

# Análise Exploratória da Qualidade dos Vinhos Tinto com Base em Propriedades Físico-Químicas e Sensoriais

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

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

3<sup>rd</sup> Caio Vinícius Pessoa Freires  
Dept. de TeleInformática  
Universidade Federal do Ceará  
Fortaleza, Brazil  
caiopeessoa145@gmail.com

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

**Abstract**—Este trabalho apresenta uma análise exploratória do conjunto de dados "Wine Quality", focada em vinhos tintos. Foram avaliadas propriedades físico-químicas e sensoriais, considerando seus principais valores estatísticos. A análise inclui quatro abordagens: univariada incondicional, univariada condicional por classe, bivariada incondicional e multivariada incondicional. O objetivo é compreender e identificar relações entre os preditores e a qualidade dos vinhos tintos, fornecendo insights para possíveis modelos preditivos.

**Index Terms**—análise exploratória de dados, vinhos tintos, 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. Para esta análise, serão considerados apenas os dados referentes aos vinhos **tintos**, uma vez que pesquisas indicam que eles são os mais consumidos e preferidos pelos brasileiros. [2]

As 11 variáveis de entrada são:

- 1) **Acidez fixa (g/L)**: ácidos naturais predominantes, influenciam frescor e aroma [3].
- 2) **Acidez volátil (g/L)**: ácidos que evaporam facilmente, impactam sabor e aroma [4].

- 3) **Ácido cítrico (g/L)**: presente em menor quantidade, equilibra acidez [3],[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  $\mu_d$ :

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

- 3) Cálculo do desvio padrão populacional  $\sigma_d$ :

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

- 4) Cálculo da assimetria (skewness)  $\gamma_d$ :

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

### III. RESULTADOS

#### A. Descrição do Dataset

A Tabela 1 apresenta um resumo estatístico das variáveis do dataset de vinhos tintos. Observa-se que as médias de atributos como ácido fixo, açúcar residual, dióxido de enxofre livre e dióxido de enxofre total possuem variações relevantes. Esses valores indicam diferenças importantes nas características físico-químicas dos vinhos, que podem influenciar diretamente a avaliação de sua qualidade. Por exemplo, a média da qualidade dos vinhos tintos é 5,87, com valor máximo observado de 8,0, mostrando que há uma faixa relativamente limitada de variação na avaliação em comparação a outros tipos de vinho.

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	ph	sulphates	alcohol	quality
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000
mean	8.3196	0.5278	0.2710	2.5388	0.0875	15.8749	46.4678	2.2087	3.3111	0.8039	10.3978	5.8362
std	1.7411	0.1791	0.1948	1.4099	0.0471	10.4602	32.8953	0.1544	0.1937	0.1085	1.1082	0.8076
min	4.0000	0.1200	0.0000	0.0000	0.0100	1.0000	6.0000	0.9907	2.7400	0.3300	1.0000	3.0000
25%	7.1000	0.3800	0.0000	1.0000	0.0700	7.0000	22.0000	0.9950	3.2100	0.5500	9.0000	5.0000
50%	7.9000	0.5200	0.2000	2.2000	0.0700	14.0000	38.0000	0.9975	3.3100	0.6200	10.2000	6.0000
75%	9.2000	0.6400	0.2500	2.6000	0.0900	21.0000	42.0000	0.9975	3.4000	0.7500	11.0000	6.5000
max	15.0000	1.5000	1.0000	15.0000	0.1100	72.0000	220.0000	1.0000	4.0100	2.0000	14.0000	8.0000

Fig. 1: Resumo estatístico das variáveis do Vinho Tinto

#### B. Análise Monovariada Incondicional

A análise monovariada incondicional permite estudar estatisticamente cada preditor individualmente. Para os vinhos tintos, observamos os histogramas de frequência, gráficos de assimetria e os valores de média, desvio padrão e assimetria apresentados na Tabela 1.

A partir desses resultados, podemos destacar:

- 1) **Média:** Os preditores com maiores valores médios nos vinhos tintos são o *Ácido Fixo* e o *Álcool*, indicando que, em geral, esses atributos apresentam níveis mais altos nos vinhos tintos.
- 2) **Desvio Padrão:** O *Dióxido de Enxofre Livre* e o *Dióxido de Enxofre Total* apresentam desvio padrão elevado, o que indica grande variabilidade entre os vinhos tintos para esses preditores.
- 3) **Assimetria:** O preditor *Sulfatos* apresenta a maior assimetria, sugerindo a presença de alguns vinhos com valores de cloreto significativamente maiores que a maioria. Além disso, a *Densidade* apresenta um valor de assimetria elevado, indicando que há vinhos tintos com densidade muito acima da média, o que contribui para a dispersão dos dados.

Preditor	Média ( $\mu$ )	Desvio Padrão ( $\sigma$ )	Assimetria ( $\gamma$ )
fixed acidity	8.3196	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	4.5467
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	0.1544	9.8039
pH	3.3111	0.1937	0.1937
sulphates	0.6581	0.1695	2.4287
alcohol	10.3978	1.1599	-0.5768

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

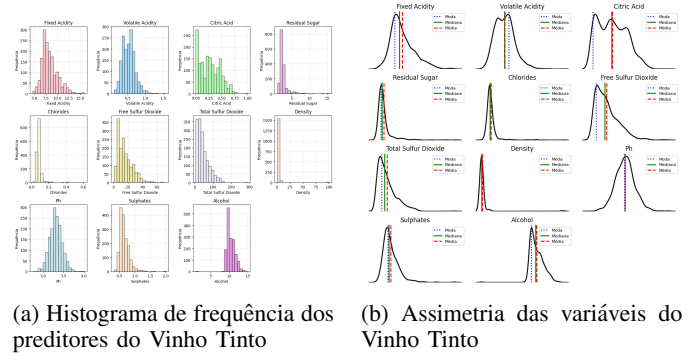


Fig. 3: Análise gráfica do Vinho Tinto

A análise dos vinhos tintos considerou tanto os histogramas de frequência quanto os gráficos de assimetria das variáveis físico-químicas, permitindo uma avaliação mais detalhada das distribuições e do comportamento dos dados.

**Acidez Fixa:** Os histogramas indicam que a acidez fixa está concentrada entre 6 e 9, com leve assimetria à direita, mostrada pelos gráficos de assimetria. Isso evidencia que a maioria dos vinhos possui acidez moderada, com poucos casos de valores mais elevados.

**Acidez Volátil:** A distribuição concentra-se entre 0,3 e 0,7, com assimetria à direita observada tanto nos histogramas quanto nos gráficos de assimetria. Valores muito altos são raros, indicando que altos níveis de acidez volátil não são comuns em vinhos tinto.

**Ácido Cítrico:** A distribuição apresenta muitos valores próximos a zero, e a assimetria reforça que alguns vinhos possuem ácido cítrico significativamente mais alto, embora não sejam a maioria.

**Açúcar Residual:** O histograma mostra forte concentração em valores muito baixos, entre 0 e 2, caracterizando vinhos predominantemente secos (até 4g de açúcar por litro, na legislação brasileira) [8]. Os gráficos de assimetria confirmam que há poucos vinhos tinto com açúcar residual mais elevado.

**Cloretos:** Distribuição extremamente concentrada e próxima a zero, com baixa assimetria, indica baixo teor de cloretos na maioria dos vinhos.

**Dióxido de Enxofre Livre:** Os histogramas e gráficos de assimetria indicam distribuição à direita, com a maioria dos vinhos até 20 mg/L e poucos valores elevados.

**Dióxido de Enxofre Total:** Apresenta padrão semelhante ao enxofre livre, mas com maior variabilidade; gráfico de assimetria confirma a tendência de valores extremos mais raros.

**Densidade:** Distribuição altamente concentrada em valores próximos de zero, com assimetria praticamente nula, evidenciando pequenas diferenças entre os vinhos.

**pH:** Distribuição aproximadamente normal, centrada entre 3 e 3,5, com gráficos de assimetria, indicando simetria da variável, ou seja, a maioria dos dados está concentrada próximo a média.

**Sulfatos:** Distribuição assimétrica à direita, concentrada entre 0,4 e 0,8. Alguns vinhos apresentam valores maiores, como evidenciado pela cauda à direita nos gráficos de assimetria.

**Álcool:** Distribuição levemente assimétrica à direita, concentrando-se entre 9 e 12%, indicando teor alcoólico moderado na maioria das amostras.

**Conclusão:** A análise conjunta dos histogramas e dos gráficos de assimetria mostra que os vinhos tintos da amostra são predominantemente secos, apresentando baixos teores de acidez volátil, cloretos e dióxido de enxofre, além de um teor alcoólico moderado. A assimetria à direita observada na maioria das variáveis indica a presença de alguns vinhos com valores atípicos, ainda que a maior parte das amostras se mantenha dentro de um perfil físico-químico consistente. Essa visão integrada permite compreender com mais clareza as características desses vinhos, sendo útil tanto para análises comparativas quanto para orientar decisões relacionadas à produção.

#### 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] PORTAL INSIGHTS. Qual o vinho preferido dos brasileiros? Disponível em: <https://www.portalinsights.com.br/perguntas-frequentes/qual-o-vinho-preferido-dos-brasileiros/>. Acesso em: 19 out. 2025.
- [3] 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.
- [4] 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.
- [5] 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.
- [6] Embrapa, "Metodologia de Análise de Vinho Tinto," [Online]. Available: <https://www.infoteca.cnptia.embrapa.br/infoteca/bitstream/doc/887323/1/Methodologiaanalisevinhotintoed012010.pdf>, acesso em: 28 set. 2025.
- [7] Família 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.