

ABORDAGEM MULTIVARIADA À TRILOGIA SOLO-VINHA-VINHO. CASO DA CASTA VINHÃO.

Ana Rita Silva¹, Ana Cristina Braga²,
Isabel Araújo³, Teresa Mota⁴, José Maria Oliveira⁵, Jorge Oliveira⁶

¹ AR Silva, aluna do mestrado Estatística de Sistemas, Universidade do Minho;
anaritavieirasilva@gmail.com

² AC Braga, DPS, Universidade do Minho; acb@dps.uminho.pt

³ Vinalia – Soluções de Biotecnologia para a Vitivinicultura, Lda;
Isabel.araujo@vinalia.com.pt

⁴ Quinta Campos de Lima; tmota@vinhoverde.pt

⁵ IBB – Institute for Biotechnology and Bioengineering, Centre of Biological
Engineering; jmoliveira@deb.uminho.pt

⁶ Sinergeo – Soluções Aplicadas em Geologia, Hidrogeologia e Ambiente, Lda.;
jorgeoliveira@sinergeo.pt



Universidade do Minho

XIX congresso
da Sociedade
Portuguesa de
Estatística

Nazaré
Portugal

28Set. - 10Out.

Índice

- **Introdução**
- **Parcela em estudo**
- **Variáveis em estudo**
- **Resultados**
 - **Análise Exploratória**
 - **Análise de Componentes Principais**
 - **Análise de *Clusters***
 - **Correlações**
- **Conclusão**
- **Bibliografia**

Introdução

- O Projecto Agrocontrol co-financiado pelo “ON.2 – O Novo Norte” e QREN através do FEDER, tem como **principal objectivo** identificar e estudar múltiplas variáveis que determinam o comportamento físico e químico dos solos, que por sua vez influenciam o desenvolvimento da videira, a qualidade final da uva e por conseguinte a qualidade dos vinhos.
- A partir dos dados obtidos em campo e laboratório, e recorrendo a técnicas de estatística multivariada e determinação de correlações, identificaram-se parâmetros do solo que influenciaram o desenvolvimento da planta em 2010 e a qualidade das uvas e respectivo vinho.

Introdução



Fig. 1 – Estação Vitivinícola Amândio Galhano (EVAG)

Parcela em estudo

A parcela em estudo situa-se na Estação Vitivinícola Amândio Galhano (EVAG), no concelho dos Arcos de Valdevez.

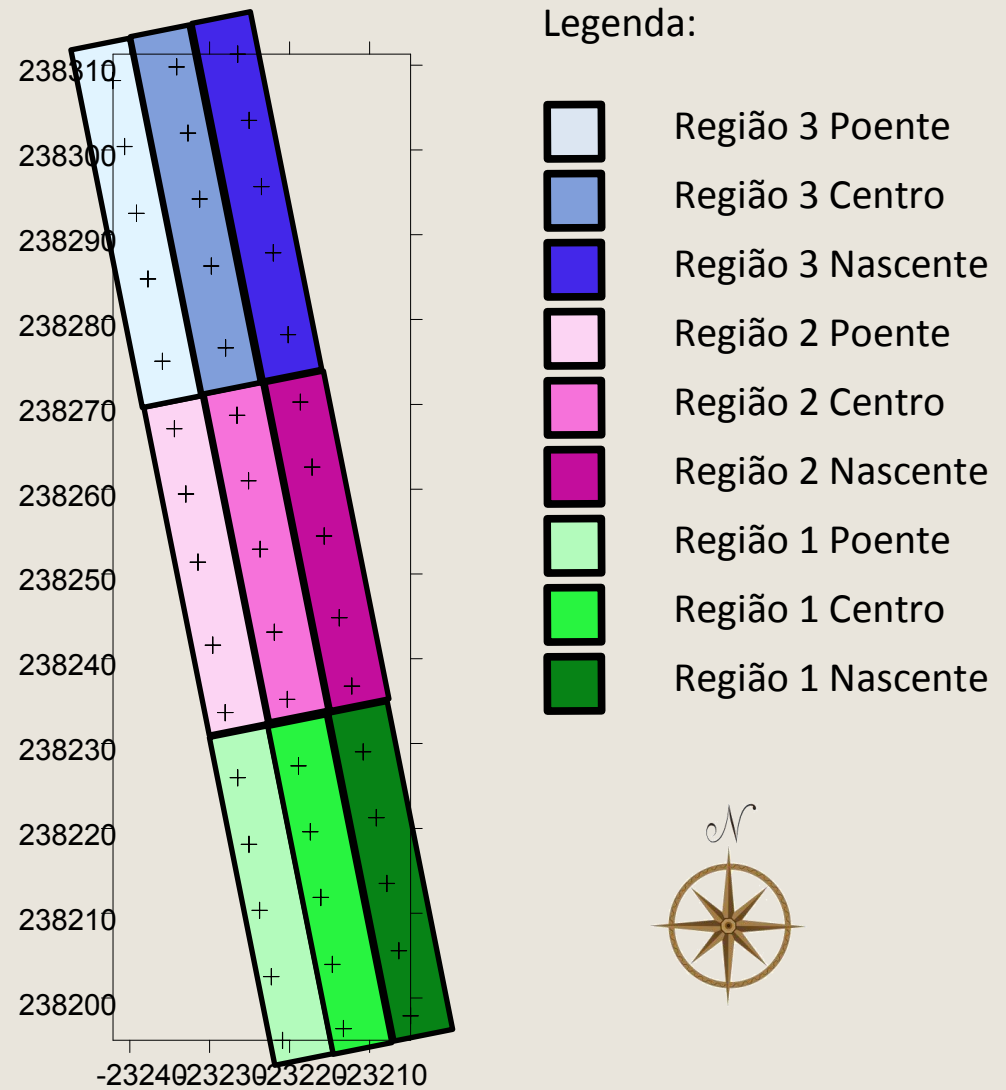


Fig. 2 - Parcela com localização dos 45 pontos georreferenciados

Variáveis em estudo

SOLO	pH; matéria orgânica (MO); densidade aparente (DA); fracção fina (FF); fracção grosseira (FG); fósforo (P2O5); potássio (K2O); cálcio (Ca); magnésio (Mg); azoto total (AzT); níquel (Ni); crómio (Cr); cádmio (Cd); nitratos (N); boro (B) e capacidade de troca catiónica (CTC)
VIDEIRA	Número de cachos por videira (Ncachos); peso médio do cacho por videira, em kg (Pcacho_kg) e peso de uvas por videira, em kg (uvas_kg_vid); número de varas por videira (Nvaras); peso das varas por videira, em kg (Pvaras_kg) e peso médio da vara por videira, em gramas (Pvara_g).
SUMO DE UVAS	Famílias de compostos do aroma na forma livre: compostos em C ₆ (FL1), álcoois (FL2), álcoois monoterpénicos (FL3), fenóis voláteis (FL4) e compostos carbonilados (FL5). Famílias de compostos do aroma na forma glicosilada: compostos em C ₆ (FG1), álcoois (FG2), álcoois monoterpénicos (FG3), óxidos e dióis monoterpénicos (FG4), norisoprenóides em C ₁₃ (FG5), fenóis voláteis (FG6) e compostos carbonilados (FG7).
MOSTO	pH do mosto (pH mosto); acidez total (Acidez_T); ácido málico (Acido_malico); ácido tartárico (Acido_tartarico); açúcares (Açucares) e teor de álcool provável (TAP).
VINHO	Avaliacao_Global e Nota_Final (limpidez_visual, cor_visual, limpidez_olfativa, intensidade_olfativa, qualidade_olfativa, qualidade_gustativa, limpidez_gustativa, intensidade_gustativa, persistência_gustativa e Avaliacao_Global)

Análise Exploratória

VALORES EM FALTA

Métodos utilizados no tratamento dos valores em falta:

- Método do vizinho mais próximo;
- Interpolação por *krigagem*;
- **Eliminação dos indivíduos com valores em falta.**

	Variáveis	Número de valores em falta	Ponto de amostragem	%
SOLO	DA	2	A42 e A45	4,44
	FF e FG	1	A14	2,22
VIDEIRA	Ncachos Peso_kg_vid Pcacho_kg nvaras pvaras_kg pvara_g	4	A13, A28, A29 e A30	8,89

Tab. 1 – Valores em falta das variáveis em estudo

Método seleccionado para a análise:

- **Eliminação dos indivíduos com valores em falta.**

Análise Exploratória

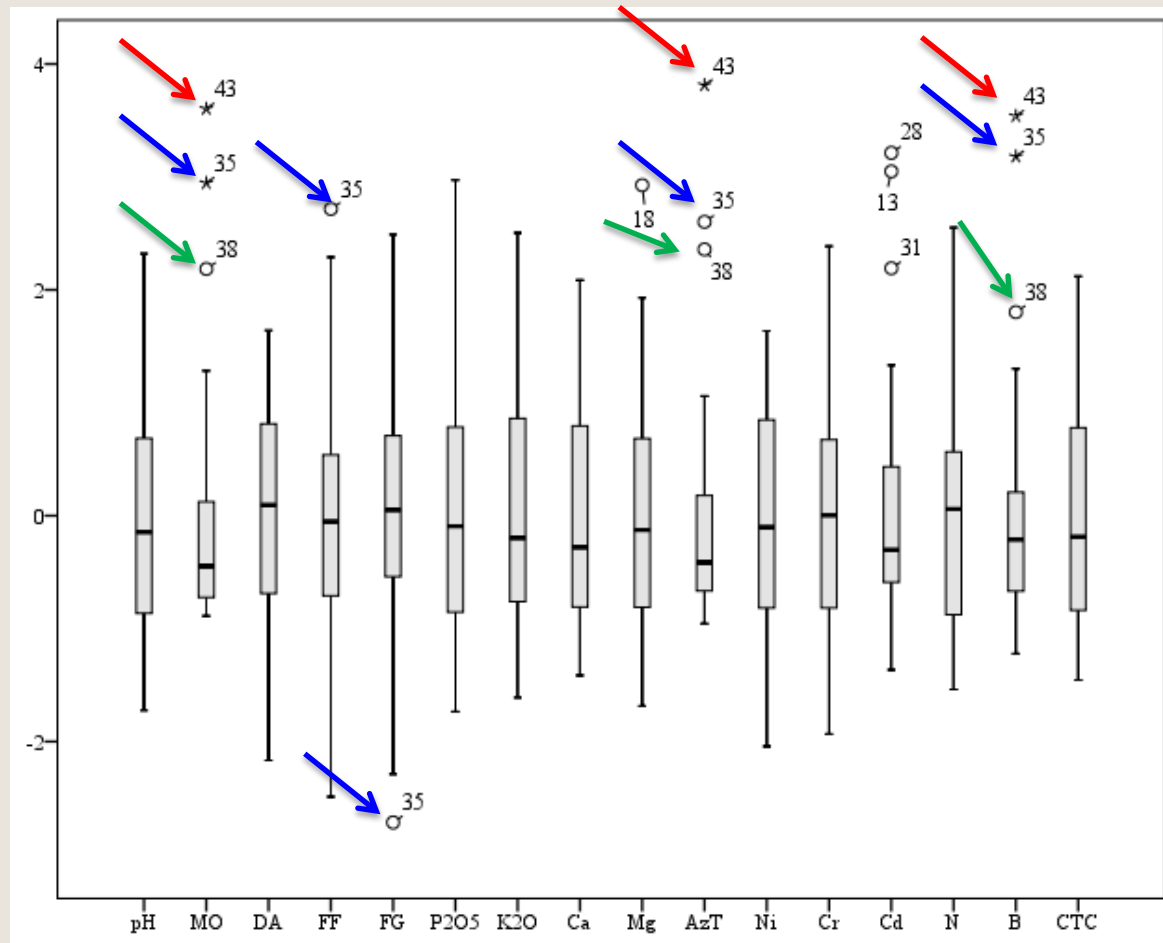


Fig. 3 - Diagrama de extremos e quartis das variáveis do solo estandardizadas

- Variáveis com *outliers*: MO, FF, FG, Mg, AzT, Cd e B.
- Os *outliers* severos aparecem nos pontos de amostragem 35 e 43.

Análise de Componentes Principais

Componente	Valores Próprios Iniciais			Soma dos quadrados dos loadings após a extração			Soma dos quadrados dos loadings após a rotação		
	Total	% de Variância	% acumulada	Total	% de Variância	% acumulada	Total	% de Variância	% acumulada
1	4,764	31,757	31,757	4,764	31,757	31,757	4,113	27,417	27,417
2	3,580	23,870	55,627	3,580	23,870	55,627	3,591	23,942	51,359
3	1,489	9,929	65,556	1,489	9,929	65,556	1,887	12,582	63,942
4	1,283	8,551	74,107	1,283	8,551	74,107	1,525	10,166	74,107
5	0,829	5,527	79,634						
6	0,814	5,426	85,060						
7	0,678	4,522	89,583						
8	0,531	3,538	93,120						
9	0,374	2,491	95,612						
10	0,336	2,242	97,854						
11	0,245	1,634	99,488						
12	0,062	0,412	99,900						
13	0,014	0,093	99,993						
14	0,001	0,007	100						
15	6,41E-15	4,27E-14	100						

Tab. 2 – Variância Total Explicada

- **Critério da variância total explicada:** retêm-se as 4 primeiras CPs que explicam aproximadamente 74,1 % da variabilidade total dos dados.
- **Critério de Kaiser:** retêm-se as 4 primeiras CPs pois têm valores próprios superiores a 1.

Análise de Componentes Principais

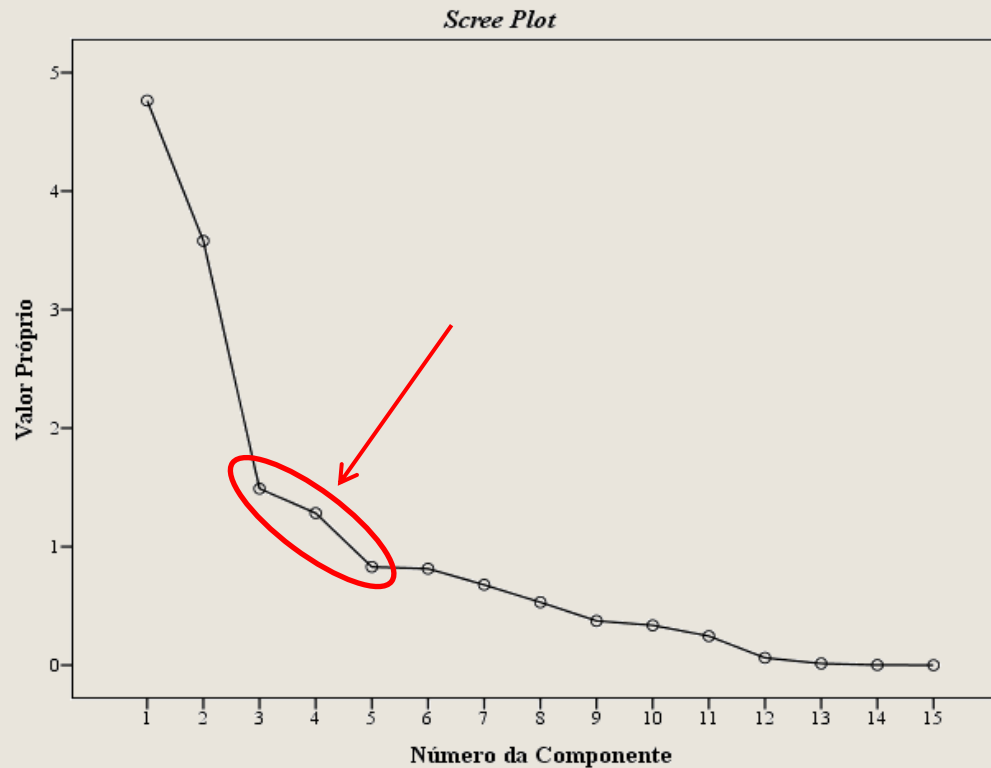


Fig. 4 - Scree Plot

- Analisando o *screeplot* pode-se observar um cotovelo com articulação de “3” a “5” e, de acordo como os dois critérios anteriores, que foram consensuais, optou-se por reter as 4 primeiras CPs.

Análise de Componentes Principais

	Componente			
	1	2	3	4
pH	0,796	-0,244	-0,167	-0,188
MO	-0,068	0,965	0,192	-0,054
DA	-0,139	0,000	-0,057	0,819
FF	-0,255	0,390	0,591	0,123
P ₂ O ₅	0,583	-0,398	0,086	0,365
K ₂ O	0,287	0,024	0,584	0,157
Ca	0,967	0,105	-0,032	-0,047
Mg	0,762	-0,169	0,082	0,278
AzT	0,032	0,965	0,171	-0,084
Ni	-0,223	0,440	0,589	-0,357
Cr	0,059	-0,135	-0,795	0,058
Cd	0,355	-0,111	0,173	0,558
N	0,567	0,407	-0,197	0,338
B	-0,048	0,961	0,193	-0,014
CTC	0,974	0,077	0,021	0,000

Tab. 3 – Matriz dos *loadings* das 4 primeiras CPs com rotação Varimax

- 1ª CP: pH, P₂O₅, Ca, Mg, N e CTC com pesos elevados positivos;
- 2ª CP: MO, AzT e B com pesos elevados positivos;
- 3ª CP: FF, K₂O, Ni com pesos elevados positivos e Cr com peso elevado negativo;
- 4ª CP: DA e Cd com pesos elevados positivos.

Análise de Componentes Principais

	Comunalidades
pH	0,755
MO	0,976
DA	0,693
FF	0,581
P ₂ O ₅	0,639
K₂O	0,448
Ca	0,949
Mg	0,693
AzT	0,968
Ni	0,718
Cr	0,657
Cd	0,480
N	0,640
B	0,964
CTC	0,955

Tab. 4 – Comunalidades após rotação Varimax

- As variáveis originais que mais informação perdem quando se transformam as quinze variáveis originais em quatro CPs são K₂O e Cd.

Análise de Componentes Principais

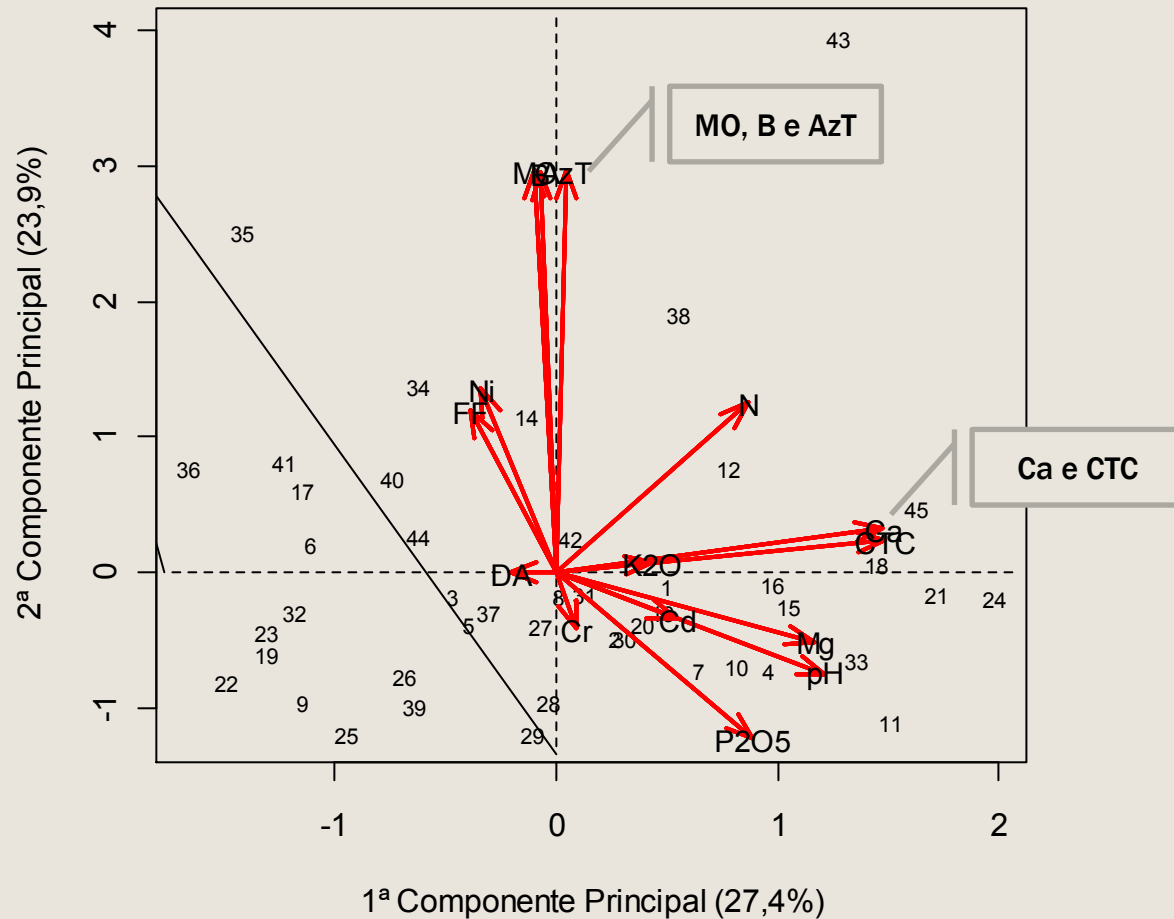


Fig. 5 - Biplot

Análise de *Clusters*

- Utilizaram-se os scores das 4 primeiras CPs para efectuar a Análise de *Clusters*.
- Medida de dissemelhança utilizada no agrupamento:
quadrado da distância euclidiana
- Métodos hierárquicos aglomerativos consensuais na composição dos *clusters*:
ligação completa, ligação média e *Ward*
- Método seleccionado:
Método de *Ward*

Análise de *Clusters*

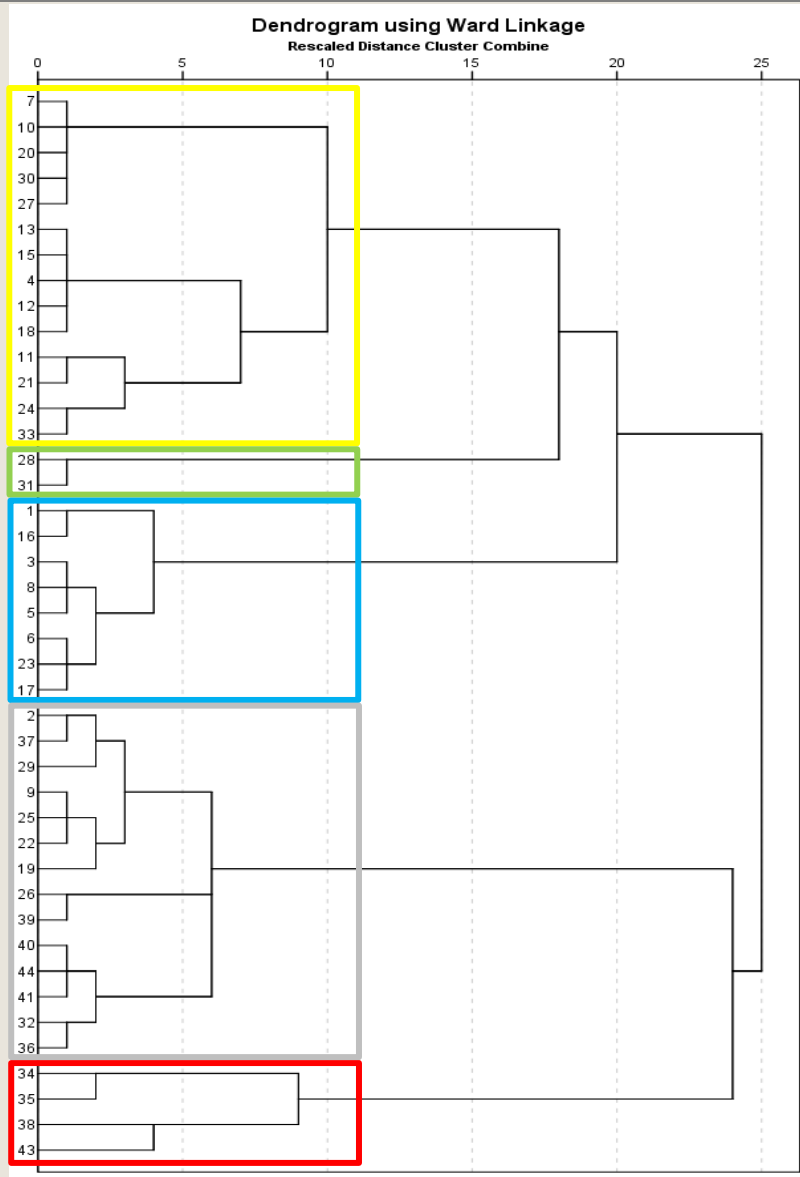


Fig. 6 - Dendrograma com base nas CPs do solo

- A visualização do dendrograma (método empírico) sugere uma divisão em **5 clusters**

Análise de *Clusters*

Número de *Clusters* (método gráfico)

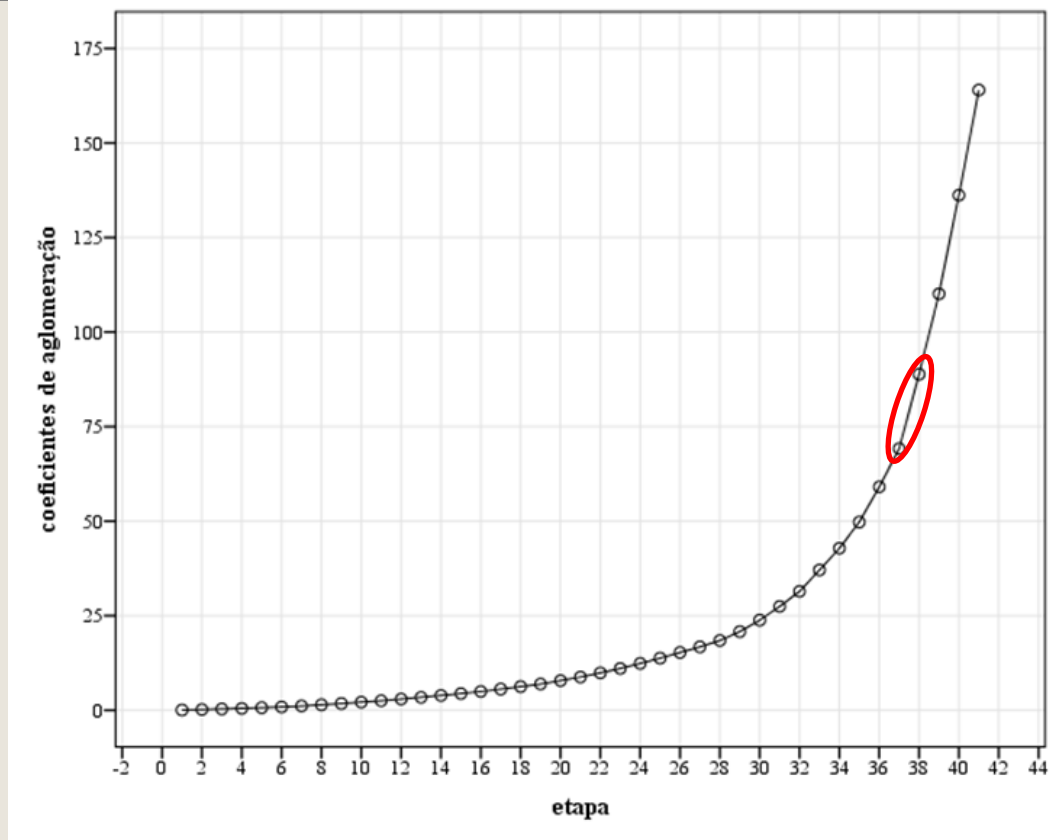


Fig. 7 – Gráfico etapa vs coeficientes de aglomeração

- O primeiro grande aumento destes coeficientes ocorre da etapa 37 para a 38.
- Por análise da tabela de agrupamento até à etapa 37 obtêm-se 5 *clusters*.

Análise de *Clusters*

Número de *Clusters* (método do R^2)

Número de <i>Clusters</i>	R-quadrado
1	0
2	0,1695
3	0,3285
4	0,4582
5	0,5782
6	0,6399
7	0,6966
8	0,7389
9	0,7739
10	0,8083

Tab. 5 – Critério R-quadrado

- Este critério sugere um elevado número de *clusters* (≈ 10).
- Um elevado nº de *clusters* torna a sua interpretação difícil.
- Optou-se pela retenção de 5 *clusters* tal como sugerem os dois critérios anteriores.

Análise de *Clusters*

Identificação da divisão da parcela de acordo com os 5 grupos retidos:



Fig. 8 - Divisão da parcela em 5 grupos

Comparação entre grupos

- Conhecidos os *clusters*, pretende-se saber se existem diferenças significativas entre eles, tendo em consideração as variáveis que traduzem a qualidade do vinho.
- Para testar a existência de diferenças entre os cinco grupos assinalados utilizou-se o teste não paramétrico de *Kruskal-Wallis*, visto que os nossos 5 grupos têm dimensão pequena e desigual.

Mosto

		<i>Clusters</i>					Estatística de teste	Graus de liberdade	Valor Prova
		1	2	3	4	5			
M O S T O	pH_mosto	3,332	3,300	3,319	3,306	3,356	1,771	4	0,778
	Acidez_T	5,670	6,127	5,699	6,676	6,418	9,878	4	0,043
	Acido_malico	1,750	2,179	1,929	2,650	2,900	15,115	4	0,004
	Acido_tartarico	4,588	4,550	4,507	4,600	4,200	5,813	4	0,214
	Açucares	215,438	216,129	216,450	217,050	216,450	1,744	4	0,783
	TAP	12,803	12,842	12,859	12,895	12,860	1,744	4	0,783

Tab. 6 – Resultados do teste de comparação para os constituintes do mosto

- Há evidências estatísticas suficientes nos dados para rejeitar a igualdade de valores médios nos 5 *clusters* em relação às variáveis Acidez_total e Acido_malico.

Videira

		Clusters					Estatística de teste	Graus de liberdade	Valor Prova
		1	2	3	4	5			
Produtividade	Ncachos	44,250	33,385	41,000	29,000	35,000	3,554	4	0,470
	Uvas_kg_vid	4,963	4,635	5,767	5,050	5,175	2,333	4	0,675
	Pcacho_kg	0,111	0,136	0,140	0,174	0,145	9,275	4	0,055
Vigor	Nvaras	34,875	29,077	32,583	51,000	29,750	6,489	4	0,165
	Pvaras_kg	2,538	2,188	2,700	3,950	2,238	2,743	4	0,602
	Pvara_g	70,570	74,654	82,148	77,451	75,187	0,541	4	0,969

Tab. 7 - Resultados do teste de comparação para as variáveis da videira

- Não se verificam diferenças significativas nos valores médios dos 5 *clusters* em relação às variáveis da videira.

Sumo das uvas

		Clusters					Estatística de teste	Graus de liberdade	Valor Prova
		1	2	3	4	5			
S U M O D A S U V A S	FL1	118,256	101,382	95,513	104,880	84,850	8,642	4	0,071
	FL2	227,601	268,426	252,554	296,290	354,260	10,834	4	0,028
	FL3	42,441	18,127	16,982	12,085	5,365	9,325	4	0,053
	FL4	6,603	7,241	8,083	7,015	10,985	8,841	4	0,065
	FL5	20,149	12,884	11,341	9,150	13,235	3,550	4	0,470
	FG1	36,695	30,620	34,707	32,145	25,250	6,254	4	0,181
	FG2	140,455	162,778	150,623	174,715	171,570	2,651	4	0,618
	FG3	3,639	3,428	4,389	3,285	3,145	5,317	4	0,256
	FG4	9,415	10,649	11,016	10,065	10,760	3,124	4	0,537
	FG5	31,951	36,276	39,999	36,630	30,940	4,500	4	0,343
	FG6	36,724	49,372	46,637	61,740	60,800	6,473	4	0,167
	FG7	0,716	1,041	0,986	1,170	1,220	5,596	4	0,231

Tab. 8 – Resultados do teste de comparação para as variáveis do sumo das uvas

- Há evidências estatísticas suficientes nos dados para rejeitar a igualdade de valores médios nos 5 *clusters* em relação à variável FL2.

Vinho

V I N H O		Clusters					Estatística de teste	Graus de liberdade	Valor Prova
		1	2	3	4	5			
	Avaliacao_Global	2,1250	2,0714	2,7143	1,5000	1,5000	11,989	4	0,017
	Nota_Final	60,6250	63,4286	69,2143	58,0000	60,5000	8,180	4	0,085

Tab. 9 – Resultados do teste de comparação para as variáveis do vinho

- Há evidências estatísticas suficientes nos dados para rejeitar a igualdade de valores médios nos 5 *clusters* em relação à variável Avaliação Global.

Solo vs Vinho

- Com o objectivo de verificar quais as variáveis do solo que influenciam a qualidade do vinho determinaram-se as correlações significativas entre as CPs e as variáveis que determinam a qualidade do vinho.
- Nesta análise foi utilizado o teste baseado no coeficiente de correlação não paramétrico de *Spearman*.

Correlações significativas

			Coeficiente de correlação de Spearman	
			CP2	CP3
MOSTO	Acidez_T	Coeficiente de correlação		0,349*
		Valor Prova		0,023
	Acido_malico	Coeficiente de correlação	0,480**	0,408**
		Valor Prova	0,001	0,007
	Acido_tartarico	Coeficiente de correlação	-0,405**	
		Valor Prova	0,008	
* Significativo a 5%				
** Significativo a 1%				

Tab. 10 – Correlações significativas entre as CPs e os constituintes do mosto

Correlações significativas

		Coeficiente de correlação de Spearman	
		CP3	
VIDEIRA	Ncachos	Coeficiente de correlação	-0,367*
		Valor Prova	0,024
* Significativo a 5%			
** Significativo a 1%			

Tab. 11 – Correlações significativas entre as CPs e o vigor e produtividade das videiras

Correlações significativas

			Coeficiente de correlação de Spearman			
			CP2	CP3	CP4	
SUMO DE UVAS	FL1	Coeficiente de correlação		-0,314*	0,317*	
		Valor Prova		0,043	0,041	
	FL2	Coeficiente de correlação	0,455**	0,331*		
		Valor Prova	0,002	0,032		
	FL3	Coeficiente de correlação		-0,344*	0,397**	
		Valor Prova		0,026	0,009	
	FL4	Coeficiente de correlação	0,530**			
		Valor Prova	0			
	FG1	Coeficiente de correlação	-0,425**	-0,378*		
		Valor Prova	0,005	0,014		
	FG4	Coeficiente de correlação			-0,458**	
		Valor Prova			0,002	
	FG5	Coeficiente de correlação			-0,372*	
		Valor Prova			0,015	
	* Significativo a 5%					
	** Significativo a 1%					

Tab. 12- Correlações significativas entre as CPs e as variáveis do sumo das uvas

Correlações significativas

		Coeficiente de correlação de Spearman	
		CP4	
VINHO	Nota_Final	Coeficiente de correlação	-0,402**
		Valor Prova	0,008
* Significativo a 5%			
** Significativo a 1%			

Tab. 13- Correlações significativas entre as CPs e as variáveis do vinho

Conclusão

- Neste estudo, a qualidade do vinho e o equilíbrio vegetativo da videira foram avaliados através das suas correlações com os parâmetros do solo simplificados em 4 CPs.
- Em relação às correlações obtidas entre a CP2 do solo e acidez do mosto, existem já estudos que referem que altos valores de MO no solo induzem alta acidez nos mostos (Delas *et al.*, 1992; Araújo, 2004).
- A determinação de correlações entre as características do solo, a qualidade dos mostos e o equilíbrio vegetativo da videira irá permitir, sempre que se justifique, correcções diferenciadas para cada local da parcela, efectuando-se deste modo, uma agricultura de forma sustentada.

Bibliografia

- Araújo I., 2004. Características Aromáticas e Cromáticas Castas Amarel e Vinhão. *Tese de Mestrado*. Universidade do Porto / Universidade Técnica de Lisboa.
- Branco J., 2004. *Uma Introdução à Análise de Clusters*. 1ª Edição. Lisboa: Sociedade Portuguesa de Estatística.
- Cadima, J., Apontamentos de Estatística Multivariada, Departamento de Matemática – Instituto Superior de Agronomia – Universidade Técnica de Lisboa, 2010
- Clímaco P., 1997. Influência da cultivar e da maturação da uva e na produtividade da videira (*Vitis vinifera* L.). *Tese de Doutoramento*. Instituto Superior de Agronomia, Universidade Técnica de Lisboa.
- Delas J., Molot C., Soyer J.P., 1992. Influence du porte-greffe et de la fertilization azotée sur la composition des baies de Merlot. *IV Simposio Internazionale di Fisiologia della Vite*, Torino, Itália, n: 397-400.
- Jolliffe, I. T., *Principal Component Analysis*, 2nd Edition, Springer, 2002
- Maroco J., 2007. *Análise Estatística com utilização do SPSS*. 3ª Edição. Lisboa: Edições Sílabo.
- Reis E., 2001. *Estatística Multivariada Aplicada*. 2ª Edição. Lisboa: Edições Sílabo.