

Análise Exploratória da Qualidade dos Vinhos Tinto e Branco

com Base em Propriedades Físico-Químicas e Sensoriais

1st Marina Vasques Rodrigues
Dept. de TeleInformática
Universidade Federal do Ceará
Fortaleza, Brazil
marinavasq18@alu.ufc.br

2nd Fábio Gabriel Esteves Ivo Gomes
Dept. de TeleInformática
Universidade Federal do Ceará
Fortaleza, Brazil
fabiogabriel@alu.ufc.br

3rd Caio Vinícius Pessoa Freires
Dept. de TeleInformática
Universidade Federal do Ceará
Fortaleza, Brazil
caiopessoa145@gmail.com

4th Fábio Agostinho da Silva Nascimento Filho
Dept. de TeleInformática
Universidade Federal do Ceará
Fortaleza, Brazil
fabinhosnf@gmail.com

Abstract—Este documento apresenta a análise exploratória do dataset "Wine Quality". São avaliadas propriedades físico-químicas e sensoriais dos vinhos, com destaque para estatísticas descritivas e visualização gráfica.

Index Terms—análise exploratória, vinhos, estatística descritiva, boxplot, histogramas

I. INTRODUÇÃO

O estudo da qualidade dos vinhos é relevante para a indústria e para consumidores. Este trabalho realiza uma análise exploratória do dataset "Wine Quality", avaliando variáveis físico-químicas e sensoriais.

II. MÉTODOS

A. Descrição do Dataset

O dataset "Wine Quality" [1] contém 6.497 amostras, sendo 1.599 vinhos **tintos** e 4.898 vinhos **brancos**. As variáveis de entrada são fatores físico-químicos (como pH e densidade), enquanto a saída representa a avaliação sensorial, obtida pela média de pelo menos três especialistas, em uma escala de 0 a 10.

As 11 variáveis de entrada são:

- 1) **Acidez fixa (g/L)**: ácidos naturais predominantes, influenciam frescor e aroma [2].
- 2) **Acidez volátil (g/L)**: ácidos que evaporam facilmente, impactam sabor e aroma [3].
- 3) **Ácido cítrico (g/L)**: presente em menor quantidade, equilibra acidez [2],[6].
- 4) **Açúcar residual (g/L)**: açúcar restante após fermentação, influencia doçura [4].
- 5) **Cloretos**: teor de cloretos, maior próximo ao mar [5].

- 6) **Dióxido de enxofre livre (mg/L)**: influência na preservação e estabilidade do vinho.
- 7) **Dióxido de enxofre total (mg/L)**: soma do livre e ligado, afeta conservação.
- 8) **Densidade**: concentração de ácidos, açúcares e outros compostos.
- 9) **pH**: nível de acidez total, influencia sabor e estabilidade.
- 10) **Sulfatos (g/L)**: contribuem para sabor e antioxidante natural.
- 11) **Álcool (% vol)**: impacto na percepção de corpo e sabor.

B. Análise Monovariada Incondicional

A análise monovariada incondicional avalia cada preditor X_d individualmente, usando todas as N observações. Os passos são:

- 1) Plotagem de histogramas (incondicional)
- 2) Cálculo da média μ_d :

$$\mu_d = \frac{1}{N} \sum_{i=1}^N X_{i,d}$$

- 3) Cálculo do desvio padrão populacional σ_d :

$$\sigma_d = \sqrt{\frac{1}{N} \sum_{i=1}^N (X_{i,d} - \mu_d)^2}$$

- 4) Cálculo da assimetria (skewness) γ_d :

$$\gamma_d = \frac{\frac{1}{N} \sum_{i=1}^N (X_{i,d} - \mu_d)^3}{\sigma_d^3}$$

III. RESULTADOS

A. Descrição do Dataset

Nas duas imagens abaixo, temos um resumo estatístico dos Vinho Tinto e Branco. Podemos observar que a média dos

preditores Ácido Fixo, Açúcar Residual, Dióxido de Enxofre Livre e Dióxido de Enxofre Total possuem uma diferença significativa. Provavelmente essas diferenças influenciam no fato de a média de qualidade dos vinhos branco ser maior que dos vinhos tintos e de a máxima do branco ser 9.00 e do tinto ser 8.00, diferença considerável.

Vinho Tinto												
	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	ph	sulphates	alcohol	quality
count	1509.000000	1509.000000	1509.000000	1509.000000	1509.000000	1509.000000	1509.000000	1509.000000	1509.000000	1509.000000	1509.000000	1509.000000
mean	0.319637	0.527821	0.270876	2.538006	0.007467	15.874822	46.467792	2.208702	3.311113	0.602148	10.387905	5.636023
std	1.741066	0.179068	0.194881	1.409028	0.047965	10.480157	32.895324	0.664060	0.154356	0.189587	1.159825	0.007569
min	4.000000	0.120000	0.000000	0.000000	0.012000	1.000000	6.000000	0.990070	2.740000	0.330000	1.000000	3.000000
25%	7.100000	0.390000	0.080000	1.900000	0.070000	7.000000	22.000000	0.995000	3.210000	0.530000	9.500000	5.000000
50%	7.900000	0.520000	0.260000	2.500000	0.070000	14.000000	36.000000	0.996750	3.210000	0.620000	10.200000	6.000000
75%	9.200000	0.560000	0.420000	2.600000	0.060000	21.000000	42.000000	0.997035	3.400000	0.730000	11.100000	6.000000
max	15.000000	1.000000	1.000000	15.000000	0.610000	72.000000	289.000000	100.369000	4.010000	2.000000	14.900000	8.000000

Vinho Branco												
	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	ph	sulphates	alcohol	quality
count	4086.000000	4086.000000	4086.000000	4086.000000	4086.000000	4086.000000	4086.000000	4086.000000	4086.000000	4086.000000	4086.000000	4086.000000
mean	6.854708	0.270241	0.334192	6.391415	0.045772	35.308085	138.360657	0.994027	3.180267	0.488647	10.514267	5.677909
std	0.843368	0.180795	0.121020	5.072058	0.021848	17.007137	42.488065	0.002991	0.151001	0.114126	1.238621	0.006329
min	3.000000	0.000000	0.000000	0.000000	0.000000	2.000000	9.000000	0.907110	2.720000	0.220000	8.000000	3.000000
25%	6.300000	0.210000	0.270000	1.700000	0.036000	23.000000	0.991723	3.090000	0.410000	0.500000	9.500000	5.000000
50%	6.300000	0.260000	0.320000	5.200000	0.042000	34.000000	0.993740	3.100000	0.470000	0.640000	10.400000	6.000000
75%	7.300000	0.320000	0.380000	9.900000	0.050000	46.000000	107.000000	0.996100	3.200000	0.550000	11.400000	6.000000
max	14.200000	1.100000	1.600000	85.000000	0.340000	289.000000	440.000000	1.039590	3.620000	1.000000	14.200000	9.000000

Fig. 1: Resumo estatístico das variáveis por tipo de vinho

B. Análise Monovariada Incondicional

Sabendo que essa análise serve para estudar estatisticamente cada preditor separadamente, temos abaixo histogramas de freqüência, gráficos de assimetria e o resultados dos valores de média, desvio padrão e assimetria.

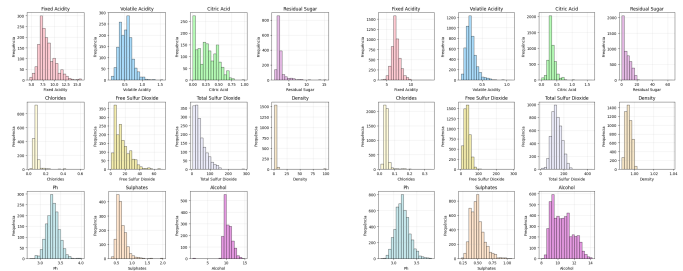
Ao analisarmos os resultados da média, desvio padrão e assimetria para os preditores dos vinhos tinto e branco, podemos fazer as seguintes observações:

- Média:** A média da Acidex Fixa e da Densidade são consideravelmente maiores no vinhos tinto, enquanto nos vinhos branco os maiores são Açúcar Residual, Dióxido de Enxofre Livre e Dióxido de Enxofre Total.
- Desvio Padrão:** Podemos observar, para ambos os tipos de vinho, que o Dióxido de Enxofre Livre e o Dióxido de Enxofre Total possuem um valor muito alto, indicando que os valores estão bem espalhados ao redor da média, ou seja, existe uma grande variedade de valores desses preditores.
- Assimetria:** Podemos observar que o Cloreto, tanto no tinto quanto no branco, é o preditor com maior assimetria, ou seja, existem alguns vinhos que possuem quantidade de cloreto muito alto comparado com os demais. Além disso, o desvio da densidade do vinho tinto é 10 vezes maior que a do branco, o que quer dizer que alguns vinhos tintos possuem valores de densidade extremamente alto em comparação com o restantes dos vinhos. Essa discrepância na densidade podemos observar na Fig.1 onde sua média do vinho tinto é 2.208702 e o valor máximo é 100.369000, ou seja, existem valores quase 50 vezes maiores que a média que causam o aumento do valor da assimetria.

Vinho Tinto			
Preditor	Média (μ)	Desvio Padrão (σ)	Assimetria (γ)
fixed acidity	0.3196	1.7411	0.9828
volatile acidity	0.5278	0.1791	0.6716
citric acid	0.2710	0.1948	0.3183
residual sugar	2.5388	1.4099	4.5407
chlorides	0.0875	0.0471	5.6803
free sulfur dioxide	15.8749	10.4602	1.2506
total sulfur dioxide	46.4678	32.8953	1.5155
density	2.2087	9.6641	9.8039
pH	3.3111	0.1544	0.1937
sulphates	0.6581	0.1695	2.4287
alcohol	10.3978	1.1599	-0.5768

Vinho Branco			
Preditor	Média (μ)	Desvio Padrão (σ)	Assimetria (γ)
fixed acidity	6.8548	0.8439	0.6478
volatile acidity	0.2782	0.1808	1.5770
citric acid	0.3342	0.1210	1.2819
residual sugar	6.3914	5.0721	1.0771
chlorides	0.0458	0.0218	5.0233
free sulfur dioxide	35.3081	17.0071	1.4067
total sulfur dioxide	138.3607	42.4981	0.3907
density	0.9940	0.0030	0.9778
pH	3.1883	0.1510	0.4578
sulphates	0.4898	0.1141	0.9772
alcohol	10.5143	1.2306	0.4873

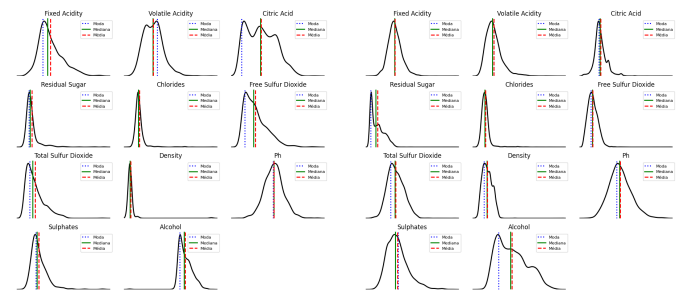
Fig. 2: Média, desvio padrão e assimetria das variáveis



Vinho Tinto

Vinho Branco

Fig. 3: Histograma de freqüência dos preditores por tipo de vinho



Vinho Tinto

Vinho Branco

Fig. 4: Assimetria das variáveis por tipo de vinho

IV. REFERÊNCIAS

- [1] P. Cortez, A. Cerdeira, F. Almeida, T. Matos, and J. Reis, "Modeling wine preferences by data mining from physicochemical properties," *Decision Support Systems*, vol. 47, no. 4, pp. 547–553, 2009.
- [2] Caveroyale, "Ácido Cítrico: Importância e Aplicações em Vinhos Premium," [Online]. Available: <https://www.caveroyale.com.br/glossario/>

acido-citrico-importancia-aplicacoes-vinhos-premium/,
acesso em: 28 set. 2025.

- [3] Caveroyale, “Acidez Volátil: Entenda seu Impacto nos Vinhos Premium,” [Online]. Available: <https://www.caveroyale.com.br/glossario/acidez-volatil-vinhos-premium/>, acesso em: 28 set. 2025.
- [4] Agrovin, “Técnicas para corrigir a acidez do vinho,” [Online]. Available: <https://agrovin.com/pt-pt/tecnicas-para-corrigir-a-acidez-do-vinho/>, acesso em: 28 set. 2025.
- [5] Embrapa, “Metodologia de Análise de Vinho Tinto,” [Online]. Available: <https://www.infoteca.cnptia.embrapa.br/infoteca/bitstream/doc/887323/1/Metodologiaanalisevinhotintoed012010.pdf>, acesso em: 28 set. 2025.
- [6] Famiglia Valduga, “A importância da acidez no vinho,” [Online]. Available: <https://blog.famigliavalduga.com.br/qual-a-importancia-da-acidez-no-vinho/>, acesso em: 28 set. 2025.