiva e grupal básica são idênticas. As diferenças entre os agrupamentos dos quadros 8 e 9, são impostas pela variável calibre (x_3), a não uniformidade de ponderação dos três atributos, e à ausência de x_1 e x_2, pois as variáveis da porosidade (x_7 e x_9) mantêm-se.

Usando o comando prune.tree, da biblioteca tree, posso "podar", da melhor maneira, a árvore da figura 13, para obter o modelo Arv3 só com os três grupos finais seguintes:

Qualidade má: 1, 2, 3, 4, 5, 6, 9, 10, 12, 13, 15, 20, 30, 31, 32, 33, 34, 37 (total:18)

Qualidade média: 7, 8, 11, 14, 16, 17, 19, 25, 27, 28, 35, 36, 39 (total: 13)

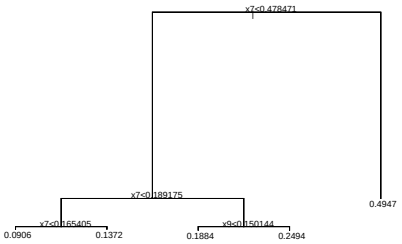
Qualidade boa: 18, 21, 22, 23, 24, 26, 29 (total: 7)

Esta árvore tem uma discrepância média residual igual a 0,01082 (=0,3786/35), e só a variável x_7 é usada nas partições. Insiro o gráfico da distribuição dos provetes pelo espaço do índice da qualidade da cortiça (y) e da variável x_7, na figura 14. Nesta, pode-se verificar que os provetes da qualidade boa se destacam das nuvens dos pontos das qualidades má e média. No quadro 10, apresento a distribuição dos resíduos da árvore em análise.

Comparando o agrupamento dos provetes, antes e depois da "poda", verifica-se que a classe má funde-se numa só, o mesmo acontece com a média, e a boa mantém a sua identidade. Quando se sujeita a uma "poda" idêntica (obter três grupos finais) a árvore descrita no quadro 8 e figura 11, ocorre o mesmo padrão de concentração de grupos (má: 21 provetes; média: 12; boa: 5), dado ter uma ramificação idêntica.

Limito-me a assinalar que emerge, aqui, a questão da ponderação dos atributos da qualidade da cortiça (e.g., BARRETO, 1986).

Um modelo de árvore existente pode servir para classificar outros conjuntos de dados, referentes a objectos similares.



Grupos	8	9	10	11	3

Figura 13 - Árvore detectada na aplicação, com ponderação desigual, do TOPSIS descrita no texto e recorrendo às variáveis x_3, x_7 e x_9

Quadro 9 - Grupos de provetes detectados pelo modelo da árvore, na aplicação, com ponderação desigual, do TOPSIS descrita no texto

Grupos	Provetes Incluídos	Total	Classificação
8	1, 2, 3, 4, 5, 9, 10, 12, 30, 33, 34 37	12	Má inferior
9	6, 13, 15, 20, 21, 32	6	Má superior
10	7, 8, 11, 19, 25, 38	6	Média inferior
11	14, 16, 17, 27, 28, 35, 36	7	Média superior
3	18, 21, 22, 23, 24, 26, 29	7	Boa

Quadro 10 - Distribuição dos resíduos e valores da discrepância média residual (dmr) do estabelecimento das árvores das figuras 11 (Arv1) e 13 (Arv2) e do modelo Arv3

Modelo	Resíduos						dmr
	Mínimo.	1º Qu.	Mediana	Média	3º Qu.	Max.	
Arv1	-0,1562	-0,03718	-0,00005498	-4,017e-018	0,01378	0,3517	0,007959
Arv2	-0,1566	-0,03735	0,004344	4,017e-018	0,01435	0,3572	0,01085
Arv3	-0,1566	-0,03504	-0,002151	1,424e-017	0,03742	0,3572	0,01082

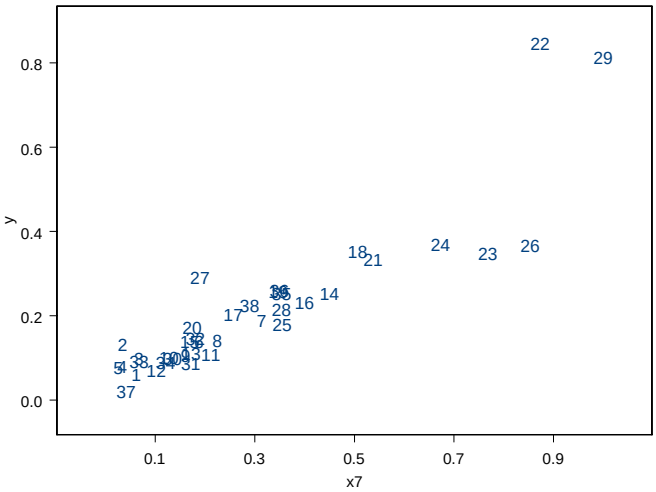


Figura 14 - Distribuição dos provetes pelo espaço do índice da qualidade da cortiça (y) e da variável x_7

Os modelos de árvore têm uma longa tradição em ecologia. Remontam à análise das associações (e.g., BARRETO, 1973), que foi o primeiro método de classificação de uso dominante, nesta disciplina.

Comentários conclusivos

Entendo justificado encerrar este texto com as seguintes asserções:

1. Os resultados que obtive aplicando diversos instrumentos de análise revelam forte coerência, e por isso não detectei nenhuma infirmação cruzada relevante.

2. Ao lidar-se com objectos que sejam descritos por atributos que representem quer "custos" quer "benefícios", é preciso ser cuidadoso na normalização prévia da matriz dos dados. A que utilizei, e descrevi, revelou-se eficaz por introduzir menor distorção na estrutura grupal da amostra (ver quadro 5) que a utilizada por PESTANA *et al.* (2005ab).

3. O algoritmo BARCOR mostrou-se eficiente e tem potencial para ser aplicado a situações formalmente semelhantes.

4. Mostrei que se podem classificar eficientemente amostras de cortiça para o

fabrico de rolhas utilizando só quatro ou cinco atributos (C5 e C4), a saber: força correspondente ao binário máximo (x_1), ângulo correspondente ao binário máximo (x_2), calibre (x_3), área máxima dos poros de barriga (x_7) e área máxima dos poros transversais (x_9). Ponderando os atributos e recorrendo ao algoritmo proposto, este número pode ser reduzido a três (x_3, x_7 e x_9).

5. Com o software de que hoje se dispõe e o actual desempenho dos computadores é possível aplicar várias rotinas de agrupamento e visualização em poucos minutos, se a amostra não for exageradamente grande (o caso analisado). No entanto, afigura-se-me que a informação aqui apresentada permite ser parcimonioso e eficiente, privilegiando a obtenção de resultados fáceis de interpretar, e sem perder informação da estrutura grupal dos dados. Admito que isto pode ser alcançado usando o método TOPSIS, eventualmente complementado, por um modelo de árvore.

6. A cortiça é um material biológico por isso a sua variação ocorre dentro de certos limites. Sugiro que especialistas desta área estabeleçam os valores máximos e mínimos possíveis de ocorrer, das diversas características, e estes sirvam para normalizar os arquétipos "melhor" e "pior" cortiça, a utilizar nas aplicações do TOPSIS ao ordenamento relativo de amostras de cortiça. Nestas circunstâncias, se também se pretender empregar uma ponderação não uniforme dos atributos sugiro o emprego do meu algoritmo COMPTO (BARRETO, 1994:87-96b). O *Software Umbela e Multat,* que divulguei em 2002, contem uma aplicação em Visual Basic 6, para correr o COMPTO.

7. A uma aplicação normalizada, no sentido do parágrafo anterior, do TOPSIS ou COMPTO, podemos aplicar um modelo de árvore e ficar a dispor de um instrumento para classificar outros conjuntos supervenientes de amostras de cortiça. Isto podia ser feito de uma maneira institucional e o modelo de árvore final seria distribuído por todos os industriais da cortiça. Passava-se a ter um modo único de classificação, e assim a obter classificações com maior comparabilidade, em todo o sector corticeiro.

8. Numa perspectiva de bom senso e abordagem mais chã, os conjuntos C4 e C5, formados pelas varáveis acima nomeadas, significam que quando pomos uma rolha numa garrafa o que é importante é a quantidade de cortiça que lá fica (x_1, x_2, x_3) ou não fica (x_7 e x_9). "Estar ou não estar cortiça, eis a questão". Com toda a simplicidade.

9. Fica por realizar a análise da possibilidade da actual prática comercial e industrial do sector corticeiro permitir acolher, fácil e generalizadamente, métodos de classificação mais exigentes que o "Critério da Pontuação Ponderada", mas, declaradamente, esta tarefa não é para mim.

Bibliografia

ALDENDERFER, M.S., BLASHFIELD, R.K., 1987. *Cluster Analysis*. Sage Publications, Inc., Thousand Oaks, California.

BARRETO, L.S., 1973. Análise das Associações da Floresta Aberta de Angola. *Revista de Ciências Biológicas*, Vol. 6, Série A:85-96, Universidade de Lourenço Marques.

BARRETO, L.S., 1986. *Ponderação de Atributos.* Centro de Estudos Florestais, INIC, Lisboa.

BARRETO, L.S., 1994. *Alto Fuste Regular. Instrumentos para a sua Gestão.* Publicações Ciência e Vida. Lda., Lisboa.

BARRETO, L.S., 2002. *Software Umbela e Multat* Departamento de Engenharia Florestal, Instituto Superior de Agronomia, Lisboa. Divulgado em disquete.

BARRETO, L.S., 2006. *O Algoritmo BARCOR: Classificação de Cortiça para Rolhas Recorrendo a Quatro Atributos de Qualidade*. "Preprint" de divulgação restrita.

DIGBY, P.G.N., KEMPTON, R.A., 1987. *Multivariate Analysis of Ecological Communities*. Chapman and Hall, London.

GIL, F., 2000. Classificações. Em *Enciclopédia Enaudi*. Volume 41. Imprensa Nacional – Casa da Moeda, Lisboa. pp. 90-110.

MAHNER, M., BUNGE, M., 1997. *Foundations of Biophilosophy*. Springer-Verlag, Berlin.

MATHSOFT, Inc., 1999a. *S-Plus 2000. Guide to Statistics,* Volume 1. Seatle.

MATHSOFT, Inc., 1999b. *S-Plus 2000. Guide to Statistics,* Volume 2. Seatle.

PEREIRA, H., 1998. O Calibre e a Porosidade como Parâmetros da Qualidade Tecnológica da Cortiça em Prancha. *Revista Florestal* **11**(2): 46-50.

PESTANA, M, LOUZADA, J., SARDINHA, R., 2005a. Classificação das Cortiças para a Produção de Rolha, Recorrendo a um Critério Objectivo. Parte I – As Rabanadas. *Silva Lusitana* **13**(1): 51-73.

PESTANA, M, LOUZADA, J., SARDINHA, R., 2005b. Classificação das Cortiças para a Produção de Rolha, Recorrendo a um Critério Objectivo. Parte II – Os Provetes (Os Quadros). *Silva Lusitana* **13**(2): 181-201.

RIPLEY, B.D., 1996.*Pattern Recognition and Neural Networks.* Cambridge University Press, Cambridge. Conforme citação em Venables e Ripley (2002:251).

SNEATH, P.H.A., SOKAL, R.R., 1973. *Numerical Taxonomy. The Principles and Practice on Numerical Classification*. W. H. Freeman and Company, San Francisco.

VENABLES, W.N., RIPLEY, B.D., 2002. *Modern Applied Statistics with S.* Fourth Edition. Springer, Berlin.

YOON, K.P., HWANG, C-L., 1995. *Multiple Attribute Decision Making. An Introduction.* Sage Publications, Inc., Thousand Oaks, California.

Entregue para publicação em Fevereiro de 2007
Aceite para publicação em Julho de 2007