

# **ANÁLISE COMPARATIVA DE PROPRIEDADES FÍSICO-QUÍMICAS E AVALIAÇÕES SENSORIAIS DE VINHOS PORTUGUESES DO TIPO VINHO VERDE**

Antonio G. S. Serrano <sup>a,\*</sup>

<sup>a</sup>Instituto de Ciências Ambientais, Químicas e Farmacêuticas, Unidade José de Filippi, Universidade Federal de São Paulo, 09972-270 Diadema - SP, Brasil

\*e-mail: [antonio.guerreiro@unifesp.br](mailto:antonio.guerreiro@unifesp.br)

## **COMPARATIVE ANALYSIS OF PHYSICAL-CHEMICAL PROPERTIES AND SENSORY EVALUATIONS OF PORTUGUESE VINHO VERDE WINE**

In this study was proposed a methodology to identify if there is any correlation between the sensory evaluation of wines and the physical-chemical properties like alcohol content, residual sugar and volatile acidity. Two large public datasets of white and red vinho verde wines are considered. A non parametric hypothesis test was applied, using Python programming language structured at the development environment Google Colaboratory running wholly in the cloud. After performing the statistical tests, it was not possible to predict if the sensory evaluation is correlated with the physical-chemical variables analysed.

Keywords: vinho verde; datasets; Python; physical-chemical; sensory evaluation.

## **INTRODUÇÃO**

Os vinhos que outrora eram reservados apenas a situações de comemoração ou vistos como um item de consumo altamente elitizado passaram gradativamente a se popularizar<sup>1</sup>. O mercado internacional de vinhos mesmo em meio à pandemia da COVID-19, que afetou drasticamente a economia global, apresentou um aumento de produção e consumo<sup>2,3</sup>.

Os vinhos portugueses do tipo Vinho Verde são típicos do Minho (região noroeste de Portugal), constituindo uma região demarcada como Indicação Geográfica e Denominação de Origem (IGDO)<sup>4</sup>, isto é, sob a guarda de um instrumento legal de propriedade industrial que objetiva distinguir a origem geográfica de um determinado produto ou serviço cujas qualidades ou

características estejam intrinsecamente atreladas à localidade do meio geográfico, considerando também fatores naturais e humanos<sup>5</sup>.

Com o crescimento do mercado global de vinhos, produtores do mundo inteiro buscam a superioridade não só no que se refere à produtividade das vinícolas, mas, principalmente, à experiência organoléptica única que o vinho possa proporcionar ao consumidor final<sup>6</sup>. Cada vez mais procura-se aproximar características físico-químicas como dulçor<sup>8</sup> (glicose ou frutose que não foram consumidas no processo fermentativo de produção vinícola), teor alcoólico<sup>9</sup> (concentração de etanol) e acidez volátil (responsável pelo aroma avinagrado que é desagradável e deletério sensorialmente ao vinho<sup>7</sup>, oriundo da fermentação acética e expresso através da concentração de ácido acético) dos patamares elevados de excelência ditados pelos profissionais da área de degustação e consumidores finais<sup>10,11</sup>.

## **PARTE EXPERIMENTAL**

### **Bancos de dados de Vinhos**

Este estudo utilizou dois bancos de dados públicos disponibilizados pela Universidade da Califórnia Irvine (UCI)<sup>12</sup> que abordam as propriedades físico-químicas e avaliações sensoriais de vinhos portugueses do tipo “Vinho Verde” das variedades tinto e branco. O projeto foi desenvolvido em *Python* e executado em nuvem via plataforma *Google Colaboratory*<sup>13</sup>.

### **Preparação dos Bancos de Dados**

Os bancos de dados foram transferidos do repositório *web* da UCI diretamente para o servidor de armazenamento de dados em nuvem *Google Drive* através de suas respectivas *urls* por meio da aplicação *Linux Wget*, operada através de programação *Shell Script* no Google Colaboratory.

A cada um dos bancos de dados foram adicionadas duas colunas contendo as respectivas variáveis categóricas: tipo de vinho (tinto ou branco - *Type: Red/White*) e classificação (bom ou ruim - *Rating: Good/Bad*), objetivando a construção de um banco de dados com parâmetros binários. O parâmetro classificação sensorial foi estabelecido por meio de um algoritmo de tomada de decisão baseado na estrutura condicional *if/else* em que vinhos com qualidade superior ou igual a 7 são classificados como bons vinhos<sup>14</sup>. Em seguida, criou-se um terceiro banco de dados através da unificação dos dois

bancos de dados supracitados, a Tabela 1 apresenta um sumário do banco de dados unificado (somente variáveis numéricas).

**Tabela 1. Medidas de tendência central do banco de dados unificado**

Atributo (unidade)	Média	Desvio Padrão	Mínimo	25%	50%	75%	Máximo
Acidez Fixa (g.dm <sup>-3</sup> )	7.22	1.30	3.8	6.4	7	7.7	15.9
Acidez Volátil (g.dm <sup>-3</sup> )	0.34	0.16	0.08	0.23	0.29	0.4	1.58
Ácido Cítrico (g.dm <sup>-3</sup> )	0.32	0.15	0	0.25	0.31	0.39	1.66
Acúcar Residual (g.dm <sup>-3</sup> )	5.44	4.76	0.6	1.8	3	8.1	65.8
Cloretos (g.dm <sup>-3</sup> )	0.056	0.035	0.009	0.038	0.047	0.065	0.611
SO <sub>2</sub> Livre (mg.dm <sup>-3</sup> )	30.53	17.75	1	17	29	41	289
SO <sub>2</sub> Total (mg.dm <sup>-3</sup> )	115.74	56.52	6	77	118	156	440
Densidade (g.cm <sup>-3</sup> )	0.995	0.003	0.987	0.992	0.995	0.997	1.039
pH	3.22	0.16	2.72	3.11	3.21	3.32	4.01
Sulfatos (g.dm <sup>-3</sup> )	0.53	0.15	0.22	0.43	0.51	0.6	2
Álcool (vol.%)	10.49	1.19	8	9.5	10.3	11.3	14.9
Qualidade	5.82	0.87	3	5	6	6	9

Os parâmetros de acidez fixa e volátil referem-se, respectivamente, à concentração de ácido tartárico e ácido acético.

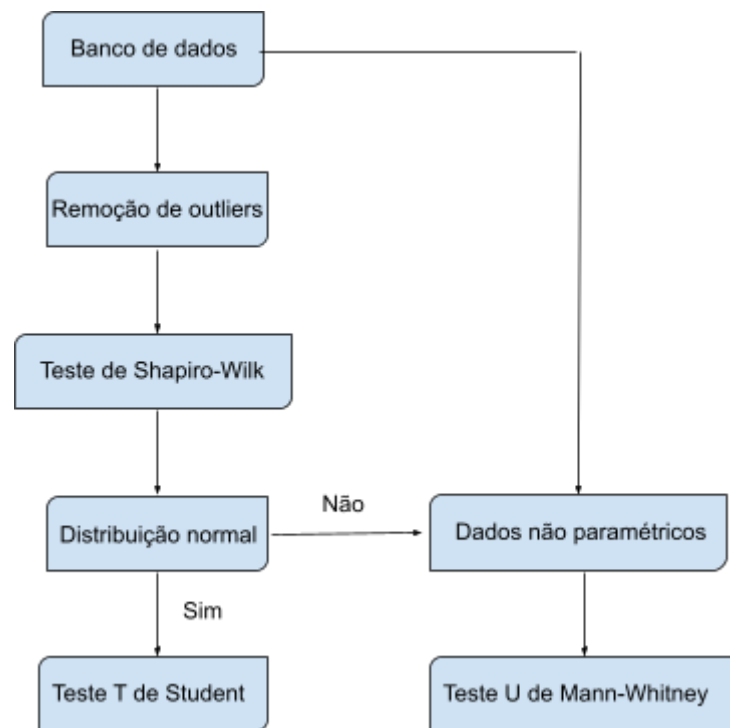
## Análise Exploratória de Dados

As variáveis submetidas à análise neste trabalho foram teor alcoólico, açúcar residual e acidez volátil porque estas são diretamente ligadas ao processo fermentativo de produção do vinho. Ademais, são parâmetros decisivos na percepção sensorial efetuada pelo profissional especializado (enólogo ou *sommelier*) e consumidor final<sup>15</sup>.

Objetiva-se avaliar se há diferença estatisticamente significativa entre as variáveis citadas quando se compara um vinho de mesmo tipo relativamente à sua classificação sensorial. Ou seja, pode-se afirmar que, por exemplo, o Teor Alcoólico médio de um tipo de vinho classificado como Bom é diferente do mesmo tipo de vinho classificado como Ruim? É possível indicar a qualidade de um vinho mediante análise relacional com estas três variáveis indicadas?

Para indicar se há uma igualdade estatisticamente significativa ou não, utiliza-se um teste T de Student, sendo requisito prévio uma distribuição normal de probabilidades ou gaussiana, caso

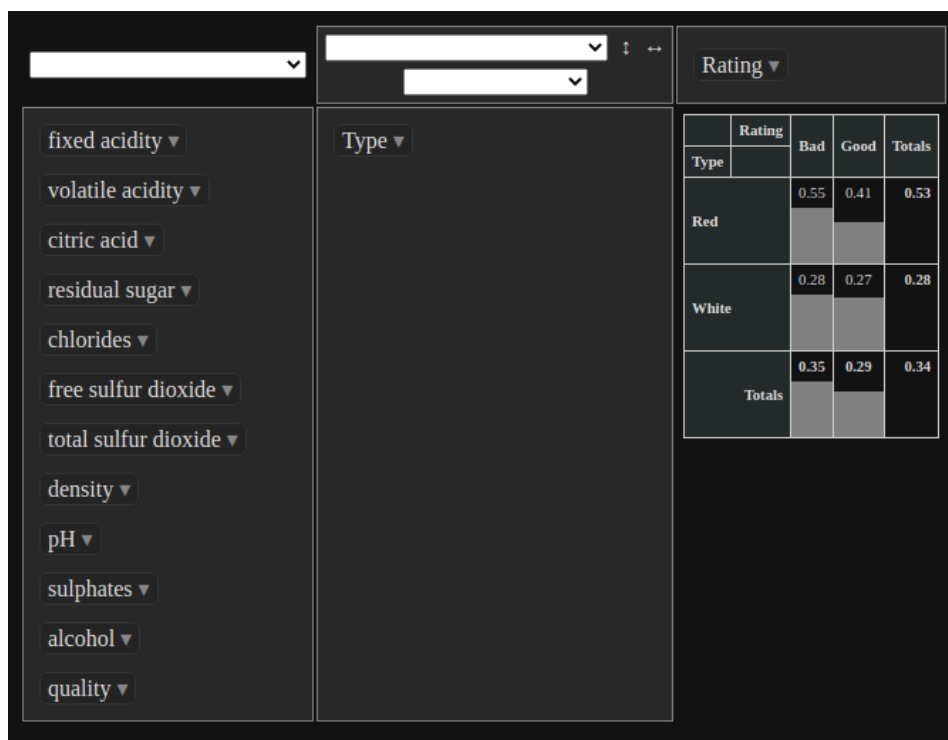
contrário, tem-se dados não paramétricos, logo, utiliza-se o teste U de Mann-Whitney. A Figura 1 indica as etapas analíticas que foram seguidas.



**Figura 1.** Esquema com as etapas da Análise Exploratória de Dados proposta

## RESULTADOS E DISCUSSÃO

A Figura 2 mostra um exemplo de tabela dinâmica interativa gerada em linguagem HTML a partir do banco de dados unificado, utilizando-se os módulos *pivottablejs* e *IPython.display*, passível também de execução em qualquer navegador de internet. Esta ferramenta apresenta uma interface gráfica interativa do tipo *drag-and-drop* (arrastar e soltar), tornando mais rápidas e intuitivas a manipulação e visualização das informações estatísticas de tendência central. Ademais, é possível observar um comparativo das variáveis relativamente ao critério de classificação sensorial.



**Figura 2.** Tabela dinâmica interativa com os valores médios de acidez volátil, segundo o tipo de vinho e classificação sensorial

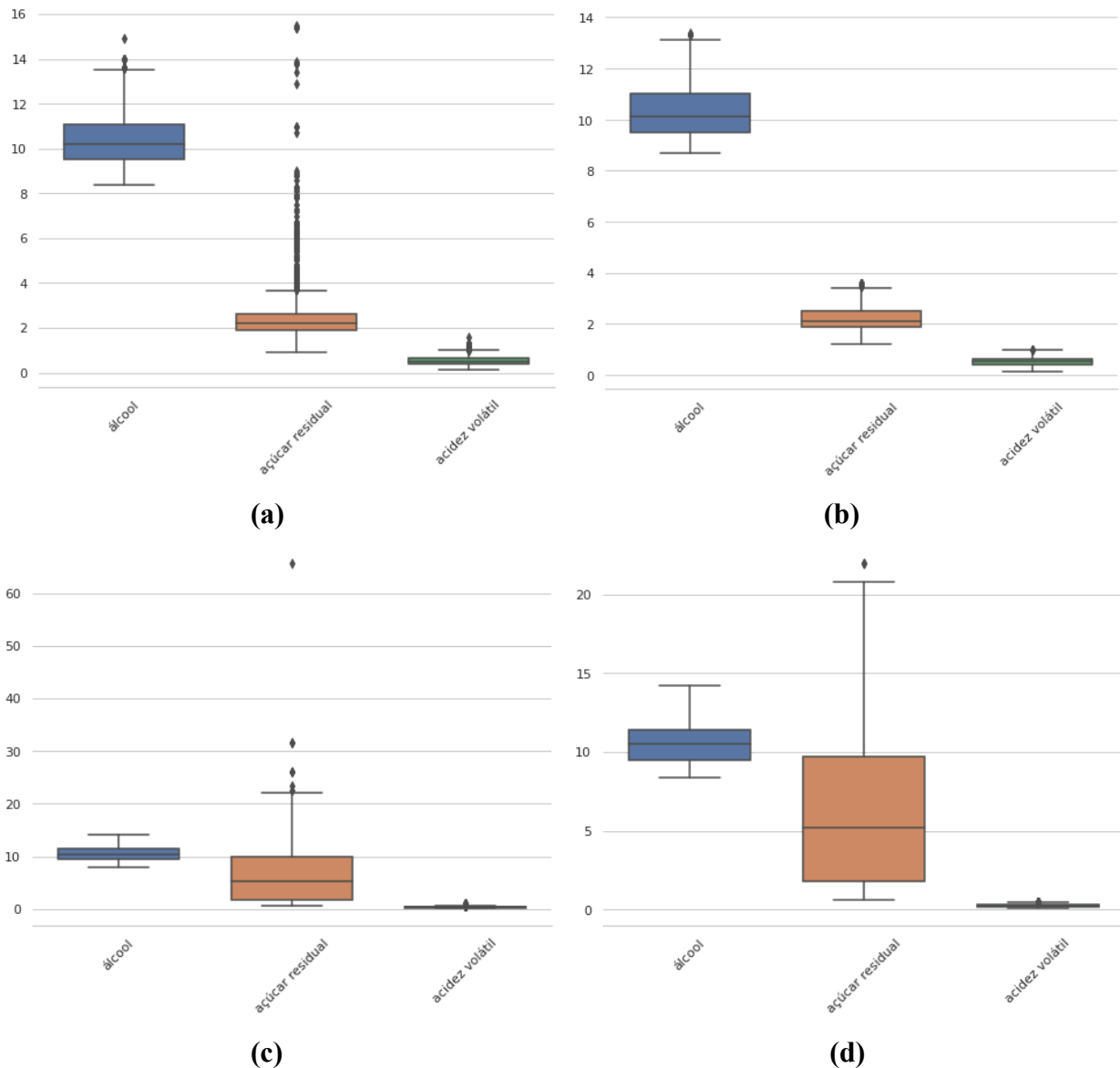
A Tabela 2 traz o recorte dos dados específico para as três variáveis de interesse, evidenciando uma grande proximidade dos valores das médias de um mesmo tipo de vinho com relação à classificação sensorial (Bom ou Ruim). Uma avaliação preliminar indicaria que vinhos de ambos os tipos classificados como Bom, por exemplo, tendem a apresentar um maior teor alcoólico.

**Tabela 2.** Tabela dinâmica com as médias de teor alcoólico, açúcar residual e acidez volátil, segundo o tipo de vinho e avaliação.

	Teor Alcoólico (vol.%)		Açúcar Residual (g.dm <sup>-3</sup> )		Acidez Volátil (g.dm <sup>-3</sup> )	
	Ruim	Bom	Ruim	Bom	Ruim	Bom
Tinto	10.25	11.52	2.51	2.71	0.55	0.41
Branco	10.27	11.42	6.70	5.26	0.28	0.27

A acidez volátil refere-se à concentração de ácido acético.

A Figura 3 indica os gráficos *boxplot* antes e após a remoção dos dados *outliers* pelo Método da Amplitude Interquartil (*IQR - Interquartile Range*).



**Figura 3.** Gráficos boxplot para teor alcoólico, açúcar residual e acidez volátil (Vinho Tinto: (a) com outliers e (b) sem outliers ; Vinho Branco: (c) com outliers e (d) sem outliers)

A Figura 4 mostra um exemplo da estruturação do código para execução do teste de Shapiro-Wilk para determinar se há uma distribuição normal de probabilidades. Para as três variáveis de interesse, não foi evidenciada uma distribuição normal ( $p < 0,05$ ). Então, os dados são não paramétricos e o teste U de Mann-Whitney é o mais apropriado.

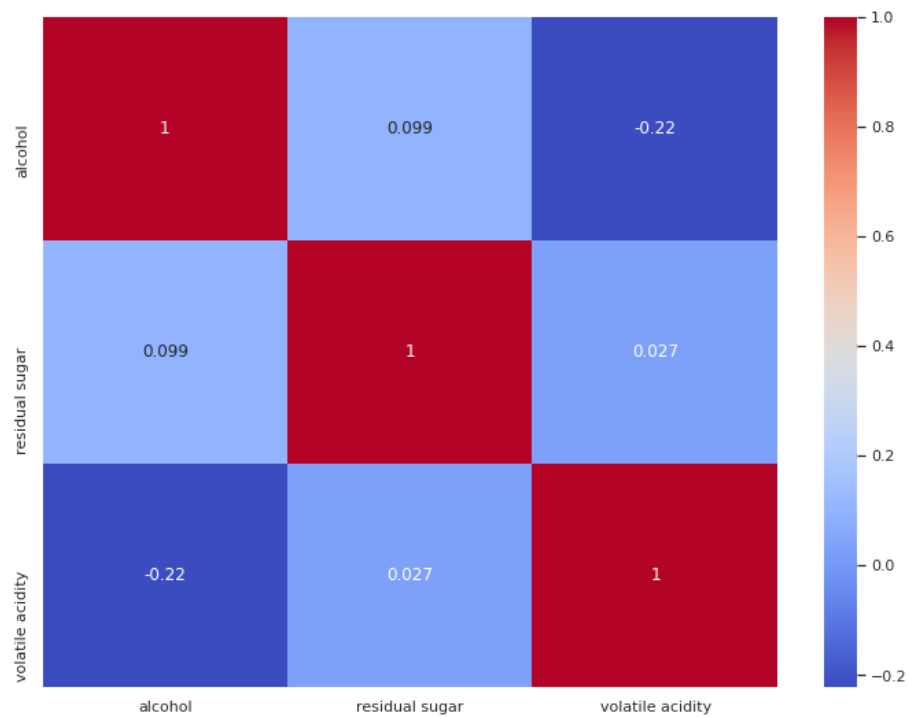
#### ▾ Teste de Shapiro-Wilk: Teor alcoólico

```
[ ] 1 shapiro_stat, shapiro_p_valor = stats.shapiro(dadosTintoClean['alcohol'])
    2
    3 print("O valor da estatística de Shapiro-Wilk = " + str(shapiro_stat))
    4
    5 print("O valor de p de Shapiro-Wilk = " + str(shapiro_p_valor))
    6
    7 #Conclusão do teste
    8 if shapiro_p_valor > 0.05:
    9
    10     print("Com 95% de confiança, os dados são similares a uma distribuição normal segundo o teste de Shapiro-Wilk!")
    11
    12 else:
    13
    14     print("Com 95% de confiança, os dados NÃO são similares a uma distribuição normal segundo o teste de Shapiro-Wilk!")
```

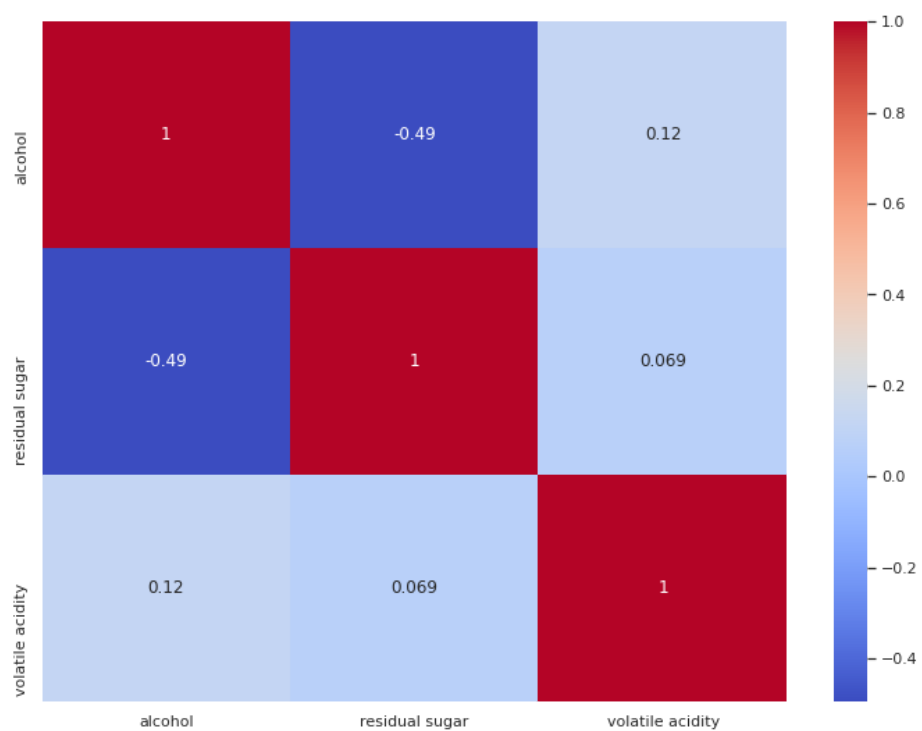
O valor da estatística de Shapiro-Wilk = 0.9296357035636902  
O valor de p de Shapiro-Wilk = 3.019555703528124e-23  
Com 95% de confiança, os dados NÃO são similares a uma distribuição normal segundo o teste de Shapiro-Wilk!

**Figura 4.** Captura de tela de um exemplo de teste de Shapiro-Wilk para normalidade efetuado através do módulo `scipy.stats` do Python, com algoritmo de tomada de decisão utilizando a estrutura condicional `if/else`

As Figuras 5 e 6 mostram os mapas de calor de correlação (*correlation heatmap*), ferramenta exploratória que permite investigar as relações de pares através do coeficiente de correlação de Pearson, que mede a intensidade do relacionamento entre duas variáveis. Quanto mais próximo dos extremos (-1 e +1), mais forte é a correlação. Quanto mais próximo do centro do intervalo, zero, mais fraca é a correlação linear. De acordo com as análises, não há nenhuma correlação linear significativa para o conjunto de variáveis avaliado.



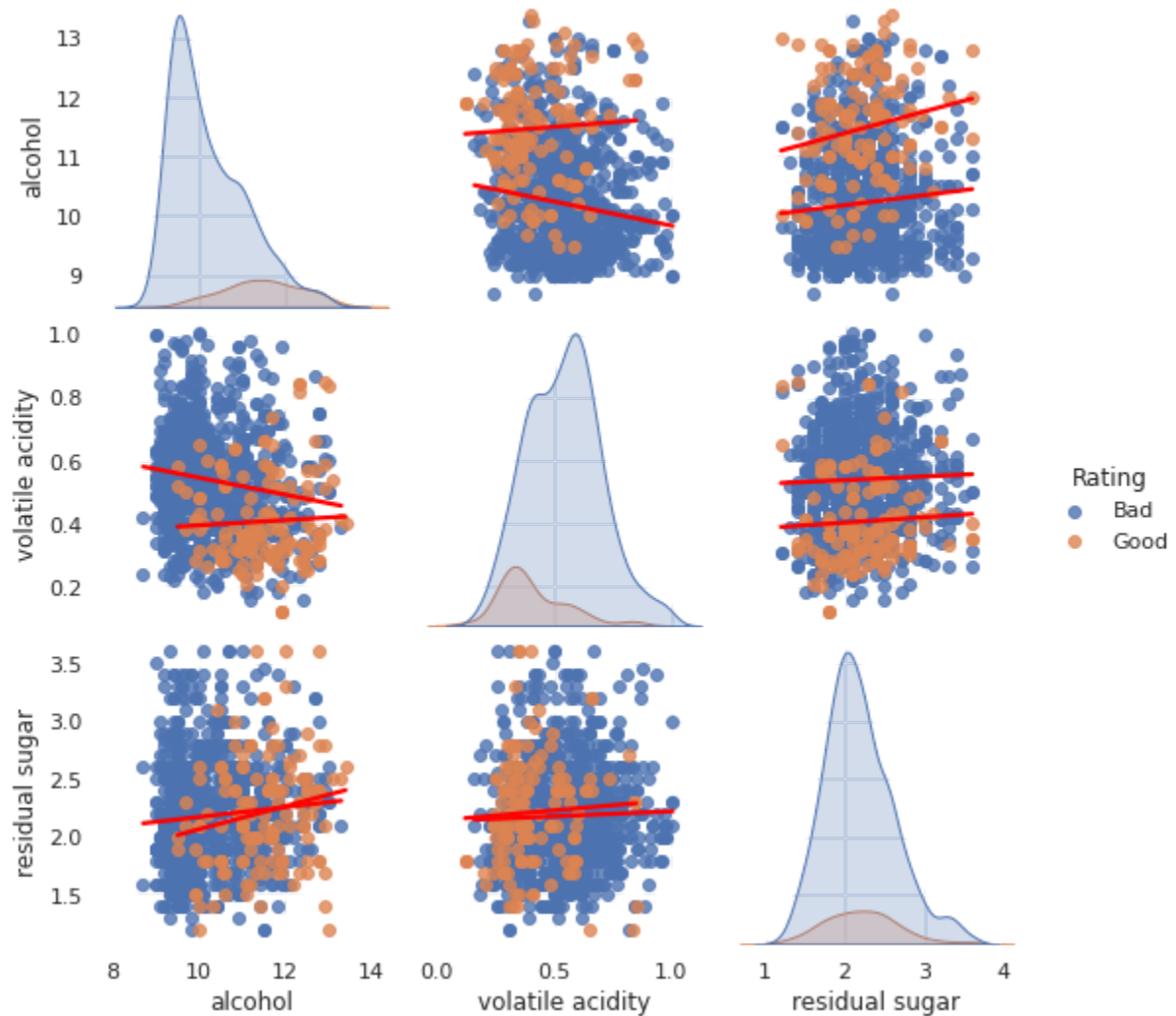
**Figura 5.** Matriz de correlação de Pearson para o banco de dados Vinho Tinto



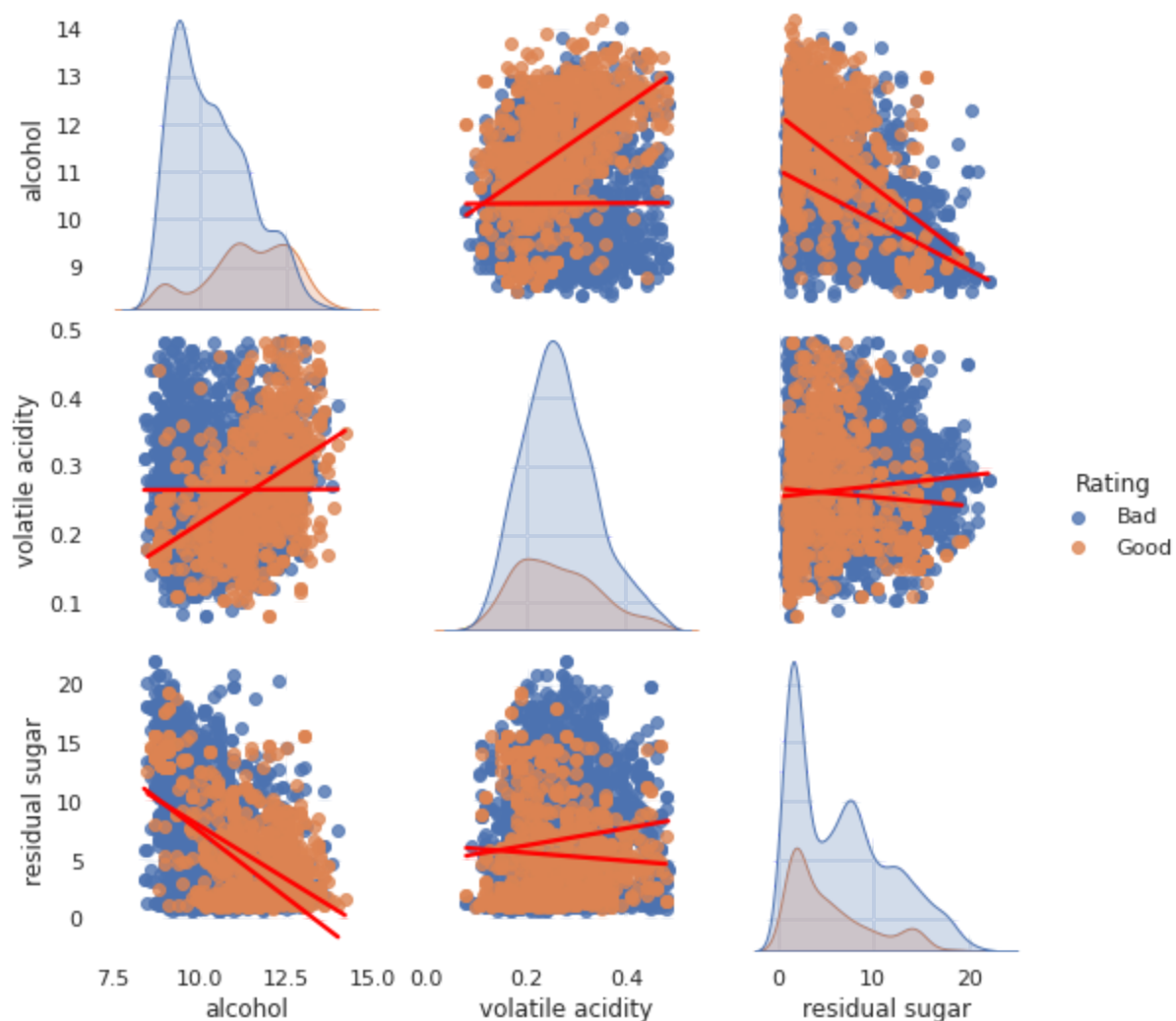
**Figura 6.** Matriz de correlação de Pearson para o banco de dados Vinho Branco



As Figuras 7 e 8 confirmam que não é possível estabelecer uma relação de normalidade e linearidade para o conjunto de variáveis analisado, pois nenhuma das distribuições de probabilidade assemelha-se a uma gaussiana e os dados encontram-se altamente dispersos ao redor da curva de regressão linear calculada.



**Figura 7.** Gráficos de regressão linear e distribuição de probabilidade para o banco de dados Vinho Tinto, correlacionando as variáveis teor alcoólico, açúcar residual e acidez volátil segundo a avaliação do vinho



**Figura 8.** Gráficos de regressão linear e distribuição de probabilidade para o banco de dados Vinho Branco, correlacionando as variáveis teor alcoólico, açúcar residual e acidez volátil segundo a avaliação do vinho

A Tabela 3 demonstra os dados para as três variáveis de interesse, evidenciando uma grande proximidade dos valores das medianas de um mesmo tipo de vinho com relação à classificação sensorial (Bom ou Ruim). Assim como ocorreu na Tabela 2, novamente, uma avaliação preliminar indicaria que vinhos de ambos os tipos classificados como Bom, por exemplo, tendem a apresentar um maior teor alcoólico.

**Tabela 3. Tabela dinâmica com as medianas de teor alcoólico, açúcar residual e acidez volátil, segundo o tipo de vinho e avaliação.**

	Teor Alcoólico (vol.%)		Açúcar Residual (g.dm <sup>-3</sup> )		Acidez Volátil (g.dm <sup>-3</sup> )	
	Ruim	Bom	Ruim	Bom	Ruim	Bom
Tinto	10	11.6	2.2	2.3	0.54	0.37
Branco	10	11.5	6	3.88	0.27	0.25

A acidez volátil refere-se à concentração de ácido acético.

A Figura 4 demonstra como foi estruturado o código para execução do teste U de Mann-Whitney para determinar se há uma igualdade das medianas para o mesmo tipo de vinho segundo a classificação sensorial. Para as três variáveis de interesse, em todos os casos foi verificado que a hipótese nula é verdadeira, isto é, não há diferença estatisticamente significativa entre as medianas ( $p > 0,05$ ).

```

1 #Codificação dos dataframes (df)
2 #R = tinto, W = branco, A = álcool, V = acidez volátil, R = açúcar residual, G = bom, B = ruim
3
4 #dfRVG
5 filterCriteria1 = (dadosTinto['Rating'] == 'Good') | (dadosTinto['volatile acidity'])
6 #dfRVB
7 filterCriteria2 = (dadosTinto['Rating'] == 'Bad') | (dadosTinto['volatile acidity'])
8
9 dfRVG = dadosTinto[filterCriteria1]
10
11 dfRVB = dadosTinto[filterCriteria2]
12
13 #mannwhitneyu(dfRVG['volatile acidity'], dfRVB['volatile acidity'])
14
15 mannwhitneyu_stat, mannwhitneyu_p_valor = stats.mannwhitneyu(dfRVG['volatile acidity'], dfRVB['volatile acidity'])
16
17 print("O valor da estatística de Mann-Whitney = " + str(mannwhitneyu_stat))
18
19 print("O valor de p de Mann-Whitney = " + str(mannwhitneyu_p_valor))
20
21 #Conclusão do teste
22 if mannwhitneyu_p_valor > 0.05:
23
24     print("Com 95% de confiança, NÃO há diferença estatisticamente significativa entre as medianas segundo o teste U de Mann-Whitney!")
25
26 else:
27
28     print("Com 95% de confiança, há diferença estatisticamente significativa entre as medianas segundo o teste U de Mann-Whitney!")

```

O valor da estatística de Mann-Whitney = 1278400.5  
O valor de p de Mann-Whitney = 0.49999235848324214  
Com 95% de confiança, NÃO há diferença estatisticamente significativa entre as medianas segundo o teste U de Mann-Whitney!

**Figura 9.** Captura de tela de um exemplo de teste U de Mann-Whitney utilizado para avaliar a igualdade das medianas efetuado através do módulo `scipy.stats` do Python, com algoritmo de tomada de decisão utilizando a estrutura condicional `if/else`

## CONCLUSÕES

Segundo as análises realizadas neste estudo, não é possível indicar a qualidade de um vinho (parâmetros binários de classificação sensorial Bom ou Ruim) segundo as variáveis teor alcoólico, açúcar residual e acidez volátil e nem estabelecer nenhuma correlação linear entre essas variáveis. Logo, pelos métodos estatísticos tradicionais, não é possível modelar estes bancos de dados de modo a correlacionar suas variáveis, exigindo uma abordagem dos mesmos através de técnicas estatísticas mais avançadas, como regressão não linear.

## REFERÊNCIAS

1. <https://g1.globo.com/economia/noticia/2021/04/20/consumo-de-vinho-tem-alta-de-18percent-em-2020-puxado-pelo-e-commerce.ghtml>, acessada em Julho de 2021.
2. [https://revistaadega.uol.com.br/artigo/consultoria-revela-crescimento-do-mercado-de-vinhos-no-mundo-e-aumento-do-consumo-na-china\\_11435.html](https://revistaadega.uol.com.br/artigo/consultoria-revela-crescimento-do-mercado-de-vinhos-no-mundo-e-aumento-do-consumo-na-china_11435.html), acessada em Julho de 2021.
3. <https://www.oiv.int/en/oiv-life/2020-a-year-of-resilience>, acessada em Julho de 2021.
4. <https://www.vinhoverde.pt/pt/sobre-o-vinho-verde>, acessada em Julho de 2021.
5. [https://manualdeig.inpi.gov.br/projects/manual-de-indicacoes-geograficas/wiki/02\\_Indica%C3%A7%C3%A3o\\_Geogr%C3%A1fica\\_eEsp%C3%A9cies\\_de\\_registro](https://manualdeig.inpi.gov.br/projects/manual-de-indicacoes-geograficas/wiki/02_Indica%C3%A7%C3%A3o_Geogr%C3%A1fica_eEsp%C3%A9cies_de_registro), acessada em Julho de 2021.
6. [https://revistaadega.uol.com.br/artigo/a-qualidade-do-vinho-brasileiro-em-numeros\\_5516.html](https://revistaadega.uol.com.br/artigo/a-qualidade-do-vinho-brasileiro-em-numeros_5516.html), acessada em Julho de 2021.
7. [https://revistaadega.uol.com.br/artigo/principais-defeitos-do-vinho\\_6468.html](https://revistaadega.uol.com.br/artigo/principais-defeitos-do-vinho_6468.html), acessada em Julho de 2021.
8. [https://revistaadega.uol.com.br/artigo/escola-do-vinho\\_12141.html](https://revistaadega.uol.com.br/artigo/escola-do-vinho_12141.html), acessada em Julho de 2021.
9. RIZZON, L. A. Metodologia para análise de vinho: Embrapa Uva e Vinho, 2010.
10. <https://sistemasdeproducao.cnptia.embrapa.br/FontesHTML/Vinho/SistemaProducaoVinhoTitulo/glossario.htm>, acessada em Julho de 2021.

11. Nota Técnica: Acidez Volátil. Instituto dos Vinhos do Douro e Porto. Disponível em: <https://www.ivdp.pt/pt/docs/Nota%20te%CC%81cnica%20acidez%20vol%C3%A1til%20Douro%202017.pdf>, acessada em Julho de 2021.
12. <https://archive.ics.uci.edu/ml/datasets/wine+quality>, acessada em Julho de 2021.
13. <https://colab.research.google.com/notebooks/welcome.ipynb?hl=pt-BR>, acessada em Julho de 2021.
14. CICCHETTI, D.V.; CICCHETTI A.F. Wine rating scales: Assessing their utility for producers, consumers, and oenologic researchers. International Journal of Wine Research 2009, 73-83.
15. MIELE, A. Técnicas de análise sensorial de vinhos e espumantes: Embrapa Uva e Vinho, 2006.