

Anexo V – Selecção de Covariáveis: Verificação da Multicolinearidade dos Dados

As considerações a efectuar na seguinte análise são procedidas a partir dos valores VIF apresentados nos seguintes *outputs*. Em função desses valores serão incluídas e excluídas variáveis na análise de regressão logística.

Cada coluna representa, na prática, um modelo de regressão linear gerado com as variáveis que apresentam valor VIF. Foram concebidas regressões lineares retirando variáveis uma a uma, de forma a estudar o comportamento que a ausência de uma variável provocava nas restantes, de modo a aferir sobre a sua situação de colinearidade. Foram ainda construídos outros modelos de regressão retirando duas ou mais variáveis, quando necessário, na mesma lógica de estudo.

V.1 – ETA Tavira

V.1.1 – Parâmetros da Etapa de Ozonização

Coefficientes^a

Modelo		Estatísticas de Colinearidade							
		VIF	VIF	VIF	VIF	VIF	VIF	VIF	VIF
1	Dose O ₃	11,757	-	10,088	11,064	8,931	9,723	-	-
	Residual O ₃	1,563	1,341	-	1,509	1,562	-	1,326	-
	TCont O ₃	14,410	13,560	13,916	-	7,623	-	-	13,413
	CT O ₃	18,829	14,303	18,823	9,961	-	9,723	1,326	13,413

a. Variável Dependente: AdenovírusAT

Considerações:

- As variáveis “Dose O₃”, “TContO₃” e “CT O₃” são multicolineares entre si. Em virtude disso, serão excluídas as variáveis “Dose O₃”, “TContO₃”.
- A variável “Residual O₃” não apresenta colinearidade com outras variáveis.
- Variáveis a incluir na regressão logística: “Residual O₃” e “CT O₃”.

V.1.2 – Parâmetros da Etapa de Cloragem

		Estatísticas de Colinearidade			
Modelo		VIF	VIF	VIF	VIF
1	Dose Cl ₂	1,009	-	1,004	1,002
	Residual Cl ₂	33,668	33,471	-	1,002
	CT Cl ₂	33,717	33,471	1,004	-

a. Variável Dependente: AdenovírusAT

Considerações:

- A variável “Dose Cl₂” não apresenta colinearidade com outras variáveis.
- As variáveis “Residual Cl₂” e “CT Cl₂” são multicolineares entre si. Em virtude disso optou-se por se excluir da análise a variável “Residual Cl₂”.
- Variáveis a incluir na regressão logística: “Dose Cl₂” e “CT Cl₂”.

V.1.3 – Parâmetros de Qualidade da Água Bruta

		Estatísticas de Colinearidade					
Modelo		VIF	VIF	VIF	VIF	VIF	VIF
1	DOC	23,445	-	176,073	7,673	9,807	-
	TOC	6,471	6,434	-	5,183	4,945	1,874
	UV254	219,829	71,944	93,578	-	4,066	1,874
	SUVA	122,437	51,213	23,313	2,265	-	-

a. Variável Dependente: AdenovírusAT

Considerações:

- As variáveis “DOC” e “TOC” são provavelmente multicolineares. Dado isso, optou-se por se excluir a variável “DOC” da análise.
- As variáveis “UV254” e “SUVA” são multicolineares entre si. Posto isso, optou-se por se excluir da análise a variável “SUVA”.
- Variáveis a incluir na regressão logística: “TOC” e “UV254”.

V.1.4 – Conjunto global de parâmetros

Aos conjunto de parâmetros até agora incluídos, juntam-se os outros também aprovados pelo teste do número de amostragens anteriormente efectuado no Anexo IV, aos quais a sua multicolinearidade não foi ainda testada (Dose PAC):

Modelo		Coeficientes ^a	
		Estatísticas de Colinearidade	
		VIF	
1	Residual O ₃	2,586	
	CT O ₃	2,811	
	Dose PAC	2,327	
	Dose Cl ₂	1,550	
	CT Cl ₂	2,074	
	TOC	2,395	
	UV254	2,611	

a. Variável Dependente: AdenovírusAT

Considerações:

- Nenhuma covariável apresenta colinearidade com outras variáveis, pela que todas se encontram aptas a integrar um modelo de regressão logística, nomeadamente as seguintes: “Residual O₃”, “CT O₃”, “Dose PAC”, “Dose Cl₂”, “CT Cl₂”, “TOC” e “UV254”.

V.2 – ETA Alcantarilha

V.2.1 – Parâmetros da Etapa de Ozonização

Modelo		Estatísticas de Colinearidade				
		VIF	VIF	VIF	VIF	VIF
1	Dose O ₃	2,178	-	1,781	2,111	2,122
	Residual O ₃	2,330	1,905	-	1,655	1,398
	TCont O ₃	5,999	5,815	4,263	-	1,738
	CT O ₃	5,225	5,091	3,163	1,513	-

a. Variável Dependente: AdenovírusAT

Considerações:

- As variáveis “Dose O₃” e “Residual O₃” não apresentam colinearidade com outras variáveis.
- As variáveis “Tcont O₃” e “CT O₃” são provavelmente multicolineares. Dado isso, optou-se por se excluir da análise a variável “Tcont O₃”.
- Variáveis a incluir na Regressão Logística: “Dose O₃”, “Residual O₃” e “CT O₃”.

V.2.2 – Parâmetros da Etapa de Cloragem

Modelo		Estatísticas de Colinearidade			
		VIF	VIF	VIF	VIF
1	Dose Cl ₂	1,208	-	1,172	1,188
	Residual Cl ₂	94,321	91,506	-	1,188
	CT Cl ₂	93,061	91,506	1,172	-

a. Variável Dependente: AdenovírusAT

Considerações:

- A variável “Dose Cl₂” não apresenta colinearidade com outras variáveis.
- As variáveis “Residual Cl₂” e “CT Cl₂” são multicolineares entre si. Em virtude disso optou-se por se excluir da análise a variável “Residual Cl₂”.
- Variáveis a incluir na regressão logística: “Dose Cl₂” e “CT Cl₂”.

V.2.3 – Parâmetros de Qualidade da Água Bruta

Coefficientes^a

Modelo		Estatísticas de Colinearidade							
		VIF	VIF	VIF	VIF	VIF	VIF	VIF	VIF
1	Turvação	2,507	-	2,446	2,458	1,585	2,383	2,267	1,314
	pH	1,371	1,338	-	1,332	1,352	1,352	1,365	1,318
	DOC	86,824	85,111	84,325	-	82,299	27,131	33,531	-
	TOC	29,132	18,416	28,721	27,614	-	25,700	24,468	2,328
	UV254	331,817	315,424	327,064	103,688	292,723	-	5,542	3,194
	SUVA	161,322	145,854	160,553	62,302	138,497	2,694	-	-

a. Variável Dependente: AdenovírusAT

Considerações:

- As variáveis “Turvação” e “pH” não apresentam colinearidade com outras variáveis.
- As variáveis “DOC” e “TOC” apresentam colinearidade com mais do que uma variável, para além uma da outra. Posto isso serão excluídas ambas, ou uma das duas em conjunto com outras mais variáveis.
- As variáveis “UV254” e “SUVA” são multicolineares entre si. Posto isso, optou-se por se excluir da análise a variável “SUVA”.
- Verifica-se que, ao excluir a variável “DOC” em simultâneo com a variável “SUVA”, o indicador *VIF* da variável “TOC” revela que esta já não apresenta colinearidade com demais variáveis. Deste modo será excluída a variável “DOC” em conjunto com a variável “SUVA” já anteriormente considerada para exclusão.
- Variáveis a incluir na regressão logística: “Turvação”, “pH”, “TOC” e “UV254”.

V.2.4 – Conjunto global de parâmetros

Aos conjunto de parâmetros até agora incluídos, juntam-se os outros também aprovados pelo teste do número de amostragens anteriormente efectuado no Anexo IV, aos quais a sua multicolinearidade não foi ainda testada (“*Dose Coagulante*” e “*Dose PAC*”):

Coefficientes^a

Modelo		Estatísticas de Colinearidade
		VIF
1	Dose O ₃	93,925
	Residual O ₃	15,354
	CT O ₃	40,332
	Dose Coagulante	157,172
	PAC	115,336
	Dose Cl ₂	8,870
	CT Cl ₂	17,899
	Turvação	28,419
	pH	113,607
	TOC	81,915
	UV254	151,327

a. Variável Dependente: AdenovírusAT

Considerações:

- Através deste *output* verifica-se que, no seu conjunto, as 11 covariáveis seleccionadas para a análise regressão logística, praticamente no seu todo, são colineares umas com as outras. Dada esta situação, o procedimento possível a executar será excluir uma ou mais covariáveis de forma gradual, de modo a reduzir a colinearidade do conjunto de variáveis para esta ETA. Deste modo, serão excluídas covariáveis pela ordem decrescente do seu valor VIF, tendo também preferencialmente em conta se as variáveis a excluir estarão ou não presentes na análise de regressão logística para a outra ETA em estudo.
- Variável a excluir na próxima regressão: “*Dose de Coagulante*”.

Coeficientes^a

Modelo		Estatísticas de Colinearidade
		VIF
1	Dose O ₃	84,951
	Resid O ₃	7,959
	CT O ₃	32,277
	PAC	28,514
	Dose Cl ₂	8,558
	CT Cl ₂	6,674
	Turvação	27,743
	pH	15,676
	TOC	27,913
	UV254 _{nm}	136,369

a. Variável Dependente: AdenovírusAT

Considerações:

- Dado que o conjunto das covariáveis ainda se apresenta no seu geral em situação de alta colinearidade, serão excluídas mais variáveis de modo a reduzir a colinearidade do conjunto de covariáveis.
- Variável a excluir na próxima regressão: “Dose O₃”.

Coeficientes^a

Modelo		Estatísticas de Colinearidade
		VIF
	Resid O ₃	4,077
	CT O ₃	15,925
	PAC	2,737
	Dose Cl ₂	6,328
	CT Cl ₂	3,664
	Turvação	5,238
	pH	4,343
	TOC	8,051
	UV254 _{nm}	26,513

a. Variável Dependente: AdenovírusAT

Considerações:

- Dado que o conjunto das covariáveis se apresenta de forma geral em situação de alta/média colinearidade, serão excluídas mais variáveis de modo a reduzir a colinearidade do conjunto de covariáveis.
- Variável a excluir na próxima regressão: “Turvação”.

Coeficientes^a

Modelo		Estatísticas de Colinearidade
		VIF
Resid O ₃		2,503
CT O ₃		12,948
PAC		2,663
Dose Cl ₂		5,008
CT Cl ₂		2,215
pH		4,111
TOC		3,926
UV254 _{nm}		10,387

a. Variável Dependente: AdenovírusAT

Considerações:

- Dado que o conjunto das covariáveis se apresenta de forma geral em situação de alta/média colinearidade, serão excluídas mais variáveis de modo a reduzir a colinearidade do conjunto de covariáveis.
- Variável a excluir na próxima regressão: “pH”.

Coeficientes^a

Modelo		Estatísticas de Colinearidade
		VIF
Resid O ₃		2,499
CT O ₃		6,584
Dose PAC		2,130
Dose Cl ₂		4,371
CT Cl ₂		2,016
TOC		3,900
UV254		6,928

a. Variável Dependente: AdenovírusAT

Considerações:

- Tendo em conta que nenhuma covariável que consta na última regressão realizada apresenta valores de *VIF* superiores a 10, pese embora algumas possuírem valores que as designam como possuindo potenciais colinearidades com outras variáveis, não será excluído mais nenhum parâmetro da análise.
- O conjunto de variáveis selecionado encontra-se assim potencialmente apto a integrar um modelo de regressão logística, sendo então composto pelas seguintes covariáveis: “Resid O₃”, “CT O₃”, “Dose PAC”, “Dose Cl₂”, “CT Cl₂”, “TOC” e “UV254”.